
Graph Theoretic Measures for Identifying Effective Blockers of Spreading Processes in Dynamic Networks

Habiba

Tanya Y. Berger-Wolf

Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan, Chicago, Illinois 60607.

Keywords: graph mining, social network analysis

HHABIBA@CS.UIC.EDU

TANYABW@UIC.EDU

1. Introduction

Many processes within a society, such as diseases, opinions, information, or behavior, spread through a network of personal interactions. Whether a phenomenon is going to spread widely within a population depends on the influence of affected individuals and the susceptibility of the unaffected ones, as well as the pattern of the interactions among individuals. Such spreading phenomena have been studied in numerous domains including epidemiology (Dezsö & Barabási, 2002; Eubank et al., 2004; Kempe et al., 2003; May & Lloyd, 2001; Morris, 1993; Newman, 2002; Pastor-Satorras & Vespignani, 2001), diffusion and adoption of novel products and technologies (Chen & Carley, 2005; Domingos, 2005; Goldenberg et al., 2001; Leskovec et al., 2006; Rogers, 2003; Young, 2006), voting and rumors (Granovetter, 1978; Moreno et al., 2004; Zanette, 2002), and numerous others.

One important question in the context of spreading processes is to find a set of individuals whose removal would result in the largest reduction of the extent of spread. We call this set of individuals “blockers” of the spread. For example, the blockers of a spreading disease are the set of individuals who are the best candidates for vaccination or quarantine. It is necessary to highlight here that while epidemiologically both quarantine and vaccination have the similar goal of reducing the extent of spread of a disease computationally they are two different problems. Quarantine is modeled by removing a node from the network while vaccination deactivates its ability to pass on a disease while leaving it in the network. Under some disease spread models those actions are equivalent.

Related questions of identifying the source of an outbreak and optimizing some outbreak detection criteria by strategically placing sensors (Berger-Wolf et al., 2005; Berry et al., 2005a; Hart et al., 2006; Eubank et al., 2004; Kempe et al., 2002; Leskovec et al., 2007) have been addressed in the context of various applications but there has been little work done for the identification of the critical individuals whose removal minimizes the spread of a phenomenon in the network once it has started spreading. The sensor placement prob-

lem can be viewed as one of identifying individuals who minimize the extent of spread while remaining in the network (somewhat similar to vaccination). In our work we focus on identifying the best blockers to be removed. The related work, including the recent developments on monotonic submodular functions (Kempe et al., 2003; Leskovec et al., 2007), is significantly different from this formulation and the computational and algorithmic results, unfortunately, do not apply. Thus, in this work we propose a set of structural network measures and a heuristic, yet systematic, approach based on those measures to identify good blockers.

An intrinsic property of populations missing in most analysis of the spreading processes on networks is the changing nature of the interactions. For example, in the context of disease spread, two individuals may be in contact for a long period of time while they are in the susceptible stage, but they may not be in close proximity of each other when either of them gets infected. Here, the explicit timing and order of interactions affect the resulting spread of the disease. Moreover, for diseases like bird flu the exact sequence of the contacts in birds is known and can be used to predict disease spread. Similarly, zebras are influenced by those with whom they have been in close proximity most recently (Fischhoff et al., 2007). Thus, a change in the behavior in zebra population will spread only over those recent contacts. Here again, the exact observed sequence of movements is available and our conclusions can be tested by comparing them with the field observations. The very nature of a spreading process implies an explicit time axis (Kempe et al., 2002). In this paper, we focus on explicitly dynamic networks, defined in Section 2.1.

2. Preliminaries and Definitions

2.1. Dynamic Network

Definition 1 Let $\{1, \dots, T\}$ be a finite set of discrete timesteps. Let $V = \{1, \dots, n\}$ be a set of individuals. Let $G_t = (V_t, E_t)$ be a graph representing a snapshot of a static network at time t : $V_t \subseteq V$, is a subset of individuals V observed at time t and an edge $(u_t, v_t) \in E_t$ if individuals u and v have interacted at time t .

A dynamic network $G = \langle G_1, \dots, G_T \rangle$ is the graph $G = (V, E)$ of the time series of graphs G_t such that $V = \bigcup_t V_t$ and $E = \bigcup_t E_t \cup \bigcup_{t-1} (v_t, v_{t+1})$.

In general, in a dynamic network both edges and individuals can change over time. However, for simplicity, we assume that the set of individuals remains the same throughout the entire time period: $V = V_1 = \dots = V_T$. However, our results easily extend to networks where this assumption does not hold. The edges E_t may be directed or undirected and weighted or unweighted. The self-edges of the type (v_t, v_{t+1}) are always directed.

2.2. Independent Cascade Spreading Model

The independent cascade model describes a spreading process of activation of individuals. In each discrete timestep, each active individual attempts to activate each of its neighbors independently with some probability of success. If an active individual succeeds those neighbors become active in the next time step. Each attempt of activation is independent of all previous attempts as well as the attempts of any other active individual to activate a common neighbor.

More formally, the input to the independent cascade model is a set $A_0 \subseteq V$ of active individuals. In a dynamic network, an active individual u_t in timestep t tries to activate each of its currently inactive neighbors v_t only once with a probability p_{u_t, v_t} , independent of all the other neighbors. If u_t succeeds in activating v_t at timestep t , then v_t will be active in step $t + 1$, whether or not $(u_{t+1}, v_{t+1}) \in E_{t+1}$. If u_t fails in activating v_t , and at any subsequent timestep u_{t+i} gets reconnected to v_{t+i} , it will again try to activate u_{t+i} . The process runs for a finite number of timesteps T or until no more activations are possible. We denote by $\sigma(A_0) = A_T$ the correspondence between the initial set A_0 and the resulting set of active individuals A_T . We call the size of the set A_T , $|A_T|$, the *extent of spread*.

The spreading process in the independent cascade model in a dynamic network is different from the static network in one important aspect. In the static case, each individual u uses all its attempts of activating each of its inactive neighbors v with the same probability $p_{u, v}$ in one timestep t . This is the timestep right after the individual u itself becomes active. After that single attempt the active individual becomes latent: that is, it is active but unable to activate others. However in the dynamic network model as defined above, the active individuals never become latent during the spreading process. For this paper, we only consider the progressive case in which an individual converts from inactive to active but never reverses (SI epidemiological model). It is a particularly important case in the context of identifying blockers since the blocking action is typically done

before any recovery.

2.3. Spread Blockers

We now formalize the notions of processes spreading in a network and individuals blocking this spread.

$Spread(\cdot)$ is a function that gives the overall average extent of spread in a network, that is, the expected number of individuals affected by a stochastic spreading process after a specified number of timesteps. The estimate of the spread is dependent on the model of the spreading process and the structure of the network. $Spread_v(\cdot)$ is the expected spread in a network, when the spreading process is initiated by the individual v . Given a model of a spreading process \mathcal{M} and a distribution of the probability of infection \mathcal{X} , we define the spreading functions as follows:

$$Spread_v : \{G, \mathcal{M}, \mathcal{X}\} \rightarrow \mathbf{R}^+$$

$$Spread(G) = \frac{1}{|V|} \sum_{v \in V} Spread_v(G, \mathcal{M}, \mathcal{X})$$

We define $Bl_X(\cdot)$ as a function that measures the reduction in the expected spread size *after removing* the set X of individuals from the network. Hence, the blocking capacity of a single individual v , $Bl_v(\cdot)$, is the reduction in expected spread size after removing individual v from the network.

$$Bl_X(G) = Spread(G) - Spread(G \setminus X).$$

$kBl(\cdot)$ is the function that finds the set of individuals of size k that results in the maximum reduction in spread in a network when that set is removed from the network. That is, this function finds the best blocker(s) in a network.

$$kBl(G) = \max_{X \subseteq V, |X|=k} Bl_X(G).$$

Thus, finding the best blockers in the network is equivalent to finding the (set of) individuals whose removal from the network *minimizes* the expected extent of spread in that network.

$$kBl(G) = \min_{X \subseteq V, |X|=k} Spread(G \setminus X).$$

2.4. Dynamic Network Measures

Standard network analysis employs various graph measures such as centrality or a clustering coefficient of a node to determine the relative importance of that node. Such highly ranked individual may be a good candidate blocker of a spreading process. We extend these measures to explicitly dynamic networks.

Dynamic Degree is the change in the neighborhood of an individual over time:

$$deg_T(u) = \sum_{1 < t \leq T} \frac{|N(u_{t-1}) - N(u_t)|}{|N(u_{t-1}) \cup N(u_t)|} * N(u_t),$$

where $N(u_t)$ is the neighborhood of u at time t .

Dynamic Average Degree of an individual is the fraction of other individuals that it interacts with over T timesteps:

$$deg_a(u) = \frac{\sum_{1 \leq t \leq T} |N(u_t)|}{T}$$

Dynamic Betweenness of an individual, similarly to the static betweenness, is the fraction of all the shortest *temporal paths* that pass through the individual. Temporal paths are formally defined in (Kempe et al., 2002; Habiba et al., 2007). Intuitively, the edges in a temporal path appear in the increasing time order. We present in detail different flavors of the traditional betweenness centrality concept in dynamic networks based on position, time, and duration of interactions among individuals in (Habiba et al., 2007).

Dynamic Closeness of an individual to all other individuals is the average *time* it takes from that individual to reach any other individual in the network. Dynamic closeness is also based on shortest temporal paths (Habiba et al., 2007) between individuals. Hence, the geodesic here is defined as the time duration of the shortest temporal path.

Dynamic Clustering Co-efficient of an individual at a given time t is the fraction its neighbors who have been neighbors among themselves in any previous timesteps:

$$CC_T(u) = \sum_{v \in N(u_T)} \frac{\left| \bigcup_{1 \leq t < T} N(v_t) \cap N(u_T) \right|}{|N(u_T)|}.$$

3. Experimental Setup

Given the dynamic network graph $G = (V, E)$ we compute the individual ranks based on graph theoretic measures defined in Section 2.4. We simulate the independent cascade spreading process on the dynamic network $G = (V, E)$ initiating it from one active individual at a time. We calculate the average extent of spread in the network over all starting individuals. For each graph measure we remove from the network the highest ranked individual in the current network based on that

measure and calculate the average extent of spread on the remaining network. We repeat this processes until $\leq 20\%$ of the individuals have been removed. We compare the reduction of the extent of spreads based on the removal of highest ranked individuals for each measure and, thus, identify most effective measures to use as indicators of blockers of spread.

4. Datasets Used in Experiments

Grevys: Populations of Grevy’s zebras (*Equus grevyi*) were observed by biologists (Sundaresan et al., 2006) in the summer 2002 in Kenya. In the resulting dynamic network, each node represents an individual animal and an interaction corresponds to physical proximity.

Onagers: Populations of onagers (*Equus hemionus*) were observed by biologists (Rubenstein et al., 2007) in January–May 2003 in India. The network is similar to that of zebras.

DBLP: We use a cleaned sample from 1967–2005 of the *Digital Bibliography and Library Project* (Ley, 2005), a Computer Science bibliography. In the dynamic network each node represents an author and an interaction is a co-authorship.

Table 1. Dynamic network dataset statistics.

Here $V = \#\text{individuals}$, $E = \#\text{edges}$, $T = \#\text{timesteps}$, $D = \text{density}$, $d = \text{diameter}$, and $p = \text{ave. path length}$.

Dataset	V	E	T	D	d	p
Grevys	28	779	44	0.52	36	4.81
Onagers	29	402	82	0.36	3	7.51
DBLP	1374	2262	38	0.09	37	5.12

5. Results and Conclusions

Figures 1, 2, and 3 show the reduction in the extent of spread for different measures based on experiments on Grevys, Onagers and DBLP datasets. The slope of the curves indicates how quickly the extent of spread reduces in a given network based on each measure. We also experimented with randomly picking and removing individuals from the network as the baseline.

The removal of the individuals with highest degree, average degree, or clustering coefficient from the network results in the quickest reduction of the spread in the network. Hence, these three measures are better candidate indicators of key blockers of spreads in a dynamic network than betweenness and closeness. As future work we plan to experiment with the extent of spreads based on the criteria of minimizing the time to spread rather than, as in this paper, the extent of spread within a fixed time. Intuitively, in that case betweenness and closeness should perform relatively better as indicator of blockers of spreads.

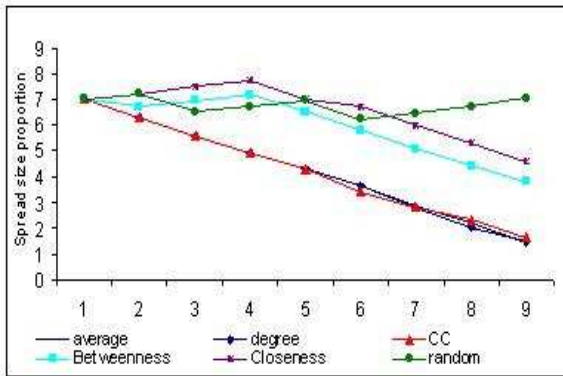


Figure 1. Independent Cascade-Grevys

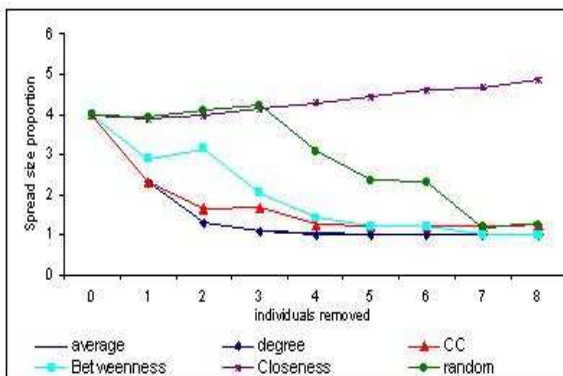


Figure 2. Independent Cascade-Onagers

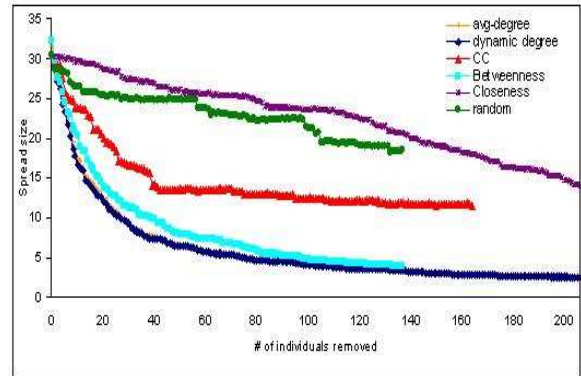


Figure 3. Independent Cascade-DBLP

References

- Berger-Wolf, T., Hart, W., & Saia, J. (2005). Discrete sensor placement problems in distribution networks. *Math. and Comp. Model.*
- Berry, J., Fleischer, L., Hart, W., Phillips, C., & Watson, J. (2005a). Sensor placement in municipal water networks. *J. of Water Resources Planning and Management*, 131.
- Carley, K. (1996). Communicating new ideas: The potential impact of information and telecommunication technology. *Tech. in Society*, 18, 219–230.
- Chen, L., & Carley, K. The impact of social networks in the propagation of computer viruses and countermeasures. *IEEE Tras on Sys, Man and Cyber*.
- Dezső, Z., & Barabási, A.-L. (2002). Halting viruses in scale-free networks. *Phys. Rev. E*, 65.
- Domingos, P. (2005). Mining social networks for viral marketing. *IEEE Intel. Sys.*, 20, 80–82.
- Eubank, S., Guclu, H., Kumar, V., Marathe, M., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 429:180–184.
- Fischhoff, I. R., Sundaresan, S. R., Cordingley, J., & Rubenstein, D. I. (2007). Habitat use and movements of plains zebra (*equus burchelli*) in response to predation danger from lions. Submitted.
- Goldenberg, J., Libai, B., & Muller, E. (2001). Using complex systems analysis to advance marketing theory development. *Academy of Mkt. Sci. Rev.*
- Granovetter, M. (1978). Threshold model for collective behavior. *American J. of Socio.*, 83, 1420–1443.
- Habiba, & Berger-Wolf, T. Y. (2007). *Influence maximization in dynamic networks* (DIMACS TR 2007-20).
- Habiba, Tantipanananadh, C., & Berger-Wolf, T. (2007). *Betweenness centrality in dynamic networks* (DIMACS TR 2007-19).
- Hart, J. B. W., Phillips, C., Uber, J. G., & Watson, J. (2006). Sensor placement in municipal water networks with temporal integer programming models. *J. Water Resources Planning and Management*, 132, 218–224.
- Kempe, D., Kleinberg, J., & Kumar, A. (2002). Connectivity and inference problems for temporal networks. *J. Comput. Syst. Sci.*, 64, 820–842.
- Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. *Proc. of the 9th ACM SIGKDD Int. Conf. on KD and DM*.
- Leskovec, J., Adamic, L. A., & Huberman, B. A. (2006). The dynamics of viral marketing. *Proc. 7th ACM EC*. (pp. 228–237).
- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., & VanBriesen, J. (2007). Cost-effective outbreak detection in networks. *Proc. 13th KDD*
- Ley, M. (2005). Digital bibliography & library project (DBLP). <http://dblp.uni-trier.de/>. A digital copy available online.
- May, R. M., & Lloyd, A. L. (2001). Infection dynamics on scale-free networks. *Phys. Rev. E*, 64.
- Moreno, Y., Nekovee, M., & Pacheco, A. F. (2004). Dynamics of rumor spreading in complex networks. *Phys. Rev. E*, 69, 066130.
- Morris, M. (1993). Epidemiology and social networks: modeling structured diffusion. *Socio. Methods and Research*, 22, 99–126.
- Newman, M. E. (2002). Spread of epidemic disease on networks. *Phys. Rev. E*, 66.
- Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86, 3200–3203.
- Rogers, E. M. (2003). *Diffusion of innovations*. Simon & Shuster, Inc. 5th edit.
- Rubenstein, D. I., Sundaresan, S., Fischhoff, I., & Saltz, D. (2007). Social networks in wild asses: Comparing patterns and processes among populations. In A. Stubbe, P. Kaczensky, R. Samjaa, K. Wesche and M. Stubbe (Eds.), *Explor. into the bio. resou. of mongolia*, vol. 10. Martin-Luther-Univ. Halle-Wittenberg.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31, 581–603.
- Sundaresan, S. R., Fischhoff, I. R., Dushoff, J., & Rubenstein, D. I. (2006). Network metrics reveal differences in social organization between two fission-fusion species, Grevy's zebra and onager. *Oecologia*.
- Young, H. P. (2006). Innovation diffusion and population heterogeneity.
- Zanette, D. H. (2002). Dynamics of rumor propagation on small-world networks. *Phys. Rev. E*, 65, 041908.