Chaos, Solitons & Fractals 72 (2015) 35–48

Contents lists available at ScienceDirect

Chaos, Solitons & Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Betweenness in time dependent networks

Ahmad Alsayed [a,b], Desmond J. Higham [a,*,1]

[a] Department of Mathematics and Statistics, University of Strathclyde, UK
[b] Department of Mathematics, Jazan University, Saudi Arabia

ARTICLE INFO

Article history:
Available online 11 February 2015

ABSTRACT

The concept of betweenness has given rise to a very useful class of network centrality measures. Loosely, betweenness quantifies the level of importance of a node in terms of its propensity to act as an intermediary when messages are passed around the network. In this work we generalize a walk-based betweenness measure to the case of time-dependent networks, such as those arising in telecommunications and on-line social media. We also introduce a new kind of betweenness measure, temporal betweenness, which quantifies the importance of a time-point. We illustrate the effectiveness of these new measures on synthetic examples, and also give results on real data sets involving voice call, email and Twitter.

1. Background material

1.1. Betweenness

This work deals with centrality measures for dynamic networks. We begin by summarizing some relevant concepts from the static network setting. Our focus is on the concept of *betweenness*, which arose in the social network analysis literature [5,22] and has become prominent across network science [16].

Loosely, betweenness quantifies the extent to which a node is relied upon when messages are passed around a network. Traditionally, shortest paths between nodes were considered, and the betweenness of node r was found by considering all other distinct nodes, $i \neq j$, and recording the proportion of shortest paths between i and j that involve node r. As pointed out by Freeman et al. [6] and by Newman [17], key messages do not necessarily follow geodesics, and hence there is scope for altering the definition in order to allow for other types of traversal through a network. In [3], a general framework was presented, based on functions of the adjacency matrix, and this is the approach that we follow here. Given an unweighted, directed network with N nodes, we let $A \in \mathbb{R}^{N \times N}$ denote the adjacency matrix, so that $(A)_{ij} = 1$ if there is an edge from i to j and $(A)_{ij} = 0$ otherwise. It then follows that the exponential, $\exp(A)$, and resolvent, $(I - \alpha A)^{-1}$, provide information about the potential for pairwise communication [2]. This can be understood by considering power series expansions of the matrix functions and noting that $(A^k)_{ij}$ counts the number of walks from i to j that involve exactly k edges. In the case of the matrix resolvent, which dates back to the work of Katz [13], the attenuation parameter, α, is chosen in the range $0 < \alpha < 1/\rho(A)$, where $\rho(\cdot)$ denotes the spectral radius.

Communicability betweenness for a general node, r, was then defined in [4] according to

$$C_N \sum \sum_{i \neq j, i \neq r, j \neq r} \frac{\exp(A)_{ij} - \exp(A - E(r))_{ij}}{\exp(A)_{ij}}, \tag{1}$$

where $C_N = \frac{1}{(N-1)^2 - (N-1)}$ is a normalizing factor. Here $E(r)$ has nonzeros only in row and column r, and in this row and column has -1 wherever A has 1; hence $A - E(r)$ is the adjacency matrix when all edges involving the node r are removed. In words, the communicability betweenness for

* Corresponding author.
[1] Supported by a Wolfson/Royal Society Research Merit Award.

http://dx.doi.org/10.1016/j.chaos.2014.12.009
0960-0779/

node r aggregates the relative decrease in exponential communicability over all other pairs of nodes when node r is removed from the network. In a similar manner, replacing the matrix exponential by the matrix resolvent, [3] defined the *resolvent betweenness* for node r as

$$C_N \sum \sum_{i\neq j, i\neq r, j\neq r} \frac{((I-\alpha A)^{-1})_{ij} - ((I-\alpha(A-E(r)))^{-1})_{ij}}{((I-\alpha A)^{-1})_{ij}}. \qquad (2)$$

We assume here that the underlying network is fully connected so that no division-by-zero issues arise in (1) and (2).

1.2. Time dependent networks

Many types of interaction have a well-defined dynamic aspect, giving rise to the study of temporal networks [12]. In this work, motivated by applications in telecommunication and on-line social media, and following the ideas in [7], we consider a fixed set of N nodes and a discrete, finite and ordered set of time points, $t_0 < t_1 < \cdots < t_M$. We then assume that the state of the network is supplied at each time t_k, as represented by an adjacency matrix, $A^{[k]}$. For example, in the Twitter context, $(A^{[k]})_{ij} = 1$ may indicate that account i sent at least one tweet to account j in the time interval $(t_{k-1}, t_k]$.

In [10] the concept of a *dynamic walk* was introduced as a means to extend centrality measures from the static case. In words, a dynamic walk of length w between a pair of nodes is any suitable traversal along w edges that respects the arrow of time – we can remain at a node and wait for an edge to appear, but we cannot go back in time and use an edge that subsequently disappeared. More precisely, a dynamic walk of length w from node i_1 to node i_{w+1} consists of a sequence of edges $i_1 \to i_2, i_2 \to i_3, \ldots, i_w \to i_{w+1}$ and a nondecreasing sequence of times $t_{r_1} \leqslant t_{r_2} \leqslant \cdots \leqslant t_{r_w}$ such that $A^{[r_m]}_{i_m, i_{m+1}} \neq 0$. Just as matrix powers can be used to count walks in the static case, dynamic walks can be counted via matrix products. It was shown in [10] that the $N \times N$ matrix

$$Q := \left(I - \alpha A^{[0]}\right)^{-1} \cdots \left(I - \alpha A^{[M]}\right)^{-1} \qquad (3)$$

is such that $(Q)_{ij}$ gives a weighted count of the number of dynamic walks of length w from node i to node j, where walks of length w are scaled by a factor α^w. This is a direct generalization of the static case described in subSection 1.1, where a single resolvent matrix was used ($M = 0$), and in order to ensure convergence of the underlying power series, we require $\alpha < 1/\max_k \rho(A^{[k]})$. Following [10] we refer to Q in (3) as the *dynamic communicability matrix*. We note that Q takes account of effects that cannot be seen through the individual snapshots, $\{A^{[k]}\}_{k=0}^{M}$, or the aggregate adjacency matrix $\sum_{k=0}^{M} A^{[k]}$. The usefulness of this concept has been illustrated on real data sets in [9,10,15,19], where Katz-style broadcast and receive centralities were computed for time-dependent networks. Similar shortest-path based measures were developed and tested in [18,20,21].

Our aim here is to use dynamic communicability as a means to quantify betweenness.

2. Temporal and nodal betweenness for dynamic networks

We will use Q in (3) as the basis for two types of betweenness measure. First, following directly from (2), we will look at the effect on communicability of removing a node for all time. Letting $E_r^{[k]}$ denote the matrix with nonzeros only in row and column r of $A^{[k]}$, and in this row and column having 1 wherever $A^{[k]}$ has 1, we see that $\bar{A}_r^{[k]} := A^{[k]} - E_r^{[k]}$ is the adjacency matrix at time point k when all edges involving the node r are removed. We then let

$$\bar{Q}_r := \left(I - \alpha \bar{A}_r^{[0]}\right)^{-1} \cdots \left(I - \alpha \bar{A}_r^{[M]}\right)^{-1}. \qquad (4)$$

In this way, $\bar{Q}_r$ has (i,j) element that quantifies the ability of node i to communicate with node j using dynamic walks that do not involve node r.

In this temporal context there is another clear sense in which betweenness can be measured. Rather than focusing on individual nodes, we may consider time points – in order to identify critical stages in the network evolution, we may record how much the dynamic communicability decreases when a time point is removed. We will let $\{\widehat{A}^{[k,q]}\}_{k=0}^{M}$ denote the adjacency matrix sequence with $A^{[q]}$ replaced by 0; that is,

$$\widehat{A}^{[k,q]} = A^{[k]}, \quad \text{for} \quad k \neq q, \quad \text{and} \quad \widehat{A}^{[q,q]} = 0.$$

We then define

$$\widehat{Q}^{[q]} := \left(I - \alpha \widehat{A}^{[0,q]}\right)^{-1} \left(I - \alpha \widehat{A}^{[1,q]}\right)^{-1} \cdots \left(I - \alpha \widehat{A}^{[M,q]}\right)^{-1}. \qquad (5)$$

In practice, since we are only concerned with comparing nodes and comparing time points based on the *relative* change that their removal causes to dynamic communicability, we are free to apply a scaling. Hence, to avoid numerical under or overflow, we will scale by $\|Q\|$, where $\|\cdot\|$ denotes the Euclidean norm. With a slight re-use of notation, we will therefore redefine $Q, \bar{Q}_r$ and $\widehat{Q}^{[q]}$ to denote these scaled versions. Setting $Q^{[-1]} = \bar{Q}_r^{[-1]} = \widehat{Q}^{[-1,q]} = I$, we therefore let, for $k = 0, 1, \ldots, M$,

$$Q^{[k]} = \frac{Q^{[k-1]}\left(I - \alpha A^{[k]}\right)^{-1}}{\|Q^{[k-1]}\left(I - \alpha A^{[k]}\right)^{-1}\|}, \qquad (6)$$

$$\bar{Q}_r^{[k]} = \frac{\bar{Q}_r^{[k-1]}\left(I - \alpha \bar{A}_r^{[k]}\right)^{-1}}{\|Q^{[k-1]}\left(I - \alpha A^{[k]}\right)^{-1}\|}, \qquad (7)$$

$$\widehat{Q}^{[k,q]} = \frac{\widehat{Q}^{[k-1,q]}\left(I - \alpha \widehat{A}^{[k,q]}\right)^{-1}}{\|Q^{[k-1]}\left(I - \alpha A^{[k]}\right)^{-1}\|}. \qquad (8)$$

Following (2), we then define the *nodal betweenness* of node r to be

$$\mathrm{NB}_r := C_N \sum \sum_{i\neq j, i\neq r, j\neq r} \frac{\left(Q^{[M]}\right)_{ij} - \left(\bar{Q}_r^{[M]}\right)_{ij}}{\left(Q^{[M]}\right)_{ij}} \qquad (9)$$

and the *temporal betweenness* of time point q to be