School of Mathematical
and Physical Sciences

Communicability Across Evolving Networks

Peter Grindrod[†] Desmond J. Higham[‡] Mark C. Parsons[§]
Ernesto Estrada[¶]

Abstract

Many natural and technological applications generate time ordered sequences of networks, defined over a fixed set of nodes; for example time-stamped information about 'who phoned who' or 'who came into contact with who' arise naturally in studies of communication and the spread of disease. Concepts and algorithms for static networks do not immediately carry through to this dynamic setting. For example, suppose A and B interact in the morning, and then B and C interact in the afternoon. Information, or disease, may then pass from A to C, but not vice versa. This subtlety is lost if we simply summarize using the daily aggregate network given by the chain A-B-C. However, using a natural definition of a walk on an evolving network, we show that classic centrality measures from the static setting can be extended in a computationally convenient manner. In particular, communicability indices can be computed to summarize the ability of

PACS: 89.75.Fb, 89.75.Hc, 84.40.Ua

1 Motivation

At the heart of network science are the well established mathematical fields of deterministic and random graph theory, with concepts such as connectedness, pathlength, diameter, degree and clique playing key roles [8, 21]. The motivation for this work is that a new type of time-dependent network-based object is emerging from a range of digital technologies that requires a fundamentally different way of thinking.

In Figure 1 we show a simple example of an evolving network, where undirected connections between a fixed set of seven nodes is recorded over three days. If we regard the links as representing communication, for example, by telephone or email, then we see that A may pass a message to C through the links A $\leftrightarrow$ B and B $\leftrightarrow$ G on day 1 and then through the links G $\leftrightarrow$ E and E $\leftrightarrow$ C on day 2. However, there is no way for C to pass a message to A. Analogously, if the links represent physical proximity, then A may pass an infection to C but C cannot cause A to be infected. This asymmetry, which arises even though each individual network is symmetric, is caused by the arrow of time. It is clear that simply aggregating the individual networks

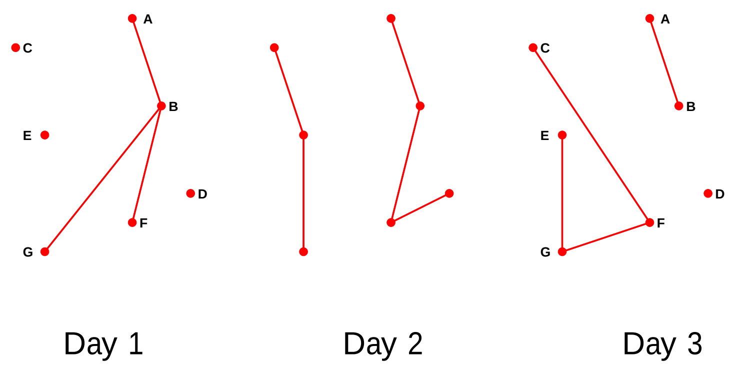

Figure 1: Simple example of an evolving network.

- networks of online social users (e.g., Facebook) interacting through messaging [25] or online chatting systems (e.g. MSN) [19],
- networks of travellers, vehicles or available routes defined over a dynamic transportation infrastructure [14, 18],
- networks describing transient social interactions over cyberspace [23],
- networks describing individuals' attendance at regularly scheduled events over time [1],
- correlated neural activity in response to a functional task [15].

respect the time ordering of the data—for example, backward and forward running clocks would be treated similarly.

Let us emphasize at this stage that unlike in the well-studied 'network growth' context, where new nodes and accompanying edges are accumulated further benefit of the walk counting approach is that the combinatorics can be conveniently described and implemented in terms of operations in linear algebra, and we will show that this feature can be carried through to the dynamic case.

We emphasize that the sequence of times $t_{r_1}, t_{r_2}, \ldots, t_{r_w}$ in Definition 1 must be nondecreasing, in order to respect the arrow of time, but

- repeated times are allowed: for example, if $r_1 < r_2 = r_3 < r_4$ then precisely two edges are followed at time t_{r_2},
- times are not required to be consecutive: for example, if $r_2 > r_1 + 1$ then the networks corresponding to times in between t_{r_1} and t_{r_2} have not been used during the walk.

Of course, depending on the application area, it may be reasonable to alter these features; forcing at most one edge per time level and/or forcing time levels to be consecutive. The ideas presented here could be adjusted accordingly.

Our key observation, which generalizes a simple result from graph theory (see, for example, [10, Lemma 1.1]) is that the matrix product $A^{[r_1]}A^{[r_2]} \ldots A^{[r_w]}$ has i, j element that counts the number of dynamic walks of length w from node

The use of the identity matrices in (2) is crucial in our target case of large, sparse networks—it allows a message to 'wait' at a node until a suitable connection appears at a later time.

Overall, as required, the matrix Q records the sum of all terms of the form (1). We may therefore use Q_{ij} as our summary of how well information can be passed from node i to node j. The nth row and column sums

$$C_n^{\text{broadcast}} := \sum_{k=1}^{N} Q_{nk} \quad \text{and} \quad C_n^{\text{receive}} := \sum_{k=1}^{N} Q_{kn} \tag{3}$$

are centrality measures that quantify how effectively node n can *broadcast* and *receive* messages, respectively[1].

Because we are interested in the relative values of the centrality measures across all nodes, rather than their absolute sizes, it makes sense to avoid under or overflow in the computation of Q using an iteration such as

$$Q^{k+1} = \quad Q$$

centrality measures reduce to multiples of the aggregate out and in degrees, shifted by unity;

$$\lim_{a \to 0^+} \frac{C_n^{\text{broadcast}} - 1}{}$$

It follows that

$$S(Q) = I + a \sum_{p=0}^{M} A^{[p]} + O(a^2), \tag{4}$$

3.2 Telecommunication Data

We now consider telecommunication data from [7]. We have daily "who phoned who" information between 106 individuals based at M.I.T. over 365

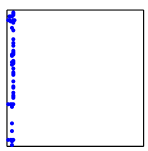

Figure 3: Adjacency matrices for the M.I.T. telecommunication data, symmetrized and aggregated into 13 sets of 28 day windows.

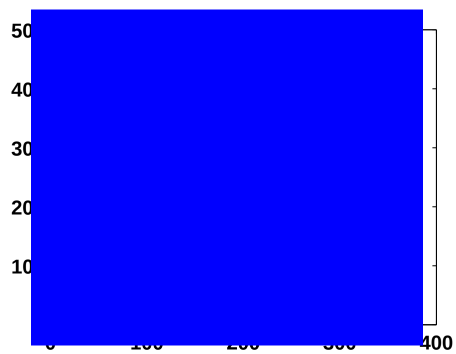

Figure 4: Daily M.I.T. telecommunication data. Upper right: total activity per day. Upper left: Broadcast versus receive centrality; correlation 0.14, top twenty overlap size 4. Lower left: broadcast centrality versus total degree; correlation 0.50, top twenty overlap size 9. Lower right: Receive centrality versus total degree; correlation 0.28, top twenty overlap size 6.

Overall, the experiments indicate that the two new measures deliver distinct information that is different from a raw degree count, and remains consistent over a range of a values.

3.3 Email Data

We now consider a public domain data set concerning email activities of Enron employees. In [5] the static, aggregate network was analysed, but here we treat it as an evolving network. We constructed daily information representing emails between 151 Enron employees, including to, cc or bcc. So $A_{ij}^{[k]} = 1$ if employee i sent at least one message to employee j on day k, but

because this type of ... eeuni1(use)-22ti(da)2use(t)27(y)unidirecti(ded,)-but)4

Figure 5: Daily M.I.T. telecommunication data. Upper: $a = 0.1$ versus $a = 0.05$; correlations are 0.93 for broadcast and 0.98 for receive, respective top twenty overlap sizes are 14 and 16. Lower: $a = 0.05$ versus $a = 0.01$; correlations are 0.82 for broadcast and 0.81 for receive, respective top twenty overlap sizes are 16 and 11.

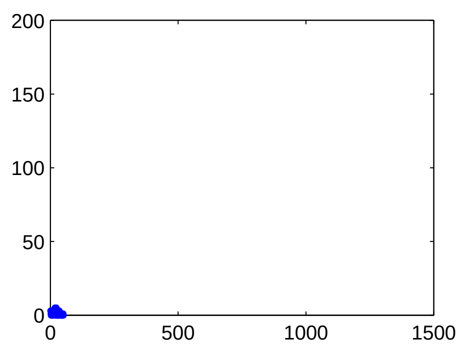

Figure 6: Results for Enron email data. Upper left: total number of edges per day. Upper right: Scatter plot of broadcast and receive centralities; correlation 0.00, top twenty overlap size 2. Lower left: Scatter plot of broadcast centrality and total out degree; correlation 0.62, top twenty overlap size 11 Lower right: Scatter plot of receive centrality and total in degree; correlation 0.28, top twenty overlap size 6.

time steps 0 up to M we have

$$E(Q|A^{[0]}) = I + a\left[R_M \circ (A^{[0]} - A^{[\infty]}) + (M+1)A^{[\infty]}\right] + O(a^2),$$

where R_M is the symmetric matrix given by $(R_p)_{ij} = (1 - (1 - \alpha_{ij} + \omega_{ij})^{M+1})/(\alpha_{ij} + \omega_{ij})$, and $\circ$ denotes componentwise multiplication. This quantifies the relative contributions to Q made by the initial condition and the long term expected *equilibrium* value for each edge. So if the observer wishes tervsumma1(givzplica-27(n263existi-4.4g7(n263dat14on263setsJ/3se)14.44d-23s63real-1imeJ/3se)1

[6] J. J. Crofts and D. J. Higham, *A weighted communicability measure applied to complex brain networks*, J. Royal Society Interface, 33 (2009).

[7] N. Eagle, A. S. Pentland, and D. Lazer, *Inferring friendship network structure by using mobile phone data*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 15274–15278.

[8] E. Estrada, M. Fox, D. J. Higham, and G.-L. Oppo, eds., *Network Science: Complexity in Nature and Technology*, Springer, Berlin, to appear, 2010.

[9] E. Estrada and N. Hatano, *Communicability in complex networks*, Phys. Rev. E, 77 (2008), p. 036111.

[10] E. Estrada and D. J. Higham, *Network propeties revealed through matrix functions*, SIAM Review, 52 (2010), pp. 696–671.

[11]

[18] L. McNamara, C. Mascolo and L. Capra, *Media Sharing based on Colocation Prediction in Urban Transport*, in Proc. of ACM 14th International Conference on Mobile Computing and Networking (Mobicom08), San Francisco, CA, September 2008, pp. 58–69.

[19] J. Leskovec and E. Horvitz, *Planetary-scale views on a large instant-messaging network*, in Proceeding of the 17th international conference on World Wide Web, Beijing, China, 2008, ACM, pp. 915–924.

[20] 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.

[21] M. E. J. Newman, *The structure and function of complex networks*, SIAM Review, 45 (2003), pp. 167–256.

[22] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, *Temporal distance metrics for social network analysis*, in Proceedings of the 2nd ACM SIGCOMM Workshop on Online Social Networks (WOSN09), Barcelona, 2009.

[23] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, *Characterising temporal distance and reachability in mobile and online social networks*, SIGCOMM Comput. Commun. Rev., 40 (2010), pp. 118–124.

[24] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, *Analysing information flows and key mediators through temporal centrality metrics*, in SNS '10: Proceedings of the 3rd Workshop on Social Network Systems, New York, NY, USA, 2010, ACM, pp. 1–6.

[25] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, *Small-world behavior in time-varying graphs*, Physical Review E, 81 (2010), p. 05510.