Simulating the Diffusion of Information: An Agent-based Modeling Approach

Cindy Hui, Mark Goldberg, Malik Magdon-Ismail, William A. Wallace
Rensselaer Polytechnic Institute, Troy, NY, USA

ABSTRACT
Diffusion occurs in various contexts and generally involves a network of entities and interactions between entities. Through these interactions, some property, e.g. information, ideas, etc., is spread through the network. This paper presents a general model of diffusion in dynamic networks. We simulate the diffusion of evacuation warnings in multiple network structures under various model settings and observe the proportion of evacuated nodes. The network dynamics occur as the result of the diffusion where nodes may leave the network after receiving the warning. We use the model to explore how the network structure, seeding strategy, network trust, and trust distribution affect the diffusion process. The effectiveness of the diffusion is a function of the network structure and seeding strategy used in delivering the initial broadcast. The simulation results reveal interesting observations on the effects of network trust and distribution of trust in the network.

Keywords: Diffusion, Dynamic networks, Social networks, Trust

INTRODUCTION
Diffusion occurs in various contexts and generally involves a network of entities and interactions between entities. These networks can consist of entities like individuals or organizations. The interactions could be physical contact, collaboration, innovation adoption, or some form of verbal or written communication depending on the circumstances. Through these interactions, some property, e.g. information, idea, innovation, disease, etc., is spread through the network. The flow of this property may in turn have an effect on the entities in the network as well as the network itself. For example, through the diffusion of a product review, individuals may become curious and browse for addition reviews on the product, or they might become convinced and adopt/purchase the product. In additional, individuals may join/form user groups to discuss the product and through these groups, form new relationships and thereby changing their social networks.

We present a general model of diffusion in dynamic networks, where the network may change as a result of the diffusion that occurs. We apply our model to the context of evacuation warnings. The network in this case is a social network of household nodes and the property being diffused is an evacuation warning. As warning messages propagate through the network, households may seek additional information, spread the information, or take action, i.e. perform evacuation. This context demonstrates one form of network dynamics where nodes may leave the network and remove their edges. Observing this form of dynamics may reveal disruptions in the spread of information, and identify properties of the network that facilitate the spread of the warning as well as identify nodes (households) that fail to evacuate. We use the model to explore how the network structure, seeding strategy, network trust, and trust distribution affect the diffusion process.

BACKGROUND AND RELATED WORK

Social networks play a significant role in the spread of information, ideas, emotions, diseases, innovations, etc. As a result, the flows of information, ideas, etc. affect the way people think, act, and bind together in a society. Modeling information flow through various social networks is an active research area, with work on diffusion of innovation and technology (Bass, 2004; Brown & Reignen, 1987; Hill, Provost, & Volinsky, 2006; Rogers, 1995; Valente, 1995; Young, 2000), viral marketing (Leskovec, Adamic, & Huberman, 2006; Leskovec, Singh, & Kleinberg, 2006), the spread of computer viruses (R Albert, Jeong, & Barabasi, 2000; Chen & Carley, 2004), and the spread of diseases (Meyers, Newman, & Pourbohloul, 2006; Morris, 2000).

The spread of infectious diseases and the spread of infectious ideas have common characteristics in terms of their diffusion process. For this reason, many diffusion models for studying the spread of ideas were developed based on models from epidemiology (Bettencourt, Cintron-Arias, Kaiser, & Chavez, 2006). Many of the epidemiology models are derived from the Susceptible/Infected/Removed (SIR) model, which was formulated by Lowell Reed and Wade Hampton Frost in the 1920s (M. E. Newman, 2002). The SIR model divides the population into three possible categories (susceptible, infected, and removed) that reflect the status of the individuals. Susceptible are individuals who are not infected but may become infected when they gain contact with an infected individual. Infected are individuals who are carrying the disease and have the potential to spread it. Removed are individuals who have either recovered from the disease or died, and cannot spread the disease. The model assigns a disease transmission probability based on a given average rate of contact, and assumes that all individuals are equally likely to become infected. Mathematical models can then be used to infer population average parameters such as contact rates and duration of infectious periods.

Many variations of the SIR model have been proposed to incorporate more realistic factors, for example implementing a social structure for contact based spread, (Barthelemy, Barrat, Pastor-Satorras, & Vespignani, 2004; Girvan & Newman, 2002; Kuperman & Abramson, 2001; Moore & Newman, 2000; Pastor-Satorras & Vespignani, 2001) or varying disease transmission probability (M. E. Newman, 2002). In general, these models depict the disease spreading process by tracking the average number of infected individuals and identifying individuals who are prone to become infected with the disease. These models can also identify the specific characteristics, at the population level, that play a significant role in the transmission process. Such characteristics include age, variable infection rates and variable infection periods. These characteristics introduce a heterogeneous population, which also leads to more complex models.

Diffusion models are used to study the adoption of products and spread of innovation influence by viewing them as a process of social interactions. The diffusion of innovation theory, introduced by Everett Rogers in 1962, defines stages of product adoption process which includes knowledge, persuasion, decision, implementation, and confirmation (Rogers, 1995). The product adoption curve classifies adopters into five categories: innovators, early adopters, early majority, late majority, and laggards. The theory suggests that the adoption curve follows an S-curve, in which a small proportion of individuals initially adopt the innovation, followed by relatively quick adoption by the early and late majority, and then levels off as the laggards finally adopt. This theory introduces the concept that for most individuals in the social network, the decision to adopt the innovation is dependent upon the other individuals in the network. Early adopters have

a profound effect on the adoption decisions of the later adopters. Recent research utilized the categories of adopters as introduced in Roger's theory to analyze how the adoption process affects the information flow of product recommendation (Song, Tseng, Lin, & Sun, 2006).
The Bass model for diffusion of innovation is a mathematical model for estimating the adoption of new products. This model introduces factors such as product market potential and interaction rate between consumers and prospective consumers into the model. Early models for innovation diffusion ignore the consumer decision-making activity that occurs in each individual. The characteristics of consumers are an important factor for product adoption. A consumer's decision is highly affected by social influences and interactions that occur over time. The concept of "word of mouth" is commonly used in marketing. It builds on the observation that a consumer's decision to accept a new product depends on what they hear from others (Goldenberg, Libai, & Muller, 2001). The concept suggests that there is a critical threshold, which assumes that the probability of adopting a product suddenly increases as a particular number of friends acquire the product.

In general, the existing diffusion models in the literature focus on two types of approaches, cascade models and threshold models. The most basic models are the Independent Cascade Model and the Linear Threshold Model. The cascade models are similar to the models of the spread of epidemic diseases (Kempe, Kleinberg, & Tardos, 2003; Leskovec, Adamic, & Huberman, 2006; Moore & Newman, 2000; M. E. Newman, 2002). In the Independent Cascade Model, each node gets a chance to influence each of its inactive neighbors with a given probability of success. If the transmission is successful, the neighbor will become active at the next time step. In general, this process continues until there are no more possible transmissions. In the Linear Threshold Model, an individual is infected based how many of their neighbors are already infected. There is a weight on the edge between two nodes, which defines a measure of influence. Each node has a threshold value, which is drawn randomly from some specified probability distribution. This threshold determines how many neighboring nodes have to be activated before the node itself becomes active. If the sum of the weights of all active neighbors exceeds the threshold, then the node will become active (Granovetter, 1978; Young, 2000). The cascade and threshold approaches form the basis for many diffusion models and extensions to these models have been made to study different diffusion processes (Delre, Jager, & Janssen, 2006; Goldenberg, Libai, & Muller, 2001; Kempe, Kleinberg, & Tardos, 2003; Leskovec, Adamic, & Huberman, 2006), as well as identifying variables that affect the diffusion process in cascade models (Centola, Eguiluz, & Macy, 2007) and observing information cascade in viral marketing (Leskovec, Adamic, & Huberman, 2006; Leskovec, Singh, & Kleinberg, 2006).

The structure of the social communication network is a very important factor in the diffusion process. At the two ends of the spectrum of graphs are regular graphs and random graphs. In regular graphs, all nodes have the same degree, i.e. every node is connected to the same number of nodes. Random graphs are generated based on some random process and are often used for proving the existence of certain graph properties. However, these graphs often do not represent how actual social networks are structured.

Scale-free networks and small-world networks are commonly used in studying social network structures. They appear to be more realistic and reflect the characteristics of biological and technological systems. In scale-free networks, most nodes have a low degree while a small subset have high degree. This addresses the phenomenon of the existence of highly connected individuals in a network. The degree distribution follows a power-law relationship in which the structural dynamics are independent of the number of nodes in the network. In small-world networks, nodes are highly clustered with small path lengths between nodes. This phenomenon is commonly found in biological, social, and synthetic systems (M. Newman, 2000; D. J. Watts, 1999) and also appears when analyzing patterns of scientific collaboration (M. E. J. Newman, 2001) and actor collaboration in films (R. Albert & Barabasi, 2002; Strogatz, 2001; Duncan J. Watts & Strogatz, 1998). Small-world networks have been used in studying algorithmic routing of messages in communication networks (Kleinberg, 2000). The speed in which information spreads is dependent on the network structure, changes with the degree of randomness in the network, and has found to increase in small-world networks (Delre, Jager, & Janssen, 2006).

Most previous research assumes the network to be static and does not consider the changes that may occur over time. Dynamic networks are becoming more prevalent in the recent research literature; for example these papers study evolving communication graphs conditioned on a static social group structure (Berger-Wolf & Saia, 2006; Cortes, Pregibon, & Volinsky, 2003; Lahiri, Maiya, Sulo, Habiba, & Berger-Wolf, 2008). In dynamic networks, nodes and edges may appear and disappear with time. There are multiple aspects of dynamics to consider. The local dynamics describe how nodes interact and how the diffusion may spread. This includes changes at the individual node level, e.g. changes in node thresholds or infection probabilities. The group dynamics describe the social group evolution that may occur over time. New nodes and edges may appear as individuals make new friends or join social groups, and/or disappear as individuals relocate or leave groups. When diffusion occurs over a social network, the dynamics of the social network determine who is interacting at each time step, which in turn determines how the diffusion may spread at that particular time step (Goldberg, Kelley, Magdon-Ismail, Mertsalov, & Wallace, 2008). In addition, the network may change due to the diffusion that occurs through the network.

There are many similarities as well as critical differences among the spread of epidemics, innovation, and the diffusion of ideas. The spread of an idea is usually an intentional act. Acquiring new ideas is often viewed to be advantageous and therefore gives a different perspective on the social interaction aspects (Bettencourt, Cintron-Arias, Kaiser, & Chavez, 2006). In the spread of idea, individuals must first hear the idea from another individual. Next, the individual needs to be convinced that there is value and significance in accepting the idea before they are willing to spread the idea. The amount of convincing required for the individual to accept an idea would depend on the characteristics of the individual as well as the nature of the diffusion process itself. For simple contagions, the spread occurs very easily between individuals. For the spread of behaviors and information, individuals might require multiple exposures before they will accept the behavior or believe the information enough to spread it.

The diffusion model presented in this paper is motivated by concepts from existing diffusion models. The key concepts found in the SIR models used in epidemiology and the standard threshold and cascade models are reflected in the framework. The diffusion model presented herein is a general framework and these particular models can be incorporated as special cases.

DIFFUSION MODEL

The diffusion process occurs on a network whose nodes represent individual entities and edges represent interactions between nodes. This network may be a directed or undirected. Through the interactions between nodes, some property is diffused through the network. We will refer to the property as messages. Messages are introduced into the network by external sources and propagated through the network as nodes interact. Each message can entail multiple sources and there is a corresponding information value for each source of the message. Since the perceived value of the message at each node may be different for each source, the messages are propagated as a vector of source and value (S,V) pairs.

External sources broadcast messages to a subset of the nodes referred to as the seed set. The messages then spread from the seed set to the rest of the network according to the diffusion process dynamics defined by the model axioms. There is a weight on each edge between two nodes. For the spread of actionable information, the weight on each edge quantifies the social relationship between two nodes based on the notion of trust (Kelton, Fleischmann, & Wallace, 2008). When nodes try to propagate a message to their neighbors, there is a probability associated with each edge determining whether the message will reach the recipient node.

Nodes have configurable attributes and the properties of each node are updated over time as interactions occur and messages are propagated. Nodes can fall in one of several states. Nodes who have not received any messages are initially in an uninformed state. As nodes become exposed to the messages, they may change from one state to another. The node state would depend on their perception of the information value in the message and their trust in the information sources and intermediate propagators.

Axioms

The model describes the diffusion process based on four axioms: Information Loss Axiom, Source Union Axiom, Information Fusion Axiom, and Threshold Utility Axiom. These axioms specify how information is propagated, how the nodes process the information they receive, and how nodes update their properties based on the information they receive.

> *Axiom 1: Information Loss Axiom*
> If (S,V) is a source-value pair at node i which is propagated to node j then the source-value pair at node j is $(S,\alpha(i,j)*V)$, where $0 \leq \alpha(i,j) \leq 1$ is the propagation loss from i to j. $\alpha(i,j)$ quantifies the trust relationship between nodes i and j.

When a message is passed from one node to another, the information value of the message is non-increasing. The information value of a message does not deteriorate over time at the node. The information value is modified only when the message is propagated. The information value of the message at the receiver node is a function of the social relationship between the sender and the receiver and not just a function of distance. The social relationship may be asymmetric, i.e. the trust weights on the edge may be different depending on the direction of the information flow.

> *Axiom 2: Source Union Axiom*
> If multiple nodes propagate a message to a node j, then the set of sources at j after propagation is the union of the set of sources already at j with the union of the set of sources arriving from the multiple nodes.

Each node stores an information set, in the form of source-value pairs, representing the information they have received. The nodes keep track of the originating source of the message and the corresponding information value as perceived by the node. At the end of each time step, the node will merge the information they received using the *Information Fusion Axiom* and update their properties using the *Threshold Utility Axiom*.

> *Axiom 3: Information Fusion Axiom*
> *A*. If a source S_i appears in multiple incoming messages with values V_i^1, V_i^2, … the information from this particular source, V_i^*, is fused into the single source-value pair (S_i , V_i^*), where $\max_k V_i^k \leq V_i^* \leq \max(\Sigma_k V_i^k,1)$. The value V_i^k corresponds to the information value of source i at node k, where $0 \leq V_i^k \leq 1$.
> *B*. Suppose that node k has source set S_1^k, S_2^k, … with information values V_1^k, V_2^k, … The fused information value at the node is at least the $\max_i V_i^k$ and at most $\Sigma_i V_i^k$ or 1. The fused value at node k is computed as $fused_k = \lambda * \Sigma_i V_i^k + (1- \lambda) * \max_i V_i^k$.

There are two components to consider when merging the information. The first part of the *Information Fusion Axiom (A)* is to fuse information from the same source appearing in multiple messages. When the same source is found in multiple messages, the combined information value for the source at the receiver node is at least the maximum of the information values for the source over all the messages and at most the sum of all the information values of the source.

The second part of the *Information Fusion Axiom (B)* outlines how to compute the fused value at a node. To specify how to combine the information values from all the different sources, we can use a weighted convex combination of the sum and maximum of the values according to a parameter λ. When λ = 0, the fused value is equal to the max of the information values. When λ = 1, the fused value is equal to the sum of the information values of all the sources, not to exceed the value of 1. Assuming that the fused value is at least the maximum of the information values suggests that having more information cannot hurt.

> *Axiom 4: Threshold Utility Axiom*
> After computing the fused value, the node state is determined based on whether the fused value exceeds certain thresholds. If the node's fused value exceeds one of the thresholds, the node will enter a new state. There is a defined behavior associated with each node state.

The node has two defined threshold levels, a lower bound and an upper bound. The lower bound determines the boundaries for when the node will acknowledge the message. The upper bound determines the boundaries for when the node will take action, i.e. spread the information. **Table 1** summarizes the possible node states along with its corresponding behaviors in the context of evacuation warnings. The lower bound threshold lies between the Disbelieved and Uninformed states, while the upper bound threshold lies between the Undecided and Believed states.

Table 1. Description of Node States

State	Description	Behaviors and Actions
Uninformed	Node has not received any messages.	No action

Disbelieved	Node has received a message but does not believe the message	No action
Undecided	Node has received the message but is uncertain of what to do	Query neighbors in the network
Believed	Node has received the message and believes the value of the message	Spread the message to its neighbors and leave the network after *x* time steps
Evacuated	Node is no longer in the network	

We assume that all nodes are initially Uninformed and have not received any messages. When nodes fall in the Undecided state, they will engage in information seeking behavior and query their neighbors in the network. The node will attempt to contact each one of their neighbors. If the communication is successful, the neighbor will send their information set to the node. When nodes enter the Believed state, they will attempt to spread the message to its neighbors for a predefined number of time steps. After an *x* number of time steps, the node will leave the network and enter Evacuated state. When the node is in Evacuated state, the node will remove all incoming and outgoing edges and become unreachable by their network neighbors.

The threshold levels, in general, may reflect the utility of the message and the resource requirements or potential risks associated with being in a certain states. The corresponding behaviors can be defined to fit the context. If the utility of the message is high, the lower bound threshold should be relatively low, since the individual node is more likely to acknowledge the message. However, if a state requires resources to be put at risk, then the threshold to enter that state should be higher. For example, if there are high costs and consequences associated with being in a Believed state, the upper bound threshold should be somewhat high. On the other hand, if the utility of the message is low, both thresholds should be relatively low. In this case, individual nodes may be more willing to acknowledge a message or take an action, i.e. propagate the message, since there are low costs and few consequences to the action.

From the perspective of information trust, the threshold levels may reflect the required confidence levels each node must have before they are willing to act on the information. The actions that the nodes will choose to take will depend on trust in the information they receive and the trust in the information source and propagators (Kelton, Fleischmann, & Wallace, 2008). The fused value of the node shows the amount of confidence the user develops as it receives addition information. The confidence describes to the node's willingness to use the information.

EXPERIMENTS

We illustrate the concepts of the model by simulating the spread of evacuation warnings in a social network. The context of the experiments are motivated by the evacuation warnings scenario described in (Hui, Goldberg, Magdon-Ismail, & Wallace, 2008) and preliminary studies can be found in (Hui, Goldberg, Magdon-Ismail, & Wallace, 2010; Hui, Magdon-Ismail, Wallace, & Goldberg, 2009). In the case of evacuations, warnings are broadcasted from information sources to the at-risk population. We assume that the initial broadcasted messages will reach a certain proportion of individuals from the population, referred to as seeds. These seeds will then attempt to propagate the evacuation warning to the rest of the population. Applying the model to the diffusion of warnings captures network dynamics as a result of the diffusion, i.e. receiving the evacuation warning may cause individuals to leave the network and disrupt the flow of information. In a structural sense, nodes are removed from the network and incoming and outgoing edges from the nodes are removed as well.

We simulate the diffusion of evacuation warnings in multiple network structures under various model settings and observe the ultimate proportion of evacuated nodes. **Table 2** summarizes the experimental parameters used in these simulations.

Table 2. Experimental Parameters

Parameter	**Values**
Network structure	Grid, Regular, Scale-free, Random, Random-Group
Seed size	5%, 10%, 20%
Seeding strategy	Random nodes, Highest degree nodes
Network trust	0.45, 0.50, 0.55
Trust scenarios	Equal trust (*A*), Higher trust within group (*B*), Randomly distributed trust (*C*)
Information fusion parameter λ	0.50, 0.75, 1.00

Network Structures

We observe the diffusion process on five different types of network structures of the same size, as summarized in **Table 3**. Each network has 100,000 nodes and similar graph densities. In the grid network, the nodes are arranged in a two-dimensional grid, where most nodes have 4 neighbors. In the random regular graph, all nodes have 4 randomly selected neighbors. The scale-free network is a network whose degree distribution follows a power law. We used the Barabasi-Albert model for generating random scale-free networks using preferential attachment (R. Albert & Barabasi, 1999, 2002), The degree distribution of the resulting graph follows a power law of the form $P(k) \sim k^{-3}$. The random graph is a Erdos-Renyi network where nodes are linked randomly with an edge probability p = 0.00004.

In the random group model, nodes are more likely to be connected to other nodes belonging to the same group than to nodes of a different group. This graph attempts to capture how individuals are more likely to communicate with certain individuals and less frequently with other individuals. Nodes are assigned to *k* groups of size *m* where the total number of nodes is $n=k*m$. The edge probability between nodes from the same group is p_s and the edge probability between nodes from different groups is p_d. Assuming that p_s is greater than p_d., we constructed a group network where $p_s = 2*p_d$.

Table 3. Summary of Network Structures

Network	**Density**
Grid	0.00003987
Regular (d=4)	0.00004000
Random (p=0.00004)	0.00004000
Scale-free (m=2, k=2)	0.00003900
Group (k=2, m=50,000)	0.00003994

Seeding Strategies

We configure five trustworthy sources (*trust_value = 0.90*). Each source will broadcast messages with high information value 0.95. We assume that the initial broadcast will reach all its intended recipients. We consider two strategies for selecting the seed set. One strategy is to randomly select a set of nodes. Another strategy is to select the set of nodes with the highest degree, i.e. most neighbors. Selecting a set of nodes with highest degree roughly corresponds to influential or popular nodes of the network as defined by their degree centrality. We simulate the broadcast of messages with initial seed set sizes of 5%, 10%, and 20% of the node population. The initial seed set is evenly divided across the five sources so that each source is connected to an equal number of nodes in the network.

Network Trust and Trust Scenarios

There is a weight on each interaction edge in the social network representing the trust between the nodes. We construct the following trust scenarios to investigate how differences in trust between nodes affect the diffusion process as well as the degree of the trust differences. We split the population of nodes into two social groups G_1 and G_2 of equal size. The trust values between nodes will be determined based on the sender node and recipient node's social group membership.

There are four types of links showing the direction of information between any two neighboring nodes in the network: (G_1 to G_1), (G_1 to G_2) , (G_2 to G_1) , and (G_2 to G_2). Each link represents the trust between the two nodes when information is transferred from the sender to the recipient.

We introduce the trust values, t_{avg}, t_{low}, and t_{high}. High trust links have value $t_{high} = t_{avg} + \varepsilon$ and low trust links have value $t_{low} = t_{avg} - \varepsilon$. The trust differential ε is the difference from the average trust t_{avg}. The trust scenarios describe how these trust values are assigned to each edge in the network. The average trust of the entire network will be approximately t_{avg} for the simulated networks since the two groups have equal sizes. We simulate the scenarios using t_{avg} equal to 0.45, 0.50, and 0.55.

Scenario A. Equal between all nodes.

This scenario represents a homogeneous network where everyone has the same trust in everyone else. There are no social groups and no differences in trust between nodes, i.e. $\varepsilon = 0$ and $t_{low} = t_{high}$.

Scenario B. Higher trust in nodes from the same group.

This scenario represents a population where people have higher trust in others who are in the same group or similar to them. This is such a case, for example, in the dissemination of hazard information in populations with ethnic groups. Individuals who belong to the same group have t_{high} in each other and have t_{low} in individuals of a different group.

Scenario C. Random trust between all nodes.

There is no structure in how trust is distributed in the network in this scenario. The values t_{high} and t_{low} are randomly assigned onto links throughout the network with probability 0.5. As a result, individuals have higher trust in about half of the population.

Information Fusion Parameter

When a source appears in multiple messages with different information values, the information is joined into a single value by taking the maximum of all the information values. For computing the information fused value at the node, we compute $fused_k = \lambda * \Sigma_i V_i^k + (1- \lambda) * \max_i V_i^k$ with $\lambda = 0.5, 0.75, 1.0$. The fused value is equal to the maximum of the computed value or 1.

Node Attributes

The node thresholds are constructed to reflect the context of evacuation warnings. There are high risks, costs, and consequences associated with evacuation warnings. The assumption is that nodes will be eager to acknowledge the warning, but may perform information seeking before ultimately believing the message and deciding to evacuate the network. The lower bound threshold will be small to demonstrate that nodes are not likely to disbelieve the warning. The difference between the upper bound and lower bound thresholds will be large, i.e. nodes will fall in Undecided state and seek for information. When a node enters Believed state, the node will spread the warning to its neighbors and will evacuate, i.e. leave the network, after 5 time steps. When a node tries to query for information or spread a message, the information is transmitted with probability 0.75.

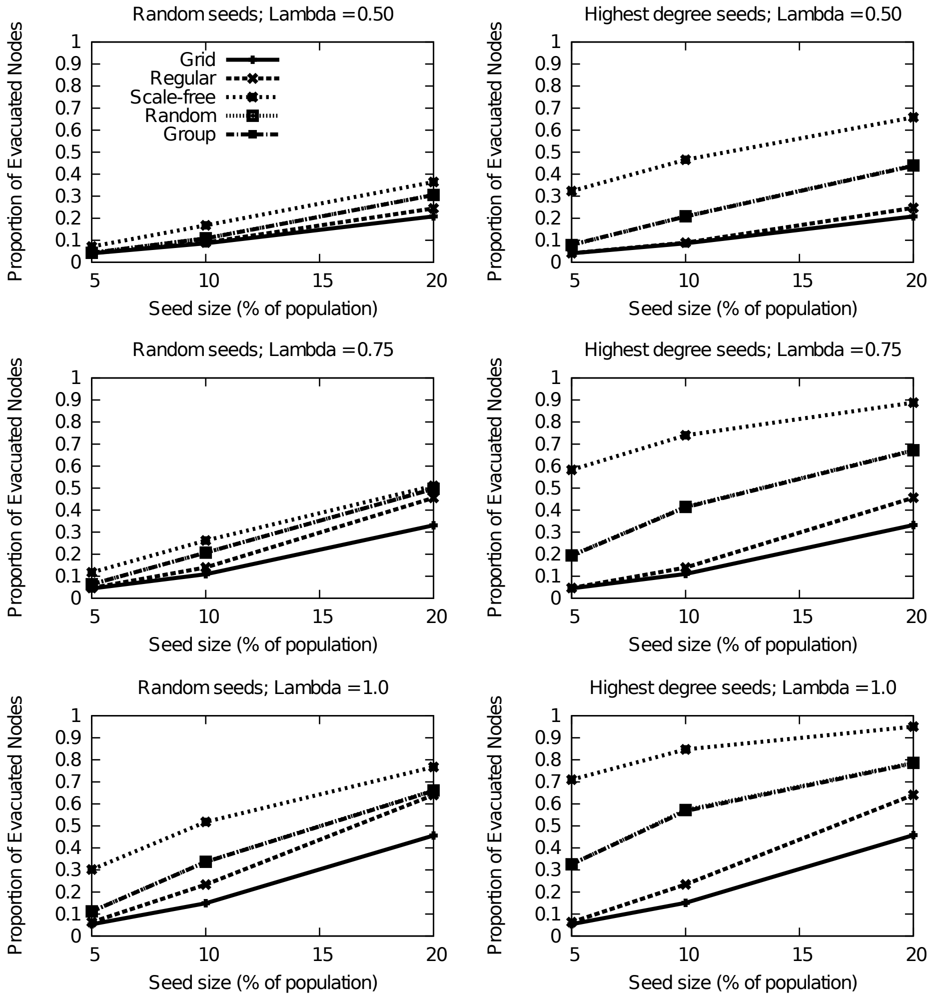
RESULTS AND DISCUSSION

Each simulation run lasts 50 time steps and is repeated 100 times. The diffusion process stabilizes within the 50 times steps where no more node state transitions occur and all Believed nodes have entered Evacuated state. Based on these experiments we report the average proportion of evacuated nodes at the end of the simulations.

Network Structures and Seeding Strategies

The results show that the network structure and the seeding strategy used both have an impact on the proportion of evacuated nodes produced. **Figure 1** shows the proportion of evacuated nodes for each network structure under *Scenario A*. Increasing the size of the seed set led to larger proportions of evacuated nodes. In these experiments, using the highest degree nodes as seeds was more beneficial in spreading the evacuation message than when using a randomly selected set of nodes. The effect of the seeding strategy is also dependent on the network structure. In the grid network and the regular network, there is little difference between the seeding strategies since most nodes share the same degree. Seeding using the highest degree nodes showed improvement in the random network and the random group network. In scale-free networks, seeding using high degree nodes results in a drastic increase in the proportion of evacuated nodes. This is due to the fact that in scale-free networks there is a set of highly connected nodes, “hubs”, whose degree exceeds the average node degree. These “hubs” are essential in the network stays connected enabling objects to flow through. The targeted removal of a few “hubs” can disconnect the network and disrupt the flow. On the other hand, the “hubs” can also good selections for seeds to spread information that would reach a larger proportion of the network.

Figure 1. Proportion of evacuated nodes as we increase the size of the seed set, for information fusion parameters λ = 0.50, 0.75, 1.0 and t_{avg} = 0.50.



Information Fusion Parameter

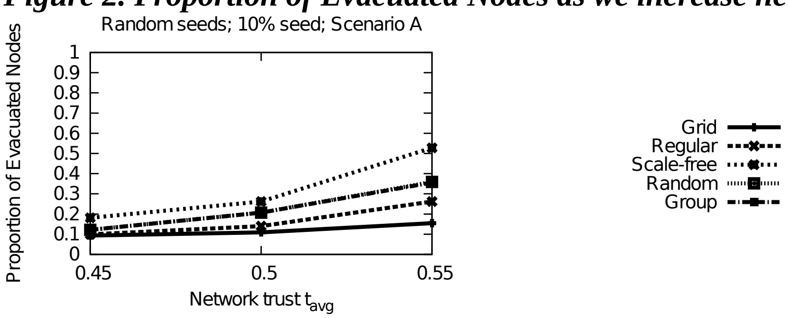
In the diffusion model, the way in which the information is fused at the node is an important parameter, i.e. the selection of the parameter λ. When we perform information fused with λ = 1, the fused value is more likely to exceed the lower and/or upper bound thresholds. As a result, the messages diffuse to a larger proportion of nodes since each recipient of information are more easily to reach an information seeking or spreading state. As we decrease the value of λ the fused value gets closer to the maximum of the individual bits of information. The parameter λ can be interpreted as the node's propensity to trust and could depend on the nature of the diffusion, e.g. how contagious it is. When λ = 0, the node only takes into account the message with the largest perceived value. In this case, the information that has most value is acknowledged and any other smaller pieces of information are discarded. New information is not useful unless it provides greater information value. When we increase λ, any piece of information provides value to the node and how much value provided depends on where λ sits between 0 and 1. In these experiments, treating information from each source independently and adding their information values results in the greatest diffusion, i.e. λ =1.

We observe that λ is interrelated with other model parameters such as node thresholds, number of sources, information value of the initial message, and the network trust. For example, in order for diffusion to occur in a setting where node thresholds are high and network trust is low, λ would have to be large in order to fuse the information values from multiple messages. In an alternate setting, if there were many sources broadcasting high-valued information in a very trustworthy network, the parameter λ may not affect the diffusion process as much. Further investigation will be done to explore the effect of λ in more complex experimental settings.

Network Structures, Trust Scenarios, and Trust Differentials

Network trust refers to the average of all the trust values on every edge in the network. When the network trust is increased, it is expected that having more trust would have a positive effect on the diffusion process. **Figure 2** presents the proportion of evacuated nodes for *Scenario A*, as we increase the network trust from 0.45 to 0.55 for the case where messages are randomly broadcasted from the sources to 10% of the network. For *Scenario A*, the network trust is equivalent to the trust value on each interaction edge in the network, since all edges have the same trust value. Increasing the network trust had the largest effect on the scale-free network. The proportion of evacuated nodes more than doubled when the network trust was increased from 0.45 to 0.55. On the other hand, increasing the network trust had a smaller effect on the proportion of evacuated nodes for the grid and regular networks.

Figure 2. Proportion of Evacuated Nodes as we increase network trust for Scenario A



The results for one of the experimental settings using the two seeding mechanisms are shown in **Table 4** and **Table 5**. The results show that *Scenarios B and C* produced larger proportions of evacuated nodes when compared to *Scenario A*, when there was equal trust among all nodes. This suggests that scenarios with differences in trust are better for diffusion than a uniform trust scenario. In addition, increasing the trust differential from ε = 0.05 to ε = 0.15 resulted in larger proportions of evacuated nodes. This observation suggests that large differences in trust may actually promote the spread of information in situations such as the one described by the experimental parameters.

In the case that ε = 0.15, when a node receives high-valued information through a high trusted edge, the node will be more likely to enter an Undecided or Believed state. The information at the recipient node is a function of the social relationship between the sender and recipient and in this case, the value of the social relationship is high and therefore, the value of the information would not depreciate as much according to the *Information Loss Axiom*. When this occurs, the node will enter the higher state in one time step and seek or spread the message at the next time step. On the other hand, if a node receives information through a low trusted edge, the node will more likely enter a Disbelieved or Undecided state, in which they would wait for information or seek for information the next time step.

Table 4. Proportion of Evacuated Nodes for Random seed, Seed size 20%, t_{avg} = 0.50, λ = 1.0

Network	*Scenario A*	*Scenario B*		*Scenario C*	
	ε = 0	ε = 0.05	ε = 0.15	ε = 0.05	ε = 0.15
Grid	0.457	0.481	0.548	0.488	0.570
Regular	0.640	0.659	0.731	0.658	0.729
Scale-free	0.768	0.798	0.848	0.799	0.848
Random	0.661	0.686	0.756	0.687	0.756
Group	0.659	0.741	0.836	0.685	0.756

Table 5. Proportion of Evacuated Nodes for Degree seed, Seed size 20%, t_{avg} = 0.50, λ = 1.0

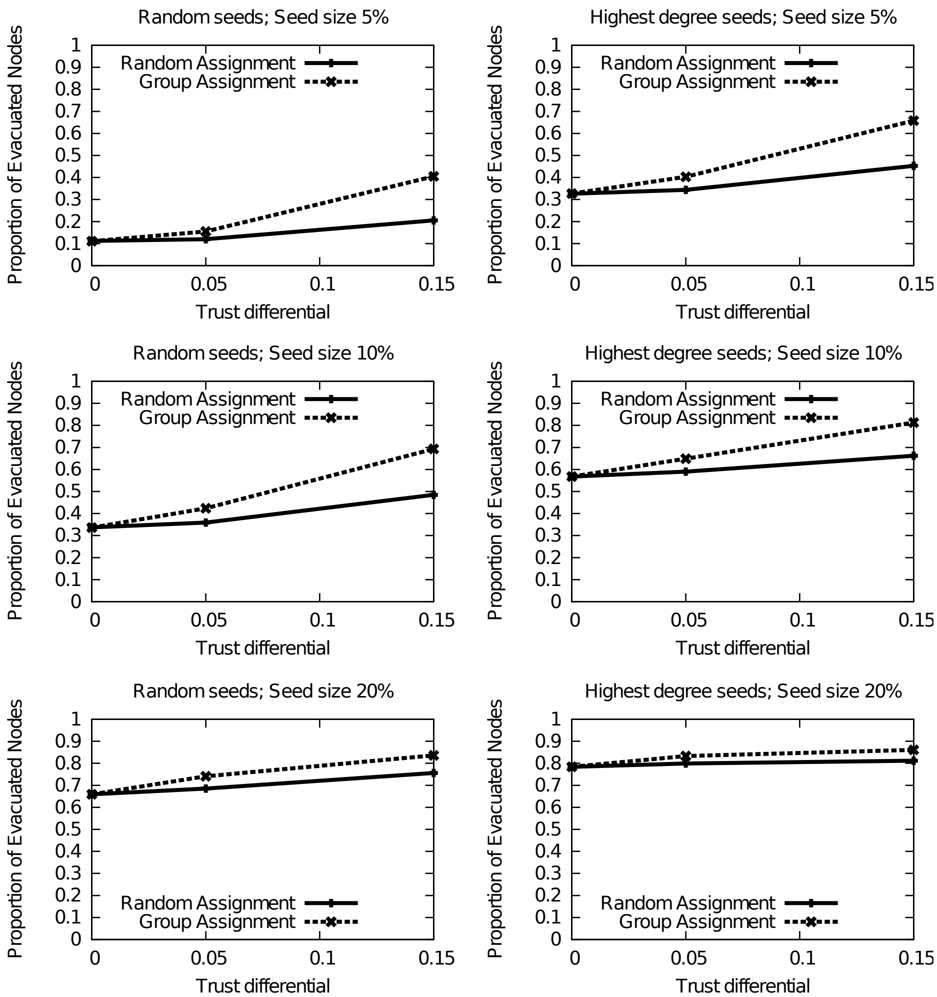
Network	*Scenario A*	*Scenario B*		*Scenario C*	
	ε = 0	ε = 0.05	ε = 0.15	ε = 0.05	ε = 0.15
Grid	0.458	0.482	0.549	0.490	0.571
Regular	0.641	0.659	0.730	0.659	0.729
Scale-free	0.951	0.960	0.948	0.960	0.948
Random	0.787	0.801	0.814	0.801	0.813
Group	0.784	0.833	0.861	0.799	0.812

Network Group Structure and Trust Scenarios

For most of the network structures, each node was randomly assigned to one of the two groups, G_1 and G_2 throughout the network. In a more realistic model, nodes from the same social group are more likely to communicate with each other and form clusters in the network. The random group network tries to captures this element where nodes from the same group are more likely to have edges connected to each other than to nodes from different groups. For the random group network used in these experiments, the nodes from group G_1 are twice as likely to be connected to others from group G_1 than from group G_2, and likewise for nodes from group G_2.

Figure 3 presents simulation results for random node assignment and the group assignment on the random group network. Group Assignment refers to the node assignment that was used for constructing the random group network, i.e. nodes are assigned to groups and the interaction edges were added based on the groups. Random Assignment refers to randomly assigning nodes to groups after the network was constructed. Note that the networks are identical except for the group memberships of the nodes.

Figure 3. Comparison of the proportion of evacuated nodes for Random Assignment and Group Assignment as we increase the trust differential and vary the seed size.



For a small seed size of 5%, Group Assignment appears to have a greater effect on the diffusion process than with Random Assignment. As the seed size increases, the effect is not as drastic. This suggests that in addition to trust differentials, the distribution of the node groups in the network can have a significant effect on the diffusion process. When nodes from the same group are clustered together, they create clusters where information is propagated with high trust. Having this structure is especially important when the initial broadcasts reaches fewer nodes and the network depends on trustworthy edges to propagated the message to the rest of the network.

We observed that in the random group network, there are more edges connecting nodes within groups than between groups. As a result, the amount of trust in the network is actually increased when nodes have higher trust within groups as described for *Scenario B*. Further studies can be done to investigate the impact of group structure and the distribution of trust values on the edges in the network. We can examine how connectivity between groups and within groups, as well as the number of groups and the group size, affects the diffusion process. Furthermore, we can simulate scenarios where groups have access to different information sources and varied trust in information sources.

CONCLUSION

We presented a general model of diffusion in dynamic networks and used the model to simulate the diffusion of evacuation warnings in various network structures. In the case of evacuation warnings, the network is dynamic as a result of the diffusion where nodes may leave the network as they receive warning information. The desired action of the diffusion is for nodes to spread the warning and eventually evacuate. We investigated how network structure, seeding strategy, network trust, and trust distribution affect the diffusion process. The effectiveness of the diffusion depends on the network structure and the seeding strategy used. The simulation results showed that differences in trust between nodes in a network led to a larger proportion of evacuated nodes when compared to equal trust between all nodes. The results provide interesting observations regarding the parameters in the model, such as information fusion, node thresholds, and network trust. Further experimentation includes investigating how the parameters relate to each other in more complex diffusion settings.

One extension to the current model is to enable nodes to go from an infected and contagious state back to an infected and not contagious state. It is possible that a node in Believed state may become Undecided after obtaining new information. Incorporating this aspect to the model would allow us to model more complex scenarios such as impeding the spread of evacuation warnings by diffusing a subsequent abort message and simulating the spread of conflicting information.

Acknowledgements
This material is based upon work partially supported through the Command, Control, and Interoperability Center for Advanced Data Analysis Center of Excellence by a grant from the Department of Homeland Security, Science and Technology Directorate, Office of University Programs, the U.S. National Science Foundation (NSF) under Grant No. IIS-0621303, IIS-0522672,IIS-0324947, CNS-0323324, NSF IIS-0634875 and by the U.S. Office of Naval Research (ONR) Contract N00014-06-1-0466, and the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

Albert, R., & Barabasi, A. L. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512.

Albert, R., & Barabasi, A. L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics, 74*(1), 47-97.

Albert, R., Jeong, H., & Barabasi, A. (2000). Error and attack tolerance of complex networks. *Nature, 406*(6794), 378-382.

Barthelemy, M., Barrat, A., Pastor-Satorras, R., & Vespignani, A. (2004). Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters, 92*(17), 178701.

Bass, F. (2004). A new product growth for model consumer durables. *Management Science, 50*(Supplement 12), 1825-1832.

Berger-Wolf, T. Y., & Saia, J. (2006, August). *A framework for analysis of dynamic social networks.* In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, Pennsylvania.

Bettencourt, L., Cintron-Arias, A., Kaiser, D., & Chavez, C. (2006). The power of a good idea: Quantitative Modeling of the spread of ideas from Epidemiological models. *Physica D, 364*, 513-536.

Brown, J., & Reignen, P. (1987). Social ties and word-of-mouth referral behaviour. *Journal of Consumer Research, 14*(3), 350-362.

Centola, D., Eguiluz, V. M., & Macy, M. W. (2007). Cascade dynamics of complex propagation. *Physica A, 374*, 449-456.

Chen, L., & Carley, K. (2004). The impact of countermeasure propagation on the prevalence of computer viruses. *IEEE Transactions on Systems, Man, and Cybernetics, 32*(2), 823-833.

Cortes, C., Pregibon, D., & Volinsky, C. (2003). *Computational methods for dynamic graphs*. Florham Park, New Jersey: AT&T Shannon Labs.

Delre, S. A., Jager, W., & Janssen, M. A. (2006). Diffusion Dynamics in Small-World Networks with Heterogeneous Consumers. *Computational & Mathematical Organizational Theory, 13*(2), 185-202.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821-7826.

Goldberg, M., Kelley, S., Magdon-Ismail, M., Mertsalov, K., & Wallace, W. A. (2008, August). *Communication Dynamics of Blog Networks.* In Proceedings of the SIGKDD Workshop on Social Network Mining and Analysis.

Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Marketing Letters, 12*(3), 211-223.

Granovetter, M. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology, 83*(6), 1420-1443.

Hill, S., Provost, F., & Volinsky, C. (2006). Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science, 21*(2), 256-276.

Hui, C., Goldberg, M., Magdon-Ismail, M., & Wallace, W. A. (2008). *Micro-Simulation of Diffusion of Warnings.* In Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management ISCRAM2008.

Hui, C., Goldberg, M., Magdon-Ismail, M., & Wallace, W. A. (2010, April 12-14). *Agent-based Simulation of the Diffusion of Warnings.* In Agent-Directed Simulation Symposium (ADS'10), as part of the 2010 Spring Simulation Multi-conference (SpringSim'10), Orlando, FL.

Hui, C., Magdon-Ismail, M., Wallace, W. A., & Goldberg, M. (2009, April 30- May 2). *The Impact of Changes in Network Structure on the Diffusion of Warnings.* In Proceedings of the Workshop on Analysis of Dynamic Networks at the SIAM International Conference on Data Mining, Sparks, NV.

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of American Society for Information Science and Technology, 59*(3), 363--374.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). *Maximizing the spread of influence through a social network.* In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC.

Kleinberg, J. (2000). *The small-world phenomenon: An algorithmic perspective*. Paper presented at the 32nd ACM Symp. on Theory of Computing.
Kuperman, M., & Abramson, G. (2001). Small world effect in an epidemiological model. *Phys. Rev. Lett., 86*(13), 2909-2912.
Lahiri, M., Maiya, A., Sulo, R., Habiba, & Berger-Wolf, T. Y. (2008, December). *The Impact of Structural Changes on Predictions of Diffusion in Networks.* In Proceedings of the 2008 IEEE international Conference on Data Mining ICDM Workshop on Analysis of Dynamic Networks, Pisa, Italy.
Leskovec, J., Adamic, L. A., & Huberman, B. (2006). *The dynamics of viral marketing.* In Proceedings of the 7th ACM Conference on Electronic Commerce (EC06), New York, NY, USA.
Leskovec, J., Singh, A., & Kleinberg, J. (2006). *Patterns of Influence in a Recommendation Network.* In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD).
Meyers, L. A., Newman, M., & Pourbohloul, B. (2006). Predicting epidemics on directed contact networks. *Journal of Theoretical Biology, 240*, 400-418.
Moore, C., & Newman, M. E. J. (2000). Epidemics and percolation in small-world networks. *Physical Review E, 61*(5), 5678-5682.
Morris, S. (2000). Contagion. *Review of Economic Studies, 67*(1), 57-78.
Newman, M. (2000). Models of the small world. *Journal of Statistical Physics, 101*(3/4), 819--841.
Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical Review E, 66*(1).
Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 404-409.
Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic Spreading in Scale-Free Networks. *Physical Review Letters, 86*(14), 3200-3203.
Rogers, E. (1995). *Diffusion of Innovations*. New York, NY: Free Press.
Song, X., Tseng, B. L., Lin, C.-Y., & Sun, M.-T. (2006, Aug). *Personalized Recommendation Driven by Information Flow.* In Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval
Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*, 268-276.
Valente, T. (1995). *Network Models of the Diffusion of Innovations*: Hampton Press.
Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology, 105*(2), 493-527.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393*, 440-442.
Young, H. P. (2000). *The Diffusion of Innovations in Social Networks* (Economics Working Paper Archive): Johns Hopkins University.