

EPJ B

Condensed Matter
and Complex Systems

EPJ.org

your physics journal

Eur. Phys. J. B (2012) 85: 26

DOI: 10.1140/epjb/e2011-20129-7

Community overlays upon real-world complex networks

X. Ge and H. Wang



Community overlays upon real-world complex networks

X. Ge¹ and H. Wang^{2,a}

¹ DaLian Maritime University, School of Information Science and technology, DaLian 116026, P.R. China

² Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, 2628 CD, Delft, The Netherlands

Received 21 February 2011 / Received in final form 25 July 2011

Published online 18 January 2012 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2012

Abstract. Many networks are characterized by the presence of communities, densely intra-connected groups with sparser inter-connections between groups. We propose a community overlay network representation to capture large-scale properties of communities. A community overlay G_o can be constructed upon a network G , called the underlying network, by (a) aggregating each community in G as a node in the overlay G_o ; (b) connecting two nodes in the overlay if the corresponding two communities in the underlying network have a number of direct links in between, (c) assigning to each node/link in the overlay a node/link weight, which represents e.g. the percentage of links in/between the corresponding underlying communities. The community overlays have been constructed upon a large number of real-world networks based on communities detected via five algorithms. Surprisingly, we find the following seemingly universal properties: (i) an overlay has a smaller degree-degree correlation than its underlying network $\rho_o(D_{l+}, D_{l-}) < \rho(D_{l+}, D_{l-})$ and is mostly disassortative $\rho_o(D_{l+}, D_{l-}) < 0$; (ii) a community containing a large number W_i of nodes tends to connect to many other communities $\rho_o(W_i, D_i) > 0$. We explain the generic observation (i) by two facts: (1) degree-degree correlation or assortativity tends to be positively correlated with modularity; (2) by aggregating each community as a node, the modularity in the overlay is reduced and so is the assortativity. The observation (i) implies that the assortativity of a network depends on the aggregation level of the network representation, which is illustrated by the Internet topology at router and AS level.

1 Introduction

A large system of elements (nodes) and their interactions or relations (links) can be represented as a network, a set of nodes interconnected by a set of links. Examples of complex networks range from biological networks and communication networks to social networks. The characterization of networks has been extensively investigated to understand the effect of the network structure on its functioning [1,2]. Graph theoretical analysis of complex networks has taken a large flight with the discovery of small-world properties [3] and the power-law degree distribution [4] in the late nineties of the last century. Most studied properties focus on the properties of individual nodes or node pairs such as the distance between a node pair and the similarity of two connected nodes in degree.

However, it is equally important to examine the large-scale properties of a network as a whole. It has been shown that many complex networks are characterized by the presence of communities or modules, densely connected groups of nodes with sparser connections between groups.

Algorithms to detect non-overlapping communities have been extensively studied aiming to improve the partitioning quality, to reduce the computational complexity as well as to incorporate extra community related information [5–8]. One fascinating observation is that community sizes, which is the number of nodes in a community, follow a power law distribution [9,10], when the network is partitioned to maximize the modularity, an measure proposed by Newman [11] to quantify the quality of a partitioning. Statistical properties of overlapping communities has been studied in e.g. [12]. Moreover, a network can be tiled instead of with communities but with boxes, where all nodes within a box are connected by a minimum distance smaller than a given l_B . Large-scale properties at the level of boxes have been studied in [13].

Many studies have focused on algorithms to detect community structures and basic community related properties such as community sizes. In this paper, we propose to systematically study large-scale properties of complex networks on a higher hierarchical level: the community overlay. A community overlay network can be constructed as shown in Figure 1: first, non-overlapping communities can be detected via partitioning algorithm in the origi-

^a e-mail: h.wang@tudelft.nl

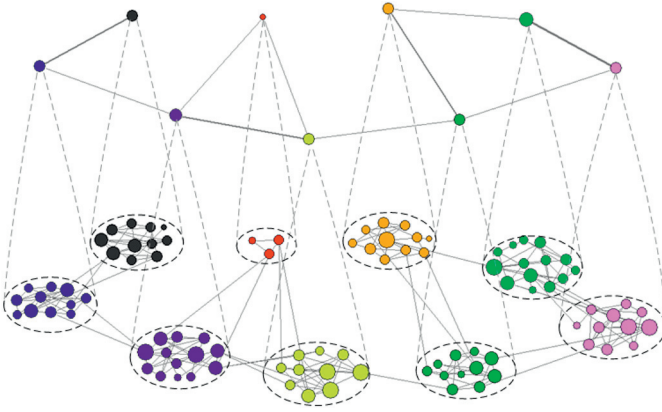


Fig. 1. (Color online) A network G with $N = 95$ nodes and $L = 213$ links and its community overlay network G_o . Nodes with the same color belong to the same community. In the overlay network, the size of a node is proportional to its node weight $w_i = n_i/N$ and the width of a link is proportional to its link weight $w_{ij} = l_{ij}/L$.

nal network, called the underlying network, $G(N, L)$ with N nodes and L links. Each community is generally a subgraph with a set of community members densely interconnected by a set of links. Furthermore, we condense each community in the underlying network as a node in the community overlay network G_o . Two nodes i and j in the overlay G_o are connected if the underlying network has a positive number $l_{ij} > 0$ of links with two end nodes residing in the corresponding two communities respectively. We assign to each link (i, j) in the overlay G_o a link weight $w_{ij} = l_{ij}/L$, and similarly, to each node i a node weight $w_i = n_i/N$, which is the percentage of nodes residing in the underlying community i . The node weight can be also any other topological measure of the corresponding community. In this way, the community overlay network G_o captures network features at the community level. It is a condensed network expression with both link and node weights. We consider non-overlapping instead of overlapping communities as the nodes of the overlay, since analytical approach tends to be feasible when nodes in a network are as independent as possible.

In this paper, we will investigate whether universal properties exist in the community overlays upon real-world complex networks. The second goal is to understand the relation between characteristics of the underlying network and of its community overlay. This may explain the general features observed in the community overlay networks, if there are.

2 Algorithms to detect communities

Finding community structure in complex networks is important to grasp inherent properties of complex networks, which can be reflected by the community overlay. However, some properties observed in the community overlay can be introduced by the intrinsic mechanism of a com-

munity detection algorithm. In order to avoid this, we select five algorithms based on different principles. We do not specify the number of communities in advance nor use extra community related information such as the node type in protein network. Hence, we select algorithms that detect the community structure according to the natural topological structure.

Table 1 lists the five algorithms we adopt to detect communities. The modularity M , proposed by Newman and Girvan [14], is a measure of the quality of a particular division of the network, which is defined in [11] as

$$M = \frac{1}{2L} \sum_{i=1}^N \sum_{j=1}^N \left(a_{ij} - \frac{d_i d_j}{2L} \right) 1_{\{i \text{ and } j \text{ belong to the same community}\}} \quad (1)$$

where a_{ij} is the matrix element of the adjacency matrix¹ A of the graph and d_i is the degree (the number of neighbors) of node i . The modularity is proportional to the number of links falling within clusters or groups minus the expected number in an equivalent network with links placed at random. Thus, if the number of links within a group is no better than random, the modularity is zero. A modularity approaching one reflects networks with strong community structure: a dense intra-group and a sparse inter-group connection pattern. If links are placed at random, then the expected number of links between node i and node j equals $\frac{d_i d_j}{2L}$, where d_j is the degree of node j . The motivation of algorithm cnm [15] is to partition a network into communities with maximum modularity.

Algorithm btw [16] computes community structure in graphs based on link betweenness. The betweenness of a link (i, j) is defined as the number of shortest paths between all pairs of nodes which run through (i, j) . If a network contains communities or groups that are only loosely connected by a few inter-group links, then all shortest paths between different communities traverse one of these few links. Thus, the links connecting communities will have a high link betweenness. By removing these links, the network can be separated into different communities.

The idea of vot algorithm [8] is that each link is regarded as a resistor with the same resistance, and we can connect a battery between two arbitrary nodes so that each node has its fixed voltage, thereby building a voltage difference. Since nodes inside a community are densely connected, their voltages tend to be close. A big voltage gap happens about halfway between the two communities, where the links are sparse and the local resistance is large. Therefore, nodes can be divided into different communities according to a voltage threshold.

Algorithm mrw [17] introduces an information theoretic approach that reveals community structure in weighted and directed networks to comprehend the multipartite organization of large-scale biological and social systems. The result is a map that both simplifies

¹ Each element a_{ij} is either 1 or 0 depending on whether node i and j are connected or not.

Table 1. Community partitioning algorithms.

Label (abbreviation)	Running time	Granularity (number of modules detected)	Ref.
CNM (cnm)	$O(N \log^2 N)$	1st largest number of modules	[15]
EdgeBetweenness (btw)	$O(kmN)$	3rd largest number of modules	[16]
Voltage (vot)	$O(N + L)$	5th largest number of modules	[8]
MRW (mrw)	$O(N \log N)$	4th largest number of modules	[17]
h-cnm	NA	2nd largest number of modules	[18]

and highlights the regularities in the structure and their relationships.

Algorithm h-cnm [18] is based on the cnm algorithm. It attempts to merge community structures in a balanced manner which dramatically speeds up the computation and slightly improves the modularity.

3 Network metrics

We will examine properties of real-world networks as well as their corresponding community overlay networks mainly via assortativity and modularity defined as (1). Since the number of communities in real-world networks, equivalently, the size of their community overlays differ from network to network, we cannot simply compare community overlays of different sizes by graph metrics such as the average distance between two nodes. Therefore, we study the metrics which are independent of the network size. “Mixing” in complex networks [19] refers to the tendency of network nodes to connect preferentially to other nodes with either similar or opposite properties. Mixing is computed [20] via the correlations between the properties of connected node pairs, such as the degree-degree correlation

$$\rho(D_{l+}, D_{l-}) = 1 - \frac{\sum_{i \sim j} (d_i - d_j)^2}{\sum_{i=1}^N d_i^3 - \frac{1}{2L} \left(\sum_{i=1}^N d_i^2 \right)^2} \quad (2)$$

where d_i is the degree (number of neighbors) of node i , $i \sim j$ represents the condition that i and j are connected, N and L are the number of nodes and links respectively. Networks, where nodes preferentially connect to nodes with (dis)similar property, are called (dis)assortative. For example, networks, where high degree nodes preferentially connect to other high degree nodes, are assortative in degree correlation $\rho(D_{l+}, D_{l-}) > 0$, whereas networks, where high degree nodes connect to low-degree nodes, are disassortative $\rho(D_{l+}, D_{l-}) < 0$. Such a correlation, called assortativity, will be examined based on the degree as well as on the node weight (e.g. the size of the underlying community) correlation in the overlay network.

One fundamental link weight related measure [21] is the link weight correlation of links incident to a node. The link weight correlation examines whether links connected to a same node tend to possess similar or dissim-

ilar link weights. Ramasco and Gonçalves [22] have proposed a measure that examines the ratio of the average link weight variance around each node to that of an ensemble of weight-reshuffled instances of the original graph. For example, the variance of the link weight around a node i can be defined as

$$\sigma_w^2(i) = \sum_{j \in \mathcal{N}(i)} \left(w_{ij} - \frac{\sum_{j \in \mathcal{N}(i)} w_{ij}}{d_i} \right)^2$$

where $\mathcal{N}(i)$ is the set of neighboring nodes of i and $\frac{\sum_{j \in \mathcal{N}(i)} w_{ij}}{d_i}$ is, thus, the average link weight of the links arriving at i . The link weight correlation is then measured as

$$\Delta_w = \frac{E_{org}[\sigma_w]}{E_{rand}[\sigma_w]} \quad (3)$$

where the average standard deviation of link weights around each node $E[\sigma_w]$ is estimated for the original graph and an ensemble of weight-reshuffled² instances. The type of link weight correlation around each node in a network is revealed by comparing with the randomized instances: positive ($\Delta_w < 1$), negative ($\Delta_w > 1$) or non-correlated ($\Delta_w = 1$).

4 Characterizing the community overlays

In this section, we examine features of the large set of real-world networks and of their community overlays. We have collected a large data set of real-world networks which represent the topology of various complex systems. Most of the data sets, that we have used, are available publicly. They are complex networks from a wide range of systems in nature and society. A network is connected if there exists a path between each pair of nodes. We consider only the networks formed by the largest connected component of our real-world networks. Furthermore, we also generate a network from each classical network model including the Barabási-Albert power law graph³, the Erdős-Rényi

² The set of L link weights are re-assigned randomly to the set of L links.

³ It starts with m nodes. At every time step, we add a new node with m links that connect the new node to m different nodes already present in the graph. The probability that a new node will be connected to node i in step t is proportional to the degree $d_i(t)$ of that node.

Table 2. Network data set. N is node number and L is edge number in network. $\rho(D_{l+}, D_{l-})$ and $\rho_o(D_{l+}, D_{l-})$ are the assortativity of the original network and of the community overlay respectively. (a) Network of American football games between Division IA colleges during regular season Fall 2000, as compiled by Girvan and Newman; (b) network of coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999; (c) symmetrized snapshot of the structure of the Internet at the level of autonomous systems, reconstructed from BGP tables posted at archive.routeviews.org; (d) Gnutella snapshots; (e) the direct airport-to-airport American mileage a maintained by the US Bureau of Transportation Statistics; (f) a graph representing the interconnection between cities in the Netherlands; (g) class collaborations and call graphs (nodes represent classes, links relations between them) for different software libraries; (h) the same as(g), but for Linux; (i) protien-protien interaction network in budding yeast; (j) network representing the neural network of *Elegans*; (k) Erdős-Rényi random graph; (l) Barabási-Albert power law graph; (m) word-adjacency network of a text in English; (n) power-grid infrastructure at three different levels of one city-area in Western Europe.

Classification	Network	N	L	$\rho(D_{l+}, D_{l-})$	$\rho_o(D_{l+}, D_{l-})$
Social	a America football	115	613	0.163	-0.119
	b High Energy Theory C	5835	13 815	0.185	-0.450
Technological	c Internet AS 20060722	22 963	48 436	-0.198	-0.508
	d Gnutella crawls	737	803	-0.193	-0.230
Transportation	e Air transportation	2179	31 326	-0.688	-0.688
	f Dutch road	29 663	34 982	0.246	-0.106
Software Networks	g My Sql	5285	11 352	-0.083	-0.465
	h Linux	29 663	34 982	-0.068	-0.482
Biological	i Yeast	1458	1948	-2.10	-0.175
	j C. Elegans	297	2148	-0.163	-0.253
Model	k Erdős-Rényi random graph	500	1446	-0.005	-0.333
	l Barabási-Albert power law graph	500	1500	-0.055	-0.538
Others	m Word Adjacencies English Book	7377	44 205	-0.237	-0.475
	n Western Europe Power Grid	3419	3953	-0.128	-0.214

random graph⁴ and the PFP graph⁵. The total number of networks reaches 82. These networks range from biological networks, technological networks to social networks as partly shown in Table 2. More descriptions about these complex networks can be found in⁶. Due to the computational complexity of algorithms of detecting communities, our data set does not include extremely large networks such as the social network of users registered to Youtube, which contains more than 10^6 nodes.

4.1 Community overlay

We use $G(N, L)$ to denote the original underlying network with N nodes and L links. Recall that each node

⁴ It can be generated from a set of N nodes by randomly assigning a link with probability p to each pair of nodes.

⁵ It grows Internet-like networks by using the two mechanisms. (1) Interactive growth: with probability p , a new node is attached to an old node in the existing system and two new internal links are added connecting this old node to two other old nodes; and with probability $1-p$, a new node is attached to two old nodes and one new internal link is added connecting one of the two old nodes to another old node. (2) Positive-Feedback Preference. The probability that a link is attached to an old node i with degree k_i is proportional to the degree of that node.

⁶ <http://www.nas.its.tudelft.nl/index.php/research/>
127

i in the community overlay G_o represents a community in the underlying graph and a link weight $w_{ij} = l_{ij}/L$ in G_o represents the percentage of links in the underlying network directly connecting the two corresponding communities. We may assign to each node in the overlay a node weight w_i , which can be the percentage of nodes/links or any other topological measure of the corresponding community. As an initial start, we consider only the percentage of nodes n_i/N or links l_i/L in a community as a potential node weight. Communities as defined are densely intra-connected and sparsely inter-connected. Thus, $\sum_{i=1}^{N_o} l_i = O(L)$ where N_o is the number of communities and $l_i = O(n_i^\alpha)$ where $\alpha \geq 1$. Furthermore, real-world networks are mostly sparse graphs where $L = O(N)$ and $\sum_{i=1}^{N_o} n_i = N$. Hence, $n_i = O(l_i)$, the percentage of nodes contained in a community is expected to be proportional to the percentage of links in this community. We examine four large networks. As shown in Figure 2, the number of nodes versus the number of links in a community is close to a linear function. In these cases, it is sufficient to consider only e.g. the percentage of nodes in a community as the node weight. Actually, we have considered both percentage of nodes and percentage of links in a community as the node weight. They always point to similar observations. In this paper, we only illustrate the results where the percentage of nodes in a community quantifies the node weight $w_i = n_i/N$ and $\sum_{i=1}^{N_o} w_i = 1$.

For each real-world network, we construct five community overlays based on the communities detected via five

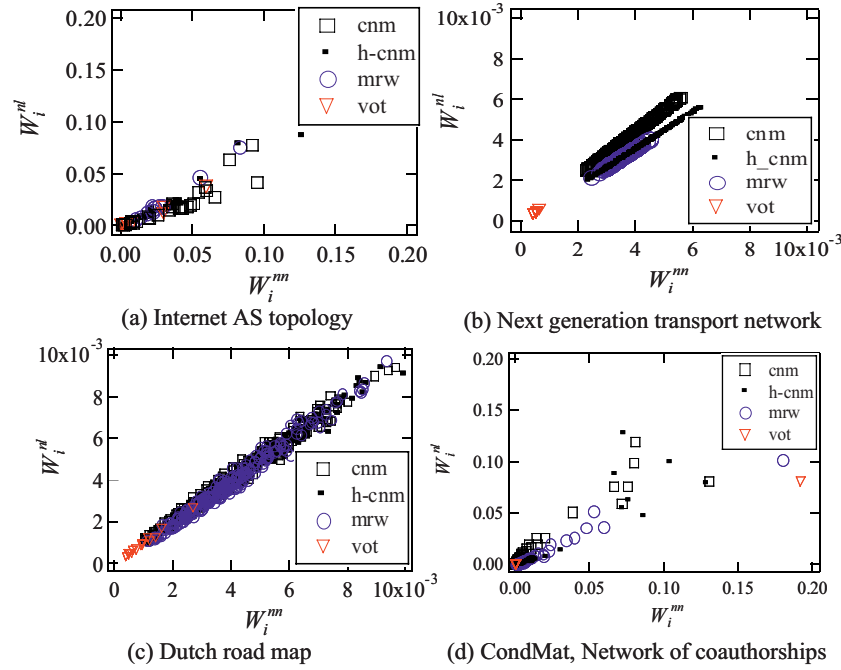


Fig. 2. (Color online) The number of links W_i^{nl} in a community versus the number of nodes W_i^{nn} in that community in four networks. Internet AS topology: $N = 22\,963$, $L = 48\,436$; next generation network: $N = 29\,902$; $L = 32\,707$; Dutch road map: $N = 32\,171$, $L = 37\,447$; CondMat, Network of coauthorships between scientists posting preprints on the Condensed Matter E-Print Archive: $N = 27\,519$, $L = 116\,181$. Algorithm cnm has been applied to detect communities.

algorithms. Community detection algorithms cnm, h-cnm, vot and mrw have been applied to all the 82 networks, while the more time consuming algorithm btw has been only used for the 44 smaller networks. We present the results of btw in the Appendix and the results of the other four algorithms in the main text so that the figures are clear.

4.2 Assortativity and modularity

Firstly, we investigate the degree-degree correlation or degree assortativity of each real-world network as well as of its 5 community overlays. As shown in Figure 3, the assortativity of the overlay is smaller than that of the underlying network $\rho_o(D_{l+}, D_{l-}) < \rho(D_{l+}, D_{l-})$. Moreover, the community overlays upon all real-world networks are mostly disassortative since $\rho_o(D_{l+}, D_{l-}) < 0$. These two observations hold surprisingly for almost all the networks that we studied, independent of the algorithms applied.

Why is the community overlay more disassortative than the underlying network $\rho_o(D_{l+}, D_{l-}) < \rho(D_{l+}, D_{l-})$? Although the influence of degree distribution on assortativity has been widely studied [23,24], this can not be explained by degree distribution alone because of the following two reasons: Firstly, the networks we considered as well as their overlays possess diverse degree distribution whereas they lead to the same observation that the overlay is more disassortative; secondly, the assortativity tends to be able to vary over a large range without changing the degree distribution [24,25].

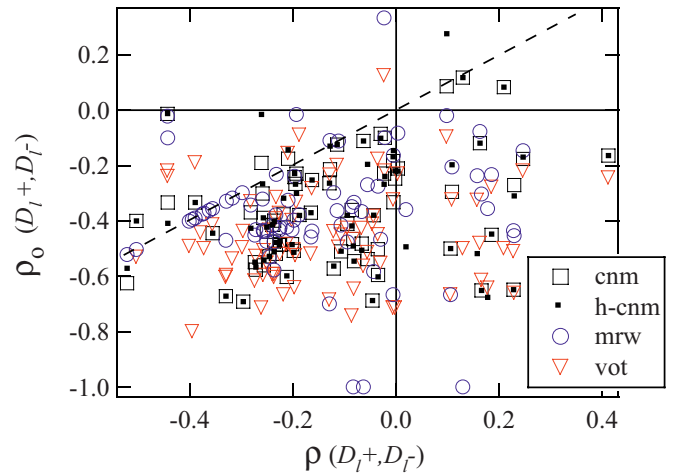


Fig. 3. (Color online) Degree-degree correlation of each real-world network $\rho(D_{l+}, D_{l-})$ and of its community overlay $\rho_o(D_{l+}, D_{l-})$. The dotted line is $\rho_o(D_{l+}, D_{l-}) = \rho(D_{l+}, D_{l-})$.

In order to explain why the overlays are more disassortative, we first investigate the relation between assortativity and modularity. Here, we illustrate the relation between modularity and assortativity of a network given a degree distribution. We have deduced a degree-preserving rewiring algorithm in [20], that, each rewiring step, either increases or decreases the assortativity of a graph. For example, the *degree-preserving assortative random rewiring* is defined as follows: Randomly select two links associated with four nodes and then rewire the two

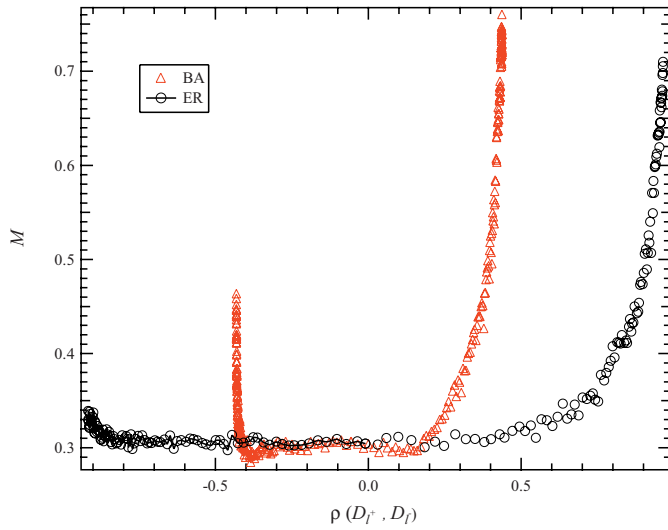


Fig. 4. (Color online) The modularity for both the Barabási-Albert scale-free graph ($N = 500, L = 1500$) and the Erdős-Rényi random graph ($N = 500, L = 1500$) as a function of the degree-degree correlation. The algorithm *cnm* has been used to detect communities.

links such that the two nodes with the highest degree and the two lowest-degree nodes are connected. If any of the new links exists before rewiring, discard this step and a new pair of links is randomly selected. Given a network, the degree-degree correlation can be tuned to range from the minimum to the maximum assortativity via degree-preserving random rewiring. We consider two representative graphs, a Barabási-Albert scale-free graph and an Erdős-Rényi random graph, each with $N = 500$ nodes and $L = 1500$ links. Figure 4 depicts the modularity as a function of the degree-degree correlation. Actually, the assortativity has been varied from the minimum to the maximum for 100 times for each graph. The modularity has been averaged over the 100 realizations. The modularity is mostly positively correlated to the degree-degree correlation, especially when the degree-degree correlation is large. However, negative correlation has been observed in Figure 4 when the assortativity approaches the minimum⁷ as well as in another example in [26]. The positive correlation between assortativity and modularity is, thus, not universal, but it has been widely observed in real-world networks and simulated networks and so far has been explained in [26–28]. Note that the network is always required to be connected during each degree-preserving random rewiring step. Similar results have been observed in [26] where network connectivity is not restricted during the rewiring.

Newman suggested that assortative mixing is one possible mechanism for community formation [27]. The other way around, if a network has an evident community structure reflected by a large modularity m , the structures within communities are relatively homogeneous. In this case, nodes with a similar degree tend to be connected,

⁷ A real-world network seldom achieves the minimum assortativity given its degree sequence as shown in [30].

which contributes to a large assortativity. Thus, a higher modularity contributes to a large assortativity as further explained in [28]. These arguments supports that social networks are in general more assortative than other real-world networks because of their modular structure, even when those networks are of different sizes and link densities [28]. It is the modular or group structure that (partially) account for the assortativity [28].

In view of the relation between modularity and degree-degree correlation, we study the modularity of the set of real-world networks as well as their community overlays. In general, the overlay has a smaller modularity than the original network $M_o < M$, as depicted in Figure 5a. After condensing each community as a node in the overlay, the community structure exhibited in the underlying graph disappears at least partially. Another explanation for the decrease of modularity in the overlay is that the modularity of a network tends to be small in a dense network [29]. As shown in Figure 5b, the overlay is mostly denser than the corresponding underlying network.

Another observation is that the modularity of the two layers are positively correlated with each other. A network with a large modularity has few links between communities. Correspondingly, the number of links in the overlay tends to be small, which may render a large modularity of the overlay [29]. An example is given in Figure 6. Two networks with distinct modularity are derived from the Netscience network by degree-preserving rewiring. The community overlay upon the network with a large modularity is shown to be apparently sparse and its modularity is large.

The overlay network is constructed by abbreviating each community in the underlying network as a node. In this way, the community structure in the overlay is less evident than in the underlying network, rendering $M_o < M$. The modularity and the degree-degree correlation tend to be positively correlated. In other words, modular structure contributes to a high assortativity. As a result, the overlay has a weaker degree-degree correlation compared to the original network. The overlay is more disassortative. Newman [31] found that technological and biological networks are disassortative while social networks are assortative. Here, we argue that the assortativity of a network also depends on the aggregated level of the network: a network is more disassortative at a higher aggregated level. One example is the Internet, a network of autonomous systems (AS) which are collections of IP networks and routers under the control of one or more network operators. The Internet at AS level, as expected, is more disassortative than that at the router level: $\rho_{AS}(D_{l+}, D_{l-}) = -0.189 < \rho_{router}(D_{l+}, D_{l-}) = -0.024$ [31,32].

4.3 Node weight related properties

The node weight in the community overlay represents the percentage of nodes contained in the underlying community. Figure 7a illustrates that the linear correlation coefficient between the degree and the weight of a node in

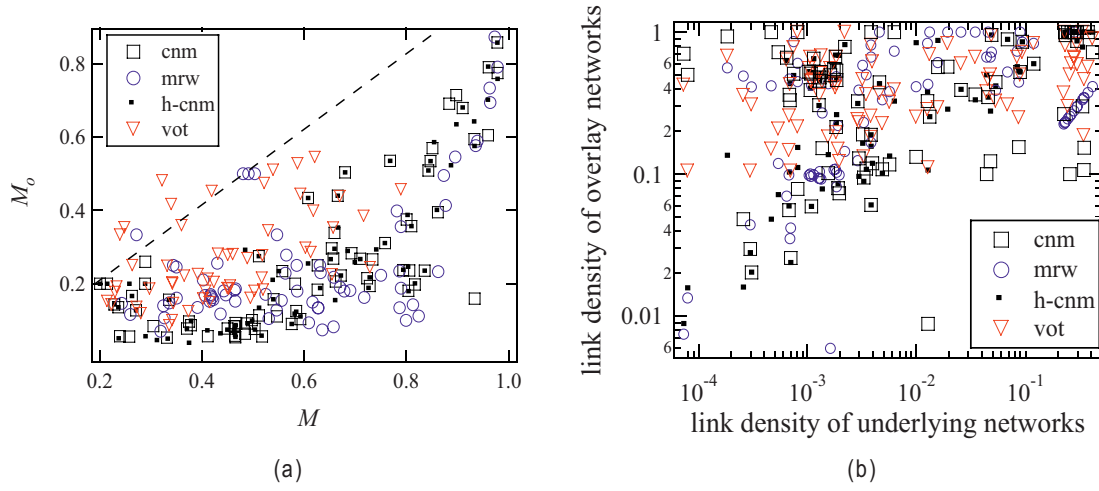


Fig. 5. (Color online) (a) Modularity of each real-world network M and its overlay community network M_o . The dotted line is $M_o = M$. (b) Link density of each real-world network and its overlay community network.

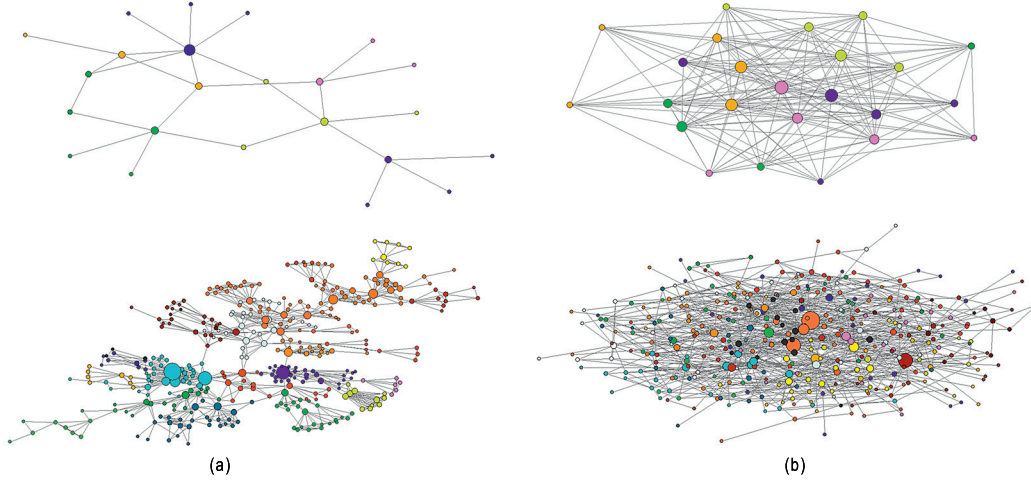


Fig. 6. (Color online) NetScience network ($N = 379, L = 914$) rewired to network (a) with a large modularity $M = 0.859$ with its community overlay $G_o(N = 23, L = 27)$, $M_o = 0.354$ and to network (b) with a small modularity $M = 0.567$ with its community overlay $G_o(N = 24, L = 180)$, $M_o = 0$.

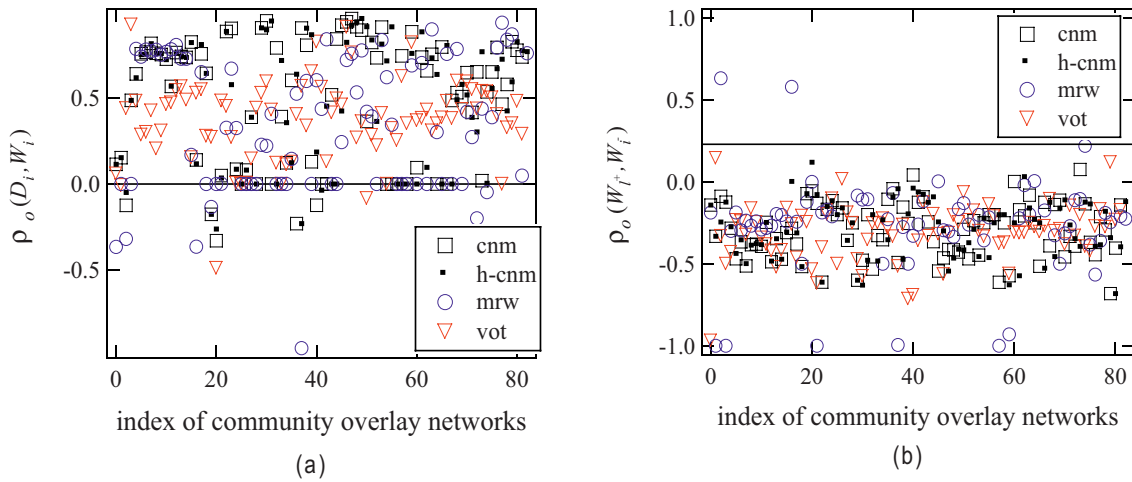


Fig. 7. (Color online) (a) The correlation coefficient $\rho_o(D_i, W_i)$ between the degree and the weight of a node in the community overlay networks. (b) The node weight-node weight correlation $\rho_o(W_{i+}, W_{i-})$ in the community overlay networks.

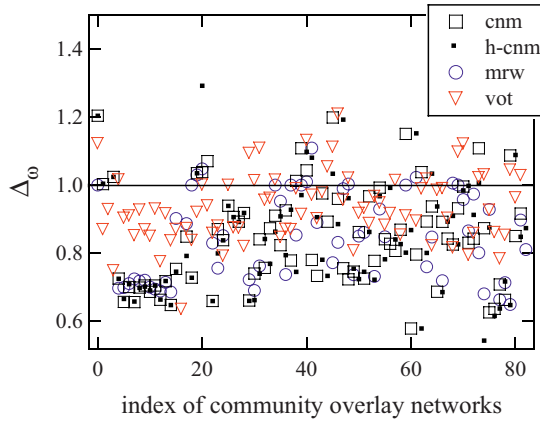


Fig. 8. (Color online) Link weight correlation Δ_w around a node in community overlay networks.

the overlays is mostly positive⁸ $\rho_o(D_i, W_i) > 0$. In other words, a community containing a large number of nodes in the underlying network tends to be directly connected to many other communities. Since the degree correlation in the overlay is mostly disassortative $\rho_o(D_{l+}, D_{l-}) < 0$ and $\rho_o(D_i, W_i) > 0$, node weight-node weight correlation⁹ $\rho_o(W_{l+}, W_{l-})$ can be expected to be disassortative, as shown in Figure 7b.

4.4 Link weight related properties

The link weight correlation around a node is examined by the metric Δ_w , as defined in (3). As depicted in Figure 8, no generic features have been observed in link weight correlation.

Finally, one may wonder whether a community with a large number of nodes inside tend to connect to other communities via many links. Thus, we examine the correlation between the node weight W_i and the average link weight $\sum_j W_{ij}/D_i$ around the node. No obvious generic correlation has been found, as shown in Figure 9. However, it seems that the community partitioning algorithm cnm by Newman favors a negative correlation: a community with a large number of nodes has, on average, a small number of links to other communities. This observation follows the concept of algorithm cnm, which maximizes the modularity (1), equivalently, minimizes the number of links between communities. Figure 9 illustrates the influence of community detection algorithms on the observed community properties, which motivates us to choose five algorithms based on different principles to observe the real universal properties on the community overlays.

⁸ Note that several overlays have a degree and node weight correlation close to zero. As shown in Figure 5b, the link density of some overlays is close to one. In this case, all the nodes have almost the same degree and there is no correlation between degree and node weight.

⁹ The node weight-node weight correlation is the linear correlation coefficient in the node weight of all connected node pairs in a network. It is also the assortativity based on node weight.

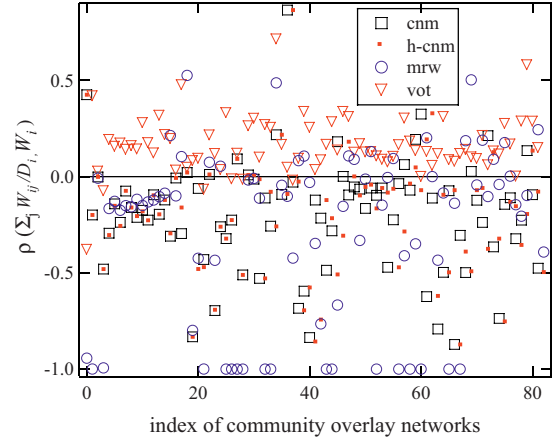


Fig. 9. (Color online) The correlation coefficient between the node weight W_i and the average link weight around the node $\sum_j W_{ij}/D_i$.

5 Conclusion

The proposed community overlay (Fig. 1) captures the interconnections of the underlying communities, the percentage of links between communities and the percentage of nodes/links or other properties of underlying communities. It allows us to systematically investigate large-scale properties of networks at community level. Different community detection algorithms return most often different community partitions. Without losing generality, we selected five classic algorithms based on distinct principles. The community structures of a large number of real-world networks have been computed by the five algorithms. Each community structure detected by each algorithm is mapped into a community overlay. The following overlay properties seems to be universal, independent of the underlying network and the detection algorithm: (a) the overlay network is more disassortative in degree-degree correlation compared to the original network; (b) a large community tends to connect to many other communities; (c) large communities tend to connect to small communities and (d) the modularity of the underlying network and of the overlay are positively correlated. Observation (c) is a natural consequence of (a) and (b). Actually, as we condense each community in the underlying network into a node in the overlay, the modularity is reduced. In other words, evident community structure hardly appears in the overlay. Modularity tends to be positively correlated with or contributes to assortativity. Thus, the assortativity in the overlay is expected to be smaller or the overlay is more disassortative. Whether a complex network is assortative or disassortative does not only depend on the mechanism based on which the network evolves. The seemingly generic observation (a) suggests that the aggregation level of a network plays also an important role in determining the assortativity of a network. This may explain the disassortativity observed in the As level Internet topology, which is a higher hierarchical level upon the router level Internet topology.

We are very grateful to Prof. Piet Van Mieghem for his valuable comments. Xin Ge is supported by his supervisor Prof. Hai Zhao and funded by China Scholarship Council (CSC).

Appendix: Results derived from algorithm btw

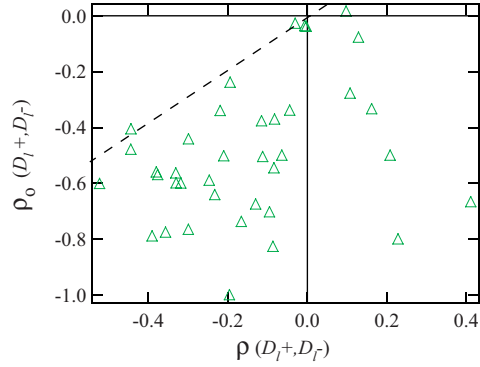


Fig. A.1. (Color online) Degree-degree correlation of each real-world network $\rho(D_{l+}, D_{l-})$ and of its community overlay $\rho_o(D_{l+}, D_{l-})$. The dotted line is $\rho_o(D_{l+}, D_{l-}) = \rho(D_{l+}, D_{l-})$.

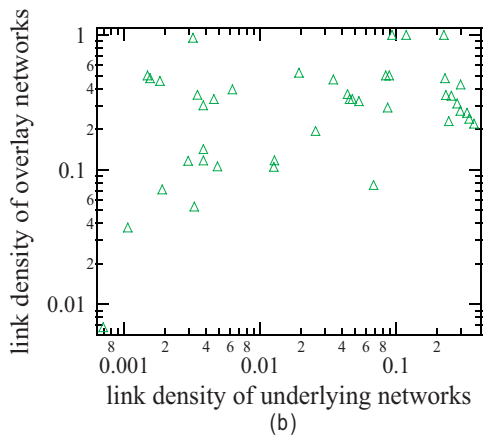
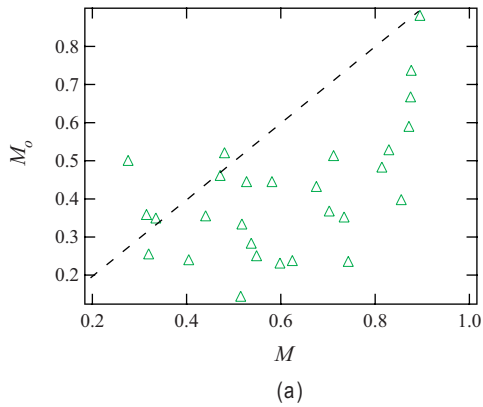


Fig. A.2. (Color online) (a) Modularity M of each real-world network and M_o of its community overlay. The dotted line is $M_o = M$. (b) Link density of each real-world network and its community overlay.

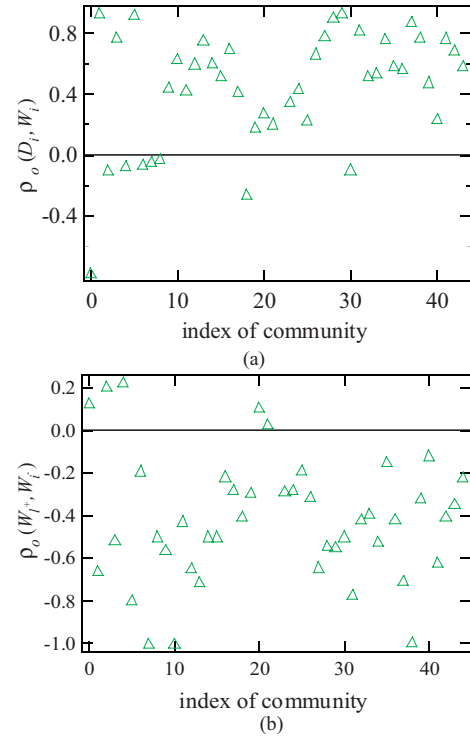


Fig. A.3. (Color online) (a) The correlation coefficient $\rho_o(D_i, W_i)$ between the degree and the weight of a node in the community overlay networks. (b) The node weight-node weight correlation $\rho_o(W_{l+}, W_{l-})$ in the community overlay networks.

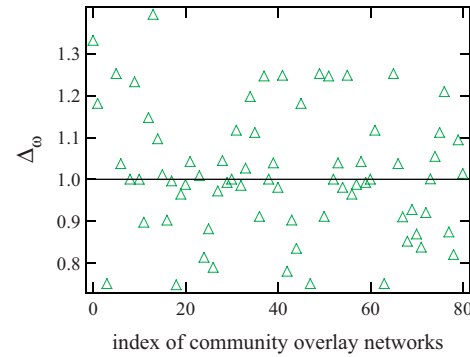


Fig. A.4. (Color online) Link weight correlation Δ_w around a node in community overlay networks.

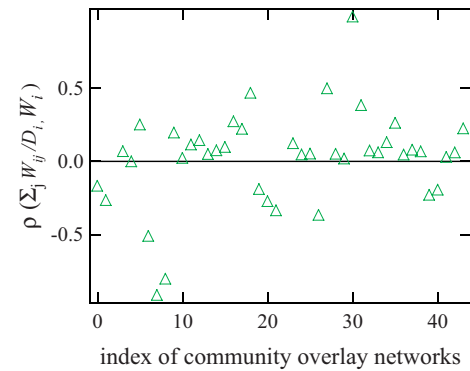


Fig. A.5. (Color online) The correlation coefficient between the node weight W_i and the average link weight around the node $\sum_j W_{ij}/D_i$.

References

1. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
2. L. da F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, Adv. Phys. **56**, 167 (2007)
3. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
4. R. Albert, A. Barabasi, Science **286**, 509 (1999)
5. D. Chen, Y. Fu, M. Shang, Physica A **388**, 2741 (2009)
6. S. Fortunato, Phys. Rep. **486**, 75 (2010)
7. G. Wang, Y. Shen, M. Ouyang, Comput. Math. Appl. **55**, 2746 (2008)
8. F. Wu, B.A. Huberman, Eur. Phys. J. **38**, 331 (2004)
9. M.E.J. Newman, Phys. Rev. E **69**, 066133 (2004)
10. A. Arenas, L. Danon, A. Díaz-Guilera, P.M. Gleiser, R. Guimer, Eur. Phys. J. B **38**, 373 (2004)
11. M.E.J. Newman, Proc. Natl. Acad. Sci. USA **103**, 8577 (2006)
12. G. Palla, I. Derényi, I. Farkas, T. Vicsek, Nature **435**, 814 (2005)
13. C. Song, S. Havlin, H.A. Makse, Nature **433**, 392 (2005)
14. M.E. Newman, M. Girvan, Phys. Rev. E **69**, 026113 (2004)
15. A. Clauset, M.E. Newman, C. Moore, Phys. Rev. E **70**, 066111 (2004)
16. M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002)
17. M. Rosvall, C.T. Bergstrom, Proc. Natl. Acad. Sci. **105**, 1118 (2008)
18. K. Wakita, T. Tsurumi, Finding Community Structure in Mega-scale Social Networks, *Proceedings of IADIS international conference on WWW/Internet 2007* (2007), pp. 153–162
19. M.E.J. Newman, Phys. Rev. E **67**, 026126 (2003)
20. P. Van Mieghem, H. Wang, X. Ge, S. Tang, F.A. Kuipers, Eur. Phys. J. B **76**, 643 (2010)
21. H. Wang, L. Douw, J. Martin Hernandez, J.C. Reijneveld, C.J. Stam, P. Van Mieghem, Phys. Rev. E **82**, 021924 (2010)
22. J.J. Ramasco, B. Gonçalves, Phys. Rev. E **76**, 066106 (2007)
23. J. Park, M.E.J. Newman, Phys. Rev. E **68**, 026112 (2003)
24. H. Wang, W. Winterbach, P. Van Mieghem, Eur. Phys. J. B **83**, 203 (2011)
25. P. Holme, Phys. Rev. E **75**, 046111 (2007)
26. P. Van Mieghem, X. Ge, P. Schumm, S. Trajanovski, H. Wang, Phys. Rev. E **82**, 056113 (2010)
27. M.E.J. Newman, M. Girvan, Mixing patterns and community structure in networks, in *Statistical Mechanics of Complex Networks*, edited by R. Pastor-Satorras, J. Rubi, A. Diaz-Guilera (Springer, Berlin, 2003)
28. M.E.J. Newman, J. Park, Phys. Rev. E **68**, 036122 (2003)
29. R. Guimerà, M. Sales-Pardo, L.A. Nunes Amaral, Phys. Rev. E **70**, R025101 (2004)
30. S. Zhou, R.J. Mondragón, New J. Phys. **9**, 173 (2007)
31. M.E. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
32. J. Zhang, H. Zhao, J. Xu, Z. Liu, Comput. Commun. **33**, 2001 (2010)