

# UCLA

## UCLA Previously Published Works

**Title**

Detecting communities based on network topology.

**Permalink**

<https://escholarship.org/uc/item/9h70h1bn>

**Journal**

Scientific reports, 4(1)

**ISSN**

2045-2322

**Authors**

Liu, Wei  
Pellegrini, Matteo  
Wang, Xiaofan

**Publication Date**

2014-07-01

**DOI**

10.1038/srep05739

Peer reviewed



## OPEN

## Detecting Communities Based on Network Topology

## SUBJECT AREAS:

COMPUTATIONAL  
SCIENCE

SCIENTIFIC DATA

CLASSIFICATION AND  
TAXONOMYWei Liu<sup>1,2</sup>, Matteo Pellegrini<sup>2</sup> & Xiaofan Wang<sup>1</sup><sup>1</sup>Department of Automation, Shanghai Jiao Tong University, Shanghai, 200240, China, <sup>2</sup>Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, 90055.Received  
20 March 2014Accepted  
27 June 2014Published  
18 July 2014Correspondence and  
requests for materials  
should be addressed to  
X.W. (xfwang@sjtu.  
edu.cn)

Network methods have had profound influence in many domains and disciplines in the past decade. Community structure is a very important property of complex networks, but the accurate definition of a community remains an open problem. Here we defined community based on three properties, and then propose a simple and novel framework to detect communities based on network topology. We analyzed 16 different types of networks, and compared our partitions with Infomap, LPA, Fastgreedy and Walktrap, which are popular algorithms for community detection. Most of the partitions generated using our approach compare favorably to those generated by these other algorithms. Furthermore, we define overlapping nodes that combine community structure with shortest paths. We also analyzed the *E. Coli* transcriptional regulatory network in detail, and identified modules with strong functional coherence.

Communities are groups that are densely connected among their members, and sparsely connected with the rest of the network. Community structure can reveal abundant hidden information about complex networks that is not easy to detect by simple observation. There are many large-scale complex networks (systems) in the real world whose structure is not fully understood. A great deal of research has been carried out to uncover the structures of these real world networks, to improve the ability to manage, maintain, renovate and control them. With the help of varied approaches, it is possible to shed light on the general structure of these networks, and further understand their function.

Network science methods have been used in various settings<sup>1,2</sup>, including social<sup>3,4</sup>, information<sup>5</sup>, transportation<sup>6</sup>, energy<sup>7</sup>, ecological<sup>8</sup>, disease<sup>9</sup>, and biological networks<sup>10–13</sup>. In most of these cases we can find clear community structures, which are usually associated with specific functions. However, to date, most detection methods have limitations, and there is still a lot of room to develop more general approaches.

At present, most methods focus on the detection of node community. One popular approach is based on the optimization of the modularity  $Q$ <sup>14,15,52,56</sup> of a sub-network. Some methods<sup>13,14,29,34,38–40</sup> force every node to be assigned to a single community. This assumption doesn't always reflect real world networks, where several overlapping communities can co-exist. For example, in social networks, a person may have family relationship circles, job circles, friend circles, social circles, hobby circles and so on. Algorithms that can discover overlapping communities<sup>16–23</sup> have been developed, and recently, methods to detect link communities<sup>20,24,25</sup> have been presented. The concept of a link community is useful for discovering overlapping communities, as edges are more likely to have unique identities than nodes, which instead tend to have multiple identities. In addition, statistical<sup>24</sup>, information-theoretic<sup>25,48,53</sup> and synchronization and dynamical clustering approaches<sup>49,50,58–60</sup> have also been developed to detect communities.

No matter which method is used to detect community structure, they should present a quantitative definition of community first, but no definition is universally accepted so far. Here we defined community based on three properties, and then propose a simple and novel method to detect communities based on network topology. The main idea is similar to “fishing”. We first use the adjacency lists of nodes (network topology) as a “fishpond”, then detect some strong sense communities from “fishpond” as “baits”, and then use these “baits” to catch the “fish” (weak sense communities). Our method is simple, stable, and easy to understand. It's a parameter free approach, and it can detect overlapping communities, isolated communities and determine the number of communities in an unsupervised manner. Moreover, our method can detect not only cohesive and large communities, but also sparse and small ones. We analyzed 16 different types of networks with our method and the results show that our approach compares favorably to other methods.

**Results**

We have tested the performance of our method with both synthetic networks and real-world networks. The size of the networks spans tens to tens of thousands of nodes. We analyzed 16 different types of networks (as shown in



Table 1 | The lists of we analyzed networks

Networks	Nodes	Edges	Reference
GR_256	256	702	[29]
LFR_128	128	597	[42]
LFR_256	256	1294	[42]
LFR_512	512	3285	[42]
Karate	34	78	[30]
Dolphins	62	159	[32]
Football	115	613	[14][31]
Pol-books	105	441	[14]
Jazz	198	2742	[36]
E. Coli	418	519	[33]
Email	1133	5451	[44]
Facebook	2888	2981	[34]
Protein	3274	8748	[43]
Power	4941	6594	[7]
Collaboration	5242	14490	[41]
PGP	10680	24316	[37]

table 1), and compared our partitions with several popular algorithms. The results are in tables 2 and 3. Below we analyze one synthetic network and four real world networks in detail.

**Synthetic Network.** The synthetic network we analyzed is the LFR benchmark network<sup>42</sup>. It contains 128 nodes and 597 edges, and includes 8 communities and 10 overlapping nodes. The average degree is 9.328, the maximum degree is 30, the minimum size of a community is 10, the maximum size of a community is 30, the mixing parameter is 0.1, and the number of memberships of overlapping nodes is 2. As shown in Fig. 1, we detected 8 communities in this network using our algorithm, and they are identical with those real communities, except for the 10 overlapping nodes. The reason is these 10 overlapping nodes do not satisfy the overlapping node definition of our algorithm, which combines the structure property with shortest path. We provide examples of other benchmark networks in table 2.

**Real-World Networks.** The Zachary karate club network is a famous empirical network. A conflict between club president John (node 34) and the instructor Mr. Hi (node 1) lead to 34 members of the university sports club to split into two groups<sup>30</sup>. As Fig. 2 (a) shows, the two communities discovered by our algorithm are identical with the groups described by Newman<sup>30</sup>. We note that some nodes should be defined as “overlapping nodes” based on the general definition. For example, node 10 and node 3 have equal numbers of connections with two communities, but they have different shortest path lengths with nodes 1 and 34, the hub nodes. As a result, we don’t define them as overlapping nodes based on our definition.

The American college football network is another widely used empirical network compiled by Newman<sup>14</sup> in 2004. There are 115 Division I-A teams that play 613 games during the regular fall season of 2000, and these teams are grouped into 11 different conferences, except for 8 independent teams. We found 11 strong sense communities using our algorithm, as shown in Fig. 2 (b). These are identical to the 11 conferences of Division I-A teams, except for the 8 independent teams, which are assigned to “The Southeast Conference”, “The Big East Conference” and “The Mid-American Conference” respectively.

The Facebook network<sup>34</sup> is a directed user–user friendships network compiled by Julian and Jure in 2012. There are 2888 users and 2981 friendships in this network. An edge indicates that the user represented by the left node is a friend of the user represented by the right node. We identified 7 strong sense communities and 9 overlapping nodes, as shown in Fig. 2 (c). The structure of this network is very clear based on visual inspection. All the friendships are established around ten users, so that we should have 10 communities. Our partitions merge community pairs involving nodes 603 and 288, 710 and 714, and 2687 and 2699. We found that all these pairs are directly connected to each other and all the overlapping nodes are of the first type.

**Biological Network.** The E. Coli transcriptional regulatory network<sup>33</sup> was compiled by Shen-Orr et.al. in 2002. There are 423 operons and 519 regulatory links as well as 5 self-regulation events. This is a directed network, and each edge is directed from an operon that encodes a transcription factor to an operon that it directly regulates (an operon is one or more genes transcribed on the same mRNA). Here we use an undirected version of the network, and analyze the network using the updated RegulonDB<sup>47</sup> 8.3.

As shown in Fig. 2 (d), the E. Coli transcriptional regulatory network is composed of 29 disconnected sub-networks and 5 isolated nodes. We detected all the disconnected parts and isolated nodes correctly using our algorithm. The largest sub-network was divided into 18 communities and 19 overlapping nodes. We analyze the 23 modules that have more than 3 members using the DAVID functional annotation tool<sup>45,46</sup>. All of the communities are functionally coherent (i.e. the genes appear to participate in a common biological process). For example, the first module contains 7 operons, which are enriched for the process “Arginine biosynthesis (p-value is 8.1E-28)”. The second module contains 8 operons (23 genes), and 22 genes are involved in “Sulfur metabolic processes (p-value is 2.1E-39)”. The results of other modules are shown in Table 4. Each module we identified has at least one transcription factor (except for module 4). We also found that the all of the overlapping nodes we discovered are of the first type. Besides gene *ecfI*, which doesn’t have any annotation, the other overlapping nodes all share different functions across modules.

Table 2 | The analyze of empirical networks

Networks	Names	Clusters	Ours		Infomap		LPA		Fastgreedy		Walktrap	
			N <sub>c</sub>	NMI	N <sub>c</sub>	NMI	N <sub>c</sub>	NMI	N <sub>c</sub>	NMI	N <sub>c</sub>	NMI
GR_256		18	18	1.0000	18	1.0000	18	1.0000	18	0.9688	18	0.9890
LFR_128		8	8	0.9184	50	0.6233	8	0.9187	7	0.8276	8	0.9200
LFR_256		14	14	0.9751	109	0.6776	14	0.9676	12	0.8498	14	0.9681
LFR_512		26	26	0.9820	219	0.7168	24	0.9304	17	0.7085	24	0.9172
Karate		2	2	1.0000	3	0.6995	3	0.6995	3	0.6925	5	0.5042
Dolphins		2	3	0.7621	6	0.5373	6	0.5699	4	0.5571	4	0.5816
Football		12	11	0.9492	12	0.9720	9	0.8947	6	0.7532	10	0.9369
Polbooks		3	3	0.5780	6	0.4935	3	0.5744	4	0.5308	4	0.5428

N<sub>c</sub> represents the number of communities detected by different algorithms, and clusters represent the number of real communities of empirical network.



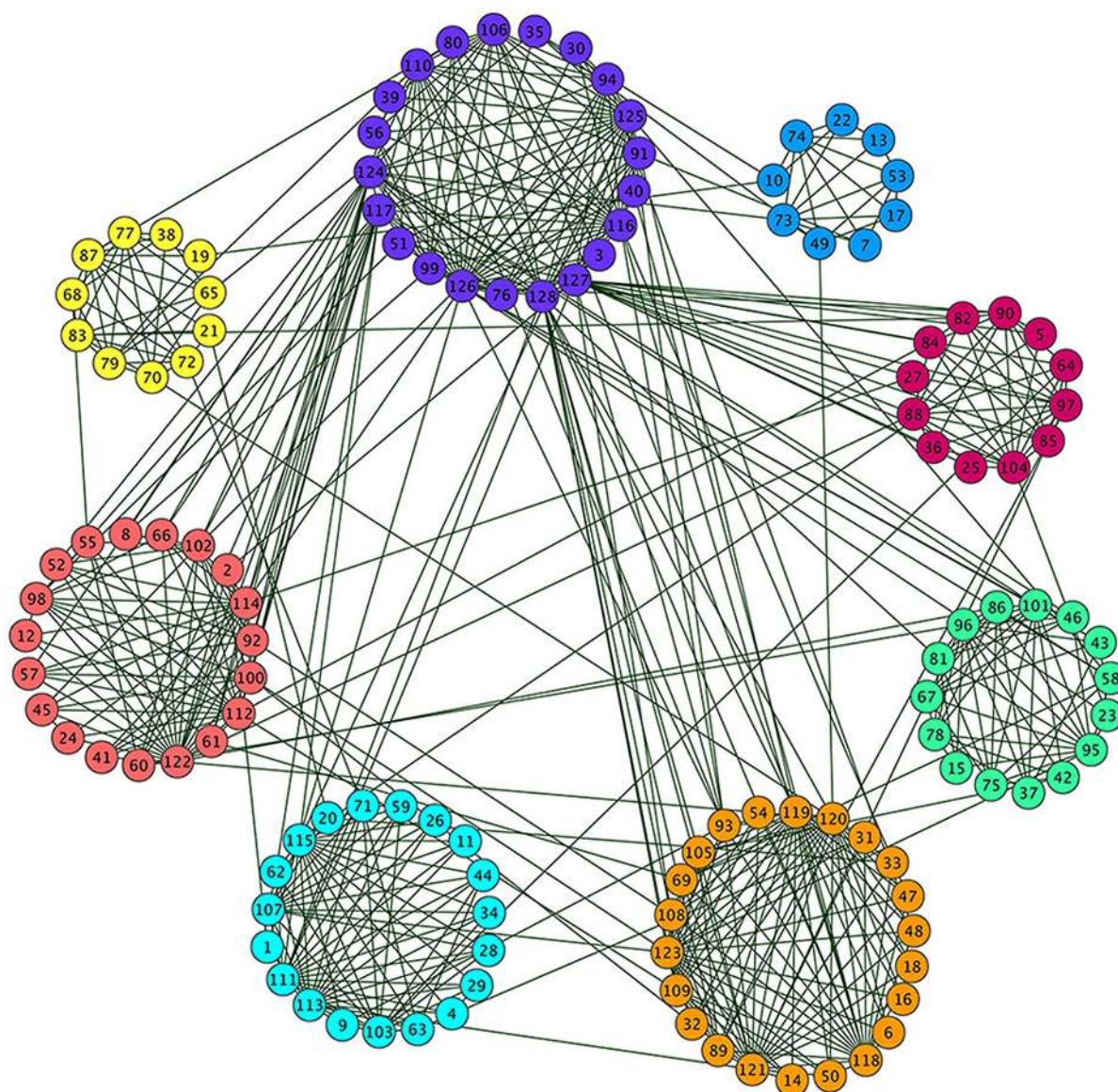
Table 3 | The analyze of unempirical networks

Networks	Ours		Infomap		LPA		Fastgreedy		Walktrap	
	$N_c$	Q	$N_c$	Q	$N_c$	Q	$N_c$	Q	$N_c$	Q
Jazz	3	0.4397	7	0.2800	3	0.4295	4	0.4389	11	0.4384
E. Coli	46	0.8902	72	0.7351	55	0.7445	45	0.7784	52	0.7463
Email	19	0.5374	69	0.5260	5	0.0320	12	0.5070	49	0.5307
facebook	7	0.7307	11	0.7961	9	0.7496	8	0.8087	6	0.6331
Protein	303	0.8673	423	0.7105	365	0.7095	203	0.7632	395	0.7291
Power	632	0.8199	487	0.8182	495	0.8043	41	0.9335	364	0.8310
Collab.	684	0.8378	716	0.7921	720	0.7950	427	0.8026	815	0.7824
PGP	770	0.8194	1065	0.8014	985	0.8023	204	0.8525	1574	0.7894

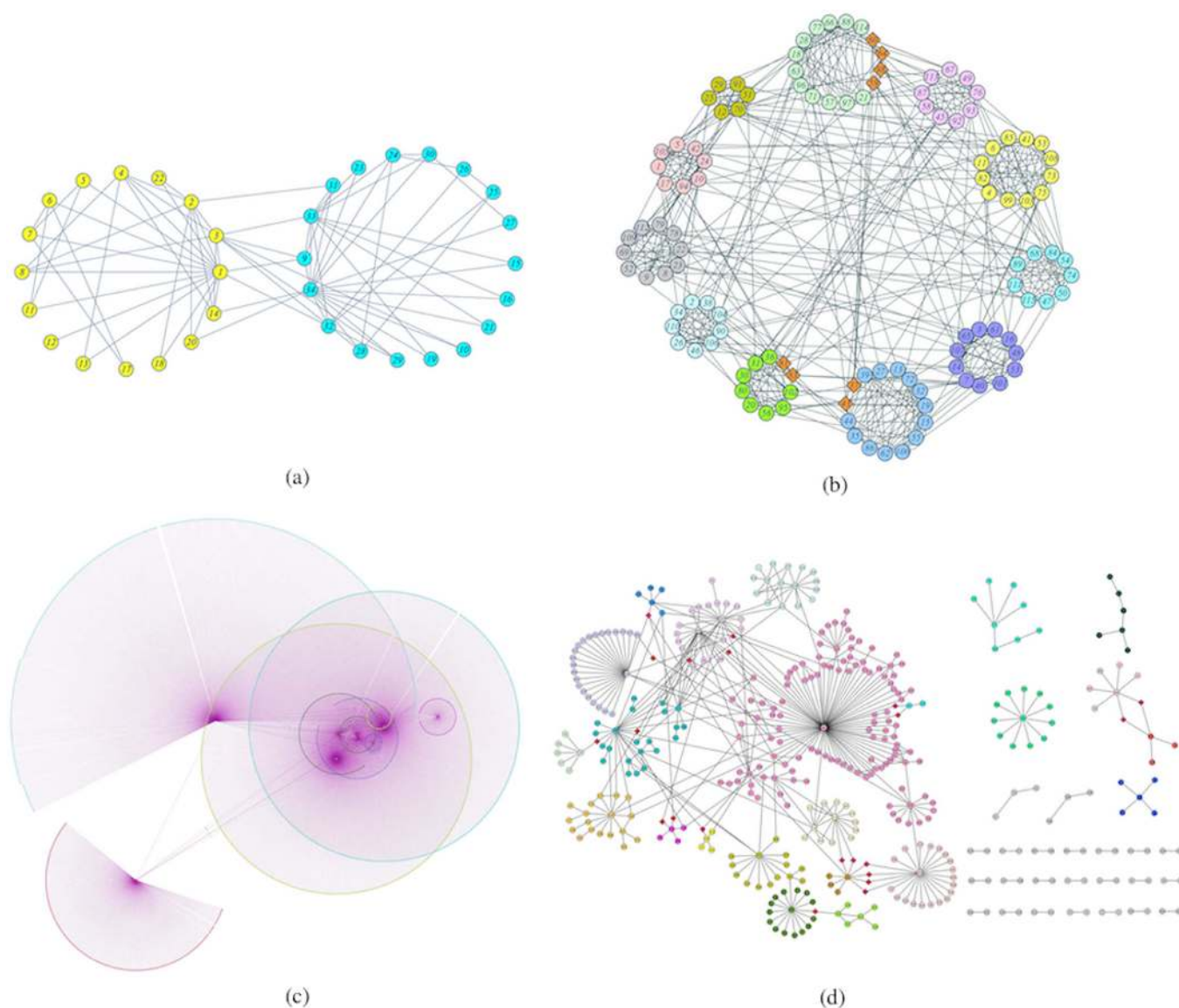
$N_c$  represents the number of communities.

**Comparison with other methods.** To test the performance of our algorithm, we compared our partitions with four popular algorithms: Infomap<sup>35</sup>, LPA<sup>57</sup>, Fastgreedy<sup>51</sup> and Walktrap<sup>55</sup>. Empirical networks are compared using NMI values, as shown in table 2. The others are

compared using the modularity Q metric, as shown in table 3. We can see that our algorithm, LPA and Walktrap all perform well on synthetic networks. However, for empirical real world networks, our algorithm and Infomap perform better than the other



**Figure 1** | A network of LFR-benchmark with 128 nodes. This network is constructed from 8 pre-assigned communities, and 10 overlapping nodes are included.



**Figure 2 | The networks of real-world.** (a) Zachary karate club network: the two communities we detected are identical with the two real communities. (b) US college football network: the 11 communities we detected are identical with the 11 real conferences except for the 8 independent teams, which are assigned to 3 different conferences respectively. (c) Facebook network: we discover 7 communities and 9 overlapping nodes (orange diamond shaped nodes) using our algorithm. (d) *E. Coli* transcriptional regulatory network: we discovered 46 modules, 5 isolated nodes and 19 overlapping nodes (red diamond shaped nodes) with our algorithm.

methods. Our algorithm and Fastgreedy always have stable higher modularity. We conclude that our algorithm performs well and is competitive with other methods.

## Discussion

Discovering complex network community structure has become an important challenge during the past decade. Several advanced algorithms have been proposed to detect community structures in complex networks, but each has limitations<sup>35</sup>. For example, some approaches don't perform well on large-scale networks, some need to pre-estimate community numbers, some can't uncover overlapping communities, some depend on multiple parameters, some are unable to discover sparse modules or small communities, some are domain-specific, work with specific structures, and still some don't generate stable partitions etc.

Our algorithm overcomes most of these limitations. It's a parameter free approach, and can find communities from adjacency lists of nodes directly. Thus it is conceptually very simple, efficient, easily

implementable and suitable for large-scale networks. As we have shown, it can be applied to networks from multiple domains. It can auto-detect the number of communities, discover isolated nodes and isolated communities, and always outputs stable partitions. Furthermore, It can offer two different kinds of overlapping communities, and detect cohesive communities, sparse communities and small communities as well. There are many sparse communities and small communities in real world networks, so it is important to be able to identify these structures. For example, sparse communities and small communities are important structures in biological networks.

We propose a novel and simple framework to detect community structure of complex networks based on network topology. Compared to popular methods previously reported in the literature, our algorithm performs competitively for both synthetic and real networks, but as we all know, it is far from providing an unforeseen breakthrough in community finding. In the future, we intend to improve the ability of our algorithm to detect the second type of overlapping nodes and discover hierarchical structure of complex

Table 4 | The function annotations of *E. coli*. modules detected by our algorithm

$I_m$	$N_o$	$N_g$	Function clusters	P-value	TFs	$N_m$
1	7	10	Arginine biosynthesis	8.1E-28	1	10
2	8	23	Sulfur metabolic process	2.1E-39	2	22
3	6	10	Carbohydrate catabolic process	6.0E-12	2	9
4	4	6	Cellular carbohydrate metabolic process	5.6E-7	0	6
5	15	51	Locomotion	2.2E-90	2	45
6	12	16	SOS response	8.5E-29	1	13
7	14	27	Amine metabolic process	7.6E-22	2	22
8	7	7	Cellular amino acid biosynthetic process	3.6E-10	2	7
9	5	5	Cytoplasm	3.6E-3	1	5
10	5	23	Ion transport	5.0E-34	1	22
11	16	24	Nitrogen compound biosynthetic process	1.9E-19	1	19
12	16	18	Cellular process	3.1E-3	5	16
13	24	34	Cellular metabolic process	6.2E-6	1	25
14	18	35	Cellular metabolic process	5.8E-10	8	32
15	5	12	Chorismate metabolic process	3.0E-25	1	11
16	8	11	Cellular aromatic compound metabolic process	2.4E-15	1	10
17	24	56	Cellular process	7.3E-4	3	15
18	25	73	Generation of precursor metabolites and energy	1.7E-69	5	58
19	10	19	Macromolecule metabolic process	1.2E-5	1	15
20	124	236	Primary metabolic process	8.9E-21	35	183
21	6	7	Monocarboxylic acid metabolic process	1.1E-8	2	6
22	6	11	Carbohydrate metabolic process	1.1E-9	1	10
23	26	71	Metabolic process	1.1E-5	9	61

$I_m$  represents the index of Module,  $N_o$  represents the number of operons each module contains,  $N_g$  represents the number of genes each module contains, TFs represents the number of TF in each module,  $N_m$  represents how many genes are matched with the DAVID *Escherichia coli* database in each module.

network. While in the current implementation we do identify most overlapping nodes, others are left out, hindering the ability to split bigger communities into smaller ones.

## Methods

We defined community based on three properties, and then propose a simple and novel framework to detect communities based on network topology. The three properties of community are community structure property, community membership property and overlapping member properties respectively. The community structure property is used to define a community, while the community membership and overlapping member properties are used to define the members of a community.

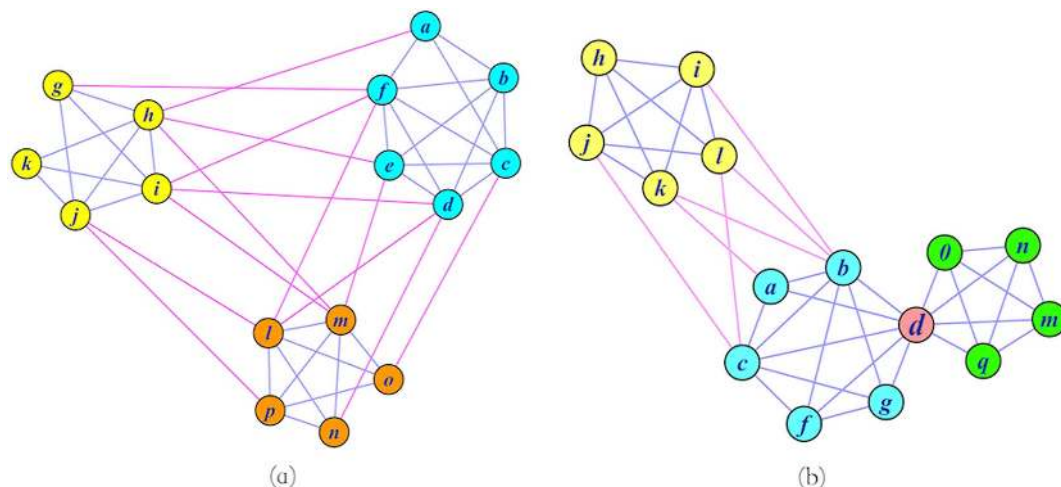
**Community structure property.** Although there is not a general definition, it is widely accepted that a community should be a sub-network that is internally densely connected, while externally sparsely connected<sup>23,24,26,27,31</sup>. Here we define two types of community structures: strong sense and weak sense communities. A sub-network is defined as a strong sense community if its internal connections are larger than its external connections. A sub-network is defined as a weak sense community, if its

internal connections are equal or smaller than its external connections, but its internal connections are larger than the connections between this sub-network and any other communities.

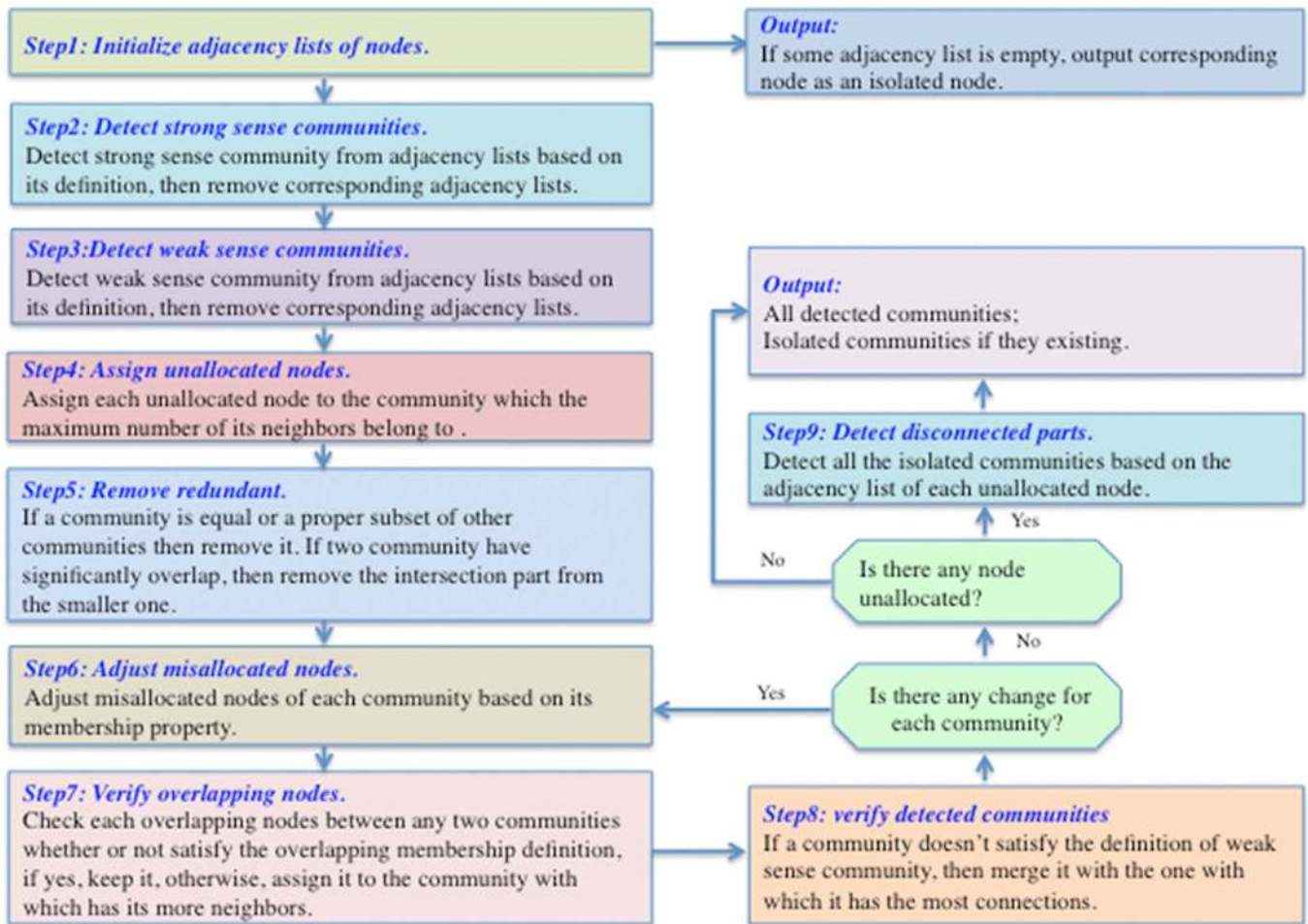
In Fig. 3 (a) we provide an example of two types of community structures. In this network, the internal connections of a community are colored blue, and the connections among communities are colored purple. Based on our definition, it's easy to see that the cyan community is a strong sense community, as it has more internal connections than external connections. While the yellow community is a weak sense community, as although the number of internal connections equals the number of external ones, its internal connections are more numerous than the connections between it and any one of the other two communities. As we can see, the orange community is also a weak sense community.

**Community membership property.** By definition, a member of a community should have more neighbors within its community than in any other community, unless it is an overlapping member. That is to say, each node should join the community, which has its maximum number of neighbors, except for overlapping nodes.

**Overlapping member property.** We define two types of overlapping nodes. One is based on the number of connections between it and corresponding communities, and



**Figure 3 | Example of community property.** (a) Community structure property: the nodes of this network divide into three clusters, the cyan cluster is a strong sense community, and the others are weak sense communities. (b) Overlapping member property: node d is a second type of overlapping node shared by cyan community and green community.



**Figure 4** | The flowchart of our algorithm. The flow chart described our algorithm step by step in detailed. It includes 8 main steps and 1 additional step if necessary.

the shortest path between it and hub members of corresponding communities, while the other is based on topology structure of community. For the first type of overlapping node, not only the number of connections between it and the corresponding communities should be equal, but also the shortest path between it and hub members of corresponding communities should be equal as well. For the second type of overlapping node, it should be tightly connected with both communities, and there should be few connections between the two corresponding communities if remove such nodes. As the network in Fig. 3 (b) shows, node “d” should be assigned to the cyan community if we don’t take into account overlapping member properties. However, node “d” is an overlapping node of the second type, connecting the cyan and green communities, based on our definition.

**Hub member.** A hub member has the most neighbors within a community.

**Algorithm.** Our algorithm is based on the idea that community structures can be detected from sub-networks by comparing the number of internal and external connections of each community, and it is mainly made up of four parts: initialize the adjacency lists of nodes, search for strong sense communities from adjacency lists, detect weak sense communities based on strong sense ones from adjacency lists, and iteratively readjust nodes to discovered communities based on their community membership property until the approach converges. The specific steps are as shown in Fig. 4. An example of the application of the algorithm to a network is shown in the supplementary materials.

Due to the fact that our algorithm is similar in spirit to the label propagation algorithm (LPA), we compare its performance with LPA. The greatest similarity between these two algorithms is that they both tend to assign each node in the network to the community with which they have the maximum number of neighbors. The main differences have to do with the fact that LPA initializes node communities, while we initialize edge communities (the adjacency list of each node denotes all its corresponding edges); LPA depends on specific random seeds, initial conditions and tie-break rules for its execution, while our approach doesn’t; Our algorithm is a deterministic algorithm, while LPA isn’t; LPA searches for communities based on label dynamic propagation and static network topology, while our algorithm searches for community only based on static network topology.

**Evaluation measures.** To evaluate the performance of our algorithm we need an approach to measure the accuracy of community partitions. However, there is no general standard technique for this, because it is difficult to know the structures of real world networks a priori. Therefore, we use the normalized mutual information (NMI) measure<sup>16</sup> to evaluate community partitions. NMI is defined as follows

$$NMI(X|Y) = \frac{H(X) + H(Y) - H(X,Y)}{(H(X) + H(Y))/2} \quad (1)$$

Where X corresponds to the real communities, Y corresponds to the predicted communities, and  $H(X)$  denotes the entropy of random community X, whereas  $H(X,Y)$  denotes the joint entropy of X and Y.

For other networks we use the modularity<sup>28</sup> measure to evaluate the quality of a partition. It is based on the intuitive idea that random networks do not exhibit community structure. Let us define a matrix  $e$  where the elements  $e_{ij}$  represents the fraction of total connections between two different communities, and the real fraction of links exclusively within a community is  $e_{ii}$ . Then the sum of any row of  $e$ ,  $a_i = \sum_j e_{ij}$  corresponds to the fraction of links connected to community  $i$ , and the expected number of intra-community links is just  $a_i^2$ . We can compare  $e_{ii}$  and  $a_i^2$  directly, and sum over all the communities in the network. This measure is known as modularity:

$$Q = \sum (e_{ii} - a_i^2) \quad (2)$$

**Complexity Analysis**

It is advantageous for an algorithm to have lower time complexity, so that it can be applied to large-scale networks. Our algorithm consists of two stages: the first stage is detecting communities, and the cost time complexity is  $O(k^2 + lkn)$ , where  $n$  is the number of nodes,  $k$  is the number of detected communities and  $l$  is the maximum size of the initial adjacency lists. The second stage is adjusting membership among communities, and has a cost time complexity  $O(vk^2 + lk^2 + kn)$ , where  $v$  is the maximum count of overlapping nodes. Thus, the total time complexity for both steps is  $O((v+l)k^2 + lkn)$ . As a result, this algorithm can be efficiently applied to a network of tens of thousands of nodes and the execution time is around an hour on a typical CPU.



1. Newman, M. E. J., Barabási, A. L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton Univ. Press, 2006).
2. Newman, M. E. J. *Networks an introduction* (Oxford University Press, 2010).
3. Palla, G., Barabási, A. L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
4. González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
5. Onnela, J. P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA* **104**, 7332–7336 (2007).
6. Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. A. N. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* **102**, 7794 (2005).
7. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
8. Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food-web structure. *Nature* **426**, 282–285 (2003).
9. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
10. Barabási, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
11. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
12. Guimerà, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
13. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
14. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
15. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
16. Lancichinetti, A., Fortunato, S. & Kertesz, J. Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys.* **11**, 033015 (2009).
17. Palla, G., Derenyi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).
18. Gregory, S. Finding overlapping communities in networks by label propagation. *New J. Phys.* **12**, 103018 (2010).
19. Lancichinetti, A. & Fortunato, S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* **80**, 06118 (2009).
20. Evans, T. S. & Lambiotte, R. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).
21. Gregory, S. An algorithm to find overlapping community structure in networks. *Lect. Notes Comput. Sc.* **4702**, 91–102 (2007).
22. Shen, H., Cheng, X., Cai, K. & Hu, M. B. Detect overlapping and hierarchical community structure in networks. *Phys. Stat. Mech. Appl.* **388**, 1706–1712 (2009).
23. Li, D. *et al.* Synchronization interfaces and overlapping communities in complex networks. *Phys. Rev. Lett.* **101**, 168701 (2008).
24. Ahn, Y. Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multi-scale complexity in networks. *Nature* **466**, 761–764 (2010).
25. Evans, T. S. & Lambiotte, R. link partitions and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).
26. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
27. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. & Parisi, D. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA* **101**, 2658–2663 (2004).
28. Danon, L., Díaz-Guilera, A., Duch, J. & Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theor. Exp.* P09008 (2005).
29. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).
30. Zachary, W. W. An Information Flow Model for Conic and Fission in Small Groups. *J. Anthropol. Res.* **33**, 452–473 (1977).
31. Evans, T. S. Clique Graphs and Overlapping Communities. *J. Stat. Mech.* P12037 (2010).
32. Lusseau, D. The Emergent Properties of a Dolphin Social Network. *Proc. Biol. Sci.* **270**, (suppl. 2), S186–S188 (2003).
33. Shen-Orr, S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
34. McAuley, J. & Leskovec, J. Learning to discover social circles in ego networks. *Adv. Neural Inf. Process. Syst.* **25**, 548–556 (2012).
35. Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. *Phys. Rev. E* **80**, 056117 (2009).
36. Gleiser, P. & Danon, L. Community structure in jazz. *Adv. Complex Syst.* **06**, 565 (2003).
37. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Phys. Rev. E* **70**, 056122 (2004).
38. Newman, M. E. J. Communities, modules and large-scale structure in networks. *Nat. Phys.* **8**, 25–31 (2011).
39. Zhang, S. & Zhao, H. Community identification in networks with unbalanced structure. *Phys. Rev. E* **85**, 066114 (2012).
40. Aldecoa, R. & Marín, I. Deciphering Network Community Structure by Surprise. *PLOS ONE* **6**(9), e24195 (2011).
41. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).
42. Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **78**, 046110 (2008).
43. Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. & David, E. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32**, D449–D451 (2004).
44. Guimerà, R., Danon, L., Diaz-Guilera, A., Giral, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **4**, 44–57 (2009).
46. Huang, D. W. & Sherman, B. T. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
47. Salgado, H. & Peralta, M. *et al.* RegulonDB (version 8.0): Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* **41**, D203–213 (2012).
48. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **105**, 1118–1123 (2008).
49. Boccaletti, S., Ivanchenko, M., Pluchino, A., Latora, V. & Rapisarda, A. Detecting complex network modularity by dynamical clustering. *Phys. Rev. E* **75**, 045102(R) (2007).
50. Oh, E., Choi, C., Kahng, B. & Kim, D. Modular synchronization in complex networks with a gauge Kuramoto model. *Europhys. Lett.* **83**, 68003 (2008).
51. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004).
52. Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005).
53. Rosvall, M. & Bergstrom, C. T. An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. USA* **104**, 7327–7331 (2007).
54. Lancichinetti, A., Radicchi, F., Ramasco, J. & Fortunato, S. Finding statistically significant communities in networks. *PLOS ONE* **6**, e18961 (2011).
55. Pons, P. & Latapy, M. Computing communities in large networks using random walks. *J. Graph Algorithms Appl.* **10**, 191–218 (2006).
56. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).
57. Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007).
58. Lancichinetti, A. & Fortunato, S. Consensus clustering in complex networks. *Sci. Rep.* **2**, srep00336 (2012).
59. Gfeller, D. & De Los Rios, P. Synchronization and Spectral Coarse Graining in Oscillator Networks. *Phys. Rev. Lett.* **100**, 174104 (2008).
60. Yuan, W. J. & Zhou, C. Interplay between structure and dynamics in adaptive complex networks: Emergence and amplification of modularity by adaptive dynamics. *Phys. Rev. E* **84**, 016116 (2011).

## Acknowledgments

This work is supported by CSC scholarship Fund of china, and the National Natural Science Foundation of China under Grant No. 61374176.

## Author contributions

W.L. analyzed data, designed and performed research. M.P., W.L. and X.F.W. discussed the results and wrote the manuscript text. All authors reviewed the manuscript.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Liu, W., Pellegrini, M. & Wang, X.-F. Detecting Communities Based on Network Topology. *Sci. Rep.* **4**, 5739; DOI:10.1038/srep05739 (2014).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>