REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
DECEMBER 2022	TECHNICAL PAPER	DECEMBER 2020	MAY 2022

4. TITLE AND SUBTITLE

POWER LAW MODELS FOR CONNECTIVITY OF RANDOM GEOMETRIC GRAPHS

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
IN-HOUSE	N/A	62788F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
		R1LP

6. AUTHOR(S)

Warren H. Debany Jr., Mark H. Linderman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIG
525 Brooks Road
Rome NY 13441-4505

8.PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10.SPONSOR/MONITOR'S ACRONYM(S)	11.SPONSOR/MONITOR'SREPORT NUMBER(S)
Air Force Research Laboratory/RIG 525 Brooks Road Rome NY 13441-4505	AFRL/RI	AFRL-RI-RS-TP-2022-021

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. PA # AFRL-2022-2503
Date Cleared: 26 MAY 2022

13. SUPPLEMENTARY NOTES

This is a work of the United States Government and is not subject to copyright protection in the United States.

14. ABSTRACT

This paper addresses the connectivity properties of ad hoc wireless networks. These networks do not rely on a fixed infrastructure to make up for the limited communication range of nodes, but instead achieve connectivity through cooperative, peer-to-peer, multi-hop routing. Ad hoc wireless networks can be used for intra-team communications by military personnel or first responders, ground-based vehicle networks, airborne networks, and even provide device-to-device communication without cell towers. Networks composed of nodes with fixed communication range can be modeled as Random Geometric Graphs (RGGs). Modeling the connectivity properties of these networks is critical to evaluating their performance as well as their vulnerability to failure or attack. Knowledge of these properties can also improve survivability by ensuring that the network is not fully connected, such as to mitigate the transmission of disease or malware. RGGs are modeled in flat and toroidal square spaces as well as in circles and rectangular spaces of various aspect ratios. They are modeled in spaces without obstructions, in spaces with regularly and randomly placed obstructions, and in lattices. RGGs are modeled largely in two-dimensional spaces but also in spaces of dimensions one through six. Two RGG growth models are considered, where a given connectivity goal is achieved either by adding nodes with fixed communication range or by finding the minimum necessary range of a fixed set of nodes. Regression analysis is applied to the outputs of simulations of RGG growth to obtain expressions that predict various statistical metrics for full connectivity, partial connectivity, and no isolated nodes. It is shown that power law (log-log) models yield accurate statistical metrics for these forms of connectivity for both growth models, although log law models performed slightly better for square lattices. The accuracies of inverse regression and reverse regression were tested.

15. SUBJECT TERMS

Random Geometric Graphs, Ad Hoc Networks, Multi-Hop Routing, Communication Link Obstructions, Power Law Models, Log Law Models, Linear Regression Models

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
U	U	U	SAR	438

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
WARREN DEBANY, JR.	**N/A**

PREVIOUS EDITION IS OBSOLETE.

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Power Law Models for Connectivity of Random Geometric Graphs

Warren H. Debany Jr., PhD
Volunteer Emeritus Corps (VEC)
Information Warfare Division
Air Force Research Laboratory (AFRL/RIG)
525 Brooks Rd.
Rome NY 13441-4505

Mark H. Linderman, PhD
Chief Scientist
Information Directorate
Air Force Research Laboratory (AFRL/RI)
525 Brooks Rd.
Rome NY 13441-4505

Abstract

This paper addresses the connectivity properties of ad hoc wireless networks. These networks do not rely on a fixed infrastructure to make up for the limited communication range of nodes but instead achieve connectivity through cooperative, peer-to-peer, multi-hop routing. Ad hoc wireless networks can be used for intra-team communications by military personnel or first responders, ground-based vehicle networks, airborne networks, and even provide device-to-device communication without cell towers. These networks can also describe the connectivity of some biological systems.

Approved for Public Release; Distribution Unlimited

Networks comprised of nodes with fixed communication range can be modeled as Random Geometric Graphs (RGGs). Modeling the connectivity properties of these networks is critical to evaluating their performance as well as their vulnerability to failure or attack. Knowledge of their properties can also be used to engineer their survivability by ensuring that the network is not fully connected, such as to mitigate the transmission of disease or malware.

In this paper, RGGs are modeled in flat and toroidal square spaces as well as in circles and rectangular spaces of various aspect ratios. They are modeled in spaces without obstructions, in spaces with regularly and randomly placed obstructions, and in lattices. RGGs are modeled largely in two-dimensional spaces but also in spaces of dimensions one through six.

Two RGG growth models are considered, where a given connectivity goal is achieved by either adding a number of nodes n with fixed communication range r (n-for-given-r growth) or finding the minimum necessary range of a fixed set of nodes (r-for-given-n growth).

Regression analysis is applied to the outputs of simulations of RGG growth to obtain expressions that predict various statistical metrics for several connectivity goals including full connectivity, partial connectivity, and no isolated nodes. It is shown that power law models yield accurate statistical metrics for these connectivity goals for both growth models, although a log law model performed slightly better for square lattices.

The accuracies of inverse regression and reverse regression for the two growth models were tested.

The connectivity properties of RGGs have been studied by many researchers. Notably, M.D. Penrose developed a broadly general theory for the connectivity of RGGs. This paper tests the results of Penrose and others for a variety of related problems.

Contents

1 Introduction to Random Geometric Graph (RGG) Connectivity **14**
1.1 Connectivity 16
1.2 Nodal Spaces 21
1.3 Border Effects: Flat vs. Toroidal Spaces 24
1.4 Distance in Flat and Toroidal Spaces 27
1.5 RGG Growth Models: n-for-given-r and r-for-given-n 30
1.6 Connectivity 31
1.7 Configurations and Geometries 33

2 Nomenclature **35**

3 Neighbors, Degree, Isolated Nodes, and Components **38**
3.1 Examples of RGGs 39
3.2 Examples of n-for-given-r Growth 47
3.3 Examples of r-for-given-n Growth 56
3.4 Estimating the Probability of No Isolated Nodes in Two-Dimensional Spaces 58
3.5 Penrose's Theorem 61
3.6 Gupta and Kumar's Theorem 69
3.7 Consideration of Border Effects 73
3.8 Koskinen's Regression Analyses 74

4 RGGs that Grow by Adding Nodes with Fixed Range: The n-for-given-r Growth Model **75**
4.1 Simulations and Data Analysis 76
4.1.1 Simulators 77
4.1.2 Simulations 78
4.1.3 Adding Nodes Can Break Full Connectivity of RGGs with n-for-given-r Growth 82
4.1.4 Sample Size for Randomly Generated Geometries . . . 86
4.1.5 Linear Regression Models, Goodness-of-Fit, and Errors 87

4.2 Flat Unit Square with n-for-given-r Growth 90
4.2.1 Flat Unit Square with n-for-given-r Growth for Mid Ranges 93
4.2.1.1 Flat Unit Square: Full Connectivity for Mid-r Set 94
4.2.1.2 Flat Unit Square: 99% Connectivity for Mid-r Set 97
4.2.1.3 Flat Unit Square: 50% Connectivity for Mid-r Set 99
4.2.1.4 Flat Unit Square: No Isolated Nodes for Mid-r Set 103
4.2.1.5 Flat Unit Square: Number of Neighbors Needed for Full Connectivity for Mid-r Set . 106
4.2.1.6 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set 108
4.2.1.7 Flat Unit Square: Testing the Linear Regression Models for Mid-r Set 114
4.2.2 Flat Unit Square: Full Connectivity for Large Ranges 119
4.2.3 Flat Unit Square with n-for-given-r Growth for Small Ranges 125
4.2.3.1 Flat Unit Square: Full Connectivity for Small-r Set 126
4.2.3.2 Flat Unit Square: No Isolated Nodes for Small-r Set 128
4.2.3.3 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-r Set 131

4.2.4 Flat Unit Square with n-for-given-r Growth for Mid and Small Ranges 136
4.2.4.1 Flat Unit Square: Full Connectivity for Mid+Small-r Set 137
4.2.4.2 Flat Unit Square: No Isolated Nodes for Mid+Small-r Set 140
4.2.4.3 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid+Small-r Set 143
4.2.5 Flat Unit Square: Comparing Linear Regression Models from Mid-r Set, Small-r Set, and Mid+Small-r Set 149

4.3 Toroidal Unit Square with n-for-given-r Growth 150
4.3.1 Toroidal Unit Square with n-for-given-r Growth for Mid Ranges 152
4.3.1.1 Toroidal Unit Square: Full Connectivity for Mid-r Set 153
4.3.1.2 Toroidal Unit Square: 99% Connectivity for Mid-r Set 156
4.3.1.3 Toroidal Unit Square: No Isolated Nodes for Mid-r Set 159
4.3.1.4 Toroidal Unit Square: Number of Neighbors Needed for Full Connectivity for Mid-r Set . 164
4.3.1.5 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set 166
4.3.2 Toroidal Unit Square with n-for-given-r Growth for Small Ranges 172
4.3.2.1 Toroidal Unit Square: Full Connectivity for Small-r Set 173
4.3.2.2 Toroidal Unit Square: No Isolated Nodes for Small-r Set 176
4.3.2.3 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-r Set 179
4.3.3 Toroidal Unit Square with n-for-given-r Growth for Mid and Small Ranges 185
4.3.3.1 Toroidal Unit Square: Full Connectivity for Mid+Small-r Set 186
4.3.3.2 Toroidal Unit Square: No Isolated Nodes for Mid+Small-r Set 189
4.3.3.3 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid+Small-r Set 192
4.3.4 Toroidal Unit Square: Comparing Linear Regression Models from Mid-r Set, Small-r Set, and Mid+Small-r Set 198

4.4 Unit Circle with n-for-given-r Growth for Mid Ranges 199
4.4.1 Unit Circle: Full Connectivity for Mid-r Set 202
4.4.2 Unit Circle: 99% Connectivity for Mid-r Set 205
4.4.3 Unit Circle: No Isolated Nodes for Mid-r Set 208
4.4.4 Unit Circle: Number of Neighbors Needed for Full Connectivity for Mid-r Set 213
4.4.5 Unit Circle: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set 215

5 RGGs that Grow by Increasing the Range of a Fixed Number of Nodes: The r-for-given-n Growth Model 220
5.1 Simulations and Data Analysis 221
5.2 Flat Unit Square with r-for-given-n Growth 223
5.2.1 Flat Unit Square with r-for-given-n Growth for Small Numbers of Nodes 224
5.2.1.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Small-n Set 225
5.2.1.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-n Set 228
5.2.2 Flat Unit Square with r-for-given-n Growth for Large Numbers of Nodes 230
5.2.2.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Large-n Set 231
5.2.2.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Large-n Set 234
5.2.3 Flat Unit Square with r-for-given-n Growth for Small and Large Numbers of Nodes 236
5.2.3.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Small+Large-n Set 237
5.2.3.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small+Large-n Set 240
5.2.4 Flat Unit Square: Comparing Linear Regression Models from Small-n Set, Large-n Set, and Small+Large-n Set 246

5.3 Toroidal Unit Square with r-for-given-n Growth 247
5.3.1 Toroidal Unit Square with r-for-given-n Growth for Small Numbers of Nodes 248
5.3.1.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Small-n Set 249
5.3.1.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-n Set 252
5.3.2 Toroidal Unit Square with r-for-given-n Growth for Large Numbers of Nodes 255
5.3.2.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Large-n Set 256
5.3.2.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Large-n Set 259
5.3.3 Toroidal Unit Square with r-for-given-n Growth for Small and Large Numbers of Nodes 262
5.3.3.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Small+Large-n Set . . 263
5.3.3.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small+Large-n Set 266
5.3.4 Toroidal Unit Square: Comparing Linear Regression Models from Small-n Set, Large-n Set, and Small+Large-n Set 273

5.4 Koskinen's Regression Analyses for RGGs in a Flat Unit Square with r-for-given-n Growth 274
5.4.1 Testing Koskinen's Estimate of the Range for a Given Number of Nodes 275
5.4.2 Testing Koskinen's Estimate of Penrose's c_P as a Function of the Number of Nodes 283
5.5 Maximum Range Required for No Isolated Nodes in Flat Unit Squares and Unit Circles with r-for-given-n Growth . . 286

6 Inverting n-for-given-r Linear Regression Models to Estimate r-for-given-n Growth **291**
6.1 Flat Unit Square: Inverse Regression for Full Connectivity . . 293
6.2 Flat Unit Square: Inverse Regression for No Isolated Nodes . 295
6.3 Toroidal Unit Square: Inverse Regression for Full Connectivity 297
6.4 Toroidal Unit Square: Inverse Regression for No Isolated Nodes 299
6.5 Observations on Inverse Regression and Reverse Regression to Estimate r-for-given-n Growth 301

7 RGGs in d-Dimensional Spaces **303**
7.1 RGGs in d-Dimensional Spaces with n-for-given-r Growth . . 304
7.1.1 One-Dimensional Spaces 305
7.1.2 Two-Dimensional Spaces 308
7.1.3 Three-Dimensional Spaces 312
7.1.4 Four-Dimensional Spaces 315
7.1.5 Comparing Number of Nodes Required in d-Dimensional Spaces 318
7.2 RGGs in d-Dimensional Spaces with r-for-given-n Growth . . 321
7.2.1 Estimating the Probability of No Isolated Nodes in d-Dimensional Spaces 322
7.2.2 Penrose's Theorem in d-Dimensional Spaces 325
7.2.3 Comparing Full Connectivity, No Isolated Nodes, and Penrose's Theorem in d-Dimensional Spaces 327
7.2.4 One-Dimensional Spaces 329
7.2.5 Two-Dimensional Spaces 336
7.2.6 Three-Dimensional Spaces 342
7.2.7 Four-Dimensional Spaces 348
7.2.8 Five-Dimensional Spaces 354
7.2.9 Six-Dimensional Spaces 359
7.2.10 Comparing Ranges Required in d-Dimensional Spaces 364

8 RGGs in Flat Rectangular Spaces with n-for-given-r Growth **370**
8.1 Unit Rectangle of Width 2 . . . 372
8.2 Unit Rectangle of Width 3 . . . 373
8.3 Unit Rectangle of Width 4 . . . 374
8.4 Unit Rectangle of Width 5 . . . 375
8.5 Comparing Connectivity of Flat Unit Rectangles of Varying Widths to Flat Unit Squares . . . 376

9 RGGs in Spaces with Obstructions and n-for-given-r Growth **380**
9.1 Flat Unit Square with Obstructions . . . 382
9.1.1 Regularly Placed Obstructions: Full Connectivity and No Isolated Nodes . . . 385
9.1.2 Regularly Placed Obstructions: Full Connectivity for Large Ranges . . . 391
9.1.3 Regularly Placed Obstructions: 99% Connectivity . . . 394
9.1.4 Randomly Placed Obstructions: 99% Connectivity . . . 398
9.1.5 Randomly Placed Obstructions: 99% Connectivity for Large Ranges . . . 402
9.1.6 Comparing Connectivity in Spaces with Regularly and Randomly Placed Obstructions . . . 405
9.2 Flat and Toroidal Lattice . . . 407
9.2.1 Flat Lattice: Full Connectivity . . . 409
9.2.2 Toroidal Lattice: Full Connectivity and No Isolated Nodes . . . 416
9.2.3 Toroidal Lattice: Estimating the Probability of No Isolated Nodes . . . 423

10 Conclusions **425**

Acknowledgment **431**

References **431**

1 Introduction to Random Geometric Graph (RGG) Connectivity

This work is based on joint research performed in 2004–2008 with Dr. Jon M. Kleinberg under the former AFRL/Cornell University Information Assurance Institute (IAI) supported by AFOSR Grant F49620-00-1-209.

People and things are increasingly connected through wireless networks. While mainstream wireless architectures have previously required devices to connect to a fixed infrastructure through base stations such as cell towers and hubs, technologies are increasingly able to support *ad hoc* wireless networks.

Ad hoc networks are comprised of nodes that communicate cooperatively with each other in peer-to-peer, multi-hop fashion. They require no fixed infrastructure and thus can operate in disadvantaged areas or unplanned scenarios. Ad hoc networks can support intra-team communications by military personnel or first responders. They can be rapidly deployed by intent or simply take advantage of a cooperative community of nodes. Mobile Ad Hoc Networks (MANETs) involve nodes that move, and Vehicular Ad Hoc Networks (VANETs) are MANETs that specifically support the mobility created by being hosted on vehicles.

Ad hoc networks can support cooperative constellations of sensors (often called "sensor nets") or devices serving as part of an Internet of Things (IoT). Cellular phones can be operated directly phone-to-phone in ad hoc mode to provide multi-hop networks without cell towers; for example, see [Asadi (2014)] [Albrecht (2021)].

This paper explores the conditions under which ad hoc networks have sufficient numbers of nodes or sufficient nodal communication range, depending on how they grow, to achieve desired conditions of connectivity with a required level of confidence. This paper does not seek to develop new theoretical approaches but rather to obtain empirical expressions that

provide the "ground truth" that would guide and validate future advances in this field of study.

For the convenience of the reader, entries in the Table of Contents are hyperlinked in the PDF document as are the labels in the text for sections, figures, expressions/equations, footnotes, and references/citations.

1.1 Connectivity

Ad hoc networks require the ability to route data from sources to their destinations but this paper focuses only on the essential ability of the nodes to communicate with each other by means of their wireless links, in short, the wireless aspect of their connectivity. The routing aspect of network connectivity is considered here only in terms of the *existence* of at least one path from any node to any other node and not on the details of the routes or algorithms that would find those routes.

It is assumed in this paper that n nodes are scattered uniformly at random positions according to a homogeneous Poisson point process [Penrose (1997)] [Bettstetter (2002b)] in a defined *space*. This is referred to in this paper as *Poisson distribution of nodes*.

Movement of nodes is not considered in this paper. For a network composed of nodes that are in motion the results obtained in this paper can considered to be based on a snapshot of the connectivity of the MANET at an instant in time.

It is also assumed in this paper that all nodes have the same bidirectional *nodal communication range* r (usually called simply the *range*). It is assumed that range is omnidirectional in all of the dimensions of the nodal space.

This paper does not deal with issues of interference, channel selection, or throughput; for example, see [Asadi (2014)] [Bettstetter (2005)] [Bicket (2005)]. Most wireless networks are based on Radio Frequency (RF) techniques but nothing in this paper precludes links based on optical, acoustic, or other means to the extent they fit the assumptions of this paper.[1]

In this model, two nodes are *connected* if and only if they are within the given range r of each other and that connection is called a *link*. Links are *line-of-sight*, meaning that any obstruction between two nodes prevents direct communication between those nodes.

[1] For example, disease transmission (discussed later in this paper) involves the transfer of pathogens or other infectious agents or contaminants. Such transfer may be highly likely if a susceptible organism is within a certain distance of an infected organism but highly unlikely at a greater distance.

Connected nodes are *neighbors* of each other and the number of neighbors of a node is the *degree* of the node.

It is also convenient to define a weaker relationship called *near neighbors*. These are nodes that are in close proximity[2] to each other. They may be within range of each other (and are thus neighbors) or just outside range of each other.

This type of graph is called a *Random Geometric Graph (RGG)*.[3] An RGG contains one or more nodes and zero or more links. Thus, a single node or a set of nodes with no links is still an RGG.

A set of nodes forms a *component*[4] if and only if, for every pair of nodes in the component, there is a *path* between them, meaning that the nodes are either connected to each other or have a chain of connected nodes between them. Components are equivalence classes in the sense that every node is part of one and only one component. An *isolated node* is the only node in its component.

The largest component in an RGG is called the *giant component (GC)*. In the early stages of RGG growth the GC may not be unique because more than one component may have the same largest size. The maximum fraction of nodes in any component is the *connectivity* of the RGG and is denoted by S.

[2] "Close proximity" is an approximate term meaning the nodes are within a small number of multiples of r of each other.

[3] In a generalization of the RGG model, M.D. Penrose explored the properties of *soft RGGs* where the probability of two nodes being connected is given by a distribution function based on their displacement from each other [Penrose (2016)]. This concept is distinguished from the *hard RGGs* studied in most work on this subject, as well as this paper, where two nodes are connected if and only if they are separated by a distance less than or equal to the hard threshold of r.

[4] Both "component" and "cluster" are commonly used terms for a set of connected nodes. "Component" is used here instead of "cluster" to avoid the impression that all nodes in a component must be in close physical proximity to each other. Nodes contained in a single component may be quite distant from each other in the space, whereas the term "cluster" might imply that they are in close physical proximity.

The condition where a single GC contains all of the nodes in the RGG (that is, $S = 1$) is *full connectivity* or *100% connectivity* and the RGG is said to be *fully connected*. Under this condition there is a path from every node to every other node in the RGG.

If an RGG is not fully connected, then it is *disconnected*. While the condition of full connectivity is the usual goal in the RGG literature, *less-than-full* or *partial connectivity* may be sufficient or even necessary. This may occur under circumstances due to power control or node/link failures; for example, see [Santi (2003)] [Ta (2009)]. In Section 9 it is shown that when randomly placed obstructions interfere with line-of-sight communication it may be impossible or impractical to achieve 100% connectivity, and so it may be necessary to settle for less-than-full connectivity.

For these reasons, models are also obtained for the probability of achieving 99% connectivity ($S \geq 0.99$), meaning the probability that the GC contains at least 99% of the nodes in the RGG. In addition, models are obtained for the probability of achieving the condition of 50% connectivity ($S \geq 0.50$) as a reasonable choice of threshold for the emergence of the GC as discussed in Section 4.2.1.3.

Another form of less-than-full connectivity is where an RGG contains *no isolated nodes*. Conditions under which RGGs contain no isolated nodes have been studied extensively as to how they relate to those for full connectivity.

The literature uses terms such as "with high probability," "the probability tends toward 1," "almost surely," and "with probability 1" to describe the likelihood that an RGG is connected.[5] Following the example of Bettstetter [Bettstetter (2004)], this paper instead uses the specific criterion that RGGs achieve a given connectivity condition with probability or confidence of at least $C = 0.99$ in place of these less definite phrases. This paper also determines the conditions under which an RGG is connected with probability of at most $C = 0.01$ when it is desirable that RGGs be disconnected.

[5] The goal of full connectivity with probability that does not just approach 1 but must achieve exactly 1 is discussed in Section 3.6.

In most cases considered in this paper the probability of RGGs being connected is estimated using percentiles of results based of a *sample size* of $N = 1{,}000$ random (or Monte Carlo) simulations per configuration. Sets of simulations with larger and smaller sample sizes are also used in a few situations.

Connectivity goals are usually stated here as achieving a connectivity condition (given as the fraction or percentage of nodes S in the GC or that there are no isolated nodes) along with the probability of achieving that condition. For example, a connectivity goal may be stated as "100% connectivity (or full connectivity) with probability $C = 0.99$."

While this paper considers the properties of RGGs abstractly the discussions are largely in the context of communications networks. RGGs have applicability to other problems, however, as they have been proposed to model reservoir computing [Aven (2020)] and model data set relationships in machine learning [Avrachenkov (2020)] [Erba (2020)].

Most of the work described in this paper is based on two-dimensional or 2D nodal spaces (meaning the dimension d of the space is two) but spaces of other dimensionalities are modeled as well. It is assumed that $d = 2$ unless otherwise specified.

The one-dimensional or 1D case applies in situations where placement constrained to be linear, as may be the case with VANETs.

The two-dimensional case applies to unconstrained placement of nodes on a surface except where node placement and links are explicitly constrained by obstructions.[6]

The three-dimensional or 3D case clearly applies to airborne nodes.

[6] Curvature of the surface, such as the curvature of the earth, is not considered here. It is assumed that the range of the nodes is small enough that plane geometry is accurate enough to model spatial relationships and it is not necessary to resort to spherical geometry.

Higher dimensions are considered in this paper primarily for comparison to results in the RGG literature but, as noted above, the spatial relationships that define RGGs have proposed to represent other abstract relationships such machine learning with data sets in very high dimensionality spaces [Avrachenkov (2020)] [Erba (2020)].

All distances in this paper are Euclidean or L^2 norm, calculated using the Pythagorean Theorem, rather than "Manhattan" distances or L^1 norm.[7] The calculation of these distances is addressed in Section 1.4 of the Introduction.

[7] Author Debany expresses regret that this statement appeared incorrectly in an earlier paper due to an editing error. See [Debany (2021)], footnote #5 on page 7.

1.2 Nodal Spaces

The connectivity of random graphs has been studied for many years. Early work, such as that by Erdős [Erdős (1959)] [Erdős (1960)] [Erdős (1983)], addressed growth of random graphs where links are randomly created between nodes without regard for physical node positions. The RGGs considered here, however, have very different properties due to the aforementioned constraints of node positions in d-dimensional spaces and their fixed ranges of communication with neighbors.

Because dimensions other than two are considered the general term *space* is used to refer to the place where RGGs are located and nodes are referred to as being *in* a given space rather than on it.

The connectivity of an RGG depends on the specific shape of the space in which the nodes are scattered. Approaches in the RGG literature generally use models where the nodes are scattered in a space that is either a *square* or a *circle*[8] and this paper considers both of these shapes. Here, the square nodal space is the *unit square* of area $A = 1$, so its sides have length $a = 1$. The circular nodal space is the *unit circle* of area $A = 1$, so its radius has length $a = 1/\sqrt{\pi} \approx 0.5642$. In this context, "unit" refers to defining the area of a space to be "unity." As described later in this section the normalized results obtained for unit spaces can be scaled to apply to spaces of any desired size and units of length.

In addition to the unit square and unit circle, this paper also considers the more general case of nodal spaces that are *unit rectangles* that have area $A = 1$. A unit rectangle's width is denoted as a_w and its height is a_h where $a_w \times a_h = 1$.

This paper also considers cases where unit square spaces contain *obstructions* or obstacles that limit the placement of the randomly placed nodes and block line-of-sight communication between nodes even when they are within range of each other and otherwise would be neighbors.

[8] The "circle model" is also called the "disk model" in the RGG literature.

A nodal space consisting of a *lattice* is discussed as well where nodes are randomly placed only at the points of the lattice and nodes can communicate with only those nodes that are in the same row or the same column of the lattice as well as being within a given range parameter ω (described in detail in Section 9.2). A lattice does not have an area associated with it but rather is characterized by its number of points. Only square lattices are considered here.

It is sufficient to use nodal spaces with unit areas because the connectivity properties derived using these models are invariant with respect to the ratio of the range to the side lengths or radius of the space. The connectivity properties derived for a unit square with sides $a = 1$ and nodes with range r are the same as those for a square space with sides a' and nodes with range r' when

$$r = \frac{r'}{a'} \quad . \tag{1}$$

Similarly, the connectivity properties derived for a unit circle with radius $a = 1/\sqrt{\pi} \approx 0.5642$ and nodes with range r are the same as those for a circular space with radius a' and nodes with range r' for which

$$r = \frac{a\, r'}{a'} \quad . \tag{2}$$

The connectivity properties derived for a unit rectangle with sides a_w and a_h (where $a_w \times a_h = 1$) and nodes with range r are the same as those for a rectangular space with sides a'_w and a'_h and nodes with range r' when r, a_w, and a_h are given by

$$\begin{aligned} r &= \frac{r'}{\sqrt{a'_w \times a'_h}} \\ a_w &= \frac{a'_w}{\sqrt{a'_w \times a'_h}} \\ a_h &= \frac{1}{a_w} \quad . \end{aligned} \tag{3}$$

The normalized values r and a do not have units associated with them, while the actual values r' and a' may have units in addition to providing scale. For example, with $a' = 10{,}000$ meters the results for the "unitless" unit square (that is, with $a = 1$) can be transformed using (3) to provide results that apply to a square that is 10,000 meters on a side.

Approaches in the RGG literature vary in their usage of actual and normalized values. Some authors generalize their models by not restricting the area of the spaces to the unit area $A = 1$ but instead use the actual range and side length or radius in their expressions and then simplify their derivations by normalizing the range as described above; for example, see [Bettstetter (2004)] and [Santi (2003)]. One must be alert to these differences and take this adjustment into account when comparing the results of this paper to other results in the RGG literature.

In two-dimensional space the nodes are located at (x, y) coordinates in the space. When two nodes are indexed as i and j they have the positions (x_i, y_i) and (x_j, y_j), respectively. The lower left-hand corner of a square or rectangle is located at the position (0, 0) and so all of the values of x and y for its nodes' coordinates are non-negative. The center of a circle is at the position (0, 0) and so the x and y values can be positive or negative. Extension of these alignments to other dimensions d are obvious.

In d-dimensional spaces the coordinates of a node are given as $(x_1, x_2, x_3, \ldots x_d)$. Nodes indexed as i and j have the positions $(x_{i,1}, x_{i,2}, \ldots x_{i,d})$ and $(x_{j,1}, x_{j,2}, \ldots x_{j,d})$, respectively.

In the generalization to spaces of dimension d, the one-dimensional "unit square" is simply a line of length $a = 1$. "Unit squares" in three-dimensional spaces are unit cubes and "unit squares" in spaces of dimensions $d \geq 4$ are unit hypercubes where all of their sides have length $a = 1$.

The term *area* is used in this paper not only in the customary sense of area in 2D, but is also used as the general term to mean length in a one-dimensional space or volume or hypervolume in dimensions greater than two.

1.3 Border Effects: Flat vs. Toroidal Spaces

Predicting the connectivity of RGGs is complicated by the consideration of *border effects*[9] in the nodal spaces.

To simplify analyses, much of the RGG literature assumes that the nodal space is *toroidal*. This is to say the top and bottom edges wrap around and the right and left edges wrap around to form a torus and thus eliminate the effect of the border. In this paper, a nodal space that is not toroidal is called *flat*.

This paper considers two kinds of unit square models: the *flat unit square* that does not wrap around at the edges and the *toroidal unit square* that wraps around at the edges. A flat unit square clearly models real-world geometries more faithfully than a toroidal unit square.

It is well-known (for example, see [Bettstetter (2002b)]) that the connectivity of an RGG depends on the mean number of neighbors each node has. In a flat space the *border nodes* (those nodes that are within distance r of the border or the edges of the space) have fewer neighbors or near neighbors on average than the non-border nodes.

In contrast to a square space, where the a torus is easily defined and used to simplify analyses, the very concept of a "toroidal circle" is problematic. If the border of a circle "wrapped around," then the border of the circle would effectively map to a single point and thus create a discontinuity. Every node within distance r of the border of the circle would be at least a near neighbor of every other such node, and every node within $r/2$ of the border would in fact be a neighbor of every other such node. Due to this discontinuity the border nodes of a "toroidal circle" would (on average) have more near neighbors than the non-border nodes, rather than fewer. For this reason, wrap-around at the border of a two-dimensional circle creates an uneven distribution of nodes rather than that expected from a homogeneous Poisson point process.

[9] Border effects are also called *boundary effects* in the literature.

Some researchers have proposed modeling a two-dimensional toroidal circle as the surface of a three-dimensional sphere where distances between nodes are measured as surface distances on the sphere; for example, see [Gupta (2000)] [Bettstetter (2004)]. The principles of spherical geometry make it obvious that such a mapping onto a curved space would introduce distortion of geometries relative to those in flat space. The amount of distortion can be roughly gauged using the following numerical examples; for simplicity rounded values are used.

Consider a three-dimensional sphere with unit area $A = 1$, so that its radius is $a = \frac{1}{2\sqrt{\pi}} = 0.2821$; this is defined to be the *unit sphere*. Choose an arbitrary reference point on the surface of the unit sphere, and then place two nodes on the surface that are at surface distance x from the reference point and at a right angle to each other. On a flat surface the distance between those two nodes would be given by the Pythagorean Theorem as $\sqrt{2\,x^2} = \sqrt{2}\,x$.

For $x = 0.1000$, on a flat surface these two nodes would be at distance 0.1414 from each other. On the curved surface of the sphere, however, their surface distance is 0.1399, which is 1.070% less than the flat distance. The discrepancy is smaller for smaller x: for $x = 0.0100$ their surface distance is 0.010% less than the flat distance and for $x = 0.0050$ their surface distance is 0.003% less than the flat distance.

It would be an interesting problem for future study to explore the impact of the local distortions of the surface distances[10] on a sphere on the connectivity of RGGs and whether the sphere would be a good representation of a "toroidal circle" for small ranges.

[10] Surface distance on the unit sphere is used in the examples above but there is another measure of distance between two nodes: the chord distance through the interior of the sphere. Chord distance between nodes on a sphere would introduce even more distortion, however, as this example shows. In a two-dimensional toroidal unit circle (disregarding the issue of the discontinuity at the border) the maximum chord distance between any two nodes would be twice the radius of the circle or $\frac{2}{\sqrt{\pi}} = 1.1284$. The maximum chord distance between any two nodes on the unit sphere would be twice the radius of the sphere or $\frac{1}{\sqrt{\pi}} \approx 0.5642$ which is half that of the two-dimensional toroidal unit circle.

Only flat unit circles and flat unit rectangles are considered in this paper, so they are referred to simply as *unit circles* and *unit rectangles*, and it is understood that there is no wraparound at their border.

When comparing the findings presented in this paper to other results in the RGG literature, it is critical to determine whether the other models assume flat spaces, toroidal spaces, or simply disregard border effects.

1.4 Distance in Flat and Toroidal Spaces

In two-dimensional flat space the Euclidean distance between two nodes indexed as i and j at positions (x_i, y_i) and (x_j, y_j) is given by the Pythagorean Theorem as the positive value of

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{4}$$

In the same way, in d-dimensional flat space the Euclidean distance between two nodes indexed as i and j at positions $(x_{i,1}, x_{i,2}, \ldots x_{i,d})$ and $(x_{j,1}, x_{j,2}, \ldots x_{j,d})$ is given by the positive value of

$$\sqrt{\sum_{d'=1}^{d} (x_{i,d'} - x_{j,d'})^2} \tag{5}$$

In one-dimensional flat space the Pythagorean Theorem is equivalent to simply taking the absolute value of the difference between the two positions in a line.

Calculating the Euclidean distance between two nodes in toroidal space is more complicated than in flat space because nodes can link to neighbors across the edges and corners of a toroidal unit square. In this paper, toroidal Euclidean distances are calculated only for the unit square. Two methods were used in the course of this work.

The first method, used for some of the simulations in two-dimensional spaces, is that described by Bettstetter [Bettstetter (2002a)] and is outlined here.

In two-dimensional toroidal space the Euclidean distance between two nodes in the unit square indexed as i and j at positions $(x_i,\, y_i)$ and $(x_j,\, y_j)$ is given by the *minimum* of the set of positive distances calculated as

$$\begin{aligned}
&\sqrt{(x_i - 1 - x_j)^2 + (y_i - 1 - y_j)^2} \quad , \\
&\sqrt{(x_i - 1 - x_j)^2 + (y_i + 0 - y_j)^2} \quad , \\
&\sqrt{(x_i - 1 - x_j)^2 + (y_i + 1 - y_j)^2} \quad , \\
&\sqrt{(x_i + 0 - x_j)^2 + (y_i - 1 - y_j)^2} \quad , \\
&\sqrt{(x_i + 0 - x_j)^2 + (y_i + 0 - y_j)^2} \quad , \\
&\sqrt{(x_i + 0 - x_j)^2 + (y_i + 1 - y_j)^2} \quad , \\
&\sqrt{(x_i + 1 - x_j)^2 + (y_i - 1 - y_j)^2} \quad , \\
&\sqrt{(x_i + 1 - x_j)^2 + (y_i + 0 - y_j)^2} \quad , \\
&\sqrt{(x_i + 1 - x_j)^2 + (y_i + 1 - y_j)^2} \quad . \qquad (6)
\end{aligned}$$

Note that the fifth distance calculation of the nine in the list above is the simple flat Euclidean distance. It may not be necessary to calculate all of the 3^d cases: if any distance is found to be less than or equal to r, then the two nodes are neighbors and no further calculations need be done. As most pairs of nodes are not neighbors, however, the full set of calculations would nearly always have to be done.

The "+ 0" terms are included for clarity and the pattern of counting in the "base 3" values of -1, 0, and 1 to extend the toroidal distance calculation to other values of d is straightforward.

The toroidal distance calculation based on this method is expensive because 3^d cases may have to be checked. This paper considers RGGs in spaces of dimensions up to six which would require 729 cases to be calculated.

A second and much faster implementation of the toroidal distance calculation was obtained, however, by observing that the distance contribution for each dimension in (5) for the flat distance can be *independently* adjusted to achieve the same computation as (6).

The independent adjustment method is implemented as follows. In d-dimensional toroidal space the Euclidean distance between two nodes indexed as i and j at positions $(x_{i,1}, x_{i,2}, \ldots x_{i,d})$ and $(x_{j,1}, x_{j,2}, \ldots x_{j,d})$ is given by the positive value of

$$\sqrt{\sum_{d'=1}^{d} z_{i,j,d'}^2} \tag{7}$$

$$\text{where } z_{i,j,d'} = \begin{cases} x_{i,d'} - x_{j,d'} - 1 & \text{if } x_{i,d'} - x_{j,d'} > 1/2, \\ x_{i,d'} - x_{j,d'} + 1 & \text{if } x_{i,d'} - x_{j,d'} < -1/2, \\ x_{i,d'} - x_{j,d'} & \text{otherwise.} \end{cases}$$

Simulations using the independent adjustment toroidal distance calculation of (7) required roughly 20% less total computation than those using (6) for two-dimensional spaces and the savings were far greater in higher-dimensional spaces due to reducing $O(3^d)$ operations to $O(d)$ operations.

1.5 RGG Growth Models: n-for-given-r and r-for-given-n

Connectivity properties of RGGs are approached in the literature from different directions based on how the RGGs grow or evolve. One model for growth asks the question "What range r is required for connectivity for a given number of nodes n in a space?" This is referred to in this paper as the *range for a given number of nodes (r-for-given-n)* growth model.

A complementary model for growth asks the question the other way around as "How many nodes n in a space are required for connectivity for a given range r?" This is referred to in this paper as the *number of nodes for a given range (n-for-given-r)* growth model.

As shown in Section 6, the n-for-given-r and r-for-given-n growth models are not quite inverses of each other although they may be at least approximately inverses under certain conditions.

Both of these models rely on answering the underlying question "What is the probability that an RGG with a given number of nodes n and a given range r is connected?"

Much of the RGG literature is based on the r-for-given-n growth model because of theorems by M.D. Penrose[11] that are discussed later in detail. These models take the approach of fixing the number of nodes and then increasing the range to the minimum value required to achieve a threshold of connectivity or the disappearance of isolated nodes.[12]

To the degree that these RGG abstractions model the real world, however, it seems more reasonable to assume that nodes with a fixed range would be added to a constellation of nodes than to assume that a fixed number of nodes would coordinate a variable yet common range. For that reason, more effort is expended in this work on the n-for-given-r growth model than on the r-for-given-n growth model, although models for both are developed and contrasted.

[11] As of 20 May 2022, Google Scholar reports that [Penrose (1999)] has been cited 584 times in the literature and [Penrose (2003)] has been cited 2,790 times.

[12] Once added, a node itself does not disappear. The term "disappearance" is used in the sense that there are no longer any isolated nodes in the RGG; that is, every node is linked to at least one neighbor.

1.6 Connectivity

Most prior research involving RGGs has sought to determine the number of nodes and the range that assure a high probability of being *connected*. In some cases, however, it may be desirable to determine the thresholds for the complementary problem of guaranteeing a high probability of the RGG being *disconnected*.

The condition of achieving a disconnected network applies in a situation such as preventing disease transmission; for example, see [Estrada (2016)] [Penrose (1997)] [Penrose (1999)].

Diversity has been shown to be an effective means for improving survivability in information systems and networks [Bailey (2005)]) as well as in biological systems. In the case of an information system or network, a node population that may share vulnerabilities as a monoculture could be made artificially diverse so that subsets of the nodes in that population that share vulnerabilities *are not* fully connected with high confidence.

For these reasons, models are presented for conditions under which the RGG is disconnected with high probability, which is defined here to mean the probability that the RGG is fully connected is at most 0.01.

This paper also obtains models for the *mean number of nodes* and the *mean range* required for connectivity.

It is straightforward to determine whether *a specific RGG* is fully connected according to the rules of its configuration but it is more difficult to determine the probability that *any randomly generated RGG* with a given configuration is connected. On the other hand, it is far easier to determine the probability of the weaker condition that any randomly generated RGG with a given configuration does not contain isolated nodes.

Because a necessary condition for full (or nearly full) connectivity is that there are no (or very few) isolated nodes, much of the literature is based on the relationship between the probability of RGGs being connected and the probability of nodes being isolated. More specifically, many models for connectivity are based on approximating or bounding the desired probability that an RGG is fully connected by the far easier-to-obtain probability that it contains no isolated nodes.

Many models involve the further step of making the assumption that having no isolated nodes is not only a necessary condition but that is effectively a *sufficient* condition for 100% connectivity as well.

Much of this paper is devoted to determining the conditions under which this approximation is reasonable. Based on simulations of RGG growth, highly accurate *Linear Regression Models (LRMs)* are obtained that do not rely on this assumption.

Empirical models are also presented for RGG growth that achieve several thresholds of less-than-full connectivity.

Many RGG studies have considered the general case of k-connectivity that requires that there are at least k paths between each pair of nodes. As noted in Section 1.1 of the Introduction, this paper is concerned with only the most basic RGG connectivity where there exists at least one path between pairs of nodes in a component; this is called *1-connectivity* or *simple connectedness* or *simple connectivity* in the literature.[13]

[13] Note that Penrose [Penrose (1999)] [Penrose (2003)] derives results for "$(k+1)$-connectivity" where $k=0$ means 1-connectivity.

1.7 Configurations and Geometries

In this paper the *configuration* of the nodal space for an RGG in a unit space is defined by the following parameters:

- Shape of the space (square, circle, rectangle, lattice).
- Whether the space is flat or toroidal.
- Number of nodes n.
- Common nodal communication range r (or ω for a lattice).
- Dimension d of the space.
- Number and placement of obstructions and fraction of area obstructed.

This paper considers unit squares and lattices that are flat or toroidal but only flat unit circles and flat unit rectangles are addressed. All of the nodal spaces have unit area other than the lattice which is parameterized differently as described in Section 9.2.

A specific simulated instance of a randomly placed set of nodes and their resulting links, in accordance with a given configuration and the rules for its flat or toroidal space, is referred to as a *geometry*. The randomly generated geometries are tested to determine whether they have reached the condition of being fully connected, partially connected, or containing no isolated nodes.

In the n-for-given-r growth model, geometries have a fixed range and RGGs grow by adding nodes until the *first* number of nodes is reached where a desired connectivity condition is achieved. It is important to set the stopping point as the first such number of nodes as is shown in Section 4.1.3 that continuing to add nodes can break that connectivity condition as well as restore it. In this paper the statistical metrics regarding the number of nodes required to achieve a given connectivity condition for RGGs with n-for-given-r growth apply to the number of nodes that first achieve that condition.

The order in which the nodes are added is critical, as the same set of nodes added in a different order would be likely to result in a different first number of nodes where the connectivity condition is reached.

In the r-for-given-n growth model, geometries have a fixed number of nodes and RGGs grow by increasing the range[14] until a desired connectivity condition is reached. In contrast to the n-for-given-r growth model, where continuing to add nodes can break and restore a connectivity condition, the minimum range that achieves the desired connectivity condition for an RGG with r-for-given-n growth is unique and increasing the range cannot disconnect components or create isolated nodes.

[14] "Increasing the range" is notional as discussed extensively in later sections.

2 Nomenclature

The following abbreviations and symbols are used in this paper. Terms defined and used in only one place usually are not included in this list.

$\ln x$, $\ln(x)$	– Natural logarithm of x or the base e log of x, where $e = 2.71828182845\ldots$. All logarithms used in this paper are natural logs.
$\exp(x)$, e^x	– Exponential function of x.
$\Gamma(x)$	– Gamma function of x.
%ile	– Percentile of a set of values; for example, "99th percentile" is written as "99%ile."
GC	– Giant Component.
IoT	– Internet of Things.
LRM	– Linear Regression Model.
MANET	– Mobile Ad Hoc Network.
R^2	– Coefficient of Determination: a commonly used measure of goodness-of-fit of a linear regression.
MAPE	– Mean Absolute Percentage Error: a measure of the predictive ability of a candidate model.
RGG	– Random Geometric Graph.
n-for-given-r	– Number of nodes for a given range. It is understood that the value is the number of nodes that *first* achieves a desired connectivity condition.
r-for-given-n	– Range for a given number of nodes. It is understood that the value is the *minimum* range that achieves a desired connectivity condition.
RF	– Radio Frequency.
SSR	– Sum of the Squared Residuals.
SSRUV	– Sum of the Squared Residuals for Untransformed Variables.
VANET	– Vehicular Ad Hoc Network.
A	– Normalized area of a nodal space: a unit square, unit rectangle, or unit circle has $A = 1$ (unit area).
a	– Normalized length of a side of a unit square space (so $a = 1$) or radius of a unit circular space (so $a = 1/\sqrt{\pi} = 0.5642\ldots$).

a_h and a_w – Normalized height and width of a unit rectangular space where $a_w \times a_h = 1$.

r – Normalized nodal communication range for a space of unit area.

(Units are not associated with A, a, a_h, a_w, or r.)

A' – Actual (not normalized) area of a nodal space: square, rectangular, or circular.

a' – Actual length of a side of a square space or radius of a circular space.

a'_h and a'_w – Actual height and width of a rectangular space where $a'_w \times a'_h = A'$.

r' – Actual nodal communication range for a space of area A'.

(Units can be associated with A', a', a'_h, a'_w, and r'.)

$\alpha_C(n)$ – Function used by Koskinen to replace c_P in Penrose's Theorem. For example: $\alpha_{0.99}(n)$ for $C = 0.99$.

C – Confidence: probability of occurrence of a specified condition.

$c_{GK}(n)$ – Function used in Gupta and Kumar's Theorem.

c_P – Value used in Penrose's Theorem.

d – Dimension of a nodal space. The space is said to be "d-dimensional."

1D, 2D, 3D, ... – One-dimensional space, two-dimensional space, three-dimensional space, ...

k – An RGG is k-connected if there are at least k paths between each pair of nodes.

L^1 – "Manhattan" distance.

L^2 – Euclidean or Pythagorean Theorem distance.

LP_i – Lattice point i.

m, b – Coefficients used in LRMs. Following a transformation of variables that yields a "straight line," m is the slope and b is the ordinate axis intercept of the line resulting from linear regression on the transformed variables.

n – Number of nodes.

N – Sample size.

M – Length of a "side" of a square lattice: a lattice contains $M \times M$ points where nodes can be placed.

ω – Nodal communication range for a lattice: a distance given as a count of lattice points.

p – Placement probability for a lattice: the probability that a point in a lattice contains a node. A point can contain at most one node.

r_{max} – Worst-case range: the maximum range required for full connectivity or no isolated nodes.

$r_{max}(n, N)$ – An r_{max} obtained for given values of n and N.

S – Fraction of nodes that are in the GC or the maximum fraction of nodes in any component if the GC is not unique.

θ_d – Volume of a "unit ball" (a line, circle, sphere, or hypersphere) with radius 1 in d-dimensional space.

(x, y) – Coordinates of a node in two-dimensional space.

(x_i, y_i) – Coordinates of node i in two-dimensional space.

$(x_1, x_2, x_3, \ldots x_d)$ – Coordinates of a node in d-dimensional space.

$(x_{i,1}, x_{i,2}, x_{i,3}, \ldots x_{i,d})$ – Coordinates of node i in d-dimensional space. **(The symbols x and y are occasionally reused without ambiguity to denote other values.)**

Mid-r set, large-r set, small-r set, mid+small-r set – Mid-range set, large-range set, small-range set, and union of mid-range and small-range sets as defined in Section 4.1.2.

Small-n set, large-n set, small+large-n set – Set of small numbers of nodes, set of large numbers of nodes, and union of small and large numbers of nodes as defined in Section 5.1.

3 Neighbors, Degree, Isolated Nodes, and Components

This section outlines relationships between the *mean degree* of nodes in an RGG (that is, the mean number of neighbors or links per node) and the probability that there are no isolated nodes. The relationship between the disappearance of isolated nodes and achieving full connectivity is discussed. Theorems from the RGG literature that deal with these factors are presented.

This section demonstrates the two models of RGG growth considered in this paper: n-for-given-r and r-for-given-n.

The conditions for the emergence of the GC are demonstrated for both types of growth.

The significance of border effects is discussed to highlight the difference between the toroidal models that allow spaces to wrap around at their edges and flat models where nodes have no neighbors beyond the border of the space.

3.1 Examples of RGGs

Figure 1 shows a geometry consisting of 38 nodes scattered randomly in a flat unit square demonstrating Poisson distribution of nodes. Even though the Poisson point process is homogeneous, there are areas in the square where nodes are clustered closely together and there are other areas that are void of nodes.[15]

Figure 2 shows the same geometry as that in Figure 1 with the common range of the nodes set at $r = 0.2000$. Dots represent the nodes, and the circles with radii r represent the extent of the range. Two nodes are connected when they are within each other's circles. The green dots and circles represent the largest component (the GC) and the red dots and circles represent smaller components.

Figure 2 is the most faithful description of the geometry but a more readable representation is shown in Figure 3 where the circles are drawn at the half-range $r/2$ instead of at the full distance r. In Figure 3, two nodes are connected when their half-range circles simply touch.[16] The GC shown in green is more clearly visible in Figure 3 than in Figure 2 and it is easier to trace a path from any node to any other node in the GC.

The links between connected nodes are shown in Figure 4. Figure 5 is simplified further by eliminating the half-range circles and showing only the links.

The discussion below may be easiest to follow using the half-range circles shown in Figure 3.

The RGG shown in Figure 3 has 35 of its 38 nodes in the GC when $r = 0.2000$. That is to say, $S = 35/38 = 0.921$: the RGG is 92.1% connected.

[15] Figure 20 in Section 3.5 demonstrates this consequence of the Poisson distribution of nodes for a larger number of nodes.

[16] See page 243 of [Bollobás (2006)] for an example of this method of representing RGGs.

The two smaller components are clearly visible in Figure 3. One component contains two nodes and one component consists of a single isolated node. The isolated node at position $(0.9901, 0.8145)$ near the upper right-hand corner is unconnected to any other node because this space is a flat unit square. If this space were instead a toroidal unit square, then the node at $(0.9901, 0.8145)$ would be connected to the nodes at $(0.0127, 0.8687)$ and $(0.0408, 0.9557)$ near the upper left-hand corner and thus would be part of the GC.

In the n-for-given-r growth model, nodes are sequentially added to the geometry. Each new node added to an RGG can produce only one of the following results: it can increase the size of an existing component by one, it can bridge two or more existing components to form a single larger component, or it can create a new isolated node. The process stops when 100% of the nodes in the geometry are part of a single GC: the condition of *full connectivity*.

This process illustrates an intrinsic difference between the n-for-given-r and r-for-given-n growth models where a fully connected RGG with n-for-given-r growth may have a new node added that is an isolated node, thus creating a disconnected RGG, whereas a fully connected RGG with r-for-given-n growth cannot become disconnected when range increases. This is discussed further in Section 4.1.3.

In the r-for-given-n growth model as described by [Penrose (1999)] [Penrose (2003)], a set of nodes is placed uniformly at random positions in a space and the range r is set at zero so that all nodes in the RGG are isolated. As the range is increased, components of connected nodes begin to form that increase in size as nearby nodes and other components become connected. The process stops when the minimum range is found that achieves the desired connectivity condition for the RGG.[17,18]

[17] The concept of the range being "increased" is notional and does not imply that the r-for-given-n growth model requires nodes with variable range. The r-for-given-n growth model is framed in this way only to define the minimum range needed to provide connectivity in an RGG with a given number of nodes.

[18] As the goal of r-for-given-n growth is only to find the minimum range r for which an RGG reaches a specific connectivity condition, the range actually being "increased" is not essential. An equivalent approach would be to start with a large initial range and decrease the range until the RGG no longer meets the desired connectivity condition. The method used for this paper was to perform a binary search to find the smallest range for which the connectivity condition is met.

This process can be observed in Figure 3. A small increase in range from 0.2000 to 0.2385 would cause the isolated node at $(0.9901, 0.8145)$ to become connected to the node at $(0.7542, 0.7794)$, thus adding it to the GC. Continuing to increase the range, at $r = 0.2527$ the node at $(0.8690, 0.1522)$ in the lower right-hand corner would be connected to the node at $(0.7184, 0.3552)$, which would cause that smaller component to be absorbed into a single GC that contains all of the nodes in the RGG, thus achieving full connectivity.

In the example described above, note that full connectivity was *not* achieved for the RGG shown in Figure 3 at the point when the last isolated node was absorbed into a component. This significance of this fact discussed in Section 3.5.

In the n-for-given-r growth model, nodes with fixed range are added to an RGG until the GC contains all of the nodes, at which point the process stops; in the r-for-given-n growth model the range for a fixed set of nodes is increased until the GC contains all of the nodes. In both models, there is a "phase transition" to the "connected regime" [Bobrowski (2017)] [Dall (2002)] in which the largest component in the RGG explodes from containing only a small fraction of the nodes in the space to the GC that encompasses the vast majority of the nodes.

Referring to Figure 5 which shows only the nodes and the links between neighbors, the serpentine path demonstrates the challenge of making any simplifying assumptions about the shape of connected components as an RGG grows. In the case of this graph, there are several sets of nodes containing as few as one or two or three nodes that, if deleted (or were not added), would partition the GC into much smaller components.[19] The serpentine path also demonstrates that nodes contained in a single component may be quite distant from each other in the space.

[19] A set of exactly R nodes that disconnects a network upon their removal from a network is called an *R-cutset* [Goldschmidt (1994)].

Looking only at the cutsets that would drastically affect the size of the GC shown in Figure 5, there is a 1-cutset consisting of the node at $(0.5862, 0.2671)$ that would cut the GC roughly in half if it were deleted (or, alternately, not added). There are two such 2-cutsets consisting of the nodes at $(0.6147, 0.4528)$ and $(0.7184, 0.3552)$ or $(0.6147, 0.4528)$ and $(0.7073, 0.4756)$, and a 3-cutset consisting of the nodes at $(0.3995, 0.3193)$ and $(0.4586, 0.2408)$ and $(0.5928, 0.1101)$. This demonstrates the potential fragility of networks in general and RGGs in particular and shows why many researchers have studied the conditions for k-connectivity where multiple paths provide resiliency to failure.

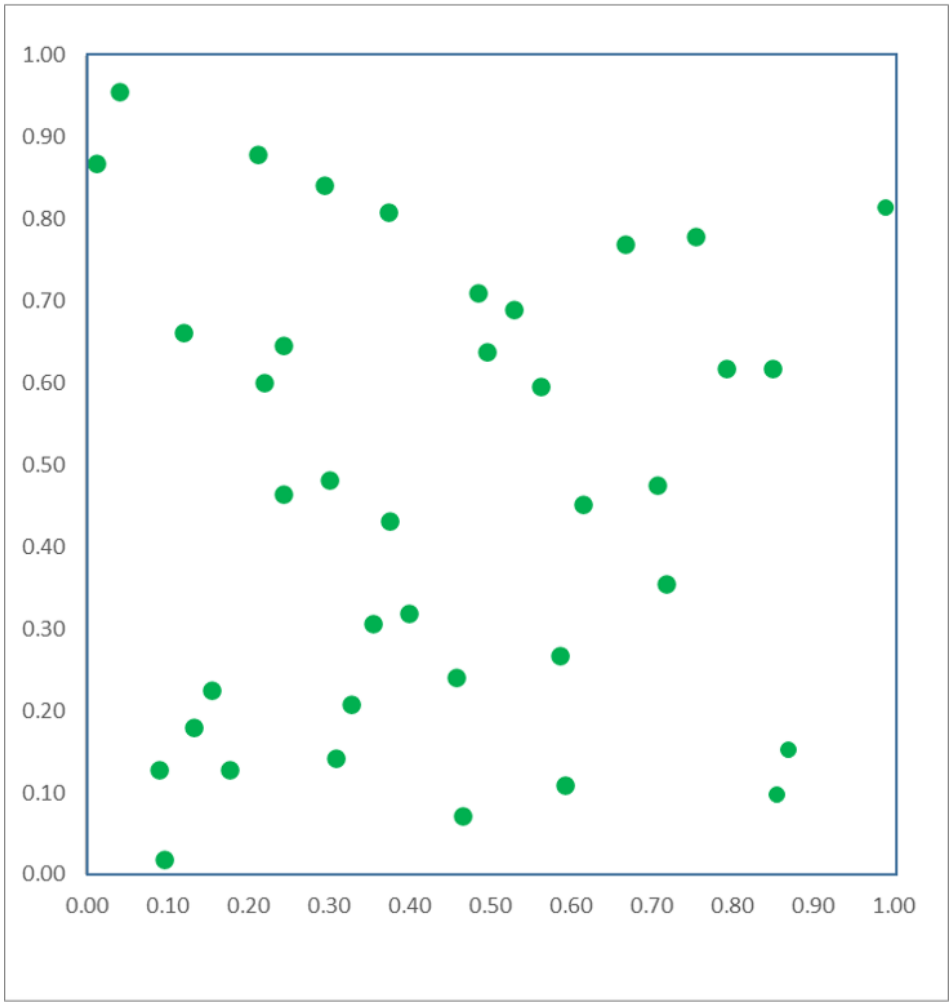

Figure 1: Geometry consisting of 38 nodes demonstrating Poisson distribution of nodes in a unit square. Communication range is unspecified in this figure.

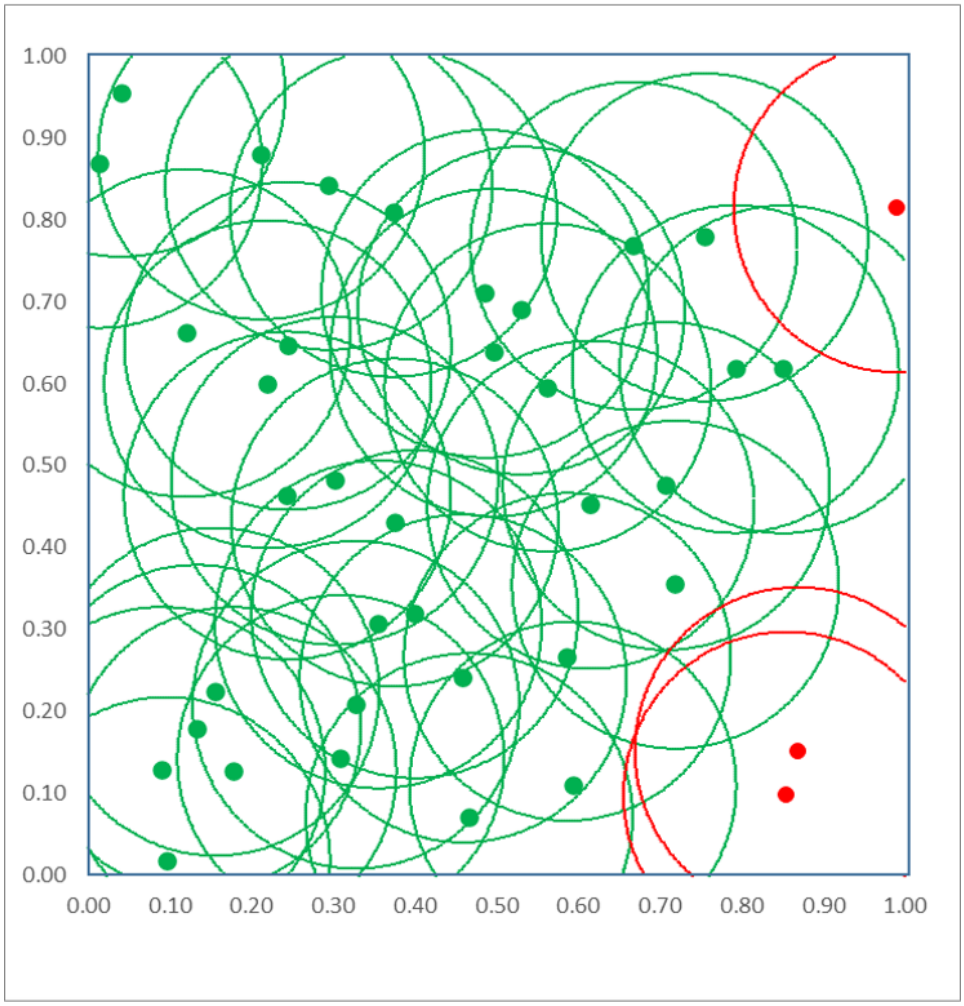

Figure 2: Geometry consisting of same 38 nodes shown in Figure 1 with communication range 0.2000 scattered in a unit square. A node is in range of another node when it is within the circle denoting the range r. See Figure 3, Figure 4, and Figure 5 for cleaner views of the geometry.

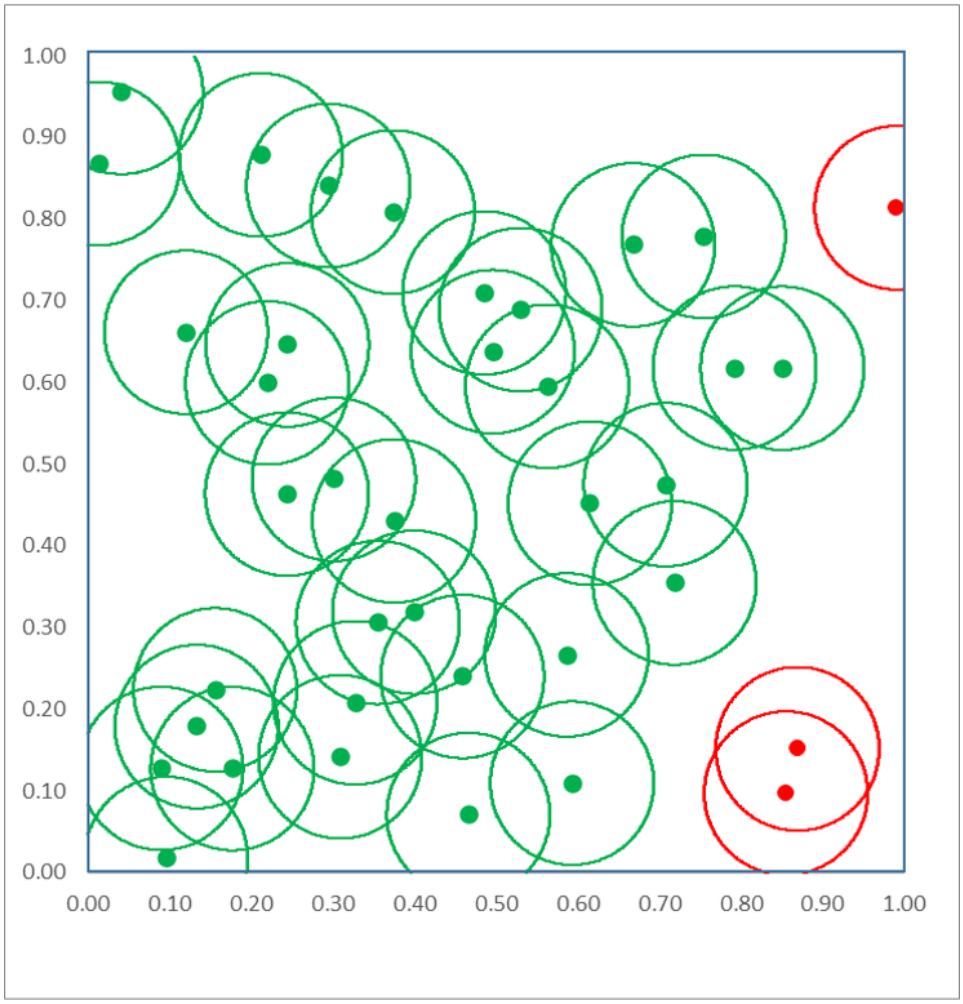

Figure 3: Same geometry as shown in Figure 2 but these circles denote only half the range. Thus, a node is in range of another node when its half-range circle touches that of the other node. There is a GC composed of 35 nodes, a component composed of two nodes, and a single isolated node.

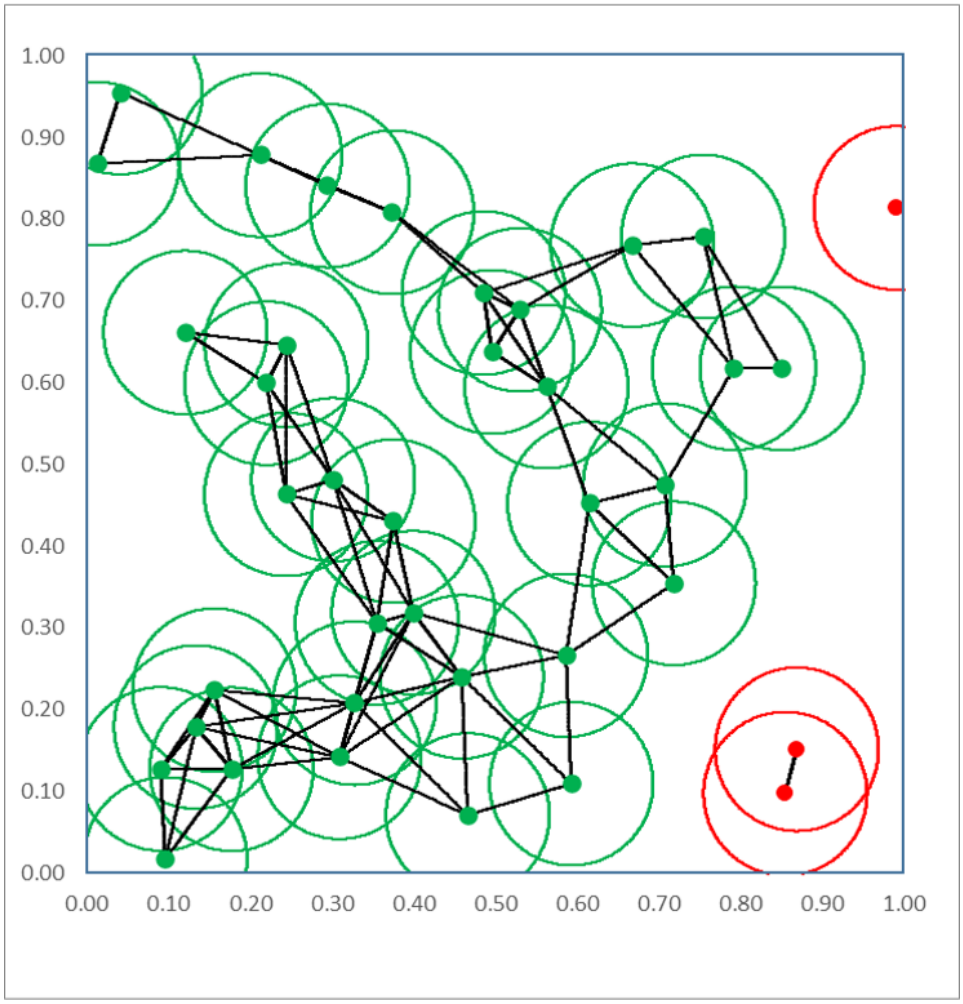

Figure 4: Same geometry as shown in Figure 2 and Figure 3, with addition of links between nodes that are connected neighbors.

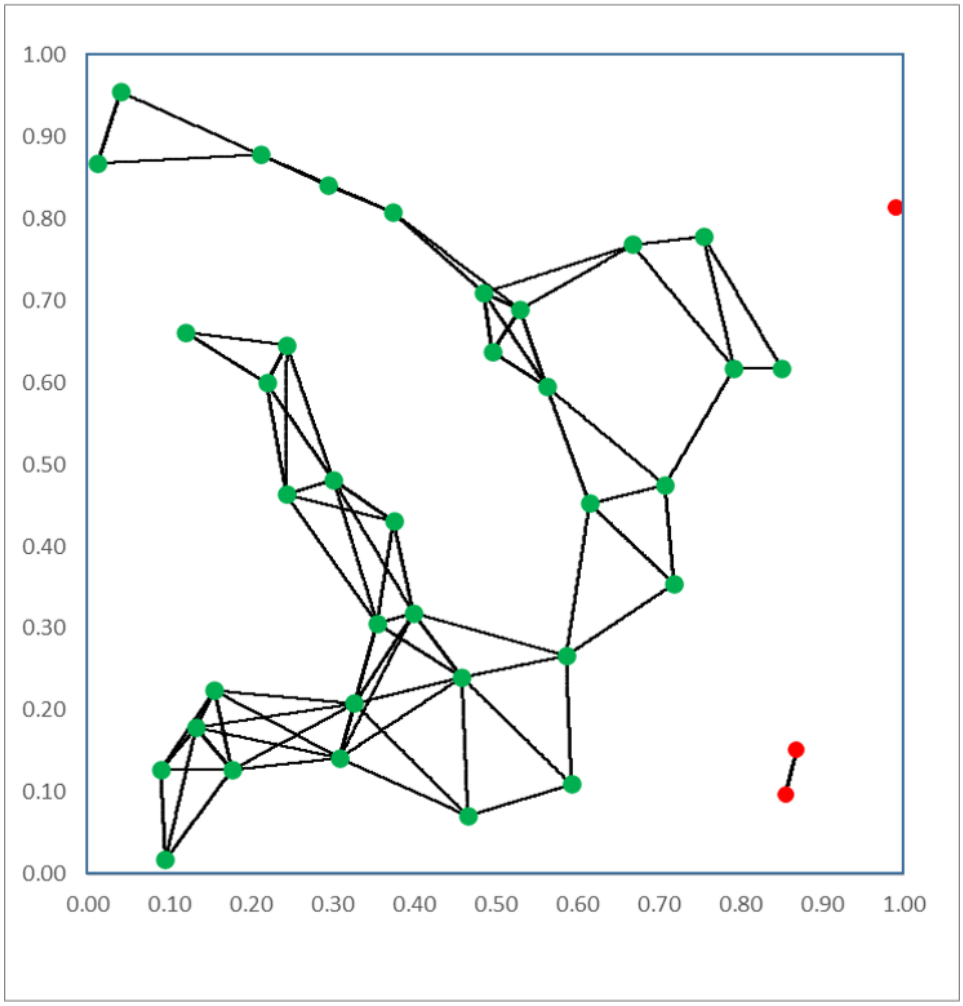

Figure 5: Same geometry as shown in Figure 2, Figure 3, and Figure 4, showing only the links between connected neighbors.

3.2 Examples of n-for-given-r Growth

The evolution of RGGs is demonstrated in Figure 6 for RGGs in a flat unit square with $r = 0.0100$ and n-for-given-r growth. The graph shows a scatter plot of S for the GC for 100 simulated geometries where the number of nodes n is increased until full connectivity is achieved. S remains small until about 10,000 nodes are in the space. In the region of $n = 10{,}000$ to 20,000 (centered at approximately 15,000 nodes) the size of the GC grows from nearly 0% to practically 100%.[20]

Figure 7 shows S as a function of the mean degree of the nodes in the RGGs for the same 100 simulations used for Figure 6. Again, there is an abrupt growth in S, which occurs between mean degrees of 4 and 5; it is centered at a mean degree of approximately 4.5 neighbors per node for RGGS in a flat unit square. These values are consistent with those observed by others [Balister (2005)] [Dall (2002)].

Figure 8 shows the number of nodes (rather than the fraction S) in the largest and second-largest components for the same simulations shown in Figure 6. As nodes are added the GC and the next-largest component are roughly the same size until the largest component begins to absorb the new nodes more frequently than the next-largest component does. Abruptly, the second-largest component begins to *decrease* in size — not because the component gets smaller but because the currently second-largest component is repeatedly absorbed into the GC, leaving a smaller component as the new second-largest component.

Figure 9 shows how the number of distinct components grows with the number of nodes. Initially, as nodes are added they are largely isolated nodes or members of small components; however, the number of components begins to fall even before the GC becomes overwhelmingly dominant. The number of distinct components falls quickly and then flattens into a narrow tail, indicating that there are a few components that remain until the

[20] The large values of S in the left-hand tail of the graph are due to S being calculated on the basis of small numbers. With only one node the RGG has "100% of its nodes" in the GC. With two nodes the RGG almost always has two GCs, each with one node. With three nodes the RGG almost always has three GCs each with one node, and so on. As the number of nodes increases, S falls to a small value until the phase transition begins.

fortuitous addition of a randomly placed node bridges them into the GC or other component.

Figure 10 shows a scatter plot of S for RGGs in a toroidal unit square with $r = 0.0100$ and n-for-given-r growth. As in the previous example, the graph shows S for 100 simulated geometries where the number of nodes n is increased until full connectivity is achieved. The midpoint of the phase change in the size of the GC occurs at slightly less than $n = 15{,}000$.

Figure 11 shows the growth of S as a function of the mean degree for RGGs in a toroidal unit square. The midpoint of the phase transition in the size of the GC occurs at approximately 4.3 neighbors per node, which is slightly less than that observed for RGGs in a flat unit square as shown in Figure 7.

Figure 12 overlays the scatter plot of S as a function of the number of nodes shown in Figure 10 for RGGs in a toroidal unit square on the values of S shown in Figure 6 for RGGs in a flat unit square. The overlay of the plots demonstrates the observation made above that RGGs in a toroidal unit square experience the explosion of the GC at slightly smaller numbers of nodes than those in a flat unit square.

Similarly, Figure 13, which overlays the scatter plot of S as a function of the mean degree shown in Figure 11 for RGGs in a toroidal unit square on the values of S shown in Figure 7 demonstrates the observation made above that RGGs in a toroidal unit square experience the explosion of the GC at slightly smaller mean degrees than those in a flat unit square.

Figure 14 and Figure 15 show the fraction of nodes and the mean degree, respectively, in the GC for RGGs in a unit circle with $r = 0.0100$ and n-for-given-r growth. The phase transition for the size of the GC occurs again at approximately $n = 15{,}000$ nodes and mean degree of 4.5 neighbors per node.

Figure 16 overlays the scatter plot of S as a function of the number of nodes shown in Figure 14 for RGGs in a unit circle on top of the values of S shown in Figure 6 for RGGs in a flat unit square. Figure 17 overlays the scatter plot of S as a function of the mean degree shown in Figure 15 for RGGs in a unit circle on top of the values of S shown in Figure 7. The close alignment of these two scatter plots for RGGs in a unit circle and in

a flat unit square indicate that there is no visibly substantial difference in the emergence of the GC between those two spaces.

The emergence of the GC in RGGs in a flat unit square with n-for-given-r growth is discussed further in Section 4.2.1.3 when the condition of 50% connectivity is considered. Figure 6 is revisited in more detail in that section.

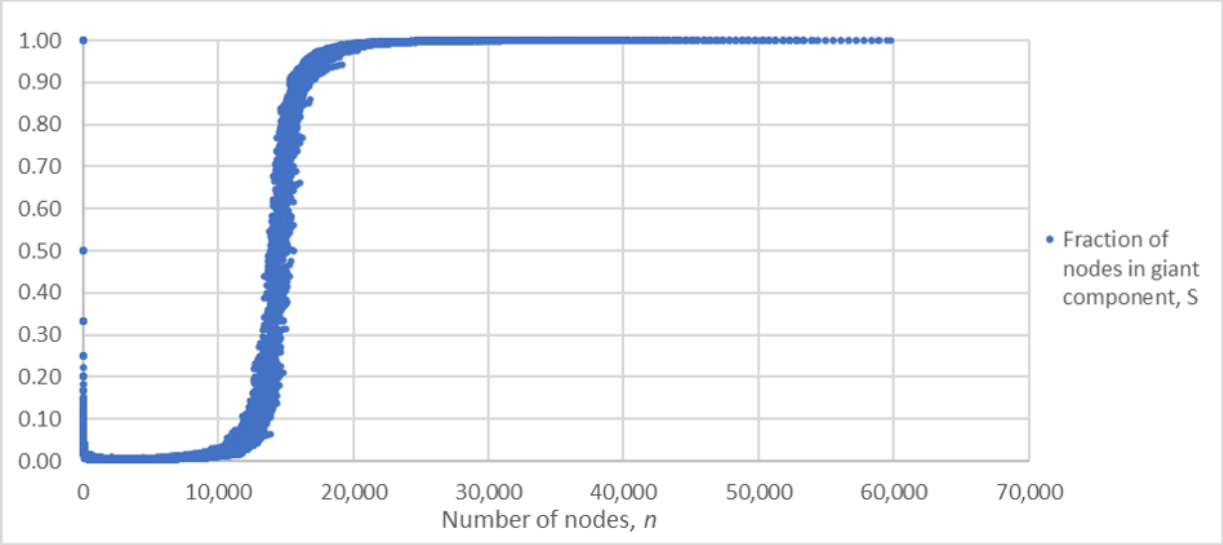

Figure 6: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a flat unit square with range 0.0100 and n-for-given-r growth. S is plotted as a function of the number of nodes.

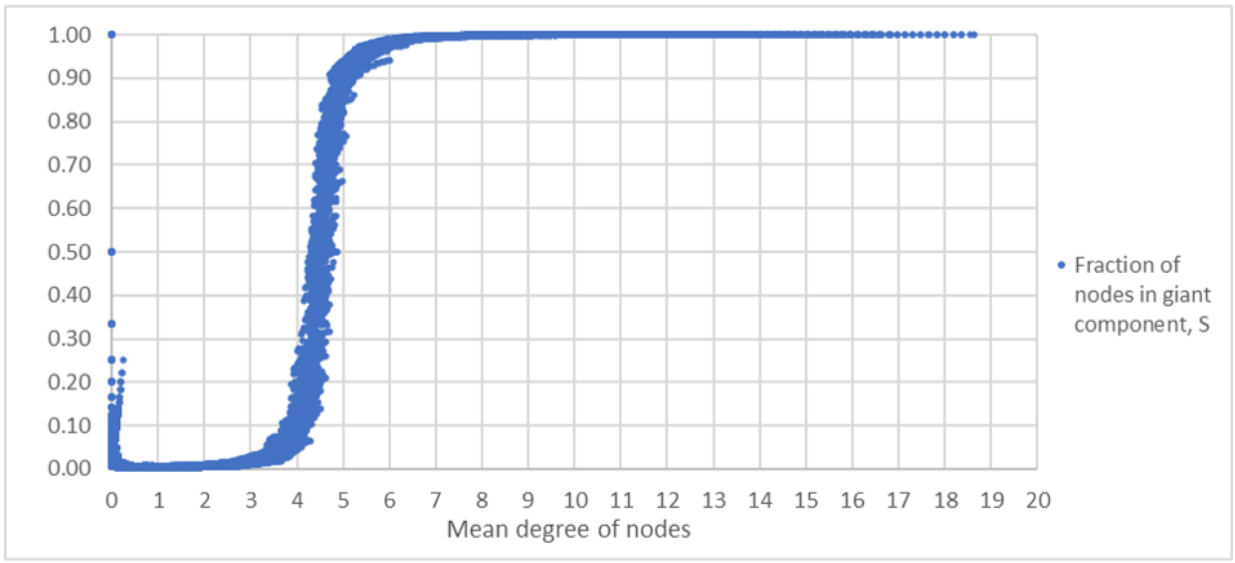

Figure 7: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a flat unit square with range 0.0100 and n-for-given-r growth. S is plotted as a function of the mean degree of the nodes.

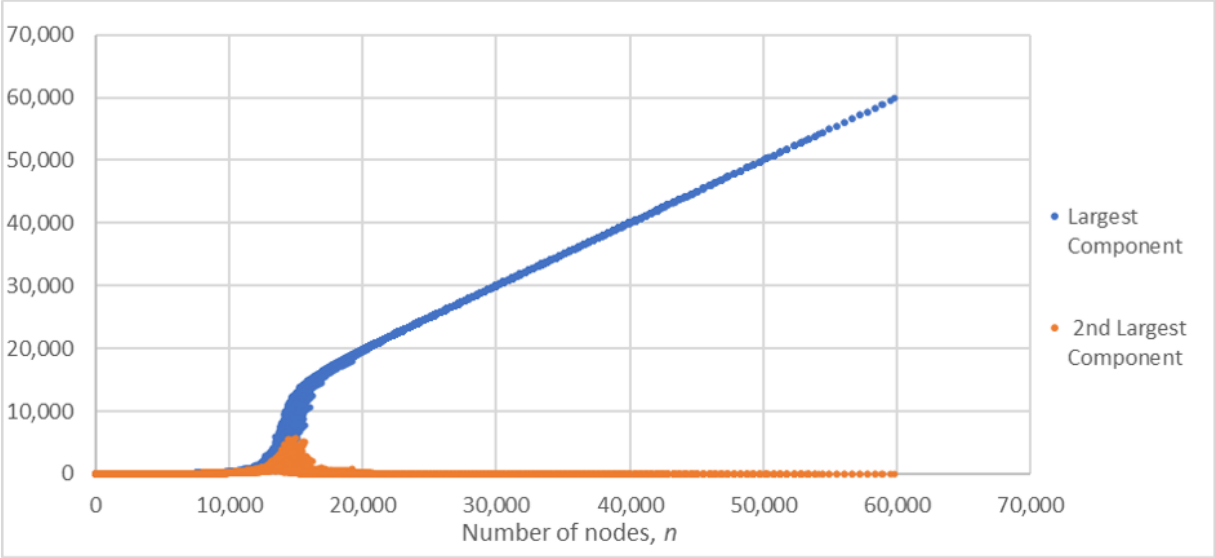

Figure 8: Number of nodes in the largest and second-largest components for the same simulated geometries used in Figure 6.

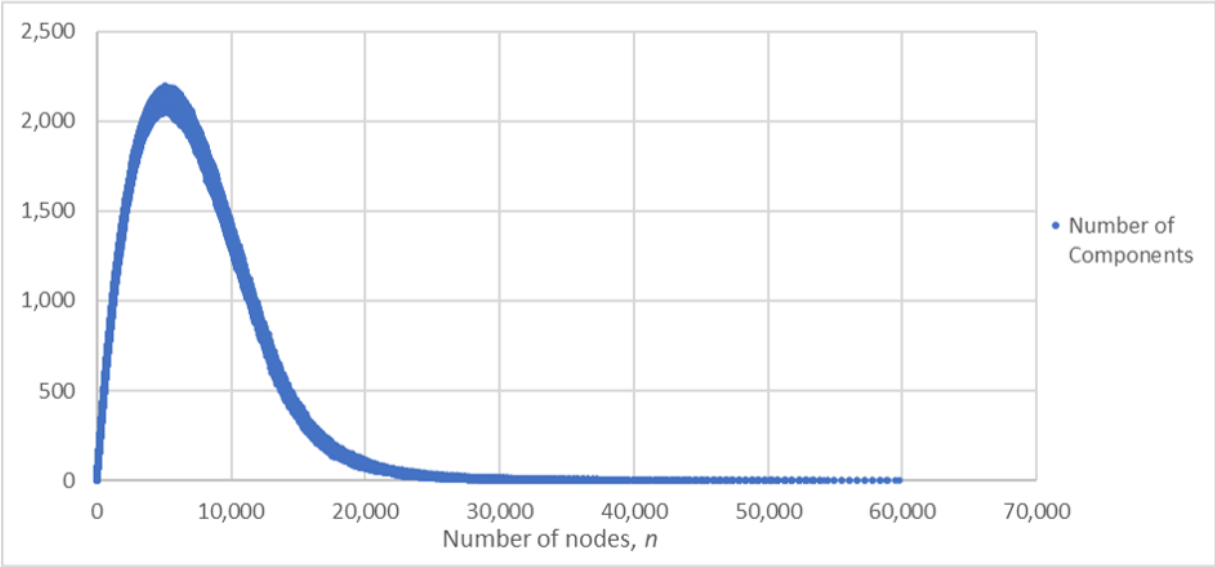

Figure 9: Number of components for the same simulated geometries used in Figure 6.

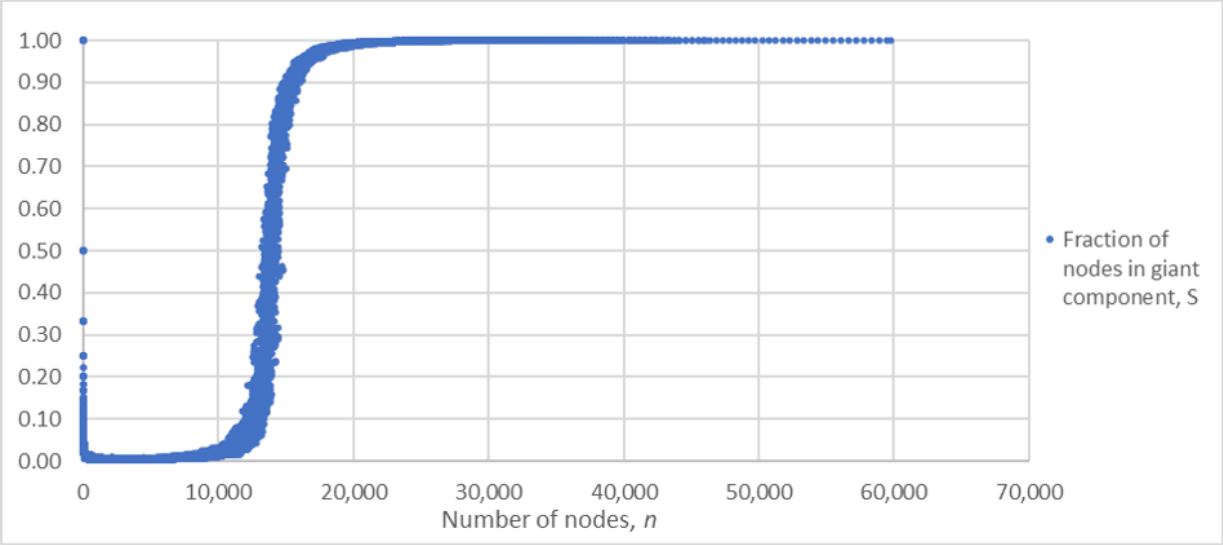

Figure 10: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a toroidal unit square with range 0.0100 and n-for-given-r growth. S is plotted as a function of the number of nodes.

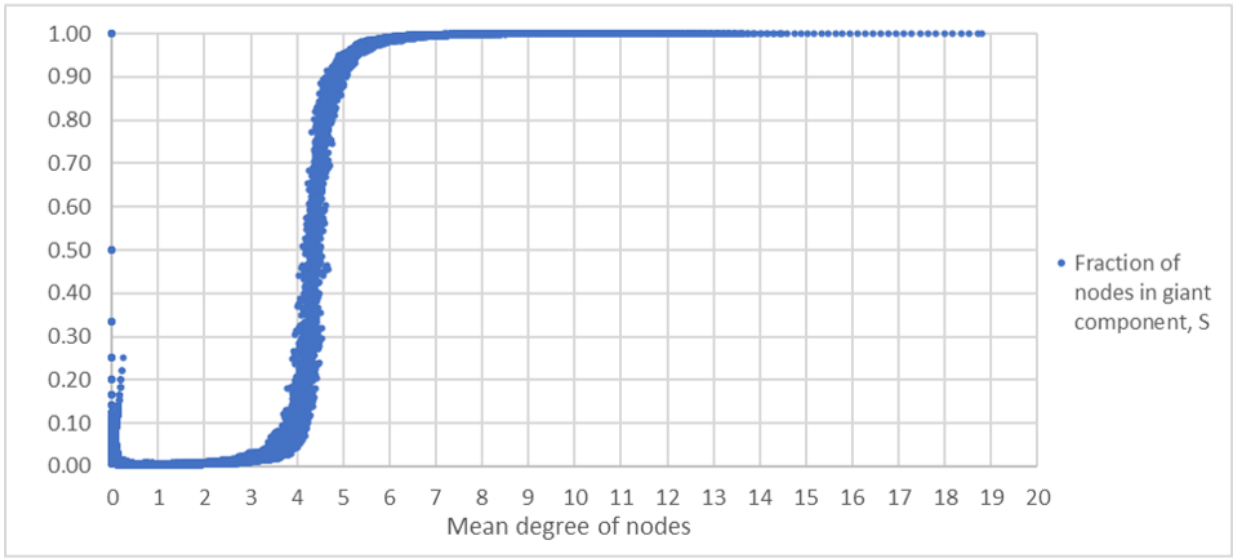

Figure 11: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a toroidal unit square with range 0.0100 and n-for-given-r growth. S is plotted as a function of the mean degree of the nodes.

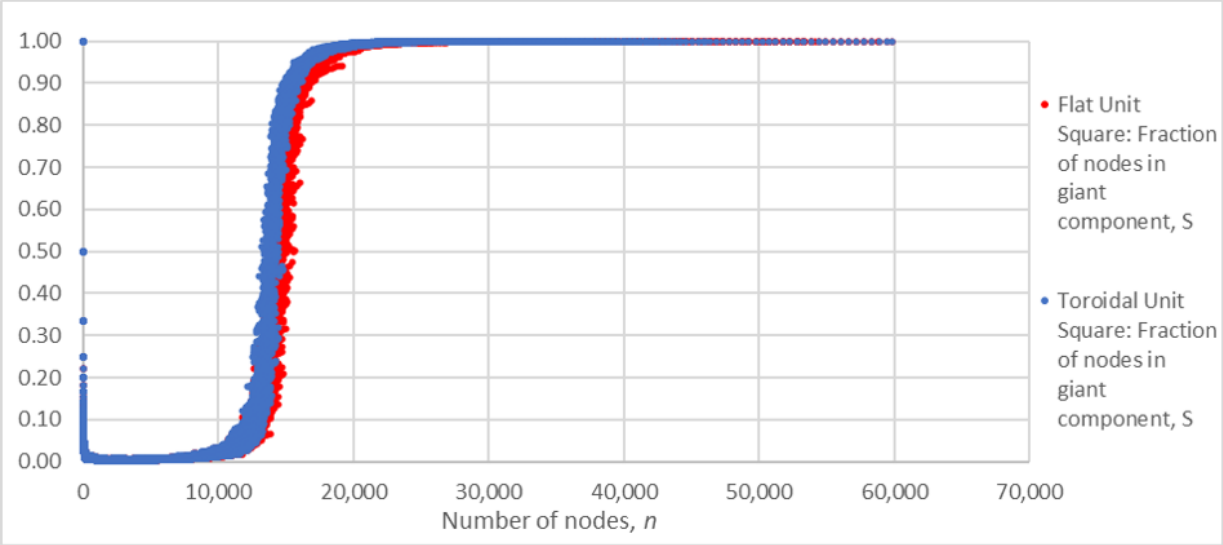

Figure 12: Overlay of scatter plots of S, as a function of number of nodes, shown in Figure 10 for RGGs in a toroidal unit square and Figure 6 for RGGs in a flat unit square.

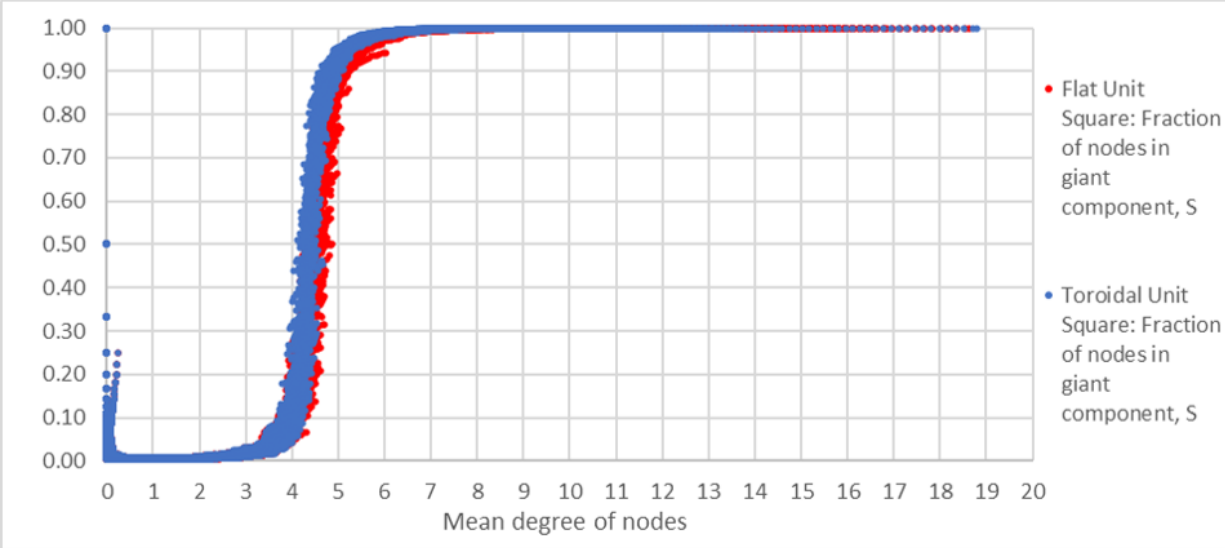

Figure 13: Overlay of scatter plots of S, as a function of mean degree, shown in Figure 11 for RGGs in a toroidal unit square and Figure 7 for RGGs in a flat unit square.

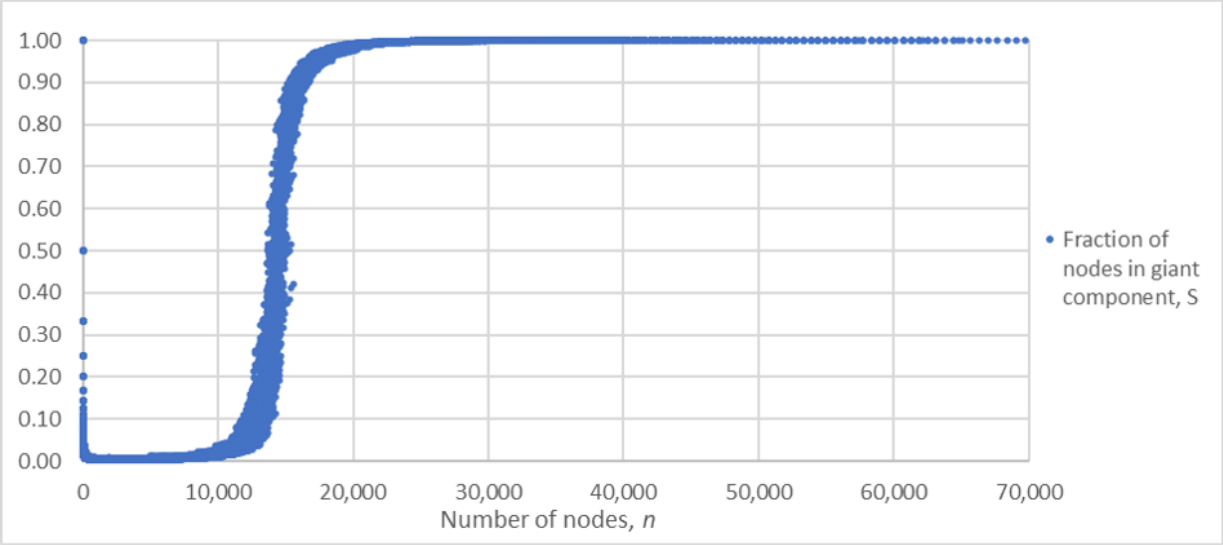

Figure 14: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a unit circle with range 0.0100 and n-for-given-r growth. S is plotted as a function of the number of nodes.

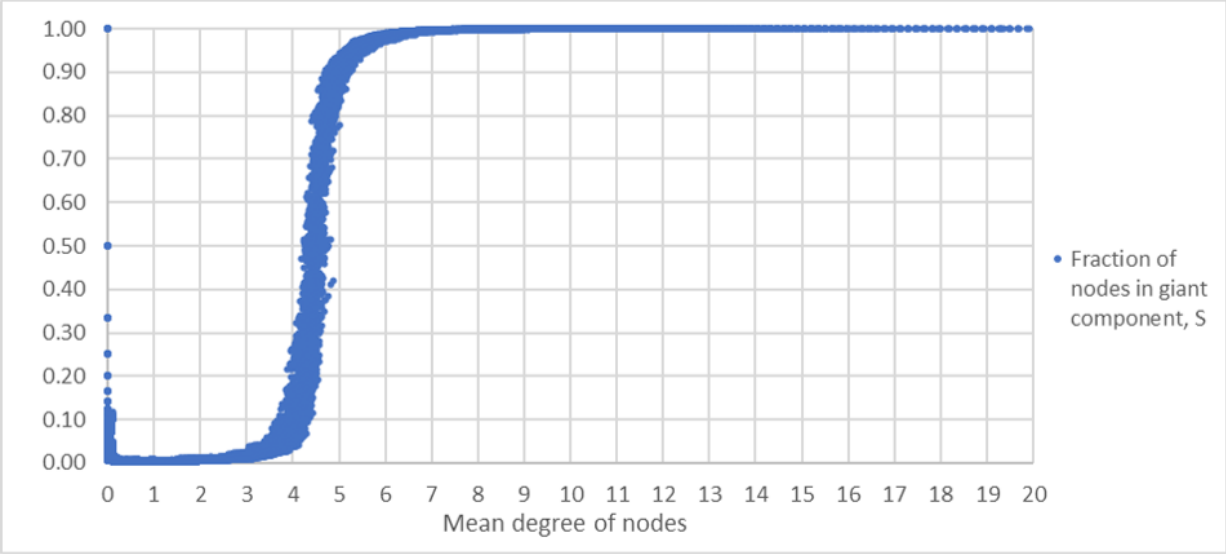

Figure 15: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a unit circle with range 0.0100 and n-for-given-r growth. S is plotted as a function of the mean degree of the nodes.

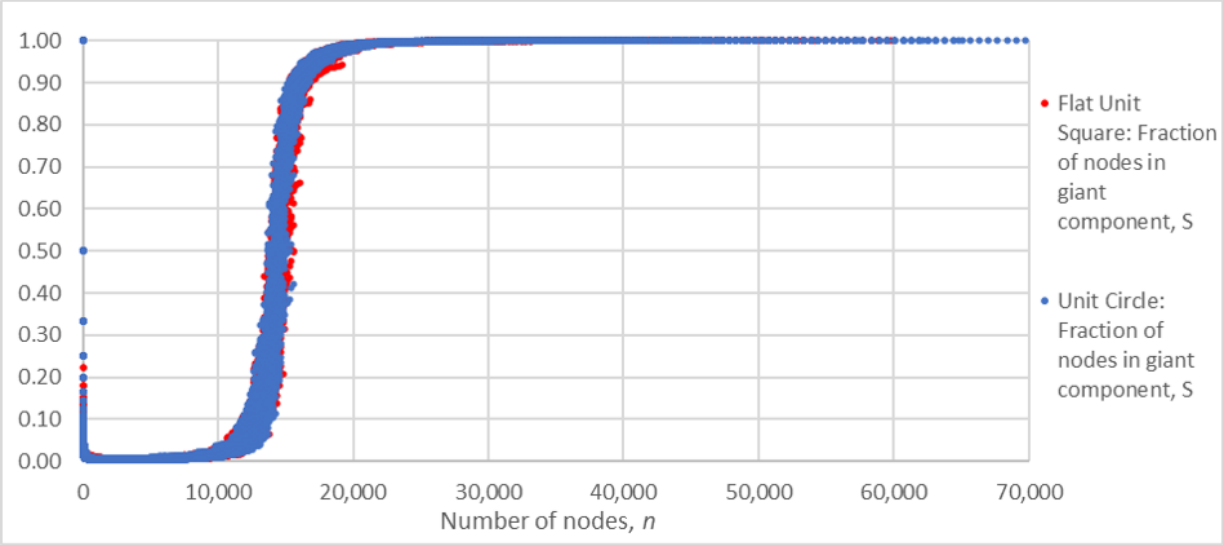

Figure 16: Overlay of scatter plots of S, as a function of number of nodes, shown in Figure 14 for RGGs in a unit circle and Figure 6 for RGGs in a flat unit square.

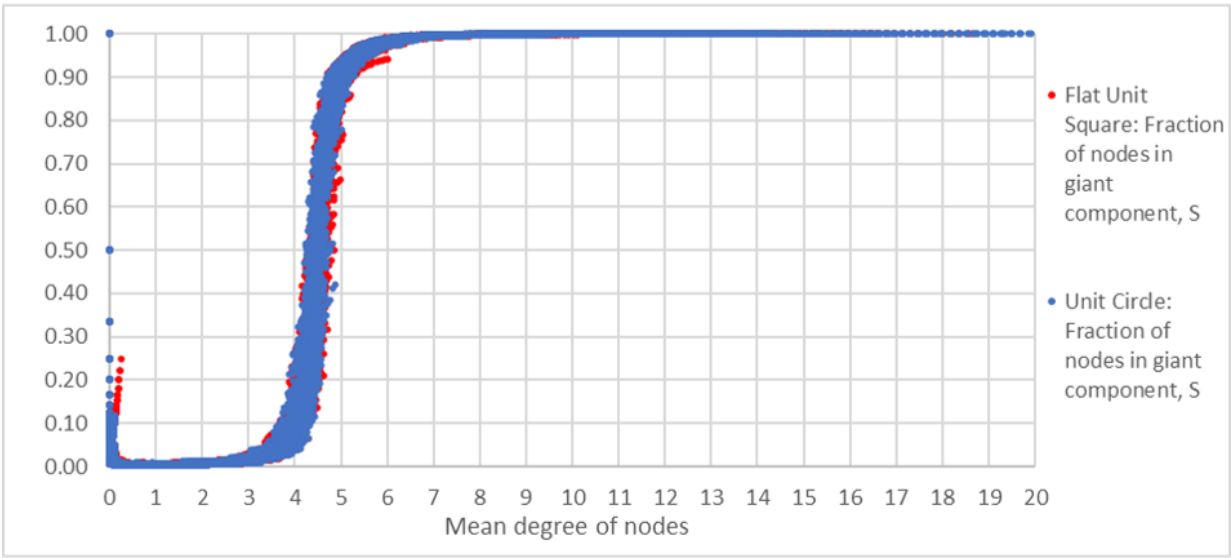

Figure 17: Overlay of scatter plots of S, as a function of mean degree, shown in Figure 15 for RGGs in a unit square and Figure 7 for RGGs in a flat unit square.

3.3 Examples of r-for-given-n Growth

Figure 18 demonstrates the evolution of the size of the GC in the r-for-given-n growth model for RGGs in a flat unit square with $n = 15{,}000$ nodes. The scatter plot of S for the GC for 100 simulated geometries, where the range r is increased until full connectivity is achieved, shows that the phase transition occurs at a range of approximately $r = 0.01$; this is consistent with this situation being the complement of the configuration for the n-for-given-r growth model used for Figure 6.

Figure 19 shows how GC size grows as a function of r for configurations in a toroidal unit square with $n = 15{,}000$ nodes. The midpoint of the phase transition occurs at slightly less than $r = 0.01$; this is consistent with the observation that the phase transition in the complementary problem shown in Figure 10 occurs at slightly less than the value of $n = 15{,}000$ nodes observed for RGGs in a flat unit square.

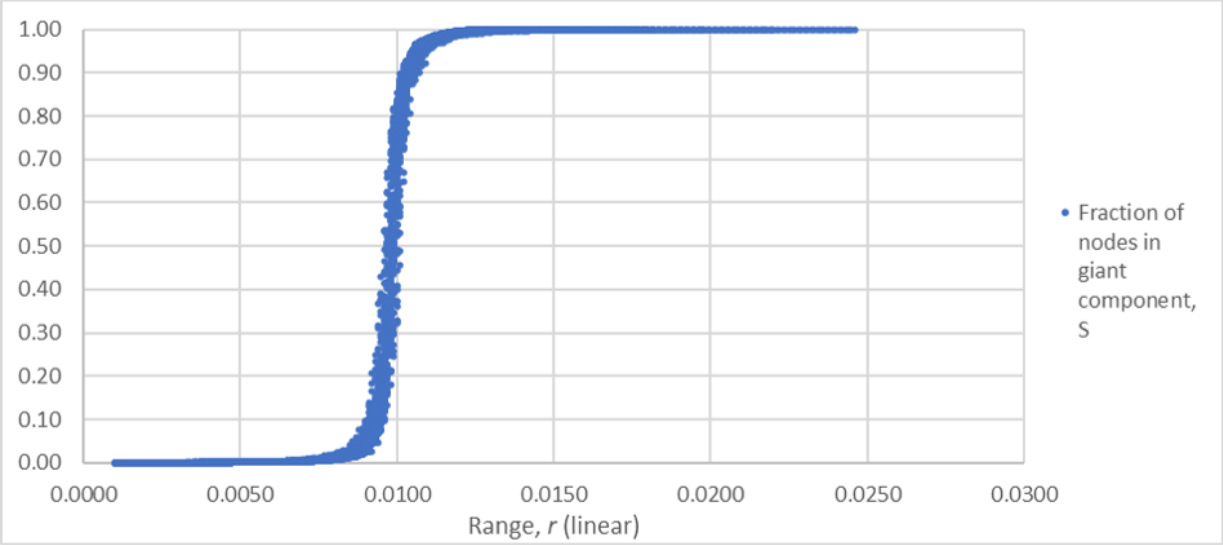

Figure 18: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a flat unit square with 15,000 nodes and r-for-given-n growth. S is plotted as a function of the range.

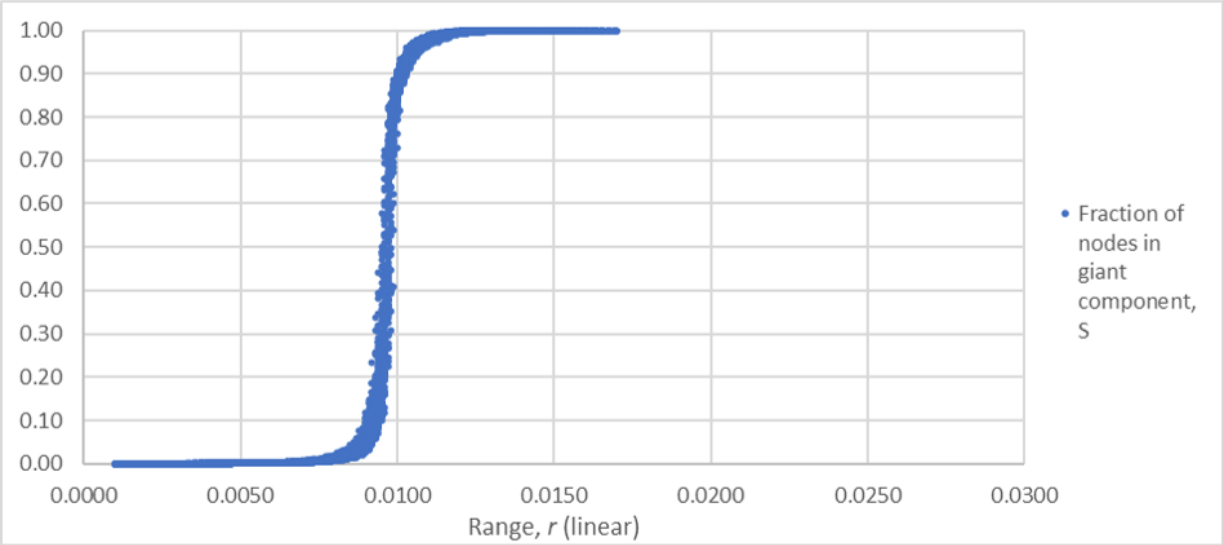

Figure 19: Fraction of nodes S in the GC for 100 simulated geometries of RGGs in a toroidal unit square with 15,000 nodes and r-for-given-n growth. S is plotted as a function of the range.

3.4 Estimating the Probability of No Isolated Nodes in Two-Dimensional Spaces

Section 3.2 and Section 3.3 focused on the size of the GC as a function of the number of nodes n in the n-for-given-r growth model or the range r in the r-for-given-n growth model; however, the phase transition in the size of the GC indicates only the *start* of the region where full connectivity (or even only 99% connectivity) occurs. In the figures shown in those sections, it can be seen that the scatter plots must continue well beyond the value of n or r at which the phase transition occurs to first achieve the condition of full connectivity.

Most work in this paper addresses the goal of determining the value of n or r for which $S = 1$ (full connectivity) with probability $C = 0.99$. Results are also obtained for the relaxed goal of $S = 0.99$ or 99% connectivity. This paper also considers the condition where it is required that RGGs be fully connected with probability $C = 0.01$, meaning the RGG is *disconnected* with probability 0.99.

A necessary condition for an RGG to be fully connected is that there are no isolated nodes. As a result, many models in the literature for RGG connectivity are based on the probability that there are no isolated nodes. This probability is a *lower bound* on the probability that an RGG is fully connected but it is frequently used *as* the probability that an RGG is fully connected under certain conditions as discussed later.

This section considers only the case of two-dimensional toroidal space. Section 7.2.1 addresses the problem of estimating the probability of no isolated nodes for the general case of d-dimensional toroidal spaces.

The simplest expression for the probability that a randomly generated geometry for an RGG configuration contains no isolated nodes is obtained when border effects can be ignored, which is true in the case of a toroidal unit square. In an RGG in a toroidal space the mean density of nodes is constant throughout the space. For unit spaces of area $A = 1$ the mean node density is given by $\frac{n}{A} = \frac{n}{1} = n$.[21]

In two-dimensional space the area of a circle with radius r is

$$\pi\, r^2 \quad . \tag{8}$$

This elementary expression for the area of a circle is introduced for consistency with Section 7.2.1 where (8) is generalized to spaces of arbitrary dimension d.

With mean node density n and omnidirectional range r the mean number of neighbors per node or mean degree μ within the circle defined by that range for any one node is

$$\mu \quad = \quad \pi\, r^2 n \quad . \tag{9}$$

With the mean number of neighbors per node given by (9) the probability that any one node has exactly y neighbors is given by the Poisson distribution:

$$\Pr\{\text{any one node has exactly } y \text{ neighbors}\} \; = \frac{\mu^y\, e^{-\mu}}{y!} \quad . \tag{10}$$

Ignoring border effects, the probability that an individual node in an RGG is isolated is given by (10) with $y = 0$ as

$$\begin{aligned} \Pr\{\text{any one node is isolated}\} \quad &= \quad e^{-\mu} \\ &= \quad e^{-\pi\, r^2 n} \quad . \end{aligned} \tag{11}$$

[21] The density of nodes is constant throughout the space when averaged throughout the space. As examples show in Section 3.1, a smaller subarea selected in the space may have a greater or lesser node density than the overall average.

Assuming statistical independence, the product of the complements of the probabilities of individual node isolation given by (11) yields the probability or confidence C of the joint event that (ignoring border effects) *no node* in an RGG is isolated:

$$\begin{aligned} C &= \Pr\{\text{no isolated nodes}\} \\ &= \left[1 - e^{-\pi\, r^2 n}\right]^n \quad . \end{aligned} \tag{12}$$

Expression (12) is an example of a solution to the underlying question of the probability of connectedness (in this case, the probability of having no isolated nodes) without consideration of the specific RGG growth model, as outlined in Section 1.5.

For the r-for-given-n growth model, (12) can be solved for r as a function of n and C:

$$r = \sqrt{\frac{-\ln(1 - C^{1/n})}{\pi\, n}} \quad . \tag{13}$$

This result was obtained in similar ways by others; for example, see Bettstetter [Bettstetter (2002a)], his expression #9. These derivations are shown to demonstrate that (12) and expressions obtained from it are strong results for the case where border effects are ignored as they depend only on the minimal assumptions of Poisson distribution of nodes and statistical independence.

There is no closed-form expression equivalent to (13) that is a solution for (12) for n as a function of r and C; however, (12) is easily solved for n by numerical means such as a binary search which was done for the work described here.

3.5 Penrose's Theorem

Penrose [Penrose (1999)] [Penrose (2003)] generalized the problem of determining the probability of node isolation and connectivity for RGGs in toroidal spaces of arbitrary dimension d for the r-for-given-n growth model. A toroidal space was assumed so that border effects could be ignored.

Penrose presented an argument that, starting with a set of nodes and a small r where all nodes are isolated and then increasing that range, an RGG becomes fully connected "at the same instant"[22] the last isolated node becomes connected "with high probability for n large" (Theorem 13.17 of [Penrose (2003)]). In other words, he argued that the disappearance of isolated nodes in an RGG in toroidal space with high probability for large n would be a sufficient condition for full connectivity as well. Penrose did not address the question as to how large n had to be for this statement to hold but this paper explores this problem extensively. Penrose's Theorem 13.17 is referred to in this paper as *Penrose's First Claim*.

For the special case of two-dimensional space and for 1-connectivity (the only kind of connectivity considered in this paper), he obtained the following result involving the probability or confidence C that an RGG has no isolated nodes (Corollary 13.20 of [Penrose (2003)]), which is stated in the notation of this paper as:

> For an RGG without border effects to contain no isolated nodes,
>
> $$C \;=\; \Pr\{\; n\pi\, r^2 - \ln n \le c_P \;\} \qquad (14)$$
>
> where C and c_P are related as
>
> $$\begin{aligned} C &= \exp(-\exp(-c_P)) \\ &= e^{-e^{-c_P}} \qquad \text{for } -\infty < c_P < \infty \\ &\quad \text{or, inversely,} \\ c_P &= -\ln(-\ln(C)) \qquad \text{for } 0 < C < 1\ . \end{aligned} \qquad (15)$$

[22] Bettstetter states this as "*at the same moment*" (italics his) [Bettstetter (2004)].

Penrose's Corollary 13.20 is referred to in this paper as *Penrose's Second Claim*.

Based on the First and Second Claims, Penrose then proposed that (14) and (15) provide sufficient conditions that an RGG in toroidal space is fully connected for large n with probability C. This is referred to in this paper as *Penrose's Third Claim*.

Penrose's three claims outlined in this section are referred to jointly in this paper as *Penrose's Theorem* for two-dimensional space.

Penrose's Theorem is explored in Section 7.2.2 for the general case of d-dimensional spaces.

The leaps made in Penrose's three claims have been used by many researchers in their models (for example, see [Bettstetter (2004)] [Koskinen (2003a)]. This paper tests these claims for RGGs in the toroidal spaces assumed by Penrose as well as under other circumstances.

Although c_P in (14) and (15) can take any real value, reasonable extreme bounds on its values can be established. The probability $C = 0.999999$ ("one in a million chance there is an isolated node") requires $c_P = 13.8155$ and $C = 0.000001$ ("one in a million chance there is no isolated node") requires $c_P = -2.6258$.

Two values of c_P used throughout this paper are $c_P = 4.6001$ for the high probability of being connected defined as $C = 0.99$ and $c_P = -1.5272$ for the high probability of being disconnected defined as $C = 0.01$.

For the r-for-given-n growth model, Penrose's (14) and (15) are equivalent to saying that an RGG contains no isolated nodes (and so by Penrose's First Claim should also be fully connected for "n large") with probability C when r is given by

$$r = \sqrt{\frac{\ln n + c_P}{\pi n}} \quad . \tag{16}$$

As previously noted for (12) in Section 3.4, there is no closed-form solution for (16) for n as a function of r and C but (16) is easily solved numerically for n using a binary search.

Penrose's First Claim that an RGG becomes fully connected "at the same instant" for large n means that the numbers of nodes and the ranges predicted by Penrose's Theorem apply not only to the condition of RGGs having no isolated nodes but the stricter condition of being fully connected as well. Thus, throughout this paper, Penrose's Theorem will be said to predict those two conditions interchangeably as the theorem is tested against the results of RGG simulations.

The vast generality of the problem solved by Penrose was accomplished through a series of limit theorems. Thus, the expectation might be that the Second Claim of Penrose's Theorem could be a weaker result than the more direct and practically assumption-free (12) and (13). The two approaches, however, provide nearly identical numerical results.

For $C = 0.99$ the relative error[23] of the value of r obtained from Penrose's (16) with respect to the value of r obtained from (13) is -3.64×10^{-5} for $n = 10$ nodes, -2.73×10^{-6} for $n = 100$, -2.18×10^{-7} for $n = 1{,}000$, and -1.82×10^{-8} for $n = 10{,}000$.

Penrose's Theorem was derived for the case where there is a high probability of RGGs being fully connected, but even for smaller values of C the two approaches still provide very close values. Noting that the largest difference occurs with small n, for $n = 10$ the relative error of the value of r obtained from Penrose's (16) with respect to the value of r obtained from (13) is -6.39×10^{-3} for $C = 0.50$ and -1.18×10^{-1} for $C = 0.01$.

[23] The term "error" is not intended to imply that one value is more correct than the other. The term is used here only to identify the type of comparison performed and the values are so close that it is immaterial as to which is chosen as the reference value for the relative error calculation.

Even more convincing than these numerical comparisons, observe that Penrose's (16) can be derived directly from (13). Expression (13) contains the term $1 - C^{1/n}$ which can be closely approximated as $-\frac{\ln C}{n}$ in the limit as n increases.[24] Replacing $1-C^{1/n}$ by $-\frac{\ln C}{n}$ in (13) yields exactly Penrose's (16).

This paper tests the validity and applicability of Penrose's Theorem in many later sections. Two points regarding Penrose's Theorem are made here in advance of the detailed results.

The first point is that the statement that an RGG is inevitably fully connected "at the same instant" the last isolated node becomes connected for "n large" in a toroidal space with r-for-given-n growth does not hold for small numbers of nodes. For small n, this statement is invalidated by the counterexample shown in Figure 3 and its other versions with $n = 38$. For that RGG, as described in Section 3.1, when the range r increased from 0.2000 to 0.2385 the last isolated node at $(0.9901, 0.8145)$ was absorbed. However, the remaining component of two nodes was not joined to the GC until r reached 0.2527 and the node at $(0.8690, 0.1522)$ was connected to the node at $(0.7184, 0.3552)$ in the GC to achieve full connectivity.

This RGG was not contrived as a counterexample to the First Claim of Penrose's Theorem but was the very first instance generated as an example of the GC for the n-for-given-r and r-for-given-n growth models; the fact that it disproved Penrose's statement above for small n was noticed only later.

Penrose's First Claim is shown not to hold for another RGG with larger n. Figure 20 shows a geometry consisting of 561 nodes scattered randomly in a flat unit square demonstrating Poisson distribution of nodes. Figure 21 shows the same geometry where the common range r of the nodes is set at 0.0700; this is another example of n-for-given-r growth where the RGG is not fully connected when the last isolated node disappears. With $r = 0.0700$ and 561 nodes in the RGG, all of the nodes are in the GC (the green dots) except for a component consisting of three nodes (the red dots) with

[24] Even though this is a known approximation, it does not seem to be listed in commonly available tables of limit theorems. It is derived by keeping the first two terms of the series expansion of the exponential function with base a, $a^x = e^{x \ln a} = \sum_{i=0}^{\infty} \frac{(x \ln a)^i}{i!}$, taking $a = C$ and $x = 1/n$. For $C = 0.99$ the relative error of the approximation is less than 0.1% for $n \geq 7$. For $C = 0.50$ the relative error of the approximation is less than 1% for $n \geq 35$ and less than 0.1% for $n \geq 347$. For $C = 0.01$ the relative error of the approximation is less than 1% for $n \geq 232$ and less than 0.1% for $n \geq 2{,}305$.

no link to the GC. When added, the 562^{nd} node (denoted with a large blue dot) would connect the smaller component to the GC, although its links are omitted for clarity.

The preceding counterexample demonstrates that the disappearance of isolated nodes does not guarantee full connectivity of RGGs with n-for-given-r growth, at least for an RGG of this size.[25] This randomly generated geometry was the third one produced for this configuration so this phenomenon is not rare.[26]

The second point is that, in spite of the counterexamples above that show Penrose's First Claim does not hold for small n, the point of full connectivity of RGGs in both the n-for-given-r and the r-for-given-n growth models may occur closely enough to the point where the last isolated node disappears in toroidal unit square RGGs that the approximation has small error for $C = 0.99$. Thus, it performs very well in exactly the situation for which it was formulated. The First Claim of Penrose's Theorem assumes that n is large but a finding shown in this paper is that the Third Claim of Penrose's Theorem works well for toroidal unit square RGGs even for small n. Penrose's Theorem is not as accurate when border effects are considered, however.

It was shown in this section that (12) and Penrose's Theorem are almost indistinguishable numerically and with the use of a single approximation they are the same expression. For that reason, results regarding the situations where Penrose's Theorem works well also apply to (12).

[25] One could point out that the counterexamples based on RGGs with only 38 or 561 nodes does not disprove Penrose's First Claim for larger RGGs. But, jumping ahead in the discussion, later sections in this paper show that the probability of RGGs in two-dimensional spaces becoming fully connected when the last isolated nodes disappeared was not observed to reach or even approach 1 within the r and n size limits of these simulations. The probability reached only 0.804 for flat RGGs and 0.832 for toroidal RGGs with $r = 0.000251$, which was the smallest range simulated for n-for-given-r growth. The probability reached only 0.789 for flat RGGs and 0.821 for toroidal RGGs with $n = 10{,}000{,}000$ nodes, which was the largest number of nodes simulated for r-for-given-n growth. By any standard, these would be considered large RGGs.

[26] More precisely, it was rather improbable that it should have taken as many as three tries to find such a case. This is discussed in detail in footnote #38 in Section 4.2.1.6.

The value c_P in (16) for toroidal unit square RGGs is independent of n and is determined only by C. However, for flat unit squares Koskinen [Koskinen (2004)] applied regression analysis to the results of simulations of RGGs with r-for-given-n growth to obtain functions of C and n to serve in place of c_P in (16) to closely estimate the range required to achieve full connectivity of RGGs in these spaces. This is outlined in Section 3.8 and discussed extensively in Section 5.4.

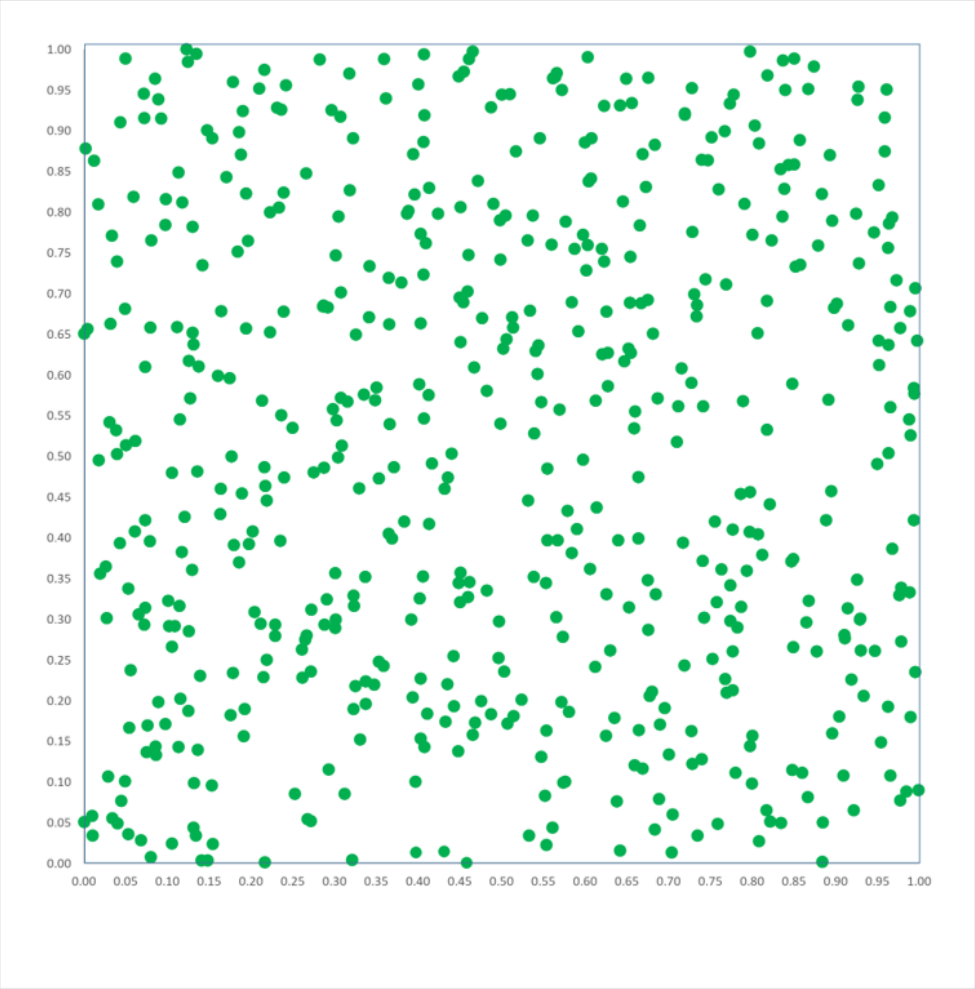

Figure 20: Geometry consisting of 561 nodes demonstrating Poisson distribution of nodes in a unit square. Communication range is unspecified in this figure.

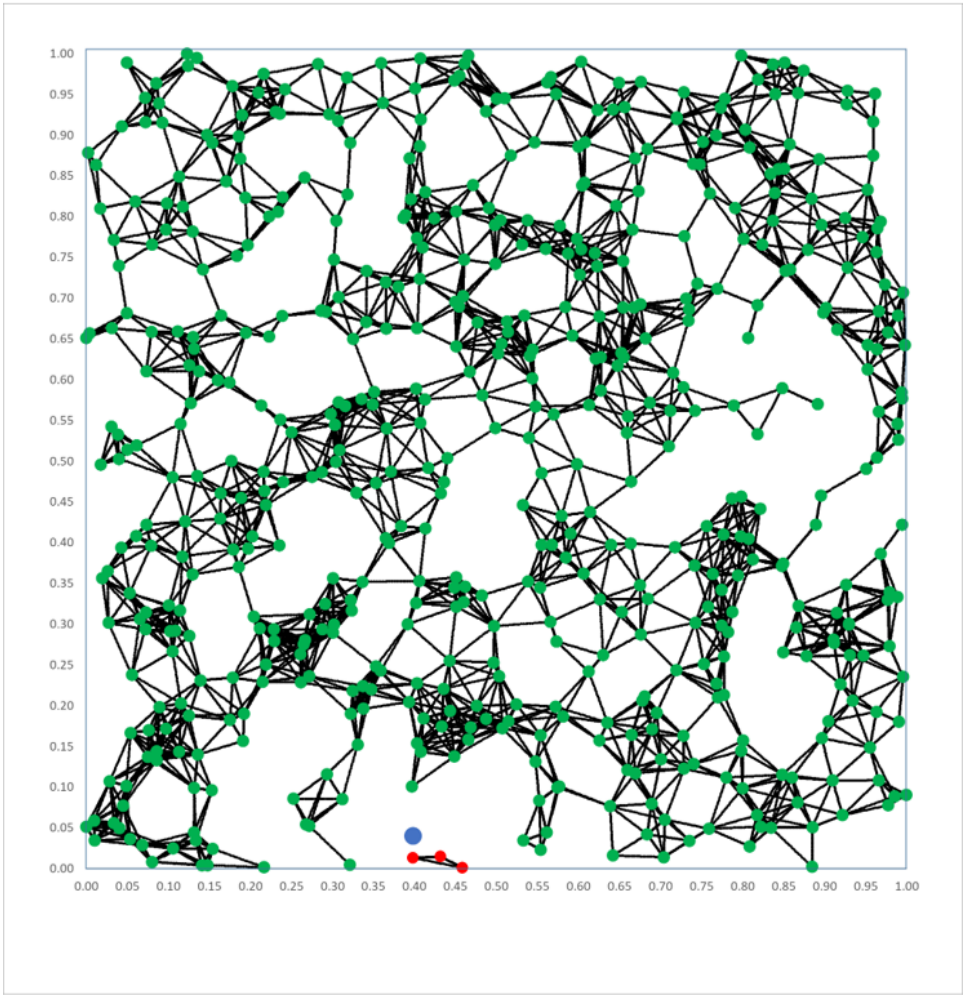

Figure 21: Example of n-for-given-r growth where the RGG is not fully connected when the last isolated node disappears. The same 561 nodes shown in Figure 20 are included in this figure. With 561 nodes and range 0.0700, all of the nodes are in the GC (green dots) except for a component consisting of three nodes (red dots). When added, the 562^{nd} node (large blue dot) would bridge the smaller component to the GC.

3.6 Gupta and Kumar's Theorem

P. Gupta and P.R. Kumar studied the connectivity of RGGs in a unit circle. They obtained the following result (Theorem 3.2 of [Gupta (1998)]) which is stated in the notation of this paper as:

> For a unit circle without border effects, an RGG with range r such that
>
> $$r = \sqrt{\frac{\ln n + c_{GK}(n)}{\pi\, n}} \tag{17}$$
>
> is fully connected with probability 1 as $n \to \infty$ if and only if $c_{GK}(n) \to \infty$.

This statement is referred to in this paper as *Gupta and Kumar's Theorem*.

Gupta and Kumar's RGG growth model is r-for-given-n. Their derivation is based on the disappearance of isolated nodes and uses percolation results by Meester and Roy [Meester (1996)] as the basis of the assumption that an RGG becomes "almost surely" fully connected when the range is increased to the point there are no isolated nodes remaining as the number of nodes goes to infinity. They disregarded border effects in the derivation of their theorem. An appendix in their paper considered border effects but that analysis was not used to modify their Theorem 3.2.

Gupta and Kumar's Theorem is frequently cited in the RGG literature,[27] but its application is problematic as their theorem does not provide a specific function $c_{GK}(n)$ to be used in (17) or otherwise describe the behavior of r as n and $c_{GK}(n)$ both go to infinity.

These questions can be answered for Gupta and Kumar's Theorem because its statement includes the strict requirement of full connectivity "with probability 1." This requirement is interpreted here in its literal sense to mean $C = 1$ or absolute certainty, and thus it must apply to *all* future cases. Consider an extreme situation where an RGG with $n \geq 2$ nodes in a

[27] As of 20 May 2022, Google Scholar reports that [Gupta (1998)] has been cited 1,893 times in the literature.

unit circle with radius $a = 1/\sqrt{\pi} \approx 0.5642$ has all of its nodes placed in one of exactly two components located at the endpoints of a maximal length chord: the diameter of the circle.[28] Because any RGG in a unit circle must be fully connected with probability 1 when the range r is at least the diameter of the circle the *worst-case* range r_{max} is equal to the diameter. Thus, $r_{max} = 2\,a = 2/\sqrt{\pi} \approx 1.1284$, for any number of nodes $n \geq 2$.[29]

The fact that such a geometry is constructed rather than occurring by chance is irrelevant because of the requirement "with probability 1." As unlikely as this random placement of nodes would be for even small values of n, this geometry provides an achievable upper bound on the value of the range that is both necessary and sufficient to guarantee full connectivity of any RGG in a unit circle with probability 1.

Stated another way, setting the range to be the diameter of the unit circle would satisfy the "if" part of their theorem (sufficiency) and no smaller range would satisfy the "only if" part (necessity).

The worst-case value of $r_{max} = 2/\sqrt{\pi}$ can be used to obtain the function $c_{GK}(n)$ that satisfies Gupta and Kumar's Theorem. Solving (17) for $c_{GK}(n)$ when $r = r_{max}$ yields

$$c_{GK}(n) = 4\,n - \ln n \quad . \tag{18}$$

Evaluating (18) for $n = 10$ and 100 yields $c_{GK}(10) = 37.70$ and $c_{GK}(100) = 395.39$.

Function $c_{GK}(n)$ defined as (18) goes to infinity as n goes to infinity, as required by Gupta and Kumar's Theorem, which validates the "only if" part of their theorem as it is the smallest function that provides r_{max}.

[28] Without loss of generality, to aid visualization imagine the RGG is rotated about its center at $(0, 0)$ such that the two components are at the antipodal (x, y) coordinates $(-a, 0)$ and $(a, 0)$.

[29] This RGG also demonstrates that a flat circle without consideration of border effects is not the same as a "toroidal circle." Without toroidal wraparound the minimum range for a flat unit circle that guarantees two components of antipodal nodes are connected is $r_{max} = 2\,a$. If the circle were toroidal, then any $r > 0$ would be sufficient to guarantee two antipodal components in this space would be connected because they would be adjacent to each other across the border. Furthermore, it can be shown that a "toroidal circle" with two or more nodes scattered anywhere in the circle would be fully connected with any $r \geq a$.

On the other hand, the "if" part of their theorem is shown not to hold because, for example, the candidate function $c_{GK}(n) = 3\,n$ goes to infinity but is not large enough (that is, its resulting range is not sufficient) to guarantee that every RGG in a unit circle is fully connected with the required "probability 1."

Gupta and Kumar's (17) resembles Penrose's (16), the only difference being that Gupta and Kumar's $c_{GK}(n)$ is for the condition of full connectivity with $C = 1$ and is a function of n where Penrose's c_P depends only on C where $C < 1$.

Penrose's derivation was for RGGs in a toroidal unit square and Gupta and Kumar's derivation was for RGGs in a unit circle without consideration of border effects but it is instructive to brush those nuances aside and contrast $c_{GK}(n)$ and c_P.

Gupta and Kumar's Theorem states that $c_{GK}(n)$ increases as n increases and must go to infinity to provide full connectivity with probability 1. Although Penrose does not address connectivity with the extreme confidence of $C = 1$, his (15) yields the value of c_P required to achieve a given level of confidence that an RGG (without border effects) is fully connected if (16) is valid.

As noted in Section 3.5, for the very high probability of full connectivity of $C = 0.999999$ Penrose's Theorem requires only $c_P = 13.8155$ regardless of n. In contrast, even for a small RGG with only $n = 10$ nodes the much larger value of $c_{GK}(10) = 37.70$ would be required by Gupta and Kumar's (18) if one accepts the argument presented above regarding the worst-case value of r_{max} for full connectivity with probability 1.

The estimate of a value of r_{max} can be reduced if a relaxed definition of "the worst-case range observed for no isolated nodes in a billion randomly generated geometries" is accepted in lieu of the actual minimum range needed to be "fully connected with probability 1"

Trials[30] were run with various numbers of nodes n placed uniformly at random positions in unit circles where the sample size (number of trials)

[30] The term *trial* is used here because these were not full simulations of RGG growth but only measurements made of the relative random placements of the nodes.

was $N = 10^9$. For the RGG generated in each trial, the minimum distance of every node to its nearest neighbor was obtained, and the maximum of those distances equaled the minimum range that would achieve no isolated nodes in that RGG; the maximum of the N minimum ranges obtained in that way can be taken as a working approximation of r_{max} for that set of trials. Let the value of r_{max} obtained for given values of n and N be denoted as $r_{max}(n, N)$.[31]

For $n = 10$ the observed value $r_{max}(10, 10^9) = 0.9867$ yields a bound of $c_{GK}(10) = 28.29$. For $n = 100$ the observed value $r_{max}(100, 10^9) = 0.4069$ yields a bound of $c_{GK}(100) = 47.40$. These values are much smaller than those obtained using the diameter of a unit circle as r_{max} and can be taken to be the worst-case values of $c_{GK}(n)$ if the largest range observed required for no isolated nodes in a billion trials is accepted as a surrogate for that needed to for any RGG to be "fully connected with probability 1."

The observed values of the worst-case ranges required resulting from these large numbers of trials are discussed further in Section 5.5.

One of the few efforts in the RGG literature to obtain expressions for a function similar to $c_{GK}(n)$ for full connectivity for various values of C less than 1 is that by H. Koskinen [Koskinen (2004)]. Koskinen applied regression analysis to the results of simulations of RGGs with r-for-given-n growth in flat unit squares to arrive at functions of n that would serve in place of Penrose's c_P in (16) or Gupta and Kumar's $c_{GK}(n)$ in (17). This is outlined briefly in Section 3.8 and discussed extensively in Section 5.4.

[31] Although these maxima would vary over multiple sets of trials the values given here were obtained simply as estimates of worst-case ranges required.

3.7 Consideration of Border Effects

Many, if not most, analyses in the RGG literature explicitly consider only toroidal spaces or simply neglect the border effects.

Aside from the work by H. Koskinen [Koskinen (2004)] mentioned in Section 3.6 and discussed extensively in later sections, one of the few exceptions is the work of C. Bettstetter [Bettstetter (2002b)] [Bettstetter (2004)] who explicitly addressed border effects for nodes in a circle.[32]

Bettstetter presented approaches for obtaining the probability that an RGG with a given r and n contains no isolated nodes based on explicit consideration of the number of neighbors within the range of border nodes as a function of those nodes' distances from the border. He presented a detailed model based on integration and also outlined a simpler model based on summing terms based on discrete rings rather than the continuous summation of an integral [Bettstetter (2002b)] [Bettstetter (2003)] [Bettstetter (2004)].

Both methods were implemented during the course of the research described here and it was observed that Bettstetter's results were always within small percentages of the results obtained from simulations.

[32] Notably, Bettstetter also analyzed the connectivity of RGGs consisting of mobile nodes.

3.8 Koskinen's Regression Analyses

As mentioned in Section 3.6, H. Koskinen appears to be one of very few researchers to apply regression analysis to the results of simulations involving RGGs. Koskinen made significant progress in two areas.

The first area was explored in his Master's Thesis [Koskinen (2003a)] and a subsequent paper [Koskinen (2003b)] where he proposed a model for a transformation of statistical metrics for the range required to achieve full connectivity of RGGs in a flat unit square with r-for-given-n growth. He obtained an LRM for the reciprocal of the square of the mean range as well as several percentiles of the range, as a function of n, that has an extraordinarily good fit over the values of n that were simulated.

The second accomplishment [Koskinen (2004)] was the development of regression estimates of functions of n to replace c_P in (16) or functions to serve as Gupta and Kumar's $c_{GK}(n)$ as defined in their Theorem 3.2 and used in (17) of this paper. Koskinen obtained empirical measurements of c_P as functions of n for RGGs in a flat unit square with r-for-given-n growth. He obtained expressions for adjustments to c_P as functions of n that achieved full connectivity with various values of C less than 1.

These two achievements are described in detail and tested in Section 5.4.

4 RGGs that Grow by Adding Nodes with Fixed Range: The n-for-given-r Growth Model

This section presents results for the n-for-given-r growth model. In this model of RGG growth, nodes with fixed range r are added uniformly at random positions in a space until a desired connectivity condition is achieved, at which point the process stops.

Expressions are obtained for several statistical metrics for the number of nodes required to *first* achieve[33] the condition of full connectivity ($S = 1$), 99% connectivity ($S = 0.99$), or having no isolated nodes.[34] The statistical metrics are the 99%ile, mean, and 1%ile number of nodes for those conditions. The shapes of the nodal spaces modeled are the two-dimensional flat unit square, toroidal unit square, and flat unit circle. These spaces do not contain obstructions.

Variations of the n-for-given-r growth model are considered in subsequent sections. Section 7 considers nodal spaces of dimensions $d = 1$ to 6, Section 8 addresses flat unit rectangular nodal spaces, and Section 9 introduces line-of-sight constraints created by regularly and randomly placed obstructions.

The empirically based approach used in this paper is to simulate many randomly generated geometries for selected configurations and use linear regression to obtain the desired predictive models.

The n-for-given-r growth model may not have received as much attention in the RGG literature because it is not as amenable to analysis as the r-for-given-n growth model is by methods such as those employed by Penrose.

[33] The split infinitive phrase "to first achieve" is not ungrammatical as there is no actual English grammar rule against its use but only a preference that it be avoided. The split infinitive construction should be used when it is the clearest and most unambiguous expression of a concept [Nicholson (1957)].

[34] The reasons for using the number of nodes that first achieve a desired connectivity condition are given in Section 4.1.3.

4.1 Simulations and Data Analysis

Section 4.1.1, Section 4.1.2, Section 4.1.3, Section 4.1.4, and Section 4.1.5 contain introductory material that is common to many other sections. Those other sections highlight where the methods used differ from those described here.

4.1.1 Simulators

Several computer programs were written to simulate and report the growth and connectivity of RGGs. All of the programs were written in the C programming language.

Two primary optimizations were used. The first improved the search for pairs of connected nodes by localizing the searches to small regions of near neighbors based on the current range. The second was that the simulators found the equivalence classes using the very fast "weighted quick-union algorithm" given as Program 1.3 by Sedgewick [Sedgewick (1998)].

The simulations of RGGs with r-for-given-n growth were implemented as binary searches using simulations of n-for-given-r growth.

The localized search was implemented as follows. When adding an n^{th} node to an RGG with range r, rather than brute-force measuring the distance from the n^{th} node to the $n-1$ previously placed nodes the unit square is subdivided into $\lfloor 1/r^2 \rfloor - 1$ bins that are $1/r$ on a side (the last bin is larger to allow for fractional bins). In this way, a node in a bin need have its distance calculated with respect to only those nodes in its own bin and adjacent bins that may contain the node's near neighbors. This extends naturally to higher dimensions.

The distance is calculated appropriately according to whether the space is flat and toroidal as described in Section 1.4.

For independent verification, where possible, RGG configurations and data reported in the literature by prior researchers were simulated and compared to the results of these simulators.

In spite of the optimizations and speedups that were incorporated the runtimes for small ranges or large numbers of nodes were quite long. Fortunately, a 16-processor and 32-processor server were made available for this work as well as other High Performance Computing (HPC) resources. Even so, many of the graphs each represent weeks of wall clock time on these servers.

4.1.2 Simulations

Four sets of ranges were used for the configurations that were simulated and tested: a mid-range set, a large-range set, a small-range set, and the union of the mid-range and small-range sets. Because the LRMs vary slightly based on the sets over which they were derived, relative errors are calculated and presented for each set of ranges.

Unless otherwise specified, each LRM for RGGs with n-for-given-r growth is based on the simulation results for the *mid-range set (mid-r set)* (also called the *mid ranges*), which consists of values of the range r decremented from 0.1000 down to 0.0050. Starting from $r = 0.1000$ the range is decreased geometrically with the multiplier $10^{-1/10}$ so there are ten sampled values of r per decade. The ranges are rounded to a precision of four places after the decimal point.

Thus, the 14 ranges in the mid-r set in decreasing order are $r = 0.1000$, 0.0794, 0.0631, 0.0501, 0.0398, 0.0316, 0.0251, 0.0200, 0.0158, 0.0126, 0.0100, 0.0079, 0.0063, and 0.0050.

As most of the research reported in the RGG literature assumes the number of nodes n is large, the majority of the configurations used in this paper are based on the mid-r set so that n will be large.

For more practical purposes smaller numbers of nodes must be considered, which means that configurations with larger ranges must be used. For this purpose, another set is used where the smallest range is 0.1000 and the largest range is selected to be slightly larger than the diagonal of a unit square which has length $\sqrt{2}$. For unit squares, any range greater than or equal to $\sqrt{2}$ would yield the same results.

Starting from $r = 0.1000$ the range is increased geometrically with the multiplier $10^{1/10}$ so there are ten sampled values of r per decade. The ranges are rounded to a precision of four places after the decimal point. The smallest range larger than $\sqrt{2}$ generated in this way is $r = 1.5849$.

Thus, the 13 ranges in the *large-range set (large-r set)* (also called the *large ranges*) in decreasing order are $r = 1.5849$, 1.2589, 1.0000, 0.7943, 0.6310, 0.5012, 0.3981, 0.3162, 0.2512, 0.1995, 0.1585, 0.1259, and 0.1000.

At the other extreme, for more theoretical purposes even larger numbers of nodes than those required for the mid-r set must be considered, which means that configurations with even smaller ranges must be used. The third set of ranges used in this paper starts from $r = 0.010000$ and the range is decreased geometrically with the multiplier $10^{1/10}$ so there are ten sampled values of r per decade but the first (meaning the largest) value included in the set is 0.005012. The ranges are rounded to a precision of six places after the decimal point.

Thus, the 14 ranges in the *small-range set (small-r set)* (also called the *small ranges*) in decreasing order are $r = 0.005012$, 0.003981, 0.003162, 0.002512, 0.001995, 0.001585, 0.001259, 0.001000, 0.000794, 0.000631, 0.000501, 0.000398, 0.000316, and 0.000251.

The union of the mid-r set and small-r set is referred to as the *mid+small-range set (mid+small-r set)* (also called the *mid and small ranges*). For consistency, all of the ranges in this set are rounded to a precision of six places after the decimal point.

In some cases, subsets of these sets of ranges are used.

The results based on the mid-r set are based on separate sets of simulations from those based on the mid+small-r set. The results for the small-r set are taken from the simulations for the mid+small-r set. Because separate sets of simulations were performed for some ranges, statistical metrics and estimates of probabilities may vary slightly throughout the sections that follow.

Except where noted the sample size for each configuration's value of r was $N = 1{,}000$ simulated geometries, which is discussed in Section 4.1.4.

Uniformly distributed random numbers used in these simulations were produced using the "ran2" random number generator given by [Press (2002)]. Press et al. made this confident statement:

> "We think that, within the limits of its floating-point precision, ran2 provides perfect random numbers; a practical definition of "perfect" is that we will pay $1,000 to the first reader who convinces us otherwise (by finding a statistical test that ran2 fails in a

> nontrivial way, excluding the ordinary limitations of a machine's floating-point representation)."

As of the time of this writing, there does not appear yet to have been a payoff against this challenge.

As a precaution against potential initialization issues such as that famously suffered when the RC4 stream cipher was misused in the Wired Equivalent Privacy (WEP) of IEEE 802.11 [Edney (2004)], a "warm up" phase is used where the first 1,000 values generated by the ran2 random number generator are discarded after it is initialized with a random seed.

The term *random seed* means "the seed for the random number generator." The random seeds were generated systematically to avoid duplication except where intended.

For the n-for-given-r simulations the simulated geometries for each value of r used different random seeds but the simulations for the flat and toroidal unit squares used the same sets of random seeds by design. The sets of nodes were generated prior to each simulation and padded so that the random number generator was synchronized at the start of the generation of each simulated geometry. This allowed the differences between flat and toroidal RGG growth for *exactly the same sets of node positions* to be observed. Thus, the comparisons of the statistical metrics between those two nodal spaces are based not only on the same *distributions* of nodes but, *a fortiori*, on the same *sets* of nodes.

Unless otherwise noted, each simulation of a geometry continued until the condition of 100% connectivity was reached. Results of simulations that are taken beyond that point are presented in Section 4.1.3. The numbers of nodes or ranges for all of the other connectivity conditions (no isolated nodes, 50% connected, or 99% connected) were found on the way to full connectivity. In this way the comparisons of those statistical metrics are also based not only on the same distributions of nodes but on the same sets of nodes.

The simulations were performed using "embarrassingly parallel processing" [Moler (1986)] [Messina (1987)] on multiprocessor servers. The full sets of 1,000 simulated geometries for each configuration's value of r were partitioned into smaller sets (20, 50, or other convenient numbers of subsets depending on which server was available). The random seeds for each set were allocated appropriately to ensure each of the 1,000 simulations were unique as well as the seeds being synchronized between the flat and toroidal simulations as outlined above.

4.1.3 Adding Nodes Can Break Full Connectivity of RGGs with n-for-given-r Growth

A fully connected RGG with n-for-given-r growth may have a new node added that is more than distance r from any node already in the GC, thus creating a disconnected RGG with an isolated node. For the same reason, a geometry with the lesser condition of no isolated nodes can also have an isolated node introduced.

The terms *break* or *broken* are used here to mean that an RGG that has reached a desired connectivity condition loses that condition with the addition of nodes.

Two issues relating to the breaking of a connectivity condition of RGGs with n-for-given-r growth are discussed in this section.

The first issue was where it was observed with n-for-given-r simulations that, early in the growth process, on occasion a small number of nodes would cluster closely enough to form a GC that contained all of the nodes added so far but would not contain the next node added. In other words, a very small RGG would be fully connected early in the growth process but the addition of another node would break invariably break that condition. This happened most often in the simulations based on the large-r set and the largest values of the mid-r set. It was also observed that even a large fully connected RGG could become disconnected with the addition of a node. The same phenomena occurred with small and large RGGs with no isolated nodes where that condition was broken with the addition of a node. These issues and their impact on the simulation process are discussed in this section.

The phenomenon where adding nodes can break a fully connected RGG or RGG with no isolated nodes with n-for-given-r growth illustrates an intrinsic difference between the two growth models. On the other hand, a fully connected RGG or RGG with no isolated nodes with r-for-given-n growth cannot become disconnected or introduce an isolated node when range increases.

To avoid the problem of an n-for-given-r growth simulation terminating with a pathologically small number of nodes, a lower limit was set on the size of a fully connected RGG such that the simulator would not stop and

report that the RGG was "fully connected" until that limit was reached. After some experimentation the *minimum size* of a fully connected RGG was set at 25 nodes for all of the n-for-given-r simulations reported in this paper except where noted. The phenomenon was infrequent enough, however, that the results of the statistical metrics were nearly indistinguishable from those obtained with the minimum size set at five or even smaller numbers of nodes.

Figure 22 shows scatter plots of two sets of simulations that report the number of nodes required to first achieve full connectivity in a flat unit square for $N = 1{,}000$ simulated geometries per configuration. The configurations used subsets of the mid-r and large-r sets (for a total of 14 values of $r = 0.0100$ up to 0.1995). The random seeds are the same for the two sets of simulations and node generation for every simulation is padded to generate the same number of nodes regardless of the number of nodes at which the simulation terminates. Blue dots represent simulations where the minimum component size is constrained to be at least 25 nodes, and red dots represent simulations where the minimum component size is allowed to be only two nodes. The blue dots overlap perfectly the red dots for each individual simulation except for the cases where the minimum component size of two nodes resulted in simulations terminating at full connectivity with small numbers of nodes.

For the 14,000 simulations depicted in Figure 22 where simulations were allowed to terminate at a minimum component size of two nodes, 327 simulations terminated with fewer than 25 nodes in the GC. There were 311 simulations that terminated with just the first two nodes added to the RGGs. The smallest range in this set of simulations where "full connectivity" was obtained with a pathologically small number of nodes was with $r = 0.0158$ where one simulation terminated upon adding just the first two nodes.

The second issue was where even a large fully connected RGG with n-for-given-r growth can be broken by the addition of a node. To determine roughly how frequently this can occur, sets of simulations were run with geometries in a flat unit square where nodes continued to be added beyond the point of first becoming fully connected. The total number of nodes added for each set of simulations was set at double the 99%ile number of nodes required to first achieve full connectivity for the given range for that set as given by (22) in Section 4.2.1.1. For each of the values $r = 0.1000$ down to

0.0200 in the mid-r set as listed below, 100 simulations were run[35] and a count was made of the number of times a fully connected RGG was broken by the addition of a node.

The results were as follows:

- For $r = 0.1000$, 9 RGGs were broken.
- For $r = 0.0794$, 5 RGGs were broken.
- For $r = 0.0631$, 12 RGGs were broken.
- For $r = 0.0501$, 10 RGGs were broken.
- For $r = 0.0398$, 13 RGGs were broken, with one RGG being broken again.
- For $r = 0.0316$, 10 RGGs were broken, with one RGG being broken again.
- For $r = 0.0251$, 11 RGGs were broken.
- For $r = 0.0200$, 7 RGGs were broken.

Based on these limited sets of simulations the indication is that in roughly 10% of the simulations a fully connected RGG with n-for-given-r growth can be broken by the addition of additional nodes. In every case, however, every RGG that was broken was restored (that is, fully connected again) at the end of the simulation.

It was noted during the development of these experiments that the percentages vary over multiple runs for a given range. It was also noted that an RGG can become disconnected and fully connected several times although those specific observations are not included in the data presented above.

Because a fully connected RGG or RGG with no isolated nodes can have that connectivity condition broken the statistical metrics obtained for any connectivity condition of RGGs with n-for-given-r growth are based on the *first* number of nodes that achieve that condition, given that the specified minimum size of the RGG has been reached as described above in this section.

[35] Only 100 simulations per value of r were run for these experiments because of the volume of output created to get these results.

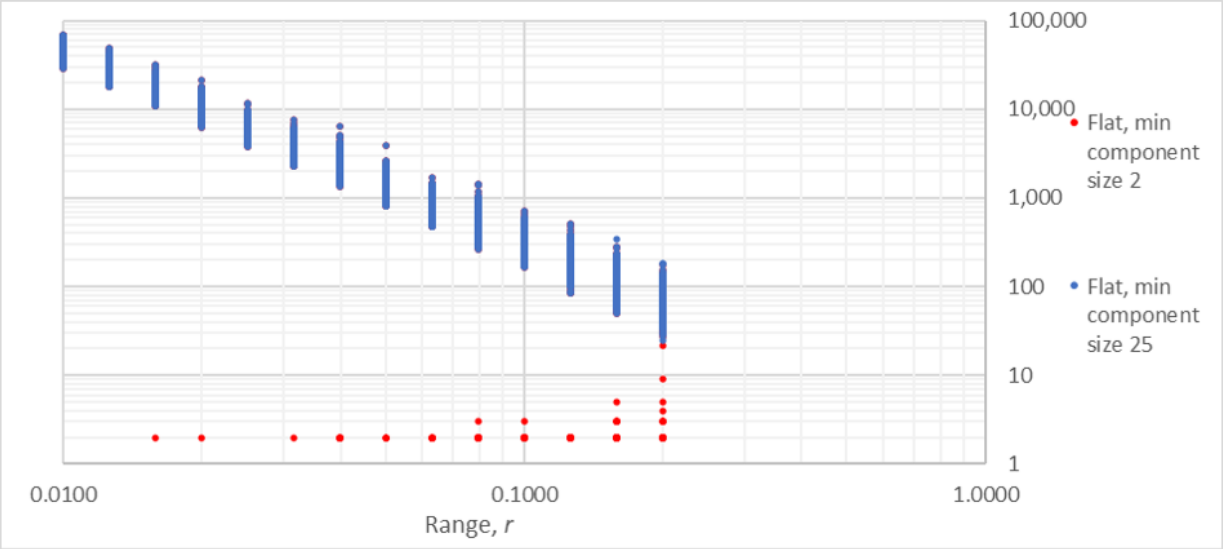

Figure 22: Two scatter plots of number of nodes required to first achieve full connectivity in a flat unit square for $N = 1{,}000$ simulated geometries per configuration. Blue dots represent simulations where the minimum component size is constrained to be at least 25 nodes, red dots represent simulations where the minimum component size is allowed to be only two nodes. Each plot contains 14,000 points; all of the red dots are covered precisely by the blue dots except for 327 simulations where the RGGs were fully connected with fewer than 25 nodes in the GC.

4.1.4 Sample Size for Randomly Generated Geometries

The decision to use a sample size of $N = 1{,}000$ randomly generated geometries for each configuration's value of r (except where noted) balanced the runtimes of the simulations against the variations in the models during the design of the simulations. The choice of $N = 1{,}000$ simulations per value of r is supported by Figure 23 which shows the ± 2 standard error bounds[36] around the *mean number of nodes* required to first achieve full connectivity in a flat unit square. The smallest points that can be plotted on the graph for the lower and upper bounds overlap so closely at each value of r that they appear to be single points. Figure 24 shows the ratio of the two standard error bounds to the mean value where the ratio drops from about 1.7% at $r = 0.1000$ to about 0.9% at $r = 0.0050$.

The equivalent statistical measure of result variation for percentiles is quite involved and is so seldom used that the additional credibility it would provide is minimal. For that reason the ± 2 standard error bounds about the mean are the primary evidence presented here that the sample size of $N = 1{,}000$ simulations per value of r is sufficient to obtain accurate statistical metrics.

However, another way to qualitatively evaluate the adequacy of sample size $N = 1{,}000$ to capture extreme values from the simulations is to use a result by S.S. Wilkes, cited by Duncan [Duncan (1986)], that gives the probability ϵ that a random sample of size N from a data set contains a specified fraction β of the range of data in the set:

$$N\,\beta^{N-1} - (N-1)\,\beta^{N} = 1 - \epsilon \quad . \tag{19}$$

Expression (19) indicates that a sample of size $N = 1{,}000$ contains 99.5% of the range of the full population with probability 0.960. The sample contains 99.6% of the range with probability 0.909 and contains 99.7% of the range with probability 0.801.

[36] The term "standard error" in statistics refers to the standard deviation of a set of samples divided by the square root of the number of samples [Altman (2015)] [Weisstein (2019a)].

4.1.5 Linear Regression Models, Goodness-of-Fit, and Errors

Section 4.2, Section 4.3, and Section 4.4 present LRMs for three nodal spaces as outlined in Section 4: the two-dimensional flat unit square, toroidal unit square, and flat unit circle. Differences between flat and toroidal unit squares are noted. For completeness the results of the LRMs are compared to those predicted by Penrose's model even though the n-for-given-r RGGs that grow by adding nodes with fixed range do not fit Penrose's approach that assumes the range increases over a fixed set of nodes; this comparison is performed again in Section 5 where the r-for-given-n growth model is used.

The statistical tool used is linear regression based on transformed input and output variables. It is shown that the log of the number of nodes given by several statistical metrics is very closely a linear function of the log of the range r. That is to say, the statistical metrics closely fit straight lines when plotted as a function of the range on log-log scales. The term *power law* is commonly used for such functions even without rigorous tests such as those outlined by Clauset et al. [Clauset (2009)].

The term *error* by itself in this paper always means *relative error*; it is explicitly stated when *absolute error* is used. The absolute error of a predicted or estimated value with respect to a reference value is (predicted − reference). The relative error is the absolute error divided by the reference value or (predicted − reference)/reference. The percentage relative error is (predicted − reference)/reference × 100%.

In this paper the predicted values are those estimated by the model and the reference values are the observed values obtained from the simulation results (the true values are of course unknown). Thus, a positive error means a predicted value overestimates an observed value and a negative error means a predicted value underestimates an observed value. The squared residuals are equal to the squared absolute errors.

The goodness-of-fit of the LRMs was determined by their *Coefficient of Determination* or R^2 values. The models' predictive or forecasting power was determined using *Mean Absolute Percentage Error (MAPE)* and the maximum relative error (called max relative error or max error in this paper) over the specified set of values of r.

R^2 is a commonly used measure of goodness-of-fit of a linear regression. It describes how much of the variation of the output variable is explained by the LRM. R^2 takes values from 0 to 1, where closer to 1 is a better fit [MathWorks (2020)]. The R^2 values presented in this paper are calculated by the spreadsheet program and are not "adjusted R^2" values. The R^2 values reported in the text of this paper are calculated using the spreadsheet formula and they sometimes differ in the third or fourth decimal place from the trend line values shown on the plots. This is a known issue with the spreadsheet program and makes no practical difference.

MAPE is the average of the absolute values of the percentage relative errors of the model's predicted values for the output variable with respect to either the true values (if known) or the observed values of the output variable [Kim (2016)]. MAPE is not used to obtain an LRM: it is used only to choose among viable candidate models on the basis of the quality of their predictive ability. Linear regression obtains coefficients for a model by minimizing the Sum of the Squared Residuals (SSR) between the estimated values from a candidate model and the observed values being modeled.

The bias reduction techniques for power law LRMs as described by Newman [Newman (1993)] were tested but had no effect on the quality of the regression estimates so they were not used for any of the LRMs presented in this work.

The MAPE and max relative error values given for each of the LRMs are calculated on the basis of the same data used to generate the LRMs. To gauge the effect of testing based on training data or overfitting to noise, in Section 4.2.1.7 MAPE and max error are calculated for the LRMs obtained for RGGs in a flat unit square for the 99%ile number of nodes to first achieve full connectivity and no isolated nodes on the basis of 32 additional runs of 1,000 simulated geometries for the mid-r set outlined in Section 4.1.2. The same is done for one run of a much larger set of 10,000 simulated geometries for each range in the mid-r set. It is shown that MAPE and max error based on the "training data set" are optimistic but not drastically so.

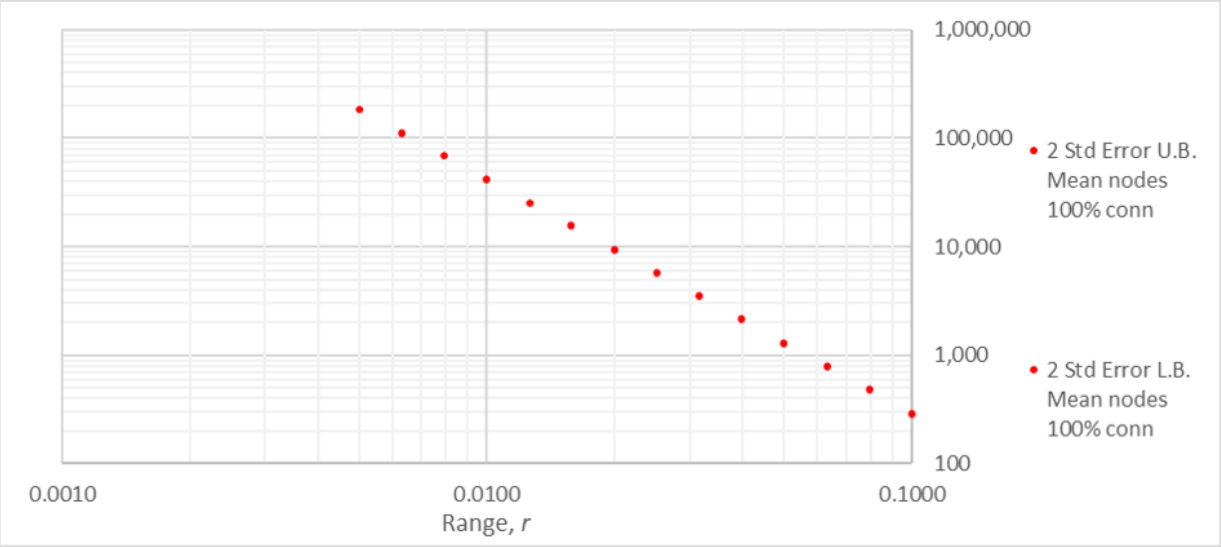

Figure 23: Upper and lower "two standard error bounds" around the mean number of nodes for full connectivity in a flat unit square for $N = 1{,}000$ simulated geometries per configuration. The upper and lower bounds are so tight they appear to be single dots on top of the mean value.

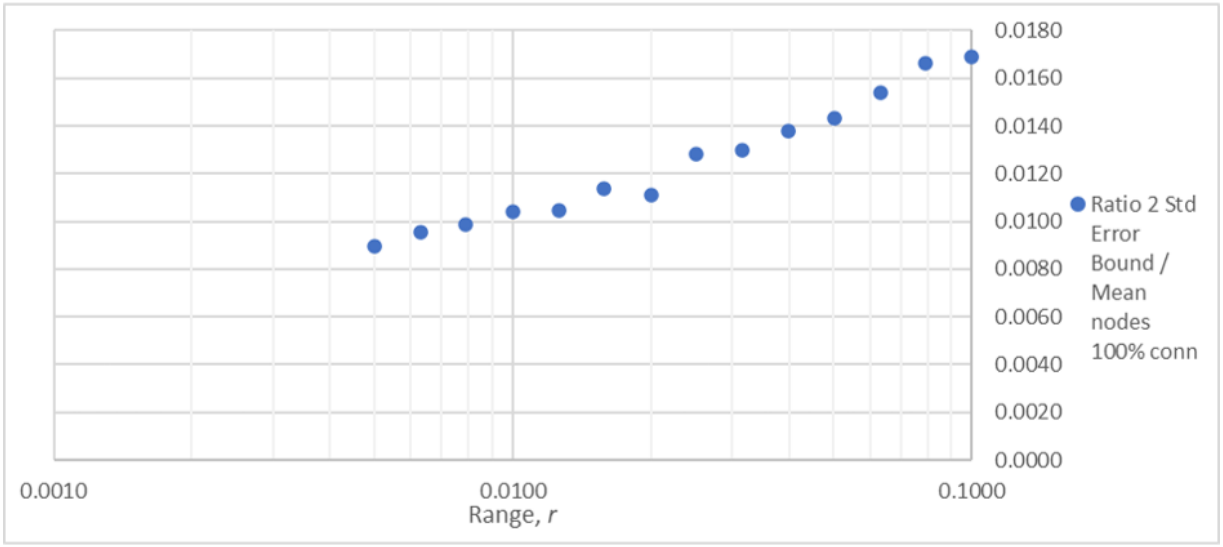

Figure 24: Ratio of the value of two standard errors to the mean number of nodes for full connectivity for the upper and lower bounds shown in Figure 23.

4.2 Flat Unit Square with n-for-given-r Growth

This section considers RGGs consisting of nodes scattered uniformly at random positions in a two-dimensional square space of area $A = 1$ with no wraparound at the border. This is a flat space so there are border effects. The RGGs evolve according to the n-for-given-r growth model.

Figure 25 demonstrates the consequence of the border effects experienced by RGGs in flat spaces. For configurations with $r = 0.0100$, it shows the position where the last node was added to achieve full connectivity for each of 1,000 simulated geometries. It can be seen that the last node added tends to be near the border of a flat unit square, although it might not itself be a border node. In the n-for-given-r growth model the last node added always bridges an existing isolated node or smaller component into the GC. The fact that border nodes have fewer neighbors on average than the non-border nodes means that they are less likely to be absorbed into the GC at the time they are added.

For each simulated geometry for a given configuration, nodes with a fixed range r are added one-by-one to determine the number of nodes n required to first achieve the condition of 100% connectivity, 99% connectivity, or no isolated nodes. The 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve each of those conditions are obtained for that configuration.

Results are given separately for the mid-r, large-r, small-r, and mid+small-r sets. Section 4.2.1 presents LRMs and other results for RGGs with n-for-given-r growth in flat unit squares based on configurations using the mid-r set. Section 4.2.2 presents LRMs for RGG configurations based on the large-r set. Section 4.2.3 presents LRMs for RGG configurations based on the small-r set and Section 4.2.4 presents LRMs for RGG configurations based on the mid+small-r set. The LRMS vary slightly based on the sets of ranges over which they were obtained. Although an LRM is valid only for the set on which it is based, the relative errors of the mid-r LRMs are compared to those of the mid+small-r LRMS for that set of ranges.

Most of the results deal with the statistical metrics for full or nearly full connectivity. Section 4.2.1.3 presents LRMs for 50% connectivity for RGGs with n-for-given-r growth in flat unit squares because of the close relationship between that connectivity condition and the phase transition where the single large GC emerges.

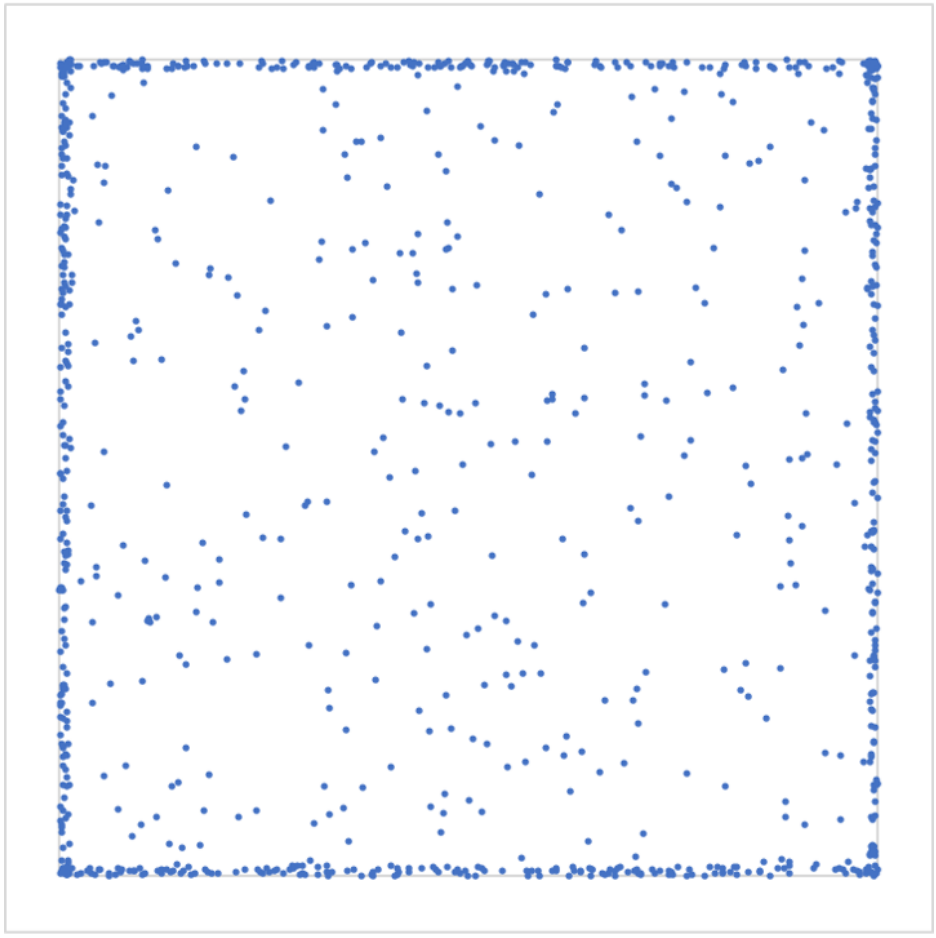

Figure 25: Position where the last node was added for full connectivity of each of 1,000 simulated geometries for RGGs in a flat unit square with range 0.0100 and n-for-given-r growth. Border effects cause the positions of the last nodes added to tend to cluster near the border. Although values on the axes are not shown, they span 0 to 1.

4.2.1 Flat Unit Square with n-for-given-r Growth for Mid Ranges

Section 4.2.1.1 through Section 4.2.1.7 present results for RGGs in flat unit squares with n-for-given-r growth using ranges in the mid-r set which is the set of ranges from 0.1000 down to 0.0050.

These sections give LRMs for full connectivity, 99% connectivity, 50% connectivity, and no isolated nodes.

They explore the relationship between connectivity and the mean degree or number of neighbors per node.

These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

The LRMs are tested to determine how much overfitting of the models to a specific data set impacts the error estimates.

4.2.1.1 Flat Unit Square: Full Connectivity for Mid-r Set

Figure 26 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a flat unit square with n-for-given-r growth for the mid-r set outlined in Section 4.1.2. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. The close fits to straight lines when plotted on these transformed axes show that the *power law* is a good approximation for the LRMs. The LRMs fit the values of the statistical metrics to equations of the form

$$\ln n = m \ln r + \ln b \tag{20}$$

which is expressed in customary power law form as

$$n = b\, r^m \quad . \tag{21}$$

The LRMs with the values of m and b replaced in (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.6571\, r^{-2.0678} \tag{22}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 1.85% and the max relative error of the value of the predicted n is -5.19%.

MAPE and max relative error given for (22) are calculated on the basis of the same data used to generate this LRM. To gauge the effect of testing based on training data or overfitting to noise, in Section 4.2.1.7 MAPE and max error for (22) are calculated again on the basis of 32 additional runs of 1,000 simulations for the mid-r set as well as a larger run of 10,000 simulations for each range in that set.

For the mean number of nodes required to first achieve full connectivity with n-for-given-r growth, the LRM is

$$n = 2.0601\, r^{-2.1522} \tag{23}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.60% and max error of the value of the predicted n is -1.42%.

It was noted that a value of n equal to the mean number of nodes required to first achieve full connectivity achieves full connectivity for about 55% to 60% of the RGGs.

For the 1%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, the LRM is

$$n = 1.0312\, r^{-2.2435} \tag{24}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.32% and max error of the value of the predicted n is 3.84%.

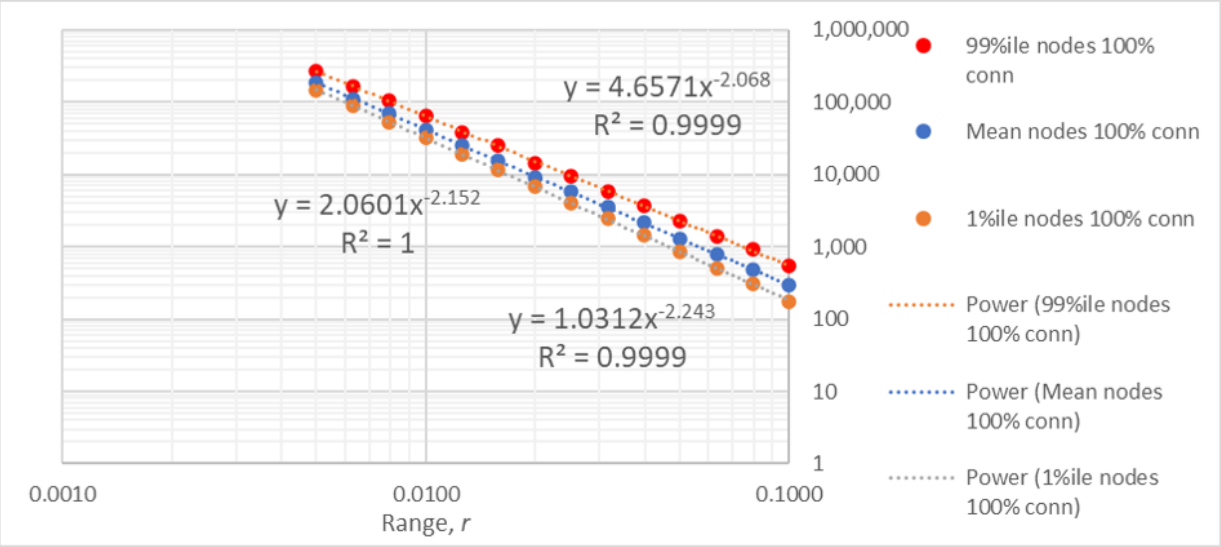

Figure 26: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth for the mid-r set.

4.2.1.2 Flat Unit Square: 99% Connectivity for Mid-r Set

As discussed in Section 1.1 of the Introduction, less-than-full connectivity may be adequate in some situations. Figure 27 shows the number of nodes required to first achieve at least 99% connectivity (the condition where $S \geq 0.99$) for RGGs with n-for-given-r growth.

This figure shows that the number of nodes required to first achieve 99% connectivity closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity with n-for-given-r growth, (21) becomes

$$n = 5.8319\, r^{-1.7908} \tag{25}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 3.27% and max error of the value of the predicted n is -7.38%.

For the mean number of nodes required to first achieve 99% connectivity with n-for-given-r growth, the LRM is

$$n = 2.7682\, r^{-1.9361} \tag{26}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.84% and max error of the value of the predicted n is -1.81%.

For the 1%ile number of nodes required to first achieve 99% connectivity with n-for-given-r growth, the LRM is

$$n = 1.5586\, r^{-2.0477} \tag{27}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.60% and max error of the value of the predicted n is 5.44%.

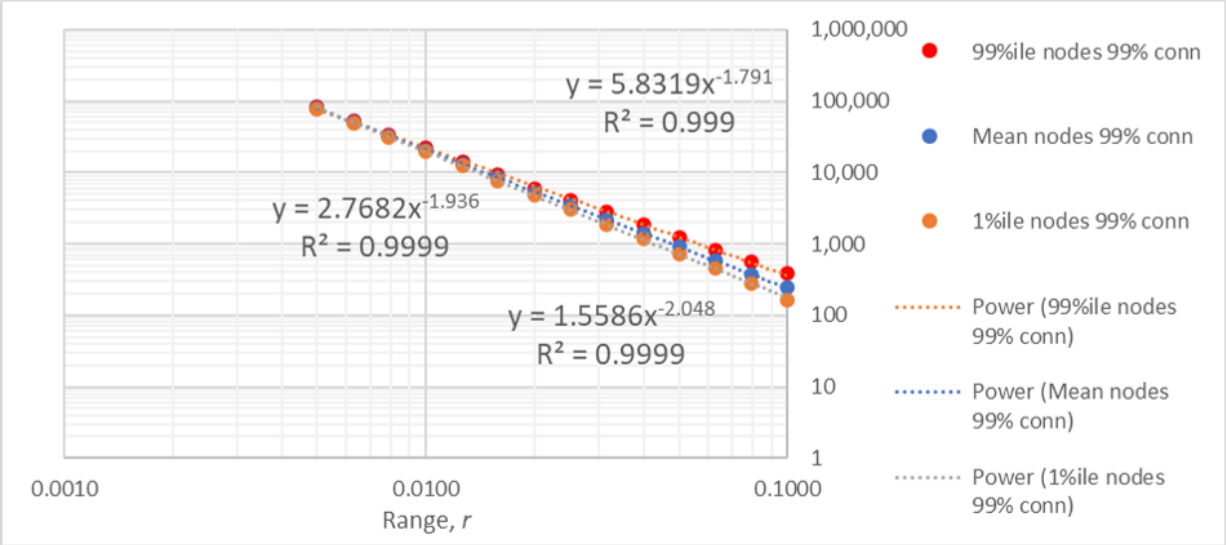

Figure 27: LRMs for 99%ile, mean, and 1%ile number of nodes for 99% connectivity of RGGs in a flat unit square with n-for-given-r growth.

4.2.1.3 Flat Unit Square: 50% Connectivity for Mid-r Set

The concept of a phase transition where small components and isolated nodes suddenly coalesce into a single large GC was discussed in Section 3.1 and examples were given in Section 3.2 for n-for-given-r growth and in Section 3.3 for r-for-given-n growth. In the RGG literature the phase transition is usually discussed in general terms of a band where the GC begins to accumulate many or most of the nodes in the RGG rather than being a discrete threshold for the fraction S of the nodes in the GC. Some analyses, however, have set specific thresholds such as $S = 0.50$, 0.95, and 0.99 [Santi (2003)] [Ta (2009)].

The emergence of the GC in RGGs with n-for-given-r growth shown in Figure 6 in Section 3.2 is very rapid between the points where S reaches 0.1 and 0.9 and those points closely bracket the narrow part of the S plot where it crosses 0.50. For that reason the number of nodes n where the value of $S = 0.50$ seems to be a reasonable threshold for the phase transition. The plot of S in Figure 6 is based on 100 simulated geometries with $r = 0.0100$ and the mean value of n for which S was above 0.50 was 14,477.64; it was noted in Section 3.2 that the midpoint was at slightly less than $n = 15{,}000$).

Figure 28 shows the number of nodes required to first achieve at least 50% connectivity (the first value of n where $S \geq 0.50$) for RGGs with n-for-given-r growth. This figure shows that the number of nodes required to first achieve 50% connectivity closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. The trend lines are given for the full mid-r set for the 99%ile and mean number of nodes to first achieve 50% connectivity but the two largest ranges (0.0794 and 0.1000) are omitted for the 1%ile statistical metric because the predicted values of n based on a power law LRM that includes those ranges deviates substantially from the observed values.

The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve 50% connectivity with n-for-given-r growth, (21) becomes

$$n = 2.2254\, r^{-1.9219} \tag{28}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.32% and max error of the value of the predicted n is -3.67%.

For the mean number of nodes required to first achieve 50% connectivity with n-for-given-r growth, the LRM is

$$n = 1.2812\, r^{-2.0275} \tag{29}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.80% and max error of the value of the predicted n is 6.29%.

For the 1%ile number of nodes required to first achieve 50% connectivity with n-for-given-r growth the LRM based on the subset of the mid-r set outlined above is

$$n = 0.8260\, r^{-2.1083} \tag{30}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 3.10% and max error of the value of the predicted n is 8.90%.

Plugging the value $r = 0.0100$ into (29), the predicted mean number of nodes required to first achieve 50% connectivity is 14,541.91, which has an error of only 0.44% from the observed mean n based on the 100 simulated geometries for that configuration.

It is perhaps surprising to have the power law model the behavior of the emergence of the GC so closely when one considers how much unpredictability there is to the shape of the GC as illustrated in the examples shown in Figure 5 and Figure 21.

Adding to the physical unpredictability of RGG growth as demonstrated in Figure 5 the numerical growth of the GC and its crossing of the threshold of 0.50 is not a smooth process. If it were a smooth process spread out over thousands of values of n, then S would tend to be close to the target value of 0.50 when it crossed that threshold; however, for the 100 simulated geometries the mean value of S when it crossed that point was not 0.50, or even a value only slightly larger, but was the far larger value 0.5652. Some of the jumps in S were quite large, as demonstrated by the fact that S for one of the simulated geometries summarized in Figure 6 jumped from 0.4629 to 0.8380 at $n = 14{,}967$ and another geometry had an even larger increase from 0.3993 to 0.7908 at $n = 14{,}607$.

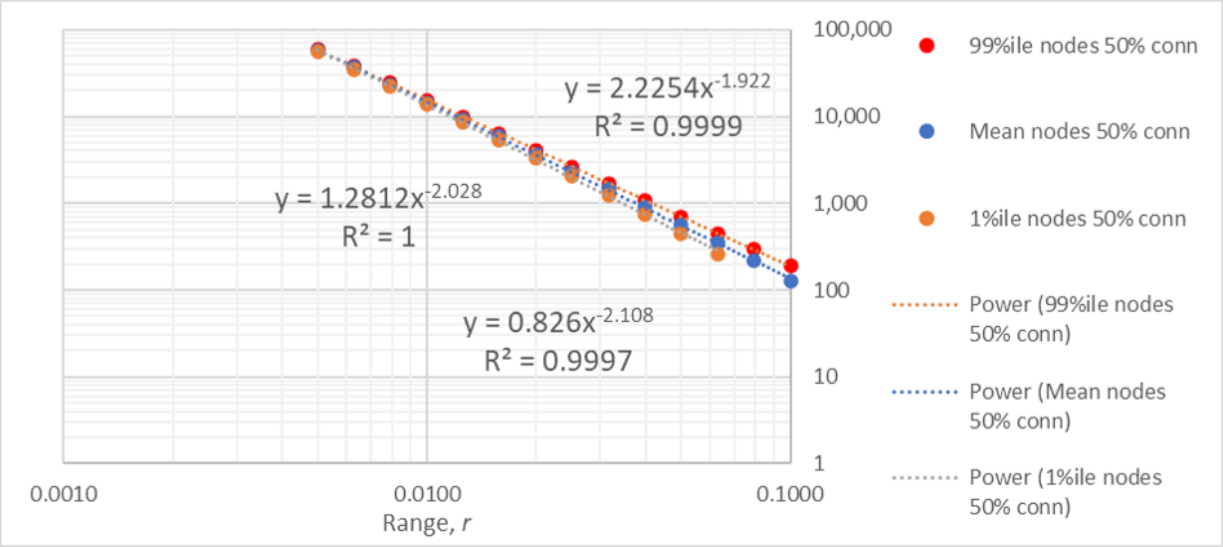

Figure 28: LRMs for 99%ile, mean, and 1%ile number of nodes for 50% connectivity of RGGs in a flat unit square with n-for-given-r growth.

4.2.1.4 Flat Unit Square: No Isolated Nodes for Mid-r Set

Figure 29 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 3.9064\, r^{-2.0969} \tag{31}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 2.25% and max error of the value of the predicted n is 6.55%.

MAPE and max error given for (31) are calculated on the basis of the same data used to generate this LRM. To gauge the effect of testing based on training data or overfitting to noise, in Section 4.2.1.7 MAPE and max error for (31) are calculated again on the basis of 32 additional runs of 1,000 simulations for the mid-r set as well as a larger run of 10,000 simulations for each range in that set.

For the mean number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 1.5378\, r^{-2.2040} \tag{32}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.42% and max error of the value of the predicted n is 3.88%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.6303\, r^{-2.3339} \tag{33}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 4.57% and max error of the value of the predicted n is 11.47%.

The number of nodes required to first achieve full connectivity in a flat unit square shown in Figure 26 and their LRMs (22), (23), and (24) are substantially different from the number of nodes required for the disappearance of isolated nodes shown in Figure 29 and their LRMs (31), (32), and (33). This emphasizes again that, for the n-for-given-r growth model, full connectivity does not necessarily occur when the last isolated node is absorbed.

This observation raises the question as to *how many* additional nodes must be added to an RGG in a flat unit square beyond the point where the last isolated node disappears. Figure 30 shows the 99%ile, mean, and 1%ile *additional* number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node.[37] The 1%ile value is exactly 1 for every range but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the mid-r set to thousands and hundreds of thousands for the smallest ranges simulated in these experiments.

Figure 30 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

[37] The count is made from the last number of nodes where there was at least one isolated node, so the minimum value is one, rather than the number of nodes since the disappearance of isolated nodes to avoid the problem of plotting the zero value on a log scale.

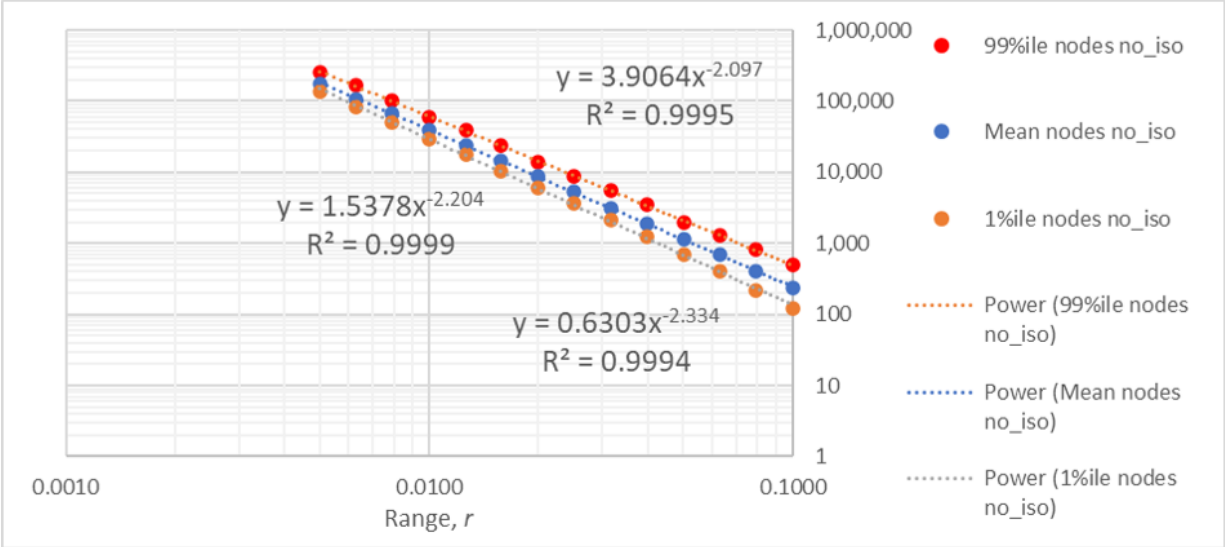

Figure 29: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth.

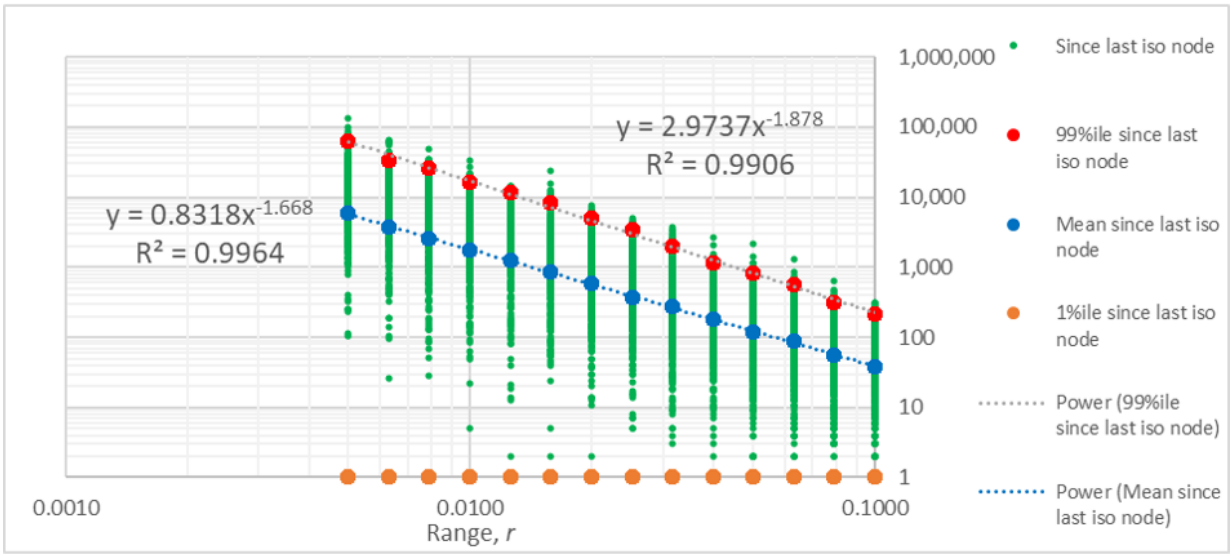

Figure 30: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.2.1.5 Flat Unit Square: Number of Neighbors Needed for Full Connectivity for Mid-r Set

The significance of the number of neighbors per node was discussed in Section 3 where it was pointed out that the phase transition from mostly disconnected to mostly connected occurs when each node has about 4.5 neighbors on average, or a mean degree of 4.5. Full connectivity, however, requires that nodes have far more neighbors on average than the degree at which the phase transition occurs. Figure 31 shows LRMs for the 99%ile, mean, and 1%ile values for the mean degree required to first achieve full connectivity of RGGs with n-for-given-r growth. Note that Figure 31 is a linear-log plot and the regression models for the statistical metrics are linear functions of the mean degree and the log of the range and thus they are modeled as log laws rather than the log-log power law functions that were shown to work well for the statistical metrics for the numbers of nodes.

Figure 32 is a scatter plot of the predicted vs. observed mean degree required to first achieve full connectivity for all of the simulated geometries using the ranges in the mid-r set. For each geometry the predicted mean degree is the value of μ given by (9) for each given range r and the number of nodes n required to first achieve full connectivity. The red line represents the value where the observed mean degree would have equaled the predicted mean degree for a simulation; it can be seen that the observed values are slightly smaller than the predicted values due to the nodes near the border of the space having fewer near neighbors on average than the non-border nodes.

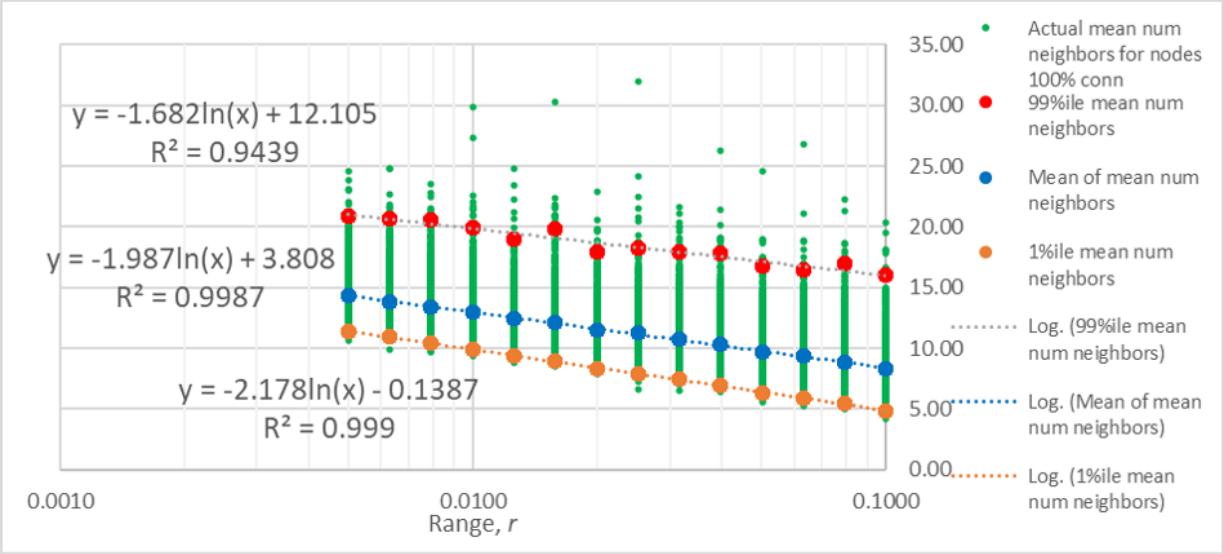

Figure 31: LRMs for 99%ile, mean, and 1%ile values for mean degree for full connectivity in a flat unit square with n-for-given-r growth.

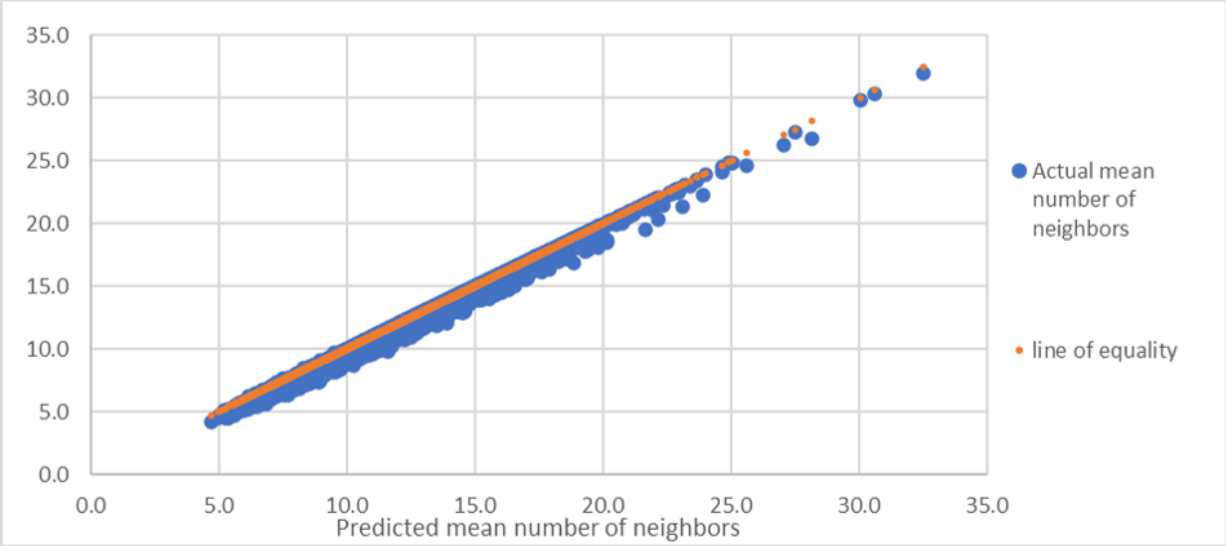

Figure 32: Observed and predicted mean degree for full connectivity in a flat unit square with n-for-given-r growth.

4.2.1.6 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a flat unit square with n-for-given-r growth using the mid-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 33 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected with n-for-given-r growth. If the First Claim of Penrose's Theorem were correct for RGGs in a flat unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a flat unit square with n-for-given-r growth, one can say that the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.328 for $r = 0.1000$ and increases to about 0.657 for $r = 0.0050$.

The LRM for this fraction closely follows a log function of the range over the mid-r set. The trend line and its LRM shown in Figure 33 provide an interpolation function for this fraction over these ranges.[38]

[38] Several similar plots are shown later in this paper but the LRM was obtained only in this case for the following purpose.

In Section 3.5 Figure 21 shows a case where an RGG is still not fully connected after the disappearance of the last isolated node. It was remarked in footnote #26 that it was unlikely that even three tries should have been needed to obtain this example. The equation for the LRM given by the trend line in Figure 33 predicts that the probability of an RGG in a flat unit square with n-for-given-r growth being fully connected when the last isolated node disappears is approximately $-0.114 \ln r + 0.0784$; this probability is about 0.382 for a configuration with $r = 0.0700$ and so the chance of finding the desired example in the first randomly generated geometry would be 0.618.

Using the geometric probability distribution for waiting time the probability of first getting such a case in the second geometry is 0.236, and the probability of first getting it in the third geometry is 0.090. The probability of not finding the desired example until a fourth or subsequent geometry is only 0.056.

Figure 33 shows that the probability of an RGG being fully connected when the last isolated node disappears increases as r decreases (and consequently the number of nodes increases) but it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a flat unit square for even the smallest range in the mid-r set.

Figure 34 relates the number of nodes needed for full connectivity to that required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. Even for small ranges, more nodes are required to first achieve full connectivity after the disappearance of isolated nodes, which is consistent with the discussion of Figure 33. The ratios may asymptotically approach 1 for even smaller ranges but the ratio is still above 1.04 for the smallest range in the mid-r set.

As noted in Section 3.5, the values $c_P = 4.6001$ for $C = 0.99$ and $c_P = -1.5272$ for $C = 0.01$ are obtained from (15).

Figure 35 shows the error of the LRM (22) given in Section 4.2.1.1 for the 99%ile number of nodes required to first achieve full connectivity of RGGs with respect to the observed number of nodes obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 4.2.1.1 that MAPE for (22) is 1.85% and max error of the predicted value of n is -5.19%.

Figure 35 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. These error values are considerably greater than those for the LRM.

Figure 36 shows the same error statistics as in Figure 35 but now for the 1%ile number of nodes required to first achieve full connectivity, which is the threshold used here for RGGs to be disconnected. Figure 36 shows the error percentages for the LRM (24) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. Again, the LRM provides close estimates while the predicted number of nodes from Penrose's expressions

has far greater errors. It must be noted, however, that Penrose's emphasis was on high levels of confidence C.

The observed values of Penrose's c_P that would satisfy (16) can be obtained from the results of the simulations for the pairs of n and r where full connectivity was first achieved and compared to the values predicted by Penrose's Theorem. Expression (16) can be solved for c_P as a function of r and n as

$$c_P = n\,\pi\,r^2 - \ln n \quad . \tag{34}$$

Figure 37 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 26.[39]

The maximum of the c_P observed for any range was 22.7870. The visible trend in Figure 37 is that all three of the statistical metrics plotted for c_P seem to decrease slightly as the range decreases and thus the number of nodes increases but even the smallest values of $c_P = 8.5345$ for the 99%ile values of n and $c_P = $-0.4440 for the 1%ile values of n are higher than Penrose's Theorem's values of $c_P = 4.6001$ required for $C = 0.99$ and $c_P = -1.5272$ for $C = 0.01$.

As a reminder, this is "off-label" use of Penrose's results. Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended solely to assist in determining when they do apply.

Koskinen [Koskinen (2004)] modeled the dependence of the observed values of c_P on n to estimate the range required to achieve full connectivity of RGGs in flat unit squares with r-for-given-n growth. This is discussed extensively in Section 5.4.

[39] The 99%ile, mean, and 1%ile values for c_P are calculated using (34) for the 99%ile, mean, and 1%ile values of n, respectively, for the given set of values of r.

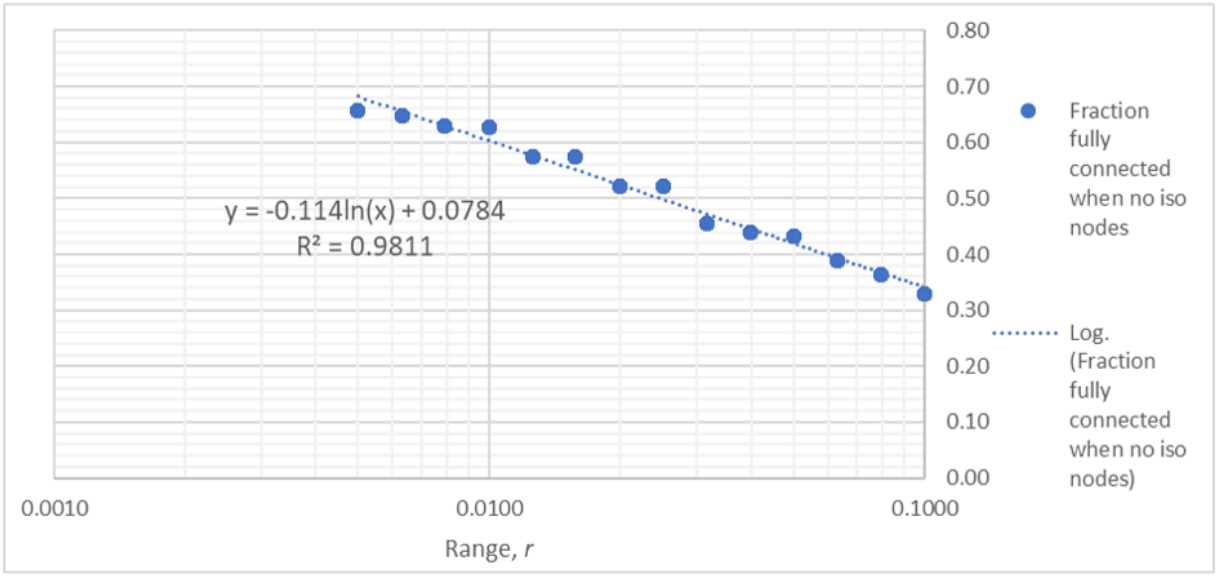

Figure 33: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a flat unit square to become fully connected with n-for-given-r growth.

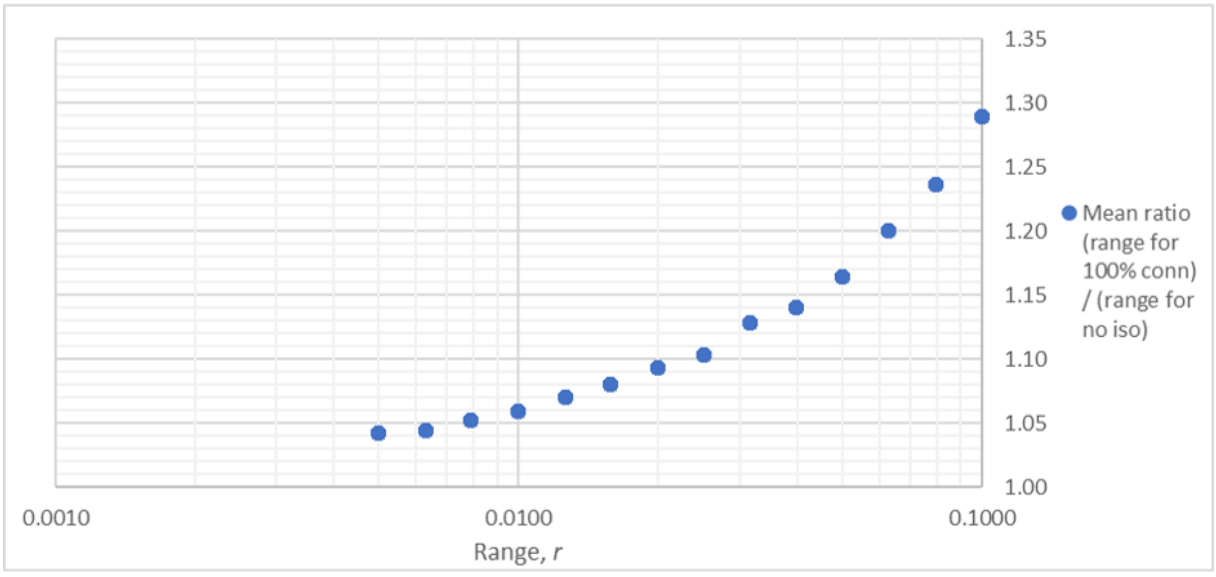

Figure 34: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a flat unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

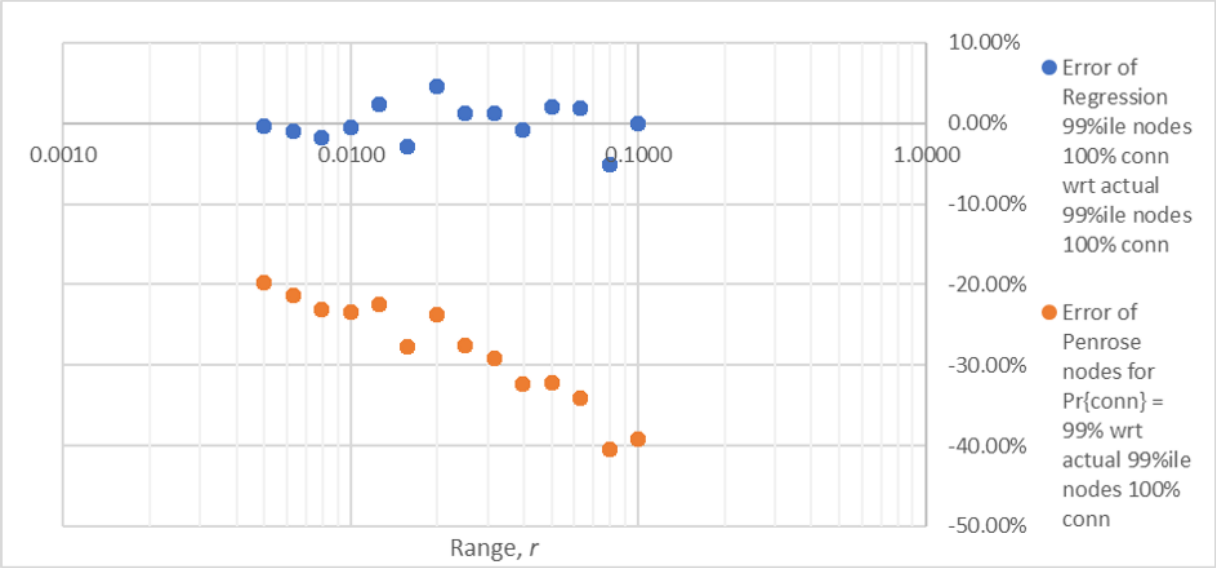

Figure 35: Relative error of LRM (22) for 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

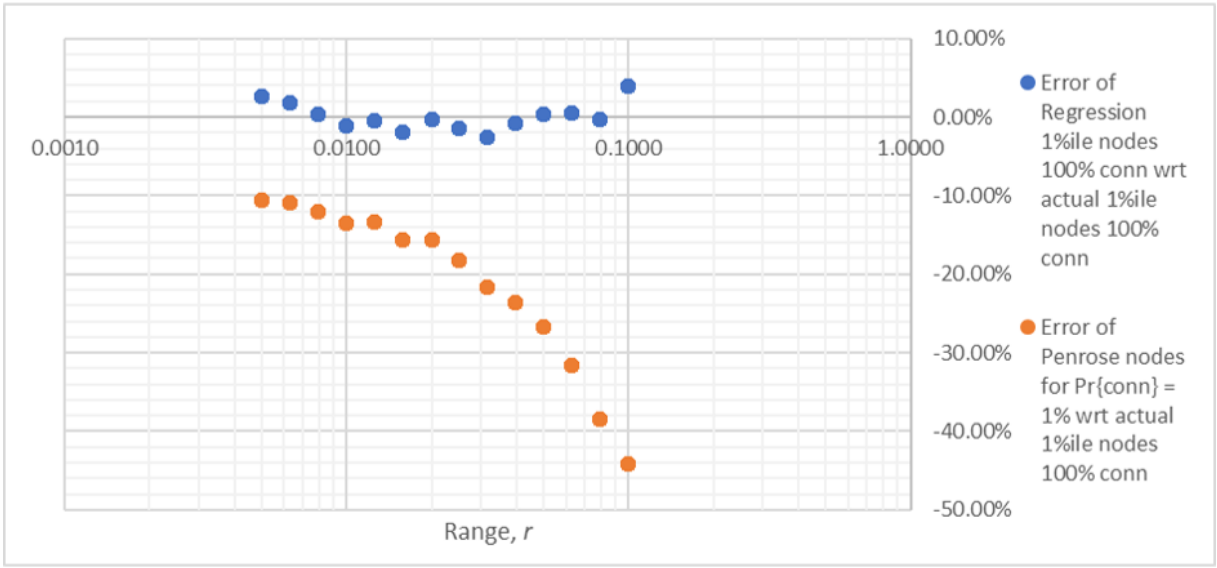

Figure 36: Relative error of LRM (24) for 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

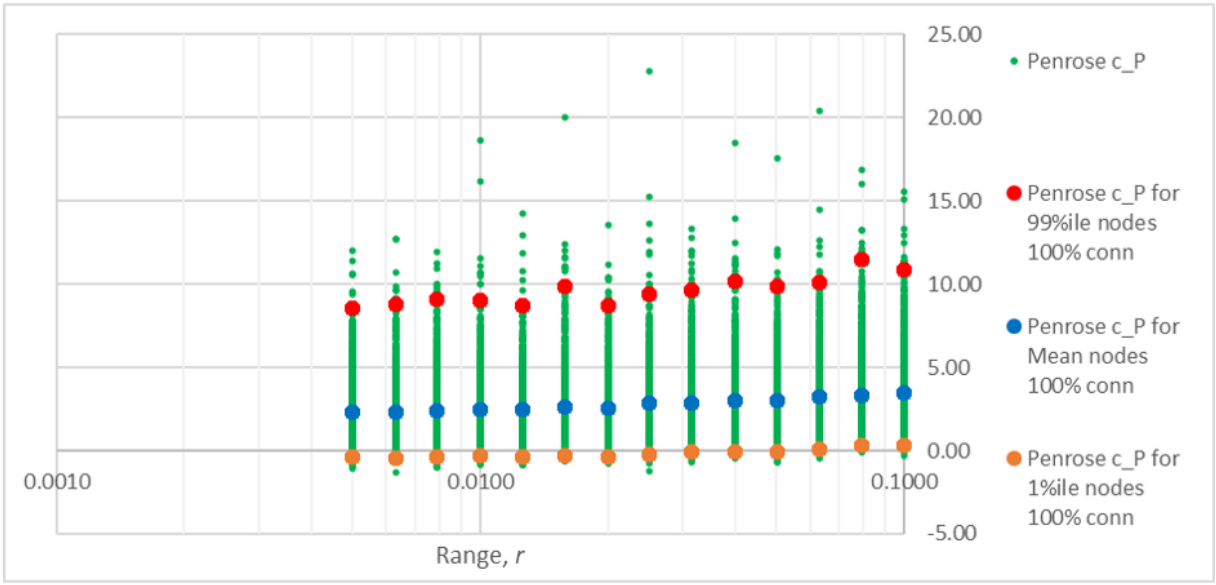

Figure 37: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.2.1.7 Flat Unit Square: Testing the Linear Regression Models for Mid-r Set

Throughout this paper the MAPE and max error values given for each of the LRMs are calculated on the basis of the same data used to generate the LRMs. In other words, these error values are based on the training data and would reasonably be expected to be lower than if the LRMs were tested on other, independently generated data sets. All data generated for a given configuration have identical distributions regardless of their random seed, as they are obtained from the same RGG growth processes, but there is inevitable overfitting due to the LRMs "learning" the noise in the measurements. The question is whether the expected variations in the statistical metrics result in excessively optimistic error values compared to those obtained from additional runs.

In this section MAPE and max error are calculated for the LRM (22) given in Section 4.2.1.1 for RGGs in a flat unit square for the 99%ile number of nodes required to first achieve full connectivity and (31) given in Section 4.2.1.4 for the 99%ile number of nodes required to first achieve no isolated nodes. Both are calculated on the basis of 32 additional runs of 1,000 simulations for the mid-r set.

MAPE and max error are also calculated for (22) and (31) for another run of 10,000 simulated geometries for each range.

Figure 38 and Figure 39 show scatter plots of the 99%ile number of nodes required to first achieve the conditions of full connectivity and no isolated nodes, respectively, for the data used to create the LRMs as well as 32 additional runs. Both scatter plots show close clustering of the 99%ile number of nodes for the 32 additional runs around the 99%ile number of nodes for the original sets of simulations. At a gross level, these results indicate that the LRMs based on the original sets would be expected to perform well for the additional runs.

More precise comparisons are shown in Figure 40 and Figure 41 for full connectivity and no isolated nodes, respectively. These scatter plots show the MAPE and max error pairs for the 99%ile number of nodes

estimated using (22) and (31) for the original sets of simulations in red and the 32 additional runs in blue.

Figure 40 and Figure 41 show that the reported MAPE and max error values for (22) in Section 4.2.1.1 for full connectivity and (31) in Section 4.2.1.4 for no isolated nodes, which were based on the original sets of simulations, were slightly lower than those for most of the 32 additional runs for those conditions.

Another assessment of the adequacy of the sample size used for the simulations is made using a single but much larger data set of simulations for RGGs in a flat unit square. One set of 10,000 simulated geometries was obtained for the mid-r set. Based on the results of that set the following observations are made.

The LRM (22) for the 99%ile number of nodes required to first achieve full connectivity yielded MAPE 1.75% and max error 4.25% for the larger data set. These values are slightly lower than the MAPE and max error reported in Section 4.2.1.1. These values are approximately double those resulting from the LRM

$$n = 4.6392\, r^{-2.0735} \tag{35}$$

obtained from the 10,000 simulated geometries for each range, which were MAPE 0.82% and max error −1.96%.

The LRM (31) for the 99%ile number of nodes required to first achieve no isolated nodes yielded MAPE 1.34% and max error −3.65% for the larger data set. As seen for the condition of full connectivity, these error values for no isolated nodes are slightly lower than the MAPE and max error reported in Section 4.2.1.4. These values are even closer to those resulting from the LRM

$$n = 3.9064\, r^{-2.0969} \tag{36}$$

obtained from the 10,000 simulated geometries for each range, which were MAPE 1.02% and max error −2.49%.

These results show that MAPE and max error based on the "training data set" tend to be optimistic but the fact that some of the additional runs of 1,000 simulations resulted in even lower MAPE and max error values means that the error performance of these LRMs is not excessively optimistic. The results of the independent run of 10,000 simulations for each range still gave small MAPE and max error for the LRMs that were developed using only 1,000 simulations.

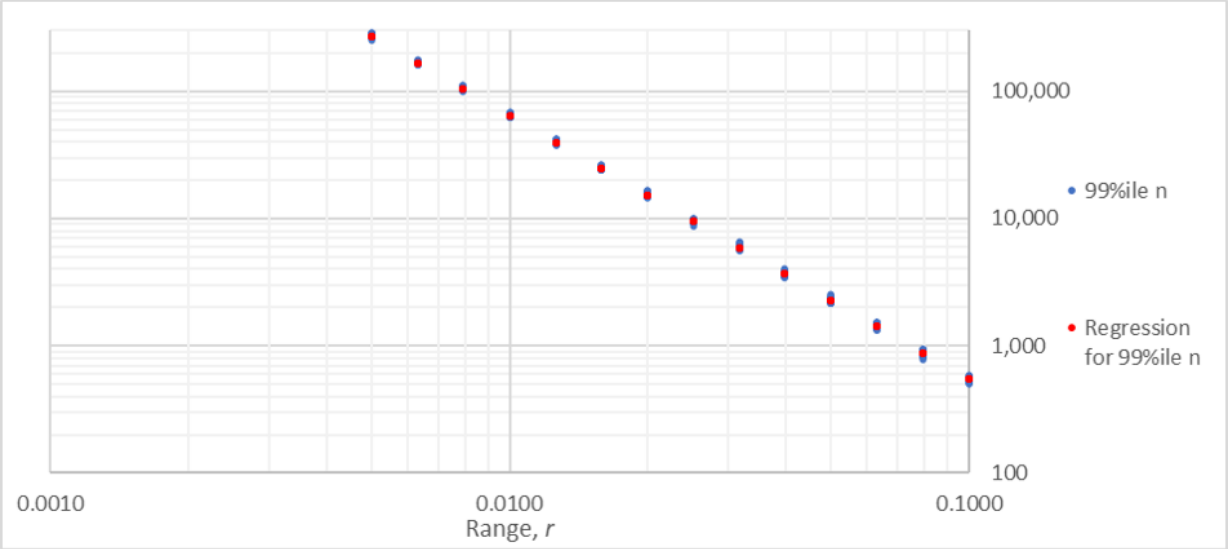

Figure 38: Plot of the 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth for the data used to create the LRMs as well as those from 32 additional runs.

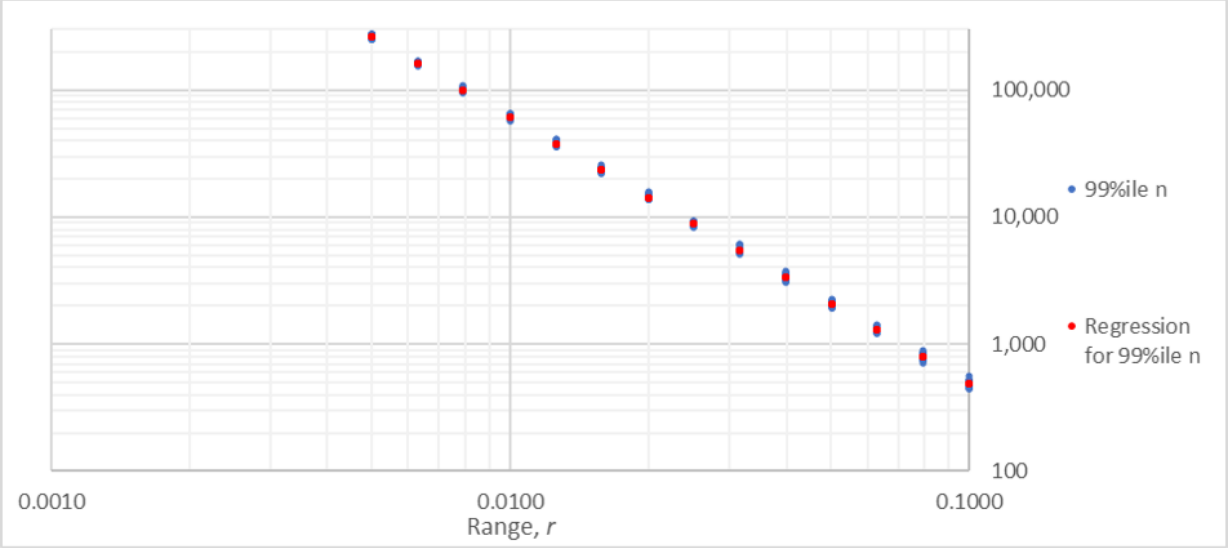

Figure 39: Plot of the 99%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth for the data used to create the LRMs as well those from 32 additional runs.

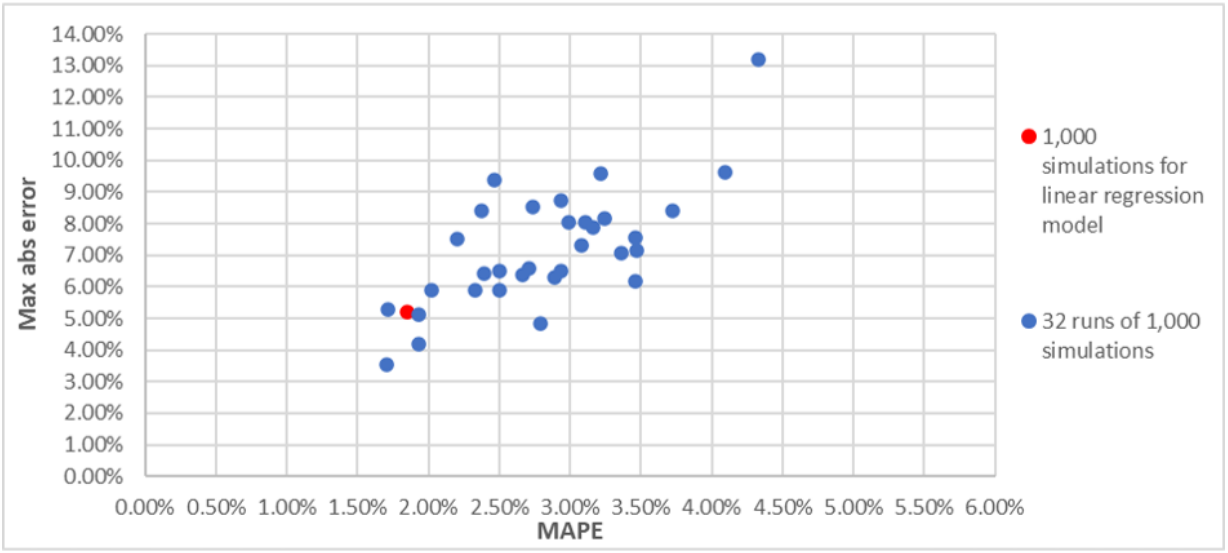

Figure 40: Scatter plot of the MAPE and max error pairs for 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth estimated using LRM (22) for the original sets of simulations and the 32 additional runs.

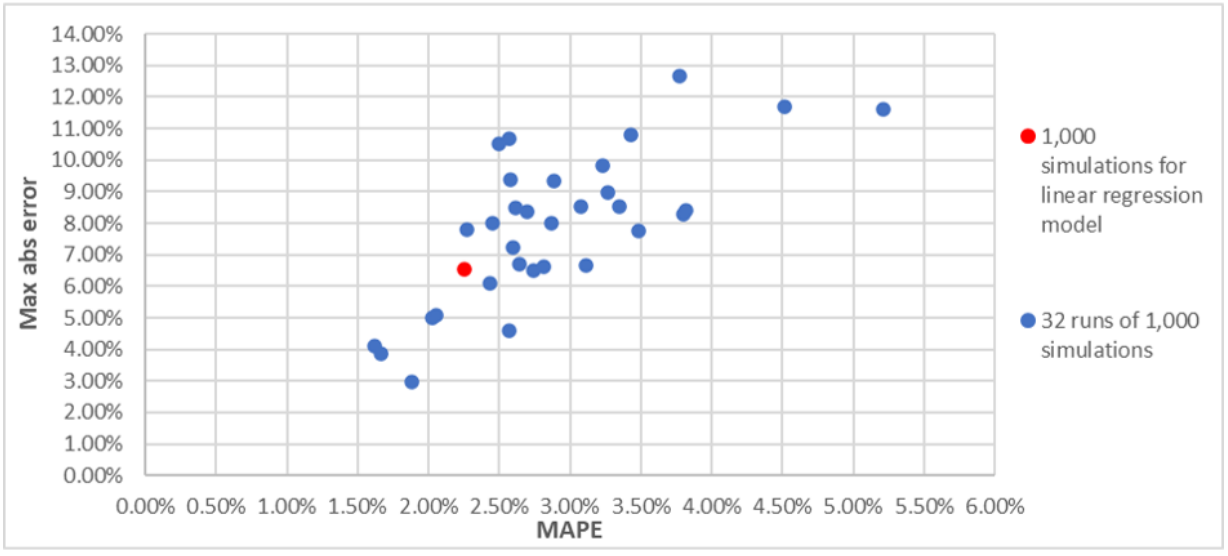

Figure 41: Scatter plot of the MAPE and max error pairs for 99%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth estimated using LRM (31) for the original sets of simulations and the 32 additional runs.

4.2.2 Flat Unit Square: Full Connectivity for Large Ranges

All of the configurations previously used for this paper were based on the mid-r set described and listed in Section 4.1.2. Here, configurations are tested with the large-r set which is the set of ranges from 1.5849 down to 0.1000.

As noted in Section 4.1.3, fully connected RGGs with n-for-given-r growth were constrained to have a minimum size of 25 nodes.

LRMs are obtained for full connectivity. These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations.

Figure 42 shows the 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve full connectivity of RGGs based on the sample of size $N = 1{,}000$ simulated geometries for each value in the large-r set. For the largest ranges in the set the three statistical metrics are essentially flat but for the smaller ranges each of the three statistical metrics appears to demonstrate power law behavior.

Figure 43 shows the LRMs for the number of nodes n required to first achieve full connectivity of RGGs for the subsets of the large-r set where the 99%ile, mean, and 1%ile statistical metrics appear to obey power laws. These ranges are $r = 0.1000$ up to 0.3981 for the 99%ile n, $r = 0.1000$ up to 0.2512 for the mean n, and $r = 0.1000$ up to 0.1995 for the 1%ile n.

The LRMs for full connectivity corresponding to the equations for the trend lines shown in Figure 43 are as follows.

For the 99%ile number of nodes required to first achieve full connectivity in a flat unit square with n-for-given-r growth based on the simulated geometries using the sets of ranges listed above, (21) becomes

$$n = 4.8769\, r^{-2.0557} \tag{37}$$

with an R^2 value of 0.9985. For the simulated values of r, MAPE is 3.42% and max error of the value of the predicted n is −5.16%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.8579\, r^{-2.1959} \tag{38}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 0.83% and max error of the value of the predicted n is 1.42%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 0.5994\, r^{-2.4812} \tag{39}$$

with an R^2 value of 0.9989. For the simulated values of r, MAPE is 1.77% and max error of the value of the predicted n is 1.42%.

The LRMs obtained from the subsets of the large-r set are slightly different from the LRMs obtained from the mid-r set but they provide similar numerical results. When the LRMs obtained in Section 4.2.1.1 for the mid-r set for full connectivity of RGGs in flat unit squares are used to estimate the number of nodes required to first achieve full connectivity with the large-r set the error values are as follows:

- For the 99%ile number of nodes, MAPE is 3.40% and max error of the value of the predicted n is -7.91%.
- For the mean number of nodes, MAPE is 2.66% and max error of the value of the predicted n is 4.03%.
- For the 1%ile number of nodes, MAPE is 8.26% and max error of the value of the predicted n is 19.88%.

These results show that the LRMs based on the mid-r set provide very good estimates for the 99%ile and mean number of nodes required to first achieve full connectivity with the large-r set and have only slightly greater error for the 1%ile number of nodes.

Figure 44 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected with n-for-given-r growth for the large-r set. Interpreting this fraction as a probability, the probability that an RGG becomes fully connected when the last isolated node is absorbed is observed to be 1 for $r \geq 0.5012$ in this set of ranges. The probability drops as the ranges decrease from 0.3981 down to 0.1995, reaching a minimum of 0.239 for this set of ranges. As r decreases from 0.1995, this probability increases and appears to track the log trend line shown Figure 33 in Section 4.2.1.6 for the mid-r set.

Figure 45 relates the number of nodes needed for full connectivity to that required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. Because fully connected RGGs were constrained to have a minimum size of 25 nodes the ratios are exactly 1 for the largest ranges. The ratio reaches its peak value of 1.550 at $r = 0.1585$, and then decreases as the range decreases.

Figure 46 shows that the LRM (22) has very small relative error for ranges $r \leq 0.3981$. The estimate based on Penrose's Theorem has much larger relative error, which would be expected for RGGs in a flat unit square.

As noted above, fully connected RGGs with n-for-given-r growth were constrained to have a minimum size of 25 nodes. Minimum sizes of two and five nodes were also tested and power law behavior was still observed. The plots of the observed numbers of nodes remained straight until they reached those minimum sizes and the m and b coefficients for those LRMs were only slightly different from those obtained using the default minimum size of 25 nodes.

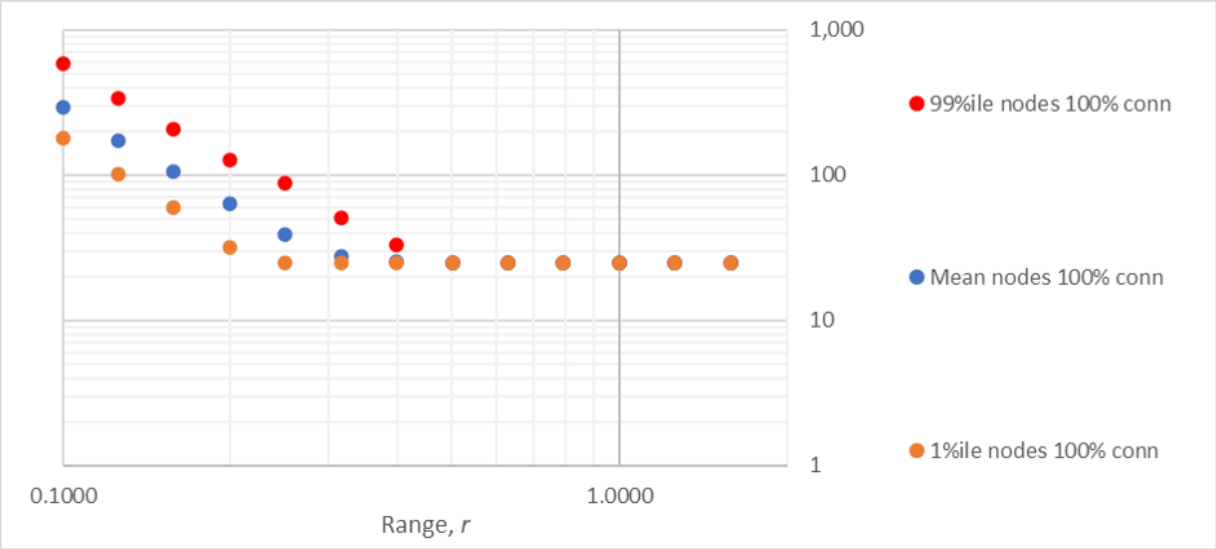

Figure 42: Values of the 99%ile, mean, and 1%ile statistical metrics for number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth for each value in the large-r set.

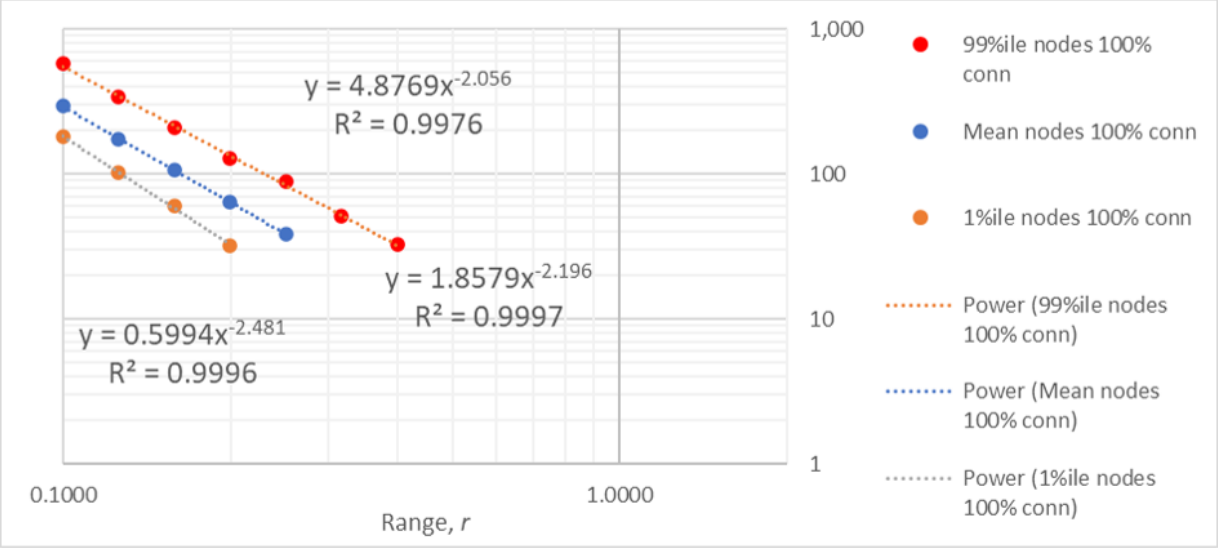

Figure 43: LRMs for the number of nodes for full connectivity of RGGs for the subsets of the large-r set where the 99%ile, mean, and 1%ile statistical metrics can be modeled as power laws.

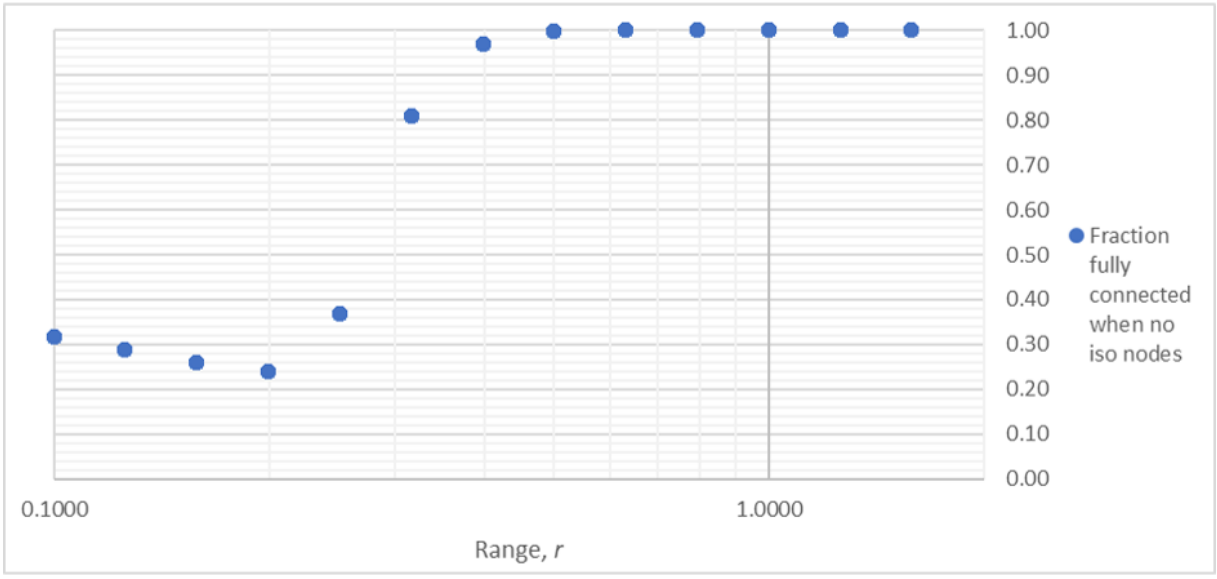

Figure 44: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a flat unit square to become fully connected with n-for-given-r growth for the large-r set.

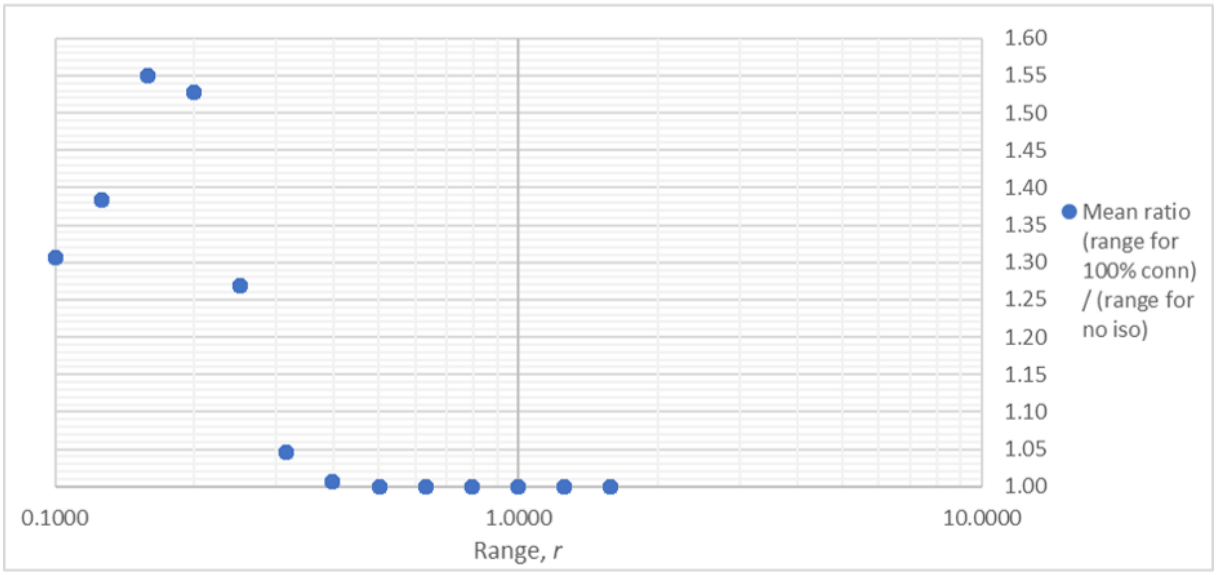

Figure 45: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a flat unit square with n-for-given-r growth for the large-r set. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

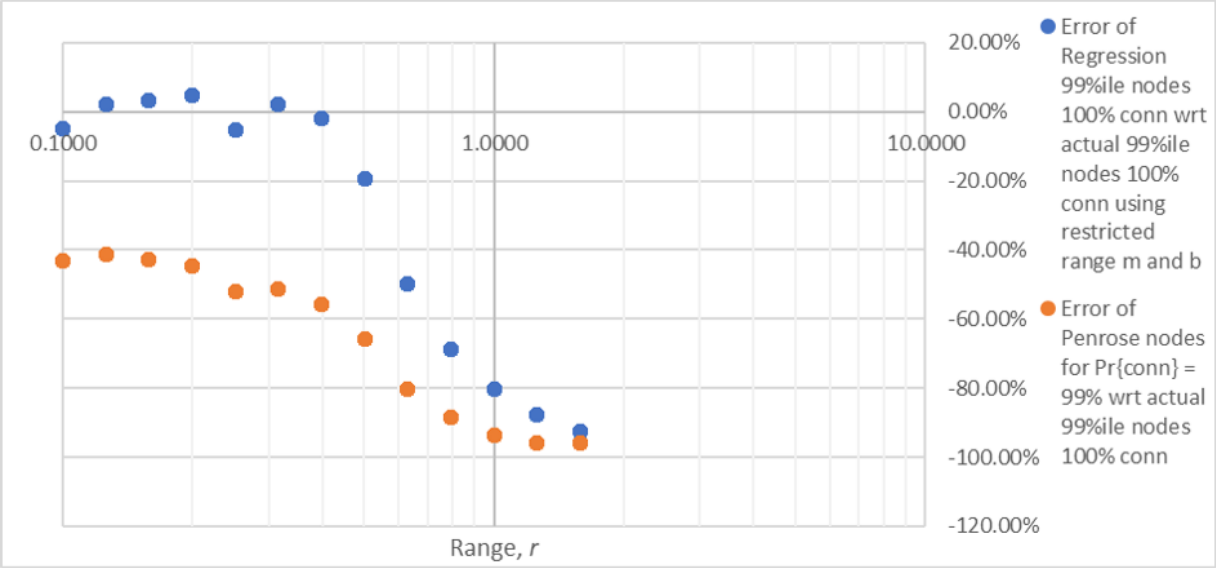

Figure 46: Relative error of LRM (22) for 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth based on the subset of the large-r set. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

4.2.3 Flat Unit Square with n-for-given-r Growth for Small Ranges

Section 4.2.3.1 through Section 4.2.3.3 repeat some of the analyses presented in Section 4.2.1 and its subsections for ranges in the mid-r set but now for the small-r set consisting of the ranges from 0.005012 down to 0.000251.

LRMs are obtained for full connectivity and no isolated nodes.

These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

4.2.3.1 Flat Unit Square: Full Connectivity for Small-r Set

Figure 47 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a flat unit square with n-for-given-r growth for the small-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.5833\, r^{-2.0767} \tag{40}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.71% and max error of the value of the predicted n is -3.23%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.5365\, r^{-2.1140} \tag{41}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.58% and max error of the value of the predicted n is 1.07%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.6880\, r^{-2.1451} \tag{42}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.96% and max error of the value of the predicted n is 2.85%.

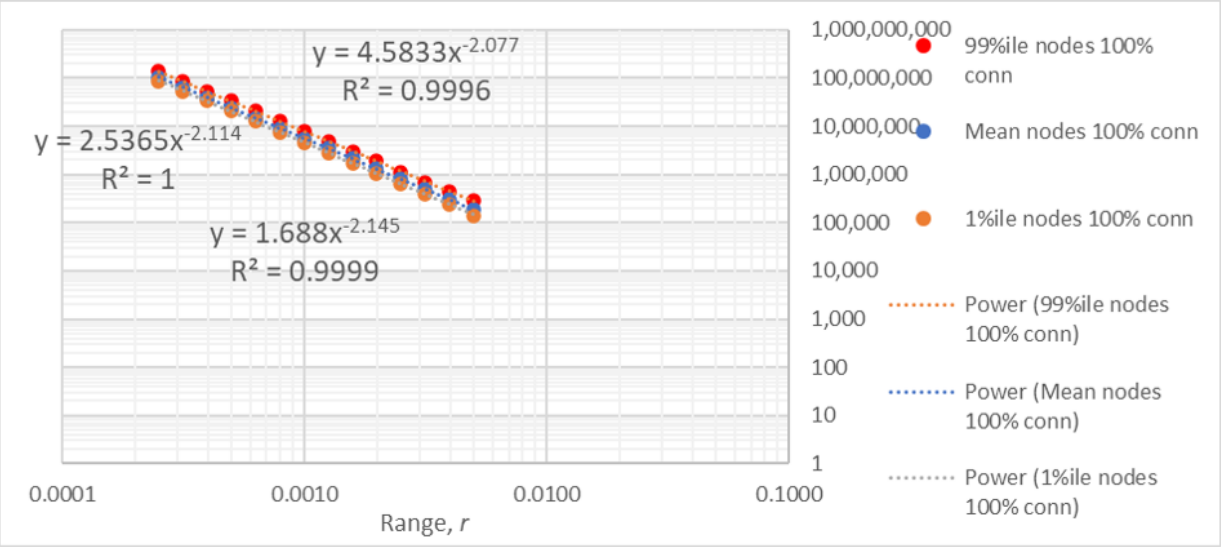

Figure 47: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth for the small-r set.

4.2.3.2 Flat Unit Square: No Isolated Nodes for Small-r Set

Figure 48 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the small-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 4.4078\, r^{-2.0801} \tag{43}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.62% and max error of the value of the predicted n is 4.13%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 2.3465\, r^{-2.1217} \tag{44}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.69% and max error of the value of the predicted n is 1.26%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.5740\, r^{-2.1515} \tag{45}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.88% and max error of the value of the predicted n is 2.42%.

It is seen for RGGs in a flat unit square that the number of nodes required to first achieve full connectivity shown in Figure 47 and their LRMs (40), (41), and (42) are larger than the number of nodes required for the disappearance of isolated nodes shown in Figure 48 and their LRMs (43), (44), and (45). Full connectivity does not necessarily occur when the last isolated node is absorbed.

Figure 49 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. The 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the small-r set to nearly a million and tens of millions for the smallest ranges simulated in these experiments.

Figure 49 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

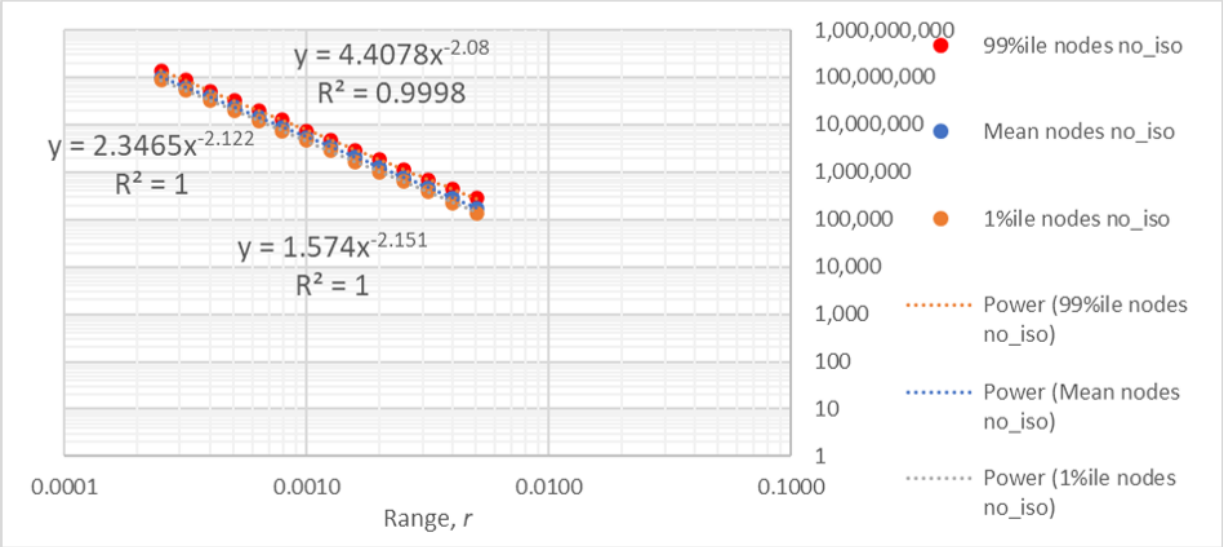

Figure 48: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth.

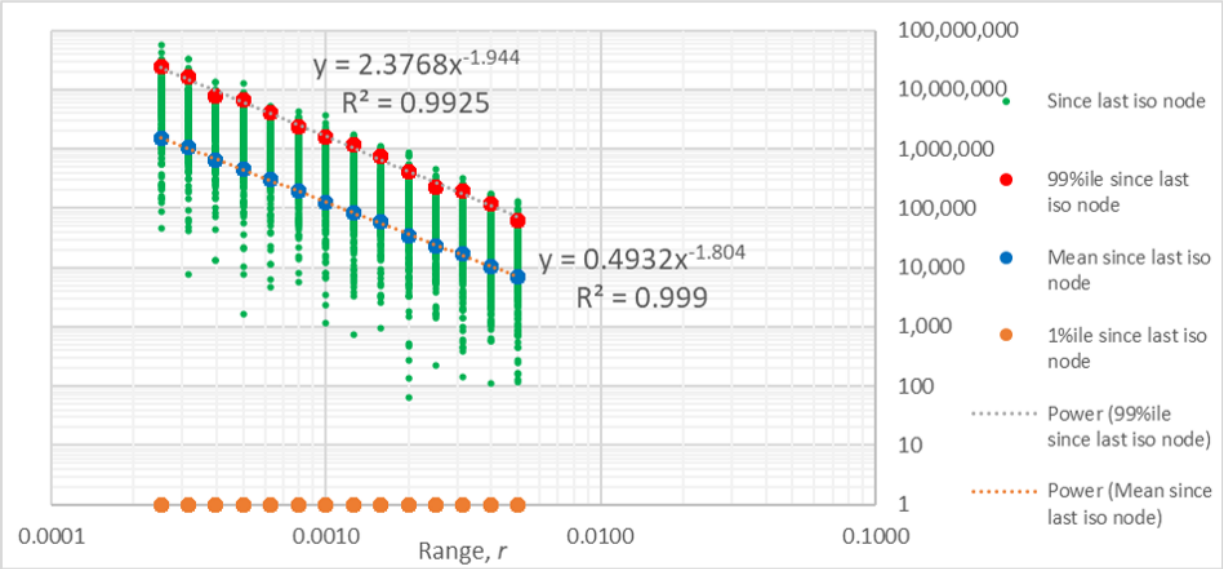

Figure 49: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.2.3.3 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a flat unit square with n-for-given-r growth using the small-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 50 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a flat unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a flat unit square with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.639 for $r = 0.005012$ and increases to about 0.804 for $r = 0.000251$.

This probability increases as r decreases (and consequently the number of nodes increases) but it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a flat unit square even for the smallest range in the small-r set.

Figure 51 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. For small ranges, very few additional nodes are required to first achieve full connectivity after the disappearance of isolated nodes. The indication is that the ratios may asymptotically approach 1 for even smaller ranges.

Figure 52 shows the error of the LRM (40) given in Section 4.2.3.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. It can be seen that the error percentages hover close to zero.

Figure 52 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. For the ranges in the small-r set, Penrose's predicted values of n have MAPE 17.72% and max error -24.46%.

Figure 53 shows the same error statistics as in Figure 52 but now for the 1%ile number of nodes required to first achieve full connectivity. Figure 53 shows the error percentages for the LRM (42) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. The power law LRM has extremely small relative error and Penrose's predicted values of n are only slightly larger. For the ranges in the small-r set, Penrose's predicted values of n have MAPE 6.02% and max error -8.68%.

Figure 54 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 47.

The maximum of the c_P observed for any range was 18.315. The visible trend in Figure 54 is that all three of the statistical metrics plotted for c_P seem to be practically flat (with some noise) as the range varies. The mean value of c_P is 8.6842 for the 99%ile values of n and -0.7545 for the 1%ile values of n.

It is noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

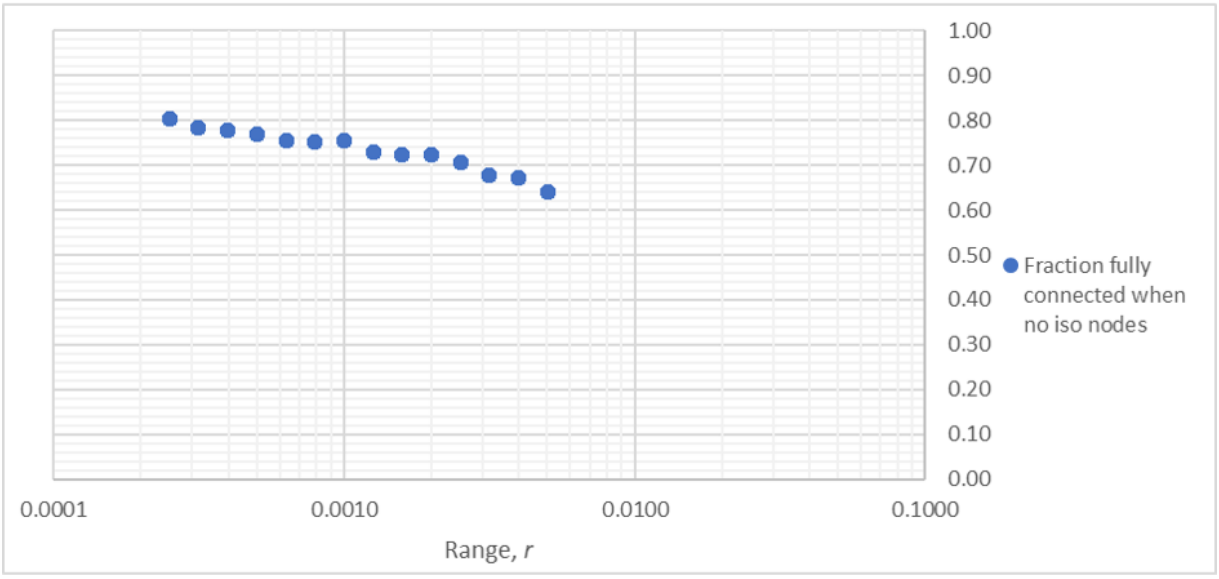

Figure 50: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a flat unit square to become fully connected with n-for-given-r growth.

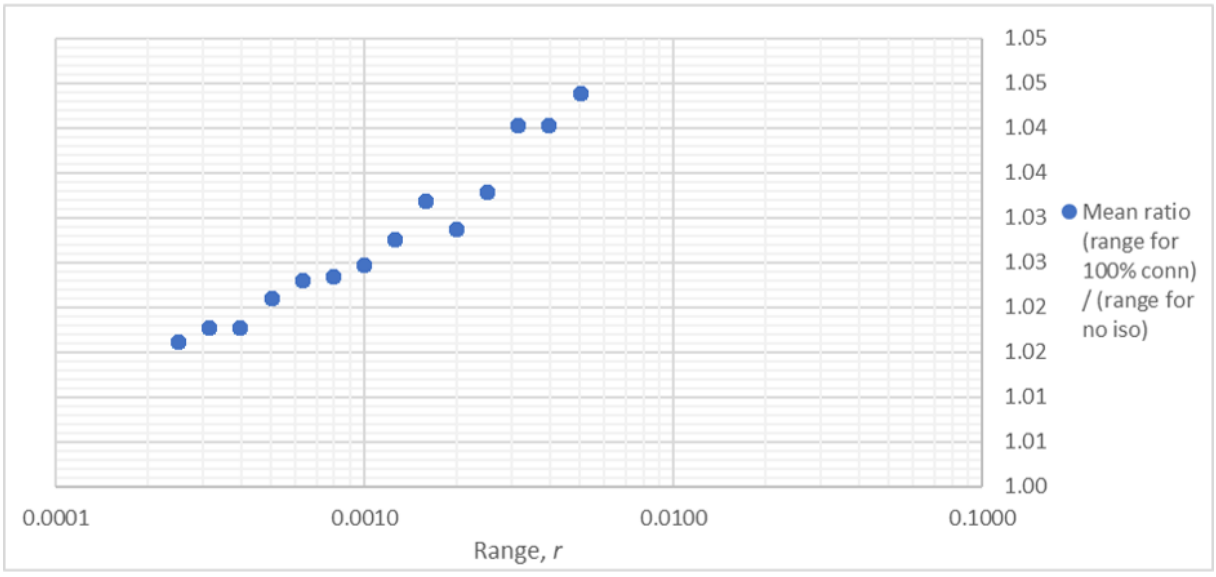

Figure 51: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a flat unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

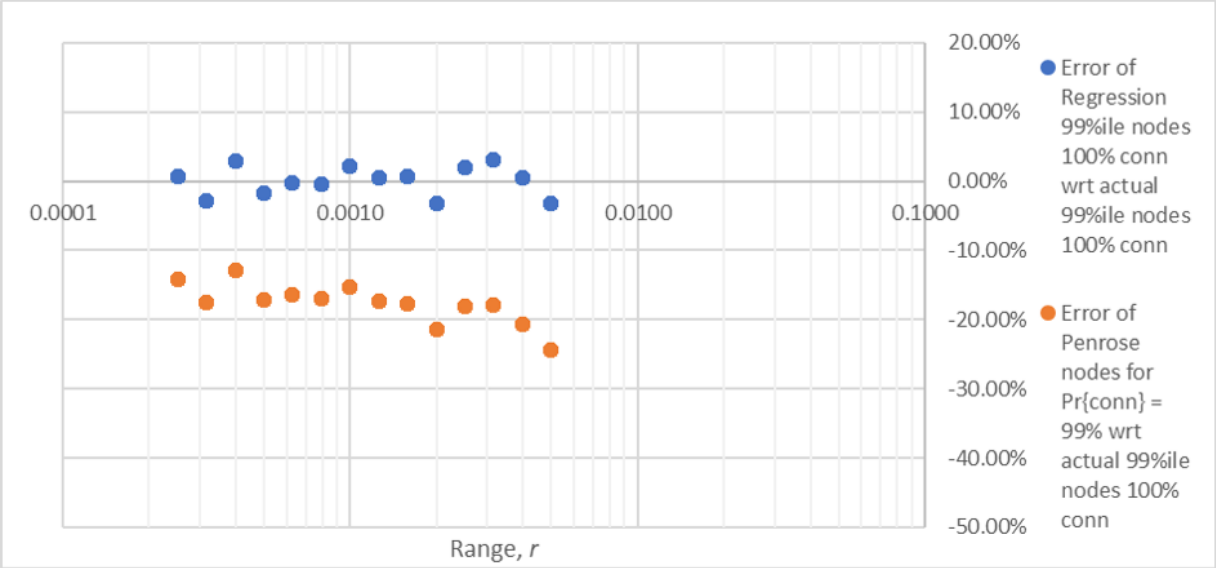

Figure 52: Relative error of LRM (40) for 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

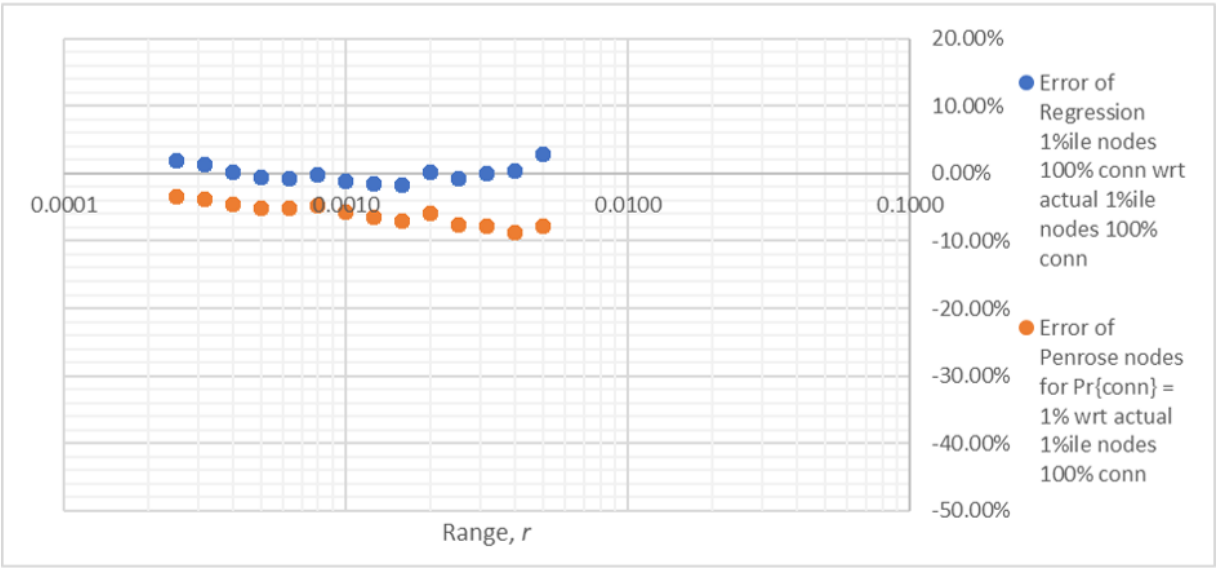

Figure 53: Relative error of LRM (42) for 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

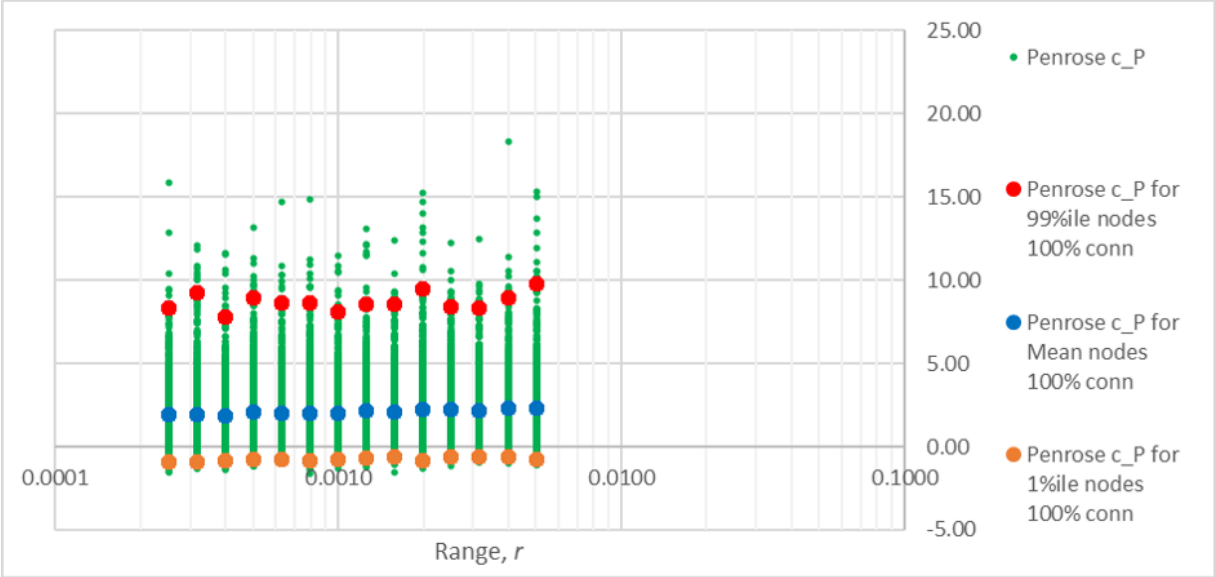

Figure 54: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.2.4 Flat Unit Square with n-for-given-r Growth for Mid and Small Ranges

Section 4.2.4.1 through Section 4.2.4.3 repeat some of the analyses presented in Section 4.2.1 and its subsections for ranges in the mid-r set and Section 4.2.3 and its subsections for ranges in the small-r set but now for ranges in the mid+small-r set which is the set of ranges from 0.100000 down to 0.000251.

LRMs are obtained for full connectivity and no isolated nodes.

These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

4.2.4.1 Flat Unit Square: Full Connectivity for Mid+Small-r Set

Figure 55 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a flat unit square with n-for-given-r growth for the mid+small-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.4322\, r^{-2.0815} \tag{46}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.15% and max error of the value of the predicted n is -5.36%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.2071\, r^{-2.1344} \tag{47}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 1.78% and max error of the value of the predicted n is 3.87%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.2867\, r^{-2.1853} \tag{48}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 3.99% and max error of the value of the predicted n is 12.01%.

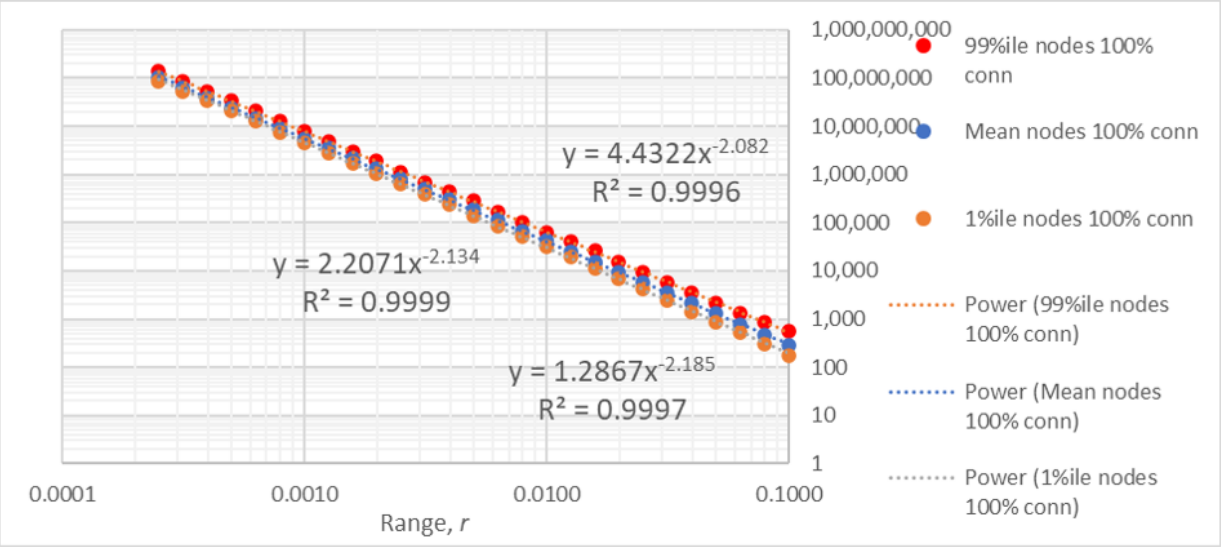

Figure 55: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth for the mid+small-r set.

4.2.4.2 Flat Unit Square: No Isolated Nodes for Mid+Small-r Set

Figure 56 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid+small-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 3.9364\, r^{-2.0967} \tag{49}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.43% and max error of the value of the predicted n is -9.04%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.8137\, r^{-2.1598} \tag{50}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 3.64% and max error of the value of the predicted n is 11.17%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.9319\, r^{-2.2293} \tag{51}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 8.03% and max error of the value of the predicted n is 30.60%.

It is seen for RGGs in a flat unit square that the number of nodes required to first achieve full connectivity shown in Figure 55 and their LRMs (46), (47), and (48) are larger than the number of nodes required for the disappearance of isolated nodes shown in Figure 56 and their LRMs (49), (50), and (51). Full connectivity does not necessarily occur when the last isolated node is absorbed.

Figure 57 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. The 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the mid+small-r set to nearly a million and tens of millions for the smallest ranges simulated in these experiments.

Figure 57 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

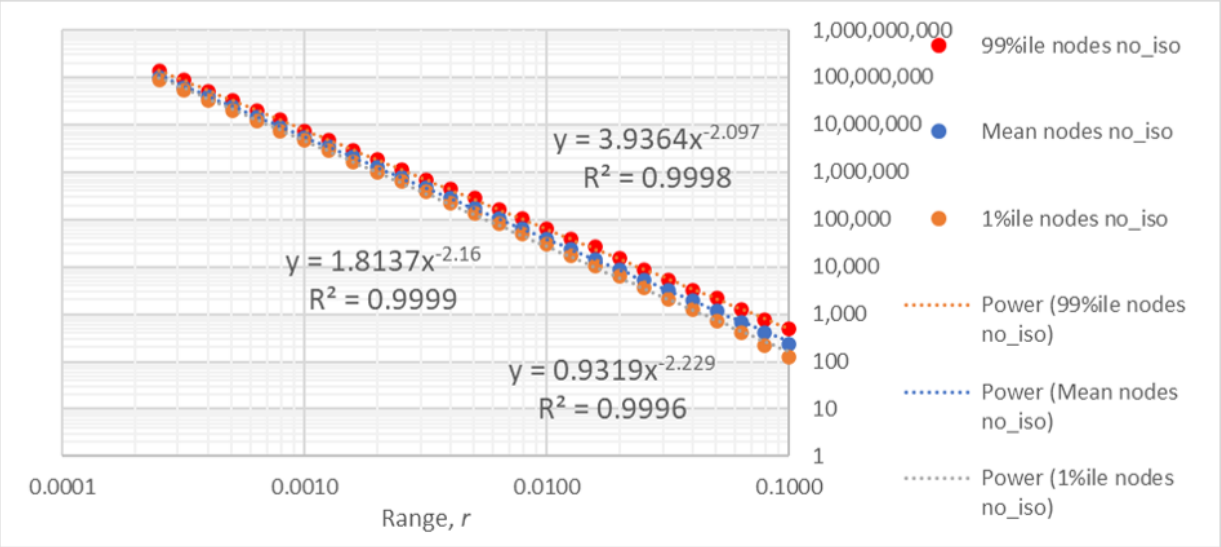

Figure 56: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth.

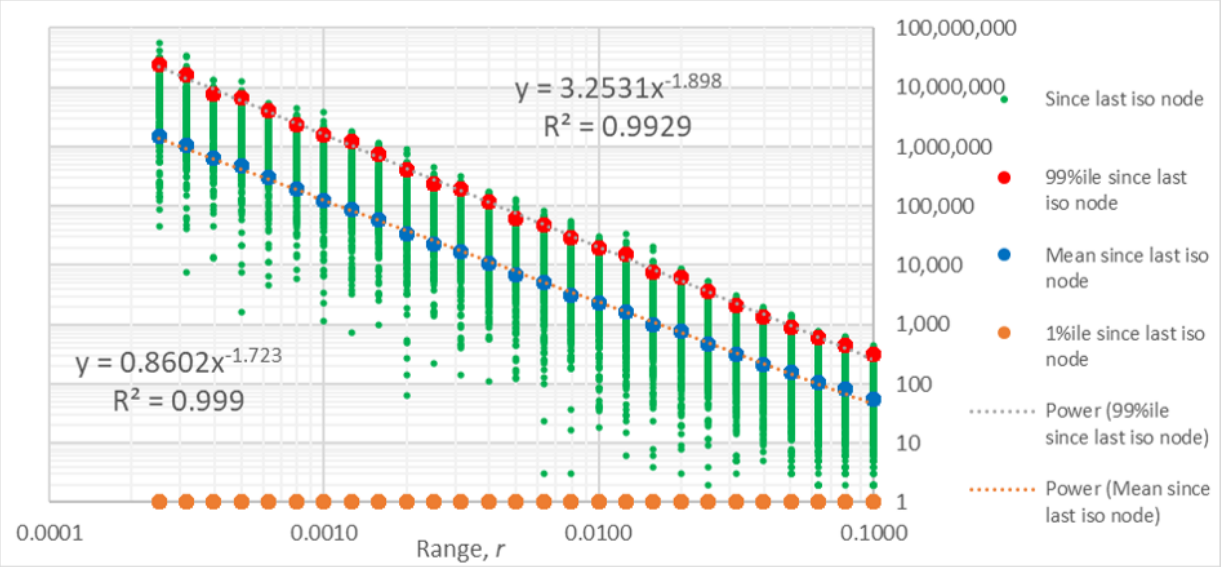

Figure 57: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.2.4.3 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid+Small-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a flat unit square with n-for-given-r growth using the mid+small-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 58 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a flat unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a flat unit square with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.314 for $r = 0.100000$[40] and increases to about 0.804 for $r = 0.000251$. As the range decreases the bend in the curve is apparent, indicating that the fraction approaches 1 more slowly for the smaller ranges.

This probability increases as r decreases (and consequently the number of nodes increases) but it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a flat unit square even for the smallest range in the mid+small-r set.

Figure 59 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r

[40] It was noted in Section 4.1.2 that separate sets of simulations were performed for some ranges. For this reason the probability that an RGG becomes fully connected when the last isolated node is absorbed was observed to be about 0.314 for $r = 0.100000$ based on the mid+small-r set, while in Section 4.2.1.6 this probability was observed to be about 0.328 for $r = 0.1000$ based on the mid-r set.

is plotted on the graph. For small ranges, very few nodes are required to first achieve full connectivity after the disappearance of isolated nodes. The ratios may asymptotically approach 1 for even smaller ranges.

While Figure 58 shows that an RGG in a flat unit square with n-for-given-r growth does not inevitably become fully connected at the same instant the last isolated node disappears, Figure 59 shows that, except for the largest ranges in the set, very few nodes must be added to the 99%ile number of nodes required to first achieve no isolated nodes to reach the 99%ile number of nodes required to first become fully connected.

Figure 60 shows the error of the LRM (46) given in Section 4.2.4.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. This figure also includes the relative error of the LRM (22) given in Section 4.2.1.1 that was obtained for the mid-r set and, although this LRM is slightly different from that obtained for the mid+small-r set, it tracks well even for the smallest ranges simulated.

Figure 60 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. It has larger relative error than the LRMs obtained for the mid+small-r and mid-r set sets.

Figure 61 shows the same error statistics as in Figure 60 but now for the 1%ile number of nodes required to first achieve full connectivity. Figure 61 shows the error percentages for the LRM (48) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$.

For the ranges in the mid+small-r set, Penrose's predicted values of both the 99%ile and 1%ile values of n have large errors that decrease as the range decreases, while the LRM appears to have increasing relative errors for the smaller ranges.

Figure 62 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 55.

The maximum of the c_P observed for any range was 21.1690. The visible trend in Figure 62 is that all three of the statistical metrics plotted for c_P decrease slightly as the range decreases.

It is noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

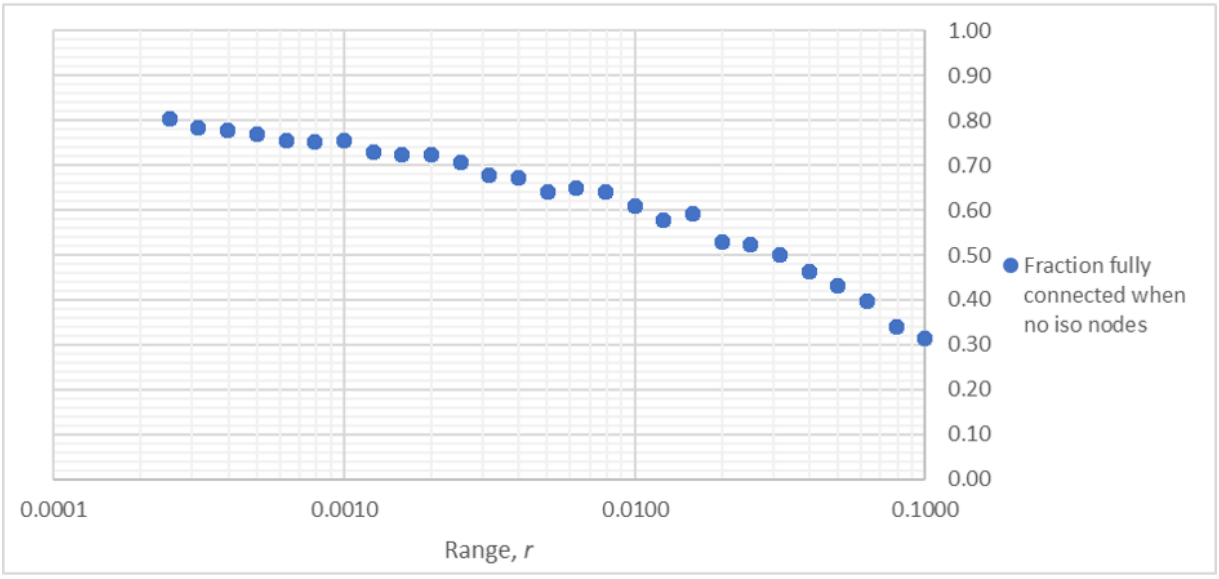

Figure 58: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a flat unit square to become fully connected with n-for-given-r growth.

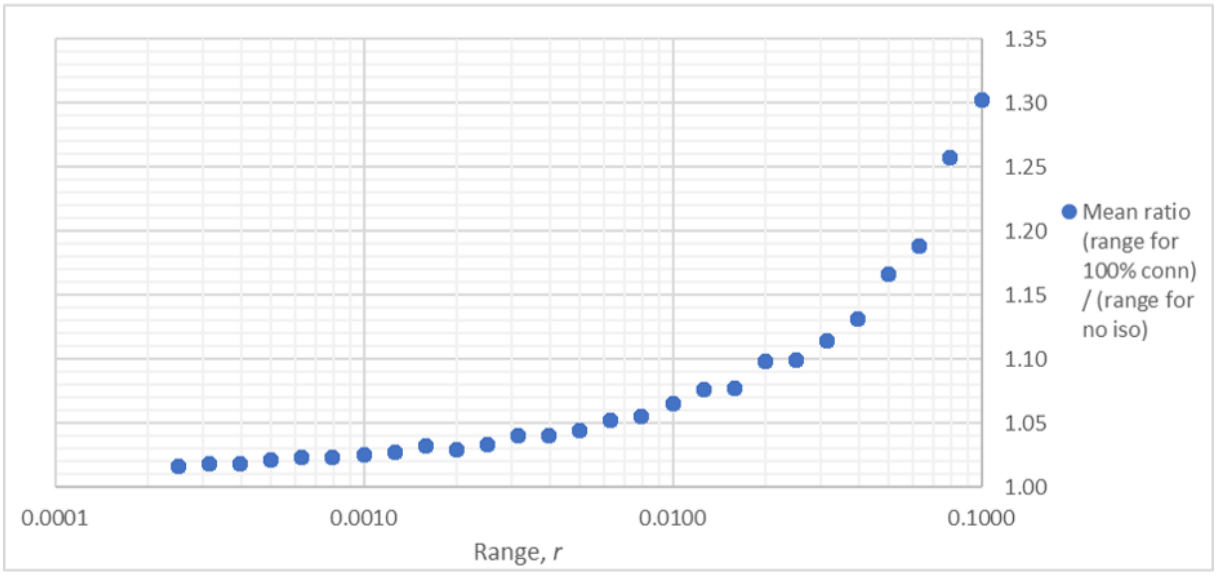

Figure 59: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a flat unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

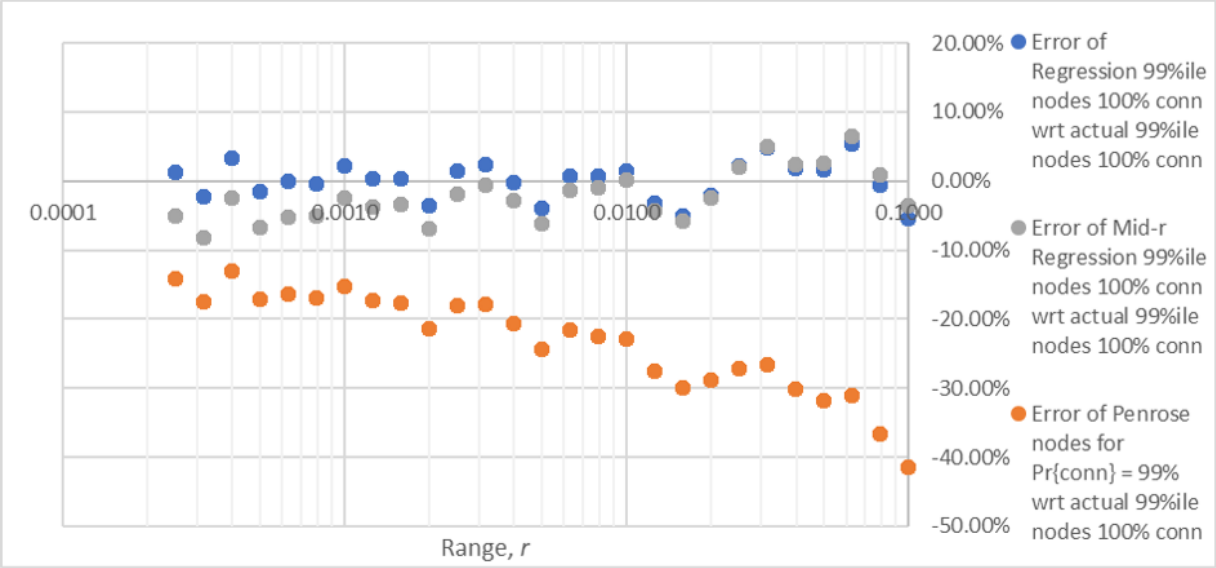

Figure 60: Relative error of LRM (46) obtained from the mid+small-r set and LRM (22) obtained from the mid-r set for 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

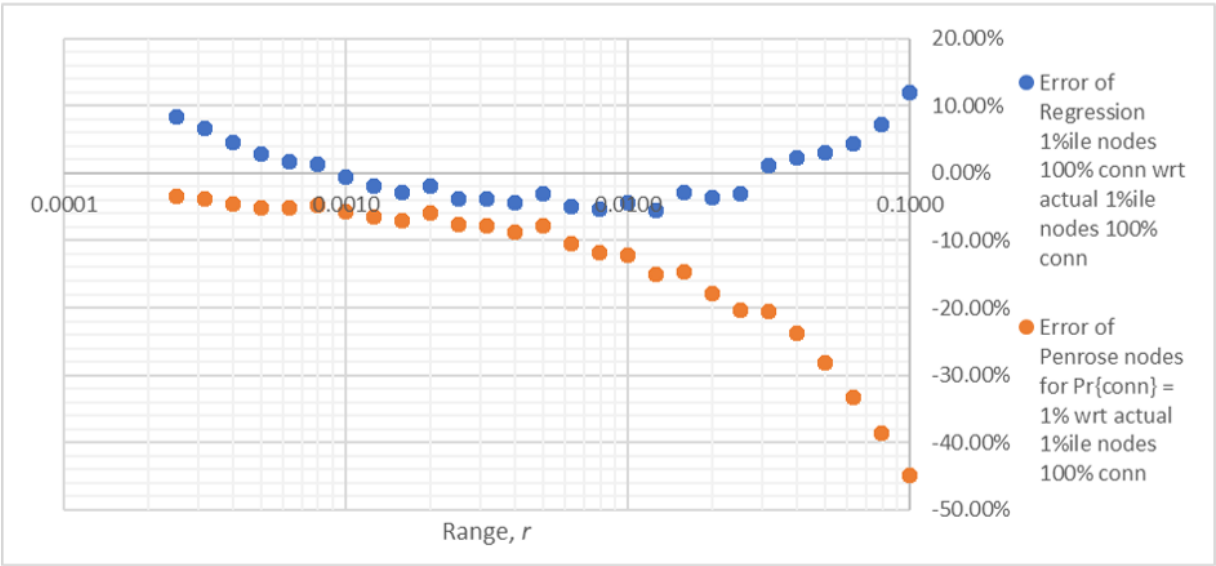

Figure 61: Relative error of LRM (48) for 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

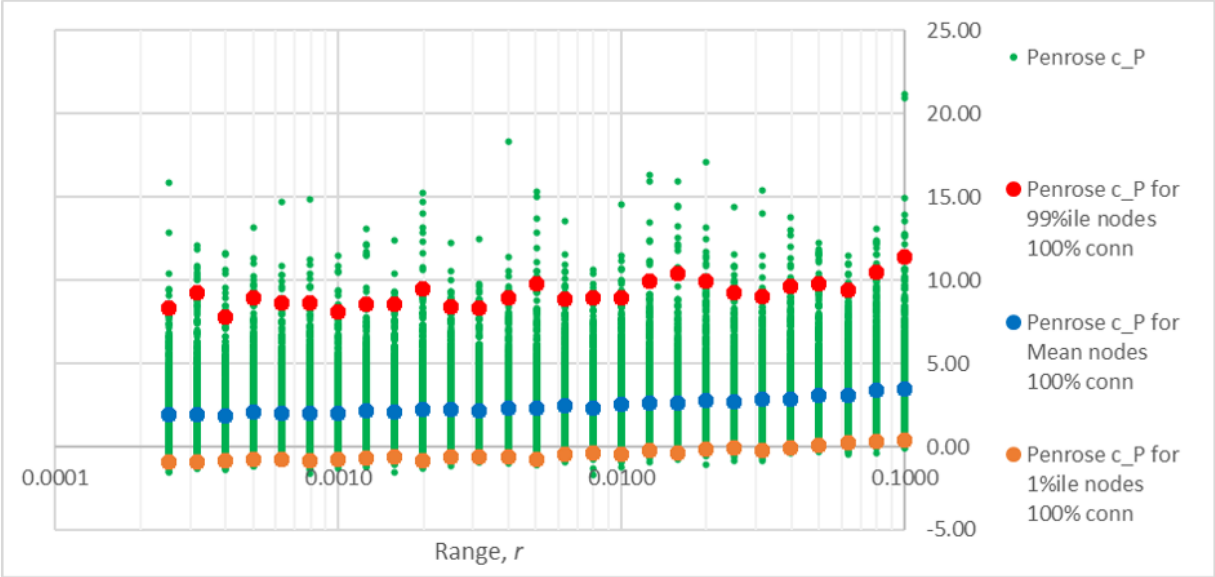

Figure 62: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.2.5 Flat Unit Square: Comparing Linear Regression Models from Mid-r Set, Small-r Set, and Mid+Small-r Set

This section compares the MAPE and max error resulting from using an LRM derived from one set of ranges to predict numbers of nodes for other sets of ranges. Only the results for the 99%ile numbers of nodes for full connectivity are compared.

- LRM (22) obtained from mid-r set:
  - Applied to mid-r set: MAPE 1.85%, max error −5.19%
  - Applied to small-r set: MAPE 2.62%, max error 6.51%
  - Applied to mid+small-r set: MAPE 2.04%, max error −6.57%
- LRM (41) obtained from small-r set:
  - Applied to mid-r set: MAPE 4.34%, max error −8.13%
  - Applied to small-r set: MAPE 1.71%, max error −3.23%
  - Applied to mid+small-r set: MAPE 1.67%, max error −4.01%
- LRM (46) obtained from mid+small-r set:
  - Applied to mid-r set: MAPE 3.65%, max error −8.13%
  - Applied to small-r set: MAPE 2.45%, max error 7.40%
  - Applied to mid+small-r set: MAPE 2.15%, max error −5.36%

All of the MAPE values for flat RGGs with n-for-given-r growth are fairly small percentages regardless of which LRM is applied to a particular set of ranges. In two cases the smallest MAPE was obtained by using the LRM derived from that set of ranges; the exception was where the small-r set MAPE was slightly smaller for the mid+small-r LRM than for the small-r LRM, even though the difference is essentially insignificant.

The smallest max error resulted from the LRM obtained from that set of ranges in all three cases.

These results are a general indication that the linearity of the log-log transformed 99%ile number of nodes for flat RGGs with n-for-given-r growth holds well over the various intervals represented by the mid-r set, small-r set, and their union.

4.3 Toroidal Unit Square with n-for-given-r Growth

This section considers RGGs consisting of nodes scattered uniformly at random positions in a two-dimensional square space of area $A = 1$ with wraparound at the border. This is a toroidal space so there are no border effects. The RGGs evolve according to the n-for-given-r growth model.

Figure 63 shows the position where the last node was added to first achieve full connectivity for each of the 1,000 simulated geometries. This figure demonstrates that the last nodes added are spread in a pattern that is more uniform than that shown in Figure 25 for flat unit squares; in fact, the pattern looks much like the Poisson distribution of nodes in RGGs as demonstrated in Figure 1 and Figure 20. This is because all nodes have the same number of neighbors on average due to wraparound at the border, so the nodes near the border of the space are as likely as any other nodes to be absorbed into the GC at the time they are added. It is in stark contrast to the situation shown in Figure 25 where the last nodes added that first achieve full connectivity in a flat unit square tend to cluster near the border of the square.

Toroidal spaces are frequently assumed in the RGG literature due to the simplification they bring to analyses. This section explores the properties of toroidal space and contrasts the results with those for RGGs in flat space.

For each simulated geometry for a given configuration, nodes with a fixed range r are added one-by-one to determine the number of nodes n required to first achieve the condition of 100% connectivity, 99% connectivity, or no isolated nodes. The 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve each of those conditions are obtained for that configuration.

Results are given separately for the mid-r, small-r, and mid+small-r sets. Section 4.3.1 presents LRMs and other results for RGGs with n-for-given-r growth in toroidal unit squares based on configurations using the mid-r set. Section 4.3.2 presents LRMs for RGG configurations based on the small-r set and Section 4.3.3 presents LRMs for RGG configurations based on the mid+small-r set.

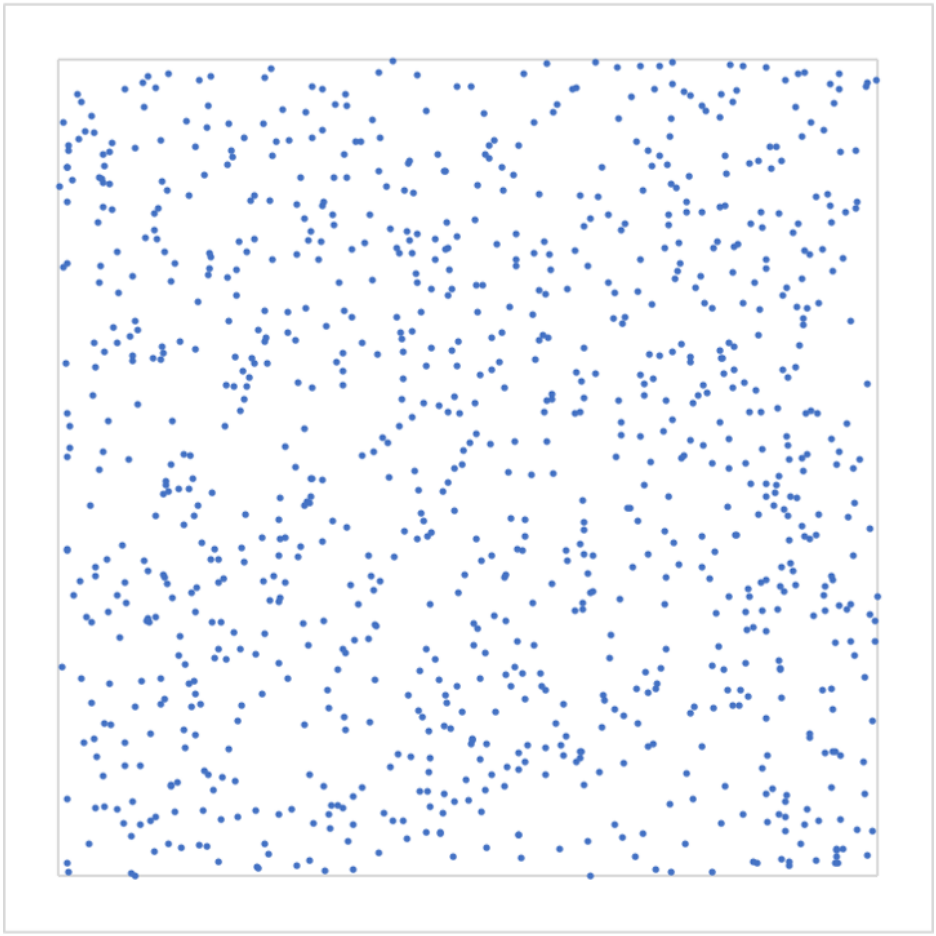

Figure 63: Position where the last node was added for full connectivity of each of 1,000 simulated geometries for RGGs in a toroidal unit square with range 0.0100 and n-for-given-r growth. The lack of border effects in a toroidal space eliminates the clustering of the last nodes added near the border. Although values on the axes are not shown, they span 0 to 1.

4.3.1 Toroidal Unit Square with n-for-given-r Growth for Mid Ranges

Section 4.3.1.1 through Section 4.3.1.5 present results for RGGs in toroidal unit squares with n-for-given-r growth using ranges in the mid-r set which is the set of ranges from 0.1000 down to 0.0050.

These sections give LRMs for full connectivity, 99% connectivity, and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

4.3.1.1 Toroidal Unit Square: Full Connectivity for Mid-r Set

Figure 64 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the mid-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. As was the case for RGGs in a flat unit square, the close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 2.2053\, r^{-2.1711} \tag{52}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.64% and max error of the value of the predicted n is -4.58%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.2306\, r^{-2.2318} \tag{53}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.80% and max error of the value of the predicted n is 4.70%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 0.7369\, r^{-2.2968} \tag{54}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 2.77% and max error of the value of the predicted n is 6.54%.

A visual comparison of Figure 26 for RGGs in a flat unit square and Figure 64 for RGGs in a toroidal unit square shows, even on a log scale where differences are compressed, that more nodes are needed to first achieve full connectivity in a flat space than in a toroidal space. This is quantified in Figure 65 that gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity in flat and toroidal spaces; note that the scale for the range is linear rather than log. Figure 65 shows that the ratios appear to fall toward 1 for all three of the statistical metrics as the range becomes very small. The ratios for all three of these statistical metrics vary roughly linearly with the range. For the largest ranges in the mid-r set the data points and their LRMs show that between roughly 25% to 75% more nodes must be added to first achieve full connectivity of RGGs in flat space than toroidal space for the three statistical metrics reported here.

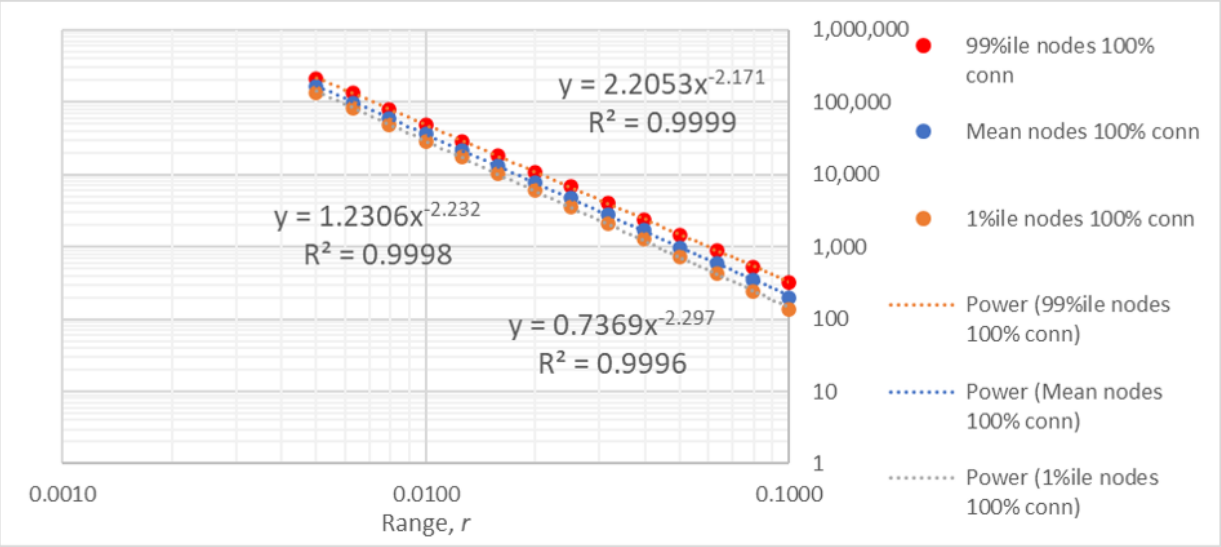

Figure 64: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the mid-r set.

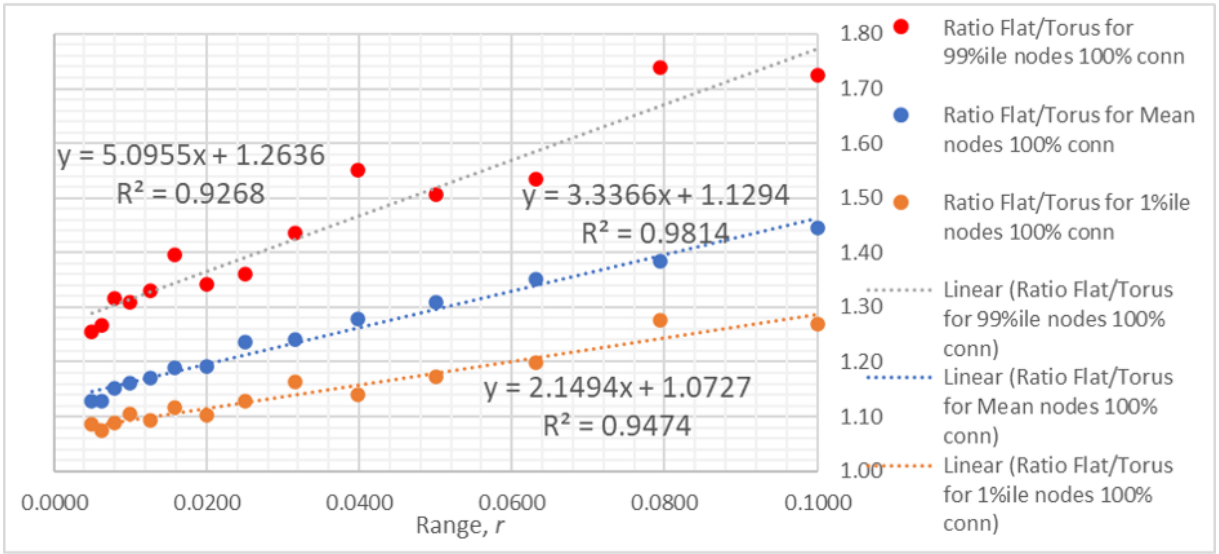

Figure 65: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in flat and toroidal spaces with n-for-given-r growth.

4.3.1.2 Toroidal Unit Square: 99% Connectivity for Mid-r Set

Demonstrating less-than-full connectivity, Figure 66 shows statistical metrics for the number of nodes required to first achieve at least 99% connectivity (the condition where $S \geq 0.99$) for RGGs in a toroidal unit square with n-for-given-r growth.

Figure 66 shows that the number of nodes required to first achieve 99% connectivity closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity with n-for-given-r growth, (21) becomes

$$n = 3.1303\, r^{-1.9110} \tag{55}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.29% and max error of the value of the predicted n is -2.83%.

For the mean number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 1.7650\, r^{-2.0227} \tag{56}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.14% and max error of the value of the predicted n is 2.84%.

For the 1%ile number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 1.0810\, r^{-2.1182} \tag{57}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 3.43% and max error of the value of the predicted n is 8.34%.

As for the case of full connectivity, a visual comparison of Figure 27 for RGGs in a flat unit square and Figure 66 for RGGs in a toroidal unit square shows that more nodes are needed to first achieve 99% connectivity in flat space than in toroidal space. Figure 67 gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to first achieve 99% connectivity in flat and toroidal spaces; note that the scale for the range is linear rather than log. Figure 67 shows that the ratio tends toward 1 for all three of the statistical metrics as the range becomes very small. The ratios for all three of these statistical metrics vary roughly linearly with the range. For the largest ranges in the mid-r set the data points and their LRMs show that between roughly 25% to 50% more nodes must be added to first achieve full connectivity of RGGs in flat space than toroidal space for the three statistical metrics reported here.

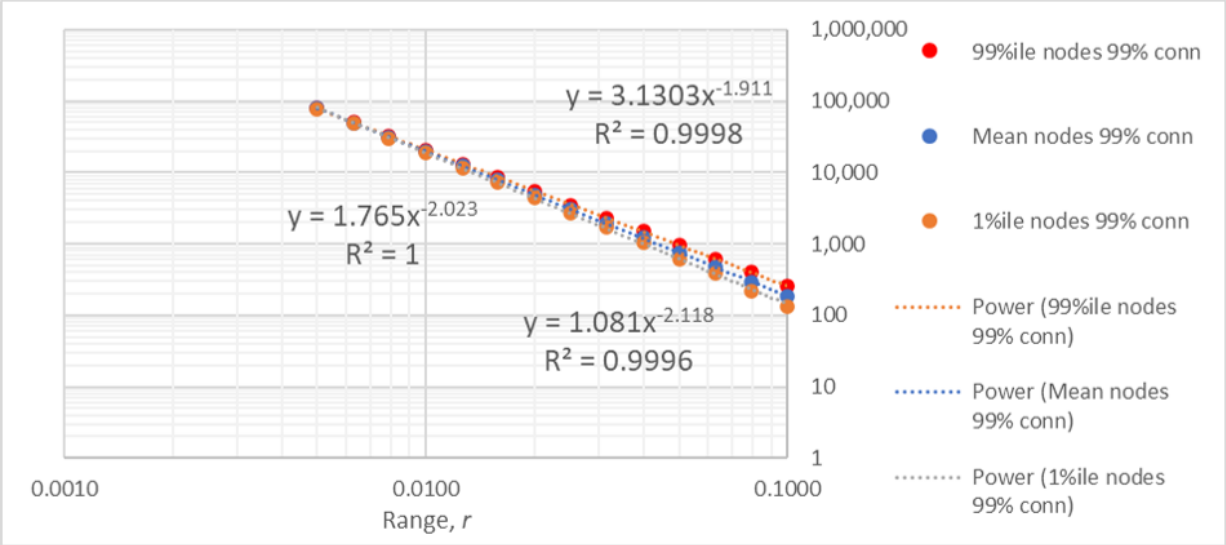

Figure 66: LRMs for 99%ile, mean, and 1%ile number of nodes for at least 99% connectivity of RGGs in a toroidal unit square with n-for-given-r growth.

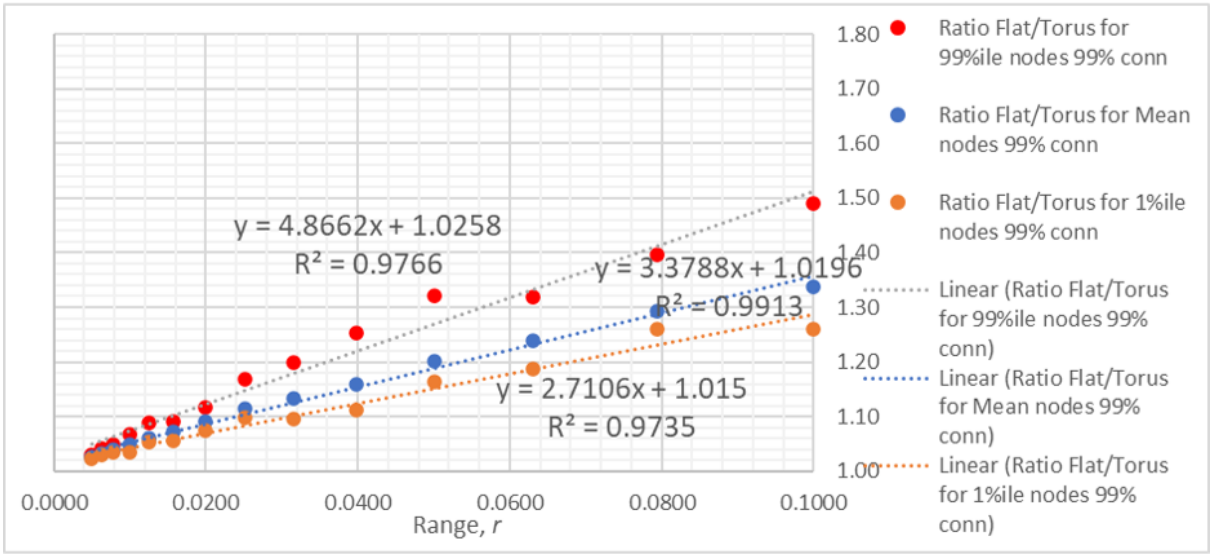

Figure 67: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for 99% connectivity in flat and toroidal spaces.

4.3.1.3 Toroidal Unit Square: No Isolated Nodes for Mid-r Set

Figure 68 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 2.0273\, r^{-2.1856} \tag{58}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.59% and max error of the value of the predicted n is 3.94%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.0082\, r^{-2.2682} \tag{59}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.85% and max error of the value of the predicted n is 7.54%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.4646\, r^{-2.3873} \tag{60}$$

with an R^2 value of 0.9989. For the simulated values of r, MAPE is 6.23% and max error of the value of the predicted n is 15.69%.

As for the cases of full and 99% connectivity, a visual comparison of Figure 29 for RGGs in a flat unit square and Figure 68 for RGGs in a toroidal unit square demonstrates that more nodes are needed to first achieve the elimination of isolated nodes in flat space than in toroidal space. Figure 69 gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to first achieve no isolated nodes in flat and toroidal spaces; note that the scale for the range is linear rather than log. Figure 69 shows that the ratio tends toward 1 for all three of the statistical metrics as the range becomes very small. The ratios for all three of these statistical metrics vary roughly linearly with the range. For the largest ranges in the mid-r set the data and their LRMs show that between roughly 25% to 65% more nodes must be added to first achieve full connectivity of RGGs in flat space than toroidal space for the three statistical metrics reported here.

The R^2 values for the LRMs for both a flat and toroidal unit square are virtually unity. For RGGs in a flat unit square, across the three connectivity conditions (full connectivity, 99% connectivity, and no isolated nodes) and the three statistical metrics (99%ile, mean, and 1%ile) reported here, MAPE varies from 0.60% to 4.57% and the absolute values of max errors vary from 1.42% to 11.47%. For RGGs in a toroidal unit square, MAPE varies from 1.14% to 6.23% and the absolute values of max errors vary from 2.83% to 15.69%. There does not seem to be a consistent trend where the LRMs for flat or toroidal spaces gives larger or smaller errors.[41]

As was observed in Section 4.2 and its subsections for RGGs in a flat unit square, it is seen for RGGs in a toroidal unit square that the number of nodes required to first achieve full connectivity shown in Figure 64 and their LRMs (52), (53), and (54) are substantially different from the number of nodes required for the disappearance of isolated nodes shown in Figure 68 and their LRMs (58), (59), and (60). Even in a toroidal space with n-for-given-r growth, full connectivity does not necessarily occur when the last isolated node is absorbed.

As discussed in Section 4.2.1.4 for RGGs in a flat unit square with n-for-given-r growth, this observation raises the question as to how many additional nodes must be added to an RGG in a toroidal unit square beyond

[41] This seems to be counter-intuitive as one might think that, due to the large impact of border effects, one or the other of the two spaces should give a better fit for the linear regression models.

the point where the last isolated node disappears. Figure 70 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. As in flat space, for RGGs in a toroidal unit square the 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the mid-r set to thousands and hundreds of thousands for the smallest ranges simulated in these experiments.

Figure 70 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

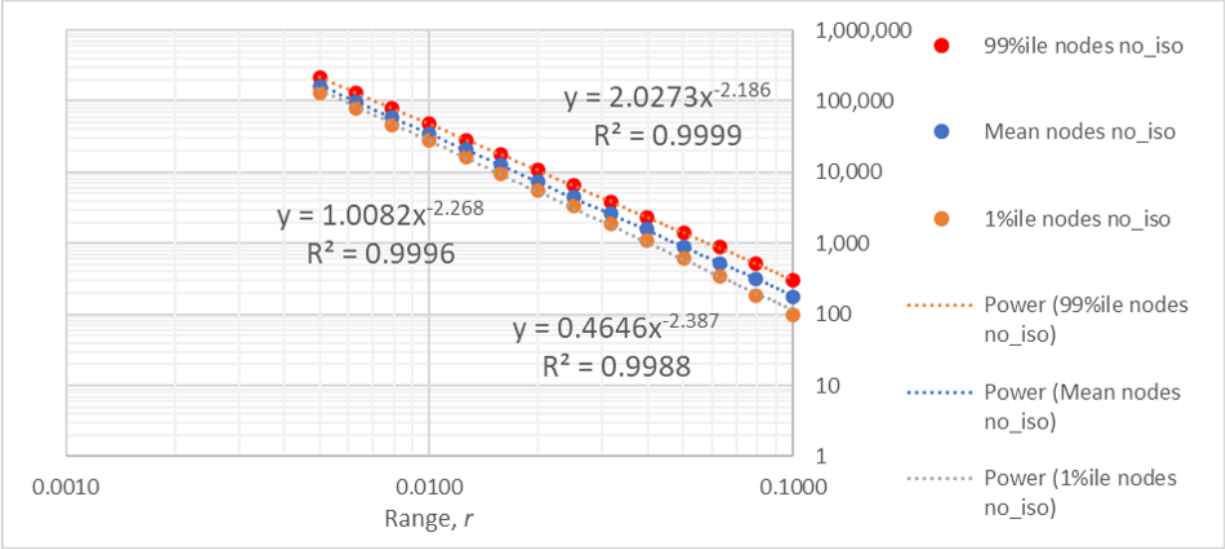

Figure 68: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a toroidal unit square with n-for-given-r growth.

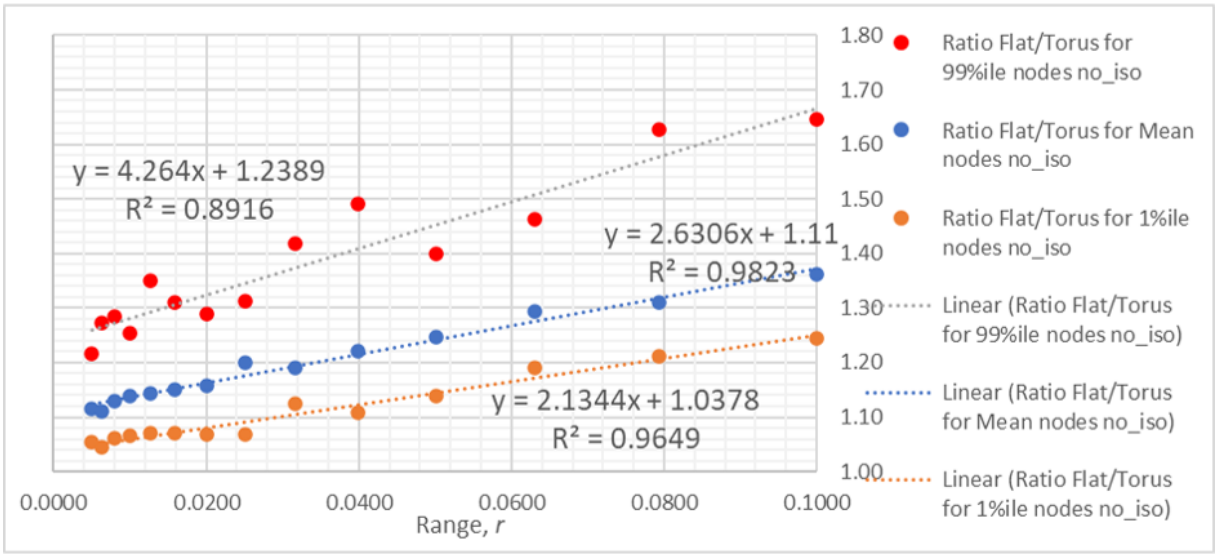

Figure 69: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for no isolated nodes in flat and toroidal spaces.

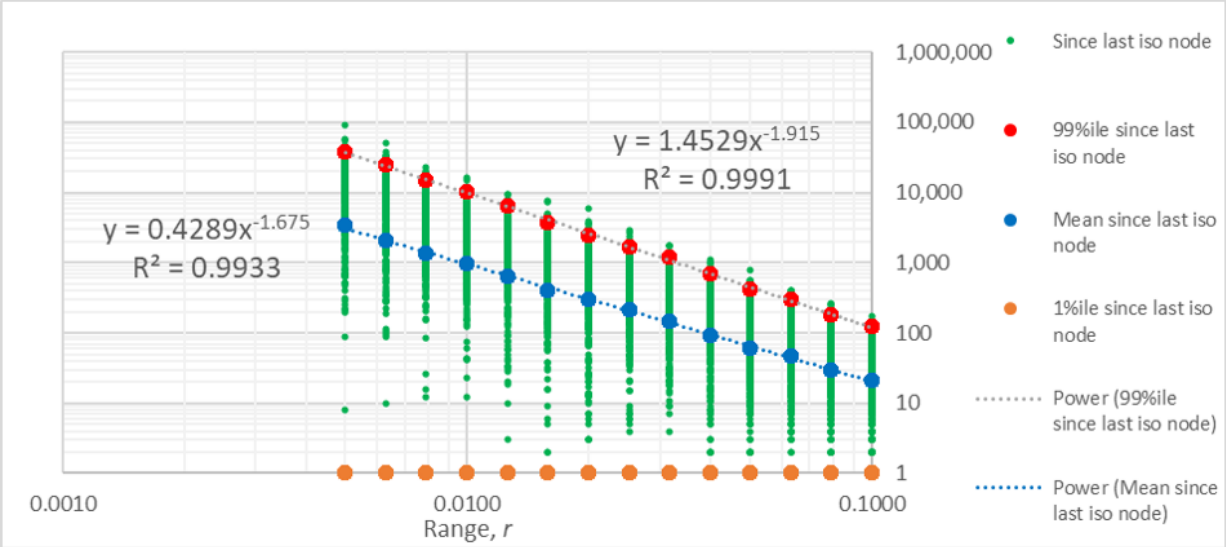

Figure 70: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.3.1.4 Toroidal Unit Square: Number of Neighbors Needed for Full Connectivity for Mid-r Set

Figure 71 shows a linear-log plot with LRMs for the 99%ile, mean, and 1%ile values for the mean degree required to first achieve full connectivity of RGGs with n-for-given-r growth. The regression models for the statistical metrics are linear functions of the mean degree and the log of the range and thus they are modeled as log laws rather than the log-log power law functions that were shown to work well for the statistical metrics for the numbers of nodes.

Figure 72 is a scatter plot of the predicted vs. observed mean degree required to first achieve full connectivity for all of the simulated geometries using the ranges in the mid-r set. For each geometry the predicted mean degree is the value of μ given by (9) for each given range r and the number of nodes n required to first achieve full connectivity. The red line represents the value where the observed mean degree would have equaled the predicted mean degree for a simulation. In contrast to the case of a flat unit square as shown in Section 4.2.1.5, for RGGs in a toroidal unit square the observed values are distributed tightly and symmetrically about the predicted values due to the nodes near the border having the same mean number of neighbors as the other nodes in the space.

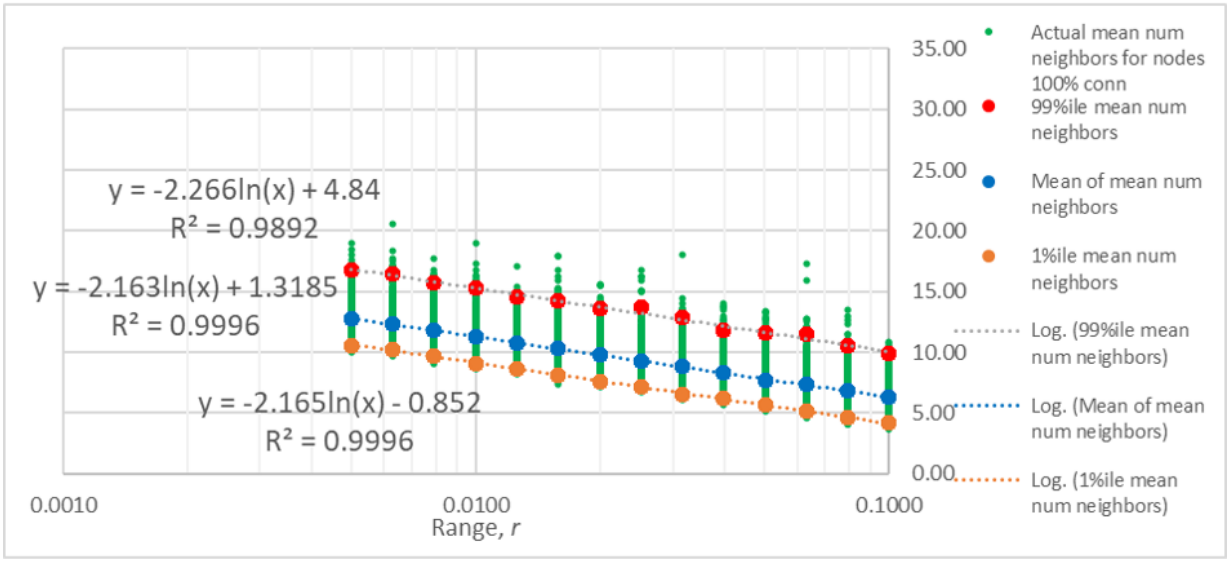

Figure 71: LRMs for 99%ile, mean, and 1%ile values for mean degree for full connectivity in a toroidal unit square with n-for-given-r growth.

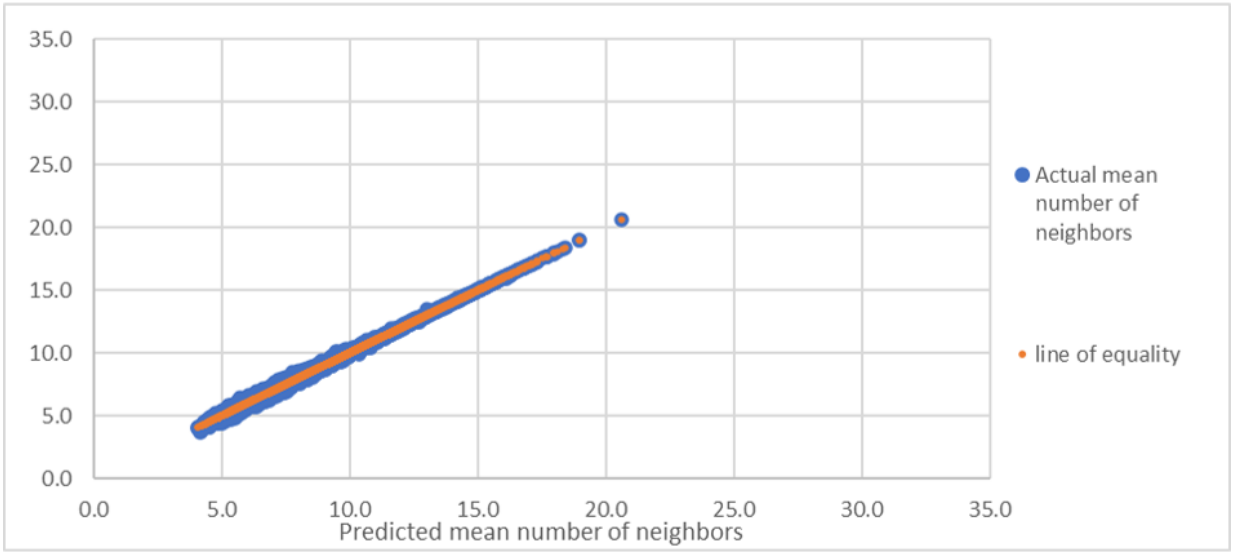

Figure 72: Observed and predicted mean degree for full connectivity in a toroidal unit square with n-for-given-r growth.

4.3.1.5 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a toroidal unit square with n-for-given-r growth for the mid-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 73 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a toroidal unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a toroidal unit square with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.382 for $r = 0.1000$ and increases to about 0.724 for $r = 0.0050$.

This probability increases as r decreases (and consequently the number of nodes increases) but, as shown for the case of a flat unit square, it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a toroidal unit square even for the smallest range in the mid-r set.

Figure 74 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. For small ranges, very few nodes are required to first achieve full connectivity after the disappearance of isolated nodes. The ratios may asymptotically approach 1 for even smaller ranges but the ratio was still observed to be 1.028 for the smallest range in the mid-r set.

While Figure 73 shows that an RGG in a toroidal unit square with n-for-given-r growth does not inevitably become fully connected at the same instant the last isolated node disappears, Figure 74 shows that for the smaller ranges in the mid-r set very few nodes must be added to the 99%ile number of nodes required to first achieve no isolated nodes to reach the 99%ile number of nodes required to first become fully connected.

Figure 75 shows the error of the LRM (52) given in Section 4.3.1.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 4.3.1.1 that MAPE for (52) is 1.64% and max error of the predicted value of n is -4.58%.

Figure 75 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. The error values are very close to those obtained for the LRM: Penrose's predicted values of n have MAPE 2.05% and max error 5.02%.

Figure 76 shows the same error statistics as in Figure 75 but now for the 1%ile number of nodes required to first achieve full connectivity, which is the threshold used here for RGGs to be disconnected. Figure 76 shows the error percentages for the LRM (54) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. The LRM provides fairly close estimates while the predicted number of nodes from Penrose's expressions have far greater errors for the largest ranges in the mid-r set. For the smallest ranges, however, Penrose's Theorem has smaller error than the LRM.

Figure 77 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 64.

The maximum of the c_P observed for any range was 10.0674. The visible trend in Figure 77 is that all three of the statistical metrics plotted for c_P seem to be practically flat (with some noise) as the range varies, ranging from 4.0554 to 4.8041 for the 99%ile values of n and from -1.2539 to -0.6160 for the 1%ile values of n. The mean value of c_P is 4.4037 for the 99%ile values of n and -1.0000 for the 1%ile values of n, which are very

close to Penrose's values of $c_P = 4.6001$ required for $C = 0.99$ and fairly close to $c_P = -1.5272$ for $C = 0.01$.

It may be surprising that there is a very close match between the observed mean value of c_P for the 99%ile number of nodes required to first achieve full connectivity and Penrose's value of c_P for $C = 0.99$ even for the larger values of r when it is observed that, for these values of r, on average between 10% and 20% additional nodes are needed to first achieve full connectivity over those required for no isolated nodes as shown in Figure 74.

It is noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

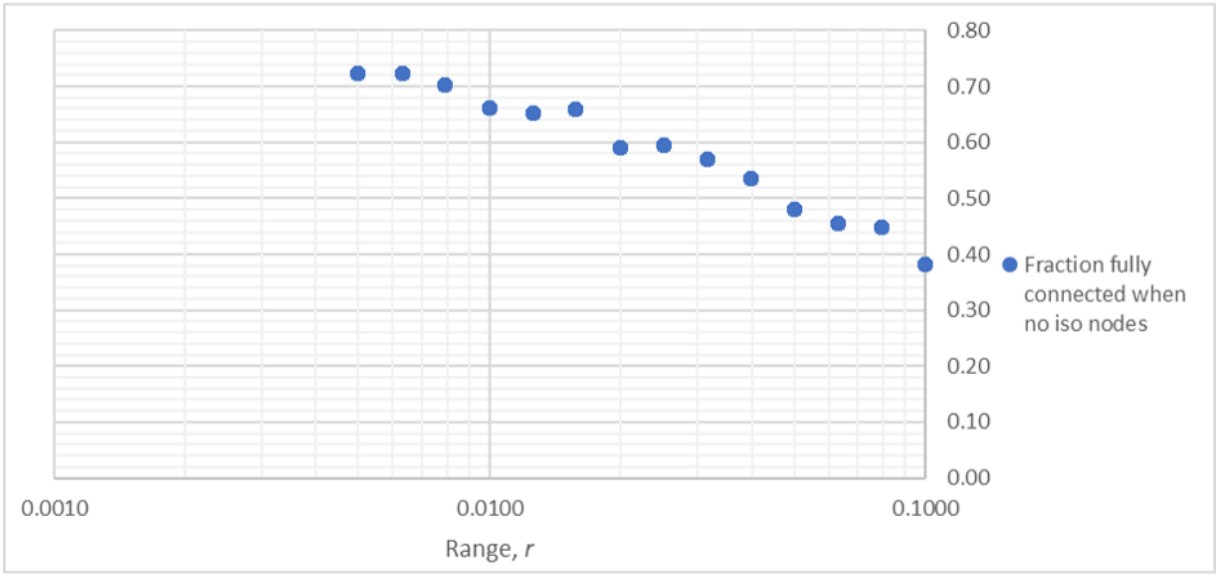

Figure 73: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a toroidal unit square to become fully connected with n-for-given-r growth.

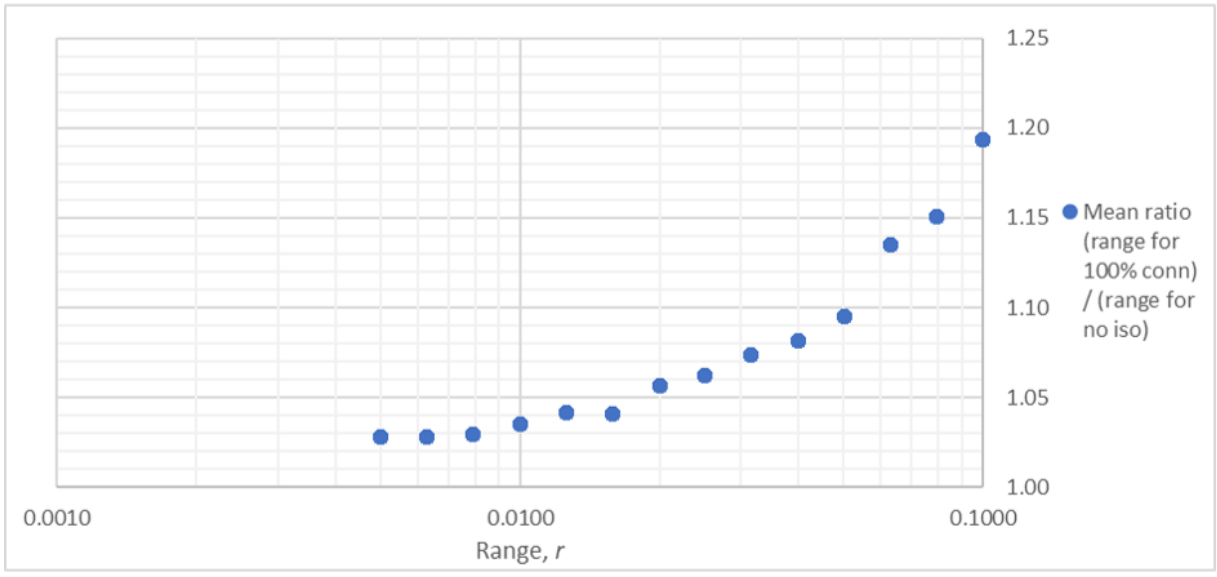

Figure 74: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a toroidal unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

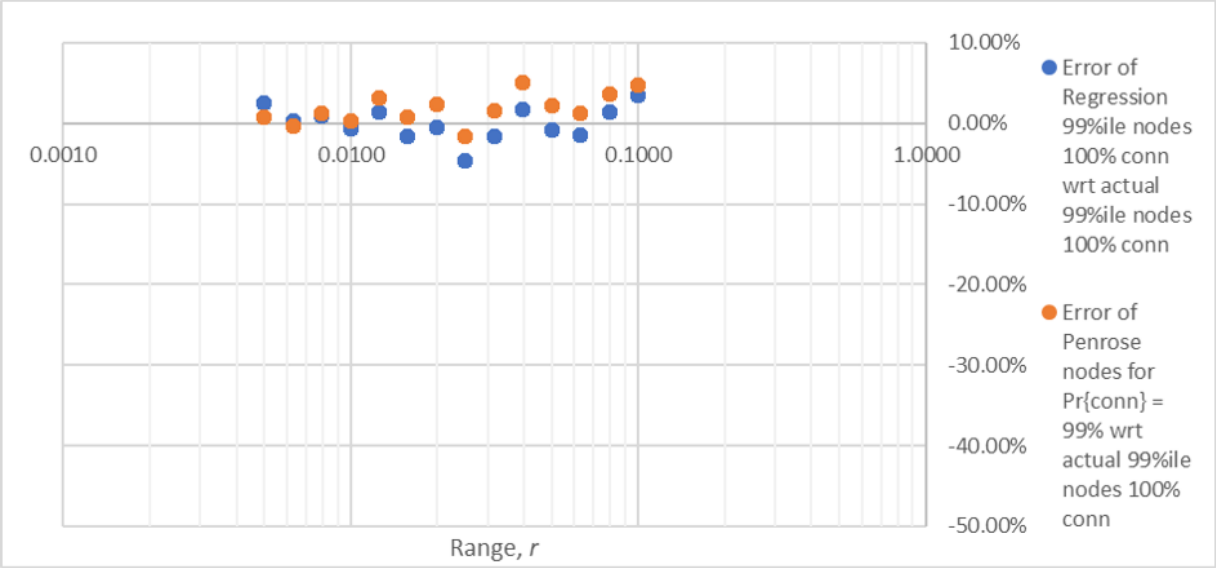

Figure 75: Relative error of LRM (52) for 99%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

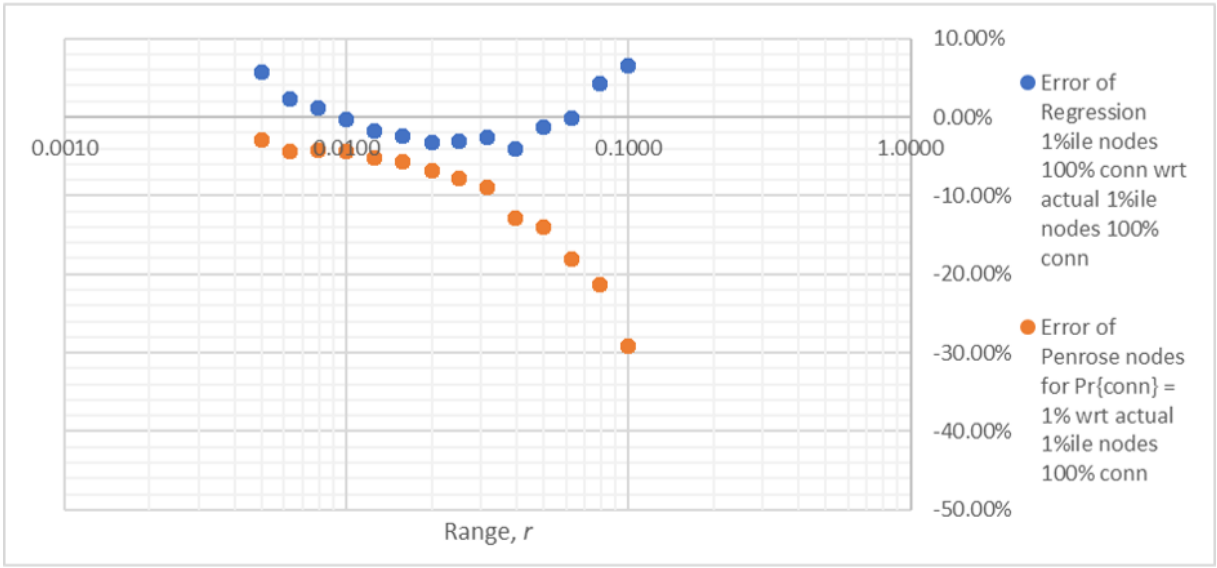

Figure 76: Relative error of LRM (54) for 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

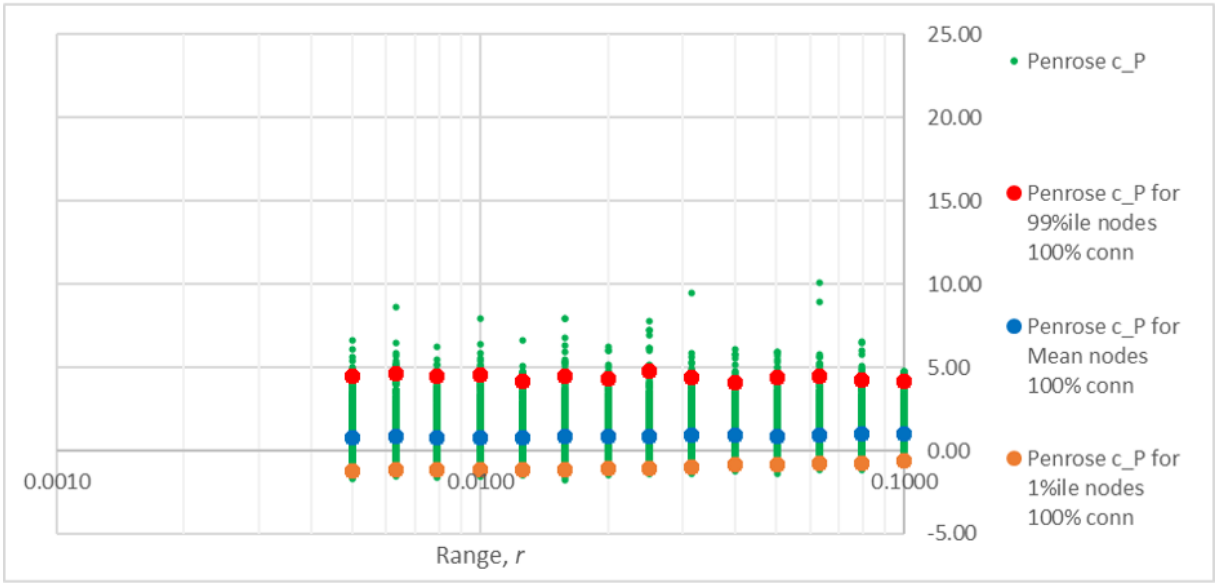

Figure 77: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.3.2 Toroidal Unit Square with n-for-given-r Growth for Small Ranges

Section 4.3.2 through Section 4.3.2.3 repeat some of the analyses presented in Section 4.3.1 and its subsections for ranges in the mid-r set but now for the small-r set consisting of the ranges from 0.005012 down to 0.000251.

LRMs are obtained for full connectivity and no isolated nodes.

These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

4.3.2.1 Toroidal Unit Square: Full Connectivity for Small-r Set

Figure 78 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the small-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 3.2288\, r^{-2.0989} \tag{61}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.33% and max error of the value of the predicted n is -3.96%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.0280\, r^{-2.1335} \tag{62}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.64% and max error of the value of the predicted n is 1.25%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.5224\, r^{-2.1539} \tag{63}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.86% and max error of the value of the predicted n is 2.48%.

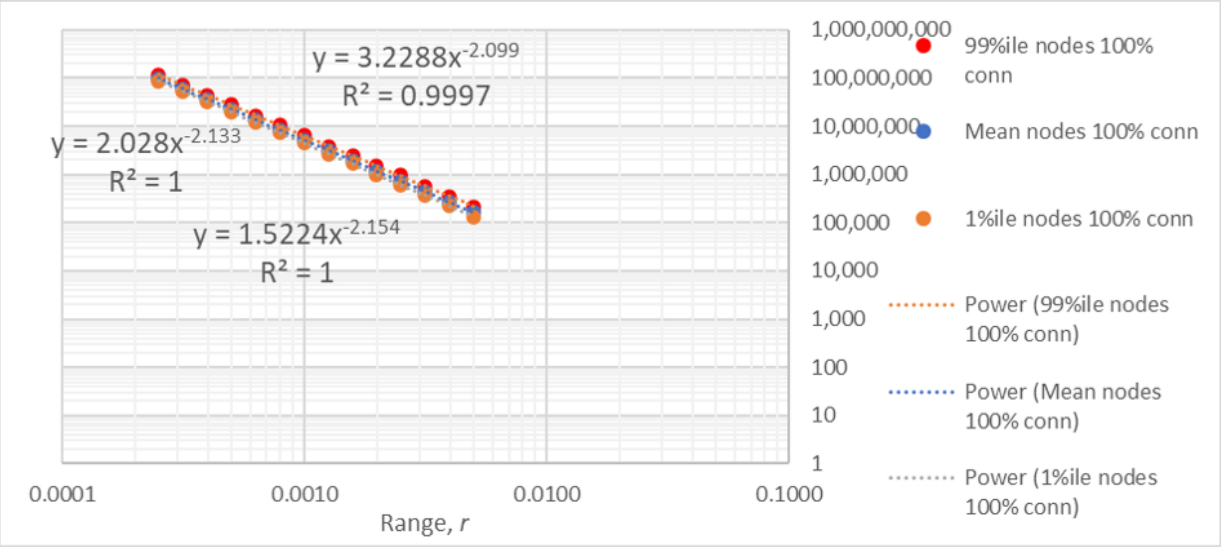

Figure 78: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the small-r set.

4.3.2.2 Toroidal Unit Square: No Isolated Nodes for Small-r Set

Figure 79 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the small-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 3.1763\, r^{-2.1000} \tag{64}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.44% and max error of the value of the predicted n is 4.10%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.9464\, r^{-2.1375} \tag{65}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.68% and max error of the value of the predicted n is 1.36%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.4037\, r^{-2.1629} \tag{66}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 1.22% and max error of the value of the predicted n is 2.74%.

It is seen for RGGs in a toroidal unit square that the number of nodes required to first achieve full connectivity shown in Figure 78 and their LRMs (61), (62), and (63) are larger than the number of nodes required for the disappearance of isolated nodes shown in Figure 79 and their LRMs (64), (65), and (66). Full connectivity does not necessarily occur when the last isolated node is absorbed.

Figure 80 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. The 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the small-r set to millions and tens of millions for the smallest ranges simulated in these experiments.

Figure 80 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

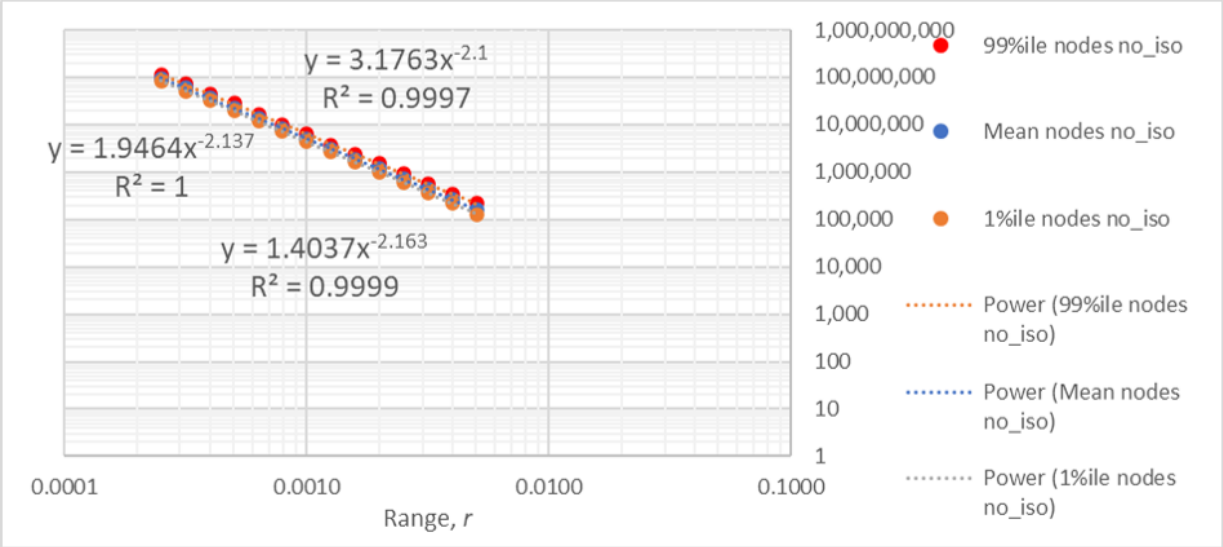

Figure 79: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a toroidal unit square with n-for-given-r growth.

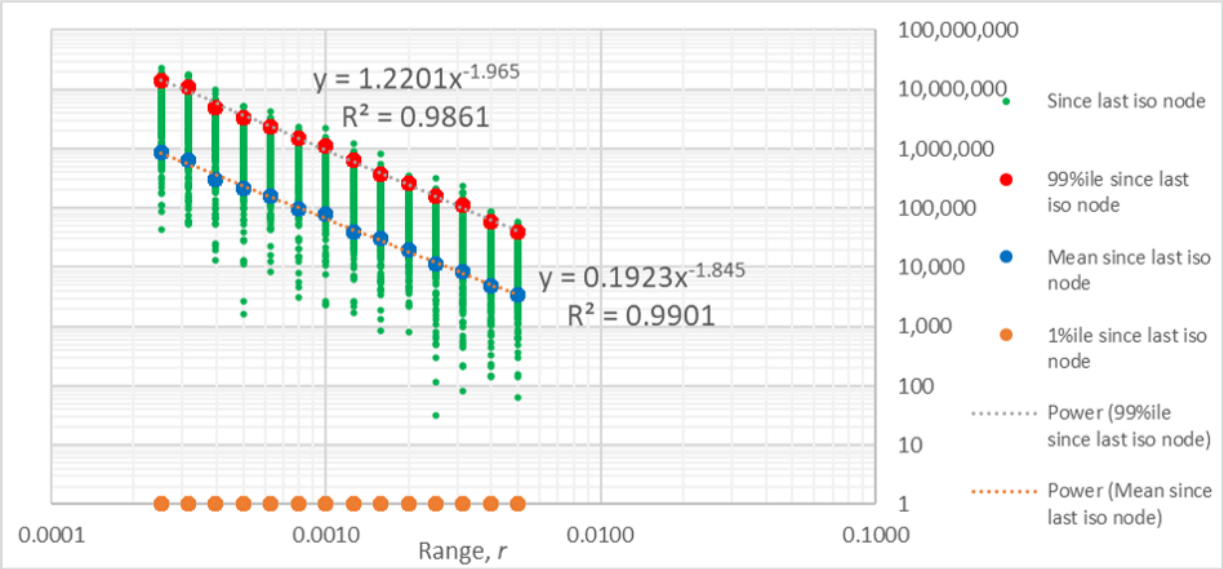

Figure 80: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.3.2.3 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a toroidal unit square with n-for-given-r growth using the small-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 81 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a toroidal unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a toroidal unit square with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.742 for $r = 0.005012$ and increases to about 0.832 for $r = 0.000251$.

This probability increases as r decreases but it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a toroidal unit square even for very small r (and thus very large n).

Figure 82 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. For small ranges, very few additional nodes are required to first achieve full connectivity after the disappearance of isolated nodes. The ratios may asymptotically approach 1 for even smaller ranges.

While Figure 81 shows that an RGG in a toroidal unit square with n-for-given-r growth does not inevitably become fully connected at the same instant the last isolated node disappears, Figure 82 shows that very

few nodes must be added to the 99%ile number of nodes required to first achieve no isolated nodes to reach the 99%ile number of nodes required to first become fully connected.

Figure 83 shows the error of the LRM (61) given in Section 4.3.2.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. It can be seen that the error percentages hover close to zero.

Figure 83 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. The error values are very small and are very close to those obtained for the LRM: for the ranges in the small-r set, Penrose's predicted values of n have MAPE 1.51% and max error 4.67%, which are only slightly larger than those obtained using the LRM (61).

Figure 84 shows the same error statistics as in Figure 83 but now for the 1%ile number of nodes required to first achieve full connectivity. Figure 84 shows the error percentages for the LRM (63) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. The relative error values are again very close to those obtained for the LRM and both are very small: for the ranges in the small-r set, Penrose's predicted values of n have MAPE 1.82% and max error -3.59% which are again only slightly larger than those obtained using the LRM (63).

Figure 85 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 78.

The maximum of the c_P observed for any range was 9.0614. The visible trend in Figure 85 is that all three of the statistical metrics plotted for c_P seem to be practically flat (with some noise) as the range varies. The mean value of c_P is 4.5156 for the 99%ile values of n and -1.3098 for the 1%ile values of n, which are very close to Penrose's values of $c_P = 4.6001$ required for $C = 0.99$ and fairly close to $c_P = -1.5272$ for $C = 0.01$.

There is a very close match between the observed mean value of c_P for the 99%ile number of nodes required to first achieve full connectivity and Penrose's value of c_P for $C = 0.99$, which would be expected because very few additional nodes are needed to first achieve full connectivity over those required for no isolated nodes as shown in Figure 82.

It is noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

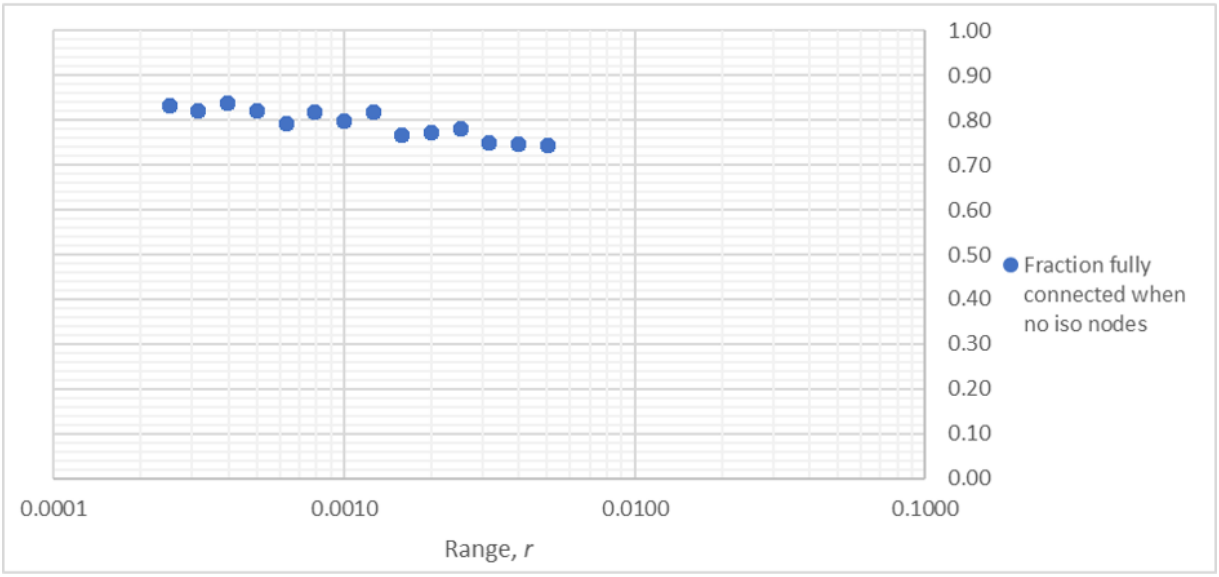

Figure 81: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a toroidal unit square to become fully connected with n-for-given-r growth.

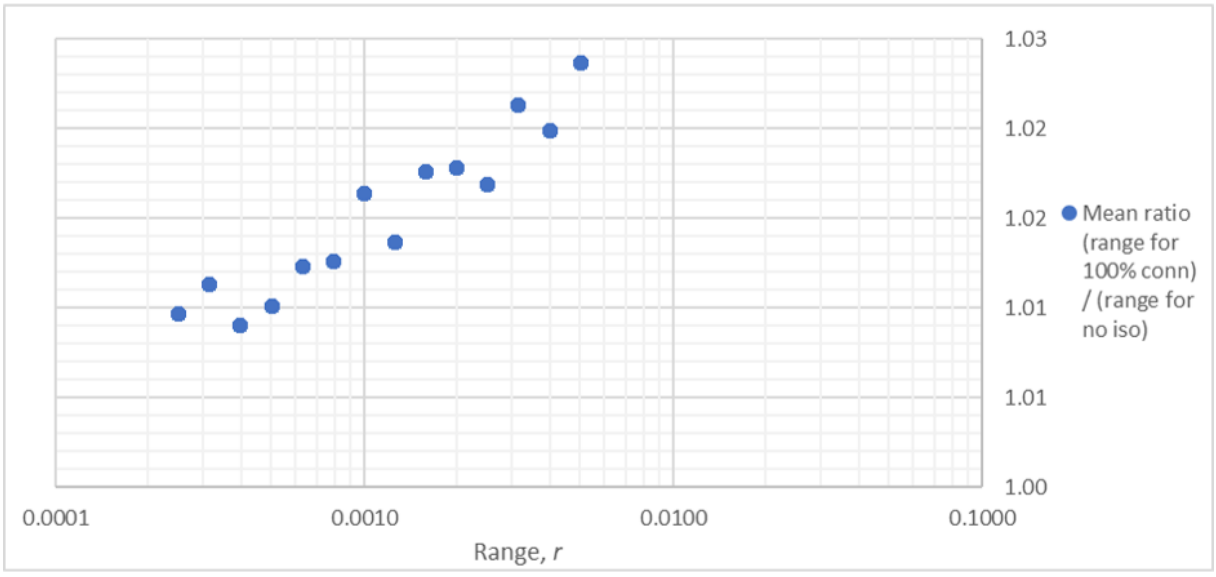

Figure 82: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a toroidal unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

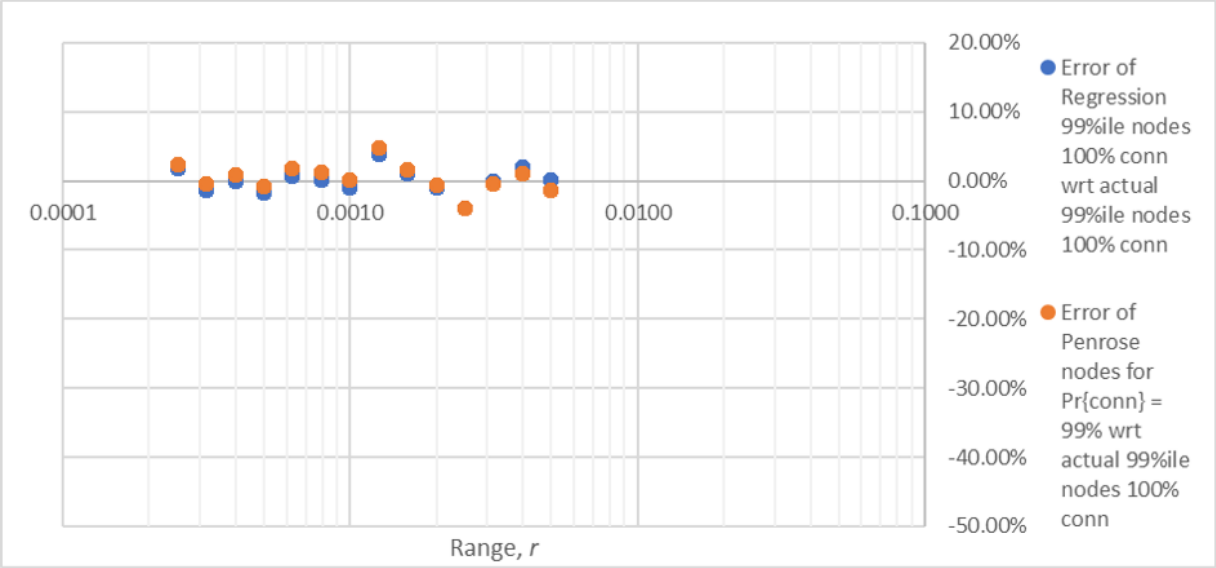

Figure 83: Relative error of LRM (61) for 99%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

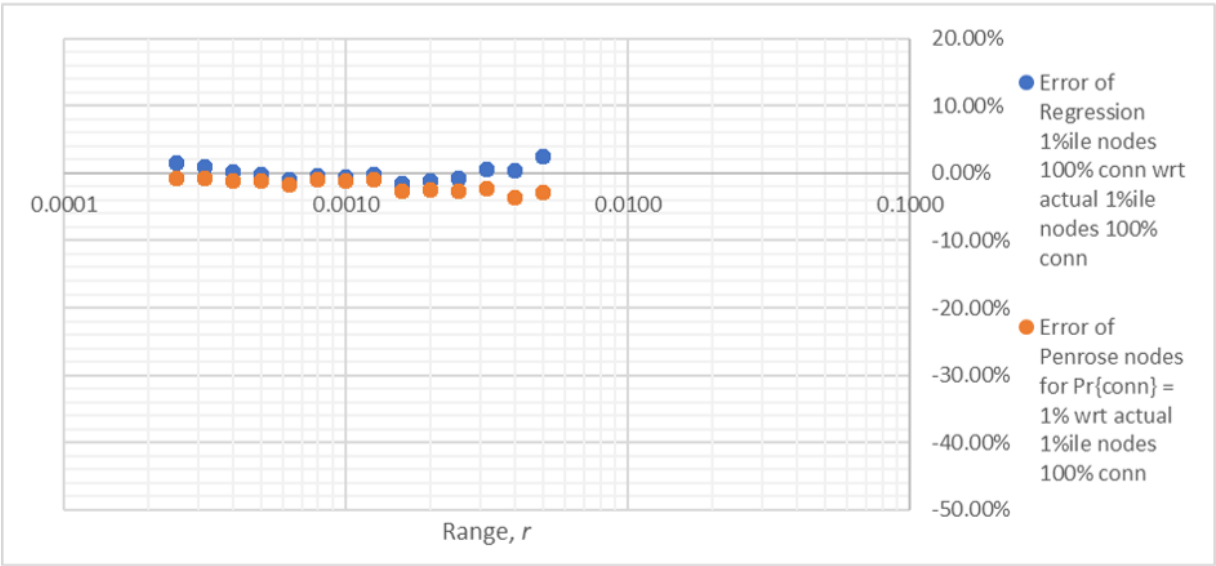

Figure 84: Relative error of LRM (63) for 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

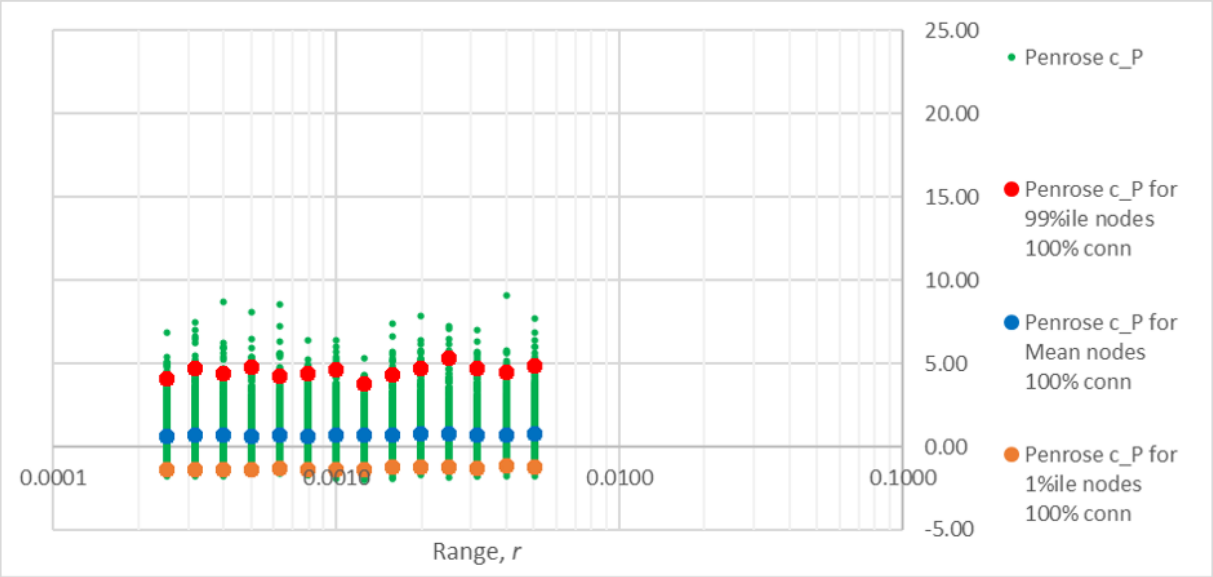

Figure 85: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.3.3 Toroidal Unit Square with n-for-given-r Growth for Mid and Small Ranges

Section 4.3.3 through Section 4.3.3.3 repeat some of the analyses presented in Section 4.3.1 and its subsections for ranges in the mid-r set and Section 4.3.2 and its subsections for ranges in the small-r set but now for ranges in the mid+small-r set which is the set of ranges from 0.100000 down to 0.000251.

LRMs are obtained for full connectivity and no isolated nodes.

These sections compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

4.3.3.1 Toroidal Unit Square: Full Connectivity for Mid+Small-r Set

Figure 86 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the mid+small-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 2.5562\, r^{-2.1330} \tag{67}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.95% and max error of the value of the predicted n is 9.11%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.5066\, r^{-2.1774} \tag{68}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 4.14% and max error of the value of the predicted n is 11.95%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 0.9940\, r^{-2.2169} \tag{69}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 6.06% and max error of the value of the predicted n is 18.70%.

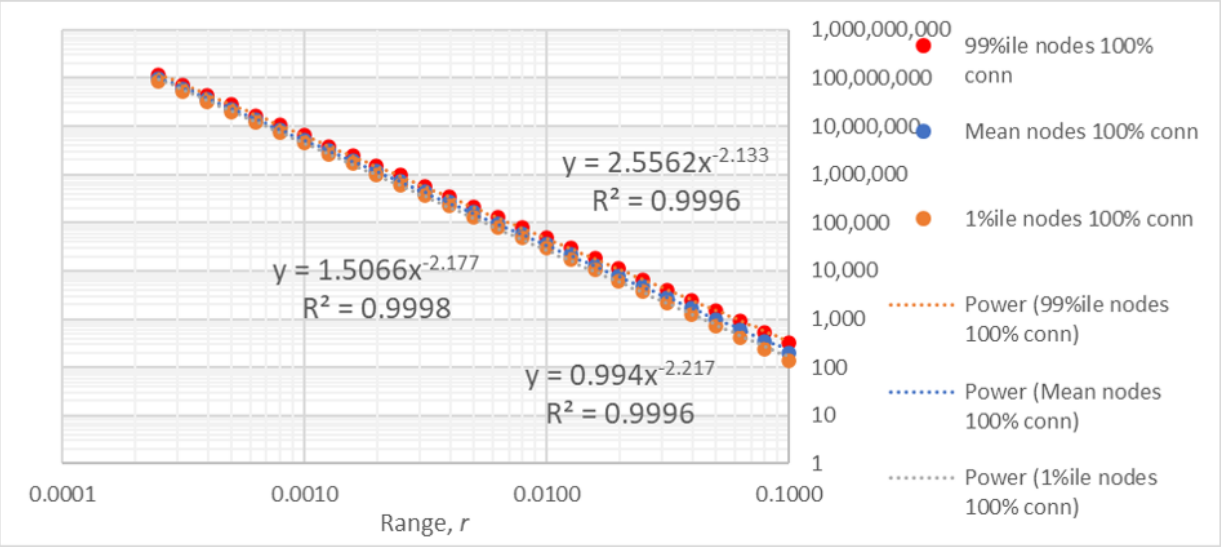

Figure 86: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the mid+small-r set.

4.3.3.2 Toroidal Unit Square: No Isolated Nodes for Mid+Small-r Set

Figure 87 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid+small-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 2.4595\, r^{-2.1373} \tag{70}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 3.02% and max error of the value of the predicted n is 9.94%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.3274\, r^{-2.1942} \tag{71}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 5.57% and max error of the value of the predicted n is 18.63%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.7326\, r^{-2.2598} \tag{72}$$

with an R^2 value of 0.9990. For the simulated values of r, MAPE is 10.56% and max error of the value of the predicted n is 53.17%.

It is seen for RGGs in a toroidal unit square that the number of nodes required to first achieve full connectivity shown in Figure 86 and their LRMs (67), (68), and (69) are larger than the number of nodes required for the disappearance of isolated nodes shown in Figure 87 and their LRMs (70), (71), and (72). Full connectivity does not necessarily occur when the last isolated node is absorbed.

Figure 88 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. The 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the mid+small-r set to millions and tens of millions for the smallest ranges simulated in these experiments.

Figure 88 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

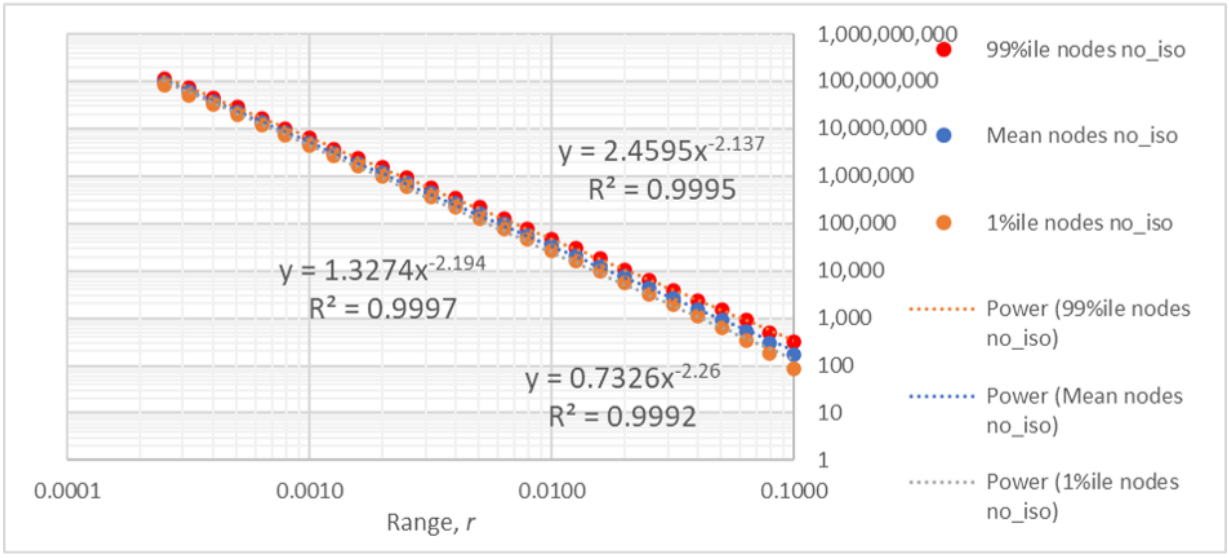

Figure 87: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a toroidal unit square with n-for-given-r growth.

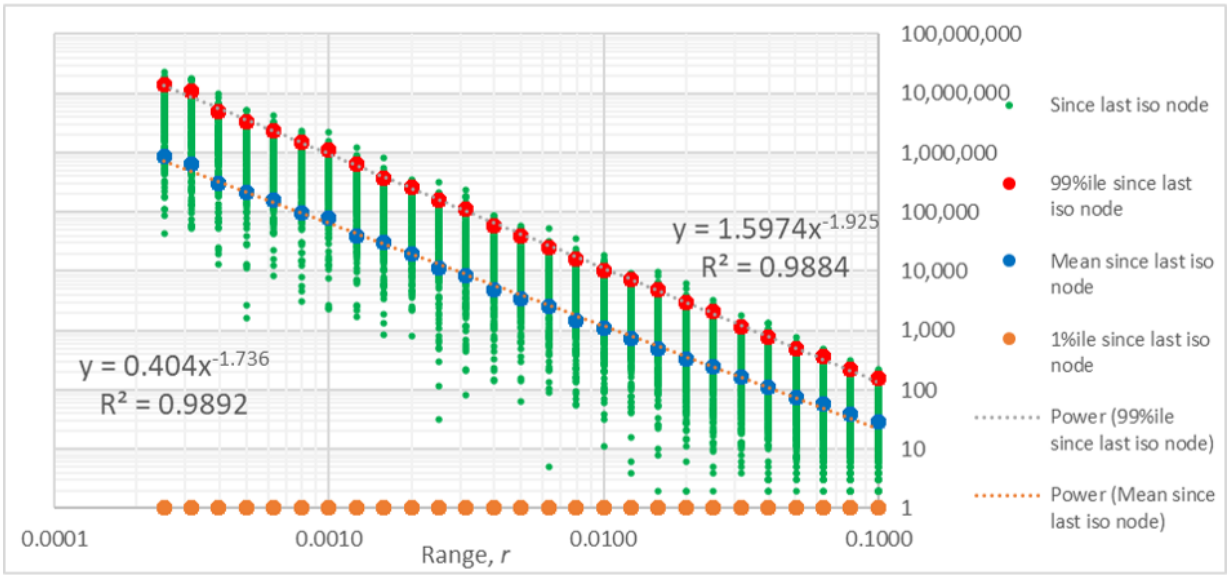

Figure 88: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

4.3.3.3 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid+Small-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a toroidal unit square with n-for-given-r growth. The following figures show results that test the accuracy of Penrose's Theorem for this situation using the mid+small-r set.

Figure 89 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a toroidal unit square with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a toroidal unit square with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.385 for $r = 0.100000$ and increases to about 0.832 for $r = 0.000251$. As the range decreases the bend in the curve is apparent, indicating that the fraction approaches 1 more slowly for the smaller ranges.

This probability increases as r decreases but it is clear that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a toroidal unit square even for very small r (and thus very large n).

Figure 90 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. For small ranges, very few additional nodes are required to first achieve full connectivity after the disappearance of isolated nodes. The ratios may asymptotically approach 1 for even smaller ranges.

While Figure 89 shows that an RGG in a toroidal unit square with n-for-given-r growth does not inevitably become fully connected at the same instant the last isolated node disappears, Figure 90 shows that, except for the largest ranges in the set, very few nodes must be added to the 99%ile number of nodes required to first achieve no isolated nodes to reach the 99%ile number of nodes required to first become fully connected.

Figure 91 shows the error of the LRM (67) given in Section 4.3.3.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. This figure also includes the relative error of the LRM (52) given in Section 4.3.1.1 that was obtained for the mid-r set; it can be seen that this LRM has increasing error as range decreases from the smallest value included in the mid-r set.

Figure 91 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. It can be seen that the error percentages for the LRM and Penrose's predicted values of n are close to each other and both hover close to zero, with Penrose's values even closer to zero than those for the LRM. For the ranges in the mid+small-r set, Penrose's predicted values of n have MAPE 1.79% and max error 5.85%

Figure 92 shows the same error statistics as in Figure 91 but now for the 1%ile number of nodes required to first achieve full connectivity. Figure 92 shows the error percentages for the LRM (69) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. For the ranges in the mid+small-r set Penrose's predicted values of n have MAPE 6.37% and max error -29.70% but the error percentages for Penrose's values are closer to zero than those for the LRM for ranges 0.012589 and smaller and the error percentages for the LRM actually get worse for the smallest ranges in the set.

Figure 93 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 86.

The maximum of the c_P observed for any range was 10.3797. The visible trend in Figure 93 is that all three of the statistical metrics plotted for c_P seem to be practically flat (with some noise) as the range varies. The mean value of c_P is 4.4690 for the 99%ile values of n and -1.1426 for the 1%ile values of n, which are very close to Penrose's values of $c_P = 4.6001$ required for $C = 0.99$ and fairly close to $c_P = -1.5272$ for $C = 0.01$.

As noted in Section 4.3.1.5 for the results of the mid-r set, it may be surprising that there is a very close match between the observed mean value of c_P for the 99%ile number of nodes required to first achieve full connectivity and Penrose's value of c_P for $C = 0.99$ even for the larger values of r when it is observed that, for these values of r, on average between 10% and 20% additional nodes are needed to first achieve full connectivity over those required for no isolated nodes as shown in Figure 90.

It is noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

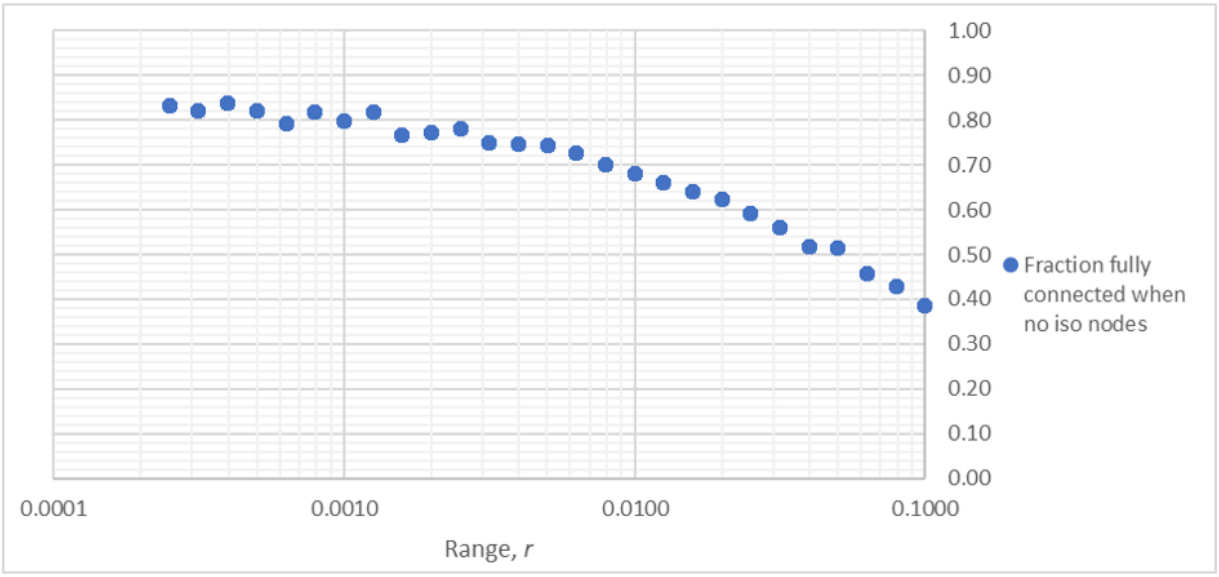

Figure 89: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a toroidal unit square to become fully connected with n-for-given-r growth.

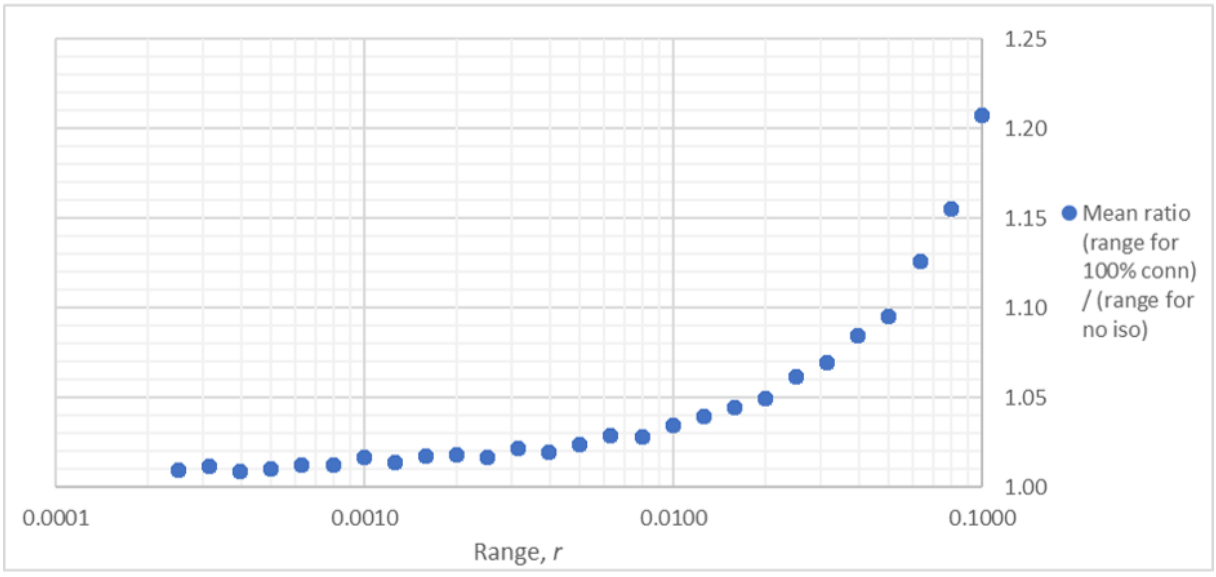

Figure 90: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a toroidal unit square with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

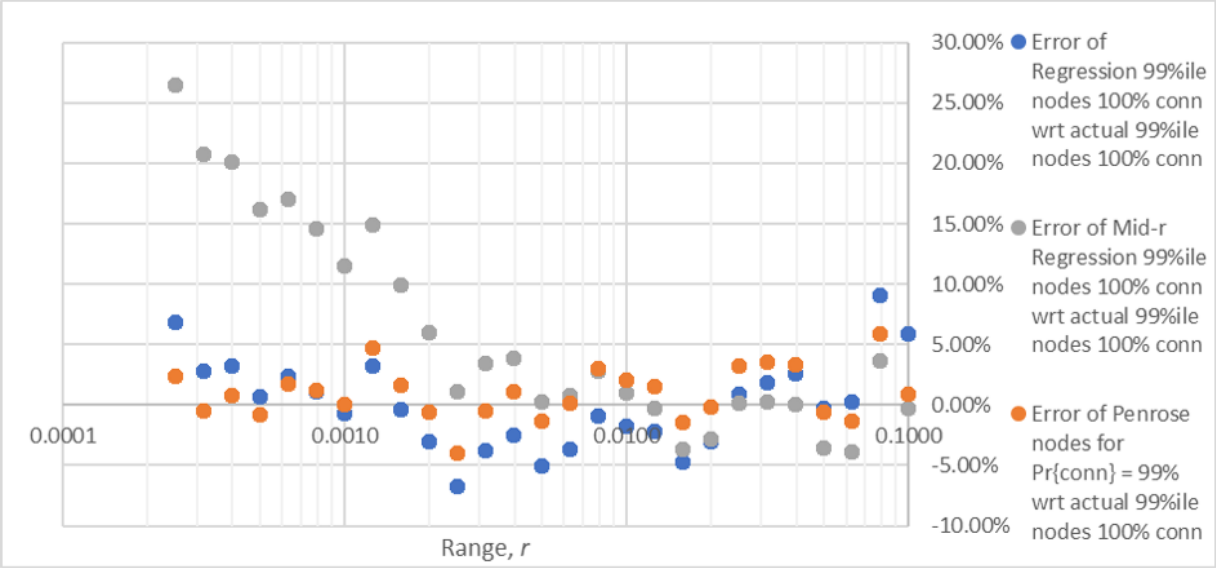

Figure 91: Relative error of LRM (67) obtained from the mid+small-r set and LRM (52) obtained from the mid-r set for 99%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

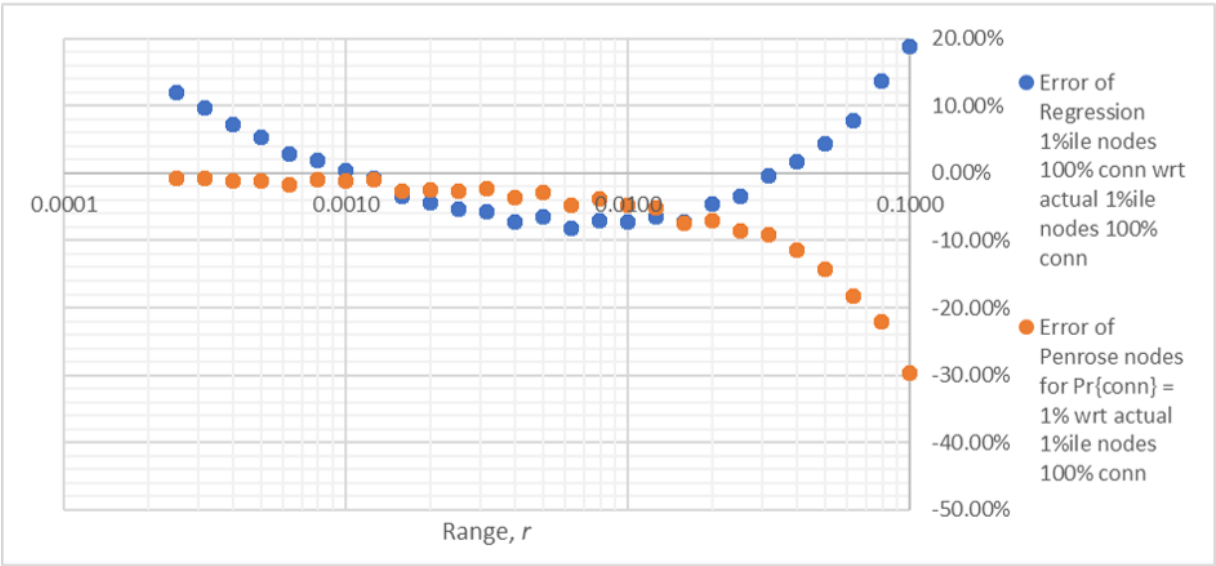

Figure 92: Relative error of LRM (69) for 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

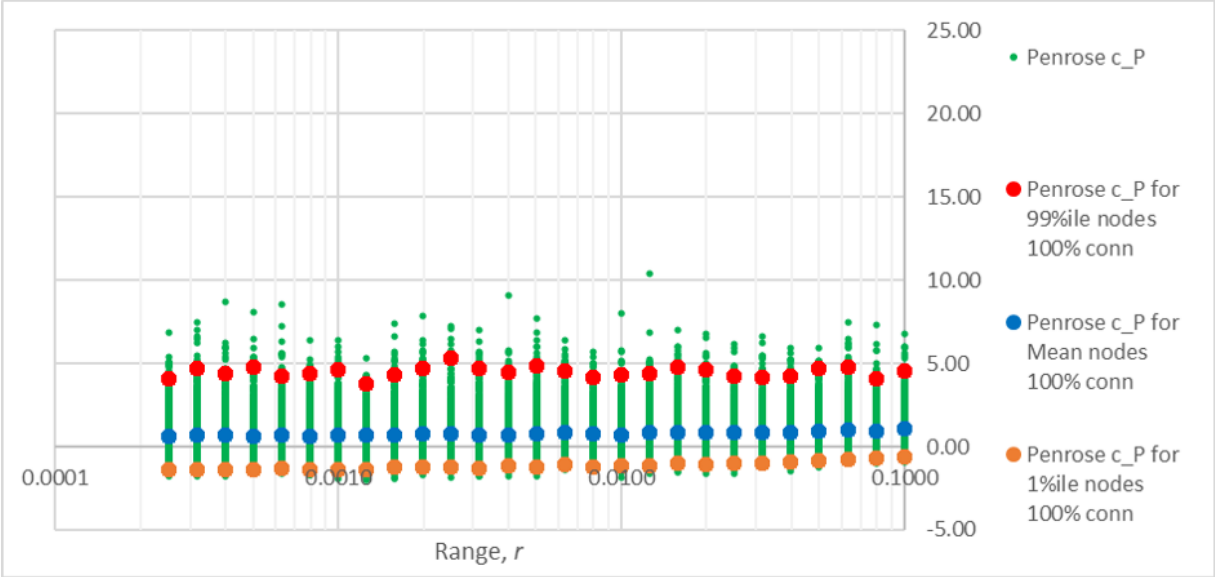

Figure 93: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

4.3.4 Toroidal Unit Square: Comparing Linear Regression Models from Mid-r Set, Small-r Set, and Mid+Small-r Set

This section compares the MAPE and max error resulting from using an LRM derived from one set of ranges to predict numbers of nodes for other sets of ranges. Only the results for the 99%ile numbers of nodes for full connectivity are compared.

- LRM (52) obtained from mid-r set:
  - Applied to mid-r set: MAPE 1.64%, max error −4.58%
  - Applied to small-r set: MAPE 11.53%, max error 28.30%
  - Applied to mid+small-r set: MAPE 3.38%, max error 9.87%
- LRM (61) obtained from small-r set:
  - Applied to mid-r set: MAPE 11.84%, max error 26.44%
  - Applied to small-r set: MAPE 1.33%, max error −3.96%
  - Applied to mid+small-r set: MAPE 3.04%, max error 6.80%
- LRM (67) obtained from mid+small-r set:
  - Applied to mid-r set: MAPE 7.00%, max error 26.44%
  - Applied to small-r set: MAPE 6.41%, max error 26.42%
  - Applied to mid+small-r set: MAPE 2.95%, max error 9.11%

For toroidal RGGs with n-for-given-r growth the smallest MAPE and smallest max error resulted when the LRM obtained from a set of ranges is applied to that set of ranges.

These results indicate that the linearity of the log-log transformed 99%ile number of nodes for toroidal RGGs with n-for-given-r growth does not strictly hold over the various intervals represented by the mid-r set, small-r set, and their union but it is still fairly good. The best results are obtained when the LRM derived from the mid-r set is used for the mid-r ranges and the LRM derived from the small-r set is used for the small-r ranges; however, the LRM derived from the mid+small-r set still performed well over that entire set of ranges.

4.4 Unit Circle with n-for-given-r Growth for Mid Ranges

This section considers RGGs consisting of nodes scattered uniformly at random positions in a circular space of area $A = 1$ (and thus radius $a = 1/\sqrt{\pi} \approx 0.5642$) with no wraparound at the border. This is a flat space so there are border effects. The RGGs evolve according to the n-for-given-r growth model. Only the mid-r set of ranges is modeled.

For each simulated geometry for a given configuration, nodes with a fixed range r are added one-by-one[42] to determine the number of nodes n required to first achieve the condition of 100% connectivity, 99% connectivity, or no isolated nodes. The 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve each of those conditions are obtained for that configuration.

For configurations with $r = 0.0100$, Figure 94 shows the position where the last node was added to first achieve full connectivity for each of 1,000 simulated geometries. As noted in Section 4.2 for RGGs in a flat unit square with n-for-given-r growth, the last node added always bridges an existing isolated node or smaller component into the GC. Although the last node added tends to be near the border the last node added might not itself be a border node. This figure demonstrates for the unit circle the consequence of the border effects experienced by RGGs in flat spaces, which is that border nodes have fewer near neighbors on average so they are less likely to be absorbed into the GC at the time they are added.[43]

[42] In the simulation of this growth process, a new node position is generated by creating an (x, y) coordinate distributed uniformly in a square with sides equal to twice the specified radius of the circle and centered at the origin. A position that falls outside the radius centered at the origin is discarded and another position is generated. The process iterates until a position within the specified radius is obtained.

An alternate procedure for creating points randomly in a circle would be to generate a magnitude uniformly at random from zero to the radius and generate an angle uniformly at random from zero to 360° but that method would not generate a uniform distribution of node positions as nodes would be closer together near the origin and farther apart as they approach the border of the circle.

[43] The circle appears smaller than the flat unit square shown in Figure 25 and the toroidal unit square shown in Figure 63 because the axes in Figure 94 span −1 to 1 rather than 0 to 1 but all three of these spaces have unit area.

The axes are not labeled in these three figures because they would have obscured the nodes shown in Figure 94, and so for consistency the labels are omitted in Figure 25 and Figure 63 as well.

A unit circle has a border length or perimeter of $2\,\pi\,a = 2\,\sqrt{\pi} \approx 3.5449$, whereas a flat unit square has a border length of 4. As a result, a flat unit square has (very roughly and depending on the range of the nodes) about 12% more border nodes which have fewer near neighbors on average than the non-border nodes.

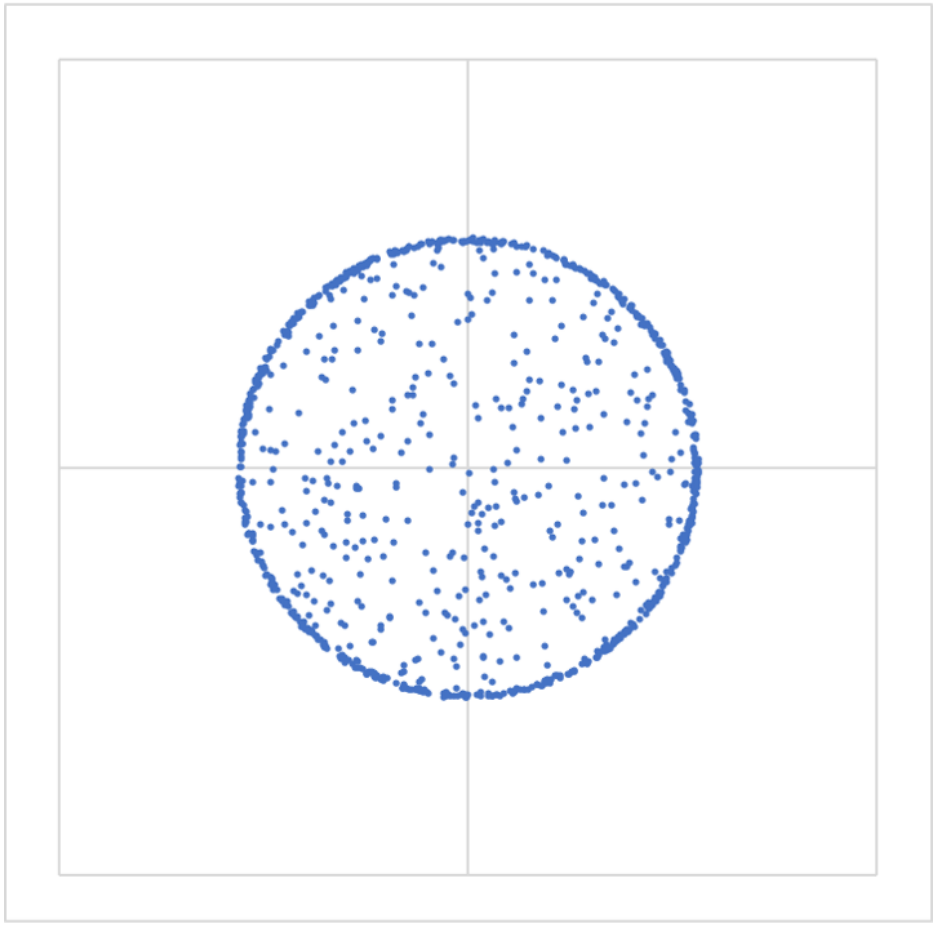

Figure 94: Position where the last node was added for full connectivity of each of 1,000 simulated geometries for RGGs in a unit circle with range 0.0100 and n-for-given-r growth. Border effects cause the positions of the last nodes added to tend to cluster near the border. Radius of circle is $a = 1/\sqrt{\pi} \approx 0.5642$ and axes of plot span -1 to 1.

4.4.1 Unit Circle: Full Connectivity for Mid-r Set

Figure 95 shows the trend lines for LRMs for the number of nodes required to first achieve full connectivity of RGGs in a unit circle with n-for-given-r growth for the mid-r set. Trend lines are shown for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r.

The trend lines are plotted on log-log scales for the range and number of nodes. As was the case for RGGs in a flat unit square, the close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs.

The LRMs with the values of m and b replaced in the power law expression (21) are as follows. For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 3.7526\, r^{-2.1039} \tag{73}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.23% and max error of the value of the predicted n is 3.30%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.8410\, r^{-2.1725} \tag{74}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.76% and max error of the value of the predicted n is 1.90%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.0098\, r^{-2.2449} \tag{75}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.64% and max error of the value of the predicted n is 4.41%.

Figure 96 gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity in a flat unit square and unit circle; note that the scale for the range is linear rather than log. Trend lines and the LRMs are shown in Figure 96 and, although the R^2 values indicate very poor fits, all of the ratios lie in a narrow band between roughly 0.98 and 1.18.

The key observation from Figure 96 is that the number of nodes required to first achieve full connectivity is generally only slightly greater in a flat unit square than in a unit circle for the largest ranges in the mid-r set and that ratio appears to decrease roughly linearly with the range. This may be because, for a given number of nodes, a flat unit square generally has more border nodes by a small percentage than does a unit circle as observed in Section 4.4.

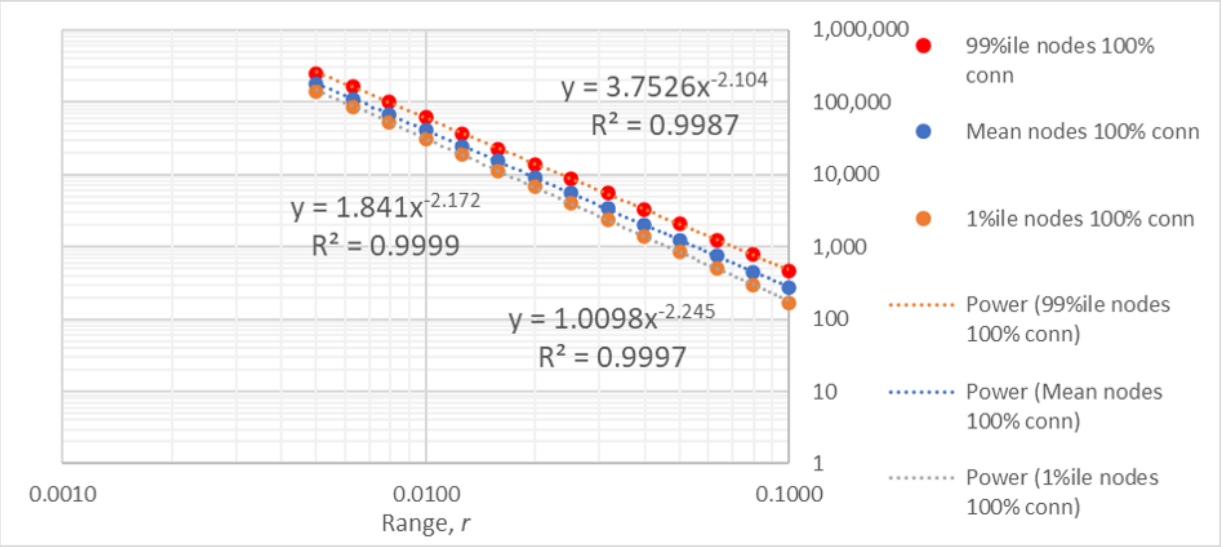

Figure 95: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a unit circle with n-for-given-r growth for the mid-r set.

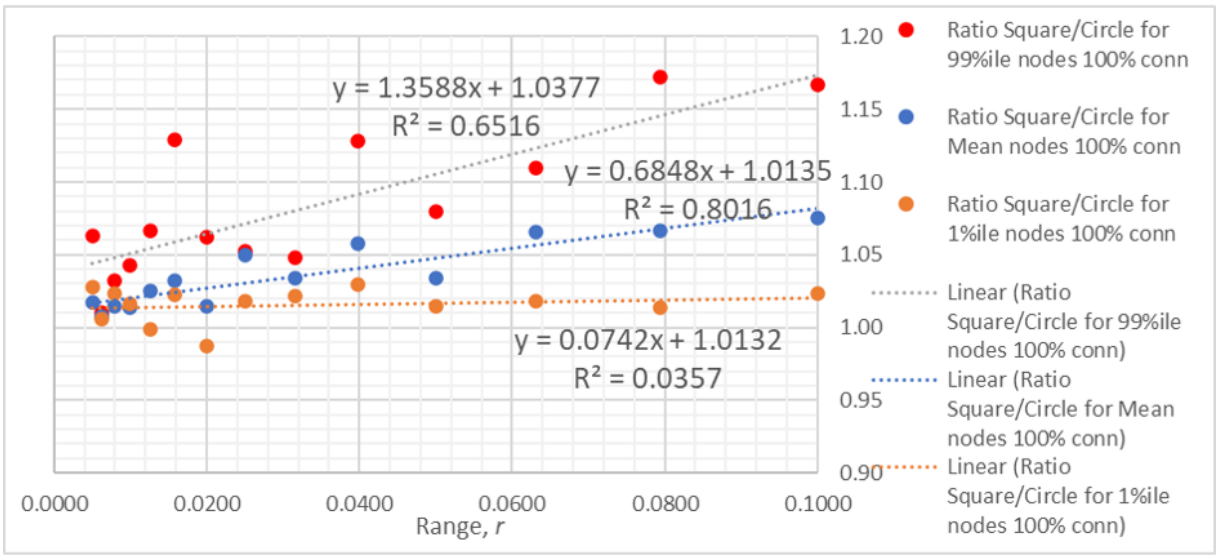

Figure 96: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square and a unit circle with n-for-given-r growth. Ratios are shown for 99%ile, mean, and 1%ile statistical metrics.

4.4.2 Unit Circle: 99% Connectivity for Mid-r Set

Demonstrating less-than-full connectivity, Figure 97 shows statistical metrics for the number of nodes required to first achieve at least 99% connectivity (the condition where $S \geq 0.99$) for RGGs in a unit circle with n-for-given-r growth using the mid-r set.

Figure 97 shows that the number of nodes required to first achieve 99% connectivity closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity with n-for-given-r growth, (21) becomes

$$n = 5.1101\, r^{-1.8168} \tag{76}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 3.19% and max error of the value of the predicted n is -7.94%.

For the mean number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 2.5799\, r^{-1.9496} \tag{77}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.62% and max error of the value of the predicted n is -1.36%.

For the 1%ile number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 1.4902\, r^{-2.0562} \tag{78}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.81% and max error of the value of the predicted n is 6.00%.

Figure 98 gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity in a flat unit square and unit circle; note that the scale for the range is linear rather than log. As was done for full connectivity, the trend lines and the LRMs are shown in Figure 98 and again the R^2 values indicate very poor fits but even so the ratios lie in a narrow band between roughly 0.99 and 1.11.

The key observation from Figure 98 is that the number of nodes required to first achieve 99% connectivity is generally only slightly greater in a flat unit square than in a unit circle for the largest ranges in the mid-r set and that ratio appears to decrease roughly linearly with the range. This may be because, for a given number of nodes, a flat unit square generally has more border nodes by a small percentage than does a unit circle as observed in Section 4.4.

This observation is in line with that made in Section 3.2 in regard to Figure 16 and Figure 17 which was there no visible difference in the emergence of the GC between RGGs in a flat unit square or a unit circle.

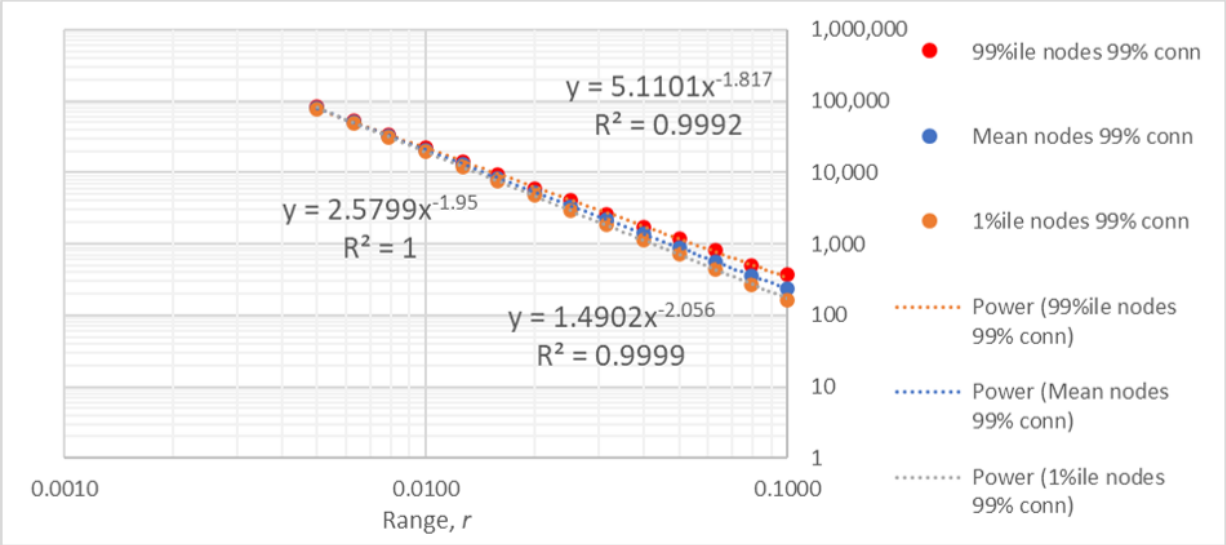

Figure 97: LRMs for 99%ile, mean, and 1%ile number of nodes for at least 99% connectivity of RGGs in a unit circle with n-for-given-r growth.

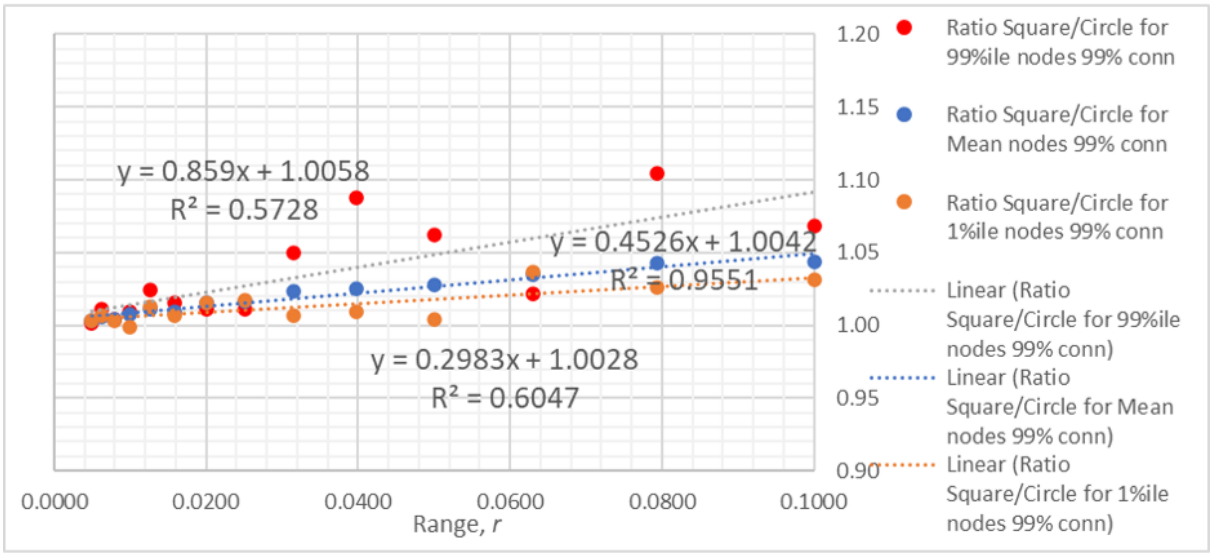

Figure 98: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for 99% connectivity of RGGs in a flat unit square and unit circle. Ratios are shown for 99%ile, mean, and 1%ile statistical metrics.

4.4.3 Unit Circle: No Isolated Nodes for Mid-r Set

Figure 99 shows that the number of nodes required to first achieve no isolated nodes in RGGs with n-for-given-r growth closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r in the mid-r set. The LRMs corresponding to the equations for the trend lines are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 3.2281\, r^{-2.1323} \tag{79}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 3.00% and max error of the value of the predicted n is 6.74%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.4150\, r^{-2.2192} \tag{80}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.00% and max error of the value of the predicted n is 5.66%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.5926\, r^{-2.3479} \tag{81}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 5.23% and max error of the value of the predicted n is 14.80%.

The R^2 values for the LRMs for a unit circle are virtually unity. Across the three connectivity conditions (full connectivity, 99% connectivity, and no isolated nodes) and the three statistical metrics (99%ile, mean, and 1%ile) reported here, MAPE for a unit circle varies from 0.62% to 5.23% and the absolute values of max errors vary from 1.36% to 14.80%.

As was observed in Section 4.2 and its subsections for RGGs in a flat unit square as well as in Section 4.3 and its subsections for RGGs in a toroidal unit square, the number of nodes required to first achieve full connectivity of RGGs in a unit circle shown in Figure 95 and their LRMs (73), (74), and (75) are substantially different from the number of nodes required for the disappearance of isolated nodes shown in Figure 99 and their LRMs (79), (80), and (81). For RGGs in a unit circle with n-for-given-r growth, full connectivity does not inevitably occur when the last isolated node is absorbed.

As discussed in Section 4.2.1.4 and Section 4.3.1.3 for RGGs in flat and toroidal unit squares with n-for-given-r growth, this observation raises the question as to how many additional nodes must be added to an RGG in a unit circle beyond the point where the last isolated node disappears. Figure 100 shows the 99%ile, mean, and 1%ile number of nodes required to first achieve full connectivity since the last number of nodes where there was at least one isolated node. The 1%ile value is again always 1 but the 99%ile and mean number of additional nodes required to be added grow from tens and hundreds for the largest ranges in the mid-r set to thousands and hundreds of thousands for the smallest ranges simulated in these experiments.

Figure 100 also shows that power law LRMs closely fit the 99%ile and mean number of additional nodes required to be added beyond the disappearance of the last isolated node to first achieve full connectivity.

Figure 101 gives the ratios of the 99%ile, mean, and 1%ile number of nodes required to eliminate isolated nodes in a flat unit square and unit circle; note that the scale for the range is linear rather than log. The trend lines and the LRMs are shown in Figure 101 and again the R^2 values indicate very poor fits but even so the ratios lie in a narrow band between roughly 0.96 and 1.20.

The key observation from Figure 101 is that the number of nodes required to eliminate isolated nodes is generally only slightly greater in a flat unit square than in a unit circle for the largest ranges in the mid-r set and that ratio appears to decrease roughly linearly with the range. This may be because, for a given number of nodes, a flat unit square generally has more border nodes by a small percentage than does a unit circle as observed in Section 4.4.

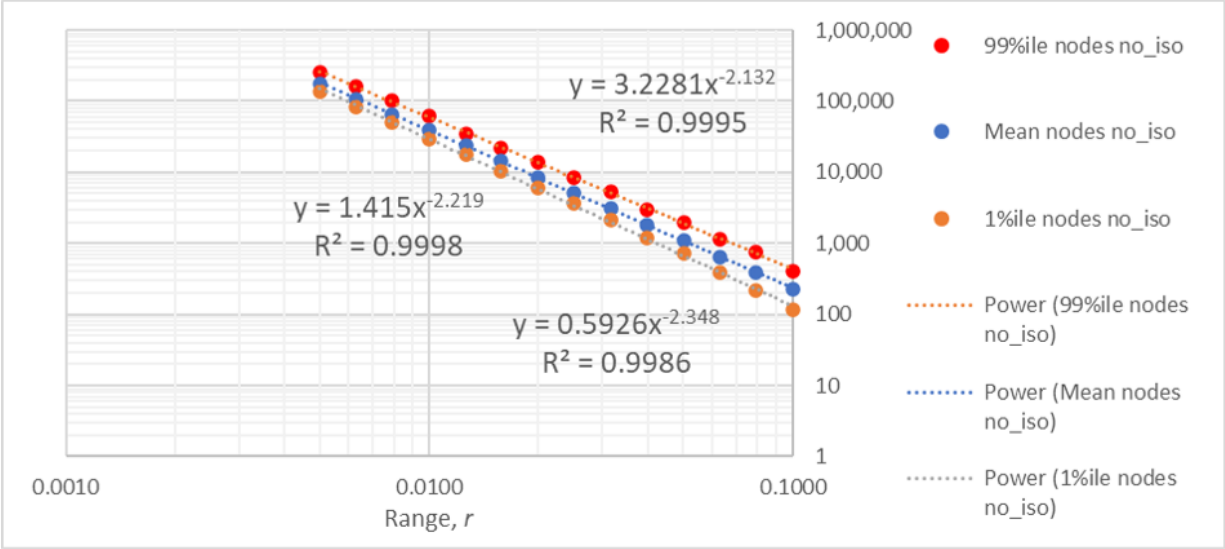

Figure 99: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a unit circle with n-for-given-r growth.

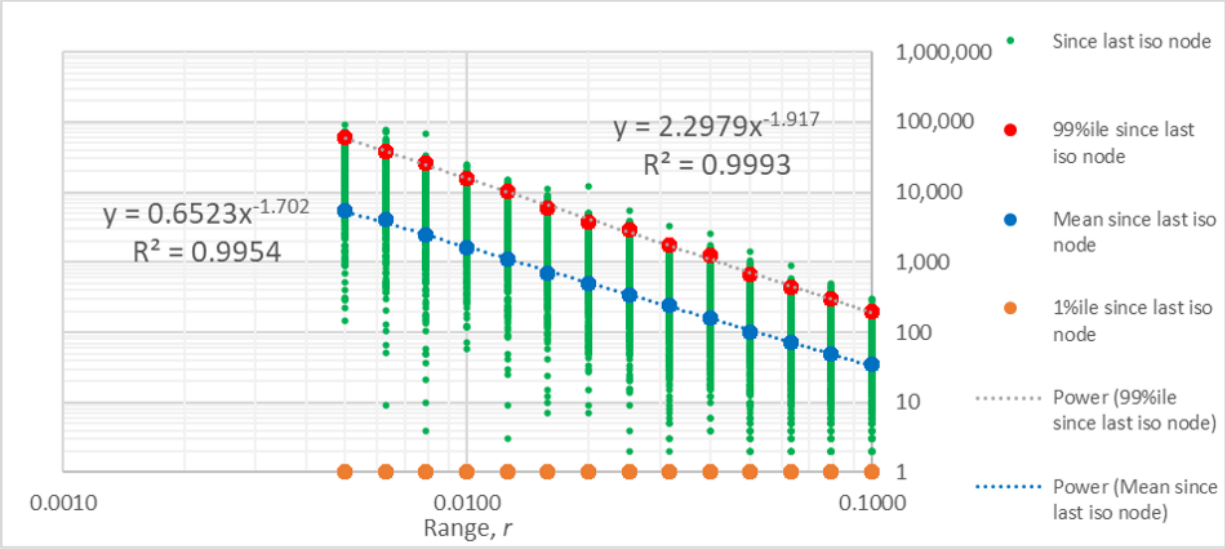

Figure 100: LRMs for 99%ile and mean number of nodes for full connectivity of RGGs in a unit circle with n-for-given-r growth since the last number of nodes where there was at least one isolated node.

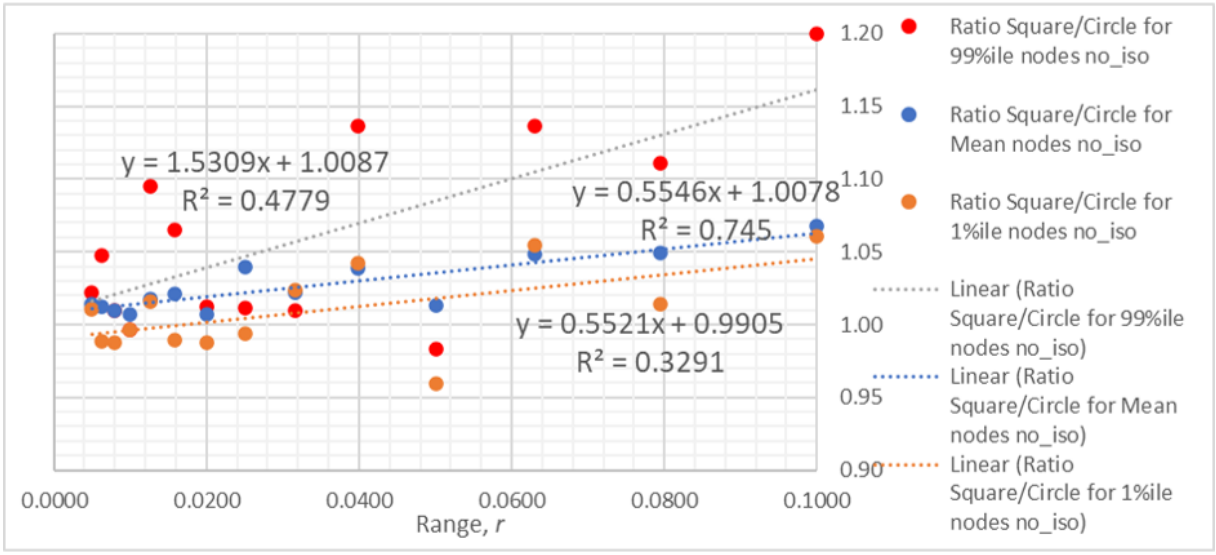

Figure 101: LRMs for ratios of 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a flat unit square and unit circle. Ratios are shown for 99%ile, mean, and 1%ile statistical metrics.

4.4.4 Unit Circle: Number of Neighbors Needed for Full Connectivity for Mid-r Set

Figure 102 shows a linear-log plot with LRMs for the 99%ile, mean, and 1%ile values for the mean degree required to first achieve full connectivity. The regression models for the statistical metrics are linear functions of the mean degree and the log of the range and thus they are modeled as log laws rather than the log-log power law functions that were shown to work well for the statistical metrics for the numbers of nodes.

Figure 103 is a scatter plot of the predicted vs. observed mean degree required to first achieve full connectivity for all of the simulated geometries using the ranges in the mid-r set. For each geometry the predicted mean degree is the value of μ given by (9) for each given range r and the number of nodes n required to first achieve full connectivity. The red line represents the value where the observed mean degree would have equaled the predicted mean degree for a simulation. As was the case for RGGs in a flat unit square as shown in Section 4.2.1.5, the observed values are slightly smaller than the predicted values due to the nodes near the border having fewer near neighbors than the other nodes in the space.

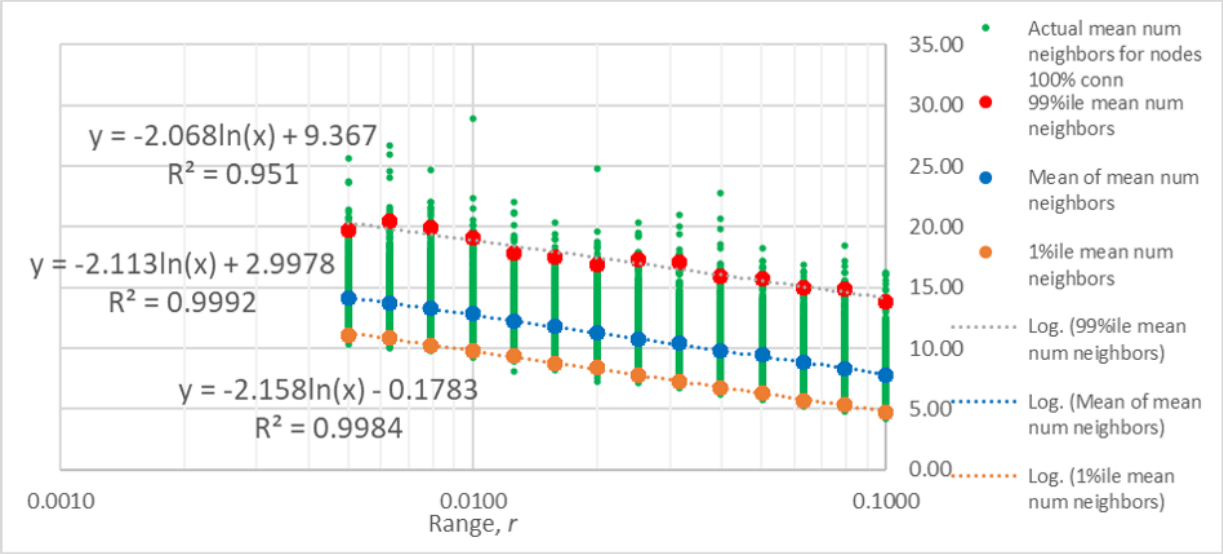

Figure 102: LRMs for 99%ile, mean, and 1%ile values for mean degree for full connectivity in a unit circle with n-for-given-r growth.

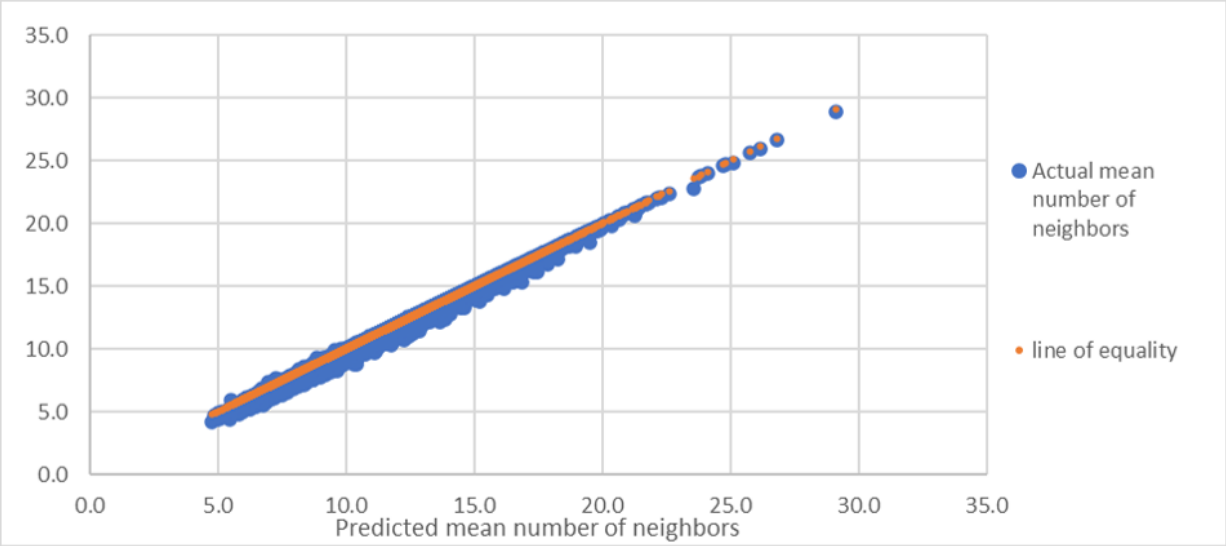

Figure 103: Observed and predicted mean degree for full connectivity in a unit circle with n-for-given-r growth.

4.4.5 Unit Circle: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Mid-r Set

It was discussed in Section 3.5 that Penrose's First Claim that an RGG becomes fully connected when the last isolated node is absorbed applies specifically to the situation of RGGs with "n large" in a toroidal space with r-for-given-n growth; his claim is tested in this section for its applicability to RGGs in a flat unit circle with n-for-given-r growth using the mid-r set. The following figures show results that test the accuracy of Penrose's Theorem for this situation.

Figure 104 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected. If the First Claim of Penrose's Theorem were correct for RGGs in a unit circle with n-for-given-r growth, then the fraction would be 1 or close to it, particularly for the smaller values of r which imply that n is large. Interpreting this fraction as a probability, for RGGs in a unit circle with n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.318 for $r = 0.1000$ and increases to about 0.694 for $r = 0.0050$.

This probability increases as r decreases (and consequently the number of nodes increases) but Figure 104 shows that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for node-by-node growth of an RGG in a unit circle even for the smallest range in the mid-r set.

Figure 105 relates the number of nodes needed for full connectivity to those required for no isolated nodes. For each randomly generated geometry for each value of r the ratio of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of r is plotted on the graph. Even for small ranges, more nodes are required to first achieve full connectivity after the disappearance of isolated nodes, which is consistent with the discussion of Figure 104. The ratios may asymptotically approach 1 for even smaller ranges but the ratio was observed to be 1.039 for the smallest range in the mid-r set.

Figure 106 shows the error of the LRM (73) given in Section 4.4.1 for the 99%ile number of nodes required to first achieve full connectivity with respect to the observed number of nodes obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 4.4.1 that MAPE for (73) is 2.23% and max error of the predicted value of n is 3.30%.

Figure 106 also shows the same error calculations for the number of nodes required to achieve full connectivity given by Penrose's Theorem with $C = 0.99$. These error values are considerably greater than those for the LRM.

Figure 107 shows the same error statistics as in Figure 106 but now for the 1%ile number of nodes required to first achieve full connectivity, which is the threshold used here for RGGs to be disconnected. Figure 107 shows the error percentages for the LRM (75) and Penrose's predicted number of nodes for full connectivity with $C = 0.01$. Again, the LRM provides close estimates while the predicted number of nodes from Penrose's expressions has far greater errors.

Figure 108 shows values of c_P for the 99%ile, mean, and 1%ile values of n, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of r, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 95.

The maximum of the c_P observed for any range was 17.6590. As was observed for RGGs in a flat unit square, the visible trend in Figure 108 is that all three of the statistical metrics plotted for c_P seem to decrease slightly as the range decreases and thus the number of nodes increases. The smallest values of $c_P = 7.3536$ for the 99%ile values of n and $c_P = $-0.7077 for the 1%ile values of n are higher than Penrose's Theorem's values of $c_P = 4.6001$ required for $C = 0.99$ and $c_P = -1.5272$ for $C = 0.01$.

It noted again that Penrose's Theorem was based on the r-for-given-n growth model and does not necessarily apply to the n-for-given-r growth model used in this section. These comparisons are intended only to assist in determining when they do apply.

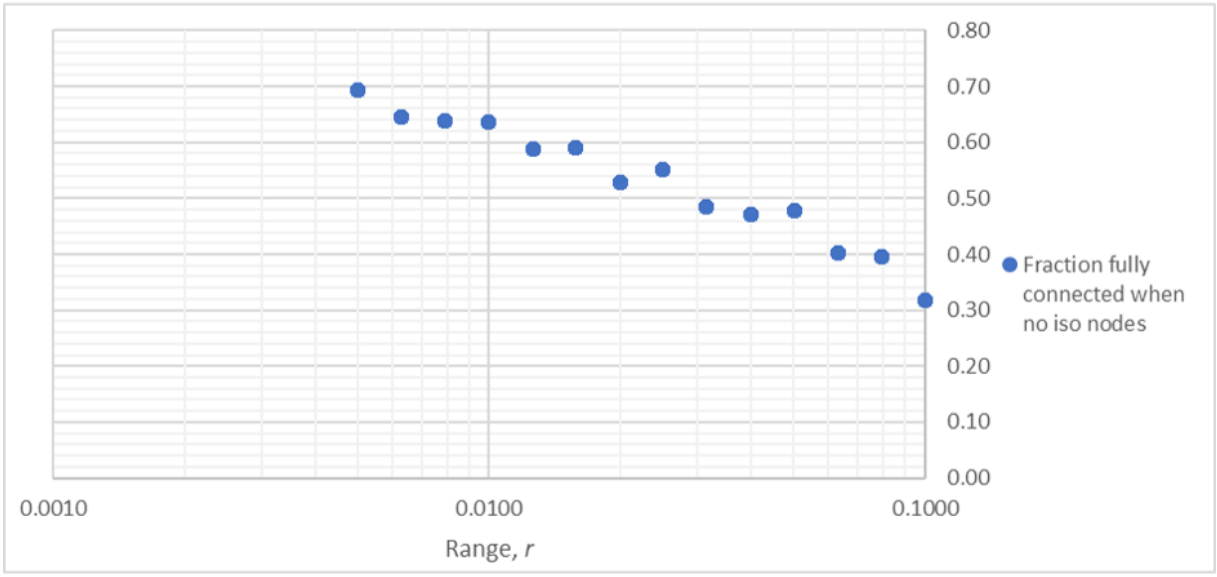

Figure 104: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a unit circle to become fully connected with n-for-given-r growth.

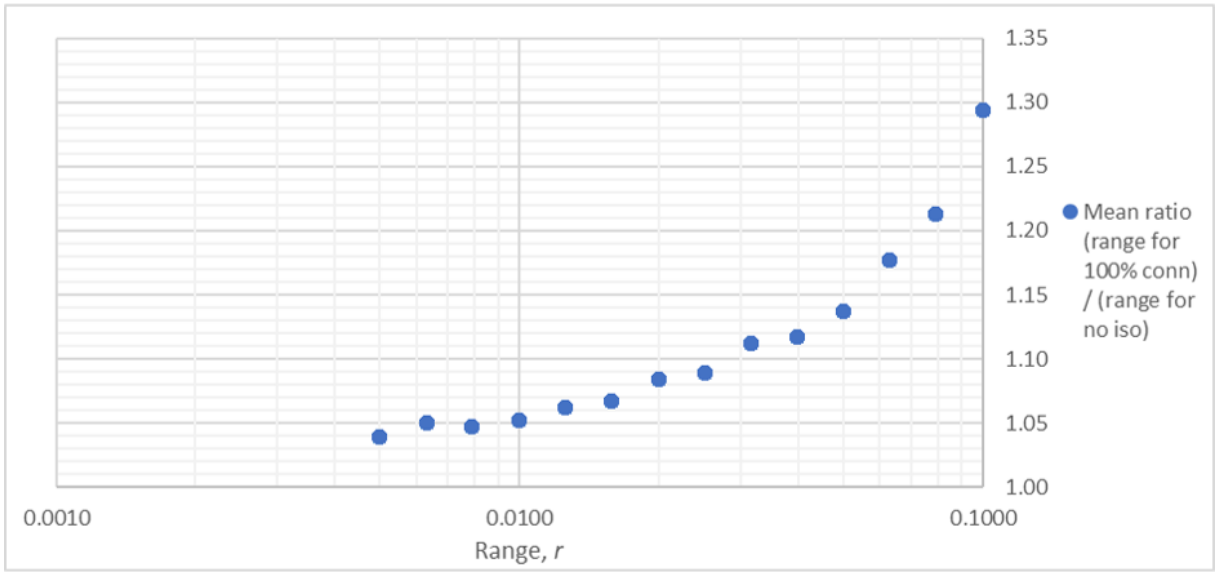

Figure 105: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in RGGs in a unit circle with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

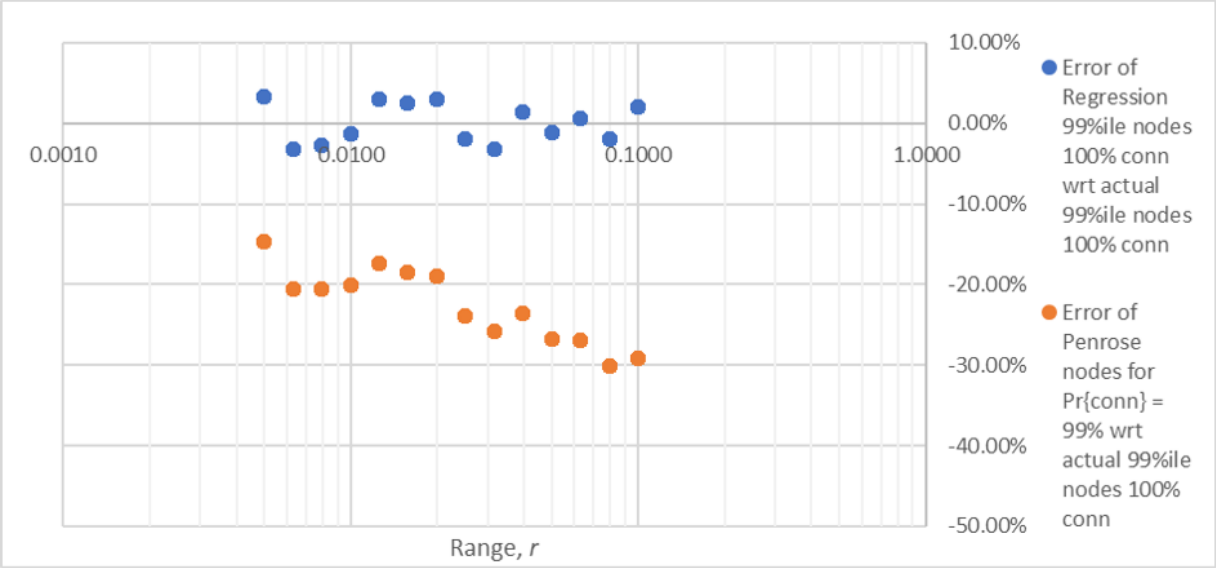

Figure 106: Relative error of LRM (73) for 99%ile number of nodes for full connectivity of RGGs in a unit circle with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.99$.

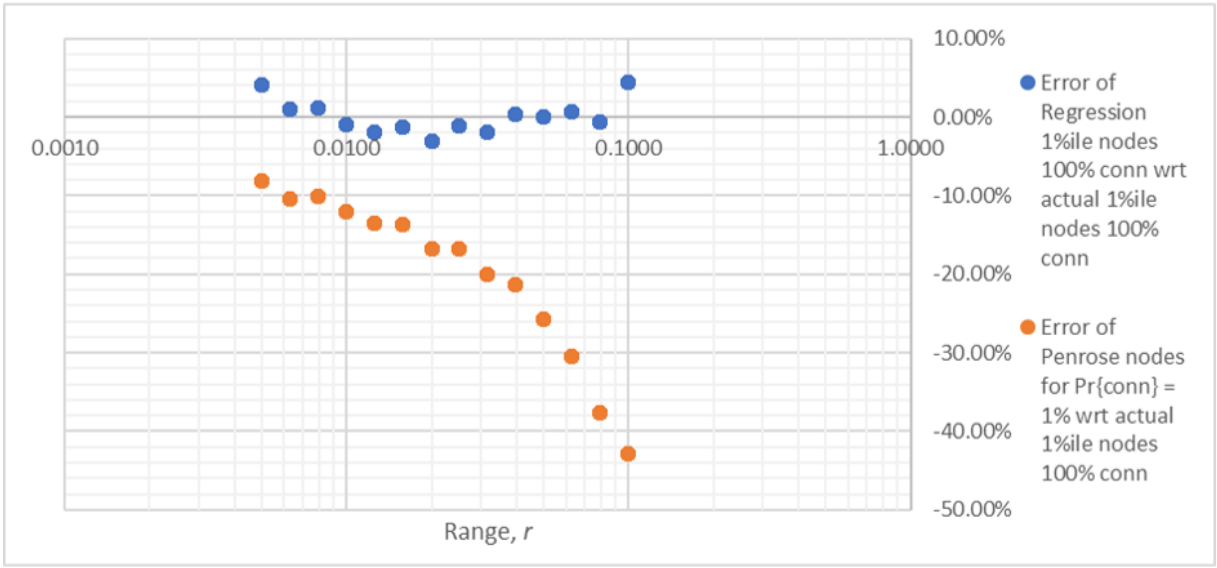

Figure 107: Relative error of LRM (75) for 1%ile number of nodes for full connectivity of RGGs in a unit circle with n-for-given-r growth. Relative error is shown for predicted n given by Penrose's Theorem with $C = 0.01$.

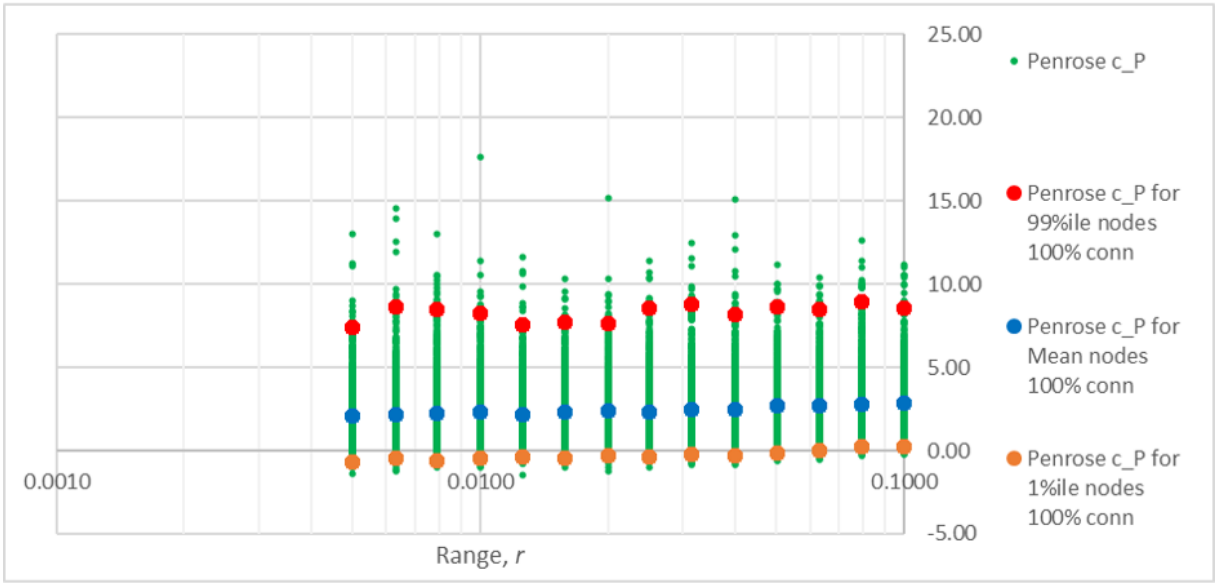

Figure 108: Values of Penrose's c_P for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a unit circle with n-for-given-r growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

5 RGGs that Grow by Increasing the Range of a Fixed Number of Nodes: The r-for-given-n Growth Model

This section presents results for the r-for-given-n growth model. Much of the introductory material for this section is common to that presented in Section 4 and its subsections. Material specific to the r-for-given-n growth model is described in Section 5.1.

In this model of RGG growth, a fixed set of nodes is scattered uniformly at random positions in a space and the minimum range r is found for which the RGG reaches the desired connectivity condition.

Expressions are obtained for the 99%ile and mean range required to achieve either the condition of full connectivity ($S = 1$) or the condition of having no isolated nodes. Only the two-dimensional flat and toroidal unit squares without obstructions were considered in this section.

5.1 Simulations and Data Analysis

As was done for the work described in Section 4.1.2 for the n-for-given-r growth model, the empirically based approach used for the r-for-given-n growth model is to simulate many randomly generated geometries for selected configurations and use linear regression to obtain the desired predictive models.

Three sets of numbers of nodes were used for these simulations: a set of small numbers of nodes, a set of large numbers of nodes, and the union of those two sets. Because the LRMs vary slightly based on the sets over which they were derived, relative errors are calculated and presented for each set of numbers of nodes.

Unless otherwise specified, each LRM is based on the simulation results for the first set: a set of numbers of nodes n incremented from 10 to 100,000. Starting from $n = 10$ the number of nodes was increased geometrically with the multiplier $10^{1/10}$ so there were ten sampled values of n per decade. The values of n are rounded up to the nearest integer (that is, the ceiling of the calculated value).

Thus, the 41 numbers of nodes in the *small numbers of nodes set (small-n set)* are $n = 10$, 13, 16, 20, 26, 32, 40, 51, 64, 80, 100, 126, 159, 200, 252, 317, 399, 502, 631, 795, 1,000, 1,259, 1,585, 1,996, 2,512, 3,163, 3,982, 5,012, 6,310, 7,944, 10,000, 12,590, 15,849, 19,953, 25,119, 31,623, 39,811, 50,119, 63,096, 79,433, and 100,000.

Due to the fact that much of the theoretical interest regarding the r-for-given-n growth model is for large numbers of nodes, the second set of numbers of nodes was used where n is incremented from 100,000 to 10,000,000. Starting from $n = 100{,}000$ the number of nodes was increased geometrically with the multiplier $10^{1/10}$ so there were ten sampled values of n per decade. The values of n are rounded up to the nearest integer.

Thus, the 21 numbers of nodes in the *large numbers of nodes set (large-n set)* are $n = 100{,}000$, 125,893, 158,490, 199,527, 251,189, 316,228, 398,108, 501,188, 630,958, 794,329, 1,000,000, 1,258,926, 1,584,894, 1,995,263, 2,511,887, 3,162,278, 3,981,072, 5,011,873, 6,309,574, 7,943,283 and 10,000,000.

In some cases, subsets or combinations of these sets of numbers of nodes are used. Notably, the third set, called the *small and large numbers of nodes set (small+large-n set)*, is the union of the small-n and large-n sets.

The simulations performed for the small-n set are separate from those performed for the small+large-n set. The results for the large-n set are taken from the simulations for the small+large-n set. As noted in Section 4.1.2 for RGGs with n-for-given-r growth, because separate sets of simulations were performed for some numbers of nodes, statistical metrics and estimates of probabilities may vary slightly throughout the sections that follow.

As described in Section 4.1.4 for the n-for-given-r growth model, the sample size for each configuration's value of n was $N = 1{,}000$ simulated geometries with r-for-given-n growth.

A binary search was used to closely estimate the smallest range that achieved the desired connectivity conditions. The simulated geometries for each value of r used different random seeds but the same sequence of seeds were used for the flat and toroidal RGGs for fully connected and no isolated nodes so that comparisons could be made on the basis of the same randomly generated nodes in each case. Node generation was not padded for the simulations of r-for-given-n growth as was done for n-for-given-r growth as described in Section 4.1.2.

LRMs and the analysis of their goodness-of-fit and predictive errors for the results of the simulations of RGGs with r-for-given-n growth was performed as described in Section 4.1.5.

5.2 Flat Unit Square with r-for-given-n Growth

This section considers RGGs consisting of nodes scattered uniformly at random positions in a two-dimensional square space of area $A = 1$ with no wraparound at the border. This is a flat space so there are border effects. The RGGs evolve according to the r-for-given-n growth model.

Results are given separately for the small-n, large-n, and small+large-n sets outlined in Section 5.1. The LRMS vary slightly based on the sets of numbers of nodes over which they were obtained. Although an LRM is valid for only the set on which it is based, the relative errors of the small-n LRMs are compared to those of the small+large-n LRMS for that set of numbers of nodes.

For each simulated geometry for a given configuration, a fixed number of nodes n are placed in the space to determine the minimum range r required to achieve the condition of 100% connectivity or no isolated nodes, and the 99%ile and mean statistical metrics of those ranges are obtained for that configuration.

5.2.1 Flat Unit Square with r-for-given-n Growth for Small Numbers of Nodes

Section 5.2.1.1 and Section 5.2.1.2 present results for RGGs in flat unit squares with r-for-given-n growth using the small-n set consisting of the numbers of nodes from 10 up to 100,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.2.1.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Small-n Set

For RGGs in a flat unit square with r-for-given-n growth, using the small-n set, Figure 109 shows the data and trend lines for the 99%ile and mean range for full connectivity and Figure 110 shows the data and trend lines for the 99%ile and mean range required to achieve no isolated nodes. These results are based on the samples of size $N = 1{,}000$ simulated geometries per value of n.

The trend lines are plotted on log-log scales for the number of nodes and range. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs for the r-for-given-n growth model, as it was for n-for-given-r. The LRMs fit the values of the statistical metrics to equations of the form

$$\ln r = m \ln n + \ln b \tag{82}$$

which is expressed in power law form as

$$r = b\, n^m \quad . \tag{83}$$

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 109, (83) becomes

$$r = 2.1290\, n^{-0.4804} \tag{84}$$

with an R^2 value of 0.9997. For the simulated values of n, MAPE is 1.87% and max error of the value of the predicted r is 6.25%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 109,, the LRM is

$$r = 1.3312\, n^{-0.4584} \tag{85}$$

with an R^2 value of 0.9998. For the simulated values of n, MAPE is 1.29% and max error of the value of the predicted r is 7.85%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 110,, the LRM is

$$r = 2.0122\, n^{-0.4765} \tag{86}$$

with an R^2 value of 0.9996. For the simulated values of n, MAPE is 1.93% and max error of the value of the predicted r is 7.60%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 110,, the LRM is

$$r = 1.1202\, n^{-0.4432} \tag{87}$$

with an R^2 value of 0.9995. For the simulated values of n, MAPE is 2.31% and max error of the value of the predicted r is 9.43%.

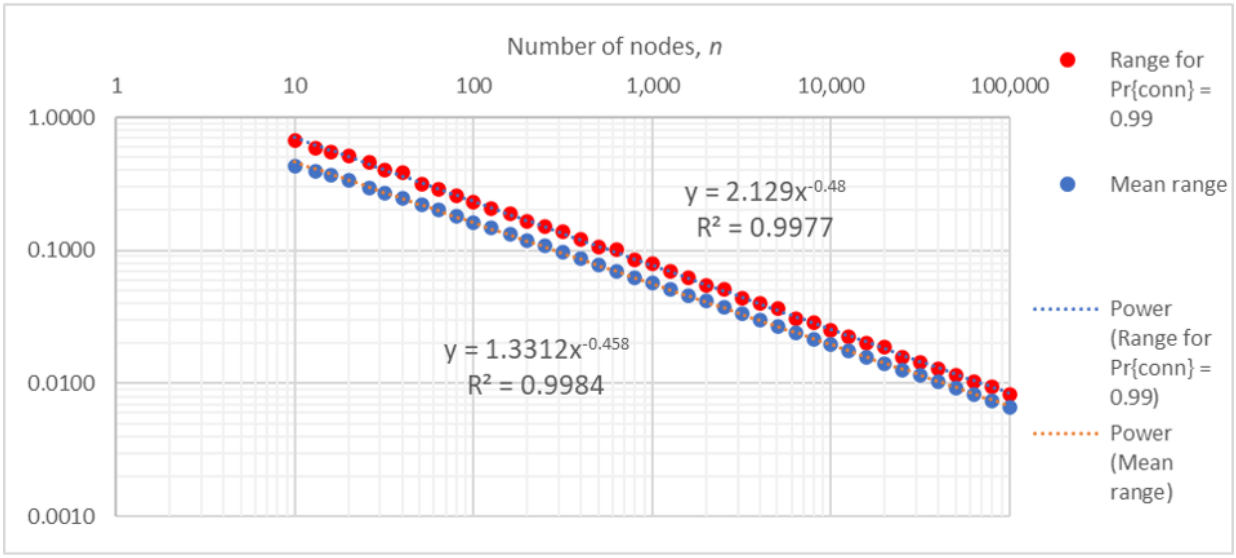

Figure 109: LRMs for 99%ile and mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth.

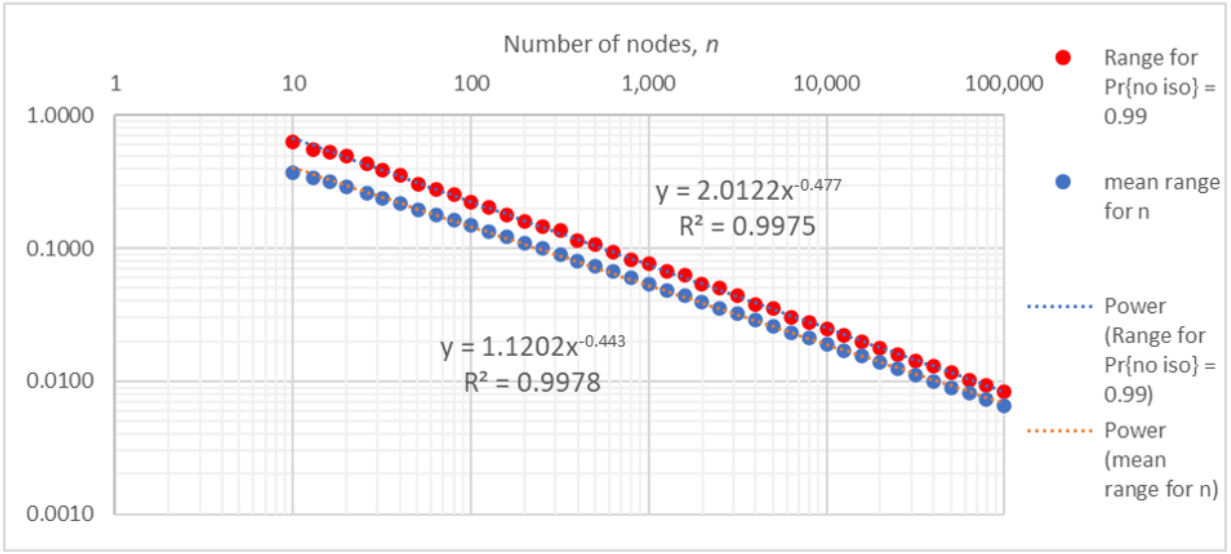

Figure 110: LRM for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth.

5.2.1.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-n Set

As discussed in Section 3.5, Penrose proposed that an RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem for RGGs in a flat unit square with r-for-given-n growth using the small-n set.

Figure 111 shows the error of the LRM (84) given in Section 5.2.1.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 5.2.1.1 that MAPE for (84) is 1.87% and max error of the predicted value of r is 6.25%.

Figure 111 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). These error values are considerably greater than those for the LRM.

Figure 112 shows the error of the LRM (86) given in Section 5.2.1.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.2.1.1 that MAPE for (86) is 1.93% and max error of the predicted value of r is 7.60%.

Figure 112 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). Even for the lesser condition of having no isolated nodes, these error values are considerably greater than those for the LRM.

As a reminder, Penrose's Theorem was based on the r-for-given-n growth model without border effects and thus does not necessarily apply to the flat space assumed in this section. These comparisons are intended only to assist in determining when they do apply.

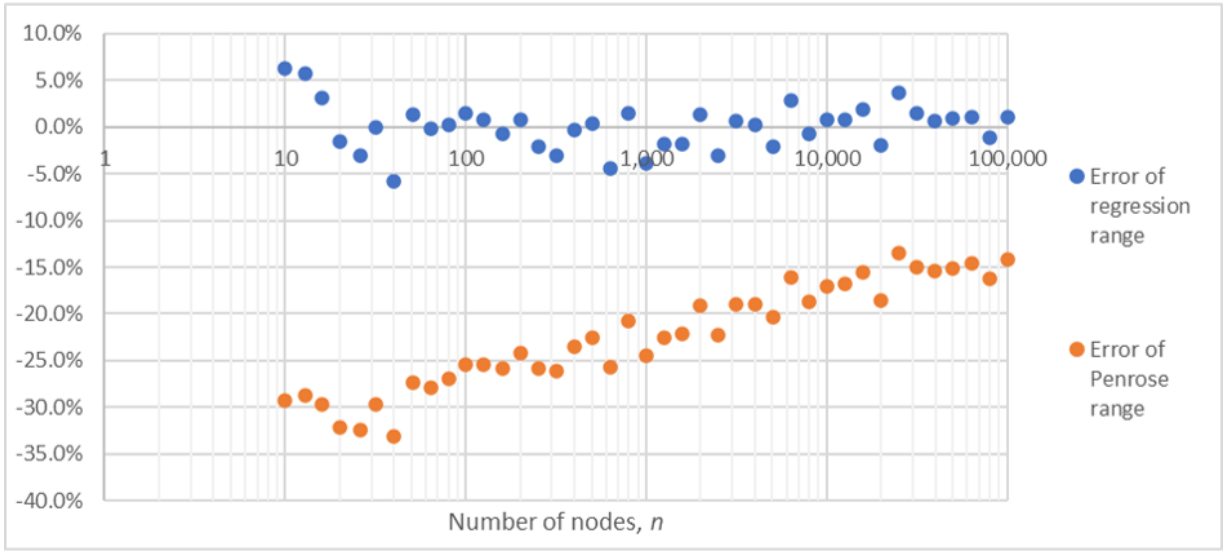

Figure 111: Relative error of LRM (84) for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

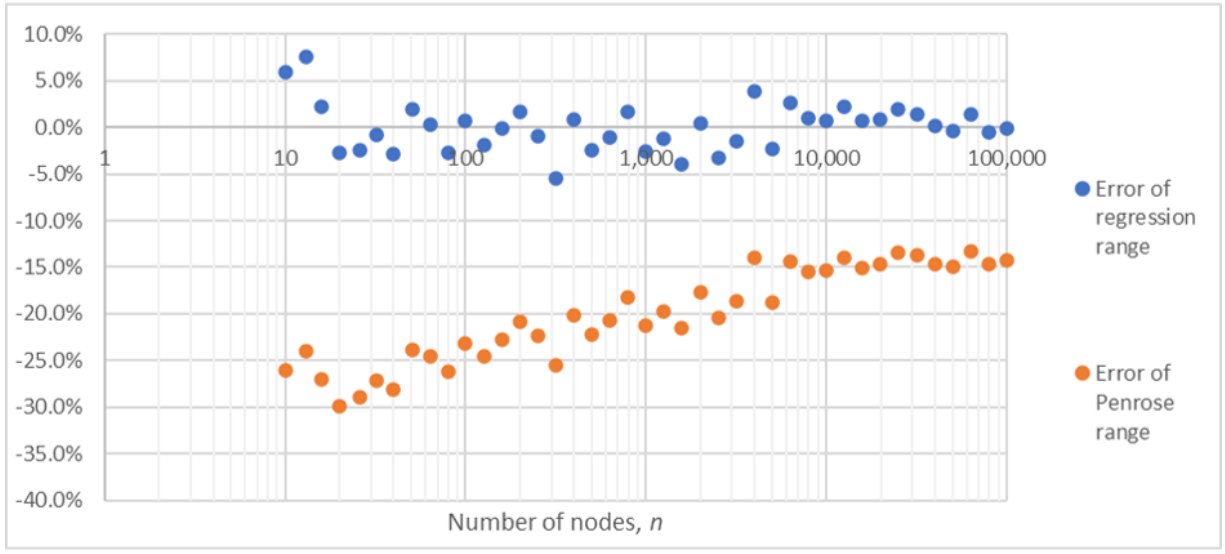

Figure 112: Relative error of LRM (86) for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

5.2.2 Flat Unit Square with r-for-given-n Growth for Large Numbers of Nodes

Section 5.2.2.1 and Section 5.2.2.2 present results for RGGs in flat unit squares with r-for-given-n growth using the large-n set consisting of the numbers of nodes from 100,000 up to 10,000,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.2.2.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Large-n Set

For RGGs in a flat unit square with r-for-given-n growth, using the large-n set, Figure 113 shows the data and trend lines for the 99%ile and mean range for full connectivity and Figure 114 shows the data and trend lines for the 99%ile and mean range required to achieve no isolated nodes. These results are based on the samples of size $N = 1{,}000$ simulated geometries per value of n.

The trend lines are plotted on log-log scales for the number of nodes and range. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs for the r-for-given-n growth model, as it was for n-for-given-r.

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 113, (83) becomes

$$r = 2.1963\, n^{-0.4845} \tag{88}$$

with an R^2 value of 0.9996. For the simulated values of n, MAPE is 1.18% and max error of the value of the predicted r is -2.32%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 113,, the LRM is

$$r = 1.5267\, n^{-0.4717} \tag{89}$$

with an R^2 value of 1.0000. For the simulated values of n, MAPE is 0.22% and max error of the value of the predicted r is -0.72%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 114,, the LRM is

$$r = 2.1196\, n^{-0.4829} \tag{90}$$

with an R^2 value of 0.9997. For the simulated values of n, MAPE is 0.93% and max error of the value of the predicted r is -2.99%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 114,, the LRM is

$$r = 1.4573\, n^{-0.4693} \tag{91}$$

with an R^2 value of 1.0000. For the simulated values of n, MAPE is 0.25% and max error of the value of the predicted r is -0.88%.

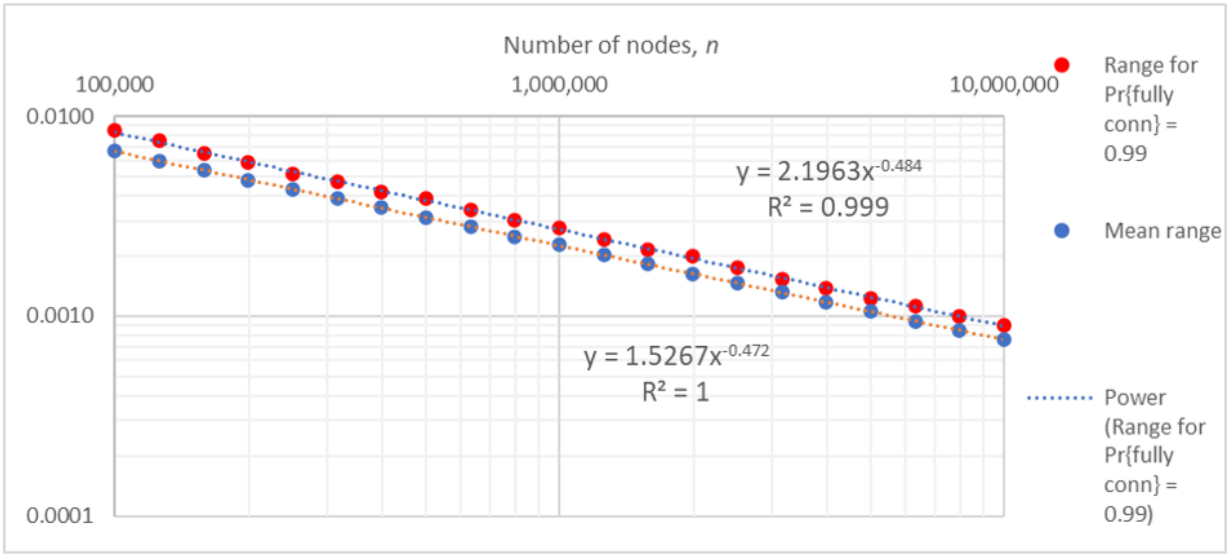

Figure 113: LRMs for 99%ile and mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth.

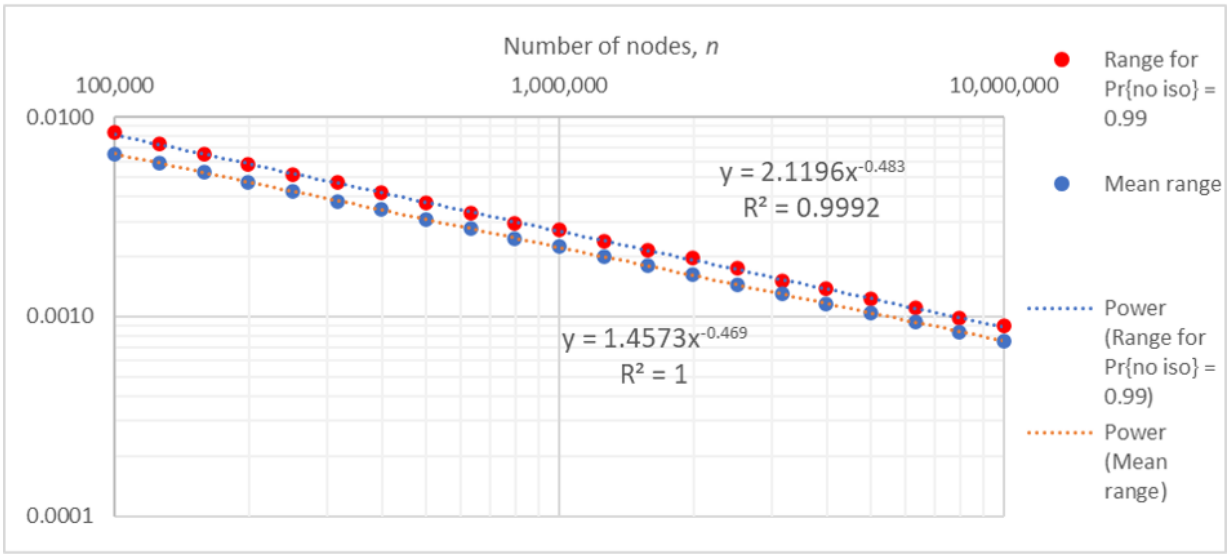

Figure 114: LRM for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth.

5.2.2.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Large-n Set

As discussed in Section 3.5, Penrose proposed that an RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem for RGGs in a flat unit square with r-for-given-n growth using the large-n set.

Figure 115 shows the error of the LRM (88) given in Section 5.2.2.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 5.2.2.1 that MAPE for (88) is 1.18% and max error of the predicted value of r is -2.32%.

Figure 115 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). These error values are considerably greater than those for the LRM.

Figure 116 shows the error of the LRM (90) given in Section 5.2.2.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.2.2.1 that MAPE for (90) is 0.93% and max error of the predicted value of r is -2.99%.

Figure 116 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). Even for the lesser condition of having no isolated nodes, these error values are considerably greater than those for the LRM.

As a reminder, Penrose's Theorem was based on the r-for-given-n growth model without border effects and thus does not necessarily apply to the flat space assumed in this section. These comparisons are intended to assist in determining when they do apply.

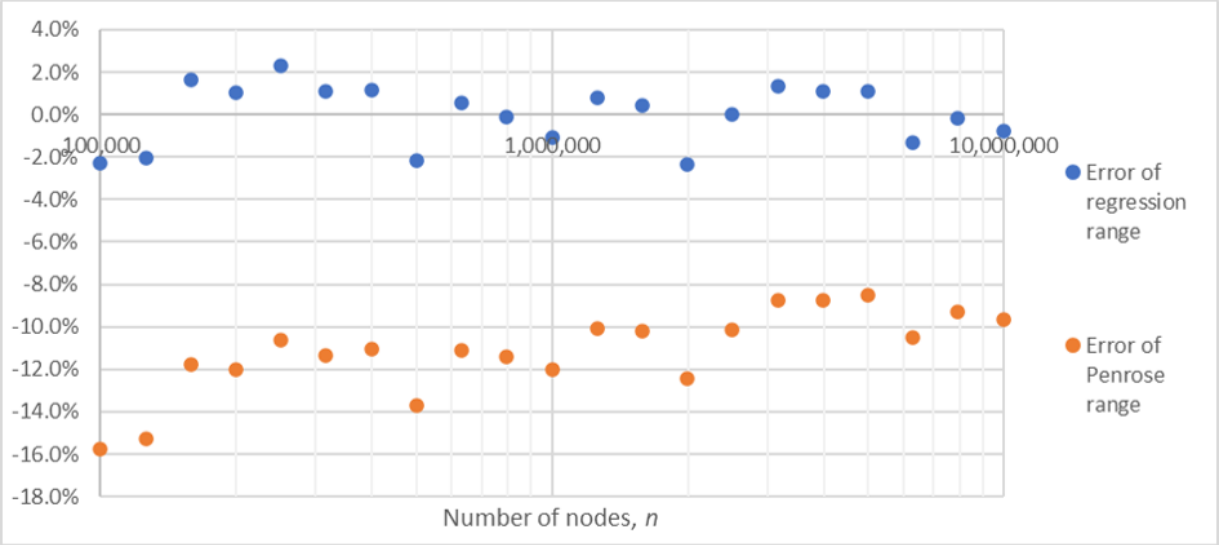

Figure 115: Relative error of LRM (88) for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

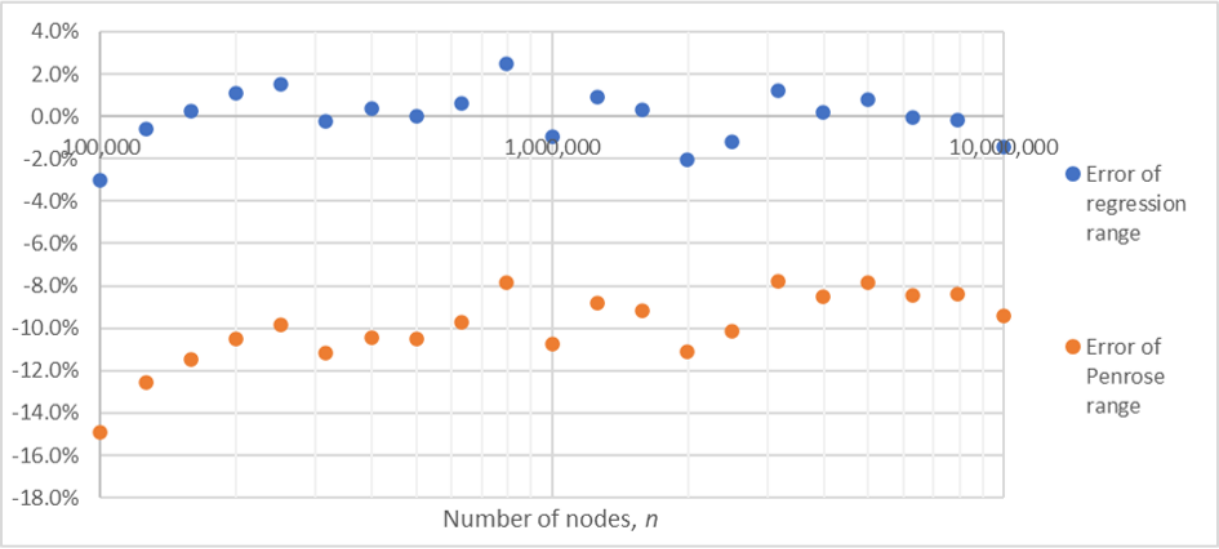

Figure 116: Relative error of LRM (90) for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

5.2.3 Flat Unit Square with r-for-given-n Growth for Small and Large Numbers of Nodes

Section 5.2.3.1 and Section 5.2.3.2 present results for RGGs in flat unit squares with r-for-given-n growth using the small+large-n set consisting of the numbers of nodes from 10 up to 10,000,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.2.3.1 Flat Unit Square: Full Connectivity and No Isolated Nodes for Small+Large-n Set

For RGGs in a flat unit square with r-for-given-n growth, using the small+large-n set, Figure 117 shows the data and trend lines for the 99%ile and mean range for full connectivity and Figure 118 shows the data and trend lines for the 99%ile and mean range required to achieve no isolated nodes. These results are based on the samples of size $N = 1{,}000$ simulated geometries per value of n.

The trend lines are plotted on log-log scales for the number of nodes and range. The close fits to straight lines when plotted on these transformed axes show that the power law is a good approximation for the LRMs for the r-for-given-n growth model, as it was for n-for-given-r.

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 117, (83) becomes

$$r = 2.1671\, n^{-0.4832} \tag{92}$$

with an R^2 value of 0.9999. For the simulated values of n, MAPE is 1.85% and max error of the value of the predicted r is 6.89%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 117,, the LRM is

$$r = 1.3700\, n^{-0.4632} \tag{93}$$

with an R^2 value of 0.9999. For the simulated values of n, MAPE is 1.61% and max error of the value of the predicted r is 10.09%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 118,, the LRM is

$$r = 2.0778\, n^{-0.4812} \tag{94}$$

with an R^2 value of 0.9999. For the simulated values of n, MAPE is 1.79% and max error of the value of the predicted r is 6.38%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 118,, the LRM is

$$r = 1.1934\, n^{-0.4536} \tag{95}$$

with an R^2 value of 0.9996. For the simulated values of n, MAPE is 3.08% and max error of the value of the predicted r is 12.87%.

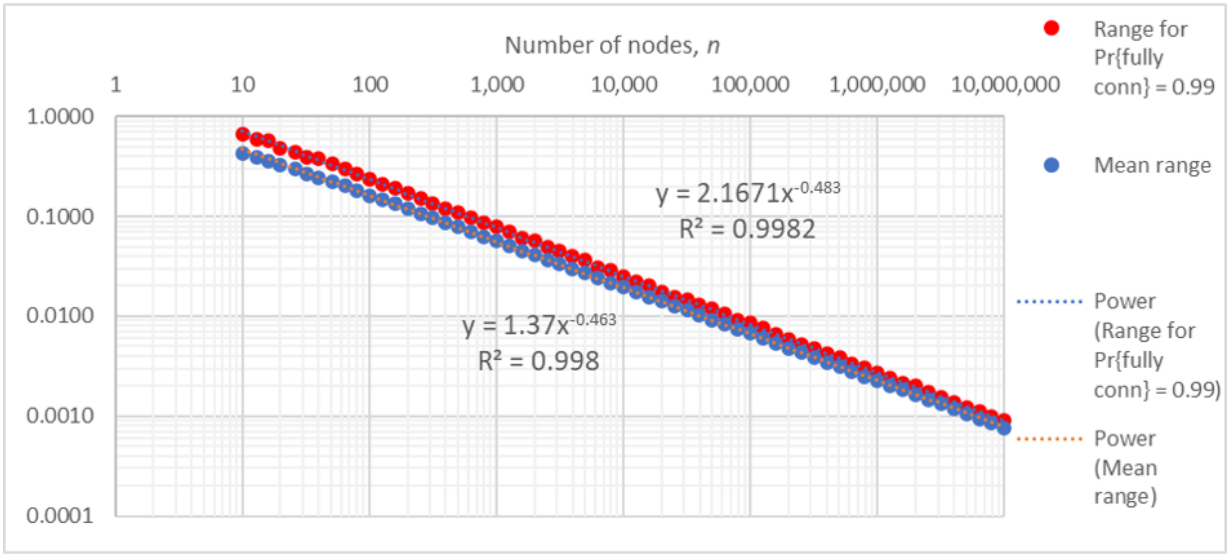

Figure 117: LRMs for 99%ile and mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth.

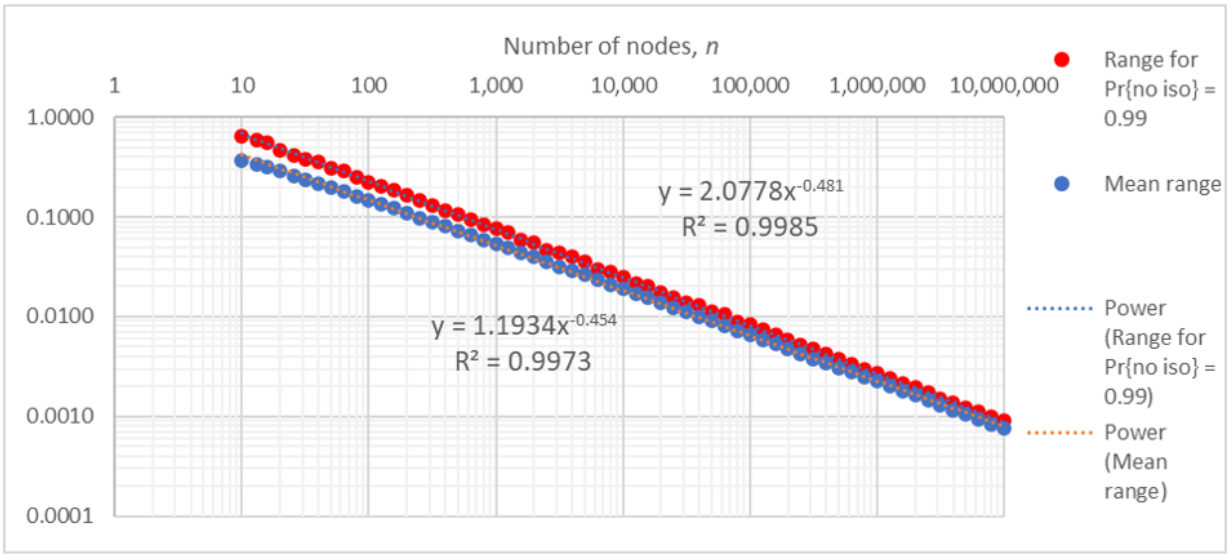

Figure 118: LRM for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth.

5.2.3.2 Flat Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small+Large-n Set

As discussed in Section 3.5, Penrose proposed that an RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem for RGGs in a flat unit square with r-for-given-n growth using the small+large-n set.

Figure 119 shows the fraction of simulations for which increasing the range to the point where the last isolated node becomes connected also causes the RGG to become fully connected.[44] If the First Claim of Penrose's Theorem were correct for RGGs in a flat unit square with r-for-given-n growth, then the fraction would be 1 or close to it for the larger values of n. Interpreting this fraction as a probability, for RGGs in a flat unit square with r-for-given-n growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.415 for $n = 10$, falls to about 0.344 for $n = 40$,[45] and increases to about 0.789 for $n = 10{,}000{,}000$.[46]

[44] For n-for-given-r growth, it is straightforward to determine when "adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected" because the number of nodes for those two conditions are integers and the comparison is exact. For r-for-given-n growth, on the other hand, determining when "increasing the range to the point where the last isolated node becomes connected also causes the RGG to become fully connected" is not as simple because the ranges are floating point numbers and they are both obtained using the inexact method of binary search.

The rule used in this paper was that the two ranges were considered to be "the same" if the range required for the last isolated node to disappear was greater than or equal to a threshold value of 0.9999 of the range required for the RGG to become fully connected.

Even though the inexact method of binary search was used, there was no case where a flat or toroidal simulation for r-for-given-n growth reported a range for which the last isolated node disappeared that was larger than the range for a fully connected RGG for the same randomly generated set of nodes.

[45] It is unclear as to why the probability that an RGG in a flat unit square in 2D with r-for-given-n growth becomes fully connected when the last isolated node is absorbed would decrease and then increase as n increases. This phenomenon is remarked upon in subsequent footnotes: see references to this footnote #45.

[46] The fraction 0.789 for $n = 10{,}000{,}000$ was obtained using the value of 0.9999 for the threshold of the ratios to be considered "the same," as discussed in footnote #44. The sensitivity of the estimate of the fraction to the threshold value can be roughly gauged as follows. The fraction 0.791 was obtained using the threshold value of 0.999 and the fraction 0.829 was obtained using the threshold value of 0.99. The shapes of the curves

Even though this probability increases as n increases, it shows that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for r-for-given-n growth of an RGG in a flat unit square.

Figure 120 relates the range needed for full connectivity to that required for no isolated nodes. For each randomly generated geometry for each n, the ratio of the range required to achieve full connectivity to the range where isolated nodes disappear is calculated, and the mean for all of the geometries for that value of n is plotted on the graph. For small numbers of nodes, considerably greater range is required to achieve full connectivity after the disappearance of isolated nodes, which is consistent with the discussion of Figure 119. The ratios may asymptotically approach 1 for even larger numbers of nodes but the ratio was observed to be 1.0102 for RGGs containing 10,000,000 nodes.

Figure 121 shows the error of the LRM (92) given in Section 5.2.3.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. It can be seen that the error percentages hover close to zero; it was noted in Section 5.2.3.1 that MAPE for (92) is 1.85% and max error of the predicted value of r is 6.89%. This figure also includes the relative error of the LRM (84) given in Section 5.2.1.1 that was obtained for the small-n set and, although this LRM is slightly different from that obtained for the small+large-n set, it closely tracks the relative error of (92) over the full small+large-n set.

Figure 121 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). These error values are considerably greater than those for the LRM.

Figure 122 shows the error of the LRM (94) given in Section 5.2.3.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.2.3.1 that MAPE for (94) is 1.79% and max error of the predicted value of r is 6.38%. This figure also includes the relative error of the LRM (86) given

were practically unchanged other than slight proportional shifts upward for the smaller threshold values.

in Section 5.2.1.1 that was obtained for the small-n set; as seen for full connectivity, this LRM for no isolated nodes closely tracks the relative error of (94) over the full small+large-n set.

Figure 122 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). Even for the lesser condition of having no isolated nodes, these error values are considerably greater than those for the LRM.

Figure 123 shows values of c_P for the 99%ile values of r, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of n, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 117. Similarly, Figure 124 shows values of c_P obtained from (34) for the r and n pairs shown in Figure 118 for no isolated nodes.

The visible trends for both fully connected RGGs and RGGs with no isolated nodes for the 99%ile values plotted for c_P in Figure 123 and Figure 124, respectively, are that there is an initial slight rise for the smaller n and then a slight falling off for the larger n. The maximum c_P value observed for any n in Figure 123 for full connectivity was 35.1087 and the maximum of c_P value observed for any n in Figure 124 for no isolated nodes also was 35.1087. The value of c_P at $n = 10{,}000{,}000$ in Figure 123 for full connectivity was 9.2802 and in Figure 124 for no isolated nodes it was 9.1242; both of these values are higher than Penrose's value of $c_P = 4.6001$ required for $C = 0.99$.

As a reminder, Penrose's Theorem was based on the r-for-given-n growth model without border effects and thus does not necessarily apply to the flat space assumed in this section. These comparisons are intended only to assist in determining when they do apply.

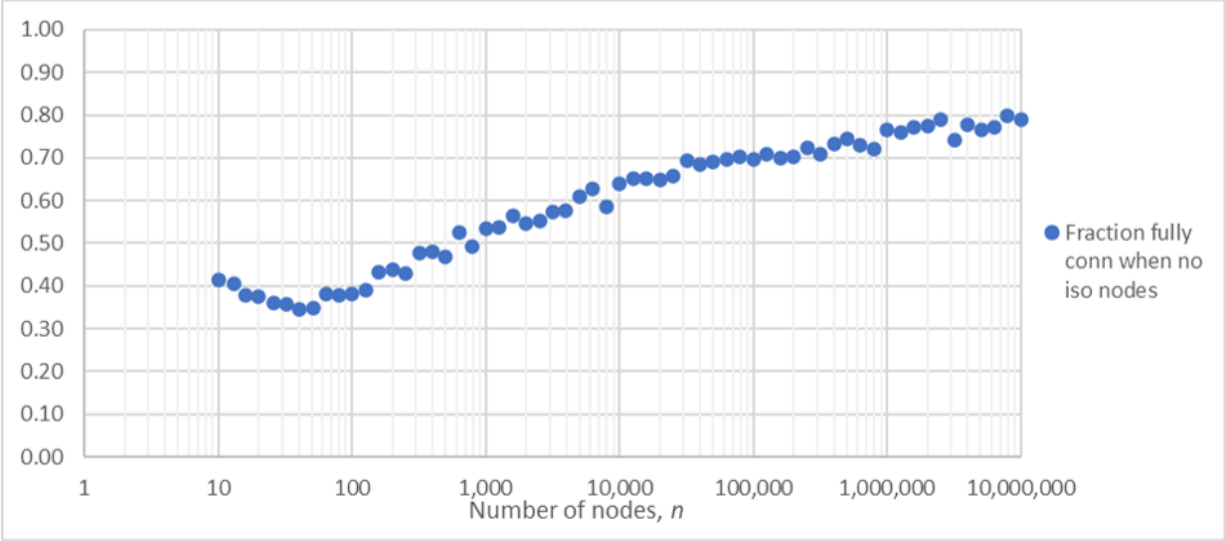

Figure 119: Fraction of simulations for which increasing the range to the point where the last isolated node becomes connected also causes RGGs in a flat unit square to become fully connected with r-for-given-n growth.

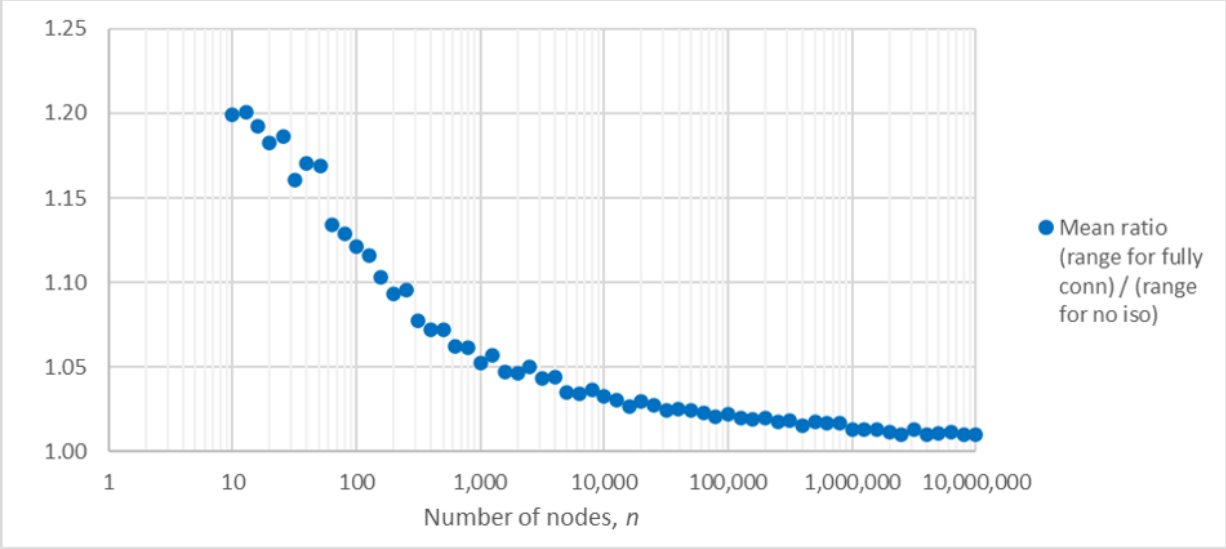

Figure 120: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in RGGs in a flat unit square with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry.

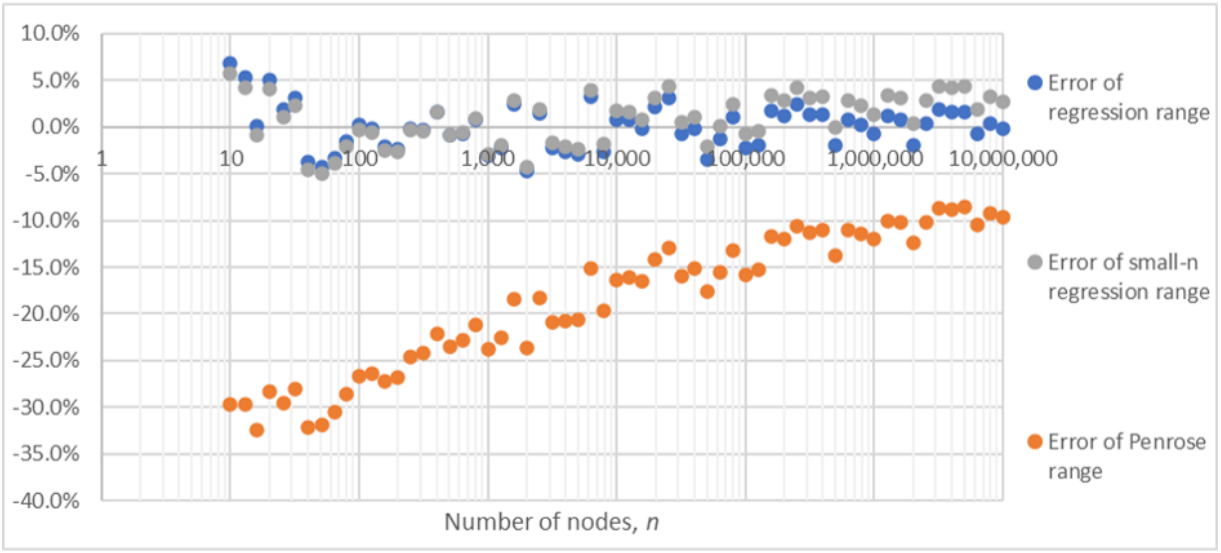

Figure 121: Relative error of LRM (92) obtained from the small+large-n set and LRM (84) obtained from the small-n set for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

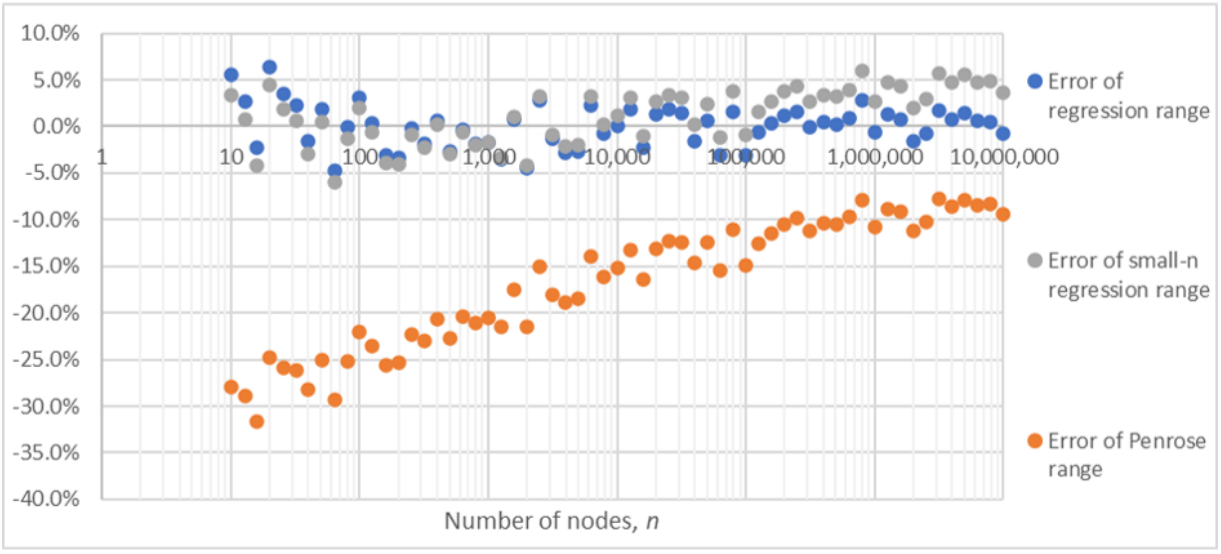

Figure 122: Relative error of LRM (94) obtained from the small+large-n set and LRM (86) obtained from the small-n set for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

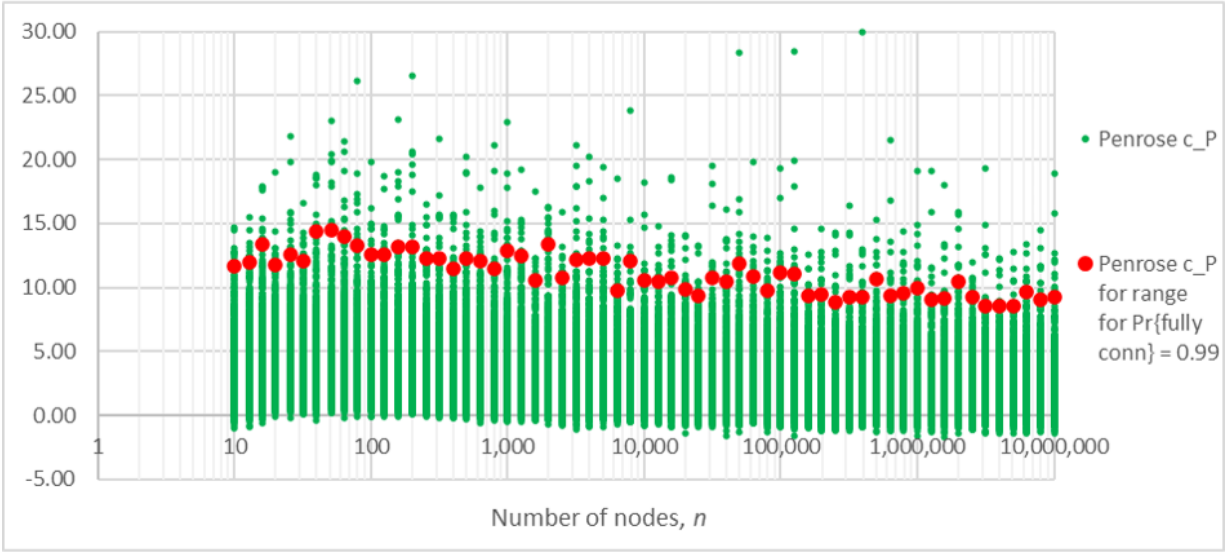

Figure 123: Values of Penrose's c_P for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

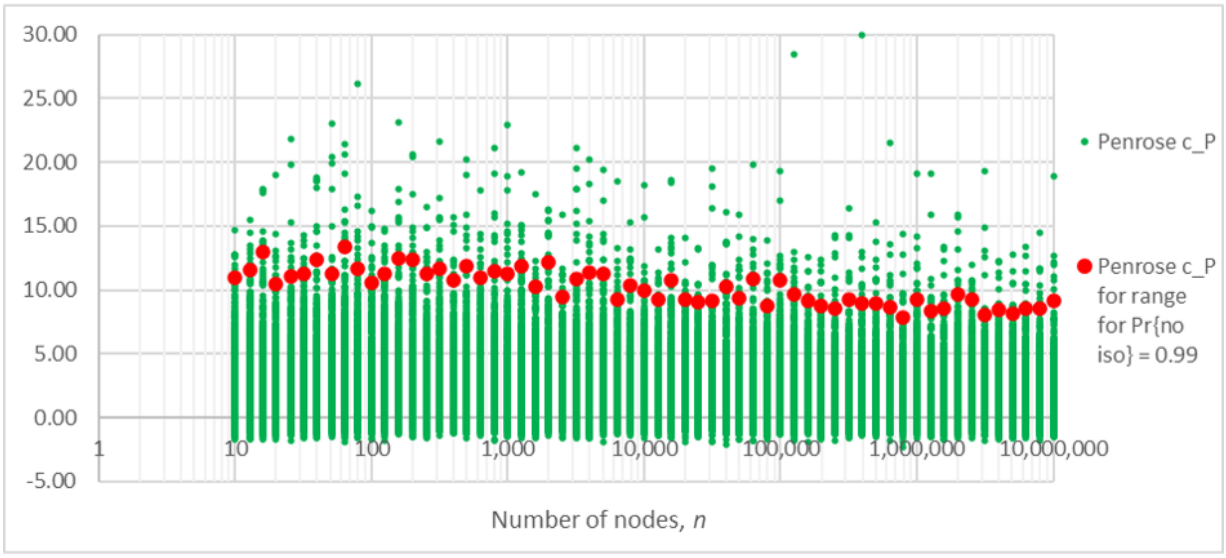

Figure 124: Values of Penrose's c_P for 99%ile range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

5.2.4 Flat Unit Square: Comparing Linear Regression Models from Small-n Set, Large-n Set, and Small+Large-n Set

This section compares the MAPE and max error resulting from using an LRM derived from one set of numbers of nodes to predict ranges for other sets of numbers of nodes. Only the results for the 99%ile range for full connectivity are compared.

- LRM (84) obtained from small-n set:
  - Applied to small-n set: MAPE 1.87%, max error 6.25%
  - Applied to large-n set: MAPE 1.99%, max error 8.59%
  - Applied to small+large-n set: MAPE 1.87%, max error 7.44%
- LRM (88) obtained from large-n set:
  - Applied to small-n set: MAPE 2.64%, max error 4.38%
  - Applied to large-n set: MAPE 1.18%, max error −2.32%
  - Applied to small+large-n set: MAPE 1.27%, max error 2.47%
- LRM (92) obtained from small+large-n set:
  - Applied to small-n set: MAPE 2.35%, max error 5.70%
  - Applied to large-n set: MAPE 1.85%, max error 8.03%
  - Applied to small+large-n set: MAPE 1.85%, max error 6.89%

For flat RGGs with r-for-given-n growth all of the MAPE values are fairly small percentages regardless of which LRM is applied to a particular set of numbers of nodes. The smallest MAPE was obtained by using the LRM derived from that set of numbers of nodes but the differences were small and in two cases there were ties for the minimum.

The smallest max error resulted from using the LRM obtained from that set of numbers of nodes except for the small+large-n set.

These results are a general indication that the linearity of the log-log transformed 99%ile ranges for flat RGGs with r-for-given-n growth holds well over the various intervals represented by the small-n set, large-n set, and their union.

5.3 Toroidal Unit Square with r-for-given-n Growth

This section considers RGGs consisting of nodes scattered uniformly at random positions in a two-dimensional square space of area $A = 1$ with wraparound at the border. This is a toroidal space so there are no border effects. The RGGs evolve according to the r-for-given-n growth model.

All of the situations considered prior to this section varied slightly from the assumptions of Penrose's work and thus the comparisons to Penrose's Theorem were "off-label" applications. The situations considered in this section and its subsections match Penrose's assumptions for RGGs in two-dimensional spaces.

Results are given separately for the small-n, large-n, and small+large-n sets.

For each simulated geometry for a given configuration, a fixed number of nodes n are placed in the space to determine the minimum range r required to achieve the condition of 100% connectivity or no isolated nodes, and the 99%ile and mean statistical metrics of those ranges are obtained for that configuration.

5.3.1 Toroidal Unit Square with r-for-given-n Growth for Small Numbers of Nodes

Section 5.3.1.1 and Section 5.3.1.2 present results for RGGs in toroidal unit squares with r-for-given-n growth using the small-n set consisting of the numbers of nodes from 10 up to 100,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.3.1.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Small-n Set

Figure 125 and Figure 126 show the trend lines for LRMs for the range required to achieve full connectivity and no isolated nodes, respectively, for RGGs in a toroidal unit square with r-for-given-n growth for the small-n set. Trend lines are shown for the 99%ile and mean statistical metrics for the range based on the sample of size $N = 1{,}000$ simulated geometries per value of n.

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 125, (83) becomes

$$r = 1.3167\, n^{-0.4490} \tag{96}$$

with an R^2 value of 0.9990. For the simulated values of n, MAPE is 3.01% and max error of the value of the predicted r is 13.71%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 125,, the LRM is

$$r = 0.9648\, n^{-0.4331} \tag{97}$$

with an R^2 value of 0.9991. For the simulated values of n, MAPE is 2.90% and max error of the value of the predicted r is 11.50%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 126,, the LRM is

$$r = 1.2839\, n^{-0.4474} \tag{98}$$

with an R^2 value of 0.9991. For the simulated values of n, MAPE is 2.85% and max error of the value of the predicted r is 12.54%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 126,, the LRM is

$$r = 0.8634\, n^{-0.4230} \tag{99}$$

with an R^2 value of 0.9986. For the simulated values of n, MAPE is 3.66% and max error of the value of the predicted r is 12.76%.

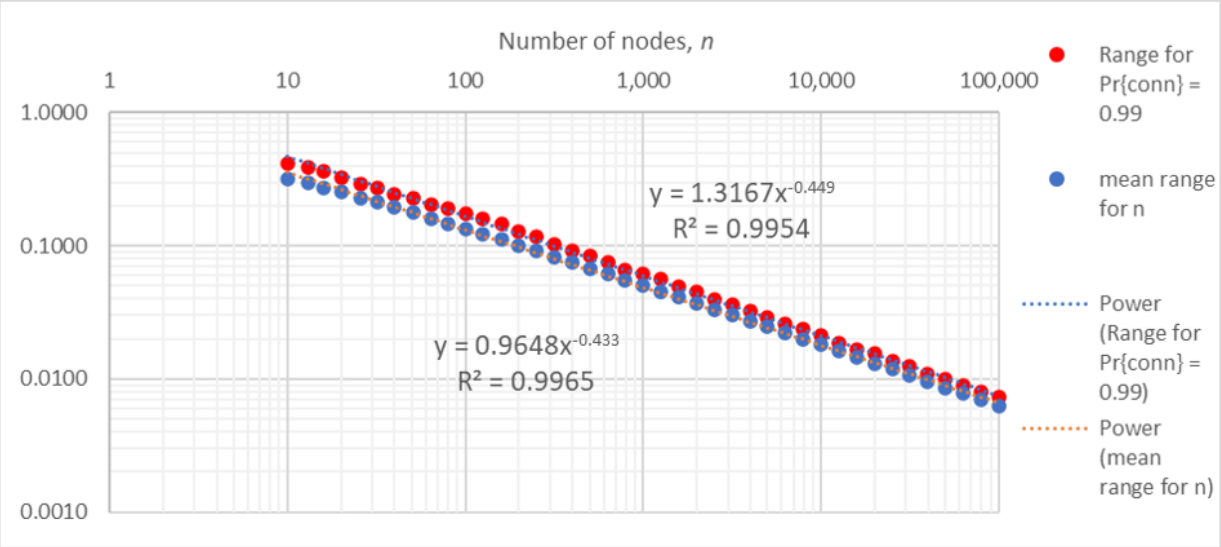

Figure 125: LRM for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth.

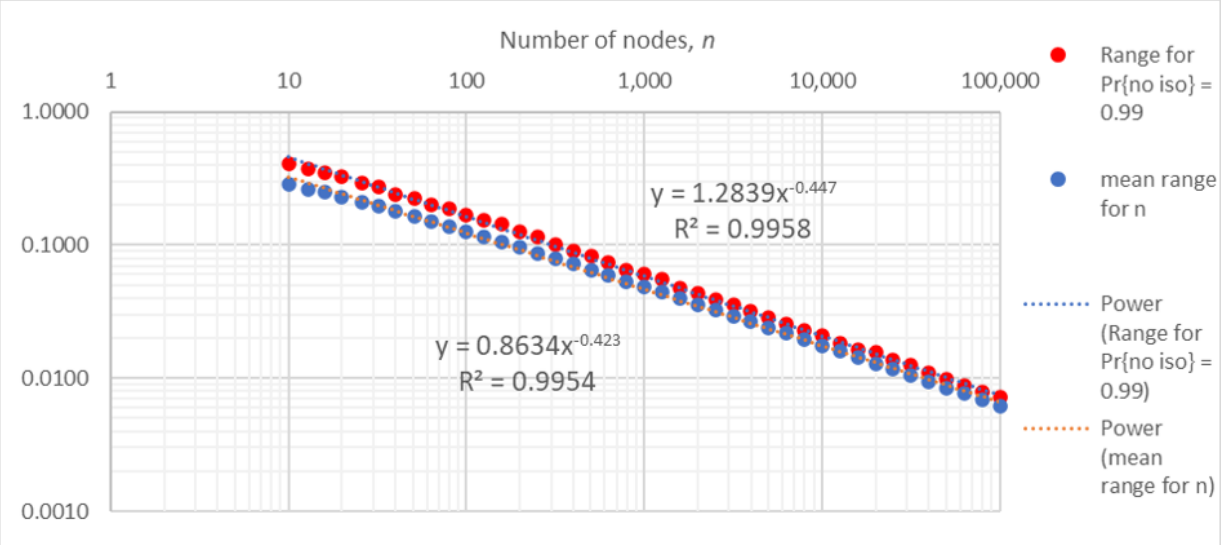

Figure 126: LRM for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth.

5.3.1.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small-n Set

As discussed in Section 3.5, Penrose proposed that a toroidal RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem in precisely this situation for the small-n set consisting of values of n up to 100,000.

Figure 127 shows the error of the LRM (96) given in Section 5.3.1.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. In spite of the fact that the probability is significant that the RGG is not inevitably fully connected at the same instant the last isolated node disappears (looking ahead to Figure 135 in Section 5.3.3.2), it can be seen that the error percentages hover close to zero even for the values of n only slightly greater than 10; it was noted in Section 5.3.1.1 that MAPE for (96) is 3.01% and max error of the predicted value of r is 13.71%.

Figure 127 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). For RGGs in a toroidal unit square with r-for-given-n growth, the predicted range using Penrose's Theorem for probability 0.99 of being fully connected has practically the same small error as the LRM based on simulated geometries. MAPE for Penrose's Theorem is 2.19% and max error of the predicted value of r is 13.84%.

Figure 128 shows the error of the LRM (98) given in Section 5.3.1.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.3.1.1 that MAPE for (98) is 2.85% and max error of the predicted value of r is 12.54%.

Figure 128 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). The predicted range resulting from Penrose's Theorem again has practically the same small error as that for

the LRM for no isolated nodes. MAPE for Penrose's Theorem is 2.66% and max error of the predicted value of r is 15.11%.

Penrose's Theorem is based on the r-for-given-n growth model without border effects and thus would be expected to be directly applicable to the toroidal space assumed in this section. Even though an RGG in a toroidal unit square does not inevitably become fully connected at the point where the last isolated node disappears, these comparisons show that the expressions (14), (15), and (16), provide excellent predictions of the minimum range required for a given number of nodes for probability 0.99 of being fully connected or having no isolated nodes. These findings are consistent with the error percentages plotted in Figure 127 and Figure 128 for $n \geq 100$.

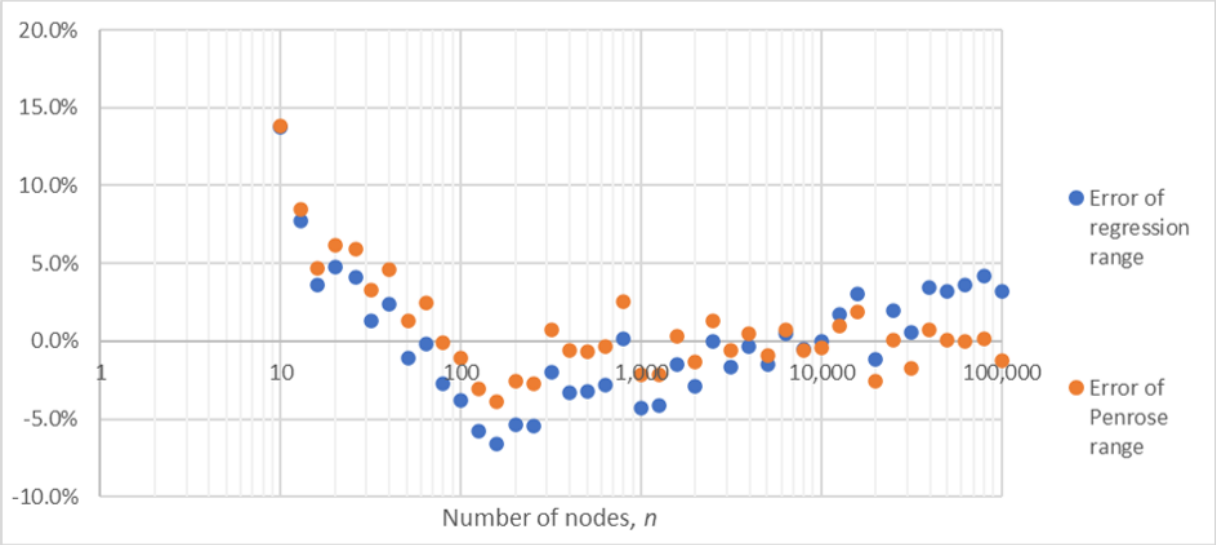

Figure 127: Relative error of LRM (96) for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

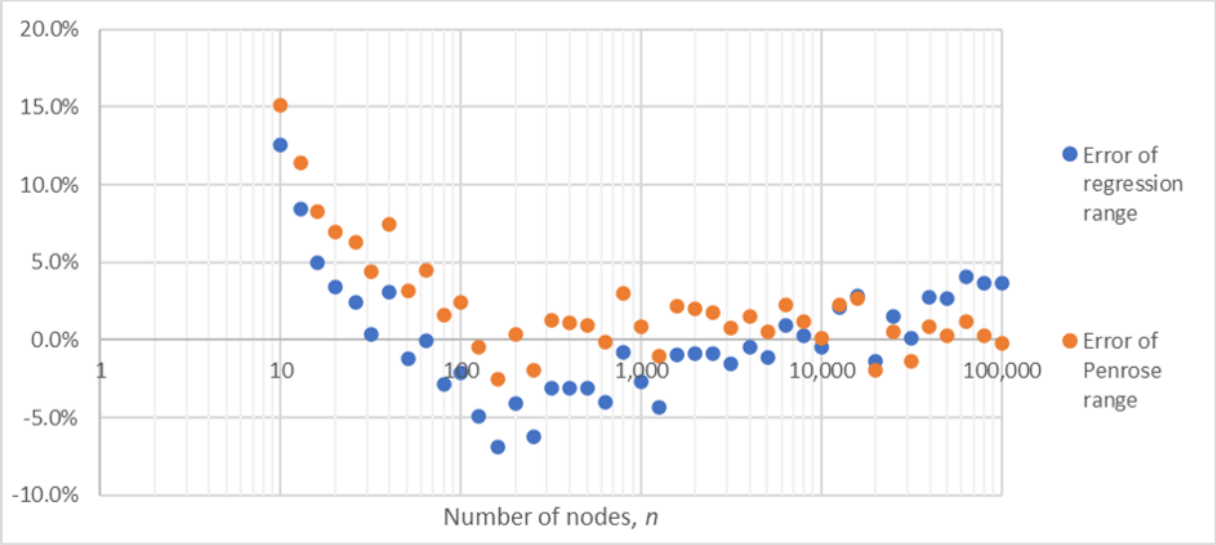

Figure 128: Relative error of LRM (98) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

5.3.2 Toroidal Unit Square with r-for-given-n Growth for Large Numbers of Nodes

Section 5.3.2.1 and Section 5.3.2.2 present results for RGGs in toroidal unit squares with r-for-given-n growth using the large-n set consisting of the numbers of nodes from 100,000 up to 10,000,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.3.2.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Large-n Set

Figure 129 and Figure 130 show the trend lines for LRMs for the range required to achieve full connectivity and no isolated nodes, respectively, for RGGs in a toroidal unit square with r-for-given-n growth for the large-n set. Trend lines are shown for the 99%ile and mean statistical metrics for the range based on the sample of size $N = 1{,}000$ simulated geometries per value of n.

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 129, (83) becomes

$$r = 1.6453\, n^{-0.4720} \tag{100}$$

with an R^2 value of 0.9999. For the simulated values of n, MAPE is 0.59% and max error of the value of the predicted r is 1.34%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 129,, the LRM is

$$r = 1.3434\, n^{-0.4660} \tag{101}$$

with an R^2 value of 1.0000. For the simulated values of n, MAPE is 0.20% and max error of the value of the predicted r is -0.52%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 130,, the LRM is

$$r = 1.6392\, n^{-0.4721} \tag{102}$$

with an R^2 value of 0.9999. For the simulated values of n, MAPE is 0.56% and max error of the value of the predicted r is -1.36%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 130,, the LRM is

$$r = 1.3086\, n^{-0.4647} \tag{103}$$

with an R^2 value of 1.0000. For the simulated values of n, MAPE is 0.23% and max error of the value of the predicted r is -0.54%.

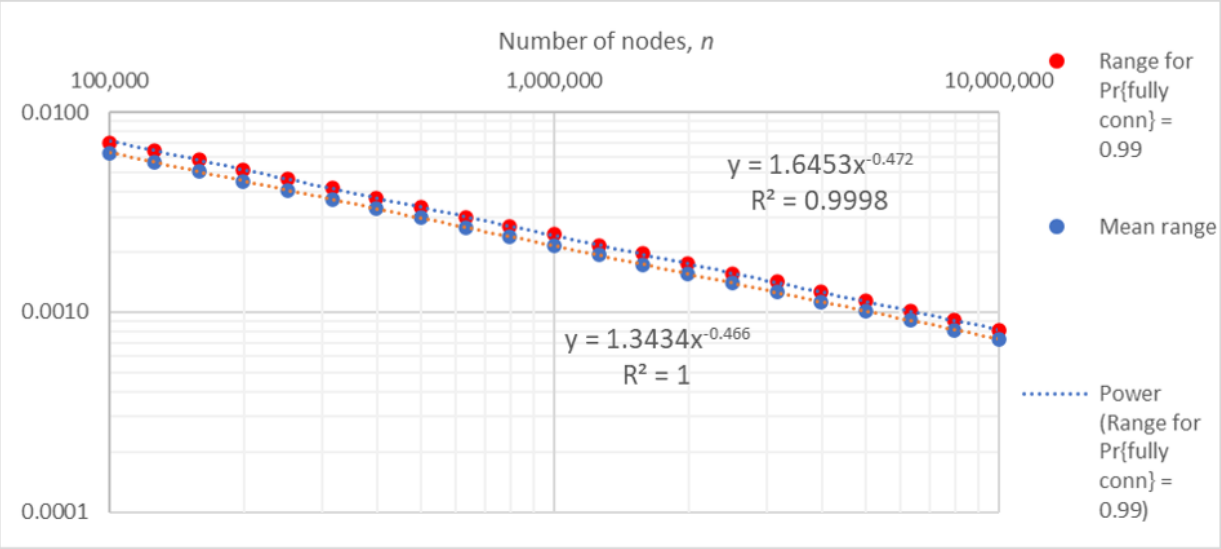

Figure 129: LRM for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth.

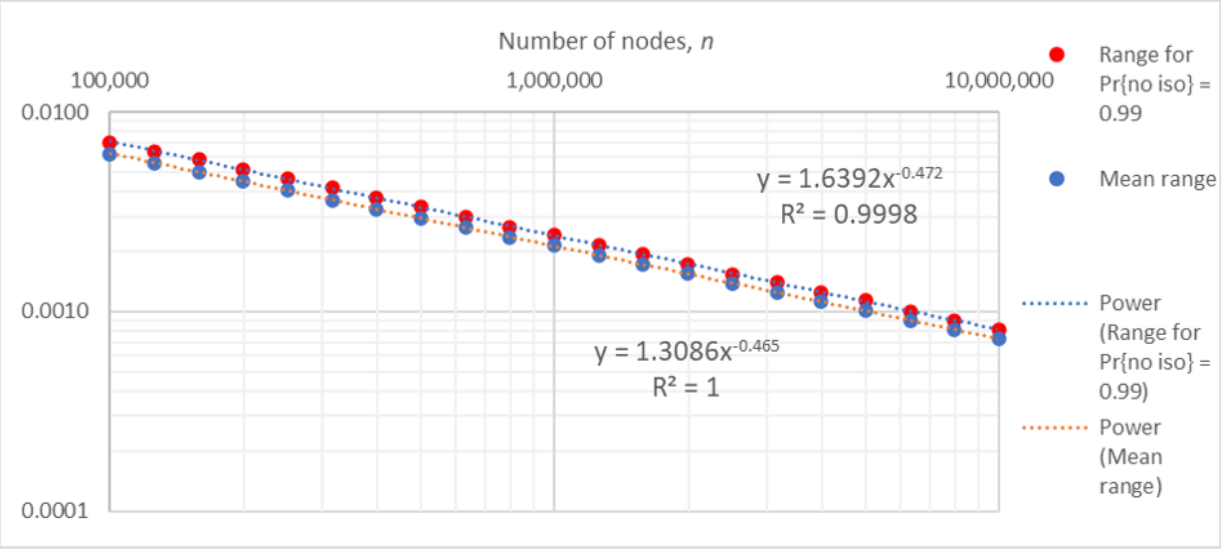

Figure 130: LRM for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth.

5.3.2.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Large-n Set

As discussed in Section 3.5, Penrose proposed that a toroidal RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem in precisely this situation for the large-n set consisting of values of n from 100,000 up to 10,000,000.

Figure 131 shows the error of the LRM (100) given in Section 5.3.2.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. In spite of the fact that the probability is significant that the RGG is not inevitably fully connected at the same instant the last isolated node disappears (looking ahead to Figure 135 in Section 5.3.3.2), it can be seen that the error percentages hover close to zero over the values of n in the large-n set; it was noted in Section 5.3.2.1 that MAPE for (100) is 0.59% and max error of the predicted value of r is 1.34%.

Figure 131 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). For RGGs in a toroidal unit square with r-for-given-n growth, the predicted range using Penrose's Theorem for probability 0.99 of being fully connected has practically the same small error as the LRM based on simulated geometries. MAPE for Penrose's Theorem is 0.60% and max error of the predicted value of r is −1.31%.

Figure 132 shows the error of the LRM (102) given in Section 5.3.2.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.3.2.1 that MAPE for (102) is 0.56% and max error of the predicted value of r is 1.36%.

Figure 132 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). The predicted range resulting from Penrose's Theorem again has practically the same small error as that for

the LRM for no isolated nodes. MAPE for Penrose's Theorem is 0.60% and max error of the predicted value of r is 1.39%.

Penrose's Theorem is based on the r-for-given-n growth model without border effects and thus would be expected to be directly applicable to the toroidal space assumed in this section. Even though an RGG in a toroidal unit square does not inevitably become fully connected at the point where the last isolated node disappears, these comparisons show that the expressions (14), (15), and (16), provide excellent predictions of the minimum range required for a given number of nodes for probability 0.99 of being fully connected or having no isolated nodes. These findings are consistent with the error percentages plotted in Figure 131 and Figure 132 for $n \geq 100$.

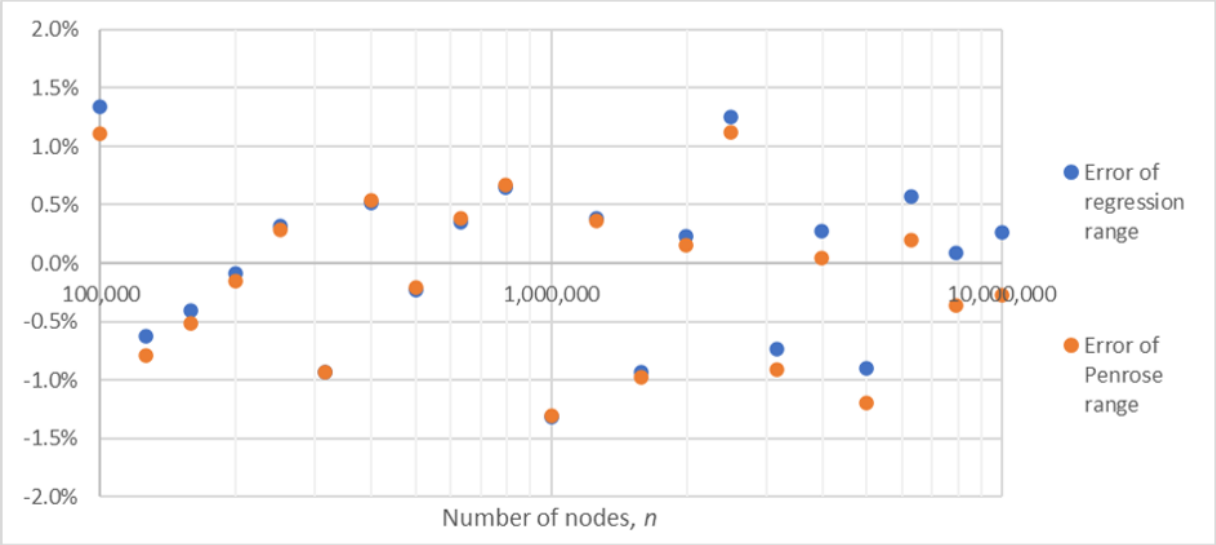

Figure 131: Relative error of LRM (100) for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

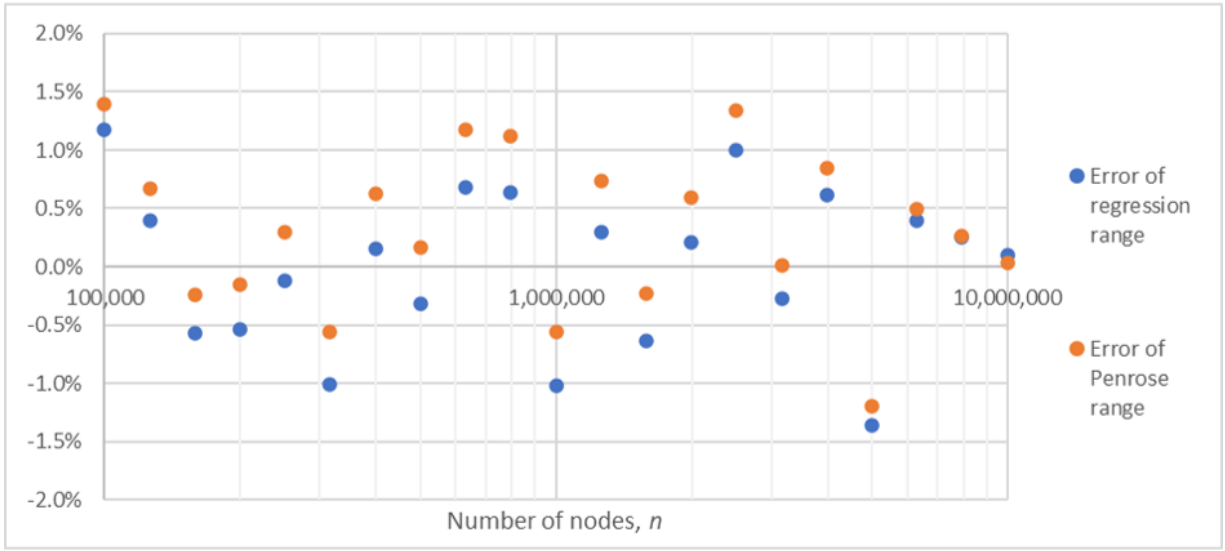

Figure 132: Relative error of LRM (102) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

5.3.3 Toroidal Unit Square with r-for-given-n Growth for Small and Large Numbers of Nodes

Section 5.3.3.1 and Section 5.3.3.2 present results for RGGs in toroidal unit squares with r-for-given-n growth using the small+large-n set consisting of the numbers of nodes from 10 up to 10,000,000.

These sections give LRMs for full connectivity and no isolated nodes.

They compare the numbers of nodes required for full connectivity and no isolated nodes obtained from simulations to Penrose's predicted numbers of nodes.

5.3.3.1 Toroidal Unit Square: Full Connectivity and No Isolated Nodes for Small+Large-n Set

Figure 133 and Figure 134 show the trend lines for LRMs for the range required to achieve full connectivity and no isolated nodes, respectively, for RGGs in a toroidal unit square with r-for-given-n growth for the small+large-n set. Trend lines are shown for the 99%ile and mean statistical metrics for the range based on the sample of size $N = 1{,}000$ simulated geometries per value of n.

The LRMs with the values of m and b replaced in (83) are as follows. For the 99%ile range required to achieve full connectivity with r-for-given-n growth as shown in Figure 133, (83) becomes

$$r = 1.3988\, n^{-0.4590} \tag{104}$$

with an R^2 value of 0.9995. For the simulated values of n, MAPE is 3.27% and max error of the value of the predicted r is 15.02%.

For the mean range required to achieve full connectivity with r-for-given-n growth as shown in Figure 133,, the LRM is

$$r = 1.0448\, n^{-0.4462} \tag{105}$$

with an R^2 value of 0.9993. For the simulated values of n, MAPE is 3.86% and max error of the value of the predicted r is 17.46%.

For the 99%ile range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 134,, the LRM is

$$r = 1.3723\, n^{-0.4579} \tag{106}$$

with an R^2 value of 0.9995. For the simulated values of n, MAPE is 3.29% and max error of the value of the predicted r is 15.43%.

For the mean range required to achieve no isolated nodes with r-for-given-n growth as shown in Figure 134,, the LRM is

$$r = 0.9572\, n^{-0.4400} \tag{107}$$

with an R^2 value of 0.9989. For the simulated values of n, MAPE is 4.95% and max error of the value of the predicted r is 19.85%.

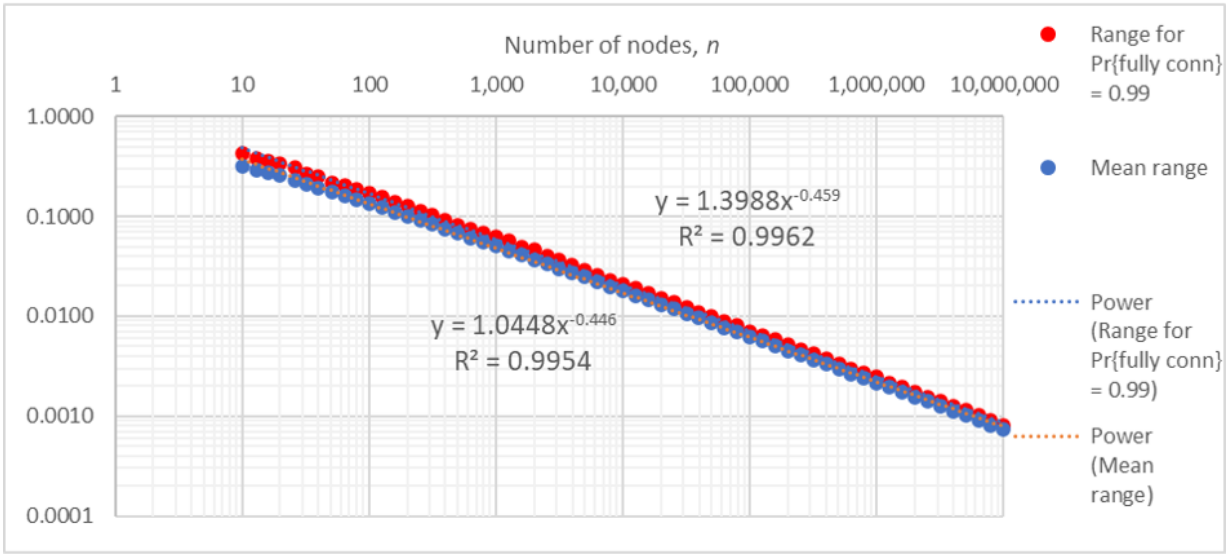

Figure 133: LRM for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth.

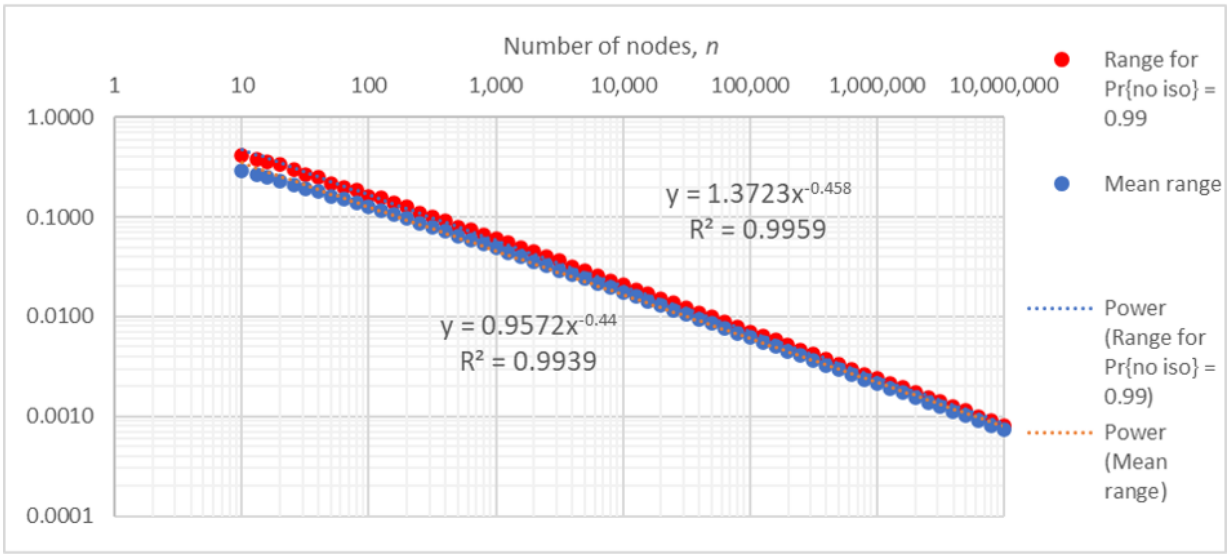

Figure 134: LRM for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth.

5.3.3.2 Toroidal Unit Square: Isolated Nodes, Connectivity, and Testing Penrose's Theorem for Small+Large-n Set

As discussed in Section 3.5, Penrose proposed that an RGG with r-for-given-n growth and "n large" becomes fully connected when the last isolated node is absorbed. The following figures show results that test the accuracy of Penrose's Theorem for RGGs in a toroidal unit square for the small+large-n set. This paper tests Penrose's Theorem for many "off-label" cases but this section tests his theorem under precisely the conditions for which it was proved for values of n up to 10,000,000; by almost any standard this would be accepted as "n large."

Figure 135 shows the fraction of simulations for which increasing the range to the point where the last isolated node becomes connected also causes the RGG to become fully connected.[47] If the First Claim of Penrose's Theorem were correct for RGGs in a toroidal unit square with r-for-given-n growth, then the fraction would be 1 or close to it for the larger values of n. The plot for this fraction for RGGs in a toroidal unit square is slightly higher than the corresponding values for RGGs in a flat unit square. Interpreting this fraction as a probability, for a toroidal unit square with r-for-given-n growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.468 for $n = 10$, falls to 0.385 at $n = 26$,[48] and increases to about 0.821 for $n = 10{,}000{,}000$.[49] Even though this probability increases as n increases, it shows again that the First Claim of Penrose's Theorem is not a sufficient condition for full connectivity for r-for-given-n growth of an RGG in a toroidal unit square.

[47] Two ranges were considered to be "the same" if the range required for the last isolated node to disappear was greater than or equal to 0.9999 times the range required for the RGG to become fully connected, as discussed in footnote #44 in Section 5.2.3.2.

[48] It is unclear as to why the probability that an RGG in a toroidal unit square in 2D with r-for-given-n growth becomes fully connected when the last isolated node is absorbed would decrease and then increase as n increases. This phenomenon was noted in footnote #45 for the same configurations in a flat unit square.

[49] The fraction 0.821 for $n = 10{,}000{,}000$ was obtained using the value of 0.9999 for the threshold of the ratios to be considered "the same," as discussed in footnote #44 in Section 5.2.3.2. The fraction 0.828 was obtained using the threshold value of 0.999 and the fraction 0.872 was obtained using the threshold value of 0.99. The shapes of the curves were nearly unchanged other than slight proportional shifts upward for the smaller threshold values.

Figure 136 relates the range needed for full connectivity to that required to achieve only the condition of no isolated nodes. For each randomly generated geometry, the ratio of the range required to achieve full connectivity to the range where isolated nodes disappear is calculated, and the mean of the ratio for that value of n is plotted on the graph. For small numbers of nodes, greater range is required to achieve full connectivity after the disappearance of isolated nodes which is consistent with the discussion of Figure 135 but the increase in range is less than that required for RGGs in a flat unit square. The ratios may asymptotically approach 1 for very large numbers of nodes, and indeed the observed ratio for these simulations is only 1.0052 when the RGGs contain 10,000,000 nodes.

Figure 137 shows the error of the LRM (104) given in Section 5.3.3.1 for the 99%ile range required to achieve full connectivity with respect to the observed range obtained from the simulated geometries. In spite of the fact that the probability is significant that the RGG is not inevitably fully connected at the same instant the last isolated node disappears, it can be seen that the error percentages hover close to zero for the values of n that are 20 and greater; it was noted in Section 5.3.3.1 that MAPE for (104) is 3.27% and max error of the predicted value of r is 15.02%. This figure also includes the relative error of the LRM (96) given in Section 5.3.1.1 that was obtained for the small-n set; it shows increasing relative error for numbers of nodes larger than those in the small-n set.

Figure 137 also shows the same error calculations for Penrose's predicted range for full connectivity with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). For RGGs in a toroidal unit square with r-for-given-n growth, the predicted range using Penrose's Theorem for probability 0.99 of being fully connected has practically the same small error as the LRM based on simulated geometries and are even closer for the largest values of n. MAPE for Penrose's Theorem is 1.53% and max error of the predicted value of r is 10.90%.

The situation for no isolated nodes is similar to that for fully connected RGGs. Figure 138 shows the error of the LRM (106) given in Section 5.3.3.1 for the 99%ile range required to achieve no isolated nodes with respect to the observed range obtained from the simulated geometries. As for the case of full connectivity, it can be seen that the error percentages for the LRM for no isolated nodes hover close to zero. It was noted in Section 5.3.3.1 that MAPE for (106) is 3.29% and max error of the predicted value of r is

15.43%. This figure also includes the relative error of the LRM (98) given in Section 5.3.1.1 that was obtained for the small-n set; as for the case of full connectivity it shows increasing relative error for no isolated nodes for numbers of nodes larger than those in the small-n set.

Figure 138 also shows the same error calculations for Penrose's predicted range for no isolated nodes with $C = 0.99$. The ranges are given by (16) with $c_P = 4.6001$ for $C = 0.99$ from (15). The predicted range resulting from Penrose's Theorem again has practically the same small error as that for the LRM for no isolated nodes. MAPE for Penrose's Theorem is 1.79% and max error of the predicted value of r is 13.17%.

Figure 139 shows values of c_P for the 99%ile values of r, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of n, that achieved full connectivity. These values of c_P are given by (34) for the r and n pairs shown in Figure 133. Similarly, Figure 140 shows values of c_P obtained from (34) for the r and n pairs shown in Figure 134 for no isolated nodes.

The trends visible in Figure 139 and Figure 140 are that the 99%ile values plotted for c_P increase slightly from $n = 10$ to about $n = 1{,}000$, and then remain practically flat (with some noise). The maximum of the c_P observed for any range in Figure 139 for full connectivity was 11.6030 and the mean value of c_P was 4.7562 for the 99%ile values of r for $n \geq 1{,}000$. The maximum of the c_P observed for any range in Figure 140 for no isolated nodes was 11.6030 and the mean value of c_P was 4.5627 for the 99%ile values of r for $n \geq 1{,}000$. With respect to Penrose's value of $c_P = 4.6001$ required for $C = 0.99$, these mean values of c_P have an error of -1.33% for full connectivity and an error of -0.81% for no isolated nodes.

As noted above, this section considers precisely the case for which Penrose's Theorem was proved. Penrose's Theorem is based on the r-for-given-n growth model without border effects and thus is directly applicable to the toroidal space assumed in this section. Furthermore, this section tests his theorem with very large values of n. Figure 137 demonstrates that Penrose's Theorem closely estimates the 99%ile range required for a two-dimensional toroidal RGG to be fully connected with r-for-given-n growth even for as few as 20 nodes. This is in spite of the fact that Figure 135 shows that with only 20 nodes the probability that such an RGG is fully connected when the last isolated node disappears is only 0.390 and Figure 136 shows that, on average, the range required for an RGG to be

fully connected is 11.83% greater than that required for the last isolated node to disappear in the same RGG.

Even though Penrose's First Claim does not hold for numbers of nodes up to 10,000,000, the empirical evidence is that his Third Claim does appear to hold and expressions (14), (15), and (16) accurately predict the range required for such an RGG to be fully connected with the high probability of 0.99 for $n \geq 20$.

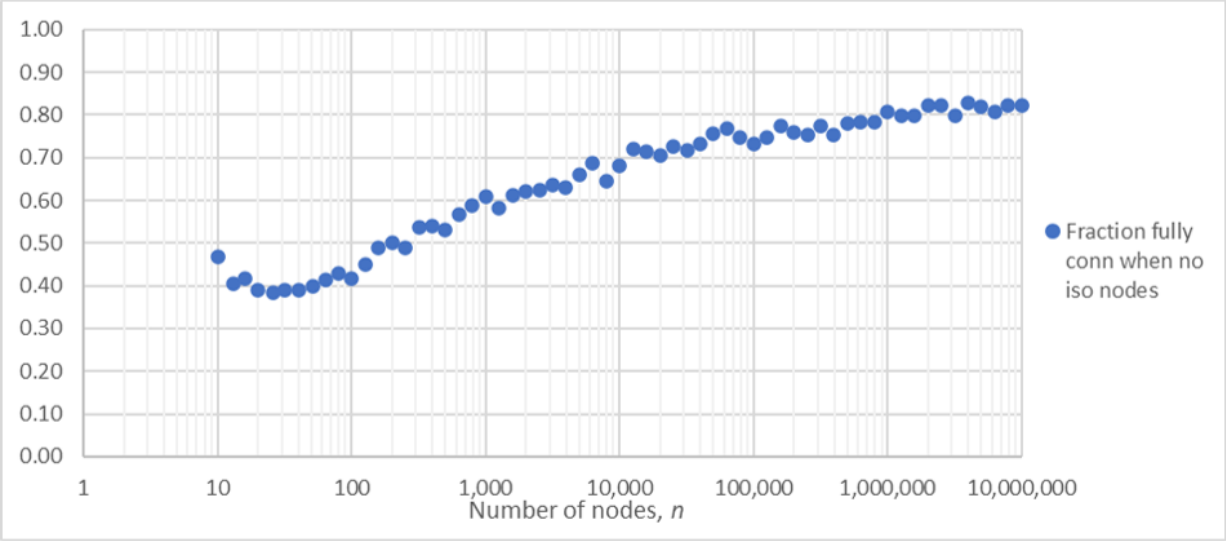

Figure 135: Fraction of simulations for which increasing the range to the point where the last isolated node becomes connected also causes RGGs in a toroidal unit square to become fully connected with r-for-given-n growth.

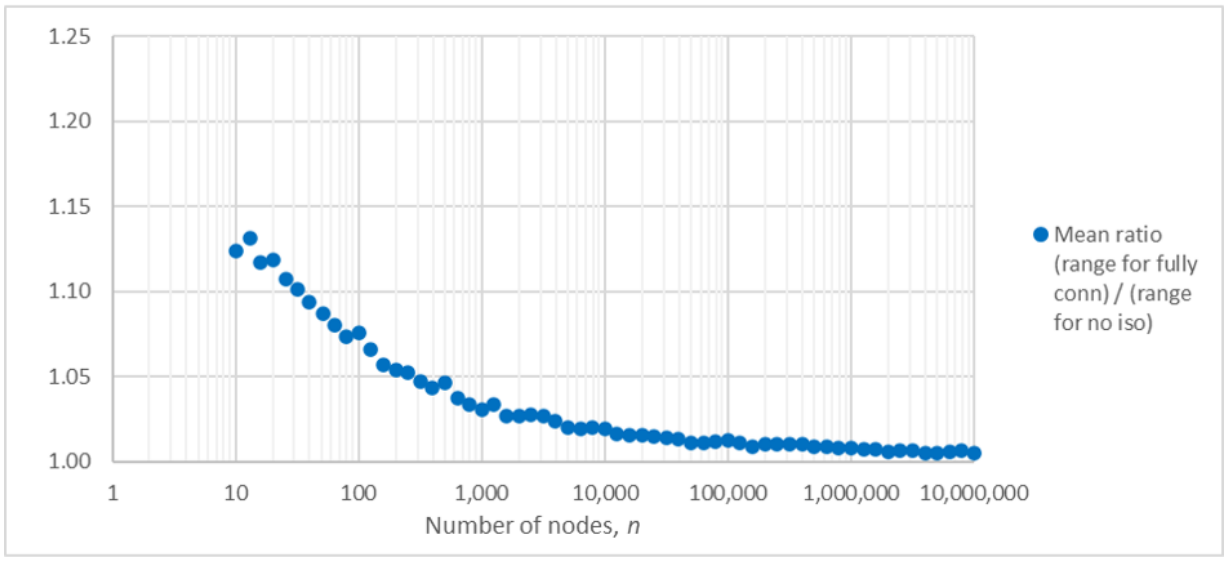

Figure 136: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in RGGs in a toroidal unit square with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry.

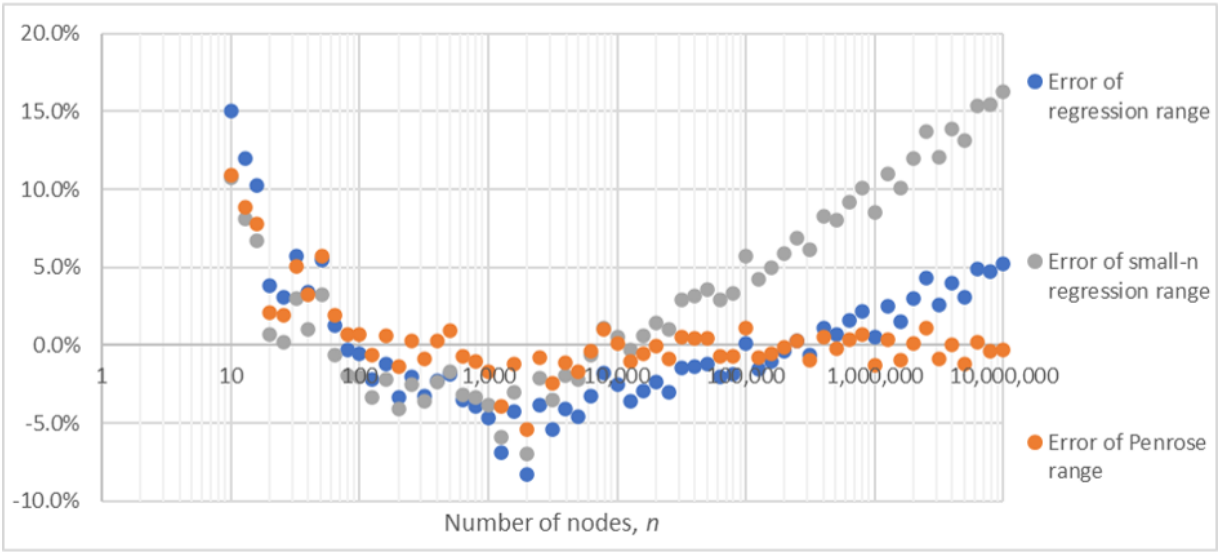

Figure 137: Relative error of LRM (104) obtained from the small+large-n set and LRM (96) obtained from the small-n set for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

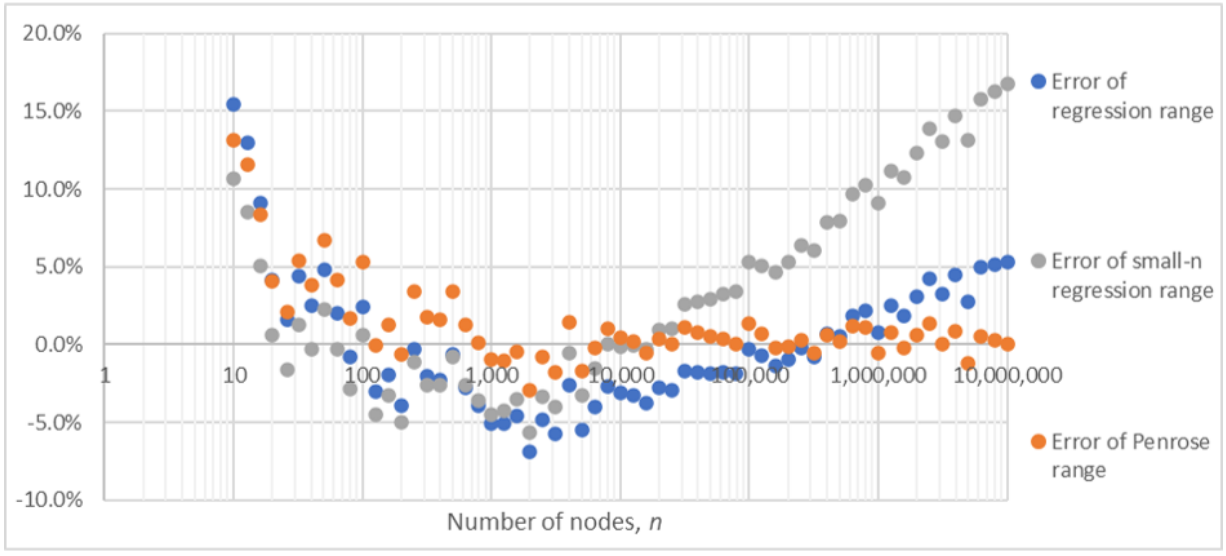

Figure 138: Relative error of LRM (106) obtained from the small+large-n set and LRM (98) obtained from the small-n set for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

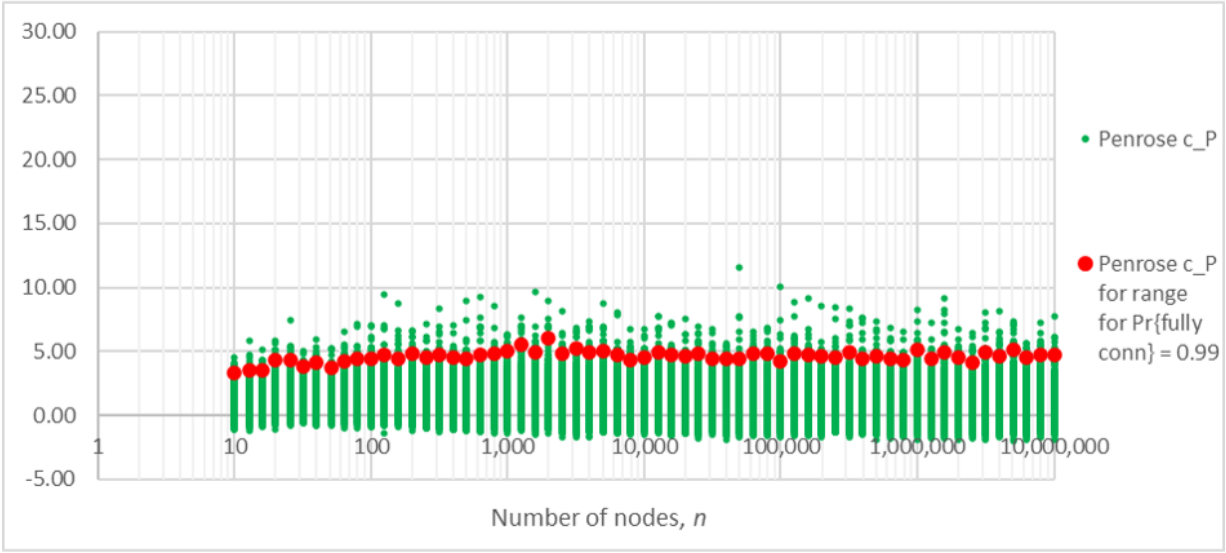

Figure 139: Values of Penrose's c_P for 99%ile range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

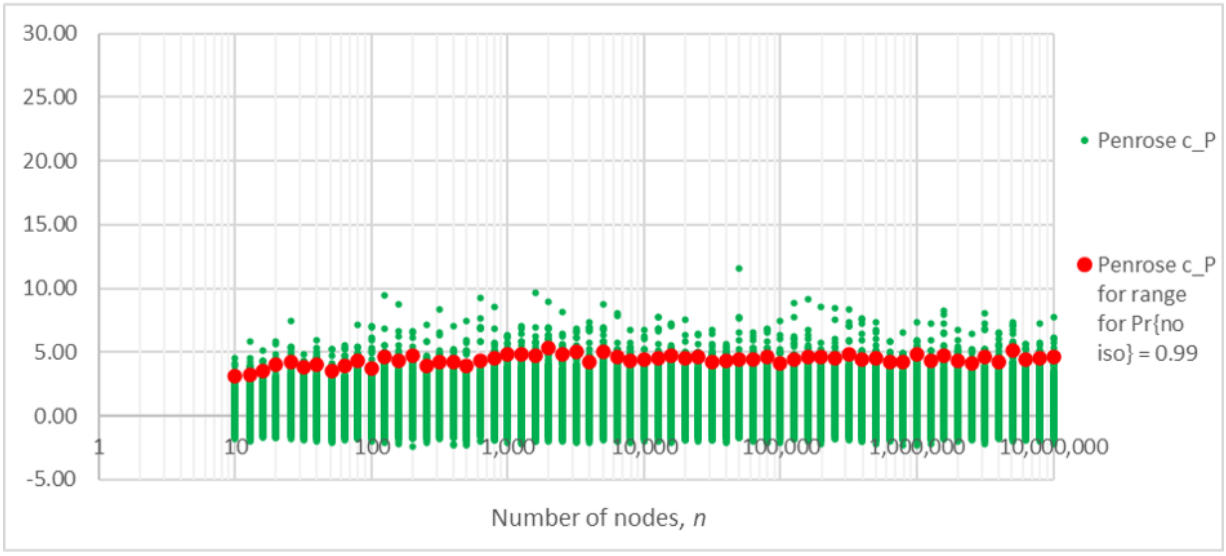

Figure 140: Values of Penrose's c_P for 99%ile range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth. The smaller green dots are the values of c_P for all n and r pairs that achieved full connectivity.

5.3.4 Toroidal Unit Square: Comparing Linear Regression Models from Small-n Set, Large-n Set, and Small+Large-n Set

This section compares the MAPE and max error resulting from using an LRM derived from one set of numbers of nodes to predict ranges for other sets of numbers of nodes. Only the results for the 99%ile range for full connectivity are compared.

- LRM (96) obtained from small-n set:
  - Applied to small-n set: MAPE 3.01%, max error 13.71%
  - Applied to large-n set: MAPE 7.06%, max error 34.76%
  - Applied to small+large-n set: MAPE 3.81%, max error 18.07%
- LRM (100) obtained from large-n set:
  - Applied to small-n set: MAPE 10.04%, max error 16.26%
  - Applied to large-n set: MAPE 0.59%, max error 1.34%
  - Applied to small+large-n set: MAPE 2.18%, max error 5.21%
- LRM (104) obtained from small+large-n set:
  - Applied to small-n set: MAPE 5.35%, max error 16.26%
  - Applied to large-n set: MAPE 4.84%, max error 31.27%
  - Applied to small+large-n set: MAPE 3.27%, max error 15.02%

For toroidal RGGs with r-for-given-n growth the smallest MAPE and smallest max error resulted when the LRM obtained from a set of numbers of nodes was applied to that set.

These results indicate that the linearity of the log-log transformed 99%ile ranges for toroidal RGGs with r-for-given-n growth does not strictly hold over the various intervals represented by the small-n set, large-n set, and their union but it is still quite good. The best results are obtained when the LRM derived from the small-n set is used for the small-n set and the LRM derived from the large-n set is used for the large-n set; however, the LRM derived from the small+large-n set still performed well over that entire set of numbers of nodes.

5.4 Koskinen's Regression Analyses for RGGs in a Flat Unit Square with r-for-given-n Growth

H. Koskinen addressed the problem of determining the conditions for the full connectivity of RGGs in a flat unit square with r-for-given-n growth. He applied regression analysis based on the results of simulations to solve two problems. The first was to estimate the range required to achieve full connectivity. The second was to obtain a function that would adjust the value of Penrose's c_P in (16) as a function of n for the difference between RGG growths in flat and toroidal spaces.

It is shown here that his results, which were obtained on the basis of simulations of RGGs containing only relatively small numbers of nodes, are extremely accurate even for large numbers of nodes.

He addressed k-connectivity of RGGs but his results only for 1-connectivity are tested here.

5.4.1 Testing Koskinen's Estimate of the Range for a Given Number of Nodes

Koskinen used the results of simulations that found the mean and several percentiles for the ranges required to achieve full connectivity as a function of the number of nodes in an RGG [Koskinen (2003a)] [Koskinen (2003b)]. The simulations involved 5,000 randomly generated geometries with the numbers of nodes $n = 5$, 10, 15, 20, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, and 350. He observed that plots of the *inverse squares* of these statistical metrics as a linear function of n formed very nearly straight lines.

A set of simulations using these configurations was performed for this paper and the 99%ile and mean range required to achieve full connectivity were obtained. The plot of the reciprocal of the squared mean range required to achieve full connectivity is shown in Figure 141 and it closely matches that shown in Figure 3.3b of Koskinen's thesis [Koskinen (2003a)].

The usual approach for linear regression using transformed variables is to find a transformation that yields an approximately straight line (or flat plane or hyperplane) and then find the coefficients for the "best straight line." Those coefficients minimize the sum of the squared residuals or SSR for the transformed variables.

Koskinen, however, found better coefficients for inverse square transformations by minimizing the SSR for the *untransformed variables*. For brevity, this method is referred to in this paper as minimizing the *Sum of the Squared Residuals for Untransformed Variables (SSRUV)*. This method gave an enormous improvement in the quality of the LRMs for the inverse square transformation.

Not already having access to a statistical analysis tool that could find LRMs for transformed variables by minimizing SSRUV rather than SSR, a short program was written for this paper that did so by a grid-based search with a small step size of 0.0001 over a wide potential range of the coefficients.[50]

[50] An optimized version of the grid-based SSRUV search algorithm that started with a coarse step size and iteratively searched smaller windows until the minimum step size of 0.0001 was reached took 0.108 CPU second to find Koskinen's coefficients for the mean range obtained for the simulations described here. For reference, simply opening

Koskinen's LRM for the mean range required to achieve full connectivity for the RGG configurations described above, as given in his thesis [Koskinen (2003a)], was

$$r = \frac{1}{\sqrt{0.3487\, n + 1.9379}} \quad . \tag{108}$$

For the simulated values of $n \leq 350$, MAPE is 0.96% and max error of the value of the predicted r is -2.53%.

Koskinen's LRM for the 99%ile range required to achieve full connectivity using the coefficients given in his thesis was

$$r = \frac{1}{\sqrt{0.1663\, n + 0.5754}} \quad . \tag{109}$$

For the simulated values of $n \leq 350$, MAPE is 1.41% and max error of the value of the predicted r is 2.90%.

His regression estimation errors were better than those achieved using the power law LRMs for the 15 values of n listed above. For the mean range required to achieve full connectivity based on the simulation results for $n \leq 350$, the power LRM was found to be

$$r = 1.1387\, n^{-0.4241} \quad . \tag{110}$$

For the simulated values of n, MAPE is 2.49% and max error of the value of the predicted r is 10.33%.

For the 99%ile range required to achieve full connectivity based on the simulation results for $n \leq 350$, the power law LRM was

$$r = 1.8748\, n^{-0.4516} \quad . \tag{111}$$

For the simulated values of n, MAPE is 2.01% and max error of the value of the predicted r is 6.55%.

a spreadsheet program required 9 CPU seconds on the same computer. The unoptimized version of the grid-based SSRUV search algorithm took 164 CPU seconds to find the same coefficients.

Figure 142 shows the relative error for the regression estimates for the mean range given by Koskinen's inverse square LRM (108) and the power law LRM (110).

Figure 143 shows the same for the 99%ile range given by Koskinen's (109) and the power law LRM (111).

For the mean range, Figure 142 demonstrates that Koskinen's inverse square LRM gives much better estimates for small numbers of nodes. For larger n, the inverse square and power law LRMs have roughly equal error but with opposite signs.

For the 99%ile range, Figure 143 shows that the inverse square LRM gives smaller errors than the estimate from the power law LRM.

Both LRMs, however, have small percentages of regression estimation error for the simulated values of n.

Given the successful results obtained by minimizing SSRUV for the inverse square LRMs, a natural step was to use the same program to find the coefficients for the power law LRMs by minimizing SSRUV instead of the SSR for the transformed variables. It was found, however, that the approach of minimizing SSRUV yielded power law LRMs that had much larger errors than those obtained by minimizing SSR and diverged even more as n increased. Thus, the usual approach of minimizing SSR based on the transformed variables gave better results for the power law LRMs.

Koskinen's approach was then applied to the results of the full set of simulations described in Section 5.1 that used the small-n set of numbers of nodes $n \leq 100{,}000$. His original expressions (108) and (109) that were derived for $n \leq 350$ were tested as well as new inverse square LRMs that were found by minimizing SSRUV based on the full set of simulations with $n \leq 100{,}000$.

The inverse square LRM for the mean range required to achieve full connectivity for the full set of simulations based on the small-n set was found to be

$$r = \frac{1}{\sqrt{0.3495\, n + 1.9611}} \quad . \tag{112}$$

For the simulated values of $n \leq 100{,}000$, MAPE is 7.56% and max error of the value of the predicted r is -19.86%.

The inverse square LRM for the 99%ile range required to achieve full connectivity obtained for the full set of simulations based on the small-n set was

$$r = \frac{1}{\sqrt{0.1708\, n + 0.5790}} \ . \tag{113}$$

For the simulated values of $n \leq 100{,}000$, MAPE is 3.58% and max error of the value of the predicted r is -9.86%.

The error values for the inverse square LRMs for the full set of simulations based on the small-n set are higher than those for the power law LRMs given in Section 5.2.1.1. For the mean range required to achieve full connectivity, the power law LRM (85) has MAPE 1.29% and max error 7.85%. For the 99%ile range required to achieve full connectivity, the power law LRM (84) has MAPE 1.87% and max error 6.25%.

Figure 144 shows plots of the errors for the mean ranges as estimated by Koskinen's inverse square LRM (108) obtained for the simulations for $n \leq 350$, as well as the inverse square LRM (112) and power law LRM (85) obtained for the simulations for $n \leq 100{,}000$. It is remarkable as to how the two inverse square LRMs overlap each other practically indistinguishably for the full data set, given that (108) was based on such relatively small values of n. Even so, the power law LRM has consistently smaller error than the inverse square LRMs for $n \geq 252$ and that gap widens greatly for the larger values of n.

Similarly, Figure 145 shows plots of the errors for the 99%ile ranges as estimated by Koskinen's inverse square LRM (109) obtained for the simulations for $n \leq 350$, as well as the inverse square LRM (113) and power law LRM (84) obtained for the simulations for $n \leq 100{,}000$. The errors for all three models are quite small. At this scale, it can be seen that the two inverse square LRMs track each other closely again for the full data set; it is surprising that (109) which was trained on small values of n has smaller error than (113) which was trained on the larger values. Again, the power law LRM has consistently smaller error than the inverse square LRMs for the larger values of n and the gap continues to widen.

The power law LRMs obtained for the smaller values of n are not shown in these plots because they did not perform at all well for the larger values of n.

It can be seen that Koskinen's LRMs give slightly smaller estimation errors for the ranges required for full connectivity for the smaller numbers of nodes but the power law LRMs provide closer and more consistent results for the larger numbers of nodes. Both models, however, have relatively small errors over the full set of the numbers of nodes modeled.

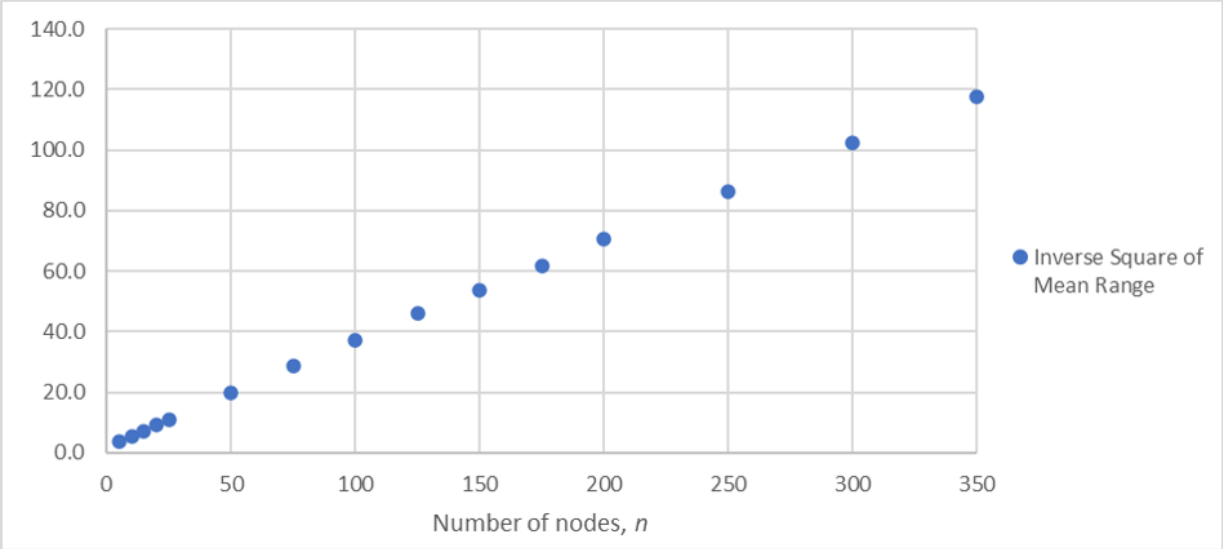

Figure 141: Inverse of the square of the mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth for each number of nodes in Koskinen's set of values of $n \leq 350$.

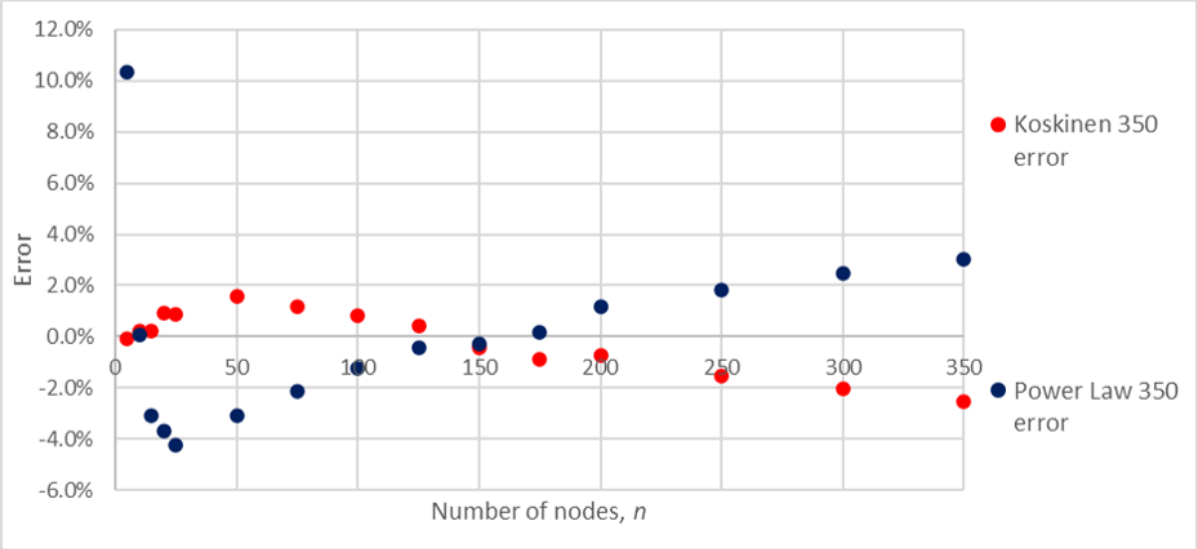

Figure 142: Relative errors for regression estimates for mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth given by Koskinen's inverse square LRM (108) and the power law LRM (110) based on Koskinen's set of values of $n \leq 350$.

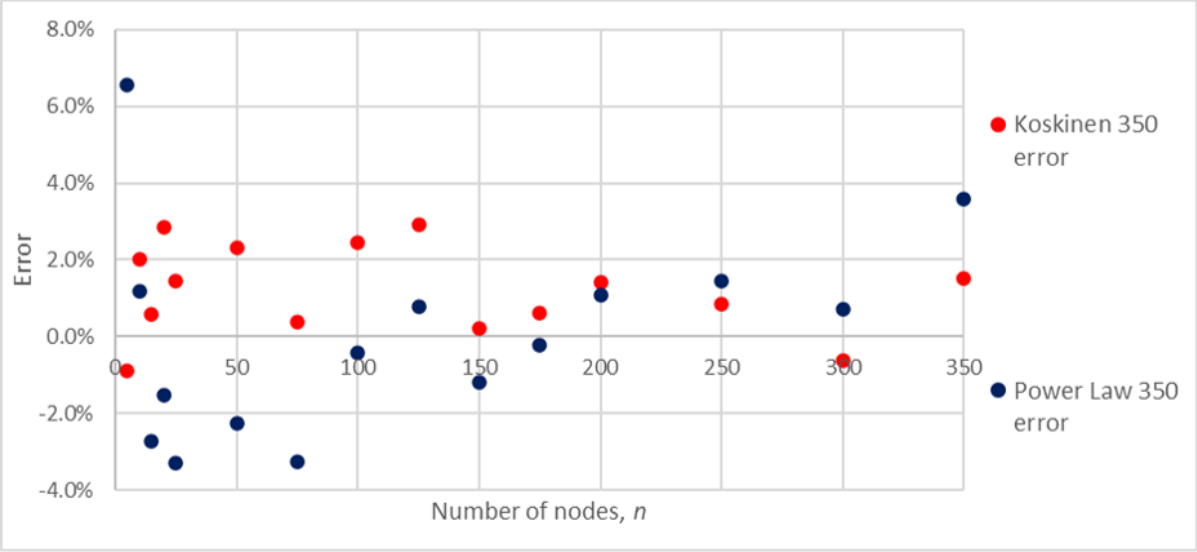

Figure 143: Relative errors for regression estimates for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth given by Koskinen's inverse square LRM (109) and the power law LRM (111) based on Koskinen's set of values of $n \leq 350$.

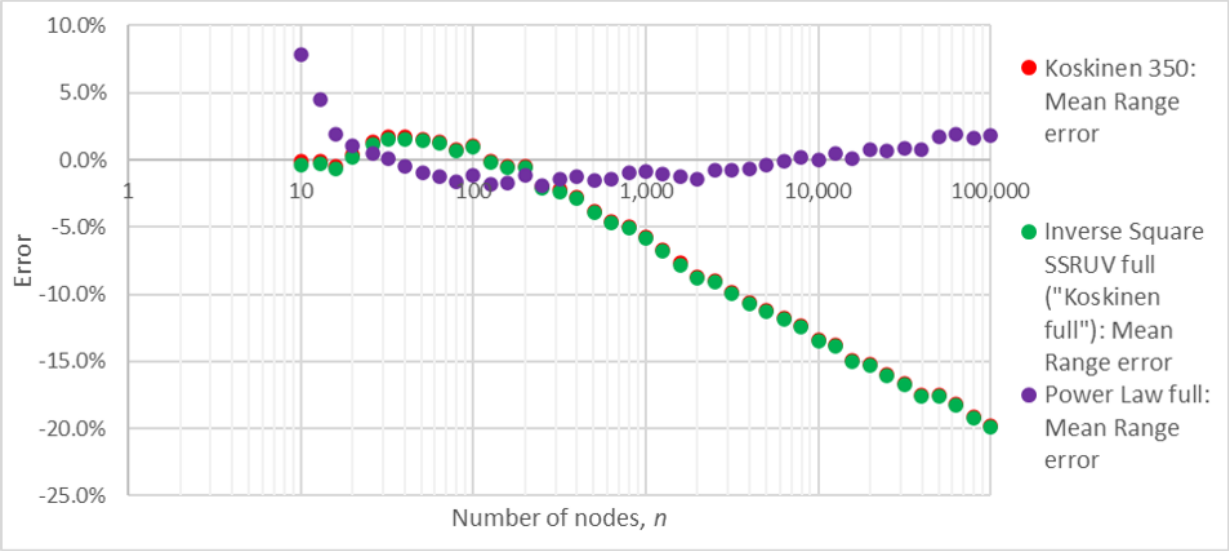

Figure 144: Relative errors for mean range for full connectivity of RGGs in a flat unit square with r-for-given-n growth for $n \leq 100{,}000$ for three models: Koskinen's inverse square LRM (108) for $n \leq 350$, and the inverse square LRM (112) and power law LRM (85) for $n \leq 100{,}000$.

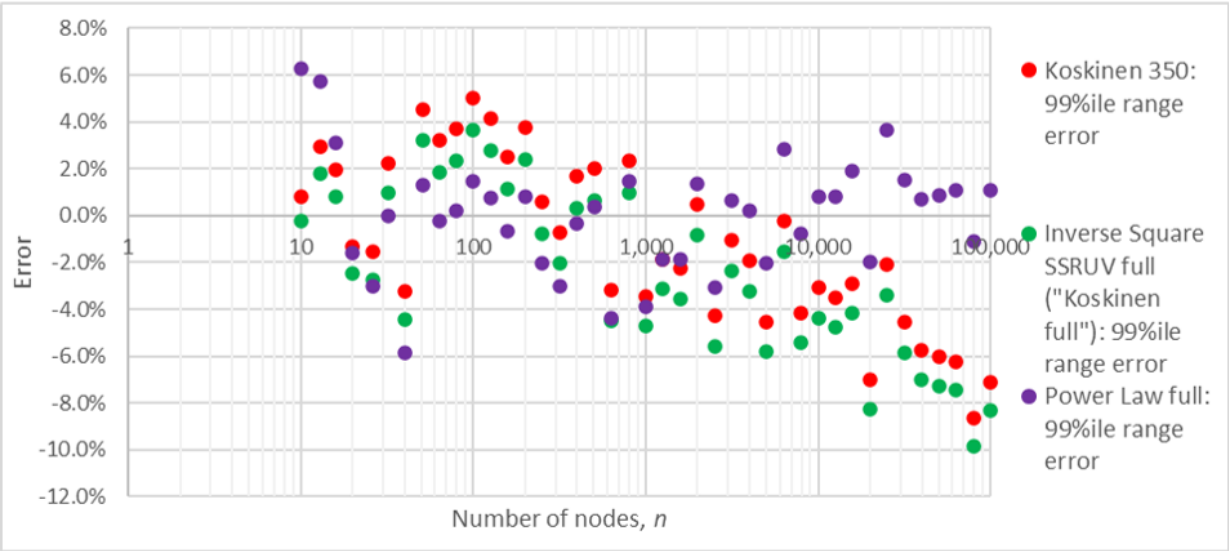

Figure 145: Relative errors for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth for $n \leq 100{,}000$ for three models: Koskinen's inverse square LRM (109) for $n \leq 350$, and the inverse square LRM (113) and power law LRM (84) for $n \leq 100{,}000$.

5.4.2 Testing Koskinen's Estimate of Penrose's c_P as a Function of the Number of Nodes

Penrose's estimate of the range required to achieve full connectivity with probability C as a function of the number of nodes n given by (15) and (16) was derived for toroidal unit squares and assumed "n large." The value c_P in (16) is independent of n and is determined only by C.

Koskinen [Koskinen (2004)], on the other hand, applied regression analysis to the results of the simulations outlined in Section 5.4.1 for numbers of nodes $n \leq 350$ to obtain functions of n to replace c_P in (16) to estimate the range required to achieve full connectivity of RGGs in flat (rather than toroidal) unit squares with r-for-given-n growth. By treating c_P as a function of n, the approach generates a function similar to that of Gupta and Kumar's $c_{GK}(n)$ but for probability C less than 1.

Koskinen defined a function $\alpha_C(n)$ to replace c_P in (16) for RGGs in a flat unit square:

$$r = \sqrt{\frac{\ln n + \alpha_C(n)}{\pi\, n}} \quad . \tag{114}$$

Note that the symbol C is used in this paper where Koskinen used the symbol q to denote the probability of being fully connected.

He proposed a regression model for $\alpha_C(n)$ of the form

$$\alpha_C(n) = c_1\, n^{-c_2} - c_3\, e^{-c_4\, n} + e^{-e^{-C}} \tag{115}$$

where c_1, c_2, c_3, and c_4 were coefficients obtained by regression for various values of C and the last term corresponds to Penrose's c_P given by (15). Thus, Koskinen obtained an adjustment to Penrose's c_P for RGGs in a flat unit square as a function of n that consists of a power law term and an exponential term.

For this paper, only the coefficients for $C = 0.99$ were tested. For $C = 0.99$, Koskinen found $c_1 = 12.2$, $c_2 = 0.0742$, $c_3 = 11.0$, and $c_4 = 0.122$, so his (115) became

$$\alpha_{0.99}(n) \quad = \quad 12.2\, n^{-0.0742} - 11.0\, e^{-0.122\, n} + e^{-e^{-0.99}} \quad . \qquad (116)$$

Figure 146 plots Koskinen's $\alpha_{0.99}(n)$ given by (116) as well as the observed 99%ile values of c_P for RGGs in a flat unit square with r-for-given-n growth for the small-n set. The curve for c_P shows a peak for the smaller values of n and then a falling off for the larger n; Koskinen modeled this shape very accurately with his $\alpha_{0.99}(n)$ using only the observed values of c_P for $n \leq 350$ that were available to him. It is remarkable that the regression coefficients derived for numbers of nodes $n \leq 350$ provide such good estimates of the values of $\alpha_{0.99}(n)$ for n up to 100,000.

The peak value of Koskinen's $\alpha_{0.99}(n)$ for these values of n is 15.8145 at $n = 32$. The exponential term falls below 5% of the term that is Penrose's $c_P = 4.6001$ for $C = 0.99$ at this value of n, and the exponential term effectively approaches zero quickly after that point. The power law term is 9.4336 at $n = 32$ which is double the value of c_P. The power law term remains dominant for all of the values of n simulated and it is 5.1923 or 13% greater than c_P for $n = 100{,}000$.

Figure 147 shows the relative error of the power law regression estimate for the 99%ile range given by the power law LRM (84) and the errors of the estimate provided by (114) that uses the values of $\alpha_{0.99}(n)$ provided by (116). It was noted in Section 5.2.1.1 that MAPE for the power law LRM (84) is 1.87% and max error of the predicted value of r is 6.25%. Expression (114) has MAPE 1.89% and max error 4.33%. The errors plotted in Figure 147 track closely for the most part, except for the smaller values of n where (114) provides closer estimates. Both sets of estimates have relatively small errors.

As was shown in Section 5.4.1 for LRMs for estimating range, Koskinen was successful in developing highly accurate regression models for an adjustment to Penrose's c_P to account for the growth of RGGs in flat unit squares. Impressively, even though his regression models were trained on simulation results for relatively small numbers of nodes they are accurate for far larger numbers of nodes as well.

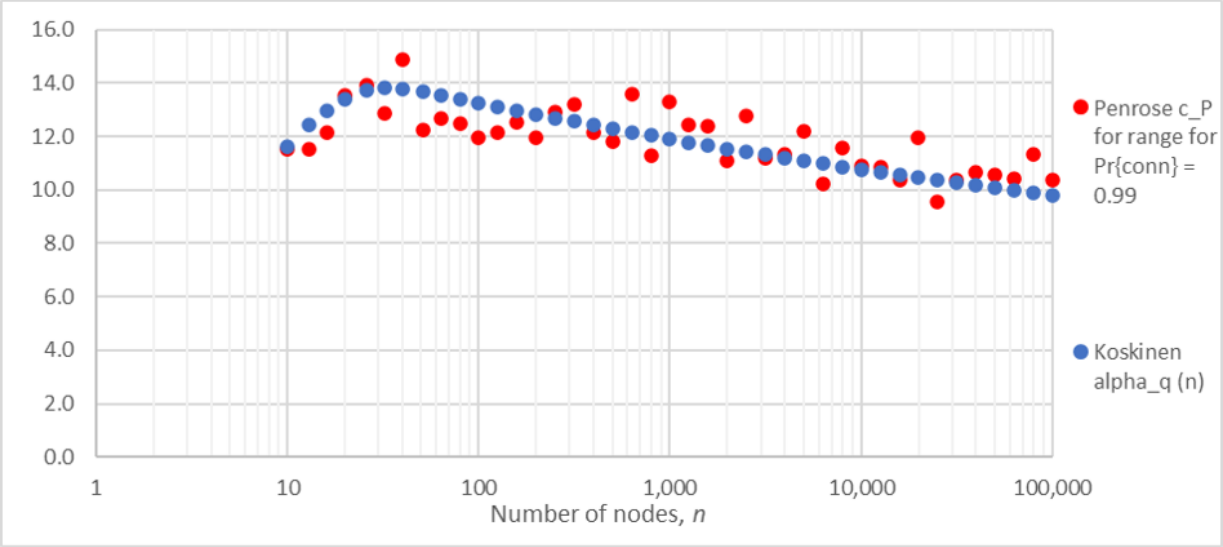

Figure 146: Values of c_P for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth, resulting from the sample of size $N = 1{,}000$ simulated geometries for each value of $n \leq 100{,}000$, as well as the values of $\alpha_{0.99}(n)$ given by the regression model (116).

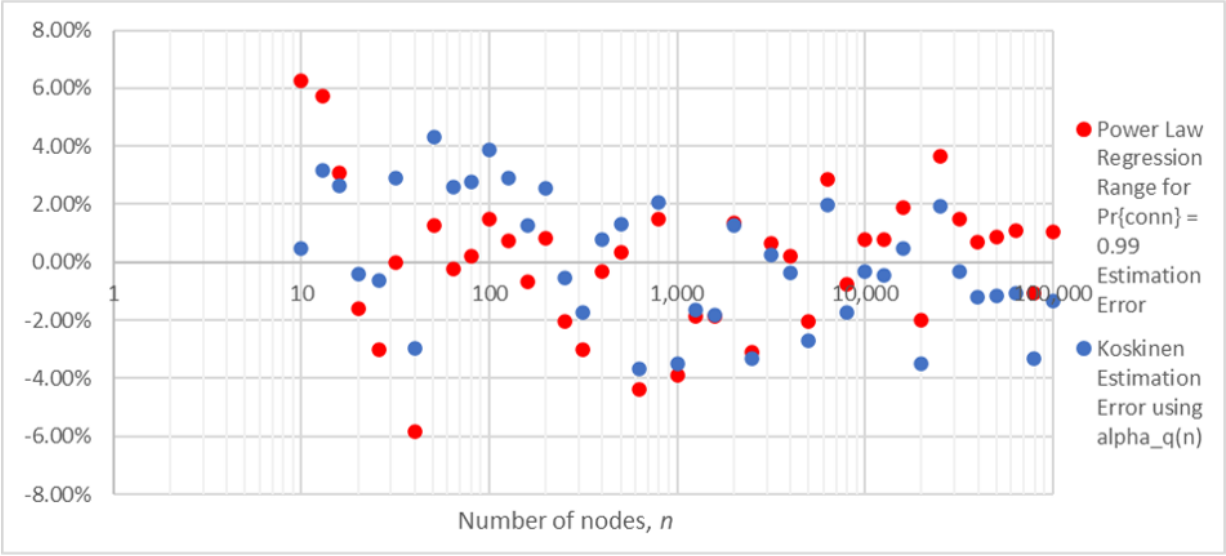

Figure 147: Relative errors for 99%ile range for full connectivity of RGGs in a flat unit square with r-for-given-n growth for $n \leq 100{,}000$ for two models. The models are the power law LRM (84) and Penrose's (16) with c_P replaced by Koskinen's $\alpha_{0.99}(n)$ given by (116).

5.5 Maximum Range Required for No Isolated Nodes in Flat Unit Squares and Unit Circles with r-for-given-n Growth

In Section 3.6 the worst-case ranges for no isolated nodes observed for a billion randomly generated RGGs in a unit circle were proposed as an approximation of the ranges that would achieve full connectivity "with probability 1." These ranges provided tighter bounds on the possible values of Gupta and Kumar's $c_{GK}(n)$ than those obtained by using the true worst-case range required which is the diameter of the unit circle.

This section explores some properties of the worst-case ranges observed for large sets of trials and compares them to the 99% values of the range for RGGs in a flat unit square with r-for-given-n growth obtained in Section 5.2.1.1.

For a set of trials with sample size N, where in each trial n nodes are randomly placed in a space to represent an RGG, let $r_{max}(n, N)$ denote the observed worst-case range required for that set. That value is obtained as follows. For the RGG generated in each of the N trials with a given n, the distance of every node to its nearest neighbor was obtained, and the maximum of those distances equaled the minimum range that would achieve no isolated nodes in that RGG. For a set of N trials with a given n, the maximum of the N minimum ranges obtained in that way is taken to be $r_{max}(n, N)$.

As noted in Section 3.6, the term *trial* is used because these were not full simulations of RGG growth but only measurements made of the relative random placements of the nodes to determine the largest range that would be needed to guarantee no isolated nodes in any of the N RGGs in a set of trials.

The numbers of nodes used were the 11 smallest values in the small-n set described in Section 5.1, which are $n = 10$, 13, 16, 20, 26, 32, 40, 51, 64, 80, 100. The sample sizes N were the eight powers of ten from 10^2 to 10^9.

The observed values of $r_{max}(n, N)$ would be expected to vary over multiple sets of trials but even so it was found that these values could be estimated using linear regression. Such an estimate of the worst-case range required for no isolated nodes could be useful as a precondition for providing very high assurance of a path between any pairs of nodes in an RGG.

Plots are omitted but it was observed that, for a given value of N, a good fit for $r_{max}(n, N)$ was a linear function of the log of n. The linearity of a log-log transformation made the power law a good fit but the linear-log transformation which yielded a log law LRM provided slightly better results and had the benefit of being a simpler model.[51]

For a flat unit square and sample size $N = 10^9$, the LRM based on the log transformation of n was found to be

$$r_{max}(n, 10^9) = -0.2612 \ln n + 1.6593 \tag{117}$$

with an R^2 value of 0.9949. For the simulated values of n, MAPE is 1.48% and max error of the value of the predicted maximum range that would be observed is −2.96%.

[51] The log-log and linear-log transformations give practically the same results in this instance because the log function does not differ very much from a straight line for the values of the range considered in this section. For values of r from 1.1 down to 0.2298 decreasing geometrically with the multiplier $10^{-1/50}$ so there are 50 sampled values of r per decade (for a total of 35 values of r), the log-log LRM has an R^2 value of 0.9594. MAPE and max relative error do not apply here because of the zero crossing of the log function. The mean absolute error is 0.0810 and the root mean squared absolute error is 0.0938. A plot of those values would show only a slight bow of the log function around a straight line.

Note, on the other hand, that the ranges obtained for the r-for-given-n growth models described in Section 5.2.1.1 would be modeled very poorly using log law LRMs. If Figure 109 were plotted on linear-log axes instead of log-log, the values would not at all closely approximate a straight line.

For a unit circle and sample size $N = 10^9$, the LRM based on the log transformation of n was found to be

$$r_{max}(n, 10^9) = -0.2632 \ln n + 1.5800 \tag{118}$$

with an R^2 value of 0.9912. For the simulated values of n, MAPE is 2.57% and max error of the value of the predicted maximum range that would be observed is −9.55%.

It was observed further that, for a given value of n, a good fit for $r_{max}(n, N)$ was also a linear function of the log of N. Based on that indication, multiple regression was used to find a fit for $r_{max}(n, N)$ as a linear function of the log of n and log of N.

For a flat unit square and sample size N, the LRM based on the log transformation of n and N was found to be

$$r_{max}(n, N) = -0.2381 \ln n + 0.0202 \ln N + 1.1808 \tag{119}$$

with an R^2 value of 0.9685. For the simulated values of n, MAPE is 5.23% and max error of the value of the predicted maximum range that would be observed is −26.35%.

For a unit circle and sample size N, the LRM based on the log transformation of n and N was found to be

$$r_{max}(n, N) = -0.2274 \ln n + 0.0172 \ln N + 1.1154 \tag{120}$$

with an R^2 value of 0.9642. For the simulated values of n, MAPE is 6.09% and max error of the value of the predicted maximum range that would be observed is −32.88%.

These LRMs provide only rough estimates of what the observed $r_{max}(n, N)$ would be. They are valid within only the extents of the values of n and N specified above and they break down outside these extents.

Figure 148 shows the observed values for $r_{max}(n, N)$ for no isolated nodes in a flat unit square for $n = 10$ to 100 and $N = 10^2$, 10^3, 10^6, and

10^9 as outlined above. The 99%ile range required for no isolated nodes as shown Figure 109 in Section 5.2.1.1 is plotted for comparison. Lines between points were added solely for readability.

The plot shows that the 99%ile range required for no isolated nodes and $r_{max}(n, 100)$ track closely as might be expected. The values of $r_{max}(n, N)$ increase with N and illustrate that they are roughly linear functions of the log of n.

For sample size $N = 10^9$, the maximum ratio of the observed $r_{max}(n, 10^9)$ to the 99%ile range was 2.13 with the mean value being 1.91 over the values of n. For the other values of N, the next largest ratios were 2.00 and 1.94.

For the small values of n used in this section, a reasonable multiplier of the 99%ile range required for no isolated nodes to estimate the range that would guarantee no isolated nodes for all practical purposes is 2. Such a range would yield a working approximation of that needed to achieve full connectivity "with probability 1."

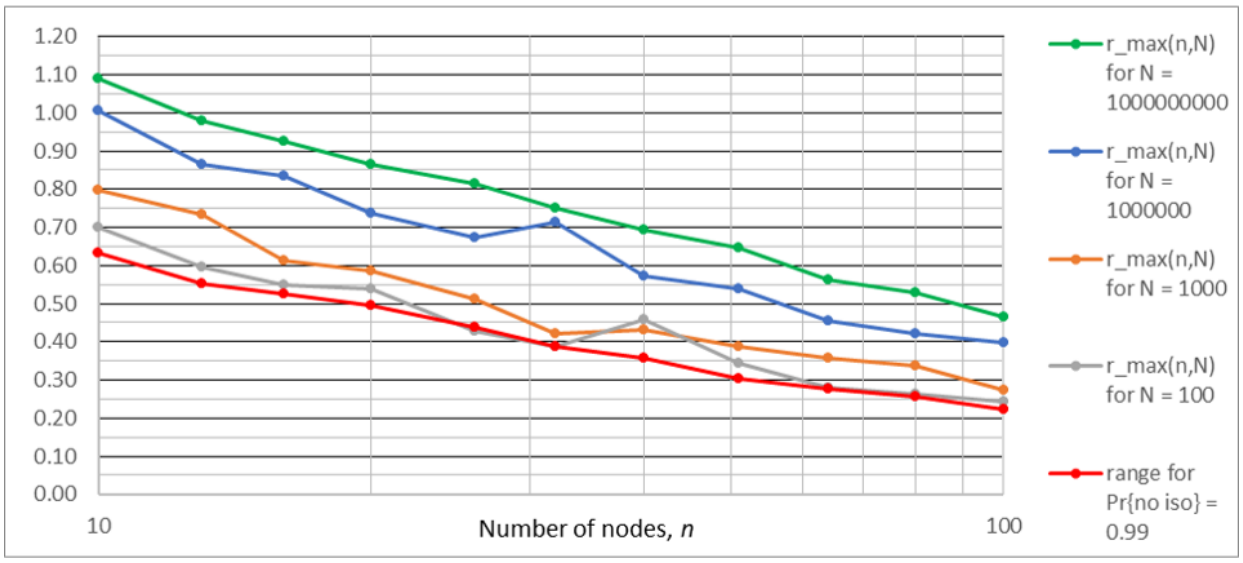

Figure 148: Ranges required for no isolated nodes in a flat unit square for $n = 10$ to 100. Plot shows 99%ile range required for no isolated nodes as given in Figure 109. Plot shows observed $r_{max}(n, N)$ for values of $N = 10^2$, 10^3, 10^6, 10^9. Lines between points were added for readability.

6 Inverting n-for-given-r Linear Regression Models to Estimate r-for-given-n Growth

The evolutionary processes for RGGs in the n-for-given-r and r-for-given-n growth models are quite different but conceptually they are inverses of each other. Only the "n-for-given-r"-to-"r-for-given-n" problem is considered here. This section explores the degree to which inverses of the LRMs for the n-for-given-r growth model based on the mid-r set can predict the behavior of the r-for-given-n growth model based on the small-n set.

Suppose that a set of output values y depends on a set of input values x, where the x and y values are already transformed so that y shows a linear relationship with x. Let $\widehat{y} = f(x)$ be the LRM obtained for these data, where $\widehat{y}$ is the estimate of the observed value of y for a given value of x. The general problem addressed here is framed as obtaining a function such that a value of the output y could be used to predict what input value of x (thus, an estimated $\widehat{x}$) would produce that value y.

In the context of this paper, the specific problem is to obtain an r-for-given-n LRM that yields an estimate of r as a function of n from an n-for-given-r LRM. Two approaches to this problem discussed by Parker et al. [Parker (2010)] are tested in this paper.

The first approach is *inverse regression*, which is defined as inverting the LRM function $f(\cdot)$ so that $\widehat{x} = f^{-1}(y)$. The second approach is *reverse regression*, which is defined as finding a new LRM $h(\cdot)$ by treating x as the output variable and y as the input variable to obtain $\widehat{x} = h(y)$. The two methods are not equivalent and would not be expected to produce the same function; that is, in general, for a given regression one would expect that $f^{-1}(y) \neq h(y)$.

One can always invert an LRM $f(\cdot)$ or perform another regression to obtain a new LRM $h(\cdot)$ but whether or not the resulting functions would give good results depends on the specific distribution of errors in the output variable y for that situation. As of 2010, Parker et al. described the statistical literature in this area as "somewhat underdeveloped." The best way to determine whether inverse regression or reverse regression works in a given instance is to apply those methods and test the results.

In this section only the process of inverse regression to obtain an r-for-given-n LRM from an n-for-given-r LRM is described in detail. Some results are summarized in Section 6.5 where the performance of reverse regression LRMs are compared to the performance of the inverse regression LRMs.

The hypothesis that an LRM can be inverted and apply to a different RGG growth model is tested for this situation by inverting the n-for-given-r power law expression (21) to obtain an r-for-given-n expression as

$$r = (\frac{1}{b})^{\frac{1}{m}} \, n^{\frac{1}{m}} \quad . \tag{121}$$

Graphically, the hypothesis that LRMs can be inverted is tested as shown in Figure 149, Figure 150, Figure 151, and Figure 152. These figures plot the n-for-given-r growth model's values for 99%ile n as a function of r and they also plot the r-for-given-n growth model's values *inversely* with the specified values of n shown as a function of the 99%ile values of r.

There is no significance to the sizes of the points plotted in these figures. The different point sizes were used solely for readability.

Specific details of each of these plots are given in their applicable sections that follow.

The values in the mid-r set used to obtain the n-for-given-r model and the 99%ile values of r resulting from the r-for-given-n model based on the small-n set do not entirely overlap. The largest values of r used for n-for-given-r growth are smaller than the smallest values of r needed for the values of n used for r-for-given-n growth. Because of this, the LRMs for the two growth models were trained over slightly different intervals of r and n.

In the sections that follow, the LRMs for the 99%ile values of n in the n-for-given-r growth model are inverted using (121) to obtain an inverse LRM for r-for-given-n growth. That expression is used to generate a set of predicted 99%ile values of r for the small-n set. MAPE and max error values are obtained with respect to the observed 99%ile ranges for those numbers of nodes.

6.1 Flat Unit Square: Inverse Regression for Full Connectivity

Figure 149 shows the 99%ile number of nodes required to first achieve full connectivity of RGGs in a flat unit square with n-for-given-r growth for the mid-r set as shown in Figure 26 in Section 4.2.1.1. Figure 149 also shows the 99%ile range required to achieve full connectivity of RGGs in a flat unit square with r-for-given-n growth for the small-n set as shown in Figure 109 in Section 5.2.1.1 but, unlike Figure 109, the points are plotted inversely as if n were the dependent variable and r were the independent variable.

It can be seen that the r-for-given-n points fall only slightly higher than the trend for the n-for-given-r points, which indicates that the inverse of the n-for-given-r LRM potentially can provide a close estimate of the r-for-given-n points.

For a flat unit square and full connectivity, (22) for the 99%ile number of nodes is inverted using (121) to obtain the inverse LRM for r-for-given-n growth as

$$r = 2.1043\, n^{-0.4836} \quad . \tag{122}$$

When (122) is used to predict the 99%ile values of the range for full connectivity for the small-n set and compared to the observed 99%ile values of r, MAPE is 3.72% and max error is −8.03%. As noted above, the n-for-given-r and r-for-given-n LRMs are trained over slightly different intervals of r and n.

For comparison, it is shown in Section 5.2.1.1 that MAPE is 1.87% and max error of the predicted value of r is 6.25% for the r-for-given-n LRM (84) over the small-n set.

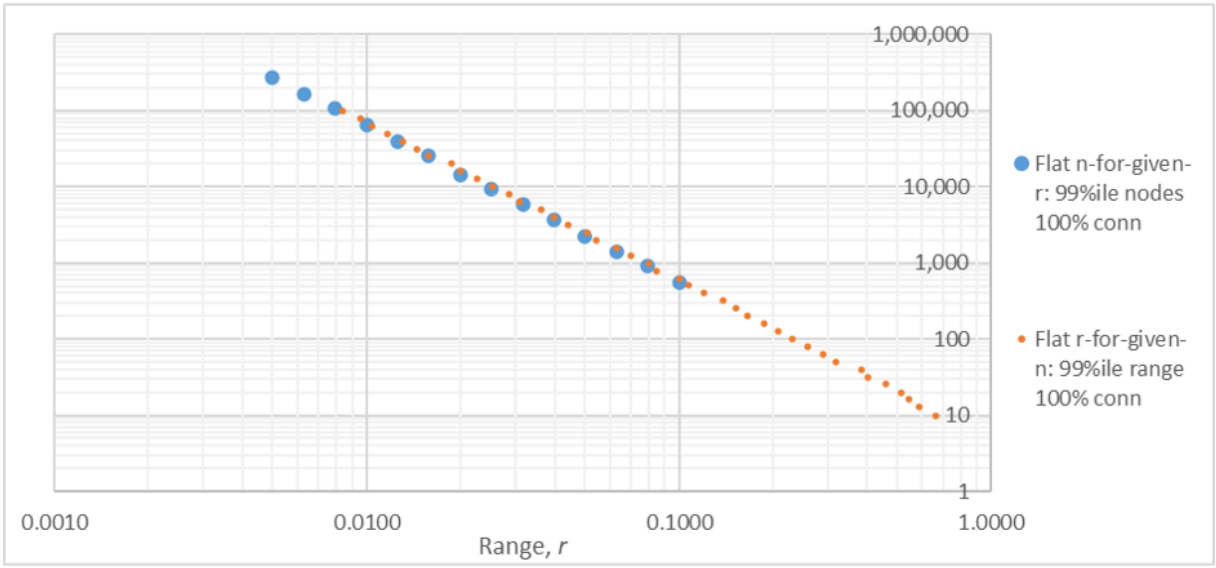

Figure 149: The 99%ile number of nodes for full connectivity of RGGs in a flat unit square with n-for-given-r growth as a function of range and the range for full connectivity of RGGs in a flat unit square with r-for-given-n growth as a function of the number of nodes but plotted inversely as if the number of nodes were a function of the range. Points are plotted as different sizes for readability.

6.2 Flat Unit Square: Inverse Regression for No Isolated Nodes

Figure 150 shows the 99%ile number of nodes required to first achieve no isolated nodes in RGGs in a flat unit square with n-for-given-r growth for the mid-r set as shown in Figure 29 in Section 4.2.1.4. Figure 150 also shows the 99%ile range required to achieve no isolated nodes in RGGs in a flat unit square with r-for-given-n growth for the small-n set as shown in Figure 110 in Section 5.2.1.1 but, unlike Figure 110, the points are plotted inversely as if n were the dependent variable and r were the independent variable.

As was the case for the condition of full connectivity, it can be seen that the r-for-given-n points in the graph for no isolated nodes fall only slightly higher than the trend for the n-for-given-r points, which again indicates that the inverse of the n-for-given-r LRM might provide a close estimate of the r-for-given-n points.

For a flat unit square and no isolated nodes, (31) for the 99%ile number of nodes is inverted using (121) to obtain the inverse LRM for r-for-given-n growth as

$$r = 1.9152\, n^{-0.4769} \quad . \tag{123}$$

When (123) is used to predict the 99%ile values of the range required to achieve no isolated nodes for the small-n set and compared to the observed 99%ile values of r, MAPE is 5.19% and max error is −10.23%.

For comparison, it is shown in Section 5.2.1.1 that MAPE is 1.93% and max error of the predicted value of r is 7.60% for the r-for-given-n LRM (86) over the small-n set.

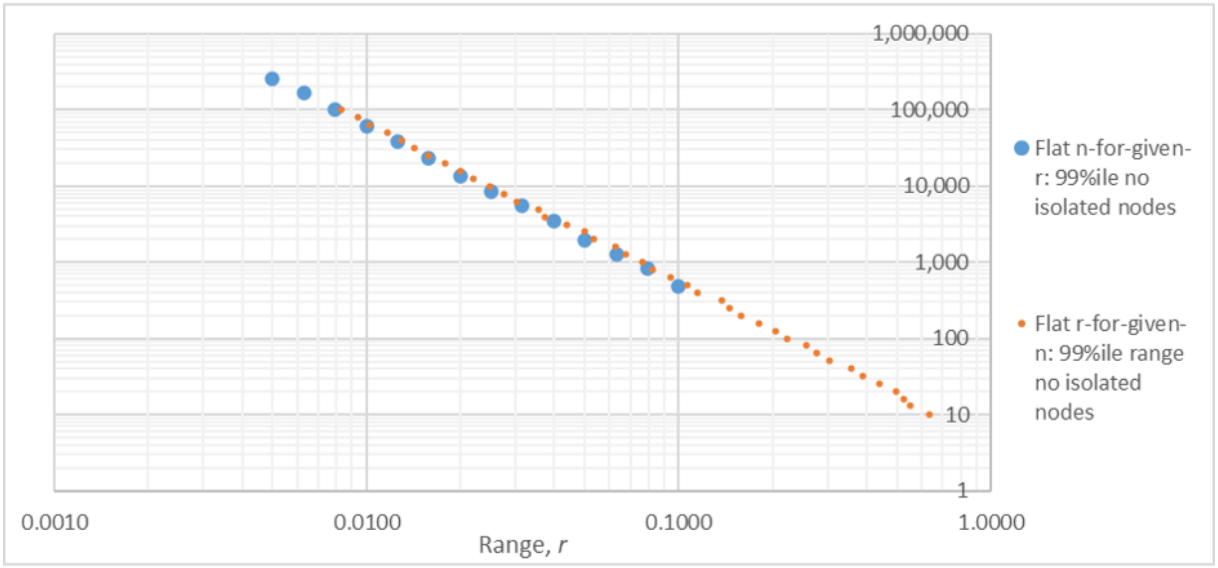

Figure 150: The 99%ile number of nodes for no isolated nodes in RGGs in a flat unit square with n-for-given-r growth as a function of range and the range for no isolated nodes in RGGs in a flat unit square with r-for-given-n growth as a function of the number of nodes but plotted inversely as if the number of nodes were a function of the range. Points are plotted as different sizes for readability.

6.3 Toroidal Unit Square: Inverse Regression for Full Connectivity

Figure 151 shows the 99%ile number of nodes required to first achieve full connectivity of RGGs in a toroidal unit square with n-for-given-r growth for the mid-r set as shown in Figure 64 in Section 4.3.1.1. Figure 151 also shows the 99%ile range required to achieve full connectivity of RGGs in a toroidal unit square with r-for-given-n growth for the small-n set as shown in Figure 125 in Section 5.3.1.1 but, unlike Figure 125, the points are plotted inversely as if n were the dependent variable and r were the independent variable.

It can be seen that the r-for-given-n points for full connectivity lie closely on the trend for the n-for-given-r points, which indicates that the inverse of the n-for-given-r LRM may provide an even closer estimate of the r-for-given-n points for RGGs in a toroidal unit square than was the case for a flat unit square.

For RGGs in a toroidal unit square and full connectivity, (52) for the 99%ile number of nodes is inverted using (121) to obtain the inverse LRM for r-for-given-n growth as

$$r = 1.4395\, n^{-0.4606} \quad . \tag{124}$$

When (124) is used to predict the 99%ile values of the range required to achieve no isolated nodes for the small-n set and compared to the observed 99%ile values of r, MAPE is 3.32% and max error is 21.05%. The largest errors occur at small values of n; when the error percentages are restricted to include only those for $n \geq 100$, MAPE is 1.57% and max error is −3.73%.

For comparison, it is shown in Section 5.3.1.1 that MAPE is 3.01% and max error of the predicted value of r is 13.71% for the r-for-given-n LRM (98) over the small-n set. In Section 5.3.1.2 it is shown that MAPE for Penrose's Theorem is 2.19% and max error of its predicted value of r is 13.84%.

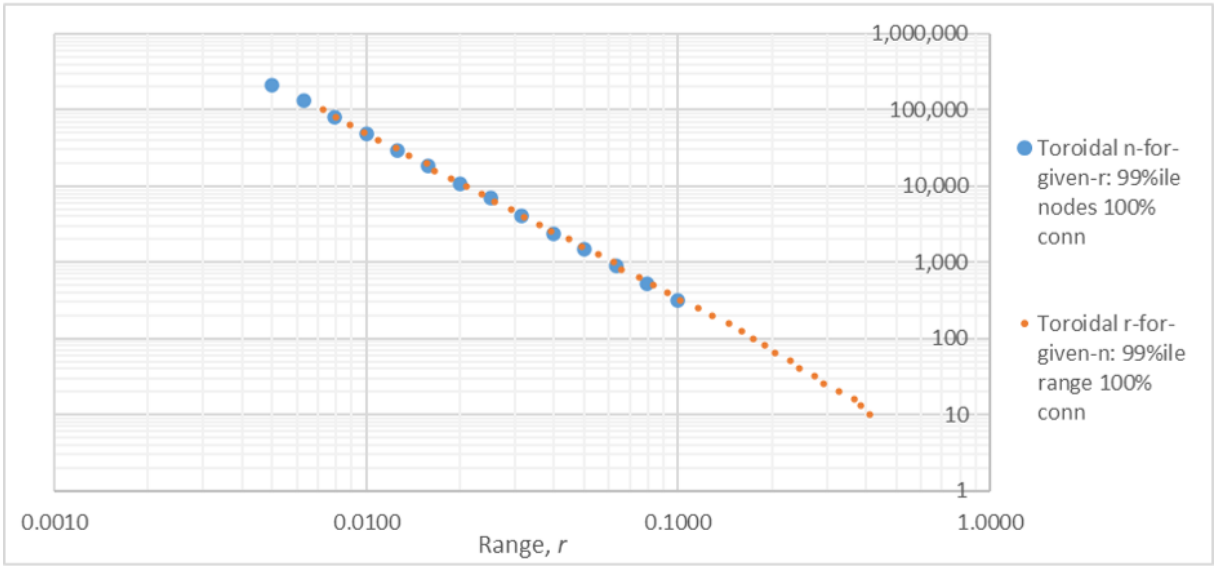

Figure 151: The 99%ile number of nodes for full connectivity of RGGs in a toroidal unit square with n-for-given-r growth as a function of range and the range for full connectivity of RGGs in a toroidal unit square with r-for-given-n growth as a function of the number of nodes but plotted inversely as if the number of nodes were a function of the range. Points are plotted as different sizes for readability.

6.4 Toroidal Unit Square: Inverse Regression for No Isolated Nodes

Figure 152 shows the 99%ile number of nodes required to first achieve no isolated nodes in RGGs in a toroidal unit square with n-for-given-r growth for the mid-r set as shown in Figure 68 in Section 4.3.1.3. Figure 152 also shows the 99%ile range required to achieve no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth for the small-n set as shown in Figure 126 in Section 5.3.1.1 but, unlike Figure 126, the points are plotted inversely as if n were the dependent variable and r were the independent variable.

As was the case for full connectivity, it can be seen that the r-for-given-n points in the graph for no isolated nodes lie closely on the trend for the n-for-given-r points, which again indicates that the inverse of the n-for-given-r LRM may provide an even closer estimate of the r-for-given-n points for RGGs in a toroidal unit square than was the case for a flat unit square.

For RGGs in a toroidal unit square and no isolated nodes, (58) for the 99%ile number of nodes is inverted using (121) to obtain the inverse LRM for r-for-given-n growth as

$$r = 1.3817\, n^{-0.4575} \quad . \tag{125}$$

When (125) is used to predict the 99%ile values of the range required to achieve no isolated nodes for the small-n set and compared to the observed 99%ile values of r, MAPE is 3.17% and max error is 18.32%. The largest errors occur at small values of n; when the error percentages are restricted to include only those for $n \geq 100$, MAPE is 1.83% and max error is −4.81%.

For comparison, it is shown in Section 5.3.1.1 that MAPE is 2.85% and max error of the predicted value of r is 12.54% for the r-for-given-n LRM (98) over the small-n set. In Section 5.3.1.2 it is shown that MAPE for Penrose's Theorem is 2.66% and max error of its predicted value of r is 15.11%.

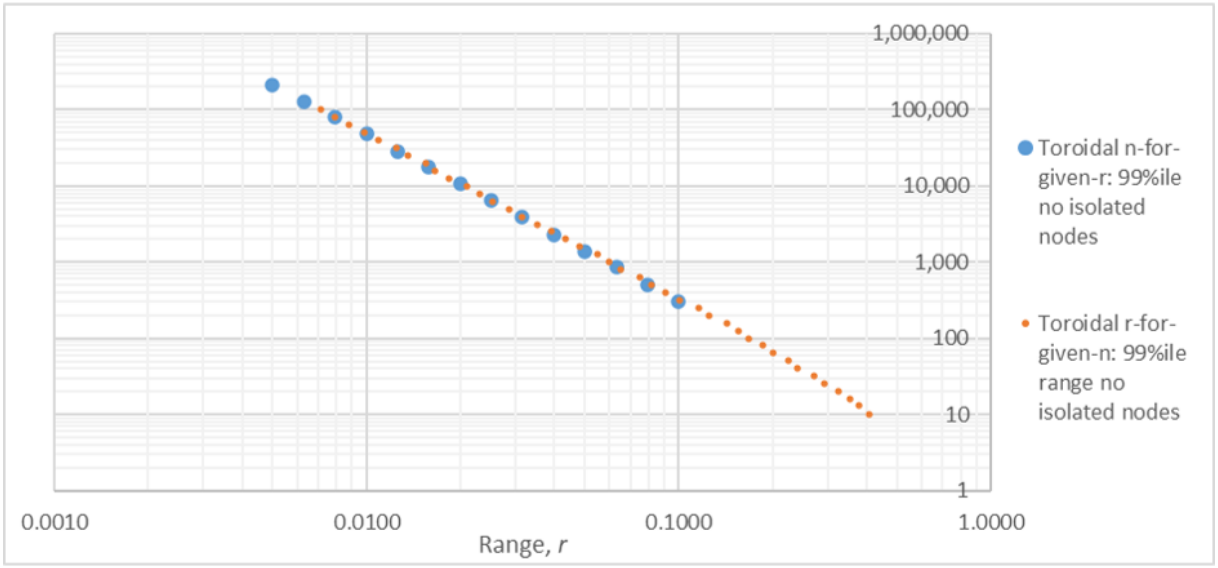

Figure 152: The 99%ile number of nodes for no isolated nodes in RGGs in a toroidal unit square with n-for-given-r growth as a function of range and the range for no isolated nodes in RGGs in a toroidal unit square with r-for-given-n growth as a function of the number of nodes but plotted inversely as if the number of nodes were a function of the range. Points are plotted as different sizes for readability.

6.5 Observations on Inverse Regression and Reverse Regression to Estimate r-for-given-n Growth

Section 6.1, Section 6.2, Section 6.3, and Section 6.4 demonstrate that inverting the 99%ile n-for-given-r LRM to obtain a 99%ile r-for-given-n LRM produces expressions that have MAPE and max relative errors that are quite close to those of the r-for-given-n LRM that is derived directly from the simulation results for the small-n set of numbers of nodes $10 \leq n \leq 100{,}000$.

Section 6.1 and Section 6.2 show for RGGs in a flat unit square that the inverse regression LRM for r-for-given-n growth has slightly larger errors than the r-for-given-n LRM derived directly from the simulation results for both full connectivity and no isolated nodes for the numbers of nodes $10 \leq n \leq 100{,}000$.

Section 6.3 and Section 6.4 show for RGGs in a toroidal unit square that the same is true but the inverse regression LRM for r-for-given-n growth has smaller errors for $n \geq 100$ than the r-for-given-n LRM derived directly from the simulation results for $10 \leq n \leq 100{,}000$, and this holds for both full connectivity and no isolated nodes.

Reverse regression was also tested. These results are summed up very briefly because reverse regression was found to produce virtually the same results as those obtained using inverse regression. The coefficients for the reverse regression LRMs for r-for-given-n growth obtained from the simulation data for the n-for-given-r growth model by treating the ranges as the output variable and the observed numbers of nodes as the input variable differed from the inverse regression LRMs only in the third or fourth place after the decimal point. The MAPE and max errors differed not at all or only in the second place after the decimal point.

These results indicate that RGGs with the same range and number of nodes tend to have similar "99%ile" properties regardless of whether the growth model involves adding nodes or increasing range to achieve full connectivity or no isolated nodes.

Under the hypothesis that statistical metrics for RGGs with r-for-given-n growth can be approximated by inverting or reversing those obtained for r-for-given-n growth, expressions can be given for the 99%ile range required for full connectivity of RGGs in a unit circle; LRMs for RGGs in a unit circle were not obtained in Section 5 as they were for flat and toroidal unit squares.

For a unit circle and full connectivity, (73) for the 99%ile number of nodes is inverted using (121) to obtain an inverse LRM for r-for-given-n growth as

$$r = 1.8749\, n^{-0.4753} \tag{126}$$

and reverse regression performed on the data from which (73) was obtained yields a reverse LRM for r-for-given-n growth as

$$r = 1.8737\, n^{-0.4752} \quad . \tag{127}$$

The inverse and reverse LRMs (126) and (127) are seen to be practically identical. Without simulations of r-for-given-n growth of RGGs in unit circles the MAPE and max error values for (126) and (127) are unknown at this time.

7 RGGs in d-Dimensional Spaces

Much of the RGG literature deals with two-dimensional nodal spaces but some prior researchers have also addressed one-dimensional spaces where the nodes of an RGG lie in a line as well as spaces of dimensions three and greater. In particular, work by M.D. Penrose [Penrose (1999)] [Penrose (2003)] addressed the broad problem of the connectivity of RGGs in toroidal unit squares in nodal spaces of arbitrary dimensions $d \geq 2$. Furthermore, Penrose considered k-connectivity which involved not only single but multiple paths between nodes in an RGG. His work was of such fundamental and general nature that the expressions collectively referred to as "Penrose's Theorem" in Section 3.5 of this paper were stated as the special case of 1-connectivity and $d = 2$, which is to say simple connectivity in two-dimensional space.

Concrete realizations of RGGs are probably limited to 1D, 2D, and 3D applications but, as noted in the Section 1.1 of the Introduction, RGGs also have the potential to be used to represent relationships in machine learning data sets in very high dimensionality spaces [Avrachenkov (2020)] [Erba (2020)]. This provides the motivation to understand the connectivity properties of RGGs even in spaces that are not physically realizable.

For the n-for-given-r growth model, Section 7.1 explores the connectivity of RGGs in flat and toroidal unit squares and obtains LRMs for the conditions of 100% connectivity and no isolated nodes for spaces of dimensions $1 \leq d \leq 4$. For the r-for-given-n growth model, Section 7.2 tests Penrose's Theorem for spaces of arbitrary dimension d against an expression for the probability of no isolated nodes as well as against simulations of RGGs in flat and toroidal unit squares for spaces of dimensions $1 \leq d \leq 6$.

7.1 RGGs in d-Dimensional Spaces with n-for-given-r Growth

The approach for estimating RGG connectivity in much of the previous literature has been based on relating the condition of being fully connected to the absorption of isolated nodes. As was shown in Section 4.2 and Section 4.3 and their subsections, RGGs with n-for-given-r growth in two-dimensional spaces in flat unit squares require proportionately more nodes to be added beyond the point where the last isolated node disappears to first achieve full connectivity than are required in toroidal unit squares to achieve full connectivity. This was demonstrated in Figure 33 and Figure 73 which showed the fraction of simulated geometries where full connectivity was first achieved when the last isolated node disappeared and in Figure 34 and Figure 74 which gave the mean ratios of the number of nodes required to first achieve those two conditions. Similar plots are given in this section.

This section presents LRMs for RGGs in flat and toroidal unit squares with n-for-given-r growth in spaces of dimensions $d = 1$, 2, 3, and 4. Only the 99%ile number of nodes statistical metric is considered for the conditions of full connectivity and no isolated nodes. The introductory material presented in Section 4 and its subsections regarding the simulations and their conditions applies here. The results for $d = 1$, 2, and 3 are based on $N = 1{,}000$ simulated geometries for the ranges in the mid-r set as defined in Section 4.1.2 which is the set of ranges from 0.1000 down to 0.0050. Due to computational constraints the results for $d = 4$ are based on a subset of the ranges in the mid-r set consisting of $r = 0.1000$ down to 0.0126.

The results for two-dimensional spaces presented in this section are based on a set of simulations independent from those used in Section 4.2 and Section 4.3 and are included as yet another assessment of the noise in the results of the simulations.

7.1.1 One-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 1$ consist of nodes in a line of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with n-for-given-r growth for the mid-r set.

The plots of the 99%ile number of nodes are not shown in this section but are included in Section 7.1.5 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 6.6334\, r^{-1.1189} \tag{128}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 1.11% and max error of the value of the predicted n is 3.22%.

For a flat space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 3.4068\, r^{-1.0981} \tag{129}$$

with an R^2 value of 0.9988. For the simulated values of r, MAPE is 3.09% and max error of the value of the predicted n is 5.61%.

For a toroidal space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, the LRM is

$$n = 3.7563\, r^{-1.1793} \tag{130}$$

with an R^2 value of 0.9991. For the simulated values of r, MAPE is 2.77% and max error of the value of the predicted n is 5.89%.

For a toroidal space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 2.5993\, r^{-1.1418} \tag{131}$$

with an R^2 value of 0.9988. For the simulated values of r, MAPE is 2.68% and max error of the value of the predicted n is 8.83%.

For one-dimensional flat and toroidal unit squares, Figure 153 shows the fraction of simulated geometries where full connectivity is achieved at the same point where the last isolated node disappears. Interpreting the fractions shown in Figure 153 as probabilities, it can be seen that there is only a very small probability that an RGG in a flat or toroidal unit square in 1D is fully connected at the same point where the last isolated node disappears for the largest ranges tested, and those probabilities fall to nearly zero for smaller ranges.

Figure 154 shows the mean ratios of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes first disappear in flat and toroidal unit squares in one-dimensional spaces for RGGs with n-for-given-r growth. The mean is taken over all of the ratios for a given value of the range for each simulated geometry. Figure 154 shows that on average, for RGGs in a flat unit square in 1D, a simulated geometry requires 60% more nodes to first achieve full connectivity than for the last isolated node to disappear for the largest range simulated and that ratio increases to well over double as the range decreases. For RGGs in a toroidal unit square, the simulated geometries require about 20% more nodes for full connectivity for the largest range simulated and the ratio increases to slightly more than double as the range decreases.

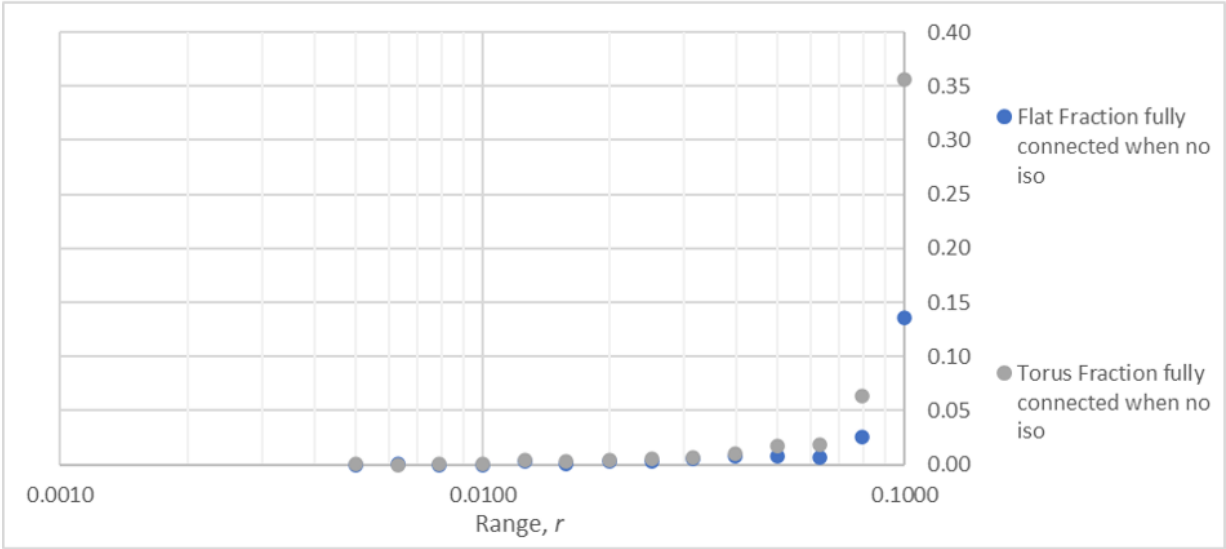

Figure 153: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 1D (i.e., simply flat and toroidal lines) to become fully connected with n-for-given-r growth.

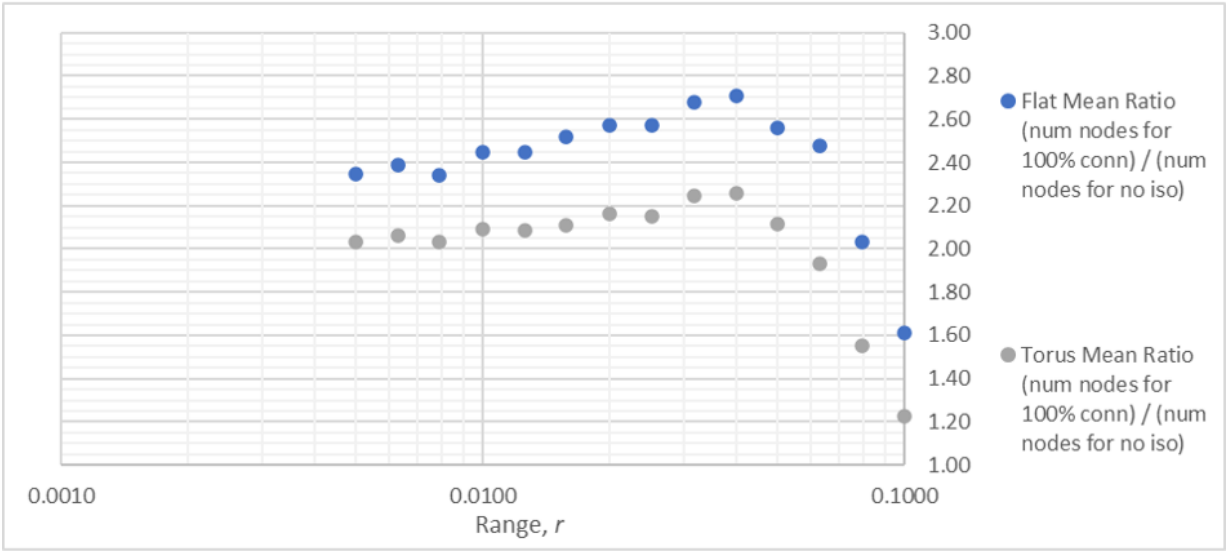

Figure 154: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in flat and toroidal unit squares in 1D with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

7.1.2 Two-Dimensional Spaces

RGGs in a unit square in spaces of dimension $d = 2$ have already been extensively studied in this paper. This section gives LRMs for RGGs in flat and toroidal spaces that were simulated for another independent sample of size $N = 1{,}000$ randomly generated geometries for the mid-r set to assess the differences in the LRMs due to noise in the data and optimistic error calculations due to overfitting.

The plots of the 99%ile number of nodes are not shown in this section but are included in Section 7.1.5 and compared to those for the other dimensions simulated.

For a flat space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.8834\, r^{-2.0593} \tag{132}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.93% and max error of the value of the predicted n is 5.99%.

For a flat space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 4.0869\, r^{-2.0882} \tag{133}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.86% and max error of the value of the predicted n is -6.80%.

For a toroidal space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, the LRM is

$$n = 2.3166\, r^{-2.1607} \tag{134}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.86% and max error of the value of the predicted n is 3.39%.

For a toroidal space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 2.1191\, r^{-2.1752} \tag{135}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.63% and max error of the value of the predicted n is -3.53%.

Arbitrarily taking the previous LRMs given in Section 4.2.1.1, Section 4.2.1.4, Section 4.3.1.1, and Section 4.3.1.3 for the 99%ile number of nodes as the reference for comparison, the relative error of the predicted number of nodes resulting from the new LRMs for the mid-r set is as follows. Compared to (22) for a flat space and full connectivity, (132) has MAPE 3.58% and max error 8.82%. Compared to (31) for a flat space and no isolated nodes, (133) has MAPE 4.14% and max error 16.55%. Compared to (52) for a toroidal space and full connectivity, (134) has MAPE 2.17% and max error 6.65%. Compared to (58) for a toroidal space and no isolated nodes, (135) has MAPE 1.97% and max error 5.99%.

It can be seen that MAPE for the new LRMs is quite small when compared to the original LRMs, although max error percentages can be in the double digits.

Comparisons similar to those given in Section 4.2 and Section 4.3 are provided for two-dimensional flat and toroidal unit squares in Figure 155 and Figure 156.

As was shown in Section 4.2 and Section 4.3, Figure 155 indicates that the probability that full connectivity is first achieved when the last isolated node disappears is at its largest value for the smallest range tested and probability curve is slightly higher for RGGs in the toroidal spaces.

Figure 156 shows the mean ratios of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes first disappear in flat and toroidal unit squares in two-dimensional spaces for RGGs with n-for-given-r growth. The mean is taken over all of the ratios for a given value of the range for each simulated geometry. Figure 156 shows that about 30% more nodes are required to first achieve full connectivity than for no isolated nodes in RGGs with large range in flat spaces of $d = 2$ and about

20% more nodes for RGGs in toroidal spaces. Both ratios decrease with the range but the ratios do not quite reach 1 for the smallest ranges tested.

The data points shown in Figure 155 and Figure 156 differ slightly from those shown in Figure 33, Figure 73, Figure 34, and Figure 74 due to the noise resulting from different simulations. The values obtained from the two sets of simulations are close, however, and the trends are clearly the same.

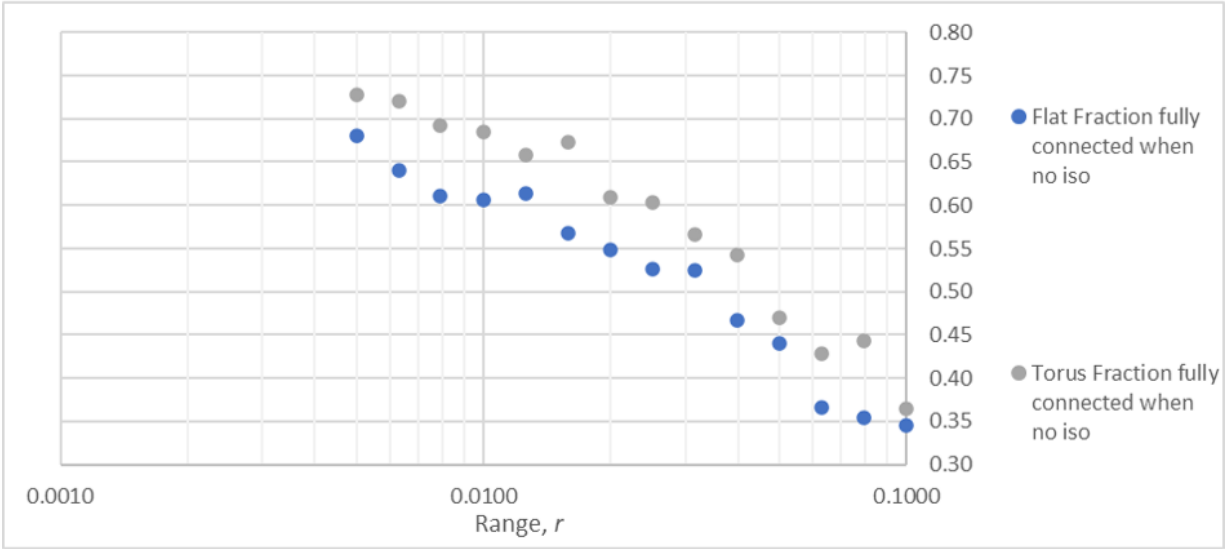

Figure 155: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 2D to become fully connected with n-for-given-r growth.

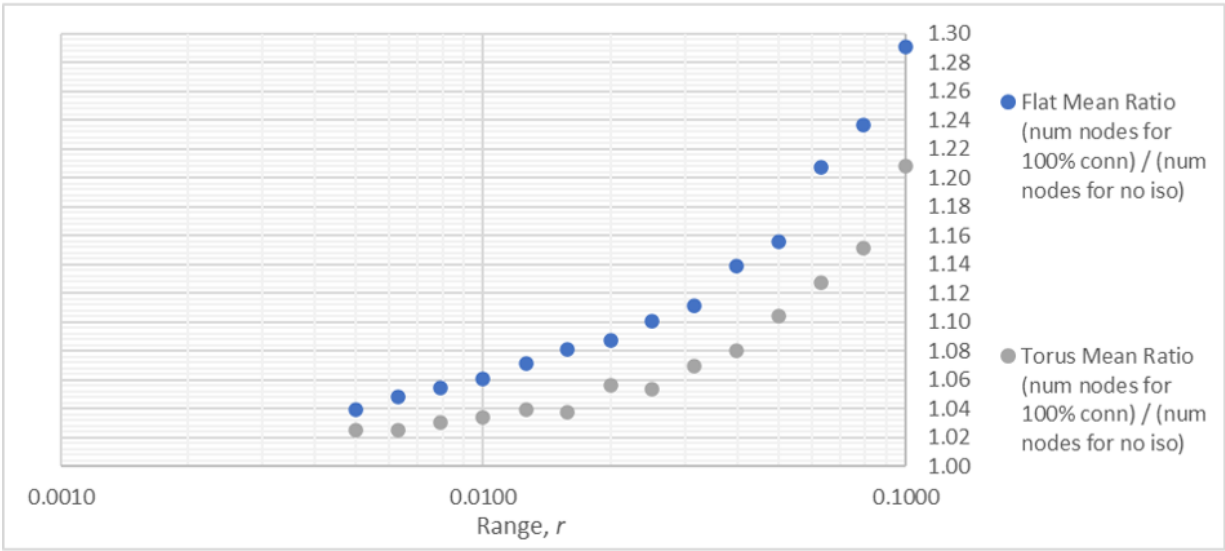

Figure 156: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in flat and toroidal unit squares in 2D with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

7.1.3 Three-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 3$ consist of nodes in a cube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with n-for-given-r growth for the mid-r set.

The plots of the 99%ile number of nodes are not shown in this section but are included in Section 7.1.5 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.9995\, r^{-3.1019} \tag{136}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.40% and max error of the value of the predicted n is 7.05%.

For a flat space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 4.7681\, r^{-3.1109} \tag{137}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.39% and max error of the value of the predicted n is -5.93%.

For a toroidal space and the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, the LRM is

$$n = 1.8837\, r^{-3.1996} \tag{138}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.25% and max error of the value of the predicted n is 6.60%.

For a toroidal space and the 99%ile number of nodes required to first achieve no isolated nodes with n-for-given-r growth, the LRM is

$$n = 1.8881\, r^{-3.1985} \tag{139}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 2.35% and max error of the value of the predicted n is 6.59%.

The results for RGGs in three-dimensional flat and toroidal unit squares are shown in Figure 157 and Figure 158.

Figure 157 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears (as predicted by Theorem 13.17 of [Penrose (2003)]) is much higher for spaces of $d = 3$ than that shown in Figure 155 for $d = 2$ for every value of the simulated range, with the probability being higher for the toroidal space. That probability increases as the range decreases for both flat and toroidal spaces. For the smallest ranges simulated, the probability of becoming fully connected when the last isolated node disappears is slightly below 95% for RGGs in a flat space and slightly above 95% for RGGs in a toroidal space.

Figure 158 shows the mean ratios of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes first disappear in flat and toroidal unit squares in three-dimensional spaces for RGGs with n-for-given-r growth. The mean is taken over all of the ratios for a given value of the range for each simulated geometry. The ratios for both flat and toroidal spaces are only slightly greater than 1 for the largest ranges simulated and they approach 1 even more closely for the smaller ranges.

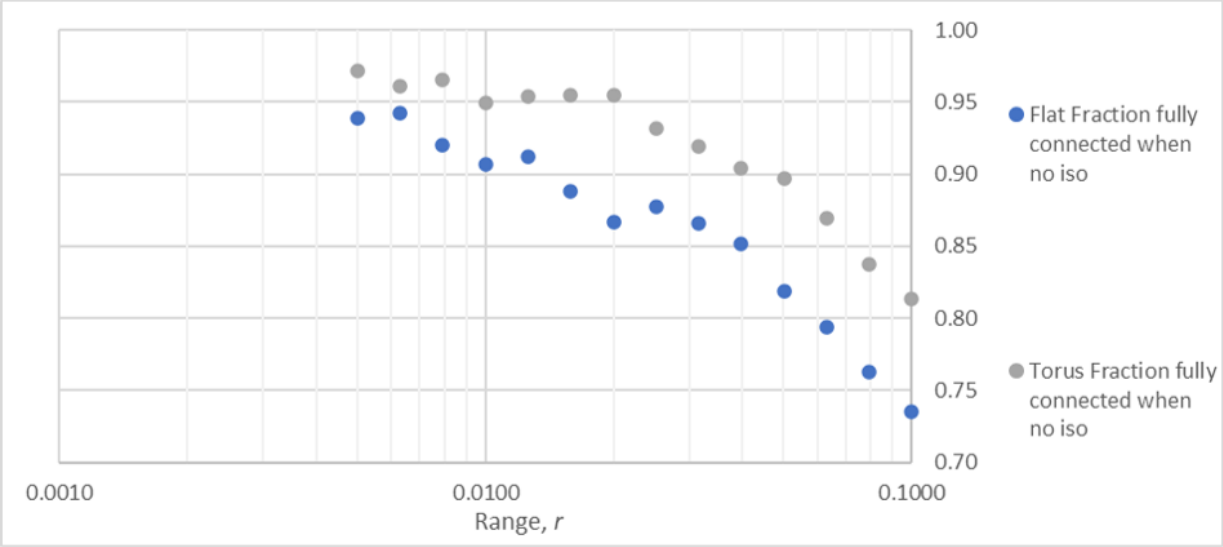

Figure 157: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 3D (i.e., flat and toroidal cubes) to become fully connected with n-for-given-r growth.

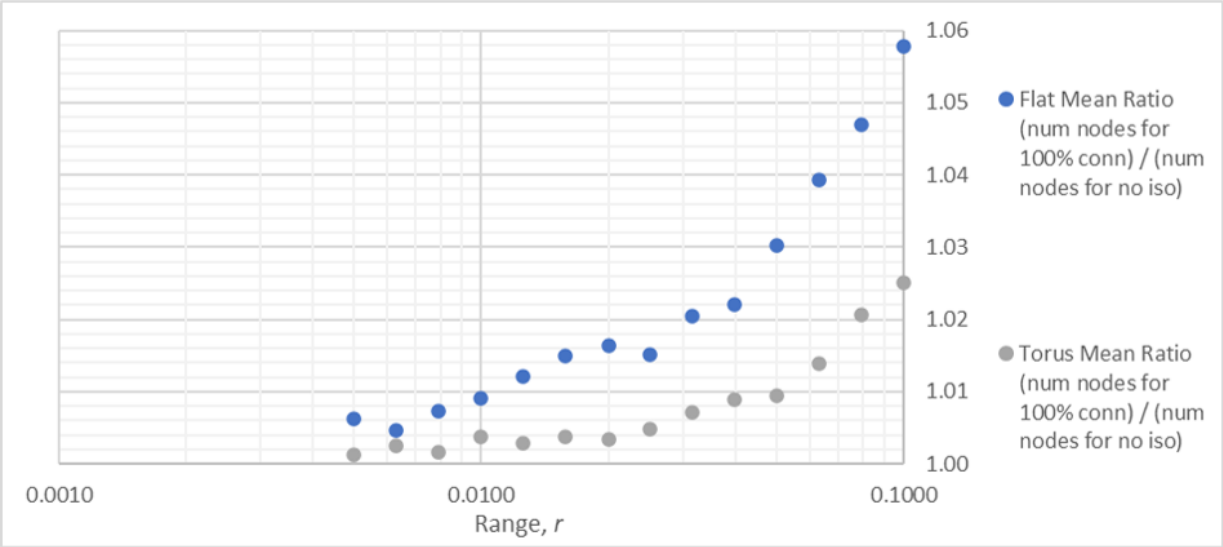

Figure 158: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in flat and toroidal unit squares in 3D with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

7.1.4 Four-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 4$ consist of nodes in a hypercube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with n-for-given-r growth for a subset of the ranges in the mid-r set. The values used were $r = 0.1000$ down to 0.0126. Smaller ranges could not be used because they required numbers of nodes that exceeded memory limitations on servers.

The plots of the 99%ile number of nodes are not shown in this section but are included in Section 7.1.5 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales. As is the case for the spaces of smaller dimension, the power law is a good approximation for the LRMs.

While the number of nodes required to first achieve no isolated nodes did not always equal the numbers required to first achieve full connectivity, this event occurred so often that the LRMs for the 99%ile number of nodes required to first achieve full connectivity and 99%ile number of nodes required to first achieve no isolated nodes for flat space were the same. This held as well for toroidal space.

For a flat space and the 99%ile number of nodes required to first achieve both full connectivity and no isolated nodes with n-for-given-r growth, (21) becomes

$$n = 7.7847\, r^{-4.0697} \tag{140}$$

with an R^2 value of 0.9974. For the simulated values of r, MAPE is 4.90% and max error of the value of the predicted n is -11.69%.

For a toroidal space and the 99%ile number of nodes required to first achieve both full connectivity and no isolated nodes with n-for-given-r growth, the LRM is

$$n = 1.7207\, r^{-4.2401} \tag{141}$$

with an R^2 value of 0.9993. For the simulated values of r, MAPE is 3.06% and max error of the value of the predicted n is -5.29%.

The results for RGGs in four-dimensional flat and toroidal unit squares are shown in Figure 159 and Figure 160.

Figure 159 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears is again much higher than for spaces of smaller dimensions for every value of the simulated range, with the probability being higher for the toroidal space. The probability is above 90% for both spaces even for the largest ranges considered.

Figure 160 shows the mean ratios of the number of nodes required to first achieve full connectivity to the number of nodes where isolated nodes first disappear in flat and toroidal unit squares in one-dimensional spaces for RGGs with n-for-given-r growth. The mean is taken over all of the ratios for a given value of the range for each simulated geometry. The ratios for both flat and toroidal spaces over the ranges simulated are practically indistinguishable from 1.

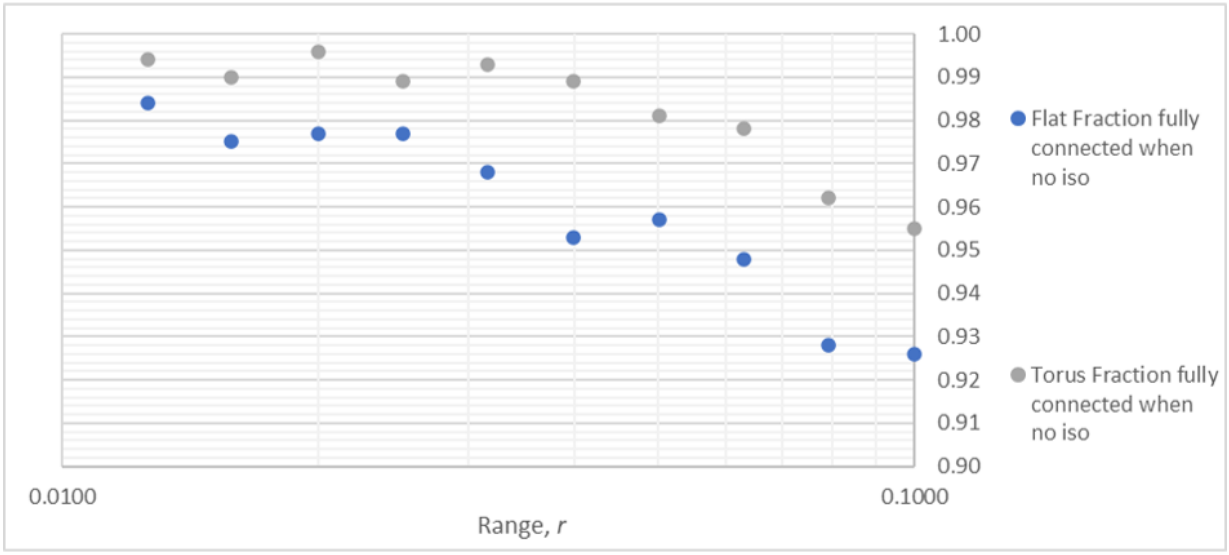

Figure 159: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 4D (i.e., flat and toroidal hypercubes) with n-for-given-r growth to become fully connected.

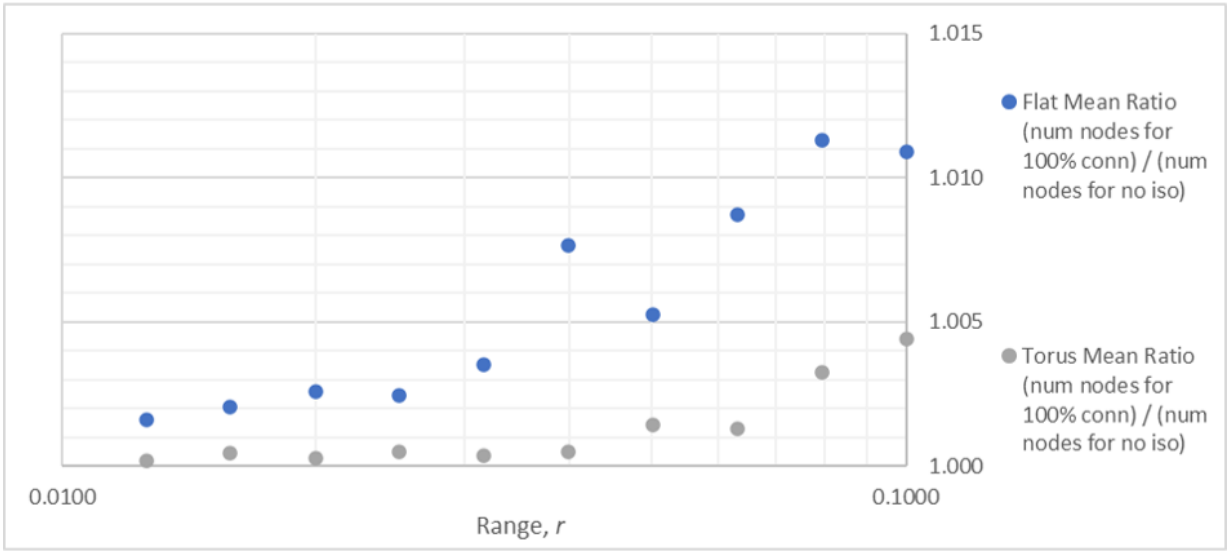

Figure 160: Mean of the ratios of number of nodes for full connectivity to the number of nodes where isolated nodes disappear in flat and toroidal unit squares in 4D with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

7.1.5 Comparing Number of Nodes Required in d-Dimensional Spaces

Figure 161, Figure 162, Figure 163, and Figure 164 summarize by type of connectivity the LRMs presented by dimension in Section 7.1.1, Section 7.1.2, Section 7.1.3, and Section 7.1.4. For n-for-given-r growth, these plots demonstrate that, as d increases, there is a huge increase in the number of nodes required to first achieve the connectivity conditions of RGGs being either fully connected or having no isolated nodes in both flat unit squares and toroidal unit squares.

The increase in the number of nodes required for connectivity is a consequence of the fact that a given number of nodes in a space spread out more as the dimensionality of a space increases. Thus, as considered by Penrose in the derivation of his theorem, more nodes are required to ensure the connectivity of an RGG.

An indication of the spreading out of nodes as the dimensionality of the space increases is shown through the example of the *mean distance between any two randomly placed nodes* in a flat unit square area of dimension d. A table of such values is given by Weisstein [Weisstein (2019b)]. For $d = 1$ the mean distance between two randomly placed nodes in a unit line is 0.3333. For $d = 2$ the mean distance between two randomly placed nodes in a unit square is 0.5214. For $d = 3$ the mean distance between two randomly placed nodes in a unit cube is 0.6617. For $d = 4$ the mean distance between two randomly placed nodes in a unit hypercube is 0.7777.

Even though the number of nodes required to first achieve connectivity increases with d, the number of nodes required to first achieve no isolated nodes more closely approaches the number needed to first achieve full connectivity for both flat and toroidal spaces as the dimensionality increases. This is demonstrated in Figure 154, Figure 156, Figure 158, and Figure 160. This phenomenon is also shown by the LRMs presented in Section 7.1.1, Section 7.1.2, Section 7.1.3, and the discussion in Section 7.1.4 by the fact that the respective m and b coefficients in the expressions for the 99%ile number of nodes required to first achieve full connectivity and no isolated nodes become closer as the dimensionality increases.

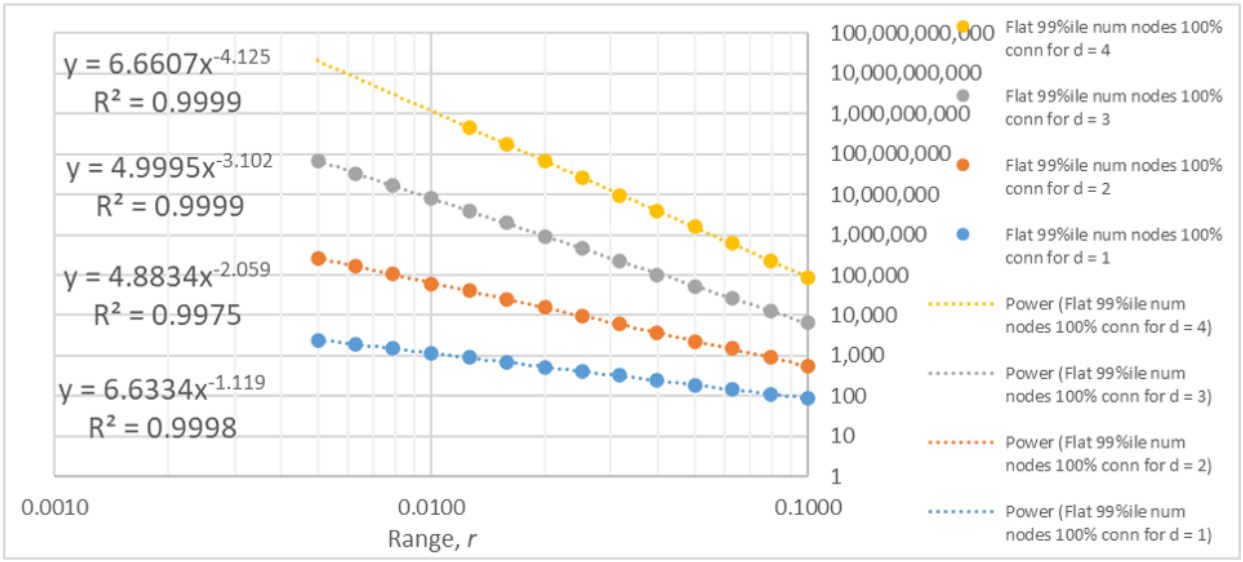

Figure 161: LRMs for 99%ile number of nodes for full connectivity of RGGs in flat unit squares in 1D through 4D for RGGs with n-for-given-r growth.

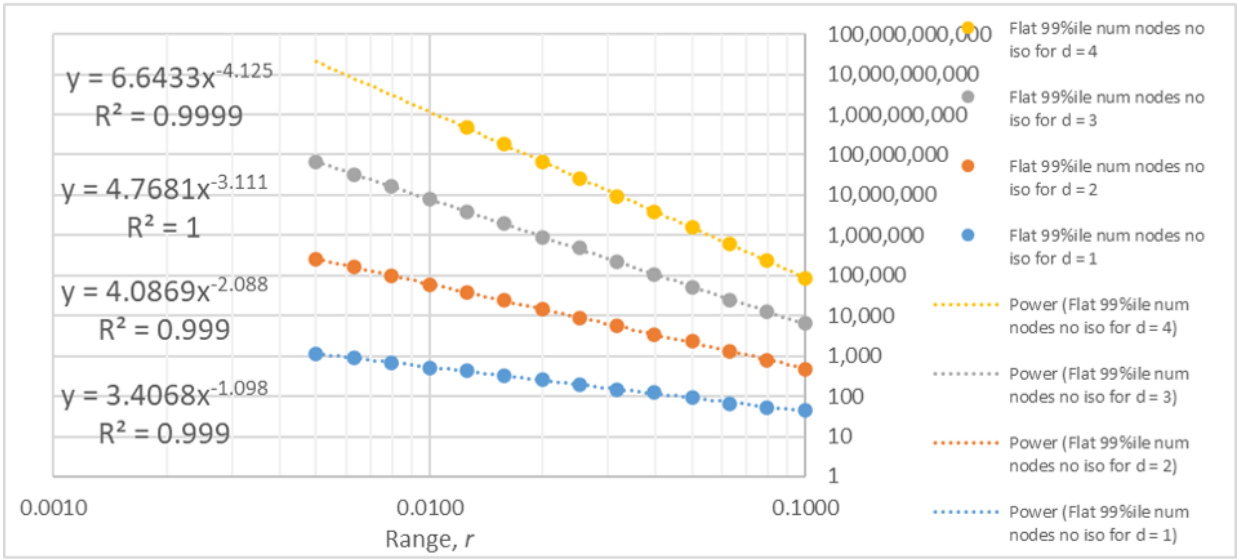

Figure 162: LRMs for 99%ile number of nodes for no isolated nodes in RGGs in flat unit squares in 1D through 4D for RGGs with n-for-given-r growth.

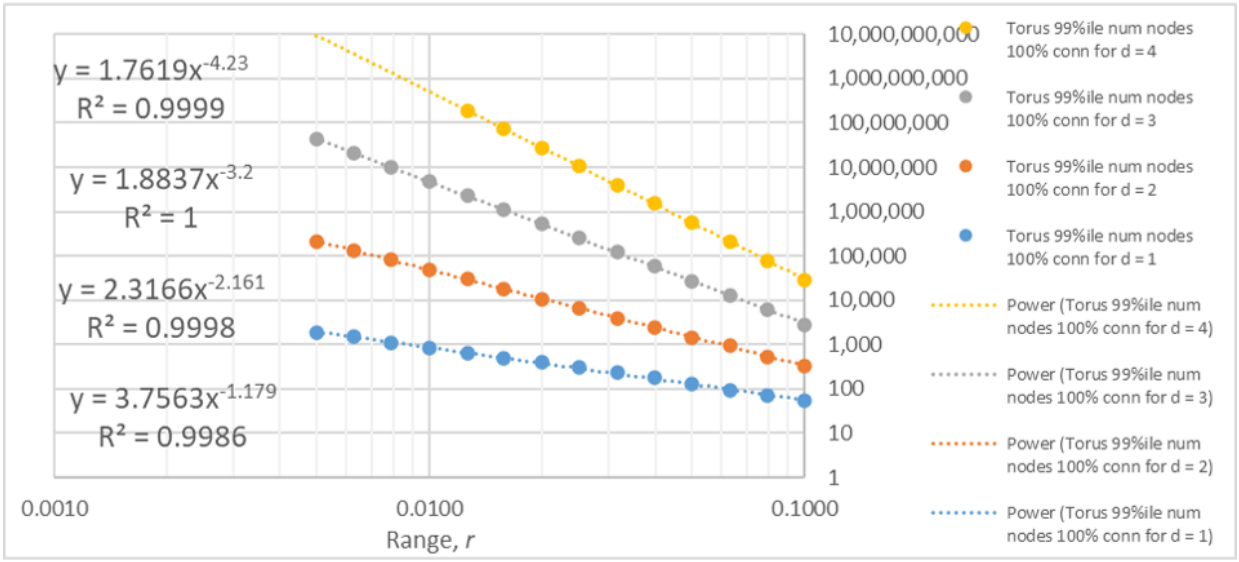

Figure 163: LRMs for 99%ile number of nodes for full connectivity of RGGs in toroidal unit squares in 1D through 4D for RGGs with n-for-given-r growth.

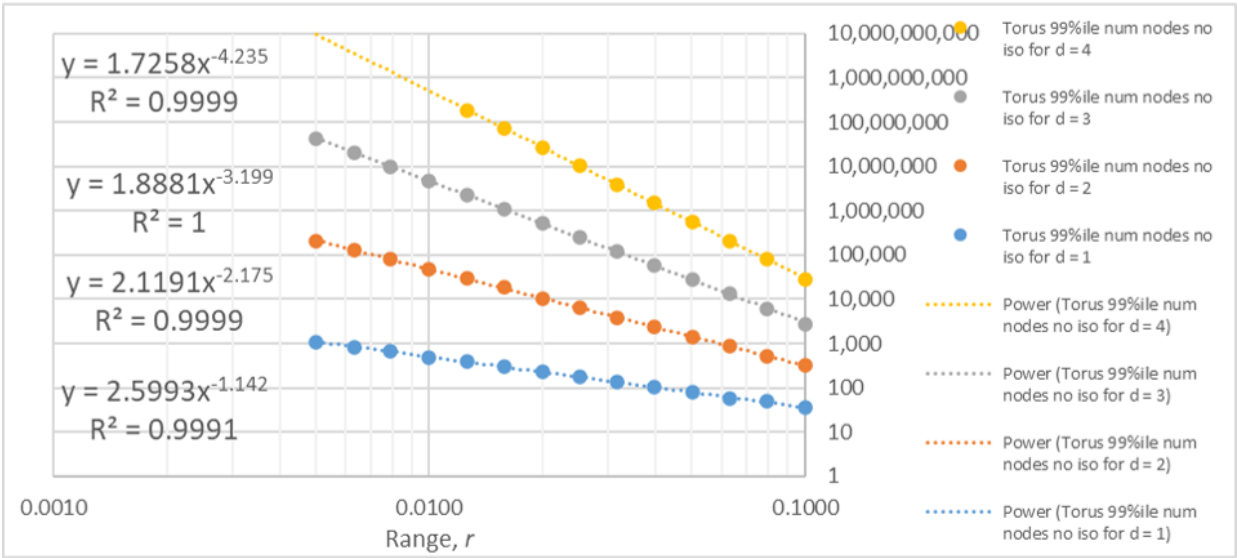

Figure 164: LRMs for 99%ile number of nodes for no isolated nodes in RGGs in toroidal unit squares in 1D through 4D for RGGs with n-for-given-r growth.

7.2 RGGs in d-Dimensional Spaces with r-for-given-n Growth

In Section 3.4 expression (13) was derived to estimate the range required for no isolated nodes for a given number of nodes with probability C for two-dimensional space. In this section the equivalent expression is derived for RGGs in d-dimensional spaces.

The estimate of the range required for no isolated nodes in d-dimensional spaces is compared to the general form of Penrose's Theorem for d-dimensional spaces.

This section considers connectivity in d-dimensional spaces for the r-for-given-n growth model where the RGGs increase the range of a fixed number of nodes until the desired connectivity conditions are reached. The statistical metric is the 99%ile range required to achieve the conditions of full connectivity and no isolated nodes.

Ratios are obtained for the ranges required to achieve full connectivity, no isolated nodes, and the values given by Penrose's Theorem in d-dimensional spaces.

7.2.1 Estimating the Probability of No Isolated Nodes in d-Dimensional Spaces

Section 3.4 presented expressions for the probability of no isolated nodes in RGGs in two-dimensional toroidal spaces. When border effects can be disregarded, (12) gives the probability or confidence C that an RGG contains no isolated nodes as a function of r and n. Expression (13) is the solution for r in terms of C and n, and n can be found as a function of C and r by solving (12) numerically. This section derives the equivalents of (12) and (13) for d-dimensional spaces.

To be consistent with Penrose's assumptions the spaces considered in this section (as was done in Section 3.4) are toroidal unit squares in 2D and their "unit square" equivalents in spaces of dimensions $d \neq 2$.

As noted in Section 1.2 of the Introduction, for $d = 1$ the space is a line of length $a = 1$ that wraps around at its ends, for $d = 3$ the space is a cube with sides of length $a = 1$ that wraps around at its opposite faces, and for $d \geq 4$ the space is a hypercube with sides of length $a = 1$ that wraps around at its opposite faces. As also noted in Section 1.2, the term *area* is used in this paper not only in the customary sense of area in 2D but is also used as the general term to mean length in a one-dimensional space or volume or hypervolume in dimensions greater than two.

In an RGG in a toroidal space of any dimension the mean density of nodes is constant throughout the space. For unit spaces of area $A = 1$ the mean density of nodes is given by $\frac{n}{A} = \frac{n}{1} = n$.

In Section 3.4 (8) stated the well-known fact that the area of a circle with radius r in two-dimensional space is given by πr^2. To find the probability of no isolated nodes in d-dimensional spaces the corresponding but less well-known expression for arbitrary dimension d is needed.

Mathematical literature refers to a circle with radius 1 in d-dimensional spaces as the *unit ball*. For $d = 1$ the unit ball is a line with "radius"

1 and so its length is 2. For $d \geq 2$ the unit ball is a circle, sphere, or hypersphere with radius 1.[52]

The volume of the unit ball in d-dimensional spaces is given by Cooper and Frieze [Cooper (2011)] in terms of the gamma function as

$$\theta_d = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \tag{142}$$

and so the "volume" of a line, circle, sphere, or hypersphere with radius r in d-dimensional spaces is

$$\theta_d \, r^d \quad . \tag{143}$$

Using more convenient alternate expressions for (142) that are also given by Cooper and Frieze [Cooper (2011)] for θ_d for odd and even values of d, familiar exact values of (143) are $2\,r$ for $d=1$ (the length of a line that has two "radii" of length r), $\pi\, r^2$ for $d=2$ (which is (8) yielding the area of a circle), and $\frac{4\,\pi}{3} r^3$ for $d=3$ (the volume of a sphere).

Continuing into higher dimensions, the volumes of hyperspheres are given by $\frac{\pi^2}{2} r^4$ for $d=4$, $\frac{8\,\pi^2}{15} r^5$ for $d=5$, and $\frac{\pi^3}{6} r^6$ for $d=6$.

With mean node density n and range r the mean degree μ_d within that range in d-dimensional spaces for any one node is

$$\mu_d \quad = \quad \theta_d \, r^d \, n \quad . \tag{144}$$

Knowing the mean number of neighbors per node μ_d the probability that any one node has exactly y neighbors is given by the Poisson distribution:

$$\Pr\{\text{any one node has exactly } y \text{ neighbors}\} \; = \frac{\mu_d^y \, e^{-\mu_d}}{y!} \quad . \tag{145}$$

[52] Note the distinction that the "unit ball" is defined to have *radius* 1 whereas the "unit circle" and "unit square" are defined to have *area* 1. Note also that the "unit ball" is not the same as the "unit sphere" introduced in Section 1.3 as a possible model of a toroidal unit circle; the unit sphere is defined to have surface area 1.

Ignoring border effects, the probability that an individual node in an RGG is isolated is given by (145) with $y = 0$ as

$$\begin{aligned} \Pr\{\text{any one node is isolated}\} &= e^{-\mu_d} \\ &= e^{-\theta_d r^d n} \quad . \end{aligned} \tag{146}$$

Assuming statistical independence, the product of the complements of the probabilities of individual node isolation given by (146) yields the probability or confidence C of the joint event that (ignoring border effects) *no node* in a d-dimensional RGG is isolated:

$$\begin{aligned} C &= \Pr\{\text{no isolated nodes}\} \\ &= \left[1 - e^{-\theta_d r^d n}\right]^n \quad . \end{aligned} \tag{147}$$

As in Section 3.4, (147) answers the question of the probability of having no isolated nodes without consideration of the specific RGG growth model in a toroidal space.

For the r-for-given-n growth model, (147) can be solved for r as a function of n and C:

$$r = \left[\frac{-\ln(1 - C^{1/n})}{\theta_d \, n}\right]^{1/d} \quad , \tag{148}$$

which is the d-dimensional form of (13) for the 2D case.

As in Section 3.4 for $d = 2$, the derivation of (147) and the expressions obtained from it depend only on the Poisson distribution of nodes and the assumption of statistical independence. It is shown in the next section that (148) yields virtually the same numerical results as the Second Claim of Penrose's Theorem for d-dimensional spaces.

7.2.2 Penrose's Theorem in d-Dimensional Spaces

Penrose's Theorem for two-dimensional space was described in Section 3.5 but it has been noted that Penrose in fact generalized the problem of determining the probability of node isolation and connectivity for spaces of arbitrary dimension d [Penrose (1999)] [Penrose (2003)]. Assuming a toroidal space so that border effects could be ignored, he obtained the following result for an RGG in a d-dimensional spaces with r-for-given-n growth and 1-connectivity (Corollary 13.20 of [Penrose (2003)]), which is stated in the notation of this paper as:

> For an RGG in d-dimensional spaces (where $d \geq 2$) without border effects to contain no isolated nodes,
>
> $$C \quad = \quad \Pr\{\ \theta_d\, r^d\, n - \ln n \leq c_P\ \} \qquad (149)$$
>
> where C and c_P are related by (15) and θ_d is given by (142).

Even though this proposition explicitly excludes one-dimensional spaces from consideration, the connectivity of RGGs in 1D is tested in Section 7.2.3.

For the r-for-given-n growth model, Penrose's (149) and (15) are equivalent to saying that an RGG in d-dimensional spaces contains no isolated nodes (and so by Penrose's First Claim should be fully connected for "n large") with probability C when r is given by

$$r \quad = \quad \left[\frac{\ln n + c_P}{\theta_d\, n}\right]^{1/d} . \qquad (150)$$

Thus, for 1-connectivity, Penrose's (149) and (150) are the d-dimensional form of what is called "Penrose's Theorem" in this paper. Specifically, (150) is the d-dimensional equivalent of (16) for the 2D case.

It is demonstrated here that (148), derived under the minimal assumptions of Poisson distribution of nodes and statistical independence, provides nearly identical numerical results as Penrose's (150) for values of C close to 1. As in Section 3.5, the probability of $C = 0.99$ (so $c_P = 4.6001$) is used for comparison below.

For two-dimensional space, it was found in Section 3.5 that the relative error of the value of r obtained from Penrose's (16) with respect to the value of r obtained from (13) is -3.64×10^{-5} for $n = 10$ nodes, and that value decreases as n increases.

For d-dimensional spaces, the relative error of the value of r obtained from Penrose's (150) with respect to the value of r obtained from (148) also decreases as n increases. In addition, the relative error for $n = 10$ nodes decreases with d, where that relative error is -7.28×10^{-5} for $d = 1$, -3.64×10^{-5} for $d = 2$, -2.43×10^{-5} for $d = 3$, -1.82×10^{-5} for $d = 4$, -1.46×10^{-5} for $d = 5$, and -1.21×10^{-5} for $d = 6$.

Finally, as described in Section 3.5 and justified in footnote #24 for two-dimensional space, replacing $1 - C^{1/n}$ by $-\frac{\ln C}{n}$ in (148) yields exactly Penrose's (150).

These results show that the d-dimensional form of the Second Claim of Penrose's Theorem yields numerical results that are practically identical to those obtained from (148) which depends only on the minimal assumptions of Poisson distribution of nodes and statistical independence.

The d-dimensional form of Penrose's Theorem is tested against simulation results in Section 7.2.3.

7.2.3 Comparing Full Connectivity, No Isolated Nodes, and Penrose's Theorem in d-Dimensional Spaces

Section 7.2.4 through Section 7.2.9 compare the ranges required to achieve full connectivity to those required to achieve no isolated nodes for flat and toroidal spaces of dimensions $d = 1, 2, \ldots 6$, as obtained from simulated geometries for RGGs with r-for-given-n growth. Section 7.2.10 compares the ranges for full connectivity and no isolated nodes to the values predicted by the Second Claim of Penrose's Theorem for those dimensions. In spite of the restriction of Penrose's Theorem to spaces of $d \geq 2$, this theorem is also tested in 1D.

As noted previously, Penrose's Theorem applies specifically to toroidal square spaces with r-for-given-n growth; however, it has also been tested in this paper for many "off-label" situations such as flat spaces, circular shapes, and n-for-given-r growth.

For two-dimensional space, Penrose's Theorem was tested in Section 5.3.1.2 Section 5.3.2.2, and Section 5.3.3.2 in its intended situation. Even though only 82.1% of simulations resulted in RGGs in toroidal space being fully connected when the last isolated node disappeared even for 10,000,000 nodes, Penrose's Theorem worked extremely well for prediction of the 99%ile range. This was shown most dramatically in Figure 137 and Figure 138.

For two-dimensional space, it was observed in Section 5.3.1.2 that Penrose's Theorem performed very well for RGGs in a toroidal unit square with r-for-given-n growth, and in Section 4.3.1.5 it was observed that the theorem performed well for n-for-given-r growth for $C = 0.99$. Furthermore, in these cases, Penrose's Theorem had small error even for values of n as small as 20 nodes in spite of the fact that its proof was for "n large."

In this section it is observed that Penrose's Theorem performed well for RGGs in toroidal spaces of dimensions other than $d = 2$, including 1D which was outside the envelope of its assumptions. It is demonstrated that Penrose's Theorem gives excellent results under the circumstances for which it was intended and proven as well as for some other conditions.

The results for two-dimensional spaces presented in this section are based on a set of simulations independent from those used in Section 5.2.1.1 and Section 5.3.1.1.

7.2.4 One-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 1$ consist of nodes in a line of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 4.8087\, n^{-0.8856} \tag{151}$$

with an R^2 value of 0.9987. For the simulated values of r, MAPE is 6.87% and max error of the value of the predicted r is 30.34%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 3.8185\, n^{-0.9341} \tag{152}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.97% and max error of the value of the predicted r is -12.19%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 2.9755\, n^{-0.8574} \tag{153}$$

with an R^2 value of 0.9982. For the simulated values of r, MAPE is 8.15% and max error of the value of the predicted r is 37.99%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 2.5676\, n^{-0.8932} \tag{154}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 5.70% and max error of the value of the predicted r is 23.71%.

The results for RGGs in one-dimensional flat and toroidal unit squares are shown in Figure 165, Figure 166, and Figure 167.

Figure 165 shows the probability that an RGG is connected "at the same instant" that the last isolated node disappears on a linear-log scale. In contrast to the behavior noted for two-dimensional space, that probability decreases as the number of nodes increases for both flat and toroidal spaces. For the largest number of nodes simulated, the probability of becoming fully connected when the last isolated node disappears is at its highest values for the smallest number of nodes simulated (0.215 for RGGs in a flat space and approximately 0.244 for RGGs in a toroidal space for $n = 10$ nodes) and are below 1% and sometimes 0% for $n \geq 252$.

Because the probabilities are so small for the larger values of n in Figure 165, Figure 166 shows the same points on log-log scales. Only the values for $n = 10$ up to 1,000 are shown to avoid using the zero-valued probabilities of becoming fully connected when the last isolated node disappears. The straight lines indicate that these probabilities can be modeled by power law relationships as shown by the trend lines and their LRMs on the plots. The two LRMs are quite similar and the approximation

$$2.7\, n^{-1.1} \tag{155}$$

would track closely through the middle of the two graphs of the probabilities.

Figure 167 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in one-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. RGGs in flat

and toroidal spaces require 80% and 60% more range to become fully connected beyond the range required for no isolated nodes for smallest values of n simulated. The ratios for RGGs in a toroidal space increases with the number of nodes and the ratios for RGGs in a flat space appear to quickly approach 2 as n increases.

The relative errors of the LRMs given in (151), (152), (153), and (154) are shown against those of Penrose's predicted ranges in Figure 168, Figure 169, Figure 170, and Figure 171, respectively. In spite of the fact that Penrose excluded the case of $d = 1$ from his theorem, it can be seen that Penrose's Theorem provides excellent predictions of the range for no isolated nodes in RGGs in both flat and toroidal spaces, and for no isolated nodes in RGGs in toroidal space Penrose's Theorem performs even better than the LRM (154) for large n.

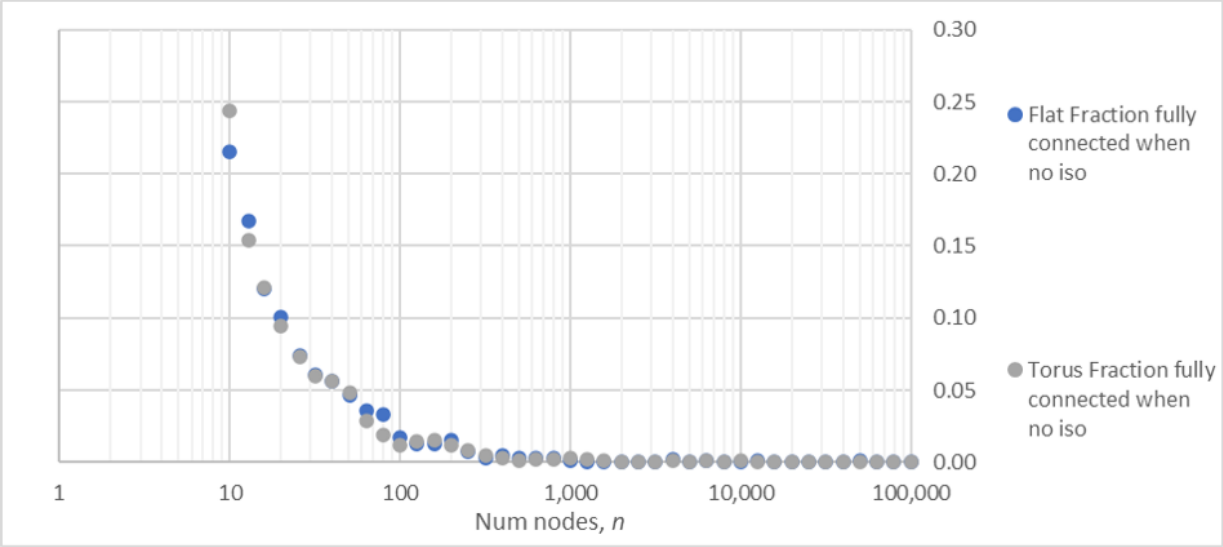

Figure 165: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 1D to become fully connected with r-for-given-n growth.

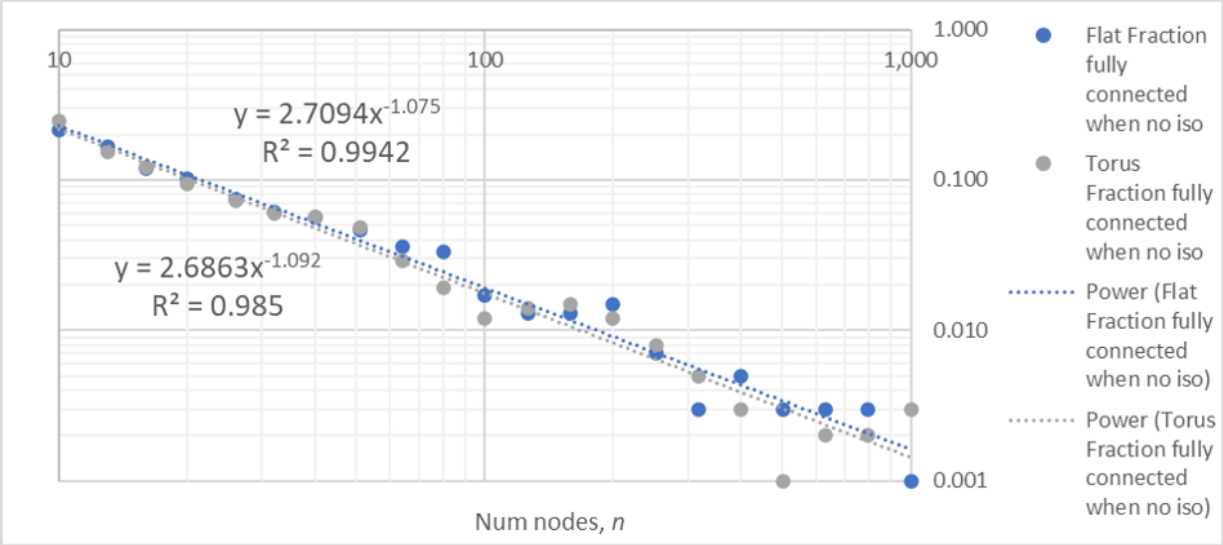

Figure 166: Same as Figure 165 but showing the fraction as a power law on log-log scales for the non-zero values of the fraction.

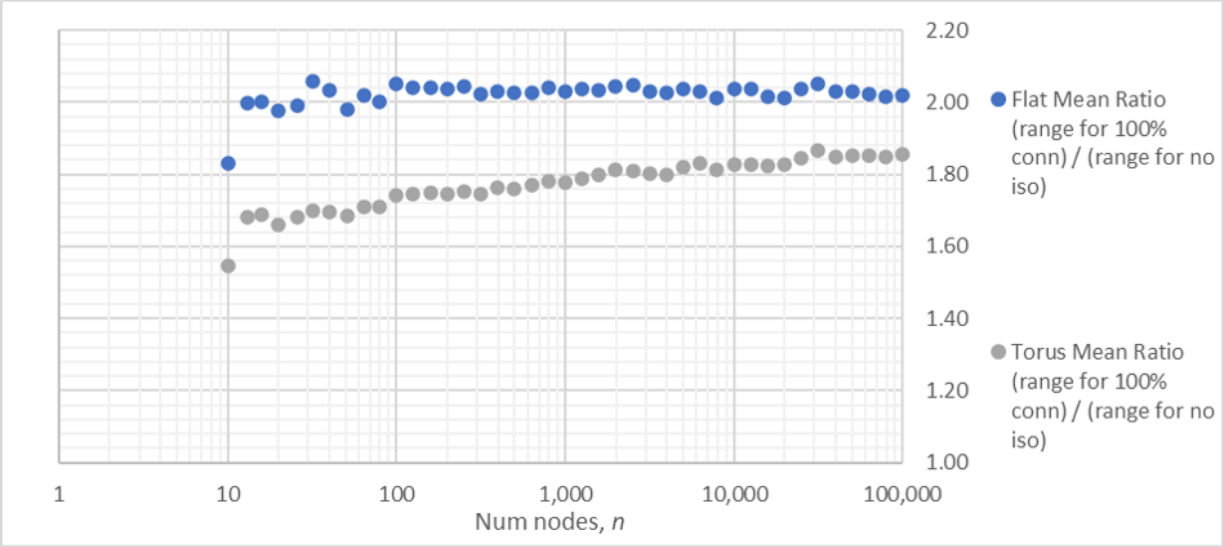

Figure 167: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 1D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

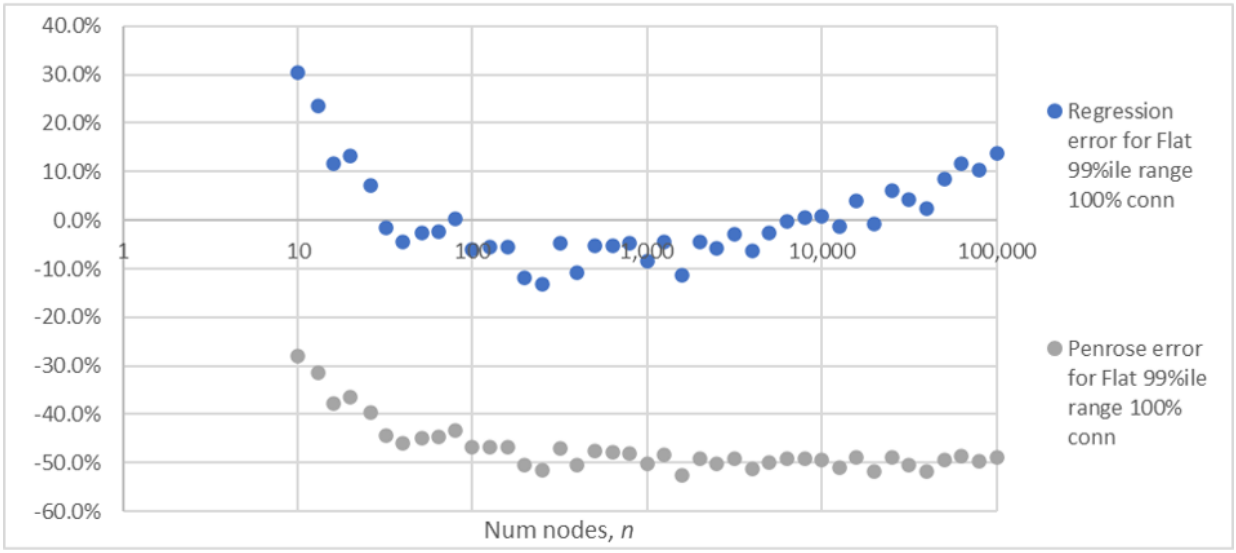

Figure 168: Relative error of LRM (151) for 99%ile range for full connectivity of RGGs in a flat unit square in 1D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

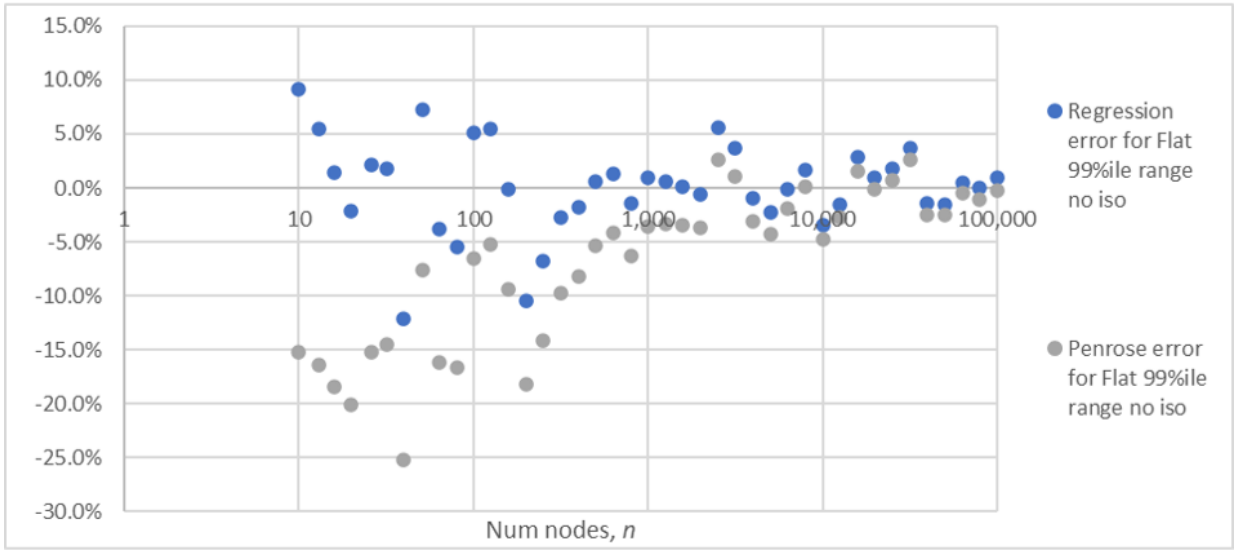

Figure 169: Relative error of LRM (152) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 1D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

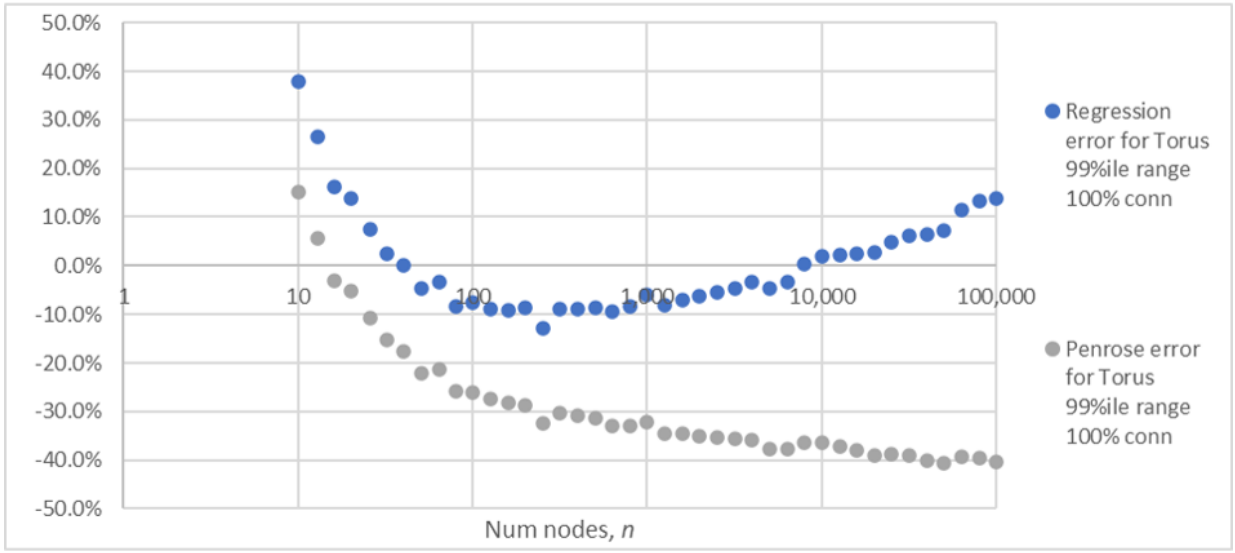

Figure 170: Relative error of LRM (153) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 1D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

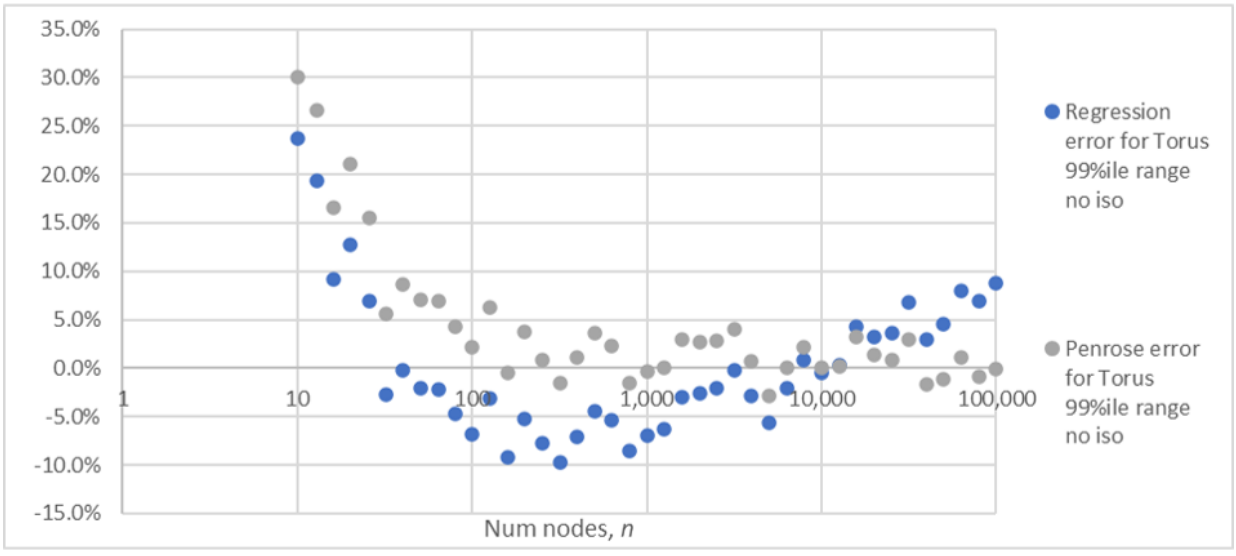

Figure 171: Relative error of LRM (154) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 1D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.5 Two-Dimensional Spaces

RGGs in flat and toroidal unit squares were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set. The results for two-dimensional spaces presented in this section are based on a set of simulations independent from those used in Section 5.2.1.1 and Section 5.3.1.1.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. As previously presented for RGGs in two-dimensional spaces in Section 5.2.1.1 and Section 5.3.1.1, the plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 2.1597\, n^{-0.4830} \tag{156}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.09% and max error of the value of the predicted r is 5.48%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 2.0518\, n^{-0.4796} \tag{157}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 1.41% and max error of the value of the predicted r is 4.29%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 1.3089\, n^{-0.4484} \tag{158}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 2.82% and max error of the value of the predicted r is 10.29%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.2728\, n^{-0.4460} \tag{159}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 2.81% and max error of the value of the predicted r is 0.84%.

The results for RGGs in two-dimensional flat and toroidal unit squares are shown in Figure 172 and Figure 173.

Figure 172 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears (as predicted by Theorem 13.17 of [Penrose (2003)]) is not 1. That probability increases as the number of nodes increases for both flat and toroidal spaces,[53] but, for the largest number of nodes simulated here the probability of becoming fully connected when the last isolated node disappears in these sets of simulations is approximately 0.693 for RGGs in a flat space and approximately 0.752 for RGGs in a toroidal space. These values are consistent with those shown in Figure 119 in Section 5.2.3.2 and Figure 135 in Section 5.3.3.2.

Figure 173 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in two-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. The ratios for both flat and

[53] It is unclear as to why the probability that an RGG in a flat or toroidal unit square in 2D with r-for-given-n growth becomes fully connected when the last isolated node is absorbed would decrease and then increase as n increases. This phenomenon was noted in footnote #45.

toroidal spaces are only slightly greater than 1 for the largest values of n simulated and they approach 1 even more closely for the smaller values of n.

The relative errors of the LRMs given in (156), (157), (158), and (159) are shown against those of Penrose's predicted ranges in Figure 174 and Figure 175 for flat unit squares and Figure 176 and Figure 177 for toroidal unit squares, respectively. These trends closely match those shown in Figure 111 and Figure 112 for flat unit squares in Section 5.2.1.2 as well as Figure 127 and Figure 128 for toroidal unit squares in Section 5.3.1.2.

The LRMs outperform Penrose's Theorem's predictions of the 99%ile range for RGGs in flat space. Both the LRMs and Penrose's Theorem have very small relative errors for RGGs in toroidal space for both full connectivity and no isolated nodes but Penrose's Theorem has consistently better performance for the largest numbers of nodes simulated.

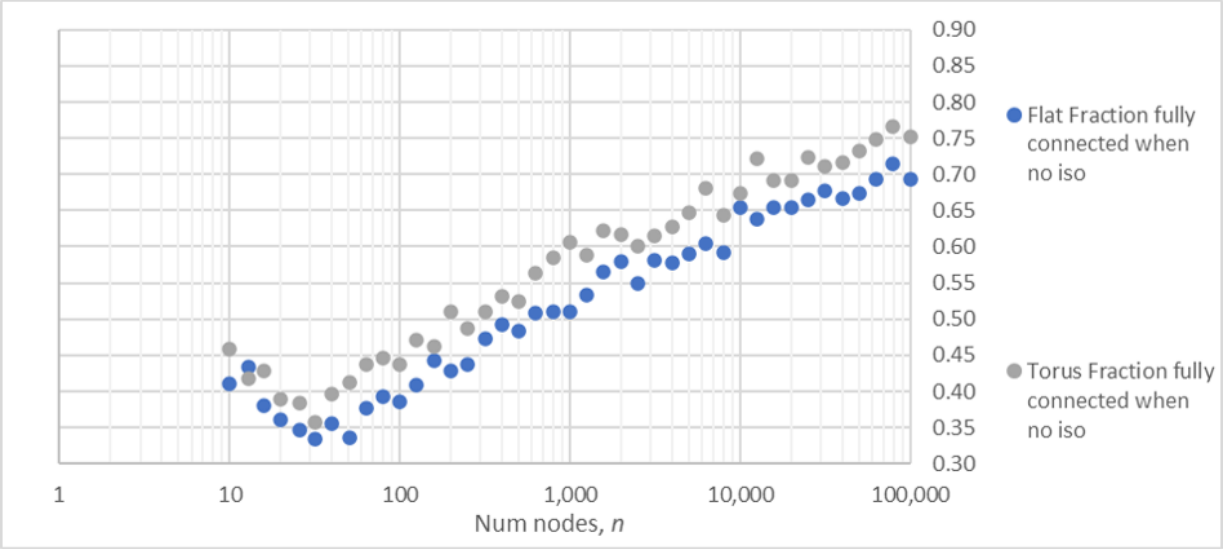

Figure 172: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 2D to become fully connected with r-for-given-n growth.

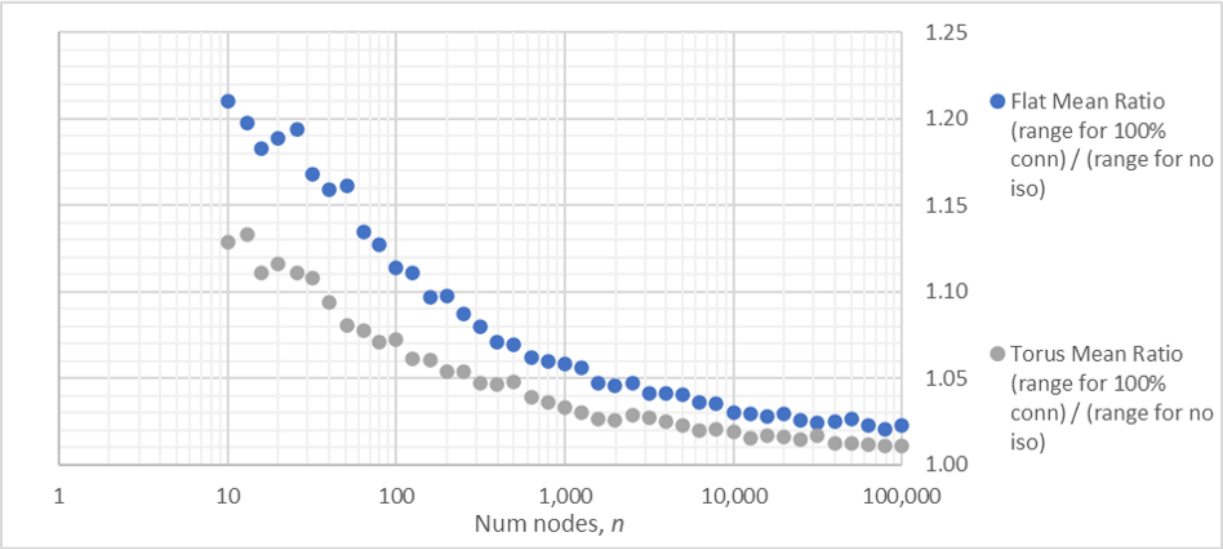

Figure 173: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 2D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

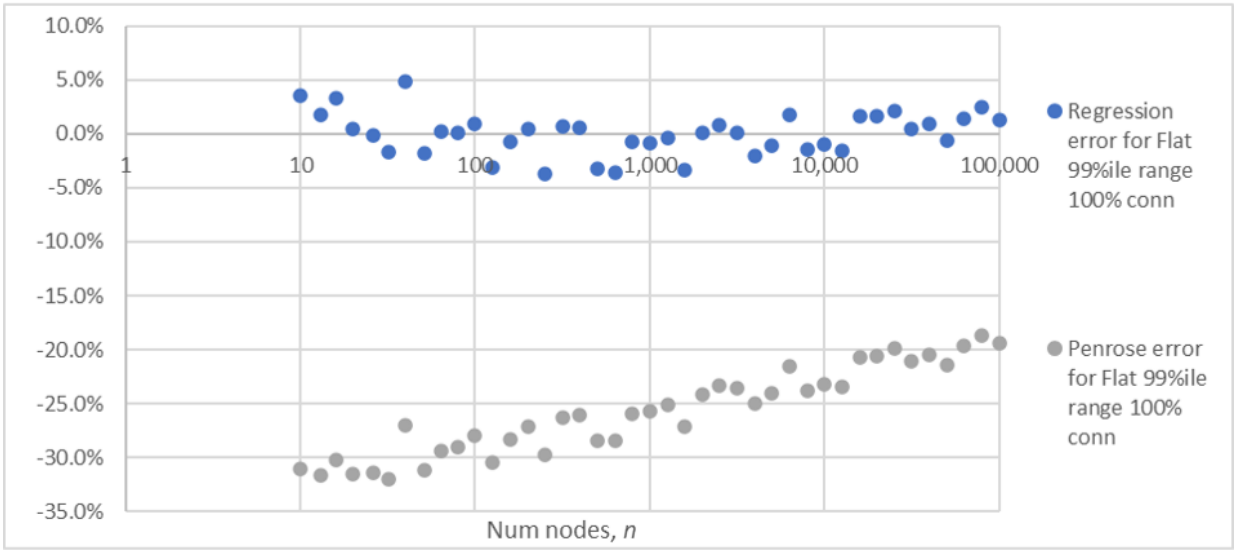

Figure 174: Relative error of LRM (156) for 99%ile range for full connectivity of RGGs in a flat unit square in 2D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

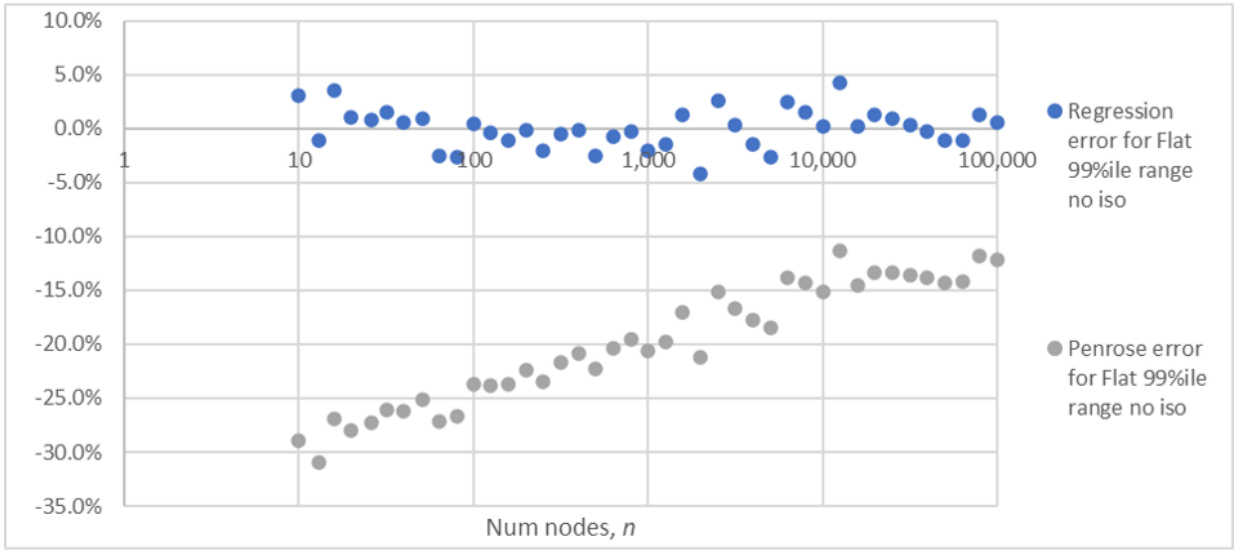

Figure 175: Relative error of LRM (157) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 2D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

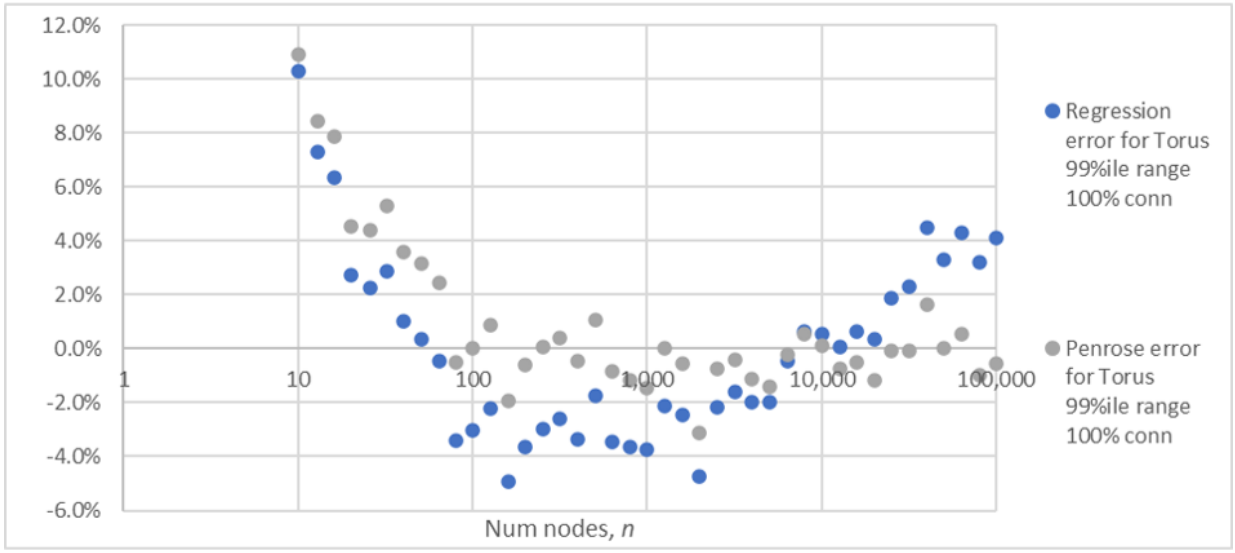

Figure 176: Relative error of LRM (158) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 2D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

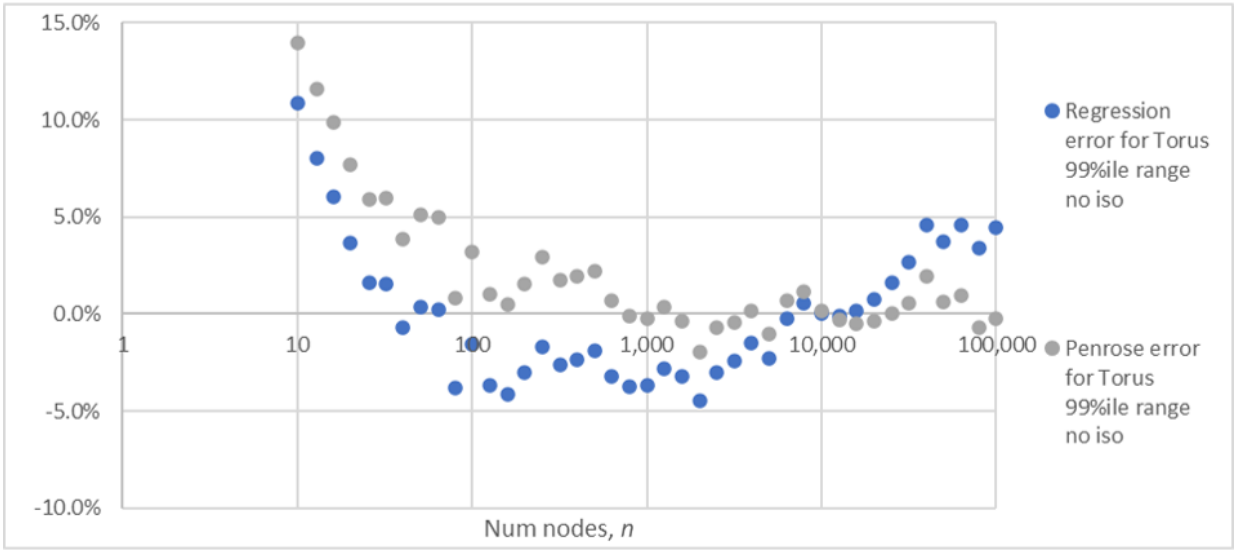

Figure 177: Relative error of LRM (159) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 2D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.6 Three-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 3$ consist of nodes in a cube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 1.7311\, n^{-0.3221} \tag{160}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 1.53% and max error of the value of the predicted r is 4.81%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.7174\, n^{-0.3216} \tag{161}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 1.51% and max error of the value of the predicted r is 4.70%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 1.0710\, n^{-0.2973} \tag{162}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 1.90% and max error of the value of the predicted r is 6.19%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.0674\, n^{-0.2971} \tag{163}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 1.91% and max error of the value of the predicted r is 7.56%.

The results for RGGs in three-dimensional flat and toroidal unit squares are shown in Figure 178 and Figure 179.

Figure 178 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears (as predicted by Theorem 13.17 of [Penrose (2003)]) is much higher for spaces of $d = 3$ than that shown in Figure 172 for $d = 2$ for every value of the simulated number of nodes, with the probability being higher for the toroidal space. That probability increases as the number of nodes increases for both flat and toroidal spaces.[54] For the largest number of nodes simulated, the probability of becoming fully connected when the last isolated node disappears is approximately 0.885 for RGGs in a flat space and approximately 0.934 for RGGs in a toroidal space.

Figure 179 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in three-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. The ratios for both flat and toroidal spaces are only slightly greater than 1 for the largest values of n simulated and they approach 1 even more closely for the smaller values of n.

[54] The probability that an RGG in a flat or toroidal unit square in 3D with r-for-given-n growth becomes fully connected when the last isolated node is absorbed does not appear to decrease and then increase as n increases as noted in footnote #45 for RGGs in 2D; however, the two graphs appear to have a flat interval at the smallest values of n.

The relative errors of the LRMs given in (160), (161), (162), and (163) are shown against those of Penrose's predicted ranges in Figure 180, Figure 181, Figure 182, and Figure 183, respectively. The LRMs outperform Penrose's Theorem's predictions of the 99%ile range for RGGs in flat space. Both the LRMs and Penrose's Theorem have very small relative errors for RGGs in toroidal space for both full connectivity and no isolated nodes but Penrose's Theorem has consistently better performance for the largest numbers of nodes simulated.

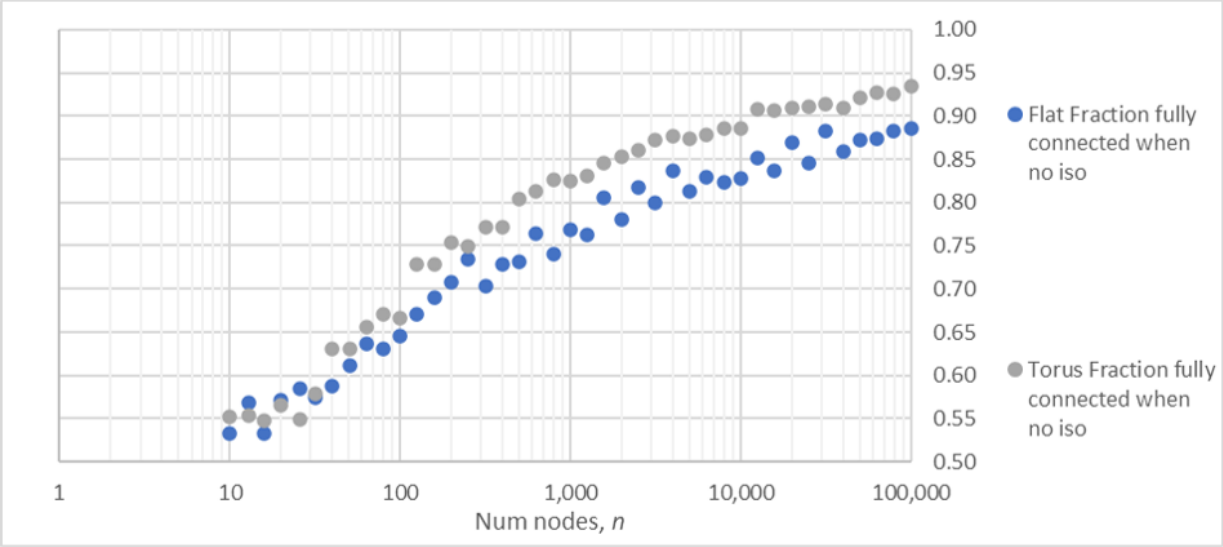

Figure 178: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 3D to become fully connected with r-for-given-n growth.

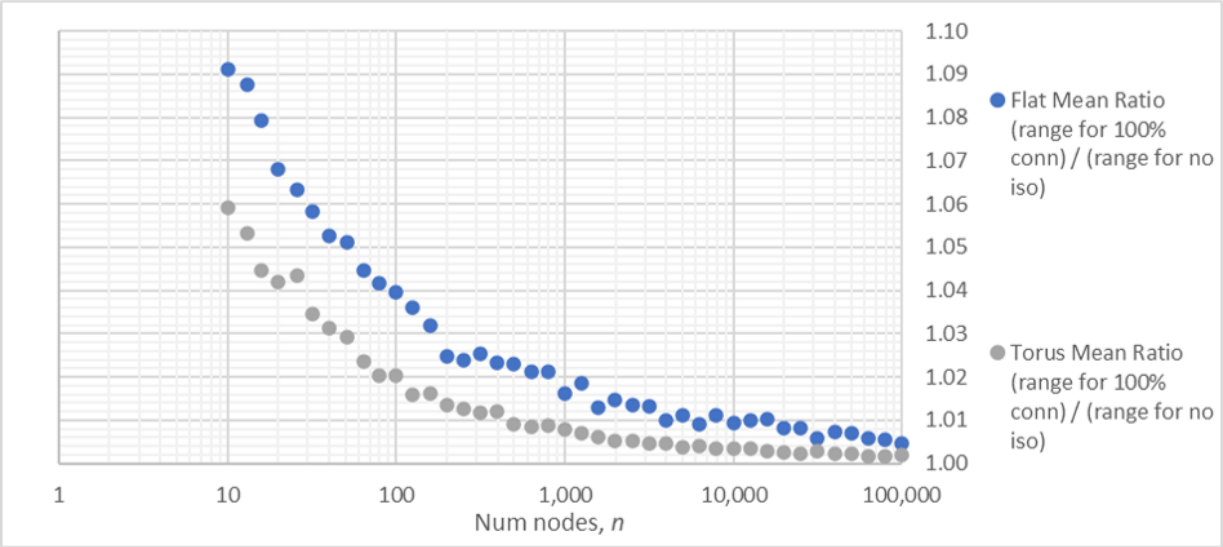

Figure 179: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 3D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

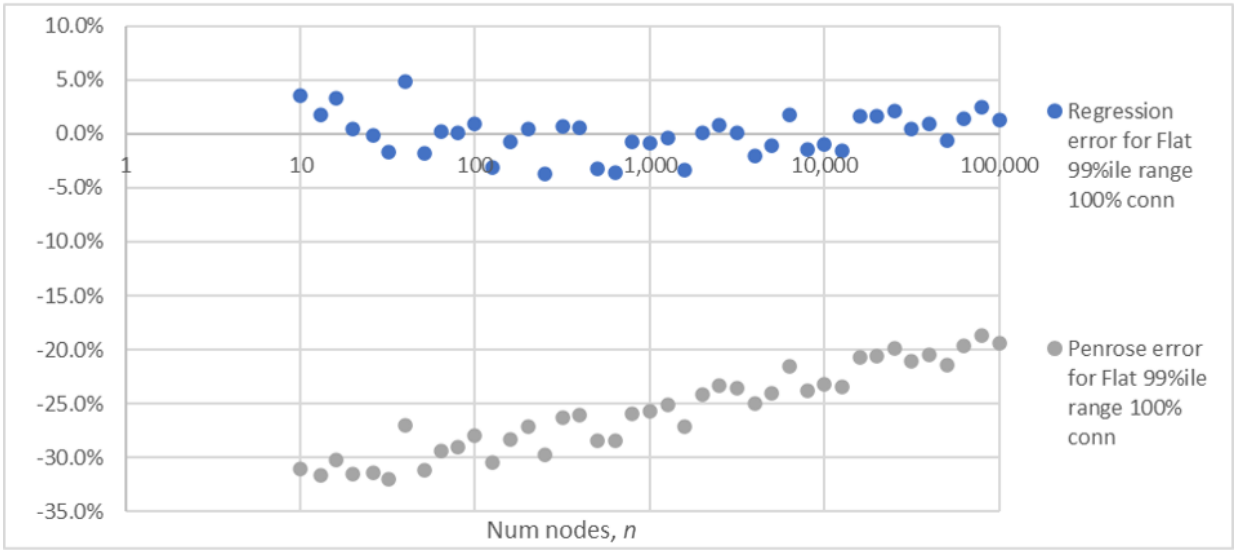

Figure 180: Relative error of LRM (160) for 99%ile range for full connectivity of RGGs in a flat unit square in 3D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

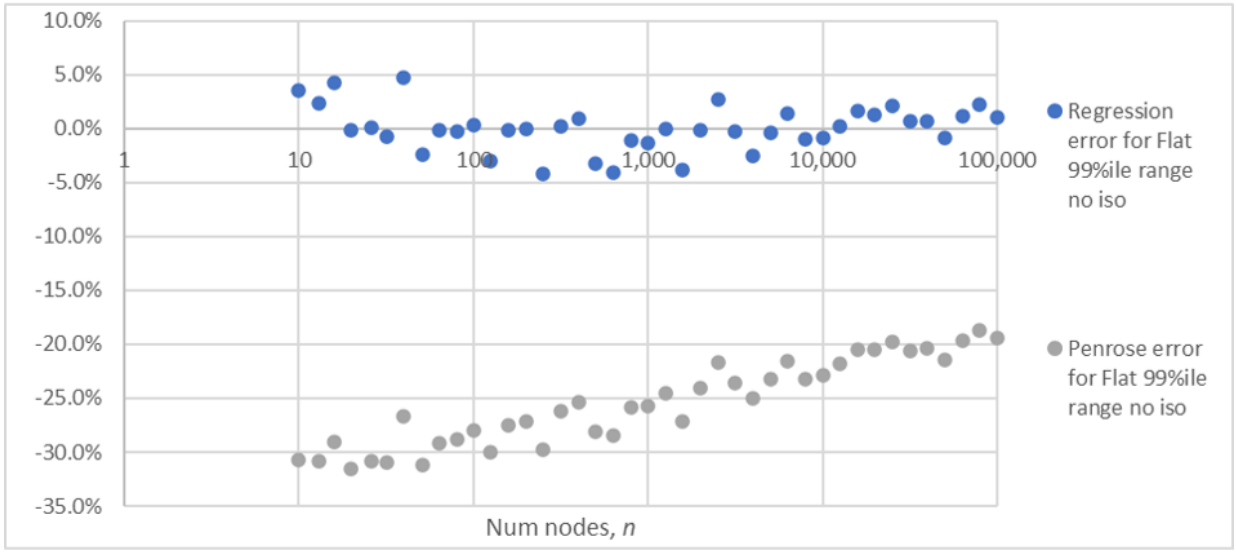

Figure 181: Relative error of LRM (161) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 3D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

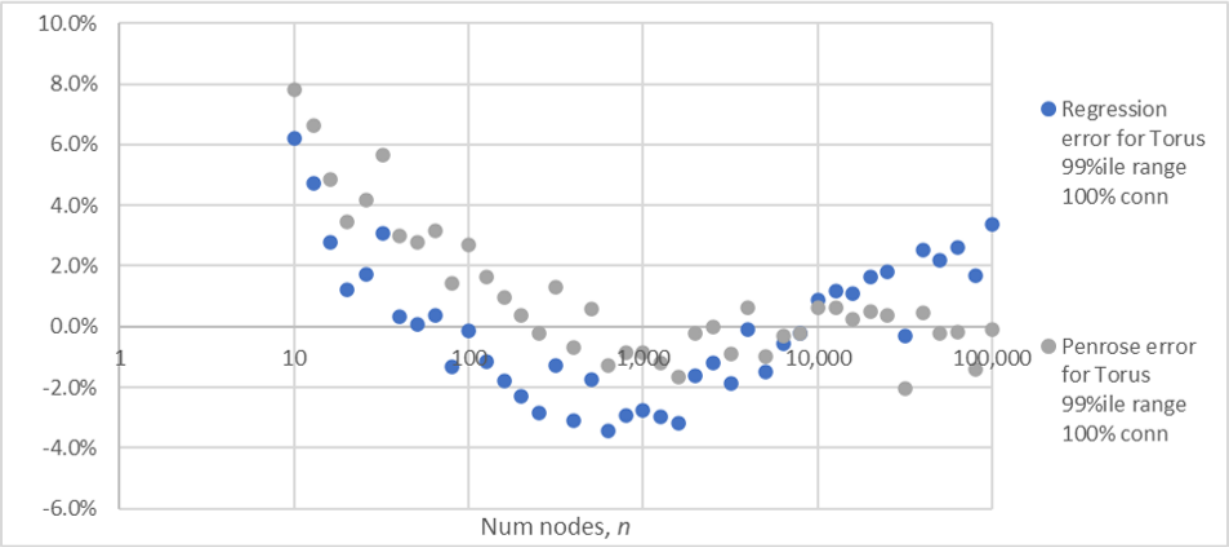

Figure 182: Relative error of LRM (162) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 3D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

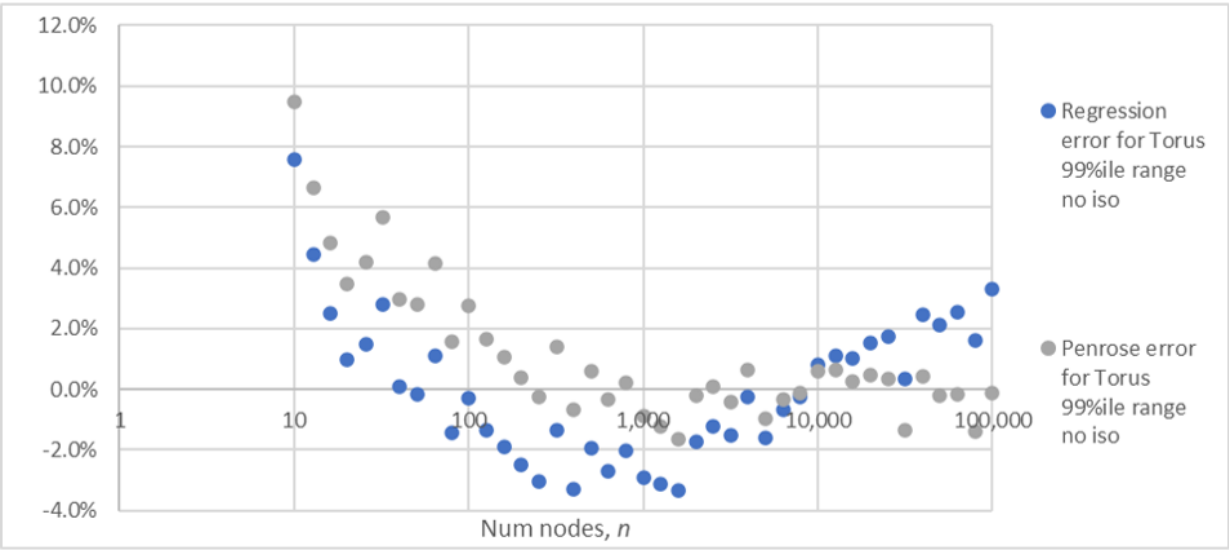

Figure 183: Relative error of LRM (163) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 3D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.7 Four-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 4$ consist of nodes in a hypercube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 1.6151\, n^{-0.2417} \tag{164}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 1.27% and max error of the value of the predicted r is 3.47%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.6088\, n^{-0.2413} \tag{165}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 1.29% and max error of the value of the predicted r is −3.65%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 1.0156\, n^{-0.2239} \tag{166}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 1.17% and max error of the value of the predicted r is 3.38%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.0128\, n^{-0.2236} \tag{167}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 1.18% and max error of the value of the predicted r is 3.49%.

The results for RGGs in four-dimensional flat and toroidal unit squares are shown in Figure 184 and Figure 185.

Figure 184 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears is again much higher than for spaces of smaller dimensions for every value of the simulated range, with the probability being higher for the toroidal space. That probability increases as the number of nodes increases for both flat and toroidal spaces.[55] For the largest number of nodes simulated, the probability of becoming fully connected when the last isolated node disappears is approximately 0.954 for RGGs in a flat space and approximately 0.975 for RGGs in a toroidal space.

Figure 185 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in four-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. The ratios for both flat and toroidal spaces are only slightly greater than 1 for the largest values of n simulated and they approach 1 even more closely for the smaller values of n.

[55] The probability that an RGG in a 4D or higher dimensional flat or toroidal unit square with r-for-given-n growth becomes fully connected when the last isolated node is absorbed does not appear to decrease and then increase as n increases as noted in footnote #45 for RGGs in 2D, nor does it exhibit a flat interval as observed for the 3D cases.

The relative errors of the LRMs given in (164), (165), (166), and (167) are shown against those of Penrose's predicted ranges in Figure 186, Figure 187, Figure 188, and Figure 189, respectively. The LRMs outperform Penrose's Theorem's predictions of the 99%ile range for RGGs in flat space. Both the LRMs and Penrose's Theorem have very small relative errors for RGGs in toroidal space for both full connectivity and no isolated nodes but Penrose's Theorem has consistently better performance for the largest numbers of nodes simulated.

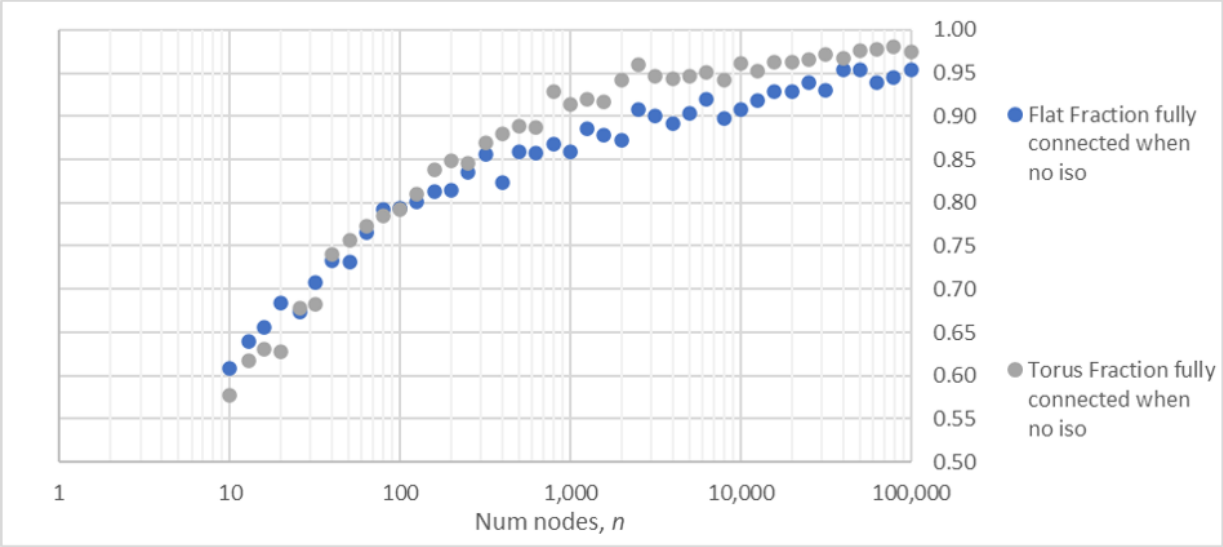

Figure 184: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 4D to become fully connected with r-for-given-n growth.

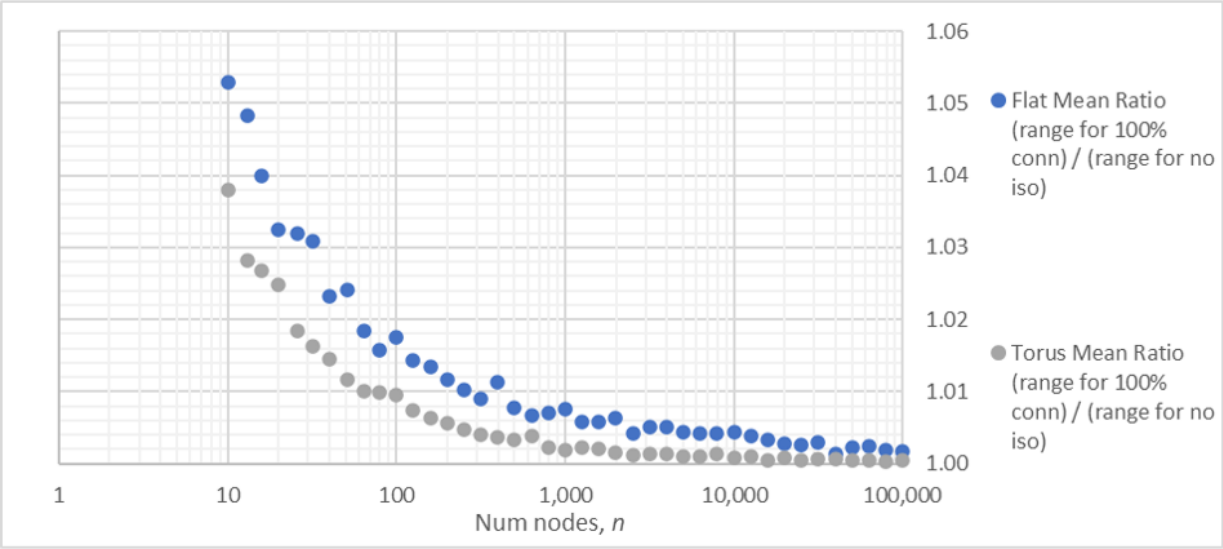

Figure 185: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 4D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

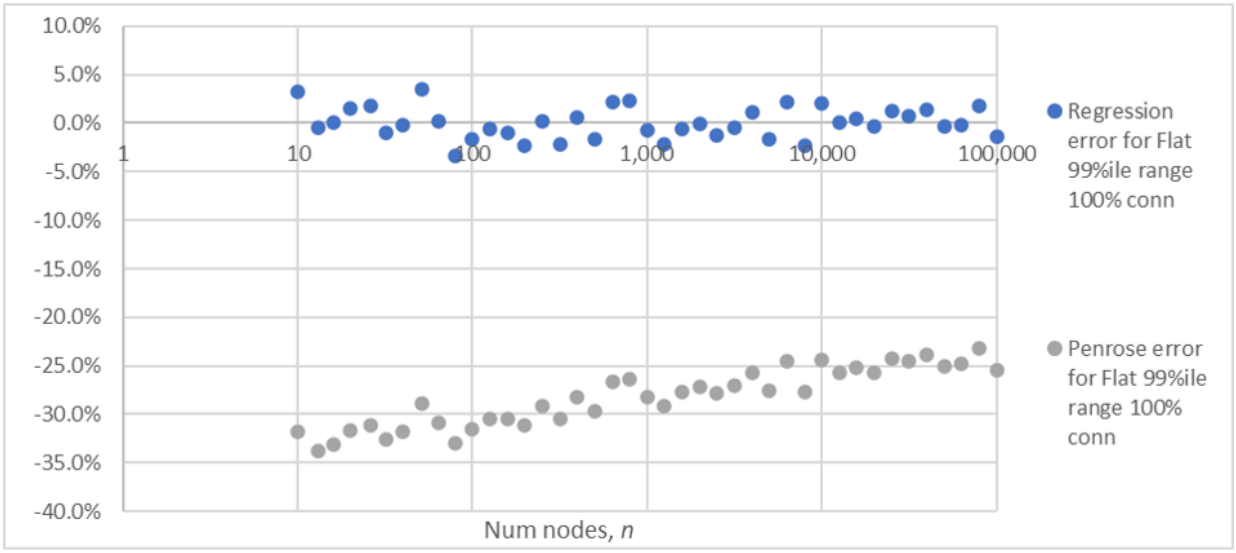

Figure 186: Relative error of LRM (164) for 99%ile range for full connectivity of RGGs in a flat unit square in 4D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

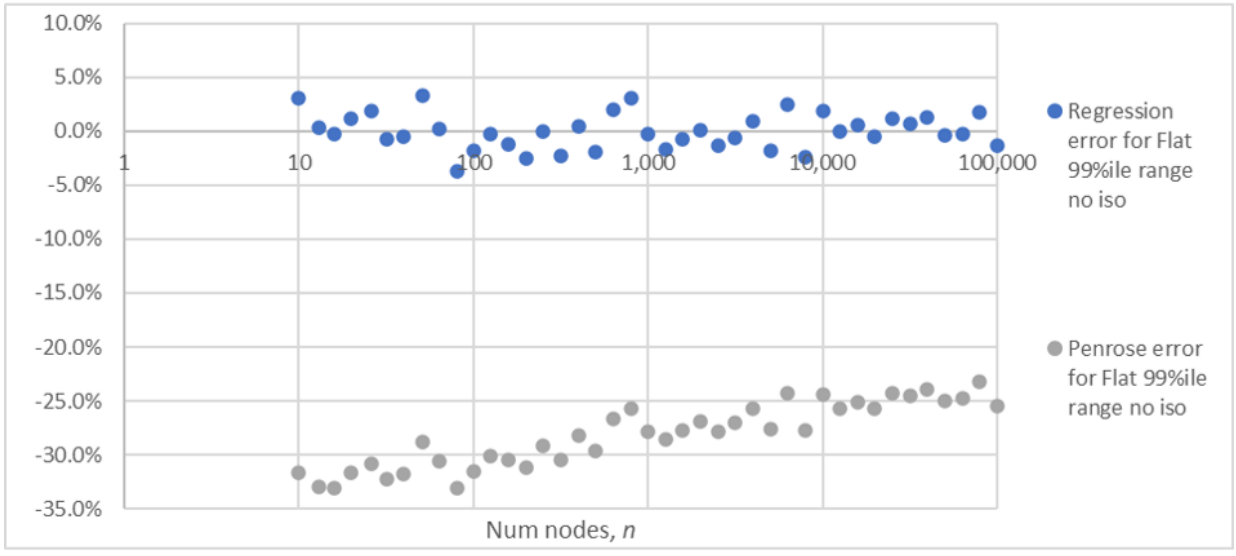

Figure 187: Relative error of LRM (165) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 4D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

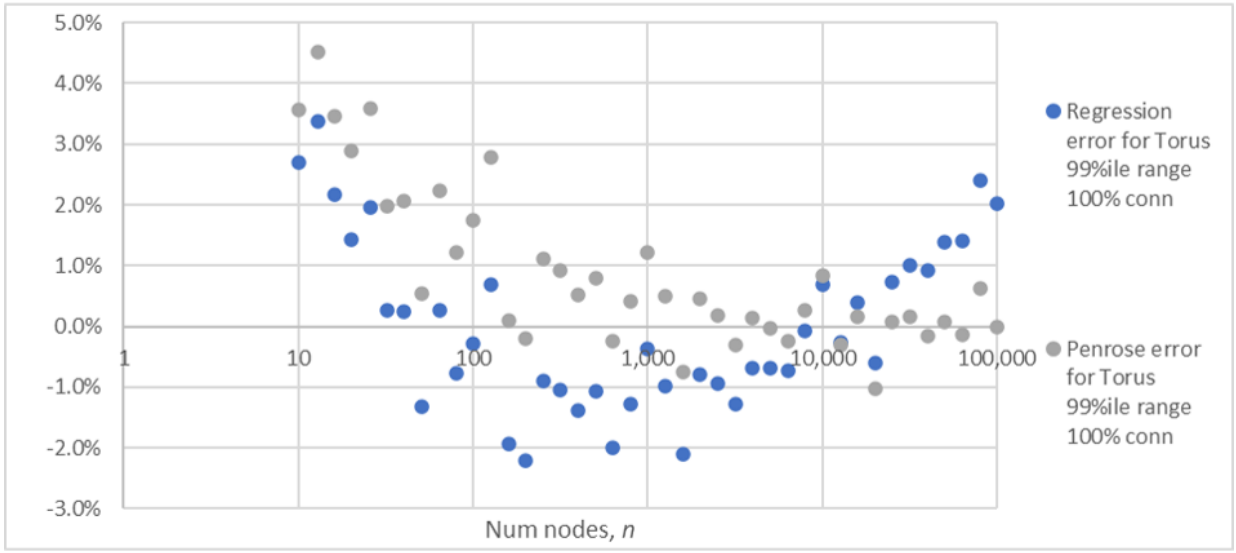

Figure 188: Relative error of LRM (166) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 4D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

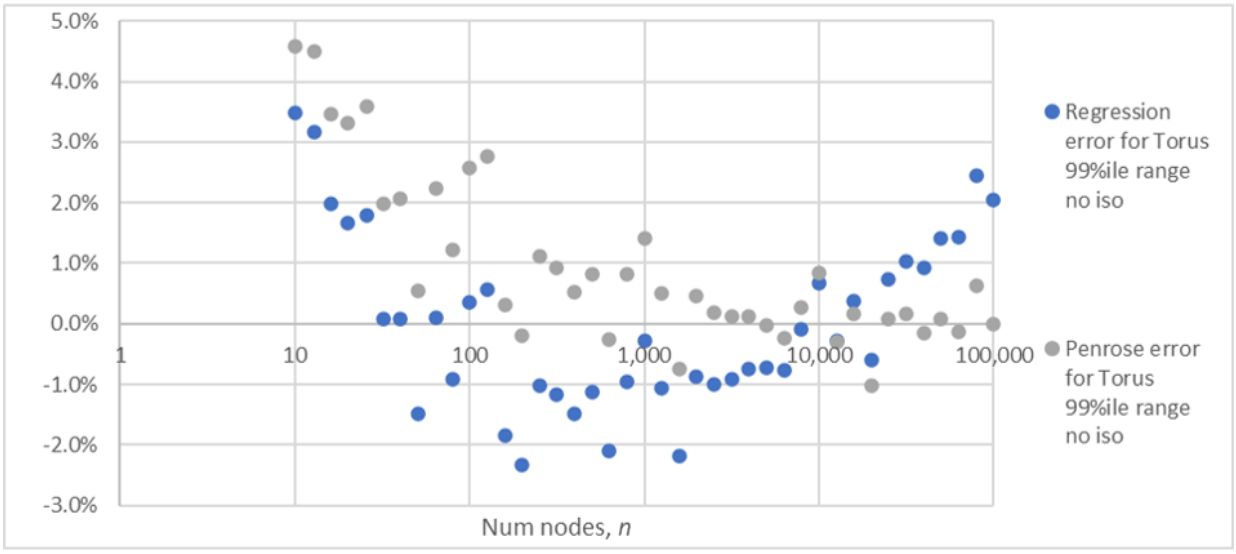

Figure 189: Relative error of LRM (167) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 4D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.8 Five-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 5$ consist of nodes in a hypercube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 1.6003\, n^{-0.1952} \tag{168}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 0.87% and max error of the value of the predicted r is -2.72%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.5982\, n^{-0.1951} \tag{169}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 0.86% and max error of the value of the predicted r is -2.79%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 1.0058\, n^{-0.1799} \tag{170}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 0.80% and max error of the value of the predicted r is 1.91%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.0042\, n^{-0.1797} \tag{171}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 0.80% and max error of the value of the predicted r is 2.19%.

The results for RGGs in five-dimensional flat and toroidal unit squares are shown in Figure 190 and Figure 191.

Figure 190 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears is again much higher than for spaces of smaller dimensions for every value of the simulated range, with the probability being higher for the toroidal space. That probability increases as the number of nodes increases for both flat and toroidal spaces. For the largest number of nodes simulated, the probability of becoming fully connected when the last isolated node disappears is approximately 0.969 for RGGs in a flat space and approximately 0.994 for RGGs in a toroidal space.

Figure 191 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in five-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. The ratios for both flat and toroidal spaces are only slightly greater than 1 for the largest values of n simulated and they approach 1 even more closely for the smaller values of n.

The relative errors of the LRMs given in (168), (169), (170), and (171) are shown against those of Penrose's predicted ranges in Figure 192, Figure 193, Figure 194, and Figure 195, respectively. The LRMs outperform Penrose's Theorem's predictions of the 99%ile range for RGGs in flat space. Both the LRMs and Penrose's Theorem have very small relative errors for RGGs in toroidal space for both full connectivity and no isolated nodes but Penrose's Theorem has consistently better performance for the largest numbers of nodes simulated.

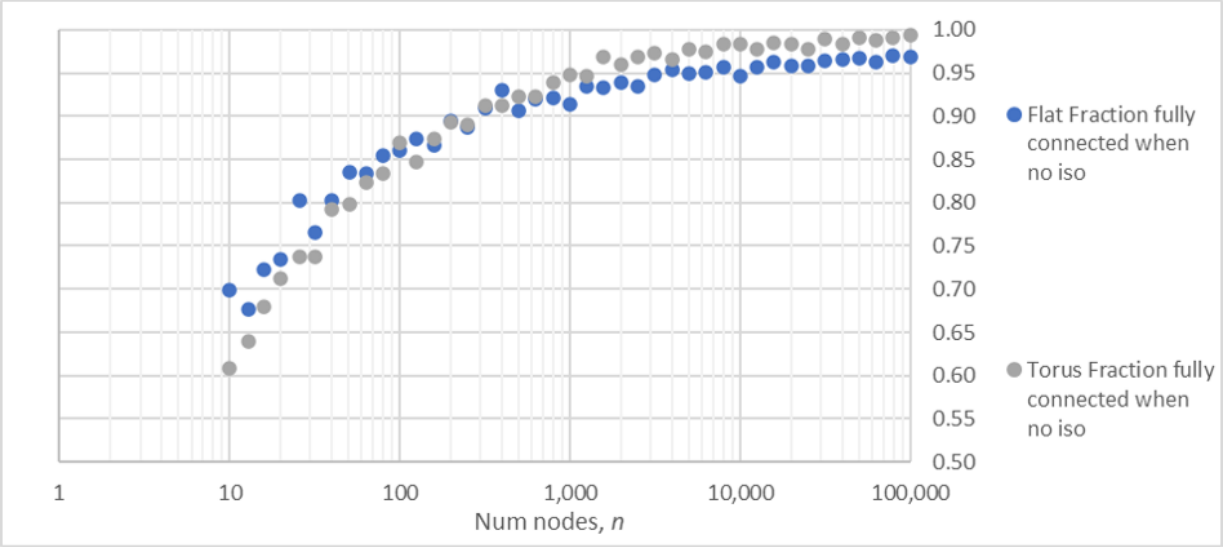

Figure 190: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 5D to become fully connected with r-for-given-n growth.

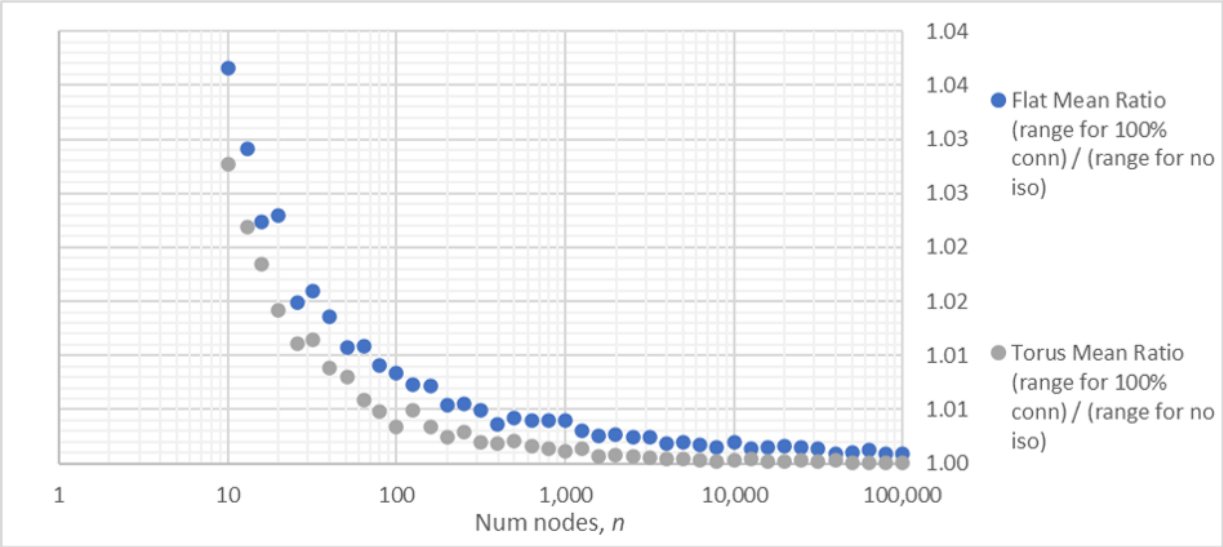

Figure 191: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 5D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

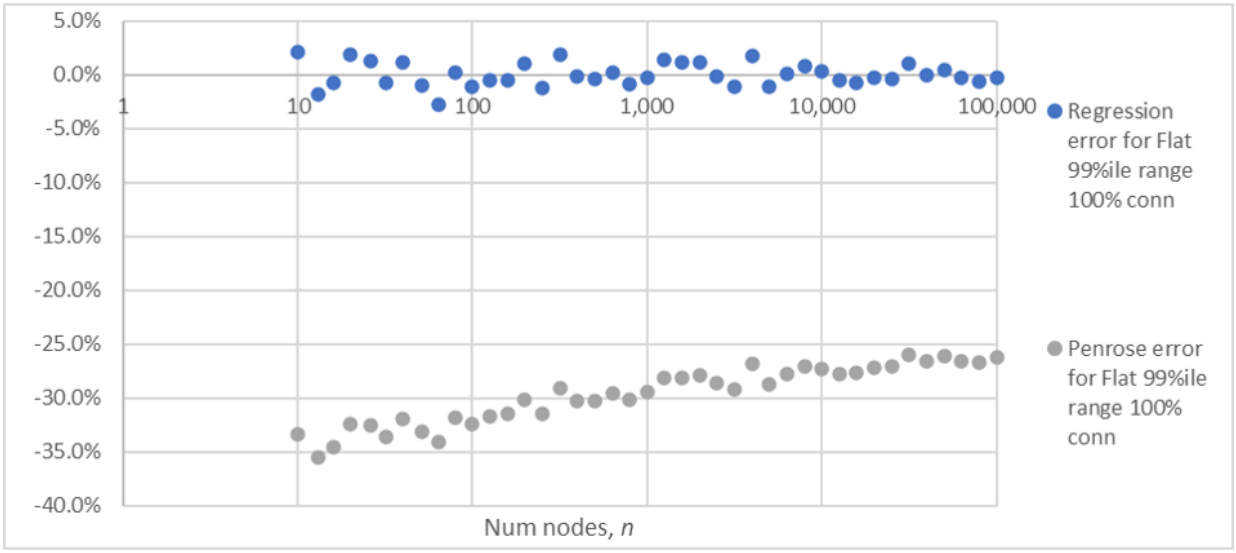

Figure 192: Relative error of LRM (168) for 99%ile range for full connectivity of RGGs in a flat unit square in 5D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

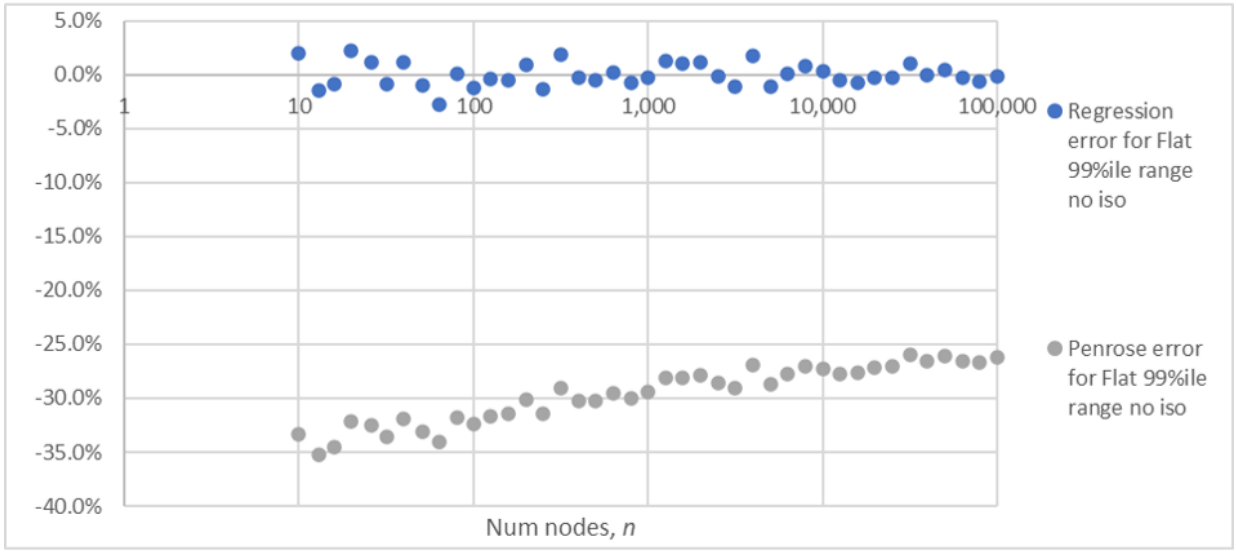

Figure 193: Relative error of LRM (169) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 5D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

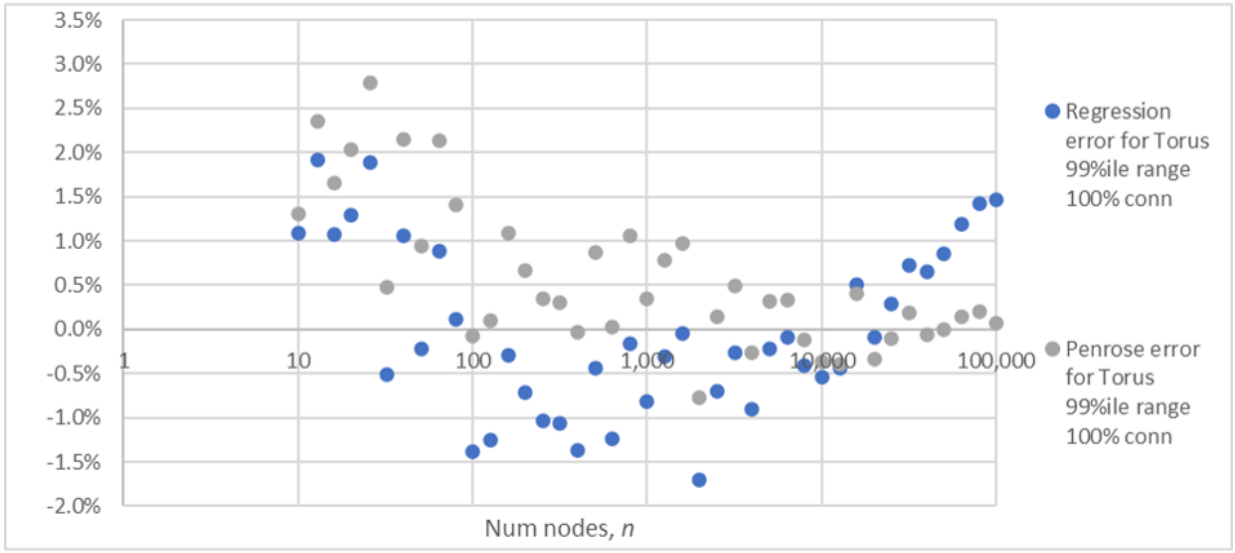

Figure 194: Relative error of LRM (170) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 5D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

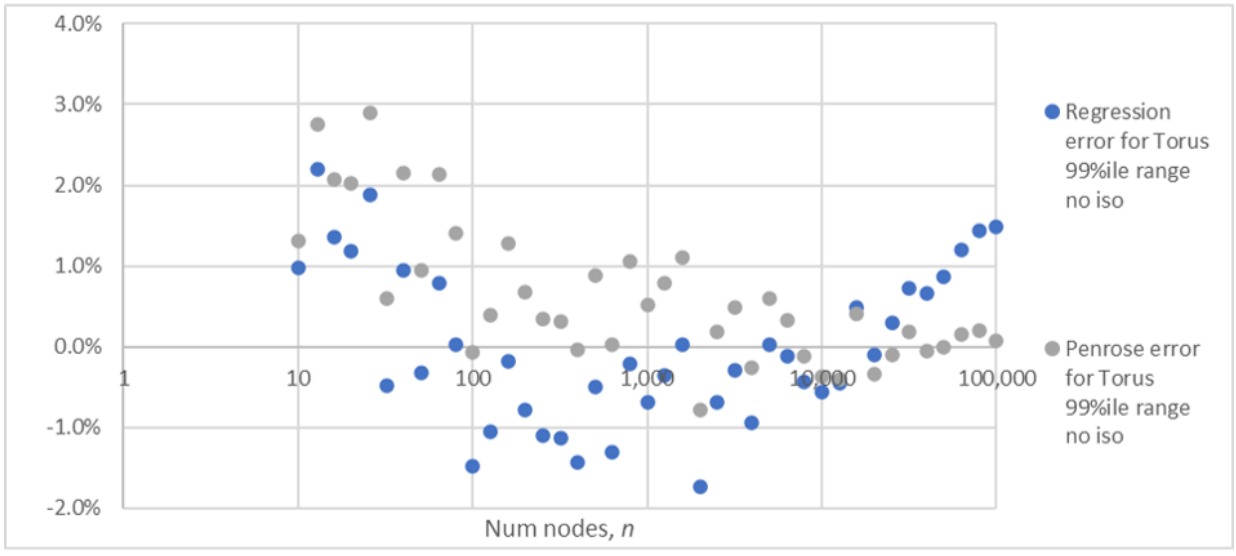

Figure 195: Relative error of LRM (171) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 5D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.9 Six-Dimensional Spaces

RGGs in a "unit square" in spaces of dimension $d = 6$ consist of nodes in a hypercube with sides of length $a = 1$. RGGs in flat and toroidal spaces were simulated for $N = 1{,}000$ randomly generated geometries with r-for-given-n growth for the small-n set.

The plots of the 99%ile range are not shown in this section but are included in Section 7.2.10 and compared to those for the other dimensions simulated. The plots and their trend lines for both full connectivity and no isolated nodes showed nearly straight lines when plotted on log-log scales so the power law fit is a good approximation for the LRMs.

For a flat space and the 99%ile range for full connectivity with r-for-given-n growth, (83) becomes

$$r = 1.5607\, n^{-0.1600} \tag{172}$$

with an R^2 value of 0.9986. For the simulated values of r, MAPE is 1.01% and max error of the value of the predicted r is 7.97%.

For a flat space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.5605\, n^{-0.1600} \tag{173}$$

with an R^2 value of 0.9986. For the simulated values of r, MAPE is 1.01% and max error of the value of the predicted r is 7.96%.

For a toroidal space and the 99%ile range for full connectivity with r-for-given-n growth, the LRM is

$$r = 1.0137\, n^{-0.1507} \tag{174}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 0.46% and max error of the value of the predicted r is 1.21%.

For a toroidal space and the 99%ile range for no isolated nodes with r-for-given-n growth, the LRM is

$$r = 1.0130\, n^{-0.1506} \tag{175}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 0.48% and max error of the value of the predicted r is 1.38%.

The results for RGGs in six-dimensional flat and toroidal unit squares are shown in Figure 196 and Figure 197.

Figure 196 shows that the probability that an RGG is connected "at the same instant" that the last isolated node disappears is again much higher than for spaces of smaller dimensions for every value of the simulated range, with the probability being higher for the toroidal space. That probability increases as the number of nodes increases for both flat and toroidal spaces. For the largest number of nodes simulated, the probability of becoming fully connected when the last isolated node disappears is approximately 0.992 for RGGs in a flat space and approximately 0.995 for RGGs in a toroidal space.

Figure 197 shows the mean ratios of the range required for full connectivity to the range required for isolated nodes to disappear in flat and toroidal unit squares in six-dimensional spaces for RGGs with r-for-given-n growth. The mean is taken over all of the ratios for a given value of the number of nodes for each simulated geometry. The ratios for both flat and toroidal spaces are only slightly greater than 1 for the largest values of n simulated and they approach 1 even more closely for the smaller values of n.

The relative errors of the LRMs given in (172), (173), (174), and (175) are shown against those of Penrose's predicted ranges in Figure 198, Figure 199, Figure 200, and Figure 201, respectively. The LRMs outperform Penrose's Theorem's predictions of the 99%ile range for RGGs in flat space. Both the LRMs and Penrose's Theorem have very small relative errors for RGGs in toroidal space for both full connectivity and no isolated nodes and their performance is nearly indistinguishable.

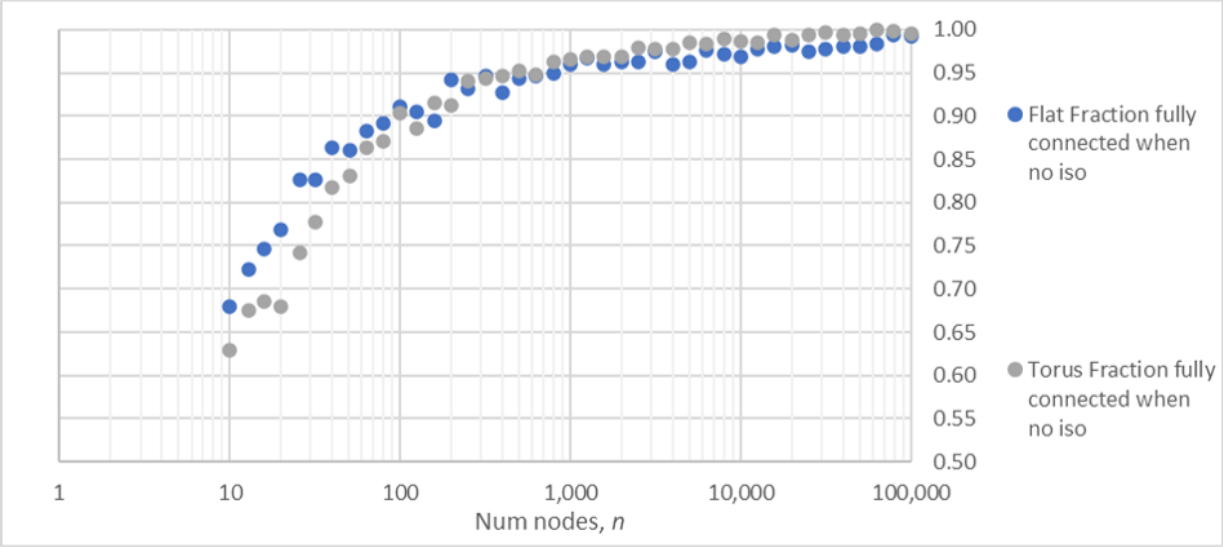

Figure 196: Fraction of simulations for which increasing the range to the point that causes the last isolated node to become connected also causes RGGs in flat and toroidal unit squares in 6D to become fully connected with r-for-given-n growth.

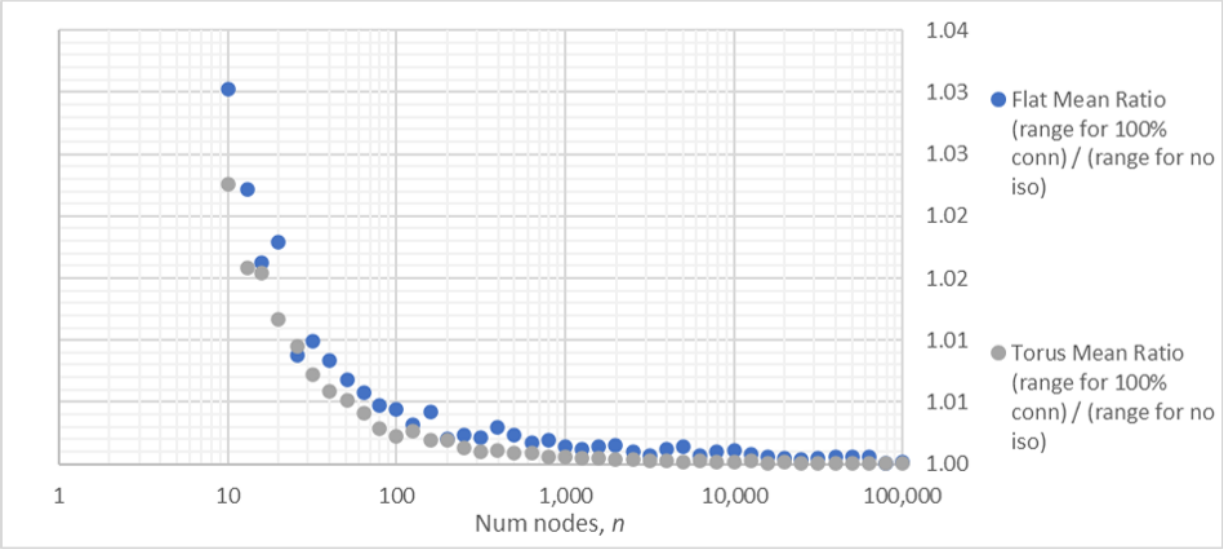

Figure 197: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in flat and toroidal unit squares in 6D with r-for-given-n growth. Mean is taken over all of the ratios for a given value of the range for each simulated geometry.

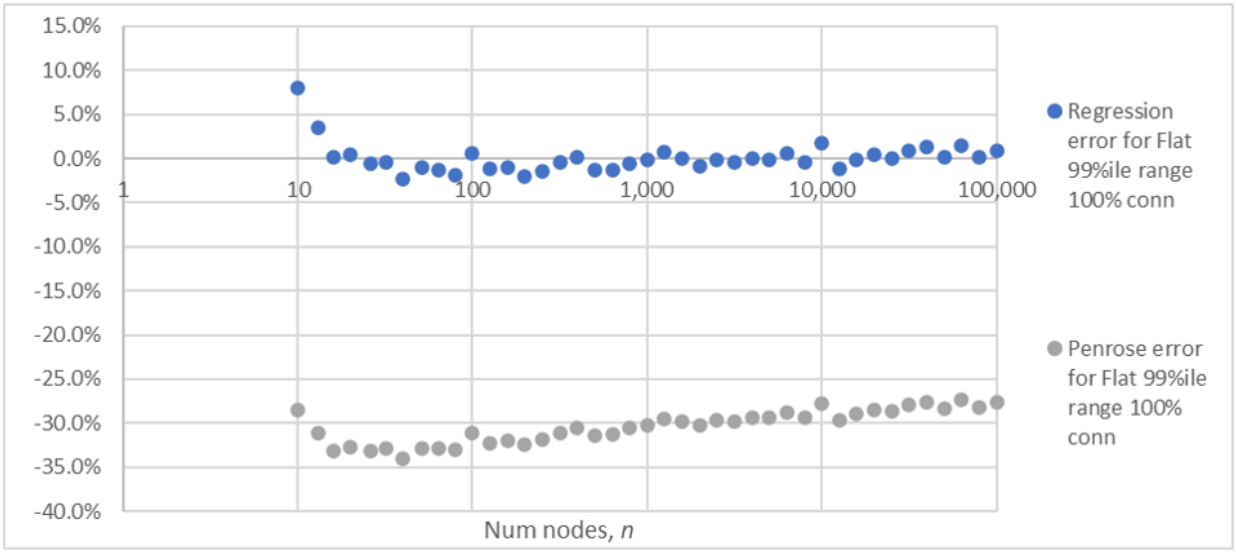

Figure 198: Relative error of LRM (172) for 99%ile range for full connectivity of RGGs in a flat unit square in 6D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

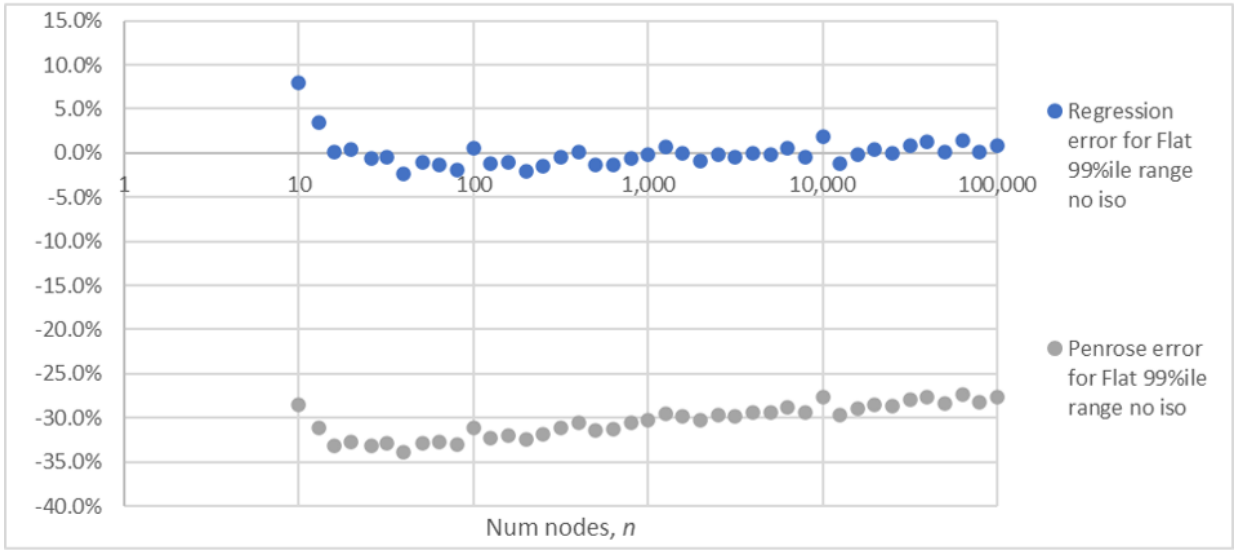

Figure 199: Relative error of LRM (173) for 99%ile range for no isolated nodes in RGGs in a flat unit square in 6D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

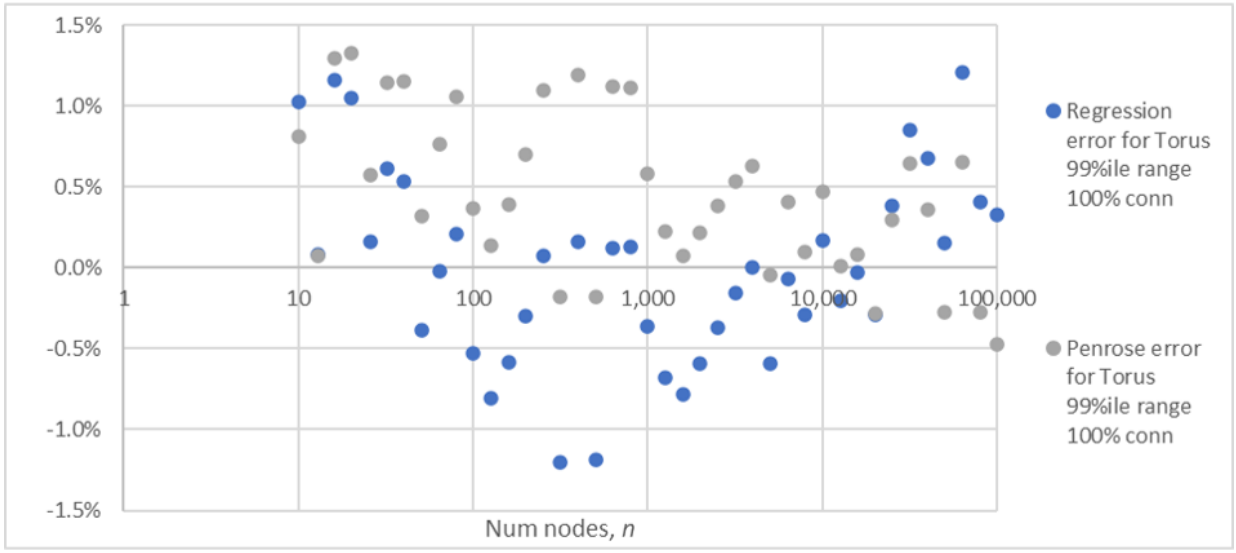

Figure 200: Relative error of LRM (174) for 99%ile range for full connectivity of RGGs in a toroidal unit square in 6D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

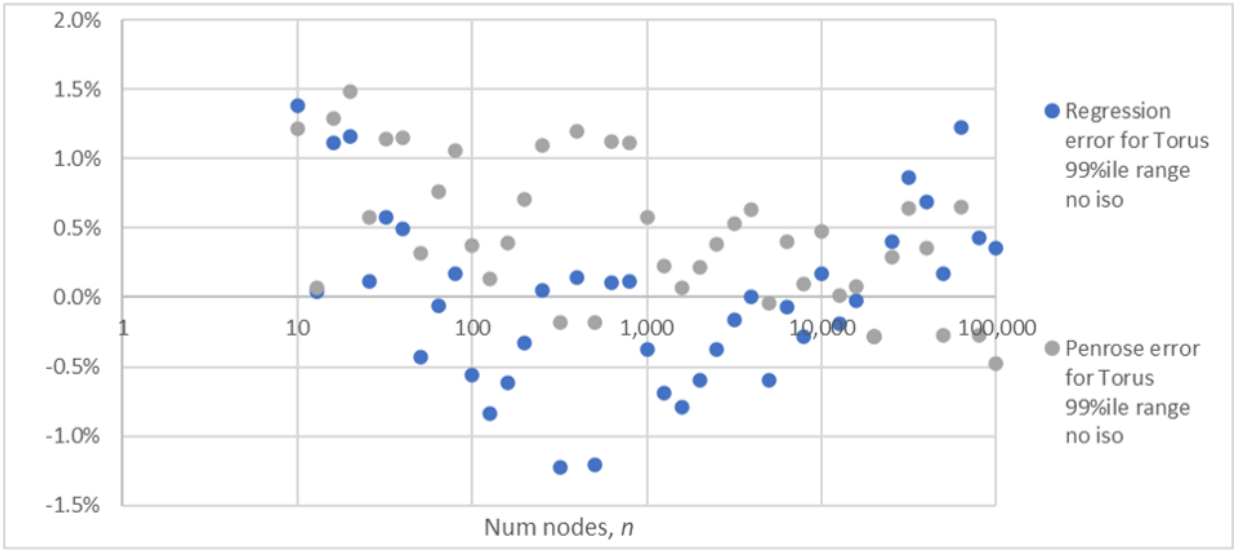

Figure 201: Relative error of LRM (175) for 99%ile range for no isolated nodes in RGGs in a toroidal unit square in 6D with r-for-given-n growth. Relative error is shown for the predicted r given by Penrose's Theorem with $C = 0.99$.

7.2.10 Comparing Ranges Required in d-Dimensional Spaces

Figure 202, Figure 203, Figure 204, and Figure 205 summarize, by type of connectivity, the LRMs presented by dimension in Section 7.2.4, Section 7.2.5, Section 7.2.6, Section 7.2.7, Section 7.2.8, and Section 7.2.9. These plots demonstrate that, as d increases, for r-for-given-n growth there is a huge increase in the ranges required to first achieve the connectivity conditions of RGGs being either fully connected or having no isolated nodes in both flat unit squares and toroidal unit squares.

As noted in Section 7.1.5 for RGGs with n-for-given-r growth, the increase in the range required for connectivity in RGGs with r-for-given-n growth is a consequence of the fact that a given number of nodes in a space spread out more as the dimensionality of a space increases. Thus, as considered by Penrose in the derivation of his theorem, greater range is required to ensure the connectivity of an RGG.

While the range required to achieve connectivity increases with d, the range required to achieve no isolated nodes more closely approaches the range that achieves full connectivity for both flat and toroidal spaces. This is demonstrated graphically in Figure 206, Figure 207, Figure 208, and Figure 209. This phenomenon is also shown by the LRMs presented in Section 7.2.4, Section 7.2.5, Section 7.2.6, Section 7.2.7, Section 7.2.8, and Section 7.2.9 by the fact that the respective m and b coefficients in the expressions for the 99%ile range for full connectivity and no isolated nodes become closer as the dimensionality increases.

Figure 206 shows the ratio of the 99%ile range required for achieve fully connected RGGs to the 99%ile range required to achieve no isolated nodes in RGGs in flat spaces in dimensions one through six and Figure 207 shows the same ratios for RGGs in toroidal spaces. These figures show that RGGs in one-dimensional spaces require an increasing fraction of range to become fully connected beyond the range where the last isolated node disappears. As n increases that fraction appears to approach approximately 2 for RGGs in flat space and approximately 1.65 for RGGs in toroidal space. These ratios are on par with those noted in Section 7.2.4 for the mean values of the ratios for the individual simulations for each configuration.

In contrast to the situation for one-dimensional spaces, the ratios of the 99%ile range required for achieve fully connected RGGs to the 99%ile range required to achieve no isolated nodes in RGGs in flat spaces in dimensions two through six are so small they have to be shown in figures Figure 206 and Figure 207 without the results for $d = 1$. For both flat and toroidal spaces, the ratios are very small even for $d = 2$ and the smallest numbers of nodes simulated, and those ratios decrease with both d and n.

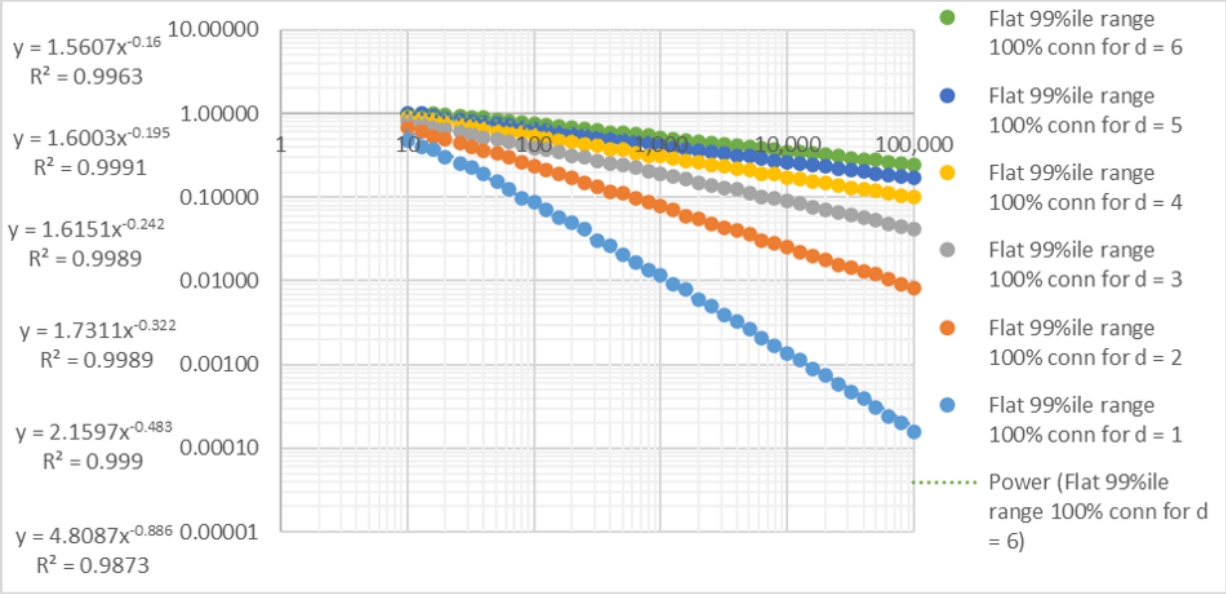

Figure 202: LRMs for 99%ile range for full connectivity of RGGs in flat unit squares in 1D through 6D for RGGs with r-for-given-n growth.

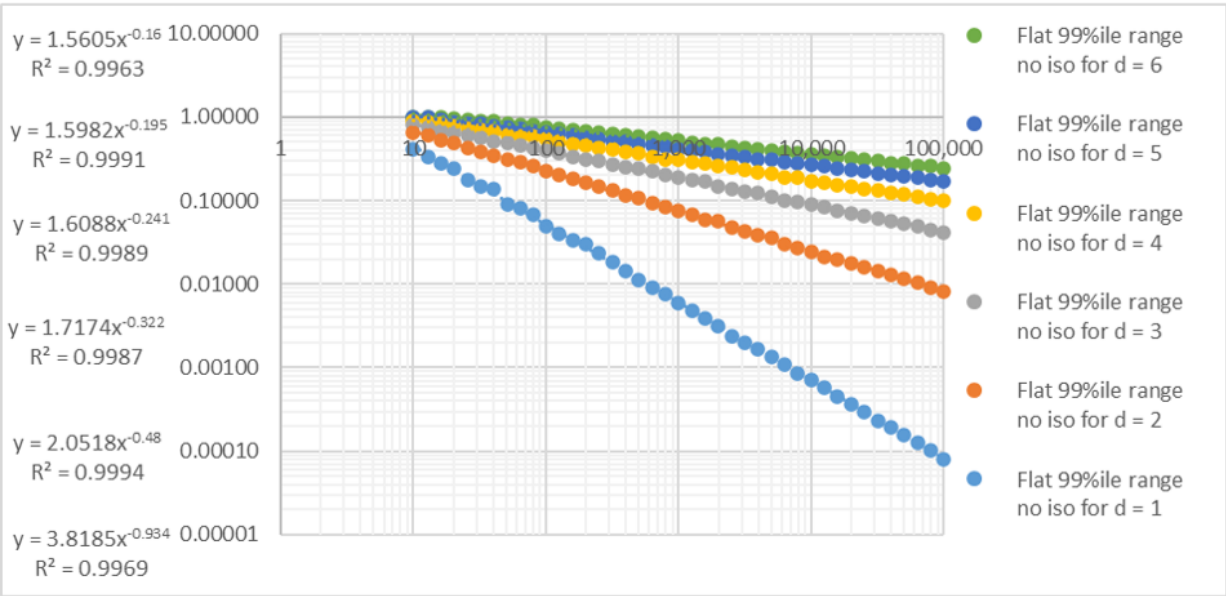

Figure 203: LRMs for 99%ile range for no isolated nodes in RGGs in flat unit squares in 1D through 6D for RGGs with r-for-given-n growth.

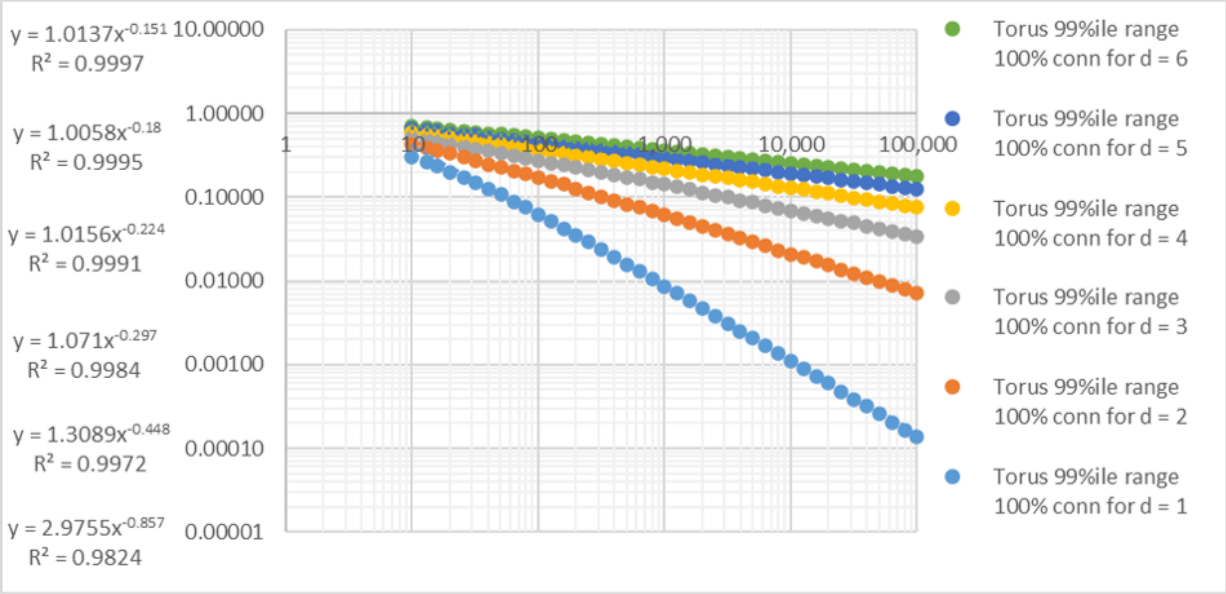

Figure 204: LRMs for 99%ile range for full connectivity of RGGs in toroidal unit squares in 1D through 6D for RGGs with r-for-given-n growth.

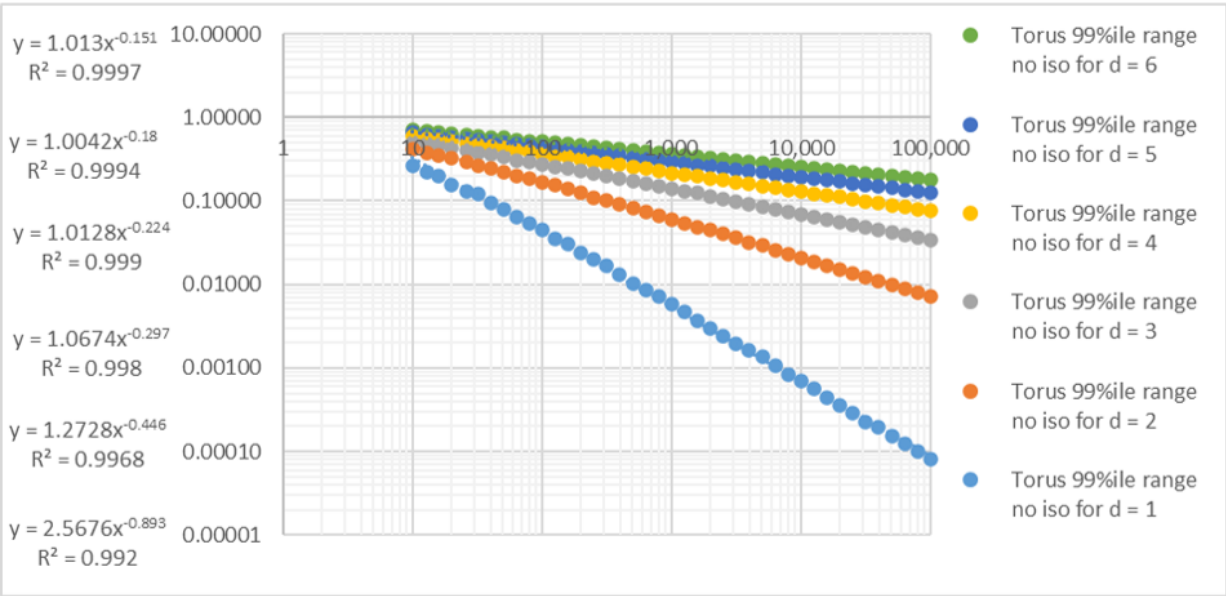

Figure 205: LRMs for 99%ile range for no isolated nodes in RGGs in toroidal unit squares in 1D through 6D for RGGs with r-for-given-n growth.

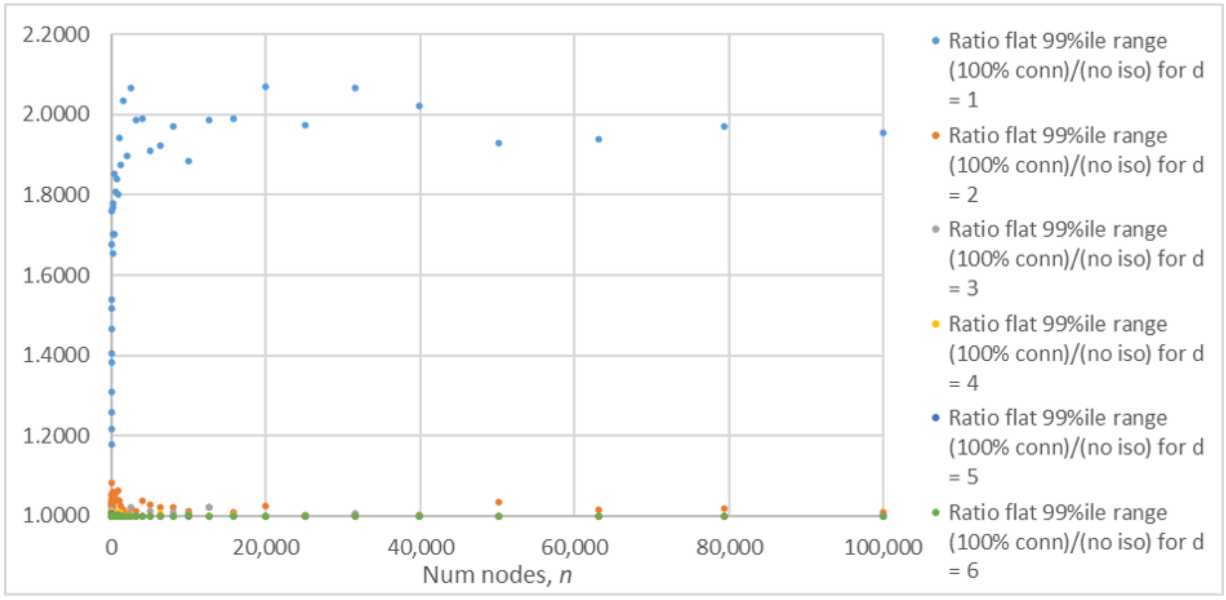

Figure 206: Ratios of 99%ile range for full connectivity to 99%ile range for no isolated nodes in RGGs in flat unit squares in 1D through 6D for RGGs with r-for-given-n growth.

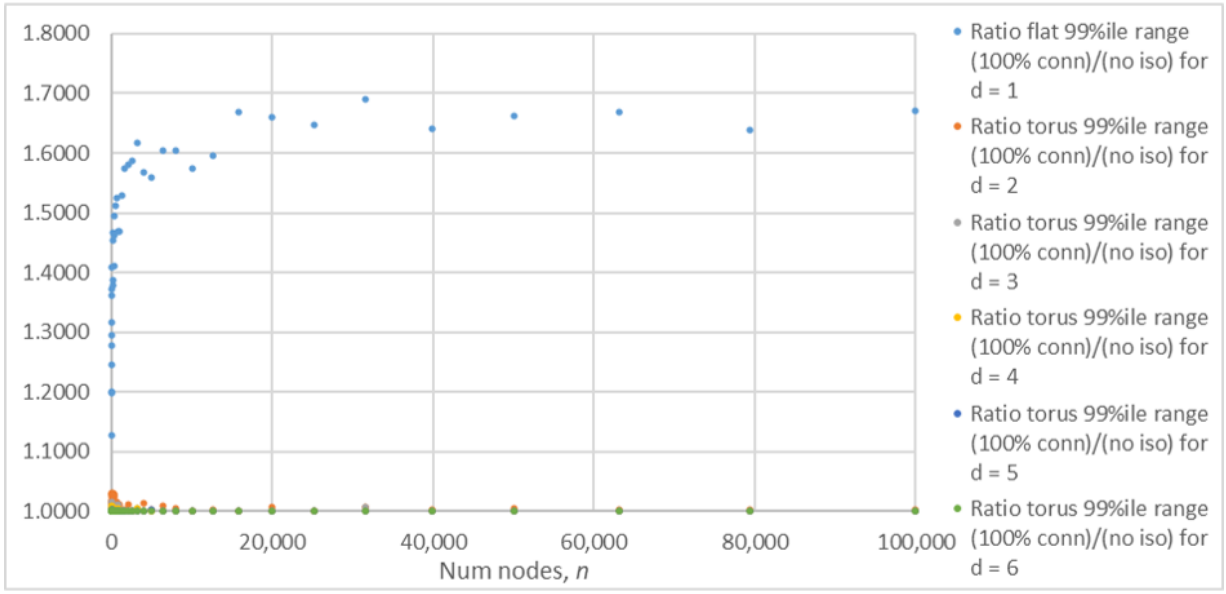

Figure 207: Ratios of 99%ile range for full connectivity to 99%ile range for no isolated nodes in RGGs in toroidal unit squares in 1D through 6D for RGGs with r-for-given-n growth.

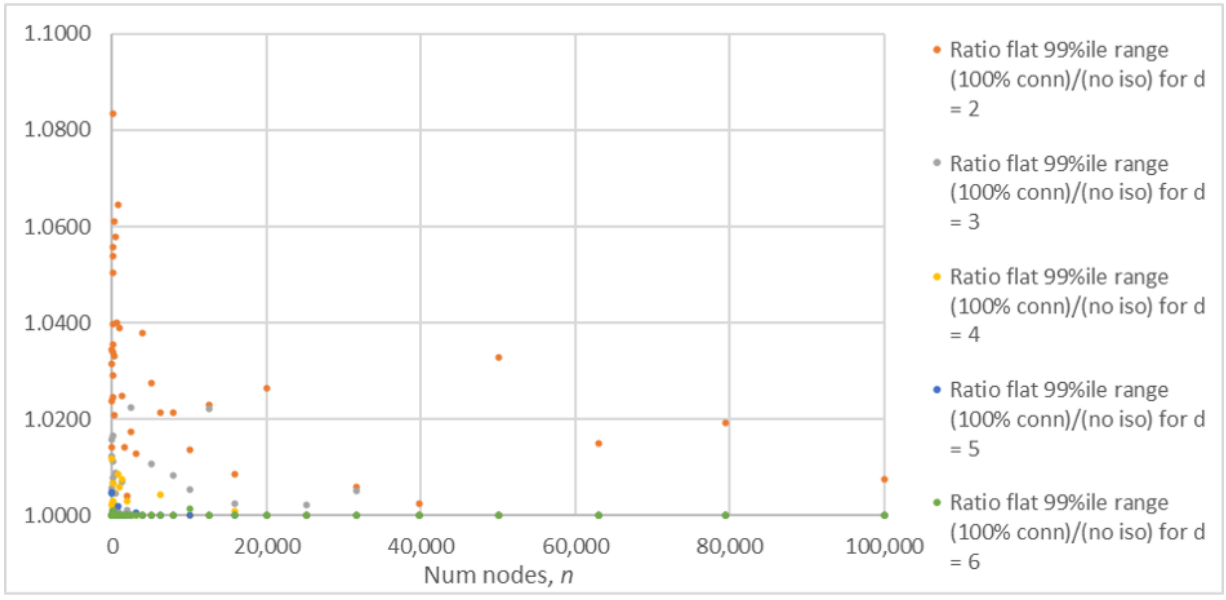

Figure 208: Same as Figure 206 for flat unit squares but omitting 1D to show the 2D through 6D results more clearly.

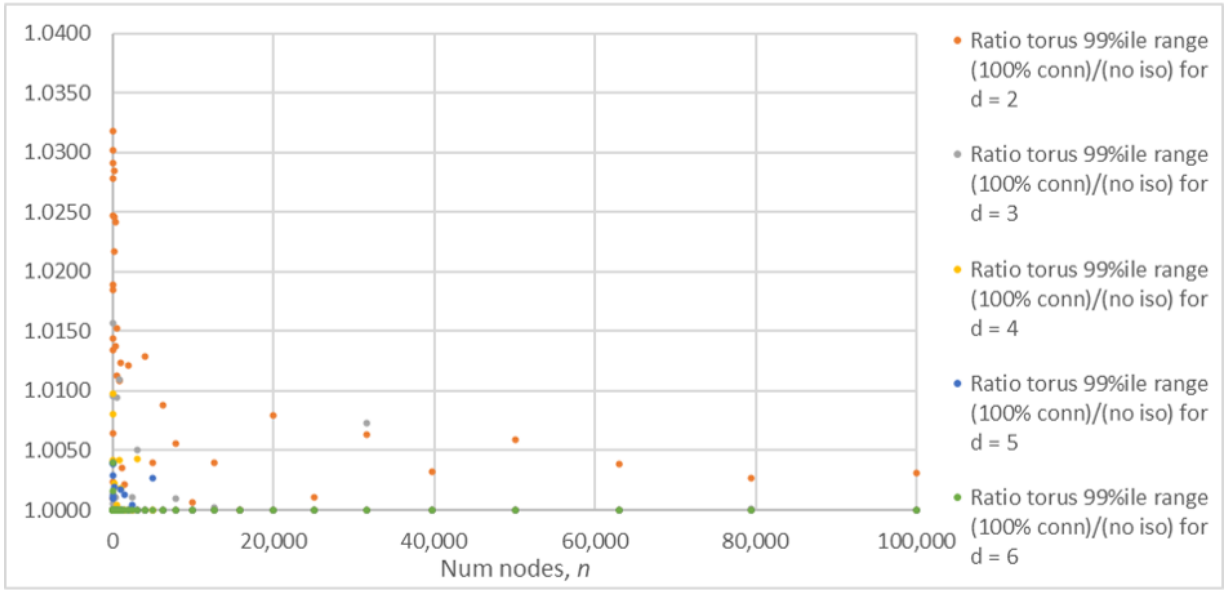

Figure 209: Same as Figure 207 for toroidal unit squares but omitting 1D to show the 2D through 6D results more clearly.

8 RGGs in Flat Rectangular Spaces with n-for-given-r Growth

Most prior research on RGGs has considered nodal spaces consisting of simple symmetric shapes such as squares and circles. Only a small amount of the literature has reported results concerning the properties of RGGs in spaces of other shapes such as rectangles; for example, see [Bettstetter (2003)] [Estrada (2015)] [Estrada (2016)]. Practical applications of the theory or empirical observations of RGG connectivity must take into account more realistic generalizations of the shapes of nodal spaces.

Estrada et al. [Estrada (2016)] addressed the problem of RGG connectivity in the context of disease transmission. In that work, it was found that elongation of an area made the RGG more resilient to the propagation of epidemics. The work by Estrada et al. shows not only that the shape of the nodal space is a significant factor for connectivity but it also provides motivation to consider applications where it is of value to know the conditions that guarantee high probability of an RGG being disconnected rather than the usual goal that it be fully or at least partially connected.

In this section simulations of n-for-given-r growth quantify the changes in the number of nodes required to first achieve connectedness in flat unit rectangles of varying width and height. The mid-r set is used for this work.

LRMs are obtained for RGGs in unit rectangles with n-for-given-r growth in two-dimensional spaces. The statistical metrics are the 99%ile, mean, and 1%ile number of nodes to first achieve the conditions of full connectivity and no isolated nodes. The plots of these data and their trend lines for full connectivity show nearly straight lines when plotted on log-log scales and closely resemble those given in Section 4.2. Thus, the power law is still a good approximation for the LRMs for RGGs with these asymmetric shapes. The plots are not shown as they provide little insight beyond that offered by the prose descriptions of the results.

All of the rectangular spaces considered here have unit area or $A = 1$. Only flat unit rectangles are modeled, so they are referred to simply as *unit rectangles*.

As noted in Section 1.2 of the Introduction, unit rectangles have width a_w and height a_h where $a_w \times a_h = 1$. Without loss of generality, the larger side length is called the width. As a rectangle is anchored with its lower left-hand corner at (0, 0), the width is the rectangle's maximum extent in the x direction and the height is its extent in the y direction. LRMs are given in this section for rectangles with width $\times$ height side lengths of $2 \times 1/2$, $3 \times 1/3$, $4 \times 1/4$, and $5 \times 1/5$; unambiguously, these unit rectangles are referred to simply by their widths 2, 3, 4, and 5, respectively. The unit rectangle of width 1 is the flat unit square studied extensively in this paper, and its LRMs for full connectivity are given in Section 4.2.1.1.

The introductory material presented in Section 4 as well as Section 4.2 and its subsections regarding the simulations and their conditions applies here.

8.1 Unit Rectangle of Width 2

This section presents LRMs for RGGs in unit rectangles with width $\times$ height side lengths of $2 \times 1/2$.

For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.3904\, r^{-2.0897} \tag{176}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 3.26% and max error of the value of the predicted n is 7.92%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.1169\, r^{-2.1517} \tag{177}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.37% and max error of the value of the predicted n is 0.93%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.1536\, r^{-2.2219} \tag{178}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.96% and max error of the value of the predicted n is -2.85%.

8.2 Unit Rectangle of Width 3

This section presents LRMs for RGGs in unit rectangles with width $\times$ height side lengths of $3 \times 1/3$.

For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 4.5324\, r^{-2.0851} \tag{179}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 2.91% and max error of the value of the predicted n is -6.74%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.3087\, r^{-2.1383} \tag{180}$$

with an R^2 value of 1.0000. For the simulated values of r, MAPE is 0.75% and max error of the value of the predicted n is 2.09%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.2954\, r^{-2.2011} \tag{181}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.50% and max error of the value of the predicted n is 2.84%.

8.3 Unit Rectangle of Width 4

This section presents LRMs for RGGs in unit rectangles with width $\times$ height side lengths of $4 \times 1/4$.

For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 5.0687\, r^{-2.0651} \tag{182}$$

with an R^2 value of 0.9996. For the simulated values of r, MAPE is 3.37% and max error of the value of the predicted n is 5.85%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.5678\, r^{-2.1204} \tag{183}$$

with an R^2 value of 0.9999. For the simulated values of r, MAPE is 1.68% and max error of the value of the predicted n is -6.34%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.5144\, r^{-2.1728} \tag{184}$$

with an R^2 value of 0.9998. For the simulated values of r, MAPE is 2.29% and max error of the value of the predicted n is -4.86%.

8.4 Unit Rectangle of Width 5

This section presents LRMs for RGGs in unit rectangles with width $\times$ height side lengths of $5 \times 1/5$.

For the 99%ile number of nodes required to first achieve full connectivity with n-for-given-r growth, (21) becomes

$$n = 5.1667\, r^{-2.0625} \tag{185}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 3.41% and max error of the value of the predicted n is -10.14%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.9528\, r^{-2.0945} \tag{186}$$

with an R^2 value of 0.9995. For the simulated values of r, MAPE is 3.21% and max error of the value of the predicted n is -12.02%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.8109\, r^{-2.1389} \tag{187}$$

with an R^2 value of 0.9997. For the simulated values of r, MAPE is 3.06% and max error of the value of the predicted n is -7.66%.

8.5 Comparing Connectivity of Flat Unit Rectangles of Varying Widths to Flat Unit Squares

The key observation regarding unit rectangles of varying widths is that the power law holds for the widths modeled even when RGGs are not in a symmetric space such as a square or circle. For unit rectangles of width 2, the MAPE and max error values were quite close to those for RGGs in a flat unit square and the error values for the mean and 1%ile number of nodes were actually slightly smaller than those for a flat unit square.

As a general trend, with exceptions, the MAPE and max error for the 99%ile, mean, and 1%ile number of nodes increased as the width increased. Even so, the MAPE values for width 5 were small percentages although the max error values were much higher than those for a flat unit square.

An RGG requires more nodes to achieve a specific connectivity condition in a unit rectangle than in a flat unit square as demonstrated in the following figures.

Figure 210 shows the ratios of the 99%ile number of nodes required to first achieve full connectivity of RGGs in unit rectangles to those required for RGGs in flat unit squares. The ratios are largest (approximately 2 and greater) for large ranges and they fall toward 1.5 or 1.6 for the smallest ranges used. The ratios are larger for the larger widths.

The situation is quite different for the ratios of the mean number of nodes required to first achieve full connectivity of RGGs in unit rectangles to those required for RGGs in flat unit squares. Figure 211 shows the ratios are again largest for the large ranges but for the smaller ranges the ratios fall quickly toward 1. Still, ratios are larger for the larger widths.

The situation is different yet again for the ratios of the 1%ile number of nodes required to first achieve full connectivity of RGGs in unit rectangles to those required for RGGs in flat unit squares is yet again different from the other cases. Figure 212 shows the ratios increase as the range decreases from approximately $r = 0.0316$. For the largest ranges simulated, the behavior is mixed with the 1%ile number of nodes for widths 2 and 3 being the smallest for the largest ranges and the 1%ile number of nodes being the

largest or close to largest for widths 4 and 5. The ratios for the narrowest unit rectangles with widths 4 and 5 show non-monotonic behavior where they have at their minimum near a range between 0.02 and 0.03.

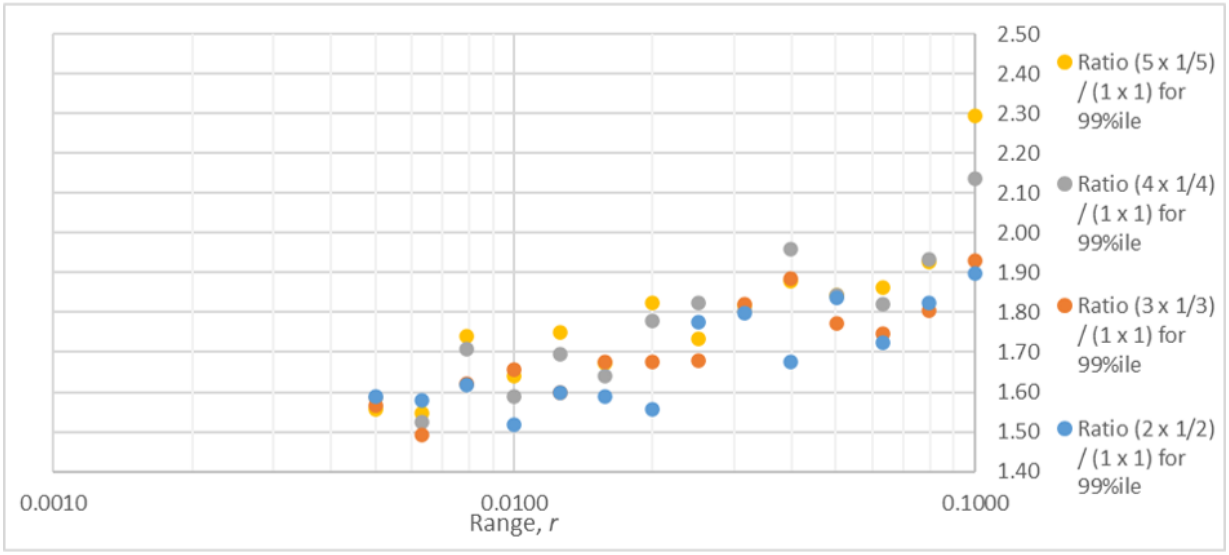

Figure 210: Ratios of 99%ile number of nodes for full connectivity of RGGs in flat unit rectangles of widths 2, 3, 4, and 5 with n-for-given-r growth to the number of nodes required for a flat unit square.

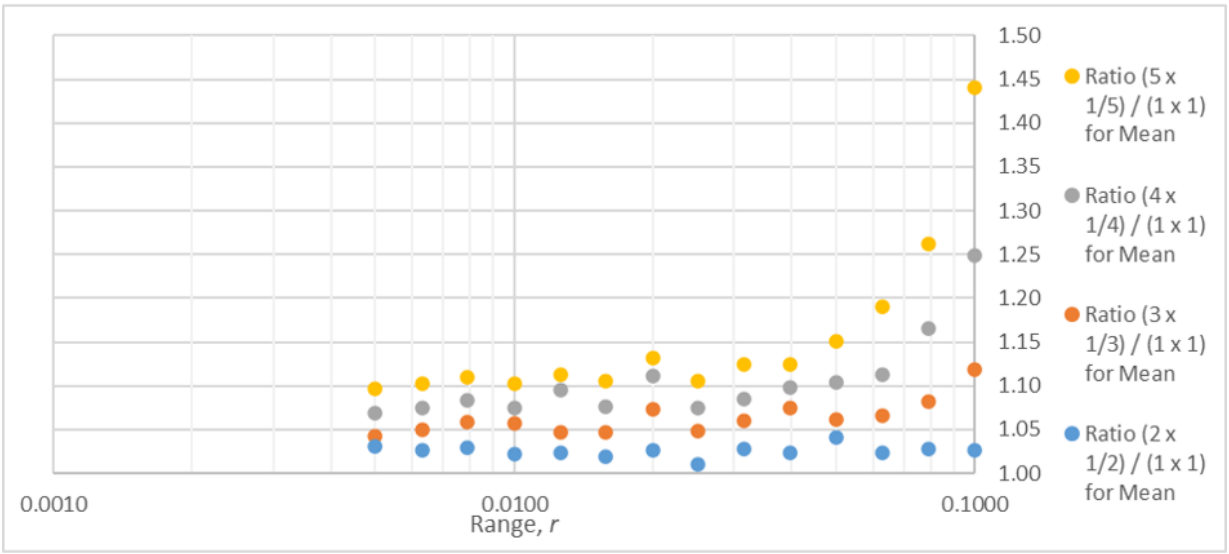

Figure 211: Ratios of mean number of nodes for full connectivity of RGGs in flat unit rectangles of widths 2, 3, 4, and 5 with n-for-given-r growth to the number of nodes required for a flat unit square.

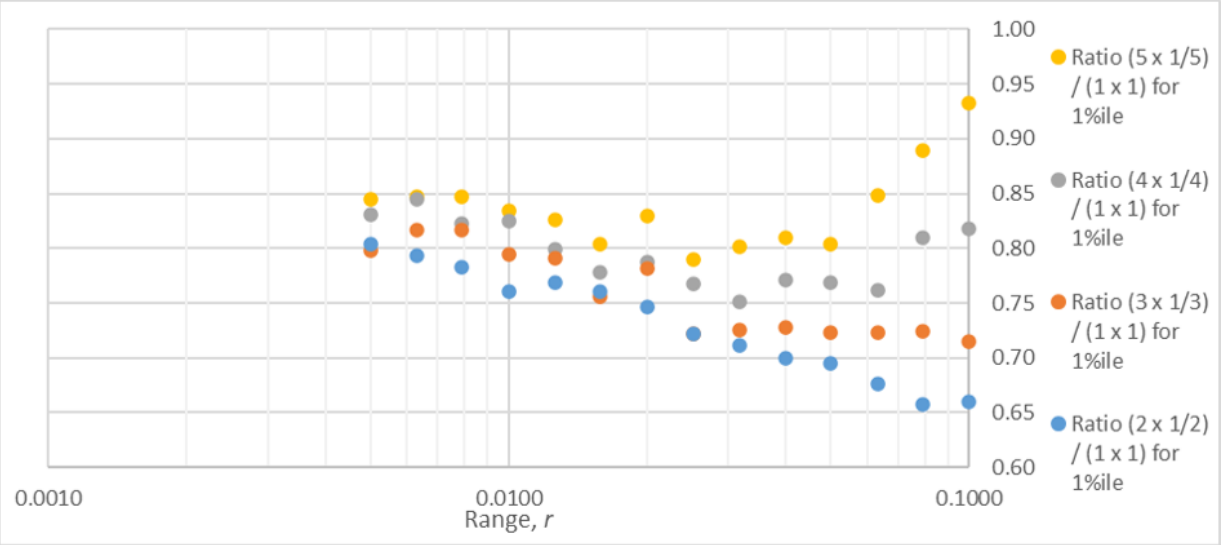

Figure 212: Ratios of 1%ile number of nodes for full connectivity of RGGs in flat unit rectangles of widths 2, 3, 4, and 5 with n-for-given-r growth to the number of nodes required for a flat unit square.

9 RGGs in Spaces with Obstructions and n-for-given-r Growth

Most prior research on RGGs has considered nodal spaces where the sole constraint on communication between nodes is that of distance. Only a small amount of the literature has investigated the properties of spaces that contain obstructions; for example, see [Aboue Nze (2012)] [Bollobás (2009)] [Czumaj (2007)] [Frieze (2007)] [Frieze (2009)].

Practical applications of the theory or empirical observations of RGG connectivity must take into account the fact that today's high data rate wireless communication systems have links that use RF or optical frequencies/wavelengths that are not only short range but are readily blocked by buildings, interior walls, or natural obstructions. Thus, such links are limited by both distance and line-of-sight.

This section reports the results of simulations involving configurations that include closed *obstructions* or obstacles that block line-of-sight between nodes in the nodal space. It is assumed that, no matter how close two nodes are, there is no link between them if line-of-sight is blocked.

Only n-for-given-r growth is addressed.

Two models of node placement are considered in this section: randomly placed nodes in a flat unit square with multiple obstructions and randomly placed nodes in flat and toroidal square lattices.

The first model is where nodes are randomly placed without constraint in a flat unit square and they are linked when they are within range r of each other. The only difference from the previous flat unit square model is that now it is assumed that links may be blocked by intervening obstructions.

The randomly placed nodes of an RGG are constrained to lie only in the area outside of the obstructions; in other words, nodes are never placed inside an obstruction.

For simplicity, the singular term "obstructed area" is used to refer to the *total* area that is obstructed and thus excluded from having nodes placed in it.

The second model is where nodes are placed at random at points of a *lattice*. Nodes in a lattice can communicate with other nodes that are at most ω positions away and are in the same row or column. This model is described in detail in Section 9.2.

Due to the infinite number of possibilities involving configurations that contain obstructions, only a few configurations have been modeled here to demonstrate the relationships between the range and the number of nodes required for connectivity.

9.1 Flat Unit Square with Obstructions

This section presents LRMs for RGGs in flat unit squares with n-for-given-r growth in two-dimensional spaces with obstructions. The statistical metrics are the 99%ile, mean, and 1%ile number of nodes to first achieve the conditions of full connectivity, 99% connectivity, and no isolated nodes. Both the mid-r and large-r sets described and listed in Section 4.1.2 were used for this work.

Only a few LRMs are provided to demonstrate that the statistical metrics appear to closely follow power laws.

The introductory material presented in Section 4 as well as Section 4.2 and its subsections regarding the simulations and their conditions applies here with the following considerations: the obstructions are square and all of the configurations modeled in this paper have 16 obstructions that, in total, occupy half the area of a flat unit square.

Two models of obstruction placement are considered. The first is where the 16 obstructions are *regularly placed* in a flat unit square. The second is where the 16 obstructions are *randomly placed* in a flat unit square.

The only constraint on random placement of the obstructions is that their borders do not overlap each other or the border of the square; however, the borders may be arbitrarily close. Because of this, a randomly placed collection of obstructions may effectively partition the RGG into mutually unreachable pockets of area that are cul-de-sacs of nodes that can, in practical terms, never[56] join the GC.

The statistical consequences of partitioning an RGG into mutually unreachable areas can be demonstrated simply by use of a series of random numbers. Suppose a randomly placed collection of obstructions in a unit square creates one or more areas totaling just 0.01 that are unreachable from the remaining 0.99 area of the square. For an RGG with n nodes,

[56] Because the borders can be arbitrarily close together but not actually touch, no portion of the space is truly unreachable from any other. Even so, the probability that there exists a connected chain of nodes between some areas may be vanishingly small resulting in an RGG having areas that are for all practical purposes unreachable from each other. This issue is addressed later in this section and in Section 9.1.3.

the probability that the RGG can be fully connected is the probability that all n nodes are placed only in the larger area; the event that all n nodes are in the smaller area can be disregarded. This can be modeled by finding the probability that n random numbers generated on the interval $[0, 1)$ are all less than or equal to 0.99.

A short C program was written that estimated this probability on the basis of 10^6 trials. First, the impact on full connectivity was estimated. For $n = 100$ nodes the probability that 100% of the nodes would be in 99% of the area was found to be 0.366. For $n = 1{,}000$ nodes the probability that 100% of the nodes would be in 99% of the area was found to be 3.4×10^{-5}. Next, the impact on partial connectivity was estimated. For $n = 100$ nodes the probability that 99% of the nodes would be in 99% of the area was found to be 0.736. For $n = 1{,}000$ nodes the probability that 99% of the nodes would be in 99% of the area was found to be 0.584.

For RGGs containing regularly placed obstructions, the condition of full connectivity was always achievable. For RGGs containing randomly placed obstructions, full connectivity was found to be unachievable for so many of the simulated geometries that, at best, only the LRMs for 99% connectivity could be obtained. This is the reason for including the LRMs for 99% connectivity of RGGs in nodal spaces with regularly placed obstructions.

In rare cases even 99% connectivity was not achievable with randomly placed obstructions. The samples of $n = 1{,}000$ simulated geometries for each value of r were partitioned into 20 sets of 50 geometries, each with its own random seed, yielding $14 \times 20 = 280$ sets and $14 \times 1{,}000 = 14{,}000$ unique geometries. A small number of the sets contained geometries that appeared not to reach 99% connectivity regardless of the number of nodes added to the RGGs, and so after a high threshold was reached those simulations were terminated and replacement sets were run with different random seeds. This happened once each for sets simulating ranges $r = 0.0079$, 0.0063, and 0.0050, so only 3/280 or 1.07% of the sets had this issue. Recognizing the fact there is a small probability that one or more additional geometries in each of those sets might also have resulted in failure to achieve 99% connectivity, an estimate of a close lower bound on the fraction of such geometries for the cases modeled here is 3/14,000 or 0.02%.

Much of the emphasis in this paper has been on determining the behavior of RGGs with large numbers of nodes because that is where most of the focus of prior work has been. For example, Penrose's Theorem was proved for "n large" and Gupta and Kumar's Theorem applies for "$n \to \infty$." However, by addressing the practical issue of obstructions, this section begins to move more toward realistic applications which would involve smaller numbers of nodes. The sections that follow give results based on both the mid-r and large-r sets.

9.1.1 Regularly Placed Obstructions: Full Connectivity and No Isolated Nodes

Figure 213 shows an instance of a flat unit square with 16 regularly placed obstructions that occupy half the area of the square. The range is $r = 0.0200$ and full connectivity was first achieved when 4,370 nodes were randomly placed in this space outside of the obstructed area.[57]

The mid-r set is used in this section. Figure 214 shows that the number of nodes required to first achieve full connectivity closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. The values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs demonstrate only that their behaviors in this case appear to obey power laws.

The LRMs for full connectivity corresponding to the equations for the trend lines shown in Figure 214 are as follows.

For the 99%ile number of nodes required to first achieve full connectivity in a flat unit square with n-for-given-r growth with 16 regularly placed obstructions occupying half the area, (21) becomes

$$n = 4.4689\, r^{-1.9500} \tag{188}$$

with an R^2 value of 0.9970. For the simulated values of r, MAPE is 8.29% and max error of the value of the predicted n is −18.72%.

[57] If one were to zoom in on the plot it would appear that some nodes overlap the borders of the flat unit square or the obstructions but this is an artifact of the limitations of the spreadsheet that generated the plot. None of the nodes are outside the square or inside the obstructions.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 2.7773\, r^{-1.9641} \tag{189}$$

with an R^2 value of 0.9962. For the simulated values of r, MAPE is 9.67% and max error of the value of the predicted n is -20.41%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 1.9052\, r^{-1.9835} \tag{190}$$

with an R^2 value of 0.9968. For the simulated values of r, MAPE is 9.10% and max error of the value of the predicted n is -17.37%.

Figure 215 shows that the number of nodes required to first achieve no isolated nodes closely follows a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. As noted above, the values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs indicate only that the general behavior may obey a power law.

The LRMs for no isolated nodes corresponding to the equations for the trend lines shown in Figure 215 are as follows.

For the 99%ile number of nodes required to first achieve no isolated nodes in a flat unit square with n-for-given-r growth with 16 regularly placed obstructions occupying half the area, (21) becomes

$$n = 2.9578\, r^{-2.0344} \tag{191}$$

with an R^2 value of 0.9982. For the simulated values of r, MAPE is 6.08% and max error of the value of the predicted n is -17.70%.

For the mean number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 1.5501\, r^{-2.0799} \tag{192}$$

with an R^2 value of 0.9982. For the simulated values of r, MAPE is 6.36% and max error of the value of the predicted n is -18.79%.

For the 1%ile number of nodes required to first achieve no isolated nodes, the LRM is

$$n = 0.7437\, r^{-2.1698} \tag{193}$$

with an R^2 value of 0.9989. For the simulated values of r, MAPE is 4.95% and max error of the value of the predicted n is -14.11%.

Figure 216 shows the fraction of simulations where adding a node that causes the last isolated node to become connected also causes the RGG to become fully connected with n-for-given-r growth. Interpreting this fraction as a probability, for RGGs in a flat unit square with 16 regularly placed obstructions occupying half the area of the square and n-for-given-r growth, the probability that an RGG becomes fully connected when the last isolated node is absorbed is about 0.107 for $r = 0.1000$ and increases to about 0.681 for $r = 0.0050$.

The ratios of the number of nodes required to first achieve full connectivity to the number of nodes required to first eliminate all isolated nodes are shown in Figure 217 are shown for the 99%ile and 1%ile values of those conditions. For the largest ranges in the set, considerably more nodes are required to first achieve full connectivity after the disappearance of isolated nodes for the 1%ile values but the ratio is smaller for the 99%ile values. The ratios for both of those goals may asymptotically approach 1 for the smallest ranges, which is consistent with Figure 216 that shows that the joint event of the disappearance of isolated nodes and achieving full connectivity becomes more frequent as the range decreases.

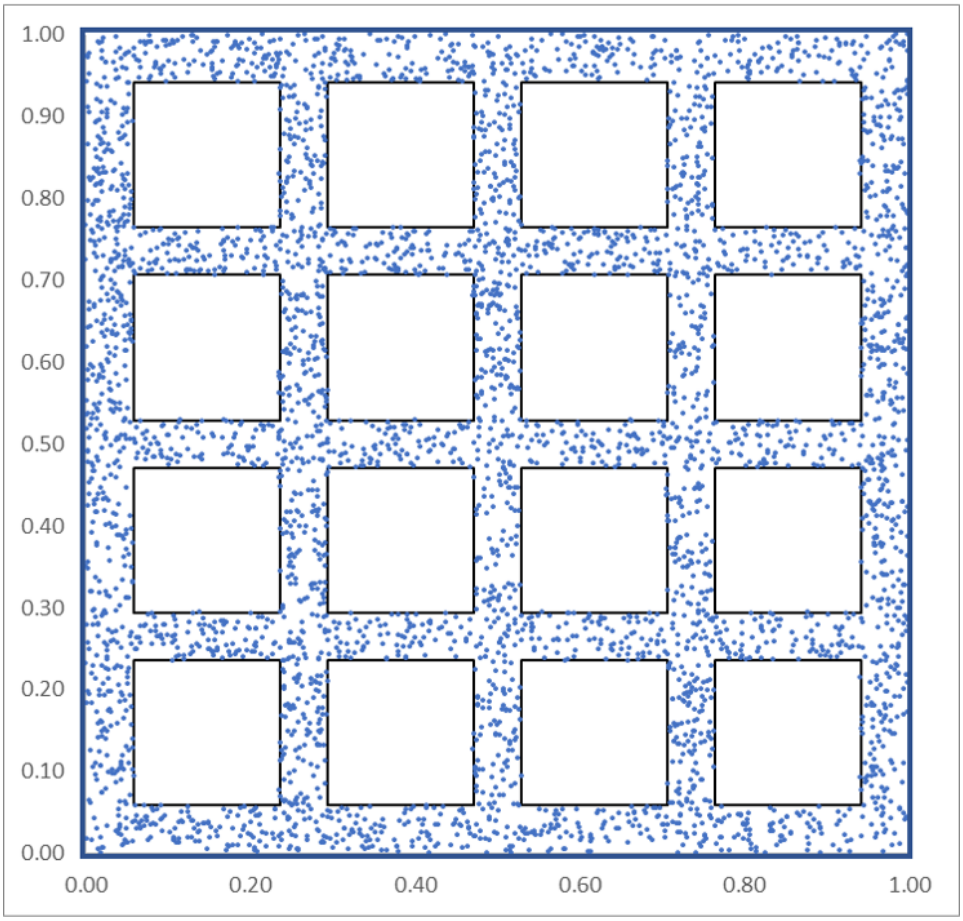

Figure 213: Geometry containing nodes with range 0.0200 scattered in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square. With n-for-given-r growth, the RGG consisting of 4,370 nodes randomly placed in the space outside of the obstructed area became fully connected with the addition of the last node.

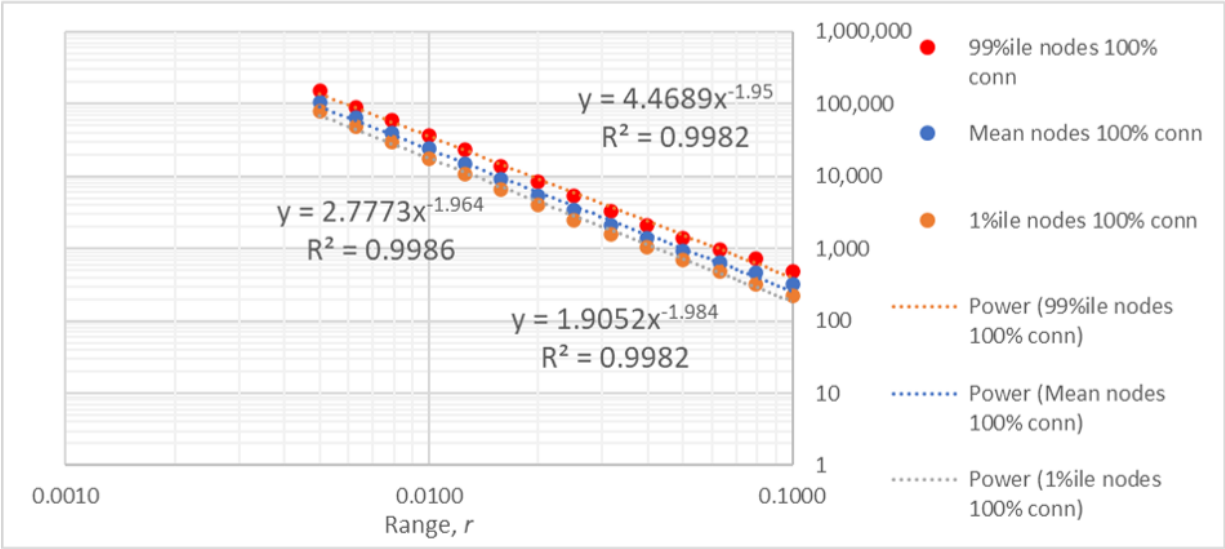

Figure 214: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth.

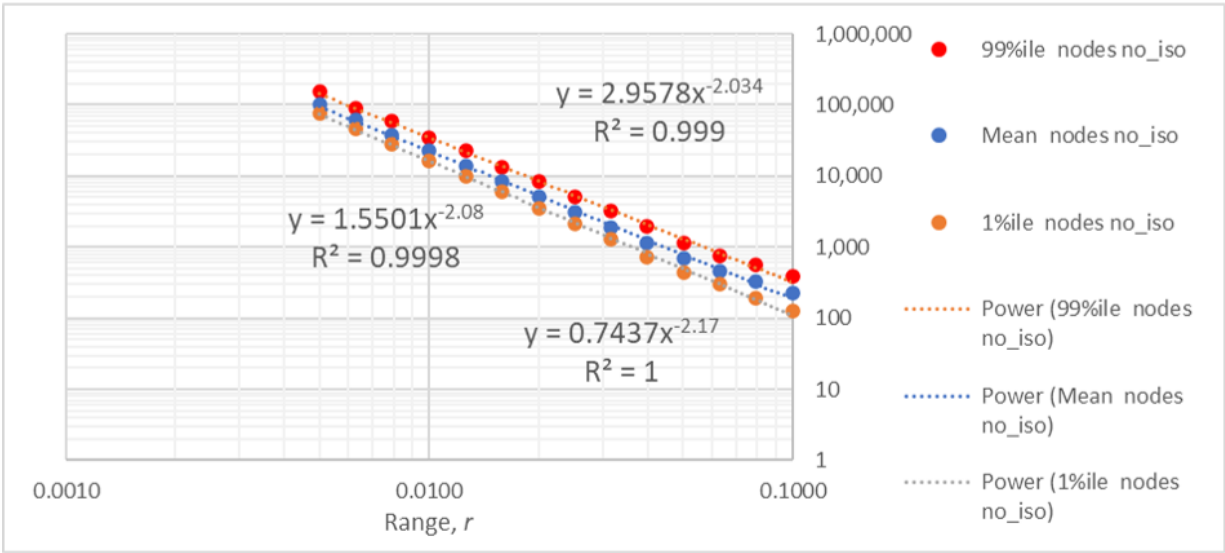

Figure 215: LRMs for 99%ile, mean, and 1%ile number of nodes for no isolated nodes in RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth.

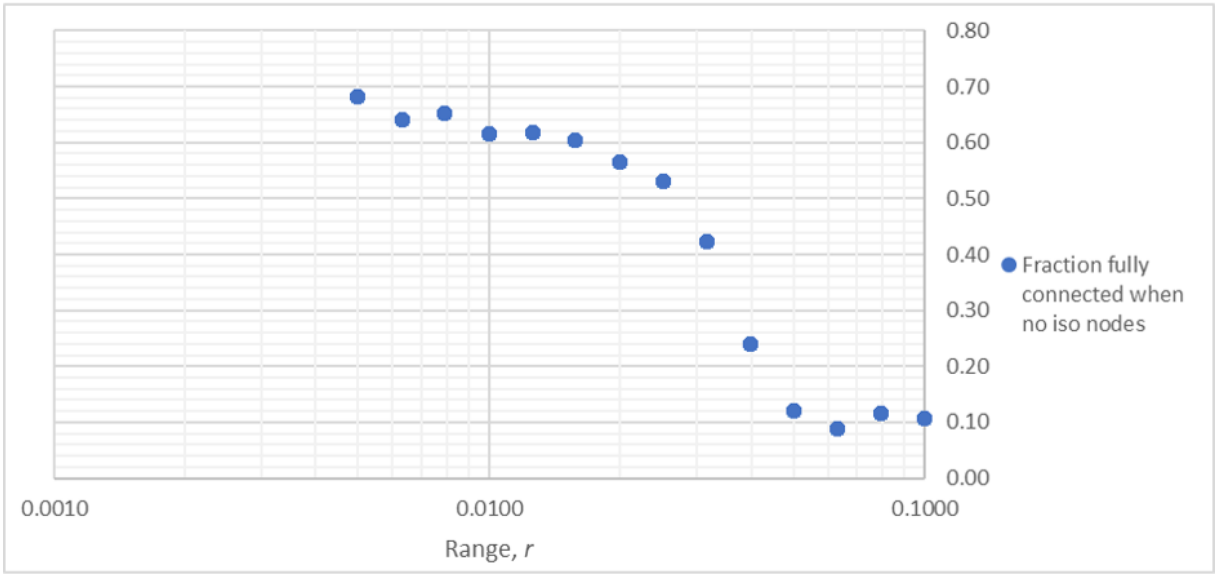

Figure 216: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square to become fully connected with n-for-given-r growth.

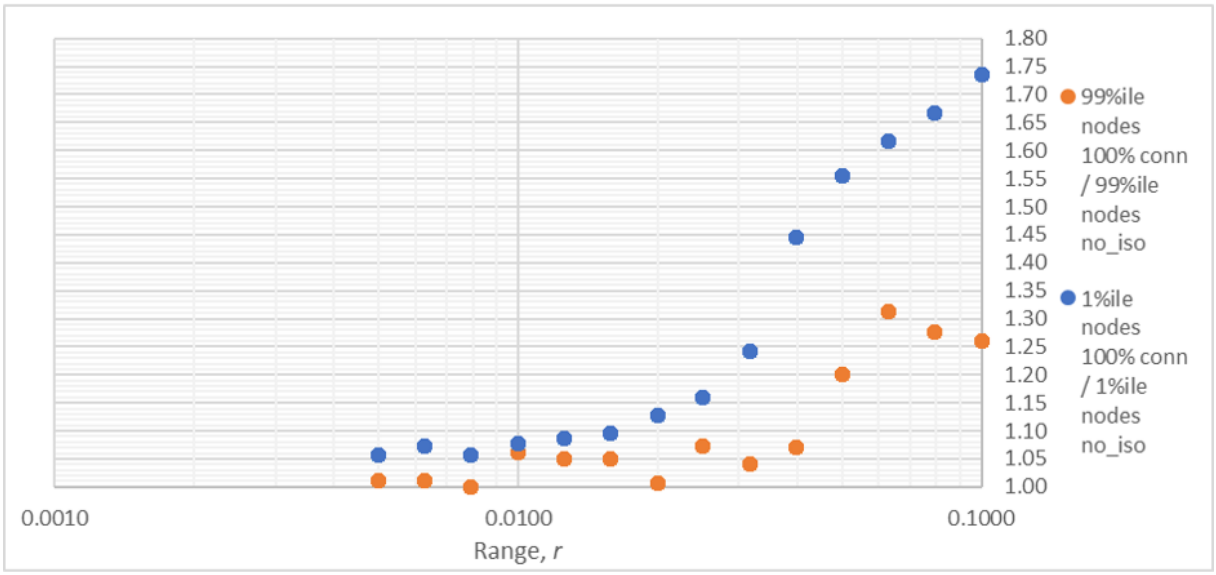

Figure 217: Ratios of 99%ile and 1%ile number of nodes for full connectivity to the number of nodes for no isolated nodes in RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth.

9.1.2 Regularly Placed Obstructions: Full Connectivity for Large Ranges

In this section configurations with much larger ranges than those used in Section 9.1.1 are modeled and tested in two-dimensional flat unit squares with regularly placed obstructions. The large-r set is used which consists of the ranges from 1.5849 down to 0.1000.

Figure 218 shows the 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve full connectivity of RGGs based on the sample of size $N = 1{,}000$ simulated geometries for each value in the large-r set. From $r = 0.1000$ up to 0.5012, the three statistical metrics appear to demonstrate power law behavior. For the largest ranges in the set, however, the three statistical metrics are essentially flat.

As noted in Section 4.1.3, fully connected RGGs with n-for-given-r growth are constrained to have a minimum size of 25 nodes which explains why the flat portion of the plot of the 1%ile statistical metric is at 25 nodes. It does not, however, explain why the flat portions of the plots occur where they do for the 99%ile and mean statistical metrics. When the minimum size of a fully connected RGG was reduced to five nodes the flat portion of the plot of the 1%ile statistical metric followed suit but the flat portions of the plots of the 99%ile and mean statistical metrics were unchanged.

Figure 219 shows the LRMs for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve full connectivity of RGGs for ranges $r = 0.1000$ up to 0.5012, for which values the statistical metrics appear to obey power laws. As noted previously, the values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs show only that the general behavior appears to obey a power law for the selected subset of ranges.

The LRMs for full connectivity corresponding to the equations for the trend lines shown in Figure 219 are as follows.

For the 99%ile number of nodes required to first achieve full connectivity in a flat unit square with n-for-given-r growth with 16 regularly placed obstructions occupying half the area, based on the simulated geometries using the subset of ranges from 0.1000 up to 0.5012, (21) becomes

$$n = 54.6503\, r^{-0.9353} \tag{194}$$

with an R^2 value of 0.9834. For the simulated values of r, MAPE is 5.10% and max error of the value of the predicted n is 11.67%.

For the mean number of nodes required to first achieve full connectivity, the LRM is

$$n = 26.8184\, r^{-1.0386} \tag{195}$$

with an R^2 value of 0.9874. For the simulated values of r, MAPE is 5.20% and max error of the value of the predicted n is −9.40%.

For the 1%ile number of nodes required to first achieve full connectivity, the LRM is

$$n = 10.9936\, r^{-1.2748} \tag{196}$$

with an R^2 value of 0.9938. For the simulated values of r, MAPE is 4.81% and max error of the value of the predicted n is −8.42%.

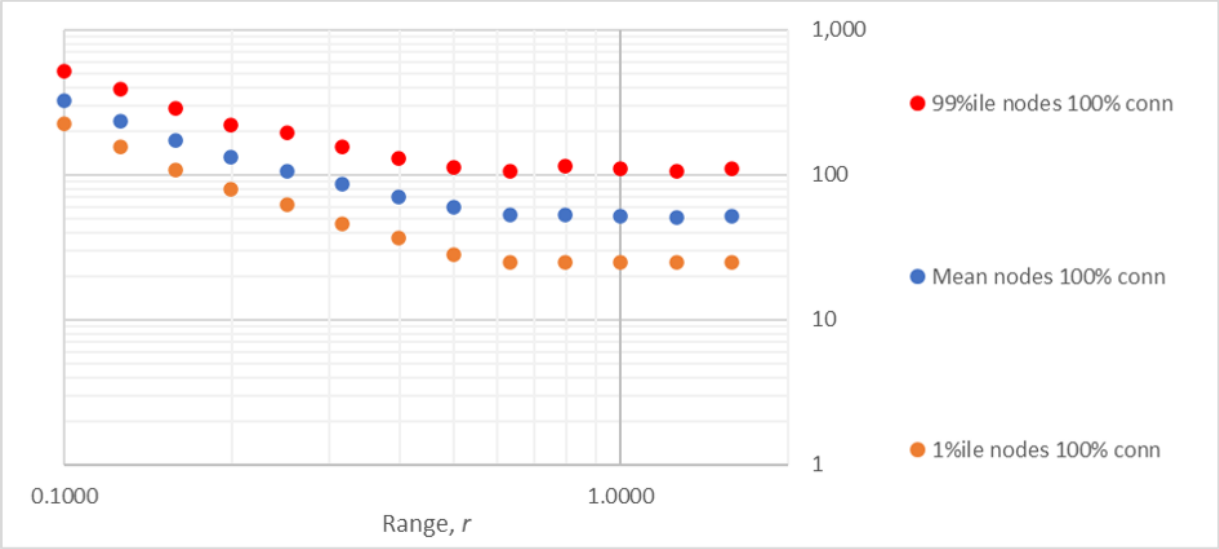

Figure 218: The 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth for the large-r set.

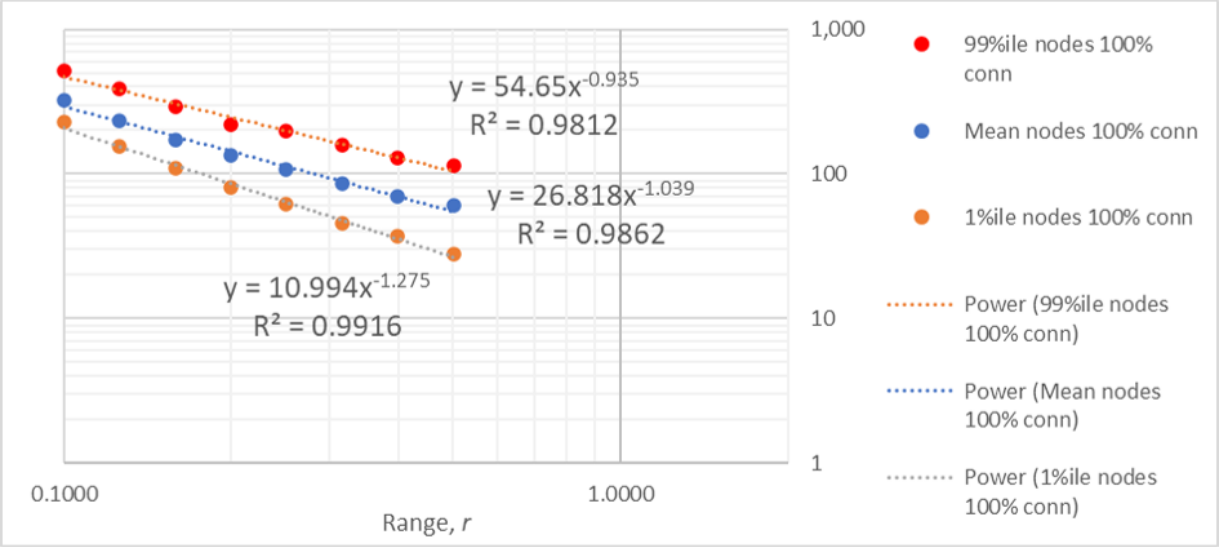

Figure 219: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth for the ranges in the large-r set where the statistical metrics appear to obey power laws.

9.1.3 Regularly Placed Obstructions: 99% Connectivity

This section presents LRMs for 99% connectivity for spaces with regularly placed obstructions. They are included for comparison to the results for spaces with randomly placed obstructions where 100% connectivity was often not achievable but 99% connectivity was achievable in all but a few cases.

Figure 220 shows an instance of a flat unit square with 16 regularly placed obstructions that occupy half the area of the square. The range is $r = 0.0200$ and 99% connectivity was first achieved when 3,700 nodes were randomly placed in this space outside of the obstructed area. There were 3,663 nodes in the GC (so $S = 0.99$), leaving 37 nodes that were isolated or part of smaller components. The red nodes are not part of the GC, and they are displayed with larger dot size to make them more visible.

Figure 221 shows that the number of nodes required to first achieve 99% connectivity appears to follow a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. As noted earlier, the values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs show only that the general behavior appears to obey a power law.

The LRMs for 99% connectivity corresponding to the equations for the trend lines shown in Figure 221 are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity in a flat unit square with n-for-given-r growth with 16 regularly placed obstructions occupying half the area, (21) becomes

$$n = 13.1123\, r^{-1.5171} \tag{197}$$

with an R^2 value of 0.9978. For the simulated values of r, MAPE is 5.46% and max error of the value of the predicted n is −12.46%.

For the mean number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 6.0477\, r^{-1.6561} \tag{198}$$

with an R^2 value of 0.9981. For the simulated values of r, MAPE is 5.77% and max error of the value of the predicted n is -12.00%.

For the 1%ile number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 3.3332\, r^{-1.7663} \tag{199}$$

with an R^2 value of 0.9989. For the simulated values of r, MAPE is 4.66% and max error of the value of the predicted n is -9.40%.

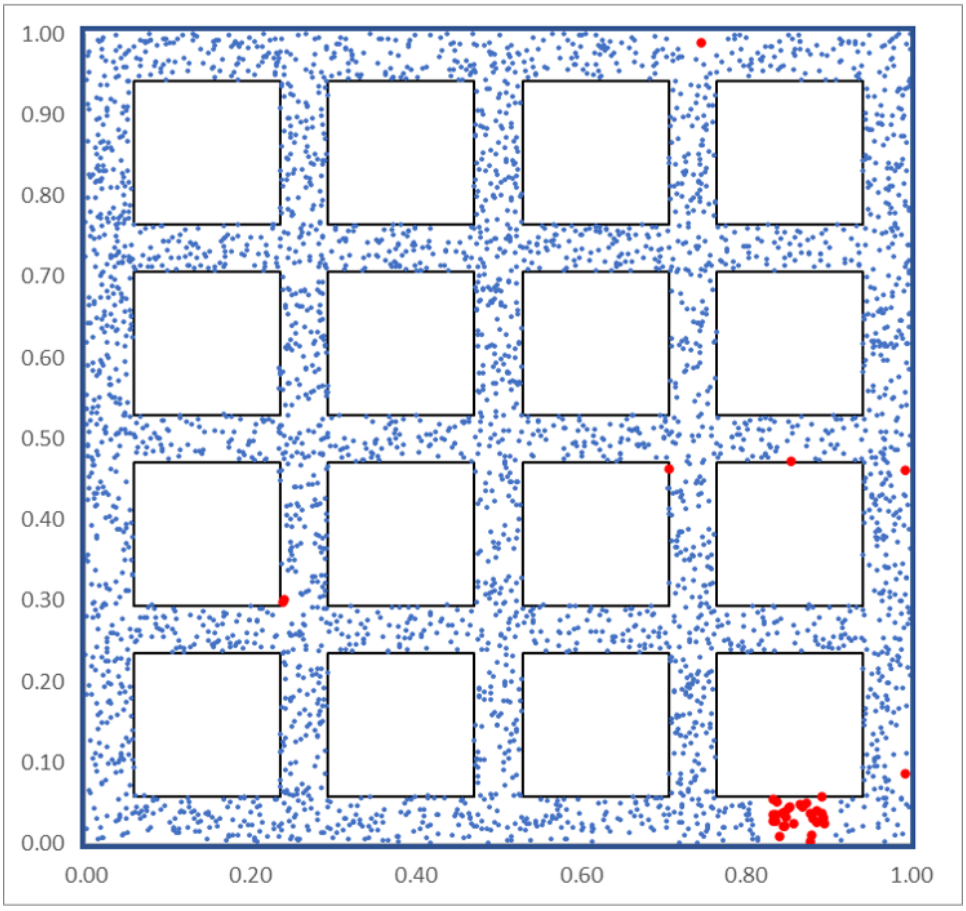

Figure 220: Geometry containing nodes with range 0.0200 scattered in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square. With 3,700 nodes randomly placed in the space outside of the obstructed area the RGG is 99% connected. The 37 nodes that are isolated or part of smaller components than the GC are represented as large red dots.

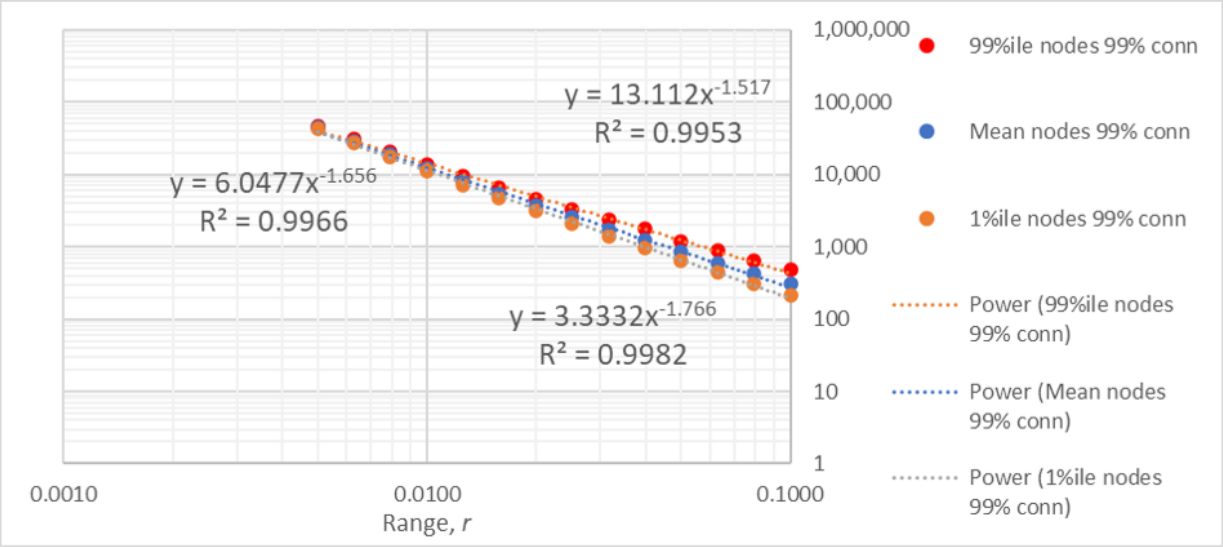

Figure 221: LRMs for 99%ile, mean, and 1%ile number of nodes for 99% connectivity of RGGs in a flat unit square with 16 regularly placed obstructions that occupy half the area of the square with n-for-given-r growth.

9.1.4 Randomly Placed Obstructions: 99% Connectivity

Previous sections considered configurations with where RGGs are in spaces with regularly placed obstructions, but this and following sections present LRMs for 99% connectivity for spaces with randomly placed obstructions. As noted in Section 9.1, full connectivity was often not achievable for these configurations but 99% connectivity was achievable in all but a few cases.

Figure 222 shows an instance of a flat unit square with 16 randomly placed obstructions that occupy half the area of the square. The range is $r = 0.0200$ and 7,826 nodes randomly placed in this space outside of the obstructed area achieved 99% connectivity. There were 7,759 nodes in the GC (so $S = 0.99$), leaving 67 nodes that were isolated or part of smaller components. The red nodes are not part of the GC, and they are displayed with larger dot size to make them more visible.

Figure 223 shows that the number of nodes required to first achieve 99% connectivity may follow a power law for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes based on the sample of size $N = 1{,}000$ simulated geometries per value of r. As noted earlier, the values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs show only that the general behavior appears to obey a power law.

The LRMs for 99% connectivity corresponding to the equations for the trend lines shown in Figure 223 are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity in a flat unit square with n-for-given-r growth with 16 randomly placed obstructions occupying half the area, (21) becomes

$$n = 103.6901\, r^{-1.3715} \tag{200}$$

with an R^2 value of 0.9832. For the simulated values of r, MAPE is 11.74% and max error of the value of the predicted n is −33.14%.

For the mean number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 22.3705\, r^{-1.4844} \tag{201}$$

with an R^2 value of 0.9994. For the simulated values of r, MAPE is 2.74% and max error of the value of the predicted n is -7.13%.

For the 1%ile number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 6.2442\, r^{-1.6734} \tag{202}$$

with an R^2 value of 0.9992. For the simulated values of r, MAPE is 3.58% and max error of the value of the predicted n is -7.71%.

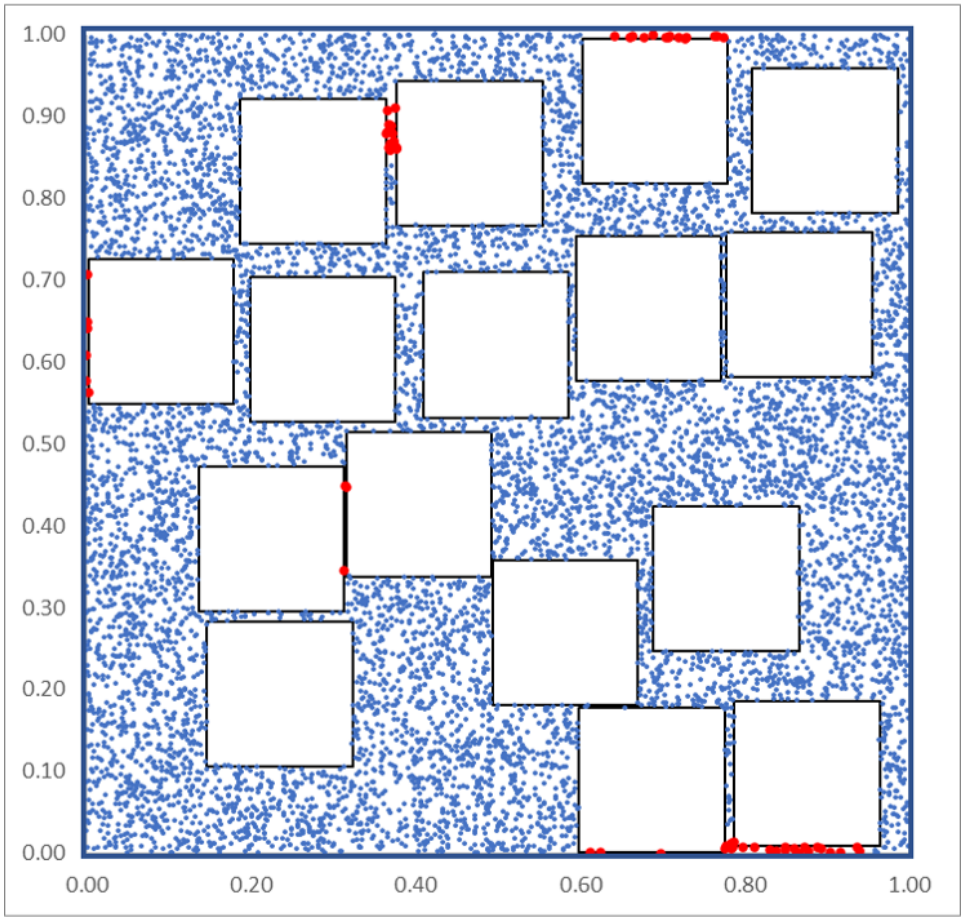

Figure 222: Geometry containing nodes with range 0.0200 scattered in a flat unit square with 16 randomly placed obstructions that occupy half the area of the square. With 7,826 nodes randomly placed in the space outside of the obstructed area the RGG is 99% connected. The 67 nodes that are isolated or part of components smaller than the GC are represented as large red dots.

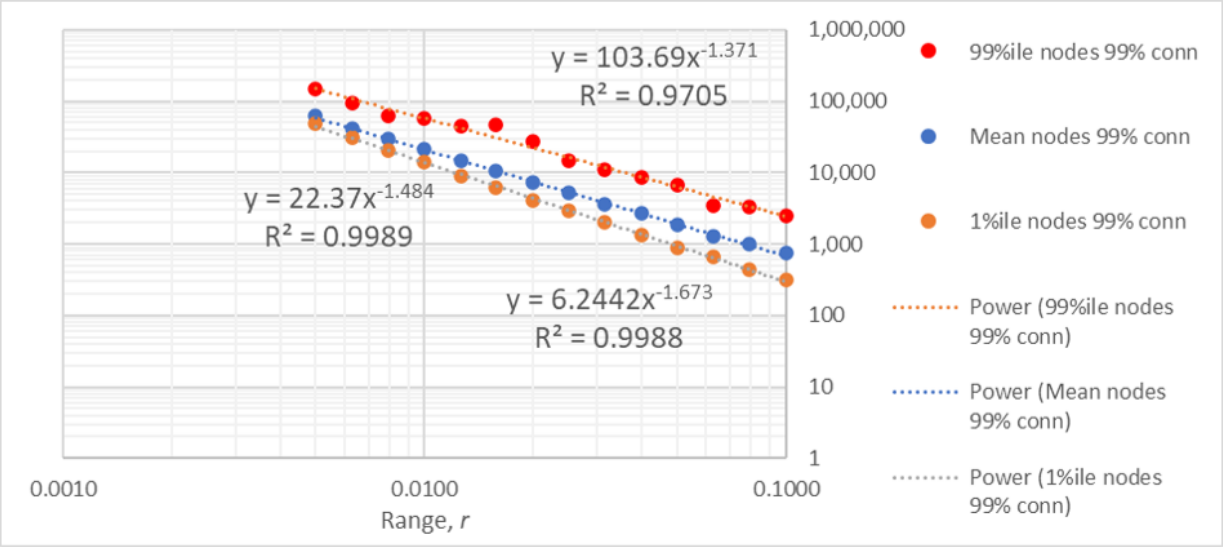

Figure 223: LRMs for 99%ile, mean, and 1%ile number of nodes for 99% connectivity of RGGs in a flat unit square with 16 randomly placed obstructions that occupy half the area of the square with n-for-given-r growth.

9.1.5 Randomly Placed Obstructions: 99% Connectivity for Large Ranges

In this section configurations with larger ranges than those used in Section 9.1.4 are considered. The large-r set is used which consists of the ranges from 1.5849 down to 0.1000.

Figure 224 shows the 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve 99% connectivity of RGGs based on the sample of size $N = 1{,}000$ simulated geometries for each value of r in the large-r set. From $r = 0.1000$ up to 0.3162 for the 99%ile and from $r = 0.1000$ up to 0.5012 for the mean and 1%ile, the three statistical metrics appear to demonstrate power law behavior. For the largest ranges in the set, however, the three statistical metrics are essentially flat.

Figure 225 shows the LRMs for the 99%ile, mean, and 1%ile statistical metrics for the number of nodes required to first achieve 99% connectivity of RGGs for the ranges where the statistical metrics appear to obey power laws. As noted previously, the values of m and b in (21) depend on the specific set of obstructions used in this case, and so these LRMs show only that the general behavior appears to obey a power law for the selected subset of ranges.

The LRMs for 99% connectivity corresponding to the equations for the trend lines shown in Figure 225 are as follows.

For the 99%ile number of nodes required to first achieve 99% connectivity in a flat unit square with n-for-given-r growth with 16 randomly placed obstructions occupying half the area, based on the simulated geometries using the sets of ranges listed above, (21) becomes

$$n = 274.4434\, r^{-1.0436} \tag{203}$$

with an R^2 value of 0.9827. For the simulated values of r, MAPE is 5.04% and max error of the value of the predicted n is −7.84%.

For the mean number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 101.4732\, r^{-0.8053} \tag{204}$$

with an R^2 value of 0.9564. For the simulated values of r, MAPE is 7.60% and max error of the value of the predicted n is 15.59%.

For the 1%ile number of nodes required to first achieve 99% connectivity, the LRM is

$$n = 17.2195\, r^{-1.1954} \tag{205}$$

with an R^2 value of 0.9899. For the simulated values of r, MAPE is 5.69% and max error of the value of the predicted n is -10.29%.

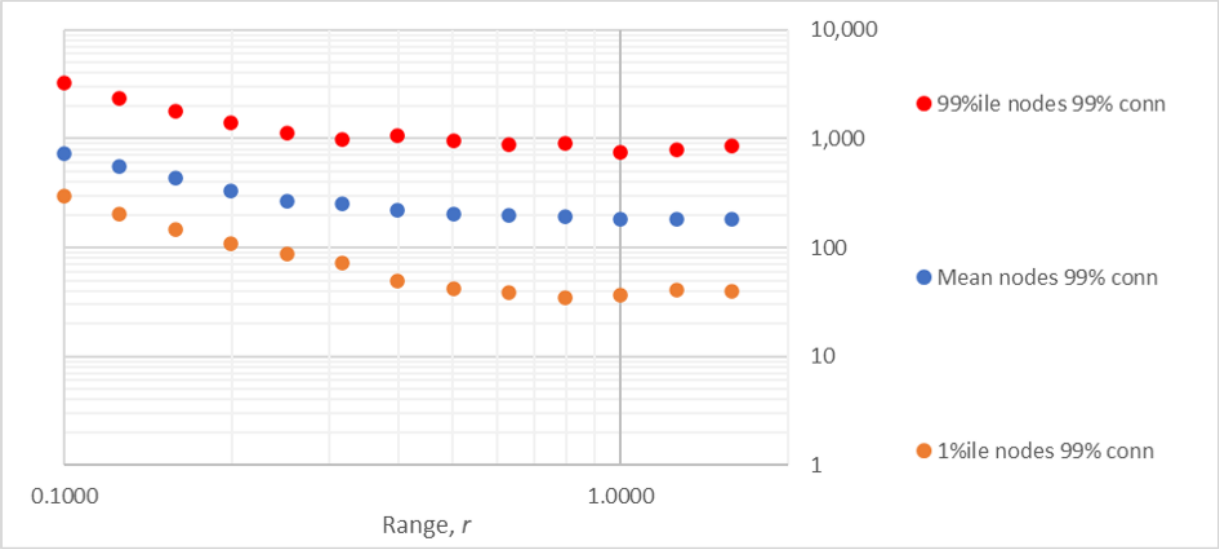

Figure 224: The 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with 16 randomly placed obstructions that occupy half the area of the square with n-for-given-r growth for the large-r set.

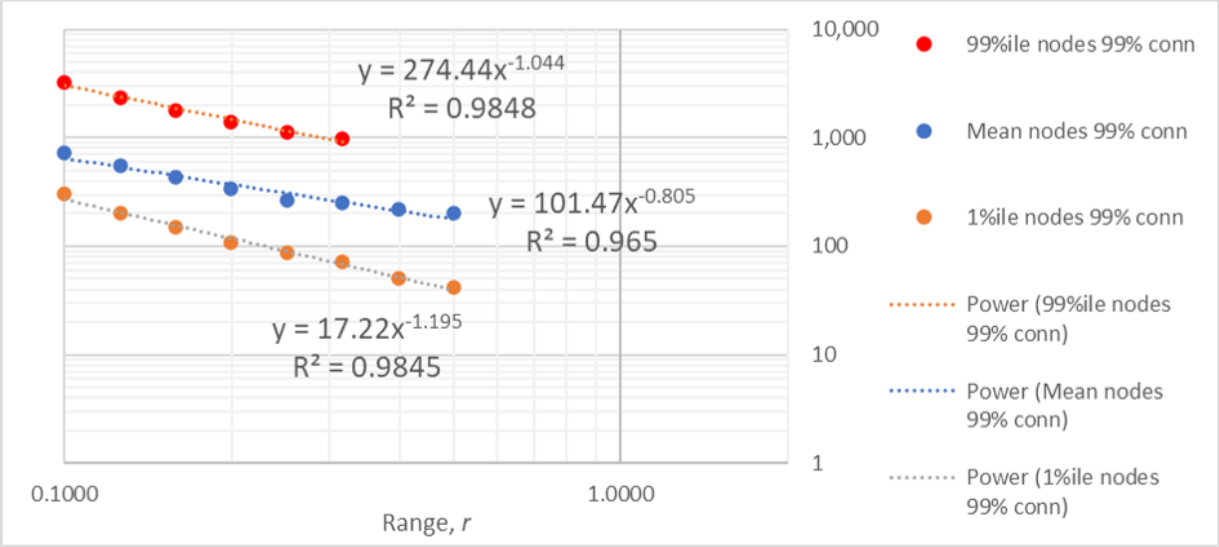

Figure 225: LRMs for 99%ile, mean, and 1%ile number of nodes for full connectivity of RGGs in a flat unit square with 16 randomly placed obstructions that occupy half the area of the square with n-for-given-r growth for the ranges in the large-r set where the statistical metrics appear to obey power laws.

9.1.6 Comparing Connectivity in Spaces with Regularly and Randomly Placed Obstructions

It has been demonstrated that, for the configurations tested in this section, the statistical metrics for the number of nodes required to first achieve connectivity of RGGs in spaces with obstructions appear to fit power law LRMs for all but the largest ranges tested.

As was observed for spaces without obstructions, the probability of an RGG in a flat space with n-for-given-r growth with regularly placed obstructions becoming fully connected when the last isolated node disappears appears to increase as the range decreases but that probability does not reach 1 even for the smallest range in the mid-r set. This determination could not be made for RGGs with randomly placed obstructions due to the fact that 100% connectivity could not be achieved in every case.

For a given range, RGGs in spaces with randomly placed obstructions require more nodes than do RGGs in spaces with the same number of regularly spaced obstructions. Figure 226 shows the ratios of the statistical metrics presented in Section 9.1.3 and Section 9.1.4. Figure 226 shows that, for the configurations considered, the ratio of the 99%ile number of nodes required to first achieve 99% connectivity for randomly and regularly placed obstructions is between 3 and 7.[58] The ratio is observed to be between 1 and 2.5 for the mean and 1%ile number of nodes and the ratios decrease as the range decreases.

[58] It is unclear as to why the graph for this ratio would appear to show a peak.

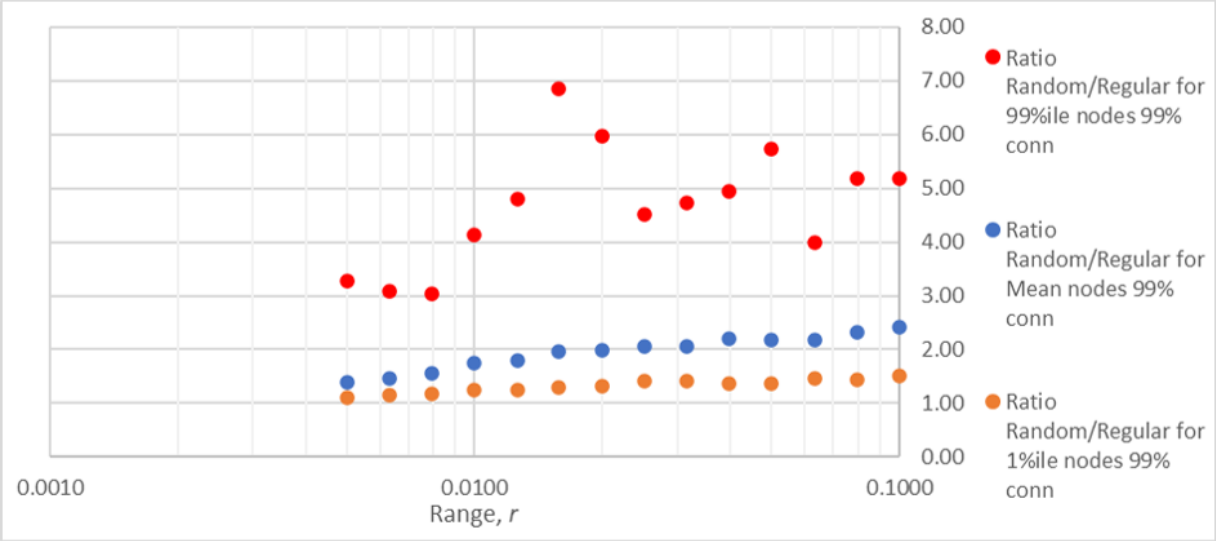

Figure 226: Ratios of 99%ile, mean, and 1%ile number of nodes for 99% connectivity of RGGs in a flat unit square with 16 regularly vs. randomly placed obstructions that occupy half the area of the square with n-for-given-r growth.

9.2 Flat and Toroidal Lattice

This section explores the conditions for connectivity of nodes randomly placed at the points of a lattice. The nodes in a lattice can communicate with other nodes if and only if they are in the same row or the same column and are within range ω of each other, where ω is an integer representing a number of intersections.

A lattice can be considered to be a special case of a space with obstructions, where the obstructions constrain the possible node positions to specific regularly spaced points and also constrain the communications paths between nodes to be only directly "up," "down," "left," or "right."

Range $\omega = 1$ means that nodes are linked if and only if they are at adjacent points of the lattice. Range $\omega = 4$ means nodes are linked if and only if they are at most four lattice points away from each other[59] in the same row or the same column.

As noted in Section 4.1.3, fully connected RGGs with n-for-given-r growth were constrained to have a minimum size of 25 nodes and the same approach was done for lattices with n-for-given-r growth. The effect of reducing that minimum size was also tested.

Only flat and toroidal square lattices with n-for-given-r growth are modeled.

Nodes are added to a lattice by selecting points randomly and "without replacement" so that a lattice point can contain at most one node.

Lattices are square and contain $M \times M$ points where nodes can be placed.

[59] "At most ω lattice points away from each other" means there are at most $\omega - 1$ lattice points between them.

RGGs in $M \times M$ lattices are simulated by using flat and toroidal unit squares containing an $(M-1) \times (M-1)$ array of regularly placed obstructions. The intersections of the rows and columns between the obstructions as well as the intersections between the obstructions and the borders constitute the M^2 lattice points where nodes can be placed.[60]

The nodal communication range ω is simulated for the lattice by calculating the equivalent value of range r for this node spacing in a unit square. For consistency, the term "n-for-given-r growth" is used even when the range is given as ω.

For a given value of M, the maximum range that a node would need to have a link with its most distant neighbor in a flat lattice is $\omega = M - 1$ and the maximum range that a node would need to have a link with its most distant neighbor in a toroidal lattice is $\omega = \lfloor M/2 \rfloor$.

The n-for-given-r growth of an RGG in a lattice is parameterized by the *placement probability* p, which is the fraction of points of the lattice that are occupied by a node. In these experiments, an RGG containing n nodes in a $M \times M$ lattice created by n-for-given-r growth in an $(M-1) \times (M-1)$ array of obstructions has a placement probability $p = n/M^2$.

[60] The total area of the obstructions is set to zero so that the nodes are correctly spaced for the flat and toroidal distance calculations; however, the nodes are still constrained to communicate only within their own rows and columns.

9.2.1 Flat Lattice: Full Connectivity

For $M = 17$, Figure 227 shows an RGG in a 17×17 flat lattice created by placing a 16×16 array of obstructions in a flat unit square.[61] For lattice range $\omega = 2$, the equivalent flat unit square range is $r = 0.1178$. Nodes were added at the points of the flat lattice until the RGG was fully connected with that value of ω. The green dots represent the 175 nodes (for a placement probability of $p = 175/17^2 = 0.606$) in the GC. The links between connected nodes are shown as well.

It is well known and was also demonstrated in Section 3.2 and Section 3.3 that the GC in a flat or toroidal unit square has a sharp phase transition for both n-for-given-r and r-for-given-n growth. The same apparently does not hold for an RGG in a flat lattice with n-for-given-r growth. Figure 228 shows a scatter plot of the fraction of nodes S in the GC for 100 simulated geometries for a configuration representing RGGs in a 17×17 flat lattice with $\omega = 5$. The plot shows S as a function of the placement probability p. In contrast to the abrupt phase transition of the GC shown in Figure 6 for an RGG in a flat unit square with n-for-given-r growth, there is no clear phase transition in the case of an RGG in a flat lattice with n-for-given-r growth for the configuration tested. The best that one could say based on examination of Figure 228 is that, for this case, the phase transition to a single large GC takes place somewhere between $p = 0.05$ and 0.15.

Figure 229 shows the fraction of simulated geometries for configurations with an RGG in a 17×17 flat lattice that became fully connected at the same time the last isolated node disappeared. Interpreting this fraction as a probability, it can be seen that for this configuration this probability falls from about 70% for the smallest range $\omega = 1$ to about 50% for the largest range $\omega = 16$. The same general trend, where this probability increases with decreasing range (and thus increasing the required number of nodes), was observed for RGGs with n-for-given-r growth in all of the spaces modeled except for the one-dimensional flat and toroidal unit squares. Corresponding with this trend, Figure 230 shows that the mean value of the ratio (for each simulated geometry) of the number of nodes required to first achieve full connectivity to the number of nodes required to first

[61] The obstructions have zero area but are plotted notionally with non-zero area to show the lattice points more clearly.

achieve no isolated nodes increases with the range ω: the additional number of nodes required to first achieve full connectivity rises from about 3% for $\omega = 1$ to about 14% for $\omega = 16$.

The 99%ile number of nodes to first achieve the conditions of full connectivity and no isolated nodes for this configuration are nearly identical, however. Both sets of values are plotted in Figure 231 but they are so close that they appear to be a single set of values.

Note that Figure 231 is a linear-log plot. The best linearity for the number of nodes plotted for both full connectivity and no isolated nodes as a function of ω was observed on a linear-log plot for $\omega = 1$ to 10, with an inflection point beginning at $\omega = 11$ after which the plots of the number of nodes flattened. Thus, the best-fitting LRMs for the number of nodes to first achieve full connectivity or no isolated nodes for RGGs in these lattices are log law functions of the range ω rather than power laws.

The LRMs for the 99%ile number of nodes to first achieve full connectivity and no isolated nodes differ so slightly that, for clarity, the trend line and equation only for full connectivity are shown in Figure 232.

For the 99%ile number of nodes required to first achieve full connectivity in an RGG in a 17×17 flat lattice with n-for-given-r growth for the values of $\omega = 1$ to 10, the LRM is

$$n = -90.1593 \ln \omega + 280.1924 \tag{206}$$

with an R^2 value of 0.9955. For the simulated values of ω, MAPE is 3.01% and max error of the value of the predicted n is -7.80%.

As noted above, the log law had the best fit for the observed data. The log law had the highest R^2 and lowest MAPE and max error of the LRMs that were considered, which was consistent with the visual appearance of the log-log (power law) and log-linear (exponential) plots. The power law and exponential LRMs were also good fits with only slightly smaller R^2 but their MAPE and max error percentages were roughly double those of the log law LRM.

For comparison, the power law LRM for an RGG in this lattice was

$$n = 316.0418\,\omega^{-0.5777} \tag{207}$$

with an R^2 value of 0.9695. For the simulated values of ω, MAPE is 5.95% and max error of the value of the predicted n is 13.27%.

When the minimum size of a fully connected RGG in a lattice was reduced to only two nodes the LRM obtained was

$$n = -90.3102\,\ln\omega + 280.0232 \tag{208}$$

which is almost indistinguishable from (206).

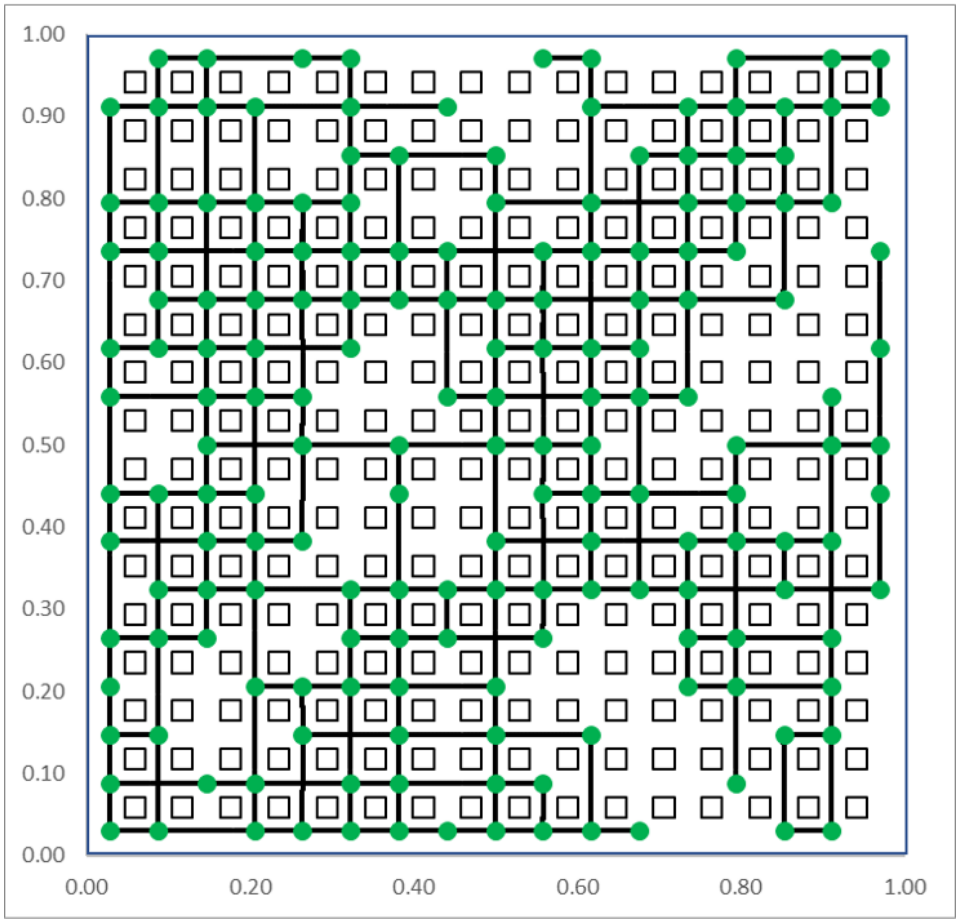

Figure 227: Geometry containing nodes with range $\omega = 2$ scattered in a 17×17 flat lattice. With n-for-given-r growth, the 175 nodes randomly placed at the points of the lattice became fully connected with the addition of the last node.

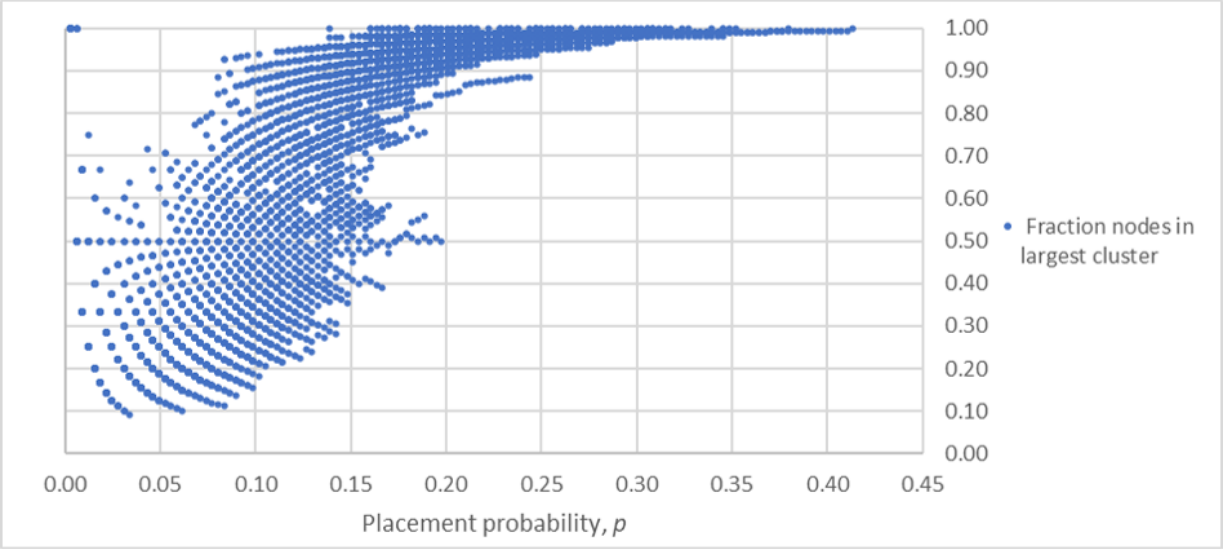

Figure 228: Fraction of nodes S in the GC for 100 simulated geometries in a 17×17 flat lattice with $\omega = 5$. S is shown as a function of the placement probability p. A sharp phase transition is not visible.

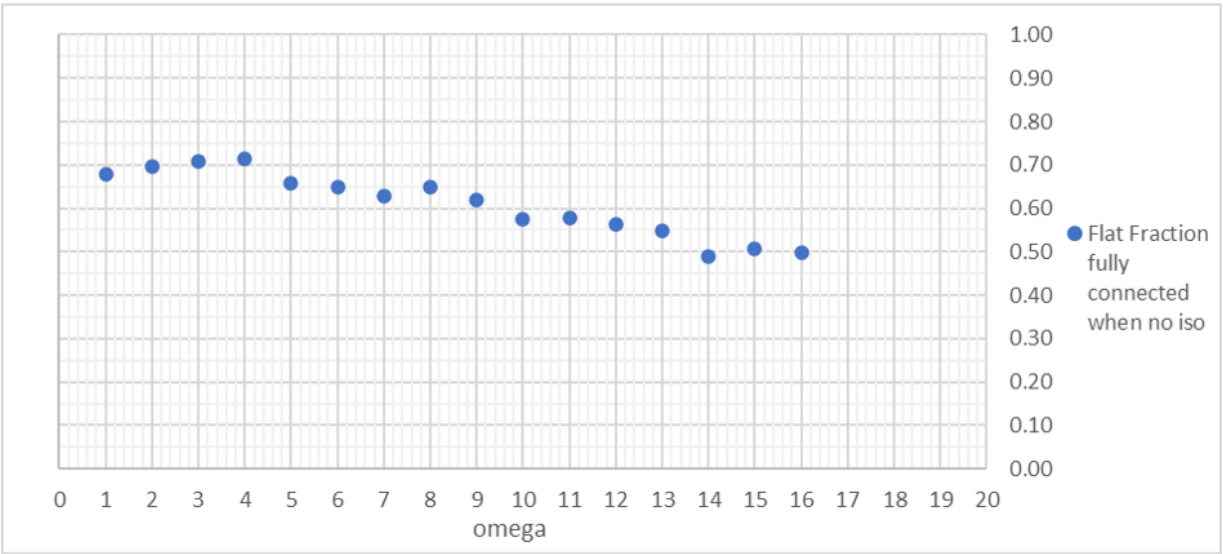

Figure 229: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes an RGG in a 17×17 flat lattice to become fully connected with n-for-given-r growth.

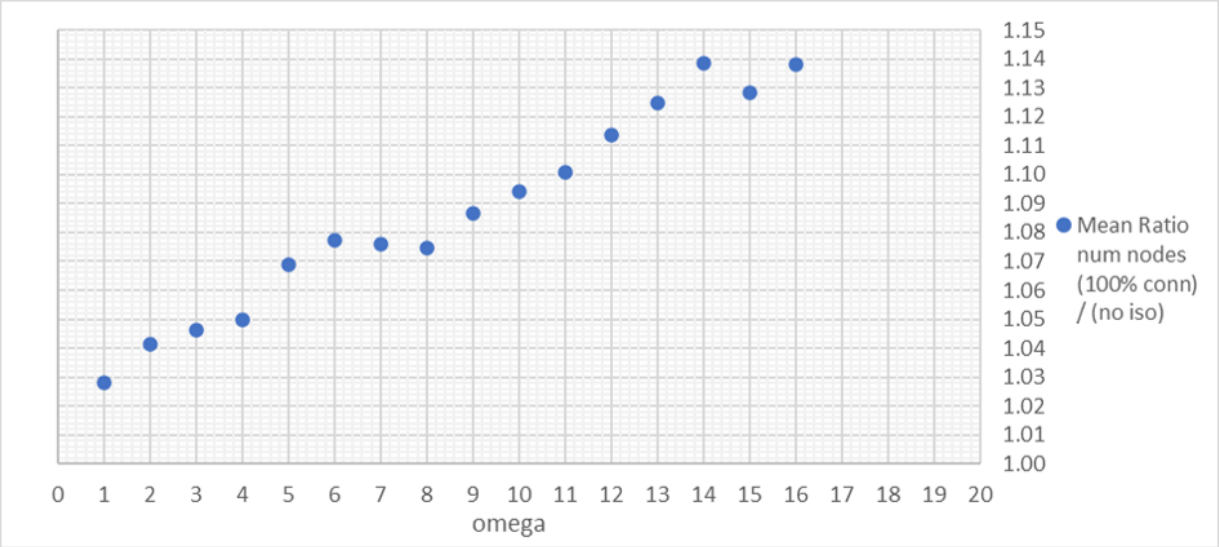

Figure 230: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in an RGG in a 17×17 flat lattice with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range ω for each simulated geometry.

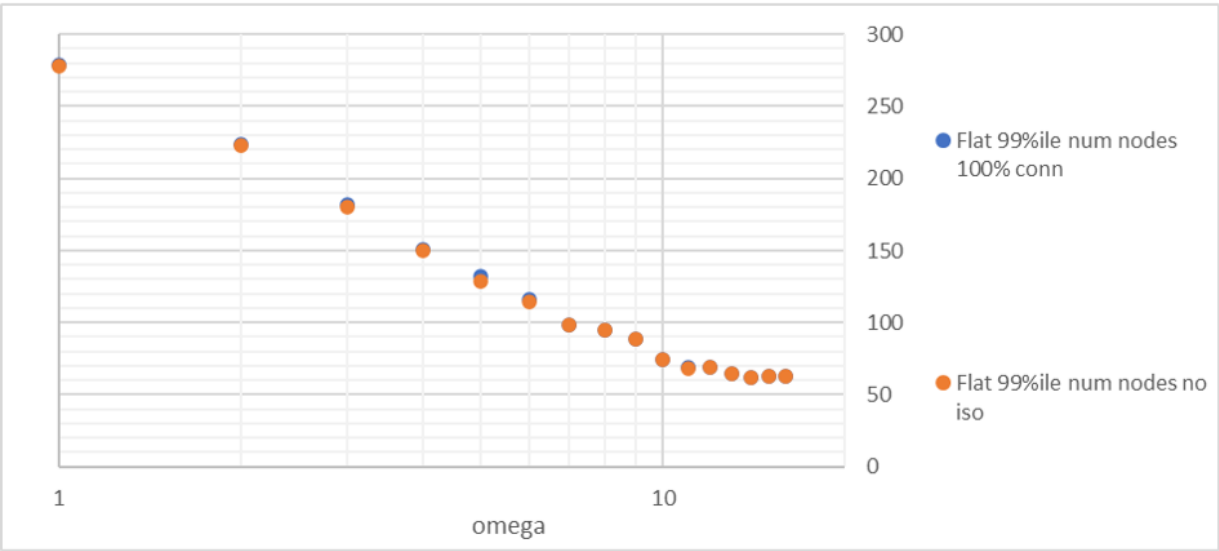

Figure 231: The 99%ile number of nodes for full connectivity and no isolated nodes in an RGG in a 17×17 flat lattice with n-for-given-r growth. The two sets of data are so close that they appear to be a single set of values.

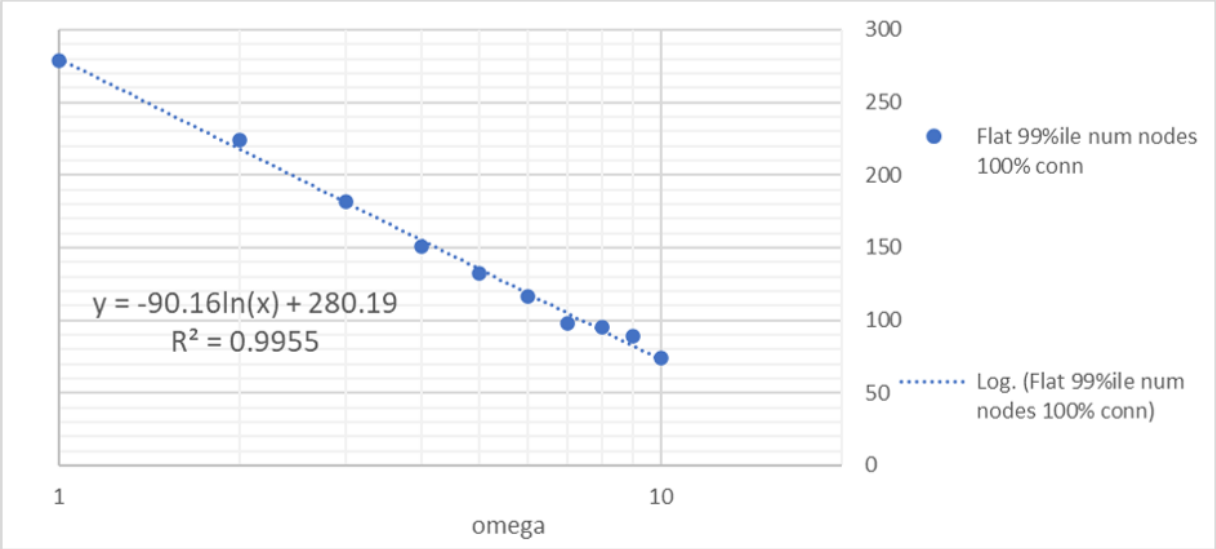

Figure 232: LRM for 99%ile number of nodes for full connectivity in an RGG in a 17×17 flat lattice with n-for-given-r growth for the ranges where the statistical metric appears to obey a power law.

9.2.2 Toroidal Lattice: Full Connectivity and No Isolated Nodes

For $M = 17$, Figure 233 shows an RGG in a 17×17 toroidal lattice created by placing a 16×16 array of obstructions in a toroidal unit square.[62] For lattice range $\omega = 2$, the equivalent toroidal unit square range is $r = 0.1178$. Nodes were added at the rows of the toroidal lattice until the RGG was fully connected with that value of ω. The green dots represent the 131 nodes (for a placement probability of $p = 131/17^2 = 0.453$) in the GC. The links between connected nodes are shown as well, including the links that cross the "border" of the lattice.

As was shown for a flat lattice, an RGG in a toroidal lattice with n-for-given-r growth does not appear to have a sharp phase transition to a single large GC. Figure 234 shows a scatter plot of the fraction of nodes S in the GC for 100 simulated geometries for a configuration representing RGGs in a 17×17 toroidal lattice with $\omega = 5$. The plot shows S as a function of the placement probability p. In contrast to the abrupt phase transition of the GC shown in Figure 10 for a toroidal unit square with n-for-given-r growth, there is no clear phase transition in the case of an RGG in a toroidal lattice with n-for-given-r growth for the configuration tested. The best that one could say based on examination of Figure 234 is that, for this case, the phase transition to a single large GC takes place somewhere between $p = 0.05$ and 0.11, which is slightly sharper than for RGGs in flat lattices.

Figure 235 shows the fraction of simulated geometries for configurations with RGGs in a 17×17 toroidal lattice that became fully connected at the same time the last isolated node disappeared. Interpreting this fraction as a probability, it can be seen that for this configuration this probability falls from about 80% for the smallest range $\omega = 1$ to about 50% for the largest range $\omega = 16$. The same general trend, where this probability increases with decreasing range (and thus increasing the required number of nodes), was observed for RGGs with n-for-given-r growth in all of the spaces modeled except for the one-dimensional flat and toroidal unit squares. Corresponding with this trend, Figure 236 shows that the mean value of the ratio (for each simulated geometry) of the number of nodes required to first achieve full connectivity to the number of nodes required to first

[62] The obstructions have zero area but are plotted notionally with non-zero area to show the lattice points more clearly.

achieve no isolated nodes increases with the range ω: the additional number of nodes required to first achieve full connectivity rises from about 2% for $\omega = 1$ to about 14% for $\omega = 16$.

The flattening of the plots for values of ω larger than 8 seen in Figure 235 is due to the fact that in a 17×17 toroidal lattice a node with any range $\omega \geq \lfloor 17/2 \rfloor = 8$ can reach any other node in its row or column.

As is the case for RGGs in flat lattices, the 99%ile number of nodes to first achieve the conditions of full connectivity and no isolated nodes for this configuration are nearly identical. Both sets of values are plotted in Figure 237 but again they are so close that they appear to be a single set of values.

Figure 237 is a linear-log plot because log law was the best fit, as it was for RGGs in a flat lattice. The best linearity for the number of nodes plotted for both full connectivity and no isolated nodes as a function of ω was observed on a linear-log plot for $\omega = 1$ to 8, with an inflection point beginning at $\omega = 9$ after which the plots of the number of nodes flattened.

For clarity, the trend line and equation only for full connectivity are given in Figure 238.

For the 99%ile number of nodes required to first achieve full connectivity in an RGG in a 17×17 toroidal lattice with n-for-given-r growth for the values of $\omega = 1$ to 8, the LRM is

$$n = -98.8628 \ln \omega + 261.8126 \tag{209}$$

with an R^2 value of 0.9943. For the simulated values of ω, MAPE is 4.38% and max error of the value of the predicted n is -12.15%.

As was the case for RGGs in a flat lattice, the log law had the highest R^2 and lowest MAPE and max error of the LRMs that were considered.

For comparison, the power law LRM for an RGG in this lattice was

$$n = 296.5598\,\omega^{-0.7038} \tag{210}$$

with an R^2 value of 0.9755. For the simulated values of ω, MAPE is 6.35% and max error of the value of the predicted n is 12.76%.

For the 99%ile number of nodes required to first achieve no isolated nodes in an RGG in a 17×17 toroidal lattice with n-for-given-r growth for the values of $\omega = 1$ to 8 the LRM (only slightly different from (209) for full connectivity) is

$$n = -99.0813\,\ln\omega + 261.9760 \tag{211}$$

with an R^2 value of 0.9947. For the simulated values of ω, MAPE is 4.22% and max error of the value of the predicted n is -11.22%.

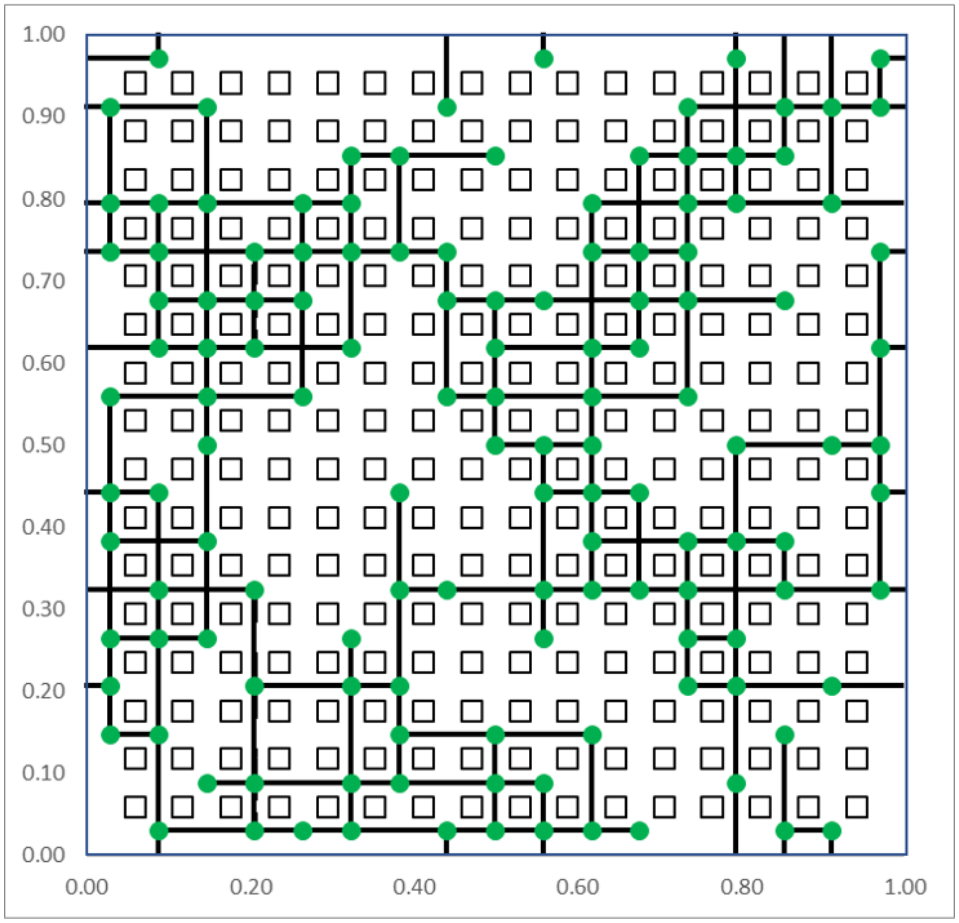

Figure 233: Geometry containing nodes with range $\omega = 2$ scattered in a 17×17 toroidal lattice. With n-for-given-r growth, the 131 nodes randomly placed at the points of the lattice became fully connected with the addition of the last node.

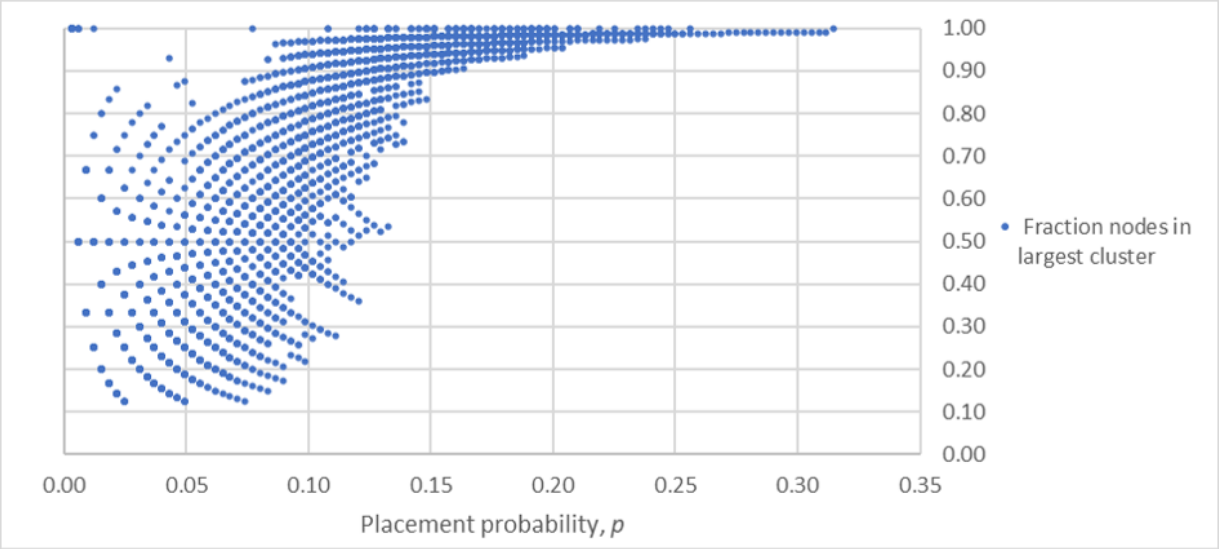

Figure 234: Fraction of nodes S in the GC for 100 simulated geometries in an RGG in a 17×17 toroidal lattice with $\omega = 5$. S is shown as a function of the placement probability p. A sharp phase transition is not visible.

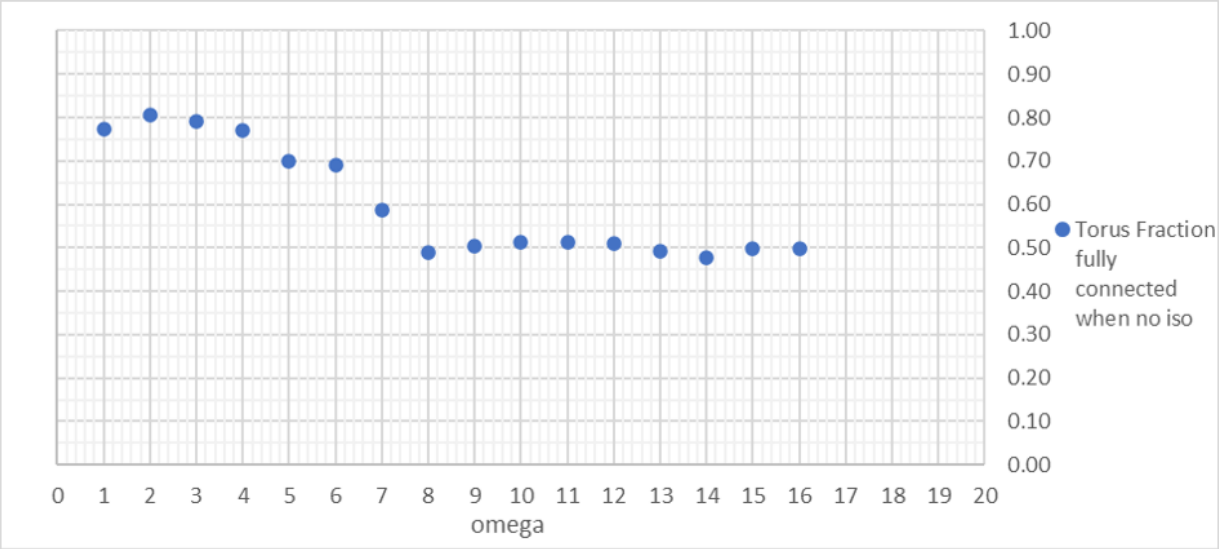

Figure 235: Fraction of simulations where adding a node that causes the last isolated node to become connected also causes an RGG in a 17×17 toroidal lattice to become fully connected with n-for-given-r growth.

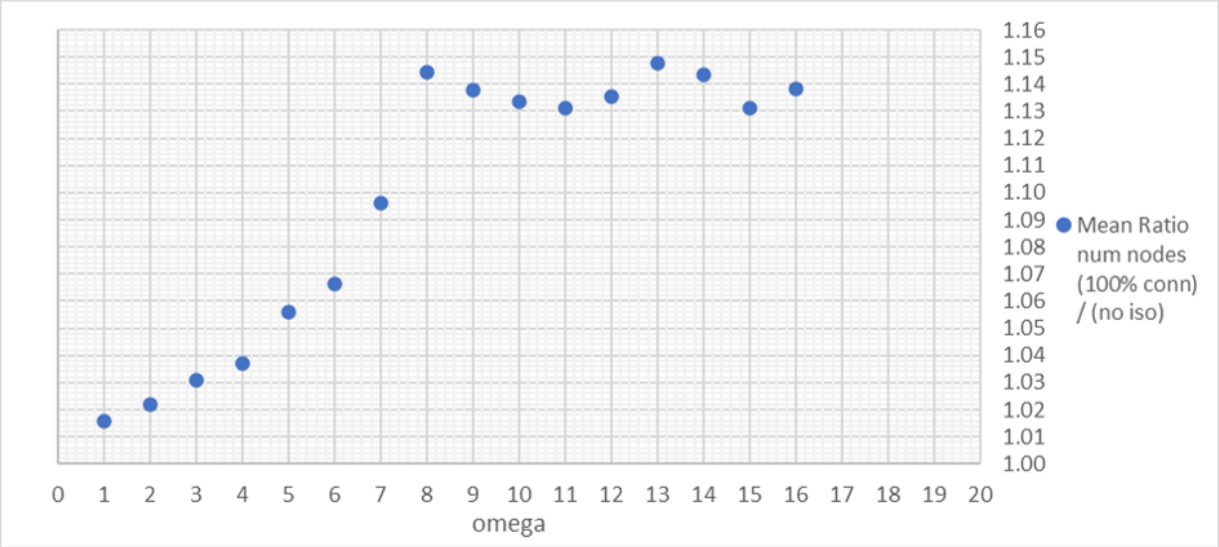

Figure 236: Mean of the ratios of range for full connectivity to the range where isolated nodes disappear in an RGG in a 17×17 toroidal lattice with n-for-given-r growth. Mean is taken over all of the ratios for a given value of the range ω for each simulated geometry.

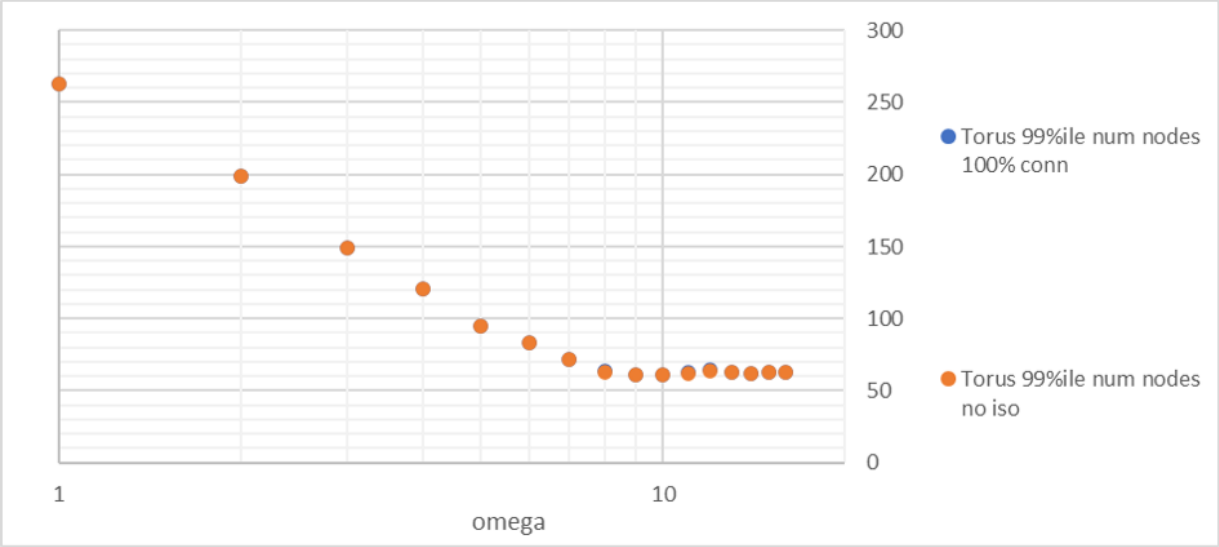

Figure 237: The 99%ile number of nodes for full connectivity and no isolated nodes in an RGG in a 17×17 toroidal lattice with n-for-given-r growth. The two sets of data are so close that they appear to be a single set of values.

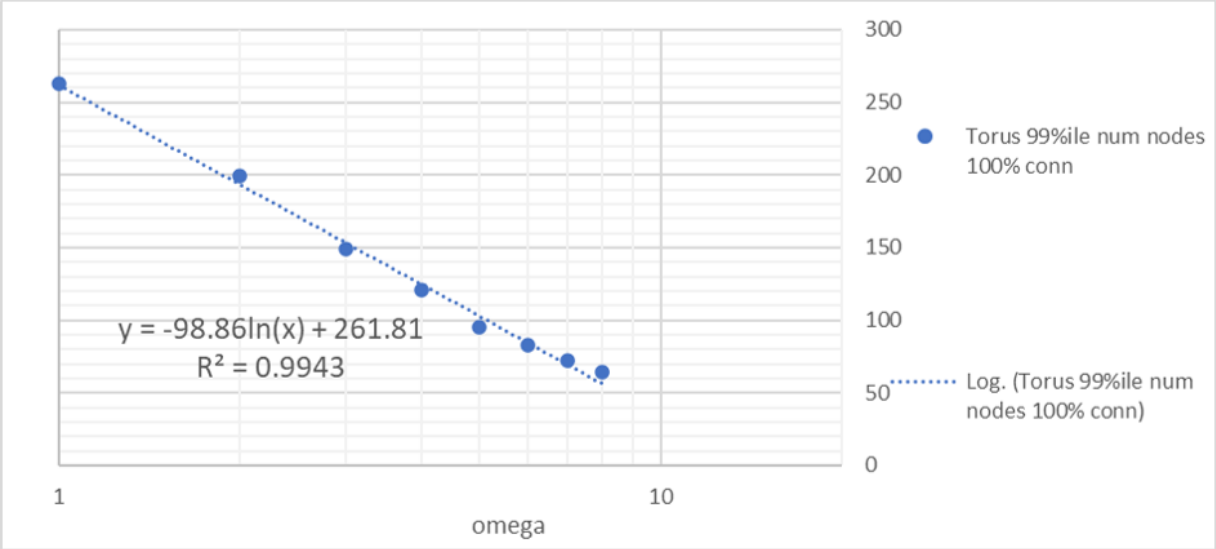

Figure 238: LRM for 99%ile number of nodes for full connectivity of an RGG in a 17×17 toroidal lattice with n-for-given-r growth for the ranges where the statistical metric appears to obey a power law.

9.2.3 Toroidal Lattice: Estimating the Probability of No Isolated Nodes

Section 3.4 obtained expressions for the probability that an RGG in a two-dimensional toroidal space has no isolated nodes and Section 7.2.1 did the same for RGGs in d-dimensional spaces. This section develops similar expressions for the case of an RGG in a toroidal lattice in two-dimensional space.

Consider an RGG in a toroidal lattice with $M \times M$ points that contains n nodes with range ω. The nodes are assumed to be randomly scattered "without replacement" on the lattice points, meaning a lattice point can contain at most one node. Any given node in the lattice can reach $4\,\omega$ lattice points; let those lattice points be identified as LP_1, LP_2, ... $LP_{4\,\omega}$.

The first step is to determine the probability that any given node has no neighbor in any of the $4\,\omega$ lattice points it can reach. Since that given node at the center of the $4\,\omega$ lattice points, there are $n-1$ points remaining in the lattice that are distributed on M^2-1 lattice points, so the probability that LP_1 is *not* occupied by a node is $1-\frac{n-1}{M^2-1}$. That leaves $n-2$ nodes distributed on M^2-2 lattice points, so the probability that LP_2 is not occupied by a node is $1-\frac{n-2}{M^2-2}$. Continuing this argument for the rest of the lattice points LP_3, LP_4, ... $LP_{4\,\omega}$ that are within range of the given node,

$$\begin{aligned}\Pr\{\text{any one node is isolated}\} &= \left(1-\frac{n-1}{M^2-1}\right)\left(1-\frac{n-2}{M^2-2}\right) \\ &\quad \cdots\left(1-\frac{n-(4\,\omega)}{M^2-(4\omega)}\right) \\ &= \prod_{i=1}^{4\,\omega}\left(1-\frac{n-i}{M^2-i}\right) \qquad (212)\end{aligned}$$

which is equivalent to (11) for RGGs in a toroidal unit square.

Assuming statistical independence, the complements of the probabilities of individual node isolation given by (212) yield the probability or confidence C of the joint event (without border effects) that none of the n nodes in an RGG in a lattice are isolated:

$$\begin{aligned} C &= \Pr\{\text{no isolated nodes}\} \\ &= \left[1 - \prod_{i=1}^{4\,\omega} \left(1 - \frac{n-i}{M^2 - i}\right)\right]^n \end{aligned} \tag{213}$$

which is equivalent to (12) for RGGs in a toroidal unit square.

There is no closed-form solution for (213) that yields n as a function of C but the smallest value of n that achieves a given value of C is easily obtained by numerical means.

For the toroidal lattice configuration with $M = 17$ used in Section 9.2.2, (213) was solved for the smallest values of n that reached $C = 0.99$. These values of n are the predicted values of the 99%ile values of n that achieve no isolated nodes for that configuration.

For $\omega = 1$ to 8, these predicted values of n have MAPE 5.38% with respect to the results of the simulations, which improved on the regression MAPE 6.44% based on (211).

The derivation of (213) for RGGs in a toroidal lattice was straightforward because the probability of a given node having no neighbors is the same regardless of a node's position in the lattice. Deriving an expression for the probability of no isolated nodes in an RGG in a flat lattice would not be as easy because the probability of a given node having no neighbors is different for lattice points that are less than distance ω from the border.

10 Conclusions

The study of the connectivity properties of Random Geometric Graphs (RGGs) is fundamental to the understanding of ad hoc networks. This paper addressed two growth models for RGGs: n-for-given-r where RGGs grow by adding nodes with a given range and r-for-given-n where the range of a fixed number of nodes increases. In both models, nodes are scattered uniformly at random positions according to a homogeneous Poisson point process in a space.

The goal for n-for-given-r growth is to find the first number of nodes n that achieve a desired connectivity condition, and the goal for r-for-given-n growth is to find the minimum range r that achieves a desired connectivity condition. The desired connectivity condition may be full connectivity, 99% or 50% connectivity, or no isolated nodes. The primary statistical metrics obtained in this paper were the 99%ile, mean, and 1%ile of the first number of nodes or smallest range that achieve those conditions.

The approach used in this paper was to simulate the growth of RGGs to obtain statistical metrics for the various connectivity conditions, and then apply linear regression to transformed variables to obtain models for those connectivity goals.

RGGs were tested in two-dimensional spaces that were square, rectangular, and circular.

They were simulated in spaces of dimension one through six.

RGGs were tested in spaces without obstructions as well as with obstructions, and where nodes were located only at the points of a lattice.

Spaces modeled were toroidal, where the edges of the space wrap around, as well as flat, where there was no wraparound.

For RGGs other than those in lattices, the primary finding of this work was that each of the statistical metrics for the number of nodes or the range required to achieve any of the connectivity conditions as a function of the range or the number of nodes, respectively, was closely modeled by a linear function of a log-log transformation, resulting in a power law Linear Regression Model (LRM).

For example, for RGGs in a flat unit square with n-for-given-r growth, LRM (22) for the 99%ile number of nodes required to first achieve full connectivity as a function of range based on the mid-r set was $n = 4.6571\,r^{-2.0678}$ which had a mean absolute percentage relative error (MAPE) of only 1.85% and maximum relative error of −5.19% over the set of ranges simulated. It was noted that an RGG in a flat unit square achieves full connectivity for about 55% to 60% of the time with n equal to the mean number of nodes required to first achieve full connectivity.

For RGGs in a toroidal unit square with r-for-given-n growth, LRM (85) for the mean range required for full connectivity as a function of the number of nodes based on the small-n set was $r = 1.3312\,n^{-0.4584}$ which had MAPE of 1.29% and max error of 7.85% over the numbers of nodes simulated.

The power law model for the mean number of nodes for 50% connectivity was shown to closely match the threshold for the giant component's phase transition from being very small to encompassing most of the nodes in an RGG with n-for-given-r growth.

The connectivity properties of RGGs in flat spaces are different from those for RGGs in toroidal spaces due to border effects. For RGGs with n-for-given-r growth, considerably more nodes are required to achieve a given connectivity condition for the same range in flat spaces than in toroidal spaces, although the ratio appears to fall toward 1 as the range decreases.

Observing that the n-for-given-r and r-for-given-n growth models are conceptually inverses of each other, it was reasonable to investigate whether an n-for-given-r linear regression model could be inverted to yield a good r-for-given-n model. While the method of inverse regression is not necessarily usable as a general rule, this approach appears to work well for this problem domain as the r-for-given-n models obtained by inverse regression varies only slightly from and had only slightly greater relative errors than the r-for-given-n models obtained by linear regression directly from the simulation data.

The related method of reverse regression was also applied to the n-for-given-r simulation data (where the range was treated as the dependent variable) and it yielded r-for-given-n models that were nearly identical to the r-for-given-n models obtained by inverse regression.

Power law models performed well for RGGs in rectangular spaces. RGGs in flat unit rectangular spaces ranging from $2 \times 1/2$ to $5 \times 1/5$ were tested and the estimates given by power law models were found to have small relative error.

RGG growth for a few configurations of spaces with regularly and randomly placed obstructions was tested, and again power law LRMs performed well.

RGGs consisting of nodes placed at the points of flat and toroidal lattices with n-for-given-r growth were tested. Although good power law models could be obtained, it was found that linear regression models for the number of nodes required to first achieve full connectivity based on linear-log transformations gave estimates that had slightly smaller relative errors than those based on log-log transformations.

It was observed that the emergence of the giant component in an RGG in a lattice with n-for-given-r growth did not have a sharp phase transition as they did for the other spaces modeled.

Much of the testing of prior researchers' work in this paper focused on that of M.D. Penrose and H. Koskinen.

Penrose developed a general theory of RGG connectivity for r-for-given-n growth in d-dimensional toroidal spaces.

An aspect of Penrose's theory referred to in this paper as "the First Claim of Penrose's Theorem" was the hypothesis for r-for-given-n growth that, as the range increases, RGGs in toroidal space with large numbers of nodes ("n large") become fully connected with high probability "at the same instant" the last isolated node becomes connected.

Another aspect of Penrose's theory referred to in this paper as his "Second Claim" was to determine the conditions under which all isolated nodes in a toroidal RGG disappear with probability C. It was found in this paper that Penrose's expressions yield virtually the same numerical results as those obtained based only on the minimal assumptions of Poisson distribution of nodes and statistical independence. Furthermore, substitution of an approximation into the expression based on these minimal assumptions gave Penrose's expression.

What is referred to in this paper as "the Third Claim of Penrose's Theorem" combined the first two claims to say that the expressions for the conditions for no isolated nodes with probability C can be taken to be those for full connectivity as well for large RGGs in toroidal space.

Penrose's Theorem has been central to the field of RGG connectivity. His results for the special case of 1-connectivity or "simple connectivity" were tested against models that matched his assumptions as well as many "off-label" cases.

Most of the work described in this paper dealt with RGGs in two-dimensional spaces but models were also developed for RGGs with n-for-given-r growth in d-dimensional spaces. It was found that more nodes were required for both full connectivity and no isolated nodes as the dimensionality of the space increased. Power law models provided good estimates for spaces in all of the dimensions tested.

Simulation results verified that the fraction of RGGs that become fully connected when the last isolated nodes disappears increases as the range decreases for n-for-given-r growth or the number of nodes increases for r-for-given-n growth for spaces of dimension two and greater. For RGGs in one-dimensional spaces, however, the probability decreased as the range decreased for n-for-given-r growth or the number of nodes increased for r-for-given-n growth.

For RGGs in two-dimensional spaces, the probability of RGGs becoming fully connected when the last isolated nodes disappeared was not observed to reach 1 within the limits of these simulations. This probability reached 0.804 for flat RGGs and 0.832 for toroidal RGGs with range 0.000251, which was the smallest range simulated for n-for-given-r growth. This probability reached 0.789 for flat RGGs and 0.821 for toroidal RGGs with 10,000,000 nodes, which was the largest number of nodes simulated for r-for-given-n growth. By any reasonable definition, 10,000,000 nodes would be considered "n large," yet the probability that an RGG with r-for-given-n growth is fully connected when the last isolated nodes disappears hovers around 80%, which effectively disproves Penrose's First Claim.

In spite of this evidence against Penrose's First Claim, Penrose's Third Claim was indeed shown to hold for the 99%ile number of nodes or range required for full connectivity for RGGs in toroidal spaces. Furthermore, not only was it found that his theorem had even smaller relative error than the power law LRMs in many cases for both n-for-given-r and r-for-given-n growth but it did so even for numbers of nodes as small as 20. Thus, "n large" was found not to be required for Penrose's predicted values of r or n to be accurate.

Although Penrose's Theorem addresses connectivity "with high probability" (defined in this paper to be with confidence $C = 0.99$) it was found that in many cases his theorem gave accurate predictions even for $C = 0.01$.

For RGGs in spaces of dimension two and greater, the fraction of RGGs that become fully connected when the last isolated nodes disappears increases with the dimensionality d for a given range for n-for-given-r growth or a given number of nodes for r-for-given-n growth.

Koskinen used data obtained from simulations of RGGs in a flat unit square with r-for-given-n growth and obtained LRMs for several statistical metrics for the inverse square of the range required for full connectivity given a number of nodes. In spite of the small numbers of nodes simulated, Koskinen's models were also extremely accurate even though the power law models proposed in this paper gave better results for larger numbers of nodes.

The empirical data generated by the simulation approach used for this paper provided "ground truth" to test theories and results that have appeared in prior RGG literature. The models presented in this paper can provide the basis for refinements of theories of the formation of the complicated structure of RGGs.

Acknowledgment

Dr. E. Paul Ratazzi provided invaluable technical advice as well as making available much of the computational resources used to perform this work.

References

[Aboue Nze (2012)] G-C. Aboue Nze and F. Guinand, "Mean Degree of Ad Hoc Networks in Environments with Obstacles," *Proceedings, 8^{th} IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2012.

[Albrecht (2021)] M.R. Albrecht, J. Blasco, R.B. Jensen, and L. Mareková, "Mesh Messaging in Large-Scale Protests: Breaking Bridgefy," *Topics in Cryptology – CT-RSA 2021* in *Lecture Notes in Computer Science*, Vol 12704, Springer, 2021.

[Altman (2015)] D.G. Altman, "Standard Deviations and Standard Errors," *BMJ*, https://doi.org/10.1136/bmj.331.7521.903, Oct 2005.

[Asadi (2014)] A. Asadi, Q. Wang, and V. Mancuso, "A Survey on Device-to-Device Communication in Cellular Networks," *IEEE Communications Surveys & Tutorials*, Vol 16, No 4, 2014.

[Aven (2020)] T. Aven, *Exploring Reservoir Computing with Spatial Constraints*, Master's Thesis, Norwegian University of Science and Technology (NTNU), 2020.

[Avrachenkov (2020)] K.E. Avrachenkov and A.V. Bobu, "Cliques in High-Dimensional Random Geometric Graphs," *Applied Network Science 5*, No 1, Springer Open, 2020.

[Bailey (2005)] M.G. Bailey, "Malware Resistant Networking Using System Diversity," *Proceedings, 6^{th} ACM Special Interest Group for Information Technology Education Conference (SIGITE05)*, 2005.

[Balister (2005)] P. Balister, B. Bollobás, and M. Walters, "Continuum Percolation with Steps in the Square or the Disc," *Random Structures & Algorithms*, Wiley, Vol 26, No 4, April 2005.

[Bettstetter (2002a)] C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multihop Network," *Proceedings, 3^{rd} ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '02)*, 2002.

[Bettstetter (2002b)] C. Bettstetter and J. Zangl, "How To Achieve a Connected Ad Hoc Network with Homogeneous Range Assignment: An Analytical Study with Consideration of Border Effects," *Proceedings,* 4^{th} *IEEE International Workshop on Mobile and Wireless Communications Networks (IWMCN)*, 2002.

[Bettstetter (2003)] C. Bettstetter and J. Eberspacher, "Hop Distances in Homogeneous Ad Hoc Networks," *Proceedings,* 57^{th} *IEEE Semiannual Vehicular Technology Conference*, 2003.

[Bettstetter (2004)] C. Bettstetter "On the Connectivity of Ad Hoc Networks," *The Computer Journal*, Oxford University Press, Vol 47, Iss 4, Jan 2004.

[Bettstetter (2005)] C. Bettstetter and C. Hartmann, "Connectivity of Wireless Multihop Networks in a Shadow Fading Environment," *Wireless Networks* 11, Springer, 2005.

[Bicket (2005)] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and Evaluation of An Unplanned 802.11b Mesh Network," *Proceedings,* 11^{th} *ACM Annual International Conference On Mobile Computing and Networking (MobiCom '05)*, 2005.

[Bobrowski (2017)] O. Bobrowski and M. Kahle, "Topology of Random Geometric Complexes: A Survey," *arXiv:1409.4734v2 [math.AT]*, 23 Jul 2017.

[Bollobás (2006)] B. Bollobás and O. Riordan, *Percolation*, Cambridge University Press, 2006.

[Bollobás (2009)] B. Bollobás, S. Janson, and O. Riordan, "Line-of-Sight Percolation," *Combinatorics, Probability, and Computing (CPC)*, Vol 18, Cambridge University Press, 2009.

[Clauset (2009)] A. Clauset, C.R. Shalizi, and M.E.J. Newman, "Power-Law Distributions in Empirical Data," *arXiv:0706.1062v2 [physics.data-an]*, Feb 2009.

[Cooper (2011)] C. Cooper and A. Frieze, "The Cover Time of Random Geometric Graphs," *Random Structures & Algorithms*, Wiley, Vol 38, Iss 3, May 2011.

[Czumaj (2007)] A. Czumaj and X. Wang, "Communication Problems in Random Line-of-Sight Ad Hoc Radio Networks," *Stochastic Algorithms: Foundations and Applications, International Symposium on Stochastic Algorithms (SAGA 2007), Lecture Notes in Computer Science*, Vol 4665, Springer, 2007.

[Dall (2002)] J. Dall and M. Christensen, "Random Geometric Graphs," *Physical Review E* 66, 016121, 2002.

[Debany (2021)] W.H. Debany, "TDoA Geolocation Accuracy as a Function of the Number of Randomly Placed Sensors," AFRL-RI-RS-TP-2021-003, DTIC #AD1126223, https://apps.dtic.mil/sti/citations/AD1126223,Mar2021.

[Dousse (2004)] O. Dousse, M. Franceschetti, and P. Thiran, "The Costly Path from Percolation to Full Connectivity," *Proceedings, Allerton Conference*, 2004.

[Duncan (1986)] A.J. Duncan, *Quality Control and Industrial Statistics*, 5^{th} Ed, R.D. Irwin, 1986.

[Edney (2004)] J. Edney and W.A. Arbaugh, *Real 802.11 Security: Wi-Fi Protected Access and 802.11i*, Pearson Education, 2004.

[Erba (2020)] V. Erba, S. Ariosto, M. Gherardi, and P. Rotondo, "Random Geometric Graphs in High Dimension," *Physical Review E*, 102, 012306, 2020.

[Erdős (1959)] P. Erdős and A. Rényi, "On Random Graphs I," *Publicationes Mathematicae Debrecen*, Vol 6, 1959.

[Erdős (1960)] P. Erdős and A. Rényi, "On the Evolution of Random Graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, Vol 5, 1960.

[Erdős (1983)] P. Erdős, E.M. Palmer, and R.W. Robinson "Local Connectivity of a Random Graph," *Journal of Graph Theory*, Vol 7, 1983.

[Estrada (2015)] E. Estrada and M. Sheerin, "Random Rectangular Graphs," *Physical Review E* 91, 042805, 2015.

[Estrada (2016)] E. Estrada, S. Meloni, M. Sheerin, and Y. Moreno, "Epidemic Spreading in Random Rectangular Networks," *Physical Review E* 94, 052316, 2016.

[Frieze (2007)] A. Frieze, J. Kleinberg, R. Ravi, and W.H. Debany, "Line-of-Sight Networks," *Proceedings,* 18^{th} *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, 2007.

[Frieze (2009)] A. Frieze, J. Kleinberg, R. Ravi, and W.H. Debany, "Line-of-Sight Networks," *Special Issue of Combinatorics, Probability, and Computing (CPC)*, Vol 18, Cambridge University Press, 2009.

[Goldschmidt (1994)] O. Goldschmidt, P. Jaillet, and R. Lasota, "On Reliability of Graphs with Node Failures," *Networks*, Vol 24, Iss 4, 1994.

[Gupta (1998)] P. Gupta and P.R. Kumar, "Critical Power for Asymptotic Connectivity in Wireless Networks," *Stochastic Analysis, Control, Optimization and Applications*, 1998.

[Gupta (2000)] P. Gupta and P.R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, Vol 46, No 2, Mar 2000.

[Kim (2016)] S. Kim and H. Kim, "A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts," *International Journal of Forecasting*, Elsevier, https://doi.org/10.1016/j.ijforecast.2015.12.003,2016.

[Koskinen (2003a)] H. Koskinen, *Connectivity and Reliability in Ad Hoc Networks*, Master's Thesis, Helsinki University of Technology, 2003.

[Koskinen (2003b)] H. Koskinen, "Statistical Model Describing Connectivity in Ad Hoc Networks," *Proceedings, Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'03)*, 2003.

[Koskinen (2004)] H. Koskinen, "Quantile Models for the Threshold Range for k-Connectivity," *Proceedings,* 7^{th} *ACM International Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM '04)*, 2004.

[MathWorks (2020)] MathWorks Documentation, "Coefficient of Determination (R-Squared)," https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html.

[Meester (1996)] R. Meester and R. Roy, *Continuum Percolation*, Cambridge University Press, 1996.

[Messina (1987)] P.C. Messina, "Emerging supercomputer architectures," DOE/ER/25009-449, California Institute of Technology, Pasadena CA USA, 1987.

[Moler (1986)] C. Moler, "Matrix Computation on Distributed Memory Multiprocessors" in *Hypercube Multiprocessors 86*, SIAM, 1986.

[Newman (1993)] M.C. Newman, "Regression Analysis of Log-Transformed Data: Statistical Bias and Its Correction," *Environmental Toxicology and Chemistry: An International Journal*, Vol 12, Iss 6, 1993.

[Nicholson (1957)] M. Nicholson, *A Dictionary of American-English Usage Based on Fowler's Modern English Usage Elements of Style*, Signet reprint of hardcover edition published by Oxford University Press, 1957.

[Parker (2010)] P.A. Parker, G.G. Vining, S.R. Wilson, J.L. Szarka, and N.G. Johnson, "The Prediction Properties of Inverse and Reverse Regression for the Simple Linear Calibration Problem," *Journal of Quality Technology* 42 (4), 2010, https://ntrs.nasa.gov/api/citations/20110016499/downloads/20110016499.pdf.

[Penrose (1997)] M.D. Penrose, "The Longest Edge of the Random Minimal Spanning Tree," *The Annals of Probability* 7(2), May 1997.

[Penrose (1999)] M.D. Penrose, "On k-Connectivity for a Geometric Random Graph," *Random Structures & Algorithms*, Wiley, Vol 15, Iss 2, Sep 1999.

[Penrose (2003)] M.D. Penrose, *Random Geometric Graphs*, Oxford University Press, 2003.

[Penrose (2016)] M.D. Penrose, "Connectivity of Soft Random Geometric Graphs," *The Annals of Applied Probability* 26(2), Apr 2016.

[Press (2002)] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, 2002.

[Santi (2003)] P. Santi and D.M. Blough, "The Critical Transmitting Range for Connectivity in Sparse Wireless Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, Vol 2, No 1, Jan-Mar 2003.

[Sedgewick (1998)] R. Sedgewick, *Algorithms in C, Parts 1–4 (Fundamentals, Data Structures, Sorting, Searching)*, 3^{rd} Ed, Addison-Wesley, 1998.

[Ta (2009)] X. Ta, G. Mao, and B.D.O. Anderson, "On the Giant Component in Wireless Multi-Hop Networks," *Proceedings, IEEE Wireless Communications and Networking Conference (WCNC)*, 2009.

[Weisstein (2019a)] E.W. Weisstein, "Standard Error," *MathWorld—A Wolfram Web Resource*: https://mathworld.wolfram.com/StandardError.html, original date of article unknown.

[Weisstein (2019b)] E.W. Weisstein, "Hypercube Line Picking," *MathWorld—A Wolfram Web Resource*: https://mathworld.wolfram.com/HypercubeLinePicking.html, original date of article unknown.