

Communicability and multipartite structures in complex networks at negative absolute temperatures

Ernesto Estrada,^{1,3,*} Desmond J. Higham,² and Naomichi Hatano³¹*Institute of Complexity Science, Department of Physics and Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, United Kingdom*²*Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, United Kingdom*³*Institute of Industrial Science, University of Tokyo, Komaba, Meguro 153-8505, Japan*

(Received 16 April 2008; published 8 August 2008)

We here present a method of clearly identifying multipartite subgraphs in a network. The method is based on a recently introduced concept of the communicability, which very clearly identifies communities in a complex network. We here show that, while the communicability at a positive temperature is useful in identifying communities, the communicability at a negative temperature is useful in identifying multipartite subgraphs; the latter quantity between two nodes is positive when the two nodes belong to the same subgraph and is negative when they do not. The method is able to discover “almost” multipartite structures, where intercommunity connections vastly outweigh intracommunity connections. We illustrate the relevance of this work to real-life food web and protein-protein interaction networks.

DOI: [10.1103/PhysRevE.78.026102](https://doi.org/10.1103/PhysRevE.78.026102)

PACS number(s): 89.75.Fb, 89.75.Hc, 87.10.-e, 02.10.Ox

INTRODUCTION

Since the publication of the seminal paper by Girvan and Newman in 2002 [1], there has been a dramatic explosion of strategies and approaches for detecting communities in complex networks [2–8]. These networks represent systems in which the nodes account for the interacting entities, such as molecules, cells, animal species, technological artifacts, etc., and the links record their interactions. A network community is a group of entities which have a large number of interactions among them but very few interactions with members of the other groups [1].

A completely contrasting type of structure can also be found in complex networks. In such structures the nodes are organized in groups of (almost) bipartite subgraphs [9,10]. Here, the phrase bipartite signifies that the (sub)graph can be divided into two distinct subsets such that all connections are from one subset to the other. In this organization a community is a collection of nodes that are not interconnected, but have many neighbors in common. These structures appear in a wide variety of complex systems. For instance, in the context of protein-protein interaction networks, the bipartite structure is intimately related to the existence of complementary binding domains in proteins [11,12] as well as to the identification of essential proteins [13]. In food webs, bipartite substructures might represent different trophic levels, such as those formed by plants and herbivores, and in social networks some of these disjoint sets can represent potential competitors as in sellers-buyers or dating networks.

The main purpose of the present paper is to identify a bipartite structure (and more generally a multipartite structure) in an arbitrary network. To do this, we introduce the concept of communicability with a negative temperature. Two of the authors recently demonstrated that we can identify communities in an arbitrary network very clearly with

the use of the communicability (with a positive temperature) [14]. Here we show that we can clearly identify a multipartite structure with the use of the communicability with a negative temperature.

PRELIMINARIES

We represent a complex network by an undirected graph $G=(V,E)$, where V and E are the sets of nodes and links, respectively. Let G have n nodes. Then the adjacency matrix of G , $\mathbf{A}(G)=\mathbf{A}$, is a square, symmetric matrix of order n , whose elements A_{ij} are ones or zeroes if the corresponding nodes are adjacent or not, respectively. This matrix has n (not necessarily distinct) real-valued eigenvalues [15], which are denoted here by $\lambda_1, \lambda_2, \dots, \lambda_N$, and assumed to be labeled in a nonincreasing manner: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$. Let ϕ_j be an orthonormal eigenvector corresponding to the eigenvalue λ_j . Then, $\phi_j(i)$ designates the component of this eigenvector to the i th node in the network. A graph is said to be bipartite if its nodes can be divided into two disjoint sets V_1 and V_2 such that every link connects a vertex in V_1 and one in V_2 , but there is no edge between two nodes in the same set.

THEORETICAL APPROACH

Let us consider the communicability between a pair of nodes p and q in the network [14],

$$G_{pq}(\beta) = (e^{\beta\mathbf{A}})_{pq} = \sum_{j=1}^n \phi_j(p)\phi_j(q)e^{\beta\lambda_j}, \quad (1)$$

where $\beta = \frac{1}{kT}$ is the inverse temperature, k is the Boltzmann constant, and T is the absolute temperature [16]. We have previously shown that Eq. (1) represents the Green’s function of the network, that is, a function which expresses how an impact propagates from one node to another node in the graph [14]. Using the quantity (1), we constructed the “communicability graph $\Theta(G)$ ” from the original graph G . The

*Corresponding author. estrada66@yahoo.com

communicability graph indicates communities in the graph G very clearly [17].

It is known from spectral clustering techniques that the eigenvectors corresponding to positive eigenvalues give a partition of the network into clusters of tightly connected nodes [18,19]. In contrast, the eigenvectors corresponding to negative eigenvalues make partitions in which nodes are not close to those which they are linked, but rather to those with which they are not linked [18,19]. Such differences have made possible the classification of complex networks into four universal classes [20]. Let us demonstrate the above statements for a cycle C_n with even number of nodes n . The adjacency matrix of a cycle is diagonalized by the eigenvectors $\{\phi_j(p) = \text{Re}e^{ik_j p / \sqrt{n}}\}$, where $\phi_j(p)$ denotes the component on the p th node of the eigenvector with the label j , and $k_j \equiv 2\pi(j-1)/n$. The corresponding eigenvalues are $\lambda_j = 2 \cos k_j$. The largest eigenvalue $\lambda_1 = 2$ is given by the eigenvector $\phi_1(p) = \text{const}$, which is a partition of the whole network C_n into just one cluster. The second largest eigenvalue $\lambda_2 = 2 \cos(2\pi/n)$ is given by the eigenvector $\phi_2(p) = \cos(2\pi p/n) / \sqrt{n}$, which is positive for almost half of the nodes and negative for the other half. In short, the second largest eigenvalue gives a partition of the network C_n into two clusters. On the other hand, the lowest eigenvalue $\lambda_{1+n/2} = -2$ is given by the eigenvector $\phi_{1+n/2}(p) = (-1)^p / \sqrt{n}$, which gives a partition of the network C_n into two subgraphs; that is, the eigenvector is positive for the nodes with even p and negative for the nodes with odd p .

From the perspective of the communicability function (1) we can say that a positive (negative) value of β increases the contribution of the positive (negative) eigenvalues to the communicability function. Then if we write the communicability function as

$$G_{pq}(\beta) = \sum_{\lambda_j < 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j} + \sum_{\lambda_j = 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j} + \sum_{\lambda_j > 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j} \quad (2)$$

we have that

$$G_{pq}(\beta > 0) \approx \sum_{\lambda_j > 0} \phi_j(p)\phi_j(q)e^{\beta\lambda_j}, \quad (3)$$

$$G_{pq}(\beta < 0) \approx \sum_{\lambda_j < 0} \phi_j(p)\phi_j(q)e^{-|\beta|\lambda_j}. \quad (4)$$

In other words, $G_{pq}(\beta > 0)$ determines a partition of the network into clusters of tightly connected nodes, which corresponds to the network communities. On the other hand, for $G_{pq}(\beta < 0)$ the network is partitioned in such a way that the nodes are close to other nodes which have similar patterns of connections with other sets of nodes, i.e., nodes to which they are structurally equivalent. In the first case, we say that the nodes corresponding to larger components tend to form *quasicliques*. That is, clusters in which every two nodes tend to interact with each other. In the second case, the nodes tend to form *quasibipartites*, i.e., nodes are partitioned into *almost*

disjoint subsets with high connectivity between sets but low internal connectivity.

Negative values of β arise when the absolute temperature is negative ($T < 0$). Note that the temperature scale from cold to hot then runs [21] $0 \text{ K}, \dots, +300 \text{ K}, \dots, +\infty \text{ K}, \dots, -\infty \text{ K}, \dots, -300 \text{ K}, \dots, -0 \text{ K}$. In the limit $T \rightarrow -0$ the largest contribution to the communicability is from the lowest eigenvalue of the adjacency matrix λ_n ,

$$\lim_{\beta \rightarrow -\infty} G_{pq}(\beta) \approx \phi_n(p)\phi_n(q)e^{-|\beta|\lambda_n}, \quad (5)$$

which is known to produce a two-coloring of the nodes [22]. In the above-mentioned example of the cycle C_n , the quantity (5) is positive when p and q have the same parity (that is, when they belong to the same subgraph) and negative when not. This implies that the sign of the communicability at a negative temperature indicates whether or not two nodes belong to the same subgraph. This is the main observation on which we develop the theoretical approach hereafter.

In order to understand the meaning of the inverse temperature β in the context of complex networks, we may express the communicability in terms of powers of the adjacency matrix,

$$(e^{\beta A})_{pq} = \sum_{k=0}^{\infty} \frac{\beta^k (A^k)_{pq}}{k!}. \quad (6)$$

Accordingly, β represents a weight given to every link of the network. This weight accounts for the ‘‘strength’’ of the interaction between the corresponding nodes in the graph. For instance, $\beta = 0$, which corresponds to the limit $T \rightarrow \infty$, corresponds to a graph with no links. This case is similar to a gas formed by monoatomic particles. On the other hand, very large values of β in the limit $T \rightarrow +0$ represent very large attractive interactions between pairs of bonded nodes in a similar manner to a solid. The new cases considered in this work, $\beta < 0$, correspond to the existence of repulsive interactions between the pairs of linked nodes, which obligates them to be in separated clusters forming bipartite structures in the network.

From now on we consider, for the sake of simplicity, the case where $\beta = -1$. Then,

$$G_{pq}(\beta = -1) = (e^{-A})_{pq} = \sum_{j=1}^n \phi_j(p)\phi_j(q)e^{-\lambda_j}. \quad (7)$$

Now, let us interpret the exponential negative adjacency matrix. First, we expand it in powers of the adjacency matrix,

$$e^{-A} = I - A + \frac{A^2}{2!} - \frac{A^3}{3!} + \dots, \quad (8)$$

which can be expressed in terms of the hyperbolic functions as

$$e^{-A} = \cosh(A) - \sinh(A). \quad (9)$$

The term $[\cosh(A)]_{pq}$ represents the weighted sum of the number of walks of even length connecting nodes p and q in the network. Similarly, $[\sinh(A)]_{pq}$ represents the weighted

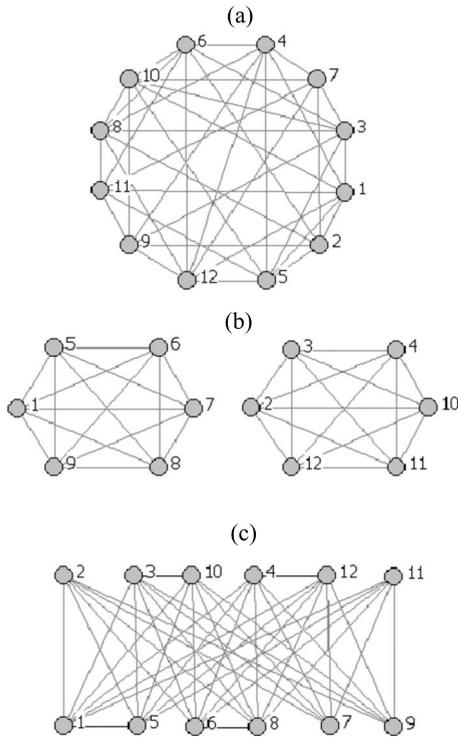


FIG. 1. (a) Quasibipartite graph having 12 nodes which was built from a complete bipartite graph with two disjoint sets of 6 nodes each (see main text). (b) The node-repulsion graph corresponding to the quasibipartite graph shown in Fig. 1(a). (c) Organization of the nodes of the graph given in Fig. 1(a) to represent the two quasibipartite clusters found by the method developed in this work. The black lines represent the intracluster connections and the gray lines the intercluster links.

sum of the number of walks of odd length connecting nodes p and q .

Let us consider a bipartite graph and let p and q be nodes which are in two different disjoint sets of the graph. Then, there are no walks of even length between p and q in the graph and

$$G_{pq}(\beta = -1) = [-\sinh(A)]_{pq} < 0. \quad (10)$$

However, if p and q are nodes in the same disjoint set, then there is no walk of odd length connecting them due to the lack of odd cycles in the bipartite graph, which makes

$$G_{pq}(\beta = -1) = [\cosh(A)]_{pq} > 0. \quad (11)$$

The above argument shows that, in general, the sign of the communicability at a negative temperature, $G_{pq}(\beta = -1) = (e^{-A})_{pq}$, gives an indication as to how the nodes can be separated into disjoint sets.

Strategy for detecting quasibipartite clusters

The relations (10) and (11) hold when the graph is bipartite. However, the arguments concerning odd vs even length walks carry through to the case of quasibipartite subgraphs. A bipartite subgraph (bipartite cluster) is a subset of nodes of the graph that can be divided into two disjoint sets. In a

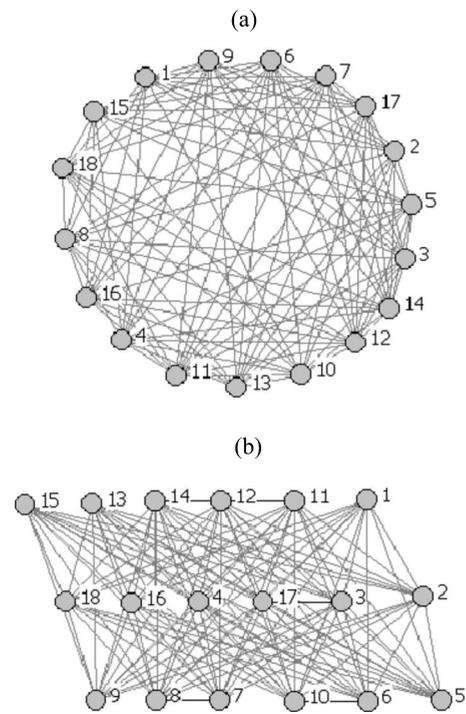


FIG. 2. (a) Quasitripartite graph having 18 nodes which was built from a complete tripartite graph with three disjoint sets of 6 nodes each (see main text). (b) Organization of the nodes of the graph given in Fig. 2(a) to represent the three quasitripartite clusters found by the method developed in this work. The black lines represent the intracluster connections and the gray lines the intercluster links.

quasibipartite subgraph we have a subset of nodes that can be divided into two *almost disjoint* sets. The term “almost disjoint” means that most of the links in the subgraph are intersets links but there are very few intraset links. In more formal terms we can define a quasibipartite cluster of nodes as follows.

Definition 1. Let $C \subseteq V$ be a cluster of nodes in the network. Then, C is a quasibipartite cluster if, and only if, $[\cosh(A)]_{pq} > [\sinh(A)]_{pq} \forall p, q \in C$.

Our strategy for detecting quasibipartite clusters in complex networks is as follows. First we start by calculating $\exp(-A)$, whose (p, q) entry gives the communicability at negative temperature between the nodes p and q in the network. We recall that the positive entries of this matrix, i.e., $G_{pq} > 0$, correspond to pairs of nodes which are in the same quasibipartite cluster. The matrix $\exp(-A)$ can be represented as a signed complete graph in which every link connecting a pair of nodes has positive ($G_{pq} > 0$) or negative ($G_{pq} < 0$) signs. We recall that a signed graph is a graph whose edges are labeled by signs. In a signed graph a positive clique is a maximal set of mutually adjacent vertices in which every pair of nodes is linked by a positive edge. That is, a positive subgraph B is said to be a positive clique if $G_{pq} > 0 \forall p, q \in B$. Then a *quasibipartite cluster* in the network is a positive clique. A clique is a maximal complete subgraph and a complete subgraph is a part of a graph in which all nodes are connected to each other.

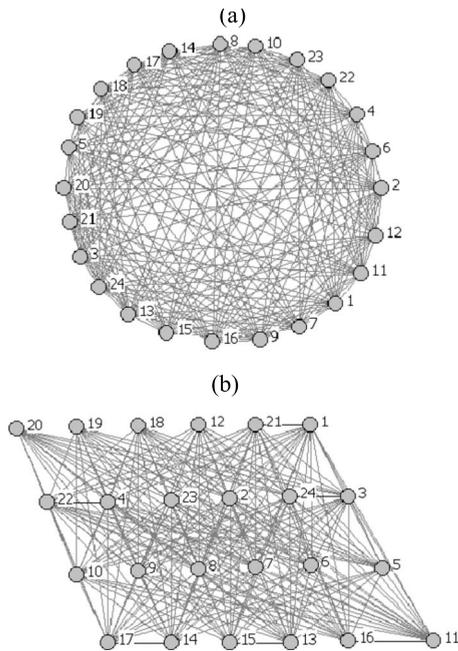


FIG. 3. (a) Quasitrapartite graph having 24 nodes which was built from a complete tetrapartite graph with three disjoint sets of 6 nodes each (see main text). (b) Organization of the nodes of the graph given in Fig. 3(a) to represent the four quasitrapartite clusters found by the method developed in this work. The black lines represent the intracluster connections and the gray lines the intercluster links.

In order to account for the inter- to intraset proportion of links in the detection of quasibipartite clusters we introduce a Heaviside step function:

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}.$$

Definition 2. The *node-repulsion graph* is a graph whose adjacency matrix is given by $\Theta[\exp(-A)]$, which results from the elementwise application of the function $\Theta(x)$ to the matrix $\exp(-A)$. A pair of nodes p and q in the node-repulsion graph $\Theta[\exp(-A)]$ is connected if, and only if, they have $G_{pq} > 0$.

Now, suppose that there is a link between the nodes p and q and there are also links between them and a third node r . This means that $G_{pq} > 0$, $G_{pr} > 0$, and $G_{qr} > 0$. Consequently, the three nodes form a positive subgraph B . As we want to detect the largest subset of nodes connected to this triple we have to search for the nodes s for which $G_{is} > 0 \forall i \in B$. Using the *node-repulsion graph*, this search is reduced to finding the cliques in a simple graph, $\Theta[\exp(-A)]$. These cliques correspond to the quasibipartite clusters of the network.

Finding the cliques in a graph is a classical problem in combinatorial optimization, which has found applications in diverse areas [23]. Here we use a well-known algorithm due to Bron and Kerbosch [24], which is a depth-first search for generating all cliques in a graph. This algorithm consumes a time per clique which is almost independent of the graph size for random graphs and for the Moon-Moser graphs of n ver-

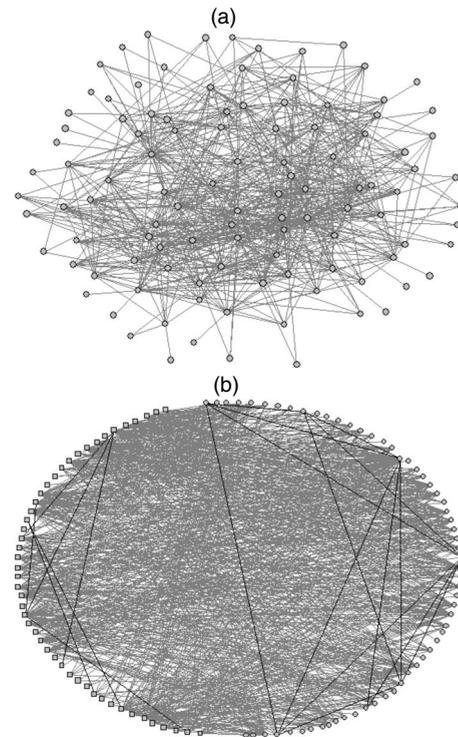


FIG. 4. (a) Network representation of the food web of Canton Creek. (b) Bipartite structure of this network as found by the method developed in this work. Nodes in each quasibipartite cluster are represented by squares and circles. The black lines represent the intracluster connections and the gray lines the intercluster links.

tices the total time is proportional to $(3.14)^{n/3}$. The Moon-Moser graphs have the largest number of maximal cliques possible among all n -vertex graphs regardless of the number of edges in the graph [25].

Our algorithm differs from those in [9,10] in a number of ways. Fundamentally, our aim is different. Rather than quantifying the overall bipartivity of the network, or of individual nodes or edges, we are looking for communities that share in a bipartite substructure. Moreover, we allow several such substructures to be present. In contrast to [9], this algorithm takes account of both odd and even length walks, avoids the need for a cutoff parameter by considering walks of all possible lengths, and does not require a complex energy landscape to be searched by a heuristic discrete optimization algorithm. This approach differs from [10] in that the *difference* between odd and even length walks is used, and we consider walks between distinct pairs of nodes rather than closed walks, allowing communities to be detected.

COMPUTATIONAL TESTS

In order to test our strategy to find quasibipartite clusters in complex networks, we start with a small quasibipartite graph. The graph was formed by placing 12 nodes into two disjoint sets of 6 nodes each (a bipartite graph) and then connecting at random some of the nodes in each disjoint set, which creates a quasibipartite structure. The graph has average degree $\langle k \rangle = 6.67$. We apply the Bron and Kerbosch (BK)

algorithm to the 0,1-adjacency matrix of the node-repulsion graph, $\Theta[\exp(-A)]$, to identify the cliques, which correspond to the quasibipartite clusters in the original graph.

In Fig. 1(a), we illustrate the graph in such a way that the quasibipartite structure is not visually apparent. This would be the case with a typical graph drawing algorithm. The node-repulsion graph $\Theta[\exp(-A)]$, which consists of two isolated components, is illustrated in Fig. 1(b). Each of these components is formed by a complete graph of 6 nodes. Then, it is evident that the BK algorithm identifies these two cliques as the only ones existing in the node-repulsion graph $\Theta[\exp(-A)]$, which indeed correspond to the two quasibipartite clusters of the graph, as illustrated in Fig. 1(c). The two almost disjoint sets are represented in two lines of nodes where the intercluster links are represented in gray and the intracluster ones in black. The intracluster links make the graph quasibipartite rather than perfectly bipartite.

Next, we create two new graphs using a similar approach. The first is a quasitripartite graph and the second is a quasitetrapartite graph. The application of the current approach clearly divides the first graph into three quasipartite clusters and the second one into four. The graphs and their partitions are illustrated in Figs. 2 and 3, respectively.

MULTIPARTITE STRUCTURE IN REAL-WORLD NETWORKS

As a proof of concept we first select a network which we already know is bipartite. It is the network of heterosexual contacts obtained empirically at the Cadham Provincial Laboratory during six months between November 1997 and May 1998 [26]. This network, consisting of 82 nodes and 84 connections, was studied by Lind *et al.* [27] where illustrations and details can be found. Our current approach clearly identifies the two bipartite clusters, one consisting of 47 nodes and the other of 35 (results not shown). The node-repulsion graph clearly identified the two isolated components.

As a second example, we studied the food web of Canton Creek, which consists primarily of invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand [28]. This network consists of 108 nodes (species) and 707 links (trophic relations). Using our current approach, we find that this network can be divided into two almost-bipartite clusters, one having 66 nodes and the other 42. Only 20 links connect nodes in the same clusters, 13 of them connect nodes in the set containing 66 nodes and the other 7 connect nodes in the set of 42 nodes. Thus 97.2% of links are connections between the two almost-bipartite clusters and only 2.8% links are intracluster connections. In Fig. 4, we illustrate the network and its quasibipartite clusters as found in the current work. Other food webs (see [29] and references therein), like that of the pelagic species from the largest of a set of 50 New York Adirondack lake food webs (Bridge Brook), a marine ecosystem on the northeast US shelf (Shelf), invertebrates in an English pond (Skipwith), and a food web like Canton Creek but in native tussock habitat (Stony stream) are also formed by two

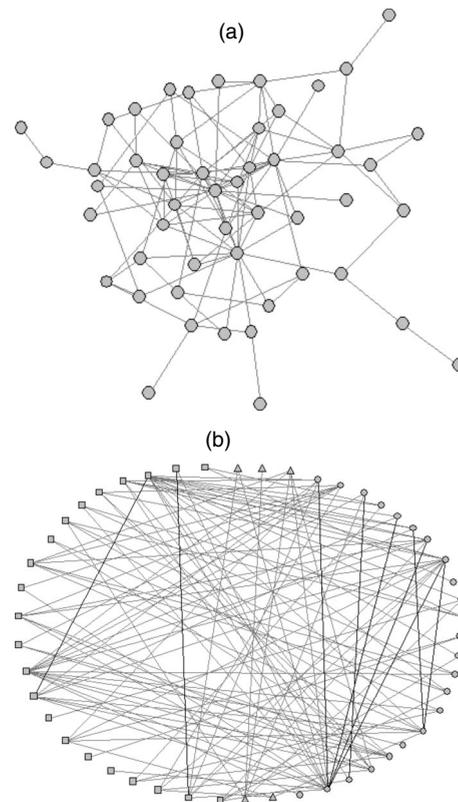


FIG. 5. (a) Network representation of the protein-protein interaction network of the Kaposi sarcoma-related herpes virus (KSHV). (b) Bipartite structure of this network as found by the method developed in this work. Nodes in each quasibipartite cluster are represented by squares and circles. Triangles correspond to nodes not in these two quasibipartite groups (see text for explanation). The black lines represent the intracluster connections and the gray lines the intercluster links

main quasibipartite clusters with no overlap between them. However, there are other food webs with a larger number of quasibipartite clusters with large overlap among them. One example is the network formed by birds and predators and arthropod prey of *Anolis* lizards on the island of St. Martin (see [29] and references therein), located in the northern Lesser Antilles (St. Martins), which has 116 quasipartite clusters.

The next example corresponds to the protein-protein interaction network (PIN) of the Kaposi sarcoma-associated herpes virus (KSHV) [30]. KSHV is a member of the γ -herpes virus subfamily associated with Kaposi sarcoma and B-cell lymphomas. Its PIN was generated by Uetz *et al.* [30] by testing 12 000 viral protein interactions involving both full-length proteins and protein fragments. From this pool of interactions, Uetz *et al.* [30] identified 123 nonredundant interacting pairs of proteins, 8 of which were self-interactions. The resulting PIN of KSHV, formed by 50 proteins and 115 interactions, is illustrated in Fig. 5(a). Some of the global topological characteristics of this PIN can be found in Uetz *et al.* [30].

Using our current approach we identify 34 quasipartite clusters in the PIN of KSHV. The proteins grouped in every cluster are given in Table I. The largest clusters are the num-

TABLE I. Quasibipartite clusters found in the protein-protein interaction network of the Kaposi sarcoma-related herpes virus (KSHV). The numbers are identification labels for the clusters and the proteins which form the cluster are given in the second column.

No.	Proteins
1	K1 K15 K3 K5 K8 K8.1 Orf41 Orf52 Orf54 Orf65 Orf67.5 Orf68 Orf74 Orf56 K12 Orf48 Orf61
2	K8.1 Orf41 Orf54 Orf65 Orf67.5 Orf68 K12 Orf48 Orf61 Orf39 Orf47
3	Orf41 Orf65 K12 Orf48 Orf61 Orf39 Orf47 Orf2
4	Orf48 Orf61 Orf39 Orf47 Orf45 Orf2 Orf28 Orf37 Orf49
5	Orf56 Orf48 Orf61 Orf45
6	K1 K15 K3 K5 K8 K8.1 Orf23 Orf41 Orf50 Orf52 Orf54 Orf65 Orf67.5 Orf68 Orf74 Orf56 K12 Orf61
7	K8.1 Orf41 Orf50 Orf54 Orf65 Orf67.5 Orf68 K12 Orf61 Orf39 Orf47
8	Orf61 Orf39 Orf47 Orf45 Orf2 Orf27 Orf28 Orf31 Orf37 Orf49 Orf59
9	Orf61 Orf45 Orf2 Orf27 Orf28 Orf31 Orf34 Orf37 Orf49 Orf59 Orf69
10	K1 K10 K11 K15 K3 K5 K8 K8.1 Orf23 Orf30 Orf41 Orf50 Orf52 Orf54 Orf58 Orf65 Orf67.5 Orf68 Orf72 Orf74 Orf56
11	K1 K10 K10.5 K11 K15 K3 K5 K8 K8.1 Orf23 Orf30 Orf41 Orf50 Orf52 Orf54 Orf58 Orf65 Orf67.5 Orf68 Orf72 Orf74
12	K10 K10.5 K11 Orf23 Orf50 Orf72 Orf36
13	K10 Orf36 Orf25
14	K1 K10.5 K11 K15 K3 K5 K8 K8.1 Orf23 Orf30 Orf41 Orf50 Orf52 Orf54 Orf58 Orf65 Orf67.5 Orf68 Orf72 Orf74 K12
15	K10.5 K8.1 Orf41 Orf50 Orf54 Orf65 Orf67.5 Orf68 K12 Orf39 Orf47
16	K1 K10.5 K11 K15 K3 K5 K8.1 Orf30 Orf41 Orf52 Orf54 Orf65 Orf67.5 Orf68 Orf74 K12 K9
17	K10.5 K8.1 Orf41 Orf54 Orf65 Orf67.5 Orf68 K12 K9 Orf39 Orf47
18	K1 K11 K15 K3 K5 K8 K8.1 Orf23 Orf30 Orf41 Orf50 Orf52 Orf54 Orf58 Orf65 Orf67.5 Orf68 Orf72 Orf74 Orf56 K12
19	K1 K11 K15 K3 K5 K8.1 Orf30 Orf41 Orf52 Orf54 Orf65 Orf67.5 Orf68 Orf74 Orf56 K12 K9
20	K1 K15 K3 K5 K8 K8.1 Orf30 Orf41 Orf52 Orf54 Orf58 Orf65 Orf67.5 Orf68 Orf74 Orf56 K12 Orf48
21	K1 K15 K3 K5 K8.1 Orf30 Orf41 Orf52 Orf54 Orf65 Orf67.5 Orf68 Orf74 Orf56 K12 K9 Orf48
22	K8.1 Orf41 Orf54 Orf65 Orf67.5 Orf68 K12 K9 Orf48 Orf39 Orf47
23	Orf41 Orf65 K12 K9 Orf48 Orf39 Orf47 Orf2
24	K9 Orf48 Orf39 Orf47 Orf45 Orf2 Orf37
25	Orf56 K9 Orf48 Orf45
26	K9 Orf45 Orf2 Orf37 Orf57
27	Orf36 Orf25 K7 Orf27 Orf29b Orf31 Orf34 Orf53 Orf57 Orf59 Orf6 Orf60 Orf62 Orf63 Orf69 Orf9
28	Orf25 Orf45 K7 Orf27 Orf28 Orf29b Orf31 Orf34 Orf37 Orf49 Orf53 Orf57 Orf59 Orf6 Orf60 Orf63 Orf69 Orf75 Orf9
29	Orf25 K7 Orf27 Orf28 Orf29b Orf31 Orf34 Orf37 Orf49 Orf53 Orf57 Orf59 Orf6 Orf60 Orf62 Orf63 Orf69 Orf75 Orf9
30	Orf45 K7 Orf2 Orf27 Orf28 Orf29b Orf31 Orf34 Orf37 Orf49 Orf53 Orf57 Orf59 Orf6 Orf60 Orf63 Orf69 Orf75 Orf9
31	Orf39 Orf47 Orf45 K7 Orf2 Orf27 Orf28 Orf31 Orf37 Orf49 Orf53 Orf59 Orf6 Orf60
32	K7 Orf2 Orf27 Orf28 Orf29b Orf31 Orf34 Orf37 Orf49 Orf53 Orf57 Orf59 Orf6 Orf60 Orf62 Orf63 Orf69 Orf75 Orf9
33	Orf48 Orf45 Orf2 Orf28 Orf37 Orf49 Orf75
34	Orf56 Orf48 Orf45 Orf75

bers 10, 11, 14, and 18 which have 21 proteins each. However, there is a very large overlap among them ranging from 66.6% for clusters 11 and 18 to 95% for the pairs of clusters (10, 11), (11, 14), (14, 18). There is another group of quasibipartite clusters containing a large number of proteins. They are the clusters 28, 29, 30, and 32. They also display very large overlapping among them, ranging from 89.5% to 94.7%. However, these two groups of clusters are completely orthogonal. That is, absolutely no overlapping exists between any of the clusters of the first group (10, 11, 14, and 18) with the clusters in the second group (28, 29, 30, and 32). Then we conclude that the PIN of KSHV can be divided into two

disjoint clusters of almost the same size, which contain 78% of the proteins in the PIN. These two clusters are illustrated in Fig. 5(b).

On the other side of the coin, there are networks displaying a huge number of small quasibipartite clusters. This is the case for those networks lacking a bipartite structure at all but having a multipartite structure. For instance, the neuronal synaptic network of the nematode *C. elegans* (see [31] and the references therein), which has 280 nodes and 1973 links, is formed by 43 753 quasipartite clusters. This network has been formerly shown to have a superhomogeneous structure [32], which explains its lack of bipartivity.

CONCLUSIONS

Given a complex network, the concept of a node-repulsion graph has intuitive interpretations in terms of (i) a Green's function at negative absolute temperature, and (ii) a measure of the discrepancy between the overall number of odd and even walks between pairs of nodes. Moreover, this concept allows for a natural, well-defined quantification of quasibipartite clusters that can be investigated with a simple, parameter-free computational algorithm. This algorithm was able to discover inherent bipartite communities in real data

sets, and hence has the potential to unlock hidden patterns at the heart of complex networks.

ACKNOWLEDGMENTS

E.E. acknowledges the support of the Royal Society of Edinburgh and the Edinburgh Mathematical Society (March 2008), and the IIS, University of Tokyo, for financial support. D.J.H. was supported by Engineering and Physical Sciences Research Council Grant No. GR/S62383/01.

-
- [1] M. Girvan and M. J. E. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
- [2] J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005).
- [3] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, *Nature (London)* **435**, 814 (2005).
- [4] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, *Physica A* **352**, 669 (2005).
- [5] R. Guimera and L. A. N. Amaral, *Nature (London)* **433**, 895 (2005).
- [6] M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
- [7] M. E. J. Newman, *Eur. Phys. J. B* **38**, 321 (2004).
- [8] L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas, *J. Stat. Mech.: Theory Exp.* (2002) P09008.
- [9] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, *Phys. Rev. E* **68**, 056107 (2003).
- [10] E. Estrada and J. A. Rodríguez-Velázquez, *Phys. Rev. E* **72**, 046105 (2005).
- [11] A. Thomas, R. Cannings, N. A. M. Monk, and C. Cannings, *Biochem. Soc. Trans.* **31**, 1491 (2003).
- [12] J. L. Morrison, R. Breitling, D. J. Higham, and D. R. Gilbert, *Bioinformatics* **22**, 2012 (2006).
- [13] E. Estrada, *J. Proteome Res.* **5**, 2177 (2006).
- [14] E. Estrada and N. Hatano, *Phys. Rev. E* **77**, 036111 (2008).
- [15] N. L. Biggs, *Algebraic Graph Theory* (Cambridge University Press, Cambridge, England, 1993).
- [16] E. Estrada and N. Hatano, *Chem. Phys. Lett.* **439**, 247 (2007).
- [17] E. Estrada and N. Hatano (unpublished).
- [18] A. J. Seary and W. D. Richards, Jr., in *Proceedings of the International Conference on Social Networks*, edited by M. G. Everett and K. Rennolds (Greenwich University Press, London, 1995), Vol. 1, p. 47.
- [19] T. Xiao and S. Gong, *Pattern Recogn.* **41**, 1012 (2008).
- [20] E. Estrada, *Phys. Rev. E* **75**, 016103 (2007).
- [21] N. F. Ramsey, *Phys. Rev.* **103**, 20 (1956).
- [22] B. Aspvall and J. R. Gilbert, *SIAM J. Algebraic Discrete Methods* **5**, 526 (1984).
- [23] I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo, in *Handbook of Combinatorial Optimization*, edited by D.-Z. Du and P. M. Pardalos, Supplement Vol. A. (Kluwer Academic Publishers, Dordrecht, 1999), pp. 1–74.
- [24] C. Bron and J. Kerbosch, *Commun. ACM* **16**, 575 (1973).
- [25] J. W. Moon and L. Moser, *Isr. J. Math.* **3**, 23 (1965).
- [26] J. L. Wylie and A. Jolly, *Sex Transm. Dis.* **28**, 14 (2001).
- [27] P. G. Lind, M. C. González, and H. J. Herrmann, *Phys. Rev. E* **72**, 056127 (2005).
- [28] C. Townsend, R. M. Thompson, A. R. McIntosh, C. Kilroy, E. Edwards, and M. R. Scarsbrook, *Ecol. Lett.* **1**, 200 (1998).
- [29] J. A. Dunne, R. J. Williams, and N. D. Martinez, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12917 (2002).
- [30] P. Uetz *et al.*, *Science* **311**, 239 (2006).
- [31] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [32] E. Estrada, *Europhys. Lett.* **73**, 649 (2006).