Neural Computing and Applications (2020) 32:10275–10301
https://doi.org/10.1007/s00521-019-04568-0

ORIGINAL ARTICLE

Consensus-based aggregation for identification and ranking of top-*k* influential nodes

Bharat Tidke[1] · Rupa Mehta[1] · Jenish Dhanani[1]

Received: 25 January 2019 / Accepted: 11 October 2019 / Published online: 6 November 2019

Abstract
Technology has continuously been a crucially influenced and acutely tangled with the progress of society. Online Social Networks (OSN) are interesting and valuable datasets that can be leveraged to improve understanding about society and to know inter-personal choices. Identification and Ranking of Influential Nodes (IRIN) is non-trivial task for real time OSN like Twitter which accustom with ever-changing network, demographics and contents having heterogeneous features such as Tweets, Likes, Mentions and Retweets. Existing techniques such as *Centrality Measures* and *Influence Maximization* ignores vital information available on OSN, which are inappropriate for IRIN. Most of these approaches have high computational complexity i.e. $O(n^3)$. This research aims to put forward holistic approach using Heterogeneous Surface Learning Features (HSLF) for IRIN on specific topic and proposes two approaches: Average Consensus Ranking Aggregation and Weighted Average Consensus Ranking Aggregation using HSLF. The effectiveness and efficiency of the proposed approaches are tested and analysed using real world data fetched from Twitter for two topics, Politics and Economy and achieved superior results compared to existing approaches. The empirical analysis validate that the proposed approach is highly scalable with low computational complexity and applicable for large datasets.

Keywords Social network · Influence analysis · Centrality measures · Surface Learning

1 Introduction

Contemporary OSN such as Facebook, LinkedIn and Twitter are quickly grown in popularity [1]. OSN allow nodes (users) to do many social interactions and activities that attracted many researchers to dedicate their study for IRIN in OSN by analysing the online connections and information propagation [2–7]. Moreover, ranking of influential nodes based on different social interactions and activities found to be very useful in various domains such as Social Sciences, Marketing, Management, and Security [8, 9].

Alp and Öğüdücü [7] claimed that 1% of nodes, those are influential have impact on the 25% of the information transmits in the society. Research in computing the influence of a node from OSN data is a notional problem that depends on the application in hand. There is no concrete portrayal for influential node [2, 10]. Consequently, novel influence metrics are continually growing, and most of them agrees on several evaluation criteria [11]. IRIN of Twitter face few challenging issues such as influence of a node is mainly affected and imitated due to structure of network topology, similarly, benchmarks of influential nodes vary for diverse applications [11]. Another challenge lies to select top-k influential nodes from numerous disagreeing list of influential nodes, generated by various techniques [10].

Rank aggregation in OSN play important role specifically in the circumstances where nodes with higher position in manifold network metrics are vital than the nodes having the highest position in one metric [12, 13]. While aggregating multiple features, most of the techniques consider significance of all these features with equal importance for IRIN and result in low accuracy. Wang et al. [13] introduced a multi-attribute ranking method based on entropy, which assign weight to different attributes related to location and neighbourhood of node in the network, by applying K-shell decomposition method. They

✉ Bharat Tidke
ds15co007@coed.svnit.ac.in

[1] Department of Computer Engineering, SVNIT, Surat, India

have shown that their approach ranks influential nodes, with low computational complexity i.e. $O(n)$. The limitation of this approach is that it is applicable for undirected graph which is not suitable for OSN like Twitter.

To address, the problem of computational complexity and significance of features, proposed approaches exploit Heterogeneous Surface Learning Features (HSLF), which consider network based feature i.e. Indegree, user based feature i.e. Likes and the content based features i.e. Mentions, Retweets and Retweet Mention Ratio. These features were used to generate weight based consensus aggregation ranking list of top-k influential nodes to solve the problem of feature significance with low computational complexity, i.e. $O(n)$.

To achieve aggregate rank of influential nodes, two approaches were proposed i.e. Average Consensus Ranking Aggregation (ACRA) and Weighted Average Consensus Ranking Aggregation (WACRA), which finally aggregate unique list of final top-k nodes. The abbreviations used throughout this paper is defined in Table 1.

Table 1 Abbreviations

SN	Social network
OSN	Online Social Network
G	Social network graph
IRIN	Identification and Ranking of Influential Nodes
HSLF	Heterogeneous Surface Learning Features
V	Set of nodes
E	Relationships among nodes
N	Node
D_{in}	Indegree
D_{c}	Degree centrality
C_{c}	Closeness centrality
B_{c}	Betweenness centrality
E_{C}	Eigenvector centrality
RTMN	Retweet and Mention Ratio
ACRA	Average Consensus Ranking Aggregation
WACRA	Weighted Average Consensus Ranking Aggregation
μ	Rank of list of influential nodes of each HSLF
N_{PR}	PageRank
k	Threshold cut off influential nodes
€	Top-k rank list
NI	Node Influence
V_j	Influential score (value) of node
V, V_j	Key-value
BC	Borda Count
MV	Borda Majority Voting
MCC	Matthews Correlation Coefficient
Avg.P	Average Precision
Avg.P@k	Average Precision for top-k nodes
MRR	Mean reciprocal rank
MRR@k	Mean reciprocal rank for top-k nodes

2 Related work

Existing research for IRIN is centred on single network (friends or followers) to compute influence of node [11]. However, in OSN like Twitter, the nodes can form network without having direct friends or follower's relationship. Therefore, multi-feature network depiction of OSN can be leveraged to acutely recognise the links between the nodes based on various features, activities and user interest [7, 14].

2.1 Centrality based approaches

Node influence is mainly calculated and replicated by the network structure of the graph [11, 15, 16]. In information and network science, the popular techniques for IRIN normally apply structural data. The following techniques based on *Centrality Measure* are normally used for IRIN [11, 17, 18].

2.1.1 Degree centrality

It measure the total number of ties a node has with other nodes. The degree centrality is simple approach and can be useful for large and complex networks. The equation for this *Centrality Measure* is as follows, where $D_c(N_k)$ is the degree of (N_k).

$$D_c(N_k) = d(N_k) \tag{1}$$

Mostly, nodes with a higher degree or more ties are more dominant in the network and be likely to have a better capability to influence others.

2.1.2 Closeness centrality

Closeness centrality give emphasis on the distance of a node from all other nodes by concentrating on the geodesic distance. In other words, it measures the magnitude requires to spread information from a specific node to all other nodes. Degree centrality is exceptional instance of Closeness centrality as only ties without geodesic distance are measured. In contrast to Degree, Closeness centrality emphases on the level of node influence over the entire network. In Eq. 2, $C_{\text{c}}(N_k)$ denotes closeness centrality and $d(N_k, N_j)$ denotes distance among two nodes in the network:

$$C_{\text{c}}(N_k) = \sum_{k=1}^{n} \frac{1}{d(N_k, N_j)} \tag{2}$$

2.1.3 Betweenness centrality

Betweenness centrality is the likelihood of node present in between shortest path when two or more nodes communicates with each other in entire network. Betweenness is vital in connecting two community or nodes and can acts a communication controller.

$$B_c(N_k) = \sum_{i,j \neq k} \frac{g_{ikj}}{g_{ij}} \tag{3}$$

In the given Eq. (3), g_{ikj} complete geodesic connecting nodes i and j, which passes from node k, g_{ij} is geodesic distance among node i and j.

2.1.4 Eigenvector centrality

It calculates importance of node as a function of the significance of its nearby nodes. Contrast to Degree centrality, it considers the perception that links to higher centrality or valued nodes are vital than nodes with lower centrality. Let assume, edge between two node i.e.v_i and v_j has weight (w_{ij}) and Z is constant, then the eigenvector centrality for node v_j can be calculated as:

$$E_C(V_j) = \frac{1}{Z} \sum_{v_i \in N(v_j)} w_{ij} \times E_c(v_j) \tag{4}$$

Liu et al. [19] proposed two approaches which modifies existing Degree centrality i.e. Weight Degree Centrality (Wdc) and Extended Weight Degree Centrality (EWdc) to identify influential node based on assortativity of the graph. However, their approach demonstrated high time complexity, which is not unsuitable for large network. Yang et al. [20] proposed a hybrid approach by applying two *Centrality Measure:* Closeness and Betweenness to identify influential opinion leaders in OSN. They improve existing Betweenness centrality by assigning weights, which is calculated using closeness centrality. However, computational complexity of computing both Betweenness and Closeness centrality is n^3. Sheikhahmadi et al. [16] proposed novel two level approach for IRIN based on neighbourhood strength using interactions among the nodes present in social network. They assigned weights to relationships based on interactions and divide network into communities considering those interactions that exist among the users. Zareie et al. [21] presented an entropy based centrality measure to enhance the spreading capability of node by considering degree of nodes and centrality of second order neighbouring node.

2.2 Iterative ranking approaches

Google's PageRank algorithm [22] was mainly exist to rank web pages, but it has been used widely in many applications.

2.2.1 PageRank

PageRank is similar to Eigenvector centrality measure with additional scaling and damping factors. PageRank use the notion of voting which differentiate it from eigenvector centrality. Supposed *df* is a damping factor (by default value is 0.85), the PR of node v_i is set to a predefined value and iteratively computed using following formula:

$$N_{PR}(v_i) = (1 - df) + \left(df \times \sum_{v_j \in N(v_i)} \frac{w_{ji} \times N_{PR}(v_j)}{\sum_{v_k \in N(v_j)} w_{jk}} \right) \tag{5}$$

Recently, Zhang et al. [23] proposed improved *Topical PageRank* which considers the aspects such as location, correlation and year of publication of Scientific Article. They calculated the topical influence of scientific articles using textual data of articles to extract specific topics and its correlations in the field of Scholarly Literature.

2.3 Holistic approaches

The research for IRIN using holistic approaches based on nodes features and contents as an alternative to *Centrality Measures* and iterative algorithms are recently gain popularity [24–26]. Many basic features related to nodes such as Favorites, Listed, Self-Mentions, Number of Followers, Number of Tweets etc. and contents related features such as Retweet, Reply, New Tweet, Topic Similarity, are used for IRIN.

Integrating multiple features is crucial for accomplishment of IRIN [27]. In recent past, techniques like Hierarchical [28], Multilayer network [8], Multi-Criteria Decision Making (MCDM) [29] and Multi-Attribute Integrated Measurement (MAIM) [13] came into existence for IRIN. Du et al. [30] proposed TOPSIS based Multi-Attribute Decision Making (MADM), assess the influence propagation ability of nodes in complex network. They used Degree centrality, Betweenness centrality and Closeness centrality to rank top-k influential nodes. Gandhi and Muruganantham [31] proposed Pugh and TOPSIS based MCDM approach for Facebook data to identify influential nodes. They utilized existing *Centrality Measure* such as Degree, Betweeness, Closeness, Eigenvector and PageRank to identify influential nodes.

Similarly, Zareie and Sheikhahmadi [28] applied TOPSIS to select influential nodes for information spreading and as a set of seeds for influence maximization. They calculated distance between two nodes to confirm that nodes have least overlap and large exposure in network space. The aforementioned limitation of time complexity due to usage of Betweenness and Closeness centrality is also applicable for all this approaches. Hung et al. [32] proposed hybrid MCDM model based on interrelationship and influential weights for online reputation management

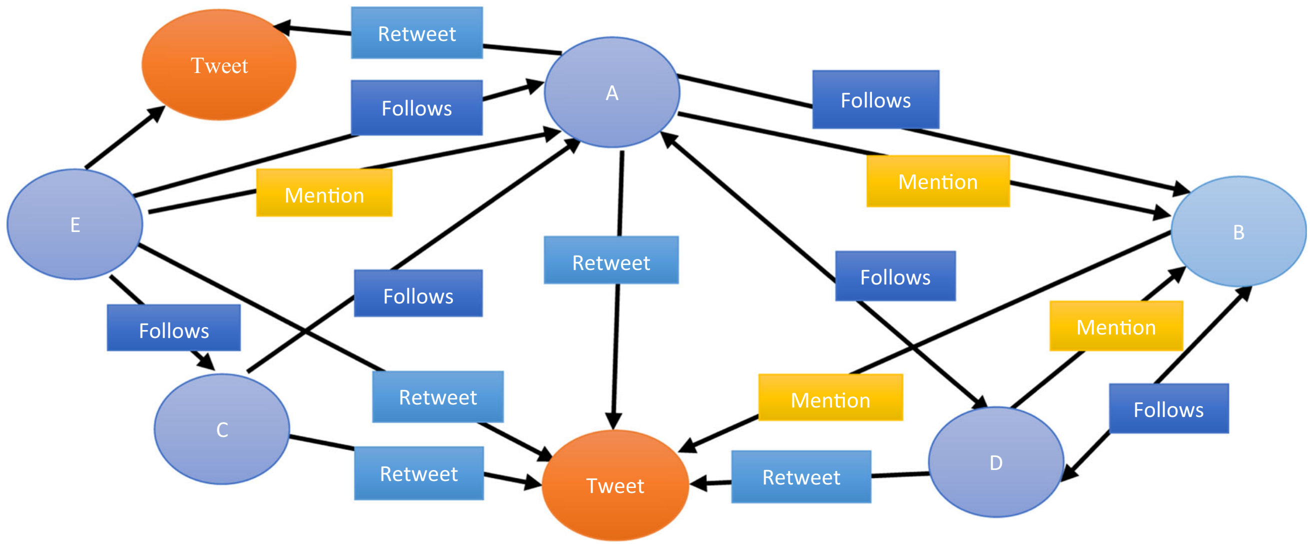


Fig. 1 Example snapshot of Twitter network

to examine the professional services of marketing. Li and Shiu [33] used various features such as nodes Liking, Network Topology, Activities in terms of activeness, interaction and social similarity to proposed influence diffusion model based on AHP (MCDM) for online. Urena et al. [34] proposed consensus model based on *Opinion Dynamics* mechanisms for Group Decision Making by forming social influence network, where individual node comprises one knowledgeable person signified by their views and influence. They showed that consensus approach correctly identified the influential like-minded nodes and remove the nodes with mischievous behaviour to reach the agreement on the decisions.

Sun et al. [35] build a *Machine Learning* approach using various features such as Tweet Count, Followers, Following, Retweet Count etc. to identify influential nodes. They compare their approach with Follower Rank and Weighted Page Rank. However, they have not mentioned how various features integrated or applied any learning model to ensure their evaluation criteria and even standard evaluation measures were not used to check correctness of their results. Basu et al. [36] utilized Degree Centrality, Pagerank, Eigenvector Centrality, Betweenness Centrality, Closeness Centrality as multi-feature criteria to compute influential nodes in Twitter OSN. They suggested a key insight while ranking influential nodes in OSN, simple and well-known approach like Borda Count (BC) performed better in terms of accuracy. They proposed consensus approach to rank influential nodes based on Multi-Features which give stability to the rank orders. The limitation of this approach is the computational complexity of *Centrality Measure*. Zahid and Swart [37] proposed Borda Majority Count which is a special case of a Range Voting, to assign seats in parliament based on various features. The consensus approach used to compute the rank of members based on Borda count and Majority voting. They showed that their approach takes less computational complexity as compare to other aggregation approach as well as provide better accuracy while selecting Member of Parliament. Xiao et al. [38] proposed two metrics i.e. RetweetRank and MentionRank on the hashtag based community to identify influential nodes. They used news as a topic for identification of influential nodes and create two separate set of influential nodes using Retweet and Mention and suggested that RetweetRank identify influential nodes whose tweets on specific topic are more important and attract other nodes. Moreover, MentionRank identify influential nodes who mostly have authority on that topic and compare their approach with Follower rank and PageRank.

The most of the techniques in literature survey either apply *Centrality Measure* [10, 11, 19] or *Iterative Ranking* [23, 39] based on implicit or explicit relation of OSN. Some of them used MCDM [32] and Aggregation [28] techniques with having high complexity. In addition, all features were assumed to have equal important, while aggregating various features. The rank order is also ignored while aggregating for IRIN [40]. To overcome aforementioned limitation next section provides a detailed methodology of proposed approaches.

3 Proposed methodology

The fundamental structure of Twitter Social Network can be signified as a graph $G = (V, E)$, where each individual node in the set V represents a tweet or user. An edge (m, n) in the set E depict certain types of relationships that form due to various interactions and activities between nodes m

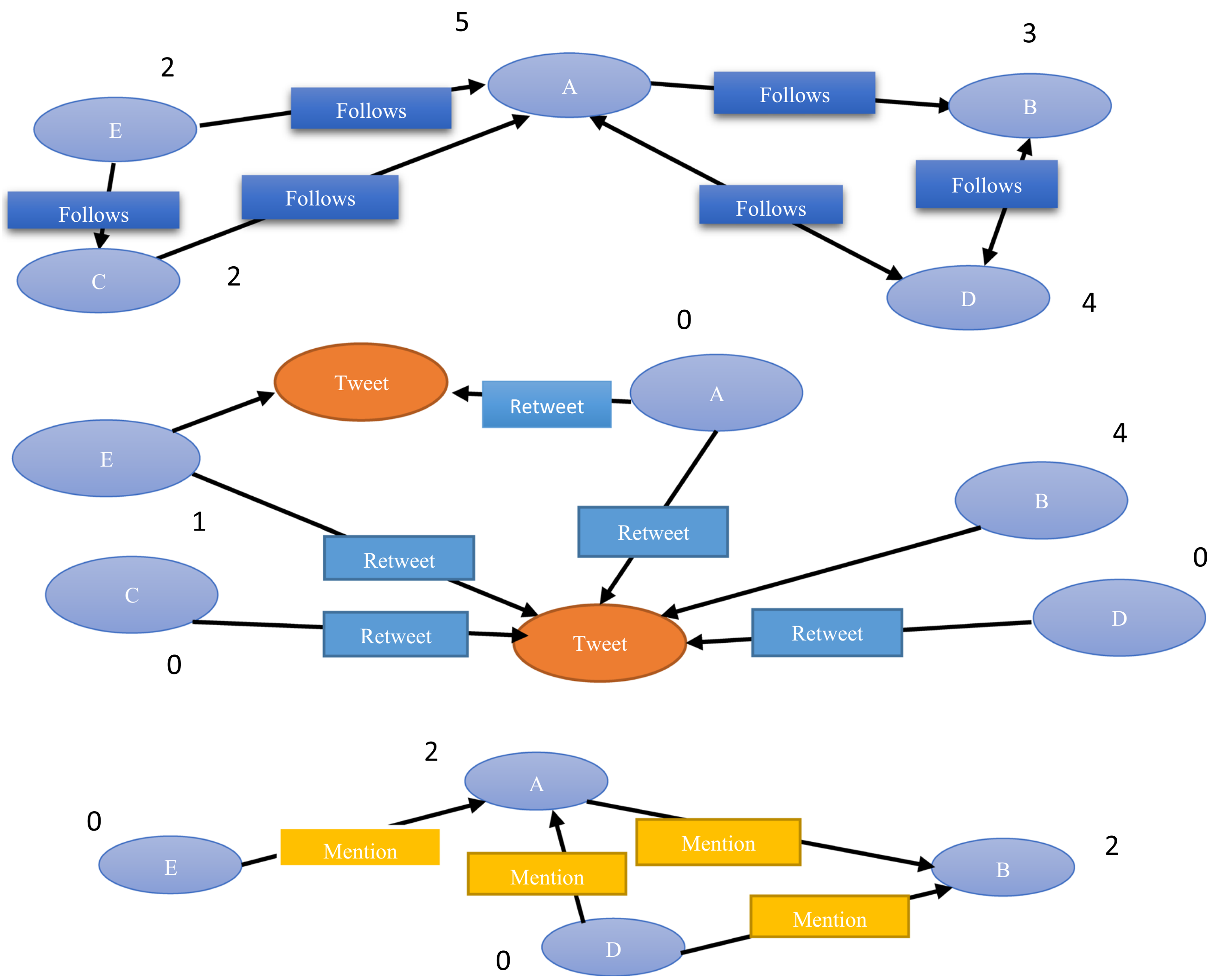


Fig. 2 Example snapshot of various individual relationship in Twitter network

and n (m and n can be user- user, user-tweet and tweet–tweet). In friends-follower's relationship, the Twitter network structure contain edges that are explicitly specified by users. However, in interactions and activities of these nodes, the edges are implicit and need to be inferred.

Figure 1 shows example of Twitter user and tweet network, where node interact with other nodes based on their activities such as posting tweet and even follow nodes whom they find important and influential. For Twitter data, Cha et al. [2] suggested that total number of followers of node A indicates A's popularity, known as indegree influence. Node A can propagate information to their followers by retweeting the tweet posted by some node. In addition, one can mention name (screenname) of other node in the tweet (Fig. 1).

The user A's name included in tweet by other node refer as @mention (@screenname), provides the significance of a A on particular topic. The number of "retweets" to the tweet posted by V indicates ability to disseminate information to other users. Researchers emphasis on inclusion of mention and retweet impact, while computing nodes influence, instead of only indegree as measures of influence [10, 11]. Figure 2, depicts the example influence score on top of each user based on their Indegree, Retweet and Mentions.

3.1 Architecture

The architecture is segregated in four phases: Data Collection, Data Storage, Compute Node Influence and Rank Aggregation for top-k influential nodes (Fig. 3).

3.1.1 Data collection

The experimental analysis of proposed algorithms was carried out using real time Twitter data. Two topics are

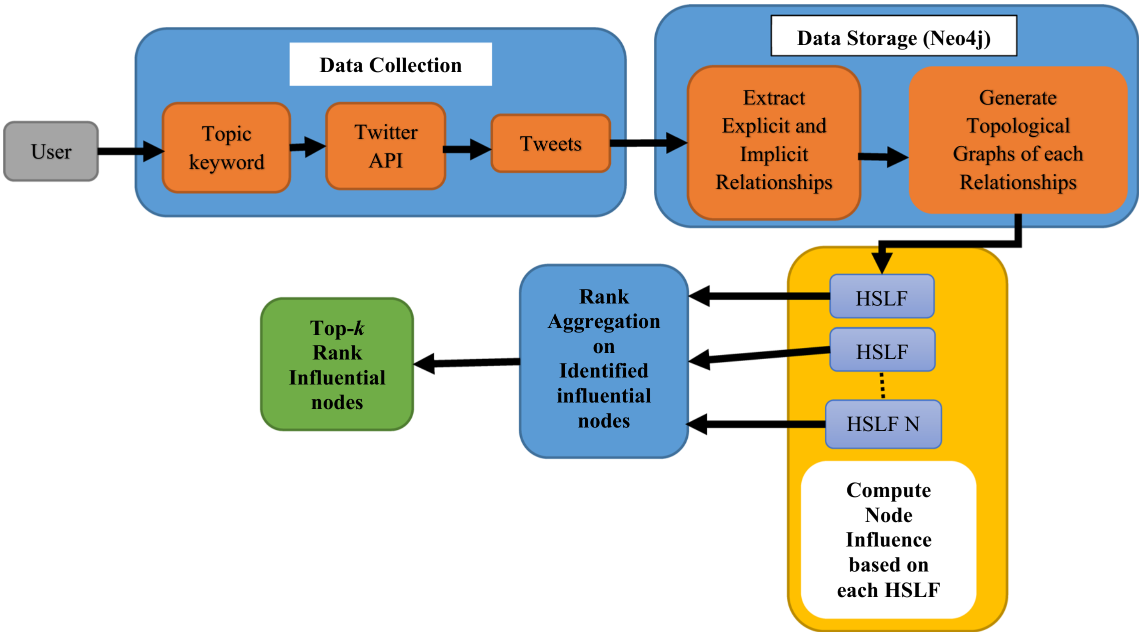


Fig. 3 Proposed architecture to identify and rank top-*k* influential nodes for Twitter data

Table 2 Keyword for fetching Twitter data

Economy	Indian_Economy, Indian_GDP, GST, RBI, Demonetisation
Politics	Indian_Politics, loksabha, Indian_Parliament, Indian_Elections

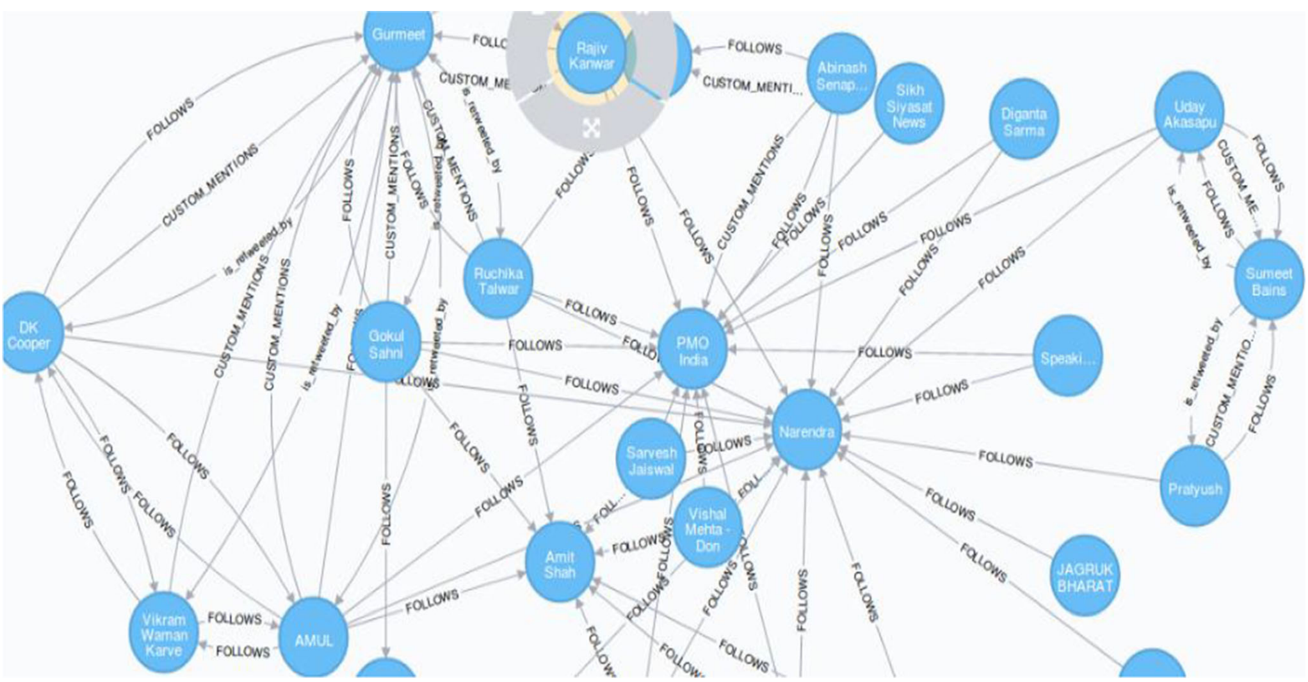


Fig. 4 Example Snapshot of Twitter data in neo4j graph database

selected i.e. Politics and Economy, as they are diverse domain. However, both topics involve few common influential nodes who have authority as well as influence on society. Using Twitter API, tweets for these topics with relevant keywords (Table 2) were collected between September 15, 2018 and November 10, 2018. Further

Table 3 Statistics of nodes and relationships of topics

Nodes	Relationships
Politics	
Tweet: 62,422	Follows: 581,761
User: 41,326	Mentions: 170,332
	Retweets: 32,814
Economy	
Tweet: 49,251	Follows: 898,576
User: 32,729	Mentions: 89,190
	Retweets:58,308

analysis was performed on the collected tweets stored in Neo4j in terms of Nodes, Relationships, Labels and Property. An example of stored graph is depicted in Fig. 4. Table 3 show statistics about the collected data set.

Further, the cypher queries for neo4j are created to generate list of influential nodes based on various HSLF. The brief explanation of various HSLF is given in Sect. 3.1.2.

3.1.2 Heterogeneous surface learning features

- *In-degree (HSLF 1)* The in-degree D_{in} of a node V_i in a social network G is the total number of incoming edges, the node has from other nodes [40].

$$D_{\text{in}}(V_i) = \sum_{v_j \in V_{\text{in}}(v_i)} w_{ji} \tag{6}$$

  where $V_{in}(v_i)$ is the number of nodes which directed to V_i
- *Likes (HSLF 2)* The HSLF likes gives the popularity of node's tweets among other nodes [2].
- *Mention (HSLF 3)* The HSLF Mention effectiveness used to propagate information to remote node rather than the neighbourhood only and give importance to the proper set of nodes that have authority on the topic or relate to it [26].
- *Retweet (HSLF 4)* The HSLF retweet effectiveness give the information on importance of node contents on topic which will be propagate by other nodes. In other words, the retweet count gives the ability of a node to produce useful information that is propagated to other node. A node can send information to her followers by retweeting posts of other nodes [26].
- *Retweet and Mention Ratio [RTMN] (HSLF 5)* The RTMN is the ratio of sum of total number of retweet (tnr) and total number of mention (tnm) obtained by a node on particular topic to the total number of tweets (ntt) he has posted on that topic [6].

$$\text{RTMN} = (\text{tnr} + \text{tnm})/\text{ntt} \tag{7}$$

After generating various list of influential nodes using HSLF, next step is to aggregate these list to generate unique single list of influential nodes based on proposed approaches.

3.1.3 Rank aggregation

Ranking aggregation technique is one of the well-known method which aggregate ranked list generated by various features [12]. The aggregation ranking can be divided into three categories i.e. Complete ranking, Fractional ranking and Top- k ranking [41]. Complete ranking assign rank to all nodes present in the network. Fractional ranking assign rank to the fraction of all the nodes, for example 1/4, 1/2 or 1/3 of nodes. The top- k ranking assign rank to the nodes based on threshold (k) given by the user. For example, if there are 100 nodes present in the network, if user wants top 5 influential nodes, then the value of k is 5. The influential nodes list generated by various feature consider only k value while generating the final list. In this paper, proposed approaches mainly focused on top-k node ranking generated from various HSLF list to generate a consensus unique single rank list.

The tradition rank aggregation algorithms such as position based Borda Count (BC) [36] and distance based Kemeny Aggregation assume that the consensus always be applied on same set of nodes (order will be different). Moreover, the Kemeny aggregation is NP hard [41] and straightforward usage of the BC method is not possible, since, in our proposed approaches HSLF generate different sets of influential nodes. For example, node X might be present in two HSLF generated list, but not in other HSLF list. Therefore, traditional rank aggregation method we cannot apply directly. To overcome the challenge, we proposed average consensus based rank aggregation (ACRA) based on key-value pair, where nodes is the key and their respective position in the list are their values.

Let V be a fixed set of nodes; we can assume for simplification, that $V = \{n1, n2, \ldots, ni\}$, where ni indicates the cardinality of V. A ranking of node influence over V is an ordered list: $\mu = (n1 > n2 > \cdots > ni)$, where $ni \in V$. As discussed earlier, proposed approaches are based on top-k ranking criteria for consensus aggregation, therefore, only top-k rank nodes which fulfils the criteria for each HSLF will be present in the top-k list denoted by μ.

Definition 1 For a given node $i \in V$ present in μ, $\mu(i)$ represents the rank of i in μ. Consider each HSLF gives their ordered list of n nodes, then, we have multiple list that can be represented as multiset rankings $\{\mu_{i1} \quad \mu_{i2} \quad \ldots \quad \mu_{in}\}$. A consensus rank-position aggregation (RPA) of μ_i is a ranking € on the set of list of top-k nodes generated by HSLF as compare to complete rankings aggregation of all nodes N, that minimizes the size of list by $\Delta(\text{€},N)$ where Δ is the subtraction of € from N and we denote the set RPA of N by $k(N)$.

Consider a scenario, where proposed HSLF features generates m ranking of n nodes to compute Node Influence (NI) as shown in Eq. (1):

$$\mathrm{NI} = \begin{array}{c} \\ \mathrm{HSLF}_1 \\ \mathrm{HSLF}_2 \\ \vdots \\ \mathrm{HSLF}_m \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} \\ \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & \mu_{mn} \end{bmatrix} \end{array} \tag{8}$$

The ith row of NI denotes the rank list generated by HSLF_i i.e. $\mu_i = [\mu_{i1} \quad \mu_{i2} \quad \ldots \quad \mu_{in}]$ and $\mu_{ij} = (1 \le \mu_{ij} \le n)$ represent the place of nodes in the ordered list. Now, we have the dataset of NI i.e. $D_{\mathrm{NI}} = (\mathrm{HSLF}_1, \ \mathrm{HSLF2}, \ldots, \mathrm{HSLF}_m)$ with various permutation of node ranking. The main objective is to generate consensus based rank aggregation of k nodes that best signifies the dataset for IRIN. To achieve these, two algorithms ACRA (Algorithm 1) and WACRA (Algorithm 2).

- ***Algorithm* 1. ACRA**

Input: Set of V nodes and tweets with their stats (Number of tweets) and relationships as a graph $G = (V, E)$ edge (m, n), and value of *k.*

Output: Top-*k* nodes using **ACRA**

Let, *n:* no. of nodes

1. **For** $1 \ to \ n$
   - Compute and Sort $HSLF_1$ (Indegree) using $equation$ 6.
   - Compute and Sort $HSLF_2$ (Likes) score
   - Compute and Sort $HSLF_3$ (Mention) score
   - Compute and Sort $HSLF_4$ (Retweet) score
   - Compute and Sort $HSLF_5$ (RTMN) score
2. **For** $input \ k$
   - Extract the top-*k* nodes from each HSLF generated list
3. **Create *NI* matrix using** $HSLF_1$ to $HSLF_m$ as $m * k$ matrix using $equation$ (8)

   For $i = 1 \ to \ m$

   For $j = 1 \ to \ k$

   - Assigned **value** to each node based on its **Position Rank** in the list

     $$Vj = k - j + 1 \tag{9}$$

     where, Vj= value of j^{th} node
   - Generate a key–value pair $< V, Vj >$

   where, key represents node and value represents its computed score for each HSLF

   For End

   For End
4. **Compute an ACRA**
   - Let P denotes the list of the aggregated weight as score for nodes having same key
   - Calculate ACRA using $equation$ 10:

     $$ACRA = \underset{\forall z \in P}{Generate} < Z_{key}, \frac{P_{score}}{m} > \tag{10}$$

   Now, Sort the ACRA in descending order based on score and fetch top-k IRIN
5. **End**

- ***Algorithm* 2. WACRA**

Input: *Step* 3 *of Algorithm* 1

Output: Top-*k* nodes using WACRA

1. **Assign weights to each node based on its threshold criteria (relevant to topical influential nodes)**

   For $i = 1\ to\ m$

   For $j\ = 1\ to\ k$

   - $HSLF_1 = 0.8 * Vj$
   - $HSLF_2 = 0.85 * Vj$
   - $HSLF_3 = 0.9 * Vj$
   - $HSLF_4 = 0.95 * Vj$
   - $HSLF_5 = Vj$
     - Generate a key–value pair $< V, Vj\ >$
     - Where key represents node and value represents its computed score

   For End

   For End

2. **Compute WACRA**
   - Let P *denotes* the list of the aggregated weights as score for pairs having same key
   - Calculate WACRA using equation (10) with assigned weights
   - Now, Sort the WACRA generated list in descending order based on score and fetch top-*k* nodes
   - **End**

To get understandings of proposed ACRA and WACRA, consider an example where 11 nodes are present (Table 4). Five HLSF were used to generate five list.

3.1.4 ACRA example

1. Step 1 and 2.
   - User input is *k = 5* and HSLF= *5* and each generates their own sorted list of rank.
2. Step 3.
   - Computing the value for above matrix using Eq. 9 in algorithm 1 for given nodes.
   - Assigned value to each node based on its Position Rank in the list.
3. Step 4.

Table 4 Example ranking list generated by $HSLF_m$

Features	1	2	3	4	5
HSLF1	A	C	E	G	H
HSLF2	D	E	C	I	J
HSLF3	D	B	A	I	C
HSLF4	A	G	F	I	C
HSLF5	K	G	A	F	C

- Compute total value of each key

$$P = \begin{bmatrix} <A,(5+3+5+3)>, <B,(4)>, <C,(4+3+1+1+1)>, \\ <D,(5+5)>, <E,(3+4)>, <F,(3+2)>, \\ <G,(2+4+4)>, <H,(1)>, <I,(2+2+2)>, \\ <J,(1)> <K,(5)> \end{bmatrix}$$

$$P = \begin{bmatrix} <A,(16)>, <B,(4)>, <C,(10)>, \\ <D,(10)>, <E,(7)>, <F,(5)>, \\ <G,(10)>, <H,(1)>, <I,(6)>, \\ <J,(1)> <K,(5)> \end{bmatrix}.$$

- Calculate ACRA

$$\text{ACRA} = \begin{bmatrix} <A,\left(\frac{16}{5}\right)>, <B,\left(\frac{4}{5}\right)>, <C,\left(\frac{10}{5}\right)>, \\ <D,\left(\frac{10}{5}\right)>, <E,\left(\frac{7}{5}\right)>, <F,\left(\frac{5}{5}\right)>, \\ <G,\left(\frac{10}{5}\right)>, <H,\left(\frac{1}{5}\right)>, <I,\left(\frac{6}{5}\right)>, \\ <J,\left(\frac{1}{5}\right)> <K,\left(\frac{5}{5}\right)> \end{bmatrix}$$

$$\mathrm{ACRA} = \begin{bmatrix} <A,(3.2)>, <B,(0.8)>, <C,(2)>, \\ <D,(2)>, <E,(1.4)>, <F,(1)>, \\ <G,(2)>, <H,(0.2)>, <I,(1.2)>, \\ <J,(0.2)>, <K,(1)> \end{bmatrix}$$

- Sort the ACRA in descending order based on score and fetch top-k IRIN

$$\mathrm{ACRA} = \begin{bmatrix} <A,(3.2)>, <C,(2)>, <D,(2)>, \\ <G,(2)>, <E,(1.4)>, <I,(1.2)>, \\ <F,(1)>, <K,(1)>, <B,(0.8)>, \\ <H,(0.2)>, <J,(0.2)> \end{bmatrix}$$

- Final output of the ACRA for Top-5 IRIN $= \{A, C, D, G, E\}$

  In ACRA, the nodes C, D, G have equal score, still D and G are places below C respectively. It also depends upon the HSLF features input or if sorted according to name. For example, if HSLF1 node is influential and have same score as other influential node of HSLF 2 then HSLF 1 node is placed above HSLF 2 or irrespective of HSLF sequence, once list is generated the node with same score is sorted according to their name. Such scenario is naïve and does not provide effective ranking of influential node. To address, the problem of nodes with equal weights as well as the priority of HSLF, proposed WACRA perform better. To get understandings of proposed WACRA approach, consider the same example as shown in the Table 5.

Table 5 Score assigned to each node for ACRA

Position	1	2	3	4	5
HSLF1	A	C	E	G	H
	Score = 5	Score = 4	Score = 3	Score = 2	Score = 1
	<A,5>	<C,4>	<E,3>	<G,2>	<H,1>
HSLF2	D	E	C	I	J
	Score = 5	Score = 4	Score = 3	Score = 2	Score = 1
	<D,5>	<E,4>	<C,3>	<I,2>	<J,1>
HSLF3	D	B	A	I	C
	Score = 5	Score = 4	Score = 3	Score = 2	Score = 1
	<D,5>	<B,4>	<A,3>	<I,2>	<C,1>
HSLF4	A	G	F	I	C
	Score = 5	Score = 4	Score = 3	Score = 2	Score = 1
	<A,5>	<G,4>	<F,3>	<I,2>	<C,1>
HSLF5	K	G	A	F	C
	Score = 5	Score = 4	Score = 3	Score = 2	Score = 1
	<K,5>	<G,4>	<A,3>	<F,2>	<C,1>

3.1.5 WACRA example

1. Step 3.

   Computing the value for above matrix using Eq. (9) in Algorithm 1, for given nodes. Table 6 show the new computed value based on weights.
2. Step 4.

$$P = \begin{bmatrix} <A,(3.75+2.55+4.5+2.85)>, <B,(3.4)>, \\ <C,(3+2.4+0.85+0.9+0.95)>, <D,(4+4.5)>, \\ <E,(2.25+3.2)>, <F,(2.7+1.9)>, <G,(1.5+3.6+3.8)>, \\ <H,(0.75)>, <I,(1.6+1.7+1.8)>, <J,(0.8)> <K,(4.)> \end{bmatrix},$$

$$P = \begin{bmatrix} <A,(13.65)>, \\ <B,(3.4)>, <C,(8.1)>, \\ <D,(8.5)>, <E,(5.45)>, \\ <F,(4.6)>, <G,(8.9)>, \\ <H,(0.75)>, <I,(5.1)>, \\ <J,(0.8)> <K,(4.75)> \end{bmatrix}$$

$$\mathrm{WACRA} = \begin{bmatrix} <A,\left(\frac{13.65}{5}\right)>, <B,\left(\frac{3.4}{5}\right)>, \\ <C,\left(\frac{8.1}{5}\right)>, <D,\left(\frac{8.5}{5}\right)>, \\ <E,\left(\frac{5.45}{5}\right)>, <F,\left(\frac{4.6}{5}\right)>, \\ <G,\left(\frac{8.9}{5}\right)>, <H,\left(\frac{0.75}{5}\right)>, \\ <I,\left(\frac{5.1}{5}\right)>, <J,\left(\frac{0.8}{5}\right)>, <K,\left(\frac{4.75}{5}\right)> \end{bmatrix}$$

$$\mathrm{WACRA} = \begin{bmatrix} <A,(2.73)>, <B,(0.68)>, \\ <C,(1.62)>, <D,(1.7)>, \\ <E,(1.09)>, <F,(0.92)>, \\ <G,(1.78)>, <H,(0.15)>, \\ <I,(1.02)>, <J,(0.16)>, <K,(0.95)> \end{bmatrix}$$

$$\mathrm{ACRA} = \begin{bmatrix} <A,(2.73)>, <G,(1.78)>, \\ <D,(1.7)>, <C,(1.62)> \\ , <E,(1.09)>, <I,(1.02)>, \\ <K,(0.95)>, F,(0.92)>, \\ <B,(0.68)>, <J,(0.2)>, <H,(0.15)> \end{bmatrix}$$

- Final output of the WACRA for Top-5 IRIN $= \{A, G, D, C, E\}$

The problem of nodes having same score have been eliminated in WACRA. In addition, the rank of nodes G is place upper as compare to C in WACRA. To verify our theoretical analysis, we perform several experiments and evaluate results with well-known measures in experiments section. The limitation of proposed approach WACRA, is that the weight assigned to each HSLF is based on intuitions as well as theory suggested in [2, 25, 26]. Comprehensive and unbiased evaluation of WACRA is still needed while assigning weights to different HSLF.

Table 6 Score assigned to each node for WACRA

Position	1	2	3	4	5
HSLF1	A Score = 3.75 <A,3.75>	C Score = 3 < C,3>	E Score = 2.25 <E,2.25>	G Score = 1.5 <G,1.5>	H Score = 0.75 <H,1>
HSLF2	D Score = 4 <D,4>	E Score = 3.2 <E,3.2>	C Score = 2.4 <C,2.4>	I Score = 1.6 <I,1.6>	J Score = 0.8 <J,0.8>
HSLF3	D Score = 4.25 <D,4.25>	B Score = 3.4 <B,3.4>	A Score = 2.55 <A,2.55>	I Score = 1.7 <I,1.7>	C Score = 0.85 <C,0.85>
HSLF4	A Score = 4.5 <A,4.5>	G Score = 3.6 <G,3.6>	F Score = 2.7 <F,2.7>	I Score = 1.8 <I,1.8>	C Score = 0.9 <C,0.9>
HSLF5	K Score = 4.75 <K,4.75>	G Score = 3.8 <G,3.8>	A Score = 2.85 <A,2.85>	F Score = 1.9 <F,1.9>	C Score = 0.95 <C,0.95>

3.1.6 Computational complexity analysis

Consider each HSLF$_m$ metrics in terms of their computational complexity. As HSLF counts only the numbers which gives linear complexity of $O(N)$ in case of number of Tweets, Retweets, Like, Mention And Followers.

Let, U: size of unique nodes in the network, A is individual node then,

- Number of A's tweets:$O(N)$
- Number of A's followers: $O(N)$
- Number of nodes retweeted A's tweets: $O(N \cdot U)$
- Number of A's tweets liked by other nodes. $O(N)$
- Number of mention A's gets from other nodes. $O(N \cdot U)$

$$\mathrm{RTMN} = \frac{O(N \cdot U) + O(N \cdot U)}{O(N)}$$

Since, the k is constant for nodes, the computational complexity of HSLF$_m$ is linear.

Now, proposed algorithm complexity is analyze: first ACRA; step 1 is just the user input and calculates HSLF$_m$. Now in step 3, the matrix is calculate $n * m$, where n is number of nodes and m is the number of HSLF. Therefore, the computational complexity is $O(n * m)$, Since m is always very small constant (in our case its 5) which gives the complexity of $O(N)$. In addition, proposed approaches applied on top-k nodes of each features, therefore, the complexity is again reduce i.e. $O\left(\frac{N}{n-k}\right)$. Step 4 count the score of each n, which again is $O\left(\frac{N}{n-k}\right)$ Therefore, the final complexity of ACRA is $O\left(\frac{N}{n-k}\right) + O\left(\frac{N}{n-k}\right) = O\left(\frac{N}{n-k}\right)$. Similarly, the computational complexity of WACRA is $O\left(\frac{N}{n-k}\right)$ or $O(k)$ where k is the value of top-k given by user same as ACRA, as it just assigns weight instead of position value.

4 Experiments

In this section, several experiments were performed to validate proposed approaches with two real Twitter datasets.

4.1 Experimental settings

All relationships in terms of graphs are stored in Neo4j graph database. Algorithms were implemented in python with py2neo library support. Experiments were run on a PC with Intel(r) core(TM) i7-7700 cpu @ 3.60 ghz dual CPU and 16 GB RAM with OS Ubuntu 16.04.5 LTS (XenialXerus).

4.2 Result analysis

For IRIN, different value of k influential node i.e. k = 25, 20, 15, 10, 5 were set beforehand. We consider three different influential measures i.e. Number of Followers [40], Retweet rank [40, 43] and Relevance Right [44] list as a benchmark list of influential nodes. These lists can act as evaluation list to check the efficiency of the proposed approaches and state-of-art approaches for top-k influential node. The results of proposed approaches i.e. ACRA (Algorithm 1) and WCRA (Algorithm 2) are compared with baseline approaches i.e. (Degree Rank, Mention Rank & Retweet Rank. [38], Borda Count (BC) [36] and Borda Majority Voting (MV) [37].

4.2.1 Identification of top-*k* influential nodes

The experimental validation for identification of influential node is carried out using three different measures i.e.

Table 7 Confusion matrix

	Predicted as positive	Predicted as negative
Actually positive	True positives (TP)	False negatives (FN)
Actually negative	False positive	True negative

Accuracy (ACC), F1-measures and Matthews Correlation Coefficient (MCC) [40] which is based on different values of Table 7.

- *Accuracy* Accuracy is mainly sensitive evaluation measure and it is basically a ratio of accurately predicted results to the total results. It evaluates the complete usefulness of the model by calculating the likelihood of the correctness for predicted output. In addition, error rate in accuracy measure is the probability of incorrect output given by prediction model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

- *F-measure* F-measure is a well-known evaluation measure for binary classification problem. Therefore, it fits the criteria for evaluating our proposed influence analysis methodology which is based on binary detection of influential nodes in a given network of well-known influential node. F-measure is a fusion of precision and recall, which are efficient metrics for influential node identification where the imbalance problem exists. The F-measure is the harmonic mean of precision and recall, i.e.

$$F\text{-measure} = \frac{2 * P * R}{(P + R)} \tag{12}$$

where $\text{precision}(P) = \frac{\text{TP}}{(\text{TP+FP})}$ and $\text{recall}(R) = \frac{\text{TP}}{(\text{TP+FN})}$.

- *Matthews Correlation Coefficient (MCC)* This measure is frequently used in binary classification problem for imbalanced dataset as it takes mutual accuracies and error rates on both values (True and False), and considers all classes of confusion matrix. MCC gives value ranges between 1 for best case output to − 1 for the worst case output. If prediction models perform arbitrarily the MCC gives value near to 0.

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{13}$$

4.2.2 Performance comparisons on politics data

Politics data from Twitter in the context of Indian user were extracted using various keywords (Table 3).

4.2.2.1 Followers The follower metric is the count of total follower of node on Twitter, irrespective of topic.

Figure 5 depict that, for Politics dataset, BC perform better in term of accuracy as compare to all other approaches when evaluate with Twitter Followers count. The BC [36] is the aggregation of four measures i.e. Degree centrality, Closeness centrality, Eigenvector centrality and PageRank based on friend's-follower relationship which offer more inclination towards indegree based measure and makes it more suitable for follower count evaluation. However, proposed ACRA and WACRA perform better as compare to other comparative approaches.

Similarly, Fig. 6, show that for Politics dataset BC perform better in term of F1 score. F1-measure do not consider TN value of confusion matrix which makes the heuristic approaches like Degree rank with very low and Retweet rank with negligible score.

However, proposed approaches gave better F1 score compare to Mention rank and MV. For MCC evaluation measure the trends of the performance remains almost similar for all approaches as Accuracy and F1-measure, where BC outperformss all other approaches (Fig. 7).

4.2.2.2 Retweet online Retweet Online is a popular and prevailing influence evaluation technique on Twitter. A node presents on Twitter can trace their rank on Retweet Online website. Retweet Online Rank considers number of followers, friends, latest retweets, and lists of a node. These

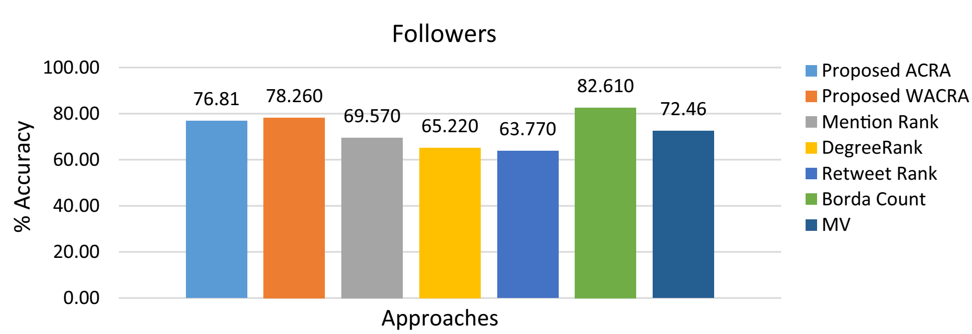


Fig. 5 Comparison among approaches in terms of accuracy to identify top-k influential nodes for politics dataset matched with Followers

Fig. 6 Comparison among approaches in terms of F1-measure to identify top-k influential nodes for politics dataset matched with Followers

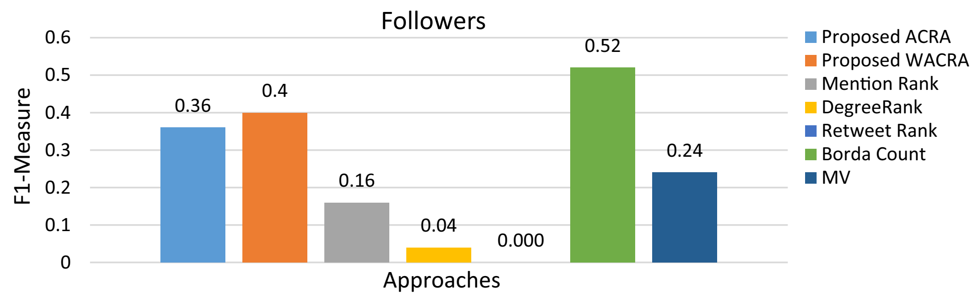
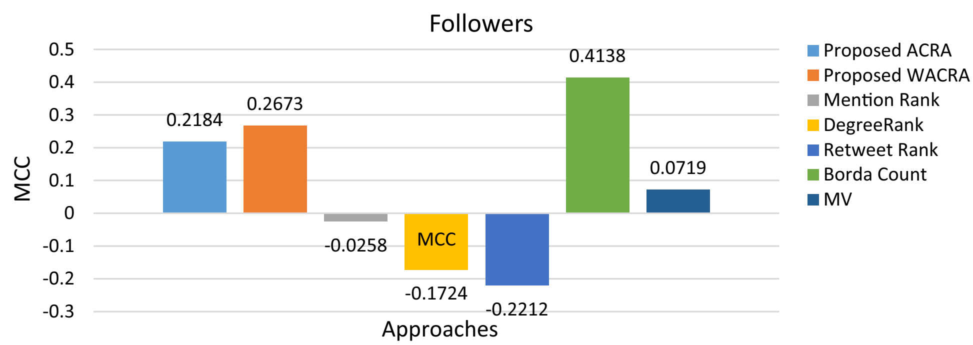


Fig. 7 Comparison among approaches in terms of MCC to identify top-k influential nodes for politics dataset matched with Followers



features are taken as standards while assigning rank. For example, if node has rank 2 means he is the second most influential node on Twitter according to Retweet Online.

Influential nodes present in Retweet Online Rank list (Table 8) for Politics topics may belongs to different domain, for example SushmaSwaraj (Politics), Akshaykumar (Bollywood) imVkohli (Cricket). These nodes are either very popular or celebrity.

Therefore, nodes with many followers and list score have upper hand as compare to nodes with less friend and low list score. It is possible that some nodes may be having more or almost same number of retweets.

Figures 8, 9, and 10 imitates the intuition, as BC performance better in term of accuracy, F1 measure and MCC as compare to all other approaches when compare with Retweet Online. However, proposed ACRA and WACRA perform better as compare to other baseline approaches.

4.2.2.3 Relevance right Relevance Right authority rank combines topic relevancy and peer recognition of node based on their contents (topical) popularity on web.

Relevance Right is better evaluation measure for topic related influential nodes as compare to Followers count and Retweet rank. Both emphasis on importance of influential nodes content and activities rather than merely popularity. The nodes that are irrelevant to the topic Politics or they are not expert of Politics will be less influential as compare to the nodes with more relevancy and authority in Politics. Figure 11 show that proposed approach ACRA and WACRA outperforms other baseline approaches with 82.61% accuracy, as both, used various topic relevancy features such as Retweets, Mention, Likes and Retweet Mention Ratio for IRIN. Though, BC aggregate various centrality and PageRank to measure influence, but mainly focuses on friend's–follower relationships, which makes them less effective while identifying topical influential or authority nodes. Moreover, heuristic approaches such as Mention Degree and Retweet rank considers only one feature for IRIN, ignoring other important features, lead to poor performance when evaluates with Relevance Right. As, MV fails to identify same number of influential nodes as compare to its counterparts as shown in Tables 8 and 9. While evaluating results, same number of candidates in other approaches were taken into consideration for accuracy, F1-measure and MCC which result in average performance of MV (Figs. 12, 13).

4.2.3 Performance comparisons on economy data

Economy data from Twitter in the context of Indian user were extracted using various keywords as depicted in Table 3. The topic Politics is somewhat generalized topic for masses and consist of influential nodes from various domains, posting tweets as opinions to show their agreement and disagreement on the topic. So, approaches utilizing Friends-Followers features dictates the accuracy as seen with BC in terms of Followers count as well as Retweet Online. However, the Economy topic needs expertise or authority for posting tweets on Twitter as compare to Politics. Therefore, the content related features like mentions retweets becomes vital while identifying influential nodes for Economy dataset.

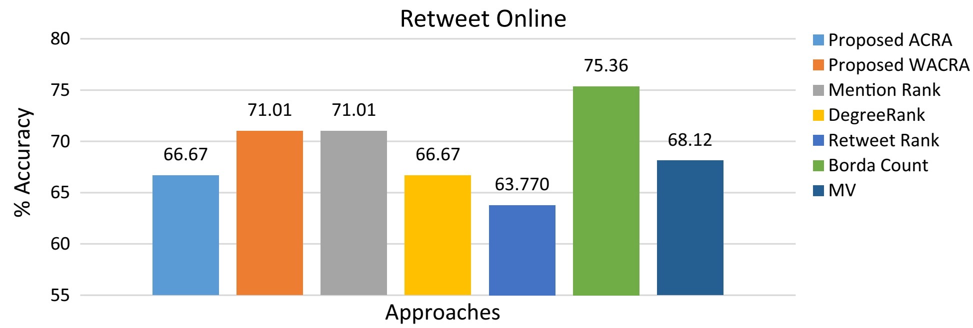


Fig. 8 Comparison among approaches in terms of accuracy to identify top-k influential nodes for politics dataset matched with Retweet Online

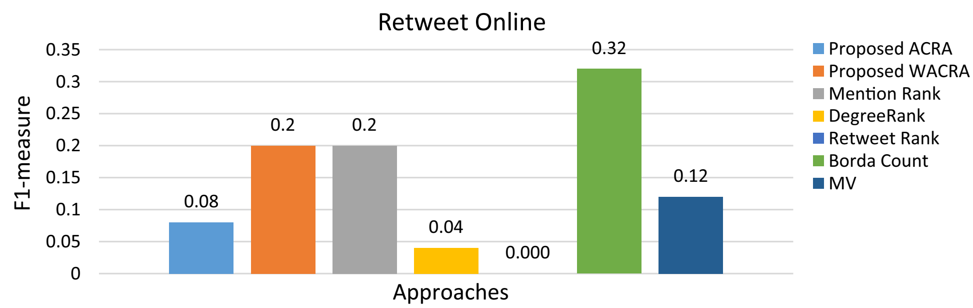


Fig. 9 Comparison among approaches in terms of F1-measure to identify top-k influential nodes for politics dataset matched with Retweet Online

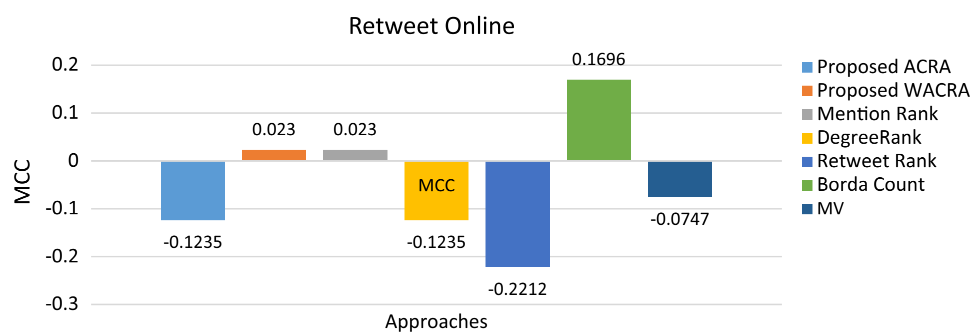


Fig. 10 Comparison Among approaches in terms of MCC to identify top-k influential nodes for politics dataset matched with Retweet Online

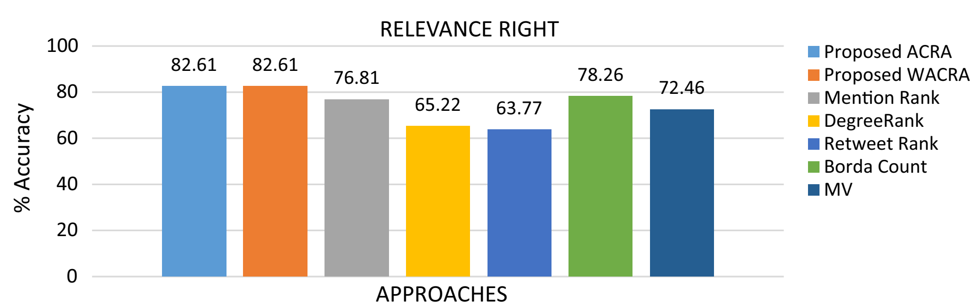


Fig. 11 Comparison among approaches in terms of accuracy to identify top-k influential nodes for politics dataset matched with Relevance Right

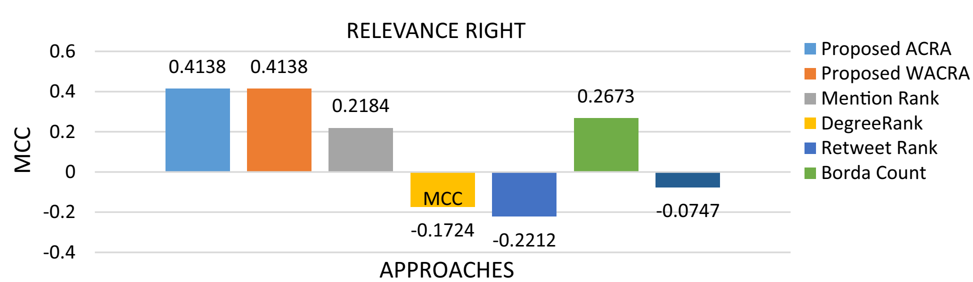


Fig. 12 Comparison among approaches in terms of F1-measures to identify top-k influential nodes for politics dataset matched with Relevance Right

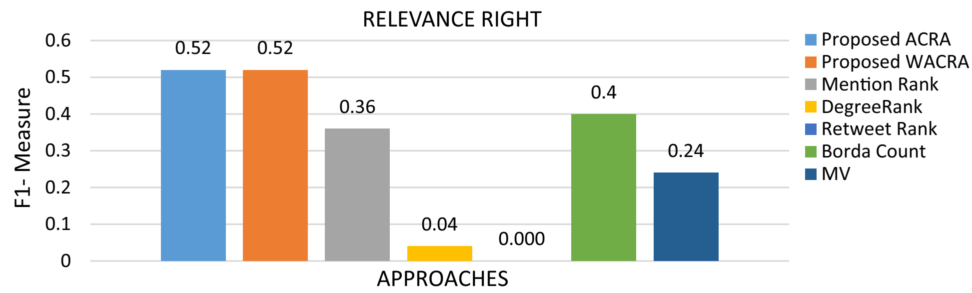


Fig. 13 Comparison among approaches in terms of MCC to identify top-k influential nodes for politics dataset matched with Followers

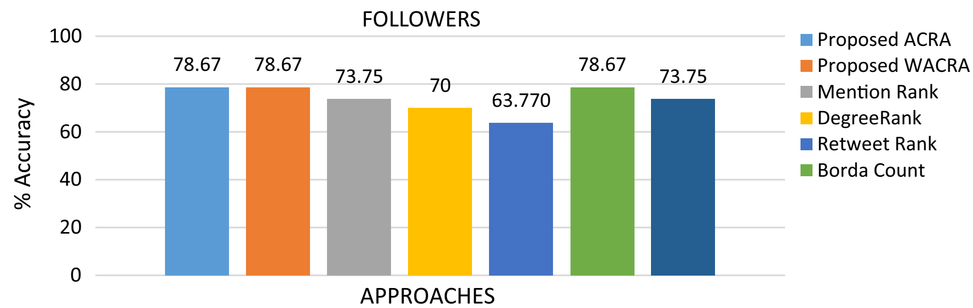


Fig. 14 Comparison among approaches in terms of accuracy to identify top-k influential nodes for economy dataset matched with Followers

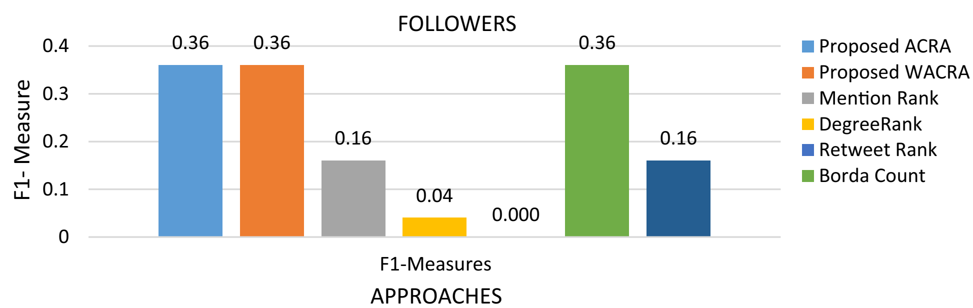


Fig. 15 Comparison Among approaches in terms of F1-measures to identify top-k influential nodes for economy dataset matched with Followers

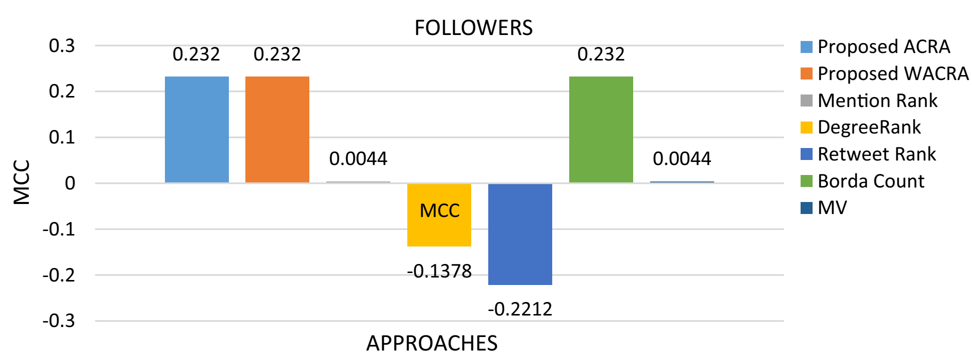


Fig. 16 Comparison Among approaches in terms of MCC to identify top-k influential nodes for economy dataset matched with Followers

Aggregation of followers and contents based features makes proposed approaches to achieve better accuracy as compare to other approaches as depicted in Fig. 14. Similarly, Fig. 15 again depicts that for Economy dataset, BC and proposed approaches performs better in term of F1-measures. Similar to Politics dataset, heuristic approaches such as Degree rank with very low and Retweet rank with negligible F1-measure score due to absence of TN value. For MCC evaluation measure there is drift in performance remains same as with Accuracy and F1-measure, where BC and our approaches outperformss all other approaches (Fig. 16).

For Retweet Online, in Politics dataset while measuring accuracy BC performs better compare to proposed approach (Fig. 17). As aforementioned influential nodes in Economy data are vital with followers and contents features score and Retweet Online evaluation consider both leads to improvement in performance of proposed

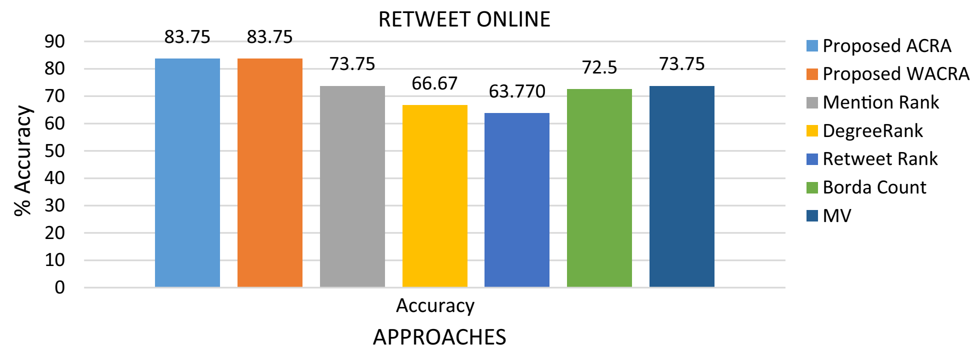


Fig. 17 Comparison among approaches in terms of accuracy to identify top-k influential nodes for economy dataset matched with Retweet Online

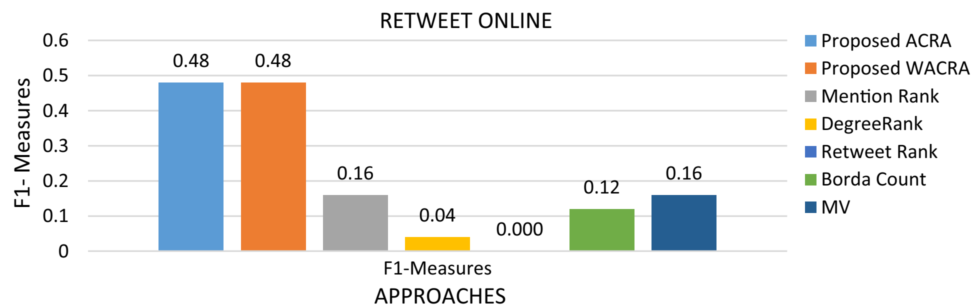


Fig. 18 Comparison among approaches in terms of F1-measures to identify top-k influential nodes for economy dataset matched with Retweet Online

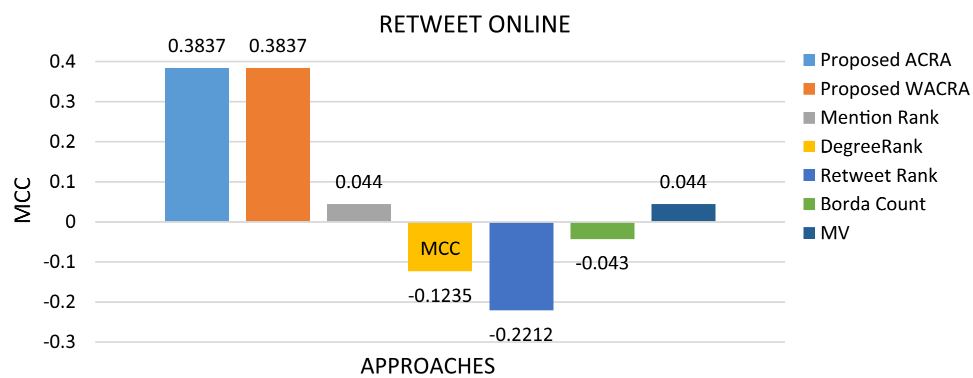


Fig. 19 Comparison among approaches in terms of MCC to identify top-k influential nodes for economy dataset matched with Retweet Online

approaches with accuracy of 83.75% compare to all other approaches.

The Mention rank and MV comes second in terms of accuracy overcoming BC. The proposed approaches show drastically improvement in performance in terms of F1-measures with score 0.48 (Fig. 18). The results show that the proposed approaches identify more correct influential nodes than any other approaches. Similarly, in case of MCC the Degree, and Retweet rank as well as BC gives negative value of MCC score (Fig. 19). The Mention rank and MV perform better than aforementioned approaches and however proposed approaches outperforms all comparative approaches.

Figure 20, show that proposed approaches ACRA and WACRA outperforms other comparative approaches with same 76.25% accuracy. The importance of contents and expertise in Economy dataset as well as the topic relevancy and content authority criteria in Right Relevancy to identify influential nodes allow proposed approaches to outperforms its counterparts in all three evaluation measures i.e. Accuracy, F1-measures and MCC (Figs. 20, 21, 22). The Mention rank performed better as compare to Degree, Retweet BC and MV.

4.2.4 Ranking of top-influential nodes

Many real world application needs the ranking of influential nodes to select the require k influential nodes according to their requisite. However, identified influential nodes provide the probability of node present in the list not its position.

To evaluate effectiveness for ranking of identified nodes, two different measures are used i.e. Average Precision (Avg.P) and Mean Reciprocal Rank (MRR) for

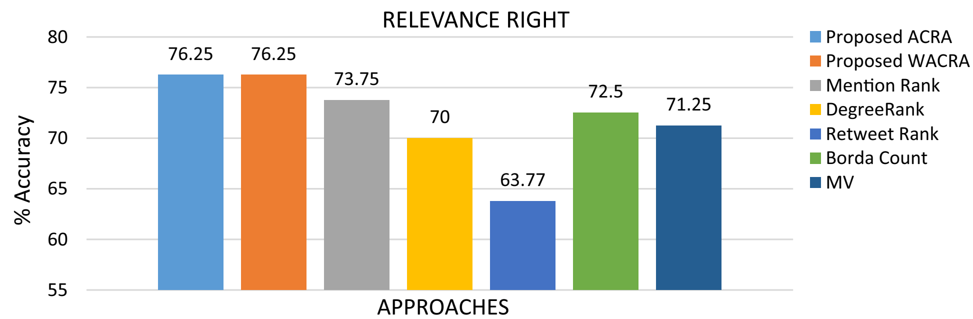


Fig. 20 Comparison among approaches in terms of accuracy to identify top-k influential nodes for economy dataset matched with Relevance Right

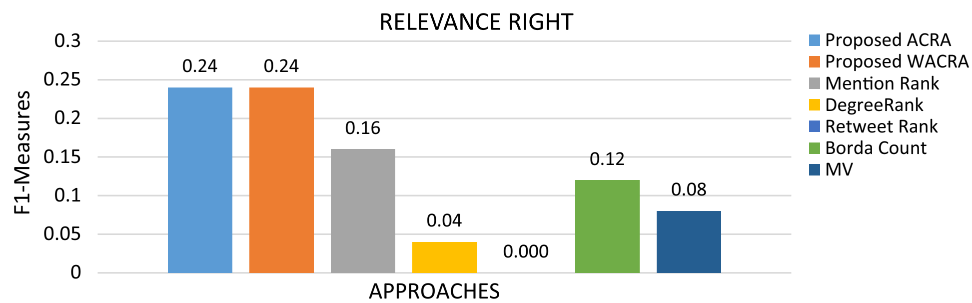


Fig. 21 Comparison Among approaches in terms of F1-measures to identify top-k influential nodes for economy dataset matched with Relevance Right

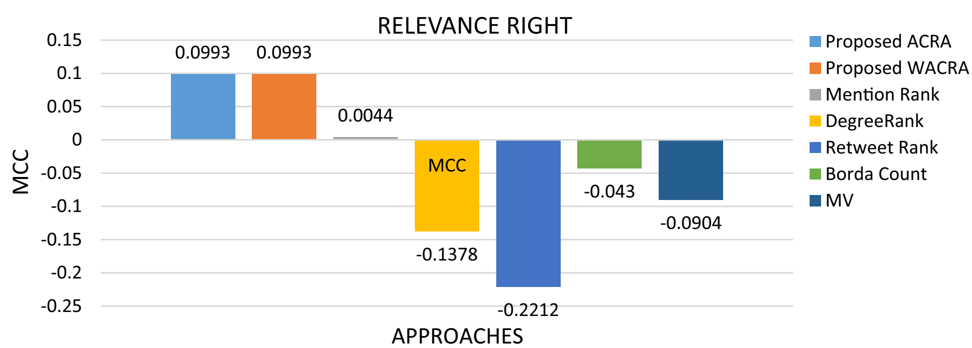


Fig. 22 Comparison among approaches in terms of MCC for identified top-k influential nodes for economy dataset matched with Relevance Right

various top-k ranks such as @5, @10….@25 for both dataset Politics and Economy [40]. Similar to identification, the cut-off value of ranking is user-defined which may be change with different application and user.

- *Average Precision (Avg.P)* Avg.P is a well-accepted evaluation measure for approaches that performs relevant information extraction such as Google search. This measure is suitable for rank top-k relevant influential nodes. For IRIN, Avg.P calculate the mean of the Precision at k. Avg.P give higher value for placing accurate influential nodes at the top. So sequence of "correct" and "incorrect" affects the final AP score,

$$\begin{aligned} \text{Avg.P@}k &= \frac{1}{n}\sum_{m=1}^{K}(P(m) \text{ if } m\text{th node is correct}) \\ &= \frac{1}{n}\sum_{m=1}^{k}P(m)\cdot \text{correct}(m) \end{aligned} \tag{14}$$

where k is list of ranked influential nodes; m is correctly ranked influential nodes and if mth item was correct (correct $(m) = 1$) else not (correct (m) = 0).

- *Mean Reciprocal Rank (MRR)* The reciprocal rank is an evaluation metric for assessing correctness of object position, predicted by some models for a given query. The reciprocal rank of a predicted output is the multiplicative inverse of the position of the first accurate response: first position gets 1, second position gets 1/2, 1/3 for object placed in third position and so on.

$$\text{R.Rank} = \begin{cases} \frac{1}{n} & \text{if output is correct} \\ 0 & \text{if output is wrong} \end{cases} \tag{15}$$

$$\text{MRR} = \frac{1}{|q|}\sum_{k=1}^{|q|}\frac{1}{\text{position}_i} \tag{16}$$

Therefore, MRR is the mean of the *R.rank* of predicted result for a given '*q*' query.

4.2.5 Performance comparisons on politics data

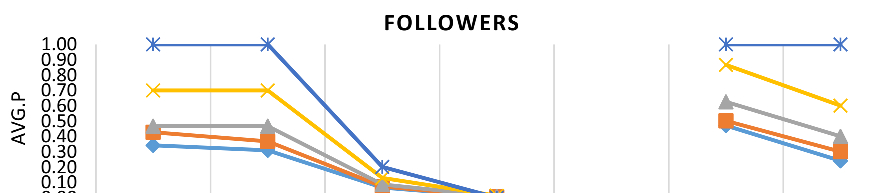
4.2.5.1 Followers The BC in terms of Avg.P@25 for Followers count (Fig. 23) perform better as compare to its counterparts. Similarly, BC performed best or equal best in all Avg.P@*k*. The main reason is that as followers based features were the main criteria for BC, the most influential nodes with higher followers has place in top position. Degree and Retweet rank perform the worst for all Avg.P@*k,* as they fail to even identify influential nodes. In case of Avg.P@25 and Avg.P@20 proposed ACRA performs better than WACRA, as weightage to indegree is very less in WACRA, the nodes with higher followers is placed in lower position which directly affects the performance of WACRA, while comparing with Followers Count evaluation measure. However, in remaining Avg.P@*k* they have equal performance score. MV rank performance improve proportionally with lower values of *k*.

Again, the BC outperformss its counterparts in terms of MRR, for each MRR@*k* for Followers (Fig. 24). The reason is same as Avg.P. Degree and Retweet rank perform the worst for all MRR@*k,* as in Avg.P.

Interestingly, proposed WACRA, in case of all MRR@*k* outperforms proposed ACRA. Though, weightage to indegree is very less in WACRA, the influential nodes position in Followers Count evaluation measures are same as position of WACRA. Mention rank show average performance and MV performance improves proportionally with lower values of *k* and it performance is better after BC and WACRA (Fig. 24).

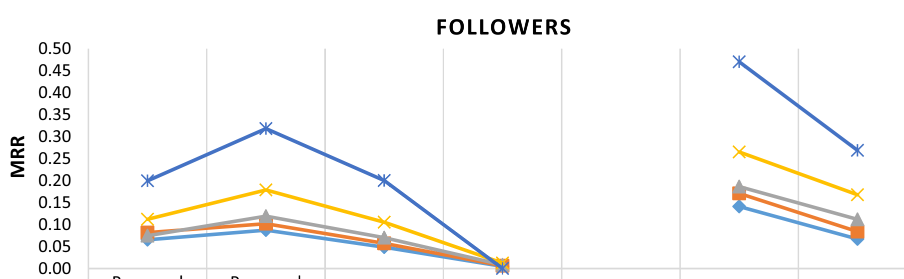
4.2.5.2 Retweet online The BC in terms of Avg.P@25 for Retweet Online perform better as compare to its counterparts (Fig. 25).

The main reason is that; Retweet Online rank consider retweets as one of the features for computing nodes influence. The remaining features are basically focused on popularity of nodes in terms of Friends, Followers and List which incline their list towards BC. However, proposed WACRA, in case of all MRR@*k* outperforms proposed ACRA with its better ranking list when compared to Retweet Online. Degree and Retweet rank again perform the worst for Avg.P@*k*. MV performance improves



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
AVG.P @25	0.341	0.309	0.065	0.005		0.470	0.240
AVG.P @20	0.426	0.368	0.073	0.006		0.501	0.300
AVG.P @15	0.467	0.467	0.086	0.008		0.625	0.400
AVG.P @10	0.700	0.700	0.129	0.013		0.866	0.600
AVG.P @5	1.000	1.000	0.200	0.000		1.000	1.000

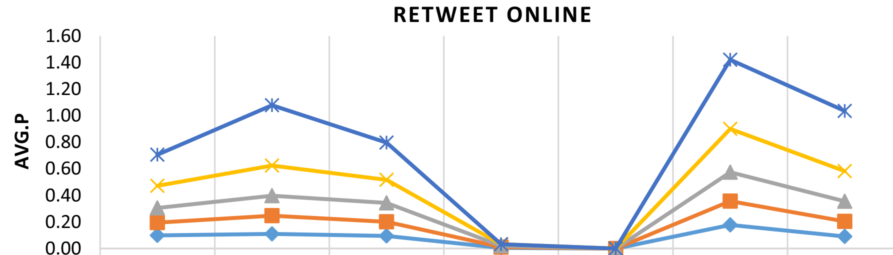
Fig. 23 Comparison Among approaches in terms of average precision (Avg.P) to rank top-*k* influential nodes for politics dataset matched with Followers



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
MRR @25	0.065	0.087	0.049	0.005		0.141	0.067
MRR @20	0.082	0.102	0.057	0.006		0.171	0.084
MRR @15	0.075	0.119	0.070	0.008		0.186	0.112
MRR @10	0.112	0.179	0.105	0.013		0.265	0.168
MRR @5	0.199	0.318	0.200	0.000		0.470	0.269

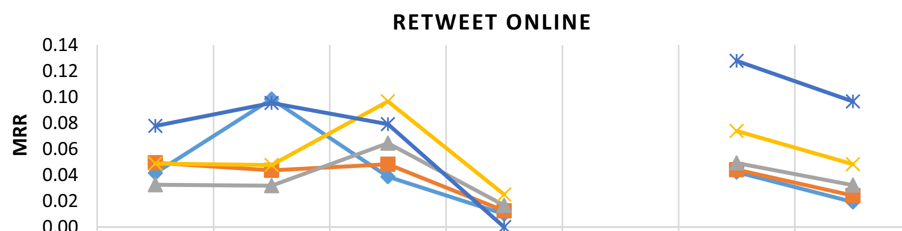
Fig. 24 Comparison among approaches in terms of mean reciprocal rank (MRR) to rank top-*k* influential nodes for politics dataset matched with Followers

Fig. 25 Comparison Among approaches in terms of average precision (Avg.P) to rank top-k influential nodes for politics dataset matched with Retweet Online



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
AVG.P @5	0.2333	0.4533	0.2800	0.0000		0.5200	0.4533
AVG.P @10	0.1667	0.2267	0.1733	0.0125		0.3267	0.2267
AVG.P @15	0.1111	0.1511	0.1422	0.0083		0.2178	0.1511
AVG.P @20	0.0958	0.1369	0.1067	0.0063		0.1780	0.1133
AVG.P @25	0.0986	0.1096	0.0944	0.0050		0.1778	0.0907

Fig. 26 Comparison Among approaches in terms of mean reciprocal rank (MRR) to rank top-k influential nodes for politics dataset matched with Retweet Online



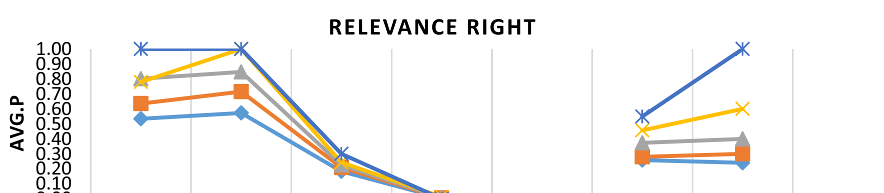
	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
MRR @25	0.0416	0.0985	0.0387	0.0100		0.0422	0.0193
MRR @20	0.0494	0.0436	0.0484	0.0125		0.0441	0.0242
MRR @15	0.0326	0.0318	0.0645	0.0167		0.0493	0.0322
MRR @10	0.0489	0.0477	0.0967	0.0250		0.0739	0.0483
MRR @5	0.0778	0.0954	0.0792	0.0000		0.1278	0.0966

proportionally with lower values of k. For MRR@20 ACRA performance better, for MRR@15 and MRR@10, Mention rank performs better (Fig. 26).

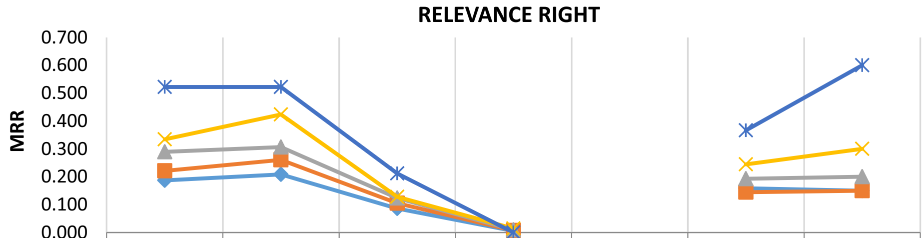
Similarly, for MRR@5, BC outperforms their counter parts. The main reason is that, MRR compare the position of individual nodes in evaluation metric, i.e. Retweet Online rank list, with the position of same node in the list of various approaches. Degree and Retweet rank again perform the worst average precision for all Avg.P@k.

4.2.5.3 Relevance right The proposed approach WACRA outperforms its counterparts in terms of Avg.P@ 25, 20, 15, 10 and equals ACRA and BC for Avg.P@5 for Relevance Right (Fig. 27). The main reason is that relevance Right consider topic relevancy and nodes popularity based on various stats. Degree and Retweet rank gave the worst for all Avg.P@k. Interestingly, proposed ACRA, outperforms all other approaches except WACRA for all Avg.P@25, 20, 15, 10 and equals ACRA and BC for

Fig. 27 Comparison Among approaches in terms of average precision (Avg.P) to rank top-k influential nodes for politics dataset matched with Relevance Right



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
AVG.P @25	0.534	0.573	0.183	0.005		0.259	0.240
AVG.P @20	0.636	0.716	0.209	0.006		0.280	0.300
AVG.P @15	0.801	0.847	0.226	0.008		0.374	0.400
AVG.P @10	0.780	1.000	0.243	0.013		0.457	0.600
AVG.P @5	1.000	1.000	0.300	0.000		0.550	1.000



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
MRR @25	0.187	0.208	0.086	0.006		0.159	0.150
MRR @20	0.221	0.260	0.105	0.007		0.144	0.150
MRR @15	0.290	0.307	0.123	0.010		0.193	0.200
MRR @10	0.334	0.423	0.128	0.014		0.244	0.300
MRR @5	0.522	0.522	0.212	0.000		0.367	0.600

Fig. 28 Comparison Among approaches in terms of mean reciprocal rank (MRR) to rank top-k influential nodes for politics dataset matched with Relevance Right

Avg.P@5 for Relevance Right. BC shows average performance and MV performance improves proportionally with lower values of k and it performed better next to WACRA and ACRA.

ACRA, outperforms all other approaches except WACRA for Av and MV for MRR@5 for Relevance Right as shown in Fig. 28. BC and Mention rank show average performance.

BC and Mention rank show below average performance as compare to their previous record due to ignorance of topic relevancy.

4.2.6 Performance comparisons on economy data

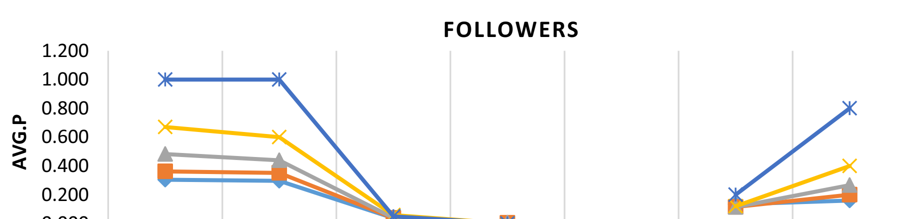
The proposed approach ACRA outperforms its counterparts in terms of Avg.P@ 25, 20, 15, 10, however, equals WACRA for Avg.P@ 5 (Fig. 29). The main reason is that ACRA give equal weightage to indegree feature as compare to WACRA which give lowest weightage to indegree. WACRA stands second best for Avg.P@ 25, 20, 15, 10. MV outperforms all other approaches except proposed approaches for all for Avg.P@k. BC and Mention rank show below average performance as compare to their previous record as Economy data required topical authority or expertise to be more influential contrast to Politics dataset.

The proposed approach WACRA outperforms its counterparts for MRR @25, 20, 15, 10, however, MV for MRR@5 outperforms its counterparts when evaluated with Followers Count for Economy dataset.

The main reason is that position of nodes in Followers Count is match more with WACRA list as compare to its counterparts for MRR@25, 20, 15, 10. Degree and Retweet rank perform worst for all MRR@k. Again, MV perform better for MRR@10 except proposed approaches. BC performs better for MRR@25, 20, 15 except proposed approaches and Mention rank show below average performance as compare to their previous record.

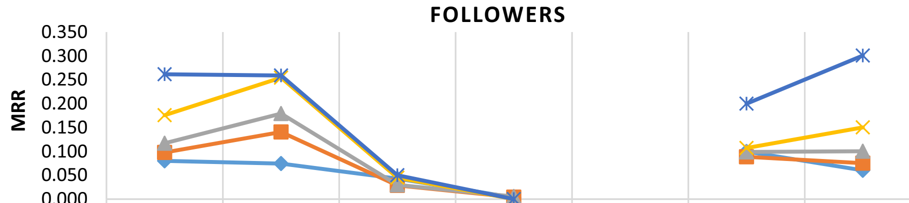
The proposed approach ACRA outperformss its counterparts in terms of Avg.P, for Avg.P @25, 20, 15, 10, and equals with WACRA for MRR@5 when evaluated with Retweet rank Online for Economy dataset (Fig. 30).

4.2.6.1 Retweet online For Retweet rank WACRA perform better as compare to its counterparts. Degree and Retweet rank gave the worst Avg.P@k. MV outperforms



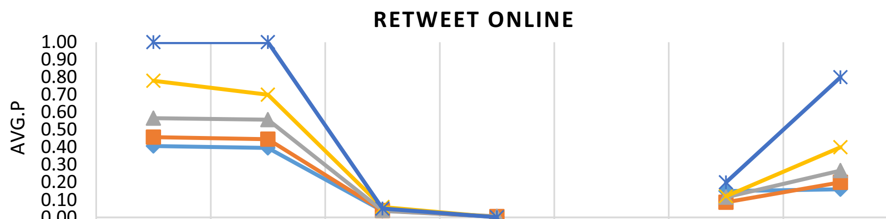
	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
AVG.P @25	0.304	0.296	0.037	0.003		0.134	0.160
AVG.P @20	0.362	0.351	0.039	0.004		0.116	0.200
AVG.P @15	0.482	0.439	0.039	0.006		0.114	0.267
AVG.P @10	0.670	0.600	0.058	0.000		0.120	0.400
AVG.P @5	1.000	1.000	0.050	0.000		0.200	0.800

Fig. 29 Comparison Among approaches in terms of average precision (Avg.P) to rank top-k influential nodes for economy dataset matched with Followers



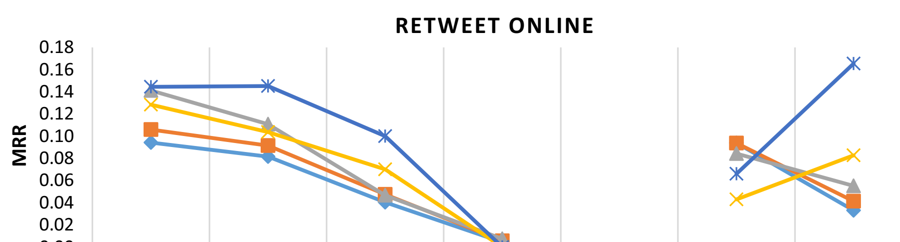
	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRan k	Retweet Rank	Borda Count	MV
MRR @25	0.080	0.074	0.043	0.003		0.100	0.060
MRR @20	0.098	0.141	0.029	0.004		0.089	0.075
MRR @15	0.117	0.179	0.030	0.006		0.099	0.100
MRR @10	0.176	0.255	0.045	0.000		0.107	0.150
MRR @5	0.261	0.259	0.050	0.000		0.200	0.301

Fig. 30 Comparison among approaches in terms of mean reciprocal rank (MRR) to rank top-k influential nodes for economy dataset matched with Followers



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRan k	Retweet Rank	Borda Count	MV
AVG.P @25	0.407	0.397	0.037	0.003		0.150	0.160
AVG.P @20	0.458	0.446	0.039	0.004		0.086	0.200
AVG.P @15	0.566	0.557	0.039	0.006		0.114	0.267
AVG.P @10	0.780	0.700	0.058	0.000		0.120	0.400
AVG.P @5	1.000	1.000	0.050	0.000		0.200	0.800

Fig. 31 Comparison among approaches in terms of average precision (Avg.P) to rank top-k influential nodes for economy dataset matched with Retweet Online



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRan k	Retweet Rank	Borda Count	MV
MRR @25	0.094	0.082	0.040	0.004		0.094	0.033
MRR @20	0.106	0.091	0.048	0.006		0.094	0.041
MRR @15	0.141	0.111	0.047	0.007		0.084	0.055
MRR @10	0.128	0.104	0.070	0.000		0.043	0.083
MRR @5	0.144	0.145	0.100	0.000		0.066	0.165

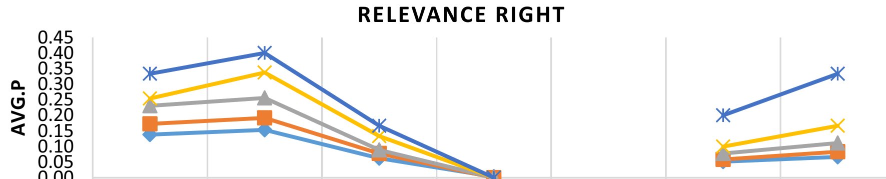
Fig. 32 Comparison among approaches in terms of mean reciprocal rank (MRR) to rank top-k influential nodes for economy dataset matched with Retweet Online

all other approaches except proposed approaches for Avg.P@ k.

The proposed approach ACRA outperforms its counterparts for MRR @25, 20, 15, 10; however, WACRA perform better for MRR@5. MV performed best for MRR@5. ACRA give same weightage to all features as compare to WACRA, which improve the performance. MRR@ 25, 20, 15, 10 give lowest weightage to indegree which is nothing but followers list. MV perform better for MRR@10 except proposed approaches. BC perform better for MRR@20, 15 except proposed approaches (Fig. 31).

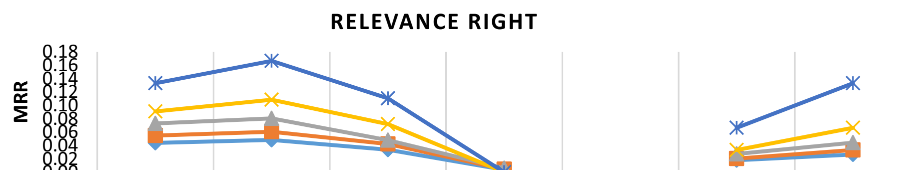
MV outperforms all other approaches except proposed approaches. Mention rank show average performance for all for MRR@ k (Fig. 32).

4.2.6.2 Relevance right The proposed approach WACRA outperforms its counterparts for Avg.P@k. ACRA perform second best for Avg.P@k. The main reason is that WACRA



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
AVG.P @25	0.138	0.153	0.062	0.003		0.051	0.067
AVG.P @20	0.173	0.192	0.078	0.000		0.058	0.083
AVG.P @15	0.230	0.256	0.089	0.000		0.078	0.111
AVG.P @10	0.254	0.337	0.133	0.000		0.100	0.167
AVG.P @5	0.333	0.400	0.167	0.000		0.200	0.333

Fig. 33 Comparison among approaches in terms of average precision (Avg.P) to rank top-k influential nodes for economy dataset matched with Relevance Right



	Proposed ACRA	Proposed WACRA	Mention Rank	DegreeRank	Retweet Rank	Borda Count	MV
MRR @25	0.044	0.048	0.034	0.004		0.018	0.027
MRR @20	0.055	0.061	0.042	0.005		0.020	0.033
MRR @15	0.073	0.081	0.048	0.006		0.027	0.044
MRR @10	0.091	0.109	0.072	0.000		0.033	0.067
MRR @5	0.133	0.167	0.111	0.000		0.067	0.133

Fig. 34 Comparison among approaches in terms of mean reciprocal rank (MRR) to rank top-k influential nodes for economy dataset matched with Relevance Right

assign higher weightage to contents related feature which are more incline toward topic relevancy same as Relevance Right as compare to ACRA which stands second best for Avg.P@ 25, 20, 15, 10 and equals with MV for MRR@5 who perform better for MRR@k except proposed approaches (Fig. 33). BC perform slightly worse as compare to its performance in Followers Count and Retweet rank. Mention rank show average performance for all for MRR@k.

All approaches show same performance when evaluate with MRR for all MRR@k as in Avg.P@k as shown in Fig. 34.

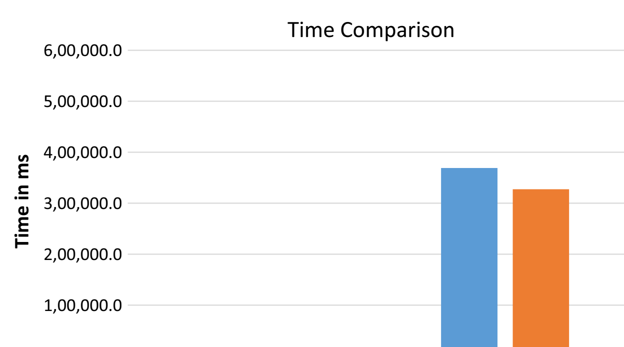
4.2.7 Time comparison

Figure 35 show the time comparison between sum of HSLF and sum of BC measures.

For both Politics as well as Economy data, there is enormous difference for time taken by HSLF compare to BC measures to generate list of influential nodes. The result show that HSLF are scalable as compare to BC measure.

5 Result discussion

Consensus aggregation approach for IRIN in topical networks is an important approach demonstrated in findings from several experimental results. It is valuable to integrate heterogeneous features based on Link, User based and Contents [10]. The proposed approaches have tendency to identify the nodes with more topical authority, in addition, to their popularity which makes them suitable for IRIN with respect to particular topic. However, the effectiveness of various features is totally depending on network structure and specific topic. It is observed from empirical analysis that indegree resembles the popularity of a node, mentions signify the tag value of a node name to make the tweet important as well as compute authority node on specific topic. Mention make information to disseminate to



	HSLF	Borda Count measures
Politics	2944	369022
Economy	2734	327381

Fig. 35 Time comparison

correct set of nodes. Retweet as a popular measure is effective in terms of count. However, the Retweet rank shown negligible presence while IRIN in our experimental analysis.

Many existing work apply distance-based consensus rank aggregation approach such as Kemeny's rule or Kendall's tau to overcome the problem of Condorcet criterion. These methods are NP-hard [37, 42], however, proposed approach WACRA gives result in polynomial time. In addition, simple average consensus aggregation approaches are robust and provide accurate results in most of the scenarios. The heuristic approaches such as Degree rank, Mention rank and Retweet rank do not augment the total task of influence as these approaches IRIN based on individual criteria and overlook the importance of combined influence. Therefore, the outcomes of the heuristic approaches fail to obtained desired results.

The aggregation of centrality performs better when compared with certain evaluation measures and may be biased towards one relationship in the network, as in case, BC performed better when compared with Followers Count. However, their computational complexity is high as shown in Fig. 35. Evaluation list used in experiment i.e. Followers Count, Retweet Online and Relevance Right are some available IRIN benchmarks, however, existing literature suggested that there is no globally accepted benchmark to verify the results of influential node. Many authors used manual annotation to verify their results [10]. Similarly, Al-Garadi et al. [45] and Alp and Öğüdücü [7] used real spread dynamics to identify influential nodes. These evaluation measures are either judges influence maximization or use retweets as measure to determine spread. In our case, the problem is focused on influential nodes that have authority over specific topic. So, Influential nodes finding and influential maximization [46] seed nodes are two different problems.

6 Conclusion

With the wide-ranging arrival of networked communication platforms such as social networks, it is crucial to improve approaches to find influential nodes from the heterogeneous network. Finding influential nodes play vital role in many applications such as Viral marketing, Smart Cities decision making and management. The diversity of contradicting features is a problem for a decision-maker while ranking influential nodes for Twitter data. To derive conclusions, there must be unbiased unique single ranking list. An evident key is to select the significant features that are capable to estimates the influence of individual nodes for particular topic. For IRIN, in Twitter data, there is no convincing observation to believe that one feature is by some means mediocre to its counterparts.

The main objective of this research is to improve scalability and accuracy for IRIN using Twitter data. A holistic approach introduced that combines the network, nodes specific and contents based features as a HSLF for IRIN. In addition, proposed two approaches Average Consensus Ranking Aggregation and Weighted Average Consensus Ranking Aggregation using these HSLF. To identify influential nodes; evaluation measures such as Accuracy, F1-measures and MCC and to rank influential nodes; evaluation measures such as Average Precision and Mean reciprocal rank are used to validate the effectiveness and efficiency of proposed approach on two topics Politics and Economy for real Twitter data. The experiments show that the proposed approaches provide scalability and better results when compared with heuristic as well as state-of-the art approaches using well known evaluation measures.

In future, different application will be exploring further to apply proposed approaches. The weight to each feature can be assign based on their correlation with actual influential nodes or entropy base as suggested in [21, 28]. Moreover, diverse features for IRIN can be develop to improve the efficiency and effectiveness for IRIN as well as on other influence domain such as the Influence Maximization and Continuous Time Diffusion Model.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

See Tables 8 and 9.

Table 8 Identified and ranked influential nodes for various approaches for politics data

S. No	ACRA	WACRA	Mention rank	Degree rank	Retweet rank	Borda count	Borda Majority Voting	Followers	Retweet Online	Relevance Right
1	RahulGandhi	narendramodi	narendramodi	shaktisinha	bad_indian_girl	narendramodi	narendramodi	narendramodi	SushmaSwaraj	narendramodi
2	narendramodi	RahulGandhi	rsprasad	HappymonJacob	satirist_indian	arunjaitley	RahulGandhi	iamsrk	Gurmeetramrahim	RahulGandhi
3	arunjaitley	Swamy39	ImaanZHazir	Akshayamukul	politics_is_lyf	SrBachchan	arunjaitley	SrBachchan	nitin_gadkari	arunjaitley
4	ArvindKejriwal	ArvindKejriwal	sagarikaghose	_NAN_DINI	Firojkh57781644	AmitShah	AmitShah	akshaykumar	narendramodi	smritiirani
5	AmitShah	arunjaitley	abhijitmajumder	ArmchairPseph	Shahzebalvi4	SushmaSwaraj	Swamy39	sachin_rt	manojkjhadu	priyankac19
6	Swamy39	AmitShah	LegalKant	Sanju_Verma_	prayag	Swamy39	ArvindKejriwal	imVkohli	MinhazMerchant	AmitShah
7	sardesairajdeep	sardesairajdeep	RahulGandhi	UnamPillai	lop_india	Maheshsisodia10		virendersehwag	sachin_rt	PiyushGoyal
8	sardanarohit	ChouhanShivraj	ManishTewari	narendramodi	pink_indian_	rajnathsingh		ArvindKejriwal	padhalikha	Swamy39
9	mvmeet	Shehla_Rashid	padhalikha	rchops	Yogi210875	smritiirani		arunjaitley	pbhushan1	sardesairajdeep
10	Shehla_Rashid	divyaspandana	bhaiyyajispeaks	KanchanGupta	kiran_patniak	RahulGandhi		AnupamPKher	akshaykumar	sureshpprabhu
11	priyankac19	sardanarohit	Shehla_Rashid	write2kill	JaiPrakashAga18	DeviRangwa		ImRo45	DrGPradhan	nitin_gadkari
12	divyaspandana	priyankac19	ShekharGupta	singhsahana	Divya_India_ask	DhanrajDara1		SushmaSwaraj	ippatel	ArvindKejriwal
13	ChouhanShivraj	ShefVaidya	WrongDoc	khalidbshah	soilditem	sairamalikpk		AmitShah	imVkohli	shekhargupta
14	ShefVaidya	myogiadityanath	Rohini_Swamy	rahultripathi	shrieas	ArvindKejriwal		FarOutAkhtar	myogiadityanath	ChouhanShivraj
15	ShashiTharoor	ShashiTharoor	ZaidZamanHamid	YusufDFI	HinduNatl	PiyushGoyal		thekiranbedi	manoharparrikar	prakashjavdekar
16	virendersehwag	rsprasad	mvmeet	DilliDurAst	Forgoten_Indian	sharmayogendr22		rajnathsingh	Swamy39	dev_fadnavis
17	rsprasad	ManishTewari	DilliDurAst	Drsunandambal	GeeGaja	AnupamPKher		yadavakhilesh	virendersehwag	swapan55
18	SrBachchan	Gurmeetramrahim	ArvindKejriwal	gauravbjp4india	1Mdkhan	BDUTT		MirzaSania	KapilMishra_IND	rammadhavbjp
19	yadavakhilesh	sagarikaghose	SushantBSinha	meenakshisharan	drsatish0380	sksurana_jantv		GautamGambhir	yadavakhilesh	vasundharabjp
20	PiyushGoyal	virendersehwag	nitin_gadkari	a_truthsayer	zak_india	Fahadnaeem_		sardesairajdeep	arunjaitley	sagarikaghose
21	Gurmeetramrahim	LegalKant	ngogoi98	NarenMenon1	GubbannaG	sardesairajdeep		smritiirani	abhijitmajumder	divyaspandana
22	sambitswaraj	bad_indian_girl	arunjaitley	sunitarora	iamcongressi1	sachin_rt		bhogleharsha	PiyushGoyal	ShefVaidya
23	TajinderBagga	priyagupta999	RKRadhakrishn	khatvaanga	SaritaTanwani	ShashiTharoor		RahulGandhi	gauravbjp4india	ShashiTharoor
24	sagarikaghose	SrBachchan	MoeedNj	EvilYindoo	_prettywitty	manoharparrikar		Swamy39	AnupamPKher	Shehla_Rashid
25	priyagupta999	mohitsmartlove	AsimAli6	india131979	Blooming_India	virendersehwag		msdhoni	bhaiyyajispeaks	ManishTewari

Table 9 Identified and ranked influential nodes for various approaches for economy data

S. no	ACRA	WACRA	Mention rank	Degree rank	Retweet rank	Borda count	Borda Majority Voting	Followers	Retweet Online	Relevance Right
1	narendramodi	narendramodi	mitchellvii	narlak	prayag	narendramodi	narendramodi	narendramodi	SushmaSwaraj	arunjaitley
2	RahulGandhi	arunjaitley	sidmtweets	barkhad	satirist_indian	BhavikaKapoor5	RahulGandhi	iamsrk	nitin_gadkari	Piyush Goyal
3	arunjaitley	RahulGandhi	anandmahindra	anshumanscribe	navinwin1985	anshumanscribe	arunjaitley	SrBachchan	narendramodi	narendra modi
4	Swamy39	Swamy39	narendramodi	meenamo	bad_indian_girl	barkhad	imVkohli	BeingSalmanKhan	sureshpprabhu	rntata2000
5	AmitShah	AmitShah	UmarKhalidJNU	sankrant	IMA_Indian	InderpalSinghD2		akshaykumar	akshaykumar	thejaggi
6	ArvindKejriwal	ArvindKejriwal	arunjaitley	pramit_b	Nagaraj35337174	s_palani		imVkohli	imVkohli	bibekdebroy
7	PiyushGoyal	PiyushGoyal	fawadchaudhry	sonaliranade	Agri_Updates	shree_sw		aamir_khan	myogiadityanath	ram_guha
8	YashwantSinha	YashwantSinha	EmergingRoy	CarolHusband	ry54351	meenamo		virendersehwag	manoharparrikar	suchetadalal
9	anandmahindra	anandmahindra	rishibagree	deepsealioness	Bhatt_Indian	sankrant		ArvindKejriwal	Swamy39	swapan55
10	imVkohli	PChidambaram_IN	Jan_Achakzai	s_palani	MANJUNATHAS21	RahulGandhi		arunjaitley	virendersehwag	menakadoshi
11	PChidambaram_IN	s_katsuragi	Mayank1029	Hitarth1987	hi38797592	pramit_b		AnupamPKher	arunjaitley	dhume
12	s_katsuragi	imVkohli	UmarCheema1	narendramodi	HinduNatl	arunjaitley		ImRo45	PiyushGoyal	nileshshah68
13	ShashiTharoor	ShashiTharoor	abhijitsapkal1	BhavikaKapoor5	MarkMichaeldom	Shehzad_Ind		chetan_bhagat	AnupamPKher	pbmehta
14	divyaspandana	divyaspandana	muglikar_	Gopalee67	pink_indian_	sonaliranade		SushmaSwaraj	ShashiTharoor	sanjay__bakshi
15	AnupamPKher	ARMurugadoss	jainarayangaur	manojladwa	JaiSing07836143	SrBachchan		AmitShah	ArvindKejriwal	mihirssharma
16	ARMurugadoss	SirPareshRawal	RahulGandhi	vasudevan_k	sbm_indian	GeorgekurianINC		rajnathsingh	chetan_bhagat	iamsamirarora
17	SirPareshRawal	KapilSibal	RohanV	ecophilo	priyaursal	Swamy39		ajaydevgn	smritiirani	vikramchandra
18	KapilSibal	AnupamPKher	PChidambaram_IN	EdAmmar	praks_Indian	Prakash12Joshi		Varun_dvn	SrBachchan	chellaney
19	1Mdkhan	kanhaiyakumar	KanchanGupta	prayag	OshoKapoor	Gopalee67		harbhajan_singh	RahulGandhi	acorn
20	GeorgekurianINC	rishibagree	RenukaJain6	sanjayuvacha	RanjanaSiroha	AmitShah		yadavakhilesh	rajnathsingh	shashitharoor
21	manoharparrikar	ITarunRathi	debjani_ghosh_	Avadhutwaghbjp	Preety_Indian	manojladwa		sardesairajdeep	AmitShah	anandmahindra
22	rishibagree	manoharparrikar	asad_sayeed	MLA_Panda	TiwariKamlesh2	SushmaSwaraj		smritiirani	sardesairajdeep	anjliraval
23	kanhaiyakumar	GeorgekurianINC	ashoswai	madskak	infobusket	smritiirani		RahulGandhi	rsprasad	nsitharaman
24	SrBachchan	SrBachchan	SadhguruJV	Brahamvakya	GeorgekurianINC	rajnathsingh		Swamy39	sambitswaraj	nandannilekani
25	ITarunRathi	1Mdkhan	Swamy39	ArnazHathiram	MohanLa37334130	PiyushGoyal		RNTata2000	BDUTT	PChidambaram_IN

References

1. Aggarwal CC (2011) An introduction to social network data analytics. In: Aggarwal C (ed) Social network data analytics. Springer, Boston, pp 1–15
2. Cha M, Haddadi H, Benevenuto F, Gummadi PK (2010) Measuring user influence in twitter: the million follower fallacy. Icwsm 10(10–17):30
3. Bello-Orgaz G, Jung JJ, Camacho D (2016) Social big data: recent achievements and new challenges. Inf Fusion 28:45–59
4. Sapountzi A, Psannis KE (2016) Social networking data analysis tools & challenges. Future Gener Comput Syst 86:893–913
5. Serrat O (2017) Social network analysis. In: Serrat O (ed) Knowledge solutions. Springer, Singapore, pp 39–43
6. Oro E, Pizzuti C, Procopio N, Ruffolo M (2018) Detecting topic authoritative social media users: a multilayer network approach. IEEE Trans Multimed 20(5):1195–1208
7. Alp ZZ, Öğüdücü ŞG (2018) Identifying topical influencers on twitter based on user behavior and network topology. Knowl-Based Syst 141:211–221
8. De Domenico M, Solé-Ribalta A, Omodei E, Gómez S, Arenas A (2015) Ranking in interconnected multilayer networks reveals versatile nodes. Nat Commun 6:6868
9. Wang Z, Du C, Fan J, Xing Y (2017) Ranking influential nodes in social networks based on node position and neighborhood. Neurocomputing 260:466–477
10. Al-Garadi MA, Varathan KD, Ravana SD, Ahmed E, Mujtaba G, Khan MUS, Khan SU (2018) Analysis of online social network connections for identification of influential users: survey and open research issues. ACM Comput Surv (CSUR) 51(1):16
11. Riquelme F, González-Cantergiani P (2016) Measuring user influence on Twitter: a survey. Inf Process Manage 52(5):949–975
12. Subbian K, Melville P (2011) Supervised rank aggregation for predicting influencers in twitter. In: 2011 IEEE third international conference on privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom). IEEE, pp 661–665
13. Wang S, Zhao J (2015) Multi-attribute integrated measurement of node importance in complex networks. Chaos Interdiscip J Nonlinear Sci 25(11):113105
14. Zareie A, Sheikhahmadi A, Jalili M (2019) Identification of influential users in social networks based on users' interest. Inf Sci 493:217–231
15. Chen Z, Liu Y, Zhu W (2016) Ranking node influence in social networks. In: 2016 15th international symposium on parallel and distributed computing (ISPDC). IEEE, pp 277–284
16. Sheikhahmadi A, Nematbakhsh MA, Zareie A (2017) Identification of influential users by neighbors in online social networks. Phys A 486:517–534
17. Borgatti SP (2005) Centrality and network flow. Soc Netw 27(1):55–71
18. Newman ME (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54
19. Liu Y, Wei B, Du Y, Xiao F, Deng Y (2016) Identifying influential spreaders by weight degree centrality in complex networks. Chaos, Solitons Fractals 86:1–7
20. Yang L, Qiao Y, Liu Z, Ma J, Li X (2018) Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm. Soft Comput 22(2):453–464
21. Zareie A, Sheikhahmadi A, Fatemi A (2017) Influential nodes ranking in complex networks: an entropy-based approach. Chaos, Solitons Fractals 104:485–494
22. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Comput Netw ISDN Syst 30(1–7):107–117
23. Zhang Y, Ma J, Wang Z, Chen B, Yu Y (2018) Collective topical PageRank: a model to evaluate the topic-dependent academic impact of scientific papers. Scientometrics 114(3):1345–1372
24. Chai W, Xu W, Zuo M, Wen X (2013) ACQR: a novel framework to identify and predict influential users in micro-blogging. In: Pacis, p 20
25. Liu N, Li L, Xu G, Yang Z (2014) Identifying domain-dependent influential microblog users: a post-feature based approach. In: AAAI, pp 3122–3123
26. Mei Y, Zhong Y, Yang J (2015) Finding and analyzing principal features for measuring user influence on Twitter. In: 2015 IEEE first international conference on big data computing service and applications (BigDataService). IEEE, pp 478–486
27. Gao C, Zhong L, Li X, Zhang Z, Shi N (2015) Combination methods for identifying influential nodes in networks. Int J Mod Phys C 26(06):1550067
28. Zareie A, Sheikhahmadi A (2018) A hierarchical approach for influential node ranking in complex social networks. Expert Syst Appl 93:200–211
29. Mardani A, Jusoh A, Zavadskas EK (2015) Fuzzy multiple criteria decision-making techniques and applications—two decades review from 1994 to 2014. Expert Syst Appl 42(8):4126–4148
30. Du Y, Gao C, Hu Y, Mahadevan S, Deng Y (2014) A new method of identifying influential nodes in complex networks based on TOPSIS. Phys A 399:57–69
31. Gandhi M, Muruganantham A (2015) Potential influencers identification using multi-criteria decision making (MCDM) methods. Procedia Comput Sci 57:1179–1188
32. Hung YH, Huang TL, Hsieh JC, Tsuei HJ, Cheng CC, Tzeng GH (2012) Online reputation management for improving marketing by using a hybrid MCDM model. Knowl-Based Syst 35:87–93
33. Li YM, Shiu YL (2012) A diffusion mechanism for social advertising over microblogs. Decis Support Syst 54(1):9–22
34. Urena R, Chiclana F, Melancon G, Herrera-Viedma E (2019) A social network based approach for consensus achievement in multiperson decision making. Inf Fusion 47:72–87
35. Sun Q, Wang N, Zhou Y, Luo Z (2016) Identification of influential online social network users based on multi-features. Int J Pattern Recognit Artif Intell 30(06):1659015
36. Basu S, Maulik U, Chatterjee O (2016) Stability of consensus node orderings under imperfect network data. IEEE Trans Comput Soc Syst 3(3):120–131
37. Zahid MA, De Swart H (2015) The borda majority count. Inf Sci 295:429–440
38. Xiao F, Noro T, Tokuda T (2014) Finding news-topic oriented influential Twitter users based on topic related hashtag community detection. J. Web Eng. 13(5&6):405–429
39. Tidke B, Mehta R, Dhanani J (2018) SIRIF: supervised influence ranking based on influential network. J Intell Fuzzy Syst 35:1225–1237
40. Tidke B, Mehta R, Dhanani J (2019) Multimodal ensemble approach to identify and rank top-k influential nodes of scholarly literature using Twitter network. J Inf Sci. https://doi.org/10.1177/0165551519837190
41. Ailon N (2010) Aggregation of partial rankings, *p*-ratings and top-*m* lists. Algorithmica 57:284. https://doi.org/10.1007/s00453-008-9211-1
42. Korba A, Clemencon S, Sibony E (2017) A learning theory of ranking aggregation. In: Proceedings of the 20thInternational Conference on Artificial Intelligence and Statistics, in PMLR, vol 54, pp 1001–1010
43. Online 1: https://www.retweetrank.com/. Accessed 10 Nov 2018

44. Online 2: https://www.rightrelevance.com/search/influencers?query=indian%20politics&instance_type=person. Accessed 14 Nov 2018
45. Al-Garadi MA, Varathan KD, Ravana SD, Ahmed E, Chang V (2016) Identifying the influential spreaders in multilayer interactions of online social networks. J Intell Fuzzy Syst 31(5):2721–2735
46. Singh SS, Kumar A, Singh K, Biswas B (2019) C2IM: community based context-aware influence maximization in social networks. Physica A 514:796–818