IMA Journal of Management Mathematics (2012) **23**, 1−15
doi:10.1093/imaman/dpr001
Advance Access publication on February 14, 2011

Models for evolving networks: with applications in telecommunication and online activities

PETER GRINDROD
Department of Mathematics and Statistics, Centre for Advanced Computing and Emerging Technologies, University of Reading, RG6 6AX, UK

AND

DESMOND J. HIGHAM*
Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK
*Corresponding author: d.j.higham@strath.ac.uk

[Received on 1 April 2010; accepted on 5 January 2011]

Driven by a range of modern applications that includes telecommunications, e-business and online social interaction, recent ideas in complex networks can be extended to the case of time-varying connectivity. Here, we propose a general framework for modelling and simulating such dynamic networks, and we explain how the long-time behaviour may reveal important information about the mechanisms underlying the evolution.

Keywords: birth and death process; graph theory; random graph; range-dependent edge.

1. Introduction

Many application areas give rise to connectivity patterns that change over time. As well as traditional contexts such as epidemiology (Kao *et al.*, 2007; Vernon & Keeling, 2009), examples are arising in modern applications such as mobile telecommunications, online trading, smart metering and online social networking (Borgnat *et al.*, 2008; Gautreau *et al.*, 2009; Geard & Bullock, 2008; Grindrod & Higham, 2010; Grindrod & Parsons, 2010; Kao *et al.*, 2007; Kleinberg, 2008; Tang *et al.*, 2010a,b). Information such as ‘who called who’, ‘who tweeted who’, ‘who facebooked who’ and ‘people who bought his book also bought. . . ’ is naturally evolving over time and cannot be fully exploited through a static representation as a single time average or snapshot. The motivation for our work is that these emerging, data-rich disciplines can generate large, highly resolved network sequences that demand new models and computational tools.

Although we will draw on concepts from the well-studied ‘network growth’ context, where new vertices and accompanying edges are accumulated (Barabási & Albert, 1999; Newman, 2003), we are concerned here with a different time-dependent scenario where the population of vertices remains fixed from the outset, and the graph evolves through the appearance (birth) or the deletion (death) of edges. Specific examples that have recently received attention include

- networks of mobile phone users with a link denoting current interaction (Tang *et al.*, 2010b),
- transportation networks defined over a dynamic infrastructure (Gautreau *et al.*, 2009),

- networks describing transient social interactions (Tang *et al.*, 2010a),
- correlated neural activity in response to a functional task (Grindrod & Higham, 2010).

Our aims here are (a) to set out and study some general options for describing such evolving networks in a stochastic setting and (b) to discuss practical challenges in interpreting and calibrating suitable models. In particular, we show how ideas from Grindrod & Higham (2010) can be extended to produce a wider class of models. We note that related models are also the subject of analysis in theoretical computer science (Avin *et al.*, 2008; Clementi *et al.*, 2008, 2009).

In Section 2, we introduce a range of models under successively less restrictive assumptions. Section 3 focusses on a particularly promising model class and in Section 4, we make some observations about long-time behaviour. In Section 5, we give some illustrative simulations on synthetic data and then show computational results on a real evolving mobile phone network. Concluding remarks appear in Section 6.

2. Stochastic models

For simplicity, we consider here undirected, unweighted graphs, defined on a set V of exactly $n > 2$ vertices, with no self-loops. Extensions to directed graphs and self-loops follow naturally. Any such graph G may be represented by a symmetric $(n \times n)$ adjacency matrix, A, with elements $A_{ij} = A_{ji} = 1$ if the edge $e = (i, j)$ is present and zero otherwise.

Let S_n denote the set of all such graphs defined over these n vertices. We have $|S_n| = 2^{\frac{n(n-1)}{2}} = M(n)$, say. An evolving graph over discrete time steps is simply an ordered sequence, $\{G_k\}$ for $k = 0, 1, 2, \ldots,$ within S_n. We think of the evolving graph as taking the particular state G_k at the kth time step, i.e. at time t_k from some monotonically increasing time sequence.

To introduce a stochastic element, suppose we have a set of conditional probabilities, defined for all possible networks, $G_{k+1} \in S_n$, given all the networks earlier within the evolving sequence: say

$$P(G_{k+1}|G_k, G_{k-1}, \ldots).$$

This set determines a probability distribution for the next element, G_{k+1}, in the sequence, given its history to date. It may be applied iteratively to generate successive further elements of the sequence.

Let us also suppose that the evolving graph respects the natural Markov property, so that the most recent graph in the sequence is sufficient to define the statistics of its future evolution, then we must have well-defined probabilities

$$P(G_{k+1}|G_k), \tag{1}$$

for all ordered pairs (G_{k+1}, G_k) in $S_n \times S_n$. Hence, a total of $2^{n(n-1)}$ probabilities is required in this general setting. This number seems prohibitively large if we wish to study general properties or calibrate models to data.

We may reduce this complexity by imposing symmetry. If there is nothing distinguished about any of the vertices that effects the evolution, then we may argue that the probabilities should not change if the vertices are permuted. (Note this is not the case for range dependent graphs introduced in Section 3.3, where the vertices are embedded within some underlying ordering). In this *permutable* case, the specification is reduced by a factor of $n!$. For n large, Stirling's approximation gives $n! = \mathrm{O}(2^{n \log_2 n})$, so this cannot provide much relief!

A more useful simplification is to assume *edge independence*. Here, at each time step, the probability for the appearance or disappearance of each edge, e, is independent of those for all other edges, yet all such probabilities are conditional on properties of G_k. So for each edge e and each G_k, we must specify the probability

$$P(e \in G_{k+1} | G_k). \tag{2}$$

Then we can reassemble the full graph transition probabilities in (1) from those for the independent edge transitions in (2), to give

$$P(G_{k+1}|G_k) = \prod_{e \in G_{k+1}} P(e \in G_{k+1}|G_k) \prod_{e \notin G_{k+1}} (1 - P(e \in G_{k+1}|G_k)). \tag{3}$$

This radically reduces the need to specify graph to graph transition probabilities since at worst we now require $n(n-1)2^{-1+n(n-1)/2}$ probabilities. In the next section, we consider how these may be defined as suitable functions of e and G_k.

We end this section by noting that a different set of simplifying assumptions were studied in Grindrod & Parsons (2010). In that case, the Markovian property was relaxed and edge dynamics were allowed to depend upon previous history. It was shown that in this scenario, it is possible for edges to become immortal persisting for all time or to become extinct. However, in order to make the analysis tractable, the authors in Grindrod & Parsons (2010) assumed that the behaviour of each edge depended only upon its own history, ruling out the types of interaction that we introduce in the next section.

3. Independent edge birth and death dynamics

We will introduce a class of independent edge models by specifying edge birth and death dynamics. Let us assume that we have well-defined terms

$$\alpha(e) = P(e \in G_{k+1} | e \notin G_k), \quad \omega(e) = P(e \notin G_{k+1} | e \in G_k) \tag{4}$$

that denote the edge birth probability and edge death probability, respectively. For simplicity, this notation suppresses their possible dependence on G_k. Then to define a model, we simply give $\alpha(e)$ and $\omega(e)$ as functions of properties of G_k, when e is not present, or present, in G_k, respectively. We will propose three useful cases.

3.1 *Births and deaths dependent upon degree*

Suppose the edge e that connects vertices v_i and v_j is not in G_k. Let d_i and d_j denote the degree of vertices v_i and v_j within G_k, respectively. Then let us define

$$\alpha(e) = F_\alpha(d_i, d_j),$$

where F_α is any continuous mapping from pairs of integers onto the interval [0,1]. In the undirected edge case that we consider in this work, symmetry demands $F_\alpha(z_1, z_2) = F_\alpha(z_2, z_1)$ for all non-negative integers z_1 and z_2. For example, F_α might be monotonically increasing in both arguments, meaning that edges are more likely to appear between vertices of higher degree. Such a case is given by

$$F_\alpha(d_i, d_j) = \frac{d_i d_j + a}{d_i d_j + a + b}$$

for positive reals a and b. This mirrors the concepts of *preferential attachment* and *assortativity* found in static models (Barabási & Albert, 1999; Newman, 2002).

Similarly, suppose e is in G_k and connects vertices v_i and v_j. Then we may define

$$\omega(e) = F_\omega(d_i, d_j),$$

where F_ω is any continuous mapping from pairs of integers onto the interval [0,1] (again such that $F_\omega(z_1, z_2) = F_\omega(z_2, z_1)$). For example, F_ω may be monotonically decreasing in both arguments, meaning that edges are less likely to disappear between vertices of higher degree.

This type of degree-dependent activity where certain individuals act as *hubs* or *authorities*, Kleinberg (1998), or, in Gladwell's terminology *connectors*, *mavens* and *salespeople*, Gladwell (2000), may be appropriate for modelling social and business networking communities, such as Linkedin, and for online marketplaces, such as ebay, where 'powersellers' may attract more business.

3.2 *Births and deaths dependent upon local clustering*

Localized clustering, the ability for connections to be transitive, is a basic ingredient of small world networks (Watts & Strogatz, 1998). In applications involving networks of social ties, it may be natural to assume that edges evolve so as to triangulate second neighbours and strengthen clique-formation (Adamic *et al.*, 2003).

Suppose some possible edge e connects vertices v_i and v_j but is not in G_k. Let r_{ij} denote the number of adjacent vertices that v_i and v_j have in common. Then r_{ij} is the i, jth element of A_k^2 (the square of the adjacency matrix for G_k). Then let us define

$$\alpha(e) = F_\alpha(d_i, d_j, r_{ij}),$$

where F_α is any continuous mapping from triples of integers onto the interval [0,1]. Note that $r_{ij} \leqslant \min\{d_i, d_j\}$. As before we require $F_\alpha(z_1, z_2, z_3) = F_\alpha(z_2, z_1, z_3)$ for all non-negative integers z_1 and z_2. For example, F_α may be monotonically decreasing in both $z_1 - z_3$ and $z_2 - z_3$, but increasing in z_3, so that the edges are likely to appear between vertices that have many adjacent vertices in common. Such a case is given by

$$F_\alpha(d_i, d_j, r_{ij}) = \frac{1 + r_{ij}}{\sqrt{1 + d_i d_j}}.$$

Similarly, suppose e is in G_k and connects vertices v_i and v_j. Then we may define

$$\omega(e) = F_\omega(d_i, d_j, r_{ij}),$$

where F_ω is any continuous mapping from triples of integers onto the interval [0,1], with $F_\omega(z_1, z_2, z_3) = F_\omega(z_2, z_1, z_3)$. For example, F_ω may be monotonically decreasing in z_3 meaning that edges are less likely to disappear between vertices of with many common adjacencies. Such a case is given by

$$F_\omega(d_i, d_j, r_{ij}) = \frac{1 + \sqrt{d_i d_j}}{1 + r_{ij}}.$$

3.3 *Births and deaths dependent upon edge range*

In some circumstances, it is reasonable to assume that connections between vertices are determined in part by their relative locations in some physical or abstract space (Kleinberg, 2000; Pržulj *et al.*, 2004). This concept of location in space may go beyond the geographical; there is evidence for a more general 'social distance' metric that, in principle, could be inferred from the network data (Watts *et al.*, 2002). More specifically, 'closeness; could be measured by combining a heterogeneous range of known attributes (Cox & Cox, 1994), e.g. via personal details supplied when subscribing to an online service. Furthermore, in some types of online activity, individuals are termed *influencers* if they are able to persuade their friends, family and colleagues to copy their choices (Technology Quarterly, 2010). This would correspond to occupying a central position in the relevant space. This differs from the type of hub concept mentioned in Section 3.1 in that the level of influence does not depend on the current degree of the node and is fixed for all time.

In the secific formalization of range dependent graphs (Grindrod, 2002, 2003; Grindrod *et al.*, 2009; Higham, 2003, 2005), the vertices are considered to have an underlying (generally unknown) ordering on the integer lattice, and the *range* of any possible edge e connecting vertices v_i and v_j, is defined by the distance $m(e) = |i - j|$. To generate an instance of the graph, each edge is then created independently, with a probability given by some predetermined function of its range.

These ideas were extended in Grindrod & Higham (2010) to the dynamic network setting. In that case, we define a range-dependent birth rate and death rate

$$\alpha(e) = F_\alpha(m(e)),$$

$$\omega(e) = F_\omega(m(e)),$$

where F_α and F_ω are continuous mappings from the integers onto the interval [0,1].

Typically, we may choose F_α to be monotonically decreasing so that longer range edges are less likely to arise. The case where

$$\frac{F_\alpha(m)}{F_\omega(m)} = \theta^{m^2},$$

for constant $0 \leqslant \theta \leqslant 1$ was studied in Grindrod & Higham (2010) and shown to be attractive from a model calibration perspective.

This case of range-dependent edge evolution is special because not only are the births or deaths each for edge independent from each other but also each edge depends only on its own immediate history (and of course its own range). In the newly proposed degree-dependent and cluster-dependent models of Sections 3.1 and 3.2, the births and deaths depend on G_k in a more sophisticated way, and the evolution of each edge cannot be determined without determining the evolution of some or all other possible edges. For the range-dependent case though, whether or not e is in G_{k+1} depends only on whether it is in G_k, and the corresponding forms assumed for $\alpha(e)$ and $\omega(e)$. As pointed out in Grindrod & Higham (2010), if $p_k(e) = P(e \in G_k)$, then

$$p_{k+1}(e) = \alpha(e)(1 - p_k(e)) + (1 - \omega(e))p_k(e).$$

So as k becomes large, such an evolving network burns out its initial starting point and we have

$$p_k(e) \to p_\infty(e) = \frac{\alpha(e)}{\alpha(e) + \omega(e)}, \quad \text{as } k \to \infty.$$

4. Long-term behaviour

Suppose we have any Markov model over S_n, where we are given or can calculate $P(G_{k+1}|G_k)$ for all pairs (G_{k+1}, G_k) in $S_n \times S_n$, as in (2) above. Let $s_m \in S_n$ denote the mth element of S_n, for

$m = 1, \ldots, M(n)$. At the kth time step, let $\mathbf{w}_k = (w_{k,1}, \ldots, w_{k,M(n)})^\top$, where $w_{k,m} = P(G_k = s_m)$. Then the $P(G_{k+1}|G_k)$ together determine a transition matrix, B, such that

$$\mathbf{w}_{k+1} = B\mathbf{w}_k.$$

Under the reasonable assumption of ergodicity, meaning that B is irreducible, this iteration tends to a unique non-negative fixed point $\mathbf{w}^\star > 0$ satisfying $B\mathbf{w}^\star = \mathbf{w}^\star$.

Let $\phi(s)$ be any binary valued feature defined for all $s \in S_n$, so that $\phi = 1$ if the graph has the feature present, and $\phi = 0$ otherwise. Then as k tends to infinity the expected value of $\phi(G_k)$ is given by

$$\langle \phi \rangle = \sum_{m=1}^{M(n)} w_m^\star \phi(s_m).$$

In particular, each possible edge, e, is present with a probability

$$p_\infty(e) = \sum_{e \in s_m} w_m^\star.$$

Let $G^\star$ denote the *long-term expected random graph*, where each edge e has a probability $p_\infty(e)$ of being present.

Suppose this model does not use some additional (imposed) knowledge that differentiates between the vertices. Of course range-dependent evolving graphs employ an imposed ordering of the vertices for example; while Barabási style aggregative graphs allow vertices to become active in some predefined order, externally imposed. But for evolving graphs having no such vertex discrimination, symmetry demands that $G^\star$ is invariant to any permutations of the vertices. Hence, all possible edges in $G^\star$ are equally likely: $G^\star$ is an Erdös–Rényi random graph, which therefore has a Poisson distribution of vertex degrees, and no scale free or small world properties (Newman, 2010). For example, this must be true of the evolving networks depending solely upon (current) vertex degrees or localized clustering coefficients, introduced in the Sections 3.1 and 3.2.

So, if we observe a large evolving graph with long-term average behaviour which has a non-Poissonian vertex degree distribution, then we know that the dynamics of any assumed underlying Markov model, of the type introduced here, must involve some extra knowledge or imposed information distinguishing the vertices. The Markovian assumption could itself be tested by computing correlations over time.

5. Numerical simulations

In Fig. 1, we show the first sixteen adjacency matrices that arise from a degree-based model, as described in Section 3.1. This represents a particular path of the stochastic process, where the network at step $k+1$ was computed by calculating the edge birth and death rates for the network at step k and using a pseudo-random number generator to insert and delete edges accordingly. The precise details are as follows. For a pair of vertices i and j, we used an edge birth probability given by

- $p_a = 0.9$ if $\min(\deg_i, \deg_j) > 8$,
- $p_b = 0.2$ otherwise,

and an edge death probability given by

- $p_c = 0.9$ if $\min(\deg_i, \deg_j) < 6$,
- $p_d = 0.2$ otherwise.

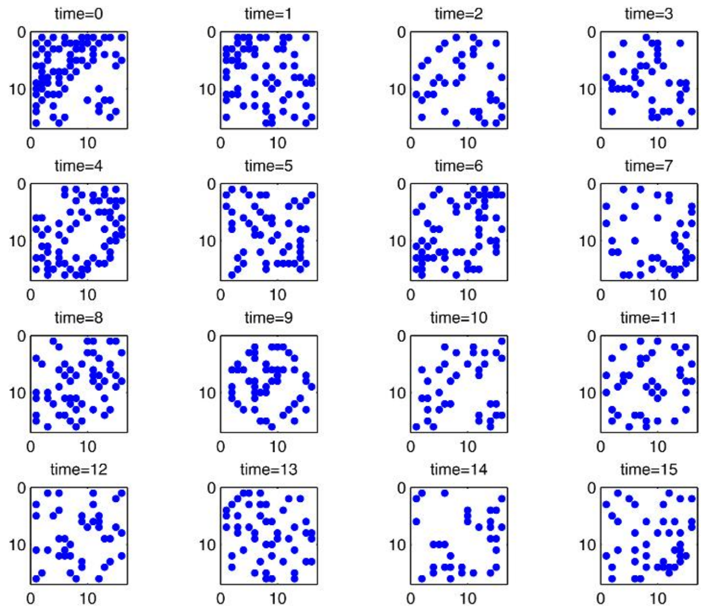

FIG. 1. Initial network adjacency matrix and first fifteen iterates for a degree-dependent evolution. Vertices are ordered according to degree at time zero.

Hence, in this model, new edges are favoured between vertices that both have relatively high degree, and existing edges involving at least one low-degree node are penalized. In order to make the adjacency matrices compact for visualization purposes, we used a relatively small number of vertices, $n = 16$. We begin, at time $t = 0$, with a sample of an Erdös–Rényi random graph with 38 edges. The vertices are ordered according to their degree in the initial network, from high to low.

After continuing the iteration for 200 more time steps, Fig. 2 shows time levels $t = 216$ to $t = 231$.

To illustrate long-time behaviour on a larger network, Fig. 3 shows the results for a case with $n = 200$ vertices. Here, the initial network, shown in the upper left picture, is chosen from the preferential attachment model (Barabási & Albert, 1999; Newman, 2002), as implemented in the function pref.m of CONTEST (Taylor & Higham, 2009). We used the same degree-based birth and death rate construct as for Figs 1 and 2, with edge birth probabilities rescaled to

- $p_a = 0.3$ if $\min(\deg_i, \deg_j) > 25$,
- $p_b = 0.05$ otherwise,

and an edge death probabilities given by

- $p_c = 0.9$ if $\min(\deg_i, \deg_j) < 10$,
- $p_d = 0.6$ otherwise.

The initial degree distribution, which by construction is scale-free, is shown in the upper right picture.

The lower left picture then shows the adjacency matrix after 10^4 time steps, with degree distribution in the lower right. We see that the scale-free pattern of the initial degree distribution is completely lost over time and a Poisson-type distribution has arisen, as predicted in Section 4. For confirmation,

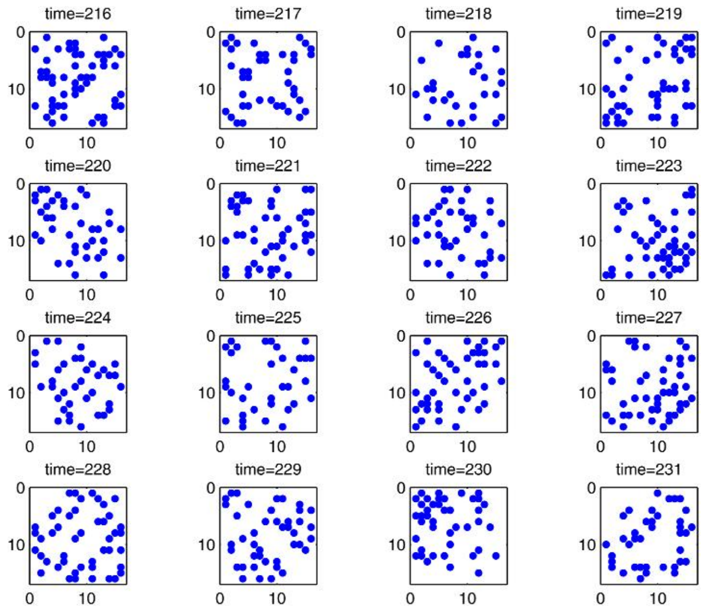

FIG. 2. Iterates at time $t = 216$ to $t = 231$ for the network sequence initiated in Fig. 1.

Fig. 4 plots the cumulative degree distribution as circles and superimposes as a solid line the cumulative Poisson distribution with the same mean of 15.3. To further quantify the goodness of fit to a Poisson distribution, Fig. 5 gives quantile–quantile plot (using qqplot from the MATLAB statistics toolbox, http://www.mathworks.com/help/toolbox/stats/5. Here, the quality of fit can be judged via closeness to the reference line (Gentle, 2009).

Figure 6 shows a different scenario where the edge evolution involves an external factor. Here, we mix together the degree-dependent and range-dependent ideas from Sections 3.1 and 3.3. For the same size $n = 200$ and the same initial network as in Fig. 3, we now use edge birth rates given by

- $p_a = \exp(-|i - j|/10)$ if $\deg_i + \deg_j > 50$,
- $p_b = 0.05$ otherwise,

and edge death rates given by

- $p_c = 0.9$ if $\max(\deg_i, \deg_j) < 30$,
- $p_a = \exp(-|i - j|/10)$ otherwise.

In this case, pairs of vertices are likely to grow a new edge if at least one of them has high degree and they are in close proximity. An existing edge connecting vertices that have degree below 30 is likely to be removed. We see that the long-time network at time 10^4 has a small number of vertices that are abundantly connected to near neighbours. The resulting degree distribution looks far from Poisson, as verified in the cumulative distribution plot of Fig. 7 and the quantile–quantile plot of Fig. 8.

As a final computational test, we consider an evolving network from Eagle *et al.* (2009). This data comes from a 'Reality Mining' study that used mobile phones to follow 106 subjects at Madras Institute of Technology over the course of the 2004–2005 academic year. Pairwise calls, SMS activity

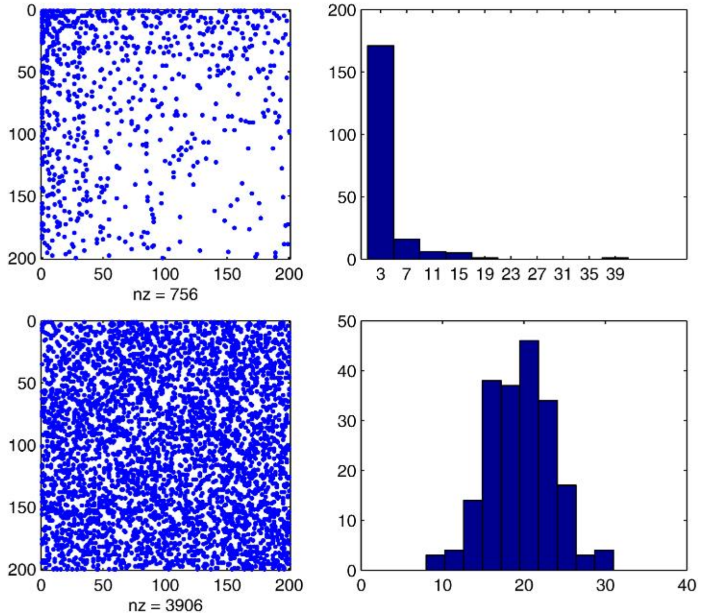

FIG. 3. An evolving network with degree-based evolution. Top left: initial network. Top right: initial degree distribution. Bottom left: network at time $t = 10^4$. Bottom right: degree distribution at time $t = 10^4$.

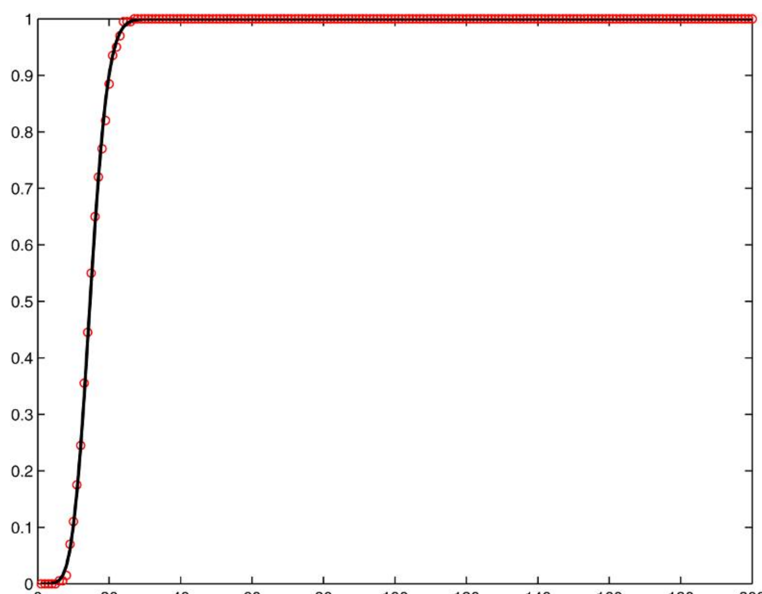

FIG. 4. Circles: cumulative degree distribution for time $t = 10^4$ network shown in Fig. 3. Solid line: interpolant through cumulative degree distribution for a Poisson random variable with matching mean of 15.3.

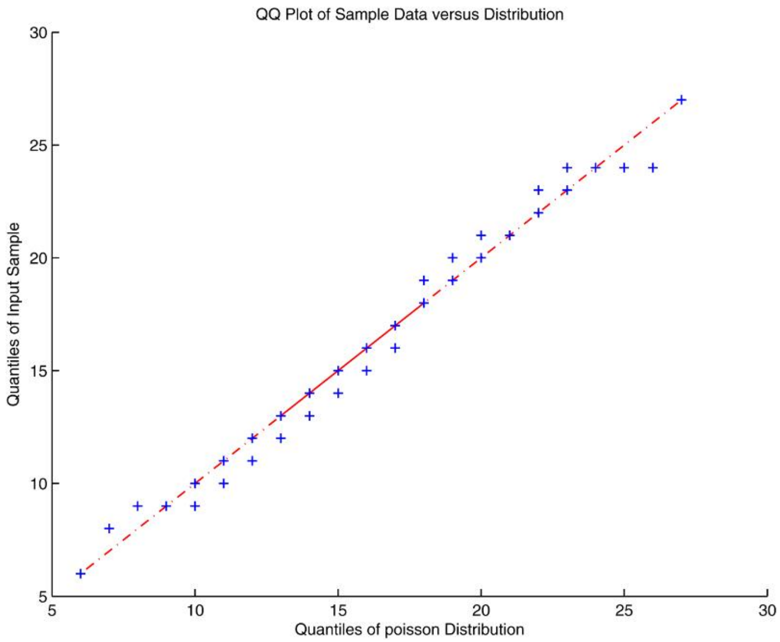

FIG. 5. Quantile–quantile plot for the degree data and Possion distribution from Fig. 4.

and proximity information were recorded. Here, we consider just the voice call component of the data, summarized into weekly activity. So a link between nodes i and j in the kth adjacency matrix indicates that at least one phone call took place between subjects i and j in week k. This represents an evolving network over 52 time points. This network sequence was also used in Grindrod & Parsons (2010), a visualization of the complete data set can be found there. Of course, understanding the mechanisms that drive this type of dynamic network has immediate benefits for designing mobile phone contracts, identifying and marketing to specific customer groups and predicting future patterns of network useage.

We attempted to calibrate this evolving network based on a range-dependent model. Following Grindrod & Higham (2010), we formed a Laplacian matrix based on the cumulative edge data and used the Fiedler vector to infer an ordering of the nodes. More precisely, we formed the symmetric matrix $L - D = W$, where W records the accumulated edge count between each pair of nodes and the degree matrix D is diagonal with $(D)_{ii} = \sum_{j=1}^{n} (W)_{ij}$. By construction, L has a smallest eigenvalue of zero. For the second smallest eigenvalue, we recorded the corresponding eigenvector υ. Then node i is placed before node j if $\upsilon_i < \upsilon_j$. It is shown in Grindrod & Higham (2010) that this ordering solves a relaxed version of the problem of finding the maximum likelihood reordering under an appropriate range-dependent model.

For each pair of nodes i and j, we estimated the birth and death probabilities (4) based on their observed frequency. However, in this context, it is not appropriate to use a simple frequency count for the edge death, s/N, where

- s is the number of edge deaths observed, and
- N is the total number of time points between the first and penultimate time on which an edge existed.

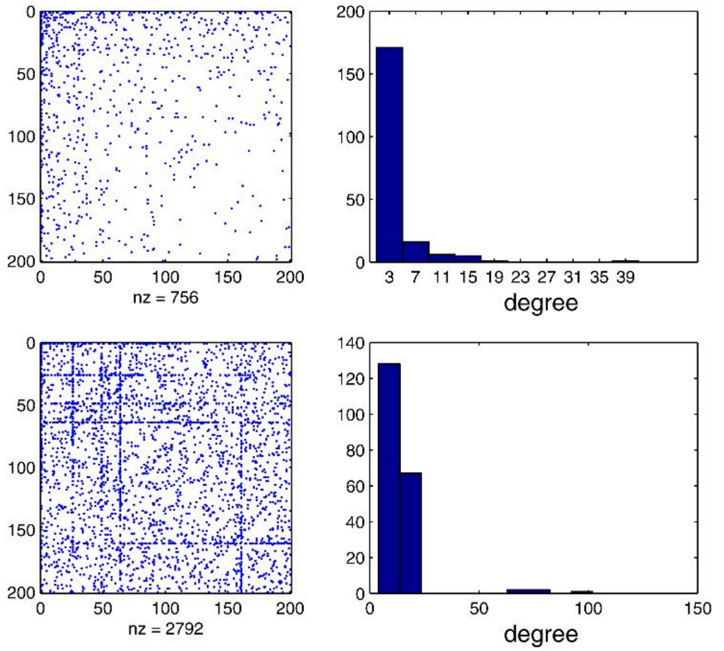

FIG. 6. An evolving network with evolution based on degree and edge range. Top left: initial network. Top right: initial degree distribution. Bottom left: network at time $t = 10^4$. Bottom right: degree distribution at time $t = 10^4$.

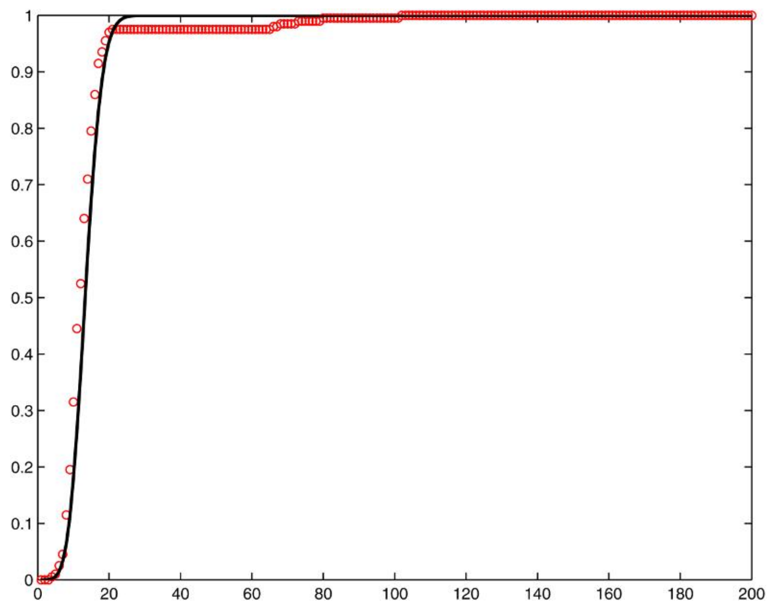

FIG. 7. Circles: cumulative degree distribution for time $t = 10^4$ network shown in Fig. 6. Solid line: interpolant through cumulative degree distribution for a Poisson random variable with matching mean of 14.

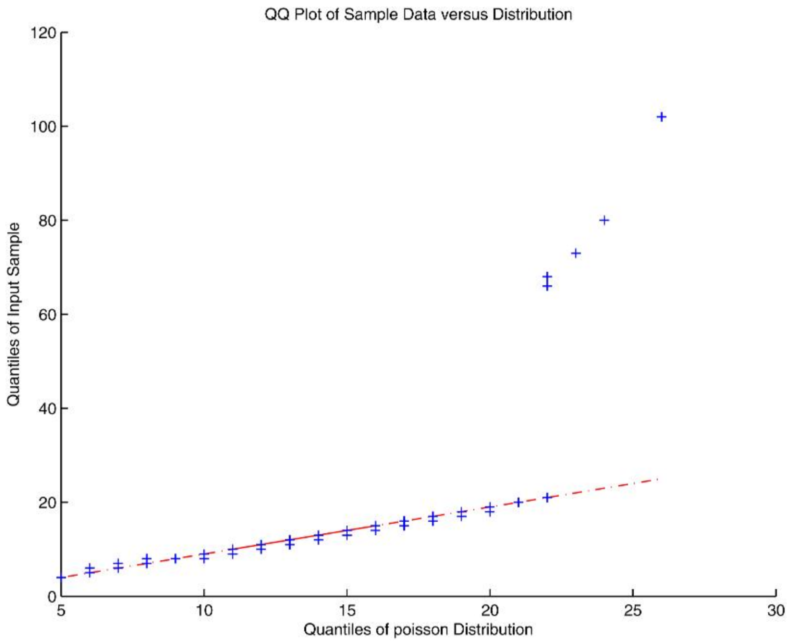

FIG. 8. Quantile–quantile plot for the degree data and Possion distribution from Fig. 7.

This is because $N = 0$ for many pairs i and j. For this reason, we used Laplace's law of succession (MacKay, 2003) to give $(s+1)/(N+2)$ as our estimate for the edge death probability. Similarly, the edge birth probability was estimated as $(s'+1)/(N'+2)$, where

- s' is the number of edge births observed, and
- N' is the total number of time points between the first and penultimate time on which no edge existed.

Figs 9 and 10 indicate the estimated birth and death probabilities, respectively, for the node ordering given by the Fiedler vector. There are two main observations to be made.

1. In both figures, the probabilities are not uniform across pairs i and j. Following the discussion in Section 4, this rules out the possibility that the data comes from the steady state of an evolving network model like those discussed in Sections 3.1 and 3.2, where nodes are not differentiated by some external property.
2. The nodes have been reordered in an attempt to reveal range dependency, so that, in this new ordering, birth and death rates depend solely on $|i - j|$. In Figs 9 and 10, this would correspond to a Toeplitz structure (common values along each superdiagonal). No such obvious pattern is observed, although there is an indication that some nodes are very active, having relatively high edge birth and death probabilities, and in many cases, this activity is localized to neighbours who are close in the new ordering. There is also evidence of clusters of high activity in blocks along the diagonal, where near-neighbours in this newly discovered ordering have strong mutual affinity. This behaviour is in broad agreement with the mixed 'range and degree dependence' model that was used for Fig. 6.

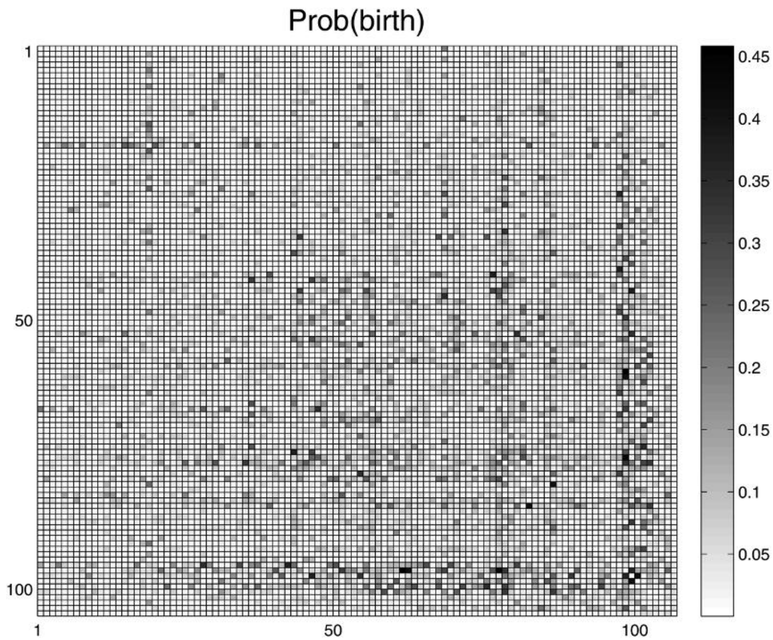

FIG. 9. Symmetric matrix whose i, j element shows the estimated birth probability in (4) for an edge between nodes i and j, assuming range-dependent activity. The underlying data came from an evolving network arising in telecommunication. Nodes are ordered via a Fiedler vector in an attempt to reveal range-dependency, as described in Grindrod & Higham (2010).

6. Discussion

Ideas from network science have proved to be useful in a range of disciplines, but we feel that there is great value to be had from moving attention away from fixed topological structures. Many applications, notably in telecommunications, social networking, online trading and utility consumption, give rise to a sequence of network 'snaphots' from an evolving system. Summarizing and quantifying the mechanisms that drive the network evolution has a clear potential to help with decision-making issue faced by professionals in areas such as online marketing, telecommunications and business development.

New challenges arise in this time-dependent context. Here, we focussed on modelling apsects—what mechanisms might govern the changes in topology? We extended the framework in Grindrod & Higham (2010), showed that broad conclusions may be drawn about steady-state behaviour and tested these ideas on real mobile phone data.

We are currently in a data rich, information poor (DRIP) era where there are huge potential benefits to be had from smarter, more strategic use of evolving network data. Among the key challenges, *calibration* and *model selection* stand out immediately—fitting and comparing models that have a tractable number of parameters in terms of their explanatory and predictive power on real data sets. An accurate, well-tuned model would not only offer a high-level summary of the nature of the interactions but would also give a powerful quantitative tool to predict future evolution and study the response of the network under 'what-if' scenarios. We hope that the framework outlined here sets the scene for a systematic modelling approach.

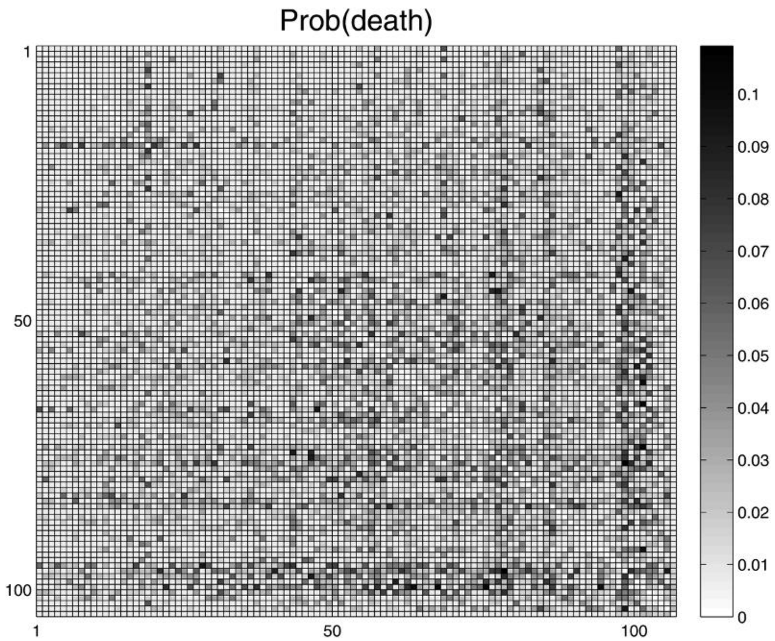

FIG. 10. Analogue of Fig. 9 for edge death probabilities.

Funding

The authors acknowledge support from the Engineering and Physical Sciences Research Council through the Mathematical Sciences programme and Bridging the Gaps initiative, and from the Research Councils United Kingdom Digital Economy programme (EP/F033036/1, EP/1016058/1 and EP/1016031/1).

REFERENCES

ADAMIC, L. A., BUYUKKOKTEN, O. & ADAR, E. (2003) A social network caught in the Web. *First Monday*, **8**.

AVIN, C., KOUCKÝ, M. & LOTKER, Z. (2008) How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). *ICALP '08: Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part I*. Berlin: Springer, pp. 121–132.

BARABÁSI, A.-L. & ALBERT, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.

BORGNAT, P., FLEURY, E., GUILLAUME, J.-L., ROBARDET, C. & SCHERRER, A. (2008) Description and simulation of dynamic mobility networks. *Comput. Netw.*, **52**, 2842–2858.

CLEMENTI, A. E., MACCI, C., MONTI, A., PASQUALE, F. & SILVESTRI, R. (2008) Flooding time in edge-Markovian dynamic graphs. *Proceedings of the 27th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC'08)*. New York: ACM Press, pp. 213–222.

CLEMENTI, A. E., PASQUALE, F. & SILVESTRI, R. (2009) MANETS: high mobility can make up for low transmission power. *Proceedings of the 36th International Colloquium on Automata, Languages and Programming: Part II*, ICALP '09. Berlin: Springer, pp. 387–398.

COX, T. F. & COX, M. A. A. (1994) *Multidimensional Scaling*. London: Chapman and Hall.

EAGLE, N., PENTLAND, A. & LAZER, D. (2009) Inferring social network structure using mobile phone data. *Proc. Natl. Acad. Sci.*, **106**, 15274–15278.

GAUTREAU, A., BARRAT, A. & BARTHELEMY, M. (2009) Microdynamics in stationary complex networks. *Proc. Natl. Acad. Sci.*, **106**, 8847–8852.

GEARD, L. N. & BULLOCK, N. L. (2008) Group formation and social evolution: a computational model. *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*. Cambridge, MA: MIT Press, pp. 197–203.

GENTLE, J. (2009) *Computational Statistics*. Berlin: Springer.

GLADWELL, M. (2000) *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown.

GRINDROD, P. (2002) Range-dependent random graphs and their application to modeling large small-world proteome datasets. *Phys. Rev. E*, **66**, 066702-1–066702-7.

GRINDROD, P. (2003) Modeling proteome networks with range-dependent graphs. *Am. J. Pharmacogenomics*, **3**, 1–4.

GRINDROD, P. & HIGHAM, D. J. (2010) Evolving graphs: dynamical models, inverse problems and propagation. *Proc. R. Soc. A*, **466**, 753–770.

GRINDROD, P., HIGHAM, D. J. & KALNA, G. (2009) Periodic reordering. *IMA J. Numer. Anal.*, **30**, 195–207.

GRINDROD, P. & PARSONS, M. (2010) Social networks: evolving graphs with memory dependent edges. *Technical Report 02*. Department of Mathematics, University of Reading, Reading, UK.

HIGHAM, D. J. (2003) Unravelling small world networks. *J. Comput. Appl. Math.*, **158**, 61–74.

HIGHAM, D. J. (2005) Spectral reordering of a range-dependent weighted random graph. *IMA J. Numer. Anal.*, **25**, 443–457.

KAO, R. R., GREEN, D. M., JOHNSON, J. & KISS, I. Z. (2007) Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J. R. Soc. Interface*, **4**, 907–916.

KLEINBERG, J. (1998) Authoritative sources in a hyper-linked environment. *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*. New York: ACM Press.

KLEINBERG, J. (2000) Navigation in a small world. *Nature*, **406**, 845.

KLEINBERG, J. (2008) The convergence of social and technological networks. *Commun. ACM*, **51**, 66–72.

MACKAY, D. J. C. (2003) *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

NEWMAN, M. E. J. (2002) Assortative mixing in networks. *Phys. Rev. Lett.*, **89**, 208701.

NEWMAN, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

NEWMAN, M. E. J. (2010) *Networks: An Introduction*. Oxford, UK: Oxford University Press.

PRŽULJ, N., CORNEIL, D. G. & JURISICA, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics*, **20**, 3508–3515.

TANG, J., MUSOLESI, M. & MASCOLO, C. & LATORA, V. (2010a) Characterising temporal distance and reachability in mobile and online social networks. *ACM SIGCOMM Comput. Commun. Rev.*, **40**, 118–124.

TANG, J., SCELLATO, S., MUSOLESI, M., MASCOLO, C. & LATORA, V. (2010b) Small world behavior in time-varying graphs. *Phys. Rev. E*, **81**, 055101.

TAYLOR, A. & HIGHAM, D. J. (2009) CONTEST: a controllable test matrix toolbox for MATLAB. *ACM Trans. Math. Softw.*, **35**, 1–17.

TECHNOLOGY QUARTERLY (2010) Untangling the social web. Software: from retailing to counterterrorism, the ability to analyse social connections is proving increasingly useful. *The Economist*.

VERNON, M. C. & KEELING, M. J. (2009) Representing the UK's cattle herd as static and dynamic networks. *Proc. R. Soc. B*, **276**, 469–476.

WATTS, D. J., DODDS, P. S. & NEWMAN, M. E. J. (2002) Identity and search in social networks. *Science*, **296**, 1302–1305.

WATTS, D. J. & STROGATZ, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.