Centrality for graphs with numerical attributes

Oualid Benyahia
Universite Jean Monnet
Laboratoire Hubert Curien UMR CNRS 5516
18 rue Lauras 42000 Saint-Etienne
Email: oualid.benyahia@univ-st-etienne.fr

Christine Largeron
Universite Jean Monnet
Laboratoire Hubert Curien UMR CNRS 5516
18 rue Lauras 42000 Saint-Etienne
Email: Christine.Largeron@univ-st-etienne.fr

***Abstract*—Identification of important actors in social networks is a hard task but with various interesting applications such as in information recommendation or for viral marketing. Existing centrality measures evaluate the importance of an actor in considering only the structural positions regardless of prior information on these actors such as their popularity, accessibility or behavior. A few measures have been proposed for weighted networks, notably the three common measures of centrality: degree, closeness, and betweenness. However, these extended versions have solely focused on the weights of ties and not on the attributes of nodes. This article proposes generalizations that combine these both aspects. We present a set of measures, based on conventional centrality indicators, suited to weighted attributed graphs where the nodes are characterized by attributes. We illustrate the benefits of this approach on real attributed graphs. Experiments have validated the contribution of the links weights and attributes, especially for the detection of information broadcasters in social networks.**

I. Introduction

Several indicators like the centrality measures, initially developed in social network analysis, allow to identify the actors which play an important role in a network. The advantage of these measures is their ease of implementation and the possibility to leverage local information rather than to exploit the whole graph. However, they have been designed for networks represented by simple graphs where the actors correspond to the nodes and their relationships to the links. Nowadays, social networks become more complex. In such context, we propose to improve the identification of important actors by leveraging not only the structural properties of the network but also information describing these actors.

Thus, the aforementioned networks, in particular information networks, can be represented by attributed weighted graphs where each node is associated with features that can be used to asses its importance or influence. Exploiting this additional information on actors in propagation models, has been efficient. Therefore, we suggest that these attributes can also be integrated in centrality measures, giving rise to new definitions more suited for attributed graphs. To identify prevalent important or influential actors in complex networks, our aim is to adapt the classical centrality indicators to handle graphs where weighted links quantify the relations between nodes and features describe the nodes.

The sequel of the article is organized as follows. Section 2 is dedicated to related works. Section 3 describes the introduced measures and section 4 presents experiment results conducted on real co-publication networks which confirm the efficiency of the proposed measures.

II. Related works

In the literature, the problem of identifying important or influential actors in a network has been explored in two main directions, firstly with diffusion models and secondly with indicators which characterize the position of an actor in the network from different points of view.

The methods in the first family dedicated to the study of propagation in social networks, are based on the *word-of-mouth* principle that the behavior or the opinion of an actor depends strongly on those of his close social circle ([1], [2]). The problem can be formulated as a discrete optimization problem, known as *Influence Maximization* which consists in finding a set of nodes (referred to as seeds) such that under the influence model, the expected number of nodes activated by the seeds is the largest possible. Kempe *et al.* have proved that this optimization problem is NP-hard, and proposed a *Greedy Approximation Algorithm* applicable to all propagation models ([3]), which guarantees that the influence spread is very close to the optimal influence spread, under the submodularity assumptions. Following the same idea, several works have exploited complementary information on nodes and obtained best results ([4], [5], [6], [7]). Consequently, even if these diffusion models are beyond the scope of this paper, we retain from these researches, the idea that the actors characteristics allow to improve the evaluation of their importance.

The second kind of approaches to detect important actors, less costly in time processing, consist in characterizing the role or the position occupied by an actor in the network by means of centrality indicators ([8], [9]). These measures allow to compare the position of an actor relatively to others from different ways. Thus, the degree centrality characterizes individuals with important number of links i.e. direct neighbors while the closeness centrality those that can easily reach other members of the network. The betweenness centrality allows to identify actors that are more likely to be intermediary between other actors in the network. Finally, the eigenvector centrality and its variant, the pageRank, detect the most influential actors who are connected to other important and highly connected members of the network. These measures can be computed on network represented as a directed or undirected graph. For example, the degree centrality can be defined by considering all neighbors of a node, its successors or its predecessors. Some measures can be applied to weighted or unweighted graphs. Thus, the eigenvector centrality is defined for graphs with weighted edges and in that case, the edge weight can be interpreted as the intensity of interactions between two nodes. Likewise the degree and the closeness centralities have been defined for weighted graphs ([10], [11]).

The vertices characteristics are not considered in previous researches, except in [12] which is probably the most related work to our own article. Indeed, the authors considered the case of attributed graphs where categorical attributes like the gender or the month of birth are used to define groups of nodes. They extend classical centrality measures by using two well-known metrics: the E-I homophily index [13], and the Gould and Fernandez brokerage metrics [14]. Specifically, the E-I index is seen as a partitioning degree centrality and thereby is generalized to other centrality measures such as the closeness centrality and the eigenvector centrality, leading to the definition of a closeness E-I and an eigenvector E-I. Likewise, the betweenness centrality is extended using the Gould and Fernandez brokerage measures which divide the ego network into five memberships. This approach is then generalized to the whole network.

In contrast with the work of [12], the approach proposed in this article, handles numerical attributes that are more likely to characterize the importance or the influence of an actor. These attributes are processed as a prior information, given under the form of numeric measures related for example to the fame, the experience or the notoriety of the actor. For example in the domain of scientific publications, the nodes attributes can be expressed by the number of publications (or co-publications), the number of collaborators (students, Phd, etc.), the number of associated scientific projects or institutions, the number of active years or the number of rewarded papers in a scientific area. In the context of online social networks, the attributes values can be for instance the age of the user account, the frequency of account activity, the number of different terminal equipments used to access the account, etc. In a movies network, the attributes characterizing an actor can be the number of movies he played in, the total budget of the movie or the actor's income, the number of prize awarded, etc.

To the best of our knowledge, none of the classical centrality measures allows to exploit these contextual information in addition to the relational information in attributed graphs, and this is what we aim to achieve in the next section.

III. MEASURING INFLUENCE BY CENTRALITY INDICATOR

An information network can be represented as an attributed graph $G = (V, E)$ where V is the set of nodes and $E \subset V \times V$ is the set of edges. We suppose that each node $v_i \in V$ is associated to a numerical vector $Y^i = (y_1^i, y_2^i, \ldots, y_L^i)^T$ describing its attributes ([15]). In the following, y_l^i is the value of the attribute $l \in L$ observed for the node v_i and w_i its global weight computed by means of the norm of the vector Y^i, thus $w_i = \|Y^i\|$. w_{ij} denotes the weight of the edge between node v_i and node v_j and describes the intensity of their relation.

The nodes attributes y_l^i, considered in this work, correspond to prior information that indicates the notoriety or importance of an actor, independently of the structure or the topology of the network. Thus, their choice depend strongly on the nature of the studied network.

In order to compare the relative position occupied by a node in a simple graph, several centrality measures have already been defined: degree centrality, closeness centrality, betweenness centrality or prestige centrality (eigenvector centrality and pagerank). In the following, we propose to extend these measures, initially defined for simple graphs, so as to take into account features or weights characterizing the nodes and the links.

A. degree centrality

The degree centrality measures the relative importance of a node by counting its direct links in the graph, after normalization by the size of the network ($Degree(vi)$) [8]. This measure, based on the local structure around the node i.e. its neighborhood, is the simplest way to measure the centrality. Degree centrality has been extended to the case of weighted and directed graphs [10]. In a directed network, a node may have a different number of outgoing and incoming ties, and therefore, the degree is split into *out-degree* (deg_{out}) and *in-degree* (deg_{in}), respectively.

$$Degree(v_i) = \frac{deg(v_i)}{|V| - 1} \tag{1}$$

In a weighted graph, the degree centrality is defined by means of links weights as follows:

$$WEDegree(v_i) = \sum_{v_j \in out(v_i)} w_{ij} \tag{2}$$

where w_{ij} is the weight of the edge between node v_i and node v_j, $out(v_i)$ is the set of successors v_j of node $v_i \in V$. In [11] another degree centrality measure is proposed for weighted graphs:

$$WEOpsahlDegree(v_i) = (deg_{out}(v_i))^{(1-\alpha)} \cdot \left(\sum_{v_j \in out(v_i)} w_{ij} \right)^{\alpha} \tag{3}$$

where the tuning parameter $\alpha \in [0, 1]$ is used to set the relative importance of the number of ties compared to the ties weights and is equal to $0,5$ to get a balance between links and their weights.

In the case of attributed graphs, we introduce three variants of these measures $WNDegree$, $WNEDegree$ and $WNEOpsahlDegree$ corresponding respectively to weighted or not weighted graphs:

$$WNDegree(v_i) = w_i \cdot Degree(v_i) \tag{4}$$

$$WNEDegree(v_i) = w_i \cdot WEDegree(v_i) \tag{5}$$

$$WNEOpsahlDegree(v_i) = w_i \cdot WEOpsahlDegree(v_i) \tag{6}$$

B. Closeness centrality

Closeness is defined as the inverse of the farness, which in turn, is the sum of distances to the other nodes. According to the definition, a node is considered as important if it can rapidly reach the other nodes of the graph ([9], [16]). The usual measure is defined by the inverse of the sum of the geodesic distances of a given node to others nodes:

$$CCentr(v_i) = \frac{1}{\sum\limits_{\substack{v_j \in V \\ j \neq i}} |ShortPath(v_i, v_j)|} \tag{7}$$

where $ShortPath(v_i, v_j)$ denotes the shortest path between node v_i and node v_j and $|ShortPath(v_i, v_j)|$ is the length of this path (number of links in the path). Conventionally, $|ShortPath(v_i, v_j)|$ is set to $|V|$ or ∞ if such path does not exist so that two nodes which belong to different components do not have a finite distance between them.

For weighted graphs, a variant that incorporates the links weights using Dijsktra algorithm is proposed in [11] but it is time-consuming and consequently not suited for online applications. For this reason, we propose to compute the sum of the normalized geodesic distances between a given node v_i and the other nodes of the network:

$$CWECentr(v_i) = \sum_{\substack{v_j \in V \\ j \neq i}} \frac{\sum\limits_{e \in ShortPath(v_i, v_j)} w(e)}{|ShortPath(v_i, v_j)|} \tag{8}$$

where $w(e)$ is the weight of the link $e \in E$ in the path. If such path does not exist, the numerator term has a null value and $|ShortPath(v_i, v_j)|$ is set to $|V|$ or ∞. For an attributed network, we propose to compute the closeness centrality in one of the following ways, depending whether the graph is weighted or not:

$$CWNCentr(v_i) = w_i \cdot CCentr(v_i) \tag{9}$$

$$CWNECentr(v_i) = w_i \cdot CWECentr(v_i) \tag{10}$$

C. Betweenness centrality

According to the betweenness measure, a node is important, if it is located on a great number of geodesics paths between the other nodes. Formally, it is equal to the number of shortest paths between all pairs of nodes that pass through that node ([9], [8]):

$$BCentr(v_i) = \sum_{\substack{(v_k, v_j) \in V \times V \\ i \neq k \neq j}} |g_{kj}(v_i)| / |g_{kj}| \tag{11}$$

where $g_{kj}(v_i)$ (respectively $|g_{kj}(v_i)|$) is the set of the shortest paths (respectively the cardinality of the set of shortest paths) between nodes v_k and v_j that pass through v_i and g_{kj} (respectively $|g_{kj}|$) is the set of all shortest paths (respectively the cardinality of the set of shortest paths) between the nodes v_k and v_j. Thereby, the more there are paths passing through a node the more it is important. The nodes with a high betweenness play an important role into the communications or transfers in the network since they control the flow between non adjacent nodes.

We adapt this measure for weighted graphs in the following manner:

$$BWECentr(v_i) = \sum_{\substack{(v_k, v_j) \in V \times V \\ i \neq j \neq k}} \frac{\sum\limits_{S \in g_{kj}(v_i)} \sum\limits_{e \in S} w(e)}{\sum\limits_{S \in g_{kj}} \sum\limits_{e \in S} w(e)} \tag{12}$$

where $w(e)$ is the weight of the link $e \in E$ in a shortest path $S \in g_{kj}$ between the two nodes v_k and v_j.

When the nodes are described by a set of attributes, the measure is weighted by the node's weight w_i, leading to different definitions depending whether the graph is weighted or not:

$$BWNCentr(v_i) = w_i \cdot BCentr(v_i) \tag{13}$$

$$BWNECentr(v_i) = w_i \cdot BWECentr(v_i) \tag{14}$$

D. Eigenvector centrality and PageRank

Eigenvector centrality is based on the idea that the score of a node is higher if it is connected to nodes having a high score than if it is connected to nodes with a low score ([17], [18]). Usually, the *eigenvector centrality* is recursively computed by the following formula:

$$EVCentr(v_i) = \frac{1}{\lambda_1} \cdot \sum_{v_j \in V} a_{ij} \cdot EVCentr(v_j) \tag{15}$$

where $A = \{a_{ij}\}$ is the adjacency matrix of the graph and λ_1 is the largest eigenvalue obtained as solution of the equation $AX = \lambda X$. The measure *EVWNECentr* is computed on the weighted adjacency matrix for the weighted graphs.

The *PageRank* is a variant of *eigenvector centrality* ([19], [18]). It was initially introduced to measure the popularity of Web pages and is usually defined by the equation:

$$\begin{aligned} PRankCentr(v_i) = &(1 - \beta) W_0 \\ &+ \beta \sum_{v_j \in V} a_{ji} \frac{PRankCentr(v_j)}{deg(v_j)} \end{aligned} \tag{16}$$

where W_0 is generally fixed uniformly and is equal to $\frac{1}{|V|}$ for all nodes.

In the case of attributed graph, we adopt a custom formulation *PRankWNECentr* of the *PageRank* ([20], [21]) in which the nodes weights w_i computed on the attributes are used instead of the uniform weights.

IV. Experimentation

Several experiments have been carried out on real attributed networks so as to evaluate the benefits of the proposed measures.

A. Dataset

For experimentations, we used the dataset from [4] which allows to generate graphs with ground truth. It has been extracted from the *Arnetminer* academic research system [1] and concerns 640134 authors and 1554643 co-publications on different topics and we have chosen three of them: *Data mining*, *Information Retrieval (IR)* and *Bayesian Network*. Formally, for a chosen research topic, we obtain a graph $G = (V, E)$ where the set of nodes V represents the authors and each author v_i is characterized by an attribute w_i corresponding to his number of publications. The set E of edges represents the co-publication links weighted by the number of articles w_{ij} co-written by the two authors v_i and v_j.

[1] http://www.arnetminer.org.

TABLE I. STATISTICS ON THE CO-PUBLICATIONS GRAPHS.

Graphs	number of Nodes	number of edges
Data-Mining	679	1687
Information Retrieaval (IR)	657	1907
Bayesian Network	554	1238

As shown in Table I, the graph associated to the Data Mining topic is an undirected graph with 679 nodes and 1687 edges. The graph dedicated to Information Retrieval has 657 nodes and 1907 edges and for the Bayesian Network topic, it has 554 nodes and 1238 co-publication links.

B. Evaluation

To assess the centrality measures defined in section III, we consider two other indicators as ground truth references of the author's influence: the *H-index* and the number of *Citations* of an author. These indicators are extracted from Arnetminer [2]. The *H-index* (HIRSCH index) [22] is an index used to estimate the rank of research scientists: an author has an index of h if h of his publications have at least h citations and the other publications have not more than h citations each.

Thereby, after obtaining centrality measures for the authors, we are able to compare the ranking of the authors according to a given measure with those provided by the two ground truth indicators: the *H-index* and the number of *Citations*. For each indicator, we chose to assess the top 20 best ranked authors. To determine the accuracy of a centrality measure, we compute the *Jaccard index* between the ordered list obtained with the centrality measure and the list ordered either by the *H-index* or the number of *Citations*.

Otherwise, we use also the *Precision* and *Recall* rates to evaluate the ranking provided by each centrality measure. As the total number of authors returned by each measure is equal to the number of influential authors defined by *H-index* (or number of *Citations*), the *Precision* and *Recall* rates are identical.

C. Results and analysis

Table II, Table III and Table IV show the results (Jaccard index, Precision/Recall) of the experiments, conducted respectively on the three graphs (Data-Mining, Information Retrieval (IR) and Bayesian Networks), for the five families of centrality measures (18 measures): Degree, Closeness, Betweenness, PageRank and Eigenvector. The results are computed by means of conventional centrality measures and by the proposed variants, denoted (*) in the tables, that take into account the weights of the links and the attributes of the nodes.

The score of the majority of centrality measures is better when compared to $H-index$ than to the number of $Citations$, notably in terms of Precision/Recall in the three graphs.

We can observe that for the Data-Mining graph, the best score according to the Jaccard index and the Precision/Recall rate are obtained by the authority measures, i.e. the conventional pageRank (**PRankCentr**) and the eigenvector centrality variant (**EVWNECentr***) if we consider the H-index as ground truth. If the number of Citations is taken as ground truth, the best scores are provided by the conventional degree centrality measures in the weighted graph (**WEDegree** and **WEOpsahlDegree**) and the new measures (**WNEDegree*** and **WNEOpsahlDegree***) as well as **EVWNECentr***. This result emphasizes the contribution of the links weights (number of co-publications between authors) and the preponderance of the authority measures to infer the influence of an author in the Data-Mining graph.

For the Information Retrieval graph, the variant measure **WNDegree*** of degree centrality obtains the best accuracy compared to the reference indicators (H-index and number of Citations). Moreover, it is followed by other measures which take also into account the nodes attributes (WNEOpsahlDegree*, CWNCentr*, BWNCentr*, BWNECentr*), including the variant CWNCentr* of the closeness indicator which gives generally low scores.

In the Bayesian Network graph, the highest accuracy is obtained by the three variants **WNEDegree***, **WNEOpsahlDegree*** and **BWNCentr*** compared to the two reference indicators (H-index and number of Citations). Thus, we can conclude that the weights and attributes allow to improve the ability to infer the influential authors.

In conclusion, the proposed variants are more suited than the usual measures to infer the influential users in weighted graphs with attributes. We must notice that the closeness centrality and its variants give generally the lowest scores because they are not well adapted for the studied graphs. In contrast, the degree centrality variants, especially **WNEDegree*** and **WNEOpsahlDegree*** that incorporate the links weights and nodes attributes, give generally a good estimation of the authors influence in the three graphs and are efficient in terms of processing times. In fact, they have the advantage of being computed locally and consequently, they can be rapidly obtained compared to the other measures. For the other measures (betweenness and eigenvector), the results are improved when the weights of the links are considered, particularly if the H-index is used as ground truth. Moreover, this improvement is more important if we take into account the attributes that describe the nodes, in addition of the weights of the links.

The computing complexity of the proposed variants is of the same order of magnitude as those of the classical centrality measures. In fact the new variants of degree, pagerank and eigenvector centralities exhibit processing times equivalent as those of their corresponding measures designed for unweighted graphs without attributes. For the degree centrality, the sum is computed over numerical weights instead of binary weights. For the eigenvector and pagerank centralities, the same iterative process is conducted to reach a final stationary distribution. For the closeness and betweenness centrality, the computing process is more complicated and the challenge is to compute the shortest paths between every pair of nodes. For these variants, we also need to get the cumulative weights of the shortest paths and then their average, and this is done in a linear time compared to conventional algorithms. This yields to an overall complexity that is in the same order of magnitude as for a simple graph.

[2]http://arnetminer.org/person-ranklist/hindex/89.

TABLE II. SCORE RESULTS OF INFLUENCE MEASURES ON DATA-MINING GRAPH.

	Jaccard (H-Index)	Jaccard (Citations)	P/R (H-Index)	P/R (Citations)
Degree	0.3333	0.1765	0.5	0.3
WNDegree(*)	0.3333	0.1765	0.5	0.3
WEDegree	0.3333	**0.2121**	0.5	**0.35**
WEOpsahlDegree	0.3333	**0.2121**	0.5	**0.35**
WNEDegree(*)	0.3793	**0.2121**	0.55	**0.35**
WNEOpsahlDegree(*)	0.3793	**0.2121**	0.55	**0.35**
CCentr	0.3333	0.1765	0.5	0.3
CWNCentr(*)	0.3333	**0.2121**	0.5	**0.35**
CWECentr(*)	0.0	0.0256	0.0	0.05
CWNECentr(*)	0.25	0.1765	0.4	0.3
BCentr	0.3333	0.1429	0.5	0.25
BWNCentr(*)	0.3793	0.1765	0.55	0.3
BWECentr(*)	0.2903	0.1765	0.45	0.3
BWNECentr(*)	0.3793	0.1765	0.55	0.3
PRankCentr	**0.4286**	0.1765	**0.6**	0.3
PRankWNECentr(*)	0.2903	0.1765	0.45	0.3
EVCentr	0.2121	0.1111	0.35	0.2
EVWNECentr(*)	**0.4286**	**0.2121**	**0.6**	**0.35**

TABLE III. SCORE RESULTS OF INFLUENCE MEASURES ON INFORMATION RETRIEVAL (IR) GRAPH.

	Jaccard (H-Index)	Jaccard (Citations)	P/R (H-Index)	P/R (Citations)
Degree	0.25	0.2121	0.4	0.35
WNDegree(*)	**0.4286**	**0.3333**	**0.6**	**0.5**
WEDegree	0.25	0.2121	0.4	0.35
WEOpsahlDegree	0.2903	0.25	0.45	0.4
WNEDegree(*)	0.2903	0.2121	0.45	0.35
WNEOpsahlDegree(*)	0.3793	0.2903	0.55	0.45
CCentr	0.25	0.2121	0.4	0.35
CWNCentr(*)	0.3793	0.2903	0.55	0.45
CWECentr(*)	0.0	0.0	0.0	0.0
CWNECentr(*)	0.2121	0.1765	0.35	0.3
BCentr	0.1765	0.1429	0.3	0.25
BWNCentr(*)	0.3793	0.2903	0.55	0.45
BWECentr(*)	0.25	0.2121	0.4	0.35
BWNECentr(*)	0.3793	0.2903	0.55	0.45
PRankCentr	0.25	0.2121	0.4	0.35
PRankWNECentr(*)	0.3793	0.25	0.55	0.4
EVCentr	0.25	0.2121	0.4	0.35
EVWNECentr(*)	0.2903	0.25	0.45	0.4

TABLE IV. SCORE RESULTS OF INFLUENCE MEASURES ON BAYESIAN NETWORKS GRAPH.

	Jaccard (H-Index)	Jaccard (Citations)	P/R (H-Index)	P/R (Citations)
Degree	0.2121	0.1429	0.35	0.25
WNDegree(*)	0.3333	0.25	0.5	0.4
WEDegree	0.25	0.1765	0.4	0.3
WEOpsahlDegree	0.2903	0.2121	0.45	0.35
WNEDegree(*)	**0.3793**	0.2903	**0.55**	0.45
WNEOpsahlDegree(*)	**0.3793**	0.2903	**0.55**	0.45
CCentr	0.2121	0.1765	0.35	0.3
CWNCentr(*)	0.3333	0.25	0.5	0.4
CWECentr(*)	0.0526	0.0256	0.1	0.05
CWNECentr(*)	0.2903	0.2121	0.45	0.35
BCentr	0.2903	0.25	0.45	0.4
BWNCentr(*)	**0.3793**	**0.3333**	**0.55**	**0.5**
BWECentr(*)	0.25	0.25	0.4	0.4
BWNECentr(*)	0.3333	0.25	0.5	0.4
PRankCentr	0.3333	0.25	0.5	0.4
PRankWNECentr(*)	0.25	0.2121	0.4	0.35
EVCentr	0.1111	0.0811	0.2	0.15
EVWNECentr(*)	0.3333	0.25	0.5	0.4

V. CONCLUSION

In this article, we propose a set of centrality measures to estimate the importance of the actors in complex social networks. These variants, induced from the classical centrality measures, exploit not only the network structure but also incorporate weights that evaluate the strength of the relationships between the actors and attributes which describe these last ones. A serie of experiments was carried out on three co-publications graphs. These experiments confirm the interest of these new measures to identify actors considered as the most influent in their research field. The results show that these variants, suited for weighted attributed graphs, are more efficient than the usual measures, in particular the weighted variants of the degree centrality, with processing times in the same order of magnitude as those required by the classical measures.

REFERENCES

[1] D. J. Crandall, D. Cosley, D. P. Huttenlocher, J. M. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *SIGKDD*, 2008, pp. 160–168.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *SIGKDD*, 2008, pp. 7–15.

[3] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD'03)*, Washington, USA, 2003, pp. 137–146.

[4] J. Tang, J. Sun, , C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 807–816.

[5] K. Saito, K. Ohara, , Y. Yamagishi, M. Kimura, and H. Motoda, "Learning diffusion probability based on node attributes in social networks," in *Proceedings of the 19th international conference on Foundations of intelligent systems*, ser. ISMIS'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 153–162.

[6] C. Lagnier, L. Denoyer, E. Gaussier, and P. Gallinari, "Predicting information diffusion in social networks using content and user's profiles," in *Proceedings of the 35th European Conference on Advances in Information Retrieval*, ser. ECIR'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 74–85.

[7] A. Guille, H. Hacid, and C. Favre, "Predicting the temporal dynamics of information diffusion in social networks," *CoRR*, vol. abs/1302.5235, 2013.

[8] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[9] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge Univ Pr, 1994.

[10] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2004.

[11] T. Opsahl, F. Agneessens, , and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245–251, jul, 2010.

[12] M. G. Everett and S. P. Borgatti, "Categorical attribute based centrality: E-I and G-F centrality." *Social Networks*, vol. 34, pp. 562 – 569, 2012.

[13] D. Krackhardt and R. N. Stern, "Informal networks and organizational crises: An experimental simulation." *Social Psychology Quarterly*, vol. 51, no. 2, pp. 123 – 140, 1988.

[14] R. V. Gould and R. M. Fernandez, "Structures of mediation: A formal approach to brokerage in transaction networks." *Sociological Methodology*, vol. 19, no. 1, p. 89, 1989.

[15] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *Proceedings of the VLDB Endowment*, vol. 2, pp. 718–729, 2009.

[16] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Operations Research*, vol. 12, no. 3, pp. 450–459, 1964.

[17] P. Bonacich and L. Paulette, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, vol. 23, no. 3, pp. 191–201, jul, 2001.

[18] R. Ghosh and K. Lerman, "Predicting influential users in online social networks," in *Proceedings of KDD workshop on Social Network Analysis (SNA-KDD)*, July 2010.

[19] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the seventh international conference on World Wide Web 7*, ser. WWW7, 1998, pp. 107–117.

[20] B. Sergey, R. Motwani, , L. Page, and T. Winograd, "What can you do with a web in your pocket?" *IEEE Data Engineering Bulletin*, vol. 21, pp. 37–47, 1998.

[21] G. Jeh and J. Widom, "Scaling personalized web search," in *WWW '03: Proceedings of the 12th international conference on World Wide Web*. New York, NY, USA: ACM Press, 2003, pp. 271–279.

[22] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46,, pp. 16 569–16 572, 2005.