IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1** (2011), pp. 25–32

Site percolation in clustered random networks

Adam Hackett[1], James P. Gleeson[1] and Sergey Melnik[1]

[1] *Department of Mathematics & Statistics, University of Limerick, Ireland*

Abstract. In [1] a model was introduced for the creation of a class of random networks with nonzero clustering. Within this model the degree distribution and clustering spectrum of a network are prescribed, and as such can be fitted to given real-world data. Here we present an analytical approach to site percolation in these networks. Theoretical predictions for the critical site occupation probability and the fractional size of the giant connected component are shown to match well with numerical simulations on both real and synthetic networks. We also demonstrate the application of our approach to Newman's triangle-based model of clustered random networks [2].

Keywords: Random networks, clustering, percolation
MSC 2000: 05C82, 91C20, 60K35

1. Introduction

According to Bollobás and Riordan [3], "Percolation theory was founded by Broadbent and Hammersley [4], in order to model the flow of fluid in a porous medium with randomly blocked channels." From a theoretical viewpoint, percolation is simply a process through which we obtain a random subgraph of a graph. The two simplest types are site percolation and bond percolation. In site percolation sites (vertices) are selected independently with uniform probability p_s. Selected sites are labelled *occupied* and p_s is called the *site occupation probability*. Similarly, for bond percolation bonds (edges) are chosen with uniform *bond occupation probability* p_b. Of particular interest is the component structure of the selected subgraph. At some critical value, termed the *percolation threshold*, p_{th}, it undergoes a phase transition from being comprised of a disconnected set of small components to containing a giant connected component (GCC) which spans a nonzero fraction of the total number of vertices N.

Concepts from percolation theory have been applied extensively to investigate the resilience of social, technological, and biological networks to both random error and intentional attack [5, 6, 7]. For example, in simple epidemiological models the bond occupation probability p_b has been related to the

average transmissibility of a disease, so that the GCC represents the expected fractional size of an epidemic outbreak [8, 9]. In the complex networks literature there exist a number of well established analytical results for percolation on synthetic networks based on the configuration model [10, 11, 12, 13]. These results were derived by assuming a locally tree-like structure for the underlying network, thus neglecting the presence of short loops (cycles). However, while that is a reasonable assumption for such networks since they are known have zero clustering in the limit of infinite network size, as pointed out by Watts and Strogatz [14] (and by Rappoport as far back as the 1940s [15]) many empirical networks contain an unexpectedly high number of triangles – fully connected subgraphs (or *cliques*) of three vertices – and thus have non-negligible levels of clustering.

In [1] Gleeson demonstrated an analytical approach to bond percolation in a class of clustered random networks with arbitrary degree distribution P_k and degree-dependent clustering coefficient c_k [16, 17]. The model used to create these networks is a generalization of that introduced by Trapman [18] where clustering was achieved by embedding cliques within a locally tree-like structure. Prior to [1] most studies of the effects of clustering on percolation had relied on numerical simulations using various algorithms to generate clustered networks, e.g. [19, 20, 21]. While Newman did derive analytical results for a bipartite graph model of highly clustered networks [22], in general those networks can not be fitted to a prescribed degree distribution P_k. Serrano and Boguñá were able to obtain approximate analytical solutions for their model [23, 24], but the networks they constructed have only weak levels of clustering $c_k < 1/(k-1)$.

In this paper we extend the analytical approach presented in [25] to obtain the size of the GCC and the position of the site percolation threshold in Gleeson's networks. We also demonstrate how this approach can be applied to Newman's triangle-based model of clustered random networks to independently derive analytical results for site percolation without referring to the so-called generating function formalism [11, 12].

2. Site percolation in Gleeson's $\gamma(k, c)$ networks

Following [1] we consider the class of random networks in which each vertex may be part of a single clique. Networks of this type can be decomposed into a disjoint set of cliques linked together by external edges (see Fig. 1 of [1]). Internal edges connect vertices in the same clique, while external edges connect one clique to another. The joint probability that a randomly chosen vertex has degree k and is a member of a c-clique is $\gamma(k, c)$. Gleeson [1] calculated

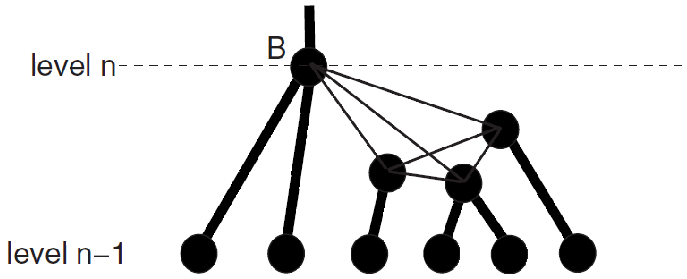

Figure 1: Tree diagram for level-by-level activation [1].

that the degree-dependent clustering coefficient in such a network is

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

where $P_k = \sum_c \gamma(k,c)$, the degree distribution, is the probability that a randomly chosen vertex has degree k.

To obtain analytical results for site percolation in these networks we apply a tree-based approach which is generalizable to a variety of cascade dynamics on networks [25] and is related to work on the random field Ising model [26]. In this approach we consider level-by-level activation of vertices (Fig. 1). The expected fractional size of the GCC, S, is simply the probability that in the steady-state of this process a randomly chosen vertex will be active.

Briefly, we find that

$$S = p_s \sum_{k,c} \gamma(k,c)[1-(1-q)^{k-c+1}(1-Q_c)], \tag{2}$$

where $Q_c = 1-(1-\bar{q}_c)^{c-1}$ is the probability that the top node (such as B) of a c-clique is activated by its clique neighbours, and

$$\bar{q}_c = p_s \sum_{k'} \frac{\gamma(k',c)}{\sum_{k''}\gamma(k'',c)}[1-(1-q)^{k'-c+1}], \tag{3}$$

is the probability that a c-clique member at the intermediate level is activated by one of its children at level $n-1$. Our system of equations is closed with the recurrence relation

$$G(q) = p_s \sum_{k,c} \Pi_{kc}[1-(1-q)^{k-c}(1-Q_c)], \tag{4}$$

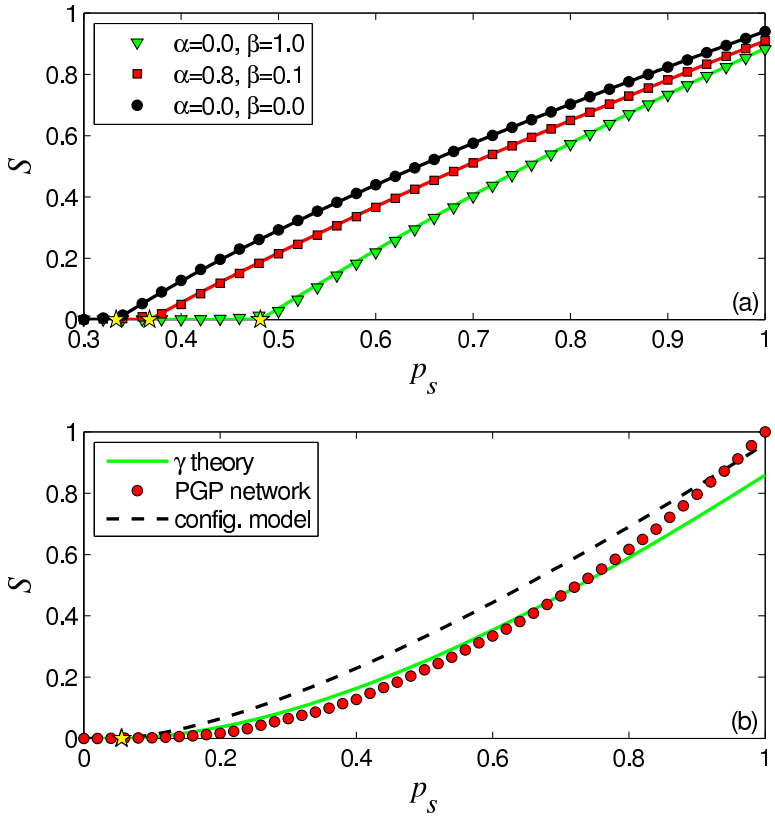

Figure 2: GCC size S vs. site occupation probability p_s for (a) synthetic networks with Poisson degree distribution; (b) pretty-good-privacy network. Percolation thresholds p_{th} marked with yellow pentagrams.

which prescribes the probability that (in an infinite network) a vertex is active, conditional on its parent (the vertex on the next highest level with which it shares a link) being inactive. The site percolation threshold is the solution for p_s of $G'(0) = 1$:

$$p_s \sum_{k,c} \Pi_{kc}[(k-c) + p_s(c-1)(z_c - c + 1)] = 1. \tag{5}$$

Comparisons between our theoretical predictions for the size of the GCC and the results of numerical simulations are shown in Fig. 2. In Fig. 2(a) we consider networks with Poisson degree distribution $P_k = z^k e^{-z}/k!$ and mean degree $z = 3$. We introduce clustering into these networks by embedding

3-cliques (triangles) and 4-cliques:

$$\gamma(k,c) = [(1-\alpha-\beta)\delta_{c1} + \alpha\delta_{c3} + \beta\delta_{c4}]P_k. \tag{6}$$

The γ-theory line in Fig. 2(b) was produced by first measuring P_k and c_k for the pretty-good-privacy network [27] and then applying the following parameterization of the joint distribution

$$\gamma(k,c) = P_k \binom{k}{c-1} g_k^{c-1}(1-g_k)^{k-c+1}, \tag{7}$$

where $g_k = \sqrt{c_k}$. As this figure illustrates the match between theory and numerical simulations on synthetic networks is excellent, while for real-world networks we attain a significant improvement over the (nonclustered) configuration model theory.

3. Application of theory to Newman's $p(s,t)$ networks

In [2] Newman generalizes the standard configuration model for the construction of random graphs with arbitrary degree distributions [10,12] by specifying both the number of single edges s_i and triangle edges t_i in which each vertex i participates. The joint probability that a randomly chosen vertex is connected to s single edges and t triangles is $p(s,t)$.

Our tree-based theory can be further generalized to obtain analytical results for site percolation in these networks. For example, we calculate that the expected fractional size of the GCC is

$$S = p_s \sum_{s,t} p(s,t)[1 - {\sigma_0}^s {\tau_0}^t], \tag{8}$$

where σ_0 and τ_0 are respectively the probability that a vertex attached to a single edge is not active and the probability that a vertex attached to a triangle edge is not active. The position of the site percolation threshold is the solution for p_s of

$$2p_s(p_s(\langle s^2 - s\rangle\langle t^2 - t\rangle - \langle st\rangle^2) - \langle s\rangle\langle t^2 - t\rangle) - p_s\langle s^2 - s\rangle\langle t\rangle + \langle s\rangle\langle t\rangle = 0,$$

where $\langle\cdot\rangle$ denotes the average over $p(s,t)$.

The comparison between this theory and numerical simulations is shown in Fig. 3. Here we plot the fractional size of the GCC as a function of site occupation probability for various values of the clustering coefficient C [12] on a synthetic network with doubly Poisson degree distribution

$$p(s,t) = e^{-\mu}\frac{\mu^s}{s!}e^{-\nu}\frac{\nu^t}{t!}, \tag{9}$$

and average degree $\mu + 2\nu = 2$.

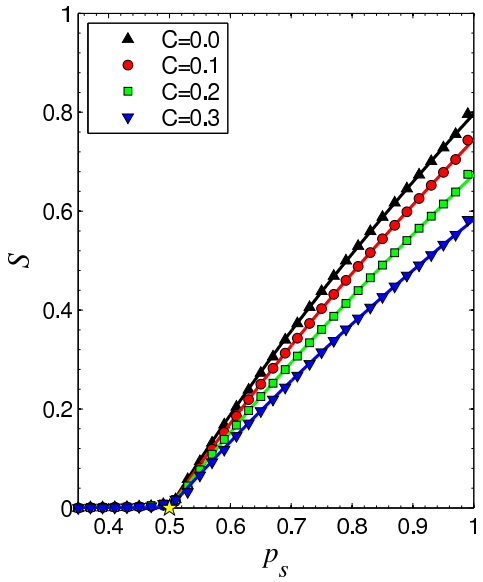

Figure 3: GCC size S vs. site occupation probability p_s for $C = 0.0, 0.1, 0.2, 0.3$ on a network with a doubly Poisson degree distribution. Percolation threshold p_{th} marked with yellow pentagram.

4. Conclusions

In summary, we have extended Gleeson's tree-based approach to threshold dynamics on complex networks [25] to investigate site percolation in the class of clustered random networks introduced in [1]. We have derived analytical results for the position of the site percolation threshold and the size of the giant connected component and shown how this approach may be further generalized to derive similar results for the class of triangle-based clustered networks described by Newman in [2]. Whilst the match between our theory and numerical simulation on synthetic networks of both types is excellent, Gleeson's networks may be parameterized to fit the degree-distribution and clustering spectrum of real-world networks, thus allowing us to achieve a significant improvement over standard nonclustered theory in this case.

Acknowledgements

This work was funded by Science Foundation Ireland under programmes 06/IN.1/I366 and MACSI 06/MI/005.

References

[1] J. P. Gleeson, *Phys. Rev. E* **80**, 036107 (2009).

[2] M. E. J. Newman, *Phys. Rev. Lett.* **103**, 058701 (2009).

[3] B. Bollobás and O. Riordan, *Percolation* (Cambridge University Press, Cambridge, 2006).

[4] R. Broadbent and J. M. Hammersley, *Proc. Cambridge. Philos. Soc.* **53**, 629-641 (1957).

[5] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).

[6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, (Oxford University Press, Oxford, 2003).

[7] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).

[8] P. Grassberger, *Math. Biosci.* **63**, 157 (1983).

[9] M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).

[10] M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).

[11] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).

[12] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).

[13] A. Vázquez and Y. Moreno, *Phys. Rev. E* **67**, 015101(R) (2003).

[14] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).

[15] A. Rapoport, *Bull. Math. Biophys.* **10**, 145 (1948).

[16] M. Á. Serrano and M. Boguñá, *Phys. Rev. E* **74**, 056114 (2006).

[17] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).

[18] P. Trapman, *Theor. Popul. Biol.* **71**, 160 (2007).

[19] K. Klemm and V. M. Eguiluz, *Phys. Rev. E* **65**, 036123 (2002).

[20] E. Volz, *Phys. Rev. E* **70**, 056115 (2004).

[21] M. Á. Serrano and M. Boguñá, *Phys. Rev. E* **72**, 036133 (2005).

[22] M. E. J. Newman, *Phys. Rev. E* **68**, 026121 (2003).

[23] M. Á. Serrano and M. Boguñá, *Phys. Rev. E* **74**, 056115 (2006).

[24] M. Á. Serrano and M. Boguñá, *Phys. Rev. Lett.* **97**, 088701 (2006).

[25] J. P. Gleeson, *Phys. Rev. E* **77**, 046117 (2008).

[26] D. Dhar, P. Shukla, and J. P. Sethna, *J. Phys. A* **30**, 5259 (1997).

[27] M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas, *Phys. Rev. E* **70**, 056122 (2004).