Euro. Jnl of Applied Mathematics (2012), *vol.* 23, *pp.* 659–668.
doi:10.1017/S0956792512000186

A model for dynamic communicators

ALEXANDER V. MANTZARIS and DESMOND J. HIGHAM

Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK
e-mail: d.j.higham@strath.ac.uk

(*Received 20 December 2011; revised 8 June 2012; accepted 11 June 2012; first published online 26 July 2012*)

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 online human behaviour concerning the existence of discussion catalysts, who initiate long threads, and online leaders, who trigger feedback. We show that the model captures an important property found in e-mail and voice call data – 'dynamic communicators' with sufficient foresight or impact to generate effective links and having an influence that is grossly underestimated by static measures based on snaphots 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 [26]. 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 e-mail activity [2, 11, 12], voice calls [6, 11, 20], online social interaction [12, 30], geographical proximity of mobile device users [16, 30], dynamic transportation infrastructure [4, 8, 23], voting and trading patterns [1, 24] 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 emphasise 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 summarise key properties of a dynamic network? Here, we develop a simple stochastic model for edge evolution motivated by (a) empirical studies from the social sciences, and (b) observations from customised centrality measures applied to large scale data sets.

In Section 2, we illustrate the importance of respecting time dependency. We use a centrality measure from Ref. [11] on a synthetic example to show that static summaries and snaphots may give a misleading impression about the influence of a node. This is followed up in Section 3 by tests on real human, social interactions where we identify *dynamic communicators*, whose impact hinges on the time dependency. Section 4 then introduces a new dynamic network model, where a built-in node hierarchy affects downstream

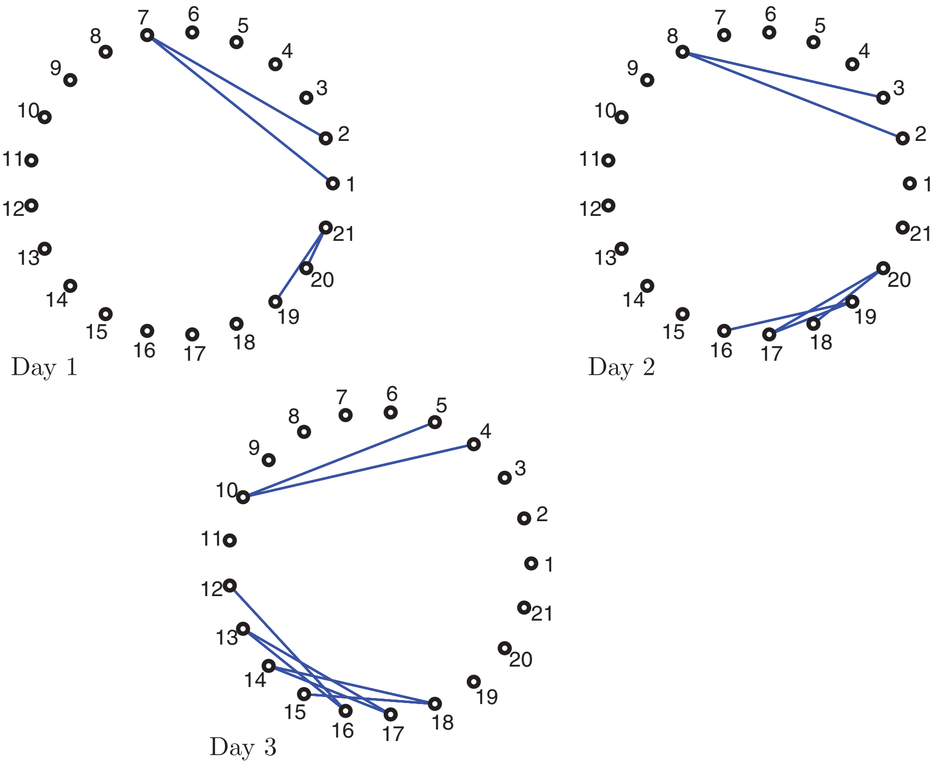


Figure 1. (Colour online) An ordered sequence of three undirected and unweighted networks.

interaction. We show that this simple feature is sufficient to generate dynamic communicators. Section 5 finishes with brief conclusions.

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 16–18 on day two and nodes 12–15 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 summarising the recent ideas from Ref. [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.

https://doi.org/10.1017/S0956792512000186 Published online by Cambridge University Press

Following the notation in Ref. [11], for a fixed set of N nodes and time points $t_0 < t_1 < \cdots < t_M$, 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 $7 \to 2$ on day one and $2 \to 8$ on day two), but there are no dynamic walks from node 8 to node 7. In Ref. [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

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

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 normalised matrix $\mathcal{Q}/\|\mathcal{Q}\|$ in order to avoid numerical under- or overflow. These measures reduce to the classic Katz centralities [17, 26] when there is a single time point.[1]

Related work in Refs. [27–29] has also devised centrality measures that respect time dependency, based on shortest paths rather than walks. We also note that paths and other graph-theoretical 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 Ref. [18]. Spread of information or disease across a time-dependent contact network was considered in Ref. [12], whereas Ref. [19] focusses on the issue of optimal routes to pass the most timely information. In Ref. [21], in the context of time-dependent links, algorithms were studied that place sensors at a subset of nodes in order to detect cascades, with the aim of minimising the expected disruption over a class of scenarios. An approach for discovering temporal communities is given in Ref. [24], with extra links being added to the network sequence in order to represent the passage of time. The ideas in Ref. [11] differ from those mentioned previously by focussing on individual nodes and all possible communication routes that respect the arrow of time. In this work, we use the data-driven tools from Ref. [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 five nodes with higher positions in terms of overall degree, seven nodes have higher positions in terms of the maximum over

[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.

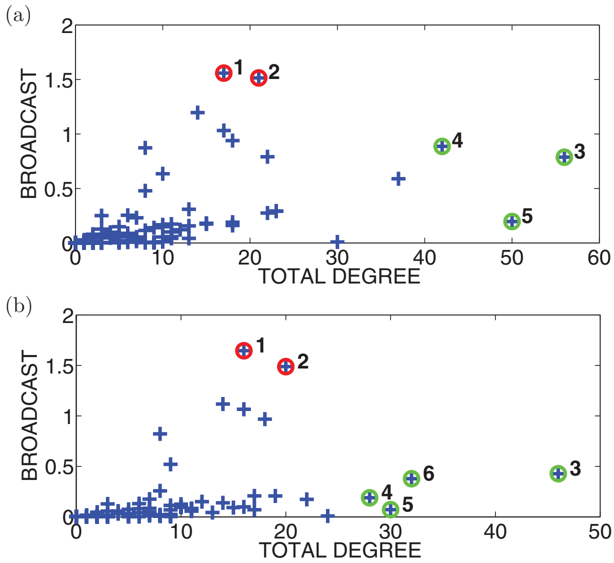


Figure 2. (Colour online) Broadcast centrality against total out degree across a 14-day subset of Enron e-mail interaction [11]. Time resolution is one day for (a) and two days for (b). The same two dynamic communicators, labelled 1 and 2, stand out in each case.

$k = 1, 2$ and 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 e-mail data [11]. In this case, there are $N = 151$ nodes, and in Figure 2(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 e-mail (`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 Figure 2(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

https://doi.org/10.1017/S0956792512000186 Published online by Cambridge University Press

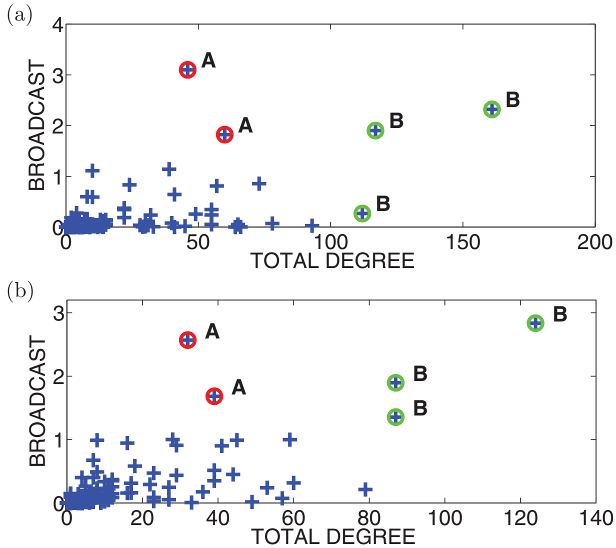


FIGURE 3. (Colour online) Broadcast centrality against total out degree across 30 days of MIT voice call data. (a) Uses a time window of a single day, and (b) two consecutive days.

to guarantee influence amongst other nodes in this time-dependent setting. To test for consistency at a different time resolution, in Figure 2(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, labelled 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 academics [6]. Figure 3(a) displays the results with a time resolution of one day and Figure 3(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 4 repeats the computations in Figures 2(a) and 3(a) with time's arrow in reverse. We see that node 2 and, especially, node 1 from the Enron e-mail data set have dramatically reduced in broadcast ability. Similarly, for the MIT voice call data, the nodes labelled A no longer stand out as dynamic communicators. This test emphasises that the timing of the interactions, in relation to follow-on activity, is a crucial component.

4 New model

The model that we propose for explaining this phenomenon can be motivated as generalising the concept of network hierarchy from the static case [25]. 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

https://doi.org/10.1017/S0956792512000186 Published online by Cambridge University Press

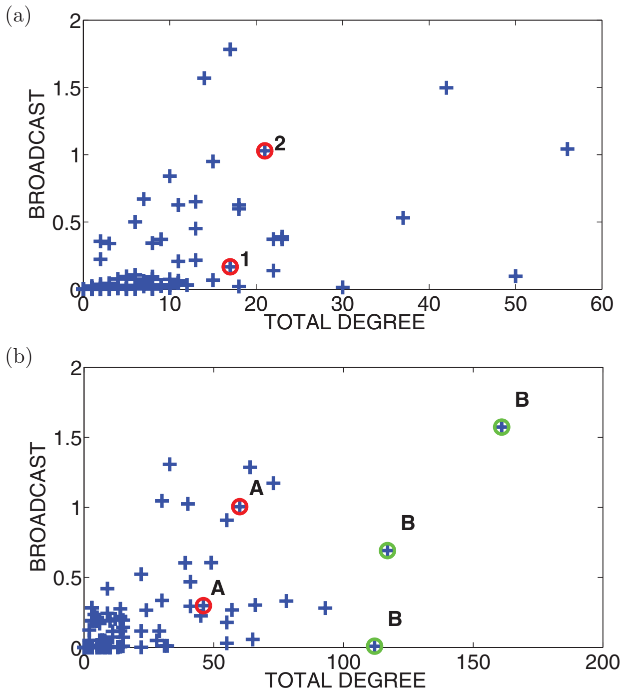


FIGURE 4. (Colour online) Repeat of the broadcast centrality *versus* total degree computations from Figures 2(a) and 3(a) with time ordering reversed. Upper: Enron data. Lower: MIT data.

mentioned previously, 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 online 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 online community who are 'responsible for the majority of messages that initiate long threads' [9]. Further, Huffaker [14] identifies *online 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 [26] 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 Refs. [2, 31] the aims of (a) identifying a key feature in dynamic human interaction data sets, and (b) offering a simple, intuitively reasonable and explanatory mechanism. However, unlike those references, we are not focussing on *when* – that is, the precise timing of events for a single player. Instead we focus on *where* – that is, 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 $0 < l_1 \leqslant l_2 \leqslant \cdots \leqslant l_N$. 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

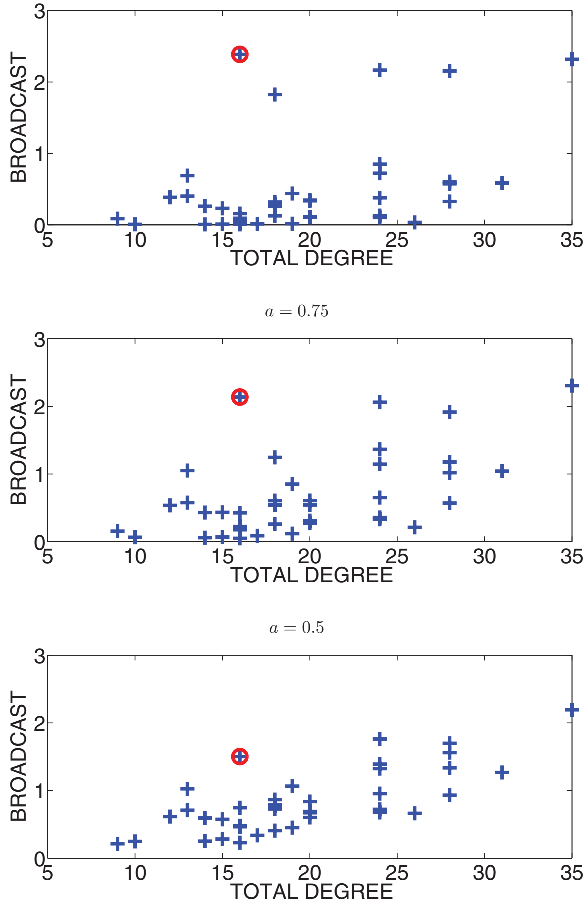
$$r_n^{[k]} := \frac{\sum_{i=1}^{N} l_i \left(A^{[k]}\right)_{in}}{1 + l_N \sum_{i=1}^{N} \left(A^{[k]}\right)_{in}},$$

  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 $r_n^{[k]}$ summarises 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 5 we show computational results for the case where $l_n = e^n$, $b = 0.01$, $c_{\mathrm{b}} = 1$ and $c_{\mathrm{r}} = 4$, with $N = 40$ nodes over 365 time points. 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 2 and node 40 in Figure 5 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



$a = 0.75$

$a = 0.5$

$a = 0.25$

FIGURE 5. (Colour online) Broadcast centrality against total degree (horizontal axis) for network sequences generated from the new dynamic communicator model.

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

https://doi.org/10.1017/S0956792512000186 Published online by Cambridge University Press

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. DJH was also supported by a Fellowship from the Leverhulme Trust.

References

[1] Bajardi, P., Barrat, A., Natale, F., Savini, L. & Colizza, V. (2011) Dynamical patterns of cattle trade movements. *PLoS ONE* **6**, e19869.

[2] Barabási, A.-L. (2005) The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211.

[3] Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M. & Grafton, S. T. (2011) Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA* **118**, 7641–7646.

[4] Berman, K. (1996) Vulnerability of scheduled networks and a generalization of Menger's theorem. *Networks* **28**, 125–134.

[5] Coles, N. (2001) It's not what you know—it's who you know that counts. Analysing serious crime groups as social networks. *Br. J. Criminol.* **41**, 580–594.

[6] Eagle, N., Pentland, A. S. & Lazer, D. (2009) Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **106**, 15274–15278.

[7] Esfandiar, P., Bonchi, F., Gleich, D., Greif, C., Lakshmanan, L. & On, B.-W. (2010) Fast Katz and commuters: Efficient estimation of social relatedness in large networks In: R. Kumar & D. Sivakumar (editors), *Algorithms and Models for the Web-Graph*, vol. 6516 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 132–145.

[8] Gautreau, A., Barrat, A. & Barthelemy, M. (2009) Microdynamics in stationary complex networks. *Proc. Natl. Acad. Sci. USA* **106**, 8847–8852.

[9] Gleave, E., Welser, H. T., Lento, T. M. & Smith, M. A. (2009) A conceptual and operational definition of 'social role' in online community. In: *Proceedings of the 42nd Hawaii International Conference on System Sciences*, Los Alamitos, CA, USA, IEEE Computer Society, pp. 1–11.

[10] Grindrod, P. & Higham, D. J. (2010) Evolving graphs: Dynamical models, inverse problems and propagation. *Proc. Roy. Soc.* A **466**, 753–770.

[11] Grindrod, P., Higham, D. J., Parsons, M. C. & Estrada, E. (2011) Communicability across evolving networks. *Phys. Rev.* E **83**, 046120.

[12] Holme, P. (2005) Network reachability of real-world contact sequences. *Phys. Rev.* E, **71**, 046119.

[13] Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.*, [online] URL: http://www.sciencedirect.com/science/article/pii/S0370157312000841.

[14] Huffaker, D. (2010) Dimensions of leadership and social influence in online communities. *Hum. Commun. Res.* **36**, 593–617.

[15] Huffaker, D., Wang, J. A., Treem, J., Ahmad, M. A., Fullerton, L., Williams, D., Poole, M. S. & Contractor, N. (2009) The social behaviors of experts in massive multiplayer online role-playing games. *IEEE Int. Conf. Comput. Sci. Eng.* **4**, 326–331.

[16] Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., Gesualdo, F., Pandolfi, E., Rav, L., Rizzo, C. & Tozzi, A. E. (2011) Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE* **6**, e17144.

[17] Katz, L. (1953) A new index derived from sociometric data analysis. *Psychometrika* **18**, 39–43.

[18] Kempe, D., Kleinberg, J. & Kumar, A. (2002) Connectivity and inference problems for temporal networks. *J. Comput. Syst. Sci.* **64**, 820–842.

[19] Kossinets, G., Kleinberg, J. & Watts, D. (2008) 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, USA, ACM, pp. 435–443.

[20] Kumpula, J. M., Onnela, J. P., Saramäki, J., Kaski, K. & Kertész, J. (2007) Emergence of communities in weighted networks. *Phys. Rev. Lett.* **99**, 228701.

[21] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N. (August 2007) Cost-effective outbreak detection in networks. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 420–429.

[22] Lu, Z., Savas, B., Tang, W. & Dhillon, I. (December 2010) Supervised link prediction using multiple sources. In: *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pp. 923–928.

[23] McNamara, L., Mascolo, C. & Capra, L. (September 2008) Media sharing based on colocation prediction in urban transport. In: *Proceedings of ACM 14th International Conference on Mobile Computing and Networking (Mobicom08)*, San Francisco, CA, pp. 58–69.

[24] Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J.-P. (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**, 876–878.

[25] Muchnik, L., Itzhak, R., Solomon, S. & Louzoun, Y. (2007) Self-emergence of knowledge trees: Extraction of wikipedia hierarchies. *Phy. Rev.* E **76**, 016106.

[26] Newman, M. E. J. (2010) *Networks an Introduction*, Oxford University Press, Oxford.

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

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

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

[30] Tang, J., Scellato, S., Musolesi, M., Mascolo, C. & Latora, V. (2010) Small-world behavior in time-varying graphs. *Phys. Rev.* E **81**, 05510.

[31] Zhao, K., Stehlé, J., Bianconi, G. & Barrat, A. (2011) Social network dynamics of face-to-face interactions. *Phys. Rev.* E **83**, 056109.