PROCEEDINGS OF THE ROYAL SOCIETY A

Proc. R. Soc. A (2010) **466**, 753–770
doi:10.1098/rspa.2009.0456
Published online 11 November 2009

Evolving graphs: dynamical models, inverse problems and propagation

By Peter Grindrod[1] and Desmond J. Higham[2,*]

[1]*Department of Mathematics and Centre for Advanced Computing and Emerging Technologies, University of Reading, RG6 6AX, UK*
[2]*Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK*

Applications such as neuroscience, telecommunication, online social networking, transport and retail trading give rise to connectivity patterns that change over time. In this work, we address the resulting need for network models and computational algorithms that deal with dynamic links. We introduce a new class of evolving range-dependent random graphs that gives a tractable framework for modelling and simulation. We develop a spectral algorithm for calibrating a set of edge ranges from a sequence of network snapshots and give a proof of principle illustration on some neuroscience data. We also show how the model can be used computationally and analytically to investigate the scenario where an evolutionary process, such as an epidemic, takes place on an evolving network. This allows us to study the cumulative effect of two distinct types of dynamics.

Keywords: birth and death process; epidemiology; network; neuroscience; random graph; reproduction rate

1. Introduction

The last decade has seen a huge rise in interest in complex networks and their applications to mass communication, social and natural phenomena. Until recently, one characteristic of such networks that is fundamental within almost all applications has received only marginal attention: that the networks may evolve (Saramäki & Kaski 2005; Kao *et al.* 2007; Borgnat *et al.* 2008; Gautreau *et al.* 2009; Vernon & Keeling 2009). This issue is distinct from network growth, or aggregative phenomena: it embodies the property that all edges within the network are transient to some extent. Very recently, general classes of dynamic networks have been proposed and studied in the theoretical computer science literature (Avin *et al.* 2008; Clementi *et al.* 2008, 2009) from a complexity theory perspective. Our work looks at complementary issues driven by the need for practical tools in modelling, calibration and data analysis.

In Borgnat *et al.* (2008), it is pointed out that network evolution can be approached from distinct directions: at the macro-level, it may be observed within real datasets by studying the time course of global parameters from one snapshot

*Author for correspondence (aas96106@maths.strath.ac.uk).

Received 27 August 2009
Accepted 13 October 2009

This journal is © 2009 The Royal Society

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

to another, or at the micro-level, dynamic properties could be ascribed to the individual birth–death rules for each edge; and that specific applications will require some analytical methods to approach an *inverse problem*: given some data from a time-dependent evolving network, how best may one represent it within a suitably defined class of models?

In this paper, we consider these problems and offer an operational approach to each.

In §2, we consider micro-to-macro-models that produce evolving networks, considered as classes of Markov processes defined over the set of all possible undirected graphs on a finite set of n vertices. This space grows as $O(2^{n^2})$, so we suggest simplifying to the case of independent dynamics for each edge.

As a further simplification, we may then allow the dynamics to depend only on the *range* of the edge. This extends the static concepts of 'lattices plus shortcuts' (Kleinberg 2000*a*; Newman *et al.* 2000; Higham 2007) and more general range-dependent random graphs (Grindrod 2002; Higham 2005) to the dynamic setting. In many applications, vertices are fixed within some Euclidean or underlying metric space and each edge has a natural 'range' representing the distance between the end vertices, which may impact on the possibility of that edge arising. For example, in traditional acquaintanceship networks physical neighbours have a good chance of knowing each other. These ideas are pervasive in the literature (e.g. the recent treatment in Franceschetti & Meester 2008). But in many applications, there is no obvious, fixed lattice topology. For example:

(i) in communication networks, individuals may be mobile,
(ii) in online social networks, or telecommunications networks, physical position does not determine the cost or likelihood of interaction,
(iii) in cyberspace, hyperlinks are not constrained by the underlying Internet connectivity structure,
(iv) in high-frequency functional connectivity networks from neuroscience, cognitive processing tasks may be distributed across the brain, and
(v) in proteomic interaction networks, connections may be caused by a combination of features, including electrostatic, hydrophobic or chemical similarities rather than any obvious sequence-level or geographical commonalities.

In all these cases, the concept of 'range' is more elusive, not corresponding directly to a simple distance, but can still have some meaning. The range of an edge reflects the transitive nature of the connection. If vertex a is connected to vertex b which is in turn connected to vertex c, how likely is it that vertex a is also connected to vertex c? If it is very likely then we will say that connection is (locally) transitive and the edge from a to c is short range, if it is very unlikely then the edge from a to c is long range. Cliques are full of short-range edges. Hence, when we are presented with data from an evolving graph as just a transient set of connections between an arbitrarily ordered list of vertices, there is potential to add insight by inferring a range for every possible edge. This is the *inverse problem* we will address: that of representing a given evolving graph as a range-dependent evolving graph within an imposed class and thus inferring a range for every possible edge.

Section 2 briefly discusses evolving networks in a general setting. In §3, we review the formal concept of range dependency, introduce the new extension to evolving networks and look at the inverse problem of discovering the ranges.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

The spectral algorithm that we develop is then illustrated in §4 and tested on some real brain activity networks. In §§5 and 6, we turn our attention to propagation problems defined on evolving graphs. Unlike percolation problems for static graphs, or dynamic propagation of infections or information on static graphs, here the dynamics and the sequential behaviour of the evolving graph interferes with the dynamics assumed for propagation of changes to vertex-dependent states. We thus consider the role of evolving graphs in transmission and threshold behaviour as well as achievable expected path lengths and shortest transit times. We believe that this work taps into an exciting and novel field for analysis with many potential applications—from information theory and neuroscience through to viral, or buzz, marketing.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

2. Independent edge-dependent dynamics

Let V_n denote a set of n labelled vertices. Let S_n denote the space of all undirected graphs defined on V_n. Then $|S_n| = 2^{n(n-1)/2}$. Any element of S_n, denoted by a, may be represented by an $n \times n$ symmetric adjacency matrix, which we will denote by A.

We will consider discrete-time Markov processes defined over S_n, whose paths consist of time-dependent sequences of elements, $\{a_j\}$ in S_n, each representing the *evolving graph* at time $t_j = j\,\delta t$. Even when we restrict to the case where the transition matrix is time independent, to fully specify such a process, in general, requires $2^{n(n-1)}$ non-negative graph-to-graph transition probabilities. We therefore continue with a simplified class of models where the time-dependent appearance or disappearance of each individual edge is governed by a random process that is independent of all other edges. More precisely, consider an evolving graph $\{a_j\}$ defined as follows. Let $\alpha(e)$ denote the probability that any edge, e, not part of the network at time t may be added to it over the time step δt. Let $\omega(e)$ denote the probability that any edge, e, that is part of the network at time t will be removed over the time step δt. So α and ω specify the birth and death probabilities, respectively, that we assume to be $O(\delta t)$ for δt small.

For any pair a, $a' \in S_n$, let $P(a'|a)$ denote the probability that $a_{j+1} = a'$ given that $a_j = a$, and let $E(a)$ denote the set of edges belonging to a. Then, the 'independent edge-dependent' model yields the graph-to-graph transition probability

$$P(a'|a) = \prod_{e \in E(a'),\ e \notin E(a)} \alpha(e) \prod_{e \notin E(a'),\ e \notin E(a)} (1 - \alpha(e)) \times \prod_{e \notin E(a'),\ e \in E(a)} \omega(e) \prod_{e \in E(a'),\ e \in E(a)} (1 - \omega(e)). \tag{2.1}$$

This expression gives the probability that exactly the right subsets of edges are added and deleted to achieve the required transition. As a result of the edge independence assumption, there are now $n(n-1)$ parameters ($\alpha(e)$ and $\omega(e)$) rather than the $2^{n(n-1)}$ required in the general case.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

It is straightforward to calibrate such a model, given sufficient data. Suppose that we observe a sequence $\{\hat{a}_j | j = 1, \ldots, J\}$. Then we may estimate the parameters $\alpha(e)$ and $\omega(e)$ independently for each edge, using Laplace's law from the sequence data. Specifically, suppose that an edge e is absent within the first $J-1$ terms of the observed sequence of graphs on exactly $M(e)$ occasions and of these graphs exactly $m(e)$ are followed by graphs that contain e (so e appears in those transitions), then we have the following estimate

$$\widehat{\alpha}(e) = \frac{m(e)+1}{M(e)+2}. \tag{2.2}$$

Similarly, suppose an edge e occurs within the first $J-1$ terms of the observed sequence of graphs on exactly $M^\star(e)$ occasions and of these graphs, exactly $m^\star(e)$ are followed by graphs that do not contain e (so e is lost in the transition). Then

$$\widehat{\omega}(e) = \frac{m^\star(e)+1}{M^\star(e)+2}. \tag{2.3}$$

We may then use these estimates in equation (2.1) to generate any graph-to-graph transition probability that is required. In particular, we could simulate and analyse sequences of networks.

3. Range dependency

(*a*) *Review of the static case*

In Grindrod (2002), the class of *range-dependent random graphs* was introduced as a parameterized model that can reproduce important properties seen in real networks. Protein–protein interaction data were used to motivate and justify the concept. Closely related models based on similar principles include:

(i) the original small world networks of Watts & Strogatz (1998), and their counterparts based on adding shortcuts rather than rewiring (Newman *et al.* 2000), where edges are either long or short range,
(ii) the two-dimensional lattice-based model of Kleinberg (2000*a*,*b*), and
(iii) the geometric model used by Pržulj and co-workers (Pržulj *et al.* 2004; Kuchaiev *et al.* 2009) to describe protein–protein interactions.

Range-dependent random graphs are best introduced by imagining the vertices set out in a line and labelled by their integer positions. To simplify things further, if n, the number of vertices, approaches infinity, then we may approximate the graph by one on infinitely many vertices $(\ldots, -2, -1, 0, 1, 2, \ldots)$, since the edge effects become less important.

Range-dependent random graphs are then defined as follows: an edge is present between any vertices i_1 and i_2 with probability $p_{i_1,i_2} = f(|i_1 - i_2|)$, where f is a given monotonically decreasing function of the edge *range* $|i_1 - i_2|$. Thus, in this model the presence or absence of an edge depends only on its range, and each edge is independent. Letting P_k be the consequent probability that any vertex has degree k, the generating function $G_0(x) = \sum_{k=1}^{\infty} x^k P_k$ can be used to study the Watts–Strogatz clustering coefficient (Watts & Strogatz 1998) and the mean degree (see Grindrod 2002 for details).

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

In the protein–protein interaction case, and in most realistic network scenarios, the vertices will be labelled in a way that does not reflect any edge range information. Therefore, there is a natural inverse problem of reordering the vertices to reveal the range dependency. Genetic algorithms (Grindrod 2002) and more efficient spectral methods (Higham 2003; Grindrod *et al.* submitted) have proved successful in this context.

Our aim is now to develop these static ideas into the evolving graph framework.

(*b*) *Evolving range-dependent random graphs*

Consider a set of n vertices labelled by location $i=1,\ldots,n$. As in §2 we consider a discrete Markov process over S_n where all edges evolve independently. Suppose further that each edge e has transition probabilities that depend only on its range: if an edge connects vertices i_1 and i_2, then we will write $k(e)=|i_1-i_2|$ to denote its range. Then, an evolving range-dependent random graph has birth and death transition probabilities

$$\alpha(e)=f_\alpha(k(e)) \quad \text{and} \quad \omega(e)=f_\omega(k(e)),$$

given by functions $f_\alpha(k)$ and $f_\omega(k)$ that map the positive integers onto $[0,1]$.

Now let $p(e,j)$ denote the probability that the edge e is present within a_j, the graph at time t_j. Then using the transition probabilities above, we have the dynamical equation

$$p(e,j+1)=f_\alpha(k(e))(1-p(e,j))+(1-f_\omega(k(e))p(e,j)).$$

A steady distribution must then satisfy

$$p_0(k(e))=\frac{\alpha(e)}{\alpha(e)+\omega(e)}=\frac{f_\alpha(k(e))}{f_\alpha(k(e))+f_\omega(k(e))}, \tag{3.1}$$

which depends only upon $k(e)$. Hence, at equilibrium any single observation of the evolving graph appears as a range-dependent random graph with each edge present according to this range-dependent probability function $p_0(k)$.

Now consider the natural inverse problem. Given an observed sequence, $\{\hat{a}_j|j=1,\ldots,J\}$, with the vertices in some given ordering $i=1,\ldots,n$, how can we best represent that evolving graph within the class of evolving range-dependent random graphs? It is clearly reasonable to reorder the vertices with a mapping $q(i)$ so as to maximize the likelihood of the actual observations.

The simplest way forward is to consider each edge in turn within the evolving sequence. Suppose that $e=(i_1,i_2)$, in the original ordering, is observed on exactly r_{i_1,i_2} occasions (and is absent on $J-r_{i_1,i_2}$ occasions). Let R denote the symmetric non-negative matrix with elements r_{i_1,i_2}. Under a reordering $q(i)$ the edge range becomes $k(e)=|q(i_1)-q(i_2)|$ and the likelihood of the observations for this edge is given by

$$p_0(|q(i_1)-q(i_2)|)^{r_{i_1,i_2}}\cdot(1-p_0(|q(i_1)-q(i_2)|))^{J-r_{i_1,i_2}}.$$

Trivially, we can rewrite this as

$$\frac{p_0(|q(i_1)-q(i_2)|)^{r_{i_1,i_2}}}{(1-p_0(|q(i_1)-q(i_2)|))^{r_{i_1,i_2}}}(1-p_0(|q(i_1)-q(i_2)|))^{J}.$$

Since all edges are independent, we may write the likelihood over the entire graph by taking a product over all possible edges to give

$$\prod_{e=(i_1,i_2)} \frac{p_0(|q(i_1)-q(i_2)|)^{r_{i_1,i_2}}}{(1-p_0(|q(i_1)-q(i_2)|))^{r_{i_1,i_2}}} \prod_{e=(i_1,i_2)} (1-p_0(|q(i_1)-q(i_2)|))^J.$$

But the second product is independent of q since all possible edges appear and are raised to the same power. Thus, the likelihood of these observations, given any q, has the proportionality

$$\mathcal{L}(q) \propto \widehat{\mathcal{L}}(q) := \prod_{e=(i_1,i_2)} \left(\frac{p_0(|q(i_1)-q(i_2)|)}{(1-p_0(|q(i_1)-q(i_2)|))} \right)^{r_{i_1,i_2}}.$$

So we should choose a reordering q to maximize $\mathcal{L}(q)$. From equation (3.1), we thus have

$$\widehat{\mathcal{L}}(q) = \prod_{e=(i_1,i_2)} \left(\frac{f_\alpha(|q(i_1)-q(i_2)|)}{f_\omega(|q(i_1)-q(i_2)|)} \right)^{r_{i_1,i_2}}. \tag{3.2}$$

Maximizing $\widehat{\mathcal{L}}$ in equation (3.2) over all reorderings q is a hard combinatoric optimization problem. We can make progress through two types of simplification.

First, we assume that the ratio of birth and death transition probabilities has the functional form

$$\frac{f_\alpha(k)}{f_\omega(k)} \propto \theta^{k^2}, \tag{3.3}$$

for some constant θ. Then taking logarithms in equation (3.2), we have

$$\log \widehat{\mathcal{L}} \propto \log(\theta) \sum_{i_1 > i_2} r_{i_1,i_2} (q(i_1)-q(i_2))^2. \tag{3.4}$$

The second step is to relax the problem so that q is allowed to be a real-valued vector. The right-hand side in equation (3.4) may be written as the quadratic form $q^{\mathrm{T}} \Delta_R q$. Here Δ_R, the Laplacian matrix associated with R, has the form D–R, where the diagonal matrix D contains the row/column sums of R. To remove shifting and scaling redundancies we also impose the constraints $\|q\|_2 = 1$ and $\sum_{i=1}^n q(i) = 0$.

We now have a tractable optimization problem. For applications where the death rate exceeds the birth rate at long range, so $\theta < 1$, it is solved by a *Fielder vector*—an eigenvector corresponding to the smallest non-zero eigenvalue of Δ_R. A reordering of the vertices can be recovered by sorting the components of q; that is, vertex i is placed before vertex j if $q_i < q_j$. This approach has been found to be effective for static networks (Van Driessche & Roose 1995; George & Pothen 1997; Ding *et al.* 2001; Higham 2003, 2005; Strang 2008; Grindrod *et al.* submitted); here we are showing that in the evolving case it is possible to justify from first principles the idea of reordering on the cumulative (non-binary) matrix R.

We point out that it is not necessary to know the actual value of θ in equation (3.3). The derivation assumes only that this functional form exists and θ is not required by the algorithm. We have found in practice that performance is not sensitive to the precise form of range dependency (Higham 2003), especially in the long range, or large k, regime. Furthermore, this approach of spectral reordering based on R can be used for any dataset, and the validity of equation (3.3) may then be tested *a posteriori*.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

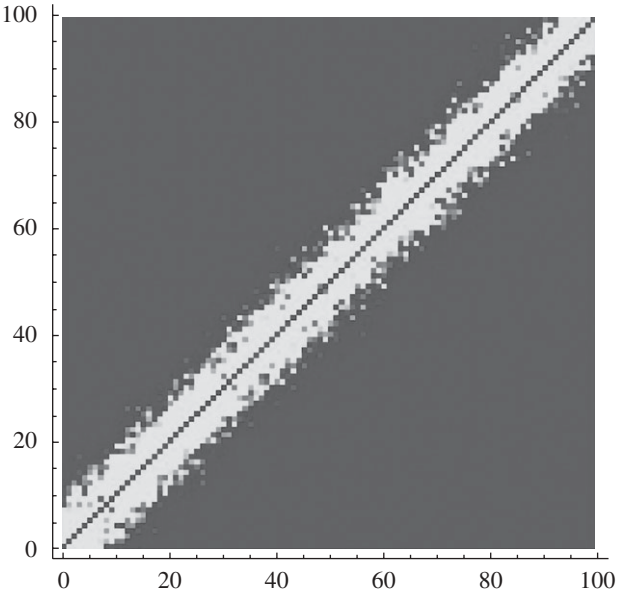

Figure 1. The reordered sum (superposition) of the evolving graph after maximizing the relaxed likelihood. Elements close to the diagonal, $i \approx j$, represent edges that are inferred to be at short range.

We also note that the reordering approach continues to make sense when $\theta > 1$ in equation (3.3). This includes the case where long ranges are very unlikely to emerge, but those that do are long lived. In this case, because $\log(\theta) > 0$, the expression in equation (3.4) is maximized by an eigenvector of the Laplacian that corresponds to a dominant eigenvalue.

4. Computational results for reordering an evolving graph

To test the reordering approach, we generated some synthetic data from the appropriate underlying model. With $n = 100$ vertices, we chose a birth rate $f_\alpha(k) = 0.1(0.98)^{k^2}$ and death rate $f_\omega(k) = 0.2$. After ordering the nodes arbitrarily and letting the network evolve for 200 time steps, we applied the reordering algorithm. In this new ordering, figure 1 shows the binarized sum of adjacency matrices over all 200 time steps; a light dot denotes that an edge was present for at least one time step. Figure 2 shows a typical member of the sequence in this new ordering. We see from figures 1 and 2 that the hidden range dependency has been revealed by the algorithm.

Next, we illustrate the algorithm in an electroencephalography application, using data from Sweeney-Reed & Nasuto (in press). Here, the measurements represent electrical activity produced by the firing of neurons within the brain over a short period of time, reflecting correlated synaptic activity caused by post-synaptic potentials of nearby cortical neurons.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

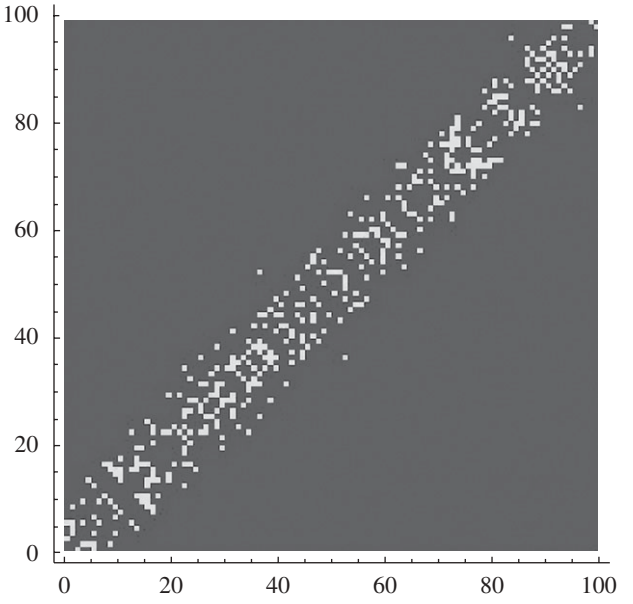

Figure 2. A typical reordered element of the evolving network. This node ordering places non-zeros close to the diagonal, revealing the propensity for short-range edges.

In these experiments, the subject carried out a specific task—tapping in time to music. We use 4 s worth of data sampled at 500 Hz, with 2 s prior to tap and 2 s after. Hence, the finger tap starts around 1000 samples into the data. Measurements were taken at each of 128 electrodes arranged at fixed points on the scalp. Figure 3 shows the electrode locations (see Sweeney-Reed & Nasuto in press for further details).

Results from nearby electrodes may possibly be correlated due to the volume integrative effects of the skin and scalp. On the other hand, transient correlations between channels corresponding to electrodes located some distance apart on the scalp may represent dynamic, synchronous, locking as some otherwise separate tasks become briefly coordinated; or the need for specific tasks that rely on a distributed processing effort. The high-resolution data allow us to examine whether such wide-scale transient correlations occur, as some separate cognitive processes become (briefly) coordinated to manage a combination of processing, sensory and motor responses that may be distributed across the cortex.

We therefore let each electrode represent a vertex within an evolving graph. We subdivided the time-dependent data into windows of 50 consecutive time steps, each lasting 0.08 s. Within each time step we first obtained the *all versus all* channel correlation matrix and defined an edge between vertex i and vertex j if this correlation exceeded 0.8. This resulted in an evolving sequence of 50 adjacency matrices.

We note that correlation between signals is a far from perfect measure, especially in the search for noisy, transient, synchronous components within time series, but it will serve for our current purpose of illustrating how naturally

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

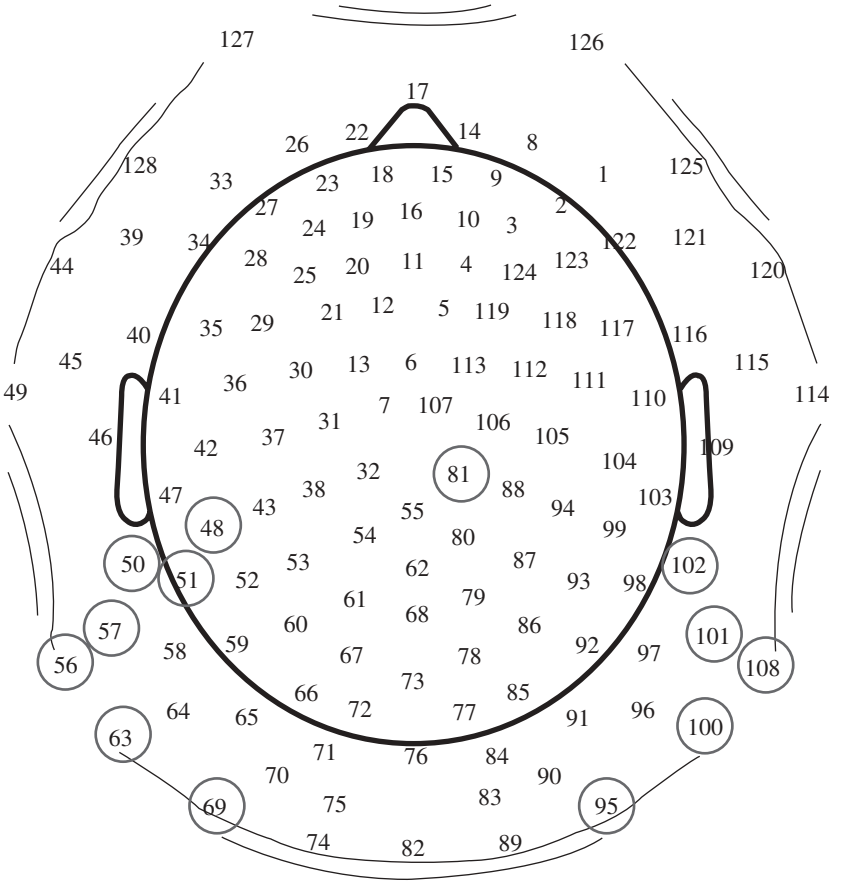

Figure 3. Location on the scalp of the 128 electrodes.

the concept and methods of evolving (organizational) graphs, introduced in this paper, may represent the coordinated emergent, transient, responses (both local and non-local) of the brain.

Figure 4 shows the evolving sequence of 50 networks. Here time increases along each row in the picture—as shown by the indicative labelling of times $1, 2, 3, 4, 5$ and $47, 47, 48, 49, 50$. The effect of the subject's 'task' after 2 s can be seen clearly in time steps $26, 27, 28$. In this figure, vertices are ordered according to the default values provided by the recording equipment, as shown in figure 3, and we note that some vertices with neighbouring indices (i.e. successive row/column indices in the adjacency matrix) correspond to electrodes that are geographically close on the scalp, but in other cases they do not. In particular, there is an artificial periodicity or 'wrap around effect' in this default vertex ordering; the earliest vertices are geographically close to the latest. This effect manifests itself through the significant non-zero blocks in the off-diagonal corners of the adjacency matrices.

Figure 5 repeats the information in figure 4, with vertices reordered via the spectral algorithm. In this case, we see that the activity has been arranged into coherent blocks. At the latter end of this new ordering (lower right corner),

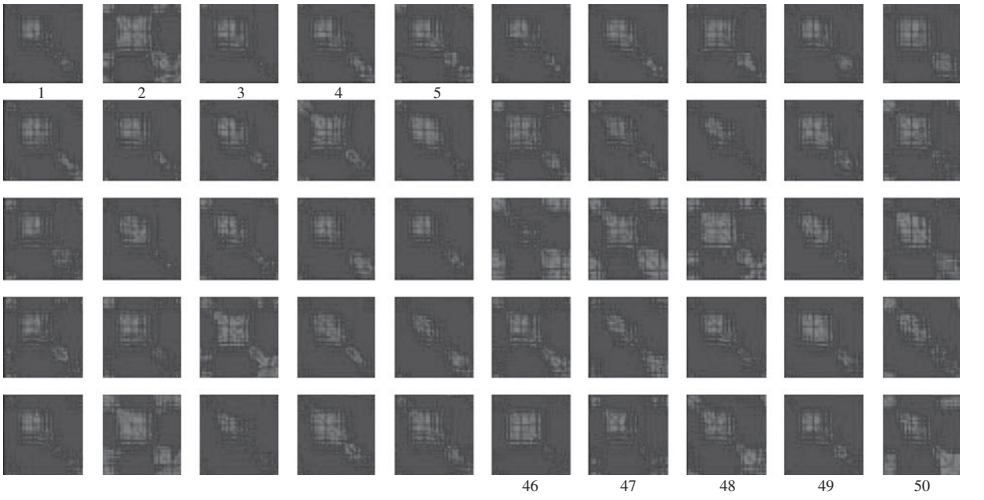

Figure 4. An evolving sequence of 50 adjacency matrices that represent correlations between brain activity at 128 regions. Vertices are ordered using the default provided by the recording equipment.

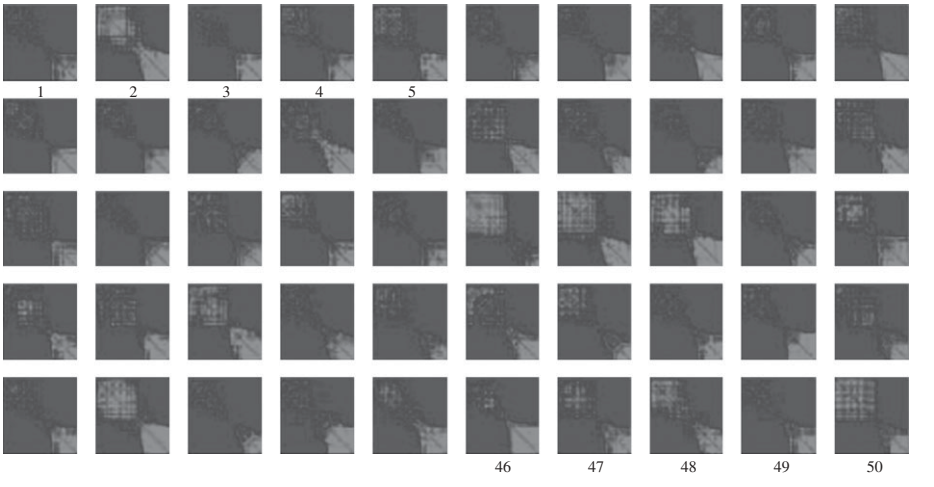

Figure 5. The adjacency matrices from figure 4, reordered according to the range dependency algorithm.

one set of vertices appears to have a consistently strong set of mutual correlations, whereas at the start (upper left corner) a more transient set of correlations is captured. The apparent periodicity from figure 4 has been removed and there is a clear propensity for 'short-range' edges, that is, connections between near neighbours.

In figure 6, we show the binarized cumulative matrix R under the new ordering. We also give the new electrode ordering; so electrodes numbered 55, 22 and 17 in the equipment are placed in positions 1, 2 and 3, respectively, and so on. Of particular interest are electrodes 81, 102, 108, 48, 101, 57, 56, 95, 50, 63, 100 and 51, which appear to provide crosstalk between the two otherwise isolated

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

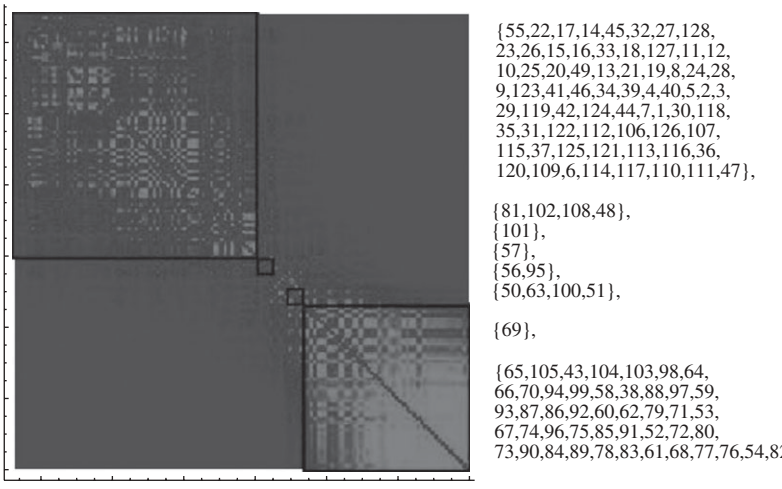

Figure 6. The cumulative edge matrix R, with blocks of highly correlated electrodes indicated.

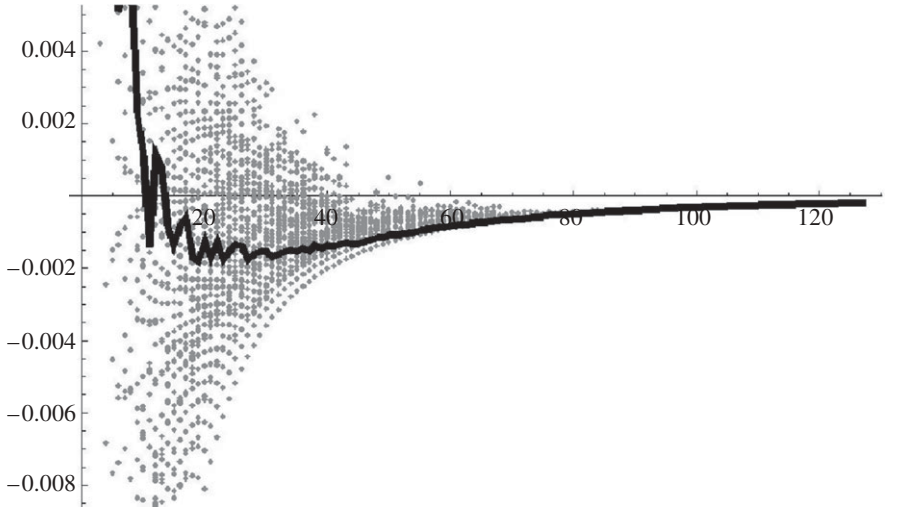

Figure 7. Scatter plot (grey) and average (dark line) of the scaled log birth data ratio $(\log(\widehat{\alpha}(k)/\widehat{\omega}(k))/k^2)$ as a function of predicted range, k.

groups. Looking at their physical locations, which are highlighted in figure 3, we see that these regions cover two physically separate areas close behind both ears, and hence it is plausible that these edges correspond to auditory activity.

As an *a posteriori* check on the relevance of the evolving network model, in figure 7 we scatter plot the values

$$\frac{1}{k^2}\log\left(\frac{\widehat{\alpha}(k)}{\widehat{\omega}(k)}\right),$$

where $\widehat{\alpha}(k)$ and $\widehat{\omega}(k)$ are computed from equations (2.2) and (2.3). If assumption (3.3) is valid, then these values provide estimates for $\log\theta$. For each predicted

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

range, k, the solid line in the figure shows the average of the scattered points, and we see that the results are consistent with the $\theta < 1$ scenario to which the algorithm applies.

Although it would of course be possible to customize this computational technique to account for the specific nature of the problem, we believe that this experiment on real neuroscience data confirms that the basic modelling and algorithmic approach has potential for understanding, calibrating and summarizing evolving networks. In particular, it allows the transient nature of the interactions to be quantified and compared across different settings.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

5. Simulating propagation within an evolving graph

In this section, we present two sets of computational results that arise when we put dynamics on the evolving network model. These results are then explained analytically in §6.

(a) *An epidemic model*

Suppose we have a binary state variable defined at each vertex and at each time step. To be specific, let this variable take values 'infected' and 'susceptible'. Initially all vertices except one are labelled susceptible.

From one time step to the next, we impose the following simple dynamics, depending on a single parameter μ.

(i) A susceptible vertex has no effect on the fate of any other vertex.
(ii) Each infected vertex passes on the infection to all of its current immediate neighbours.
(iii) Having passed on the infection, each infected vertex becomes susceptible with probability $1-\mu$, or else remains infected, with probability μ.

Intuitively, we foresee two possibilities depending on the infection dynamics and the evolving network dynamics: (i) that the evolving graph is sparse so the infection spreads only for large μ (where the infection tends to remain at individual vertices for a long time and these vertices eventually acquire edges), but not for small μ, or (ii) the evolving network is dense enough to support spreading for all values of μ—even when $\mu = 0$ the infection lasts for just one time step at each infected vertex. In the special case where $\mu = 1$ and every α is positive, we are in the *successive percolation* or *flooding* (Clementi *et al.* 2008, 2009) regime where all vertices are certain to become infected.

The novelty in this area lies in the dynamic coupling between the evolution of the contact network and the time course of the infection, in contrast to most of the existing work in this field, which has been carried out with percolation type models or susceptible, infected, recovered (SIR) dynamics on static graphs.

In figure 8, we depict a threshold case, where the elements of the evolving (range-dependent) network are individually very sparse; the expected degree is 0.79. All elements are highly disconnected. In this example, we use the same evolving network in each case with 100 vertices, for which $f_\alpha(k) = 0.1(0.9)^{k^2}$ and $f_\omega(k) = 0.5$. We seed the infection at a single vertex (vertex 40), with a different value of μ for each run. We emphasize that recovery at each infected vertex is decided independently at random in all cases (a Poisson process).

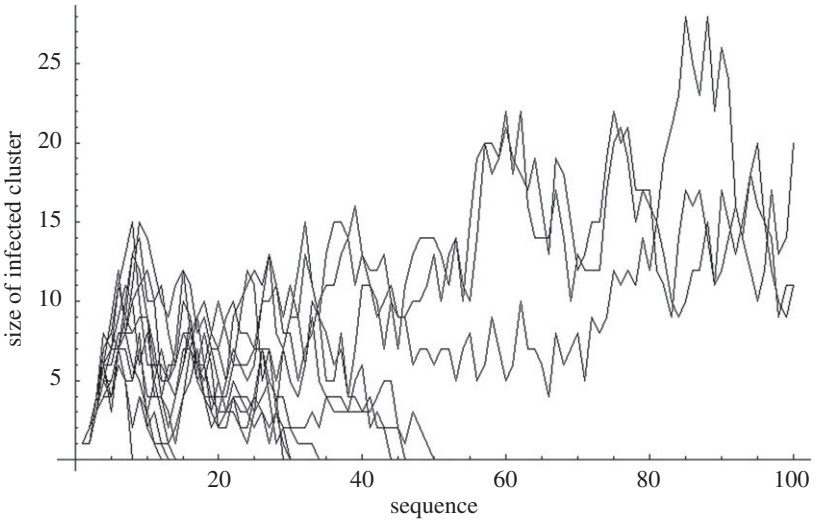

Figure 8. Propagation of infection through the evolving graph as recovery rate lessens.

We employ values for μ between 0.1 (where the infection dies out quickly) and 0.625, where it propagates over the whole evolution. The threshold was observed to be approximately between 0.55 and 0.575. So, we have an example of how occasional long-range transient connections between otherwise short-range, isolated communities may propagate disease providing that the infectious period for individuals (and hence of the isolated communities) is long term enough.

(*b*) *A message passing viewpoint*

Suppose that $\mu = 1$, and that we are transmitting a message rather than an infection. Starting out from a specific vertex, called the sender, at each time step the message is relayed to any vertices directly connected to previous message holders. Hence, the message spreads according to the successive edges added at each time step to the existing message holders. Eventually, all vertices will receive the message (since $\mu = 1$ so a vertex never forgets the message). Suppose also that a vertex is designated as the desired *receiver*. Then, in the range-dependent graph setting, there is a natural range between sender and receiver, which we will denote by k_{mess}. We will say the minimum number of time steps needed for the first arrival of the message at the receiver is the *shortest transmission time* (STT). It is clearly of interest to study the distribution of the STT and related quantities, such as the natural 'path length' measured by taking whichever message-holding vertex is currently closest in range to the receiver. This path length is important if, for example, the message propagation from one vertex to another along any single edge during any single time step is noisy, so that a 'Chinese Whisper' effect accentuates such noise.

For $f_\alpha(k) = 0.1(0.9)^{k^2}$ and $f_\omega(k) = 0.5$, we have an average degree at any time step given by $z = 0.792$. For $n = 200$ we selected both a sender vertex and a receiver vertex randomly and carried out an experiment with a fresh evolving graph in each case. Then over 17 300 such experiments, we obtain the results shown in figure 9 for the mean, mode and distribution of the STT.

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

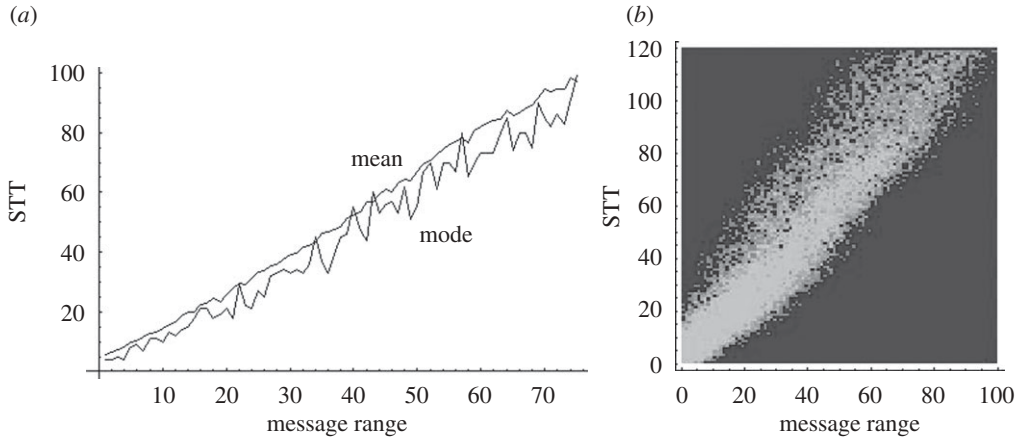

Figure 9. (*a*) Mean and mode of the STT versus message range, and (*b*) density plot for STT versus message range, obtained from 17 300 experiments.

6. Analysing propagation within an evolving graph

We now show that the evolving random network model used in figures 8 and 9 is sufficiently compact to allow for some theoretical analysis.

(*a*) *Analysis of the message passage simulations*

In order to understand the behaviour seen in figure 9, we will make two simplifying assumptions.

(i) At each time step, the nearest vertex to the receiver, say A, either remains the nearest vertex, or the new nearest vertex arises from an edge that has appeared from vertex A. We may then assume that each new edge utilized is independent of the previous ones (since we make no assumption about any edge leaving A, prior to the arrival of the message at A).
(ii) The 'edge effects' caused by the requirement that the message exactly reaches the receiver can be ignored, so that we can focus on the general phase where the message progresses as quickly as possible away from the sender.

In any element of the evolving graph, we have the equilibrium probability that any edge of length k is present at a particular time step given by equation (3.1).

Consider any particular vertex that the message has reached, and let $\pi(k)$ denote the probability that the longest range of any edge connected to it is exactly equal to k. Then, we have directly

$$\pi(0) = \prod_{j=1}^{\infty} (1 - p_0(j))^2$$

and

$$\pi(k) = (2p_0(k) - p_0(k)^2) \prod_{j=1}^{\infty} (1 - p_0(k+j))^2, \quad k \geq 1.$$

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

The sequence $\{\pi(k)\}$ sums to one, and we can use these forms to calculate the expected range of the longest range edge connected to any vertex ($=\sum k\pi(k)$).

Similarly, let $\pi^\star(k)$ denote the probability that the longest range of any edge connected to the particular vertex *in the direction of the receiver* is exactly equal to k. Then, we have

$$\pi^\star(0) = \prod_{j=1}^{\infty}(1 - p_0(j))$$

and

$$\pi^\star(k) = p_0(k)\prod_{j=1}^{\infty}(1 - p_0(k+j)), \quad k \geq 1.$$

The expected range of the longest range edge (LRE) in the direction of the receiver is thus $\sum k\pi^\star(k)$.

For the choices of $f_\alpha(k)$ and $f_\omega(k)$ in figure 9, the expected range of the longest range edge is given by 1.3598 in any direction. Similarly, the expected range of the longest range edge in the direction of the receiver is given by LRE = 0.773.

Under our simplifying assumption that edges transmit the message in the direction of the receiver, we have

$$\mathrm{STT} = \frac{k_{\mathrm{mess}}}{\mathrm{LRE}} = 1.293k_{\mathrm{mess}}.$$

This explains the slope of the mean and mode curves in figure 9. Over long message ranges we may appeal to the central limit theorem: the message propagates according to a process that could be subdivided into independent increments, for much shorter message ranges, successively drawn from a distribution of paths for the shorter ranges. Thus, the message arrival behaviour and time resembles that for a diffusion–advection process.

Of those successive time steps, we expect that a fraction $\pi^\star(0) = 0.657$ made no progress towards the receiver—the range of the longest edge in the receiver's direction being zero—for that time step at the closest vertex.

Hence, the message travels a total message range of k_{mess}, over 1.293 k_{mess} successive time steps, along approximately $1.293(1-0.657)k_{\mathrm{mess}} = 0.442k_{\mathrm{mess}}$ edges. Notice that the expected range of the longest range edge in the direction of the receiver, given that it is not zero, is simply $\sum k\pi^\star(k)/(1-\pi^*(0))$, which is approximately 2.25.

(*b*) *Analysis of the epidemic simulations*

For the computations shown in figure 8, we observed a threshold value for μ governing overall growth or decay of the disease. To understand this behaviour and estimate the critical μ, we consider, at equilibrium, a single infected vertex, A. Our aim is to estimate the number of new vertices that A infects before its eventual recovery. The expected degree of node A on the next step is the sum over the probability of all possible edges,

$$\hat{d} := 2\sum_{k=1}^{\infty} p_0(k), \tag{6.1}$$

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

where $p_0(k)$ is defined in equation (3.1). Some of the vertices that A points to may already be infected, of course. In particular, the most likely already-infected neighbour is the vertex, say B, that infected A. Because the edge death rate, f_ω, is constant in this experiment, this value gives the probability that the edge connecting B to A remains over one step. So the probability that the edge A–B remains *and* vertex B has not recovered is $f_\omega\mu$. Hence, correcting $\hat{d}$ in equation (6.1) to discount 'infection' of an already-infected node B, we arrive at

$$\tilde{d} := \hat{d} - f_\omega\mu$$

as our approximation to the expected number of new infections caused by A after one step.

Now let us consider new edges appearing at subsequent steps. Suppose we consider any such edge of length k. This edge is not present on the first time-step post-infection with probability $1 - p_0(k)$. The probability that it *first* appears as a new edge on exactly the jth time-step post-infection ($j = 2, 3, \ldots$) is thus

$$(1 - p_0(k))(1 - f_\alpha(k))^{j-2} f_\alpha(k).$$

(This expression arises as the product of the probabilities that the edge (i) is missing on the first step, (ii) does not appear in steps 2 through to $j - 1$, and (iii) appears at step j.) The expected number of edges connecting to vertex A for the first time at the jth time-step post-infection is thus

$$2\sum_{k=1}^{\infty}(1 - p_0(k))(1 - f_\alpha(k))^{j-2} f_\alpha(k).$$

The probability that vertex A is still infectious at the jth step following infection is μ^{j-1}. So, summing over all times, our overall approximation of the expected number of new vertices that A infects before recovery is

$$R(\mu) := \tilde{d} + 2\sum_{j=2}^{\infty}\mu^{j-1}\sum_{k=1}^{\infty}(1 - p_0(k))(1 - f_\alpha(k))^{j-2} f_\alpha(k). \tag{6.2}$$

This quantity is analogous to the basic *reproduction rate*, usually called R_0, that arises in classical epidemiology (Diekmann *et al.* 1990), and we are therefore interested in the critical value $R(\mu) = 1$.

For the case $f_\omega = 0.5$, $f_\alpha(k) = 0.1(0.9)^{k^2}$ used in figure 8, we show in figure 10 the behaviour of $R(\mu)$ in equation (6.2). Solving numerically, we find that $R(\mu) = 1$ at $\mu = 0.575$, which matches well with the experimentally observed threshold.

7. Discussion

Many new challenges arise when we move from fixed networks to the more general case where connections are time dependent. In this work, we highlighted some fundamental tasks in the modelling and calibration of evolving networks and proposed a novel range-dependent birth and death mechanism. This framework permits a variety of evolutionary behaviours but also imposes enough structure to make it possible for some analysis and algorithm development. We showed that the evolving network model adds value to some real brain activity data, but we believe

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

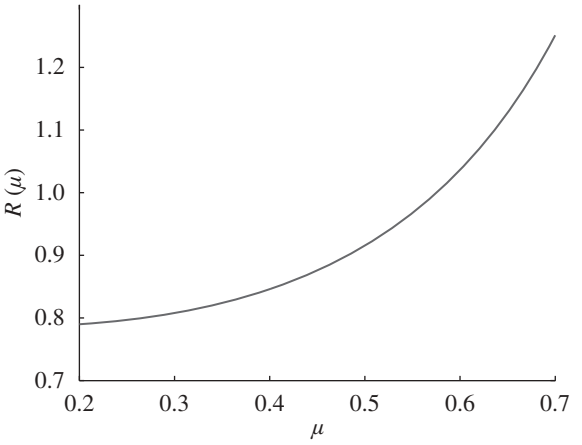

Figure 10. Plot of $R(\mu)$ in equation (6.2) for the experiment in figure 8.

that, due to the generic nature of the modelling assumptions, the same approach will be applicable to many other scenarios where sparse network 'snapshots' are observed, including applications in telecommunications, sociology and business.

Although our evolving network model was set up as a discrete-time Markov chain, we remark that an analogous continuous time framework could be developed, which might be more suited to rigorous probabilistic analysis.

The evolving network model can also be combined with a time-dependent process that acts *over* the network. We showed via simulation and analysis that the overall behaviour of the resulting stochastic process depends strongly on the interaction between the two sets of dynamics. There is much scope here for adding realistic topological dynamics to the traditional static network models of disease propagation and message passing.

We thank Slawomir Nasuto for making available some data from Sweeney-Reed & Nasuto (in press) and a template for figure 3, and for helpful discussions about the computational results in §4. The authors acknowledge support from EPSRC through project grant GR/S62383/01 and Bridging the Gaps 'Cognitive System Science' grant EP/F033036/1, and from MRC through project grant G0601353.

References

Avin, C., Koucký, M. & Lotker, Z. 2008 How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). In *ICALP '08: Proc. of the 35th Int. colloquium on automata, languages and programming, Part I*, pp. 121–132. Berlin, Germany: Springer-Verlag.

Borgnat, P., Fleury, E., Guillaume, J.-L., Robardet, C. & Scherrer, A. 2008 Description and simulation of dynamic mobility networks. *Comp. Netw.* **52**, 2842–2858. (doi:10.1061/j.comnet.2008.06.077)

Clementi, A. E., Macci, C., Monti, A., Pasquale, F. & Silvestri, R. 2008 Flooding time in edge-Markovian dynamic graphs. In *Proc. of the 27th Annual ACM SIGACT-SIGOPS Symp. on principles of distributed computing* (*PODC'08*), pp. 213–222. New York, NY: ACM Press.

Clementi, A. E., Monti, A., Pasquale, F. & Silvestri, R. In press. Information spreading in stationary Markovian evolving graphs. In *Proc. of the 23rd IEEE Int. Parallel and Distributed Processing Symp.* (*IPDPS'09*).

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021

Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382. (doi:10.1007/BF00178324)

Ding, C. H. Q., He, X., Zha, H., Gu, M. & Simon, H. D. 2001 A min-max cut algorithm for graph partitioning and data clustering. In *Proc. of the 1st IEEE Conf. on data mining*. pp. 107–114. Washington, DC: IEEE Computer Society.

Franceschetti, M. & Meester, R. 2008 *Random networks for communication: from statistical physics to information systems*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

Gautreau, A., Barrat, A. & Barthelemy, M. 2009 Microdynamics in stationary complex networks. *Proc. Natl Acad. Sci. USA* **106**, 8847–8852. (doi:10.1073/pnas.0811113106)

George, A. & Pothen, A. 1997 An analysis of spectral envelope reduction via quadratic assignment problems. *SIAM J. Matrix Anal. Appl.* **18**, 706–732. (doi:10.1137/S089547989427470X)

Grindrod, P. 2002 Range-dependent random graphs and their application to modeling large small-world proteome datasets. *Phys. Rev. E* **66**, 066 702-1–066 702-7. (doi:10.1103/PhysRevE.66.066702)

Grindrod, P., Higham, D. J. & Kalna, G. Submitted. Periodic reordering, *IMA J. Numer. Anal.*

Higham, D. J. 2003 Unravelling small world networks. *J. Comp. Appl. Math.* **158**, 61–74. (doi:10.1016/S0377-0427(03)00471-0)

Higham, D. J. 2005 Spectral reordering of a range-dependent weighted random graph. *IMA J. Numer. Anal.* **25**, 443–457. (doi:10.1093/imanum/dri003)

Higham, D. J. 2007 A small world phenomenon for Markov chains. *SIAM Rev.* **49**, 91–108. (doi:10.1137/060664987)

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. (doi:10.1098/rsif.2007.1129)

Kleinberg, J. 2000*a* Navigation in a small world. *Nature* **406**, 845. (doi:10.1038/35022643)

Kleinberg, J. 2000*b* The small-world phenomenon: an algorithmic perspective. In *Proc. 32nd ACM Symp. on theory of computing*, pp. 163–170.

Kuchaiev, O., Rašajski, M., Higham, D. J. & Pržulj, N. 2009 Geometric de-noising of protein–protein interaction networks. *PLoS Comput. Biol.* **5**, e1000454. (doi:10.1371/journal.pcbi.1000454)

Newman, M. E. J., Moore, C. & Watts, D. J. 2000 Mean-field solution of the small-world network model. *Phys. Rev. Lett.* **84**, 3201–3204. (doi:10.1103/PhysRevLett.84.3201)

Pržulj, N., Corneil, D. G. & Jurisica, I. 2004 Modeling interactome: scale-free or geometric? *Bioinformatics* **20**, 3508–3515. (doi:10.1093/bioinformatics/bth436)

Saramäki, J. & Kaski, K. 2005 Modeling epidemics with dynamic small-world networks. *J. Theor. Biol.* **234**, 413–421. (doi:10.1016/j.jtbi.2004.12.003)

Strang, G. 2008 *Computational science and engineering*. Wellesley, MA: Wellesley-Cambridge Press.

Sweeney-Reed, C. M. & Nasuto, S. J. In press. Detection of neural correlates of self-paced motor activity using empirical mode decomposition phase locking analysis. *J. Neurosci. Methods* **184**, 54–70.

Van Driessche, R. & Roose, D. 1995 An improved spectral bisection algorithm and its application to dynamic load balancing, *Parallel Comput.* **21**, 29–48. (doi:10.1016/0167-8191(94)00059-J)

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. (doi:10.1098/rspb.2008.1009)

Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)

Downloaded from https://royalsocietypublishing.org/ on 28 November 2021