

A Model for Dynamic Communicators

Alexander V. Mantzaris and Desmond J. Higham

Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK

(Received 12 December 2011)

We develop and test an intuitively simple dynamic network model to describe the type of time-varying connectivity structure present in many technological settings. The model assumes that nodes have an inherent hierarchy governing the emergence of new connections. This idea draws on newly established concepts in on-line human behavior concerning the existence of discussion catalysts, who initiate long threads, and on-line leaders, who trigger feedback. We show that the model captures an important property found in email and voice call data—‘dynamic communicators’ with sufficient foresight or impact to generate effective links have an influence that is grossly underestimated by static measures based on snapshots or aggregated data.

Key Words: Social networks, complex networks, stochastic models

1 Introduction

Random graph models and centrality measures have provided extremely useful tools in network science [25]. However, the fundamental ideas in this area are tied to the concept

of a single, static network. Many emerging network data sets are *dynamic*; links between nodes may appear and disappear in a time-dependent manner. Examples arise naturally when we measure email activity [2, 12, 11], voice calls [6, 11, 20], online social interaction [12, 29], geographical proximity of mobile device users [16, 29], dynamic transportation infrastructure [4, 8, 21], voting and trading patterns [1, 23] and neural activity [3, 10], and also when *link prediction* [7, 22] is required.

For this reason, new models and algorithms are needed to address dynamic structures [10]. We emphasize that in this setting it is not just the ‘final state’ of an iterative process that is of interest. Instead we are concerned with the real-time dynamics—what mechanisms drive the continual change in topological structure, and how do we summarize key properties of a dynamic network?

2 Background and Motivation

Figure 1 shows a hypothetical scenario of communication between a set of 21 nodes over three days. This undirected, unweighted, network sequence has been constructed so that node 21 does not appear to be unusually important when we consider any single day, or the aggregate over the three days. However, closer inspection shows that the *timing* of node 21’s links is special. A message from node 21 may reach nodes 19 and 20 on day one, nodes 16 to 18 on day two and nodes 12 to 15 on day three. We could interpret node 21 as being an *influential* player—when other individuals receive a message that can be traced back to node 21 they burst into action and pass the message on. Alternatively, we could interpret node 21 as being a *knowledgable* player who can accurately predict, and thereby exploit, the future network structure, perhaps from experience, expertise or

insider information. In this work, after summarizing the recent ideas from [11] that allow us to quantify the intuitive notion that node 21 is special, we confirm that the same phenomenon is seen in real communication data. We then introduce a new, general dynamic network model, based on simple but intuitively reasonable principles, that captures the effect.

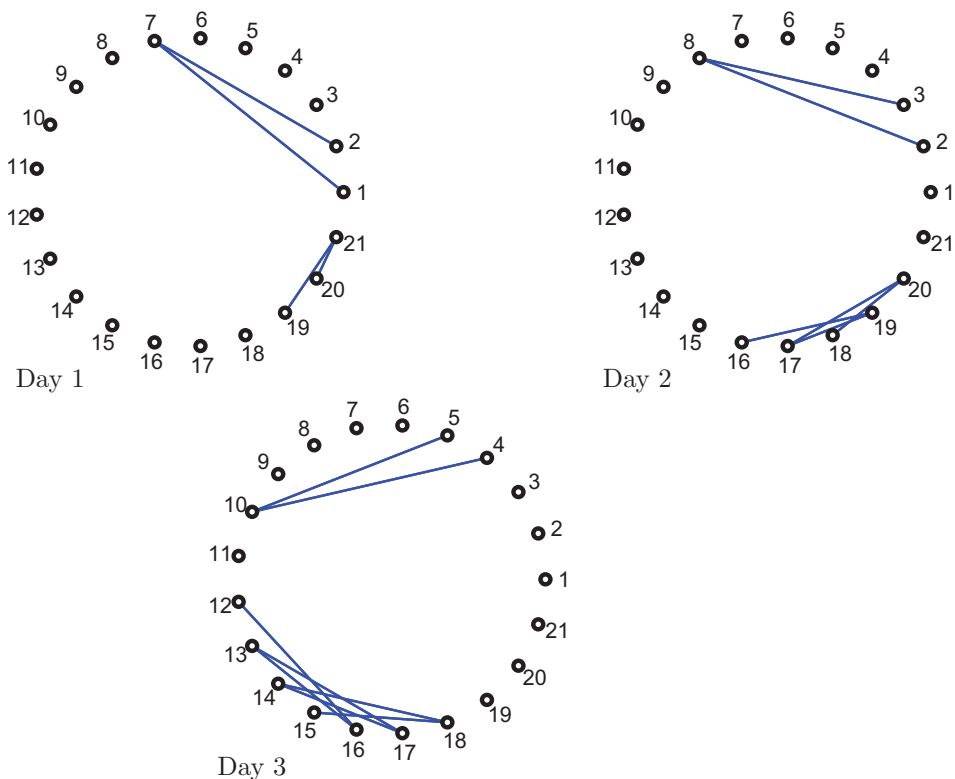


FIGURE 1. An ordered sequence of three undirected and unweighted networks.

Following the notation in [11], for a fixed set of N nodes and time points $t_0 < t_1 < \dots < t_M$, we consider an ordered sequence of unweighted graph adjacency matrices $A^{[k]} \in \mathbb{R}^{N \times N}$, so that $(A^{[k]})_{ij} = 1$ if there is a link from node i to node j at time t_k and $(A^{[k]})_{ij} = 0$ otherwise. A *dynamic walk of length w* is any traversal along w edges, where the appearance of the edges must respect the arrow of time. We note that even

in the case of undirected networks, where each $A^{[k]}$ is symmetric, dynamic walks lack symmetry in general. For example, in Figure 1 there is a dynamic walk of length two from node 7 to node 8 (using $7 \rightarrow 2$ on day one and $2 \rightarrow 8$ on day two) but there are no dynamic walks from node 8 to node 7. In [11] it was shown how to compute the matrix $\mathcal{Q} \in \mathbb{R}^{N \times N}$ for which $(\mathcal{Q})_{ij}$ is 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 a^w . Here a is an appropriately chosen fixed parameter that downweights the contribution of longer walks. The corresponding row and column sums

$$C_n^{\text{broadcast}} := \sum_{k=1}^N \mathcal{Q}_{nk} \quad \text{and} \quad C_n^{\text{receive}} := \sum_{k=1}^N \mathcal{Q}_{kn} \quad (2.1)$$

are centrality measures that quantify how effectively node n can *broadcast* and *receive* dynamic messages. In practice, because we are typically concerned with ranking the nodes, it is preferable to compute with a normalized matrix $\mathcal{Q}/\|\mathcal{Q}\|$ in order to avoid numerical under or overflow. These measures reduce to the classic Katz centralities [17, 25] when there is a single time point ¹.

Related work in [26, 27, 28] has also devised centrality measures that respect time dependency, based on shortest paths rather than walks. We also note that paths and related concepts under time-dependent connectivity have been studied previously. Berman [4] considered dynamic networks where each edge has a start and finish time, and looked at global connectivity issues. Related work in the case where each edge exists at a single instant of time appeared in [18]. Spread of information or disease across a time-dependent

¹ As discussed in the original work of Katz [17], the downweighting parameter a may also be interpreted as the probability that a message successfully traverses an edge.

contact network was considered in in [12], whereas [19] focuses on the issue of optimal routes to pass the most timely information. An approach for discovering temporal communities is given in [23], with extra links being added to the network sequence in order to represent the passage of time. The ideas in [11] differ from those mentioned above by focusing on individual nodes and all possible communication routes that respect the arrow of time. In this work we use the data-driven tools from [11] to motivate and test a new mathematical model for network evolution.

We can now quantify our intuitive arguments concerning the role of node 21 in Figure 1. This node has the largest dynamic broadcast centrality, $C_n^{\text{broadcast}}$, (using $a = 0.5$) while ranking much lower according to static measures: there are 5 nodes with higher positions in terms of overall degree, 7 nodes have higher positions in terms of the maximum over $k = 1, 2, 3$ of Katz centrality at each time point (using $a = 0.4$), and node 21 lies in 8th position in terms of Katz centrality on the aggregate network $\sum_{k=1}^3 A^{[k]}$ (using $a = 0.3258$). We will use the term *dynamic communicator* to describe a node of this type; that is, having excellent centrality in the dynamic sense that is not apparent when we consider only snapshot or aggregate views of the network sequence. These players can distribute a message, or spread a disease, across the network in a manner that efficiently exploits the transient nature of the links. For simplicity we focus here on broadcast centrality, but we note later that a symmetry argument allows us to cover the case of receiving.

3 Practical Observations

We continue by showing that dynamic broadcasters can be found consistently in real communication data sets. For Figure 2 we use two weeks of Enron email data [11]. In this case there are $N = 151$ nodes and in the upper picture a) we use a time resolution $t_{i+1} - t_i$ of one day—we have $M + 1 = 14$ time points, and $(A^{[k]})_{ij} = 1$ signifies that at least one email (to, cc or bcc) was sent from person i to person j on day k . The horizontal axis records the total out degree, that is, the aggregate bandwidth generated over the whole time period, for each person. The vertical axis represents the broadcast centrality $C_n^{\text{broadcast}}$ from (2.1). Five nodes are highlighted and labelled in subfigure a). Nodes 1 and 2 have the highest broadcast centrality but modest total out degree—these are examples of dynamic communicators. Follow-up analysis shows that they correspond to an executive and the vice president, who can be speculated as having a large influence. The nodes labelled 3, 4 and 5 have high bandwidth but relatively poor broadcast centrality. Nodes 3 and 4 correspond to traders in the company and node 5 has an unknown role. The figure highlights that an exceptionally high out degree is neither necessary nor sufficient to guarantee influence amongst other nodes in this time-dependent setting. To test for consistency at a different time resolution, in subfigure b) we split the same data differently, with $t_{i+1} - t_i$ covering a two day period, giving $M + 1 = 7$ time points made from pairs of consecutive days. The same two dynamic communicators are observed. An extra node, labeled number 6, has also emerged as another example with high bandwidth but relatively poor broadcast centrality. This node corresponds to an employee.

Figure 3 shows results of a similar experiment using 30 days of voice call data between

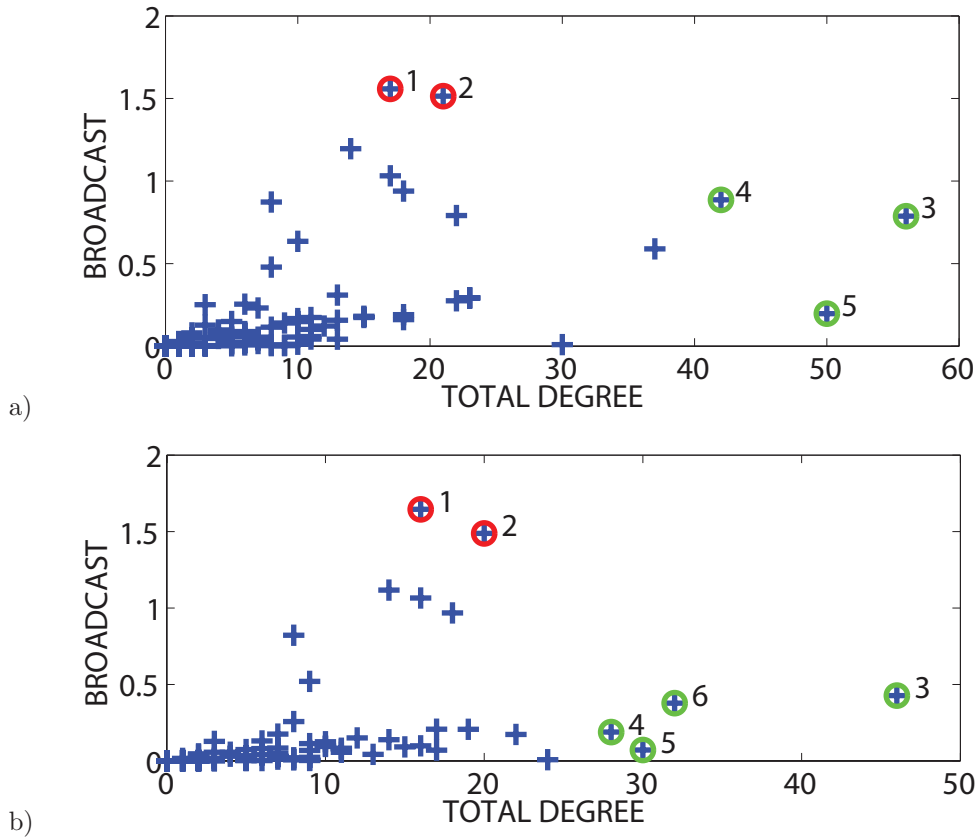


FIGURE 2. Broadcast centrality against total out degree across a 14 day subset of Enron email interaction [11]. Time resolution is one day for subfigure a) and two days for subfigure b). The same two dynamic communicators, labeled 1 and 2, stand out each time.

academics [6]. Subfigure a) displays the results with a time resolution of one day and subfigure b) uses pairs of days. We regard the nodes labelled ‘A’ as dynamic communicators due to their large broadcast measure and low total degree. By contrast, nodes labelled ‘B’ have a relatively high total degree, so their ability to broadcast is much less surprising. As in Figure 2, changing the time resolution has not affected which nodes emerge as dynamic communicators.

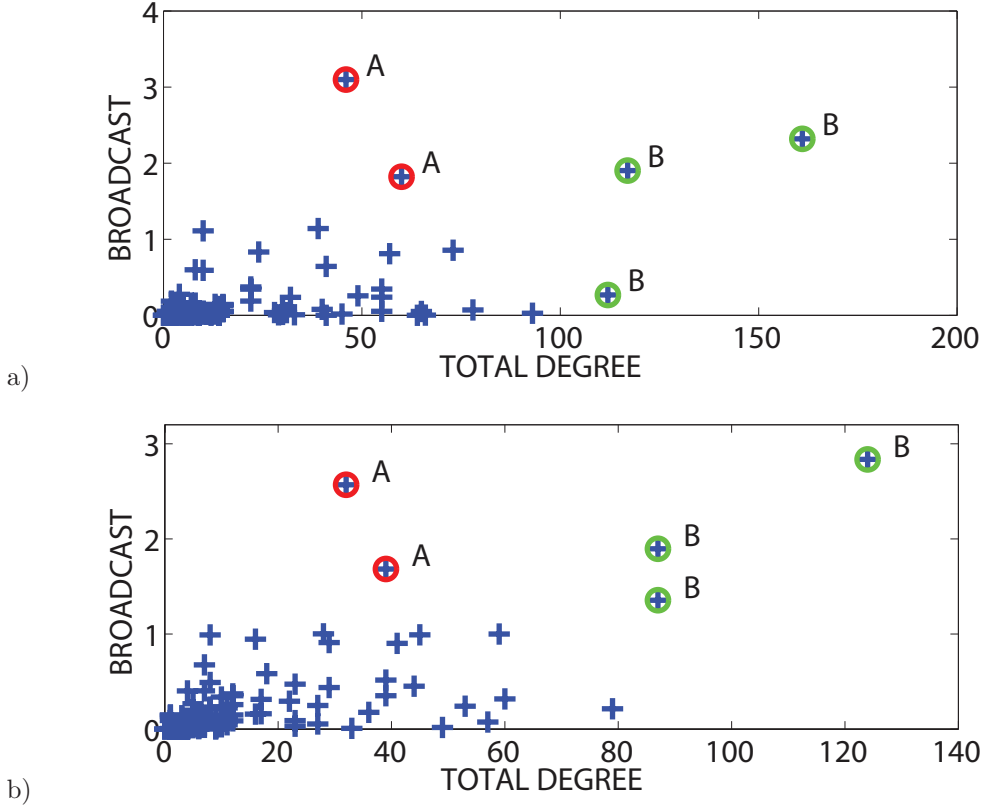


FIGURE 3. Broadcast centrality against total out degree across 30 days of MIT voice call data.

Subfigure a) uses a time window of a single day and subfigure b) two consecutive days.

4 New Model

The model that we propose for explaining this phenomenon can be motivated as generalizing the concept of network hierarchy from the static case [24]. In the new dynamic setting, we assume that there is an underlying hierarchy such that some nodes have enhanced importance, causing their links to have a knock-on effect at future times. (As mentioned earlier, an alternative interpretation with ‘prescience’ instead of ‘importance’ is also possible.) This hierarchy may arise through an imposed chain of command, as in business or military organisations, through a more subtle structure of the type observed

in social or criminal networks [5], or may be earned through completion of tasks, as in on-line gaming [15]. The idea that hierarchy can impact communication structure is intuitively reasonable, and is supported in the social sciences by, for example, the empirical discovery of *discussion catalysts* in an on-line community who are “responsible for the majority of messages that initiate long threads” [9]. Further, Huffaker [14] identifies *on-line leaders* who have the ability to “trigger feedback, spark conversations within the community, or even shape the way that other members of a group ‘talk’ about a topic.” We will incorporate these ideas into a discrete time, discrete space Markov chain, in order to build on the successful ‘random graph’ models [25] from the static setting, and the more general dynamic framework of [10].

Compared with the static case, relatively little attention has been paid to developing mathematical descriptions of temporal networks [13]. From a modelling perspective, our work shares with [2, 30] the aims of (a) identifying a key feature in dynamic human interaction data sets and (b) offering a simple, intuitively reasonable, explanatory mechanism. However, unlike those references, we are not focusing on *when*—i.e., the precise timing of events for a single player. Instead we focus on *where*—i.e., the particular pairs of players involved in each interaction—while accounting for the time ordering of the events.

Our model begins by assigning a fixed level of importance, l_n , to each node n . We order the nodes so that $0 < l_1 \leq l_2 \leq \dots \leq l_N$. The key concept in the model is that from one time point to the next, a node is responsive to the total importance of its current links—messages received from more highly ranked nodes are more likely to generate follow-on communication. More precisely, given the time t_k network, $A^{[k]}$, we generate $A^{[k+1]}$ as

follows. For each node, n , new, undirected, links appear in the row and column of n of $A^{[k+1]}$ as the result of two processes; basal and responsive.

- *Basal*: with probability b node n generates a fixed number c_b of links, with the new neighbours chosen uniformly and independently at random. Otherwise no basal links are generated from node n .

- *Responsive*: with probability

$$r_n^{[k]} := \frac{\sum_{i=1}^N l_i (A^{[k]})_{in}}{1 + l_N \sum_{i=1}^N (A^{[k]})_{in}}$$

node n generates a fixed number c_r of links, with the new neighbours chosen uniformly and independently at random. Otherwise no responsive links are generated from node n .

Here $0 < b < 1$ and the positive integers c_b and c_r are fixed parameters. The probability $r_n^{[k]}$ summarizes the current importance of connections involving node n relative to the maximum possible value, with a shift of one added in the denominator to deal with unconnected nodes. Repeated edges are of course removed and for simplicity we consider links to be undirected.

In Figure 4 we show computational results for the case where $l_n = e^n$, $b = 0.01$, $c_b = 1$ and $c_r = 4$, with $N = 40$ nodes over 365 timepoints. The node at the top of the hierarchy, $n = 40$, is marked with a circle. Results are shown for scaling parameters $a = 0.75, 0.5, 0.25$ in order to check for consistency against the choice of downweighting. In each case node 40 operates as a dynamic broadcaster: despite ranking 26th in terms of aggregate degree, it is able to communicate effectively across the network in a well-

defined temporal sense, since its links carry a level of importance that creates a knock-on effect.

Because of the combinatorial, walk-counting derivation of \mathcal{Q} in (2.1) it is straightforward to check that the broadcast and receive centralities for each node are swapped when we reverse the direction of each link and also reverse the arrow of time. It follows that node 21 in Figure 1 and node 50 in Figure 4 would become examples of *dynamic receivers* if we supply the adjacency matrices in reverse order. For these people, without taking full account of the temporal connectivity patterns, we could easily underestimate their ability to accumulate information, or their chance of becoming infected. We argued that the dynamic broadcasters may be the nodes with added importance, in the sense that their links automatically generate a follow-on response, or added predictive power, in the sense that they preferentially link to nodes that are about to become active. There are similar passive and active explanations for the existence of dynamic receivers. They may have added importance, in the sense that they are the preferred point of contact for any node in the network that is currently bursting with information or passing on requests for advice—for example, in massive multiplayer online role-playing games a preference has been observed for players to send messages to higher level players [15]. More actively, dynamic receivers may have added *global, historical knowledge*, in the sense that they know which nodes are currently most informative and deliberately form links with them.

5 Conclusions

Our main aim in this work was to propose and study a new model that describes the dynamic appearance and disappearance of connections in an evolving network. The model

quantifies the intuitively simple notion of an underlying hierarchy of nodal importance, prescience or global knowledge. In practice this may arise directly through an imposed managerial or chain-of-command structure, or more subtly through social status, or intelligence. Computational simulations confirmed that the new model captures an effect that can be found in communication data—certain individuals, referred to here as *dynamic communicators*, are able to punch above their weight in the sense that standard centrality measures based on snapshots or aggregate summaries of network activity grossly underestimate their ability to interact with other members of the community.

Following on from this work, there is great potential for testing for the existence of dynamic communicators in other classes of evolving network, calibrating models of this type against real data, and investigating the role of dynamic communicators when there is a second source of dynamic behaviour taking place over the evolving network structure, such as a stochastic susceptible/infected/recovered disease propagation model.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council and the Research Councils UK Digital Economy Programme, under grant EP/I016058/1.

References

- [1] P. BAJARDI, A. BARRAT, F. NATALE, L. SAVINI, AND V. COLIZZA, *Dynamical patterns of cattle trade movements*, PLoS ONE, 6 (2011), p. e19869.
- [2] A.-L. BARABÁSI, *The origin of bursts and heavy tails in human dynamics*, Nature, 435 (2005), pp. 207–211.

- [3] D. S. BASSETT, N. F. WYMBS, 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., 108 (2011), p. doi: 10.1073/pnas.1018985108.
- [4] K. BERMAN, *Vulnerability of scheduled networks and a generalization of Menger's Theorem*, Networks, 28 (1996), pp. 125–134.
- [5] N. COLES, *It's not what you know-it's who you know that counts. Analysing serious crime groups as social networks*, Brit. J. Criminology, 41 (2001), pp. 580–594.
- [6] N. EAGLE, A. S. PENTLAND, AND D. LAZER, *Inferring friendship network structure by using mobile phone data*, Proc. Nat. Acad. Sci., 106 (2009), pp. 15274–15278.
- [7] P. ESFANDIAR, F. BONCHI, D. GLEICH, C. GREIF, L. LAKSHMANAN, AND B.-W. ON, *Fast Katz and commutators: Efficient estimation of social relatedness in large networks*, in Algorithms and Models for the Web-Graph, R. Kumar and D. Sivakumar, eds., vol. 6516 of Lecture Notes in Computer Science, Springer Berlin/Heidelberg, 2010, pp. 132–145.
- [8] A. GAUTREAU, A. BARRAT, AND M. BARTHELEMY, *Microdynamics in stationary complex networks*, Proc. Nat. Acad. Sci., 106 (2009), pp. 8847–8852.
- [9] E. GLEAVE, H. T. WELSER, T. M. LENTO, AND M. A. SMITH, *A conceptual and operational definition of 'social role' in online community*, vol. 0, Los Alamitos, CA, USA, 2009, IEEE Computer Society, pp. 1–11.
- [10] P. GRINDROD AND D. J. HIGHAM, *Evolving graphs: Dynamical models, inverse problems and propagation*, Proc. Roy. Soc. A, 466 (2010), pp. 753–770.
- [11] P. GRINDROD, D. J. HIGHAM, M. C. PARSONS, AND E. ESTRADA, *Communicability across evolving networks*, Physical Review E, 83 (2011), p. 046120.
- [12] P. HOLME, *Network reachability of real-world contact sequences*, Physical Review E, 71 (2005).
- [13] P. HOLME AND J. SARAMÄKI, *Temporal Networks*, ArXiv e-prints, (2011).
- [14] D. HUFFAKER, *Dimensions of leadership and social influence in online communities*, Human Communication Research, 36 (2010), pp. 593–617.

- [15] D. HUFFAKER, J. A. WANG, J. TREEM, M. A. AHMAD, L. FULLERTON, D. WILLIAMS, M. S. POOLE, AND N. CONTRACTOR, *The social behaviors of experts in massive multi-player online role-playing games*, Computational Science and Engineering, IEEE International Conference on, 4 (2009), pp. 326–331.
- [16] L. ISELLA, M. ROMANO, A. BARRAT, C. CATTUTO, V. COLIZZA, W. VAN DEN BROECK, F. GESUALDO, E. PANDOLFI, L. RAV, 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), p. e17144.
- [17] L. KATZ, *A new index derived from sociometric data analysis*, Psychometrika, 18 (1953), pp. 39–43.
- [18] D. KEMPE, J. KLEINBERG, AND A. KUMAR, *Connectivity and inference problems for temporal networks*, J. Comput. Syst. Sci., 64 (2002), pp. 820–842.
- [19] G. KOSSINETIS, J. KLEINBERG, AND D. WATTS, *The structure of information pathways in a social communication network*, in Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '08, New York, NY, USA, 2008, ACM, pp. 435–443.
- [20] J. M. KUMPULA, J. P. ONNELA, J. SARAMÄKI, K. KASKI, AND J. KERTÉSZ, *Emergence of Communities in Weighted Networks*, Phys. Rev. Lett., 99 (2007), pp. 228701+.
- [21] 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.
- [22] Z. LU, B. SAVAS, W. TANG, AND I. DHILLON, *Supervised link prediction using multiple sources*, in Data Mining (ICDM), 2010 IEEE 10th International Conference on, Dec. 2010, pp. 923 –928.
- [23] P. J. MUCHA, T. RICHARDSON, K. MACON, M. A. PORTER, AND J.-P. ONNELA, *Commu-*

- nity structure in time-dependent, multiscale, and multiplex networks*, *Science*, 328 (2010), pp. 876–878.
- [24] L. MUCHNIK, R. ITZHAK, S. SOLOMON, AND Y. LOUZOUN, *Self-emergence of knowledge trees: extraction of wikipedia hierarchies*, *Phy. Rev. E*, 76 (2007), p. 016106.
- [25] M. E. J. NEWMAN, *Networks an Introduction*, Oxford University Press, Oxford, 2010.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] J. TANG, S. SCELLATO, M. MUSOLESI, C. MASCOLO, AND V. LATORA, *Small-world behavior in time-varying graphs*, *Phys. Rev. E*, 81 (2010), p. 05510.
- [30] K. ZHAO, J. STEHLÉ, G. BIANCONI, AND A. BARRAT, *Social network dynamics of face-to-face interactions*, *Physical Review E*, 83 (2011), pp. 056109+.

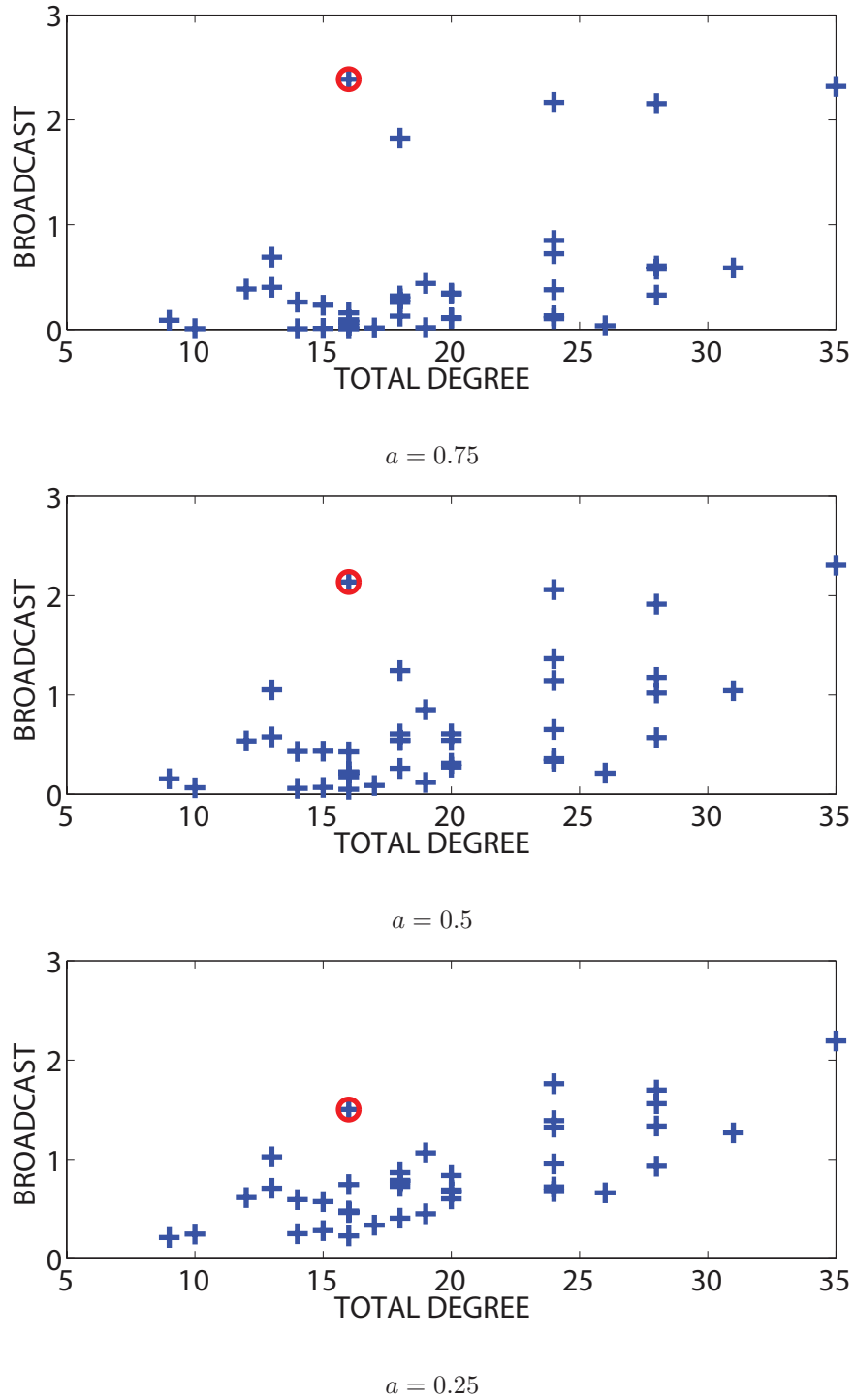


FIGURE 4. Broadcast centrality against total degree (horizontal axis) for network sequences generated from the new dynamic communicator model.