Modeling Spread of Ideas in Online Social Networks

Muhammad A Ahmad[1] **Ankur Teredesai**[2]

[1]Department of Computer Science and Engineering,
University of Minnesota,
Minnesota, MN 55455, USA
[2]Institute of Technology,
University of Washington,
Tacoma, WA 98402, USA
Email: mahmad@cs.umn.edu, ankurt@u.washington.edu

Abstract

Internet based online social networks collectively facilitate the spread of ideas. Hence, to understand how social networks evolve as a function of time, it is critical to learn the relationship between the information dissemination pathways or flows and the type of ideas being disseminated. We first classify the spread of ideas into two types based on their rate and nature of proliferation; fads and non-fads. A 'fad' refers to an idea that quickly becomes popular in a culture, remains popular for a brief period, and then loses popularity dramatically. We then model the information dissemination pathways for both these types of ideas. Our results indicate that the proliferation of information in a network strongly correlates with the the type of idea, the degree of participation of the nodes, and a node's availability *i.e.,* presence. Further we derived that after reaching a certain saturation point, a fad exhibits periodic spreading behavior implying that a fad rarely completely disappears from a network. We use data from an instant messaging network community to verify the proposed theoretical modeling framework.

Keywords: Simulation, Social Networks, Information Dissemination, Instant Messaging, Memes, Social Network Analysis.

1 Introduction

In the real world, information disseminates because of information asymmetry. In economics, this information asymmetry as discussed by Arrow (1963) is referred to as the situation when one party to a transaction has more or better information than the other party. Hence, when ideas spread, and information disseminates, information asymmetry decreases. Moreover, information asymmetry has recently been noted to be on the decline thanks to the Internet. The Internet facilitates users that are unknowlegeable to acquire heretofore unavailable information very rapidly. Think of how easy it is to get costs of competing insurance policies, various dealer's quotes for the same car, etc. (Levitt S., & Dubner, S., 2005). Thus, information dissemination or the motivation to spread ideas is the central doctrine behind online social networking.

Existing efforts that model information dissemination assume that the underlying social networks are static *i.e.,* the topology of the network remains constant. In a dynamic social network the edges between the nodes are not fixed and can change over the course of time. Thus, the topology reflects the participation of the nodes in the network. The focus of this paper is to develop a novel framework that attempts to explain the phenomenon of information dissemination in dynamic online social networks. Instant Messaging (IM) networks are examples of such dynamic online social networking environments. It is a very popular way of computer-based communication. The task of studying information dissemination in dynamic networks has gained even more importance from a cybersecurity perspective, since it can used to study information flows in terrorist networks. A feature that terrorist organizations share with highly dynamic networks like IM networks is that the unavailability (removal) of several actors does not make much difference in spreading information. For terrorist networks, this is so because the network is structured to minimise the loss of utility and sustain the network inspite of removal of a few actors (Erickson 1981). The information pathways evolve and adapt as a response to anti-terrorist activities (Sageman 2004) with the consequence that the topology of the network does not remain constant.

There have been few prior attempts to propose appropriate modeling solutions for information flow prediction problems in dynamic social networks. Domingos et al mine the network value of customers using a Markov chaining process (Domingos, P., & Richardson, M. 2001). Kempe et al (2003) provide a theoretical perspective to study the spread of influence through a social network. A key difference and advancement in our efforts being that our methods are not constrained by the availability of connectionist information. *i.e.,* we do not presume that the network topology is a known parameter of the modeling algorithm. In fact, we compare the results of our algorithm for a connections-known network and a network where the nodes can randomly become active or passive (real-network). For theoretical grounding, we borrow the notions of susceptibility and transmissibility from epidemiology; and adapt them to our problem domain. We use a real world dataset, consisting of IM status logs, and IM user behavior data to calibrate our theoretical models. We should also emphasize that the current study is exploratory in nature and thus the goal is to find general trends in proliferation of information in social networks instead of making particular (and speculative) predictions.

To provide some background, IM can be defined

Ankur M. Teredesai is a faculty member on leave from Rochester Institute of Technology, Rochester, NY.

as a communications service that enables its users to create a kind of private chat room with another individual in real time over the Internet if both of them are using the same service. Users can initiate chat sessions with other people in their buddylist. IM technology lets users communicate across networks, in remote areas, and in a highly pervasive and ubiquitous manner. Industrial and governmental organizations are very interested in understanding the nature of broad knowledge-sharing networks that exist within their organizations. IM communication is fast becoming a standard platform for such networks (Teredesai, A., Resig, J., Dawara, S., & Homan, C., 2004). To preserve the privacy of IM users, neither the connectionist data consisting of the user's buddylist nor the content of the chat-sessions was collected. Consequently, any modeling has been restricted to user status logs which are made publicly available as traffic data through open standards. Also, this anonymized IM network status-log data is publicly available (imscanwebsite 2006).

This article is further divided into the following sections. Section 2 talks about related work in Social Network Analysis, section 3 provides the theoretical background and introduces key concepts. Section 4 gives the formalization for the general model, section 5 introduces a variant of the current model that can be used to study fads. Finally experiments and conclusion are covered in sections 6 and 7 respectively.

2 Related Work

The field of social network analysis has been gaining tremendous importance in recent years. The scale-free nature of the link distribution in the World-Wide Web indicates that collective phenomena play a previously unsuspected role in the development of the web. While earlier formalization and modeling was on random graphs, Albert et al note that we need to look beyond the traditional random graph models to gain a better understanding of the web's topology in order to design effective strategies for making information widely accessible (Albert, R., Jeong, E., & Barabasi, A-L. 1999). One can expand the question and ask, Is the same true of large scale social networks such as IM networks? Thus, the question of information dissemination is strongly linked to the study of changes in network topology and vice-versa.

Prior to our attempt, Kautz et al attempted to model the dissemination of information in social networks using collaborative filtering techniques and adopted a simulation approach to validate the results (Kautz, H., Selman, B.& Shah, M., 1997). Our current work is along a similar vein, although the setting and the nature of our network puts different constraints in modeling and formalization. Another interesting area where our current research can be extended is to address the problem of discovering connection subgraphs as networks evolve over time. Previously, Faloutsos et al. (2004) provided a fast algorithm for discovery of connection subgraphs for static networks. Given the discrete nature of IM Networks, Faloutsos's fast algorithm can be implemented at each time-step iteratively to determine the change in the connection subgraphs. This can lead to discovery of change points or prominent events within the social network evolution. Our network model is based upon the work by Moore and Newman (2000) which was itself built upon a previous work by Newman and Watts (1998). Watts and Strogatz (1998) developed a small-world model where any two individuals in the network have only a small degree of separation between them. Moore and Newman used a variant of this model to study disease transmission in small-world networks. In this paper we use the Moore and Newmann's model as a basis for our models and experiments.

Karp et al. studied epidemic algorithms for the lazy transmission of updates to distributed copies of a database (Karp, R., Schindelhauer, C., Shenker, S. & Vocking, B., 2000). Kempe et al. explored gossip protocols to study the dynamic behavior of a network in which information is changing continuously over time (Kempe, D., Kleinberg, J., & Demers, A., 2003). These attempts are noteworthy but the scope of our paper is different since we are more interested in models for information dissemination.

3 Theoretical Background

Watts and Strogatz (1998) developed a small-world model where any two individuals in the networks only have a small degree of separation between them. Moore and Newman used a variant of this model to study disease transmission in small-world networks (2000). In this paper we use a variant of their model as a basis for our modeling. Moore and Newman's network can be constructed as follows: Consider a k dimensional lattice. For simplification consider k=1. Each vertex in the network is connected to all its neighboring sites. Any two different are then randomly connected with an edge. The process is continued until a small world network is obtained. The proof that this network is a small world network is beyond the scope of this paper and is discussed by Moore and Newman.

As stated earlier a major difference between an IM network and traditional networks is that in the IM networks any arbitrarily chosen node in the network cannot be always be assumed to be an active participant in the network. This is so because the person corresponding to the node may go offline, thus effectively severing all links with its neighbors for the time period that she is offline. In the case of the spread of infectious diseases this is akin to an infected person being physically absent and thus leaving the social network and then rejoining later. In IM networks this situation happens fast enough and often enough to be noticeable even in small intervals of time. In our network this situation is reflected by IM status changes. The following concepts will be useful in studying information dissemination in dynamic online social networks.

Susceptibility: We define susceptibility σ as the probability that an individual exposed to a meme will be infected by the meme and will herself become a 'carrier'. Given any subject or topic, people with different backgrounds are likely to have varying opinions on the subject. A person usually adopts a meme if it conforms with the world view that the person already holds. It is also possible that memes can change its state during transmission but in the current scenario we assume that the memes are fixed. In the current context, a higher probability for σ denotes that the new piece of information conforms well with views a person already hold. On the other hand a low probability would imply the opposite. To reflect this situation each node in our network is assigned a random value between 0 and 1 for, susceptibility σ.

Transmissibility: We define Transmissibility τ as the probability that whenever two nodes A and B come into contact, and A is already infected, then the meme will be transmitted to B.

In an ideal network (there is only one status *i.e.,* the transmissibility will be perfect for all individuals since everyone will want to share the information with all their neighbors. However, in our model which more closely mimics a real IM network, transmissibility depends upon the agent being active or inactive.

Status: The status s_t of an IM user at time t is an element of the set $\{s_{online}, s_{idle}, s_{away}, s_{offline}\}$ which can be mapped to active and inactive nodes $\{1, 0\}$.

Origin: The node from which the meme originates will be referred to as the origin.

Active Node: A node at time t is said to be active if the corresponding user is online.

3.1 IM Network as a Graph

Consider a network of IM users represented by V nodes, let the links between the IM users be represented by the edge-set E and let A be the set of users who have adopted the meme.

At the beginning of the experiment one of the nodes, say v_B is chosen at random and is flagged as a carrier. v_B will henceforth be referred to as the origin. The node v_B is then placed in the set A and all its neighbors are selected and the meme is said to be transmitted from v_B to a neighbor v_C if the joint probability $p(C \cap B)$, described in section 4, is greater than a predefined threshold T. The joint probability is given by equation (1):

$$p(C \cap B) = p(C|B).p(C) \quad (1)$$
$$p(C \cap B) = p(B).p(C) \quad (2)$$

For the next iteration and all subsequent iterations the elements of set A are selected and an attempt is made to transmit the meme to their respective neighbors until the network gets saturated with the meme or all subsequent transactions result in a probability that is less than or equal to the threshold.

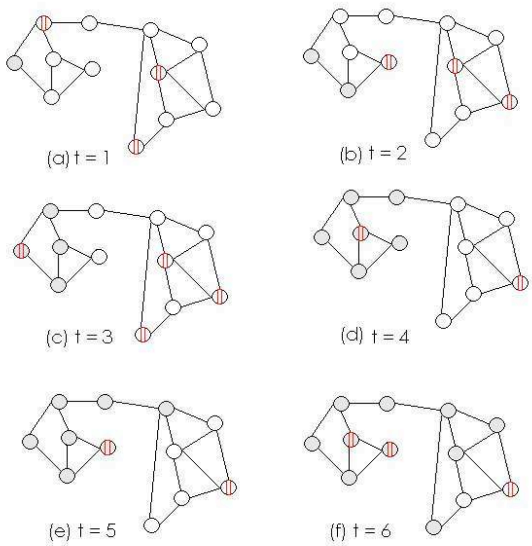

Figure 1: Proliferation of a meme in the network over four iterations. The gray nodes represented are flagged nodes while the nodes with red bars are inactive.

4 The General Information Dissemination Model

Now we consider the question that at time t what would be the number of nodes reachable from the origin? In other words, how many nodes will be infected at time t. This quantity can be termed as **reachability**. The set of such agents would form a subgraph R. Consider the ideal case where transmissibility and susceptibility are perfect *i.e,* equal to one. In this case the number of iterations required to saturate the the network with the meme would be equal to the eccentricity of the origin (initially flagged node). This is so because after t iterations, all the nodes with distance $\leq t$ will be reachable in an ideal network and eccentricity just defines the maximum distance between the origin and any other node in the network. In the best case scenario, where eccentricity is equal to radious, reachability would be equal to the radius of the network while in the worst case it would be equal to the diameter of the network when eccentricity would be equal to the diameter. In the ideal network, at time t, the set R would be equal to A since the probability for the transmission of the meme would be either equal (when $T = 1$) to or greater than the threshold. For the ideal case, after t iterations, A_t can be given as:

$$A_t = \{\ v_t\ :\ v_t\ \in\ V,\ d(v_t, v_0) \leq t\} \quad (3)$$

Let η_i be the set of neighbors of $v_i \in A_t$ such that $\eta_i = \{x : x \notin A_t,\ d(x, v_i) = 1\}$ then A_t for the ideal case can also be given as:

$$A_t = A_{t-1} \cup \left(\bigcap_{i=1}^{k} \eta_i \right) \quad (4)$$

Let us now consider the generalized case where the susceptibility is different for each agent and the transmissibility changes with time. In the case of the IM network the transmissibility is given by the status of the user. Let the paths between node A and B be represented by $\rho(A, B)$ then for the generalized case, if i is the iteration then A_t can be given as:

$$A_t = \left\{ \begin{array}{c} v_t\ :\ v_t\ \in\ V,\ d(v_t, v_0) \leq t, \\ \wedge v_t \in \rho(v_0, v_i),\ T_i \geq T \end{array} \right\} \quad (5)$$

Equation 4 morphologically remains the same with the difference that $\eta_i = \{x : x \notin A_t,\ d(x, v_i) = 1\, s_{(i-1)t} = 1\}$ Consequently the set A_t of users who have adopted the meme at time t will be different for the ideal case and the IM case, specifically: $A_{IM} \subseteq A_{IDEAL}$.

Now we consider the question that given a vertex v_t and the origin, what is the probability that after t iterations v_t will also be infected? To avoid cumbersome notation we introduce the notation α_{it} such that at time t, α_{it} is 1 if $v_{it} \in A_t$, and 0 otherwise. Hence α_{it} represents all the nodes that are in A_t at time t. Since the transmissibility τ and the susceptibility σ are independent of one another, the probability a_{i_t} that at time t node v_i will be infected with the meme can be given as:

$$a_{i_t} = \tau_{i_t}.\sigma_{i_t} \quad (6)$$

The above equation is however an oversimplification since it does not take into account the state of the neighbors of the node or even the rest of the network. This can be remedied as follows. Given any two nodes v_i and v_j, the probability of transmission or infection from v_i and v_j is a conditional probability as defined above and also depends upon the set of paths between v_i and v_j. First consider the simplest case where v_0 and v_1 are neighbors, then the probability that at time t, v_1 has been infected can be determined as:

$$p_{1_t} \ = \ max(\alpha_{1_t}, a_{1_t}.a_{0_t}) \quad (7)$$

Now consider the more complex case where two paths from v_2 to v_0 via v_1 and also via v_3 and there are four nodes v_0, v_1, v_2, v_3 then:

$$p_{2_t} = max(\alpha_{2_t}, a_{2_t}.a_{1_t}.a_{0_t}, a_{3_t}.a_{2_t}.a_{0_t}) \quad (8)$$

Now consider the generalized case where there are n paths from any vertex to any other node, then the probability is given by:

$$p_{n_t} = max\left\{ \begin{array}{c} \alpha_{n_t}, \prod_{i=1}^{n} \rho_{1i_t}, \prod_{i=1}^{n} \rho_{2i_t}, \\ \ldots, \prod_{i=1}^{n} \rho_{(n-1)i_t}, \prod_{i=1}^{n} \rho_{ni_t} \end{array} \right\} \quad (9)$$

Equation 9 gives the probability that after t iteration the node v_n will also be a infected. Notice that in case of the IM Network many of the path are invalidated if one or more corresponding users of the network in the path ρ_i are offline when the meme is supposed to reach them. Now we consider the question, how does the connectivity of the network affect the acceptance of the meme in the network? First we enunciate the following definitions:

Articulation Node: is a node whose removal from a graph disconnects the graph.

Separate: A set of nodes (or edges) of a graph G is said to separate two nodes u and v of G if the removal of these elements from G produces a graph that lie in different components.

Menger's Theorem: Let u and v be non-adjacent nodes in a graph G, then the minimum number of nodes that separate u and v (Let M be the set of such nodes) is equal to the maximum number of disjoint paths in G.(Weisstein 2006)

Consider a graph which is defined by the nodes and the corresponding edges that are reachable from the origin v_0 at time t, for the IM Network. After each iteration, the number of cliques or components can increase, decrease or remain constant depending the number of agents who are offline. Alternatively the topology of the network can even become like a disconnected graph if many users go offline. Hence the number of nodes that are reachable from the origin v_0 may change since there could be multiple paths of length t connecting the origin and other nodes. Additionally the user who goes offline is represented by a cut node or a set of such nodes S which could separate many other nodes and thus effectively changing the number of nodes that are reachable from R.

In the context of the IM network, Menger's Theorem implies that the number of disjoint paths between different nodes change as the topology of the network changes. Consequently the probability of transfer directly depends upon the minimum size of set M. This can be illustrated by considering the case where the size of M is small and the IM users corresponding to the elements of M go offline before they are infected by the meme. Thus all the nodes from the uninfected part of the graph that are connected by these nodes to the infected side are effectively disconnected from the infected side of the graph. Even if one has a large graph one can still end up in a similar situation. This can be illustrated as follows: Consider a nodes v_i and suppose that S exists, then for$x : x \in M,\ x \notin\ A_t\ \forall\ t\ \geq\ d(v_0, v_i)$ we get $p_it = 0$. Hence having a highly connected node (hub) in the u-v path is not sufficient to ensure that most of the nodes are infected in a timely manner. This leads to the conclusion that the time at which the hubs go offline or comes online is equally crucial.

5 The Fads Model

One of the assumptions made in our model is that any node that is successfully infected by a meme will always remain infected. If this assumption is dropped then the afore described model can be modified to study "fads" as well. A fad or a "craze" can be defined as a meme such that the meme no longer has a binary value of 1 or 0 but can take a whole range of values. However the value associated with a fad decreases over the course of time unless the node is re-exposed to it. Two variants of the Fad model are considered: The first case is the one in which just as in the original model a node can infect all of its neighbors. In the second case the chance of a fad spreading decreases if it an node is reexposed to it. Additionally the assumption is made that the decay rate of the fads decrease even when the node is inactive.

When a node is exposed to a fad, the value of the fad is initially set to 1. However the passage of each iteration in which the node is not reexposed to the fad causes the value associated with the fad to decrease by a specified decay rate. It should be noted that a decay rate of 0 is equivalent to the original information dissemination model. If the fad's value for a particular node drops to 0, that node will stop exposing others to the fad. Additionally fads decay even when the user is not online. If d is the decay rate then the value of a fad after i iterations can be given as:

$$f = max(1 - d.(i - k), 0),\ 0 \leq\ k\ \leq\ i \quad (10)$$

where k is the last iteration when the node is reexposed to the meme. The two versions of the model are described below:

Fixed Transmission Case: In the fixed transmission case, the value of a fad does not affect its transmissability. The assumption from the original model that an infected node will always expose each other node it comes in contact with to the fad is maintained.

Weakening Transmission Case: In the weakening transmission case, a node is no longer guaranteed to expose every node with which it communicates to the fad. Transmission is instead a random event whose probability is equal to the fad's value in the currently infected node. Each neighbor node of an infected node has an equal chance of getting infected at each iteration.

6 Experiments

We conducted a series of simulation-based experiments to recreate the process of evolution of a social network *i.e.,* how information disseminates in such networks and how fads get proliferated in such networks.

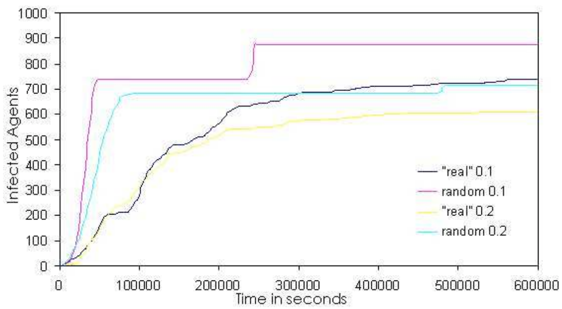

Figure 2: The average results for an actual network for the actual data and random data for a small world network.

6.1 Dataset Overview

The IM status-log data was used as a base for determining the transmissibility of the nodes. Thus an

IM user was randomly assigned to a node and the status of the IM user was used for driving the simulation at each interval. The annonymized dataset used in these experiments was collected by Teredesai et al.(2004) Each user is given a unique identifier in place of their actual screen name.

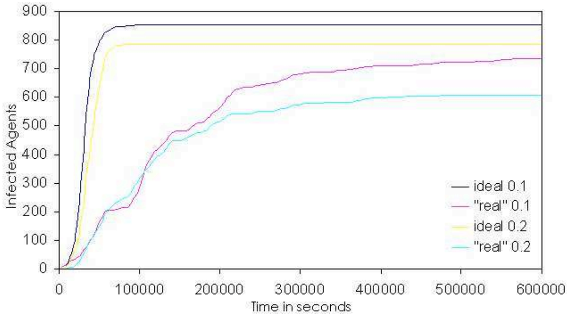

Figure 3: The average results for an ideal and a actual network for a small world network.

6.2 Results for the General Model

In order to compare IM stats in controlled data vs random behavior of agents we conducted additional experiments. A few comparisons of these experiments are described in figure 3 which gives the average results for 100 runs for the ideal network and the actual network for a small world network. It is clear from the figure that the ideal network quickly reaches the saturation point while it takes longer for a actual network to get saturated and the infection is slower. This should not come as a surprise since the situation is analogous to what happens in real life *i.e.,* even though the graph indicates that there is a tipping point, the idea is not readily adopted by everyone in the network. Additional experiments were performed to contrast the results for IM status driven data vs. randomly generated data to see if the results obtained from the previous experiments were not artifacts of the experimental arrangement. Some of the representative results of these experiments are given in Figure 2.

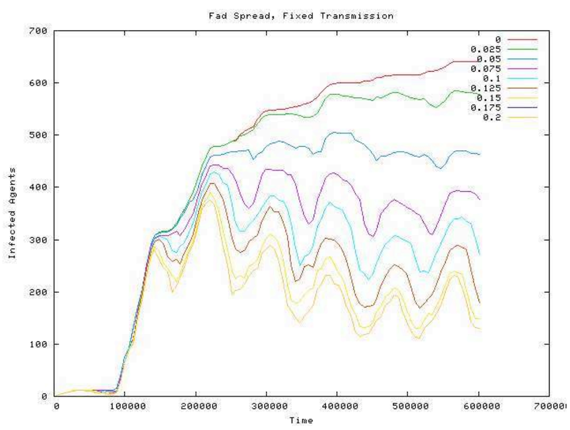

Figure 4: Fixed Transmission Results, varying decay rates.

In is interesting to note that the simulation with randomly generated status data converges earlier as compared to the actual IM Status data. Given this behavior, it is conjected that while the offline/online usage patterns exhibited by IM users are somewhat

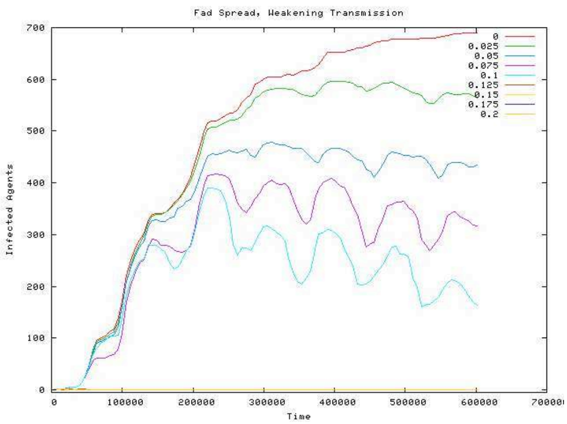

Figure 5: Weakening Transmission Results, varying decay rates.

fixed that leads towards an early saturation, the randomly generated allows those nodes to infect others that would otherwise be offline most of the time. The results also indicate that the status of the IM users is not completely random.

6.3 Results for the Fad Model

A dataset of 996 randomly chosen IM users was used for this series of experiments. Representative results from these experiments and some observations are presented below.

6.3.1 Fixed Transmission Case

The fixed transmission case shows that the count of infected nodes tends to level off at a position lower than the ideal. The leveling-off point comes earlier as the decay rate increases. Higher decay rates also begin to produce sinusoidal-looking curves in the infection count. It could be conjectured that this phenomenon could be mimicking the daily cycle of users going offline during the night. The conjecture is in line with the intuition that fads go in and out of vogue not only depending upon how actively interested people are interested in them but also how much cost they can incur. With very high decay rates, we see that the average infection count begins to slowly drop off after an initial peak. The dataset does not, however, extend far enough in time to allow us to determine if this is another cyclical variation or a permanent loss.

6.3.2 Weakening Transmission Case

It is observed that in the weakening transmission model the network behaves similarly to the original information dissemination model for low decay rates. However the final count of infected nodes is less as compared to the original model if the infection rates are set high. For very high values the initially infected node does not even maintain the fad long enough to infect many of its neighbors thus causing the total count of infected nodes to rapidly fall to 0.

7 Conclusion

In this paper we presented a formal framework to understand information dissemination of fads and non fads in dynamic online social networks. We developed a basic models based on probabilistic interaction between IM users. It was observed that the dissemination of information in a dynamic network depends upon the level of participation of the nodes in the network. We compared this process for a real world IM Network and an Ideal IM Network where all the users are online most of the time and readily accept an idea when they come across it. It was discovered that not only the connectivity of some of the nodes (hubs) determine how fast the meme is proliferated but also the time-span in which the corresponding person is online or offline. Another important factor in proliferation is the size of the set S as defined in Menger's Theorem. It was also observed that after initial quick proliferation the extent of proliferation of fads is periodical in nature. In conclusion, there is significant scope and need for developing effective models to study the spread of information in online social networks and we formulated two such models for the IM based social networks in this paper.

8 Acknowledgments

We would like to acknowledge Alexander Jarocha-Ernst at the Rochester Institute of Technologyfor suggesting improvements to the fads model and implementing the 'fads' part of the model.

References

Albert, R., Jeong, E., & Barabasi A-L. (1999) Diameter of the World Wide Web Nature 401 130-131,

Carley, K. M., Prietula, Michael J. & Lin, Zhiang (1998) Political Science Convention, April 3-6, Chicago, IL. Design Versus Cognition: The interaction of agent cognition and organizational design on organizational performance. Journal of Artificial Societies and Social Simulation vol. 1, no. 3, http://www.soc.surrey.ac.uk/JASSS/1/3/4.html

Domingos, P., & Richardson, M., (2001) Mining the network value of customers, KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining 57–66 San Francisco, California, http://doi.acm.org/10.1145/502512.502525 ACM Press

Erickson, B., (1981)Secret societies and social structure. Social Forces, 60(1):188210.

Faloutsos C., McCurley K. S., & Tomkins A., (2004) Fast discovery of connection subgraphs, KDD '04: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining, 118–127, Seattle, WA, USA, http://doi.acm.org/10.1145/1014052.1014068,ACM Press.

IMSCAN Website (http://lacuna.rit.edu)

Karp, R., Schindelhauer, C., Shenker, S. & Vocking, B., Randomized (2000) Rumor Spreading. 41st IEEE Symposium on Foundations of Computer Science.

Kautz, H., Selman, B.& Shah, M., (2000) Referral Web: combining social networks and collaborative filtering, Communications of the ACM, v.40 n.3, p.63-65.

Kempe, D., Kleinberg, J., & Tardos, E., (2003) Maximizing the spread of influence through a social network KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.

Kempe, D., Kleinberg, J., & Demers, A.,(2003) Spatial gossip and resource location protocols. Proc. 33rd ACM Symposium on Theory of Computing, 2003 137–146 Washington, D.C. http://doi.acm.org/10.1145/956750.956769 ACM Press

Moore, C., & Newman, M., (2000) Epidemics and Percolation in Small-World Networks, Cristopher , Santa Fe Institute Working Papers 00-01-002

Sageman, M. (2004) Understanding Terror Networks. University of Pennsylvania Press.

Teredesai, A., Resig, J., Dawara, S., & Homan, C., (2004) Extracting Social Networks from Instant Messaging Populations. KDD 2004 Link Discovery Workshop.

Teredesai, A., Resig, J.,(2004) A Framework for Mining Instant Messaging Services. SIAM DM 2004 Workshop on Link Analysis, Counter-Terrorism & Privacy.

Arrow, K. J.(1963) Uncertainty and the Welfare Economics of Medical Care. American Economic Review.

Levitt S., & Dubner, S., (2005) Freakonomics: A Rouge Economist Explores the Hidden Side of Everything. William Morrow/HarperCollins.

Watts, D. J., & Strogatz, S. H., (1998) Collective dynamics ofsmall-world networks, Nature 393, 440442.

Weisstein, E. W., "Menger's n-Arc Theorem." From MathWorld–A Wolfram Web Resource. (http://mathworld.wolfram.com/Mengersn-ArcTheorem.html)

Young, H. P., (2000) The Diffusion of Innovations in Social Networks, Economics Working Paper Archive 437, The Johns Hopkins University,Department of Economics.