# CHARACTERISTICS OF ROBUST COMPLEX NETWORKS

by

ALI SYDNEY

B.S., United States Naval Academy, Maryland, 2007

---

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2009

Approved by:

Major Professor

Dr. Caterina Scoglio

# Copyright

# Abstract

In network theory, a complex network represents a system whose evolving structure and dynamic behavior contribute to its robustness. The study of complex networks, though young, spans diverse domains including engineering, science, biology, sociology, psychology, and business, to name a few. Regardless of the field of interest, robustness defines a network's survivability in the advent of classical component failures and at the onset of cryptic malicious attacks.

With increasingly ambitious initiatives such as GENI and FIND that seek to design future internets, it becomes imperative to define the characteristics of robust topologies, and to build future networks optimized for robustness. This thesis investigates the characteristics of network topologies that maintain a high level of throughput in spite of multiple attacks. To this end, we select network topologies belonging to the main network models and some real world networks. We consider three types of attacks: removal of random nodes, high degree nodes, and high betweenness nodes. We use elasticity as our robustness measure and, through our analysis, illustrate that different topologies can have different degrees of robustness. In particular, elasticity can fall as low as $0.8\%$ of the upper bound based on the attack employed. This result substantiates the need for optimized network topology design. Furthermore, we implement a tradeoff function that combines elasticity under the three attack strategies and considers the cost of the network. Our extensive simulations show that, for a given network density, regular and semi-regular topologies can have higher degrees of robustness than heterogeneous topologies, and that link redundancy is a sufficient but not necessary condition for robustness.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

This work would not be a reality without the support and encouragement of my Advisor, Dr Caterina Scoglio, under whose supervision I began and completed this thesis. Dr Scoglio's engaging personality constantly motivated me to achieve excellence by ensuring all fundamental concepts were thoroughly understood. She is a true "sailor" at heart and, just like my superiors at the US Naval Academy, she reinforced the ideals of "work hard-play hard". For this reason, I submitted a conference paper and had the privilege to present my work in Japan. Additionally, I recently submitted my first journal paper.

First, I would like to thank God for giving me the ability and allowing me the opportunity to pursue my dreams and aspirations. I would also like to thank my colleagues and friends, Phillip Schumm and Mina Youssef, for their "indoctrination" into this work and their continued efforts to realize its completion. Their constant reviews and constructive criticisms have enabled me to strive to achieve a higher quality of work. I am also grateful for the insights and contribution of Dr Robert Kooij from Delft University of Technology. Finally, I would like to thank Dr Gruenbacher and Dr Soldan for providing an assistanceship. My success would not be possible without their support.

I cannot end without thanking my family: Victoria, Andrew, Curby Dwaine, and Tayilah Sydney, whose constant thoughtfulness and love has provided this thirst for excellence. I am also appreciative of my older brother's example in being my role model, Dr Curby Dwaine Sydney, who never rests until his objective has been achieved. Furthermore, I would like to thank my church family, especially the Khodras, for their constant prayers. Finally, I want to thank Netta Khodra for her support and encouragement throughout these years.

# Chapter 1

# INTRODUCTION

Why study future network topologies? For one, we have experienced several moderate sized failures and thus, large failures are inevitable. In particular, the 2006 earthquake in Taiwan disrupted undersea fiber optic communication lines and as a result, banks from South Korea to Australia suffered massive interruptions [1]. Though this represents a direct network failure, failures can also occur indirectly. For example Code Red, a computer virus that incapacitated numerous networks, resulted in a global loss of 2 billion US dollars [2]. Furthermore, in 2004, Sassar virus disruptions accounted for the halt on maritime operations in the UK, the halt on railway operations in Australia, and the interruptions in hospital facilities in Hong Kong [3]. The US General Accounting Office estimated 250,000 annual attacks on Department of Defense networks [4]. Objectives range from theft to immobilization of entire networks. Another riveting example stems from a series of cascading failures in 2003 that resulted in a blackout in the Northeastern states [5]. A similar phenomenon occurred the very same year in Italy and left 56 million residents without power for 9 hours [6]. Our daily routines would cease to exist should network topologies disintegrate. Thus, as failures and attacks increase, it is imperative to design future topologies robust against unforeseen catastrophes for future network initiatives.

Amongst other definitions, a network can be robust if disconnecting components is difficult. However, we define robustness as the ability of a network to maintain its total throughput under node and link removal. The former definition is based on topological characteristics, while the latter also considers traffic flow.

Approaches for determining the robustness of graphs has evolved from simple graph theoretic concepts that highlight the connectivity of a graph [7] to more recent concepts that consider the spectrum of a graph [8]. However, these measures are unable to capture our definition of robustness. For this reason, we use elasticity as our measure of robustness; it meets the functional requirements of capturing throughput under node and link removal.

The importance of this thesis stems from our objective to extract the characteristics of robust networks. With these results, we seek to produce future robust network topologies. Thus, to realize our first goal, we

1. use the metric elasticity as a measure of robustness of a network
2. establish the upper bound for elasticity
3. assess elasticity for diverse network models
4. present correlations between elasticity and selected network metrics
5. develop and implement a function that considers the tradeoff between elasticity and network cost
6. extract characteristics of networks that make them robust

The rest of this thesis is structured as follows. Chapter 2 gives a brief overview on measures for robustness while Chapters 3 and 4 provide a more detailed review of metrics and distributions for characterizing the structure and behavior of a network. Chapter 5 presents the network models from which networks will be selected to assess their elasticity. In Chapter 6, we review elasticity, our robustness measure, and provide analytical and numerical approaches to obtain the upper bound. In Chapter 7, we assess the elasticity of each network, implement a tradeoff function that considers elasticity under the three removal strategies, and discuss the characteristics which make a network robust. Finally, we discuss the benefits and shortcomings of elasticity and highlight our future initiatives to characterize the robustness of complex networks in Chapter 8.

# Chapter 2

# DEFINITION OF ROBUSTNESS

What is robustness? In complex networks, robustness can encompass properties ranging from redundancy and diversity, to concepts such as the ability to operate under perturbation or the efficiency of feedback control mechanisms. In this work, we define robustness as the ability of a network to maintain its total throughput under removal of nodes (and links). Our definition directly considers flows on a network. Properly defined, flows can represent intuitive concepts such as traffic flow on a road network, or an abstract concept as the flow (or transmission) of a disease in an epidemic.

Though flows provide an additional perspective in evaluating the behavior of a network, we cannot overlook the role of structural properties of a network; both are necessary to extract features of a network. For this reason metrics to measure the robustness can be classified into two key groups: metrics that consider a network's structure (the interaction of network components), and metrics that consider a network's function (contributing factors to the network's operation and behavior). Throughout this thesis, the terms network and graph will be used interchangeably. However, from a graph theory perspective, a graph is a collection of edges and vertices while a network contains information relevant to a system (flows, capacities etc.)

### 2.0.1 Background and Related Work

The classical approach for determining robustness of networks entails the use of basic concepts from graph theory. For instance, the connectivity of a graph is an important, and probably the earliest, measure of robustness of a network [7]. Node (link) connectivity, defined as the size of the smallest node (link) cut, determines in a certain sense the robustness of a graph to the deletion of nodes (links). However, the node or link connectivity only partly reflects the ability of graphs to retain certain degrees of connectedness after deletion. Other improved measures were introduced and studied, including super connectivity [9], conditional connectivity [10], restricted connectivity [11], fault diameter [12], toughness [13], scattering number [14], tenacity [15], expansion parameter [16], and isoperimetric number [17]. In contrast to node (link) connectivity, these new measures consider both the cost to damage a network and how badly the network is damaged.

Subsequent measures consider the size of the largest connected component as nodes are attacked [18]. Furthermore, percolation models were used to assess the damage incurred by random graphs [19]. From spectral analysis, experimentalists consider the second smallest Laplacian eigenvalue as a measure of how difficult it is to break the network into components [8].

The measures reviewed thus far consider the network structure to assess robustness. However, more recent efforts have incorporated the behavior of the network [20, 21]. More precisely, the authors maximized flows in the network while imposing constraints on routers and links.

Other metrics in networking literature include the average node degree [22], betweenness [23], heterogeneity [24], and characteristic path length [25]. In this thesis, our results show significant corelations between elasticity and some of these metrics which will be used to characterize the robustness of networks.

Chapter 3 provides a more thorough analysis of the various metrics to assess the robustness of a network.

# Chapter 3

# METRICS TO ASSESS ROBUSTNESS

## 3.1 CHARACTERIZING THE STRUCTURE

Combinatorics, a branch of theoretical mathematics, considers metrics to characterize the network's structure. Metrics such as edge expansion, node expansion, spectral gap, natural connectivity, and algebraic connectivity, are metrics to measure the robustness of a graph with respect to its connectivity (for two nodes $s$ and $t$, connectivity defines the existence of a path between these nodes. Otherwise, they are disconnected.) In this case, robustness assesses the difficulty of disconnecting a network into components (a component is subgraph in which a path exists between any source-destination node $s$ and $t$). Applying this definition of robustness, a fully connected graph, formally known as a mesh or complete graph, is more robust than a star graph. However, given the extensive costs in implementing a mesh network, combinatorics considers a robust graph as one, which though sparse, exhibits high connectivity: expander graphs fall in this category. In the following sections, several metrics will be presented with their respective definitions, and for some metrics, a discussion on their applicability in certain cases. Table 3.1 highlights the most frequently used variables throughout this work.

| Variables | Definitions |
|---|---|
| $n$ | node |
| $N$ | total number of nodes |
| $m$ | link |
| $M$ | total number of links |
| $k_n$ | degree of node $n$ |
| $\bar{k}$ | average node degree |
| $P(k)$ | Probability given node has degree $k$ |
| $n(k)$ | Number of nodes with degree $k$ |

**Table 3.1**: *Definitions of the most common variables used throughout this work*

### 3.1.1 Average Degree

Average node degree $\bar{k}$, of graph $G$, where $G = (V, E)$, is the average number of edges connecting vertices in $G$. Equation 3.1 computes the average node degree.

$$\bar{k} = \frac{2M}{N} = \frac{1}{N}\sum_{n=1}^{N} k_n = \sum_{k=0}^{k_{\max}} kP(k) \tag{3.1}$$

where $k_{max}$ is the maximum node degree.

### 3.1.2 Diameter

Diameter is the longest shortest path (or the longest graph geodesic), between any source-destination node in graph $G$. It can be calculated by finding the shortest path between source node $s$ and destination node $t$, $\forall$ $(s, t) \in V$: Diameter then is the longest of all these paths. Several algorithms, including Breath-first search, Depth-first search, and Floyd-Warshall, can compute this metric [26]. Radius is another metric related to diameter: unlike diameter, it is the shortest of the set of all longest shortest paths from (or to) all nodes.

### 3.1.3 Characteristic Path Length

Average distance, characteristic path length, and average shortest path, all refer to the same metric. Thus, they all have similar definitions: the expected shortest distance between two nodes. More specifically, it is the sum of all source-destination shortest paths divided by the number of node pairs in the network.

### 3.1.4 Clustering Coefficient

For an undirected graph, Watts and Strogatz's clustering coefficient $C_n$ in [27], for a given node $n$, is defined as

$$C_n = \frac{2m_{ij}}{k_n(k_n-1)} \qquad n = 1, 2, .., N, \forall\,(i,j)\,\epsilon V \tag{3.2}$$

where $m_{ij}$ is the number of links among all neighbors of node $n$. Thus, the clustering coefficient of a graph, $C_G$, is given by

$$C_G = \frac{1}{N}\sum_{n=1}^{N} C_n \qquad 0 \leq C_G \leq 1 \tag{3.3}$$

The clustering coefficient assesses how likely it is for a node and its neighbors to form a mesh. Intuitively, higher clustering implies higher path diversity and higher interconnection. For the study of virus propagation in the SIR epidemic model, clustering (that partitions the network) provides the benefit of reducing the epidemic's size [28]. However, viruses spread faster in highly clustered networks. The examples in Figures 3.1 and 3.2 serve as an example to calculate clustering coefficient of a node

- For $G$ in Figure 3.1, where $G = (V, E)$, nodes $\{2, 3, 4\} \in V$ form the neighborhood of node $1$. The edges in this neighborhood are $\{(2, 3)(2, 4)(3, 4)\} \in E$. The Maximum number of edges possible between these nodes is $\frac{k_1(k_1-1)}{2} = 3$. Thus, $C_1 = 1$

- In Figure 3.2, the set of nodes in the neighborhood of node $1$ is the same as in Figure 3.1. However, there is only one edge in this neighborhood: $\{(3, 4)\} \in E$. The Maximum number of edges possible between these nodes is 3. Thus, $C_1 = \frac{1}{3}$

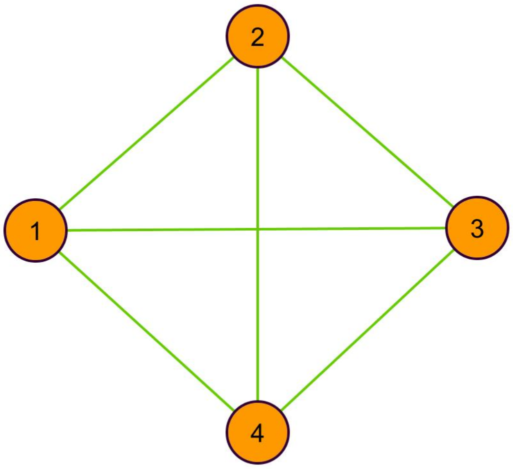

**Figure 3.1**: *Computing clustering coefficient for node 1*

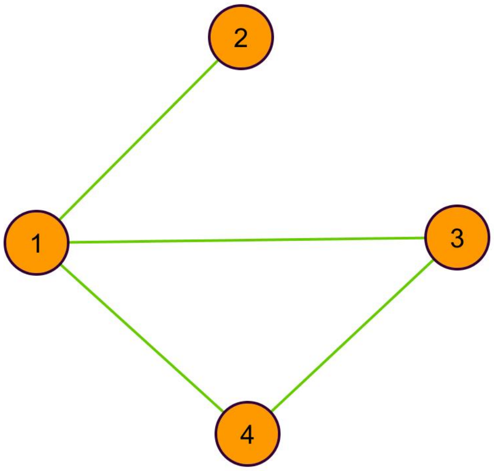

**Figure 3.2**: *Computing clustering coefficient for node 1*

### 3.1.5 Network Centralization

Network centralization $C_g$ is a measure of network centrality in social networks [29]. This metric was motivated by research interests in the impact of centrality on abstract concepts like leadership. For this reason, highly centralized networks have the highest values of centralization. Centralization ranges from $0$ to $1$ with a value of $1$ assigned to a star topology and $0$ to a mesh topology. Network centralization is computed as

$$C_g = \frac{n}{n-2}\left(\frac{k_{max}}{n-1} - Density\right) \tag{3.4}$$

where Density is defined $\frac{\sum_{i,j} 2A_{i,j}}{n(n-1)}$. For graph $G = (V, E)$, the adjacency matrix $A_{ij} = 1 \; \forall \; \{i, j\}$ $\in E$, if $\exists$ an edge between nodes $i$ and $i$. Otherwise, $A_{ij} = 0$.

### 3.1.6 Network Heterogeneity

Network heterogeneity $H_g$, favors networks with hubs. Hence, networks with an increasingly hub-like structure have a higher value. Though this metric detects hubs, it is unable to capture their position. As will be shown in Chapter 7, the location of hubs are crucial to the robustness of networks under targeted attacks. Heterogeneity for a graph $G$ is computed in [24] as

$$H_g = \frac{\sqrt{variance(k)}}{\bar{k}} \tag{3.5}$$

### 3.1.7 Assortativity Coefficient

Assortativity coefficient $r$, is defined in [30] as the correlation of node degrees to pairs of nodes. Positive values indicate a correlation between nodes of similar degree. Conversely, negative values indicate a correlation between nodes of dissimilar degrees. Assortativity coefficient is defined as

$$r = \frac{\sum_i j_i k_i - M^{-1} \sum_i j_i \sum_{i'} k_{i'}}{\sqrt{\left[\sum_i j_i^2 - M^{-1} \left(\sum_i j_i\right)^2\right] \left[\sum_i k_i^2 - M^{-1} \left(\sum_i k_i\right)^2\right]}} \qquad -1 \leq r \leq 1 \tag{3.6}$$

where $j_i$ and $k_i$ are the degrees of nodes $j$ and $k$ excluding any edge between these two nodes.

### 3.1.8 Edge Expansion

Edge expansion $h(G)$, also known as the edge expansion coefficient in [31] or the isoperimetric number of a graph $G$, is defined in [32, 33] as

$$h(G) = \min_{1 \leq |S| \leq \frac{n}{2}} \frac{|\partial(S)|}{|S|} \tag{3.7}$$

| $i$ | $\partial\left(S_i\right)$ | $\partial\left\|\left(S_i\right)\right\|$ | $\left\|S_i\right\|$ | $h\left(S_i\right)$ |
|---|---|---|---|---|
| 1 | $\{(2,4),(1,3)\}$ | 2 | 2 | 1 |
| 2 | $\{(2,4),(3,4),(3,5)\}$ | 3 | 2 | 1.5 |
| 3 | $\{(3,4)\}$ | 1 | 4 | 0.25 |

**Table 3.2**: *Edge Expansion calculation for sets $S_1$, $S_2$, and $S_3$ in Graphs $G_1$ and $G_2$*

where the minimum is over all sets $S$ with at most $\frac{n}{2}$ nodes. $\partial(S)$ is a set of edges with exactly one endpoint in $S$. Table 3.2 provides expansion calculations for some components, $S$, of the graphs in Figures 3.3 and 3.4. These calculations simply demonstrate the evaluation of edge expansion.

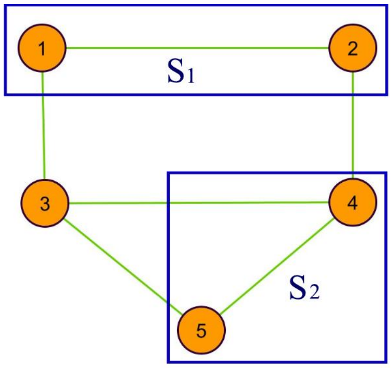

**Figure 3.3**: *Graph $G_1$ for which edge expansion is calculated for sets $S_1$ and $S_2$*

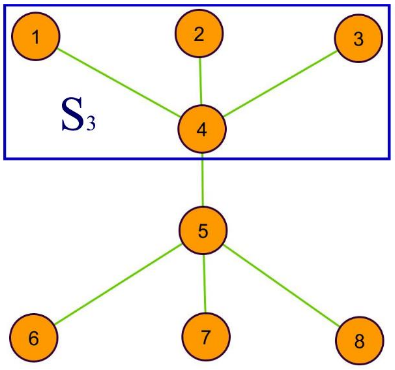

**Figure 3.4**: *Edge expansion decreases in Graph $G_2$, as a network has a more hub-like connectivity (i.e. nodes are added in a preferential attachment manner)*

| $i$ | $\Gamma(S_i)$ | $\Gamma\|(S_i)\|$ | $\|S_i\|$ | $g_\alpha(S_i)$ |
|---|---|---|---|---|
| 4 | $\{(4)\}$ | 1 | 3 | 0.33 |
| 5 | $\{(4,5)\}$ | 2 | 3 | 0.67 |
| 6 | $\{(4)\}$ | 1 | 4 | 0.25 |

**Table 3.3**: *Vertex Expansion calculation for sets $S_4$, $S_5$, and $S_6$ in Graphs $G_4$, $G_5$, and $G_6$*

### 3.1.9 Vertex Expansion

The $\alpha$-vertex expansion $g_\alpha(G)$ for graph $G$ is defined as

$$g_\alpha(G) = \min_{1 \le |S| \le \alpha n} \frac{|\Gamma(S)|}{|S|} \tag{3.8}$$

where $\Gamma(S)$ is a set of vertices with at least 1 neighbor in $S$. Figure 3.5 shows the initial condition of graph $G_4$, and Figures 3.6 and 3.7 show how edge expansion varies when $G_4$ is slightly modified to represent a more connected version $G_5$ in 3.6 and a less connected version $G_6$ in Figure 3.7. For the following graphs, $\alpha = \frac{1}{2}$. Table 3.3 contains vertex edge expansion for $S_4$, $S_5$, and $S_6$

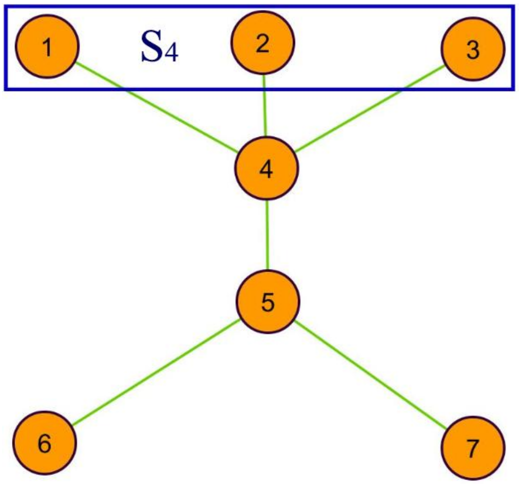

**Figure 3.5**: *This graph shows the initial condition of graph $G_4$*

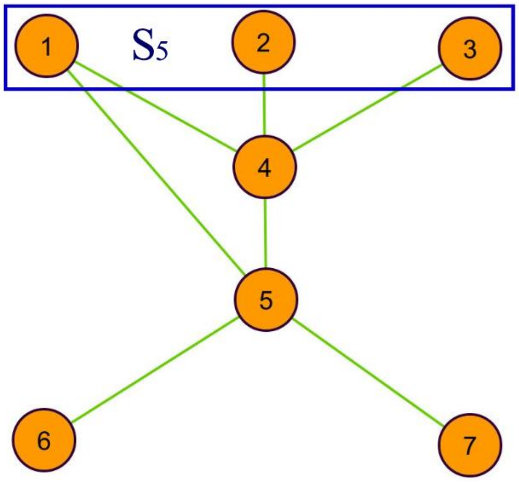

**Figure 3.6**: *Edge* $(1, 5)$ *was added to* $G_4$ *to produce graph* $G_5$

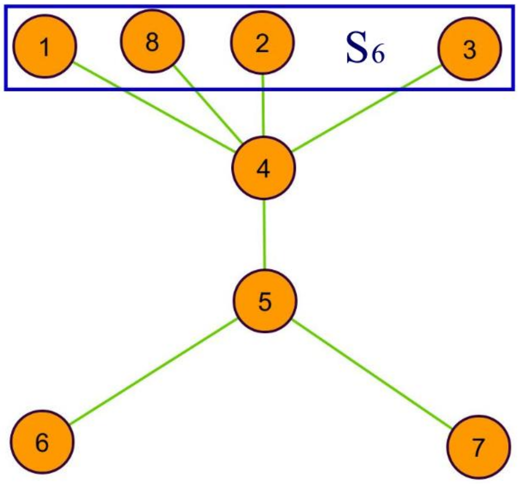

**Figure 3.7**: *Node 8 was added to* $G_4$ *to produce graph* $G_6$

### 3.1.10 Spectral Expansion

Given a regular graph (regular implies that all nodes have the same node degree), spectral gap, referred to as a measure of spectral expansion, is derived from the eigenvalues of the adjacency matrix $A$. Since $A$ is symmetric, it has full rank of $n$ and thus, $n$-real eigenvalues: $\lambda_0 \geq \lambda_1 \geq ... \geq \lambda_{n-1}$. For a regular graph where $\lambda_0 = k$, spectral gap $= k - \lambda_1$. However, spectral gap is generally referred to as $\lambda_0 - \lambda_1$. Normalized eigenvalues are obtained by $\frac{A}{k}$ (The resulting eigenvalues range from -1 to 1) [34].

When the graph is regular, a higher value of spectral gap is preferred for robustness: This implies a higher connectivity. For example, in Table 3.4, the spectral gap of the 6-node mesh graph is higher than that of the cubic 6-node graph and hence, the mesh graph is more robust than the cubic graph.

| Graphs | $\lambda_0$ | $\lambda_1$ | spectral gap |
|---|---|---|---|
| 6-node mesh | 1 | 0.200 | 0.800 |
| 6-node cubic | 1 | 0.677 | 0.333 |

**Table 3.4**: *Analysis of spectral gap for two normalized, regular graphs*

### 3.1.11 Relationship between edge, vertex, and spectral expansion

Edge and vertex expansion are directly related and provide the same general result on the robustness of a graph (if the graph is regular, edge, vertex, and spectral expansion, all provide similar qualitative results). For any graph $G = (V, E)$, the author of [31] gives the following result

$$h(G) \geq \frac{g_1(G) - 1}{2} \tag{3.9}$$

Additionally, if $G$ is $d$-regular, spectral gap sets the bounds for edge expansion as follows:

$$\frac{1}{2}(d - \lambda_1) \leq h(G) \leq \sqrt{2d(d - \lambda_1)} \tag{3.10}$$

### 3.1.12 Conductance

Conductance $\Phi$ for graph $G = V, E$ provides meaningful information on the connectivity within a network. Among several applications, conductance can measure the speed at which information traverses a network (social or telecommunication), the rate of an epidemic propagation (biological), and the performance of routing algorithms (Internet) [35]. From a structural perspective, a high value of conductance inherently signals a highly connected graph. For a cut $\left(S, \bar{S}\right)$, as conductance defined in [36], has been reduced to

$$\Phi\left(S\right) = \frac{\sum_{i \in S, j \in \bar{S}} A_{ij}}{\min\left(A\left(S\right), A\left(\bar{S}\right)\right)} \tag{3.11}$$

where $A\left(S\right) = \sum_{i \in S} \sum_{j \in V} A_{ij}$. Therefore, the conductance of the entire graph is given as:

$$\Phi_G = \min_{S \subseteq V} \Phi\left(S\right) \tag{3.12}$$

Additionally, $\frac{h(G)}{k}$ ($h(G)$ is the edge expansion) gives the conductance for a $k$-regular graph.

### 3.1.13 Natural Connectivity

From a spectral analysis approach, researchers in [37] endeavor to use an average of the spectrum of eigenvalues as a more accurate measure of robustness. This averaged spectrum is called the natural connectivity, $\bar{\lambda}$, is defined as:

$$\bar{\lambda} = \ln\left(\frac{\sum_{i=1}^{N} \exp^{\lambda_i}}{N}\right) \tag{3.13}$$

where $\lambda_i$ defines the eigenvalues of the adjacency matrix. As the number of nodes in a graph increases, finding eigenvalues becomes computationally expensive. Among existing methods to compute eigenvalues, the Power Iteration method (which is used by Google to calculate the page rank of documents [38]), provides only the dominant eigenvalue and its corresponding eigenvector. This method is seldom used due to its limited output. However, matrix theory shows that shifting properties can be applied to obtain other eigenvalues. The Jacobi method is another approach to extract eigenvalues and eigenvectors from a matrix. However, as the size of the graph grows the convergence time for these algorithms increase. Therefore, finding eigenvalues of complex networks is a non-trivial task.

### 3.1.14 Algebraic Connectivity

Experimentalists in [8, 39] consider the second smallest Laplacian eigenvalue $\mu_2$, introduced as the algebraic connectivity [40], as a measure of how difficult it is to break the network into components. As $\mu_2$ increases, the connectivity of a graph also increases, and it becomes increasingly difficult to fragment this graph. However, as shown in Figures 3.8 through 3.11, when flows are introduced in these networks, the results of algebraic connectivity are inconsistent with our definition of robustness (the ability for a network to maintain its total level of flow under increasing removal of nodes). More specifically, the comparison in Figures 3.8 and 3.9 shows that algebraic connectivity is consistent in discriminating between the robustness of the two topologies. However, in Figures 3.10 and 3.11, the results are inconsistent.

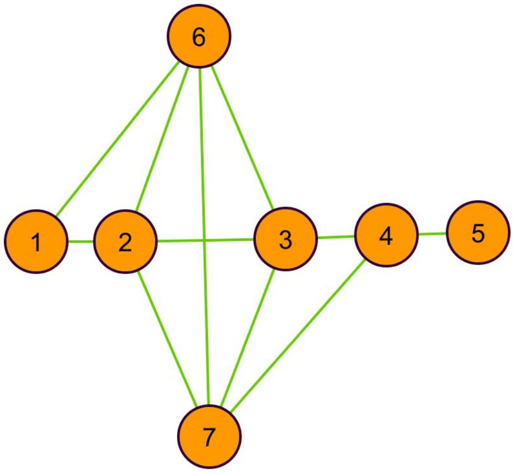

**Figure 3.8**: *For network 1, $\mu_2$=0.677*

Based on the values of $\mu_2$ shown in Figures 3.8 and 3.9, network 1 is more robust than network 2 (its value of $\mu_2$ is higher than that of network 2). Initially, 42 origin-destination O-D (origin-destination) flows exist in both networks. In network 1, if one node is attacked to effectuate

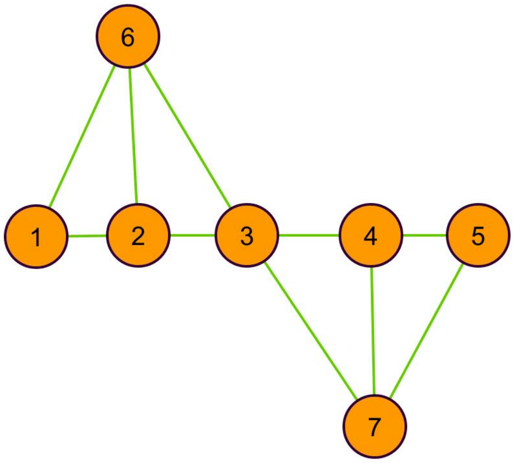

**Figure 3.9**: *For network 2, $\mu_2$=0.586*

maximum damage (node 4), 20 flows will continue to be delivered, which corresponds to $48\%$ of the initial throughput. In network 2, if the central node is targeted (node 3), 12 flows will continue to be delivered, which corresponds to $30\%$ of the initial throughput. Thus, the throughput and consequently the elasticity (described subsequently) of network 1 are greater than that of network 2. Likewise, $\mu_2$ for network 1 is greater than that of network 2. In this case, the second smallest eigenvalue captures our definition of robustness in that a larger $\mu_2$ implies a more robust topology.

The networks in Figures 3.10 and 3.11 serve as counterexamples. In network 3, there exist 30 initial O-D flows. If the central node is removed (node 3), you can continue to route 20 flows: $67\%$ of the initial throughput. In network 4, however, there is no central node and in the case where the removal of one node inflicts maximum damage on the network (node 1), out of 306 initial flows 240 will be delivered, which corresponds to $78\%$ of the initial throughput. Thus, network 4 is more robust than the one in network 3.

However $\mu_2$ for network 4 is considerably lower than that of network 3. On this premise, $\mu_2$ does not consistently capture our definition of robustness. Additionally, should one node become disconnected, $\mu_2$ becomes 0, which according to our definition, is too coarse a measure to capture a network's robustness.

Therefore, based on our observations, algebraic connectivity can extract the robustness of

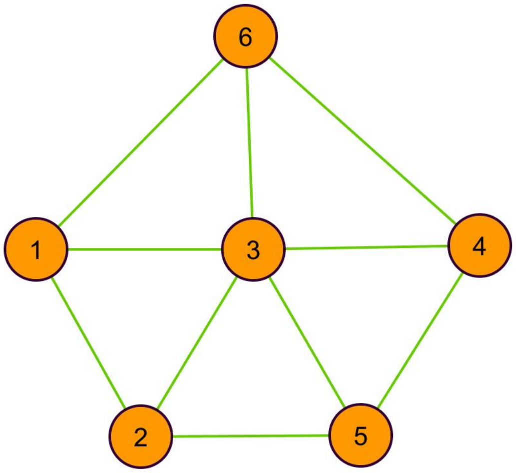

**Figure 3.10**: *For network 3,* $\mu_2$*=2.382*

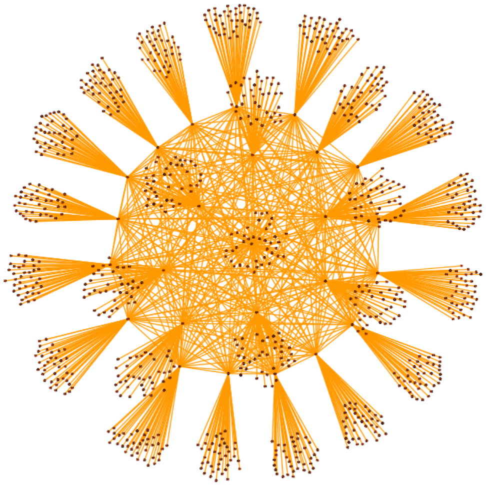

**Figure 3.11**: *For network 4,* $\mu_2$*=0.890*

a graph with respect to its connectivity. However, in considering network behaviors, algebraic connectivity is unable to produce consistent results. For this reason, there exists a need for metrics to assess behaviors occurring in a network.

## 3.2 CHARACTERIZING THE BEHAVIOR

Complex systems are intricate, not only in their structure, but also in their underlying behavior. For this reason, they require metrics, which consider behaviors in the network to determine their robustness. The authors of [41, 20, 25] proved that metrics that determine a network's structural properties are insufficient to determine the network's functional properties. More specifically in [18], it was concluded that the Internet exhibited this, "'Robust yet Fragile,"' nature ("Robust yet Fragile" implies that the Internet is robust against random attacks but fragile to targeted attacks). This conclusion was based on experiments performed using structural metrics. However, authors in [41, 20, 25] showed that the Internet's robustness was due to its self-organizational properties and its fragilities were due to Internet hijackings. Additionally, they proved that though the Internet's topology indeed was scale-free, as claimed in [18], the high node degree hubs were on the periphery of the network. Therefore, attacks to highly connected vertices would only affect the network locally: the core would remain unaffected. In summary, results in [41, 20, 25] proves that metrics which capture a network's robustness with respects to its connectivity are insufficient and generally provide misleading conclusions about a network's robustness with respect to its behavior.

### 3.2.1 Network Performance

Researchers in [20] have introduced two metrics to compare topologies with similar node degree distributions but different topological structure:

1. Network likelihood evaluates the physical connectivity of system.
2. Network performance evaluates the system's behavior.

Performance $P(G)$ is a measure the maximum throughput under gravity flows (In the gravity flow model, the traffic demand $x$ between node $i$ and $j$ is the product of these demands $x_i x_j$).

$$P\left(G\right) = \max_{\rho} \sum_{i,j} X_{i,j} \quad s.t. \quad RX \leq B \tag{3.14}$$

- $\rho$ is a global constant
- $x_i$ and $x_j$ are the traffic demands at nodes $i$ and $j$
- $X_{i,j} = \rho x_i x_j$

- $X_{i,j}$ - traffic flow between source node $i$ and destination node $j$
- $R$ - routing matrix using standard shortest path ($R_{kl}$ = 1 if flow $l$ passes through router $k$)
- $X$ is a vector of all $X_{i,j}$ flows
- $B$ is a vector of all router bandwidth capacities

The above formulation maximizes the total throughput in the network subject to capacity constraints on the nodes: this maximized total throughput is achieved by maximizing the constant $\rho$, which increases all flows in proportion to each other.

Irrespective of the field of study, networks generally contain constraining components. For the case where a node is the constraining element, network performance is a suitable metric to evaluate such topologies. This is the case in [20]. Assuming over provisioned link capacities, it is logical to define a metric for the constraining component of this network: the router. However, this constraint may not exist for topologies within different fields of study. Below are a few of such cases:

### Links as constraining factors

In building a network to model epidemic spread in biology, a node could be a member from the species under investigation (an animal, human etc.) and a link could exist if two nodes interact as defined by temporal or spatial parameters [42]. The flow can then be modeled as the transmission of the disease from one node to another, given the epidemic front propagates throughout the region under investigation. In this scenario, there is no direct constraint on the number of flows through the node. However, there exists an abstract constraint on the links, which takes the form of a probability: $Prob(i/j)$-the probability that node $i$ will be infected given that its neighbor $j$ is infected. The actual details are more involved and not the focus of this work. This analogy simply serves the purpose of presenting a case where links rather than nodes are the constraining factors of a network.

Therefore, when nodes are not the constraining element, a new metric elasticity (introduced subsequently in Chapter 5, page 26), is more applicable.

# Chapter 4

# DISTRIBUTIONS TO ASSESS ROBUSTNESS

## 4.1 CHARACTERIZING THE STRUCTURE

Distributions obtained from graphs can provide meaningful information on the connectivity within the graph. Some of the more frequent distributions are presented in this section.

### 4.1.1 Node Degree

The degree of a node $i$ is the number of nodes adjacent to $i$ or the number of edges connected to $i$. For a directed network, where edges are unidirectional, the indegree of node $i$ is the number of edges connected to node $i$, emanating from $i$'s neighbors. Conversely, the outdegree of node $i$ is the number of edges from node $i$ to its neighbors. Hence, the node degree distribution $P(k)$ gives the probability that a randomly selected node has degree $k$. Given a graph $G$ with size $N$, the degree distribution is a power-law if $P(k) \sim k^{-\gamma}$, where $\gamma > 1$ [22]. Furthermore, the power law distribution cuts-off at the maximum degree, $k_{cut-off} = n^{\frac{1}{\gamma-1}}$. The node degree distribution is defined as,

$$P(k) = \frac{n(k)}{N} \qquad\qquad k = 0, 1, .., k_{max} \tag{4.1}$$

The node degree distribution allows comparison between different graphs. For example, from a binomial distribution, one can potentially deduce that no highly connected nodes exist. This typifies randomly generated ER graphs models. On the other hand, power-law distributions (or heavy tailed distributions) have highly connected hubs. This is true of naturally occurring phenomenon such as the Pareto income principle (introduced by Vilfredo, from observing that 20 percent of the people in Italy own 80 percent of the land - 'the 80-20 rule') in addition to dynamical systems in physics which exhibit the self-organized critical (SOC) property.

### 4.1.2 Topological Coefficients

For a given node $i$, topological coefficient $T_n$, provides information on the tendency of $i$ to share its neighbors. This has direct implications to the particular field of interest. For instance, in the study of epidemics, the more "'sharing"' that occurs between partners and their neighbors, the higher the likelihood of virus transmission and propagation. In social and media networks, where information dissemination is of utmost importance, an objective to consider could be to maximize the topological coefficients of nodes. Topological coefficient is defined in [43] as:

$$T_n = \frac{avgJ\left(n,m\right)}{k_n} \tag{4.2}$$

where $J\left(s,t\right) = 1$ if node $t$ shares at least one common neighbor with $s$. If $s$ and $t$ are also neighbors, $J\left(s,t\right)$ is incremented by 1. Figure 4.1 shows the calculation for $T_n$ where $n$ is node 4: $J\left(4,2\right) = 1$, $J\left(4,3\right) = 2$, and $J\left(4,5\right) = 2$. Hence, $T_4 = \frac{1.67}{3}$. If node $n$ has one or no neighbors, $T_n = 0$.

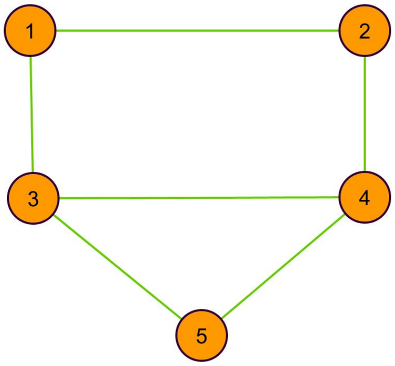

**Figure 4.1**: *Sample calculations for evaluating topological coefficient of node 4*

### 4.1.3 Stress

Stress $S_v$, is a centrality index which indicates the number of shortest $s$-$t$ paths traversing node $v$ and defined as [44]:

$$S_v = \sum_{st} \alpha_{st}(v) \tag{4.3}$$

where $\alpha_{st}(v)$ is the number of shortest paths between $s$ and $t$ going through $v$.

### 4.1.4 Betweenness

The shortest path betweenness for node $v$ $B_v$, is the number of $s$-$t$ (source-destinaion) paths traversing $v$ divided by the total number of $s$-$t$ paths. Betweenness differentiates between the importance of nodes in providing connectivity to other nodes in a network. Thus, if only one $s$-$t$ path exists such that this path traverses node $v$, then the betweenness of node $v$ would be high. However, if there are multiple $s$-$t$ paths and only one traverses node $v$, the betweenness of node $v$ would be low.

Betweenness is defined in [44] as:

$$B_v = \sum_{st} \frac{\alpha_{st}(v)}{\alpha_{st}} \tag{4.4}$$

where $\alpha_{st}$ is the number of shortest paths between nodes $s$ and $t$ and $\alpha_{st}(v)$ is the number of shortest paths between $s$ and $t$ going through $v$.

### 4.1.5 Closeness Centrality

Closeness $C_s$, is a centrality index that indicates proximity of a node. Among other applications, it can indicate the speed at which information travels from node $s$ to $t$ in a network. Therefore, the closeness distribution gives the proximity of nodes with neighbors of degree $k$. Closeness, in [45], is defined for node $s$ as:

$$C_s = \frac{1}{avgL\left(s,t\right)} \tag{4.5}$$

where $L\left(s,t\right)$ is the shortest distance between nodes $s$ and $t$.

### 4.1.6 Average Clustering Coefficient

In this section we refer back to the metric, average clustering coefficient, presented in 3.1.4 on page 7. For this metric, the average clustering coefficient distribution provides the average clustering coefficient for nodes with $k$ neighbors.

### 4.1.7 Neighborhood Connectivity

For a given vertex $i$ with $k$ neighbors, neighborhood connectivity $N_c$ defines the average size of the neighborhood of these $k$ neighbors. Therefore, theneighborhood connectivity distribution gives the average size of the neighborhood of a node with $k$ neighbors, for all nodes in the network.

In Table 4.1, for each node, its respective neighbor and the size of its neighbor's neighborhood is given. Furthermore, the result of network connectivity for each node (for node $i$ with $k$ neighbors, network connectivity is the average size of $k$) is presented on the row in which this node is first introduced. Finally, Figure 4.3 shows the distribution for $N_c$.

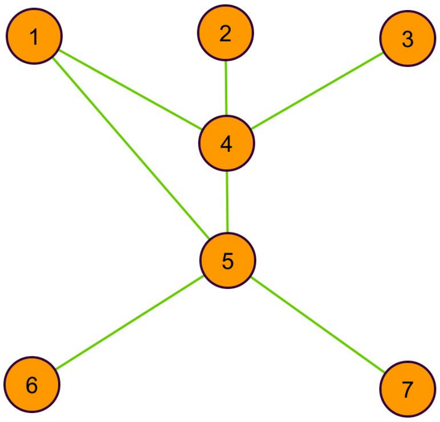

**Figure 4.2**: *Graph for demonstrating the calculation of network connectivity*

| **Node** | **Neighbor** | **Size of Neighborhood** | $N_c$ |
|---|---|---|---|
| 1 | 4 | 4 | 4 |
| 1 | 5 | 4 | – |
| 2 | 4 | 4 | 4 |
| 3 | 4 | 4 | 4 |
| 4 | 1 | 2 | 2 |
| 4 | 2 | 1 | – |
| 4 | 3 | 1 | – |
| 4 | 5 | 4 | – |
| 5 | 1 | 2 | 2 |
| 5 | 4 | 4 | – |
| 5 | 6 | 1 | – |
| 5 | 7 | 1 | – |
| 6 | 5 | 4 | 4 |
| 7 | 5 | 4 | 4 |

**Table 4.1**: *Analysis of network connectivity for the graph in Figure 4.2*

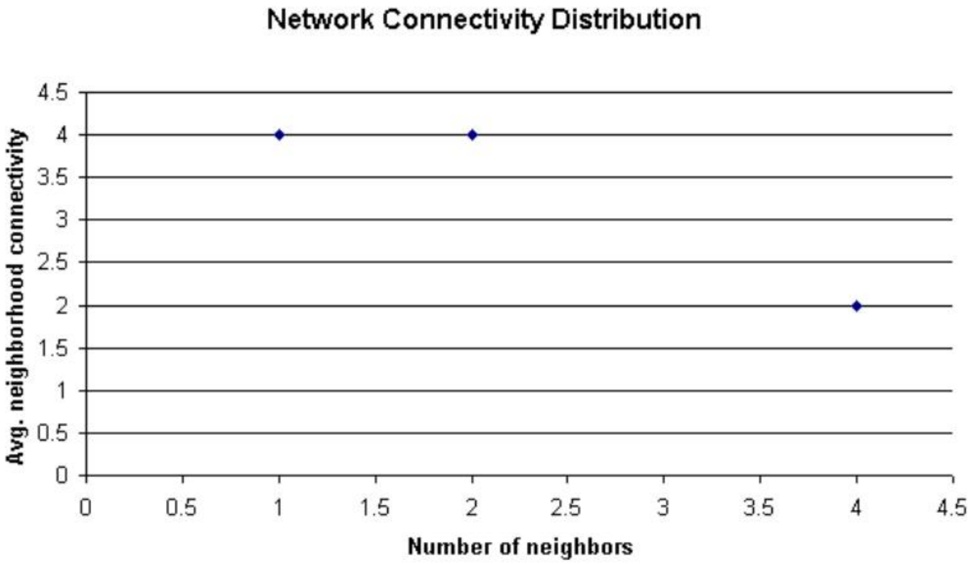

**Figure 4.3**: *Network connectivity distribution for the graph in 4.2*

# Chapter 5

# NETWORK MODELS

This chapter reviews the six models from which 18 topologies were selected. They include networks from random models, Watts-Strogatz models, preferential attachment models, near-regular models, trade-off and optimization models, and real-world models. For each topology, some of the more common properties are shown in Table A.1.

**Table 5.1**: *Network characteristics where ASP is the average shortest path and Het is heterogeneity*

| Networks | ♯ Nodes | ♯ Links | Density | Diameter | ASP | Het |
|---|---|---|---|---|---|---|
| Gi-dense | 1000 | 4505 | 0.00902 | 7 | 3.391 | 0.331 |
| MySpace | 955 | 10976 | 0.02409 | 4 | 2.013 | 2.027 |
| Watts-Strogatz 1 | 1000 | 3000 | 0.00601 | 7 | 4.14 | 0.301 |
| PA 2 | 1000 | 2964 | 0.00593 | 6 | 3.534 | 1.109 |
| Gi-sparse | 1000 | 2009 | 0.00402 | 12 | 5.154 | 0.491 |
| PA 1 | 1000 | 1981 | 0.00397 | 8 | 4.177 | 1.185 |
| Watts-Strogatz 2 | 1000 | 2000 | 0.004 | 9 | 5.294 | 0.37 |
| YouTube | 1089 | 1576 | 0.00266 | 12 | 5.096 | 1.319 |
| Flickr | 967 | 1515 | 0.0032 | 12 | 4.624 | 1.394 |
| meshcore | 1000 | 1275 | 0.00255 | 3 | 2.911 | 3.796 |
| near-regular 2 | 992 | 3781 | 0.00769 | 31 | 14.706 | 0.133 |
| HOT 2 | 1000 | 1049 | 0.0021 | 12 | 7.144 | 1.892 |
| ringcore | 1000 | 1000 | 0.002 | 14 | 8.196 | 3.122 |
| HOT 1 | 939 | 988 | 0.00224 | 10 | 6.812 | 2.032 |
| PA-sparse | 1000 | 1049 | 0.0021 | 14 | 5.793 | 1.892 |
| Abilene | 886 | 896 | 0.00229 | 10 | 6.95 | 2.09 |
| near-regular 1 | 992 | 1921 | 0.00391 | 61 | 21 | 0.089 |

## 5.1 Random models

A random graph is obtained by random addition of links between $n$ vertices. Erdos-Renyi's (ER) stochastic model is one of the most studied of these models. In the construction of an ER graph $G(N, E)$, $E$ edges are connected at random to $N$ nodes [19]. For this model, each of the $\frac{N(N-1)}{2}$ edges have an equal probability of being selected. However, this thesis considers the Gilbert (Gi) model $G(N, p)$, a modified version of the ER model where edges are connected to vertices with a probability of $p$. Below are a few key properties of random graphs

- The average node degree determines the connectivity of the graph. Therefore, if $\bar{k} < 1$, a disconnected components exist. At $\bar{k} = 1$, a phase transition occurs, and a giant component exists when $\bar{k} > 1$ [19].

- The node degree exhibits a binomial distribution and thus, given $N$ nodes and a probability of $p$,

$$P\left(k\right) = \binom{N-1}{k} p^{k} \left(1 - p\right)^{N-1-k}. \tag{5.1}$$

  However, the model in this thesis was based on the Poisson distribution, an approximiation of the binomial distribution when the limit of N is large and $pN = \bar{k}$ [19].

$$P\left(k\right) = e^{-\bar{k}} \frac{\bar{k}^{k}}{k!} \tag{5.2}$$

- As $k$ becomes large, the degree distribution decays exponentially

For the Gi-dense and Gi-sparse networks used in this thesis, $p = 0.0091$ and $0.004094$ respectively [46]. Figure 5.1 shows the Gi-sparse network.

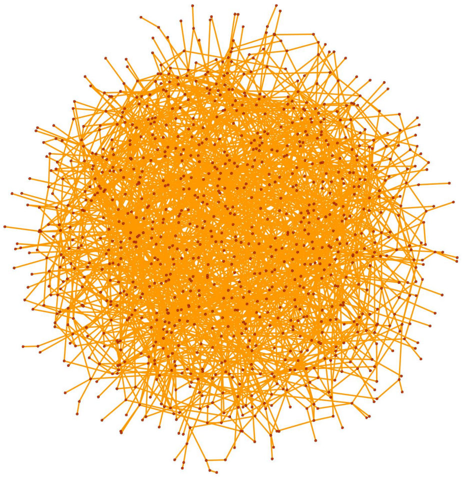

**Figure 5.1**: *The Gi-sparse network with size* $N = 1000$ *and average degree* $\bar{k} = 4.018$

## 5.2 Watts-Strogatz model

The Watts-Strogatz model is constructed by interpolating between a regular ring lattice and a random network. The construction begins with a ring of $N$ nodes. Each node is then connected to its $k$ nearest neighbors. Then in a clockwise manner, node $i$ is selected. The edge that connects to $i$'s nearest neighbor is randomly rewired with a probability of $p$ (or left untouched with a probability of $1 - p$ ) considering the constraint that no self-loops or duplicate loops can exist. This procedure is repeated cyclically for each successive node until node $i$ is once again selected. At this point, the edge that connects to $i$'s second nearest neighbor undergoes similar rewiring procedures. This cycle of node selection and rewiring recurs until the edge that connects all nodes $i$ to their furtherest neighbor is considered [27].

In the Watts-Strogatz model, the parameter $p$ determines the level of randomness in the graph while maintaining the initial number of nodes and links [19]. For intermediate values of $p$, Watts-Strogatz models produces a Small-world network, which captures the high clustering properties of regular graphs and the small characteristic path length of random graph models. Figure 5.2 shows three snapshots of graphs obtained for different values of $p$.

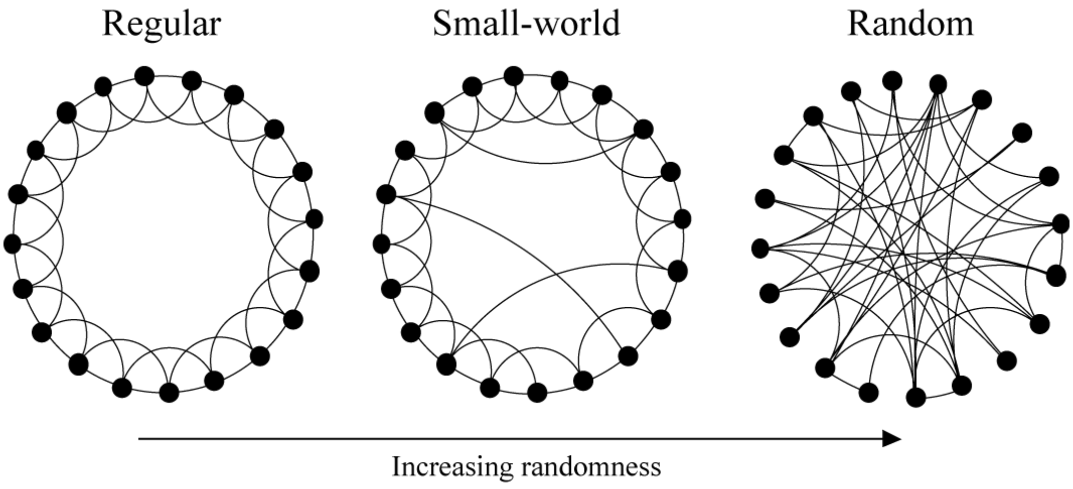

**Figure 5.2**: *The construction of Watts-Strogatz model. For the regular graph $p = 0$. The random graph is obtained at $p = 1$ and for intermediate values of $p$, a Small-world network is realized [27].*

For the Watts-Strogatz (W-S) 1 and 2 networks used in this thesis, the rewiring probability was $0.3$ and $0.5$ [46]. Figure 5.3 shows the W-S 1 network.

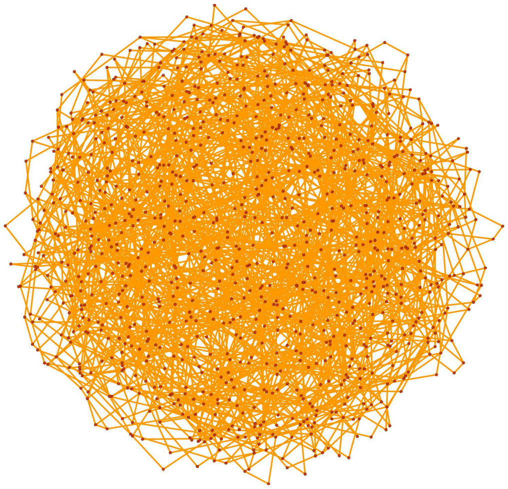

**Figure 5.3**: *The W-S 1 network with $N = 1000$ and $\bar{k} = 4$*

## 5.3 Preferential Attachment Model

Preferential Attachment (PA) models changed the norm that complex networks are associated with complete randomness and highlighted an emerging class of topologies associated with a heavy tailed node degree distribution [47]. This distribution, also known as a power-law distribution, can be reviewed in Subsection 4.1.1 page 20. These networks pervade numerous real world domains. For example, within the sphere of social networks, an individual with few friends is more likely to form a new friendship with a more popular person. Likewise, new Internet websites will more likely establish ties with the most popular websites.

From their origin, preferential attachment (PA) models have been considered vulnerable to targeted attacks while robust to random failures [47]. This model constitutes popular nodes called "hubs", which have a large number of neighbors compared to other nodes with few neighbors. The rules for construction are governed by the two key principles of growth and preferential attachment. The initial number of nodes at construction must be greater than two and each node must have at least one neighbor. At each time step, a new node is added to the graph. The probability of attracting this new node is determined by the node degree of preexisting nodes. Thus, the higher the node degree of preexisting nodes, the higher the probability of attracting new nodes. The attachment probability is given in [46] by

$$P\left(k_i\right) = \frac{k_i}{\sum_{j=0}^{N} k_j} \tag{5.3}$$

where $P(k_i)$ is the probability that a new node will connect to an existing node $i$ with degree $k_i$. For this work, the PA 1, PA 2, and PA-sparse networks were constructed using the Barabasi-Albert Scale-free model [46, 20]. Figure 5.4 shows the PA-sparse network.

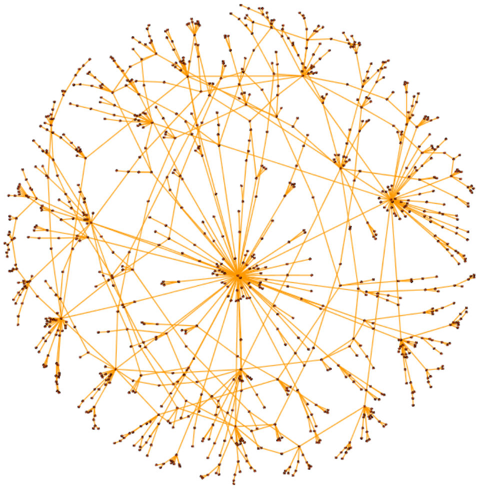

**Figure 5.4**: *The PA-sparse network with* $N = 1000$ *and* $\bar{k} = 2.098$

## 5.4 Near-Regular Models

The near-regular (n-r) networks are best visualized in a planar, grid-like fashion. The n-r 1 network is composed of a 31 by 32 grid where node $i$ is connected to node $j$ if $j$ is a distance $d = 1$ unit: $1$ unit is the regular distance among nodes in the grid. The structure of n-r 2 is similar to that of n-r 1. However, in addition to $d = 1$ unit, all nodes within a distance of $d = \sqrt{2}$ units are connected. Figure 5.5 shows the n-r 1 network.

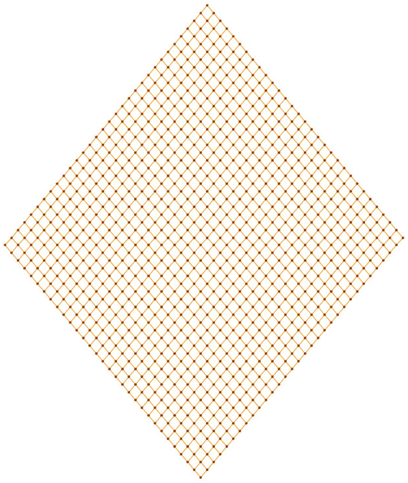

**Figure 5.5**: *The n-r 1 network with* $N = 992$ *and* $\bar{k} = 3.87$

## 5.5 Trade-off and Optimization Models

The authors of [48] introduce networks with bimodal degree distributions optimized to minimize the impact of random attacks. The meshcore and ringcore topologies shown in Figures 5.6 and 5.7 represent this model. The Heuristically Optimized Trade-off (HOT) network presents a simple model for Internet growth [49, 20]. The HOT 1 and 2 networks represent this model. Figure 5.8 shows the HOT 2 network.

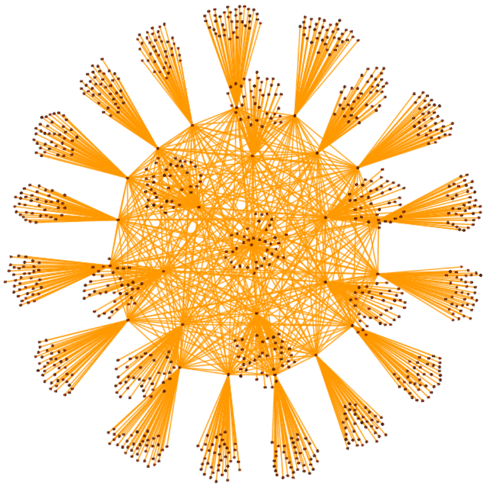

**Figure 5.6**: *The meshcore network with* $N = 1000$ *and* $\bar{k} = 2.55$

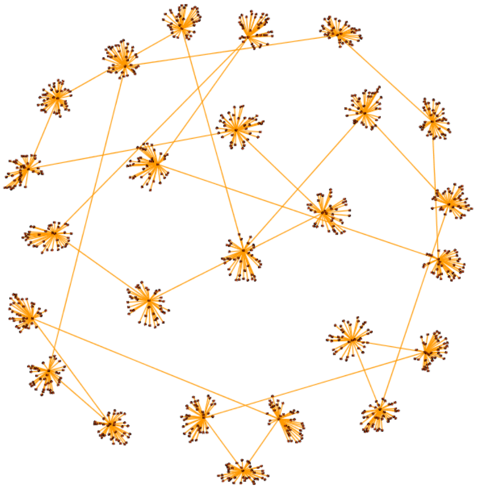

**Figure 5.7**: *The ringcore network with* $N = 1000$ *and* $\bar{k} = 2$

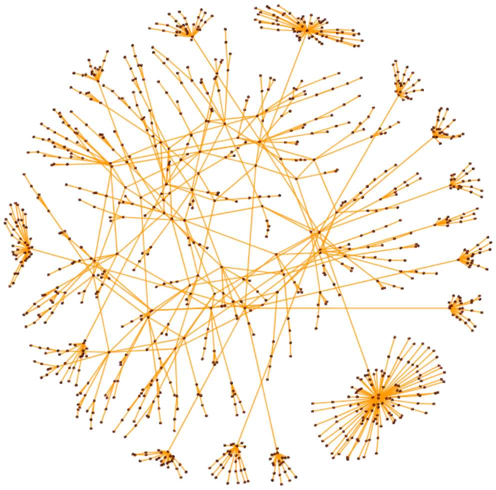

**Figure 5.8**: *The HOT 2 network with* $N = 1000$ *and* $\bar{k} = 2.098$

## 5.6 Real-World Models

Online social networking connects individuals with common interests. This thesis features the MySpace, YouTube, and Flickr networks. These networks were obtained via snowball sampling and have been rescaled [50]. The Abilene network in Figure 5.9 was built using the Abilene core while customers and peer networks were each replaced with a gateway router [20].

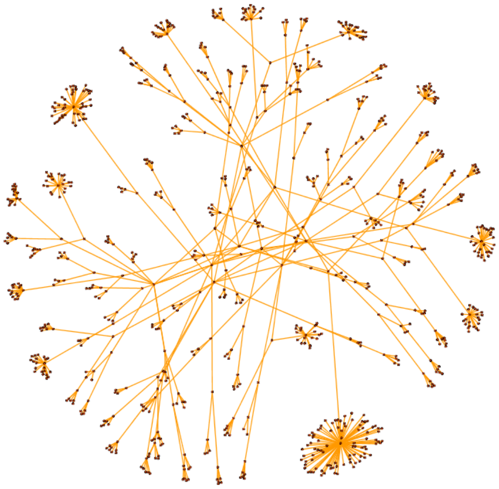

**Figure 5.9**: *The Abilene network with* $N = 1000$ *and* $\bar{k} = 2.022$

# Chapter 6

# ELASTICITY: Characterizing the behavior

## 6.1 MOTIVATION AND CONTRIBUTION

The study of robustness is fundamental to numerous network research problems using approaches that amplify internal behaviors of a network. To this end, we first define robustness as the ability of a network to maintain its total level of flow under increasing removal of nodes and consequently links, and second, we consider flows as the behaviors occuring within a network. These two components form the foundation of our new metric, elasticity. Inspired by the performance metric in [20], elasticity considers flows within a giant component as well as flows within fragmented components. This latter consideration is necessary because depending on the phenomena under investigation, nodes within segmented components can interact and contribute to the robustness of a network. For instance, in the event of inter-AS interruptions, intra-AS communications persists and as such, a level of robustness exists. Similarly, in the epidemic domain, virus propagation continues within subcomponents even though a "giant component" ceases to exist. In other words, elasticity considers the contribution of isolated components to the robustness of the network.

The initial level of flow in a network determines its initial robustness. As nodes are removed, the level of flow can change considerably and as a result, topologies which appear robust at the onset may potentially disintegrate as few nodes are removed. Conversely, topologies which appear less robust at the onset can become more robust eventually, as they are better able to maintain the total level of flow. Thus, elasticity considers the robustness of a network as the aggregation of local behaviors resulting from the removal of each node.

Figure 6.1 is a hypothetical case that serves the purpose of demonstrating the evaluation of elasticity E, which is the area under the curve of throughput vs the fraction of nodes removed. More importantly, the throughput and fraction of nodes removed are normalized to compare topologies of varied nodes and links. Hence, $G_1$ and $G_2$ have the same initial throughput. Initially, $G_1$ has a the higher total throughput. However, when subsequent nodes are removed, $G_2$ has the higher throughput at each iteration. In the complete analysis, $G_2$ has a higher elasticity thus, it is more robust than $G_1$.

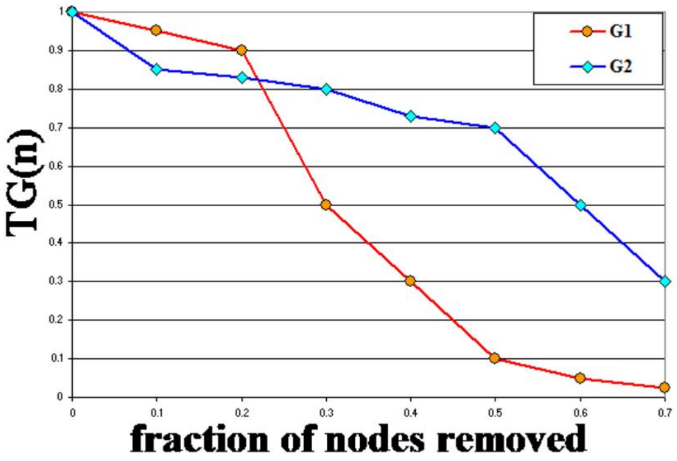

**Figure 6.1**: *The evaluation of elasticity for two generic graphs $G_1$ and $G_2$*

In Section 6.2, we propose a new metric elasticity that provides one number to describe the robustness of a network. Furthermore, we establish the upper bound on elasticity and select the most appropriate routing algorithm for elasticity.

## 6.2 Definition of elasticity

For a network $G$, having no loops or parallel links, elasticity $E(G)$ is a measure of the overall robustness. As shown in Figure 6.2, elasticity is the area under the curve of throughput versus the fraction of nodes removed. The throughput is normalized to compare networks of different magnitudes, and at each iteration, it is recalculated at the removal of each node. Initially, $T_G(0) =$

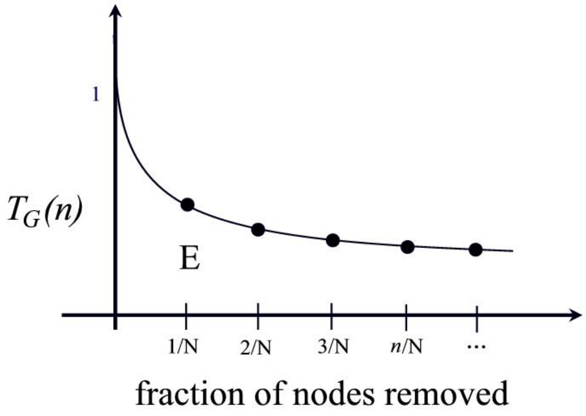

**Figure 6.2**: *The evaluation of elasticity*

1 which accounts for the normalized throughput. This value decreases as $\frac{n}{N}$ of nodes are removed, and therefore, elasticity (E) provides a measure of robustness at any point of node removal.

Therefore, when $\zeta$ nodes have been removed, elasticity can be computed as

$$E\left(\frac{\zeta}{N}\right) = \frac{1}{2N}\sum_{n=0}^{\zeta}\left(T_G\left(\frac{n}{N}\right) + T_G\left(\frac{n+1}{N}\right)\right) \tag{6.1}$$

where $T_G(\frac{n}{N})$ is the throughput at each interval when $n$ nodes are removed. $N$ is the total number of nodes in the network and $0 \leq (\zeta, n) \leq N$. At each iteration, the throughput is computed as

$$T_G\left(t\right) = \frac{\max_{\rho}\sum_{i,j} X_{i,j}\left(t\right)}{\alpha} \quad s.t. \quad LX \leq B\left(t\right) \tag{6.2}$$

where $t = \frac{n}{N}$ and $\rho$ is a constant used to vary the proportion of flows in network. $\alpha$ is the unnormalized initial throughput and $X_{i,j}\left(t\right)$ is the traffic flow between source node $i$ and destination node $j$. $L$ is the routing matrix, $X$ is a vector of all $X_{i,j}\left(t\right)$ flows, and $B\left(t\right)$ is a vector of all link bandwidth capacities.

### 6.2.1 Upper bound for Elasticity

**Analytical results**

In this section, we consider the mesh network as the topology which provides the highest elasticity under all attack strategies for any given network. We assume homogeneous flows where each flow has a value of $1$. Additionally, each link has a capacity of $1$ and $X_{ij}(t)$ can be $1$ or $0$ depending on whether or not a flow exists between nodes $i$ and $j$. With these assumptions, we proceed to determine the upper bound for elasticity.

**Theorem.** Given a mesh network with $N$ nodes, and assuming homogeneous flows and link capacities of $1$, then $\lim_{N\to\infty} E(N) = \frac{1}{3}$.

*Proof.* Elasticity can be formulated using both discrete and continuous approaches. At each iteration when a node is removed, the throughput is given by

$$T_G(t) = \frac{(N-n)(N-n-1)}{N(N-1)} \tag{6.3}$$

where $t = \frac{n}{N}$.

*Discrete Elasticity (trapezoidal integration).* For a given network of size $N$, Equation 6.4 computes elasticity when $\zeta$ nodes have been attacked.

$$E(\zeta) = \frac{1}{N}\left(\frac{1}{2} + \sum_{n=1}^{\zeta-1}\frac{(N-n)(N-n-1)}{N(N-1)} + \frac{(N-\zeta)(N-\zeta-1)}{2N(N-1)}\right) \tag{6.4}$$

where $1 \leq \zeta \leq N-1$. Equation 6.5 computes the total elasticity for a network with $N$ nodes when all $N$ nodes are progressively removed.

$$E(N) = \frac{1}{N}\left(\frac{1}{2} + \sum_{n=1}^{N-1}\frac{(N-n)(N-n-1)}{N(N-1)}\right) \tag{6.5}$$

*Continuous Elasticity.* Equation 6.6 gives the formulation of elasticity for the continuous case. Similar to the discrete case, Equation 6.7 computes elasticity for a given mesh network with size of $N$ where $\zeta$ nodes have been removed and Equation 6.8 computes the total elasticity for a mesh network with $N$ nodes. As the size of the network grows, Equation 6.9 then provides the upper bound on elasticity when all $N$ nodes are removed.

$$E\left(t\right) = \int_0^t T_G\left(\tau\right)d\tau, \quad 0 \leq t \leq 1 \tag{6.6}$$

$$E\left(\zeta\right) = \frac{N\left(N-1\right)\zeta + \frac{1}{2}\left(1-2\zeta\right)\zeta^2 + \frac{1}{3}\zeta^3}{N^2\left(N-1\right)} \tag{6.7}$$

$$E\left(N\right) = \frac{1}{3} - \frac{1}{6N} - \frac{1}{6N^2} \tag{6.8}$$

Therefore,

$$\lim_{N\to\infty} E(N) = \frac{1}{3} \tag{6.9}$$

Q.E.D.

**Numerical Results**

Figure 6.3 compares the convergence rate of the discrete and continuous cases when $\zeta$ nodes have been attacked from a mesh network where $N = 20$. As depicted, both approaches converge at the onset of node removal.

Figure 6.4 compares the convergence rate of elasticity for the discrete case in Equation 6.5 to the continuous case in Equation 6.8. As shown, both cases converge for a mesh network with 10 nodes.

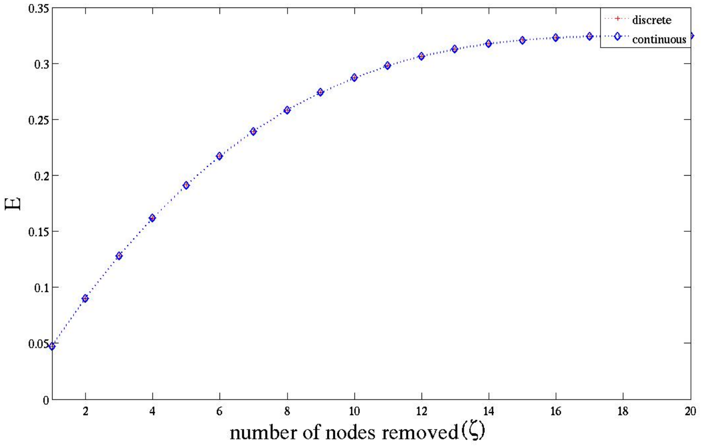

**Figure 6.3**: *Comparison of the convergence rates of elasticity, from Equations 6.4 and 6.7, where $\zeta$ nodes have been attacked*

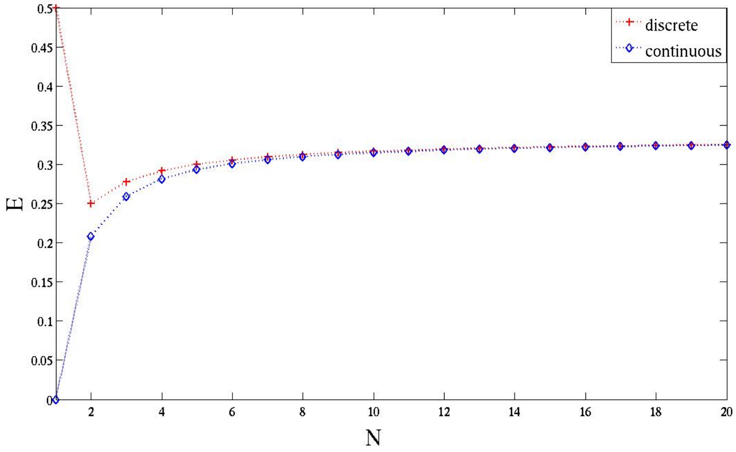

**Figure 6.4**: *Comparison of the convergence rates of elasticity, from Equations 6.5 and 6.8, for a network of size $N$*

Given the definition of elasticity, Subsection 6.3 explores models to implement a working prototype.

## 6.3 Model for Elasticity

Algorithm 1 shows the generic model for elasticity. As shown, the results from this model depends on the routing algorithm selected. For this reason, three routing approaches are explored: 1) Optimization (heterogeneous traffic matrix), 2) Dijkstra's standard shortest path (heterogeneous traffic matrix), and 3) Dijkstra's standard shortest path (homogeneous traffic matrix). In all approaches, link capacities were homogeneous.

**Algorithm 1** Elasticity Model

```
Adjacency Matrix=[]
while ($T_G(n) > tolerance$)
for $i = 1$ to $N$ do
    Execute routing algorithm
    Locate maximum congested link
    Calculate $T_G$(n)
    Remove node/link
end for
end while
Compute $E(G)$
```

**Routing Model 1: Optimization (heterogeneous traffic matrix)**

The Objective Function (Function 6.10) of the optimization problem maximizes the individual flow between all pairs of nodes. Equations 6.11-6.13 are the main constraints to the optimization problem. Equation 6.11 ensures that each node sends $\delta$ unit of traffic to every other node, while Equation 6.12 represents the balance of the incoming and outgoing traffic demands through any node in the network. Inequality 6.13 represents the capacity constraint on each link, and Equation 6.14 computes the utilization of each link.

$$Maximize\ \delta \tag{6.10}$$

Subject to

$$\sum_{j \in N} flow_{s,j,s} = \delta(N-1)\ \forall s \tag{6.11}$$

$$\sum_{i \in N} (flow_{i,j,s} - flow_{j,i,s}) = \delta\ \forall s, j, j \neq s \tag{6.12}$$

$$\sum_{s \in N} flow_{i,j,s} \leq capacity_{i,j}\ \forall i, j, i \neq j \tag{6.13}$$

$$utilization_{i,j} = \sum_{s \in N} flow_{i,j,s}\ \forall i, j \tag{6.14}$$

Algorithm 2 provides elasticity using the optimization approach discussed previously.

**Algorithm 2** Optimization

```
while $Connected := True$ do
    $capacity_{i,j} := 1$
    $demand_{i,j} := 0$
    while $\sum_{i,j} capacity_{i,j} \neq 0$ do
        Solve the optimization problem
        Update the demand between nodes that are connected with non-zero capacity links
        $demand_{i,j} := demand_{i,j} + \delta$
        $capacity_{i,j} := capacity_{i,j} - utilization_{i,j}$
    end while
    Remove one node (or a group of nodes)
end while
```

**Routing Model 2: Dijkstra's algorithm (heterogeneous traffic matrix)**

The second approach realizes Dijkstra's algorithm. As shown in Algorithm 3, flows traverse the shortest path from source to destination. This algorithm has a running time $O(n^2)$. However, when the heterogeneous traffic matrix is considered, the running time increases to $O(n^3)$.

**Algorithm 3** Dijkstra's shortest path algorithm [26]

```
begin
$S := 0; \bar{S} := N$
$d(i) := \infty$ for each node $i \in N$
$d(s) := 0$ and pred(s) $:= 0$
while $|S| < n$ do
begin
let $i \in \bar{S}$ be a node for which $d(i) = \min\{d(j) : j \in \bar{s}\}$
$S := S \cup \{i\}$;
$\bar{S} := \bar{S} - \{i\}$;
for each $(i, j) \in A(i)$ do
  if $d(j) > d(i) + c_{ij}$ then $d(j) := d(j) + c_{ij}$ and pred($j$) $:= i$
end for
end
```

**Routing Model 3: Dijkstra's algorithm (homogeneous traffic matrix)**

This approach also revolves around Algorithm 3 and has a running time $O(n^2)$. However, a homogeneous traffic matrix was implemented. Given these three models, Subsection 6.3.1 evaluates each and selects the most feasible.

### 6.3.1 Evaluation of Routing Models

Figure 6.5 shows the three networks for which elasticity was computed: Net 1, Net 2, and Net 3. For these three networks, we compare the results of elasticity provided by each routing model. More precisely, the following steps were taken for each network:

1. Elasticity was computed using model 1 (optimization with a heterogeneous traffic matrix)

2. Elasticity was computed using model 2 (Dijkstra's approach with a heterogeneous traffic matrix)

3. Elasticity was computed using model 3 (Dijkstra's approach with a homogeneous traffic matrix)

Additionally, each approach targeted, first, nodes with highest degree and, second, nodes with the highest betweenness.

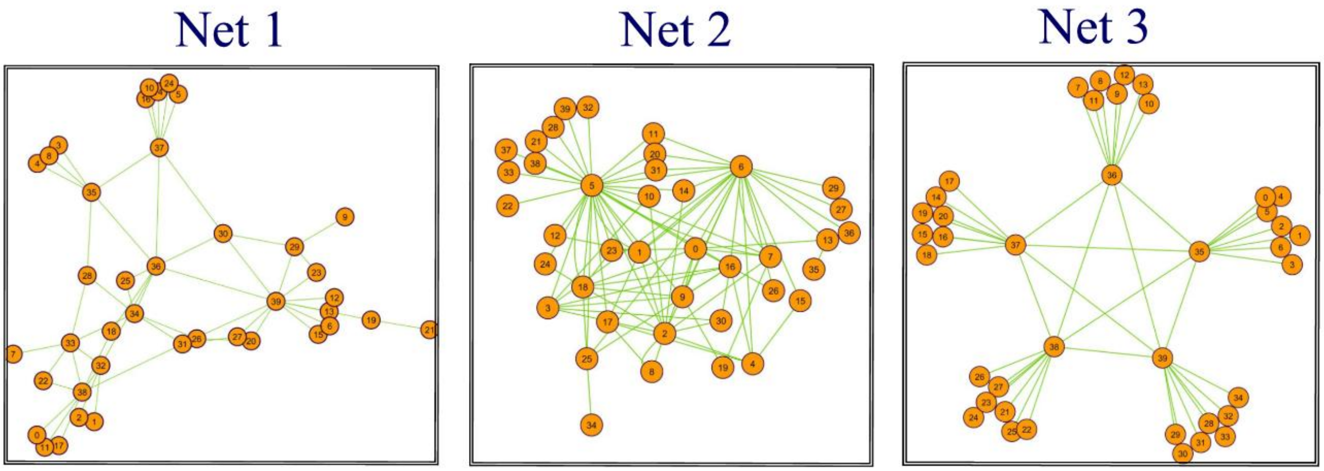

**Figure 6.5**: *Three networks for which elasticity was evaluated*

Figure 6.6 shows the throughput degradation as nodes with highest degree are attacked in Net 1. The throughput values were normalized using the total capacity of the network to compare the three different removal strategies. As depicted, the optimization approach produces the highest elasticity, followed by Dijkstra's heterogeneous approach and finally, Dijkstra's homogeneous approach. This trend was observed for each network under both attack strategies. However, under certain circumstances where the network has low connectivity, the elasticity results were identical for both Dijkstra's "heterogeneous" and optimization models.

For each of the three routing model, each network was given a rank of 1, 2 or 3, based on its value for elasticity: 1 as the highest and 3 as the lowest. Table 6.1 displays the rankings for each network under highest node degree attack. As shown, elasticity was highest for Net 1, followed by Net 2, and finally, Net 3, for each approach. Though the values were different for the highest betweenness attack strategy (not shown), the rankings were similar to that of Table 6.1. The implications of these results are presented at the conclusion of this section.

We then observed that the criteria for node addition to the shortest path could potentially affect the results of elasticity. More specifically, in Algorithm 3, nodes are added to the shortest path if the following optimality condition is satisfied:

$$d\left(j\right) > d\left(i\right) + c_{ij} \tag{6.15}$$

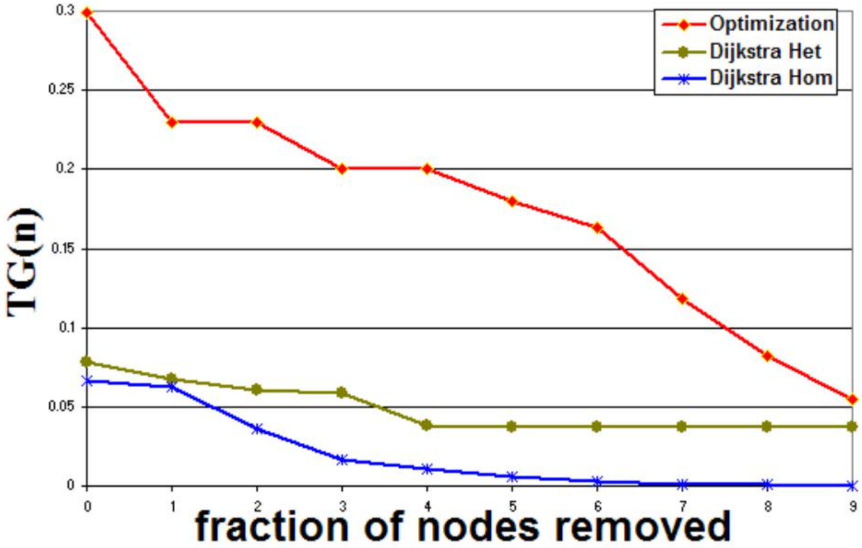

**Figure 6.6**: *The throughput degradation as nodes with highest degree are attacked for Net 1*

**Table 6.1**: *Elasticity comparison for all networks under* **highest node degree** *attack*

| Approach | Rank 1 | Rank 2 | Rank 3 |
|---|---|---|---|
| 1 | Net 1 | Net 2 | Net 3 |
| 2 | Net 1 | Net 2 | Net 3 |
| 3 | Net 1 | Net 2 | Net 3 |

where $d(j)$ is the distance label at node $j$ and $c_{ij}$ is the cost of moving from node $i$ to $j$.

However, if there are several nodes $j$, such that each node equally satisfies this condition, the next node added to the shortest path is selected sequentially. To investigate the impact of this constraint on the results of elasticity, we modified Dijkstra's algorithm to relax the sequential constraint by randomly selecting the next node $j$ that will be added to the shortest path. The resulting changes are reflected in Algorithm 4. For each network, the result of Algorithm 4 were compared to that of Algorithm 3.

**Algorithm 4** Dijkstra's "Modified" shortest path algorithm

```
begin
$S := 0; \bar{S} := N$
$d(i) := \infty$ for each node $i \in N$
$d(s) := 0$ and pred(s) $:= 0$
while $|S| < n$ do
begin
let $i \in \bar{S}$ be a node for which $d\,(i) = \min\,\{d\,(j) : j \in \bar{s}\}$
$S := S \cup \{i\}\,;$
$\bar{S} := \bar{S} - \{i\}\,;$
for each $(i, j) \in A\,(i)$ do
    $X_i = j, \forall j \in N$ which satisfy the optimality condition
    $j_{selected} = rand\,(X_i)$
    then $d\,(j) := d\,(j) + c_{ij}$ and pred($j$) $:= i$
end for
end
```

For Algorithm 4, we conducted 100 sample runs and averaged elasticity for each network. For each run, we observed that the number of flows in the bottleneck link varied. However, the difference between Algorithm 4 and Algorithm 3 were negligible. Hence, the rankings shown in Table 6.1 remain the same.

### 6.3.2 Conclusion

In this section, we introduced a new metric elasticity as a measure of robustness for a given network.

### Research Method

More importantly, we motivated the need for elasticity and detailed our contributions to the field. To this end, we 1) defined elastic, 2) derived its upper bound and 3) presented the model to compute elasticity. Finally, we observed a dependance of elasticity on the selected routing algoriithm. Thus, we investigated the results of elasticity using three possible routing models:

1. An Optimization algorithm based on a $heterogeneous$ traffic matrix
2. Dijkstra's algorithm based on a $heterogeneous$ traffic matrix and
3. Dijkstra's algorithm based on a $homogeneous$ traffic matrix

Each approach was applied to three networks, namely Net 1, Net 2, and Net 3, using the following node-attack strategies:

- Highest degree and
- Highest betweenness.

Dijkstra's algorithm was then modified to include randomness in selection of nodes added to the shortest path.

### Results

The following results were obtained:

1. The mesh network provided the upper bound of $\frac{1}{3}$ on elasticity
2. The optimization model produced the highest elasticity followed Dijkstra's "heterogeneous" model and finally, Dijkstra's "homogeneous" model.
3. Though the addition of randomness to Dijkstra's algorithm varied the value of the bottleneck link, the average value of elasticity was close to that of the non-random version. Thus, the rankings were identical to that of Table 6.1

4. Quantitatively, all approaches produced different values for elasticity. However, qualitatively, the results were similar in that, all approaches were consistent in ranking the different networks from highest to least elasticity. Consequently, the least costly method was selected to implement elasticity: Dijkstra's algorithm.

# Chapter 7

# EXPERIMENTAL RESULTS

In this Section, we evaluate elasticity for a set of selected topologies. A more detailed analysis of the characteristics of these topologies is provided in Tables A.1 and A.2 found in the Appendix. First, we compute the elasticity of all networks under each attack strategy and second, we implement a tradeoff function that combines the elasticities obtained for each attack strategy and penalizes networks for having excess links.

## 7.1 Elasticity of Networks Under Three Attack Strategies

In the subsequent sections, Elasticity R, Elasticity D, and Elasticity B refer to elasticity under the following three attack strategies:

1. removal of random nodes (Elasticity R)
2. removal of highest degree nodes (Elasticity D)
3. removal of highest betweenness nodes (Elasticity B)

Table 7.1 ranks all networks in descending order of magnitude based on the number of links and the scores for elasticity under the three strategies. As shown, the mesh network is the most robust under all strategies. This is expected, as it sets the upper bound on elasticity. Under random attacks, the elasticity for the Gi-dense and MySpace networks are in proximity to that of the mesh network. As cost is a critical factor in network design, it is financially sensible to implement the latter two topologies rather than the mesh because Table 7.1 shows that the MySpace and Gi-dense networks can provide about $94\%$ of the elasticity that the mesh provides while only using about $1\%$ of the links.

| Nets. | links | Nets. | Elas. R | Nets. | Elas. D | Nets. | Elas. B |
|---|---|---|---|---|---|---|---|
| mesh | 499500 | mesh | 0.3333 | mesh | 0.3333 | mesh | 0.3333 |
| MySpace | 10976 | MySpace | 0.3119 | n-r 2 | 0.2426 | Gi-dense | 0.2390 |
| Gi-dense | 4505 | Gi-dense | 0.3111 | Gi-dense | 0.2082 | W-S 2 | 0.1770 |
| n-r 2 | 3781 | PA 2 | 0.2743 | MySpace | 0.1721 | MySpace | 0.1719 |
| W-S 2 | 3000 | W-S 2 | 0.2703 | W-S 2 | 0.1640 | W-S 1 | 0.1260 |
| PA 2 | 2964 | PA 1 | 0.2677 | n-r 1 | 0.1342 | Gi-sparse | 0.1010 |
| Gi-sparse | 2009 | Gi-sparse | 0.2520 | W-S 1 | 0.1170 | PA 2 | 0.0719 |
| W-S 1 | 2000 | W-S 1 | 0.2490 | Gi-sparse | 0.1143 | PA 1 | 0.0558 |
| PA 1 | 1981 | n-r 2 | 0.2316 | PA 2 | 0.0644 | YouTube | 0.0332 |
| n-r 1 | 1921 | Flickr | 0.2211 | PA 1 | 0.0535 | Flickr | 0.0315 |
| YouTube | 1576 | YouTube | 0.2132 | YouTube | 0.0371 | n-r 2 | 0.0246 |
| Flickr | 1515 | meshcore | 0.1997 | Flickr | 0.0285 | n-r 1 | 0.0178 |
| meshcore | 1275 | HOT 2 | 0.1623 | HOT 1 | 0.0129 | meshcore | 0.0083 |
| HOT 2 | 1049 | PA-sparse | 0.1537 | HOT 2 | 0.0095 | HOT 1 | 0.0059 |
| PA-sparse | 1049 | HOT 1 | 0.1405 | Abilene | 0.0093 | HOT 2 | 0.0048 |
| ringcore | 1000 | ringcore | 0.1290 | meshcore | 0.0083 | PA-sparse | 0.0039 |
| HOT 1 | 988 | Abilene | 0.1280 | PA-sparse | 0.0045 | Abilene | 0.0031 |
| Abilene | 896 | n-r 1 | 0.1016 | ringcore | 0.0040 | ringcore | 0.0026 |

**Table 7.1**: *Networks sorted in descending order for number of links, Elasticity R (Elas. R), Elasticity D (Elas. D), and Elasticity B (Elas. B)*

The subsequent Subsections show correlations for elasticity under the specified attack strategy.

### 7.1.1 Correlation Between Elasticity and Number of Links

From Table 7.1, it is notable that for all removal strategies the MySpace, Gi, PA, and Watts-Strogatz networks all vie for the highest elasticity. This phenomenon can be explained by considering the large number of links of these networks. Figures 7.1, 7.2, and 7.3 confirm this propensity and depict elasticity under random, targeted, and highest betweenness attacks respectively. The networks are color-coded to represent two classes of networks: 1) The heterogeneous class with varying intensities of blue represents networks with a power-law distribution, and 2) the semi-regular class, further broken down into deterministic and random networks, has varying shades of red and is indicative of networks with a poisson degree distribution. This tendency for elasticity to increase as the number of links increase is not always the case. For instance, in Figure 7.1, though the n-r 2 network has about $3900$ edges with an elasticity of about $0.235$, the PA 1 network has $2000$ edges and an elasticity of $0.26$. Thus, a large number of edges is not necessary even if it is a sufficient condition for high elasticity.

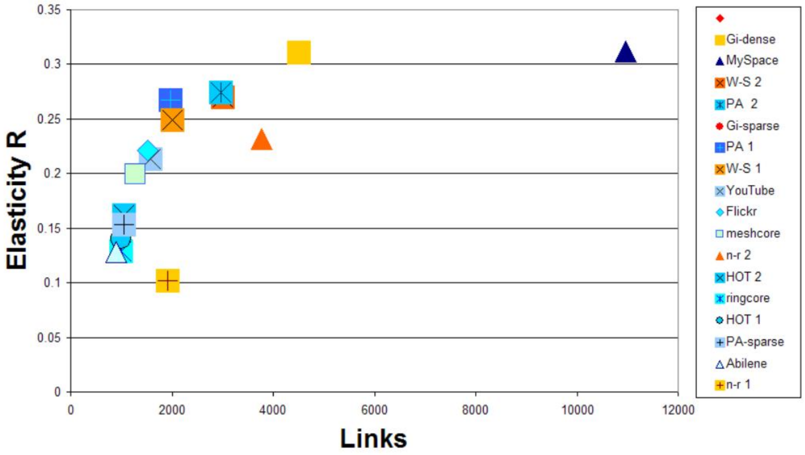

**Figure 7.1**: *Elasticity R vs number of links for each network in Table A.1*

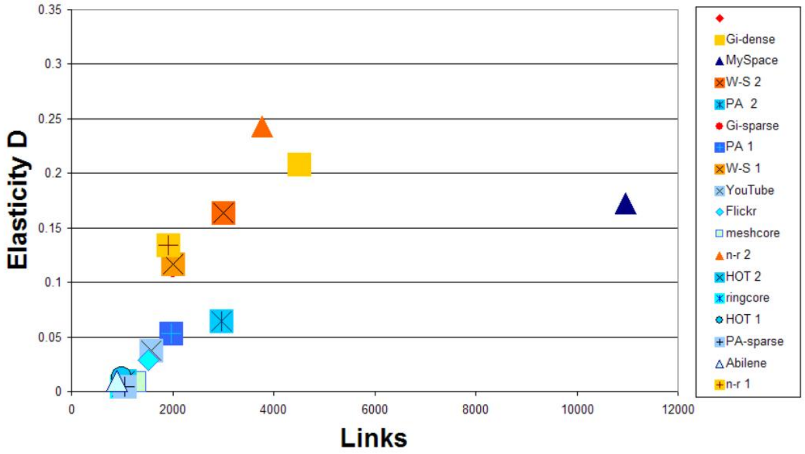

**Figure 7.2**: *Elasticity D vs number of links for each network in Table A.1*

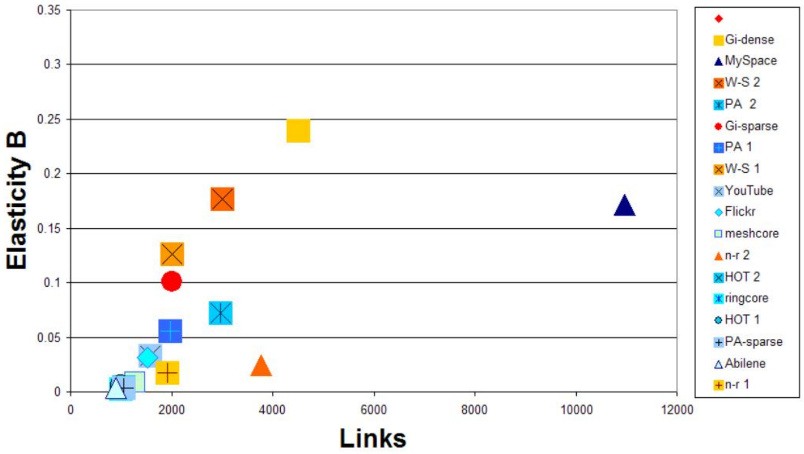

**Figure 7.3**: *Elasticity B vs number of edges for each network in Table A.1*

Under financial constraints, meticulous consideration must be given to the architecture of the network. Table 7.1 shows that under random attack, elasticity can be as low as $30.5\%$ of the upper bound. This sharply declines to $1.2\%$ for highest degree attacks and $0.8\%$ for highest betweenness attacks. For this reason, the design of a robust topology is of utmost importance to obtain high elasticity. For example, the HOT 1 and PA-sparse networks have the same number of links and also the same number of nodes. Though Figures 7.4 and 7.5 show almost identical degree distributions for these networks, their response to attack differs [20]. Under random attacks, the PA-sparse provides $9.76\%$ more than the HOT 1 network. In the PA-sparse network, low degree nodes outnumber high degree nodes (hubs) and hence, the probability of attacking hubs is lower than that of attacking other nodes. This is also the case for the HOT 1 network. However, the ratio of low degree nodes to hubs is higher in the PA-sparse network than in the HOT 1 network. As a result, Elasticity R for the PA-sparse network is higher than the HOT 1 network. However, the HOT 1 topology provides three times as much Elasticity D as the PA-sparse network. The PA-sparse network is more susceptible to this attack because the hubs in this network facilitate interconnection and are vital to the elasticity of the network. However, the hubs in the HOT 1 network are located on the periphery and are less critical to interconnections [18].

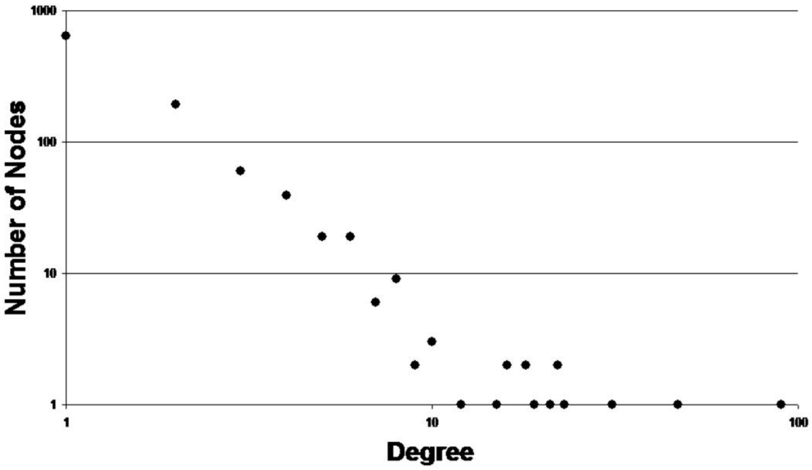

**Figure 7.4**: *Node degree distribution for the PA-sparse network*

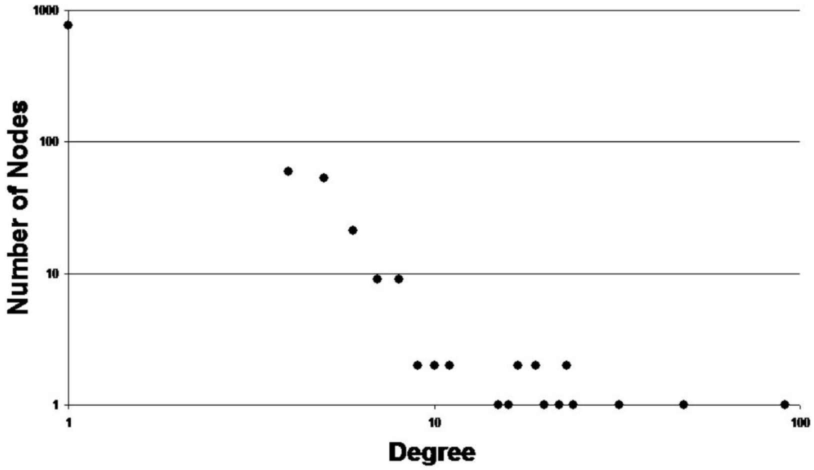

**Figure 7.5**: *Node degree distribution for the HOT 1 network*

For highest betweenness attack, the elasticity of both networks decreases even more. It is notable that from highest degree to highest betweenness attack, the elasticity provided by HOT 1 exhibits a $54.3\%$ decrease whereas that provided by PA 1 exhibits a much smaller decrease of $13.3\%$. This can be interpreted from Figures 7.6 and 7.7 that show the betweenness distribution for the PA-sparse and HOT 1 networks. From Figure 7.6, nodes with the highest degrees have the highest betweenness. Thus, damage incurred under highest betweenness attacks is almost similar to that under highest degree attack. However, for the HOT 1 network there is a large decrease in elasticity from highest degree attack to highest betweenness attacks. Figure 7.7 shows that nodes with the highest betweenness tend to have lower degrees. This nodes facilitate interconnection within the network. Thus, attacks on these nodes are more detrimental than high degree attacks.

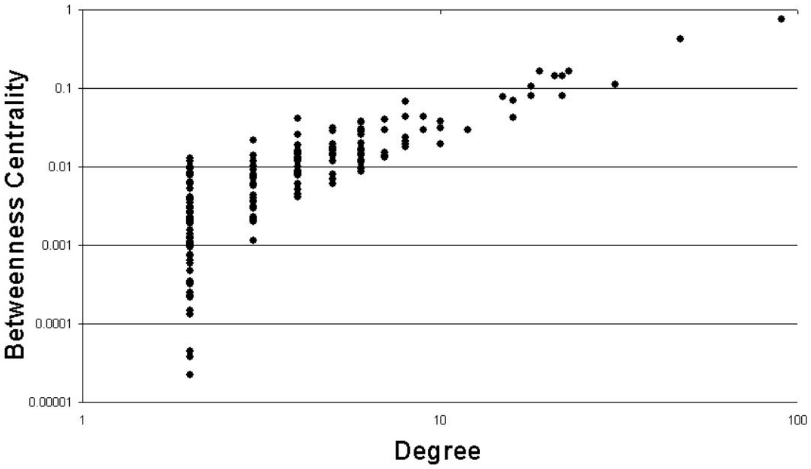

**Figure 7.6**: *Betweenness distribution for the PA-sparse network*

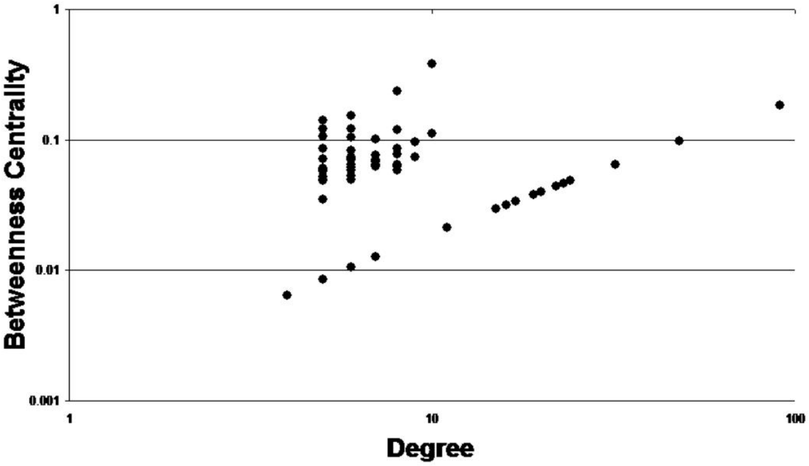

**Figure 7.7**: *Betweenness distribution for the HOT 1 network*

### 7.1.2 Correlation Between Elasticity and Heterogeneity

Figures 7.8 through 7.15 show other notable correlations for robust topologies. More precisely, Figures 7.8, 7.9, and 7.10 illustrate the effect of heterogeneity on the elasticity of a network. The interpretation of these Figures is that non-heterogeneous, or rather homogeneous networks, have a proclivity for higher levels of elasticity. These include the variations of Watts-Strogatz's small world models, the random models, and the near-regular models, which Table A.2 gives the heterogeneity scores from $.089$ to $.491$. The implications of these results are far reaching where network structure is concerned. Although these topologies are sufficiently costly, in addition to the fact that they may fail to capture the properties of real world networks, their topological structure offers remarkable resilience to attacks.

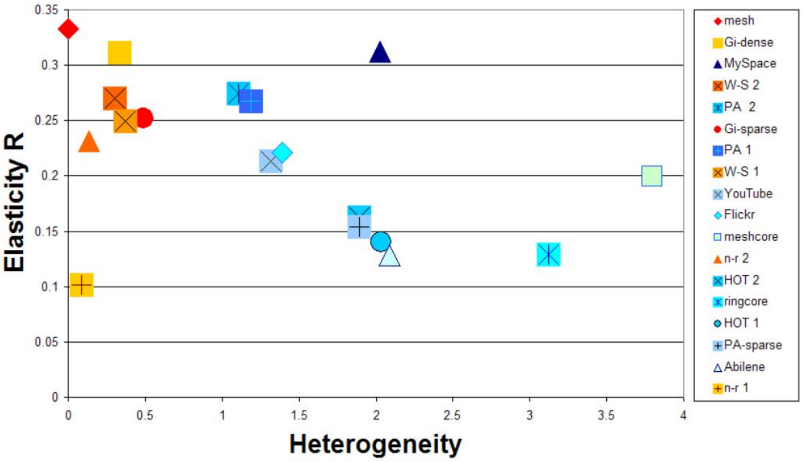

**Figure 7.8**: *Elasticity R vs heterogeneity for each network in Table A.1*

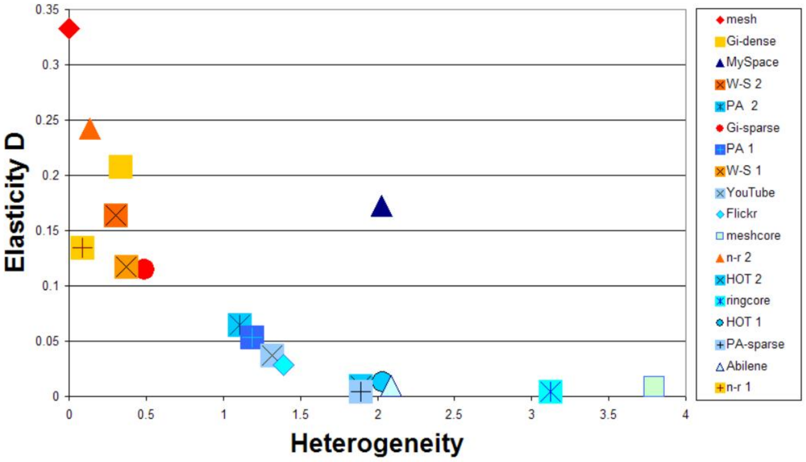

**Figure 7.9**: *Elasticity D vs heterogeneity for each network in Table A.1*

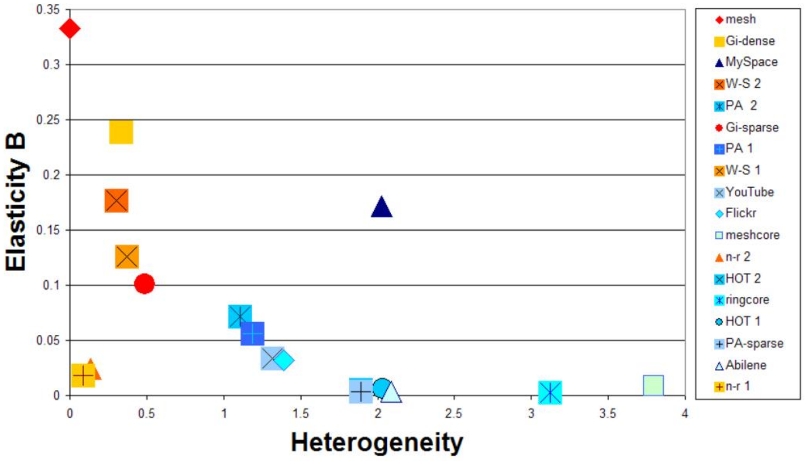

**Figure 7.10**: *Elasticity B vs heterogeneity for each network in Table A.1*

Figures 7.11 and 7.12 compare the degree distribution of the W-S 2 and Abilene networks respectively. These Figures provide an in-depth understanding of the effects of attacks on the respective networks. The W-S 2 network is a representative of the random, semi-regular class of topologies, and as shown in Figure 7.11, the majority of nodes tend to have a degree close to the average degree. Therefore, the damage incurred under highest node degree and highest betweenness attacks is comparable. From Table 7.1, W-S 2 has elasticity scores of $0.164$ and $0.177$ for highest degree and highest betweenness attacks. This is the case for the other networks within this class.

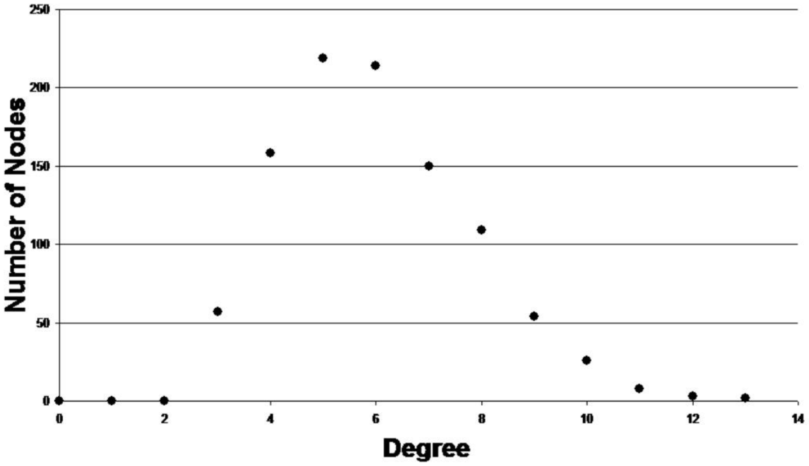

**Figure 7.11**: *Node degree distribution of Watts-Strogatz 2 network*

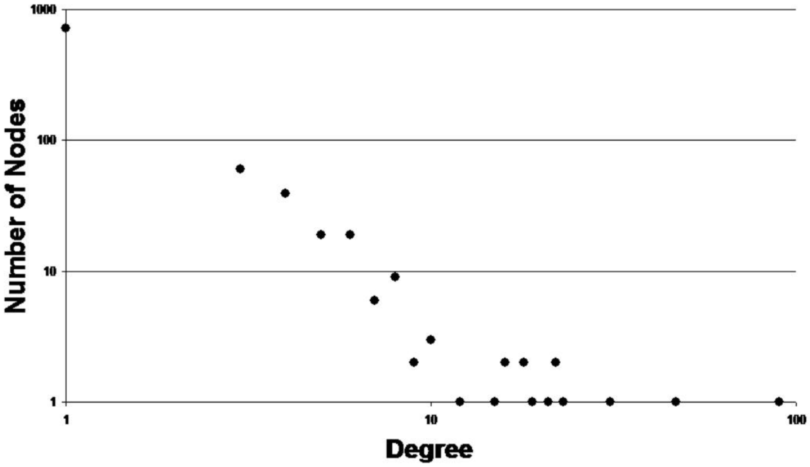

**Figure 7.12**: *Node degree distribution of Abilene network*

A representative of the deterministic, semi-regular networks, n-r 1 maintains its elasticity under random and high degree attacks: a notable property from networks within the semi-regular class. This result can be understood by the almost constant node degree. However, as nodes are removed under highest betweenness attacks, core nodes appear and are destroyed. For n-r 1, elasticity decreases considerably from highest degree attacks to highest betweenness attacks by $35\%$. Thus, although homogeneous topologies are sufficiently costly, in addition to the fact that they may fail to capture the properties of some real world networks, their topological structures offer remarkable resilience to attacks.

Figure 7.12 shows the degree distribution for the Abilene network: a representative of the heterogeneous class of networks. Based on the “type” of heterogeneous network under investigation, the impact of highest degree attacks can vary. On the one hand, networks like Abilene avoid cataclysmic damage under high degree attack because the hubs are located on the periphery of the network and thus, highest degree attack has miminal effect on the overall operation of this network. However, heterogeneous networks like PA-sparse are severely damaged because the hubs are critical and hold the network together [18]. Thus, the location of hubs in the network affects its elasticity.

### 7.1.3 Correlation Between Elasticity and Characteristic Path Length

The characteristic path length tells the expected distance, in number of hops, from a given source node $s$ to a destination node $t$. Figures 7.13, 7.14, and 7.15 show that the characteristic path length tends to be negatively correlated with elasticity. This is not a necessary condition as Figure 7.13 provides instances where a network with high characteristic path length can have a higher elasticity than a network with a smaller characteristic path length. For example, from Table A.2 the n-r 2 network has a large average distance and from Table 7.1, it has an elasticity of $.2316$ under random attack. As a counterexample, the Abilene network has small characteristic path length but an elasticity of $.128$ under random attack. As discussed previously, elasticity has the tendency to increase as the number of links in the network increases. If the number of nodes in a given network is kept constant as the number of links increase, path diversity will eventually increase. As a result, network congestion decreases which ultimately increases elasticity.

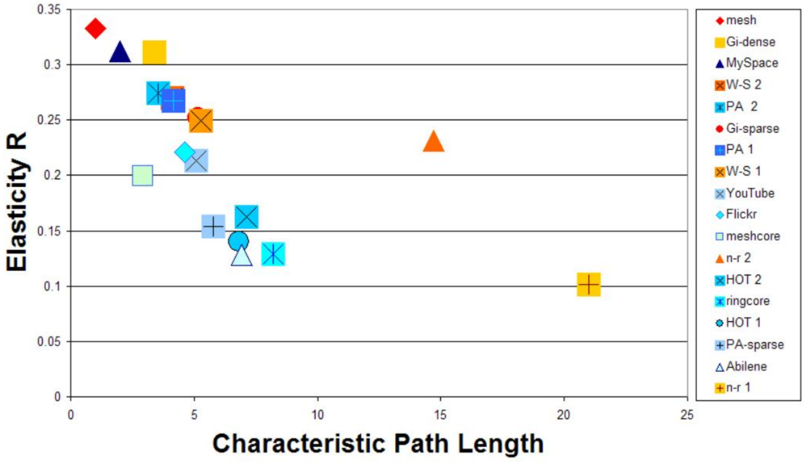

**Figure 7.13**: *Elasticity R vs characteristic path length for each network in Table A.1*

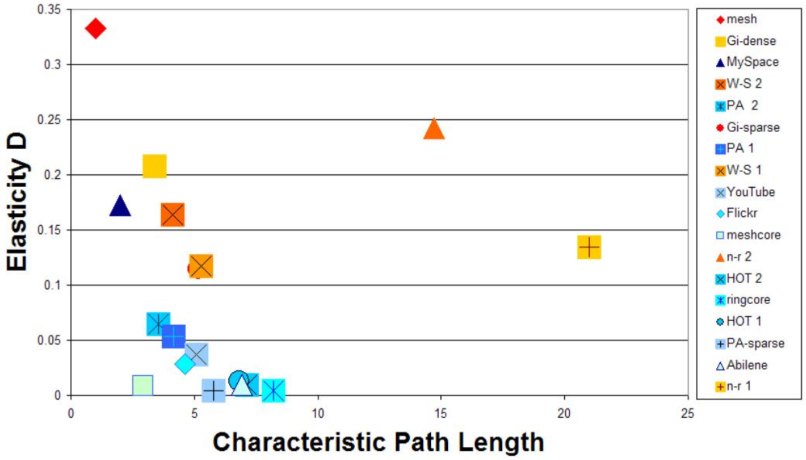

**Figure 7.14**: *Elasticity D vs characteristic path length for each network in Table A.1*

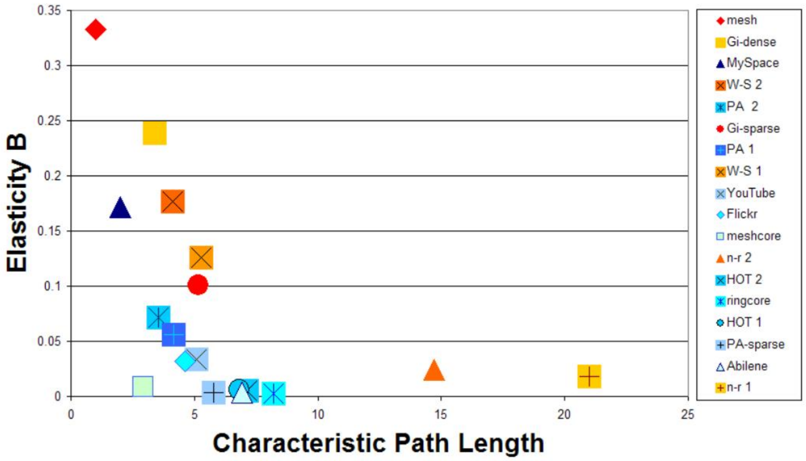

**Figure 7.15**: *Elasticity B vs characteristic path length for each network in Table A.1*

## 7.2 Elasticity of Networks with Tradeoff Function Applied

To compensate for the tradeoff between elasticity and number of links, we introduce a tradeoff function $Re\,(G)$ that provides robustness with respect to elasticity. For a given network $G$, our robustness measure can be computed as

$$Re\,(G) = \alpha ElasticityR + \beta ElasticityD + \delta ElasticityB - \gamma density' \tag{7.1}$$

where $0 \leq (\alpha, \beta, \delta, \gamma) \leq 1$, $0 \leq density' \leq 1$ and $density' = 1 - e^{-\frac{1}{2}\frac{(M-(N-1))}{N}}$. The $\frac{1}{2}$ factor determines the rate at which $density'$ changes with respect to the number of excess links. $\alpha$, $\beta$,$\delta$, and $\gamma$ are tolerance parameters and, as such, represent the tolerance of a network towards random, targeted, and highest betweenness attacks. $M$ is the total number of links and $M - (N-1)$ represents the number of excess links in a network: these are links which exceed the threshold necessary to obtain $1$ connected component with N nodes.

The implications of this approach are considerable. This function facilitates independence for constructing networks based on a projected need. Thus, a network engineer who envisions persistent, random attacks would consider a high value of $\alpha$. Similarly, $\beta$ or $\delta$ would dominate where targeted attacks or highest betweenness attacks respectively are predominant. Moreover, $\gamma$

| **Networks** | **Number of links** | $R_e$ |
|---|---|---|
| HOT 2 | 1049 | 0.1519 |
| PA-sparse | 1049 | 0.1374 |
| ringcore | 1000 | 0.1351 |
| Abilene | 896 | 0.1342 |
| HOT 1 | 988 | 0.1330 |
| Watts-Strogatz 1 | 2000 | 0.0982 |
| meshcore | 1275 | 0.0874 |
| YouTube | 1576 | 0.0828 |
| ER sparse | 2009 | 0.0708 |
| Flickr | 1515 | 0.0340 |
| PA 1 | 1981 | -0.0110 |
| Watts-Strogatz 2 | 3000 | -0.0210 |
| ER dense | 4505 | -0.0684 |
| near-regular 1 | 1921 | -0.1206 |
| PA 2 | 2964 | -0.2150 |
| near-regular 2 | 3781 | -0.2561 |
| MySpace | 10976 | -0.3388 |

**Table 7.2**: *Ranking for topologies according to the cost function* $R_e$

could be varied based on financial constraints.

For each network, we obtained $R_e$ for tolerance values of $\alpha$, $\beta$, $\delta$, and $\gamma = 1$; Table 7.2 depicts the rankings of each topology with their respective number of links and $R_e$ scores. The common tolerance values facilitate an unbiased analysis of robustness by providing equal likelihood of occurrence to each attack strategy. In addition, these rankings represent the case where networks are completely penalized for having excess links and as a result, the structure of the network plays a more significant role to determine the robustness of the network. From Table 7.2, there exists three distinct groups with respect to the values of $R_e$. Networks from HOT 2 to HOT 1 form the first group and have the highest $R_e$. The next group contains networks from Watts-Strogatz 1 to Flickr with mid-range values of $R_e$, and the final group includes the networks from PA 1 to MySpace with the lowest values of $R_e$.

Figure 7.16 shows the "HOT" 2 topology: The highest ranking topology under homogeneous tolerance parameters. From Table A.1, this network has only $50$ excess links and avoids a high penalty for the existence of excess links. Though "HOT 2" exhibits power-law properties, the hubs are located on the periphery. From Table 7.1, "HOT 2" realizes a higher elasticity than the meshcore network. However, under highest betweenness attacks, the meshcore exhibits a higher elasticity than the HOT topology. Figure 7.17 shows the betweenness distribution for the "HOT 2" topology. On the one hand, nodes with the highest betweenness have node degrees ranging from about 6 to 10. The majority of these nodes are located in the core of the network and facilitate communication between other nodes. On the other hand, nodes with lowest and highest degrees are located on the network periphery. In all, though "HOT 2" has an admirable structure for targeted attacks and is virtually cheap, it is more liable to fail under highest betweenness attacks. However, considering the values for the tolerance parameters discussed previously, the "HOT 2" network is the most suitable.

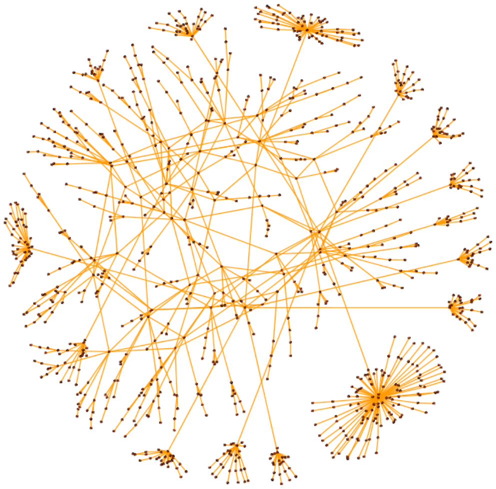

**Figure 7.16**: *'HOT 2", the highest ranking network as shown in Table 7.2*

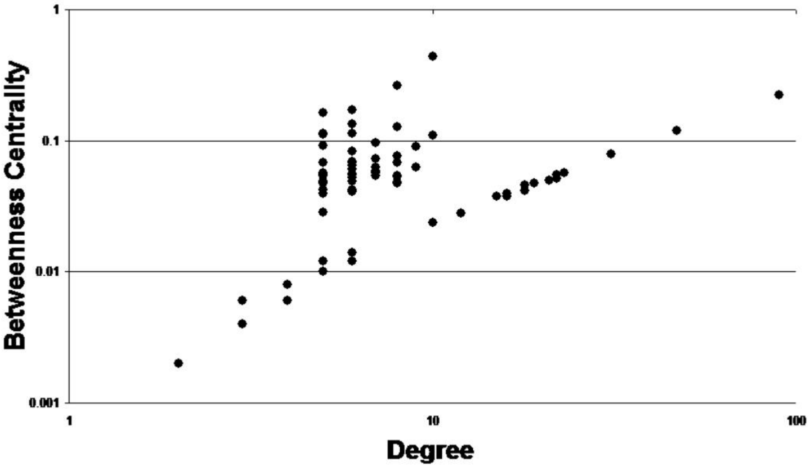

**Figure 7.17**: *Betweenness distribution for the "HOT 2" network*

## 7.3 Tradeoff Between Characteristic Path Length and Heterogeneity

The ideal network to provide elasticity tends to exhibit a low score for heterogeneity and a short characteristic path length. In all networks, the mesh has the shortest characteristic path and the lowest score for heterogeneity and hence, it boasts the highest elasticity. However, this high elasticity comes at a very high cost which network designers are unwilling to consider. For this reason, it is imperative to consider a tradeoff between a short characteristic path length and a low score for heterogeneity. Figure 7.18 can be interpreted as a decrease in the characteristic path length such that the network becomes more heterogeneous decreases elasticity. However, a decrease in both components results in an increase in elasticity.

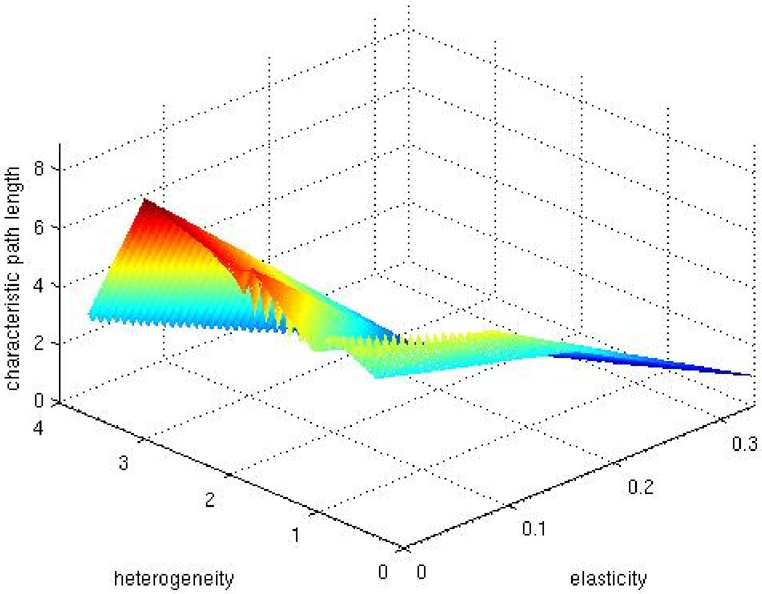

**Figure 7.18**: *Elasticity increases as characteristic path length and network heterogeneity decrease*

# Chapter 8

# CONCLUSIONS AND OUTLOOK

This thesis endeavors to extract the characteristics of robust complex networks. To this end, we define robustness as the ability for a network to maintain its total throughput under removal of nodes and consequently, links. This definition motivates the introduction of a new metric, namely elasticity, which provides one distinct value as the measure of robustness of a network. In addition, we theoretically derived an upper bound of $\frac{1}{3}$ on elasticity, and illustrated the utility of elasticity on 18 networks from six different network models under random, highest degree, and highest betweenness attacks.

For our experimental results, we conducted two experiments to evaluate elasticity. Our first experiment realized elasticity under the aforementioned attack strategies. Under all attack strategies, the mesh network outperformed all other networks because it sets the upperbound for elasticity. However, the MySpace and ER dense networks were able to provide $94\%$ of the upper bound with as few as $2\%$ and $1\%$ respectively, of the number of links in the mesh network. With cost as a deterrent to implementing a mesh network, designers would sensibly gravitate towards implementing the latter two networks.

Networks such as MySpace, Gi, PA, and W-S exhibited high elasticity due to the exceedingly large number of links compared to the other sparse networks. One reason for this behavior emanates from the increased path diversity that results from having a large number of links. This reduces congestion and, as a result, increases elasticity. However, we showed that increasing the number of links was not a necessary condition for high elasticity. To this end, we demonstrated that though the PA 1 network has half the number of edges as the n-r 2 network, it provides up to $7.5\%$ more elasticity than the n-r 2 network.

Though increased links tend to increase elasticity, under realistic financial constraints, careful design of the network architecture has far-reaching repurcussions. Under random attack, elasticity can be as low as $30.5\%$ of the upper bound. This sharply declines to $1.2\%$ for highest degree attacks and $0.8\%$ for highest betweenness attacks. Thus, designing the network that maximizes elasticity is of utmost importance. In line with [18, 20], we used elasticity to show that though two networks have similar number of nodes, links, and degree distribution, the location of the hubs influences the robustness of a network. In particular, we used the PA sparse network to show that under random attacks the elasticity is higher than under highest degree attack. This occurs because the proportion of hubs to less critical nodes is quite low. Thus, the probability of randomly attacking critical nodes is low and hence, elasticity is high. However, under high degree attack, the PA is destroyed at an alarming rate because nodes that facilitate interconnection are attacked. Furthermore, under highest betweenness attacks, the PA network has an elasticity score that is similar to that under highest degree attack. This can be explained using the betweenness distribution which shows that the highest degree nodes are also the nodes with the highest betweenness. Thus, either attack results in a comparative value for elasticity. The HOT 1 network, on the other hand, remains robust under highest degree attacks. This network has the high degree nodes on the periphery, and thus, highest degree attack affects this network minimally. However, under highest betweenness attacks, HOT 1 exhibits a $54\%$ decrease with respect to highest degree in elasticity compared to the $13.3\%$ experienced by the PA 1 network.

We showed that networks with lower heterogeneity tend to have higher elasticity. To illustrate this concept, we classified topologies into two classes. The semi-regular class, which includes the random and deterministic subclasses, exhibits a Poisson node degree distribution. Heterogeneous networks fall into the second class. The node degree distribution for the random, semi-regular class shows that most nodes tend to have degrees in proximity to the average node degree. As a result, random and highest degree attacks gives similar values for elasticity. The deterministic, semi-regular class exhibits this same property. However, under highest betweenness attacks, the core nodes are destroyed, and the network falls apart. The amount of damage incurred by heterogeneous networks depends on the structure of the architecture. As discussed previously, the locations of the hubs influence how robust the network is toward high degree attacks.

We also showed that networks with lower characteristic path length tend to have higher elasticity. As the number of links in the network increase, the characteristic path length decreases which inevitably increases path diversity. With more paths available, congestion is decreased and thus elasticity increases.

We then implemented a tradeoff function $Re$ that provides a measure of robustness with respect to elasticity. More specifically, this function summed up elasticity under all removal strategies and penalized networks for having excess links. In addition, the inclusion of tolerance parameters $\alpha$, $\beta$, $\delta$, and $\gamma$ in the tradeoff function provides network engineers the liberty to carefully tailor their networks to best meet their needs. We implemented this tradeoff function for $\alpha = \beta = \delta = \gamma = 1$. Finally, we showed that networks with high elasticity had short characteristic path length and low heterogeneity.

Elasticity is defined and computed under simple assumptions. As an example, it is dependent on the routing algorithm used, which can perhaps alter current network rankings. However, elasticity provides benefits which are far-reaching. More precisely, it identifies key characteristics of robust complex networks: A short characteristic path length, low heterogeneity, and strategically located links to facilitate a "homogeneous" core such that if hubs should be added, they should be placed on the periphery of the network to provide added resilience against targeted attacks.

For our future work, we intend to incorporate expander graphs in our evaluation and formulate a working definition of the core and periphery to include details about the size and characteristics. Armed with this knowledge, we seek to combine particular graphs to extract the essential components to increase elasticity. Finally, we will develop heuristics to build graphs such that elasticity is maximized.

# Bibliography

[1] ITPRO. [Online]: http://www.itpro.co.uk/100966/taiwan-quake-exposes-weakness-of-undersea-data-lines, 2006.

[2] CNN. [Online]: http://archives.cnn.com/2001/TECH/2001.

[3] BBC. [Online]: http://news.bbc.co.uk/2/hi/technology/3682537.stm, 2004.

[4] Computer Security. [Online]: http://ftp.fas.org/irp/gao/aimd-96-108.htm, 1996.

[5] US Canada Power System Outage Task Force. August 14th blackout: Causes and recommendations, (2003).

[6] Swiss Federal Office of Energy. Report on the Blackout in Italy on September 28 (2003).

[7] H. Frank and I. Frisch, "Analysis and design of survivable networks," in *IEEE Transactions on Communications Technology COM-18*, p. 567, 1970.

[8] A. Jamakovic and S. Uhlig, "Influence of the network structure on robustness," in *Networks, 2007. ICON 2007. 15th IEEE International Conference*, pp. 278–283, 2007.

[9] D. Bauer, F. Boesch, C. Suffel, and R. Tindell, "The theory and application of graphs," pp. 89–98, 1981.

[10] F. Harary, "On conditional edge-connectivity of graphs," in *Networks*, vol. 13, p. 346, 1981.

[11] A. H. Esfahanian and S. Hakimi, "On computing a conditional edge-connectivity of a graph," in *Journal of Information processing Letters*, vol. 27, p. 195, 1988.

[12] M. Krishnamoorth and B. Krishnamirthy, "Fault diameter of interconnection networks," in *Computers and Mathematics with Applications*, vol. 13, p. 577, 1987.

[13] V. Chvatal, “Tough graphs and hamiltonian circuits,” in *Discrete Math*, vol. 5, p. 215, 1973.

[14] H. Jung, “On a class of posets and the corresponding comparability graphs,” in *Journal of Combinatorial Theory B*, vol. 24, p. 125, 1978.

[15] M. Cozzen, D. Moazzami, and S. Stueckle, “Seventh international conference on the theory and applications of graphs,” p. 11111122, 1995.

[16] M. Alon, “Eigenvalues and expanders,” in *Combinatorica*, vol. 6, p. 83, 1986.

[17] B. Mohar, “Isoperimetric numbers of graphs,” in *Journal of Combinatorial Theory Series B*, vol. 47, p. 274, 1989.

[18] R. Albert and H. J. L. Barabasi, “Error and attack tolerance of complex networks,” in *Nature*, vol. 406, pp. 378–382, 2000.

[19] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Process on Complex Networks*. Cambridge University Press, 2008.

[20] D. Alderson, L. Li, and C. D. W. Willinger, “Understanding internet topology: Principles, models, and validation,” in *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. 13, pp. 1205–1218, 2005.

[21] A. Sydney, C. Scoglio, P. Schumm, and R. Kooij, “Elasticity: Topological characterization of robustness in complex networks,” in *IEEE/ACM Bionetics*, 2008.

[22] P. Mahadevan, D. Krioukov, M. Fomenkov, and B. Hauffaker, “Lessons from three views of the internet topology,” in *arXiv.org:cs/0508033*, 2005.

[23] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, “Systematic topology analysis and generation using degree correlations,” in *ACM/SIGCOMM*, vol. 47, p. 274, 2006.

[24] J. Dong and S. Horvath, “Understanding network concepts in modules,” in *BMC Systems Biology*, vol. 1, 2007.

[25] D. L. Alderson, “Catching the network science bug: Insight and opportunity for the or researcher,” in *INFORMS*, vol. 56, p. 10471065, 2008.

[26] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Asia Limited: Prentice-Hall, 2005.

[27] D. Watts and S. Strogatz, “Collective dynamics of ’small-world’ networks,” in *Nature*, vol. 393, pp. 440–442, 1998.

[28] M. Newman, “Properties of highly clustered networks,” in *Physical Review E*, vol. 68, 2003.

[29] L. Freeman, “Centrality in social networks: Conceptual clarifications,” in *Social Networks*, vol. 1, pp. 215–239, 1978.

[30] M. Newman, “Mixing patterns in networks,” in *Physical Review E*, vol. 67, 2003.

[31] N. Alon, “On the edge-expansion of graphs,” in *Combinatorics, Probability and Computing*, vol. 6, pp. 146–152, 1997.

[32] B. Mohar, “Isoperimetric numbers of graphs,” in *Journal of Combinatorial Theory Series B*, vol. 47, pp. 274–291, 1989.

[33] P. Golovach, “Computing the isoperimetric number of a graph,” in *Cybernetics and System Analysis*, vol. 30, 1994.

[34] N. Linial and A. Wigderson, “Expander graphs and their applications,” in *Lecture*, 2003.

[35] F. Leighton and S. Rao, “Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms,” in *Journal of the ACM*, vol. 46, pp. 787–832, 1999.

[36] C. Gkantsidis, M. Mihail, and A. Saberi, “Conductance and congestion in power law graphs,” in *ACM SIGMETRICS*, vol. 1, pp. 148 –159, 2003.

[37] J. Wu, Y. Tan, H. Deng, B. L. Y. Li, and X. Lv, “Spectral measure of robustness in complex networks,” in *arXiv:0802.2564*, 2008.

[38] I. Ilse and R. M. Wills, “Analysis and computation of google’s pagerank,” in *IMACS*, 2005.

[39] F. R. K. Chung, “Spectral graph theory,” in *CBMS REgional Converence SEries in Mathematics*, vol. 92, 1997.

[40] M. Fiedler, “Algebraic connectivity of graphs,” in *Czechoslovak Math Journal*, 1973.

[41] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, “The robust yet fragile nature of the internet,” in *PNAS*, vol. 102, pp. 14497–14502, 2005.

[42] J. D. Murray, “Modeling the spread of rabies,” in *American Scientist*, pp. 280–284, 1987.

[43] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Keitzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker, “A human protein-protein interaction network: a resource for annotating the proteome,” in *Cell*, vol. 122, pp. 957–968, 2005.

[44] U. Brandes, “A faster algorithm for betweenness centrality,” in *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[45] M. Newman, “A measure of betweenness centrality based on random walks,” in *Social Networks*, vol. 27, pp. 39–54, 2005.

[46] NWBTEAM. Network Workbench Tool, Indiana University, Northeastern University, and University of Michigan. http://nwb.slis.indiana.edu, 2008.

[47] L. B. R. Albert and E. Bonabeau, “Scale-free networks,” in *Scientific American*, vol. 288, pp. 60–69, 2003.

[48] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H. E. Stanley, “Optimization of network robustness to waves of targeted and random attacks,” in *PHYSICAL REVIEW E*, vol. 71, 2005.

[49] A. Fabrikand, E. Koutsoupias, and C. Papadimitriou, “Heuristically optimized trade-offs: A new paradigm for power laws in the internet,” in *ICALP*, vol. 2380, pp. 110–122, 2002.

[50] P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat, “Orbis: Rescaling degree correlations to generate annotated internet topologies,” in *ACM/SIGCOMM*, 2007.

# Appendix A

# APPENDIX

| Networks | $\sharp$ Nodes | $\sharp$ Links | Density | Diameter | $\lambda_2$ | Spectral Gap |
|---|---|---|---|---|---|---|
| ER dense | 1000 | 4505 | 0.00902 | 7 | 0.7942 | 0.3688 |
| MySpace | 955 | 10976 | 0.02409 | 4 | 0.709 | 1.6753 |
| Watts-Strogatz 1 | 1000 | 3000 | 0.00601 | 7 | 1.1619 | 0.2113 |
| PA 2 | 1000 | 2964 | 0.00593 | 6 | 1.2514 | 0.2794 |
| ER sparse | 1000 | 2009 | 0.00402 | 12 | 0.1508 | 1.9223 |
| PA 1 | 1000 | 1981 | 0.00397 | 8 | 0.5443 | 1.8339 |
| Watts-Strogatz 2 | 1000 | 2000 | 0.004 | 9 | 0.5043 | 0.1427 |
| YouTube | 1089 | 1576 | 0.00266 | 12 | 0.0547 | 1.9704 |
| Flickr | 967 | 1515 | 0.0032 | 12 | 0.099 | 1.9413 |
| meshcore | 1000 | 1275 | 0.00255 | 3 | 0.3869 | 1.7948 |
| near-regular 2 | 992 | 3781 | 0.00769 | 31 | 0.0038 | 0.0038 |
| HOT 2 | 1000 | 1049 | 0.0021 | 12 | 0.0062 | 1.9979 |
| ringcore | 1000 | 1000 | 0.002 | 14 | 0.0016 | 1.9998 |
| HOT 1 | 939 | 988 | 0.00224 | 10 | 0.0079 | 1.9973 |
| PA sparse | 1000 | 1049 | 0.0021 | 14 | 0.0151 | 0.0076 |
| Abilene | 886 | 896 | 0.00229 | 10 | 0.0063 | 1.9975 |
| near-regular 1 | 992 | 1921 | 0.00391 | 61 | 0.0096 | 0.0025 |

**Table A.1**: *Network characteristics Table A.1*

| Networks | Conductance | Average shortest path | Clustering coefficient | Heterogeneity |
|---|---|---|---|---|
| ER dense | 0.2756 | 3.391 | 0.0098192 | 0.331 |
| MySpace | 0.4335 | 2.013 | 0.328865 | 2.027 |
| Watts-Strogatz 1 | 0.2109 | 4.14 | 0.0419 | 0.301 |
| PA 1 | 0.2482 | 3.534 | 0.02818 | 1.109 |
| ER sparse | 0.1695 | 5.154 | 0.00638 | 0.491 |
| Watts-Strogatz 2 | 0.1787 | 4.177 | 0.02697 | 1.185 |
| watts1 | 0.152 | 5.294 | 0.003214 | 0.37 |
| YouTube | 0.1251 | 5.096 | 0.0083 | 1.319 |
| Flickr | 0.137 | 4.624 | 0.015 | 1.394 |
| meshcore | 0.1275 | 2.911 | 0.14132 | 3.796 |
| near-regular 2 | 0.0241 | 14.706 | 0.45127 | 0.133 |
| HOT 2 | 0.0392 | 7.144 | 0 | 1.892 |
| ringcore | 0.0021 | 8.196 | 0 | 3.122 |
| HOT 1 | 0.0311 | 6.812 | 0 | 2.032 |
| PA sparse | 0.0401 | 5.793 | 0.00278 | 1.892 |
| Abilene | 0.0136 | 6.95 | 0.007556 | 2.09 |
| near-regular 1 | 0.016 | 21 | 0 | 0.089 |

**Table A.2**: *Network characteristics Table A.2*