Spectral measures of bipartivity in complex networks

Ernesto Estrada[1,*] and Juan A. Rodríguez-Velázquez[2]

[1]*Complex Systems Research Group, X-rays Unit, RIAIDT, Edificio CACTUS, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain*

[2]*Department of Mathematics, University Carlos III de Madrid, 28911 Leganés (Madrid), Spain*

We introduce a quantitative measure of network bipartivity as a proportion of even to total number of closed walks in the network. Spectral graph theory is used to quantify how close to bipartite a network is and the extent to which individual nodes and edges contribute to the global network bipartivity. It is shown that the bipartivity characterizes the network structure and can be related to the efficiency of semantic or communication networks, trophic interactions in food webs, construction principles in metabolic networks, or communities in social networks.

PACS number(s): 89.75.Hc, 89.75.Fb, 87.10.+e, 89.20.−a

I. INTRODUCTION

The study of complex networks has become an important area of multidisciplinary research involving physics, mathematics, biology, social sciences, informatics, and other theoretical and applied sciences. The importance of this field resides in the existence of a unifying language to describe disparate real-world systems that are of great relevance in modern society, ranging from the Internet or powergrids to metabolic or protein interaction networks (PINs) [1–6]. Recently, "bipartivity" has been proposed as an important topological characteristic of complex networks [7]. A network (graph) $G=(V,E)$ is called *bipartite* if its vertex set V can be partitioned into two subsets V_1 and V_2 such that all edges have one endpoint in V_1 and the other in V_2. There are numerous natural systems that are modeled as bipartite networks, such as reaction networks or "two-mode" networks, in which two disjointed sets of nodes are related by links representing the relationship between the elements of both classes [8–11]. For instance, they can represent authors that cite (or are cited by) papers, people that belong to institutions, cities that have certain services, or voting results of delegates concerning certain proposals. Holme *et al.* [7] pointed out several areas for the potential application of a quantitative measure of bipartivity, such as network studies of sexually transmitted diseases, trade networks of buyers and sellers, "genealogical" networks of disease outbreak, and food webs [7]. Just to give some examples of the relevance of a measure of network bipartivity we will mention the following scenarios. Let us consider the study of a sexually transmitted disease. It is known that the transmission rates for homosexual and heterosexual contacts differ [12]. Consequently, the transmission of this disease will depend on how bipartite the corresponding network is. In other words, having an idea of the bipartivity of sexual networks, we will have an idea on the rate of spreading of a sexually transmitted disease.

Another scenario in which the analysis of network bipartivity can be of great utility is for the study of information and communication networks. In a dictionary, for instance, all entries should be related in a self-referential way showing a large transitivity between triples of words. In communication networks, like a network of airports or the Internet, the network bipartivity indicates that two separate groups exist, where direct communication is only possible between nodes in the different groups. The lack of direct communication between "members" of the same group is an indication of the lack of efficiency of such networks. If we are considering, for instance, a bipartite network of U.S. airports in which airplanes can fly from the east to the west coast but no east-east or west-west allowed, then it is obvious that we have to fly from New York to San Francisco and from there to Miami because no direct flight from New York to Miami can exist. The ideal case coincides with the least bipartite network in which each pair of airports are connected by a flight, i.e., a complete graph. This is an exaggerated, but illustrative example about the importance of bipartivity, for the design of an efficient communication network as well as about the importance of having effective ways of measuring bipartivity in complex networks.

We introduce here a spectral measure of bipartivity for complex networks that is easy to compute, changes monotonically with changes in network bipartivity, and allows the calculation of individual node contributions to global bipartivity. We analyze the spectral measure of bipartivity by considering several real-world networks, which include information, biological, social, and technological networks.

II. THEORETICAL METHODS

Our approach to define a measure of network bipartivity is based on the concept of closed walks. A walk of length r is a sequence of (not necessarily different) vertices $v_1, v_2, \cdots, v_r, v_{r+1}$ such that for each $i=1,2,\cdots,r$ there is a link from v_i to v_{i+1}. A closed walk (CW) is a walk in which $v_{r+1}=v_1$. A CW is called odd (even) if r is odd (even). A cycle is a CW in which all vertices in $v_1, v_2, \cdots, v_r, v_{r+1}$ are different. The theoretical motivation of our measure of bipar-

tivity arose from the following result, which follows straightforwardly from known results in graph theory [13]: A nontrivial graph (without self-loops) is bipartite if and only if it contains no odd closed walks.

Let $G=(V,E)$ be a network with N nodes and having eigenvalues of the adjacency matrix $\sigma=(\lambda_1,\cdots,\lambda_N)$ [14]. The subgraph centralization measure has been defined as a weighted sum of the so-called spectral moments of the adjacency matrix $\mu_l=\Sigma_{j=1}^{N}(\lambda_j)^l$. It is obvious that $\mu_0=N$, i.e., it is equal to the number of nodes in the network, $\mu_1=0$ for nonweighted networks, and μ_l $(l\neq 0,1)$ represents the number of CWs of length l in the network. Accordingly, the subgraph centralization for a network $G=(V,E)$ is equal to [14]:

$$\langle SC\rangle = SC(G) = \frac{1}{N}\sum_{l=1}^{\infty}\frac{\mu_l}{l!} = \frac{1}{N}\sum_{j=1}^{N}e^{\lambda_j}. \quad (1)$$

The subgraph centralization $\langle SC\rangle$ can be expressed as the sum of two contributions, one coming from odd and the other from even CWs

$$\langle SC\rangle = \frac{1}{N}\sum_{j=1}^{N}[\cosh(\lambda_j)+\sinh(\lambda_j)] = \langle SC\rangle_{even}+\langle SC\rangle_{odd}. \quad (2)$$

If $G(V,E)$ is bipartite, then $\langle SC\rangle_{odd}=1/N\sum_{j=1}^{N}\sinh(\lambda_j)=0$, because there are no odd CWs in the network and, therefore,

$$\langle SC\rangle = \langle SC\rangle_{even} = \frac{1}{N}\sum_{j=1}^{N}\cosh(\lambda_j). \quad (3)$$

Consequently, the proportion of even CWs to the total number of CWs is a measure of the network bipartivity

$$\beta(G) = \frac{\langle SC\rangle_{even}}{\langle SC\rangle} = \frac{\langle SC\rangle_{even}}{\langle SC\rangle_{even}+\langle SC\rangle_{odd}} = \frac{\sum_{j=1}^{N}\cosh(\lambda_j)}{\sum_{j=1}^{N}e^{\lambda_j}}. \quad (4)$$

It is evident that $\beta(G)\leq 1$ and $\beta(G)=1$ if, and only if, G is bipartite, i.e., $\langle SC\rangle_{odd}=0$. Furthermore, as $0\leq\langle SC\rangle_{odd}$ and $\sinh(\lambda_j)\leq\cosh(\lambda_j)$, $\forall\lambda_i$, then $\frac{1}{2}<\beta(G)$ and $\frac{1}{2}<\beta(G)\leq 1$. The lower bound is reached for the least possible bipartite graph with N nodes, which is the complete graph K_N. As the eigenvalues of K_N are $N-1$ and -1 (with multiplicity $n-1$) [13], then $\beta(G)\rightarrow\frac{1}{2}$ when $N\rightarrow\infty$ in K_N. This lower bound coincides with that given by Holme *et al.* [7] in the $N\rightarrow\infty$ limit for their measures.

Despite the fact that both measures coincide in the extreme values, they show different values for the rest of networks. Consider, for instance, a process in which new edges are successively added to a bicomplete graph $K_{N1,N2}$, which has two disjointed sets of nodes V_1 and V_2 of cardinality N_1 and N_2, respectively. We will obtain the least bipartite graph K_N, $N=N_1+N_2$, by joining together all of the N_1 nodes of V_1 and all of the N_2 nodes of V_2. The addition of one edge to V_1 (V_2) will introduce N_2 (N_1) triangles to the network. In the

case of star graphs $K_{1,N2}$, the addition of one edge to V_2 introduces only one triangle to each graph, which make the graphs with larger N_2 more bipartite because the proportion of even to total CWs, $\beta(G)$, increases. In this case, $\beta(G)$ coincides with b_1, which is one minus the proportion of frustrated to total number of edges in the network. It can be seen that b_1 drops dramatically in K_3 (the triangle) despite it being very close to bipartite, indicating that the aforementioned proportion is the important aspect for bipartivity and not the number of frustrated edges to be removed to make the graph bipartite. Thus, both measures follow the same trend for $K_{1,N2}$ as shown in Fig. 1(a). However, both measures give different trends if we consider graphs of the type $K_{N1,N2}$ $(N_1\neq 1, N_2\neq 1)$. In Fig. 1(b), we illustrate this situation by adding edges to the graph $K_{2,3}$. Here the addition of one edge to V_2 produces two triangles, while its addition to V_1 produces three. Thus, the numbers of frustrated and total edges are the same, but the proportion of even to total CWs is not. As a consequence, b_1 gives the same value for these pairs of graphs. However, $\beta(G)$ shows that the graphs having the frustrated edge joining pairs of nodes at V_2 are more bipartite than those in which the frustrated edge is joining nodes at V_1. This result can be straightforwardly generalized to any kind of network.

A desired property for $\beta(G)$ is that it changes monotonically as the bipartivity of the graph changes. Let G be a noncomplete graph and let e be an edge of the complement of G. Let $G+e$ be the graph obtained by adding the edge e to G. In this situation, there exist real and nonnegative numbers a and b, such that $SC_{even}(G+e)=SC_{even}(G)+a$ and $SC_{odd}(G+e)=SC_{odd}(G)+b$. Notice that a is the contribution of edge e to SC_{even} and b is the contribution of edge e to SC_{odd}. Thus, with the above notation, if $b\geq a$, then $\beta(G)\geq\beta(G+e)$. That is, as $a+b/2\geq a$ and $SC_{even}(G)\geq SC(G)/2$ then $(a+b)\times SC_{even}(G)\geq a\times SC(G)$. The addition of $SC_{even}(G)\times SC(G)$ to both terms and further reordering gives $SC_{even}(G)[SC(G)+a+b]\geq SC(G)[SC_{even}(G)+a]$ and, consequently,

$$\beta(G) = \frac{SC_{even}(G)}{SC(G)} \geq \frac{SC_{even}(G)+a}{SC(G)+a+b} = \beta(G+e), \quad (5)$$

which proves the monotony of the change for the spectral bipartivity measure as can be seen in Fig. 1(b).

The contribution of node i to network bipartivity $\beta(i)$ can be obtained by using the subgraph centrality of node i [15] as follows:

$$SC(i) = \sum_{j=1}^{N}[v_j(i)]^2 e^{\lambda_j}, \quad (6)$$

where $(v_1,v_2,\ldots,v_n)$ is an orthonormal basis of R^N composed by eigenvectors of the adjacency matrix associated with the eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_N$, and $v_j(i)$ is the ith component of v_j. Hence, $\beta(i)$ is given by

$$\beta(i) = \frac{\sum_{j=1}^{N}[v_j(i)]^2\cosh(\lambda_j)}{\sum_{j=1}^{N}[v_j(i)]^2 e^{\lambda_j}}. \quad (7)$$
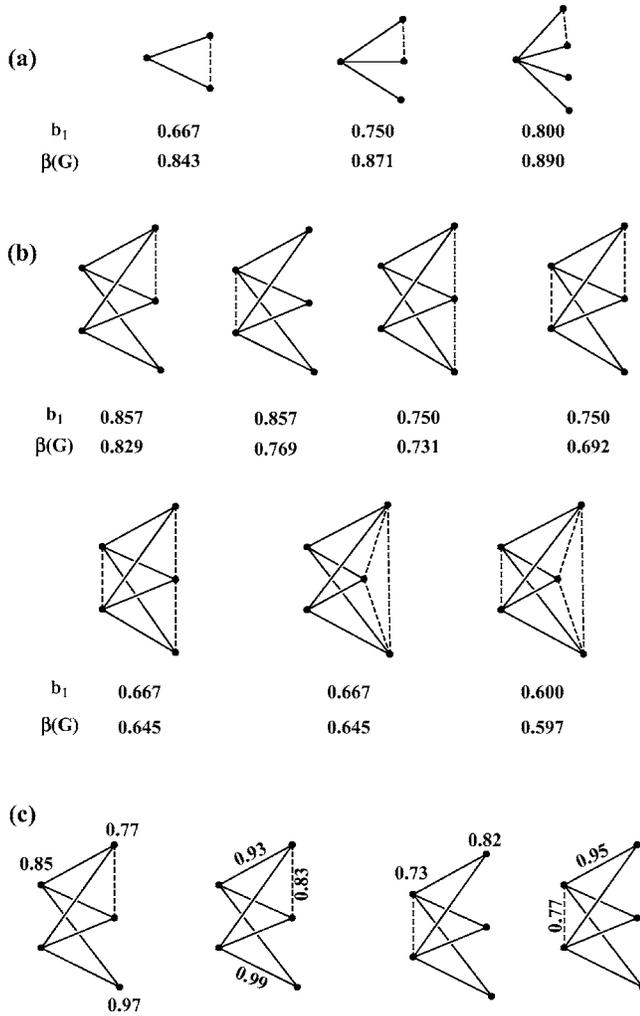
FIG. 1. Some graphs used in the discussion of global and local spectral bipartivity measures. (a) Graphs obtained by the addition of one frustrated edge (dotted line) to stars $K_{1,N2}$. Only one triangle is introduced to each graph after the addition of the new links. Thus, the proportion of even to total CWs (even+odd), $\beta(G)$, increases as N_2 increases and the networks with larger N_2 are more bipartite. The proportion of frustrated to total number of edges decreases as N_2 increases and b_1 follows the same trend that $\beta(G)$. (b) Values of network bipartivity, b_1 and $\beta(G)$, for the graphs produced by successive addition of frustrated edges to $K_{2,3}$. It is observed that $\beta(G)$ shows a monotonically decreasing trend from the bipartite graph to the complete graph. b_1 does not differentiate between pairs of graphs having the same number of frustrated edges despite the fact that they show different proportions of even to total number of CWs. (c) Values of $\beta(i)$ and $\beta(e)$ for the first two graphs in (b). The lowest values of these measures indicate those nodes and/or links which contribute more to the nonbipartivity of the graph. Despite the fact that some links with values different from one are not frustrated edges, their being removed reduces the number of odd CWs, e.g., triangles, increasing the bipartivity of the network.

On the other hand, we can make use of our finding about the monotony of $\beta(G)$ to calculate the contribution of an edge e to the network bipartivity $\beta(e)$. Let e be an edge of the network G and let $G-e$ be the network obtained by removing e from G. Then, $\beta(e)$ is given by $\beta(e)$

$= 1 - [\beta(G-e) - \beta(G)]$ {the formula $1-[\cdots]$ is used to make these values follow the same trend as those of $\beta(i)$}. In Fig. 1(c), we illustrate the values of $\beta(i)$ and $\beta(e)$ for the nodes and links of the first two graphs of Fig. 1(b).

III. SPECTRAL BIPARTIVITY IN REAL-WORLD COMPLEX NETWORKS

A total of 17 complex networks were studied, including 2 semantic networks, one based on Roget's Thesaurus of English (Roget) [16] and the other based on the Online Dictionary of Library and Information Science (ODLIS) [17] (two words are connected if one is used in the definition of the other); 4 social networks that include a scientific collaboration network in the field of computational geometry [18], inmates in prison [19], injecting drug users (IDUs) that have shared a needle in the last six months [20], and the friendship network between members of a karate club [21]; 7 biological networks including the protein-protein interaction network (PIN) of yeast compiled by Bu *et al.* [22] on data obtained by von Mering *et al.* [23]; the direct transcriptional regulation between genes in yeast [24], and 4 food webs representing trophic relations in different ecosystems: Coachella Valley [25], Grassland [26], Ythan Estuary [27], and El Verde Rainforest [28]; 5 technological networks, one based on the airport transportation network in the U.S. in 1997 [29], the Internet at the autonomous systems (AS) level as from April 1998 [30], and 3 electronic sequential logic circuits parsed from the ISCAS89 benchmark set, where nodes represent logic gates and flip flops [31].

The results of calculations are given in Table I. There are four networks in which the low values of $\beta(G)$ are indicative of the efficient construction of these networks. These are the two semantic networks, the transportation network of U.S. airports, and the Internet at AS. As we have previously remarked, in semantic networks, like Roget and ODLIS, all individual entries must be bootstrapped from other entries in a self-referential way [32], which immediately precludes bipartivity from these semantic networks. In transportation or communication networks, a significant degree of bipartivity is translated into a low efficiency in traveling between the nodes located in the same disjointed set, which makes the network inefficient. However, bipartivity can also be a desired property in technological networks, as demonstrated by the high bipartivity observed for the three electronic circuits studied.

The four social networks analyzed show very different bipartivity values. While the karate club and prison networks reveal certain bipartivity, the IDUs and the collaboration network show values of $\beta(G)=0.5$. The bipartivity observed for the Karate Club network can be rationalized by the fact that there are two main disjointed *hubs* in the network: the club's instructor and the club's president. The rest of the nodes, which form the other set, show a low average degree (3.84), indicating that there is not a high number of links between them. At the other extreme is the collaboration network, which has a value of $\beta(G)=0.5$. This network consists of clusters of fully connected nodes, formed by coauthors of a

TABLE I. Values of the spectral network bipartivity measure, $\beta(G)$, for complex networks of different types and sizes. The number of nodes (N) and edges (E) are given along with clustering coefficients. The correlation coefficients (R^2) of the linear regression between bipartivity and clustering coefficients are also given in order to show their linear independence.

Type	Network	N	E	$\beta(G)$	$C^{(1)}(G)^{a}$	$C^{(2)}(G)^{a}$
Information	Roget	994	3640	0.529	0.162	0.134
	ODLIS	2898	16376	0.500	0.351	0.056
Social	Karate club	34	78	0.597	0.588	0.256
	Prison	67	142	0.698	0.330	0.288
	Drugs	616	2012	0.500	0.722	0.368
	Geom	3621	9461	0.500	0.679	0.219
Biological	Coachella	30	241	0.500	0.707	0.697
	Grassland	75	113	0.743	0.497	0.174
	Stony stream	112	830	0.815	0.076	0.020
	El Verde	156	1439	0.500	0.231	0.232
	Trans-yeast	662	1062	0.960	0.092	0.016
	PIN-yeast	2224	6608	0.500	0.201	0.102
Technological	USAir97	332	2126	0.500	0.749	0.396
	Internet-1998	3522	6324	0.502	0.340	0.014
	Electronic1	122	189	0.948	0.064	0.344
	Electronic2	252	399	0.950	0.060	0.310
	Electronic3	512	819	0.952	0.058	0.290
R^2					0.450	0.100

[a]$C^{(1)}(G)$ as defined by Watt and Strogatz and $C^{(2)}(G)$ defined as three times the number of triangles divided by the number of connected triples in the network (see [1] for definitions).

particular paper, which are interconnected and make the network nonbipartite.

Holme *et al.* [7] considered the study of food webs as a potential area for the application of bipartivity measures. This supposition is based on the idea that the simplest picture of a food web can be represented as different "trophic" levels where species in one level predate species located at the level below, producing networks with a high degree of bipartivity. This appears to be the case for Grassland and Stony stream, both of which have a significant degree of bipartivity. Grassland represents a good example of a network where trophic levels are responsible for the bipartivity observed. The trophic relations observed are only interclasses and form an almost bipartite graph (see Fig. 2). The situation is quite different for the other two food webs, which show a very low degree of bipartivity as a consequence of their larger number of trophic relations between species. Similar results ($\beta \approx 0.5$) are obtained for other food webs, such as Little Rock, Scotch Broom, and Ythan Estuary with parasites (data not shown). Therefore, we believe that bipartivity is not a general characteristic of ecological systems despite the fact that food webs with a pronounced degree of bipartivity can be found as a consequence of the trophic relations between classes.

The metabolic network of yeast is the most bipartite network of all those studied here. This finding can be explained by the construction of the network, which is based on two sets of nodes—one representing regulating genes and the other representing regulated genes—with connections between both sets. In contrast, the PIN of yeast shows a low degree of bipartivity despite its low clustering, which indicates that odd cycles larger than triangles play an important role in the interactions between proteins in this organism.

The utility of the local bipartivity index $\beta(i)$ lies in the possibility of identifying those nodes and links that contribute significantly to the bipartivity in a network. Removing them will significantly affect the bipartivity degree of the whole network. For instance, removing the node with the lowest contribution to $\beta(i)$ in Grassland increases its bipartivity from $\beta(G)=0.734$ to $\beta(G)=0.794$, and this example can reach $\beta(G)=0.863$ by removing the three nodes with the lowest contribution to $\beta(i)$. It is possible to find numerous practical applications for the detection of node and/or link bipartivity in real-world networks. For instance, this approach can be applied in the field of sexually transmitted diseases, where nodes with different contributions to the network bipartivity can play different roles in the transmission of such diseases [12].

IV. CONCLUSIONS

Network bipartivity is a topological characteristic that cannot be accounted for by other structural measures, such as clustering coefficients (see Table I). On the other hand, the
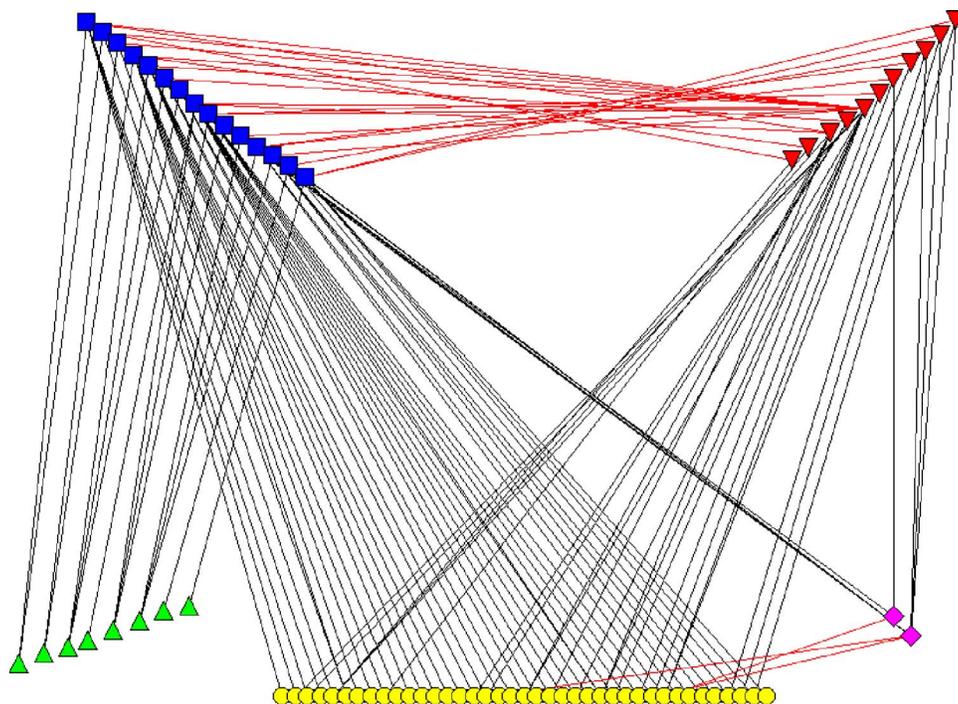
FIG. 2. (Color online). Grassland food web represented as an almost bipartite network. Up-triangles (green) represent grass, (blue) squares represent herbivores, (yellow) circles represent parasitoids, (red) down triangles represent hyperparasitoids, and (pink) diamonds represent hyper-hyperparasitoids. The network bipartivity is illustrated as two sets of nodes situated at the top and bottom of the graphic. Links between nodes in disjointed sets are the vertical ones (black) and those between nodes in the same set are the horizontal ones (red).

physical consequences of network bipartivity depend on the particular network and processes which are studied and cannot be generalized even for networks in the same fields. While social networks of friendships are expected to be non-bipartite due to the propensity for two of one's friends to also be friends of each other, in sexual networks bipartivity can arise from homosexual contacts. A similar situation occurs in food webs, where in some systems, the trophic relations between species in different trophic levels can introduce bipartivity to the network—a situation that does not occur if the species are in the same trophic level. The exception appears to be communication and/or information systems, in which the lack of bipartivity represents a measure of efficiency in the network construction.

The study of complex networks has proved that in order to understand the evolution of the processes taking place in such systems, it is necessary to characterize the topological properties of these networks. In this way, the clustering co-efficient, assortativity coefficient, diameter, centrality measures, etc., are well-known network properties giving important information about the architecture of complex networks. Bipartivity measures, such as the one introduced by Holme *et al.* [7] and the one presented here, are a new measurable characteristic of networks, not accounted by any other network parameter, which should be considered in the study of complex systems represented by networks. Network bipartivity is of great importance for the study of sexual networks, genealogical networks of a disease outbreak, food webs, technological networks, and so forth.

ACKNOWLEDGMENTS

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[2] S. H. Strogatz, Nature (London) **410**, 268 (2001).

[3] R. Albert and A-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[5] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[6] A-L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).

[7] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, Phys. Rev. E **68**, 056107 (2003).

[8] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A-L. Barabási, Nature (London) **407**, 651 (2000).

[9] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, UK, 1994).

[10] S. A. Morris, e-print cond-mat/0501386 (unpublished).

[11] S. A. Souma, Y. Fujiwara, and H. Aoyama, Physica A **324**, 396 (2003).

[12] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, Oxford, UK, 1991).

[13] C. Chartrand and L. Lesniak, *Graphs and Digraphs*, 3rd edition (Chapman and Hall, New York, 1996).

[14] D. Cvetković, P. Rowlinson, and S. Simić, *Eigenspaces of Graphs* (Cambridge University Press, Cambridge, UK, 1997).

[15] E. Estrada and Juan A. Rodríguez-Velázquez, Phys. Rev. E **71**, 056103 (2005).

[16] Roget's Thesaurus of English Words and Phrases (2002). Project Gutenberg. http://gutenberg.net/etext/22

[17] ODLIS: Online Dictionary of Library and Information Science (2002). http://vax.wcsu.edu/library/odlis.html

[18] B. Jones, Computational Geometry Database, February 2002. http://compgeom.cs.edu/~jeffe/compgeom/biblios.html

[19] D. MacRae, Sociometry **23**, 360 (1960).

[20] Data for this project was provided in part by NIH Grant Nos. DA12831 and HD41877. Those interested in obtaining a copy of this data should contact Professor James Moody (moody.77@sociology.osu.edu).

[21] W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[22] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, Nucleic Acids Res. **31**, 2443 (2003).

[23] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork, Nature (London) **417**, 399 (2002).

[24] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).

[25] P. H. Warren, Oikos **55**, 299 (1989).

[26] N. D. Martinez, B. A. Hawkins, H. Ali Dawah, and B. P. Feifarek, Ecology **80**, 1044 (1999).

[27] M. Huxman, S. Beany, and D. Raffaelli, Oikos **76**, 284 (1996).

[28] *The Food Web of a Tropical Rainforest*, edited by R. B. Waide and W. B. Reagan (University of Chicago Press, Chicago, 1996).

[29] Taken from http://vlado.fmf.uni-lj.si/pub/networks/data/default.htm as collected from North American Transportation Atlas Data (NORTAD): http://www.bts.gov/publications/north_american_transportation_atlas_data/

[30] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[31] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[32] M. Sigman and G. A. Cecchi, Proc. Natl. Acad. Sci. U.S.A. **99**, 1742 (2002).