



OPEN

## On the statistical significance of communities from weighted graphs

Zengyou He<sup>1,2</sup>✉, Wenfang Chen<sup>1</sup>, Xiaoqi Wei<sup>1</sup> & Yan Liu<sup>1</sup>

Community detection is a fundamental procedure in the analysis of network data. Despite decades of research, there is still no consensus on the definition of a community. To analytically test the realness of a candidate community in weighted networks, we present a general formulation from a significance testing perspective. In this new formulation, the edge-weight is modeled as a censored observation due to the noisy characteristics of real networks. In particular, the edge-weights of missing links are incorporated as well, which are specified to be zeros based on the assumption that they are truncated or unobserved. Thereafter, the community significance assessment issue is formulated as a two-sample test problem on censored data. More precisely, the Logrank test is employed to conduct the significance testing on two sets of augmented edge-weights: internal weight set and external weight set. The presented approach is evaluated on both weighted networks and un-weighted networks. The experimental results show that our method can outperform prior widely used evaluation metrics on the task of individual community validation.

Community detection is a fundamental issue in network data analysis. It aims at dividing nodes in a network into different groups called communities. It is expected that there should be more edges within each community and few edges across different communities. The community detection procedure has been widely used in many fields such as social science, biology, medicine, and chemistry<sup>1</sup>.

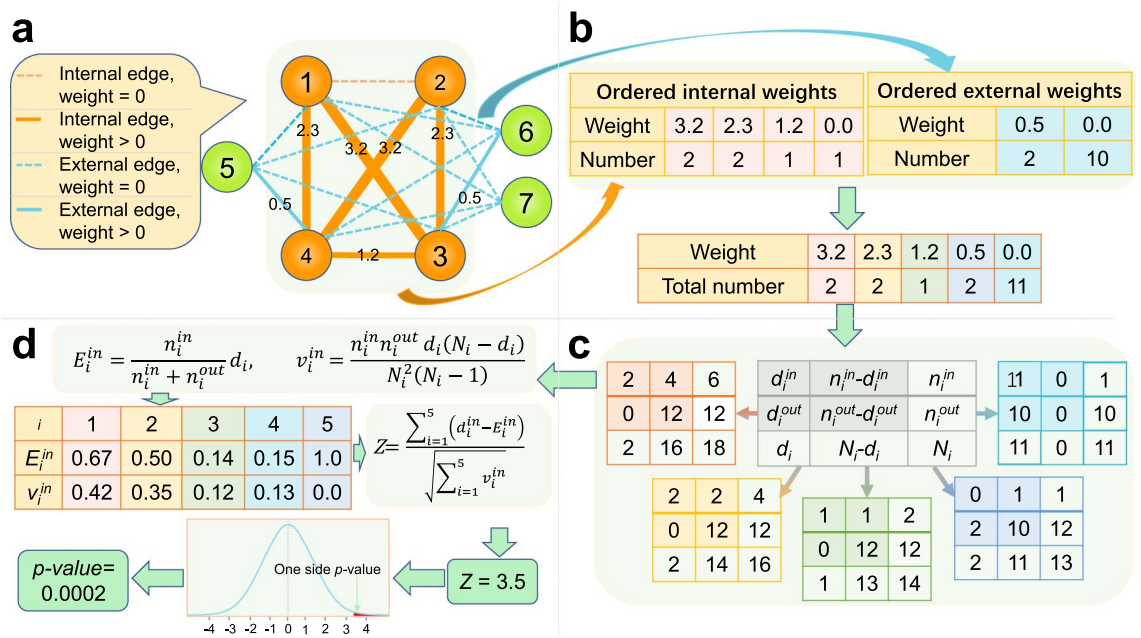
During the past decades, numerous community detection algorithms have been developed from different perspectives<sup>1–4</sup>. Despite these developments, the issue of deciding whether a derived community is real or not is far from being resolved. Such a community validation issue fits naturally into a framework of hypothesis testing, in which the null hypothesis is that the target community is not real.

In fact, many metrics such as modularity and conductance have been proposed for assessing the goodness of a potential community<sup>5</sup>. However, most of these metrics are not developed based on a rigorous significance testing procedure<sup>6</sup>. Theoretically, the realness of a community should be an analytical problem relative to some particular definitions of communities. Towards this direction, several research efforts have been conducted to analytically assess the realness of one candidate community, such as OSLOM<sup>7,8</sup>, ESSC<sup>9</sup>, DSC<sup>10</sup>, CCME<sup>11</sup>, and FOCS<sup>12</sup>. Among these methods, only OSLOM and CCME focus on validating a community in weighted networks. Unfortunately, in both OSLOM and CCME, the statistical significance of a target community is assessed through the probability of association between each node and the community<sup>13</sup>. One method that can directly test the realness of a community in edge-weighted graphs is still not available.

We formulate the community significance assessment problem in edge-weighted networks as a non-parametric two-sample test issue on censored data. In this paper, the edge-weights are assumed to be non-negative and continuous. The network structure and edge-weights may contain substantial measurement errors during the network inference process<sup>14</sup>. Based on this observation, we model each edge-weight as a censored observation in survival analysis<sup>15</sup>. If there is no edge between two nodes, the corresponding edge-weight is 0, which is either unobserved or truncated. Consequently, we construct two groups of edge-weights: one group is composed of edge-weights within the community and another group is composed of edge-weights between nodes in the community and remaining nodes outside the community. If the target community is not a real community, it is reasonable to expect that there is no difference between these two groups. Therefore, we can utilize a distribution-free two-sample test procedure in censored data analysis to assess the statistical significance of the candidate community. In this paper, we choose the popular Logrank test<sup>16</sup> to fulfill this task.

One of the characteristics of the presented formulation is that the unreliability of edge-weights is fully incorporated into the model. Meanwhile, it provides a general framework for validating weighted communities from

<sup>1</sup>School of Software, Dalian University of Technology, Dalian 116024, China. <sup>2</sup>Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116024, China. ✉email: zyhe@dlut.edu.cn



**Figure 1.** The main workflow of our method. (a) The network with augmented edge-weights for missing links. Each solid line represents a real edge and each dashed line denotes an augmented edge that is not included in  $E$ . In this network, there are totally 7 nodes and the candidate community is the sub-graph induced by the node set  $\{1,2,3,4\}$ . The size of each line is proportional to the corresponding edge-weight and the weights of augmented edges for missing links are zeros. (b) The ordered edge-weights. For the given candidate community in (a), we can construct two sets of edge-weights. One set is composed of the internal edge-weights and another set is composed of the external edge-weights. We can order these two sets to check if their empirical cumulative distribution functions are different. If the candidate community is not a real one, it can be expected that these two sets of weights have the same distribution. In Logrank test, these two sets are merged to generate a combined ordered edge-weight list. (c) Contingency tables for distinct weights. For each of the five distinct weights in (b), we can construct a contingency table. For instance, there are 2 edges whose weights are 3.2 and these two edges are internal edges, the first column of the leftmost table is (2, 0, 2). The second column of the leftmost table is (4, 12, 16) since there are 4 (and 12) internal (and external) edges whose weights are less than 3.2. (d) The final test statistic and its formula. Based on the contingency tables in (c), we can calculate the test statistic  $Z$  and the corresponding  $p$ -value to quantify the statistical significance of the candidate community. The Bonferroni correction is further applied to deliver an adjusted  $p$ -value, which is obtained by multiplying the original  $p$ -value with number of all possible communities of the same size.

a new angle. In particular, for un-weighted networks, we can reveal some interesting connections between our formulation and the community evaluation metric in Ref.<sup>17</sup>. We evaluate our method on both weighted networks and un-weighted networks. Experiments demonstrate that our method is comparable with prior state-of-the-art metrics on individual community assessment.

The rest of this article is arranged as follows. In the next section, our formulation is described in detail. Thereafter, the experimental results are presented and some discussions are given in the last two sections.

## Method

**Notations.** Given a weighted network  $G = (V, E, W)$ , where  $V$  is the node set,  $E$  is the edge set, and  $W$  is the set of positive weights. For a given subset of nodes  $S$  ( $S \subseteq V$ ), its induced subgraph  $G[S]$  can be regarded as a candidate community. All edges incident on the nodes in  $S$  could be divided into two groups: the set of internal edges within  $G[S]$  that are incident on two nodes from  $S$  and the set of external edges of  $G[S]$  which are incident on one node from  $S$  and another node from  $V \setminus S$ .

**Formulation.** To test if a given candidate community in the weighted network is a real one, we present a formulation based on the non-parametric two-sample test for censored data. The main workflow of our method is shown in Fig. 1, which will be elaborated below.

In the first step of our method, we regard each edge-weight as a censored observation. As shown in Fig. 1 a, the set of edge-weights is augmented by including the weights of missing links. The edge-weights of these missing links are 0s since they are either truncated or unobserved. Then, we can construct two augmented sets of edge-weights: the set of internal edge-weights  $W_S^{in} = \{w_1^{in}, w_2^{in}, \dots, w_{\binom{|S|}{2}}^{in}\}$  and the set of external edge-weights

$$W_S^{out} = \{w_1^{out}, w_2^{out}, \dots, w_{|S|(|V|-|S|)}^{out}\}.$$

Let  $F_{in}(\cdot)$  and  $F_{out}(\cdot)$  be complementary cumulative distribution functions for internal and external weights, respectively. Then, the community assessment issue can be modeled as a two-sample test problem, in which the hypotheses under consideration are:

$$H_0: F_{in}(w) = F_{out}(w);$$

$$H_1: F_{in}(w) > F_{out}(w), \text{ for at least one } w, \text{ where } w \geq 0 \text{ is the non-negative weight.}$$

Note that a larger edge-weight indicates a stronger connection between two corresponding nodes and both  $F_{in}(\cdot)$  and  $F_{out}(\cdot)$  are complementary cumulative distribution functions. If  $H_0$  is violated and  $H_1$  holds, then there will be more internal edges with positive edge-weights and the internal edges are associated with larger weights. Hence, the proposed two-sample test is capable of quantifying the realness of a target community in a statistically sound manner.

To solve the above two-sample test issue, many effective methods can be utilized. Here we adopt the Logrank test due to its popularity in censored data analysis. As shown in Fig. 1b,  $W_S^{in}$  and  $W_S^{out}$  are first merged to generate a new set  $W_S$  and then all edge-weights in  $W_S$  are sorted in a non-increasing order. For each distinct edge-weight  $w_i$  in  $W_S$ , we can construct a  $2 \times 2$  table as shown in Fig. 1c. In Fig. 1c,  $d_i^{in}$ ,  $d_i^{out}$  and  $d_i$  denote the number of internal edges, the number of external edges and the number of edges in  $W_S$  whose edge-weights are  $w_i$ . Meanwhile,  $n_i^{in}$ ,  $n_i^{out}$  and  $n_i$  denote the number of internal edges, the number of external edges and the number of edges in  $W_S$  whose edge-weights are not bigger than  $w_i$ .

Based on above notations, the test statistic  $Z$  of Logrank test is:

$$Z = \frac{\sum_{i=1}^q (d_i^{in} - E_i^{in})}{\sqrt{\sum_{i=1}^q v_i^{in}}}, \tag{1}$$

where  $q$  is the number of distinct edge-weights in  $W_S$ ,  $N_i = n_i^{in} + n_i^{out}$ ,  $E_i^{in} = \frac{n_i^{in}}{N_i} d_i$  and  $v_i^{in} = \frac{n_i^{in} d_i (N_i - d_i) n_i^{out}}{(N_i)^2 (N_i - 1)}$ . When  $H_0$  is true,  $Z$  approximately follows a  $N(0, 1)$  distribution. In Eq. (1), each  $(d_i^{in} - E_i^{in})$  can be regarded as the ‘‘observed minus expected’’ difference with respect to the number of internal edges of a specific weight. Thus, a real community should be associated with a large  $Z$  statistic. Based on this approximation, the corresponding  $p$ -value can be obtained to assess the statistical significance of  $G[S]$ .

In regard to the combinatorial nature of community detection, the community significance assessment issue is actually a multiple hypothesis testing problem. Hence, we need to conduct a multiple testing correction by calculating an adjusted  $p$ -value for each candidate community. The most popular method for multiple testing correction is probably the Bonferroni correction approach, in which the original  $p$ -value is multiplied by the number of tested hypotheses to obtain an adjusted  $p$ -value. In our context, the number of tested hypotheses can be calculated as the number of possible communities of the same size. More precisely, for a given community  $G[S]$  of size  $|S|$ , its adjusted  $p$ -value is calculated as  $p_{adj}(G[S]) = \min\{1, p(G[S]) \times \binom{|V|}{|S|}\}$ , where  $p(G[S])$  is original  $p$ -value of  $G[S]$  and  $|V|$  is the number of nodes of the graph. In the experiment, the adjusted  $p$ -value is used instead of the original  $p$ -value in our method.

**Un-weighted special case.** For un-weighted networks, the edge-weight is either 1 or 0, and thus  $q = 2$  in Eq. (1). For the weight  $w_1 = 1$ , we have:  $n_1^{in} = \binom{|S|}{2}$ ,  $n_1^{out} = |S|(|V| - |S|)$ , and thus

$$E_1^{in} = \frac{n_1^{in}}{N_1} d_1 = \frac{\binom{|S|}{2}}{\binom{|S|}{2} + |S|(|V| - |S|)} d_1, \quad v_1^{in} = \frac{n_1^{in} n_1^{out} d_1 (N_1 - d_1)}{(N_1)^2 (N_1 - 1)} = \frac{\binom{|S|}{2} |S|(|V| - |S|) d_1 \left[ \binom{|S|}{2} + |S|(|V| - |S|) - d_1 \right]}{\left[ \binom{|S|}{2} + |S|(|V| - |S|) \right]^2 \left[ \binom{|S|}{2} + |S|(|V| - |S|) - 1 \right]}.$$

For the weight  $w_2 = 0$ , we have:  $n_2^{in} = d_2^{in}$ ,  $n_2^{out} = d_2^{out}$ ,  $N_2 = d_2$ , and thus  $E_2^{in} = \frac{n_2^{in}}{N_2} d_2 = d_2^{in}$ ,  $v_2^{in} = 0$ . Thus, Eq. (1) can be rewritten as:

$$\begin{aligned} Z &= \frac{(d_1^{in} - E_1^{in}) + (d_2^{in} - E_2^{in})}{\sqrt{(v_1^{in} + v_2^{in})}} = \frac{(d_1^{in} - E_1^{in})}{\sqrt{v_1^{in}}} = \frac{\left( d_1^{in} - \frac{\binom{|S|}{2}}{\binom{|S|}{2} + |S|(|V| - |S|)} d_1 \right)}{\sqrt{\frac{\binom{|S|}{2} |S|(|V| - |S|) d_1 \left[ \binom{|S|}{2} + |S|(|V| - |S|) - d_1 \right]}{\left[ \binom{|S|}{2} + |S|(|V| - |S|) \right]^2 \left[ \binom{|S|}{2} + |S|(|V| - |S|) - 1 \right]}}} \\ &= \sqrt{\frac{\left( \frac{\binom{|S|}{2}}{d_1} (|S|(|V| - |S|)) \left[ \binom{|S|}{2} + |S|(|V| - |S|) - 1 \right] \right)}{\left[ \binom{|S|}{2} + |S|(|V| - |S|) - d_1 \right]}} \left[ \frac{d_1^{in}}{\binom{|S|}{2}} - \frac{d_1^{out}}{|S|(|V| - |S|)} \right] \\ &= \sqrt{\frac{\left( \frac{\binom{|S|}{2}}{(|S|(|V| - |S|)) d_1} \left[ \binom{|S|}{2} + |S|(|V| - |S|) - 1 \right] \right)}{}} |S|(|V| - |S|) \left[ \frac{d_1^{in}}{\binom{|S|}{2}} - \frac{d_1^{out}}{|S|(|V| - |S|)} \right]. \end{aligned} \tag{2}$$

In Ref.<sup>17</sup>, a new community evaluation metric has been presented, which can be expressed as follows based on our notations:

$$Z' = |S|(|V| - |S|) \left[ \frac{d_1^{in}}{\frac{|S|^2}{2}} - \frac{d_1^{out}}{|S|(|V| - |S|)} \right]. \quad (3)$$

Thus, we can get the mathematical relation between  $Z$  (our test statistic for un-weighted networks) and  $Z'$  (the community evaluation metric in Ref.<sup>17</sup>):

$$Z \approx Z' \sqrt{\frac{\binom{|S|}{2} \left[ \binom{|S|}{2} + |S|(|V| - |S|) - 1 \right]}{(|S|(|V| - |S|))d_1 \left[ \binom{|S|}{2} + |S|(|V| - |S|) - d_1 \right]}}. \quad (4)$$

As shown in Eq. (4), we can quantitatively establish the connection between the special case of our method for un-weighted networks and one popular metric in the literature.

## Results

To test whether the presented  $p$ -value is effective on community evaluation, we conduct a series of experiments according to the pipeline shown in Fig. 2. Firstly, existing community detection algorithms are employed to produce a set of identified communities on networks with ground-truth communities. Then, we use both internal validation metrics (e.g. our method, modularity) and external validation metrics (e.g. precision, recall) to quantitatively validate each identified community. Since external validation metrics are calculated based on the ground-truth information, which can be used as the “gold standard”. In other words, one internal validation metric is a good community validation index if it is highly correlated with each external validation metric on the assessment of identified communities. Based on this assumption, we calculate the Pearson’s correlation coefficient between two vectors (one is generated from an internal validation metric and another one is produced by an external validation metric), where each vector is composed of the validation index values on a set of identified communities. Finally, the Friedman test<sup>18</sup> and three post-hoc tests: the Nemenyi test<sup>19</sup>, the Bonferroni–Dunn test<sup>20</sup> and the Holm’s step-down test<sup>21</sup> are employed to check if our method is significantly better than other popular internal validation metrics.

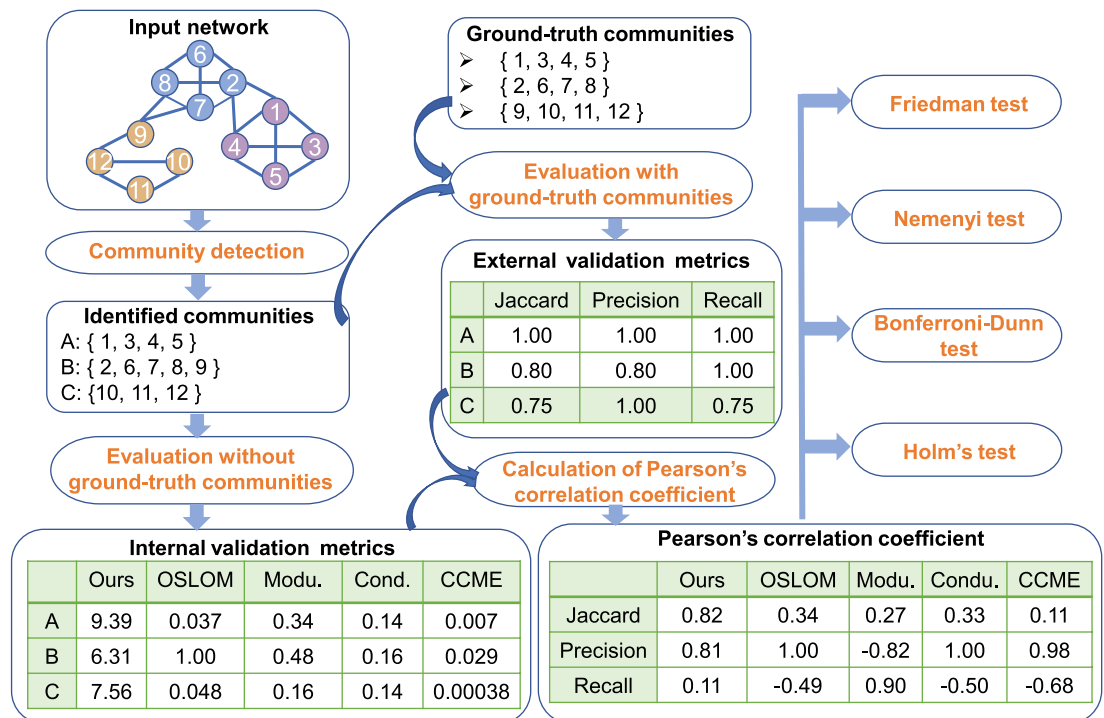
**Data sets.** In our experiment, we use two groups of data sets. One group is composed of four weighted PPI (Protein–Protein Interaction) networks: Collins2007<sup>22</sup>, Gavin2006<sup>23</sup>, Krogan2006\_core<sup>24</sup>, Krogan2006\_extended<sup>24</sup>. There are three sets of ground-truth communities for these weighted PPI networks, where each set is collected from one of the following databases of protein complexes: CYC2008<sup>25</sup>, MIPS<sup>26</sup> and SGD<sup>27</sup>. There are 408, 203 and 323 ground-truth communities in these three sets, respectively. Another group is composed of six real un-weighted networks: Karate<sup>28</sup>, Football<sup>29</sup>, Personal Facebook (Personal)<sup>9</sup>, Political blogs (PolBlogs)<sup>30</sup>, Books about US politics (PolBooks)<sup>31</sup>, and Railways<sup>32</sup>. The topological characteristics of six real un-weighted networks are provided in Supplementary Table S1.

**Parameter setting.** We choose three classical community detection methods: SLPaw<sup>33</sup>, Infomap<sup>34</sup> and Louvain<sup>35</sup> to detect communities. In our experiment, we run these three methods with their default parameter settings for weighted graphs.

**Experiment.** We compare our method with four internal metrics: conductance<sup>36</sup>, modularity<sup>37</sup>,  $p$ -value in OSLOM<sup>7</sup> and  $p$ -value in CCME<sup>11</sup>. The  $p$ -value of OSLOM is obtained with its default setting and the  $p$ -value of a community in CCME is the maximal  $p$ -value of all nodes within the community. Since Conductance, the  $p$ -values in our method, OSLOM and CCME are negatively correlated with Jaccard coefficient, Precision and Recall, we use the negative Pearson’s correlation coefficient for these three metrics in the performance comparison. Then, for the set of reported communities on each data set from each community detection algorithm, we can use the Pearson’s correlation coefficient with respect to each external validation metric to check which internal validation metric is better. More precisely, a better internal validation metric should have a larger correlation coefficient. The detailed results are recorded in the Supplementary Tables S2–S5. From these tables, we can obtain the rank distribution for each internal validation metric, where a larger correlation coefficient will be assigned to a smaller rank. The rank distributions on weighted PPI networks and un-weighted networks in terms of box plots are provided in Fig. 3a,b, respectively.

From Fig. 3, it can be observed that our method can achieve the smallest average rank among five internal validation metrics. To check if our method is really better than the other four internal validation metrics, we first apply the Friedman test to assess the null hypothesis that all methods have the same rank. The  $\chi_F^2$  value in the Friedman test on weighted PPI networks and un-weighted networks is 200.5704 and 45.6889, respectively. This means that the performance gaps among different internal validation metrics are statistically significant when the significance level is specified to be 0.05. Then, we further employ the Nemenyi test, the Bonferroni–Dunn test and the Holm’s step-down test to compare our method with each competing internal validation metric in a pair-wise manner.

In Table 1, we record the rank difference between our method and each competing method on both weighted PPI networks and un-weighted networks. As shown in Table 1, the rank difference values between our method



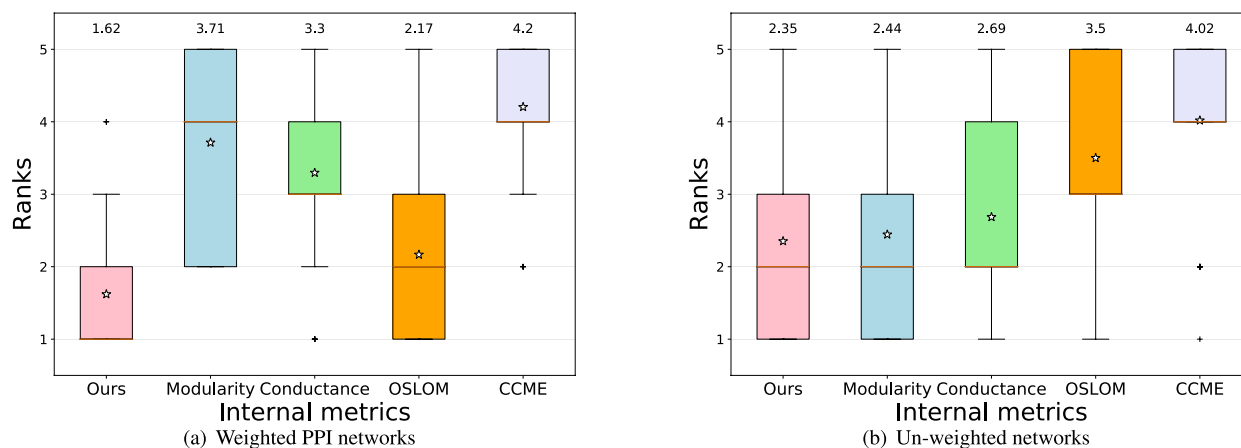
**Figure 2.** The main workflow of our experiments. Given a network with three ground-truth communities: {1, 3, 4, 5}, {2, 6, 7, 8} and {9, 10, 11, 12}, three communities are detected by the SLPaw algorithm:  $A = \{1, 3, 4, 5\}$ ,  $B = \{2, 6, 7, 8, 9\}$ ,  $C = \{10, 11, 12\}$ . Each identified community can be assessed using both the internal validation metric and the external validation metric. For example, the negative log value of our method and the precision for the identified community A is 9.39 and 1, respectively. Then, a vector for each metric on the set of identified communities is readily available. For instance, the validation index vector for our method and the Jaccard coefficient is (9.39, 6.31, 7.56) and (1, 0.8, 0.75), respectively. Consequently, the Pearson's correlation coefficient is calculated between each pair of vectors: one from an internal validation metric and another one from an external validation metric. Based on the correlation coefficients, four statistical tests are further applied to check if our method is really better than other internal validation metrics on the task of individual community assessment.

and Modularity, Conductance and CCME on weighted PPI networks are larger than the critical difference (CD) thresholds for both the Nemenyi test and the Bonferroni–Dunn test when the significance level is 0.05. This indicates that our method is significantly better than Conductance, Modularity and CCME under these two tests. On the un-weighted networks, our method is significantly better than OSLOM and CCME according to the Bonferroni–Dunn test and the Nemenyi test.

In Table 2, we list the  $p$ -values for the average rank difference between our method and each competing method based on the Holm's step-down test. Besides, the adjusted significance level for each position after sorting the  $p$ -values in a non-decreasing order is provided as well. As shown in Table 2, all the  $p$ -values on PPI networks are smaller than the corresponding adjusted significance levels. This indicates the superiority of our method over other four metrics on weighted networks is also confirmed by the Holm's step-down test. Similar to results in Table 1, we can claim that our method is significantly better than OSLOM and CCME on un-weighted networks based on the hypothesis testing results in Table 2.

## Discussion

We have presented a general approach for assessing the statistical significance of a community from weighted networks. In this new formulation, the weights of missing links are set to be zeros and all edge-weights are treated as truncated observations. Based on this assumption, the community validation issue is modeled as a two-sample test problem on censored data. The presented formulation provides a general framework for community validation from a significance testing perspective. Based on this framework, we can either reveal the rationale underlying some existing community validation metrics or develop new community evaluation measures.



**Figure 3.** The rank distribution for each internal validation metric on (a) weighted PPI networks and (b) Un-weighted networks. Here the box plot is used to depict the distributions: minimum (lower line), lower quartile (lower edge of the box), median (center line), upper quartile (upper edge of the box) and maximum (upper line). Meanwhile, mean (star) and outlier (cross) are also plotted.

	Weighted PPI networks	Un-weighted networks
RD (Ours, modularity)	2.0926	0.0926
RD (Ours, conductance)	1.6759	0.3333
RD (Ours, OSLOM)	0.5463	1.1481
RD (Ours, CCME)	2.5833	1.6667
CD (Nemenyi)	0.8301	0.5870
CD (Bonferroni–Dunn)	0.7601	0.5375

**Table 1.** The average rank difference  $RD(a, b)$  between two methods, where  $a$  denotes our method and  $b$  represents a competing internal validation metric. In last two rows, the critical difference (CD) thresholds at the significance level  $\alpha = 0.05$  for the Nemenyi test and the Bonferroni–Dunn test are listed, respectively.

	Weighted PPI networks	Adjusted $\alpha$	Un-weighted networks	
P (Ours, modularity)	0	0.0125	2.16E–08	P (Ours, CCME)
P (Ours, CCME)	0	0.0167	8.06E–05	P (Ours, OSLOM)
P (Ours, conductance)	3.33E–15	0.0250	0.1367	P (Ours, conductance)
P (Ours, OSLOM)	5.56E–03	0.0500	0.3805	P (Ours, modularity)

**Table 2.** The  $p$ -value  $P(a, b)$  for the rank difference between two methods under the Holm’s step-down test on weighted PPI networks (the left column) and un-weighted networks (the right column), where  $a$  denotes our method and  $b$  represents a competing internal validation metric. In the middle column, the adjusted significance level at  $\alpha = 0.05$  for each position after sorting the  $p$ -values in a non-decreasing order is listed.

Received: 17 July 2021; Accepted: 21 September 2021  
Published online: 13 October 2021

## References

- Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
- Orman, G. K., Labatut, V. & Cherifi, H. Comparative evaluation of community detection algorithms: A topological approach. *J. Stat. Mech. Theory Exp.* **2012**, P08001. <https://doi.org/10.1088/1742-5468/2012/08/p08001> (2012).
- Dao, V. L., Bothorel, C. & Lenca, P. Community structure: A comparative evaluation of community detection methods. *Netw. Sci.* **8**, 1–41. <https://doi.org/10.1017/nws.2019.59> (2020).
- Jebabli, M., Cherifi, H., Cherifi, C. & Hamouda, A. Community detection algorithm evaluation with ground-truth data. *Physica A* **492**, 651–706. <https://doi.org/10.1016/j.physa.2017.10.018> (2018).
- Chakraborty, T., Dalmia, A., Mukherjee, A. & Ganguly, N. Metrics for community analysis: A survey. *ACM Comput. Surv.* **50**, 1–37 (2017).
- Zhang, P. & Moore, C. Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18144–18149 (2014).
- Lancichinetti, A., Radicchi, F., Ramasco, J. J. & Fortunato, S. Finding statistically significant communities in networks. *PLoS ONE* **6**, e18961 (2011).

8. Lancichinetti, A., Radicchi, F. & Ramasco, J. J. Statistical significance of communities in networks. *Phys. Rev. E*. <https://doi.org/10.1103/PhysRevE.81.046110> (2010).
9. Wilson, J. D., Wang, S., Mucha, P. J., Bhamidi, S. & Nobel, A. B. A testing based extraction algorithm for identifying significant communities in networks. *Ann. Appl. Stat.* **8**, 1853–1891 (2014).
10. He, Z., Liang, H., Chen, Z., Zhao, C. & Liu, Y. Detecting statistically significant communities. *IEEE Trans. Knowl. Data Eng.* <https://doi.org/10.1109/TKDE.2020.3015667> (2020).
11. Palowitch, J., Bhamidi, S. & Nobel, A. B. Significance-based community detection in weighted networks. *J. Mach. Learn. Res.* **18**, 6899–6946 (2018).
12. Palowitch, J. Computing the statistical significance of optimized communities in networks. *Sci. Rep.* **9**, 1–10 (2019).
13. He, Z., Liang, H., Chen, Z., Zhao, C. & Liu, Y. Computing exact p-values for community detection. *Data Min. Knowl. Discuss.* **34**, 833–869 (2020).
14. Newman, M. E. Network structure from rich but noisy data. *Nat. Phys.* **14**, 542–545 (2018).
15. Klein, J. P. & Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data* (Springer, 2003).
16. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50**, 163–170 (1966).
17. Zhao, Y., Levina, E. & Zhu, J. Community extraction for social networks. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7321–7326 (2011).
18. Mahrer, J. M. & Magel, R. C. A comparison of tests for the k-sample, non-decreasing alternative. *Stats Med.* **14**, 863–871 (2010).
19. Nemenyi, P. Distribution-free multiple comparisons. *Biometrics* **18**, 263 (1962).
20. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).
21. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
22. Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6**, 439–450 (2007).
23. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
24. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
25. Pu, S., Wong, J., Turner, B., Cho, E. & Wodak, S. J. Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res.* **37**, 825–831 (2009).
26. Mewes, H.-W. *et al.* MIPS: Analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
27. Hong, E. L. *et al.* Gene Ontology annotations at SGD: New data sources and annotation methods. *Nucleic Acids Res.* **36**, D577–D581 (2008).
28. Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473 (1977).
29. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826 (2002).
30. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 us election: Divided they blog. In *Proc. 3rd International Workshop on Link Discovery*, 36–43 (2005).
31. Krebs, V. Social network analysis software & services for organizations, communities, and their consultants (2013). <http://www.orgnet.com>
32. Chakraborty, T., Srinivasan, S., Ganguly, N., Mukherjee, A. & Bhowmick, S. On the permanence of vertices in network communities. In *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1396–1405 (2014).
33. Xie, J., Szymanski, B. K. & Liu, X. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In *IEEE 11th International Conference on Data Mining Workshops*, 344–349 (2011).
34. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118–1123 (2008).
35. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
36. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000).
37. Newman, M. E. Analysis of weighted networks. *Phys. Rev. E* **70**, 056131 (2004).

## Acknowledgements

This work has been supported by the Natural Science Foundation of China under Grant Nos. 61972066 and 61572094 and the Fundamental Research Funds for the Central Universities (No. DUT20YG106).

## Author contributions

Z.H. and W.C. designed research, performed research, and wrote the paper; X.W. and Y.L. analyzed data.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99175-2>.

**Correspondence** and requests for materials should be addressed to Z.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021