HellRank: A Hellinger-based Centrality Measure for Bipartite Social Networks

Seyed Mohammad Taheri
Hamidreza Mahyar
Mohammad Firouzi
Ali Movaghar

Received: 2016-10 / Accepted: date

Abstract Measuring centrality in a social network, especially in bipartite mode, poses several challenges such as requirement of full knowledge of the network topology and lack of properly detection of top-k behavioral representative users. In this paper, to overcome the aforementioned challenging issues, we propose an accurate centrality measure, called HellRank, to identify central nodes in bipartite social networks. HellRank is based on the Hellinger distance between two nodes on the same side of a bipartite network. We theoretically analyze the impact of the Hellinger distance on a bipartite network and find an upper and lower bounds for this distance. The computation of HellRank centrality measure can be distributed by letting each node uses only local information on its immediate neighbors and therefore do not need a central entity to have full knowledge of the network topological structure. We experimentally evaluate performance of the HellRank measure in correlation with other centrality measures on real-world networks. The results show partial ranking similarity between the HellRank and the other conventional metrics according to the Kendall and Spearman rank correlation coefficient.

Keywords Bipartite Social Networks · Top-k Central Nodes · Hellinger Distance · Recommender Systems

1 Introduction

Social networks have become more popular recently with hundreds of millions of users. With the growth of spreading information across social networks, the question of “how to measure relative importance of users in a social network?” has become increasingly challenging and interesting as important users are more likely to be infected or to infect a large number of users. Understanding users’ behaviors when they connect to social networking sites creates opportunities for richer studies of social interactions. Also, finding a subset of users to statistically represent the original social network is a fundamental issue in Social Network Analysis. This small subset of users (more behavioral representative users) usually plays an important role in influencing the social dynamics on behavior and structure.

Centrality measure, a term in social network analysis, is widely used to quantify the relative importance of nodes within a network. The most central nodes are often the nodes that have more weight, both in terms of the number of interactions as well as the number of connections to other nodes [62]. In social network analysis, such centrality notion is used to identify influential users [62,73,77], as the influence of a user is the ability to popularize a particular content in the social network. To this end, various centrality measures have been proposed over the years to rank the network nodes according to their topological and structural properties [11,23,79]. These measures can be considered as several points of view with different computational complexity, ranging from low-cost measures (e.g., Degree centrality) to more costly measures (e.g., Betweenness and Closeness centralities) [72,52]. Moreover, [64] concluded that centrality may not be restricted to shortest paths. Furthermore, the global topological structure of
Fig. 1: Simple Bipartite graph $G = (V_1, V_2, E)$ with two different nodes set $V_1 = \{A, B, C, D\}$, $V_2 = \{1, 2, 3, 4, 5, 6, 7\}$ and link set $E$ that each link connects a node in $V_1$ to a node in $V_2$.

many networks is initially unknown, while these structural network metrics require full knowledge of the network topology [72,47].

In one hand, many real-world social networks have a bi-modal nature that allows the network to be modeled as a bipartite graph (see Figure 1). In a bipartite network, there are two types of nodes and the links can only connect nodes of different types [79]. Social Recommender System is one of the most important systems that can be modeled by the bipartite networks in which two types of nodes are users and items. In such systems, the centrality measure can have different interpretations from conventional centrality measures such as Betweenness, Closeness, Degree and PageRank [32]. The Structural metrics, such as Betweenness and Closeness centrality, are known as the most common central nodes’ identifier in one-mode networks, although in bipartite social networks they are not usually appropriate in identifying central users that are perfect representative for the bipartite network structure. For example, in a social recommender system modeled by network graph in Figure 1, user $D \in V_1$ is associated with items that have too low connections and have been considered less often by other users; meanwhile user $D$ is considered as the most central node based on these common centrality metrics, because it has more connections. However user $B \in V_1$ is much more real representative than $D$ in the network.

On the other hand, the common centrality measures typically defined for non-bipartite networks. But for using these measures in bipartite networks, some projection methods have been introduced to convert bipartite networks to one-mode [80,61]. In these methods, a bipartite network is projected by considering one of the two node sets and if each pair of these nodes shares a neighbor in the network, two nodes will be connected in the projected one-mode network[37,41]. One of the major challenges is that every link in a real network is formed independently, but this does not happen in the projected one-mode network. Because of lack of independency in the formation of links in the projected network, analysis of the metrics that use the random network [7] as a basis for their approach, is difficult. Because the classic random networks are formed by the assumption of links being independent from each other [56]. The second challenge is that the projected bipartite network nodes tend to form Cliques. A clique is a fully connected subset of nodes that all of its members are neighbors. As a result, the metrics that are based on triangles (i.e., a clique on three nodes) in the network, can be inefficient (such as structural holes or clustering coefficient measures) [42,56].

Despite the fact that the projected one-mode network is less informative than its corresponding bipartite representation, some of the measures for one-mode networks have been extended to bipartite mode [33,56]. Also, because of requirement of full knowledge of network topology and lack of proper measure for detection of more behavioral representative users in bipartite social network, the use of conventional centrality measures in the large-scale bipartite social networks for using in the recommender systems is a challenging issue. However, In order to overcome the aforementioned challenging issue and retain the original information in bipartite networks, the need for proposing an accurate centrality measure in such networks seems essential [31, 41].

Motivated by these observations and taking into account users’ importance indicators for detection of central nodes in Social Recommender Systems, we introduce a new centrality measure, called HellRank, to identify central nodes in the bipartite social networks. HellRank is based on the Hellinger distance[54], on one type of f-divergence measure, that indicates structural similarity of each node to other network nodes. So, this distance-based measure is accurate for detection of more behavioral representative nodes. We empirically show that nodes with high HellRank centrality measure have relatively high Degree, Betweenness and PageRank centrality measures in bipartite networks. In the proposed measure, despite of different objectives to identify central nodes, there is a partial correlation between HellRank and other common metrics.

This paper is organized as follows. In the next Section, we discuss the related work on behavioral representative and influence identification mechanisms and centrality measures for bipartite networks, and then highlight the research gap between our objectives and previous work. In Section 3, we introduce our proposed...
measure to solve the problem of centrality measure in bipartite networks. Experimental results and discussions are presented in Section 4. We conclude our work and discuss the future works in Section 5.

2 Related Work

There are several relevant studies on social influence analysis and the problem of important users in three different categories. First, in Section 2.1, we study existing works on behavioral representative users detection methods from social networks. Second, we review previous mechanisms for identifying influential users in social networks by considering the influence as a measure of the relative importance, in Section 2.2. Third, we focus in more details on centrality measures for bipartite networks, in Section 2.3.

2.1 Behavioral Representative Users Detection

Unlike influence maximization, in which the goal is to find a set of nodes in a social network who can maximize the spread of influence [13,30], the objective of behavioral representative users is to identify a few average users who can statistically represent the characteristics of all users [36]. Another type of related work is social influence analysis. [3] and [63] proposed methods to qualitatively measure the existence of influence. [14] studied the correlation between social similarity and influence. [66] presented a method for measuring the strength of such influence. The problem of sampling representative users from social networks is also relevant to graph sampling [39,50,71]. [5] presented a comprehensive view of user behavior by characterizing the type, frequency, and sequence of activities users engage in and described representative user behaviors in online social networks based on clickstream data. [24] found significant diversity in end-host behavior across environments for many features, thus indicating that profiles computed for a user in one environment yield inaccurate representations of the same user in a different environment. [49] proposed a methodology for characterizing and identifying user behaviors in online social networks.

However, most existing works focused on studying the network topology and ignored the topic information. [65] aimed to find representative users from the information spreading perspective and [2] studied the network sampling problem in the dynamic environment. [58] presented a sampling-based algorithm to efficiently explore a user’s ego network respecting its structure and to quickly approximate quantities of interest. [17] focused on use the social structure of user community, in addition to the users profiles and previous behaviors, as an additional source of information in building recommender systems. [68] presented a formal definition of the problem of sampling representative users from social network.

2.2 Identifying Influential Users

[25] studied how to infer social probabilities of influence by developing an algorithm to scan over the log of actions of social network users using real data. [6,67] focused on influence maximization problem to model the social influence on large networks. TwitterRank, as an extension of PageRank metric, was proposed by [74] to identify influential users in Twitter. [12] used the Susceptible-Infected-Recovered (SIR) model to examine the spreading influence of the nodes ranked by different influence measures. [75] identified influencers using joint influence powers through Influence network. [81] identified influence users by using user trust network. [40] proposed weighted LeaderRank technique by replacing the standard random walk to a biased random walk. [69] proposed a new approach to incorporate users’ reply relationship, conversation content and response immediacy to identify influential users of online health care community. [19] used multi-attribute and homophily characteristics in new method to identify influential nodes in complex networks.

In the specific area of identifying influential users in networks with two mode nature, [4] presented a dynamical model for rewiring in bipartite networks and obtained time-dependent degree distributions. [41] defined a bipartite clustering coefficient by taking differently structured clusters into account, that can find important nodes across communities. The concept of clustering coefficient will be discussed in further details in the following Section.

2.2.1 Clustering Coefficient

This measure shows nodes tendency to formation of clusters together and has attracted a lot of attention in both empirical and theoretical works. In many real-world networks, especially social networks, nodes incline to cluster in densely connected groups [57]. Many measures have been proposed to examine the tendency of nodes in networks. In particular, global clustering coefficient provides a overall assessment of clustering in the network [43], and local clustering coefficient evaluate clustering value of the immediate neighbors of a node in the network [18].
The global clustering coefficient is the fraction of 2-path (i.e., three nodes connected by two links) that are closed by the presence of a link between the first and the third node in the network. The local clustering coefficient is the fraction of links among a node’s interactions over the possible number of links between them [18, 56].

Due to structural differences in bipartite networks, applying these general clustering coefficients directly to a two-mode network, is not clearly appropriate [9], thus the common metrics were extended or redefined, and different clustering measures were defined in the bipartite networks. In one of the most common clustering coefficients in bipartite networks, 4-period density is measured instead of triangles [59, 78]. However, this measure could not consider triple closure concept in bipartite networks. In one-mode graphs, Degree centrality of node \( i \) is defined as the fraction of links among a node’s interactions to all others in the network with \( t \) others in the network. However in bipartite graphs, Degree centrality of node \( i \) is defined as the number of node’s connections to members on other side. Degree centrality is normalized for easier comparison. Hence, degree for each node is divided on the size of the other node set. Let \( d_i^* \) be the normalized Degree centrality of node \( i \) that is equal to [9, 20]:

\[
d_i^* = \frac{d_i}{n_i}, \quad d_j^* = \frac{d_j}{n_j}; \quad i \in V_1, j \in V_2
\]

As the network size becomes increasingly large, employing Degree centrality is the best option [76]. On the other hand, this centrality is based on a highly local view around each node, while we need more informative measures that can further distinguish among nodes with the same degree [29].

2.3.2 Closeness Centrality

The standard definition of Closeness \( c_i \) for node \( i \) in one-mode networks, refers to the sum of geodesic distances from node \( i \) to all \( n - 1 \) others in the network with \( n \) nodes [60]. For bipartite networks, this measure can be calculated using the same approach, but the main difference is in the normalization amount. Let \( c_i^* \) be the normalized Closeness centrality of node \( i \in V_1 \) is equal to [9, 20]:

\[
c_i^* = \frac{n_2 + 2(n_1 - 1)}{c_i}; \quad i \in V_1
\]

\[
c_j^* = \frac{n_1 + 2(n_2 - 1)}{c_j}; \quad j \in V_2
\]

For the example bipartite network shown in Figure 1, normalized Closeness centrality of the nodes A, B, C and D are respectively equal to 0.35, 0.61, 0.52 and 0.68. It specifies that node D is the most central node which says that Closeness centrality cannot help us much objective to find more behavioral representative nodes in bipartite social networks.

2.3.3 Betweenness Centrality

Betweenness centrality of node \( i, b_i \), refers to the fraction of shortest paths in the network that pass through node \( i \) [21]. In bipartite networks, maximum possible Betweenness for each node is limited by relative size of two nodes sets, as introduced by [10]:

\[
b_{\text{max}}(V_1) = \frac{1}{2} [n_2^2(s + 1)^2 + n_2(s + 1)(2t - s - 1) - t(2s - t + 3)]
\]
where \( s = (n_1 - 1) \mod n_2 \) and \( t = (n_1 - 1) \mod n_2 \); and

\[
b_{\text{max}}(V_2) = \frac{1}{2} n_1^2 (p + 1)^2 + n_1 (p + 1) (2r - p - 1) - r (2p - r + 3)
\]

where \( p = (n_2 - 1) \mod n_1 \) and \( r = (n_2 - 1) \mod n_1 \).

For the example bipartite network shown in Figure 1, normalized Betweenness centrality of the nodes A, B, C and D are respectively equal to 0, 0.45, 0.71 and 0.71. It specifies that nodes C and D are the most central nodes which says that Betweenness centrality cannot help us much objective to find more behavioral representative nodes in bipartite social networks.

### 2.3.4 Eigenvector and PageRank Centrality

Another important centrality measure is eigenvector centrality, which is defined as principal eigenvector of adjacency matrix of the network. A node’s score is proportional to the sum of the scores of its immediate neighbors. This measure exploits the idea that nodes with connections to high-score nodes are more central [8]. The eigenvector centrality of node \( i \), \( e_i \), is defined as [20]:

\[
e_i = \lambda \sum a_{ij} e_j
\]

where \( \text{Adj}(G) = (a_{ij})_{i,j=1}^{n} \) denotes the adjacency matrix of the network with \( n \) nodes and \( \lambda \) is the principal eigenvalue of the adjacency matrix. Our interest regarding to the eigenvector centrality is particularly focused on the distributed computation of PageRank[76]. The PageRank is a special case of eigenvector centrality when the adjacency matrix is suitably normalized to obtain a column stochastic matrix [76]. The PageRank vector \( R = (r_1, r_2, \ldots, r_n)^T \) is the solution of the following equation:

\[
R = \frac{1 - d}{n} 1 + d LR
\]

where \( r_i \) is the PageRank of node \( i \) and \( n \) is the total number of nodes, \( d \) is a damping factor sets around 0.85 and \( L \) is a modified adjacency matrix, such that \( l_{i,j} = 0 \) if and only if node \( j \) does not have a link to \( i \) and \( \sum_{i=1}^{n} l_{i,j} = 1 \). \( l_{i,j} = \frac{a_{i,j}}{d_j} \), and \( d_j = \sum_{i=1}^{n} a_{i,j} \) is the out-degree of node \( j \) [55].

For the example bipartite network shown in Figure 1, normalized PageRank centrality of the nodes A, B, C and D are respectively equal to 0.05, 0.11, 0.08 and 0.21. It specifies that node D is the most central node which says that PageRank centrality cannot help us much objective to find more behavioral representative nodes in bipartite social networks.

### 3 Proposed Method

In this paper, we want to identify more behavioral representative nodes in bipartite social networks. To this end, we propose a new similarity-based centrality measure. Since the similarity measure is usually inverse of the distance metrics, we first choose a suitable distance measure, namely Hellinger distance (Section 3.1). Then we apply this metric to bipartite networks. After that, we theoretically analyze the impact of the distance metric in the bipartite network. Next, we generate a distance matrix on one side of the network. Finally, we compute the score of each node, called HellRank, accordingly to this matrix. As a result, the nodes with high HellRank centrality are more behavioral representative nodes in bipartite social networks.

#### 3.1 Select A Well-Defined Distance Metric

When we want to choose a base metric, an important point is whether this measure is based on a well-defined mathematical metric. Since we want to introduce a similarity based measure for each pair of nodes in the network. We choose a proper distance measure as base metric, because the similarity measures are in some sense the inverse of the distance metrics. A true distance metric must have several main characteristics. A metric with these characteristics on a space induces topological properties like open and closed sets, which leads to the study of more abstract topological spaces. [27] introduced the following definition for a distance metric.

**Definition 1** A metric space is a set \( X \) that has a notion of the distance function \( d(x, y) \) between every pair of points \( x, y \in X \). A well-defined distance metric \( d \) on a set \( X \) is a function \( d : X \times X \rightarrow \mathbb{R} \) such that for all \( x, y, z \in X \):

1. **Positive Definiteness:** \( d(x, y) \geq 0 \) and \( d(x, y) = 0 \) if and only if \( x = y \);
2. **Symmetry:** \( d(x, y) = d(y, x) \);
3. **Triangle Inequality:** \( d(x, z) \leq d(x, y) + d(y, z) \).

We define our distance function as the difference between probability distribution for each pair of nodes based on f-divergence function, which is defined by [15]:

**Definition 2** An f-divergence is a function \( D_f(P||Q) \) that measures the difference between two probability distributions \( P \) and \( Q \). For a convex function \( f \) such that \( f(1) = 0 \), the f-divergence of \( Q \) from \( P \) is defined as:

\[
D_f(P||Q) = \int_{\Omega} f\left(\frac{dP}{dQ}\right)dQ
\]
where $\Omega$ is a sample space, which is the set of all possible outcomes.

In this paper, we use one type of f-divergence metric, called Hellinger distance (aka Bhattacharyya distance), that was introduced by Ernst Hellinger in 1909 [54]. Although in probability theory and information theory, Kullback-Leibler divergence [35] is a more common measure of difference between two probability distributions, it does not satisfy both the symmetry and triangle inequality conditions [51]. Thus, this measure is not intuitively appropriate to explain similarity in our problem. As a result, we choose Hellinger distance to quantify the similarity between two probability distributions [51].

For two discrete probability distributions $P = (p_1, \ldots, p_m)$ and $Q = (q_1, \ldots, q_m)$, in which $m$ is length of the vectors, Hellinger distance is defined as:

$$D_H(P||Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{m} (\sqrt{p_i} - \sqrt{q_i})^2}$$

It is directly related to the Euclidean norm of the difference of the square root of vectors, i.e.,

$$D_H(P||Q) = \frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{Q} \|_2$$

3.2 Apply to Bipartite Networks

In this Section, we want to apply the Hellinger distance to a bipartite network for measuring the similarity of the nodes on one side of the network. Assume that $x$ is a node in a bipartite network in which its neighborhood is $N(x)$ and its degree is $\deg(x) = |N(x)|$. Suppose that the greatest node degree of the network is $\Delta$. Let $L_i$ be number of $x$’s neighbors with degree of $i$. Suppose the vector $L_x = (l_1, \ldots, l_\Delta)$ be the non-normalized distribution of $L_i$ for all adjacent neighbors of $x$. Now, we introduce the Hence Hellinger distance between two nodes, $x$ and $y$, on one side of the bipartite network as follows:

$$d(x, y) = \sqrt{2} \ D_H(L_x \parallel L_y)$$

The function $d(x, y)$ represents the difference between two probability distribution of $L_x$ and $L_y$. To the best of our knowledge, this is the first work that introduces the Hellinger distance between each pair of nodes in a bipartite network, using degree distribution of neighbors of each node.

3.3 Theoretical Analysis

In this Section, we first express the Hellinger distance for all positive real vectors to show that applying this distance to bipartite networks still satisfies its metricity (lemma 1) according to Definition 1. Then, we find upper and lower bound for the Hellinger distance between two nodes of bipartite network.

**Lemma 1** Hellinger distance for all positive real vectors is a well-defined distance metric function.

**Proof** Based on the true metric properties in Definition 1, for two probability distribution vectors $P$ and $Q$, the following holds:

$$D_H(P||Q) \geq 0$$
$$D_H(P||Q) = 0 \iff P = Q$$
$$D_H(P||Q) = d_H(Q||P)$$

If we have another probability distribution $R$ similar to $P$ and $Q$, then according to the triangle inequality in norm 2, we have:

$$\frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{Q} \|_2 \leq \frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{R} \|_2 + \| \sqrt{Q} - \sqrt{R} \|_2$$

$$\rightarrow D_H(P||Q) \leq D_H(P||R) + D_H(Q||R)$$

that shows triangle inequality in Hellinger distance for all positive real vectors is a well-defined distribution metric function.

Using this distance measure, we have the ability to detect differences between local structures of nodes. In other words, this distance expresses similarity between the local structures of two nodes. If we normalize the vectors (i.e., sum of the elements equals to one), then differences between local structures of nodes may not be observed. For example, there does not exist any distance between node $x$ with $\deg(x) = 10$ that its neighbors’ degree are 2 and node $y$ with $\deg(y) = 1$ that its neighbor’s degree are 2. Therefore, our distance measure with vectors normalization is not proper for comparing two nodes.

Then, we claim that if the difference between two nodes’ degree is greater (or smaller) than a certain value, the distance between these nodes cannot be less (or more) than a certain value. In other words, their local structures cannot be similar more (or less) than a certain value. In the following theorem, we find upper and lower bound for the Hellinger distance between two nodes on one side of a bipartite network using their degrees difference.

**Theorem 1** If we have two nodes $x$ and $y$ on one side of a bipartite network, such that $\deg(x) = k_1$, $\deg(y) = k_2$, and $k_1 \geq k_2$, then we have a lower bound for the distance between these nodes as:

$$d(x, y) \geq \sqrt{k_1} - \sqrt{k_2}$$

and an upper bound as:

$$d(x, y) \leq \sqrt{k_1} + k_2$$
Proof To prove the theorem, we use the Lagrange multipliers. Suppose \( L_x = (l_1, \ldots, l_x) \) and \( L_y = (h_1, \ldots, h_y) \) are positive real distribution vectors of nodes \( x \) and \( y \). We can minimize the distance between these nodes by solving \( \min_{L_x, L_y} \sqrt{2} D_H(L_x || L_y) \), which is equivalent to find the minimum square of their distance:

\[
\min_{L_x, L_y} 2 \left( D_H(L_x || L_y) \right)^2 = \min_{L_x, L_y} \sum_{i=1}^{\Delta} \left( \sqrt{l_i} - \sqrt{h_i} \right)^2
\]

So, Lagrangian function can be defined as follows:

\[
F(L_x, L_y, \lambda_1, \lambda_2) = \sum_{i=1}^{\Delta} \left( \sqrt{l_i} - \sqrt{h_i} \right)^2 + \lambda_1 (k_1 - \sum_{i=1}^{\Delta} l_i) + \lambda_2 (k_2 - \sum_{i=1}^{\Delta} h_i)
\]

Then, we take the first derivative with respect to \( l_i \):

\[
\frac{\partial F}{\partial l_i} = 1 - \frac{\sqrt{l_i}}{\sqrt{h_i}} - \lambda_1 = 0 \Rightarrow h_i = \frac{l_i}{(1 - \lambda_1)^2}
\]

Due to \( \sum_{i=1}^{\Delta} l_i = k_1 \) and \( \sum_{i=1}^{\Delta} h_i = k_2 \), we have:

\[
\sum_{i=1}^{\Delta} h_i = k_2 \Rightarrow \sum_{i=1}^{\Delta} \frac{l_i}{(1 - \lambda_1)^2} = k_2 \Rightarrow (1 - \lambda_1) = \pm \sqrt{\frac{k_1}{k_2}}
\]

But in order to satisfy \( \sqrt{h_i} = \frac{\sqrt{l_i}}{k_i} \), the statement \( 1 - \lambda_1 \) must be positive, thus:

\[
h_i = \frac{l_i}{(1 - \lambda_1)^2} = l_i \frac{k_2}{k_1}
\]

After derivation with respect to \( h_i \), we also reach similar conclusion. If this equation is true, then equality statement for minimum function will occur, as:

\[
\min_{L_x, L_y} 2 \left( D_H(L_x || L_y) \right)^2 = \sum_{i=1}^{\Delta} \left( \sqrt{l_i} - \sqrt{k_2/k_1} \right)^2 = \sum_{i=1}^{\Delta} l_i (1 - \sqrt{k_2/k_1})^2 = (1 - \sqrt{k_2/k_1})^2 \sum_{i=1}^{\Delta} l_i = (1 - \sqrt{k_2/k_1})^2 k_1 \Rightarrow
\]

\[
\min_{L_x, L_y} \sqrt{2} D_H(L_x || L_y) = \sqrt{k_1} (1 - \sqrt{k_2/k_1}) = \sqrt{k_1} - \sqrt{k_2}
\]

In the case of second equation for upper bound, as we know, the following statement is true for any \( p_i, p_j \) and \( q_i, q_j \):

\[
(\sqrt{p_i} - \sqrt{q_i})^2 + (\sqrt{p_j} - \sqrt{q_j})^2 \leq (\sqrt{p_i} + \sqrt{q_j} - 0)^2 + (\sqrt{q_i} + q_j - 0)^2 = p_i + p_j + q_i + q_j
\]

Suppose in our problem, \( p_i = l_i, p_j = l_j \), and \( q_i = h_i, q_j = h_j \), then this inequality holds for any pair of elements in \( L_x \) and \( L_y \). Eventually we have:

\[
d(x,y) \leq \sqrt{(\sqrt{k_1} - 0)^2 + (\sqrt{k_2} - 0)^2} = \sqrt{k_1 + k_2}
\]

We can conclude that it is not possible for any pair of nodes on one side of the bipartite network that their distance to be more than a certain value by increasing their degrees.

As a result, we found upper and lower bound for the Hellinger distance between two nodes on one side of a bipartite network using their degrees difference.

### 3.3.1 An Example with Probabilistic View

In this example, we want to analysis the similarity among nodes based on Hellinger distance information in an artificial network. We examine how we can obtain required information for finding similar nodes to a specific node \( x \) as the expected value and variance of the Hellinger distance. Suppose that in a bipartite artificial network with \( n_1 \) nodes in one side (\( V_1 \)) and \( n_2 \) nodes in other side (\( V_2 \)), nodes on one side is connected to nodes on other side using Erdős-Rényi model \( G(m,p) \). In other words, there is an link with probability \( p \) between two sides of the network. Distribution function \( L_x \) of node \( x \in V_1 \) can be expressed as a Multinomial distribution-form as:

\[
P(l_1, \ldots, l_\Delta | \text{deg}(x) = k) = P(L_x | \text{deg}(x) = k) = \left( \begin{array}{c} k \\ l_1, \ldots, l_\Delta \end{array} \right) \prod_{i=1}^{\Delta} P_i^{l_i}
\]

where \( P_i = \binom{n_2 - 1}{i - 1} p^{i-1} (1 - p)^{n_2 - i} \) is a Binomial distribution probability \( B(n_2, p) \) for \( x \)'s neighbors that their degree is equal to \( i \).

According to the Central Limit Theorem [28], Binomial distribution converges to a Poisson distribution \( \text{Pois}(\lambda) \) with parameter \( \lambda = (n_2 - 1)p \) and the assumption that \( (n_2 - 1)p \) is fixed and \( n_2 \) increases. Therefore, average distribution of \( P(L_x | \text{deg}(x) = k) \) will be \( \mu = (kp_1, kp_2, \ldots, kp_\Delta) \). In addition, degree distribution in Erdős-Rényi model converges to Poisson distribution by increasing \( n_1 \) and \( n_2 \) (\( \lambda = n_1p \) for one side of network and \( \lambda = n_2p \) for another one).
Limit of average distribution of \( P(L_x|\text{deg}(x) = k) \) by increasing \( \Delta \), approaches \( k \) times of a Poisson distribution. Thus, normalized \( L_x \) vector is a Poisson distribution with parameter \( \lambda = (n_2 - 1)p \). To find a threshold for positioning similar and closer nodes to node \( x \), we must obtain expectation and variance of the Hellinger distance between \( x \) and other nodes in node set \( V_1 \). Before obtaining these values, we mention the following lemma to derive equal expression of Hellinger distance and difference between typical mean and geometric mean.

**Lemma 2** Suppose two distribution probability vectors \( P = (p_1, \ldots, p_m) \) and \( Q = (q_1, \ldots, q_m) \) that \( P \) is \( k_1 \) times of a Poisson distribution probability vector \( P_1 \sim \text{Poisson}(\lambda_1) \) and \( Q \) is \( k_2 \) times of a Poisson distribution probability vector \( P_2 \sim \text{Poisson}(\lambda_2) \). The square of Hellinger distance between \( P \) and \( Q \) is calculated by:

\[
D_H^2(P \parallel Q) = \frac{k_1 + k_2}{2} - \sqrt{k_1 k_2 (1 - e^{-\frac{1}{2}(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2})}
\]

**Proof** The squared Hellinger distance between two Poisson distributions \( P_1 \) and \( P_2 \) with rate parameters \( \lambda_1 \) and \( \lambda_2 \) is [70]:

\[
D_H^2(P_1 \parallel P_2) = 1 - e^{-\frac{1}{2}(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2}
\]

Therefore, the squared Hellinger distance for probability vectors \( P \) and \( Q \), will be equal to \((\sum_{i=1}^{m} p_i = k_1, \sum_{i=1}^{m} q_i = k_2)\):

\[
D_H^2(P \parallel Q) = \frac{1}{2} \sum_{i=1}^{m} (\sqrt{p_i} - \sqrt{q_i})^2
\]

\[
= \frac{1}{2} \sum_{i=1}^{m} (p_i + q_i - 2\sqrt{p_i q_i})
\]

\[
= \frac{k_1 + k_2}{2} - \sqrt{k_1 k_2 (1 - e^{-\frac{1}{2}(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2})}
\]

However, in the special case of \( \lambda_1 = \lambda_2 \), we have:

\[
D_H^2(P \parallel Q) = \frac{k_1 + k_2}{2} - \sqrt{k_1 k_2}
\]

It means that the squared Hellinger distance is equal to difference between typical mean and geometric mean.

To calculate the second moment of distance between node \( x \in V_1 \) and any other nodes \( z \in V_1 \) in the same side of the bipartite network based on the lemma 2, we have:

\[
E_{z \in V_1} [d^2(x, z)] = E \left[ 2 D_H^2(L_x \parallel L) \right]
\]

\[
= \sum_{i=1}^{n_2} \left( \frac{e^{-n_2 p (n_1 p)^i}}{(n_1 p)!} (k + i - 2\sqrt{k_i}) \right)
\]

\[
\approx \frac{n_2}{(n_1 p)!} \left( e^{-n_2 p (n_1 p)^i} (k + i - 2\sqrt{k_i}) \right)
\]

Where \( L = (L_z | z \in V_1) \) and the infinite can be approximated by \( n_2 \) elements. Similarly, for distance expectation we have:

\[
E \left[ \sqrt{2} D_H(L_x \parallel L) \right] \approx \sum_{i=1}^{n_2} \left( \frac{e^{-n_2 p (n_1 p)^i}}{(n_1 p)!} \sqrt{(k + i - 2\sqrt{k_i})} \right)
\]

In addition, variance can also be obtained based on these calculated moments:

\[
Var_{z \in V_1} (d(x, z)) = E_{z \in V_1} [d^2(x, z)] - (E_{z \in V_1} [d(x, z)])^2
\]

Hence, using these parameters, the required threshold for finding similar nodes to a specific node \( x \), can be achieved. If we want to extend our method to more complex and realistic networks, we can assume that distribution \( L_x \) is a multiple of Poisson distribution (or any other distribution) vector with parameter \( \lambda_x \), in which \( \lambda_x \) can be extracted by either the information about structure of the network or appropriate maximum likelihood estimation for node \( x \). Therefore, the threshold will be more realistic and consistent with the structure of the real-world networks.

### 3.3.2 Generalization to Weighted Bipartite Networks

The introduced distance metric function can be extended to weighted networks. The generalized Hellinger distance between two nodes of the weighted bipartite network can be considered as:

\[
d(x, y) = \sqrt{2} D_H(W_x \parallel W_y)
\]

where \( W_x = (w_1, \ldots, w_\Delta) \), \( w_i = \sum_{j \in N(x), \text{deg}(j) = 1} w_j \), and \( w_j \) is the vector of weights on the links of the network.

### 3.4 Rank Prediction via HellRank

In this Section, we propose a new Hellinger-based centrality measure, called HellRank, for the bipartite networks. Now, according to the Section 3.2, we find the Hellinger distances between any pair of nodes in each side of a bipartite network. Then we generate an \( n_1 \times n_1 \) distance matrix \( (n_1 \), the number of nodes in one side of network). The Hellinger distance matrix of \( G \) shown...
in Figure 1 is as follows:

\[
\text{Hell-Matrix}(G) = \begin{pmatrix}
A & 0 & 0.42 & 0.54 & 1 \\
B & 0.42 & 0 & 0.12 & 0.86 \\
C & 0.54 & 0.12 & 0 & 0.82 \\
D & 1 & 0.86 & 0.82 & 1 \\
\end{pmatrix}
\]

According to the well-defined metric features (in Section 3.1) and the ability of mapping to Euclidean space, we can cluster nodes based on their distances. It means that any pair of nodes in the matrix with a less distance can be placed in one cluster by specific neighborhood radius. Moreover, By summing inverse of elements for each row in the distance matrix, we get final similarity score for each node of the network, namely HellRank as:

\[
\text{Hell-Score}(x) = \sum_{x,z \in V_1} \frac{1}{d(x, z)}
\]

Because a similarity measure is usually in some sense the inverse of a distance metric: they take on small values for dissimilar nodes and large values for similar nodes. The nodes in one side with higher similarity scores represent more behavioral representation of that side of the bipartite network. In other words, these nodes are more similar than others to that side of the network. HellRank actually indicates structural similarity for each node to other network nodes. For example network shown in Figure 1, according to Hellinger distance matrix, normalized HellRank of nodes A, B, C, and D are respectively equal to 0.71, 1, 0.94, and 0.52. It is clear that among all of the mentioned centrality measures in Section 2.2, only HellRank considers node ‘B’ as a more behavioral representative node. Hence, sorting the nodes based on their HellRank measures, we have a better rank prediction for nodes of the network, i.e., the nodes with high HellRank is more similar to other nodes. In addition, we find nodes with less scores to identify very specific nodes which are probably very different from other nodes in the graph. The nodes with less HellRank are very dissimilar to other nodes on that side of the bipartite network.

4 Experimental Evaluation

In this Section, we experimentally evaluate performance of the proposed HellRank measure in correlation with other centrality measures on real-world networks. After summarizing datasets and evaluation metrics used in the experiments, the rest of this Section addresses this goal. Finally, we present a simple example of mapping the Hellinger distance matrix to the Euclidean space to show clustering nodes based on their distances.

4.1 Datasets

To examine a measure for detection of central nodes in a two-mode network, South Davis women [16], is one of the first and the most common bipartite datasets, recently used. This network has a group of women and a series of events as two sides of network. A woman linked to an event if she presents at that event. Another data set used in the experiments is OPSAHL-collaboration network [53], which contains authorship links between authors and publications in the arXiv condensed matter Section (cond-mat) with 16726 authors and 22015 articles. A link represents an authorship connecting an author and a paper.

4.2 Evaluation Metrics

One of the most popular evaluation metrics for comparison of different node ranking measures is Kendall’s rank correlation coefficient (\(\tau\)). In fact, Kendall is a non-parametric statistic that is used to measure statistical correlation between two random variables [1]:

\[
\tau = \frac{N_{<\text{concordant pairs}>} - N_{<\text{discordant pairs}>}}{\frac{1}{2}n(n-1)}
\]

where \(N_{<S>}\) is the size of set \(S\).

Another way to evaluate ranking measures is binary vectors for detection of top-\(k\) central nodes. All of vector’s elements are zero by default and only top-\(k\) nodes’ values are equal to 1. To compare ranking vectors with the different metrics, we use Spearman’s rank correlation coefficient (\(\rho\)) that is a non-parametric statistics to measure the correlation coefficient between two random variables[38]:

\[
\rho = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}}
\]

where \(x_i\) and \(y_i\) are ranked variables and \(\overline{x}\) and \(\overline{y}\) are mean of these variables.

4.3 Correlation between HellRank and Common Measures

We implement our proposed HellRank measure using available tools in NetworkX [26]. To compare our measure to other common centrality measures such as Latapay Clustering Coefficient [Section 2.2.1], Bipartite Degree [Section 2.3.1], Bipartite Closeness [Section 2.3.2], Bipartite Betweenness [Section 2.3.3] and PageRank [Section 2.3.4], we perform the tests on Southern Davis Women dataset. In Figure 2, we observe the obtained ratings of these metrics (normalized by maximum value).
for 18 women in the Davis dataset. In general, approximate correlation can be seen between the proposed HellRank metric and the other conventional metrics in the women scores ranking. It shows that despite different objectives to identify the central users, there is a partial correlation between HellRank and other metrics.

Figure 3 shows scatter plots of standard metrics versus our proposed metric, on Davis bipartite network. Each point in the scatter plot corresponds to a women node in the network. Across all former metrics, there exist clear linear correlations between each two measures. More importantly, because of the possibility of distributed computing to calculate HellRank for each node, this metric can also be used in billion-scale graphs, while many of most common metrics such as Closeness or Betweenness are limited to small networks[72]. We observe high HellRank nodes have high bipartite Betweenness, bipartite Degree and bipartite Closeness. This reflects that high HellRank nodes have higher chances to reach all nodes within short number of steps, due to its larger number of connections. In contrast with high HellRank nodes, low HellRank nodes have various Latapy CC and projected Degree values. This implies that nodes that are hard to be differentiated by these measures can be easily separated by HellRank.

To more analysis of the correlations between measures, we use Kendall between ranking scores provided by different methods in Table 1 and Spearman’s rank correlation coefficient between top $k = 5$ nodes in Table 2. These tables illustrate the correlation between each pair in bipartite centrality measures and again emphasizes this point that despite different objectives to identify the central users, there is a partial correlation between HellRank and other common metrics.

In the next experiment, we compare the top-$k$ central users rankings produced by Latapy CC, PageRank, Bipartite and projected one-mode Betweenness, Degree, Closeness and HellRank with different values of $k$. We employ Spearman’s rank correlation coefficient measurement to compute the ranking similarity between two top-$k$ rankings. Figure 4 presents result of Spearman’s rank correlation coefficient between the top-$k$ rankings of HellRank and the other seven metrics, in terms of different values of $k$. As shown in the figure, the correlation values of top $k$ nodes in all rankings, reach almost constant limit at a specific value of $k$. This certain amount of $k$ is approximately equal 4 for all metrics. This means that the correlation does not increase at a certain threshold for $k = 4$ in the Davis dataset.

To evaluate the HellRank on a larger dataset, we repeated all the mentioned experiments for the arXiv cond-mat dataset. The scatter plots of standard metrics versus HellRank metric can be seen in Figure 5. The results show that there exist almost linear correlations between the two measures in Bipartite Betweenness, Bipartite Degree and PageRank. In contrast to these metrics, HellRank has not correlation with other metrics such as Bipartite Closeness, Latapy Clustering Coefficient and Projected Degree. This implies that nodes that are hard to be differentiated by these metrics, can be separated easily by HellRank metric.

Moreover, Spearman’s rank correlation with different values of $k$ in arXiv cond-mat dataset can be seen in Figure 6. We observe the correlation values from top $k$ nodes in all rankings, with different values of $k$, reach almost constant limit at a specific value of $k$. This certain amount of $k$ approximately equals to 1000 for all metrics except Bipartite Closeness and Latapy CC metrics. This means that the correlation does not increase at a certain threshold for $k = 1000$ in the arXiv cond-mat dataset.

4.4 Mapping the Hellinger Distance Matrix to the Euclidean Space

Since we have a well-defined metric features and ability of mapping the Hellinger distance matrix to the Euclidean space, other experiment that can be done on this matrix, is clustering nodes based on their distance. ThisHellinger distance matrix can then be treated as a valued adjacency matrix and visualized using standard graph layout algorithms. Figure 7 shows the result of such an analysis on Davis dataset. This figure is a depiction of Hellinger Distance for each pair of individuals, such that a line connecting two individuals indicates that their Hellinger distance are less than 0.50. The diagram clearly shows the separation of Flora and Olivia, and the bridging position of Nora.

5 Conclusion and Future Work

In this paper, we proposed HellRank centrality measure for properly detection of more behavioral representative users in bipartite social networks. As opposed to previous work, by using this metric we can avoid projection of bipartite networks into one-mode ones, which makes it possible to take much richer information from the two-mode networks. The computation of HellRank can

\[ \text{In a valued adjacency matrix, the cell entries can be any non-negative integer, indicating the strength or number of relations of a particular type or types [34].} \]
be distributed by letting each node uses only local information on its immediate neighbors. To improve the accuracy of HellRank, we can extend the neighborhood around each node. The HellRank centrality measure is based on the Hellinger distance between two nodes of a bipartite network and we theoretically find an upper and lower bounds for the distance.

We experimentally evaluated HellRank on the Southern Women Davis dataset and the results showed that Brenda, Evelyn, Nora, Ruth, and Theresa should be considered as important women. Our evaluation analyses depicted that the importance of a woman does not only depend on her Degree, Betweenness, and Closeness centralities. For instance, if Brenda with low Degree centrality is removed from the network, the information would not easily spread among other women. As another observation, Dorothy, Olivia, and Flora have very low HellRank centralities. These results are consistent with the results presented in Bonacich (1978), Doreian (1979), and Everett and Borgatti (1993).

As a future work, more meta data information can be taken into account besides the links in a bipartite network. Moreover, we can consider a bipartite network as a weighted graph in which the links are not merely binary entities, either present or not, but have associated a given weight that record their strength relative to one another. Furthermore, as HellRank measure is proper for detection of more behavioral representative users in bipartite social network, we can use this measure in Recommender Systems. In addition, we can detect top k central nodes in a network with indirect measurements and without full knowledge the network topological structure, using compressive sensing theory.

References

1. Abdi, H.: The Kendall Rank Correlation Coefficient. Encyclopedia of Measurement and Statistics pp. 508–510 (2007). URL https://www.utdallas.edu/ herve/Abdi-KendallCorrelation2007-pretty.pdf

2. Ahmed, N.K., Neville, J., Kompella, R.: Network sampling: From static to streaming graphs. ACM Transactions on Knowledge Discovery from Data (TKDD) 8(2), 7 (2014)

3. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 7–15. ACM (2008)

4. Beguerisse Díaz, M., Porter, M.A., Onnela, J.P.: Competition for popularity in bipartite networks. Chaos 20(4), 043,101 (2010). DOI 10.1063/1.3475411. URL http://www.ncbi.nlm.nih.gov/pubmed/21198071

5. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, pp. 49–62. ACM (2009)

6. Bharathi, S., Kempe, D., Salek, M.: Competitive influence maximization in social networks. Internet and Network Economics pp. 306–311 (2007)

7. Bollabás, B.: The evolution of random graphs. Transactions of the American Mathematical Society 286(1), 257–257 (1984). DOI 10.1090/S0002-9947-1984-0756039-5. URL http://www.citeulike.org/group/3509/article/4012374

8. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. The Journal of Mathematical Sociology 2(1), 113–120 (1972). DOI 10.1080/0022250X.1972.9989806

9. Borgatti, S.P., Everett, M.G.: Network analysis of 2-mode data. Social Networks 19(3), 243–269 (1997). DOI 10.1016/s0378-8733(96)00301-2. URL <Go to ISI>://WOS:A1997XE81300004

10. Borgatti, S.P., Halgin, D.S.: Analyzing Affiliation Networks. The Sage Handbook of Social Network Analysis (1941), 417–433 (2011)

11. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in Twitter: The million follower fallacy. Icwsm 10(10-17), 30 (2010)

12. Chen, D.B., Gao, H., Li, L., Zhou, T.: Identifying influential nodes in large-scale directed networks: The role of clustering. PLoS ONE 8(10), e77,455 (2013). DOI 10.1371/journal.pone.0077455

13. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 199–208. ACM (2009)

14. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th
HellRank: A Hellinger-based Centrality Measure for Bipartite Social Networks

1st workshop on Social network systems, pp. 1–6. ACM (2008)

50. Maiya, A.S., Berger-Wolf, T.Y.: Benefits of bias: Towards better characterization of network sampling. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 105–113. ACM (2011)

51. Mandl, P.: Asymptotic Statistics 2., vol. 42 (1986). DOI 10.2307/2530729

52. Mururanganathan, A., Road, R.G., Nadu, T., Road, R.G., Nadu, T.: Ranking the Influence Users in a Social Network Site Using an Improved Topsis 73(1), 1–11 (2015)

53. Newman, M.E.: The structure of scientific collaboration networks. Proc Natl Acad Sci U S A 98(2), 404–409 (2001). DOI 10.1073/pnas.021544898(2)[pii]

54. Nikulin, M.: Hellinger distance. Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer (2001). DOI 10.1016/0378-8733(89)90016-6

55. Okamoto, K., Chen, W., Li, X.Y.: Ranking of closeness centrality for large-scale social networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5059 LNCS, 186–195 (2008)

56. Opsahl, T.: Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks 35(2), 159–167 (2013). DOI 10.1016/j.socnet.2011.07.001

57. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. Social Networks 31(2), 155–163 (2009). DOI 10.1016/j.socnet.2009.02.002

58. Papagelis, M., Das, G., Koudas, N.: Sampling online social networks. IEEE Transactions on knowledge and data engineering 25(3), 662–676 (2013)

59. Robins, G.L., Alexander, M.: Small Worlds Among Interlocking Directors: Network Structure and Distance in Bipartite Graphs. Computational & Mathematical Organization Theory 10(1), 69–94 (2004). DOI 10.1023/B:CMOT.0000032580.12184.c0. URL http://www.springerlink.com/content/q567773rj070xvj2/

60. Sabidussi, G.: The centrality index of a graph. Psychometrika 31(4), 581–603 (1966)

61. Sawant, S., Pai, G.: Yelp Food Recommendation System. Stanford Machine Learning Project (CS229) (2013). URL http://cs229.stanford.edu/proj2013/SawantPai-YelpFoodRecommendationSystem.pdf

62. Silva, A., Guimarães, S., Meira, W., Zaki, M.: ProfileRank. Proceedings of the 7th Workshop on Social Network Mining and Analysis - SNAKDD ’13 pp. 1–9 (2013). DOI 10.1145/2501025.2501033. URL http://dl.acm.org/citation.cfm?doid=2501025.2501033

63. Singla, P., Richardson, M.: Yes, there is a correlation from social networks to personal behavior on the web. In: Proceedings of the 17th international conference on World Wide Web, pp. 655–664. ACM (2008)

64. Stephenson, K., Zelen, M.: Rethinking centrality: Methods and examples. Social Networks 11(1), 1–37 (1989). DOI 10.1016/0378-8733(89)90016-6

65. Sun, K., Morrison, D., Bruno, E., Marchand-Maillet, S.: Learning representative nodes in social networks. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 25–36. Springer (2013)

66. Tang, J., Sun, J., Wang, C., Yang, Z.: Social influence analysis in large-scale networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 807–816. ACM (2009)

67. Tang, J., Sun, J., Wang, C., Yang, Z.: Social influence analysis in large-scale networks. Proceedings of SIGKDD’09 pp. 807–816 (2009). DOI 10.1145/1557019.1557108

68. Tang, J., Zhang, C., Cai, K., Zhang, L., Su, Z.: Sampling representative users from large social networks. In: AAAI, pp. 304–310. Citeseer (2015)

69. Tang, X., Yang, C.C.: Ranking User Influence in Healthcare Social Media (. ACM Trans. Intell. Syst. Technol. 3(4), 1–21 (2012). DOI 10.1145/2337542.2337558

70. Torgersen, E.: Comparison of statistical experiments. 36. Cambridge University Press (1991)

71. Ugander, J., Karre, B., Backstrom, L., Kleinberg, J.: Graph cluster randomization: Network exposure to multiple universes. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 329–337. ACM (2013)

72. Wehmuth, K., Ziviani, A.: DACCER: Distributed Assessment of the Closeness Centrality Ranking in complex networks. Computer Networks 57(13), 2536–2548 (2013). DOI 10.1016/j.comnet.2013.05.001

73. Wei, D., Deng, X., Zhang, X., Deng, Y., Mahadevan, S.: Identifying influential nodes in weighted networks based on evidence theory. Physica A: Statistical Mechanics and its Applications 392(10), 2564–2575 (2013). DOI 10.1016/j.physa.2013.01.054

74. Weng, J., Lim, E.P., Jiang, J.: TwitterRank : Finding Topic-sensitive Influential Twitterers. New York Paper 504. 261–270 (2010). DOI 10.1145/1718487.1718520. URL http://portal.acm.org/citation.cfm?id=1718520

75. Weng, J., Xie, L., J, Liu, R.Y.K., Liao, S.S.Y.: Discovering target groups in social networking sites: An effective method for maximizing joint influential power. Electronic Commerce Research and Applications 11(4), 318–334 (2012). DOI 10.1016/j.elerap.2012.01.002

76. You, K., Tempo, R., Qiu, L.: Distributed Algorithms for Computation of Centrality Measures in Complex Networks pp. 1–14 (2015). URL http://arxiv.org/abs/1507.01694

77. Yustiawan, Y., Maharani, W., Gozali, A.A.: Degree Centrality for Social Network with Opsahl Method. In: Proceedings of SIGKDD’09 pp. 807–816 (2009). DOI 10.1145/1557019.1557108

78. Zhou, T., Ren, J., Medo, M., Zhang, Y.C.: Bipartite network projection and personal recommendations in reciprocal and bipartite social networks-An online dating case study. IEEE Intelligent Systems 30(4), 25–35 (2014). DOI 10.1109/MIS.2013.104

79. Zhao, K., Wang, X., Yu, M., Gao, B.: User recommendations in reciprocal and bipartite social networks-An online dating case study. IEEE Intelligent Systems 29(2), 27–35 (2014). DOI 10.1109/MIS.2013.104
Fig. 2: Comparison between rankings for all women in the Davis dataset based on various centrality metrics.
Fig. 3: The correlations between HellRank and the other standard centrality metrics on Davis (normalized by max value)
Fig. 4: Spearman’s rank correlation with different values of $k$ in Davis dataset
Fig. 5: The correlations between HellRank and the other standard centrality metrics on Opsahl (normalized by max value)
Fig. 6: Spearman’s rank correlation with different values of $k$ in arXiv cond-mat dataset
Fig. 7: Mapping Hellinger distance matrix to Euclidean Space. A tie indicates that the distance between two nodes is lesser than 0.50