CENTRALITY MEASURE BASED ON CLUSTERINGS IN THE EFFECTIVE RESISTANCE EMBEDDING SPACE

KEVIN BUI[*], JOE FENG[*], ALTO SENDA[*], JUSTIN WANG[*]

February 28, 2015

Abstract. Centrality measures are useful tools in network analysis since they quantify a node's importance in a graph modeling a network. The definition of importance varies based on context, so choosing an appropriate centrality measure is essential. For larger networks in particular, we need to be able to find the most important nodes among hundreds or thousands of them. We propose a new centrality measure called k-means centrality based on applying k-means to the effective resistance embedding. The effective resistance embedding is derived from using Spielman-Srivastava's algorithm [3] in order to compute the effective resistance of a graph when representing it as an electrical network. k-means centrality measures the connectedness of one node to all other clusters on the graph computed by k-means. This measure has potential use in social network analysis since it determines the number of communities a node is a part of within a graph. We compare the results of our new measure to those of the already established betweenness and closeness centrality measures and their current-flow variants. Numerical results show that k-means centrality produces results similar to other centrality measures. However, they show that it is not the most discriminative measure, but it has the highest correlation with the degrees of the nodes.

Key words. network-means centrality, electrical network, k-means, spectral graph theory

1. Introduction. The concept of centrality is investigated in order to answer the question, "What is the most important nodes in a network?" Centrality is an important measure used in social, biological, communication, and transportation networks since it helps analyze the relative structural prominence of nodes in the network. Especially in social network analysis it can measure the most influential or the most connected person depending on the context of the network. Two most frequently utilized centrality measures in this area are betweenness centrality and closeness centrality since they both account for the information that travels on the shortest paths of the network. However, since both centrality measures ignore information spread along non-shortest paths, current-flow betweenness centrality and current-flow closeness centrality were proposed by Brandes and Fleischer [1] to account for that. These variants are based on representing a graph as an electrical network where electrical current is propagated throughout the whole network.

In this paper, we review betweenness, closeness, and their current-flow variants and we also propose a new centrality measure called k-means centrality which is based on applying the k-means algorithm to the effective resistance embedding. The effective resistance embedding is a higher dimensional embedding of the nodes derived from computing the effective resistance of a graph representing an electrical network. Finally, we present numerical results comparing the proposed centrality measure to the aforementioned measures.

2. Centrality Measures. Throughout the paper, we consider only the graphs $G = (V, E)$ that are simple, undirected, and connected and that have $n \geq 3$ nodes.

2.1. Centrality Measures Based on Shortest Paths. Based on information travelling on shortest paths of a network, betweenness centrality and closenness centrality are the most popular yet most basic measures used in network analysis.

Betweenness centrality [2, 1] $c_B : V \to \mathbb{R}_{\geq 0}$ measures the extent to which a node lies on the shortest paths between other nodes. It is defined as

$$c_B(v) = \frac{1}{n_B} \sum_{s,t \in V} \frac{\sigma_{st}(v)}{\sigma_{s,t}}, \tag{2.1}$$

where $\sigma_{st}(v)$ is the number of shortest paths from s to t containing v as an inner vertex, $\sigma_{s,t}$ is number of shortest paths from s to t, and $n_B = (n-1)(n-2)$ is a normalizing constant.

On the other hand, closeness centrality $c_C : V \to \mathbb{R}$ [2, 1] measures the mean distance from a node to other nodes. It is defined as

$$c_C(v) = \frac{n_C}{\sum_{t \neq v} d_G(v,t)} \tag{2.2}$$

where $n_C = n - 1$ is a normalizing constant and $d_G(v,t)$ denotes the length of the shortest path between v and t.

[1]Department of Mathematics, UCLA, 520 Portola Plaza, Mathematical Sciences Building 6363, Los Angeles, CA 90095-1555, email: xyz@ucla.edu

[2]Second Department, UCLA, 520 Portola Plaza, Mathematical Sciences Building 6363, Los Angeles, CA 90095-1555

Measures based on shortest paths do not account for the spread of information along non-shortest paths. Consequently, they are a poor metric for systems in which information flow behaves this way. Resolving this issue, current-flow betweenness centrality and current-flow closeness centrality account for the spread of information along non-shortest paths. Before we review their definitions, we will discuss how a graph can be represented as an electrical network.

2.2. Graph as an Electrical Network [1]. A graph can be represented as an electrical network $N = (G; c)$ with positive edge weights $c : E \to \mathbb{R}_{>0}$ indicating the conductance of an edge. Since we are dealing with undirected graphs in this paper, we have $c(e) = 1$.

To represent how current flows throughout an electrical network, a vector $b : V \to \mathbb{R}$ called supply defines where current externally enters and leaves it. A node $v \in V$ is a source if $b(v) > 0$ and is a sink if $b(V) < 0$. By conservation, $\sum_{v \in V} b(v) = 0$ is required. For simplicity, we only consider the unit current entering the network with a single source and a single sink, so

$$b_{st}(v) = \begin{cases} 1, & \text{if } v = s \\ -1, & \text{if } v = t \\ 0 & \text{otherwise} \end{cases}$$

To account for the direction of the current, each edge is given an arbitrary orientation. Furthermore, we denote $\vec{e}$ as the directed edge corresponding to the orientation of $e \in E$ and $\vec{E}$ as the set of all oriented edges. Now we will define $x : \vec{E} \to \mathbb{R}$ as the current of the graph and introduce a lemma relating to it.

DEFINITION 1. *Let $N = (G; c)$ be an electrical network with supply b. A vector $x : \vec{e} \to \mathbb{R}$ is called a* current *if it satisfies the following:*

1. *Kirchoff's Current Law*

$$\sum_{(v,w) \in \vec{E}} x(v, w) - \sum_{(u,v) \in \vec{E}} x(u, v) = b(v) \quad \text{for all} \quad v \in V \tag{2.3}$$

2. *Kirchoff's Potential Law*

$$\sum_{i=1}^{k} x(\vec{e_i}) = 0 \quad \text{for every cycle} \quad e_1, ..., e_k \quad \text{in} \quad G. \tag{2.4}$$

LEMMA 1. *For an electrical network $N = (G; c)$ and any supply b, there is a unique current $x : \vec{E} \to \mathbb{R}$.*

A value $x(\vec{e}) > 0$ indicates that the current is flowing in the direction of $\vec{e}$; $x(\vec{e}) < 0$ indicates that the current is flowing against the direction of $\vec{e}$. For an st-supply, we denote x_{st} as the st-current. Furthermore, current is related to potential difference $\hat{p} : \vec{E} \to \mathbb{R}$ such that $\hat{p}(\vec{e}) = x(\vec{e})/c(e)$ for all $e \in E$. A vector $p : V \to \mathbb{R}$ assigns absolute potentials if $\hat{p}(v, w) = p(v) - p(w)$ for all $(v, w) \in \vec{E}$. We use $\hat{p}_{st}$ and p_{st} to indicate that the potential differences and the absolute potentials respectively are based on the st-supply.

LEMMA 2. *Let $N = (G; c)$ be an electrical network with supply b. For any fixed vertex $v_1 \in V$ and constant $p_1 \in \mathbb{R}$, there are unique absolute potentials $p : V \to \mathbb{R}$ with $p(v_1) = p_1$.*

LEMMA 3. *The absolute potentials of an electrical network $N = (G; c)$ with supply b are exactly the solutions of $Lp = b$ where L is the Laplacian matrix defined as follows:*

$$L(v, w) = \begin{cases} \sum_{e: v \in e} c(e), & \text{if } v = w \\ -c(e), & \text{if } e = v, w \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

for all $v, w \in V$.

Lemma 3 shows that computing the absolute potentials of an electrical network is possible, but since L is singular, there are multiple assignments of absolute potentials for a given supply b. However, by Lemma 2, we can set $p(v_1) = 0$ to solve for only one assignment given the fixed vertex ordering $v_1, v_2, ..., v_n$. To do so, we omit the

row and column corresponding to v_1 in L to obtain $\widetilde{L} \in \mathbb{R}^{n-1\times n-1}$. Since $\widetilde{L}$ is positive definite and regular, we obtain

$$p = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \widetilde{L}^{-1} \end{pmatrix} b \tag{2.6}$$

With all the necessary theory established, we now review the current-flow variants of betweenness and closeness.

2.3. Current-Flow Betweenness Centrality. The amount of st-current flowing through a node is the analog of the number of shortest paths passing through it for betweenness centrality. Given a supply b and the st-current, we define throughput of a vertex $v \in V$ as

$$\tau_{st}(v) = \frac{1}{2}\left(-|b(v)| + \sum_{e:v\in e} |x(\vec{e})| \right) \tag{2.7}$$

to represent the amount of current flowing through v. The throughput is well-defined because of Lemma 1, where x is unique for a given supply b.

Current-flow betweenness centrality $c_{CB}(v) : V \to \mathbb{R}_{\geq 0}$ is defined as

$$c_{CB}(v) = \frac{1}{n_B} \sum_{s,t\in V} \tau_{st}(v) \quad \text{for all} \quad v \in V \tag{2.8}$$

where $n_B = (n-1)(n-2)$[1]. Because calculating throughput for current-flow betweenness is computationally difficult, we use a randomized approximation algorithm summarized below to approximate the centrality measure.

Algorithm 1 Randomized approximation scheme for current-flow betweenness [1]

Require: electrical network $N = (G; c)$, threshold $\epsilon > 0$, constant l
  return current-flow betweenness approximation $c'_{CB} : V \to \mathbb{R}_{\geq 0}$
  $c'_{CB} \leftarrow \mathbf{0}$ and $k \leftarrow l \cdot \lceil (c^*/\epsilon)^2 \log n \rceil$ where $c^* = n(n-1)/n_B$.
  for $i = 1, ..., k$ **do**
    select $s \neq t \in V$ uniformly at random and solve $Lp = b_{st}$ according to Equation 2.6
    for $v \in V \setminus \{s, t\}$ **do**
      for $e \in \{v, w\} \in E$ **do**
        increase $c'_{CB}(v) byc(e) \cdot |\widetilde{p}(v) - \widetilde{p}(w)| \cdot c^*/2k$
      end for
    end for
  end for

2.4. Current-Flow Closeness Centrality. Current-flow closeness centrality $c_{CC} : V \to \mathbb{R}_{>0}$ is defined by

$$c_{CC}(s) = \frac{n_C}{\sum_{t\neq s} p_{st}(s) - p_{st}(t)} \quad \text{for all} \quad s \in V \quad [1]. \tag{2.9}$$

Current-flow centrality is well-defined by Lemma 2 since any two absolute potentials differ only by an additive constant. The term $p_{st}(s) - p_{st}(t)$ is an alternative measure of distance between s and t given the unit st-current, making it a current-flow variant of closeness centrality.

To compute the effective resistance between any two nodes, we use Spielman-Srivastava's algorithm summarized below. Since the solution X is a $n \times O(\log n)$ matrix, each row can be interpreted as an embedding of the corresponding vertex to the $O(\log n)$-dimensional Euclidean space. We call X the effective resistance embedding. Because the effective resistance embedding is higher dimensional, we may apply the k-means algorithm to obtain graph clusters. This motivates us to propose a new centrality measure.

2.5. k-Means Centrality. By applying the k-means algorithm to the effective resistance embedding, we obtain a partitioning of the nodes into k sets: $\{S_1, S_2, ..., S_k\}$. Based on the results, we define a new centrality measure called k-means centrality. We define k-means centrality $c_K(s) : V \to \mathbb{R}_{\geq 0}$ as

$$c_K(s) = \sum_{k=2}^{L} \sum_{t\neq s} \mathbf{1}_{\{\text{edge } st \text{ is a cut edge in } \{S_1^{(k)}, S_2^{(k)}, ..., S_k^{(k)}\}\}} \quad \text{for all} \quad s \in V. \tag{2.10}$$

where a cut edge is an edge whose nodes are part of two different partition sets and $\mathbf{1}$ is the indicator function. This centrality measure computes the propensity of a node to lie on a cut edge when partitioning the graph into k clusters for different values of k ranging from 2 to some constant L.

Algorithm 2 Approximating Effective Resistances [3]

Require: Laplacian matrix L, incident matrix B, random Johnson-Lindenstrauss projection of size $m \times O(\log n)$
return effective resistance between nodes i and j
Solve the system $LX = B^T Q$
Compute $\|X_i - X_j\|_2$ where X_i is the ith-row of X for effective resistance $p_{st}(i) - p_{st}(j)$

3. Numerical Experiments. In this section, we compute the centrality measures on two synthetic data and two real data and compare them. We also analyze how discriminative each measure is and how correlated it is to the degrees of the nodes. From our results, we would like to see how effective k-means centrality is as a centrality measure compared to the others.

The visual representations of the graphs we produce were created by using Pajek, so nodes with larger centrality have larger size and they are closer to violet on the spectrum. Likewise, nodes with smaller centrality have smaller size and they are closer to red on the spectrum.

3.1. A Toy Network. We compute the centrality measures for the graph shown in Figure 3.1. Because nodes A and B are most central, they are expected to have high centrality values. Furthermore, being a symmetric graph, opposite nodes are expected to have the same centrality values. The results are shown in Figure 3.2.

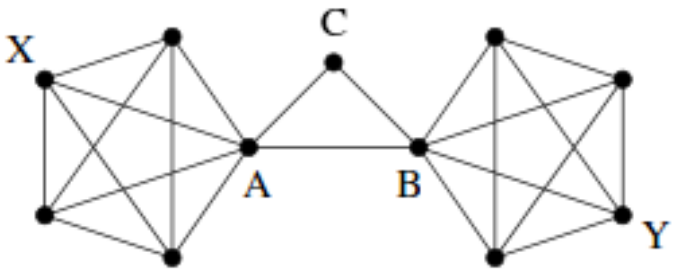


FIG. 3.1. *Toy Graph*

As expected, nodes A and B show higher centrality values than the other nodes for all five centrality measures. However, current-flow closeness centrality and k-means centrality lack symmetry in their values. The asymmetry for current-flow closeness is attributed to the Spielman-Srivastava's algorithm used to approximate effective resistance. The result shows that the algorithm performs poorly on small networks. As for k-means centrality, the random initial assignments of centroids for every run of the k-means algorithm may have caused the asymmetry.

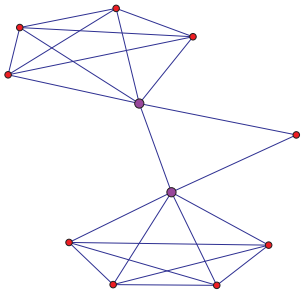
3.2. The "Tripartite" Graph. We compute the centrality measures on what we call the "tripartite" graph, which is shown in Figure 3.3. This graph has three major clusters of nodes: both outer clusters have 100 nodes and the inner cluster has 50 nodes. Within each cluster, nodes are connected to another with high probability, and the nodes of the outer clusters are connected to the nodes of the inner cluster with also a high probability. In addition, there is a small number of edges between the two outer clusters. The results of the "tripartite" graph is shown in Figure 3.4.

We observe that betweenness centrality performs exceptionally poorly on this graph because most of the nodes have low centrality values. This result is attributed to the edges between the outer clusters since they form some of the shortest paths of the graph. Moreover, we see that current-flow closeness centrality does not have as many nodes with low centrality values and nodes with high centrality values as the other centrality measures. Most of its nodes have intermediate centrality values instead. All the other centrality measures produce similar results.

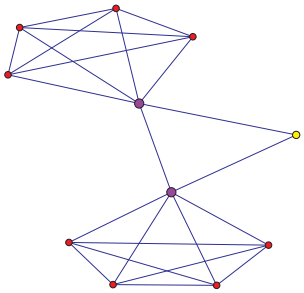
Because the graph have 250 nodes, we need to perform further tests to better analyze the centrality measures. We perform two tests: cluster recovery and average distance.

The cluster recovery test attempts to separate the graph based on the centrality measures into two partitions: the partition corresponding to the outer clusters and the partition corresponding to the inner cluster. We run k-means on the vector of centrality values with $k = 2$, thus producing a vector of indices of values either 1 or 2. The vector of the centrality values are ordered from 100 nodes of the left outer cluster, 50 nodes of the inner cluster, and finally 100 nodes of the right outer cluster. Thus a centrality measure performs well on this test if the first 100

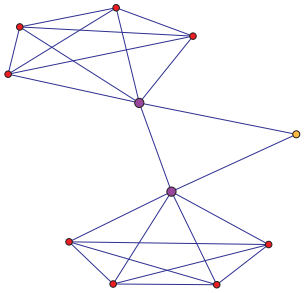
(a) Betweenness Centrality of Toy Graph



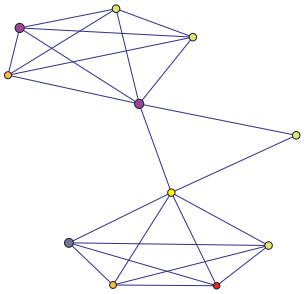
(b) Closeness Centrality of Toy Graph



(c) Current-Flow Betweenness Centrality of Toy Graph



(d) Current-Flow Closeness Centrality of Toy Graph



(e) k-Means Centrality of Toy Graph

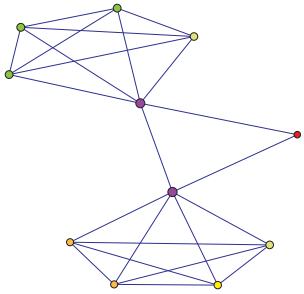


FIG. 3.2. *Centrality Measures of Toy Graph*

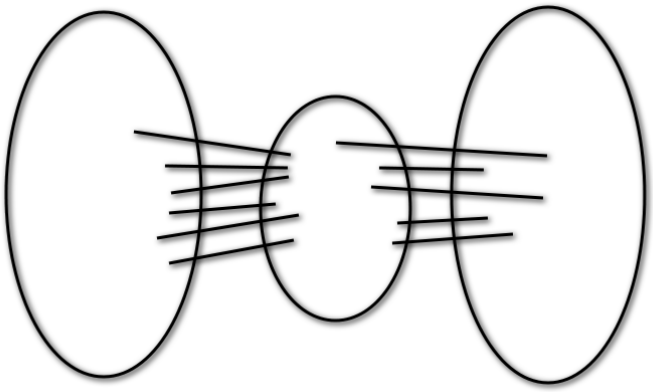


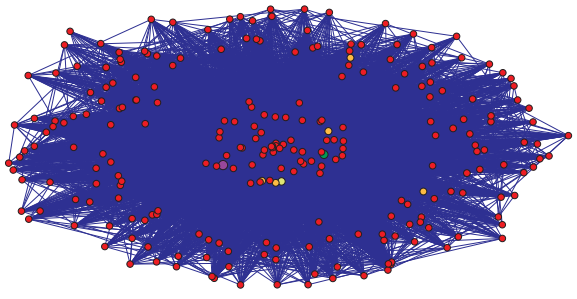
FIG. 3.3. *Tripartite Graph*

nodes and the last 100 nodes are clustered in one group and the rest in another group. The results for this test are shown in Figure 3.5.

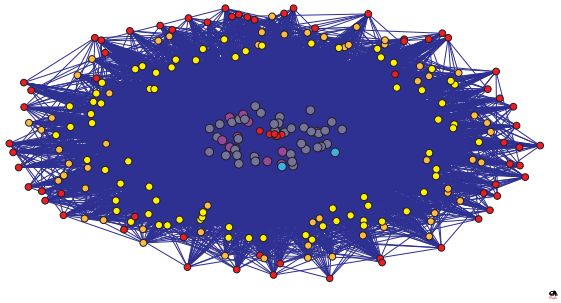
According to our results, the centrality scores of closeness, current-flow betweenness, and k-means centrality

were successfully partitioned into two representative clusters of the "tripartite" graph. Betweenness clustered majority of the nodes into one group. Current-flow closeness was successful in classifying the middle 50 nodes into one group, but some of the outer nodes, corresponding to the first and last 100 clusters, are clustered in the same group. This shows that current-flow closeness is vulnerable to noise, which are the edges between the outer clusters of the "tripartite" graph.

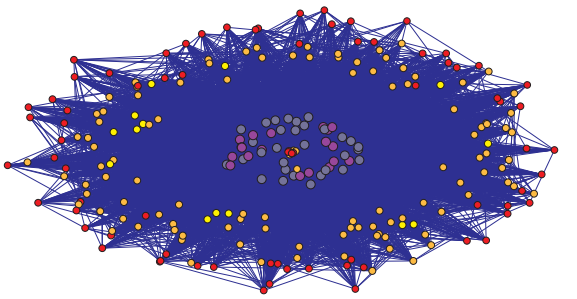
(a) Betweenness Centrality of "Tripartite" Graph



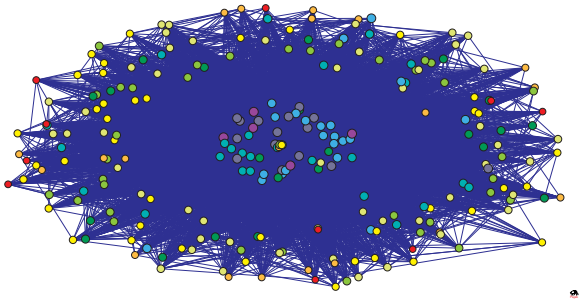
(b) Closeness Centrality of "Tripartite" Graph



(c) Current-Flow Betweenness Centrality of "Tripartite" Graph



(d) Current-Flow Closeness Centrality of "Tripartite" Graph



(e) k-Means Centrality of "Tripartite" Graph

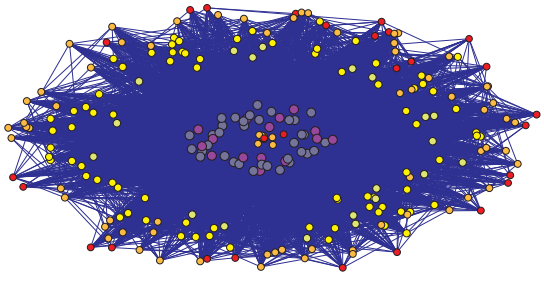
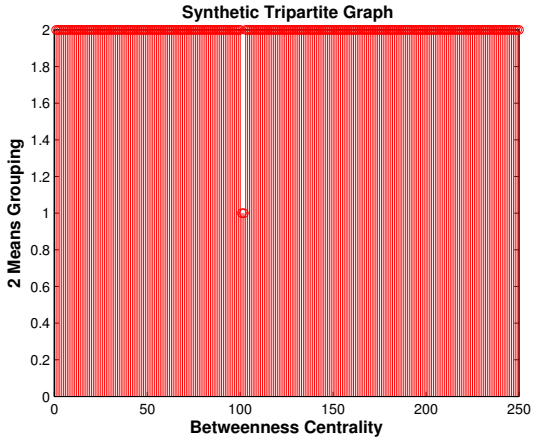


FIG. 3.4. *Centrality Measures of "Tripartite" Graph*

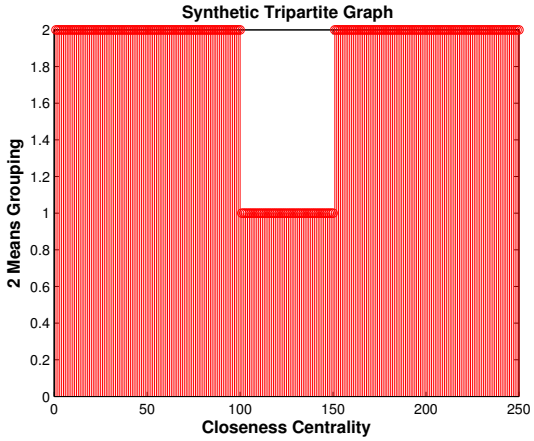
Our next test is the average distance test, where we sort the vector of centrality values in descending order. In the sorted vector, we expect that the first 50 nodes to correspond to the nodes of the inner cluster and the last 200 nodes to correspond to nodes of the outer clusters of the "tripartite" graph because of their expected centrality values. For every cut point p, which is an index of the vector that divides it into two groups, we the compute average pairwise distances σ_1 and σ_2 of the two groups. Then we plot $\lambda = \frac{\sigma_1+\sigma_2}{2}$ for each cut point p. We expect that the minimum value of λ to be at $p = 50$ since that is where the vector is divided into two groups of high centrality values and low centrality values, that is the inner cluster and the outer clusters, respectively. As a result, the difference in centrality values within each group should be small as compared to other instances for different value p. The test is explained in more details as follows:

1. Divide the vector of centrality values into two groups by the index p, so their sizes are p and q such that $p + q = n$.
2. Compute the distance matrices D_1 and D_2 based on Euclidean norm for both groups and compute the

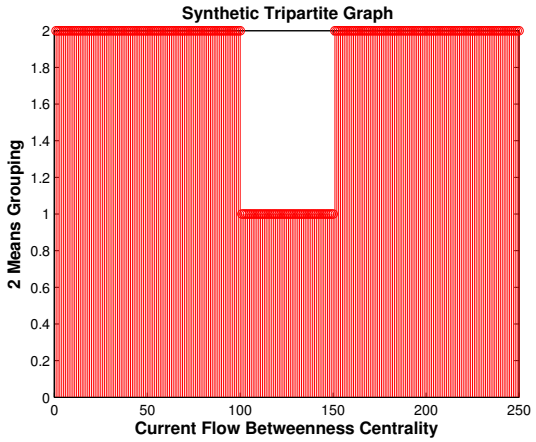
(a) Betweenness Centrality Partitioned into 2 Clusters



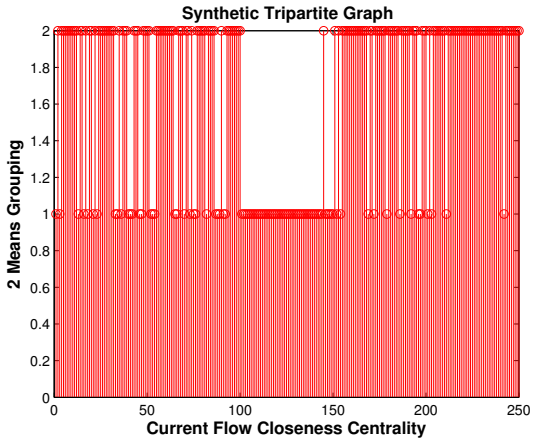
(b) Closeness Centrality Partitioned into 2 Clusters



(c) Current-Flow Betweenness Centrality Partitioned into 2 Clusters



(d) Current-Flow Closeness Centrality Partitioned into 2 Clusters



(e) k-Means Centrality Partitioned into 2 Clusters

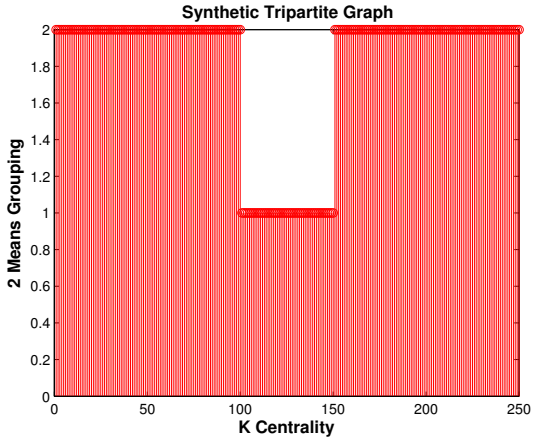


FIG. 3.5. *Cluster Recovery Test on "Tripartite" Graph*

average pairwise distances σ_1 and $\sigma 2$ where

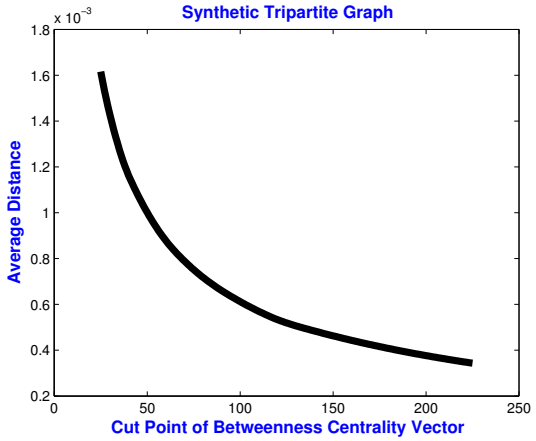
$$\sigma_1 = \frac{\sum_{k=1}^{p}[D_1 \cdot \mathbb{1}]_k}{\binom{p}{2}} \qquad \sigma_2 = \frac{\sum_{k=1}^{q}[D_2 \cdot \mathbb{1}]_k}{\binom{q}{2}} \tag{3.1}$$

where $\mathbb{1}$ is the ones vector.

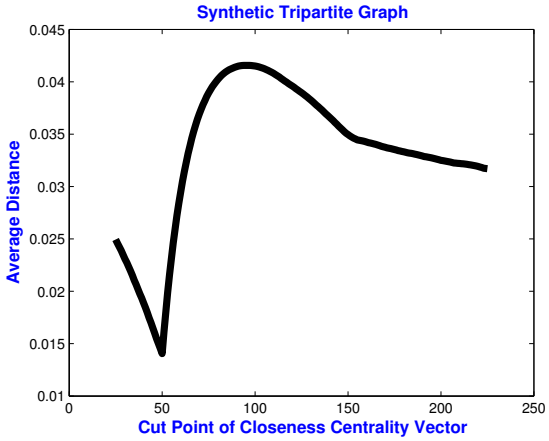
3. Compute $\lambda = \frac{\sigma_1+\sigma_2}{2}$ and plot it with its corresponding cut point p.

The plots of the test are displayed in Figure 3.6.

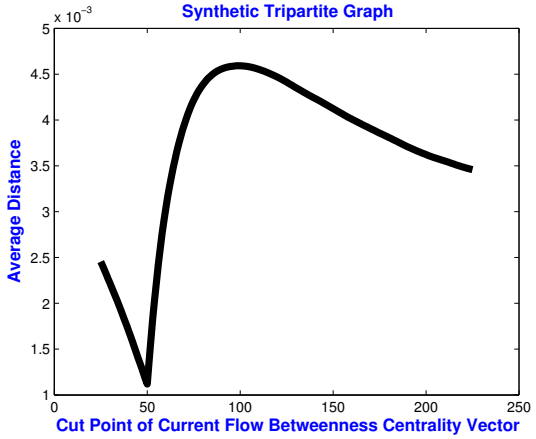
(a) Average Distance for Betweenness Centrality



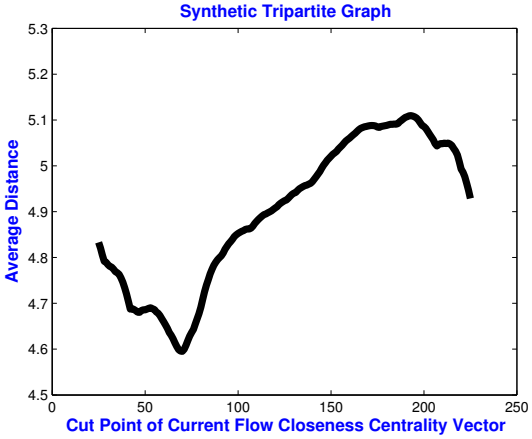
(b) Average Distance for Closeness Centrality



(c) Average Distance for Current-Flow Betweenness



(d) Average Distance for Current-Flow Closeness



(e) Average Distance for k-Means Centrality

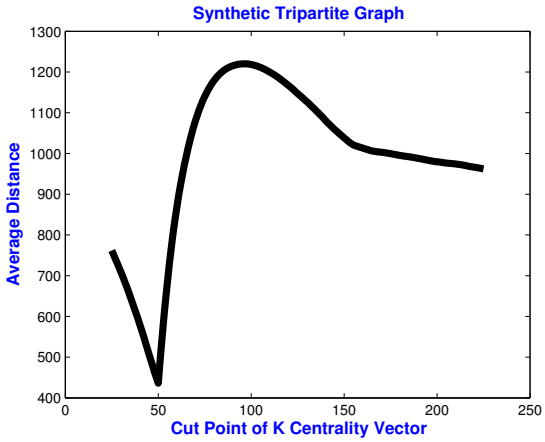


FIG. 3.6. *Average Distance Test on "Tripartite" Graph*

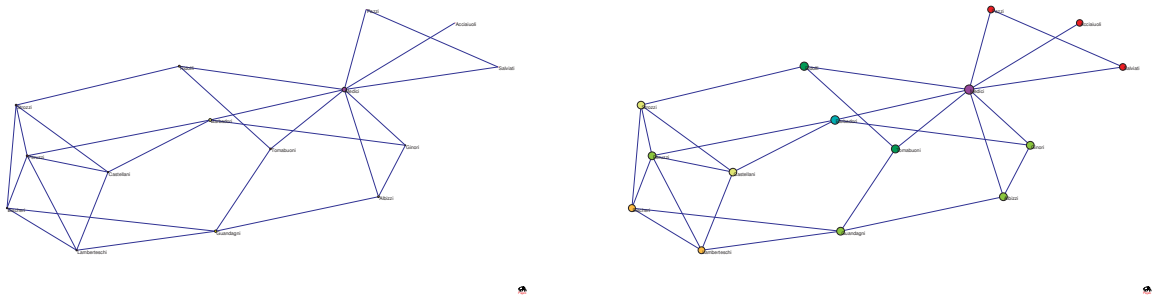
The results show that closeness centrality, current-flow betweenness, and k-means centrality performed well in this test since λ obtain its minimum at $p = 50$. Because there are edges between the outer clusters of the "tripartite" graph, λ for current-flow closeness does not obtain its minimum at $p = 50$. Furthermore, λ for betweenness centrality is unable to obtain its minimum $p = 50$ because it does not account for the information spread on non-shortest paths.

3.3. Florentine Network. Now we compute the centrality measures on the Florentine network, consisting of the fifteen most influential families of 15th century Florentine, Italy. Because the Medici family was considered the most influential at the time, its centrality values is expected to be high for any centrality measure. The results of the centrality measures are shown in Figure 3.7.

From our results, we see that the purple node with largest size in the graph is the Medici family for all centrality measures. However, for betweenness centrality and k-means centrality, the other nodes do not have much influence on the graph as evident by their sizes. On the other hand, the other centrality measures show that the other nodes have some considerable influences.

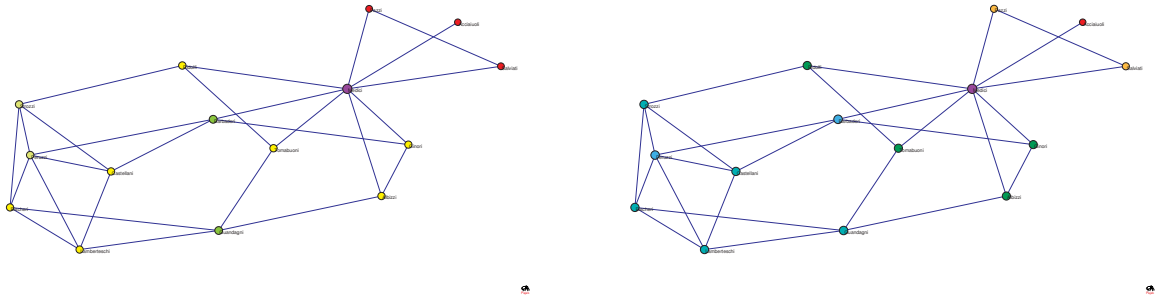
(a) Betweenness Centrality of Florentine Network

(b) Closeness Centrality of Florentine Network



(c) Current-Flow Betweenness Centrality of Florentine Network

(d) Current-Flow Closeness Centrality of Florentine Network



(e) k-Means Centrality of Florentine Network

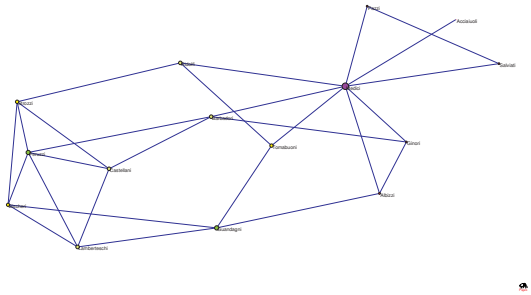


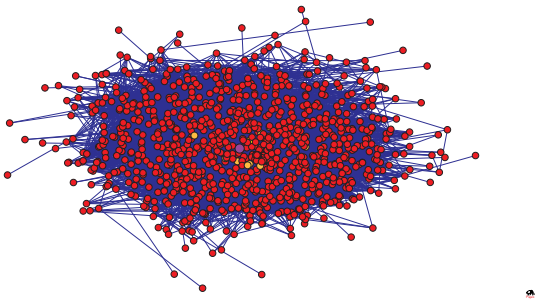
FIG. 3.7. *Centrality Measures on Florentine Network*

3.4. Reed College's Social Network. For a larger real data set, we compute the centrality measures on the Facebook friends network of Reed College containing 962 nodes. The results are shown in Figure 3.8

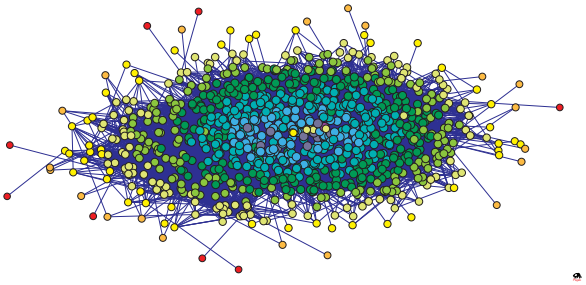
Similarly to the "tripartite" graph, we observe that betweenness centrality performs poorly since majority of the nodes have low centrality values. The other four centrality measures have varying degrees of success of distinguishing the nodes. Closeness and current-flow closeness centrality have very few nodes of low centrality values and more nodes of intermediate centrality values. Current-flow betweenness and k-means centrality have similar results, but current-flow betweenness has more nodes with low centrality values.

3.5. Discriminative Measure. A centrality measure is considered useful if it is able to discriminate different node properties in a graph. In other words, it is able to detect nodes with considerable, but not high or low, amount of influences. To determine how well a centrality measure discriminates the nodes, we compute the discrimination measure. To obtain the discrimination measure δ, we compute the distance matrix D based on the Euclidean norm

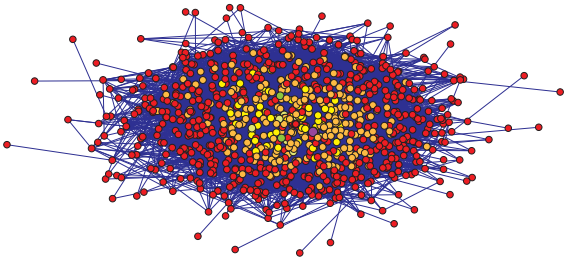
(a) Betweenness Centrality of Reed Network



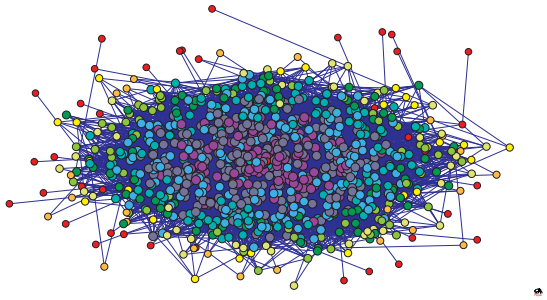
(b) Closeness Centrality of Reed Network



(c) Current-Flow Betweenness Centrality of Reed Network



(d) Current-Flow Closeness Centrality of Reed Network



(e) k-Means Centrality of Reed Network

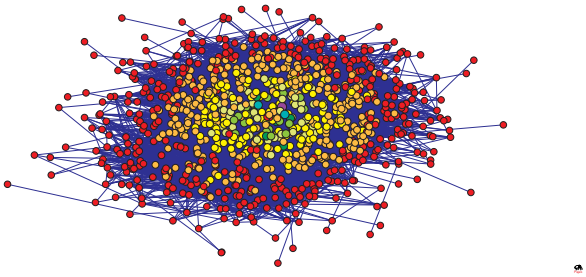


FIG. 3.8. *Centrality Measures on Reed Network*

from the vector of centrality values so that we have

$$\delta = \sum_{i=1}^{n} \left[\frac{v}{\| v \|} \right]_i \qquad \text{where} \quad v = \frac{D * \mathbb{1}}{n} \tag{3.2}$$

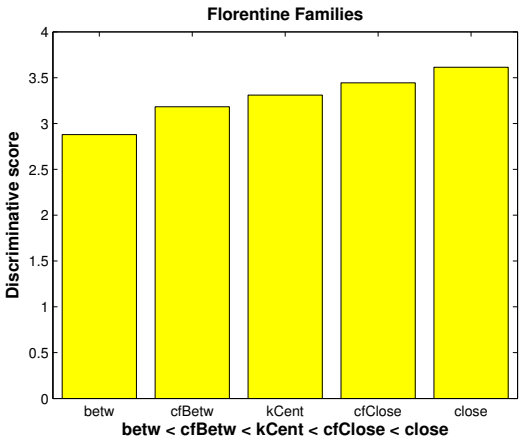
In short, the discriminative measure is the sum of the normalized vector of averaged pairwise distances. Thus, a larger discrimination measure indicates that the centrality measure better distinguishes different node properties. We compute the discriminative measure of the centrality measures for both Florentine network and Reed network and their results are shown in Figure 3.9

From the results, the ranking of the centrality measures from least discriminative to most discriminative is as follows for both networks: betweenness, current-flow betweenness, k-means centrality, current-flow closeness, and closeness centrality. This shows that the discrimination measure is consistent across networks. In the graphs, k-means centrality ranks in the middle, so it measures less discriminatively than closeness and current-flow closeness centrality but more discriminatively than betweenness and current-flow betweenness centrality. Furthermore, the discriminative measure for betweenness centrality verifies that it is not at all a useful centrality measure.

3.6. Correlation with Node Degrees. We compute the correlation between the values of the centrality measures and the node degrees. If the correlation with node degrees is high, this makes a centrality measure ineffective since computing the degrees would be more efficient for network analysis. The results are shown in Figure 3.10.

Unlike the case of discriminative measure, we obtain different rankings from lowest to highest correlation between Florentine and Reed networks since a particular centrality measure's propensity to match node degree

(a) Discriminative Measures of Florentine Network



(b) Discriminative Measures of Reed Network

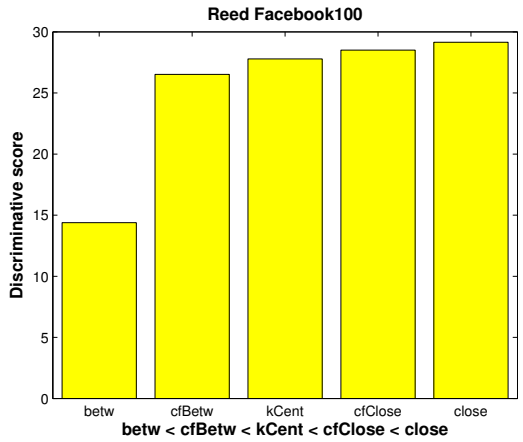
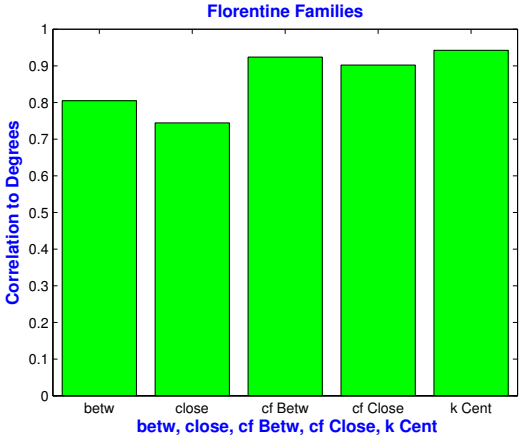


FIG. 3.9. *Discriminative Measures of Centrality Measures on Real Networks*

(a) Correlation Between Centrality Measures and Node Degrees of Florentine Network



(b) Correlation Between Centrality Measures and Node Degrees of Reed Network

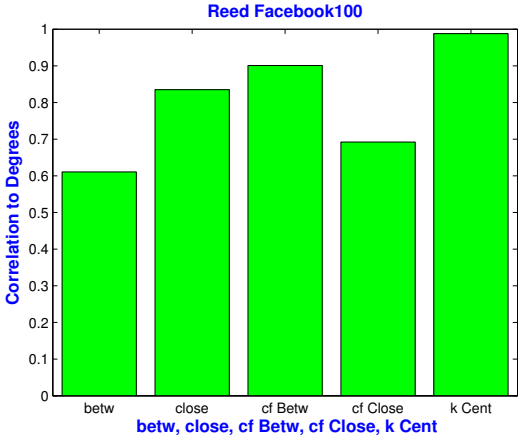


FIG. 3.10. *Correlation Between Centrality Measures and Node Degrees on Real Networks*

depends on the graph structure. For the Florentine network, the ranking are the following: closeness, betweenness, current-flow closeness, current-flow betweenness, and k-means centrality. For the Reed network, we have betweenness, current-flow closeness, closeness, current-flow betweeness, and k-means centrality. It is logical that k-means centrality would have a high correlation to node degree because it measures the tendency of a node to lie on a cut edge between two clusters of a network, where a cut node would be related to the number of edges adjacent to that node.

4. Summary and conclusion. After considering the results of the proposed k-means centrality on synthetic and real networks, we observe that it performs similarly to the existing centrality measures. k-means centrality can be utilized as a middle ground for the other centrality measures we compared it with. We also see that it is highly correlated to node degrees, which suggest that it might not be an effective centrality measure. However, as we mention in the introduction, a proper choice of centrality measure depends on the context for which it is being used. In our work, we focus mainly on the discriminative aspect of our new measure. It is certainly possible that the k-centrality would fare better under different circumstances. In particular, since this centrality is derived from the k-means algorithm, it may be fruitful to explore its results on other networks that have known distinct clusters within them.

REFERENCES

[1] Ulrik Brandes and Daniel Fleischer, *Centrality measures based on current flow*, Springer, 2005.
[2] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, USA, 2010.
[3] Daniel A. Spielman and Nikhil Srivastava, *Graph sparsification by effective resistances*, SIAM J. Comput, (2011), p. 19131926.