On the Topology of Overlay-Networks

Thomas Fuhrmann
Institut für Telematik, Universität Karlsruhe (TH), Germany

***Abstract*— Random-graph models are about to become an important tool in the study of wireless ad-hoc and sensor-networks, peer-to-peer networks, and, generally, overlay-networks. Such models provide a theoretical basis to assess the capabilities of certain networks, and guide the design of new protocols. Especially the recently proposed models for so-called small-world networks receive much attention from the networking community.**

This paper proposes the use of two more mathematical concepts for the analysis of network topologies, dimension and curvature. These concepts can intuitively be applied to, e.g., sensor-networks. But they can also be sensibly defined for certain other random-graph models. The latter is non-trivial since such models may describe purely virtual networks that do not inherit properties from an underlying physical world. Analysis of a random-graph model for Gnutella-like overlay-networks yields strong indications that such networks might be characterized as a sphere with fractal dimension.

Index terms — Network Topology, Random-Graphs, Small-World Networks, Peer-to-Peer Computing, Sensor-Networks

I. Introduction

Network topology is an important aspect for the performance and resilience properties of networks. Many communication networks are especially designed to have a favorable topology, e.g, to provide redundancy and avoid performance bottlenecks. But often not only technical reasons determine the networks' topologies. Other forces, e.g., economical forces, drive the network formation process, too. Hence, topology needs to be also studied analytically, e.g., to create topology models that can be used in network simulators.

Only recently has it been discovered that the Internet, like many other networks, has a power-law structure, namely in the degree distribution of its autonomous systems. A similar power-law structure has been discovered in other networks formed on top of the Internet, i.e. *overlay-networks*, e.g., the world-wide-web link-graph and peer-to-peer networks. Such overlay-networks are especially interesting since their topology is not necessarily determined by external forces but can be chosen by the protocol itself. E.g., resilient overlay networks need to find a topology that is redundant with respect to the underlying network infrastructure's links, and file sharing peer-to-peer protocols work best, if their topology reflects the servants' capabilities to serve their peers.

Wireless ad-hoc and sensor networks, on the other hand, are subject to a completely different topology forming process. Here, only nodes within the restricted reach of their radio link can be directly connected. As a result, such networks acquire their topology from the environment in which the nodes operate. This paper presents an interesting similarity between the graphs formed by sensor networks and certain random-graph overlay-networks, namely Gnutella-like peer-to-peer networks. It shows that both types of network graphs can be sensibly assigned a dimension and curvature. These concepts can serve as general means to classify random-graph models. In particular, they allow a distinction between models that are often viewed as equivalent.

In order to provide a common ground for the subsequent presentation, section II starts by giving a brief overview of different established models for random graphs, a research field that has only recently been adopted from mathematics and statistical physics to network research, and that is still not treated consistently in the networking literature. Section III then introduces the concept of dimension and curvature into the study of random graphs. Both concepts describe the number of network nodes reachable in a given number of hops. Especially, the concept of curvature naturally explains recent measurements in the field of peer-to-peer networks, that were sometimes incorrectly interpreted as effects of disconnected, i.e. hidden, subgraphs. Section IV briefly illustrates both concepts by applying them to a sensor network model. There, both concepts have an intuitive meaning whose understanding is helpful for the transfer of the concepts to random-graphs.

Section V discusses a typical small-world network model, the Barabasi-Albert model, in its application to Gnutella-like overlay-networks. It is shown that this model fails to reflect the structure observed in the Gnutella network. Hence, an alternative model is proposed that better describes recent measurements of the degree distribution of Gnutella-like peer-to-peer networks. Section VI supplements this analysis with a study of the distance distribution resulting from these models. Using the concept of dimension and curvature this extended analysis shows that the structure of Gnutella-like networks might be best understood as a sphere with fractional dimension. Section VII finally concludes with an outlook on future work.

II. Random graphs

Random graphs were introduced by Erdös and Renyi [1], [2]. They studied the probability space of graphs with a constant number of vertices and edges, and of graphs into which edges are introduced with a constant probability. Watts and Strogatz, on the other hand, studied the process of randomly rewiring a regular graph [3]. Barabasi and Albert, introduced yet another model. They studied graphs that are built up gradually by adding new vertices and edges so that the probability of an existing vertex gaining one of the new edges is proportional to its degree [4], [5].

Although, in the literal sense all of these models are *random graphs*, namely probability spaces built from a set of graphs, often only the Erdös-Renyi models are called random graphs, while the Watts-Strogatz and the Barabasi-Albert type graphs are often termed *small-world networks*, in allusion to the work by Milgram [6] who studied social networks. The idea behind this term is the comparatively small average path-length between two arbitrary nodes in such networks. To correct a common misunderstanding, Erdös-Renyi random graphs, too, show the same small-world characteristic. So, small-world networks are small when compared to regular graphs, not when compared to Erdös-Renyi random graphs.

The Barabasi-Albert model is of special interest for the study of computer networks, especially of peer-to-peer networks because the process of adding nodes to an existing network models the growth of many such networks: [7] gives the argument that a useful web-site or one that is en vogue is referenced more often than an uninteresting page. The same argument was put forth for the autonomous systems of the Internet [8], [9], and the Gnutella network [10]. There, nodes linked to many other nodes spread the knowledge about their existence in the Gnutella overlay-network more efficiently and have thus a higher probability that newly connecting servants connect to them. This mechanism is more closely studied in sectionV where yet another model is proposed that better reflects the measured properties of the Gnutella network.

III. Topology, curvature and dimension

If mathematical strictness is omitted, *topology* can be be said to describe the structure of a set without requiring a metric. To this end, the concept of a *neighborhood* is used. Typically, one would think of infinite sets, although one can construct topologies for finite sets, too. If a set is equipped with a *metric*, a neighborhood of an element x can be defined as the subset whose elements have less than a given distance from x. Varying this distance and varying x then yields the neighborhoods required to define a topology. In other words, a metric can directly induce a topology.

In computer networks, various properties can be used to define a metric, e.g., hop-count and transmission delay. While a computer scientist might think of a metric being such a concrete property, a mathematician thinks of a metric as a mere mapping that assigns distances to pairs of points with no other implied meaning. In order to qualify as a metric in the mathematical sense, such a mapping needs to be symmetric in its arguments and it needs to satisfy the triangle inequality. Luckily, both approaches coincide when a network has bidirectional links and employs shortest path routing.

In theoretical physics, the term *metric* is closely linked to the study of so-called *manifolds*, a mathematical structure that generalizes the concept of a vector space. A smooth surface, e.g., of a sphere or a torus, gives a good intuition of a two-dimensional manifold. As has been said above, the topology of a manifold can be derived from its metric. In addition to its topology a (differentiable) manifold is also equipped with *curvature* and *torsion*. Both, too, can be derived from the metric. With a fair amount of simplification one can say that the curvature is determined by the excess or deficit content of an infinitesimal piece of surface. Imagine, e.g., a piece of paper that was soaked with a spill of water and now starts to bend because the fibers extend locally at the soaked spot. Similarly, a circle with a given radius encloses a smaller amount of the surface if it is drawn on a sphere than on a flat piece of paper. Conversely, the enclosed surface is larger for a circle drawn on a saddle, i.e. a surface of negative curvature.

The relation to computer networks becomes immediately clear when one considers the number of network nodes that can be reached from a given node within a certain number of hops. This is the natural analogon of surface content and a typical property that is studied with random graphs and small-world networks. Recently, some confusion arose from the question why the number of nodes reachable by flooding with a given time-to-live (TTL) did not increase with the TTL to the extend that was naively expected. It was even speculated that the missing nodes got somehow isolated from the network. The concept of curvature, however, demonstrates that deficits in the size of reachable areas are quite natural. There is no hidden land on earth. The earth's surface is smaller than naively calculated from a flat map because the earth is a sphere and not a flat disk. The very same argument applies to the size of networks, like, e.g., Gnutella.

Before analyzing the size and structure of such random graphs in detail, another fundamental mathematical concept needs to be introduced, the *dimension* of a set. It, too, deals with the rate of increase in reachable areas when the maximal traversed distance is increased. Without mathematical strictness, a set can be defined to have dimension d if the size of the area reachable within a distance r increases (for $r \to 0$) proportional to r^d. E.g., the area of a disk in a flat two-dimensional manifold increases with πr^2. On the unit sphere, it increases with $4\pi \sin^2 \frac{r}{2}$ which, for $r \to 0$ is again πr^2. So, both cases yield expectedly $d = 2$. [11] gives many illustrative examples for fractal sets, i.e. sets with non-integer dimension.

Summarizing this short overview, one can say that a topology describes a set without reference to a metric. With a metric, a set can be described further: On small scales, the rate of increase of the reachable parts of a set is measured by its dimension. On larger scales, the excess or deficit in the size of the reachable parts is described by the curvature. Of course, this summary is rough and lacks mathematical strictness. But it can serve as motivation for the terminology used in the following sections.

IV. Dimension and Curvature of Sensor Networks

Although, these concepts are well-known tools in mathematics and theoretical physics, its use for the study of networks seems to be largely unknown. While questions of connectedness in ad-hoc and sensor networks and the associated phase-transition have been studied with random-graph approaches already [12], [13], [14], the use of concepts like dimension and especially curvature has, to the best of our knowledge, not been considered so far.

In order to motivate the use of these concepts for the classification of overlay-network models, we give an intuitive example for sensor networks first. Section VI will then apply the same method and concept to a random graph model showing that certain types of overlay-networks can be understood as fractal sets with curvature.

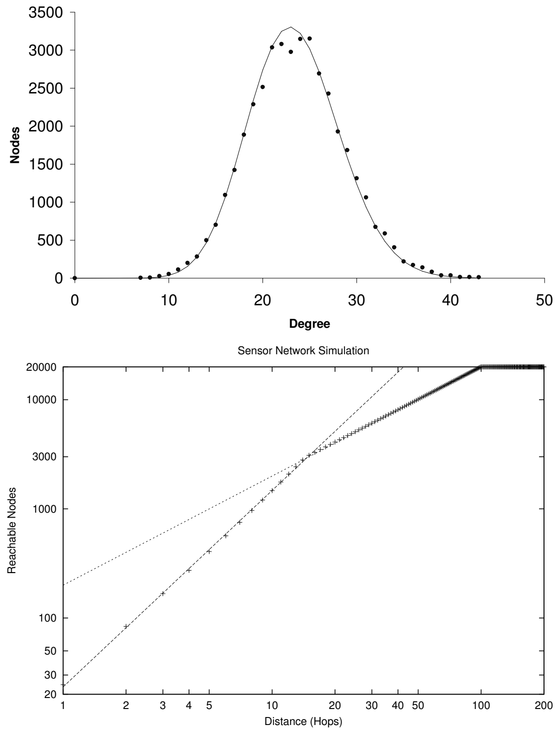

Fig. 1. Degree distribution (top) and node distribution (bottom) from a 20 000 nodes simulation of the sensor network model on a 30 × 200 torus.

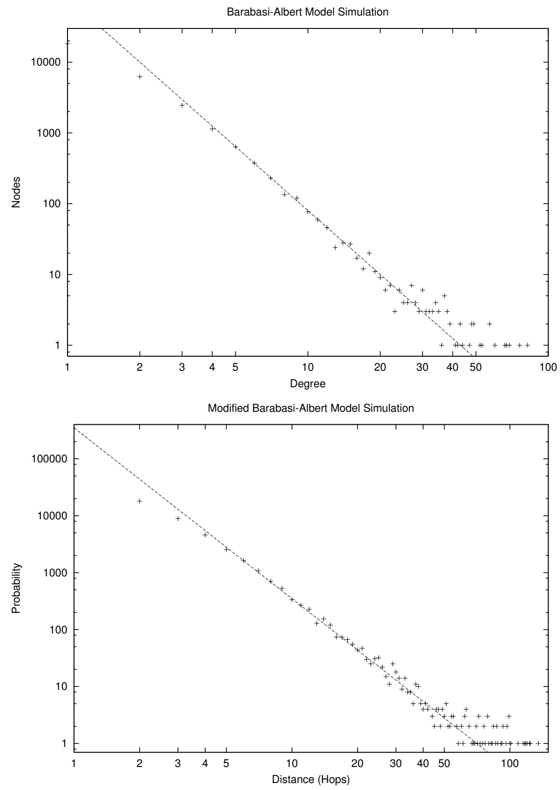

Fig. 2. Degree distribution from a 40 000 nodes simulation of the Barabasi-Albert model (left) and the modified Barabasi-Albert model (right)

Consider two-dimensional surface, e.g., a sphere or torus, that is covered with an equal distribution of nodes, that can communicate with one-another via a wireless link. Each node directly connects to all neighbors that are not farther than a fixed given distance (*sensor network model*). The question that is studied here is: *How many nodes can be reached within a given number of hops?*

Figure 1 shows the result of a simple simulation of this model on a torus. 20 000 nodes are equally distributed on a torus consisting of 30 × 200 unit squares. Each node's radio link covers a unit disk. The upper graph shows the simulated degree distribution as compared to the theoretically expected Poisson distribution. The lower graph shows how the average number of reachable nodes increases with the distance. The graph clearly shows the three regimes: at small distances the torus appears to be flat, at larger distances it appears to be a "'one-dimensional"' pipe, and at very large distances complete coverage of the whole torus ends the increase in nodes.

This sensor-network example is intuitive and straightforward. But astonishingly, the same kind of analysis can be applied to overlay-networks, too. Clearly, there, the graph structure does not necessarily reflect the topology of an underlying environment since overlay-networks are virtual networks. Nevertheless, as viewed from within the overlay network its topology can be very similar to the sensor-network example that was just presented.

V. Modeling Gnutella-like networks

Recently, the great interest in models for peer-to-peer networks, especially for Gnutella-like networks, and the discovery that many real-life networks can be described by random-graph models with power-law degree distribution has lead to the conviction that Gnutella-like networks can be modeled by such graphs, too. On the other hand, recent measurements seem to contradict this assumption [10], [15].

The simulation experiments described below show that both conclusions are only partly correct. They indicate that the measured properties of Gnutella can in fact be described by a model with a power-law degree distribution. But the required model differs from the Barabasi-Albert model of the world-wide-web link graph. In this section the degree distribution of the two model types is analyzed. In the following section the difference between the two models is further studied with the notion of dimension and curvature illustrated above.

The Barabasi-Albert model builds up a random graph by gradually adding new vertices and edges such that the probability of an existing vertex gaining one of the new edges is proportional to its degree. A mathematical analysis of this model [16] predicts a degree distribution of the resulting network that follows a power law $P(degree = x) \propto x^{\alpha}$ with $\alpha = -3$. This prediction is in good accordance with simulations (see [16] and figure 2).

Unlike this model, with Gnutella, servants create more than one initial link to the Gnutella overlay-network. This behavior can be simulated by a modified Barabasi-Albert model where new nodes create more than one initial connection. Figure 2 shows that this modification does not change the power-law structure of the resulting graph.

Recent measurements, however, found that the degree distribution in Gnutella does not obey a pure power-law as would follow from the Barabasi-Albert model [10], [15]. Even more, unlike the world-wide-web, Gnutella can be assumed to respect some amount of locality in its network since the targets for new links for a node are found via the *ping-pong* algorithm that seeks for nodes in a TTL-limited neighborhood of the respective node.

We hence propose the following model to describe the behavior of Gnutella and similar networks. This model also incorporates the recent measurement results for the *ping-pong* mechanism and the uptime of Gnutella servants [17]:

> *The network model consists of two types of nodes: Persistent nodes maintain long-lived connections to the Gnutella overlay network. Non-persistent nodes, e.g., servants with dial-up connections to the Internet, connect and disconnect rather frequently. (Note that in typical dial-up networks, reconnecting servants obtain a new IP address and thus appear as new nodes in the network.)*
> *Upon connection, each node creates one link to some node of the network. This node is chosen as a* 3-neighbor *of a randomly selected node. The term* 3-neighbor *describes a node that is at most three hops away from the randomly selected node.*

In Gnutella neighbors are discovered by the ping-pong mechanism. For the model presented here, all the details of this mechanisms are neglected. Only the principal mechanism of neighbor-mediated link creation is maintained. This is an important difference to the BARABASI-ALBERT model that does not respect any locality in the creation of new links and that is hence *not* suited to model Gnutella-like networks. This difference is most important for the following refinement step that models the connectivity strengthening of a ripening Gnutella network:

> *While non-persistent nodes are assumed to create all links upon initial connection to the network, persistent nodes remain connected long enough to create additional links to other nodes. However, the rate of this connection growth will typically be assumed to be small.*

Figure 3 shows results from a simulation run for this model that contained 40 000 nodes of which 15 000 were assumed to be persistent. These nodes gained at least 5 additional links each. The total number of links per node was limited to at most 100.

In the plot in figure 3 one can clearly distinguish two components: For small degrees (one to three links per node) the network exhibits a power-law for the degree distribution. This is in accordance to the expectations for small-world networks. At higher degrees the distribution rises again and reaches a local maximum at a degree of about 10. Beyond that maximum the degree distribution again follows the expected power-law. Both power-laws have the same parameter, here, $\alpha = 3.8$.

This pattern found by the simulation of the proposed model reflects the structure found in measurements of the Gnutella network [10], [15]. In these publications, however, the conclusion was drawn that a ripened Gnutella network did not follow a power-law for the degree distribution. With the model described above it now becomes clear, that this conclusion is not correct and that Gnutella-like network can in fact effortlessly be described by a two-regime structure of persistent and non-persistent nodes, both of which exhibit a power-law for the degree distribution. The exact shape of the degree distribution is governed by the ratio of persistent and non-persistent nodes.

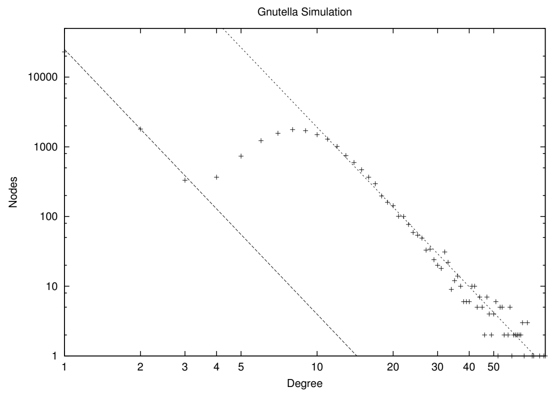

Fig. 3. Degree distribution from the simulation of the model for a Gnutella-like random graph (40 000 nodes of which 15 000 are persistent)

The following section will analyze further properties of this model. This will demonstrate that the simple notion of Gnutella-like networks being small-world networks with power-law degree distribution does not suffice to describe the properties of such networks.

VI. CURVATURE AND DIMENSION OF RANDOM GRAPHS

So far, the discussion has only considered the degree distribution of the two network models. The term *small-world* network, however, addresses the question of distance, measured as hop count. It seems that both issues are not always clearly distinguished in the literature. Therefore, the following discussion will address the question of the *dimension* of a graph. This is the natural analogon of the concept of a dimension of a fractal set. It, too, obeys a power law. But this power law must not be confused with the power law of the degree distribution discussed in the previous section.

As described above, the dimension describes — on small scales — the increase in the number of reachable nodes when the time-to-live is increased. In order to sensibly assign a dimension to a graph this increase must follow a power law, $n \propto r^d$, where d is the dimension of the graph. Hence, of course, not all graphs have a dimension, e.g., a tree cannot be assigned a dimension.

Following the analogy from section III, on larger scale, the increase in reachable nodes can be described as governed by the curvature of the graph. Again, not for all graphs this definition is mathematically sensibly possible. Even more, for a graph model that does not single out certain regions of the graph, the curvature must be expected to be constant, so that the graph needs to described as a *sphere*. (Mathematically, a surface of constant positive curvature is a sphere.) From this follows that

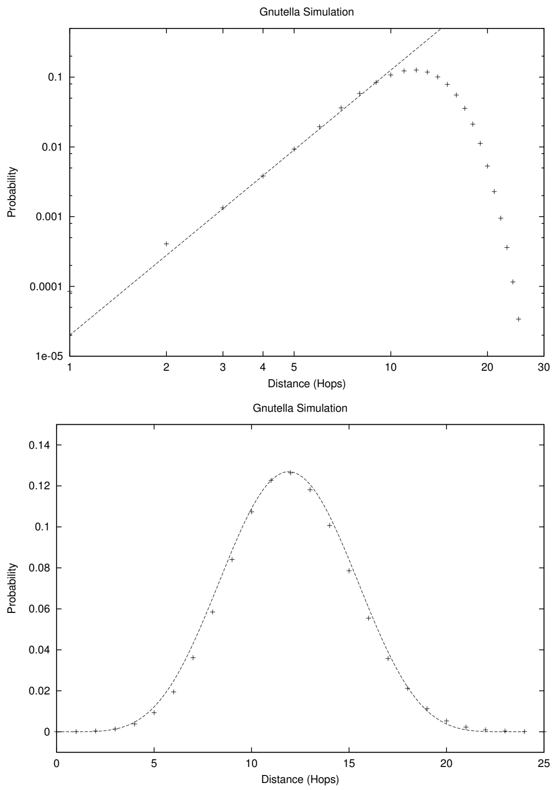

Fig. 4. Simulation of a Gnutella-like random graph: Probability distribution for distances between pairs of nodes as compared to the theoretical prediction. (Both graphs show the same data, see text.)

the number of nodes reachable within time-to-live r is proportional to $sin(\frac{r}{2})^d$. Curvature, more exactly, constant curvature, is thus another important criteria to distinguish graphs.

In the following, we present the simulation results for the Barabasi-Albert model, the modified Barabasi-Albert model, and our Gnutella model. They will show that the latter model indeed leads to a sphere structure with fractional dimension. The two other models can still be assigned a dimension, but they fail to be describable by a constant curvature. Since both concepts, dimension and curvature, lead to important topological consequences for a network, this analysis is a valuable tool for the study of network models.

A. Gnutella-like Graphs

Figure 4 shows the comparison of the described simulation with the theoretical prediction from a graph with fractal dimension and constant curvature. Both viewgraphs show the same simulation of 40 000 nodes presented above.

Since for practical purposes, it is convenient to measure the rate of increase in the number of reachable nodes, the graphs do *not* show cumulated node counts, but the number of nodes that have exactly the given distance from a randomly selected node.

The upper graph demonstrates that, up to about 8 to 10 hops, the distribution is described by a power-law. The fitted line's slope directly yields the dimension of the graph, namely 4.8. The lower graph shows that the simulation outcome is also well described by the assumption of a constant curvature. The fit corresponds to a quadrant length of 11.9 hops, where quadrant length means half the maximum distance within the graph. (In graph theory the maximum distance within the graph is called *diameter*. In combination with the sphere analogon this term is misleading since the geometrical and graph theoretical diameter differ by a factor of $\frac{\pi}{2}$.)

Both, dimension and quadrant length, depend on the parameters of the simulation:

Persistent nodes	Total nodes	Dimension	Quadrant length
25000	80000	5.1	12.6
25000	60000	5.1	12.2
25000	40000	4.8	12.2
15000	40000	4.8	11.9
5000	40000	4.8	11.9
5000	25000	4.4	11.2
5000	10000	4.1	10.2

The nature of this dependence needs to be explored by further studies beyond the scope of this paper.

B. Barabasi-Albert Graphs

Doing the same analysis for the Barabasi-Albert models yields a different result. Figure 5 shows a simulation of 40 000 nodes. The graph has a dimension of 4.55, but there is a significant deviation from the constant curvature model. Up to about 12 hops the model is still in good accordance with a constant curvature with quadrant length 11.0. But there is a significant higher fraction of nodes with a distance of 15 and more hops than could be expected based on the constant curvature model.

With the modified Barabasi-Albert model, the constant curvature assumption fails completely (cf. fig. 6). The resulting graph has a dimension of 7.0, but the distance distribution falls off rapidly at about 6 hops instead of showing a quadrant length of 7.3 as is indicated at small distances.

This analysis shows that the structure of an overlay-network has important consequences for the node distribution. If a network is known to have a sphere structure, the node distribution can be well predicted. This knowledge can then be used to choose the protocol parameters for optimized performance. If that knowledge lacks or, even worse, the wrong model is chosen, performance is greatly degraded. As can be seen from the latter example, small-world networks can be easily misjudged with regard to the expected node distance. As a consequence, the network is either unnecessarily flooded (distances overestimated) or, e.g., a search fails because too few nodes receive a message (distance underestimated).

VII. Conclusion and Outlook

This paper has illustrated how the mathematical concepts of dimension and curvature can be applied to the study of random graphs and especially to Gnutella-like overlay networks. These concepts are intuitively derived from a sensor-network model, where nodes are equally distributed on a surface, e.g., on a sphere. Although the direct intuition does not carry over to an overlay-network, i.e. a potentially purely virtual network, these concepts yield a simple model that describes the distribution of the node-to-node distances found in certain random

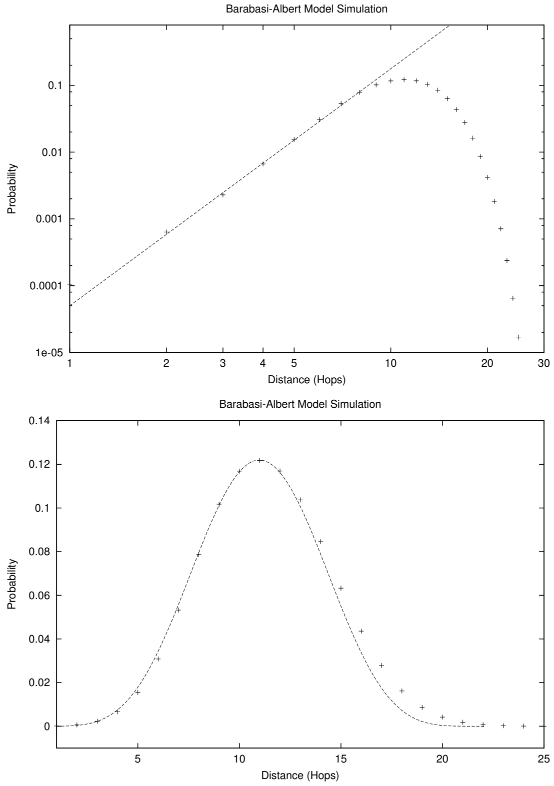

Fig. 5. Simulation of a Barabasi-Albert random graph: Probability distribution for distances between pairs of nodes as compared to the theoretical prediction. (Both graphs show the same data, see text.)

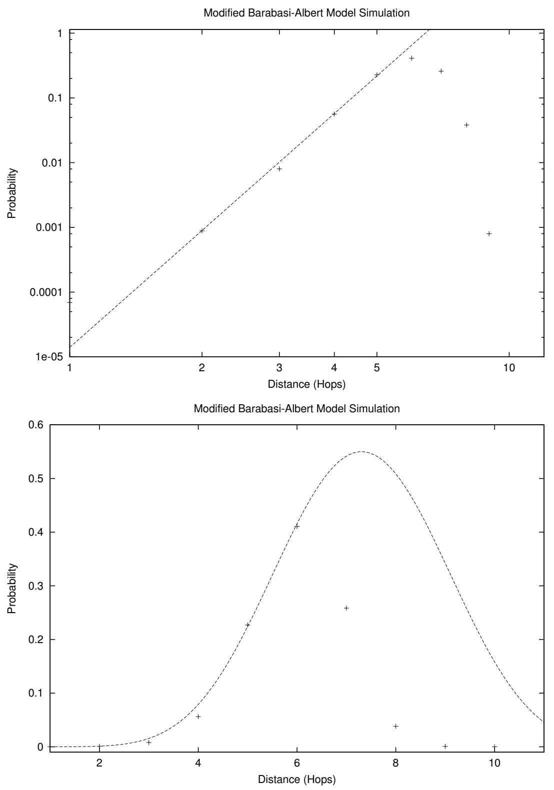

Fig. 6. Simulation of a modified Barabasi-Albert random graph with two initial links for each node (60 000 nodes). Clearly, the simulation outcome cannot be explained by the curvature assumption (plotted curve).

graphs. Besides its general importance as a classifying property of random-graphs, this property is also very important for the understanding and improvement of many peer-to-peer networks, especially Gnutella-like networks.

This paper has also demonstrated how the presented concepts of a dimension and curvature of a graph together with the proposed model for Gnutella-like networks lead to the picture of *Gnutella as a sphere with fractal dimension*. The fact that the model's degree distribution well reflects the measurements by RIPEANU indicates that this model might actually describe the properties of the Gnutella network better than established models. However, before this conclusion can be drawn, further measurements of the dimension and curvature, as defined in this paper, are needed to confirm the simulation results presented in this paper.

REFERENCES

[1] P. Erdös and A. Renyi, "On random graphs", *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.
[2] P. Erdös and A. Renyi, "On the evolution of random graphs", *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks", *Nature*, vol. 393, pp. 440–442, 1998.
[4] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science*, vol. 286, pp. 509–512, 1999.
[5] R. Albert and A.-L. Barabasi, "Topology of evolving networks: Local events and universality", *Physical Review Letters*, vol. 85, no. 24, 2000.
[6] S. Milgram, "The small world problem", *Psychol. Today*, vol. 2, pp. 60–67, 1967.
[7] R. Albert, H. Jeong, and A.-L. Barabasi, "Diameter of the world wide web", *Nature*, vol. 401, pp. 130–131, Sept. 1999.
[8] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology", in *SIGCOMM*, 1999, pp. 251–262.
[9] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in internet topologies revisited", in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002)*, 2002, vol. 2, pp. 608–617.
[10] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design", *IEEE Internet Computing Journal (Special issue on peer-to-peer networking)*, vol. 6, no. 1, 2002.
[11] B. B. Mandelbrot, *The Fractal Geometry of Nature*, H. W. Freeman and Company, New York, 1977.
[12] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks", in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, New York, 2002, vol. 2, pp. 1079 –1088.
[13] A. Faragó, "Scalable analysis and design of ad hoc networks via random graph theory", in *Proceedings of the 6th international workshop on Discrete algorithms and methods for mobile computing and communications*, Atlanta, Georgia, 2002, pp. 43 – 50.
[14] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network", in *Proceedings of the third ACM international symposium on Mobile ad hoc networking & computing*, Lausanne, Switzerland, June 2002, pp. 80–91.
[15] M. Ripeanu and I. Foster, "Mapping gnutella network: Macroscopic properties of large-scale peer-to-peer systems", in *Proceedings of the First International Workshop on Peer-to-Peer Systems (IPTPS'02)*, Cambridge, Massachusetts, Mar. 2002.
[16] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks", *Review of Modern Physics*, vol. 74, no. 47, 2002.
[17] K. Tutschku and H. deMeer, "A measurement study on the dynamics of gnutella overlays", in *Proceedings of Kommunikation in Verteilten Systemen*, Leipzig, Germany, February 26-28 2003, pp. 295–306.