4-2014

Robustness of complex networks to global perturbations

Samuel Heiserman
Binghamton University--SUNY, sheiser1@binghamton.edu

ROBUSTNESS OF COMPLEX NETWORKS TO GLOBAL PERTURBATIONS

BY

SAMUEL HEISERMAN

BA, Binghamton University 2012

MS, Binghamton University 2014

Submitted in partial fulfillment of the requirements for

the degree of Master of Science in Systems Science

in the Graduate School of

Binghamton University

State University of New York

2014

Accepted in partial fulfillment of the requirements for

the degree of Master of Science in Systems Science

in the Graduate School of

Binghamton University

State University of New York

2014

April 23, 2014

Dr. Hiroki Sayama, Committee Chair

Departments of Bioengineering & Systems Science and Industrial Engineering, Binghamton University

Dr. Harold Lewis, Committee Member

Department of Systems Science and Industrial Engineering, Binghamton University

Dr. Mohammad Khasawneh, Committee Member

Department of Systems Science and Industrial Engineering, Binghamton University

Abstract

This thesis studies the robustness of complex dynamical networks with *non-trivial topologies* against *global perturbations,* following Robert May's seminal work on network stability, in order to find critical stability thresholds of global perturbations and to determine if their impact varies across different network topologies. Numerical analysis is used as the primary research method. Dynamical networks are randomly generated in the form of a coefficient matrix of stable linear differential equations. The networks are then inflicted with global perturbation (i.e., addition of another random matrix with varying magnitudes) and their stabilities are tested for each perturbation magnitude, to determine at what scale of global perturbation they are jarred to instability.

The results show a monotonic decrease of the instability threshold over increasing link density for all network topologies. For a given link density, random regular networks show highest robustness against global perturbation, closely followed by Watts-Strogatz small-world networks and Erdos-Renyi random graphs, and then Barabasi-Albert scale-free networks are least robust among the four topologies tested. Fully connected networks used in May's original work are found to be consistently unstable in the presence of global perturbation of any magnitude. These findings offer useful implications for the robustness and sustainability/vulnerability of real-world complex networks with nontrivial topologies.

I would like to dedicate this project in two parts.

The first is to the many people who got me here. A condensed list could be as follows: Doctoral students Tom Raway and Rian Shams for their great insights and proficiencies in refining the code; Professors Hal Lewis and Hiroki Sayama for their remarkable investments of time and patience in the project and in me as a student; and my parents Maura and Arthur for almost everything else.

The second is to the endeavor of modeling and understanding global impacts and systemic risk in all forms. Amongst all the uncertainty and speculation surrounding the 2008 housing crash, it seems objectively clear that a lack of knowledge about the robustness of a complex system overall can lead in many cases to the system's collapse.

When human society grows to a size and scale of resource demand where it becomes dependent on vast networks to manage and distribute these resources, it becomes imperative for our social function that we explore what is capable of destabilizing them. Under the modern condition of ecological and financial volatility, our collective security is only as robust as our knowledge of these fragilities. In order to bolster or design-out the weak points of these complex systems we all depend on, we first have to find them.

Acknowledgements

I would like to acknowledge the keen eye of my advisor Dr. Hiroki Sayama for matching my broad conceptual focus with May's work and the analytical methods it employs, and the guidance and insights of Doctoral candidate Tom Raway for his continued help throughout the project.

Table of Contents

List of Figures

List of Abbreviations

BA – Barabasi-Albert scale-free network model

CG – Complete (fully connected) graph model

ER – Erdos-Renyi random graph model

GP – Global Perturbation

RR – Random regular graph model

WS – Watts-Strogatz small-world network model

List of Notations

n – network size (i.e., number of nodes)

A – an $n \times n$ matrix formed by: $A = a\,E - I$

a – *link strength* parameter

E – an $n \times n$ random matrix whose elements are sampled from a uniform distribution [-1,1] where edges exist in the network (otherwise 0; the network topology is determined by the network model as well as n and C)

C – link density parameter

I – an $n \times n$ identity matrix

B – an $n \times n$ matrix formed by: $B = A + \tau P$

τ – global perturbation magnitude

P – another $n \times n$ random matrix whose elements are sampled from a uniform distribution [-1,1] where edges exist in the network (otherwise 0)

Chapter 1: Introduction

The objective of this project is to expand the known set of methods for testing network robustness by introducing a new type of perturbation, one designed to affect systems globally. This *global perturbation* (GP) is defined by its simultaneous effect on all network links, and is tested for its effect on network stability across a range of *link strength a* values, over a set of network models whose structure is observed in real world networks. This model set consists of: Barabasi-Albert (BA) scale-free, Watts-Strogatz (WS) small-world, Erdos-Renyi (ER) random, and random regular (RR) network models. The general motivation of this research is to derive potential implications for real world networks which have demonstrated themselves vulnerable to these sorts of sudden, system-wide changes to their environments.

1.1 *Domain Contexts*

Two leading domain contexts relevant to the topic of network stability are ecological community webs and financial transaction networks, due to the necessity of their robust functioning for the health and stability of our human societies. Both of these systems have exhibited volatility in the face of certain environmental changes, including global changes which affect each network as a whole. In the ecological context, changes to key factors such as temperature and resource availability have the potential to destabilize local food webs, as they interfere with the connections between interdependent species and trophic levels (Allesina et al., 2012).

In the financial context, changes to factors that affect these networks globally, such as central banks' interest rates and the federal governments' tax policies, can have significant effects on the stability of national economies overall. This became greatly significant for much of the US population when the sub-prime housing bubble burst in 2008, a situation due at least in part to the de-regulation of banks' lending policies and of the derivatives markets during the 1990's (Squartini et al., 2013). Although the global perturbations within the model studied in this thesis are highly abstract compared to these real-world networks, there may be some significant patterns to find within the range of global perturbation magnitudes and/or across the network models.

1.2 *May's Model*

This investigation of network stability is conceptually built upon a model proposed by mathematician and theoretical biologist Robert May in his seminal paper "Will a Large Complex Network Be Stable" (May, 1972). May's purpose was to explore possible relationships between networks' stability and their complexity, by testing the stability of randomly generated networks across a range of network sizes n, connectivities C and link strengths a. Network size n is the number of nodes a network is composed of, connectivity C is the number of its links, and link strength a is the average magnitude of the weights of those links. May found that stability was much more likely in networks which maintained certain balances between connectivity C and connection weight a, concluding these precise points to be critical stability thresholds, below which a network will likely remain stable and above which it will likely collapse.

1.3 *The Proposed Model*

Though the model proposed in this thesis is similar to May's in its methods of network generation and stability testing, it diverges from it and expands upon it in several key ways: the first is the introduction of the global perturbation (GP). While May studied his set of networks by varying their network traits $\{N,C,a\}$, he did not consider any external perturbations added to them. In this thesis, a set of stable networks will be disrupted by introducing system-wide tremors in the form of global perturbations (GP's) to the networks' link strengths a. For each network, the magnitude of the perturbations administered is increased with each stable result, until a destabilizing magnitude (critical τ value) is found. The second way the proposed model differs from May's is the range of network types it performs these tests on. While his model generates and tests only random graphs, the proposed model incorporates a set of non-trivial network models: Barabasi-Albert scale-free networks and Watts-Strogatz small-world networks, as well as Erdos-Renyi random graphs and random regular graphs. Complete graphs (in which all nodes are linked to each other) are also tested in the preliminary stability testing, but were found unstable at such a high rate that the model was excluded from the main experiment. The final way the proposed model differs from May's is the size of the networks it tests, i.e., $n = 1000$ each. In his 1972 work published in Nature, he referred to the network sizes used by his predecessors, Gardner and Ashby, of $n = \{4,6,8\}$, distinguishing the networks he would be testing as 'larger', $n > 10$ (May 1972). This thesis studies significantly larger networks than those studied in their papers.

The goal for this numerical analysis is to discover which network models are most volatile to which levels of perturbation magnitude, and how this robustness performance varies over a range of link strengths a. The future purpose of this model could be to serve as theoretical bases for forming hypotheses about which real world networks are safe to change in this sudden, global way and which should be changed more gradually or piece by piece. The lack of understanding about the depth and nature of robustness in our social and ecological networks has left us vulnerable to major collapses in these systems, and it seems that unless our set of measures and forecasts of robustness becomes itself more robust, we are bound to get hit by more unforeseen failures, and at greater costs.

Chapter 2: Background and Related Work

2.1 *Graph Theory and Network Science*

This section introduces the field of Network Science, which forms the basis of this thesis. A network, or ‘graph’ in mathematical terms, is in essence a collection of entities connected to each other. These entities, whether they represent people, ideas, computers, cells, nations or any form of interacting component are referred to generally as ‘nodes’ (or ‘vertices’) and connected to each other by sets of ‘links’ (or ‘edges’).

Network Science is a new interdisciplinary field which has evolved rapidly in the last two decades around the pursuit of modeling the complex webs of connection that compose our natural, technical and social environments. The field has been defined by the United States National Research Council as “the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena" (National Research Council, 2006), and as this description implies it has a very broad impact-scope, with applications ranging across industrial sectors, government services and academic fields alike (Strogatz, 2001). The subject of investigation may be a power or telecommunications grid, a neuron or species interaction web, a social advocacy group or political faction, or any other system whose aggregate behavior is based on the interactions of its component parts. These real world systems in all of their complexity cannot be fully characterized even by the most sophisticated models available, as there is always more going on than is understood. Inescapable as that is, within the vast and mostly uncharted field of complex systems science this network approach has proven itself highly effective, especially in systems whose key factors are better understood and thus more fit for mathematical abstraction (Strogatz, 2001).

One organization, or taxonomy of many networks explored within Network Science, is given below (Dodds, 2014):

1. *Physical Networks* (whose structures are physically embedded in the external world)
    1. Types:
        1. Distribution (branching)
        2. Re-distribution (cyclical)
    2. Examples:
        1. Riverways
        2. Neural Pathways
        3. Trees and Leaves
        4. Blood Pathways
        5. Power Grids
        6. Roadways
        7. The Internet
2. *Interaction Networks* (maps of interactions between organisms of different scale)
    1. Examples:
        1. The Blogosphere
        2. Biochemical Networks

3. Gene-Protein Networks
4. Food Webs
5. WWW Hyperlinks
6. Phone Calls
7. Airline Routes
8. The Media
9. Sexual Relationships
10. Friendships & Acquaintances
11. Boards & Directions
12. Social Media (facebook, twitter, etc)
13. Creative Networks (webs of artistic collaboration)

3. *Relational Networks* (webs of concepts, or human interaction with resource supplies)
    1. Examples:
        1. Consumer Purchases
        2. Thesauri (words connected by similarity of meaning)
        3. Knowledge/Databases and Ideas
        4. Metadata - Tagging (such as on flickr)
        5. Search of Scientific Materials (webs of clicks between subjects online)

The study of real world networks conducted in this new field of Network Science has mathematical roots dating back to the '7 Bridges of Konigsberg' problem, proposed by mathematician Leonhard Euler in 1735. The problem was as follows: The city of Konigsberg, Prussia lay on either side of the Pregel river and included two islands, connected to each other and to the mainland by seven bridges (as shown in Figure 1). The problem was to find a route that crossed each bridge exactly once, without crossing the water by any other means (Briggs, 1986). Euler proved that there was in fact no possible solution, as each of the four land masses was touched by an odd number of bridges when an even number would be needed for all land masses (possibly except for two of them). Along with this finding, Euler also critically observed that to approach this problem required knowledge only of the connections between the land masses (the nodes and links) and nothing else, allowing him to abstract the original detailed map of the city to a simple diagram of dots and lines, an object which became mathematically referred to as a graph, as shown on the right in Figure 1.

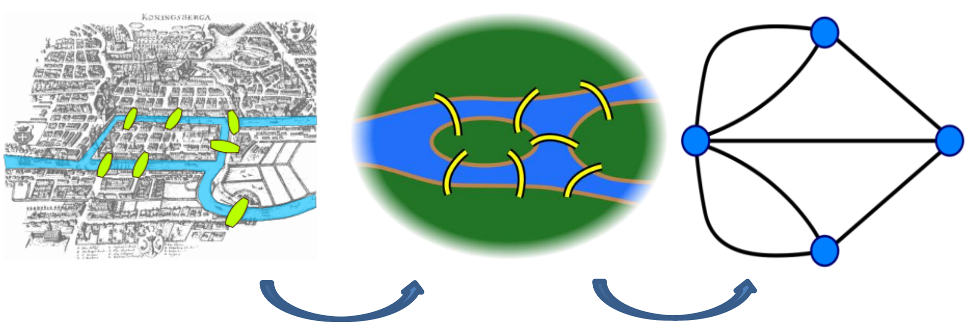


Figure 1. Emergence of Graph Theory (Briggs, 1986)

This valid formalization of the bridges and islands as 'edges' and 'vertices' laid the groundwork for a new branch of mathematics now known as Graph Theory, which models pairwise relationships between given entities by representing them graphically as collections of nodes connected by sets of links (Briggs, 1986). These sets of nodes and their links (mathematically referred to as vertices and their edges) together form the mathematical objects known as 'graphs', formalized as $G = [V,E]$, where V is a set of vertices and E as a set of edges.

Graph Theory has been interdisciplinary since its inception with Euler, as several of its foremost early applications were conducted in different fields. The term 'graph' itself was coined by a mathematician named James Joyce Sylvester in an analogy he posed between "quantic invariants" and "co-variants" of algebraic and molecular diagrams (Briggs, 1986). One of today's biggest applications of Network Science, social network analysis, originated with psychologist James Monroe and his introduction of the 'sociogram', a graph based depiction of the social structure of boys and girls in an elementary school. This new approach to studying social ties was renowned at the time, and was published in the New York Times in 1933 (Briggs, 1986).

2.2_*Network Models*

The modern renaissance of graph theory began with its evolution from a tool for static relational mappings to one for state of the art dynamical network models when it was brought together with probability theory by mathematicians Paul Erdos and Alfred Renyi to form the Erdos-Renyi random graph model in 1959 (Briggs, 1986). Within this model, network links are formed at random, as each potential link is given an equal probability of being formed (Costa et al, 2007).

This method of generating networks is known as the Erdos-Renyi model. It produces what are also called 'random' networks, which constitute one of an established set of network models, mechanisms by which networks of different structure are generated (Costa et al, 2007). The discovery of network models which take on the structures observed in real world networks, such as Watts's and Strogatz's 'Small-World' and Barabasi's and Albert's 'Scale-Free' networks, has been greatly responsible for validating network modeling as a critical new approach to designing, maintaining and generally understanding networks of all types. The scale-free property for example, which was discovered by Barabasi and Albert in 1999 in the distribution of hyperlinks on the World Wide Web, indicates when identified in a network that the network will be much more vulnerable to targeted attacks on its hubs than to random failure of its nodes (Albert et al, 2000).

This development in the field has also put Network Science on the map in the public discourse, with network terms such as 'degrees of separation', 'viral' content and system 'hubs' commonly understood due to their sheer ubiquity and universally intuitive nature (Albert et al, 2000).

2.3_*Network Properties*

In this section I'll introduce some terminology used to characterize network traits. One main advantage of networks is its broad accessibility to newcomers, as its vocabulary can be interpreted by each person in whatever context is most comfortable to them. The measurement of *network size,* for example, can be interpreted as how many people are at a party, cars are on a highway, school are in a city or proteins are at work in a metabolic process. The simple question is: how many nodes are there in the system. The network *density* is a measurement of how many links exist between the nodes compared to the maximum number possible if each node were linked to each other (Costa et al., 2007). This could be interpreted as the density of close friendships held between students in a class, predator/prey ties held between species in a community food web, or collaborations forged between members of Congress.

The *directedness* of a network's links indicates which way(s) the links connect, whether a given link will only transmit from node A to node B but not back from B to A. This can be easily interpreted in the context of roadway networks, as many roads can only go one way. Likewise the *weight* of a network's links indicates how strong or dense a connection is. In the same context, this could represent how much traffic there is on a given road, or what the speed limit is. The *average degree* of a network simply measures the average number of links (the *degree*) held by each node (Costa et al., 2007). This is closely related to network *density*, though it differs in that it does not consider the number of possible links. The *average path length* of a network measures how many steps compose the shortest path between one node and another on average, found simply by averaging the number of steps (also known as the 'degree of separation' in laymen's terms) it takes to connect every combination of two nodes in a network (Costa et al., 2007). The network *diameter* is the longest among all of these shortest paths (Costa et al., 2007).

A node's *clustering coefficient* is a measurement of how connected its neighbors are. In other words, how many of your friends are friends with each other. This is calculated as a ratio of existing ties between neighbors to the maximum possible number of ties. The *clustering coefficient* of a network is found by taking the average of those of all its nodes (Costa et al., 2007).

Another key tool used to analyze networks is *node centrality*, which yields levels of influence that individual nodes have on the network. This study of centrality is largely comprised of a set of four main measures: *degree* centrality, which tracks how many connections each node has; *closeness* centrality, which tracks the distance in space each node is from all others on average; *betweenness* centrality, which tracks what proportion of all shortest paths each given node is part of; and *eigenvector* centrality, which tracks how well connected a node's connections are by measuring the *average degree* of its neighbors. These centrality measures are used, for example, to identify which members of a dangerous group should be monitored, captured or executed in order to gain the most relevant information or optimally disrupt the network's functioning (Barabasi, 2009). Figure 2 summarizes some of this discussion of network models and properties.

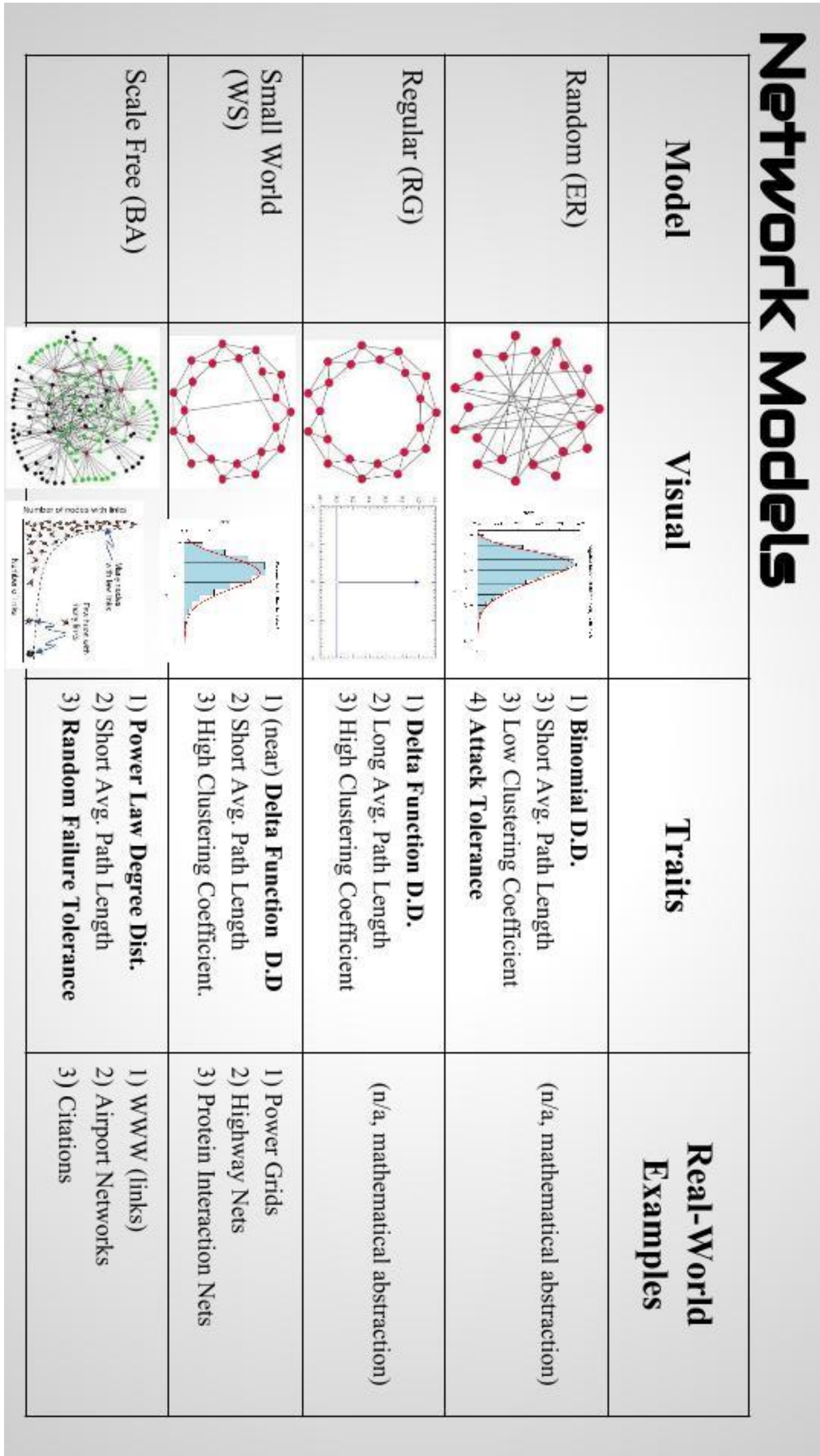


Figure 2. Network Models

A more exhaustive list of network properties is given below (Costa et al., 2007):

- *Degree Distribution*
- *Assortativity/Homophily* (disproportionate connectivity between nodes with similar degree)
- *Motifs* (recurring substructures within networks)
- *Modularity* (subgrouping or 'community' structure formation within a network)
- *Concurrency* (simultaneity of connections between nodes over time)
- *Hierarchical Scaling* (hierarchical structure of nodes over multiple scales)
- *State-Topology Coevolution* (the interdependence of a network's structure and the activity taking place on it)
- *Robustness* (the ability of a network to maintain stability despite failures or perturbations)

2.4 *Dynamical Systems and Stability*

In order to fully grasp the essence of network stability one must first have a firm grasp on May's original model. This calls for a basic understanding of Dynamical Systems Theory, a branch of mathematics which deals with systems' autonomous change over time. Autonomous change implies that each of the system's updates is based only on its own properties, deemed 'autonomous' as it is not affected by any external factors. The field was conceived within the development of Newtonian mechanics and carries up through modern theories of nonlinear dynamics, focusing on systems' underlying dynamical mechanisms, rather than just properties of static observations.

A dynamical system is a system whose state is entirely described by a finite set of variables, and whose behavior is entirely determined by predetermined rules. So what happens on Day 2 depends completely on a defined set of traits and what their values were on Day 1. Examples include: motion of celestial bodies, simple pendulum swinging, population growth, and behavior of two agents in a negotiation such as the Prisoner's Dilemma.

One basic trait of dynamical systems is how they are formulated mathematically, whether in discrete or continuous time. Both May's model and the model studied in this thesis are built to operate in continuous time. This means that the equations composing the models are differential equations, in the general form of $dx/dt = F(x,t)$ where F is some function determining the rule that the system's behavior will follow. This is distinct from discrete time models, in which time is broken into discrete steps, the comprising equations of which are difference equations and take the general form $x_t = F(x_{t-1},t)$ where x_t is the variable describing the state of the system at time step t.

Another fundamental trait of these systems' mathematical formulation is whether they are linear or nonlinear. Linear equations are desired across all areas of applied mathematics because they are familiar and well behaved. Linear dynamical systems are sure to be analytically solvable and to show either convergence to an equilibrium point (exponential decay), divergence from an equilibrium point (exponential growth), periodic oscillation, or some combination thereof. Nonlinear systems on the other hand are not so kind, and our understanding of them is less well defined in many cases. These systems are often not analytically solvable, and tend to show much more complex and mysterious behaviors than their linear counterparts.

The network trait which is focused on in both May's work and the proposed model is their stability. According to May, the collection of interactions between ecosystem species, which his networks were meant to model, generally follow a set of 'quite' non-linear first order differential equations (May, 1972). Despite this nonlinear structure, there is a linear method which can be used to carry out stability testing, known as Linear Stability Analysis. This is done by examining the behavior of the system just around its equilibrium points, such that the stability of the equilibrium point is characterized by the equation: $dx/dt = Ax$ (May, 1972). This means that the change in x (the set of disturbed populations) can be represented by a matrix A, wherein the interactions between each species near equilibrium are linearly approximated, multiplied by a vector x.

This process is carried out as follows: First the function is linearized at this equilibrium point to produce a matrix (called a Jacobian matrix), and then the matrix's eigenvalues are checked to see if any have values greater than zero for their real parts. The reason the system's eigenvalues around these steady points are referred to is that they indicate whether the system, after impact from some perturbation, is gravitating back towards the equilibrium point and stability or moving further away and towards instability. For the eigenvalues' real parts the distance of the system's state from equilibrium is increasing (destabilizing) when the real part is greater than 0, and decreasing (stabilizing) when it is less than 0. If all eigenvalue's real parts have negative values, the system is moving back toward equilibrium and can be considered stable. This criterion of linear stability is used in both May's work and the proposed model, though it is just one of a set of methods used to analyze stability in networks (Ellens and Kooji, 2013).

2.5 *May's Model in Depth*

The model on which this thesis is based was designed by theoretical biologist and ecologist Robert May in the early 1970's, to test the relationship between network complexity and stability for larger networks than had been investigated by his predecessors, M.R. Gardner and W.R. Ashby. May aimed to find out whether the key finding of their work, the sharp transition from stability to instability observed at certain critical thresholds of complexity, scaled to networks of greater size. The model was conceived in the domain context of ecological food webs with many interacting species, and the variables comprising the system were: number of species *n*, average density of links between species *C*, and link strength *a*, representing how heavily dependent species were on each other (May, 1972). This is summarized in figure 3.

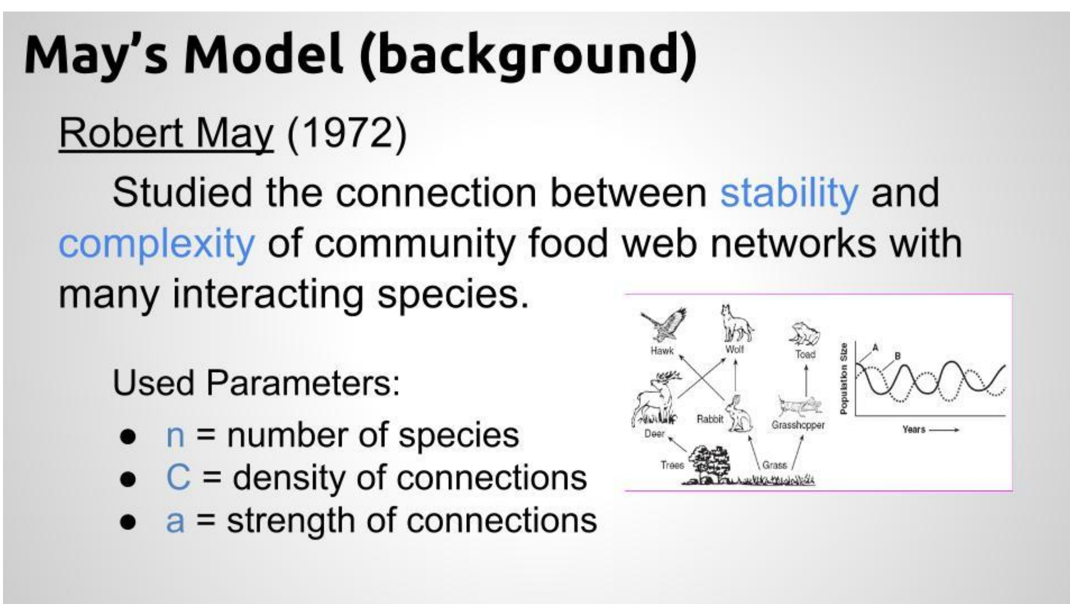


Figure 3. May's Model

This information of which species interact with which others can be drawn directly from a diagram of a food web (or 'trophic' web), though for the sake of mathematical generalizability May chose to computationally generate this data, making several simplifying assumptions in constructing the model:

1. Each individual species on its own will maintain a stable population over time.
2. Each matrix element (encoding an interaction between species) is assigned from a distribution of random numbers between -1 and 1 of mean value 0, making each interaction equally likely to be positive or negative.
3. Link strength a is uniformly scaled for all interactions.

The model is formalized as follows:

$$A = aE - I$$

The matrix A represents an $n \times n$ matrix of interactions between species. It is composed of the matrix E (an $n \times n$ random matrix whose elements are sampled from a uniform distribution [-1,1]) times link strength a, minus the $n \times n$ identity matrix I. The topological structure of May's networks followed the random graph model.

Once the matrix has been generated, the stability of the system is tested. A system is declared stable if and only if all of the eigenvalues of matrix A have negative real parts. For each value of n, C and a, a probability is found that a matrix with those traits will correspond to a stable system, denoted by: $P(n, C, a)$. This is also done for each combination of a and n, denoted by $P(n, a)$. This analysis uncovered critical stability thresholds, where stability was overwhelmingly likely in matrices with: $a < \frac{n^{-1/2}}{2}$ and equally unlikely in matrices with: $> \frac{n^{-\frac{1}{2}}}{2}$. This transition was very sharp and consistent due to the statistical fact that "although individual matrix elements are liable to have any value, by the time one has an $n \times n$ matrix with n^2 such statistical elements, the total system has relatively well defined properties" (May, 1972). A nearly identical stability threshold is found when network density C is introduced, measured as a ratio of actual links to topologically possible links within a network. Matrices with $a < nC^{-1/2}$ were surely stable, while those with $a > nC^{-1/2}$ were surely not.

In uncovering sharp transitions to instability beyond these critical thresholds, May's results did indeed concur with Gardner and Ashby's, showing that their finding of sharp transition to instability did indeed scale to larger networks. These results suggest that too large of a density or link strength (C or a) is detrimental to a network's stability, and this effect is more pronounced with larger n. This is a balance that is observed in many real world ecosystems (Allesina et al., 2012). The other noteworthy result of his model was that webs were much more likely to be stable if they were arranged in "blocks", as a web of 12 species had a much higher stability rate when organized into three separate 4-species communities. This trait of modularity is also observed in many real world ecosystems, and is believed to contribute significantly to networks' robustness in general (May et al., 2008).

2.6 *Network Robustness*

Our lives are composed of series of networks; from the infrastructural networks that provide us with water, electricity, communications and transportation to our institutional networks that govern the creation and flow of money and laws, to our interpersonal social networks describing who we work and spend time with, to the ecosystems which continually generate the natural resources we need to survive. As crucial as this web of webs is, its consistent performance can lead many to take it for granted, especially among the fortunate few in the first world who have reaped its benefits throughout our lives nearly without interruption. It is when interruptions do occur, as has happened in recent years in the forms of power blackouts, financial crashes, and ecosystem biodiversity loss and resource depletion due to over-consumption and pollution, that we are forced to step back and reexamine the integrity of the systems we have created and their effect on the natural systems which created us. It is for this process of reexamination that one can see the greatest real world implication of the field of network stability and robustness.

Robustness is the "ability of a network to continue performing well when it is subject to failures or attacks" (Ellens and Kooji, 2013). There are general network traits which have been found to foster robustness across networks of diverse domains. Another, much more recent work of May's entitled "Ecology for Bankers", suggests that ecological food webs of interacting species and financial networks of interacting banks are both more robust when they are structurally more modular, redundant, and disassortative (May et al., 2008). Modularity is the "degree to which the nodes of a system can be decoupled into relatively discrete components" (May et al., 2008). This quality is expressed by the presence of community structures, clusters within networks characterized by high connectivity between their internal nodes and sparse connectivity to those outside the community. This principle is applied in the context of forest fire management, in which forests are preventatively divided into distinct modules in order to limit fires' potential to spread (May et al., 2008).

Redundancy is equivalent to the availability of alternative pathways between nodes. This concept is key to the structural function of the Internet. The messages sent over the web reach their destinations with such high reliability not because every router is impregnable, but because of the algorithms' ability to find alternate routes when encountering defective ones. It is because of this, along with the Internet's scale-free structure, that a random failure of 80% of all web sites would still not crash the system completely (Albert et al., 2000).

Dissortativity has to do with the connectivity between highly connected 'large' nodes and much lesser connected 'small' nodes. In dissortative networks, large nodes have their connections "disproportionately with small nodes", while the small nodes "connect with disproportionately few large ones" (May et al., 2008). This is observed in the ecological network structures of plants and pollinators and both marine and freshwater food webs, as well as in the Fedwire interbank payment network, as "large banks were disproportionately connected to small banks and vice versa" (May et al., 2008).

The success of these three traits to produce more robust network structures within social, natural, and technological domains seems to provide solid grounds for their prioritization as characteristics to be built into the design of networks of all types in the future. The research question addressed in this project, of the effect of global perturbations on network stability, will also yield broadly applicable knowledge about network robustness that will be useful to the design and security of future networks of all types.

Chapter 3: Methods

This section lays out in detail the design of the numerical experiment conducted in this thesis, along with the algorithm by which it was conducted. To set the stage for this the model's basic objective and assumptions will be re-overviewed briefly:

The objective of the model is to test the robustness of four network models by tracking their stability under the influence of global perturbations:

1. Barabasi-Albert scale-free networks
2. Watts-Strogatz small-world networks
3. Erdos-Renyi random graphs
4. Random regular graphs

Some simplifying assumptions were made in the formulation of both May's model and the proposed model, as follows:

1. The nodes comprising the networks affect each other linearly by the equation $dx/dt = Ax$, as characteristic to the method of linear stability analysis.
2. Each node would return to stable equilibrium if not influenced by other nodes.
3. Each link strength takes an initial value between -1 and 1, which is generated from a random $n \times n$ matrix of mean 0 and multiplied by a.
4. Each link's value is equally likely to be positive (+) or negative (-)

5. The global perturbations affect only links existing within the networks, as the matrices' 0-values (representing no connection between nodes) remain 0 after perturbation.

The algorithm for numerical analysis is implemented in the *Python 2.7* environment. The first step is to generate the matrix A, which is done by taking the adjacency matrix of a randomly generated graph of a given network model from within the NetworkX library (Hagberg et al., 2008). This adjacency matrix consists only of 1's and 0's, so the next step is to replace all the 1's with randomly generated values between -1 and 1, to capture the equal likelihood of negative and positive influence between nodes as is in May's model. The identity matrix is then subtracted to form the matrix A. Next this matrix A is tested for its stability, by checking if all of its eigenvalues are negative in their real parts. If it is found to be stable, it is passed into the next loop to form a perturbed matrix B, or if it is found unstable the 'while' loop repeats until a stable A-matrix is found.

With a stable A found, the next matrix B is constructed as the sum of the stable A and a perturbation matrix P (a random real $n \times n$ matrix whose components are random numbers sampled from the range [-1, 1]) multiplied by a perturbation coefficient τ. This product, within the equation $B = A + \tau P$ constitutes a global perturbation because it is being inflicted to the entire A matrix. The role of τ is to set the magnitude of this global perturbation. B matrices are then iteratively formed and tested for stability across ranges of τ for the purpose of identifying a critical τ value, the minimal level of perturbation force to destabilize each matrix. These τ ranges are applied and averaged over 30 repeats for each B matrix in order to counterbalance the model's stochastic nature. The maximum eigenvalue's real part is stored for each B and plotted over each τ value, to reveal at which point the network is jarred to instability, as measured by the lowest level of τ to destabilize each given network.

This process is carried out over three ranges of τ, first stepping logarithmically in base 2 from 0.01 to 2.56, to reveal approximately what τ magnitude first brings about instability. Once this critical τ vicinity is identified, a finer search is conducted by applying a linear range to the area on the log scale around which instability first arose, stepping in increments of 0.1. For example, if the logarithmic range revealed instability first arising between the values of 0.32 and 0.64, the new linear range over which the B matrices would be re-tested would be all τ values between 0.16 and 0.64 in steps of 0.1. This would yield a range of {0.16, 0.26, 0.36, 0.46, 0.56}. The final τ range is made up of a finer run over the same linear range, by stepping in increments of 0.01. This is obviously done to increase precision.

Once this data has been generated for each network model, it is presented on a 2D plot for the more refined linear τ range, in which the x-axis is the link strength a and the y-axis is the minimum τ value associated with each unstable B matrix (critical τ) over the 30 repeats. The initial exponential range is not included in the final results because it shows only an extremely broad range of τ values, and thus produces nearly identical plots each containing just a single point. This process is depicted in its entirety in pseudo-code in Figure 4.

Phase 1: *Preliminary experiment* test

```
For each network model Do
    For each n-value Do
        For each a-value Do
            unstable_count = 0
            Repeat × 10
                generate adjacency matrix A of given network model
                replace 1's with random values from -1 to 1
                subtract I from A
                calculate A's eigenvalues
                If maximum eigenvalue real parts > 0 then
                    network is unstable, unstable_count +=1
            Output a, unstable_count
        Output n, a, unstable_count
    Plot network model, n, a, unstable_count
```

Phase 2: *Main experiment*

```
n = 1000
set a to a specific value
For each network model Do
    While network is unstable Do
        generate adjacency matrix A of given network model
        replace 1's with random values from -1 to 1
        subtract I from A
        calculate A's eigenvalues
        If maximum eigenvalue real part <= 0 then
            network is stable, exit loop
    Initialize critical_τ_list
    Repeat ×30
        set log τ range
        Repeat ×3 (for log, linear and linear2 ranges):
            For τ in τ range:
                create P matrix as a random real n×n matrix
                create B-matrix as A + τP, changing only A's non-zero
                values
                If maximum eigenvalue real part of B > 0 then
                    network just got destabilized; configure the next τ
                    range and then exit loop
        Append τ to critical_τ_list
```

Figure 4. *Algorithm Pseudo-Code*

The preliminary experiment was conducted to test how stable networks are in general for given network models and parameter values. After the relevant network models and parameter values were determined, the main experiment was conducted to find critical global perturbation magnitudes (referred to as the critical τ) for each network model.

The preliminary experiment was carried out by generating *A* matrices and testing their stability over a linear range of *a* values between 0.1 and 1 and *n* values between 200 and 2000, and tracking how often for each set of a network model and parameter values (*n*, *a*) the generated *A*'s were found to be unstable. Unlike in May's model, here the link density *C* is held constant, because sparse connectivity (low *C*) is characteristic of scale-free network models. May was not limited in this way because he was working with dense random graphs *C*. Because the scale-free model is structurally limited in its set of possible *C* values, and the models should be evaluated with equal parameter values (*a*, *C*), *C* was set to 2 for all models in this thesis. This signifies that two links exist in a network for every node, on average. This did not apply to the complete graph (CG) model.

Within the main experiment, the value of *a* was set to range from 0 to 0.5 in steps of 0.1 and the *n* was set to 1000, as it was the greatest *n* value to fall within the desired computational time, as projected by the preliminary experiment results. The final results, which this process produced, are discussed in the next section.

Chapter 4: Results and Discussions

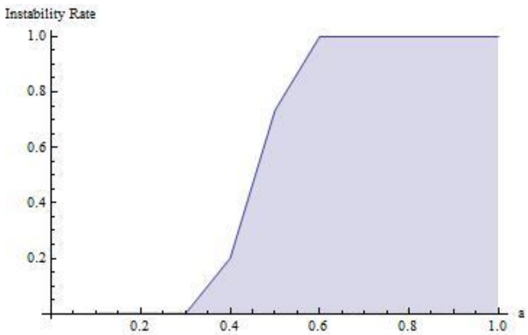
This section contains the results yielded from running the network models for both the *preliminary* and *main* experiments. The complete graph (CG) model was not run for the overall main experiment because of the consistent instability found from its preliminary experiment runs.

4.1 *Preliminary Experiment*

The preliminary experiment was conducted for each of the five network models, and revealed that the WS, BA, ER and RR models transition from virtually all stable at $a = 0.4$ to virtually all unstable by $a = 0.6$, while the CG model was found almost always unstable across the whole a range. The n range, conversely, made no significant difference to the proportion of generated A matrices found to be unstable. This is shown in Figures 6, 8, 10, 12 and 14.

Figures 5, 7, 9, 11, 13 and 15 show the rates of instability over a linear range of a values from 0.1 to 1.0 with step size 0.1, one figure for each of the five network models at $n = 1000$. Figures 6, 8, 10, 12 and 14 show these same rates over linear ranges of both a and n, with the same a range and an n range of 200 to 2000 with step size 200.

Scale-Free Networks (BA)



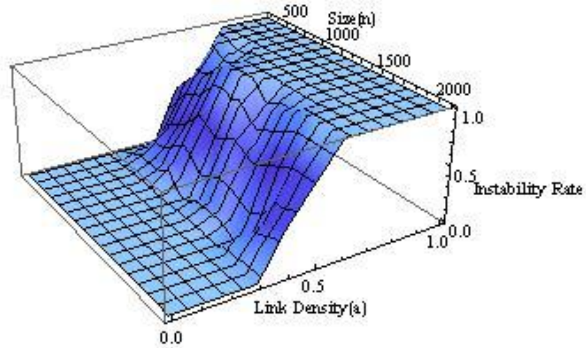
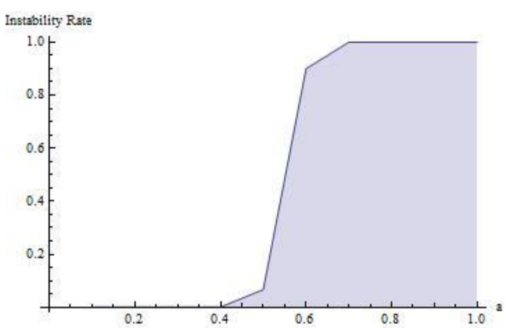


Figure 5. Instability rate of Barabasi-Albert scale-free networks for varying *a* values, plotted in 2D. *n* = 1000, *C* = 2.

Figure 6. Instability rate of Barabasi-Albert scale-free networks for varying *a* and *n* values, plotted in 3D. *C* = 2.

Small-World Networks (WS)



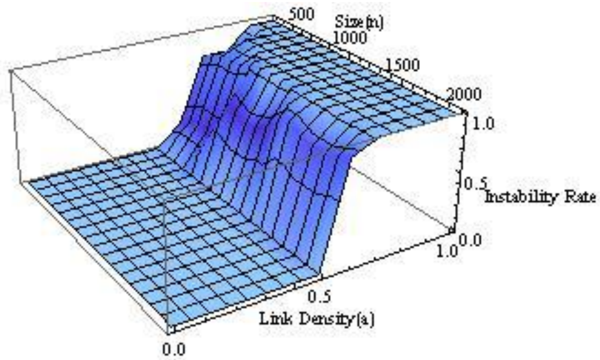


Figure 7. Instability rate of Watts-Strogatz small-world networks for varying *a* values, plotted in 2D. *n* = 1000, *C* = 2.

Figure 8. Instability rate of Watts-Strogatz small-world networks for varying *a* and *n* values, plotted in 3D. *C* = 2.

Random Regular Graphs (RR)

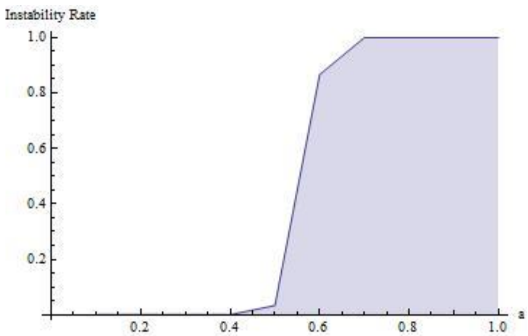


Figure 9. Instability rate of random regular networks for varying *a* values, plotted in 2D. *n* = 1000, *C* = 2.

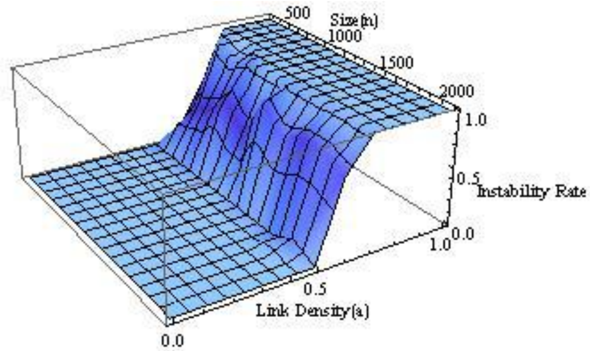


Figure 10. Instability rate of random regular networks for varying *a* and *n* values, plotted in 3D. *C* = 2.

Complete Graphs (CG)

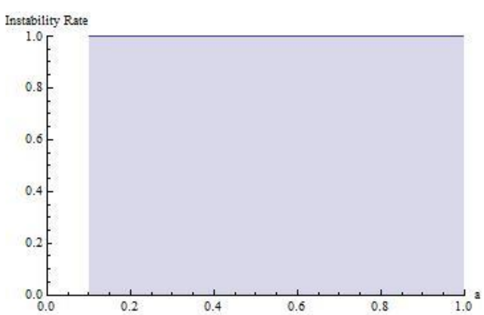


Figure 11. Instability rate of complete networks for varying *a* values, plotted in 2D. *n* = 1000, *C* = 2.

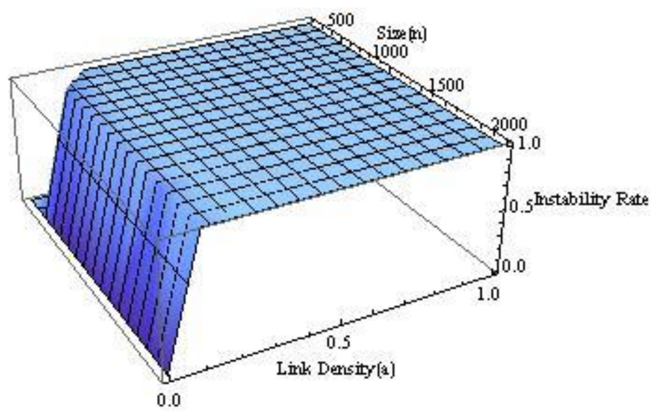


Figure 12. Instability rate of complete networks for varying *a* and *n* values, plotted in 3D. *C* = 2.

Erdos-Renyi Random Graphs (ER)

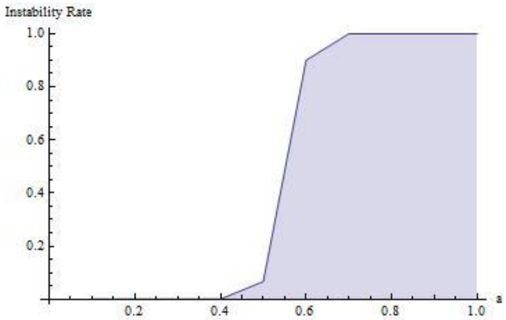


Figure 13. Instability rate of Erdos-Renyi random networks for varying *a* values, plotted in 2D. n = 1000, C = 2.

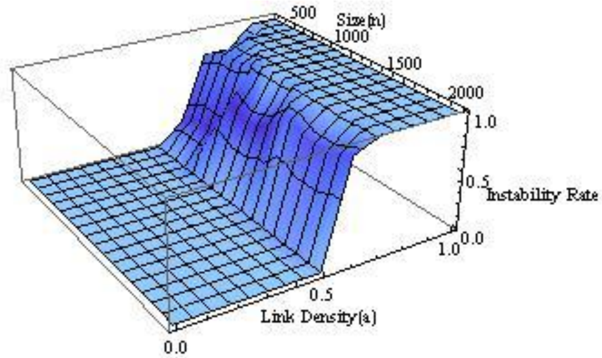


Figure 14. Instability rate of Erdos-Renyi random networks for varying *a* and *n* values, plotted in 3D. C = 2.

4.2 *Main Experiment*

Figure 15 shows the robustness performance (critical τ value) of each model over a range of *a* values from 0.1 to 0.5.

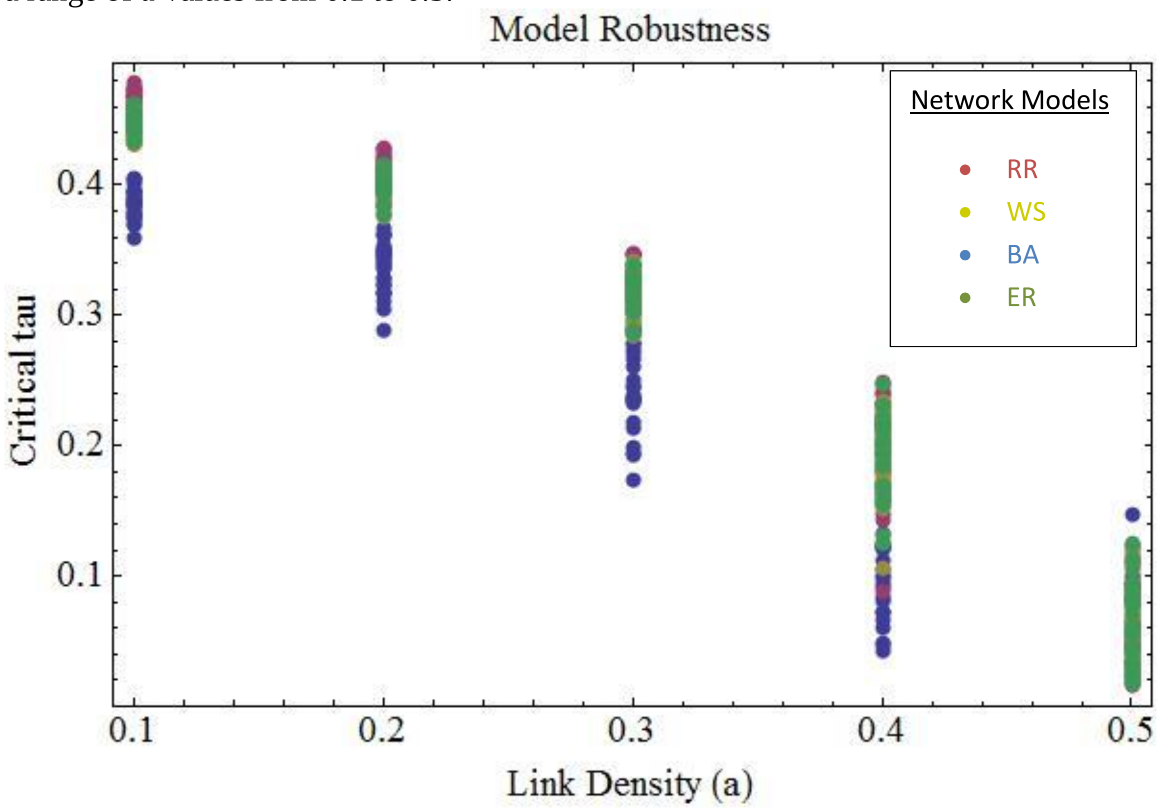


Figure 15. Robustness performance of RR, WS, BA and ER graphs for varying *a* values, plotted in 2D. n = 1000, C = 2.

In Figure 16, a slight horizontal shift was used to illustrate the models' performance more clearly.

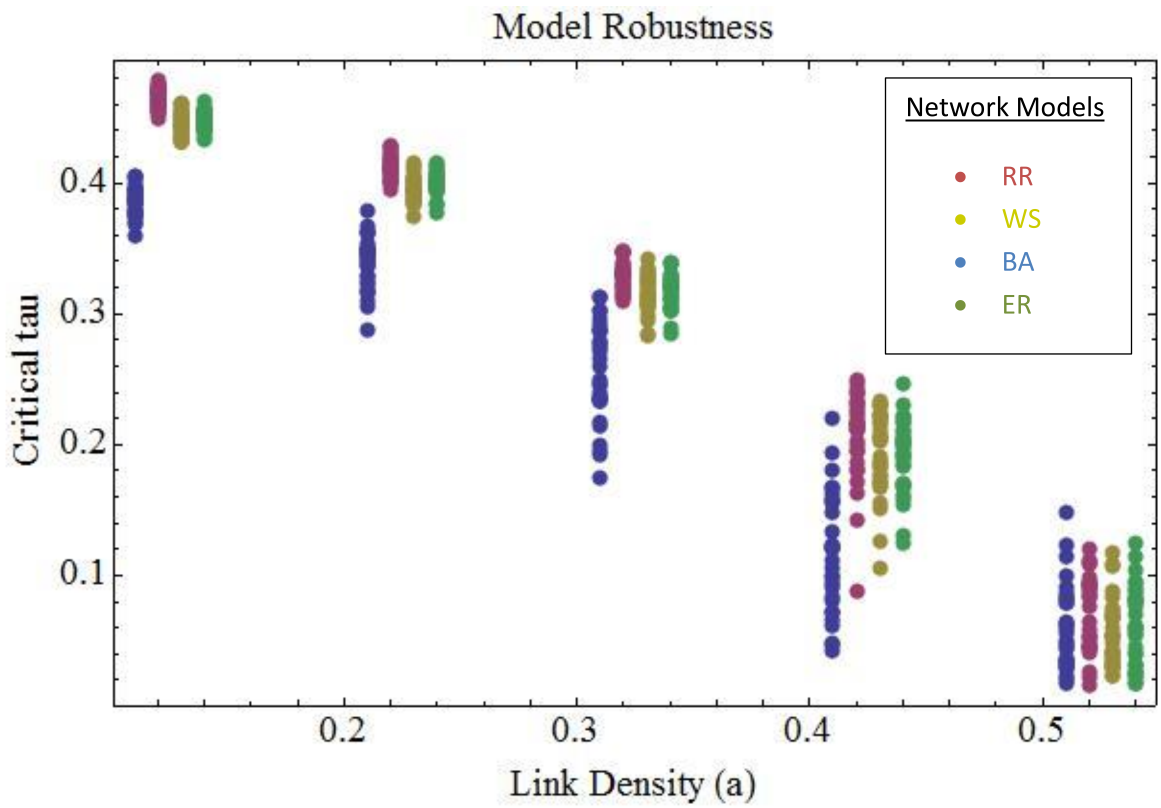


Figure 16. Shifted robustness performance of RR, WS, BA and ER graphs for varying *a* values, plotted in 2D. n = 1000, C = 2.

Figure 17 shows the curves that were fit to each network model along with their corresponding R^2 values.

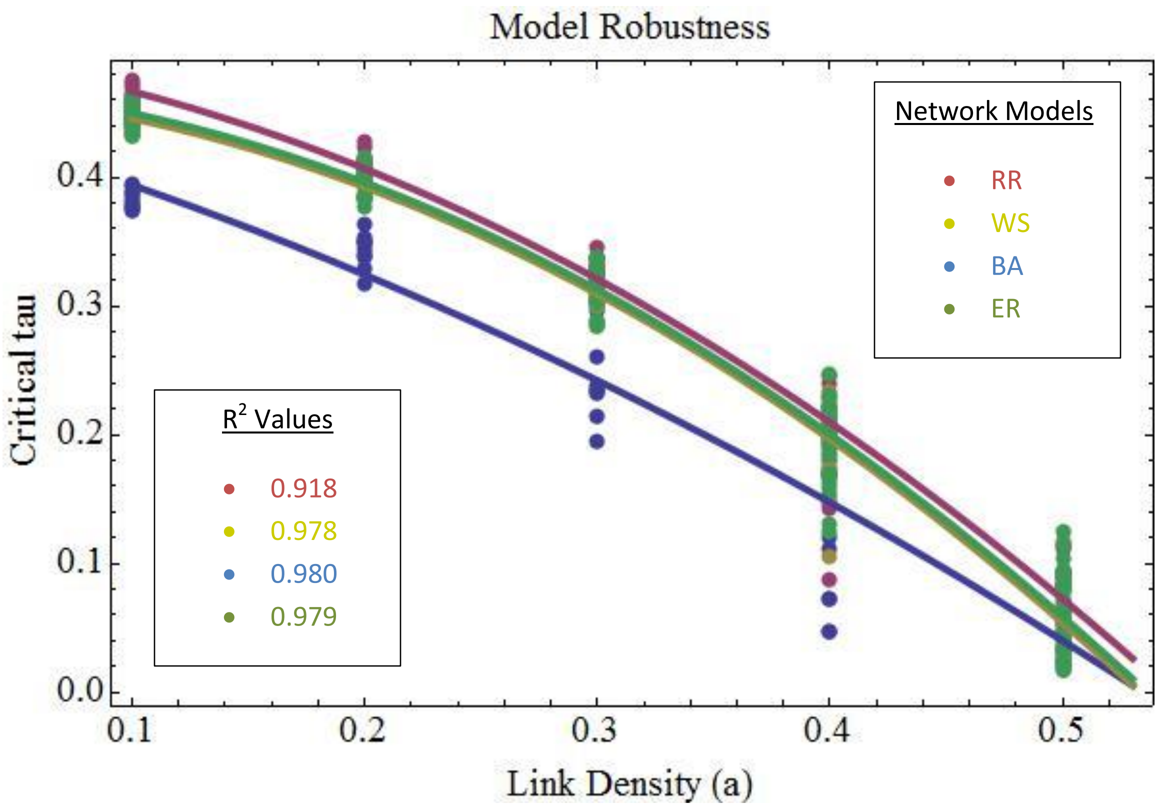


Figure 17. Curve fitting of robustness performance of RR, WS, BA and ER graphs for varying *a* values, plotted in 2D. n = 1000, C = 2.

Figure 18 shows the variances observed in critical τ value for each model, across the range of *a*.

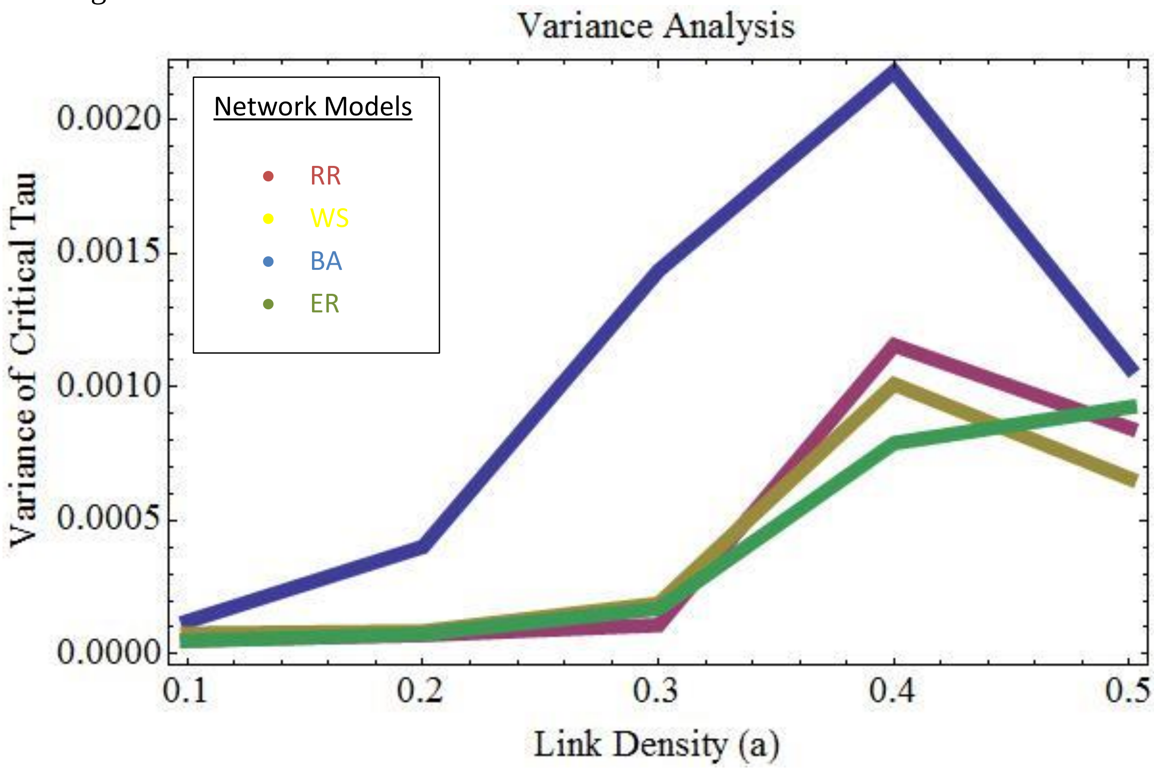


Figure 18. Variance of robustness performance of RR, WS, BA and ER graphs for varying *a* values, plotted in 2D. n = 1000, C = 2.

Figures 19-28 show results of ANOVA and Tukey post-test values for the network models' robustness performance, measured by critical τ, at each value of a.

	DF	Sum of Sq.	Mean Sq.	F-ratio	P-value
Model	3	0.106	0.035	468.007	$1.29*10^{-64}$
Error	116	0.009	0.00007	--	--
Total	119	0.115	--	--	--

Figure 19. ANOVA, a = 0.1

Models → {{1,2}, {1,3}, {2,3}, {1,4}, {2,4}}

Figure 20. Tukey Post-Test, a = 0.1

For figure 20, a Tukey post-test was performed generating sets of models shown above. These sets of models are those whose means have be found to be statistically different from each other. This indicates that there is significant difference between: BA and RR, BA and WS, RR and WS, BA and ER and RR and ER. The exception found at this value of *a* is between the WS and ER models.

	DF	Sum of Sq.	Mean Sq.	F-ratio	P-value
Model	3	0.096	0.032	201.988	$6.987*10^{-46}$
Error	116	0.018	.0001	--	--
Total	119	0.114	--	--	--

Figure 21. ANOVA, a = 0.2

Models → {{1,2}, {1,3}, {2,3}, {1,4}, {2,4}}

Figure 22. Tukey Post-Test, a = 0.2

For Figure 22, a Tukey post-test was performed generating sets of models shown above. These sets of models are those whose means have be found to be statistically different from each other. The results for this *a* value are identical to prior *a*, as significant differences are found between: BA and RR, BA and WS, RR and WS, BA and RR and ER, (etc). The exception found is once again between the WS and ER models.

	DF	Sum of Sq.	Mean Sq.	F-ratio	P-value
Model	3	0.099	0.033	69.745	$7.49*10^{-26}$
Error	116	0.095	0.0004	--	--
Total	119	0.154	--	--	--

Figure 23. ANOVA, a = 0.3

Models → {{1,2}, {1,3}, {1,4}}

Figure 24. Tukey Post-Test, a = 0.3

For Figure 24, a Tukey post-test was performed generating sets of models shown above. These sets of models are those whose means have be found to be statistically different from each other. This indicates that there is significant difference between just three pairs of models: BA and RR, BA and WS, and BA and ER.

	DF	Sum of Sq.	Mean Sq.	F-ratio	P-value
Model	3	0.149	0.0496	38.617	$2.22*10^{-17}$
Error	116	0.149	0.001	--	--
Total	119	0.298	--	--	--

Figure 25. ANOVA, a = 0.4

Models → {1,2}, {1,3}, {1,4}

Figure 26. Tukey Post-Test, a = 0.4

For Figure 26, a Tukey post-test was performed generating sets of models shown above. These sets of models are those whose means have be found to be statistically different from each other. The results for this *a* value are identical to prior *a*, as significant differences are found between just three pairs of models: BA and RR, BA and WS, and BA and ER.

	DF	Sum of Sq.	Mean Sq.	F-ratio	P-value
Model		0.004	0.0001	1.38	0.252
Error		0.102	0.0008	--	--
Total		0.105	--	--	--

Figure 27. ANOVA, a = 0.5

Models → {none}

Figure 28. Tukey Post-Test, a = 0.5

For Figure 28, a Tukey post-test was performed generating sets of models shown above. These sets of models are those whose means have be found to be statistically different from each other. At this largest *a* value there are no longer any statistical differences between any pairs of models, indicating that these models become more uniform as well as less robust as *a* increases.

Chapter 5: Conclusions

5.1 *Main Findings*

There are several findings which seem clearly inferable from these results. The first is the sharp transition to instability found for four of the five network models in the middle *a* range, around 0.4 to 0.6, and that this transition remains virtually unaffected over a range of network sizes *n*. This does not hold, however, for the complete graph CG model, since it produced consistently unstable networks across all *a* values. This is the reason that the CG model is not included in the final results, because its unstable nature would have disallowed it from completing the m*ain experiment* within any reasonable computation time. The BA scale-free graph also is distinguished, for its greater fragility in comparison to the WS, ER and RR models. This is evident, as the instability rate for BA rises more gradually than the others and beginning earlier in the *a* range, as shown in Figures 5 and 6.

From the *main experiment* results for the remaining four models, there appears to be a monotonic decrease for all network models as link strength *a* increases. Variance of critical τ values also grows until a = 0.4, spiking from 0.2 to 0.3 for the BA model and from 0.3 to 0.4 for ER, RR and WS. This increasing variance creates the dissolved looking clusters most visible at a = 0.4 in Figure 16. There are statistically significant differences in robustness performance shown between the models. The ANOVA analysis shows that the statistical difference between the models diminishes as *a* increases. This is evident as the number of models which are found different from each other decreases from 5 out of 6 possible pairs at a = 0.1 and 0.2, to 3 out of 6 at a = 0.3 and 0.4, to 0 at a = 0.5. There also were statistically significant trends in robustness performance between network models. The RR model has the highest critical τ values overall and is therefore most robust, followed closely by the WS and ER models, which behave nearly identically throughout. Along with showing the greatest fragility during the *preliminary experiment*, the BA model is also found least robust and most variant in the *Main experiment*, distinguished from the others by a much wider gap in both respects.

These large variances in robustness performance, observed within network models and within a values, shows that performance is being significantly influenced by network traits which are not being controlled in this model. The structure of links in these networks is subject to change with the inherent randomness. This is most pronounced within the BA model, as shown in Figure 18, indicating that network link structure matters for robustness performance.

5.2 *Future Directions*

This wide variance in performance observed within the current parameters presents an opportunity for future work, to investigate what unknown network connectivity properties are most associated with robustness performance. This would involve tracking robustness across ranges of other network parameters, such as connectivity and clustering coefficient. This knowledge could potentially be used to enhance network design by optimizing for robustness to global perturbations. If it were observed, for example, that BA networks with smaller clustering coefficients (fewer, more powerful hubs) were less robust, that may have implications for designed systems following the BA model (such as the airport network in the United States). This finding could imply that the airport network is maximally stable to global perturbations within a certain range of clustering coefficient, and could account for this in its design process. The broader question being introduced is: What specific, unknown structural differences are causing the variance in performance observed between and within network models?

The connection between this abstract model and the real world systems it was inspired by could also be strengthened by applying it to networks which are more qualitatively and quantitatively similar to real world networks. This could mean testing networks which are topologically dynamic and interdependent with nonlinear relationships between nodes. Furthermore, linear stability analysis using eigenvalues is just one means of measuring network stability. This criterion could be expanded to include other forms of stability tests.

Finally, it may be practical to implement some form of evolutionary computation to scan the vast search space of possible networks and evolve maximally robust networks of each model, tracking fitness across numerous network parameters along with the global perturbation robustness (critical τ) measurement.

5.3 *Limitations*

It is difficult to determine from these results, how exactly scale-free, small-world, random and random regular graphs should be designed or otherwise treated differently in real world contexts to maintain maximal stability. Given the abstract nature of these networks and of the stability testing itself, it can only be determined at this point that networks of any model are made more fragile and variant in their response to global perturbations with the greater their link density *a*, and that there is a clear relationship in critical stability thresholds to this perturbation between the four models, with: RR as the most robust, followed by WS and ER close behind, and then BA most fragile by a wider margin.

This modeling approach has many pros, such as its speed, simplicity and flexibility, though its simple and abstract mathematical nature saddles it with limitations – those of its traits which do not match up with real world domain contexts. The proposed model contains many such limitations, which will be approximated with the following list:

1) The nature of linear stability analysis defines the relationships between each node to be linear. This is of course not the case in many real world systems.
2) As in May's model the magnitude of the links' densities is uniformly scaled by *a*. This ensures that all the links' densities are on the same order of magnitude on average, which often is not the case in real world contexts.
3) As in May's model each node of each network is set to return to stability by default if left unaffected. This is a trait held by neither financial firms nor ecosystem species, as these organisms would not survive without ecological interaction.

4) As in May's model each link is equally likely to be positive or negative, another simplification not observed in real world contexts.
5) The global perturbation inflicted to destabilize the *A* matrices only comes in the form of addition. In real world contexts global perturbations do not only have additive effects, but rather can take numerous and unknown forms.
6) The global perturbation affects only the previously existing links of each network, meaning that no new connections are formed from these perturbations. This is not the case within either ecology or finance, as environmental changes often create new links among firms and species.
7) All stability tests performed within the model are done on static networks which have been generated by Python's NetworkX library, whereas the real world systems whose stability is crucial for societal function (whether financial, ecological, infrastructural or any other) are obviously neither static nor computationally generated.

There are surely many more limitations inherent to this modeling approach. The hopeful assumption in carrying out this project was that the results it yields may still have some practical relevance to the real world systems it was inspired by.

Works Cited

Allesina, Stefano, and Si Tang. "Stability criteria for complex ecosystems." *Nature* 483 (2012):6 205-208. Print.

Squartini, Tiziano, Iman Van Lelyveld, and Diego Garlaschelli. "Early-warning signals of topological collapse in interbank networks." *Scientific Reports* 3 (2013): 1-8. Print.

May, Robert "Will a Large Complex System be Stable?." *Nature* 238.5364 (1972): 413-414. Print.

Committee on Network Science for Future Army Applications (2006). *Network Science*. National Research Council.

Strogatz, Steven H.. "Exploring complex networks." *Nature* 410.6825 (2001): 268-276. Print.

Dodds, Peter. "More Introductioneering to Complex Networks." Complex Networks. University of Vermont. College of Engineering and Mathematical Sciences, Burlington. 16 Jan. 2014. Class lecture.

Biggs, N., E. Lloyd, and R. Wilson. "250 Years of graph theory 1736–1986." *Journal of Graph Theory* 10.3 (1986): iv-vi. Print.

Costa, L. Da F., F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. "Characterization of complex networks: A survey of measurements." *Advances In Physics* 56.1 (2007): 167-242. Print.

Albert, Reka, Jeong Hawoong, and Albert-Laszlo Barabasi. "Error and attack tolerance of complex networks." *Nature* 406 (2000): 378-381. Print.

Barabasi, A.-L.. "Scale-Free Networks: A Decade and Beyond." *Science* 325.5939 (2009): 412-413. Print.

Ellens, W., and R. E. Kooij. "Graph measures and network robustness." *arXiv preprint arXiv:1311.5064* (2013).

May, Robert M., Simon A. Levin, and George Sugihara. "Complex systems: Ecology for bankers." *Nature* 451.7181 (2008): 893-895. Print.

Hagberg, Aric, Pieter Swart, and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. No. LA-UR-08-05495; LA-UR-08-5495. Los Alamos National Laboratory (LANL), 2008.