

# Identifying Central Nodes in Directed and Weighted Networks

Sharanjit Kaur<sup>1</sup>

Department of Computer Science,  
Acharya Narendra Dev College,  
University of Delhi, Delhi

Ayushi Gupta<sup>2</sup>

Department of Computer Science,  
Sri Guru Tegh Bahadur Khalsa College,  
University of Delhi, Delhi

Rakhi Saxena<sup>\*3</sup>

Department of Computer Science,  
Deshbandhu College,  
University of Delhi, Delhi

**Abstract**—An issue of critical interest in complex network analysis is the identification of key players or important nodes. Centrality measures quantify the notion of importance and hence provide a mechanism to rank nodes within a network. Several centrality measures have been proposed for un-weighted, un-directed networks but applying or modifying them for networks in which edges are weighted and directed is challenging. Existing centrality measures for weighted, directed networks are by and large domain-specific. Depending upon the application, these measures prefer either the incoming or the outgoing links of a node to measure its importance. In this paper, we introduce a new centrality measure, *Affinity Centrality*, that leverages both weighted in-degrees as well as out-degrees of a node's local neighborhood. A tuning parameter permits the user to give preference to a node's neighbors in either incoming or outgoing direction. To evaluate the effectiveness of the proposed measure, we use three types of real-world networks - migration, trade, and animal social networks. Experimental results on these weighted, directed networks demonstrate that our centrality measure can rank nodes in consonance to the ground truth much better than the other established measures.

**Keywords**—Centrality; weighted network; directed network; migration network; world input output trade network; community structure

## I. INTRODUCTION

Data analysts from diverse domains represent relationships or ties between entities using graph-based network models. The semantic meaning of nodes and ties is, however, domain-specific; in social networks where nodes represent individuals, ties might represent friendship or face-to-face communication [17], [2] whereas, in web networks, ties signify the existence of hyperlinks between web pages [16]. In most real-world networks, ties are characterized by their strength as well as direction. For instance, in world trade networks, where links between nations represent the exchange of commodities, tie strength is the cash flow and its direction indicates either import or export [6]. When both the strength and direction of ties are available, modeling data as *weighted, directed network* can be more elucidative and revelatory.

Network models are generally deployed to explain or predict the behavior of entities [11]. One key requirement in these applications is to determine the 'most important' or 'central' node in a network. A centrality measure quantifies this notion of node importance and provides a means to rank nodes based on their importance. Central nodes are useful in varied applications such as predicting most cited authors [22], determining influential spreaders for product advertisement in

online social networks [12], [25], detecting influential criminals [9], performing resilience analysis of power grid networks [13], locating key areas of activity in the urban infrastructure of a city [1], and traffic sampling for intrusion detection [28].

Several centrality measures have been formulated to quantify the notion of central nodes in un-weighted/ weighted, un-directed networks and are surveyed in [7], [3], [4], [5]. However, quantification of node centrality is more challenging in complex weighted and directed networks due to the dynamic effect of weighted reciprocal links on its computation. Very few measures exist for such networks, and the area remains under-explored.

### A. The Problem and Motivation

PageRank (PR) proposed by Brin and Page to rank web pages is a popular and effective centrality measure [20], and there exist variations and extensions of PR for weighted, directed networks [27], [30]. These measures quantify the importance of a web page by iterative counting of the number and quality of its incoming links. The underlying assumption is that more important web pages have more *incoming* links from other central web pages. The problem is that this assumption, though correct for web pages, may not be valid for other domains. For example, in the migration networks, a state's importance in the network is affected not only by the incoming migrant population but also by the outgoing migrants from that state.

A pair of centrality measures that consider both incoming and outgoing links are computed through the Hyperlink-Induced Topic Search (HITS) algorithm for web pages. However, this method delivers two metrics - hub score and authority score [14]. A good hub page has *outgoing* links to many good authorities; a good authority page has *incoming* links from many good hub pages. Similarly, the recently proposed Bi-directional h-index also presents two measures,  $h_{in}$ -index and  $h_{out}$ -index that give preference to incoming and outgoing links, respectively [29].

This raises a critical question regarding the importance of *incoming* versus *outgoing* links when computing the relative importance of a node. We conjecture that, in some domains, incoming links have more impact than outgoing links, whereas, in others, it is vice versa. This trade-off offers the opportunity to define a novel measure that can tune the relative importance between incoming and outgoing ties.

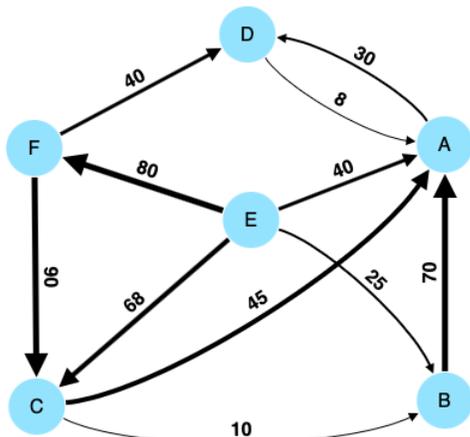


Fig. 1. Example Network.

Consider the example network shown in Fig. 1 modeled as a citation network where nodes are authors and weighted incoming link from author A to author B indicates the number of times A has cited B. In citation networks, importance of an author is commensurate with the number of citations, therefore incoming links should be given preference for computing centrality. Highly cited authors are more important, and if the citations are from other highly cited authors, then the importance should increase proportionately. In the example network, author A is the most central by virtue of receiving the highest incoming links (citations). Authors C and D receive citations from two authors each. Although author C gets more citations compared to D, the centrality of D should be high because of being cited by highly cited authors compared to C.

On the other hand, if the network in Fig. 1 is an organizational network of employees and weighted outgoing link represents the number of tasks assigned by employee A to employee B, then outgoing links should be preferred for computing the importance. An employee at higher position supervises a large number of employees and has the privilege to assign more tasks to them. Such an employee has higher importance in the organization compared to others. Following this hypothesis, employee E is the most central because this node has maximum outgoing links. Between employees F and C with equal number of outgoing links, employee F should be considered more important than C because F receives tasks from other important employees.

In the same vein, for applications such as analysis of trade or migration networks, both incoming and outgoing links could be given user specified weightage.

Recognizing these requirements, we propose a new centrality measure called *Affinity Centrality* that determines the importance of a node based on preference and influence proportions of its local network. We propose an intuitive upgradation of simple yet powerful weighted degree centrality by incorporating neighbors' attachment with the node. The quantum of centrality contributed by a node's neighbor is decided by the relative proportion of its incoming/outgoing

interactions. A tuning parameter permits the user to flexibly assign more weightage to either the in-neighbors or the out-neighbors of a node. Our centrality measure leverages only local node topology, which distinguishes it from well-established PageRank and HITS methods. Despite its simplicity, the measure is able to rank nodes in a better consonance to the ground truth than these established measures.

### B. Our Contributions

We introduce Affinity Centrality ( $\mathcal{AC}$ ), a centrality measure for weighted and directed networks. The summary of contributions follows.

- We propose a tunable centrality measure for quantifying the importance of a node by combining the advantages offered by its neighbors' topology via incoming and outgoing links (Section III).
- We perform an extensive evaluation of  $\mathcal{AC}$  on real-world migration and trade networks and compare its effectiveness with well established centrality measures (Section IV-B).
- We demonstrate empirically the effect of the tuning parameter in capturing the relative importance of the incoming versus outgoing ties (Section IV-C).
- We evaluate the role of central nodes delivered by the proposed centrality measure on the community structure of real-world networks (Section IV-D).

### C. Organization of the Paper

The paper is organized as follows: after a survey of centrality measures for weighted and directed networks in Section II, we present the proposed centrality measure ( $\mathcal{AC}$ ) in Section III. Section IV presents empirical investigations followed by conclusions in Section V.

## II. RELATED WORK

Vast literature exists for centrality measures designed for un-weighted and un-directed networks [5], [19], [12], [10], [18]. However, computing centrality for weighted and directed networks still faces some gaps in terms of incorporation of the direction of interactions in the computation. We briefly describe the existing work for directed and weighted networks by dividing them into two categories viz. i) Local-neighborhood based and ii) Global network structure based measures.

In the local-neighborhood based class, a node's importance is computed based on its interaction with  $l$ -hop neighbors where  $l$  indicates the number of hops. Opsahl et al. proposed a generalized centrality method to incorporate impact of degree along with the strength of interactions using a tuning parameter which can be tuned to give importance to either of the two aspects [19]. However, the proposed mechanism considers either incoming or outgoing direction in computation. Neighborhood centrality computes importance based on the centrality of a node and its 2-hop neighbors' centrality for un-directed and weighted networks [18]. The absence of direction in computation reduces its applicability to directed networks.

Global network structure based methods consider the influence of all nodes on the importance of a pivot node. Two

established algorithms in this category are HITS [14] and PageRank [20], that measure the probability of a random walker visiting a node on the web to assign a rank. HITS gives two scores Hub and Authority based on the direction considered whereas PageRank does ranking using incoming interactions only. Various extensions to these two algorithms have been proposed to extend them to directed weighted networks [29], [30], [24]. Zhang et al. proposed a weighted Pagerank algorithm for directed networks that incorporated the role of a node's degree, its strength and the node information using a tuning parameter to compute its rank [30]. Wang et al. modified the efficiency centrality for un-directed and weighted networks and incorporated both the degree and distance of all the nodes in a network [26]. Singh et al. proposed hybrid node-weighted centrality measures based on closeness and decay measures and made use of node information along-with edge weight to identify important nodes [23]. However, the high computational complexities of global network based algorithms make them unsuitable for large networks.

Designing an effective ranking measurement to capture the importance of nodes in a directed and weighted network is still an open challenge. Our proposed measure *Affinity Centrality* fills this gap by encapsulating both types of interactions along with their strength in the computation of the topological significance of a node in the network.

### III. AFFINITY CENTRALITY FOR WEIGHTED AND DIRECTED NETWORK

This section describes the proposed centrality measure called *Affinity Centrality* that leverages auxiliary information in a node's 1-hop neighborhood to determine its importance.

#### A. Notations used

Let  $G(V, E)$  be a weighted and directed network of order  $N = |V|$ , size  $M = |E|$  where  $V$  denotes the vertex set and  $E$  denotes the edge set. The network  $G$  can be represented by an asymmetric weighted adjacency matrix  $W := (w_{ij})$  of size  $N \times N$  where  $N = |V|$ . Each element  $w_{ij} \in \mathbb{Z}^+$  represents the strength of the interaction from node  $i$  to  $j$  and  $w_{ij} = 0$  represents no interaction. We use  $w_{i \rightarrow j}$  and  $w_{i \leftarrow j}$  to refer to the strength of outgoing and incoming ties of node  $i$ , respectively.

Let  $O_i$  denotes the total strength of all the outgoing ties and  $I_i$  denotes the total strength of all the incoming ties of a node  $i$  i.e  $I_i = \sum_j w_{i \leftarrow j}$  and  $O_i = \sum_j w_{i \rightarrow j}$ . Hence, total edge weight  $T = \sum_i I_i = \sum_i O_i$ . In case when weights are unknown,  $W = A$  where  $A := (a_{ij})$  is the standard adjacency matrix having  $a_{ij} = 1$  if nodes  $i$  and  $j$  are adjacent, otherwise 0. Notations used in the paper are detailed in the Table I for ready reference.

#### B. Constituents of Node Importance

A directed and weighted network, in general, consists of asymmetric interactions, and the direction of an interaction along with its weight emulates the preferential attachment of individuals in their neighborhood [8], [27], [19], [30]. The importance of a node depends upon its bonding with its local neighbors, which depends upon the strength and direction of the interactions. We refer to weight on an incoming edge as *in-strength* and weight on an outgoing edge as *out-strength*.

TABLE I. INTERPRETATION OF NOTATIONS USED

Notation	Interpretation
$G$	Network
$V$	Vertex set
$E$	Edge set
$N$	Order of $G$
$M$	Size of $G$
$W$	Weighted adjacency matrix
$T$	Total incoming/outgoing strength
$O_i$	Total outgoing strength of node $i$
$I_i$	Total incoming strength of node $i$
$L_i$	1-hop neighbors of node $i$
$w_{i \rightarrow j}$	Strength of outgoing tie from node $i$ to node $j$
$w_{i \leftarrow j}$	Strength of incoming tie of node $i$ from node $j$

We define the following two components based on the link interactions of a node with its direct neighbors.

- i. *Preference*: Favors from in-neighbors indicate their endorsement for an individual. For a node  $i$ ,  $\frac{w_{i \leftarrow j}}{O_j}$  determines the endorsement from its neighbor  $j$ . The higher the value, the more preferentially attached the node  $i$  is with node  $j$ . Also, the influence gained through preferential attachment increases if the endorsement for the node  $j$  within its local neighborhood is high too. In other words, resources gained by an individual show its power which is captured by its total in-strength ( $I_j$ ). Formally, preference ( $\beta_i$ ) of a node  $i$  with neighborhood set  $L_i$  is defined in the Eq. 1.

$$\beta_i = \frac{1}{T} \sum_{j \in L_i} \frac{w_{i \leftarrow j}}{O_j} I_j \quad (1)$$

- ii. *Influence*: The strength of the outgoing ties of a node  $i$  demonstrates its influence on its neighbors and captures its endorsement (preferences) for others. A higher value of  $\frac{w_{i \rightarrow j}}{I_j}$  indicates a high influence of node  $i$  on node  $j$ . Also, the influence of a node  $i$  propagates in the network through its neighbors, which is captured through their out-strength  $O_j$ . Collective endorsement of the neighbors along with an individual's support impacts its influence on others. Formally, influence  $\gamma_i$  of a node  $i$  is computed as given in the Eq. 2.

$$\gamma_i = \frac{1}{T} \sum_{j \in L_i} \frac{w_{i \rightarrow j}}{I_j} O_j \quad (2)$$

#### C. Affinity Centrality

The importance of a node depends upon its structural position in the network which depends upon its interactions with neighbors. We compute the proposed *affinity centrality* ( $AC$ ) by incorporating effect of *preference* and *influence* of neighbors on the node  $i$ , using a tuning parameter  $\theta \in [0, 1]$  (Eq. 3). Note that  $\theta$  gives flexibility to the end-user to include either of the in-strength and out-strength or both based on the application need.

$$AC_i = \theta \beta_i + (1 - \theta) \gamma_i \quad (3)$$

Using  $\theta = 1$  will reveal the influence of in-degree neighborhood on a node's *affinity*, in contrast to  $\theta = 0$ ,

that captures influence using its endorsement for neighbors. Using  $\theta = 0.5$  will incorporate the role of both *influence* and *preference* on the node's position in the network structure. The higher the position is, the more powerful/important that node is. For example, in trading, the importance of a supplier is dependent on imports as well as exports. Importing from the established suppliers increases its endorsement, whereas exporting to powerful vendors improves its position in the trade. Hence,  $\theta = 0.5$  is recommended in such scenarios. In case, influence is to be captured purely on the basis of imports/exports, then  $\theta = 0/1$  is recommended.

To substantiate the argument, we rank nodes of the example network (Fig. 1) by computing  $\mathcal{AC}$  with varying values of  $\theta$ . The ranks are shown in Table II. With  $\theta = 0$ , only outgoing links are considered for capturing centrality; hence node E is assigned the highest rank, and node F is ranked above C. With  $\theta = 1$ , outgoing links are ignored resulting in node A being ranked highest, and node D ranked above C. The results validate the motivation (see subsection I-A) and establish the theoretical formulation of the proposed centrality measure.

TABLE II. RANKING BY  $\mathcal{AC}$  MEASURE USING  $\theta = \{0, 0.5, 1\}$  OF NODES IN EXAMPLE NETWORK (FIG. 1)

Node	$\theta = 0$	$\theta = 0.5$	$\theta = 1$
A	5	1	1
B	4	6	5
C	3	3	3
D	6	4	2
E	1	2	6
F	2	5	4

#### D. Algorithmic Complexity

As the method exploits information of the nodes' neighborhood to quantify centrality, the computational complexity is  $O(M)$ . The proposed method is effective for handling large networks due to its  $O(M + N)$  storage space requirements.

### IV. EXPERIMENTAL EVALUATION

The goal of this section is to assess the performance of the proposed Affinity Centrality ( $\mathcal{AC}$ ) on the basis of the following questions.

- i. How effective is the ranking delivered by  $\mathcal{AC}$  measure?  
We inspect this question in Section IV-B using six weighted and directed real networks for which ground-truth can be crafted.
- ii. How does in-strength and out-strength impact the ranking computed by  $\mathcal{AC}$  measure?  
This investigation is done to demonstrate the role of *preference* and *influence* components on the importance of a node using a small sized weighted and directed network.
- iii. What is the role of topological central nodes on the community structure?  
We examined this question by extracting communities of the six networks and studied their evaluation in terms of important nodes delivered by the proposed centrality measure.

We evaluated the performance of  $\mathcal{AC}$  measure by comparing its results with two simple and widely used local centralities viz. Weighted in-degree (WI) and Weighted out-degree (WO). We also compare the results with the two global structure based algorithms viz. PageRank [20] and HITS [14] for weighted, directed networks.

We implemented our proposed measure  $\mathcal{AC}$  and variation of degree centrality - Weighted in-degree (WI) and Weighted out-degree (WO) in Python (64bits, v 3.6.9) and executed on Intel Core i3-4005U CPU @1.70GHz with 4GB RAM. We used the modules PageRank and HITS of the graph library networkx<sup>1</sup> of Python for comparison. Results of the experimentation are discussed in the following sub-sections after the description of networks used.

#### A. Real-world Networks

We consider three types of directed and weighted networks - migration, trade and animal networks to investigate the effectiveness and stability of the proposed  $\mathcal{AC}$  measure. Description of these publicly available networks are detailed below.

- i. We use Indian migration data of Census 2001 and Census 2011<sup>2</sup> for two factors viz. business and education to study the role of movement of population on state dynamics. For year  $X \in \{2001, 2011\}$ , we extracted the number of migrants for all possible pairs of 27 Indian states for each factor separately and created datasets named as *EducationX* and *BusinessX*. A dataset is mapped to a network by representing a state as a node and connecting a pair of states by a directed edge where direction captures the movement of migrants from the source state to the destination state. The absence of an edge between any two states indicates no migration between them. The number of migrants between two states is used as edge weight.
- ii. The World Input Output Network (WION) represents the transaction volumes (in million United States Dollars) between 56 sectors of 43 countries in the world<sup>3</sup>. A few of the sectors to name are - Water Transport, Construction, Telecommunications, Real estate activities, Publishing activities, Education, Fishing and Aquaculture. The transaction volumes of these sectors from a country (say A) to country (say B) are added to get the total transaction volume and vice versa. The countries are represented using the country codes which are used worldwide, for instance - CHN represents China, ESP represents Spain, and so on. Note that the data of the country named as ROW(rest of world) is not considered as it indicates transaction volumes to and from the rest of the world and is not associated with a particular pair of countries. We use WION data for two years - 2004 and 2014 which are named WION2004 and WION2014. Each dataset is mapped to a network by taking a country as a node. Two nodes are connected by a directed edge if and only if there is an import/export transaction between them and the volume of the transaction is used as the edge weight.

<sup>1</sup><https://networkx.org/>

<sup>2</sup><https://censusindia.gov.in/>

<sup>3</sup><http://www.wiod.org/database/wiots16>

TABLE III. PROPERTIES OF SIX WEIGHTED AND DIRECTED NETWORKS, CC: AVERAGE CLUSTERING COEFFICIENT, WD: WEIGHTED DENSITY

Category	Network	CC	Diameter	Transitivity	Reciprocity	WD	Size	Order
Migration	Business2001	0.92	2	0.91	0.94	738	617	27
	Education2001	0.95	2	0.95	0.95	565	662	27
	Business2011	0.91	2	0.90	0.92	758	609	27
	Education2011	0.94	2	0.94	0.94	621	646	27
Trade	WION2004	0.98	2	0.95	0.96	2078	1715	43
	WION2014	0.98	2	0.98	0.99	4013	1768	43
Animal	Moreno-Rhesus	0.61	4	0.59	0.76	2.7	111	16

iii. The Moreno Rhesus monkey grooming network represents a network of 16 monkeys<sup>4</sup>. The network consists of 16 nodes representing the monkeys and a weighted edge from a monkey (say A) to another monkey (say B) represents the number of times the monkey A groomed monkey B. We use this small network to demonstrate the role of the introduced tuning parameter ( $\theta$ ) on the node centrality.

Topological and structural properties of these networks are given in the Table III.

### B. Effectiveness of the Affinity Centrality

We study the effectiveness of the proposed centrality measure  $\mathcal{AC}$  by using six real networks in two categories: migration and trade. The topological characteristics of the networks are detailed in Table III. To evaluate the performance of the centrality measures, we consider the Gross Domestic Product (GDP) of a state/country as ground truth because it is the most commonly used measure of economic activity and stability during a period of time (typically 1 year). Higher GDP of a state/country indicates richness in terms of resources and services and raises the living standards of its residents by offering more jobs, business opportunities, etc..

We compare the performance of the  $\mathcal{AC}$  measure with four popular centrality measures viz, i) In-strength (WI) ii) Out-strength (WO) iii) Weighted Pagerank (PR) and Weighted HITS (both hub and authority scores). For each centrality measure, we rank nodes such that rank 1 is assigned to the largest value and so on. Spearman's rank correlation coefficient [21] is used to find the correlation of computed node ranks with ground-truth ranks, where correlation value indicates the ability of centrality measures to deliver correct ranks.

Table IV shows the ranking assigned by different measures to the top-10 ranked nodes as per GDP (first column in the table) for six networks. Ranking by  $\mathcal{AC}$  measure is computed using  $\theta = 0.5$  to include an equal, fair proportion of the two components viz. *preference* and *influence* on a node's importance. The last row of the table shows the correlation value. For all networks excluding Business2001,  $\mathcal{AC}$  identifies the top-ranked nodes most accurately. Also, rankings assigned by  $\mathcal{AC}$  are in better agreement with the ground truth compared to other measures, as indicated by the largest correlation (shown in bold). This is attributed to the inclusion of the relative importance of two components in capturing importance. Hence,  $\mathcal{AC}$  stands out as the most effective performer for capturing importance in weighted directed networks.

### C. Effect of Tuning Parameter

This section demonstrates empirically the effect of the tuning parameter ( $\theta$ ) in capturing the relative importance of incoming versus outgoing ties using the animal network shown in Fig. 2 where edge size reflects the proportional edge weight. We use  $\theta \in \{0, 0.5, 1\}$  to show the role of both components *preference* and *influence* on the node ranks (shown in the Table V). We examine the ego-networks of three nodes M1, M4, M5 (Fig. 3) to study the role of  $\theta$  on a node's importance. In the figures,  $I$  indicates total in-strength and  $O$  indicates total out-strength of the neighbor.

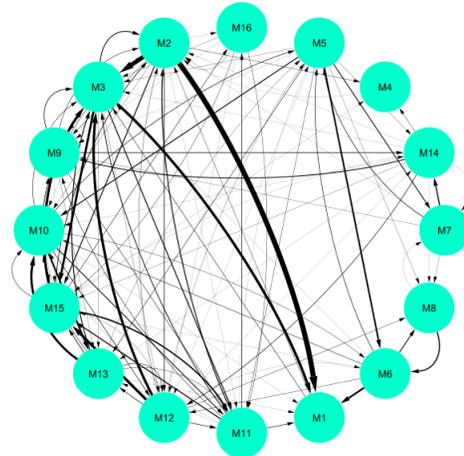


Fig. 2. Moreno-Rhesus Animal Network.

i. Table V shows that node M1 is assigned extreme ranks for  $\theta = 0$  and  $\theta = 1$ , where value 0 or 1 indicates inclusion of a node's *preference* or *influence* component in centrality computation. Consider Fig. 3a for analysis of ranks for different  $\theta$  values. When  $\theta = 0$  is used, measure  $\mathcal{AC}$  assigns a low rank to M1 because of its low influence on its outgoing neighbors M2 and M3 which themselves are less preferred nodes in their neighborhood (low value of  $I$ ). On the other hand,  $\theta = 1$  results in a high rank because of the high preferences of its three neighbors (M2, M3, M6) for node M1 where preferences are in proportion to in-strength of neighbors (Fig. 3a). High proportions of in-strength of three neighbors (M2, M3, M6) indicate their high preferences for the node which improves the rank of M1. Also assigning equal weightage to both components of importance ( $\theta = 0.5$ ) results in a middle rank as both the preference from the neighbors and the influence to neighbors together impact its importance.

<sup>4</sup><http://konect.cc/networks>

TABLE IV. TOP-10 STATES/COUNTRIES RANKED BY THEIR GDP AND THEIR RANKS BY SIX CENTRALITY MEASURES

(A) BUSINESS2001

State	AC	WI	WO	PR	Hub	Authority
Maharashtra	2	2	4	2	3	2
Uttar Pradesh	3	11	1	10	2	16
West Bengal	4	3	7	4	7	3
Andhra Pradesh	9	7	12	5	15	7
Tamil Nadu	12	13	6	9	9	11
Gujarat	1	1	5	1	5	1
Karnataka	8	4	13	3	14	5
Rajasthan	5	17	2	16	1	17
Madhya Pradesh	14	10	11	11	6	9
Punjab	10	16	8	14	12	13
<b>Correlation</b>	<b>0.90</b>	0.73	<b>0.90</b>	0.78	0.85	0.72

(B) BUSINESS2011

State	AC	WI	WO	PR	Hub	Authority
Maharashtra	1	1	4	1	5	2
Uttar Pradesh	3	7	2	7	2	13
Andhra Pradesh	9	6	12	4	11	6
Tamil Nadu	12	13	7	11	9	10
Gujarat	2	2	5	2	4	1
West Bengal	6	3	6	5	6	3
Karnataka	8	4	13	3	8	4
Rajasthan	4	17	1	14	1	16
Madhya Pradesh	11	11	11	8	7	7
Kerala	17	19	14	17	12	20
<b>Correlation</b>	<b>0.91</b>	0.79	0.89	0.82	<b>0.91</b>	0.79

(C) EDUCATION2001

State	AC	WI	WO	PR	Hub	Authority
Maharashtra	1	1	16	1	20	2
Uttar Pradesh	2	4	1	6	1	7
West Bengal	8	6	7	11	8	5
Andhra Pradesh	12	13	4	9	7	14
Tamil Nadu	9	9	9	7	13	13
Gujarat	11	11	5	5	3	12
Karnataka	4	3	13	2	6	3
Rajasthan	14	12	6	13	5	10
Madhya Pradesh	7	5	12	4	10	4
Punjab	15	16	10	12	14	17
<b>Correlation</b>	<b>0.81</b>	0.70	0.78	0.77	0.73	0.65

(D) EDUCATION2011

State	AC	WI	WO	PR	Hub	Authority
Maharashtra	1	1	13	1	16	2
Uttar Pradesh	2	4	1	6	1	5
Andhra Pradesh	10	11	5	7	10	15
Tamil Nadu	8	8	12	4	13	14
Gujarat	13	10	8	5	3	11
West Bengal	11	9	9	14	11	8
Karnataka	4	3	15	2	8	3
Rajasthan	7	7	4	11	5	7
Madhya Pradesh	9	6	7	9	6	4
Kerala	14	22	3	18	4	22
<b>Correlation</b>	<b>0.85</b>	0.75	0.78	0.79	0.75	0.71

(E) WION2004

Country	AC	WI	WO	PR	Hub	Authority
USA	1	1	1	1	3	1
JPN	6	7	5	8	5	7
DEU	2	2	2	2	2	2
GBR	3	4	3	4	6	4
FRA	4	3	4	3	7	3
CHN	5	5	8	5	8	6
ITA	8	6	7	6	9	9
ESP	10	9	13	9	15	10
CAN	7	8	6	7	1	5
KOR	12	13	12	13	11	12
<b>Correlation</b>	<b>0.94</b>	0.91	0.93	0.91	0.92	0.91

(F) WION2014

Country	AC	WI	WO	PR	Hub	Authority
USA	1	1	1	1	3	1
CHN	3	3	3	3	2	2
JPN	6	6	7	9	6	7
DEU	2	2	2	2	4	3
GBR	5	5	6	5	8	8
FRA	4	4	5	4	10	5
BRA	19	18	19	21	16	15
ITA	10	8	10	6	12	11
RUS	13	19	11	16	14	20
IND	23	20	23	23	18	17
<b>Correlation</b>	<b>0.91</b>	0.90	<b>0.91</b>	0.87	0.90	0.90

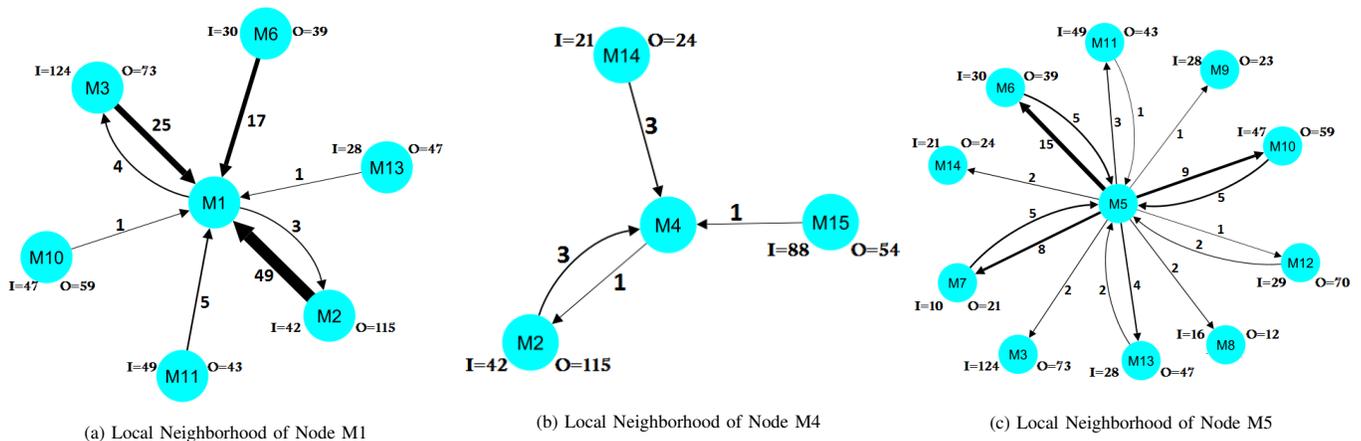


Fig. 3. Directed Weighted Local Networks of Nodes M1, M4 and M5 of Moreno-Rhesus Animal Network, *I*: Total in-Strength of the Node, *O*: Total Out-Strength of the Node.

TABLE V. RANKINGS BY  $\mathcal{AC}$  MEASURE USING  $\theta = \{0, 0.5, 1\}$  FOR MORENO-RHESUS ANIMAL NETWORK

Monkey	$\theta = 0$	$\theta = 0.5$	$\theta = 1$
M1	15	6	2
M2	1	2	4
M3	3	1	1
M4	16	16	16
M5	5	9	12
M6	9	10	8
M7	12	13	15
M8	13	14	13
M9	11	11	7
M10	2	4	6
M11	8	5	5
M12	4	7	9
M13	7	8	10
M14	10	12	11
M15	6	3	3
M16	14	15	14

- ii. In contrast to node M1, the ranking of node M4 remains the same (although low) for all cases as shown in Table V. It is always ranked lowest because of its minimal influence on its out-going neighbor (M2) and low proportion of preferences from its incoming neighbors M2, M14, M15 having high values of I and O (Fig. 3b).
- iii. Node M5 has a large in-degree and out-degree compared to node M1, but with low interaction strength. When  $\theta = 0$  is used, the high influence of node M5 on its influential neighbors with high out-strength rises its rank (Fig. 3c). For  $\theta = 1$ , the node is ranked 12<sup>th</sup>, which is comparatively lower than the rank assigned to node M1. A lower rank is attributed to lower preferences from its neighbors although the value of in-degree is high (low value of in-strength and I). When  $\theta = 0.5$ , node M5 is ranked in the middle due to the cumulative effect of both components on its structural position in its egonet.

#### D. Central Nodes and Community Structure

The objective of this section is to detail the role of central nodes delivered by the proposed measure  $\mathcal{AC}$  on the evolution of communities. Communities provide a good insight into the connection patterns and binding among nodes. We use community detection module *Rbpots* [15] of library *CDLIB*<sup>5</sup> in Python to extract communities from the directed and weighted networks. Extracted communities are plotted using *Paintmaps*<sup>6</sup> where a color scheme is used to differentiate communities based on the interaction strength of underlying nodes. We executed the *Rbpots* module to identify communities in four networks of migration class and two networks of trade class. The plots are shown in Fig. 4. We compared the extracted communities for two different years under the same class to understand their evolution in terms of change in node ranking. We describe below our observations using the top-10 nodes delivered by the measure  $\mathcal{AC}$  (Table VI) for two categories of networks.

- i. **Business-based migration network:** Fig. 4a and 4b show the communities for networks Business2001

and Business2011. All states retained their community membership in 2001 and 2011, except Uttar Pradesh. Three top-ranked states Gujarat, Maharashtra, and Uttar Pradesh in 2001 resulted in a bigger community (red color) whereas Delhi (rank=7) created a community with its neighborhood in the northern Indian region (Table IV (A)). A significant increase in the *preference* term of Uttar Pradesh caused an upgrade in its rank such that it appears in the top-10 states in place of Punjab whose *preference* increased slightly and *influence* decreased quite significantly from 2001 to 2011. This led to the removal of Uttar Pradesh from the red colored community in 2001. This splitting may be attributed to a significant increase in incoming migrants in Uttar Pradesh in 2011 compared to Punjab in the same year.

- ii. **Education-based migration network:** Fig. 4c and 4d show the communities for networks Education2001 and Education2011 where the number of communities remains the same. In 2011, the *preference* of West Bengal decreased significantly while its *influence* increased slightly due to which the pink colored community in 2001 split into two parts. In 2011, the *preference* of Andhra Pradesh increased significantly, and thereby it is placed in the top-10 ranks. This resulted in the merging of the eastern states of the split community with the community formed around Andhra Pradesh.
- iii. **Trade network:** We identified communities for two trade networks WION2004 and WION2014 to understand the role of central nodes in the community formation (Fig. 5a and 5b). Three communities are formed with the only change in membership of Brazil. The rightmost two columns of Table VI show the top-10 ranked countries for both the networks. China (CHN) is ranked higher by the measure for WION2014 network, which is attributed to a significant increase in both its *preference* and *influence* terms from 2004 to 2014. Also, an increase in the trade interactions of Korea with other countries leads to a notable increase in both its *preference* and *influence* and thus it is placed in top-10 ranks delivered by the proposed measure in 2014 (see column WION2014 of Table VI). In the same year, Brazil's interactions with Asian countries like China, India, Japan etc. increased compared to North American countries. The reduced trade with Northern American countries compared to Asian countries not only enhanced its rank, but pulled it out from its earlier community and placed it with communities in Asia (see Fig. 5b).

#### E. Discussion

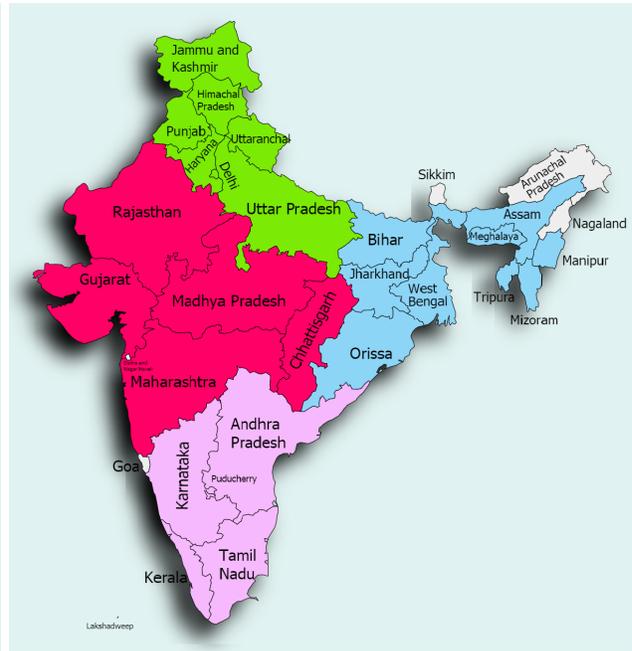
The proposed centrality measure  $\mathcal{AC}$  captures important nodes effectively compared to other centrality measures in Migration and Trade networks. This is attributed to incorporation of both in-strength and out-strength of the node with its 1-hop neighbors by setting tuning parameter  $\theta = 0.5$ . On the other hand, the compared centrality measures either consider in-strength or out-strength in computation which results in loss of information and thereby deteriorates the results.

<sup>5</sup><https://cdlib.readthedocs.io/en/latest/>

<sup>6</sup><https://paintmaps.com>



(a) Business 2001



(b) Business 2011



(c) Education 2001



(d) Education 2011

Fig. 4. Communities Extracted for Four Indian Migration Networks.

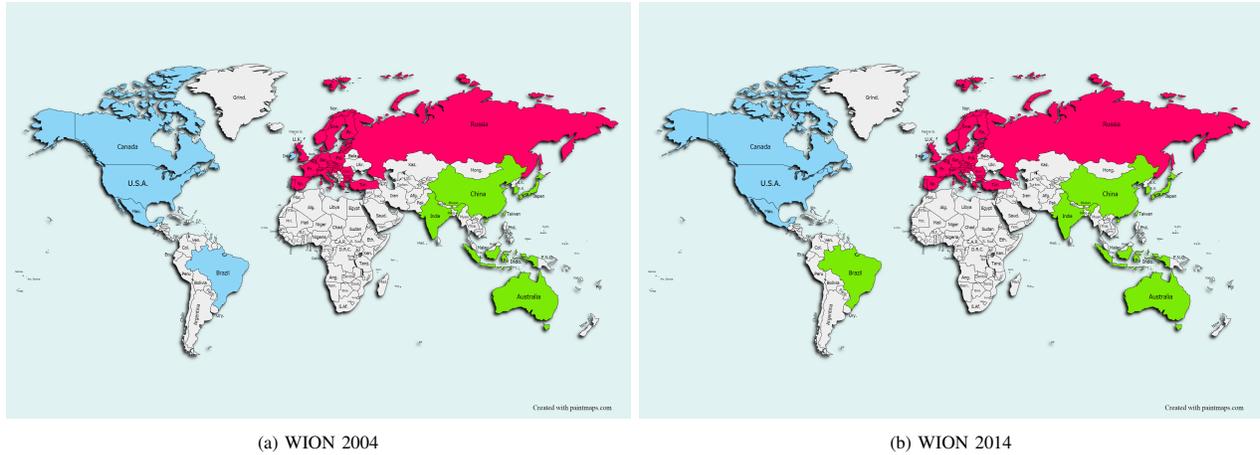


Fig. 5. Communities Extracted for Two Trade Networks: WION2004 and WION2014.

TABLE VI. TOP-10 RANKED NODES BY  $\mathcal{AC}$  MEASURE USING  $\theta = 0.5$  FOR MIGRATION AND TRADE NETWORKS

Migration Networks				Trade Networks	
Business2001	Business2011	Education2001	Education2011	WION2004	WION2014
Gujarat	Maharashtra	Maharashtra	Maharashtra	USA	USA
Maharashtra	Gujarat	Uttar Pradesh	Uttar Pradesh	DEU	DEU
Uttar Pradesh	Uttar Pradesh	Delhi	Delhi	GBR	CHN
West Bengal	Rajasthan	Karnataka	Karnataka	FRA	FRA
Rajasthan	Bihar	Bihar	Bihar	CHN	GBR
Bihar	West Bengal	Jharkhand	Jharkhand	JPN	JPN
Delhi	Delhi	Madhya Pradesh	Rajasthan	CAN	CAN
Karnataka	Karnataka	West Bengal	Tamil Nadu	ITA	NLD
Andhra Pradesh	Andhra Pradesh	Tamil Nadu	Madhya Pradesh	NLD	KOR
Punjab	Haryana	Kerala	Andhra Pradesh	ESP	ITA

Results on Animal network demonstrate that quantum of associations (*Preference* and *Influence*) of a node with its neighbors impacts its centrality in the network. Hence, the inclusion of incoming and outgoing interactions in the computation is application-specific and the centrality measure needs to incorporate them as per the application need. The tuning parameter  $\theta$  of the proposed metric provides this facility. Also, central nodes as determined by  $\mathcal{AC}$  play a vital role in binding their neighbors together in communities.

## V. CONCLUSION AND FUTURE WORK

In this paper, we propose a centrality measure, called *Affinity Centrality* ( $\mathcal{AC}$ ), to determine key nodes in weighted, directed networks. The importance of a node is computed by leveraging its *preference* for and *influence* on its local neighborhood. These components of  $\mathcal{AC}$  are computed by summing proportionate in-strength and out-strength of 1-hop neighbors, respectively. A tuning parameter  $\theta \in \{0, 1\}$  gives flexibility to the end-user to include either the incoming or the outgoing or both types of neighbors depending on the application domain. To verify the effectiveness of  $\mathcal{AC}$ , we used three types of weighted and directed real-world networks - migration, trade, and animal social networks.

An empirical study based on these diverse networks demonstrates the effectiveness and superior performance of  $\mathcal{AC}$  compared to the prevailing centrality measures. As for future work, we expect to further generalize the proposed metric by incorporation of  $l$ -hop neighborhood. However, the present

work can be substantially improved to provide a generalized centrality metric that not only consider strength and direction but also incorporate topological characteristics of the network as well as auxiliary node information to deliver a more robust ranking.

## REFERENCES

- [1] Taras Agryzkov, Leandro Tortosa, José F Vicent, and Richard Wilson. A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science*, 46(4):668–689, 2019.
- [2] Lars Backstrom, Daniel Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume 2006, pages 44–54, January 2006.
- [3] Francis Bloch, Matthew O. Jackson, and Pietro Tebaldi. Centrality measures in networks, 2021.
- [4] Paolo Boldi and Sebastiano Vigna. Axioms for Centrality. *Internet Mathematics*, 10(3-4):222–262, 2014.
- [5] Stephen P. Borgatti and Martin G. Everett. A Graph-theoretic Perspective on Centrality. *Social Networks*, 28(4):466–484, 2006.
- [6] Federica Cerina, Zhen Zhu, Alessandro Chessa, and Massimo Riccaboni. World input-output network. *PLOS ONE*, 10(7):1–21, July 2015.
- [7] Kousik Das, Sovan Samanta, and M. Pal. Study on centrality measures in social networks: a survey. *Social Network Analysis and Mining*, 8:1–11, 2018.
- [8] Easley David and Kleinberg Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, USA, 2010.

- [9] Ricardo Lopes de Andrade, Leandro Chaves Rêgo, Ticiania L. Coelho da Silva, José Antônio F. de Macêdo, and Wellington C.P. Silva. Energy disruptive centrality with an application to criminal network. *Communications in Nonlinear Science and Numerical Simulation*, 99:105834, 2021.
- [10] Linton C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1979.
- [11] Matthew O. Jackson. A survey of network formation models: Stability and efficiency. In Gabrielle Demange and Myrna Wooders, editors, *Group Formation in Economics: Networks, Clubs, and Coalitions*, page 11–57. Cambridge University Press, 2005.
- [12] Sharanjit Kaur, Rakhi Saxena, and Vasudha Bhatnagar. Leveraging hierarchy and community structure for determining influencers in networks. In *19th International Conference on Big Data Analytics and Knowledge Discovery*, volume 10440 of *Lecture Notes in Computer Science*, pages 383–390. Springer, 2017.
- [13] Dong Hwan Kim, Daniel A. Eisenberg, Yeong Han Chun, and Jeryang Park. Network topology and resilience analysis of south korean power grid. *Physica A: Statistical Mechanics and its Applications*, 465:13–24, 2017.
- [14] Jon M. Kleinberg. Hubs, authorities, and communities. *Journal of the ACM*, 46(5):604–632, September 1999.
- [15] E.A. Leicht and M.E.J. Newman. Community structure in directed networks. *Physical Review Letters*, March 2008.
- [16] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.
- [17] Jure Leskovec and Julian McAuley. Learning to discover social circles in ego networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [18] Ying Liu, Ming Tang, Tao Zhou, and Younghae Do. Identify influential spreaders in complex networks, the role of neighborhood. *Physica A: Statistical Mechanics and its Applications*, 452:289–298, June 2016.
- [19] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, July 2010.
- [20] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.
- [21] Simo Puntanen, editor. *Spearman Rank Correlation Coefficient*, pages 502–505. Springer New York, New York, NY, 2008.
- [22] Rakhi Saxena, Sharanjit Kaur, and Vasudha Bhatnagar. Social centrality using network hierarchy and community structure. *Data Mining and Knowledge Discovery*, 32(5):1421–1443, 2018.
- [23] Anuj Singh, Rishi Ranjan Singh, and S. R. S. Iyengar. Node-weighted centrality: a new way of centrality hybridization. *Computational Social Networks*, November 2020.
- [24] Semanur Soyyigit and Çiğdem Boz. Global input - output analysis: A network approach. *Yıldız Social Science Review*, December 2017.
- [25] Min Wang, Wanchun Li, Yuning Guo, Xiaoyan Peng, and Yingxiang Li. Identifying influential spreaders in complex networks based on improved k-shell method. *Physica A: Statistical Mechanics and its Applications*, 554:124229, 2020.
- [26] Yunchuan Wang, Shasha Wang, and Yong Deng. A modified efficiency centrality to identify influential nodes in weighted networks. *Pramana*, 92:1–11, April 2019.
- [27] W. Xing and A. Ghorbani. Weighted pagerank algorithm. In *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004.*, pages 305–314, 2004.
- [28] Seunghyun Yoon, Taejin Ha, Sunghwan Kim, and Hyuk Lim. Scalable traffic sampling using centrality measure on software-defined networks. *IEEE Communications Magazine*, 55(7):43–49, 2017.
- [29] Li Zhai, Xiangbin Yan, and Guojing Zhang. Bi-directional h-index: A new measure of node centrality in weighted and directed networks. *Journal of Informetrics*, 12(1):299–314, 2018.
- [30] Panpan Zhang, Tiandong Wang, and Jun Yan. Pagerank centrality and algorithms for weighted, directed networks with applications to world input-output tables, 2021.