

Comment on “Properties of highly clustered networks”

Istvan Z. Kiss^{1,} & Darren M. Green²*

¹ *Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9RF, UK*

² *Institute of Aquaculture, University of Stirling, Stirling, Stirlingshire, FK9 4LA, UK*

* corresponding author
email: i.z.kiss@sussex.ac.uk

Abstract

We consider a procedure for generating clustered networks previously reported by Newman (M.E.J. Newman, Phys. Rev. E **68**, 026121 (2003)). In the same study, clustered networks generated according to the proposed model, have been reported to have a lower epidemic threshold under SIR-type network epidemic dynamics. By rewiring networks generated by this model, such that the degree distribution is conserved, we show that the lower epidemic threshold can be closely reproduced by rewired networks with close to zero clustering. The reported lower epidemic threshold, can be explained by different degree distributions observed in the networks corresponding to different levels of clustering. Clustering results in networks with high levels of heterogeneity in node degree, a higher proportion of nodes with zero connectivity, and links concentrated within highly interconnected components of small size. Hence, networks generated by this model differ in both clustering and degree distribution, and the lower epidemic threshold is not explained by clustering alone.

PACS number(s): 89.75.Hc, 64.60.Ak, 87.19.Xx, 87.23.Ge

1 Introduction

Many systems can naturally be modelled as a collection of interacting units, represented as a network of nodes connected by edges [2]. Such a representation has attracted interest in epidemiology for describing the dynamics of infectious disease transmission in complex populations (for a review see [3]). Much research has focused on understanding the implications of network properties on the epidemic processes that they support. Heterogeneity in node connectivity (degree) and clustering (i.e. to what degree common neighbours of a node are themselves neighbours) have perhaps attracted the most interest [3,4,5].

To study the theoretical effects of varying one network property (e.g. clustering), one would ideally like to generate multiple networks with all properties identical, except the property of interest. This is easier to say than do, as in practice, different network properties may constrain each other, or not be independent. In this note, we show that the clustered network model proposed by Newman [1] generates networks where higher clustering results in higher heterogeneity in node degree with a higher proportion of nodes with zero connections. This leads to the appearance that clustering decreases the epidemic threshold, whereas link rewiring to eliminate clustering in turn causes, contrarily, a decrease in the epidemic threshold.

2 Results

Newman described an algorithm for network construction that allows for a variable clustering coefficient and variable degree distribution while keeping the mean degree ($\langle k \rangle$) constant [1]. In analysing the dynamics of SIR-type epidemic processes [6] on such networks, he concluded that clustering decreases the epidemic threshold in these networks. In contrast, Keeling [7] concluded that clustering *increases* the epidemic threshold. Keeling’s model maintains a degree distribution that is close to Poisson for different levels of clustering. This allows node-level properties (i.e. the degree distribution) to be preserved between clustered and random networks, though higher-level network properties such as network diameter are not controlled for. Here, we consider further the behaviour of Newman’s model in terms of degree distribution and epidemic dynamics.

We use a definition of clustering that is easily interpretable in an epidemiological context. The number of ‘triples’ in a network, Λ , is the count of the permutations of nodes UVW where V is connected to both U and W . In ‘triangles’ Δ , U must also be connected to W . The clustering coefficient is defined as $C = \frac{\Delta}{\Lambda}$.

Networks were constructed using the method given by Newman [1]. A group size $\nu = 10$ was used throughout with a desired network size of $n' = 5000$ nodes and a desired mean degree of $k = 5$. Clustering was controlled by varying the number of groups per node, μ . The number of groups, g was calculated as $\lceil n'\mu/\nu \rceil$ and the required number of nodes as $n = \lceil g\nu/\mu \rceil$. For all parameter values used, $n = n'$. Following Newman [1], connection probability was thus given by $p = k / [\mu(\nu - 1)]$ and the estimated clustering coefficient as $C = p / [1 + \mu(\nu - 1)/(\nu - 2)]$. Both $\langle k \rangle$ and C were measured on the generated networks to test for agreement with the above.

For each of the g groups, ν nodes were chosen, without replacement, from the population of n nodes. Thus, the numbers of groups enjoyed by each node i is distributed as $\text{Binomial}(g, \nu/n)$. Then, the number of groups each pair of nodes i and j have in common c_{ij} was determined. A link was placed between i and j with probability $1 - (1 - p)^{c_{ij}}$. Matrix A represents the adjacency matrix thus generated where $A_{ij} = 1$ indicates a link, and zero otherwise. All networks were undirected with $A_{ij} = A_{ji}$ and $A_{ii} = 0$.

The node degree distribution for various levels of clustering is shown in Fig. 1, in each case averaged across 100 networks. The proportion of zero-degree nodes and heterogeneity in node degree increases with clustering. These are interdependent since the average number of links per node $\langle k \rangle$ is kept constant. Hence in the most clustered networks, edges connect a limited number of nodes in tightly-connected groups with many nodes having no

contacts at all.

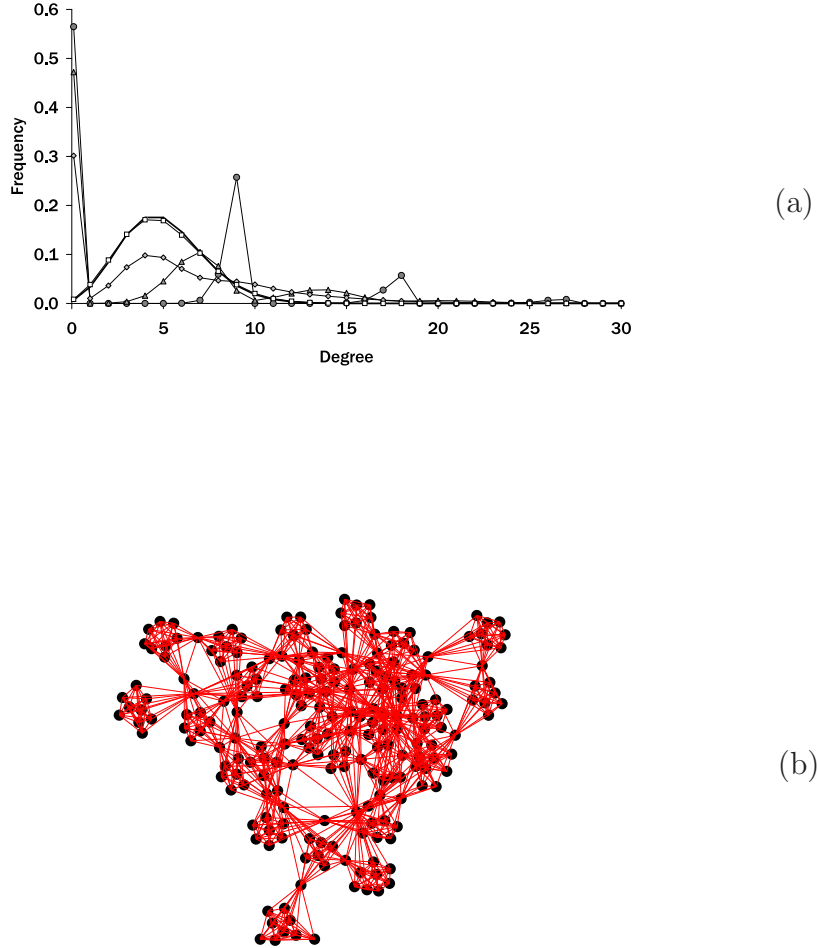


Fig. 1: (a) Degree distribution for networks generated by the clustering algorithm for four different levels of clustering. Square: $C = 0.0$; lozenge: $C = 0.2$; triangle: $C = 0.4$; circle: $C = 0.6$; thick line: Poisson distribution with mean $\lambda = 5$. $n = 5000$. (b) Clustered network with $C = 0.6$ and $n = 500$, with singleton nodes ($\sim 0.56 \times n$) removed. Interconnected groups can be clearly seen, as can occasional nodes with multiple group membership.

It is well known that the variance of the degree distribution is correlated with the basic reproduction number R_0 , and that networks with highly heterogeneous degree distributions have a low epidemic threshold [4,5]. To tease

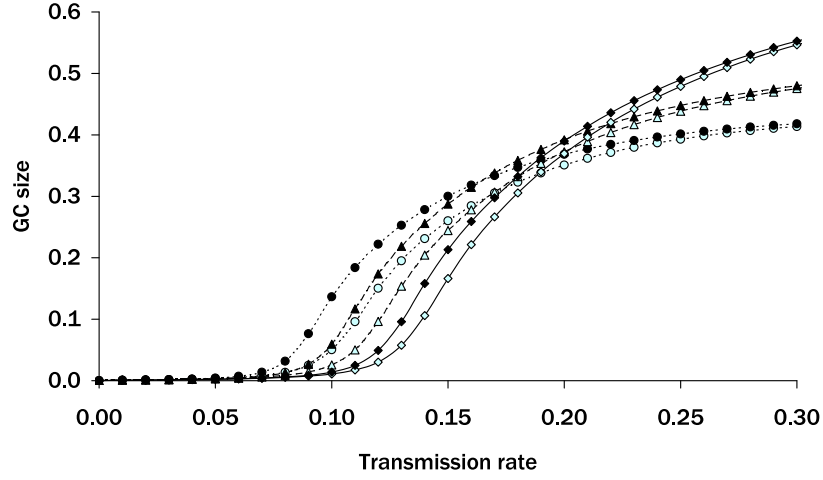


Fig. 2: Giant component size versus transmission rate for clustered and rewired networks with $n = 5000$. Lozenge/solid line: $C = 0.2$; triangle/dashed: $C = 0.4$; circle/dotted: $C = 0.6$; open symbols: clustered networks; filled symbols: rewired networks.

apart the confounded effects of degree distribution and clustering in our networks, we used a rewiring procedure that kept the degree distribution and number of triples Δ unchanged whilst reducing clustering to close to zero by reducing the number of triangles Δ . Repeatedly, four unique nodes u, v, w , and x were chosen, such that $A_{uv} = 1, A_{wx} = 1, A_{ux} = 0$ and $A_{wv} = 0$. These edges were rewired so that $A_{uv} = 0, A_{wx} = 0, A_{ux} = 1$ and $A_{wv} = 1$. Each of the 100 networks for each level of clustering was subject to $50n$ cycles of rewiring and the resulting clustering coefficients recalculated.

In Fig. 2 the size of the giant connected component is plotted for different levels of clustering for both Newman's original model and the same networks with clustering removed by rewiring. This figure quantitatively demonstrates Newman's result of a lower epidemic threshold with the clustered networks, with small numerical differences, but clearly shows that the effect can also be reproduced by the unclustered, rewired, networks. The effect of clustering is seen to be to *increase* the epidemic threshold when the degree distribution is conserved between clustered and unclustered networks. Therefore, the lower epidemic threshold in Newman's clustered network model is a result of the combination of clustering and changing degree distribution. This is not surprising since unclustered networks with high heterogeneity have been shown to have vanishingly small epidemic thresholds in the limit of infinite variance in node degree distribution [4,5]. We argue that the effect observed by Newman is more likely to be a result of the changing degree distribution

rather than clustering.

References

- [1] M.E.J. Newman, Phys. Rev. E **68**, 026121 (2003).
- [2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
- [3] M.J. Keeling, Theo. Pop. Biol. **67**, 1 (2005).
- [4] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
- [5] R.M. May and A.L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
- [6] O. Diekman and J.A.P. Heesterbeek, *Mathematical epidemiology of infectious diseases*, John Wiley & Son (2000).
- [7] M.J. Keeling, Proc. Roy. Soc. Lond. B **266**, 859 (1999).