



# Disease spread in small-size directed networks: Epidemic threshold, correlation between links to and from nodes, and clustering

Mathieu Moslonka-Lefebvre<sup>a,b,c</sup>, Marco Pautasso<sup>c,\*</sup>, Mike J. Jeger<sup>c</sup>

<sup>a</sup> Department of Biology, Université Paris-Sud XI, 91405 Orsay, France

<sup>b</sup> Biochemistry and Bio-engineering Department, ENS Cachan, 61 av President Wilson, 94230 Cachan, France

<sup>c</sup> Division of Biology, Imperial College London, Silwood Park, Ascot SL5 7PY, UK

## ARTICLE INFO

### Article history:

Received 13 January 2009

Received in revised form

15 April 2009

Accepted 7 June 2009

Available online 21 June 2009

### Keywords:

Commercial transport

Disease management

Food webs

Initial conditions

Spread of ideas

Susceptible–infected–susceptible (SIS)

## ABSTRACT

Network epidemiology has mainly focused on large-scale complex networks. It is unclear whether findings of these investigations also apply to networks of small size. This knowledge gap is of relevance for many biological applications, including meta-communities, plant–pollinator interactions and the spread of the oomycete pathogen *Phytophthora ramorum* in networks of plant nurseries. Moreover, many small-size biological networks are inherently asymmetrical and thus cannot be realistically modelled with undirected networks. We modelled disease spread and establishment in directed networks of 100 and 500 nodes at four levels of connectance in six network structures (local, small-world, random, one-way, uncorrelated, and two-way scale-free networks). The model was based on the probability of infection persistence in a node and of infection transmission between connected nodes. Regardless of the size of the network, the epidemic threshold did not depend on the starting node of infection but was negatively related to the correlation coefficient between in- and out-degree for all structures, unless networks were sparsely connected. In this case clustering played a significant role. For small-size scale-free directed networks to have a lower epidemic threshold than other network structures, there needs to be a positive correlation between number of links to and from nodes. When this correlation is negative (one-way scale-free networks), the epidemic threshold for small-size networks can be higher than in non-scale-free networks. Clustering does not necessarily have an influence on the epidemic threshold if connectance is kept constant. Analyses of the influence of the clustering on the epidemic threshold in directed networks can also be spurious if they do not consider simultaneously the effect of the correlation coefficient between in- and out-degree.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Epidemic models assuming regularly or randomly connected individuals are now involving more complex networks (Keeling, 2005; May, 2006; Jeger et al., 2007). Compared to regular lattices, epidemics in small-world networks are facilitated by long-distance connections (Moore and Newman, 2000). In scale-free networks of infinite size, epidemics lack a threshold, which implies that even pathogens with a low probability of transmission will persist (Pastor-Satorras and Vespignani, 2001). Whether these findings also apply to complex networks of small size is still unclear. This is an important knowledge gap as small-size networks are relevant to many epidemics spreading within subgroups of individuals (Liu et al., 2004; Guimarães et al., 2007; Sun and Gao, 2007; Pellis et al., 2009).

\* Corresponding author. Tel.: +44 020 759 42533.

E-mail address: [m.pautasso@ic.ac.uk](mailto:m.pautasso@ic.ac.uk) (M. Pautasso).

Networks are not only relevant for epidemiology but have found application in a variety of biological systems (Proulx et al., 2005; Bascompte, 2007; Jeger et al., 2007). Networks with size of the order of magnitude of hundreds of nodes, in turn, are relevant for many current issues in ecology. Examples include closely interacting primate groups (Dunbar, 1993), social networks of manakins (Ryder et al., 2008), cavity-nesting community webs (Blanc and Walters, 2007), diseases of bumble bee colonies (Otterstatter and Thomson, 2007) and vascular epiphytes on host tree species (Lobel et al., 2006; Burns, 2007). The importance of small-size networks in biology is further shown by plant–pollinator interactions (e.g. Olesen et al., 2006; Nielsen and Bascompte, 2007), mycorrhiza, rhizomorphs, and plant pathosystems (Southworth et al., 2005; Lamour et al. 2007; Brooks et al., 2008), and food webs (e.g. Montoya and Solé, 2002; Neutel et al., 2007). Although networks of small size are relevant both for epidemiology and for ecology and evolution (e.g. Hanski and Ovaskainen, 2000; Brooks, 2006; Ings et al., 2009), there has been surprisingly little theoretical work investigating whether results

obtained for large-size and infinite networks also apply to small-size networks. We know for example that heterogeneity in the contact structure can markedly lower invasion thresholds in networks of large size (Boguna et al., 2004; Colizza and Vespignani, 2007; Jeger et al., 2007), but there is still only patchy knowledge about the dynamics and properties of small-size networks of various structures. This is of concern, given the several applications of small-size networks in natural systems and also given that the rapid globalization and structural changes of trade interactions are increasing the relevance of complex networks of small size for the invasion biology of exotic organisms (Jones and Baker, 2007; Dehnen-Schmutz et al., 2007; Brenn et al., 2008). A real world application of directed networks of small size in epidemic controlling are trade movements of infected material amongst plant nurseries (e.g. Bandyopadhyay and Frederiksen, 1999; Anderson et al., 2004; Slippers et al., 2005). A recent example is given by the spread in regional networks of plant nurseries and garden centres of *Phytophthora ramorum*, the oomycete causing Sudden Oak Death in the West Coast of the USA and leaf blight and dieback in many ornamental shrubs both in America and Europe (Werres et al., 2001; Prospero et al., 2007; Xu et al., 2009). Given the wide range of species affected, movements of infected plant material in the horticultural trade have the potential to make this emerging plant disease outbreak even more widespread (Holdenrieder et al., 2004; Frankel, 2008; Grünwald et al., 2008). There is thus the necessity to control the network of plant nurseries and retail outlets trading ornamental species susceptible to *P. ramorum* in an effective and efficient way.

In this study, we investigated whether heterogeneity in the contact structure and the presence of short-cuts as in small-world networks still make a difference to epidemic development in small-size, directed networks. Directed networks, given the more complicated adjacency matrices, have been used to model epidemics relatively rarely (Newman et al., 2001; Meyers et al., 2006; Park and Kim, 2006; Kenah and Robins, 2007), but are relevant to many real-world situations with asymmetries in contact structures, and deserve more study for various network sizes, structures, and levels of connectance. We have shown elsewhere (Pautasso and Jeger, 2008) that heterogeneity in the contact structure still affects the epidemic threshold even in the case of networks of one hundred nodes, but inquire here whether variations in the epidemic threshold (the boundary between no epidemic and an epidemic) for different network structures and at different levels of connectance can be explained by the correlation coefficient between links to and from nodes and by the clustering coefficient of the network. Previous work in percolation theory suggests that the correlation coefficient between links to and from nodes is important in the case of heterogeneity in the contact structure (Schwartz et al., 2002; Woolhouse et al., 2005; Kao et al., 2006), but it is unclear whether or not this importance is maintained for small-size networks with different (i) size, (ii) structure (local, random, small-world, and scale-free), (iii) levels of connectance, and (iv) clustering.

## 2. Materials and Methods

We simulated disease spread and establishment in networks of 100 and 500 nodes. For both network sizes, we used six kinds of structure: (1) local (nearest-neighbour transmission), (2) random (nodes connected with probability  $p$ ), (3) small-world (local networks rewired with short-cuts), and scale-free structure (see Jeger et al., 2007 for a visualization). For scale-free networks, we considered separately networks with in- and out-degree of nodes (4) positively, (5) not, and (6) negatively correlated. The networks were directed, i.e. a link from node  $a$  to node  $b$  did not imply the

reverse connection (as e.g. in Newman et al., 2001; Boguna and Serrano, 2005; Meyers et al., 2006; Park and Kim, 2006). Directed networks are realistic approximations of many real world systems; wherever a directed link does not entail the reverse connection (asymmetrical interactions), from food webs to plant–animal mutualistic networks, from infectious disease epidemiology to the spread of information (Bascompte et al., 2003; da Gama and Nunes, 2006; Bode et al., 2008; Thebault and Fontaine, 2008).

For each network structure, 100 replicates were built in MATLAB at each level of connectance (for 100 nodes: 100, 200, 400, and 1000 links; for 500 nodes: 500, 2500, 5000, and 10 000 links; Table 1). Rather than keeping connectance constant between the two network sizes, we needed a lower level of connectance for 500 nodes (500 links, i.e. connectance = 0.002) to obtain results similar to the lower level of connectance for 100 nodes (100 links, i.e. connectance = 0.01). To obtain a disappearance of a significant correlation between epidemic threshold and correlation links in and out for the larger network size (500 nodes), it was necessary to use a lower connectance level that at network size equal 100 nodes. We also left out results from the highest level of connectance for 500 nodes (25 000 links, i.e. connectance = 0.1) as in this case networks were overly connected so that differences between network structures were likely to be small. Local networks were built starting from a regular ring with 100 (for 100 nodes) and 500 (for 500 nodes) links more than the target number of links and by randomly generating 100 or 500 gaps, respectively. Random digraphs were generated using the  $G(N, M)$  model where  $M$  directed links are placed randomly and independently between the  $N$  nodes of the graph. Small-world networks were built with the Watts and Strogatz (1998) algorithm and a rewiring coefficient of 0.25. This rewiring coefficient allowed the construction of small-world networks with clustering intermediate between those of random and of local networks. Small-world networks are networks with high clustering compared to random networks and small shortest path length compared to local networks, and these conditions were respected in our case. Scale-free networks were built with a preferential attachment algorithm, starting with a seed network and based on five parameters adding nodes and/or links depending on the in-, out-, and total degree of existing nodes:

- (1) The parameter  $a$  added both a node  $i$  and an arrow  $(i, j)$  according to the in-degree of an existing node  $j$ . This step was repeated  $a$ -times.
- (2) The parameter  $b$  added both a node  $i$  and an arrow  $(j, i)$  according to the out-degree of an existing node  $j$ . This step was repeated  $b$ -times.
- (3) The parameter  $c$  added both a node  $i$  and an arrow  $(i, j)$  or  $(j, i)$  according to the total degree of an existing node  $j$ . The choice of the direction of each arrow added, i.e.  $(i, j)$  or  $(j, i)$ , was equally probable. This step was repeated  $c$ -times.
- (4) The parameter  $d$  only added an arrow  $(i, j)$  and not a node according to the out-degree of an existing node  $i$  and to the

**Table 1**

Number of links ( $L$ ) and connectance ( $C = L/N^2$ ) for the two network sizes studied (number of nodes =  $N$ ).

$N$	100			
$L$	100	200	400	1000
$C$	0.01	0.02	0.04	0.1
$N$	500			
$L$	500	2500	5000	10 000
$C$	0.002	0.01	0.02	0.04

in-degree of an existing node  $j$  ( $i \neq j$ ). This step was repeated  $d$ -times.

- (5) The parameter  $e$  only added an arrow ( $i, j$ ) and not a node according to the total degrees of existing nodes  $i$  and  $j$  ( $i \neq j$ ). This step was repeated  $e$ -times.

Epidemic development was deterministic, with discrete time-step and governed by the probabilities of infection transmission between nodes ( $p_t$ ) and of infection persistence in a node ( $p_p$ ). The transmission probability  $p_t$  was either zero (unconnected nodes) or a value constant for different links in a network but variable amongst network replicates (in order to work at the threshold conditions in each replicate; see Pautasso and Jeger, 2008). The persistence probability  $p_p$  combined in one single parameter the length of infectiousness, detection and control measures. We also set  $p_p$  to be the same for all nodes. Both  $p_t$  and  $p_p$  are real variables, going from 0 to 1. This can be a realistic assumption for many ecological networks, wherever persistence and transmission are not either switched on or off, but can assume any value between these two extremes. We assumed all nodes to be of equal capacity and kind (differences between nodes are thus entirely due to their in- and out-degree). At each iteration, the contact structure of the network realization was maintained exactly the same. Networks were not necessarily fully connected, so it is possible that at the lower levels of connectance not all nodes could be reached from all nodes.

For each iteration, we obtained the infection status of a given node  $P_{i(x)}$  in the following way:

$$P_{i(x)} = \sum p_{t(x,y)} P_{i(y)}$$

for  $y$  going from 1 to 100, where  $p_t$  refers to the connection of the node  $x$  from a node  $y$ , and  $P_{i(y)}$  is the infection status of the node  $y$  at the previous iteration. At the beginning of the epidemic  $P_{i(x)}$  was set to zero for all nodes except for the starting node of the epidemic, with  $P_{i(i)} = 1$ . For the connection of a node with itself,  $p_p$  was used instead of  $p_t$ . The biological motivation for self-loops is that nodes which have become infected by a pathogen have a certain probability to remain infected due to the persistence of inoculum through time. The model was thus a susceptible–infected–susceptible (SIS) model. This can be a realistic assumption for many epidemiological systems, wherever nodes are still at risk even after eradication of a disease outbreak if complete immunization is not possible and if there is a continuing trade or contact with susceptible material or inoculum (Jeger et al., 2007).

The development of the epidemic was assessed on the basis of the sum of  $P_{i(x)}$  across all nodes and on the basis of the number of nodes with  $P_{i(x)}$  higher than an arbitrary value (0.01). The epidemic was started with a single infection of a single node, as the threshold conditions were not affected by whether epidemics are started with a single or with multiple infections (unpublished observations). Also, results were consistent using a different starting probability of infection. Although the starting node had a marked influence on the epidemic size at equilibrium (Pautasso and Jeger, 2008), making the epidemic start from different nodes did not affect the threshold conditions ( $p_p^*$  and  $p_t^*$ ) which define a boundary between no epidemic and an epidemic. Given that there is a linear threshold in a graph of  $p_p^*$  as a function of  $p_t^*$  (Pautasso and Jeger, 2008), we worked at  $p_p^* = 0$  and assessed the threshold only in terms of  $p_t^*$ .

The clustering coefficient  $C_i$  of a node  $i$  which is part of a digraph characterizes the extent to which nodes adjacent to any node  $i$  are adjacent to each other. More precisely,

$$C_i = \frac{\text{Card}(E_{\Gamma i})}{\text{Card}(V_{\Gamma i}) \times [\text{Card}(V_{\Gamma i}) - 1]}$$

where  $\text{Card}(X)$  symbolises the cardinality of the set  $X$ , i.e. the number of elements of  $X$ . The neighbourhood  $\Gamma_i$  of a vertex  $i$  is the

digraph that consists solely of the set  $V_{\Gamma i}$  of vertices connected from and/or to  $i$  (not including  $i$  itself) and of the set  $E_{\Gamma i}$  of all arrows connecting such vertices. The average clustering  $C$  of a digraph is the average of the clustering of each node of this digraph.

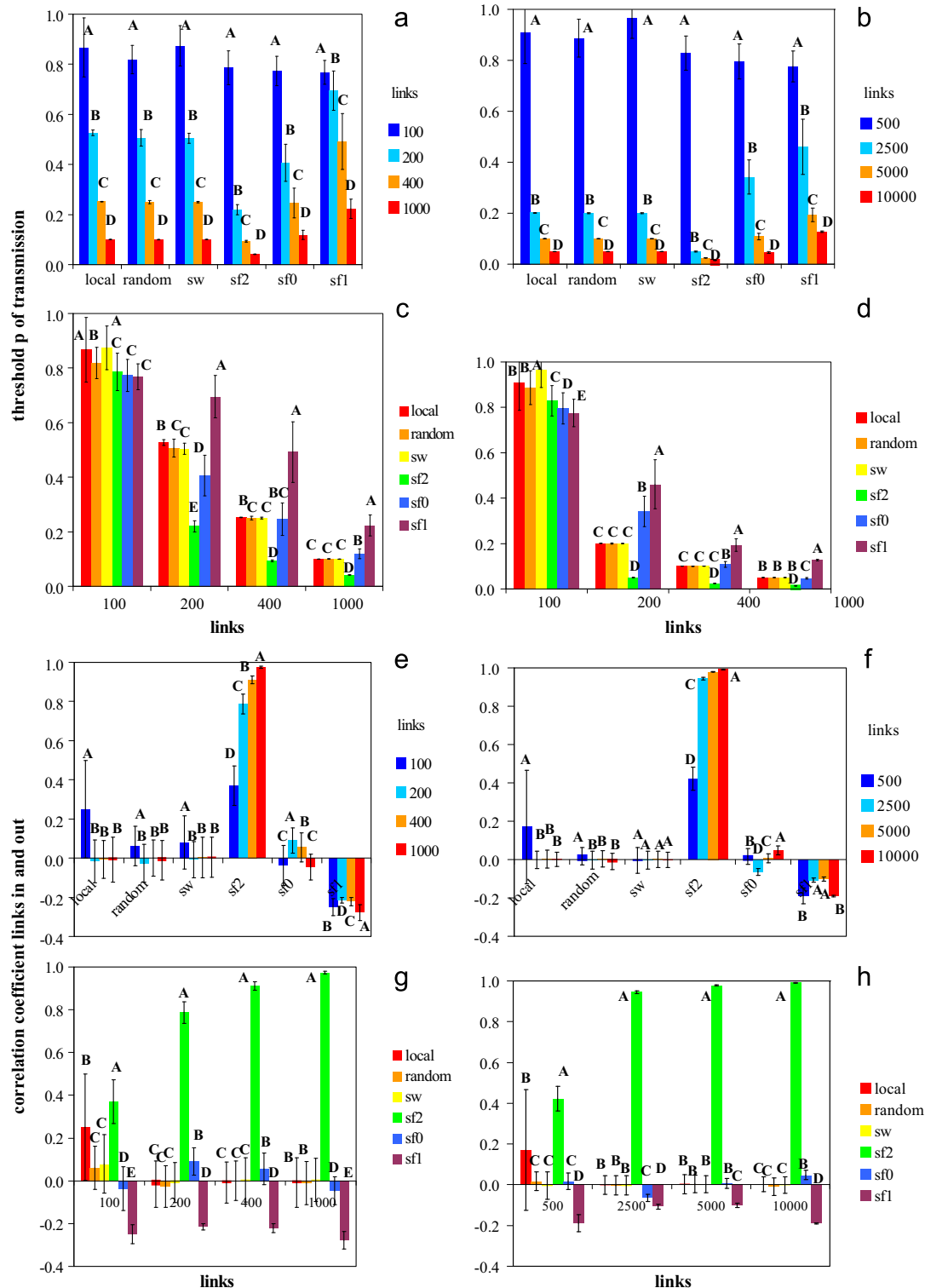
Analysis of variance (ANOVA) of the epidemic threshold and of the correlation coefficient between in- and out-degree of the 100 nodes of the network replicates for the different levels of connectance (within a network structure) and for the different types of network structure (at a given level of connectance) was carried out in SAS 9.1 (proc ANOVA). The same package (proc GLM) was used for multi-variate regressions of the threshold  $p_t^*$  against the correlation coefficient between in- and out-degree and the average clustering coefficient of the 100/500 nodes of the network replicate for a given network structure and level of connectance.

### 3. Results

The threshold  $p_t^*$  significantly decreased with increasing connectance for all structures and with both network sizes (Fig. 1a and b). With the exception of the lowest connectance level for both network sizes, two-way scale-free networks showed a significantly lower and one-way scale-free networks a significantly higher threshold than all other structures (Fig. 1c and d). For network size of 100 nodes, random networks showed a significantly lower threshold than local networks, but not at the highest connectance (Fig. 1c). This result was not confirmed at the larger network size, where random networks had an epidemic threshold not significantly different than the one of local networks (Fig. 1d). Small-world networks showed a threshold not significantly different from random networks (except at the lowest connectance for both network sizes). The threshold of uncorrelated scale-free networks, at all connectance levels and for both network sizes, lay between those for two-way and one-way scale-free networks (Fig. 1c and d).

The correlation coefficient between in- and out-degree of the nodes of the networks was, by definition, positive for two-way scale-free networks, not significantly different from zero for uncorrelated scale-free networks (except, by chance, at some levels of connectance; Fig. 1e and f), and negative for one-way scale-free networks. This coefficient was also clustered around zero for local, small-world and random networks. There was a significant increase in the correlation coefficient between in- and out-degree for two-way scale-free networks with increasing connectance for both network sizes (Fig. 1g and h).

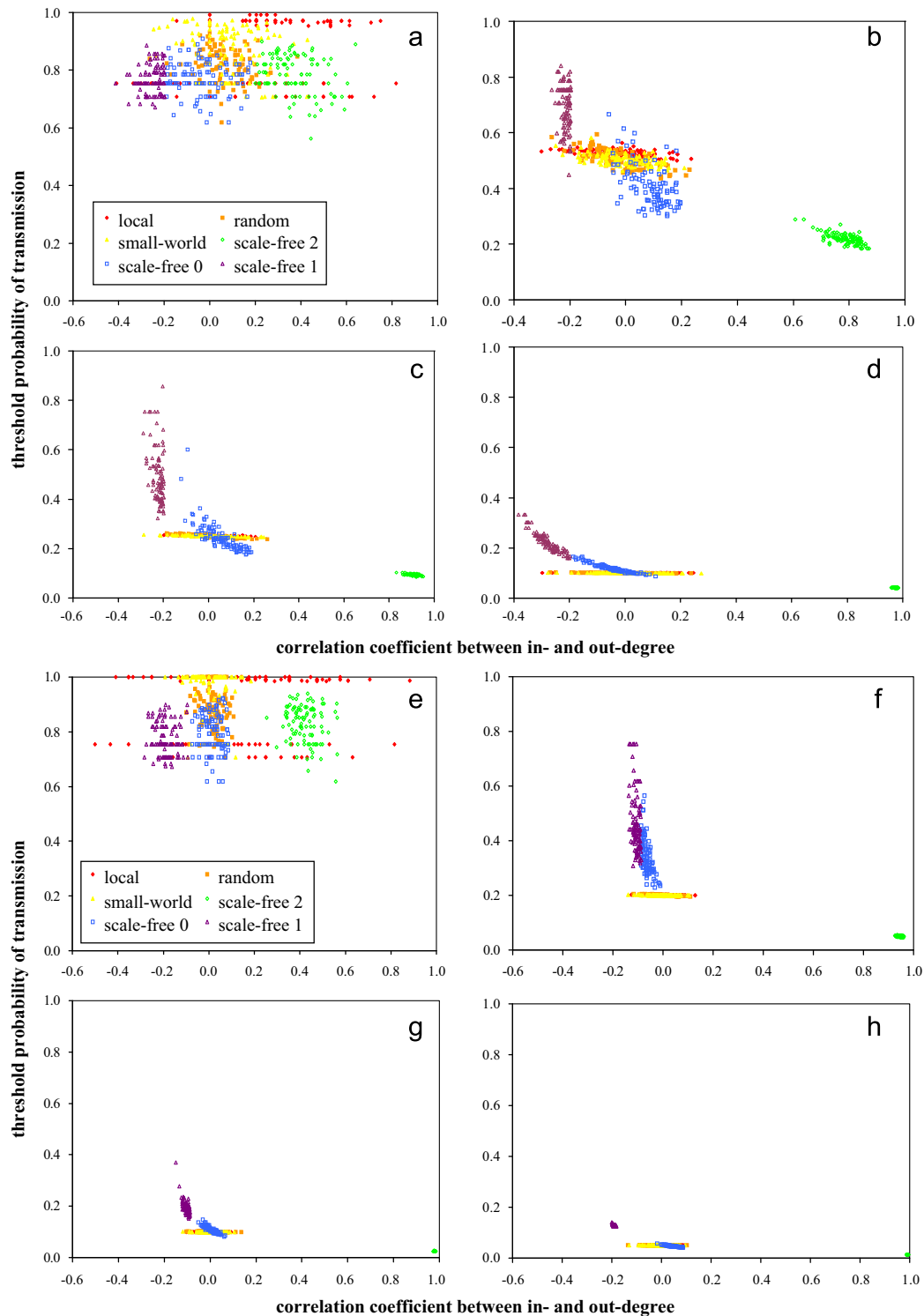
With the exception of the lowest connectance, there were generally significantly negative relationships between threshold  $p_t^*$  and the correlation coefficient between in- and out-degree for all structures and with both network sizes (Fig. 2; Tables 2 and 3). At the lowest level of connectance, these relationships were significant only for small-world and two-way scale-free networks for the network size of 100 nodes. The proportion of variance in threshold  $p_t^*$  explained by the correlation between the in- and the out-degree increased with connectance for all structures and for both network sizes, but with exception of two-way scale-free networks. Apart from one-way scale-free networks for 500 nodes and two-way scale-free networks for both network sizes, at the highest connectance these proportions of variance were substantial (between 0.84 and 0.93 for 100 nodes, and between 0.56 and 0.95 for 500 nodes). However, apart from one-way scale-free networks at the network size of 100 nodes, the slopes of these relationships tended to become flatter with increasing connectance (Tables 2 and 3). This is a consequence of the overall lower threshold  $p_t^*$  at higher connectance (Fig. 2).



**Fig. 1.** Threshold probability of transmission for the different (a, b) levels of connectance and (c, d) network structures, and correlation coefficient between in- and out-degree of the nodes of the 100 replicates for the different (e, f) levels of connectance and (g, h) network structures. Network size = 100 (a, c, e, g) and 500 (b, d, e, h) nodes. Error bars are standard deviations of 100 replicates. Different letters show significant differences (ANOVA,  $p < 0.05$ ) within (a, b, e, f) levels of connectance for a given network structure, and (c, d, g, h) network structures for a given level of connectance.

Including the clustering as an additional explaining factor in the regression of the threshold  $p_t$  against the correlation coefficient between links to and from nodes did not affect the previous results. The clustering increased with increasing level of connectance for all structures and for both network sizes (Figs. 3

and 4). By definition, at a given connectance, the clustering was higher in local than in small-world networks, and in small-world compared to random networks (Figs. 3 and 4). The clustering also decreased from two-way to uncorrelated and to one-way scale-free networks (Figs. 3 and 4). For both network sizes, the



**Fig. 2.** Threshold probability of transmission for the different network replicates as a function of the correlation coefficient between in- and out-degree of the nodes for the six network structures. Network size = 100 nodes (number of links = (a) 100, (b) 200, (c) 400, and (d) 1000) and 500 nodes (number of links = (e) 500, (f) 2500, (g) 5000, and (h) 10 000).

clustering was significantly negatively related to the threshold  $p_t$  at the lowest level of connectance for local and small-world networks, and at all levels of connectance for two way scale-free networks (Fig. 3; Tables 2 and 3).

For both network sizes, the relationship of the correlation coefficient between links to and from nodes with the clustering was significantly positive for all scale-free networks (except at the

lowest level of connectance; Fig. 4; Tables 4 and 5). However, the proportion of variance in the correlation coefficient between in- and out-degree explained by the clustering was substantial for uncorrelated and one-way, but not for two-way scale-free networks. For both network sizes, there was also a significantly positive correlation between these two variables for random networks at the two highest connectance levels (Tables 4 and 5).



**Table 2**

Proportion of variance explained, slope and associated *p* value of the regression of threshold probability of transmission as a function of (a) the correlation coefficient between in- and out-degree and (b) the clustering for the 100 replicates of the six network structures (local, random, small-world, two-way, uncorrelated, and one-way scale-free) at four levels of connectance (100, 200, 400, and 1000 links; network size = 100 nodes).

	100			200			400			1000		
	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>
<i>l</i>	0.17	+0.05	−5.45	0.33	−0.06	+0.09	0.58	−0.01	+0.03	0.85	−0.001	+0.000
	<i>p</i>	0.30	0.001		0.001	0.15		0.001	0.03		0.001	0.98
<i>r</i>	0.02	−0.07	−0.76	0.54	−0.25	+0.04	0.74	−0.06	+0.02	0.86	−0.009	−0.003
	<i>p</i>	0.23	0.41		0.001	0.89		0.001	0.79		0.001	0.86
sw	0.22	−0.16	−4.63	0.27	−0.11	−0.02	0.62	−0.02	−0.00	0.84	−0.002	+0.001
	<i>p</i>	0.003	0.001		0.001	0.86		0.001	0.82		0.001	0.45
sf2	0.11	−0.17	−3.24	0.55	−0.23	−0.30	0.34	−0.09	−0.02	0.25	−0.04	−0.01
	<i>p</i>	0.008	0.04		0.001	0.001		0.001	0.05		0.001	0.001
sf0	0.01	−0.04	−0.90	0.28	−0.59	−0.27	0.59	−0.69	+0.28	0.93	−0.27	+0.02
	<i>p</i>	0.49	0.48		0.001	0.61		0.001	0.33		0.001	0.45
sf1	0.04	+0.12	−1.52	0.03	−0.86	+1.66	0.23	−3.31	+4.09	0.87	−0.92	+0.09
	<i>p</i>	0.27	0.15		0.12	0.24		0.001	0.005		0.001	0.35

**Table 3**

Proportion of variance explained, slope and associated *p* value of the regression of threshold probability of transmission as a function of (a) the correlation coefficient between in- and out-degree and (b) the clustering for the 100 replicates of the six network structures (local, random, small-world, two-way, uncorrelated, and one-way scale-free) at four levels of connectance (500, 2500, 5000, and 10 000 links; network size = 500 nodes).

	500			2500			5000			10 000		
	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>	$r^2$	<i>a</i>	<i>b</i>
<i>l</i>	0.15	+0.00	−32.7	0.66	−0.01	−0.00	0.48	−0.001	−0.001	0.56	−0.0001	−0.0001
	<i>p</i>	0.97	0.001		0.001	0.63		0.001	0.68		0.001	0.25
<i>r</i>	0.02	−0.21	+3.43	0.77	−0.04	−0.02	0.87	−0.01	−0.03	0.94	−0.002	−0.003
	<i>p</i>	0.21	0.50		0.001	0.89		0.001	0.31		0.001	0.66
sw	0.24	−0.13	−26.7	0.66	−0.01	+0.00	0.68	−0.002	−0.003	0.87	−0.0004	−0.0001
	<i>p</i>	0.20	0.001		0.001	0.85		0.001	0.15		0.001	0.76
sf2	0.08	−0.20	−18.1	0.50	−0.06	−0.09	0.60	−0.03	−0.02	0.38	−0.04	−0.004
	<i>p</i>	0.06	0.03		0.001	0.001		0.02	0.001		0.001	0.001
sf0	0.01	−0.12	−6.20	0.41	−3.66	+5.74	0.83	−0.64	+0.44	0.95	−0.16	+0.07
	<i>p</i>	0.48	0.44		0.001	0.01		0.001	0.001		0.001	0.001
sf1	0.01	−0.12	−5.11	0.15	−4.16	+2.59	0.55	−2.28	+0.75	0.40	−0.63	+0.11
	<i>p</i>	0.38	0.49		0.001	0.33		0.001	0.07		0.001	0.07

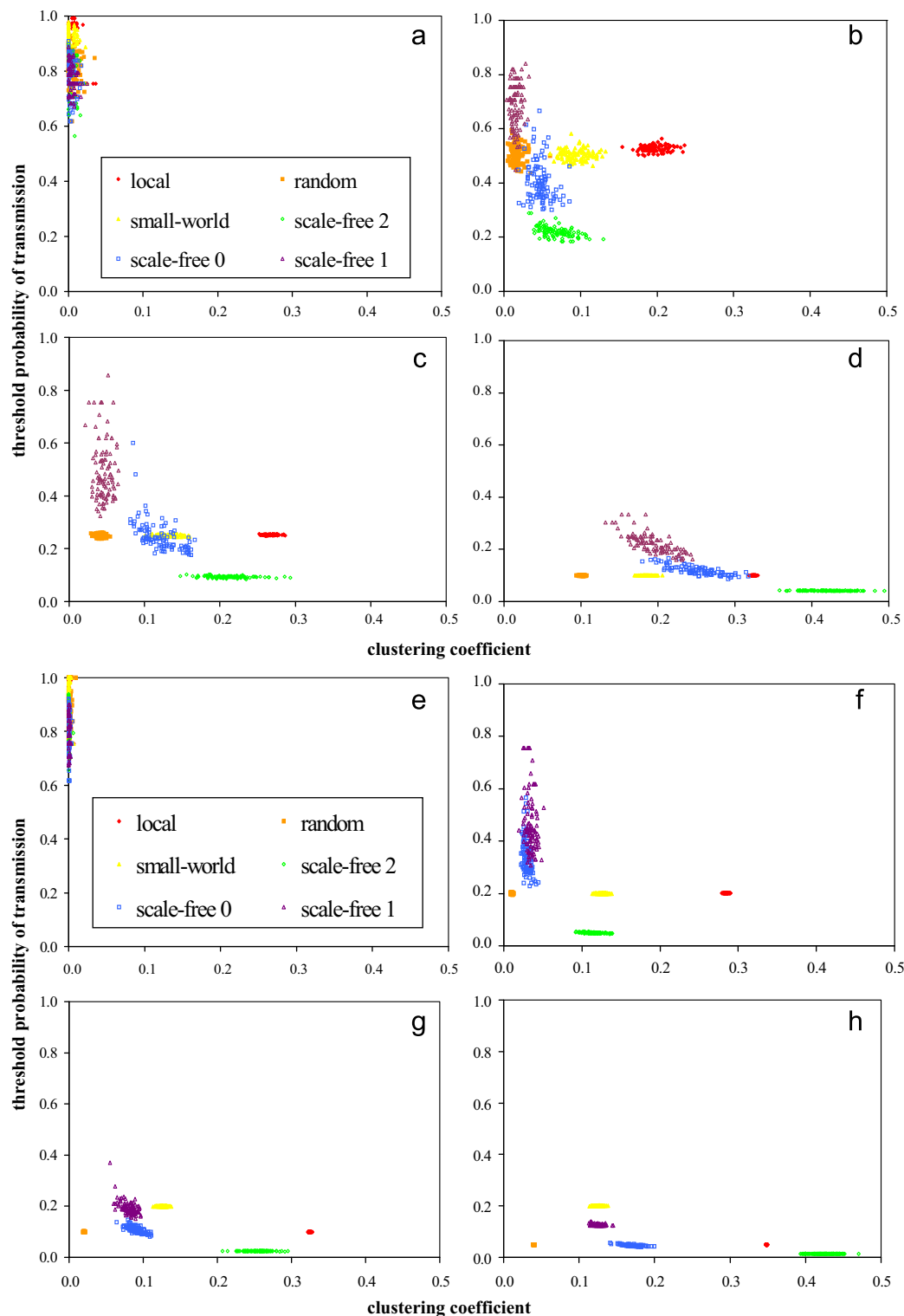
#### 4. Discussion

Networks of small size have biological significance in a variety of ecological fields. Examples include meta-populations, mutualistic, and antagonistic interactions (Dunne et al., 2002; Lundgren and Olesen, 2005; Brooks, 2006; Pautasso et al., 2008; Thebault and Fontaine, 2008). In spite of the relevance of small-size networks for many issues in natural sciences, it is not clear whether theoretical results derived from analyses of large-scale complex networks apply also to small-size networks (Guimarães et al., 2007). Moreover, much work in network epidemiology has focused on undirected networks (e.g. Keeling, 2005; Shirley and Rushton, 2005; May, 2006), whereas many of the biological networks mentioned in Introduction are inherently asymmetrical and thus can only be realistically modelled with directed networks.

Our analysis shows that in SIS models of epidemics in directed networks of small size, the threshold is lower for scale-free network structures only if there is a positive correlation between in- and out-degree of nodes (Fig. 2). This finding is in good agreement with results obtained in percolation theory for large-size networks (Schwartz et al., 2002). Our work further shows that

when this correlation between in- and out-degree of nodes is negative (one-way scale-free networks), for small-size networks the epidemic threshold is higher than in non-scale-free networks. This result is broadly independent of the connectance level and of the network size, although it breaks down in case of sparsely connected networks. For sparsely connected networks, differences in clustering amongst networks can become important.

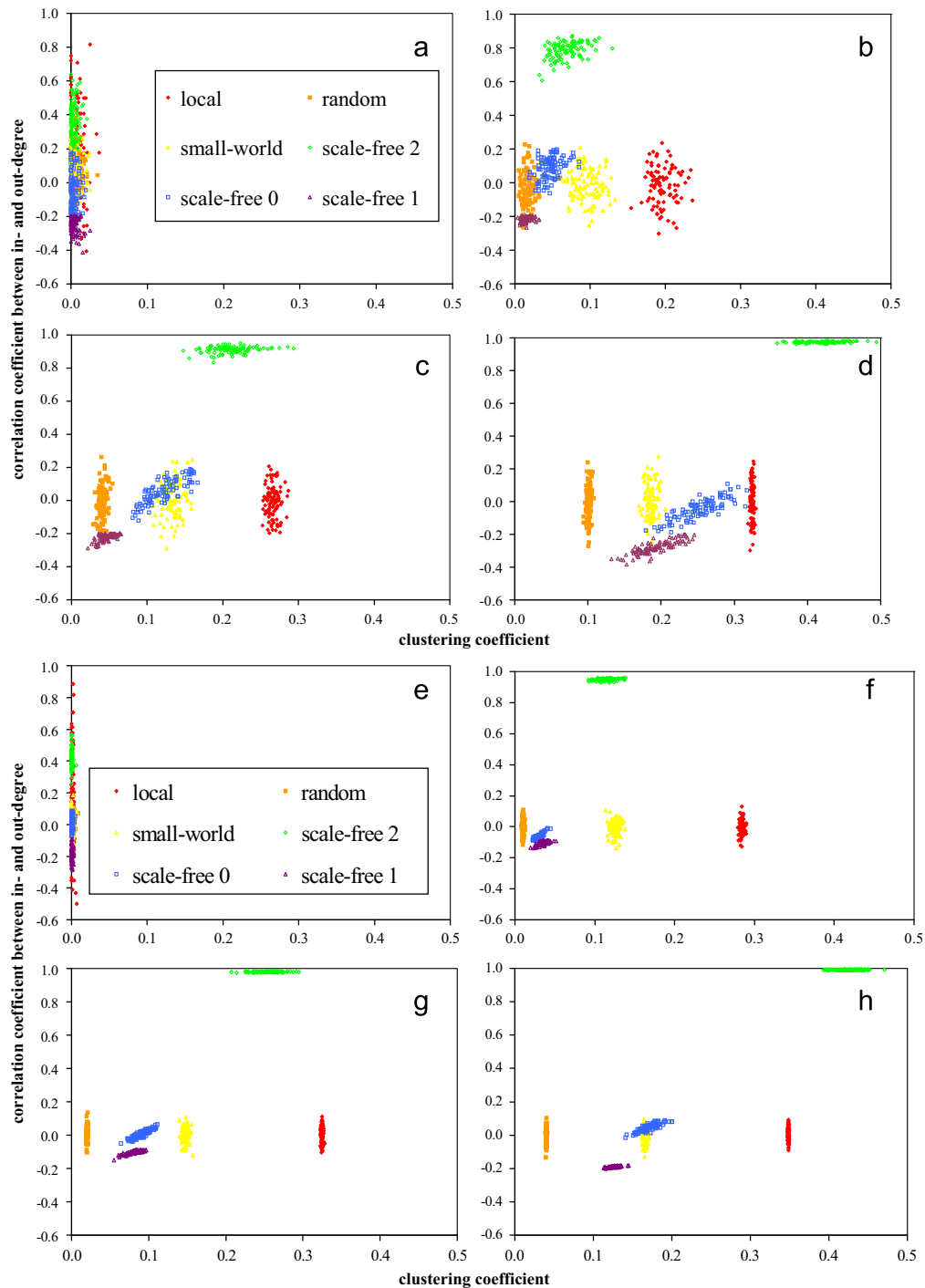
Clustering has been shown to be a network feature relevant to the development and control of epidemics (e.g. Eames and Keeling, 2003; Keeling, 2005; Kiss et al., 2005; Hartvigsen et al., 2007; Turner et al., 2008; Miller, 2009). More clustered networks are believed to be less prone to invasion by a pathogen, as this will be likely to become confined inside clusters of connected nodes (Keeling, 2005; Naug, 2008). Clusters can also slow down epidemic development in the first phases of epidemics (e.g. Szendroi and Csanyi, 2004). In scale-free networks of infinite size, the presence of high local clustering has been shown to make it possible for a non-null epidemic threshold to be present (Eguiluz and Klemm, 2002; but see Serrano and Boguna, 2006). Clustering in large-scale scale-free networks has been shown to decrease the size of epidemics, but also to decrease their threshold (Newman, 2003).



**Fig. 3.** Threshold probability of transmission for the different network replicates as a function of the average clustering coefficient of the nodes for the six network structures. Network size = 100 nodes (number of links = (a) 100, (b) 200, (c) 400, and (d) 1000) and 500 nodes (number of links = (e) 500, (f) 2500, (g) 5000, and (h) 10000).

In the case of our small-size networks, the influence of clustering on the epidemic threshold is dwarfed by the one of the correlation coefficient between in- and out-degree. The negative correlation between epidemic threshold and clustering for uncorrelated and one-way scale-free networks at the highest levels of connectance (Fig. 3) can be explained by the positive correlation of the correlation coefficient between in- and out-degree with the

clustering in the same networks (Fig. 4). Unlike Eames (2008) for undirected networks, we do not observe any differences in the influence of clustering on epidemic development depending on whether contacts are regular or random. In our small-size, directed networks, with the exception of sparsely connected local networks, clustering has no influence on the epidemic threshold within both local and random networks (Fig. 3).



**Fig. 4.** Correlation coefficient between in- and out-degree for the different network replicates as a function  $n$  of the average clustering coefficient of the 100 nodes for the six network structures. Network size = 100 nodes (number of links = (a) 100, (b) 200, (c) 400, and (d) 1000) and 500 nodes (number of links = (e) 500, (f) 2500, (g) 5000, and (h) 10 000).

Our main conclusion is that, in directed networks, analyses of the influence of clustering on the epidemic threshold can be spurious if they do not consider simultaneously the effect of the correlation coefficient between in- and out-degree. In some cases (sparsely connected local, small-world, and two-way scale free networks), we observe a negative correlation between epidemic threshold and clustering and no significant effect of the correlation coefficient between in- and out-degree on the threshold, suggesting that in extreme situations the correlation coefficient between in- and out-degree may not play the same role as for large-scale networks. We also point out that analyses which show

that clustering is negatively related to the epidemic threshold (e.g. Britton et al., 2008) need to take the connectance level into account, as (i) the clustering increases with increasing connectance, (ii) the epidemic threshold decreases with increasing connectance, so that (iii) if connectance increases the epidemic threshold is bound to be negatively correlated with the clustering. Our results show, however, that, when keeping connectance constant, clustering and epidemic threshold are not necessarily related.

Epidemics in scale-free networks are now commonly investigated (e.g. Boccaletti et al., 2006; Masuda and Konno, 2006;



**Table 4**

Proportion of variance explained, slope and associated  $p$  value of the regression of the correlation coefficient between in- and out-degree as a function of the clustering for the 100 replicates of the six network structures (local, random, small-world, two-way, uncorrelated, and one-way scale-free) at four levels of connectance (100, 200, 400, and 1000 links; network size = 100 nodes).

	100			200			400			1000		
	$r^2$	slope	$p$	$r^2$	slope	$p$	$r^2$	slope	$p$	$r^2$	slope	$p$
$l$	0.06	−7.40	0.01	0.00	−0.01	0.99	0.00	+0.26	0.85	0.00	−0.93	0.88
$r$	0.00	−0.65	0.69	0.01	+0.14	0.37	0.07	+5.37	0.008	0.06	+10.1	0.01
sw	0.00	−0.84	0.68	0.01	−0.49	0.41	0.03	+1.55	0.09	0.00	+0.33	0.83
sf2	0.00	+1.46	0.56	0.18	+1.13	0.001	0.04	+0.16	0.04	0.06	+0.05	0.01
sf0	0.00	−0.03	0.98	0.13	+1.81	0.001	0.65	+2.61	0.001	0.70	+1.83	0.001
sf1	0.05	−2.02	0.03	0.08	+0.71	0.005	0.42	+1.50	0.001	0.69	+1.37	0.001

**Table 5**

Proportion of variance explained, slope, and associated  $p$  value of the regression of the correlation coefficient between in- and out-degree as a function of the clustering for the 100 replicates of the six network structures (local, random, small-world, two-way, uncorrelated, and one-way scale-free) at four levels of connectance (500, 2500, 5000, and 10 000 links; network size = 500 nodes).

	500			2500			5000			10 000		
	$r^2$	slope	$p$	$r^2$	slope	$p$	$r^2$	slope	$p$	$r^2$	slope	$p$
$l$	0.05	−45.5	0.03	0.00	+1.61	0.48	0.00	−3.40	0.46	0.00	−4.90	0.52
$r$	0.01	+3.08	0.32	0.02	+7.42	0.21	0.09	+22.4	0.002	0.06	+25.0	0.01
sw	0.00	+2.67	0.59	0.00	+0.05	0.96	0.00	+0.02	0.98	0.00	−0.84	0.60
sf2	0.00	−0.37	0.96	0.08	+0.19	0.005	0.09	+0.04	0.002	0.01	+0.01	0.33
sf0	0.00	+2.82	0.55	0.73	+3.33	0.001	0.79	+2.18	0.001	0.71	+1.71	0.001
sf1	0.00	−1.63	0.75	0.54	+1.60	0.001	0.74	+1.10	0.001	0.55	+0.45	0.001

Colizza and Vespignani, 2008), but there has been less attention to directed scale-free networks. In this analysis, we confirm the importance of the correlation coefficient between in- and out-degree for epidemics in directed scale-free networks (Woolhouse et al., 2005). We also provide evidence for a different behaviour of the clustering in directed scale-free networks where there are different correlation coefficients between in- and out-degree. Although there is a positive correlation of the clustering with the correlation coefficient between in- and out-degree for all scale-free network types (except at the lowest connectance level), only for uncorrelated and one-way scale-free networks has this correlation a substantial  $r^2$  (for 500 nodes: between 0.54 and 0.79, Table 5). For two-way scale-free networks, the correlation between clustering and correlation coefficient between links in and out of nodes is significant but negligible (again for 500 nodes,  $r^2$  values are between 0.01 and 0.08, Table 5). Further work is needed to investigate whether results obtained for undirected scale-free networks apply to all types of directed scale-free networks.

Our results are essentially independent of the network size used (100 and 500 nodes). We have avoided using a lower number of nodes than 100 as it is likely that differences between network structures will tend to become blurred for even smaller networks. However, an interesting question would be at which small size heterogeneity in the contact structure stops having a significant influence on network properties and epidemic processes. A remarkable result is also that network breakdown at the lowest connectance level (which causes for example the disappearance of a significant relationship between epidemic threshold and correlation coefficient between in- and out-degree) happens at two different connectance levels (0.01 for 100 nodes and 0.002 for 500 nodes, Table 1) for the two network sizes. More research is needed to assess which other processes are not independent of the interaction between network size and connectance.

In conclusion, our study confirms the result obtained for large-size networks that regardless of the size of the network, the epidemic threshold is negatively related to the correlation

coefficient between in- and out-degree for all structures. However, this does not hold if small-size networks are sparsely connected. In this case, clustering plays a significant role (Table 2). In small-size networks, when the correlation between links in and out is negative (one-way scale-free networks), the epidemic threshold can be higher than in non-scale-free networks. Contrary to what is found in analyses of large-size networks, clustering does not necessarily have an influence on the epidemic threshold of small-size networks if connectance is kept constant. Analyses of the influence of the clustering on the epidemic threshold in directed networks can also be spurious if the effect of the correlation coefficient between in- and out-degree is disregarded.

## Acknowledgements

Many thanks to T. Harwood, O. Holdenrieder, J. Parke, M. Shaw, J. Tufto, F. Van den Bosch, and X. Xu for discussions and insights, and to T. Matoni and anonymous reviewers for helpful comments on a previous version of the ms. This study was funded by the Department for Environment, Food and Rural Affairs, the Rural Economy and Land Use Programme, UK, and the French Ministry of National Education and Research.

## References

- Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R., Daszak, P., 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19, 535–544.
- Bandyopadhyay, B., Frederiksen, R.A., 1999. Contemporary global movement of emerging plant diseases. *Ann. New York Acad. Sci.* 894, 28–36.
- Bascompte, J., 2007. Networks in ecology. *Basic Appl. Ecol.* 8, 485–490.
- Bascompte, J., Jordano, P., Melian, C.J., Olesen, J.M., 2003. The nested assembly of plant–animal mutualistic networks. *Proc. Natl. Acad. Sci. USA* 100, 9383–9387.
- Blanc, L.A., Walters, J.R., 2007. Cavity-nesting community webs as predictive tools: where do we go from here? *J. Ornithol.* 148, S417–S423.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U., 2006. Complex networks: structure and dynamics. *Phys. Rep.* 424 (4–5), 175–308.

- Bode, M., Burrage, K., Possingham, H.P., 2008. Using complex network metrics to predict the persistence of metapopulations with asymmetric connectivity patterns. *Ecol. Model.* 214, 201–209.
- Boguna, M., Serrano, M.A., 2005. Generalized percolation in random directed networks. *Phys. Rev. E* 72, 016106.
- Boguna, M., Pastor-Satorras, R., Vespignani, A., 2004. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B* 38, 205–209.
- Brenn, N., Menkis, A., Grünig, C.R., Sieber, T.N., Holdenrieder, O., 2008. Community structure of *Phialocephala fortinii* s. lat. in European tree nurseries, and assessment of the potential of the seedlings as dissemination vehicles. *Mycol. Res.* 112, 650–662.
- Britton, T., Deijfen, M., Lageras, A.N., Lindholm, M., 2008. Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* 45, 743–756.
- Brooks, C.P., 2006. Quantifying population substructure: extending the graph-theoretic approach. *Ecology* 87, 864–872.
- Brooks, C.P., Antonovics, J., Keitt, T.H., 2008. Spatial and temporal heterogeneity explain disease dynamics in a spatially explicit network model. *Am. Nat.* 172, 149–159.
- Burns, K.C., 2007. Network properties of an epiphyte metacommunity. *J. Ecol.* 95, 1142–1151.
- Colizza, V., Vespignani, A., 2007. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.* 99, 148701.
- Colizza, V., Vespignani, A., 2008. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* 251 (3), 450–467.
- da Gama, M.M.T., Nunes, A., 2006. Epidemics in small world networks. *Eur. Phys. J. B* 50, 205–208.
- Dehnen-Schmutz, K., Touza, J., Perrings, C., Williamson, M., 2007. A century of the ornamental plant trade and its impact on invasion success. *Divers. Dist.* 13, 527–534.
- Dunbar, R.I.M., 1993. Coevolution of neurological size, group size and language in humans. *Behav. Brain Sci.* 16, 681–735.
- 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, 12917–12922.
- Eames, K.T.D., 2008. Modelling disease spread through random and regular contacts in clustered populations. *Theor. Pop. Biol.* 73, 104–111.
- Eames, K.T.D., Keeling, M.J., 2003. Contact tracing and disease control. *Proc. R. Soc. London B* 270, 2565–2571.
- Eguiluz, V.M., Klemm, K., 2002. Epidemic threshold in structured scale-free networks. *Phys. Rev. E* 89, 108701.
- Frankel, S.J., 2008. Sudden oak death and *Phytophthora ramorum* in the USA: a management challenge. *Aust. Plant Path.* 24, 282–284.
- Grünwald, N.J., Goss, E.M., Press, C.M., 2008. *Phytophthora ramorum*: a pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Mol. Plant Path.* 9, 729–740.
- Guimarães, P.R., de Menezes, M.A., Baird, R.W., Lusseau, D., Guimaraes, P., dos Reis, S.F., 2007. Vulnerability of a killer whale social network to disease outbreaks. *Phys. Rev. E* 76, 042901.
- Hanski, I., Ovaskainen, O., 2000. The metapopulation capacity of a fragmented landscape. *Nature* 404, 755–758.
- Hartvigsen, G., Dresch, J.M., Zielinski, A.L., Macula, A.J., Leary, C.C., 2007. Network structure, and vaccination strategy and effort interact to affect the dynamics of influenza epidemics. *J. Theor. Biol.* 246, 205–213.
- Holdenrieder, O., Pautasso, M., Weisberg, P.J., Lonsdale, D., 2004. Tree diseases and landscape processes: the challenge of landscape pathology. *Trends Ecol. Evol.* 19, 446–452.
- Ings, T.C., Montoya, J.M., Bascompte, J., Bluethgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.L., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M., Van Veen, F.F.J., Warren, P.H., Woodward, G., 2009. Ecological networks—beyond food webs. *J. Anim. Ecol.* 78, 253–269.
- Jeger, M.J., Pautasso, M., Holdenrieder, O., Shaw, M.W., 2007. Modelling disease spread and control in networks: implications for plant sciences. *New Phytol.* 174, 279–297.
- Jones, D.R., Baker, R.H.A., 2007. Introductions of non-native plant pathogens into Great Britain, 1970–2004. *Plant Pathol.* 56, 891–910.
- Kao, R.R., Danon, L., Green, D.M., Kiss, I.Z., 2006. Demographic structure and pathogen dynamics of the network of livestock movements in Great Britain. *Proc. R. Soc. London B* 273, 1999–2007.
- Keeling, M.J., 2005. The implications of network structure for epidemic dynamics. *Theor. Pop. Biol.* 67, 1–8.
- Kenah, E., Robins, J.M., 2007. Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. *J. Theor. Biol.* 249, 706–722.
- Kiss, I.Z., Green, D.M., Kao, R.R., 2005. Disease contact tracing in random and clustered networks. *Proc. R. Soc. London B* 272, 1407–1414.
- Lamour, A., Termorshuizen, A.J., Volker, D., Jeger, M.J., 2007. Network formation by rhizomorphs of *Armillaria lutea* in natural soil: their description and ecological significance. *FEMS Microb. Ecol.* 62, 222–232.
- Liu, J.Z., Wu, J.S., Yang, Z.R., 2004. The spread of infectious disease on complex networks with household-structure. *Physica A* 341, 273–280.
- Lobel, S., Snäll, T., Rydin, H., 2006. Metapopulation processes in epiphytes inferred from patterns of regional distribution and local abundance in fragmented forest landscapes. *J. Ecol.* 94, 856–868.
- Lundgren, R., Olesen, J.M., 2005. The dense and highly connected world of Greenland's plants and their pollinators. *Arct. Antarct. Alp. Res.* 37, 514–520.
- Masuda, N., Konno, N., 2006. Multi-state epidemic processes on complex networks. *J. Theor. Biol.* 243, 64–75.
- May, R.M., 2006. Network structure and the biology of populations. *Trends Ecol. Evol.* 21, 394–399.
- Meyers, L.A., Newman, M.E.J., Pourbohloul, B., 2006. Predicting epidemics on directed contact networks. *J. Theor. Biol.* 240, 400–418.
- Miller, J.C., 2009. Spread of infectious disease through clustered populations. *J. Roy. Soc. Interface*, doi:10.1098/rsif.2008.0524, in press.
- Montoya, J.M., Solé, R.V., 2002. Small world patterns in food webs. *J. Theor. Biol.* 214, 405–412.
- Moore, C., Newman, M.E.J., 2000. Epidemics and percolation in small-world networks. *Phys. Rev. E* 61, 5678–5682.
- Naug, D., 2008. Structure of the social network and its influence on transmission dynamics in a honeybee colony. *Behav. Ecol. Sociobiol.* 62, 1719–1725.
- Neutel, A.M., Heesterbeek, J.A.P., van de Koppel, J., Hoenderboom, G., Vos, A., Kaldeewey, C., Berendse, F., de Ruiter, P.C., 2007. Reconciling complexity with stability in naturally assembling food webs. *Nature* 449, 599–602.
- Newman, M.E.J., 2003. Properties of highly clustered networks. *Phys. Rev. E* 68, 026121.
- Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2001. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64, 026118.
- Nielsen, A., Bascompte, J., 2007. Ecological networks, nestedness and sampling effort. *J. Ecol.* 95, 1134–1141.
- Olesen, J.M., Bascompte, J., Dupont, Y.L., Jordano, P., 2006. The smallest of all worlds: pollination networks. *J. Theor. Biol.* 240, 270–276.
- Otterstatter, M.C., Thomson, J.D., 2007. Contact networks and transmission of an intestinal pathogen in bumble bee (*Bombus impatiens*) colonies. *Oecologia* 154, 411–421.
- Park, S.M., Kim, B.J., 2006. Dynamic behaviors in directed networks. *Phys. Rev. E* 74, 026114.
- Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200–3203.
- Pautasso, M., Jeger, M.J., 2008. Epidemic threshold and network structure: the interplay of probability of transmission and of persistence in small-size directed networks. *Ecol. Compl.* 5, 1–8.
- Pautasso, M., Harwood, T., Shaw, M.W., Xu, X., Jeger, M.J., 2008. Epidemiological modeling of *Phytophthora ramorum*: network properties of susceptible plant genera movements in the nursery sector of England and Wales. In: *Proceedings of Sudden Oak Death III Science Symposium*, PSW-GTR-214, USDA Forest Service, pp. 257–264.
- Pellis, L., Ferguson, N.M., Fraser, C., 2009. Threshold parameters for a model of epidemic spread among households and workplaces. *J. R. Soc. Interface*, doi:10.1098/rsif.2008.0493, in press.
- Prospero, S., Hansen, E.M., Grünwald, N.J., Winton, L.M., 2007. Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Mol. Ecol.* 16, 2958–2973.
- Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353.
- Ryder, T.B., McDonald, D.B., Blake, J.G., Parker, P.G., Loiselle, B.A., 2008. Social networks in the lek-mating wire-tailed manakin (*Pipra filicauda*). *Proc. R. Soc. London B* 275, 1367–1374.
- Schwartz, N., Cohen, R., ben-Avraham, D., Barabasi, A.L., Havlin, S., 2002. Percolation in directed scale-free networks. *Phys. Rev. E* 66, 015104.
- Serrano, M.A., Boguna, M., 2006. Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.* 97, 088701.
- Shirley, M.D.F., Rushton, S.P., 2005. The impact of network topology on disease spread. *Ecol. Compl.* 2, 287–299.
- Slippers, B., Stenlid, J., Wingfield, M.J., 2005. Emerging pathogens: fungal host jumps following anthropogenic introduction. *Trends Ecol. Evol.* 20, 420–421.
- Southworth, D., He, X.H., Swenson, W., Bledsoe, C.S., Horwath, W.R., 2005. Application of network theory to potential mycorrhizal networks. *Mycorrhiza* 15, 589–595.
- Sun, H.J., Gao, Z.Y., 2007. Dynamical behaviors of epidemics on scale-free networks with community structure. *Physica A* 381, 491–496.
- Szendroi, B., Csanyi, G., 2004. Polynomial epidemics and clustering in contact networks. *Proc. R. Soc. London B* 271, S364–S366.
- Thebault, E., Fontaine, C., 2008. Does asymmetric specialization differ between mutualistic and trophic networks? *Oikos* 117, 555–563.
- Turner, J., Bowers, R.G., Clancy, D., Behnke, M.C., Christley, R.M., 2008. A network model of *E. coli* O157 transmission within a typical UK dairy herd: the effect of heterogeneity and clustering on the prevalence of infection. *J. Theor. Biol.* 254, 45–54.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.
- Werres, S., Marwitz, R., Veld, W.A.M.I., De Cock, A.W.A.M., Bonants, P.J.M., De Weerd, M., Themann, K., Ilieva, E., Baayen, R.P., 2001. *Phytophthora ramorum* sp. nov., a new pathogen on *Rhododendron* and *Viburnum*. *Mycol. Res.* 105, 1155–1165.
- Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Liu, W.C., Mellor, D.J., Thomas, M.R., 2005. Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* 1, 350–352.
- Xu, X.M., Harwood, T.D., Pautasso, M., Jeger, M.J., 2009. Spatio-temporal analysis of *Phytophthora ramorum* cases in England and Wales (2003–2006). *Ecography* 32, 504–516.