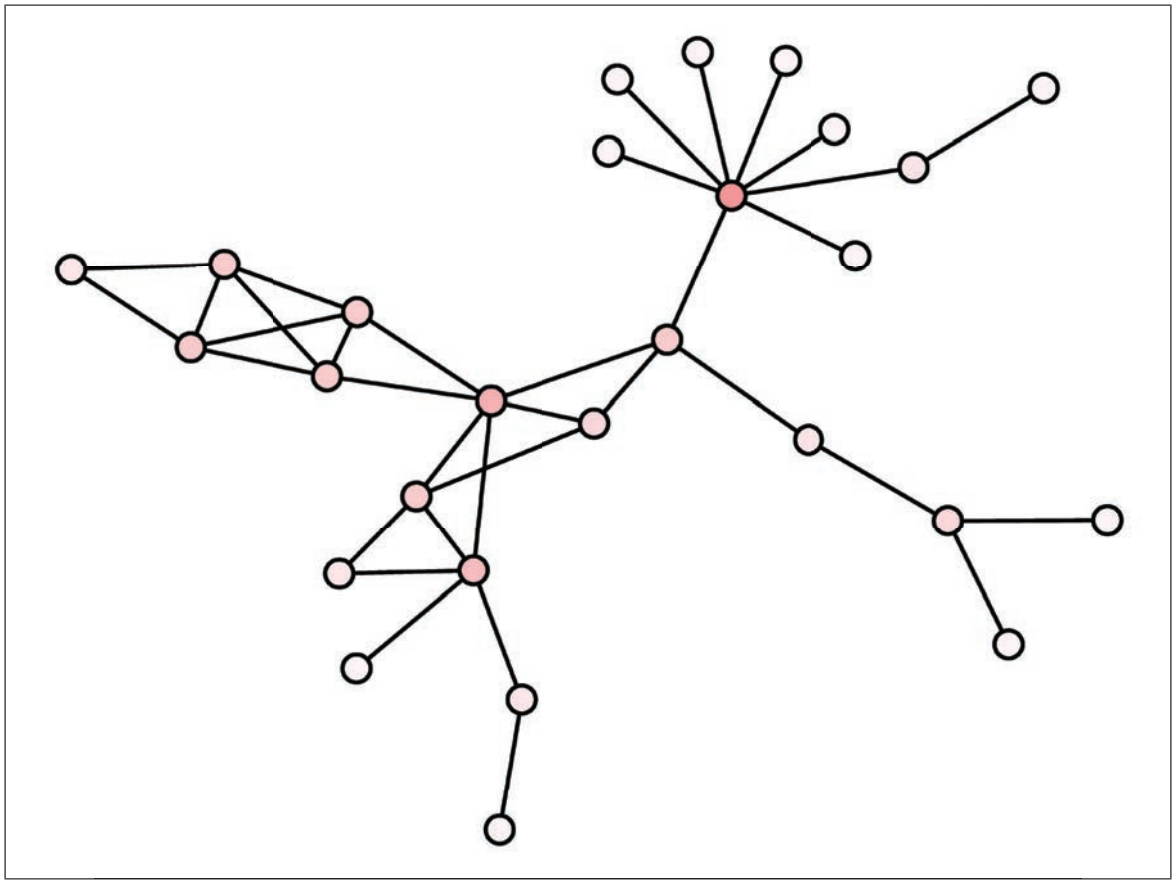

Research Report

DON SNYDER, CHRISTIAN JOHNSON, KATHERINE C. HASTINGS, BART E. BENNETT,
LANCE MENTHE, JOSHUA STEIER

# A Primer on Graph-Theoretic Models and Metrics

## Using Graphs for Modeling and Assessing Robustness



For more information on this publication, visit **www.rand.org/t/RRA1506-1**.



Published by the RAND Corporation, Santa Monica, Calif.

# About This Report

The objective of this project was to provide tools to improve the rigor, reproducibility, and scalability of analytical tools that support cybersecurity assessment, nuclear surety, and nuclear safety certification.

There are two intended audiences for this report. The first audience is those providing analytical support to assess the robustness of systems and missions. The analytical setting that motivated this research was the assessment of cybersecurity, nuclear surety, and nuclear safety. But the work can be applied broadly to any analysis of resiliency in which graphs (networks) are a suitable model. The second audience is those consuming analysis to inform decisions. This group includes authorizing officials for cybersecurity and certifiers for nuclear surety and safety. These decisionmakers will often be presented with analyses of robustness. The more that they know about the strengths and weaknesses of graphical methods, the better informed their decisions can be.

The research reported here was commissioned by the U.S. Air Force Program Executive Office for Strategic Systems and conducted within the Force Modernization and Employment Program of Project AIR FORCE as part of a fiscal year 2022 project, "Algorithms to Support a Nuclear Unified Certification Strategy."

## Project AIR FORCE

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the Department of the Air Force's (DAF's) federally funded research and development center for studies and analyses, supporting both the United States Air Force and the United States Space Force. PAF provides the DAF with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. Research is conducted in four programs: Strategy and Doctrine; Force Modernization and Employment; Resource Management; and Workforce, Development, and Health. The research reported here was prepared under contract FA7014-16-D-1000.

Additional information about PAF is available on our website:

http://www.rand.org/paf/

This report documents work originally shared with the DAF on March 25, 2022. The draft report, dated March 2022, was reviewed by formal peer reviewers and DAF subject-matter experts.

## Acknowledgments

We thank Major General Anthony Genatempo for sponsoring the work. Brigadier General (select) Jason Bartolomei and David Wright provided day-to-day support critical to the execution of the research. (Ranks are those at the time of this research.) We also deeply appreciate the collegial interactions with the entire Sentinel (formerly Ground Based Strategic Deterrent) Unified Certification Strategy team, too numerous to cite individually.

At RAND, we thank Jennifer Brookes and Jim Powers for reviewing an earlier draft of this report. We thank Teddy Parker and Matthew Sargent for formal reviews.

That we received help and insights from those acknowledged above should not be taken to imply that they concur with the views expressed in this report. We alone are responsible for the content, including any errors or oversights.

# Summary

As weapon systems become more complex, the need increases for mathematical tools to assess their properties with the requisite rigor. The increasing digitization of engineering records in single, authoritative databases enables the use of more-sophisticated modeling tools. For nuclear programs, the need for improved assessment tools is most acute in the areas of nuclear certification, nuclear surety, and cybersecurity risk assessment. This report is a primer on the use of graph (network) theory to support such analysis. The goals are to stimulate analysts on the power of graph theory to support nuclear certification, nuclear surety, and cybersecurity risk assessment and to give those receiving such analysis a better context for understanding how to use the resultant insights.

Our work summarizes the key areas of graph theory pertinent to nuclear certification, nuclear surety, and cybersecurity risk assessment. Much of the discussion is a critical summary of the literature, some of which are quite recent discoveries. We combine the insights from the literature with our own experiences in modeling similar problems with graph theory.

This primer explains when a graph is a useful model of a system and when it is not. It gives a few examples of constructing simple graphs, emphasizing that even a single system can be usefully represented by a graph in different ways depending on what question is being addressed. There is no universal graph representation of any system. We go on to show that graphs can be usefully applied to model systems and processes that are not evidently networks, such as safety and insider threat assessments.

We review the principal metrics for expressing how important individual nodes and links are in a graph. Each of the metrics measures a different aspect of a node or link and therefore provides a different insight. Which to use depends on the question being addressed, and when to use and not use each metric is discussed.

The overall structure of a graph, called its *topology,* governs the behavior of the system in important ways. We discuss three common topological types expected to arise in nuclear certification, nuclear surety, and cybersecurity risk assessment. We show how each plays roles in the robustness of the system and in how flows move across a graph. It is important to ensure flow in some circumstances (e.g., information dissemination) and to impede flow in others (e.g., stopping malware propagation). We also discuss community structures (modularity) in a graph, which are structures larger than the scale immediately surrounding a node but smaller than the global structure. In graphs with strong modularity, both robustness and flow are highly controlled by this intermediate (mesoscopic) structure.

The more that decisionmakers for nuclear certification, nuclear surety, and cybersecurity risk assessment and analysts closely collaborate, the more rigorous and useful graph theoretic modeling—and all modeling—can be to support decisions.

# Contents

# Figures and Table

## Figures



## Table

# Chapter 1. Overview

> Complex systems routinely exhibit behaviors that are difficult to predict or even to understand. The challenge of representing, analyzing, designing, and optimizing such systems demands the power and conceptual rigor of mathematical approaches such as modeling.
>
> —John M. Borky and Thomas H. Bradley[1]

As the U.S. Department of Defense expands its adoption of digital engineering,[2] and specifically model-based systems engineering (MBSE),[3] new opportunities arise to exploit the resulting rich data environment for purposes beyond system design. MBSE environments comprise data elements, attributes of those data (often called *stereotypes*), and relationships among those data. These data characteristics are exactly the ones that describe the interrelationships of networks. Networks—called *graphs* in mathematics—are models of systems that depict the relationships (links) among entities (nodes).

Graphs can capture a variety of relationships—and the strength of those relationships—among the associated entities, including but not restricted to connections among components, interdependencies of elements, how information or power flows through a system, and sequencing of events. When a system can be represented as a graph, a number of powerful mathematical theorems and algorithms become available for the rigorous analysis of the system. Graphs are particularly strong models for assessing the ability of a system to experience the failure of its entities or their relationships and still maintain acceptable function, which we call the *robustness* of the system. Mathematics offers many tools for assessing the robustness of graphs.

The purpose of this report is to present guidance on how MBSE-type information can be used to assess the robustness of various aspects of weapon systems using graphs. The presentation is based on a literature review of applied graph theory and our collective experiences in modeling systems with graphs. We show how to select an appropriate graphical representation of a system for given problem types, how such critical attributes as robustness can be measured, how to

---

[1] John M. Borky and Thomas H. Bradley, *Effective Model-Based Systems Engineering*, Springer, 2019, pp. 1–2.

[2] *Digital engineering* is "[a]n integrated digital approach that uses authoritative sources of systems' data and models as a continuum across disciplines to support lifecycle activities from concept through disposal." Defense Acquisition University, *Glossary of Defense Acquisition Acronyms and Terms*, July 21, 2020, p. 2.

[3] *MBSE* is "[t]he formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases." International Council on Systems Engineering, *Systems Engineering Vision 2020*, Version 2.03, INCOSE-TP-2004-004-02, September 2007, p. 15.

select the appropriate tools to assess robustness, and what insights can, and cannot, be gleaned from graphical analysis.

The presentation in this report is not an introduction to graph theory or network science. For that, we refer the reader to standard texts.[4] We review selected topics that we assess are most directly applicable to using MBSE-type data for analyzing robustness of systems. The definitions and observations are heuristic rather than expressed in rigorous mathematical terms. Parts of the report assume knowledge of matrix algebra. The purpose of the report is to summarize what is known from mathematical rigor and empirical studies, not to prove the results.

## Motivation

The motivation for this primer is to provide more rigorous and scalable tools for assessing the nuclear surety, nuclear safety, and cybersecurity of the Sentinel intercontinental ballistic missile program (LGM-35A), which was formerly called the Ground Based Strategic Deterrent. All three of these assessments involve evaluating robustness of systems or processes. The Sentinel program is a full replacement of the nuclear-armed Minuteman III intercontinental ballistic missile system. The program is designing and producing more than 600 missiles, refurbishing 450 operationally configured launch facilities (silos), designing and producing at least 24 launch control centers, and creating the requisite weapon system command and control, as well as producing numerous support equipment and facilities for test and operations support.

Sentinel presents two critical challenges for assessing nuclear surety, nuclear safety, and cybersecurity. The first is the size and complexity of the program. Being so large and involving so many interrelated systems, doing assessments "by hand" is not practicable. The second is the paramount importance of the nuclear surety, nuclear safety, and cybersecurity of such a destructive weapon system. It is imperative that the best possible assessments of robustness be undertaken for nuclear systems. Graphical models provide the ability to both handle the complexity and provide the mathematical rigor needed for the assurance demanded for nuclear-armed systems.

Our goals are (1) to present the power of graph analysis to those responsible for nuclear surety, nuclear safety, and cybersecurity so that they are aware of these tools that can help them achieve the desired analytical rigor when assessing complex systems and can better interpret analysis presented to them, and (2) to summarize for the supporting analysts some of the more recently discovered insights from the literature that should be of use in Sentinel. Although devised for the application to Sentinel, the guidance in this primer is applicable to programs beyond Sentinel.

---

[4] Mark Newman, *Networks*, 2nd ed., Oxford University Press, 2018; Katharina A. Zweig, *Network Analysis Literacy: A Practical Approach to the Analysis of Networks*, Springer, 2016; and David Easley and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010, are good, recent textbooks surveying the elements of graph theory in practical contexts.

## Graphs as Models

As a point of departure, it is important to clarify that systems are not graphs. Graphs are abstractions of systems in the form of a model. The word *network*, synonymous with the word *graph*, is nevertheless often used to describe systems. We hear of information technology networks, transportation networks, power distribution networks, and so forth. These are fine colloquial expressions. But for the level of precision needed in this primer, the internet, for example, is a real, physical set of routers, cables, transmitters, and other components. The internet is not a graph (network). It is, of course, quite amenable to being represented as a graph. When well constructed, a graph captures key attributes of a system such as the internet, but any graph simplifies a system and thereby loses a lot of detail.

The utility of a graphical representation depends on how well what the graph includes expresses the attributes of interest and whether what the graph excludes is minimally relevant to the problem being assessed. As we argue, systems such as the internet—or any system—are best represented by different graphs depending on the problem being addressed. If we are concerned about cyberattacks against specific types of routers, we might choose a graph that represents the logical connection of those routers with other key elements of the internet. If we are concerned with the consequences of the loss of a bridge, cable landing station, or tunnel through which internet cables transit, we might choose a graph that represents the physical layout of key components of the internet. A graph is a model, and the model should be tailored to the problem being addressed.[5]

MBSE databases are useful repositories of detailed data, attributes of data, and interrelationships of data for a program *from which useful models can be constructed*. These databases, although often structured as a graph, are not themselves models tailored to a specific problem. A secondary benefit of this primer will be in helping analysts identify what information should be captured and maintained in MBSE databases to support model building to address critical problems. But the primary focus of this report will be how to form a useful graphical representation of a system to inform a particular problem and which mathematical tools can be

[5] For recent applications relevant to Sentinel, see Peter Davison, Bruce Cameron, and Edward F. Crawley, "Technology Portfolio Planning by Weighted Graph Analysis of System Architectures," *Systems Engineering*, Vol. 18, No. 1, 2015; Willie K. Harrison, "The Role of Graph Theory in System of Systems Engineering," *IEEE Access*, Vol. 4, 2016; Matthew W. Potts, Pia Sartor, Angus Johnson, and Seth Bullock, "A Network Perspective on Assessing System Architectures: Foundations and Challenges," *Systems Engineering*, Vol. 22, No. 6, 2019; Matthew W. Potts, Pia A. Sartor, Angus Johnson, and Seth Bullock, "A Network Perspective on Assessing System Architectures: Robustness to Cascading Failure," *Systems Engineering*, Vol. 23, No. 5, 2020; Mareike Bockholt, and Katharina Anna Zweig, "Towards a Process-Driven Network Analysis," *Applied Network Science*, Vol. 5, No. 1, 2020; and Bryan M. O'Halloran, Nikolaos Papakonstantinou, Kristin Giammarco, and Douglas L. Van Bossuyt, "A Graph Theory Approach to Predicting Functional Failure Propagation During Conceptual Systems Design," *Systems Engineering*, Vol. 24, No. 2, 2021.

brought to bear on that problem. Forming the most useful model for a given problem is part skill and part art, so we will provide examples as we discuss the concepts.[6]

## Model Formulation

The first step in model formulation is to define the problem. What insights are needed, and what decisions will be informed by the analysis? In the case of assessing robustness, the key questions to answer are: What aspects of the system are to be assessed for robustness? And, against what is the system to be robust? For example, an analyst might want to assess the robustness of the system design to adversary cyberattacks. Or they might want to assess the robustness of the system to physical attacks or accidental physical failure of collocated subsystems. Or they want to assess the robustness of the flow of information or power through the system when nodes or links are lost. Or they might want to assess the robustness of nuclear surety against an insider threat. The potential problems that can be assessed are nearly endless. Each of these problems will require a different and tailored graphical model of the system.

The analyst should, at this stage, state as well as can be articulated for the model of the system what data elements are being captured, along with their interrelationships, and any potentially important system attributes that could be lost in the model formulation. The fidelity of a model depends on the problem at hand: A particular graph might be quite good for assessing the robustness to one threat and quite poor for another. It is also vital at this stage to explicitly state any assumptions that are made in formulating the model of the system in the form of a graph. These assumptions will be necessary inputs to the validation of the model and the analytical results derived from it and, ultimately, how well the analysis responds to the decisionmaker's question and how assiduously the result supports the final decisions.

Note that sound modeling practice moves from defining the problem to be assessed to selecting the appropriate model formulation and not the other way around. It is poor modeling practice to approach a problem with a predetermined model (e.g., a graph) and assessment measures and calculational methods (elements of graph theory). The model and tool should be tailored to the specific problem. In the case of graph theory, it is a tool that is especially well suited for assessing the ability of a system to experience failure of its entities or their relationships and still maintain acceptable function, which is why we limit the discussion in this report to the assessment of robustness.

Modeling systems as graphs will be examined in Chapter 2.

---

[6] Classic surveys of the practice of modeling include C. C. Lin and L. A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*, Macmillan Publishing Co., 1974; Edward A. Bender, *An Introduction to Mathematical Modeling*, John Wiley & Sons, 1978; Thomas L. Saaty and Joyce M. Alexander, *Thinking with Models: Mathematical Models in the Physical, Biological, and Social Sciences*, Pergamon Press, 1981; Rutherford Aris, *Mathematical Modelling Techniques*, Dover, 1994; and A. C. Fowler, *Mathematical Models in the Applied Sciences*, Cambridge University Press, 1997.

## Graph Measures

In many branches of science, theory and experiment have verified that a known, finite set of variables fully characterizes a system. For example, in classical mechanics, a finite set of parameters of a system fully characterizes that system, and any future state of the system can be calculated from it.[7] For graphs, there is no similar single measure that fully describes the robustness of the graph. The reason is that graphs depict various phenomena, and depending on which phenomena is of interest, a different measure, or set of measures, is needed. What is important in one context might not be important in another. Choosing the most-appropriate measures requires both skill and art. It is one of our principal goals of this primer to make this process of selecting measures easier for the analyst and easier for the consumer of the analysis to understand and appropriately use results to inform decisions.

The robustness of a graph depends on attributes at both the local level and the global level. At the local level, robustness of a system against the loss of individual nodes and links is quantified by *centrality metrics*. Centrality metrics are attributes of individual nodes and/or links. At the global level, robustness of a system is reflected in the overall structure of the graph, called its *topology*, and is measured by a variety of *topological measures*. Topological measures are properties of the graph as a whole. Although topological considerations should generally precede considerations of individual components, we will discuss centrality metrics first because an understanding of centrality metrics is necessary to understand some of the measures of topology.

### *Centrality Metrics*

Centrality metrics assign a scalar value to each node or link in a graph that expresses some sense of its importance to the overall robustness of the graph. Centrality metrics generally express the consequences of the loss of an individual node or link to some overall functionality of the graph and, hence, how fragile (anti-robust) the graph is to that loss.

We will discuss four topics pertaining to a number of commonly used centrality metrics:

- what the metric measures, expressed in words, and what use it is best suited to
- how the metric is expressed mathematically
- how computationally intensive it is to calculate the metric, which provides insight into whether it is practical for use in large, complex graphs
- one or more examples of the use of the metric.

Centrality metrics are the topic of Chapter 3.

[7] We note that the precision to which those parameters are known limits the precision to which future states can be calculated. For chaotic systems, this limit can be severe.

### *Topological Measures*

Topological measures capture how the global structure of a graph determines how the system that the graph represents functions. These measures help to answer such questions as: Are there nodes or links that can be targeted that disproportionately cause failure? In a graph that depicts flow of some entity, such as information or power, when parts of the system fail, will there be other meaningful parts that operate well and others that operate poorly? What is the likelihood that failure in one part of the graph will cause cascading failures in others? Many topological measures are expressed as the statistical distribution of centrality metrics, which is why we discuss centrality metrics before we discuss topological measures.

Measures of topology are the topic of Chapter 4.

Chapter 5 summarizes the conclusions reached and offers final remarks on the use of graphs for assessing robustness and the choice of measures of robustness.

# Chapter 2. Model Formulation

> Mathematical modeling is a subject that is difficult to teach. It is what applied mathematics . . . is all about, and yet there are few texts that approach the subject in a serious way. Partly, this is because one learns it by practice: There are no set rules, and an understanding of the 'right' way to model can only be reached by familiarity with a wealth of examples.
>
> —A. C. Fowler[8]

A model is a representation of a system that highlights specific features, reduces the size or complexity of the system, or allows key components to be examined and analyzed without the constraints that may be imposed by costs, latency, or risks in the real world. For example, if performed directly (without using a model), exploring options to increase the robustness of a communication system may be prohibitively expensive or overly time-consuming, or it may undesirably interfere with operations on the actual communication system. Exploring options using a model of the system can provide the desired insights without these impositions.

Models are never exact replicas of actual systems. They purposefully omit certain components and functions of a system not relevant to the problem being addressed. In this way, models are always wrong in some way but may be sufficiently right for their intended purpose.[9] Indeed, models arguably are useful *only* insofar as they fulfill specific purposes. With the right choice of formalism, however, a model that performs sufficiently like the system itself to inform a particular question can be constructed and thus can provide insights regarding that question.

The inherent incompleteness of models is perhaps the primary reason that model formulation—the process of abstracting the real world into a simplified form for the desired purpose—is so difficult. Throughout the formulation process, the analyst must ask: Has everything necessary to meet the purpose been included? Has everything unnecessary been excluded so that the model truly focuses on what is most important? The process of model formulation is therefore often iterative and often leads to a richer understanding of the system and the associated analytic questions.

Graphs are a desirable modeling formalism because a large body of algorithms and statistical machinery exists to analyze them. But many other formalisms exist beyond representation by graphs. When considering whether any particular modeling formalism should be used, there are many considerations: Which formalism provides the greatest flexibility in representing the

[8] Fowler, 1997, p. 3.

[9] Said another way by statistician George Box, "Essentially all models are wrong, but some are useful." Ron Wasserstein, "George Box: A Model Statistician," *Significance*, Vol. 7, No. 3, 2010, p. 134.

system? Which has tools that will allow me to answer the desired questions about the system? Which imposes the fewest constraints on the analysis? Is the model sufficiently parsimonious to be practicable?

In this chapter, we explain when a graph is an appropriate formalism to model a system, and we underscore that a system may often be represented as a graph in multiple ways: A system does not have an intrinsic graph, and there is no such thing as a universal graph representation of a system. We show as well that graphs are a versatile modeling formalism that can represent many different systems for many different purposes, including systems that are not obviously networks.

## When a Graphical Model Is Appropriate

In common parlance, a graph refers to any quantitative diagram (such as a plot of a function), but in the mathematical field of graph theory, a graph is a particular class of mathematical objects. Broadly speaking, a graph represents relationships among entities. The entities are represented by *nodes* (also called *vertices*), and the relationships among them are represented as the *links* (also called *edges* or *arcs*) that connect the nodes. Graphs are often an appropriate representation of a system when the critical attributes of the questions being posed are well characterized by *discrete, well-defined entities* in the system interacting in some *well-defined manner*. Examples include traffic flowing over a road network[10] and power flowing through electrical circuits—but they can also represent more-abstract relational data, such as project schedules and social networks.

### *Computer Networks*

Computer networks are examples of systems that are amenable to representation as a graph. The graph formalism is appropriate because a computer network consists of discrete routers connected by distinct communication lines. There is no ambiguity in what constitutes a router or a communications line, although the physical nature of the communication lines might differ (e.g., copper cable, fiber optic, or wireless).

The purpose of modeling a computer network as a graph might be to answer some of the following questions: How robust is communication from one router to another across the network to loss of a subset of routers? In what way might we expect malware to propagate (infecting routers and transmitting over links)?

[10] Indeed, the first application of what would become graph theory was a road network problem: Leonhard Euler's solution to the Seven Bridges of Königsberg in 1736. See Norman L. Biggs, E. Keith Lloyd, and Robin J. Wilson, *Graph Theory, 1736-1936*, Clarendon Press, 1986.

### *Power Grids*

Another example of a physical network that is well-suited to a graph representation is an electrical power grid. Consider the question of how to efficiently detect intentional or unintentional disruption. It would be prohibitively expensive to install monitoring software on every single power line to watch for fluctuations. The total number of power stations, transformers, and generators is smaller than the number of separate power lines—there are generally about 80–85 percent as many power stations as power lines[11]—so installing links at every power station would be less expensive, but likely still very costly. However, because every power line need be monitored *only at one end*, it may be possible to cover the entire network with fewer monitors.

This problem has been addressed by representing the power grid as a graph, where the nodes represent power stations, generators, and buses, and the links represent power lines. This representation is appropriate because the stations, generators, and buses are a well-defined, discrete set of components, as are the power lines that join them. Using the graph formalism, determining the minimum set of monitors proves to be an example of a well-known—albeit difficult[12]—problem known as the *minimum vertex cover* problem. This method has been used to demonstrate that a surprisingly small number of software installation sites can be used to monitor the Western Interconnection of North American power grid.[13]

## Modeling a System Using Graphs

### *Directed Versus Undirected Graphs*

Although graphs can represent a wide variety of systems, there are rules that limit their construction. The most important rule is that the relationships modeled by the graph must be binary or pairwise.[14] In other words, a link cannot connect *directly* to other links and a node cannot connect *directly* to other nodes; rather, each link must join exactly two nodes, and each

---

[11] Sinan G. Aksoy, Emilie A. H. Purvine, Eduardo Cotilla Sanchez, and Mahantesh Halappanavar, "A Generative Graph Model for Electrical Infrastructure Networks," *Journal of Complex Networks*, Vol. 7, No. 1, 2019, Table 1.

[12] This is a difficult problem that is the dual to the maximum clique problem. See Qinghua Wu and Jin-Kao Hao, "A Review on Algorithms for Maximum Clique Problems," *European Journal of Operational Research*, Vol. 242, No. 3, 2015.

[13] Éric Filiol and Cécilla Gallais, "Optimization of Operational Large-Scale (Cyber) Attacks by a Combinatorial Approach," in Information Resources Management Association, ed., *Cyber Warfare and Terrorism Concepts, Methodologies, Tools, and Applications*, IGI Global, 2020.

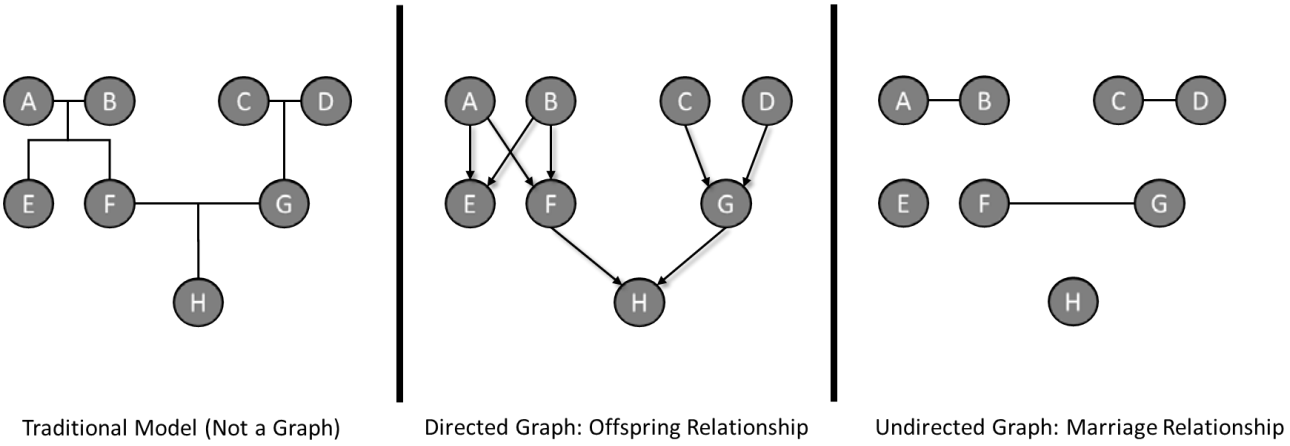
[14] So-called hypergraphs exist where the links can be more like sheets or membranes that join multiple points at the same time, but the statistical machinery for analyzing them is much poorer.

pair of nodes may be joined by no more than one link.[15] Note that while links must join to nodes—they cannot lead nowhere—the reverse is not true: Isolated nodes may be (and often are) part of a graph. Links can have differing weights, reflecting, for example, differing flow capacities between nodes.

If the relationships being modeled treat both nodes symmetrically, such as "these two stations are connected by track," or "these two people are acquainted," or "these two computers can exchange information," then this yields an *undirected* graph. The connections in these cases are depicted by links that join the nodes. The examples discussed previously would most likely be modeled as undirected graphs. On the other hand, if the relationships treat the two nodes differently, such as "Alice sends a letter to Bob," or "Task A must be completed before Task B," or "Annapolis provides electrical power to Baltimore," then this produces a *directed graph* (or sometimes *digraph*) and the typical representation of the link is an arrow.

For this reason, not everything that looks like a graph is a graph—and deciding whether to use a directed or undirected graph is an important modeling choice. For example, as shown in Figure 2.1, the classic representation of a family tree is *not* a graph, but multiple useful graphs can be extracted from it.

**Figure 2.1. Family Tree Representations**




The representation on the left of Figure 2.1 violates the required structure of graphs because links join directly to other links. This occurs in large part because we are trying to encode two different kinds of relationships within the same graph: marriages and offspring. We can, however, easily extract two proper graphs from the family tree. The diagram in the middle of

[15] Technically speaking, graph theory allows for multiple links between nodes, and even links that join a node back to itself, but almost all the algorithms discussed in this report forbid such loops. Where such representations are unavoidable, one can create dummy nodes to build richer structures within this constraint.

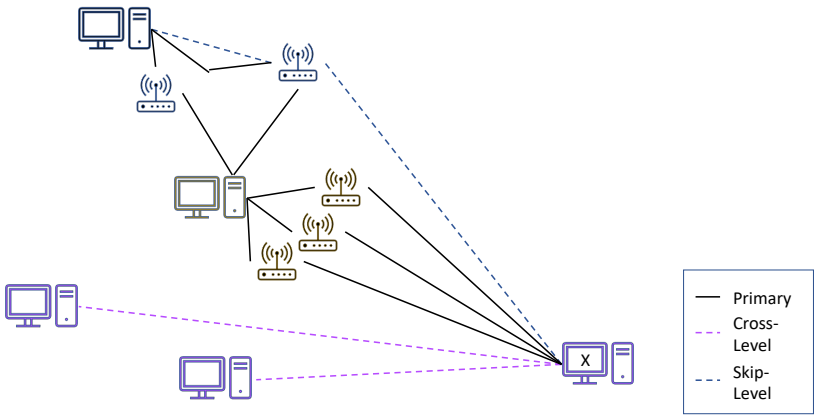
Figure 2.1 depicts a directed graph showing offspring relationships, whereas the one on the right depicts an undirected graph showing marriages.

Note also that the length and positioning of the links in the diagram is irrelevant. All that matters is which nodes they join. As a reminder, the reason for representing a system with the graph formalism is not to build a pretty picture but to facilitate an analytic purpose, such as understanding a system's connectivity, robustness, and vulnerabilities. We usually think of this as a logical rather than a physical representation.

### *Creating Different Graphs to Model the Same System*

A single system that can be represented as nodes and interconnecting links can often be usefully represented as a graph in more than one way. Which way to model the system depends on the question the model is meant to inform. Consider a simple computer network. Figure 2.2 shows one way to model the computer network as a graph. This representation emphasizes the logical interconnections among the various computers. Nodes indicate computers and links indicate the independent paths along which information flows among them. The sense in which these links are independent is that they use different cables, different wireless means, or different protocols. They might use different commercial carriers. This type of logical graph model is useful for examining the robustness of a system to the loss of one of the types of links depicted or how malware might propagate through the computer network.

**Figure 2.2. Logical Graph Model of a Computer Network**

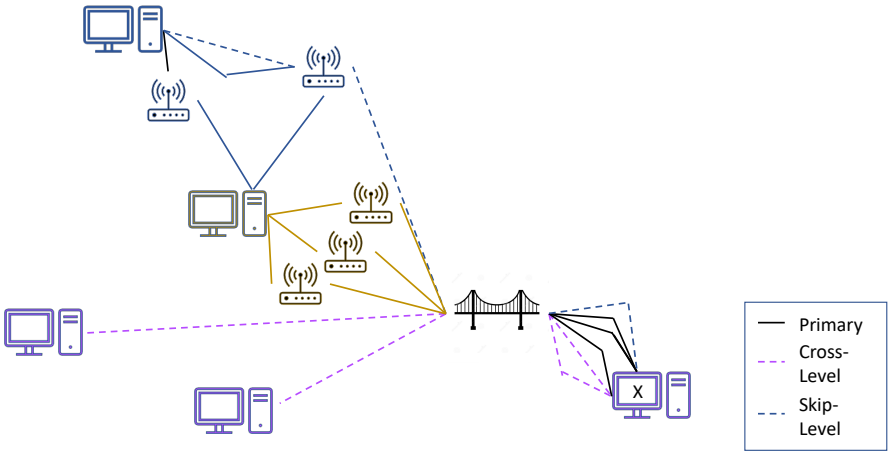



But the computer systems and the communications channels that connect them are also real, physical entities located somewhere in space.[16] Consequently they are also subject to attacks and failures associated with their physical location. In Figure 2.2, the computer on the lower right

[16] Or, in the case of wireless connections, electromagnetic waves produced and received by physical entities that traverse physical pathways.

appears to have three independent links to the computer in the center. However, suppose those three links are at some area bundled into a common cable on a bridge to cross a river. If that bridge fails, all three links fail simultaneously. This circumstance is shown in a physical graph model in Figure 2.3. This model captures the physical layout, in particular the fact that the links to the computer on the lower right all pass through a common physical conduit.

**Figure 2.3. Physical Graph Model of a Computer Network**




As an example, in July 2001, a freight train caught fire in the Howard Street Tunnel in Baltimore, Maryland. Multiple internet service providers supplied independent connections on the east coast of the United States at the time. But many used the Howard Street Tunnel as a conduit for their cables. The result was a widespread internet interruption across many suppliers. And because the tunnel was a convenient conduit, power, water, and phone connections were also interrupted. A failure to understand the physical layout and rely only on logical graph models led to surprise at the extent of the loss of services from one accident.[17] Like links, nodes too can coincide physically in ways that control interesting failure modes. If several computers are in the same building, an unmitigated power interruption to that building or a kinetic attack against that building can remove those nodes simultaneously. What appeared to be redundancy in a logical graph layout was not redundant in a physical graph layout. To model the effects of

[17] Lyndsey Layton and Don Phillips, "Train Sets Tunnel Afire, Shuts Down Baltimore," *Washington Post*, July 19, 2001, 2001; Andrew Ratner, "Train Derailment Severs Communications," *Baltimore Sun*, July 20, 2001; Hilary C. Styron, *CSX Tunnel Fire, Baltimore, Maryland*, U.S. Fire Administration, U.S. Department of Homeland Security, USFA-TR-140, July 2001; Mark R. Carter, Mark P. Howard, Nicholas Owens, David Register, Jason Kennedy, Kelley Pecheux, and Aaron Newton, *Effects of Catastrophic Events on Transportation System Management and Operations, Howard Street Tunnel Fire, Baltimore City, Maryland—July 18, 2001*, U.S. Department of Transportation, July 2002.

physical attacks and geographically correlated failures, a graph model that captures the physical and geographic layout is necessary.

These examples demonstrate that multiple graphs and multiple models may be needed to address various purposes and questions in the same context. In the example in Figures 2.2 and 2.3, if malware propagation is the problem to analyze, the logical representation of Figure 2.2 is appropriate. If the consequences of an attack against the bridge is the problem, the physical representation of Figure 2.3 is appropriate. Furthermore, the necessary elements represented in the models—both directly associated with the context and indirectly associated with the surrounding environment—differ based on the desired purpose. We have also shown that multiple graphs may be needed to represent different aspects of the same problem. Indeed, without these multiple graphs, a significant part of the problem may be inadvertently ignored to the detriment of the conclusion reached. Asking the right questions and probing beyond the surface are essential parts of formulating the right model to achieve the desired results.

### *Key Questions*

In sum, the key questions for the analyst seeking to create a graph from data are the following:

1. What system is to be modeled?
2. What questions does the analyst seek to answer about this system?
3. What relationships are necessary to capture in the model?
    a. What discrete, distinct entities are best represented as the nodes?
    b. Are the interrelationships among the entities directed or undirected?
    c. Do these interrelationships obey the binary/pairwise restrictions?

## Versatility of Graph Representations

Some systems (or processes) are more amenable to being modeled as a graph than others. Systems that are called networks are good examples—computer networks, airline route networks, and so on. But graph models can offer insight for a wide variety of settings that are less obviously networks. We previously described how the scheduling of tasks for nuclear certification, nuclear surely, and cybersecurity risk assessment can be modeled by graphs and how those graphs can be used to better manage these processes.[18] In the remaining part of this chapter, we introduce examples in which graph models can be usefully employed for nuclear certification, nuclear surety, and cybersecurity risk assessment—but in settings that are less obvious than computer networks.

[18] Don Snyder, Christian Johnson, Parousia Rockstroh, Lance Menthe, and Bart Bennett, *Graph Theoretic Algorithms for the Ground Based Strategic Deterrent Program: Prioritization and Scheduling*, RAND Corporation, RR-A583-1, 2021.

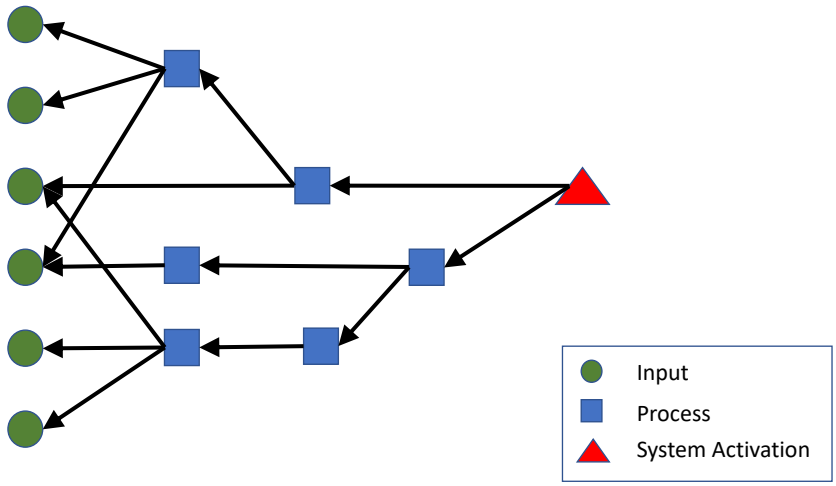
### *System Safety*

One way to mitigate the safety risk of a system operating outside of desired conditions is to require a specific set of conditions to be met before the system can operate. Such safety systems exist in a variety of applications, such as door-interlock mechanisms for microwaves, ignition systems for automobiles, and launch protocols for intercontinental ballistic missile systems, the latter being an extreme case because of the consequences of inadvertent launch. In this section, we examine a simple safety system and use a different kind of graph to examine its ability to prevent inadvertent activation of the system.

Inputs and process sequences required for safe system activation can be represented in a dependency graph, a type of directional graph or digraph. Nodes represent inputs and processes (called tasks), such as decisions made, communication received, keys turned, passwords entered, switches activated, sensors stimulated, or any number of other tasks. These graphs use arrows rather than lines for the links. Arrows point at nodes that must be completed before the node at the tail of the arrow can start. A task that first requires multiple prior tasks to complete will have arrows starting at the node and pointing at each of the dependent task nodes.

Such an example is shown in Figure 2.4. Although inputs (shown as green circles) may appear scattered throughout such a graph, we have collected them on the left. System activation is shown as a red triangle on the right. In between are nodes representing various tasks that must be initiated and completed before final activation can occur. This example demonstrates how a very different kind of problem from those discussed above can be represented as a graph. We have used additional features of graph theory, the directed arrow, and the digraph, to enlarge the set of problems that can be represented as graphs. Furthermore, when added as an additional layer to the logical and physical graphs modeling workflow in a system, the trade-offs between system effectiveness and safety—the need for the system to function when desired but not to inadvertently activate when not desired—can be examined and balanced.

**Figure 2.4. Example of a Dependency Graph for System Activation Safety**



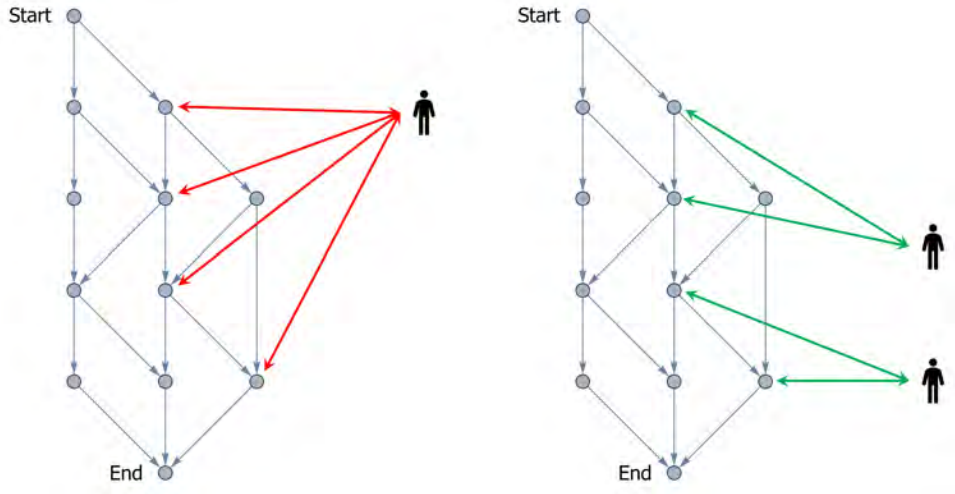
### *Insider Threat*

An insider threat is "the potential for an individual who has or had authorized access to an organization's assets to use that access, either maliciously or unintentionally, to act in a way that could negatively affect the organization."[19] It may seem unusual to suggest that insider threat be represented by a graph. Although various aspects of the insider threat problem may be better suited to other means of examination and modeling, two examples show how this class of problems can usefully be formulated as a graph. In the first, we show how the development of a computer network can be aided using a graph. In the second, we show how an insider might be detected using a graph.

#### Design

Suppose there is a process that for safety, surety, or cybersecurity reasons, leaders require more than one person to cooperate to complete. Figure 2.5 shows such a process modeled as a directed graph of tasks. Various paths are possible from the start to the end. One of these paths (sequences of tasks) from the top to the bottom must be completed for the process to execute successfully. Each task must be completed by a person authorized for the task. The graph on the left side of Figure 2.5 shows a circumstance in which a single individual has been granted authorization to perform enough of the tasks to complete the process, violating the desired two-person rule. The graph on the right side of the figure shows an acceptable authorization, such that the two individuals can complete the process as a team, but neither can complete the process alone.

**Figure 2.5. Example of Analyzing Insider Threats via Graphs**




[19] CERT National Insider Threat Center, *Common Sense Guide to Mitigating Insider Threats*, 6th ed., Carnegie Mellon University, Software Engineering Institute, Technical Note CMU/SEI-2018-TR-010, December 2018, p. ix.

The example in Figure 2.5 is simple enough that the authorizations and process execution abilities are visually evident. But in complex cases involving a large number of tasks with complicated sequencing and a large number of individuals, the analysis can be quite complicated. Analysis by hand can lead to errors and oversights. To ensure rigor and completeness, graph models can address such questions as: Can the authorized individuals execute the task under all required circumstances? How many individuals must collude to perform an unauthorized process?

### Detection

Another application of graph theory to the insider threat is the detection of unauthorized behavior.[20] Information flow, access to systems, and other behavior among individuals and systems, databases, and so forth can be modeled as a graph. The behaviors can be depicted as links, and the individuals, systems, databases, and other entities as nodes. Patterns of acceptable behavior can be established in advance, and deviations from acceptable behavior noted. Or patterns of behavior can be monitored over time, and departures from nominal behavior noted. Graphical models can assist with analysis of deviations from acceptable behavior and help identify insider threats.

**Key Points About Modeling a System as a Graph**

- For a graph to be a good representation of a system, its key characteristics being studied must be discrete, well-defined entities that have discrete, well-defined interactions.
- A system does not possess a single graph representation but rather has multiple potential graph representations depending on the problem being investigated, each giving distinct insights.
- Graphs can be useful models of some systems that are not obviously networks.

## Summary

Graphs are a powerful means for analyzing a variety of systems and systems of systems. However, abstracting to a graph representation, like any modeling formulation, should be done carefully—thoroughly enough to contain the significant elements relevant to the analytic purpose but as simple as possible. Like other forms of mathematical modeling, there is art in formulating a model that is as simple as possible yet captures the key elements of the problem being addressed. The process often involves some degree of iteration—formulating a model, doing some analysis, learning and discovering more about the model, and then refining the model.

[20] See You Chen, Steve Nyemba, Wen Zhang, and Bradley Malin, "Specializing Network Analysis to Detect Anomalous Insider Actions," *Security Informatics*, Vol. 1, No. 1, 2012. Other applications can be found in Ramkumar Chinchani, Anusha Iyer, Hung Q. Ngo, and Shambhu Upadhyaya, "Towards a Theory of Insider Threat Assessment," *Proceedings of the 2005 International Conference on Dependable Systems and Networks*, 2005; and Anagi Gamachchi, Li Sun, and Serdar Boztas, "A Graph Based Framework for Malicious Insider Threat Detection," *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

The examples provided in this chapter show only a few of the many potential applications of graph theory. Graphs reduce a system to nodes and links, but this simplicity also includes a remarkable flexibility in representation. The power of graph methods are most evident for systems of great size and complexity. In the next two chapters, we discuss concepts and methodologies in graph theory that can be used to explore and analyze a graph and quantitatively respond to decisionmaker questions.

# Chapter 3. Centrality Metrics

> There are many possible definitions of importance and there are correspondingly many centrality measures for networks.
>
> —Mark Newman[21]

It is often useful to know how important a node or link is to the overall functioning of a system. When modeled as a graph, a number of metrics are available for assessing this relative importance. A metric that assigns a scalar value to a node or link is called a *centrality metric*. The word centrality reflects the indirect effects of that node or link on the functioning of the overall system modeled by the graph. It does not mean how central the node or link might be in a visual plot of a graph. Visualizations of a graph place nodes at various locations that might not reflect their centrality in the sense of importance. Hence, depending on the visualization algorithms used, highly important nodes might or might not plot toward the center.

In this chapter, we review some of the most commonly used centrality metrics that we expect would be useful for Sentinel program analysis. These centrality metrics are also the most commonly available in software packages for graph analysis. We address the questions: When is it most useful to employ one centrality metric versus another? What meaning does each have, and what are the limitations of each?

We make no attempt to provide an exhaustive list of all metrics that have been proposed. The presentation is heuristic and often based on empirical observations rather than mathematically rigorous definitions and proofs. Given the complexity of many real graphs, few formally provable conclusions about centrality have been discovered. Many of the most useful results are empirical. When proofs are available, they add little to the general guidance for the use and interpretation of these concepts, so we omit the formality.

[21] Newman, 2018, p. 159.

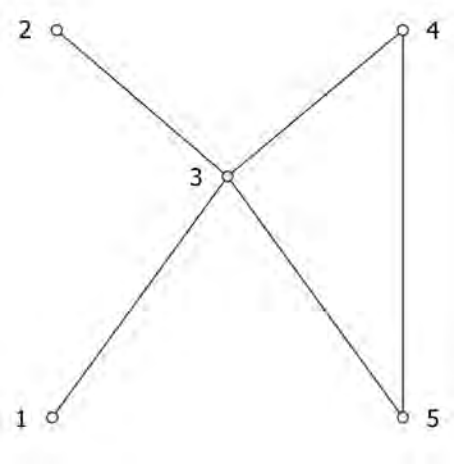
## Principal Centrality Metrics

### *Degree Centrality*

#### Description

The *degree* of a node in an undirected graph—graphs in which the links have no direction—is the number of links from that node to other nodes.[22] Degree centrality is one of the most commonly used centrality metrics. It is intuitive and easily calculated. For example, in Figure 3.1, nodes 1 and 2 both have degree one as they each have a single link connecting them to other nodes in the graph. Node 3 has the highest degree (four) because it has four links connecting it to each of the other nodes in the graph.

**Figure 3.1. Example Graph for Degree Centrality**




To express degree mathematically, we first introduce the concept of the adjacency matrix **A** of a graph, defined such that $A_{ij} = p$ indicates that nodes $i$ and $j$ are connected by $p$ links. Typically, $p \in \{0, 1\}$. For an undirected graph, the adjacency matrix is symmetric with zero diagonal (nodes are not considered to be linked to themselves).[23] The degree centrality $k_i$ for node $i$ in a graph with $n$ nodes is then given by:

$$k_i = \sum_{j=1}^{n} A_{ij}.$$

[22] Degree centrality can be extended to directed graphs, in which links have direction. In directed graphs, each node has an in-degree (the number of links pointed toward the node) and an out-degree (the number of links emanating from the node).

[23] It is permissible that graphs can contain nodes that have links to themselves, but we do not consider such graphs in this report.

We note that the magnitude of degree centrality is a function of the size of the graph as suggested by the summation index in the mathematical definition.[24] If it is desirable to have a measure that is independent of graph size, then it is appropriate to scale the metric based on graph size. In a graph with $n$ nodes, a given node $i$ can be connected to at most $n-1$ other nodes, so the maximum degree of any node is $n-1$. Thus, we can normalize the degree centrality measure by dividing by $n-1$:

$$\bar{k}_i = \frac{\sum_{j=1}^{n} A_{ij}}{n-1}.$$

The resulting measure $\bar{k}_i$ represents the *proportion* of other points to which $i$ is adjacent.

### Application

One application of degree centrality is to estimate the importance of a node in the spreading of an epidemic. The propagation of computer malware on an information technology network is an example of epidemic spreading, a topic that we treat in more detail in the next chapter. One estimate of the relative importance of a given computer in malware propagation is the square of its degree centrality. In this model, the likelihood of any computer to be infected is proportional to the number of computers it is directly connected to—its degree centrality.[25] And the likelihood that it will pass on the malware to another computer is also proportional to the number of computers to which it is directly connected—its degree centrality. Hence, a computer's relative role in malware propagation is the square of its degree centrality. This measure is intuitive, easy to quantify, quickly calculated, can be calculated for any node without knowing the full graph structure, and expresses the first-order role of epidemic spreading for some purposes.

Not all connections among computers are equally likely to transmit malware. For example, some malware might propagate via email and others via direct connections on a local area network. The concept of degree centrality can be extended to incorporate this variance if the relative probabilities are known and can be collected. To incorporate this additional level of detail, the graph model can be enhanced to add weights to the links that are proportional to the likelihood of particular malware transmission between each pair of computers, giving a weighted graph. In the adjacency matrix, instead of 0 for no interaction and 1 for interaction, the values of $p$ would reflect the relative probabilities of malware propagation. The resulting weighted degree centrality would, then, express this more nuanced view of connections.

Degree centrality can be extended to directed graphs. In this case, degree centrality has alternative implementations and interpretations because the direction of the links must be considered. That is, one must distinguish between incoming and outgoing links. Although degree

---

[24] We use the term *size* to mean the total number of nodes in a graph. In some mathematical literature, the term *order* is used for this value, and *size* is used to denote the number of links.

[25] Note that what constitutes a connection can vary with context. For a computer virus that predominantly spreads via email, the number of email addresses in a computer's contact list would be a good measure of degree centrality.

centrality in a directed graph can be calculated as the sum of indegree and outdegree, for example,

$$k_i = k_i^{\text{in}} + k_i^{\text{out}},$$

it may also be calculated as the average of indegree and outdegree

$$k_i = \frac{k_i^{\text{in}} + k_i^{\text{out}}}{2},$$

which aligns with the concept of degree in an undirected graph. One may also wish to consider the indegree and outdegree centralities separately: Nodes with a high indegree centrality are well-positioned to receive information, and those with high outdegree centrality can be viewed as key sources of information.

### Computational Complexity

Algorithms for solving a problem are typically characterized by three main types of operations: assignments (such as assigning some value to a variable), arithmetic operations (addition, subtraction, multiplication, and division), and logical operations (comparison of two values). The total number of operations performed by the algorithm determines the time it takes the algorithm to run. This computational run time is also referred to as the *computational complexity* of the algorithm. Computational complexity can be reported in one of three ways—the best-case, average-case, and worst-case computational run times. Computational complexity determines how practicable a given metric is for graphs of large size. We present complexity analysis in this report in terms of the worst case, which is denoted by $\mathcal{O}(\cdot)$.

From the definition, we see that calculation of degree centrality for each node requires summation of $n$ values, specifically the entries in the adjacency matrix indicating whether the node is connected to each of the other nodes in the graph. This calculation must be done for all $n$ nodes, thus requiring $n \times n = n^2$ operations. We denote the computational complexity of this calculation as $\mathcal{O}(n^2)$. In the case of sparse graphs, storing the graph as a list of links rather than an adjacency matrix is more efficient from both a storage and run time perspective. An algorithm that considers each link exactly once can be used to determine degree centrality in $\mathcal{O}(m)$ time, where $m$ is the number of links.

### Limitations and Additional Considerations

The degree centrality takes into account only the neighbors of a given node. It therefore reflects only the local structure of the graph in the immediate vicinity of a node and does not contain any information about the position of the node in the larger structures of the graph. For graphs in which the larger structure plays a large role in the functioning of the overall system that the graph represents, the degree centrality is a weak metric of the importance of a node to that function. We present such examples in the next chapter along with more-sophisticated metrics.

Another limitation of the degree centrality is that it assumes that all nodes are qualitatively the same. For a computer network, it means that if a computer is connected directly to five other computers of little significance to the functioning of a mission, this node is considered more important than a computer attached to just one computer, even if that computer is the central controller for the mission. This circumstance points to the need to have a measure that goes beyond merely counting the number of connections that a node has and also incorporates the relative importance of the nodes to which it is connected. To capture these deeper effects, a more-sophisticated metric is needed. The eigenvector centrality metric is one such measure.

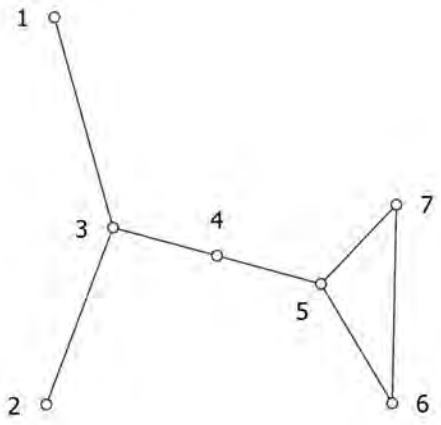
### *Eigenvector Centrality*

#### Description

The *eigenvector* centrality metric provides a measure of the influence of a node within a graph by considering not only the degree of the node but also the centrality of its adjacent nodes.[26] We consider here only undirected graphs that are *connected*, which means that for every pair of nodes in the graph, there exists a path that connects those two nodes along links in the graph. The greater the centrality of its neighbors, the greater the centrality of the node itself.

Consider the graph in Figure 3.2. Both node 3 and node 5 have degree three. However, node 5 is connected to two nodes of degree two while node 3 is connected only to nodes of degree one. Thus, node 5 has a greater eigenvector centrality than node 3.

**Figure 3.2. Example Graph for Eigenvector and Betweenness Centralities**




[26] Phillip Bonacich, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology*, Vol. 2, No. 1, 1972.

The eigenvector centrality of a node is proportional to the combined centralities of its neighbors (and the neighbors to their neighbors, and so on). Mathematically, this suggests that the eigenvector centrality $c_i$ for node $i$ in a graph with $n$ nodes can be represented by

$$c_i = \lambda^{-1} \sum_{j:(i,j) \text{ is a link}} c_j = \lambda^{-1} \sum_{j=1}^{n} A_{ij} c_j$$

where $\lambda^{-1}$ is a proportionality constant. (Here $A_{ij}$ is either 1 or 0, indicating that nodes $i$ and $j$ are either adjacent or they are not. Parallel links connecting the same two nodes are not accounted for.) This relation can equivalently be written by the matrix equation:

$$\lambda \mathbf{c} = \mathbf{A}\mathbf{c}.$$

A nonzero solution vector $\mathbf{c}$ to this equation is called an *eigenvector* of $\mathbf{A}$, and $\lambda$ is its corresponding *eigenvalue*. While an $n \times n$ matrix can have up to $n$ distinct eigenvalues, each with a corresponding eigenvector, in this context, we are interested in the eigenvector corresponding to the maximal eigenvalue because the entries of this eigenvector are guaranteed to be nonnegative.[27] An eigenvector corresponding to the maximal eigenvalue is typically referred to as the *principle eigenvector*. The values of the principle eigenvector define the relative eigenvector centralities for the nodes of a graph.[28]

### Application

As discussed earlier, the relative importance of each computer in malware propagation is dependent on those computers to which it is connected and their respective connections. A computer that is connected to other computers that are highly connected would be considered more important in the transmission of malware, should it become infected, than one that is connected to computers that are poorly connected since the malware is more likely to spread to more computers (and at a faster rate) in the former case. As a result, the malware may negatively affect a larger portion of the overall network in a shorter period of time before its detection. In the latter case, subsequent transmission may be reduced or eliminated (or at least slowed) because there are fewer paths along which transmission may occur. It is important to emphasize that eigenvector centrality reveals the *relative* importance of nodes in epidemic spreading (or any other application), but there is no absolute interpretation of this value for any computer (node).

---

[27] For a (real) nonnegative square matrix, the Perron-Frobenius theorem guarantees the existence of an eigenvector of the maximal eigenvalue that contains only nonnegative entries. See C. R. MacCluer "The Many Proofs and Applications of Perron's Theorem," *SIAM Review*, Vol. 42, No. 3, 2000.

[28] For additional discussion of eigenvector centrality metrics, see Phillip Bonacich, "Power and Centrality: A Family of Measures," *American Journal of Sociology*, Vol. 92, No. 5, 1987; and Phillip Bonacich, "Some Unique Properties of Eigenvector Centrality," *Social Networks*, Vol. 29, No. 4, 2007.

### Computational Complexity

There are a variety of algorithms for calculating the eigenvalues and eigenvectors of a matrix. Which to use depends on the attributes of the matrix. For graphs, the adjacency matrix is real and (for undirected graphs) symmetric. The calculational speed depends on the speed of convergence of the algorithm, which in turn depends on the details of the matrix, so there is no fixed calculational burden for a matrix of a given size. Given these attributes, the computational complexity of calculating all the eigenvectors of a graph to adequate precision is approximately $\mathcal{O}(n^3)$.[29]

### Limitations and Additional Considerations

A drawback of the eigenvector centrality is that most of the weight of the eigenvector is concentrated in a few nodes, particularly nodes that act as hubs.[30] Most nodes have centrality closer to zero, which makes it difficult to determine the importance of these nodes. The reason for this skewed distribution of centrality is that centrality tends to be *reflected* back to nodes—the centrality of a given node depends on the centrality of its neighbors, and their centrality depends on that of the original node.

This reflection of centrality back to nodes can lead to disproportionate accumulation of eigenvector centrality near high-degree nodes, or even within particular communities in graphs that exhibit strong community structure.[31] Graphs that exhibit strong community structure are called *modular* graphs. The topic of modularity is one of overall graph structure and therefore discussed in the next chapter. For highly modular graphs, the individual eigenvectors tend to correspond to one of the communities. Therefore, the eigenvector centrality will be biased toward one community and will not properly reflect the importance of nodes in other communities. The eigenvector centrality metric should be used with caution in highly modular graphs.[32]

---

[29] Andreas Baltz and Lasse Kliemann, "Spectral Analysis," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005.

[30] K.-I. Goh, B. Kahng, and D. Kim, "Spectra and Eigenvectors of Scale-Free Networks," *Physical Review E*, Vol. 64, No. 5, 2001.

[31] A *community* within a graph is a subset of nodes that exhibit denser connections among themselves than to the rest of the graph.

[32] Martin G. Everett and Stephen P. Borgatti, "Extending Centrality," in Peter J. Carrington, John Scott, and Stanley Wasserman, eds., *Models and Methods in Social Network Analysis*, Cambridge University Press, 2005; Dirk Koschützki, Katharina Anna Lehmann, Leon Peeters, Stefan Richter, Dagmar Tenfelde-Podehl, and Oliver Zlotowski, "Centrality Indices," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005.

In directed graphs, the condition of the Perron-Frobenius theorem[33]—that **A** be irreducible—does not necessarily hold, so the existence of a maximal eigenvalue and corresponding principle eigenvector—which are necessary to describe the centralities of nodes—is not guaranteed. For this reason, eigenvector centrality is typically used only for undirected graphs. There are variations on eigenvector centrality that are more appropriate for directed graphs (e.g., the Katz centrality).[34]

### *Betweenness Centrality*

#### Description

In contrast to degree and eigenvector centralities, which focus on the local structure surrounding a node, *betweenness* centrality takes into account where the node resides in the global structure of the graph. Betweenness centrality can be thought of as measuring the likelihood of a node acting as a bridge or a conduit between two components of a graph. That is, nodes with higher betweenness centrality are more likely to act as bridges in the graph and are thought of as having a high degree of control over traffic in the graph. (For simplicity, we discuss betweenness as a property of nodes. The concepts can be equally applied to links.) Such nodes can also be bottlenecks for the graph because a great deal of traffic flows through them. In Figure 3.2, node 4 acts as a bridge between the left and right components of the graph and controls communication between the two components. Node 4 has high betweenness centrality.

Although the concept of betweenness was originally introduced by Bavelas,[35] it was Freeman who formally defined it and provided a mathematical definition.[36] White and Borgatti expanded Freeman's concept to directed graphs.[37] The most common definition of betweenness centrality is the fraction of the shortest paths between all nodes in a graph that pass through a node of interest. In mathematical terms, if we let $i$ and $j$ denote any pair of nodes, define $\sigma_{ij}$ as the number of all shortest paths between $i$ and $j$, and define $\sigma_{ij}(k)$ as the number of shortest paths between $i$ and $j$ that pass through node $k$, then the betweenness centrality of $k$ is given by the fraction of shortest paths between $i$ and $j$ that pass through $k$:

$$c_k = \sum_{i \neq j \neq k} \frac{\sigma_{ij}(k)}{\sigma_{ij}}.$$

---

[33] The condition of the Perron-Frobenius theorem is that **A** be irreducible, and it can be shown that **A** is irreducible if and only if the associated graph is strongly connected, which means there exists a path between each pair of nodes. This condition does typically hold in (connected) directed graphs.

[34] Leo Katz, "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, Vol. 18, No. 1, 1953.

[35] Alex Bavelas, "A Mathematical Model for Group Structures," *Applied Anthropology*, Vol. 7, No. 3, 1948.

[36] Linton C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, Vol. 40, No. 1, 1977.

[37] Douglas R. White and Stephen P. Borgatti, "Betweenness Centrality Measures for Directed Graphs," *Social Networks*, Vol. 16, No. 4, 1994.

Like degree centrality, betweenness centrality is dependent on the size of the graph as suggested by the summation indices, which sum over pairs of nodes. This metric can also be normalized to a produce a metric independent of graph size by scaling by the maximum possible betweenness value. This value is equal to the maximum number of shortest paths that can pass through a node and is achieved when a given node lies on all shortest paths between all other pairs of nodes in the graph. Thus, it is equivalent to the number of node pairs, not including the node of interest itself, within the graph and is expressed as $(n-1)(n-2)/2$ for undirected graphs and $(n-1)(n-2)$ for directed graphs. [38]

### Application

Betweenness is an estimate of the importance of a node to the overall flow through a graph.[39] The main application of the betweenness centrality is to identify which nodes are choke points in the flow of an entity through the graph. In the case of an internet protocol router network, routers with high betweenness are expected to handle the highest volume of packets. Failure of these nodes might affect the flow of packets across the network more than nodes of low betweenness. As we discuss in the next chapter, this metric is useful in mitigating epidemic propagation in graphs, an example of which is the spreading of computer malware.

### Computational Complexity

Calculating the betweenness centrality metric requires the calculation of shortest paths between all pairs of nodes in a network. In particular, it requires the calculation of the number of shortest paths between all pairs of nodes that pass through a given node. Brandes developed an ingenious algorithm that is much faster than the direct calculation of all paths. For a graph with $n$ nodes and $m$ links, Brandes' algorithm runs in $\Theta(nm)$ time for unweighted graphs and in $\Theta(n^2 \log(n) + nm)$ time for weighted graphs.[40]

### Limitations and Additional Considerations

Although betweenness centrality captures the role of a node in the global structure of a graph, it has a few limitations. Many nodes do not lie on a shortest path between any two other nodes,

---

[38] For a given node, $(n-1)(n-2)$ represents the number of pairs of other nodes in the graph, taking the order of the nodes, and thus the orientation of the connecting link, into account (e.g., the node pair (1,2) and (2,1) are distinct because they represent different orientations of a link joining nodes 1 and 2). In an *undirected* graph, the order of the nodes is not important in identifying pairs of nodes (e.g., node pairs (1,2) and (2,1) are equivalent); thus, the number of unique pairs of other nodes is given by $(n-1)(n-2)/2$.

[39] For details on the exact selection of centrality metrics according to the nature of the flow in a graph, see Stephen P. Borgatti, "Centrality and Network Flow," *Social Networks*, Vol. 27, No. 1, 2005.

[40] Ulrik Brandes, "A Faster Algorithm for Betweenness Centrality," *Journal of Mathematical Sociology*, Vol. 25, No. 2, 2001; Riko Jacob, Dirk Koschützki, Katharina Anna Lehmann, Leon Peeters, and Dagmar Tenfelde-Podehl, "Algorithms for Centrality Indices," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005; Ulrik Brandes, "On Variants of Shortest-Path Betweenness Centrality and Their Generic Computation," *Social Networks*, Vol. 30, No. 2, 2008.

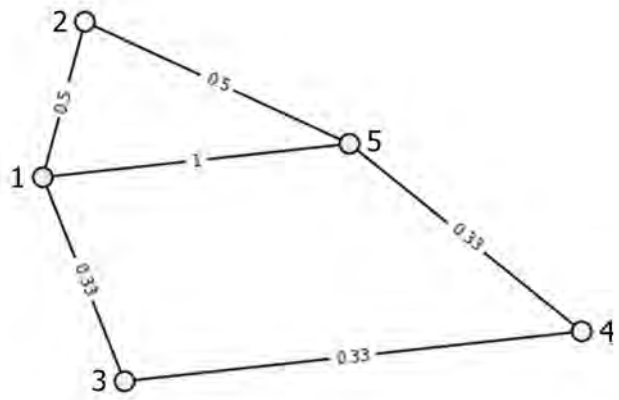
thus resulting in many nodes with a betweenness value of zero. In large graphs, such nodes can be a large proportion of all nodes. In these situations, the betweenness centrality metric may be of little use with respect to a large portion of the network and does not provide a unique ranking of nodes.

Another limitation is that the betweenness centrality assumes that flows take the shortest path in a graph. In real-world situations, this condition does not always hold.[41] When transiting from one node to another, the flowing entity might not have the ability to know what the shortest path is, and alternative paths can in some real-world circumstances be nearly as adequate as a shortest path. Packet switching in internet protocol networks is an example in which the betweenness centrality does not always capture the amount of flow through a node (or link). Nevertheless, the betweenness centrality approximation of the importance of flow through a node has proven useful in a wide variety of contexts in which shortest paths are not demonstrably characteristic of the flow phenomenon, some of which we describe in the next chapter.

A final concern is that, in some circumstances, rounding errors during calculations can introduce unwanted imprecision in the betweenness centrality. Brandes' algorithm relaxed prior assumptions that links were either present or not present and allowed for the assignment of weights to links to account for the fact that a communication between two nodes may be quicker along paths with more intermediate nodes that are strongly connected (e.g., have more-frequent contact) than paths with fewer weakly connected intermediate nodes. However, the algorithm identifies multiple shortest paths only if they have *exactly* the same distance, and a computer's representation of fractions can present issues. For example, consider the graph in Figure 3.3. There are three paths from node 1 to node 5 ($1 \rightarrow 5$, $1 \rightarrow 2 \rightarrow 5$, and $1 \rightarrow 3 \rightarrow 4 \rightarrow 5$), all of which should be considered shortest paths with a distance (weight) of one. However, while paths $1 \rightarrow 5$ and $1 \rightarrow 2 \rightarrow 5$ have exactly the same distance (weight) of one, the third path $1 \rightarrow 3 \rightarrow 4 \rightarrow 5$ does not have a distance (weight) of exactly one as a computer calculates this to be 0.3333 + 0.3333 + 0.3333 = 0.9999, so the third path is considered the only shortest path in the network.

[41] M. E. J. Newman, "A Measure of Betweenness Centrality Based on Random Walks," *Social Networks*, Vol. 27, No. 1, 2005.

**Figure 3.3. Shortest Paths in a Network**




## Supplementary Metric

### *Cutsets*

Although less commonly used and less frequently available in software packages, metrics based on cutsets of graphs have been successfully applied to mission assurance and are therefore potentially useful tools for Sentinel program analysis and related activities.[42]

#### Description

The method of cutsets can be used to identify single points of failure, pairs of nodes or links that cause failure, triplets of nodes or links that cause failure, and so on. The method emerges from one of the most central and powerful theorems of graph theory—Menger's theorem. Menger's theorem states that the number of independent paths between two nodes in a graph is exactly equal to the size of the minimum cutset of the graph.[43] To explain what that means, we need to describe cuts in graphs.

A *cut* is a partition of a graph into two disconnected components. Associated with each cut is a *cutset*, which can be a set of links (*link cutset*) or nodes (*node cutset*) whose removal from the graph results in the disconnected components. A *minimum cutset* is the minimum set of links (or nodes) that must be removed to result in disconnected components.[44] The terms *link connectivity*

---

[42] Snyder et al., 2022.

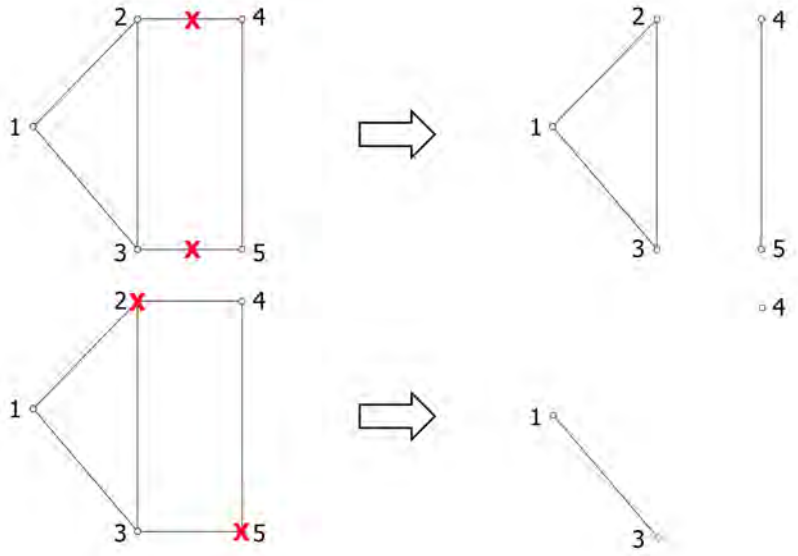
[43] Béla Bollobás, *Modern Graph Theory*, Springer, 1998, p. 75; Reinhard Diestel, *Graph Theory*, 5th ed., Springer, 2017, p. 67.

[44] Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Inc., 1993, pp. 27–28.

and *node connectivity* are often used to refer to the size of the minimum cutset and are measures of a graph's robustness to link (or node) removal.

The graphs in Figure 3.4 illustrate the concept of a cutset. In the top graph, the removal of the links connecting nodes 2 and 4 (typically denoted by (2,4)) and nodes 3 and 5 (denoted as (3,5)) separate the graph into two components—nodes 1, 2, and 3 are separated from nodes 4 and 5. In the bottom graph, the removal of nodes 2 and 5 disconnect the graphs into separate components. Both cases reflect a minimum cutset because it is not possible to separate the graph into two disconnected components via the removal of either a single link or a single node.

**Figure 3.4. Examples of Link and Node Cutsets**




In many graphs, the connection between two nodes is vitally important, for example, the start and end of a process. Which nodes (functions) or links (connections) are most critical to completing the process? If we denote the nodes that must remain connected $i$ and $j$, a cutset is the set of links (nodes) whose removal completely separates node $i$ from node $j$ so that there is no path between them. Such a cutset is called an $i-j$ *cutset*. In Figure 3.4, the set of links {(2,4), (3,5)} is a 1–4 link cutset because the removal of those two links separate nodes 1 and 4. Similarly, the node set {2,5} is a 1–4 node cutset because its removal from the network also separates nodes 1 and 4.

### Application

The method of cutsets is particularly useful in two settings: (1) when it is desired to quantify the number of independent, redundant paths between two nodes in a graph and (2) when it is useful to bin nodes (links) into those that are single points of failure, those that are pairs that cause failure upon removal, triplets that cause failure upon removal, and so on. Said another

way, the first cutset gives all nodes (links) that are single points of failure. The second cutset gives all pairs of nodes (links) that cause failure upon removal. The third cutset gives all triplets of nodes (links) that cause failure upon removal, and so on, as far as desired or as can be specified. The cutsets provide a ranking of nodes (links) for criticality to the graph connecting the start and end nodes.

As mentioned, a common setting for using the cutset method is to analyze a process with unique starting and ending points on a graph. Nodes and links between the start and completion represent individual functions or tasks (the nodes) and their interactions or sequencing (the links). Cutsets have been applied to the problem of assessing mission assurance. The approach depicted missions graphically by decomposing missions using mission thread analysis. In the context of cybersecurity, the decomposition can also be done via Functional Mission Analysis-Cyber.[45] By using the cutset method, each node was ranked in tiers to identify most critical sets of functions.[46] Critical sets of functions could then be supplemented via redundancy or defended with higher priority than other functions.

### Computational Complexity

For a graph of $n$ nodes and $m$ links, the minimum node cutsets can be found in approximately $\mathcal{O}(n \log^2 n)$ time[47]—a recent improvement on the previously best and commonly used algorithm by Karger that ran in $\mathcal{O}(m \log^3 n)$.[48]

### Limitations and Additional Considerations

One of the limitations of node cutset analysis is that frequency (e.g., of the activity occurrence represented by the node) is not taken into account. Thus, cutset analysis may identify a node as highly critical (e.g., a single point of failure) even though the corresponding activity rarely occurs and, thus, rarely has the potential to impede (mission or project) completion. With regard to link cutset analysis, such frequency considerations can be captured via link capacities and accounted for in the analysis.

Cutset analysis can also be the victim of combinatorial explosion depending on the structure of the graph, which can present issues, particularly, with storage. Consider the graph in Figure

---

[45] William Young and Nancy G. Leveson, "An Integrated Approach to Safety and Security Based on Systems Theory: Applying a More Powerful New Safety Methodology to Security Risks," *Communications of the ACM*, Vol. 57, No. 2, 2014.

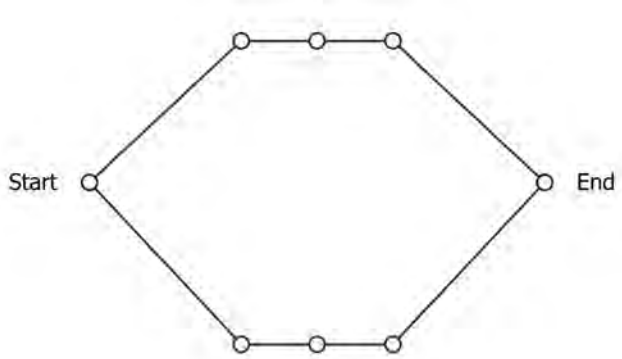
[46] For a full discussion of this example, see Snyder et al., 2022.

[47] Parinya Chalermsook, Jittat Fakcharoenphol, and Danupon Nanongkai, "A Deterministic Near-Linear Time Algorithm for Finding Minimum Cuts in Planar Graphs," *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, January 2004. This problem is one of active research. Malinowski (2016) proposed a new algorithm, but the exact computational speed is not specified (Jacek Malinowski "A New Efficient Algorithm Generating All Minimal S-T Cut-Sets in a Graph-Modeled Network," *AIP Conference Proceedings*, Vol. 1738, No. 1, 2016).

[48] David R. Karger, "Minimum Cuts in Near-Linear Time," *Journal of the ACM*, Vol. 47, No. 1, 2000.

3.5 that consists of two parallel paths representing two ways in which a project (mission) can successfully be completed. Approaching this from a node cutset perspective, in order for the project (mission) to be disrupted, one activity from each path must be disrupted. Here, there are three activities along each path that could be disrupted, which implies a total of $3^2 = 9$ different combinations of tasks that could be interrupted (or targeted) to prevent completion. As the number of paths and nodes increases so, too, does the number of (minimum) cutsets. While determining the size of such a cutset is possible, identifying (and storing) all of the different cutsets can be problematic.

**Figure 3.5. Example Demonstrating the Combinatorial Nature of Cutsets**




## Summary of Centrality Metrics

Figure 3.6 provides a visual summary of the centrality metrics that have been presented. The figure highlights the fact that different centrality metrics express different ways in which a node might be considered important. The node with highest degree has, by definition, the most connections to other nodes. In that sense, it is the most important. But its position near the periphery of the graph reduces its overall role in flow through the graph, and the nodes to which it is connected are not themselves very highly connected. The node with highest betweenness sits centrally in the flow among many nodes in the graph, particularly flows that connect pairs of nodes that are distant from one another on the periphery of the graph. The node with highest eigenvector centrality does not have the most connections to other nodes, but the nodes to which it is connected are relatively highly connected. In this sense, the node with highest eigenvector centrality resides at a critically important position in the graph. This graph illustrates that different centrality metrics often identify and emphasize different nodes as important within a graph, so it is necessary to understand what each metric tells us about a node's importance to ensure proper selection in a given context.

Table 3.1 provides a summary of the key features of the centrality metrics discussed. Recall that degree centrality considers only the immediate environment of a node. The eigenvector centrality takes into account the environment of its neighbors, and so on, but the effect of distant

nodes on the metric is very small. The betweenness and cutsets/connectivity, however, consider the global network structure in which the node is embedded. While most centrality metrics can be applied to both directed and undirected graphs, eigenvector centrality is most appropriate for undirected graphs. However, variations exist for directed cyclic networks.

**Figure 3.6. Graphical Summary of Centrality Metrics**

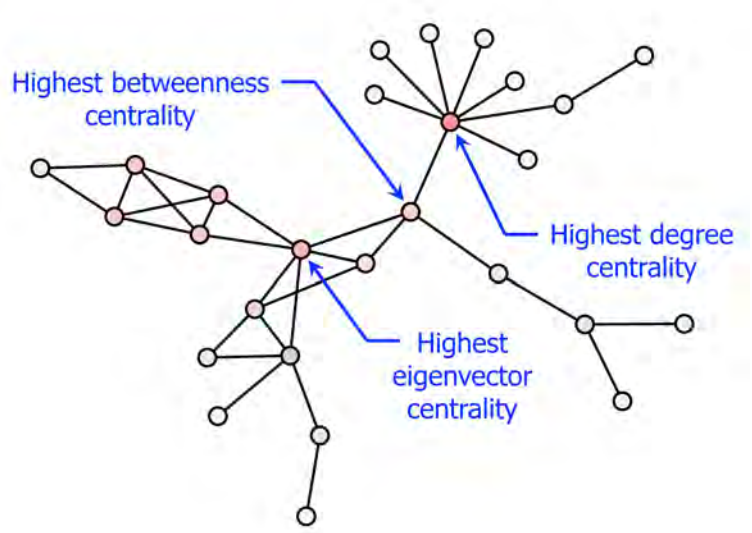



**Table 3.1. Summary of Centrality Metrics**

| Metric | What It Tells Us | Level | Network Type | Run Time |
|---|---|---|---|---|
| Degree | How many direct connections a given node has to others | Local | Directed or undirected | $\mathcal{O}(n^2)$ |
| Eigenvector | The level of influence a node may have over the entire network | Mostly local | Undirected[a] | $\mathcal{O}(n^3)$ |
| Betweenness | Which nodes are bridges between nodes or components | Global | Directed or undirected | $\Theta(n^2 \log n + nm)$ weighted graph $\Theta(nm)$ unweighted graph |
| Cutsets/Connectivity | Which nodes, upon removal, reduce the number of redundant paths | Global | Directed or undirected | $\mathcal{O}(n \log^2 n)$ |

[a] Variations exist for directed, cyclic networks. Eigenvector centrality is not applicable to directed, acyclic networks.

With respect to complexity, Table 3.1 reveals that the run time of the centrality algorithms is heavily influenced by the number of nodes in the network. In fact, both degree and eigenvector centrality scale only with the number of nodes and are not explicitly dependent on the number of

links in the graph.[49] Computational complexity of both betweenness and cutsets/connectivity measures is affected by the number of links in the graph, which is reasonable since both of these metrics consider the global structure of the graph as defined by links and paths. While betweenness can run in $\mathcal{O}(nm)$ time in unweighted graphs, the density of the graph will affect efficiency; dense graphs (i.e., $m \sim n^2$) will exhibit run times closer to $\mathcal{O}(n^3)$.

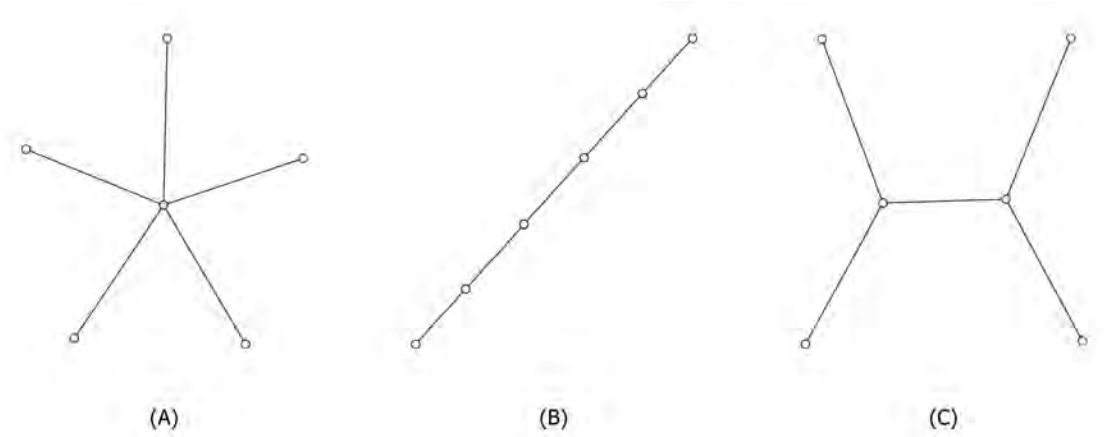
[49] Recall that, in sparse graphs, it is more efficient to represent the graph as a list of links, so the computational complexity of degree centrality is dependent only on the number of links.

# Chapter 4. Topological Measures

> One of the problems in dealing with a concept as general and inclusive as that of a structure is that no single example can suggest more than a fragment of the full concept, so any good example is in danger of being perceived as more representative than it possibly can be.
>
> —Charles E. Rickart[50]

Glancing at a small graph, it is evident that the overall structure of the graph matters. In Figure 4.1, three graphs are depicted, all with six nodes and five links. Each has obviously different structures, and those structures heavily influence connectivity and flow. Hence the structure is important in determining what failure modes the graph is robust against. The graph on the left (A) is in the form of a star. The center node is necessarily involved in the connection or flow between any other nodes. The center node is unique and of high importance by any measure. All other nodes are structurally identical to one another. The middle graph (B) is in the form of a line, a formal sequence. The two nodes at the end are structurally identical, and the nodes in between them are all structurally identical to one another. The graph on the right (C) is a slightly more complicated structure; two central nodes act as a collective core to a star.

**Figure 4.1. Simple Graph Structures**




For small graphs like these, the structures and how they control the behavior of the system that the graph models are generally easily recognizable. But for graphs with large size and complicated patterns of interconnections, these structural features become very difficult to discern visually. Furthermore, the structures that the human eye is drawn to can often be biased

[50] Charles E. Rickart, *Structuralism and Structures: A Mathematical Perspective*, World Scientific, 1995, p. 11.

by the plotting algorithm used to draw the graph, which is shown visually in the next section. Measures are needed for expressing the structures of graphs, called *topologies*, as well as what they mean with respect to robustness.

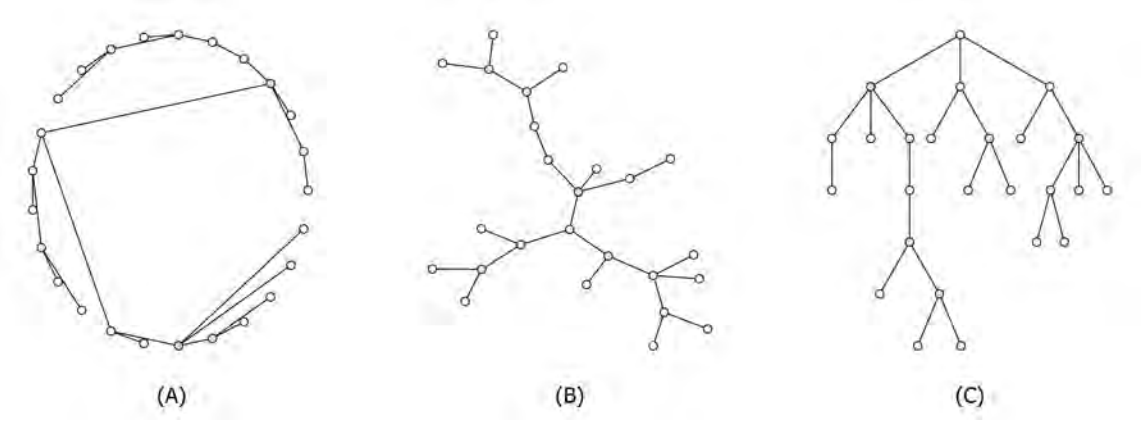
## Individual Topological Measures

### *The Fragility of Trees*

The first topological category of graphs that we discuss is an intuitive one: *trees*. To define a tree, we need to first define a few terms. A *path* in a graph is a sequence of links connecting two nodes. The *length* of a path is the number of links the path traverses.[51] A path of length greater than or equal to three that has the same node as the starting and ending point of the path is called a *cycle*. A tree is a connected graph with no cycles.[52] That is, no path of any length forms a loop from any node back to that same node in the graph.

Figure 4.2 shows a tree with 25 nodes using three different plotting algorithms. All three are commonly used plotting algorithms, and all three depictions in Figure 4.2 are of the exact same graph and therefore have identical topology. The plotting algorithm used to generate the one on the left (A) places all nodes on a circle. Although useful in some contexts, this plotting algorithm obscures the branching structure that is characteristic of trees. The one in the middle (B) uses a common algorithm called *spring embedding*. The branching structure is more evident than in (A), but the method is not tailored to illustrate the branching nature of the graph. Only the one on the right (C) is designed to emphasize the tree (branching) structure.

**Figure 4.2. An Example of a Graph with Tree Topology**




[51] If the graph has weighted links, the length is equal to the sum of the weights along a path.

[52] A *connected graph* is one in which every node is connected to every other node by some path. Put another way, no nodes or sets of nodes are isolated with no links to the rest of the graph.

Trees often occur as representations of branching processes that always go forward in time. Process flows are therefore often trees. Data buses often take the approximate structure of a tree. The graph illustrated in Figure 4.2 could represent a sequence of process assignments and dependencies in an organization. The node at the top of representation C could represent an office assigning tasks to three subordinate offices, shown on the next row of nodes down. Reading downward, each row of nodes could depict the dependency of a task on the row above. Each branch completes the full task when it ends in a node of degree 1 (nodes of degree 1 in a tree are appropriately called *leaves*). The full set of tasks represented by the graph are complete when all leaves are complete.

For our purposes, processes that can be well modeled by a tree are not *structurally* robust. It is a fundamental result of graph theory that every tree with $n$ nodes has exactly $n - 1$ links.[53] Therefore, any pair of nodes in a tree is connected by exactly one path.[54] This lack of path redundancy renders the topology fragile. As can be seen in Figure 4.2, not every plot of a graph will reveal this fragility, and therefore, checking for tree structure is insightful for learning whether a system represented by a graph is devoid of redundant paths.

### *The Failure Modes of Complex Graphs*

Trees are a very specific graph topology that occur in a few specific contexts. More commonly, large, complex graphs have richer connection patterns. It was a revolutionary discovery in 2000 that the topology of complex graphs has profound control over how they fail and, therefore, the nature of their robustness. By *failure* we mean that the phenomenon represented by the graph abruptly drops in performance. If the graph represents connections among computers, failure might entail an abrupt drop in the ability for two computers in the network to communicate, which means that numerous nodes are isolated from other nodes. The defining topological characteristic that governs robustness that has been empirically investigated is the probability distribution function of the degree centrality of the nodes. The *degree distribution* $P(k)$ of an undirected graph is the probability distribution composed of the probabilities that any randomly selected node has degree $k$.

Two degree distributions have been carefully studied in the literature. For a heuristic definition, the first topology is a class of graphs in which nodes are connected to other nodes with equal probability.[55] These graphs are called random graphs or Erdős-Rényi graphs, named after two pioneers in graph theory who defined them and derived foundational proofs about them.[56] Erdős-Rényi graphs have a Poisson degree distribution, which becomes symmetric about

---

[53] Bollobás, 1998, p. 11, Corollary 8.

[54] Diestel, 2017, p. 14, Theorem 1.5.1.

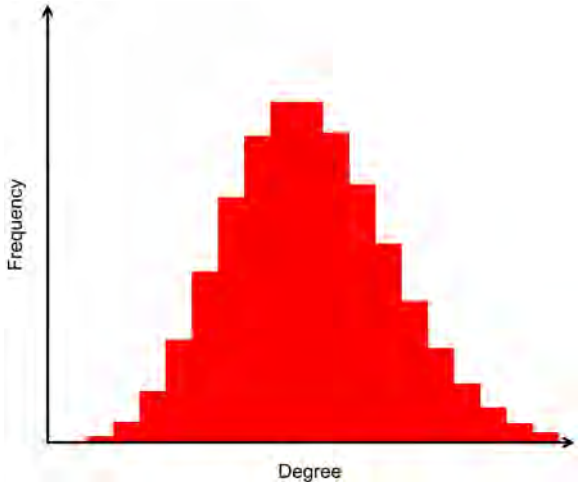
[55] In the mathematical literature, these topologies are often defined as ensemble averages, but this precision is not needed for the purpose of this report.

[56] P. Erdős and A. Rényi, "On Random Graphs I," *Publicationes Mathematicae (Debrecen)*, Vol. 6, 1959.

the mean node degree as the graph size grows large. A typical example of the degree distribution for an Erdős-Rényi graph is shown in Figure 4.3.

As can be seen in Figure 4.3, there is a typical type of node, and nodes of that type are the most common and have an average degree centrality. Because of the symmetry, the number of nodes of very low degree centrality is about the same as the number of nodes with very high degree centrality.

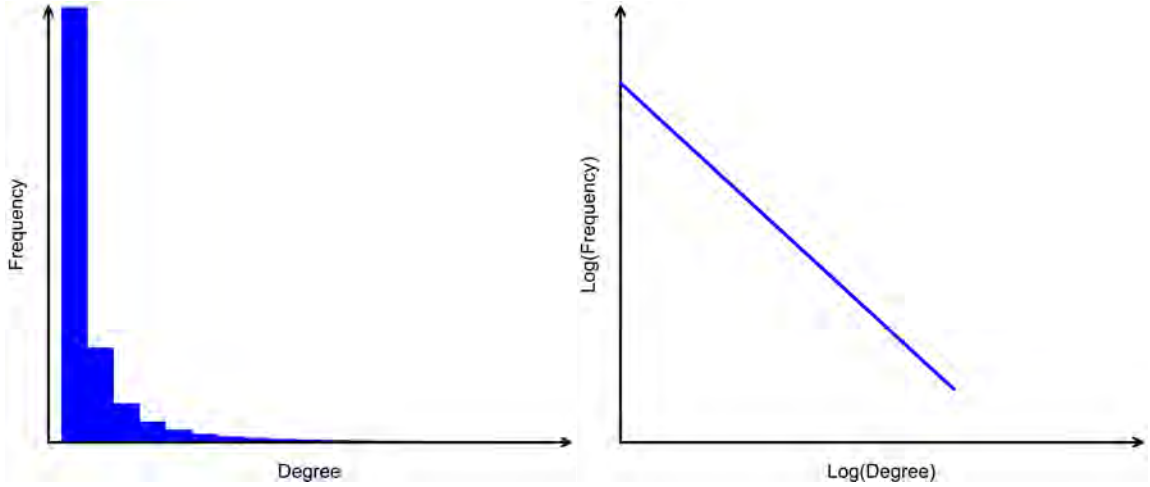
**Figure 4.3. Degree Distribution of an Erdős-Rényi Topology**




The second topology of interest is one in which the degree distribution is described by a power law. One way in which a power-law degree distribution can arise is when the probability of the attachment of new nodes to a graph is proportional to the degree of existing nodes in the graph. Hence, highly connected nodes get more nodes connected to them over time relative to poorly connected nodes. These graphs are often called *scale-free* or *fat-tailed* graphs. Figure 4.4 shows the degree distribution of a typical scale-free graph. There is an enormous number of nodes with very low degree and very few nodes of extremely high degree, as seen in the histogram on the left of the figure. Because the degree distribution is governed by a power law, a plot of the logarithm of frequency against the logarithm of degree yields a straight line, as shown on the right of the figure.[57]

[57] The statistical distribution of centrality metrics other than degree centrality can also be used to characterize topology. Betweenness centrality has also been explored, but if a graph has a power-law distribution for degree centrality, it also has a power-law distribution for betweenness centrality. Because degree centrality is so simple to compute and to use to generate synthetic graphs, it is nearly ubiquitously used. See M. Barthélemy, "Betweenness Centrality in Large Complex Networks," *The European Physical Journal B*, Vol. 38, 2004.

**Figure 4.4. Degree Distribution of a Scale-Free Topology**



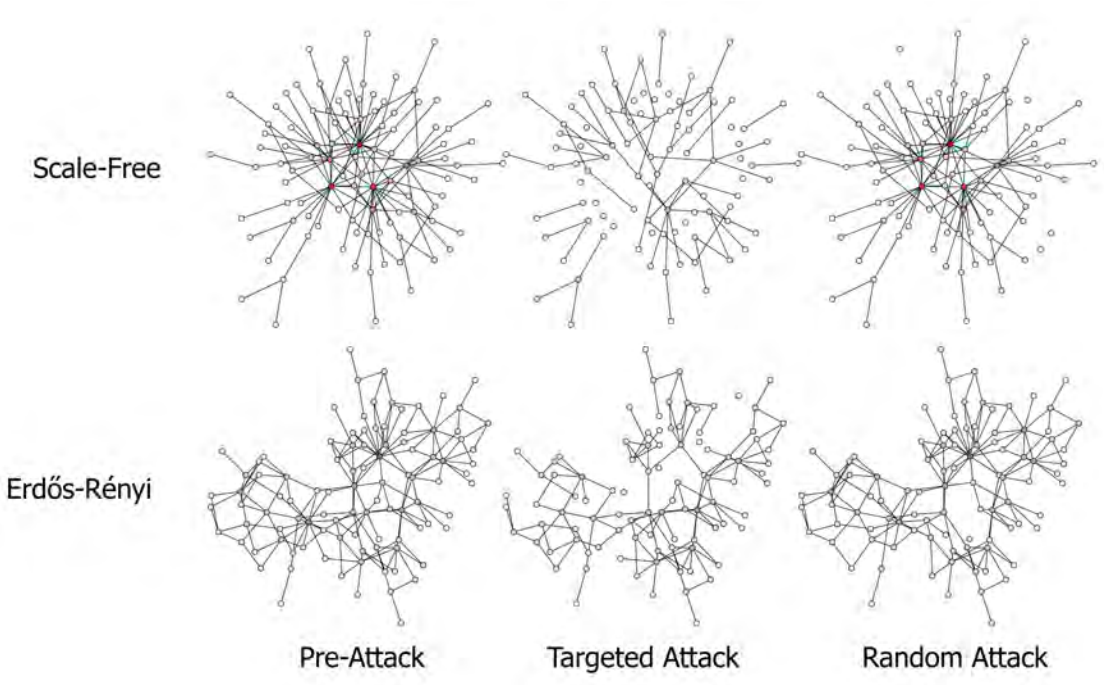

Erdős-Rényi and scale-free graphs fail in different ways upon the loss of nodes depending on attack type. The difference of interest is in how each topology fails when under targeted versus random attacks. A targeted attack removes the nodes in order of importance (by some centrality metric). A random attack removes nodes randomly, targeting all nodes with equal probability.

Nodes of average degree dominate the behavior of Erdős-Rényi graphs. A random attack will preferentially hit the nodes of the most abundant degree, which are the average nodes. A targeted attack will also focus on those same nodes since they are the most central to the connectivity of the graph. Hence, Erdős-Rényi graphs fail roughly the same for targeted and random attacks.

In contrast, scale-free topologies fail quite differently when subjected to targeted and random attacks. The most-important nodes are those of very high degree. Since there are proportionately few of them, they are unlikely to be hit by random attacks.[58] But the few high-degree nodes in a scale-free network are extremely important to the network, so targeted attacks against them are highly effective at disconnecting the graph. Hence, whereas random and targeted attacks have (roughly) similar effects in disconnecting a Erdős-Rényi graph, they are profoundly different scale-free graphs, and those differences increase as the size of a scale-free graph grows. Figure 4.5 illustrates these points for two synthetically generated graphs.

[58] Réka Albert, Hawoong Jeong, and Albert-László Barabási, "Error and Attack Tolerance of Complex Networks," *Nature*, Vol. 406, No. 6794, 2000; Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs," *Physical Review Letters*, Vol. 85, No. 25, 2000; Reuven Cohen, Keren Erez, Daniel ben-Avraham, and Shlomo Havlin, "Resilience of the Internet to Random Breakdowns," *Physical Review Letters*, Vol. 85, No. 21, 2000; and Reuven Cohen, Keren Erez, Daniel ben-Avraham, and Shlomo Havlin, "Breakdown of the Internet Under Intentional Attack," *Physical Review Letters*, Vol. 86, No. 16, 2001.

**Figure 4.5. Topological Determination of Failure Modes**




NOTE: The scale-free graph has a heterogeneity parameter of 7.95 with 100 nodes and 137 links. The Erdős-Rényi graph has a heterogeneity parameter of 5.18 with 100 nodes and 177 links. Even with graphs of just 100 nodes and a small difference in the heterogeneity parameter, the topological control of failure modes is visually evident. The different topological response increases with both graph size and difference in heterogeneity parameter.

The graphs in the upper row are scale free and those in the lower row are Erdős-Rényi. The graphs of both types are shown in the left column in their preattack states. The networks are the same size and have the same number of links. Only the topology differs. The middle column shows the graphs under targeted attack with the removal of the five nodes of highest degree. The right column shows the graphs under random attack, with the removal of five randomly selected nodes. In this illustrative example, it can be seen that the Erdős-Rényi graph fails similarly against both attack types, whereas the scale-free graph is robust against random attacks (failures) but susceptible to targeted attacks.

For measuring robustness, the key question is whether a real graph displays either of these failure characteristics or somewhere in between. Real graphs often do not show ideal Poisson or power-law degree distributions, either because they are neither of these ideals or the nature of the degree distribution is obscured by finite graph size, noise, or truncation effects.[59] We seek a measure that captures where a real graph (and the system it models) lies between these two end members.

[59] See, for example, Gavin S. Hartnett, Edward Parker, Timothy R. Gulden, Raffaele Vardavas, and David Kravitz, "Modelling the Impact of Social Distancing and Targeted Vaccination on the Spread of COVID-19 Through a Real City-Scale Contact Network," *Journal of Complex Networks*, Vol. 9, No. 6, 2021.

Whether the degree distribution is exactly Poisson or power-law is not what is important. What is important is how tightly the degree is distributed around the mean. The more tightly the degree is distributed, the more it is said to be *homogeneous*. The more fat-tailed, the more the degree distribution is said to be *heterogeneous*. A useful measure for robustness of the topology along these lines is its heterogeneity of the degree distribution. A parameter from statistics that measures the position in the range between homogeneous (e.g., Erdős-Rényi) and heterogeneous (e.g., power-law or fat-tailed) topologies is the ratio of the second to the first moments of the degree distribution, called the *heterogeneity parameter*:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\sum_k k^2 P(k)}{\sum_k kP(k)}.$$

For homogeneous graphs, $\kappa \sim \langle k \rangle$; for heterogeneous graphs, $\kappa \to \infty$ as the number of nodes in the graph $n \to \infty$. Therefore, a heuristic definition of heterogeneous graphs is those graphs with $\kappa \gg \langle k \rangle$.[60] Heterogeneous graphs are expected to be robust to random failure but susceptible to targeted attacks. Homogeneous graphs with $\kappa \sim \langle k \rangle$ are expected to fail similarly for random and targeted attacks. The designer can adjust graph topology to tailor the robustness to the desired heterogeneity characteristics.

For the behavior of random removal of nodes, there is a simple relationship between the heterogeneity parameter and the failure characteristics. The fraction of randomly selected nodes $f_c$ that must be removed to shatter a graph into small disconnected components can be estimated by[61]

$$f_c = 1 - \frac{1}{\kappa - 1}.$$

This relationship is plotted in Figure 4.6 and is a measure of robustness of graphs to random attacks. As $\kappa \to \infty$, $f_c \to 1$, which means that all nodes must be removed to shatter the graph. For finite values of the heterogeneity parameter, $f_c < 1$, and a finite number of nodes removed will shatter the graph.[62]

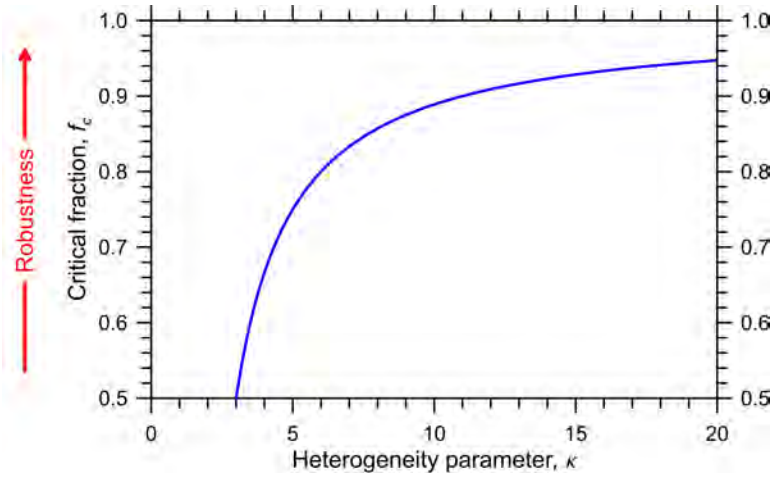
---

[60] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani, "Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks," *Physical Review Letters*, Vol. 92, No. 17, 2004; Alain Barrat Marc Barthélemy, and Alessandro Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008, pp. 37–43; Rinku Jacob, K. P. Harikrishnan, R. Misra, and G. Ambika, "Measure for Degree Heterogeneity in Complex Networks and Its Application to Recurrence Network Analysis," *Royal Society Open Science*, Vol. 4, No. 1, 2017; Keith M. Smith and Javier Escudero, "Normalised Degree Variance," *Applied Network Science*, Vol. 5, No. 1, 2020.

[61] In the physics literature, $f_c$ is called the *percolation threshold* of the graph and is defined as the transition when the giant component of the graph no longer scales linearly with the size of the graph.

[62] Barrat, Barthélemy, and Vespignani, 2008, pp. 126−129. See also Nicole Balashov, Reuven Cohen, Avieli Haber, Michael Krivelevich, and Simi Haber, "Optimal Shattering of Complex Networks," *Applied Network Science*, Vol. 4, No. 1, 2019.

**Figure 4.6. The Fraction of Nodes That Must Be Removed to Shatter a Graph**




The result in Figure 4.6 is strictly valid only for networks in which the connections among nodes of highest degree are not correlated. One important measure of correlation within graphs is the degree of *assortative mixing*. *Assortative* graphs—ones in which nodes are preferentially connected to other nodes of the same degree—are known to be more robust to node failure than disassortative graphs.[63] (*Disassortative graphs*—ones in which nodes are preferentially connected to dissimilar nodes—are empirically less common than assortative or uncorrelated graphs.) Empirical studies have shown that technical networks tend to be slightly disassortative, whereas social networks tend to be assortative (i.e., well-connected people tend to associate with other well-connected people, and hermits preferentially associate with hermits).[64] Although assortative mixing renders graphs slightly more robust to node failure, high assortative mixing in the pre-attack topology of graphs can impair the graph's functionality relative to topologies with lesser assortativity, but the degree of that impairment depends on the exact functionality needed from the system that the graph represents.[65]

[63] Assortativity is most commonly measured by the Pearson correlation coefficient, which is negative for disassortative graphs, positive for assortative graphs, and zero for uncorrelated graphs (M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, Vol. 89, No. 20, 2002).

[64] Newman, 2002; Rogier Noldus and Piet Van Mieghem, "Assortativity in Complex Networks," *Journal of Complex Networks*, Vol. 3, No. 4, 2015.

[65] For additional details, see Masaya Murakami, Shu Ishikura, Daichi Kominami, Tetsuya Shimokawa, and Masayuki Murata, "Robustness and Efficiency in Interconnected Networks with Changes in Network Assortativity," *Applied Network Science*, Vol. 2, No. 1, 2017.

**Key Points About the Failure Modes of Complex Graphs**

The heterogeneity parameter κ is a measure of how a graph responds to random and targeted attacks.

- Graphs with $\kappa \sim \langle k \rangle$ are expected to fail similarly for random and targeted attacks.
- Graphs with $\kappa \gg \langle k \rangle$ are expected to be robust against random attacks (or failures) but susceptible to targeted attacks.

### *Virus Propagation on Graphs*

Other attributes of interest are how fast an entity can propagate through a graph and, if mitigation strategies are put in place to retard the flow of that entity, whether and how the propagation can be restrained. This topic often falls under the rubric *epidemic spreading* because the topic was first investigated in the context of the spreading of contagious disease. The mathematics of epidemic spreading applies to a number of phenomena beyond the spreading of disease, including the spreading of rumors, how an overload in a power system can lead to cascading failures, and the contamination of computer systems by malware. Like robustness, epidemic spreading depends strongly on graph topology, specifically the graph's heterogeneity.

Exactly how an entity propagates depends on the detailed characteristics of the entity being modeled. For example, models for the spreading of disease focus on whether an infected person, after recovery, becomes immune and can therefore no longer transmit the disease and, if so, how long that recovery takes and how long immunity lasts. Models for the propagation of computer malware take a similar frame but focus more on whether computers of specific configurations are susceptible, whether patches have been installed (if relevant), and whether human behavior plays a role. Quantitative understanding of epidemic spreading is, therefore, highly contextual.

Nevertheless, some insightful generalizations can be made about the role of graph topology.[66] The very attributes of heterogeneous graphs that make them robust to random failures also make them excellent for the spreading and diffusion of biological and computer viruses. One key observation relates to the fraction of nodes in the graph that must be immunized to stop the further propagation of a virus, called the *immunization threshold*. This threshold is considerably lower for homogeneous graphs than heterogeneous ones. For fully heterogeneous graphs of large size, nearly all the nodes must be immunized to halt spreading.[67] A second observation is that the time to infect nodes in the graph is inversely proportional to the heterogeneity parameter.[68] This

---

[66] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani, "Epidemic Processes in Complex Networks," *Reviews of Modern Physics*, Vol. 87, No. 3, 2015, is an excellent, thorough review.

[67] Romualdo Pastor-Satorras and Alessandro Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, Vol. 86, No. 14, 2001.

[68] Barthélemy et al., 2004; Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani, "Dynamical Patterns of Epidemic Outbreaks in Complex Heterogeneous Networks," *Journal of Theoretical Biology*, Vol. 235, No. 2, 2005.

result means that, for highly heterogeneous networks of large size, the time to infect all the nodes in the graph approaches zero. The situation is much more manageable for homogeneous graphs.

At the same time that the topology of heterogeneous graphs promotes viral propagation, it also provides fruitful avenues for immunization strategies. First consider homogeneous graphs. For these graphs, the nodes of average degree centrality are also the most-abundant and most-important nodes for viral transmission. Since random immunization will preferentially select nodes of average degree, a strategy of immunizing randomly selected nodes in a homogeneous graph is the best strategy, although not a highly effective one.

A random immunization strategy fails, however, for heterogeneous graphs. In heterogeneous graphs, nodes of average degree play no special role in the functioning of the system the graph models. It is the nodes of the highest degree that dominate the propagation of viruses, and these are unlikely to be selected at random. An effective immunization strategy must target the most-important nodes in the graph, which is to say, the nodes with the highest centrality metrics. The more that is known of the graph structure, the better the selection can be made of the most appropriate centrality metric.[69] The most direct and easily calculated centrality metric is the degree centrality. As argued in the previous chapter, an estimate of the role that a node plays in epidemic spreading is roughly the degree centrality squared. Targeting nodes with high degree centrality has been shown empirically to be an effective immunization strategy on uncorrelated, heterogeneous networks.[70]

But degree centrality reflects only the local situation of a node. Epidemic spreading is a flow in the graph and consequently depends on the global structure in which a node is embedded, not just its local environment. The betweenness centrality captures how a node's position within the overall graph structure affects its role in flow, making it a stronger choice for targeted immunization than degree centrality. Targeting nodes with high betweenness centrality has been empirically shown to generally be a better immunization strategy on heterogeneous graphs than degree centrality.[71] Another observation is that the importance of a node for epidemic spreading is not just related to its connectivity but also to whether it sits in the core or periphery of the graph. Therefore, the centrality metric called the k-*core* (or k-*shell*) metric identifies nodes of high importance for epidemic spreading.[72] The drawback of betweenness and *k*-core centralities

---

[69] Pastor-Satorras et al., 2015.

[70] Romualdo Pastor-Satorras and Alessandro Vespignani, "Immunization of Complex Networks," *Physical Review E*, Vol. 65, No. 3, 2002; Zoltán Dezső and Albert-László Barabási, "Halting Viruses in Scale-Free Networks," *Physical Review E*, Vol. 65, No. 5, 2002.

[71] Chao Gao, Jiming Liu, and Ning Zhong, "Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis," *Knowledge and Information Systems*, Vol. 27, No. 2, 2011. See also Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han, "Attack Vulnerability of Complex Networks," *Physical Review E*, Vol. 65, No. 5, 2002.

[72] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse, "Identification of Influential Spreaders in Complex Networks," *Nature Physics*, Vol. 6, No. 11, 2010.

is the need to know the detailed structure of the full graph and the computational time, especially for the betweenness centrality.

Even when the full details of graph structure are unknown, if a graph is known or suspected to be highly heterogeneous, strategies can be pursued that exploit the heterogeneity of the graph to find the most-important nodes. Some ingenious methods have been devised for doing so along the lines of using selective immunization of nodes via biased sampling methods.[73] Although a program office is likely to know the details of any system architecture, some cases in which the architecture changes rapidly over time or the details are proprietary might call for employing one or more of these biased sampling methods.

The graph topology, therefore, provides information on how a computer virus, for example, will propagate through a network. How fast will it propagate through a system? Is the modularity of the graph sufficient to contain viruses to acceptable enclaves? What components should be prioritized for monitoring or patching? How fast does patching need to respond to halt virus spreading? Does the topology need to be adjusted to produce more-acceptable outcomes? These questions can be answered using graph theoretic tools.

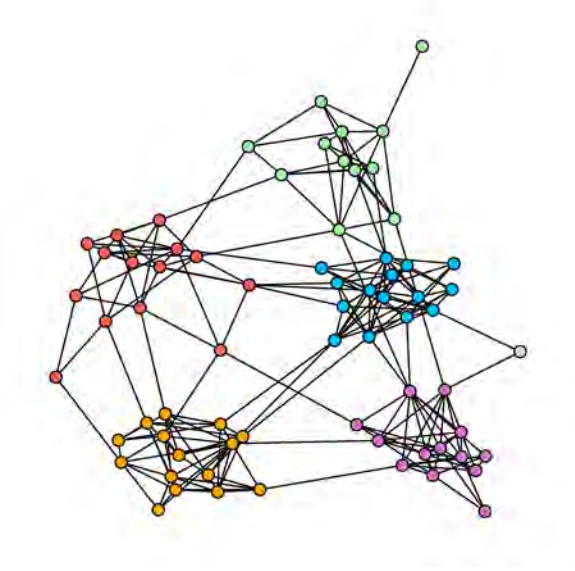
### *Modularity of Graphs*

Beyond heterogeneity, the degree to which a graph topology has strong community structures—often called *modularity*—plays a strong role in both the robustness and epidemic spreading in graphs. This mesoscopic level of structure lies between the local structure around individual nodes, such as degree centrality, and the global structure captured by such measures as the heterogeneity parameter. For the same global structure as given by a heterogeneity parameter, two graphs can exhibit different robustness and epidemic spreading characteristics if they differ in modularity at the mesoscopic level. Figure 4.7 shows a heterogeneous graph with high modularity; communities are indicated by different colored nodes.

Graphs with high modularity, such as the one shown in Figure 4.7, are characterized by densely connected communities that are weakly linked to one another. Both social and technological networks can display strong community structure. The links between communities are called *bridges*, and the nodes that define these links are called *bridging nodes*. Bridges and bridging nodes play important roles in modular graphs.

[73] Reuven Cohen, Shlomo Havlin, and Daniel ben-Avraham, "Efficient Immunization Strategies for Computer Networks and Populations," *Physical Review Letters*, Vol. 91, No. 24, 2003; J. Gómez-Gardeñes, P. Echenique, and Y. Moreno, "Immunization of Real Complex Communication Networks," *European Physical Journal B*, Vol. 49, No. 2, 2006.

**Figure 4.7. An Example of a Graph with High Modularity**



The more modular the graph topology, the more this mesoscopic structure determines the robustness of the graph. In many real modular graphs, the most effective way to shatter the graph is to target the bridges or bridging nodes.[74] After such an attack, the graph is thus roughly broken into its individual communities. Flow across the whole graph is stopped, but communication within the communities remains fairly robust. Targeting nodes of high degree centrality in highly modular graphs can be less effective even in highly heterogeneous graphs since high degree nodes can often cluster in the same communities (the topology of Facebook connections is one example).[75] Graph topologies designed with controlled modularity can be exploited to control which nodes can communicate with one another after attack and where to take risk in the loss of connections.

For the same reason that modularity can dominate robustness characteristics, so too can it dominate epidemic spreading. Bridges and bridging nodes play central roles in epidemic spreading. In some cases, epidemic spreading happens first within communities and then among communities in a graph with high modularity, therefore giving the graph two epidemic thresholds.[76] In graphs with dense connections in the communities, the communities play a lesser

[74] Quang Nguyen, Tuan V. Vu, Hanh-Duyen Dinh, Davide Cassi, Francesco Scotognella, Roberto Alfieri, and Michele Bellingeri, "Modularity Affects the Robustness of Scale-Free Model and Real-World Social Networks Under Betweenness and Degree-Based Node Attack," *Applied Network Science*, Vol. 6, No. 1, 2021.

[75] Clara Stegehuis, Remco van der Hofstad, and Johan S. H. van Leeuwaarden, "Epidemic Spreading on Complex Networks with Community Structures," *Scientific Reports*, Vol. 6, Article 29748, 2016.

[76] Bnaya Gross and Shlomo Havlin, "Epidemic Spreading and Control Strategies in Spatial Modular Network," *Applied Network Science*, Vol. 5, No. 1, 2020.

role in epidemic spreading; a dominant role is played by bridges and bridging nodes.[77] Therefore, immunization strategies for highly modular heterogeneous graphs typically target bridges and bridging nodes, thus confining virus spreading to communities.[78]

### *Cascading Failures in Graphs*

Some flows across graphs can be redistributed over the nodes and links depending on the available paths. The power grid and the internet are examples. If flow of electricity or data packets can take three paths before a failure, but only two afterward, the load on the nodes and links in those two paths increases. The capacities of those nodes and links are, in general, limited. The failure of a node or link can, therefore, increase the loads on other nodes and links to the degree that they fail, thereby increasing the loads on the remaining nodes and links to the point that more fail, and so on. This failure mode is called *cascading failure* or *avalanching failure*.[79] Heterogeneous graphs are generally more susceptible to cascading failures than homogeneous graphs.[80] Results can be counterintuitive: Some attacks on the power grid are expected to cascade with higher likelihood if the nodes bearing lower loads fail rather than those bearing highest load.[81] Such results stress the importance of analysis. Cascades in graphs are complex and require tailored analysis. Cascades (avalanches) can be arrested by various methods, including preemptive removal of selected nodes or links after the first failure but before the cascade advances,[82] or by protecting key nodes and links that can be identified without knowledge of the global topology.[83]

---

[77] Stegehuis, van der Hofstad, and van Leeuwaarden, 2016.

[78] Marcel Salathé and James H. Jones, "Dynamics and Control of Diseases in Networks with Community Structure," *PLoS Computational Biology*, Vol. 6, No. 4, 2010. See Hocine Cherifi, Gergely Palla, Boleslaw K. Szymanski, and Xiaoyan Lu, "On Community Structure in Complex Networks: Challenges and Opportunities," *Applied Network Science*, Vol. 4, No. 1, 2019, for a review of methods.

[79] Lucas D. Valdez, Louis Shekhtman, Cristian E. La Rocca, Xin Zhang, Sergey V. Buldyrev, Paul A. Trunfio, Lidia A. Braunstein, and Shlomo Havlin, "Cascading Failures in Complex Networks," *Journal of Complex Networks*, Vol. 8, No. 2, 2020.

[80] Adilson E. Motter and Ying-Cheng Lai, "Cascade-Based Attacks on Complex Networks," *Physical Review E*, Vol. 66, No. 6, 2002; Liang Zhao, Kwangho Park, and Ying-Cheng Lai, "Attack Vulnerability of Scale-Free Networks Due to Cascading Breakdown," *Physical Review E*, Vol. 70, No. 3, 2004; E. J. Lee, K.-I. Goh, B. Kahng, and D. Kim, "Robustness of the Avalanche Dynamics in Data-Packet Transport on Scale-Free Networks," *Physical Review E*, Vol. 71, No. 5, 2005.

[81] Jian-Wei Wang and Li-Li Rong, "Cascade-Based Attack Vulnerability on the US Power Grid," *Safety Science*, Vol. 47, No. 10, 2009.

[82] Adilson E. Motter, "Cascade Control and Defense in Complex Networks," *Physical Review Letters*, Vol. 93, No. 9, 2004.

[83] Alex Smolyak, Orr Levy, Irena Vodenska, Sergey Buldyrev, and Shlomo Havlin, "Mitigation of Cascading Failures in Complex Networks," *Scientific Reports*, Vol. 10, Article 16124, 2020.

## Mitigating Topological Fragility

There is no single way to measure robustness.[84] Just as there are many ways to measure the centrality of a node or link because there are a variety of ways in which nodes and links might be important, so too are there many ways to measure robustness of a graph topology because there are many insults to a system that the system needs to be robust against.[85] And a measure that expresses robustness to an insult to a graph of one topology might not be the right measure for the robustness against the same insult to a graph of a different topology. There is no formulaic way to measure robustness; measures must be selected based on the problem at hand.

For the same reasons, there are no formulaic solutions for designing systems to be robust. Solutions seeking improvements to the robustness of systems depend on a number of factors, including what the system is to be robust against, the topology of the graphical models of the system, how much risk is acceptable and of what kinds, what costs can be borne, and other constraints. In the context of robustness of complex graphs against the removal of nodes and links, two cases have been well studied—homogeneous and heterogeneous topologies.

We noted above that homogeneous and heterogeneous networks fail quite differently under random and targeted attacks. Large heterogeneous graphs are extremely robust against random attacks but can be susceptible to targeted attacks. Homogeneous graphs are better against targeted attacks but not as robust against random attacks. Is there a graph topology that combines these two behaviors and has better performance against random attacks than homogeneous graphs, yet avoids the susceptibility against targeted attacks that heterogeneous graphs can exhibit?

One approach for enhancing system robustness through adjustments to graph topology seeks this optimal compromise. Because there is no single way to measure robustness, there is no single way to define this objective function for optimization. Several measures have been proposed and investigated analytically. Fortunately, the optimum topologies of uncorrelated complex graphs between the extremes of homogeneous and heterogeneous are not very sensitive to the choice of robustness measure. Where costs or other constraints do not readily permit adding nodes or links to bolster robustness, a topological rewiring of a graph with the same number of nodes and links can optimize the robustness of the system against a given definition of robustness.[86]

---

[84] Seyedmohsen Hosseini, Kash Barker, and Jose E. Ramirez-Marquez, "A Review of Definitions and Measures of System Resilience," *Reliability Engineering & System Safety*, Vol. 145, January 2016.

[85] Such studies as Xiangrong Wang, Ling Feng, Robert E. Kooij, and Jose L. Marzo, "Inconsistencies Among Spectral Robustness Metrics," in: Trung Q. Duong, Nguyen-Son Vo, and Van Ca Phan, eds., *Quality, Reliability, Security and Robustness in Heterogeneous Systems: Proceedings of the 14th EAI International Conference, Qshine 2018*, Springer, 2019, that see differences among measure of robustness as "inconsistencies" fail to recognize that each metric measures a different attribute and gives unique insights.

[86] André X. C. N. Valente, Abhijit Sarkar, and Howard A. Stone, "Two-Peak and Three-Peak Optimal Complex Networks," *Physical Review Letters*, Vol. 92, No. 11, 2004; G. Paul, T. Tanizawa, S. Havlin, and H. E. Stanley,

A second approach is to judiciously add nodes, links, or both to enhance the robustness of a system. Like the previous case, the exact optimal locations depend on the choice of robustness measure. The locations also depend on the costs of adding nodes or links.[87]

Mitigations for graphs with high modularity must focus on the community structure of these graphs. Robustness to node or link removal and epidemic spreading are mathematically similar problems. Therefore, immunization strategies for epidemic spreading on modular complex graphs give considerable insights into strategies for enhancing the robustness of networks, especially those that represent process of flow across the graph.

## Insights from Topological Measures

The behavior of systems that are well modeled by graphs depends strongly on the topology of the graph. Graphs capture interdependencies of components of systems. They are excellent means by which to understand indirect effects—those effects induced on one part of a system by another, sometimes remote component. These indirect effects include failure modes, propagation of information or viruses, and cascading failures.

Graphs display an uncountable number of topologies. We focused in this chapter on trees and complex graphs. We chose trees because they are common in the context of problems likely to be faced in Sentinel. We discussed complex graphs, with an emphasis on heterogeneity, assortativity, and modularity, for two reasons. They are common, and they require the power of mathematical graph theory to understand.

Most empirically studied large graphs are complex, and of the complex graphs, the vast majority are heterogeneous—they have fat-tailed degree distributions. Among the better studied heterogeneous graphs are the internet,[88] the World Wide Web, social networks, protein interactions in cells, airline route structures, food webs, citations in the scientific literature, electronic circuits, neural networks, and metabolic reactions.[89] Heterogeneous topologies arise

---

"Optimization of Robustness of Complex Networks," *European Physical Journal B*, Vol. 38, No. 2, 2004; G. Paul, T. Tanizawa, S. Havlin, and H. E. Stanley, "Optimization of Robustness of Complex Networks" (Erratum), *European Physical Journal B*, Vol. 48, No. 1, 2005.

[87] Yukio Hayashi and Jun Matsukubo, "Improvement of the Robustness on Geographical Networks by Adding Shortcuts," *Physica A*, Vol. 380, Issue C, 2007; Navid Ahmadian, Gino J. Lim, Jaeyoung Cho, and Selim Bora, "A Quantitative Approach for Assessment and Improvement of Network Resilience," *Reliability Engineering & System Safety*, Vol. 200, August 2020; Masaki Chujyo and Yukio Hayashi, "A Loop Enhancement Strategy for Network Robustness," *Applied Network Science*, Vol. 6, No. 1, 2021.

[88] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, "On Power-Law Relationships of the Internet Topology," *Computer Communication Review*, Vol. 29, No. 4, 1999; Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener, "Graph Structure in the Web," *Computer Networks*, Vol. 33, Nos. 1–6, 2000; Romualdo Pastor-Satorras and Alessandro Vespignani, *Evolution and Structure of the Internet*, Cambridge University Press, 2004.

[89] Réka Albert and Albert-László Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, Vol. 74, No. 1, 2002; M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, Vol. 45, No. 2, 2003; Albert-László Barabási and Eric Bonabeau, "Scale-Free Networks," *Scientific*

when a new node attaches preferentially to well connected nodes rather than randomly. This mechanism explains why many, but not all, large, complex graphs display heterogeneity. Homogeneous topologies often occur when some strong constraint limits the nature of connections. Whereas the airline route structure in the United States is strongly heterogeneous, the interstate highway system is homogeneous.[90] Airlines can fly from any airport to any other airport without much regard of airports in between. But any highway connecting New York City with Los Angeles sensibly goes through the major cities along the way. Hence there are airline hubs at major airports and the graph is heterogeneous. But there are no hubs for highways, and the graph is homogeneous.

Complex graphs display a wide variety of complex behaviors. These behaviors—central to understanding robustness, failure modes, and propagation of information and viruses—are impossible to discern for large, complex graphs without the mathematical tools of graph theory. These tools, some of which we described in this chapter, offer the ability to scale to large systems and to understand the complexity in reducible terms, as well as offer a level of rigor that intuition and analysis by hand cannot provide.

The choices of graph representation, centrality metrics, and measures that depend on topology, such as robustness or epidemic diffusion, are contextual. The mathematical tools of graph theory are powerful and provide unique insights into the systems that they model. But there is no formulaic approach to this analysis. There is no substitute for expertise, and analysts will need to be consulted to tailor the analysis to the problem at hand and perform the calculations. Decisionmakers who consume this analysis will need at least a rudimentary understanding of the underlying concepts. It is our goal that this chapter is a strong foundation for that understanding of topological controls.

---

*American*, May 2003; Guido Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology*, Oxford University Press, 2007.

[90] Barabási and Bonabeau, 2003.

# Chapter 5. Summary and Concluding Remarks

> The world is becoming more and more complicated. We are rapidly approaching an unbearable and even dangerous situation: Only a small fraction of the population will know what is going on and the great majority will be condemned to be merely passive spectators. Mathematics is the only subject broad enough to prevent this.
>
> —Arthur Engel[91]

The theory of constructing graph models and analyzing graphs is too vast to reduce to a short primer. It is also impossible to reduce graph analysis to a formulaic approach. Analysts will need to take the concepts, metrics, and techniques described in this report and extend them as needed by building on them and consulting the literature. Our hope is that this brief introduction will provide ideas regarding how graphs can be used to add rigor to nuclear certification, nuclear surety, and cybersecurity risk assessment activities. The analysts will need to understand in detail what the certifiers and risk assessors want to know about the system. The certifiers and risk assessors must be able to shape how they communicate their problems for the analysts. The closer the collaboration between those making decisions and those performing analysis to inform those decisions, the better that graph models can be constructed and analyzed to support nuclear certification, nuclear surety, and cybersecurity risk assessment.

The metrics for centrality of nodes and links and the measures of topological control over robustness and epidemic spreading are only samples of the mathematical apparatuses available to analyze graphs. The centrality metrics that we presented—degree, betweenness, and eigenvector—are the most commonly available in software packages that have applicability to the types of problems encountered in assessing a weapon system. We also described a centrality metric that is less commonly seen—cutset analysis—because it is well suited to assess the number of independent paths through a system and ranks sets of nodes or links by how critical they are to reducing the number of independent paths. The number of topologies is nearly boundless. The ones that we discussed—trees and homogeneous and heterogeneous complex graphs—are expected to be common in weapon system settings. As analysts explore new graph representations and as certifiers and risk assessors define specific needs, the set of tools needed will build on those presented here.

It is our hope that this primer gives fresh ideas to analysts for how to more rigorously perform nuclear certification, nuclear surety, and cybersecurity risk assessment on large, complex systems and that those who receive that analysis can better interpret them.

---

[91] Arthur Engel, "The Relevance of Modern Fields of Applied Mathematics for Mathematical Education," *Educational Studies in Mathematics*, Vol. 2, No. 2–3, 1969, p. 257.

# Abbreviation

| | |
|---|---|
| MBSE | model-based systems engineering |

# References


Ahmadian, Navid, Gino J. Lim, Jaeyoung Cho, and Selim Bora, "A Quantitative Approach for Assessment and Improvement of Network Resilience," *Reliability Engineering & System Safety*, Vol. 200, August 2020.

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Inc., 1993.

Aksoy, Sinan G., Emilie A. H. Purvine, Eduardo Cotilla Sanchez, and Mahantesh Halappanavar, "A Generative Graph Model for Electrical Infrastructure Networks," *Journal of Complex Networks*, Vol. 7, No. 1, 2019.

Albert, Réka, and Albert-László Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, Vol. 74, No. 1, 2002.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási, "Error and Attack Tolerance of Complex Networks," *Nature*, Vol. 406, No. 6794, 2000.

Aris, Rutherford, *Mathematical Modelling Techniques*, Dover, 1994.

Balashov, Nicole, Reuven Cohen, Avieli Haber, Michael Krivelevich, and Simi Haber, "Optimal Shattering of Complex Networks," *Applied Network Science*, Vol. 4, No. 1, 2019.

Baltz, Andreas, and Lasse Kliemann, "Spectral Analysis," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005.

Barabási, Albert-László, and Eric Bonabeau, "Scale-Free Networks," *Scientific American*, May 2003.

Barrat, Alain, Marc Barthélemy, and Alessandro Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.

Barthélemy, M., "Betweenness Centrality in Large Complex Networks," *The European Physical Journal B*, Vol. 38, 2004.

Barthélemy, Marc, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani, "Dynamical Patterns of Epidemic Outbreaks in Complex Heterogeneous Networks," *Journal of Theoretical Biology*, Vol. 235, No. 2, 2005.

Barthélemy, Marc, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani, "Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks," *Physical Review Letters*, Vol. 92, No. 17, 2004.

Bavelas, Alex, "A Mathematical Model for Group Structures," *Applied Anthropology*, Vol. 7, No. 3, 1948.

Bender, Edward A., *An Introduction to Mathematical Modeling*, John Wiley & Sons, 1978.

Biggs, Norman L., E. Keith Lloyd, and Robin J. Wilson, *Graph Theory, 1736-1936*, Clarendon Press, 1986.

Bockholt, Mareike, and Katharina Anna Zweig, "Towards a Process-Driven Network Analysis," *Applied Network Science*, Vol. 5, No. 1, 2020.

Bollobás, Béla, *Modern Graph Theory*, Springer, 1998.

Bonacich, Phillip, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology*, Vol. 2, No. 1, 1972.

Bonacich, Phillip, "Power and Centrality: A Family of Measures," *American Journal of Sociology*, Vol. 92, No. 5, 1987.

Bonacich, Phillip, "Some Unique Properties of Eigenvector Centrality," *Social Networks*, Vol. 29, No. 4, 2007.

Borgatti, Stephen P., "Centrality and Network Flow," *Social Networks*, Vol. 27, No. 1, 2005.

Borky, John M., and Thomas H. Bradley, *Effective Model-Based Systems Engineering*, Springer, 2019.

Brandes, Ulrik, "A Faster Algorithm for Betweenness Centrality," *Journal of Mathematical Sociology*, Vol. 25, No. 2, 2001.

Brandes, Ulrik, "On Variants of Shortest-Path Betweenness Centrality and Their Generic Computation," *Social Networks*, Vol. 30, No. 2, 2008.

Broder, Andrei, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener, "Graph Structure in the Web," *Computer Networks*, Vol. 33, Nos. 1–6, 2000.

Caldarelli, Guido, *Scale-Free Networks: Complex Webs in Nature and Technology*, Oxford University Press, 2007.

Callaway, Duncan S., M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs," *Physical Review Letters*, Vol. 85, No. 25, 2000.

Carter, Mark R., Mark P. Howard, Nicholas Owens, David Register, Jason Kennedy, Kelley Pecheux, and Aaron Newton, *Effects of Catastrophic Events on Transportation System Management and Operations, Howard Street Tunnel Fire, Baltimore City, Maryland—July 18, 2001*, U.S. Department of Transportation, July 2002.

CERT National Insider Threat Center, *Common Sense Guide to Mitigating Insider Threats*, 6th ed., Carnegie Mellon University, Software Engineering Institute, Technical Note CMU/SEI-2018-TR-010, December 2018.

Chalermsook, Parinya, Jittat Fakcharoenphol, and Danupon Nanongkai, "A Deterministic Near-Linear Time Algorithm for Finding Minimum Cuts in Planar Graphs," *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, January 2004.

Chen, You, Steve Nyemba, Wen Zhang, and Bradley Malin, "Specializing Network Analysis to Detect Anomalous Insider Actions," *Security Informatics*, Vol. 1, No. 1, 2012.

Cherifi, Hocine, Gergely Palla, Boleslaw K. Szymanski, and Xiaoyan Lu, "On Community Structure in Complex Networks: Challenges and Opportunities," *Applied Network Science*, Vol. 4, No. 1, 2019.

Chinchani, Ramkumar, Anusha Iyer, Hung Q. Ngo, and Shambhu Upadhyaya, "Towards a Theory of Insider Threat Assessment," *Proceedings of the 2005 International Conference on Dependable Systems and Networks*, 2005.

Chujyo, Masaki, and Yukio Hayashi, "A Loop Enhancement Strategy for Network Robustness," *Applied Network Science*, Vol. 6, No. 1, 2021.

Cohen, Reuven, Shlomo Havlin, and Daniel ben-Avraham, "Efficient Immunization Strategies for Computer Networks and Populations," *Physical Review Letters*, Vol. 91, No. 24, 2003.

Cohen, Reuven, Keren Erez, Daniel ben-Avraham, and Shlomo Havlin, "Breakdown of the Internet Under Intentional Attack," *Physical Review Letters*, Vol. 86, No. 16, 2001.

Cohen, Reuven, Keren Erez, Daniel ben-Avraham, and Shlomo Havlin, "Resilience of the Internet to Random Breakdowns," *Physical Review Letters*, Vol. 85, No. 21, 2000.

Davison, Peter, Bruce Cameron, and Edward F. Crawley, "Technology Portfolio Planning by Weighted Graph Analysis of System Architectures," *Systems Engineering*, Vol. 18, No. 1, 2015.

Defense Acquisition University, *Glossary of Defense Acquisition Acronyms and Terms*, July 21, 2020.

Dezső, Zoltán, and Albert-László Barabási, "Halting Viruses in Scale-Free Networks," *Physical Review E*, Vol. 65, No. 5, 2002.

Diestel, Reinhard, *Graph Theory*, 5th ed., Springer, 2017.

Easley, David, and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

Engel, Arthur, "The Relevance of Modern Fields of Applied Mathematics for Mathematical Education," *Educational Studies in Mathematics*, Vol. 2, No. 2–3, 1969.

Erdős, P., and A. Rényi, "On Random Graphs I," *Publicationes Mathematicae (Debrecen)*, Vol. 6, 1959.

Everett, Martin G., and Stephen P. Borgatti, "Extending Centrality," in Peter J. Carrington, John Scott, and Stanley Wasserman, eds., *Models and Methods in Social Network Analysis*, Cambridge University Press, 2005.

Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos, "On Power-Law Relationships of the Internet Topology," *Computer Communication Review*, Vol. 29, No. 4, 1999.

Filiol, Éric, and Cécilla Gallais, "Optimization of Operational Large-Scale (Cyber) Attacks by a Combinatorial Approach," in Information Resources Management Association, ed., *Cyber Warfare and Terrorism Concepts, Methodologies, Tools, and Applications*, IGI Global, 2020.

Fowler, A. C., *Mathematical Models in the Applied Sciences*, Cambridge University Press, 1997.

Freeman, Linton C., "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, Vol. 40, No. 1, 1977.

Gamachchi, Anagi, Li Sun, and Serdar Boztas, "A Graph Based Framework for Malicious Insider Threat Detection," *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

Gao, Chao, Jiming Liu, and Ning Zhong, "Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis," *Knowledge and Information Systems*, Vol. 27, No. 2, 2011.

Goh, K.-I., B. Kahng, and D. Kim, "Spectra and Eigenvectors of Scale-Free Networks," *Physical Review E*, Vol. 64, No. 5, 2001.

Gómez-Gardeñes, J., P. Echenique, and Y. Moreno, "Immunization of Real Complex Communication Networks," *European Physical Journal B*, Vol. 49, No. 2, 2006.

Gross, Bnaya, and Shlomo Havlin, "Epidemic Spreading and Control Strategies in Spatial Modular Network," *Applied Network Science*, Vol. 5, No. 1, 2020.

Harrison, Willie K., "The Role of Graph Theory in System of Systems Engineering," *IEEE Access*, Vol. 4, 2016.

Hartnett, Gavin S., Edward Parker, Timothy R. Gulden, Raffaele Vardavas, and David Kravitz, "Modelling the Impact of Social Distancing and Targeted Vaccination on the Spread of COVID-19 Through a Real City-Scale Contact Network," *Journal of Complex Networks*, Vol. 9, No. 6, 2021.

Hayashi, Yukio, and Jun Matsukubo, "Improvement of the Robustness on Geographical Networks by Adding Shortcuts," *Physica A*, Vol. 380, Issue C, 2007.

Holme, Petter, Beom Jun Kim, Chang No Yoon, and Seung Kee Han, "Attack Vulnerability of Complex Networks," *Physical Review E*, Vol. 65, No. 5, 2002.

Hosseini, Seyedmohsen, Kash Barker, and Jose E. Ramirez-Marquez, "A Review of Definitions and Measures of System Resilience," *Reliability Engineering & System Safety*, Vol. 145, January 2016.

International Council on Systems Engineering, *Systems Engineering Vision 2020*, Version 2.03, INCOSE-TP-2004-004-02, September 2007.

Jacob, Riko, Dirk Koschützki, Katharina Anna Lehmann, Leon Peeters, and Dagmar Tenfelde-Podehl, "Algorithms for Centrality Indices," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005.

Jacob, Rinku, K. P. Harikrishnan, R. Misra, and G. Ambika, "Measure for Degree Heterogeneity in Complex Networks and Its Application to Recurrence Network Analysis," *Royal Society Open Science*, Vol. 4, No. 1, 2017.

Karger, David R., "Minimum Cuts in Near-Linear Time," *Journal of the ACM*, Vol. 47, No. 1, 2000.

Katz, Leo, "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, Vol. 18, No. 1, 1953.

Kitsak, Maksim, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse, "Identification of Influential Spreaders in Complex Networks," *Nature Physics*, Vol. 6, No. 11, 2010.

Koschützki, Dirk, Katharina Anna Lehmann, Leon Peeters, Stefan Richter, Dagmar Tenfelde-Podehl, and Oliver Zlotowski, "Centrality Indices," in Ulrik Brandes and Thomas Erlebach, eds., *Network Analysis: Methodological Foundations*, Springer, 2005.

Layton, Lyndsey, and Don Phillips, "Train Sets Tunnel Afire, Shuts Down Baltimore," *Washington Post*, July 19, 2001.

Lee, E. J., K.-I. Goh, B. Kahng, and D. Kim, "Robustness of the Avalanche Dynamics in Data-Packet Transport on Scale-Free Networks," *Physical Review E*, Vol. 71, No. 5, 2005.

Lin, C. C., and L. A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*, Macmillan Publishing Co., 1974.

MacCluer, C. R., "The Many Proofs and Applications of Perron's Theorem," *SIAM Review*, Vol. 42, No. 3, 2000.

Malinowski, Jacek, "A New Efficient Algorithm Generating All Minimal S-T Cut-Sets in a Graph-Modeled Network," *AIP Conference Proceedings*, Vol. 1738, No. 1, 2016.

Motter, Adilson E., "Cascade Control and Defense in Complex Networks," *Physical Review Letters*, Vol. 93, No. 9, 2004.

Motter, Adilson E., and Ying-Cheng Lai, "Cascade-Based Attacks on Complex Networks," *Physical Review E*, Vol. 66, No. 6, 2002.

Murakami, Masaya, Shu Ishikura, Daichi Kominami, Tetsuya Shimokawa, and Masayuki Murata, "Robustness and Efficiency in Interconnected Networks with Changes in Network Assortativity," *Applied Network Science*, Vol. 2, No. 1, 2017.

Newman, M. E. J., "Assortative Mixing in Networks," *Physical Review Letters*, Vol. 89, No. 20, 2002.

Newman, M. E. J., "A Measure of Betweenness Centrality Based on Random Walks," *Social Networks*, Vol. 27, No. 1, 2005.

Newman, M. E. J., "The Structure and Function of Complex Networks," *SIAM Review*, Vol. 45, No. 2, 2003.

Newman, Mark, *Networks*, 2nd ed., Oxford University Press, 2018.

Nguyen, Quang, Tuan V. Vu, Hanh-Duyen Dinh, Davide Cassi, Francesco Scotognella, Roberto Alfieri, and Michele Bellingeri, "Modularity Affects the Robustness of Scale-Free Model and Real-World Social Networks Under Betweenness and Degree-Based Node Attack," *Applied Network Science*, Vol. 6, No. 1, 2021.

Noldus, Rogier, and Piet Van Mieghem, "Assortativity in Complex Networks," *Journal of Complex Networks*, Vol. 3, No. 4, 2015.

O'Halloran, Bryan M., Nikolaos Papakonstantinou, Kristin Giammarco, and Douglas L. Van Bossuyt, "A Graph Theory Approach to Predicting Functional Failure Propagation During Conceptual Systems Design," *Systems Engineering*, Vol. 24, No. 2, 2021.

Pastor-Satorras, Romualdo, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani, "Epidemic Processes in Complex Networks," *Reviews of Modern Physics*, Vol. 87, No. 3, 2015.

Pastor-Satorras, Romualdo, and Alessandro Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, Vol. 86, No. 14, 2001.

Pastor-Satorras, Romualdo, and Alessandro Vespignani, *Evolution and Structure of the Internet*, Cambridge University Press, 2004.

Pastor-Satorras, Romualdo, and Alessandro Vespignani, "Immunization of Complex Networks," *Physical Review E*, Vol. 65, No. 3, 2002.

Paul, G., T. Tanizawa, S. Havlin, and H. E. Stanley, "Optimization of Robustness of Complex Networks," *European Physical Journal B*, Vol. 38, No. 2, 2004.

Paul, G., T. Tanizawa, S. Havlin, and H. E. Stanley, "Optimization of Robustness of Complex Networks" (Erratum), *European Physical Journal B*, Vol. 48, No. 1, 2005.

Potts, Matthew W., Pia Sartor, Angus Johnson, and Seth Bullock, "A Network Perspective on Assessing System Architectures: Foundations and Challenges," *Systems Engineering*, Vol. 22, No. 6, 2019.

Potts, Matthew W., Pia A. Sartor, Angus Johnson, and Seth Bullock, "A Network Perspective on Assessing System Architectures: Robustness to Cascading Failure," *Systems Engineering*, Vol. 23, No. 5, 2020.

Ratner, Andrew, "Train Derailment Severs Communications," *Baltimore Sun*, July 20, 2001.

Rickart, Charles E., *Structuralism and Structures: A Mathematical Perspective*, World Scientific, 1995.

Saaty, Thomas L., and Joyce M. Alexander, *Thinking with Models: Mathematical Models in the Physical, Biological, and Social Sciences*, Pergamon Press, 1981.

Salathé, Marcel, and James H. Jones, "Dynamics and Control of Diseases in Networks with Community Structure," *PLoS Computational Biology*, Vol. 6, No. 4, 2010.

Smith, Keith M., and Javier Escudero, "Normalised Degree Variance," *Applied Network Science*, Vol. 5, No. 1, 2020.

Smolyak, Alex, Orr Levy, Irena Vodenska, Sergey Buldyrev, and Shlomo Havlin, "Mitigation of Cascading Failures in Complex Networks," *Scientific Reports*, Vol. 10, Article 16124, 2020.

Snyder, Don, Elizabeth Bodine-Baron, Dahlia A. Goldfeld, Bernard Fox, Myron Hura, Mahyar A. Amouzegar, and Lauren Kendrick, *Cyber Mission Thread Analysis: A Prototype Framework for Assessing Impact to Missions from Cyber Attacks to Weapon Systems*, RAND Corporation, RR-3188/1-AF, 2022. As of January 9, 2024: https://www.rand.org/pubs/research_reports/RR3188z1.html

Snyder, Don, Christian Johnson, Parousia Rockstroh, Lance Menthe, and Bart Bennett, *Graph Theoretic Algorithms for the Ground Based Strategic Deterrent Program: Prioritization and Scheduling*, RAND Corporation, RR-A583-1, 2021. As of January 9, 2024: https://www.rand.org/pubs/research_reports/RRA583-1.html

Stegehuis, Clara, Remco van der Hofstad, and Johan S. H. van Leeuwaarden, "Epidemic Spreading on Complex Networks with Community Structures," *Nature Scientific Reports*, Vol. 6, Article 29748, 2016.

Styron, Hilary C., *CSX Tunnel Fire, Baltimore, Maryland*, U.S. Fire Administration, U.S. Department of Homeland Security, USFA-TR-140, July 2001.

Valdez, Lucas D., Louis Shekhtman, Cristian E. La Rocca, Xin Zhang, Sergey V. Buldyrev, Paul A. Trunfio, Lidia A. Braunstein, and Shlomo Havlin, "Cascading Failures in Complex Networks," *Journal of Complex Networks*, Vol. 8, No. 2, 2020.

Valente, André X. C. N., Abhijit Sarkar, and Howard A. Stone, "Two-Peak and Three-Peak Optimal Complex Networks," *Physical Review Letters*, Vol. 92, No. 11, 2004.

Wang, Jian-Wei, and Li-Li Rong, "Cascade-Based Attack Vulnerability on the US Power Grid," *Safety Science*, Vol. 47, No. 10, 2009.

Wang, Xiangrong, Ling Feng, Robert E. Kooij, and Jose L. Marzo, "Inconsistencies Among Spectral Robustness Metrics," in: Trung Q. Duong, Nguyen-Son Vo, and Van Ca Phan, eds., *Quality, Reliability, Security and Robustness in Heterogeneous Systems: Proceedings of the 14th EAI International Conference, Qshine 2018*, Springer, 2019.

Wasserstein, Ron, "George Box: A Model Statistician," *Significance*, Vol. 7, No. 3, 2010.

White, Douglas R., and Stephen P. Borgatti, "Betweenness Centrality Measures for Directed Graphs," *Social Networks*, Vol. 16, No. 4, 1994.

Wu, Qinghua, and Jin-Kao Hao, "A Review on Algorithms for Maximum Clique Problems," *European Journal of Operational Research*, Vol. 242, No. 3, 2015.

Young, William, and Nancy G. Leveson, "An Integrated Approach to Safety and Security Based on Systems Theory: Applying a More Powerful New Safety Methodology to Security Risks," *Communications of the ACM*, Vol. 57, No. 2, 2014.

Zhao, Liang, Kwangho Park, and Ying-Cheng Lai, "Attack Vulnerability of Scale-Free Networks Due to Cascading Breakdown," *Physical Review E*, Vol. 70, No. 3, 2004.

Zweig, Katharina A., *Network Analysis Literacy: A Practical Approach to the Analysis of Networks*, Springer, 2016.