

Giant Clusters in Random Ad Hoc Networks

G. Németh* and G. Vattay

Communication Networks Laboratory, Eötvös University, Budapest, Hungary

(Dated: October 27, 2018)

The present paper introduces ad hoc communication networks as examples of large scale real networks that can be prospected by statistical means. A description of giant cluster formation based on the single parameter of node neighbor numbers is given along with the discussion of some asymptotic aspects of the giant cluster sizes.

PACS numbers: 89.75.Hc, 05.40.-a, 64.60.Cn

I. INTRODUCTION

Nowadays, natural and designed networks are in the focus of research in different scientific disciplines. Using computers the amount of available empirical data on real world networks has been increased during the past few years. Examples of real networks include the World Wide Web[1, 2], the Internet [3, 4, 5, 6, 7], collaboration networks of movie actors and scientists [8, 9, 10], power grids [11, 12] and the metabolic network of living organisms [13, 14, 15, 16].

Random graphs are natural candidates for the description of the topology of such large systems of similar units. In [17, 18, 19] the authors have developed a model – which assumes each pair of the graph’s vertices to be connected with equal and independent probabilities – that treats a network as an assembly of *equivalent units*.

This model, introduced by the mathematicians Erdős and Rényi, has been much investigated in the mathematical literature[20, 21]. However, the increasing availability of large maps of real-life networks has indicated that the latter structures are fundamentally correlated systems, and in many respects their topologies deviate from the uncorrelated random graph model.

Two classes of models, commonly called the *small-world graphs*[11, 12, 22] and the *scale-free networks*[23,

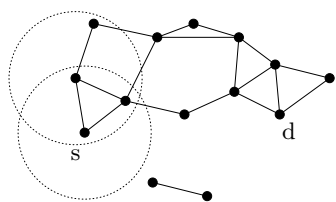


FIG. 1: Nodes and connections of an example ad hoc network. The transmission range is the same for all nodes – it is noted by the dotted circle for two of the nodes. The shortest path between the *s* source and *d* destination users touches 3 intermediate nodes, and there is an alternative route of 6 hops, which has no common intermediate nodes with the first.

24] have been developed to capture the clustering and the power law degree distribution present in real networks[1, 3, 8, 9, 10, 11, 12, 13, 23, 24, 25, 26, 27, 28, 29].

Here we present *ad hoc networks*[30] as new examples of real structures that can be investigated using the above network models. Ad hoc networks arise in next generations of communication systems and hereby we try to summarize the principal characteristics of such systems. In the ad hoc scheme users communicate by means of short range radio devices, which means that every device can connect to those devices that are positioned no farther than a finite maximum geometrical range. We call this range the given device’s *transmission range* and the exact value of this range may depend on the transmitter’s power and various other physical parameters. See Fig. 1 for an example ad hoc network topology. Neighbor nodes talk the way ordinary radios – like CBs – do, however communication between non-neighboring users is also possible. The latter case is accomplished by sending the information from the source user to the destination hop by hop, through intermediate nodes. If the density of users in the area is high compared to their transmission ranges, it is highly possible that more than one alternative route exists between two users. This last feature can be exploited in the case if the shortest route is overloaded or broken, or if the system allows splitting the information flow into separate parallel flows. Moreover, the users are free to move randomly and organize themselves arbitrarily; thus, the network’s topology may change rapidly and unpredictably. Such a network may operate in a stand-alone fashion, or may be connected to the Internet.

Giant clusters in ad hoc networks are made interesting because a communication network provides a meaningful service only if it integrates as many users as possible within the covered area (e.g., 99% may be considered a good coverage). In this paper we introduce a fractal model, that duplicates the giant component formation in ad hoc networks in an area inlaid with obstacles, partially screening radio transmission. Our main result is that in such networks the giant component size can be described by a single parameter: the average number of neighbors a node has. The rest of this paper is structured as follows. Section II gives a detailed description of our random ad hoc network model. In Sections III and IV we delve into the topology differences between

*Electronic address: gab@cni.elte.hu

random graphs and graphs built using our model. Section V shows the numerical simulation results supporting these analyses.

II. THE RANDOM AD HOC NETWORK MODEL

A wireless ad hoc network consists of a number of radio devices, also referred to as "nodes" in the following. Every node may be connected to one or more other nodes in her vicinity; the actual set of connections depends on the distance of the nodes. In a static environment these connections define the topology of the system; if the nodes allowed to move then the topology may change, however at any given point of time there is still a well defined topology available.

To be precise we define a *random ad hoc network* as a set of uniformly distributed nodes on the arena of the unit Euclidean square $[0; 1] \times [0; 1]$ with connections between pairs of them. The connections are two-way in the sense that if node A can communicate to node B , then node B is also able to communicate to node A .

Two nodes are connected if the geometrical distance of the two is less than a certain value r_t , that is the nodes can communicate up to their "transmission range". We represent a realization of such a system using an undirected graph $G(V, E)$, where the vertices and the edges denote the nodes and the two-way connections respectively. Sometimes a graph resulting this way is referred to as a *geometric random graph* or *GRG*. Note that there are no loops and no multiple edges in G : *a*) a node should not communicate to itself; and *b*) if two nodes are neighbors, then technically there is no sense to open a second communication channel between them.

Furthermore, all the length parameters in the system are made dimensionless as follows. Length is measured as the multiples of the *unit radius* r_0 , which is in turn defined by the share of the whole area for each node:

$$r_0 := \sqrt{\frac{A}{N\pi}} \quad (1)$$

where A is the size of the arena. The ratio of the transmission range and the unit radius is called the *normalized transmission range* and noted by:

$$r_n := \frac{r_t}{r_0} \quad (2)$$

As mentioned in the Introduction, a communication network may deliver meaningful service only if the network is connected, or at least has a vast subset that is connected. Our work is focused on examining the criteria for giant cluster formation and in particular in networks with fractal connectivity properties.

In the following we give a short overview of networks on random graphs and afterwards we turn to our model of fractal ad hoc connectivity.

III. CONNECTIVITY IN RANDOM NETWORKS

After distributing and connecting the nodes as described previously, the largest connected component of G can be determined. Let S be this components' size fraction:

$$S := \frac{\text{nodes in the largest component}}{N}$$

which quantity is obtained by counting. This quantity is of particular importance because the network gets fully connected if S diverges and for this end we are to investigate its relationship with other network parameters.

In [31] the authors present the theory of *random graphs*[18] of arbitrary degree distribution. Among others, an exact result for the component sizes is given, which we shall cite here. It is shown, that the average component size diverges if

$$\sum_{k=0}^{\infty} k(k-2)p_k = 0$$

holds, where p_k is the degree distribution of vertices in G . Let us use here the actual distribution of our ad hoc network: it is easily seen that the probability distribution of the number of nodes contained in any disc with radius r_n is the Poisson-distribution with expectation value of r_n^2 . It means that

$$p_k = \frac{(r_n^2)^k}{k!} e^{-r_n^2} \quad (3)$$

is the probability that a vertex will have $k-1$ neighbors (the -1 is because the node itself does not count for a neighbor). Applying the result in [31], one can derive the relationship of the size of the giant component S and the transmission range:

$$r_n^2 = \frac{\log(1-S)}{-S} \quad (4)$$

It shall be noted here that while (4) holds for random networks and – as it is to be shown in Section V – for fractal ad hoc networks, the r_n-S relationship is different for the finite range ad hoc case, however, the latter is to be discussed in a separate paper.

IV. THE FRACTAL AD HOC NEIGHBORSHIP ALGORITHM

The results of the previous section apply for scenarios where the arena is "flat": that is the only limit to build a connection between two nodes is their geometrical distance. In the present section we introduce the idea of generalized obstacles that can screen nodes from each other even if they are positioned within transmission range. This change produces graphs with extended

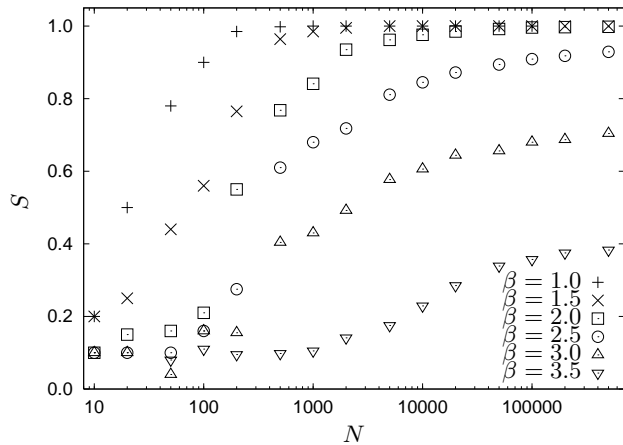


FIG. 2: Giant component sizes for various values of N and β . Note that $S(N)$ reaches 1 for $\beta \leq 2$, yet $\lim_{N \rightarrow \infty} S(N, \beta) = S_{\max}(\beta) < 1$ for $\beta > 2$.

spatial structure which is why we call the algorithm fractal.

The obstacles are adopted by changing the algorithm for edge generation. Now two nodes within the transmission range will be connected with a probability which is given as the function of their geometrical distance. For every two nodes $u, v \in V$ let $p(\text{dist}(u, v))$ be the probability that an edge $e_{uv} \in E$ connecting them is set up. For the description of our obstacles we use a long tailed probability function which is implied by the picture of a hilly landscape, where the possibility of connections drops with the increasing geometrical distance between the nodes, however long range connections are still possible:

$$p(r) = \frac{a}{\left(1 + \frac{r}{r_0\beta}\right)^\beta} \quad (5)$$

with parameter values $a > 0$ and $\beta > 0$.

Performing computer simulations of networks connected according to (5), one obtains different results, as β changes. On Fig. 2 we compared the resulting giant cluster sizes for different β values. At lower parameter values $S(N)$ saturates to $S = 1$ – all nodes become elements of the giant cluster above a certain finite node number. For $\beta = 2.5$ and above S still converges to a finite value, however the limit now is strictly less than 1. It means that networks with such parameter values will not become fully connected even at large node numbers, moreover, the proportion of the largest connected subgraphs drops with β worse than linearly. In the rest of this Section we try to interpret this dual behavior of $S(\beta)$.

It is easy to imagine that the more connections the nodes have in average, the larger the giant cluster grows. More accurately we state that the *average vertex degree* $\langle C \rangle$ determines the cardinality of the largest connected

subgraph in G . Clearly, if $\langle C \rangle = 0$, then every connected component contains a single node, and in the $N \rightarrow \infty$ limit S becomes 0. Also, if $\langle C \rangle$ diverges or even if only a single node is connected to all the others, the graph obviously gets fully connected. Based on these considerations we are to examine $\langle C \rangle$ in detail.

Vertex degree in G can be calculated by fixing a single node and totaling the $\langle C_r \rangle$ expectation value of the number of neighbors that reside exactly at the distance r away from the fixed one. Assuming that the density of nodes is constant (N/A), $\langle C_r \rangle$ can be expressed by multiplying the average number of nodes in distance r and the probability (5):

$$\langle C_r \rangle = \frac{2r\pi}{A} N p(r)$$

Now if $\bar{\rho} = N/A$, the average vertex degree is

$$\langle C \rangle = \int_{\mathfrak{A}} \langle C_r \rangle dr = \int_{\mathfrak{A}} p(r) 2\pi\bar{\rho}r dr \quad (6)$$

where \mathfrak{A} represents the physical boundaries of the arena. As there are no nodes outside this region, thus the integral shall be 0 outside \mathfrak{A} .

In general solving (6) yields

$$\langle C \rangle = a2\pi\bar{\rho} \frac{r_0\beta}{1-\beta} \times \left[r \left(1 + \frac{r}{r_0\beta}\right)^{1-\beta} - \frac{r_0\beta}{2-\beta} \left(1 + \frac{r}{r_0\beta}\right)^{2-\beta} \right]_{\mathfrak{A}} \quad (7)$$

However the expectation value of $\langle C \rangle$ is dependent on the value of β . Accordingly, our discussion is separated into several cases.

a $\beta > 2$. In this case (7) can be evaluated for \mathfrak{A} being the interval $r \in [0; \infty)$ in the limit where $r_0 \rightarrow 0$:

$$\langle C \rangle = \frac{a2\pi\bar{\rho} \cdot r_0^2\beta^2}{(1-\beta)(2-\beta)} \equiv \frac{2a\beta^2}{(\beta-1)(\beta-2)} \quad (8)$$

Furthermore, knowing that

$$\lim_{\alpha \rightarrow \infty} \frac{1}{(1+x)^\alpha} = e^{-\alpha x}$$

in the $\beta \rightarrow \infty$ limit (8) becomes

$$\langle C \rangle = a2\pi\bar{\rho}r_0^2 \equiv 2a$$

b $\beta = 1$ or $\beta = 2$. (7) diverges logarithmically in r , thus C does not have an expectation value.

c $\beta < 2$ and $\beta \neq 1$. Here $\langle C \rangle$ will diverge as $N \rightarrow \infty$, however unlikely to the previous case we try to determine the $\langle C(N) \rangle$ relation. First let us rewrite (7) as

$$\langle C \rangle = \frac{a2\pi\bar{\rho} \cdot r_0\beta}{(1-\beta)(2-\beta)} \left[\frac{r(1-\beta) - r_0\beta}{\left(1 + \frac{r}{r_0\beta}\right)^{\beta-1}} \right]_{\mathfrak{A}} \quad (9)$$

Concerning the r -dependence in $[\dots]_{\mathfrak{A}}$ we can assume that there is a maximal transmission range r_{\max} such that for transmission ranges $r > r_{\max}$ the contribution of the integrand in (6) is negligible. This way the $[\dots]_{\mathfrak{A}}$ part of (9) can be estimated as

$$[\dots]_{\mathfrak{A}} \simeq -\frac{r_0\beta}{\left(1 + \frac{r}{r_0\beta}\right)^{\beta-1}} + \frac{r_{\max} \cdot (1-\beta)}{\left(1 + \frac{r_{\max}}{r_0\beta}\right)^{\beta-1}} \quad (10)$$

Now if $r_0 \rightarrow 0$ (which happens to be the case at sufficiently large node numbers) the first term in (10) vanishes and the $+1$ becomes negligible in the denominator of the second term. After substituting this second term and simplifying the expression, (7) finally becomes

$$\langle C \rangle \simeq \frac{a2\pi\bar{\rho}}{2-\beta} \left(\frac{r_0\beta}{r_{\max}}\right)^{\beta} \cdot r_{\max}^2$$

The N -dependence of $\langle C \rangle$ can be derived from here by substituting definition (1), $\bar{\rho} = N/A$ and the fact that $r_{\max}^2 \propto A$. By these means the above expression yields:

$$\langle C \rangle \propto N^{1-\frac{\beta}{2}} \quad (11)$$

To summarize, if $\beta > 2$, then a finite neighbor count is expected, and thus such networks are not going to be fully connected (see again Fig. 2). On the other hand, if $\beta < 2$, then $\langle C \rangle$ diverges exponentially with increasing node numbers, which in theory leads to fully connected networks at large N , and means, that the more nodes are in the system, the larger the fraction of connected nodes is to become.

V. SIMULATION RESULTS

We carried out computer simulations to illustrate our findings, especially Eqs. (8) and (11). During a simulation run we first pick the random coordinates for the N nodes. Second the probability p is calculated according to (5), using the input parameters a, β and r . Then for every two nodes a uniform random number $\xi \in [0; 1]$ is generated and compared to p : for cases $\xi < p$ an edge connecting those two nodes is recorded. Finally we count the component sizes and take the largest of these. The output of the simulation run is the average vertex degree, $\langle C \rangle$, and the largest components' size, S . Note the analogy with (3) if using the fact, that here $\langle C \rangle = r_n^2$.

As the first test we recorded the giant cluster size vs. transmission range relationship. Data points were obtained by repeated runs, changing only the amplitude parameter a of (5) in an appropriate interval (e.g., $a \in [1; 9]$ for the $\beta = 2.5$ case). Fig. 3 illustrates that in a network connected using the fractal neighborhood algorithm

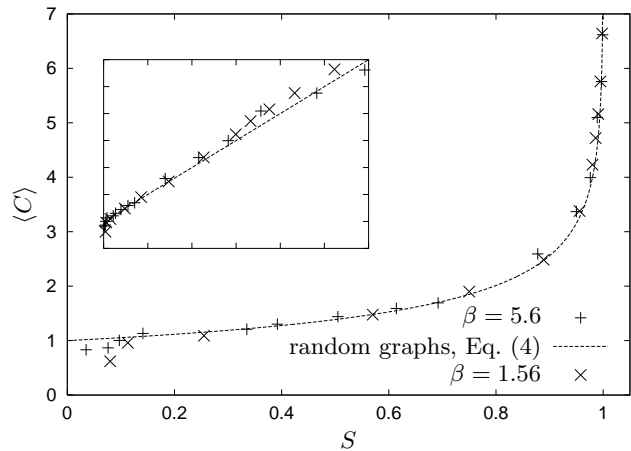


FIG. 3: Simulations of ad hoc networks using the fractal neighborhood algorithm with parameter values both $\beta < 2$ and $\beta > 2$ yield the same giant cluster size vs. average vertex degree as random graphs. Inset displays the same plots with $-\log(1-S)/S$ as the abscissa.

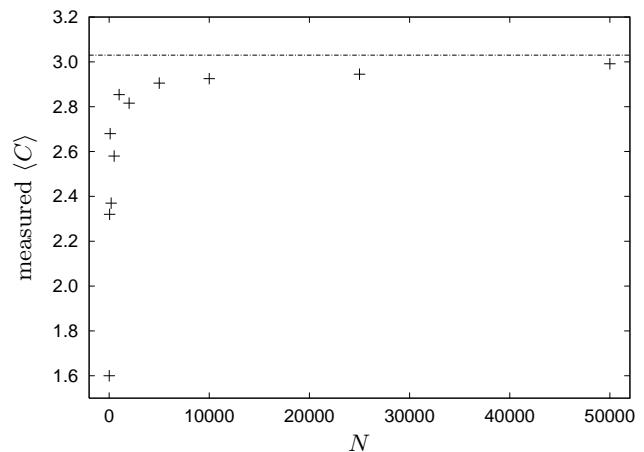


FIG. 4: Average vertex degree of ad hoc graphs for $\beta > 2$. Data points were acquired using $a = 0.8$, $\beta = 5.6$; the dashed line yields the analytical result $\langle C \rangle = 3.03$, which shall hold in the $N \rightarrow \infty$ limit.

the observable $S - \langle C \rangle$ relationship matches the equivalent analytical result for random graphs for both relevant cases a and c in Section IV.

On the other hand, the behavior of $\langle C \rangle$ turns out to be sensible to the value of β , as it was expected. Let us start with the case $\beta > 2$. Fig. 4 presents the simulation results for networks connected as by (5), using $a = 0.8$ and $\beta = 5.6$. According to (8), the average vertex degree is expected to be

$$\langle C \rangle = \frac{2 \cdot 5.6^2}{4.6 \cdot 3.6} \simeq 3.03$$

in this case. It is clearly seen on the Figure that increas-

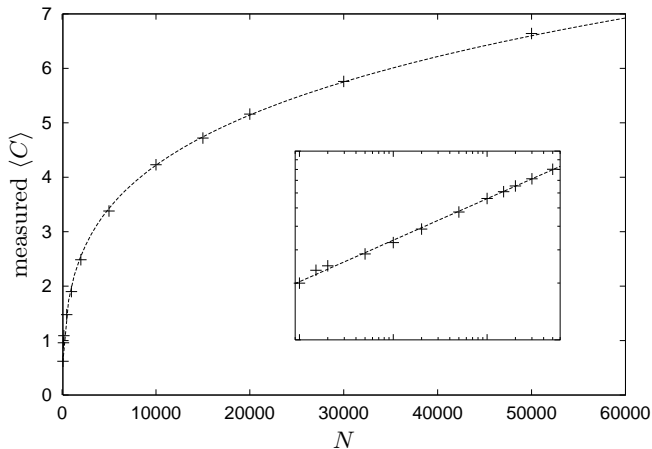


FIG. 5: Divergence of average vertex degrees with N for $\beta < 2$. Crosses: data points for $a = 0.1$, $\beta = 1.56$; dashed line is $\propto N^{1-\beta/2} = N^{0.22}$. Inset displays the same plot with both axes logarithmic.

ing N , the simulation output converges to the analytical result.

Now let us turn to the $0 < \beta < 2$ case. On Fig. 5 the data obtained for $a = 0.1$ and $\beta = 1.56$ is shown along with a numeric function fit according to (11):

$$\langle C(N) \rangle = c_0 \cdot N^{1-\frac{1.56}{2}} + c_1$$

(the parameters turn out to be $c_0 = 0.74$ and $c_1 = -1.38$). The simulations agree with the $N^{1-\beta/2}$ divergence well, as calculated in Section IV.

Figs. 4 and 5 now illustrate the differing S -behavior presented on Fig. 2, as a data set for $\beta = 5.6$ would converge to some connectivity $< 20\%$ even at very large N , while the one for $\beta = 1.56$ is clearly reaching $S = 1$ for node numbers in the magnitude of several thousands.

VI. CONCLUSIONS

In the present paper we have investigated the connected components that are produced in random ad hoc networks. Based on the results, the number of nodes needed for a given connectivity ratio can be estimated. Thus our results may hint about the usefulness of random fractal ad hoc networks.

We modified the conventional connection function and made long range connections possible. This way the producing networks become extended in their spatial structure, as thought the network is situated in an area with obstacles screening some of the transmissions. We have found that a single parameter – the average neighbor count $\langle C \rangle$ – can characterize the proportion of the largest connected subnetwork. We have also seen that depending on the connection function parameters, this proportion can be either bounded or unbounded as the system size N is increased. For both cases $\langle C(N) \rangle$ was derived analytically and confirmed by simulations.

Acknowledgements

The authors thank the support of Hungarian National Science Found OTKA T37903 and T32437.

-
- [1] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
 - [2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins, and J. Wiener, in *Proc. of the 9th Int. World-Wide Web Conf.* (2000), <http://www.almaden.ibm.com/cs/k53/www9.final>.
 - [3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).
 - [4] A. Medina, I. Matta, and J. Byers, *ACM SIGCOMM Computer Communication Review (CCR)* **30**(2), 18 (2000).
 - [5] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
 - [6] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Breakdown of the internet under intentional attack*, cond-mat/0010251.
 - [7] Y. Shavitt, X. Sun, A. Wool, , and B. Yener, *Computing the Unmeasured: An Algebraic Approach to Internet Mapping*, Tech. Rep. Technical Report 10009674-000214-01TM, Lucent Technologies (2000).
 - [8] A. L. Barabási, H. Jeong, E. Ravasz, Z. Néda, T. Vicsek, and A. Schubert, *Evolution of the scientific collaboration network*, to be published.
 - [9] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **98**, 404 (2001).
 - [10] M. E. J. Newman, *Who is the best connected scientist? A study of scientific coauthorship networks*, cond-mat/0011144.
 - [11] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
 - [12] D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness (Princeton Reviews in Complexity)* (Princeton University Press, Princeton, New Jersey, 1999).
 - [13] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, *Nature* **407**, 651 (2000).
 - [14] P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell, *Small Cycles in Small Worlds*, Tech. Rep. Working Paper no. 00-10-058, Santa Fe Inst. (2000).
 - [15] J. M. Montoya and R. V. Solé, *Small world patterns in food webs*, Tech. Rep. Working Paper no. 00-10-059, Santa Fe Inst. (2000).
 - [16] R. V. Solé and J. M. Montoya, *Complexity and fragility in ecological networks*, Tech. Rep. Working Paper no. 00-10-060, Santa Fe Inst. (2000).
 - [17] P. Erdős and A. Rényi, *Publicationes Mathematicae* **6**, 290 (1959).

- [18] P. Erdős and A. Rényi, *Publ. of the Math. Inst. of the Hung. Acad. of Sci.* **5**, 17 (1960).
- [19] P. Erdős and A. Rényi, *Acta Math. Sci. Hung.* **12**, 261 (1961).
- [20] B. Bollobás, *Random Graphs* (London, 1985).
- [21] B. Bollobás and A. G. Thomason, *The Annals of Discrete Mathematics* pp. 47–97 (1985).
- [22] M. Barthélémy and L. A. N. Amaral, *Phys. Rev. Lett.* **82**, 3180 (1999).
- [23] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [24] A.-L. Barabási, R. Albert, and H. Jeong, *Physica* **A272**, 173 (1999).
- [25] S. Redner, *Eur. Phys. J. B* **4**, 131 (1998).
- [26] L. A. Adamic and B. A. Huberman, *Nature* **401**, 131 (1999).
- [27] J. Kleinberg, S. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proc. of the Int. Conf. on Comb. and Comp.* (1999).
- [28] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, *Classes of behavior of small-world networks*, *cond-mat/0001458*.
- [29] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, *Phys. Rev. E* **64**, 026704 (2001).
- [30] IETF Mobile Ad-hoc Networks Working Group, <http://www.ietf.org/html.charters/manet-charter.html>.
- [31] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Random graphs with arbitrary degree distribution and their applications*, *cond-mat/0007235*.