



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A matrix iteration for dynamic network summaries

Citation for published version:

Grindrod, P & Higham, DJ 2013, 'A matrix iteration for dynamic network summaries', *Siam review*, vol. 55, no. 1, pp. 118-128. <https://doi.org/10.1137/110855715>

Digital Object Identifier (DOI):

[10.1137/110855715](https://doi.org/10.1137/110855715)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Siam review

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A Matrix Iteration for Dynamic Network Summaries*

Peter Grindrod[†]
Desmond J. Higham[‡]

*Submitted to the SIAM Review Research Spotlights section as a short,
Timely Communication.*

Abstract. We propose a new algorithm for summarizing properties of large-scale time-evolving networks. This type of data, recording connections that come and go over time, is generated in many modern applications, including telecommunications and online human social behavior. The algorithm computes a dynamic measure of how well pairs of nodes can communicate by taking account of routes through the network that respect the arrow of time. We take the conventional approach of downweighting for length (messages become corrupted as they are passed along) and add the novel feature of downweighting for age (messages go out of date). This allows us to generalize widely used Katz-style centrality measures that have proved popular in network science to the case of dynamic networks sampled at nonuniform points in time. We illustrate the new approach on synthetic and real data.

Key words. centrality, communicability, dynamic network, Katz centrality, social network analysis, telecommunication, resolvent

AMS subject classifications. 65F60, 91D30, 05C82

DOI. 10.1137/110855715

1. Introduction. Large-scale complex networks arise in a range of natural and technological settings [7, 19], and they pose many challenges to applied and computational mathematicians. Many key ideas in this area came from the field of social network analysis [23], at a time when interactions were typically recorded link by link in the field and systems involved tens or perhaps hundreds of players. Nowadays, however, the same principles are being applied to networks in biology, online behavior, and telecommunications, where data involving millions of nodes or more can be generated and stored automatically.

The motivation for our work is that many emerging interaction data sets involve an element of time [11]. For example, human social contact can be monitored in relation to emails and phone calls [6, 22], online chats [12], allegiances during massive online gaming [20], and physical proximity [13]. In neuroscience we can record which

*Received by the editors November 18, 2011; accepted for publication (in revised form) May 20, 2012; published electronically February 7, 2013. This work was supported by the Engineering and Physical Sciences Research Council and the Research Councils UK Digital Economy program through the MOLTEN (Mathematics Of Large Technological Evolving Networks) project, EP/I016058/1, and HORIZON, EP/G065802/1.

<http://www.siam.org/journals/sirev/55-1/85571.html>

[†]Department of Mathematics, University of Reading, Reading, UK.

[‡]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK (d.j.higham@strath.ac.uk). The work of this author was supported by a Fellowship from the Leverhulme Trust.

brain regions have correlated activity as a task is performed [1] and in e-business we may be told that “people who bought this book also bought...” [16]. In these cases, links are transient, and after dividing the time axis into discrete units (seconds, minutes, hours, ...) we obtain a *time-ordered sequence of networks*. Our aim here is to present an algorithm that is able to summarize key features in this type of data in a manner that (a) respects the time-dependency and (b) generalizes previously developed algorithms.

We focus on *centrality measures* that identify key nodes. Many alternatives have been proposed and evaluated for static networks [3, 7, 19]. Our approach is motivated by the original work of Katz [14], and in the time-dependent context we have in mind questions such as

- who is currently most likely to know the latest news, import the latest software virus, or catch the latest strain of influenza?
- who is currently most effective at broadcasting the latest news, spreading the latest software virus, or passing on the latest strain of influenza?

In the next section, we set up the notation and motivate a new algorithm, which is then described in section 3. A key feature of the algorithm is illustrated on specially constructed synthetic networks in section 4, and two types of real interaction data are used in section 5. We finish with a brief discussion in section 6.

2. Notation and Background. Suppose we have a time-ordered sequence of unweighted graphs defined over a set of N nodes. Given the time points $t_0 < t_1 < \dots < t_M$, we let $A^{[k]}$ denote the adjacency matrix for the network at time t_k . So the i, j entry of $A^{[k]}$ equals 1 if there is a link from node i to j at time t_k and equals zero otherwise. Directed links, where $A_{ij}^{[k]} \neq A_{ji}^{[k]}$, are allowed, but we have no self-loops, so $A_{ii}^{[k]} \equiv 0$.

We let $\Delta t_i := t_i - t_{i-1}$ denote the spacing between successive time points. We do not assume that the time points are equally spaced. Nonuniform spacing is natural, for example, if we have time-stamped emails or text messages with each $A^{[k]}$ recording one event.

In this setting, we can envisage excursions around the network that respect the arrow of time. For example, if A links with B today and B links with C tomorrow, then there is a natural route from A to C but (in the absence of any other links) there is not a natural route from C to A. With this philosophy, *dynamic walks* were defined in [10] as follows.

DEFINITION 1. 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 $A_{i_m, i_{m+1}}^{[r_m]} \neq 0$.

This concept was used to motivate the definition of a *dynamic communicability matrix*, which we write here as $\mathcal{Q}^{[M]}$, where, more generally,

$$(2.1) \quad \mathcal{Q}^{[k]} = \left(I - aA^{[0]}\right)^{-1} \left(I - aA^{[1]}\right)^{-1} \dots \left(I - aA^{[k]}\right)^{-1}.$$

The parameter a is assumed to satisfy $a < 1/\max_k \rho(A^{[k]})$, where $\rho(\cdot)$ denotes the spectral radius. This ensures that the resolvents in (2.1) exist and may be expanded according to $(I - aA^{[p]})^{-1} = I + aA^{[p]} + a^2(A^{[p]})^2 + \dots$. It is then straightforward to see that $(\mathcal{Q}^{[k]})_{ij}$ is a weighted sum of the number of dynamic walks from i to j using the ordered sequence $\{A^{[0]}, A^{[1]}, A^{[2]}, \dots, A^{[k]}\}$, where the number of walks of length w is scaled by a^w . The key idea here is that each possible walk around the network

from node i to node j adds to the communicability measure, but longer walks are less influential than shorter walks. In the static case, where $k = 0$, this leads to the classical Katz centrality [14], and we note that Katz also offered the interpretation that a represents the probability that a message successfully traverses an edge.

It is important to note that the use of $\mathcal{Q}^{[M]}$ is strongly tied to the idea of a *start point* t_0 and an *end point* t_M . Any walk that took place in the time period t_0, \dots, t_M has equal influence. Also, by construction, the elements in $\mathcal{Q}^{[M]}$ are nonnegative and nondecreasing with k , so that pairs of nodes cannot become less communicative over time. These features are appropriate in some applications, e.g., if the networks represent functional connectivity between brain regions in the course of a well-defined task [1]. However, there are many applications where we are interested in the *current* and *recent* activity, but not in the activity that took place a long time ago—messages go out of date, rumors lose their timeliness, and some viruses become less infectious.

In this work we take the view that it is of interest to know whether node i *recently* had the opportunity to get a message to node j using *short walks*. At one extreme the matrix $A^{[k]}$ gives us the most localized picture, telling us what is possible using single steps with only today's connectivity. At the other extreme, the matrix $\mathcal{Q}^{[k]}$ gives us the most historical view, telling us what is possible using all walks over all the connections that ever existed up to the current time. In the next section we present a matrix iteration that interpolates between these two extremes.

3. New Iteration. Suppose we have several months' worth of hourly email or phone activity, starting from some arbitrary day zero. It would be of interest to compute a time-dependent "running summary" of communicability between pairs of nodes. Here, at each point in time, we wish to quantify the capability of node i to pass messages to node j , where

- (i) as we discussed in the previous section and has been used in the derivation of many centrality measures for static networks [4, 5, 7, 8, 14, 19], *long walks are less important than short walks*, but also, in this time-dependent setting,
- (ii) *walks that started recently are more important* than walks that started a long time ago.

Part (ii) is the novel feature that we incorporate in this work.

These requirements motivate the idea of a *running dynamic communicability matrix*, $\mathcal{S}^{[k]}$, based on two parameters, $a \in (0, 1)$ and $b > 0$. Here, as in (2.1), a is used to downweight walks of length w by the factor a^w . To explain the new parameter b , we refer to the current *age*, t , of a dynamic walk as the time that has elapsed since the walk began. The parameter b is then used to further downweight by the age-dependent factor e^{-bt} . So b is used to filter out "old" activity.

We therefore propose the following iteration, where, for convenience, $\mathcal{S}^{[-1]} = 0$,

$$(3.1) \quad \mathcal{S}^{[k]} = (I + e^{-b\Delta t_k} \mathcal{S}^{[k-1]})(I - aA^{[k]})^{-1} - I, \quad k = 0, 1, 2, \dots$$

To understand how this works, we can expand the right-hand side of (3.1) as

$$(3.2) \quad aA^{[k]} + a^2 A^{[k]^2} + \dots + a^r A^{[k]^r} + \dots$$

$$(3.3) \quad + e^{-b\Delta t_k} \mathcal{S}^{[k-1]}$$

$$(3.4) \quad + e^{-b\Delta t_k} \mathcal{S}^{[k-1]} aA^{[k]} + e^{-b\Delta t_k} \mathcal{S}^{[k-1]} a^2 A^{[k]^2} + \dots + e^{-b\Delta t_k} \mathcal{S}^{[k-1]} a^r A^{[k]^r} + \dots$$

This leads to the following interpretation.

- The terms in (3.2) give a length-weighted count of all walks that start and finish at the current time, t_k .

- The term in (3.3) deals with all “old” walks that do not involve time t_k . These get downweighted by the time factor $e^{-b\Delta t_k}$, because the age of each such walk has increased by Δt_k .
- The terms in (3.4) deal with all walks that began at an earlier time but make use of one or more edges at the current time, t_k . The factor $e^{-b\Delta t_k}$ is used again, because the age of each such walk has increased by Δt_k . Then we have a length-downweighting factor a^r if r new edges are used.

So, overall, we get what we wanted:

The i, j element of the matrix $\mathcal{S}^{[k]}$ records a scaled count of the number of dynamic walks from i to j that can be taken with the time-ordered sequence $A^{[0]}, A^{[1]}, A^{[2]}, \dots, A^{[k]}$. The scaling comprises the product of (a) a factor a^w for walks of length w and (b) a factor e^{-bt} for walks that began t time units ago.

The factor $e^{-b\Delta t_k}$ in (3.1) may be interpreted as the probability that a message does not become “irrelevant” (or a virus doesn’t mutate into a harmless form) over a time length Δt_k . We also note that the iteration automatically incorporates the case of nonuniform time spacing.

There are, of course, many other possible choices for downweighting. On static networks the inverse factorial $1/w!$ for walks of length w , which leads to the matrix exponential, has proved popular [7]. However, the factor a^w is particularly convenient when we move into the realm of dynamic networks, since the basic law of indices $a^{w_1}a^{w_2} = a^{w_1+w_2}$ allows us to do the combinatorics across time points through simple matrix products in (2.1). The same reasoning also explains why the functional form $e^{-b\Delta t}$ was chosen in (3.1).

When $b = 0$ (no downscaling in time), we essentially recover the original iteration (2.1) from [10]; we have $\mathcal{Q}^{[k]} = I + \mathcal{S}^{[k]}$. On the other hand, for $b = \infty$, that is, $e^{-b\Delta t_k} \equiv 0$ (complete downscaling in time), we revert to Katz static centrality with $\mathcal{S}^{[k]} = (I - aA^{[k]})^{-1} - I$.

A simple variation of (3.1) restricts our attention to walks using at most one edge per time point. (For example, the time taken to pass a message along an edge may be comparable with a typical Δt .) In this case, we just replace the resolvent in (3.1) by its first two terms:

$$(3.5) \quad \mathcal{S}^{[k]} = (I + e^{-b\Delta t_k} \mathcal{S}^{[k-1]})(I + aA^{[k]}) - I.$$

More generally, the resolvent expansion could be truncated after more terms if it is appropriate to impose some other upper limit on the number of edges traversed in a single time point.

We also point out that in the case of discrete, instantaneous interaction, such as email and text, it is attractive to take Δt so small that each $A^{[k]}$ records a single event. In this way, the requirement for the networks to be unweighted does not incur a loss of detail.

We may summarize down to the level of nodal information by aggregating the ability of a node to communicate with, or receive communication from, every other node. In this way,

$$(3.6) \quad \mathcal{S}^{[k]} \mathbf{1} \quad \text{and} \quad \mathcal{S}^{[k]T} \mathbf{1},$$

where $\mathbf{1} \in \mathbb{R}^N$ denotes the vector of ones, give running versions of the dynamic *broadcast* and *receive* communicabilities introduced in [10]—the i th components indicate the current propensity for node i to act as a source or sink, respectively, of information.

4. Illustration on Synthetic Data. We now give an artificially constructed example that illustrates the new algorithm. Our network sequence has 31 nodes over 64 time points. We regard the time points as representing days, with Δt equal to one day, and suppose that over a certain time period, day 17 to day 48, the links are the consequence of rumors that originate from node 1. These rumors cascade from node 1 across the network: a node that receives a rumor passes it on to a new recipient the next day. In this manner node 1 gets messages across the network with very little effort in a way that is difficult to unravel from either snapshots or aggregate data. It is the timing and local follow-on effect of the links that distinguishes node 1 from its peers. We will show that this well-hidden and transient broadcasting role is not revealed by existing centrality measures, but can be uncovered by a suitable choice of parameters in the new algorithm.

To begin, we construct directed networks on days 1 to 16 and days 49 to 64 that consist of “noise” in the form of independent Erdős–Rényi graphs, so each directed edge appears on each of these days with independent probability $p = 2/N$. Each node therefore has an average of two outgoing edges on these days. To describe days 17 to 48, consider a directed binary tree with the obvious labeling: node 1 at level 1, nodes 2 and 3 at level 2, nodes 4 to 7 at level 3, nodes 8 to 15 at level 4, and nodes 16 to 31 at level 5. We switch on different levels at various times.

- L1: Level 1 switched on: there are connections $1 \rightarrow 2$ and $1 \rightarrow 3$.
- L2: Level 2 switched on: there are connections $2 \rightarrow 4$ and $2 \rightarrow 5$, and $3 \rightarrow 6$ and $3 \rightarrow 7$.
- L3: Level 3 switched on: there are connections $4 \rightarrow 8$ and $4 \rightarrow 9$, up to $7 \rightarrow 14$ and $7 \rightarrow 15$.
- L4: Level 4 switched on: there are connections $8 \rightarrow 17$ and $8 \rightarrow 18$, up to $15 \rightarrow 30$ and $15 \rightarrow 31$.

In this way, we use the directed binary tree structure to cascade rumors that start at node 1. At day 17, we use L1 and at day 18 we use L2. Then from days 19 to 48 we alternate between “L1 plus L3” and “L2 plus L4.” So on odd days there is a new rumor from node 1 that is passed to level 2, and the previous rumor (from two days ago) is passed from level 3 to level 4. On even days the most recent rumor is passed from level 2 to level 3, and the previous rumor is passed from level 3 to level 4.

Note that by “staggering” the cascade, we get quite a subtle effect over this time period. In particular, node 1 does not have high bandwidth: it is switched off every other day, and even when it is switched on, it has the same number of outward links as four other nodes.

For this data, the biggest spectral radius over all days is $\max_k \rho(A^{[k]}) = 2.7$, so the upper limit for a in the resolvent is 0.37. We use $a = 0.3$ in these tests.

Figure 4.1 shows the rank of node 1 over time, as measured by the running broadcast communicability from (3.6). A rank of 1 means that node 1 is the best broadcaster, and a rank of 31 means it is the worst. The upper picture uses $b = 0.6$, and in this case the algorithm has picked out the special nature of node 1 over the period where the rumor cascading takes place. From around day 20, the node builds up in rank, until it becomes one of the top performers (despite its modest bandwidth). After day 48, when the cascading stops, the node loses rank considerably. The middle picture uses $b = 0.1$. In this case, we are close to the iteration (2.1) from [10]. We are not discounting old walks sufficiently to capture the cascading effect—the “noise” at the start of the time period continues to influence the result. The lower picture shows $b = 1$. Here we are close to using static Katz centrality in each snapshot—old walks are heavily penalized and we are viewing each day almost in isolation. Because

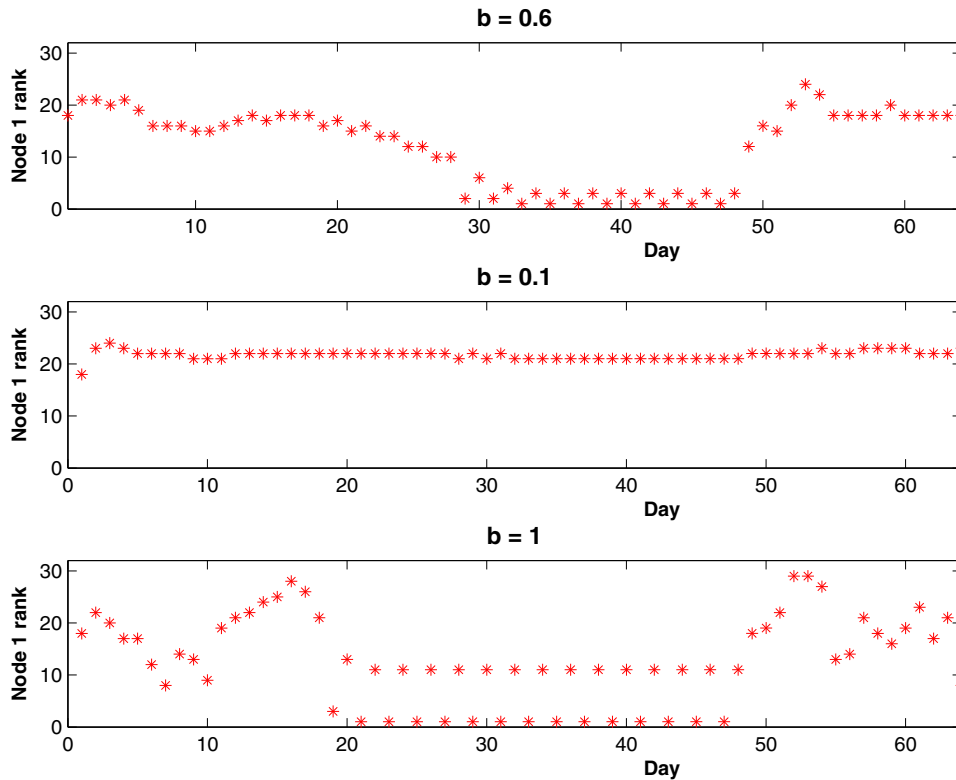


Fig. 4.1 Rank of node 1 in the synthetic test (rank = 1 denotes that this node is currently the best broadcaster). Upper: $b = 0.6$. Middle: $b = 0.1$. Lower: $b = 1$. Cascade takes place over days 17 to 48.

we localize so strongly in time, node 1 immediately reduces in rank on days 18, 20, 22, \dots , 48, where it has no activity. If we take this to the extreme case of $e^{-b\Delta t} = 0$, then, during the cascade period, when node 1 switches on it will share the highest rank with the level 3 nodes (nodes 4, 5, 6, 7), and when it switches off it will share the lowest rank with the level 3 and level 5 nodes.

The question of whether properties like those of node 1 in this example are of interest, and, if so, how to quantify them, clearly depends on the application. In general, this issue requires us to understand, or estimate, the natural time scales at work; that is, how quickly relevance decays over time. This is reflected in the fact that the plots in Figure 4.1 are sensitive to the parameter b (whereas tests in [10] found that a could be changed by an order of magnitude without a major effect on the node rankings). Our aim here is to present the new algorithm and show that it can discover interesting “hidden” features if b is chosen appropriately. Automating the choice of b via problem-specific knowledge or data-driven analysis is an important area for future work. The second experiment in section 5 suggests one approach.

5. Real Human Interaction Data. We now illustrate the new iteration on two real social interaction data sets. The first experiment gives a feel for the overall smoothing effect of the time-downweighting parameter b . In the second experiment

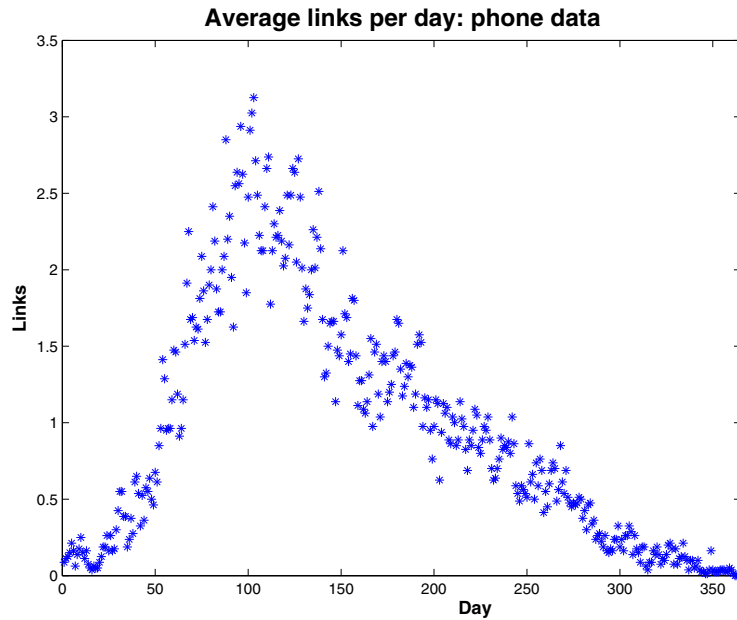


Fig. 5.1 “Reality Mining” data. Average links per node on each day.

we focus on the use of $\mathcal{S}^{[k]}$ to define the time-dependent broadcast and receive centralities (3.6) and judge the quality of these measures for one-day-ahead predictions of Katz centrality. We emphasize that these computations are presented for illustrative purposes—quantifying the effectiveness of network summaries and centrality measures is not a well-defined task, and it necessarily depends on the type of data under consideration and the issues of interest.

We begin with the “Reality Mining” telecommunication data from [6]. In this case there are 106 nodes and we summarize into 365 days: $(A^{[k]})_{ij} = (A^{[k]})_{ji} = 1$ means that nodes i and j communicated by telephone at least once on day k . Here $\max_k \rho(A^{[k]}) = 8.2$, so the upper limit for a is 0.12, and we use $a = 0.05$.

In Figure 5.1 we show how the average number of links per node varies over time. We see a build up of activity that peaks at around day 100. Figure 5.2 summarizes the network communicability by showing the evolution of $\|S^{[k]}\|$, scaled by its maximum, where $\|\cdot\|$ denotes the Euclidean norm. The upper picture uses $b = \infty$ (more precisely, $e^{-b\Delta t} = 0$ in (3.1)), which corresponds to the static Katz centrality applied at each daily snapshot. We see that this measure closely follows the overall links-per-day structure across time. The middle picture uses $b = 0.2$, so $e^{-b\Delta t} = e^{-b} = 0.82$, allowing us to capture the effect of earlier activity propagating across the network. We see that the running centrality measure presents a much smoother temporal summary and highlights a later period, from around days 140 to 200, as the most influential, with distinct bimodal structure. There is a “lag time” of around 50 days between peaks for the static and running summaries. This lag time was found to depend quite strongly on the parameter b , but the bimodal structure of the running summary was persistent: the lower picture uses $b = 0.4$, so $e^{-b\Delta t} = e^{-b} = 0.67$, and we see that the lag is much shorter. With b increased to 0.6 the picture becomes very similar to the $b = \infty$ case.

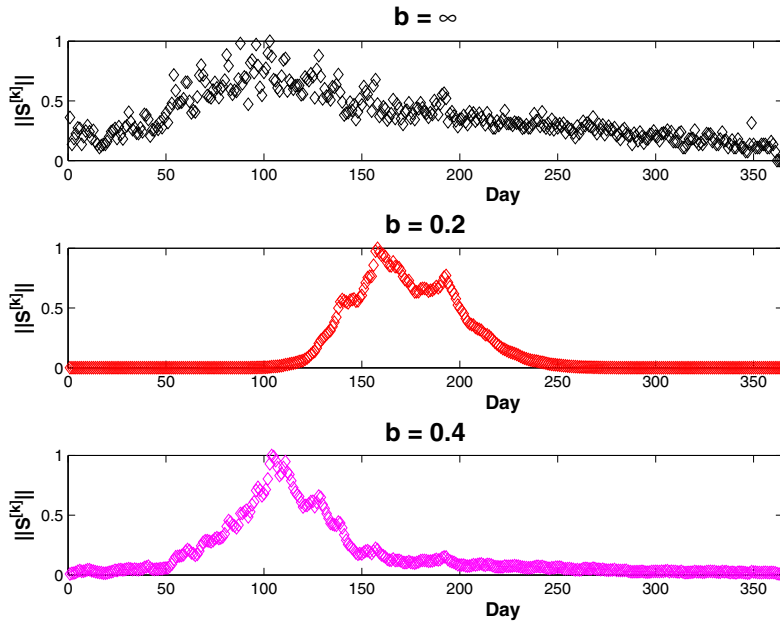


Fig. 5.2 Test on “Reality Mining” data. Upper: size of the running dynamic communicability matrix, $\mathcal{S}^{[k]}$, when $b = \infty$, equivalent to Katz centrality. Middle: $b = 0.2$. Lower: $b = 0.4$.

The second experiment uses email data between Enron employees [15], as studied previously, for example, in [10, 21]. We have 151 nodes in the network with interaction summarized over 1138 days. Because email is a one-way form of communication, we use directed links, with $(A^{[k]})_{ij} = 1$ if person i emailed person j at least once on day k . We regard To:, cc:, and bcc: as equivalent, for simplicity. Figure 5.3 displays the average number of links per node over time. We have $\max_k \rho(A^{[k]}) = 3.93$, giving an upper limit of 0.25 for a , and we choose $a = 0.2$. Our aim is to use current information $\{A^{[0]}, A^{[1]}, A^{[2]}, \dots, A^{[k]}\}$ to predict the next day’s behavior. We focus on centrality and target tomorrow’s Katz-style broadcast and receive centralities, given by

$$(5.1) \quad (I - aA^{[k+1]})^{-1}\mathbf{1} \quad \text{and} \quad (I - aA^{[k+1]})^{-T}\mathbf{1},$$

respectively. At each time point $k = 0, 1, \dots, M - 1$ we test whether the dynamic broadcast communicabilities assigned to each node by (3.6) match tomorrow’s actual Katz broadcast centralities in (5.1) by computing the Pearson correlation coefficient between the two vectors. We sum the correlation coefficients over all time points (ignoring cases where the adjacency matrix, $A^{[k+1]}$, was empty) to give an overall quantification of how well the dynamic summary predicts future behavior, with a larger value indicating better performance. We also do the same for receive centrality.

The upper and lower pictures in Figure 5.4 give the results for broadcast and receive centralities, respectively, as a function of the time-downweighting parameter, b . These are normalized to have a maximum of 1. Here, $b = \infty$ corresponds to using only today’s network to predict tomorrow’s and $b = 0$ corresponds to letting all earlier days have equal influence. We see that the best performance arises with an intermediate case of $b \approx 0.3$, where, for example, walks that started three days ago are

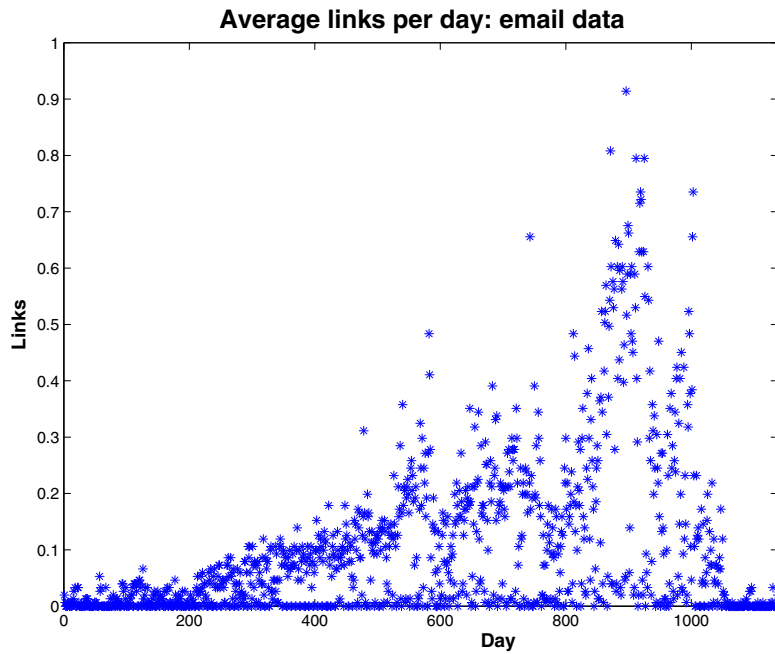


Fig. 5.3 *Enron email data. Average links per node on each day.*

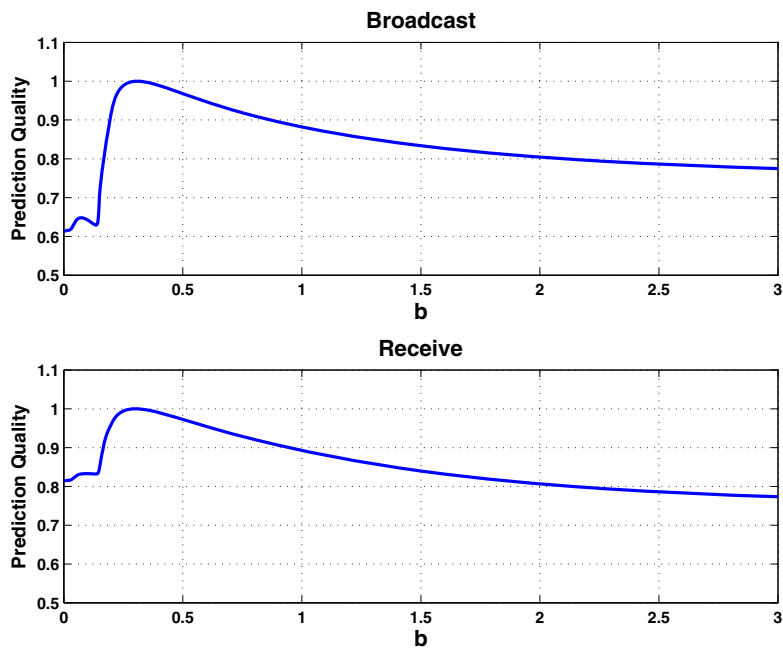


Fig. 5.4 *Test on Enron email data. Upper: prediction quality for broadcast centrality as a function of the downweighting parameter b . Lower: receive centrality.*

downweighted by $e^{-3b} \approx 0.41$ and walks that started two weeks ago are downweighted by $e^{-14b} \approx 0.015$. This test indicates how the choice of b can be informed by studying a quantity of interest on real data.

6. Discussion. We believe that the new iteration (3.1) can form the basis of many useful tools aimed at analyzing temporal networks. For example:

- If the iteration is applied in real time, the resulting centrality measures (3.6) can identify the latest “hot” players and those currently on rapid upward or downward trajectories.
- The dynamic summary $\mathcal{S}^{[k]}$ can be used to predict likely upcoming links (estimating future bandwidth) or to interfere effectively with the network evolution (suggesting new friends in an online social network or recommending movies in a rental service). Here the static Katz measure, corresponding to $b = \infty$, has proved successful [17].
- The behavior of $\mathcal{S}^{[k]}$ for appropriate time-dependent random graph models, for example, from the class in [9], could provide a baseline against which to calibrate unusual network behavior.
- By applying a clustering method to the running dynamic communicability matrix $\mathcal{S}^{[k]}$ at each time point, we could detect sets of nodes that organize into tightly connected communities and then disperse over time. This would give an alternative to the approach in [18].
- The running dynamic broadcast and receive centralities in (3.6) could be used to rank nodes and compare them with known hierarchical structure—for example, job status within a company—in order to see who is punching above their weight.
- In general, $\mathcal{S}^{[k]}$ will evolve into a dense matrix, so very-large-scale applications are not feasible. It is therefore of interest to devise fast and low storage algorithms that use approximations of $\mathcal{S}^{[k]}$, along the lines of those developed in the static case [2].

Acknowledgment. The Enron email data was originally collected and prepared by members of the CALO Project (A Cognitive Assistant that Learns and Organizes). We obtained the data from the website <http://www.cs.cmu.edu/~enron/>.

REFERENCES

- [1] D. S. BASSETT, N. F. WYMBES, M. A. PORTER, P. J. MUCHA, J. M. CARLSON, AND S. T. GRAFTON, *Dynamic reconfiguration of human brain networks during learning*, Proc. Nat. Acad. Sci. USA, 18 (2011), pp. 7641–7646.
- [2] F. BONCHI, P. ESFANDIAR, D. F. GLEICH, C. GREIF, AND L. V. S. LAKSHMANAN, *Fast matrix computations for pair-wise and column-wise commute times and Katz scores*, Internet Math., 8 (2012), pp. 73–112.
- [3] S. BORGATTI, K. CARLEY, AND D. KRACKHARDT, *On the robustness of centrality measures under conditions of imperfect data*, Social Networks, 28 (2006), pp. 124–136.
- [4] S. BORGATTI, A. MEHRA, D. BRASS, AND G. LABIANCA, *Network analysis in the social sciences*, Science, 323 (2009), pp. 892–895.
- [5] J. J. CROFTS AND D. J. HIGHAM, *Googling the brain: Discovering hierarchical and asymmetric network structures, with applications in neuroscience*, Internet Math. (Special Issue on Biological Networks), 7 (2011), pp. 1–22.
- [6] N. EAGLE AND A. PENTLAND, *Reality mining: Sensing complex social systems*, Personal Ubiquitous Comput., 10 (2006), pp. 255–268.
- [7] E. ESTRADA, *The Structure of Complex Networks*, Oxford University Press, Oxford, 2011.
- [8] E. ESTRADA, N. HATANO, AND M. BENZI, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119.

- [9] P. GRINDROD AND D. J. HIGHAM, *Evolving graphs: Dynamical models, inverse problems and propagation*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 466 (2010), pp. 753–770.
- [10] P. GRINDROD, D. J. HIGHAM, M. C. PARSONS, AND E. ESTRADA, *Communicability across evolving networks*, Phys. Rev. E, 83 (2011), 046120.
- [11] P. HOLME AND J. SARAMÄKI, *Temporal networks*, Phys. Rep., 519 (2012), pp. 97–125.
- [12] D. HUFFAKER, *Dimensions of leadership and social influence in online communities*, Human Commun. Res., 36 (2010), pp. 593–617.
- [13] L. ISELLA, M. ROMANO, A. BARRAT, C. CATTUTO, V. COLIZZA, W. VAN DEN BROECK, F. GESUALDO, E. PANDOLFI, L. RAVÀ, C. RIZZO, AND A. E. TOZZI, *Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors*, PLoS ONE, 6 (2011), e17144.
- [14] L. KATZ, *A new index derived from sociometric data analysis*, Psychometrika, 18 (1953), pp. 39–43.
- [15] B. KLIMT AND Y. YANG, *Introducing the Enron corpus*, in First Conference on Email and Anti-Spam (CEAS), Mountain View, CA, 2004.
- [16] J. LESKOVEC, L. A. ADAMIC, AND B. A. HUBERMAN, *The dynamics of viral marketing*, in Proceedings of the 7th ACM Conference on Electronic Commerce (EC '06), ACM, New York, 2006, pp. 228–237.
- [17] Z. LU, B. SAVAS, W. TANG, AND I. DHILLON, *Supervised link prediction using multiple sources*, in Proceedings of the 10th IEEE International Conference on Data Mining (ICDM), 2010, pp. 923–928.
- [18] P. J. MUCHA, T. RICHARDSON, K. MACON, M. A. PORTER, AND J.-P. ONNELA, *Community structure in time-dependent, multiscale, and multiplex networks*, Science, 328 (2010), pp. 876–878.
- [19] M. E. J. NEWMAN, *Networks: An Introduction*, Oxford University Press, Oxford, 2010.
- [20] M. SZELL, R. LAMBIOTTE, AND S. THURNER, *Multirelational organization of large-scale social networks*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 13636–13641.
- [21] J. TANG, M. MUSOLESI, C. MASCOLO, V. LATORA, AND V. NICOSIA, *Analysing information flows and key mediators through temporal centrality metrics*, in Proceedings of the 3rd Workshop on Social Network Systems (SNS '10), ACM, New York, 2010, pp. 1–6.
- [22] J. TANG, S. SCCELLATO, M. MUSOLESI, C. MASCOLO, AND V. LATORA, *Small-world behavior in time-varying graphs*, Phys. Rev. E, 81 (2010), 05510.
- [23] S. WASSERMAN AND K. FAUST, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, UK, 1994.