

Parameterisation of Keeling's network generation algorithm

Jennifer Badham ^{a,*}, Hussein Abbass ^a, Rob Stocker ^a

^a*Artificial Life and Adaptive Robotics Laboratory, School of ITEE, Australian Defence Force Academy, Northcott Drive, Canberra 2600*

Abstract

Simulation is increasingly being used to examine epidemic behaviour and assess potential management options. The utility of the simulations rely on the ability to replicate those aspects of the social structure that are relevant to epidemic transmission. One approach is to generate networks with desired social properties.

Recent research by Keeling and his colleagues has generated simulated networks with a range of properties and examined the impact of these properties on epidemic processes occurring over the network. However, published work has included only limited analysis of the algorithm itself and the way in which the network properties are related to the algorithm parameters.

This paper identifies some relationships between the algorithm parameters and selected network properties (mean degree, degree variation, clustering coefficient and assortativity). Our approach enables users of the algorithm to efficiently generate a network with given properties, thereby allowing realistic social networks to be used as the basis of epidemic simulations. Alternatively, the algorithm could be used to generate social networks with a range of property values, enabling analysis of the impact of these properties on epidemic behaviour.

Key words: Disease spread, Transmission networks, Clustering, Assortativity, Social networks

PACS: 89.75.Hc, 87.23.Ge, 87.19.Xx

* Corresponding author.

Email addresses: research@criticalconnections.com.au (Jennifer Badham), h.abbass@adfa.edu.au (Hussein Abbass), r.stocker@adfa.edu.au (Rob Stocker).

1 Introduction

Historically, statistical analysis of data has been used to investigate biological relationships such as that between water quality and disease. More recently, a different type of data has been added to the biological toolkit, where relationships can be directly visualised as networks. For example, protein interactions for gene regulation has been modelled as a network (Jeong et al., 2001). Network analysis can then be used to increase understanding of these relationships.

One major area of research using network analysis is disease transmission. Epidemiologists have developed mathematical models to estimate the size and other key features of an epidemic, assuming homogeneity in social structure (Kermack and McKendrick, 1927). Elements of social structure have been introduced for models where the network of contacts is fundamental to transmission, such as those that deal with sexually transmitted disease (Kermack and McKendrick, 1927; Becker, 1973; Lajmanovich and Yorke, 1976; Nold, 1980; Gupta et al., 1989; Newman, 2002). Network methods potentially enable the full impact of social structure to be incorporated in epidemic research through simulation.

Detailed demographic and other information can be used to synthesise the population of interest and their activities so as to derive contact information (for example, the EpiSims project (Eubank, Kumar, Marathe, Srinivasan, and Wang, Eubank et al.)). Alternatively, there are several algorithms available that enable a network to be generated with specific values of some social network properties (including Erdős and Rényi (1960); Molloy and Reed (1995); Watts and Strogatz (1998); Barabási and Albert (1999); Newman and Park (2003)). However, the literature on designing algorithms for generating networks is still in its infancy and there is a need for better algorithms that can easily generate networks that match real world desired properties. Common characteristics of social networks (defined in Section 2) include positively skewed degree distribution and higher assortativity and clustering coefficient than would be expected from the degree distribution (Newman and Park, 2003).

One algorithm has been developed and used by Keeling and his colleagues to examine the impact of varying values of mean degree, degree variation and clustering coefficient on epidemic processes occurring over the network (Keeling, 1999; Eames and Keeling, 2002; Read and Keeling, 2003; Keeling, 2005). However, there has been limited analysis of the algorithm itself and the way in which the network properties are related to the algorithm parameters. Understanding such relationships enables the algorithm to be used to efficiently generate networks with desired properties.

2 Social network properties

In a social network, nodes (or vertices or actors) represent individual people and edges (or links) represent the relationship of interest. For epidemic behaviour, that relationship is sufficient contact to permit contagion in some undefined period of time.

The degree of a node is the number of edges that it shares with other nodes. The degree distribution is then the frequency distribution of different degrees across the nodes in the network. As with any distribution, key measures include mean and variance of degree. In addition, real world network degree distributions are typically highly positively skewed (Newman, 2003).

Clustering is a measure of network transitivity, the extent to which neighbours of a node are neighbours of each other. It is defined as follows (Watts and Strogatz, 1998):

Suppose that a vertex v has k_v neighbours [that is, nodes with which it shares an edge]; then at most $k_v(k_v - 1)/2$ edges can exist between them Let C_v denote the fraction of these allowable edges that actually exist. Define C (the clustering coefficient) as the average of C_v over all v .

Each node contributes equally to the mean, regardless of its degree. By convention, a node with degree of 0 or 1 is assigned clustering of 0. The theoretical range is the interval $[0,1]$.

Degree assortativity of a network (denoted r) is defined in (Newman, 2002) as *simply the Pearson correlation coefficient of the degrees at either ends of an edge*. The theoretical range is the interval $[-1,1]$.

Social networks have higher assortativity and clustering coefficient values than would be expected in a network with randomly created edges constrained by the degree distribution (Newman and Park, 2003). Consequently, network generation algorithms that focus on degree distribution may not adequately represent social structures.

3 Algorithm description

In Waxman (1988), a network is generated by randomly locating nodes in a nominal space, creating edges based on the distance between pairs of nodes and then discarding the space. The algorithm used by Keeling and his colleagues is conceptually similar but additionally uses focal points to introduce more complex social structure.

The algorithm has several separate steps and four flexible parameters, as well as the number of nodes and giant component threshold. We use a slightly different presentation and notation from Keeling's to make the later discussion on standardisation easier to follow. However, the method is identical to the published algorithm.

Initially, N nodes and F focal points are randomly located in a periodic square two dimensional space with side length \sqrt{N} . Each node is moved toward the nearest focal point some proportion (m here, f in Keeling papers) of its distance from that focal point. The movement of the nodes toward the focal points increases the local density of nodes near focal points and decreases the distance between nodes that are near the same focal point.

The next step in the algorithm creates edges between nodes based on the Euclidean distance (d) between the two nodes and two parameters, referred to as height (H) and variance (V_k here, V in Keeling papers). The parameter names reflect the Gaussian inspiration for the function. Edge probability is given by:

$$P(\text{edge}) = \min \left\{ \begin{array}{l} H \exp\left(\frac{-d^2}{2V_K}\right) \\ 1 \end{array} \right. \quad (1)$$

Note that, despite the Gaussian influence, this is not a probability density function. In particular, H and V_K are set independently so the total probability over all distances is not 1.

Following creation of the edges, the information about location in the square space is discarded. If the giant component of the network includes some threshold proportion of nodes (90% is used in the literature), the giant component is retained as the generated network. If the giant component is too small, the algorithm is restarted.

An example network is shown at Figure 1. The algorithm parameters used to generate this network are defined later.

4 Interpretation and standardisation of parameters

In order to examine the impact of parameters for different sized networks, some standardisation was undertaken. The parameters can be considered in three pairs. The scale of the network is set by the number of nodes (N) and the giant component threshold. This pair needs no modification as they set the scale required.



Fig. 1. A single instance of a network generated with the Keeling algorithm and parameters $N=1\ 000$, $f=0.1$, $m=0.3$, $H=1.80$ and $V=0.0395$. The network is drawn using the Kamada-Kawai layout (Kamada and Kawai, 1988) within Pajek (Batagelj and Mrvar, 1998) software.

The second pair of parameters concerns focal points, their number (F) and their impact on node location (m). These focal points enable representation of *places where people congregate, and could represent schools and workplaces, or family groups, depending on their number* (Keeling, 2005, pg 3). The parameter m is already independent of scale. However, the density of focal points is F/N per unit area. This means that the same number of focal points could be sparse or dense, depending on N , leading to large or small impacts on distances between nodes. Replace F with a parameter for the density of focal points ($f = F/N$). For a given f and m , the same local pattern of nodes will be created regardless of the total number of nodes (N).

The final parameters concern edge density (H, V_K), with the probability of an edge also based on the distance between nodes in the notional space. Recall that the space is discarded after the network is generated.

The H parameter in the edge probability function represents the probability of an edge between two identically located nodes. This is already independent of scale. Note that if $H > 1$, there is some threshold distance such that an edge must be formed between any pair of nodes closer than that distance.

The V_K parameter provides a distance over which edge creation occurs and is inspired by the Gaussian representation of the variance of node pair distances. *Small values of $[V_K]$ increase the density of short edges relative to longer ones* (Waxman, 1988). If V_K is very small, $P(\text{edge})$ is low for all but very small values of d . In contrast, if V_K is very large, the value of d has minimal impact on the probability and the probability of an edge approaches H for all node pairs. In a sense, V_K joins the communities that form around each focal point.

As published, the variance (V) parameter has no size-independent interpretation. However, it is possible to standardise this parameter in a way that is both consistent with the Gaussian inspiration and allows experiments using either formulation to be compared. There are two standardisation approaches.

As included in the published formula, a pair of nodes in the same absolute locations in two different sized networks would have the same probability of an edge being created between them. On the other hand, (Waxman, 1988) standardised distances as a proportion of the maximum distance.

Clearly, a larger network has many more pairs of nodes and a larger mean distance between them. Consider the variance of the distance between pairs of nodes where the nodes are placed randomly in 2D square wrapped space of side length \sqrt{N} , ignoring the effect of the focal points. Variance of node pair distance is given by:

$$\begin{aligned}\text{var}d_{ij} &= \sigma_d^2 \\ &= \frac{1}{N_d} \sum_{ij} d_{ij}^2 - \bar{d}^2\end{aligned}\quad (2)$$

where d_{ij} is distance node i to node j and N_d is the number of node pairs, given by:

$$\begin{aligned}d_{ij} &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\ N_d &= \frac{N(N-1)}{2}\end{aligned}$$

Consider the first term (sum of distance squared). As the space is wrapped, the selection of (0,0) is arbitrary. Locate (0,0) at node i and reset (0,0) for any given node i . Further, there are four identical quadrants with (on average over an ensemble of nodes constructed in the same way) $(N-1)/4$ nodes and maximum coordinate of $\sqrt{N}/2$. In addition, there are N statistically identical points (to be considered as node i) and each distance should only be included once. Thus:

$$\begin{aligned}\sum_{ij} d_{ij}^2 &= \frac{N}{2} \sum_j d_j^2 \quad x_j, y_j \in [-\sqrt{N}/2, \sqrt{N}/2] \\ &= 2N \sum_j (x_j^2 + y_j^2) \quad x_j, y_j \in [0, \sqrt{N}/2]\end{aligned}\quad (2a)$$

Consider the second term (mean distance). On average, the mean distance across the space is the mean within a quadrant and the selection of node i is arbitrary.

$$\begin{aligned}\bar{d}_{ij} &= \bar{d}_j \\ &= \frac{4}{N-1} \sum_j \sqrt{x_j^2 + y_j^2} \quad x_j, y_j \in [0, \sqrt{N}/2]\end{aligned}\quad (2b)$$

Substitute (2a) and (2b) into (2) and take the limit for large N with continuous distribution of nodes:

$$\sigma_d^2 = \frac{4}{N-1} \int_0^{\sqrt{N}/2} \left[\int_0^{\sqrt{N}/2} (x^2 + y^2) dy \right] dx - \left[\frac{4}{N-1} \int_0^{\sqrt{N}/2} \left[\int_0^{\sqrt{N}/2} \sqrt{x^2 + y^2} dy \right] dx \right]^2 \quad (3)$$

Solving the integrals gives:

$$\begin{aligned} \sigma_d^2 &= \frac{4}{(N-1)} \frac{N^2}{24} \\ &\quad - \left[\frac{4}{(N-1)} \frac{N^{3/2}}{24} \left[\log(1 + \sqrt{2}) + \sqrt{2} \right] \right]^2 \\ &= \frac{N^2}{6(N-1)} \times \\ &\quad \left[1 - \frac{N}{6(N-1)} \left[\log(1 + \sqrt{2}) + \sqrt{2} \right]^2 \right] \\ &\approx 0.0203 N \end{aligned} \quad (4)$$

Throughout this paper, the parameter V is standardised so that it represents a multiple of the expected node pair distance variance. The interpretation of the standardised V is that $V=1$ is a neutral position that reflects the expected distance between nodes that are randomly located with density of 1 per unit area (ignoring the effect of focal points). The V parameter then indicates whether local edges are to be relatively more ($V < 1$) or less ($V > 1$) preferred, but has no absolute interpretation.

For the actual edge creation function, V_K must be calculated as:

$$V_K = V \frac{N^2}{6(N-1)} \times \left[1 - \frac{N}{6(N-1)} \left[\log(1 + \sqrt{2}) + \sqrt{2} \right]^2 \right] \quad (5)$$

That is:

$$V_K \approx 0.0203 NV \quad (6)$$

5 Experimental design

Four networks were constructed for each parameter set with between 3 and 7 values for each parameter as follows:

$N = 250, 500, 1000$ $f = 0.05, 0.1, 0.2$ $m = 0.1, 0.3, 0.5, 0.7$ $H = 0.4, 0.6, 0.8, 1, 1.5, 2$ $V = 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5$

Of the 6 048 experiments, 24 networks were not able to be created (that is, the giant component represented less than 90% of the nodes). These were all in the experiments with $N=250$ and $V=0.25$. Within this group, the algorithm was more likely to fail with a higher value for m and lower values for H and f .

For each network successfully created, several properties were calculated. These were mean degree, variance of degree (and coefficient of variation of degree), clustering coefficient and assortativity.

6 Impact of Parameters

The relationship between parameters and properties was investigated using summary tables, charts and correlations. Each property of interest (mean degree, degree variance, clustering coefficient and degree assortativity) was examined separately for each of the five algorithm parameters. Any relationship identified is, of course, only valid over the range of parameters considered. Results are summarised at Table 1.

The most straightforward relationship identified is that mean degree (denoted z) is approximately proportional to NHV , as shown in Figure 2. This is equivalent to proportionality between the probability that an edge is created and HV , as there are $N-1$ opportunities for each node to make edges.

A broad range of values was covered by the experiments, with mean degree ranging between 3 and 508. A linear regression model was fitted to the (natural) log transformed data. The best fit model ($p < 0.01$, $R^2 = 0.993$) is given by:

$$z = 0.148 N^{0.97} H^{0.89} V^{0.94}$$

The move proportion (m) has a consistent influence on mean degree in that mean degree increases as m increases for all parameter sets. However, except for the parameter sets with the smallest values of both N and V , the change in mean degree is very small across the range of m values. There is a slight

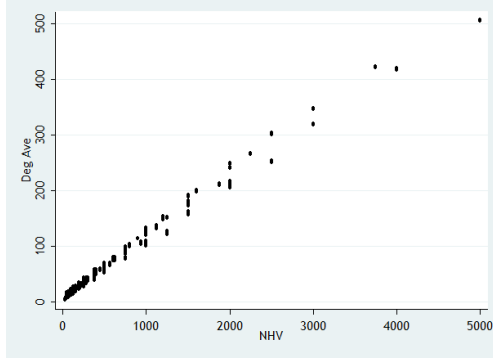


Fig. 2. Relationship between mean degree and the product of three algorithm inputs: nodes (N), height (H) and variance (V). Note that the values of N , H and V are constrained, so each point may represent many results from the 6 024 simulations plotted.

tendency for mean degree to decrease with increases in focal point density (f), but this relationship is inconsistent. While both of these parameters are statistically significant in a log transformed model ($p < 0.01$), the regression coefficients are less than 5% of the coefficients for the other parameters (which are of the same magnitude) and explanatory power is increased only slightly (to $R^2 = 0.994$). These parameters are therefore not included in the model.

Assuming proportionality, a useful rule of thumb is:

$$z \approx 0.1168 NHV \quad (7)$$

This rule underestimates mean degree at low parameter values and overestimates at high values. A different constant is therefore likely to be necessary for research using networks of more than 5 000 nodes.

The second network property considered is variance of degree (denoted $\text{var}(k)$). This property is highly correlated with mean degree (correlation coefficient of 0.87) and is proportional to NHV but with a smaller constant. Hence, any attempt to match a desired degree variance will also affect mean degree. Unlike mean degree, the focal point parameters (f and m) do have an impact so there is some capacity to tune degree variance. However, the variance of $\text{var}(k)$ is relatively high and the impact of these focal point parameters is inconsistent in both size and direction. Thus, there is limited capacity to control the algorithm to generate networks with specific $\text{var}(k)$ based on the sample size used. Instead, 'tuning' could be achieved by generating a large number of networks and selecting an appropriate subset.

The third property examined is mean clustering coefficient (denoted CC). Across the tested parameters, this coefficient showed a substantial range, from 0.09 to 0.85 (potential values are 0 to 1 inclusive). It is most strongly related

to only the H parameter, with a correlation coefficient of 0.91 and a value of approximately $0.30H$. As a result, clustering coefficient can be adjusted without changing mean degree by changing H to the appropriate level and then setting V so that HV remains constant.

Further tuning can be achieved with the f and m parameters, but only for low values of V (< 1). In this range, clustering coefficient can be increased by increasing the influence of focal points by decreasing f or, to a lesser extent, by increasing m .

Over the parameter values tested, assortativity ranged between 0 and 0.9, suggesting a broad scope of values are possible provided positive assortativity is desired. Assortativity increases with higher values of H and/or lower values of V . The focal point parameters (f and m) affect assortativity but the direction and size of their influence is inconsistent between various combinations of the other parameters.

Table 1
Summary of parameter impact on network property

Parameter	z	$\text{var}(k)$	CC	r
$\uparrow N$	\uparrow	\uparrow	-	-
$\uparrow H$	\uparrow	\uparrow	\uparrow	\uparrow
$\uparrow V$	\uparrow	\uparrow	-	\downarrow
$\uparrow f$	-	-	$V < 1 \downarrow$	any
$\uparrow m$	-	-	$V < 1 \uparrow$	any

In general, three of the input parameters (N , H and V) have both a substantial and predictable impact on all four considered network properties (mean degree, variation in degree, clustering coefficient and assortativity). The other two input parameters (f and m) affect some network properties but the effect is generally unpredictable and small, though a larger sample of networks may reveal a trend. Thus, there are insufficient degrees of freedom to generate networks with specific values of each potential network property.

However, using the main parameters of N , H and V , there is clear potential to generate networks with specific values of: number of nodes, mean degree and whichever is preferred of clustering coefficient and assortativity. Some tuning of the other property may also be available by varying f and m . Starting values can be obtained from any two equations from:

$$\begin{aligned}
z &= 0.1168 NHV && \text{(Mean degree)} \\
CC &= 0.3 H && \text{(Clustering)} \\
r &= 0.225 H - 0.046 V && \text{(Assortativity)}
\end{aligned} \tag{8}$$

7 Capacity to match published real world networks

Relevant properties from 27 published networks (of which 10 are social networks) have been collated by Newman (Table 10 of Newman, 2003). Of these, only the company directors network is undirected, has positive assortativity, fewer than 10 000 nodes and values for all properties. However, the degree distribution reflects the underlying bipartite structure and is substantially skewed (Newman et al., 2001, pg 411) so is not suitable as a potential match.

Three other networks were identified as potential match candidates: Physics coauthorship has relatively high values for both coefficients ($z=9.27$, $CC=0.56$, $r=0.363$); Biology coauthorship has (almost) median values for both coefficients, and a large numerical distance between the two coefficient values ($z=15.53$, $CC=0.60$, $r=0.127$); and email address books has relatively low values for both coefficients ($z=3.38$, $CC=0.13$, $r=0.092$).

While none of these networks met all conditions, they provide a range of real world mean degree, clustering coefficient and assortativity coefficient triples. The physics coauthorship network was selected to assess the capacity of the algorithm to match specific network properties, but reduced in size to 5 000 nodes. Initial parameter values were chosen to prioritise clustering coefficient or assortativity coefficient as two separate experiments.

The starting values for H and V are calculated using the relationships at (8) Arbitrary initial values of f (0.1) and m (0.5) were selected for all experiment series. Algorithm parameters and network properties for the sets of experiments are at Tables 2 and 3 respectively.

For the clustering coefficient prioritised experiment, initial parameter values were $N=5\ 000$, $H=1.87$, $V=0.0085$, $f=0.1$ and $m=0.5$. Ten networks were generated with these parameters and mean degree was higher than target. Parameters H and V were reduced until mean degree and clustering coefficient were reasonably close to target. Several alternative values of f and m were tested to improve assortativity targeting while maintaining mean degree and clustering coefficient but no improvements occurred.

During the assortativity focused experiment, it became clear that the iden-

tified relationships between parameters and properties break down at very small values of V . In particular, the clustering coefficient increases with V instead of having no significant association. Figure 3 shows this relationship for parameter V in the range 0.002 to 0.02 for various values of the H parameter.

Preliminary analysis indicates that the particular value of V at which the relationships break down depends on the number of nodes and other parameters (particularly H) as can be seen in Figure 3. One possible reason is that very low values of V lead to a high proportion of nodes with degree of 0 or 1 which, by definition, have a clustering coefficient of 0. A giant component is not able to be formed.

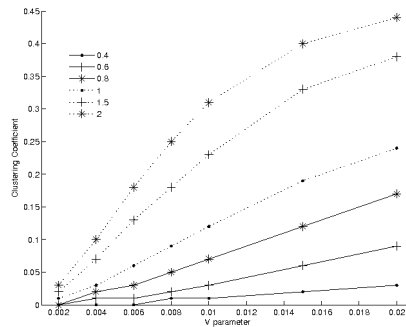


Fig. 3. Clustering coefficients where V parameter is very small, various values of H parameter. All generated networks have 1 000 nodes and are retained regardless of giant component size. Mean values plotted from sets of $f=0.05, 0.1$ and 0.2 ; $m=0.1, 0.3, 0.5$ and 0.7 ; 3 networks per parameter set.

Very small values of V are more likely to be required for sparse networks, where mean degree is low and/or number of nodes is high. With the limitation of low V values in mind, a further set of experiments attempted to match property values for the physics coauthorship network, but limited to 1 000 nodes (also displayed at Tables 2 and 3) to provide a greater effective range for V . This enabled networks to be generated that matched mean degree and assortativity.

Table 2

Physics coauthorship matching experiments: Inputs to algorithm

Experiment set	f	m	H	V
$N=5\ 000$, CC focus, initial	0.1	0.5	1.87	0.0085
$N=5\ 000$, CC focus, final	0.05	0.5	1.25	0.0054
$N=1\ 000$, CC focus, final	0.1	0.3	1.80	0.0395
$N=1\ 000$, r focus, final	0.1	0.3	1.00	0.066

The example network previously shown at Figure 1 uses the final algorithm parameters identified for the 1 000 node network with clustering coefficient

Table 3

Physics coauthorship matching experiments: Network properties

Experiment set	Runs	z	CC	r
Target properties		9.27	0.56	0.363
$N=5\ 000$, CC focus, initial	10	12.36	0.86	0.64
$N=5\ 000$, CC focus, final	50	9.40	0.57	0.52
- coefficient of variation		1.1%	0.7%	4.4%
$N=1\ 000$, CC focus, final	50	9.21	0.56	0.50
- coefficient of variation		1.9%	1.9%	10.1%
$N=1\ 000$, r focus, final	50	9.33	0.40	0.38
- coefficient of variation		1.8%	1.9%	15.7%

given priority.

8 Discussion

The Keeling algorithm is able to generate networks with a given number of nodes and mean degree (or degree variance), and either clustering or assortativity coefficient, with some flexibility in the third property. In general, assortativity and clustering coefficients are similar and there is only limited capacity to control these properties independently. In using the algorithm, care must therefore be taken to attribute an apparent effect to the appropriate property, particularly for the mean degree and degree variance pair, or assortativity and clustering pair. In addition, over the parameter range explored, only positive values of assortativity could be realised.

Further, although mean degree is controllable, the degree of each node arises from the algorithm input parameters. The nodes and focal points are distributed uniformly at random, and the movement and edge creation parameters apply equally to all nodes. Thus, the degree distribution is approximately binomial. The mean degree distribution calculated over 100 networks with the $N=1\ 000$, r focus, final parameters is at Figure 4 .

The algorithm is straightforward to implement and consistent in the properties of the networks produced with the same input parameters, particularly for mean degree and clustering coefficient. While assortativity is less stable, a high proportion of networks have values close to the target.

We identify relationships that allow fast selection of suitable initial parameters

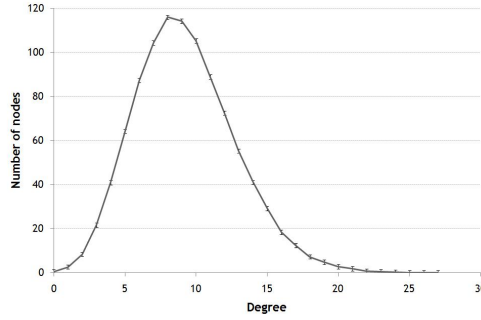


Fig. 4. Average degree distribution over 100 network instances with parameters $N=1\ 000$, $f=0.1$, $m=0.3$, $H=1.80$ and $V=0.0395$. Error bars indicate 95% confidence interval.

and parameter modifications to obtain particular property values if required. The impacts of the H and V parameters on the network properties are consistent and predictable. Initial values for these parameters can be selected with the appropriate pair of relationships from (8).

However, the paper has also identified a key limitation in the networks able to be produced. The algorithm becomes unpredictable with very low values of the V parameter. Further investigation is required to specify the boundaries of this unpredictability, but a potential boundary is where the proportion of nodes with degree 0 or 1 is significant or where the giant component makes up only a small proportion of the network. This limitation may make the algorithm impractical for simulating large networks with low mean degree.

This paper has focussed on the practical issues associated with using the algorithm to generate networks with specific properties. There are several interesting questions not explored here concerning other characteristics of the algorithm, such as the conditions under which a giant component is formed and the limiting behaviour at extreme values of the input parameters.

The other key area for further research is extension to control of the shape of the degree distribution. One potential approach is to assign target degrees to each node and incorporate the target degree into the edge creation function. A more skewed degree distribution may also maintain network connectivity at relatively low degree, thus also enabling the simulation of large networks with low mean degree.

Acknowledgement

The authors would like to thank the referees for detailed and constructive suggestions and Ruhul Sarker for several useful discussions. This research is primarily funded by a University College Postgraduate Research Scholar-

ship and supported by the ARC Centre for Complex Systems grant number CEO0348249.

References

- Barabási, A.-L. and R. Albert (1999). Emergence of scaling in random networks.
- Batagelj, V. and A. Mrvar (1998). Pajek - program for large network analysis.
- Becker, N. (1973). Carrier-borne epidemics in a community consisting of different groups.
- Eames, K. and M. Keeling (2002). Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases.
- Erdős, P. and A. Rényi (1960). On the evolution of random graphs.
- Eubank, S., V. Kumar, M. Marathe, A. Srinivasan, and N. Wang. Structure of social contact networks and their impact on epidemics.
- Gupta, S., R. Anderson, and R. May (1989). Networks of sexual contacts: implications for the pattern of spread of HIV.
- Jeong, H., S. Mason, A.-L. Barabási, and Z. Oltvai (2001). Lethality and centrality in protein networks.
- Kamada, T. and S. Kawai (1988). An algorithm for drawing general undirected graphs.
- Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions.
- Keeling, M. (2005). The implications of network structure for epidemic dynamics.
- Kermack, W. and A. McKendrick (1927). Contributions to the mathematical theory of epidemics - I.
- Lajmanovich, A. and J. Yorke (1976). A deterministic model for gonorrhoea in a nonhomogeneous population.
- Molloy, M. and B. Reed (1995). A critical point for random graphs with a given degree sequence.
- Newman, M. (2002). Assortative mixing in networks.
- Newman, M. (2003). The structure and function of complex networks.
- Newman, M. and J. Park (2003). Why social networks are different from other types of networks.
- Newman, M., S. Strogatz, and D. Watts (2001). Random graphs with arbitrary degree distributions and their applications.
- Nold, A. (1980). Heterogeneity in disease transmission modelling.
- Read, J. and M. Keeling (2003). Disease evolution on networks: the role of contact structure.
- Watts, D. and S. Strogatz (1998). Collective dynamics of 'small-world' networks.
- Waxman, B. (1988). Routing of multipoint connections.