View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AIS Electronic Library (AISeL)

Communications of the Association for Information Systems

Volume 43 Article 21

9-2018

# A New Influence Measure Based on Graph Centralities and Social Network Behavior Applied to Twitter Data

Romain Boulet
*University of Lyon*, romain.boulet@univ-lyon3.fr

Jean-Fabrice Lebraty
*University of Lyon*

Follow this and additional works at: https://aisel.aisnet.org/cais

Recommended Citation
Boulet, Romain and Lebraty, Jean-Fabrice (2018) "A New Influence Measure Based on Graph Centralities and Social Network Behavior Applied to Twitter Data," *Communications of the Association for Information Systems*: Vol. 43 , Article 21.
DOI: 10.17705/1CAIS.04321
Available at: https://aisel.aisnet.org/cais/vol43/iss1/21

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Communications of the Association for Information Systems

Research Paper DOI: 10.17705/1CAIS.04321 ISSN: 1529-3181

# A New Influence Measure Based on Graph Centralities and Social Network Behavior Applied to Twitter Data

**Romain Boulet**
University of Lyon, UJML3
iaelyon School of Management, Magellan
*romain.boulet@univ-lyon3.fr*

**Jean-Fabrice Lebraty**
University of Lyon, UJML3
iaelyon School of Management, Magellan

**Abstract:**

In this paper, we use graph theory to explore concepts of influence in socialized groups. When analyzing social networks, centrality indicators make it possible to assess the power of an individual. We discuss various centrality indicators and focus on degree and betweenness. After observing a strong correlation between them, we propose defining new assessments based on a decorrelation method that we characterize from different mathematical perspectives (algebraic, probabilistic, and statistical). We apply this theoretical framework to a network of tweets about the Uber versus taxi conflict, which took place in June, 2015, and for which we detected different influential individuals.

**Keywords:** Social Networks, Graph Theory, Measures, Influence.

This manuscript underwent peer review. It was received 12/20/2016 and was with the authors for 9 months for 2 revisions. Thomas Case served as Associate Editor.

# 1 Introduction

In April, 2016, a French startup called Linkurious attracted worldwide celebrity from offering a method of visualizing graphs that enabled data analysts to analyze a large number of documents from a Panamanian law firm. Linkurious' method helped investigators reveal the links between those involved in the Panama Papers case. This example illustrates how globalization has expanded social networks and emphasizes the importance of having methods that give meaning to them and that can analyze the movement of information, rumors, and disinformation (Qin, Cai, & Wangchen, 2015).

However, while much work has developed ways to analyze social networks (Cao, Basoglu, Sheng, & Lowry, 2015; Cross, Borgatti, & Parker, 2002; Freeman, 2004; Scott, 2000, Granovetter, 1973), which refer to social structures that comprise ties (or links) between social actors, the unprecedented scale of social networking applications that we see today (more than two billion people use them; indeed, since they appeared around 10 years ago, they have dethroned the Web and email to represent the major category of applications on the Internet) and advances in processing power that allow small organizations and even individuals to analyze large volumes of data call for new analytical methods.

As such, in this paper, we develop new methods to analyze social networks. Specifically, we use graph theory as the basis for our work as many previous studies have done (Abilhoa & de Castro, 2014; Nettleton, 2013). We also focus on the concept of influence given that researchers have highlighted the relationship between social networks and influence in finding that "fake news" spreads faster than true news (Langin, 2018). Overall, we address the following research question:

> **RQ**: Can we define a new measure of influence based on graph theory?

This paper proceeds as follows: in Section 2, we review social networking concepts and graph theory (including centralities in graphs). In Section 3, we develop our theoretical framework. In developing this framework, we 1) observe that various centrality measures that graph theory provides (e.g., degree or betweenness) correlate to one other and, thus, propose a corrected betweenness measure to identify what we call "quiet relays", 2) propose a measure for a message to trigger behavior in this kind of network, and 3) introduce a unified influence measure that considers the two previous measures and the measure of degree in graph theory. In Section 4, we describe a particular situation (i.e., actions against Uber), and we apply the methodology developed in Section 3 to Twitter data. Subsequently, we show ways to use graph theory to analyze influence in social networks. Finally, in Section 5, we discuss our study's managerial implications and strategies for increasing influence in social networks and conclude the paper.

# 2 Literature Review

## 2.1 Using Graph Theory to Analyze Social Networks

Graph theory, which researchers in France such as Claude Berge (1958) in his book *Graphs and Hypergraphs* developed in the second half of the 20th century, provides the mathematical basis for analyzing social networks. The rise of social network analysis took place in the same period. Sociologists such as White, Granovetter, Burt, or Degenne and Forsé (1994) defined and studied the key concepts, such as weak links (Granovetter, 1973), centralities (Freeman, 1979), and structural holes (Burt, 1992). The introduction of mathematical tools, based on graph theory or matrices (both of which fall under algebraic graph theory), allowed researchers to document sociological concepts and develop methods at the crossroads of mathematics, computer science, and sociology. Two sets define a simple graph: a set of vertices (or nodes) and a set of edges (or links) that connect two vertices. In social networks, vertices represent individuals and edges represent (social) links between individuals. In this section, we discuss the common properties of social networks, various concepts of centralities, and the notion of community.

A social network is not a stochastic network: individuals do not randomly create the links between them but forge them through "a social process". As such, one can expect that all types of social networks—whatever their nature (friendship, professional, etc.)—take a common form. Most importantly, all these networks have low density (the graphs are sparse), a proportionally small number of links, and many have only 10 percent or less of all possible links. A network that features the small-world effect—which Travers and Milgram (1969) demonstrate and Watts (2003) explains well—represents a common structural form. Such networks feature a low average length of the shortest paths (a small number of intermediaries is required to connect two individuals) and a high degree of local connectivity (friends of my friends are also my friends). Researchers have also identified other structural constants, such as the presence of a giant connected component and the power-law modeling degree distribution (Barabási, 2009).

An individual's position in a network can give the individual a level of power or influence, which the concept of centrality addresses. Researchers have defined centrality in several ways that correspond to a specific notion of power. First, degree centrality, measured by the number of incident links on a node that represents an individual, assesses an individual's popularity: an individual has power because the individual has many social ties. Simply, the individual is popular. If all individuals with high degree centralities are tightly interconnected, they form a "rich club", which means a unique central and influential cluster (in the sense of degree centrality) in the network (Colizza, Flammini, Serrano, & Vespignani, 2006). But situations in which one does not need to have a high degree centrality to have power exist, such as when one occupies a strategic position in the network. Betweenness centrality estimates this situation, and one determines it by calculating, for a given node, the proportion of the shortest paths that pass through the node. Thus, a node with high betweenness centrality is one through which a significant number of shortest paths pass—a key node in connectivity and network communication. Being at the network center also provides power: the node can quickly contact other network members. Closeness centrality, which refers to the inverse of the average distance from one node to other nodes, measures the average proximity of a node to the other nodes. Eigenvector centrality (or PageRank) measures centrality such that the centrality of an individual is proportional to the centralities of individuals to which the individual is linked (my influence depends on the influence of my friends); eigenvector centrality relies on Perron-Frobenius theorem and is obtained from the eigenvector associated with the largest eigenvalue of the adjacency matrix of the graph.

Because social networks show a small-world effect (they feature a higher density of local rather than global connections), one can expect communities to emerge (i.e., sets of nodes that have a high density of links but few links between the sets). The abundance of methods and algorithms to detect communities makes it impossible to comprehensive review here. Nonetheless, we can mention some typical methods; Fortunato (2010) provides an excellent overview of community-detection methods. The spectral partitioning method (Luxburg, 2007) is based on embedding the nodes in a Euclidean space thanks to eigenvectors associated to small eigenvalues of the Laplacian matrix associated with the graph. The algorithm of Girvan and Newman (2002, 2004) comprises an optimization of modularity that measures the quality of partitioning of a network into communities. Pons and Latapy (2005) used random walks in the graph to detect communities. Other methods build on stochastic block models (Amini, Chen, Bickel, & Levina, 2013), overlapping communities (Ahn, Bagrow, & Lehmann, 2010), and the use of node attributes (Yang, McAuley, & Leskovec, 2013). Breaking down a network into communities is not unique: for instance, one can view an individual as belonging to several communities or as belonging to one community formed from several closely linked communities.

## 2.2 Social Networks

In this section, we first discuss some classic networks that have contributed to advances in graph structural analysis. These networks, which mostly date from the 20th century, do not rely on the Internet. As such, we subsequently focus on Internet-based applications designed for social networks and studies performed on these networks.

The analysis methods for social networks result from studies that mainly sociologists performed and that dealt with structural analysis such as the small-world phenomenon and community detection. For instance, to apply small-world phenomenon theories, Duncan Watts relied on many networks, such as the large social network of collaborations between Hollywood actors, and showed that, in a network with more than 200,000 nodes and 0.027 percent density, the median of shortest paths was 3.65 (Watts & Strogatz, 1998). Networks such as the karate club (Zachary, 1977) or the network of meetings between American football teams (Park & Newman, 2005) have a known community structure; therefore, researchers have widely used them in the search for partitioning methods thanks to the possibility of a comparison of the results with the known structure (Newman, 2004). Barabási and Bonabeau (2003) demonstrated that many networks are scale free, such as the network of scientific collaborations (Barabási et al., 2002) and the network of sexual relationships (Freiesleben de Blasio, Svensson, & Liljeros, 2007). Finally, we note that social networks are networks of interactions between people in the same way as biological networks are networks of interactions between proteins, as neural networks are networks of interactions between neurons, as power grids are networks of interactions between generators or distribution substations, and so on. By studying large networks of interactions, researchers can develop (generic) methodologies applicable to social networks.

The advent of the Internet has contributed to the rise of social networks and various data on cybernetic social networks such as email contact networks (Ebel, Mielsch, & Bornholdt, 2002; Newman, 2002), Web networks (Albert, Jeong, & Barabási, 1999), peer-to-peer networks (Latapy & Magnien, 2006), instant messaging networks (Smith, 2002), dating networks (Holme, Edling, & Liljeros, 2004) and social networks (Csanyi & Szendroi, 2004). In particular, Facebook represents a prominent example, and many researchers have studied it (Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008; Traud, Mucha, & Porter, 2012). Likewise, they have also recently begun to examine Twitter due to its own growth across the world and the ease with which one can extract data from it (due to its relatively accessible API). For instance, Kumar, Morstatter, and Liu (2014) detail how to efficiently crawl Twitter data. Studies around the Twitter network cover data analysis and network analysis including structural analysis (Myers, Sharma, Gupta, & Lin, 2014), community detection (Ozer, Kim, & Davulcu, 2016), and evolution of centralities (Pereira, Amo, & Gama, 2016).

Using information technologies as the media for social networking turns networks into social media. As Sung and Hwang (2014) point out, social media have both public and private aspects. The private aspect refers to the fact that users invest heavily in posting various elements of their private lives, and the public aspect reflects the impact the private accounts can have on a public audience. Each user in a social network can observe a competition for popularity (measured by the number of followers or retweets) and the desire to become an influencer or have their 15 minutes of fame. Moreover, these networks continue to grow in importance for spreading information. Traditional and online social media are now interlinked: on the one hand, Twitter accounts deliver news from traditional media to Twitter users and, on the other, traditional media address topics that first arose on Twitter. Thus, traditional media no longer have complete control over spreading information. Now, users can play a role in it as well. They can create, forward, and receive information—whether facts, rumors, or fake news. In this context, users with a large influence have a high impact on the nature of information spread on these media. Thus, measuring influence is essential to understand why a rumor may become viral.

## 2.3 Measuring Influence

One way to measure influence involves using the notion of centralities in a network as we discuss in Section 2.1. The measures associated with this notion allow one to highlight popular users, relay users, or central users. Although essential to understand a network's topology, these centrality measures do not sufficiently measure influence. Indeed, they only provide information about a user's position in a network but do not measure if this position has an influential role in it. A user's number of followers or retweets measures popularity but may not be enough to measure influence. Moreover, in Section 3.1, we present the possibility of bias (in interpreting the centralities) based on correlated measures. Thus, one cannot use only these measures to assess influence and must adapt and enrich them to better catch the notion of influence in a network.

The concept of influence comes from the social sciences (particularly psychology). As Almgren and Lee (2016) point out, we can define influence as any change to perceptions, attitudes, or behavior in an individual from interacting with another individual or group in a specific context. At the individual level, it affects the way people decide (Ariely, 2012). At the organizational level, it shapes sensemaking for employees as stakeholders (Weber & Glynn, 2006). Obviously, one has to consider ethical issues in order to separate influence and manipulation (Sunstein, 2016). Managers find interest in the extent to which one can measure influence. Because the notion of influence builds on information, one may consider social media as an innovation when studying this notion. Moreover, measuring influence in a social media may serve as a key component in managerial decision making. For all these reasons, the IS discipline should study influence and propose solutions to evaluate or assess influence, which we do in this paper.

Twitter represents a particularly interesting example of social media. Originally a microblogging medium, Twitter allows users to post and share messages that have no more than 140 characters (though this number has changed for some languages; for instance, as of 2018, Twitter now allows up to 280 characters for English). Due to this brevity, Twitter has become the media of immediacy. Individuals on Twitter have first reported many events, such as Sohaib Athar (an IT consultant) who first reported the military operation that lead to the death of Osama bin Laden mere minutes into the operation. As of 2018, Twitter has about 320 million active users (high but well below the one billion mark), and Katy Perry's account has the most followers (109 million followers as at May, 2018). Similarly, the most retweeted tweet in 2015 involved a boy band and had more than 700,000 retweets. Besides these high statistics, one should know that Twitter provides an API that automatically recovers, through scripts, all publicly published data, though it does have some limitations.

# 3 Theoretical Framework

## 3.1 Correlated Measures

The different notions of centralities (see Section 2.1) can correlate with one another and, in particular, can all correlate with degree. On the one hand, an individual with a high degree is more likely than an individual with low degree to be in a position of relay (i.e., with high betweenness centrality). On the other hand, an individual with a high degree implies that the individual is at path-length distance 1 from a large number of individuals, which decreases the individual's average length of shortest paths to other individuals and, therefore, increases closeness centrality.

In this paragraph, we focus on degree and betweenness centralities. Keep in mind that one can consider an individual with a high degree to be "popular" and an individual with high betweenness to be an "intermediary" or "relay individual" between parts of the network.

Most importantly, researchers have often observed a correlation between the different centrality measures (Valente, Coronges, Lakon, & Costenbader, 2008), which implies that an individual with a high degree has high betweenness. In particular, one can see the correlation in scale-free networks whose degree distribution follows a power law. Indeed, in this case, the probability that a node has degree k is proportional to k-γ, and, on average, the measure of betweenness of a node of degree k is proportional to k-η, where η = (γ-1)/(δ-1) with δ ≈ 2.2 (Barthélemy, 2004).

In this paper, we offer an approach to detect individuals with high betweenness that a high degree does not necessarily induce. We propose a new measure called "corrected betweenness" to remove the influence that degree has on betweenness centrality. Specifically, we expose the mathematical foundations to different point of views, which measure makes it possible to correct a linear dependence between centralities measured by significant correlation coefficients (and close to ±1). As Valente et al. (2008) explains and as our data shows (see Section 4.3), correlated centrality measures are often observed in networks.

## 3.2 Decorrelation Method

Consider the two variables *X* and *Y* whose observations on n individuals are known. The two variables may not be independent (e.g., if they are correlated).

It then becomes interesting to decorrelate them; specifically, to transform *Y* into a new measure $\tilde{Y}$ uncorrelated with *X*. Methods such as principal component analysis (Hotelling, 1933; Saporta, 1990) allow decorrelation but only by transforming all variables; in our case, we want to keep *X* and transform *Y*. The tool we develop to do so resembles whitening or prewhitening techniques, which many researchers in the field of signal processing use to provide uncorrelated data as Kulkarni and von Storch (1995) do for a AR(1) process.

In this section, we discuss decorrelation from several viewpoints.

### 3.2.1 Algebraic Approach

Let *X* and *Y* be two vectors of $\mathbb{R}^n$. The $\mathbb{R}^n$ space can be endowed with the following scalar product:

$$\langle X, Y\rangle = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \ \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad (1)$$

We then define a new "linearly corrected" measure $\tilde{Y}$ from the influence of *X* by the projection of *Y* on the orthogonal of *X*:

$$\tilde{Y} = Y - \ \langle Y, \frac{X}{\sqrt{\langle X, X\rangle}}\rangle \frac{X}{\sqrt{\langle X, X\rangle}} \qquad (2)$$

### 3.2.2 Probabilistic Interpretation

Let $\Omega$ be the set of individuals (such as the nodes of the network), and let $X{:}\Omega \to \mathbb{R}$ and $Y{:}\Omega \to \mathbb{R}$ be two random variables. Because the $\Omega$ set has a finite cardinal number *n*, the spaces of random variables from $\Omega$ to $\mathbb{R}$ and $\mathbb{R}^n$ are isomorphic. We can, therefore, unambiguously identify the random variable *X* and the vector of $\mathbb{R}^n$ *(X(1), X(2), …, X(n))* that we will also call *X*. The scalar product that Equation 1 shows then corresponds to the covariance between the *Y* and *X* variables, and, through the bilinearity of covariance, we get $cov(\tilde{Y}, X) = 0$, which means that the $\tilde{Y}$ and *X* variables are uncorrelated.

### 3.2.3 Statistical Interpretation

When the *X* and *Y* variables, from which we obtained the measure of *n* individuals, are correlated (correlation coefficient is significantly non-zero), we can perform a linear regression and obtain $Y = aX + b + \varepsilon$, where $a$ is estimated by least squares, which means that $a = r\frac{s_Y}{s_X}$, where r represents the correlation coefficient $r = \frac{cov(Y,X)}{s_Y s_X}$ and $s_Y$ and $s_X$ represent the standard deviations of *Y* and *X*, respectively. To remove the factor *X* from *Y*, we calculate $Y - aX$:

$$\begin{aligned} Y - aX &= Y - r\frac{s_Y}{s_X}X \\ &= Y - \frac{cov(Y,X)}{{s_X}^2}X \\ &= Y - cov\left(Y, \frac{X}{s_X}\right)\frac{X}{s_X} \end{aligned} \tag{3}$$

However, the covariance corresponds to the scalar product that Equation 1 shows and, therefore, $s_X = \sqrt{\langle X, X \rangle}$, so $Y - aX$ is equal to $\tilde{Y}$ from Equation 2. Thus, $\tilde{Y}$ and X are uncorrelated.

### 3.2.4 Decorrelation Matrix

Let *A* be the matrix with two rows and *n* columns, whose first and second rows correspond to *n* observations of the random variable *X* and *Y,* respectively. Decorrelating *A* comes down to determining a decorrelation matrix *W* such that the rows of *WA* correspond to two uncorrelated vectors. Various decorrelation methods exist (Kessy, Lewin, & Strimmer, 2018), but we only want to transform the second row of *A*.

In view of the previous arguments, the $W = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}$ matrix with $a = -\frac{s_Y}{s_X} \times cor(X, Y)$ is such that the rows of *WA* are uncorrelated and the first row of *A* is preserved; as such, it represents our decorrelation matrix.

We conclude this point with a comment. Although the $\tilde{Y}$ and *X* variables are uncorrelated, they are not necessarily independent, which can be shown with a chi-square test of independence. One can also observe this non-independence when *X* is the vector of the degrees and *Y* is the vector of betweenness. A node of degree 1 necessarily has 0 betweenness and, therefore, has a value of $\tilde{Y}$ equal to $-\langle Y, \frac{X}{\sqrt{\langle X,X \rangle}} \rangle \frac{1}{\sqrt{\langle X,X \rangle}}$ or $-\langle Y, \frac{X}{\langle X,X \rangle} \rangle$.

## 3.3 The "Corrected Betweenness" for Identifying "Quiet Relays"

Let *D* and *B*, which correspond to the degree and betweenness centralities of *n* individuals in the network, be the vectors of $\mathbb{R}^n$. Thus, we define the *corrected betweenness measure* as:

$$\widetilde{B} = B - \frac{s_B}{s_D} \times cor(D, B) \times D, \tag{4}$$

where $s_B$ and $s_D$ are the standard deviations of *B* and *D*, and $cor(D, B)$ is the correlation coefficient between *B* and *D*.

Thus, this corrected betweenness measure shows the individuals who link several communities (this betweenness measure does not require one to detect communities beforehand) without necessarily having a high degree. Compared to the usual betweenness measure, it makes it possible to decrease the rankings of individuals with high degrees and, therefore, to bring forward individuals with lower degrees. It builds up individuals with high betweenness and low degree.

We call individuals with high corrected betweenness measures *quiet relays*. One can see quiet relays in a social network as not necessarily popular relays. Their betweenness centrality is higher than that which their degree could provide.

Figure 1 shows an example of quiet relay. In the network that this figure depicts, we can see three groups of five vertices (which are complete graphs) and three vertices A, B, and C that link these groups. Vertices A and B have a high degree (10), whereas C has a low degree (2). Vertices A and B have the highest betweenness centrality, whereas vertex C is ranked fifth (after vertices A and B and the two vertices of degree 7). However vertex C is ranked first on the corrected betweenness measure.

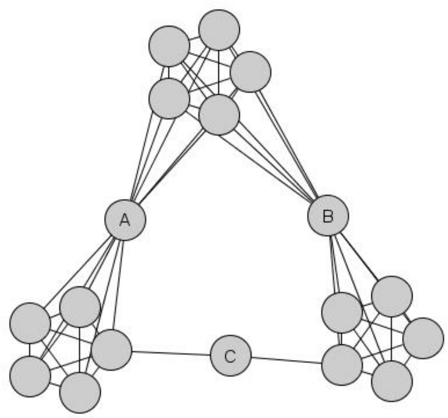


**Figure 1. Example of High Degree Vertices with a High Betweenness (A and B) and a Quiet Relay (C)**

Note that the corrected betweenness may be non-positive, though it is not an issue because we are interested in the ranking (one can add a constant to make it positive without changing the ranking).

### 3.4 A "Behavioral Triggering Measure" in Social Networks

The quiet relays that we discuss in Section 3 concern the dissemination of information but not necessarily influence in the sense of "behavioral influence". In this section, we propose an influence measure for a Twitter network or, more precisely, to a network of retweets. However, one could easily adapt this measure to other kinds of networks (we give some examples at the end of this section)

We consider a tweet to have influence if users read and find it relevant (retweets) and if the tweet causes other users to write their own tweets. So, for each individual *i*, we establish an initial influence measure, *Z*, by counting the number of users who write a tweet after having retweeted *i*. This measure is, of course, correlated with the degree. An individual with a widely retweeted tweet is more likely to see a tweet from those who retweet the original tweet than an individual with a poorly retweeted tweet

In terms of graph-theoretic vocabulary, let *G* be the graph of Twitter users linked by retweets *v* be a vertex, $N_v$ be its neighborhood (i.e., the subgraph of *G* induced by the vertices linked to *v*), and $d_v$ be its degree (thus, there are $d_v$ vertices in $N_v$). Given a tweet written by *v*, we consider the subgraph of $N_v$ induced by Twitter users having written a tweet, and we are interested in the number of vertices in this graph that we denote with $z_v$.

Obviously, $z_v$ is linked to $d_v$: it is lower than or equal to $d_v$, and the higher $d_v$ is, the higher $z_v$ is likely to be. As such, we need to apply the decorrelation that we discuss in Section 3.2 to define a *corrected* $z_v$ measure that we denote with $\tilde{z}_v$ and that we call the "behavioral triggering measure". Figure 2 illustrates this relation between $d_v$ and $z_v$ and, in particular, the fact that high degree vertex (such as vertex A) is more likely to have reactions to their tweets.

One can adapt this approach to analyze other social networks. For instance, when analyzing a Facebook network, one can replace "retweet" with "like" and "writing a tweet" with "writing a comment".

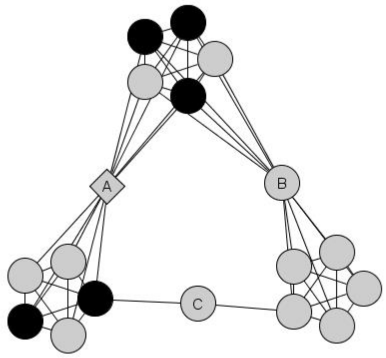


**Figure 2. Theoretical Example of Relation between $d_v$ and $z_v$[1]**

### 3.5 A New and Unified Measure of Influence

In their paper, He, Hu, Shi, and Liu (2014) propose measuring (in terms of influence) the impact that a message of recrimination can have on members of a social network. To do so, they use a model that considers three dimensions: 1) message quality, 2) transmission speed, and 3) the degree of user interaction. The user interaction dimension refers to the user's position in the social network. From these three dimensions, the authors propose a formula (p. 6042) that seems applicable to Twitter after adjusting it. The main adjustment involves how to determine message quality. Twitter has short messages that often include objects such as videos, images, and links. These objects are qualitative and, therefore, measured differently. Also, we suggest that one should not measure a tweet's instrinsic quality. Instead, we suggest measuring its success through its number of retweets (and/or commented retweets and "likes"). The other two dimensions being valid, we adjust He et al.'s model in several ways.

Our first dimension measures influence through tweet success using degree centrality. We denote $d_i$ as the degree of the i[th] individual and *D* as the vector of the $d_i$.

Our second dimension corresponds to He et al.'s (2014) third dimension: the degree of user interaction. In this case, one can interpret it as the ability to reach diverse communities (i.e., the ability to be an intermediary). We calculate this dimension using corrected betweenness that we discuss in Section 3.3. We denote $\tilde{b}_i$ as the value of this measure for the i[th] individual and $\tilde{B}$ as the vector of the $\tilde{b}_i$.

Our third dimension measures influence through tweet transmission and reflects He et al.'s (2014) second dimension. We assess it with the behavioral triggering concept that we discuss in Section 3.4. We denote $\tilde{z}_i$ as the value of this measure for the i[th] individual and $\tilde{Z}$ as the vector of the $\tilde{z}_i$.

Using the theoretical framework that we describe in Section 3, we define the second and third dimensions orthogonally to the first.

A linear combination of the three measures above then determines the measure of overall influence of the i[th] individual:

$$\alpha d_i + \beta \tilde{b}_i + \gamma \tilde{z}_i, \qquad (5)$$

Finally, we need to set the values of the coefficients $\alpha, \beta, \gamma$. We propose setting them so that the vectors *D*, $\tilde{B}$, and $\tilde{Z}$ are standardized (based on the scalar product that Equation 1 defines), which, from a statistical viewpoint, results in reduction (divide by the standard deviation). This reduction erases a difference in order of magnitude between the measures *D*, $\tilde{B}$, and $\tilde{Z}$. Therefore, we set $= \frac{1}{\sqrt{<D,D>}}$, $\beta = \frac{1}{\sqrt{<\tilde{B},\tilde{B}>}}$, $\gamma = \frac{1}{\sqrt{<\tilde{Z},\tilde{Z}>}}$. These coefficients have the advantage that they remain unchanged if one adds a constant to *D*, $\tilde{B}$, or $\tilde{Z}$ (a translation does not change the variance), and the ranking that this new measure obtains remains unchanged if one adds a constant to *D*, $\tilde{B}$, or $\tilde{Z}$.

[1] Assumes that 50 percent of people who retweet tweets have written a tweet. The vertices show twitter users, and the links show retweets; black vertices show Twitter users who wrote a tweet after retweeting a tweet that the diamond vertex wrote.

If one deals with a social network for which one cannot or does not need to compute a behavioral triggering measure, then we could set $\gamma = 0$ (i.e., consider only the two first dimensions: the degree and the corrected betweenness).

# 4 The Uber Case

## 4.1 Case Description and Data Collection

Founded in 2009, Uber is an American company that helps people to find a means of urban transportation in a more efficient way than existing methods (taxis and public transportation). Uber developed a smartphone application to link seekers (customers) and providers (officially registered drivers or just employees/Uber subscribers) using geolocation services that smartphones can use. In addition, Uber established at least two major partnerships: with Google and Microsoft and its Bing mapping service. Extremely practical, the Uber service expanded quickly to offer its service in more than 300 cities in six years and achieved a valuation of more than US$50 billion dollars. Of course, this sudden competition led existing stakeholders to react in an often aggressive manner. For instance, on 26 June, 2015, twelve French trade unions launched an indefinite nationwide taxi strike. This action against Uber and similar services frequently led to skirmishes with and violence against some Uber drivers. The media then provided the stage on which battles between the various groups (e.g., pro-taxi/anti-Uber stakeholders, pro-Uber/anti-taxi stakeholders, and various citizens and journalists) fought.

We used Twitter to collect our data due to its immediacy. Also, we accessed its API through NVivo 11 software. We extract tweets that appeared between 23 June and 3 July, 2015, with the keywords "#UberPOP" and "#TAXI". In total, we collected 11,308 tweets. We then exported this data to an Excel spreadsheet, processed it with R, and viewed it with Gephi.

From this data set, we built the following network: network nodes represent Twitter users who tweeted or retweeted a tweet on the Uber/taxi conflict. We considered two Twitter users to have a link if one retweeted the other.

## 4.2 Structural Analysis of the Network

We conducted a structural analysis to better understand the network more broadly. It had 6,605 nodes and 8,609 links, which amounted to a 0.04 percent link density; note that the literature on network analysis commonly features low link densities.

This network had a low average length of shortest paths (4.2) and a characteristic length in the same range (4.1). These measures can be considered as low because they are lower than two reference measures: 1) the well-known "six degrees of separation" and 2) the theoretical value measured with a random graph, which is $\frac{\ln(6605)}{\ln\left(\frac{2\times 8609}{6605}\right)} = 9.17$.

However, this retweet network was not a small-world network because local connectivity, measured by clustering coefficients, was not high. Keep in mind that, in a small-world network, the conditional probability that two nodes are linked, given that they are linked to a third, is much higher than the probability that any two nodes are linked in the network (the latter being equal to the density). Here, because the network had only 106 triangles, the conditional probability that two nodes were linked, given that they are linked to a third (the second clustering coefficient), was 0.015 percent—the same order of magnitude (or even smaller) than the density.

This network was also scale free. As Figure 3 shows, the degree distribution followed a power law, which implies that the betweenness measure was likely correlated with the degree measure (see Section 4.3).

This power law of degree distribution also implies that the network had many low degrees and few high degrees, which means that the network contained hubs. However, the hubs did not form a rich club because the six nodes with the highest degrees had no connections among them, the 10 highest degree vertices induced only one link in the network. The network also had a negative (−0.2) assortativity measure (homophily of degrees) (Newman, 2002), which means that the nodes with high degrees not only did not connect to each other but also tended to link to nodes with low degrees.

Research on communities has used the standard Clauset-Newman-Moore (Clauset, Newman, & Moore, 2004) method. We found that communities form around hubs that have many dangling nodes (nodes with only one link). This result concurs with our previous findings that show the presence of few nodes with high degrees and many nodes with low degrees (scale free), the lack of clustering (triangles allow the emergence of groups dense in links), the absence of a rich club, and a negative assortativity measure. Figure 4 summarizes these results: it shows various communities (one color per community) with nodes with high degrees around which many nodes with low degrees gravitate.

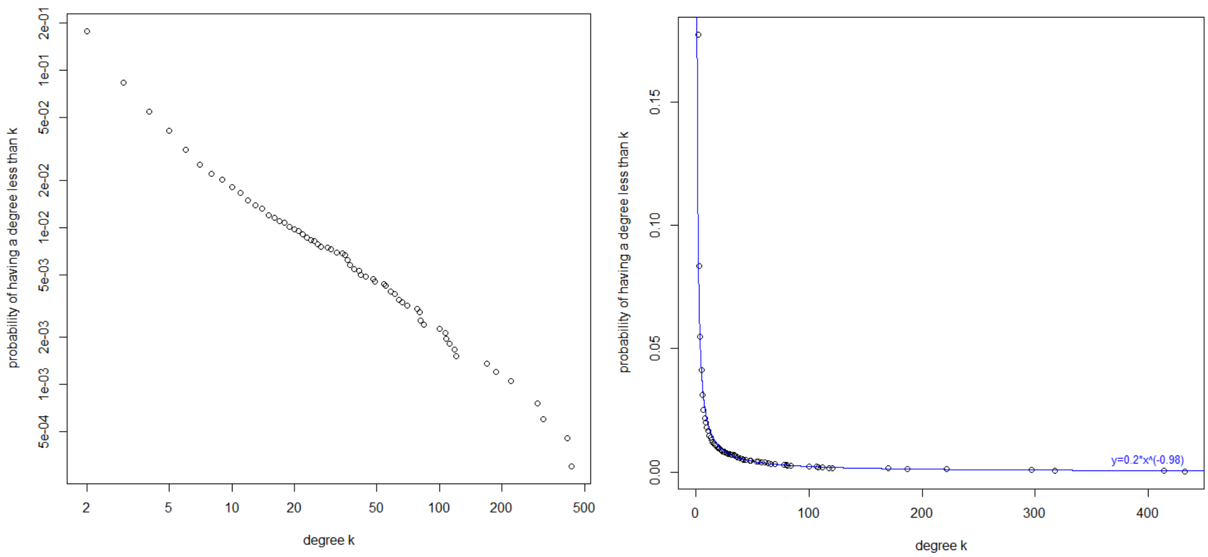


**Figure 3. Degree Distribution on a Logarithmic Scale (Left) and Power Law Adjustment (Right)[2]**

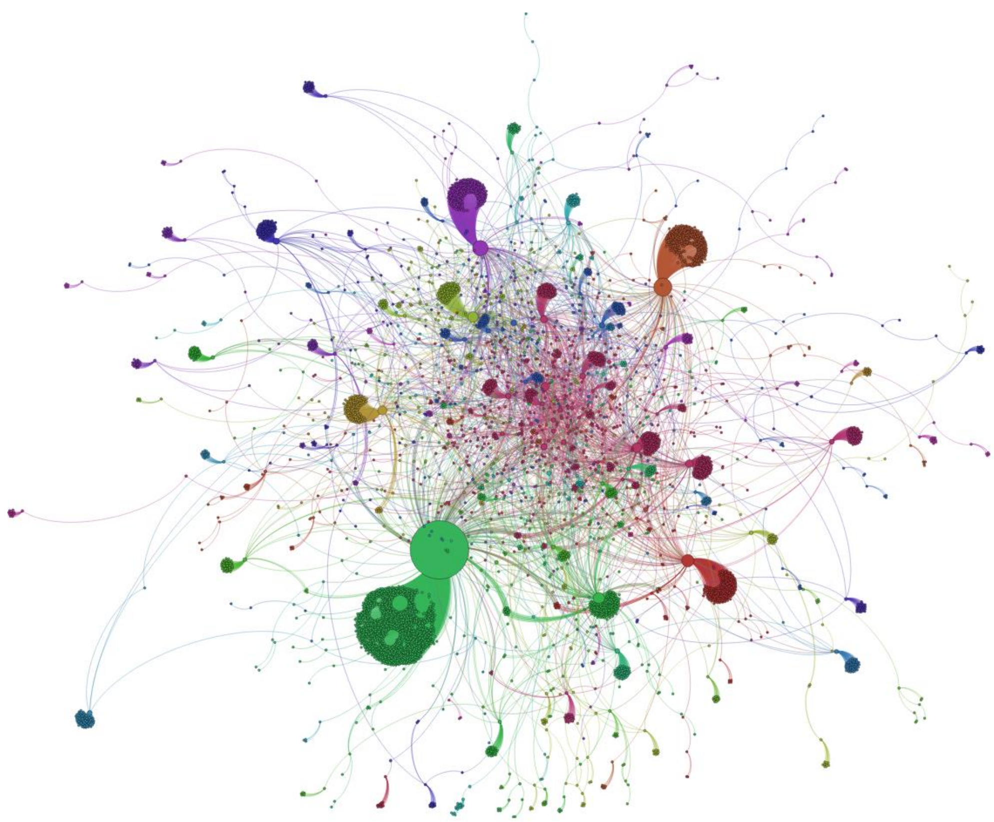

**Figure 4. Representation of Network of Retweets for Uber Data[3]**

[2] The node with the highest degree does not appear in this analysis given its atypically high value (1837 compared to 432 for the node with the second highest degree).

### 4.3 Popular Users and Quiet Relays

After grasping the overall structure of the network in Section 4.2, in this section, we focus on the key individuals in the network. We start by showing individuals with high degree and individuals with high betweenness (as we define them in Section 2.1). Subsequently, after noting a link between these two measures, we compute the corrected betweenness that we discuss in Section 3.3.

The three Twitter users with the highest degree were:

- @lhenault (or L0uis Hénaµlt), a data scientist who posts humorous tweets. His degree in the network was 1837. The largest vertex in Figure 4 represents him (in green at the bottom left).
- @Torped00 (or Fred Zoullissimmo), who posts humorous tweets and had 84.1K tweets. His degree in the network was 432. The large vertex in brown at the top-right in Figure 4 represents him.
- @jul_mm (Julien Migaud-Muller), a journalist whose tweets provide information on developing stories. His degree in the network was 414. The large vertex in purple at the top in Figure 4 represents him.

These three individuals with high degrees also had the highest betweenness: @lhenault, with an atypically high betweenness measure, remained in first place, though @jul_mm and @Torped00 switched places for the second and third spots. Specifically, @lhenault had more than four times the betweenness measure that @Torped00 had, and it followed the same pattern as for degree (i.e., @lhenault had a betweenness measure 3.75 times higher than the second highest betweenness measure).

Thus, we can see that the degree and betweenness measures in this network had a strong correlation. The Pearson correlation coefficient was 0.987 (0.995 if we removed @lhenault), and Spearman's rank correlation coefficient was 0.986 (0.990 if we removed @lhenault).

We confirm this strong link between degree and betweenness by simulating a random graph whose nodes represent the individuals in our network who retain their degrees. The algorithm rewires the links. As such, we can compare our betweenness measures with measures determined with a random graph in which the nodes have the same degree as in the one we examined. Of 1,000 simulations, in 100 percent of cases, @lhenault had the highest betweenness; in 78.5 percent of cases, @Torped00 and @jul_mm had the second and third highest betweenness, respectively; in 21.5 percent of the remaining cases, @jul_mm took second place and @Torped00 took third place. Thus, in 100 percent of cases, as in our real network, the three individuals with the highest degrees also had the highest betweenness.

The three quiet relay users with the highest corrected between measure (see Section 3.3) were:

- @Daturaparano (Datura Gouyou), who had 36 retweets in our network, including 35 between 9:39 and 9:54 on 26 June.
- @PrunelleTika (Dreba Tika from CIV), who had 32 retweets in our network, including 25 between 10:44 and 10:55 on 25 June.
- @FidelinShana (Shanon), who had 21 retweets in our network, all between 13:30 and 13:37 on 26 June.

These users played a relay role by disseminating information. This dissemination occurred through a high number of retweets in a short time. Further, we found that the three users (@Daturaparano, @PrunelleTika, and @FidelinShana) had links to 14, 13, and 7 communities (see Section 4.2), respectively, in a quieter manner than the three users with high degrees because, although @lhenault, @Torped00, and @jul_mm had links to 32, 21, and 21 communities, respectively, @Daturaparano, @PrunelleTika, and @FidelinShana each had twelve times fewer links than @Torped00.

[3] The size of the node corresponds to its degree. The colors correspond to the communities extracted by the Clauset-Newman-Moore algorithm.

### 4.4 Behavioral Triggers

The degree distribution followed a power law, and many individuals had low degrees (which includes a degree of 1). Thus, although retweeting demonstrates reading and an interest in the relayed tweet, it does not necessarily meaningfully indicate influence. Thus, we computed the behavioral triggering measure (see Section 3.4) to find users (and, therefore, tweets) that influenced other users in the sense that they activated other users to write a tweet.

Recall that our behavioral triggering measure is decorrelated to the degree; otherwise, in our case study, the Pearson correlation coefficient between degree and the number of reactions would have been 0.869, which would have highlighted the need to decorrelate the latter from the former.

The three individuals with the highest behavioral triggering measure were:

- @jul_mm (Julien Migaud-Muller), a journalist whose tweets provide information on a developing story.
- @PerrineST (Perrine Stenger), a pop culture journalist.
- @omercuriot (Olivier Mercuriot), an international development director of a company.

Interestingly, these three users differed from the quiet relays. Further, with the exception of @jul_mm, they differed from the popular individuals as well.

In the network we examined, this measure shows users who tended to create information: two journalists and a director of a company.

### 4.5 Influent Individuals

In Sections 4.3 and 4.4, we highlight the top three users in the popularity, quiet relay, and behavioral trigger dimensions. With the exception of @jul_mm, the individuals in each dimension differed.

Following Section 3.5, we combine these three dimensions to obtain a unique indicator of influence. As a result, the three highest influential individuals were:

- @jul_mm (Julien Migaud-Muller), who had the highest behavioral triggering measure.
- @lhenault (or L0uis Hénaµlt), who had the highest degree.
- @Daturaparano (Datura Gouyou), who had the highest corrected betweenness.

Unsurprisingly given the results concerning the popularity and the behavioral trigger measures, @jul_mm took the first position. @lhenault and @Daturaparano—a popular user and quiet relay, respectively—took the second and third spots. As a result, this ranking includes the three dimensions that make up the influence measure.

## 5 Discussion and Conclusion: Influence Management Strategy for Social Networks

### 5.1 Implications of Detecting Quiet Relays

This paper has two main managerial implications for someone who wants to have an influence in a social network. The first implication concerns the cost of influencing and the second one concerns quality.

In order to have an influence in a social network, one should target high-value individuals (i.e., people who have the largest number of followers and who others frequently retweet). However, one must pay a price to get in touch with these people. First of all, since high-value targets have many interactions, one needs to expend much effort catch their attention. Second, because they know that they have a central position in the network and that their actions have an impact, they may be suspicious of any popup contact who arises out of the blue. Finally, they may ask for a reward for relaying one's message. For these reasons, one may find interest in discovering relays, who typically do not know about their relevant and valued position. One will incur a lower cost in connecting with a relay but receive a similar efficiency in growing one's influence.

From a qualitative point of view, we observed that these quiet relays conduct reasoned actions: in other words, that they more carefully select the messages they retweet. Having both a high betweenness and a low degree shows that the individual has the ability to connect two communities with different specificities. This qualitative ability makes these relays highly valuable for an influencer who wants to reach multiple communities with a single contact. They could even serve as valuable assets for a company.

Detecting quiet relays has a theoretical interest, too. It could introduce a change of perspective opposed to Barabási and Albert preferential attachment mechanism (Barabási & Albert, 1999). In a preferential attachment model, a new node entering the network will create links with existing nodes according to a probability proportional to their degree. In other words, it will link preferably with the "popular" nodes (i.e., nodes with a high centrality measure). It entails that popular nodes tend to become more and more popular. As a result, a positioning strategy based on quiet relays (with, for instance, the creation of the same links of them) takes a counter approach to the preferential attachment.

This change of perspective (focus on quiet relays instead of popular users) can induce a paradigm shift in online social networks. Indeed, highlighting quiet relays is against the current global trend of social networks users where the influence of an individual is solely based on measures related to the degree centrality (number of retweets, number of followers on Twitter or number of friends on Facebook) and, therefore, on the quantity and not necessarily the quality. Introducing a new indicator related to the relay position of the user could better highlight users with strategic links and slow down the competition for popularity.

## 5.2 Increasing One's Influence

Users can increase their influence in the three dimensions that we discuss in Section 3.5. For the first dimension (influence measured by the number of retweets), messages meant to appear humorous seem receive more retweets. For the second dimension (corrected betweenness), a user needs to have links to different user communities but manage the links sparingly (it is ineffective to tweet or retweet randomly and massively, which the first dimension considers). For the third dimension (transmission), the message must create information, so tweets from journalists ranked highly.

To reduce others' betweenness centrality and increase one's own, one needs to connect two subnetworks that a only single individual previously connected. However, individuals can find doing so difficult in practice if they lack an overall vision of the network. In this paper, we demonstrate that retweeting popular messages from different backgrounds makes it possible to become a quiet relay.

One needs to search for influential individuals to study the virality of a phenomenon to understand or use a dissemination process, such as information cascade models (Zhu, Wang, Wu, & Zhu, 2014). In our case, a tweet does not necessarily launch a cascade; rather, an event can, too. A tweet, therefore, represents an early witness to the event. We noted an increase in the number of messages in the early morning (at about 7:00 a.m.), which might correspond to newscasts. Virality makes it possible to measure the event's influence.

Li et al.'s (2014) approach for identifying influential users to predict their impact on the network uses the concepts of information inventor and information spreader that we find in our analyses with those who create information (Section 4.4) and those who pass it on (Section 4.3), respectively. As such, we suggest that being active in these two aspects increases visibility and influence.

## 5.3 Conclusion

We used tools from graph theory to measure influence in a network. By showing that the main concepts of centrality can correlate, we propose new measures that address that correlation. These new measures reveal three types of influential users, which previous work has also shown. We demonstrate the existence of widely retweeted individuals (particularly those who use humor), "quiet relays" (who pass on information), and those who create information. We consider these three dimensions together to create a unique influence measure and, thus, reveal these three profiles. Figure 5 summarizes our contribution.

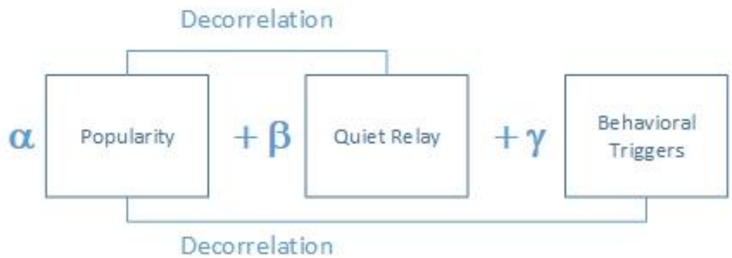


**Figure 5. Assessing Influence in a Social Network**

To illustrate this figure, in this case we have (see Section 3.5) $\alpha = 3.942 \times 10^{-2}, \beta = 3.903 \times 10^{-5}$ and $= 5.255$ . We take the example of the user @jul_mm (see Figure 6). He ranked third for popularity with a degree centrality of 414, fourth among the quiet relays with a corrected betweenness of 409128.2, and first for the behavioral triggers with a measure of 6.538. As a result, he had an influence measure of 66.65 and ranked first.

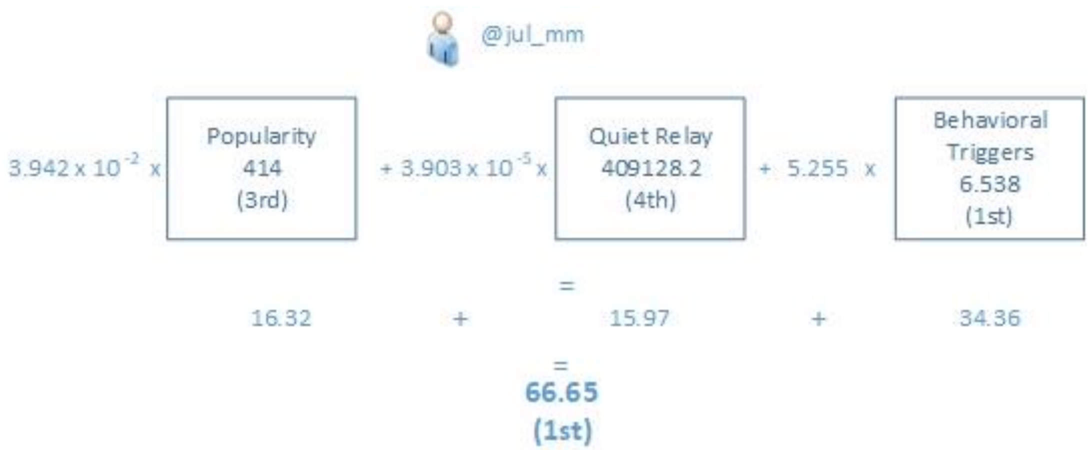


**Figure 6. Example of Computation and Decomposition of Influence for the User @jul_mm**

This new measure has double potential interest. On one hand, a company that wants to track its activities on social network can use the measure to determine relevant users in the network and define a stable measure of performance for its actions. Many software programs provide ways to determine the key influencers, but they do not determine how they find these users and, moreover, they show only whether a user is or is not an influencer. Here, we provide three types of influential users and a unique way to assess an event's influence. On the other hand, for a company that wants to, for example, counter a negative rumor we provide a map of the relevant users that they may target to do so. And, once again, with our unique measure, this company can see when the influence goes back to an acceptable threshold.

As we mention in Section 2.3, when describing the concept of influence, one needs to consider ethical issues. There is a fine line between well-meaning influence and evil manipulation. One needs to link influence management to ethics when studying the former (Mingers & White, 2010).

In addition, in the near future, we hope to work on the strategies that allow users to amplify positive influence and cope with negative propaganda. We can propose the following key elements of a new method. First, one could identify three groups of users: influential, potentially influential, and non-influential individuals. Second, one could analyze the characteristics of these individuals and their environment, such as network-based characteristics for the former (e.g., centrality and clustering coefficient) and additional pieces of information on the nodes such as location, time of tweet, number of hashtags in the tweet, and content for the latter. Finally, we will need works that focuses on identifying a significant link between one or more of these characteristics and the influence potential. This link could support efficient strategies to increase influence.

## References

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? The structure of the Twitter follow graph. In *Proceedings of the International Conference on World Wide Web* (pp. 493-498).

Abilhoa, W. D., & de Castro, L. N. (2014). A keyword extraction method from Twitter messages represented as graphs. *Applied Mathematics and Computation*, *240*, 308-325.

Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, 466(7307), 761-764.

Albert, R., Jeong, H., & Barabási, A. L. (1999). The diameter of the World Wide Web. *Nature*, *401*, 130-131.

Almgren, K., & Lee, J. (2016). An empirical comparison of influence measurements for social network analysis. *Social Network Analysis and Mining*, *6*(1-52), 1-18.

Amini, A. A., Chen, A., Bickel, P. J., & Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, *41*(4), 2097-2122.

Ariely, D. (2012). *Predictably irrational: The hidden forces that shape our decisions* (2nd Ed.). London, UK: Harper Collins.

Barabási, A.-L. (2009). Scale-free networks: A decade and beyond. *Science*, *325*(5939), 412-413.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509 LP-512.

Barabási, A.-L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 60-69.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, *311*(3-4), 590-614.

Barthélemy, M. (2004). Betweenness centrality in large complex networks. T*he European Physical Journal B: Condensed Matter and Complex Sys*tems, *38*(2), 163-168.

Berge, C. (1958). *Théorie des graphes et ses applications*. Paris: Dunod.

Burt, R. S. (1992). *Structural holes: the social structure of competition*. Cambridge, MA: Harvard University Press.

Cao, J., Basoglu, K. A., Sheng, H., & Lowry, P. B. (2015). A systematic review of social networks research in information systems: Building a foundation for exciting future research. *Communications of the Association for Information Systems*, 36, 727-758.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*, 66111.

Colizza, V., Flammini, A., Serrano, M. A., & Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature Physics*, *2*, 110-115.

Cross, R., Borgatti, S., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*(2), 25-46.

Csanyi, G., & Szendroi, B. (2004). Structure of a large social network. *Physical Review E*, *69*, 036131.1-036131.5.

Degenne, A., & Forsé, M. (1994). *Les réseaux sociaux.* Paris: Armand Colin.

Ebel, H., Mielsch, L.-I., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, *66*, 35103.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75-174.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215-239.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver: Empirical Press.

Freiesleben de Blasio, B., Svensson, A., & Liljeros, F. (2007). Preferential attachment in sexual networks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 10762-10767.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821-7826.

Girvan, M., & Newman, M. E. J. (2004). Finding and evaluating community structure in network. *Physican Review*, *69*(2), 26113.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, *78*(6), 1360-1380.

He, J., Hu, M., Shi, M., & Liu, Y. (2014). Research on the measure method of complaint theme influence on online social network. *Expert Systems with Applications*, *41*(13), 6039-6046.

Holme, P., Edling, C. R., & Liljeros, F. (2004). Structure and time-evolution of an Internet dating community. *Social Networks*, *26*, 155-174.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417-441.

Kessy, A., Lewin, A., & Strimmer, K. (2018). Optimal whitening and decorrelation. *The American Statistician*, 1-6.

Kulkarni, A., & von Storch, H. (1995). Monte Carlo experiments on the effect of serial correlation on the Mann-Kendall test of trend. *Meteorologische Zeitschrift*, *4*(2), 82-85.

Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics*. New York, NY: Springer.

Langin, K. (2018). Fake news spreads faster than true news on Twitter—thanks to people, not bots. *Science*. Retrieved from http://www.sciencemag.org/news/2018/03/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots

Latapy, M., & Magnien, C. (2006). Measuring fundamental properties of real-world complex networks. *arXiv.* Retrieved from https://arxiv.org/abs/cs/0609115

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*(4), 330-342.

Li, J., Peng, W., Li, T., Sun, T., Li, Q., & Xu, J. (2014). Social network user influence sense-making and dynamics prediction. *Expert Systems with Applications*, *41*(11), 5115-5124.

Luxburg, U. (2007). A Tutorial on Spectral Clustering. *Statistics and Computing*, *17*(4), 395-416.

Mingers, J., & White, L. (2010). A review of the recent contribution of systems thinking to operational research and management science. *European Journal of Operational Research*, *207*(3), 1147-1161.

Nettleton, D. F. (2013). Data mining of social networks represented as graphs. *Computer Science Review*, *7*, 1-34.

Newman, M. E. J. (2002). Assortative Mixing in Networks. Physical Review Letters, *89*(20).

Newman, M. E. J. (2004). Detecting community structure in networks. *European Physical Journal B*, *38*(2), 321-330.

Ozer, M., Kim, N., & Davulcu, H. (2016). Community detection in political Twitter networks using Nonnegative Matrix Factorization methods. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*

Park, J., & Newman, M. E. J. (2005). A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, *10*.

Pereira, F. S. F., Amo, S. d., & Gama, J. (2016). Evolving centralities in temporal graphs: A Twitter network analysis. In *Proceedings of the IEEE International Conference on Mobile Data Management* (vol. 2, pp. 43-48).

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks (long version). *Journal of Graph Algorithms and Applications*, *10*(2), 191-218.

Qin, Z., Cai, J., & Wangchen, H. Z. (2015). How rumors spread and stop over social media: A multi-layered communication model and empirical analysis. *Communications of the Association for Information Systems*, *36*, 369-391.

Saporta, G. (1990). *Probabilités, analyse des données et statistique.* Paris: Editions Technip.

Scott, J. P. (2000). *Social network analysis: A handbook* (2nd ed.). Thousand Oaks, CA: Sage.

Smith, R. D. (2002). Instant messaging as a scale-free network. *arXiv*. Retrieved from https://arxiv.org/abs/cond-mat/0206378

Sung, M., & Hwang, J.-S. (2014). Who drives a crisis? The diffusion of an issue through social networks. *Computers in Human Behavior*, *36*(0), 246-257.

Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. New York, NY: Cambridge University Press.

Traud, A., Mucha, P., & Porter, M. (2012). Social structure of Facebook networks. *Physica A*, *391*, 4165-4180.

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, *32*(4), 425-443.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections*, *28*(1), 16-26.

Watts, D. J. (2003). *Small worlds: The dynamics of networks between order and randomness.* Princeton, NJ: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*, 440-442.

Weber, K., & Glynn, M. a. (2006). Making sense with institutions: Context, thought and action in Karl Weick’s theory. *Organization Studies*, *27*(11), 1639-1660.

Yang, J., McAuley, J., & Leskovec, J. (2013). Community detection in networks with node attributes. In *Proceedings of the International Conference on Data Mining* (pp. 1151-1156).

Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, *33*, 329-364.

Zhu, T., Wang, B., Wu, B., & Zhu, C. (2014). Maximizing the spread of influence ranking in social networks. *Information Sciences*, *278*, 535-544.

## About the Authors

**Romain Boulet** is associate professor in mathematics at the IAE Lyon School of Management and is a member of the Magellan Research Center. His research areas include graph theory (especially algebraic graph theory) and network analysis, with a particular interest in interdisciplinary collaboration. He has focused on the analysis of legal networks and organized workshops on complexity and public policy.

**Jean-Fabrice Lebraty** is a professor in management at the IAE Lyon School of Management. He is a member of the Magellan Research Center and specializes in Information System Management. His main areas of interest are extreme decision-making and crowdsourcing. He is also a member of the French Foundation for Management Education (FNEGE), serving as international program manager. Since 2015, he has been involved in *Communications of the Association for Information Systems* as the "French Papers" editor.

Copyright © 2018 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.