Detecting degree symmetries in networks

Petter Holme
Department of Computer Science, University of New Mexico, Albuquerque, New Mexico 87131, USA
(Received 2 May 2006; published 14 September 2006)

The surrounding of a vertex in a network can be more or less symmetric. We derive measures of a specific kind of symmetry of a vertex which we call *degree symmetry*—the property that many paths going out from a vertex have overlapping degree sequences. These measures are evaluated on artificial and real networks. Specifically we consider vertices in the human metabolic network. We also measure the average degree-symmetry coefficient for different classes of real-world network. We find that most studied examples are weakly positively degree symmetric. The exceptions are an airport network (having a negative degree-symmetry coefficient) and one-mode projections of social affiliation networks that are rather strongly degree symmetric.

DOI: 10.1103/PhysRevE.74.036107 PACS number(s): 89.75.Fb, 89.75.Hc

I. INTRODUCTION

With the advent of modern database technology numerous large scale network datasets have been made available. This development has triggered a surge of activity in studies of statistical network properties [1–3]. The underlying idea of these studies is that the network structure (the way the networks differ from completely random networks) contain some information of the function, both locally and globally, of the network. Hence a common theme in these works has been the development of structural measures to characterize network structure. In this paper we propose and evaluate a measure of a previously unstudied network structure—a special case of symmetry we call *degree symmetry*. In geometry an object is symmetrical if it is invariant to rotations, reflections, and so on. In networks, with no given geometrical embedding, these concepts must be relaxed. Furthermore, we would like to have a continuous measure saying not only if a vertex is a local center of symmetry or not, but also how symmetric the vertex is. The aspect of symmetry we address is, roughly speaking, that if you look at the object (network in our case) in different ways from a symmetric vertex it still looks the same. The process of "looking" will in our case be walking along paths (non-self-intersecting sequences of edges). Furthermore, since degree (number of neighbors) is commonly regarded as the most fundamental quantity relating a vertex to its function, we say two vertices "look the same" if they have the same degree. We will thus derive our measure by performing walks along all paths from a vertex and compare the sequence of degrees of the vertices along these paths. The situation we have in mind is depicted in Fig. 1—all paths from the central vertex have degree sequences starting with (3,2,…), thus the central vertex is highly degree symmetric.

The rest of the paper is organized as follows: First we give a detailed derivation of the degree-symmetry coefficient (in two different versions, appropriate for different needs). Then we evaluate these on example networks and a biochemical network. Finally we discuss the average degree symmetry of different classes of real-world networks.

II. DERIVATION OF THE MEASURE

We will consider the network represented by a graph $G=(V,E)$ of N vertices, V, and M edges, E. For a vertex i to have high degree symmetry it has, as mentioned, to have many paths with the same sequence of degrees. We will use a cutoff l for the pathlength and consider only paths of that length. The reason for this cutoff is threefold: First, in all (with possibly some curious exception) network processes, a vertex is more affected by its closest surroundings then vertices further away. Thus one would like to have a lower weight on the contribution from distant vertices. Second, the number of vertices n steps away grows fast with the distance from i. For finite networks this means that the paths soon reach the periphery of the network where unwanted finite-size effects set in. Third, for computation speed, one benefit from a cutoff.

Assume there are p paths of length l from a vertex i. We then denote the degree sequences of these paths

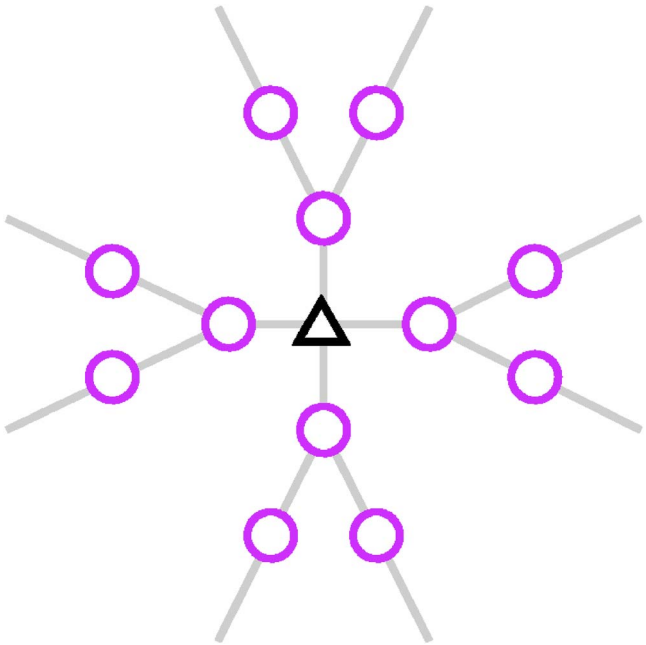


FIG. 1. (Color online) Illustrations of degree symmetry. Consider paths of length-2 (i.e., $l=2$). All paths out from the central (black) vertex have the degree sequence (3,2) meaning the central vertex has high degree symmetry.

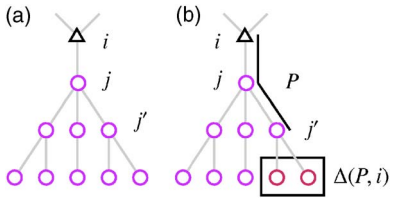


FIG. 2. (Color online) Illustrations of concepts in the derivation of the degree symmetry coefficient. (a) illustrates the branching number. Consider paths of length-3 out from i. The branching number of the path (i,j) is five (there are five paths from i of length-3 that goes through j). The branching number at j' is two. (b) shows the set $\Delta(P,i)$, where P is the path (i,j,j').

$$Q_l(i) = \{[k(v^1_{1,i,l}), \cdots, k(v^l_{1,i,l})] \ ,$$

$$\vdots \ ,$$

$$[k(v^1_{p,i,l}), \cdots, k(v^l_{p,i,l})]\}, \tag{1}$$

where $k(v)$ denotes the degree of a vertex v and $v^j_{m,i,l}$ is the jth vertex of along the mth path of length l leading out from i. Then if there are unexpectedly many vertices at the same (j) index in the sequence with the same degree, the vertex i is a local center of degree symmetry. A rough symmetry measure would thus be to count the fraction of index pairs with the same degree, i.e.,

$$\frac{\tilde{s}_l(i)}{\Lambda} = \sum_{0 \leqslant n < n' \leqslant p} \sum_{j=1}^{l} \delta[k(v^j_{n,i,l}), k(v^j_{n',i,l})], \tag{2}$$

where

$$\Lambda = (l-1)\binom{p}{2} \quad \text{and} \quad \delta(x,y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases} \tag{3}$$

This measure is very crude and lacks many desired statistical features. For example, all paths that go via a particular neighbor of i will give a contribution to the sum. In practice this means that vertices with a high degree vertex rather far from itself (but closer that l) will trivially have a high $\tilde{s}_l(i)/\Lambda$. A first step would thus be to omit the contribution of vertices occurring in many sequences of $Q_l(i)$ at a specific index. That is, for all $l' \in (0,l)$ one wants to exclude the terms

$$\sum_{n,n'} \delta[k(v^1_{n,i,l}), k(v^1_{n',i,l})], \tag{4}$$

where n and n' are indices of paths that are identical to the first l' steps, from Eq. (2). Let $S_l(i)$ denote the number of such terms.

To calculate $S_l(i)$ consider a path $P=(i,\ldots,j)$ of length $l' < l$. Let $b_l(P,i)$ be the number of paths from i of length l that start with the path P. We call $b_l(P,i)$ the *branching number* of P, see Fig. 2(a). All pairs of paths starting with P will contribute to $\tilde{s}_l(i)$ a distance l' from i (since they all pass through j). Let $\Delta(P,i)$ be the set of neighbors to j that is not on the path P from i to j, see Fig. 2(b). [The number of elements in $\Delta(P,i)$ is thus $k_j - 1$.] This situation gives a contribution

$$S_l(P,i) = \binom{b_l(P,i)}{2} + \sum_{j \in \Delta(P,i)} S_l[(P,j),i] \tag{5}$$

from vertices of indices in the interval $[l',l]$ of $Q_l(i)$ to $\tilde{s}_l(i)$, where (P,j') denotes the path $(i,\ldots,j,j')$.

To further improve the measure one would like to, assuming some null model, subtract the expected random contribution to $\tilde{s}_l(i)/\Lambda$. If this can be achieved one would have a symmetry coefficient $s_l(i)$ that is zero when the symmetry is what can be expected from the null model, larger if i is a center of unexpectedly high symmetry, and less than zero if i is degree antisymmetric. A final symmetry coefficient could thus be written

$$s_l(i) = \frac{\tilde{s}_l(i) - S_l(i)}{\Lambda - S_l(i)} - \nu, \quad \text{provided } \Lambda > S_l(i), \tag{6}$$

where ν is the expected value of $[\tilde{s}_l(i) - S_l(i)]/[\Lambda - S_l(i)]$ in a null model. $\Lambda = S_l(i)$ can only happen if there is one or no path of length l. In both these cases the degree symmetry concept makes no sense so, if $\Lambda = S_l(i) \in \{0,1\}$, we set $s_l(i) = 0$. The null model we assume is random constrained on the degree distribution of the network. That is, given the fraction p_k of k-degree vertices the network is as random as possible. As it turns out ν is tricky to calculate analytically. There are two ways to proceed—either one calculates an approximative ν or one obtains ν via averaging $[\tilde{s}_l(i) - S_l(i)]/[\Lambda - S_l(i)]$ over realizations of the null model. Except being more accurate, the latter approach has the advantage of giving an error estimate of $s_l(i)$—one can by specifying a p value define significantly symmetric, or antisymmetric, vertices. We will use both approaches: The approximative method for analyzing example networks and the numerical method for analyzing real-world data.

We obtain an approximative value of ν, ν^{app}, by assuming ν is approximately equal to the probability that a pair of vertices, reached by walking along paths, is the same. Note that, since there are k ways into a degree-k vertex, when following a path the probability to reach a degree-k vertex is

$$\frac{kp_k}{\sum_{k'} k' p_{k'}} = \frac{kp_k}{\langle k \rangle}. \tag{7}$$

Thus the probability ν^{app} that two vertices of the same degree is reached by following different paths is

$$\nu^{\text{app}} = \sum_k p_k \left(\frac{kp_k}{\langle k \rangle} \right)^2 = \frac{1}{\langle k \rangle^2} \sum_k k^2 p_k^3. \tag{8}$$

One reason this approach is not exact is that the number of terms in the expression for $\tilde{s}_l(i)$ increases with the degree of the j in $\Delta(P,i)$ of Eq. (5). There are other higher-order effects related to other correlations between the path structure and the degree of the vertices.

To summarize we have two measures of local vertex symmetry, one approximative

$$s_l^{\mathrm{app}}(i) = \frac{\tilde{s}_l(i) - S_l(i)}{\Lambda - S_l(i)} - \frac{1}{\langle k\rangle^2}\sum_k k^2 p_k^3, \tag{9}$$

and one based on Monte Carlo sampling

$$s_l^{\mathrm{MC}}(i) = \frac{\tilde{s}_l(i) - S_l(i)}{\Lambda - S_l(i)} - \left\langle \frac{\tilde{s}_l(i) - S_l(i)}{\Lambda - S_l(i)} \right\rangle. \tag{10}$$

The sampling is conveniently done by random rewiring the edges of the original network [4].

III. ALGORITHM

The heart of the algorithm, as suggested in the preceding section, is a depth-first search with depth l. When returning along the traced out paths the branching number can be calculated recursively through

$$b_l(P,i) = \begin{cases} 1 & \text{if } P \text{ has length } l, \\ \displaystyle\sum_{j' \in \Delta((P,j'),i)} b_l[(P,j'),i] & \text{otherwise.} \end{cases} . \tag{11}$$

$S_l(P_i)$ can be calculated simultaneously using Eq. (5). A slight complication is that the same vertex may appear in different branches of the depth first search while calculating b and $\tilde{s}$. For small cutoff values this is easy to handle: For $l=2$ it does not affect the calculation at all. For $l=3$ one would only have to keep different depths [of Eqs. (5) and (11)] separate. For the calculation of $\tilde{s}_l(i)$ the terms of $Q_l(i)$ must be stored. Since the number of paths p grows fast with l, this can be quite a constraint for a large l. Luckily it suffices to store a histogram $h(l',k)$ counting the number of vertices of degree k at position l' of the paths $Q_l(i)$. p (and thus Λ) can be calculated as the number of time the depth l of the depth first search is reached. The running time of the algorithm is $O(p)$. A mean field approximation for networks with few triangles gives $O(p) \approx O(\langle k\rangle^l)$.

IV. EXTENSIONS AND CONSIDERATIONS

The method outlined above can quite straightforwardly be extended to network with directed edges, distinct types of edges or (integer) edge weights.

Imagine a network with z different edge sets $E_1, \ldots, E_z$. Such networks frequently occur in cellular biochemistry—e.g., protein interaction networks where different types of protein interaction can be recorded [5], or gene regulation networks where the edges can be activating or inhibitory. One sensible way to extend the above procedure is to use the union of the edges as your graph but to say two pairs of vertices in $Q_l(i)$ are identical if their degrees with respect to all of the networks are the same. To formalize this $Q_l(i)$ would be generalized to

$$Q_l(i) = \{[\mathbf{k}(v^1_{1,i,l}), \cdots, \mathbf{k}(v^l_{1,i,l})] ,$$

$$\vdots ,$$

$$[\mathbf{k}(v^1_{p,i,l}), \cdots, \mathbf{k}(v^l_{p,i,l})]\}, \tag{12}$$

where $\mathbf{k}(v)$ is a vector with v's degrees with respect to the different edge types, and the δ function of Eq. (4) would be one if the arguments are equal at all their indices, and zero otherwise. The ν^{app} must be redefined too,

$$\nu^{\mathrm{app}} = \frac{1}{\langle k\rangle^2}\sum_{k',k''} k' p_{k'} k'' p_{k''} \prod_{i=1}^{z}\sum_{j=1}^{z} p_i(k_j|k') p_i(k_j|k''), \tag{13}$$

where $p_i(k|k')$ is the conditional probability that a vertex has degree k with respect to edge set E_i given that its degree in the union network is k'. The case of a directed network can be treated similarly—one consider paths following edges in both directions but a vertex pair gives a contribution to $\tilde{s}$ only if both the in and out degrees are the same.

The approach of Sec. III can straightforwardly be applied to networks where multiple edges are allowed. Since multiple edges can be used to model weighted graphs [6] the generalization to weighted graphs (at least where edge weights represent the probability of following an edge) is simple. The other aspect of multigraphs, self-edges, is trivially dealt with—by the requirement that paths should not intersect themselves a self-edge will never be followed and can thus be omitted already when the graph is constructed.

The overlap required for a vertex pair to be considered equal in the calculation of the symmetry coefficient is rather strict. Sometimes one would like to treat two paths as similar even if their degrees differs slightly. Particularly, this applies to broad degree distributions. The functional difference between degree-2 and degree-3 vertices may be significant; but whether a vertex has degree 1002 or 1003 probably does not matter. To achieve such a relaxation one can construct a integer sequence $K_1 < K_2 < \ldots$ and let

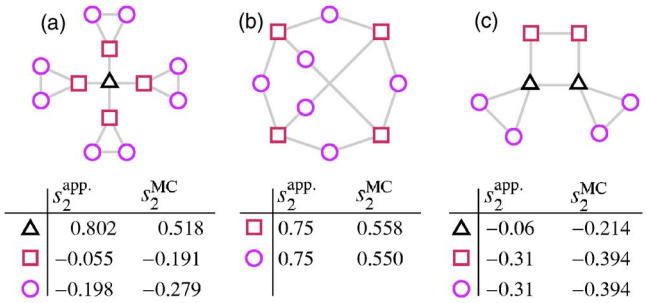
$$\delta(k,k') = \begin{cases} 1 & \text{if } K_i \leqslant k,k' < K_{i+1} \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases} . \tag{14}$$

That is one constructs a series of equivalence classes of vertices. For a power law, or similarly broad, degree distributions one can let $K_{i+1} - K_i$ increase exponentially with i. In this case one also must modify the definition of ν^{app},

$$\nu^{\mathrm{app}} = \frac{1}{\langle k\rangle^2}\sum_i \left(\sum_{K_i \leqslant k < K_{i+1}} p_k\right)\left(\sum_{K_i \leqslant k < K_{i+1}} k p_k\right)^2. \tag{15}$$

V. DEGREE SYMMETRIES OF EXAMPLE NETWORKS

In this section we evaluate the measure, for example, networks and real-world networks. We will use the smallest nontrivial cutoff $l=2$ throughout this section. Most conclusions hold for $l=3$ or 4.



	$s_2^{\text{app.}}$	s_2^{MC}
△	0.802	0.518
□	−0.055	−0.191
○	−0.198	−0.279

	$s_2^{\text{app.}}$	s_2^{MC}
□	0.75	0.558
○	0.75	0.550

	$s_2^{\text{app.}}$	s_2^{MC}
△	−0.06	−0.214
□	−0.31	−0.394
○	−0.31	−0.394

FIG. 3. (Color online) Degree symmetries of small example networks. (a) is consistent with the example Fig. 1(a). (b) is an example of a graph with only positive degree symmetries. (c) shows a graph with only negative degree symmetries. The cutoff length $l=2$ is used.

A. Small test graphs

To get a feeling for the s_l measure we start by considering a few small test networks, see Fig. 3. In Fig. 3(a) we display a network with the same degree symmetry, with respect to the central vertex (triangle), as Fig. 1. As expected the central vertex has a strong degree symmetry coefficient. To carry through the calculation of Eq. (9) once we obtain the degree distribution $p_2=8/13$, $p_3=4/13$, and $p_4=1/13$ giving $\nu^{\text{app}}=165/832\approx 0.198$. All length-2 paths out from the central vertex have the degree sequence (3,2) so $\tilde{s}_2(\Delta)=4$, $S_2(\Delta)=4$, and $\Lambda=28$ giving $s_2^{\text{app}}(\Delta)=667/832\approx 0.802$. The degree-3 vertices (squares) have two degree sequences of their outgoing paths (4,3) and (2,2), whereas paths from degree-2 vertices (triangles) have degree sequences (3,4) and (2,3). This difference is larger than expected from the null model (random networks with eight degree-2 vertices, four degree-3 vertices, and one degree-4 vertex), thus the negative s_2 values for these vertices.

In Fig. 3(b) we show a graph where all vertices have positive degree-symmetry coefficient. Paths from degree-2 vertices have only the degree sequence (3,2) and paths from degree-3 vertices have only the degree sequence (2,3). Thus, for every vertex, the view of degrees along the path out to the rest of the network is the same no matter which direction one looks in from that vertex. A radically different view is seen in Fig. 3(c). In this case the vertices have three distinct positions in the network. The vertices marked with squares have degree-2 and -4 outgoing paths of degree sequences (2,4), (4,4), (4,2), and (4,2). The circles, despite their different network position (as being part of triangles), have the same set of degree sequences for their paths of length-2. The degree-3 vertices have six length-2 paths: three having the degree sequence (2,2), three having degree sequence (4,2). It is easy to convince oneself that this is close to as dissimilar a network with four degree-2 and two degree-4 vertices can be. Consequently all vertices have negative degree-symmetry indices. It is worth pointing out that Fig. 3(c) possesses other symmetries than degree symmetry. The layout has, for example, reflexive symmetry along a vertical axis. We emphasize that such symmetries would need to be captured by other measures.

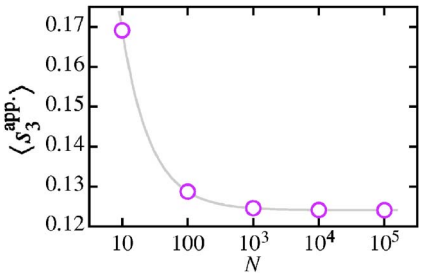


FIG. 4. (Color online) The average approximative symmetry coefficient for $l=3$ and random graphs with $M=2N$. The line is a fit to a power-law decay form $(0.124+0.435N^{-1.02}$, to be exact).

B. Regular networks

If all vertices have the same degree a network is called *regular* [7]. Then by definition all paths are known to fully overlap. This trivial overlap should be canceled in our symmetry measure so $s_l(i)=0$ for all l and i. Since $S_l(i)$ is the number of terms in $\tilde{s}_l(i)$ and all these terms are one we have $S_l(i)=\tilde{s}_l(i)=\Lambda$. Furthermore, $\nu^{\text{app}}=1$ which gives $s_l(i)$ for all vertices and cutoff lengths.

C. Random graphs

Next we evaluate the average approximative symmetry coefficient $\langle s^{\text{app}} \rangle$ for random graphs [7]—graphs obtained by successively adding M edges between N vertices with the restriction that no multiple edge, or self-edge, may occur. Such networks have no correlations at all and can serve as a reference point for neutrality [2]. Ideally we would like such networks to, on average, have a degree-symmetry coefficient of zero. As seen in Fig. 4 $\langle s_l^{\text{app}} \rangle$ converge to a small but positive value. The decay is roughly inversely proportional to N—the same scaling as the fraction of triangles in the network—which suggests that the presence of triangles, and perhaps other short cycles, is an important source of finite size effects of s_l^{app}. We conclude that the Monte Carlo sampling measure s_l^{MC} (or a more elaborate measure) is needed if one wants to compare different networks. If, on the other hand, one aims to compare different vertices of the same network the faster $s_l^{\text{app}}(i)$ calculation is sufficient. This is not an uncommon situation in the design of network measures. Another example of this where neutrality is nonzero in the large-N limit is *modularity*, measuring subgraphs that are densely connected within but not between each other [8].

VI. DEGREE SYMMETRIES OF REAL NETWORKS

In this section we apply our measures to real-world networks. First we take a look at the symmetry coefficients of specific vertices in the metabolic network of humans, then we look at the average symmetry coefficients of various classes of networks.

A. Human metabolic networks

An important use of statistical graph theory is to characterize chemical reaction networks. Of many possible network

representations [9] we let vertices be chemical substances, and for all reactions of an organism we link substrates with products. For example, the hypothetical reaction $A+B \leftrightarrow C+D$ would contribute with the edges (A,C), (A,D) and (B,C), (B,D) to the metabolic network. The data is derived from the KEGG database, and described in detail in Ref. [10]. Since the degree distributions of metabolic networks are highly skewed [11] we use a exponentially increasing set of intervals as equivalence classes [as discussed in the connect of Eq. (14)]: $K_n=2^n$.

It has been argued that degree is strongly related to the function of the chemical substance [11,12]. This means that the degree symmetry potentially can give additional information about the function of the vertices. For the human metabolic network, and $l=2$, roughly half of the vertices have a p value of less than 5% (i.e., in the null-model sampling of the calculation of s_2^{MC}, less than 5% or more than 95% of the values of

$$\frac{\tilde{s}_l(i) - S_l(i)}{\Lambda - S_l(i)} \tag{16}$$

are smaller than the value of the real network). In Fig. 5(a) we show the 2-neighborhood of one vertex with significantly higher s_2^{MC} than expected; Fig. 5(b) depict the 2-neighborhood of a vertex with significantly higher s_2^{MC}. The reason these particular vertices are used as examples is that their 2-neighborhoods are of appropriate sizes, neither too big, nor too small, to be displayed and described. Spermine, Fig. 5(a), is a substance with high degree symmetry—$s_2^{\mathrm{MC}}=0.89\pm0.02$. Both its neighbors are in the same degree-equivalence class of vertices with degree 4 to 7. Of vertices two steps away from spermine there is also a significant overlap with two (out of four) neighbors to the neighbor spermidine being in the equivalence class defined by degrees in the interval [8,16); whereas two vertices are in the equivalence class of degrees in [4,8). The three paths from spermine via *S*-adenosylmethioninamine also contribute to the overlap in the two steps from spermine as two vertices (methylthioadenosine and spermindine) have degrees in the same equivalence class. The neighborhood of C04850, seen in Fig. 5(b), is visually less balanced and also having a negative degree symmetry—$s_2^{\mathrm{MC}}=-0.11\pm0.01$. We note that there are some vertex pairs in the second neighborhood whose degree-classes overlap, but apparently this is not enough to make the symmetry coefficient nonnegative.

B. Average symmetry values

So far we have discussed degree symmetries of vertices. In this section we average s_l over V to obtain a graph-wide measure for degree symmetry. In Table I we display values of s_2^{MC} for a number of different network types. Some of these have highly skewed degree distributions. For these, the exponentially increasing degree equivalence classes of Sec. VI A are appropriate. Since we intend to compare all networks we use the same equivalence classes for all networks. The first observation is that almost all networks have a positive average symmetry coefficient. The only clear exception

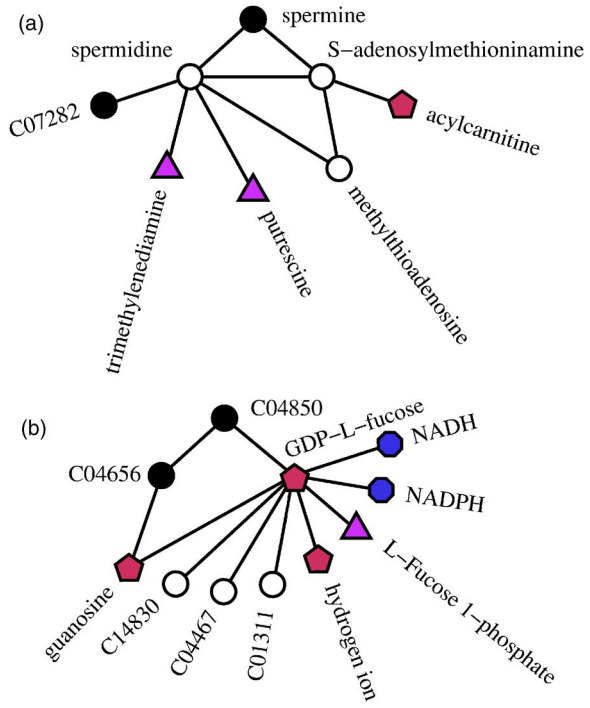


FIG. 5. (Color online) The 2-neighborhood of spermine—a vertex with high degree symmetry—(a), and C04850—a vertex with low degree symmetry—(b), in the human metabolic network. The symbols indicate the equivalence classes defined by exponentially growing intervals. Filled circles have degree 2, unfilled circles have degree 4 or 5, a vertex symbolized by an n-gon have degree in the interval $[2^n,2^{n+1})$. In case the chemical names are overly long the KEGG codes are given ("C" and five digits): C07282 represents eIF5A-precursor-deoxyhypusine, C04850 represents 1,3-β-*D*-galactosyl-(α-1,4-*L*-fucosyl)-*N*-acetyl-*D*-glucosaminyl-*R*, C04556 represents 4-amino-2-methyl-5-phosphomethylpyrimidine, C04467 represents α-*L*-fucosyl-1,2-β-*D*-galactosyl-*R* and C01311 represents 1,4-β-*D*-galactosyl-(α-1,3-*L*-fucosyl)-*N*-acetyl-*D*-glucosaminyl-*R*.

is the airport network. This means that if you start a two-leg airplane trip at a particular airport, choosing between two random itineraries (without caring about the frequency of flights), then the probability of the airports along these itineraries being different in number of connections is smaller than in a random network. The strongest degree symmetries are found in one-mode projections of social affiliation networks. Note that the other social networks, derived from questionnaires and electronic communication does not have such strong symmetry coefficients. In one-mode projections high-degree vertices are known to have strong tendency to attach to other high-degree vertices, and low-degree vertices to attach to other low-degree vertices—so-called assortative mixing [26]. If this property is strong there will be regions of vertices with high-degree and other regions with low-degree vertices. The paths within these regions would also have similar degree sequences. Thus high assortative mixing can be related to high degree symmetry, the first causing the second or vice versa. They are, of course, not equivalent—e.g., the example network with all vertices having positive symmetry coefficients [Fig. 3(b)] is maximally disassorta-

TABLE I. The network sizes N and M and the average numerical degree-symmetry coefficient s_2^{MC} of real-world 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. In the street networks the vertices are Swedish city-street segments connected if they share a junction. In the airport network the vertices are American airports and edges represent a regular, nonstop route. In the citation networks the vertices are papers and two papers are connected if one cites the other. The "scientometrics" network consists of papers from the journal *Scientometrics*. The "small-world" network are all papers citing Ref. [13] or having the phrase "small world" in the title. The board of directors and Ajou student networks are derived from one-mode projections of affiliation networks (where edges goes from persons to corporate boards and university classes, respectively). The Ajou student network is averaged over graphs of 16 semesters. One edge represents two students taking at least three classes together that semester. The high school networks are gathered from questionnaires—an edge means that two persons have listed each other as acquaintances. It is averaged over 84 individual schools. In the electronic communication networks one edge represent that at least one of the vertices has contacted the other over some electronic medium. 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 interact (the two graphs correspond to two different types of experiments determining the interaction edges). The metabolic networks consist of chemical substances and edges are constructed as described in Sec. VI A. Values for animal metabolism is averaged over six networks, fungi metabolism is averaged over two, and bacteria metabolism is averaged over 96 networks.

Network		Reference	N	M	s_2^{MC}
Geographical networks	Interstate highways		935	1315	0.016±0.003
	Streets, Stockholm	[16]	3325	5100	0.014±0.003
	Streets, Malmö	[16]	1868	3026	0.020±0.003
	Streets, Göteborg	[16]	1258	1516	0.026±0.003
	Airport	[14]	332	2126	−0.0573±0.0002
Citation networks	Scientometries	[15]	2728	10398	0.015±0.020
	Small world	[15]	233	994	0.007±0.002
One-mode projections of affiliation networks	Board of direction	[17]	6193	43074	0.175±0.004
	Ajou University students	[18]	7285±128	75898±6566	0.13±0.01
Acquaintance networks	High school friendship	[19]	571±43	1104±60	0.020±0.002
Electronic communication networks	E-mail	[20]	3186	31856	−0.01±0.01
	Internet community	[21]	28295	115335	0.01898±0.0001
Food webs	Little Rock lake	[22]	92	960	0.042±0.001
	Ythan estuary	[23]	134	593	0.027±0.002
Neural network	*C. elegans*	[24]	280	1973	0.0839±0.0001
Biochemical networks	*S. cervisiae* protein	[5,25]	4580	7434	0.0205±0.0001
	S. cervisiae genetic	[5,25]	4580	5129	0.0996±0.0001
	Animal metabolism	[10]	1621±123	4662±473	0.02±0.01
	Plant metabolism, *A. thaliana*	[10]	1561	4302	0.0133±0.0003
	Fungi metabolism	[10]	1281±97	3654±289	0.03±0.02
	Bacteria metabolism	[10]	1070±35	2776±109	0.018±0.002

tively mixed (in the sense of Ref. [26]). Where the weak symmetry coefficients of other networks come from is outside the scope of this investigation. One possible explanation would be that functional units [27] might often be degree-symmetric centers.

VII. SUMMARY AND CONCLUSIONS

We have derived a measure for a specific notion of symmetry in networks—the property that the paths out from a vertex have overlapping degree sequences. The measure is designed so that random networks, conditioned only to have the same set of degrees as the original network, have the value zero. We propose two versions of the symmetry coefficient, the first being approximately zero for random networks, the second requiring a randomization procedure (and thus longer simulation time) but being more accurately zero for random networks. The measure was evaluated on example graphs. We show that they are able to detect vertices in degree symmetric, and potentially functionally meaningful positions in the human metabolic network. The average degree symmetry of various networks were also investigated. We found almost all networks having a weakly positive degree coefficient. The exceptions being the network of Ameri-

can airports and their interconnections (having a negative degree-symmetry coefficient) and one-mode projections of social affiliation networks (having rather strongly positive values). Our measure is not the first to be based on a the properties of paths going out from a vertex. For example, people have been using path counts for assessing the functional similarity of pairs of vertices [28–30]. In social network studies such measures are commonly called "egocentric" [31].

Symmetry concepts have been successfully utilized in many field of physics. We believe degree symmetry, and other classes of network symmetries, will be a fruitful direction of future network studies. Degree symmetry is in particular, we believe, an important concept for networks where degree is strongly related to the function of the vertex. Two open questions from this study is what causes the rather ubiquitous weakly positive degree symmetries, and what process in the airline decision making that causes the negative average symmetry coefficient of the airline network.

ACKNOWLEDGMENTS

The author acknowledges financial support from the Wenner-Gren foundations and help with data acquisition from: Gerald Davis, Jean-Pierre Eckman, Michael Gastner, Mikael Huss, Beom Jun Kim, Sungmin Park, and Martin Rosvall. This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by Grant No. P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is made to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[3] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[4] J. M. Roberts, Jr., Soc. Networks **22**, 273 (2000).
[5] P. Holme and M. Huss, J. R. Soc., Interface **2**, 327 (2005).
[6] M. E. J. Newman, Phys. Rev. E **70**, 056131 (2004).
[7] S. Janson, T. Łuczac, and A. Ruciński, *Random Graphs* (Wiley, New York, 1999).
[8] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Phys. Rev. E **70**, 025101(R) (2004).
[9] J. Zhao, H. Yu, J. Luo, Z. W. Cao, and Y.-X. Li, Chinese Science Bulletin **51**, 1529 (2006).
[10] M. Huss and P. Holme, e-print q-bio.MN/0603038; (http://www.genome.jp/).
[11] H. Jeong, B. Tombor, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[12] R. Guimerà and L. A. Nunes Amaral, Nature (London) **433**, 895 (2005).
[13] S. Milgram, Psychol. Today **2**, 60 (1967).
[14] Data obtained from http://vlado.fmf.uni-lj.si/pub/networks/pajek/data/gphs.htm
[15] The citation networks were obtained from http://vlado.fmf.uni-lj.si/pub/networks/data/cite/. These networks are the result of searches in the database WebofScience used with the permission of ISI Philadelphia.
[16] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[17] G. F. Davis, M. Yoo, and W. E. Baker, Strat. Org. **1**, 301 (2003).
[18] P. Holme, S. M. Park, B. J. Kim, and C. R. Edling, e-print cond-mat/0411634.
[19] P. Bearman, J. Moody, and K. Stovel, Am. J. Sociol. **110**, 44 (2004).
[20] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14333 (2004).
[21] P. Holme, C. R. Edling, and F. Liljeros, Soc. Networks **26**, 155 (2004).
[22] N. D. Martinez, Ecol. Monogr. **61**, 367 (1991).
[23] S. J. Hall and D. Raffaelli, J. Anim. Ecol. **60**, 823 (1991).
[24] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, Philos. Trans. R. Soc. London, Ser. B **314**, 1 (1986).
[25] P. Pagel *et al.*, Bioinformatics **21**, 832 (2004).
[26] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[27] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[28] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, SIAM Rev. **46**, 647 (2004).
[29] G. Jeh and J. Widom, Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining (Edmonton, 2002), pp. 538–543.
[30] E. A. Leicht, P. Holme, and M. E. J. Newman, Phys. Rev. E **73**, 026120 (2006).
[31] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, 1994).