Metrics for Community Analysis: A Survey

TANMOY CHAKRABORTY, Indraprastha Institute of Information Technology, Delhi (IIIT-D), India
AYUSHI DALMIA, IBM Research, India
ANIMESH MUKHERJEE and NILOY GANGULY, IIT Kharagpur, India

Detecting and analyzing dense groups or *communities* from social and information networks has attracted immense attention over the last decade due to its enormous applicability in different domains. Community detection is an *ill-defined problem*, as the nature of the communities is not known in advance. The problem has turned even more complicated due to the fact that communities emerge in the network in various forms such as disjoint, overlapping, and hierarchical. Various heuristics have been proposed to address these challenges, depending on the application in hand. All these heuristics have been materialized in the form of new *metrics*, which in most cases are used as optimization functions for detecting the community structure, or provide an indication of the goodness of detected communities during evaluation. Over the last decade, a large number of such metrics have been proposed. Thus, there arises a need for an organized and detailed survey of the metrics proposed for community detection and evaluation. Here, we present a survey of the start-of-the-art metrics used for the detection and the evaluation of community structure. We also conduct experiments on synthetic and real networks to present a comparative analysis of these metrics in measuring the goodness of the underlying community structure.

CCS Concepts: • **Human-centered computing → Social network analysis**; • **Information systems →** *Clustering*;

Additional Key Words and Phrases: Metrics, community discovery, community evaluation

ACM Reference format:
Tanmoy Chakraborty, Ayushi Dalmia, Animesh Mukherjee, and Niloy Ganguly. Metrics for Community Analysis: A Survey. *ACM Comput. Surv.* 50, 4, Article 54 (August 2017), 37 pages.
https://doi.org/10.1145/3091106

1 INTRODUCTION

Analyzing the community structure of networks has received tremendous attention over the last decade across different disciplines such as computer science, physics, biology, and sociology. Although community is an *ill-defined* concept [43], a general consensus suggests that community structure is a decomposition of nodes into groups, where within a group nodes are highly connected, and across groups nodes are loosely connected [49]. Communities are formed due to the structural or functional similarities among the vertices in the network [94]. Therefore,

Both the authors have equal contribution. A part of the research was done when the first author was at University of Maryland, College Park, USA.

Authors' addresses: T. Chakraborty is with Dept. of Computer Science & Engg., Indraprastha Institute of Information Technology, Delhi, India; email: tanmoy@iiitd.ac.in; A. Dalmia is with IBM Research, Delhi, India; email: ayushidalmia@in.ibm.com; A. Mukherjee and N. Ganguly are with Dept. of Computer Science & Engg., Indian Institute of Technology Kharagpur, India; emails: {animeshm, niloy}@cse.iitkgp.ernet.in.

https://doi.org/10.1145/3091106

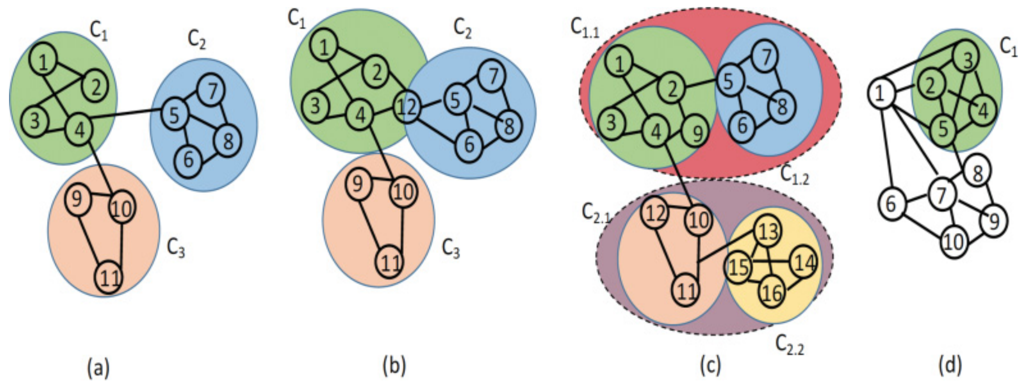

Fig. 1. Toy example illustrating different types of communities: (a) nonoverlapping, (b) overlapping, (c) hierarchical, and (d) local.

understanding the decomposition of nodes into communities in a network provides a high-level view of the formation of the network structure through the interactions of nodes having an identical nature [1, 11, 39, 80].

Communities in real-world networks are of different kinds: disjoint or nonoverlapping (e.g., students belonging to different disciplines in an institute) [43], overlapping (e.g., a person having membership in different social groups on Facebook) [16, 136], hierarchical (e.g., cells in the human body form tissues that in turn form organs and so on) [9], and local (e.g., a person having *uneven* interaction between certain members within a social group in Facebook) [111]. Figure 1 illustrates a toy example of these four types of communities. In case of nonoverlapping communities, each node belongs to only one community, while for overlapping communities, nodes can belong to multiple communities (e.g., node 12 in Figure 1(b) belongs to both C_1 and C_2). Hierarchical communities are those structures where one can view the graph at multiple levels and choose the community structure at a particular hierarchy. Depending on the application at hand, we can view the network at a particular hierarchical level. Local communities are communities that do not show any structure when viewed globally but possess different structures from a local perspective. Note that in this survey, we mostly focus our discussion on nonoverlapping and overlapping community structures because these are the most general and widely used form of community structures.[1]

The task of community analysis is performed in two separate phases: first, *detection* of meaningful community structure from a network, and second, *evaluation* of the appropriateness of the detected community structure. Since there is a lack of consensus on the definition of a community structure, multiple ideas emerged, which in turn resulted in different definitions of community structure. Every definition of community is justified in terms of different metrics formulated. Therefore, analyzing such metrics is crucial to understanding the development of the studies pertaining to community analysis.

The ill-posed definition of the community structure makes the evaluation framework complicated in the sense that there is no universally accepted metric for evaluating a community detection algorithm. There has been plenty of metrics proposed in the literature that are qualified to evaluate a community detection algorithm on their own merit. Interestingly, these quality metrics

[1]Metrics that are used for disjoint and overlapping communities can also be applicable for local communities and hierarchical communities (at different levels separately).

not only are useful to evaluate the community structure but also can be incorporated into various algorithms to discover the community structure of a network. The evaluation of the communities detected by an algorithm from a network becomes easier if the actual (ground-truth) community structure of the network is known a priori. In this direction, attempts were made to construct artificially generated networks with an inherent community structure [71]. To make a correspondence between the detected and the ground-truth community structures, a few metrics were borrowed from the literature of *clustering* in data mining [12] and reformulated by incorporating the network information.

Although there have been studies surveying different approaches mostly on the *detection* of nonoverlapping [43, 69, 116] and overlapping [57, 136] communities separately, no attempts have been initiated to understand thoroughly *the metrics used to design the algorithms and to evaluate the goodness of a community structure*. Metrics help us in understanding the goodness of an algorithm in a quantitative manner. We believe that evaluating a community detection algorithm may be difficult because it requires selection of a quality function among the many that are contradictory in nature. The network structure and the shape of the underlying community structure may drastically differ from one case to another. It is therefore of utmost importance to select the right metric depending on the purpose and properties of the given network. Moreover, most of the metrics are extensions of some old measures. Therefore, we need to understand the evolutionary route of a derived metric and its possible extensions.

Here, we conduct a comprehensive survey on the state-of-the-art metrics used for detecting and evaluating the community structure. More importantly, we bring together metrics related to all the classes of community structures (disjoint, overlapping, local, hierarchical, etc.) into a single article that can help us understand the derivation of one measure into another, with special emphasis on metrics for nonoverlapping and overlapping community detection. Specifically, in case of metrics related to community detection, we study traditional metrics like Modularity and its variants, as well as more recently proposed metrics including Permanence, Surprise, Significance, and Flex. In case of community evaluation, we study various state-of-the-art metrics such as Normalized Mutual Information, Purity, Rand Index, and F-Measure. Finally, a comparative analysis is presented based on the experiments conducted on synthetically generated and real-world networks.

Note, unless otherwise stated, we use the following terms interchangeably in this article: graph and network; node and vertex; edge and link; community, cluster, and partition.

2 METRICS FOR DISCOVERING COMMUNITY STRUCTURE

The aim of community detection is to discover inherent community structure from a network. However, as mentioned before, the definition of a community is not well defined. What is a good community? Let us consider a network $G(V, E)$, where V denotes the set of vertices and E denotes the set of edges. One can have an exponentially large number of possible community structures in the graph G. Finding these community structures by optimizing a metric, say, modularity, is an NP-Hard problem [28]. Moreover, not all partitions of a graph are equally good. In order to obtain the best partition of the graph and thus significant communities, most of the community detection algorithms aim to optimize a *goodness metric* that essentially indicates the quality of the communities detected from the network. The goal of the community detection algorithm would be to obtain the best partition of the network that would optimize the metric. A wide variety of such metrics were proposed that can detect the quality of the communities obtained for a given partition. This section presents a detailed study on the large number of goodness metrics proposed by the graph mining algorithms for community detection. Note that apart from their extensive usage for community detection in different algorithms, these metrics are popularly used to measure the goodness of a discovered community structure. We shall discuss this usage in Section 3.

2.1 Metrics Used for Non-Overlapping Community Detection

The nonoverlapping community detection algorithms aim at partitioning the vertices of a network, $G(V, E)$ into K nonempty, mutually exclusive groups, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$, such that each vertex must be a part of *exactly one* community, that is, $|\omega_1| + |\omega_2| + \cdots + |\omega_K| = |V|$. $|\omega_i|$ indicates the number of nodes in the community ω_i.

2.1.1 Simple Metrics Based on Topological Properties of the Network. Various simple metrics capturing the topological properties of the network have been proposed in the past that are used to compute the quality of communities detected.

Let us assume a function $f(\omega)$ that indicates the goodness of the community on the basis of the connectivity of nodes in it. [138] summarized these scoring functions and grouped them into the following four broad classes:

(A) *Functions considering the internal connections only*: The following functions consider the edges within the community ω:

- **Internal density:** $f(\omega) = \frac{|E_\omega^{in}|}{|\omega|(|\omega|-1)/2}$ computes the density of edges in a community and is calculated as the ratio of the number of edges completely internal to ω and the total number of possible edges within the community [110].
- **Edge inside:** $f(\omega) = |E_\omega^{in}|$ measures the total number of edges completely internal to ω [110].
- **Average degree:** $f(\omega) = 2|E_\omega^{in}|/|\omega|$ computes the average degree of the nodes of ω by considering edges within it only [110].
- **Fraction over median degree (FOMD):** $f(\omega) = \frac{|u:u\in\omega, |(u,v):v\in\omega| > d_m|}{|\omega|}$ computes how tightly knit the community is by computing the number of nodes in ω that have internal degree greater than d_m, the median value of the degree of all the nodes in V. This score is normalized by the total number of nodes within the community.
- **Triangle Participation Ratio (TPR):** It computes the fraction of nodes that are part of a triangular motif or triad: $f(\omega) = \frac{|u:u\in\omega, \{v,w\in\omega, (u,v)\in E, (u,w)\in E, (v,w)\in E\} \neq \phi|}{|\omega|}$ for ω.

For Figure 1(a), the Internal Density, Edge Inside, Average Degree, FOMD, and TPR for community C_1 are 0.67, 4, 2, 0.25, and 0, respectively.

(B) *Functions considering the external connections only*: The following functions consider the edges outside the community ω:

- **Expansion:** $f(\omega) = |E_\omega^{out}|/|\omega|$ computes for each node the number of attached edges pointing outside ω. [110].
- **Cut Ratio (or Ratio Cut):** $f(\omega) = \frac{|E_\omega^{out}|}{|\omega|(N-|\omega|)}$, similar to internal density, computes the fraction of the edges out of all possible edges that are leaving outside ω [132, 133].

For Figure 1(a), the Expansion and Cut Ratio for community C_1 are given as 0.5 and 0.07, respectively.

(C) *Functions considering internal and external connections*: The following functions consider the edges that are within as well as outside the community ω:

- **Conductance**: $f(\omega) = \frac{|E_\omega^{out}|}{2|E_\omega^{in}|+|E_\omega^{out}|}$ for a community measures the ratio of the total number of edges of ω pointing to the other external communities and the total number of edges connected with ω [120].

- **Normalized cut**: $f(\omega) = \frac{|E_\omega^{out}|}{2|E_\omega^{in}|+|E_\omega^{out}|} + \frac{|E_\omega^{out}|}{2(m-|E_\omega^{in}|)+|E_\omega^{out}|}$ normalizes the cut score [59, 120].
- **Maximum-ODF (Out Degree Fraction)**: $f(\omega) = max_{u\in\omega} \frac{|(u,v)\in E: v\notin\omega|}{d(u)}$ is the maximum of the fraction of edges connected with a node in ω pointing outside the community ω [41].
- **Average-ODF**: $f(\omega) = \frac{1}{|\omega|}\sum_{u\in\omega} \frac{|(u,v)\in E: v\notin\omega|}{d(u)}$ is the average fraction of edges connected with nodes in ω that point outside the community ω [41].
- **Flake-ODF**: $f(\omega) = \frac{|u: u\in\omega, |(u,v)\in E: v\in\omega| < d(u)/2|}{|\omega|}$ is an interesting measure. It computes the fraction of nodes in ω, whose number of edges completely internal to the community is less than the number of edges pointing to external communities [41].

For Figure 1(a), the Conductance, Normalized cut, Maximum-ODF, Average-ODF, and Flake-ODF for community C_1 are $0.2, 0.29, 0.5, 0.125$, and 0, respectively.

(D) *Functions considering the model of a network*:

- **Modularity**: For a given community structure, it captures the following intuition: out of the total number of edges in the graph, how many of them are completely internal to a community with respect to the original graph, and how would this number differ if we consider a null graph, that is, a random graph with an identical degree distribution as the original graph? One can choose the null graph in an arbitrary manner from a wide range of possibilities. Usually, the null graph considered has an equivalent degree distribution as of the given graph. For an unweighted and undirected network, modularity is defined as

$$Q_{ud} = \sum_{\omega\in\Omega}\left[\frac{|E_\omega^{in}|}{m} - \left(\frac{|E_\omega^{in} + E_\omega^{out}|}{2m}\right)^2\right]. \tag{1}$$

An alternative way of defining the modularity of a graph is as follows [96]:

$$Q_{ud} = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{d(i)d(j)}{2m}\right]\delta_{\omega_i,\omega_j}, \tag{2}$$

where $\delta_{\omega_i,\omega_j}$ is the Kronecker delta function, which returns 1 if $\omega_i = \omega_j$, and 0 otherwise. The value of modularity lies between -1 and 1. A higher value of modularity indicates a strong community structure. A large number of algorithms have been developed with the objective of optimizing modularity value [13, 22, 32, 36, 38, 117, 129].

The Modularity for community C_1 in Figure 1(a) is 0.75.
[138] further defined four goodness metrics for a community that capture the network structure:

- **Separability** considers the idea that a good community should be highly distinct from the remaining portion of the network [43, 120]. Thus, following this principle, a good community ω should have more edges completely internal to the community and fewer edges external to the community. Separability computes the ratio of the total number of internal edges and the total number of external edges of the community and is given by $f(\omega) = |E_\omega^{in}|/|E_\omega^{out}|$.
- **Density** is similar to separability but considers only the internal degree. The intuition is that for a community to be considered good, it should be well connected [43]. It computes out of the total number of possible edges what fraction of edges exist completely internal to the community ω and is given by $f(\omega) = 2|E_\omega^{in}|/(|\omega|(|\omega|-1))$.
- **Cohesiveness** takes into consideration the inherent topological structure of the community. Ideally, a community having an even and uniform internal connection is considered

to be good; that is, a good community should not be easily split into its corresponding sub-communities. This notion is captured by the conductance, given by $f(\omega) = min_{\omega' \subset \omega} \Phi(\omega')$, where $\Phi(\omega')$ indicates conductance of ω' measured with respect to the subgraph induced by the vertices in ω. A group should possess high cohesiveness in order to qualify as a "good community"; that is, the number of internal connections should be large as this implies that one would need to delete a large number of edges before the community dissolves into disconnected components [75].

- **Clustering Coefficient** considers the intuition that pairs of nodes with common neighbors have a higher possibility to be connected with each other [131].

For Figure 1(a), the Separability and Density of community C_1 are given by 2 and 0.67, respectively. The Cohesiveness and the Clustering Coefficient of community C_1 are given by 0.5 and 0, respectively.

In another paper, [75] used two other topological metrics to measure the quality of a community:

- **Volume**: It is measured by the sum of degrees of all the vertices in ω, that is, $\sum_{u \in \omega} d(u)$.
- **Edges Cut**: $|E_{\omega}^{out}|$ is based along the lines of cohesiveness. It is based on the intuition that a good community should have fewer external connections. It measures the number of edges needed to be removed for disconnecting the community ω from the remaining graph.

For Figure 1(a), the Volume and Edges Cut for community C_1 are 10 and 2, respectively.

2.1.2 Popular and Complex Metrics Exploring Network Properties in Depth. So far we have discussed the quality metrics based on simple graph features. In the following, we focus on some of the popular metrics proposed in the literature. Modularity is the most widely used metric to detect the strength of the communities. Introduced in the seminal paper [97], the principal idea behind modularity is that the number of intercommunity edges for a given graph must be greater than the number of edges for a null graph. The definition of modularity suggested by [97] in Equation (1) was applicable only to *unweighted* and *undirected* graphs. Several modifications and extensions to modularity were suggested in the existing literature of community detection. The modifications cater to the specific tasks and the type of graphs one may intend to analyze.

Modularity for weighted graphs: [93] proposed a simple extension of the existing definition of modularity for weighted graphs. One can view a weighted graph as a multigraph with multiple edges between a pair of nodes. After mapping the weighted graph to a multigraph, it can be easily shown that Equation (2) is a generalized formula for modularity. For a weighted network, instead of A_{ij}, we use W_{ij}, representing the weight of the link connecting nodes i and j. The degree d(i) is replaced with the strength s(i) of node i, which is the sum of the degrees of the adjacent nodes. In order to normalize the equation, the number of edges $m = |E|$ in Equation (2) is substituted by the sum W of the weights of all edges. Thus, in the modified equation, instead of the expected degree, we compare the actual weight W_{ij} with the product $\frac{s(i)s(j)}{2W}$. This product essentially computes the expected weight of the edge (i, j) in the null graph of modularity. The equation of modularity for a weighted network can be written as

$$Q_w = \frac{1}{2|W|} \sum_{ij} \left[W_{ij} - \frac{s(i)s(j)}{2|W|} \right] \delta_{\omega_i, \omega_j}. \tag{3}$$

Modularity for directed graphs: [4] proposed extensions to modularity for directed graphs. For a directed edge, the in-degree and the out-degree of the end vertices determine the probability of the orientation in one of the two possible directions. For the directed graphs, the formulation of the modularity metric can be defined as follows:

$$Q_d = \frac{1}{m}\sum_{ij}\left[A_{ij} - \frac{d(i)^{out}d(j)^{in}}{m}\right]\delta_{\omega_i,\omega_j}, \tag{4}$$

where $d(i)^{in}$ and $d(i)^{out}$ are the in-degree and out-degree of node i, respectively. For a graph whose edges are directed and weighted, Equations (3) and (4) can be aggregated as follows, which is the most general form of modularity:

$$Q_{gen} = \frac{1}{|W|}\sum_{ij}\left[W_{ij} - \frac{s_i^{out}s_j^{in}}{|W|}\right]\delta_{\omega_i,\omega_j}. \tag{5}$$

Similarity-based modularity: [40] proposed similarity-based modularity, which is robust to the communities in a graph having dense interconnections. Instead of considering intercommunity and intracommunity edges as the criteria of partitioning, as considered in the traditional definition of modularity, they proposed a more general concept, *similarity* $S(i,j)$, to measure the graph partition quality. The notion of similarity between two nodes i and j is given by the number of shared neighbors normalized by the number of neighbors of each vertex as shown here:

$$S(i,j) = \frac{|\Gamma_i \cap \Gamma_j|}{\sqrt{|\Gamma_i||\Gamma_j|}}. \tag{6}$$

For a partition Ω, the similarity-based modularity (Q_s) is given as

$$Q_s = \sum_{\omega\in\Omega}\left[\frac{\sum_i\sum_j S(i,j)\delta(i,\omega)\delta(j,\omega)}{\sum_i\sum_j S(i,j)} - \left(\frac{\sum_i\sum_j S(i,j)\delta(i,\omega)}{\sum_i\sum_j S(i,j)}\right)^2\right], \tag{7}$$

where

$$\delta(i,\omega) = \begin{cases} 1, & \text{if } i \in \omega \\ 0, & \text{otherwise.} \end{cases}$$

Motif modularity: All the metrics mentioned previously consider nodes and edges as the smallest units in a network. However, [5] argued that "network motif" should be considered as the smallest nontrivial structural unit of a network, based on which they redefined modularity. The proposed definition of motif modularity captures the following intuition: a good community will "contain" a larger number of motifs than a null model that is generated using a similar degree of distribution but in a random fashion. They listed a triangular modularity for a partition Ω, which considers triangular motifs and is given by

$$Q_{mm}^{\triangle}(\omega) = \left[\frac{\sum_{ijk} A_{ij}(\omega)A_{jk}(\omega)A_{ki}(\omega)}{\sum_{ijk} A_{ij}A_{jk}A_{ki}} - \frac{\sum_{ijk} n_{ij}(\omega)n_{jk}(\omega)n_{ki}(\omega)}{\sum_{ijk} n_{ij}n_{jk}n_{ki}}\right], \tag{8}$$

where $A_{ij}(\omega) = A_{ij}\delta_{(\omega_i,\omega_j)}$, $n_{ij} = d(i)d(j)$, $n_{ij}(\omega) = n_{ij}\delta_{\omega_i,\omega_j}$, and $\delta_{\omega_i,\omega_j} = 1$ if both vertices i and j belong to the same community, and 0 otherwise.

Max-Min modularity: [27] argued that users should have the ability to customize the definition of modularity, which is solely dependent on the network structure. They presented a new measure, called max-min modularity, which considers both the structural properties of the network, Q_{Max} (same as modularity), and user-defined constraints in finding communities, Q_{Min}. The second part is controlled by the users depending on their needs. The Max-Min modularity is given by

$$\begin{aligned} Q_{Max\_Min} &= Q_{Max} - Q_{Min} \\ &= \sum_{ij}\left[\frac{1}{2m}\left(A_{ij} - \frac{d(i)d(j)}{2m}\right) - \frac{1}{2m'}\left(A'_{ij} - \frac{d(i)'d(j)'}{2m'}\right)\right]\delta_{\omega_i,\omega_j}. \end{aligned} \tag{9}$$

The second part tries to minimize the modularity of the complement of the graph G, denoted by G', constructed by taking into consideration the user-defined criteria U. The higher the value of Q_{Max_Min}, the better the community division over Q_{gen}.

Influence-based modularity: [48] proposed influence-based modularity, where they claimed that "a community is composed of individuals who have a greater capacity to influence others within their community than outsiders." Contrary to the existing modularity, where only one hop neighbor of a node is considered, they considered n-hop neighbors, where n is a user-defined parameter. The influence is measured using an *influence matrix* P, which captures the number of *n-hop*-length paths between a pair of nodes i and j in the graph. Consider $\bar{P}$ as the expected capacity to influence. Influence-based modularity is then given by

$$Q_i = \sum_{ij}[P_{ij} - \bar{P_{ij}}]\delta_{\omega_i,\omega_j}. \tag{10}$$

Diffusion-based modularity: [66] argued that the direction of an edge in the network is often ignored in the definition of modularity. Although past works [4, 98] tried to capture the direction of edges, they do not share a common definition of the community structure in directed networks. [66] proposed a modified definition of modularity, which, according to them, is the first generalized modularity metric. They considered diffusion in directed graphs, after taking inspiration from Google's PageRank algorithm. Unlike PageRank, which considers the importance of nodes, they proposed *LinkRank*, which indicates the importance of links. For an edge connecting nodes i and j, the LinkRank can be computed as the probability that a random walker is walking from node i to node j in the stationary state. By using LinkRank, the modified definition of modularity can be written as

$$Q_{lr} = \sum_{ij}[L_{ij} - E(L_{ij})]\delta_{\omega_i,\omega_j}, \tag{11}$$

where $E(L_{ij})$ is the expected value of LinkRank with respect to the null model L_{ij}.

Dist-modularity: [79] claimed that the previous null models used in the original modularity are not good representations of real-world networks and thus the result is less accurate in original modularity. A common feature of many real-world networks is similarity attraction; that is, nodes similar to each other have a higher chance of getting connected. They extended the existing definition of modularity and proposed Dist-modularity by introducing a new null model where the expected number of edges is given by

$$P_{ij}^{Dist} = \frac{\tilde{P_{ij}} + \tilde{P_{ji}}}{2}, \tag{12}$$

where,

$$\tilde{P_{ij}} = \frac{d(i)d(j)e^{-(s_{ij}/\sigma)^2}}{\sum_{v\in V} d(v)e^{-(s_{iv}/\sigma)^2}}. \tag{13}$$

Here s_{ij} denotes the similarity distance between v_i and v_j—the smaller the s_{ij}, the more similar are the two nodes i and j. σ is a parameter of the quality metric that controls how fast or slow the function $e^{-(s_{ij}/\sigma)^2}$ decreases. Using the new null model, dist-modularity is given by

$$Q_{Dist} = \frac{1}{2m}\sum_{ij}\left[A_{ij} - P_{ij}^{Dist}\right]\delta_{\omega_i,\omega_j}. \tag{14}$$

Limits of modularity: Given a network, the knowledge of the size of the communities is not known a priori. Modularity and its variants discussed so far are not robust in nature and would fail to capture communities of all types in a network. [44] discussed the limits of modularity. They wrote, "modularity optimization may fail to identify modules smaller than a scale which depends on the total size of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined." The problem is known as the *resolution limit* of modularity. [52] studied Modularity at a deeper level and pointed out the degeneracy problem. Modularity may not have a global minima. An exponential number of many community structures may be possible to discover from a network whose topological structures are highly different from each other but exhibit modularity values very close to the optimum. This is often known as the *degeneracy of solution*. It tends to occur for directed, weighted, bipartite, and even multiscale generalizations of modularity. However, several modifications of modularity have been proposed in the past to address the limits of modularity. In the following section, we briefly discuss the modifications proposed that overcome the limits of the original definition.

Metrics proposed to overcome the limitations of modularity: Modifications of modularity's null model were introduced by [82] and [90]. [82] identified the sensitivity of modularity toward larger communities. The value of modularity tends to decrease with the increase in the large number of small communities since the actual number of edges might decrease compared to the expected number of edges in the graph. It might lead to a bias toward detecting large communities. In order to address this, they proposed a modified form of modularity by improving the predicted fraction of edges within communities. The method is to create a group of random networks with identical degree distribution as the original network and with the constraint that multiple edges and self-connections are forbidden. They first calculated the probability of an edge between any pair of nodes in an ensemble network. The predicted fraction of edges within each community in a random network is given by the sum of these probabilities over each pair of nodes within the community. This sum (denoted by f_ω) replaces the second element of modularity Q in Equation (1). Thus, **modified modularity** is given by

$$Q_{rnl} = \sum_{\omega\in\Omega}\left[\frac{|E_\omega^{in}|}{m} - f_\omega\right]. \tag{15}$$

[90] addressed the limits of modularity by proposing a local definition of modularity. Here, they computed the expected number of edges in a local manner instead of computing it globally. For computing the expected number of edges, they considered only a part of the subgraph induced by the community and the neighboring communities. Their intuition behind considering this local subgraph to compute the expected number of edges is as follows: a randomized model would assume that the probability of a vertex to connect to any other vertex is equal. However, in reality, a community is associated with neighboring communities only. This gives a better approximation of the real-world random network. On a directed graph, the **localized modularity** LQ is given by

$$LQ = \sum_{\omega\in\Omega}\left[\frac{|E_\omega^{in}|}{L_{\omega n}} - \frac{|E_\omega^{in}||E_\omega^{out}|}{L_{\omega n}^2}\right], \tag{16}$$

where $L_{\omega n}$ is the total number of edges in the subgraph comprising community ω and its neighbor communities. It is important to note that unlike modularity, the localized modularity does not have any upper bound.

In order to address the resolution limit, [112] suggested tuning the parameters contributing to the null model. They scaled the topology by a factor r by adding self-loops of the same magnitude r to the vertices.

[77] proposed another metric using average modularity degree to evaluate the quality of communities, called the **modularity density** or **D-value** [26, 50, 51, 89, 103, 142]. For a partition Ω, the modularity density is represented by

$$D = \sum_{\omega \in \Omega} d(G_\omega), \tag{17}$$

where $d(G_\omega)$ indicates the average modularity degree of subgraph $G_\omega = (V_\omega, E_\omega)$ given by

$$d(G_\omega) = d_{in}(G_\omega) - d_{out}(G_\omega), \tag{18}$$

where $d_{in}(G_\omega)$ and $d_{out}(G_\omega)$ are the average inner and outer degrees of subgraph G_ω, respectively.

[6] viewed the resolution limit of modularity as a feature of the quality function and proposed a modification to the modularity function that allows viewing the network at multiple resolutions. The modularity of the network at scale r is given by

$$Q_r = \sum_{\omega \in \Omega} \left[\frac{|E_\omega^{in}| + |V_\omega| r}{|E| + |V| r} - \left(\frac{|E_\omega^{in} + E_\omega^{out}| + |V_\omega| r}{2|E| + |V| r} \right)^2 \right]. \tag{19}$$

The resolution limit of the method is expected to be solved after the introduction of a modified version that allows one to tune the resolution parameters. However, [70] showed that multiresolution modularity suffers from two opposite coexisting problems: if we use a low-resolution value, the measure would try to merge small subgraphs and thus allow communities larger than a specific size only; when the resolution value is high, the measure would try to split large communities, leading to a large number of small communities. Thus, the resolution limit of modularity still remains largely unresolved.

[143] argued that simply counting the internal edges of a community is not enough to capture the topology of a network. They proposed a new parameter, called minimal diameter D_ω, which is defined as the average minimal path for all pairs of vertices in a given module. The new definition of modularity metric is then given as

$$Q_d = \sum_{\omega \in \Omega} \left[\frac{|E_\omega^{in}|}{D_\omega} - \left(\frac{|E_\omega^{in} + E_\omega^{out}|}{2|E|} \right)^2 \times \frac{1}{\tilde{D_\omega}} \right], \tag{20}$$

where D_ω is the observed diameter for the community ω and can be computed easily. $\tilde{D_\omega}$ is the expected diameter for a randomized graph and is computed by the methods discussed by [47].

[123] addressed the resolution limit of modularity by proposing the quality metric **modularity intensity**. Modularity intensity is proposed along the lines of **cohesiveness** and thereby captures the community evolutionary process. The goal is to give higher scores to those communities that are hard to split into two or more subcommunities.

[30] addressed the multiresolution problem of modularity. They first proposed modularity with **split penalty** given by

$$Q_s = Q_{ud} - SP. \tag{21}$$

The split penalty addresses the problem of favoring small communities by measuring the negative effect of edges that join the nodes of other communities and is given by

$$SP = \sum_{\omega_i \in \Omega} \left[\sum_{\substack{\omega_j \in \Omega \\ \omega_j \neq \omega_i}} \frac{|E_{\omega_i,\omega_j}|}{2m} \right], \tag{22}$$

where $|E_{\omega_i,\omega_j}|$ is the number of edges connecting community ω_i and community ω_j for an unweighted network. For a weighted network, the number of edges is replaced by the sum of weights of edges. We can use Equation (1) and Equation (22) in Equation (21) to obtain the modified modularity with split penalty. Using split penalty alone may lead to large communities, which may not be desirable. In order to address this, they introduced **community density**. For undirected networks, the proposed metric **modularity density** is given by

$$Q_{uds} = \sum_{\omega_i \in \Omega} \left[\frac{|E^{in}_{\omega_i}|}{m} d_{\omega_i} - \left(\frac{|E^{in}_{\omega_i} + E^{out}_{\omega_i}|}{2m} d_{\omega_i} \right)^2 - \sum_{\substack{\omega_j \in \Omega \\ \omega_j \neq \omega_i}} \frac{|E_{\omega_i,\omega_j}|}{2m} d_{\omega_i,\omega_j} \right], \tag{23}$$

where d_{ω_i} is the edge density of ω_i and d_{ω_i,ω_j} is the pair-wise density between ω_i and ω_j.

[141] proposed a new community metric given as follows:

$$\Phi(\Omega) = \Phi_1(\Omega) - \Phi_2(\Omega), \tag{24}$$

where

$$\Phi_1(\Omega) = \sum_{\omega \in \Omega} \frac{|E^{in}_{\omega} + E^{out}_{\omega}|}{N_\omega}, \Phi_2(\Omega) = \sum_{\omega_i \in \Omega} \sum i \neq j \frac{|E_{\omega_i,\omega_j}|}{N_{\omega_i,\omega_j}}. \tag{25}$$

The function $\Phi_1(\Omega)$ defines the sum of the average degrees in each subnetwork and $\Phi_2(\Omega)$ defines the sum of the average number of connections between subnetworks. It is easy to see that for community identification, our goal is to both maximize $\Phi_1(\Omega)$ and minimize $\Phi_2(\Omega)$.

[88] addressed the resolution limit of modularity by proposing a new quality metric called **Z-Modularity**. For a division Ω, they proposed a quantity to compute how much the partition Ω is statistically rare by considering the internal edges of the communities. They considered the following generative mechanism to create edges over V:

"First N edges are placed over V at random with the same distribution of vertex degree. Then, the probability that the edge is placed within communities is given by

$$p = \sum_{\omega \in \Omega} \left(\frac{D_\omega}{2m} \right)^2, \tag{26}$$

where D_ω is the sum of degrees of all the nodes in ω. Note that this edge generation process is the same as the null-model used in the definition of modularity, with the exception of the sample size. Unlike the null-model, the sample size N is not necessarily equal to the number of edges m. Let X be a random variable denoting the number of edges generated by the process within communities. Then, X follows the binomial distribution $B(N, p)$. By the central limit theorem, when the sample size N is sufficiently large, the distribution of X/N can be approximated by the normal distribution $N(p, p(1-p)/N)$."

Considering this edge generation process, they quantified the statistical rarity of partition Ω using the Z-score:

$$Z(\Omega) = \frac{\sum_{\omega \in \Omega} \frac{|E_\omega^{in}|}{m} - \sum_{\omega \in \Omega} \left(\frac{|D_\omega|}{2m}\right)^2}{\sqrt{\sum_{\omega \in \Omega} \left(\frac{|D_\omega|}{2m}\right)^2 \left(1 - \sum_{\omega \in \Omega} \left(\frac{|D_\omega|}{2m}\right)^2\right)^2}}. \tag{27}$$

Recently, [135] proposed multiresolution frameworks by using a generalized "self-loop rescaling strategy" to address the resolution limit of modularity.

Adaptive scale modularity: [127] proposed six properties (see Section 1 of SI Text for details) as axioms for community-based quality functions viz. "scale invariance," "permutation invariance," "monotonicity," "richness," "locality," and "continuity." They analyzed these six properties and found that modularity follows three properties: scale invariance, permutation invariance, and continuity. In order to adapt modularity to obey the remaining three properties, they proposed adaptive scale modularity. Instead of normalizing the edge-weights with the sum of the edge-weights, they proposed to normalize it using a constant M. However, if one only keeps the factor M, the fixed scale modularity does not obey monotonicity. The parameter $\gamma \geq 2$ ensures that the adaptive scale modularity obeys all six axioms. The proposed metric is give by

$$Q_{asm} = \sum_{\omega \in \Omega} \left[\frac{|E_\omega^{in}|}{M + \gamma(|E_\omega^{in} + E_\omega^{out}|)} - \left(\frac{|E_\omega^{in} + E_\omega^{out}|}{M + \gamma(|E_\omega^{in} + E_\omega^{out}|)} \right)^2 \right]. \tag{28}$$

Until now, we have discussed modularity and various adaptations of modularity to incorporate different properties of the network and the community. The metrics discussed so far follow the underlying principle of modularity—for a good community, the number of edges internal to a community should be higher than the same in a random graph that follows the same degree distribution as the original graph. We now pay attention to other quality metrics, each of which consider different aspects of a network.

Community Score: [104] proposed a simple but effective goodness function that intends to maximize the in-degree of the vertices inside a community explicitly and minimize the out-degree of these nodes implicitly. Consider a community ω where μ_i denotes the fraction of edges attached with node i connecting other nodes in ω, that is, the internal degree of node i. Next, they considered the "power mean" of ω of order r, denoted as $M(\omega)$, which is given by

$$M(\omega) = \frac{\sum_{i \in \omega} (\mu_i)^r}{N_\omega}. \tag{29}$$

Considering $|E_\omega| = |E_\omega^{in}| + |E_\omega^{out}|$, the score of a community ω is given by $score(\omega) = M(\omega)|E_\omega|$. The community score of the graph corresponding to a partition Ω is given by

$$CS = \sum_{\omega \in \Omega} score(\omega). \tag{30}$$

SPart: [31] proposed a new fitness function, SPart. It is a node-based metric that considers the contribution of each node and its neighbors to compute the quality of the partition. For a node v belonging to a community ω, it takes the internal $d(v)^{in}$ and external $d(v)^{out}$ degree into consideration. Thus, the strength of a node is computed using the following measure:

$$SNode(v) = \frac{d(v)^{in} - d(v)^{out}}{|\omega|}. \tag{31}$$

The higher the value of this metric for a node, the more important the node is for the community ω. The fitness or goodness of a community ω is evaluated by considering the strength $SNode(v)$ of each node $v \in \omega$ (first-level nodes) along with the strengths of all nodes $w \in \omega$ (second level nodes) that are linked with v. Thus, the following measure computes the goodness of a community:

$$SComm(\omega) = \sum_{v \in \omega} \left[SNode(v) + \frac{1}{2} \sum_{\substack{w \in \omega \\ A_{vw} = 1}} SNode(w) \right]. \tag{32}$$

The overall fitness of a particular partition Ω is given as follows:

$$SPart(\Omega) = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} SComm(\omega_i) \frac{\nu(\omega_i)}{|\omega_i|}, \tag{33}$$

where $\nu(\omega_i)$ is the ratio of the number of edges within the community ω_i to the total number of edges.

Significance: [126] proposed a new line of thought that instead of figuring out whether a "fixed" community containing at least E internal edges exists with a minimum probability, one might want to shift the attention to find out the probability of finding a community with at least E internal edges in a *random graph*. Now, one can reduce the problem of computing the probability of finding a certain community to the goal of finding certain dense subgraphs in a random graph. The problem of finding certain dense subgraphs can be computed as follows: in a random graph $\mathcal{G}$ of size n and density p, the probability that a subgraph S of size n_ω and density q appears is given asymptotically as follows:

$$Pr(S(n_\omega, q) \subseteq \mathcal{G}(n, p)) = e^{\Theta(-\binom{n_\omega}{k} D(q||p))}, \tag{34}$$

where $D(q||p)$ is the Kullback-Leibler divergence. For a community Ω, one can compute the probability for the community to be contained in a random graph as follows:

$$Pr(\Omega) = \prod_{\omega} exp(-\binom{n_\omega}{2} D(p_c||p)). \tag{35}$$

Significance is then given by

$$S(\Omega) = -\log Pr(\Omega) = \sum_{c} \binom{n_\omega}{2} D(p_\omega||p). \tag{36}$$

Permanence: [17, 23, 24] proposed permanence, a vertex-based community quality metric to quantify the probability of a vertex to remain in the community in which it is assigned and the degree by which it is "pulled" by the neighboring communities. The permanence of a vertex v compares the internal neighbors of the vertex, $I(v)$ with the maximum number of neighbors to one of the external communities of v, that is, $E_{max}(v)$. This is normalized by $d(v)$. Permanence also captures how well the vertex is connected within the community via internal clustering coefficient $c_{in}(v)$, defined by the fraction between the actual number of connections among the internal neighbors[2] of v and the total number of possible edges among them. A high internal clustering coefficient indicates the presence of a clique-like structure, which again increases the chance of the node to be a part of a community. If v is a part of a singleton community, $C_{in}(v)$ is treated as 0. Mathematically, the permanence of a vertex v is given by

$$Perm(v) = \left[\frac{I(v)}{E_{max}(v)} \times \frac{1}{d(v)} \right] - [1 - c_{in}(v)]. \tag{37}$$

[2]Internal neighbors of v are all the neighbors that belong to the same community where v belongs.

If $E_{max}(v) = 0$, the first component of Equation (37) is treated as $\frac{I(v)}{D(v)} = 1$. To compute the permanence of the entire network, one needs to first sum the permanence of all the vertices in the network and normalize it with the number of vertices. The permanence of a network provides a quantitative way to estimate to what extent the vertices in the given partition are bounded to their respective communities. The value of permanence ranges from −1 to +1. Permanence has been proved to ameliorate the limitations of modularity [19, 24, 76].

Surprise: Another community quality metric, Surprise, proposed by [3], follows the idea of modularity by comparing a network parameter with a null model. In the proposed metric, for a given community of the network, one needs to compare the distribution of the nodes and edges in communities with respect to a null model. The idea is that in a null model, edges between nodes would emerge randomly as opposed to a network with a good community. The difference in the observed and expected distribution of nodes and edges is computed using KL-divergence. To this end, it uses the following cumulative hypergeometric distribution:

$$S = \sum_{j=p}^{Min(M,|E|)} \frac{\binom{M}{j}\binom{F-M}{|E|-j}}{\binom{F}{|E|}}, \tag{38}$$

where E and F are the observed and the maximum possible edges in a network, respectively; M is the maximum possible number of intracommunity edges for a given community; and p is the total number of intracommunity links observed in that community. Thus, surprise measures how unlikely ("surprising") that distribution is. Unlike modularity, surprise also considers the role of nodes within each community along with the role of edges. Qualitatively, surprise performs better than modularity. This claim was later confirmed by [2, 42]. In practice, it is not straightforward to work with, nor is it simple to implement in an optimization procedure, mainly due to numerical computational problems.

[125] proposed an asymptotic approximation of surprise by allowing the links to be withdrawn with replacement. Thus, **asymptotical surprise** is given by

$$S = \sum_{j=p}^{Min(M,|E|)} \binom{M}{j} \langle q \rangle^j (1 - \langle q \rangle)^{M-j}, \tag{39}$$

where $q = \frac{\sum_{\omega \in \Omega} |E_\omega^{in}|}{|E|}$ and $\langle q \rangle$ indicates the expected value of q.

Communitude: [87] quantified the community degree of ω in terms of the fraction of the internal edges of the subgraph induced by ω. Inspired by the Z-score (presented in Equation (27)), they further estimated a similar probability distribution of the fraction of the internal edges of the subgraph generated using the nodes in ω. Next, they quantified the community degree of ω in terms of the fraction of the internal edges of the subgraph using the Z-score as follows:

$$com(\omega) = \frac{\frac{|E_\omega^{in}|}{m} - \left(\frac{2|E_\omega^{in} + E_\omega^{out}|}{2m}\right)^2}{\sqrt{\left(\frac{2|E_\omega^{in} + E_\omega^{out}|}{2m}\right)^2 \left(1 - \left(\frac{2|E_\omega^{in} + E_\omega^{out}|}{2m}\right)^2\right)}}. \tag{40}$$

If $\omega = \phi$, $com(\omega) = 0$. The upper bound for the metric is 1. Note, this quality function can be viewed as a modified version of the function given by the Z-score, which is normalized by the standard deviation of the fraction of the number of internal edges of the subgraph.

[34] used **compactness,** which calculates the speed with which information diffuses in a community. In order to compute this metric, one begins with the node that is at the boundary and moves to compute the number of edges reached per time step assuming the information is transmitted without any loss. The intuition is similar to defining a community along the following

Table 1. Notations Used in This Survey

Graph specific	
$G(V, E)$	A network with sets of vertices V and edges E
A	Adjacency matrix of network G
N	$N = \|V\|$, number of vertices in G
m	$m = \|E\|$, number of edges in G
Γ_u	Neighbors of node u
$d(u)$	Degree of vertex u
$d(u)^{in}$	In-degree of vertex u
$d(u)^{out}$	Out-degree of vertex u
$C_{in}(u)$	Internal clustering coefficient of vertex u (i.e., fraction of actual number of edges and possible number of edges among the internal neighbors of u, where internal neighbors of u are those that are part of the same community u belongs to)
Community specific	
Ω	Detected nonoverlapping community structure, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$
Ψ	Detected overlapping community structure, $\Psi = \{\psi_1, \psi_2, \ldots, \psi_K\}$
C	Ground-truth community structure $C = \{c_1, c_2, \ldots, c_J\}$
N_{ω_j}	Number of vertices in community ω_j
$N_{\omega_i c_j}$	$\|\omega_i \cap c_j\|$, number of vertices present in both ω_i and c_j communities
E_{ω}^{in}	Set of edges completely internal to community ω
E_{ω}^{out}	Set of edges from community ω and to other outside communities

principle: a good community has a group of nodes within which information transmits quickly and reaches everyone. They recently used it as a community-level quality function to measure the goodness of the detected community [35].

The **local internal clustering coefficient** [131] is also used as a quality metric in [35]. The idea behind this metric is that neighbors of a vertex belonging to the same community should also be connected to each other.

A summary of the metrics discussed in this subsection can be found in the SI text (Table 1).

2.2 Metrics for Overlapping and Fuzzy Community Detection

In the previous section, we discussed the community structure where a node can belong to a single community only. However, in real-world datasets, the community boundaries might not be so crisp since a node can have membership in multiple communities at the same time. For instance, a user in a social network can participate in several groups such as a friendship circle, family circle, and so forth; a researcher may be active in several areas at the same time; a person can participate in discussion in several blogs or forums. Moreover, in social networks, a vertex can be a part of unlimited groups because there is no restriction on the number of groups a person can belong to. This phenomenon also happens in other networks, such as citation networks [18] and biological networks [124], where a node might have multiple functions. Therefore, overlapping community structure is a natural phenomenon in real networks, and thus there has been an increasing interest in discovering communities that are not necessarily disjoint.

More formally, given a graph $G(V, E)$, the task of overlapping community detection is to group the vertices V into nonempty sets $\Psi = \{\psi_1, \psi_2, \ldots, \psi_{|\Psi|}\}$, where a vertex can belong to more than one set, that is, $|\psi_1| + |\psi_2| + \cdots + |\psi_{|\Psi|}| \geq |V|$.

Note that in this case, a node can either be a part of a community *completely* or not (binary decision), resulting a *crisp overlapping community*. On the other hand, several attempts have been conducted with the idea that a node can belong to a community with a certain probability. This is called *fuzzy overlapping community* structure. In this section, we discuss the metrics associated with both of these notions of overlapping community structure detection.

Although an initial attempt for overlapping community detection was done by [101], [140] introduced the original notion of overlapping community structure. Consider a set of overlapping communities $\Psi = \{\psi_1, \psi_2, \ldots, \psi_{|\Psi|}\}$, where a node i can be a part of multiple communities given by a vector of belonging coefficients $(\alpha_{i\psi_1}, \alpha_{i\psi_2}, \ldots, \alpha_{i\psi_{|\Psi|}})$. A belonging coefficient $\alpha_{i\psi}$ indicates how strongly node i belongs to community ψ. Therefore, the following constraints are assumed to hold:

$$0 \leq \alpha_{i\psi} \leq 1 \quad \forall i \in V, \forall \psi \in \Psi \qquad and \qquad \sum_{\psi \in \Psi} \alpha_{i\psi} = 1. \tag{41}$$

Modularity: Using the notion of fuzzy overlap, [140] defined the membership of each community as $\bar{V_\psi} = \{i | \alpha_{i\psi} > \lambda, i \in V\}$, where λ is a threshold based on which a soft assignment is converted into a final community structure. They proposed a generalized notion of modularity as follows:

$$Q_{ov}^{z} = \sum_{\psi \in \Psi} \left[\frac{|E_\psi^{in}|}{m} - \left(\frac{2|E_\psi^{in}| + |E_\psi^{out}|}{2m} \right)^2 \right], \tag{42}$$

where $|E_\psi^{in}| = \sum_{i,j \in \psi} ((\alpha_{i\psi} + \alpha_{j\psi})/2) W_{ij}$ and $|E_\psi^{out}| = \sum_{i \in \psi, j \in N - N_\psi} (\alpha_{i\psi} + (1 - \alpha_{j\psi})/2) W_{ij}$. W_{ij} indicates the weight on the edge between nodes i and j. This modified modularity was used to give a relative membership of the nodes.

[92] extended modularity for computing the goodness of an overlapping partition Ψ by replacing the Kronecker delta function in Equation (2) and proposed a fuzzy variant of modularity as follows:

$$Q_{ov}^{F} = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d(i)d(j)}{2m} \right] s_{ij}, \tag{43}$$

where

$$s_{ij} = \sum_{\psi \in \Psi} \alpha_{i\psi} \alpha_{j\psi}. \tag{44}$$

The goal is thus to find a fuzzy partition such that the difference between the actual similarity among the vertices given by the adjacency matrix and the fuzzy similarity given by the belongingness vector is minimized.

A similar goodness metric was proposed by [119], where they extended modularity as follows:

$$Q_{ov}^{S} = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d(i)d(j)}{2m} \right] \alpha_{i\psi} \alpha_{j\psi}. \tag{45}$$

[118] proposed to optimize Equation (46), to discover both the hierarchical and overlapping communities together by introducing a belonging coefficient. The belonging coefficient $\alpha_{i\psi}$ of a

node i for a given community is redefined as the number of communities O_i to which it belongs. The extended modularity is given by

$$Q_{ov}^{E} = \frac{1}{2m} \sum_{\psi \in \Psi} \sum_{ij} \left[A_{ij} - \frac{d(i)d(j)}{2m} \right] \frac{1}{O_i O_j}. \tag{46}$$

[99] proposed the following metric in terms of a function F:

$$Q_{ov}^{N} = \frac{1}{2m} \sum_{\psi \in \Psi} \sum_{ij} \left[A_{ij} F(\alpha_{i\psi}, \alpha_{j\psi}) - \frac{d(i)d(j) \left(\sum_{v \in V} F(\alpha_{v\psi}, \alpha_{j\psi}) \right) \left(\sum_{v \in V} F(\alpha_{i\psi}, \alpha_{v\psi}) \right)}{2mN^2} \right], \tag{47}$$

where $F(\alpha_{i\psi}, \alpha_{j\psi})$ can be defined by any of the following: a product $\alpha_{i\psi}\alpha_{j\psi}$, an average $(\alpha_{i\psi} + \alpha_{j\psi})/2$, a maximum $max(\alpha_{i\psi}, \alpha_{j\psi})$, or any other suitable function.

[71] maximized the following local fitness function in their proposed algorithm LFM to obtain natural communities:

$$f(\psi) = \frac{|E_{in}|^{\psi}}{(|E_{in}|^{\psi} + |E_{out}|^{\psi})^{\alpha}}, \tag{48}$$

where α indicates the resolution parameter that is used to control the community size.

[105] used the definition of community score (Equation (30)) introduced in [104] on the line graph of a given network G.

[74] defined a crisp overlapping goodness measure for a partition Ψ as follows:

$$Q_{ov}^{crisp} = \frac{1}{|\Psi|} \sum_{\psi \in \Psi} Q_{\psi}. \tag{49}$$

The modularity Q_{ψ} for a given community ψ is given by

$$Q_{\psi} = \frac{|E_{\psi}^{in}| + |E_{\psi}^{out}|}{N_{\psi}(N_{\psi} - 1)/2} \frac{1}{N_{\psi}} \sum_{i \in N_{\psi}} \frac{\sum_{j \in N_{\psi}, i \neq j} A_{ij} - \sum_{j \notin N_{\psi}} A_{ij}}{d(i) s_i}. \tag{50}$$

The number of communities in which node i belongs is given by s_i.

[25] proposed using the modified modularity Q_{ov}^{Ψ} for weighted networks defined as

$$Q_{ov}^{\Psi} = \frac{1}{2m} \sum_{\psi \in \Psi} \sum_{ij} \left[A_{ij} - \frac{d(i)d(j)}{2m} \right] \alpha_{i\psi} \alpha_{j\psi}, \tag{51}$$

where $\alpha_{i\psi} = \frac{k_{i\psi}}{\sum_{\psi \in \Psi} k_{i\psi}}$ is the strength of community membership exhibited by node i for community ψ, and $k_{i\psi} = \sum_{j \in \psi} w_{ij}$ is the sum of weights of edges from i into community ψ.

[58] proposed the following modification to the fitness function, given by Equation (48):

$$f(\psi) = \frac{|E_{in}|^{\psi} + 1}{(|E_{in}|^{\psi} + |E_{out}|^{\psi})^{\alpha}}, \tag{52}$$

which allows singleton communities.

[29] extended the existing definition of modularity density introduced in [30]. They proposed the following goodness function:

$$Q_{ov}^{MD} = \sum_{\psi \in \Psi} \left[\frac{|E_{\psi}^{in}|}{m} d_{\psi} - \left(\frac{2|E_{\psi}^{in}| + |E_{\psi}^{out}|}{2m} d_{\psi} \right)^2 - \sum_{\psi' \in \Psi, \psi \neq \psi'} \frac{|E_{\psi,\psi'}|}{2m} d_{\psi,\psi'} \right],$$
$$d_{\psi} = \frac{2|E_{\psi}^{in}|}{\sum_{i,j \in \psi, i \neq j} f(\alpha_{i\psi}, \alpha_{j\psi})}, \tag{53}$$
$$d_{\psi,\psi'} = \frac{|E_{\psi,\psi'}|}{\sum_{i \in \psi, j \in \psi'} f(\alpha_{i\psi}, \alpha_{j\psi})}$$

where $|E_{\psi}^{in}| = \frac{1}{2} \sum_{i,j \in \psi} f(\alpha_{i\psi}, \alpha_{j\psi}) A_{ij}$, $|E_{\psi}^{out}| = \sum_{i \in \psi} \sum_{\substack{\psi' \in \Psi \\ \psi \neq \psi' \\ j \in \psi'}} f(\alpha_{i\psi}, \alpha_{j\psi'}) A_{ij}$, and $|E_{\psi,\psi'}| = \sum_{i \in \psi, j \in \psi'} f(\alpha_{i\psi}, \alpha_{j\psi'}) A_{ij}$.

Flex: [37] introduced a metric called flex that tries to maximize internal edges in a community along with the local clustering coefficient of each community. In order to compute flex for a given partition, one needs to consider the *Local Contribution* of a node i to a community ψ, which is given as follows:

$$LC(i,\psi) = \lambda * \triangle(i,\psi) + (1-\lambda) * N(i,\psi) - \kappa * \wedge(i,\psi), \tag{54}$$

where $\triangle(i,\psi)$ is the ratio between the number of triangles node i is a part of (transitivity of node i) within ψ and the total number of triangles formed by i considering the entire network. The expression $N(i,\psi)$ indicates the fraction of neighbors of node i internal to community ψ. Finally, $\wedge(i,\psi)$ measures the ratio between the number of wedges (open triangle or two-hop path) node i is a part of within the community ψ and the total number of wedges containing node i in the entire network. The parameters λ and κ are weights that are used to tune the importance of each term in the expression.

Once the local contribution of each node with respect to each community is computed, the *Community Contribution* (CC) of a community ψ is then defined as

$$CC(\psi) = \sum_{i \in \psi} LC(i,\psi) - \frac{N_{\psi}{}^{\gamma}}{N}, \tag{55}$$

where γ is the penalization weight used to avoid the trivial solution. The *flex* value of a given community structure Ψ is given by

$$Flex(\Psi) = \frac{1}{N} \sum_{\psi \in \Psi} CC(\psi). \tag{56}$$

Further, a two-step process was proposed by [10] that maximizes the following local density function:

$$f(\psi) = \frac{|W_{in}|^{\psi}}{|W_{in}|^{\psi} + |W_{out}|^{\psi}}. \tag{57}$$

A modified version was introduced by [65] after including the probability of edge e_p, where the parameter λ manages the behavior of the algorithm when the network is sparse:

$$f(\psi) = \frac{|W_{in}|^{\psi}}{|W_{in}|^{\psi} + |W_{out}|^{\psi}} + \lambda e_p. \tag{58}$$

A summary of the metrics discussed in this subsection can be found in the SI text (Table 2).

We list the advantages, disadvantages, and time complexity of the quality metrics discussed in Sections 2.1 and 2.2 in the SI text (Table 5). We further classify the community goodness metrics based on the type of network: directed, undirected, weighted, and so on. A summary of this classification can be found in the SI text (Table 6). We also classify different community detection algorithms based on the metrics they optimize in the SI text (Table 7).

2.3 Other Metrics for Community Detection

The majority of the community quality detection metrics that have been proposed in the literature pertain to overlapping and nonoverlapping communities only. However, attempts have been made to propose community quality metrics depending on the application at hand or the network properties. A detailed explanation of all such metrics can be found in the SI text (Section 2).

3 METRICS FOR COMMUNITY EVALUATION

Once the communities from a network are detected using a community detection algorithm, the next task is to evaluate the detected community structure. The evaluation becomes easier if the actual community structure of the network (often known as the "ground-truth" community structure) is available. In such cases, various metrics can be used to measure the similarity between the discovered community structure and the ground-truth structure. We refer to these metrics as ground-truth-based validation metrics. If the algorithm is able to detect a community structure that has high resemblance with the ground truth, the algorithm is well accepted and can be used further for other networks where the underlying ground-truth communities might not be available.

3.1 Ground-Truth-Based Validation Metrics for Nonoverlapping Community Structure

In this section, we discuss the metrics used to measure the similarity between the detected solution and the ground-truth community structure for the case of nonoverlapping community detection. Note that most of these metrics are borrowed from the literature of "clustering" in data mining.

Let us recall the notations to be used in this section again: given a graph $G(V, E)$, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$ is the set of detected communities, and $C = \{c_1, c_2, \ldots, c_J\}$ is the set of ground-truth communities. $N = |V| = \sum_{k=1}^{K} |\omega_k| = \sum_{j=1}^{J} |c_j|$ is the total number of nodes, $N_{c_j} = |c_j|$, and $N_{\omega_i c_j} = |\omega_i \cap c_j|$.

[78, 81] introduced **Purity**, where each detected community is assigned to the ground-truth label that is most frequent in the community. Formally, it is measured as

$$Purity(\Omega, C) = \frac{1}{N} \sum_{k} \max_{j} N_{\omega_k, c_j}. \tag{59}$$

The maximum value of purity is 1, indicating an exact match between two partitions. The minimum value is 0, indicating that the partitions are completely opposite. Note that purity is an asymmetric metric; that is, $Purity(\Omega, C)$ and $Purity(C, \Omega)$ are not equal. In general, the former measure is used more often and is known as "simply purity," whereas the latter definition is known as "inverse purity" [7]. [36] discussed the bias of simple purity on the number and size of the communities. However, this remark is not validated for inverse purity.

One can get high purity if the number of communities is large; in particular, the purity value goes up to 1 when each node is assigned to its own community. In contrast, the inverse purity tends to favor detecting a small number of large-size communities. In the utmost case, the algorithm assigns all the nodes to a single community. Thus, purity can not be used as a tradeoff between the quality and the number of communities. To solve this problem, an additional constraint was introduced in [95]—when a discovered community is majority in many ground-truth communities, we consider all the nodes involved in the community as misclassified. In such a case, a new measure is adopted, called **F-Measure** [7]. It is calculated as the harmonic mean of simple purity and inverse purity:

$$F - Measure = \frac{2 \cdot Purity(\Omega, C) \cdot Purity(C, \Omega)}{Purity(\Omega, C) + Purity(C, \Omega)}. \tag{60}$$

Another interpretation of community is to consider it as a set of decisions, one for each pair of nodes in the network [61]. Two nodes will be assigned to the same community if and only if they both have the same label in ground truth. A true positive (TP) decision indicates that two nodes that are part of the same ground-truth community are assigned to the same detected community. A true negative (TN) decision indicates that two nodes that do not share any ground-truth community are assigned to different communities in the discovered community structure. Here we can observe two kinds of errors. A false positive (FP) indicates that two nodes belonging to different ground-truth communities are mistakenly assigned to the same detected community. Similarly, a false negative (FN) indicates that two nodes that are part of the same ground-truth community are mistakenly assigned to different communities in the detected community structure. The percentage of decisions that are correct is measured by the **Rand Index** (RI) as follows:

$$RI = \frac{TP + TN}{TP + FP + FN + TN}. \tag{61}$$

The RI gives equal weight to FPs and FNs. If $\beta > 1$, F_β will penalize FNs more than FPs, thus producing more weight to recall (R) as follows:

$$P = \frac{TP}{TP + FP}; \ R = \frac{TP}{TP + FN}; \ F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}. \tag{62}$$

Although the RI is more stringent and reliable, it has a few shortcomings. In brief, the RI can result in having biases, which may mislead the results in certain applications. However, there are other external measures that can produce accurate results, such as the Jaccard coefficient [56], the FowlkesMallows index [46], the Minkowski measure [64], and the τ statistics [63].

In community detection, the **Adjusted Rand Index** (ARI), the chance-corrected version of the RI, is often used [61]. The ARI is less sensitive to the number of communities [128]. The idea of "chance correction" is a general formulation and can be applied for any measure H:

$$H_c = \frac{H - E(H)}{H_{max} - E(H)}, \tag{63}$$

where H_c is the chance-corrected measure, H_{max} is the maximal value of H, and $E(H)$ is the expected value for some null model. [61] mentioned that if the partitions are generated randomly with constraints on the number and size of the communities, the expectation for the number of pairs in a community intersection $\omega_i \cap c_j$ is given by

$$E\binom{N_{\omega_i, c_j}}{2} = \binom{N_{\omega_i}}{2}\binom{N_{c_j}}{2} \Big/ \binom{N}{2}. \tag{64}$$

By replacing in Equation (63) and after some simplifications, we get the final ARI:

$$ARI(\Omega, C) = \frac{\sum_{ij}\binom{N_{\omega_i, c_j}}{2} - \sum_i \binom{N_{\omega_i}}{2} \sum_j \binom{N_{c_j}}{2} \Big/ \binom{N}{2}}{\frac{1}{2}\left(\sum_i \binom{N_{\omega_i}}{2} + \sum_j \binom{N_{c_j}}{2}\right) - \sum_i \binom{N_{\omega_i}}{2} \sum_j \binom{N_{c_j}}{2} \Big/ \binom{N}{2}}. \tag{65}$$

The ARI is symmetric and ranges from -1 (both communities are completely different) to 1 (both communities are exactly similar). The ARI value of (below) 0 indicates that the similarity between Ω and C is equal (less) than the expected value from two random communities.

Normalized Mutual Information (NMI) [45, 81, 122], another alternative information-theoretic metric, is defined as follows:

$$NMI(\Omega, C) = \frac{I(\sigma, C)}{[H(\sigma) + H(C)]/2}, \tag{66}$$

where I is mutual information:

$$I(\Omega, C) = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N} \, log \frac{N|\omega_k \cap c_j|}{|\omega_k||c_j|}. \tag{67}$$

H is entropy as defined as follows:

$$H(\Omega) = -\sum_k \frac{|\omega_k|}{N} \, log \frac{\omega_k}{N}. \tag{68}$$

NMI is always a number between 0 and 1. A major problem of NMI is that it is not a *true metric*; that is, it does not follow triangle inequality (see Section 3.2 of SI text).

In contrast, **Variation of Information** (VI) [67, 85] or shared information distance obeys the triangle inequality. It is defined as

$$\begin{aligned} VI(\Omega, C) &= -\sum_{i,j} r_{ij}[\log(r_{ij}/p_i) + \log(r_{ij}/q_j)] \\ &= H(\Omega) + H(C) - 2I(\Omega, C), \end{aligned} \tag{69}$$

where $p_i = |\omega_i|/N$, $q_j = |c_j|/N$ and $r_{ij} = |\omega_j \cap c_j|/N$.

However, [100] argued (see Section 3.1 of SI text) that the traditional metrics often ignore the network topology. To formulate more realistic metrics, [100] proposed to use both the traditional metrics and network properties. However, they also noticed that this additional information leads to a more complicated process due to the "multiplicity of values." Recently, [68] proposed the modification of the traditional measures with the network information—modified purity, modified ARI, and modified NMI.

To begin with, the purity of *a node* is defined for a community structure Ω with respect to another community structure C:

$$Purity(u, \Omega, C) = \delta(arg \max_k N_{\omega_i, c_k}, j), \tag{70}$$

where $u \in \omega_i$ and $u \in c_j$. δ is the Kronecker delta function; therefore, $\delta(x, y) = 1$ if $x = y$, and 0 otherwise. The value of purity is therefore binary: 1 if the community in C containing u is the majority in that of Ω containing u, and 0 otherwise. Then the purity of a part ω_i relative to a partition C can be measured by averaging the purity of its constituent nodes: $Purity(\omega_i, C) = \frac{1}{|\omega_i|} \sum_{u \in \omega_i} Purity(u, \Omega, C)$. So, for all the communities in Ω with respect to C, we get $Purity(\Omega, C) = \sum_i \sum_{u \in \omega_i} \frac{1}{N} Purity(u, \Omega, C)$. Note that the purity of each node is normalized by N. In order to consider the topological information, [68] proposed to replace this uniform weight by a value w_u to penalize misclassification more strongly related to topologically more important nodes in the network. Therefore, the **Modified Purity** can be defined as follows:

$$Purity_M(\Omega, C) = \sum_i \sum_{u \in \omega_i} \frac{w_u}{w} Purity(u, \Omega, C), \tag{71}$$

where $w = \sum_v w_v$, that is, sum of all the weights of the nodes. This normalization allows keeping the measure between 0 and 1. Similarly, using the modified definition of purity, we can obtain a modified F-measure using Equation (60).

However, since the Rand Index is designed on the basis of pairwise comparisons, it is difficult to remove the individual contributions per node, as we have seen earlier. Therefore, [68] used similarity for pairs of nodes and proposed to multiply the weights corresponding to two nodes: $w_u w_v$. Then, for any subset of nodes S, it can be translated as $W(S) = \sum_{u,v \in S} w_u w_v$. Therefore, from Equation (65), the **Modified ARI** can be obtained as follows:

$$ARI_M(\Omega, C) = \frac{\sum_{ij} W(\omega_i \cap W(c_j) - \sum_j W(\omega_i) W(c_j)/W(V)}{\frac{1}{2}\left(\sum_i W(\omega_i) + \sum_j W(c_j) \right) - \sum_i W(\omega_i) \sum_j W(c_j)/W(s)}. \tag{72}$$

Since the assumption behind NMI is that all nodes have the similar probability of $1/N$ to be selected, [68] replaced it by the node-specific weight w_u. One can then measure a joint probability distribution $p'_{ij} = \sum_{u \in \omega_i \cap c_j} w_u / W$. Therefore, Equations (67) and (68) can be replaced as follows:

$$I(\Omega, C) = \sum_k \sum_j \frac{W(\omega_k \cap c_j)}{W} \, log \frac{W \cdot W(\omega_k \cap c_j)}{W(\omega_k) W(c_j)} \tag{73}$$

$$H(\Omega) = -\sum_k \frac{W(\omega_k)}{W} \, log \frac{W(\omega_k)}{W}. \tag{74}$$

By replacing the previous two equations in Equation (66), one can get the **Modified NMI**.

All the modified metrics discussed previously require the definition of an individual weight w_u for node u. [68] considered three types of weights: (1) degree measure, $w_u = d_u / \underset{v}{max}\,(d_v)$, where d_u is the degree of u; (2) embeddedness measure [72], $w_u = e_u/d_u$, where e_u is the internal degree of u in its own community; and (3) weighted embeddedness measure, $w_u = e_u / \underset{v}{max}\,(d_v)$.

[8] proposed **Edit Distance** between a pair of community structures to measure the similarity. The idea is to compute the minimum number of transformations required to move one partition to another. Although it was proved to be intelligible, it requires a one-to-one matching. If the communities merge or split, one cannot use this metric (see Section 3.3 in SI text).

A summary of the metrics discussed in this subsection can be found in the SI text (Table 3).

3.2 Ground-Truth-Based Validation Measures for Overlapping Community Structure

Let us again recall that for a network $G(V, E)$, $\Psi = \{\psi_1, \psi_2, \ldots, \psi_K\}$ is the set of detected communities, and $C = \{c_1, c_2, \ldots, c_J\}$ is the set of ground-truth communities. $N = |V| = |\cup_{k \in K} \psi_k| = |\cup_{j \in J} c_j|$ is the total number of nodes, $N_{c_j} = |c_j|$, and $N_{\omega_i c_j} = |\omega_i \cap c_j|$.

NMI was further extended for overlapping community structure [84]. For each node i in the detected community structure Ψ, its community membership information can be captured by a binary vector of size $|\Psi|$, where $(x_i)_k$ is 1 if node i belongs to the k^{th} cluster ψ_k, and otherwise 0. The k^{th} entry of this vector can be viewed as a random variable X_k, whose probability distribution is given by $P(X_k = k) = N_k/N$, $P(X_k = 0) = 1 - P(X_k = 1)$, where $N_k = |\psi|$. The same is applicable for the random variable Y_l of l^{th} community in C. The empirical marginal and joint probability distributions, $P(X_k)$ and $P(X_k, Y_l)$, respectively, are used to further define entropy $H(X)$ and $H(X_k, Y_l)$. The conditional entropy of a community X_k, given Y_l, is calculated as $H(X_k|Y_l) = H(X_k, Y_l) - H(Y_l)$. The entropy of X_k with respect to the entire vector Y is based on the best matching between X_k and any component of Y given by

$$H(X_k|Y) = min_{l \in 1,2,\ldots,|C|} H(X_k|Y_l). \tag{75}$$

The normalized conditional entropy of a community X with respect to Y is

$$H(X|Y) = \frac{1}{C} \sum_k \frac{H(X_k|y)}{H(X_k)}. \tag{76}$$

Similarly, we can define $H(Y|X)$. Then the NMI for overlapping community, **Overlapping Normalized Mutual Information** (ONMI), for two community structures Ω and C can be defined as

$ONMI(X|Y) = 1 - [H(X|Y) + H(Y|X)]/2$. ONMI can be easily reduced to NMI when there is no overlap in the network.

The formulation of ARI (discussed earlier) in the overlapping setting is the **Omega index** [33, 91]. It is also based on the agreement between the pairs of nodes in two community structures. Here, we consider a pair of nodes to be in agreement if both of them are assigned into the same number of communities (or possibly none). In other words, it considers the number of pairs of nodes—belonging to no communities, assigned together in a single community, assigned into exactly two communities, and so on. The Omega index is defined in the following way [54, 58]:

$$Omega(\Psi, C) = \frac{Omega_u(\Psi, C) - Omega_e(\Psi, C)}{1 - Omega_e(\Psi, C)}. \tag{77}$$

The unadjusted Omega index $Omega_u$ is defined as

$$Omega_u(\Psi, C) = \frac{1}{M} \sum_{j=1} max(|\Psi|, |C|) |t_j(\psi_i) \cap t_j(c_j)|, \tag{78}$$

where $M = N(N-1)/2$, indicates all possible edges; $t_j(C)$ is the set of pairs that appear exactly j times in a community C. With respect to the null mode, the expected Omega index $Omega_e$ is given by

$$Omega_e(\Psi, C) = \frac{1}{M^2} \sum_{j=1} max(|\Psi|, |C|) |t_j(\psi_i)| \cdot |t_j(c_j)|. \tag{79}$$

A maximum value of the Omega index is 1, indicating exact correspondence between the detected and ground-truth community structures.

[14] extended the concept of RI by making it able to evaluate an overlapping partition of a dataset. Further, [15] proposed another effective solution, named **Generalized External Index** (GEI), to the problem of comparing two overlapping partitions by deriving the concepts of agreements and disagreements for each individual pair of nodes (*i*,*j*). To do so, the following auxiliary definitions are needed:

- $\alpha_\Psi(i, j)$: Number of communities shared by nodes i and j in partition Ψ
- $\alpha_C(i, j)$: Number of communities shared by nodes i and j in partition C
- $\beta_\Psi(i)$: Number of communities in Ψ that node i is a part of, minus 1
- $\beta_C(i)$: Number of communities in C that node i is a part of, minus 1

Based on these definitions, the agreements a_G and disagreements d_G associated to pair (i, j) are defined as

$$a_G(i, j) = min\{\alpha_\Psi(i, j), \alpha_C(i, j)\} + min\{\beta_\Psi(i), \beta_C(i)\} + min\{\beta_\Psi(j), \beta_C(j)\}, \tag{80}$$

$$d_G(i, j) = abs[\alpha_\Psi(i, j) - \alpha_C(i, j)] + abs[\beta_\Psi(i) - \beta_C(i)] + abs[\beta_\Psi(j) - \beta_C(j)]. \tag{81}$$

These measures, in turn, can be used to define the generalized external index for comparing overlapping partitions:

$$GEI(\Psi, C) = \frac{a_G}{a_G + d_G}. \tag{82}$$

[62] proposed another extension of RI, named **Fuzzy Rand Index**. They considered the RI as a distance measure. Given a fuzzy partition $P = \{P_1, P_2, \dots P_k\}$ of V, each element $v \in V$ can be characterized by its membership vector $P(v) = \{P_1(v), P_2(v), \dots, P_k(v)\} \in [0, 1]^k$, where $P_i(v)$ indicates the extent of v's membership in the i^{th} community P_i. A fuzzy relation on V can be formulated using the similarity of the corresponding membership vector of vertices: $E_p(u, v) = 1 - ||P(u) - P^{'}(v)||$. Now, the "concordance" of two fuzzy community structures Ψ and C can be generated as follows: let us consider a pair (u, v) as being concordant as long as both Ψ

and C agree on their extent of equivalence. One can further consider the degree of concordance as $1 - |E_\Psi(u,v) - E_C(u,v)| \in [0,1]$. Analogously, the level of discordance is $|E_\Psi(u,v) - E_C(u,v)|$. Therefore, the distance measure on fuzzy community structures is then defined as follows:

$$d(\Psi, C) = \frac{\sum_{u,v \in V} |E_\Psi(u,v) - E_C(u,v)|}{n(n-1)/2}. \tag{83}$$

Likewise, $1 - d(\Psi, C)$ corresponds to the extent of normalized concordance and, therefore, is another generalization of the original Rand index.

[139] used the average **F1-score** to measure the equivalence of two overlapping partitions. Essentially, it measures the average of the following matching scores: the F1-score of the best-similar ground-truth community to each detected community, and the F1-score of the best-similar detected community to each ground-truth community as follows:

$$F1 = \frac{1}{2}\left(\frac{1}{|\Psi|}\sum_{\psi_i \in \Psi} F1(\psi_i, C_{g(i)}) + \frac{1}{|C|}\sum_{c_i \in C} F1(\Psi_{g'(i)}, c_i)\right), \tag{84}$$

where the best matching g and g' is defined as follows: $g(i) = \text{argmax}_j\, F1(\Omega_i, C_j)$, $g'(i) = \text{argmax}_j\, F1(\psi_j, C_i)$, and F1-score $F1(C_i, C_j)$ between two communities C_i and C_j is the harmonic mean of their precision and recall.

[139] also used **number of communities** to be the relative accuracy between the discovered and the actual number of communities: $1 - \frac{|\Psi| - |C|}{2|C|}$.

Precision, Recall, and F-measure (see Equation (60)) are also used to compare two overlapping partitions [134]. [130] used three measures: **sensitivity,** measured by the fraction of actual overlapping nodes detected by the algorithm; **specificity,** measured by the actual nonoverlapping nodes detected by the algorithm; and **accuracy,** measured by the sum of sensitivity and specificity after considering them equally.

Note that for all of these metrics, higher values mean more "accurately" detected communities. An upper bound of 1 is obtained when the detected community structure perfectly matches with the ground-truth community structure.

A summary of the metrics discussed in this subsection can be found in the SI text (Table 4).

4 EXPERIMENTS AND RESULTS

In this section, we demonstrate the efficacy of the state-of-the-art community scoring functions as an indicator to measure the goodness of the community structure. In particular, we concentrate on the metrics used for evaluating nonoverlapping and overlapping community structures. First, we discuss the benchmark datasets used in this experiment. Following this, we elaborate on the experimental setup and the results for nonoverlapping and overlapping community structures.

4.1 Benchmark Datasets

We take both the synthetic and the real-world networks with the known community structure.

4.1.1 Datasets with Nonoverlapping Community Structure. We examine a set of artificially generated networks and three real-world complex networks used in [23].

Synthetic networks: We select the LFR benchmark model [71] to generate synthetic networks whose underlying community structures are known a priori. We can directly control the following parameters associated with the model to create various networks: number of nodes n, average degree of nodes k, and maximal degree k_{max}; an exponent β for the distribution of the community size; an exponent γ for the degree distribution; and mixing coefficient μ. The mixing coefficient μ represents the desired ratio of the inter- and intracommunity edges. In this experiment, we vary

Table 2. Description of Real-World Networks with Nonoverlapping Community Structure

Network	n	e	k	k_{max}	c	n_c^{max}	n_c^{min}	f
Football	115	613	10.57	12	12	13	5	0
Railway	301	1,224	6.36	48	21	46	1	0.301
Coauthorship	103,677	352,183	5.53	1,230	24	14,404	34	0.001

n: number of nodes; e: number of edges; c: number of communities; k: average degree of nodes; k_{max}: maximum degree of nodes; n_c^{min}: smallest size of the community; n_c^{max}: largest size of the community; f: percentage of nodes for which the ground truth is unknown.

the number of nodes (n) and mixing coefficient (μ) to generate different network structures. For all other parameters, we consider the default values as mentioned in [71]. Note that for each parameter configuration, we generate 100 LFR networks, and the values in all the experiments are reported by averaging the results.

Real-world networks: We use three real-world networks whose properties are summarized in Table 2.

Football network [49] is a network of American football games between Division IA colleges. The nodes in the network are the football teams, and edges represent regular-season games between the two teams. The teams are divided into conferences (indicating communities), and games happen more frequently between members of the same conference than between members of different conferences.

Railway network consists of nodes representing railway stations in India [23]. Two nodes are connected by an edge if the corresponding stations are connected by at least one train. Here the communities are states/provinces in India because it is more likely that the number of trains is higher within a state than across states.

Coauthorship network is generated from the computer science citation dataset released by [20, 21]. Here each node corresponds to an author, and two authors are connected if they have written at least one paper together. Each author is associated with a research field (such as AI, Algorithms, Databases) on which he or she wrote most of the papers. The fields act as the ground-truth communities since authors tend to collaborate with each other more frequently within the same field than across fields.

4.1.2 Datasets with Overlapping Community Structure. We also examine various synthetic and real-world networks whose ground-truth communities are overlapping in nature.

Synthetic networks: We use the same LFR benchmark networks proposed by [71]. Along with the other parameters mentioned earlier, we can control two other parameters, namely, the percentage of overlapping nodes O_n and the number of community memberships of a node O_m. We vary the following parameters depending on the experimental need: n, μ, O_n, and O_m.

Real-world networks: We use three real networks with known overlapping ground-truth community structures [86, 138]. We can see that the ground-truth communities of a large number of nodes for LiveJournal and Youtube are not known. To use such incomplete ground-truth efficiently, there can be two possibilities: either the nodes without ground-truth community labels (and their attached edges) are removed from the original network and then the algorithms are run, or the algorithms are run on the original network first and during comparison the accuracy can be checked based on only those nodes whose ground-truth labels are available. Since the first approach might lead to a sparse network, we adopted the second approach in this article. The properties of these networks are summarized in Table 3.

Table 3. Description of the Real-World Networks with Overlapping Community Structure

Network	n	e	C	ρ	S	O_m	f
LiveJournal	3,997,962	34,681,189	310,092	0.536	40.02	3.09	71.29
Amazon	334,863	925,872	151,037	0.769	99.86	14.83	5.28
Youtube	1,134,890	2,987,624	8,385	0.732	43.88	2.27	71.29

n: number of nodes; e: number of edges; C: number of communities; ρ: average edge density per community; S: average community size; O_m: average number of communities a node is a part of; f: percentage of nodes for which the ground truth is unknown.

LiveJournal network contains nodes that are users in the LiveJournal blogging community and edges are friendship relationships. This site also allows users to form a group, which other members can then join. The groups defined by the users are considered as ground-truth communities.

Amazon network is based on the "Customers Who Bought This Item Also Bought" feature of the Amazon website. Here the nodes are the products, and one edge is drawn between two nodes if the corresponding products are copurchased quite frequently. The ground-truth community structure is marked by the product category provided by Amazon.

Youtube network contains nodes corresponding to the users on Youtube, and edges are formed due to the friendship with each other. The ground-truth community structure is marked by the user-defined groups.

Initially, we removed all nodes that are part of small ground-truth communities (communities with size less than five). We further observed that in most real-world networks, the coverage of the *entire network* by the ground-truth information is not available. Therefore, for the purpose of evaluation, we considered the following approach: we ran community detection algorithms on the original network and detected the community structure. Then we measured the similarity between the detected and the ground-truth communities based on only those nodes whose ground-truth labels are available.

Note that there has been a recent debate that ground-truth communities based on the metadata information of vertices may not reflect the true notion of communities in the network [60, 102]. Therefore, metadata and ground-truth labels should be treated differently. However, since most of the research on community detection dealt with the standard networks and the ground-truth communities mentioned earlier, we considered them in this article without delving further into this debate.

4.2 Experimental Setup

In the presence of ground-truth communities, the *validation metrics* (discussed in Section 3) compare the detected community structure with the available ground-truth results. This method of evaluation is considered to be more accurate and thus preferred widely. However, as mentioned earlier, for most of the real-world networks, the underlying ground-truth community structure is unknown. In that case, we evaluate the goodness of the detected communities using the *scoring metrics* (discussed in Section 2). There have been studies showing that in the absence of ground truth, if a majority of the scoring metrics rank a particular partition structure high, then the claim that the partitioning is good can be done more confidently [108, 109, 121]. However, it is not clear which scoring metric is most effective in this ranking task.

In this section, we design our experimental setup along the line proposed in [121] to understand which scoring metric highly correlates with the validation metrics. The one that has the highest correlation with the validation metrics can be used to measure the goodness of the detected community structure in the absence of ground truth.

We adopt the experimental setup discussed in [121]. Let us assume that $SM = \{SM_i\}$ and $VM = \{VM_j\}$ are sets of scoring and validation metrics, respectively. $CD = \{CD_1, CD_2, \ldots, CD_k\}$ is the set of k community detection algorithms. We perform the following steps:

(i) For each network, we execute k algorithms present in CD and obtain k different community structures.
(ii) For each of these community structures, we compute all the scoring metrics in SM separately.
(iii) The algorithms in CD are then ranked based on the value of each of these SM metrics separately, with the highest rank given to the highest value.
(iv) The community structures are further compared with the ground-truth labels of the network in terms of each of the validation metrics in VM separately.
(v) The algorithms are ranked again based on the values of each validation metric (highest value/best match has the best rank).
(vi) Finally, we obtain Spearman's rank correlation between the rankings obtained for each of the scoring metrics SM_i (step (iii)) and each of the ground-truth validation metrics VM_j (step (v)).

The validation metrics (namely, NMI, ONMI, ARI, etc.) are generally used to measure the correspondence between the detected and the ground-truth community structures; whereas the scoring metrics (such as modularity, conductance, etc.) assign a score to each community structure without knowing the underlying ground truth. Therefore, these two measures are orthogonal. However, since validation metrics are more reliable, the ranking obtained from a good scoring function should be more correlated with the same obtained from the validation metrics. We compare the relative ranks instead of the absolute values, because the ranges of the values are not equivalent, and we are more interested in checking the relative ordering of the algorithms obtained from each metric, and thus the goodness of the scoring metrics.

4.3 Comparison of Nonoverlapping Community Scoring Metrics

We compare the performance of seven state-of-the-art community scoring metrics used as goodness measures for nonoverlapping community structure: modularity (Mod), modularity density (MD), conductance (Con), communitude (Com), asymptotic surprise (Sur), significance (Sig), and permanence (Perm). These metrics form the set SM as mentioned in Section 4.2. For detecting communities from the synthetic and real-world networks, we use eight algorithms from the four most popular groups mentioned by [43] (representing CD in Section 4.2): (1) *modularity based:* FastGreedy [95], Louvain [13], CNM [32]; (2) *random walk based:* WalkTrap [107]; (3) *information theoretic based:* InfoMod [114] and InfoMap [115]; and (4) *genetic algorithm based:* GA-NET [103] and MOGA-NET [106]. To match the detected output with the ground-truth community structure, we consider five validation measures (representing VM in Section 4.2): variation of information (VI), normalized mutual information (NMI), adjusted Rand index (ARI), F-measure (F), and purity (Pu).

Figure 2 presents a comparative result of the seven scoring metrics for different LFR networks with nonoverlapping community structure. In most of the cases, a general trend is observed: permanence turns out to be superior among all, which is followed by modularity, although there are few exceptions where modularity outperforms others. In most cases, communitude stands as a third ranked metric, followed by modularity density and surprise. Conductance consistently performs the worst among all the metrics. In a few cases, we notice that while all the metrics show a decline, permanence tends to increase (Figures 2(d) and 2(e)) or remain consistent (Figure 2(b)).

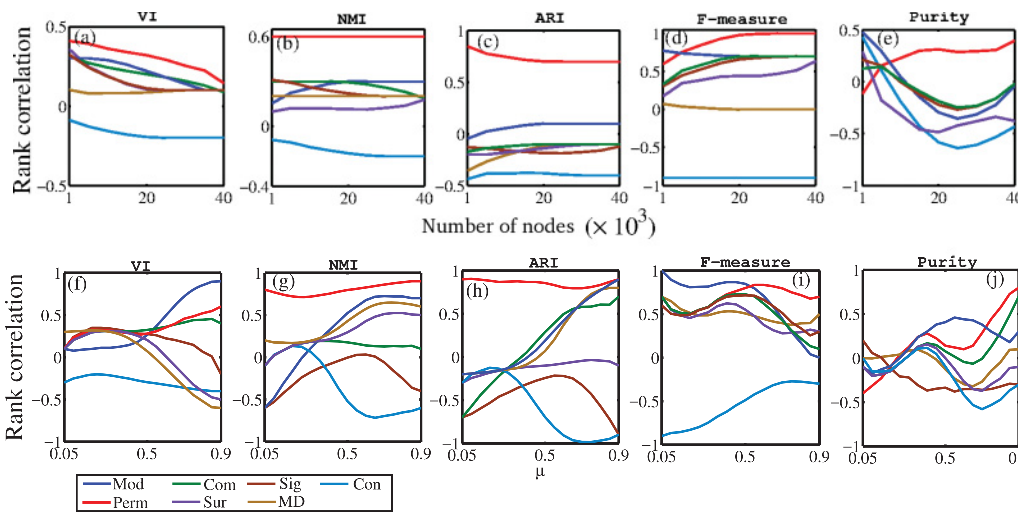

Fig. 2. (Color online) Spearman's rank correlation among the results obtained from seven scoring metrics and five validation measures for LFR networks with nonoverlapping community structure ((upper panel) varying the number of nodes, (lower panel) varying the value of μ).

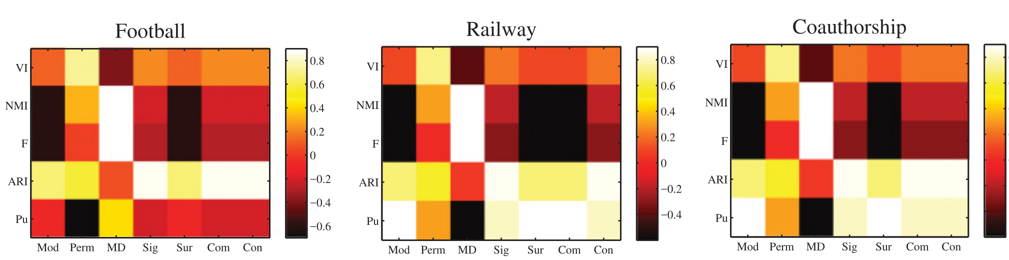

Fig. 3. (Color online) Heatmap depicting the Spearman's rank correlation among the results obtained from seven scoring metrics and five validation measures for real-world networks with nonoverlapping community structure.

Figure 3 presents a heatmap depicting the rank correlation for real-world networks. We notice that for the football network, modularity density outperforms others with the average rank correlation of 0.37 (over all the validation measures), followed by permanence (0.13), significance (0.08), communitude (0.07), conductance (0.04), modularity (-0.11), and surprise (-0.11). For the railway network, the result is slightly different, where permanence (0.37) outperforms others. For the coauthorship network, which is reasonably sparse and constitutes weaker community structure, permanence (0.37) turns out to be the best, followed by significance (0.27), communitude (0.27), and conductance (0.27). In short, on average, permanence performs better than other state-of-the-art metrics irrespective of the underlying network structure and validation measures.

4.4 Comparison of Overlapping Community Scoring Metrics

We further compare the performance of the five overlapping community scoring metrics: Q_{ov}^{Z} (Equation (42)), Q_{ov}^{N} (Equation (47)), Q_{ov}^{S} (Equation (45)), Q_{ov}^{MD} (Equation (53)), and flex (Equation (56)). These metrics form the set SM, mentioned in Section 4.2. For the purpose of evaluation, we take four ground-truth-based measures (representing VM): ONMI, Omega index, generalized

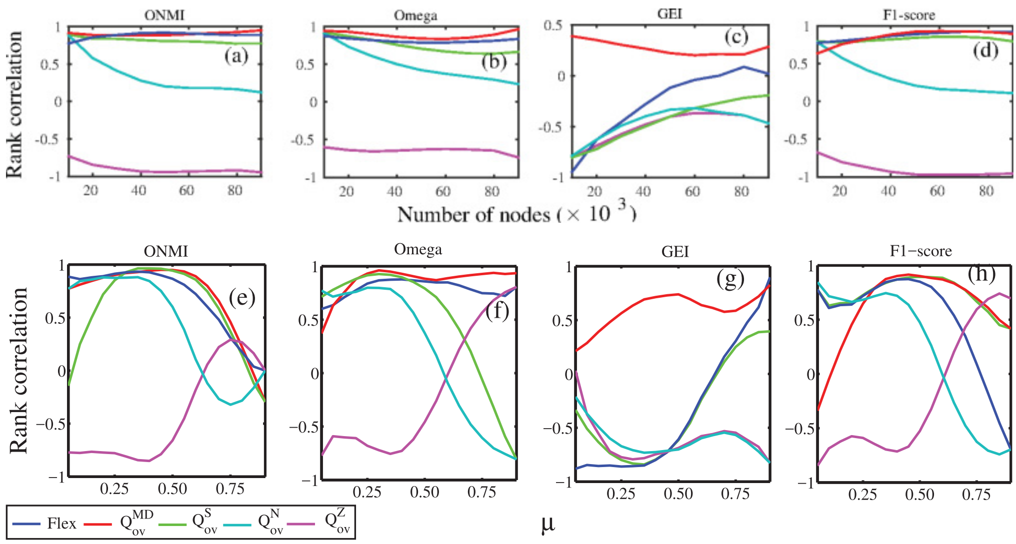

Fig. 4. (Color online) Spearman's rank correlation among the results obtained from five overlapping community scoring metrics and four validation measures for LFR networks with overlapping community structure: varying the number of nodes n ($\mu = 0.3$, $O_m = 5$, $O_n = 10\%$); mixing parameter μ ($n = 10,000$, $O_m = 5$, $O_n = 10\%$).

external index (GEI), and F1-score. We detect the overlapping community structure using seven popular algorithms that cover different types of overlapping community detection heuristics mentioned in [136]: (1) *local expansion and optimization:* OSLOM [73] and EAGLE [118]; (2) *agent-based dynamical algorithms:* COPRA [53] and SLPA [137]; (3) *fuzzy detection using mixture model:* MOSES [83] and BIGCLAM [139]; and (4) *genetic algorithm based:* GA-NET+ [105]. These algorithms form the set *CD*. The experiment discussed in Section 4.2 is repeated to check which one among *SM* highly corresponds to the results obtained from VM.

Figure 4 shows the results for the LFR networks by varying different parameters, that is, n and μ. We also vary the parameters O_m and O_n (see Figure 3 in the SI text). For most of the cases, Q_{ov}^{MD} seems to be the best, which is followed by flex, Q_{ov}^{S}, Q_{ov}^{N}, and Q_{ov}^{Z}. Most surprisingly, if we look at the trends carefully in Figure 4, we notice that the pattern obtained by comparing with GEI is significantly different from the others. This indicates that the GEI-based validation measure may not be a good performance indicator for community evaluation.

The heatmaps in Figure 5 show the performance of the scoring metrics for real-world networks. We compute the correlation of the rank of the algorithms as discussed in Section 4.2. For the LiveJournal, Amazon, and Youtube networks, the average correlations (over all validation measures) are reported sequentially (delimited by comma): flex (0.16, 0.26, 0.08), Q_{ov}^{Z} (-0.27, -0.09, -0.43), Q_{ov}^{MD} (0.16, 0.46, 0.19), Q_{ov}^{S} (0.05, 0.39, -0.29), and Q_{ov}^{N} (0.16. -0.37, -0.15). While the correlation seems to be positive (almost neutral) for flex and Q_{ov}^{MD}, Q_{ov}^{Z} and Q_{ov}^{N} seem to be negatively correlated with the validation metrics. In short, although Q_{ov}^{MD} seems to have higher correlation with the validation metrics, there is no metric that performs well on all kinds of networks.

4.5 Intuitive Justification of the Superiority of Permanence and Q_{ov}^{MD}

The theoretical justification behind the superior performance of permanence is mentioned in [23]. The intuitive idea is that since it considers the internal degree and total degree of a vertex, it

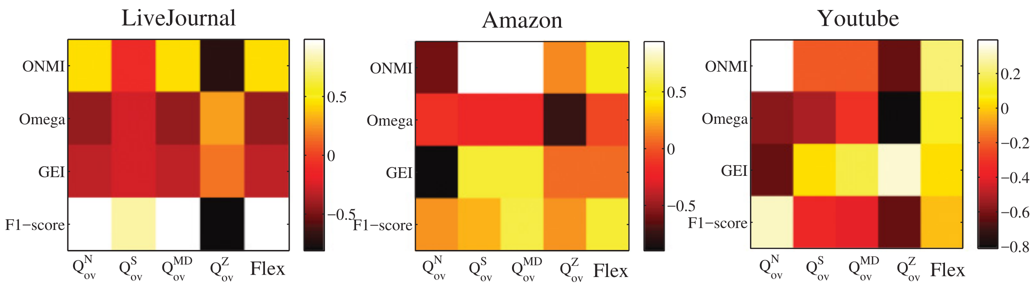

Fig. 5. (Color online) Heatmap depicting the Spearman's rank correlation among the results obtained from five overlapping community scoring metrics and four validation measures for real-world networks with overlapping community structure.

captures the notion of modularity to some extent. Further, it considers how the internal neighbors of a vertex are connected, which provides additional information about the stability of a vertex within a community. Moreover, it reduces the tie-breaking situation while assigning a vertex to a community. [23] provided two conditions to show how a vertex is assigned to a community.

Let vertex v be connected to α (β) nodes in community A (B), and these α (β) nodes form the set N_α (N_β). The number of vertices in community A is $(x + \alpha)$, and in community B is $(y + \beta)$. Let the average internal degree of a vertex $a \in N_\alpha$ and a vertex $b \in N_\beta$, before v is assigned to any of the communities, be I_α and I_β, respectively. Let the average internal clustering coefficient of the neighboring nodes in communities A and B be C_A and C_B, respectively. If v is added to communities A (B), then the average internal clustering coefficient of v becomes C^v_A (C^v_B), and the average internal clustering coefficient of the nodes in $N_\alpha(N_\beta)$ becomes C^α (C^β).

Condition I. If $\alpha = \beta$ *and* $C^\beta = C_B\frac{I_\beta-1}{I_\beta+1}$, *then communities A, B, and v will remain separate rather than v joining community A, if* $\alpha(\frac{2C_A-1}{I_\alpha+1}) + (1 - C^v_A) \geq \frac{1}{2\alpha}$.

Condition II. Joining v to community A gives higher permanence than merging the communities A, B, and v if $C^\beta = C_B$, *and* $(\frac{\gamma}{(\gamma+1)\beta} + \frac{C^v_A(2\gamma+1)-C^v_B}{(\gamma+1)^2} - \frac{\beta}{I_\beta+1}) > 1$, *where* $\gamma = \alpha/\beta$, *and also if* $C^\beta = C_B\frac{I_\beta-1}{I_\beta+1}$, *and* $(\frac{\gamma}{(\gamma+1)\beta} + \frac{C^v_A(2\gamma+1)-C^v_B}{(\gamma+1)^2} + \frac{\beta(2C_B-1)}{I_\beta+1}) > 1$.

More explanations can be obtained in [23].

On the other hand, extended modularity density Q^{MD}_{ov} considers two additional components, *split penalty* and *community density*, into Newman's modularity. Split penalty is the fraction of edges that connect nodes of different communities. Community density includes internal community density and pair-wise community density. These constraints reduce the resolution limit problem and provide additional benefit to handle the tie-breaking situation during node-to-community assignment [29].

5 CONCLUSION

Detecting communities has been a well-studied problem. However, despite such a vast extent of research in the detection and analysis of community structure, researchers are often in doubt when selecting an appropriate measurement metric. In this survey, we briefly introduced the problem of community detection. Based on the type of communities detected, we classified the quality metrics into different categories (nonoverlapping and overlapping). Next, for each category, we performed experiments to evaluate the goodness of these metrics and discussed the promising metrics. Most of the community scoring metrics are also used to evaluate the community structure. We hope that presenting all kinds of metrics together will enable the readers to understand the evolution chain of these metrics and provide them with the opportunity to select the right metric in the right context.

Three major observations derived from this review are noted as follows:

- In reviewing the quality metrics, we observed that the most popular and widely accepted metric in the literature of community analysis is Newman-Grivan's *modularity*, which also lays the foundation for other metrics. A large number of modified forms of modularity have been proposed.
- Although the drawbacks of modularity have been addressed several times, there are rare occasions where a completely new understanding of a community structure has been presented. Therefore, most of the exploration is along a narrow direction. It would be interesting to view communities in novel ways. Recent explorations in this direction include surprise, significance, and permanence.
- Our experimental results indicate that permanence and extended modularity density (Q_{ov}^{MD}) are most appropriate in measuring the quality of a community structure compared to the other competing metrics for nonoverlapping and overlapping community detection, respectively.

The field of community analysis is still in an exploratory stage, despite such an enormous literature over the last decade. The notion of community structure proposed by physicists [43, 49] is mostly criticized by computer scientists [23, 138, 139] by analyzing groups/modules present in real-world networks. Further, the debate on whether user-generated groups in real social/information networks should be treated as ground truth for evaluation of community detection algorithms has recently been initiated [60, 102] and needs more investigation. However, we believe that the current survey would unfold a complete landscape of how a community structure has been pursued so far by different quantitative metrics, which may help researchers overcome the existing drawbacks and put an end to this ever-growing area of research.

An interesting question is yet to be addressed: given a network, do we apply algorithms for nonoverlapping detection or overlapping community detection? There is no such metric/algorithm that is able to detect nonoverlapping as well as overlapping community structure depending on the network topology without knowing the type of the underlying community structure. Moreover, there is very limited literature addressing the effect of various network noises on the behavior of the metric [23]. [55, 113] explored the significance of community structures only for the case of nonoverlapping community detection, which is based on the notion of modularity, although the same is yet to be explored for overlapping community structure and for other goodness metrics. Last but not least, we believe that metrics are the core component of the community detection algorithms, and this review would have a huge implication in understanding the plethora of research in the area of community analysis.

REFERENCES

[1] E. Abbe and C. Sandon. 2015. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS'15)*. 670–688. DOI: http://dx.doi.org/10.1109/FOCS.2015.47

[2] R. Aldecoa and I. Marín. 2013. Surprise maximization reveals the community structure of complex networks. *Scientific Reports* 3 (2013).

[3] R. Aldecoa and I. Marín. 2011. Deciphering network community structure by surprise. *PloS One* 6, 9 (2011), e24195.

[4] A. Arenas, J. Duch, A. Fernndez, and S. Gmez. 2007. Size reduction of complex networks preserving modularity. *New Journal of Physics* 9, 6 (2007), 176.

[5] A. Arenas, A. Fernndez, S. Fortunato, and S. Gmez. 2008. Motif-based communities in complex networks. *Journal of Physics A: Mathematical and Theoretical* 41, 22 (2008), 224001.

[6] A. Arenas, A. Fernndez, and S. Gmez. 2008. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics* 10, 5 (2008), 053039.

[7] J. Artiles, J. Gonzalo, and S. Sekine. 2007. The SemEval-2007 WePS evaluation: Establishing a benchmark for the web people search task. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval'07)*. Association for Computational Linguistics, 64–69.

[8] T. Aynaud and J.-L. Guillaume. 2010. Static community detection algorithms for evolving networks. In *Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'10)*. 513–519.

[9] M. F. Balcan and Y. Liang. 2013. Modeling and detecting community hierarchies. In *Proceedings of the 2nd International Conference on Similarity-Based Pattern Recognition (SIMBAD'13)*. Springer-Verlag, Berlin,160–175.

[10] J. Baumes, M. K. Goldberg, M. S. Krishnamoorthy, M. Magdon-Ismail, and N. Preston. 2005. Finding communities by clustering a graph into overlapping subgraphs.*IADIS AC* 5 (2005), 97–104.

[11] A. R. Benson, D. F. Gleich, and J. Leskovec. 2016. Higher-order organization of complex networks. *Science* 353, 6295 (2016), 163–166. DOI: http://dx.doi.org/10.1126/science.aad9029

[12] P. Berkhin. 2006. A survey of clustering data mining techniques. In *Grouping Multidimensional Data - Recent Advances in Clustering*. 25–71.

[13] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. L. J. S. Mech. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics* 2008 (2008), P10008.

[14] R. J. G. B. Campello. 2007. A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. *Pattern Recognition Letters* 28, 7 (2007), 833–841.

[15] R. J. G. B. Campello. 2010. Generalized external indexes for comparing data partitions with overlapping categories. *Pattern Recognition Letters* 31, 9 (2010), 966–975.

[16] A. Chakraborty, S. Ghosh, and N. Ganguly. 2012. Detecting overlapping communities in folksonomies. In *23rd ACM Conference on Hypertext and Social Media*. 213–218.

[17] T. Chakraborty. 2015. Leveraging disjoint communities for detecting overlapping community structure. *Journal of Statistical Mechanics: Theory and Experiment* 2015, 5 (2015), P05017.

[18] T. Chakraborty and A. Chakraborty. 2013. OverCite: Finding overlapping communities in citation network. In *International Conference on Advances in Social Network Analysis and Mining (ASONAM'13)*. 1124–1131.

[19] T. Chakraborty, S. Kumar, N. Ganguly, A. Mukherjee, and S. Bhowmick. 2016. GenPerm: A unified method for detecting non-overlapping and overlapping communities. *IEEE Trans. Knowl. Data Eng.* 28, 8 (2016), 2101–2114. DOI: 10.1109/TKDE.2016.2554119

[20] T. Chakraborty, S. Sikdar, N. Ganguly, and A. Mukherjee. 2014. Citation interactions among computer science fields: A quantitative route to the rise and fall of scientific research. *Social Network Analysis Mining* 4, 1 (2014), 187.

[21] T. Chakraborty, S. Sikdar, V. Tammana, N. Ganguly, and A. Mukherjee. 2013. Computer science fields as ground-truth communities: Their impact, rise and fall. In *Advances in Social Networks Analysis and Mining 2013 (ASONAM'13)*. 426–433.

[22] T. Chakraborty, S. Srinivasan, N. Ganguly, S. Bhowmick, and A. Mukherjee. 2013. Constant communities in complex networks. *Scientific Reports* 3 (2013).

[23] T. Chakraborty, S. Srinivasan, N. Ganguly, A. Mukherjee, and S. Bhowmick. 2014. On the permanence of vertices in network communities. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. ACM, New York,1396–1405.

[24] T. Chakraborty, S. Srinivasan, N. Ganguly, A. Mukherjee, and S. Bhowmick. 2016. Permanence and community structure in complex networks. *CoRR* abs/1606.01543 (2016).

[25] D. Chen, M. Shang, Z. Lv, and Y. Fu. 2010. Detecting overlapping communities of weighted networks via a local algorithm. *Physica A: Statistical Mechanics and its Applications* 389, 19 (2010), 4177–4187.

[26] G. Chen, Y. Wang, and J. Wei. 2013. A new multiobjective evolutionary algorithm for community detection in dynamic complex networks. *Mathematical Problems in Engineering* 2013 (2013), 1–7.

[27] J. Chen, O. R. Zaane, and R. Goebel. 2009. Detecting communities in social networks using max-min modularity. In *SDM*. SIAM, 978–989.

[28] M. Chen, K. Kuzmin, and B. K. Szymanski. 2014. Community detection via maximization of modularity and its variants. *IEEE Transactions on Computational Social Systems* 1, 1 (March 2014), 46–65.

[29] M. Chen, K. Kuzmin, and B. K. Szymanski. 2014. Extension of modularity density for overlapping community structure. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'14)*. 856–863.

[30] M. Chen, T. Nguyen, and B. K. Szymanski. 2013. On measuring the quality of a network community structure. In *2013 International Conference on Social Computing (SocialCom'13)*. 122–127.

[31] C. Chira, A. Gog, and D. Iclanzan. 2012. Evolutionary detection of community structures in complex networks: A new fitness function. In *IEEE Congress on Evolutionary Computation (CEC'12)*. 1–8.
[32] A. Clauset, M. E. J. Newman, and C. Moore. 2004. Finding community structure in very large networks. *Physical Review E* 70, 6 (2004), 066111.
[33] L. M. Collins and C. W. Dent. 1988. Omega: A general formulation of the rand index of cluster recovery suitable for non-disjoint solutions. *Multivariate Behavioral Research* 23, 2 (1988), 231–242.
[34] J. Creusefond, T. Largillier, and S. Peyronnet. 2014. Finding compact communities in large graphs. *CoRR* abs/1410.2105 (2014).
[35] J. Creusefond, T. Largillier, and S. Peyronnet. 2015. On the evaluation potential of quality functions in community detection for different contexts. *CoRR* arXiv:1510.01714 (2015).
[36] L. Danon, A. Daz-Guilera, J. Duch, and A. Arenas. 2005. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment* 2005, 9 (2005), P09008.
[37] F.-O. de Frana and G.-P. Coelho. 2015. A flexible fitness function for community detection in complex networks. In *Complex Networks VI*, Giuseppe Mangioni, Filippo Simini, Stephen Miles Uzzo, and Dashun Wang (Eds.). Studies in Computational Intelligence, Vol. 597. Springer International Publishing, 1–12.
[38] J. Duch and A. Arenas. 2005. Community detection in complex networks using extremal optimization. *Physical Review E* 72, 2 (2005), 027104.
[39] B. Faltings, K. Leyton-Brown, and P. Ipeirotis (Eds.). 2012. *ACM Conference on Electronic Commerce (EC'12)*. ACM. http://dl.acm.org/citation.cfm?id=2229012
[40] Z. Feng, X. Xu, N. Yuruk, and T. A. J. Schweiger. 2007. A novel similarity-based modularity function for graph partitioning. In *Data Warehousing and Knowledge Discovery*, Ilyeal Song, Johann Eder, and Thomanh Nguyen (Eds.). Lecture Notes in Computer Science, Vol. 4654. Springer, Berlin, 385–396.
[41] G. W. Flake, S. Lawrence, and C. Lee Giles. 2000. Efficient identification of web communities. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'00)*. ACM, New York, 150–160.
[42] T. Fleck, A. Kappes, and D. Wagner. 2014. Graph clustering with surprise: Complexity and exact solutions. In *Theory and Practice of Computer Science (SOFSEM'14*), Viliam Geffert, Bart Preneel, Branislav Rovan, Jlius tuller, and Amin Tjoa (Eds.). Lecture Notes in Computer Science, Vol. 8327. Springer International Publishing, 223–234.
[43] S. Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3–5 (2010), 75–174.
[44] S. Fortunato and M. Barthlemy. 2007. Resolution limit in community detection. *PNAS* 104, 1 (2007), 36–41.
[45] S. Fortunato and A. Lancichinetti. 2009. Community detection algorithms: A comparative analysis: Invited presentation, extended abstract. In *4th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'09)*. 27.
[46] E. B. Fowlkes and C. L. Mallows. 1983. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association* 78, 383 (1983), 553–569.
[47] A. Fronczak, P. Fronczak, and J. A. Hołyst. 2004. Average path length in random networks. *Physical Review E* 70, 5 (Nov.2004), 056110.
[48] R. Ghosh and K. Lerman. 2010. Community detection using a measure of global influence. In *Proceedings of the 2nd International Conference on Advances in Social Network Mining and Analysis (SNAKDD'08)*. Springer-Verlag, Berlin,20–35.
[49] M. Girvan and M. E. Newman. 2002. Community structure in social and biological networks. *PNAS* 99, 12 (June 2002), 7821–7826.
[50] M. Gong, Q. Cai, Y. Li, and J. Ma. 2012. An improved memetic algorithm for community detection in complex networks. In *2012 IEEE Congress on Evolutionary Computation*. IEEE, 1–8.
[51] M. Gong, B. Fu, L. Jiao, and H. Du. 2011. Memetic algorithm for community detection in networks. *Physical Review E* 84, 5 (2011), 056101.
[52] B. H. Good, Y.-A. de Montjoye, and A. Clauset. 2010. Performance of modularity maximization in practical contexts. *Physical Review E* 81, 4 (Apr. 2010), 046106.
[53] S. Gregory. 2010. Finding overlapping communities in networks by label propagation. *New Journal of Physics* 12, 10 (2010), 103018.
[54] S. Gregory. 2011. Fuzzy overlapping communities in networks. *Journal of Statistical Mechanics: Theory and Experiment* 2011, 2 (2011), P02017.
[55] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. 2004. Modularity from fluctuations in random graphs and complex networks. *Physical Review E* 70, 2 (2004), 025101.
[56] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. 2001. On clustering validation techniques. *Journal of Intelligent Information Systems* 17, 2–3 (Dec. 2001), 107–145.

[57] S. Harenberg, G. Bello, L. Gjeltema, S. Ranshous, J. Harlalka, R. Seay, K. Padmanabhan, and N. Samatova. 2014. Community detection in large-scale networks: A survey and empirical evaluation. *WIREs Computer Statistics* 6 (2014), 426–439.

[58] F. Havemann, M. Heinz, A. Struck, and J. Glser. 2011. Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels. *Journal of Statistical Mechanics: Theory and Experiment* 2011, 1 (2011), P01023.

[59] D. S. Hochbaum. 2010. Polynomial time algorithms for ratio regions and a variant of normalized cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 5 (2010), 889–898.

[60] D. Hric, R. K. Darst, and S. Fortunato. 2014. Community detection in networks: Structural communities versus ground truth. *Physical Review E* 90, 6 (Dec. 2014), 062805. DOI: http://dx.doi.org/10.1103/PhysRevE.90.062805

[61] L. Hubert and P. Arabie. 1985. Comparing partitions. *Journal of Classification* 2, 1 (1985), 193–218.

[62] E. Hullermeier and M. Rifqi. 2009. A fuzzy variant of the Rand index for comparing clustering structures. In *IFSA/EUSFLAT Conference*, Joo Paulo Carvalho, Didier Dubois, Uzay Kaymak, and Joo Miguel da Costa Sousa (Eds.). 1294–1298.

[63] A. K. Jain and R. C. Dubes. 1988. *Algorithms for Clustering Data*. Prentice-Hall,Upper Saddle River, NJ.

[64] D. Jiang, C. Tang, and A. Zhang. 2004. Cluster analysis for gene expression data: A survey. *IEEE Transactions on Knowledge and Data Engineering* 16, 11 (Nov. 2004), 1370–1386.

[65] S. Kelley. 2011. *The Existence and Discovery of Overlapping Communities in Large-Scale Networks*. BiblioBazaar. https://books.google.com/books?id=mW4CywAACAAJ.

[66] Y. Kim, S.-W. Son, and H. Jeong. 2010. Finding communities in directed networks. *Physical Review E* 81, 1 (Jan 2010), 016103.

[67] A. Kraskov, H. Stgbauer, R. G. Andrzejak, and P. Grassberger. 2003. Hierarchical clustering based on mutual information. *CoRR* q-bio.QM/0311039 (2003).

[68] V. Labatut. 2013. Generalized measures for the evaluation of community detection methods. *CoRR* abs/1303.5441 (2013).

[69] A. Lancichinetti and S. Fortunato. 2009. Community detection algorithms: A comparative analysis. *Physical Review E* 80, 5 (Nov 2009), 056117.

[70] A. Lancichinetti and S. Fortunato. 2011. Limits of modularity maximization in community detection. *Physical Review E* 84, 6 (Dec 2011), 066122.

[71] A. Lancichinetti, S. Fortunato, and J. Kertész. 2009. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* 11, 3 (2009), 033015.

[72] A. Lancichinetti, M. Kivel, J. Saramki, and S. Fortunato. 2010. Characterizing the community structure of complex networks. *PLoS ONE* 5, 8 (2010), e11976.

[73] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato. 2011. Finding statistically significant communities in networks. *PLoS ONE* 6, 4 (2011), e18961.

[74] A. Lázár, D. Ábel, and T. Vicsek. 2010. Modularity measure of networks with overlapping communities. *EPL (Europhysics Letters)* 90, 1 (2010), 18001.

[75] J. Leskovec, K. J. Lang, and M. Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*. ACM, New York, 631–640.

[76] X. Li, B. Wu, Q. Guo, X. Zeng, and C. Shi. 2015. Dynamic community detection algorithm based on incremental identification. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW'15)*. 900–907.

[77] Z. Li, S. Zhang, R.-S. Wang, X.-S. Zhang, and L. Chen. 2008. Quantitative function for community detection. *Physical Review E* 77, 3 (2008), 036109.

[78] T. Y. Lin, S. Ohsuga, C.-J. Liau, and X. Hu (Eds.). 2006. *Foundations and Novel Approaches in Data Mining*. Studies in Computational Intelligence, Vol. 9. Springer.

[79] X. Liu, T. Murata, and K. Wakita. 2012. Extending modularity by capturing the similarity attraction feature in the null model. *arXiv preprint arXiv:1210.4007* (2012).

[80] M. W. Mahoney, A. Dasgupta, K. J. Lang, and J. Leskovec. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6, 1 (2009), 29–123.

[81] C. D. Manning, P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York.

[82] C. P. Massen and J. P. K. Doye. 2005. Identifying communities within energy landscapes. *Physical Review E* 71, 4 (Apr 2005), 046101.

[83] A. McDaid and N. Hurley. 2010. Detecting highly overlapping communities with model-based overlapping seed expansion. In *ASONAM*. Washington, DC,112–119.

[84] A. F. McDaid, D. Greene, and N. J. Hurley. 2011. Normalized mutual information to evaluate overlapping community finding algorithms. *CoRR* abs/1110.2515 (2011).

[85] M. Meilă. 2007. Comparing clusterings—An information based distance. *Journal of Multivariate Analysis* 98, 5 (May 2007), 873–895.
[86] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC'07)*. ACM, New York, 29–42. DOI: http://dx.doi.org/10.1145/1298306.1298311
[87] A. Miyauchi and Y. Kawase. 2015. What is a network community?: A novel quality function and detection algorithms. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM'15)*. ACM, New York, 1471–1480.
[88] A. Miyauchi and Y. Kawase. 2015. Z-score-based modularity for community detection in networks. *CoRR* abs/1501.01909 (2015).
[89] C.-H. Mu, J. Xie, Y. Liu, F. Chen, Y. Liu, and L.-C. Jiao. 2015. Memetic algorithm with simulated annealing strategy and tightness greedy optimization for community detection in networks. *Applied Soft Computing* 34 (2015), 485–501.
[90] S. Muff, F. Rao, and A. Caflisch. 2005. Local modularity measure for network clusterizations. *Physical Review E* 72, 5 (2005), 056107.
[91] G. Murray, G. Carenini, and R. Ng. 2012. Using the omega index for evaluating abstractive community detection. In *Proceedings of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization*. Association for Computational Linguistics, Stroudsburg, PA, 10–18.
[92] T. Nepusz, A. Petróczi, L. Négyessy, and F. Bazsó. 2008. Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E* 77, 1 (Jan. 2008), 016107.
[93] M. E. J. Newman. 2004. Analysis of weighted networks. *Physical Review E* 70, 5 (Nov. 2004), 056131.
[94] M. E. J. Newman. 2004. Detecting community structure in networks. *EPJB* 38, 2 (March 2004), 321–330.
[95] M. E. J. Newman. 2004. Fast algorithm for detecting community structure in networks. *Physical Review E* 69, 6 (June 2004), 066133.
[96] M. E. J. Newman. 2006. Modularity and community structure in networks. *PNAS* 103, 23 (2006), 8577–8582.
[97] M. E. J. Newman and M. Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69, 2 (Feb. 2004), 026113.
[98] M. E. J. Newman and E. A. Leicht. 2007. Mixture models and exploratory analysis in networks. *PNAS* 104, 23 (2007), 9564–9569.
[99] V. Nicosia, G. Mangioni, V. Carchiolo, and M. Malgeri. 2009. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment* 2009, 03 (2009), P03024. http://stacks.iop.org/1742-5468/2009/i=03/a=P03024.
[100] G. K. Orman, V. Labatut, and H. Cherifi. 2012. Comparative evaluation of community detection algorithms: A topological approach. *Journal of Statistical Mechanics* 8 (2012), p08001.
[101] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 7043 (2005), 814–818.
[102] L. Peel, D. B. Larremore, and A. Clauset. 2016. The ground truth about metadata and community detection in networks. *CoRR* abs/1608.05878 (2016). http://arxiv.org/abs/1608.05878
[103] C. Pizzuti. 2008. GA-net: A genetic algorithm for community detection in social networks. In *PPSN*, Gnter Rudolph, Thomas Jansen, Simon M. Lucas, Carlo Poloni, and Nicola Beume (Eds.), Vol. 5199. Springer, 1081–1090.
[104] C. Pizzuti. 2008. GA-net: A genetic algorithm for community detection in social networks. In *Parallel Problem Solving from Nature PPSN X*, Gnter Rudolph, Thomas Jansen, Nicola Beume, Simon Lucas, and Carlo Poloni (Eds.). Lecture Notes in Computer Science, Vol. 5199. Springer, Berlin, 1081–1090.
[105] C. Pizzuti. 2009. Overlapped community detection in complex networks. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*. 859–866.
[106] C. Pizzuti. 2012. A multiobjective genetic algorithm to find communities in complex networks. *IEEE Transactions on Evolutionary Computation* 16, 3 (June 2012), 418–430.
[107] P. Pons and M. Latapy. 2006. Computing communities in large networks using random walks. *Journal of Graph Algortihms and Applications* 10, 2 (2006), 191–218.
[108] R. Rabbany, M. Takaffoli, J. Fagnan, O. R. Zaıane, and R. J. G. B. Campello. 2013. Communities validity: Methodical evaluation of community mining algorithms. *Social Netw. Analys. Mining* 3, 4 (2013), 1039–1062.
[109] R. Rabbany and O. R. Zaane. 2015. Generalization of clustering agreements and distances for overlapping clusters and network communities. *Data Mining and Knowledge Discovery* 29, 5 (2015), 1458–1485.
[110] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. 2004. Defining and identifying communities in networks. *PNAS* 101, 9 (2004), 2658.
[111] J. Reichardt and S. Bornholdt. 2004. Detecting fuzzy community structures in complex networks with a potts model. *Physical Review Letters* 93, 21 (Nov 2004), 218701.

[112] J. Reichardt and S. Bornholdt. 2006. Statistical mechanics of community detection. *Physical Review E* 74, 1 (July 2006), 016110.
[113] J. Reichardt and S. Bornholdt. 2006. When are networks truly modular?*Physica D Nonlinear Phenomena* 224 (December 2006), 20–26.
[114] M. Rosvall and C. T. Bergstrom. 2007. An information-theoretic framework for resolving community structure in complex networks. *PNAS* 104, 18 (2007), 7327.
[115] M. Rosvall and C. T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *PNAS* 105, 4 (2008), 1118–1123.
[116] S. E. Schaeffer. 2007. Survey: Graph clustering. *Computer Science Review* 1, 1 (Aug. 2007), 27–64. DOI: http://dx.doi.org/10.1016/j.cosrev.2007.05.001
[117] P. Schuetz and A. Caflisch. 2008. Multistep greedy algorithm identifies community structure in real-world and computer-generated networks. *Physical Review E* 78, 2 (2008), 026112.
[118] H. Shen, X. Cheng, K. Cai, and M.-B. Hu. 2009. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications* 388, 8 (2009), 1706–1712.
[119] H. Shen, X. Cheng, and J.-F. Guo. 2009. Quantifying and identifying the overlapping community structure in networks. *Journal of Statistical Mechanics: Theory and Experiment* 2009, 7 (2009), P07042.
[120] J. Shi and J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 8 (Aug. 2000), 888–905.
[121] K. Steinhaeuser and N. V. Chawla. 2010. Identifying and evaluating community structure in complex networks. *Pattern Recognition Letters* 31, 5 (April 2010), 413–421.
[122] A. Strehl and J. Ghosh. 2003. Cluster ensembles — A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3 (March 2003), 583–617.
[123] P. G. Sun, L. Gao, and Y. Yang. 2013. Maximizing modularity intensity for community partition and evolution. *Information Science* 236 (July 2013), 83–92.
[124] M. T. Thai and P. M. Pardalos. 2011. *Handbook of Optimization in Complex Networks: Theory and Applications*. Vol. 57. Springer Science & Business Media.
[125] V. A. Traag, R. Aldecoa, and J.-C. Delvenne. 2015. Detecting communities using asymptotical surprise. *Physical Review E* 92, 2 (Aug. 2015), 022816.
[126] V. A. Traag, G. Krings, and P. Van Dooren. 2013. Significant scales in community structure. *Scientific Reports* 3 (2013).
[127] T. van Laarhoven and E. Marchiori. 2014. Axioms for graph clustering quality functions. *Journal of Machine Learning Research* 15 (2014), 193–215.
[128] N. X. Vinh, J. Epps, and J. Bailey. 2009. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In *ICML*. ACM, New York, 1073–1080.
[129] K. Wakita and T. Tsurumi. 2007. Finding community structure in mega-scale social networks. *CoRR* abs/cs/0702048 (2007).
[130] Q. Wang and E. Fleury. 2012. Fuzziness and overlapping communities in large-scale networks. *Journal of Universal Computer Science* 18, 4 (2012), 457–486.
[131] D. J. Watts and S. H. Strogatz. 1998. Collective dynamics of'small-world' networks. *Nature* 393 (1998), 440–442.
[132] Y. C. Wei, C. K. Cheng, San Diego. Dept. of Computer Science University of California, and Engineering. 1990. *Ratio Cut Partitioning for Hierarchical Designs*. Dept. of Computer Science and Engineering, University of California, San Diego.
[133] Y.-C. Wei and C.-K. Cheng. 1989. Towards efficient hierarchical designs by ratio cut partitioning. In *1989 IEEE International Conference on Computer-Aided Design, 1989 (ICCAD'89). Digest of Technical Papers*. 298–301. DOI: http://dx.doi.org/10.1109/ICCAD.1989.76957
[134] J. J. Whang, D. F. Gleich, and I. S. Dhillon. 2013. Overlapping community detection using seed set expansion. In *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management (CIKM'13)*. ACM, New York, 2099–2108.
[135] J. Xiang, Y.-N. Tang, Y.-Y. Gao, Y. Zhang, K. Deng, X.-K. Xu, and K. Hu. 2015. Multi-resolution community detection based on generalized self-loop rescaling strategy. *Physica A: Statistical Mechanics and its Applications* 432 (2015), 127–139.
[136] J. Xie, S. Kelley, and B. K. Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys* 45, 4, Article 43 (Aug. 2013), 35 pages.
[137] J. Xie and B. K. Szymanski. 2012. Towards linear time overlapping community detection in social networks. In *PAKDD*. Kuala Lumpur, Malaysia, 25–36.
[138] J. Yang and J. Leskovec. 2012. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics (MDS'12)*. ACM, New York, 3:1–3:8.

[139] J. Yang and J. Leskovec. 2013. Overlapping community detection at scale: A nonnegative matrix factorization approach. In *WSDM*. ACM, New York, 587–596.
[140] S. Zhang, R.-S. Wang, and X.-S. Zhang. 2007. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A: Statistical Mechanics and Its Applications* 374, 1 (2007), 483–490.
[141] S. Zhang and H. Zhao. 2012. Community identification in networks with unbalanced structure. *Physical Review E* 85, 6 (Jun 2012), 066114.
[142] X.-S. Zhang, R.-S. Wang, Y. Wang, J. Wang, Y. Qiu, L. Wang, and L. Chen. 2009. Modularity optimization in community detection of complex networks. *EPL (Europhysics Letters)* 87, 3 (2009), 38002.
[143] Y. Zhen-Qing, Z. Ke, H. Song-Nian, and Y. Jun. 2012. A new definition of modularity for community detection in complex networks. *Chinese Physics Letters* 29, 9 (2012), 098901.

Received May 2016; revised April 2017; accepted May 2017