5-2015

A Model of Activity and Intervention Across Social Networks

Allison Marie Heming
University of Tennessee - Knoxville, aheming@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Allison Marie Heming entitled "A Model of Activity and Intervention Across Social Networks." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Charles R. Collins, Major Professor

We have read this thesis and recommend its acceptance:

Judy D. Day, Ohannes Karakashian

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Model of Activity and Intervention Across Social Networks

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Allison Marie Heming
May 2015

To my family and the friends who have become family.

Acknowledgments

To my parents and brother who have kept me sane, supported me, believed in me, and listened to me complain for countless hours about a dream they would never let me give up on. I would not have been able to make this happen if it was not for your unrelenting support, even when you had absolutely no idea what I was talking about. I would not be where I am today if you did not encourage me to always follow my dreams. Thank you.

To my support system in northeast Ohio, you all have inspired me and provided me guidance. I cannot thank you enough for helping me find myself and my dream. Stefanie Arnold, you are remembered, loved, and missed dearly.

To the fellow graduate students who have studied countless hours with me, cried with me, laughed with me, supported me, understood me, and became my family. You all mean the world to me and have made this journey so much more than I expected. Amanda Diegel, Brittany Stephenson, Jacob Khale, Jordan McCarter, Josh Lipsmeyer, Joe Dawes, Kelly Rooker, Khoa Dinh, Leah McConoughey, Natalie Green, Paisleigh Kelley, and Rachel Hanson, thank you.

Last, but certainly not least, so my support system here in Knoxville: Donna Braquet, Vice Chancellor Rickey Hall, the Student Success Center staff, everyone at the OUTreach Center and Chancellor's Commission for LGBT People, you all are an inspiration. To my advisor Dr. Collins, who helped push me through and provided guidance along the way, and Pam Armentrout who is one of the most amazing women on the planet. Pam, this journey would not have been what it was for me without your kind words, positivity, and generosity. Thank you.

"Mathematics may not teach us how to add love or how to minus hate, but it gives us every reason to hope that every problem has a solution"

Abstract

Social network analysis is a growing field used to measure connectivity and activity of people and communities. We develop a model that creates a network and measures the overall activity of that network. We then apply interventions to this model and measure the change in overall activity. Through an optimization process we are able to determine the best course of action that minimizes or maximizes the overall activity of the network.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Social network analysis is an up and coming field of study. Being able to quantify and model real-world interaction in various forms has proved to be an incredibly important tool in many ways; seeing professional, social network, gives us the ability to see how elements are connected, trends form, interactions change, and so much more. Modeling social networks provides us the opportunity to simplify existing complex networks and provides a mathematical framework in which we can examine properties of the network that might be useful or of interest. Through producing a model of a social network, we are able to predict various properties and outcomes. This thesis will discus some of these various models and apply them in an effort to measure the spread of information and activity throughout 15 weeks on a college campus of about 25,000-30,000 students, faculty, and staff members. We begin with a general overview of graph theory and terms that are not only used throughout the paper but add to the general understanding. We use standard graph theory terminology that is both common and intuitive to promote easy reading and understanding. Then we progress into discussing several different models of social networks and how we applied these models to our network. Lastly, we will discuss our actual models and results as well as potential future work and applications.

1.1 Preliminary Definitions and Notions

We begin by reviewing several standard mathematical definitions. We let $\mathbb{N}$ denote the set of natural numbers, including zero. $\mathbb{Z}$ denotes the set of integers. Logarithms written as 'log' are taken at base 2, where as the natural logarithm is denoted as 'ln'. A set $\mathcal{A} = \{A_1, A_2, ..., A_k\}$ of disjoint subsets of a set A is a *partition* of A if $A = \bigcup_{i=1}^{k} A_i$ and $A_i \neq \emptyset$ for every i. Another partition $\{A_1', ..., A_l'\}$ of A *refines* the partition $\mathcal{A}$ if each A_i' is contained in some A_j.When A is a subset we let $[A]^k$ denote the set of all k-element subsets of A.

The definitions used are based on R. Diestel's *Graduate Texts in Mathematics: Graph*

Theory [2]. A *graph* is a pair $G = (V, E)$ of sets satisfying $E \subseteq [V]^2$, meaning that the elements of E are 2-element subsets of V. To assist with notation we should assume that $V \cap E = \emptyset$. The elements of set V are the *nodes* (or *vertices* or *points*)of the graph G with the elements of set E are the *edges* (or *lines*). The standard way to visualize a graph is drawing a dot for each node and joining the dots by lines which serve as edges. The way in which the graph is drawn, such as which nodes are connected and in what manner, are considered irrelevant, what is relevant is which pairs of nodes form edges and which do (see e.g. Figure 1.1).

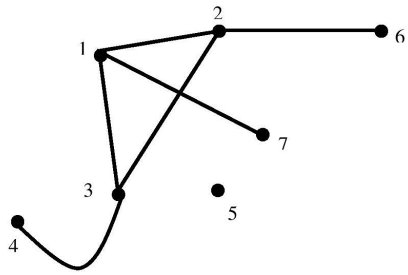

Figure 1.1: The graph on V={1,. . . , 7} with an edge set of E={{1,2},{1,3},{1,7},{2,3},{2,6},{3,4}}

A graph with the vertex set V is said to be a graph *on* V. The vertex set of graph G is commonly denoted as $V(G)$, with its edge set denoted as $E(G)$. It is important to note that these conventions are independent of the names of the nodes and edges sets in general; that is the node set F of graph $W = (F, B)$ is still referred to as $V(W)$, not as $F(W)$. The distinction between a graph and its node or edge set is not always made, thus we might refer to a node $v \in G$ instead of $v \in V(G)$. An edge, e, can be denoted similarly, $e \in G$.

The *order* of a graph G is number of nodes that G has and is written as $|G|$ with the number of edges denoted as $||G||$. Graphs can be finite or infinite depending upon their order. All graphs in this paper will be finite unless otherwise stated. Graphs can also be empty, in which we denote that as $\emptyset$ and call it an *empty graph*. A graph of order 0 or 1 is called a *trivial graph*.

A node is *incident* with an edge e if $v \in e$, meaning that e is an edge at v. Two nodes incident with an edge are its *endvertices* or *ends*, and an edge *joins* it ends. An edge x, y is typically denoted as xy (or yx). We have an $X - Y$*edge* if $x \in X$ and $y \in Y$. $E(X, Y)$ denotes that set E which is the set of all $X - Y$ edges. The set of all edges in E at a node v is denoted by $E(v)$.

Two nodes $1, 2$ of G in Figure 1.1 are *adjacent* (or *neighbors*) if 12 is an edge of G. The two edges $e \neq h$ are *adjacent* if they have an end in common. Nodes or edges that are pairwise non-adjacent are called *independent.* A set of nodes or edges are said to be *independent* (or *stable*) if no two of its elements are adjacent. A graph, G, is said to be *complete* if all nodes of G are pairwise adjacent, thus if a graph has no independent elements it is complete. A complete graph on n nodes is a K^n, see e.g. Figure 1.2.

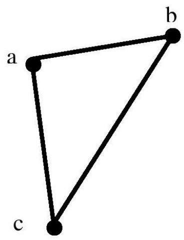

Figure 1.2: A K^3 graph: a complete graph on 3 nodes

Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. We let $G \cup G' = (V \cup V', E \cup E')$ and $G \cap G' = (V \cap V', E \cap E')$. G and G' are *disjoint* if $G \cap G' = \emptyset$. If $V' \subseteq V$ and $E' \subseteq E$, then G' is a *subgraph* of G. If G' is a subgraph of G that makes G a *supergraph* of G', denoted as $G' \subseteq G$. Alternatively we say that G contains G'. If $G' \subseteq G$ and G' contains all the edges $xy \in E$ with $x, y \in V'$, then we say that G' is an *induced subgraph* of G. Therefore, V' *induces* (or *spans*) G' in G, denoted $G' = G[V']$. We say that $G' \subseteq G$ is a *spanning* subgraph of G if V' spans all of G ($V' = V$).

We will now discuss, in more detail, the degree of a node. Let $G = (V, E)$ be a non-empty graph. The set of neighbors of a node, v in G is denoted by $N_G(v)$, more concisely $N(v)$. The *degree*, $d_G(v) = d(v)$, is defined as the number $|E(v)|$ of edges at v. This definition means that the degree is equal to the number of neighbors of v for the standard definition we have of a graph. This notion does not hold true for multigraphs, which we discuss later on. A node that has degree 0 is *isolated.* We denote the *minimum* *degree* of a graph G as $\delta(G) = min\{d(v)|v \in V\}$ and the *maximum degree* of a graph G as $\Delta(G) = max\{d(v)|v \in V\}$. If all of the nodes of the graph have the same degree , k, then the graph is *k-regular* (or simply *regular*). We can use the following equation to obtain the *average degree* of G:

$$d(G) = \frac{1}{|V|} \sum_{v \in V} d(v). \tag{1.1}$$

We now turn our attention to notions of paths, cycles and connectivity. A *path* is a non-empty graph that is of the form:

$$V = \{x_0, x_1, ..., x_k\} \qquad E = \{x_0x_1, x_1x_2, ..., x_{k-1}x_k\},$$

where the x_i's are all distinct. The nodes x_0 and x_k are linked by P and are called its *ends* with the nodes $x_1, x_2, ..., x_{k-1}$ are the *inner* nodes of P. The number of edges of a path is the path's *length*, and the path of length k is denoted by P^k, see e.g. Figure 1.3. Since k can be 0 we notice that when this happens we have $P^0 = K^1$.

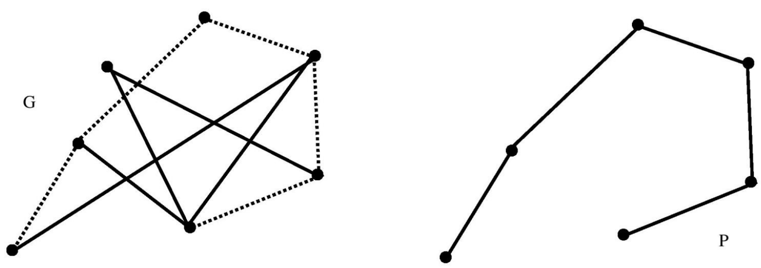

Figure 1.3: A path $P = P^5$ in G

Given sets A, B of nodes, we call $P = x_0, ..., x_k$ an $A - B$ *path* if $V(P) \cap A = \{x_0\}$ and $V(P) \cap B = \{x_k\}$. Two or more paths are *independent* if none of them contain an inner node of another, thus two $a - b$ paths are independent if and only if a and b are their only common nodes. The *average shortest path* is the average number of steps along the shortest path for all possible pairs of nodes.

A graph $C = P + x_{k-1}x_0$ is a *cycle* if $P = x_0...x_{k-1}$, with $k \geq 3$, is a path. The *length* of a cycle is determined by the number of edges. The cycle of length k is called a $k-cycle$ and is denoted by C^k. The minimum length of a cycle in a graph G is the *girth* and denoted $g(G)$. The cycle of maximum length in graph G is the *circumference.* If the graph does not contain a cycle then the girth is ∞ and the circumference is 0. An *induced cycle* in G is a cycle in G forming an induced subgraph, thus an induced cycle is one that contains no chords. A *chord* is an edge that joins two nodes of a cycle but is not an edge of the cycle, see e.g. Figure 1.4. The *distance*, $d_G(x, y)$, in G of two nodes x, y is the length of the shortest $x - y$ path in G. If a shortest path does not exist in G then we say that $d_G(x, y) = \infty$. A node is *central* in G if its greatest distance from any other node is as small as possible. This distance is the *radius* of G which is denoted by $\text{rad}(G)$.

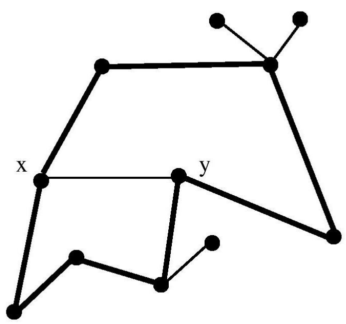

Figure 1.4: A cycle C^8 with chord xy, and induced cycles C^4, C^4

Consider a non-empty graph $G = (V, E)$, we say that G is *connected* if any two of its nodes are linked by a path in G. If $U \subseteq V(G)$ and $G[U]$ is connected, then we call U connected in G. A maximal connected subgraph of G is called a *component* of G, as seen in Figure 1.5.

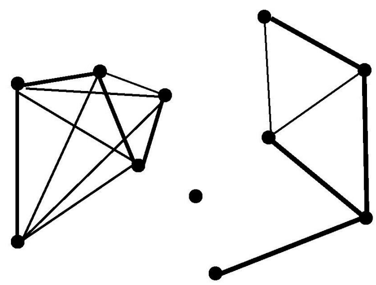

Figure 1.5: A graph with three components, and a minimal spanning connected subgraph in each component

A node that separates two other nodes of the same component is a *cutvertex* and an edge separating its ends is a *bridge*, see e.g. Figure 1.6. We call G *k-connected* (for $k \in \mathbb{N}$) if $|G| > k$ and $G - X$ is connected for ever set $X \subseteq V$ and with $|X| < k$. Alternatively, no two nodes of G are separated by fewer than k other nodes. With this in mind, we see that every non-empty graph is 0-connected and the 1-connected graphs being the non-trivial connected graphs. The greatest integer k such that G is k-connected is the *connectivity* of G denoted $\kappa(G)$. We see that G is disconnected if and only if $\kappa(G) = 0$.

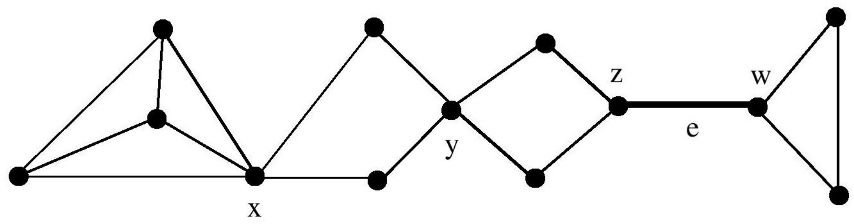

Figure 1.6: A graph with cutvertices w, x, y, z and bridge $e = zw$

We will now discuss *l-partite* graphs. Let $l \geq 2$ be an integer. A graph $G = (V, E)$ is called l-partite if V has a partition into l classes in which every edge has its ends in different classes, with nodes in the same partition class being non-adjacent. A l-partite graph is considered *complete* when every two nodes from different partition classes are adjacent. We call these complete l-partite graphs for all l together *complete multipartite* graphs and we denote them as $K_{n_1,\ldots,n_l}$; if $n_1 = \ldots = n_l := s$ which is abbreviated to K^l_s. This is intended to play off of earlier notation such that if we replace every node of a K^l by an independent s-set we have K^l_s, thus each partition class of a K^l_s graph contains exactly s nodes. Figure 1.7 below is an illustration of l-partite graphs as well as the notation discussed.

We conclude our introduction to graph theory with notions that begin to set up the types of graph being dealt with directly. A *directed graph* is a pair of (V, E) of disjoint sets, of nodes and edges, together with two maps init:$E \to V$ and ter: $E \to V$. These maps assign every edge e an *initial node* (init(e) and a *terminal node* (ter(e). An edge e is said to be directed from init(e) to ter(e). A directed graph can have more than one edge connecting two nodes, called *multiple edges*. If init(e) = ter(e), the edge e is

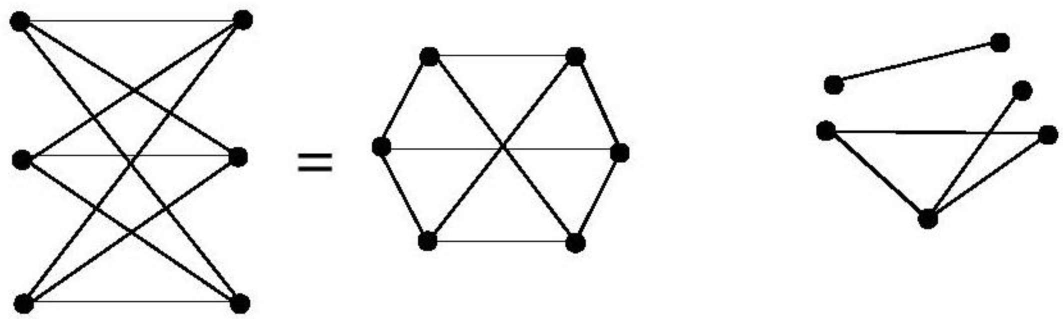

Figure 1.7: The two graphs on the left are $K_{3,3} = K_3^2$ and the graph on the right is a 3-partite graph

called a *loop*. A directed graph, denoted D, is an *orientation* of an undirected graph G if $V(D) = V(G)$ and $E(D) = E(G)$, and if init(e), ter(e)=x,y for every edge $e = xy$. In other words, oriented graphs are directed graphs without loops or multiple edges. A *multigraph* is a pair (V, E) of disjoint sets (nodes and edges) together with a map $E \to V \cup [V]^2$ assigning to every edge either one or two nodes, its ends. Multigraphs can have loops and multiple edges, thus multigraphs can be thought of as a directed graph whose directions have been 'forgotten'. We denoted edges, e, of multigraphs as $e = xy$ even though that edge is no longer unique.

Chapter 2

Various Network & Graph Structures

Now that we have a brief overview of general graph theory terms and notions we will discuss network and graph structures. There are quite a few competing theories as far as which model resembles real life social networks the best. We will only discuss random networks in generality, and mention briefly other network models that are used to model social networks and their short comings.

2.1 Erdös and Rényi Random Graph

We begin with the Erdös and Rényi random graphs. These two gentlemen laid the foundation for the theory of random graphs in their seminal papers [3][4][5].

Definition 1. *Let N be the number of nodes. An **Erdös Rényi simple random graph** is when nodes are connected by edges, such that each pair of nodes i, j has a connecting edge with independent probability, p.*

This random graph is the simplest model which has been well studied and researched, but does not work nicely for real-world networks. The biggest downfall of this simple model is that the degree distribution is not similar to the degree distribution for real-world networks. *Degree distribution* is the frequency count of the occurrence of each degree. This can be calculated for undirected graphs, as well as directed graphs. Directed graphs can be calculated as in-degree distribution and out-degree distribution.

In order to understand just why this simple random graph is a poor fit to model real-world networks we consider a node in a random graph. Using the definition of a random graph we know that this node connects with equal probability, p, to each of the other $N - 1$ nodes present in the graph, where N denotes to total number of nodes in the graph. Using this notion we see that the probability, p_k that it has degree exactly k

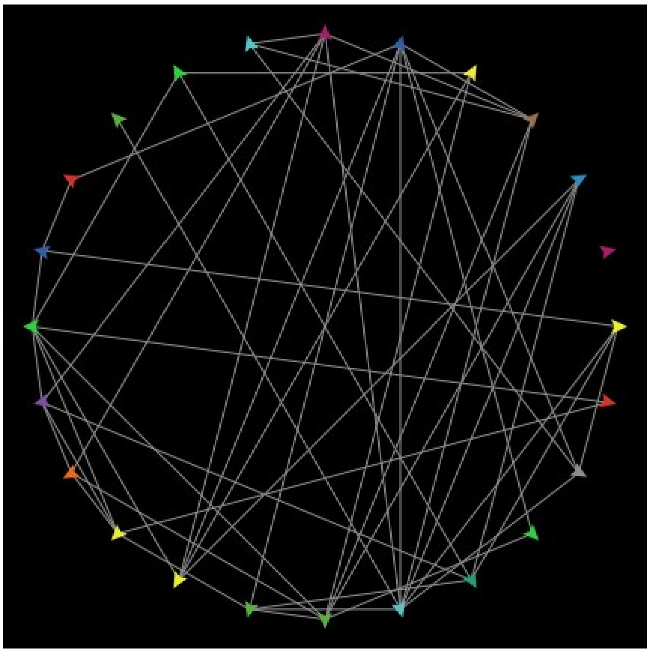

Figure 2.1: This is an Erdös and Rényi random graph with 24 nodes and $p = 0.15$

is given by a binomial distribution [6].

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.1}$$

The average degree of a node of a simple random graph is $z = (N-1)p$, which can also be written as the follows[6]:

$$p_k = \binom{N-1}{k} [\frac{z}{N-1-z}]^k [1 - \frac{z}{N-1}]^{N-1} \simeq \frac{z^k}{k!} e^{-z}. \tag{2.2}$$

Thus taking the limit of the approximate equality we obtain a Poisson distribution. Therefore if we have a large random network, that network would then have a Poisson degree distribution, which is not conducive to modeling real-world networks. One might ask given these important downfalls why should we even consider these simple random graphs? Well, many properties of a simple random graph can be calculated easily and exactly, contrary to other models.

The degree distribution of Eroös and Rényi random graph gives us extremely well connected nodes thus we see an emergence of a giant component within the network. College campuses tend to not be well connected networks. Rarely are students well connected to each other, in that every student knows every other student. While there might be a few highly connected people (nodes), the amount of people that each person knows is minimal in relation to the amount of people at the institution. Figure 2.2

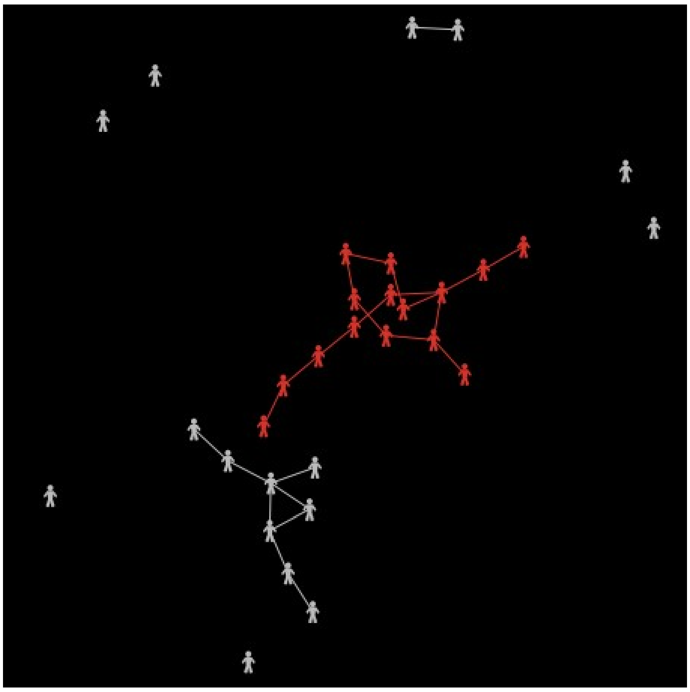

Figure 2.2: This is an Erdös and Rényi random graph with 31 nodes.

provides an example of what a giant component is. The connected people that are red in color are the giant component. The people that are white are those who are not connected to the giant component in any way. While the graph has 31 nodes (people), 15 of them are in the giant component. As time continues, more and more nodes connect, meaning that the giant component would eventually contain all 31 nodes and the degree of each node would continue to increase. This means that the average shortest path to get from one node to another would decrease. In other words, over time not only does the graph become completely connected but the path to get from one node to any other given node gets shorter in length.

2.2 Alternative Random Graph Structures

Since Erdós and Rényi published their work, the theory and development of networks analysis is growing to be able to constantly better model, various types of networks.

Several of these are static geographic, preferential attachment, and small world networks. These are not the only network models that have been developed, but these are the most current and relevant to the structure built for this thesis.

2.2.1 Static Geographic Network

The static geographical model is when each node connects to a specific number of its closest neighbors. This takes some of the randomness out of an ER random graph and attributes more to the geography of the nodes to each other.

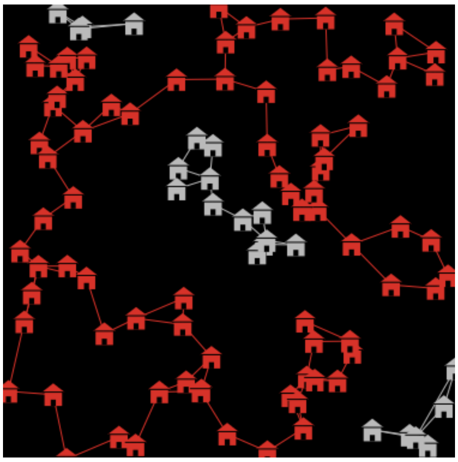

Figure 2.3: Static geographic network in which the number of closest neighbors is 2 and p=0.4. The red houses represent the longest path.

In a college setting a static geographic model could be relevant with the close proximity students, faculty, and staff have to each other. The only difference is that the university setting is not static. All attendees of a university tend to be moving quite a bit, different spaces for classes, eating in different locations, living in one location. This model presents an alternative possibility to a random simulation, but has its limitations as discussed.

2.2.2 Preferential Attachment

Preferential attachment occurs when new nodes prefer to attach to a well-connected node over less-well connected nodes. This network is examined in the Barabasi-Albert model

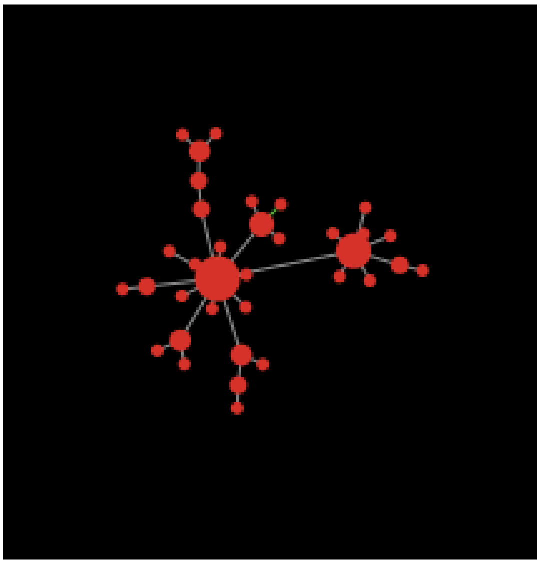

Figure 2.4: This is preferential attachment graph of 35 nodes. The bigger the node the higher connectivity.

which was first used to describe the skewed degree distribution of the World Wide Web. This model uses preferential attachment in that each node connects to other nodes with a probability that is proportional to their degree. In order for this process to work there would need to already be a subgraph in place in order to build the preferential attachments off of the existing graph.

2.2.3 Small World Networks

The idea of small world networks began with Milgram's experiment in the 1960's. This experiment determined that the average path length to get mail from Nebraska to a specific person in Boston was 6.5. This experiment was duplicated in 2003, with email by Dodds, Muhamad, and Watts. This experiment with 18 targets, 13 countries, over 600,000 people and 24, 163 message chains provided an average shortest path of 4.

Small world networks are a way to depict a big network that has an interesting local structure. A good example of a small world network is if one breaks the entire world into a network. In a small world network each node is wired in which all local nodes are connected. The nodes are then rewired according to some rule that is specified. The rule can range from optimizing for a particular property or adding edges with some probability. Watts and Strogatz were the first to make a simple model that reconciled

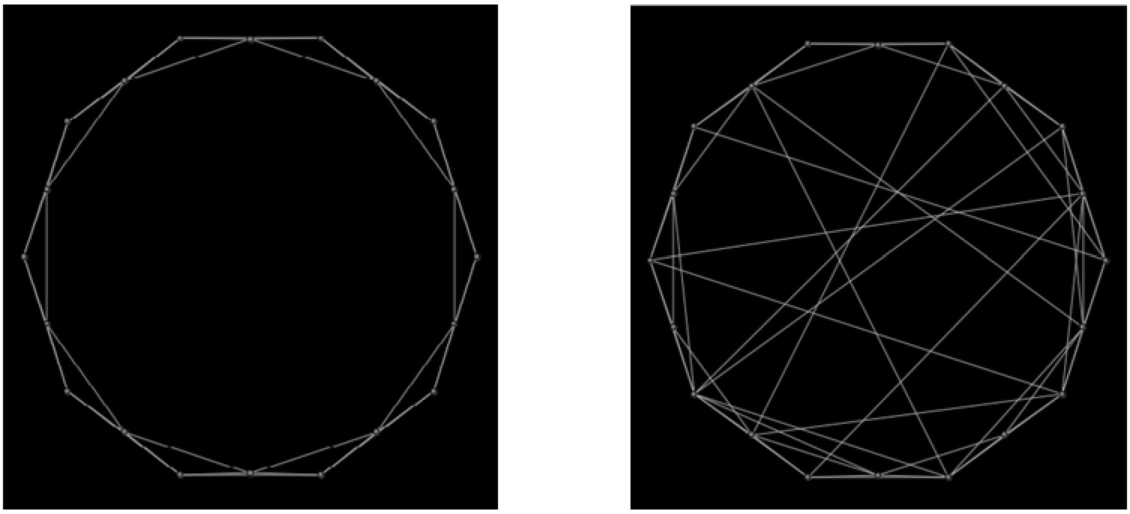

Figure 2.5: The left picture is a Watts & Strogatz small world network with 20 nodes. The right is a rewiring by adding more edges with p=.30 and ave. shortest path=2.21.

two observations, a high clustering and short average paths. Other mathematicians and scientist have improved on this model by incorporating geography and hierarchy social structure or evolving the model from different constraints. This helps improve the model for optimization and various other things.

Chapter 3

General Network Structure

This chapter discusses the general development of the network we are utilizing, how we came about structuring the network, what it means to be an intervention, and how we measure the effectiveness of these interventions. This thesis focuses on the discussion of practicality of modeling a social network on a college campus and the implications of practices of administration at various institutions will happen in a later section. This chapter is designed to give a general view of the network and measurements.

3.1 Development

Institutions typically offer a wide variety of student groups that span a wide range of interests; athletics, government, professional affiliations, major specific, special interest, Greek life, and many more. We broke these numerous groups into 5 main types with varying reaction to the activity of other groups, the outward response to other groups, long term activism, and connectivity and influence between groups; all with respect to how they would respond to a campus issue/policy change.

The level of reaction to another group's activity is described through the α variable. This is measured by a range of five α values, that span no reaction to high reaction to a particular group's activity. The outward response to a different group is taken in to account by the variable β. This is measured by a range of four β values that denote no public group activity and significant public group activity pertaining to an campus issue/policy change. The long term activity of a group is quantified by a δ variable which can be one of four quantities. It is important to note that this δ variable is a decay variable because it is subtracted from the overall activity. Thus the bigger the δ the lower the long term interest.

Group Type 1 is a group that does not have a high reaction to the activity of other groups, does not react publicly, and does not have a long term interest. This group type could be considered a group that centers around supporting an athletic program and their general interest in social issues that are happening on campus. If an issue does not directly impact the groups mission of supporting a specific athletic program, then that group would have no outward response (β), or in any of the responses of the groups they are connected to already (α). Group Type 2 has minimal outward response, minimal reaction to its connected groups, and low long term interest. This could be a political organization such as the College Republicans or College Democrats, who do not have a keen interest in the social issue, but have more of an interest in the election of leaders of the particular party. Thus the overall interest in social issues is low and the group is not particularly interested in what its connections are doing with the issue either, but does not completely ignore their connections like Type 1. Since this group type has low interest in general the long term interest is also low.

Group Type 3 has a high reaction to other group's activities, responds with high outward response, and is able to maintain this interest level throughout. This makes Type 3 similar to a Student Government Association. SGA has a high level of outward activity pertaining to events or changes impacting campus by nature of the organization, and thus they have a high reaction to other activity as well. SGA is able to maintain a high level of activity throughout a time span because of the number of people that the organization has and the general nature of the organization. Group Type 4 are groups that have a high reaction to other group activities with a low outward response levels and a high long term activity level. This makes Type 4 similar to a special interest groups similar to Sexual Empowerment and Awareness at Tennessee (SEAT), volOUT, or Progressive Student Alliance(PSA) who is really interested in specific topics and can maintain a level of visibility and passion. Group Type 5 has a moderate level of reaction to other groups expression of activity, a high outward response to a policy change and low levels of long term activity. Type 5 is to resemble Greek life organizations; social and service sororities and fraternities.

Once the different values were determined for α, β, and δ we needed to determine connections from group to group. These connections come in through the e_{ab} variable is the relative amount of influence and connection from group b on group a, and only has four value possibilities. The code for how the connections were determined can be found in the Appendix. It is important to note that this connections in this model are completely arbitrary and same group types have the potential to be connected, but are not automatically connected. In other words, no group types have a connection of 1. Similarly it is equally as important to note that not all groups are necessarily connected. The connections e_{ab} are put into a matrix, E, for general understanding and simplicity.

The overall activity of a groups is measured by equation 3.1, where t is the time variable (in weeks) and X is the measure of overall activity of some group. The notation X_a^{t+1} is the measure of activity of group a at time $t+1$, similarly α_a is the group's reaction to any expression of activity, β_b is a measure of how public group b is with their activity, and e_{ab} is the amount of influence/connection from group b and a.

$$X_a^{t+1} = X_a^t + \sum_a \alpha_a e_{ab} \beta_b X_b^t - \delta_a X_a^t \tag{3.1}$$

This equation is what is manipulated through the interventions in various ways which is discussed in the next section. We also see that X_a^t is a vector that is updated each week with the most current value over 16 weeks. This finite time situation creates a discrete model that allows us to examine the change in X_a over time and be able to characterize this in a graph similar to Figure 4.2.

3.2 Interventions Methods

Interventions are anything that changes the general behavior, positively or negatively, of the groups over time. This analysis considers five different interventions that interact with various aspects of equation 3.1. Interventions are only active in changing behavior for a specific time frame. This means that the variable that is impacted by the intervention would be changed as denoted for a specific period of time, and then go back to its normal state. An intervention is additive in nature when affecting a variable that is not the overall activity level of a group, while an intervention that is multiplicative in nature affects only the activity level directly, the X variable.

Interventions were determined based on some general patterns that happen on a college campus as far as different events that might illicit a response of a student organization. The variable in which was affected as determined to provide some variation in the network to examine the influence of each. Cost, time, and number of events are all constraints, as often times institutions are seeking to maximize outcome with minimal cost, time, and resources.

3.3 Effectiveness

Table 3.1: How effectiveness if measured

Scenario	How measured
Maximizing	$\sum_{a,t}[X_a^t]_+$
Minimizing	$\sum_{a,t}[X_a^t]_+$

The overall effect of the interventions is measured in various ways depending upon what scenario we are working with. For any scenario pertaining to minimizing or maximizing overall activity of each of the student groups we will use the sum of overall positive activity of all student groups (values below zero are set to zero). We will do our best to achieve a minimum or maximum while considering our cost, time, and number of events constraints. The values associated with each group type can be found in Table 4.3. We focus on optimizing the overall activity level, either by minimizing or maximizing. The algorithm evaluates each possible set of interventions and measures the sum of the overall activity level. If the sum is higher, considering the maximization problem, that becomes the new output that is compared to the activity level at the next set of interventions. The Matlab code for this process can be found in the Code section of the Appendix. The set of interventions that provides the highest sum of activity is the best set of interventions for that particular number of groups, group types, etc. We determine the effectiveness of the interventions by comparing the sums that were discussed in Table 3.1 through calculating the percent change. The percent change will essentially allow us to compare the two processes, no intervention and optimal intervention, in such a way that we can see how much overall activity has changed utilizing the optimal interventions.

Chapter 4

Specific Network Structure

Now that the general overview of the network and the structure has been discussed, we will discuss actual numerical values given to each of these variables for the specific model this thesis utilizes. For this model, we have chosen α to be the following five values $\alpha = [0.0, 0.025, 0.050, 0.075, 0.10]$, where zero is no reaction to activity by another group and 0.10 is a strong outward reaction to the activity of another group. The outward response of a group is specified to be $\beta = [0.0, 0.08333, 0.16667, 0.25]$, where zero is no outward response and 0.25 is a significant outward response. Our decay term δ consists of the following four values, $\delta = [0.64, 0.16, 0.04, 0.01]$. These variables come together to form five specific group types. These group types can be found in the Table 4.1.

Table 4.1: Group type values

	Variables		
Group Type	α	β	δ
1	0.0	0.0	0.64
2	0.025	0.08333	0.16
3	0.1	0.16667	0.01
4	0.075	0.08333	0.04
5	0.05	0.25	0.16

Now that group types are established the connections that build the network need to be determined. As discussed in the previous chapter, the variable is e_{ab} for connectivity and relative influence of group b on group a, $e_{ab} \in [0.0, 0.10, 0.30, 0.40]$. These values are important because groups do not have to be connected since 0 is a potential value for e_{ab} and groups are not automatically connected in that 1 is not a potential value for e_{ab}. This means that not all groups, or nodes, are directly connected to each other, providing some interesting dynamics. This could make one group essential to the amount of activity by having it be the only link from one set of groups to another by creating a cut node and a bridge as in Figure 1.6. This could also create a structure that we see in Figure 2.2 and Figure 2.3 where sections of the graph are not attached to the giant component, which would change the way the groups interact and ultimately the overall activity level of each group.

The process of connectivity in the network discussed in this paper is as follows. The e_{12}, the edge between a group of type 1 and type 2, e_{23}, e_{34}, and e_{45} are all 0.1. Groups of the same type (e.g. e_{11}, e_{22}, etc.) have an edge value of 0.3. The edge between groups of type 1 and groups of type 4, e_{14}, is 0.4. Similarly, $e_{25} = 0.4$. Any other combination of group types does not have an edge, e.g. $e_{13} = 0$. This process is outlined in the makenetwork.m piece of Matlab code.

Deciding how many of each group type there are as well as the initial start value of each group are decided by two different random number generators that are controlled by seeds. The initial start value of each group is decided by the seed i, initial, while the distribution of group types is controlled by the seed gtd, group type distribution. These seeds are reset before each run discussed below in help insure that the process can be duplicated and to create a control environment that can be utilized for comparing different runs that change other variables (e.g. cost limit). The Matlab code for these two random number generators and their seeds can be found in the makegroups.m file.

This model has 5 specific types of interventions that can be utilized to change the

Table 4.2: Intervention types

Intervention	1	2	3	4	5
Name	Email	Focus group	Visit	Conference/ event	Web post
Variable changed	X and δ	β	X	α	δ
Amount of change	$X*1.1$ and $\delta/0.2$	$\beta + 0.2$	$X * 1.4$	$\alpha + 0.3$	$\delta - 0.5$
Cost	$1	$100	$60	$300/$850	$1
Probability	0.1	0.6	0.4	0.3	0.2

behavior of a group in a particular way. Variables of equation 3.1 were changed in either a multiplicative or additive manner by interventions. You can find these explained in Table 4.2. If an intervention is applied, it remains for 3 weeks and then the variable that is changed from the intervention returns to its normal state. It is possible for an intervention to not be needed, meaning no intervention provides a higher activity level, compared to an intervention, for that specific period in time. It is also important to note, that there is a probability that each intervention is applied. This means that two groups of the same type might not necessarily respond the same to the same intervention. In fact the way it is setup in this model, one group who reacts differently does so by not respond to the intervention at all. The code pertaining to the types of interventions, cost, influence on variables, and time frame is discussed in the Matlab code file in the Appendix.

The optimization of this process occurs by ticking through each set of interventions.. We begin by recording the initial value with no interventions, [0,0,0,0,0]. This is then saved as our best value. If the algorithm encounters a set of interventions that provides a higher(or lower) value then that value is then saved as the best overall value and that set of interventions becomes the new optimal set of interventions. This process runs until it has successfully gone through all possible sets of interventions, all 5^5 of them.

4.1 Scenario 1

We begin by examining the activity level of 20 groups on campus, randomly generated, and a time span of 16 weeks. The connections between each group can be found in the

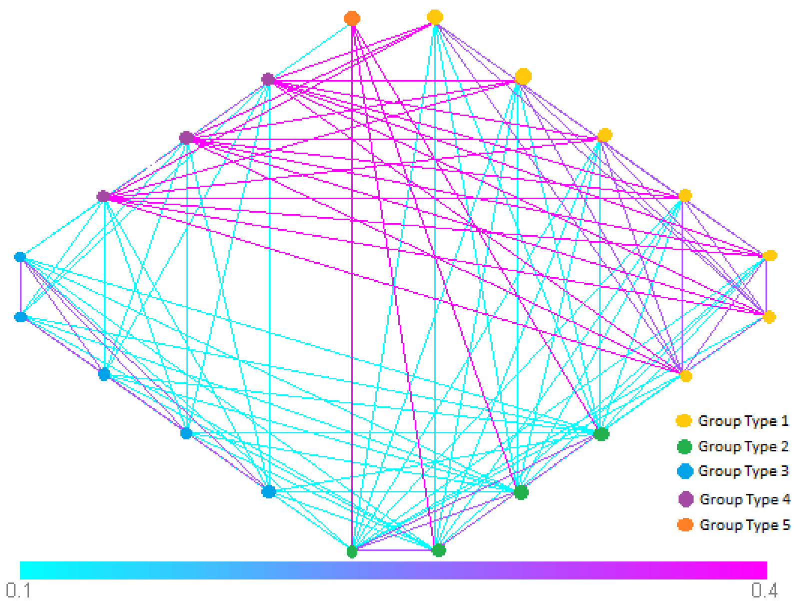

Figure 4.1: Graph created from the adjacency matrix, E

matrix E, which is displayed by the weighted adjacency matrix graph in Figure 4.1. This allows us to have a pictorial representation of our connectivity rules that were discussed above. We can see in Figure 4.1 that while not every node on our graph is connected, we have a highly connected graph. The break down of group types can been seen in Table

Table 4.3: Summary table for Scenario 1

Number of Groups	Group Type	Number of Each Type	Initial Value	Optimal Interventions	Percent Change
	1	7	85.2381	[4,2,3,0,4]	140.95 %
	2	4			
20	3	5			
	4	3			
	5	1			

4.3, with the overall activity level throughout a typical semester, without interventions, is 85.2381. Therefore, if an institution did not change anything pertaining to the groups or the way the campus climate is presently, then the overall activity of all groups would be 85.2381, with an average group activity of 4.26. The right part of Figure 4.2 is the graph of this overall activity as it occurs throughout the 16 weeks. In this graph we can clearly see the different group types, as they are depicted by similar lines with different initial values.

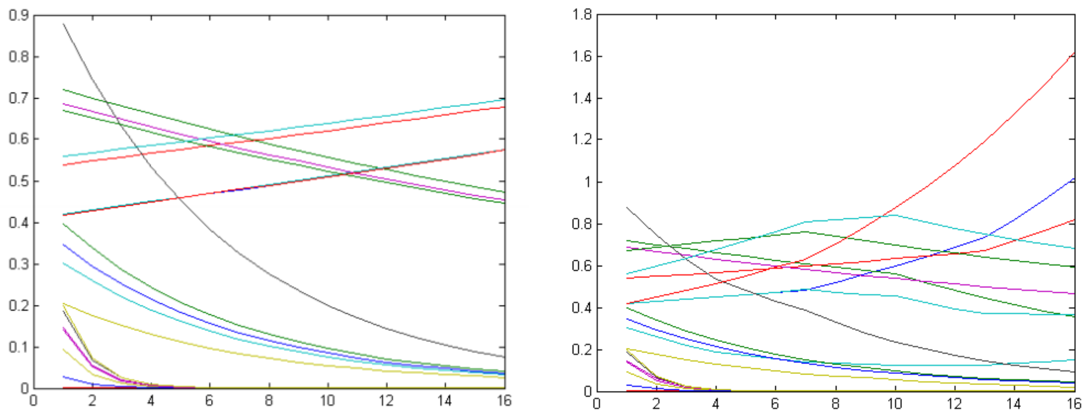

Figure 4.2: Graph of Scenario 1 without interventions(left) & with interventions (right)

We seek to maximize the overall activity level by implementing an intervention every three weeks while considering a cost limit of $10,000 and time of 16 weeks. The best set of interventions as found by this model is the intervention set of $[4, 2, 3, 0, 4]$. This means that given the parameters of cost and time the best set of interventions to maximize overall activity level is a conference or campus event at weeks 1 and 13, a focus group at week 4, visits to student groups at week 7, and nothing at week 10. This provides an overall activity output of 205.3810, with an average group activity of 10.269. We can see the changes that are produced by the interventions in the graph on the right in Figure 4.2.

Table 4.4: Summary of results for the group type distribution found in table 4.3

Random Seed Value	Initial Value	Optimal Interventions	Percent Change
i=1	85.2381	[4,2,3,0,4]	104.95 %
i=2	83.5887	[4,2,3,0,4]	157.38 %
i=3	68.7217	[4,2,3,0,4]	124.49 %
i=4	89.3464	[4,2,3,0,4]	148.71 %
i=5	67.8284	[4,2,3,0,4]	108.17 %

In Table 4.4 we examine the stability of the model when varying the initial values. We see that while the percent change fluctuates, the set of interventions remains the same. The optimal set of interventions, a campus event, focus group, group visit, and campus event, at least doubles overall student group activity on campus. This means that this group distribution that is in Table 4.3 does not fluctuate much with respect to changing the initial value. This could be because of the particular group distribution that is presented here, which is a topic for future research.

4.1.1 Variations of Group Type Distribution

Table 4.5: Summary of the distribution of group types for gtd=7

Number of Groups	Group Type	Number of Each Type
20	1	4
	2	4
	3	4
	4	3
	5	5

We now take some time to look at a more even distribution of group types, meaning the distribution is not heavily skewed to one group type or another, as displayed in Table 4.5. We continue to have 20 groups with the group type distribution that is generated when gtd=7 and i=1. We can also observe the results, initial values, percent change, and the optimizing set of interventions for every run in Table 4.6. This table shows that for this particular group type distribution the sensitivity to different initial values for the activity of each group comes into play. We see the commonalities are starting off the semester with a campus event or conference and following that up with focus groups of students and web posts.

Table 4.6: Summary of results for group type distributions of Table 4.5

Seed Value	Initial Value	Optimal Interventions	Percent Change
i=1	65.3489	[4,2,4,4,5]	107.86 %
i=2	78.3603	[4,2,4,4,5]	135.92 %
i=3	85.7172	[4,2,5,2,5]	130.67 %
i=4	105.4307	[4,2,5,2,5]	107.87 %
i=5	92.9353	[4,2,5,2,5]	124.59 %

We examine the sensitivity of the model with various group distributions and the same initial value seed in Table 4.7. We see that for different group type distributions we have a slightly different set of optimal interventions. These interventions all begin with hosting a campus event and then holding a focus group. This is interesting in that we see the group type distribution changes the initial starting value, even though the random seed is the same for each case. It is not surprising that this would occur, given Equation 3.1 and the interaction of all of the variables. Having different distributions of group types provide different alpha, beta, delta, and connectivity values. This would cause the initial value to fluctuate.

4.1.2 Cost Limit

We consider the same situation above, where the group type distribution is as displayed in Table 4.5. We change our cost limit to $1,000 as well as the cost for a campus event,

Table 4.7: Summary table for various group type distributions within Scenario 2

Number of Groups	Group Type	Number of Each Type	Initial Value	Optimal Interventions	Percent Change
20	1 2 3 4 5	7 4 5 3 1	85.2381	[4,2,3,0,4]	104.95 %
20	1 2 3 4 5	4 4 7 3 2	92.0355	[4,2,2,0,2]	244.96 %
20	1 2 3 4 5	4 4 7 3 2	92.5356	[4,2,2,0,2]	183.57 %
20	1 2 3 4 5	5 3 3 3 6	66.9838	[4,2,2,1,2]	131.07 %
20	1 2 3 4 5	3 5 5 4 3	62.4465	[4,2,3,3,5]	119.23 %

intervention 4, to be $850 instead of $300. We compare this to the $10,000 cost limit with the $850 cost of the campus event. The comparison of the results can be found in Table 4.8. It is evident that changing the cost limit impacts the optimal intervention strategy.

Table 4.8: Comparison of interventions and percent change depending on cost limit

	Cost limit $1,000		Cost limit $10,000	
Random Seed Values	Optimal Interventions	Percent Change	Optimal Interventions	Percent Change
i=1,gtd=7	[4,5,2,1,5]	79.66 %	[4,2,5,2,5]	120.65 %
i=2,gtd=7	[4,5,2,1,5]	89.83 %	[4,2,4,4,5]	140.74 %
i=3,gtd=7	[4,3,5,5,5]	83.16 %	[4,2,5,2,5]	149.37 %
i=4,gtd=7	[4,3,5,5,5]	89.09 %	[4,2,5,2,5]	121.40 %
i=5,gtd=7	[4,3,5,5,5]	79.14 %	[4,2,5,2,5]	149.82 %

Regardless of the cost limit, the optimal intervention strategies begin with having an all campus event pertaining to a specific change or topic. After this initial intervention at $t = 1$ it is strictly dependent upon the cost limit as the next course of action, which is as expected, however it is important to note that regardless of cost limit a focus group and multiple web post are in the follow-up after the initial event.

The impact of the initial value in what interventions are optimal and the combination of those interventions are evident. We see in Table 4.8 that for one group of initial values the best set of interventions is a campus event, web post, focus group, email, and web post, in that order. A different set of initial values provides an optimal intervention set of campus event, group visit, and 3 web posts as optimal. These are two distinct set of interventions that occur for the same set of group distributions but different initial values.

4.2 Scenario 2

Table 4.9: Summary table for Scenario 2

Number of Groups	Group Type	Number of Each Type	Initial Value	Optimal Interventions	Percent Change
50	1	14	281.8699	[4,4,2,2,4]	546.01 %
	2	9			
	3	8			
	4	9			
	5	10			

Scenario 2 consists of 50 groups, cost limit of \$10,000, time of 16 weeks, and the original initial activity values. The group distribution, provided by gtd=1, can be found in Table 4.9. We use these in Table 4.10 to examine the impact of the initial activity level of the same group distribution. We see in Table 4.10 that regardless of the randomly generated initial value, we obtain the same set of optimal interventions of a campus event/conference at weeks 1, 4 and 13, and a focus group at weeks 7 and 10. In Table 4.10 we see that that regardless of initial value generated, we obtain the same set of optimal interventions of a campus event/conference at weeks 1, 4 and 13, and a focus group at weeks 7 and 10. Given that our cost limit is such that cost is not a constraint, this

Table 4.10: Summary of Results for gtd=1

Random Seed Value	Initial Value	Optimal Interventions	Percent Change
i=1	281.8699	[4,4,2,2,4]	546.01 %
i=2	231.1008	[4,4,2,2,4]	476.46 %
i=3	230.9180	[4,4,2,2,4]	487.52 %
i=4	249.1189	[4,4,2,2,4]	462.78 %
i=5	216.9733	[4,4,2,2,4]	467.77 %

set of interventions is as expected. In general it is much easier to keep student activity up when the university itself is holding more campus activities, events, or conferences around a topic.

4.2.1 Variations of Group Type Distributions

We also notice that regardless of the group type distribution we have the same general set of optimal interventions in that multiple campus events and focus groups. While the percent change varies among the various group type distributions, there is an incredible increase in activity based on the interventions regardless. An interesting dynamic that is happening in Table 4.11 is the actual variation of percentage change. It is expected that with the same relatively similar distribution of group types and the exact same set of interventions, that the percentage change would be similar as well. We notice in Table 4.11 in row 2 a different variation of the same two interventions as all of the other group type distributions in Table 4.11.

4.2.2 Cost Limit

We again reduce the cost limit to $1,000 instead of $10,000 and utilize intervention 4, a campus conference or event, at the cost of $850 instead of $300. We can view the results in Table 4.12. We see that the optimal set of interventions remains constant for each cost limit set regardless of the random seed of the initial value of the group types. With the cost limit at $1,000 a campus event or conference can only happen one time with follow-ups happening in forms of visiting student groups, sending out emails, and making a web

Table 4.11: Summary table for various group type distributions within Scenario 2

Number of Groups	Group Type	Number of Each Type	Initial Value	Optimal Interventions	Percent Change
50	1 2 3 4 5	14 9 8 9 10	281.8699	[4,4,2,2,4]	546.01 %
50	1 2 3 4 5	9 11 18 7 5	445.3319	[2,4,2,2,4]	584.37 %
50	1 2 3 4 5	6 15 15 7 7	375.1804	[4,4,2,2,4]	818.78 %
50	1 2 3 4 5	11 6 11 13 9	312.1336	[4,4,2,2,4]	425.48 %
50	1 2 3 4 5	10 12 10 7 11	234.4767	[4,4,2,2,4]	710.36 %

post. This group of interventions substantially increases the overall activity by more than doubling it in every case considered. When the cost limit was $10,000, the optimal set of interventions contains the more expensive interventions in hosting campus events several times as well as several focus groups of students on campus. These interventions increased activity more than five times what the initial activity level happening on campus. While this second set of interventions discussed might be ideal in maintaining a high overall campus activity, it is not realistic to the higher education environment, nor is the set of interventions feasible.

Table 4.12: Comparison of interventions and percent change depending on cost limit

	Cost limit $1,000		Cost limit $10,000	
Random Seed Values	Optimal Interventions	Percent Change	Optimal Interventions	Percent Change
i=1,gtd=5	[4,5,3,5,3]	182.91 %	[4,4,2,2,4]	710.36 %
i=2,gtd=5	[4,5,3,5,3]	168.90 %	[4,4,2,2,4]	666.70 %
i=3,gtd=5	[4,5,3,5,3]	170.01 %	[4,4,2,2,4]	666.35 %
i=4,gtd=5	[4,5,3,5,3]	181.67 %	[4,4,2,2,4]	705.96 %
i=5,gtd=5	[4,5,3,5,3]	197.28 %	[4,4,2,2,4]	758.37 %

4.3 Scenario 3

Table 4.13: Summary of Minimization Results

Scenario	Initial Value	Optimal Interventions	Percent Change
1	65.3489	[4,0,0,5,1]	-12.80 %
2	234.4767	[5,5,1,0,5]	-3.57 %

It would be intuitive for the lack of intervention to naturally be the lowest amount of activity produced by the model. We consider a minimization of specific group type distributions of Scenario 1 and Scenario 2.

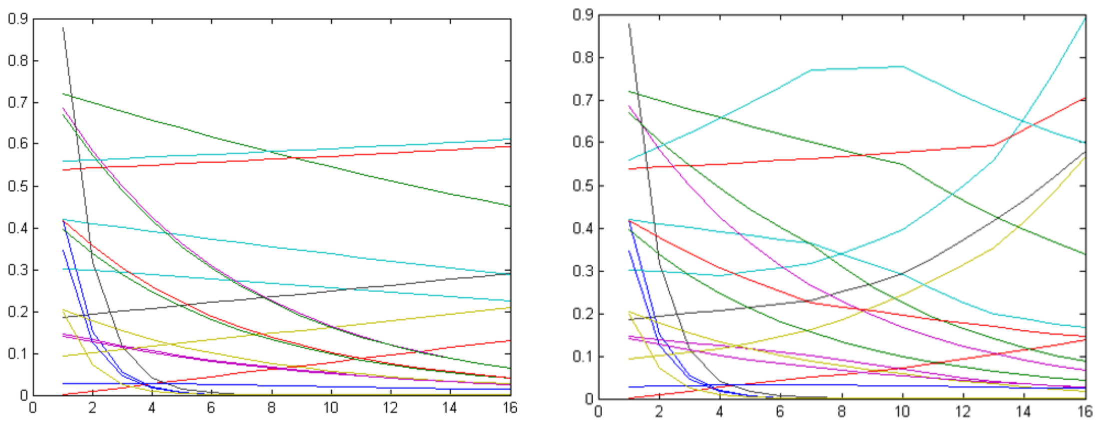

Figure 4.3: Graph of the minimization of Scenario 1 without interventions(left) & with interventions(right)

Here we begin by having 20 groups utilizing the same distribution as Table 4.5. We see that by minimizing the overall activity level we obtain an optimal intervention set of interventions that is not [0,0,0,0,0], but [4,0,0,5,1]. Then we examine 50 groups utilizing the same distribution as the last row in Table 4.11. Again we see that the optimal set of interventions to minimize activity is [5,5,1,0,5]. Thus there is a set of interventions that create a lower activity level than doing nothing. This could be any number of reasons, as often times administration will make changes and they create difficult environments for some groups to be successful; changing procedure for funding, re-establishing a group, etc. Therefore there are groups that end up doing worse after changes then they would have done with no changes.

Chapter 5

Conclusion

This model was developed in an attempt to quantify overall activity of students and student groups on campus. The ability to measure this activity is significant to being able to determine the best way to inform students and all of campus of a policy change. We began developing the model with a general idea of what the overall activity on a college campus could look like. This development style means that the values that were chosen for this specific model and network structure were completely arbitrary. It is important to note that these numbers provided the best overall picture of what was generally expected and then examined with the interventions. The data for the interventions was also arbitrary to investigate the best course of action.

The model developed in Chapter 4 provided some interesting general knowledge about the best set of interventions. The results that are obtained from the specific network structure suggest that regardless of cost limitations, initial start value, and group type distribution, offering a big campus event or conference in the beginning of the semester will help increase overall activity. This can be followed up by various interventions depending up on the cost limitations of the university, but the main interventions to follow typically range from a focus group, web post, or email. The main point is that in order to increase overall activity it is recommended to offer a big campus event early to reach as many people as possible with the information. This can be done through an event at welcome week just before the semester starts.

5.1 Future Research

This model would be much more reliable and realistic if research was conducted to produce the level of response to a group's actibity (α), how each group responds publicly (β), the lack of interest over time (δ), and how connected each group is to one another variables($e_a b$). These values are expected to be unique at every campus and to vary from year to year. Thus it would be advantageous to develop research to provide some average values over a span of several years for different size and types of institutions. Being able to provide a model for various enrollment data and Carnegie classifications would be ideal to be able to produce a relatively accurate model that could be applied to an institutions without having to spend a year researching their specific institution.

The model developed here is a highly connected model. It is realistic to assume that not all campuses or student organizations are highly connected. We could see some of these dynamics occur here at University of Tennessee when considering the communication between main campus and the agricultural campus. It would be interesting to see if the set of interventions change with a less connected campus dynamic. It would also be of interest how much the percent interest would change if the connectivity is less. This could be done by changing the code in the makenetwork.m file that produces the E matrix with respect to each group type. It would also be interesting to change connectivity levels in general to see how decreasing e_{ab} could potentially impact the model. The last possibility for future work, pertaining to edges, is having a directed graph model. The way group a reacts to group b may not be the same way that group b reacts to group a, meaning that $e_{ab} \neq e_{ba}$. This could be explored through a directed graph scenario.

One could also vary the cost limitations to more than \$10,000 and \$1000. It would be more accurate to be able to put in a programs actual budget restrictions to be able to provide a more accurate picture of the best ways to intervene for that particular program. The cost of each intervention could also vary depending on the institution and what you are wanting to offer at that particular intervention. These considerations could change the cost of each intervention substantially. There are also plenty of interventions that were not even considered; sending letters or mail, holds on registration accounts, monetary action, tabling at various events, etc. One could also consider doing interventions more frequently than every 3 weeks or less frequently than 3 weeks. Along with cost, exploring if there is any connection between an even group type distribution and the sensitivity of the model to initial values would be interesting future work. One could also explore the probability an intervention happens to a group as something that might want to be controlled or might want to vary more.

Lastly, one could consider doing a comparison of information spread and overall activity. Just because a group knows about a policy change, event, or something happening it does not mean that they will respond to it. Being able to compare the amount of student groups informed to the amount of student groups who actually act can be incredibly important for institutions. This could provide information on which groups to selectively reach out to with changes because they will be the most active about spreading the message or interacting with other groups or communities. This could be of particular interest when the university is seeking to change policies that could experience some negative press or backlash from the student body.

References

[1] Aiello, W., Chung, F., & Lu, Linyuan. (2001). A random graph model for power law graphs. *Experimental Mathematics*, 10(*1*), 53-66.

[2] Diestel, R. (1997). *Graduate texts in mathematics: Graph theory.* New York, NY: Springer-Verlang.

[3] Erdös, P. & Rényi, A. (1959). On random graphs, *Publicationes Mathematicae (Debrecen)*, 6, 290-297.

[4] Erdös, P. & Rényi, A. (1960). On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17-61.

[5] Erdös, P., & Rényi, A. (1961). On the strength of connectedness of random graphs, *Acta Math. Acad. Sci. Hungar*, 12, 261-267.

[6] Newman, M. E., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *PNAS*, 99(*1*), 2566-2572.

Appendix

```
%mastervalue for alpha, beta, delta pertaining to each group type
function [mastera,masterb,masterd,numtypes] = masterabd(key)
    %vector contains the alpha values which is a groups interest
    masteralpha = .1*[0.0 0.25 0.5 0.75 1.0];
    %vector contains the beta values which is a groups outward response
    masterbeta = .25*[ 0.0 1/3 2/3 1] ;
    %vector containes the delta values which is decay over time
    masterdelta =([.8 .4 .2 .1]).^2;
if (nargin==0), key = 1; end
if (key == 1)
    numtypes = 5;
    %alpha is group i's interest in the topic
    mastera = masteralpha([1 2 5 4 3]);
    %beta is group j's outward response
    masterb = masterbeta([1 2 3 2 4]);
    %delta is the decay of group interest over time
    masterd = masterdelta([1 2 4 3 2]);
elseif (key == 2)
end
```

```
%makegroups.m: generates how many of each group type and their properties
function C = makegroups(numgrp,key,Seed,Seed2)
if (nargin==1), key = 1; end
[mastera,masterb,masterd,numtypes] = masterabd (1);
[s1,s2]=RandStream.create('mlfg6331_64','NumStreams',2);
%specifies random stream with seed for C(g).val(1)
s1=RandStream('mt19937ar','seed',Seed);
%specifies random stream with seed for random generation of group types,
%C(g).type
s2=RandStream('mt19937ar','seed',Seed2);
for g=1:numgrp
    C(g).type = randi(s2,[1,numtypes]);
    C(g).alpha = mastera(C(g).type);
    C(g).beta = masterb(C(g).type);
```

```
    C(g).delta = masterd(C(g).type);
    C(g).val(1)= rand(s1,1) ;
end


%makenetwork.m contains how each e_{ab} is decided
function [Edges] = makenetwork(C,key)
if (nargin==1), key = 1; end
masteredge = (1/2)*[0 1/5 3/5 4/5 5/5];
[~,~,~,numtypes] = masterabd (1);
    for ta =1:numtypes
        for tb =1:numtypes
            if ta==tb ;
                E(ta,tb) = masteredge(3);
            elseif  abs(ta-tb)==1
                E(ta,tb) = masteredge(2);
            elseif abs(ta-tb)==3
                E(ta,tb) = masteredge(4);
            else
                E(ta,tb) = masteredge(1);
            end
        end
    end
    n=length(C);
    for ga=1:n
       for  gb=1:n
            Edges(ga,gb) = E(C(ga).type ,C(gb).type);
       end
    end


%runit1: runs the model from intitial time to final time
function C = runit1(C,Edges,t0,tf)
numgrp = length(C);
X = zeros(1,numgrp);  % pre-allocate for holding community values
```

```
for t = t0:tf
    for ga = 1:numgrp
        X(ga) = C(ga).val(t) - (C(ga).delta)*C(ga).val(t);
        for gb = 1:numgrp
            X(ga) = X(ga) + Edges(ga,gb)*C(ga).alpha*C(gb).beta*C(gb).val(t);
        end
        C(ga).val(t+1) = X(ga);
    end
end
end


%sampleplot: provides graph of overall activity level
function sampleplot(C)
% a simple graph of the evolution of the values for all the
% groups over the full time period.
numgrp = length(C);  % number of groups
for g = 1:numgrp
    y = C(g).val;
    t = (1:length(y));
    %semilogy(t,y)  % semilog plot just cause the values grow exponentially
    plot(t,y)
    if (g==1), hold all, end  % force all graphs on the same axis
end
hold off


% master script to set up the data, run the model, and process the results
%without any interventions
numgrp = 50;  % number of groups
key = 1;      % which group data to use
Seed=i;       %random seed, change the 1 to change the seed
Seed2=gtd;
% build the groups (communities)
C = makegroups(numgrp,key,Seed,Seed2);
```

```
% build the network
E = makenetwork(C);
% run the model from week 1 to week 15
C = runit1(C,E,1,15);
% process results
sampleplot(C)
```

```
% makeinterventions.m:creates the key/table of the possible interventions
function I = makeinterventions(key)
if (nargin==1), key = 1; end
if (key==1)  % positive interventions
    I(1).type = 1;  % changes X and delta
    I(1).name = 'email';
    I(1).effect = 1.1;  % multiplicative impact
    I(1).cost = 1;
    I(1).percentage = 0.1;

    I(2).type = 2;  % changes beta
    I(2).name = 'focus group';
    I(2).effect = 0.2;  % additive impact
    I(2).cost = 100;
    I(2).percentage = 0.6;

    I(3).type = 3;  % changes X
    I(3).name = 'visit';
    I(3).effect = 1.4;  % multiplicative impact
    I(3).cost = 60;
    I(3).percentage = 0.4;

    I(4).type = 4;  % changes alpha
    I(4).name = 'campus-wide event';
    I(4).effect = 0.3;  % additive impact
    I(4).cost = 300;
    %I(4).cost = 850;
```

```
    I(4).percentage = 0.3;

    I(5).type = 5;  % changes delta
    I(5).name = 'web post';
    I(5).effect = 0.05;  % additive impact
    I(5).cost = 1;
    I(5).percentage = 0.2;
end


%applyintervention.m : changes the community parameters (C),
%according to the intervention (ints)
function C = applyintervention(C,t,ints,Inters,apply);
% from the key Inters
% apply=1 means apply, apply=0 means unapply (remove)
% Inters tells the
%   type (name, variable changed)
%   the amount of effect
%   the percentage of groups effected
%   cost (not used in this routine)
if ints>0
    itype = Inters(ints).type;
    ieff = Inters(ints).effect;
    iper = Inters(ints).percentage;
else
    itype = 0;
    iper = 0;
end
numgrps = length(C);
for i = 1:numgrps
    if rand(1)<= iper  % random chance of impact
        switch itype
            case 0 % do nothing
            case 1 % email, change X and delta
                if (apply)
```

```
                C(i).val(t) = C(i).val(t)*ieff;
                C(i).delta = C(i).delta/ieff;
            else
                C(i).delta = C(i).delta*ieff;
            end
        case 2 % focus group, change beta
            if (apply)
                C(i).beta = C(i).beta + ieff;
            else
                C(i).beta = C(i).beta - ieff;
            end
        case 3 % visit, change X
            if (apply)
                C(i).val(t) = C(i).val(t)*ieff;
            end
        case 4 % campus-wide event, change alpha
            if (apply)
                C(i).alpha = C(i).alpha + ieff;
            else
                C(i).alpha = C(i).alpha - ieff;
            end
        case 5 % web post, change delta
            if (apply)
                C(i).delta = C(i).delta - ieff;
            else
                C(i).delta = C(i).delta + ieff;
            end
        end
    end
end
```

```
%withinterventions.m
% for a given set of communities (C), with network (E)
% run the model applying a set of interventions (ints)
```

```
% run over 3 week intervals (3x5 = 15 weeks total)
% described via the Intervention Key (Inters)
function C = withinterventions(ints,C,E,Inters)
startweek = 1;
for i = 1:5
    endweek = i*3;
    C = applyintervention(C,startweek,ints(i),Inters,1);
    % apply the intervention
    C = runit1(C,E,startweek,endweek);  % run for 3 weeks
    C = applyintervention(C,startweek,ints(i),Inters,0);
    % unapply the intervention
    startweek = endweek + 1;
end
```

```
%nextinter.m file
% given a set of interventions m values in range 0 .. n
% produce the 'next' one by incrementing the last and
% carrying as needed
function ints = nextinter(ints,n)
m = length(ints);
ints(1) = ints(1) + 1;  % increment first one
for j = 1:m-1  % carry as needed
    if ints(j)>n
        ints(j) = 0;
        ints(j+1) = ints(j+1) + 1;
    end
end
% note if ints(m)>n we are done
```

```
%score.m: this file scores the overall activity level
function v = score(C,key)
% compute the results over the 15 week period
% according to key
```

```
if (nargin==1), key = 1; end

numgrps = length(C);

if (key==1)  % sum positives over all 15 weeks
    v = 0;
    for i = 1:numgrps
        v = v + sum(max(C(i).val,0));
    end
elseif (key>=10) % count above a threshold (key-10)
    threshold = key - 10;
    for i = 1:numgrps
        v = v + sum(C(i).val>threshold);
    end
end


% master script to set up the data, run the model, and optimize
numgrp = 50;  % number of groups
key = 1;      % which group data to use
keyi = 1;     % which set of interventions to load
keys = 1;     % which scoring mechanism to use
Seed = i;      % key for random number generator
Seed2=gtd;
costlimit = 10000;  % maximum cost allowed
% build the groups (communities)
C = makegroups(numgrp,key,Seed,Seed2);
% build the network
E = makenetwork(C);
% get the intervention key
Inters = makeinterventions(keyi);
numint = length(Inters);  % number of different interventions
% run once to set valuesb
C = withinterventions(zeros(1,5),C,E,Inters);
```

```
bestscore = score(C,keys);
bestint = zeros(1,5);
disp('Score with No Interventions:')
disp(bestscore)
% save copy of original community
C0 = C;
sampleplot(C0)
% systematic optimization scheme
tct = 0;
ints = zeros(1,5);
done = 0; % track progress
while ~done
    % compute a set of interventions below the cost limit
    cost = costlimit+1;
    while (cost>costlimit)
        ints = nextinter(ints,numint);
        if (ints(end)>numint)  % all done
            break
        end
        cost = 0;
        for i = 1:length(ints);
            if ints(i)>0
                cost = cost + Inters(ints(i)).cost;
            end
        end
    end
    if ints(end)>numint % done
        break
    end
    C = withinterventions(ints,C0,E,Inters);
    val = score(C,keys);
   sampleplot(C)
      if (val<bestscore)%for minimum
        %(val>bestscore) for maximum
        bestscore = val;
```

```
        bestint = ints;
        disp(val)
    end
    tct = tct + 1;
end
```

Vita

Allison Heming was born in Cincinnati, Ohio. She graduated summa cum laude from Lake Erie College in 2012, with a Bachelor of Science degree in Mathematics. Allison, has a passion for mathematics and higher education which brought her to the University of Tennessee, Knoxville, where she enrolled as a graduate student. She also worked as a Graduate Teaching Assistant in the Student Success Center as an Academic Coach and Supplemental Instruction Coordinator. Upon graduation she plans to apply her mathematical skills and higher education experience to bridge the assessment gap at colleges and universities.