

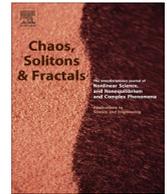


ELSEVIER

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.

© 2015 Elsevier Ltd. All rights reserved.

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

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

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.

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

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

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, $t_0 < t_1 < \dots < t_M$. 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 $i_1 \rightarrow i_2, i_2 \rightarrow i_3, \dots, i_w \rightarrow i_{w+1}$ and a nondecreasing sequence of times $t_{r_1} \leq t_{r_2} \leq \dots \leq t_{r_w}$ such that $A_{i_m, i_{m+1}}^{[r_m]} \neq 0$. 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

$$Q := (I - \alpha A^{[0]})^{-1} \dots (I - \alpha A^{[M]})^{-1} \quad (3)$$

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

$$\bar{Q}_r := (I - \alpha \bar{A}_r^{[0]})^{-1} \dots (I - \alpha \bar{A}_r^{[M]})^{-1}. \quad (4)$$

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 $\{\hat{A}^{[k,q]}\}_{k=0}^M$ denote the adjacency matrix sequence with $A^{[q]}$ replaced by 0; that is,

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

We then define

$$\hat{Q}^{[q]} := (I - \alpha \hat{A}^{[0,q]})^{-1} (I - \alpha \hat{A}^{[1,q]})^{-1} \dots (I - \alpha \hat{A}^{[M,q]})^{-1}. \quad (5)$$

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 $\hat{Q}^{[q]}$ to denote these scaled versions. Setting $Q^{[-1]} = \bar{Q}_r^{[-1]} = \hat{Q}^{[-1,q]} = I$, we therefore let, for $k = 0, 1, \dots, M$,

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

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

$$\hat{Q}^{[k,q]} = \frac{\hat{Q}^{[k-1,q]} (I - \alpha \hat{A}^{[k,q]})^{-1}}{\|\hat{Q}^{[k-1,q]} (I - \alpha \hat{A}^{[k,q]})^{-1}\|}. \quad (8)$$

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

$$NB_r := C_N \sum_{i \neq j, i \neq r, j \neq r} \frac{(Q^{[M]})_{ij} - (\bar{Q}_r^{[M]})_{ij}}{(Q^{[M]})_{ij}} \quad (9)$$

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

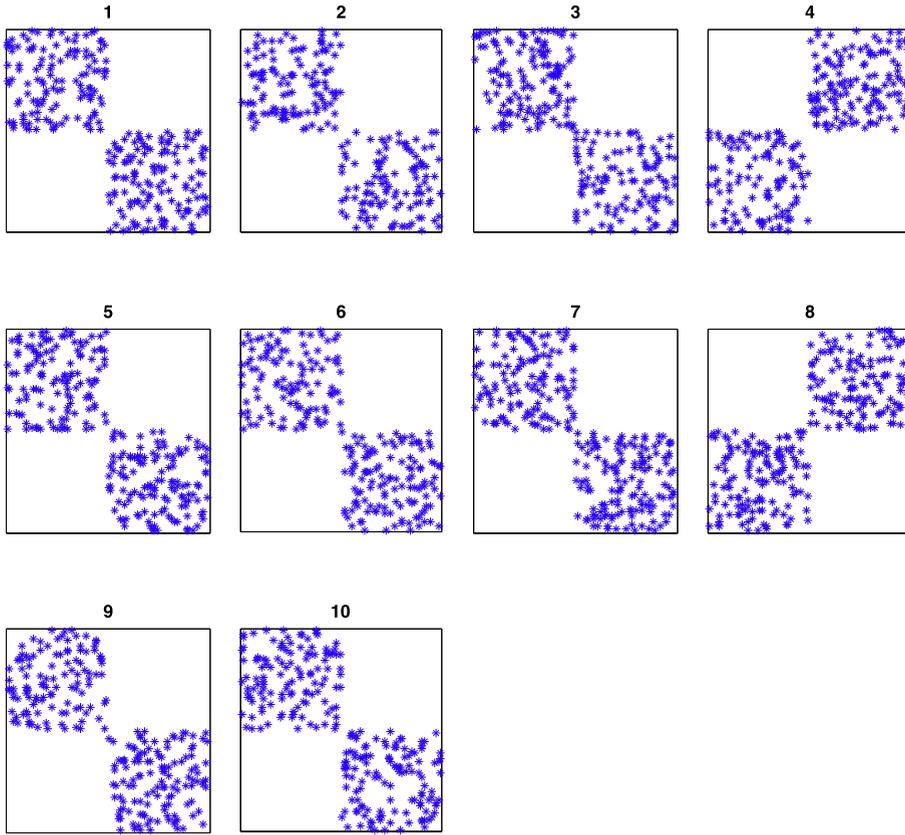


Fig. 1. A sample of the dynamic network process used to illustrate temporal betweenness. Here we have 10 days. The pictures show a typical non-zero pattern in the unsymmetric adjacency matrices. Nodes are ordered so that group A appears before group B. The undirected block diagonal structure on days 1, 2, 3, 5, 6, 7, 9, 10 arises because only intra-group links are generated. The off-diagonal block diagonal structure on days 4 and 8 arises because only inter-group links are generated.

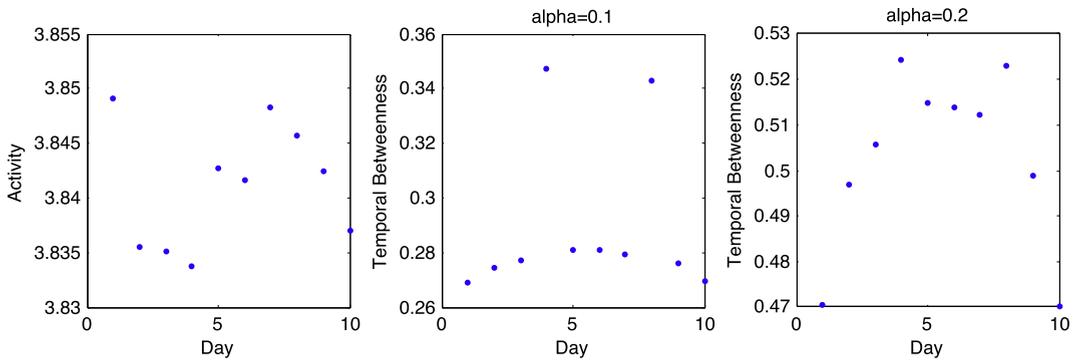


Fig. 2. Results for the network sequence illustrated in Fig. 1, averaged over 1000 runs. Left: total activity at each day. Middle: temporal betweenness for each day with $\alpha = 0.1$. Right: temporal betweenness for each day with $\alpha = 0.2$.

$$TB^{[M,q]} := C_N \sum_{i \neq j} \frac{(Q^{[M]})_{ij} - (\widehat{Q}^{[M,q]})_{ij}}{(Q^{[M]})_{ij}}. \quad (10)$$

We note that in (9) and (10), a zero appearing in the denominator is always accompanied by a zero in the numerator. Here, and throughout this work, we use the convention that $0/0 = 0$.

To get some insight into these new definitions, we will analyze their behavior in the limit $\alpha \rightarrow 0^+$. We begin with a lemma concerning a product of matrix resolvents.

Lemma 2.1. Suppose we have a finite sequence of square matrices $A^{[0]}, A^{[1]}, \dots, A^{[M]}$, each in $\mathbb{R}^{N \times N}$. Let

$$R^{[k]} = (I - \alpha A^{[0]})^{-1} (I - \alpha A^{[1]})^{-1} \dots (I - \alpha A^{[k]})^{-1}.$$

As $\alpha \rightarrow 0^+$, we have

$$R^{[k]} = I + \alpha \sum_{p=0}^k A^{[p]} + \alpha^2 \sum_{p=0}^k \sum_{h=p}^k A^{[p]} A^{[h]} + O(\alpha^3).$$

Proof. The result follows straightforwardly by induction. At the initial time point, as $\alpha \rightarrow 0^+$, we have

$$(I - \alpha A^{[0]})^{-1} = I + \alpha A^{[0]} + \alpha^2 (A^{[0]})^2 + O(\alpha^3),$$

giving the result for $k = 0$. Now, suppose the result is true for the number of factors equal to $0, 1, 2, \dots, k - 1$. Then

$$\begin{aligned} R^{[k]} &= R^{[k-1]} (I - \alpha A^{[k]})^{-1} = \left(I + \alpha \sum_{p=0}^{k-1} A^{[p]} + \alpha^2 \sum_{p=0}^{k-1} \sum_{h=p}^{k-1} A^{[p]} A^{[h]} \right. \\ &\quad \left. + O(\alpha^3) \right) \left(I + \alpha A^{[k]} + \alpha^2 (A^{[k]})^2 + O(\alpha^3) \right) = I + \alpha \sum_{p=0}^k A^{[p]} \\ &\quad + \alpha^2 \sum_{p=0}^k \sum_{h=p}^k A^{[p]} A^{[h]} + O(\alpha^3), \end{aligned}$$

as required.

We now characterize the two new betweenness measures in this limit. \square

Theorem 2.1. Assume that $\sum_{p=0}^M (A^{[p]})_{ij} \neq 0$, for all $i \neq j$, so that every edge appears at least once. Then, as $\alpha \rightarrow 0^+$,

$$\lim_{\alpha \rightarrow 0^+} \frac{NB_r}{\alpha} = C_N \sum_{i \neq j, i \neq r, j \neq r} \frac{\sum_{p=0}^M \sum_{h=p}^M (A^{[p]} A^{[h]} - \bar{A}_r^{[p]} \bar{A}_r^{[h]})_{ij}}{\sum_{p=0}^M A_{ij}^{[p]}}, \tag{11}$$

and

$$\lim_{\alpha \rightarrow 0^+} TB^{[M,q]} = C_N \sum_{i \neq j} \frac{A_{ij}^{[q]}}{\sum_{p=0}^M A_{ij}^{[p]}}. \tag{12}$$

Proof. Noting that, for $i \neq r$ and $j \neq r$,

$$\sum_{p=0}^M (A^{[p]})_{ij} = \sum_{p=0}^M (\bar{A}_r^{[p]})_{ij},$$

we see from Lemma 2.1 that

$$\sum_{i \neq j, i \neq r, j \neq r} \frac{Q_{ij}^{[M]} - (\bar{Q}_r^{[M]})_{ij}}{Q_{ij}^{[M]}} = \frac{\left(\alpha \sum_{p=0}^M \sum_{h=p}^M (A^{[p]} A^{[h]} - \bar{A}_r^{[p]} \bar{A}_r^{[h]}) + O(\alpha^2) \right)_{ij}}{\left(\sum_{p=0}^M A^{[p]} + O(\alpha) \right)_{ij}}.$$

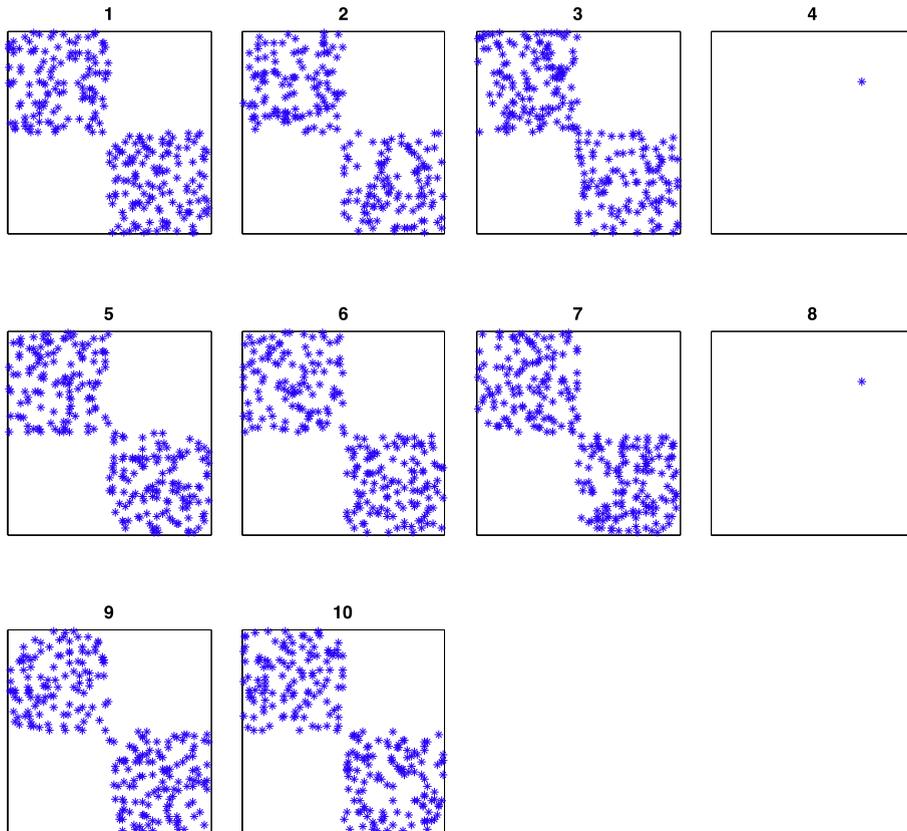


Fig. 3. A sample of the dynamic network process used to illustrate nodal betweenness. We have 10 days. The pictures show the non-zero pattern each day in the unsymmetric adjacency matrices. Nodes are ordered so that group A appears before group B. The undirected block diagonal structure on days 1, 2, 3, 5, 6, 7, 9, 10 arises because only intra-group links are generated. A single, directed edge is inserted deterministically from node 25 to node 75, on days 4 and 8. Hence, without node 25 or 75, the two groups would be disconnected, and far fewer dynamic walks would exist.

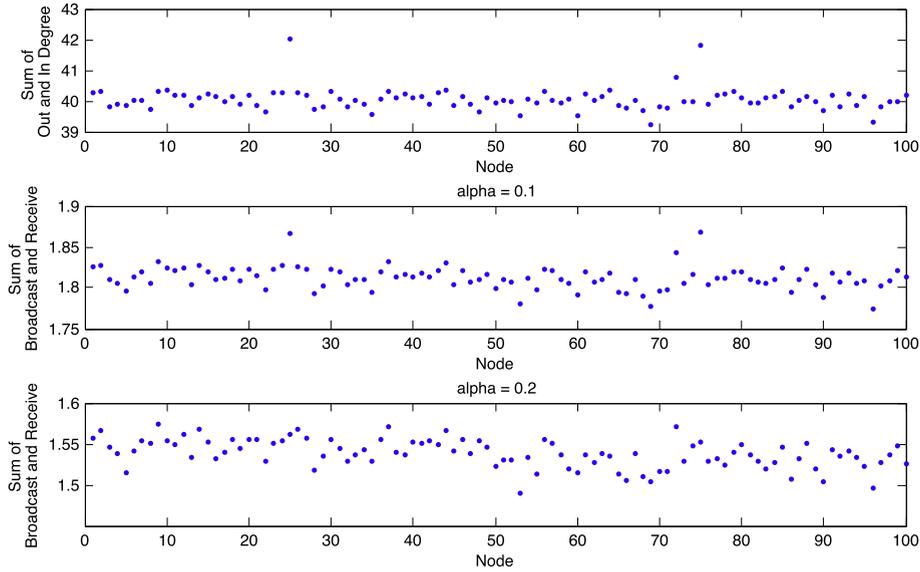


Fig. 4. Results for the network sequence illustrated in Fig. 3, averaged over 1000 runs. Upper: sum of aggregate out and in degree for each node. Middle: sum of final-time broadcast and receive centrality for each node when $\alpha = 0.1$. Lower: sum of final-time broadcast and receive centrality for each node when $\alpha = 0.2$.

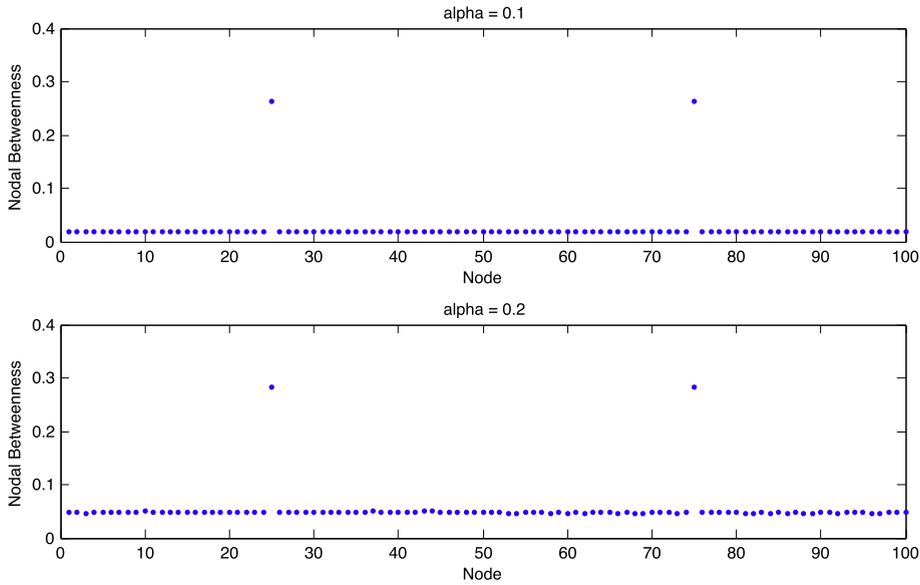


Fig. 5. Nodal betweenness results for the network sequence illustrated in Fig. 3, averaged over 1000 runs. Upper picture: nodal betweenness for each node when $\alpha = 0.1$. Lower picture: nodal betweenness for each node when $\alpha = 0.2$.

The result (11) then follows.

Similarly,

$$\sum_{p=0}^M (A^{[p]})_{ij} - (\widehat{A}^{[p,q]})_{ij} = (A^{[q]})_{ij},$$

and Lemma 2.1 gives

$$\sum_{i \neq j} \frac{Q_{ij}^{[M]} - \widehat{Q}_{ij}^{[M,q]}}{Q_{ij}^{[M]}} = \sum_{i \neq j} \frac{(A^{[q]} + O(\alpha))_{ij}}{(\sum_{p=0}^M A^{[p]} + O(\alpha))_{ij}},$$

leading to (12). \square

The results in Theorem 2.1 have a natural interpretation. The $\alpha \rightarrow 0^+$ limit focusses on short walks, and in (11) the leading term in an expansion of nodal betweenness is seen to aggregate the ratios over all pairs of nodes i, j of (a) the number of dynamic walks of length two from i to j involving node r and (b) the total number of dynamic walks from i to j of length two. In this limit we also see from (12) that the temporal version aggregates the ratio of (a) walks of length one from i to j at time q and (b) the total number of walks of length one from i to j . We emphasize that Theorem 2.1 is to be regarded as a

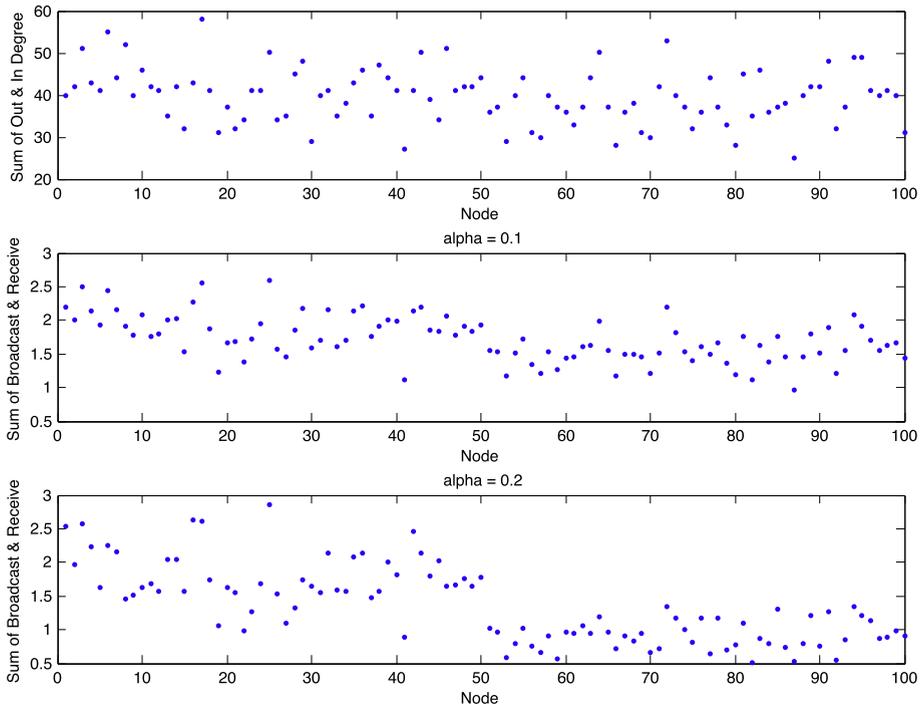


Fig. 6. Results for a single instance of the network sequence illustrated in Fig. 3. Upper: sum of aggregate out and in degree for each node. Middle: sum of final-time broadcast and receive centrality for each node when $\alpha = 0.1$. Lower: sum of final-time broadcast and receive centrality for each node when $\alpha = 0.2$.

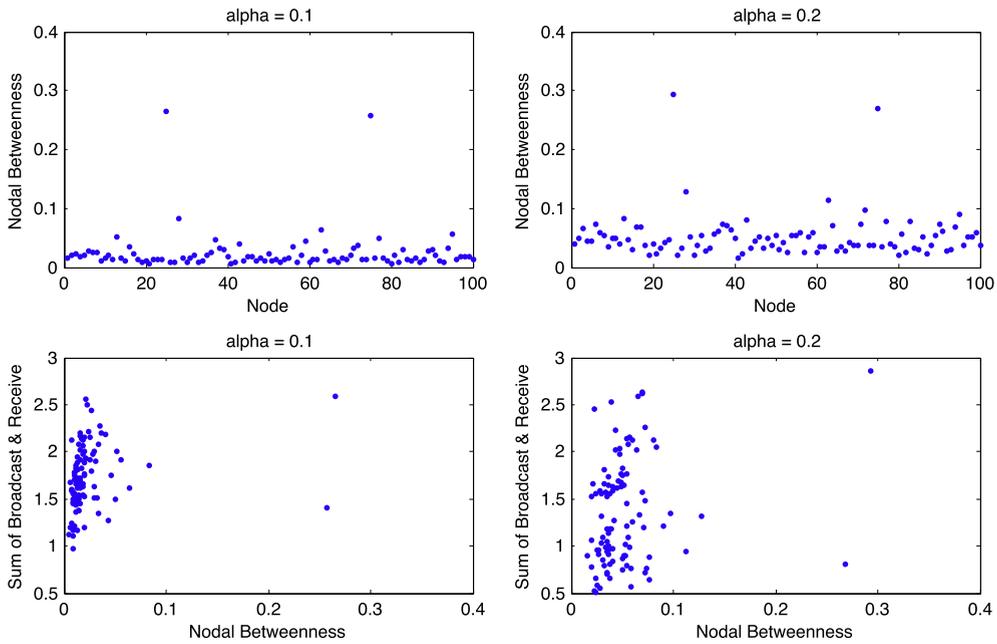


Fig. 7. Results for a single instance of the network sequence illustrated in Fig. 3. Upper left: nodal betweenness with $\alpha = 0.1$. Upper right: nodal betweenness with $\alpha = 0.2$. Lower left: sum of broadcast and receive plotted against nodal betweenness for $\alpha = 0.1$. Lower right: sum of broadcast and receive plotted against nodal betweenness for $\alpha = 0.2$.

consistency check. In practice the use of a nonzero α value, which thereby allows longer dynamic walks to be incorporated, is a key strength of the new measures.

3. Synthetic tests

In this section we give results on two computational experiments with specially constructed dynamic networks. The purpose of these tests is to confirm that the new nodal and temporal betweenness measures are able to extract the relevant features, when they are known to be present in the data.

In the first test, we focus on temporal betweenness. To begin, we consider 100 nodes, split into two groups, A and B, each of size 50. To be concrete, we regard the time points as successive days. On the first day, we allow nodes within group A to have directed edges. Each such edge is chosen independently with probability 0.05. Similarly, directed edges of the same type are generated within group B. Hence the overall network consists of two disconnected, directed Erdos–Renyi style networks. The upper left picture in Fig. 1 illustrates this idea. This construction procedure, using independent samples, is applied at days 1, 2, 3, 5, 6, 7, 9, 10. At days 4 and 8, we reverse the process: there are no links within group A, nor within group B, but directed links are inserted between nodes in A and nodes in B with independent probability 0.05. In this way, on each of the ten days nodes have the same in and out degree distribution, but days 4 and 8 play special roles in allowing information to pass between the two groups.

Within this framework, we show results averaged over 1000 independent runs. We found that $1/\max(\rho(A^{[k]})) \approx 0.26$, and hence we test with $\alpha = 0.1$ and $\alpha = 0.2$. In the

left hand picture of Fig. 2, we show the overall activity on each day, which we define as $\|A^{[k]}\|$. Note that the vertical scale is very restricted, and the variation across time is simply a reflection of the finite number of network samples. In particular, days 4 and 8 are not singled out by this measure. The middle and right hand pictures in Fig. 2 illustrate the temporal betweenness at each day, for $\alpha = 0.1$ and $\alpha = 0.2$, respectively. We see that in both cases the measure is able to highlight the key roles of days 4 and 8 – these are the instances where nodes have an opportunity to build new routes into the opposite groups. Day 4 is rated slightly higher than day 8. This can be explained by the fact that the earlier “bridging” edges have greater potential to participate in time-dependent walks that make use of subsequent edges. We also note from Theorem 2.1 that in the limit $\alpha \rightarrow 0^+$ the temporal betweenness measure would not be able to highlight days 4 and 8, since the number of edges, on average, is equal each day. It is by choosing $\alpha > 0$, and thereby allowing walks of length greater than one to have some influence, that we reach the intuitively reasonable conclusion that days 4 and 8 are special.

The second test illustrates nodal betweenness. As in the first test, we have 100 nodes split into two equal sized groups, and 10 days. On days 1, 2, 3, 5, 6, 7, 9, 10, we use the same construction as above: directed links are inserted within group A and within group B with independent probability 0.05. On days 4 and 8 there is a single directed edge from node 25 to node 75. Hence, nodes 25 and 75 have very similar aggregate degrees to the other nodes, but they have a unique ability to send and receive, respectively, messages from group A to group B. Removing either node would completely disconnect the two groups, in terms of dynamic walks. Fig. 3 shows a sequence of adjacency matrices for this test. As above, we have

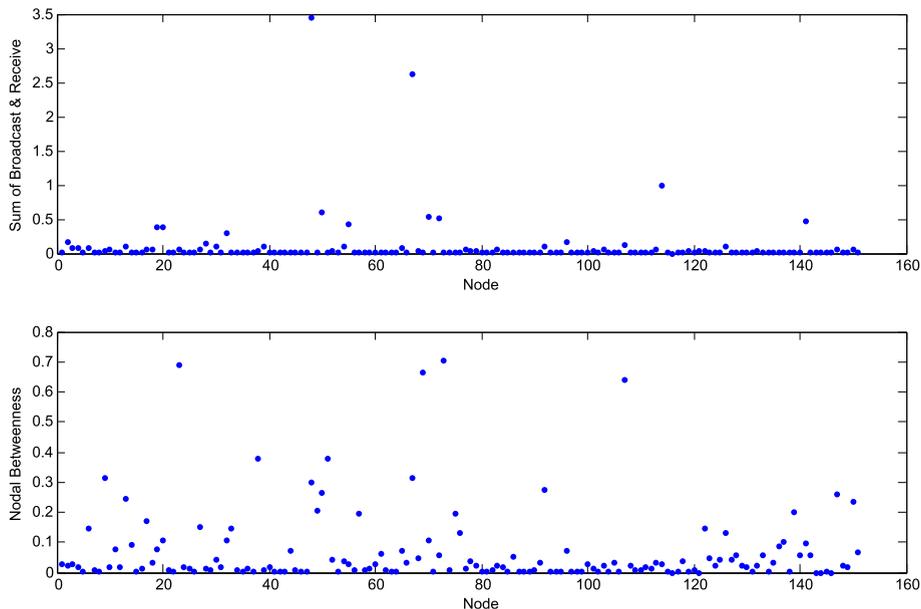


Fig. 8. Nodal betweenness results for Enron data set, consisting of a time dependent, directed set of email interactions. Upper: sum of final-time broadcast and receive centrality for each node. Lower: nodal betweenness for each node.

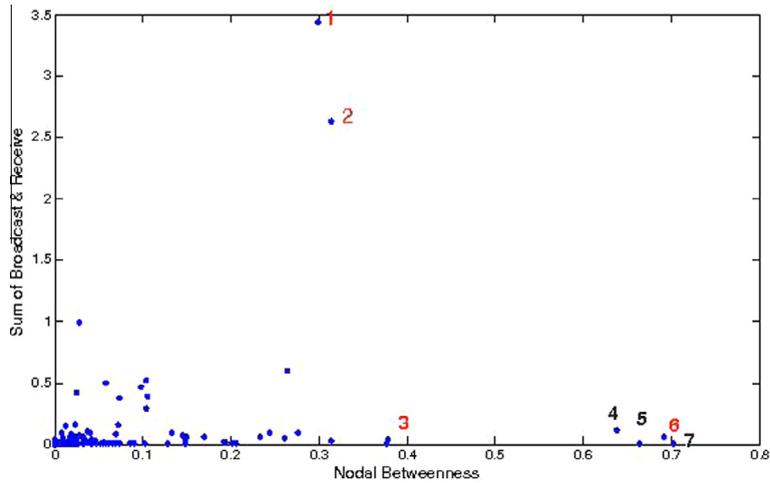


Fig. 9. Further nodal betweenness results for Enron data set: scatter plot of nodal betweenness versus sum of broadcast and receive centrality. The seven nodes with highest nodal betweenness are labelled, and the roles of the corresponding employees are discussed in the text.

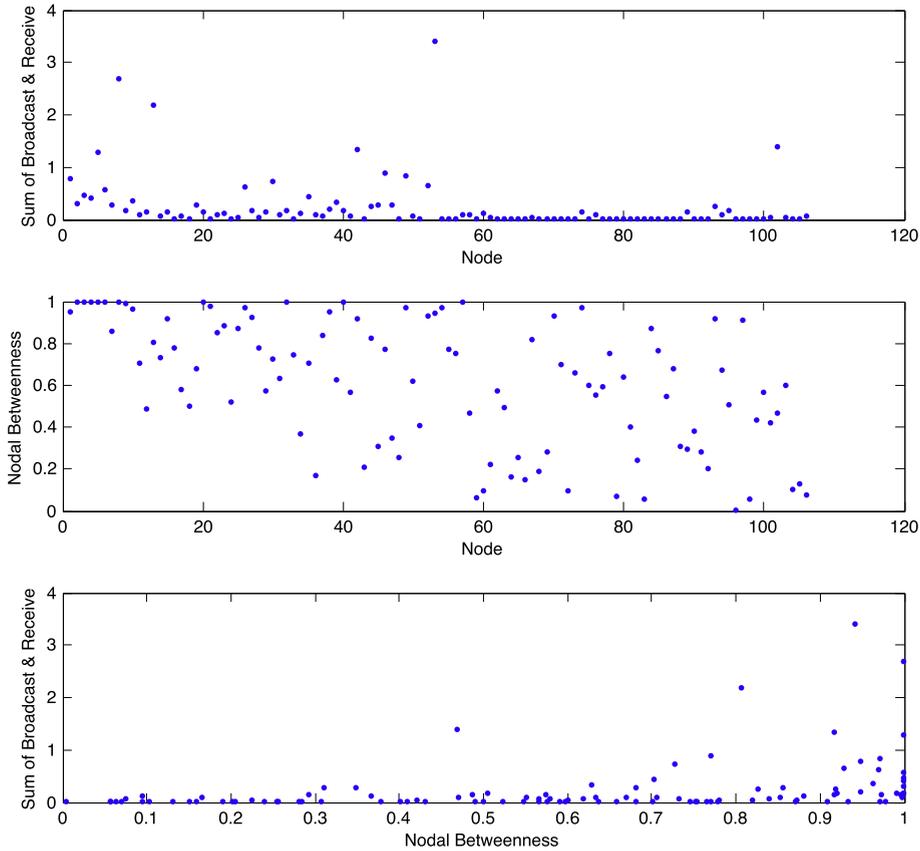


Fig. 10. As for Figs. 8 and 9, using the MIT data set, which represents undirected voice call interactions.

$1 / \max(\rho(A^{[k]})) \approx 0.26$, and show results for $\alpha = 0.1$ and $\alpha = 0.2$, averaged over 1000 independent runs.

The upper picture in Fig. 4 shows, for each node, the time-aggregate of the sum of out and in degree; that is, the total number of links entering or leaving each node. Note that the vertical axis has a very small range, and

the two special nodes have values that are roughly two more than the rest (since they are given an extra link on days 4 and 8 and otherwise have the same degree distribution as the remainder). For $\alpha = 0.1$, the middle picture shows the sum of final-time broadcast and receive centrality, which quantifies how effectively a node can

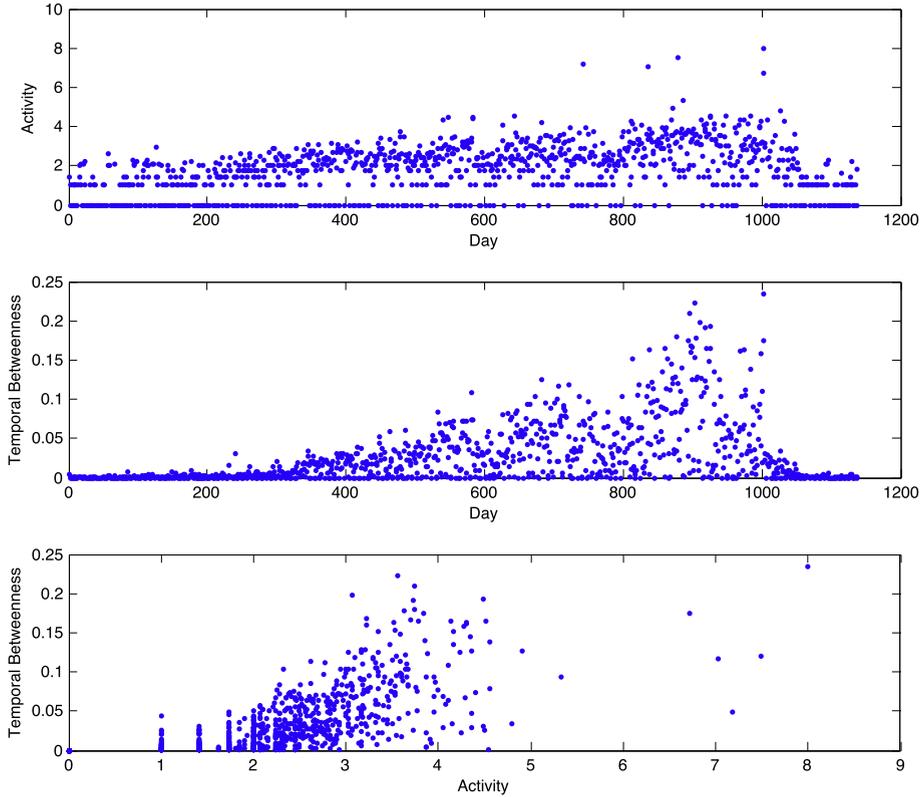


Fig. 11. Temporal betweenness results for Enron data. Upper: total activity each day. Middle: temporal betweenness each day. Lower: total activity versus temporal betweenness.

broadcast and receive dynamic messages; for node n this is given by

$$\sum_{k=1}^N Q_{nk} + \sum_{k=1}^N Q_{kn}.$$

The lower picture repeats this information for $\alpha = 0.2$. We see that for $\alpha = 0.1$, nodes 25 and 75 have a slightly higher ranking than the remainder, but this effect is much less pronounced with $\alpha = 0.2$. Intuitively, the smaller α value emphasizes short walks, notably walks of length one, so the marginally higher degree of nodes 25 and 75 has an effect.

In Fig. 5, we show the nodal betweenness for $\alpha = 0.1$ (upper) and $\alpha = 0.2$ (lower). We see that in both cases the special nature of nodes 25 and 75 is highlighted.

Fig. 6 repeats the experiment in Fig. 4 for the case of a single sample from the network sequence. In this case, the natural variation in aggregate degree across the nodes is sufficient to hide the special nature of nodes 25 and 75 – they have no extra ability to act as broadcasters or receivers. The upper pictures in Fig. 7, however, show that the nodal betweenness measure, with $\alpha = 0.1$ (left) or $\alpha = 0.2$ (right), is able to identify nodes 25 and 75 – there is a large reduction in dynamic communicability between general nodes when these two are removed. The extra information contained in nodal betweenness centrality is further confirmed in the lower pictures of Fig. 7, where we plot

the sum of broadcast and receive against nodal betweenness.

4. Tests on externally-supplied data

In this section we apply the new measures to real networks and look at their correlation with other types of centrality.

4.1. Email and voice call data

The “Enron” data set is based on company emails between 151 Enron employees over a period of 1138 days [14]. It has been widely used as a source of dynamic interactions [8,10,21]. In this case we have a directed edge $A_{ij}^{[k]} = 1$ if person i sent at least one email to person j on day k . The “MIT” data set concerns daily mobile telephone interaction between 106 people over 365 days [1]. Here, we have $A_{ij}^{[k]} = A_{ji}^{[k]} = 1$ if i and j interacted by phone during day k .

The upper picture in Fig. 8 shows the sum of final-time broadcast and receive centrality for each node in the Enron data set. We used $\alpha = 0.1$, which is below the upper limit of $1/\max(\rho(A^{[k]})) = 0.23$. The lower picture shows the corresponding nodal betweenness centrality. In Fig. 9, we scatter plot the sum of broadcast and receive against the

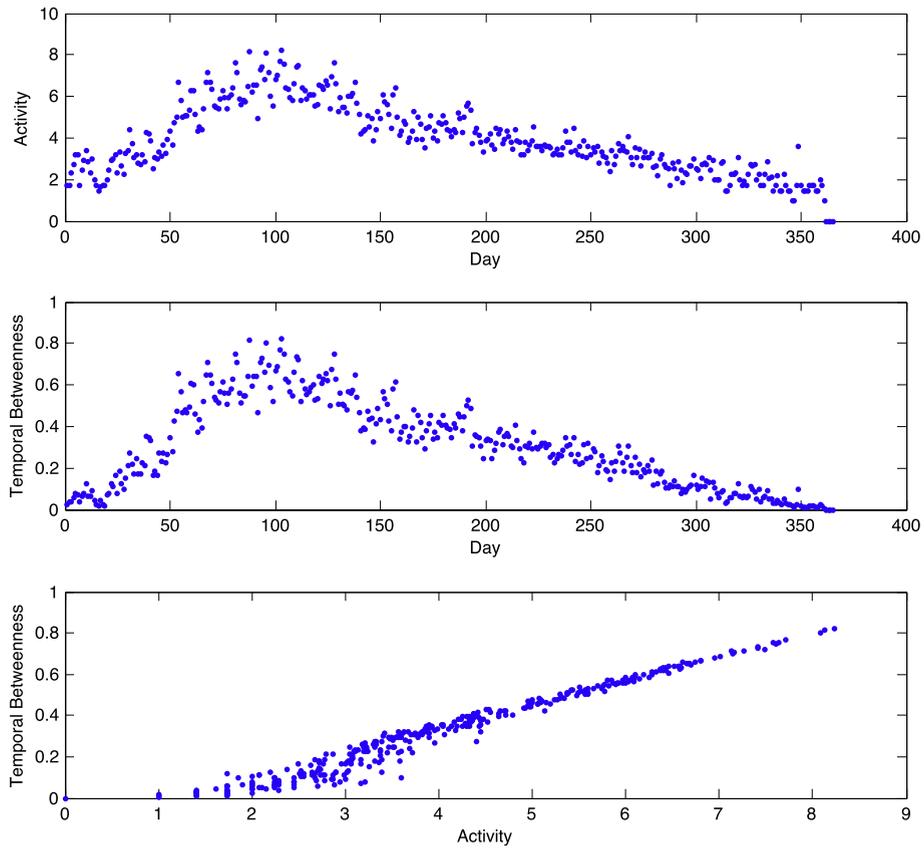


Fig. 12. Temporal betweenness results, as for Fig. 11, using the MIT Data.

Table 1

Centralities on the Twitter data. Kendall tau correlation in upper triangle. Spearman rho correlation in lower triangle.

	Out degree	Dyn. receive	Dyn. broadcast	Nodal betweenness
Out degree		0.63	0.42	0.61
Dyn. receive	0.49		0.66	0.71
Dyn. broadcast	0.75	0.80		0.63
Nodal bet.	0.48	0.75	0.83	

Table 2

Overlaps from the Twitter data. Overlap between top 10 in upper triangle. Overlap between top 20 in lower triangle.

	Out degree	Dyn. receive	Dyn. broadcast	Nodal betweenness
Out degree		3	6	5
Dyn. receive	5		4	3
Dyn. broadcast	7	16		7
Nodal bet.	6	17	17	

Table 3

Overlap amongst top ten for each of the four centrality measures against the average over five experts.

	Out degree	Dyn. receive	Dyn. broadcast	Nodal betweenness
Overlap	4	2	3	4

Table 4

Account IDs in rank order from 1 to 10. Column 1: average over five experts. Column 2: out degree. Column 3: dynamic broadcast. Column 4: nodal betweenness.

Average expert	Out degree	Dynamic broadcast	Nodal betweenness
397	74	74	398
362	34	398	397
398	362	362	74
341	370	34	345
289	358	358	373
345	71	302	375
462	345	397	362
212	398	352	380
71	352	373	385
18	484	380	358

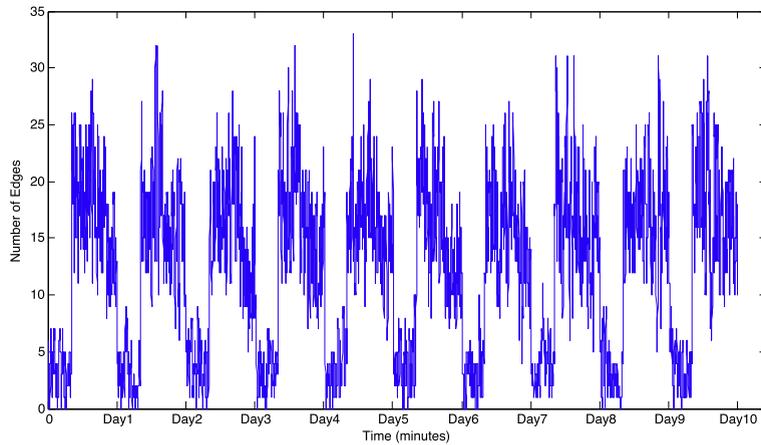


Fig. 13. IEEE data, for each one minute interval, we show the number of (undirected) edges as a function of time.

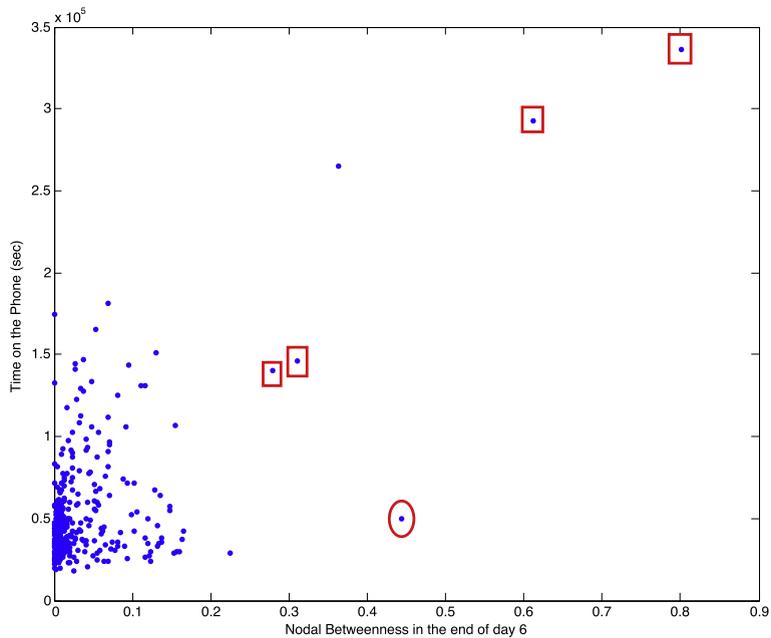


Fig. 14. IEEE data. Based on days 1 to 6: for each node we show nodal betweenness on horizontal axis against time on the phone in seconds on vertical axis. The call leader, ID 200, is marked with a red circle and the other important nodes are marked with red rectangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nodal betweenness. The picture makes it clear that centrality in terms of nodal betweenness is a different attribute to centrality in terms of dynamic communicability. The Spearman rho correlation between the two centrality vectors is 0.23 and the Kendall tau index is 0.45. Fig. 9 also contains labels for seven nodes that are highlighted by the analysis. Nodes labelled 1 and 2 have the highest sum of final-time broadcast and receive centrality and also rank 9th and 7th, respectively, in terms of nodal betweenness. These nodes correspond to an executive and the vice president, who therefore hold influential positions in the company. The nodes labelled 3, 4, 5, 6 and 7 have highest nodal betweenness but are not significant in terms of broadcast and receive centrality. Among this set of five, nodes 3 and 6 are known to correspond to a vice president and a director, and hence nodal betweenness is seen to extract important information.

Fig. 10 repeats the tests in Figs. 8 and 9 for the MIT data set. Here, $1/\max(\rho(A^{(k)})) = 0.12$, and we used $\alpha = 0.1$. In this case we have similar correlation coefficients to those from Fig. 9; a Spearman rho correlation of 0.34 and a Kendall tau index of 0.38.

In Fig. 11, the upper picture shows the total Enron activity for each day. In this case there are many days where $A^{(k)} = 0$. The middle picture shows the temporal betweenness for each day. The lower picture compares these two measures with a scatter plot, and we see that a day with large temporal betweenness is typically not a day with high activity. Fig. 12 shows the same information for the MIT data. In this case we see a strong level of consistency between overall activity and temporal betweenness. One possible explanation for this contrast between the Enron email and MIT phone results is that telephone conversations are more personal and consistent – interactions take place within well-established groups, so removing a day

from the network does not dramatically affect the overall dynamic communicability in the sense of removing possible bridge nodes that offer short-cuts between communities – instead, the effect of the removal is broadly consistent with the number of edges removed.

4.2. Twitter data

Next, we apply nodal betweenness to a Twitter data set that was used by Laflin et al. [15]. This data was collected by the authors of [15] in order to mimic a typical challenge faced by a digital advertising/marketing agency – a consumer-facing retail client wishes to know which Twitter accounts have the most influence in a given area. These accounts are then good targets for relationship-building and information seeding. The Twitter data has 20 time windows and 590 active nodes with nonzero out degree.

First, we compared nodal betweenness measures with out degree, dynamic receive and dynamic broadcast measures. We used $\alpha = 0.9/\max_k \rho(A^{(k)})$ in the resolvent-based measures. Table 1 shows the Kendall tau and Spearman rho correlation coefficients between each pair of centralities, and, similarly, Table 2 shows the top ten and twenty overlap. We see that the four measures are quite similar in these respects.

This Twitter data set is unusual in the sense that an external expert-driven ranking of the nodes is available. In [15] five social media experts were asked to examine the Twitter data (including the content of the tweets), and to identify and rank the most important nodes. After converting the five expert views into a single list, in Table 3 we show how many Twitter accounts in the top ten lists for out degree, dynamic receive, dynamic broadcast and nodal betweenness overlap with the entries in this overall expert top ten. We see that nodal betweenness performs better in

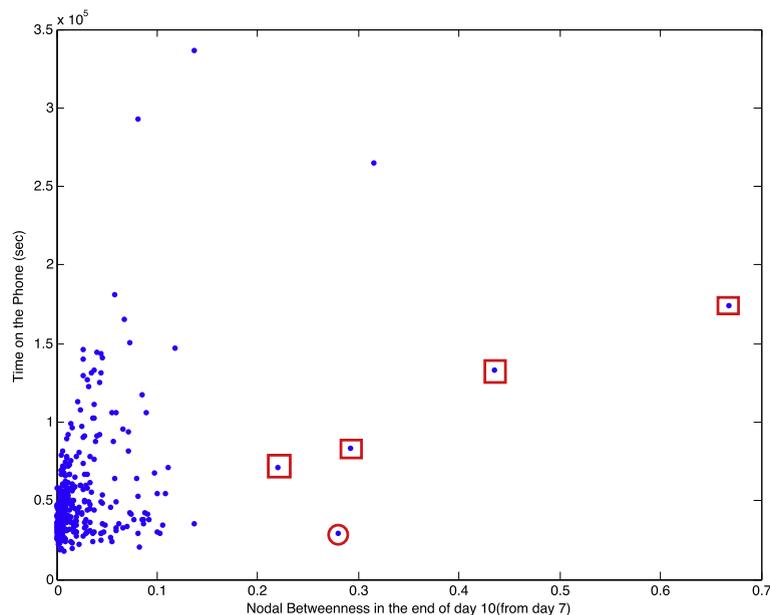


Fig. 15. As for Fig. 14, from day 7 to the end of day 10. Here the markings relate to the new IDs.

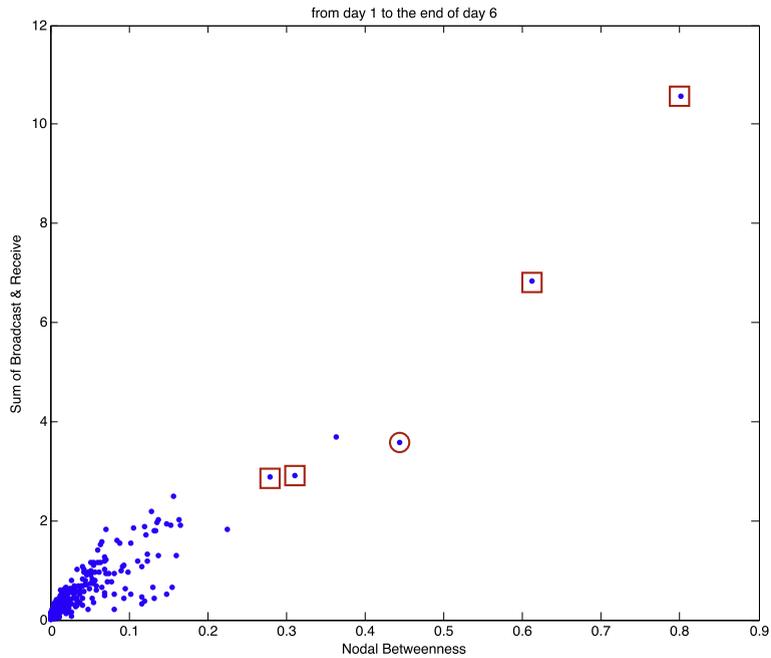


Fig. 16. IEEE data. Horizontal axis is nodal betweenness, as shown in Fig. 14. Vertical axis is sum of broadcast and receive. Symbols are as in Fig. 14.

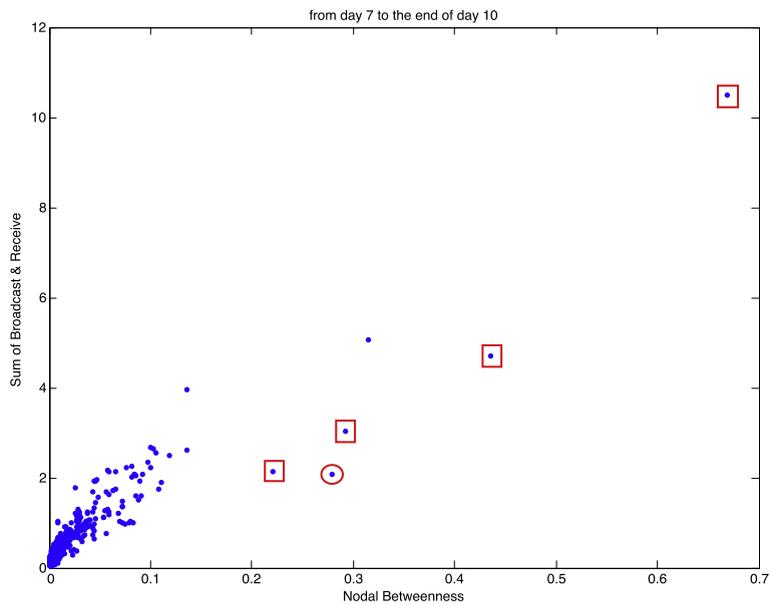


Fig. 17. IEEE data. As for Fig. 15, from day 7 to the end of day 10. Symbols relate to the new IDs, as in Fig. 15.

this regard than the two other dynamic centrality-based measures. For further information, the anonymized accounts IDs for these top ten lists are recorded in Table 4. We see that there is quite a difference between the out degree and nodal betweenness lists, with nodal betweenness giving 1st and 2nd place to two nodes, 398 and 397, that the experts ranked among their top three.

In summary, we conclude that the new nodal betweenness measure offers novel and relevant information in this setting.

4.3. IEEE VAST 2008 challenge

In this section we test nodal betweenness on a time-stamped voice call interaction data set that was supplied for test purposes as part of the IEEE VAST 2008 Challenge [11]. This data has also been studied in [9]. Here, we have a set of 9,834 interactions between 400 mobile phones over a ten day period in June 2006. For each call we have IDs for the send and receive nodes, a start time and the duration in seconds. The competition designers suggested

that the node with ID of 200 is the leader in an important community.

To see the general behavior during the ten day period, Fig. 13 shows the number of interactions as a function of time. The activity follows a natural pattern; users are least active at night and most active in the middle of the day.

Based on the data analyzes submitted by challenge teams, it is likely that between days 1 to 6, inclusive, node ID 200 does indeed control an influential subnetwork involving nodes with IDs 1, 2, 3, and 5. However, most competition entries argued that from day 7 to day 10, the important nodes change their device IDs; 200 changed to 300, and the others changed to 306, 309, 360 and 397. Hence, these nodes from the key subgroup over this period.

In our tests, we work in units of days, so $(A^{[k]})_{ij} = (A^{[k]})_{ji} = 1$ if nodes i and j conversed during day k . This gives 10 time points, and we take $\alpha = 0.9/\max_k \rho(A^{[k]}) = 0.13$. Fig. 14 scatter plots over the nodes IDs (i) the aggregate time on the phone in seconds against (ii) nodal betweenness; in both cases from day 1 to the end of day 6. In this plot, the call leader, ID 200, is marked with a red circle and the other important nodes known to be under the leader's control, IDs 1, 2, 3 and 5, are marked with red rectangles. We see in this figure that the important nodes rank very highly with respect to nodal betweenness, forming 5 of the top 6. By contrast, only two of these IDs have very high aggregate phone time, and, in particular, the leader has a modest level of activity, making this measure a poor predictor of importance.

Fig. 15 repeats the test in Fig. 14 over the remaining time period: from day 7 to the end of day 10. Here, the new IDs of the key group are marked in the plot. As in the previous graph, final-time nodal betweenness is seen to be much more useful than aggregate activity in terms of identifying these important nodes. In particular, the new call leader, ID 300, marked with a red circle, has very low overall activity on the phone but ranks 5th out of 400 for nodal betweenness. The other new important IDs have 5 out of the top 6 rankings for nodal betweenness.

Next, we compare nodal betweenness results with broadcast plus receive centrality. Fig. 16 shows the sum of broadcast and receive centrality against nodal betweenness for each node, using the time period from day 1 to the end of day 6. We have a high correlation coefficient of 0.94 between the two measures and also a high Kendall tau of 0.73. Hence, both measures are able to identify the important nodes.

Fig. 17 repeats the experiment in Fig. 16, using the data running from day 7 to the end of day 10. We see that in this case the nodal betweenness measure outperforms the sum of broadcast and receive centrality. In particular, the new call leader, ID 300, marked with a red circle, has a very modest broadcast and receive centrality, but ranks 5th out of 400 for nodal betweenness.

5. Conclusions

Our aims here were to develop and test concepts and algorithms concerning betweenness centrality for dis-

crete-time dynamic networks. Our approach was based on a Katz-style walk-counting approach that has a sound combinatorial basis. Further, since the resulting algorithms require linear systems to be solved that have the same sparsity structure as the underlying networks, it is feasible to compute these new measures on large data sets. A key novelty of our work was the introduction of a temporal betweenness measure that quantifies the importance of each time-point in terms of the global message-passing capability of the dynamic network. Our computational tests confirmed that the new dynamic centrality measures can reveal important insights.

References

- [1] Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. *Proc Nat Acad Sci* 2009;106:15274–8.
- [2] Estrada E. *The structure of complex networks*. Oxford University Press; 2011.
- [3] Estrada E, Higham DJ. Network properties revealed through matrix functions. *SIAM Rev* 2010;52: 696–671.
- [4] Estrada E, Higham DJ, Hatano N. Communicability betweenness in complex networks. *Physica A* 2009;388:764–74.
- [5] Freeman LC. A set of measures of centrality based on betweenness. *Sociometry* 1977;40:35–41.
- [6] Freeman LC, Borgatti SP, White DR. Centrality in valued graphs: a measure of betweenness based on network flow. *Soc Networks* 1991;13:141–54.
- [7] Grindrod P, Higham DJ. Evolving graphs: dynamical models, inverse problems and propagation. *Proc R Soc, Ser A* 2010;466:753–70.
- [8] Grindrod P, Higham DJ. A matrix iteration for dynamic network summaries. *SIAM Rev* 2013;55:118–28.
- [9] Grindrod P, Higham DJ. A dynamical systems view of network centrality. *Proc R Soc, Ser A* 2014;470:20130835.
- [10] Grindrod P, Parsons MC, Higham DJ, Estrada E. Communicability across evolving networks. *Phys Rev E* 2011;83:046120.
- [11] Grinstein GG, Plaisant C, Laskowski SJ, O'Connell T, Scholtz J, Whiting MA. Vast 2008 challenge: introducing mini-challenges. *IEEE* 2008:195–6.
- [12] Holme P, Saramäki J. Temporal networks. *Phys Rep* 2011;519:97–125.
- [13] Katz L. A new index derived from sociometric data analysis. *Psychometrika* 1953;18:39–43.
- [14] Klimt B, Yang Y. In: *Introducing the Enron corpus*. First conference on email and anti-spam (CEAS). CA: Mountain View; 2004.
- [15] Laffin P, Mantzaris AV, Grindrod P, Ainley F, Otley A, Higham DJ. Discovering and validating influence in a dynamic online social network. *Soc Network Anal Min* 2013;3:1311–23.
- [16] Newman M. *Networks: an introduction*. Oxford University Press; 2010.
- [17] Newman MEJ. A measure of betweenness centrality based on random walks. *Soc Networks* 2005;27:39–54.
- [18] Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G, Latora V. Graph metrics for temporal networks. In: Holme P, Saramäki J, editors. *Temporal networks. Understanding complex systems*. Springer; 2013. p. 15–40.
- [19] Rogers T. Null models for dynamic centrality in temporal networks, *J Complex Networks*, in press.
- [20] Tang J, Musolesi M, Mascolo C, Latora V. Temporal distance metrics for social network analysis. In: *Proceedings of the 2nd ACM SIGCOMM workshop on online social networks (WOSN09)*, Barcelona; 2009.
- [21] Tang J, Musolesi M, Mascolo C, Latora V, Nicosia V. *Analysing information flows and key mediators through temporal centrality metrics*. In: *3rd ACM EuroSys workshop on social network systems*, Paris, France; 2010.
- [22] Wasserman S, Faust K. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press; 1994.