

SOFTWARE ARTICLE

BMTK: a toolkit for determining modules in biological bipartite networks

Bei Wang, Jinyu Chen and Shihua Zhang*

NCMIS, CEMS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zsh@amss.ac.cn

Received October 3, 2017; Revised November 11, 2017; Accepted November 13, 2017

Background: Module detection is widely used to analyze and visualize biological networks. A number of methods and tools have been developed to achieve it. Meanwhile, bipartite module detection is also very useful for mining and analyzing bipartite biological networks and a few methods have been developed for it. However, there is few user-friendly toolkit for this task.

Methods: To this end, we develop an online web toolkit BMTK, which implements seven existing methods.

Results: BMTK provides a uniform operation platform and visualization function, standardizes input and output format, and improves algorithmic structure to enhance computing speed. We also apply this toolkit onto a drug-target bipartite network to demonstrate its effectiveness.

Conclusions: BMTK will be a powerful tool for detecting bipartite modules in diverse bipartite biological networks.

Availability: The web application is freely accessible at <http://www.zhanglabtools.net/BMTK>.

Keywords: network biology; module detection; biological bipartite networks

Author summary: Networks are becoming important tools for representing biological systems. BMTK is designed to detect bipartite modules for resolving bipartite biological networks. BMTK provides a uniform operation platform and visualization function.

INTRODUCTION

Module (or community) detection is useful for both decomposing and visualizing various types of biological association data such as detecting dense communities in genetic interaction networks, finding patterns representing protein complexes and pathways, mining tight interaction patterns in phosphorylation networks and so on [1–15]. Thus, a number of tools have been developed to identify modules such as CFinder [16], MINE [17], Linkcomm [18] and SurpriseMe [19]. Besides, several Cytoscape plugins have also been designed for such tasks [20].

Module detection in bipartite biological networks is also a key and valuable issue. For example, in transcriptional regulatory network, mining tight interactions between transcriptional factors (TFs) and target genes

could be perceived as a bipartite module detection problem, where TFs represent one type of node set and target genes represent another node set. Thus, it is an urgent task to develop efficient module detection methods for bipartite biological networks. A few methods have been developed for this task such as AdaptiveBrim [21], BCFinder [22], BiLPA [23], BiSBM [24], LeadingEigenvector [25], and LPBrim [26]. In addition, some are implemented as a plugin of Cytoscape, like NeMo [27], which combines hierarchical agglomerative clustering with neighbor-sharing score to detect diverse network communities. However, there is few user-friendly online tools to implement diverse methods and facilitate their applications. To this end, we design a web toolkit BMTK (bipartite module detection toolkit) to implement seven bipartite module detection methods. This toolkit unifies operation platform of different methods, standardizes

their input and output formats, improves program structures to speed up their computations, and provides online visualization function. BMTK will be a powerful tool for detecting bipartite modules in diverse bipartite biological networks.

METHODS

Here, we introduce the main seven methods implemented in BMTK. AdaptiveBrim employs BRIM (bipartite, recursively induced modules) with an adaptive and heuristic strategy to maximize the modularity to get an optimal partition [21]. The basic idea includes doubling the number of modules until modularity decreases and then using a bisection method to identify the maximal value of modularity between current and previous partition. BCFinder is based on an extension of the k -clique community detection algorithm. It detects all kinds of bicliques $K(m, n)$ and merges them to form communities [22]. BiLPA employs a heuristic adapted label propagation algorithm to optimize a new quantitative function called bipartite partition density to identify bipartite network modules [23]. BiSBM implements a projection-free and stochastic block model, which can be applied into k -partite networks directly [24]. It is a variant of the classic Kernighan-Lin algorithm. LeadingEigenvector adopts a spectral approach to determine the leading non-negative eigenvector of the modularity matrix of a network to extract corresponding dense modules [25]. LPBrim first employs a fast label propagation (LP) approach to find an initial module partition efficiently and then adopt a recursive induced method BRIM to determine the final module division [26]. This algorithm is designed for detecting modular structure in large-scale bipartite networks within reasonable time. LocalBM uses a greedy growth strategy to maximize a cohesiveness function. This strategy has been adopted for protein complex extraction in unipartite biological networks [28].

Implementation

BMTK is a user-friendly online software implemented by JDK1.6 (Figure 1 and Supplementary Materials) [21]. It is designed to pack diverse algorithmic implementations in Matlab, C++ or others into a unified Java package via mixed programming techniques. In BMTK, we first provide users a brief introduction, and further provide a tutorial on how to use BMTK including uploading a recognizable bipartite network file, running a method, and demonstrating and downloading identified modules. The key part is about running one of the methods. Users could select any of them by clicking corresponding method label.

We uniform the input formats of methods in BMTK

which supports either adjacency matrix or edge list of a bipartite network. The toolkit will automatically extract required information and transform it into corresponding format for each method which has its own input format originally.

After uploading the data, BMTK turns into a data summary page which provides some basic information about the uploaded bipartite network, including the number of nodes and edges, density of this network, and an interactive visualization of the network by Cytoscape Web. Next, by clicking the link below the summary, BMTK starts to apply a user-preferred method to the input network, and then automatically turns into the result page with a table containing the basic statistics of all modules. Moreover, we provide visualization of the detected modules by Cytoscape Web. Besides, we also give a global visualization of partition result in two figures by BiMat [29]. At last, all results can be downloaded for local analysis.

More detailed description about BMTK is introduced as follows.

Input data

We provide two input formats for a bipartite network in BMTK:

1) Edge list of the bipartite network (.txt)

The file contains two node columns separated by single tab or space. Nodes in each column belong to part $U = \{u_1, u_2, \dots\}$ and part $V = \{v_1, v_2, \dots\}$ of the bipartite network, respectively. Each row indicates one edge.

2) Adjacency matrix of the bipartite network (.txt)

The file contains the bipartite adjacency matrix, whose rows and columns respectively correspond to the nodes in two parts, and the values in it must be binary (0 or 1), where 1 indicates one edge exists.

Run

1) Open <http://zhanglabtools.net/BMTK>. Click “Run” section located in menu on the left, then select a method (e.g., AdaptiveBrim).

2) Upload network file through clicking “Browse” and select file format via “format” option.

3) After submitting data, the page will jump to “data summary”, which shows essential information of the bipartite network and check the reasonability of key parameters. Then click “Running” to detect modules of a given bipartite network.

4) After a while (the computing time depends on characteristics of the input network), the page will show basic statistics of the identified modules, reordered adjacency matrix and layout of the bipartite network based on the detected modules. The user can download

A Methods: AdaptiveBrim

AdaptiveBrim

This algorithm employs BRIM (bipartite, recursively induced modules) with an adaptive and heuristic strategy to maximize the modularity for determining the optimal bipartite network partition [21]. The basic idea includes doubling the number of modules until modularity decreases and then using a bisection method to identify the maximal value of modularity between current and previous partition.

Upload data.

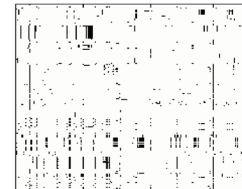
network file:

format: edgelist

B >> edgelist format

```
hsa:10008 D00294
hsa:10008 D02356
hsa:10060 D02356
hsa:10369 D00332
hsa:10369 D00437
hsa:10369 D00512
⋮ ⋮
```

C >> data summary



Rows: 201
Columns: 208
Edges: 1473
Density: 0.0352

D >> output result

"information.txt"

P = 201. - The number of nodes in part U.
Q = 208. - The number of nodes in part V.
NC = 24. - The number of detected modules.
Qb = 0.6775. - Standard modularity function: $(1/\text{num_edges}) * \text{trace}(rr' * bb * tt)$.
Qr = 0.62661. - $(\text{interactions inside modules} - \text{interactions outside modules}) / \text{total interactions}$.
Time = 2.3788. - Run time.

"moduleMembership.txt" "moduleSubnetwork.txt"

Node	Module ID	Node	Node	Module ID
hsa:11280	1	hsa:11280	D00110	1
hsa:3741	1	hsa:6331	D00110	1
hsa:3743	1	hsa:6336	D00110	1
⋮	⋮	⋮	⋮	⋮
hsa:2899	24	hsa:2899	D02546	24
hsa:2900	24	hsa:2900	D02546	24
hsa:2901	24	hsa:2901	D02546	24
D00110	1	hsa:6530	D00110	NaN
D00354	1	hsa:6531	D00110	NaN
D00358	1	hsa:6532	D00110	NaN
⋮	⋮	⋮	⋮	⋮
D01256	22	hsa:3752	D05156	NaN
D00537	23	hsa:9312	D05453	NaN
D02546	24	hsa:6530	D05458	NaN

E >> figures

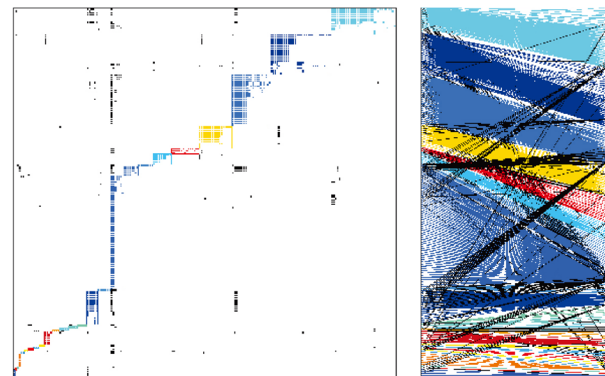


Figure 1. Illustration of BMTK. (A) Choose a method and there is a brief introduction about it. Then upload a network file, select the format of it and click "Submit". (B) There are two input formats for a bipartite network. One is the adjacency matrix and the other is the edge list. (C) After submitting the network data, the software will give a data summary and a black-white heatmap showing the adjacency matrix. (D) The result files include "information.txt", "moduleMembership.txt" and "moduleSubnetwork.txt". (E) Visualization of detected modules in two plots. One is about the reordered adjacency matrix, and the other is the layout of the given bipartite network based on the detected modules. If network size exceeds 500×500, we suggest users to use "moduleSubnetwork.txt" file to achieve module visualization in Cytoscape.

the module member nodes and module subnetworks stored in "moduleMembership.txt" and "moduleSubnetwork.txt", respectively.

Overlap test

We build an "overlap test" function in BMTK. It gives the consistency degree of modules detecting by two methods

in a heatmap. Hypergeometric test is used for all pairwise modules to test whether their overlaps are significant or not.

Tutorial and Demo

We have shown a detailed tutorial about BMTK. You can use the bipartite network via selecting "edgelist.txt" or "matrix.txt" in the "Demo" to test any method in BMTK.

Application

We first apply BMTK onto a set of simulated data with sizes of 100×120 , 250×300 , 500×600 and $1,000 \times 1,200$ and diverse numbers of embedded modules (Table 1 and Tables S1–S3). The links between nodes in each network is randomly generated with probability 0.3 within modules and 0.03 for others. The adjust rand index (ARI), a measure of agreement, is adopted to measure the module partition results against a given criterion. Generally, we find that BiLPA tends to find more detailed modules in each case. As to the time complexity, we can see that AdaptiveBrim is the most efficient one. LeadingEigenvector and LocalBM perform well for all situations. BiSBM is different from other five methods. It detects modules in two parts of a network separately with the given numbers of modules in both parts. Thus, it does not reveal the link affiliation with respect to modules in both parts. Generally, all the methods perform differently and user may need to select one method based on some domain knowledge or experience for further analysis. As BCFinder is based on an extension of the k -clique community detection algorithm which differs from other methods, we do not show its results in tables.

Moreover, we believe that bipartite module detection in a drug-target network will be helpful for understanding the modular characteristics of such a network (Figure 2A). Here, we apply BMTK to a drug-target bipartite network (Supplementary Figure S1) [30]. This network includes

208 drugs, 201 targets and 1,473 drug-target interactions. We detect a set of drug-target modules with each method (Table 2), and analyze the functions of drugs and targets (Supplementary Figures S2–S7 and Tables S4–S10). We find that in the same detected module, drugs often function in the same pathway and so do targets. For example, LeadingEigenvector identifies 14 modules, which contain 13 drugs and eight targets on average (Figure 2A and Supplementary Table S8).

For the left module in Figure 2B, it closely relates to antidiabetics (named as antidiabetics module). This module contains 13 drugs and three targets, in which 11 drugs are blood sugar related agents. Specifically, based on the KEGG DRUG database, D00593, D00336, D01599, D00380, D00219, D00379, D00335 and D01799 all belong to hypoglycemic agents. D00594, D01111 and D01854 are all antidiabetics. For the drug D01828 (amlexanox, an approved small molecule to treat aphthous ulcers and asthma), a previous study reported that using this drug for obese mice elevates energy expenditure resulting in improved insulin sensitivity and decreased steatosis. Thus, it is a promising candidate for obesity treatment. For the drug D00418, it is a kind of ATP-sensitive potassium (K_{ATP}) channel blocker. In pancreatic beta-cells, K_{ATP} channels regulate glucose-dependent insulin secretion. That is, when K_{ATP} channels open, beta-cells hyperpolarize and insulin secretion is suppressed. Thus, such K_{ATP} blockers could serve as the target for the treatment of type II diabetes. For

Table 1 Brief summary of detected modules for a simulated network of size $1,000 \times 1,200$ with 20 embedded modules

Method	#Modules	Modularity	ARI	Time (s)
AdaptiveBrim	7	0.2529	0.3452	5.7237
BiLPA	100	0.1270	0.3860	2648.5000
BiSBM	20, 20	N/A	N/A	2371.8400
LeadingEigenvector	20	0.2903	0.9675	1000.9164
LPBrim	13	0.2535	0.3716	37.4403
LocalBM	22	0.2964	0.9690	152.6102

Table 2 Brief summary of detected drug-target modules using six methods in BMTK

Method	#Modules	#Drugs	#Targets	Modularity
AdaptiveBrim	24	6	4	0.68
BCFinder [$K = (5, 5)$]	5	16	16	0.45
BiLPA	17	10	6	0.68
BiSBM*	Drug part, 17 Target part, 17	10	10	–
LeadingEigenvector	14	13	8	0.70
LPBrim	17	8	6	0.69
LocalBM	29	6	3	0.53

#Modules: the number of identified drug-target modules; #Drugs/#Targets: the number of drugs and targets in each module on average; Modularity: a quantitative index for module partition; *: note that the modules detected by BCFinder are related with the parameter K (here we just provide the results under $K = (5, 5)$); BiSBM may give different numbers of modules for both parts.

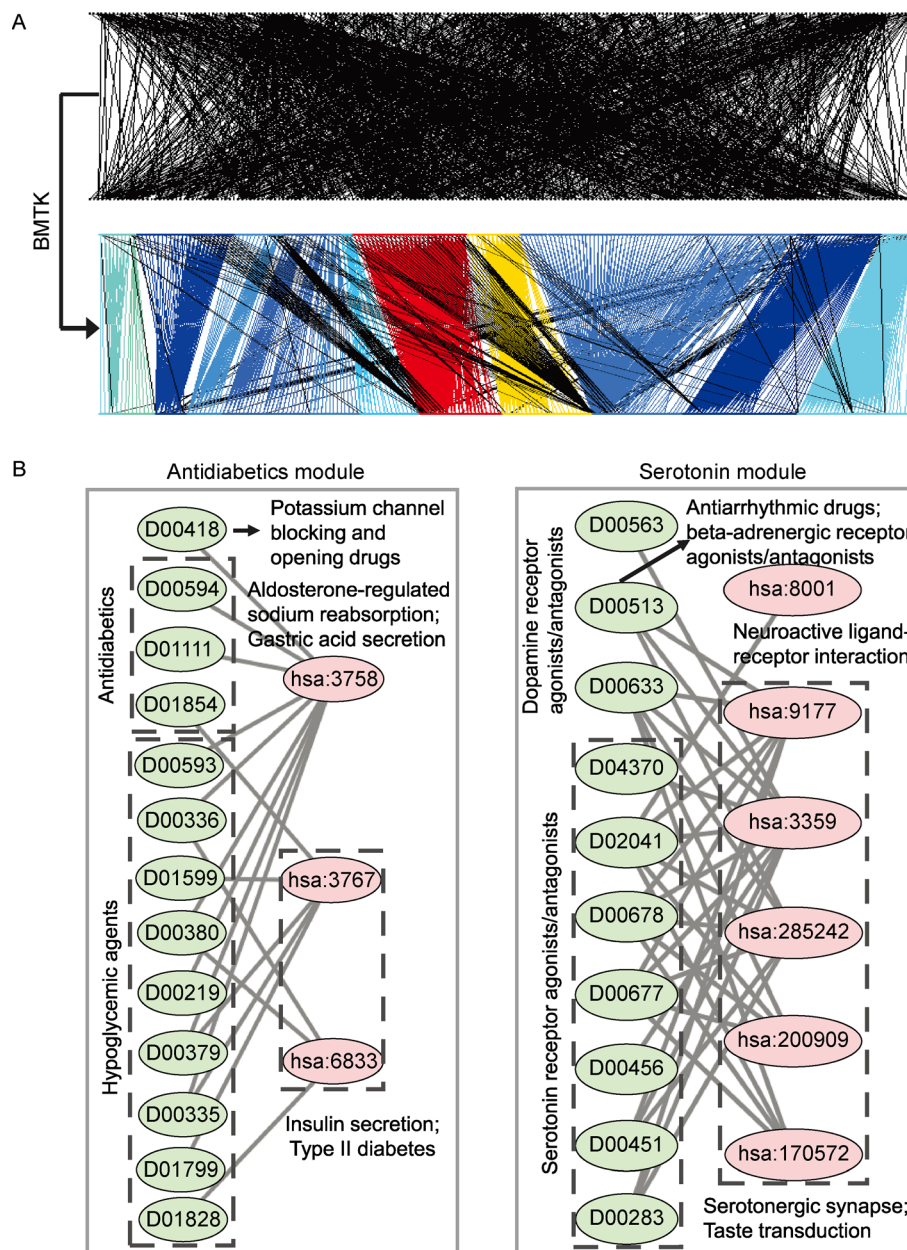


Figure 2. The bipartite modules computed by LeadingEigenvector. (A) A drug-target bipartite network with 14 identified drug-target modules marked in different color. (B) Two representative modules. The left one is closely related to antidiabetics, and the right one mainly involves in serotonin. The drugs and targets circled in dash lines have the same functions as described.

the targets hsa:3767 and hsa:6833, they are the members of type II diabetes mellitus and insulin secretion pathways. hsa:3785 (KCNQ2) encodes the potassium channel protein KvLQT2, which is also associated with insulin secretion stated as before. Besides, in the original drug-target bipartite network, it links to nine drugs, eight of which are hypoglycemic agents or antidiabetics. This also indicates that hsa:3785 probably involves in decreasing blood sugar levels. In summary, this module structure

detected by BMTK is very biologically meaningful and helpful to explore the relationships between drugs and target genes. Interestingly, the ninth detected module by BiLPA is the same as this one; and for LPBrim and AdaptiveBrim, this module is separated into the same two modules, one of which contains one target (hsa:3767) and its linked five drugs and the other one contains two targets (hsa:6833 and hsa:3758) and their linked eight drugs.

CONCLUSION

Here we develop a web toolkit BMTK, implementing seven methods originally programmed in different environments. It is a convenient and powerful tool for bipartite network module detection. It unifies input data format for all methods, optimizes each method, provides necessary data analysis, result visualization, and downloads. Moreover, we test it on a real drug-target bipartite network to show its effectiveness.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at 10.1007/s40484-018-0132-y.

ACKNOWLEDGEMENTS

This work has been supported by the National Natural Science Foundation of China (Nos. 61621003, 61422309, 61379092 and 11661141019), the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB13040600) and CAS Frontier Science Research Key Project for Top Young Scientist (QYZDB-SSW-SYS008).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Bei Wang, Jinyu Chen and Shihua Zhang declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- Carrella, D., Napolitano, F., Rispoli, R., Miglietta, M., Carissimo, A., Cuttillo, L., Sirci, F., Gregoretti, F. and Di Bernardo, D. (2014) Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics*, 30, 1787–1788
- Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñoz-Rascado, L., García-Sotelo, J. S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., *et al.* (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.*, 41, D203–D213
- Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489, 91–100
- Ryan, C. J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S. R., *et al.* (2012) Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell*, 46, 691–704
- Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., Bader, G. D. and Ferrin, T. E. (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, 12, 436
- Dinkel, H., Chica, C., Via, A., Gould, C. M., Jensen, L. J., Gibson, T. J. and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites – update 2011. *Nucleic Acids Res.*, 39, D261–D267
- Su, G., Kuchinsky, A., Morris, J. H., States, D. J. and Meng, F. (2010) GLay: community structure analysis of biological networks. *Bioinformatics*, 26, 3135–3137
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13, 2498–2504
- Frost, A., Elgort, M. G., Brandman, O., Ives, C., Collins, S. R., Miller-Vedam, L., Weibezahn, J., Hein, M. Y., Poser, I., Mann, M., *et al.* (2012) Functional repurposing revealed by comparing *S. pombe* and *S. cerevisiae* genetic interactions. *Cell*, 149, 1339–1352
- Zhang, S., Jin, G., Zhang, X. S. and Chen, L. (2007) Discovering functions and revealing mechanisms at molecular level from biological networks. *Proteomics*, 7, 2856–2869
- Li, Z., Zhang, S., Wang, R. S., Zhang, X. S. and Chen, L. (2008) Quantitative function for community detection. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 77, 036109
- Zhang, S., Ning, X. M., Ding, C. and Zhang, X. S. (2010) Determining modular organization of protein interaction networks by maximizing modularity density. *BMC Syst. Biol.*, 4, S10
- Zhang, S., Tian, D., Tran, N. H., Choi, K. P. and Zhang, L. (2014) Profiling the transcription factor regulatory networks of human cell types. *Nucleic Acids Res.*, 42, 12380–12387
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, 327, 425–431
- Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., Assmus, H. E., Andrade-Navarro, M. A. and Wanker, E. E. (2011) A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.*, 4, rs8
- Adamcsek, B., Palla, G., Farkas, I. J., Derényi, I. and Vicsek, T. (2006) CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22, 1021–1023
- Rhrissorakkrai, K. and Gunsalus, K. C. (2011) MINE: module identification in networks. *BMC Bioinformatics*, 12, 4581
- Kalinka, A. T. and Tomancak, P. (2011) linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*, 27, 2011–2012
- Aldecoa, R. and Marin, I. (2014) SurpriseMe: an integrated tool for network community structure characterization using Surprise maximization. *Bioinformatics*, 30, 1041–1042
- Szalay-Bekő, M., Palotai, R., Szappanos, B., Kovács, I. A. and Papp, B., Csermely, P. (2012) ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics*, 28, 2202–2204
- Barber, M. J. (2007) Modularity and community detection in bipartite networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*

- 76, 066102
22. Lehmann, S., Schwartz, M. and Hansen, L. K. (2008) Biclique communities. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 78, 016108
 23. Li, Z. P., Wang, R.-S., Zhang, S. and Zhang, X.-S. (2016) Quantitative function and algorithm for community detection in bipartite networks. *Inf. Sci.*, 367-368, 874–889
 24. Larremore, D. B., Clauset, A. and Jacobs, A. Z. (2014) Efficiently inferring community structure in bipartite networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 90, 012805
 25. Newman, M. E. J. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103, 8577–8582
 26. Liu, X. and Murata, T. (2009) Community detection in large-scale bipartite networks. *Information and Media Technologies*, 1, 50–57
 27. Rivera, C. G., Vakil, R. and Bader, J. S. (2010) NeMo: network module identification in Cytoscape. *BMC Bioinformatics*, 11, S61
 28. Nepusz, T., Yu, H. and Paccanaro, A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods*, 9, 471–472
 29. Flores, C. O., Poisot, T., Valverde, S. and Weitz, J. S. (2016) BiMat: a MATLAB package to facilitate the analysis of bipartite networks. *Methods Ecol. Evol.*, 7, 127–132
 30. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W. and Kanehisa, M. (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24, i232–i240