# The Effect of Network Topology in Social Diffusion: A System Dynamics Approach

**Micah McCutchan and Enrique Campos-Náñez**
*Dept. of Engineering Management and Systems Engineering*
*The George Washington University*
*1776 G Street, Suite 161*
*Washington, DC, 20052*
*Phone: 202-994-1245, Fax: 202-994-0245*
*e-mail: micah.mccutchan@gmail.com, ecamposn@gwu.edu*

**Abstract**

Several classical system dynamics models, such as models of disease spread, and technology adoption, are built under assumption of a homogeneous population. These assumptions have been recently challenged by recent results showing that the degree distributions of many social and natural networks, such as the so-called scale-free networks, exhibit long-tailed degree distributions. This paper adopts a system dynamics approach to replicate *preferential attachment*, one of the network dynamics mechanisms known to produce power-scale distributions. We then study the diffusion processes on these networks, e.g. epidemics, product adoptions. We consider a basic compartment model (Susceptible- Infected) and apply scale free network topology in place of the random network topology that is traditionally assumed. The resulting model is used to assess the effect of the topology on the diffusion of attributes throughout the network.

***Keywords***: Network dynamics, diffusion, SIR models, scale-free networks.

# 1 Introduction

The purpose of this paper is to show that systems dynamics techniques can be used to understand the importance of network topology in policy development. Rigorous mathematical treatment of the issues raised herein has been avoided to broaden the intended audience. We are motivated to use systems dynamics techniques because they can make highly complex mathematical problems accessible to decision makers and policy developers.

Networks are found everywhere and on every scale. Whether we are focused on neural networks in the brain, sexual networks[1], or the World Wide Web[2], we find that groupings of objects are defined not only in terms of the objects themselves, but also in the way that they are connected. Whether the context is disease epidemiology, product marketing, or understanding the rise and fall of political and religious movements, many authors (Barabàsi[3], Albert, Watts[4]) have shown that knowledge of the network's topology is key to understanding its dynamical behavior.

Through modeling and simulation, we would like to be able to answer questions about the importance of network topology in policy analysis. For example, if a public health organization were tasked with distributing only a limited number of vaccines with maximum effectiveness, would it be equally effective to distribute them randomly or in some directed fashion? Or in

another problem context, could an understanding of social network topology be used to develop policies to prevent the spread of religious fanaticism? Throughout this study, we may adopt several problem contexts (disease, political campaigns, etc.) to make the analysis conceptually tractable; however, we ask the reader to maintain a broad minded perspective and recognize that the implications of this preliminary analysis extend further than the specific results discussed herein.

# 2 Background

## *2.1 Networks*

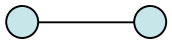
### *2.1.1 Networks and Attributes*

A network consists of a group of objects connected in some fashion. Objects can be people, neurons, computers, or geographic locations. Each object can be distinguished by one or more attributes or states. In building an epidemiological model, we would describe objects as either healthy or infected, while a political campaign model describing the same community or group of objects (people) would describe nodes as republican, democrat, or simply decided or undecided. The same network of objects can look very different depending on the object attributes in which we are interested.

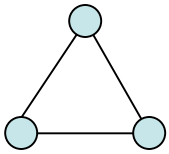
In many networks connections between objects play an important role in determining their state. Nodes will adopt different attributes or states based on the attributes or states of the nodes to which they are connected. For example, the chance that someone will contract a sexually transmitted disease has everything to do with whether his/her sexual partners are healthy or infected. Links can be characterized by interactions (verbal, sexual) or by literal connections (wires in a device, or telephone lines). When a node has made a transition from one state to another because of its links to other nodes in the network, that node is said to have *adopted* a particular trait. This study will focus on questions of adoptership and the diffusion of traits throughout the network. We will address the degree to which network topology should be considered in developing policy.

### *2.1.2 Terminology*

A *node* is an object in the network. As mentioned above, at any time *t* a node can exhibit one or more attributes or traits. Nodes are connected by *links*. In our models, nodes change state based on the number of links they have and the state of the nodes to which they are connected. A node with n links is said to have *degree n*. Two simple networks are shown in Figure 1 below:



Two Node Network where each node is of degree 1



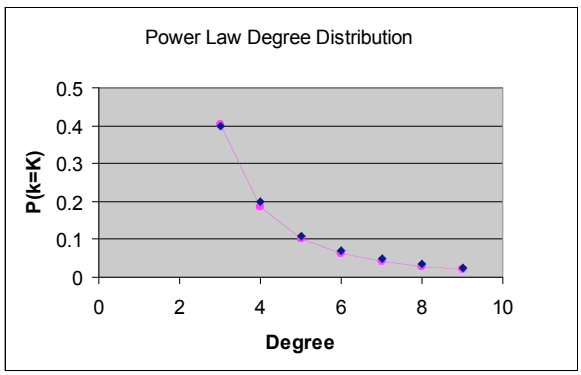
Three node network where each node is of degree 2

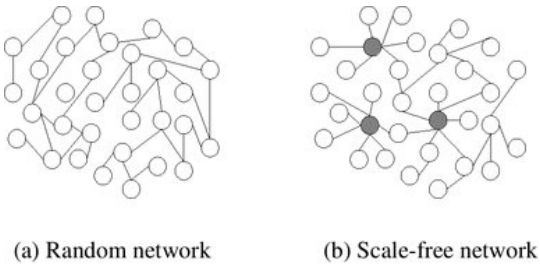
**Figure 1: Simple Networks**

All nodes in the network need not necessarily have the same degree; i.e. some nodes may be more highly connected than others. The structure of the network can be more easily understood by creating a histogram, where the number of nodes of each degree are counted and plotted in increasing order from 1 to n. This histogram can be used to create a probability mass function that describes the degree distribution of the network where *P(k)* gives the probability of finding a node of degree *k*. This profile can be used to characterize the connectivity of network as a whole and to determine its similarity or dissimilarity to other networks.

### *2.1.3 Degree Distributions*

The Erdõs-Rényi (ER) model is often used to generate complex networks. The ER model assumes a collection of *N* nodes where each pair of nodes is connected with probability *p*. This creates a random graph with a degree distribution that is strongly peaked at its average value <*k*> and decays exponentially for large degree values. For the ER model, the degree of many nodes in the network can be reasonably characterized by the expected value of the degree distribution. This characteristic has been assumed to be applicable for many networks and has been broadly used in epidemiological (ex. SIR model)[5] and other network studies. Recently, studies have shown that this assumption may not be valid in a significant number of networks[3,4]. Social networks, neural networks, and computer networks to name a few have been shown to have connectivity distribution that looks quite different from the normal degree distribution[6]. In these networks, the connectivity distribution looks more like a power law (see Figure 2).



(a) Random network (b) Scale-free network

**Figure 2: Power Law Degree Distribution, Random and Scale Free Networks**

Note that in networks with power-degree distributions low degree nodes are common but there exist a few nodes of high degree that are very well connected. Highly connected pivotal nodes or

"hubs" are not recognized when a ER model is assumed, but we shall see that these highly connected nodes are pivotal in the diffusion of attributes through the network.

The critical difference between random networks and scale-free networks is that the ER model degree distribution can be reasonably well characterized by its average value. Networks with power distributions are harder to characterize because the average value of the degree distribution tells us very little about the majority of nodes in the network. They are called "scale-free networks" in part because statistics like averages and variances tell us very little about the majority of the nodes.

## *2.2 Traditional Epidemiological Models and Scale Free Networks*

The Susceptible-Infectious Model (SI-Model) is one of the simplest epidemiological models[7]. We will briefly adopt a disease epidemiology terminology, but we would like to remind the reader that the concepts are applicable in broader contexts. The model divides a sample population into two categories: persons susceptible to infection and those that are already infected.

In the traditional SI-Model, the growth of the adopter population is a function of the infectivity *i,* the contact rate *c*, and the fraction of the population that has already been infected. The infection rate is given as:

$$IR = ciS\left(\frac{I}{N}\right) \tag{1}$$

Where $S$ is the susceptible population, $I$ is the infected population and $N$ is the total population ($N=S+I$). The contact rate $c$ can be assumed to represent the average number of people that a susceptible person interacts with per time step, where $i$ is the likelihood that one of those interactions will result in an infection. Note that the infections are stochastic, meaning that chance determines if a new infection will occur from an interaction with an infected node. The model becomes more complicated when the total population grows over time, or if people are only infectious for a period and then recover from the disease.

In this traditional model, it is assumed that the connectivity structure of the population can be characterized by a single value ($c$). If the model were used to study sexually transmitted disease, it would assume that you were almost equally likely to have the same number of sexual contacts with people of every sex, age, location, attractiveness, wealth, and social status. The point is made simply to illustrate that actual human interaction may have a structure that is more complex than that assumed in the model above and should be considered in more complicated models. Social networks such as the network of sexual relationships are scale free[3,4], implying that there is a significant number of persons with a very large number of sexual contacts, affecting the dynamics of diffusion in the network.

We proceed as follows. In Section 3, we discuss our approach to capturing scale-free network dynamics using standard system dynamics techniques. Then, in Section 4, we use this structure

to model the dynamics of infection diffusion in such a network, and compare this to the results of the classical SI model. Later in Section 5, we model market competition assuming only word-of-mouth dynamics, but exploiting the network topology. We show that in these cases knowledge of network structure is a key factor in policy analysis. Finally, we discuss the relevance of these results for the system dynamics community, in general, and policy analysis of these issues.

# 3 Scale-Free Network Model

## *3.1 Description*

First, we develop a simulation that generates Barabàsi-Albert (BA) scale-free networks using stock and flow models, based on the *preferential attachment* mechanism. The BA network is among the simplest scale free models and can be generated recursively in the following way[8]:

> …starting with a small number ($m_0$) of vertices, at every time step $t$ we add a new vertex with $m(\leq m_0)$ edges that link the new vertex to $m$ different vertices already present in the system. To incorporate preferential attachment, we assume that the probability $\prod$ that a new vertex [connects to a node] $i$ depends on the degree $k_i$ of that vertex, so that
>
> $$\Pi(k_i) = k_i / \sum_j k_j \qquad (2)$$
>
> After $t$ steps the model leads to a random network with $t+m_0$ vertices and $mt$ edges. [Note: In practice, the initial distribution of nodes must be accounted for in the calculation of the number of edges that the network will have at time $t$, so the actual number of edges may be slightly higher than $mt$.]

We can capture the preferential attachment dynamic in a stock and flow model, by using a separate stock to hold the number of nodes in the network with a certain degree, shown in figure (3) using iThink's array diagram. As new nodes enter the network they will establish $m$ new links. The effect of these new links is promoting $m$ nodes from a given degree, to the next degree. These nodes are chosen according to Equation (2). Formally, we have

$$Degree[n](t+1) = Degree[n](t) + \mathrm{Pr}omote[n-1](t) - Conservation[n](t),$$

$$P\left(\mathrm{Pr}omote[n](t) = k\right) = \binom{m}{k} p^k (1-p)^{m-k},$$

$$Conservation[n](t) = \mathrm{Pr}omote[n+1](t),$$

where $p = \dfrac{nDegree[n](t)}{\sum_{i=m_0}^{\bar{m}} iDegree[i](t)}$.

The stock and flow diagram of the model is shown below, using iThink's array feature to denote multiple (indexed) diagram elements. In this case, the elements *Degree n, Conservation of Flow, Promote, Cumulative Pn, Pn, and Weighted Membership* are indexed by the degree number.

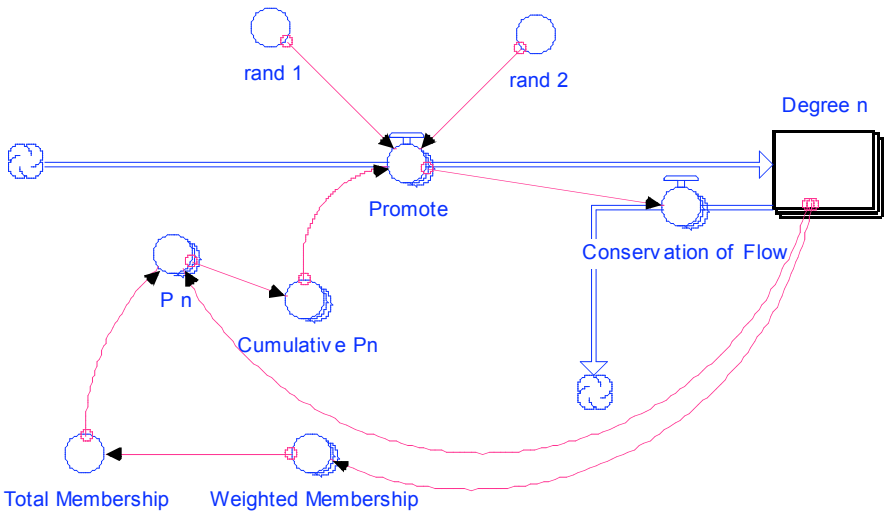



**Figure 3: iThink model of preferential attachment dynamics**

For each simulation in this section, the model begins with $m_0$=3 vertices, each of degree 2 ($k_2$ =3), and adds a new node to the network at each time step. Each new node is allowed to establish two connections to existing nodes. For each time step the probability that an entering node will attach to an existing node of degree *i* is given by Equation (2); this method of attachment favors highly connected nodes.

## *3.2 Validation and Assessments of the Model*

We will attempt to verify that the model does in fact generate an approximate B-A network by comparing the connectivity distribution for the network simulation with the analytic solution for the B-A network[9]. The analytic solution yields the following connectivity distribution:

$$P(k) = 2m^2 k^{-\gamma} \tag{3}$$

indicating the probability of being a node with degree k, where m is the number of attachments a node makes to existing nodes upon entrance to the network, and $\gamma$ is the scaling exponent that determines the shape of the connectivity distribution. The average connectivity for this distribution can be shown to be

$$\bar{k} = \sum_{k=m}^{\infty} kP(k) = 2m \tag{4}$$

We should see the analytic degree distribution reduce to $P(k) = 8k^{-3}$ for our model.

Five simulations were performed; each was allowed to run for 1000 time steps, generating a network with 1003 nodes and 2003 links. After each simulation, the final connectivity distribution was recorded. The number of nodes of each degree were averaged across the five simulation runs to find a general degree distribution which is shown in the table below.

**Table 1. Empirical distribution of nodes with a given degree after 1000 nodes are added.**

| Degree | $D_n$ |
|---|---|

| | |
|---|---|
| 3 | 399 |
| 4 | 200 |
| 5 | 109 |
| 6 | 70 |
| 7 | 49 |
| 8 | 34 |
| 9 | 23 |
| >=10 | 95 |

This distribution was fit to Eq 1.2 using a least squares regression. Regression showed the empirical value for $\gamma = 2.71$. Although this does not match the analytic solution of $\gamma = 3$, we can be reasonably sure that the distribution can be modeled with a power law because the R-sq value for the regression was $R^2$=.98.

Because of the disparity between the theoretical and empirical values, we sought to verify that the simulation had reached steady state by creating a dynamic measure of gamma as the simulation progressed. The graph below shows that even after 32,000 time steps, the simulation had not reached the steady state, theoretical distribution.

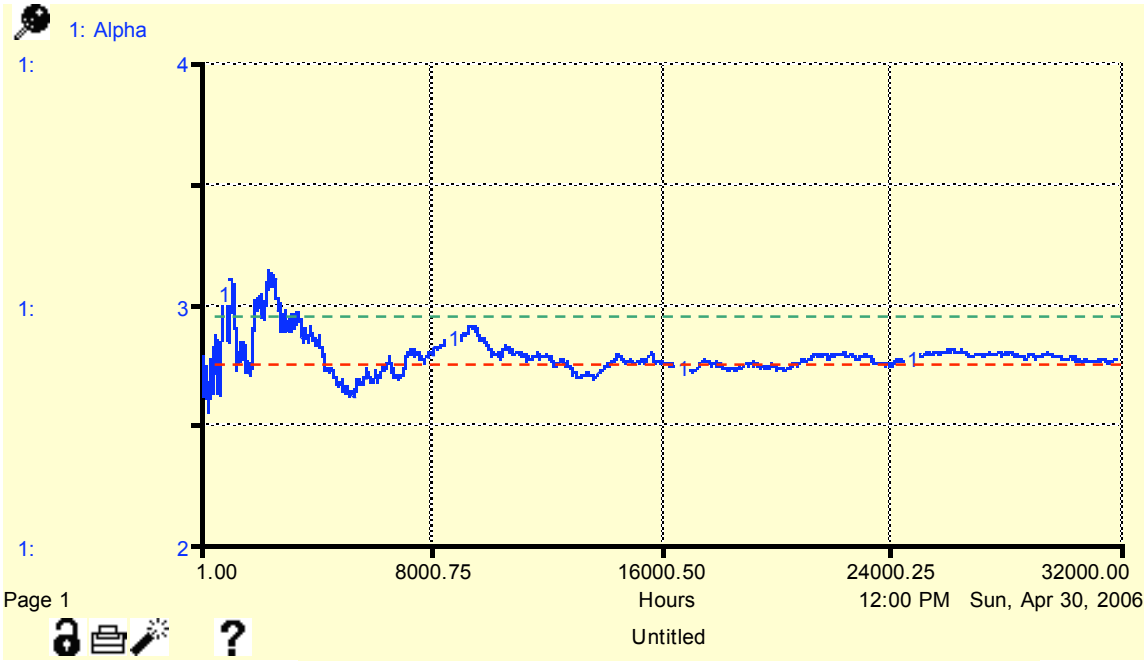



**Figure 4: Calculation of empirical distribution shape parameter**

We are encouraged that several real-world networks have connectivity distributions with gamma values close to 2.71 (See Appendix B- for a table of actual gamma values). The deviation from the analytical solution in our model is not outside the variance in B-A network topology that can be found in the real world. For this reason we accept that the model generates a reasonable BA network and turn to the next stage of the analysis.

### *3.3 Discussion*

The preferential attachment model may exhibit different behavior from the analytical solution for several reasons. Figure 4 shows that the steady state for the system may not have been reached. Because we are still in a stage of transient behavior, deviations from the analytical solution can be expected. Additionally, our model does not distinguish nodes that have degree higher than $n$=10. One of the most predominant features of the scale free network is that its degree distribution is 'heavy tailed' meaning that the distribution has significant mass in the higher degrees. Combining all of the nodes $n > 10$ into the same category biases the preferential attachment structure in the model. High degree nodes are underweighted in the Equation (2), biasing the attachment structure to lower degree nodes. This may also partially explain the deviation from the analytical solution.

As with any model, there are several simplifying assumptions that have been made and should be explicitly discussed so that the model can be applied appropriately. This model assumes a steady growth rate for the network. In fact, networks may tend have growth bursts and periods of relatively constant size. This may affect the dynamics of the network and should be considered as the results of this model are applied. Additionally, this network model does not allow existing nodes to form links. The only new links in the network are formed entering nodes. Additionally once links have been formed, they are assumed to last indefinitely. Incorporating transient links into the model would significantly complicate its dynamics. In general, we feel the value of this model is in its simplicity and hence its accessibility; as is always the case, systems in the real world may be significantly more complicated.

# 4 Scale Free-Attach 2 SI Model

The next step is to adapt the model to incorporate two distinct populations (susceptible and infected) and to establish structures by which infections occur. The "Degree n" bin array from the previous section will be renamed "Potentials" (Susceptible) and a second set of bins is added to the model and will be labeled "Adopters" (Infected). The adopter bin works the same way as the potentials bin with the value of $A_n$ representing the number of infected nodes in the network with degree n. The valve "Infections" moves a node from the potential population to the adopter population when it is infected.

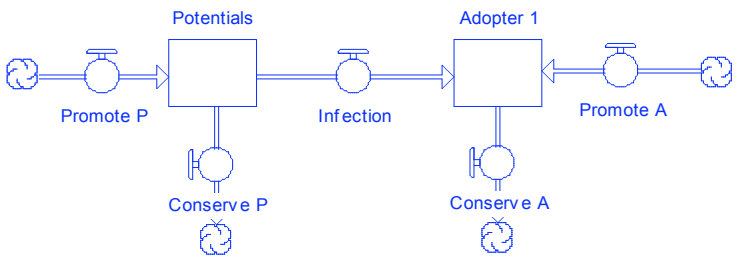



**Figure 5. Basic stock and flow structure for the SI-Scale-free model.**

All nodes entering the network do so via the potential population by preferential attachment structure (Equation (2)). Because the network is separated into potentials and adopters, the probability of attaching to a degree *i* node is now,

$$\Pi(k) = \frac{(A_k + P_{ki})}{\sum_{j=1}^{\bar{m}} (A_j + P_j)} \tag{5}$$

Much the same as epidemiological infections, making a link with an infected node does not guarantee that an additional infection will occur. When a node establishes a link, the probability that the link is infectious is simply the number of potentially infected links in the network divided by the total number of links, which we shall call $\theta$. We define the infectivity to be the probability that an infected link will result in a new infection and denote it by $\lambda$. Therefore, the product $\theta\lambda$ gives the probability of a randomly chosen link resulting in a new. With this in mind, we find that the probability that a node of degree *n* will become infected is simply

$$P_I(n) = 1 - (1 - \theta\lambda)^n \tag{6}$$

At each time step every potential node is susceptible to new infections. Nodes either become infected or remain in the potential population, so each time step can be considered a series of Bernoulli Trials where the probability of infection is given by the equation above. Consequently, the number of degree *n* nodes that are infected at each time step can be considered a binomial random variable with $P_n$ independent trials and probability $P(I|k_n)$ of success. The expected number of infections, per degree, at each time step is

$$E[I_n] = P_n \cdot P_I(n) \tag{7}$$

Highly connected nodes are more likely to become infected than less well connected nodes.

## *4.1 Validation and Assessment of the Model*

Our first step is to verify the qualitative relationship between the infectivity and the growth of the adopter population.

An exact analytical solution for the infection rate as function of time will be difficult to obtain. Instead, we note that the infection rate at each time step will equal the sum over all degrees for the infection value,

$$IR = \sum_n E[I_n] \tag{8}$$

and that each term in the sum increases its value with the infectivity λ. Because there are no recoveries and the infection rate always remains positive, we predict that the adopter population growth rate will increase with the infectivity λ.

| | |
|---|---|
| 0.003 | Green |
| 0.004 | Orange |
| 0.005 | Yellow |

| λ | Color |
|---|---|
| 0.000 | Blue |
| 0.001 | Red |
| 0.002 | Pink |

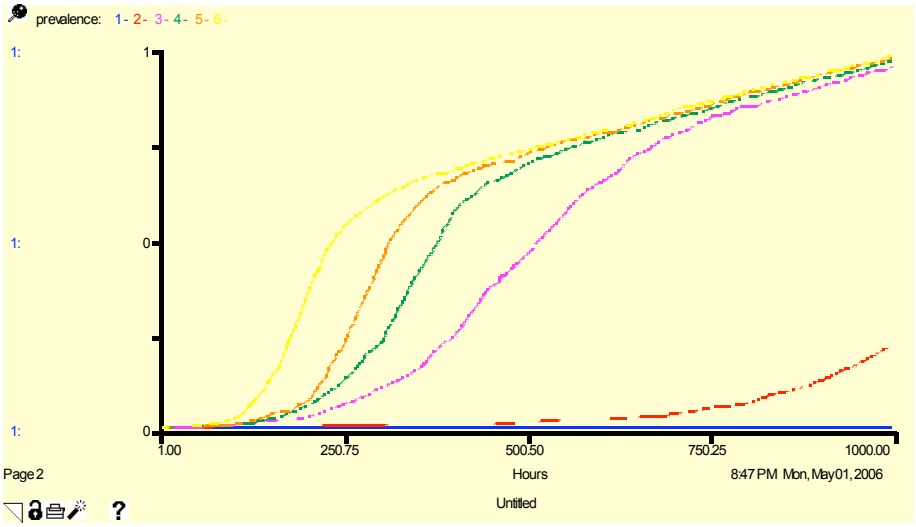



**Figure 6: Infection Prevalence In Scale Free Network for different infectivity values.**

Figure 6 shows that the rate at which the adopter population grows increases with the infectivity λ. When the graph above is compared to a traditional SI-model that assumes random network topology, we verify that network topology significantly affects the growth rate of the epidemic.

A traditional SI-model was constructed with an initial susceptible population of 1000 and 3 initial infections. A single individual was added to the susceptible population at every time step and the simulation was allowed to run for 1000 iterations. The contact rate was assumed to be the average connectivity from the analytic solution to the BA-network. The same infectivity values were used as in the Table that accompanies Figure 6.

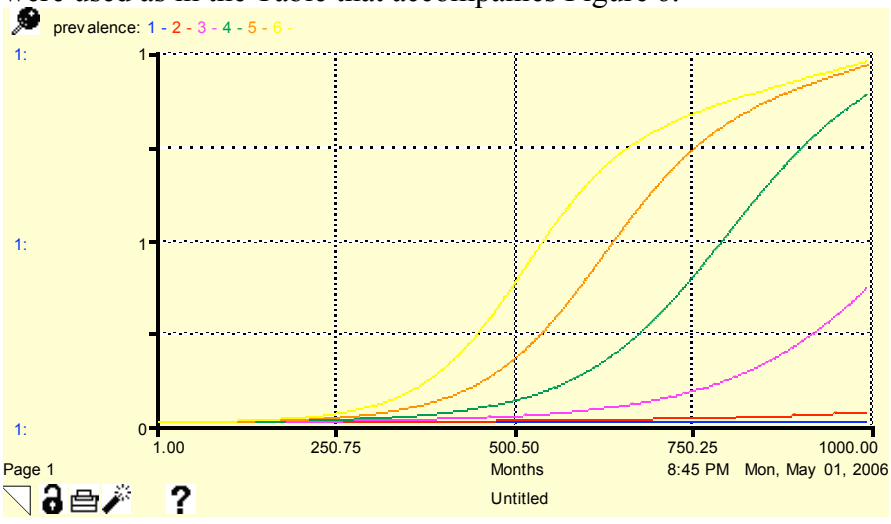



**Figure 7: Growth of the adopter population for the same infectivity values shown in Figure 6 using the traditional S-I model with random network topology.**

The comparison indicates that stochastic infections may propagate more quickly through scale free networks than in random networks.

## *4.2 Sensitivity to the Initial Distribution of Infected Individuals*

Next, we explored the affect of the initial distribution of infections on the growth of the adopter population. Two scenarios were considered, one where the adopters were selected randomly over the connectivity distribution of the initial potential population, such that the initial number of adopters of each degree is given by,

$$A_n = I_0 \Pi_n \tag{9}$$

where $I_0$ is the total number initial adopters. In the second scenario, the most highly connected $I_0$ members of the network are chosen as the initial adopters. More specifically, the number of initially infected individuals of each degree is determined by starting at the highest degree and moving $I_0$ individuals to the adopter population. If $I_0 > P_{\bar{m}}$ then the remaining $I_0 - P_{\bar{m}}$ infections are taken from the next $P_{\bar{m}-1}$ individuals with a smaller degree. This process repeats until $I_0$ of the most highly connected nodes are infected.

**Table 2:Summary of results of the SI Scale-free network where we measure the mean and standard deviation of the time at which the adopters overcome potential persons.**

| Simulation Time Step Where $\sum_{n=2}^{10} A_n \geq \sum_{n=2}^{10} P_n$ | | | | |
|---|---|---|---|---|
| | Random Adopters | High Degree Adopters | Random Adopters | High Degree Adopters |
| $I_0$ | 95% Confidence Interval, μ | | 95% Confidence Interval, $\sigma^2$ | |
| 5 | [1105, 1215] | [903, 959] | [10335, 29447] | [2636, 7511] |
| 10 | [895, 966] | [758,795] | [4407, 12558] | [1160, 3304] |
| 50 | [568, 597] | [477, 502] | [756, 2153] | [541, 1540] |

The second and third columns of Table 2, indicate that in a scale-free network, the initial distribution of infected nodes does affect the growth of the adopter population. This is not true in traditional epidemiological models where all nodes are assumed to have the same number of relationships. In scale free networks, infecting the highest degree nodes will shorten the time needed for the attribute to be adopted by more than 50% of the population. This may have broad implications in disease epidemiology, network security, intelligence communities, and marketing activities. For example, sexual networks have been shown to have scale free topology, public health professionals may have reason to focus interventions on the most sexually active segments of the population. Similarly, IT security professionals may want to focus their security efforts on the most highly connected computers or websites and marketing professionals may want to focus promotional efforts on celebrities and other highly connected people. The policy implication may be that controlling adoptership in networks depends on their topology; in scale free networks the most highly connected nodes are key to either promoting adoptership or preventing it.

The variance in the growth of the adopter population has equally important policy implications. At the 5% significance level, we can say that the growth of the adopter population has greater variance with randomly distributed infections than with the infection of only the most highly connected individuals (except in the case where 0.5% of the initial network is infected where we can not reject the hypothesis that the variances are the same). This indicates that network topology may also affect have policy implications on predictability. Because there was more variance in the growth of the adopter population when infections were randomly distributed

throughout the network, policies that are directed at highly connected individuals may play out more predictably.

### *4.3 Discussion*

Equation (7) describes the number of infections per degree as a Bernoulli random variable. Software constraints led us to use the Poisson distribution instead. The two distributions converge for very large numbers of trials, but the discrepancy may have been significant for the number of 'trials' in our model (persons in the potential population of the nth degree typically ranged from [10, $10^3$].

The assessment of the parameter $\lambda$ for this model is intended to demonstrate to the reader that the infectivity parameter affects the growth of the adopter population. The comparison with the growth of the adopter population in the random network SI model is intended to demonstrate that the two models are not equivalent. The scale free model may not exhibit faster growth of the adopter population in all cases. In-depth analysis is needed to complete a proper comparison of the models and their dynamics.

Similarly the second part of the assessment is intended to demonstrate that the initial distribution of adopters affects the growth of the adopter population. Confidence intervals shown in Table 2 have been calculated for a single sized network with a constant infectivity parameter. The effects of the initial distribution should be considered on additional scales for a range of different sized networks. Additionally, there may be interaction effects between the infectivity parameter and the initial distribution of adopters. In practice, they should not only be studied independently, but also in combination.

## 5 Competing Infections

### *5.1 Description*

This model includes a second adopter population and was built to examine competition in the same scale free network. In the second adopter population, Adopter 2, infections occur by the same method as in the previous model, with each adopter type having its own infectivity ($\lambda_1$ and $\lambda_2$) and initial distribution of infections $A1_n$ and $A2_n$ with [$3<n<10$]. New nodes are still incorporated into the existing network by preferential attachment, although equation (2) must be slightly modified to include members of the $A2_n$ network. The model structure is shown below and a detailed explanation of the model can be found in the appendix.

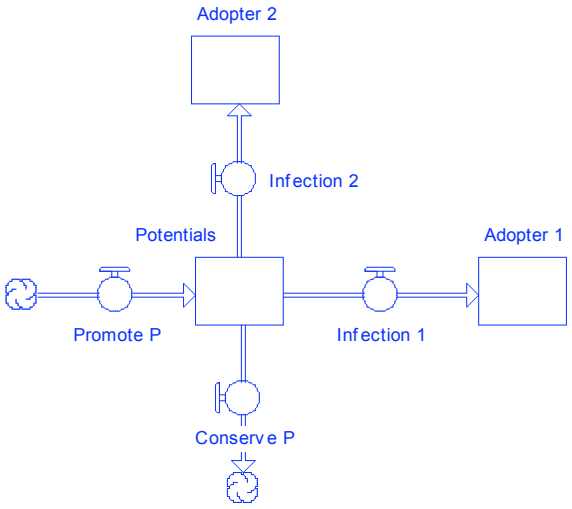



**Figure 8. Stock and flow diagram for a system with two competing adoption processes.**

## *5.2 Validation and Assessment*

The purpose of this analysis section is to begin to understand the effect of both the initial distribution of adopters and the infectivity for competition in the same scale-free network. This model may have implications when two entities are competing for adoptership, such as businesses competing in the same marketplace, or political candidates and campaign activities.

This preliminary analysis can be used as a fist step to understanding the effect of highly connected individuals such as celebrities in marketing activities or powerful lobbyists in political campaigns. Political campaigns provide an excellent framework to start to ask these types of questions and so we will use that particular context for the bulk of this analysis, but it should be recognized that the results can be extended, by analogy to many other contexts as well.

Consider the start of a local campaign in a small town somewhere in the United States. There is some segment of the population, say 33% that are already aware of the issues and the two political candidates, this group comprises the initial potential network. As the election draws closer, the pool of people who become interested in the campaign increases and more and more people join the potential population. When a member of the community has selected their candidate, they are said to have adopted that candidate and they will move to the adopter population *A1* or *A2* depending on their choice. The infectivity parameter $\lambda$ will represent in the likelihood that a voter will be 'infected' with a candidate's message and will offer their vote; but in general the infectivity parameters could take on several interpretations. In further analysis, $\lambda$ could be considered a function of the effectiveness of the candidate's message, personal charisma or possibly the size of the financial resources available. The initial distribution of adopters could be campaign staff, friends and family.

We can now use the model to answer questions about David and Goliath type campaign battles. For example, how difficult is it for an unknown candidate to capture more votes than a well connected incumbent? How much more infectious must the challenger's message be to regain

the ground lost by the incumbent's name recognition and political connections? Or in the event that each candidate's message has the same appeal, how many more people must the challenger contact to make up for the advantage afforded by the pre-existing relationships of the incumbent?

In this analysis A1 will be the adopter population of the well-connected incumbent with infectivity parameter $\lambda_1$. The challenger will be represented with A2 adopter network, with infectivity parameter $\lambda_2$. The fist step in this analysis is to get an idea of the sensitivity of the competition to differences in the infectivity parameters. For example:

> Does the challenger's message need to be twice as infections as the incumbent's to make a clear difference in the outcome, or do smaller changes have significant effects?

For this part of the analysis we let $\lambda_1 = 0.001$ for all model runs; this will be used as the comparison value. As a preliminary analysis strategy, we made adopter 2 twice as infections as adopter 1 such that $2\lambda_1 = \lambda_2$. We ran 30 simulations, each for 2000 time steps. At the end of each simulation, we will record the candidate who has captured more of the population. The number of times that the more infectious candidate wins the election will be recorded. This process was repeated for the range of values shown in Table 3.

**Table 3: Mean and Variance for the probability that the candidate with higher infectivity captures more of the potential population**

| $\mu$ describes winning percentage of candidate 2 | | | | | |
|---|---|---|---|---|---|
| $\Lambda_1$ | $\lambda_2$ | X-bar | 95%CI $\mu$ | S | 95%CI $\sigma^2$ |
| 0.0010 | .0010 | 0.533 | [0.315, 0.752] | 0.506 | [0.16, 0.46] |
| 0.0010 | .0011 | 0.800 | [0.624, 0.976] | 0.407 | [0.11, 0.3] |
| 0.0010 | .0012 | 0.700 | [0.499, 0.901] | 0.465 | [0.14, 0.39] |
| 0.0010 | .0013 | 0.933 | [0.824, 1.043] | 0.252 | [0.04, 0.11] |
| 0.0010 | .0014 | 0.967 | [0.888, 1.045] | 0.181 | [0.02, 0.059] |
| 0.0010 | .0015 | 0.967 | [0.888, 1.045] | 0.181 | [0.02, 0.059] |
| 0.0010 | .0020 | 1 | [1,1] | 0 | [0,0] |

The table shows that the competition was clearly sensitive to differences in the infectivity larger than 30% at the 95% significance level. This gives us a basis to be able to start comparisons in the competition model.

The next question we'd like to answer in the context of the campaign battle scenario is:

> Given that the incumbent has connections to the top 1% most connected people in the community, how much more infectious must the underdog's message be to have a chance of winning the election? [given that the challenger starts the campaign with a random 1% of the population as supporters]

As a preliminary analysis, we will start the campaign simulation with equal infectivity (i.e. $\lambda_1 = \lambda_2 = .001$) and gradually increase the challenger's infectivity. At each level, we performed 30 simulations and recorded the percentage where the challenger was able to beat the incumbent. Our plan was to gradually increase the challenger's infectivity until the 95% mean confidence interval for the election outcome captured the value .5. This would give a preliminary indication of the magnitude of the impact of the infectivity parameter verses the initial distribution of

supporters. To our surprise, even an infectivity of $\lambda_2 = 1$ was not enough for the challenger to catch the incumbent by the end of the simulation. In fact, the incumbent was victorious in all 30 trials, even when there was a 100% chance that a potential voter would support the challenger if they came into contact with one or more supporters.

Next, we turned to the question of the size of the initial supporter network:

> Given that the incumbent had was supported by the top 1% most connected individuals, what size would the underdog's randomly distributed network of supporters be to offset the incumbent's advantage?

We will gradually increase the size of the underdog's "grass roots" support network and record the fraction of the time that the *incumbent* is ahead at the end of the simulation. We will continue to increase the size of the challenger's network until the 95% confidence interval for the incumbent's win percentage no longer contains the value .5. Values for the trials are shown in Table 4 below.

**Table 4: Mean and Variance for 'Incumbent' win percentage, given a highly connected network of supporters representing 1% of the total population. Compared against various sizes of the challenger's randomly distributed initial network of supporters.**

| Initial Supporters as a percentage of total initial population | | | | | | | |
|---|---|---|---|---|---|---|---|
| Highly Connected | Grassroots | Victory Percentage (Incumbent $A_1$) | | | | | |
| Incumbent ($A_1$) | Challenger ($A_2$) | Sample Mean | 95%CI $\mu$ | | Sample St. Dev. | 95%CI $\sigma^2$ | |
| 1% | 1.5% | 63% | [.422 | .845] | | [0.15 | 0.433] |
| 1% | 2% | 60% | [.385 | .815] | | [.16 | .488] |
| 1% | 3% | 20% | [.025 | .375] | | [.10 | .30] |

This preliminary analysis has important implications in policy development. In the context of the campaign battle, this analysis may indicate that a challenger should focus on the size of his/her grassroots network more than the infectivity of the message.

## *5.3 Discussion*

Table 3 was assembled determine the sensitivity of the competition to differences in the infectivity values of the competitors. Similar analysis should be repeated in multiple networks of different sizes, and with a range of comparison values to determine if the significance level here is the same across all values. This analysis was intended to show one possible method by which such distinctions could be made and should not be considered a universal result.

Similarly, the second part of the assessment is not intended to show that the 'challenger' population would never be able to compete with a well-connected competitor. The results should not be generalized at this point and a rigorous analysis that includes multiple sized networks and a range of initial adopter populations is needed to determine the general behavior of the system. The reader should take from the discussion the point that in this model, for the

parameters chosen, there is a point where the connectivity of initial adopter population makes the challenger unable to catch up, even with an equally sized random distribution of initial adopters and an infectivity value of 1. The third part of the assessment should be treated similarly in terms of the generality of the result and the need for further analysis. The analysis shown is intended to support the general theme that network topology plays a significant role in diffusion behavior in networks.

# Conclusions

Systems Dynamics tools can be used to present the importance of network topology policy development. We have shown that these techniques can provide insight into the effects of infectivity parameters and initial distribution of adopters on the diffusion of properties through a network. Section 3 demonstrates that uncomplicated systems dynamics models can be used to present the preferential attachment mechanism and to verify that the models generated by the simulations are similar to those that are found in many real-world problem contexts. Section 4 showed that systems dynamics concepts can be used to demonstrate that the diffusion of attributes through a network was different depending on the network topology that was assumed. Section 5 showed that similar techniques could be used to explain the importance of highly connected nodes and their affect on competition in established networks. Each of these concepts has been made more accessible to decision makers who may not have the background needed to understand more analytical treatments of these issues.

This study is intended to demonstrate the value of systems dynamics techniques in making highly technical concepts accessible to decision makers. The implication is that systems dynamics should be developed so that the benefits of advanced network theories are not only understood by highly specialized, technically trained individuals. Policy development is very difficult in the midst of systems and networks that behave counterintuitivly. Systems dynamics can be used to explain the existence of non-linear behaviors, and the magnitude of their effects. Bridging this gap is essential for intelligent policy development.

Additional work needs to be completed to make recent advancements in network theory available to decision makers at all levels. Systems dynamics has demonstrated enormous potential to bridge the gap between esoteric theories and mainstream thinking. Specifically, along the lines of this paper, these models should be adapted to incorporate more complex mechanisms of networks and social diffusion. Temporary adoptership should be incorporated as in the Susceptible-Infected-Recovered model widely used in disease epidemiology. More complex network generating algorithms should be incorporated into these models, including those that allow existing nodes to establish new links, and those that allow links to exist for a finite period. The dynamics of attachment (network growth) are varied and complicated; systems dynamics may be used to make them increasingly accessible.

The network models presented herein will certainly need to be adapted to alternative problem contexts in order to reap their full demonstrative benefits. Modeling religious fanaticism and acute infection epidemics (avian flu) seem to be two areas that are at the forefront of national and

international concern. There may be great benefit in adapting these techniques to allow decision makers and policy developers to gain insight into the mechanics of these problems.

# Appendix A: Scale Free Attach 2 Model

The model consists of the following pieces:

Stocks

Degree n – there are 9 bins in this array, the value of *bin(i)* gives the number of nodes in the network with degree *i* for *2< i <10* with the exception that the value of *bin(10)* is the number of nodes in the network with degree >10.

Valves

Promote – When a node makes a new link, this valve array allows the node to advance in degree from *bin(n-1)* to *bin(n)* such that the value of *bin(n)* increases by 1. Because each new node is allowed to make two new links upon entering the network, each of the 9 valves can take on the value 0,1 or 2 at each time step subject to the constraint that the $\sum_i P_i < 2$ for each time step *t*.

Conserve - When a node makes a new link and advances in degree from *bin(n-1)* to *bin(n)*, this valve array allows the value of bin(n-1) to decrease its value by 1. It is true that for each C(i), C(i)=P(i+1).

Converters

Pn – is an array converter. For each Degree(i) the converter takes the value of $\prod(k_i)$ from Equation 1.1.

Cumulative Pn- is an array converter. For each Degree(i) the converter takes the value of $\sum_{i=2}^{i} \Pi(k_i)$ which is the probability that a new node will make an attachment to an existing node with Degree < 2.

Total membership and Weighted membership are simply perform intermediate calculations for Pn

Rand1 and Rand2- are random variables taken from Uniform(0,1) distributions respectively. They are used to determine which attachments will an entering node will establish when it joins the network. The first attachment is determined by Rand1. At every time step, Rand1 produces a number between 0 and 1. This value is checked against the values Cumulative Pn array at that time step. The entering node will establish a link with an existing node of degree *i* if Cumulative P(i-1)<rand1<Cumulative P(i), at which point Promote(i+1)=1 and Conserve(i)=1. The same is true for the second connection with Rand2.

# Appendix B: Shape Parameters for Scale Free Networks

| Network | Size | Gamma | Reference |
|---|---|---|---|
| Citation | 783,339 | 3 | S. Tedner, "How popular is Your Paper? An Empirical Study of the Citation Distribution" European Phys. J. B4, 131-135 (1998) |
| Words, Concurrence | 460,902 | 2.7 | R., Ferrer I Cancho and R. V.Sole, "The Small World of Human Language" Proc. Poy. Soc., London B 268, 22 61 2001 |
| Words, Synonyms | 22,311 | 2.8 | S Yook, H Jeong, & A. L. Barabasi, unpublished (2001) |
| Phone Calls | $53 \times 10^6$ | 2.1 | W., F. Chung, & L. Lu, Proceedings of the 32nd ACM Symposium on the Theory of Computing, ACM, New York, p. 271 (2000) |
| Sexual Contacts | 2810 | 3.4 | F. Liljeros, C.R. Edling, L.A.N. Amaral, H. E. Stanley, & Y. Aberg, Nature, London, 411, 907 (2001) |

# References


[1] F. Liljero, C.R. Edling, L.A.N. Amaral, H.E. Stanley & Y. Aberg, Nature, London 411, 907 (2001)
[2] R. Albert, H Joeng, & A.-L Barabàsi, Diameter of the World Wide Web, Nature 400 130-131 (1999)
[3] A.-L Barabàsi, "Linked", 2003, Penguin Books Ltd., USA
[4] D.J. Watts, "Six Degrees" 2003, W.W. Norton & Company, NYC/London
[5] Kermack, W. and A McKendrick (1927) Contributions to the mathematical theory of epidemics, Proceedings of the Royal Society, 115A 700-721
[6] A.-L Barabàsi, "Emergence of Scaling in Complex Networks", from S. Bornholdt, H.G. Schuster (Eds.) Handbook of Graphs and Networks, WILEY-VCH 2003
[7] J. D. Sterman, Business Dynamics, McGraw-Hill, 2000, USA
[8] A.-L Barabàsi and R. Albert, Emergence of Scaling in Random Networks, *Science* **286** (1999), 509-512
[9] R. Pastor-Satorras, A. Vespignani, Epidemics and immunization in scale-free networks, from S. Bornholdt, H.G. Schuster (Eds.) Handbook of Graphs and Networks, WILEY-VCH 2003