GENERALIZED NETWORK MEASURES BASED ON MODULUS OF FAMILIES OF WALKS

HEMAN SHAKERI\textsuperscript{1}, PIETRO POGGI-CORRADINI\textsuperscript{2}, CATERINA SCOGLIO\textsuperscript{3}, AND NATHAN ALBIN\textsuperscript{2}

Abstract. The modulus of a family of walks quantifies the richness of the family by favoring many short walks over fewer longer ones. In this paper we investigate various families of walks in order to introduce new measures for quantifying network properties using modulus. The proposed new measures are compared to other known quantities such as current-flow closeness centrality, out-degree centrality, and current-flow betweenness centrality. Our proposed method is based on walks on a network, and therefore will work in great generality. For instance, the networks we consider can be directed, multi-edged, weighted, and even contain disconnected parts. Examples are provided to show the effectiveness of our measures.

1. Introduction

This paper studies applications of modulus of families of walks on networks. This is a discrete analog of the classical theory of modulus of curve families in complex analysis [1]. Although modulus on networks has been studied under several different guises, see for example [12], [10], [25], [16], it is not as well understood as in the continuum setting.

Our study of modulus of walks on networks originated from [11] in which the authors compared it to a new geometric measure they called “epidemic quasimetric”. Thereafter, research in [4] showed that modulus can be considered as a convex optimization problem that can be solved effectively. Continuity and smoothness properties of modulus on networks were considered in [2] and the dual problem was explored further in [3].

In this paper we explore the versatility of modulus of families of walks, demonstrating that it provides a powerful approach to the study of networks. We describe different problems that can be handled by various classes of families of walks. Furthermore, we propose measures based on these families that can be applied in a general framework, handling directed or undirected, weighted or unweighted, and disconnected networks, while the amount of information extracted from a network can be adjusted with high accuracy. We apply the proposed measures to problems such as detecting influential parts of networks, ranking most important nodes in networks, and mitigating the spread of infection in networks. These measures capture the importance of a node on a network by considering only parts of the network most strongly influenced by the node. Thus, the computation of these proposed measures can be done in a decentralized manner, allowing an efficient parallel implementation for large networks.

The paper is organized as follows. First, we introduce our notation and necessary background on modulus of families of walks. Then, we define our proposed measures in detail and compare them to different conventional measures. Finally, we apply these new measures in various examples and applications.

Key words and phrases. Centrality Measure, Modulus, Family of Walks.

The first three authors are supported by NSF grant n. 1201427.
2. Notations, definitions and algorithms

Let $G = (V, E)$ be a network with nodes $V$ and links $E$. Using standard terminology, a walk $\gamma$ on a network is represented as a finite string of alternating nodes and links $v_1e_1v_2e_2v_3\ldots e_rv_{r+1}$, with the property that $v_i$ and $v_{i+1}$ are linked by $e_i$ for $i = 1, 2, \ldots, r$. In either case, we require that $r \geq 1$, so that a walk is considered to traverse at least one link in the network.

We define the $\rho$-length of a walk $\gamma$ as

$$\ell_\rho (\gamma) := \sum_{e \in E} N(\gamma, e) \rho(e)$$

where $\rho : E \rightarrow [0, \infty)$ is a density, interpreted as a penalization or cost that the walk $\gamma$ must pay for traversing link $e$ once, and $N(\gamma, e)$ is the number of times $\gamma$ traverses $e$.

When $p_0(e) \equiv 1$, $\ell_{p_0}$ represents the hop-length of $\gamma$. We define the $\rho$-length of a family of walks $\Gamma$ as

$$\ell_\rho (\Gamma) = \inf_{\gamma \in \Gamma} \ell_\rho (\gamma).$$

Let $p \geq 1$ and let $w : E \rightarrow (0, \infty)$ be a positive weight function representing a generalized conductivity. We define the energy of a density $\rho$ as

$$E_{p,w}(\rho) = \sum_{e \in E} w(e) |\rho(e)|^p.$$

A density $\rho$ is admissible for a family of walks $\Gamma$ if $\ell_\rho (\Gamma) \geq 1$. Let $A(\Gamma)$ be the set of all admissible densities for $\Gamma$. Then $\text{Mod}_{p,w} (\Gamma)$ is defined as

$$\text{Mod}_{p,w} (\Gamma) = \inf_{\rho \in A(\Gamma)} E_{p,w}(\rho) = E_{p,w}(\rho^*) .$$

The existence of an extremal density $\rho^*$ is proven in [2, Lemma 2.1]. To simplify notation, the subscript $w$ will be omitted unless needed.

**Proposition 2.1.** For any finite network $G$, the following properties hold:

(a) **p-Monotonicity:** For any $1 \leq p < \infty$ and any family of walks $\Gamma$, the extremal densities satisfy $0 \leq \rho^*(e) \leq 1$ for all $e \in E$. Thus, for $1 \leq p \leq q$, we have $\text{Mod}_q (\Gamma) \leq \text{Mod}_p (\Gamma)$.

(b) **$\Gamma$-Monotonicity:** For any $1 \leq p < \infty$, if $\Gamma' \subseteq \Gamma$, then $\text{Mod}_p (\Gamma') \leq \text{Mod}_p (\Gamma)$.

(c) **w-Monotonicity:** For any $p \geq 1$ and any family of walks $\Gamma$, if $w$ and $w'$ are positive edge weights with $w \leq w'$ then $\text{Mod}_{p,w} (\Gamma) \leq \text{Mod}_{p,w'} (\Gamma)$.

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

(e) **Countable Subadditivity:** For any sequence $\{\Gamma_i\}_{i=1}^\infty$ of families of walks,

$$\text{Mod}_p \left( \bigcup_{i=1}^\infty \Gamma_i \right) \leq \sum_{i=1}^\infty \text{Mod}_p (\Gamma_i).$$

(f) **Extension Rule:** Given two families of walks, $\Gamma$ and $\Gamma'$, if for all $\gamma \in \Gamma$, there exists $\sigma \in \Gamma'$ such that $\sigma$ is subwalk of $\gamma$ (i.e., $N(\sigma, e) \leq N(\gamma, e)$ for every $e \in E$), then we have $\text{Mod}_p (\Gamma) \leq \text{Mod}_p (\Gamma')$.

(g) **Parallel Rule:** Given two families $\Gamma_1$ and $\Gamma_2$, such that $N(\gamma_1, e)N(\gamma_2, e) = 0$ for every $e \in E$, $\gamma_1 \in \Gamma_1$ and $\gamma_2 \in \Gamma_2$. Then $\text{Mod}_p (\Gamma_1 \cup \Gamma_2) = \text{Mod}_p (\Gamma_1) + \text{Mod}_p (\Gamma_2)$.
Proof. For (a), see [2, Lemma 2.2 and Theorem 5.5]. For (b), (d)–(f), see [4, Proposition 3.4 and Section 5.5].

To prove (c), note that \( w \) does not affect the admissible set \( A(\Gamma) \). Moreover, for any \( \rho \in A(\Gamma) \), \( E_{p,w}(\rho) \leq E_{p,w'}(\rho) \).

For (g), since the statement is slightly different than in [4], we provide a proof. By (e), we know \( \text{Mod}_{p}(\Gamma) \leq \text{Mod}_{p}(\Gamma_{1}) + \text{Mod}_{p}(\Gamma_{2}) \).

Let \( E_{i} \) be the set of links in \( E \) such that \( N(\gamma,e) \neq 0 \) for some \( \gamma \in \Gamma_{i} \). Note that \( E_{1} \cap E_{2} = \emptyset \) by hypothesis. Given \( \rho \in A(\Gamma) \), define \( \rho_{i} = \rho \cdot 1_{E_{i}} \) for \( i = 1, 2 \), where \( 1_{E_{i}} \) is the indicator function for \( E_{i} \). Then \( \rho_{i} \in A(\Gamma_{i}) \) for \( i = 1, 2 \), and \( E_{p}(\rho) \geq E_{p}(\rho_{1}) + E_{p}(\rho_{2}) \). Taking the infimum of both sides results in \( \inf_{\rho \in A(\Gamma)} E_{p}(\rho) \geq \inf_{\rho_{1} \in A(\Gamma_{1})} E_{p}(\rho_{1}) + \inf_{\rho_{2} \in A(\Gamma_{2})} E_{p}(\rho_{2}) \). Therefore, \( \text{Mod}_{p}(\Gamma) \geq \text{Mod}_{p}(\Gamma_{1}) + \text{Mod}_{p}(\Gamma_{2}) \). \( \Box \)

2.1. Connecting families. A special case consists of families \( \Gamma(A,B) \), comprised of all walks that start on \( A \subset V \) and end on \( B \subset V \setminus A \) in the network \( G \). We will often abbreviate \( \text{Mod}_{2}(\Gamma(A,B)) \) simply by writing \( \text{Mod}_{2}(A,B) \) or \( \text{Mod}_{2}(s,t) \) if \( A = \{s\} \) and \( B = \{t\} \).

On undirected networks, \( 2 \)-Modulus of connecting families is the same as effective conductance, as described in [10]. More precisely, if \( u : V \to \mathbb{R} \) is a function, define its gradient to be the density \( \rho_{u} : E \to \mathbb{R}^{+} \) on links \( e = (x,y) \):

\[
\rho_{u}(e) := |u(x) - u(y)|. 
\]

Given two subsets \( A, B \subset V \) and functions \( u|_{A} \leq 0, u|_{B} \geq 1 \), define the capacity \( \text{Cap}(A,B) \) to be

\[
\text{Cap}(A,B) := \inf_{u|_{A} \leq 0, u|_{B} \geq 1} E_{2}(\rho_{u}). 
\]

The capacity \( \text{Cap}(A,B) \) is sometimes referred to as the effective conductance between \( A \) and \( B \). Then, the following holds.

Lemma 2.2. We always have

\[
\text{Cap}(A,B) = \text{Mod}_{2}(A,B). 
\]

Proof. For a proof see [4] and [11]. \( \Box \)

2.2. Modulus of family of walks visiting a set of intermediate nodes. Another interesting family of walks is the via family \( \Gamma_{\text{via}}(A,B;C) \) [4], which represents the family of all walks that start from a set of nodes \( A \), visit another set \( C \subset V \setminus A \), and end on nodes \( B \subset V \setminus (A \cup C) \).

Here, we introduce a modification of this family more well suited to the betweenness centrality measures defined in Section 3.3.2. Let \( A \subset V \) and let \( c \in V \setminus A \). We define the betweenness family \( \Gamma_{b}(A;c) \) as the family of walks originating at some node \( a \in A \), visiting node \( c \) and then terminating at a node in \( A \setminus \{a\} \). In other words,

\[
\Gamma_{b}(A;c) = \bigcup_{a \in A} \Gamma_{\text{via}}(a,A \setminus \{a\};c). 
\]

In the sequel, we will write \( \text{Mod}_{p}(\Gamma_{\text{via}}(A,B;C)) \) as \( \text{via} \text{Mod}_{p}(A,B;C) \) and \( \text{Mod}_{p}(\Gamma_{b}(A;c)) \) as \( \text{BMod}_{p}(A;c) \).
2.3. **Interpreting modulus as a measure of the richness of a family of walks.**

In order to measure the richness of a family of walks, we want to balance the number of different walks with relatively little overlap and how short their lengths are. For example, in a connecting family of walks that connects two sets of nodes, we want to value many short walks. \( \text{Mod}_p(\Gamma) \) provides this measure, and by varying the values of \( p \) more emphasis can be placed on properties such as the number of walks or their length and bottlenecks, see [2]. On undirected networks, when \( p = 2 \) and the family \( \Gamma \) is the connecting family between two nodes, then \( \text{Mod}_2(\Gamma) \) recovers effective conductance, see Lemma 2.2. Therefore, we primarily restrict ourselves to \( p = 2 \) in this paper. However, we will include families \( \Gamma \) that are not connecting and we address networks that can be directed.

![Diagram of modulus values](image)

**Figure 1.** 2-Modulus of the family of connecting walks from node 0 (cyan node) to node 8 (yellow node); all links have weights 1, except link (3, 6) in (C). When the family is enriched, modulus increases. Directions are shown by thicker stubs.

For example, in Figure 1, \( \Gamma_0 \) is the connecting family between node 0 and node 8. Here the networks are directed. Comparing Figure 1A to Figure 1B, we see that every walk from 0 to 8 in the former contains a subwalk in the latter, thus the modulus increases by the extension rule (Proposition 2.1 (f)). In Figure 1C, the weight of link (3, 6) is doubled to 2 and modulus increases as it must by \( w \)-monotonicity (Proposition 2.1 (c)). Figure 1D differs...
from Figure 1B in that the number of walks is higher than before, and modulus increases, demonstrating \(\Gamma\)-monotonicity (Proposition 2.1 (b)). The comparison between Figures 1C and 1D is more subtle; the relationship between the moduli is nontrivial since none of the monotonicity properties apply.

2.4. **Computing the modulus.** The numerical results in the examples that follow were produced by a Python implementation of the simple algorithm described in [4]. This algorithm exploits the \(\Gamma\)-Monotonicity property of the modulus (Proposition 2.1b) by building a subset \(\Gamma' \subseteq \Gamma\) so that \(\text{Mod}(\Gamma') \approx \text{Mod}(\Gamma)\) to a desired accuracy. In short, the algorithm begins with \(\Gamma' = \emptyset\), for which the choice \(\rho \equiv 0\) is optimal, and repeatedly adds violated constraints to \(\Gamma'\), recomputing the optimal \(\rho\) each time. The algorithm terminates when all constraints are satisfied to a given tolerance.

The two key ingredients for implementing this algorithm are a solver for the convex optimization problem (2.4) and a method for finding violated constraints (i.e., for finding paths with \(\rho\)-length less than one). In our implementation, the optimization problem is solved using the cvxopt package [8] and the violated constraint search is performed using a modified version of the implementation of Dijkstra’s algorithm included in Networkx [15].

Although simple, this algorithm is adequate for computing the modulus in the examples presented here, requiring on the order of seconds to perform the computation in Figure 7 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 graphs.

3. **Network measures based on various families of walks**

3.1. **Centrality.** There are many centrality measures for evaluating the importance of nodes in networks. The simplest measure is degree, which considers immediate neighbors. Degree ignores the rest of the network and hence cannot be a proper measure for evaluating the importance of a node in the whole network. Another centrality measure is closeness centrality, which measures the closeness of a node in a network by computing the reciprocal of the sum of shortest-path distances from the given node to all other nodes [22]:

\[
C_c(v) = \frac{1}{\sum_{u \in V} d(v, u)}
\]

where \(d(v, u)\) is the distance between node \(v\) and \(u\).

As described in [18], any reasonable centrality measure should increase when more links are added to the network. For connecting families of walks, modulus does have this behavior because of Proposition 2.1 (b). The main purpose of this paper is to introduce new centrality measures based on modulus of families of walks. For instance, a simple first attempt is to define

\[
C(v) := \sum_{i \in V} \text{Mod}_2(v, i)
\]

where \(\text{Mod}_2(v, i)\) is the 2-Modulus of the family of all walks from node \(v\) to node \(i\).

Note that modulus of connecting families is a measure of proximity rather than a distance, so this is roughly analogous to the classical measure (3.1).
However, in order to obtain this centrality measure for all nodes, we need to compute 2-Modulus, $n^2$ times. We propose a more efficient measure based on modulus below in Section 3.3.

Remark 3.1. Note that in a directed network we could also define $C_{in}(v) = \sum_{i \in V} \text{Mod}_2(i, v)$, which measures the richness of walks that are reaching $v$, and $C_{out}(v) = \sum_{i \in V} \text{Mod}_2(v, i)$, which computes the influence of node $v$ over the network. In this paper we will mostly be concerned with the latter measure.

3.2. Betweenness centrality. Betweenness centrality evaluates the prominence of $v$ in the transmission of information, disease, signals, etc., between pairs of nodes.

A popular betweenness centrality for a node $v$ is defined by the fraction of shortest paths between pairs of nodes that pass through the node $v$:

\[
C_B(v) = \sum_{s \neq t \neq v \neq s} \frac{\sigma_{st}(v)}{\sigma_{st}}
\]

where $\sigma_{st}$ is the number of shortest paths between node $s$ and node $t$ and $\sigma_{st}(v)$ is the size of the subset of such shortest paths that visit $v$ [13].

An analogous measure of betweenness centrality could be defined using via Mod, which computes the richness of walks between $s$ and $t$ that pass through node $v$ (see Section 2.2):

\[
BC(v) := \sum_{s \neq t \neq v \neq s} \frac{\text{via Mod}_2(s, t; v)}{\text{Mod}_2(s, t)}
\]

A naive implementation of this formula requires $|V|^3 + |V|^2$ modulus computations. Although not much is currently known about the computational complexity of modulus of general families of walks, in our experience computing the modulus of connecting and via families is reasonably fast due to the applicability of Dijkstra’s algorithm in these cases (see [4]) and the fact that these families contain very few important walks (see [3]). Nevertheless, for efficiency it is desirable to reduce the number of modulus computations and, therefore, we propose the following more efficient measures.

3.3. Efficient measures for centrality based on modulus. The centrality measures introduced above consider all pairs and triples of nodes, which can be infeasible for large networks. In applications the entire scope of network data cannot always be obtained, and even if it is acquired, such an extensive volume of data is sometimes unnecessary. Each node $v$ will influence some nodes more than others. Therefore, we restrict our analysis of centrality for $v$ to a portion of network around $v$. The following section describes a technique used for this purpose.

3.3.1. Shell-Centrality. For a node $v \in V$, $S(v, k)$ is the set containing nodes $y$ such that \{\(y \in V : d(y, v) = k\}\}, namely all the nodes with discovery time $k$. We call this the $k$-th shell around $v$. If the context is clear, we simply write $S_k$.

We are interested in the 2-Modulus of all walks from node $v$ to the shell $S_k$, which according to Lemma 2.2, $\text{Mod}_2(v, S_k)$ is a measure of conductance between node $v$ and the set $S_k$. Note that $\text{Mod}_2(v, S_1)$ is equal to the out-degree of $v$, corresponding to the influence of $v$ on its immediate neighbors. (The extremal density $\rho^*$ in this case gives value 1 to every out-link from $v$, and value 0 to every other link.)
The modulus from \( v \) to the shell \( S_k \) is a measure of the importance of node \( v \) out to radius \( k \) in the network. When a node has a persisting effect on larger and larger radii, it will have an overall greater closeness centrality (Figure 2).

![Figure 2](image)

**Figure 2.** (A) A directed network; (B) Plot of \( \text{Mod}_2(v, S_i) \) with various shells \( i \), given three different nodes for the network in (A). The Red node maintains its influence over the network and reaches farther, while the Blue node can influence only its first two shells. Out-degree of Blue is more than Black and it is more influential on shells 1 and 2 but Black node has influence over a larger part of the network, therefore Black node has higher centrality than Blue and Red is the more central than both Blue and Black. Link directions are shown by thicker stubs.

Our proposed measure for the centrality of a node \( v \), which we call *shell centrality*, is

\[
C_{\text{shell}}(v) := \sum_{i=1}^{d(v)} \text{Mod}_2(v, S_i)
\]

where \( d(v) \in \{1, 2, \ldots, \epsilon(v)\} \) is a cutoff to be determined later, that can vary from node to node. The largest \( d(v) \) is called the eccentricity of \( v \), i.e., \( \epsilon(v) \), is the maximum hop-length between \( v \) and any other node in \( G \).

**Example 3.2.** In order to compare the proposed measure with the exact expression of \( C \) in (3.2), we compute \( C \) and \( C_{\text{shell}} \) for a weighted and directed network found in [23] (Figure 3A). Each node in this network represents a rhesus monkey, and links between nodes represent observed grooming behavior. The direction of each link indicates the act of grooming. The correlation between \( C \) and \( C_{\text{shell}} \) is 98\%, thereby demonstrating that the measure obtained almost the same results as (3.2) with less computation costs. In Figure 3B, the centrality value of nodes is normalized by their maximum value, therefore for a node \( v \), \( 0 \leq \hat{C}(v), \hat{C}_h(v) \leq 1 \).
Figure 3. Correlation between closeness centralities measured by (3.2) and (3.5) for the Rhesus Network in [23]. Link directions are shown with thick link heads.

Example 3.3. In this example, we compute $\hat{C}_h$ with (3.5) for four weighted and directed networks, each differing from the previous one by one link. In Figure 4, the size of each node is scaled by its centrality, as computed by (3.5). The centrality value can be observed inside the nodes. Changes of nodes centralities provide interesting, and in some cases significant, assessments of a node’s influence on the entire network. In Figure 4A, the yellow node is the most central node, thereby influencing the entire network more than any other node. For the network in Figure 4B, the direction of the link from the yellow node to the magenta node has been changed and the centrality updated, with the result that the centrality of the magenta node is now the highest. As shown in Figure 4B, the white and magenta nodes have identical (weighted) out-degree centrality, but the white node cannot influence the network on its left side. If a link is added from white to magenta (Figure 4C), then the white node becomes the most central node. The addition of another link with a new node at the tail of the network, as shown in Figure 4D, only changes centrality of the cyan node, while centrality of the other nodes stays almost constant. With out-degree centrality the cyan node would be the most central node.

Example 3.4. As mentioned, a network can be huge and acquisition or consideration of all network data is not always possible. Therefore, there is a trade-off between the amount of information extracted and computational costs. For example, in Figure 5A, we consider a random directed geometric network and plot the correlation between the centrality $C_{shell}$ computed with $d(v) = \epsilon(v)$ in (3.5), and the same centrality computed with different cutoffs $d(v) = r$ for various radii $r$ of shells. By increasing these cutoff radii, we obtain increasingly better correlation, but after having reached $r = 6$ it seems that increasing the cutoff becomes unnecessary (Figure 5B). This reflects the fact that, after 6 hops from each node in this network, the importance of the node starts to decay rapidly throughout the entire network (see Figure 2B) and hence considering the first 6 shells is enough.
Remark 3.5. For a general network, it is difficult to predict the proper cutoff for a given node. In practice, we introduce a tolerance that is used to stop whenever \( \text{Mod}_2(v, S_k) \) is less than a given value.

3.3.2. Betweenness centrality measure. Consider the visiting family of walks described in Section 2.2, in which walks begin from a node \( a_i \in A \), visit node \( v \) and return to a different node \( a_j \in A \) where \( i \neq j \), we called this family \( \Gamma_{\text{via}}(A, A, v) \). We choose \( A \) to be a proportion of the most central nodes using our centrality \( C_{\text{shell}} \) from (3.5). Then, we set

\[
BC_{\text{shell}}(v) = \text{via} \text{ Mod}_2(A, A, v)
\]

Note that only one modulus computation is involved in (3.6) for each node. The number of nodes considered in \( A \) vary with the type of network, but \( BC_{\text{shell}} \) generally provides good results even when considering a handful of nodes. Also, if \( A \) happens to include all the neighbors of \( v \), then \( BC_{\text{shell}} \) simply gives the degree of \( v \).

Example 3.6. By Lemma 2.2, for an undirected network the 2-Modulus of a family of connecting walks between two nodes is equal to effective conductance. Thus, one might expect the betweenness measure (3.6) to be related to the well-known current-flow betweenness centrality (CFBC). However, the modulus-based centrality measure is more general in that it is not restricted to connected, undirected networks. In this example, we provide evidence that \( BC_{\text{shell}} \) and CFBC are linearly correlated by considering both a random geometric network and a random scale-free network.

We calculate (3.6) for a geometric network, as shown in Figure 6A (see [21]) and a scale-free network, as shown in Figure 6B (see [5]). In Figures 6C and 6D we plot the correlation...
between well-known current-flow betweenness centrality, as computed in [6], and our measure of betweenness centrality $BC_{shell}$ when the number of most central nodes chosen for the set $A$ varies.

For the geometric network in Figure 6A, there are only a few important nodes and they are scattered. So when the size of $A$ increases past a certain level, $A$ starts to include nodes on the periphery of the network and as shown in Figure 6C the correlation starts to decrease slightly.

In the scale-free network in Figure 6B, central nodes are more accurately identifiable and they are more concentrated. So including more nodes in $A$ leads to better and better correlation. However, a relatively small set $A$ (here about 10% of the nodes in the network) can already give high correlation.

4. FURTHER RESULTS AND APPLICATIONS

In order to evaluate the effectiveness of our proposed measures (3.5) and (3.6), we compare them to various conventional network measures. First, we consider undirected, unweighted, and connected (simple) networks, and compare our measure $C_{shell}$ to a well-known measure, current-flow closeness centrality, demonstrating that they lead to similar results. Then, we consider general networks where current-flow centrality cannot be applied, and thus illustrate the advantages of using 2-Modulus centrality measures. Finally, we give two applications.

4.1. Undirected networks. For connected undirected networks there are numerous centrality methods (see [18]). In particular, the symmetry of the Laplacian matrix allows one to use measures such as current-flow closeness centrality (CFC) [6].
In order to evaluate the performance of $C_{shell}$ in (3.5), we compare it to (CFC) in a simple geometric network with 60 nodes as shown in Figures 7A and 7B. These figures illustrate the measured proposed centrality and its correlation with current flow centrality. The correlation of these measures is 0.97, implying very similar rankings. This means that our measure is at least as good as (CFC) for simple networks.
Figure 7. Closeness centrality measured with $C_{shell}$ and correlation with current flow closeness centrality [6] for two networks (A) and (B) with node size. In both cases, (C) and (D), the correlation is 0.97.

4.2. Directed networks. There are fewer closeness centrality measures for directed networks compared to undirected networks, and most of the measures focus on local information of nodes. There are other measures that can be applied for directed networks, such as Pagerank and Katz centrality (for a good review, see [19]), but because they have some shortcomings in directed networks when there is a lack of mixing, we chose to compare our measure to out-degree centrality.

In Figures 8A and 8B we compare 2-Modulus centrality and out-degree centrality for two random directed networks, showing these centralities using the size of the nodes.

As depicted in these figures, out-degree centrality emphasizes the local importance of nodes, while 2-Modulus closeness centrality takes a broader perspective of the network. Consequently, nodes of the network that cannot reach most of the network have less importance in 2-Modulus centrality; however, in out-degree centrality nodes can have high centrality if
they have high out-degree, as shown in Figure 8C in which nodes that have high out-degree centrality and small 2-Modulus centrality corresponded to nodes of the network that did not significantly influence the entire network.

4.3. Ranking of most influential nodes. Comparing nodes with a low number of heavily weighted links to nodes with a high number of more lightly weighted links, is usually a challenge for measures such as out-degree centrality [20]. However, 2-Modulus centrality does not have this problem, because it is not a local property. For instance, a heavily weighted link might lead to a smaller portion of the network.

Here we consider the network in [14] that consists of relationships between a group of 32 scientists. In this network, directed links are weighted by the number of sent messages between each pair of researchers.
Opsahl et al. ranked the nodes in this network with a centrality measure that upgraded out-degree centrality that can be tuned between the number of outward links and the sum of out-weights [20]. Since this centrality measure considers only links to the nearest neighbors, it ignores most of the network structure. For example, ranking errors occur when a strong link is directed to a dead end in the network (or to an unimportant part of the network) but a weaker link is directed to important parts of the network.

In Table 1, we propose a new ranking performed based on 2-Modulus centrality with no concern for balancing between the number of paths and their weight strengths and with an eye on the node position in the network.

4.4. Suppressing epidemics. Detection of the most influential nodes is critical in some applications. Vaccination is commonly used to mitigate the spread of an infectious disease. However, it is not always possible to vaccinate the entire population. Therefore, determining the best sub-population to vaccinate is a challenge due to network complexity ([7] and [26]). In this section, we show that the proposed measure can be efficiently used to vaccinate a fraction of highly central nodes, especially for directed networks with mesoscopic structure. A majority of real networks are formed by connecting clusters of sub-networks, such as communities.

Each community contains its own structure and connects to others with a different structure. However, local measures, such as degree centrality, cannot capture these higher order structures.

In this study, we considered a random directed network consisting of clusters with internal Poisson degree distribution, which are connected to each other by another Poisson distribution [24]. In order to consider a directed version of these networks, we chose a direction for each link at random. Figure 9A shows a network generated in this way with 200 nodes and 8 modules. We compute 2-Modulus centrality with a cutoff of 4 and out-degree centrality. As presented in Figure 9B, nodes with identical out-degree centrality can have a different role in the network based on 2-Modulus centrality.

We consider an SIR (susceptible-infected-recovered) epidemic process on this network with infection rate $\beta = 0.5(\text{day})^{-1}$ and recovery rate $\delta = .2(\text{day})^{-1}$, starting with two initial infections. In the SIR model, each node can be either susceptible (S), infected (I), or recovered (R) (immune) [19, Chapter 17]. A susceptible node can become infected if it is neighboring an infected node. The infection process of a node $i$, with one infected neighbor is a Poisson process with transition rate $\beta$. Infection processes are considered to be stochastically independent of each other. In addition to the infection process, a recovery process also occurs. An infected node recovers and becomes immune with a exponentially distribute recovery rate $\delta$. The main characteristics of the model and a node transition graph are shown in Figure 9C. $Y_i$ is the number of infected neighbors of node $i$.

After vaccinating the first 50 nodes with the highest centrality in both measures, we ran several simulated epidemics (hundreds) with the parameters specified above and computed the average fraction of susceptible (not yet infected) individuals for each day of the outbreak [9]. As shown in Figure 9D, 2-Modulus centrality pinpointed the most effective nodes better and allowed a more successful mitigation of the outbreak.

5. Conclusion

In this paper we introduced general centrality measures based on modulus of families of walks. These measures provide information about nodes using knowledge from the entire
Figure 9. (A) Random modular network, (B) Out-degree centrality and 2-Modulus centrality measured for each node in network (A), (C) Node transition graph for SIR model, (D) Comparison of vaccination strategies based on 2-Modulus centrality and out-degree centrality for an SIR epidemic in network (A). The fraction of susceptible population ($S$) at the end of the outbreak for 2-Modulus centrality vaccinated people is larger than the fraction obtained by vaccinating using out-degree centrality.
network, while keeping computational costs low and without requiring acquisition of data from the entire network. These methods can be applied to very general networks, whether weighted, directed, multi-edged, or disconnected. We also presented several applications of our proposed measure to identify influential parts of a network and node ranking, as well as for mitigating epidemics. Considering different families of walks and their modulus can provide additional insights into solving other problems on networks. For example, currently, we are studying the family of all loops and using its modulus to get a better understanding of the structure of communities in general networks.

References


1 Department of Electrical and Computer Engineering, Rathbone Hall, Kansas State University, Manhattan, KS 66506, USA
   E-mail address: heman@ksu.edu
   E-mail address: caterina@ksu.edu

2 Department of Mathematics, Cardwell Hall, Kansas State University, Manhattan, KS 66506, USA
   E-mail address: albin@math.ksu.edu
   E-mail address: pietro@math.ksu.edu