© 2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purpose or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This is the author's version of the work. The definite version is published in Proc. IEEE International Conference on Communications (ICC), Cape Town, South Africa, May 23-27, 2010. ieeexplore.ieee.org.

On the Degree Distribution of k-Connected Random Networks

Christian Bettstetter ^{1,2}, Johannes Klinglmayr ¹, and Stefan Lettner ¹

¹ University of Klagenfurt, Mobile Systems Group, Institute of Networked and Embedded Systems, Austria

² Lakeside Labs GmbH, Klagenfurt, Austria

Abstract—In performance evaluations of communication and computer networks the underlying topology is sometimes modeled as a random graph. To avoid unwanted side effects, some researchers force the simulated topologies to be connected. Consequently, the resulting distribution of the node degrees does then no longer correspond to that of the underlying random graph model. Being not aware of this change in the degree distribution might result in a simulation pitfall. This paper addresses the question as to how serious this pitfall might be.

We analyze the node degree distribution in connected random networks, deriving an approximation for large networks and an upper bound for networks of arbitrary order. The tightness of these expressions is evaluated by simulation. The analysis of the distribution for large networks is extended to k-connected graphs.

Results show that specific restricted binomial distributions match the actual degree distribution better than the random graph degree distribution does. Nevertheless, the pitfall of not being aware of the change in the distribution seems not to be a serious mistake in typical setups with large networks.

Index Terms—Network theory, connectivity, network modeling, random networks, degree distribution, network simulation techniques, simulation pitfalls.

I. Introduction and Motivation

The topology of communication and computer networks is sometimes modeled as a random graph. Random graphs are in general constructed by starting with a set of nodes and adding links between node pairs according to some rule. A simple approach is to employ Erdős-Rényi graphs, where a link between any pair of nodes in the graph exists with a certain probability, and the existence of a given link is independent of the existence of other links (see [1], [2]).

Such random graphs capture an important feature of real-world networks: the small-world effect. They do not capture, however, other real-world effects, such as link correlations between adjacent nodes (needed e.g. to model wireless networks) and the existence of hubs (nodes with many links) observed in several types of networks including the Web. To model these and other structural properties of real-world networks, a variety of more refined models were defined (see [3], [4]).

Despite these advances in modeling the structure of real-world networks, Erdős-Rényi graphs remain a commonly used topology model in communications engineering. They are used to evaluate basic performance tradeoffs of new networking algorithms, and are employed as reference topologies when analyzing the impact of topology properties on certain algorithms. Topics under investigation include the spreading of worms and viruses in the Internet [5], search and replication in peer-to-peer networks [6], and probabilistic flooding [7].

An important topology property of a network is the degree distribution, i.e., the probability mass function of the number of links of a node. It has an impact on various network properties, such as connectivity, network resilience, capacity, and message propagation. The degree distributions resulting from various random graph models are well-known. The Erdős-Rényi model creates graphs with a binomial degree distribution; more sophisticated models often target at heavy-tail distributions, i.e., networks in which hubs occur with a non-zero probability.

To analyze the performance of algorithms and protocols, a typical approach is to create a set of random graphs in Monte-Carlo simulations and use these as underlying network topologies. To avoid unwanted side effects, some researchers force the generated networks to be *connected*, i.e., all disconnected topologies created by the random process are discarded in the simulation. The algorithm under investigation is run only on the connected topologies. Such simulation methodology is used in several fields of communications engineering. Examples include the analysis of wavelength requirements in optical networks [8], distributed algorithms in ad hoc networks [9], flooding and network coding [10], and traffic engineering [11].

The degree distribution conditioned to the fact that the network is connected is, however, no longer equivalent to the degree distribution of the random graph model. As a simple example, an Erdős-Rényi graph may have isolated nodes (nodes without links), while a connected Erdős-Rényi graph does not have even a single isolated node.

The goal of this paper is to make researchers aware of this potential pitfall and analyze the underlying problem using methods from stochastic graph theory. Questions are: What is the degree distribution of a *connected* random network? Is it a serious mistake to assume the degree distribution of the underlying random graph model? Focusing on Erdős-Rényi graphs, we (a) analyze the degree distribution in networks with many nodes and (b) derive an upper bound for the degree distribution in networks with an arbitrary number of nodes. The tightness of the expressions is evaluated by simulation. We finally extend the analysis to k-connected networks, taking into account robustness against failures.

II. BACKGROUND ON GRAPH THEORY

A graph G is a set of nodes connected by links. The number of nodes in G is the order n of the graph. An empty graph has n nodes but no links. A random graph is a graph generated by a random process, where a commonly used process is as follows [2]: starting with an empty graph with n

nodes, a link between any node pair is added with probability p and independent of other links. This model results in a random graph G(n,p) where each of the possible links occurs independently with probability p. Such a graph is commonly called $\operatorname{Erd\mathscr{S}s-R\acute{e}nyi}$ graph or simply random graph. The set of all graphs that can be generated by this random process is denoted $\mathcal{G}(n,p)$ or \mathcal{G} for short.

The number of links of a node is its degree d, where $d \in \{0, \ldots, n-1\}$. A node with d=0 is isolated. The minimum node degree $d_{\min}(G)$ of a graph G is the smallest degree over all nodes in G. In a random graph, each node has a random degree represented by the random variable D. The probability that an arbitrary node in a random graph G(n,p) has a certain degree d is given by a binomial distribution, i.e.,

$$\mathbb{P}\left[D=d\right] = \mathcal{B}_d\left(n-1,p\right) := \binom{n-1}{d} p^d \left(1-p\right)^{n-1-d}.$$
(1)

A path is a sequence of consecutive links in a graph. Two nodes are *connected* if there is a path between them. A graph is *connected* if there is a path between all pairs of nodes in the graph. It is called disconnected otherwise. A *connected component* is a maximal connected subgraph of G, i.e., a subgraph in which any two nodes are connected by a path, and to which no more nodes of G can be added without losing the connectivity of the subgraph. A graph is k-connected ($k \in \mathbb{N}$) if there are at least k node-disjoint paths between each pair of nodes. Equivalently, a graph is k-connected if and only if no set of (k-1) nodes exists whose removal would disconnect the graph. The event that a graph G is connected (or k-connected) is called "G con" (or "G k-con").

III. DEGREE DISTRIBUTION OF A CONNECTED RANDOM NETWORK

We are interested in the degree distribution of a random network given that the network is connected. In other words, we would like to determine the conditional distribution $\mathbb{P}[D=d\mid G \text{ con}]$. This means that among all graphs in \mathcal{G} we only take the connected ones and analyze their degree distribution. In this section, we first give an approximation for large networks (many nodes). Second, we derive an upper bound for the conditional distribution that is also for small networks (arbitrary number of nodes). Finally, we assess the tightness of both expressions by simulation.

A. Asymptotic Behavior of the Degree Distribution

The condition that a graph G has no isolated node is a necessary but not sufficient condition for G to be connected. This observation is also true for random graphs, where the set of connected graphs in G is a subset of the set of graphs without isolated nodes. Hence, $\mathbb{P}[G \text{ con}] \leq \mathbb{P}[D_{\min}(G) > 0]$ holds.

The relationship between the two events "G con" and "no isolated node in G" was analyzed in the theory of random graphs. An important observation is the following (see [12] and [13]): if the number of nodes n is sufficiently large, *almost every graph* in G without any isolated node is also connected. In other words, the random graph process yields only very few

disconnected graphs with isolated nodes. The two events are eventually equal in a probabilistic sense, and it is legitimate to exchange one for the other. In mathematical terms,

$$\lim_{M \to \infty} \mathbb{P}[(G \text{ con}) = (D_{\min} > 0)] = 1, \qquad (2)$$

where $p \leq (\ln n + \ln \ln n)/n$. Applying this insight to our problem, we can state

$$\left| \mathbb{P}\left[D = d \mid G \text{ con} \right] - \mathbb{P}\left[D = d \mid D_{\min}(G) > 0 \right] \right| \le \varepsilon \quad (3)$$
 with $\varepsilon \to 0$ as $n \to \infty$.

Using the definition of conditional probability, we have

$$\mathbb{P}[D = d \mid D_{\min}(G) > 0] = \frac{\mathbb{P}[(D = d) \cap (D_{\min} > 0)]}{\mathbb{P}[D_{\min} > 0]}.$$
(4)

The probability that there is no isolated node can be bounded as follows (Chapter 3.1 in [13] for a degree 0):

$$\mathbb{P}\left[D_{\min} > 0\right] \ge 1 - \lambda_0 \quad \text{with} \quad \lambda_0 = p^{n-1} \,. \tag{5}$$

Applying this inequality in (4) and exploiting the fact that $\mathbb{P}\left[(D=d) \cap (D_{\min}>0)\right] \leq \mathbb{P}\left[D=d\right]$, we can estimate

$$\mathbb{P}[D = d \mid D_{\min}(G) > 0] \le \begin{cases} 0 & \text{if } d = 0 \\ \frac{\mathbb{P}[D = d]}{1 - \lambda_0} & \text{if } d > 0 \end{cases} . (6)$$

As $n \to \infty$, we have $\lambda_0 \to 0$.

In summary, for large n, the distribution $\mathbb{P}\left[D=d \mid G \text{ con}\right]$ can be approximated by $\mathcal{B}_d\left(n-1,p\right)$ for d>0, and is 0 for d=0. This binomial distribution can be further approximated by a Poisson distribution $\mathbb{P}\left[D=d\right]=\mathcal{P}_d\left(\lambda\right)\coloneqq\frac{\lambda^d}{d!}\,e^{-\lambda}$ with $\lambda=np$. As a rule of thumb, this approximation is good if $n\geq 100$ and $np\leq 10$ [14]; hence, $\mathbb{P}\left[D=d\mid G \text{ con}\right]\approx\frac{\lambda^d}{d!}\,e^{-\lambda}$ for d>0, and is 0 for d=0.

B. An Upper Bound for the Degree Distribution

Let us now derive an upper bound for the conditional degree distribution that is valid for arbitrary n. Writing

$$\mathbb{P}\left[D = d \mid G \text{ con}\right] = \frac{\mathbb{P}\left[\left(D = d\right) \cap \left(G \text{ con}\right)\right]}{\mathbb{P}\left[G \text{ con}\right]}, \quad (7)$$

our approach is to overestimate the numerator and give an exact solution for the denominator. By basic probability theory we have $\mathbb{P}\left[(D=d) \cap (G \text{ con})\right] \leq \mathbb{P}\left[D=d\right]$, which yields

$$\mathbb{P}\left[D = d \mid G \text{ con}\right] \le \frac{\binom{n-1}{d} p^d \left(1 - p\right)^{n-1-d}}{\mathbb{P}\left[G(n, p) \text{ con}\right]} \tag{8}$$

for $d \in \mathbb{N}$, and zero for d = 0.

The probability that a random graph G(n,p) is connected can be calculated using the recursive approach by Gilbert [2]. Let us consider an arbitrary node in a graph. The node belongs to a connected component of order i, with $i \in \{1, ..., n\}$, if the node is connected to exactly (i-1) nodes of the graph. The probability for this event is $[2]^1$

$$\alpha(n, p, i) := \binom{n-1}{i-1} \mathbb{P}\left[G(i, p) \text{ con}\right] (1-p)^{i(n-i)} . \quad (9)$$

¹There are $\binom{n-1}{i-1}$ possibilities to choose the (i-1) nodes that are connected to the given node. The term $(1-p)^{i(n-i)}$ denotes the probability that none of the nodes within a set of i nodes has a link to any of the other (n-i) nodes of the graph.

The given node is connected to either $0, 1, \ldots, (n-2)$ or (n-1) nodes, which means that $\sum_{i=1}^{n} \alpha(n, p, i) = 1$. As $\alpha(n, p, n) = \mathbb{P}[G(n, p) \text{ con}]$, we finally obtain

$$\mathbb{P}\left[G(n,p) \text{ con}\right] = (10)^{n-1} = 1 - \sum_{i=1}^{n-1} \binom{n-1}{i-1} \mathbb{P}\left[G(i,p) \text{ con}\right] (1-p)^{i(n-i)}.$$

This result can be applied in (8) to yield an upper bound for the degree distribution of connected random graphs.

C. Simulation-Based Analysis of the Degree Distribution

To evaluate the tightness of the approximation and the upper bound for the degree distribution, a number of simulations are performed. Different random graph models $\mathcal{G}(n,p)$ are used, whose parameters are listed in Table I. These models are selected for the following reasons: First, we are interested in comparing three network orders, namely n = 10, 100, and 1000. Second, we are interested in comparing different connectivity levels, where we choose a medium connectivity level $(\mathbb{P}[G \text{ con}] = 65\%)$ and a high connectivity level $(\mathbb{P}[G \text{ con}] =$ 95 %). To obtain the appropriate connectivity level for given n, the link probability p leading to the appropriate fraction of connected graphs (with a maximum absolute error of $\pm 1\%$) is chosen. Using the software R with igraph, at least 10^6 graphs are generated for each (n, p)-pair, and a normalized degree histogram of all connected graphs is created. The term δ in Table I gives the fraction of graphs that are disconnected but have a non-zero minimum degree.

TABLE I
SIMULATED RANDOM GRAPH MODELS

	n	p	$\mathbb{P}\left[G\text{ con}\right]$	δ
\mathcal{G}_a :	10	0.300	64.9%	$4 \cdot 10^{-2}$
\mathcal{G}_b :	10	0.445	95.1%	$2 \cdot 10^{-3}$
\mathcal{G}_c :	100	0.0536	65.1%	$4 \cdot 10^{-3}$
\mathcal{G}_d :	100	0.0737	95.0%	$1 \cdot 10^{-4}$
\mathcal{G}_e :	1000	0.00773	65.0%	$5 \cdot 10^{-4}$
\mathcal{G}_f :	1000	0.00984	95.0%	$1\cdot 10^{-5}$

Fig. 1 shows the results, comparing four metrics:

- simulated ("real") degree distribution of all connected graphs,
- approximation for the asymptotic degree distribution of all connected graphs, given by the restricted binomial distribution shown on the right hand side of (6),
- upper bound for the degree distribution of all connected graphs, given in (8), and
- degree distribution of all graphs (binomial distribution).

We calculate two mean square errors: MSE_a of the restricted binomial distribution compared to the simulated distribution, and MSE_b of the binomial distribution compared to the simulated distribution. The squared sums are taken over all degree values $d = 0, \ldots, n-1$.

Fig. 1(a) shows the results for $\mathcal{G}_a = \mathcal{G}(10, 0.300)$, where 65 % of the generated graphs are connected. The MSE_a of

the restricted binomial distribution is about half the MSE_b of the binomial distribution $\mathcal{B}_d(9, 0.300)$ of all graphs. The upper bound is not tight with these parameters.

Fig. 1(b) shows the distribution for $\mathcal{G}_b = \mathcal{G}(10, 0.445)$. Here, 95% of the simulated graphs are connected. The MSE decreases by one order of magnitude; the relative difference between MSE_a and MSE_b increases. Both the restricted binomial distribution and the upper bound yield good approximations.

Figs. 1(c) and 1(d) show the results for n=100 nodes with a connectivity rate of $65\,\%$ and $95\,\%$. As expected, the restricted binomial distribution gets closer to the real distribution. As for n=10, the upper bound is weak for a low connectivity, whereas it gives a reasonable approximation for a high connectivity. This behavior is supported further in the simulations of n=1000 nodes (see Figs. 1(e) and 1(f)).

In conclusion, the restricted binomial distribution has the advantage of being closer to the real distribution especially for small networks. The bound has the advantage of always giving an upper limit for the probability of a certain degree; it is, however, a weak approximation for weakly connected networks.

IV. DEGREE DISTRIBUTION OF A k-CONNECTED RANDOM NETWORK

We are now interested in the degree distribution conditioned by the fact that a network is k-connected (i.e., there are at least k node-disjoint paths between each node pair). Such topologies are motivated by the demand for robustness against node and link failures or the request for multipath routing. For wireless multihop networks, topology control algorithms have been invented to adjust the power of devices to maintain k-connectivity [15]. In the following, we focus on the asymptotic behavior and the distribution obtained by simulations.

A. Asymptotic Behavior of the Degree Distribution

In a k-connected graph, each node must have at least k neighbors. Hence, the condition $d_{\min}(G) \geq k$ is necessary but not sufficient for G to be k-connected. For random graphs, the set of k-connected graphs is a subset of the graphs with a minimum degree of at least k. Thus, $\mathbb{P}[G \ k\text{-con}] \leq \mathbb{P}[D_{\min}(G) \geq k]$. Beyond this, the theory of random graphs was able to generalize the statement that "almost every random graph without any isolated node is connected" (Section III-A) to k-connectivity: almost every random graph in which each node has at least k neighbors is k-connected. If n is large and $p \leq (\ln n + k \ln \ln n)/n$, then $\mathbb{P}\big[(G \ k\text{-con}) = (D_{\min}(G) \geq k)\big] \to 1$ (see Ch. 7.2 in [13]). Applying this to our problem, if the network has sufficiently many nodes, we have

$$\left| \mathbb{P}[D = d \mid G \text{ } k\text{-con}] - \mathbb{P}[D = d \mid D_{\min}(G) \ge k] \right| \le \varepsilon \text{ (11)}$$

with $\varepsilon \to 0$ as $n \to \infty$.

To deal with $\mathbb{P}[D_{\min}(G) \geq k]$, we introduce the random variable X_i denoting the number of nodes with degree i in a random graph. From Chapter 3.1 in [13], we know

$$\mathbb{P}\left[X_i = 0\right] \ge 1 - \lambda_i \tag{12}$$

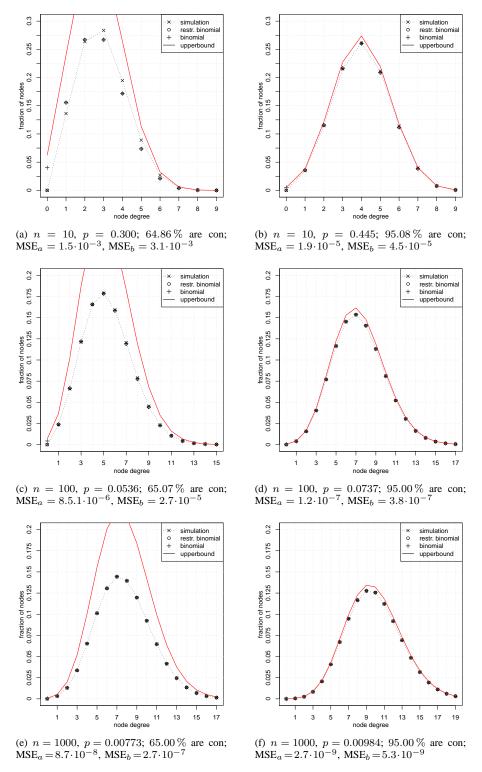


Fig. 1. Degree distribution of connected random networks.

with λ_i given by a negative binomial distribution, i.e.,

$$\lambda_{i} = \mathcal{NB}_{i} (n-1, p) := {n+i-2 \choose i} p^{n-1} (1-p)^{i}.$$
 (13)

A graph with minimum degree k has $X_i=0$ for all $i\in[0,..,k-1]$, the probability for this event being $\prod_{i=0}^{k-1}\mathbb{P}\left[X_i=0\right]\geq\prod_{i=0}^{k-1}(1-\lambda_i)$.

Applying these results in (11) yields

$$\mathbb{P}[D = d \mid D_{\min}(G) \ge k] \le \begin{cases} 0 & \text{if } d < k \\ \frac{\mathbb{P}[D = d]}{\prod_{i=0}^{k-1} (1 - \lambda_i)} & \text{if } d \ge k \end{cases}$$
(14)

For a graph G(n,p), the term $\mathbb{P}\left[D=d\right]$ is the binomial distribution (1), and the denominator goes to 1 for $n\to\infty$. Thus, for large n, the distribution $\mathbb{P}\left[D=d\mid G \text{ k-con]}\right]$ can be approximated by $\mathcal{B}_d\left(n-1,p\right)$ for $d\geq k$, and is 0 for d< k.

B. Simulation-Based Analysis of the Degree Distribution

Let us evaluate the tightness of the restricted binomial distribution (14). The *Library of Efficient Data Types & Algorithms* (LEDA) generates random graphs and verifies k-connectivity. We focus on 2- and 3-connected networks using the models \mathcal{G}_b , \mathcal{G}_d , and \mathcal{G}_f (Table I), and analyzing at least 10^6 random graphs. The results are shown in Fig. 2.

We first interpret the results for n=10. About 64% of the graphs in \mathcal{G}_b are 2-connected; their degree distribution is given in Fig. 2(a). Clearly, the MSE to the binomial distribution is much larger than it is for 1-connected graphs in the same set. The restricted binomial distribution yields a fair approximation; its MSE is, however, one order of magnitude higher than for 1-connected graphs. Fig. 2(b) considers 3-connected networks in \mathcal{G}_b (11% of all graphs). The error of the binomial distribution increases with increasing connectivity. The error of the restricted binomial distribution also increases but yields a significantly better approximation.

Increasing n to 100 and finally to 1000 (Figs. 2(c)+(e)), the fraction of 2-connected graphs is only slightly reduced. In both cases, the restricted binomial distribution offers a reasonable approximation of the simulated distribution. Clearly, the approximation becomes tighter (MSE_a decreases) as n increases. The same is true for the degree distribution of 3-connected networks in Figs. 2(d) and (f).

We interpret the results as follows: First, the restricted binomial distribution always yields a lower MSE than the binomial distribution (MSE $_a$ < MSE $_b$). Even for small networks, it is better to assume (14) than the binomial distribution. Second, the MSE of the restricted binomial distribution decreases for increasing n (cp. figures in one column vertically). Third, the MSE of the restricted binomial distribution increases for increasing k (cp. e.g. Figs. 1(b), 2(a), 2(b)).

V. CONCLUSIONS

Motivated by the fact that disconnected random topologies are discarded in simulations of communication and computer networks, we analyzed the degree distribution of connected and k-connected random graphs.

The main contribution is an approximation for the degree distribution of k-connected graphs with many nodes. This expression yields a better approximation than the original random graph degree distribution for most degree values; it even yields a reasonable approximation for small networks (here n=10). Furthermore, an upper bound for the degree distribution of 1-connected networks has been derived. It gives an upper limit for the probability of a given degree but represents a worse approximation, especially for weakly connected networks.

From a practical point of view, the results show the pitfall of not being aware of the change in the distribution does not seem to be a serious mistake in typical setups with many nodes. The difference is visible mainly in small networks.

ACKNOWLEDGMENTS

This work was performed in the research cluster Lakeside Labs. It was partly funded by the European Regional Development Fund, the Carinthian Economic Promotion Fund (KWF), and the state of Austria under grants 19964/15122/21646 and 20214/17094/24770. The authors would like to thank Sérgio Crisóstomo and Udo Schilcher for technical discussions and Kornelia Lienbacher for proofreading.

REFERENCES

- P. Erdős and A. Rényi, "On random graphs," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.
- [2] E. N. Gilbert, "Random graphs," Annals Math. Stat., vol. 30, pp. 1141– 1144, Dec. 1959.
- [3] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, eds., The Structure and Dynamics of Networks. Princeton Univ. Press, 2006.
- [4] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge Univ. Press, 2008.
- [5] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. IEEE Infocom*, (Miami, FL, USA), Mar. 2005.
- [6] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," in *Proc. SIGMETRICS*, (Marina Del Rey, CA, USA), June 2002.
- [7] S. Crisóstomo, U. Schilcher, C. Bettstetter, and J. Barros, "Analysis of probabilistic flooding: How do we choose the right coin?," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, (Dresden, Germany), June 2009.
- [8] S. Baroni and P. Bayvel, "Wavelength requirements in arbitrarily connected wavelength-routed optical networks," *Journal of Lightwave Techn.*, vol. 15, pp. 242–251, Feb. 1997.
- [9] J. Wu and H. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," in *Proc. Workshop Discrete Alg. Meth. Mob. Comput. Commun.*, (Seattle, WA, USA), ACM, Aug. 1999.
- [10] S. Crisóstomo, J. Barros, and C. Bettstetter, "Flooding the network: Multipoint relays versus network coding," in *Proc. IEEE Intern. Conf. Circuits Sys. Commun. (ICCSC)*, (Shanghai, China), May 2008.
- [11] S. Raza and C.-N. Chuah, "Interface split routing for finer-grained traffic engineering," *Perform. Eval.*, vol. 64, no. 9-12, pp. 994–1008, 2007.
- [12] B. Bollobás and A. G. Thomason, "Random graphs of small order," Random Graphs, Annals Discr. Math., pp. 47–97, 1985.
- [13] B. Bollobás, Random Graphs. Cambridge Univ. Press, 2 ed., Jan. 2001.
- [14] P. L. Gatti, Probability Theory and Mathematical Statistics for Engineers. Spon Press, 2005.
- [15] N. Li and J. C. Hou, "FLSS: A fault-tolerant topology control algorithm for wireless networks," in *Proc. ACM MobiCom*, (Philadelphia, PA, USA), Sept. 2004.

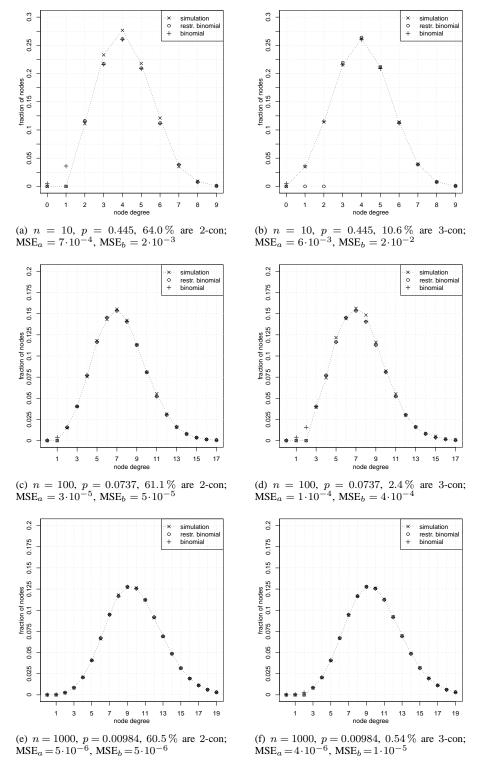


Fig. 2. Degree distribution of 2- and 3-connected random networks.