International Conference on Information and Communication Technology and Its Applications (ICTA 2016)
Federal University of Technology, Minna, Nigeria
November 28 – 30, 2016

A Survey of Influential Nodes Detection Methods in Mobile Phone Network

Elizabeth N. Onwuka, Bala A. Salihu, and Sheriff Murtala
Department of Telecommunications Engineering, Federal University of Technology, Minna, Nigeria

***Abstract*—The number of mobile phone users is increasing tremendously. The social interaction between these mobile phone users can be represented using social network graphs. This type of study has very important applications in various areas especially in the detection of criminal groups who also use these devices to interact and plan their activities. Moreover, the study of identifying influential nodes in social network of any kind is currently receiving attention in the research arena. This is because identification of influential nodes of any network is significant to understanding the network. This becomes very important if the network in question is a criminal network, considering the insecurities of the current time. In this paper, a survey of influential nodes detection methods is carried out, we first define the problems associated with influential nodes detection and then examine various methods of identifying influential nodes. We also consider techniques employed in analysing users in the mobile phone network.**

Keywords*—social network; mobile phone network; influential nodes detection; centrality measures;

I. Introduction

Since the invention of mobile communication and other services attached to it, many people find it better and cheaper to communicate using the medium than wired communication thereby attracting more subscribers to use mobile communication network. A survey carried out by international telecommunication union (ITU) shows that the population of mobile phone subscribers increased from 738 million in the year 2000 to 7 billion in 2015 and within this same time the proportion of population covered by a 2G mobile cellular network increased from 58% to 95% with more remote areas captured [1]. In developing countries, at least one member of every household communicates using a mobile phone. Each subscriber enjoys making calls and receiving calls from other users and enjoys the same for short messages and Internet services. Telecommunication networks have really made the world a global village in the sense that peoples' social reach has expanded even across borders. The log of activities of each user is stored on the user's phone and also recorded with the Mobile Network Operators (MNOs). The information collected by the MNOs is referred to as Call Detail Record (CDR).

CDR contains metadata that describes a specific instance of a telecommunication transaction (calls, messages and Internet services) but does not include the content of that transaction, for example, CDR for a particular call contains both the caller and receiver's number, the time stamp (date and time), the duration of the call, the base station ID of the caller's location and other related information. CDR may capture thousands or millions of users within a specific time and place and it can be used to create a network of mobile phone subscribers. CDR is a huge repository of human behavioural data and it belongs to the group of data being currently described as Big Data. The information from CDR reveals the inter-relationship network between mobile phone subscribers at various spheres, generally called social networks. A mobile phone network is a social structure that represents the interconnection of mobile phone subscribers based on call detail record (CDR). An example of a social network of mobile phone users is shown in Fig. 1. The idea of forming a social interaction between mobile phone users support researchers in the different area of studies like personal mobility prediction, fraud detection in telecommunication [2], urban planning and development, geographical partitioning [3] and intelligence gathering for national security [4].

Human beings normally form groups or clusters based on certain commonalities. These groups (called communities) also reflected on the communication data. Networks are made up of communities and in each community, there are nodes with varying degrees of influence, these nodes are called influential nodes. A good area of application of mobile phone network is in the detection of influential mobile subscribers. For instance, a network of mobile phone subscribers which is created by collecting call record information from a reasonable number of actors (that act as seeds) will consist of different communities and some users within these communities will influence other users either positively or negatively. The major problem in this area is how to accurately determine the genuine influential individuals in a social network. In this paper, we present an overview of various ways of finding influential nodes in a social network.

The remainder of this paper is organised as follows: Section II provides a brief background and review of related works. Section III describes different methods of identifying influential nodes in a mobile phone network. Finally, we conclude this paper in Section IV.

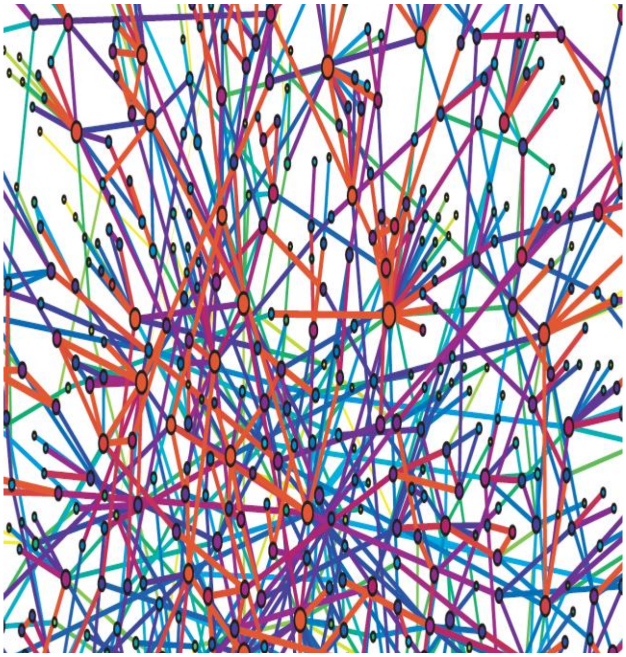

Figure 1. A snapshot of a mobile phone network sample with circles indicating mobile phone users and edge weight colour coded from yellow (weak link) to red (strong link) [5]

II. RELATED WORK

2.1 Background

There is a rapidly growing literature on influential nodes discovery in social networks, which indicates that a lot of study had been carried out in this field [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. However, due to the challenges of getting mobile phone data, little studies have been carried out on discovering communities and important mobile subscribers in the mobile phone network. A mobile phone network is treated like any other social network that has a tree network structure. Social network is usually modelled as a graph, G=(V,E), where V is a set containing all nodes (actors) in the network and E is also a set containing all edges (links) between two elements (pairs) of set V. If the direction of the edges is considered the graph is said to be directed or undirected if otherwise. Also, when the corresponding weight, W of the edges is considered, the graph, G=(V,E,W) is said to be weighted or binary (unweighted) if otherwise. A simple description of how an undirected and weighted graph is modelled from set V, E and W is shown in Fig. 2.

Exploring social network data requires basic concepts of graph representation, analysis and visualisation [18]. These concepts include centrality measures, shortest path problems, clustering techniques and network density. This is necessary when interpreting the result in order to have a good understanding of the social interactions between nodes in a network. Due to the rich resources in social network analysis, it serves as a tool for analysing and visualising big data [19]. Some major areas of study in the social network analysis are community structure, detection of cliques and discovery of key nodes and neighbours. Recently, more attention has been given to the detection of influential nodes in the social network. This is added to the fact that researchers and investigators have taken full advantage of social network analysis to unravel the operation of terrorists and criminals [4]. Crime investigation application becomes more necessary now that communication networks have changed the way people live and transact business. It is intuitively believed that criminals rely on this network for planning criminal activities of all sorts.

In our study, we focus on identifying important and interesting nodes in a mobile phone network and we discuss some of the previous studies that have been done in this research area by first looking at the major problems and concepts employed in the detection of important nodes and different approaches that had been applied so far.

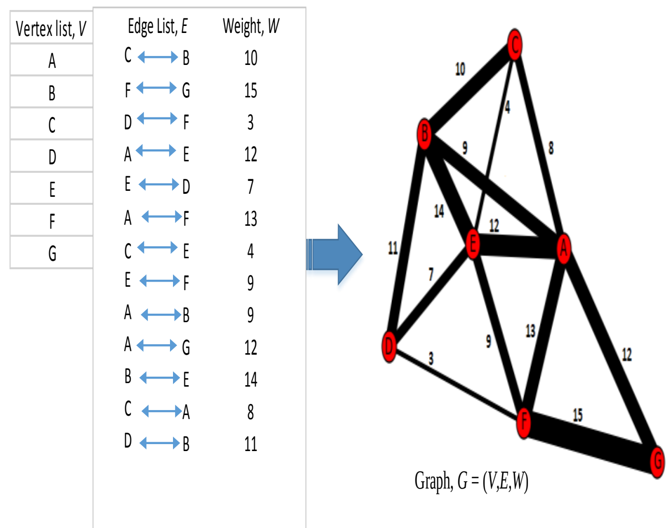

Figure 2. Modelling graph from a set of vertex, edge and weight.

2.2 Influential Nodes Detection Problem

Influential nodes are set of nodes whose roles are very important in the spread of influence across the network. These nodes have the tendency to influence other nodes either constructively or destructively. Influential nodes and "key nodes" seem to be the same. According to Borgatti [6], influential nodes' problem can either be a key player problem positive (KPP-Pos.) or key player problem negative (KPP-Neg.). KPP-Pos. is defined with respect to the way key nodes are connected and integrated into the network, while KPP-Neg. is defined in relation to the network reliance on its key nodes to sustain its connectedness.

Recently, [7][16] presented an overview of existing techniques of finding important and influential nodes in social networks. In this paper, we extend the study of Probst by reviewing more novel approaches in finding prominent nodes in the social network with emphasis on mobile phone network. For clarity, we classify some of the previous work on influential nodes detection into two categories: centrality and non-centrality measures.

1) Centrality measures

In graph theory and network analysis, the most important tool is centrality measure. Centrality measures are considered as structural measures of influence that indicate a user's position in a social network. Degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality are the four widely used centrality measures in determining the relative importance of a user within a network. Although these measures have limitations, they have been proven to be the basis of other methods of identifying key nodes based on specific purposes within a social network [20].

a) Degree Centrality: is defined as the number of edges incident upon a user. In other words, this measure indicates how many nodes can be directly reached by a particular node. The degree centrality of a user, v is given by:

$$DC(v) = \deg(v) \tag{1}$$

$$\deg(v, G) = |\{u \in V : (u, v) \in E\}| \tag{2}$$

Nodes with high degree centrality score might be considered influential. The flaw of this centrality measure is that it relies on direct connections between nodes. Using this individual centrality alone to determine the key nodes will result in the selection of nodes that only have a high number of direct connections.

b) Closeness Centrality: According to Bavelas, closeness centrality of a user as the reciprocal of the sum of its distances from all other nodes [21]. This measure is effective in describing the hierarchy among members of a group and can also be used to indicate how fast a user can reach every other user in the network. The closeness centrality of a user, v is given by:

$$CC(v) = \left(\sum_{u} d(v, u) \right)^{-1} \tag{3}$$

where $(u, v) \in E$ and $\sum d(v, u)$ is the sum of the length of all shortest paths from all other nodes from node "v". Okamoto et al. [22] introduced an efficient algorithm for discovering the top k-highest closeness centrality nodes called TOPRANK. The algorithm is made up of the approximate algorithm and exact algorithm. The approximate algorithm is applied to identify the top nodes with high centrality scores while the exact algorithm is used to rank the detected nodes. [23] presented a closeness centrality algorithm that efficiently determines the closeness scores of each user any time the social network structure is modified. The changes involve the insertion of a new edge or the removal of an existing edge.

c) Betweenness Centrality: Betweenness-based centrality measures were first introduced by Freeman in [24]. The author discovered that it is important to generalise the concept of point centrality and structural properties of the social network from past study[21]. Betweenness centrality expresses the number of times a user acts as a bridge along the shortest path between two other nodes. The betweenness centrality of a user, v is given by

$$BC(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4}$$

where σ_{st} is the number of shortest paths in the graph, G between nodes "s" and "t"; $\sigma_{st}(v)$ is the number of shortest paths in the graph, G between nodes "s" and "t" that pass through user "v". Nodes with high betweenness are responsible for controlling the spread of information across the graph. However, they might not be responsible for causing maximum disconnection (fragment) within the network [6]. Brandes also presented an algorithm for computing the betweenness index of a large number of nodes [25].

d) Eigenvector Centrality: Eigenvector centrality (also called eigencentrality) is a measure of how well a particular user is connected to other influential nodes. This is one of the oldest centrality measures developed to assist the social analyst in recognising the behaviour of people [26]. To determine eigenvector centrality, it is imperative to first find the adjacency matrix, A of the graph, G. Given $A = a(v, u)$ for a binary network.

Where:

$$a(v,u) = \begin{cases} 1, & \text{if the link between node } v \text{ and } u \text{ exist,} \\ 0, & \text{otherwise.} \end{cases}$$

The eigenvector centrality of a node, v is mathematically defined as:

$$x_v = \frac{1}{\lambda} \sum_{u \in M(v)} x_u \tag{5}$$

The boundary of the summation, is all members of the set of neighbours of v, M(v). In matrix representation, eigenvector centrality is expressed mathematically as:

$$x_v = \frac{1}{\lambda} A x_u \tag{6}$$

Where λ is the eigenvalue (constant) and x_v is the corresponding eigenvector of the adjacency matrix, A. Eigenvector centrality is much related to Katz centrality [27], a universality of degree centrality. Katz centrality measures the relative influence of nodes within a social network by determining the number of the immediate neighbours (first-degree nodes) and also all further nodes in the network that connect to the node under consideration through these close neighbours.

e) Other Centrality-Based Approaches: The number of centrality measures extends beyond the four metrics discussed earlier. It is quite interesting that most of the new measures were related one way or the other to the four most popular centrality measures with a little modification. Stephenson and Zelen suggested a new centrality measure called Information centrality [28]. This centrality has statistical theory background that considers all the path signals in a social network. Although this centrality is applied to undirected networks only, it supersedes other traditional centrality measures in identifying the most central nodes. In [29], the authors proposed an optimal inter-centrality measure, which takes into account both the user's centrality and its impact on the centrality of the other nodes. Using a criminal network, the authors' findings indicated that the key criminal is criminal with the highest optimal inter-centrality in the network and the removal of this criminal would greatly reduce the crime rate. In [30] and [31], the authors extended the work of Ballester et al., by presenting an inter-centrality with key group dimension. The modified inter-centrality explores the key group whose nodes are different from the nodes with highest individual inter-centralities.

Ilyas and Radha introduced a new centrality called principal component centrality (PCC), a variant of eigenvector centrality [32]. PCC is based on principal component analysis (PCA) and karhunen loeve transform (KLT) which handles graph adjacency matrix as a covariance matrix. Contrary to eigenvector centrality, PCC provides more features for centrality computation. Moreover, an investigation was carried out to detect influential nodes in two separate datasets using eigenvector centrality and principal component centrality [33]. Their results showed that eigenvector centrality considered the most influential node within the largest community in a network and consequently ranks the neighbours of the influential node and ignores other nodes in the remaining small communities that have low eigenvector scores. In the case of PCC, it considered both the nodes in the largest community and other nodes with zero eigenvalues in small communities.

Despite the introduction of these new centrality measures, the fact still remains that an individual centrality measure might not be the most appropriate for identifying influential nodes in a social network. A centrality measure is applied depending on a specific role of nodes in the network. For instance, nodes that are influential gossipers or most spreaders of virus function as information regulators in the network. Another different purpose is identifying nodes that can maximally disrupt the social network. The irregularities in some of these individual centrality measures have open up fascinating research fields on group and improved centrality measures that can be universal in identifying the most influential nodes [34][35]. Some studies also considered combining two or more centralities measures in getting a general set of influential nodes. Sathik and Rasheed proposed an algorithm to identify sets of key players based on centrality measures [36]. The authors addressed the key player problems [6], using closeness centrality, degree centrality and betweenness centrality. Zaman [37] in his (unpublished) PhD thesis recommended a new centrality measure known as rumor centrality that determines the source of rumor (gossip) and how influence (gossip) spread across a microblogging website (twitter).

Lately, in order to adequately discover real influential nodes. Ahsan et al. described a scheme that combines closeness centrality, betweenness centrality and eigenvector centrality to determine the influence factor of actors in an online social network obtained from Facebook [9]. The study shows that these three centrality measures are important in measuring the influence of each user and as well as the influence of the entire social network. Srinivas and Velusamy [10] presented an algorithm that combines degree centrality and clustering coefficient to discover influential nodes in three different datasets collected from Facebook. The clustering coefficient feature is used to enhance the traditional degree centrality. According to their study, the nodes with the least degree indicated that they are more connected and thus, the most influential. The algorithm is found to be effective in discovering important nodes. In [38], a criminal network is constructed and analysed using PageRank and other centrality measures to identify key criminals.

2) Non-centrality measures

In this subsection, we consider previous studies that employed other techniques different from centrality approach in detecting influential nodes.

a) Information theory: Shetty and Adibi [39] presented an information theory approach called graph entropy to discover the dominant nodes within a social network. The graph entropy of the entire network is determined every time a user is removed from the network. The removal of nodes that cause great disorder in the graph entropy are seen as influential nodes. Furthermore, [40] described an entropy measure based on Shannon measure of uncertainty. This entropy is applied to networks whose traffic moves along paths by transference. One great advantage of this measure is nodes with betweenness score of zero can actually have reasonable entropy values. In another developmental study, Ortiz-Arroyo et al. [41] proposed two entropy-based measures called connectivity entropy and centrality entropy. The authors further demonstrated how the entropy-based measure can effectively solve the two key player problems [6]. Although their result is similar to that of combinatorial optimisation algorithms the major setback for entropy approach is it cannot work on large graphs. The approach is computationally difficult to implement because it requires a path finding algorithm to simplify the operation of the entropy centrality.

b) Activity Based: This measure focuses on the activities between nodes in a social network. Goldenberg et al.[42] claimed that some individuals are more important than others based on their activity. The authors proposed an activity based measure for identifying influential nodes by selecting hub with high in-degree and out-degree in a directed network. Heidemann et al. revealed that not all connection links are active in a social network and the active links are insignificant [43]. They proposed an undirected activity based measure by modifying the PageRank algorithm for discovering important nodes in the network. The modified PageRank also considers the weighted edges between the nodes.

c) User Preferences and Attributes: Zhang et al.[11] developed an algorithm while trying to solve the influence maximisation problem called greedy algorithm based on user preferences (GAUP). The algorithm is applied along with an extended independent cascade (EIC) model and takes the user preferences into the influence diffusion process, thus making it the first algorithm that mines the top-k influential nodes based on user preferences while trying to solve influence maximisation problem. This algorithm surpasses the existing greedy algorithm that uses independent cascade (IC) model. Canali and Lancellotti [44] presented a numerical method of detecting key nodes in a social network by considering the nodes' attributes (personal information, nodes preferences, activities, uploaded contents, content accesses). The desired attributes of the nodes were collected and summed up using principal component analysis (PCA). The dimension of the PCA is matched with the nodes attributes and the dimension that greatly reflect the nodes' attributes is considered in defining a metric that determines the influential nodes in the social network. For each metric selected, the authors ranked the nodes and selected the top influential nodes accordingly.

d) Profile based characteristics: Nodes interactions are characterised as either popular, active or both. Eirinaki et al. introduced a measure named ProfRank [12], that uses the popularity and activity characteristics to identify the most influential nodes within a social network. The authors also showed that the measure performed better than betweenness centrality and Pagerank.

e) Influence Graph: Agarwal et al.[13] argued that influential bloggers are not necessarily active bloggers (nodes) on a blogging website. The authors defined statistical properties (recognition, activity generation, novelty and eloquence) that are related to influence between a blog post and proposed an influence graph model that measures the influence of each blog post across the community blog site. The influence is dependent on the influence flow and additive weight function that regulates a number of comments. The influence score is assigned to each blog's post. The maximum influence score is selected as the reference and its influence score as the blogger index. Using this index, the bloggers are ranked accordingly to the index and the most influential bloggers are the top ranked.

f) ShaPley value-based: Narayanam and Narahari, while trying to solve the influence maximisation problem proposed an algorithm called ShaPley value-based Influential Nodes (SPINs)[14]. Eventually, SPINs solved the top-k nodes problem and λ-coverage problem. The top-k problem involves the discovery of a set of influential nodes for maximising the spread of information. The λ-coverage problem focuses on the identifying the set of influential nodes having least cardinality with which it is possible to influence a fixed percentage, λ of the nodes in the social network through the process of diffusion. The authors showed that SPIN is computationally efficient when compared to other existing greedy algorithms.

g) Association Rule Learning: Erlandsson et al. discovered influential nodes in a social network by applying asocial rule learning. "Association rule learning is a machine learning technique that aims to find out how one item affects another by analysing how frequently certain items appear together in a specific dataset [15]. 'Association rule learning is carried out by applying two norms, namely, support and confidence. Support specifies the proportion of such items, while confidence specifies how many times those rules in the whole dataset are accurate. The influential nodes are listed as nodes with a confidence level of 95% and above. The technique is easy to implement and proven to be similar to PageRank and degree centrality.

III. IDENTIFYING INFLUENTIAL NODES IN A MOBILE PHONE NETWORK

In this section, research methods that are applied in detecting influential nodes in mobile phone networks were discussed. Over time, researchers attempted to study and analyse Call Detail Record (CDR) [5][45]. However, Mobile Network Operator(s) (MNOs) are strongly reluctant to release mobile phone data to the public due to privacy issue. In cases where CDR is released to third parties, MNOs might conceal the identity of their users or non-disclosure agreement (NDA's) contract is involved in protecting customers' privacy. But if the agreement fails, another way to collect CDR from mobile phone users is to develop a programme that extract users' log of activities. The programme is usually installed on user's mobile phone and each information retrieved from the user is stored in a dedicated database [46]. The process is expensive and takes longer time but the result worth it.

Kiss and Bichler [47] compared the performances of seven existing centrality measures including SenderRank, a new technique which was developed by the authors, to identify influential users in a social network constructed from a dataset collected from a telecom company. SenderRank and Out-degree centrality performed well in determining the most central nodes in a network of calls from a telecommunication company. In [48], degree centrality and betweenness centrality were combined with various seeding strategies to discover prominent nodes in an anonymized mobile phone network and two other social networks.

Catanese et al. [49] proposed a tool called LogAnalysis for scientific analysis of real phone call networks. This social network analysis tool provides both statistical and visual representation of real mobile phone network. Different centrality measures are featured in the statistical operation of this tool and they are used in ranking users according to how important they are in the phone call networks. The application of this tools is not only restricted to phone call networks but can also be applied in investigating criminal networks [50].

Han et al. [17] presented a program called iWander that runs on the mobile device of users (nodes). Random walk messages are sent to the users at fixed length of time. iWander determines the most influential users by computing the centrality of each node based on the total random walk messages received by each node. The authors carried out a theoretical analysis of the program and showed that influential nodes identified by iWander can regulate the spread of communicable diseases and can further be used to avoid a total epidemic. We summarised these methods in Table 1.

TABLE I. INFLUENTIAL NODES DETECTION TECHNIQUES IN MOBILE PHONE NETWORK

Authors	Technique	Components	Mobile phone Dataset
Kiss and Bichler[25]	Centrality measures and non-centrality	In-degree, out-degree, betweenness, closeness, Pagerank, SenderRank	Yes
Hinz *et al*., [27]	Centrality measures and non-centrality	Centrality measures and seeding techniques.	Yes
Catanese *et al*., [46]	*LogAnalysis*	Centrality measures.	Yes
Han *et al*., [48]	*iWander*	Centrality measures based on random walk sampling.	Yes

IV. CONCLUSION

We presented a background study on the detection of influential nodes and considered the key player problems as our focus in this study. We also discussed some of the previous studies related to discovering the most important and interesting nodes in a social network. We observed that centrality measures are more general and effective in the detection of influential nodes in social networks and mobile phone network. Though, results of using non-centrality measures are comparable to centrality measures approach. Thus, the idea of combining two or more centrality measures would improve detection and solve the influential nodes' detection problem.

It would be quite interesting if this study can cover other areas of communication network of wireless sensors, internet hubs and access points.

REFERENCES

[1] 'The world in 2014: ICT Facts and Figures. International Telecommunication Union.'

[2] C. A. R. Pinheiro, 'Community Detection to Identify Fraud Events in Telecommunications Networks', 2012.

[3] V. D. Blondel, A. Decuyper, and G. Krings, 'A survey of results on mobile phone datasets analysis', *EPJ Data Science*, vol. 4, no. 1, p. 1, 2015.

[4] J. D. Farley, 'Breaking Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assessment and Decision Making)', *Studies in Conflict & Terrorism*, vol. 26, no. 6, pp. 399–411, Nov. 2003.

[5] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. De Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, 'Analysis of a large-scale weighted network of one-to-one human communication', *New Journal of Physics*, vol. 9, no. 6, p. 179, 2007

[6] S. P. Borgatti, 'Identifying sets of key players in a social network', *Computational and Mathematical Organization Theory*, vol. 12, no. 1, pp. 21–34, Apr. 2006.

[7] F. Probst, 'Customer Relationship Management in a Digitally Connected World', 2013.

[8] X. Zhang, J. Zhu, Q. Wang, and H. Zhao, 'Identifying influential users in complex networks with community structure. Knowledge-Based Systems', *Knowledge-Based Systems*, vol. 42, pp. 74–84, 2013

[9] M. Ahsan, T. Singh, and M. Kumari, 'Influential user detection in social network during community detection', in *Cognitive Computing and Information Processing (CCIP), 2015 International Conference on*, 2015, pp. 1–6.

[10] A. Srinivas and R. L. Velusamy, 'Identification of influential users from social networks based on Enhanced Degree Centrality Measure', in *Advance Computing Conference (IACC), 2015 IEEE International*, 2015, pp. 1179–1184.

[11] Y. Zhang, Z. Wang, and C. Xia, 'Identifying Key Users for Targeted Marketing by Mining Online Social Network', 2010, pp. 644–649

[12] M. Eirinaki, S. P. S. Monga, and S. Sundaram, 'Identification of influential social networkers', *Int. J. Web Based Communities*, vol. 8, no. No. 2, pp. 136–158, 2012.

[13] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, 'Identifying the influential bloggers in a community', in *Proceedings of the 2008 international conference on web search and data mining*, 2008, pp. 207–218.

[14] R. Narayanam and Y. Narahari, 'A Shapley Value-Based Approach to Discover Influential Users in Social Networks', *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, pp. 130–147, Jan. 2011.

[15] F. Erlandsson, P. Bródka, A. Borg, and H. Johnson, 'Finding Influential Users in Social Media Using Association Rule Learning', Entropy, vol. 18, no. 5, p. 164, Apr. 2016.

[16] S. Singh, N. Mishra, and S. Sharma, 'Survey of various techniques for determining influential users in social networks', in Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013 International Conference on, 2013, pp. 398–403.

[17] B. Han, J. Li, and A. Srinivasan, 'Your friends have more friends than you do: Identifying influential mobile users through random-walk sampling', *IEEE/ACM Transactions on Networking*, vol. 22, no. 5, pp. 1389–1400, 2014.

[18] A. Abraham, Ed., *Computational Social Networks*. London: Springer London, 2012.

[19] M. Lieberman, 'Visualising big data: Social network analysis', in *Digital research conference*, 2014.

[20] D. M. B. Friedl and J. Heidemann, 'A critical review of centrality measures in social networks.', *Business & Information Systems Engineering*, vol. 2, no. 6, pp. 371–385, 2010.

[21] A. Bavelas, 'Communication patterns in task-oriented groups., 22(6)'. J. Acoust. Soc. Am, 1950.

[22] K. Okamoto, W. Chen, and X.-Y. Li, 'Ranking of closeness centrality for large-scale social networks', in *International Workshop on Frontiers in Algorithmics*, 2008, pp. 186–195.

[23] Ahmet Erdem Sarýyu¨ce, Kamer Kaya, Erik Saule, and U¨ mit V. C¸ atalyu¨rek, 'Incremental closeness centrality in distributed memory.', *Parallel Computing*, vol. 47, pp. 3–18., 2015.

[24] L. C. Freeman, 'A Set of Measures of Centrality Based on Betweenness', *Sociometry*, vol. 40, no. 1, p. 35, Mar. 1978.

[25] U. Brandes, 'A faster algorithm for betweenness centrality*', *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001

[26] J. R. Seeley, 'The net of reciprocal influence: A problem in treating sociometric data.', *Canadian Journal of Psychology*, vol. 3, pp. 234–240, 1949.

[27] L. Katz, 'A new status index derived from sociometric analysis', *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[28] K. Stephenson and M. Zelen, 'Rethinking centrality: Methods and examples', *Social Networks*, vol. 11, no. 1, pp. 1–37, 1989.

[29] C. Ballester, A. Calv´o-Armengol, and Y. Zenou, '"Who"s who in crime networks. Wanted: the key player'.', *Discussion Paper 4421.*, vol. 4421, no. CERP, 2004.

[30] S. Sarangi and E. Unlu, 'Key players and key groups in teams: A network approach using soccer data.' 2010.

[31] U. Temurshoev, 'Who's Who in Networks-Wanted: The Key Group', *Available at SSRN 1285752*, 2008.

[32] M. U. Ilyas and H. Radha, 'A KLT-inspired user centrality for identifying influential neighborhoods in graphs', in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, 2010, pp. 1–7.

[33] M. U. Ilyas and H. Radha, 'Identifying influential users in online social networks using principal component centrality', in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–5

[34] M. G. Everett and S. P. Borgatti, 'The centrality of groups and classes', *The Journal of mathematical sociology*, vol. 23, no. 3, pp. 181–201, 1999.

[35] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Shokrollahi, 'Improving detection of influential nodes in complex networks', Physica A: Statistical Mechanics and its Applications, vol. 436, pp. 833–845, Oct. 2015.

[36] M. M. Sathik and A. A. Rasheed, 'A centrality approach to identify sets of key players in an online weblog', *International Journal of Recent Trends in Engineering*, vol. 2, 2009.

[37] T. R. Zaman, 'Information extraction with network centralities: finding rumor sources, measuring influence, and learning community structure', Massachusetts Institute of Technology, 2011

[38] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. Mccoy, 'Constructing and Analyzing Criminal Networks', 2014, pp. 84–91.

[39] J. Shetty and J. Adibi, 'Discovering important users through graph entropy the case of enron email database', in *Proceedings of the 3rd international workshop on Link discovery*, 2005, pp. 74–81.

[40] F. Tutzauer, 'Entropy as a measure of centrality in networks characterized by path-transfer flow', *Social Networks*, vol. 29, no. 2, pp. 249–265, May 2007.

[41] D. Ortiz-Arroyo and D. M. A. Hussain, 'An Information Theory Approach to Identify Sets of Key Players', in *Intelligence and Security Informatics*, vol. 5376, D. Ortiz-Arroyo, H. L. Larsen, D. D. Zeng, D. Hicks, and G. Wagner, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 15–26

[42] J. Goldenberg, S. Han, D. R. Lehmann, and J. W. Hong, 'The role of hubs in the adoption process', *Journal of Marketing*, vol. 73, no. 2, pp. 1–13, 2009

[43] J. Heidemann, M. Klier, and F. Probst, 'Identifying key users in online social networks: A PageRank based approach', 2010.

[44] C. Canali and R. Lancellotti, 'A quantitative methodology based on component analysis to identify key users in social networks', *International Journal of Social Network Mining*, vol. 1, no. 1, pp. 27–50, 2012.

[45] F. C. Perkins III, *System and method for processing call detail records*. Google Patents, 2002

[46] A. McDiarmid, S. Bell, J. Irvine, and J. Banford, 'Nodobo: Detailed mobile phone usage dataset', *Unpublished paper, accessed at http://nodobo. com/papers/iet-el. pdf on*, pp. 9–21, 2013.

[47] C. Kiss and Bichler Martin, 'Identification of influencers—measuring influence in customer networks', *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.

[48] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker, 'Seeding strategies for viral marketing: An empirical comparison', *Journal of Marketing*, vol. 75, no. 6, pp. 55–71, 2011.

[49] S. Catanese, E. Ferrara, and G. Fiumara, 'Forensic analysis of phone call networks', *Social Network Analysis and Mining*, vol. 3, no. 1, pp. 15–33, Mar. 2013.

[50] E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara, 'Detecting criminal organizations in mobile phone networks', *Expert Systems with Applications*, vol. 41, no. 13, pp. 5733–5750, 2014.