Simulating Social Information Diffusion Using a Synthetic Population

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, VA, USA
Email: {apolloni, kchannak, ldurbeck, maleq, ckuhlman, blewis, swarup}@vbi.vt.edu

Word of mouth, rumors, and gossip are forms of interpersonal communication which can affect many human affairs: they shape public opinion, impact financial markets, create awareness or panic during disease outbreaks, etc. The diffusion of information through these channels resembles epidemic processes and is strongly affected by the structure of the social network. Unlike epidemic processes, however, communication flows depend strongly on similarity in social status, personality type, shared opinions or beliefs, etc.

There have been many attempts to model these phenomena (e.g., see [1]), but in all these cases assumptions are made regarding

1. the construction of the social network, which is considered static, or evolving according a specific set of rules, and

2. the probability of transmission, which is considered constant and independent of the particular interactors.

In reality, the network changes over time due to people's activity patterns. This requires embedding information about temporal ordering and "liveness" of links in the social network model. At the same time communication occurs only if a certain relation between individuals exists and only if interactors have enough time for conversation. This requires a more complex information diffusion model that takes into account the demographic variables describing the interactors.

In this work, we use a synthetic population, representing Montgomery County in Virginia, for simulating diffusion of information by word of mouth in a *realistic* social network. Data from the U.S. Census, activity and mobility surveys, and maps and regional infrastructure information have been used to generate the synthetic population. Individuals are endowed with demographic characteristics, their households are located based on census block and street data, and an activity routine (i.e. a sequence of the different daily activities with start and end time evaluated in seconds from midnight) is associated with each individual according to survey data. Activities are geographically located based on street data and individuals are assigned to the location at a specific time based on the activity routine. The number of people with whom individuals are actually in contact depends on the time of the day, the size of a location and its internal divisions, called sublocations (e.g., rooms in a building). The result is a dynamic contact network whose time evolution depends on individuals' activities.

Our information diffusion model makes the following assumptions:

- Individuals can interact only when they are collocated.
- The probability with which an individual conveys information to another individual is related to the similarity between the individuals, where similarity is based on social status (demographic variables).
- Information transmission requires contact of a minimum duration, and the probability of transmission increases monotonically with the duration of contact beyond this minimal value. This simulates the fact that information transmission is not always immediate, but varies with the topic, i.e., some subjects might be broached only in a lengthy interactions, whereas others might be discussed even in short encounters.

Numerical simulations have been performed using EpiFast, which is a simulation engine for modeling dynamic processes on very large social networks that has been developed over several years at the Network Dynamics and Simulation Science Lab. It has mainly been used to model disease spread in network models of US cities and states, which can consist of several million nodes (representing people) and tens of millions of edges (representing contacts between people) [2]. We have now extended it to model cultural processes like the spread of gossip and rumors.

The simulation is initialized by randomly choosing a small subset of the population to be *informed*, while the rest are *uninformed*. We track the number of people newly informed at every time step. We assume that once people become informed, they stay in that state, and continue to try to inform other people over the entire duration of the simulation. The results show several interesting phenomena. Children turn out to play a major role in the spread of information through the network, mainly through interactions at school. The results show how different initial conditions (e.g., selecting people from different demographic sets) affect the duration and the size of diffusion within and across demographic sets. The results are also used to identify the set of most important activities and groups of people that influence the rate of spread.

References

[1] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili (2007), Theory of rumour spreading in complex social networks, *Physica A* 374(1), pp. 457-470.

[2] K. Bisset, J. Cheng, X. Feng, V. S. Anil Kumar, and M. V. Marathe (2009), EpiFast: A fast algorithm for large scale realistic epidemic simulations on distributed memory systems, In *Proceedings of 23rd ACM International Conference on Supercomputing (ICS'09)*, pp. 430-439.