Finding Spread Blockers in Dynamic Networks

Habiba*[1], Yintao Yu†[2], Tanya Y. Berger-Wolf‡[1], and Jared Saia§[3]

[1] University of Illinois at Chicago, {hhabib3, tanyabw} @uic.edu
[2] University of Illinois at Urbana-Champaign, yintao@uiuc.edu
[3] University of New Mexico, saia@cs.unm.edu

Abstract. Social interactions are conduits for various processes spreading through a population, from rumors and opinions to behaviors and diseases. In the context of the spread of a disease or undesirable behavior, it is important to identify *blockers*: individuals that are most effective in stopping or slowing down the spread of a process through the population. This problem has so far resisted systematic algorithmic solutions. In an effort to formulate practical solutions, in this paper we ask: *Are there structural network measures that are indicative of the best blockers in dynamic social networks?* Our contribution is two-fold. First, we extend standard structural network measures to dynamic networks. Second, we compare the blocking ability of individuals in the order of ranking by the new dynamic measures. We found that overall, simple ranking according to a node's static degree, or the dynamic version of a node's degree, performed consistently well. Surprisingly the dynamic clustering coefficient seems to be a good indicator, while its static version performs worse than the random ranking. This provides simple practical and locally computable algorithms for identifying key blockers in a network.

1 Introduction

How can we stop a process spreading through a social network? This problem has applications to diverse areas such as preventing or inhibiting the spread of diseases [7, 26, 40], computer viruses[4] [8, 22], rumors, and undesirable fads or risky behaviors [23, 24, 37, 38]. A common approach to spread inhibition is to identify key individuals whose removal will most dampen the spread. In the context of the spread of a disease, it is a question of finding individuals to be

* Work supported in part by the Fulbright fellowship.
† Work performed in part while being a visiting student at the University of New Mexico.
‡ Work supported in part by the NSF grant IIS-0705822 and NSF CAREER Award 0747369.
§ Work supported in part by the NSF grant IIS-0705822, NSF CAREER Award 0644058, and an AFO MURI award.

[4] In particular, we are concerned with computer malware that spreads through social networks, such as email viruses and worms, cell-phone viruses, and other related malware such as the recent MySpace worm.

quarantined, inoculated, or vaccinated so that the disease is prevented from becoming an epidemic. We call this set of key individuals the *blockers* of the spreading process.

There has been significant previous work related to studying and controlling the spread of dynamic processes in a network [9–11, 16, 18, 22, 23, 26, 35, 40, 43, 44, 46, 47, 51, 54, 57, 59, 60, 67]. Unfortunately, these results have three properties rendering them ineffective for identifying good blockers in large networks. First, many proposed algorithms focus on a slightly different objective: they aim to identify nodes that will be most effective in *starting* the spread of a process rather than blocking it [44, 47]; or alternatively, nodes that would be most effective in sensing that a process has started to spread, and where the process initiated [9–11]. In this paper, we are focused specifically on identifying those nodes that are good blockers. Second, algorithms proposed in previous work all require computationally expensive calculations of some global properties over the entire network, or rely on expensive, repeated stochastic simulations of the spread of a dynamic process. In this paper, we present heuristics that identify good blockers quickly, based only on local information.

Finally, perhaps the most critical problem in previous work is the omission of the dynamic nature of social interactions. The very nature of a spreading process implies an explicit time axis [52]. For example, the flow of information through a social network depends on who starts out with the information when, and which individuals are in contact at the starting point with the information carrier [43]. In this paper, we consider explicitly dynamic networks, defined in Section 3.1. In these networks, we study the social interactions over a finite period of time, measured in discrete timesteps.

The main contributions of this paper are summarized below.

- We formally define dynamic networks in Section 3.1. This representation of networks encompasses the traditional "aggregate" view of networks defined in Section 3.2 and adds the explicit temporal component to the interactions. The time axis is necessary since most spreading processes take place on networks that evolve over time.
- We formally define the problem of identifying key spread blockers in networks in Section 3.3.
- We modify various network measures, such as the centrality measures and clustering coefficient, to incorporate the dynamic nature of the networks (Section 3.4).
- We compare the reduction in the extent of spread based on removing individuals from a network in the ranking order imposed by various network measures. We identify measures that consistently give a good approximation of the best spread blockers.
- We compare the difference in the sets of top blockers identified by various measures.
- We extensively evaluate our methods on real networks (Section 5). We use the Enron email network dataset, the MIT Reality Mining dataset, DBLP co-authorship network, animal population networks of Grevy's zebras, Plains zebras, and onagers.

Ultimately, we show that the dynamics of interactions matters, and moreover that simple local measures, such as degree, are highly indicative of an individual's capacity to prevent the spread of a phenomenon in a population. The implication of our results are that there are practical scalable heuristics for identifying quarantine and vaccination targets in order to prevent an epidemic.

2 Related Work

Dynamic phenomena such as opinions, information, fads, behavior, and disease spread through a network by contacts and interactions among the entities of the network. Such spreading phenomena have been studied in a number of domains including epidemiology [22, 26, 40, 51, 54, 57, 59], diffusion of technological innovations and adoption of new products [7, 16, 18, 23, 24, 35, 38, 44, 46, 60, 67], voting, strikes, rumors [36, 37, 53, 68], as well as spread of contaminants in distribution networks [8–11, 46] and numerous others.

One of the fundamental questions about dynamic processes is: Which individuals, if removed from the network, would block the spread of such process? Several previous results have addressed the problem of identifying such individuals [26, 40, 43]. Eubank et al. [26] experimentally show that global graph theoretic measures like expansion factor and overlap ratio are good indicators for devising vaccination strategies in static networks. Cohen et al. [21] propose another immunization strategy based on the aggregate network model. In particular, they propose an efficient method of picking high degree nodes in a network to immunize, thus inhibiting the spread of disease. Kempe et al. [43] show that a variant of the blocker identification problem is NP-hard. While these problems and suggested approaches are similar to finding good blockers in a network, unfortunately, there are critical differences that make these results inappropriate for our formulation. First of all our objective is to minimize the expected extent of spread in a network. We do not make any assumption about the source of the spread. Second, almost all the above methods simplify the spreading process by ignoring the time ordering of interactions.

There has also been significant related work on the problem of determining where to place a small number of detectors in a network so as to minimize the time required to detect the spread of a dynamic process, and, ideally, also the location at which the spread began. Berger-Wolf et al. [9] give algorithms for the problem of minimizing the size of the infected population before an outbreak is detected. Berry et al. [10, 11] give algorithms to strategically place sensors in utility distribution networks to minimize worst case time until detection. In [47], Leskovec et al. demonstrate that many objectives of the detection problem exhibit the property of submodularity. They exploit this fact to develop efficient and elegant algorithms for placing detectors in a network. While the detection problem is related to the problem of blocking a process, it is only concerned with detecting a spreading process once, whereas a good blocker prevents multiple spreading paths. Moreover, the algorithms proposed for the detection problem

all require global information and work only for a stable, relatively unchanging network.

Another related problem is that of identifying nodes in a network that are most critical for *spreading* a dynamic process. Kempe et al. [44] show that identifying key spreaders – individuals that help to propagate the spread most – is NP-hard, but admits a simple greedy $(1-1/e)$-approximation. Later, Mossel and Roch [55] showed that the general case of finding a set of nodes with the largest "influence" is NP-hard, and has a $(1 - 1/e - \varepsilon)$ approximation algorithm. Unfortunately, this approximation algorithm is computationally intensive. Strong inapproximability results for several variants of identifying nodes with high influence in social networks have been shown in [19]. Asur et al. in [5] present an event based characterization of critical behavior in interaction graphs for the purposes of modeling evolution, link prediction, and influence maximization.

Finally, Aspnes et al. [4] have studied the inoculation problem from a graph theoretic perspective. They show that finding an optimum inoculation strategy is related to the sum-of-squares partition problem. Moreover, they show that the social welfare of an inoculation strategy found when each node is a selfish agent can be significantly less than the social welfare of an optimal inoculation strategy.

3 Definitions

Populations of individuals interacting over time are often represented as networks, or graphs, where the nodes correspond to individuals and a pairwise interaction is represented as an edge between the corresponding individuals. The idea of representing societies as networks of interacting individuals dates back to Lewin's earlier work of group behavior [48]. Typically, there is a single network representing all interactions that have happened during the entire observation period. We call this representation an *aggregate network* (Section 3.2). In this paper we use an explicitly dynamic network representation (Section 3.1) that takes the history of interactions into account.

3.1 Dynamic Network

We represent dynamic network as a series $\langle G_1, \ldots, G_T \rangle$ of *static* networks where each G_t is a snapshot of individuals and their interactions at time t. For this work, we assume that the time during which the individuals are observed is finite. For simplicity, we also assume that the time period is divided into discrete steps $\{1, \ldots, T\}$. The nontrivial problem of appropriate time discretization is beyond the scope of this paper. We assume that an interaction between a pair of individuals takes place within one timestep.

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

time t. Further, for all $v \in V$ and $t \in \{1, \ldots, T-1\}$ the edges $(v_t, v_{t+1}) \in E$ are directed self edges of individuals across timesteps.

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

The definition is equivalent to an undirected multigraph representation in [43].

Figure 1 shows an example of several dynamic networks that have the same unweighted aggregate network representation.

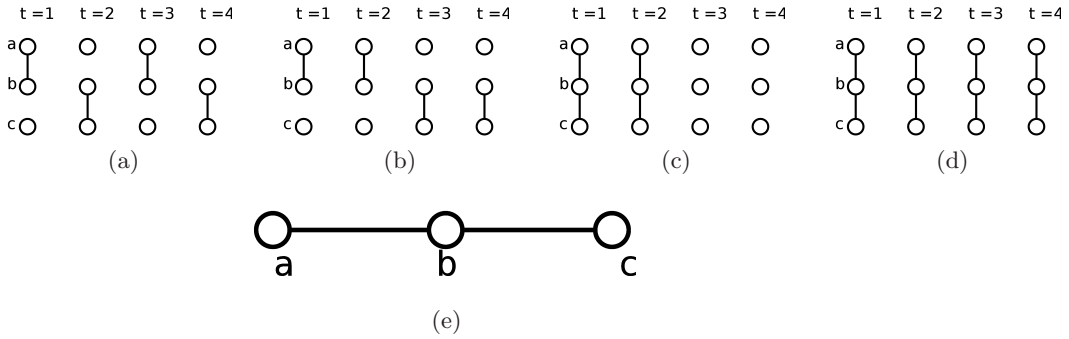

Fig. 1. Example of several dynamic networks that have the same unweighted aggregate network representation. Figures (a)–(d) show a dynamic networks of three individuals interacting over four timesteps. The solid line edges represent interactions among individuals in a timestep. Empty circles are individuals observed during a timestep. While at any given timestep some individuals may be unobserved, the particular example shows all the individuals being observed at all timesteps. Figure (e) shows an unweighted aggregate network that has the same interactions as every dynamic network in the example. Figures (a)–(c) have the multiplicity two of each edge while figure (d) has the multiplicity four for every edge in the aggregate representation.

3.2 Aggregate Network

The aggregate network is the graph $G_A = (V, E)$ of individuals V and their interactions E observed over a period of time. In this representation an edge exists between a pair of individuals if they have ever interacted during the observed time period. Multiple interactions between a pair of individuals over time are represented as a single, possibly weighted, edge or multiple edges between them. This representation provides an *aggregate* view of the population where the information about the timing and order of interactions is discard. In this work we represent aggregate networks as multigraphs.

Definition 2. *Let $\{1, \ldots, T\}$ be a finite set of discrete timesteps. Let V_t be the set of individuals observed at time t and let E_t be the set of interactions among*

individuals V_t at t. Then the aggregate graph $G_A = (V, E)$ of such a network is the set of individuals V and interactions E such that $V = \bigcup_t V_t$ and $(u, v) \in E$ if $\exists(u_t, v_t)$ at some timestep $t \in \{1, \ldots, T\}$.

Using this aggregate network model, the structure and properties of many social networks have been studied from different perspectives [6, 12, 13, 15, 41]. However, as we have mentioned, this and other similar models do not explicitly consider the temporal aspect of the network.

3.3 Spread Blockers

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

$Spread(.)$ 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 within a specified number of timesteps. The estimate of the spread is dependent on the model of the spreading process, the structure of the network, and, of course, the number of timesteps under consideration. $Spread_v(.)$ 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 $\mathcal{X} : E \to [0, 1]$, we define the spreading functions as follows:

$$Spread_v : \{Networks \times Spread\,Models \times Probability \times Time\} \to \mathbb{R}^+$$

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

The limit equilibrium state of spread is denoted by

$$Spread(G, \mathcal{M}, \mathcal{X}) = Spread(G, \mathcal{M}, \mathcal{X}, \infty) \tag{2}$$

For a fixed spread model, probability distribution and a time period we will use the overloaded shorthand notation $Spread(G)$.

We define $Bl_X(.)$ 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, $Bl_v(.)$, is the reduction in expected spread size after removing individual v from the network.

$$Bl_X : \{Networks \times Spread\,Models \times Probability \times Time\} \to \mathbb{R}^+$$

$$Bl_X(G) = Bl_X(G, \mathcal{M}, \mathcal{X}, T) = Spread(G) - Spread(G \setminus X). \tag{3}$$

$kBl(.)$ is the function that finds the maximum possible reduction in spread in a network when set of individuals of size k is removed from the network. Notice,

that the value of this function is always at least k. The argmax of this function finds the best blocker(s) in a network.

$$kBl(G) = \max_{X \subseteq V, |X|=k} Bl_X(G). \tag{4}$$

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.

$$kBl(G) = Spread(G) - \min_{X \subseteq V, |X|=k} Spread(G \setminus X). \tag{5}$$

This definition of the individuals' blocking capacity by removal corresponds in the disease spread context to the quarantine action. Vaccination or inoculation leave the node in the network but deactivate its ability to propagate the spread. For the Independent Cascade model of spread (Section 3.5) the two actions are equivalent at the abstract level of estimating the spread and identifying blockers in networks.

Since no good analytical approaches are known for identifying blockers in networks, in this paper we focus on examining the possibility of using structural network measures as practical indicators of nodes' blocking ability. We next briefly define the structural measures used in this paper.

3.4 Network Structural Measures

In network analysis various properties of the graph representing the population are studied as proxies of the properties of the individuals, their interactions, and the population itself. For example, the degree, various centrality measures, clustering coefficients, or the eigen values (PageRank) of the nodes have been used to determine the relative importance of the individuals, *e.g.,* [17, 42]. Betweenness centrality has been used to identify cohesive communities [33] and the distributions of shortest path lengths employed to measure the "navigability" of the network [66]. These and many other graph theoretic measures have been translated to many social properties [50, 56, 57].

The ability of an individual to block a process spreading over a network can be seen as another such social property. Graph measures such as clustering and assortative mixing coefficients have been used to design local vaccination strategies [40]. However, it is not clear that those are the best network measures to be used as an indicator of a node which is a good blocker. In this paper we evaluate the power of all the standard network measures of a node to indicate the blocking ability of the corresponding individual. Moreover, we extend the standard static measures to reflect the dynamic nature of the underlying network. We examine the following measures: degree, average degree, betweenness, closeness centralities and clustering coefficient. We modify those to incorporate the time ordering of the interactions.

We use the following terms interchangeably in this paper: individuals or nodes are the vertices of the network and interactions are edges that can be both

directed or undirected. Neighbors of a node, $N(.)$, are the set of nodes adjacent to it. The subscript T with a function name indicates the dynamic variant of the function.

We now state the standard network measures for aggregate networks and define corresponding measures for dynamic networks. We focus first on the global measures that summarize the entire network and then address local measures that characterize a node.

Global Structural Properties.

Density is the proportion of the number of edges $|E|$ present in a network relative to the possible number of edges $\binom{|V|}{2}$.

$$D(G) = \frac{|E|}{\binom{|V|}{2}}. \tag{6}$$

Dynamic Density is the average density of an observed time snapshot.

$$D_T(G) = \frac{1}{T} \sum_{1<t\leq T} D(G_t). \tag{7}$$

In the example in Figure 1, the density of the aggregate network in (e) is 2/3. However, the dynamic density of the networks (a), (b), and (c) is 1/3 while the dynamic density of (d) is 2/3.

Path between a pair of nodes u, v is a sequence of distinct nodes $u = v^1, v^2, \ldots, v^p = v$ with every consecutive pair of nodes connected by an edge$(v^i, v^{i+1}) \in E$.

Temporal Path between u, v is a time respecting path in a dynamic network. It is a sequence of nodes $u = v^1, \ldots, v^p = v$ where each $(v^i, v^{i+1}) \in E$ is an edge in E_t for some t. Also, for any i, j such that $i + 1 < j$, if $v^i \in V_t$ and $v^j \in V_s$ then $t < s$. The length of a temporal path is the number of timesteps it spans. Note, that this definition allows only immediate neighborhood of a node to be reached within one timestep.

In the example in Figure 1, while there is a path from c to a in the aggregate network (e), there is no temporal path from c to a in the dynamic network (b). All the temporal paths from a to c in the dynamic networks (a)–(d) are of length 2.

Diameter is the length of the longest shortest path. In dynamic networks, it is the length of the longest shortest temporal path.

Local Node Properties.

Degree of a node is the number of its unique neighbors. It is perhaps the simplest measure of the influence of an individual: the more neighbors one has, the higher the chances of reaching a larger proportion of a population.

Dynamic Degree is the change in the neighborhood of an individual over time, the rate at which new friends are gained. Let $N(u_t)$ be the neighborhood of individual u at timestep t. The relative change in the neighborhood is then:[5]

$$\frac{|N(u_{t-1}) \triangle N(u_t)|}{|N(u_{t-1}) \cup N(u_t)|} |N(u_t)|. \tag{8}$$

The Dynamic Degree DEG_T of u is the total accumulated rate of friend addition.

$$DEG_T(u) = \sum_{1<t\leq T} \frac{|N(u_{t-1}) \triangle N(u_t)|}{|N(u_{t-1}) \cup N(u_t)|} |N(u_t)|. \tag{9}$$

Note, that here we consider a friend to be "new" if it was not a friend in the previous timestep. The definition is easily extended to incorporate a longer term memory of friendship. The dynamic degree captures the gregariousness of an individual, an important quality from a spreading perspective.

Dynamic Average Degree is the average over all timesteps of the interactions of an individuals in each timestep:

$$AVG\text{-}DEG(u) = \frac{1}{T} \sum_{1\leq t\leq T} DEG(u_t). \tag{10}$$

where, $DEG(u_t)$ is the size of the neighborhood of u at timestep t.

The dynamic degree, unlike its standard aggregate version, carries the information of the timing of interactions and is sensitive to the order, concurrency and delay among the interactions. For example, in Figure 1, the degree of the node b in the aggregate network (e) is 2. However, its dynamic degree in (a) is 3, in (b) is 1, and in (c) and (d) is 0. The dynamic average degree, on the other hand does not change when the order of interactions in a dynamic network is perturbed. It just tells us the average connectivity of an individual in the observed time period. In all the dynamic networks (a)–(c) the average dynamic degree of b is 1, while in (d) it is 2.

Nodes in Neighborhood (NNk) is the number of nodes in the local k-neighborhood of an individual. The number of nodes in the 1-neighborhood is precisely the degree of an individual. We extend the measure by considering the 2- and 3-neighborhoods of each individual.

Edges in Neighborhood (ENk) is the number of edges in the local k-neighborhood of an individual. We compute the edges in neighborhood for 1-, 2- and 3-hop neighborhoods of each individual. This measure loosely captures the local density of the neighborhood of an individual.

Betweenness of an individual is the sum of fractions of all shortest paths between all pairs of individuals that pass through this individual. It is a parameter that measures the importance of individuals in a network based on their position on the shortest paths connecting pairs of non-adjacent individuals [3, 31, 32].

[5] Here $\triangle$ denotes the symmetric difference of the sets

Dynamic Betweenness of an individual is the fraction of all shortest *temporal paths* that pass through it. Intuitively, the edges in a temporal path appear in the increasing time order. This concept of betweenness incorporates the measure of a delay between interactions as well as the individual being at the right place *at the right time*. 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 [39]. In this paper, for technical reasons, we use the concept of *temporal betweenness*. Let g_{st} be the number of shortest temporal paths between s and t, $g_{st}(u)$ of which pass through u. Then the *temporal betweenness centrality*, $B_T(u)$, of a node u is the sum of fractions of all s-t shortest temporal paths passing through the node u:

$$B_T(u) = \sum_{s \neq t \neq u} \frac{g_{st}(u)}{g_{st}}. \tag{11}$$

Closeness of an individual is the average (geodesic) distance of the individual to any other individual in the network [32, 62].

Dynamic Closeness of an individual is the average *time* it takes from that individual to reach any other individual in the network. Dynamic closeness is based on shortest temporal paths and the geodesic is defined as the time duration of such paths. Let $d_T(u, v)$ be the length of the shortest temporal path from u to v. Following the definition in [62] we define dynamic closeness as follows.

$$C_T(u) = \frac{1}{\sum_{v \in V \setminus \{u\}} d_T(u, v)}. \tag{12}$$

Clustering Coefficient of an individual is the fraction of its neighbors who are neighbors among themselves [58].

Dynamic Clustering Coefficient is the sum of the fractions of an individuals' neighbors who have been friends among themselves in previous timesteps. That is, the dynamic clustering coefficient measures how many of your friends are already friends. Let $CF(u_t)$ be the number of friends of u that are already friends among themselves by timestep t. Then the dynamic clustering coefficient is defined as follows.

$$CC_T(u) = \sum_{0 \leq t < T} \frac{CF(u_t)}{|N(u_t)|(|N(u_t)| - 1)}. \tag{13}$$

Consider the example in Figure 2. The clustering coefficient of all three nodes in the static network is the same and equals to 1. However, the situation in the two dynamic networks is completely different. In network (a) the dynamic clustering coefficient of nodes a and c is 0 while that of the node b is 1. In network (b), on the other hand, the dynamic clustering coefficient of all the nodes is 0 since when b meets a and c they still don't know each other.

Apart from the measures defined above we also compute PageRank [14] of nodes.

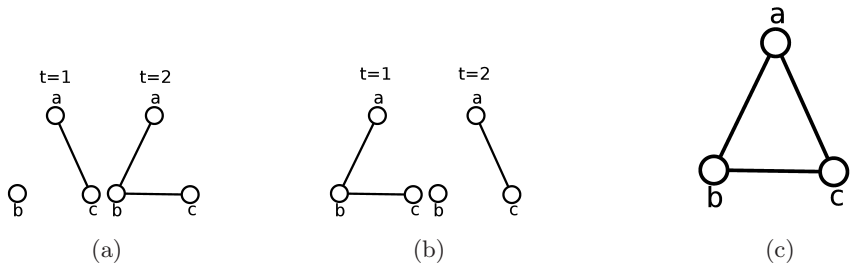

Fig. 2. Example of two dynamic networks (a) and (b) that have the same aggregate network representation (c).

3.5 Spreading Model

A propagation process in a network can be described formally using many models of transmission over the edges in that network. The fundamental assumption of all such models is that the phenomenon is spreading over and only over the edges in the network and, thus, the topology of the network defines the dynamics of the spread. For this paper we use the Independent Cascade model of diffusion in networks. Independent Cascade is one variant of the conditional decision model [34, 65]. The spreading phenomenon cascades through the network bases the the simplifying assumption that each individual base their decision to adopt or reject the spreading phenomenon on the status of each of its neighbors independently. The independent cascade model was first introduced in [34, 35] in the context of word-of-mouth marketing. This is also the most commonly used simple model to study disease transmission in networks [22, 51, 54, 57, 59] and is closely related to the simplest Susceptible-Infectious-Recovered (SIR) models from epidemiology [2]. In the Independent Cascade model, transmission from one individual to another happens independently of interactions those individuals have with all the other individuals.

The Independent Cascade model describes a spreading process in terms of of two types of individuals, active and inactive. The process unfolds in discrete synchronized timesteps. In each timestep, each active individual attempts to activate each of its inactive neighbors. The activation of each inactive neighbor is determined by a probability of success. If an active individual succeeds in affecting any of its neighbors, 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, let $G_D = (V, E)$ be a dynamic network, $A_0 \subseteq V$ be a set of active individuals, and p_{uv} be the probability of influence of u on v. For simplicity, we assume p is uniform for all V and remains fixed for the entire period of simulation. The uniform probability values also ensure that we test how the blocking ability of individuals depends solely on the structure of the network,

controlling for other parameters that may affect this ability. An active individual $u_t \in A_0$ at timestep t tries to activate each of its currently inactive neighbors v_t with a probability p, 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 the still inactive v_{t+i}, it will again try to activate v_{t+i}. The process runs for a finite number of timesteps T. 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 aggregate network in one important aspect. In the aggregate case, each individual u uses all its attempts of activating each of its inactive neighbors v with the same probability p 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 (no recovery in the epidemiological model). It is a particularly important case in the context of identifying blockers since the blocking action is typically done before any recovery.

4 Experimental Setup

We evaluate the effectiveness of each of the network structural measures as indicators of individual's blocking capacity under the Independent Cascade spreading model.

4.1 The Protocol

For each measure and for each dynamic network dataset, we follow the following steps:

1. Order the individuals $0, \ldots, |V-1|$ according to the ranking imposed by the measure.
2. For $i = 0$ to $|V-1|$ do:
   (a) Remove node i from $G = (V, E)$.
   (b) Estimate the extent of spread in $G \setminus i$ by averaging over stochastic simulations of Independent Cascade model initiated at each node in turn, 3000 iterations for each starting node.[6]
   (c) If the extent of spread is less than 10% of the nodes in the original network then STOP.

[6] Which is more than sufficient for the convergence.

We compare the power of each measure to serve as a proxy indicator for the blocking ability of an individual based on the number of individuals that had to be removed in the ordering imposed by that measure in order to achieve this reduction to 10%.

4.2 Probability of activation

We conducted the Independent Cascade spreading experiments on a variety of networks with diverse global structural properties such as density, diameter, and average path length. In each network, we assigned a different probability of activation based on the structure of the network. The probability value that for some networks facilitated propagation of spread to only a small portion of the nodes for other networks resulted in immediate spread to the entire network. The following is the procedure we used to find a meaning full probability of activation for a given network.

1. For a given $G = (V, E)$, run the Independent Cascade Spreading process with $p = 1$. Note, that this is a deterministic process.
2. Calculate the average extent of spread S in $G = (V, E)$. This is the average size of a connected component in G.
3. Rerun the spreading process while setting $p < 1$. Calculate the average extent of spread in the network. Repeatedly reduce p until the average extent of spread is half of S.
4. Set probability of activation for G equal to p.

We use the following measures for comparison: dynamic and aggregate versions of degree, betweenness, closeness centralities, and clustering coefficient, as well as the average dynamic degree (turnover rate). For the global measures of betweenness and closeness we locally approximate them within 1-, 2-, and 3-hops neighborhoods. For the datasets with directed interactions we also use page rank and approximate it within 1-, 2-, and 3-neighborhoods as well. We also rank individuals based on their neighbors within 1-, 2-, and 3- hops of nodes and edges. Overall, we experimented with 26 different measures.

We compare the structural measures to a random ordering of nodes as an upper bound and the best blockers identified by an exhaustive search as the lower bound.

4.3 Lower Bound: Best Blockers

We identify the best blockers one at a time using exhaustive search over all the individuals. To find one best blocker, we remove each individual, in turn, from the network and estimate the extent of spread using stochastic simulations of the Independent Cascade model in the remaining network. The best blocker, then, is the individual whose removal results in the minimum extent of spread after removal. We then repeat the process with the remaining individuals. This process imposes another ranking on the nodes.

Optimally, one needs to identify the *set* of top k blockers. However, this problem is computationally hard and an exhaustive search is infeasible. We have conducted limited experiments on the datasets considered in this paper and in all cases the set of iterative best k blockers equals to the set of top k blockers. This preliminary result warrants future investigation and rigorous evaluation.

5 Datasets

We now describe the datasets used in the experiments.

Grevy's: Populations of Grevy's zebras (*Equus grevyi*) were observed by biologists [29, 30, 61, 63] over a period of June–August 2002 in the Laikipia region of Kenya. Predetermined census loops were driven on a regular basis (approximately twice per week) and individuals were identified by unique stripe patterns. Upon sighting, an individual's GPS location was taken. In the resulting dynamic network, each node represents an individual animal and two animals are interacting if their GPS locations are the same. The dataset contains 28 individuals interacting over a period of 44 timesteps.

Onagers: Populations of wild asses (*Equus hemionus*), also known as onagers, were observed by biologists [61, 63] in the Little Rann of Kutch, a desert in Gujarat, India, during January–May 2003. These data are also obtained from visual scans, as in Grevy's zebra case. The dataset contains 29 individuals over 82 timesteps.

DBLP: This data set is a sample of the *Digital Bibliography and Library Project* [49]. This is a bibliographic dataset of publications in Computer Science. We use a cleaned version of the data from 1967–2005. In the dynamic network each node represents an individual author and two authors are interacting if they are co-authors on a paper. A year is one timestep. The sample we used contains 1374 individuals and 38 timesteps. We use this dataset to compare the dynamic and the static networks.

The DBLP dataset is sparse, with many small connected components. In fact, the average size of a connected component (using temporal paths) is $.03 \times |V|$. Thus, the expected extent of spread in this network cannot exceed 3%. For DBLP we set the stopping criterion for removing blockers from the network at 1% of the population being affected, rather than the 10% used for other datasets.

Reality Mining: The Reality Mining experiment is one of the largest mobile phone projects attempted in academia. These are the data collected by MIT Media Lab at MIT [25]. They have captured communication, proximity, location, and activity information from 100 subjects at MIT over the course of the 2004-2005 academic year. These data represent over 350,000 collective hours ($\sim$ 40 years) of human behavior.

Reality Mining data are collected by recording the bluetooth scans of each device every five minute. We have quantized the data to 4 hours interval for the dynamic network representation of the network based on the analysis by [20].

Enron: The Enron e-mail corpus is a publicly available database of e-mails sent by and to employees of the now defunct Enron corporation[7]. Timestamps, senders and lists of recipients were extracted from message headers for each e-mail on file. We chose a day as the timestep, and a directed interaction is present if an e-mail was sent between two individuals.
We used the version of the dataset restricted to the 150 employees of Enron organization who were actually subpoenaed. The raw Enron corpus contains 619,446 messages belonging to 158 users [1, 45].

UMass: Co-location of individuals in a population of students at the University of Massachusetts Amherst; data collected via portable motes(available with a full description at http://kdl.cs.umass.edu/data/msn/msn-info.html).

The following table provides a summary of the statistics of the networks we use in our experiments.

	V	E	T	D	D_T	d	d_T	p	p_T	r	r_T
Grevy's	28	779	44	0.30	0.52	4	36	1.84	4.81	518	432
Onagers	29	402	82	0.36	0.24	3	74	1.66	7.51	756	617
DBLP	1374	2262	38	0.002	0.09	15	37	5.54	5.12	900070	58146
Enron	147	7406	701	0.04	0.14	6	618	2.66	461.24	19620	16474
MIT	96	67107	2940	0.68	0.18	2	315	1.32	4.21	9120	9114
UMass	20	2664	693	0.72	0.35	2	8	1.28	3.71	380	374

Table 1. Dynamic network dataset statistics. Here V is the number of individuals, E is the number of edges, T is the number of timesteps, D is density and D_T is dynamic density, d is the diameter within a connected component and d_T is the dynamic diameter,p is average shortest path length and p_T is the average temporal shortest path length, and r is no. of reachable pairs and r_T is the number of temporally reachable pairs.

6 Results and Discussion

For each of the datasets we have evaluated all the structural network measures to determine how effectively they serve to identify good blockers. To recap, we rank nodes by each measure and remove them from the network in that order. After removing each node we measure the expected extent of spread in the network using simulations. We compare the effect of each measure's ordering to that of a random ordering and the brute force best blockers ordering. Figure 3 shows results for two datasets, Onagers and Enron, that are representative of the results on all the datasets. The results for the other datasets are omitted due to space limitations. For all the plots, the x-axis is the number of individuals removed

[7] Available with a full description at http://www.cs.cmu.edu/∼enron/

and the y-axis shows the corresponding extent of spread. The lower the extent of spread after removal, the better is the blocking capacity of the individuals removed. Thus, the curves lower on the plot correspond to measures that serve as better indicators of individuals' blocking power.

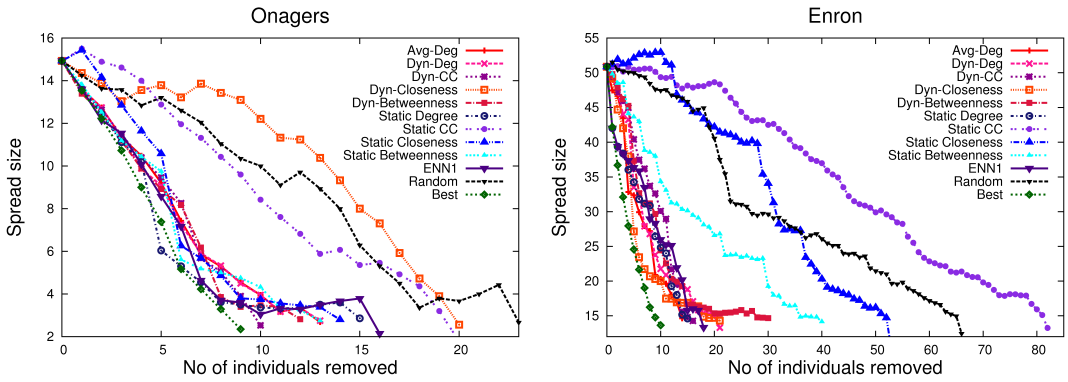

Fig. 3. [Best viewed in color.] Comparison of the reduction of extent of spread after removal of nodes ranked by various measures in Onagers and Enron datasets.

The comparison of all the measures showed that four measures performed consistently well as blocker indicators: degree in aggregate network, the number of edges in the immediate aggregate neighborhood (local density), dynamic average degree, and dynamic clustering coefficient. This is good news from the practical point of view of designing epidemic response strategies since all the measures are simple, local, and easily scalable. Figure 4 shows the results of the comparison of those four best measures, as well as the best possible and random orderings, for all the datasets. Surprisingly, while the local density and the dynamic clustering coefficients seem to be good indicators, the aggregate clustering coefficient turned out to be the worst, often performing worse than a random ordering. Betweenness and closeness measures performed inconsistently. PageRank did not perform well in the only dataset with directed interactions (Enron)[8]. As seen in Figure 4, the ease of blocking the spread depends very much on the structure of the dynamic network. In the two bluetooth datasets, MIT Reality Mining and UMass, all orderings, including the random, performed similarly. Those are well connected networks, as evident by the large difference between the dynamic diameter and the average shortest temporal path. The only way to reduce the extent of spread to below 10% of the original population seems to be trivially removing nearly 90% of the individual population. On the other hand, Enron and DBLP, the sparsely connected datasets, show the opposite trend of being easily blockable by a good ranking measure.

[8] On undirected graphs, PageRank is equivalent to degree in aggregate network

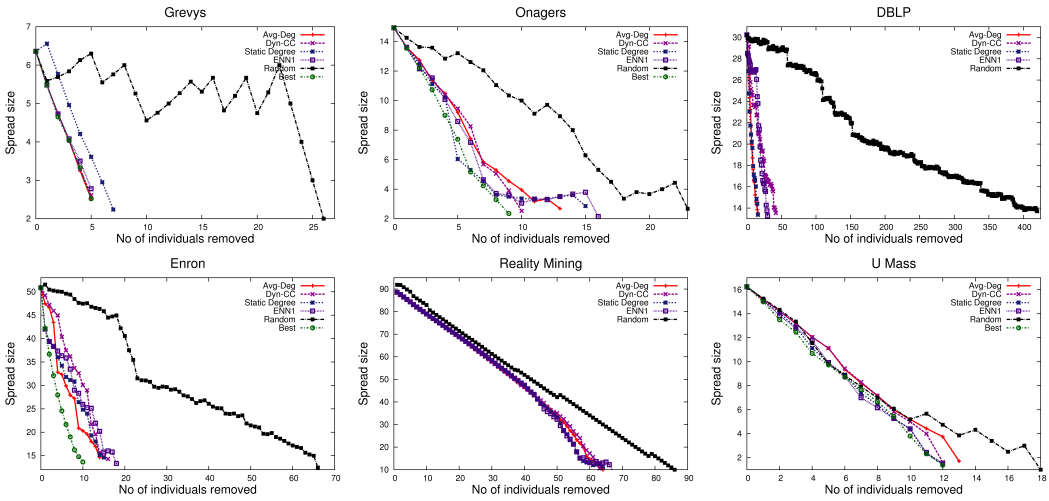

Fig. 4. [Best viewed in color.] Comparison of the reduction of the extent of spread after removal of nodes ranked by the best 4 measures. The x-axis shows the number of individuals removed and the y-axis shows the average spread size after the removal of individuals.

When rankings of different measures result in a similar blocking ability we ask whether it is due to the fact that the measures rank individuals in a similar way and, thus, identify the same set of good blockers or, rather, different measures identify different sets of good blockers. To answer this question, we compared the sets of the top ranked blockers identified by the four best measures as well as the best possible ordering. We compute the average rank difference between the sets of individuals ranked top by every two measures. Table 2 shows the pairwise difference in ranks. In general, there is little correspondence between the rankings imposed by various measures. The only strong relationship, as expected, is between the number of edges in the neighborhood of a node and its degree in the aggregate network.

We further explore the difference in the sets of the top ranked individuals by computing the size of the common intersection of all the top sets ranked by the four measures and the best possible ranking. We use the size of the set determined by the best possible ordering as the reference set size for all measures. Table 3 shows the size of the common intersection for all datasets. Again, we see a strong effect of the structure of the network. The MIT Reality Mining and the UMass datasets have the largest intersection size. On the other hand, in DBLP the four measures produced very different top ranked sets, yet all four measures were extremely good indicators of the blockers. In other networks, while there are some individuals that are clearly good blockers according to all measures, there is a significant difference among the measures. Overall, these results lead to two future directions: 1) investigating the effect of the overall network structure

Dataset	Best vs AvgDEG	Best vs DynCC	Best vs DEG	Best vs ENN1	AveDEG vs DynCC	AvgDEG vs DEG	AvgDEG vs ENN1	DynCC vs DEG	DynCC vs ENN1	DEG vs ENN1
Grevy's	4.5	4.64	4.79	3.86	4.5	2.86	2.64	5.57	5	1.14
Onagers	3.59	4.48	3.31	3.52	4.69	4.14	2.97	6.07	6	2
DBLP	-	-	-	-	430.76	71.3	78.49	434.21	428.25	77.22
Enron	21.95	50.01	27.29	21.02	46.37	22.56	21.93	44.35	44.95	25.32
MIT	-	-	-	-	4.88	14.4	14.48	14.33	14.27	2.25
UMass	4.6	4.6	3	2.7	0	3.3	3.1	3.3	3.1	1

Table 2. Average rank difference between the rankings induced by every two of the best four measures.

Dataset	Set size	Inter. size	Inter. frac
Grevy's	5	2	.40
Onagers	9	3	.33
DBLP	16	0	0
Enron	13	4	.31
Reality Mining	60	48	.80
UMass	12	10	.83

Table 3. The size of the common intersection of all the top sets ranked by the four measures and the best ranking. Set size is the size of the sets determined by the best blocking ordering. The size of the intersection is the number of the individuals in the intersection and the Intersection fraction is the fraction of the intersection of the size of the set.

on the "blockability" of the network; and 2) designing consensus techniques that combine rankings by various measures into a possible better list of blockers.

7 Conclusions and Future Work

In this paper we have investigated the task of preventing a dynamic process, such as disease or information, from spreading through a network of social interactions. We have formulated the problem of identifying good blockers: nodes whose removal results in the maximum reduction in the extent of spread in the network. In the absence of good computational techniques for finding such nodes efficiently, we have focused on identifying structural network measures that are indicative of whether or not a node is a good blocker. Since the timing and order of interactions is critical in propagating many spreading phenomena, we focused on explicitly dynamic networks. We, thus, extended many standard network measures, such as degree, betweenness, closeness, and clustering coefficient,

to the dynamic setting. We also approximated global network measures locally within a node's neighborhood. Overall, we considered 26 different measures as candidate proxies for the blocking ability of a node.

We conducted experiments on six dynamic network datasets spanning a range of contexts, sizes, density, and other parameters. We compared the extent of spread while removing one node at a time according to the ranking of nodes imposed by each measure. Overall, four structural measures performed consistently well in all datasets and were close to identifying the overall best blockers. These four measures were node degree, number of edges in node's neighborhood, dynamic average degree, and dynamic clustering coefficient. The traditional aggregate clustering coefficient and dynamic closeness performed the worst, often worse than a random ordering of nodes. All four best measures are local, simple, and scalable, thus, potentially can be used to design good practical epidemic prevention strategies. However, before such policy decisions are made, we need to verify that our results hold true in other, larger and more complete datasets and for realistic disease spread models.

The striking disparity between the performance of the dynamic and aggregate clustering coefficient indicates the necessity of taking the dynamic nature of interactions explicitly into consideration in network analysis. Moreover, this disparity justifies the extension of traditional network measures and methods to the dynamic setting. In future work, we plan to further investigate the informativeness of a range of dynamic network measures in various application contexts.

We have also compared the sets of nodes ranked at the top by various measures. Interestingly, in the networks in which it was difficult to block dynamic spread, all the measures resulted in very similar rankings of individuals. In contrast, in the networks where the removal of a small set of individuals was sufficient to reduce the spread significantly, the best measures gave very different rankings of individuals. Thus, there seems to be a dichotomy in the real-world networks we studied. On one hand, there are dense networks (e.g. MIT Reality Mining and UMass datasets) in which it is inherently challenging to block a spreading process and all measures perform similarly badly. On the other hand, there are sparse networks where it seems to be easy to stop the spread and there are many ways to do it. In future work, we will investigate the specific global structural attributes of a network that delineate this difference between networks for which it is hard or easy to identify good blockers.

The comparison of the top ranked sets also shows that while there may be some common nodes ranked high by all measures, there is a significant difference among the measures. Yet, all the rankings perform comparably well. Thus, there is a need to test a consensus approach that combines the sets ranked top by various measures into one set of good candidate blockers. This is similar to combining the top k lists returned as a web search result [27].

This paper focused on the practical approaches to identifying good blockers. However, the theoretical structure of the problem is not well understood and so far has defied good approximation algorithms. Recent developments in the

analysis of nonmonotonic submodular functions [28, 64] may be applicable to variants of the problem and may result in good approximation guarantees.

References

1. J. Adibi. Enron email dataset. [ONLINE]. http://www.isi.edu/∼adibi/Enron/Enron.htm.
2. R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992.
3. J. Anthonisse. The rush in a graph. *Amsterdam: Mathematische Centrum*, 1971.
4. J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*, 72(6):1077–1093, Sept. 2006.
5. S. Asur, S. Parthasarathy, and D. Ucar. An event-based framework of characterizing the evolutionary behavior of interaction graphs. In *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007.
6. A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614, August 2002.
7. E. Berger. Dynamic monopolies of constant size. *J. Combin. Theory Series B*, 83:191–200, 2001.
8. N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 301–310, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.
9. T. Berger-Wolf, W. Hart, and J. Saia. Discrete sensor placement problems in distribution networks. *Mathematical and Computer Modelling*, 2005.
10. J. Berry, L. Fleischer, W. Hart, C. Phillips, and J. Watson. Sensor placement in municipal water networks. *Journal of Water Resources Planning and Management*, 131(3), 2005a.
11. J. Berry, W. Hart, C. Phillips, J. G. Uber, and J. Watson. Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management*, 132(4):218–224, 2006.
12. K. Börner, L. Dall'Asta, W. Ke, and A. Vespignani. Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. In *Complexity, Special issue on Understanding Complex Systems*, 10(4):57–67, 2005.
13. K. Börner, J. Maru, and R. Goldstone. The simultaneous evolution of author and paper networks. *PNAS*, 101(Suppl 1):5266–5273, 2004.
14. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW7: Proceedings of the 7th International Conference on World Wide Web 7*, pages 107–117, Amsterdam, The Netherlands, The Netherlands, 1998. Elsevier Science Publishers B. V.
15. A. Broido and K. Claffy. Internet topology: connectivity of IP graphs. In *Proceedings of SPIE ITCom*, 2001.
16. K. Carley. Communicating new ideas: The potential impact of information and telecommunication technology. *Technology in Society*, 18(2):219–230, 1996.
17. I. Carreras, D. Miorandi, G. Canright, and K. Engøo-Monsen. Eigenvector centrality in highly partitioned mobile networks: Principles and applications. *Studies in Computational Intelligence(SCI)*, 69:123–145, 2007.

18. L. Chen and K. Carley. The impact of social networks in the propagation of computer viruses and countermeasures. *IEEE Trasactions on Systems, Man and Cybernetics*, forthcoming.
19. N. Chen. On the approximability of influence in social networks. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1029–1037, 2008.
20. A. Clauset and N. Eagle. Persistence and periodicity in a dynamic proximity network. Unpublished manuscript.
21. R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 2003.
22. Z. Dezsö and A.-L. Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65(055103(R)), 2002.
23. P. Domingos. Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20:80–82, 2005.
24. P. Domingos and M. Richardson. Mining the network value of customers. In *Seventh International Conference on Knowledge Discovery and Data Mining*, 2001.
25. N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Journal of Personal and Ubiquitous Computing*, 2006.
26. S. Eubank, H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:429:180–184., Nov 2004. Supplement material.
27. R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. In *SODA '03: Proc, 14th ACM-SIAM Symposium on Discrete Algorithms*, pages 28–36, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.
28. U. Feige, V. Mirrokni, and Vondrák. Maximizing non-monotone submodular functions. In *Foundations of Computer Science(FOCS)*, 2007.
29. I. R. Fischhoff, S. R. Sundaresan, J. Cordingley, H. M. Larkin, M.-J. Sellier, and D. I. Rubenstein. Social relationships and reproductive state influence leadership roles in movements of plains zebra (*Equus burchellii*). *Animal Behaviour*, 73(5):825–831, 2007.
30. I. R. Fischhoff, S. R. Sundaresan, J. Cordingley, and D. I. Rubenstein. Habitat use and movements of plains zebra (*Equus burchelli*) in response to predation danger from lions. *Behavioral Ecology*, 18(4):725–729, 2007.
31. L. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.
32. L. C. Freeman. Centrality in social networks: I. conceptual clarification. *Social Networks*, 1:215–239, 1979.
33. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:8271–8276, 2002.
34. J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.
35. J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review*, 2001.
36. M. Granovetter. The strength of weak ties. *American J. Sociology*, 78(6):1360–1380, 1973.
37. M. Granovetter. Threshold models of collective behavior. *American J. Sociology*, 83(6):1420–1443, 1978.
38. D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04: Proc. 13th Intl Conf on World Wide Web*, pages 491–501, New York, NY, USA, 2004. ACM Press.

39. Habiba, C. Tantipanananadh, and T. Y. Berger-Wolf. Betweenness centrality in dynamic networks. Technical Report 2007-19, DIMACS, 2007.
40. P. Holme. Efficient local strategies for vaccination and network attack. *Europhys. Lett.*, 68(6):908–914, 2004.
41. J. Hopcroft, O. Khan, B. Kulis, and B. Selman. Natural communities in large linked networks. In *Proc. 9th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining*, pages 541–546, 2003.
42. F. Jordán and J. Benedek, Z.and Podani. Quantifying positional importance in food webs: A comparison of centrality indices. *Ecological Modelling*, 205:270–275, 2007.
43. D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. *J. Comput. Syst. Sci.*, 64(4):820–842, 2002.
44. D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining*, 2003.
45. B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. In *Proceedings of the European Conference on Machine Learning*, 2004.
46. J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 228–237, New York, NY, USA, 2006. ACM Press.
47. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, and J. VanBriesen. Cost-effective outbreak detection in networks. In *Proc. 13th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining*, 2007.
48. K. Lewin. *Principles of Topological Psychology.* New York: McGraw Hill, 1936.
49. M. Ley. Digital bibliography & library project (DBLP). `http://dblp.uni-trier.de/`, December 2005. A digital copy of the databse has been provided by the author.
50. F. Liljeros, C. Edling, and L. N. Amaral. Sexual networks: Implication for the transmission of sexually transmitted infection. *Microbes and Infection*, 2003.
51. R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Physical Review E*, 64(066112), 2001.
52. J. Moody. The importance of relationship timing for diffusion. *Social Forces*, 2002.
53. Y. Moreno, M. Nekovee, and A. F. Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 69(6):066130, 2004.
54. M. Morris. Epidemiology and social networks:modeling structured diffusion. *Sociological Methods and Research*, 22(1):99–126, 1993.
55. E. Mossel and S. Roch. On the submodularity of influence in social networks. In *The Annual ACM Symposium on Theory of Computing(STOC)*, 2007.
56. M. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
57. M. E. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(016128), 2002.
58. M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, 64:016131, 2001.
59. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, Apr 2001.
60. E. M. Rogers. *Diffusion of Innovations.* Simon & Shuster, Inc., 5th edition, 2003.
61. D. I. Rubenstein, S. Sundaresan, I. Fischhoff, and D. Saltz. Social networks in wild asses: Comparing patterns and processes among populations. In A. Stubbe,

P. Kaczensky, R. Samjaa, K. Wesche, and M. Stubbe, editors, *Exploration into the Biological Resources of Mongolia*, Martin-Luther-University Halle-Wittenberg, (10): 159–176, 2007.
62. G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31:581–603, 1966.
63. S. R. Sundaresan, I. R. Fischhoff, J. Dushoff, and D. I. Rubenstein. Network metrics reveal differences in social organization between two fission-fusion species, Grevy's zebra and onager. *Oecologia*, 151:140–149, 2007.
64. T. Vredeveld and J. Lenstra. On local search for the generalized graph coloring problem. *Operations Research Letters*, 31:28–34, 2003.
65. D. Watts. A simple model of global cascades on random networks. *PNAS*, 99:5766–5771, 2002.
66. D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.
67. H. P. Young. Innovation diffusion and population heterogeneity. Working paper, 2006.
68. D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Phys. Rev. E*, 65(4):041908, Mar 2002.