PARAMETRIZING RANDOM TOPOLOGY GENERATING FUNCTIONS

Thorsten Kisner, Firoz Kaderali
Department of Communication Systems
Faculty of Mathematics and Computer Science
FernUniversität in Hagen, Germany
{thorsten.kisner,firoz.kaderali}@fernuni-hagen.de

Helge Hemmer
Department of Neurocomputing
Ruhr-Universität Bochum, Germany
helge.hemmer@ruhr-uni-bochum.de

ABSTRACT

Random graphs are a popular method to model the interconnections in different domains, ranging from social networks, IP networks up to peer-to-peer networks. For modelling and simulation often different distinct graphs are required, which must have a minimum number of links. By choosing graph generation paramaters for random graphs resulting into a low number of links, the risk arises not to generate a connected graph, which is inapplicable for further simulations. After demonstrating the question of choosing convenient parameters for the graph generation process using random graphs as an example, we concentrate on a popular class of random graphs introduced by Waxman. We show that other investigations in literature do not comply with our constraints and propose simple manageable approximations of each characteristic for random and Waxman graphs, which can be used practically to parametrize the graph generation process.

KEY WORDS

Modelling, Random Graphs, Waxman Graphs, Connectivity, Simulation

1 Introduction

During the generation of network structures consisting of nodes and edges, the interconnection process between specific nodes is a fundamental requirement for a number of fields in science, commonly controlled by a well known rule-set. In this paper, we focus on special types of networks which are represented by random graphs, where the nodes are connected to each other by a given probability or a probability function. A lot of methods have been published to create graph topologies with statistical link probabilities. The basic random graph is described by Erdös and Renyi as a statistical process in which the points in the graph are connected step by step [6]. Each edge has the same probability and in each step (unless the points are not connected yet) a random number generator is used to evaluate the probability. We demonstrate the challenge of choosing the best parameters with the constraint of a connected graph and minimum links in this generation process and extend our results to another generation process described by Waxman [21]. Waxman uses a method in which the distance between each node pair affects the probability of being connected.

The question we are trying to answer is the following: *How have the parameters of the generation process to be chosen in order to achieve a connected graph of* n *nodes whilst using a minimum number of edges?*

This question is not purely academic because there are many fields of research in which random graphs are used and our results can aid those dealing with these kinds of topologies to understand the effects the parameters have on the generation process in creating topologies which are easier to compare to each other. For example, a lot of real world networks are successfully modelled on random graph generation processes yet there is still a lot of research to be carried out in this field.

For instance, [16] examined the properties of social network structures by using a generalized Erdös Renyi random graph. Although this piece of work concentrates on social networks, the authors are convinced, that many other real world networks could be modelled in the same way. A more specific analysis carried out Malarz et. al. closely studied gossip in social networks [13], while the effects of clustering in an underlying social network is analyzed. The spread of diseases is addressed in [3]. The network model used for this is a random intersection graph, where a node represents a person, every one belonging to a number of groups, and a link can only be established between people who are in at least one predefined group.

Far away from topics mentioned before, random connected models are also used in the area of neuroscience. In [12] biologically motivated neural structure are used for detection of signalling between neural fields. In this particular model, there are two types of columns, macro- and minicolumns. The neurons of the minicolumns are randomly connected to each other in the first simulation step.

Another field of research in which random graphs are used is the modelling and simulation of sensor networks, often with the focus of self-organizing and routing aspects. In [11] a scenario with a high number of cheap sensors, connected to each other by a wireless ad-hoc network infrastructure is described where these sensors are dropped out of a plane to monitor the environment. It becomes clear that random networks are a good choice to model the resulting topology.

Random graphs are also very important in the modelling of communication networks such as the internet, ad-hoc networks and P2P systems. For an overview of applications using random graphs please refer to [18].

In the context of communication infrastructures, the aspect of IT security profits in many ways from the detailed and realistic simulation of the underlaying network, especially when computer viruses [5] or distributed denial of service attacks are taken into consideration, which are investigated and attempts are made to prevent them. The simulations of both viruses which spread and denial of service attacks can be based on random graph topologies [10].

The performance of networks is investigated in [8], where the authors present another point of view on random graph generating processes. The results are quite interesting; Routing algorithms which use the 'geometric distance method', registered an increase in performance after adding a few random links. In contrast, the performance is decreased after adding random links using 'shortest path metric' routing.

As you can see, random graph structures can be found in many different areas of research, for the most part when connected systems are being investigated. We propose a model to define the parameters for two different graph generation processes keeping in mind the minimum number of links. We are dealing with probabilities which means that there will be connected graphs with a lower number of links using parameters below our presented threshold. However, in many cases, especially when the generation of hundreds or thousands of topologies are needed for test runs, the generation of a connected graph has to be ensured.

2 Related Work

The aforementioned random graphs are, as a result of their great significance, subject to a number of studies. Rate equations known from statistical physics are used in [1] to discuss the extremal properties of random structures e.g. trees, graphs and networks. In [17] classical random graphs are compared to scale-free networks in a performance analysis of a TCP/IP network, resulting in an improved performance for the random structure when traffic is high. A higher number of routers ensure a shorter way between two nodes in a random structure which provides a more balanced trafficload.

XinPing [22] introduces a network generation method that creates a graphs from a predefined degree distribution. Classic random graphs generally use a Poisson degree distribution, which are not always a good approximation of real networks because they often follow a power-law degree sequence. The robustness when faced with network failures is addressed by Newman et. al. in [4], by exploring the flow on random graphs. Their work also leaves the traditional path of Poisson degree distributions and observes graphs with general distribution of degrees. They stated that while power-law networks like the internet are quiet robust with regard to the extraction of randomly picked nodes, they are, however, fragile when dealing with the removal of nodes with many connections.

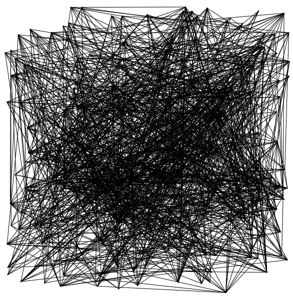
The subject of this paper is closely related to research carried out by Naldi [15], where the connectivity of Waxman-type network structures is investigated for different side ratios of the base area (we are using a quadratic area, see figure 1) and a model for the density distribution of the link lengths is proposed.

When observing networks generated on the basis of the Waxman algorithm, the problem of chosing values for the constants α and β (see eq. 4) arises, because there is no 'standard configuration' for this. While Waxman himself used to set both α and β to 0.4, others worked with very different parameter sets or ranges. For example Zegura et. al. [23] encountered the same problem which we would like to address and simply have chosen α = 0.2 and β = 0.15 with no further explanatory statements. In order to solve these kinds of problems, we present a mathematical relationship between the values of α and β in order to find the optimal values under the constraints of a connected graph and the minimum number of links.

3 Generators

Topology genrators are mainly proposed [14] to generate structures which represent the internet topology and can be distinguished into three special classes: *Random Generators* link two nodes dependent on a random function, *Structural Generators* benefit from the hierarchical properties of the internet and try to recreate this structure and *Degree-based Generators* which reproduce the degree distribution of all nodes, i. e. the probability of a node having a certain numbers of links.

In this paper we deal with random generators and focus on a simple random graph generation function [6] and the Waxman generator [21]. Figure 1 shows exemplary graphs for both types of generation[1].


(a) Pure Random Generator

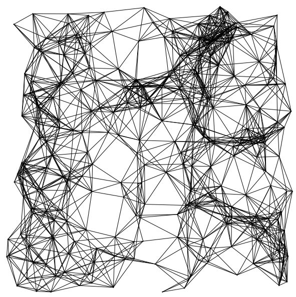

(b) Waxman Generator

Figure 1: Examples of Random Graphs

3.1 Random Generator

One of the simplest ways to create a random graph is to place the desired number of nodes on a base area with a

[1]With 250 nodes and following parameters: a) Pure Random: $p = 0.04$ b) Waxman: $\alpha = \beta = 0.5, L_{max} = 10000$.

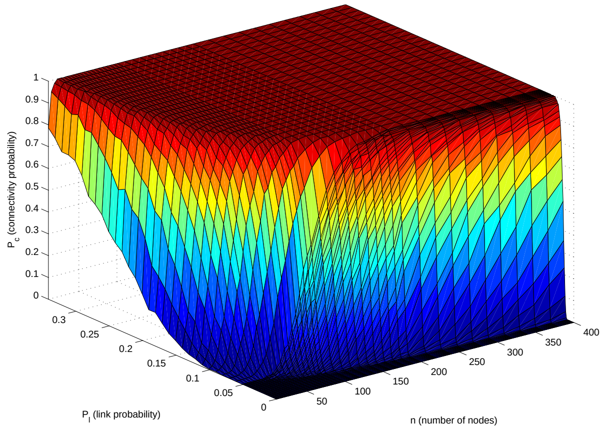


Figure 2: Connectivity Probability of Random Generator

uniform distribution of their positions. Edges between all pairs of nodes are added with a given probability P_l. The number of edges is straight proportional to P_l and quadratic to the number of nodes, which limit the maximum possible edges to $\frac{n(n-1)}{2}$.

Gilbert introduced a recursive algorithm to compute the connectivity probability for homogeneous graphs [9], the algorithm is initialized with $f_1 = 1$.

$$f_n = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} f_k \cdot (1 - P_l)^{k(n-k)} \tag{1}$$

The approximation $f_n \approx 1 - n(1 - P_l)^{n-1}$ of (1) is only valid for a large link probability and a large number of nodes and can not be used for our problem because we get very low probabilities for a large number of nodes.

Unfortunately this recursive algorithm is not easy to compute for a large number of nodes. The first problem is the binomial coefficient

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)! \cdot (n-k))!} = \prod_{i=1}^{k-1} \frac{n-i}{i} \tag{2}$$

which results in a very large value. This value is multiplied with the very low value $(1 - P_l)^{k(n-k)}$. Even double-precision 64 bit floating-point computation [19] result in rounding errors, which are fatal for the recursive algorithm.

In order to evaluate the recursive algorithm for high n, special classes like MPFR[2] [7] or `java.math.BigDecimal` need to be used, which provides decimal numbers of arbitrary precision and correct rounding. This (required) precision has some disadvantages, the most significant being the calculation time. For $n > 5000$ it took hours on actual desktop computer hardware[3] to calculate the probability of a connected graph for a given P_l. Even if this probability can be calculated

[2]Multiple-Precision Binary Floating-Point Library With Correct Rounding

[3]Intel Core 2 Duo 2,6 GHz, 4GB RAM

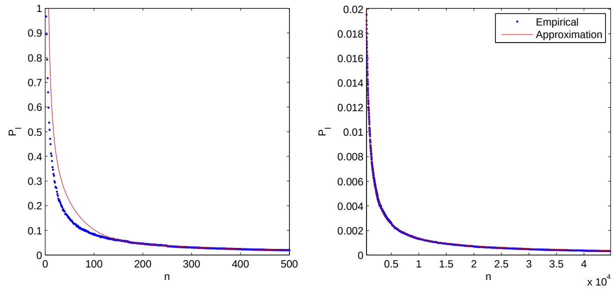


Figure 3: Characteristic of $P_c = 0.99$ with approximation

we still do not know if this is the minimum P_l needed to generate a connected graph. Our question is not *What is the connectivity probability P_c for a given P_l* but *What minimum P_l must be chosen in order to achieve a graph which is connected with the probability P_c*?

An empirical evaluation of the connectivity for various numbers of nodes n and edge connection probability is shown in figure 2. For $n > 200$ the abrupt slope of the connectivity probability P_c going from $0 + \delta$ to $1 - \delta$ with $\delta \rightarrow 0$ is remarkable and is pointed out in [2] as typical for random graph models.

In order to generate a graph with the minimum number of links which is, nevertheless, still fully connected the upper edge of this slope must be looked at. Figure 3 shows this edge for a number of nodes $2 < n < 4.5 \cdot 10^4$ and $P_c = 0.99$ and two different scales of the y-axis to get a better impression of the data. The development of these values seems to have a a power-law behavior $f(x) \sim x^\beta$ and even with the simple power-law distribution $f(x) = \alpha \cdot x^\beta$ we were able to obtain good approximation results. The addition of a Gaussian term led to an even better approximation for higher values, but we have to accept the disadvantage of a poor approximation for $n < 100$.

$$f_{\alpha_1,\beta,\alpha_2,\vartheta,\tau}(n) = \alpha_1 \cdot n^\beta + \alpha_2 \cdot e^{\vartheta(n-\tau)^2} \tag{3}$$

There is an almost perfect fit for the values $n > 100$, the curves on the right of figure 3 are almost identical for empirically estimated data and for our approximation. The values of each coefficient of eq. 3 are given in table 1.

Parameter	Value
α_1	5,3413448
β	-0,8985608
α_2	0,0856744
ϑ	-0,0001659
τ	2,4613273

Table 1: Values for approximation (eq. 3) in figure 3

With these coefficients and our approximation function $f_{\alpha_1,\beta,\alpha_2,\vartheta,\tau}(n)$ it is possible to choose a low link probability P_l resulting in a lower number of links but nevertheless still have a connected graph with probability P_c.

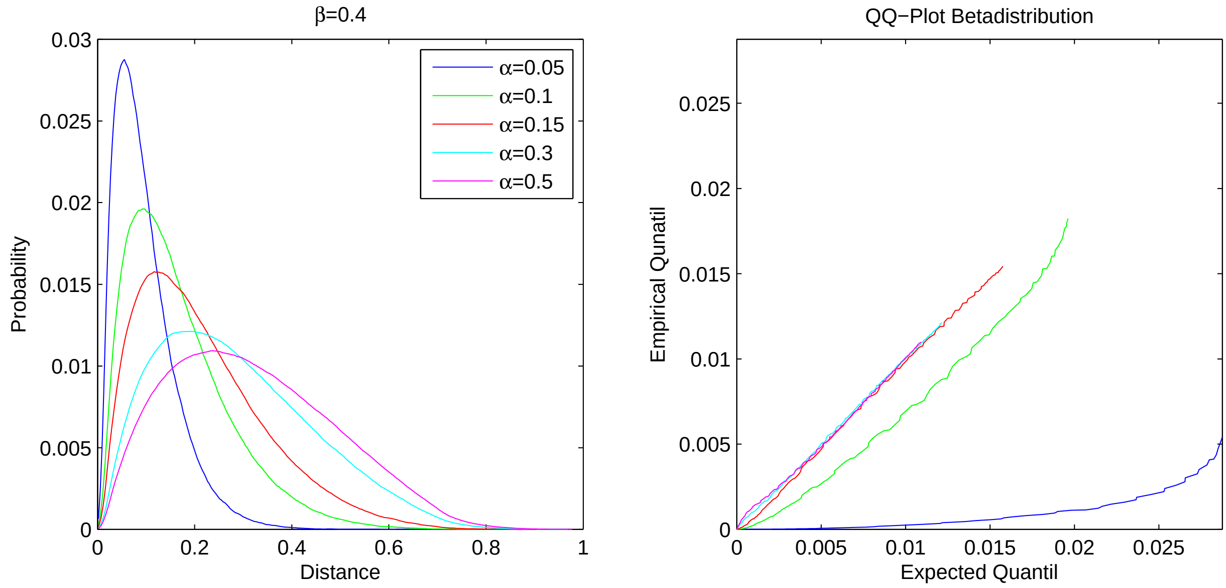


Figure 4: Link length distribution and QQ-Plot for corresponding Beta Distribution

3.2 Waxman Generator

In [21] Waxman proposed a model to generate network topologies. In his model nodes are distributed uniformly on a base area. The maximum possible distance is L and is determined by the geometric shape of the base area. With the probability $P(u, v)$ edges are created between the node pairs u and v.

$$P(u, v) = \beta \cdot e^{\frac{-d(u,v)}{\alpha L}} \tag{4}$$

The probability of two nodes u and v being connected depends on the Euclidean distance $d(u, v)$ between them and is influenced by α and β, where β with $0 < \beta \leq 1$ directly affects the link density. The ratio between short and long links can be controlled by α (see figure 4 on the left) in the range of $0 < \alpha \leq 1$. The number of links for $n = 5000$ is demonstrated in figure 5 for the allowed range of α and β. The number of links ranges upto a factor of 10, which does not seem very much, however there are many graph properties and metrics which have a quadratic complexity in calculation time [20].

Naldi [15] proposes the Beta Distribution $d_{\alpha,\beta}(x)$ to model the probability density function (pdf) of the distances between each pair of nodes.

$$d_{\alpha,\beta}(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{5}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du \tag{6}$$

With eq. 4 and eq. 5 the probability P_l for two generic nodes to be connected is estimated to be

$$P_l = \int_0^1 P(u, v) \cdot d_{\alpha,\beta}(x) dx \tag{7}$$

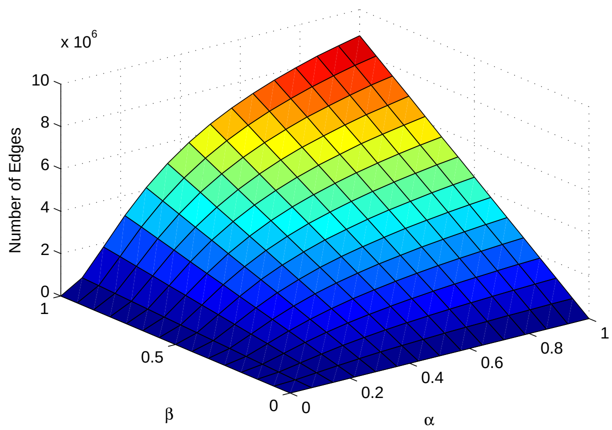


Figure 5: Number of edges in Waxman graph ($n = 5000$)

which can be resolved to $P_l = \beta \cdot \Theta(\alpha; \alpha + \beta; -1/\alpha)$. These assumptions are carried out in [15] to analytically obtain the connectivity of a Waxman graph for different side ratios of the base area (we only use a quadratic base area). The Beta Distribution is assumed as suitable and Quantil-Quantil (QQ) plots with a near linear gradient of the expected and empirical quantil are given to prove the assumption.

We also evaluated the Beta Distribution, but varied α (the pdf is independent of β and n). Figure 4 shows the pdf of the distance and we agree with [15] that the Beta Distribution is a very good approximation for the pdf of the node distances, but only for $\alpha > 0.15$. For lower values the Beta distribution becomes inferior and is really not suitable for $\alpha < 0.1$. We are interested in a low number of links (resulting in low α and β, see figure 5) which means that we can not utilize the results of [15], which finally uses the

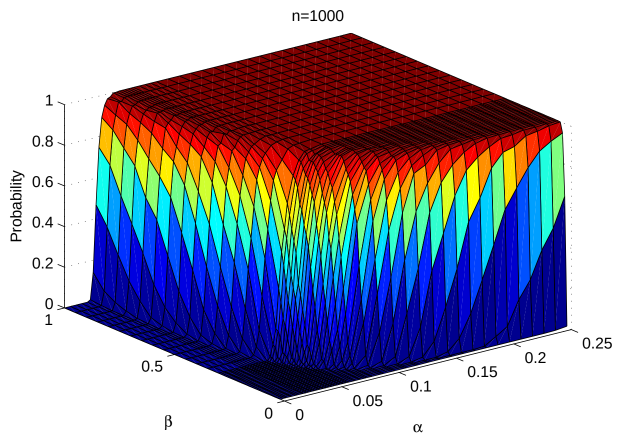


Figure 6: Connectivity Probability of Waxman graph

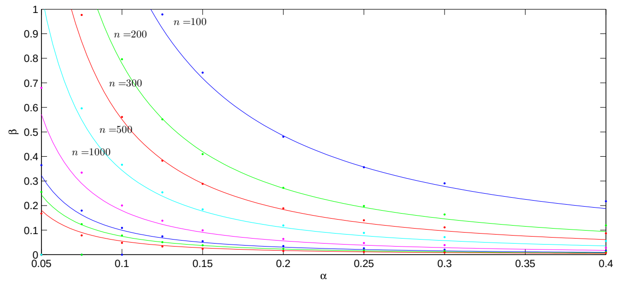


Figure 7: Connectivity of Waxman graph with $P_c = 0.99$

unwanted recursive algorithm (eq. 1) of Gilbert [9].

As for random graphs, the distribution of the upper edge of the connectivity probability of Waxman graphs can be described by a power law relationship between α and β

$$\beta(\alpha) = \rho \cdot \alpha^{\sigma} \tag{8}$$

and is shown in figure 8 for different n. The values of ρ and σ, derived from the empirical determined relationship between α and β in fig. 7, are summarized in table 2.

The next step is the evaluation of the characteristics of ρ and σ depending on the number of nodes. Figure 8 shows a plot of the values from table 2. We identified again a power-law distribution, but added μ as a third parameter

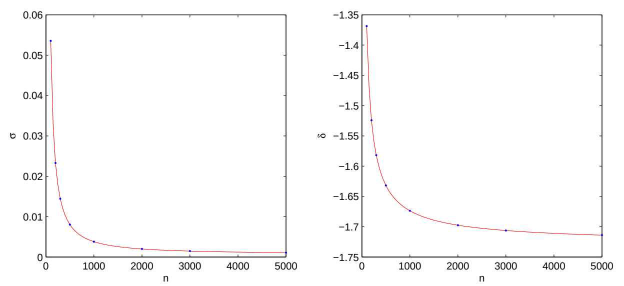


Figure 8: Characteristic for α and β dependent on n

n	ρ	σ
100	0,053536	-1,368692
200	0,023282	-1,524131
300	0,014424	-1,581728
500	0,008017	-1,631808
1000	0,0038	-1,673658
2000	0,001995	-1,697441
3000	0,001467	-1,706253
5000	0,001063	-1,713916

Table 2: Values of approximation (eq. 8) in figure 7

to get better results.

$$f(n) = \kappa \cdot n^{\lambda} + \mu \tag{9}$$

This equation is the beginning of our modelling process with the parameters κ, λ and μ defined in table 3.

	κ	λ	μ
ρ	14,831458	-1,223871	0,000637
σ	15,380093	-0,815294	-1,728748

Table 3: Parameters of eq. 9 to retrieve ρ and σ

How should these results be interpreted? If a Waxman graph with a minimum number of links as well as with a high probability of being fully connected ($P_c = 0.99$) has to be generated with n nodes, ρ and σ must be computed using eq. (9) with the values of table 3. After choosing the desired link length distribution (figure 4) with α, the corresponding β can be computed using eq. 8.

4 Conclusion and Future Work

We empirically evaluated the connectivity of about $9 \cdot 10^6$ unique random and about $12 \cdot 10^6$ unique Waxman graphs to obtain the connectivity characteristics for different values of P_l, α and β for several n. With these characteristics we derived two models to obtain the ceiling for each characteristic for a specified connectivity probability P_c. We demonstrated that the Beta distribution can not always be used to describe the link length distribution for Waxman graphs.

For a random graph the probability P_l for the graph generation process can be directly obtained by our proposed model for a particular n. The values for the two parameters α and β of the Waxman graph are obtained indirectly. For a particular n the characteristic curve of β which is dependent on α can be calculated. After choosing α, the corresponding β can be easily read from the characteristic curve. Using these derived values for the graph generation process ensures the minimum number of links while maintaining the probability P_c of a connected graph.

Future work will look at the improvement of our model with the comprehension of a variable parameter

P_c. Samples have confirmed the assumption that the characteristic of the connectivity of the the Waxman graph will change, but we are surprised that this change is only marginal. We are optimistic that further investigation will be comprise a varibale P_c in the model, so far P_c is fixed.

We are also experimenting with different derivations of a Waxman graph. For example, placing the nodes not only on a rectangular base area but also on a sphere surface where the distance between two points is a part of the greater circle and results in a more balanced node distance distribution without the tendency toward a long-tailed distribution. Another derivate is a combination of Waxman graphs and random graphs we called p-Waxman where a link is only created if the distance is below a geometry independent L_{max} and the probability below a threshold p.

References

[1] E. Ben-Naim, P. L. Krapivsky, and S. Redner. Extremal Properties of Random Structures. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors, *Complex Netwokrks*, volume 650 of *Lecture Notes in Physics, Berlin Springer Verlag*, pages 211–233, 2004.

[2] Bela Bollobas. *Random Graphs*. Cambridge University Press, 2001.

[3] T. Britton, M. Deijfen, A. Nordvall Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *ArXiv e-prints*, 708, August 2007.

[4] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85(25):5468–5471, Dec 2000.

[5] Moez Draief, Ayalvadi Ganesh, and Laurent Massoulié. Thresholds for virus spread on networks. In *valuetools '06: Proceedings of the 1st international conference on Performance evaluation methodolgies and tools*, page 51, New York, NY, USA, 2006. ACM.

[6] P. Erdös and A. Renyi. On random graphs i. *Publ. Math.*, Debrecen 6:290–297, 1959.

[7] Laurent Fousse, Guillaume Hanrot, Vincent Lefèvre, Patrick Pélissier, and Paul Zimmermann. Mpfr: A multiple-precision binary floating-point library with correct rounding. *ACM Trans. Math. Softw.*, 33(2):13, 2007.

[8] Henryk Fuks and Anna T. Lawniczak. Performance of data networks with random links. *Mathematics and Computers in Simulation*, 51:103, 1999.

[9] E. N. Gilbert. Random graphs. *Annals of Mathematical Statistic*, 30:1141–1144, 1959.

[10] Y. Kim and A. Helmy. ATTENTION: ATTackEr traceback using MAC layer abNormality detecTION. *ArXiv Computer Science e-prints*, August 2005.

[11] R. Krishnan and D. Starobinski. Message-efficient self-organization of wireless sensor networks. In *IEEE Wireless Communications and Networking Conference*, pages 1603– 1608 vol.3, 2003.

[12] J. Lücke and C. von der Malsburg. Rapid correspondence finding in networks of cortical columns. In *Proc. ICANN*, volume 4131 of *LNCS*, pages 668 – 677. Springer, 2006.

[13] K. Malarz, Z. Szvetelszky, B. Szekf, and K. Kulakowski. Gossip in Random Networks. *Acta Physica Polonica B*, 37:3049–+, November 2006.

[14] Alberto Medina, Anukool Lakhina, Ibrahim Matta, and John Byers. Brite: An approach to universal topology generation. In *International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS'01)*, Cincinnati, Ohio, USA, August 2001.

[15] Maurizio Naldi. Connectivity of waxman topology models. *Computer Communications*, 29(1):24–31, 2005.

[16] M. E. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proc. National Acadademy of Sciences U S A*, 99 Suppl 1:2566–2572, February 2002.

[17] G. Peli and G. Papp. Are Scale Free Networks Better? *ArXiv Condensed Matter e-prints*, January 2003.

[18] H. Reittu and I. Norros. Random graph models of communication network topologies. *ArXiv e-prints*, 712, December 2007.

[19] David Stevenson. Ieee task p754: A proposed standard for binary floating-point arithmetic. *Computer*, 14(3):51–62, 1981.

[20] Hongsuda Tangmunarunkit, Ramesh Govindan, Sugih Jamin, Scott Shenker, and Walter Willinger. Network topologies, power laws, and hierarchy. *SIGCOMM Computer Communication Review*, 32(1):76–76, 2002.

[21] Bernard M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, 1988.

[22] Xu Xinping. Connection probability for random graphs with given degree sequence. *ArXiv e-prints*, 708, August 2007.

[23] Ellen W. Zegura, Kenneth L. Calvert, and Samrat Bhattacharjee. How to model an internetwork. In *IEEE Infocom*, volume 2, pages 594–602, San Francisco, CA, March 1996. IEEE.