PHYSICAL REVIEW E **72**, 046111 (2005)

Core-periphery organization of complex networks

Petter Holme
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA
(Received 3 June 2005; published 12 October 2005)

Networks may, or may not, be wired to have a core that is both itself densely connected and central in terms of graph distance. In this study we propose a coefficient to measure if the network has such a clear-cut core-periphery dichotomy. We measure this coefficient for a number of real-world and model networks and find that different classes of networks have their characteristic values. Among other things we conclude that geographically embedded transportation networks have a strong core-periphery structure. We proceed to study radial statistics of the core, i.e., properties of the n neighborhoods of the core vertices for increasing n. We find that almost all networks have unexpectedly many edges within n neighborhoods at a certain distance from the core suggesting an effective radius for nontrivial network processes.

DOI: 10.1103/PhysRevE.72.046111 PACS number(s): 89.75.Hc, 89.75.Fb

All systems consisting of pairwise-interacting entities can be modeled as networks. This makes the study of complex networks a very general and interdisciplinary area of statistical physics [1]. One of the most important gains of the recent wave of statistical network studies is the quantification of large-scale network topology. Now, with the use of just a few words and numbers, one can state the essential characteristics of a huge network. A possible large-scale design principle is that one part of the network constitutes a densely connected core that also is central in terms of network distance, and the rest of the network forms a periphery. In, for example, a network of airline connections you would most certainly pass such a core airport on any many-flight itinerary. It is known that a broad degree distribution can create a core having such properties [2]. In this paper we address the question if there is a tendency for such a structure in the actual wiring of the network. For example, if one assumes degree to be, to a large extent, an intrinsic property of the vertices, then is the network organized with a distinct core-periphery structure or not? To give a quantitative answer to this question our first step is to find a core with the above mentioned properties—being highly interconnected and having a high closeness centrality [3] (the inverse average distance between vertices of a subgraph and the rest of the vertices). Once such a subgraph is identified we calculate its closeness centrality relative to the graph as a whole, and subtract the corresponding quantity for the ensemble of random graphs with the same set of degrees as the original network (cf. Ref. [4]). If the resulting coefficient is positive the network shows a pronounced core-periphery structure. Once the core and periphery are distinguished one may proceed to investigate their function by looking at the statistical properties of the n neighborhoods (the set of vertices on distance n) of the core vertices. This paper starts by defining the core-periphery coefficient and measure it for real-world networks of numerous types, then proceeds by discussing and measuring radial statistics.

We assume the network to be represented as a graph $G=(V,E)$ with a set V of N vertices and a set E of M undirected and unweighted edges. Since our analysis requires the network to be connected we will henceforth identify G with the largest connected component of the network. We also remove self-edges and multiple edges.

The notion of network centrality is a very broad and many measures have been proposed to capture different aspects of the concept [20]. One of the simplest quantities is the closeness centrality [3]

$$C_C(U) = (\langle\langle d(i,j)\rangle_{j\in V\setminus\{i\}}\rangle_{i\in U})^{-1} \tag{1}$$

of a subgraph U, where $d(i,j)$ is the graph distance between i and j. So we require a core to be a subgraph U with high $C_C(U)$, but also to be a well-defined cluster. Now, if there are many facets of the centrality concept, there are even more algorithms to identify graph clusters, each being a *de facto* cluster definition [1]. For simplicity we choose the most rudimentary cluster definition—the set of k cores. A k core is a maximal subgraph with the minimum degree k. To calculate a sequence of k cores is computationally cheaper (linear in M [21]) than more elaborate clustering algorithms. So we let our core $V_{\text{core}}(G)$ be the k core with maximal closeness and define the core-periphery coefficient c_{cp} as

$$c_{\text{cp}}(G) = \frac{C_C[V_{\text{core}}(G)]}{C_C[V(G)]} - \left\langle \frac{C_C[V_{\text{core}}(G')]}{C_C[V(G')]} \right\rangle_{G'\in\mathcal{G}(G)}, \tag{2}$$

where $\mathcal{G}(G)$ is the ensemble of graphs with the same set of degrees as G. The sequence of k cores is not necessarily unique. We maximize $C_C(U)$ over m_{seq} different sequences. The m_{null} elements of $\mathcal{G}(G)$ can be obtained by randomization of G in time and space of the order of M [22]. We use $m_{\text{null}}=1000$ and $m_{\text{seq}}=10$ for networks with $N<5000$, and $m_{\text{null}}=50$ and $m_{\text{seq}}=3$ for $N\geqslant 5000$.

The correlation of degrees at either side of an edge is an often studied quantity [4,23]. If there is a tendency for well-connected vertices to connect to each other then there will be highly interconnected clusters in the graph—one of the requirements for a core-periphery structure in our sense. This resemblance to the core-periphery structure makes the degree correlations an interesting reference structure. A common way to quantify the average degree-degree correlations is to measure the *assortative mixing coefficient* [23],

TABLE I. The network sizes N and M, the core-periphery coefficient c_{cp}, and the relative assortative mixing coefficient Δr for a number of networks. In the interstate network the vertices are American interstate highway junctions and two junctions are connected if there is a road with no junction in between. The pipeline network is a similar network of junctions and gas pipes. In the airport data (obtained from IATA www.iata.org) the vertices are airports and the edges represent airport pairs with a nonstop flight connection. The Internet figures are averages of 15 AS-level graphs constructed from traceroute searches. The arXiv, board of directors, and Ajou students are constructed one-mode projections from affiliation networks (where links goes from persons to e-prints, corporate boards, and university classes, respectively). The student network is averaged over graphs for 16 semesters. In the electronic communication networks one edge represent that at least one of the vertices has contacted the other over some electronic medium. In the nd.edu data the vertices are HTML documents and the edges are hyperlink. The citation graph is constructed from preprints in the field of high-energy physics [5]. The food webs are networks of water-living species and an edge means that one species prey on the other. For the protein networks an edge means that two proteins bind to each other physically. The metabolic and "whole cellular" networks consist of chemical substances and edges indicating that one molecule occur in the same reaction as the other (the values for these networks are averages over 43 organisms from different domains of life).

Network		Ref.	N	M	c_{cp}	Δr
Geographical networks	interstate highways		935	1315	0.231(1)	0.0851(5)
	pipelines	[6]	2999	3079	0.180(2)	0.073(2)
	streets, Stockholm	[7]	3325	5100	0.255(1)	0.080(1)
	Airport		449	2795	0.0523(3)	0.0910(3)
	Internet	[8]	1968(66)	4051(121)	0.045(2)	0.009(3)
One-mode projections of affiliation networks	arXiv	[9]	48561	287570	−0.08(3)	0.361(3)
	board of directors	[10]	6193	43074	−0.037(2)	0.280(2)
	Ajou University students	[11]	7285(128)	75898(6566)	−0.08(1)	0.66(4)
Electronic communication	email, Ebel *et al.*	[12]	39592	57703	−0.229(4)	−0.001(4)
	Internet community, pussokram.com	[13]	28295	115335	−0.183(5)	−0.005(5)
Reference networks	WWW, nd.edu	[14]	325729	1090108	−0.027(3)	−0.003(3)
	HEP citations		27400	352021	−0.10(1)	0.03(1)
Food webs	Little Rock Lake	[15]	92	960	0.005(6)	−0.0141(6)
	Ythan Estuary	[16]	134	593	−0.020(1)	−0.0153(9)
Biochemical networks	*Drosophila* protein	[17]	2915	4121	−0.035(2)	0.003(1)
	S. cervisiae protein	[18]	3898	7283	−0.249(1)	−0.069(1)
	metabolic networks	[19]	427(27)	1257(88)	−0.002(6)	0.006(1)

$$r = \frac{4\langle k_1 k_2\rangle - \langle k_1 + k_2\rangle^2}{2\langle k_1^2 + k_2^2\rangle - \langle k_1 + k_2\rangle^2}, \quad (3)$$

where k_i is the degree of the ith argument of a edge as it appear in a list of E. Now, our null model is a random graph conditioned to have the same degree sequence as the original graph. So just as for c_{cp}, we consider the deviation from our null model and measure

$$\Delta r(G) = r(G) - \langle r(G')\rangle_{G' \in \mathcal{G}(G)}. \quad (4)$$

In Table I c_{cp} and Δr are displayed for a number of real-world networks. We find that the core-periphery structure and relative degree-degree correlations follow the different classes of networks rather closely. Furthermore the core-periphery structure and degree-degree correlations seem to be quite independent network structures in practice. For example, geographically embedded networks have a clear core-periphery structure and weakly positive degree-degree correlations, whereas social networks derived from affiliations have slightly negative c_{cp} values but very high Δr values. Most geographically embedded networks have the function of transporting, or transmitting, something between the vertices. Networks with a well-defined core (which most paths pass through) and a periphery (covering most of the area) are known to have good performance with respect to communication times [6]. Also networks of airline traffic [24] and the hardwired Internet [8] are known to have well-defined cores due to traffic-flow optimization. The class of one-mode projection networks (social networks constructed by linking people that participate in something—movies, scientific research, etc.—together) show slightly negative c_{cp} values. This can be explained by that there is a grouping of the people on the basis of specialization (and, in student networks, also in grade) and thus no well-defined core. The vertices in electronic communication networks are also people but the network structures are quite different; the degree-degree correlation is typically slightly negative, as is the core-periphery coefficient. Information networks where the edges refer to supporting information sources can be expected to be grouped into topics, thus the negative c_{cp}. Food webs are other stratified networks where a lack of a well-defined core seems natural. The biochemical networks all show negative c_{cp} values.

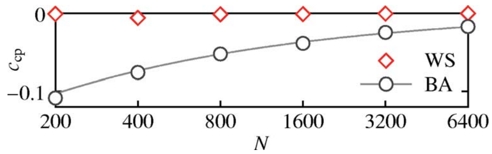

FIG. 1. (Color online) Core-periphery structure of model networks. All networks have $M=2N$. The core-periphery model has the parameter $f_{core}=0.96$ (i.e., the intended core consists of 4% of the vertices) and $\gamma=3$. All values are averaged over 10^4–10^5 network realizations. The BA-model line is a fit to a power-law form $\alpha_0+\alpha_1 N^{-\alpha_2}$ [this fit gives $c_{cp}(\infty)=\alpha_0=0.004(9)$].

In addition to the real-world networks of Table I we also measure the core-periphery coefficient for a few network models. For simple random graphs [20] where N vertices are randomly connected by M edges, defining an ensemble $\mathbb{G}(N,M)$ of graphs, $\mathcal{G}(G)$ is precisely the elements of $\mathbb{G}(N,M)$ with the same degree sequence as G. So on average, c_{cp} will be zero for random graphs. A popular network model is the Barabási-Albert (BA) model [25] where the graphs are grown by iteratively adding new vertices with edges to old vertices with a probability proportional to the degree of the old vertices. In Fig. 1 we see that c_{cp} tends to zero (or a value very close to zero) for BA model networks. The BA model has an assortative mixing coefficient r that tends to zero as N grows [23]. From this one sees that the high-degree vertices are not more interconnected than can be expected from their degrees. We also investigate the Watts-Strogatz' small-world network model [26] were one end of the edges of a circulant [20] is rewired with a certain probability (0.01 in our case). Just as for the BA model c_{cp} converges to zero (see Fig. 1). This is not so surprising, in the WS model's starting point, the circulant, every vertex is in the same position. The rewiring procedure does not aggregate vertices to a well-defined core either.

A well-defined core is a useful starting point for a radial examination of the network. By plotting quantities averaged over the n neighborhoods of the core vertices as functions of n one can get an idea of the respective purposes of the core and periphery. This kind of statistics is naturally more sensible the larger c_{cp} is, but even for slightly negative c_{cp} values it may be informative. To get a rough view of the radial network organization we plot the average degree of the vertices in the n neighborhood of core vertices as a function of n in Fig. 2. We include the corresponding results for our null model. The core vertices themselves almost always get higher average degree for the null model than the real-world networks (5–10 % higher for the networks of Fig. 2). For the first neighborhood the situation is reversed—the real-world networks have higher $\langle k\rangle$ than the null model. Then the degrees are decreasing monotonically; typically faster for the null model networks. One can imagine different functions of the peripheral vertices—either they are just conveying information, traffic, etc., to and from the core; or they are, just as the core vertices, involved in the general network processes, only less intensely. To understand this we measure the average value of the quantity,

$$\mu(i,n)=M(K_n(i))/\mathbb{E}M(K_n(i)) \tag{5}$$

over the core vertices; $M(K_n(i))$ is the number of edges within i's n neighborhood $K_n(i)$ and $\mathbb{E}M(K_n(i))$ is the expected number of edges in a set of vertices of the same degrees as $K_n(i)$ in a random graph of the same degree sequence as the original graph G. To calculate $\mathbb{E}M$ we rely on the same random sampling as for the c_{cp} calculation. To save time one can calculate $\mathbb{E}M(K)$ as the average number of edges within the original subgraph K at the same time as the $\mathcal{G}(G)$ sampling of the c_{cp} calculation. In Figs. 2(d)–2(f) we diagram $\langle\mu\rangle(n)$ for our three example networks. Since the core is constructed to be highly interconnected it is no surprise that $\langle\mu\rangle$ has a peak for small n. For the metabolic network of Fig. 2(f) this peak is small. This is due to the exceptionally high degrees ~55 of the core vertices (including substrates such as H_2O and adenosine triphosphate)—even in the null model networks this set of vertices will, for

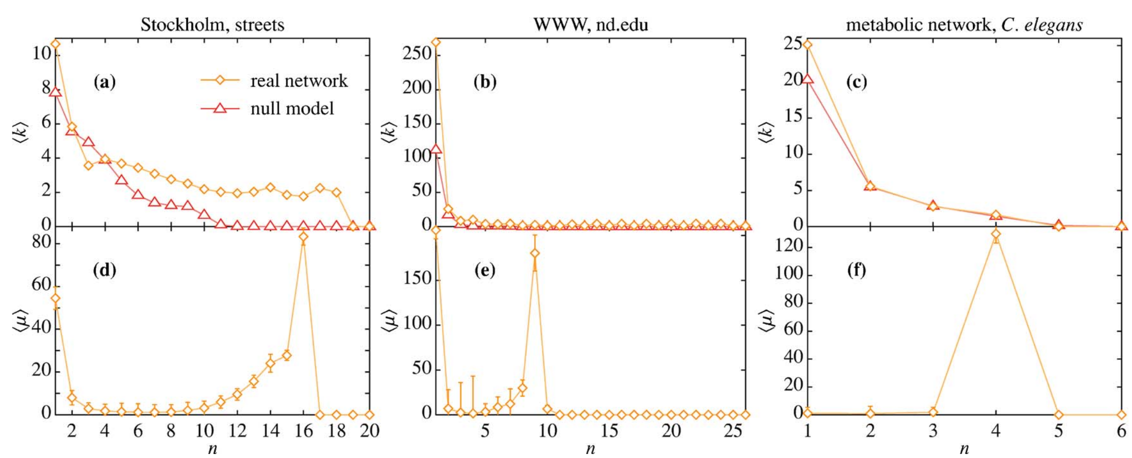

FIG. 2. (Color online) Radial statistics for three real-world networks. (a)–(c) show the average degree $\langle k\rangle$ of the n neighborhoods of the core vertices as a function of n for three real world networks: a network of streets in Stockholm, Sweden [7], a network of hyperlinked web pages [14], and the metabolic network of *C. elegans* [19]. Curves for our null model are included. In (d)–(f) we plot μ—the number of edges with in the n neighborhood relative to the expected number of edges given the degree sequence of the n neighborhood and the graph as a whole. Lines are guides for the eyes.

combinatorial reasons, be highly interconnected. For intermediate n the $\langle \mu \rangle$ values are of the order of unity. But as n increases, $\langle \mu \rangle$ grows to a sharp peak before it eventually drops to zero. This seems like a rather ubiquitous feature (at least it is present in almost all networks of Table I). We interpret this as that the periphery has both the two functions listed above: To a certain distance from the core (defined by the peak) vertices have similar function and are for this reason connected; beyond this distance the network consists only of cycle-free branches. This dichotomy—the network inside and outside of the peak radius—is yet more distinct than the core vs periphery as defined above. On the other hand, the outside is functionally rather trivial and (in all cases we study) smaller than the inside.

Many networks have subgraphs with very different characteristics and function. Perhaps the simplest division of a network is that into a core and a periphery. The core concept has been used in various senses in the past; typically it is defined as a subgraph which is most tightly connected [27] or a most central [2]. Here we use the rather strong precepts that a core should be both highly interconnected and central. We propose a coefficient c_{cp} to quantify this idea—a structural measure to complement quantities such as the clustering and assortative mixing coefficients. Different types of networks have their characteristic c_{cp} values: Geographically embedded networks typically have positive c_{cp} (a possible effect of their communication-time optimization). Social network, on the other hand, typically have slightly negative c_{cp} values despite their positive degree-degree correlations. We show that c_{cp} for model networks such as the Erdös-Rényi, Barabási-Albert, and Watts-Strogatz models goes to zero (or at least to a very small value) as the network size increases. (Preliminary studies show that it is possible to construct networks with a positive c_{cp} in the large system limit.) Once the core of a network is found one can construct a radial image of the network by plotting quantities averaged over the n neighborhoods of the core vertices as a function of n. One such quantity we study is $\mu(n,i)$—the relative number of edges within the n neighborhood of i to the expected number of edges in a subgraph of the same set of degrees in the null model. $\langle \mu \rangle$ shows, almost ubiquitously, a peak at intermediate n. We interpret this peak as an effective radius of the network. Much remains to be done in terms of characterizing the cores and peripheries of complex networks. We believe this dichotomy and the radial imagery we present are very useful tools to understand the large-scale architecture of such networks.

We thank M. Newman for valuable discussions, and H. Ebel, J.-P. Eckman, M. Gastner, J. Goodwin, A. Grönlund, M. Lokner, B. J. Kim, M. Newman, S. M. Park, S. Praszalowicz, M. Rosvall, and C. Wollter for help with data acquisition. Support from the Wenner-Gren Foundation is gratefully acknowledged.

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[2] F. Chung and L. Lu, Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002).
[3] G. Sabidussi, Psychometrika **31**, 581 (1966).
[4] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[5] http://www.cs.cornell.edu/projects/kddcup/datasets.html.
[6] M. T. Gastner and M. E. J. Newman, cond-mat/0409702 (unpublished).
[7] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[8] R. Pastor-Santorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge Univeristy Press, Cambridge, 2004).
[9] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[10] G. F. Davis, M. Yoo, and W. E. Baker, Strat. Org. **1**, 301 (2003).
[11] S. M. Park, P. Holme, and B. J. Kim, Sae Mulli **49**, 399 (2004).
[12] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103 (2002).
[13] P. Holme, C. R. Edling, and F. Liljeros, Soc. Networks **26**, 155 (2004).
[14] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[15] N. D. Martinez, Ecol. Monogr. **61**, 367 (1991).
[16] S. J. Hall and D. Raffaelli, J. Anim. Ecol. **60**, 823 (1991).
[17] L. Giot *et al.*, Science **302**, 1727 (2003).
[18] P. Pagel *et al.*, Bioinformatics **21**, 832 (2004).
[19] H. Jeong, B. Tombor, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[20] F. Buckley and F. Harary, *Distance in Graphs* (Addison-Wesley, Redwood City, 1989).
[21] D. Fernholz and V. Ramachandran, Tech. Rep. TR04-13, University of Texas at Austin (2004).
[22] J. M. Roberts Jr., Soc. Networks **22**, 273 (2000).
[23] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[24] R. Guimerà and L. A. N. Amaral, Eur. Phys. J. B **38**, 381 (2004).
[25] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[26] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[27] M. G. Everett and S. P. Borgatti, Soc. Networks **21**, 397 (1999).