

Sick and edgy: walk-counting as a metric of epidemic spreading on networks

Dennis C. Wylie and Wayne M. Getz

J. R. Soc. Interface 2009 **6**, 897-907 first published online 16 December 2008
doi: 10.1098/rsif.2008.0423

Supplementary data

["Data Supplement"](#)

<http://rsif.royalsocietypublishing.org/content/suppl/2009/03/27/rsif.2008.0423.DC1.html>

References

[This article cites 32 articles, 4 of which can be accessed free](#)

<http://rsif.royalsocietypublishing.org/content/6/39/897.full.html#ref-list-1>

Rapid response

[Respond to this article](#)

<http://rsif.royalsocietypublishing.org/letters/submit/royinterface;6/39/897>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *J. R. Soc. Interface* go to: <http://rsif.royalsocietypublishing.org/subscriptions>

Sick and edgy: walk-counting as a metric of epidemic spreading on networks

Dennis C. Wylie* and Wayne M. Getz

Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720-3112, USA

A network structure metric is herein suggested for the investigation of the behaviour of epidemic spreading processes in general network-structured populations. This simple measure, based on the algebraic powers of the adjacency matrix associated with the network in question, is shown to admit a heuristic interpretation as a representation of a spreading process similar to standard epidemic models. It is further shown that the values of this metric may be of use in understanding the dynamic pattern of epidemic spread on networks of greatly varying structural properties (e.g. the degree distribution, the assortativity/dissortativity and the clustering).

Keywords: network epidemiology; infectious disease modelling; complex networks

1. INTRODUCTION

The last decade has seen an explosion of interest in the properties of ‘complex networks’ (Albert & Barabasi 2002; Boccaletti 2006; Newman *et al.* 2006). The origin of the new focus may be traced essentially to two papers presenting the canonical examples of complex networks: first, the small-world (SW) networks of Watts & Strogatz (1998), and second, the scale-free network model of Barabasi & Albert (1999). These networks are considered ‘complex’ particularly in comparison with the Erdos–Renyi (ER) random graphs (Bollobas 2001), which have a somewhat longer history in the literature.

Network structure has been identified as an important consideration in different aspects of many areas of scientific investigation, including biochemistry and molecular biology (Jeong 2000; Giot 2003; Barabasi & Oltvai 2004), the Internet (Faloutsos *et al.* 1999; Vazquez *et al.* 2002; Eriksen 2003) and ecology (Sole & Montoya 2001; Dunne *et al.* 2002; Montoya & Sole 2002; Sole & Bascompte 2006). Theoretical studies have also highlighted the impact of network structural features on complex dynamical systems, in terms of dynamical synchronization (Nishikawa 2003; Chavez 2005; Arenas *et al.* 2006; Boccaletti 2006) and in the integration of multiple multistable subswitches (Wylie 2007).

This paper is concerned primarily with *epidemic spreading processes* on networks (Sattenspiel & Simon 1988; Hess 1996; Ball *et al.* 1997; Watts & Strogatz 1998; Pastor-Satorras & Vespignani 2001*a*; Moreno *et al.* 2002; Newman 2002; Gross *et al.* 2006). The mathematical modelling of the epidemiology of infectious diseases is an old and established subject (Diekmann & Heesterbeek 2000) that has until recently

focused on ‘macroscopic’ differential equation models making some type of mean field/homogeneous mixing assumption (Hethcote 2000; Getz & Lloyd-Smith 2006). The flexibility of this approach has been increased through the use of multicompartment models that allow a population to be divided into subgroups interacting in very distinct ways (Getz 2006).

Nevertheless, the explicit incorporation of an underlying ‘microscopic’ network structure into epidemic models has yielded some very interesting connections and results. Grassberger (1983) compared epidemic spreading on a lattice network with percolation processes of interest to statistical physicists. More recently, Pastor-Satorras & Vespignani (2001*a*) showed that the standard picture of a minimum infectivity threshold required for epidemic outbreak does not hold for the models of infectious disease spreading on scale-free network topologies (this result has been recently extended to scale-free networks with clustering as well; Serrano & Boguna 2006).

The strategy for many theoretical network epidemiology papers is to suggest a specific (set of) algorithm(s) (or dynamics) for constructing the underlying networks, and then, either through simulation or analysis, to see how the process plays out on the resulting graphs (Andersson 1998; Watts & Strogatz 1998; Pastor-Satorras & Vespignani 2001*b*; Gross *et al.* 2006). Various structural metrics may be employed to offer insights into how varying some part of the network construction algorithm leads to a particular change in the pattern of spread of disease, but this is often not done in any systematic manner. That is, it is frequently possible to imagine other networks that yield the same values for the particular metric(s) employed, but which will yield dramatically different ‘spreading dynamics’. (Consider, for example, two networks with the same degree distribution but very different patterns of clustering.)

*Author for correspondence (dennisw@socrates.berkeley.edu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2008.0423> or via <http://journals.royalsociety.org>.

Alternatively, some papers develop an analytical framework for dealing with certain restricted classes of networks (generally those described either by the degree distribution only (Moreno *et al.* 2002; Newman 2002) or by the degree distribution and the degree–degree correlations, but no other features (Moreno *et al.* 2003)) without the need to specify exactly how such networks might be constructed. Still, the results connecting the input values of the structural parameters to the output measures of epidemic spreading on such networks are in these cases technically restricted to those special networks whose structures may be described entirely by the specific metrics considered (i.e. the degree distribution and the degree–degree correlations).

By contrast, we propose and explore a structure metric that is more directly relevant to spreading processes on general networks than the most commonly available alternatives. Also, it satisfies other useful properties, such as being computationally cheap to evaluate and admitting an intuitive interpretation.

2. METHODS

2.1. Requirements for proposed metric

In order to give more precise meaning to our claim of defining a structure metric directly relevant to the behaviour of epidemic spreading processes on networks, it is necessary to consider what specific requirements must be imposed on such a hypothetical metric.

First, the metric should mimic the time course of spreading processes by consisting of a sequence of increasingly global values, loosely interpretable as the number of nodes (individuals) infected at successive time steps. In a finite network, this sequence should be subject to saturation at what is expected to be the upper limit to the size of the epidemic. Second, a versatile metric should be able to deal with complexities beyond the binary states of susceptible (S) or infected (I), for example, by including recovered/resistant states (R) that may or may not revert to susceptibility again (i.e. it would be useful if the measure could discriminate among the so-called SI, SIS, SIR and SIRS processes). Here, as a start, we focus our attention on a two-state model of the SIS type, but hopefully pointing the way to the development of measures useful for more complex problems.

Even restricting to SIS processes, one must still consider the continuum of the possible parameter values characterizing different SIS models. We confine our attention to two parameters: the transmission probability p_I that a susceptible node is infected by an infected neighbour in any one time step, and the recovery probability per time step p_S that an infected node recovers and reverts to susceptibility. Thus, we seek a metric that incorporates these two parameters, p_I and p_S , in an appropriate manner.

Finally, of course, the most important requirement must be the predictive power of the proposed metric when applied to epidemic spreading processes. For this paper, the processes to be predicted will be discrete time SIS simulations, as described in §2.6.

What sort of metric might satisfy these criteria? Given that the spreading phenomena must propagate along the paths present in the network, one might consider quantities derived from the numbers of distinct paths connecting all possible pairs of nodes. Unfortunately, these quantities are not particularly easy to calculate. However, a slight generalization from paths to walks—which differ from paths only in that they allow repetition of the same node at different steps—leads to significant mathematical simplification, as discussed in §2 in the electronic supplementary material.

2.2. Wellness and unwellness metrics

We hypothesize that the k -walk counts $(A^k)_{ij}$ from node i to j (for more details see §2 in the electronic supplementary material) provide useful information regarding the likelihood of transmission of infection from individual i to j over a k - and p_S -dependent time period. Here, $(A^k)_{ij}$ signifies the (ij) th entry of the adjacency matrix A raised to the k th power (i.e. matrix multiplied by itself k times). As discussed in §2 in the electronic supplementary material, $(A^k)_{ij}$ is also equal to the number of distinct walks of exactly k steps starting at node i and ending at node j : hence, the description ‘ k -walk counts’.

Noting that p_S^{-1} defines the mean time for an infected individual to remain infectious before returning to susceptible status, it is natural to relate $(A^k)_{ij}$ to the level of disease transmission over k such ‘generation times’ p_S^{-1} . Thus, we seek an estimate for the probability $p_i(j; k)$ of individual j being infected at time step $T_k(p_S) = \text{ROUND}[k * p_S^{-1}]$ ($\text{ROUND}[x]$ denotes rounding x to the nearest integer) of the form (the motivation for the introduction of $f = (1 - g)$ will be made clear below)

$$p_i(j; k) \sim g((A^k)_{ij}), \quad \text{where } g((A^k)_{ij}) = 1 - f((A^k)_{ij}). \quad (2.1)$$

Obviously, this requires that $f((A^k)_{ij}) \in [0, 1]$. We seek also to impose the condition that the $f((A^k)_{ij})$ satisfy

$$f((A^{k+1})_{ij}) = F(\{(A^k)_{iq} | q \leftrightarrow j\}), \quad (2.2)$$

i.e. the predicted infection status of j at time $(k+1)$ depends only on the predicted infection statuses of those individuals q neighbouring j at the previous time k . This condition amounts to demanding that the quantities $f((A^k)_{ij})$ form a localized dynamical system with discrete time index k .

Note that the interval $[0, 1]$ is closed with regard to binary multiplication (i.e. when both the numbers to be multiplied are taken from $[0, 1]$) and to being raised to any positive power (but not with regard to addition). Thus, an obvious manner in which to simultaneously satisfy the saturation condition $f((A^k)_{ij}) \in [0, 1]$ and the localized dynamics condition described by equation (2.2) would be to impose the stronger condition

$$f((A^{k+1})_{ij}) = \prod_{\{q | q \leftrightarrow j\}} f((A^k)_{iq})^{c(p_I, p_S)}, \quad (2.3)$$

where $c(p_I, p_S)$ is a constant (i.e. independent of $(A^k)_{ij}$) incorporating information regarding the transmission rate of the infection. Note that the presence of

uninfected neighbours at time k does not influence the infection status of node j at time $(k+1)$ in the dynamics resulting from equation (2.3), since a definitely uninfected node q must have $f((A^k)_{iq})=1$. This is the motivation for using f as defined in equation (2.1) instead of g , since the latter unduly complicates the expression of equation (2.3).

Is there a simple choice for a function f satisfying equation (2.3)? In order to answer this question, consider the following equivalent form of equation (2.3):

$$f\left(\sum_{\{q|q\leftrightarrow j\}}(A^k)_{iq}\right)=\prod_{\{q|q\leftrightarrow j\}}f((A^k)_{iq})^{c(p_I,p_S)}. \quad (2.4)$$

That is, in algebraic terminology, f must be a homomorphism from the additive real numbers to the (positive) multiplicative real numbers. This essentially requires that f be an exponential function,

$$f((A^k)_{ij})=\exp(-c(p_I,p_S)^k(A^k)_{ij}). \quad (2.5)$$

In this manner, we are led to consider what we refer to as the ‘wellness metric’,

$$w_{ij}(k)=\exp\left[-\left(\frac{p_I}{p_S}\right)^k(A^k)_{ij}\right], \quad (2.6)$$

where $c(p_I,p_S)$ has been specified as (p_I/p_S) . This is the simplest possible form for c satisfying the requirement that for a large (both in the number of nodes and average number of connections $\langle d \rangle$ per node) random graph, the wellness metric should be able to accurately predict the epidemic threshold condition (in this case, corresponding to the more commonly written $R_0 > 1$)

$$\frac{p_I\langle d \rangle}{p_S} > 1. \quad (2.7)$$

That equation (2.6) does indeed reproduce this result for large random graphs can be demonstrated by considering the fact that the dominant eigenvalue of the adjacency matrix A of such a network must be approximately $\langle d \rangle$ (Farkas 2001). Since A is a symmetric matrix, it can be diagonalized by an orthogonal matrix O , $A=O^TDO$, where D is the diagonal matrix of eigenvalues of A . Then $A^k=O^TD^kO$, so that $(A^k)_{ij}=\sum_m O_{mi}(\lambda_m)^k O_{mj}$, where $\lambda_m=D_{mm}$ is the m th eigenvalue of A . For large k then, $(A^k)_{ij}=\langle d \rangle^k \times [O_{1i}O_{1j} + \text{terms of order } (\lambda_2/\langle d \rangle)^k]$, where λ_2 is the second largest magnitude eigenvalue of A , which is generally significantly smaller than $\langle d \rangle$ for a large random graph. Approximating $A^k=\langle d \rangle^k O_{1i}O_{1j}$, it is then apparent that, if $(p_I\langle d \rangle/p_S) < 1$, $w_{ij}(k)$ from equation (2.6) will converge to 1 as k grows, while if instead $(p_I\langle d \rangle/p_S) > 1$, $w_{ij}(k)$ will decrease to 0.

Defining the ‘unwellness’ $u_{ij}(k)$ by (so that $u_{ij}(k)=g((A^k)_{ij})$ as defined in equation (2.1))

$$u_{ij}(k)=1-w_{ij}(k), \quad (2.8)$$

the sum $\sum_j[u_{ij}(k)]$ is then taken as a crude indicator of the expected total level of infection after $T_k=\text{ROUND}[k^*p_S^{-1}]$ time units resulting from an initial infection of individual i . Thus, the quantity

$$I_0(k)=n\langle u(k) \rangle, \quad (2.9)$$

where

$$\langle u(k) \rangle = \frac{1}{n^2} \sum_{i,j} u_{ij}(k) = \frac{1}{n^2} \sum_{i,j} [1 - w_{ij}(k)]$$

is here interpreted as affording a prediction of the expected total level of infection after T_k time units as a result of the initial infection of a single individual when that individual’s identity is undetermined.

2.3. Unwellness connectivity variance versus node variance

In equation (2.9), bracket notation was introduced for the average of the quantities $u_{ij}(k)$ over the indices i and j . That is, $\langle u(k) \rangle$ was introduced as an average over the degree to which all of the nodes are ‘connected’ to each other by k -walks filtered through the (p_I/p_S) -dependent exponential saturation function defining the unwellness metric. In this sense, one could think of $\langle u(k) \rangle$ as the average value of the ‘unwellness-connectivities’, or for short, u -connectivities $u_{ij}(k)$.

One could think of $\langle u(k) \rangle$ equally accurately, however, as the average

$$\langle u(k) \rangle = (1/n) \sum_i u_i(k), \quad (2.10)$$

where

$$u_i(k) = \frac{1}{n} \sum_j u_{ij}(k).$$

In this interpretation, $\langle u(k) \rangle$ represents the average over all nodes i of what we will term the ‘unwellness node value’ (or u -node value) $u_i(k)$. This u -node value then represents the average degree to which a particular node i is connected to the other nodes of the network by k -walks (again filtered by the saturation function of the unwellness metric). Heuristically speaking, the u -node value $u_i(k)$ may be thought of as a measure of the average unwellness of any other node in the network as a result of the initial infection of the single node i .

The reason for introducing this alternative interpretation of $\langle u(k) \rangle$ becomes clear if one focuses on the *second* moments, instead of the mean or first moment of the distribution of the values $u_{ij}(k)$. That is, the u -connectivity second moment,

$$\langle u(k)^2 \rangle = \frac{1}{n^2} \sum_{i,j} [u_{ij}(k)^2] \quad (2.11)$$

is a very different quantity than the u -node second moment,

$$\{u(k)^2\} = \frac{1}{n} \sum_i [u_i(k)^2], \quad (2.12)$$

where the u -connectivity second moment $\langle u(k)^2 \rangle$ relates to the variation in walk connectivity over all node pairs (i,j) and the u -node second moment $\{u(k)^2\}$ relates instead to the variation over single nodes i of overall connectivity to the remainder of the network.

The manner in which the different types of network structures vary with respect to these measures is discussed further in §2.6.

2.4. Modified unwellness metric

In order to construct an improved epidemic metric, it should be noted that the dynamical system described by equation (2.6) ('dynamic' in the light of its equivalence to equation (2.3), discussed in §2.2) does differ qualitatively from the usual epidemic models in one very important way: the 'wellness' of individual j at time $(k+1)$ may be arbitrarily close to 0 even if only one of its neighbours at time k has wellness differing from 1. That is, the interpretation of $w_{ij}(k+1)$ as the probability that individual j is not infected at time $(k+1)$ suffers from the unrealistic feature that this probability may be arbitrarily small, given even just one neighbour q with a sufficiently small value of $w_{iq}(k)$.

We therefore seek to modify equation (2.9) for the predicted level of infection based on the unwellness metric to address this problem. Following the discussion above, it seems reasonable to suspect that the u -connectivity second moment $\langle u(k)^2 \rangle$ (equation (2.11)) might be a particularly relevant quantity to consider. Since $u_{ij}(k) \leq 1$, it follows that $\langle u(k)^2 \rangle$ must be less than or equal to $\langle u(k) \rangle$, though typically $\langle u(k)^2 \rangle$ is much less than $\langle u(k) \rangle$ unless a substantial number of the quantities $u_{ij}(k)$ are sufficiently close to unity (requiring $w_{ij}(k) \sim 0$). Note that this is precisely the condition in which the unwellness metric is expected to overpredict the transmission.

We also incorporate the u -node second moment $\{u(k)^2\}$ (equation (2.12)) related to the variance in the network unwellness resulting from different choices of the initial infected node. Particularly, in the case where these sorts of variations in the unwellness quantities $u_{ij}(k)$ are of similar or greater magnitude to their average value (e.g. in the scale-free networks with large degree variance), it seems highly relevant to consider this quantity.

We find that both of these quantities may be profitably incorporated into the *modified unwellness metric*,

$$\begin{aligned} I_1(k) &= n\langle u(k) \rangle \exp \left[\alpha + \beta \frac{\langle u(k)^2 \rangle}{\langle u(k) \rangle} + \gamma \frac{\sqrt{\{u(k)^2\}}}{\langle u(k) \rangle} \right] \\ &= \exp(\alpha) \left[\langle u(k) \rangle + \beta \langle u(k)^2 \rangle + \gamma \sqrt{\{u(k)^2\}} + \dots \right]. \end{aligned} \quad (2.13)$$

The parameters α , β and γ appearing in equation (2.13) were fit by linear regression, as described in §2.8. The specific form of equation (2.13) (particularly, the choice of taking the square root of $\{u(k)^2\}$ but not $\langle u(k)^2 \rangle$) was chosen based on the resulting quality of the regression fit.

2.5. Application of unwellness metric to disease persistence

In §2.2–2.4, we develop the unwellness metric as a predictor of the dynamic progression of infection on networks for those cases in which the disease does not go extinct within a few generations of introduction. It is perhaps even more important, however, to gain insights into how the network structure influences the likelihood of such early extinction as opposed to persistence.

This motivates us to apply a similar regression strategy to that used in obtaining the 'modified unwellness metric' to assess the degree to and manner in which the unwellness metric may be used as a predictor of the persistence of infection.

Consider first the logistic model

$$P_{\text{persistence}} = \text{logit} \left(\kappa + \lambda \langle u(k_{\text{early}}) \rangle + \mu \frac{\langle u(k_{\text{early}})^2 \rangle}{\langle u(k_{\text{early}}) \rangle} + \nu \frac{\sqrt{\{u(k_{\text{early}})^2\}}}{\langle u(k_{\text{early}}) \rangle} \right), \quad (2.14)$$

where $P_{\text{persistence}}$ is the probability that an infection initiated at a single randomly chosen node will persist for several (here, 25) generation times and k_{early} is a fixed (low) value of the generation time index k . From the results of the regression analysis performed (as described in §2.8) on simulated SIS networks, the optimal value of k_{early} appears to be 2; this value is assumed below. This regression analysis also indicates that the u -connectivity second moment term $\langle u(k_{\text{early}})^2 \rangle / \langle u(k_{\text{early}}) \rangle$ is not a significant predictor of persistence rates. Thus, we consider instead the reduced logistic model

$$P_{\text{persistence}} = \text{logit} \left(\kappa + \lambda \langle u(k_{\text{early}}) \rangle + \nu \frac{\sqrt{\{u(k_{\text{early}})^2\}}}{\langle u(k_{\text{early}}) \rangle} \right). \quad (2.15)$$

2.6. Network construction

Four different classes of networks are considered in this paper, as follows.

- (i) *ER random graphs* with n nodes of average degree d were constructed by including each edge $i \leftrightarrow j$ with (independent) probability $(d/(n-1))$.
- (ii) *SW networks* with n nodes of (even) average degree d with rewiring probability p . These networks were constructed as in Watts & Strogatz (1998): after arranging the n nodes in a ring and connecting each node to the $(d/2)$ nearest nodes on either side of it, each edge is randomly rewired with probability p to have one of its ends connected to a new node chosen with uniform probability from the set of all nodes in the system, which are not already neighbouring the edge's fixed end.
- (iii) *Barabasi–Albert scale-free networks* with n nodes of average degree d . These networks were constructed as in Barabasi & Albert (1999). Starting with $(d+1)$ fully connected nodes, new nodes are sequentially added to the network, with each new node bringing with it $(d/2)$ new edges. Each of these $(d/2)$ new edges has one terminus fixed at the new node j , with the other terminus selected to be previously existing node i (assuming i is not already a neighbour of j) with relative probability

Table 1. Values and descriptions of the control parameters used in generating the structures of adjustable networks. (As defined in §2.6. Also, the SIS simulations (see §2.7) for regression fits discussed in §2.8 were done on each of 10 randomly generated networks from each of the $(4 \times 3 \times 3) = 36$ combinations of the values used cited here.)

control parameters	values used	comments
d_{sf}	0 2 6 10	increasing d_{sf} leads to increased degree variance
$(p_{\text{sort}}, N_{\text{sort-swap}})$	(0, 0) (1, 10^5) (-1, 10^5)	$p_{\text{sort}} > 0$ generates assortative networks; $p_{\text{sort}} < 0$ generates disassortative networks
$(C, N_{\text{cluster-swap}}, N_{\text{cluster-sort-swap}})$	no clustering—(1, 0, 0) moderate clustering—(20, 500, 10^4) severe clustering—(20, 100, 2000)	network clustered by division into C communities

$d_{i(j)}$, where $d_{i(j)}$ is the degree of node i with respect to the network made up of the first $(j-1)$ nodes only. The network is completed after the addition of the n th node (and its corresponding edges).

- (iv) *Adjustable networks* with n nodes of average degree d , described by several additional parameters (table 1) controlling the degree distribution, the community structure/clustering and the assortativity of the networks. The parameter C controls the number of distinct communities into which the network will be divided, with the level of interconnection between distinct communities increasing with the parameter $N_{\text{cluster-swap}}$ (see the electronic supplementary material). In conjunction with these clustering parameters, the parameter d_{sf} adjusts the variance of the degree distribution (increasing d_{sf} increases this variance). Finally, the sign of the parameter p_{sort} controls whether the network will be assortative ($p_{\text{sort}} > 0$) or disassortative ($p_{\text{sort}} < 0$), with the magnitude of p_{sort} adjusting the magnitude of the degree-degree correlations. The parameters $N_{\text{sort-swap}}$ and $N_{\text{cluster-sort-swap}}$ govern the number of iterations used in the assortativity adjustment algorithm (see the electronic supplementary material); as long as these are set to sufficiently high values, the desired p_{sort} -dependent assortativity patterns should be generated. The details of the algorithm used to construct these adjustable networks are described in §3 in the electronic supplementary material. Figure 1 shows how the standard deviation of the degree distribution, the transitivity and the assortativity (as measured by the Pearson correlation coefficient of the degree of the node at one end of an edge with that of the degree of the node at the other) vary with regard to the parameters in table 1 for the adjustable networks described here. (The values shown in figure 1 are averaged over 10 networks generated at each distinct combination of control parameters.) It is immediately apparent that the various parameters used to adjust the network structure do not work entirely independently of one another (as judged by these structural metrics). For instance, the variance of the degree distribution depends on both d_{sf} and on the

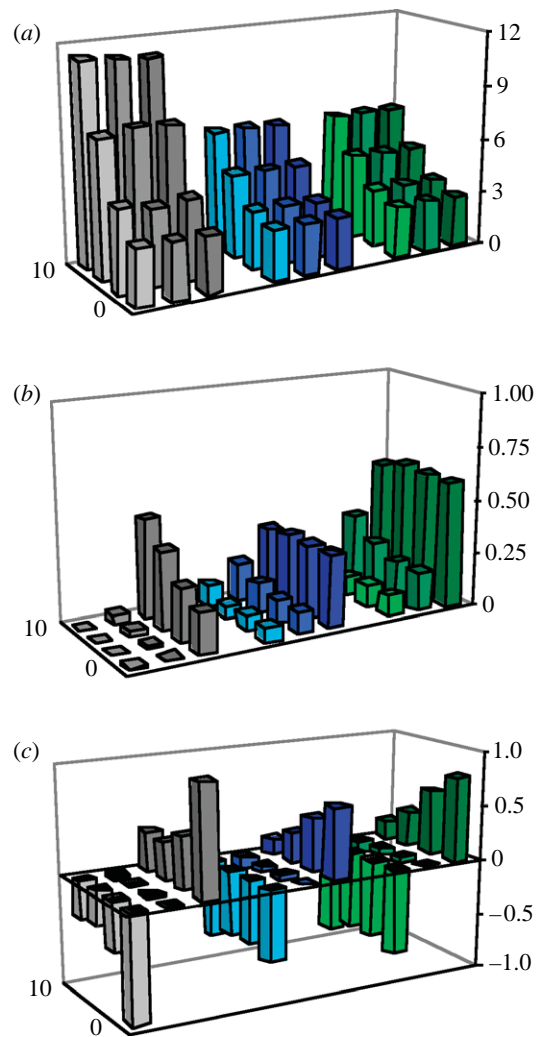


Figure 1. Properties ((a) the standard deviation of the node degree, (b) the network transitivity (clustering) and (c) the correlation of degree of neighbouring nodes (assortativity)) of ‘adjustable networks’ (§2.6) for various values of control parameters (table 1). The parameter d_{sf} increases through the sequence 0, 2, 6, 10 from front to back in each plot. The value of p_{sort} increases moving from left to right within each group of three similarly coloured bars through the sequence (-1, 0, 1). Moving across the three groups of bars (again from left to right): grey bars indicate no community structure; blue bars indicate moderate clustering into communities; and green bars indicate severe clustering into communities. All networks considered here are composed of 1000 nodes of average degree 10.

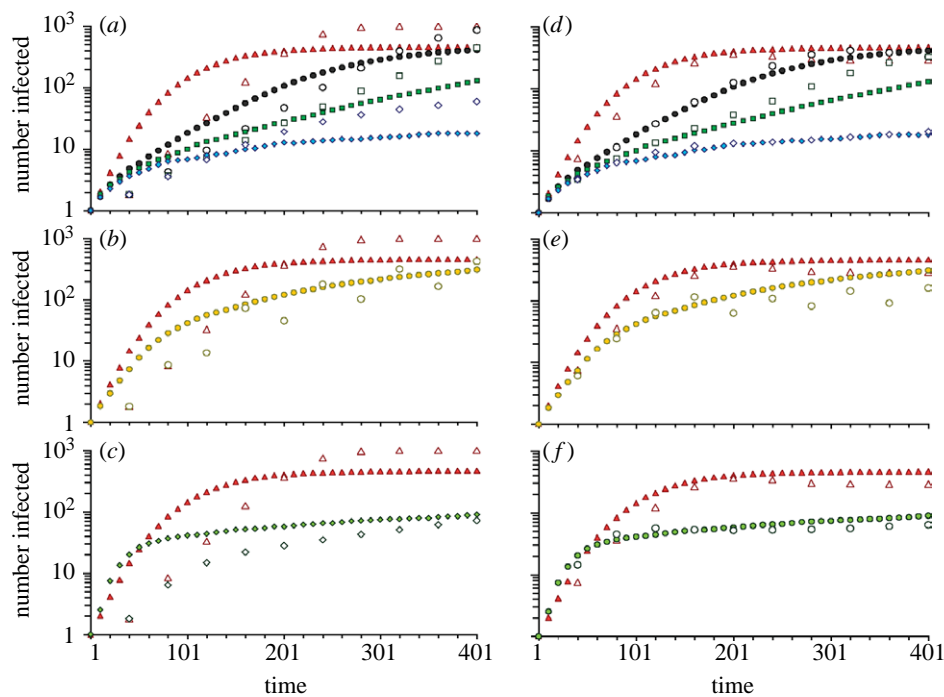


Figure 2. Comparison of the SIS simulation data (filled symbols) with the (a–c) unmodified and (d,e) modified unwellness metric predictions (open symbols). All networks are composed of 1000 nodes of average degree 10. (a,d) Red, scale free; black, random graph; light green, SW (rewiring probability $p=0.1$); blue, regular lattice. (b,e) Red, scale free; gold, dissortative scale free. (c,f) Red, scale free; green, assortative scale free.

number of communities the network is divided into, while the network transitivity depends very strongly on the assortativity rewiring parameter p_{sort} as well as the number of communities.

2.7. SIS simulations

The parameters defining the discrete-time network SIS models simulated to generate the data appearing in figures 2–4 are presented in table 2. For all of these simulations, the state of the dynamic system may be described as a network of labelled nodes. The network structure itself is here always taken to be constant (i.e. the network edges are fixed in time), but the ‘labels’ associated with the nodes—in this case, susceptible (S) or infected (I)—are functions of the discrete time index. The stochastic dynamics of the infection status of the nodes may then be described by giving the (independent) probabilities per time step of the various types of state-variable transitions available. There are only two basic types of transitions for the network SIS systems simulated herein.

- (i) $S \rightarrow I$ transitions. If node j is in the state S and has exactly m infected neighbours at time step t , node j will become infected at time step $(t+1)$ with probability $(1 - (1 - p_I)^m)$. Otherwise, node j will remain in state S.
- (ii) $I \rightarrow S$ transitions. If node j is in the state I at time t , it will transition to state S at time $(t+1)$ with probability p_S . Otherwise, node j will remain in the state I.

It should be emphasized that the discrete time step used for these simulations is not the same as the ‘generation time’ (which is equal to p_S^{-1} times the time

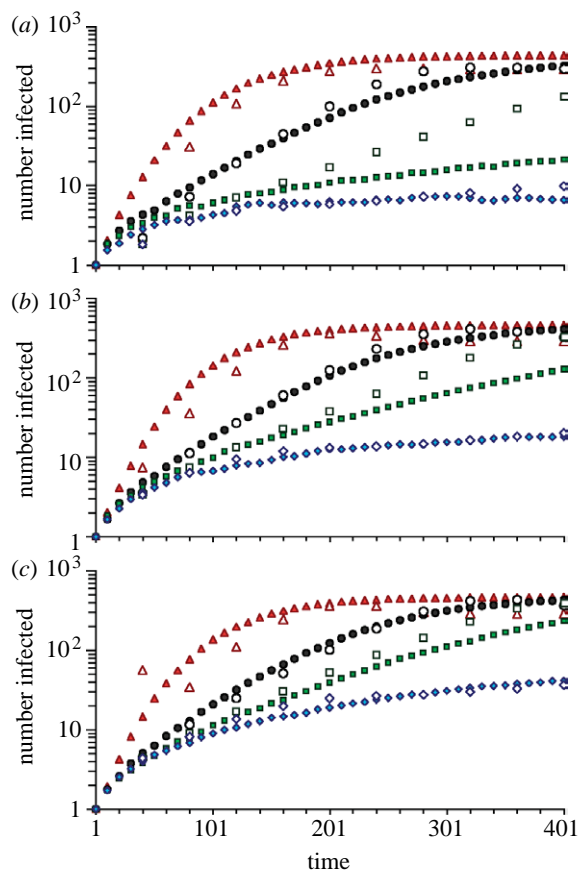


Figure 3. Comparison of the SIS simulation data (filled symbols) with the modified unwellness metric predictions (open symbols) as network edge density is varied. (a–c) Red, scale free; black, random graph; light green, SW (rewiring probability $p=0.1$); blue, regular lattice. (a) Average degree 4, (b) average degree 10 and (c) average degree 20.

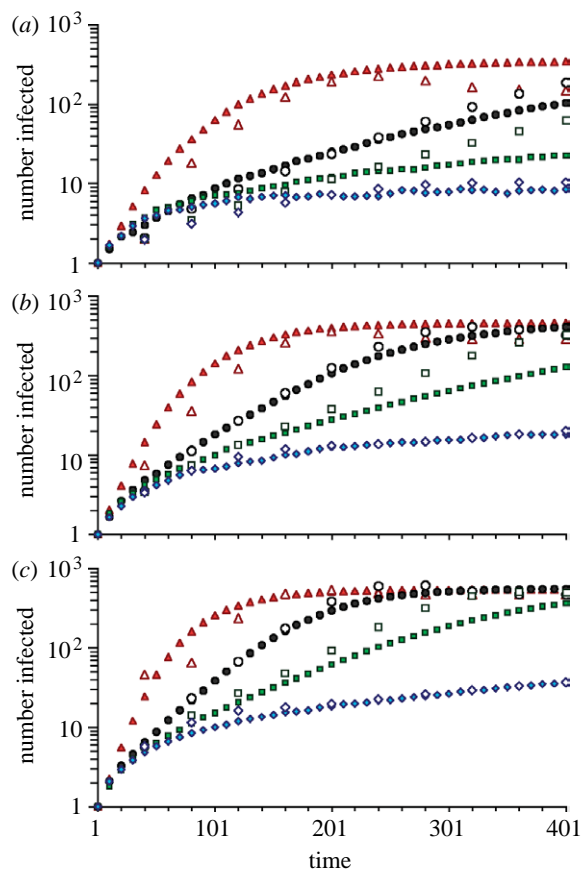


Figure 4. Comparison of the SIS simulation data (filled symbols) with the modified unwellness metric predictions (open symbols) as the disease transmissibility p_I is varied. All networks are composed of 1000 nodes of average degree 10. (a–c) Red, scale free; black, random graph; light green, SW with rewiring probability $p=0.1$; blue, regular lattice. (a) Reduced transmissibility $p_I=0.0375$ ($p_I(d)/p_S=1.5$). (b) Base case transmissibility $p_I=0.05$ ($p_I(d)/p_S=2.0$). (c) Increased transmissibility $p_I=0.0625$ ($p_I(d)/p_S=2.5$).

Table 2. Results of the regression fit (\pm s.e.) for the modified unwellness metric. (See §2.4.)

case	α	β	γ
$p_I=0.0375$	-1.007 ± 0.034	-2.617 ± 0.038	1.721 ± 0.030
base case ($p_I=0.05$)	-0.401 ± 0.035	-2.282 ± 0.023	1.444 ± 0.029
$p_I=0.0625$	0.367 ± 0.036	-2.197 ± 0.019	1.068 ± 0.030

step employed within these simulations) discussed in the context of the wellness metrics.

2.8. Regression for modified wellness metric and persistence probability models

In order to fit the parameters α , β and γ of the modified unwellness metric (see §2.4), 360 networks were constructed according to the adjustable network method described in §2.6 with systematically varied values of the network structure control parameters (table 1; figure 1). For each of the networks thus constructed, 50 SIS simulations were performed as described in §2.7. Those trajectories resulting from these stochastic simulations

Table 3. Results of the regression fit (\pm s.e.) for logistic model for persistence probability based on the unwellness metric. (See §2.5.)

case	κ	λ	ν
$p_I=0.0375$	-0.779 ± 0.090	0.337 ± 0.032	-0.992 ± 0.082
base case ($p_I=0.05$)	0.538 ± 0.085	0.070 ± 0.016	-0.958 ± 0.076
$p_I=0.0625$	0.922 ± 0.095	0.047 ± 0.010	-0.949 ± 0.085

that did not result in rapid extinction (i.e. within 1000 time steps, or 25 generation times) of the infection were then used to estimate the parameters α , β and γ of the modified wellness metric. This was done by linear regression using the model

$$\log\left(\bar{i}\left(\frac{k}{p_S}\right)\right) - \log(\langle u(k) \rangle) = \alpha + \beta \frac{\langle u(k)^2 \rangle}{\langle u(k) \rangle} + \gamma \frac{\sqrt{\{u(k)^2\}}}{\langle u(k) \rangle} + \varepsilon,$$

with random error term ε and with $\bar{i}(k/p_S)$ denoting the number of infected nodes at time step (k/p_S) averaged over those trajectories with persistent infection. The results of this fit are shown in table 2.

The parameters κ , λ and ν of the logistic model for the probability of disease persistence were estimated similarly from the same simulation data. The results of this fit are indicated in table 3.

3. RESULTS

Note that in all of figures 2–4, the simulation data presented resulted from running 250 trajectories on a single representative network of each type indicated and averaging the infection levels at each time step only over those trajectories exhibiting disease persistence through the entire 1000 time unit window run. The networks and simulations used to generate figures 2–4 were not used in the regression fits described in §§2.4 and 2.5.

3.1. Unmodified unwellness metric results

Figure 2a–c compares the results of SIS simulations (described in more detail in §2.7) on a variety of networks with the quantities $I_0(k)$ (§2.2). These results indicate that while the unwellness metric given by equation (2.9) does capture some of the qualitative variation in epidemic progression on distinct network structures, structural features beyond $\langle u(k) \rangle$ must be important to take into account as well.

The quantities $I_0(k) = n\langle u(k) \rangle$ track the average results of the SIS simulations well for some of the network types at early times, but overshoot the SIS simulations at later times. Particularly, with regard to the fact that the unwellness metric results asymptotically saturate at $u_{ij}=1$ for all i and j for those (connected) networks pictured, figure 2a–c demonstrates the consequences of the difference in the dynamics of the unwellness metric from the SIS simulations described in §2.7. The best match between the unwellness metric results and the SIS simulations is on the relatively featureless random

graphs. The unwellness metric tends to show less difference in the epidemic spreading rate (in either direction) between the random graphs and other types of network structure than do the SIS simulations (figure 2*a*). This discrepancy is especially pronounced for the SW networks: while clustering appears to have the same qualitative effect of slowing down spreading in both the unwellness metric and the SIS simulations, the degree of slowdown in the simulations is considerably greater than it is for the unwellness metric.

3.2. Modified unwellness metric results

Figure 2*d–f* shows the modified unwellness metric $I_1(k)$ (§2.4) fits rather well for a variety of networks, particularly the Erdos random graph, regular lattice, scale free and scale-free assortative networks. On the SW network featuring both a high degree of clustering and the ‘SW’ property of low characteristic path length, the modified wellness metric predicts the SIS simulation data very well over the first four to five generation times, but overpredicts the epidemic spread beyond this point. Finally, on the dissortative scale-free network, the prediction develops an interesting (but pathological) oscillatory behaviour starting at the fifth generation.

Figure 3 compares the results of the modified unwellness metric model of equation (2.13) with networks as the density of network connections is varied (in these cases, with the disease transmissibility parameter varying inversely so as to hold the expected number of secondary infections resulting from a randomly selected initially infected node constant). Two features of the results shown in figure 3 merit special attention. First, it is apparent that the model predictions generally fare better in comparison with the simulation data in denser networks than sparser networks. This trend is most noticeable for the highly clustered SW and lattice networks. Second, the predicted values for the first-generation time index $k=1$ overshoot the simulated values on the high degree-variance scale-free network topology at higher connectivity densities. Interestingly, as the generation time index k increases past 1, this behaviour rapidly vanishes.

Robustness of the fit parameters α , β and γ with respect to the ratio (which becomes equal to the epidemic threshold R_0 in the limit of large random graphs with high average connectivity)

$$\frac{p_I \langle d \rangle}{p_S}$$

was also studied through variation of the disease transmissibility p_I (table 2; figure 4). For the values of transmissibility considered, all three model parameters were significant, with β consistently negative and γ consistently positive. The parameter γ associated with the u -node second moment term appears to be more important for lower values of the transmissibility. This may reflect the larger importance of the bias of persistent disease trajectories towards those initiated from more well-connected nodes when disease extinction is a greater possibility (i.e. when transmissibility is lower). The quantitative value of the fit parameter β is considerably more robust to the disease transmissibility (table 2).

Finally, it is interesting to note that the results for the modified unwellness metric predictions with increased transmissibility show a qualitatively similar, albeit smaller, overshoot for the scale-free networks at the first-generation time index $k=1$ as appeared in the increased edge density situation.

3.3. Results of unwellness-based logistic model for persistence

The results of the regression fit of the model specified by equation (2.15) in §2.5 are shown in table 3. Table 4 displays the resulting predictions contrasted with the simulated data for various network types (again, these simulations were not used in the regression analysis determining the values of κ , λ and ν). The assortativity/dissortativity of the network structure exhibits a very large impact on the probability of persistence, which the model does a good job of capturing, largely through the u -node second moment term. Smaller effects seem to be associated with the clustering and the degree distribution.

Especially in combination with high clustering, reduced network density (again varying the infection transmissibility inversely to hold the product of connectivity and transmissibility constant) can lead to sharply decreased persistence probability. This is demonstrated by the simulation results in table 4. Also apparent in this table is the breakdown of the logistic model of equation (2.15) when applied to such sufficiently sparse (e.g. average degree 4 in the data presented in table 4) and clustered networks. Conversely, at higher network densities (e.g. networks of average degree 20 in table 4), the logistic model for persistence probability appears to perform much more reliably.

The parameter λ associated with the $\langle u(k_{\text{early}}) \rangle$ (recall from §2.5 that $k_{\text{early}}=2$ here) term is found to be positive, suggesting that the networks for which the unwellness metric predicts faster epidemic spread tend to be more likely to sustain disease persistence. Interestingly, from the robustness with respect to transmissibility (and hence, with respect to the ratio $(p_I \langle d \rangle / p_S)$) results displayed in table 5, it is apparent that the impact of this unmodified unwellness metric term is much larger for relatively lower transmissibility.

The parameter ν associated with the u -node second moment term $\{u(k_{\text{early}})^2\}^{1/2} / \langle u(k_{\text{early}}) \rangle$ is found to be negative (and fairly robust to variation in the disease transmissibility). While considered in isolation, this might be taken to suggest that the networks in which there is especially large variation in the secondary infections produced as a result of infecting different initial nodes tend towards decreased propensity for disease persistence, examination of the results of simulations on (non-assortative) scale-free networks shows that this is not the whole story. In this case, in particular, the effects of the relatively large u -node second moment term are offset by a coinciding increase in the u -connectivity first moment term $\langle u(k_{\text{early}}) \rangle$.

However, shifting attention to the assortative scale-free network simulation results, it becomes apparent that the combination of highly variable degree distribution and assortative degree correlations can result in even

Table 4. Comparison of the SIS simulation data (\pm s.e. of estimate) and the unwellness metric-based logistic model predictions for disease persistence probability. (See §2.5. Note that the base case networks have average degree 10, sparse networks have average degree 4 and dense networks have average degree 20.)

network type	persistence in SIS simulations	predicted persistence	ratio of simulated : predicted persistence
scale free	0.472 ± 0.032	0.445	1.06
random graph	0.424 ± 0.031	0.456	0.93
small world ($p=0.1$)	0.404 ± 0.031	0.458	0.88
regular lattice	0.404 ± 0.031	0.458	0.88
SF assortative	0.288 ± 0.029	0.245	1.17
SF dissortative	0.444 ± 0.031	0.507	0.87
scale free (sparse)	0.456 ± 0.032	0.441	1.03
random graph (sparse)	0.372 ± 0.031	0.433	0.86
small world (sparse)	0.280 ± 0.028	0.449	0.62
regular lattice (sparse)	0.140 ± 0.022	0.449	0.31
SF assortative (sparse)	0.204 ± 0.026	0.209	0.97
SF dissortative (sparse)	0.508 ± 0.032	0.511	0.99
scale free (dense)	0.468 ± 0.032	0.443	1.06
random graph (dense)	0.448 ± 0.031	0.462	0.97
small world (dense)	0.484 ± 0.032	0.462	1.05
regular lattice (dense)	0.472 ± 0.032	0.461	1.02
SF assortative (dense)	0.292 ± 0.029	0.286	1.02
SF dissortative (dense)	0.532 ± 0.032	0.496	1.07

Table 5. Comparison of the SIS simulation data (\pm s.e. of estimate) and the unwellness metric-based logistic model predictions for disease persistence probability. (See §2.5. Note that the base case networks have transmissibility $p_1=0.05$; all the networks indicated have average degree 10. The parameters for logistic model fits for each transmissibility case are indicated in table 3.)

network type	persistence in SIS simulations	predicted persistence	ratio of simulated : predicted persistence
scale free	0.472 ± 0.032	0.445	1.06
random graph	0.424 ± 0.031	0.456	0.93
small world ($p=0.1$)	0.404 ± 0.031	0.458	0.88
regular lattice	0.404 ± 0.031	0.458	0.88
SF assortative	0.288 ± 0.029	0.245	1.17
SF dissortative	0.444 ± 0.031	0.507	0.87
scale free ($p_1=0.0375$)	0.352 ± 0.030	0.382	0.92
random graph ($p_1=0.0375$)	0.296 ± 0.029	0.267	1.11
small world ($p_1=0.0375$)	0.244 ± 0.027	0.260	0.94
regular lattice ($p_1=0.0375$)	0.120 ± 0.021	0.257	0.47
SF assortative ($p_1=0.0375$)	0.224 ± 0.026	0.153	1.47
SF dissortative ($p_1=0.0375$)	0.348 ± 0.030	0.439	0.79
scale free ($p_1=0.0625$)	0.572 ± 0.031	0.554	1.03
random graph ($p_1=0.0625$)	0.592 ± 0.031	0.557	1.06
small world ($p_1=0.0625$)	0.604 ± 0.031	0.557	1.08
regular lattice ($p_1=0.0625$)	0.528 ± 0.032	0.555	0.95
SF assortative ($p_1=0.0625$)	0.316 ± 0.029	0.337	0.94
SF dissortative ($p_1=0.0625$)	0.600 ± 0.031	0.604	0.99

larger $\{u(k_{\text{early}})^2\}^{1/2}/\langle u(k_{\text{early}}) \rangle$ values without proportionally increased $\langle u(k_{\text{early}}) \rangle$ values, and in this case there is indeed a dramatically lowered likelihood of disease persistence.

4. DISCUSSION AND CONCLUSIONS

The ‘unwellness metric’, incorporating u -connectivity, the u -connectivity second moment and the u -node second moment, is introduced in this paper as a way of directly relating network structure data to epidemic spreading processes. This metric may be calculated in an entirely straightforward manner from the adjacency matrix of a network and herein is shown to contain a great deal of information regarding

the rate of spreading of infection, as well as the likelihood of persistence of infection—at least, for networks with a sufficiently high edge density (approx. 10 edges per node).

It is worthwhile to pause for a moment to consider the relationship of the unwellness metric to a few more familiar measures of network structure. The networks with more variability in the distribution of node degrees (e.g. the scale-free networks) tend to lead to larger average values of the u -connectivity $\langle u(k) \rangle$ (figure 2*a*), while more highly clustered networks tend towards smaller average u -connectivity. These particular trends in u -connectivity appear to qualitatively reflect the different speeds with which the infections spread on such networks (figure 2*a*).

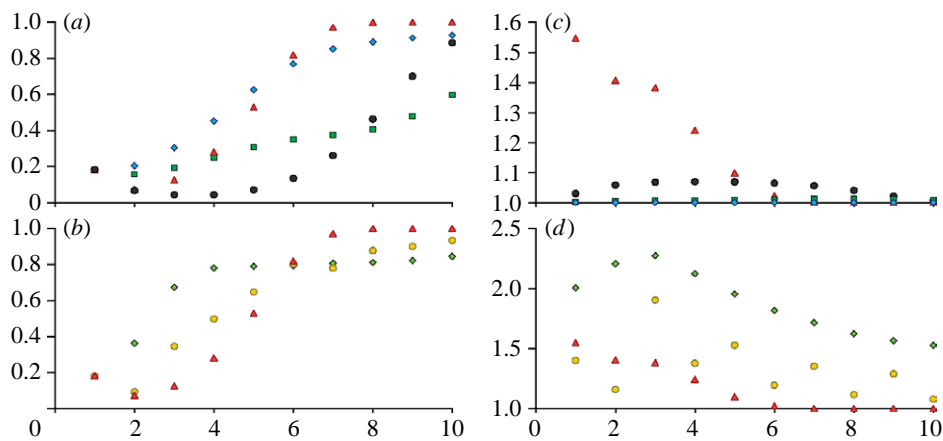


Figure 5. Plots of (a,b) $\langle u(k)^2 \rangle / \langle u(k) \rangle$ and (c,d) $\{u(k)^2\}^{1/2} / \langle u(k) \rangle$ (vertical axes) versus k (horizontal axes) for various network types. All networks are composed of 1000 nodes of average degree 10. (a,c) Red, scale free; black, random graph; light green, SW with rewiring probability $p=0.1$; blue, regular lattice. (b,d) Red, scale free; gold, dissortative scale-free network; green, assortative scale free.

Somewhat surprisingly, the scale-free networks with either assortative or dissortative degree correlations were here found to exhibit lower u -connectivity values than those with no such correlations (figure 2*b–c*). While this phenomenon does seem to qualitatively match with the infection simulation results at times following the initial epidemic explosion, it does not capture the greater speed of outbreaks during the early transient phase in assortative networks (figure 2*c*). For this case, it appears that the increased variation, not just the decreased mean, of the distribution of u -connectivities $u_{ij}(k)$ must be taken into account in understanding the transmission dynamics.

Two distinct measures of such variation were proposed in §2.4, the u -connectivity second moment $\langle u(k)^2 \rangle$ and the u -node second moment $\{u(k)^2\}$. The u -connectivity second moment is associated with the variation in the strength of coupling of nodes taken in pairs, while the u -node second moment corresponds with the variation in connectivity of individual nodes to their surroundings. Large variation in the degree distribution tends to lead to relatively large values of the u -connectivity second moment compared with random graph structure, but at early generation times k , the network clustering leads to an even larger increase in this quantity (figure 5*a,b*). On the other hand, while, as one might intuit, more heterogeneous degree distributions lead to much larger values of the u -node second moment (figure 5*c,d*; again, for low k), clustering tends to depress this particular measure. Assortative degree correlations tend to further increase both the measures of variation for low k (figure 5). Dissortativity tends to lead to larger values of the u -connectivity second moment at low k , but exhibits a somewhat more complex pattern for the u -node second moment (figure 5).

The unwellness metric, modified to incorporate the variation measures as discussed, was used to predict both the probability that a single initial infection would lead to a persistent endemic infection (tables 4 and 5) and the transient time course of initial epidemic outbreak in persistent trajectories (figures 2–4). Both infection persistence and speed of outbreak tend to increase with the average u -connectivity. The speed of outbreak tends

to be reduced for networks with larger u -connectivity second moments, but these quantities exhibited no significant effect on the persistence probability. Finally, networks for which the u -node second moment was particularly pronounced tended to show increased outbreak speed but decreased likelihood of persistence.

The results presented herein pertain to the SIS-type epidemic spreading processes for which recovery from infection provides negligible immunity to re-infection. The extension of the unwellness metric approach to the SIRS-type epidemic models poses some interesting challenges. The simplest option for such extension would be to apply the regression strategy employed in the modified unwellness metric discussed here to a set of SIRS models. In this manner, one might get a sense of how the structural features introduced here (the average u -connectivity, the u -connectivity second moment and the u -node second moment) influence the dynamics of infection in such models. One complication with this approach is that it would require multiple regressions to be done for situations involving different characteristic time scales for the loss of immunity (i.e. $R \rightarrow S$).

A more complex approach to extending the unwellness metric to SIRS processes would be to try and replace the basic definition of $w_{ij}(k)$ (equation (2.6)) with a quantity naturally incorporating the $R \rightarrow S$ rate parameter. For instance, one might try and modify $w_{ij}(k)$ by subtracting out a quantity representing a measure of the number of immune individuals who might have, in the absence of immunity, been expected to have become infected in the interval between time index $(k-1)$ and time index (k) . Such a modification of $w_{ij}(k)$ would have to be done with care in order to preserve the various properties discussed in §§2.1 and 2.2 (especially boundedness, i.e. $0 < w_{ij}(k) < 1$).

It is hoped that the results of this study will prove useful in translating the data on social network structures into practicable predictions regarding the effectiveness of proposed intervention strategies such as vaccination, isolation and quarantine strategies that alter the network structures in ways that reduce transmission rates, possibly leading to the eradication of particular infectious diseases. In particular, it is worth emphasizing that the

unwellness metric may be easily applied to networks with arbitrary clustering and degree correlation patterns. It also bears mentioning that this is not only a useful property for application to empirical network data, but also for application with regard to the network structures produced by dynamical network models which may not lend themselves to simple analytical solution.

This work was supported by a James S. McDonnell Foundation 21st Century Science Innovation Award and NIH grant no. GM83863 to W.M.G.

REFERENCES

- Albert, R. & Barabasi, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)
- Andersson, H. 1998 Limit theorems for a random graph epidemic model. *Ann. Appl. Prob.* **8**, 1331–1349. (doi:10.1214/aoap/1028903384)
- Arenas, A., Diaz-Guilera, A. & Perez-Vicente, C. J. 2006 Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.* **96**, 114102-1–114102-4. (doi:10.1103/PhysRevLett.96.114102)
- Ball, F., Mollison, D. & Scalia-Tombo, G. 1997 Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–89. (doi:10.1214/aoap/1034625252)
- Barabasi, A.-L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
- Barabasi, A.-L. & Oltvai, Z. N. 2004 Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113. (doi:10.1038/nrg1272)
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. 2006 Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308. (doi:10.1016/j.physrep.2005.10.009)
- Bollobas, B. 2001 *Random graphs*, 2nd edn. Cambridge, UK: Cambridge University Press.
- Chavez, M., Hwang, D.-U., Amann, A. & Hentschel, H. G. E. 2005 Synchronization is enhanced in weighted complex networks. *Phys. Rev. Lett.* **94**, 218701-1–218701-4. (doi:10.1103/PhysRevLett.94.218701)
- Diekmann, O. & Heesterbeek, J. A. P. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Chichester, UK: Wiley.
- Dunne, J. A., Williams, R. J. & Martinez, N. D. 2002 Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12 917–12 922. (doi:10.1073/pnas.192407699)
- Eriksen, K. A., Simonsen, I., Maslov, S. & Sneppen, K. 2003 Modularity and extreme edges of the Internet. *Phys. Rev. Lett.* **90**, 148701-1–148701-4. (doi:10.1103/PhysRevLett.90.148701)
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. 1999 On power-law relationships of the Internet topology. In *Proc. Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communication*, vol. 29, pp. 251–262.
- Farkas, I. J., Derényi, I., Barabási, A. L. & Vicsek, T. 2001 Spectra of “real-world” graphs: beyond the semicircle law. *Phys. Rev. E* **64**, 026704-1–026704-12. (doi:10.1103/PhysRevE.64.026704)
- Getz, W. M. & Lloyd-Smith, J. O. 2006 Basic methods for modeling the invasion and spread of contagious disease. In *Disease evolution: models, concepts, and data analysis*, vol. 71 (eds Z. Feng, U. Dieckmann & S. A. Levin). AMS/DIMACS series, pp. 87–109. Providence, RI: American Mathematical Society.
- Getz, W. M., Lloyd-Smith, J. O., Cross, P. C., Bar-David, S., Johnson, P. L., Porco, T. C. & Sánchez, M. S. 2006 Modeling the invasion and spread of contagious disease in heterogeneous populations. In *Disease evolution: models, concepts, and data analysis*, vol. 71 (eds Z. Feng, U. Dieckmann & S. A. Levin). AMS/DIMACS Series, pp. 113–144. Providence, RI: American Mathematical Society.
- Giot, L. *et al.* 2003 A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736. (doi:10.1126/science.1090289)
- Grassberger, P. 1983 On the critical behavior of the general epidemic process and dynamical percolation. *Math. Biosci.* **63**, 157–172. (doi:10.1016/0025-5564(82)90036-0)
- Gross, T., D’Lima, C. J. D. & Blasius, B. 2006 Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.* **96**, 208701-1–208701-4. (doi:10.1103/PhysRevLett.96.208701)
- Hess, G. 1996 Disease in metapopulation models: implications for conservation. *Ecology* **77**, 1617–1632. (doi:10.2307/2265556)
- Hethcote, H. W. 2000 The mathematics of infectious disease. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. 2000 The large-scale organization of metabolic networks. *Nature* **407**, 651–654. (doi:10.1038/35036627)
- Montoya, J. M. & Sole, R. V. 2002 Small world patterns in food webs. *J. Theor. Biol.* **214**, 405–412. (doi:10.1006/jtbi.2001.2460)
- Moreno, Y., Pastor-Satorras, R. & Vespignani, A. 2002 Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529. (doi:10.1140/epjb/e20020122)
- Moreno, Y., Gomez, J. B. & Pacheco, A. F. 2003 Epidemic incidence in correlated complex networks. *Phys. Rev. E* **68**, 035103-1–035103-4. (doi:10.1103/PhysRevE.68.035103)
- Newman, M. E. J. 2002 Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128-1–016128-11. (doi:10.1103/PhysRevE.66.016128)
- Newman, M., Barabasi, A.-L. & Watts, D. J. 2006 *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.
- Nishikawa, T., Motter, A. E., Lai, Y. & Hoppensteadt, F. C. 2003 Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? *Phys. Rev. Lett.* **91**, 014101-1–014101-4. (doi:10.1103/PhysRevLett.91.014101)
- Pastor-Satorras, R. & Vespignani, A. 2001a Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **63**, 066117-1–066117-8. (doi:10.1103/PhysRevE.63.066117)
- Pastor-Satorras, R. & Vespignani, A. 2001b Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)
- Sattenspiel, L. & Simon, C. P. 1988 The spread and persistence of infectious diseases in structured populations. *Math. Biosci.* **90**, 341–366. (doi:10.1016/0025-5564(88)90074-0)
- Serrano, M. A. & Boguna, M. 2006 Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.* **97**, 3200. (doi:10.1103/PhysRevLett.97.088701)
- Sole, R. V. & Bascompte, J. 2006 *Self-organization in complex ecosystems*. Princeton, NJ: Princeton University Press.
- Sole, R. V. & Montoya, J. M. 2001 Complexity and fragility in ecological networks. *Proc. R. Soc. B* **268**, 2039–2045. (doi:10.1098/rspb.2001.1767)
- Vazquez, A., Pastor-Satorras, R. & Vespignani, A. 2002 Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E* **65**, 066130-1–066130-12. (doi:10.1103/PhysRevE.65.066130)
- Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of small-world networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
- Wylie, D. C. 2007 Linked by loops: switch integration in complex dynamical systems. See <http://arXiv:q-bio/0704.3640v2>.