A Study of Information Diffusion over a Realistic Social Network Model

Andrea Apolloni, Karthik Channakeshava, Lisa Durbeck, Maleq Khan,
Chris Kuhlman, Bryan Lewis, and Samarth Swarup
Network Dynamics and Simulation Science Laboratory
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, USA
Email: {apolloni, kchannak, ldurbeck, maleq, ckuhlman, blewis, swarup}@vbi.vt.edu

***Abstract*—Sociological models of human behavior can explain population-level phenomena within social systems; computer modeling can simulate a wide variety of scenarios and allow one to pose and test hypotheses about the social system. Here we model and examine the spread of information through personal conversations in a simulated socio-technical network that provides a high degree of realism and a great deal of captured detail. To our knowledge this is the first time information spread via conversation has been modeled against a statistically accurate simulation of people's daily interactions within a specific urban or rural environment, capturing the points in time and space at which two people could converse, and providing a realistic basis for modeling human behavior during face-to-face interaction.**

We use a probabilistic model to decide whether two people will converse about a particular topic based on their similarity and familiarity. Similarity is modeled by matching selected demographic characteristics, while familiarity is modeled by the amount of contact required to convey information. We report our findings on the effects of familiarity and similarity on the spread of information over the social network. We resolve the results by age group, daily activities, time, household income, household size and examine the relative effect of these factors. For informal topics where little familiarity is required, shopping and recreational activities predominate; otherwise, home, work, and school predominate. We find that youths play a significant role in spreading information through a community rapidly, mainly through interactions in schools and recreational activities.

I. Introduction

How is social science aided by mathematical models? Epstein [1] has addressed this question, indicating different reasons to build social models: mathematical models could explain (but not predict) the emergence of collective phenomena, or capture qualitative behaviors of phenomena, or guide data collection by indicating controlling parameters underlying certain phenomena. Despite the possible outcomes and the recent interest in modeling social systems it has been pointed out that in many cases these models lack realism and do not take ground realities into account [2], [3]. This is particularly true in the context of socio-physics where extreme simplifications relegate the application of models and their results to a narrow range of circumstances.

Two directions for improvement have been described regarding social network models and their intricacy. Although the discussion here focuses on socio-physics and opinion formation, these directions for improvement can be extended also in the context of other agent-based models:

Moving beyond static social networks: Despite significant progress in describing network models, many models still consider a static underlying network, or a network changing according to specific update rules that do not change during the simulation. This static network approximation is useful in some cases, such as when we consider pure statistical models, power grid failures, flows of information on communication and transportation infrastructures. What distinguishes these cases is the relative timescale of the dynamical process: the network changes slowly enough relative to the process that the network can be assumed fixed. Such an assumption does not hold for many interesting problems. For time-sensitive information, such as news spreading at school about a sickness, or updates for a popular product, the speed with which the information propagates requires a dynamic social network. In such cases the topology of the network is evolving in time with agents' activities, and is affected by the agents' activities in response to their perceptions about the global situation, as demonstrated in the case of epidemic outbreaks in urban areas [4].

Richer description for agents and their interactions: When describing agents' interactions, we are faced with the problem of compressed dimensions, leading in many cases to an over-simplification, e.g., randomly assigning properties to agents. The interaction between agents is oftentimes reduced to the change of one or two characteristics of an agent, e.g., its opinion and its influence over other agents. More accurate agent-based models would be extended in several directions, considering more characteristics to change, allowing individual variations of interaction (specifying internal states of the agents) and considering the strength of links [3]. A richer model of the individual and interactions between individuals would ideally involve a multidisciplinary approach, collecting data and theories of social interaction, individual behavior, and other relevant phenomena from all fields of inquiry involved, and avoiding over-simplification.

Our Contributions: In this article, we address the first point by introducing a dynamic, interaction-based approach to describe the spread of information through conversations between pairs of individuals within a social network, capturing and utilizing the specific times of the interactions. We incorporate the second point to some extent in describing the interactions between individuals. We concentrate on information spread through conversation because of its importance in many human affairs. One-on-one conversations are examined here as a mechanism by which information is spread through the social network. As such, this phenomenon could shape political opinion, create awareness, alert community members to danger, or impact market behaviors. It also exhibits a resemblance to epidemic processes, allowing us to utilize much of our group's prior work in this area.

The diffusion of information in our study is based on two observations: i) two people are more likely to converse if they are similar to each other, a correspondence that we estimate based on demographic information about each person; ii) two people are more likely to broach a topic given sufficient time together, and talk about certain topics requires greater familiarity between people. These are two basic assumptions in social science and diffusion of innovation theory that allow the distinction between strong and weak ties between people. Since our agent-based model takes into account, in great detail, the mobility and the activities of an agent, the social network changes every minute and possesses characteristics different from theoretical ones. Also, the mechanisms by which we construct the population help to ensure that we are not lacking in detail and accuracy in the models of the individuals themselves or the demographics of the region under study.

In this work, our particular contribution is in the interaction model used, as well as the study and analysis we performed of this interaction model and its outcomes. Our approach can thus be considered as responsive to all the major points of improvement we identified. Numerical simulations for this study is performed using the *EpiFast* simulation engine [5]. EpiFast was developed in our research group to simulate epidemic outbreaks and propagation in large urban or rural contexts, based on community models formed from merging together sets of data from different sources, using a method that provides anonymous, synthetic individuals but a population statistically indistinguishable from the real one.

Paper Outline: In Section II, we summarize prior work on homophilic interaction between people, the strength of links and rumor spreading. Section III describes our methods, including the model of information spread over dynamic social networks via conversations, a description of EpiFast and our modifications of it, and the social network and community that we study. The experiments we conduct, the results, and analysis are presented in Section IV. We conclude and discuss future directions in Section V.

II. Related Work

The study of information spread and propagation of ideas, and influence in a social network has a long history in the social sciences [6]–[8]. With the advent of computers with sufficient storage and computational power, this network diffusion process has become an emerging research area in computer science [9]–[14]. The research questions that can be answered include the understanding of the extent to which people are affected by friends, the extent to which word-of-mouth information and rumors spread, and ultimately how the dynamics of adoption are likely to unfold within the underlying social network.

Rogers, in his book [6], develops a systematic analysis of all the processes related to the diffusion of information (or innovations) in a social network. He distinguishes the particular roles played by agents in the network according to their positions, elucidates the roles of homophily and heterophily, and shows imitation as the principal mechanism for adoption. The theory is supported by a collection of in-field experience.

Nekovee et al. have modeled rumor spreading where the population has been divided in three categories: ignorants, spreaders, and stiflers, using a model similar to an epidemiological one [15]. Contact between an ignorant and a spreader could increase the number of spreaders, and contact between a spreader and a stifler increases the number of stiflers, with some probabilities. They performed a systematic study of threshold properties of the process on different classes of static networks.

One of the attempts to consider time-evolving networks and rumor spreading was the work by Agliari et al. [16]. In their model, the agents are in a square lattice (grid) and interact with each other through their random movements within the region. They showed that, starting from a single informed agent, the time it takes for information to reach the entire population has a power law distribution with respect to the size of the population and the size of the grid, and that the average degree of information depends on the size of the grid.

Richardson and Domingos [13] studied the network diffusion process in the context of viral marketing to determine a cost-effective marketing plan. They present a probabilistic viral marketing model and apply the model on a knowledge-sharing web site, Epinions. Further, they posed a well-defined algorithmic problem: which set of a given number of individuals (vertices) in a social network should be seeded with the product information (or the idea) so that it will be adopted by a large fraction of the network, where the estimated extents to which individuals influence each other are given. Kempe, Kleinberg, and Tardos [10] studied this problem further and showed that it is NP-hard. They also presented a greedy heuristic-based approximation algorithm with performance guarantee 63% of the optimal. Randall et al. [11] studied the feedback effect of social ties and similarity and their interactions using data from two online communities. They also examined prediction of an individual's behavior using similarity and the social network formed based on previous interactions and observed that people encounter each other due to overlap in their interests (similarity). Interestingly, they found the consequences of these encounters can lead to further similar interests that are visible many months later.

Due to the difficulty in collecting data about an individual's social network and activities, most previous work uses less complicated and less realistic social networks than we have used here: online community networks [11], [13], static networks [6], or networks constructed from simple models [16]. We therefore study network diffusion processes over a different sort of network: a large-scale social network modeling a specific locale—its roads, its buildings, its households, its people–and in which temporal information about persons' whereabouts is maintained and used.

III. The model

Likeness, or the homophily/heterophily of individuals and the existence of weak links in a network are all recognized as fundamental factors for diffusion of informations or idea in a social network [17]. Homophily represents gradations of similarity between two individuals; heterophily represent gradations of difference. In general, two kinds of homophilies are distinguished [18]— *status homophily* related to informal or formal social status, and *value homophily* based on beliefs, values and attitudes. Status homophily tends to stratify societies into groups, mostly according to socio-demographics characteristics; value homophily is related to internal status. A well accepted assumption is that homophily enhances communication, making it more effective, and at the same time communication increases homophily, or the degree of similarity. The homophily principle is a well accepted principle within the scientific community, and it has been validated in different surveys.

In order to have diffusion in a network, a certain degree of heterogenity among individuals should be present. In a society where communication is restricted only to self-similar individuals, the spread of information is restricted; consider for example the restrictiveness of the caste system. We can classify the links formed by highly similar agents as strong links, and those formed by dissimilar agents as weak links. Weak links are important because information flows along weak ties. Consider the contrary: information cannot migrate from one group to another without conversations between heterogeneous contacts. Thus, in a highly restricted group everybody shares the same amount of information, and a prime channel for gaining new information is an individual's more dissimilar acquaintances: they usually have information that individual does not already possess. This basic mechanism has been formulated by Granovetter in [19], and classifies the strength of a link based on the frequency it occurs. Weak links are then links that seldom appear in the network but allow information to flow between distant/dissimilar groups. If these links were removed from the network the result would be a disconnected set of cliques.

We consider diffusion of information via a simulation of conversation within a realistic social environment as a proxy to reality. The social environment includes daily activities of each individual: where he/she goes, when, with whom he/she interacts, and for how long. We consider an individual who has been informed, and continues to spread the information in the network. The probability that communication can take place between individuals is biased by the following factors:

- **Homophily degree and social influence:** Each individual is associated with a vector of characteristics which is compared with the neighbor's characteristics. This is reminiscent of the Axelrod model [20]. We consider in this experiment only demographic information. This choice is related to the data in possession and also because values, beliefs and attitudes can be derived from this set of information.
- **Temporal order and spatial proximity:** Only agents who are colocated can communicate. We do not consider the use of cell phone or other communication devices. The temporal order is derived from the individual's daily activities and they continuously interact with others only at activity locations, in the order defined by their activity schedule, at particular times when the activity is being performed, and for durations of the activities. For every edge between individuals, we associate a *liveness* function which reflects the contact time and duration when the edge is active, determined by the activities of the individuals.
- **Duration of contact:** Another important characteristic influencing the effectiveness of communication, is the duration of contact—longer contact durations imply a higher probability for information transmission. Contact duration also quantifies the strength of the link.

In this paper, we assume that once an individual is informed, he/she stays informed for the rest of the simulation. Further, a newly informed individual can spread the information to other uninformed people immediately.

A. Social Network Generation

In this section, we provide a detailed description of the social network we create and use for this study. The steps we use to generate the social network follow:

Step 1: *Creation of a synthetic population*. A synthetic population consisting of individuals and households is generated from census data. Locations are assigned to each household according to the census bloc information and are geolocated using the NavTeq street data. Demographics are assigned to each individual in the population based on household data.

Step 2: *Creation of activities for the individuals*. Using the data from the activity survey, a sequence of activities—a time ordered daily schedule of activities—is associated and classified according to household demographics. Each activity sequence is then associated to a member of the household subject to the member's demographics. Each activity in the sequence is denoted with a start time and duration for the activity, location where the activity is performed and an activity type (work, shop, school etc.). Activity locations are again derived from the survey based on information on distance travelled for particular activities. Each location, based on activity type is assigned within a certain distance from the household location. An individual's demographic and the attractors associated with locations are used to assign a location for the individual's activity.

Step 3: *Creation of the social network*. The social network is created by using the interactions of the synthetic population at the activity locations. We use the activities to determine the occupancy of each location and model *sub-locations* to model the interactions within each location. The number of sub-locations is determined by the occupancy of the location at every unit of time and we assume a certain occupancy for every sub-location (for example, 25 individuals per sub-location). The schedules specify the start and duration of contact for each link that is formed between individuals in each sub-location. Currently, we are assuming that all the individuals within the same sub-location form a clique.

Although, we can create different social networks by varying the number of individuals in each sub-location, this is not the focus of this paper. Since our model considers that interactions occur when individuals are co-located in each sub-location, the synthetic activities cause a side-effect. We have long durations during late evening when all household members can potentially interact for extended durations, sometimes from 11 PM untill 6 AM. This is a result of our inheriting network construction methods that model such things as aerosolized pathogens, which can be transmitted from person to person while they sleep; however, people do not converse while they sleep. Ignoring this discrepancy would create unrealistic information spread occurring at odd times of the day. Therefore, we add a model of sleep intervals for the synthetic population to capture this sleep effect as outlined in [21]. Individuals have particular cyrcadian periods, with different sleep times and duration and they depend on factors such as age, genetics, job, and social factors. Sleeping time and duration are considered as independent characteristics and are functions of age and sex. Links are considered inactive if at least one of the two individuals is sleeping.

B. Modeling Interactions

Based on their activity profiles, if two individuals are present at the same place at any time, then a time-dependent interaction exists between the two. If, however, there is no common location, then no information is passed between the two individuals. Given that two individuals coexist at some time, the probability that an informed individual conveys information to a receptive individual is determined by similarity. Although, the synthetic population is endowed with all the demographic characteristics provided by Census, just three of them are used to quantity the similarity between two individuals: income and size of the household in which the individual resides, and the individual's age. The choice has been made because all of the other demographic characteristics are related to these three as a linear combination, and also because it has been found that reaction to external information (i.e., for health related issues) depends on an individual belonging to these groups [22], [23]. For each pair of interacting individuals, the probability p of information transfer is specified as 1/3 for each parameter that they have in common. The weight of an edge w_{ij} between nodes i and j of the network is the sum of the probabilities contributed from the 3 parameters; i.e., $w_{ij} = \sum_{k=1}^{3} p_k$. For example, if we consider the interaction between a parent and child of the same household, then $w_{ij} = 2/3$. If two individuals share the same group for all three parameters, then $w_{ij} = 1$. Two individuals in different groups for all three parameters never transfer information, regardless of the number of times they interact, and hence $w_{ij} = 0$.

C. The Simulator

To study information spread in the social network created as outlined above, we use a modified version of *EpiFast* [5]. EpiFast is a discrete time parallel simulator and implements the Reed-Frost [24] model of diffusion, with four states corresponding to the epidemiological states of susceptible, exposed, infectious, and removed of the SEIR model. The epidemiological states (used for human epidemics) become *receptive*, *incubating*, *transmitting* and *removed*, respectively, for studying information diffusion. For this study, we only use the receptive and transmitting states; i.e., once a receptive (uninformed) node receives the information, it retains the knowledge and can convey it immediately through all of its interactions for the remainder of the simulation.

In the context of information dissemination, a couple of important aspects need to be considered and required modifications to the original EpiFast implementation. EpiFast is required to incorporate time-dependent edge profiles, resulting in a dynamic network topology. Secondly, receiving information from any source takes a certain amount of time. For example, it takes individuals to be interacting for a certain duration before conveying information. We model this by considering an *interaction threshold* (referred as ndt in the rest of the paper). EpiFast was modified to include this feature for the study. For example, the effect of $ndt = 40$ minutes, is that interactions between informed and uninformed individuals that lasts 40 minutes, do not propagate information along that edge, even with $w_{ij} = 1$.

IV. Experiments and Results

In this section, we outline the details of the experimental evaluations for studying information dissemination in realistic environments. The use of Epifast allows us the flexibility to perform detailed analysis of the rumor spreading process—demographic stratum, time and current activity when an individual is informed. Further, using such simulation studies, we can determine the effect when a certain group(s) are better informed in comparison with others. We follow a top-down approach in describing the results. First, we present the overall spread characteristics. Second, we present the results of the information flow in different groups (according to age, household income, and size). Third, we present the importance of each activity location type on the spread process. Finally, we evaluate a special case where we inform a certain group, youngsters (age 0-18 years) to study the activity locations where they acquire the information. The same kind of study can be conducted for other groups, but this analysis is not presented in this paper.

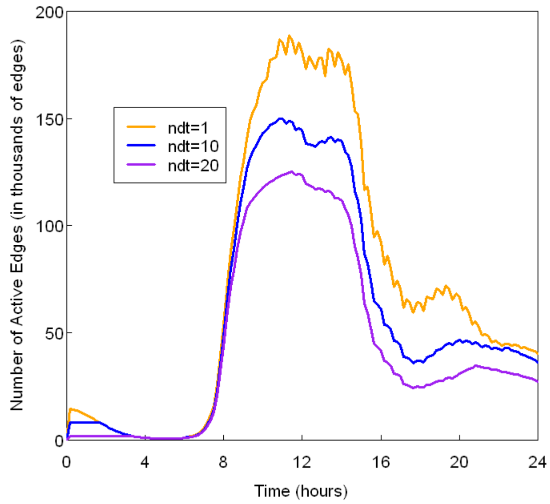


Fig. 1. Number of links created as a function of time during one day, showing three differeint minimum contact duration thresholds, ndt/10.

A. Experiment Setup

Synthetic Population: We chose the population of Montgomery County in Virginia for study. The dataset for the county contains 74,376 individuals. The population consist of almost 70% adults and young adults (age between 19 and 64 years) while younsters constitute 20% of the population. Households mostly consist of couples or couples with just one child. About 66% of the households have annual incomes of at most \$50,000.

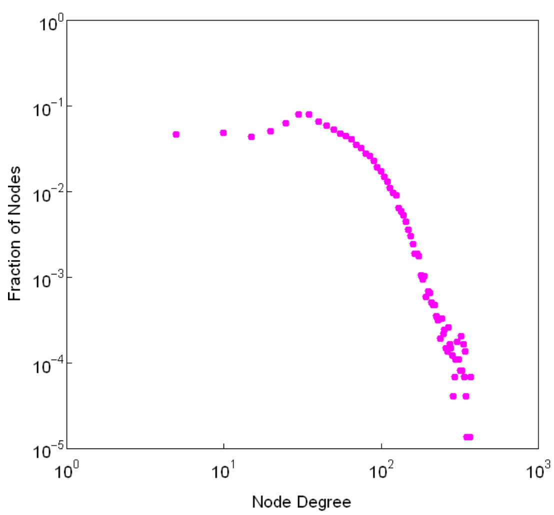


Fig. 2. Cumulative degree distribution (log-log scale). The degree of each node is the number of total contacts during a day.

During a day of activity almost 1,800,000 links are created, during which conversations can occur. Most of them last at most one hour, and almost 80% of the total are created while people are shopping or doing recreational activities. Figure 1 shows the number of links existing at different times during the day. We notice that at the beginning of the day few links are present, while most of the links are active during working and school time (9 A.M. to 3 P.M.). Due to the different activities many links are still alive at midnight. The effect of variation in ndt is a reduction in links from 9 A.M. till midnight, but it does not qualitatively change the behavior. The effect of modeling sleep causes a reduction in the duration of night time links. So, when ndt value is increased, the number of night time links reduce.

TABLE I
MAJOR DEMOGRAPHIC STRATA

Age	%	Hh Income	%	Hh Size	%
0-18	20	0-25 k	33	1	11
18-35	39	25-50 k	30	2-3	52
35-64	32	50-75 k	20	> 4	37
>64	9	>75 k	17		

The cumulative degree distribution is shown in Figure 2. The degree for an individual is the number of co-located people in the sub-location. The cumulative degree clearly follows a power law distribution.

Population Stratification: We divide the population into strata based on the major demographic factors: age, household income and household size. For each factor, we consider a value range to establish the *similarity* between two potential conversants. Table I (Hh refers to *household* throughout the paper) shows the value ranges for each factor and group size as a percentage of the population. We combine the value ranges across all demographic factors to derive 48 different groups.

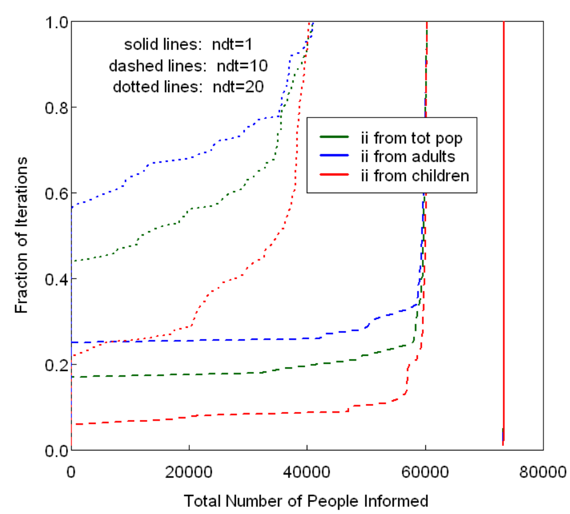


Fig. 3. Final informed size probability among different groups with different seed conditions.

Study Description: We define *outbreak* and *outbreak size* as the extent to which information spreads in the network. We observe this as an outcome of selecting certain parameter values. A *seed* indicates an individual or individuals who initially acquire the information from mass media, or other means. We choose the seed randomly (oblivious to the group the individual belongs to), or seed an individual from a particular group. The experiments are conducted by considering a normative day—activities remain the same for all days—with links in the locations randomized for each day. The discrete time unit (tu) of 10 minutes is used in EpiFast for this study (six time increments in a simulation represent an hour). All simulations started at midnight, and most simulations were 3 days in length, or 432 tu. The activity profiles obtained by **Step 2** in Section III-A were discretized into 10 minute intervals, and interactions of 5 minutes or less were ignored. Hence, the discrepancy can be as great one tu. The experiments are performed with varying ndt values of 1, 10, 20 tu corresponding to 10, 100 and 200 minutes, respectively. We repeat the simulations for 100 *iterations*; i.e., for a set of initial conditions, 100 diffusion processes were executed for sampling purposes. The process of information diffusion from an informed node i to an uniformed node j, at a particular time t in the simulation, requires the following conditions be met: (1) the edge weight w_{ij} is greater than a randomly generated number between 0 and 1; (2) the edge (i, j) exists at time t; and (3) the duration of the current interaction is greater than or equal to the ndt value of node j. If any of these conditions is not met, then the information is not transfered.

B. Results

Effect of changing seed and ndt on final informed size. Figure 3 shows the final outbreak size for different iterations for varying ndt and seed conditions. The line styles indicate the ndt values and line color indicates the seed conditions. Green indicates a randomly selected seed; blue corresponds to a seed selected from adults, and red indicates a youngster selected as the initial infected.

When $ndt = 1$ the variation of the outbreak size is independent of the seed—in all cases the entire population is informed. Increasing ndt causes the following effects: (1) raises the variability of the final outbreak size across iterations, (2) amplifies the difference between seeds, and (3) changes the probability that the information spread even takes off. The variability can be ascribed to choosing 1 informed individual initially and is not observed in the case for $ndt = 1$. Increasing ndt splits the social network (Figure 1 shows link numbers for different ndt), reducing the probability that the spread even takes off in some cases. For example, when a random seed (oblivious of the group) is chosen with ndt=20, this probability is approximately 0.4.

Effect of ndt on the information spread. Figure 4 shows the new infections for different values of ndt over a three day period. Even though the overall informed numbers are the same, there is considerable variation within each of the three runs across all iterations (due to single seed). For $ndt = 1$, clearly, most people are informed during the first day with the peak during the second half of the day. This number decreases, a smaller peak is observed during the second day and, few people are informed during the third day.

Figure 5 shows the number of newly informed people with respect to the activity type. Increasing ndt changes the way people receive information due to the pruning of weak links. When $ndt = 1$, a majority of people are informed while recreating, shopping or at home (Figure 5(a)). Increasing ndt causes information spread at school, home, and work to predominate. Further, increasing ndt shifts the peak of newly informed to Day 2 and 3, and Figures 5(b) and 5(c) show this effect at various locations. Increasing ndt prunes weak links formed at recreation and shopping locations, resulting in a smaller subset of people informed during day 1. These people get informed later—on Day 2 or 3—at home, work and school, where ties are strong.

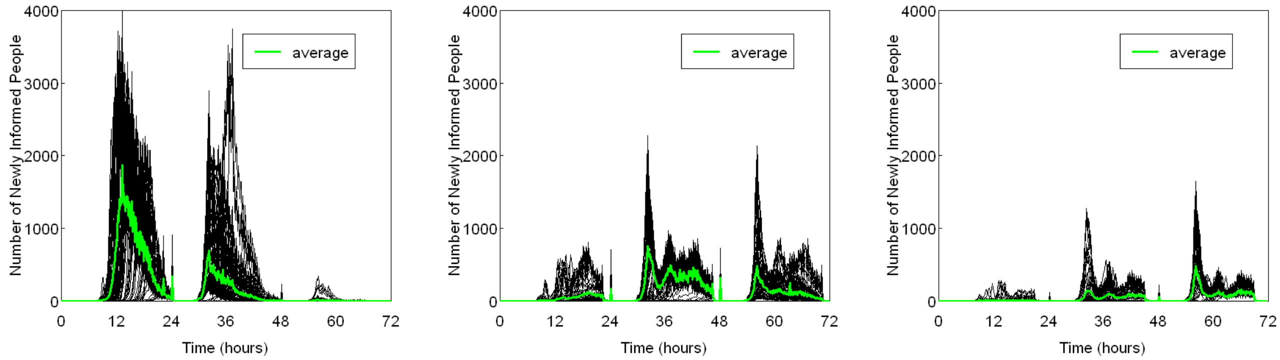
Figure 6 shows the fraction of people informed in each demographic factor as a function of time. For $ndt = 1$ all the people present within the value range for each demographic factor are informed after ≈ 40 hours. Further, age and Hh income show a similar trend for all value ranges, whereas Hh size $= 1$ has a slower spread. Similarity across value ranges in Figure 6(b) indicates that Hh income does not play a differentiating role in the process. For $ndt = 10$ or $ndt = 20$, there are still some uninformed individuals in all value ranges. In Figure 6(a) the spike in the spread for youngsters on the second and third day is due to interaction in schools and a reservoir of informed students built up during the previous day. A similar spike is not observed in Figure 6(b) because of the equal distribution of youngsters across varying income ranges. The effect of the distribution of youngsters in the Hh with size > 1 is the origin of the disperate spreading characteristics (Figure 6(c)).

Effect of activity location on information spread in youngsters. As seen from Figure 6(a), youngsters play a important role in information flow. Figure 7 shows the fraction of informed younsters at different activity locations. For youngsters, communication occurs mostly at school, independent of the ndt values. So, more youngsters get informed at school than at home. This is due to the fact that the probability of interaction with similar individuals is higher, and the duration of interaction is long enough to spread the information (*strong ties*).

V. Conclusions

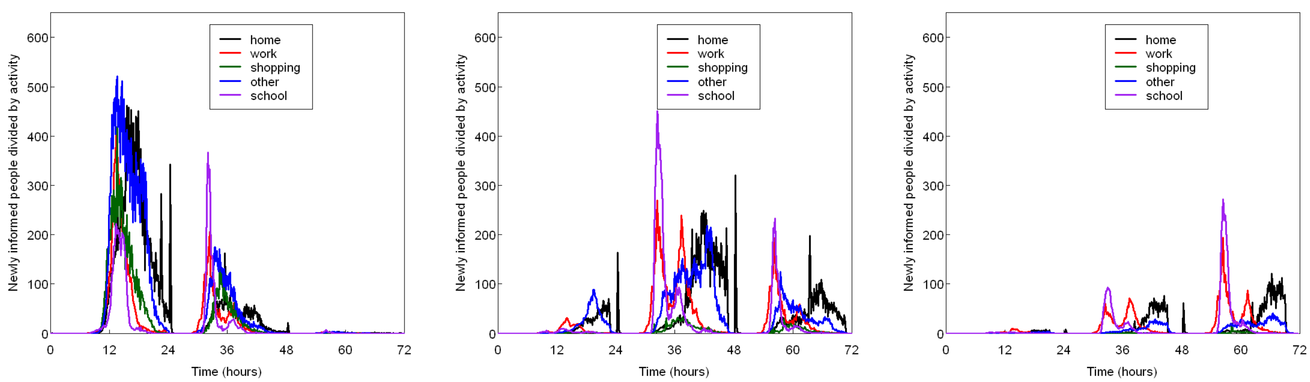
In this article, we have presented an interaction-based approach for studying the spread of rumors in a synthetic population under realistic conditions. Individuals in the synthetic population are endowed with demographic characteristics and a daily routine that creates a dynamic social network. We develop a interaction model based on similarity between communicating agents and the duration of contact. While the former allows us to model the homophily principle, the latter can be related to the importance/strength of the links.

We make the following conclusions from the study we present in this paper. If the information can be transmitted



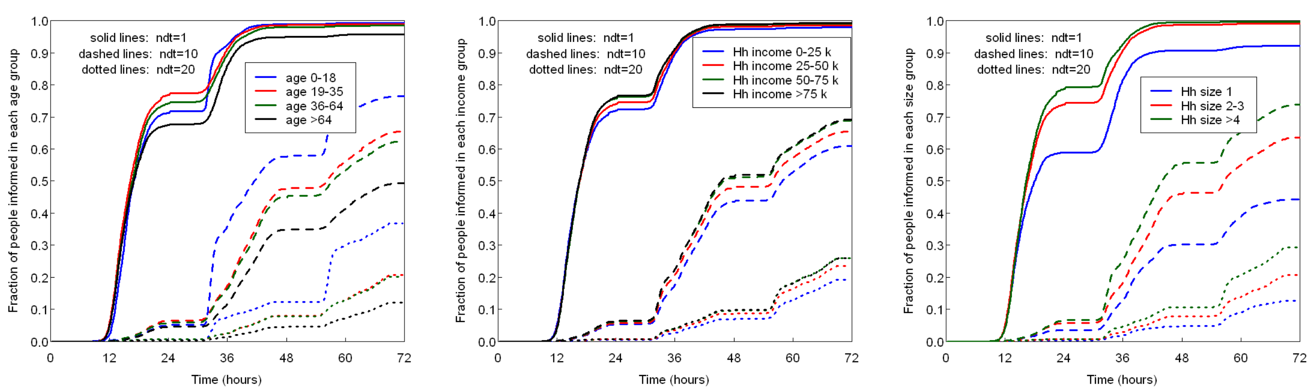
(a) ndt=1 (b) ndt=10 (c) ndt=20

Fig. 4. Comparison epicurves for different values of ndt=1,10,20



(a) ndt=1 (b) ndt=10 (c) ndt=20

Fig. 5. Newly informed people distributed according to activity



(a) Age (b) Household Income (c) Household Size

Fig. 6. Fraction of informed people in each value range of all the demographic factors (age, Hh income and Hh size.)

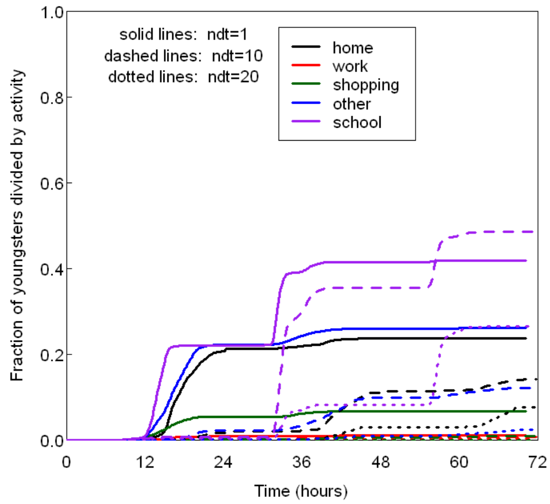


Fig. 7. Informed youngsters distribution according to activity

during a short conversation (say, with duration 10-20 minutes), recreation and shopping are activities where information spread is greatest. Otherwise, locations having stronger ties, such as home, work, and school, predominate the other locations in spreading the information. We also find that youngsters get informed mostly at school.

Our future works include the study of propagation of specific types of information, such as fashion, political opinion, and health related issues. The nature of information can affect the dynamics of adoption by the agents. For example, political opinion may be important to adults, whereas fashion or product usage may be more appealing to youngsters.

Acknowledgments

We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their advice and aid. This work has been partially supported by NSF Nets GrantCNS-0626964, NSF HSD Grant SES-0729441, CDC Center of Excellence in Public Health Informatics Grant 2506055-01, NIH-NIGMS MIDAS project 5 U01 GM070694-05, DTRA CNIMS Grant HDTRA1-07-C-0113 and NSF NETS CNS-0831633.

References

[1] J. M. Epstein, "Why model?" *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 4, p. 12, 2008.

[2] S. Moss and B. Edmonds, "Towards good social science," *Journal of Artificial Societies and Social Simulation*, vol. 8, no. 4, p. 13, 2005.

[3] P. Sobkowicz, "Modelling opinion formation with physics tools: Call for closer link with reality," *Journal of Artificial Societies and Social Simulation*, vol. 12, no. 1, p. 11, 2009.

[4] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modeling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, pp. 180–184, May 2004.

[5] K. Bisset, J. Cheng, X. Feng, V. S. A. Kumar, and M. V. Marathe, "Epi-Fast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems," in *Proceedings of the XXIII International Conference on Supercomputing*, June 2009.

[6] E. M. Rogers, *Diffusion of innovations*. New York: Free Press of Glencoe, 1962.

[7] J. Brown and P. Reinegen, "Social ties and word-of-mouth referral behavior," *Journal of Consumer Research*, vol. 14, no. 3, pp. 350–362, 1987.

[8] J. Coleman, H. Menzel, and E. Katz, *Medical Innovations: A Diffusion Study*. Bobbs Merrill, 1966.

[9] P. Domingos, "Mining social networks for viral marketing," *IEEE Intelligent Systems*, vol. 20, no. 1, pp. 80–82, 2005.

[10] D. Kempe, J. M. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.

[11] D. J. Crandall, D. Cosley, D. P. Huttenlocher, J. M. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2008, pp. 160–168.

[12] D. Kempe, J. M. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *ICALP*, 2005, pp. 1127–1138.

[13] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 61–70.

[14] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," in *Proceedings of the ACM Conference on Electronic Commerce*, 2006, pp. 228–237.

[15] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, "Theory of rumour spreading in complex social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 457 – 470, 2007.

[16] E. Agliari, R. Burioni, D. Cassi, and F. M. Neri, "Efficiency of information spreading in a population of diffusing agents," *Physical Review E*, vol. 73, no. 4, p. 046138, 2006.

[17] M. Everett and D. K. Bhowmik, "Homophily-heterophily: Relational concepts for communication research," *The Public Opinion Quarterly*, vol. 34, no. 4, pp. 523–538, 1970.

[18] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, pp. 415–444, 2001.

[19] M. S. Granovetter, "The strength of weak ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[20] R. Axelrod, "The dissemination of culture - a model with local convergence and global polarization." *Journal of Conflict Resolution*, vol. 41(2), pp. 203–226, 1997.

[21] T. Roenneberg, T. Kuehnle, M. Juda, T. Kantermann, K. Allebrandt, M. Gordijn, and M. Merrow, "Epidemiology of the human circadian clock," *Sleep Medicine Reviews*, vol. 11, no. 6, pp. 429 – 438, 2007.

[22] C. Barrett, K. Bisset, J. Leidig, A. Marathe, and M. Marathe, "Estimating the impact of public and private strategies for controlling an epidemic: A multi-agent approach," July 2009, to Appear.

[23] R. Blendon, L. S. Koonin, J. M. Benson, M. S. Cetron, W. E. Pollard, E. W. Mitchell, K. J. Weldon, and M. J. Herrmann, "Public response to community mitigation measures for pandemic influenza," *Emerging Infectious Diseases*, vol. 14, pp. 778–786, 2008.

[24] H. Abbey, "An examination of the Reed-Frost theory of epidemics," *Human Biology*, vol. 24, pp. 201–233, 1952.