

# Tuning degree distributions of scale-free networks

C.C. Leary<sup>1,2</sup>, M. Schwehm<sup>\*1</sup>, M. Eichner<sup>1</sup>, H.P. Duerr<sup>1</sup>

<sup>1</sup> Department of Medical Biometry, University of Tübingen, Westbahnhofstraße 55, 72070 Tübingen, Germany

<sup>2</sup> Department of Mathematics, State University of New York at Geneseo, 1 College Circle, Geneseo, NY 14454, USA

\* Corresponding author: [schwehm@informatik.uni-tuebingen.de](mailto:schwehm@informatik.uni-tuebingen.de)

(Received 20 February 2006)

## ABSTRACT

Scale-free networks are characterized by a degree distribution with power-law behavior and have been shown to arise in many areas, ranging from the World Wide Web to transportation or social networks. Degree distributions of observed networks, however, often differ from the power-law type and data based investigations require modifications of the typical scale-free network.

We present an algorithm that generates networks in which the skewness of the degree distribution is tuneable by modifying the preferential attachment step of the Barabási-Albert construction algorithm. Skewness is linearly correlated with the maximal degree of the network and, therefore, adequately represents the influence of superspreaders or hubs. By combining our algorithm with work of Holme and Kim, we show how to generate networks with skewness  $\gamma$  and clustering coefficient  $\kappa$ , over a wide range of values.

Keywords: scale-free networks, degree distribution, clustering, superspreaders, hubs, simulation, skewness, beta distribution

PACS numbers: 89.75 Hc, 87.23 Ge, 89.65 -s

## I. INTRODUCTION

Since the development of effective computational tools, intense research has broadened our understanding of the structure of networks. A fascinating mix of theoretical and observational work has pointed out differences and commonalities among networks ranging from the World Wide Web to transportation networks to social networks associated with sexually transmitted diseases [1-7].

Theoretical work on networks has shown that the underlying contact structure has a crucial role to play in the transmission of a disease across a network [8-13]. Scale-free contact networks, in which the probability that a vertex has degree  $k$  is proportional to  $k^{-\alpha}$ , have been shown to provide particularly fertile grounds for disease transmission. Investigations into the SARS outbreak of 2002-2003 have also served to increase our appreciation of the importance of superspreaders, individuals who are connected to an unusually large number of other individuals, in disease spread [14].

Watts and Strogatz introduced the clustering coefficient in their seminal 1998 paper [1]. For a vertex  $v$  with  $n$  neighbors, the clustering coefficient of  $v$  is defined to be

$\kappa_v = \frac{E_v}{{}_n C_2}$ , where  $E_v$  is the number of edges between the  $n$  neighbors of vertex  $v$

and  ${}_n C_2 = \frac{n(n-1)}{2}$  is the maximum number of edges possible among those  $n$

neighbors. Then  $\kappa$ , the mean clustering coefficient, is the average of  $\kappa_v$  across all vertices  $v$  in the network.

The Barabási-Albert preferential attachment algorithm generates scale-free networks, networks in which the probability that a randomly chosen vertex has degree  $k$  is proportional to  $k^{-\alpha}$ , with  $\alpha$  between 2 and 3 [15-17]. These networks include a heavy tail of high-degree vertices. As modified by Holme and Kim, among others [18-20], the Barabási-Albert algorithm can be used to create scale-free networks with a range of mean clustering coefficients. In this paper we outline an algorithm that generates networks in which the degree distribution is tuneable as well.

## II. METHODS

A contact network of 10,000 individuals was created using a generalization of the Holme and Kim [20] tuneable clustering version of the Barabási-Albert scale-free network generation algorithm. An initial collection of 8 vertices was connected in a cycle. After this, the remaining 9992 vertices were connected one at a time to the existing network. Vertex  $v$  was attached by first choosing a vertex  $w$  from the existing network using the preferential attachment scheme described below. Vertices  $v$  and  $w$  were connected (bidirectionally), and then  $v$  was connected to three other vertices chosen independently by either a triad formation step (with probability  $c$ ) or by the preferential attachment scheme (with probability  $1-c$ ). Construction was completed by revisiting the initial eight vertices and connecting each of them to three other vertices either by triad formation or preferential attachment.

The preferential attachment algorithm used was a generalization of that described by Barabási and Albert [15]. In the classic BA algorithm, the probability that a vertex  $w$  of degree  $k$  is chosen to be attached to vertex  $v$  is proportional to  $k$ . We altered this algorithm in the following way: The  $n$  vertices  $w_i$  in the existing network were ordered by their degree from largest to smallest, and then a random number  $r$  was chosen from a tuning distribution with range  $[0,1]$  and expected value  $\mu_T$ , as described below. Vertex  $v$  was then attached to vertex  $w_j$  if

$$\frac{\sum_{i=1}^{j-1} \text{deg}(w_i)}{\sum_{i=1}^n \text{deg}(w_i)} \leq r < \frac{\sum_{i=1}^j \text{deg}(w_i)}{\sum_{i=1}^n \text{deg}(w_i)}.$$

The tuning distributions used for Figs. 1-4 were simple linear distributions, with

probability density functions  $f_m(x) = \begin{cases} mx + (1 - \frac{1}{2}m) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$ , with expectation

$\mu_T = \frac{1}{2} + \frac{1}{12}m$ . The tuning distribution  $f_0$  with  $m=0$  and  $\mu_T = \frac{1}{2}$  is the uniform distribution on the interval  $[0,1]$ , and thus this parameter setting (with  $c=0$ ) corresponds to the original BA preferential attachment scheme. The tuning

distribution used for Fig. 5 was the beta distribution given by

$$f(x) = \frac{(x - x_l)^{\alpha-1} (x_u - x)^{\beta-1}}{B(\alpha, \beta) (x_u - x_l)^{\alpha+\beta-1}},$$

with parameters  $\alpha > 0$  and  $\beta > 0$ , in which  $x_l$  and  $x_u$  represents the lower and upper bound of the distribution, respectively (here:  $x_l = 0$  and  $x_u = 1$ ) and  $B(\alpha, \beta)$  denotes the beta function. The beta distribution offers more flexibility than the linear distributions and can produce almost all shapes, e.g. uniform ( $\alpha = 1$  and  $\beta = 1$ ), U-shaped ( $\alpha < 1$  and  $\beta < 1$ ), bell-shaped ( $\alpha > 1$  and  $\beta > 1$ ), right-skewed ( $\alpha > 1$  and  $\beta < 1$ ), or left-skewed ( $\alpha < 1$  and  $\beta > 1$ ).

The triad formation steps were used to induce clustering in the network. When vertex  $v$  underwent a triad formation step, a vertex  $w$  was chosen uniformly at random from among the set of vertices adjacent to  $v$ . Then a vertex  $u$  was chosen from the neighbors of  $w$  and vertex  $v$  and vertex  $u$  were connected (assuming that they were not already connected).

The algorithm was executed 100 times for each  $\mu_T \in \{0.4, 0.45, 0.5, 0.55, 0.6\}$  and each  $c \in \{0, 0.25, 0.5, 0.75, 1\}$ , yielding 2500 total networks for the analysis. In order to test the sensitivity of our results to the initial configuration of vertices, a further 2500 networks were generated beginning with a complete graph on nine vertices (which maintains an average degree of 8 in the network).

### III. RESULTS

We report results from the networks in which the initial eight nodes were connected in a ring. The results for networks that were seeded with a complete graph were similar.

Independent of the parameter settings, the 10,000 vertices in the network had an average of 8 contacts. As the tuning distribution used in the preferential attachment steps of the network creation algorithm influenced the amount of preference gained by high-degree vertices, altering this distribution allowed us to influence the degree distribution of the resulting network, as illustrated in Fig. 1. In the following analysis (i) the subscript  $T$  refers to a tuning distribution while a subscript  $D$  refers to the degree distribution of a network, (ii) we report the skewness of a distribution as  $\gamma^3$  and call its cube root  $\gamma$  the root skewness of the distribution and (iii) we refer to the degree distribution generated by the classic scale-free algorithm of Barabási and Albert ( $\mu_T = 1/2$ ,  $c = 0$ ) as the scale-free (SF) degree distribution. Realizations of 100 SF degree distributions show a normally distributed root skewness with mean  $\gamma_D = 2.26$ . Varying the clustering probability does not substantially change the skewness (Fig. 2), e.g. even for the maximum value of the clustering parameter ( $c=1$ ) the mean root skewness increases only to  $\gamma_D = 2.29$ .

#### A. Overskewness & underskewness

We call a degree distribution overskewed if the skewness exceeds that of the SF degree distribution, and underskewed if the skewness falls short of this standard skewness. Overskewed degree distributions were produced by applying tuning distributions with expectation  $\mu_T < 1/2$ , which make connections to vertices with high degree more likely (Fig. 1, row 1 and Fig. 2). Similarly, underskewed degree distributions were produced by applying tuning distributions with expectation  $\mu_T > 1/2$ , which make connections to vertices with few contacts more likely (Fig. 1, row 3 and Fig. 2). Hence, the root skewness of the degree distribution is inversely related to the expectation of the tuning distribution. The slope of this linear relationship strongly depends on the clustering (e.g. for  $c = 0$ :  $\gamma_D = -9.4\mu_T + 7$ , for  $c = 1$ :  $\gamma_D = -2.6\mu_T + 3.6$ ). Overskewness is in most cases associated with vertices of

very high degree, and those vertices can play the role of superspreaders in models of the spread of disease or computer viruses across a network.

The cube root skewness of the degree distribution  $\gamma$  is highly correlated with  $\sqrt[3]{D_{\max}}$  as well as with  $\log(D_{\max})$ , where  $D_{\max}$  is the degree of the vertex with highest degree in the network. For the networks under consideration here, the linear relationship is stronger using the logarithmic transformation, with least squares fit  $\log_{10}(D_{\max}) = 0.72 + 0.8\gamma_D$ . The coefficient of determination  $r^2 = 0.98$  indicates that  $\log(D_{\max})$  and root skewness can be used interchangeably to characterize the variability in the degree distribution of networks generated by our algorithm.

## B. Clustering

Holme and Kim [20] reported a strong linear correlation between clustering probability  $c$  and mean clustering coefficient  $\kappa$ , but the relationship shows a slight curvature which can be fitted more adequately by exponential curves of the form  $\kappa = ae^c + b$ , with estimates  $a \in [0.25, 0.28]$  and  $b \in [-0.18, -0.24]$  (Fig. 3).

The relationship between clustering and degree skewness is slightly more complicated and can best be understood by referring to the SF degree distribution. For underskewed degree distributions, increasing the clustering parameter increased the skewness of the degree distribution. For overskewed degree distributions the clustering parameter lowered the skewness of the degree distribution. Thus in all cases increasing the clustering parameter tended to make the resulting degree distribution more like the SF degree distribution, as indicated by the arrows in Fig. 2.

With this said, it is notable that the algorithm produced a variety of degree distributions with the same mean clustering, as shown in Fig. 4. To produce degree distributions with the same skewness and different mean clustering required the use of different tuning distributions.

## IV. DISCUSSION

We have exhibited an algorithm that produces contact networks with varying skewness and clustering. Investigations into real networks may demand other tuning distributions by which the degree distribution can be modified.

Using beta distributions, which can be uniform, left- or right-skewed, or even U-shaped, can provide more flexible tuning and produce a wide variety of degree distributions. Two examples, using symmetric beta distributions with expectation  $\mu_T = 1/2$ , are illustrated in Fig. 5. Such variations of the beta distribution can even lead to non-monotonic degree distributions. With a U-shaped beta distribution, for example, contacts are preferentially attached to vertices with lowest and highest degrees, increasing the prevalence of low-degree as well as the prevalence of high-degree vertices (superspreaders, hubs). Bell-shaped beta distributions, on the other hand, promote vertices with intermediate degree while preventing the occurrence of hubs. These examples show that beta tuning distributions should be capable of producing most degree distributions observed in real networks.

We have also shown that in a wide range of degree distributions, the skewness of the distribution is strongly correlated with the maximum degree of the network. Thus much of the information contained in the global measure (skewness) is already contained in a local measure (degree), although finding the value of the largest degree in an actual network may be very difficult.

As epidemiologists and other researchers continue to apply network techniques to a variety of areas, increased flexibility in network algorithms will be necessary. The results reported here provide a measure of that flexibility with little overhead.

## ACKNOWLEDGEMENTS

This work has been supported by EU projects SARScontrol (FP6 STREP; contract no. 003824) (HPD) and INFTRANS (FP6 STREP; contract no. 513715) (CCL, MS) and by the National Science Foundation (NSF 0436298) (CCL). The work was done partially while the author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2005. The visit was supported by the Institute.

Figure 1

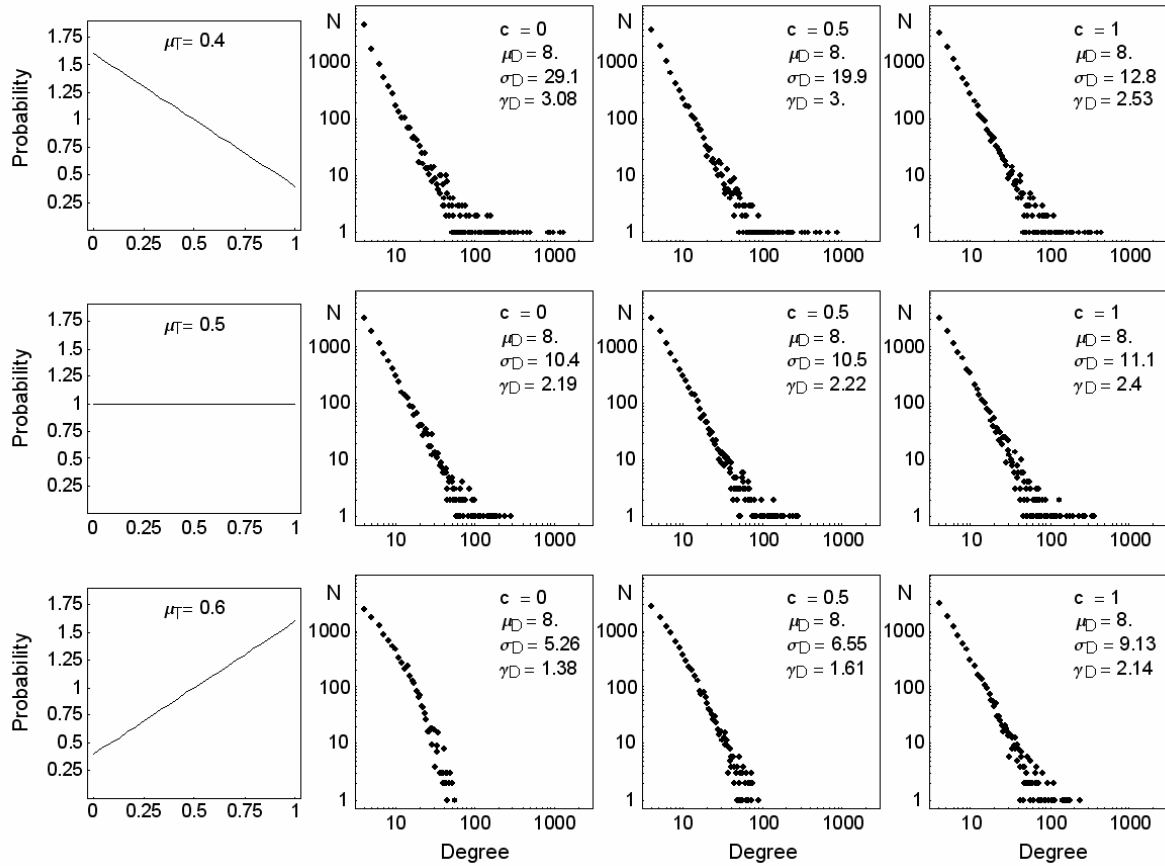


FIG. 1. Examples of degree distributions of different networks, for a subset of tuning distributions (column 1) and clustering parameters  $c$  (columns 2-4, with  $c=0$ ,  $c=0.5$ ,  $c=1$ ). Moments for the tuning distribution (subscript  $T$ ) and for the degree distributions (subscript  $D$ ) are represented by  $\mu$ : mean,  $\sigma$ : standard deviation and  $\gamma$ : cube root of the skewness. Scalefree degree distributions with power-law behavior result from the special case  $\mu_T = 1/2$ , representing a uniform tuning distribution. Underskewed degree distributions result from  $\mu_T > 1/2$  and overskewed degree distributions from  $\mu_T < 1/2$  (see also Fig. 2).

Figure 2

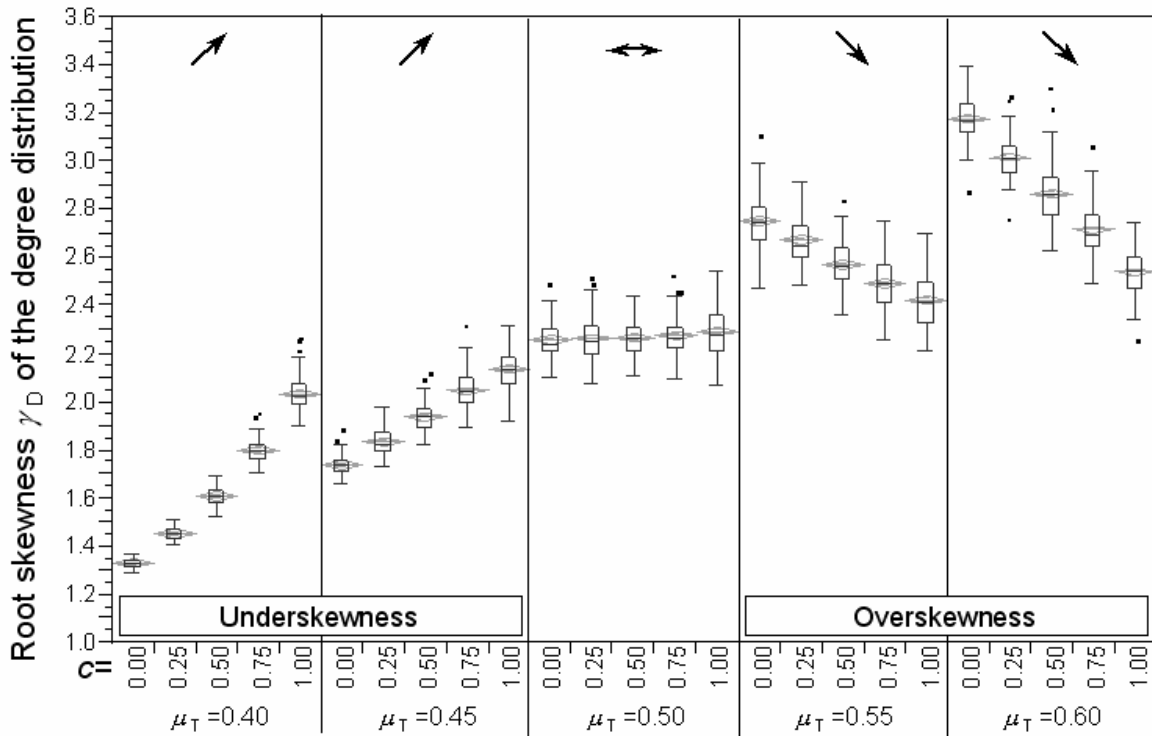


FIG. 2. Skewness distributions of different networks, grouped by the mean  $\mu_T$  of the tuning distribution and clustering parameter  $c$ . For each of the 25 parameter combinations, 100 network realizations were simulated. The skewness distribution of the corresponding degree distributions is represented by a box & whiskers plot, the mean  $\pm 1.5$  (standard error of the mean) (diamonds) and outliers (dots). Quartiles are represented by the lower and upper bounds of the box with the median as the line between them. Whiskers represent the quartiles  $\pm 1.5$  (interquartile range) and outliers are defined to lie outside this range. The skewness distributions are largely symmetric, so means and corresponding medians are close to each other. SF degree distributions with power-law behaviour result from the special case  $\mu_T = 1/2$ , representing a uniform distribution. Underskewed degree distributions result from  $\mu_T > 1/2$  and overskewed degree distributions from  $\mu_T < 1/2$  (cf. Fig. 1). Arrows in the upper part of the graph indicate the effect of clustering on the skewness of the degree distribution. Clustering shifts the skewness towards the standard, i.e. it increases the skewness of underskewed degree distributions (ascending arrows) and it decreases the skewness of overskewed degree distributions (descending arrows).

Figure 3

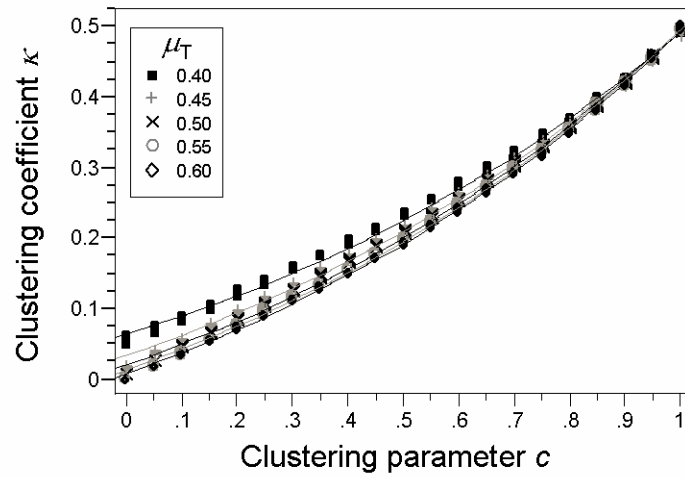


FIG. 3. Relationship between  $\kappa$ , the mean clustering coefficient and  $c$ , the clustering parameter. For each value of  $c \in \{0, 0.05, 0.1, \dots, 1\}$  and each value of  $\mu_T$ , 10 networks were generated. Tuning distributions with differing expectation  $\mu_T$  are indicated by differing markers as shown in the inset. Exponential least squares fits are grouped by  $\mu_T$ .

Figure 4

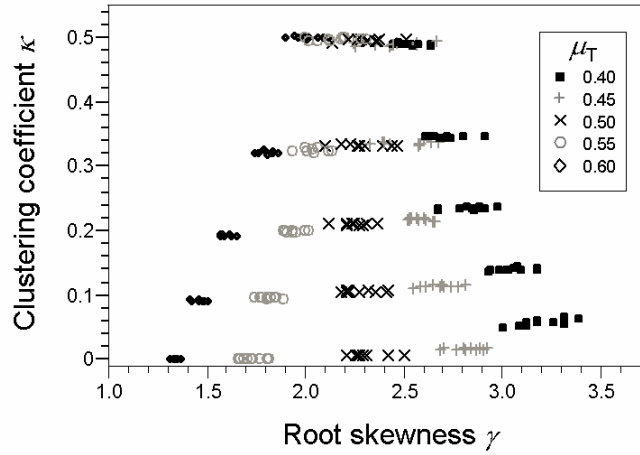


FIG. 4. Relationship between  $\kappa$ , the mean clustering coefficient and  $\gamma$ , the root skewness of the degree distribution, over 250 networks generated using the parameter values described in the main text. The horizontal bands correspond to the different values of the clustering parameter  $c \in \{0, 0.25, 0.5, 0.75, 1.0\}$  (from bottom to top). Tuning distributions with differing expectation  $\mu_T$  are indicated by differing markers as shown in the inset.

**Figure 5**

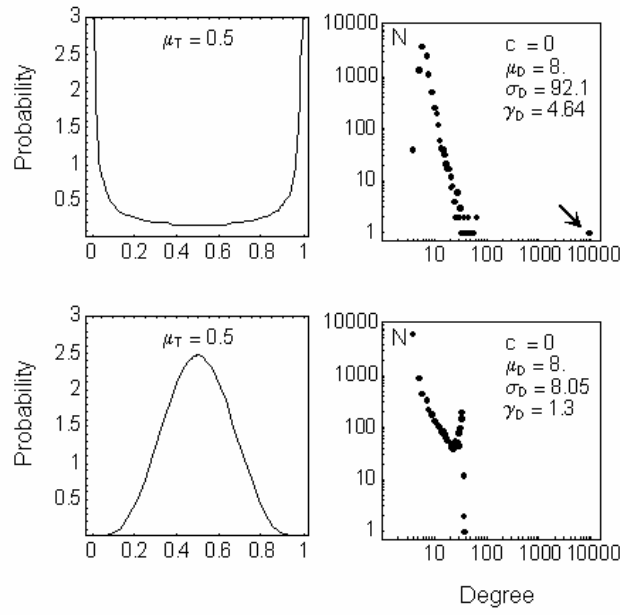


FIG. 5. Examples for degree distributions (right column) of different networks, using two beta distributions (left column) as tuning distributions with parameters  $\alpha = \beta = 0.1$  (upper row),  $\alpha = \beta = 5$  (lower row) and  $c = 0$ . Moments for the tuning distribution (subscript  $T$ ) and for the degree distributions (subscript  $D$ ) are represented by  $\mu$ : mean,  $\sigma$ : standard deviation and  $\gamma$ : cube root of the skewness. The superspreader (hub) is indicated by an arrow.

## REFERENCES

- [1] D.J. Watts and S.H. Strogatz, *Nature* **393**, 440 (1998).
- [2] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, and S.S. Manna, *Phys. Rev. E* **67**, 036106(R) (2003).
- [3] M.E.J. Newman and J. Park, *Phys. Rev. E* **68**, 036122 (2003).
- [4] V. Latora and M. Marchiori, *Physica. A* **314**, 109 (2002).
- [5] R. Guimera, S. Mossa, A. Turttschi, and L.A.N. Amaral, *P. Natl. Acad. Sci. USA* **102**, 7794 (2005).
- [6] F. Liljeros, C.R. Edling, L.A. Amaral, H.E. Stanley, and Y. Aberg, *Nature* **411**, 907 (2001).
- [7] A. Schneeberger, C.H. Mercer, S.A. Gregson, N.M. Ferguson, C.A. Nyamukapa, R.M. Anderson, A.M. Johnson, and G.P. Garnett, *Sex. Transm. Dis.* **31**, 380 (2004).
- [8] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, *J. Theor. Biol.* **235**, 275 (2005).
- [9] Z. Dezso and A.L. Barabasi, *Phys. Rev. E* **65**, 055103(R) (2002).
- [10] R.M. May and A.L. Lloyd, *Phys. Rev. E* **64**, 066112 (2001).
- [11] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **26**, 521 (2002).
- [12] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [13] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 035108(R) (2002).
- [14] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, and W.M. Getz, *Nature* **438**, 355 (2005).
- [15] A.L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [16] R. Albert and A.L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
- [17] P.L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [18] K. Takemoto and C. Oosawa, *Phys. Rev. E* **72**, 046116 (2005).
- [19] E. Volz, *Phys. Rev. E* **70**, 056115 (2004).
- [20] P. Holme and B.J. Kim, *Phys. Rev. E* **65**, 026107 (2002).