Network clustering and community detection using modulus of families of loops

Heman Shakeri,[1,*] Pietro Poggi-Corradini,[2] Nathan Albin,[2] and Caterina Scoglio[1]
[1]*Electrical and Computer Engineering Department, Kansas State University, Manhattan, Kansas 66506, USA*
[2]*Mathematics Department, Kansas State University, Manhattan, Kansas 66506, USA*
(Received 1 September 2016; published 17 January 2017)

We study the structure of loops in networks using the notion of modulus of loop families. We introduce an alternate measure of network clustering by quantifying the richness of families of (simple) loops. Modulus tries to minimize the expected overlap among loops by spreading the expected link usage optimally. We propose weighting networks using these expected link usages to improve classical community detection algorithms. We show that the proposed method enhances the performance of certain algorithms, such as spectral partitioning and modularity maximization heuristics, on standard benchmarks.

DOI: 10.1103/PhysRevE.95.012316

I. INTRODUCTION

Real networks contain closely connected subnetworks with local structural patterns characterized by their richness of loop [1]. Loops offer more pathways within them compared to treelike topologies; thus rich loop structures improve network robustness [2] and impact propagating and transporting processes in networks [3]. Previous approaches on analysis of loop structures focus on loops with lengths of order 3–5 separately [4,5] and a few such as Refs. [6,7] emphasize the role of higher order loops to characterize their overall structures. We consider assessing loop structures in the network, with any order and altogether, and apply our tool for analyzing network transitivity (known as clustering coefficient) and providing more information for community detection algorithms.

Our goal is to study loop structures in the network using the concept of modulus of loop families developed in Refs. [8–10]. Modulus is a way of measuring the richness of certain families of objects on a network, such as loops, walks, trees, etc., and is a discrete analog of the classical theory of modulus of curve families in complex analysis [11]. Although modulus on networks is not a new concept (see Refs. [12,13]), it is not as well developed as in the continuum setting. In Ref. [8], the authors showed that modulus is a standard convex optimization problem. Continuity and smoothness properties of modulus on networks were considered in Ref. [9]. A probabilistic interpretation provided in Ref. [10].

Modulus is a versatile tool to analyze networks. Different types of families of walks can be used to learn about different aspects of the network. In Ref. [14], we introduced centrality measures based on various families of walks that can be computed on directed or undirected, weighted or unweighted, and even disconnected networks. These measures do not necessarily have to consider the whole network. We applied them to detect influential sections of the network, ranking the nodes, and we explored applications to improve vaccination strategies for reducing the risk of epidemics. The applications to epidemic spreading were further studied in Ref. [15], where the authors used modulus to analyze the concept of epidemic hitting time.

Our main contributions in this paper are introducing a generic approach to analyze loop structures in the network that consider local loop topologies with an eye on the entire network. We quantify richness of loops and introduce a clustering measure based on that. Moreover, we find the probablity of usage of each link in important loops and use it as a measure of affinity between nodes to enhance network partitioning.

This paper is organized as follows. First, we introduce our notation and the necessary background on modulus of families of loops. Then, we define our proposed methods to measure clustering in the network. Next, we show how to preprocess a network in order to improve partitioning techniques such as Fiedler vector bisection and the modularity maximization heuristics. Finally, we discuss other potential applications.

II. NOTATIONS AND DEFINITIONS

Let $\mathcal{G} = (V,E)$ be a network with nodes V and links E. A *walk* is a string of nodes $\gamma = v_0 v_1 \dots v_n$ on $\mathcal{G}$ with the property that consecutive nodes v_i and v_{i+1} are linked in the network. A walk $\gamma = v_1 v_2 v_3 \dots v_r$, is a *simple loop* if the nodes v_i are all distinct, except that $v_r = v_1$. We call $\mathcal{L}$ the family of all loops in $\mathcal{G}$. Other possible loop families are loop families rooted at a given node v or link e; we write $\mathcal{L}^v$ or $\mathcal{L}^e$ in that case.

Given a density $\rho : E \to [0,\infty)$, interpreted as a penalty or cost the walker must pay for traversing link e, we define the ρ *length* of a loop γ as

$$\ell_\rho(\gamma) := \sum_{e\in\gamma} \rho(e). \tag{1}$$

When $\rho_0(e) \equiv 1$, then $\ell_{\rho_0}(\gamma)$ represents the hop length of γ. Likewise, given a family of loops $\mathcal{L}$ we set $\ell_\rho(\mathcal{L}) := \min_{\gamma\in\mathcal{L}} \ell_\rho(\gamma)$. We introduce a $|\mathcal{L}| \times |E|$ matrix $\mathcal{N}$ such that each row corresponds to a loop $\gamma \in \mathcal{L}$ and is the indicator function $\mathbb{1}_{e\in\gamma}$.

Let $w : E \to (0,\infty)$ be a positive weight function. Then, for $1 < p < \infty$, $\mathrm{Mod}_{p,w}(\mathcal{L})$ is defined as

$$\mathrm{Mod}_{p,w}(\mathcal{L}) = \min_{\{\rho | \ell_\rho(\mathcal{L})>0\}} \frac{\mathcal{E}_{p,w}}{\ell_\rho(\mathcal{L})^p}, \tag{2}$$

where $\mathcal{E}_{p,w}(\rho) = \sum_{e\in E} w(e)\rho(e)^p$ is the energy of the density ρ. In this paper, we work with an equivalent form of (2) defined as in Ref. [8]:

$$\mathrm{Mod}_{p,w}(\mathcal{L}) = \min_{\{\rho | \mathcal{N}\rho \geqslant 1\}} \mathcal{E}_{p,w}(\rho) = \mathcal{E}_{p,w}(\rho^*). \tag{3}$$

*Corresponding author: heman@ksu.edu

We call a density ρ with $\mathcal{N}\rho \geqslant 1$ admissible ρ for a family of loops $\mathcal{L}$. For example, if $\mathcal{G}$ is a tree, $\text{Mod}_p(\mathcal{L}) = 0$ by property (d) below; if $\mathcal{G}$ is an unweighted complete graph, then $\text{Mod}_p(\mathcal{L}) = \frac{1}{3^p}\binom{n}{2}$.

For a finite network $\mathcal{G}$, the following properties hold (see Refs. [8,14]):

(a) **p monotonicity:** The extremal densities satisfy $0 \leqslant \rho^*(e) \leqslant 1$ for all $e \in E$. Thus, for $1 \leqslant p \leqslant q$, we have $\text{Mod}_q(\mathcal{L}) \leqslant \text{Mod}_p(\mathcal{L})$.

(b) **$\mathcal{L}$ monotonicity:** If $\mathcal{L}' \subset \mathcal{L}$, then $\text{Mod}_p(\mathcal{L}') \leqslant \text{Mod}_p(\mathcal{L})$.

(c) **w monotonicity:** If w and w' are positive link weights with $w \leqslant w'$ then $\text{Mod}_{p,w}(\mathcal{L}) \leqslant \text{Mod}_{p,w'}(\mathcal{L})$.

(d) **Empty family:** If $\mathcal{L} = \emptyset$, then $\text{Mod}_p(\mathcal{L}) = 0$.

(e) **Countable subadditivity:** For any sequence $\{\mathcal{L}_i\}_{i=1}^{\infty}$ of families of loops,

$$\text{Mod}_p\left(\cup_{i=1}^{\infty}\mathcal{L}_i\right) \leqslant \sum_{i=1}^{\infty}\text{Mod}_p(\mathcal{L}_i).$$

The properties above allow quantification of the richness of various family of loops; i.e., a family with many short loops has a larger modulus than a family with fewer and longer loops. In particular, $\mathcal{L}$ monotonocity and subadditivity often define a notion of capacity on the set of loops in a network. For the rest of this paper, we consider $p = 2$ due to its physical and probabilistic interpretations as well as computational advantages; for instance, in this case (3) is a quadratic program.

A. Interpreting loop modulus as a measure of the richness of a family of loops

In order to measure the richness of a family of loops, we want to balance the number of different loops with relatively little overlap vs how many short loops there are in the family.

We demonstrate this in Fig. 1. For the square in Fig. 1(a), the family $\mathcal{L}$ consists of a single loop; hence $\text{Mod}_2(\mathcal{L}) = 0.25$. In Fig. 1(b), the weight of one link is doubled and modulus increases to $\text{Mod}_2(\mathcal{L}) = 0.285$, as it must, by w monotonicity [property (c)]. The network in Fig. 1(c) has more loops than the one in Fig. 1(a) and modulus increases to $\text{Mod}_2(\mathcal{L}) = 0.5$, demonstrating $\mathcal{L}$ monotonicity [property (b)]. Comparing Figs. 1(c) to 1(d), we see that they have the same number of loops, but in Fig. 1(d) they are longer and thus the modulus decreases to $\text{Mod}_2(\mathcal{L}) = 0.455$.

B. Probability interpretation of loop modulus

For $p = 2$ the modulus problem in (3) is

$$\min_{\{\rho|\mathcal{N}\rho \geqslant \mathbf{1}\}} \rho^T\rho. \tag{4}$$

We consider the Lagrangian for (4):

$$L(\rho,\lambda) = \rho^T\rho - \lambda^T(\mathcal{N}^T\rho - \mathbf{1}), \tag{5}$$

where $\lambda \in \mathbb{R}^{\mathcal{L}}_{\geqslant 0}$ is the Lagrange multipliers. It is easy to show that $\rho = \mathbf{1}$ is an interior point for the feasible region of (4), and thus strong duality holds (Slater's condition [16]). Minimizing

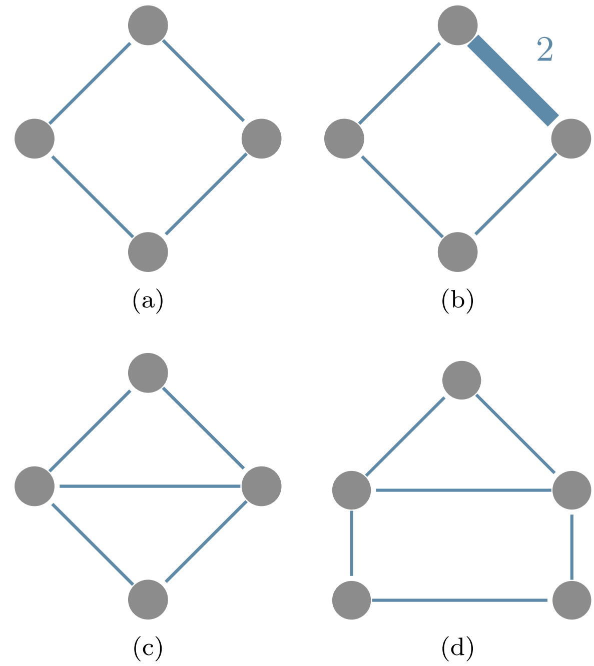


FIG. 1. Loop modulus for some networks demonstrating how modulus can quantify the richness of loops. (a) $\text{Mod}_2(\mathcal{L}) = 0.25$. (b) Weight of a link is doubled, modulus increase by w monotonicity: $\text{Mod}_2(\mathcal{L}) = 0.285$. (c) Increasing number of short loops the modulus increases by $\mathcal{L}$ monotonicity: $\text{Mod}_2(\mathcal{L}) = 0.5$. (d) Loops are longer than in panel (c) and modulus decreases: $\text{Mod}_2(\mathcal{L}) = 0.455$.

L in ρ gives

$$\rho^*(e) = \frac{1}{2}\sum_{\gamma\in\mathcal{L}}\lambda^*(\gamma)\mathbb{1}_{e\in\gamma} \tag{6}$$

and the dual problem

$$\max_{\lambda \geqslant 0}\left(\lambda^T\mathbf{1} - \frac{1}{4}\lambda^T C\lambda\right), \tag{7}$$

where C is the *overlap matrix* for $\mathcal{L}$. Namely,

$$C(\gamma_i,\gamma_j) = \sum_{e\in E}\mathcal{N}(\gamma_i,e)\mathcal{N}(\gamma_j,e) = |\gamma_i \cap \gamma_j|$$

measures the overlap of two loops.

We define a probability mass function (pmf) $\mu \in \mathcal{P}(\mathcal{L}) := \{\mu \in \mathbb{R}^{\mathcal{L}}_{\geqslant 0} : \mu\mathbf{1} = 1\}$ that defines a random loop $\underline{\gamma} \in \mathcal{L}$ with

$$\mu(\gamma) = \Pr(\underline{\gamma} = \gamma). \tag{8}$$

Writing $\lambda = \nu\mu$ for a non-negative scalar ν and a pmf μ (7) becomes

$$\max_{\nu \geqslant 0}\left(\nu - \frac{\nu^2}{4}\min_{\mu\in\mathcal{P}(\mathcal{L})}\mu^T C\mu\right). \tag{9}$$

The maximum in (9) occurs when

$$\nu^* = 2\left(\min_{\mu\in\mathcal{P}(\mathcal{L})}\mu^T C\mu\right)^{-1}. \tag{10}$$

Substituting (10) in (9), we get that $\nu^* = 2\,\text{Mod}_2(\mathcal{L})$ and

$$\text{Mod}_2(\mathcal{L})^{-1} = \min_{\mu\in\mathcal{P}(\mathcal{L})}\mu^T C\mu = \mathbb{E}_{\mu^*}|\underline{\gamma_i} \cap \underline{\gamma_j}|$$

Algorithm 1. Approximating densities for $\text{Mod}_2(\mathcal{L})$ with tolerance $0 < \epsilon_{\text{tol}} < 1$ [8]

1: $\rho \leftarrow 0$; $\rho_0 \leftarrow \mathbf{1}$
2: $\mathcal{L}' \leftarrow \emptyset$
3: $\gamma \leftarrow$ *Shortest Loop*(ρ_0)
4: **while** $\exists\gamma$ such that $\ell_\rho(\gamma) \leqslant 1 - \epsilon_{\text{tol}}$ **do**
5: $\quad\mathcal{L}' \leftarrow \mathcal{L}' \cup \{\gamma\}$
6: $\quad\rho \leftarrow \text{argmin}\{\mathcal{E}_2(\rho) : \mathcal{N}\rho \geqslant \mathbf{1}\}$
7: **end while**

for an optimal μ^*, where $\mathbb{E}_{\mu^*}|\underline{\gamma_i} \cap \underline{\gamma_j}|$ is the minimum expected overlap of two independent, identically distributed random loops with pmf $\mu^* \in \mathcal{P}(\mathcal{L})$.

Moreover by (6), the exremal density satisfies

$$\rho^*(e) = \text{Mod}_2(\mathcal{L})\mathbb{E}_{\mu^*}[\mathcal{N}(\underline{\gamma},e)],$$

where $\mathbb{E}_{\mu^*}[\mathcal{N}(\underline{\gamma},e)] = \sum_{\gamma\in\mathcal{L}} \mathcal{N}(\gamma,e)\mu^*(\gamma)$ is the expected usage of link e in loop $\underline{\gamma}$. Therefore, the optimal measures μ^* are related to the optimal density ρ^* as follows:

$$\frac{\rho^*(e)}{\text{Mod}_2(\mathcal{L})} = \mathbb{P}_{\mu^*}(e \in \underline{\gamma}). \qquad (11)$$

We call $\mathbb{P}_{\mu^*}(e \in \underline{\gamma})$ the *expected usage* of link e.

Moreover, one can always find an optimal measure μ^* that is supported on a minimal set of loops of cardinality bounded above by $|E|$; see Ref. [10, Theorem 3.5]. We think of these loops as *important loops* that play a role in the optimization problems as active constraints.

C. Approximating the modulus

The numerical results in the examples that follow are produced by a PYTHON implementation of the simple algorithm described in Ref. [8]. This algorithm exploits the $\mathcal{L}$ monotonicity [property (b)] of the modulus by building a subset $\mathcal{L}' \subseteq \mathcal{L}$ so that $\text{Mod}_2(\mathcal{L}') \approx \text{Mod}_2(\mathcal{L})$ to a desired accuracy [8, Theorem 9.1]. In short, the algorithm begins with $\mathcal{L}' = \emptyset$, for which the choice $\rho \equiv 0$ is optimal, and a loop is inserted with the shortest hop length, which then repeatedly adds violated constraints to $\mathcal{L}'$ and determines the optimal ρ each time. The algorithm terminates when all constraints are satisfied to a given tolerance (tol) (Algorithm 1).

The two key ingredients for implementing this algorithm are a solver for the convex optimization problem (3) and a method for finding violated loops, i.e., with ρ length less than one. In our implementation, the optimization problem is solved using an active set quadratic program [17] and the violated constraint search is performed using a modified version of the breadth-first search from each node that has a cutoff $1 - \text{tol}$ and reports the first backward link that forms a loop less than the cutoff.

Although simple, this algorithm is adequate for computing the modulus in the examples presented here, on a Linux operating computer with Intel core i7 (and 2.80 GHz base frequency) processor, for example. More advanced parallel primal-dual algorithms are currently under development to treat modulus computations on larger networks.

III. CLUSTERING MEASURE WITH MODULUS OF FAMILY OF LOOPS

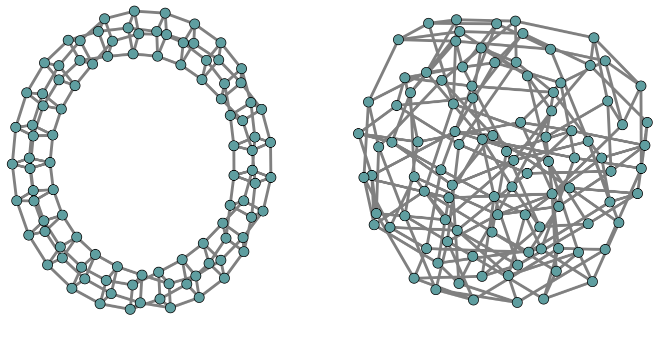
Complex networks exhibit properties such as the small-world phenomenon [18], scale-free degree distribution [19], and local clustering of nodes [18]. In social networks, when two individuals are acquainted it is probable that they have another friend in common, resulting in properties of homophily for the network. For example, in friendship networks people introduce their friends to each other. This transitivity property makes the real world networks different from synthetic random networks [20]. However, this clustering tendency is difficult to quantify.

A proposed measure of clustering for a node v [18] is to compute the fraction of links between neighbors of v that actually are in the network, over all possible ones. The authors in Ref. [21] pointed out the importance of closed paths (loops) in the cluster and discussed computation of the clustering coefficient using the density of loops with length 3 (triangles). Because this measure fails to describe the clustering of gridlike parts of the network, the authors improved the measure by counting quadrilaterals—loops with length 4 or *mutuality* in Ref. [20]—and proposed a new measure that considers different types of quadrilaterals. Similarly, Ref. [5] addresses bipartite networks that lack triangles, and thus the standard clustering coefficient is not useful. In Refs. [5,22,23] the authors emphasize the importance of longer loops in the network. The authors in Ref. [24] showed that clustering coefficient measures are highly correlated with degree, and they proposed a measure that preserves the degree sequence for the maximum possible links among neighbors of node v, thus avoiding correlation biases. Kim *et al.* introduced a *local cycling coefficient* that quantifies local circle topologies by averaging the inverse length of loops passing the nodes [7]. They average this coefficient for all nodes to derive the degree of circulation in the network.

The authors in Ref. [25] introduce a version of clustering coefficient that considers weighted network, and in Ref. [26] propose a way to measure a general clustering coefficient for weighted and directed networks.

Numerous versions of clustering coefficients for different types of networks expose the need for a generalized measure that works for a wide range of applications. We apply the concept of modulus of families of loops as a tool to study structural properties of network clustering. In this section we show that analysis of loops using modulus provides a general approach to the study of network clustering properties. We also propose an alternate clustering measure that can explain situations that conventional methods struggle to handle.

A network has a high clustering measure when most of the links are included in short loops that also visit nearby links. The standard method of counting triangles considers the smallest loops, while other methods consider the next shortest loops, i.e., quadrilaterals. A method must be devised to compare these loops and evaluate the combined influence to improve clustering measures [20]. The previous section introduced a way to



(a) (b)

FIG. 2. (a) A grid network with deg $= 4$ and 100 nodes; (b) a random regular network with deg $= 4$ and 100 nodes. The proposed clustering measures are $C(\mathcal{G}_{\text{grid}}) = 56.25\%$ and $C(\mathcal{G}_{\text{reg}}) = 34\%$. Classical clustering coefficient gives zero for the grid and 2.4% for the regular network, and average square clustering coefficients are 14.7% for the grid and 0.4% for the regular network.

evaluate a family of loops using a modulus. Therefore, we propose a comprehensive modulus-based measure of clustering.

The classical clustering coefficients that measure triangle density are usually normalized by comparing the links in the networks (that form triangles) with all possible links between nodes, i.e., all possible triangles in the corresponding complete graph. Most real networks are far from being complete graphs (even locally), and therefore classical coefficients usually have small values that are correlated to the degree of the node [24].

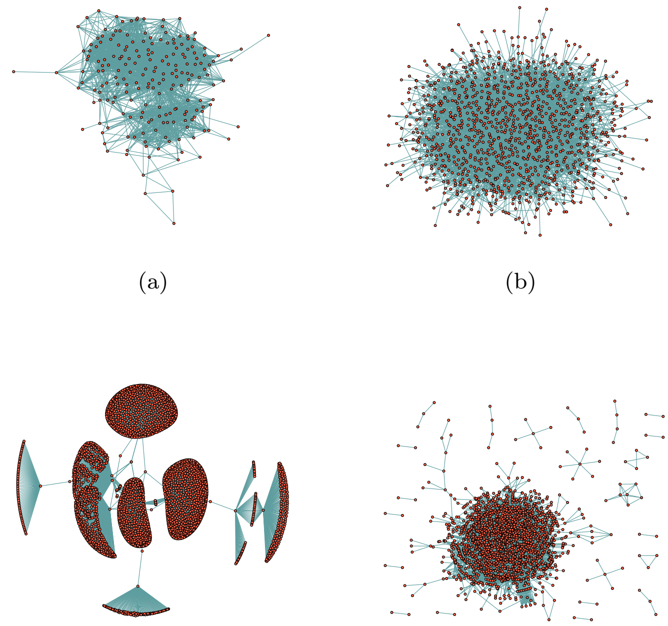
We normalize our clustering measure using the probabilistic interpretation in (11). Modulus tries to spread expected usage as much as possible among the links of the network in order to minimize the expected overlap. However, the expected link usages are not always uniform. Define a uniform density $\rho_u(e) \equiv 1/3$ that is always admissible for loop modulus because it penalizes all loops at least 1. So its energy $\mathcal{E}_2(\rho_u) = |E|/9$ gives an upper bound for $\text{Mod}_2(\mathcal{L})$.

Therefore, our proposed clustering measure takes the following form:

$$C_{\text{loop}}(\mathcal{G}) := \frac{9}{|E|}\,\text{Mod}_2(\mathcal{L}), \tag{12}$$

where C_{loop} is a measure of richness of actual link participation in important loops over the ideal case that all links participate equally in triangles. For example, consider a grid as in Fig. 2(a) with 100 nodes and 200 links. We compare its loop modulus with that of a random regular network with the same number of nodes and same degree as shown in Fig. 2—these networks behave similar to the two extremes of small world networks [18]. Since the classical methods use the number of triangles in a network, they give zero clustering coefficient to the grid and 2–3% to the random regular network. The grid has a square clustering coefficient of 14.7% and the random regular network square clustering is close to zero (we use square clustering introduced in Ref. [5]). For each network in Figs. 2(a) and 2(b),

$$\text{Mod}_2\,\mathcal{L}_{\text{grid}} = 10.8 \quad \text{and} \quad \text{Mod}_2\,\mathcal{L}_{\text{reg}} = 7.8.$$



(c) (d)

FIG. 3. (a) Jazz musicians network [27] with $C_{\text{loop}} = 10.0\%$, average triangle density $C = 52.0\%$, and average square clustering 6.66%. (b) Email communication network in University Rovira i Virgili in Spain with $C_{\text{loop}} = 13.8\%$, average triangle density $C = 16.6\%$, and average square clustering 1.46% [28]. (c) An excerpt of Facebook network with $n = 2888$ and $m = 2981$. Edges represent friendships between nodes [30] with $C_{\text{loop}} = 3.7\%$, average triangle density 0.03%, and average square clustering 0.07%. (d) Friendship network of the website hamsterster.com [31], with $n = 1\,858$ and $m = 12\,534$. The clustering in the network is $C_{\text{loop}} = 6.22\%$. The classical clustering coefficient (transitivity) is 9.04% and average square clustering coefficient is 6.78%.

Therefore, $C_{\text{loop}}(\mathcal{G}_{\text{grid}}) = 54\%$ which means the network is highly clustered and $C_{\text{loop}}(\mathcal{G}_{\text{reg}})$=34% is less clustered than grid.

In some cases, our proposed measure gives different conclusions than the classical cluster coefficients. For example, let us compare the networks (a) and (b) in Fig. 3. Network (a) is a collaboration network between jazz musicians [27] and network (b) is an email communication network at the University Rovira i Virgili in Spain [28]. In the email communication network a very rich core is balanced by many stems on the periphery and the loop clustering measure is slightly higher than for the jazz network. This goes in the opposite direction of the classical clustering coefficient result [29]. For the piece of the Facebook network in Fig. 3(c) [30], the loop clustering value is slightly greater than the classical case, reflecting a certain amount of tightly knit communities. Finally, in the friendship network for the website hamsterster [31], the clustering measure and classical clustering coefficient give similar results.

Furthermore, we can isolate the contribution of triangles, squares, and higher order loops by considering modulus of subfamilies of $\mathcal{L}$. This can be done assuming a hop-length cutoff for γ in Algorithm 1. Moreover, the property of subadditivity [property (e)] gives an upper bound for the aggregate effects.

IV. WEIGHTING TO ENHANCE COMMUNITY DETECTION ALGORITHMS

Communities in networks are defined as groups of nodes that are closely knit together relative to the rest of the network. Real world networks, for example, social networks [32] and biological networks [33], comprise densely connected parts that are loosely connected with each other. Finding these communities is crucial in analyzing the collective behavior of the network or in order to be able to make assumptions (meta population). These communities can be disjoint or overlapping. For a comprehensive review of the literature on this subject, see Ref. [34].

Radicchi *et al.* count the number of short loops that pass each link as a local measure for clustering [35]. To extend the method in Ref. [35] for low clustered networks, Vragovic *et al.* [36] consider general loops (with any length) passing the nodes to detect cluster nodes; however, compared to standard clustering methods, its results are not satisfying [34].

The authors in Ref. [37] define a new weighting for the network to improve modularity maximization methods for finding communities with sizes smaller than the resolution limit [38]. The weighting for a link comes from how many loops with length 3 and 4 it forms with the adjacent links. They show the effectiveness of their method on Lancichinetti, Fortunato, and Radicchi (LFR) benchmark networks. Also the authors in Ref. [39] propose weighting the network with a combination of link centrality [40] and *common neighbor ratio* to enhance community identification. Community detection in directed networks is a challenging problem [41]. Reference [42] improved community detection in directed networks by weighting the network. The authors considered seven different types of triangles and their respective contributions to the community structure.

When a pair of nodes are in the same group it is more likely to have strong flow of communication among each other together with others in their group, and information tends to stay within communities. This emphasizes the importance of having many nonoverlapping short loops.

Analyzing loops in a network provides information about the cluster structure and emphasizes the importance of links in these clusters. By (11) the extremal density $\rho^*(e)$ measures the amount of important loops (see Sec. II B) passing through link e (expected usage). Assuming members in the community shares a lot of cycles between themselves, thus $\rho^*(e)$ serves as a measure of affinity for the nodes connected by e. In other words, nodes on important loops are well connected to the rest of the group. In this section, we show that indeed preprocessing the network using $\rho^*(e)$ can improve network partitioning.

After we compute loop modulus for a network, the extremal density $\rho^*(e)$ gives generic information about the structure of communities that contains many short loops and the importance of links in these clusters that generalize methods in Refs. [35,36]. We can substantially improve the performance of some partitioning methods such as spectral partitioning or modularity maximization heuristics by preprocessing the network into a weighted network with link weights $\rho^*(e)$. We can apply our methods to any weighted and directed network.

As the first example, we consider Zachary's karate club [43], a friendships network at a university karate club with 34 members; see Fig. 4(a). A conflict between the instructor and the club's president split the club into two groups. Finding the communities in this network is a basic benchmark test for partitioning algorithms [44, Chapter 9].

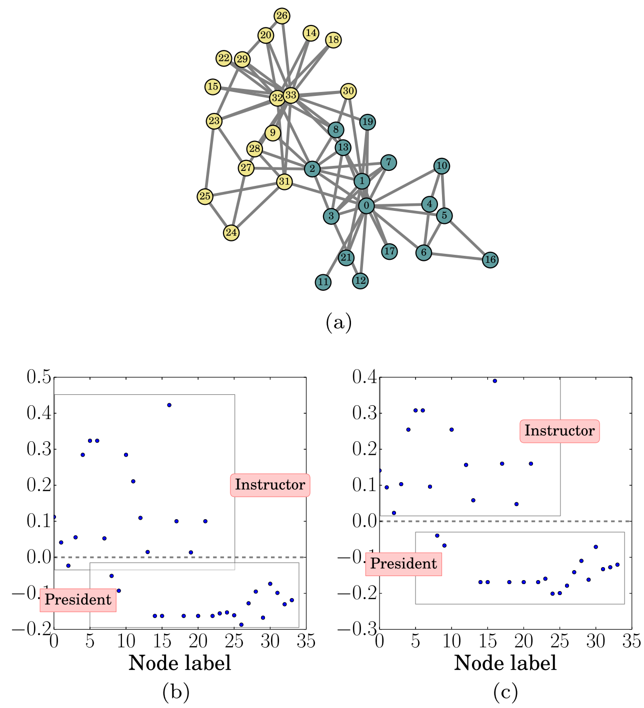


FIG. 4. (a) Zachary's karate club network [43] with the groups split after conflict. (b, c) Fiedler vector values corresponding with the node labels. (b) Spectral partitioning of Zachary's karate club network [43]; node 3 is wrongly partitioned. (c) Spectral partitioning of the same network weighted by loop modulus where nodes are correctly partitioned.

To bisect this network, we use Fiedler vector bisection [45] on both weighted and unweighted networks in Figs. 4(b) and 4(c). In the unweighted case, the bisection method failed to separate a node correctly and there are two nodes that are very close to the other cluster. Our weighting method does this clustering with complete accuracy.

It may be useful to allow for overlapping communities. For instance, a node can be a member of different communities, such as family, sport club, workplace, etc. [46]. Although bisection methods alone are unable to detect overlapping communities, we see that loop modulus can augment these methods by distinguishing nested partitions in networks with overlapping communities in the next example. Figures 5(a)–5(c) show a network that is partitioned by Palla *et al.* [47]. We compute the Fiedler vector in both unweighted and weighted cases. As shown, the unweighted method failed to separate C and D overlapping communities, while the weighted method does distinguish them with the overlapping part.

To show the effectiveness of the weighting method in a more standard fashion, we consider two popular heuristics for modularity maximization: the greedy modularity optimization method by Clauset, Newman, and Moore (CNM) [48] and the Louvian method [49] on the LFR benchmarks [50]. The

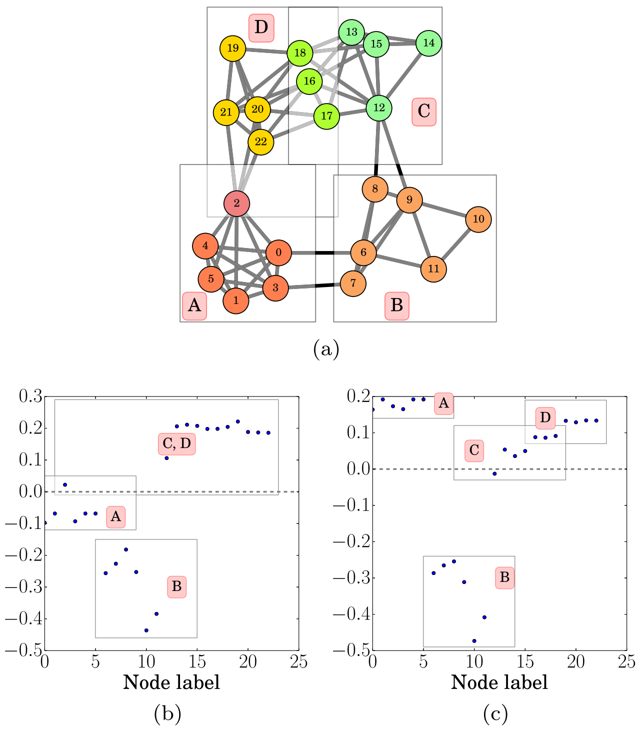


FIG. 5. (a) A network partitioned by Palla *et al.* [47]. Nodes 16, 17, and 18 are shared between C and D groups and node 2 is shared between D and A groups. (b) Fiedler vector of the network. (c) Fiedler vector of the weighted network by loop modulus where overlapping groups can be distinguished.

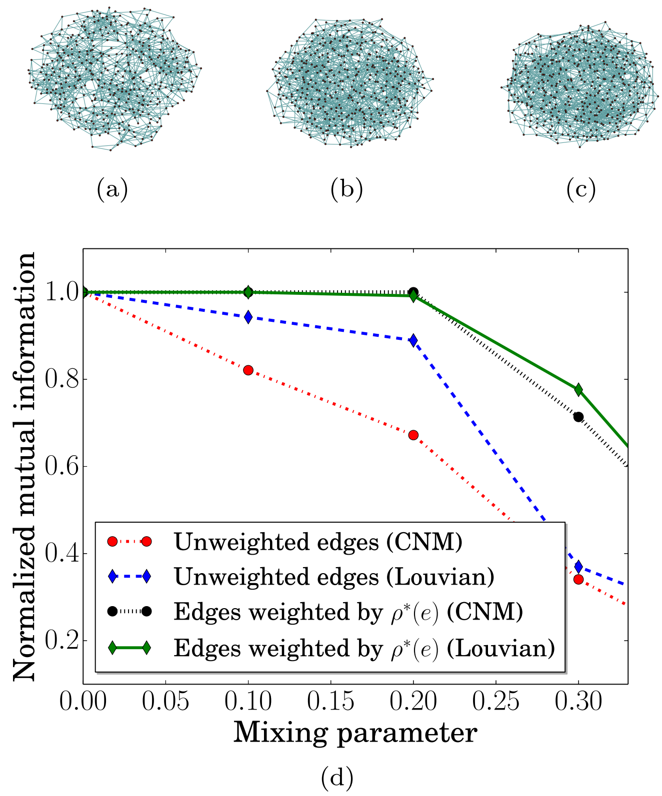


FIG. 6. (a)–(c) Networks are produced by LFR benchmark with size 400 nodes, mean degree 5, maximum degree 10, and community sizes ranging from 20 to 40. The mixing rates μ, for adjusting ratio of intracommunities links over all links, are 0.1, 0.2, and 0.3. (d) The plot depicts the normalized mutual information for community memberships found by greedy modularity optimization (CNM) and the Louvian method. Both the CNW and Louvian methods perform better on reweighted networks.

LFR benchmarks allow the user to specify the community size distribution along with the degree distribution, offering more realistic benchmarks than the Girvan-Newman benchmarks [51]. We show that reweighting the network, using $\rho^*(e)$ from loop modulus, improves both CNM and Louvian substantially.

In Figs. 6(a)–6(c), three networks are produced by the LFR benchmark with 400 nodes, mean degree 5, maximum degree 10, and community sizes ranging from 20 to 40 nodes. The interconnectedness of various communities is measured by the mixing rate μ. We plot the mutual information [52] for both the derived membership from CNM and Louvian on each network and the weighted version and compare them to the ground truth from LFR in Fig. 6. As we observed, both the CNW and Louvian algorithms perform better on reweighted networks using modulus.

V. CONCLUSION

In this paper, we use modulus of family of loops to analyze loop structures in networks. We showed that loop modulus quantifies the richness of loops in the network and we used it to measure clustering. The extremal densities found for loop modulus represent the probability of link participation in important loops. We showed that performance of community detection methods such as spectral bisection and modularity maximization partitioning can be improved by weighting networks with their extremal densities derived from loop modulus. Although, we present some applications of loop modulus, analyzing loop structures on the network can expose different aspects of the network, such as various dynamics on the network, e.g., synchronization and propagation [53–55], as well as analyzing complexity of networks [56].

ACKNOWLEDGMENTS

The authors are thankful to anonymous reviewers for their insightful comments that improved the initial manuscript and provided directions for future work. Thanks also go to Michael Higgins and his research group for their valuable suggestions. This work is funded by the National Science Foundation under Grant No. DMS-1515810.

[1] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[2] S. Mugisha and H.-J. Zhou, Phys. Rev. E **94**, 012305 (2016).
[3] T. Petermann and P. De Los Rios, Phys. Rev. E **69**, 066116 (2004).
[4] M. E. Newman, SIAM Rev. **45**, 167 (2003).

[5] P. G. Lind, M. C. González, and H. J. Herrmann, Phys. Rev. E **72**, 056127 (2005).
[6] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
[7] H.-J. Kim and J. M. Kim, Phys. Rev. E **72**, 036109 (2005).
[8] M. Goering, N. Albin, P. Poggi-Corradini, C. Scoglio, and F. Darabi Sahneh, *2015 49th Asilomar Conference on Signals, Systems and Computers* (IEEE, USA, 2015), pp. 1317–1322.
[9] N. Albin, M. Brunner, R. Perez, P. Poggi-Corradini, and N. Wiens, Conform. Geom. Dyn. Am. Math. Soc. **19**, 298 (2015).
[10] N. Albin and P. Poggi-Corradini, arXiv:1605.08462.
[11] L. V. Ahlfors, *Conformal Invariants: Topics in Geometric Function Theory* (McGraw-Hill, New York, 1973), pp. ix+157.
[12] R. J. Duffin, J. Math. Anal. Appl. **5**, 200 (1962).
[13] O. Schramm, Isr. J. Math. **84**, 97 (1993).
[14] H. Shakeri, P. Poggi-Corradini, C. Scoglio, and N. Albin, J. Comp. Appl. Math. **307**, 307 (2016).
[15] M. Goering, F. D. Sahneh, N. Albin, C. Scoglio, and P. Poggi-Corradini, arXiv:1511.07893.
[16] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[17] D. Goldfarb and A. Idnani, Math. Program. **27**, 1 (1983).
[18] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[19] A. Barabasi and R. Albert, Science **286**, 509 (1999).
[20] M. E. Newman, Social Networks **25**, 83 (2003).
[21] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 183 (2004).
[22] P. G. Lind and H. J. Herrmann, New J. Phys. **9**, 228 (2007).
[23] A. Fronczak, J. A. Hołyst, M. Jedynak, and J. Sienkiewicz, Phys. A (Amsterdam, Neth.) **316**, 688 (2002).
[24] S. N. Soffer and A. Vazquez, Phys. Rev. E **71**, 057101 (2005).
[25] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz, Phys. Rev. E **75**, 027105 (2007).
[26] T. Opsahl and P. Panzarasa, Social Networks **31**, 155 (2009).
[27] P. M. Gleiser and L. Danon, Adv. Complex Syst. **6**, 565 (2003).
[28] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).
[29] J. Kunegis, arXiv:1402.5500.
[30] J. J. McAuley and J. Leskovec, in *NIPS* 548 (2012).
[31] Hamsterster friendships network dataset, KONECT, http://konect.uni-koblenz.de/networks/petster-friendships-hamster.
[32] G. C. Homans, *The Human Group* (Routledge, New York, 2013), Vol. 7.
[33] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[34] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[35] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Nat. Acad. Sci. USA **101**, 2658 (2004).
[36] I. Vragović and E. Louis, Phys. Rev. E **74**, 016105 (2006).
[37] J. W. Berry, B. Hendrickson, R. A. LaViolette, and C. A. Phillips, Phys. Rev. E **83**, 056119 (2011).
[38] S. Fortunato and M. Barthelemy, Proc. Nat. Acad. Sci. USA **104**, 36 (2007).
[39] A. Khadivi, A. Ajdari Rad, and M. Hasler, Phys. Rev. E **83**, 046104 (2011).
[40] L. C. Freeman, Sociometry **40**, 35 (1977).
[41] F. D. Malliaros and M. Vazirgiannis, Phys. Rep. **533**, 95 (2013).
[42] C. Klymko, D. Gleich, and T. G. Kolda, arXiv:1404.5874 (2014).
[43] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).
[44] A.-L. Barabási, Philos. Trans. R. Soc. London A **371**, 20120375 (2013).
[45] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, UK, 2010).
[46] G. Palla, A.-L. Barabási, and T. Vicsek, Nature (London) **446**, 664 (2007).
[47] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) **435**, 814 (2005).
[48] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).
[49] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech. (2008) P10008.
[50] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).
[51] M. Girvan and M. E. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
[52] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech. (2005) P09008.
[53] Z. Li, Z. Duan, G. Chen, and L. Huang, IEEE Trans. Circuits Syst. I **57**, 213 (2010).
[54] Y. Kuramoto, *International Symposium on Mathematical Problems in Theoretical Physics* (Springer, New York, 1975), pp. 420–422.
[55] P. Van Mieghem, Computing **93**, 147 (2011).
[56] C. T. Butts, J. Math. Soc. **24**, 273 (2000).