NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

THE DISTANCE CENTRALITY: MEASURING STRUCTURAL DISRUPTION IN A NETWORK

by

Jonathan W. Roginski

June 2018

Dissertation Supervisor: Ralucca Gera

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2018	**3. REPORT TYPE AND DATES COVERED** Dissertation

4. TITLE AND SUBTITLE THE DISTANCE CENTRALITY: MEASURING STRUCTURAL DISRUPTION IN A NETWORK	**5. FUNDING NUMBERS**
6. AUTHOR(S) Jonathan W. Roginski	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This research provides an innovative approach to identifying the influence of vertices on the topology of a graph by introducing and exploring the neighbor matrix and distance centrality. The neighbor matrix depicts the "distance profile" of each vertex, identifying the number of vertices at each shortest path length from the given vertex. From the neighbor matrix, we can derive 11 oft-used graph invariants. Distance centrality uses the neighbor matrix to identify how much influence a given vertex has over graph structure by calculating the amount of neighbor matrix change resulting from vertex removal. We explore the distance centrality in the context of three synthetic graphs and three graphs representing actual social networks. Regression analysis enables the determination that the distance centrality contains different information than four current centrality measures (betweenness, closeness, degree, and eigenvector). The distance centrality proved to be more robust against small changes in graphs through analysis of graphs under edge swapping, deletion, and addition paradigms than betweenness and eigenvector centrality, though less so than degree and closeness centralities. We find that the neighbor matrix and the distance centrality reliably enable the identification of vertices that are significant in different and important contexts than current measures.

14. SUBJECT TERMS network, graph, neighbor matrix, distance centrality, graph topology, attack and defense, disruption, percolation, robustness, simulation	**15. NUMBER OF PAGES** 117
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

THE DISTANCE CENTRALITY: MEASURING STRUCTURAL DISRUPTION IN A NETWORK

Jonathan W. Roginski
Lieutenant Colonel, United States Army
MS, United States Military Academy, 1996

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN APPLIED MATHEMATICS

from the

NAVAL POSTGRADUATE SCHOOL
June 2018

Approved by:

Ralucca Gera
Department of Provost
Dissertation Supervisor

Carlos F. Borges
Department of
Applied Mathematics

Craig W. Rasmussen
Department of
Applied Mathematics

Gerry Baumgartner
Integrated Satellite and
Link Communications Division

David L. Alderson
Department of
Operations Research

Sean F. Everton
Department of
Defense Analysis

David Arney,
United States Military
Academy

Approved by: Craig W. Rasmussen, Chair, Department of Applied Mathematics

Douglas Moses, Vice Provost of Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This research provides an innovative approach to identifying the influence of vertices on the topology of a graph by introducing and exploring the neighbor matrix and distance centrality. The neighbor matrix depicts the “distance profile” of each vertex, identifying the number of vertices at each shortest path length from the given vertex. From the neighbor matrix, we can derive 11 oft-used graph invariants. Distance centrality uses the neighbor matrix to identify how much influence a given vertex has over graph structure by calculating the amount of neighbor matrix change resulting from vertex removal. We explore the distance centrality in the context of three synthetic graphs and three graphs representing actual social networks. Regression analysis enables the determination that the distance centrality contains different information than four current centrality measures (betweenness, closeness, degree, and eigenvector). The distance centrality proved to be more robust against small changes in graphs through analysis of graphs under edge swapping, deletion, and addition paradigms than betweenness and eigenvector centrality, though less so than degree and closeness centralities. We find that the neighbor matrix and the distance centrality reliably enable the identification of vertices that are significant in different and important contexts than current measures.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

CC	Closeness Centrality
DC	Distance Centrality
Deg	Degree Centrality
DoD	Department of Defense
EVC	Eigenvector Centrality
JCMCC	Journal of Combinatorial Mathematics and Combinatorial Computing
NPS	Naval Postgraduate School
SVD	Singular Value Decomposition
U.S.	United States
VEO	violent extremist organization
VIF	Variance Inflation Factor
WMD	weapons of mass destruction

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

The sentiment that it takes a village to raise a child certainly applies to the community required to guide a doctoral student successfully to commencement of life as a PhD.

First and foremost, I have to thank God. My time as a doctoral student gave me opportunities to rely each on the Father, His Son, Jesus Christ, and the Holy Spirit. The Father with his guiding hand both when things were going well and when they were not. Jesus Christ when I needed someone to give me strength. And the Holy Spirit to buttress my faith and give me clarity and access to my knowledge when I needed it the most. Thank you.

Quite literally, none of this work would have been possible without the vision and support of my advisor, Professor Ralucca Gera. Her willingness to explore, knowledge of the craft, and patience in mentoring were indispensable qualities in this journey. She has enough energy for two people, which was sometimes a good thing. If I found myself flagging, I could always draw energy from her drive. To describe all the reasons I am grateful to have worked for her would be another dissertation. So for now, I'll just say "thank you."

Professor Alderson's perspective made a profound impact on the way I see analysis and writing. It was my honor to work with Professor Arney; his leadership and guidance have left an indelible mark upon network science. I will be forever indebted to Professor Borges for his experience in life and math. A master motivator and empathetic counselor, his deft touch—not always soft and gentle—always seemed to be what I needed when I needed it. I appreciate Dr. Baumgartner's interdisciplinary insights and willingness to fund part of this work. Professor Everton's knowledge in the types of analysis done throughout this work was invaluable in identifying the state of the art so that we may advance it. Professor Rasmussen's understanding and tutoring made me a better mathematician and writer. And last but certainly not least, I appreciate the time Professor Bill Fox took out of his schedule to participate in my qualifying exams and dissertation defense. Your support to me has been so important since it began in 1994. Thank you.

Thank you to my family. You are my driving force. My raison d'être. I know these past years have been hard and we've all been through so much. Thank you for your hugs. My prayers are for many more. Thank you and I love you.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

You cannot solve the problems of today with the thinking used to create them.
-Attributed to Albert Einstein

The evolution of human intellect has long been defined by the progress of our science. From the Stone Age to the Bronze Age to the Iron Age, man progressed from the knowledge to fashion stone into tools to the knowledge to smelt iron. Later, we progressed from the Age of Industry defined by the First and Second Industrial Revolutions to the Information Age. New technology and one-off craftsmanship gave way to the scientific and technological advances of mass production. From quantities of scale arose unprecedented amounts of data and information which characterized our world until recent times. In the last quarter of the 20th century, a new science began to gain traction. Led by physicists, an interdisciplinary group of researchers began to see the world through the lens of connections between entities. Network science surfaced and we realized that man had transitioned from the Information Age to the Age of Connectivity.

The human experience at every level is networked, from networks of molecules that form our bodies through biological processes and laws to the electrical impulses transiting a network of synapses to move our bodies to the neural networks that define our sentience. We nourish our bodies through taking our place at the top of a complex food web (as in Figure 1.1) and gain fulfillment through interacting with others in social networks. We earn money in the midst of information and professional networks and spend it transiting road and air networks while on vacation.

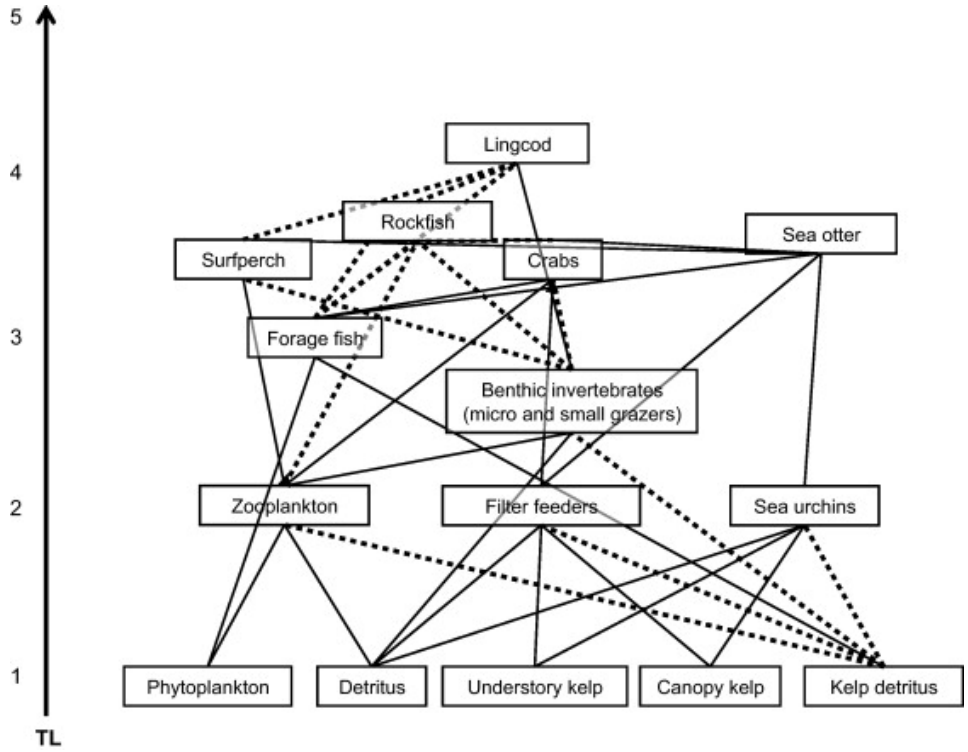


Figure 1.1: A Representation of the Kelp Forest Food Web, such as in the Monterey Bay. Source: [1].

The networks through which we navigate on a daily basis are vast, dynamic, and intricate. Given any one of these characteristics, these networks would not readily lend themselves to visualization and intuition, yet often all three manifest simultaneously. To make sense of such networks—called *complex networks*—network analysts must leverage modern techniques that "embrace the complexity" [2], [3] of the systems being analyzed, rather than introducing artificiality through dis-aggregation and simplification.

This research began viewing these problems of complexity with a United States Department of Defense (DoD) perspective. The Department of Defense is in the business of change management. On the attack, we seek to create maximal change in the network that characterizes our enemy. In defense, we guard our own network so any attack on our network realizes the smallest possible change. In humanitarian assistance and disaster response, our goal is to return the affected networks to their original state. We may aspire to create stability in a volatile environment by introducing new network elements or increasing the capacity of extant entities. In each of these scenarios, understanding network topology directly contributes to decisions made about critical resource allocation.

How do we conduct the analysis and make recommendations when structure matters? Currently, there are descriptive statistics at the graph and vertex level. Graph-level statistics provide insight into such qualities as how large the graph is (number of vertices and edges), number of steps to transit the graph (diameter), and relationships between friends (clustering). These statistics provide no insight into which individual vertices are most important. There is a suite of network centrality measures that describe which vertices are the most important in certain contexts. Degree centrality is a popularity measure that highlights which entity has the most direct connections to other entities. Vertices with high eigenvector centrality are those that are connected to important vertices. Betweenness centrality identifies which vertices in a network are on the most shortcuts between other vertices. Closeness centrality determines which vertices are the fewest steps from other vertices. Though these measures are useful in certain contexts, they do not provide insight into the extent to which a given entity exerts influence over network-level structure. It is here this work makes its contribution, solving the problem of identifying individual vertices with global structural impact.

This analysis is important to the military and beyond because the first action matters. Specifically, making the right action, first, matters. Army Doctrine Publication 3-0 (Operations) includes "surprise" as one of its 12 principles of joint operations and "exploiting the initiative" as part of its definition of the operational art [4]. The Chairman of the Joint Chiefs of Staff's National Military Strategy is guided by the President's National Security Strategy, and the mandate for network science techniques is prevalent through both documents. In Section II of the National Security Strategy, titled "Security," there are several subsections. For each, we interpret a corresponding complex network problem in which structure, surprise, and initiative matter [5].

- *Strengthen Our National Defense*. At its most abstract, the military is a multi-layer network of capabilities that reside in the Army, Navy, Air Force, and Marines. These capabilities are matched to the mandates of protecting citizens, preserving stability, providing humanitarian assistance and disaster relief, and building capacity in the networks of partner nations.
- *Reinforce Homeland Security*. We must simultaneously build resilience in the networks that define our quality of life, break down the networks that define terror,

extremist, and other illicit networks.

- *Combat the Persistent Threat of Terrorism*. The environment that enables terrorism is a network representing the convergence of ideological, financial, environmental, and survival elements. For the long term eradication of terrorism and extremism to be possible, we must effect fundamental changes to this network structure, which can only be done through understanding the relationships between elements.
- *Build Capacity to Prevent Conflict*. Where stability exists, conflict evaporates. Stability in an area may be modeled as a hierarchical network of needs. Instability persists where the needs of a population are not being met, identified through network analysis.
- *Prevent the Spread and Use of Weapons of Mass Destruction*. Proliferation and spreading are a well-known class of network problems, usually seen in the context of disease spreading; such techniques may be extended to inform prevention of WMD proliferation.
- *Confront Climate Change*. Though attempting to establish causation for climate change is a controversial topic, maintaining the security of our population against natural disasters is not. Network insights enable proactive protection of our coastal regions, infrastructure, and property against natural disasters springing from long term climate change or immediate-term events such as hurricanes, floods, fires, etc.
- *Assure Access to Shared Spaces*. Cyber (and cyberspace), space, air, and oceans connect the world and its people by enabling transit, communication, and resource flow. Understanding these flow networks enables the United States to be a leader in maintaining appropriate access for all to the Global Commons.
- *Increase Global Health Security*. As previously mentioned, disease proliferation is a well-known topic of network study. The United States led an effort to take a network approach to fighting the Ebola virus, which arguably help mitigate and localize the effects of this potential pandemic. Similar approaches will continue to increase global health security in the future.

The National Military Strategy details three military objectives, defined next, with the author's network interpretation [6].

- *Deter, Deny, and Defeat State Adversaries*. It is by identifying the United States'

relative over-match in capabilities to its adversary, the U.S. "[inflicts] damage of such magnitude as to compel the adversary to cease hostilities or render it incapable of further aggression" [6]. This over-match can be determined by analyzing the United States' capability network in relation to its adversary.

- *Disrupt, Degrade, and Defeat Violent Extremist Organizations (VEOs).* The defeat of VEOs is a network problem at every level from connections between people and leaders, planning, operations, support, finance, movement, and influence.
- *Strengthen our Global Network of Allies and Partners.* Considering the United States' network of allies is an abstraction of social network analysis. Such analysis with respect to allies and partners strengthens the team with which we operate in every geographic area and operational environment.

In this work, we take a graph-theoretic approach that may be applied to the network interdiction problem, and extended to other problems such as the influence of one entity over other entities in the network. We introduce the neighbor matrix, which identifies the "distance profiles" of each vertex in the network. In this context, distance is the "shortest path," or the fewest number of steps to transverse from one vertex to another. For each vertex, the neighbor matrix identifies the number of vertices that are one step, two steps, all the way out to the greatest number of steps possible in the network, the network diameter. The neighbor matrix contains a significant amount of structural information, including 11 characteristics often used to describe network topology. We use that topological information to introduce the distance centrality. Vertices with high distance centrality are those whose removal results in the largest change in the neighbor matrix. This neighbor matrix change measures the vertex's global structural influence by revealing the change one vertex has on the distance profiles of all other vertices.

We demonstrate the utility of the distance centrality by showing it highlights different vertices as important than betweenness, closeness, degree, and eigenvector centralities. Correlation and regression analysis show that distance centrality contains different elements of information than each of the other four centralities or the centralities in combination. The most important information provided by the distance centrality may be that it influences shortest paths like betweenness and closeness centrality, but it controls for redundancy. When there are multiple shortest paths of the same distance between

vertices, it appears important in the other centralities, but not in distance centrality. Distance centrality is able to identify the potential "load sharing" that can occur when a vertex is removed and there are multiple shortest paths between vertices. Removal of this vertex does not change the structure of the graph, even though it may be close to other vertices, on many shortest paths, or of high degree. We highlight this contribution in several synthetic and actual social networks.

There are operations research-specifically optimization-and network science methods that apply to the problem of network interdiction. Optimizers may solve such problems by minimizing the maximum flow of the commodity across the network [7]. These problems may be extended to situations in which the owner of the network can take steps to protect or "harden" the network before an attack [8]. There are graph-theoretic and network science techniques that provide insight into the vulnerability of a network to random and targeted attacks [9]. Both of these technique may be used individually or in concert to provide insight into the problems previously outlined. The approach developed in this work provides solutions to problems of network structure, which are not directly answered through optimization techniques and pinpoints areas of structural weakness that are not directly answered through current graph-theoretic approaches..

Though many of the findings herein specifically relate to Department of Defense situations, we recognize the applicability of this analysis well beyond the DoD. Though in the DoD we attack other military structures, we see a direct analog to attacking disease processes—disrupting the disease process network so the disease becomes incompatible with life. Though we may defend a unit in the DoD, we must protect power infrastructure from failure and rolling brownouts or blackouts. Police may respond to an area afflicted by riots and need to restore order—restoring the area networks to normalcy. A company may look to diversification to protect against a market downturn.

We seek knowledge about network structure that provides insight into which entities exert the most influence over network topology. We thoroughly examine the property of distance in networks and use this property to understand network structure, including the impact of individual vertices on that structure. We find that the innovative techniques outlined in this work bring to light network insights currently overlooked.

The chapters that follow include a background chapter which serves to provide definitions for terminology used throughout this work. The background chapter also provides an overview of certain linear algebra, graph theory, complex networks, and statistics concepts used later in the text. Chapter 3 introduces and explores the neighbor matrix, viewing graphs through the lens of the distances between vertices in the graph. Chapter 4 introduces the distance centrality, which measures the amount of change removing a vertex causes on the neighbor matrix, or all the distances in the graph. Chapter 4 also highlights the unique contribution of the distance centrality in the context of three synthetic networks and three social networks. Chapter 5 is an analysis of the robustness of the distance centrality, comparing the amount distance centrality changes when small perturbations are made to a graph to betweenness, closeness, degree, and eigenvector centralities. Finally, Chapter 6 highlights some of the directions for extending these rich topics.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Background

Parts of this chapter have been accepted to be published in the *Journal of Combinatorial Mathematics and Combinatorial Computing* [10].[1]

Chapter 2 is intended to define terminology as that is not standard across all graph theory and network science literature. This chapter is also intended to make the rest of the work accessible to those readers without little background in graph theory, linear algebra, complex networks, and statistics. Section 2.1 overviews the terms and concepts from these disciplines used throughout the remaining chapters of this work. As much of this work pertains to graph and network information encoded into matrices, Section 2.2 surveys some of the matrices currently used to describe graphs. Section 2.3 describes metrics used to describe complex networks, as well as the mechanics behind synthetic network formation models and a description of social networks used to demonstrate utility of methods and techniques developed in this research. Since network analysis and interdiction is provided as a motivation for many Department of Defense problems, Section 2.4 provides an overview of previous graph theoretic (network science) and operations research (optimization) work in this area. Finally, Section 2.5 provides an overview of statistics techniques germane to this work. Readers with working knowledge of these areas may skip directly to subsequent chapters without loss of context.

2.1 Definitions

The following three subsections include definitions of terms from graph theory, linear algebra, and complex networks terms appear later in the work.

[1]This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States. JCMCC will claim and protect its copyright in international jurisdictions where permission from JCMCC must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

2.1.1 Graph Theory

This work is limited to analysis of *simple, connected* graphs (no multiple edges nor loops) $G = (V(G), E(G))$, where $V(G)$ is the vertex set and $E(G)$ the edge set [11]. We annotate the vertices of a graph $v_1, v_2, \ldots, v_n$, where n is the number of vertices in the graph. Two vertices are called *adjacent* if there is an edge joining them, otherwise they are *nonadjacent* [11]. The *chromatic number* of a graph is the minimum number of colors used to color vertices of a graph so that no two adjacent vertices share the same color [11]. A *cycle* is a sequence of adjacent vertices that begins and ends at the same vertex, but repeats no other vertex [11]. A graph with no cycles is called *acyclic* and a connected, acyclic graph is called a *tree* [11]. The *girth* of a graph is the length of the graph's shortest cycle; the girth of an acyclic graph is infinity [11]. A *circuit* is a sequence of adjacent vertices that begins and ends at the same vertex, but may repeat vertices [11]. A graph H is a *subgraph* of G (denoted $H \subseteq G$) if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ [11]. If $H \subseteq G$ and either $V(H)$ is a proper subset of $V(G)$ or $E(H)$ is a proper subset of $E(G)$, then H is a proper subgraph of G, $H \subset G$ [11].

Two graphs G and H are *isomorphic* if there is a one-to-one correspondence ϕ from $V(G)$ to $V(H)$ such that $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$ [11], [12]. A *graph invariant* is a property that has the same value for every pair of isomorphic graphs [11]. An *automorphism* is an isomorphism from a graph G to itself [11]. Since the identity and inverse are both automorphisms and the composition of two automorphisms is itself an automorphism, the set of all automorphisms of a graph G forms a *group* under the operation of composition [11]. We denote the *automorphism group* of G by $Aut(G)$ [11]. Suppose v is a vertex of graph G. The set of all vertices to which v may be mapped by an automorphism of G in an *orbit* of G [11]. The automorphism relating two vertices is an equivalence relation resulting in equivalence classes that are the *orbits* of G [11]. We define the orbit of vertex i in graph G as $o(i_G)$. If G contains a single orbit, then G is *vertex-transitive* [11].

A *triangle* occurs in a graph when a path of length two is closed; that is, when a path of length two is associated with a cycle of length three [11]. A *component* of G is a connected subgraph of G that is not a proper subgraph of any other connected subgraph of G [11]. The connectivity number $\kappa(G)$ of a graph is the minimum number of vertices that

must be removed from a connected graph to create multiple components [11]. A graph G is said to be $k-connected$ if and only if $\kappa(G) \geq k$ [11]. The connectivity number $\lambda(G)$ is the minimum number of edges whose removal from a connected graph will create multiple components [11]. A graph G is said to be $k-connected$ if and only if $\lambda(G) \geq k$ [11].

A shortest path between two nonadjacent vertices is called a *geodesic* [11]. The length of a geodesic between two vertices is the *distance* between the vertices [11]. The *average distance in G* is computed as the fraction of all pairwise distances out of all possible distances, $\frac{2}{n(n-1)} \sum_{i \neq j} d(v_i, v_j)$ [11]. By convention, $d(x,y) = \infty$ if x and y are in different components. For a given vertex x in graph G, y is a *k-step (or k-hop) neighbor* of x if $d(x,y) \leq k$ for $1 \leq k \leq n-1$ [11]. Wu and Dai [13] introduced a specification of this measure: a vertex x is an *exact k-hop neighbor* of y if $d(x,y) = k$ for some $1 \leq k \leq n-1$.

The *eccentricity* of a vertex v_i, $e(v_i)$, is the maximum distance from v_i to any other vertex in G: $e(v_i) = max\{d(v_i, v_j) | v_j \in V(G)\}$ [11]. The *radius of G* is the minimum eccentricity among the vertices of G: $rad(G) = min\{e(v_i) | v_i \in V(G)\}$ [11]. The *diameter of G* is the maximum eccentricity among the vertices of G: $diam(G) = max\{e(v_i) | v_i \in V(G)\}$ [11]. The *center of G*, $Cen(G)$, is the subgraph induced by those vertices of G having minimum eccentricity: $Cen(G) = G[\{v_i \in V(G) | e(v_i) = rad(G)\}]$ [11]. The *periphery of G*, $Per(G)$, is the subgraph induced by those vertices of G having maximum eccentricity: $Per(G) = G[\{v_i \in V(G) | e(v_i) = diam(G)\}]$ [11]. The k^{th} *power* of an undirected graph G is the graph $G^k = (V(G), E(G^k))$, where $E(G^k) = \{v_i v_j | d_G(v_i, v_j) \leq k\}$ [11].

A vertex v_i's *degree* denotes the number of vertices to which v_i is adjacent [11]. The degree of vertex i in graph G is denoted by $deg_G(i)$ [11]. The *degree sequence of G* of a graph is an integer sequence $d_1, d_2, \ldots, d_n$ where $n = |V(G)|$ and d_i is the degree of vertex i [11]. We will use the convention that the sequence is non-increasing. The *density* of a graph is the ratio of the number of edges actually present in the graph to the number of possible edges [11]:

$$\frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}.$$

2.1.2 Linear Algebra

There is a significant body of graph theoretical research that brings the disciplines of algebra and graph theory together. Aptly named algebraic graph theory, foundational texts include works by Godsil and Biggs [14], [15]. The works of Chung, Cvetković et al. [16]–[19], were most helpful for understanding a branch of algebraic graph theory called spectral graph theory, leveraging eigenvalues and eigenvectors in the endeavor to "[unlock] a thousand secrets about graphs" [16]. Finally, this work uses Leon's eighth edition text [20] and the fourth edition of Golub and VanLoan [20], [21] to guide a broader linear algebraic exploration of certain matrices that describe graphs.

A norm may be used to determine the "size" of a vector or matrix and is a tool used in this work to determine the distance between matrices. To understand norms in terms of vectors (consider matrices to simply be vectors in $m \times n$ space), we begin with a description of what it means to be a vector space. A set V, defined under the operations of addition and scalar multiplication forms a *vector space* if the following axioms are satisfied [20]:

A1: Vector addition is commutative; that is, for every pair of vectors, $\mathbf{x}$ and $\mathbf{y}$ in V, the order of addition does change the result of the addition.

A2: Vector addition is associative; that is, for any three vectors in V $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, the three vectors may be grouped in any way and the result of the vector addition will be the same.

A3: The additive identity, $\mathbf{0}$ exists, so the sum of any vector in V with $\mathbf{0}$ results in the original vector.

A4: The additive inverse exists, so any vector added to its additive inverse results in the additive identity.

A5: Scalar multiplication is distributive; that is, the product of a scalar with the sum of vectors yields the same result as summing the product of multiplying the scalar by each vector.

A6: Multiplication of the sum of scalars by a vector is also distributive; that is the product of a sum of scalars with a vector yields the same result as summing the product of each scalar with the vector.

A7: Multiplication of a product of scalars by a vector is also distributive. Multiplying a vector by the product of scalars gives the same answer as multiplying the product of

each scalar with the vector.

A8: The multiplicative identity, 1 exists so that multiplication of any vector by 1 yields the original vector [20].

The vector space V is a *normed linear space* [20] if, for every vector $\mathbf{v}$ in V, there is an associated real number $||\mathbf{v}||$, called the *norm* of $\mathbf{v}$, satisfying:

1. $||\mathbf{v}|| \geq 0$ with equality if and only if $\mathbf{v} = 0$
2. $||\alpha\mathbf{v}|| = |\alpha|||\mathbf{v}||, \forall$ scalars α
3. $||\mathbf{v}+\mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||, \forall \ \ \mathbf{v}, \mathbf{w} \in V$ (Triangle Inequality).

There are many quantities that fit the definition of a norm. The $1-norm$ is the sum of absolute values of the vector entries (or in a matrix, the max of column sums of absolute values of entries). The $\infty-norm$ is the maximum of the absolute values of the vector entries (or in a matrix, the max of row sums of absolute values of entries). In this work, we use the Frobenius norm due to its familiarity as a distance calculation in $\mathbb{R}^2$ [20]:

$$||\cdot||_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 \right)^{\frac{1}{2}}.$$

An *eigenvalue* of a matrix A is a scalar λ, that coupled with its associated *eigenvector* $\mathbf{x}$, satisfies Equation 2.1 and equivalently, Equation 2.2 [20]. Since $|A - \lambda I|$ is *singular*, or not invertible, $|A - \lambda I| = 0$, the *characteristic equation*, with the *characteristic polynomial* those terms left of the equals sign [20].

$$A\mathbf{x} = \lambda\mathbf{x}, \tag{2.1}$$

$$(A - \lambda I)\mathbf{x} = \mathbf{0}. \tag{2.2}$$

The *spectrum* of graph G, denoted by $Sp(G)$, comprises the full set of eigenvalues associated with the graph's adjacency matrix or Laplacian matrix [16]–[19]. We write $Sp(G) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ to represent the spectrum of an adjacency or Laplacian matrix,

which is synonymous with the spectrum of a graph. We use the technique of *singular value decomposition*, or *SVD*, in an effort to achieve some of the analytical richness spectral techniques provide for square matrices. Singular value decomposition is a matrix factorization of the form seen in Equation 2.3 [20], [21].

$$A = U\Sigma V^T. \tag{2.3}$$

All the matrices we will consider are $m \times n$ matrices with $m > n$. The SVD results in the original matrix A being factored into the product of three matrices:

- An $m \times m$ orthogonal matrix U (i.e., the columns form a set of orthogonal unit vectors in $\mathbb{R}^m$)
- A diagonal $m \times n$ matrix Σ with diagonal entries that satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. These diagonal entries are called *singular values* and are the square root of corresponding eigenvalues of $A^T A$: $\sigma_i = \sqrt{\lambda_i}$, where λ_i is an eigenvalue of $A^T A$, $\forall i \ \ (1 \leq i \leq n)$.
- An $n \times n$ orthogonal matrix V^T: where the columns of V form a set of orthogonal unit vectors in $\mathbb{R}^n$.

2.2 Matrix Representations of Graphs

Matrices are useful structures in the representation, storage, and analysis of graphs and networks. There are many different matrices that encode information about graphs and networks from a variety of perspectives. Some matrices, such as the adjacency matrix, degree matrix, and Laplacian enumerate adjacencies. Others, such as the distance and reciprocal distance matrix identify the number of steps or hops between vertices. This section begins with an illustration depicting several different matrices and how they relate to each other, then offers a description of each matrix. The chapter is not an exhaustive survey of matrix representations of graphs. Other matrices, such as the modularity matrix used to describe communities in graphs and considered outside the scope of this research are not included.

Figure 2.1 describes the flow of Section 2.2. From a graph, the adjacencies between

vertices may be identified, resulting in the adjacency matrix. The adjacency matrix may be summarized with the degree matrix and both used to calculate the Laplacian. The distance matrix may be formed by operating on the adjacency matrix or by counting the the shortest paths between all pairs of vertices in the graph. From the distance matrix, we may calculate the reciprocal distance matrix and the neighbor matrix. The neighbor matrix may also be calculated from the adjacency matrix or from the graph itself, using an all-pairs shortest paths algorithm.

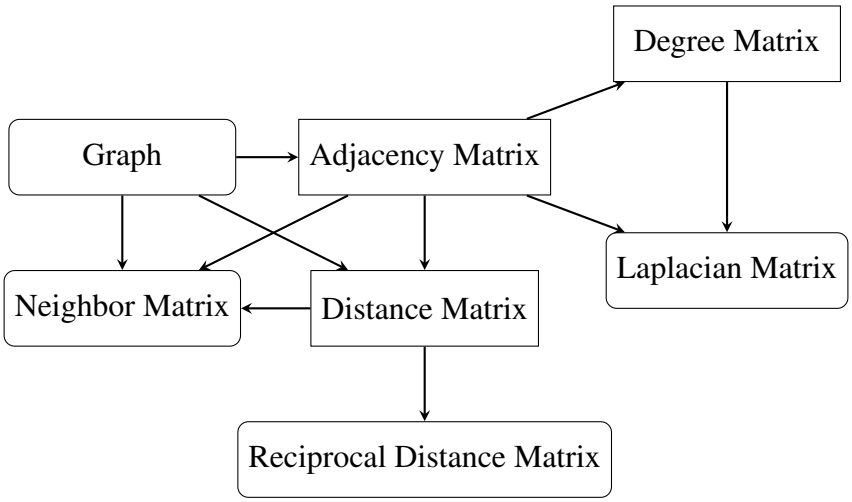


Figure 2.1: Some Matrices Used to Describe Graphs

Throughout Section 2.2, we will use Figure 2.2, which depicts a graph G, to provide examples of these matrices.

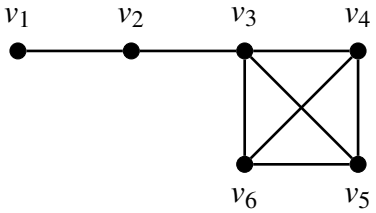


Figure 2.2: Graph G, Used to Highlight Matrix Representations of Graphs

The *identity matrix* has the value 1 as each diagonal value, with 0 elsewhere [20]. A *permutation matrix*, P is a "permuted" identity matrix, in which two rows have been exchanged [20]. The set of permutations includes the identity permutation, in which no

rows have been exchanged. So, where n is the number of rows in the identity matrix, there are $n!$ permutation matrices. The transpose of a permutation matrix is its inverse, so $P^T P = I$, where I is the identity matrix.

2.2.1 The Matrices of Adjacencies

The adjacency matrix, degree matrix, and Laplacian summarize the adjacencies of a graph.

The Adjacency Matrix

The *adjacency matrix* may be considered the "first matrix of graph theory" for its simplicity, descriptive power of graphs, and its place as the foundational element from which other graph-related matrices are derived. The adjacency matrix, $A(G)$, is an $n \times n$ matrix, with entries a_{ij} described in Definition 1 [11], [16]–[19].

Definition 1.

$$a_{ij} = \begin{cases} 1, & \textit{vertex } i \textit{ adjacent to vertex } j \\ 0, & \textit{otherwise} \ . \end{cases}$$

In the case of a simple, connected graph (such as those studied in this work), the adjacency matrix is symmetric [11], [16]–[19]. For the graph depicted in Figure 2.2, Figure 2.3 is the corresponding adjacency matrix, $A(G)$.

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 2.3: Adjacency Matrix for Graph G

We use the common convention of associating row i (or column i) of the adjacency matrix with vertex i of its described graph. Given an adjacency matrix, it is straight forward to draw an instance of a graph. However, if n is the number of vertices, a graph may be re-labelled in $n!$ ways—each potentially resulting in a different adjacency matrix for an isomorphic graph. The re-labeling of these vertices is the same as swapping respective

rows and columns in the adjacency matrix, which may be accomplished through the use of permutation matrices, P [20], [21]. Then, as shown in Equation 2.4, we have two different adjacency matrices A(G) and B(G) which represent the same graph, an issue we address later in this work

$$A(G) = P \cdot B(H) \cdot P^T. \tag{2.4}$$

As the adjacency matrix fully characterizes a graph, it is rich in information. Upon inspection, we note properties such as the number of vertices, edges, and graph *density*, which is the percentage of possible edges actually present in the graph. Through the analysis of the adjacency matrix's characteristic polynomial and spectrum, we are able to uncover more information about the graph, such as graph connectivity, the number of components, the number of triangles, girth g, number of circuits of a certain length, bounds on the chromatic number, and many more characteristics. The works by Biggs, Chung, and Cvetković present a detailed treatment of spectral graph theory [15]–[18].

The Degree Matrix

The *degree matrix* for graph G, $Deg(G)$, is an $n \times n$ diagonal matrix with diagonal entries Deg_{ij} that represent the vertex's degree (adjacency matrix row sum), see Equation 2.5 [14], [15].

Recall: entries of the adjacency matrix A are denoted a_{ij}.

$$Deg_{ij} = \begin{cases} \sum_{i=1}^{n} a_{ij}, & i = j \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

For the graph depicted in Figure 2.2, we show the corresponding degree matrix, $Deg(G)$, in Figure 2.4.

$$Deg(G) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

Figure 2.4: Degree Matrix for Graph G in Figure 2.2

The Laplacian Matrix

The *Laplacian matrix* for graph G, $L(G)$, may be defined in two ways, the second a normalized version of the first [16]–[19].

First, $L(G) = Deg(G) - A(G)$, where entries l_{ij} are defined by Equation 2.6 [18]. Recall Equation 2.5 for the calculation of $Deg(G)_i$.

$$l_{ij} = \begin{cases} Deg(G)_{ij}, & i = j \\ -1, & i \neq j \text{ (vertex } i \text{ adjacent to vertex } j) \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

For the graph depicted in Figure 2.2, we show the corresponding Laplacian matrix, $L(G)$, in Figure 2.5.

$$L(G) = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 4 & -1 & -1 & -1 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & -1 & -1 & 3. \end{bmatrix}$$

Figure 2.5: Laplacian Matrix for Graph G in Figure 2.2

The entries of the Laplacian may be normalized, with entries as shown in Equation 2.7 [16].

$$l_{ij} = \begin{cases} 1, & i = j \text{ and } d_i \neq 0 \\ -\frac{1}{\sqrt{Deg(G)_i \cdot Deg(G)_j}}, & \text{vertex } i \text{ adjacent to vertex } j \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

The Laplacian has all the information of the adjacency matrix and more. It is the matrix most often used when conducting graph spectral analysis, especially in the areas of algebraic connectivity, congruence, and equivalence [15]–[18], [22].

2.2.2 The Matrices of Distances

The matrices described in Subsection 2.2.1 answered the question: is there an edge between vertex i and vertrex j? A natural extension of this question is to inquire: how far apart are vertices i and j? The matrices in this subsection answer that very question.

The Distance Matrix

We begin with the distance matrix, defined in [23] many applications, such as in [24], [25]. This matrix should not be confused with the Euclidean Distance Matrix in [26]–[28], which identifies the Euclidean distance between n points in p dimensions. We will refer to the formerly-introduced *distance matrix* for graph G, $D(G)$, as an $n \times n$ matrix in which each entry d_{ij} is the length of the shortest path from vertex i to vertex j; $d_{ij} = 0$ when $i = j$. Direction of travel does not change the distance traversed, so $d_{ij} = d_{ji}$ and the distance matrix is symmetric.

For the graph depicted in Figure 2.2, the corresponding distance matrix, $D(G)$, is depicted by Figure 2.6.

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 3 & 3 \\ 1 & 0 & 1 & 2 & 2 & 2 \\ 2 & 1 & 0 & 1 & 1 & 1 \\ 3 & 2 & 1 & 0 & 1 & 1 \\ 3 & 2 & 1 & 1 & 0 & 1 \\ 3 & 2 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 2.6: Distance Matrix for Graph G in Figure 2.2

In [25], Mihalić et al. provide a detailed survey of the distance matrix. Not only does the paper highlight the wide application of the distance matrix from rigorous mathematically-oriented fields like applied mathematics, physics, and chemistry but also fields less rigorously mathematically oriented such as anthropology, psychology and history. The work details applications of invariants such as polynomials, spectra, and determinants in the field of chemistry.

The Reciprocal Distance Matrix

The *reciprocal distance matrix* for graph G, $RDM(G)$, is an $n \times n$ matrix in which each entry rdm_{ij} is the reciprocal of the length of the shortest path from vertex i to vertex j, with the convention of $rdm_{ij} = 0$ when $i = j$ [29]. For the graph depicted in Figure 2.2, the corresponding reciprocal distance matrix, $RDM(G)$ is shown by Figure 2.7.

$$RDM(G) = \begin{bmatrix} 0 & 1 & 0.5 & 0.\overline{3} & 0.\overline{3} & 0.\overline{3} \\ 1 & 0 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0 & 1 & 1 & 1 \\ 0.\overline{3} & 0.5 & 1 & 0 & 1 & 1 \\ 0.\overline{3} & 0.5 & 1 & 1 & 0 & 1 \\ 0.\overline{3} & 0.5 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 2.7: Reciprocal Distance Matrix for Graph G in Figure 2.2

The reciprocal distance matrix was defined by Ivanciuc et al. [29] to more accurately model the interactions of molecules in chemistry. The Wiener Index W is a topological index introduced to predict the boiling points of molecules comprised of single-bonded hydrogen and carbon atoms [30]. This index may be calculated by taking the half-sum of distance matrix entries. The authors of [29] observed that W increases with distance between molecules. However, interactions should decrease with an increase in distance between molecules. This observation resulted in the introduction of the reciprocal distance matrix to align matrix entries with the manifestation of physical phenomena.

2.3 Networks

As defined in Subsection 2.1.1, simple graphs are ordered pairs of vertices and edges. At their most abstract level, networks may be represented as graphs. Networks include more

information than these simple graphs, including descriptive information about graph elements, such as the weight or strength of connections and vertex heterogeneity. Where graphs are typically static, networks may be dynamic, growing and/or changing in accordance with some rule set that is either known or unknown. Networks may model systems that exhibit the nonlinear interactions, self-organization, and emergent behaviors that characterize *complex systems* [31]. Such networks are *complex networks*. The primary reference used herein for network and complex network terminology is Newman [32].

A *stub* is a vertex and associated edges that are not yet connected to other vertices [32]. Stubs are most commonly used in graph formation algorithms, such as the configuration model of random graph generation in which a random graph is generated from a given degree sequence by "connecting" stubs [32]. The *average degree of G* is

$$\frac{\sum_{v_i \in V(G)} deg(v_i)}{n} = \frac{2m}{n},$$

where m is the number of edges in the graph [11], [32].

The *average clustering coefficient* is the ratio of triangles in the graph to the number of connected triples (i.e., connected subgraphs on three vertices) [32]:

$$\text{Average Clustering Coefficient} = \frac{\text{number of triangles x 3}}{\text{number of connected triples}}.$$

Pearson's correlation coefficient in graphs is a measure of *assortative mixing*, i.e., the extent vertices with high degree are adjacent to each other [32]. See Newman's paper [33] for a detailed treatment of the topic. The *s-metric* determines the extent to which the graph being examined has a "hub-like" core; see [34], [35].

A variety of "centrality" measures serve to provide insight into which vertices are the most *influential* in a graph [32], [36]–[38]. The word influential is emphasized because there is no commonly accepted definition in a graph topological context (and one is not offered here). We will leverage a basic understanding of "importance" in this work: a vertex with high centrality is more "important" in some sense than a vertex of smaller centrality. The *degree* of vertex i is sometimes referred to as *degree centrality* of vertex i (degree centrality is often normalized through division by $n-1$). *Eigenvector centrality* is

calculated using an iterative process that assigns importance to a vertex based upon the importance of its neighbors. *Closeness centrality* is a measure of the distance from a vertex to all other vertices. It has many formulations, all related to calculation of the inverse mean distance from a vertex to all other vertices. *Betweenness centrality* is the extent to which a given vertex lies on the shortest paths to other vertices. See Figure 2.8. Vertex v_{10} has the highest degree centrality, while v_9 has the highest closeness centrality, v_8 the highest betweenness, and v_4 the highest eigenvector centrality.

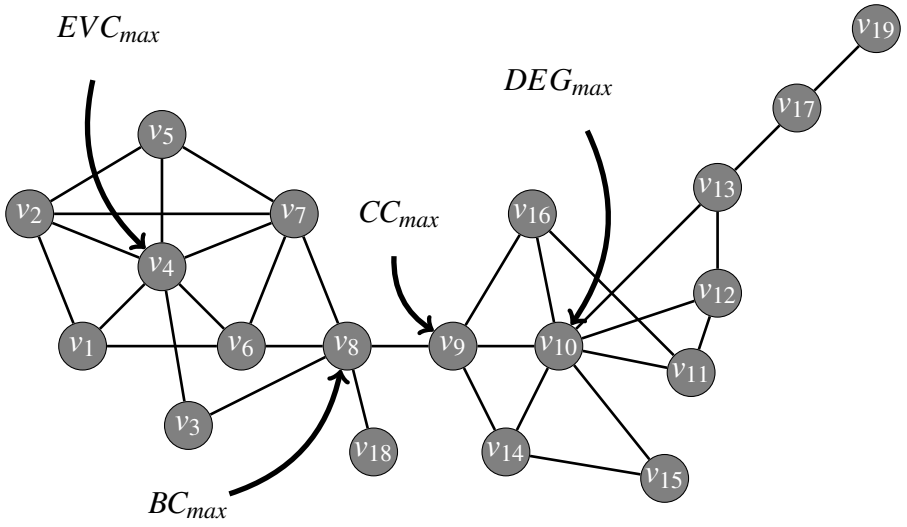


Figure 2.8: Graph H, Used to Illustrate Centrality Measures. Adapted from [39].

We may be interested in how the values of measures, metrics, and properties change with changes in the graph or network being studied. One method for examining these changes is "site" or "bond" *percolation* in which vertices or edges are deleted or added to the network, respectively.

2.3.1 Synthetic Graph Formation Models

This research uses six graphs and networks to demonstrate the validity of its findings, three synthetic graphs and three actual social networks. These graphs and networks are described in the following subsections.

Erdős-Rényi Random Graph Model

The paper "On Random Graphs I," published in 1959 by Erdős and Rényi is considered by many to be a foundational work in the field of network science [40]. An Erdős and Rényi random graph is one of the $\binom{\binom{n}{2}}{m}$ graphs with n vertices and m edges, chosen with uniform random probability [40]. Newman refers to this graph as $G(n,m)$ and discusses its relationship to a slightly different model, $G(n,p)$, introduced by Gilbert [41]. In this model, the number of edges is not fixed but the probability of connection between any two vertices is fixed (p) [41], [42].

Watts-Strogatz Small World Model

Watts and Strogatz observed that few interesting networks in our world are either purely regular (every vertex having the same number of neighbors) or completely random (as in the Erdős-Rényi model of graph formation) [43]. They introduced a model that begins with a *ring lattice* in which vertices are arranged in a ring and each vertex is adjacent to the next k vertices. The edges in the graph are then "re-wired" to introduce disorder [43]. From a ring lattice composed of n vertices and k edges per vertex, each edge is re-wired at random with probability p [43]. This model results in a lattice when $p = 0$, a completely random graph with $p = 1$, and something in between with $0 < p < 1$. These graphs can be highly clustered and still have small path lengths, combining characteristics of lattices and random graphs. The name "small world" network stems from the game in which two seemingly unrelated entities are counter-intuitively few "steps" from each other in the network [43].

Barabási-Albert Preferential Attachment Model

The preferential attachment model of graph formation was popularized by Barabási and Albert in 1999 [44]. This model of graph formation and growth has been observed to well-describe many complex systems, including:

- The Internet and World Wide Web
- Collaboration and Citation Networks (e.g., 6 Degrees of Separation)
- Transportation Networks (e.g., Airlines)
- Protein-Protein Interactions

The preferential attachment model results in degree sequences that follow a power law distribution [44]. That is, there are many low degree vertices and a relatively small number of high degree vertices, called hubs [44]. Though Barabási and Albert are synonymous with preferential attachment, it should be noted that both the phenomena of power law distributions in networks and preferential attachment had been previously noted in other works. In 1896, Vilfredo Pareto observed that 80% of the wealth of a country was held by 20% of its people [45]. Yule [46] and Simon [47] recognized the power law distribution as underlying many phenomena, such as plant mutations, populations of cities, distribution of incomes, and distribution of species among genera. Zipf is credited with recognizing power laws in the distribution of word frequencies and city sizes [48]. Finally, Price's "cumulative advantage" model is similar to preferential attachment in that the probability of a new vertex connecting to an existing vertex is related to the degree of the existing vertex and is updated after each vertex enters [49].

The formulation of the Barabási-Albert model is well-suited to generating large-scale models which reflect contemporary phenomena. As with the Price model, entities entering a network are proportionately more likely to form an initial connection with an entity which is itself well-connected [44]. Suppose we begin with a network that has n_0 vertices. At every time step, a vertex v_n enters with $m \leq n_0$ edges that links v_n to m other vertices in the network [44]. The probability Π that the new vertex forms a connection with an existing vertex i is proportional to the degree of that existing vertex, k_i, so that $\Pi(v_n, i) = \frac{k_i}{\sum_{j=1}^{n} k_j}$. After t time steps, this model leads to a graph with $t + n_0$ vertices and mt edges [44].

In summary, Table 2.1 reports some of the descriptive statistics of the synthetic graphs used throughout this work.

Table 2.1: Synthetic Graphs Analyzed

Graph	# Vertices	# Edges	Diameter
ER Random Graph	90	140	9
WS Small World	100	200	8
BA Preferential Attachment	100	291	5

2.3.2 Actual Social Networks

In addition to the classes of synthetic graphs outlined in Subsection 2.3.1, this work includes analysis of several actual social networks. As defined in Wasserman and Faust, social networks identify the relationships between people [50], enabling a network science approach to understanding social phenomena. See Table 2.2 for summary statistics.

Table 2.2: Actual Graphs Analyzed

Graph	# Vertices	# Edges	Diameter
Karate Club	34	78	5
Noordin Terror	145	560	6
Gera LinkedIn	180	412	13
Roginski LinkedIn	474	5,263	9

W.W. Zachary introduced a social network depicting the karate club of a university in the 1970*s* [51]. This network has 34 vertices and 78 edges and is depicted in Figure 4.7.

Naval Postgraduate School students compiled the Noordin Top terrorist network from the 2006 International Crisis Group report: *Terrorism in Indonesia: Noordin's Networks* [52]. This network has 145 vertices, 560 edges, and is shown in Figure 4.9.

For the final network analyzed in this work, the author used a web scraping tool offered by SociLab to gather personal connection data from LinkedIn to compile his own and Advisor's LinkedIn networks [53]. The web scraping tool gathered the subject's list of connections, then gathered information about which of those connections were themselves connected. This resulted in a network referred to as an "ego" network in sociology. Analysis of these networks was completed after removal of the original subject vertex.

2.4 Graph and Network Interdiction Models

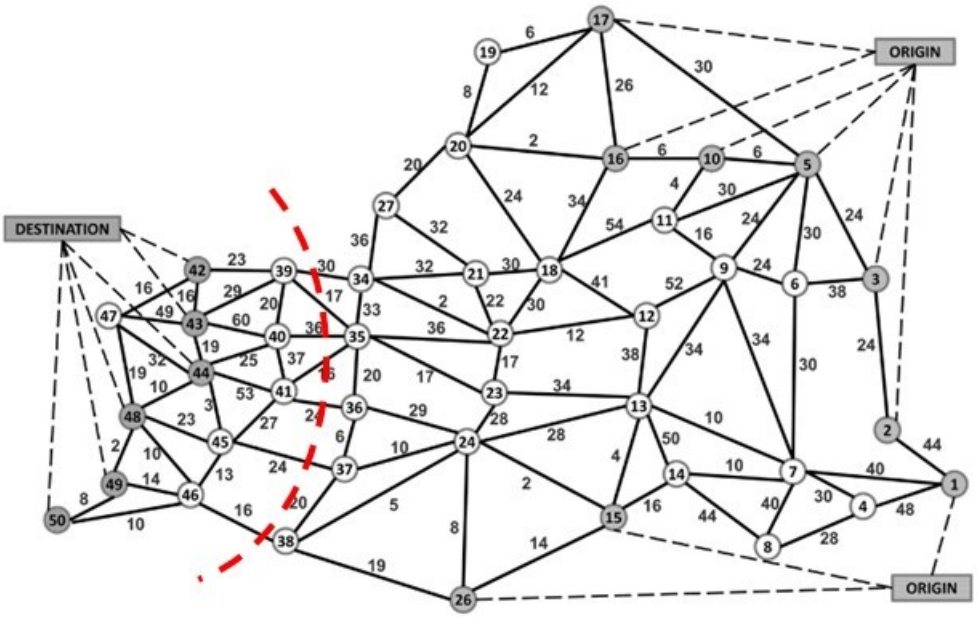
Chapter 1 introduced this work as beginning with military problems of complexity. At the heart of many such problems is a complex network that models friendly, enemy, or environmental considerations. When given a mission to accomplish, the plan is to result in the generation of some desired effect in the operational environment. Often, we direct influence toward the parts of the network deemed "important" in some way. Identifying network entities with which to interact and determining potential effects can be the result

of analysis using graph theory (network science) and/or operations research (network optimization) techniques. Though fundamentally different, the two disciplines are related in that they both provide insight into seeking or guarding against changes in the functionality of the system modeled by the graph or network.

Graph theoretic models analyze the effect of potential action on the structure of the graph, limited to vertices, edges, and graph topology [54]–[56]. Classic examples include the reliability of probabilistic graphs in which vertices and edges are active with a certain probability, including application in building reliable automated systems [57], relay circuits [58], communication networks [59], highway transportation system [60]. Wilkov provides a detailed survey of early usage of graph theory techniques to determine the reliability of graphs underlying computer networks based upon certain requirements for the connectivity of all vertex pairs, minimum graph diameter, minimum number of vertices needed to disconnect a graph [61]. The common direction taken in graph theory is to analyze the effect a small interdiction has on graph invariants [62], such as domination number [63] and independence number [64].

A key difference between graphs and networks is the information they represent. As defined in Section 2.1, graphs are ordered pairs of vertices and edges, whereas networks have additional information attributed to the vertices and edges. For instance, a vertex may represent a storage facility with a certain associated capacity and an edge may correspond to a road that has some maximum flow describing how much commodity may transit. Ahuja, Magnanti, and Orlin [65] provide an overview of the large body of work in operations research on these *network flow* problems.

The study of network flows gives rise to questions about their capacity, i.e., the maximum flow they can carry. One of the earliest systems studied is a network developed by Harris and Ross modeling the Soviet rail system (see Figure 2.9), published in a 1955 RAND report [66]. A key insight from this project is that the maximum flow through a network is equivalent to the minimum cut (i.e., bottleneck). Dantzig states “the [maximum flow-minimum cut] problem. . . arises naturally in the study of transportation networks [67]”; Alderson et al. [68] review this history while studying this Soviet rail system from a network interdiction perspective..



The network was formulated by Harris et al. and modified by Alderson et al. Edges cut as a result of the max flow-min cut algorithm added by the author.

Figure 2.9: Network Describing Flow of Commodities through Soviet Rail System. Adapted from [66], [68].

The next two sections summarize historic research in network reliability, robustness, and interdiction from perspectives of graph theory and network optimization.

2.4.1 Graph Theoretic Approach to Network Reliability and Interdiction

- Computing. Combines the languages and ideas of neurology, electrical engineering, and mathematics to develop "automata," that form the building blocks of a network. Each automaton is a vertex with several inputs and outputs that are the edges. The automaton receives signals from the input edges which are activated or not using the "Sheffer Stroke" Boolean function, also known as "NAND," which means at least one input is false. After some time delay, the automaton activates a subset of output edges. Von Neumann groups the automata into "organs" and considers the impact of failure of the automata on the overall system. Specifically, his analysis serves to inform the construction of redundant, multiplex systems systems that after accruing

ε in error at the level of basic automaton, achieve a total system system error less than $\delta \geq \varepsilon$ [57]. Pierce expanded upon Von Neumann's work to account for unequal weighting of inputs [69].

- Reliable electrical circuits. Moore and Shannon achieve similar system results to Von Neumann with more than two orders of magnitude less redundancy in circuitry using relays instead of the neurons and organs developed by Von Neumann [58]. In essence, these relays replace Von Neumann's automata nodes with relays comprised of switches in series and parallel that have a greater probability of remaining closed (transmitting signal), given similar error rates required by the analysis of the Sheffer Stroke organ.
- Graph Theory. Moskowitz generalized the work of Von Neumann and Moore and Shannon to a more abstract graph theoretical notion of redundancy that includes, but is not limited to computing and electronic circuitry [70].
- Communication Networks. When considering communication networks as *probabilistic* networks in which vertices and edges are subject to random removal (erasure), Jacobs discussed two criteria for reliability: probability that a path exists between all vertex pairs (denoted S) and the probability that a path exists between a pair of randomly selected vertices (denoted W). This work provides the graph density, defined as the number of edges divided by the number of vertices, that results in a specified probability S or W. Jacobs provides a closed form solution for small (less than 10-node) graphs and bounds for large (more than 50-node) graphs [59]. Wing and Demetriou calculate the probability of at least one path being closed between two random vertices using exhaustively using examination of connected subgraphs, disconnected subgraphs, and with estimates using Monte Carlo simulation [71]. Frank defines vulnerability criteria for several classes of communication networks with a large number of vertices [72] and Boesch thoroughly investigates bipartite and complete bipartite graphs to derive a class of "optimally damage-resistant network[s]" [73].
- Network Attack. Butler examines the connectivity of planar and nonplanar graphs after removal of vertices, edges, and combinations of both. Butler's definition of "strong" includes the number of vertices (or edges) required to separate the graph into multiple components [74]. Frank and Frisch discuss the implications of

building a network that can survive an enemy attack or natural disaster, including similar metrics to Jacobs, in addition to thresholds for the order of the largest connected subgraph and length of the longest remaining geodesic [75]. In addition to graph-level measures, Freeman introduced several metrics in a class of "centrality" measures that identify vertices that are important in some identified sense [36], [76]. These measures have been studied thoroughly in many contexts [37], [38], [77]–[82]. Many of the metrics analyzed from the graph theoretic perspective may be found in standard graph theory texts [11], [24], [83].

- Modern Work. With the recent advent of network science exploration that was popularized by the works of Watts and Strogatz, Barabási and Albert, and many others, examining network vulnerability from a graph theoretic perspective continued to be important. A major difference between modern and classical work is the presence of computing power. Whereas numerical means had often eschewed in favor of analytical, numerical methods now take the forefront in modern work due to the relative power and availability of computing resources. Bollobas et al. define network *robustness* as resistance to random damage and *vulnerability* as resistance to targeted attack [84]. Bollobas et al. define the *Linear Chord Diagram* model of graph formation as a mathematically rigorous model inspired by the Barabási-Albert preferential attachment model [44], of which Bollobas et al. are critical [84]. The teams of Bollobas et al. and Albert et al. found similar results: that networks with a power law degree distribution are robust against random node failure and vulnerable to targeted attacks, when compared to random graphs possessing a Poisson degree distribution [9], [84]. Estrada expanded this work [85].
- Network Deception. Matisziw et al. used a simulation approach that is different from both the graph theory approach and network optimization, but has heavy reliance on the contemporary availability of computer hardware and software availability [86]. Tong et al. developed a methodology to "immunize" a network from attack by choosing the most important vertices in the network and reducing their vulnerability to attack [87].

Though this is not an exhaustive survey of graph theory approaches to characterizing approaches to network robustness, vulnerability, attack, and defense, it serves as an indicator of the robust exploration in this area of research.

2.4.2 An Optimization Approach to Network Interdiction

Operations research practitioners often solve problems of network attack and defense using an optimization approach. For example, the analyst may be given a task to maximally disrupt the flow of a commodity across a given network. With the goal of selecting edges of certain capacities to interdict subject to a budget, he or she would formulate an objective function to minimize the maximum flow of the commodity across the network [88]. This work has been generalized in many ways, including the inclusion of full and partial interdiction of edges, interdicting edges with one or more resources, interdicting multiple commodities, and the usage of multiple sources (where the commodity begins) and sinks (where the commodity ends) [7]. The process of an attacker interdicting a network is considered an "attacker-defender" problem, in which the "defender" may only be a user and not take any actions to protect the network. This work has been extended to a "defender-attacker-defender" model in which the user of the network can take action to protect the network before an attack [8].

Attacker-Defender Models

Attacker-defender models are bilevel optimization problems (typically of the *max-min* form) in which one player (the "attacker") desires the disruption of a network operated by another (the "defender"). More specifically, the attacker seeks the network components (nodes or arcs) to target in order to maximize the minimum operating cost achievable by the defender. Attempts to understand the importance of individual arcs in supporting overall network flow are more than 50 years old [88]–[90]. The modern study of network interdiction problems was led by Wood et al. [7], [91]–[93], with applications to critical infrastructures and other systems led by Brown et al. [94]–[96].

Defender-Attacker-Defender Models

Similar to attacker-defender models, defender-attacker-defender models involve two players who engage in sequential play, now in three stages. More specifically, the defense first selects a strategy that will make the defended network maximally resistant to the impending attack. More precisely, in its *min-max-min* form, the defender seeks to minimize the maximum cost that the attacker can impose on the operation of the system.

Subsequent to the defense action, the problem is the same as the previously-introduced attack-defense problem [8], [97].

2.5 Statistics

Statistics is a discipline which enables assertions and conclusions to be made in the face of variation and uncertainty [98]. Both of these qualities are key to distinguishing complex networks from the graphs studied in graph theory. As stated in Section 2.3, complex networks might vary over time and often arise from data collected with a degree of uncertainty stemming from randomness or a lack of understanding the rules governing formation of the particular network. Devore provides the statistical foundation for this work in [98].

2.5.1 Correlation

Correlation measures the strength of the relationship between two variables, but does not indicate predictive power. For example, ice cream consumption and swimming pool usage in a small town may be highly correlated, but it would be strange to state that swimming pool usage causes ice cream consumption (or make the converse statement). We measure correlation by calculating the Pearson's correlation coefficient r as in Equation 2.8, where $\bar{x}$ and $\bar{y}$ are means of the random variables x and y, respectively.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \tag{2.8}$$

2.5.2 Regression

Regression measures the predictive power one or more independent variables has(have) in determining the value of a dependent variable. Such models have linear and nonlinear forms. In Chapter 4, we use single and multiple linear regression techniques to determine causality between measures.

The single linear regression model takes the form $y = \beta_0 + \beta_1 x + \varepsilon$, where β_0 and β_1 are the y-intercept and slope of the best-fit line through data points *(x,y)*. The quantity ε is a

random variable, assumed to be normally distributed with expected value 0 and variance σ^2.

The multiple linear regression model takes the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$, where β_0 is the y-intercept and $\beta_1, \beta_2, \ldots, \beta_k$ are coefficients that represent a 1-unit change in y when all other β_{k-1} variables are held fixed. The quantity ε is a random variable, with expected value 0 and variance σ^2. For the purposes of calculating confidence and prediction intervals, we assume ε is normally distributed.

We will now bring the graph theoretic, linear algebra, complex networks, and statistics ideas discussed in this chapter to bear. Graph theory, linear algebra, and complex networks will provide insight into graph structure and identification of those vertices which have greatest influence over graph topology. Linear algebra, complex network, and statistic tools will enable the evaluation of our introduced metric to gauge the level of unique information it contributes and how much that information changes with small changes in the graph.

CHAPTER 3:
The Neighbor Matrix

Parts of this chapter have been accepted to be published in the *Journal of Combinatorial Mathematics and Combinatorial Computing* [10].[2]

Matrices such as the adjacency, distance, reciprocal distance, walk, reachability, and Laplacian offer well-studied, compact structures that represent graph information [11], [22], [25], [27]–[29], [32], [83], [99]. As described in Chapter 2, two of these matrix representations that form the foundation of this work are the adjacency matrix and the distance matrix. The adjacency matrix enumerates the edges between vertices. The degree sequence is one partial summary of the adjacency matrix: a list of the adjacency matrix row sums. The distance matrix shows more information, namely the length of a shortest path between vertex pairs. We introduce and explore the *neighbor matrix*, a new matrix that extends the degree sequence to capture the distribution of each distance in a graph from 1 to k, where k is the graph diameter.

Figure 3.1 shows the topological richness of the neighbor matrix. On the left in the figure is a depiction of the degree sequence of a Barabási-Albert graph on 100 vertices. The right figure depicts the distances accounted for by the neighbor matrix. The $distance-1$ neighbors are exactly the degree sequence of the graph, depicted in black on the right. The illustration shows $distance-2$ neighbors in green, followed by $distance-3$ neighbors in tan, and $distance-4$ neighbors in brown. We depict the small number of $distance-5$ neighbors in the foreground of the picture, in grey.

In this chapter, we use combinatorial techniques to show the neighbor matrix to include many of the statistics and topological characteristics currently used to describe graphs and to provide algebraic and algorithmic techniques to calculate the neighbor matrix.

[2]This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States. JCMCC will claim and protect its copyright in international jurisdictions where permission from JCMCC must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

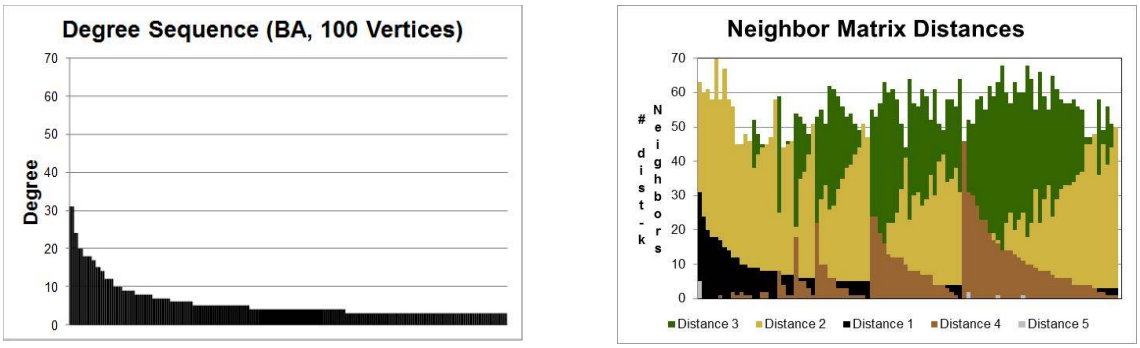


Figure 3.1: Comparing Topological Information of Neighbor Matrix to Degree Sequence

3.1 Defining the Neighbor Matrix

The following definitions are alternative (and more intuitive) terminology to the exact k-hop neighbor. The definitions will facilitate development of the neighbor matrix.

Definition 2. *Let G be a graph with $u, v \in V(G)$. Vertex u is a* distance-k neighbor *of v if and only if $d(u,v) = k$, where $k = diam(G)(1 \leq k \leq n-1)$.*

Definition 3. *Let G be a graph, with $v \in V(G)$. The* distance-k neighborhood *of v is*

$$N^{dist-k}(v) = \{u \in V(G) : d(u,v) = k\}.$$

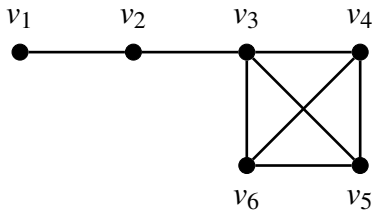
Definition 4. *Let G be a graph with $V(G) = \{v_1, v_2, \ldots, v_n\}$. The neighbor matrix,*

$$X^{dist-k}(G) = [x_{ij}], (1 \leq i \leq n, 1 \leq j \leq k),$$

where $n = |V(G)|$, $k = diam(G)$ $(1 \leq k \leq n-1)$ and $x_{ij} = |N^{dist-j}(i)|$, $i \in V(G)$; $1 \leq j \leq k$. We sort the rows of X^{dist-k} in a reverse lexicographic manner by organizing the rows of $X^{dist-k}(G)$ in non-increasing order of the entries in the first column.

The entries x_{ij} represent the count of dist-j neighbors of vertex i. This is the same number defined by Bloom et al [100], and later by Buckley and Harary [24] as the $(d_{ij})^{th}$ entry of the *distance degree sequence*. In addition, we define $X_i^{dist-k}(G)$ to be row i of the neighbor matrix associated with graph G, the row associated with $v_i \in V(G)$.

Figure 3.2 depicts a graph G and its corresponding neighbor matrix $X^{dist-3}(G)$.



$$X^{dist-3}(G) = \begin{array}{c} \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_2 \\ v_1 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 4 & 1 & 0 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \\ 2 & 3 & 0 \\ 1 & 1 & 3 \end{pmatrix} \end{array}$$

Figure 3.2: Graph G with Associated Neighbor Matrix

The neighbor matrix is a multi-dimensional analog of the degree sequence. As the degree sequence summarizes the adjacencies in the adjacency matrix, the neighbor matrix summarizes the distributions of each geodesic length found in the distance matrix. See Definition 4 and Figure 3.2.

We return to the Barabási-Albert graph of Figure 3.1 for another visualization of the Neighbor Matrix in Figure 3.3. The boldfaced rows of the Neighbor Matrix in the left of the Figure 3.3 are overlayed with a yellow rectangle in the right illustration. We consider a neighbor matrix row to be a *distance profile* for a vertex, describing in detail the number of steps or hops it take the given vertex to reach all the others in its component.

$$X^{dist-5}(G) = \begin{array}{c} \\ 3 \\ 4 \\ 2 \\ \mathbf{6} \\ \mathbf{1} \\ \mathbf{5} \\ 8 \\ \vdots \\ 13 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 31 & 63 & 5 & 0 & 0 \\ 24 & 60 & 15 & 0 & 0 \\ 20 & 61 & 18 & 0 & 0 \\ \mathbf{18} & \mathbf{70} & \mathbf{23} & \mathbf{0} & \mathbf{0} \\ \mathbf{18} & \mathbf{58} & \mathbf{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{17} & \mathbf{58} & \mathbf{23} & \mathbf{1} & \mathbf{0} \\ 15 & 67 & 17 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ 3 & 50 & 45 & 1 & 0 \end{pmatrix} \end{array}$$

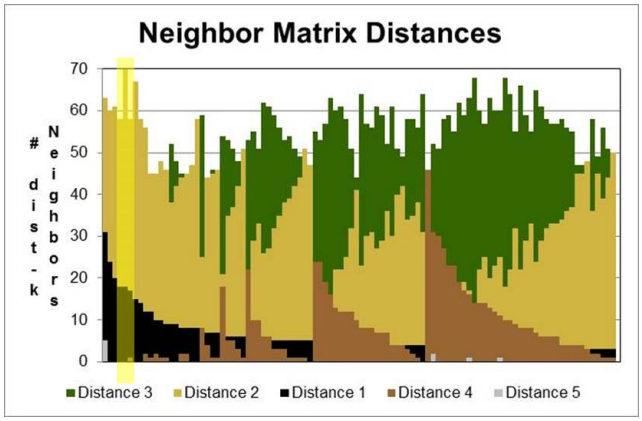


Figure 3.3: A Visualization of Neighbor Matrix Rows

Lemma 1. *There is a unique neighbor matrix $X^{dist-k}(G)$ for each graph G.*

Proof: Suppose otherwise that a given graph can have two different neighbor matrices. Let $K^1 = X_1^{dist-k}(G)$ and $K^2 = X_2^{dist-k}(G)$ such that $K^1 \neq K^2$. Then, there is an entry $k_{ij}^1 \neq k_{ij}^2$. By construction and ordering of the neighbor matrix, vertex i in G simultaneously has two different numbers of neighbors at distance j, which is not possible. ∎

Given Lemma 1, the following observation is immediate:

Proposition 1. *Let G and H be two graphs. If $G \cong H$, then $X^{dist-k}(G) = X^{dist-k}(H)$.*

Proof: By definition of isomorphism, there is a one-to-one correspondence ϕ from $V(G)$ to $V(H)$ such that $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$. Therefore, the distance in G from vertex i to vertex j is the same as the distance in H from $\phi(i)$ to $\phi(j)$, $\forall i, j \in G$. It follows directly by construction of the neighbor matrix that $X^{dist-k}(G) = X^{dist-k}(H)$. ∎

Remark 1. *The converse of Theorem* 1 *is false. That is,*

$$X^{dist-k}(G) = X^{dist-k}(H) \implies G \cong H.$$

Proof: We present an example depicting two infinite classes of non-isomorphic graphs with the same neighbor matrix below.

We define graphs G and H in accordance with Table 3.1. Vertices 1 to n define a copy of K_n. In Figure 3.4, $n = 8$.

Table 3.1: Non-isomorphic Graphs G and H with Identical Neighbor Matrices

	G	H
vertex $n+1$ adjacent to	$n, n-1, 3, 4$	$n, 2, 3, 4$
vertex $n+2$ adjacent to	$1, 2, 3, 4$	$1, 2, 3, 4$
vertex $n+3$ adjacent to	$n, n-1, n-2, n-3$	$n, n-1, n-2, n-3$
vertex $n+4$ adjacent to	$1, 2, n-2, n-3$	$1, n-1, n-2, n-3$

Figure 3.4 is an illustration of the graphs constructed in Table 3.1. The bold, red edges are those edges which are different between graphs G and H.

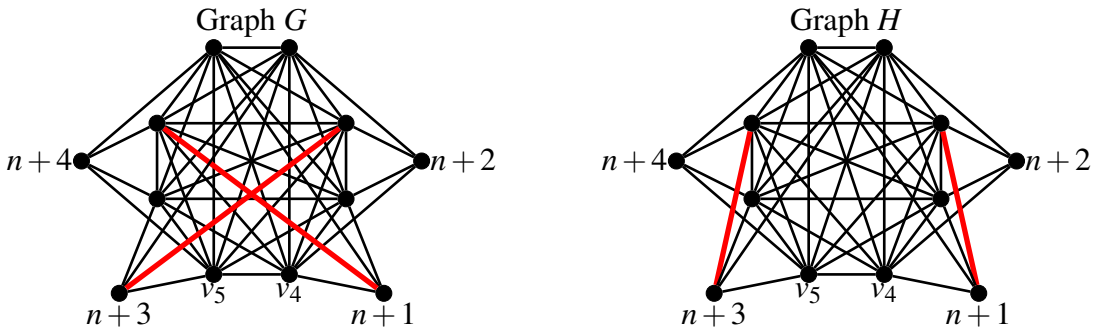


Figure 3.4: Graphs G and H

It is easy to show that $X^{dist-k}(G) = X^{dist-k}(H)$. In graph G, each degree-4 vertex has a maximum of two dist-1 neighbors shared with any other degree-4 vertex. In graph H, each degree-4 vertex has one vertex with which it shares three dist-1 neighbors. Therefore $G \ncong H$. ∎

It is clear that the implication stated in Proposition 1 holds for trees. Buckley and Harary [24] provide a counterexample to Remark 1 which may be extended to show that identical neighbor matrices do not imply isomorphism for trees. As algorithms exist to construct neighbor matrices, the contrapositive of Theorem 1 is a method to verify the non-isomorphism of two graphs.

The beauty of the newly introduced neighbor matrix extends from how naturally it captures the topology of an arbitrary graph, coupled with the simplicity of calculating matrix entries. A neighbor matrix of a graph captures each vertex's "view" of the graph through shortest paths, thus reaching every vertex of the graph, implicitly accounting for edges, cycles and other subgraphs of G (see Figure 3.5, two views of the Petersen graph).

3.2 Graph Invariants from the Neighbor Matrix

From the topological information encoded in the neighbor matrix, we may extract several graph invariants, as shown in Theorem 1.

Theorem 1. *For a simple graph G the following topological information can be observed from the neighbor matrix* $X^{dist-k}(G) = [x_{ij}], (1 \leq i \leq n, 1 \leq j \leq k)$*:*

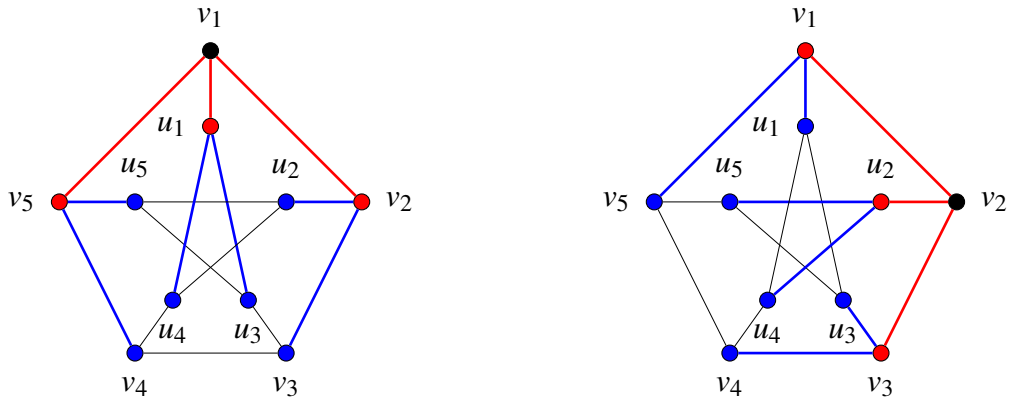


Figure 3.5: Two Views of Distance-1 and Distance-2 Neighbors (from v_1 and v_2) in the Petersen Graph

(a) *G is connected if and only if* $\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij} = n(n-1)$.

- *The number of components in G,* $\kappa(G)$*, is* $\sum_{d=1}^{n-1} \frac{|\mathcal{S}_d|}{d+1}$*, where*

$\mathcal{S}_d = \{v_i : \sum_{j=1}^{k} v_{ij} = d\}$.

(b) *The radius of the graph is* $Rad(G) = \max\{j : x_{ij} \neq 0, \forall i, 1 \leq i \leq n\}$.

(c) *The center of the graph is* $Cen(G) = G[\{v_i : x_{i\,\mathrm{rad}(G)} \neq 0, x_{ij} = 0, \forall j > rad(G)\}]$.

(d) *The closeness centrality of vertex i in a connected graph is*

$$CC_i = \frac{n-1}{\sum_{j=1}^{k} j \cdot x_{ij}}.$$

(e) *The average distance between vertices in a connected graph G is*

$$\frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{k} (j \cdot x_{ij}) = \frac{1}{n(n-1)} \sum_{j=1}^{k}\sum_{i=1}^{n} (j \cdot x_{ij}).$$

(f) *Given* $\mathrm{diam}(G) = k$ *(the number of columns in* $X^{dist-k}(G)$*),* $(1 \leq k \leq n-1)$*, the graph periphery is* $Per(G) = G[\{v_i : x_{ik} \neq 0 \ \ \forall i, \ \ 1 \leq i \leq n\}]$.

(g) *The first column is a representation of the degree sequence, in which the degree centrality of vertex i is the first entry in row i,* x_{i1}*. The number of edges in the graph is*

the half-sum of the entries in the first column:

$$m = \frac{1}{2} \cdot \sum_{i=1}^{n} x_{i1}.$$

- *The density of the graph (with m defined above) is defined as in Section 2.3.*

(h) For each column $j \in X^{dist-k}(G)$, $\sum_{i=1}^{n} x_{ij} = 2\left(|E(G^j)| - |E(G^{j-1})|\right)$.

(i) For each s $\left(1 \leq s \leq \text{diam}(G)\right)$, *the number of edges in the power graph* G^s *is given by*

$$\sum_{j=1}^{s} \sum_{i=1}^{n} x_{ij} = 2|E(G^s)|.$$

Proof:

(a) Suppose G is connected. Fix a vertex v, and note that $\forall u \in V(G-v)$, a geodesic exists such that $d(u,v) = j$ $(1 \leq j \leq k)$. There are $n-1$ choices for u in $G-v$. By construction of $X^{dist-k}(G)$, the length of a geodesic between all vertex pairs $u, v \in V(G)$ is counted. So, for each of the n rows of the matrix there are $n-1$ distance entries, totalling $n(n-1)$. For the converse, assume, to the contrary, that G is not connected and $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} = n(n-1)$. Then, there is a component G' containing q vertices $(1 \leq q \leq n-1)$; we choose $v \notin V(G')$. There are at most $n-1-q$ vertices adjacent to v. By construction of $X^{dist-k}(G)$ the entries of the row corresponding to vertex v sum to $n-1-q$, and also there are at most $n-2$ geodesics that are counted in the other rows of $X^{dist-k}(G)$. Therefore $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} \leq (n-1)(n-2) + (n-1-q) < n(n-1)$, which contradicts the initial statement that $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} = n(n-1)$. ∎

- We proved in (a) that a connected graph (i.e., graph with 1 component) has n neighbor matrix row sums of $n-1$. It follows that each connected component of order d will have d neighbor matrix row sums of $d-1$. Therefore, given $\mathcal{S}_d = \{v_i : \sum_{j=1}^{k} v_{ij} = d\}$, $\forall d : 1 \leq d \leq n-1$, the number of connected

components in G of order d is

$$\kappa(G)_d = \frac{|\mathcal{S}_d|}{d+1},$$

and, $\kappa(G) = \sum_{d=1}^{n-1} \kappa(G)_d$. ∎

By construction of X^{dist-k}, each value x_{ij} is the number of geodesics originating at vertex i of length $j\,(1 \leq j \leq k)$ with different terminal vertices. We weight the entries of the k-matrix row by multiplying each value x_{ij} by its associated distance, j. As seen in (a), the number of possible geodesics in G in a simple, connected graph is $n(n-1)$.

(d) After summing over row i, division by the number of geodesics provides the weighted average which is the average distance in G from vertex i to all other vertices. As stated in Section 2.3, the closeness centrality is the inverse of this average distance. ∎

(e) After summing over rows and columns, division by the number of geodesics provides the weighted average which is the average distance between vertices in G. ∎

(h) The left hand side of the equation double counts the sum of distance j neighbors $(1 \leq j \leq \text{diam}(G))$ from each vertex in G. The right hand side does the same by calculating the number of vertex pairs of distance at most j apart, and then it removes the number of vertex pairs of distance at most $j-1$. ∎

(i) The left hand side of the equation double counts the sum of distance $1, 2, \ldots, j$ neighbors $(1 \leq j \leq \text{diam}(G))$ from each vertex in G. The right hand side does the same by calculating the number of vertex pairs of distance at most j apart. ∎

3.3 Graph Orbits and the Neighbor Matrix

As stated in Subsection 2.1.1, vertices in the same orbit of a graph form an equivalence class. These equivalence classes are related to neighbor matrix rows, as stated in Proposition 2.

Proposition 2. *Given vertices v_i and v_j in graph G we have*

$$o_G(i) = o_G(j) \implies X_i^{dist-k}(G) = X_j^{dist-k}(G).$$

Proof: Suppose otherwise, that v_i and v_j are in the same orbit and $X_i^{dist-k}(G) \neq X_j^{dist-k}(G)$. Then, there is at least one distance value f for which v_i has more (or less) $distance - f$ neighbors than v_j. Therefore, there is no automorphism that maps v_i to v_j in $Aut(G)$. This is a contradiction, as v_i and v_j are in the same orbit, such a permutation of vertices must indeed exist in $Aut(G)$. ∎

Remark 2. *The converse of Proposition 2 is false. That is,*

$$X_i^{dist-k}(G) = X_j^{dist-k}(G) \not\implies o_G(i) = o_G(j).$$

Proof: The graph G in Figure 3.6 shows an example of a graph with a cycle. Note that vertices v_i and v_j have identical neighbor matrix rows. There are two different paths from vertex v_i to vertex v_k, a distance-2 neighbor. All paths from vertex v_j to distance-2 neighbors are unique; therefore there is no automorphism in $Aut(G1)$ that maps v_i to v_j and the two vertices are in different orbits.

Consider graph T in Figure 3.6. Note that vertices v_y and v_z have identical neighbor matrix rows. The subgraphs rooted at vertices v_y and v_z, respectively each have two distance-2 neighbors. In the subgraph rooted at v_y, the paths to the two distance-2 neighbors share a common vertex. In the subgraph rooted at v_z, the paths to the two distance-2 neighbors do not share a common vertex. Therefore, there is no automorphism in $Aut(T)$ that maps v_y to v_z and the two vertices are in different orbits. ∎

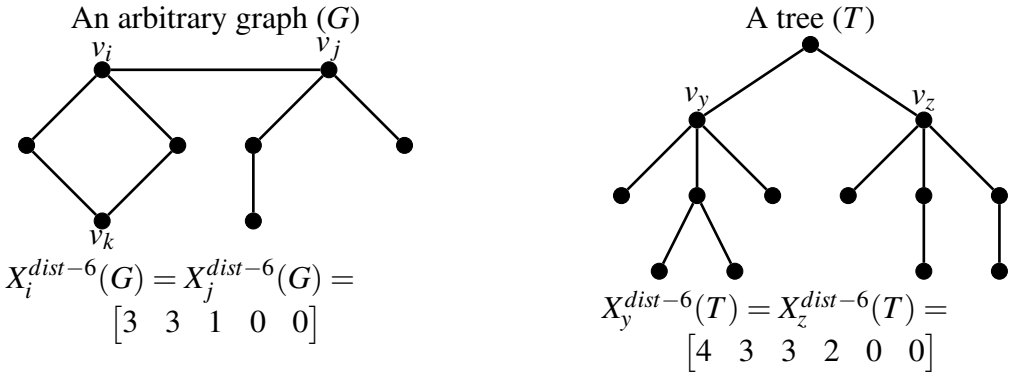


Figure 3.6: Graphs G and T, Counterexamples to Proposition 2

3.4 Finding the Neighbor Matrix

The neighbor matrix may be fully determined using powers of adjacency matrices, or any of a variety of computer algorithms that determine all-pairs shortest paths. Propositions 3 and 4 present two algebraic methods of obtaining a neighbor matrix representation of G. For these propositions we use X_A^{dist-k} to refer to the neighbor matrix obtained from an adjacency matrix A of G, and $\vec{1}$ to represent the $n \times 1$ vector whose entries are 1.

Proposition 3. *Each neighbor matrix can be obtained using adjacency matrices of powers of the original graph G, as shown in equation* 3.1. *Entries of* -1 *in* $A(G^i) - A(G^{i-1})$ *are replaced by* 0 *before multiplication by* $\vec{1}$*:*

$$X_A^{dist-k} = \left[A(G)\vec{1} \quad \left(A(G^2) - A(G)\right)\vec{1} \quad \ldots \quad \left(A(G^k) - A(G^{k-1})\right)\vec{1} \right]. \tag{3.1}$$

Proof: Notice that from the definition of G^k, the adjacency matrix of G^k, namely $A(G^k)$, includes entries of 1 for all pairs of vertices of distance k or less in the original graph G. Therefore, when we subtract $A(G^{k-1})$ from $A(G^k)$ and replace the -1 entries by 0, we are left with a $(0, 1)$-matrix with 1 entries representing pairs of vertices exactly distance k apart in G. Multiplication by $\vec{1}$ creates a vector consisting of the number of vertices at distance k from each vertex in G. ∎

We let $\mathcal{A}^j$ be the Boolean matrix obtained from $\mathcal{A}^j = (A^j)^b - \mathcal{A}^{j-1}$, where $j \geq 2$, by replacing the value of -1 and diagonal entries with zero; $\mathcal{A}^1 = A$, the adjacency matrix.

Proposition 4. *Each unsorted neighbor matrix, $X^{dist-k(u)}$ can be obtained using the boolean matrices $\mathcal{A}^j (1 \leq j \leq k)$, as shown in equation* 3.2

$$X_{\mathcal{A}}^{dist-k(u)}(G) = \left[\mathcal{A}^1\vec{1} \quad \mathcal{A}^2\vec{1} \quad \mathcal{A}^3\vec{1} \quad \ldots \quad \mathcal{A}^k\vec{1}\right]. \tag{3.2}$$

Proof: Notice that multiplying the adjacency matrix $\mathcal{A}^1$ of G by the column vector $\vec{1}$ we obtain the first column of the matrix $X^{dist-k}(G)$, the degree sequence. The second column of $X^{dist-k}(G)$ is $\mathcal{A}^2\vec{1} - \mathcal{A}\vec{1}$, as it counts all the vertices of distance 2 or less, and then it subtracts the vertices of distance 1, i.e. the neighbors of each fixed vertex in $V(G)$, as shown in Theorem 1. Similarly, $\mathcal{A}^k\vec{1} - \mathcal{A}^{k-1}\vec{1}$ counts the number of vertices k hops away, and it subtracts the number of vertices $k-1$ hops away. ∎

After a reverse lexicographic sort, $X^{dist-k(u)}$ becomes the neighbor matrix X^{dist-k}.

We next show that the neighbor-matrices X^{dist-k} obtained from different adjacency matrices of a graph G are related through the same permutation matrices as the adjacency matrices themselves.

Proposition 5. *Given A and B, two different adjacency matrices of graph G (so $B = P \cdot A \cdot P^T$, for some permutation matrix P),*

$$X_B^{dist-k}(G) = P \cdot X_A^{dist-k}(G) \cdot P^T.$$

Proof: Let A and B be two different adjacency matrices of G, so $B = P \cdot A \cdot P^T$ for some permutation matrix P. Since P is an orthogonal matrix, we have that $B^i = (PAP^T)^i = PA^iP^T$, for $1 \leq i \leq k$. Therefore $B^j - B^{j-1} = P(B^j - A^{j-1})P^T$ for each $j (1 \leq j \leq k)$, which implies that $X_B^{dist-k}(G) = P \cdot X_A^{dist-k}(G) \cdot P^T$. ∎

3.5 Conclusion

This chapter introduced and explored the neighbor matrix as an algebraic structure that contains significant graph descriptive and topological information. We proved this topological richness by proving the presence of 11 graph invariants in the neighbor matrix and relating the neighbor matrix to graph orbits. Though it does not inform isomorphism in arbitrary graphs, the neighbor matrix does provide a technique to verify the non-ismorphism of two graphs. The neighbor matrix has the potential to enable greater understanding of "graph space," as it is simultaneously more compact and richer in information than current structures used in graph exploration. For example, the degree sequence has long been used as a foundational element for the exploration of families of graphs. Since the neighbor matrix extends the degree sequence through all the distances that comprise the graph, it promises to enhance our current capability to model, analyze, and understand graphs in all distance dimensions.

CHAPTER 4:
Distance Centrality: Neighbor Matrix in Network Attack and Defense

Chapter 3 defined the neighbor matrix, which we showed to include significant topological information. We use the neighbor matrix in Chapter 4. This chapter is motivated by the desire to gain insight into which network elements (specifically vertices) are most significant in the maintenance of the network's internal structure. Though we consider this problem through the lens of the Department of Defense (DoD), we recognize implications and applications far beyond the DoD.

Suppose one organization decides to attack another. An effective planning tool might consider an adversary as a network. The attacker might strike the vertex or vertices in the network that result in the maximum internal structural change. A defense-minded organization would endeavor to solve the converse problem. It would consider its own network and "harden" those vertices most critical to its internal structure. The idea of considering a target network–be it your own or another –and focusing resources on critical elements extends naturally to the creation or maintenance of stability, network monitoring, and influence generation [7], [8], [10], [90]–[97], [101]–[103].

It has been documented that current metrics and techniques fall short of adequately quantifying network internal structure because many different graphs and networks have the same descriptive characteristics (e.g., degree sequence, clustering, average path length, etc.) [34], [35]. The goal of this research is to use a combination of established and innovative techniques to support the decision maker by better characterizing the structure of large, complex networks and pinpointing the vertices critical to the internal structure of such networks.

4.1 Limitations of Current Measures

A variety of measures currently exist to describe the qualities of a graph. The average distance between vertices is a measure of how "far apart" vertices in the graph are [28],

[29], [32]. The clustering coefficient describes graph transitivity, answering the question about the extent to which those vertices adjacent to a given vertex are themselves adjacent [32]. The Pearson Coefficient measures the extent to which vertices of similar degree are adjacent [32]. The s-metric takes the Pearson coefficient one step further and measures the extent to which high degree vertices (or, hubs) are adjacent [34], [35]. Many more descriptors exist. No individual metric or aggregation of metrics has proved to quantify graph internal structure well enough to inform decisions about how to create or mitigate change to graph topology. The metric and methodology in this work also does not fully characterize a graph or network, but it identifies vertices whose importance is under-reported by other measures.

The minimum requirement to inform decisions about graph structure is to identify when graphs are the "same" or "different." Current measures fail even this test. Table 4.1 describes 14 graphs that are clearly non-isomorphic. In Table 4.2, the entries in bold are the graph invariants previously discussed that have identical values when calculated from different graphs.

Table 4.1: Graphs Analyzed

Graph	Description	Graph	Description
$B_{7,7}$	2 x K_7 connected by P_7	$W_{1,20}$	wheel on 21 vertices
$K_{1,15}$	star on 16 vertices	P_{16}	path on 16 vertices
$W_{1,31}$	wheel on 32 vertices	$B_{10,12}$	2 x K_{10} connected by P_{12}
K_{32}	complete graph, 32 vertices	$K_{16,16}$	comp. bipartite, 32 vertices
CL_{32}	circular ladder, 32 vertices	C_{32}	cycle on 32 vertices
H_{32}	hypercube on 32 vertices	$L_{21,11}$	K_{21} connected to P_{11}
P_{32}	path on 32 vertices	$K_{1,31}$	star on 32 vertices

4.2 Sensitivity to Differences in Graph Structure

The neighbor matrix defined in Chapter 3 highlights difference in graph structure between many classes of graphs, even where previous measures do not. One measure of this is the neighbor matrix's norm. The norm of a matrix determines the "length" of the matrix. In this work, we use the Frobenius norm:

Table 4.2: Graph Comparison Metrics

Graph	AVG Distance	AVG Cluster Coeff	Pearson Correl Coeff	s-metric (Norm'd)
$B_{7,7}$	5.0	**0.640**	0.719	0.981
$W_{1,20}$	1.81	**0.640**	**-0.333**	0.323
$K_{1,15}$	**1.875**	**0**	-1	0.133
P_{16}	5.330	**0**	-0.077	0.981
$W_{1,31}$	**1.875**	0.648	**-0.333**	0.207
$B_{10,12}$	7.323	0.613	0.866	0.990
K_{32}	1.0	1.0	*	**1**
$K_{16,16}$	1.484	**0**	*	**1**
CL_{32}	4.645	**0**	*	**1**
C_{32}	8.258	**0**	*	**1**
H_{32}	2.581	**0**	*	**1**
$L_{21,11}$	4.105	0.653	0.942	0.998
P_{32}	11.0	**0**	-0.033	0.992
$K_{1,31}$	1.94	**0**	-1	0.064

*The Pearson correlation coefficient is undefined in regular graphs because the denominator of the calculation is a variance of zero.

$$||X^{dist-k}||_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{k} |x_{ij}|^2 \right)^{\frac{1}{2}}. \tag{4.1}$$

In Table 4.3, we see the neighbor matrix Frobenius norm is unique for each graph, verifying that the graphs are not isomorphic, even other classical measures are the same. As shown in Table 3.1 and Figure 3.4, nonisomorphic graphs can have the same neighbor matrix, therefore comparing neighbor matrix norms cannot confirm isomorphism between graphs. It can however determine structural difference between two graphs in examples that analysis of contemporary measures cannot.

4.3 The Distance Centrality

As described in Chapter 2, Subsection 2.3, the centrality of a vertex serves to provide insight into how *important* a given vertex is, from a certain perspective [32]. The following centrality measures are four of the most commonly used.

Table 4.3: Graph Comparison Metrics, With Neighbor Matrix Frobenius Norm

Graph	AVG Distance	AVG Cluster Coeff	Pearson Correl Coeff	s-metric (Norm'd)	X^{dist-k} Frobenius Norm
$B_{7,7}$	5.0	**0.640**	0.719	0.981	43.36
$W_{1,20}$	1.81	**0.640**	**-0.333**	0.323	79.75
$K_{1,15}$	**1.875**	**0**	-1	0.133	56.39
P_{16}	5.330	**0**	-0.077	0.981	17.55
$W_{1,31}$	**1.875**	0.648	**-0.333**	0.207	159.8
$B_{10,12}$	7.323	0.613	0.866	0.990	77.05
K_{32}	1.0	1.0	*	**1**	175.4
$K_{16,16}$	1.484	**0**	*	**1**	124.1
CL_{32}	4.645	**0**	*	**1**	60.66
C_{32}	8.258	**0**	*	**1**	44.18
H_{32}	2.581	**0**	*	**1**	89.62
$L_{21,11}$	4.105	0.653	0.942	0.998	116.0
P_{32}	11.0	**0**	-0.033	0.992	38.37
$K_{1,31}$	1.94	**0**	-1	0.064	170.0

*The Pearson correlation coefficient is undefined in regular graphs because the denominator of the calculation is a variance of zero.

- *Degree centrality*: the degree of vertex, may be normalized through division by $n-1$.
- *Eigenvector centrality*: calculated by an interative process that assigns value to a vertex based on the value of its neighbors.
- *Closeness centrality*: measures the distance from a vertex to all other vertices, calculated as the inverse mean distance.
- *Betweenness centrality*: the extent to which a given vertex lies on the shortest paths between vertex pairs.

We leverage the idea of the matrix norm computation representing a matrix's "length" to identify structurally-important vertices. We calculate the matrix of differences between the entries in the original neighbor matrix and the entries in the neighbor matrix after the given vertex is removed from the network. Taking this difference requires some matrix manipulation to ensure proper dimensionality and that entries in the vertex removal neighbor matrix correspond to the entry in the original neighbor matrix, in all cases. First,

we conduct a row swap, bringing the row representing the vertex to be removed to the n^{th} row in the matrix. After vertex removal, the neighbor matrix has $n-1$ rows, so we add a row of zeros to this neighbor matrix to ensure both matrices have the same number of rows. In the instance that the vertex removal reduces the diameter of the graph (i.e., it is a cut vertex), we add columns of zeros to the vertex removal neighbor matrix so it has the same number of columns as the original neighbor matrix. In the case where removing the vertex causes the diameter to increase, we add columns of zeros to the original neighbor matrix so it has the same number of columns as the vertex removal neighbor matrix. Now with the difference matrix calculated, we determine the matrix norm. Finally, we normalize this quantity to a value between -1 and 1 through division by the norm of the original neighbor matrix.

As this measure results in a determination of which vertices whose removal have the greatest (and least) change in a matrix's distribution of distances, it is a centrality measure, like those previously listed. We call this centrality measure **Distance Centrality**, defined in Equation 4.2.

$$DC_i = \frac{||X^{dist-k}(G) - X^{dist-k}(G-v_i)||_F}{||X^{dist-k}(G)||_F} \in (0,1) \tag{4.2}$$

Values of DC_i are close to 1 for vertices with high topological significance and close to 0 for vertices of relatively limited topological significance.

4.4 Quantifying Vertex Influence with Distance Centrality

We have shown the neighbor matrix to possess significant topological information in Chapter 3 and used the neighbor matrix to derive the distance centrality, leveraging topological richness of the neighbor matrix toward identifying which vertices have the most influence over a graph's distribution of distances or "distance profiles." In this section, we apply the distance centrality using a network "attack" schema. Our goal is to pinpoint the vertices that exert the greatest influence over a graph's topology.

We calculate the Distance Centrality of a given vertex i, DC_i, by taking the following steps.

1. Calculate the neighbor matrix of the original graph.
2. Remove each vertex (with replacement).
3. Calculate the neighbor matrix after vertex removal.
4. Take the difference between the original neighbor matrix and the neighbor matrix after vertex removal.
5. Calculate the norm of the difference matrix.
6. Divide by the norm of the original neighbor matrix.

Because distances in graphs are determined by shortest paths, we see the Distance Centrality as a mechanism for determining which vertices exert the most (and least) influence over the collection of "shortcuts" in the graph. Analysis of the graphs in Section 2.5 demonstrate the Distance Centrality's utility in identifying vertices of topological significance that are overlooked or under-reported by other centrality measures.

We depict the graphs in the section using Gephi Graph Visualization and Manipulation software, version 0.9.1. The vertices in the graphs are sized by their betweenness centrality (i.e., larger betweenness value, larger-sized vertex) and colored by modularity classes, or communities. After using the "Force Atlas 2" layout, the author arranged the vertices manually.

Erdős-Rényi Random Graph

Figure 4.1 depicts an Erdős-Rényi Random Graph with 90 vertices and 140 edges. The vertex labels are that vertex's distance centrality ranking; v_1 has the highest distance centrality. Table 4.4 shows the ten vertices with the highest distance centralities with the rankings of those vertices in betweenness, closeness, degree and eigenvector centralities.

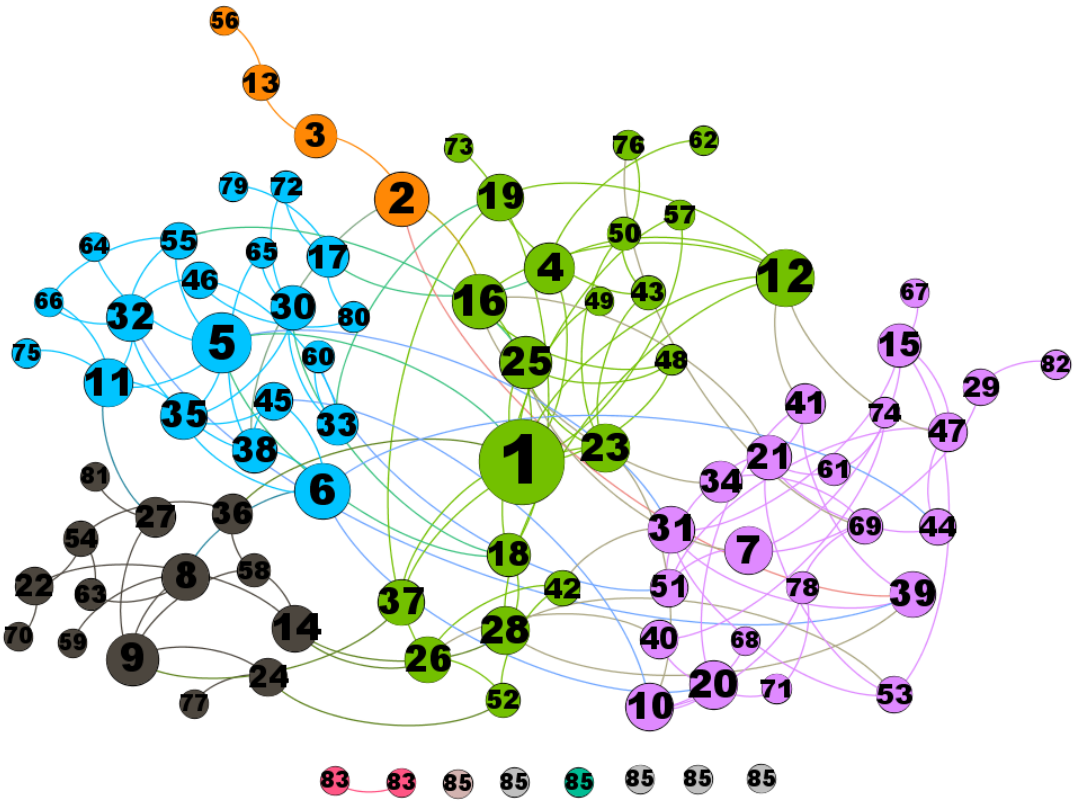


Figure 4.1: An Erdős-Rényi Random Graph on 90 Vertices and 140 Edges

Vertex 2–the second row of Table 4.4–is ranked 6, 16, 20, and 23 in betweenness, closeness, eigenvector , and degree centralities, respectively. Vertex 2 is a cut vertex; removing it causes an increase in number of components from 8 to 9 and a reduction in diameter from 9 to 8. Removal of vertex 3 in Table 4.4 causes a change in the graph similar to that caused by removing vertex 2, but with the creation of a smaller disconnected component. Figure 4.2 illustrates the adjacencies of vertex 2, showing that vertex 2 has neighbors in four of the five largest communities in the graph and that its removal further disconnects the graph.

Table 4.4: Centrality Rankings for an Erdős-Rényi Random Graph on 90 Vertices and 140 Edges

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	1	1	1
2	6	16	20	23
3	28	62	63	56
4	9	16	13	23
5	2	3	4	2
6	4	8	18	4
7	14	42	35	23
8	18	39	50	23
9	7	36	40	9
10	15	34	38	9

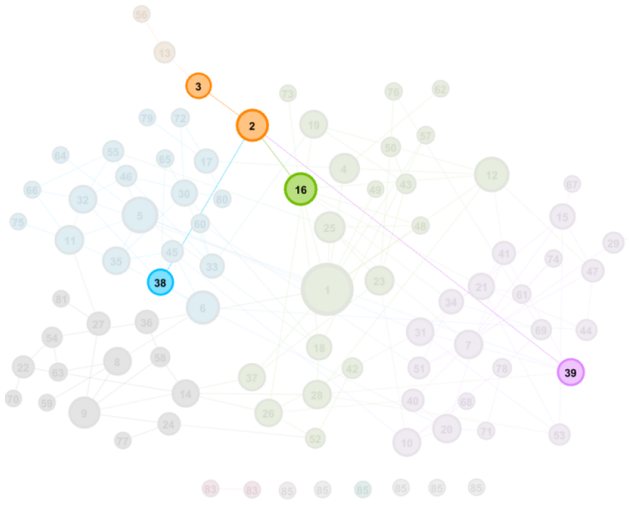


Figure 4.2: Highlighting the Inter-community Adjacencies of Vertex 2

Watts-Strogatz Small World Model

Figure 4.3 depicts a graph formed using the Watts-Strogatz small world graph formation model. Once again, the vertex label for each vertex is the distance centrality ranking of the respective vertex; v_1 has the highest distance centrality. Table 4.5 shows the ten vertices with the highest distance centralities with the rankings of those vertices in betweenness, closeness, degree, and eigenvector centralities.

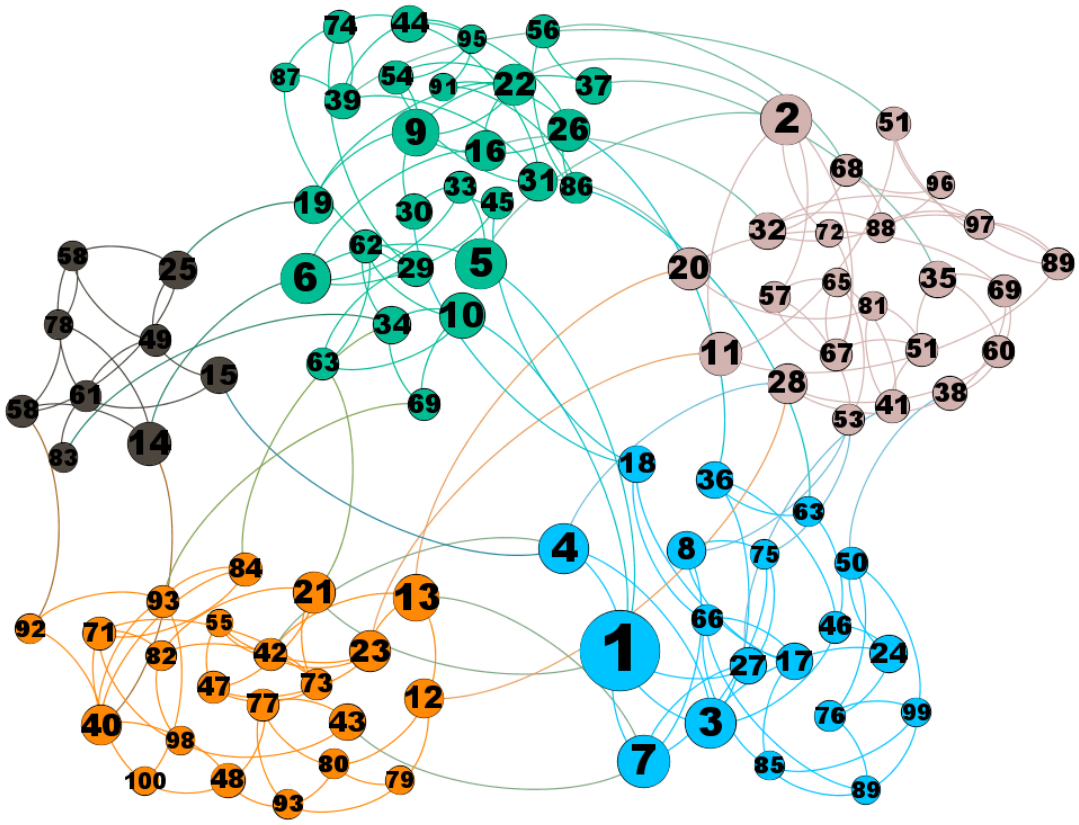


Figure 4.3: A Watts-Strogatz Small World Graph on 100 Vertices and 200 Edges

Table 4.5: Centrality Rankings for a Watts Strogatz Small World Graph on 100 Vertices

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	1	1	1
2	3	6	2	10
3	5	5	5	5
4	7	4	5	7
5	4	2	22	11
6	6	8	5	37
7	2	3	2	2
8	23	25	22	13
9	8	14	2	31
10	10	12	5	6

Vertex 8 is ranked 23, 25, 22, and 13 in betweenness, closeness, degree, and eigenvector centralities, respectively. Figure 4.4 illustrates the adjacencies of vertex 8. Vertex 8 looks rather unimpressive in Figure 4.4. The topological importance of this vertex is hidden in the complexity of the graph. In reality, the combination of unique connections between vertex 8 and its neighbors causes the removal of this vertex to create the 8^{th} most change in the distances between the vertices of this graph.

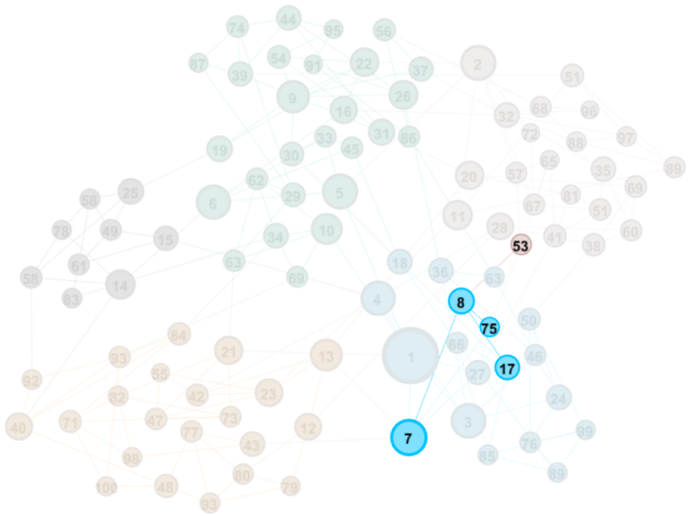


Figure 4.4: Highlighting the Adjacencies of Vertex 8

Barabási-Albert Preferential Attachment Model

Figure 4.5 depicts a graph formed using the Barabási-Albert preferential attachment model. The vertex labels are that vertex's distance centrality ranking; v_1 has the highest distance centrality. Table 4.6 shows the ten vertices with the highest distance centralities with the rankings of those vertices in betweenness, closeness, degree and eigenvector centralities.

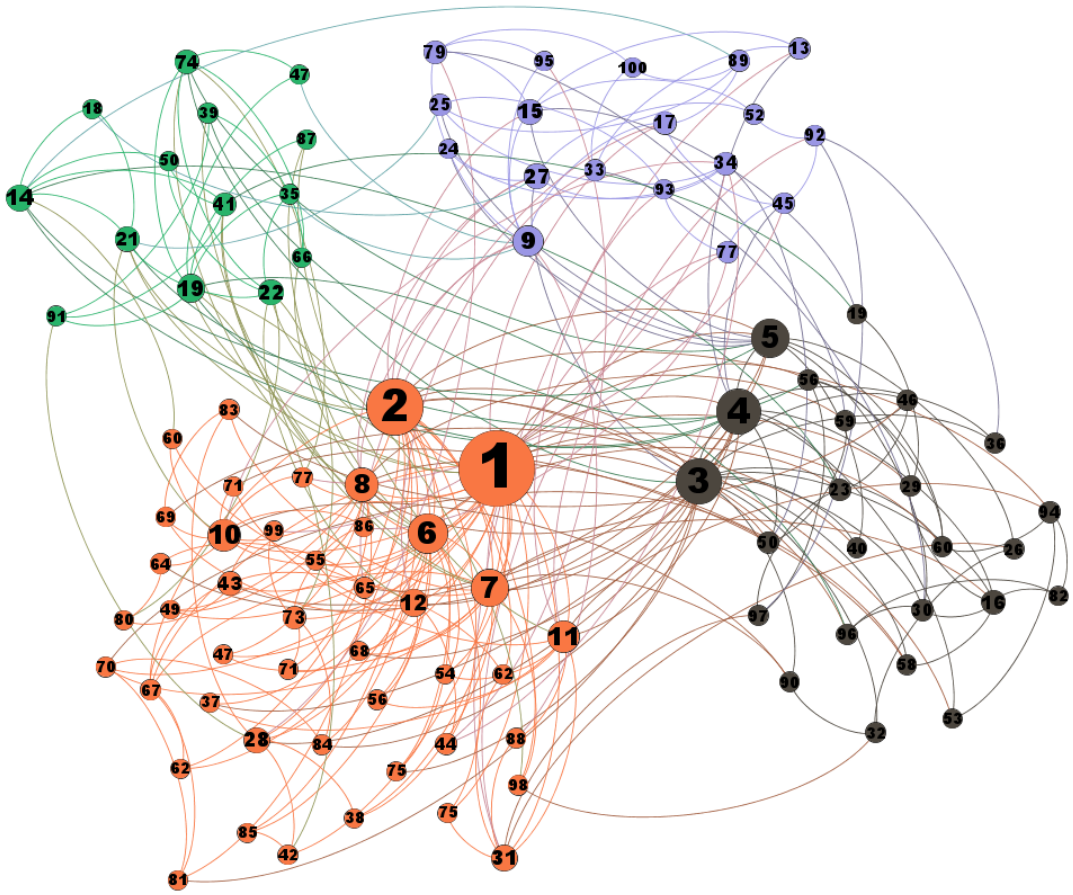


Figure 4.5: Barabási-Albert Graph on 100 Vertices and 291 Edges

Table 4.6: Centrality Rankings for a Barabási-Albert Graph on 100 Vertices

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	1	1	1
2	2	2	2	2
3	3	4	3	3
4	4	3	5	4
5	6	5	7	7
6	5	6	4	4
7	7	7	6	6
8	8	8	8	8
9	11	12	15	11
10	9	9	9	9

Vertex 9 is ranked 11, 12, 15, and 11 in other betweenness, closeness, eigenvector, and degree centralities, respectively. Figure 4.6 illustrates the adjacencies of vertex 9, showing that vertex 9 has neighbors in each of the graph's four communities. Vertex 9 is a "shortcut" from each community to all the other communities.

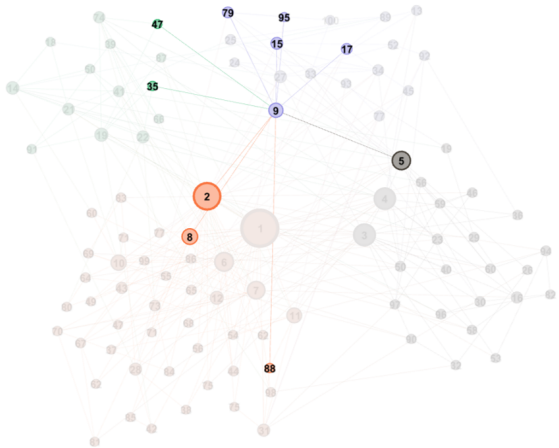


Figure 4.6: Highlighting the Inter-community Adjacencies of Vertex 9

Karate Club

Figure 4.7 depicts the relationships between the members of a college karate club in the late 1970s [51].

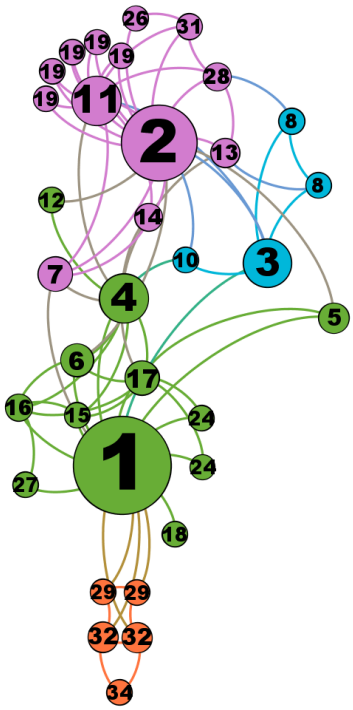


Figure 4.7: Zachary's Karate Club

Table 4.7 shows the top ten vertices in distance centrality, with the corresponding rankings in betweenness, closeness, eigenvector, and degree centralities.

Figure 4.8 illustrates the adjacencies of vertex 5, showing that vertex 5 has adjacencies directly linking it the two largest communities in the graph. Vertex 5 is an example of a vertex that is under-emphasized by other centrality measures. It is a "shortcut" from large community at the bottom of the graph to the large community at the top of the graph. It is ranked #5 in distance centrality, but 9, 8, 13, 17 in betweenness, closeness, eigenvector, and degree centrality, respectively.

Table 4.7: Karate Club Centrality Rankings

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	1	2	2
2	2	3	1	1
3	5	4	9	6
4	4	2	3	4
5	9	8	13	17
6	8	5	7	8
7	6	5	6	8
8	16	22	31	17
9	18	22	32	17
10	19	13	16	17

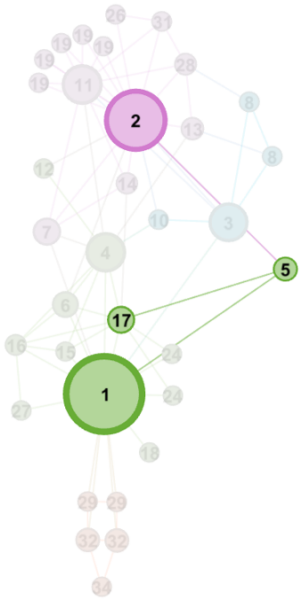


Figure 4.8: Highlighting the Adjacencies of Vertex 5

Noordin Top Terrorist Network

Figure 4.9 depicts the relationships between members an international terrorist group led by Noordin Mohammed Top [52].

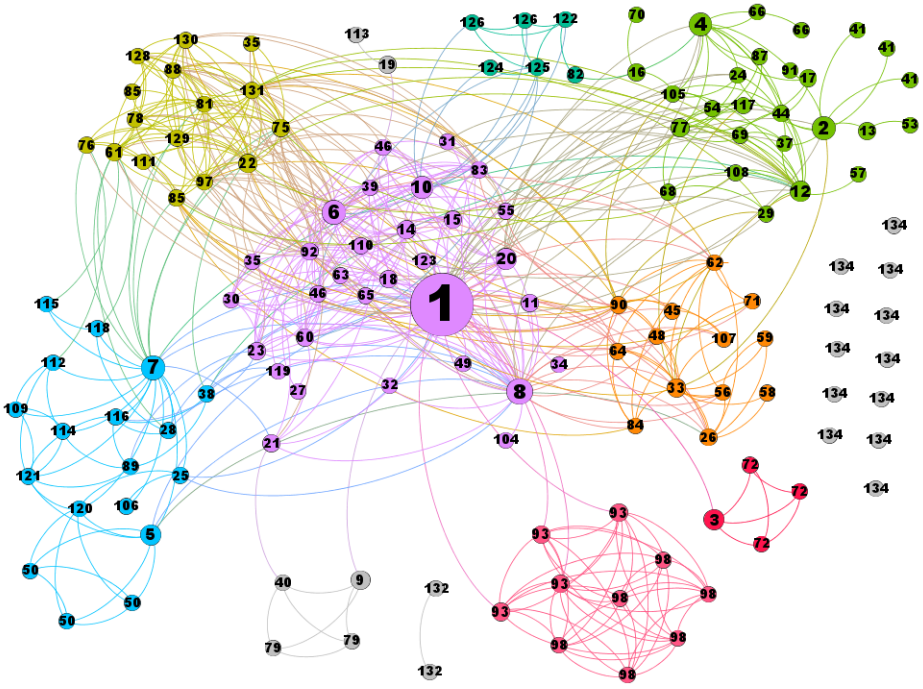


Figure 4.9: Noordin Top Terror Network

Table 4.8 depicts the top ten vertices in various centrality measures. As this graph contains isolated vertices, we do not compute eigenvector centrality.

Figure 4.10 illustrates the adjacencies of vertex 2. This vertex is a cut vertex, the removal of which results in the greatest number of isolated vertices of any vertex in the network. It is the second-highest ranked vertex in distance centrality and #5, #25, #53, #34 in betweenness, closeness, eigenvector, and degree centrality, respectively. Vertex 3 is depicted in Figure 4.11. The removal of this vertex results in a triangle being isolated from the graph. This vertex the third-highest ranked vertex in distance centrality and #11, #64, #78, #78 in betweenness, closeness, eigenvector, and degree centrality, respectively.

Table 4.8: Centrality Rankings for Noordin Top Terror Network

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	1	1	1
2	5	25	53	34
3	11	64	78	78
4	6	19	42	23
5	9	29	52	41
6	4	3	2	3
7	3	10	29	6
8	2	2	3	2
9	12	64	77	78
10	7	43	26	14

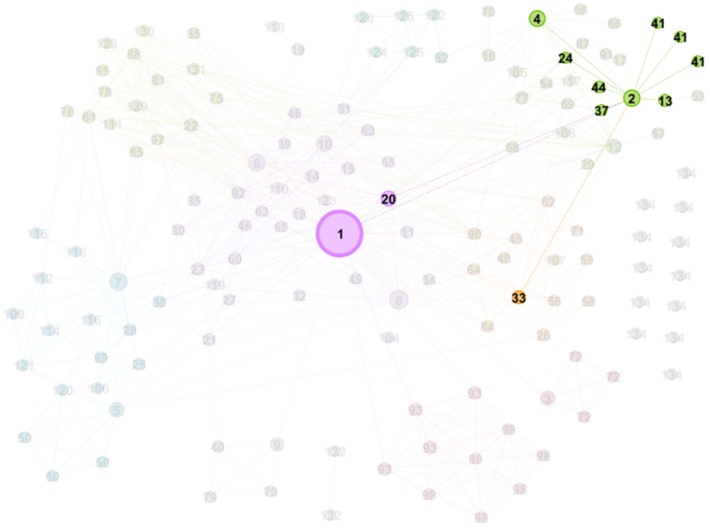


Figure 4.10: Highlighting Vertex 2 of the Noordin Network

Roginski LinkedIn Professional Network

Figure 4.12 depicts a subset of the author's professional relationships, as captured by scraping LinkedIn data [53].

Table 4.9 depicts the top ten vertices in distance centrality, with respective betweenness, closeness, eigenvector, and degree centrality measures included.

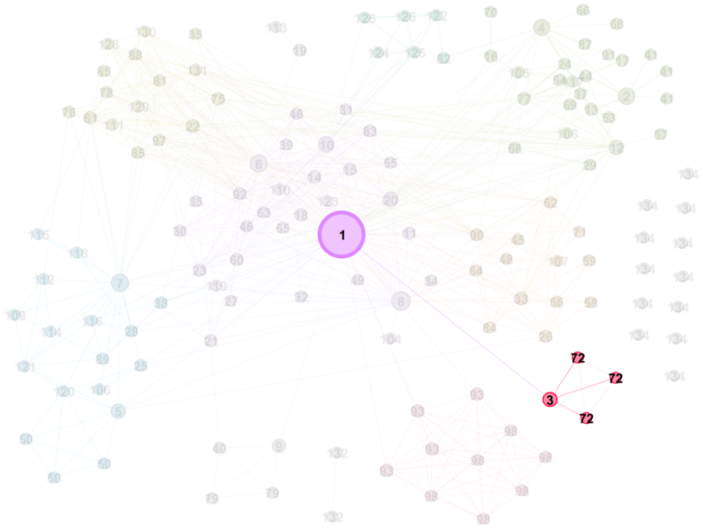


Figure 4.11: Highlighting Vertex 3 of the Noordin Network

Table 4.9: Roginski LinkedIn Network

DC	BC	CC	Deg	EVC
Entries are Vertex Labels i is i^{th}-Ranked DC				
1	1	261	323	326
2	2	15	114	108
3	3	396	412	294
4	7	419	428	301
5	5	7	67	38
6	55	235	296	248
7	23	98	173	213
8	10	31	158	108
9	15	30	55	56
10	29	399	413	338

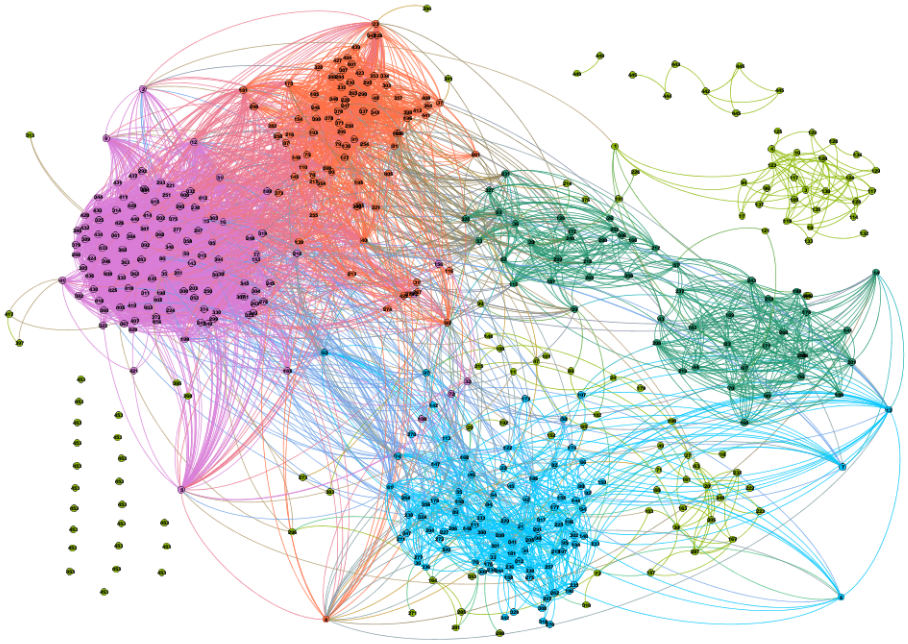


Figure 4.12: Roginski LinkedIn Network

Figure 4.13 illustrates the adjacencies of vertex 1. This vertex is a cut vertex, the removal of which results in the diameter of the graph decreasing from 9 to 7 as this vertex is the only bridge gateway to the circled community of 27 vertices.

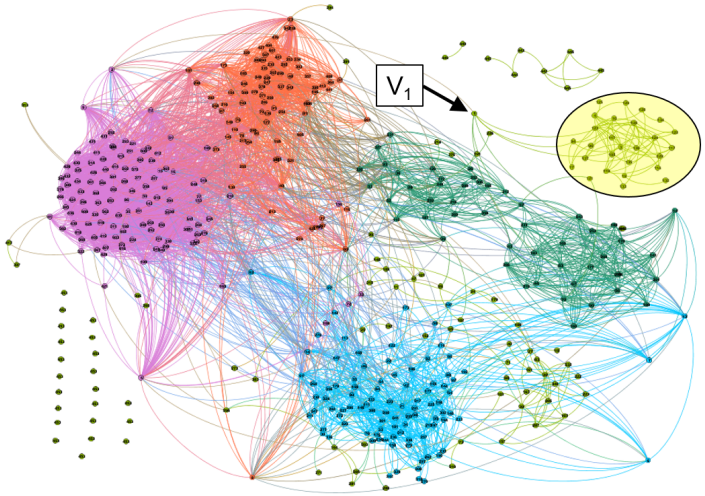


Figure 4.13: Highlighting Vertex 1 of the LinkedIn Network

This section presented the utility of the distance centrality in identifying vertices which influence distances between vertices in graphs. Even in graphs that are small enough to visualize and small enough to analyze with local resources, graph complexity quickly overwhelms the human ability to spot vertices important to graph distance. This problem is compounded when the graphs are too large to visualize and too large to exhaustively analyze with local resources.

4.5 The Distance Centrality: Not Just Another Centrality Measure

A multitude of centrality measures exist to provide insight into many different problems. In this section, we use correlation and regression to show how distance centrality provides a useful addition to the body of centrality knowledge that is not wholly contained in the measures already established.

We use the following graphs to explore the correlation and predictive relationships between existing centrality measures and the distance centrality. These graphs include synthetic graphs, developed algorithmically as described in Subsection 2.3.1. In addition, we include graphs that represent social networks which developed naturally as defined in the associated references.

- Synthetic Graphs
  - Erdős-Rényi Random Graph: 90 vertices and 100 edges
  - Barabási–Albert Preferential Attachment Graph: 100 vertices and 291 edges
  - Watts-Strogatz Small World Graph: 100 vertices and 200 edges
- Social Networks
  - Karate Club: represents a karate club at a university in the 1970*s*, 34 vertices and 78 edges [104].
  - Noordin Top Terrorist Network: from the 2006 International Crisis Group report titled *Terrorism in Indonesia: Noordin's Networks*, 145 vertices and 560 edges [52].
  - LinkedIn Network: a professional social network with 474 vertices and $5,263$ edges [105].

4.5.1 Correlation Analysis

As defined in Subsection 2.5.1, correlation measures the extent to which large values of one variable correspond with large values of another variable.

See Tables 4.10 and 4.11 for a depiction the correlations between betweenness, closeness, degree, eigenvector, and distance centralities for each of the graphs in 4.5. As previously noted by Valente et.al. in [80], these centrality values are correlated at different levels.

Table 4.10: Centrality Measure Correlation for Synthetic Graphs

Synthetic Graphs

	Erdős-Rényi						Watts Strogatz				
	DC	BC	CC	Deg	EVC		DC	BC	CC	Deg	EVC
DC	1	0.75	0.86	0.75	0.58		1	0.97	0.82	0.59	0.67
BC	0.75	1	0.61	0.91	0.81		0.97	1	0.83	0.68	0.68
CC	0.86	0.61	1	0.77	0.75		0.82	0.83	1	0.57	0.79
Deg	0.75	0.91	0.77	1	0.85		0.59	0.68	0.57	1	0.64
EVC	0.58	0.81	0.75	0.85	1		0.67	0.68	0.79	0.64	1

Barabási–Albert

	DC	BC	CC	Deg	EVC
DC	1	0.97	0.80	0.91	0.89
BC	0.97	1	0.84	0.98	0.92
CC	0.80	0.84	1	0.88	0.96
Deg	0.91	0.98	0.88	1	0.94
EVC	0.89	0.92	0.96	0.94	1

Table 4.11: Centrality Measure Correlation for Social Networks

Social Networks

	Karate Club						Noordin Top				
	DC	BC	CC	Deg	EVC		DC	BC	CC	Deg	EVC
DC	1	0.94	0.74	0.80	0.73		1	0.78	0.72	0.62	0.44
BC	0.94	1	0.72	0.92	0.80		0.78	1	0.34	0.72	0.49
CC	0.74	0.72	1	0.77	0.91		0.72	0.34	1	0.63	0.62
Deg	0.80	0.92	0.77	1	0.80		0.62	0.72	0.63	1	0.91
EVC	0.73	0.80	0.91	0.80	1		0.44	0.49	0.62	0.91	1

LinkedIn®

	DC	BC	CC	Deg	EVC
DC	1	0.76	0.48	0.13	0.04
BC	0.76	1	0.26	0.30	0.17
CC	0.48	0.26	1	0.63	0.50
Deg	0.13	0.30	0.63	1	0.92
EVC	0.04	0.17	0.50	0.92	1

Using the "rules of thumb" provided in Devore's text (see Table 4.12), we describe these correlations in Tables 4.10 and 4.11, by considering the averages of the correlation values these tables [98]. We describe the correlation between the distance centrality and all other centrality measures considered to be either "strong" or "moderate." Further exploration of these relationships is warranted. We begin by using betweenness, closeness, degree, and eigenvector centralities to predict distance centrality using multiple linear regression in Subsection 4.5.2.

Table 4.12: Devore's "Rules of Thumb" for Describing Correlations

	Lower Bound	Upper Bound
Weak	0	0.5
Moderate	0.5	0.8
Strong	0.8	1

Table 4.13: Describing Correlation Between Centrality Measures

Synthetic Graphs					
	DC	BC	CC	Deg	EVC
DC		strong	strong	moderate	moderate
BC	strong		moderate	strong	strong
CC	strong	moderate		moderate	strong
Deg	moderate	strong	moderate		strong
EVC	moderate	strong	strong	strong	
Social Networks					
DC		strong	moderate	moderate	weak
BC	strong		weak	moderate	weak
CC	moderate	weak		moderate	moderate
Deg	moderate	moderate	moderate		strong
EVC	weak	weak	moderate	strong	

4.5.2 A Regression Approach to Predicting Distance Centrality

The correlation analysis in Subsection 4.5.1 indicated the potential for a causal relationship between the distance centrality and four existing centrality measures considered. We use the technique of multiple linear regression to confirm these relationships and demonstrate that the distance centrality includes information which may be thought of as a combination of the information found in the other centrality measures.

Table 4.14 shows the extent to which each of betweenness, closeness, degree, and eigenvector centrality explain the variation in the distance centrality data associated with the graphs introduced in Section 4.5. The table depicts r^2 values that range from 0.19 to 0.96, with an average value of 0.67 which means that an average of 67% of the variability of the distance centrality data associated with the six networks we are analyzing is explained by single effects.

Table 4.14: Coefficients of Determination for Single Effect Linear Regression Models Predicting Distance Centrality

r^2 for Synthetic Graphs

	BC	CC	Deg	EVC
Erdős-Rényi	0.56	0.75	0.56	0.34
Watts-Strogatz	0.91	0.67	0.35	0.45
Barabási–Albert	0.94	0.64	0.83	0.79

r^2 for Social Networks

Karate Club	0.89	0.55	0.64	0.54
Noordin Top	0.61	0.51	0.38	0.19
LinkedIn	0.58	0.22	0.02	0.00

Naturally, we add predictors to the model in an effort to increase the model's predictive power. Table 4.15 contains the results of running stepwise multiple regression models using the JMP statistical package [106], version 12. Distance centrality is the response variable, with predictors being betweenness, closeness, degree, and eigenvector centrality. With the inclusion of additional regressors to the model, we expect the coefficient of determination's value to increase and indeed it does, $2-58\%$ over the single effect models in Table 4.14. However, the maximum variance inflation factor (VIF) indicates multicollinearity between the regression variables which must be remedied for the model's predictive power not to be compromised. There are several means by which multicollinearity may be addressed, but the most appropriate mechanism for the data we have is variable removal.

Variable removal is simply re-running the model without a variable which is a statistically significant regressor. Table 4.16 shows the results of running the same multiple regression procedures as previously, removing the noted variable for those graphs which have a

Table 4.15: Regression Model Results Using Four Centralities to Predict Distance Centrality

Synthetic Graphs

	Erdős-Rényi	Watts-Strogatz	Barabási–Albert
Variables Included	BC, CC, Deg, EVC	BC, Deg	BC, CC, Deg, EVC
r^2	0.92	0.92	0.98
Maximum VIF	8.24	2.46	38.0

Social Networks

	Karate Club	Noordin Top	LinkedIn
Variables Included	BC, CC, Deg	BC, CC, Deg, EVC	All
r^2	0.98	0.96	0.84
Maximum VIF	15.6	18.5	5.91

maximum VIF greater than 10. The results in Table 4.16 show that across a wide range of networks the distance centrality contains information which is not wholly contained in the other commonly used centrality measures, granting insight into why the distance centrality identifies difference vertices as important than the other centrality measures explored.

Table 4.16: Final Regression Results

Synthetic Graphs

	Erdős-Rényi	Watts-Strogatz	Barabási–Albert
Variable Removed	None	CC, EVC	Closeness
Variables Included	BC, CC, Deg, EVC	BC, Deg	BC, Deg
r^2	0.92	0.92	0.96
Maximum VIF	8.24	2.46	6.67

Social Networks

	Karate Club	Noordin Top	LinkedIn
Variable Removed	Betweenness	Betweenness	None
Variables Included	CC, Deg, EVC	CC, Deg, EVC	All
r^2	0.97	0.92	0.84
Maximum VIF	8.21	8.15	5.91

4.6 Conclusion

The motivation for this chapter is attack and defense of networks by identifying those vertices with high influence over the structure of a network. These vertices should be interdicted or attacked if the goal is to cause structural change or degradation in the network. Conversely, those vertices should be hardened or protected if the goal is safeguard against structural change.

We contributed to the body of graph comparison knowledge by introducing the distance centrality as a mechanism to compare graphs by analyzing the changes to graph distance profiles resulting from removing vertices (with replacement) from the graph. This technique and the newly-introduced distance centrality provides a measure shown through correlation and regression analysis to include topological information found in four commonly-used centralities: betweenness, closeness, degree, and eigvector centrality. The distance centrality may be used to indicates if an attack on a network would indeed cause change and if so, where the strike should occur. Consider the Noordin Top Terrorist network. As shown in Figures 4.10 and 4.11, the most effective attacks would cause very little change, isolating individual vertices or a triangle. This is a situation which distance centrality analysis indicates that a single strike would likely be ineffective. Another option should be pursued to generate effects on the network.

In those networks wherein an attack may be successful in generating structural change, analysis using the distance centrality points to where the attack should occur (or be guarded against), even when those results are counter intuitive. Figure 4.8 shows where a single strike on a vertex of low degree would generate the 4^{th} highest structural change. Analysis of the Facebook graph informs us to stay away from the vertex that intuition (and other centralities) guides us to, but rather interdict another. Insight into whether an attack strategy should be pursued and if so, where, is an important contribution to decision support in a world of complex networks which intuition routinely fails on its own and must be buttressed through analytical means.

CHAPTER 5:
Robustness of the Distance Centrality

We defined the Neighbor Matrix in Chapter 3 as a mathematical structure that contains significant graph topological information in a compact structure. We then used the Neighbor Matrix to define the Distance Centrality in Chapter 4, a measure that identifies vertex importance through the examination of vertex percolation effect on remaining geodesics in the graph. Next, we will explore how much this distance changes under small perturbations in the studied networks. The motivation behind this exploration is that we live, and the Department of Defense operates, in an uncertain world. Data collected may be incomplete and/or have error. Indeed, a reason for the discipline of statistics is to provide interval solutions based upon data sampled from a population that may not be studied exhaustively. In this chapter, we simulate error in data collection through the use of network simulation to further explore the distance centrality by identifying the measure's "robustness;" that is, answering the question: "how much does distance centrality change when compared to the change in betweenness, closeness, and eigenvector centrality, when perturbations are made to the original network?"

5.1 Survey of Network Robustness and Reliability

The subject of network robustness has received considerable scrutiny in recent years. This work builds upon some of the techniques that have guided the body of knowledge thus far.

5.1.1 Robustness in Network Samples

In 2003, Costenbader and Valente used network sampling to identify the change in centrality measures. The study used bootstrap sampling procedures (repeatedly sampling various percentages of collected data, with replacement, many times) to determine how much centrality measures changed, in value and rank-ordering. They found a "relatively high" correlation between actual and bootstrapped values from six of the eleven centrality measures analyzed, sampling at ten percent increments from 80& to 40% of the original data. [78].

In 2005, Everett and Borgatti compared the values of betweenness centrality in an ego network (social network comprised of an individual, or ego, his or her distance-1 neighbors, and adjacencies between distance-1 neighbors) to that of the entire network in which the ego exists. They found the values of ego-betweenness and betweenness within the larger network to be highly correlated (> 0.80) [107].

In 2007, Kim and Jeong sampled from Barabási-Albert preferential attachment models at various levels to discover vertices with higher centrality, relative to others in the network, were more robust against uncertainty in data collection; that is, a vertex appearing important in a sample was likely important in the ground truth network. They observed that closeness-based measures may be the most reliable measures under sampling uncertainty [108].

The paragraphs above are summaries of the significant work done in the area of analyzing the robustness of network measures under the uncertainty generated by surveys and samples of surveys. Additional work may be found at [82], [109]–[112].

5.1.2 Robustness Against Error Introduction

Bolland studied the political influence network of Chillicothe, Ohio, by collecting a network of 40 influential individuals. He calculated degree, closeness, and betweenness centrality on the original network and compared those results to results after introducing error. He conducted 100 replications of switching edges with probabilites $0.02, 0.05, 0.10, 0.15, 0.20$. He also conducted an experiment in which he added links to a four vertices, one of high and medium centralies, respectively, and two with low centrality. Overall, Bolland found each of the centrality measures to be correlated with each other, with the high degree of correlation to be between degree and closeness centrality, with betweenness centrality to be less correlated and more volatile under experimental changes [77].

Borgatti et al. explored Erdős-Rényi random graphs at eight levels of network density (1, 2, 5, 10, 30, 50, 60, 90%), six levels each (0 , 1 , 5, 10, 25, 50%) of four kinds of introduced error (node removal, node addition, edge removal, edge addition) and four network sizes. The experiment sought to evaluate change in betweenness, closeness,

degree, and eigenvector centralities as error was introduced into the graphs. They found each centrality measure to have behavior similar to the thers as error was introduced, with betweenness centrality being slightly more sensitive to change than the other three [79].

Franz, et al. built upon the work of centrality robustness against error introduction, extending the analysis to small world networks, scale free, core-periphery, and cellular networks [81], [113], [114]. This analysis found network topology to have a significant effect on the accuracy of centrality measures. That is, uncertainty or error in data collection impacts the accuracy of identifying a key player in some network topologies more than in others. And, types of error have different effects on different ground-truth topologies.

5.2 Methodology

This section describes the steps we took to gain insight into the Distance Centrality's robustness. We used three simulation experiments, listed in increasing order of graph change: re-wiring networks, removing edges, and adding edges. Subsequent to each simulation experiment, we analyze data collected to determine the change in distance, betweenness, closeness, and eigenvector centrality.

5.2.1 Simulation

We noted that the sampling and percolation (vertex or edge) approaches to robustness analysis were well documented in literature. [77]–[79], [81], [82], [107]–[114] In this work we also chose to analyze the effect of re-wiring on networks, an approach we have not seen applied to the topic of centrality measure robustness. We see the re-wiring approach as a useful contribution because it models a phenomenon we've not yet seen modeled: a sensor detecting emissions from a source, but is being unable to pinpoint the source's target.

Re-wiring

For each of the following graphs, we examine the change in centrality caused by a re-wiring of edges in the graph, as in Section 5.2.2.

- Erdős-Rényi random graph on 90 vertices and 140 edges
- Barabási-Albert preferential attachment model on 100 vertices and 291 edges
- Watts Strogatz small world model model on 100 vertices and 200 edges
- the 19-vertex, 32-edge graph introduced in Chapter 2 to describe the centrality measures we are using (hereafter referred to as the "centrality graph"), and
- Zachary's Karate Club graph.

In this experiment we choose to keep the degree of each vertex the same after the re-wiring. We use the double_edge_swap algorithm in Python 3.5.2's Networkx, version 1.11-5, which works as illustrated in Figure 5.1. This algorithm removes two edges (as in the thick, blue edges on the figure's left) and adds two edges to maintain a constant degree (as in the thick green edges on the figure's right).

It is possible for this algorithm to return a graph that is isomorphic to the original graph. This is addressed is Section 5.2.3.

Because the degree remains constant throughout this exploration, degree centrality remains constant and is therefore omitted from centrality analysis in this section.

Figure 5.1: A Depiction of the Networkx Double Swap Re-wiring Algorithm

Edge Removal

For this direction, we explore the following two graphs using the edge removal paradigm outlined in Subsection 5.2.2.

- the "centrality graph", and
- Zachary's Karate Club graph.

5.2.2 Simulation and Data Collection

Re-Wiring

We re-wired the original graphs in two experiments. The first experiment resulted in the smallest amount of perturbation to the original graph, re-wiring a single edge, then two, three, four, and five edges. In each instance, we created 100 new graphs resulting from the re-wirings. For each of those graphs, we calculated distance, betweenness, closeness, and eigenvector centrality for every vertex. We repeated these steps for an experiment with larger graph change, re-wiring percent increments of edges from one to five. In situations where the percentages resulted in other than an integer, we took the ceiling of the calculated decimal.

Edge Removal

We calculated distance, betweenness, closeness, degree, and eigenvector centralities for every vertex after removing every edge from the graph (with replacement), every pair of edges, and a random sample of 5000 3-, 4-, and 5-edge subsets of each graph's edge set (with replacement). Because the random generation of edge subsets resulted in some duplication, there were instances in which the actual number of graphs used was less than the target number. See Table 5.1 for the number of graphs used for each edge removal scenario.

Table 5.1: Number of Graphs Sampled for Edge Removal Experiment

	Centrality Graph		Zachary's Karate Club Graph
1 Edge Swap	32*		78*
2 Edge Swap	496*		3003*
3 Edge Swap	4960*		4972
4 Edge Swap	4986		5000
5 Edge Swap	5000		5000

*Exhaustive in that all possible n-tuples of edges were removed in the experiment.

Edge Addition

Using the graph in Figure 5.2, we add every possible edge that does not currently exist, then every possible pair of edges. Next, we select a random sample of 5000 3-, 4-, and 5-tuples of possible edges that do not already exist in the graph and for each graph

instance, calculate distance, betweenness, closeness, degree, and eigenvector centralities for every vertex.

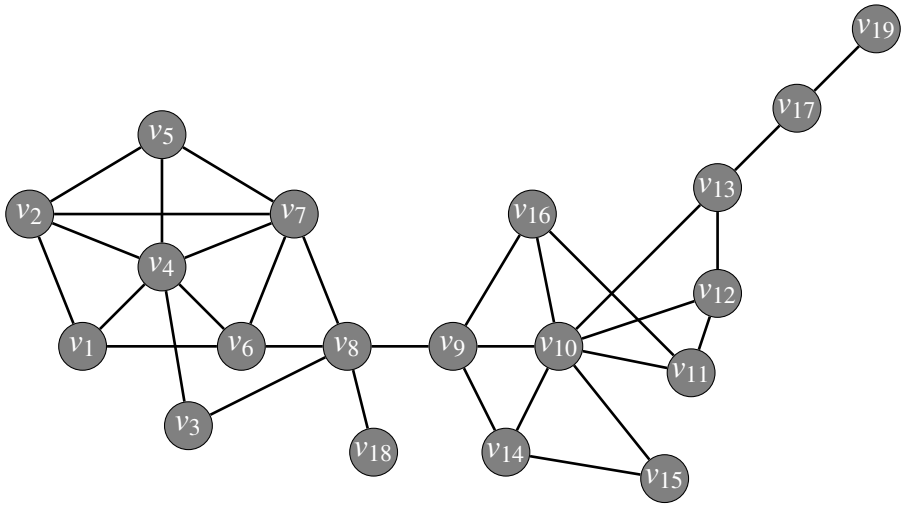


Figure 5.2: Graph Used for Edge Addition

5.2.3 Calculating Change in Centrality Measures

For every level of each experiment, we calculated the four centrality measures for all vertices and collect those measures in four $n \times 1$-dimensional vectors (n is the number of vertices). We then calculate the cosine similarity of each vector to the corresponding vector of centralities for the original graph, defined as the cosine of the angle between the two vectors.

During this stage of the process, we identify replications in which eigenvector centrality failed to converge and eliminate those replicates from further analysis.

Given $\vec{v}_o$ as the original vector of centralities and $\vec{v}_r$ the vector of centralities for a given replicate, we use the standard formulation for the cosine similarity, the cosine of the angle between vectors:

$$Sim_{Cosine} = \frac{\vec{v}_o \cdot \vec{v}_r}{||\vec{v}_o||||\vec{v}_r||}.$$

A cosine similarity near 1 means the two vectors are nearly the same; therefore the

experiment conducted resulted in very little structural change in the network. A cosine similarity near 0 indicates the vectors are nearly orthogonal, or very different and the experiment caused significant topological change to the network.

5.2.4 Summary Statistics

After gathering the cosine similarity data for each replication, we calculate the mean, standard deviation, median, and quartiles of the cosine similarity data. The mean and standard deviation provide insight into the amount of change in the centrality vector after re-wiring and the dispersion of that change data. We use median and quartile data to generate boxplots for each situation, a visual depiction of results. Based upon the aforementioned articles, we expect to see a high cosine similarity and low standard deviation, as centrality measures have been showed to be robust against graph change.

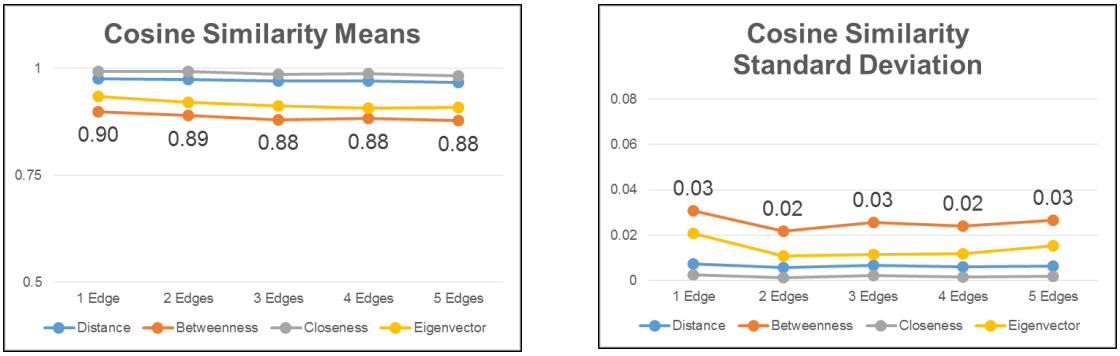
5.3 Results

Our analysis upheld previous findings, in terms of verifying the robustness of betweenness, closeness, and eigenvector centralities against small changes in graph structure. For each graph, we know the standard deviation of closeness centrality values is typically small relative to the other centrality measures, due to the nature of it calculation. This has long been identified as a weakness because relatively small changes in value can result in large change in rank ordering relative to other vertices, artificially inflating the impact of the network change when viewed from a relative perspective [32]. With this information, we expect closeness centrality to vary the least, and it did. The distance centrality consistently ranked second to closeness centrality in highest cosine similarity and lowest standard deviation (or third when degree centrality was included in the edge removal simulation); however, it has none of the limitations of closeness centrality in terms of limitation in range of potential centrality values. Betweenness and eigenvector centrality were least robust against change as seen in these experiments.

5.3.1 Re-wiring $1-5$ Edges

Figure 5.3 depicts the average of cosine similarity mean and standard deviation across completing one, two, three, four, and five double edge swaps, as in the first experiment

described in Section 5.2.



*Degree centrality not included because vertex degree held constant in this experiment.

Figure 5.3: Aggregate Summary Statistics for Swapping 1-5 Edges

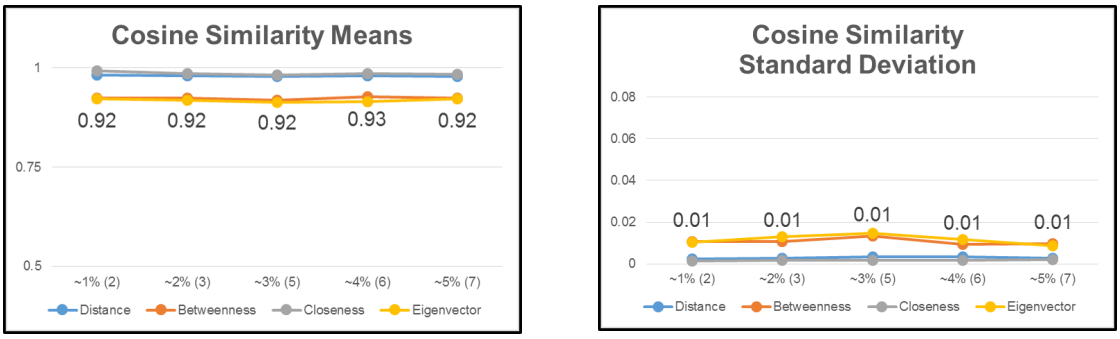
We note very little change in cosine similarity mean values in this experiment. All cosine similarity measures are above 0.88, which corresponds to an average angle between centrality vectors of about 0.495 radians or 28 degrees. The distance centrality means are the second highest (with closeness centrality highest). The distance centrality's standard deviation is the second lowest (with closeness centrality the lowest). We expected that the values of closeness centrality would not change dramatically because the nature of its calculation results in a narrow band of possible values for each graph [32]. The distance centrality may take the full range of possible values from $0-1$; yet, it is still the second highest in cosine-similarity and second-lowest in standard deviation. See Table 5.2 for greater detail about the results of this experiment.

Table 5.2: Aggregate Summary Statistics for Experiment 1: Double Edge Swapping of 1-5 Edges

	Cosine Similarity Mean					Standard Deviation			
	DC	BC	CC	EVC		DC	BC	CC	EVC
1 Edge Swap	0.975	0.899	0.993	0.934		0.007	0.031	0.003	0.021
2 Edge Swap	0.973	0.890	0.993	0.920		0.006	0.022	0.001	0.011
3 Edge Swap	0.970	0.880	0.991	0.912		0.007	0.026	0.002	0.011
4 Edge Swap	0.970	0.883	0.992	0.907		0.006	0.024	0.002	0.012
5 Edge Swap	0.967	0.878	0.990	0.909		0.006	0.027	0.002	0.016

5.3.2 Re-wiring 1%-5% Edges

Figure 5.4 depicts the average of cosine similarity mean and standard deviation across completing double edge swaps of one, two, three, four, and five percent of the graph's edges, as in the first experiment described in Section 5.2.



*Degree centrality not included because vertex degree held constant in this experiment.

Figure 5.4: Aggregate Summary Statistics for Swapping 1%-5% Edges

Again, we note very little change in cosine similarity means with all values above 0.92, which corresponds to an average angle between centrality vectors of about 0.403 radians or 23.090 degrees. The distance centrality means are the second highest, almost the same as closeness centrality highest. The distance centrality's standard deviation is the second lowest, again almost the same as with closeness centrality. See Table 5.3 for greater detail about the results of this experiment.

Table 5.3: Aggregate Summary Statistics for Experiment 1: Double Edge Swapping of 1-5 Edges

	Cosine Similarity Mean					Standard Deviation			
	DC	BC	CC	EVC		DC	BC	CC	EVC
1% Edge Swap	0.982	0.924	0.995	0.922		0.002	0.011	0.001	0.10
2% Edge Swap	0.980	0.924	0.993	0.919		0.003	0.011	0.002	0.013
3% Edge Swap	0.979	0.918	0.991	0.914		0.003	0.013	0.002	0.015
4% Edge Swap	0.980	0.927	0.993	0.916		0.003	0.009	0.002	0.012
5% Edge Swap	0.979	0.923	0.992	0.922		0.003	0.010	0.002	0.009

5.3.3 Removing Edges

The following sections result from analyzing the robustness of centrality measures through removing edges from the graph. We choose the graphs with smaller edge counts

for this analysis to enable as thorough exploration of the graph space as possible. Where it is computationally feasible, we remove all edges, pairs or edges, and combinations of n-tuples (up to $n = 5$) from the graph.

Centrality Graph

Figure 5.5 depicts the cosine-similarity means and standard deviations across removal of one, two, three, four, and five edges from the centrality graph, as described in the second experiment of Section 5.2.

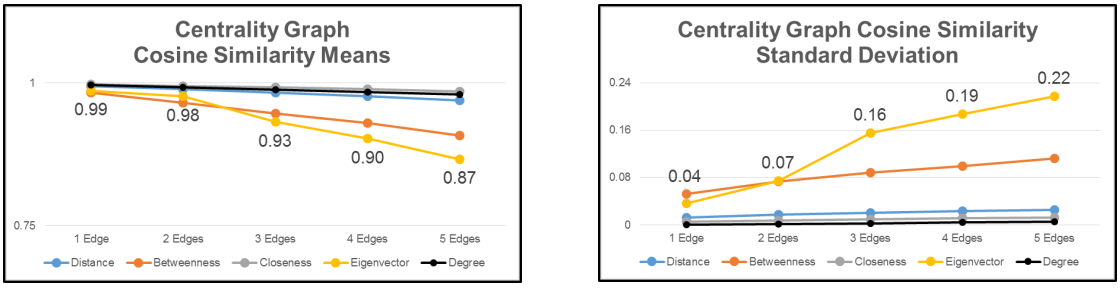


Figure 5.5: Summary Statistics for Centrality Graph

We note very little change in cosine-similarity means in this experiment. However, the statistics for eigenvector centrality resulted from analysis of a slightly different sample than the others. Eigenvector centrality failed to converge for between 33-44% of the samples. See Table 5.4. This lack of computational stability might have contributed to the relative volatility in eigenvector centrality.

Table 5.4: Convergence Summary for Eigenvector Centrality Calculation

Edges Removed	# Total Samples	# Successful Calculation	# Failed to Converge	% Successful
1	32*	18	14	56.25
2	496*	304	192	61.29
3	4960*	3035	1925	61.19
4	4986	3231	1756	64.78
5	5000	3382	1618	67.64

*Exhaustive in that all possible n-tuples of edges were removed in the experiment.

The means of all cosine similarity measures are above 0.867, which corresponds to an average angle between centrality vectors of about 0.522 radians or 29.908 degrees. The

distance centrality means are the third highest, with similar values to the highest and second highest values, closeness and degree centralities, respectively. The distance centrality's standard deviation is the third lowest, with closeness and degree centrality lower. See Table 5.5 for greater detail about the results of this experiment.

Table 5.5: Summary Statistics for Experiment 2: Removal of 1-5 Edges from Centrality Graph

Cosine Similarity Mean					
Edges Removed	DC	BC	CC	EVC	Deg
1	0.994	0.983	0.998	0.987	0.996
2	0.989	0.965	0.995	0.977	0.993
3	0.983	0.947	0.992	0.932	0.989
4	0.977	0.930	0.989	0.902	0.985
5	0.970	0.908	0.985	0.867	0.980
Cosine Similarity Standard Deviation					
1	0.012	0.052	0.005	0.037	0.000
2	0.017	0.073	0.008	0.074	0.002
3	0.021	0.087	0.009	0.156	0.003
4	0.023	0.099	0.011	0.188	0.004
5	0.026	0.112	0.013	0.218	0.006

Karate Club Graph

Figure 5.6 depicts the cosine-similarity means and standard deviations across removal of one, two, three, four, and five edges from the graph depicting Zachary's Karate Club, as described in the second experiment of Section 5.2.

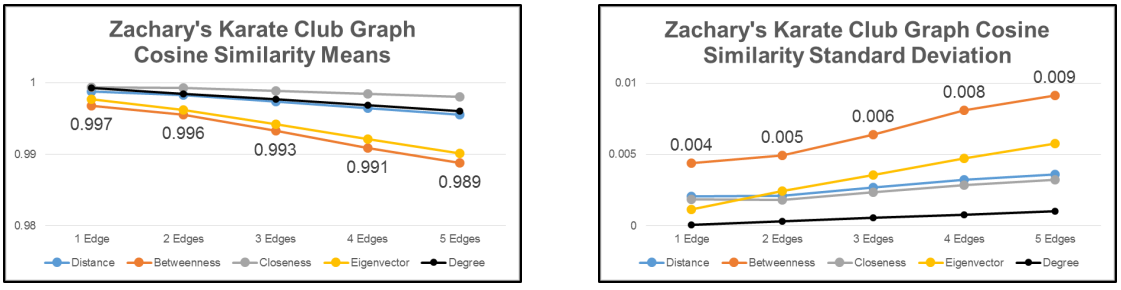


Figure 5.6: Summary Statistics for Zachary's Karate Club Graph

Once again, we note very little change in cosine-similarity means in this experiment. All means are above 0.989, corresponding to an average angle between centrality vectors of

about 0.148 radians or 8.480 degrees. The distance centrality means are the median values fr the five centralities, higher than eigenvector and betweenness centralities and lower than degree and closeness centralities. See Table 5.6 for greater detail about the results of this experiment.

Table 5.6: Summary Statistics for Experiment 2: Removal of 1-5 Edges from Karate Club Graph

Cosine Similarity Mean					
Edges Removed	DC	BC	CC	EVC	Deg
1	0.999	0.997	0.999	0.997	0.999
2	0.998	0.996	0.999	0.996	0.998
3	0.997	0.993	0.999	0.994	0.998
4	0.996	0.991	0.998	0.992	0.997
5	0.996	0.989	0.998	0.990	0.996
Cosine Similarity Standard Deviation					
1	0.002	0.004	0.002	0.001	0.000
2	0.002	0.005	0.002	0.002	0.000
3	0.003	0.006	0.002	0.004	0.001
4	0.003	0.008	0.003	0.005	0.001
5	0.004	0.009	0.003	0.006	0.001

5.3.4 Adding Edges

The following section results from analyzing the robustness of centrality measures through adding edges to the graph shown in Figure 5.2. We add all possible edges and pairs of edges, then combinations of n-tuples (up to $n = 5$) to the graph.

Figure 5.7 depicts the cosine-similarity means and standard deviations across addition of one, two, three, four, and five edges to the centrality graph, as described in the third experiment of Section 5.2.

We note the most change in cosine-similarity means for betweenness centrality, with the other measures experiencing relatively little change.

Although betweenness centrality shows the largest change in cosine similarity mean, all cosine similarity measures are above 0.779, which corresponds to an average angle between centrality vectors of about 0.678 radians or 38.847 degrees. The distance

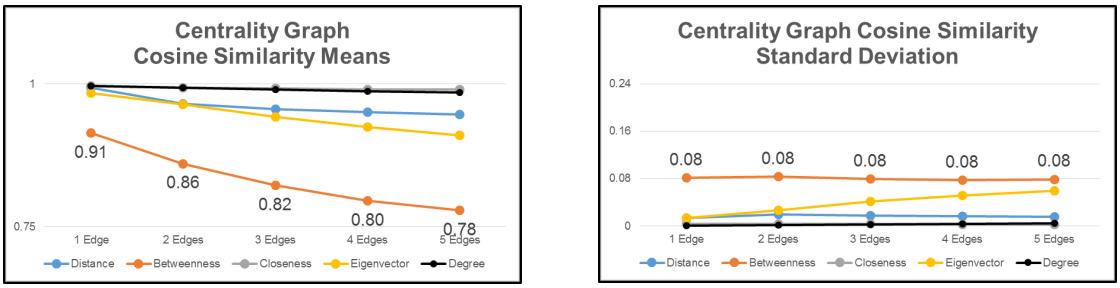


Figure 5.7: Summary Statistics for Edge Addition

centrality means are again the third highest, with similar values to the highest and second highest values, closeness and degree centralities, respectively. The distance centrality's standard deviation is the third lowest, with closeness and degree centrality lower. See Table 5.7 for greater detail about the results of this experiment.

Table 5.7: Summary Statistics for Experiment 3: Addition of 1-5 Edges to Centrality Graph

Cosine Similarity Mean					
Edges Added	DC	BC	CC	EVC	Deg
1	0.994	0.914	0.996	0.984	0.997
2	0.965	0.860	0.994	0.964	0.994
3	0.957	0.947	0.992	0.942	0.991
4	0.950	0.822	0.991	0.924	0.988
5	0.947	0.796	0.990	0.911	0.985
Cosine Similarity Standard Deviation					
1	0.013	0.082	0.003	0.014	0.000
2	0.020	0.083	0.003	0.027	0.002
3	0.018	0.079	0.003	0.041	0.003
4	0.016	0.078	0.003	0.051	0.003
5	0.056	0.079	0.003	0.059	0.004

5.4 Conclusion

In this chapter, we analyzed approximately $34,500$ instances of several different examples of graphs and found that the distance centrality is relatively robust against small changes to the edge set in graphs, when compared to betweenness, closeness, eigenvector, and degree centralities.

The first experiment conducted introduced a condition of graph re-wiring, in which edges were swapped to maintain value of each vertex's degree. Under this paradigm, $1-5$ edges were swapped in one experiment, followed by the swapping of $1-5\%$ of edges. We found distance centrality to be more robust against change than betweenness and eigenvector centrality, and slightly less robust than closeness centrality. In all cases, the mean cosine similarity of the original vector of each vertex's centralities to the vector of centralities from re-wiring the graph was second-highest for distance centrality, with closeness centrality being the highest. The standard deviation of distance centrality was also the second-lowest to closeness centrality in all instances but swapping 2 and 3 edges of the Karate Club graph, in which eigenvector centrality had a lower standard deviation.

In the second experiment, the analysis of the centralities in the Karate Club graph and the centrality graph under the edge-removal paradigm yielded similar results to re-wiring, in terms of the distance centrality's robustness. This time, with the addition of degree centrality, distance centrality exhibited the median robustness of the five centrality measures considered. Distance centrality had a lower mean cosine similarity value (and high cosine similarity standard deviation) than degree and closeness centrality, but a higher value than that of betweenness or eigenvector centralities.

This analysis provides the final part of the distance centrality "story." We saw in Chapter 4 that the distance centrality was derived from the neighbor matrix, in which a significant amount of graph topology is encoded. The distance centrality uses this topological information to identify vertices that are influential on the topology of a graph, in the sense that removal of these vertices causes greatest change to the distances which characterize the given graph. Finally, our analysis shows that the distance centrality is of a similar level of robustness to uncertainty (better than betweeness and eigenvector and worse than degree and closeness) than that of contemporary centrality measures.

CHAPTER 6:
Future Work

This research has identified the neighbor matrix as a compact mathematical structure in which significant information about a graph's structure is encoded. We used the neighbor matrix to define and calculate the distance centrality to identify those vertices with greater influence over the structure of the graph. Then, we showed the distance centrality to be a measure robust against uncertainty in data collection and/or changes in the edge set of the graph. Through the course of this work, we identified many questions that remain unanswered and extensions left yet unexplored. A summary of these questions and topics follows.

6.1 Graphicality of the Neighbor Matrix

In this work, we identified graphs of interest and analyzed their associated neighbor matrices. A standard question in graph theory is: How can we go the other way? That is, given an $m \times n$ matrix, how can we determine the matrix is graphical and therefore it *be* a neighbor matrix? The Havel-Hakimi Theorem algorithm identifies whether or not a sequence of integers is graphical; i.e., if it represents the degree sequence of a graph [11]. Just as the neighbor matrix generalizes the degree sequence from 1 to k dimensions, the answer to this question would generalize the Havel-Hamiki Theorem to several dimensions. Then, given that an $m \times n$ matrix is a neighbor matrix, how can we realize all possible graphs corresponding to the neighbor matrix?

6.2 Isomorphism

We showed in Sections 3.1, Remark 4.2 that it is possible for two non-isomorphic graphs to have the same neighbor matrix. When must graphs with identical neighbor matrices be isomorphic?

6.3 Diversity in Graph Space

We define a graph space as the set of all non-isomorphic graphs with the same neighbor matrix columns. For example, two graphs with neighbor matrices possessing identical first columns are members of the same graph space, as are graphs with the first two, or first three neighbor matrix columns, all the way out to the graph space containing all graphs with identical neighbor matrices. Given this definition, how does the neighbor matrix enable the description of graph space?

- How much diversity is there in the space of graphs with the same neighbor matrix columns? The same neighbor matrix?
- What changes may be made in a graph so there is no change in the neighbor matrix?
- What is the largest change possible in a graph that maintains an identical neighbor matrix?
    - Articulating the definition to the term "largest change" is another open problem.

A related problem is to find the number of neighbor matrix columns needed to obtain a good estimate of the network's topology, much like polynomial approximations can describe non-polynomial functions.

6.4 Expanding the Configuration Model of Network Formation

The configuration model of network formation accepts the degree sequence of a network and uses it to form a random graph. The model forms "stubs," or "half-edges" are created to match the degree of each vertex in the degree sequence. Then, the stubs are joined together to make half-edge whole. [32] Self loops are generally allowed; however, multiple edges are not. If the desired output is a simple graph, the self loops may be removed; however, that practice will lead to the original degree sequence being inexactly replicated.

We see the potential to create of specification of the configuration model that generates a random graph using several columns of the neighbor matrix as and input, rather than only using the first neighbor matrix column (the degree sequence).

6.5 Calculation of Distance Centrality on Different Types of Graphs

Our current work focused solely on the calculation of distance centrality for simple graphs. However, the calculation of this metric is easily extended to other types of graphs.

- Directed Graphs. Although the adjacency matrix for directed graphs is usually asymmetric, calculation of the neighbor matrix and therefore distance centrality is exactly the same.
- Weighted Edges. Under this scenario, neighbor matrix calculation can no longer be accomplished through raising the adjacency matrix to powers; rather, it must be calculated from the distance matrix. Distance centrality calculation is the same as in the simple graph case.
- Multiple Edges. Multiple edges can be counted as weights and solved as above.
- Self Loops. Under this scenario, neighbor matrix calculation can no longer be accomplished through raising the adjacency matrix to powers; rather, it must be calculated from the distance matrix. Distance centrality calculation is the same as in the simple graph case.

6.6 Sequential Removal of Vertices

It will be useful to explore the effect on a graph's distance profiles of sequential vertex removal. Suppose we have the capability to interdict multiple vertices. Thought we may be tempted to interdict the first n-ranked vertices. Suppose vertex i has the highest distance centrality in graph G and vertex j has the second-highest. It may be the case that vertex j does not have the highest distance centrality in graph $G - i$. Therefore, to create the most change in the graph's neighbor matrix we may not choose to remove the n vertices with largest distance centrality in graph G, but rather the vertex with the highest distance centrality after the previously top-ranked vertex was removed.

6.7 Calculation of Distance Centrality for Sets of Vertices

In this work, we calculated the distance centrality as amount of neighbor matrix change resulting from removal of one vertex. We may use similar methodology to calculate an

analogous centrality measure for sets of vertices. Vertices that comprise the set can be chosen through exhaustive enumeration of combinations of vertices, chosen randomly, or with some deliberate methodology.

Suppose V is a non-empty set of vertices. See Equation 6.1 for the formulation.

$$DC_V = \frac{||X^{dist-k}(G) - X^{dist-k}(G-V)||_F}{||X^{dist-k}(G)||_F} \in (0,1) \tag{6.1}$$

We discussed the requirement of "zero padding" to ensure appropriate dimensionality for matrix subtraction in Section 4.3. Note that for vaild matrix subtraction, the matrix $G-V$ will need to be "padded" with the number of rows of zero equal to the cardinality of V.

6.8 Temporal Analysis

Calculating the distance centrality for large graphs is computationally expensive. Given the values of the current metrics at times $1,2,\ldots,t$, how can they be quickly updated at some time $t+1$, rather than recalculated from first principles?

6.9 Conclusion

Confronting the problem of discovering which vertices in a graph exerted the most influence over graph structure, we viewed distance as a proxy measure for topology. From this perspective arose the neighbor matrix. For each vertex, the neighbor matrix shows the number of vertices at every distance from 1 to k, where k is the diameter of the graph, the longest shortest path. We use the neighbor matrix to calculate the distance centrality, a measure of how much a vertex's removal changes the distance profiles in the neighbor matrix. We showed the distance centrality to highlight vertices as important that were overlooked by other centrality measures, then validated the uniqueness of that information with correlation and regression analysis of several networks. Finally, we found that the distance centrality was more robust against small changes in graphs than betweenness and eigenvector centralities (though less so than degree and closeness centralities). In summary, the neighbor matrix and distance centrality both contribute to the body of knowledge informing decision makers about appropriate choices influencing networks when network structure matters.

List of References

[1] M. J. Espinosa-Romero, E. J. Gregr, C. Walters, V. Christensen, and K. M. Chan, "Representing mediating effects and species reintroductions in ecopath with ecosim," *Ecological Modeling*, vol. 222, no. 9, pp. 1569–1579, 2011.

[2] D. L. Alderson, "OR FORUM—Catching the 'network science' bug: Insight and opportunity for the operations researcher," *Operations Research*, vol. 56, no. 5, pp. 1047–1065, 2008.

[3] D. C. Arney, W. Pulleyblank, and K. Coronges, "Integrating information sciences: Operations research, computer science, computational linguistics, analytics, network science, computational sociology, and applied mathematics," *Phalanx*, vol. 46, no. 4, pp. 29–35, 2013.

[4] Army Doctrine Publication 3-0: Operations. (October 2011). Army Publishing Directorate. [Online]. Available: https://www.army.mil/e2/downloads/rv7/info/references/ADP_3-0_ULO_Oct_2011_APD.pdf. Accessed May 9, 2015.

[5] National Security Strategy. (February 2015). WhiteHouse.gov. [Online]. Available: http://nssarchive.us/wp-content/uploads/2015/02/2015.pdf. Accessed: September 16, 2016.

[6] The National Military Strategy of the United States of America. (June 2015). JCS.mil. [Online]. Available: http://nssarchive.us/wp-content/uploads/2015/02/2015.pdf. Accessed: September 16, 2016.

[7] R. K. Wood, "Deterministic network interdiction," *Mathematical and Computer Modelling*, vol. 17, no. 2, pp. 1–18, 1993.

[8] D. Alderson, G. Brown, W. Carlyle, and R. Wood, "Solving defender-attacker-defender models for infrastructure defense," in *Operations Research, Computing and Homeland Defense*, K. Wood and R. Dell, Eds. Hanover, MD: Institute for Operations Research and the Management Sciences, 2011, pp. 28–49.

[9] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.

[10] J. W. Roginski, R. M. Gera, and E. C. Rye, "The neighbor matrix: Generalizing the degree distribution," *arXiv preprint arXiv:1510.06952*, 2015.

[11] G. Chartrand and P. Zhang, *A first Course in Graph Theory*. North Chelmsford, MA, USA: Courier Corporation, 2012.

[12] R. C. Read and D. G. Corneil, "The graph isomorphism disease," *Journal of Graph Theory*, vol. 1, no. 4, pp. 339–363, 1977.

[13] J. Wu and F. Dai, "Efficient broadcasting with guaranteed coverage in mobile ad hoc networks," *IEEE Transactions On Mobile Computing*, vol. 4, no. 3, p. 1, 2005.

[14] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. New York, NY, USA: Springer Science & Business Media, 2013, vol. 207.

[15] N. Biggs, *Algebraic Graph Theory*. New York, NY, USA: Cambridge university press, 1993.

[16] F. R. Chung, *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society, 1997, vol. 92.

[17] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs*. Cambridge, MA, USA: Academic Press, 1979, vol. 87.

[18] D. M. Cvetković, P. Rowlinson, and S. Simic, *Eigenspaces of Graphs*. New York, NY, USA: Cambridge University Press, 1997, no. 66.

[19] D. M. Cvetković, P. Rowlinson, and S. Simić, *An Introduction to the Theory of Graph Spectra*. Cambridge, England, UK: Cambridge University Press, 2010.

[20] S. J. Leon, *Linear Algebra With Applications*. Upper Saddle River, NJ, USA: Prentice Hall, 2010, vol. 8.

[21] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: JHU Press, 2013, vol. 4.

[22] R. Merris, "Laplacian matrices of graphs: a survey," *Linear algebra and its applications*, vol. 197, pp. 143–176, 1994.

[23] S. L. Hakimi and S. S. Yau, "Distance matrix of a graph and its realizability," *Quarterly of Applied Mathematics*, vol. 22, no. 4, pp. 305–317, 1965.

[24] F. Buckley and F. Harary, *Distance in Graphs*. Boston, MA, USA: Addison-Wesley Longman, 1990.

[25] Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić, and N. Trinajstić, "The distance matrix in chemistry," *Journal of Mathematical Chemistry*, vol. 11, no. 1, pp. 223–258, 1992.

[26] J. C. Gower, "Properties of euclidean and non-euclidean distance matrices," *Linear Algebra and its Applications*, vol. 67, pp. 81–97, 1985.

[27] S. Lele, "Some comments on coordinate-free and scale-invariant methods in morphometrics," *American Journal of Physical Anthropology*, vol. 85, no. 1, pp. 407–417, 1991.

[28] S. Lele, "Euclidean distance matrix analysis (edma): Estimation of mean form and mean form difference," *Mathematical Geology*, vol. 25, no. 5, pp. 573–602, 1993.

[29] O. Ivanciuc, T.-S. Balaban, and A. T. Balaban, "Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices," *Journal of Mathematical Chemistry*, vol. 12, no. 1, pp. 309–318, 1993.

[30] H. Wiener, "Structural determination of paraffin boiling points," *Journal of the American Chemical Society*, vol. 69, no. 1, pp. 17–20, 1947.

[31] H. Sayama, *Introduction to the Modeling and Analysis of Complex Systems*. Binghamton, NY, USA: Open SUNY Textbooks, 2015.

[32] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, 2010.

[33] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.

[34] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: definition, properties, and implications," *Internet Mathematics*, vol. 2, no. 4, pp. 431–523, 2005.

[35] D. L. Alderson and L. Li, "Diversity of graphs with highly variable connectivity," *Physical Review E*, vol. 75, no. 4, p. 046102, 2007.

[36] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.

[37] P. Bonacich, "Power and centrality: A family of measures," *American journal of sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[38] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social networks*, vol. 28, no. 4, pp. 466–484, 2006.

[39] D. Ortiz-Arroyo, *Discovering Sets of Key Players in Social Networks*. New York, NY, USA: Springer, 2010.

[40] P. Erdős and A. Rényi, "On random graphs i," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[41] E. N. Gilbert, "Random graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

[42] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hungar. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[43] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[44] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[45] V. Pareto, *Cours d'economie politique: professe a l'Universie de Lausanne [A Political Economy Course: Taught at the University of Lausanne]*. F. Rouge, 1896, vol. 1.

[46] G. U. Yule, "A mathematical theory of evolution, based on the conclusions of dr. jc willis, frs," *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character*, vol. 213, pp. 21–87, 1925.

[47] H. A. Simon, "On a class of skew distribution functions," *Biometrika*, vol. 42, no. 3/4, pp. 425–440, 1955.

[48] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[49] D. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, vol. 27, no. 5, p. 292, 1976.

[50] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge university press, 1994, vol. 8.

[51] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of anthropological research*, pp. 452–473, 1977, downloaded from M.E.J. Newman's website at http://www-personal.umich.edu/ mejn/netdata/ in October 2015.

[52] S. F. Everton, *Disrupting Dark Networks*. New York, NY, USA: Cambridge University Press, 2012, no. 34.

[53] L. N. Visualization and Analysis. (November 2009). Craig Tutterow. [Online]. Available: http://socilab.com/#home. Accessed: February 15, 2016.

[54] R. D. Dutton, S. R. Medidi, and R. C. Brigham, "Changing and unchanging of the radius of a graph," *Linear algebra and its applications*, vol. 217, pp. 67–82, 1995.

[55] T. W. Haynes, L. M. Lawson, R. C. Brigham, and R. D. Dutton, "Changing and unchanging of the graphical invariants: minimum and maximum degree, maximum clique size, node independence number and edge independence number," *Congressus Numerantium*, vol. 72, pp. 239–252, 1990.

[56] O. Vacek, "Diameter-invariant graphs," *Mathematica Bohemica*, vol. 130, no. 4, pp. 355–370, 2005.

[57] J. Von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," *Automata studies*, vol. 34, pp. 43–98, 1956.

[58] E. F. Moore and C. E. Shannon, "Reliable circuits using less reliable relays," *Journal of the Franklin Institute*, vol. 262, no. 3, pp. 191–208, 1956.

[59] I. M. Jacobs, "Connectivity in probabilistic graphs," Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, MA, USA, Tech. Rep. RLE-TR-356-04739034, 1960.

[60] E. P. Durbin, "An interdiction model of highway transportation," RAND Corporation, Santa Monica, CA, Tech. Rep. AD0639602, 1966.

[61] R. Wilkov, "Analysis and design of reliable computer networks," *IEEE Transactions on Communications*, vol. 20, no. 3, pp. 660–678, 1972.

[62] F. Harary, "Changing and unchanging invariants for graphs," *Bull. Malaysian Math. Soc*, vol. 5, pp. 73–78, 1982.

[63] M. A. Henning, "Graphs with large total domination number," *Journal of Graph Theory*, vol. 35, no. 1, pp. 21–45, 2000.

[64] M. M. Halldorsson and J. Radhakrishnan, "Improved approximations of independent sets in bounded-degree graphs via subgraph removal," *Nord. J. Comput.*, vol. 1, no. 4, pp. 475–492, 1994.

[65] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[66] T. Harris and F. Ross, "Fundamentals of a method for evaluating rail net capacities," RAND Corporation, Santa Monica, CA, Tech. Rep. AD0093458, 1955.

[67] G. Dantzig and D. Fulkerson, "On the max flow min cut theorem of networks," RAND Corporation, Santa Monica, CA, Tech. Rep. AD0605014, 1955.

[68] D. Alderson, G. Brown, W. Carlyle, and L. Cox, "Sometimes there is no 'most vital' arc: Assessing and improving the operational resilience of systems," *Military Operations Research*, vol. 18, no. 1, pp. 21–37, 2013.

[69] W. H. Pierce, *Failure-tolerant computer design*. Cambridge, MA USA: Academic Press, 1965.

[70] F. Moskowitz, "The analysis of redundancy networks," *Transactions of the American institute of electrical engineers, part i: Communication and electronics*, vol. 77, no. 5, pp. 627–632, 1958.

[71] O. Wing and P. Demetriou, "Analysis of probabilistic networks," *IEEE Transactions on Communication Technology*, vol. 12, no. 3, pp. 38–40, 1964.

[72] H. Frank, "Vulnerability of communication networks," *IEEE Transactions on Communication Technology*, vol. 15, no. 6, pp. 778–789, 1967.

[73] F. Boesch and R. Thomas, "On graphs of invulnerable communication nets," *IEEE Transactions on Circuit Theory*, vol. 17, no. 2, pp. 183–192, 1970.

[74] D. E. Butler, "Communication networks with specified survivability," Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, USA, Tech. Rep. R-359, 1967.

[75] H. Frank and I. Frisch, "Analysis and design of survivable networks," *IEEE Transactions on Communication Technology*, vol. 18, no. 5, pp. 501–519, 1970.

[76] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[77] J. M. Bolland, "Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks," *Social networks*, vol. 10, no. 3, pp. 233–253, 1988.

[78] E. Costenbader and T. W. Valente, "The stability of centrality measures when networks are sampled," *Social networks*, vol. 25, no. 4, pp. 283–307, 2003.

[79] S. P. Borgatti, K. M. Carley, and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data," *Social networks*, vol. 28, no. 2, pp. 124–136, 2006.

[80] T. W. Valente, K. Coronges, C. Lakon, and E. Costenbader, "How correlated are network centrality measures?" *Connections (Toronto, Ont.)*, vol. 28, no. 1, p. 16, 2008.

[81] T. L. Frantz, M. Cataldo, and K. M. Carley, "Robustness of centrality measures under uncertainty: Examining the role of network topology," *Computational & Mathematical Organization Theory*, vol. 15, no. 4, pp. 303–328, 2009.

[82] A. Landherr, B. Friedl, and J. Heidemann, "A critical review of centrality measures in social networks," *Business & Information Systems Engineering*, vol. 2, no. 6, pp. 371–385, 2010.

[83] F. Harary, *Graph Theory*. Boston, MA, USA: Addison-Wesley, 1969.

[84] B. Bollobás and O. Riordan, "Robustness and vulnerability of scale-free random graphs," *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2004.

[85] E. Estrada, "Network robustness to targeted attacks. the interplay of expansibility and degree distribution," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 52, no. 4, pp. 563–574, 2006.

[86] T. C. Matisziw, A. T. Murray, and T. H. Grubesic, "Exploring the vulnerability of network infrastructure to disruption," *The Annals of Regional Science*, vol. 43, no. 2, pp. 307–321, 2009.

[87] O. the vulnerability of large graphs, "Tong, hanghang and prakash, b aditya and tsourakakis, charalampos and eliassi-rad, tina and faloutsos, christos and chau, duen horng," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 1091–1096.

[88] L. Ford and D. Fulkerson, "Flows in networks," RAND Corporation, Santa Monica, CA, Tech. Rep. R-375-PR, 1962.

[89] R. Wollmer, "Removing arcs from a network," *Operations Research*, vol. 12, no. 6, pp. 934–940, 1964.

[90] A. W. McMasters and T. M. Mustin, "Optimal interdiction of a supply network," *Naval Research Logistics (NRL)*, vol. 17, no. 3, pp. 261–268, 1970.

[91] A. Washburn and K. Wood, "Two-person zero-sum games for network interdiction," *Operations research*, vol. 43, no. 2, pp. 243–251, 1995.

[92] K. J. Cormican, D. P. Morton, and R. K. Wood, "Stochastic network interdiction," *Operations Research*, vol. 46, no. 2, pp. 184–197, 1998.

[93] E. Israeli and R. K. Wood, "Shortest-path network interdiction," *Networks*, vol. 40, no. 2, pp. 97–111, 2002.

[94] G. G. Brown, W. M. Carlyle, J. Salmeron, and K. Wood, "Analyzing the vulnerability of critical infrastructure to attack and planning defenses," in *Emerging Theory, Methods, and Applications*. INFORMS, 2005, pp. 102–123.

[95] G. Brown, M. Carlyle, J. Salmerón, and K. Wood, "Defending critical infrastructure," *Interfaces*, vol. 36, no. 6, pp. 530–544, 2006.

[96] G. G. Brown, W. M. Carlyle, R. C. Harney, E. M. Skroch, and R. K. Wood, "Interdicting a nuclear-weapons project," *Operations Research*, vol. 57, no. 4, pp. 866–877, 2009.

[97] D. L. Alderson, G. G. Brown, and W. M. Carlyle, "Operational models of infrastructure resilience," *Risk Analysis*, vol. 35, no. 4, pp. 562–586, 2015.

[98] J. Devore, *Probability and Statistics for Engineering and the Sciences*. Boston, MA, USA: Cengage Learning, 2004.

[99] R. Axelrod, *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton, NJ, USA: Princeton University Press, 2015.

[100] G. S. Bloom, J. W. Kennedy, and L. V. Quintas, "Some problems concerning distance and path degree sequences," in *Graph Theory*. Springer, 1983, pp. 179–190.

[101] J. W. Roginski and E. C. Rye, "Evaluating structural disruption in adaptive networks," 2014, invited talk presented November 6 in the Military Applications Cluster at the INFORMS Annual Meeting in San Francisco, CA.

[102] J. W. Roginski, "A needle in a haystack: A fresh perspective on selecting influential vertices in complex networks," 2015, invited talk presented June 1 in the Network Science for National Defense Symposium at NETSCI 2015 in Zaragoza, Spain.

[103] J. W. Roginski, "Network attack & defense: A graph comparison approach," 2015, invited talk presented November 1 in the Military Applications Cluster at the INFORMS Annual Meeting in Philadelphia, PA.

[104] S. A. general purpose network analysis and graph mining library in C++. (November 2009). Jure leskovec and rok sosič. [Online]. Available: http://snap.stanford.edu/snap. Accessed: June 16, 2014.

[105] J. W. Roginski, "Who's adding value to your personal and professional networks?" 2016, slides presented April 4 at the XXXVI Sunbelt Conference of the International Network for Social Network Analysis, Newport Beach, CA.

[106] SAS Institute Inc., *JMP Software, Version 12*, Cary, NC, 2015. [Online]. Available: http://www.jmp.com/

[107] M. Everett and S. P. Borgatti, "Ego network betweenness," *Social networks*, vol. 27, no. 1, pp. 31–38, 2005.

[108] P.-J. Kim and H. Jeong, "Reliability of rank order in sampled networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 55, no. 1, pp. 109–114, 2007.

[109] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose." in *ICWSM*, 2013.

[110] T. Wey, D. T. Blumstein, W. Shen, and F. Jordán, "Social network analysis of animal behaviour: A promising tool for the study of sociality," *Animal behaviour*, vol. 75, no. 2, pp. 333–344, 2008.

[111] J. A. Smith and J. Moody, "Structural effects of network sampling coverage i: Nodes missing at random," *Social networks*, vol. 35, no. 4, pp. 652–668, 2013.

[112] P. V. Marsden, "Network data and measurement," *Annual review of sociology*, vol. 16, no. 1, pp. 435–463, 1990.

[113] S. P. Borgatti and M. G. Everett, "Models of coreperiphery structures," *Social networks*, vol. 21, no. 4, pp. 375–395, 2000.

[114] E. M. Airoldi and K. M. Carley, "Sampling algorithms for pure network topologies: A study on the stability and the separability of metric embeddings," *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 13–22, 2005.

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California