IRis

Northeastern University

Physics Faculty Publications

Department of Physics

February 01, 2005

Inhomogeneous evolution of subgraphs and cycles in complex networks

Alexei Vázquez
University of Notre Dame

J. G. Oliveira
University of Notre Dame and Campus Universitário de Santiago

Albert-László Barabási
University of Notre Dame

Recommended Citation

Vázquez, Alexei; Oliveira, J. G.; and Barabási, Albert-László, "Inhomogeneous evolution of subgraphs and cycles in complex networks" (2005). *Physics Faculty Publications*. Paper 116. http://hdl.handle.net/2047/d20000689

This work is available open access, hosted by Northeastern University.

PHYSICAL REVIEW E **71**, 025103(R) (2005)

Inhomogeneous evolution of subgraphs and cycles in complex networks

Alexei Vázquez,[1] J. G. Oliveira,[1,2] and Albert-László Barabási[1]
[1]*Department of Physics and Center for Complex Network Research, University of Notre Dame, Indiana 46556, USA*
[2]*Departamento de Física, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*
(Received 18 October 2004; published 24 February 2005)

Subgraphs and cycles are often used to characterize the local properties of complex networks. Here we show that the subgraph structure of real networks is highly time dependent: as the network grows, the density of some subgraphs remains unchanged, while the density of others increase at a rate that is determined by the network's degree distribution and clustering properties. This inhomogeneous evolution process, supported by direct measurements on several real networks, leads to systematic shifts in the overall subgraph spectrum and to an inevitable overrepresentation of some subgraphs and cycles.

DOI: 10.1103/PhysRevE.71.025103 PACS number(s): 89.75.Hc, 87.10.+e, 89.75.Fb

Subgraphs, representing a subset of connected vertices in a graph, provide important information about the structure of many real networks. For example, in cellular regulatory networks feed-forward loops play a key role in processing regulatory information [1], while in protein interaction networks highly connected subgraphs represent evolutionary conserved groups of proteins [2]. In a similar vein, cycles, a special class of subgraphs, offer evidence for autonomous behavior in ecosystems [3], cyclical exchanges give stability to social structures [4], and cycles contribute to reader orientation in hypertext [5]. Finally, understanding the nature and frequency of cycles is important for uncovering the equilibrium properties of various network models [6].

Motivated by these practical and theoretical questions, recently a series of statistical tools have been introduced to evaluate the abundance of subgraphs [1,2,7] and cycles [8–11], offering a better description of a network's local organization. Yet, most of these methods were designed to capture the subgraph structure of a specific snapshot of a network, characterizing static graphs. Most real networks, however, are the result of a growth process, and continue to evolve in time [12]. While growth often leaves some of the network's global features unchanged, it does alter its local, subgraph-based structure, potentially modifying everything from subgraph densities to cycle abundance. Yet, the currently available statistical methods cannot anticipate or describe such potential changes.

In this paper we show that during growth the subgraph structure of complex networks undergoes a systematic reorganization. We find that the evolution of the relative subgraph and cycle abundance can be predicted from the degree distribution $P(k)$ and the degree-dependent average clustering coefficient $C(k)$. The results indicate that the subgraph composition of complex networks changes in a very inhomogeneous manner: while the density of many subgraphs is independent of the network size, they coexist with a class of subgraphs whose density increases at a subgraph-dependent rate as the network expands. Therefore, in the thermodynamic limit a few subgraphs will be highly overrepresented [1], a prediction that is supported by direct measurements on a number of real networks for which time-resolved network topologies are available. This finding questions our ability to characterize networks based on the subgraph abundance obtained from a single topological snapshot. We show that a combined understanding of network evolution and subgraph abundance offers a more complete picture.

Subgraphs. We consider subgraphs with n vertices and $n-1+t$ edges, whose central vertex has links to $n-1$ neighbors, which in turn have t links among themselves [Fig. 1(a)]. The total number of n-node subgraphs that can pass by a node with degree k is $\binom{k}{n-1}$. Each of these n-node subgraphs can have at most $n_p=(n-1)(n-2)/2$ edges between the $n-1$ neighbors of the central node. The probability that there is an edge between two neighbors of a degree k vertex is given by the clustering coefficient $C(k)$. Therefore, the probability to obtain t connected pairs and n_p-t disconnected pairs is given by the binomial distribution of n_p trials with probability $C(k)$. The expected number of (n,t) subgraphs in the network is obtained after averaging over the degree distribution, resulting in

$$N_{nt} = g_{nt} N \sum_{k=1}^{k_{\max}} P(k) \binom{k}{n-1} \binom{n_p}{t} C(k)^t [1 - C(k)]^{n_p - t}, \quad (1)$$

where $k_{\max}$ is the maximum degree and the geometric factor g_{nt} takes into account that the same subgraph can have more

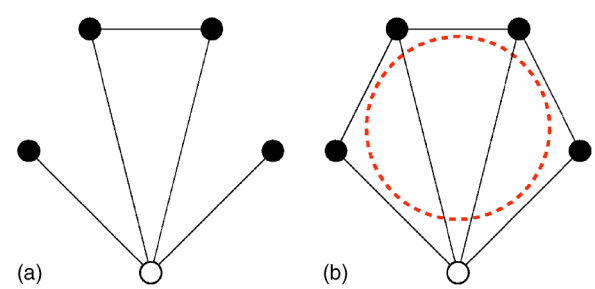

FIG. 1. Examples of subgraphs and cycles with a central vertex. The subgraph shown in (a) has $n=5$ vertices and $n-1+t=5$ edges, where $t=1$ represents the number of edges connecting the neighbors of the central vertex (empty circle) together. In (b) we show a subgraph with $t=3$ edges among the neighbors, such that the central vertex and its neighbors form a cycle of length $h=5$, highlighted by the dotted circle.

 ©2005 The American Physical Society

TABLE I. Characteristic exponents of the investigated real networks and the deterministic model. The exponents are defined through the scaling of the degree distribution $P(k)\sim k^{-\gamma}$, the clustering coefficient $C(k)=C_0k^{-\alpha}$, with $C_0\sim N^{\theta}$, the largest degree $k_{max}\sim N^{\delta}$, and the number of h cycles $N_h/N\sim N^{\zeta_h}$.

Network	γ	α	δ	θ	ζ_3	ζ_5	ζ_5
Coauthorship	2.4	0.0	0.6	0.00	0.6	1.6	2.6
Internet	2.2	0.75	1.0	0.20	0.3	0.7	1.2
Language	2.7	1.0	0.40	0.68	0.7	1.4	2.0
Model	2.6	1	0.63	0	0	0	0

than one central vertex. For instance, a triangle will be counted three times since each vertex is connected to the others, therefore $g_{31}=1/3$. For networks where $P(k)\sim k^{-\gamma}$ and $C(k)\sim k^{-\alpha}$, where γ and α are the degree distribution and clustering hierarchy exponents, $k_{max}\ll 1$, Eq. (1) predicts the existence of two subgraph classes [7],

$$\frac{N_{nt}}{N}\sim\begin{cases} C_0^t k_{\max}^{n-\gamma-\alpha t}, & n-\gamma-\alpha t>0, \quad \text{type I},\\ C_0^t, & n-\gamma-\alpha t<0, \quad \text{type II}.\end{cases} \tag{2}$$

Therefore, for the type I subgraphs the N_{nt}/N density increases with increasing network size, and N_{nt}/N is independent of N for type II subgraphs. In the following we provide direct evidence for the two subgraph types in several real networks for which varying network sizes are available: coauthorship network of mathematical publications [13], the autonomous system representation of the Internet [14,15], and the semantic web of English synonyms [16]. In each of these networks the maximum degree increases as $k_{max}\sim N^{\delta}$. We estimated δ from the scaling of the degree distribution moments with the graph size, $\langle k^n\rangle\sim N^{\delta(n+1-\gamma)}$, with $n=2$, 3, 4. Furthermore, we find that C_0 from $C(k)=C_0k^{-\alpha}$ also depends on the network size as $C_0\sim N^{\theta}$, where θ can be estimated using $C_0=\Sigma_{k\geq 2}C(k)/\Sigma_{k\geq 2}k^{-\alpha}$, giving a better estimate than a direct fit of $C(k)$. The exponents characterizing each network are summarized in Table I.

In Fig. 2 we show the density of all five vertex subgraphs ($n=5$) as a function of t. For the Internet and language networks C_0 increases with N, therefore the subgraph's density increases with the network size for all subgraphs. This consequence of the nonstationarity of the clustering coefficient is subtracted by normalizing N_{nt} by C_0^t. For the coauthorship graph with $\alpha=0$ (Table I), only type I subgraphs are observed, as predicted by (2). In contrast, for the Internet and semantic networks $\alpha>0$, therefore the overrepresented type I phase is expected to end approximately at the phase boundary predicted by (2). Indeed, to the left of the arrow denoting the $n-\gamma-\alpha t$ phase boundary we continue to observe a systematic increase in N_{5t}/NC_0^t, as expected for type I subgraphs. In contrast, beyond the phase boundary the subgraph densities obtained for different network sizes are independent of N, collapsing into a single curve.

We compared our predictions with direct counts in a growing deterministic network model [17] as well, characterized by a degree exponent $\gamma=1+\ln 3/\ln 2\approx 2.6$ and a

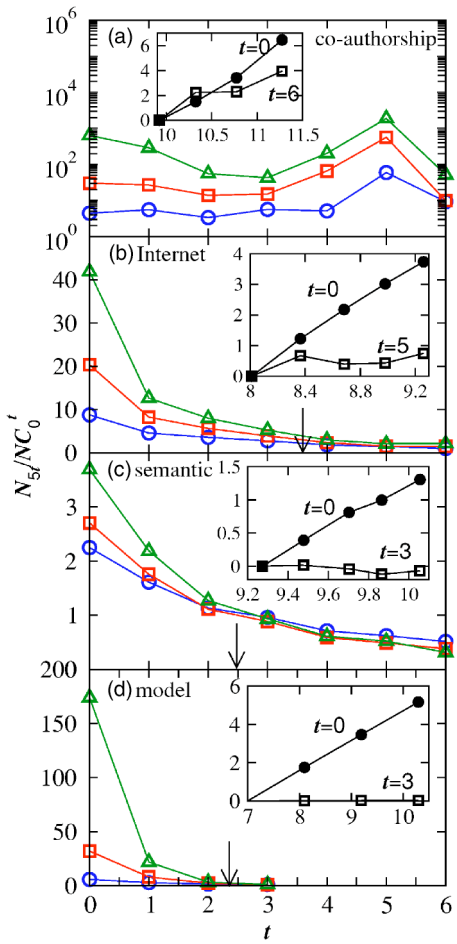

FIG. 2. Number of $(n=5,t)$ subgraphs for the (a) coauthorship, (b) Internet, and (c) semantic networks, and the (d) deterministic model as a function of t. Different symbols correspond to different snapshots of the networks evolution, from early stage (circles) to intermediate (squares) and current (i.e., largest) (triangles). N_{nt} depends strongly on t (spanning several orders of magnitude) making it difficult to observe the N dependence. Thus we normalized all the quantities (N_{5t}, C_0, and N) to the first year available. The arrows correspond to the phase boundary $5-\gamma-\alpha t=0$, with type I and II subgraphs to the left and right of the arrow, respectively. The insets show the system size dependence were we plot $\log N_{5t}$ vs $\log N$ for different values of t.

degree-dependent clustering coefficient $C(k)=C_0k^{-\alpha}$, with $C_0=2$ and $\alpha=1$. In Fig. 2(d) we show the number of $(n=5,t)$ subgraphs for different values of t and graph sizes. The arrow indicating the predicted phase transition point $n-\gamma-\alpha t=0$ clearly separates the type I from the type II subgraphs, a numerical finding that is supported by exact calculations as well. Note that only one type II $n=5$ subgraph is present in the deterministic network, due to its particular evolution rule.

Cycles. The formalism developed above can be generalized to predict cycle abundance as well. Consider the set of centrally connected cycles shown in Fig. 1(b). If the central vertex has degree k, we can form $\binom{k}{h-1}$ different groups of h vertices, $h-1$ selected from its k neighbors and the central vertex. Each ordering of the $h-1$ selected neighbors corre-

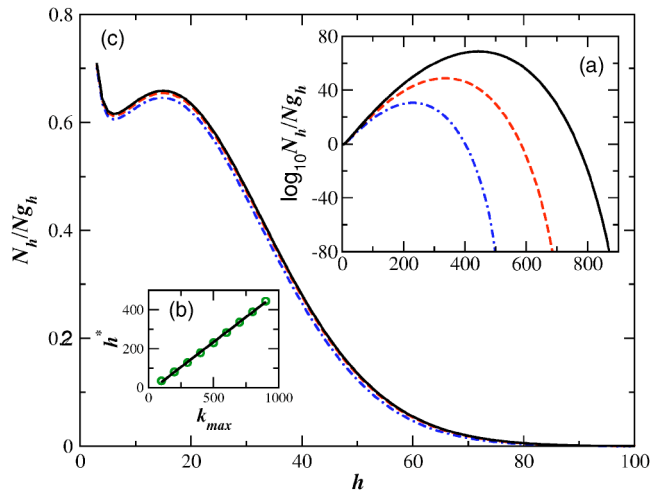

FIG. 3. Number of h cycles as computed from (3), using $\gamma=2.5$, (a) $C_0=2$ and $\alpha=1.1$, and $k_{max}=500$ (dashed-dotted), 700 (dashed), and 900 (solid), (b) h value at which N_h has a maximum as a function of k_{max}, (c) $C_0=1$ and $\alpha=0.9$.

sponds to a different cycle, therefore we multiply with half of the number of their permutations $(h-1)!/2$ (assuming that 123 is the same as 321). Finally, to obtain the number of h cycles we multiply the result with the probability of having $h-2$ edges between consecutive neighbors, $C(k)^{h-2}$, and sum over the degree distribution $P(k)$, finding

$$\frac{N_h}{N} = g_h \sum_{k=h-1}^{k_{max}} P(k) \frac{(h-1)!}{2} \binom{k}{h-1} C(k)^{h-2}, \tag{3}$$

where g_h is again a geometric factor correcting multiple counting of the same cycle. Note that (3) represents a lower bound for the total number of h cycles, which also includes cycles without a central vertex. Depending on the values of h, γ, and α the sum in (3) may converge or diverge in the limit $k_{max}\to\infty$. When it converges, the density of h cycles is independent of N (type II), otherwise it grows with N (type I). Since in preferential attachment models without clustering the density of h cycles decreases with increasing N [18], we conclude that clustering is the essential feature that gives rise to the observed high h-cycle number in such real networks such as the Internet [8]. To further characterize the cycle spectrum, we need to distinguish two different cases, $0<\alpha<1$ and $\alpha\geqslant 1$.

$0<\alpha<1$. In the $k_{max}\to\infty$ limit the cycle density follows

$$\frac{N_h}{N} \sim \begin{cases} C_0^{h-2}, & h<h_c, \\ C_0^{h-2} k_{max}^{(1-\alpha)(h-h_c)}, & h>h_c, \end{cases} \tag{4}$$

where $h_c=(\gamma-2\alpha)/(1-\alpha)$. Therefore, large cycles $(h>h_c)$ are abundant, their density growing with the network size N. As $\alpha\to 1$ the threshold $h_c\to\infty$, therefore the range of h for which the density is size independent, expands significantly.

Direct calculations using (3) show that N_h exhibits a maximum at some intermediate value of h [see Fig. 3(a)], already reported for the deterministic model [10]. The maximum represents a finite-size effect, as the characteristic cycle length h^*, corresponding to the maximum of N_h, scales as $h^*\sim k_{max}$ [Fig. 3(b)]. Yet, next we show that this behavior is not generic, but depends on the value of α.

$\alpha\geqslant 1$. For all $\gamma>2$ only type II subgraphs are expected $(N_h/N\sim C_0^{h-2})$, as suggested by the divergence of h_c in the $\alpha\to 1$ limit. If $C_0>1$ the number of h cycles continues to exhibit a maximum and the characteristic cycle length h^* scales as $h^*\sim k_{max}$. If $C_0<1$, however, the number of h

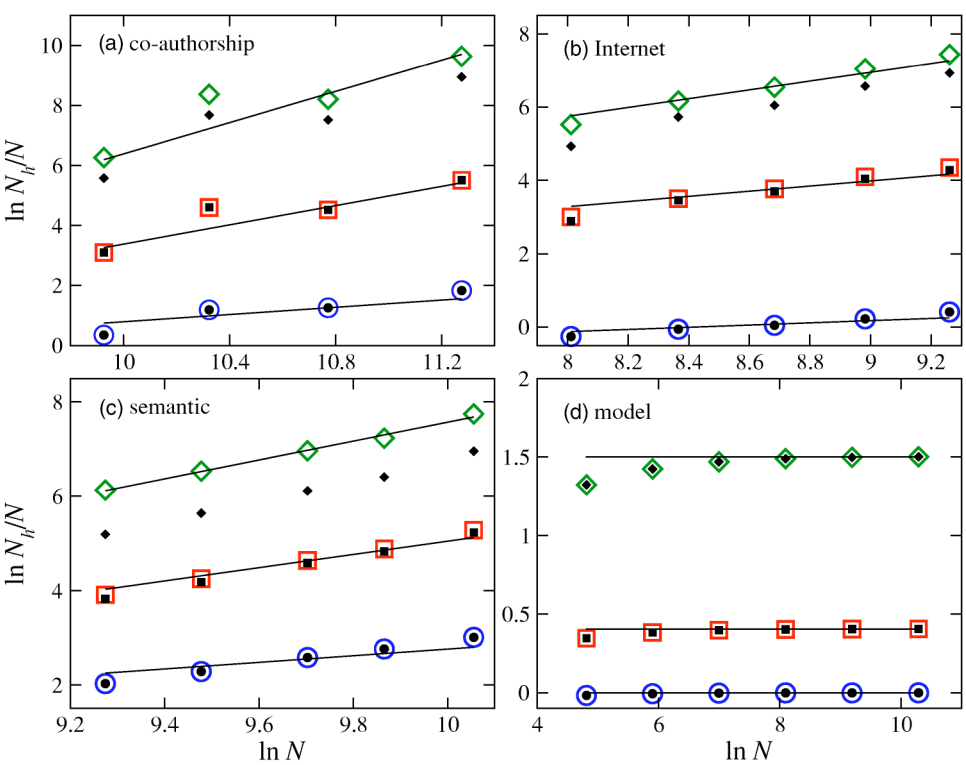

FIG. 4. Density of all (open symbols) and centrally connected (filled symbols) cycles with $h=3$ (circles), 4 (squares), and 5 (diamond) cycles as a function of the graph size. The continuous lines corresponds with our predictions (Table I).

cycles decrease with h, although a small local minima is seen for small cycles. More important, in this case N_h/N is independent of the network size [see Fig. 3(c)], in contrast with the size dependence observed earlier [Fig. 3(a) and [10]]. Thus, for networks with $\alpha>1$ or $\alpha=1$ and $C_0<1$ the cycle spectrum is stationary, independent of the stage of the growth process in which we inspect the network.

Our predictions for the cycle abundance are based on centrally connected cycles, in which a central vertex is connected to all vertices of the cycle [Fig. 1(b)]. In the following we show that our predictions capture the scaling of all h cycles as well, and not only those that are centrally connected. For this, in Fig. 4 we plot the number of $h=3$, 4, 5 cycles (i.e., all cycles as well as those that are centrally connected) as a function of the graph size for the studied real and model networks, together with our predictions (continuous line). First we note that in many cases ($h=3$ and 4) the full cycle density and the density of the centrally connected cycles overlap. In the few cases ($h=5$) where there are systematic differences between the two densities the N dependence of the two quantities is the same, indicating that our calculations correctly predict the scaling of all cycles.

For the coauthorship and Internet graphs $\alpha<1$ and $h_c<3$, therefore the $h=3$, 4, 5 cycles are predicted to be in the type I regime ($h>h_c$). In this case $N_h/N\sim N^{\zeta_h}$, where $\zeta_h=\theta(h-2)+\delta(1-\alpha)(h-h_c)$. For the language graph $\alpha=1$, therefore $\zeta_h=\theta(h-2)$. For the deterministic model a direct count of the h cycles reveals that they are of type II, i.e., their density is independent of N[10], in agreement with our predictions for $\alpha\geqslant 1$. These predictions are shown as continuous lines in Fig. 4, indicating a good agreement with the real measurements.

Our results offer evidence of a quite complex subgraph dynamics. As the network grows, the density of the type II subgraphs remains unchanged, being independent of the system size. In contrast, the density of the type I subgraphs increases in an inhomogeneous fashion. Indeed, each (n,t) subgraph has its own growth exponent ζ_{nt}, which means that their density increases in a differentiated manner: the density of some type I subgraphs will grow faster than the density of the other type I subgraphs. Thus, inspecting the system at several time intervals one expects significant shifts in subgraphs densities. As a group, with increasing network size the type I graphs will significantly outnumber the constant density type II graphs. Therefore, the inspection of the graph density at a given moment will offer us valuable, but limited, information about the overall local structure of a complex network. However, $P(k)$ and the $C(k)$ functions allow us to predict with high precision the future shifts in subgraph densities, indicating that a precise knowledge of the global network characteristics is needed to fully understand the local structure of the network at any moment. These results will eventually force us to reevaluate a number of concepts, ranging from the potential characterization of complex networks based on their subgraph spectrum to our understanding of the impact of subgraphs on processes taking place on complex networks [19,20].

ACKNOWLEDGMENTS

A.V. and A.-L.B. were supported by NSF Grant No. DMS 0441089. J.G.O. acknowledges financial support of FCT (Portugal), Grant No. SFRH/BD/14168/2003. We wish to thank J.-P. Eckmann and D. Sergi for useful discussions on the subject.

[1] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002); R. Milo, S. S. Shen-Orr, S. Itzkovitz, and U. Alon, Science **298**, 824 (2002).
[2] S. Wuchty, Z. Oltvai, and A.-L. Barabási, Nat. Genet. **35**, 118 (2003).
[3] R. E. Ulanowicz, Math. Biosci. **65**, 219 (1983).
[4] P. Bearman, Am. J. Sociol. **102**, 1383 (1997).
[5] M. Bernstein, ACM Comput. Surv. **31**, 19 (1999).
[6] E. Marinari and R. Monasson, e-print cond-mat/0407253.
[7] A. Vázquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. Oltvai, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **101**, 17940 (2004).
[8] G. Bianconi, G. Caldarelli, and A. Capocci, e-print cond-mat/0310339.
[9] G. Bianconi, G. Caldarelli, and A. Capocci, e-print cond-mat/0408349.
[10] H. D. Rozenfeld, J. E. Kirk, E. M. Bollt, and D. ben Avraham, e-print cond-mat/0403536.
[11] D. Sergi, e-print cond-mat/0412472.
[12] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2001).
[13] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A **311**, 590 (2002).
[14] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[15] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[16] S. H. Yook and H. Jeong (unpublished).
[17] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **65**, 066122 (2002).
[18] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
[19] Z. Toroczkai, B. Kozma, K. E. Bassler, N. Hengartner, and G. Korniss, Nature (London) **428**, 716 (2004).
[20] T. Petermann and P. de los Rios, Phys. Rev. E **69**, 066116 (2004).