PHYSICAL REVIEW E **73**, 025101(R) (2006)

Statistics of cycles in large networks

Konstantin Klemm and Peter F. Stadler
Department of Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany
(Received 5 July 2005; published 3 February 2006)

The occurrence of self-avoiding closed paths (cycles) in networks is studied under varying rules of wiring. As a main result, we find that the dependence between network size N and typical cycle length is algebraic, $\langle h\rangle \propto N^{\alpha}$, with distinct values of α for different wiring rules. The Barabasi-Albert model has $\alpha=1$. Different preferential and nonpreferential attachment rules and the growing Internet graph yield $\alpha<1$. Computation of the statistics of cycles at arbitrary length is made possible by the introduction of an efficient sampling algorithm.

DOI: 10.1103/PhysRevE.73.025101 PACS number(s): 89.75.Hc, 02.70.Uu, 89.20.Hh

Physics research into graphs and networks has begun to provide a common framework for the analysis of complex systems in diverse areas including the Internet, biochemistry of living cells, ecosystems, and social communities [1–3]. The graph representation of these systems as discrete units coupled by links (nodes and edges) exhibits a large set of scaling phenomena including power-law distributed number of connections per node [4], fractal dimension [5], and hierarchy of modules [6]. After observation and classification of graph structure [7], enhanced understanding of complex systems is gained by considering the interplay between structure and dynamics. Few studies take into account the complete scenario of mutual influence between an evolving network and the states of the elements it connects: in the Internet, for instance, the wiring between routers determines the dynamic distribution of data traffic, which in turn drives an ongoing adaptation of the wiring structure [8]. Many results have been obtained for the reduced scenario of dynamics on a fixed network with predefined characteristics. Large-scale behavior of diffusion, synchronization, and the contact processes on realistic networks is qualitatively different from the solutions of the mean-field approximation or on periodic lattices [9–11].

Rather than dynamics on networks, we here consider dynamics of networks which is the complementary reduction of the above scenario. We ask which local rules for establishing connections between elements lead to which overall network properties. Much of the previous work in this direction has been focused on the scale-free nature of many natural and technical networks [1]. The fraction of nodes with a given number of edges, called degree k, decays as a power law, $P(k)\sim k^{-\gamma}$ for large k. This scaling has been found to result from preferential attachment of additional edges to elements that are well connected already [4].

Beyond the degree distribution, the identification of typical subnetworks reveals a nontrivial structure. Significantly frequent subnetworks are called motifs [12]. The most attention has been given to the motif of three fully interconnected nodes. An abundance of such triangles is taken as an indicator of local organization and clustering [13]. The triangle is the smallest element of the set of cycles that form a relevant class of subnetworks. The full range analysis of cycles from length three up to system size gives insight into network structure at all scales. Previous work on cycles includes approximations for the system size scaling of the number $c(h)$ of cycles of length h for various types of artificial networks [14–18]. It has been speculated [19] that for generic networks the distribution $c(h)$ becomes sharply peaked in the limit of large number of nodes, $N\to\infty$. For the position of the peak, an algebraic growth has been conjectured $\langle h\rangle \sim N^{\alpha}$ with an exponent $\alpha\leqslant 1$ as the leading characteristic [18].

Verification of these fundamental conjectures, validity checks of the analytical approximations, and comparisons with real-world networks have been difficult so far, since an efficient method for finding the length distribution of all cycles in a given network has been lacking. The much more modest computational task Hamilton cycle, to decide whether a given graph of N nodes has at least one cycle of maximal length N, falls into the class of NP-complete problems that are widely believed to be computationally intractable. Probably no algorithm exists that solves the Hamilton cycle in a number of steps that is bounded by a polynomial in N [20]. Exact counting of all cycles is feasible only in special cases, including planar graphs, cf. [21] and references therein. In the general case, approximation by efficient sampling appears to be the only possibility to numerically investigate long cycles. Taking a step in this direction, Rozenfeld and co-authors have introduced a stochastic search for cycles [19] as self-avoiding random walks on the network. Although the method allows for a quick scan of cycles on small networks, larger systems cannot be treated as the probability of finding a given cycle is strongly suppressed with growing cycle length.

Here we suggest an alternative method based on a Markov-chain Monte Carlo algorithm. Cycles are treated as discrete microstates of a physical system at equilibrium. Elementary transitions between cycles, the analogs of single spin flips in a spin system, are defined as an addition or removal of short detours with minimal change to cycle length. By considering cycle length as energy, generic Monte Carlo procedures from statistical mechanics become applicable. Temperature is defined in the usual way and allows us to tune the sampling on preferably long or short cycles. After introducing the algorithm in detail, we test its accuracy for a set of networks where the cycle length distribution is directly

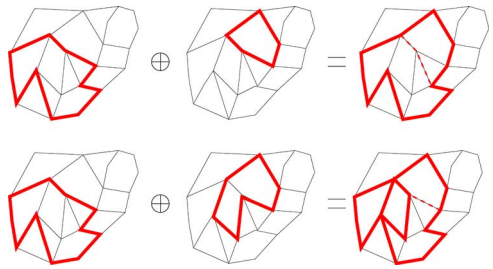

FIG. 1. (Color online) (a) Summation of two cycles resulting in a new cycle. Edges contained in either addend are contained in the sum. Edges present in both addends (dashed lines) cancel out. (b) Example of a sum of two cycles that is not a cycle itself.

accessible for comparison. We apply the algorithm to models of growing networks and find the growth exponent of the mean-cycle length. Finally, we test scaling of the number of cycles in the growing Internet.

The formulation of the algorithm uses the following basic notions of cycle space. We treat a subgraph X as the set of edges it contains. If X is a cycle, the cardinality $|X|$ is the cycle length. The sum of two subgraphs X and Y is defined as $X \oplus Y = (X \cup Y) \setminus (X \cap Y)$, i.e., an edge is contained in the sum if it is in one of the addends but not in both. The sum $X \oplus Y$ of two cycles X and Y is again a cycle if X and Y intersect in a suitable way, see Fig. 1. We generate a Markov chain of cycles $(C_0, C_1, C_2, \ldots)$ as follows. The initial condition is the empty graph $C_0 = 0$ at $t = 0$. At each step a cycle S is drawn at random from a set M of initially known cycles (the choice of M is described below). If the proposal $C' = C_t \oplus S$ is a cycle or the empty graph, it is accepted with probability

$$P_{\text{accept}} = \min\{\exp[-\beta(|C'| - |C_t|)], 1\}. \tag{1}$$

In case of acceptance we set $C_{t+1} = C'$, otherwise $C_{t+1} = C_t$. This is the Metropolis update scheme [22] with inverse temperature β and cycle length as energy. *Detailed balance* holds just as for the standard Metropolis algorithm [23].

Throughout this paper, we take M as the set of *short* (isometric) cycles of the given graph. A cycle S is short if for all nodes x and y on S, a shortest path between x and y lies also in S. As a nonshort cycle has at least one shortcut between two of its vertices, it can be decomposed into two shorter cycles that overlap on the shortcut. Typically for each nonshort cycle C one finds cycles S and C' such that S is short and $|C'| < |C|$. Applying the decomposition recursively, one sees that every cycle C occurs in a sequence $0, C_1, C_2, \ldots$ with $C_i \oplus C_{i+1} \in M$ and $|C_i| < |C_{i+1}|$. Thus taking as the possible "moves" M, the set of short cycles, not only ensures that every cycle can be reached (ergodicity). In this case, the resulting energy landscape does not have any local minima other than the unique global minimum, which is the empty graph at $E = 0$. There are exceptional graphs where the decomposability does not hold for one particular cycle. The exceptions appear to be irrelevant for the applications here as our numerical results remain unchanged when M is expanded to include more and longer (nonshort) cycles.

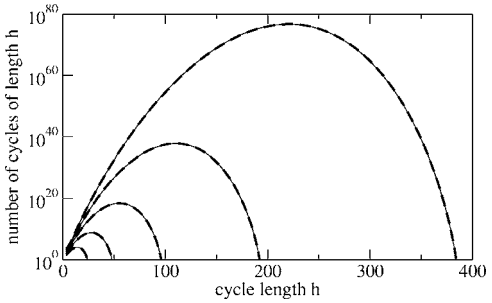

FIG. 2. Number $c(h)$ of cycles of length h estimated by the MC sampling algorithm (thick dashed curves) and the exact values from iterating Eq. (2) (thin solid curves). Studied networks are generations $n = 4, \ldots, 8$ (system sizes $N = 42$, 123, 366, 1095, and 3283 vertices) from the deterministic growth model [2]. The supplement [23] describes the generation of the histograms from runs at different temperatures.

Having defined the algorithm, we are going to study the formation of cycles as networks are growing under various attachment rules. To this end, we let the network grow to a certain size, keep it fixed while estimating the statistics of cycles, then let the network grow to a larger size, estimate the statistics again, and so forth. We first apply this procedure to a set of networks where exact computation of $c(h)$ is feasible. The pseudofractal scale-free web (PF) by Dorogovtsev and Mendes [24] grows deterministically by iterative triangle formation as follows. Start at generation $n = 0$ with two vertices connected by an edge. To obtain generation $n+1$, for each edge xy present in generation n add a new vertex z and the edges xz and yz, such that each existing edge xy becomes part of an additional triangle xyz. The calculation of $c(h)$ is particularly simple because each cycle has a unique predecessor in the previous generation, given by following direct links xy instead of the additional "detours" via z. A cycle of length h in generation n produces 2^h cycles in generation $n+1$ as the result of h binary decisions to follow the detour or the original direct edge. The histogram of cycle lengths iterates as

$$c^{(n+1)}(h) = \sum_{l=3}^{h} \binom{h}{h-l} c^{(n)}(l) \tag{2}$$

for $l \geqslant 4$ and $c^{(n+1)}(3) = c^{(n)}(3) + 3^n$. The result of the numerical iteration of these equations up to generation $n = 8$ is shown in Fig. 2, together with the results from the Monte Carlo method. The relative deviation of the sampling estimate of $c^{(n)}(h)$ from the exact value is below 25% for all cycle lengths h and all generations n. In particular, the unique cycle of maximum length $h_{\max} = 3 \times 2^n$ is detected. The method approximates the true numbers of cycles with high precision.

Next, we apply the algorithm to study the system size dependence of the cycle length distribution of stochastically growing artificial networks. All networks initiate as two vertices coupled by an edge. The networks grow by iterative attachment of vertices until a desired size N is reached. At each iteration, one new vertex z and two new edges xz and yz

TABLE I. Networks from different attachment rules and the resulting scaling exponents γ for the tail of the degree distribution and α for the growth of the cycle lengths. The last column displays the symbol used in the figure in the supplement [23].

Rule	Indep/Tri	Hom/Pref	α	γ	
IH [8]	Independent	Homogeneous	1.010(4)	∞	$\square$
IP [8]	Independent	Preferential	0.969(5)	3	$\bigcirc$
TH	Triangle	Homogeneous	0.722(5)	∞	∇
TP [24]	Triangle	Preferential	0.644(9)	3	$\triangle$
PF [23]	Triangle	Preferential	0.635(1)	2.59	$\diamond$
	Internet		0.76(4)	2.22(1)	*

are generated. We are interested in the influence of different attachment mechanisms on the cycle length distribution. To this end, we consider the following probabilistic rules for selection of the nodes x and y to which the new node z attaches. Independent homogeneous (IH) attachment: Draw x and y randomly (with equal probabilities) and independently from the set of nodes; if $x=y$, discard this choice and repeat. Independent preferential (IP) attachment: Draw an edge randomly (all edges having equal probability) and take as x one of the end vertices chosen with equal probability; draw another edge to find y analogously; if $x=y$, discard this choice and repeat. Triangle forming preferential (TP) attachment: Draw an edge randomly and take its two end vertices as x and y. Triangle forming homogeneous (TH) attachment: Draw an edge randomly, take x and y as its end vertices, and accept this choice with probability $1/[\deg(x)\deg(y)]$; otherwise reject and repeat.

Rule IP is equivalent to choosing nodes with probability proportional to degree [4], so-called preferential attachment. It generates scale-free networks with degree exponent $\gamma=3$. Rule TP implements preferential attachment with the additional constraint that x and y be connected; it is the stochastic version of the pseudofractal (PF) scale-free web [25] defined above. The resulting networks are scale-free with $\gamma=3$. The homogeneous attachment rule (IH) [4] leads to networks with exponentially decaying degree distribution ($\gamma=\infty$). The fourth rule introduced here combines triangle formation with homogeneous (TH) attachment by explicitly canceling out the degree dependence in the selection probability. We have checked that this rule generates an exponential degree distribution.

The mean-cycle length increases algebraically with system size,

$$\langle h\rangle \sim N^{\alpha}, \tag{3}$$

with the exponent $\alpha\in[0,1]$ depending on the attachment rule [23] The variance of the cycle length distribution increases algebraically with the same exponent α. Therefore the ratio of mean and variance is practically constant [23]. Considering the degree exponent γ and the cycle growth exponent α for each type of network (Table I), several observations are worth mentioning. Homogeneous attachment with triangle formation leads to a nontrivial cycle growth exponent $\alpha\approx 0.72$ even in the absence of scaling in the degree distribution $\gamma=\infty$. Networks grown stochastically with triangle formation and preferential attachment (rule TP) have the same exponent $\alpha\approx 0.64$ as the deterministic counterpart (rule PF) while the degree exponents under these two rules are clearly different. Analogously, in the absence of triangle formation (rules IH and IP) the same cycle growth exponent $\alpha\approx 1.0$ is obtained regardless of the degree exponents $\gamma\in\{3,\infty\}$.

Finally, we consider cycles in an evolving real-world network. The Internet at the level of Autonomous Systems is a growing scale-free network with degree exponent $\gamma=2.22(1)$ [26,27]. Here we analyze snapshots of the network with sizes from $N=3015$ nodes (November, 1998) to $N=10\,515$ nodes (March, 2001) [28]. We find that during this time the mean-cycle length grows from 264.9 to 757.8. As in the artificial growing networks, the growth is algebraic. The growth exponent is estimated as $\alpha=0.76(4)$ by a least squares fit. More detailed analysis is performed on the number $c(h,N)$ of cycles of given length h at system size N, cf. supplement [23]. We observe a scaling

$$c(h,N) \sim N^{\xi(h)}, \tag{4}$$

with an exponent $\xi(h)$ that depends linearly on h with a slope

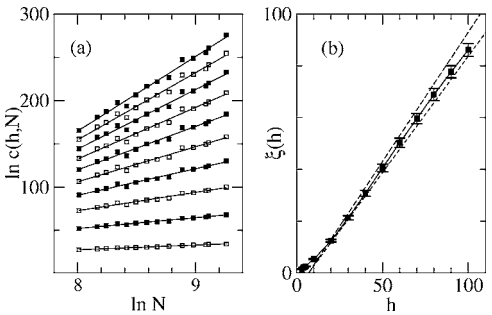

FIG. 3. Evolution of cycles in the growing Internet at the Autonomous Systems level. (a) The number of cycles of given length h as a function of system size N for $h=10,20,30,\ldots,100$ (squares, bottom to top). The straight lines are best fits of the form $c(h,N)\propto N^{\xi(h)}$. (b) Growth exponents $\xi(h)$ as defined in Eq. (4) obtained as slopes of the fitted lines in (a). Error bars of exponents indicate standard error from the fit. Dashed lines have slopes 1.0 and 0.9.

close to unity. Figure 3(b) shows that

$$\xi(h) \approx h \tag{5}$$

for not too small lengths $h \geqslant 10$. The scaling behavior is in qualitative agreement with the prediction from the first order approximation by Bianconi *et al.* [29], assuming that the Internet is a random network with a given scale-free degree distribution.

Many more results have been obtained that are not discussed in the present paper. Studying the influence of geometric constraints on networks, we have found the influence of the embedding dimension d on the cycle-length distribution of random graphs. While lowering d introduces locality and thus favors the existence of small cycles, the mean-cycle length $\langle h \rangle$ is practically invariant under changing dimension $d > 1$. Further, the present sampling method is applicable to arbitrary properties of cycles in graphs, including functions that depend explicitly on edge and node labels.

In summary, we have analyzed the evolution of cycles in growing networks. While the mean-cycle length grows with a characteristic exponent α, the relative width of the length distribution tends to zero as the system size increases. Thus, in agreement with an earlier speculation [18], the exponent α is found to be the most relevant quantity for the evolution of cycle space. In the scale-free model by Barabási and Albert [4] as well as the growth model with random homogeneous attachment, cycles are space filling ($\alpha = 1.0$), i.e., cycle length is proportional to system size. In model networks with explicit formation of triangles and in the Internet, however, cycles grow slower than the system as a whole. This class of networks having $\alpha < 1$ also includes single-scale networks with $\gamma = \infty$. Our study suggests that the cycle growth exponent serves as a useful characterization of growing networks independent of the degree exponent γ. An open question concerns universality. Can α be altered continuously by tuning parameters or does it assume distinct values, dividing the set of growing networks into universality classes?

We are grateful to C. P. Bonnington, J. Leydold, and A. Mosig for inspiring discussions. This work was supported by the DFG Bioinformatics Initiative BIZ-6/1-2.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] *Handbook of Graphs and Networks—From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin 2002).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[6] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[7] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[8] K.-I. Goh, B. Kahng, and D. Kim, e-print cond-mat/0410078.
[9] B. Kozma, M. B. Hastings, and G. Korniss, Phys. Rev. Lett. **95**, 018701 (2005).
[10] F. M. Atay, J. Jost, and A. Wende, Phys. Rev. Lett. **92**, 144101 (2004).
[11] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[12] R. Milo, R. S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[13] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[14] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
[15] E. Marinari and R. Monasson, J. Stat. Mech.: Theory Exp., P09004 (2004).
[16] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. E **71**, 026129 (2005).
[17] A. Vázquez, J. G. Oliveira, and A.-L. Barabási, Phys. Rev. E **71**, 025103(R) (2005).
[18] G. Bianconi and M. Marsili, J. Stat. Mech.: Theory Exp., P06005 (2005).
[19] H. D. Rozenfeld, J. E. Kirk, E. M. Bollt, and D. ben-Avraham, J. Phys. A **38**, 4589 (2005).
[20] M. R. Garey and D. B. Johnson, *Computers and Intractibility. A Guide to The Theory of NP-Completeness* (W. H. Freeman, San Francisco, 1979).
[21] M. M. Sysło, SIAM J. Comput. **10**, 797 (1981).
[22] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).
[23] See EPAPS Document No. E-PLEEE8-73-R22602 for more details on the algorithm and results. For more information on EPAPS, see http://www.aip.org/pubserves/epaps.html.
[24] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **65**, 066122 (2002).
[25] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[26] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[27] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[28] http://www.cosin.org.
[29] G. Bianconi, G. Caldarelli, and A. Capocci, Phys. Rev. E **71**, 066116 (2005).