DIMACS Technical Report 2007-19
November 2007

Betweenness Centrality Measure in Dynamic Networks

by

Habiba[1]
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607

Chayant Tantipathananandh[2]
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607

Tanya Y. Berger-Wolf[3]
Department of Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607

[1]Work supported in part by the Fulbright Fellowship.
[2]Work supported in part by the Microsoft award 14936.
[3]Work supported in part by the Microsoft award 14936.

ABSTRACT

In this paper we propose three methods of measuring betweenness of individuals in networks which are best modeled as graphs with explicit *time ordering* on their edges. The betweenness centrality index is one of the basic measure in the analysis of social networks, but most of the work done for measuring the betweenness index of individuals is based on the *aggregate* representation of the network. Many network problems are based on fundamental relationship involving time. We incorporate the time factor in the aggregate graph representation of social networks to create *dynamic* networks. We define and measure the betweenness in this dynamic framework. We compare the three betweenness with the standard betweenness measure for the same network. We show that by incorporating the exact times of interactions among individuals in a network, we can better study the betweenness of individuals in the underlying network.

1 Introduction

In this paper, we extend the study of betweenness centrality of individuals in social networks to networks which are explicitly dynamic.

The idea of representing societies as networks of interacting individuals dates back to Lewin's earlier work of group behavior [27]. In this model individuals are represented as nodes and interactions in the observed time period are represented as edges between individuals. This representation provides an *aggregate* view of the population. Using this aggregate network model, the structure and properties of many social networks have been studied from different perspectives [3, 5, 6, 8, 22]. However, this model and other similar analogies do not explicitly consider the temporal aspect of the network.

Many network problems are based on fundamental relationship involving time. Consider, for example the problem of modeling the flow of information [9, 10, 12, 21, 24], spreading of diseases in a population [13, 16, 23, 26, 30], viral marketing [14, 15], and transportation networks [1, 4]. For all such domains the evolution of the network over time plays a key role.

In this paper, we present a *dynamic network* model which explicitly evolve over time. The central motivation of our work in this paper is the following question: How does the betweenness centrality for individuals in a network varies when we impose the additional constraint of time ordered interactions?

Betweenness centrality is a parameter that measures the importance of individuals in a network based on their position in the shortest paths connecting a pair of non-adjacent individuals [2, 19, 20]. We present different flavors of the traditional betweenness centrality concept in dynamic networks based on position, time, and duration of interactions among individuals (Section 6). In addition, we experimentally compare the betweenness centrality of individuals in the aggregate representation with that of the dynamic representation for the same networks (Section 7). Our results show that the two network models are different based on the betweenness index of individuals in the network. The dynamic network takes into consideration the additional temporal aspect of the underlying network, hence, it is a better representation of the network.

2 Background- Graph Theoretic Notation of Social Networks

A social network can be viewed in several ways. One of the most useful views is as a *graph* of *individuals* joined by edges representing *interactions* among individuals. The types of interactions vary with domains. For example, they can be co-membership [28], proximity [29], or informal groups [21] among many other types of interactions.

In an instance of a social network graph we can study a single or multiple types of interactions. In this work, for a social network, we consider only a single type of interaction among the individuals.

An interaction is *direct*, if two individuals are connected by an edge. These individuals are called adjacent individuals. An interaction is *indirect*, when two individuals are not adjacent. In terms, for any two non-adjacent individuals, there is a *path* between that pair of individuals if they can be connected through one of more adjacent individuals [33]. The shortest path between a pair of non-adjacent individuals is a path with the least number of intermediate individuals or of the smallest weight, if weights are assigned to edges. The length of such a path is known as the *geodesic* of the pair of individuals [33]. We focus on quantifying the importance of those individuals that lie on the shortest paths between non-adjacent individuals. Anthonisse [2] and later Freeman [19, 20] were the first to quantify the importance of an individual based on its presence on the shortest paths of other individuals. This is known as the betweenness or betweenness centrality of an individual.

In this paper, we present several methods to measure the betweenness centrality of individuals in *dynamic networks*– which are explicitly based on the time ordering of interactions. In the next section we formally define such networks.

3 Dynamic Networks

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 $\{1, \ldots, T\}$.

A dynamic network can be visualized 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.

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 a snapshot of a static network at time t. $V_t \subseteq V$, is a subset of individuals V observed at time t. $(u_t, v_t) \in E_t$ if individuals u and v have interacted at time t and for all $v \in V$ and $t \in \{1, \ldots, T-1\}$, $(v_t, v_{t+1}) \in E$ are directed self edges of individuals across timesteps.*

A dynamic network $G = \langle G_1, \ldots, G_T \rangle$ 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 the above definition it is assumed that an interaction between a pair of individuals takes place within one timestep. Figure 1(a) illustrates a dynamic network of four individuals interacting over five timesteps. The solid line edges represent interactions among individuals in a timestep. The dotted lines are directed self edges from an individual to itself across timesteps. Empty circles are individuals observed during a timestep and filled circles are individuals not observed in a particular timestep.

The above dynamic network model is equivalent to an undirected multigraph representation [25].

We define $G = (V, E)$ as an undirected graph of V individuals and edge set E. Each edge $(u, v) \in E$ is labeled with a time label $\lambda(u, v)$ specifying the time at which its endpoints u and v have interacted. Thus, one can view a dynamic network as the pair (G, λ), where λ is a function from the edge set to integers between 1 and T. We call λ as a *time labeling* of G.

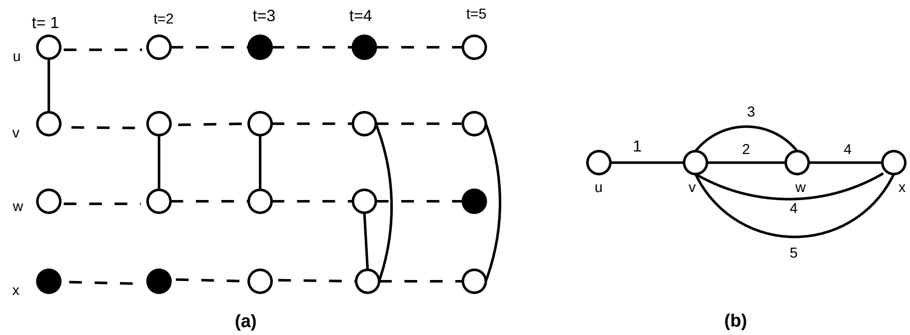

Figure 1: (a)A time series dynamic network of 4 individuals interacting over 5 timesteps. (b) Dynamic labeled multigraph of 4 individuals interacting over 5 timesteps.

Definition 2. *Let $\{1, \ldots, T\}$ be a finite set of discrete timesteps. Let $V = \{1, \ldots, n\}$ be a set of individuals observed at any time between 1 and T. E is the multiset of edges that occur during the time $\{1, \ldots, T\}$. A function λ maps each edge to an integer value between $\{1, \ldots, T\}$, $\lambda : E \to \{1, \ldots, T\}$. A* dynamic labeled multigraph*, is a pair (G, λ). Where $G = (V, E)$ and λ is a function that maps each edge in E to the timestep in which that edge occurred.*

Figure 1(b) gives the multigraph representation of the dynamic network in Figure 1(a).

Proposition 1. *The multigraph representation (G, λ) of dynamic networks is equivalent to the time series representation $G = \langle G_1, \ldots, G_T \rangle$ of dynamic networks.*

Proof.

- The time series representation of a dynamic network can be reduced to the labeled multigraph representation of a dynamic network. For the trivial case, if timesteps are equal to one timestep t, that is $G = G_t$ for the time series dynamic network G. Then, the multigraph representation for G is $G'' = (V'', E'')$ where V'' is the set of individuals observed at timestep t and E'' is the set interactions that took place at time t, and $\forall (u'', v'') \in E''$, $\lambda(u'', v'') = t$.

  For the nontrivial case of a subset $G_i, \ldots, G_k$ of the time series dynamic network $G = \langle G_1, \ldots, G_T \rangle$, the equivalent multigraph is $G'' = (V'', E'')$ such that $V'' = \bigcup_{i \le t \le k} V_t$ and $E'' = \sum_{i \le t \le k} E_t$. Each $(u'', v'') \in E''$ has a label $\lambda(u'', v'') = t$ if $(u, v) \in E_t$ where $i \le t \le k$ in the time series graph.

- The labeled multigraph representation of a dynamic network can be reduced to the time series dynamic network.

  Consider a subset $G' = (V', E')$ of a labeled multigraph $G = (V, E)$. We reduce G' to a time series network G'' such that, $\forall (u', v') \in E'$ and $u', v' \in V'$, if $\lambda(u', v') = t$, then, $u', v' \in V_t''$ and $(u', v') \in E_t''$, where, $G_t'' = (V_t'', E_t'')$ is the static graph of individuals and their interactions at time t. The time series graph G'' is the series of static graphs G_t''.

$\square$

In both representations of dynamic networks, we assume that interactions take place within a timestep. The second representation is useful as it makes it easier to use standard graph traversal algorithms like breadth first search and depth first search for different operations.

4 Temporal Paths

The notion of a path is fundamental to most of graph related measures, from connectivity problems and spanning trees to flows and cuts, all are based on paths. Paths that take into account the time labeling on edges are known as *temporal paths* [25].

A temporal path is a *time respecting* path if the time labels of the sequence of edges on the path are nondecreasing. A path is *strictly time respecting* if the time labels of the edges are increasing.

In this paper, we assume paths are strictly time respecting which is implied by our assumption that each interaction takes one timestep.

Definition 3. *A* temporal path $p(u, v)$ *between a pair of individuals* u *and* v*, is a sequence of edges connecting them. Formally,*

$$p(u, v) = \{(v_0 = u, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n), (v_n, v_{n+1} = v)\}$$

such that, $\forall (v_j, v_{j+1}), (v_i, v_{i+1}) \in p(u, v)$ *and* $\forall i < j, \lambda(v_i, v_{i+1}) < \lambda(v_j, v_{j+1})$.

Figure 2 shows all temporal paths between two individuals u and x based on the dynamic network in Figure 1.

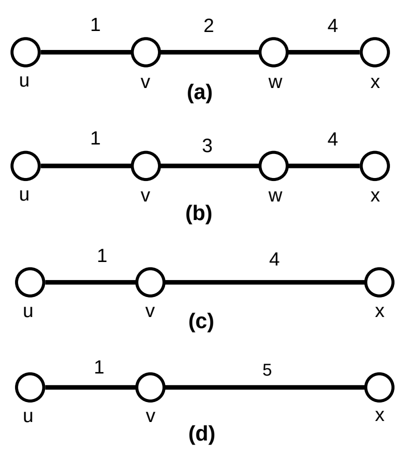

Figure 2: Simple temporal paths between u and x based on the dynamic graph in 1

5 Measures

Centrality relies on measures of geodesics and shortest paths. In this section, we define each of these for dynamic networks. We use the labeled multigraph representation of dynamic networks for defining geodesic, simple temporal path, and link path. We use the time series representation of dynamic networks for the definition of shortest temporal trails.

5.1 Geodesic

Traditionally, geodesic is the shortest distance from one individual to another individual where distance is some function of edges on the shortest path. Usually, distance is the number of edges on the path when the graph is unweighted.

In dynamic networks, geodesic is the length of the shortest temporal path. The length of temporal path is defined in two ways.

1. If we do not take into account delays between any two consecutive interactions then length $d(u, v)$ of a path $p(u, v)$ is the number of edges on the path, assuming that each edge or interaction takes one timestep.

$$d(u, v) = |\{(v_0 = u, v_1), \ldots, (v_{n-1}, v_n = v)\}|$$

   where $(v_0 = u, v_1)$ is the first edge on the path starting at u and $(v_{n-1}, v_n = v)$ is the last edge ending at v.

2. Alternatively, the length of a path is the time it takes for an interaction to take place between a pair of non-adjacent individuals. Simply, the length is the time difference of the first and last interactions on the path. The delays are implicitly embedded in this distance measure between the two individuals.

$$d(u, v) = \lambda(v_{n-1}, v_n = v) - \lambda(v_0 = u, v_1) + 1$$

   where $\lambda(v_{n-1}, v_n = v)$ and $\lambda(v_0 = u, v_1)$ are time labels of the last and first interaction respectively on the path $p(u, v)$ between u and v.

We can see that the first definition of length of a path is equivalent to the length of a path in a simple unweighted aggregate graph. But unlike the aggregate graph, there is a path between any two individuals only if it is temporal.

In this paper, for all measurements involving paths, lengths, and distances we take into account the delay factor. Thus, the geodesic $g(u, v)$ of two non-adjacent individuals u and v in a dynamic network is defined as:

$$g(u, v) = \begin{cases} 1 & : \quad (u, v) \in E \\ \lambda(v_{n-1}, v_n = v) - \lambda(v_0 = u, v_1) + 1 & : \quad (v_{n-1}, v_n), (v_0 = u, v_1) \in E \end{cases}$$

Geodesic between two individuals in a dynamic network can be interpreted differently based on the timing of interactions, the duration of interactions, and the number of individuals involved in the interaction. We show that for all three factors, the value of the geodesic is the same. Hence, all three can be independently used for defining the shortest paths. All these factors have different relevance and meaning in various domains. Hence, we think it is necessary to consider each one of them.

5.2 Shortest Simple Temporal Path

Shortest Simple Temporal Path $p_s(u,v)$, between a pair of individuals u and v is the shortest time respecting path between those individuals, with each intermediate individual present atmost once. The length of the shortest simple temporal path is the geodesic $g(u,v)$.

$$p_s(u,v) = \begin{cases} (u,v) & : \quad (u,v) \in E \\ \{(v_0 = u, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n), (v_n, v_{n+1} = v)\} & : \quad (v_i, v_{i+1}) \in E \end{cases}$$

Figure 2(a), (b), and (c) show the shortest simple path between two individuals u and x for the dynamic network in Figure 1. Note that although Figure 2(d), is a temporal path between u and x, it is not the shortest because of the time it takes for the interaction to take place.

5.3 Shortest Link Path

Shortest Link Path $p_l(u,v)$, is the shortest simple temporal path with minimum number of individuals on the path. The significance of this type of path is that it reduces the dependency of the source and destination on intermediate individuals.

$$p_l(u,v) = min|p_s(u,v)|$$

Figure 2(b) shows the shortest link path between individuals u and x for the dynamic network in Figure 1. Note, that the geodesic of all the paths in Figure 2 are the same but none of the other two are the shortest link paths.

5.4 Shortest Temporal Trails

Another way of measuring the significance of intermediate individuals on the shortest paths of non-adjacent individuals is in terms of the ratio of the time spent on an intermediate node to the total length of the path. Note, that the length of the path is defined in terms of time for the temporal paths. It is obvious that an individual that retains the information for the longest time in a shortest path is more significant than others in some ways. We modify the betweenness centrality definition based on the time spent on each individual in the shortest paths of any pair of individuals.

By accounting for delay at each intermediate individual, the paths are no longer simple. The delay on each intermediate individual is essentially a self loop around that individual. These routes in temporal networks are called temporal trails[1]. A temporal trail $p_d(u, v)$ between u and v is a sequence of edges $p_d(u, v) = \langle (v_0 = u, v_1), (v_1, v_2), \ldots, (v_{n-1}, v_n), (v_n, v_{n+1} = v)\rangle$, such that $\lambda(v_{t-1}, v_t) < \lambda(v_t, v_{t+1})$ for $t = 1, \ldots, n-1$.

The shortest such trails are the one with the smallest geodesic. We call them the *Shortest Temporal Trails*. The important thing is that these shortest trails have the same geodesic as the shortest simple paths and shortest link paths between the same pair of individuals.

Based on the above definitions of dynamic graphs, shortest temporal paths, shortest temporal trails, and geodesics we now give definitions of the betweenness centrality in dynamic networks.

6 Betweenness Centrality in Dynamic Graphs

6.1 Temporal Betweenness Centrality

Temporal betweenness centrality measures the importance of individuals based on their position in the shortest temporal paths of all other nodes. Temporal betweenness centrality $B_T(v)$ of node v is defined as:

Definition 4. *Let g_{st} be the number of shortest temporal paths between s and t. Let $g_{st}(v)$ be the number of shortest temporal paths between s and t that pass through v. Let $B_{T(st)}(v)$ be the fraction of shortest temporal (s, t) paths passing through v. Then, the temporal betweenness centrality $B_T(v)$, of a node v is defined as the sum of fraction of all shortest temporal paths passing through the node v between all pairs of nodes. Formally,*

$$B_T(v) = \sum_{s \neq t \neq v} B_{T(st)}(v) = \sum_{s \neq t \neq v} \frac{g_{st}(v)}{g_{st}}.$$

6.2 Delay-Betweenness Centrality

The delay-betweenness centrality, $B_D(v)$, of individual v, is defined as:

Definition 5. *Let nst_{st} be the number of shortest trails from s to t. Let $nst_{st}(v)$ be the number of time steps of delay of v that all shortest trails from s to t. Let $B_{D(st)}(v)$ denote the delay-dependency of (s, t) on v. The delay-betweenness centrality $B_D(v)$ of a vertex v is the sum of all delay-dependencies $B_{D(st)}(v)$ of all other node pairs (s, t). Formally,*

$$B_D(v) = \sum_{s,t: s \neq t \neq v} B_{D(st)}(v) = \sum_{s,t: s \neq t \neq v} \frac{nst_{st}(v)}{nst_{st}}.$$

[1]A trail may repeat vertices but must not repeat edges

In the dynamic network in Figure 1, there are three shortest temporal trails from u to x as illustrated in Figure 3. Individual v appears for five time steps in these trails. Therefore, the delay-dependency of u and x on v is $B_{D(ux)}(v) = \frac{5}{3}$

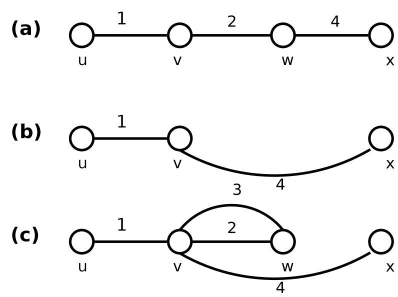

Figure 3: The shortest temporal trails from u to x in the dynamic network in Figure 1.

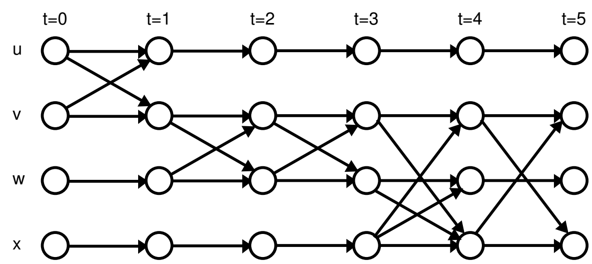

Figure 4: The DAG of the dynamic network in Figure 1.

To compute the delay-dependency $B_{D(st)}(v)$, we create a directed acyclic graph (DAG) $G'' = (V'', E'')$ from the dynamic network $G = (V, E)$ (Section 3) as follows. For each node $v \in V$, we create a node $v_t \in G''$ for $t = 0, 1, \dots, T$ if v is observed at time t. We create v_t for all $t = 0, ..., T$. We add an edge $(v_{t-1}, v_t) \in E$ for $t = 1, 2, \dots, T$. Then, for each undirected edge $(u, v) \in E$ incident on $u, v \in V$, we create two directed edges (u_{t-1}, v_t) and (v_{t-1}, u_t) in E'' where $t = \lambda(e)$. Lastly, for each $v \in V$, we create two dummy nodes, a source v_{in} and a sink v_{out}, and connect them to all v_t of v with edges (v_{in}, v_t) and (v_t, v_{out}). Figure 4 illustrates the DAG for the dynamic network in Figure 1. The dummy nodes are suppressed in Figure 4.

Lemma 1. *In a given dynamic network $G = (V, E)$, each temporal trail from u to v, $u, v \in V$, corresponds to a unique simple path from u_{in} to v_{out} in the DAG of G created as described above. Let n be the length of the temporal trail and n' the length of the corresponding simple path in the DAG. We also have that $n = n' - 2$.*

See appendix A.1 for proof.

Now, we compute $B_{D(st)}(v)$ using the Dynamic Programming in a way similar to the Brandes' algorithm [7]. The delay-betweenness $B_D(v)$ of v can be computed as the sum of the delay-dependency $B_{D(s\bullet)}(v)$ of a node s on a node v. Formally,

$$B_D(v) = \sum_{s:s\neq v} B_{D(s\bullet)}(v).$$

To compute $B_{D(s\bullet)}(v)$, let $D_s = (V'', E''_s)$, $s \in V$, denote the DAG returned by Dijkstra's traversing algorithm [11] starting at $s_{in} \in V''$. To simplify the notation, we use $(u, v) \in D_s$ to denote every edge in E of the multigraph that is on some shortest temporal trail starting at s. Note that we can count the number of temporal trails from s to v, nst_{sv}, while running Dijkstra's traversing algorithm.

Theorem 2. *The following recurrence can be used to compute the delay-dependency of s on v.*

$$B_{D(s\bullet)}(v) = \sum_{(v,w)\in D_s} \frac{nst_{sv}}{nst_{sw}} \left(1 + B_{D(s\bullet)}(w)\right).$$

See Appendix A.2 for proof.

Moreover, we can compute $B_{D(s\bullet)}$ on the DAG instead, using the standard graph traversal algorithm. By traversing the D_s and computing $B_{D(s\bullet)}(v_t)$, $v_t \in V''$ in post-order (i.e., the Dynamic Programming with memoization), we have the delay-dependency of s on node v at time t when it terminates. Then, for each $v \in V$, we compute the summation $B_{D(s\bullet)}(v) = \sum_{t=0}^{T} B_{D(s\bullet)}(v_t)$

In section 7 we experimentally compare the betweenness centrality for the dynamic and aggregate networks.

7 Experimental Results

In this section, we compare the three dynamic betweenness measures with the traditional betweenness measure in networks. We use the following data sets for our experiments.

7.0.1 Grevys

Populations of Grevy's zebras (*Equus grevyi*) were observed by biologists [17, 18, 31, 32] 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 data set contains 28 individuals interacting over a period of 44 timesteps.

Figure 5 shows the comparison of the four betweenness indices of individuals in the Grevys data set. The static betweenness of individuals is very low because the aggregate graph of the 44 timesteps results in two almost complete cliques of individuals. Hence, with aggregate graph we cannot tell which individuals are relatively more important for any kind of interaction. However, the dynamic betweenness indices show that certain individuals are more important. We infer that most of these individuals are not just the ones who are on the shortest paths between other pairs of individuals but they are the ones who appear interact with the source individual in earlier timesteps as compared to the rest of the individuals in the same shortest path. They play a key role in linking the source to the sink and thus are more important.

7.0.2 Onagers

Populations of wild asses (*Equus hemionus*), also known as onagers, were observed by biologists [31, 32] in the Little Rann of Kutch, a desert in Gujarat, India, during January–May 2003. This data is also obtained from visual scans, as in Grevy's zebra case. The data set contains 29 individuals over 82 timesteps.

Figure 6 shows the comparison of the four betweenness indices of individuals in the Onagers data set. This data set is relatively sparse. Hence, overall certain individuals are important even in the aggregate graph for interactions among non-adjacent individuals. The delay betweenness and the shortest simple path betweenness show a similar pattern of behavior in individuals. We can infer from the results, that individuals who are central and important for passing messages in an interaction between a pair of non-adjacent individuals are also retaining it for the longest time. Hence, those individuals are significant than others.

7.0.3 DBLP

This data set is a sample of the *Digital Bibliography and Library Project* [28]. This is a bibliographic data set 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.

Figure 7 shows the comparison of the four betweenness indices of individuals in the DBLP data set. Again, there is a strong co-relation between the delay betweenness and shortest simple temporal path betweenness of individuals. Another interesting observation is that the link betweenness corresponds to the former two betweenness strongly. What we can infer about the network is that it has high clustering coefficient. That is, the individuals who have co-authored among themselves also collaborated quite frequently among their mutual co-authors. Thus, an individual that initiates the collaboration with a certain new individual is vital for introducing the new individual to the rest of the group. The new individual may eventually co-author with many other individuals in the rest of the group but the one which introduces it to others is more significant than others for creating the link. By a new individual, we mean an individual that appears for the first time in the time series of the observed time period as compared to other individuals that have already been observed in previous timesteps.

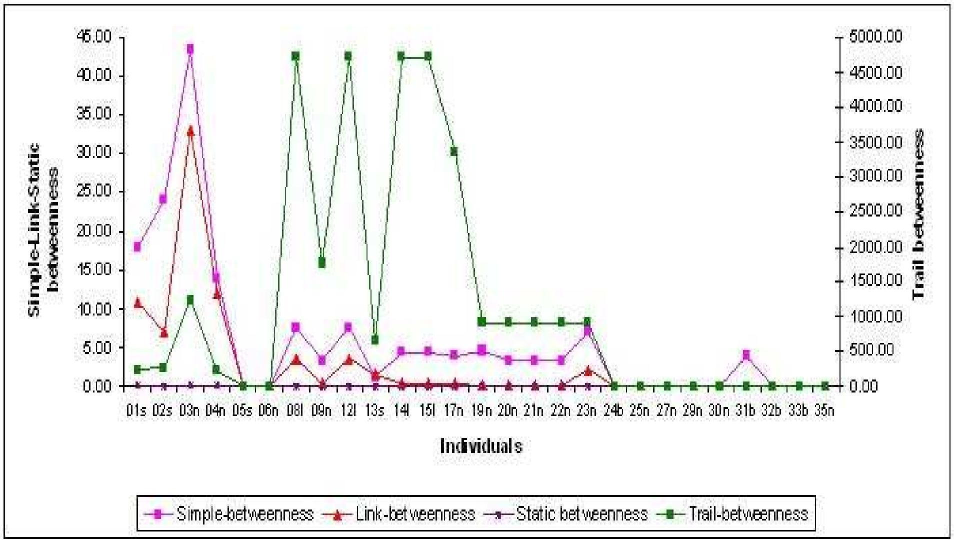

Figure 5: Grevys: Comparison-Simple, Link, Delay, Static Betweenness

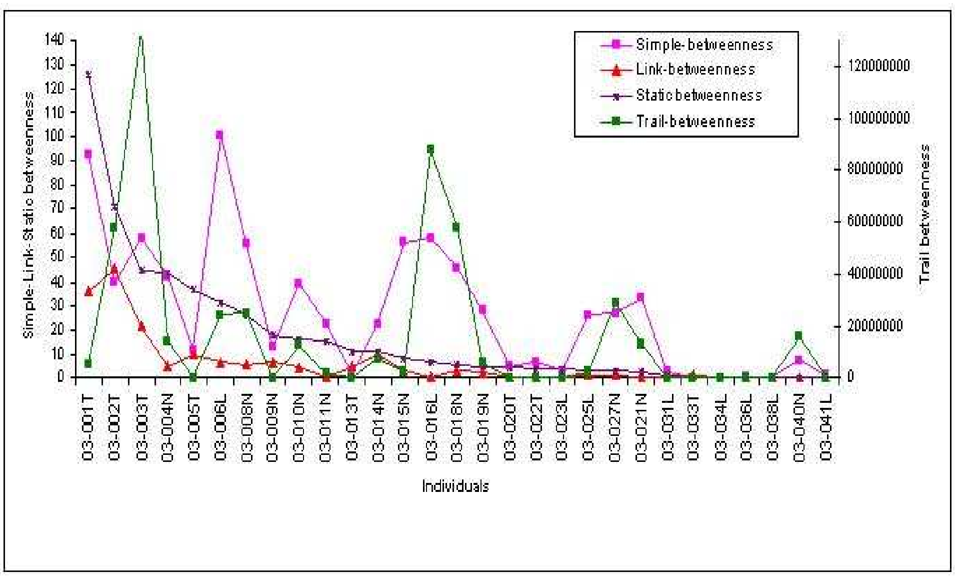

Figure 6: Onagers: Comparison-Simple, Link, Delay, Static Betweenness

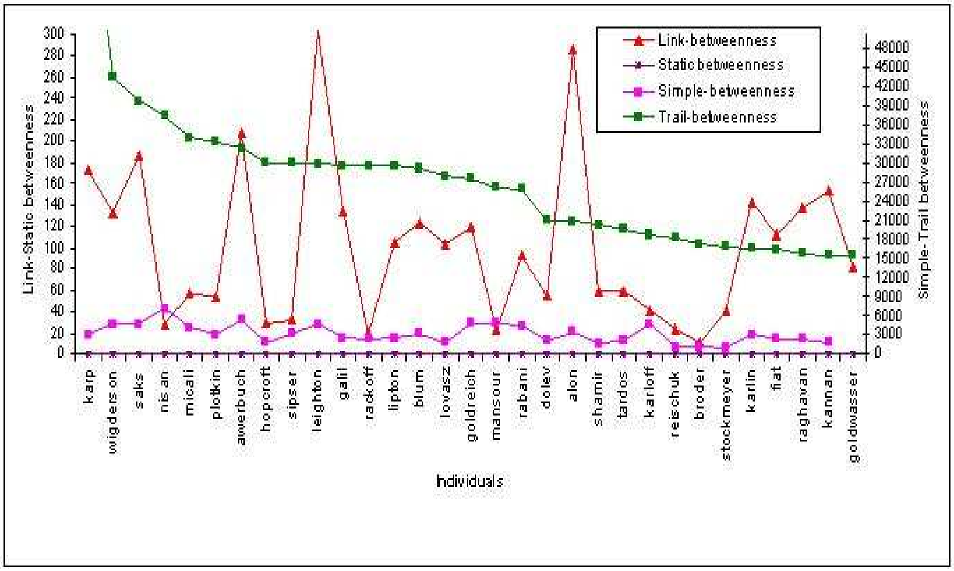

Figure 7: DBLP: Comparison-Simple, Link, Delay, Static Betweenness

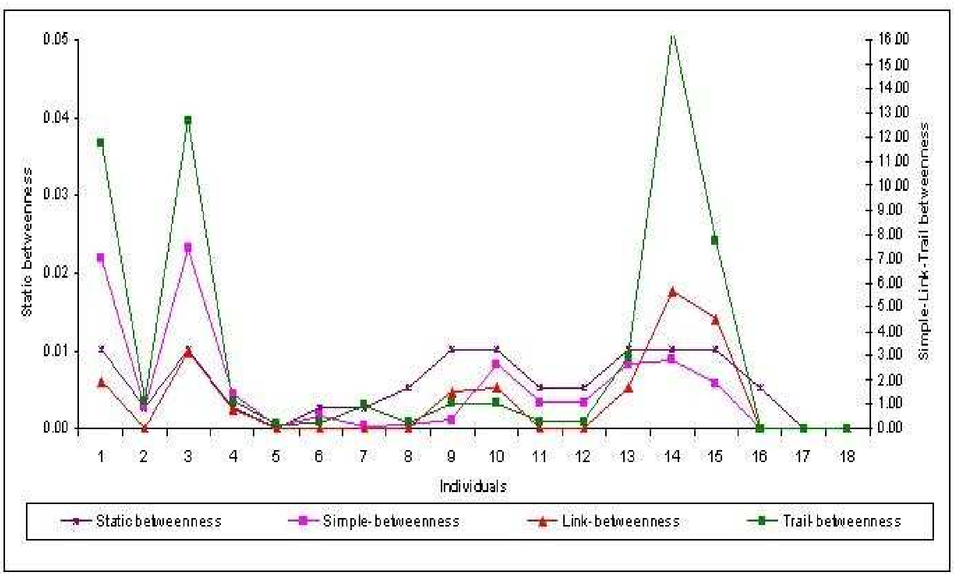

Figure 8: Southern Women: Comparison-Simple, Link, Delay, Static Betweenness

7.0.4 Southern Women

This is the data set of social activities of 18 women observed over a period of nine months in Natchez, Mississippi [21]. This data set is based on the work of five ethnographers in the 1930s who reported a comparative study of social class in black and in white society [21]. During the observed period, various subset of women met in a series of 14 informal social events.

Figure 8 shows the comparison of the four betweenness indices of individuals in the Southern Women data set. We can observe that all four types of betweenness strongly co-relate to each. From this correspondence we deduce that the group of observed individuals have interacted in a persistent pattern. Thus, the topologies of the resulting aggregate and dynamic graph are quite similar.

8 Conclusions

In this paper, we present different methods of measuring the importance of individuals based on their betweenness in dynamic social networks. We argue that for networks that are based on the time ordering of interactions, it is necessary to incorporate the time factor in the graphical representation of the network. We first give formal framework of such dynamic networks. We formulate different methods of measuring basic parameters like paths, trails, and geodesics based on the dynamic networks framework. We then present different methods of evaluating the betweenness centrality of individuals based on shortest temporal paths and shortest temporal trails.

We experimentally compare the betweenness centrality of individuals using shortest simple temporal paths, shortest link paths, and delay betweenness with the traditional betweenness in the static graphs for same networks. We graphically show the variations in the four measures.

We believe that for problems that are best modeled using a network with an explicit time ordering on its edges, our methods of measuring betweenness gives a more accurate picture of the individuals in the network. We think that other than position of individuals in the shortest paths, the time at which the intermediate individuals appear in the shortest path are vital for measuring the significance of intermediate individuals.

There are many possible extensions of our work. Other than betweenness, we can define the degree and closeness centrality measures for dynamic networks. Similarly, network measures like clustering coefficients, cliques, connectivity, and many more can be formally studied in explicitly dynamic networks.

References

[1] J. Aizen, D. Huttenlocher, J. Kleinberg, and A. Novak. Traffic-based feedback on the web. *Proceedings of the National Academy of Sciences*, 101(Suppl.1):5254–5260, 2004.

[2] J.M. Anthonisse. The rush in a graph. *Amsterdam: Mathematische Centrum*, 1971.

[3] 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.

[4] K. Berman. Vulnerability of scheduled networks and a generalization of menger's theorem. *Networks*, 28, 1996.

[5] K. Börner, L. DallAsta, 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*, 2006. in press.

[6] Katy Börner, Jeegar Maru, and Robert Goldstone. The simultaneous evolution of author and paper networks. *PNAS*, 101(Suppl 1):5266–5273, 2004.

[7] U. Brandes. A faster algorithm for betweenness centrality. *Jouranl of Mathematical Sociology*, 25:163–177, 2001.

[8] A. Broido and K. Claffy. Internet topology: connectivity of ip graphs. In *Proceedings of SPIE ITCom*, 2001.

[9] K. Carley. Communicating new ideas: The potential impact of information and telecommunication technology. *Technology in Society*, 18(2):219–230, 1996.

[10] 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.

[11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, editors. *Introduction to Algorithms*. The MIT Press, 2001.

[12] D. P. Croft, R. James, P.O.R. Thomas, C. Hathaway, D. Mawdsley, K.N. Laland, and J. Krause. Social structure and co-operative interactions in a wild population of guppies (*poecilia reticulata*). *Behavioural Ecology and Sociobiology*, In Press.

[13] Zoltán Dezsö and Albert-László Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65(055103(R)), 2002. DOI: 10.1103/PhysRevE.65.055103.

[14] P. Domingos. Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20:80–82, 2005.

[15] P. Domingos and M. Richardson. Mining the network value of customers. In *Seventh International Conference on Knowledge Discovery and Data Mining*, 2001.

[16] S. Eubank, H. Guclu, V.S. Kumar, M.V. 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.

[17] 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*, 2006. Submitted.

[18] 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. Submitted.

[19] L.C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[20] L.C. Freeman. Centrality in social networks: I. conceptual clarification. *Social Networks*, 1:215–239, 1979.

[21] Linton Freeman. Finding social groups: A meta-analysis of the southern women data. In Ronald Breiger, Kathleen Carley, and Philippa Pattison, editors, *Dynamic Social Network Modeling and Analysis*. The National Academies Press, Washington, D.C., 2003.

[22] J. Hopcroft, O. Khan, B. Kulis, and B. Selman. Natural communities in large linked networks. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 541–546, 2003.

[23] M. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B*, 266:859–867, 1999.

[24] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[25] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and inference problems for temporal networks. *J. Comput. Syst. Sci.*, 64(4):820–842, 2002.

[26] M. Kretzschmar and M. Morris. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.*, 133:165–195, 1996.

[27] K. Lewin. *Principles of Topological Psychology*. New York: McGraw Hill, 1936.

[28] Michael 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.

[29] Network Dynamics and Simulation Science Laboratory, 1880 Pratt Dr, Building XV, Blacksburg, VA, 24061. *Synthetic Data Products for Societal Infrastructures and Proto-Populations: Data Set 1.0, NDSSL-TR-06-006*. `http://ndssl.vbi.vt.edu/Publications/ndssl-tr-06-006.pdf`.

[30] M. E. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(016128), 2002. DOI: 10.1103/PhysRevE.66.016128.

[31] 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*, volume 10. Martin-Luther-University Halle-Wittenberg, 2007. In press.

[32] Siva R. Sundaresan, Ilya R. Fischhoff, Jonathan Dushoff, and Daniel I. Rubenstein. Network metrics reveal differences in social organization between two fission-fusion species, Grevy's zebra and onager. *Oecologia*, 2006. doi 10.1007/s00442-006-0553-6.

[33] S. Wasserman and Faust K. *Social Network Analysis.* Cambridge University Press, Cambridge, MA, 1994.

A Proofs

A.1 Proof of Lemma 1

Proof. Let $P = \{e_1, e_2, \ldots, e_n\}$ be a temporal trail in a dynamic network G. Let $s(e_i)$ and $t(e_i)$ denote the two nodes that e_i connects. Since P is a temporal trail, $t(e_i) = s(e_{i+1})$ and $\lambda(e_i) < \lambda(e_{i+1})$ for all $i = 1, 2, \ldots, T-1$. First, we show that every two consecutive edges e_i, e_{i+1} in P have two corresponding edges in a path in the DAG of G. Recall that we create the DAG such that each edge $e \in E$ has two corresponding edges (u_{t-1}, v_t) and (v_{t-1}, u_t) in the DAG, where $s(e) = u, t(e) = v$ and $t = \lambda(e)$. Therefore, for every two consecutive edges e_i, e_{i+1} in P, where $s(e_i) = u, s(e_{i+1}) = v, t(e_{i+1}) = w$ and $\lambda(e_i) = t$, there is a path u_{t-1}, v_t, w_{t+1} in the DAG. By induction, there is a path from $x_{\lambda(e_1)-1}$ to $y_{\lambda(e_n)}$ where $x = s(e_1)$ and $y = t(e_n)$. Since there are edges $(x_{in}, x_{\lambda(e_1)-1})$ and $(y_{\lambda(e_n)}, y_{out})$ in the DAG, there is a path from x_{in} to y_{out}. Obviously, the two edges, $(x_{in}, x_{\lambda(e_1)-1})$ and $(y_{\lambda(e_n)}, y_{out})$, make $n = n' - 2$. □

A.2 Proof of Theorem 2

Proof. We have essentially to prove that Brandes' recurrence [7] works with shortest temporal trails on a multigraph. Let $nst_{st}(u, v)$ denote the number of shortest temporal trails from s to t that contains an edge e where $(s(e), t(e)) = (u, v)$ (note that, since G is a multigraph, there can be more than one edge between u and v, and $nst_{st}(u, v)$ counts all shortest temporal trails that contain one of such edges). We have that

$$\begin{aligned}
B_{D(s\bullet)}(v) &= \sum_{t\in V: t\neq s\neq v} B_{D(st)}(v) = \sum_{t\in V: t\neq s\neq v} \frac{nst_{st}(v)}{nst_{st}} \\
&= \sum_{t\in V: t\neq s\neq v} \sum_{w:(v,w)\in D_s} \frac{nst_{st}(v,w)}{nst_{st}} \\
&= \sum_{w:(v,w)\in D_s} \sum_{t\in V: t\neq s\neq v} \frac{nst_{st}(v,w)}{nst_{st}}
\end{aligned}$$

Let $B_{D(st)}(u,v)$ denote the delay-dependency of s and t on all edges between u and v. That is, $B_{D(st)}(u,v) = nst_{st}(u,v)/nst_{st}$. Therefore, we have that

$$\sum_{w:(v,w)\in D_s} \sum_{t\in V: t\neq s\neq v} \frac{nst_{st}(v,w)}{nst_{st}} = \sum_{w:(v,w)\in D_s} \sum_{t\in V: t\neq s\neq v} B_{D(st)}(v,w).$$

Out of nst_{sw} shortest trails from s to w, there are nst_{sv} of such trails that go to v first and then go to w. Therefore, out of $nst_{st}(w)$ shortest trails from s to t that contain w, there are $\frac{nst_{sv}}{nst_{sw}} \cdot nst_{st}(w)$ of such trails that go to v first and then go to w. Therefore, we have that the delay-dependency $B_{D(st)}(v,w)$ of s and t on v and w is

$$B_{D(st)}(v,w) = \begin{cases} \frac{nst_{sv}}{nst_{sw}} & \text{if } w = t \\ \frac{nst_{sv}}{nst_{sw}} \cdot \frac{nst_{st}(w)}{nst_{st}} = \frac{nst_{sv}}{nst_{sw}} \cdot B_{D(st)}(w) & \text{otherwise.} \end{cases}$$

Substituting into the above summation, we have

$$\begin{aligned}
&\sum_{w:(v,w)\in D_s} \sum_{t\in V: t\neq s\neq v} B_{D(st)}(v,w) \\
&\quad = \sum_{w:(v,w)\in D_s} \left(\frac{nst_{sv}}{nst_{sw}} + \sum_{t\in V\setminus\{w\}: t\neq s\neq v} \frac{nst_{sv}}{nst_{sw}} \cdot B_{D(st)}(w) \right) \\
&\quad = \sum_{w:(v,w)\in D_s} \frac{nst_{sv}}{nst_{sw}} \left(1 + \sum_{t\in V\setminus\{w\}: t\neq s\neq v} B_{D(st)}(w) \right) \\
&\quad = \sum_{w:(v,w)\in D_s} \frac{nst_{sv}}{nst_{sw}} \left(1 + B_{D(s\bullet)}(w) \right)
\end{aligned}$$

$\square$