Small world graphs by iterated local edge formation

Ph. Blanchard and T. Krueger
Faculty of Physics and Research Center Bielefeld-Bonn-Stochastics, University of Bielefeld, D-33615 Bielefeld, Germany

A. Ruschhaupt
Departamento de Química-Física, Universidad del País Vasco, Bilbao, Spain
(Received 23 July 2004; published 27 April 2005)

We study graphs obtained by successive creation and destruction of edges into small neighborhoods of the vertices. Starting with a circle graph of large diameter we obtain small world graphs with logarithmic diameter, high clustering coefficients, and a fat tail distribution for the degree. Only local edge formation processes are involved and no preferential attachment was used. Furthermore, we found an interesting phase transition with respect to the initial conditions.

DOI: 10.1103/PhysRevE.71.046139 PACS number(s): 89.75.Hc, 89.65.−s

I. INTRODUCTION

In 1998 Watts and Strogatz introduced in a seminal paper small world networks and provided a simple model of such graphs [1,2]. The key property of small world graphs is the simultaneous presence of small diameter (at most logarithmic in the vertex size) and high clustering coefficient. Extensive investigations over the last decade of real networks like the World Wide Web, citation networks, friendship networks, and many others have shown that these graphs are really of small world type. But there is a second remarkable property shared by most real networks, namely, that they have a power-law-like fat tail distribution (see [3] for an excellent recent survey about the whole subject).

There are many models that produce graphs with small diameter and power law distribution. If additionally local search rules are incorporated one easily obtains high clustering as well. But until recently it was widely believed that for the small diameter property a certain amount of essentially independent random edge formations was necessary, as in classical random graphs. To obtain a scale-free distribution for the degree the basic conviction was that the random edge formation should be biased according to the principle of "the richer you are the richer you get" [4] where richness is just measured by the degree of a vertex. Several authors [6–11] have recently proposed models where iterated local edge formation processes are explicitly built in. Some of them, like [6], keep the number of nodes fixed as we do, but mostly evolutionary graphs are considered. In all these models the emergence of small world networks can be seen and furthermore the resulting graphs show a fat tail degree distribution. In [7] some mean field equations for an evolutionary version of the model in [6] are given, supporting the conjecture that an iterated local edge formation process provides a mechanism for scale-freeness as well. However, in all these models a positive fraction of the edges is necessarily generated at random (this is hard to avoid in evolutionary graphs since for the first edge a new incoming vertex has to make, no neighbors or "friend" exist yet). Since these randomly generated edges cause already a small diameter substructure we were motivated to study a model where no long ranging random edges at all are formed. To illuminate the effect of local edge formation processes we introduce in the present paper a model that is entirely based on local search rules and admits all the three above mentioned properties. The results are mainly based on numerical simulations.

II. THE ITERATED "MY FRIENDS ARE YOUR FRIENDS" PRINCIPLE

Looking at the edge formation process in social networks there are essentially two ways to get a new contact. The first is by the random event of meeting somebody, for instance, in a train or airplane. Humans refer to this kind of relation creation often as fate. Second, there are contacts which are created by the local rule "let me introduce you to one of my friends," certainly a very common process in real life. This type of relation is well known in social sciences where it is called transitivity [7]. In the language of graphs this translates to the formation of an edge between two vertices, say x and y, which had distance 2 with respect to the underlying graph metric. We refer to this kind of edge formation as the "my friends are your friends" principle (FFP). As already mentioned in the Introduction growing graphs which were partially built by using the FFP have already been described and analyzed in [5]. In the following we want to describe a model based on the iteration of this principle—the IFF model.

III. THE MODEL DESCRIPTION AND RESULTS

In detail, we want to discuss the following model of a time dependent random graph space $\mathcal{G}_t(N,t_0)$ with N vertices and an explicit integer parameter t_0 tuning the total number of edges. We start with a circular chain of N vertices with the usual nearest neighbor edges, i.e., vertex 1 is connected to vertices N and 2, vertex 2 is connected to vertices 1 and 3, and so on. We call this configuration a circle and refer to the corresponding edges as fixed base edges. The random graph space $\mathcal{G}_t(N,t_0)$ will be defined through the following algorithm. In each of the first t_0 time steps, a vertex, say x, is chosen randomly. From the set of neighbors $N_1(x,t)$ of this vertex (at the given time t) one element, say y, is chosen

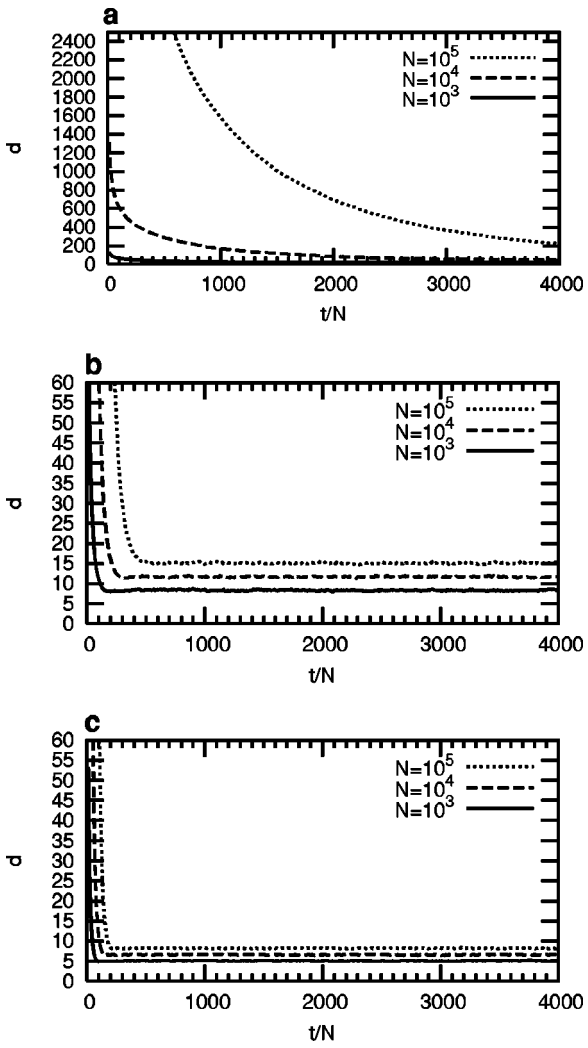

FIG. 1. Diameter of $\mathcal{G}_t(N,t_0)$, d, versus time, averaged over 100 realizations; $t_0/N=$ (a) 2, (b) 4, and (c) 10.

randomly, and then another element, say z, randomly from $N_1(y,t)/\backslash\{x\}$—the neighbor set of y at time t with the exclusion of x. If there is no edge between z and x a new one is created between the two vertices, otherwise nothing happens. In the first case we say that x has created a FF edge to z, or z was chosen by x. After repeating t_0 times the above procedure one continues for $t>t_0$ in the same way with the additional rule that whenever a vertex, say x, has created a new FF edge an FF edge containing x is deleted randomly. Note that the total number of edges stays therefore constant for $t>t_0$ and the expected value is proportional to t_0.

A. The diameter

The first surprising observation is the collapse of the diameter from const$\times N$ to const$\times \log N$ for sufficiently large t, t_0, and typical elements of $\mathcal{G}_t(N,t_0)$. This happens despite the fact that only local edge formations were used. Figure 1 shows the diameter versus t for different N and t_0. Our simulations indicate that in the case $t_0/N \gtrsim 2$ one gets an asymptotic regime in which the diameter is very small and

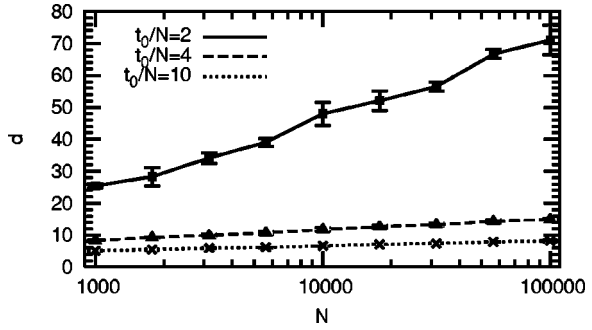

FIG. 2. Asymptotic diameter of the circle system (points connected with lines) calculated at $t/N=10^4$ ($t_0/N=2$) or $t/N=4\times10^3$ ($t_0/N=4, t_0/N=10$), averaged over 100 realizations, with error bars.

shows a logarithmic dependence on N. Figure 2 shows the asymptotic diameter for three different values of t_0/N as a function of N. (The error is defined by the absolute difference between the value averaged over ten realizations and that averaged over 100 realizations.)

The asymptotic diameter is approximately proportional to $\log N$ for $t_0/N>2$, which is one of the characteristics of small world graphs. In the case $t_0/N=2$ an asymptotic regime with still small diameter but large fluctuations starts approximately at $t/N\approx10^4$. Note that this value is much bigger than the collapse time in the cases $t/N=4,10$. It seems that $t_0/N\approx2$ is just the borderline where a collapse of the diameter appears. This corresponds to the case where the mean number of nonbase edges equals N. For smaller values of t_0 we found the diameter to stay of order $f(t_0)N$ with an almost linear function f (see Fig. 3).

This indicates an interesting phase transition in the total number of edges reminiscent of the famous phase transition in the size of the largest component in classical random graph spaces like $G(N,M)$—the space of all graphs with N vertices and M edges equipped with the uniform probability measure.

Figure 4 shows one realization of the time evolution of a typical graph from $\mathcal{G}_t(N,t_0)$ with $t_0/N=4$ and $N=10^3$. It illustrates how the diameter of the graph becomes very small only by using the IFF principle. Before the asymptotic state is reached an interesting symmetry breaking can be observed: the appearance of a few components with long range edges (taking the circle as a reference system) and hence small diameter. The components themselves are only connected via the circle skeleton. It takes a relative long time until the components finally merge and the circle is filled uniformly with edges. The same phenomenon can be seen for larger values of N and different t_0/N.

It is interesting to compare our model with a version of the random graph space $G(N,M)$ where an underlying circle is added to keep the graph connected. To be precise let $G_c(N,M)$ be the random graph space where M edges are randomly added to a circle graph with N vertices. Figure 5 shows the diameters $\bar{d}_{\text{random}}$ of the average of a few samples of $G_c(N,M)$ as a function of M/N. Here the diameter clearly decays like (const$\times N)/M$ and reaches the $\log N$ regime already for very small values of M/N ($\sim$0.05). Note that this

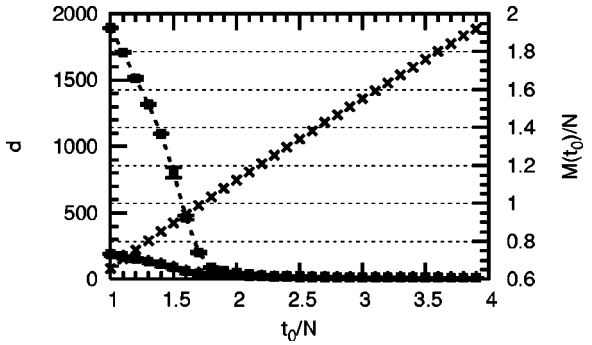

FIG. 3. Transition in the circle system. Left axis: mean diameter d calculated at $t/N=10^4$ for $N=10^3$ (triangles with error bars connected by a solid line), and $N=10^4$ (boxes with error bars connected by a dashed line). Right axis: effective nonbase edges $M(t_0)$ (crosses) versus t_0/N calculated at $t/N=10^4$ for $N=10^4$. Averaged over 100 realizations.

value is much smaller than the threshold value $M/N\sim 0.5$ for the emergence of a giant component in $G(N,M)$. Let us remark that the $G_c(N,M)$ model is very close in spirit to the original small world model of Watts and Strogatz. The main difference is with respect to the clustering coefficient since the underlying skeleton—the circle—has zero clustering coefficient. Replacing the standard circle with next neighbor connections by one where also next-next neighbor connections are edges would give an essentially equivalent model to the Watts-Strogatz one.

B. Fat tail of the degree distribution

Another unexpected property of our model is the fat tail of the degree distribution. Figures 6(a) and 6(b) show the degree distribution for $N=10^5$ and $t_0/N=4(10)$ at various times.

Let us compare the degree distribution in the asymptotic regime with those of the random graph $G_c(N,M(t_0))$, where $M(t_0)$ is the expectation of the number of edges in the IFF random graph space $\mathcal{G}_t(N,t_0)$ [see Fig. 3 for a plot of $M(t_0)/N$]. The resulting degree distributions are shown in Figs. 6(a) and 6(b) by big dashed lines. At the beginning (that is for $t\sim t_0$) the IFF model has essentialy the degree distribution of $G_c(N,M(t_0))$ but in the asymptotic regime the distributions differ drastically since the IFF model has a fat tail.

We want to close this section with a remark about the degree preferences if a vertex x was chosen in the formation

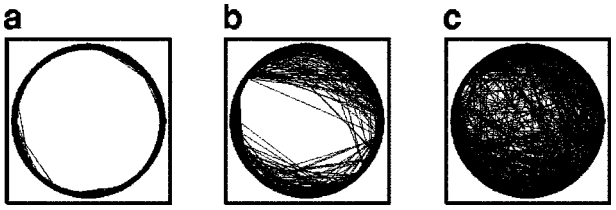

FIG. 4. Realization of a graph evolution using the circle base edge system; $N=10^3$, $t_0/N=4$, $t/N=$ (a) 60, (b) 100, and (c) 140.

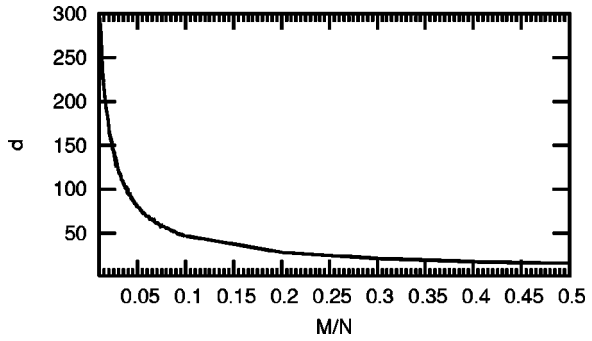

FIG. 5. Diameter $\overline{d}_{\text{random}}$ of $G_c(N,M)$ versus M/N, averaged over 100 realizations.

of a new edge. Although never explicitly included in the model there is strong numerical evidence (see Fig. 7) that the probability of a vertex to be chosen is proportional to its degree. That is one of the basic assumptions in the Albert-Barabási model. A heuristic explanation for this property in our model is given by the following argument based on two independence assumptions. Let N_1 and N_2 be the expected values for the first and second neighborhood sizes. Assume that the conditional expectation of $N_2(x)$ for x having degree k equals $k(N_1-1)$ and assume further that for $z\in N_2(x|d(x)=k)$ the expected value of $N_2(z)$ equals N_2. Clearly, with these conditions one can easily compute the probability of a vertex x with degree k to be chosen within the process of an edge formation; it is equal to $[k(N_1-1)/N](1/N_2)$ and hence is proportional to k.

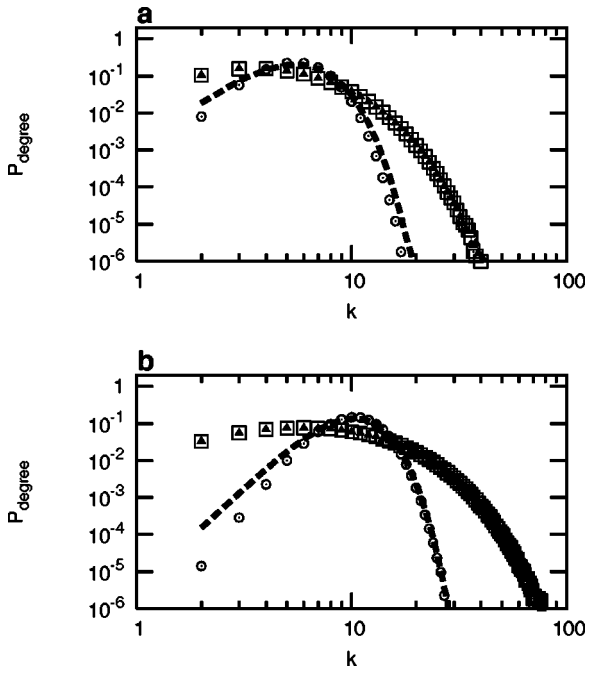

FIG. 6. Degree distribution using the circle system, $N=10^5$: (a) $t_0/N=4$, $t/N=4$ (circles), 600 (boxes), and 4000 (triangles); (b) $t_0/N=10$, $t/N=10$ (circles), 600 (boxes), and 4000 (triangles). The big dashed line shows the degree distribution of $G_c(N,M(t_0))$ $M(4N)/N=2.0$, $M(10N)/N=4.4$, all values averaged over 100 realizations.

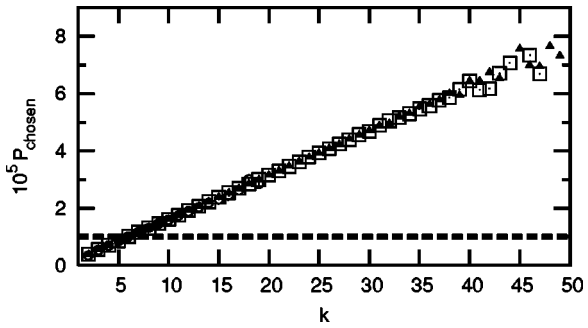

FIG. 7. Probability of a vertex to be chosen versus its degree using the circle base edge system; $N=10^5$, $t_0/N=4$, and $t/N=4$ (circles), 600 (boxes), and 4000 (triangles); the big dashed line shows the constant probability in $G(N,M)$; all values averaged over 100 realizations.

C. Clustering properties

A basic quantity to measure the local clustering around a vertex x is the number of triangles $C_3(x)$ containing x as a vertex. Note that for classical random graph spaces the expectation of this number is is of order N^{-1}. We are interested in the averaged number of triangles in the asymptotic regime of the IFF model. Figure 8 shows the mean number of triangles C_3 for all vertices with a fixed degree for $t/N=600$ and 4000. The mean number of triangles per vertex (independent of its degree) is shown by the big symbols on the figure frame. The plot for the two time values coincides practically. The number of triangles increases nearly linearly with the degree k. The reason for the fluctuations for high k is the low number of vertices with such high degrees. Figure 8 shows also the mean clustering coefficient C_c which is directly connected to the mean number of triangles; namely, the clustering coefficient $C_c(x)$ of a vertex x with degree k is given by the normalized triangle coefficient $2C_3(x)/k(k-1)$.

There is another interesting quantity, defined via the second shortest distance between two vertices, which characterizes additional clustering properties. For pairs of vertices with a common edge let $d_2(x,y)$ be the distance between x and y after removal of the connecting edge. Define for a given graph G the quantity $\varphi^{(n)}(G)$ as the fraction of pairs of vertices with distance 1 whose d_2 distance is larger than n [for a given random graph space $\mathcal{G}$ let $\varphi^{(n)}(\mathcal{G})$ be the expectation of $\varphi^{(n)}$]. In the case when the diameter of a random graph space is small due to the presence of independently generated edges the $\varphi^{(2)}$ value is usually very large since no short second-shortest paths between two vertices connected by an independently generated edge exist. This remains also true for the Watts-Strogatz small world graph. In our model we get by the very construction process a small value of $\varphi^{(2)}$ similar to the situation met in real networks (see Table I).

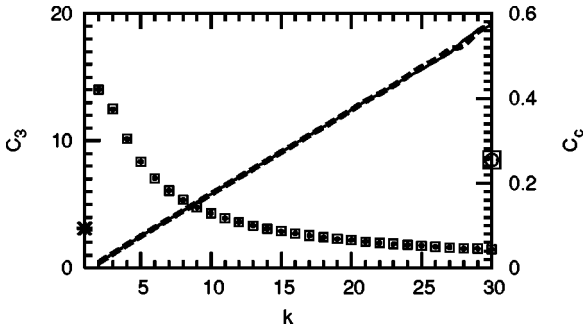

FIG. 8. Mean number of triangles C_3 and mean clustering coefficient C_c per vertex versus degree k; $N=10^5$, $t_0/N=4$; the big symbols with error bars show the mean values of all vertices independent of their degree; all values averaged over 100 realizations; the basic edges system forms a circle. Number of triangles (left axis): $t/N=600$ (dashed line and plus sign), 4000 (solid line and cross). Clustering coefficient (right axis): $t/N=600$ (boxes), 4000 (circles).

TABLE I. Mean value of $\varphi^{(n)}$ of the circle system, calculated with $N=10^5$ and after t steps, averaged over 100 realizations.

t_0/N	t/N	$\varphi^{(2)}$	$\varphi^{(3)}$
4	600	0.28±0.01	0.15±0.01
4	4000	0.28±0.01	0.15±0.01
10	600	0.30±0.01	0.16±0.01
10	4000	0.30±0.01	0.16±0.01

D. Replacing the circle by a torus

Up to now, the system of base edges forms a circle. It is natural to ask if one gets qualitatively the same results with a two-dimensional system of base edges like a torus lattice. This means that at the beginning the vertices are connected as a torus or a lattice with periodic boundaries. The algorithm to generate (and define) $\mathcal{G}_t(N,t_0)$ remains the same.

For the diameter one gets again an asymptotic regime with collapse up to logarithmic size but the asymptotic starts earlier than in the "circle" case. Even for $t_0/N=2$ we get an asymptotic regime starting at $t/N\approx 200$. Figure 9 shows the asymptotic diameter in the case of a torus base for various values of t_0 and N.

There is another difference from the one-dimensional case which can be seen in the realization shown in Fig. 10: the graph evolution is more uniform with a two-dimensional base edge system than with a one-dimensional one (compare to Fig. 4).

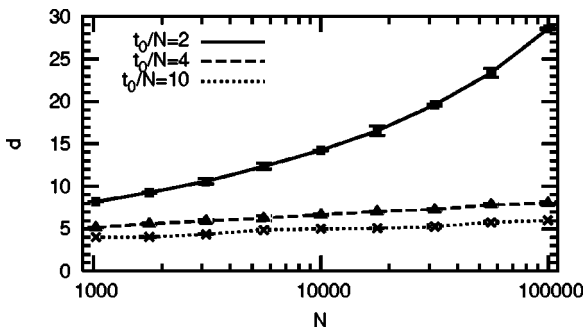

FIG. 9. Asymptotic diameter of the torus system (points connected with lines) calculated at $t/N=4\times10^3$, averaged over 100 realizations, with error bars.

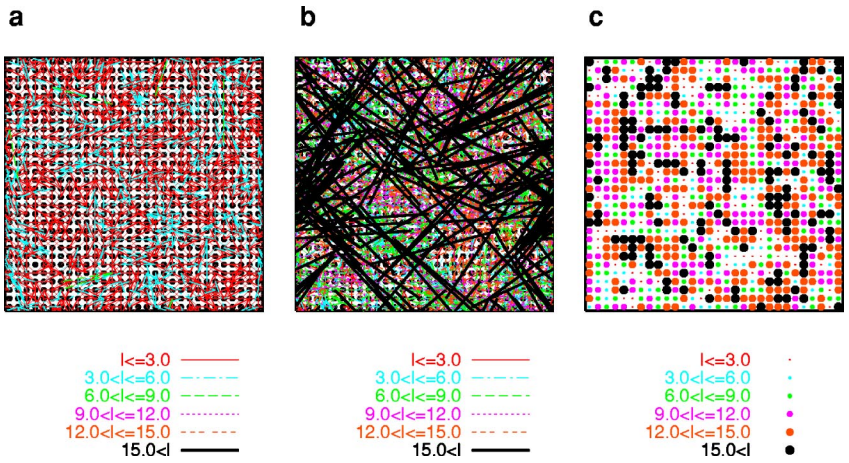

FIG. 10. (Color online) Realization using the torus system; $N=30^2$, $t_0/N=4$. (a), (b) The edges at different times are plotted. The dots characterize the vertices, the gray shade (color) of an edge characterizes its length l: $t/N=$ (a) 4, (b) 50. (c) The dots characterize the vertices, the gray shades (colors) characterize the length l of the longest edge of the vertex; $t/N=50$.

We obtain qualitatively the same results as in the one-dimensional case with respect to the degree distribution and the clustering properties for $N=300^2$ and $t_0/N=4$ and 10. In the asymptotic regime one sees again a clear fat tail distribution. The distribution of the number of triangles and the clustering coefficient for $N=300^2$ and $t_0/N=4$ gives qualitatively the same results as in the "circle" case, i.e., the distributions are nearly equal for $t/N=600$ and 400 and we have a linear dependence of the number of triangles on the degree k.

E. Removal of the base edge system

Finally we want to discuss the stability of the graphs in the asymptotic regime under removal of the circle edges or torus edges, i.e., at $t/N=1000$, we remove the base edges and examine in what way the graph splits into connected components and how these components evolve by continuing using our algorithm. Figure 11 shows the number of connected components with different size s (number of vertices) and different diameter d for $t_0/N=4$ directly after deleting the base edges at $t/N=1000$ for the "circle" and the "torus" cases. There is still one big component. The time evolution of the biggest component can be seen in Fig. 12 which shows its size s and diameter d versus time. An interesting result is that the diameter of the biggest component is nearly constant for $1500<t/N<4000$, nearly independent of t_0/N and also nearly independent of the use of a circle or a torus as the base edge system. Let us note that without the skeleton there can be no merging of different components by the very nature of the IFF principle. Therefore, as can easily be seen, the only stable asymptotic configurations consist of components that are totally connected (i.e., every vertex has an edge to all other vertices in the component) since no further changes in such components can happen. But it is very likely that the time scale until this phenomenon can be seen is huge compared to the time scales we were studying.

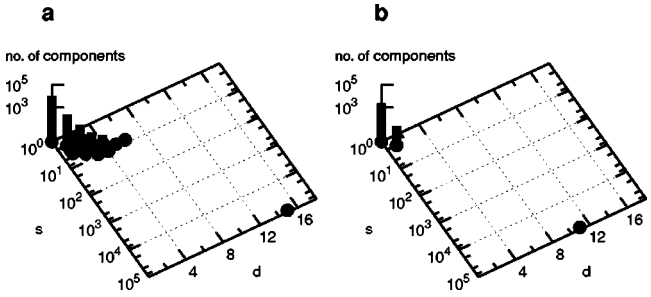

FIG. 11. Number of components with a given number of vertices and diameter, after deleting the basic edges. $t/N=1000$, $t_0/N=4$, one realization; (a) circle, $N=10^5$; (b) torus, $N=300^2$.

To compare with the classical random graph situation observe that the removal of the circle in $G_c(N,M)$ gives just the model $G(N,M)$. The component distribution for this random graph space is well known. Especially for $M>N/2$ there is always a giant component of size $K(M)\times N$ with $K(M)$ being an M-dependent constant. But in contrast to our model,

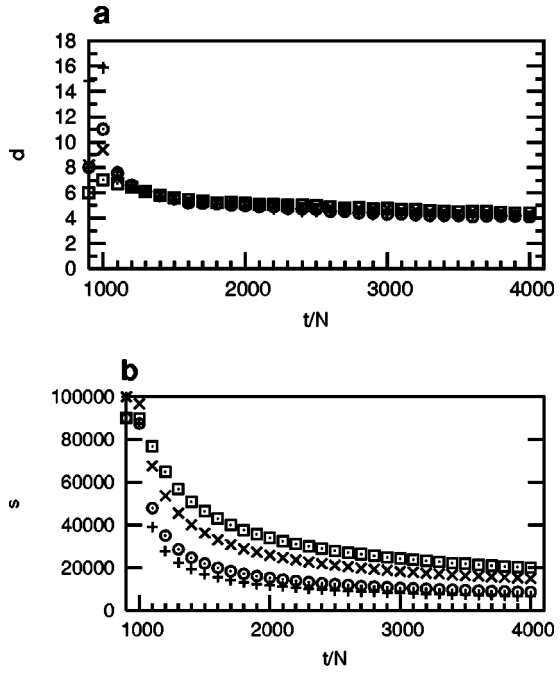

FIG. 12. Parameters of the biggest component after deleting the basic edges; (a) time dependence of its diameter d; (b) time dependence of its size s; circle, $N=10^5$, $t_0/N=4$ (plus signs), $t_0/N=10$ (crosses); torus, $N=300^2$, $t_0/N=4$ (circles), $t_0/N=10$ (boxes). The symbols on the left vertical axes indicate the values of the parameters directly before deleting the basic edges; all values averaged over 100 realizations.

the diameter d of the biggest component depends via $K(M)$ on $M(t_0)$.

IV. SUMMARY AND OUTLOOK

We have presented a model where the iteration of the "my friends are your friends" principle has produced random graphs with small world properties (logarithmic diameter, high clustering) and a fat tail distribution for the degree. The model used a fixed regular graph of large diameter as a skeleton, reminiscent of a pregiven geographical structure with pregiven neighborhood relations for vertex pairs. But in contrast to the small world graphs by Watts and Strogatz, no random, global edge formation processes are involved to obtain small diameter. The high clustering is an immediate consequence of the FF principle. The fat tail distribution is surprising since no preferential choice mechanisms are contained in the graph generation algorithm. So far our investigations are entirely numerical and clearly a more theoretical explanation of the observed phenomena is the desired next step.

There are several natural variants of the model which we would like to mention briefly.

First, in our model the total number of edges is kept fixed after the buildup phase. Instead of that one could use probabilistic rules which keep the number of edges only fixed in the mean. This would match better real situations where the FF principle is of relevance.

Second, as in the Albert and Barabási network [4], a growing number of vertices could be considered. Growth should happen here in the form of offspring of already existing vertices to be able to apply the FF principle to the "newcomers." In this situation it could well be the case that even without the iteration of the FF principle—every vertex when entering the network forms just once a number of edges according to the FF rule—small world graphs plus a fat tail for the degree distribution are obtainable. This model variant is actually the typical situation for the growth and formation of the network of citations and collaborations.

ACKNOWLEDGMENT

We would like to thank the DFG-Research Group 399 "Spectral Analysis, Asymptotic Distributions, and Stochastic Dynamics" for the support.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton University Press, Princeton, N.J., 1999).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] Ph. Blanchard and T. Krueger, J. Stat. Phys. **114**, 1399 (2004).
[6] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[7] A. Vazquez, Phys. Rev. E **67**, 056104 (2003).
[8] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[9] J. Jost and M. P. Joy, Phys. Rev. E **66**, 036126 (2002).
[10] G. Csanyi and B. Szendroi, e-print cond-mat/0305580.
[11] E. Volz, e-print cond-mat/0405381.