How clustering affects the bond percolation threshold in complex networks

James P. Gleeson, Sergey Melnik, and Adam Hackett
Department of Mathematics & Statistics, University of Limerick, Limerick, Ireland
(Received 21 December 2009; revised manuscript received 16 April 2010; published 18 June 2010)

The question of how clustering (nonzero density of triangles) in networks affects their bond percolation threshold has important applications in a variety of disciplines. Recent advances in modeling highly clustered networks are employed here to analytically study the bond percolation threshold. In comparison to the threshold in an unclustered network with the same degree distribution and correlation structure, the presence of triangles in these model networks is shown to lead to a larger bond percolation threshold (i.e. clustering *increases* the epidemic threshold or *decreases* resilience of the network to random edge deletion).

DOI: 10.1103/PhysRevE.81.066114 PACS number(s): 89.75.Hc, 64.60.aq, 64.60.ah, 87.23.Ge

I. INTRODUCTION

Clustering (or *transitivity*) in a complex network refers to the propensity of two neighbors of a given node to also be neighbors of each other, thus forming a triangle of edges within the graph. In a recent paper [1], Newman proposes a model of random networks with clustering which permits analytical solution for many important properties. An alternative model, based on embedding cliques in a locally tree-like structure, was subsequently proposed by one of us [2]. One of the most important predictions of these models is the effect of network clustering on the bond percolation process, which is a topic of considerable interest [3–12].

The bond percolation problem for a network may be stated as follows: each edge of the network graph is visited once, and *damaged* (deleted) with probability $1-p$. The quantity p is the *bond occupation probability* and the non-damaged edges are termed *occupied*. In an infinite graph, the size of the *giant connected component* (GCC) of the graph becomes nonzero at some critical value of $p>0$: this critical value of p is termed the *bond percolation threshold*, denoted p_{th}. The bond percolation problem has applications in epidemiology, where p is related to the average transmissibility of a disease and the GCC represents the size of an epidemic outbreak [13,14], and in the analysis of technological networks, where the resilience of a network to the random failure of links is quantified by the size of the GCC [7]. Analytical solutions for percolation on randomly wired networks and on correlated networks are well known [15–20], but these cases have zero clustering in the limit of infinite network size.

Newman solves the bond percolation problem within his model [1] and considers the effect of clustering on the bond percolation threshold. He gives an example where clustering decreases the value of p_{th} within the context of a certain set of networks which all share the same average degree (see Fig. 2 of [1]). However, Newman notes that the networks in his comparison set, while having the same average degree, do not all have the same degree distribution (see Sec. III for further discussion of this point). Miller [21] recently showed analytically that within the model [1] the bond percolation threshold in a clustered network is greater than the corresponding threshold in an unclustered network with the same degree distribution and correlation structure. A similar conclusion was reported by Kiss and Green [10] based on their numerical simulations using Newman's clustered bipartite graph model [3]. In this paper we focus on networks generated by the clique-based model [2] and show that the effect of clustering is qualitatively similar to that determined by Miller for the triangle-based model [1], i.e., the presence of clustering *increases* the bond percolation threshold (and hence the epidemic threshold) when networks with the same degree distribution and correlation structure are compared. We emphasize that the degree-degree correlation structure in the clustered network includes nontrivial correlations beyond nearest neighbors, and we consider the implications of this fact.

We begin by introducing the recently published models for clustered random networks, and in Sec. II we apply these to random regular graphs. Networks with heterogeneous degree distributions are examined in Secs. III and IV, and conclusions are drawn in Sec. V Extended mathematical calculations are relegated to the appendices.

We first briefly review two recent models for infinite random networks with non-zero clustering. The fundamental quantity describing the networks of [2] is the joint probability distribution $\gamma(k,c)$, giving the probability that a randomly chosen node has degree k and is a member of a c clique (a fully connected subgraph of c nodes). In these networks, nodes may be part of at most one clique. Nodes which are members of a c clique have $c-1$ edges linking them to neighbors within the same clique. They also have an additional $k-(c-1)$ neighbors who are not in the same clique as themselves (note $\gamma(k,c)=0$ for $c>k+1$ since nodes in a c clique must have at least $c-1$ neighbors). Edges which are not internal to a clique are termed *external links*. The degree distribution P_k of the network (probability that a random node has k neighbors) is obtained from γ by averaging over all possible clique sizes:

$$P_k = \sum_{c=1}^{k+1} \gamma(k,c) = \sum_{c} \gamma(k,c) \tag{1}$$

and the degree-dependent clustering coefficient c_k [28] is given in terms of γ by

$$c_k = \sum_{c} \frac{\gamma(k,c)}{P_k} \frac{(c-1)(c-2)}{k(k-1)}, \tag{2}$$

see [2] for details. The overall network clustering coefficient C [22] is then $C=\Sigma_{k\geq 2} P_k c_k$.

Analytical results for the giant connected component size are given in [2] and the bond percolation threshold $p_{th}^{(\gamma)}$ is shown to be the solution of the following polynomial equation for p:

$$\frac{1}{z_e}\sum_{k,c}(k-c+1)\gamma(k,c)[p(k-c)+(z_c-c+1)D_c(p)]=1. \tag{3}$$

Here z_e is the average number of external links per node: $z_e=\Sigma_{k,c}(k-c+1)\gamma(k,c)$, z_c is the average degree of nodes in cliques of size c: $z_c=\Sigma_k k\gamma(k,c)/\Sigma_k\gamma(k,c)$, and $D_c(p)=p\Sigma_{m=1}^{c}(m-1)P(m|c)$ are polynomial functions of p. The functions $P(m|c)$ give the probability that a node in a c clique belongs (after the deletion of edges with probability $1-p$) to a connected cluster of m nodes within the clique, including itself; these polynomial functions of p are defined and tabulated in [3].

A different approach to modeling local clustering is taken in Newman's model [1] (see also [21]). The joint distribution $p_{s,t}$ gives the probability that a randomly chosen node is connected to s single edges (similar to the external links of the γ-theory networks) and to t triangles. The degree distribution is then given by

$$P_k=\sum_{s,t}p_{s,t}\delta_{k,s+2t} \tag{4}$$

and the clustering coefficient, GCC size, and bond percolation threshold (denoted $p_{th}^{(N)}$ for Newman's model) may all be determined analytically (see [1,21] and Appendix A).

It is instructive to compare the constraints imposed on the network structure in each of these models. In Newman's model, a k-degree node may be a member of up to $\lfloor k/2\rfloor$ disjoint triangles, and thus have a local clustering coefficient of up to $1/(k-1)$ if k is even, or up to $1/k$ if k is odd. In contrast, nodes in the γ-theory networks can be members of only a single clique, but using large cliques can give arbitrarily high clustering. In Sec. II, we show that both models imply p_{th} is increased by clustering on random regular graphs—this has recently been demonstrated for the case of triangle-based networks [1] by Miller [21], but we focus on the case of higher-clustering γ-theory networks. A special class of clustered networks are those whose nodes may belong to at most one triangle. Both models [1,2] are applicable to networks in this class, and in Sec. III (see, for example, Fig. 3) we illustrate the interaction between clustering and correlation common to both models of clustering.

II. RANDOM REGULAR GRAPHS

In this Section we restrict our attention to random z-regular graphs, i.e., random graphs in which all nodes have the same degree z. As shown in [18] random graphs with zero clustering (in the limit $N\to\infty$ of infinite number of nodes) may be generated using the configuration model [23,24], for which the percolation threshold is given in terms of the degree distribution P_k as

$$p_{th}^{(1)}=\frac{\sum_k kP_k}{\sum_k k(k-1)P_k}. \tag{5}$$

For random regular graphs the degree distribution is simply $P_k=\delta_{k,z}$, and the zero-clustering percolation threshold is $p_{th}^{(1)}=\frac{1}{z-1}$.

Next we employ Eq. (3) to consider the effect of non-zero clustering in regular networks generated using the algorithm of [2]. In [2] a parametrization of $\gamma(k,c)$ is suggested which is consistent with Eq. (1) and allows the clustering to be easily adjusted:

$$\gamma(k,c)=P_k\binom{k}{c-1}g_k^{c-1}(1-g_k)^{k-c+1}. \tag{6}$$

This is a binomial distribution of the probability mass for k-degree nodes across the c-clique classes for c from 1 to $k+1$, governed by the parameter g_k. Substituting Eq. (6) into Eq. (2) gives the remarkably simple relation $c_k=g_k^2$ between the degree-dependent clustering coefficient and the parameter g_k. For the random regular graphs under consideration here, $\gamma(k,c)$ is nonzero only for $k=z$ and setting $g_z=\sqrt{C}$ in Eq. (6) allows us to investigate regular graphs with clustering coefficient C covering the full range [0,1].

Figure 1(a) compares the bond percolation threshold $p_{th}^{(\gamma)}$ in clustered γ-theory networks [determined by numerical solution of the polynomial Eq. (3), using parametrization Eq. (6)] with the zero-clustering threshold $p_{th}^{(1)}=1/(z-1)$. We also show (magenta dash-dot curves) the percolation threshold $p_{th}^{(N)}$ given by Newman's model [1], and the symbols show the threshold $p_{th}^{(b)}$ found from an earlier bipartite-graph model of clustering [3], see Appendix A for details. It is clear that all three clustering models give thresholds which are larger than $p_{th}^{(1)}$ for $C>0$, i.e., *clustering increases the bond percolation threshold in these random regular graphs*. Support for this statement in the case of γ-theory networks is given in Appendix B. The corresponding result for $p_{th}^{(N)}$ follows from the recent work of Miller [21].

Analytical expressions determining the size S of the giant connected component in γ-theory networks are also given in [2] and Fig. 1(b) shows S as a function of bond occupation probability p for $z=4$, using parametrization Eq. (6). As already noted, increased clustering leads to higher values of the transition point $p_{th}^{(\gamma)}$, but also leads to smaller GCC sizes. Having established that the presence of clustering increases p_{th} in several models of clustered regular graphs, in the remainder of this paper we will consider how diversity of node degrees also plays an important role.

III. HETEROGENEOUS NETWORKS

Networks with a range of node degrees may be characterized at first order by their degree distribution P_k or, at second order, by the joint probability $P(k,k')$ that a randomly chosen edge links vertices of degree k and k'. Analytical results for the percolation threshold are known for the ensembles of networks described fully by P_k [17] or by $P(k,k')$ [19] with

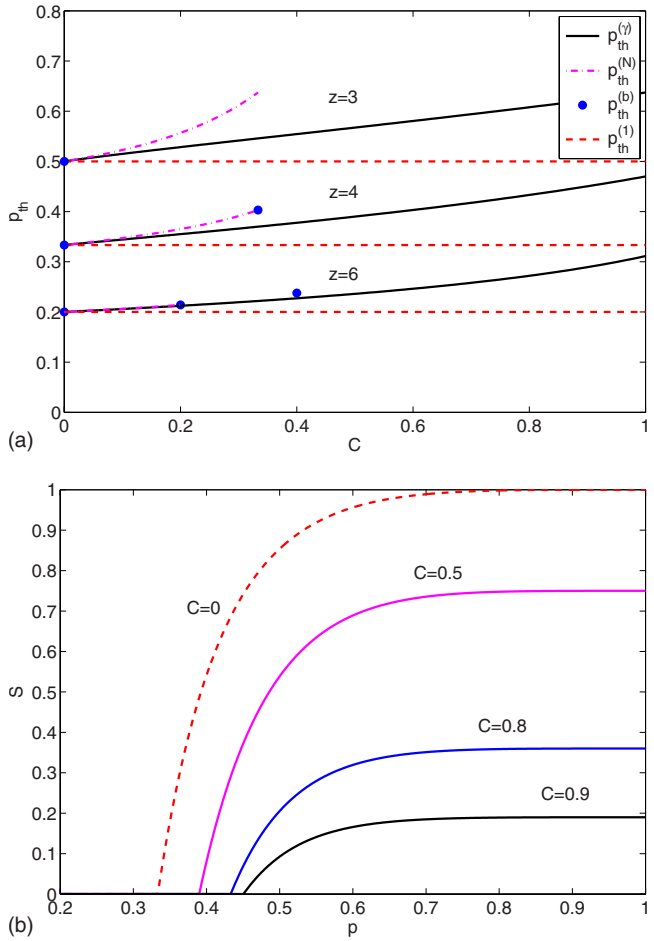

FIG. 1. (Color online) (a) Bond percolation threshold in z-regular graphs with clustering C, generated using the algorithms of [2] ($p_{th}^{(\gamma)}$, black solid) [1], ($p_{th}^{(N)}$, magenta dash-dot), and [3] ($p_{th}^{(b)}$, blue symbols). For comparison, the threshold $p_{th}^{(1)}$ in an unclustered z-regular graph is shown by the red dashed line. Note $p_{th}^{(\gamma)}=p_{th}^{(N)}=p_{th}^{(b)}=p_{th}^{(1)}$ when $C=0$, but the clustered cases all have p_{th} values exceeding $p_{th}^{(1)}$ when $C>0$. Values of z are $z=3$ (top), $z=4$ (middle), and $z=6$ (bottom). (b) Sizes of GCC $S(p)$ in $z=4$ γ-theory regular graphs with clustering coefficients as shown.

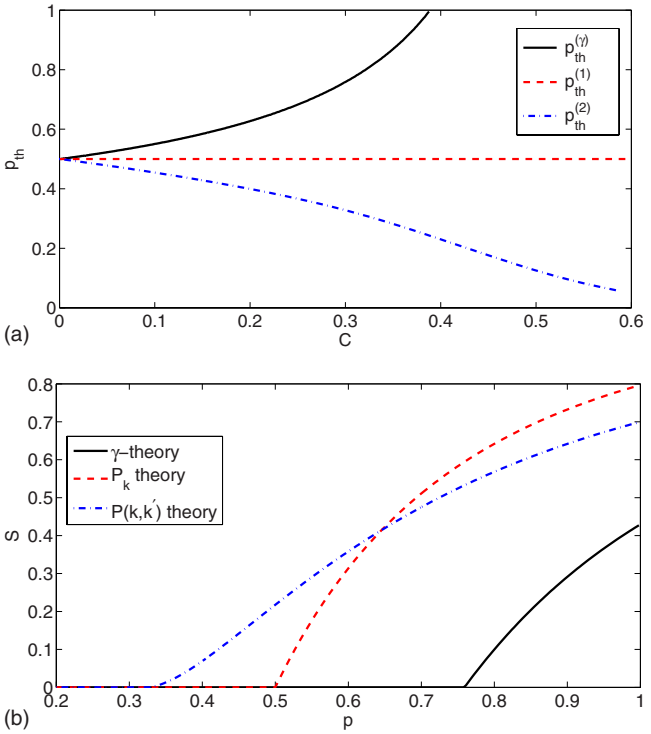

FIG. 2. (Color online) (a) Bond percolation threshold in γ-theory networks with Poisson degree distribution, $z=2$, and clustering C (black solid). For comparison, also shown is the threshold $p_{th}^{(1)}$ in an unclustered network with same degree distribution (red dashed), and the threshold $p_{th}^{(2)}$ in an unclustered network with the same degree-degree correlations (blue dash-dot) as the γ network. (b) Sizes of GCC $S(p)$ for the case $C=0.3$ in γ-theory networks (black solid), and in unclustered networks with the same degree distribution (red dashed), or same degree-degree correlations (blue dash-dot).

respective thresholds denoted $p_{th}^{(1)}$ and $p_{th}^{(2)}$, see Eq. (5) and Appendix C.

In this section we compare the bond percolation threshold $p_{th}^{(\gamma)}$ for various clustered networks with the values $p_{th}^{(1)}$ and $p_{th}^{(2)}$ corresponding to zero-clustering networks with the same degree distribution, or same degree-degree correlations as the clustered network. Our first example is a Poisson random network with degree distribution $P_k=e^{-z}z^k/k!$ and mean degree $z=2$. Figure 2(a) compares $p_{th}^{(\gamma)}$ from Eq. (3) with $p_{th}^{(1)}=1/z$ and $p_{th}^{(2)}$, the latter being determined using the joint distribution $P(k,k')$ for γ-theory networks derived in Appendix C. The clustering level of the γ-theory networks is controlled using the parametrization Eq. (6), with $g_k=\sqrt{C/(1-P_0-P_1)}$ for all k, so that the average clustering coefficient $\Sigma_{k\geq 2}P_k c_k$ is equal to C. Note that the $p_{th}^{(1)}$ line (and $p_{th}^{(2)}$ curve) show the thresholds in unclustered networks with the same degree distribution [and $P(k,k')$ distribution] as the γ-theory network with clustering C.

We see that $p_{th}^{(\gamma)}$ is larger than both of the zero-clustering thresholds $p_{th}^{(1)}$ and $p_{th}^{(2)}$, consistent with our claim that clustering increases the bond percolation threshold. The fact that $p_{th}^{(2)}$ is less than $p_{th}^{(1)}$ is due to the assortativity of the γ-theory networks, see Appendix C and [20].

Figure 2(b) shows the GCC size S in the γ-theory network (black solid curve) as a function of p for clustering $C=0.3$. Also shown are the GCC sizes in a zero-clustering network with the same degree distribution P_k (red dashed curve) and with the same $P(k,k')$ distribution (blue dash-dot curve). This figure can be compared to Fig. 2 of [1] where higher-clustering cases seem to have lower percolation thresholds than the zero-clustering case. However, it should be noted that the focus in [1] is on a different comparison to that undertaken here. The cases plotted in Fig. 2 of [1] are generated from a double Poisson $p_{s,t}$ distribution [see Eq. (13) of [1]] and all share the same mean degree z, but not the same degree distribution. In short, we compare clustered networks with unclustered versions with the same P_k (or $P(k,k')$), while Newman's comparison in [1] retains a common form for the joint distribution $p_{s,t}$, but does not conserve the degree distribution. A similar analysis applies to Fig. 2 of [3], where again it may be shown that the clustered networks used have percolation thresholds larger than those of unclus-

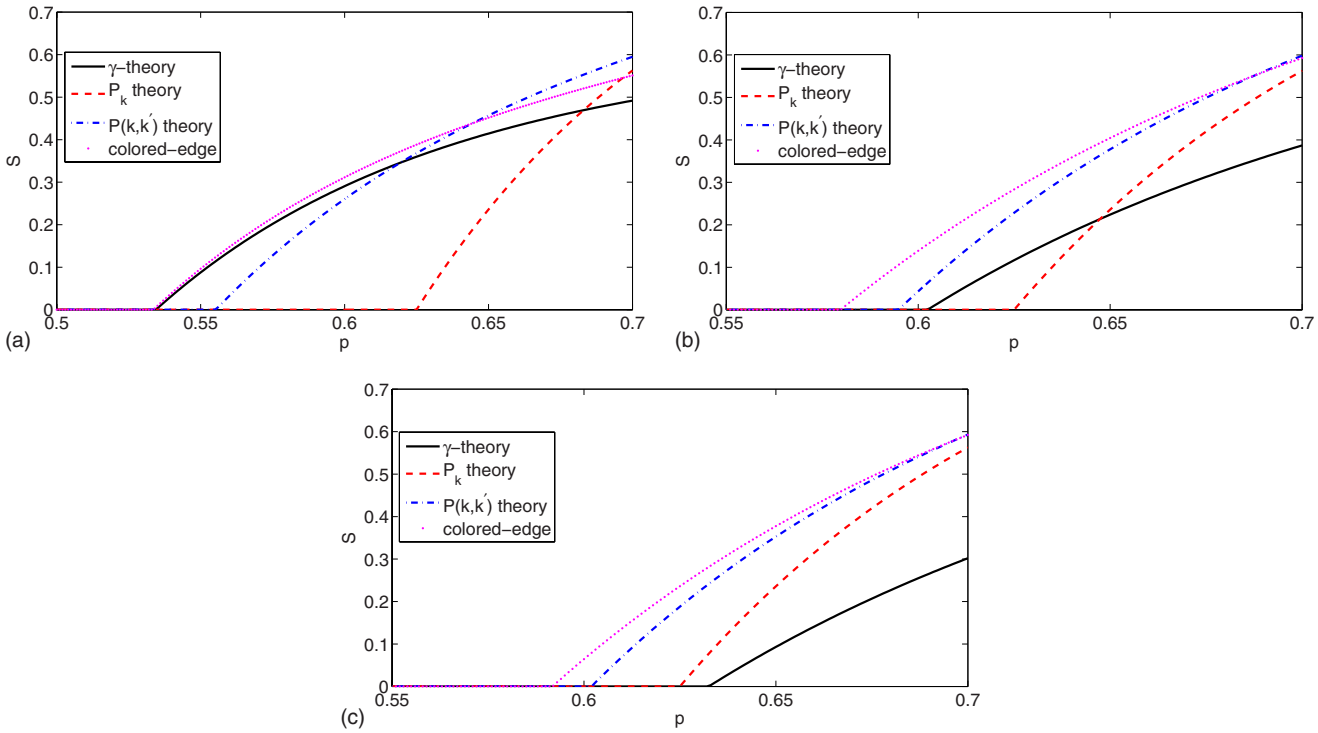

FIG. 3. (Color online) Sizes of GCC $S(p)$ for the γ-theory networks defined by (7) (black solid) and in unclustered networks with the same degree distribution (red dashed), or same degree-degree correlations (blue dash-dot). The magenta dotted curve is for the colored-edge (unclustered) networks defined in Sec. IV. Parameters are $\alpha=0.9$, with (a) $\beta=0.1$, (b) $\beta=0.4$, and (c) $\beta=0.5$.

tered networks with the same degree distribution. In fact this has been demonstrated numerically by Kiss and Green [10], who compared the GCC sizes for the networks of [3] with the GCC sizes in rewired versions of these networks.

Having examined the results for regular graphs and Poisson random networks, one might be tempted at this point to conclude that $p_{th}^{(\gamma)}$ is always greater than $p_{th}^{(1)}$ and $p_{th}^{(2)}$. However, the situation is rather more complicated than this, as demonstrated in Fig. 3 and discussed (for Newman's triangle-based networks) in [21]. To facilitate analysis, and to enable the application of both the γ theory [2] and Newman's theory [1], we restrict our attention now to the special class of networks in which each node has either zero local clustering, or is part of a single triangle. In terms of the γ theory, this means $\gamma(k,c)=0$ unless $c=1$ or $c=3$. For Fig. 3, we have also used a particularly simple degree distribution, with exactly half the nodes having degree $k=2$ and the other half having degree $k=3$. The networks examined are thus described with the theoretical models as follows

$$\gamma(2,1)=p_{2,0}=\frac{1}{2}(1-\alpha);\quad \gamma(2,3)=p_{0,1}=\frac{1}{2}\alpha,$$

$$\gamma(3,1)=p_{3,0}=\frac{1}{2}(1-\beta);\quad \gamma(3,3)=p_{1,1}=\frac{1}{2}\beta, \qquad (7)$$

with the parameters α and β controlling the level of clustering for each degree class.

Figure 3 shows that $p_{th}^{(\gamma)}$ (which equals $p_{th}^{(N)}$ in this special class of networks) may lie either below [Fig. 3(a)] or above [Fig. 3(c)] the zero-clustering thresholds $p_{th}^{(1)}$ and $p_{th}^{(2)}$. Recall our claim is that the presence of triangles increases p_{th} relative to its value in unclustered networks with the same degree distribution *and same correlation structure*. In the next section we show that the correlation structure in these examples is not fully described by only nearest-neighbor correlations as given by $P(k,k')$. When, as described in Sec. IV, the correlation structure is fully matched but clustering eliminated, the GCC size $S(p)$ is given by the magenta (dotted) curve in Fig. 3. Note the transition point for the black (solid) curve is larger in all cases than the transition point for the magenta curve, supporting our claim. Detailed analysis of the correlation structure for these cases is given in Sec. IV and Appendix E.

IV. UNCLUSTERED NETWORKS WITH CORRELATION STRUCTURE

In this section we restrict our attention to the special class of γ-theory networks wherein nodes are members of either one clique or of none, and all cliques are of equal size $c=\bar{c}$ (the example in Sec. III used $\bar{c}=3$), i.e.,

$$\gamma(k,c)=P_k(1-\alpha_k)\delta_{c1}+P_k\alpha_k\delta_{c\bar{c}}, \qquad (8)$$

for a prescribed degree distribution P_k, and with α_k determining the level of clustering for degree-k nodes. Note that the theoretical approaches of [1,2] both apply in the case $\bar{c}=3$.

To understand the correlation structure of these networks we visualize each edge of a network as being colored either

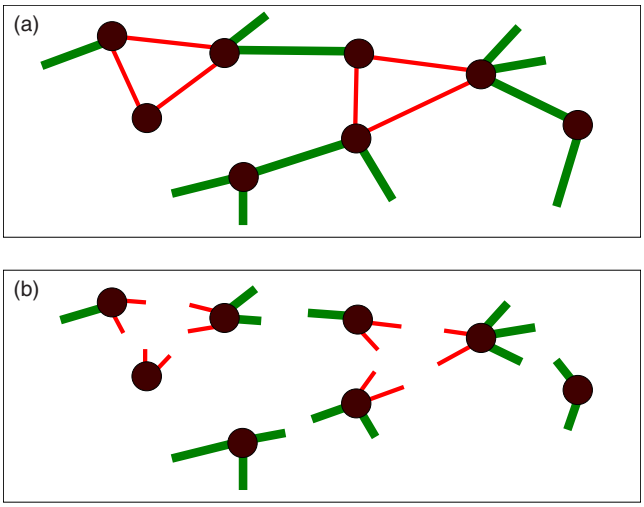

FIG. 4. (Color online) Segment of a clustered network with clique edges colored red (thin lines) and external links colored green (thick lines). After breaking each edge to obtain colored stubs as in (b), a realization of a colored-edge network is created by randomly connecting pairs of stubs of the same color.

green or red (compare to the approach for the triangle-based Newman model taken recently in [21]). The rule for edge coloring is simple: all edges which form part of a $\bar{c}$ clique are colored red, while the remaining edges (the external links in the γ-theory notation) are all colored green, see Fig. 4(a) for an example with $\bar{c}=3$. Now consider the following rewiring process, which preserves the correlation structure, but destroys the clustering within the network. First, break each edge into two end-stubs with each stub retaining the color of the original edge. We now have N isolated "hedgehog" nodes, each with a set of colored stubs as its "spines," see Fig. 4(b). The network is then reconnected together by randomly selecting pairs of green stubs to be joined with a green edge, and similarly randomly pairing red stubs with red edges. The construction method for the original γ-theory (or Newman theory) involves a similar joining of like-colored stubs, except that the randomly chosen red stubs are gathered into $\bar{c}$-cliques. By simply joining pairs of red stubs at random we retain the degree-degree correlation structure (including correlations beyond nearest neighbor) of the γ-theory network, but eliminate triangles (in the $N\to\infty$ limit). The resulting network, which we dub the *colored-edge network*, has properties which are influenced by the fact that red and green stubs are not randomly distributed among the nodes. Taking $\bar{c}=3$ for example, each node is a member of 0 or 1 triangle, so we know that each node must have either exactly zero or exactly two red edges linked to it, while a node of degree k has either k or $k-2$ green edges. These constraints mean the correlation structure of the colored-edge network is not completely described only by the nearest-neighbor correlations [i.e., by the $P(k,k')$ distribution of Appendix C]. A worked example showing this correlation structure is given in Appendix D.

Despite the non-trivial correlation structure, the lack of clustering permits the application of standard tree-based approaches to find the GCC size and the bond percolation threshold $p_{th}^{(ce)}$ for colored-edge networks generated from γ-theory networks with the single non-trivial clique class $c=\bar{c}$ (see [21] for the case $\bar{c}=3$, and Appendix E for the general $\bar{c}$ case). The magenta dotted curve in Fig. 3 shows the GCC size for the colored-edge networks. In Appendix E, we show analytically that $p_{th}^{(ce)}\leq p_{th}^{(\gamma)}$, i.e., that the clustering in the original network causes it to have an increased bond percolation threshold compared to the colored-edge network with the same correlation structure. However, the relative ordering of $p_{th}^{(ce)}$ and $p_{th}^{(1)}$ (or $p_{th}^{(2)}$)—and hence the ordering of $p_{th}^{(\gamma)}$, $p_{th}^{(1)}$, $p_{th}^{(2)}$—depends on the details of the correlation structure beyond nearest-neighbors, so the fact that $p_{th}^{(\gamma)}$ exceeds $p_{th}^{(ce)}$ does not guarantee it will exceed $p_{th}^{(2)}$, see Fig. 3(a) for an example. Further work is needed to elucidate the effects of the correlation structure on p_{th} in these unclustered networks, but we believe the effect of clique-based clustering has now been clearly separated from this question.

V. CONCLUSIONS

We have shown that within the context of the clique-based model of [2], clustering increases the bond percolation threshold in comparison with its value for networks with (i) the same degree distribution and (ii) the same correlation structure. In Sec. II, we used three different approaches for constructing random regular networks with clustering, and confirmed that p_{th} is increased by the presence of clustering, both in triangle-based networks (as shown in [21]) and also in the highly-clustered clique-based models of [3] (as first demonstrated in [10]) and [2] (see Fig. 1 and Appendix B). In Secs. III and IV, we highlighted the importance of condition (ii) by showing that the nth-nearest-neighbor correlations affect p_{th} even in the absence of clustering, i.e., networks with identical nearest-neighbor correlations (as given by the $P(k,k')$ distribution) can have differing p_{th} due to correlations beyond nearest neighbor. The nth-nearest-neighbor correlations are therefore also important when investigating the effects of clustering within various models. When these correlations are fully accounted for, our result remains valid (see Fig. 3 and Appendix E).

What should be our intuitive understanding of the effects of clustering? We believe the correct viewpoint was in fact given by Newman [1] when discussing the giant component size in the case $p=1$: "the triangles that give the network its clustering contain redundant edges that serve no purpose in connecting the giant component together." In other words, the redundant edges cause the GCC size in a clustered network to be smaller than (or at most equal to) the GCC of an unclustered network with the same correlation structure, thus explaining the observation that clustering decreases the value of $S(1)$ in the Newman model [21]. All our results indicate that in fact $S^{(\gamma)}(p)\leq S^{(ce)}(p)$ for all p in $[0,1]$, i.e., that clustering reduces the GCC size for all values of p (compared, as usual, to an unclustered network with same correlation structure), not just for $p=1$. Our main result, that $p_{th}^{(\gamma)}\geq p_{th}^{(ce)}$, may be seen as a simple consequence of this fact: since the GCC size in the clustered network is smaller than (or at most equal to) that in the unclustered network for all p, the transition point where the clustered GCC size becomes nonzero must

be larger than the transition point for the unclustered network. We therefore believe that Newman's explanation of clustering as adding redundant edges reveals the essence of the matter.

In the recent paper [21], Miller independently derives the triangle-based clustering model of [1]. He also demonstrates that within the context of this model, clustering increases the bond percolation threshold in the same sense as claimed here (i.e., when compared to an unclustered network with identical correlation structure). Our work is complementary to [21], since we show that the qualitative effect of clustering seen in triangle-based networks (i.e. clustering increases p_{th}) is also present in more heavily clustered networks described by clique-based theory (compare our results in Appendixes B and E with those in [21]). Our result in Appendix E is limited to the case where only a single nontrivial clique class is present in the network. In order to extend the analysis to networks with a variety of clique sizes, such as those used in Fig. 2, it would be necessary to extend the coloring of edges so that each distinct clique class has a unique color for its edges. Such an extension is beyond the scope of the present work, but we hypothesize that the result would be qualitatively similar to that highlighted here.

The application of these results to real-world networks remains a significant challenge. In this paper, it was possible to separate the effects of clustering and the related correlation structure within the theoretical models [1,2], but it is not clear how this might be attempted for a given real-world network or indeed for other theoretical models with clustering. Nevertheless, the understanding that within the models [1,2] clustering leads generically to an increase in the bond percolation threshold (when comparing networks with the same degree distributions and correlation structures) marks, we believe, an important step forward.

ACKNOWLEDGMENTS

This work was funded by Science Foundation Ireland under Grants No. 06/IN.1/I366 and No. MACSI 06/MI/005.

APPENDIX A: OTHER CLUSTERING MODELS

Newman's results [1] may be used to derive the following polynomial equation for the bond percolation threshold $p=p_{th}^{(N)}$ in networks described by the joint distribution $p_{s,t}$ (see also [21]):

$$2p(1+p-p^2)(p(\langle s^2-s\rangle\langle t^2-t\rangle-\langle st\rangle^2)-\langle s\rangle\langle t^2-t\rangle) - p\langle s^2-s\rangle\langle t\rangle+\langle s\rangle\langle t\rangle=0, \tag{A1}$$

where s and t specify, respectively, the number of single edges and triangle edges attached to a vertex, and $\langle\cdot\rangle$ denotes the average over the joint distribution $p_{s,t}$. For random z-regular graphs we assume the following distribution of probability mass:

$$p_{s,t}=\binom{\lfloor z/2\rfloor}{t}g^t(1-g)^{\lfloor z/2\rfloor-t} \quad \text{for } t=0 \text{ to } \left\lfloor\frac{z}{2}\right\rfloor, \tag{A2}$$

and calculate the clustering C in terms of the single parameter g using the results of [1]. The magenta dash-dotted curves in Fig. 1 show $p_{th}^{(N)}$ as a function of C.

Another analytically solvable case of clustered random regular graphs is provided by Newman's bipartite graph model [3]. In this model, nodes may be part of some number of groups (cliques), and the structure may be represented as a bipartite graph with links between nodes (individuals) and the groups (cliques) of which they are members. In general this model cannot be fitted to desired degree distributions, but the special case of z-regular graphs may be produced by taking the distribution of group sizes to be $s_n=\delta_{n,\nu}$, and the number of groups in which a node partakes to be distributed as $r_m=\delta_{m,\mu}$, where integers ν and μ satisfy the relation $(\nu-1)\mu=z$. For the case $z=6$, for example, there exist 3 such (ν,μ) pairs: (2,6), (3,3), and (4,2), leading to respective clustering coefficients of 0, 1/5, and 2/5. The formulas given in [3] allow us to calculate the bond percolation threshold for each of these cases, and the results are plotted with symbols in Fig. 1(a). Consistent with the models of [1,2], the percolation threshold is clearly increased above its unclustered value in this model.

APPENDIX B: CLUSTERING INCREASES p_{th} IN RANDOM REGULAR GRAPHS

Here we demonstrate that for random z-regular graphs generated using the γ-theory [2], the bond percolation threshold $p_{th}^{(\gamma)}$ is larger than the value $p_{th}^{(1)}=1/(z-1)$ for an unclustered network. We show this for a general $\gamma(k,c)$ distribution (with nonzero values only for $k=z$ since the graphs are regular), so the result is not dependent on a particular parametrization such as Eq. (6).

Note from Eq. (3) that $p_{th}^{(\gamma)}$ is the solution of the polynomial equation $F(p)=1$ where

$$F(p)=\frac{1}{z_e}\sum_c (z-c+1)\gamma(z,c)[p(z-c)+(z-c+1)D_c(p)], \tag{B1}$$

with $z_e=\Sigma_c(z-c+1)\gamma(z,c)$. We use the following two properties of the polynomials $D_c(p)$: (a) $D_c(p)$ is a monotonically increasing function of p on the interval [0,1] with $D_c(0)=0$, and (b) $D_c(p)$ is bounded above by

$$D_c(p)\le\frac{p^2(c-1)}{1-p(c-2)} \tag{B2}$$

for all p with $0\le p\le\frac{1}{c-2}$.

By property (a), the polynomial $F(p)$ defined in Eq. (B1) is monotonically increasing in p, with $F(0)=0$. Since $F(p_{th}^{(\gamma)})=1$, we can guarantee that $p_{th}^{(1)}\le p_{th}^{(\gamma)}$ by showing that $F(p_{th}^{(1)})\le 1$. Using property (b), we have that for $p\le\min_c[1/(c-2)]$,

$$F(p) \leq \frac{1}{z_e}\sum_c (z-c+1)\gamma(z,c) \times \left[p(z-c) + \frac{(z-c+1)p^2(c-1)}{1-p(c-2)}\right]. \quad \text{(B3)}$$

Substituting $p=p_{th}^{(1)}=1/(z-1)$ [note this p obeys $p\leq 1/(c-2)$ for all relevant cliques classes since $c\leq z+1$ in a z-regular graph] simplifies the right-hand side to yield

$$F(p_{th}^{(1)}) \leq \frac{1}{z_e}\sum_c (z-c+1)\gamma(z,c) = 1, \quad \text{(B4)}$$

hence implying that $p_{th}^{(\gamma)} \geq p_{th}^{(1)}$ as desired.

APPENDIX C: DEGREE-DEGREE CORRELATIONS IN γ-THEORY NETWORKS

The ensemble of networks characterized by $\gamma(k,c)$ is constructed as described in [2]. To determine the degree-correlation matrix $P(k,k')$ we calculate the probability that a randomly-chosen edge of the network joins together nodes of degree k and k'. The construction algorithm for the $\gamma(k,c)$ network is based on specifying stubs (half-edges) as either *external stubs* or *c-clique stubs*. Since each k-degree node in a c-clique has $k-c+1$ external stubs and $c-1$ c-clique stubs, the number of external edges in the network (half the number of external stubs) is given by

$$E_e = \frac{N}{2}\sum_{k,c}(k-c+1)\gamma(k,c), \quad \text{(C1)}$$

where N is the number of nodes. Similarly, the total number of c-clique edges is

$$E_c = \frac{N}{2}\sum_k (c-1)\gamma(k,c), \quad \text{for } c>1. \quad \text{(C2)}$$

The sum over all c-clique classes, plus the external edges, gives the total number E of edges in the network:

$$E = E_e + \sum_{c>1} E_c = \frac{1}{2}Nz. \quad \text{(C3)}$$

Therefore a randomly chosen edge of the network is an external edge with probability $E_e/E \equiv \alpha^{(1)}$ and is a c-clique edge with probability $E_c/E \equiv \alpha^{(c)}$. Then the global $P(k,k')$ matrix may be written as

$$P(k,k') = \frac{E_e}{E}P_e(k,k') + \sum_{c>1}\frac{E_c}{E}P_c(k,k') = \alpha^{(1)}P_e(k,k') + \sum_{c>1}\alpha^{(c)}P_c(k,k'), \quad \text{(C4)}$$

where $P_e(k,k')$ is the probability that a randomly chosen external edge joins nodes of degrees k and k', and $P_c(k,k')$ is similarly defined for c-clique edges.

Suppose first that the chosen edge is an external edge. Since external edges are composed of randomly-connected external stubs, the probability that an end vertex is of degree k is

$$s_k^{(1)} = \sum_c \frac{(k-c+1)\gamma(k,c)}{\sum_{k',c'}(k'-c'+1)\gamma(k',c')} \quad \text{(C5)}$$

and the probability that the chosen external edge links nodes of degrees k and k' is

$$P_e(k,k') = s_k^{(1)}s_{k'}^{(1)}. \quad \text{(C6)}$$

If the chosen edge is a c-clique edge, the probability that an end-vertex is of degree k is

$$s_k^{(c)} = \frac{(c-1)\gamma(k,c)}{\sum_{k''}(c-1)\gamma(k'',c)} = \frac{\gamma(k,c)}{\sum_{k''}\gamma(k'',c)}, \quad \text{(C7)}$$

and the probability that the chosen c-clique edge links nodes of degree k and k' is

$$P_c(k,k') = s_k^{(c)}s_{k'}^{(c)} \quad \text{for } c>1. \quad \text{(C8)}$$

Inserting Eqs. (C6) and (C8) into Eq. (C4) enables us to write the global $P(k,k')$ matrix for the network as

$$P(k,k') = \alpha^{(1)}s_k^{(1)}s_{k'}^{(1)} + \sum_{c>1}\alpha^{(c)}s_k^{(c)}s_{k'}^{(c)} = \sum_{c\geq 1}\alpha^{(c)}s_k^{(c)}s_{k'}^{(c)}. \quad \text{(C9)}$$

We can then calculate $p_{th}^{(2)}$, the bond percolation threshold in an unclustered network with the same degree-degree correlations as the original network [19,25], as $p_{th}^{(2)}=1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the matrix $\mathbf{C}$ with entries given by

$$C_{k,j} = \frac{(j-1)}{\sum_{k'}P(k,k')}P(k,j). \quad \text{(C10)}$$

Moreover, we can see that γ-theory networks are necessarily assortative by showing that

$$\sum_{k,k'}kP(k,k')k' - \left[\sum_{k,k'}kP(k,k')\right]^2 \geq 0. \quad \text{(C11)}$$

This quantity determines the sign of the Pearson correlation coefficient r defined in Eq. (3) of [25], with positive values corresponding to assortative networks. Using Eq. (C9), the left-hand side of Eq. (C11) may be written as

$$\sum_c \alpha^{(c)}x_c^2 - \left(\sum_c \alpha^{(c)}x_c\right)^2, \quad \text{(C12)}$$

where $x_c=\Sigma_k k s_k^{(c)}$ and $\Sigma_c\alpha^{(c)}=1$, so this expression may be rewritten as

$$\frac{1}{2}\sum_{c,c'}\alpha^{(c)}\alpha^{(c')}(x_c-x_{c'})^2. \quad \text{(C13)}$$

Since all $\alpha^{(c)}$ terms are non-negative the inequality Eq. (C11) must hold, and the γ-theory networks are assortative.

We emphasize the fact that assortativity follows here directly from the decomposition Eq. (C9) of $P(k,k')$ into disjoint parts, each of which has the form of a randomly-

connected network. In Newman's recent clustering model [1], for example, there are also two types of links: those which are edges of triangles, and those which are not. Stubs of each of these two types are randomly connected to stubs of the same type—it follows that the $P(k,k')$ matrix for Newman's theory must be of the general form Eq. (C9), and therefore networks generated by his model must also be assortative.

APPENDIX D: EXAMPLE OF CORRELATION IN COLORED-EDGE NETWORKS

We consider a particular example of the non-trivial correlation structure of the colored-edge networks described in Sec. IV (and further analyzed in Appendix E). Consider a colored-edge network corresponding to the example Eq. (7), where half the nodes are of degree $k=2$ and half are of degree $k=3$. We choose parameters $\alpha=0$ and $\beta=1$, which means that every $k=2$ node has two green stubs, and every $k=3$ node has 1 green and 2 red stubs. Pairs of green stubs are chosen at random to form green edges, and similarly for red stubs/edges. The nearest-neighbor correlations are given by the $P(k,k')$ matrix defined in Eq. (C9); for the parameters chosen here we have $P(2,2)=4/15$, $P(2,3)=P(3,2)=2/15$, and $P(3,3)=7/15$.

Let us now consider degree correlations beyond nearest neighbors. Specifically, we choose a node of degree 3 and examine the fraction of its second neighbors which are also of degree 3 (ignoring cycles in the $N\to\infty$ limit). We denote this quantity $Q(3|3)$, as it is the probability that node A has a second neighbor of degree 3, given that node A itself has degree 3.

Since the degree distribution of first-neighbors of A is given exactly by

$$P(k|3)=\frac{P(k,3)}{\sum_{k'}P(k',3)} \quad \text{for } k=2,3, \tag{D1}$$

it is tempting to calculate second-neighbor correlations under the Markovian assumption that the network is completely described by its $P(k,k')$ distribution. This assumption underlies the calculation of the threshold we denote as $p_{th}^{(2)}$, and if applied to our example would estimate the value of $Q(3|3)$ by

$$\sum_{k'}P(3|k')P(k'|3)=\frac{55}{81}. \tag{D2}$$

However, the coloring of the edges implies that the true nth-nearest-neighbor correlation structure is not adequately described by $P(k,k')$ for $n>1$. To show this, we now calculate the exact value of $Q(3|3)$ and show that it differs from the Markovian-assumption estimate Eq. (D2). First, note that since all $k=3$ nodes have 1 green stub (as well as 2 red stubs) and all $k=2$ nodes have 2 green stubs, traveling along a random green edge will lead to a $k=3$ node with probability 1/3, and to a $k=2$ node with probability 2/3. Similarly, traveling along a random red edge leads to a $k=3$ node with probability 1.

Let us start at the $k=3$ node called A, and enumerate all possible paths leading to degree-3 second neighbors of A, thus calculating $Q(3|3)$. A fraction 1/3 of A's first neighbors are accessed via green edges, with the remaining fraction 2/3 being accessed by traveling along a red edge. Suppose first that we travel along a green edge from A. With probability 1/3 the green edge leads to a $k=3$ neighbor, otherwise the neighbor has $k=2$. If the neighbor has $k=3$, and noting that we arrived at him along a green edge, his connections to second neighbors of A are necessarily along red edges, and so these second neighbors have degree $k=3$ with probability 1. On the other hand, if the first neighbor of A has $k=2$, the access to A's second neighbor along this path must be along a green edge, and so the second neighbor found on this path is of degree 3 with probability 1/3.

To summarize so far: starting from a $k=3$ node A we can find degree-3 second neighbors of A by proceeding

(i) along a green edge (prob 1/3) via a $k=3$ first neighbor (prob 1/3) and then along a red edge (prob 1). Total probability: 1/9.

(ii) or, along a green edge (prob 1/3) via a $k=2$ first neighbor (prob 2/3) and then along a green edge (prob 1/3). Total probability: 2/27.

Similar arguments show that the remaining possible paths proceed from A

(iii) along a red edge (prob 2/3) via a first neighbor of degree-3 (prob 1) and then either along a red edge (prob 1/2) to a $k=3$ node (prob 1), or along a green edge (prob 1/2) to a $k=3$ node (prob 1/3). Total probability: 4/9.

Summing over all possible paths we obtain

$$Q(3|3)=\frac{1}{9}+\frac{2}{27}+\frac{4}{9}=\frac{17}{27}, \tag{D3}$$

which differs from the value 55/81 obtained in Eq. (D2) under the Markovian approximation. We conclude that in colored-edge networks (and hence in the γ-theory clustered networks) nth-nearest-neighbor correlations beyond $n=1$ are not completely described by the $P(k,k')$ distribution under the Markovian assumption.

APPENDIX E: PERCOLATION IN COLORED-EDGE NETWORKS

We consider bond percolation in an unclustered network of N nodes (in the $N\to\infty$ limit), composed of two types of edges (green or red) as described in Sec. IV. Such networks may be created by considering a γ-theory network with only one nontrivial clique class $c=\bar{c}$ and with the internal c-clique edges colored red while the external links are colored green, see Fig. 4 for an example with $\bar{c}=3$. A similar idea is used in [21] for Newman's triangle-based networks [1]. The total number of green stubs (half-edges) is

$$N\sum_{k,c}(k-c+1)\gamma(k,c) = N\sum_{k}k\gamma(k,1)+N\sum_{k}(k-\bar{c}+1)\gamma(k,\bar{c}), \tag{E1}$$

and the total number of red stubs is

$$N\sum_k (\bar{c}-1)\gamma(k,\bar{c}), \tag{E2}$$

since any node with red stubs has exactly $\bar{c}-1$ of them. Green stubs are randomly linked to green stubs, and similarly for red stubs. As in [2,12], we define a node as active if it is part of the GCC, and assume all nodes are initially inactive. Using a tree structure, define q_g as the probability that a node with a green edge linking to its parent is active, and q_r is the corresponding probability for a node with a red edge leading to its parent. Then standard arguments (see, for example [12,26]) lead to the following self-consistent equations for q_g and q_r:

$$q_g = G(q_g,q_r)$$

$$q_r = R(q_g,q_r), \tag{E3}$$

where the functions G and R are defined as

$$G(q_g,q_r) = \sum_{k,c} \frac{(k-c+1)\gamma(k,c)}{z_e} \times [1-(1-pq_g)^{k-c}(1-pq_r)^{c-1}], \tag{E4}$$

$$R(q_g,q_r) = \sum_k \frac{\gamma(k,\bar{c})}{\sum_{k'}\gamma(k',\bar{c})} \times [1-(1-pq_g)^{k-\bar{c}+1}(1-pq_r)^{\bar{c}-2}]. \tag{E5}$$

Similarly, the final density of active nodes, i.e., the GCC size, is given by

$$S = \sum_{k,c} \gamma(k,c)[1-(1-pq_g)^{k-c+1}(1-pq_r)^{c-1}]. \tag{E6}$$

The percolation threshold point is determined by standard cascade condition arguments [26] applied to the system Eqs. (E3)–(E5). Defining $\mathbf{B}$ as the matrix

$$\mathbf{B} = \frac{1}{p}\left[\begin{array}{cc} \frac{\partial G}{\partial q_g} & \frac{\partial G}{\partial q_r} \\ \frac{\partial R}{\partial q_g} & \frac{\partial R}{\partial q_r} \end{array}\right]\Bigg|_{q_g=q_r=0}, \tag{E7}$$

it has elements given explicitly by

$$B_{11} = \frac{1}{z_e}\sum_{k,c}(k-c+1)(k-c)\gamma(k,c)$$

$$B_{12} = \frac{(\bar{c}-1)}{z_e}\sum_k (k-\bar{c}+1)\gamma(k,\bar{c})$$

$$B_{21} = \frac{1}{\sum_{k'}\gamma(k',\bar{c})}\sum_k (k-\bar{c}+1)\gamma(k,\bar{c})$$

$$B_{22} = \bar{c}-2. \tag{E8}$$

Note that the partial derivatives defining the elements of $\mathbf{B}$ are all evaluated at $q_g=q_r=0$, and these partial derivatives are all proportional to p. Thus the division by p in the definition Eq. (E7) gives the p-independent elements Eq. (E8). The percolation threshold in the colored-edge network is given by $p_{th}^{(ce)}=1/\lambda_{\max}$ where $\lambda_{\max}$ is the larger of the eigenvalues of $\mathbf{B}$, i.e.,

$$p_{th}^{(ce)} = \frac{2}{B_{11}+B_{22}+\sqrt{(B_{11}-B_{22})^2+4B_{12}B_{21}}}. \tag{E9}$$

Since all the B_{ij} elements are non-negative (recall from Sec. I that $\gamma(k,c)=0$ if $c>k+1$), we have the bound

$$p_{th}^{(ce)} \le \frac{1}{B_{22}}, \tag{E10}$$

(this follows by noting $\sqrt{(B_{11}-B_{22})^2+4B_{12}B_{21}} \ge B_{22}-B_{11}$) which we will use below.

Next we show that $p_{th}^{(ce)} \le p_{th}^{(\gamma)}$ for networks of this type. From Eq. (3), note that $p_{th}^{(\gamma)}$ is the solution of the polynomial equation $H(p)=1$, where

$$H(p) = \frac{1}{z_e}\sum_{k,c}(k-c+1)\gamma(k,c) \times [p(k-c)+(z_c-c+1)D_c(p)] = B_{11}p + \frac{1}{\bar{c}-1}B_{12}B_{21}D_{\bar{c}}(p), \tag{E11}$$

and B_{ij} refers to the entries of the non-negative matrix $\mathbf{B}$ above. Following the arguments of Appendix B, we will show that $H(p_{th}^{(ce)}) \le 1$ by using the bound Eq. (B2) on $D_{\bar{c}}(p)$. This gives

$$H(p) \le B_{11}p + B_{12}B_{21}\frac{p^2}{1-p(\bar{c}-2)} \tag{E12}$$

for all p such that $0 \le p \le \frac{1}{\bar{c}-2}$.

Noting that $B_{22}=\bar{c}-2$, we see from Eq. (E10) the inequality $p_{th}^{(ce)} \le 1/(\bar{c}-2)$ is obeyed and so we may apply Eq. (E12) with $p=p_{th}^{(ce)}$. Substituting $p=p_{th}^{(ce)}$ from Eq. (E9) [with Eq. (E8)] into Eq. (E12) and simplifying yields

$$H(p_{th}^{(ce)}) \le 1, \tag{E13}$$

and the result $p_{th}^{(ce)} \le p_{th}^{(\gamma)}$ follows.

[1] M. E. J. Newman, Phys. Rev. Lett. **103**, 058701 (2009).
[2] J. P. Gleeson, Phys. Rev. E **80**, 036107 (2009).
[3] M. E. J. Newman, Phys. Rev. E **68**, 026121 (2003).
[4] T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm, J. Appl. Probab. **45**, 743 (2008).
[5] M. Á. Serrano and M. Boguñá, Phys. Rev. Lett. **97**, 088701 (2006).
[6] M. Á. Serrano and M. Boguñá, Phys. Rev. E **74**, 056114 (2006).
[7] M. Á. Serrano and M. Boguñá, Phys. Rev. E **74**, 056115 (2006).
[8] K. T. D. Eames, Theor Popul. Biol. **73**, 104 (2008).
[9] J. C. Miller, J. R. Soc., Interface **6**, 1121 (2009).
[10] I. Z. Kiss and D. M. Green, Phys. Rev. E **78**, 048101 (2008).
[11] P. Trapman, Theor Popul. Biol. **71**, 160 (2007).
[12] J. P. Gleeson and S. Melnik, Phys. Rev. E **80**, 046121 (2009).
[13] P. Grassberger, Math. Biosci. **63**, 157 (1983).
[14] 161M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[15] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[16] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[17] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[18] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[19] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).
[20] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **78**, 051105 (2008).
[21] J. C. Miller, Phys. Rev. E **80**, 020901(R) (2009).
[22] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[23] E. A. Bender and E. R. Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978).
[24] B. Bollobás, Eur. J. Comb. **1**, 311 (1980).
[25] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[26] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).
[27] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[28] The *local clustering coefficient* for a node A is defined as the fraction of pairs of neighbors of node A which are also neighbors of each other [22], and the *degree-dependent clustering* c_k is the average of the local clustering coefficient over the class of all nodes of degree k [6,27].