Article:

Dahka, Z.Y., Haji Heydari, N. orcid.org/0000-0003-3663-5254 and Rouhani, S. (2020) User response to e-WOM in social networks : how to predict a content influence in Twitter. International Journal of Internet Marketing and Advertising, 14 (1). pp. 19-47. ISSN 1477-5212

https://doi.org/10.1504/IJIMA.2019.10024249

User response to e-WOM in social networks: how to predict a content influence in Twitter

Zohreh Yousefi Dahka

Faculty of Management,
University of Tehran,
Gisha Bridge, Chamran Highway,
Tehran, Iran
Email: usefie.z@ut.ac.ir

Nastaran Hajiheydari*

Sheffield University Management School,
University of Sheffield, UK
Email: n.hajiheydari@sheffield.ac.uk
*Corresponding author

Saeed Rouhani

Faculty of Management,
University of Tehran,
Gisha Bridge, Chamran Highway,
Tehran, Iran
Email: srouhani@ut.ac.ir

Abstract: The purpose of this research is to find influential factors on electronic word-of-mouth effectiveness for e-retailers in Twitter social media, applying data mining and text mining techniques and through R programming language. The relationship between using hashtag, mention, media and link in the tweet content, length of the content, the time of being posted and the number of followers and followings with the influence of e-WOM is analysed. 48,129 tweets about two of the most famous American e-retailers, Amazon and eBay, are used as samples; results show a strong relationship between the number of followers, followings, the length of the content and the effectiveness of e-WOM and weaker relevance between having media and mention with e-WOM effectiveness on Twitter. Findings of this paper would help e-retailing marketers and managers to know their influential customers in social media channels for viral marketing purpose and advertising campaigns.

Keywords: electronic word-of-mouth; e-WOM; social media; e-retailing; content influence; data mining; Twitter; text mining.

Reference to this paper should be made as follows: Dahka, Z.Y., Hajiheydari, N. and Rouhani, S. (xxxx) 'User response to e-WOM in social networks: how to predict a content influence in Twitter', *Int. J. Internet Marketing and Advertising*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Zohreh Yousefi Dahka has been graduated from the Shahid Beheshti University in Electrical Engineering for her BSc and completed the requirements of the MSc in MBA, Management Information Systems, from the Tehran University. Her thesis title was 'Determining effective factors on the success of electronic word-of-mouth, in Twitter social media and in the scope of e-retailing: an approach to data mining and text mining techniques' and the current paper is the outcome of the mentioned thesis. She has worked as an Expert and Site Analyst at the TOSAN (banking and payment solution provider) and Product Manager at the Praham company. Her research interests include information systems, business intelligence and decision making by means of data mining and text mining.

Nastaran Hajiheydari is an Associate Professor of Business at the University of Tehran, Iran. She received her MSc and PhD in Information Technology and System Management from the University of Tehran and pursues her research in Information Systems and Digital Business. She has a professional background that spans a broad spectrum of skills and ability to serve in both academia and industry. Her last ten-years academic experience includes teaching in Faculty of Management at University of, Tehran; as well as consulting and executive program development activities in Iranian public and private sector organisations. She has published several papers in reputable journals such as *Information Systems Journal*, *Kybernetes* and *Journal of Modelling in Management*. Her overall research interest is focused in digital business modelling, mobile business, social networks and large-scale data analysis application in business. She has also been a keynote speaker at several conferences, and program/general chair of several others.

Saeed Rouhani obtained a BS in Industrial Engineering from the Iran University of Science and Technology, Tehran, Iran, in 2003 and MA in Information Technology Management from the Allameh Tabataba'i University, Tehran, Iran, in 2005. He received a PhD in Systems Engineering from the Iran University of Science and Technology in 2011. He is currently an Assistance Professor in Faculty of Management, University of Tehran, Iran. His research interests include enterprise resource planning systems, business intelligence, information systems, and decision making. He has published ten books and presented more than 30 papers at different conferences and in acclaimed journals.

1 Introduction

The power of electronic word-of-mouth (e-WOM) on firms' market performance has recently attracted great research attention (e.g., Chen and Xie, 2008; Ho-Dac et al., 2013). Numerous studies on the effect of e-WOM on the performance of companies have been conducted due to the proliferation of e-WOM on social media. Prior studies have strongly verified that e-WOM is an important source of information for consumers and helps them with making purchase decisions (Chen and Xie, 2008; Chevalier and Mayzlin, 2006; Dellarocas et al., 2007; Duan et al., 2008a, 2008b; Rui et al., 2011). The mentioned studies have been conducted on different channels such as online review sites (Liu, 2006), Twitter (Asur and Huberman, 2010; Rui et al., 2011), blogs (Qin, 2011), and many other social media channels.

Controlling and leading e-WOM, as a marketing tool and influencing consumer-to-consumer communication to achieve positive purchase decisions, are difficult tasks for entrepreneurs and marketers (Kozinets et al., 2010; Kleina et al., 2015). One of the challenges faced by the current research is the complexity of measuring e-WOM effectiveness caused by the high number of effective factors involved in this process, such as direct customer experiences, attractive messages, celebrity endorsements, and consumer involvement (Dwyer, 2007). On the other hand, managers are generally unfamiliar with the key factors that can encourage online social network members to take part in an e-WOM process. Among all the e-WOM channels, social media have become among the most common e-WOM channels because of their ubiquity, mobility, and interactivity (French and Read, 2013). These features of social media enable users to communicate and connect with one another more frequently and more closely (Laroche et al., 2013; Kleina et al., 2015). So, an increased number of customers make purchase decisions based on social media referrals that are the recommendations of other consumers or the ones impressed by the contents related to a company or its products. Consequently, social-network-based communication is becoming a necessity for companies to sustain success in the current competitive markets.

Twitter is one of the main online social-network-based services in the world. It is a microblog service, where various users including celebrities, news organisations, marketers, and ordinary users find creative ways of expressing themselves in short 140-character or fewer messages, called tweets, and connect with a community of followers. Due to statistics portal, Statista, from the 1st quarter of 2010 to the 4th quarter of 2016, the social media platform the microblogging service averaged at 319 million monthly active users.

On the other side, the advent and development of the internet is enhancing e-shopping and e-selling, which is an alternative channel beside traditional shopping channels; therefore, the current research aims to investigate the factors effective on the success of e-WOM in Twitter social media, one of the most worldwide microblogs, and in the scope of e-retailing. It would give this chance to e-retailing managers and marketers to know the features of a good content for e-WOM and as a result for viral marketing and increase the influence and success of their contents on the social media. It also gives the chance to know the customers who have an influence on their brands on social media channels, based on features like the number of followings and followers and the features of their contents and even cooperate with them in order to have better advertising campaigns. Companies can also cooperate with these customers in producing a common content for e-WOM (Jin and Phua, 2014).

As a whole, the current research mainly concentrates on this scope according to the growth of online retailers and the competition among them, the development of social media communication among consumers, alongside the necessity for companies to harness user-generated e-WOM to better understand consumers so as to create competitive advantage (Dey et al., 2011). In order to be distinct, the current research uses different large-scale data analysis techniques by using data mining techniques and the tools of R programming language leading to the reduction in errors made by tools. The results would be verified being compared to the results from previous studies of the field, which might have been done by other tools; This study also concentrates on e-WOM effectiveness so the definition of influence is different in this research: it is the number of re-tweets plus the number of favourites because being liked or shared means the content has been seen and regarded; current research is done on tweets made by ordinary users not the companies themselves, so the effect of popularity that causes users to share or like a post regardless of the content, would decrease; This paper focuses on e-retailing sector, because e-WOM is more effective on online businesses, since if effective, they can easily lead to a transaction; Text mining is also done in this study. This could demonstrate the main topics of the tweets that are considered as e-WOM; considering time and both the number of followings and followers as they can influence the number of favourites and re-tweets differently is another source of originality in this study.

2 Literature review

2.1 Electronic word-of-mouth

E-WOM is the user's ideas about the company and its product or service evaluations by talking and texting each other in the virtual world (King et al., 2014; Barreto, 2014). E-WOM is different from traditional WOM in various perspectives, including:

1 several customers can receive the same message, which can be retrieved from anywhere at any time (Henning-Thurau et al., 2004)

2 e-WOM can be measured more accurately and controlled more conveniently than its traditional counterpart (Cheung and Thadani, 2012).

Henning-Thurau et al. (2004) defined e-WOM as: "any positive or negative statement made by potential, actual or former customers about a product or company, which is made available to a multitude of people and institutions via the internet." E-WOM is distributed through different internet channels such as consumer forums (web-based consumer-opinion platforms used by consumers to publicise and communicate their opinions), boycott websites, personal e-mails, chat rooms, and instant messages (Cheung and Thadani, 2012). Unlike the traditional WOM, which required face-to-face or telephone communication, e-WOM is text-based with a sender typing a message on the internet and a number of users reading this message on the internet.

The huge quantity of text-based e-WOM messages available on the internet has caused both opportunities and challenges for researchers and managers. One opportunity for instance can be free access to voluminous, authentic consumer information. On the other side, it can be hard to correctly utilise the huge amount of unstructured textual information (Lehto et al., 2007), because it would not be easy to review this huge amount of data and it is quite challenging to find a specific pattern due to not having a specific structure.

2.2 E-WOM and e-retailing

E-WOM is establishing itself as a more influential marketing tool compared to traditional WOM because of its speed, convenience, large domain, and absence of face-to-face human communication and pressure (Phelps et al., 2004; Lee et al., 2013). Recommendations and complaints are among the most commonly used e-WOM formats.

In a survey of 2,000 internet users, more than 70% of the respondents reported that online reviews had a significant influence on their purchase decisions and they also noted that online consumer reviews had a greater influence than the reviews made by professionals (comScore, 2007).

Some prior studies have investigated market performance (i.e., product revenues and diffusion) in relation to e-WOM. Studies (e.g., Chevalier and Mayzlin, 2006; Dellarocas et al., 2007; Joeckel, 2007) have proved by relatively consistent evidence that the volume and valence of e-WOM are significantly effective on product sales and stock prices. As an instance, due to Chevalier and Mayzlin (2006), favourable consumer ratings were positively related to book sales. Chu and Kim (2011) suggested that social network factors (e.g., tie strength) affected consumers' e-WOM behaviours. Sparks et al. (2013) found that both the origin and content of online reviews influenced consumers' opinions, attitudes and purchase decisions. E-businesses present a more widespread variety of products and services than do traditional businesses. According to this feature, product choices available on the Internet overwhelm consumers by causing purchase decisions to get harder. E-WOM has become increasingly fundamental in enabling consumers to make purchase decisions more conveniently and get less confused. Nielson demonstrated that 92% of 28,000 internet users in 56 countries rely on the recommendations of friends and their families, and 70% of them rely on online consumer reviews (Chaney, 2012).

2.3 E-WOM and social media

Since the advent of social network analysis, there have been efforts to identify the most effective factors on social networks. The extension of the internet and technology advancement has caused online social networks to become challenging cases of study based on big data content and complex interpersonal relationships coming together. Knowing the influence of users and being able to predict it, can be useful for many applications, such as viral marketing, controlling e-WOM process and analysis of consumer's ideas about company and its products [information propagation (Golbeck and Hendler, 2006); searching (Lada, 2005), expertise recommendation (Song et al., 2006), social customer relationship management (Li et al., 2014), percolation theory (Morone and Makse, 2015), etc.].

Social media have become of the most prevalent e-WOM channels because of their unique specifications such as ubiquity, mobility, and interactivity (French and Read, 2013; Zmuda, 2012). Social media enable users to communicate and connect with each other more frequently, closely and easily (Laroche et al., 2013; Kleina et al., 2015). According to Boyd and Ellison (2007), social networking websites are described as web-based services, which provide internet users an opportunity to have their own profiles and networks via friends lists while allowing them to reach others. Therefore, consumers can find a chance to talk about the product information that affects their purchase decisions with their friends on social media (Kozinets et al., 2010). Social media users can either create their own contents related to their opinions which can be in the form of posts, comments, likes, or sharing other users' posts. In fact, they generate their own networks together with other users with whom they have similar ideas about a product or service, so finding potential and loyal consumers can be more convenient for companies through this media.

2.4 Social media and user influence

In order to know what happens in e-WOM processes and what are the influential factors, it is necessary to know the specifications of the users who have the power to influence purchase decisions of other users because of their authority, knowledge, position or relationship (BusinessDictionary, 2015), known as influencers in literature. Based on Verhoef et al. (2013), social networks involve value co-creation because neither are all the contents made by users equally interesting, nor are all the customers equally influential due to their different specification.

According to the theory of influential, only a few members of a society are able to be impressive when presenting ideas to other users. So, by satisfying a few influencers and therefore at a lower cost, we could achieve extremely positive results due to the effects of influencers on a big network (Messias et al., 2013). Users are considered active when their activity in a social network is constant and frequent in a period of time, regardless of the attention they receive for their participation. Considering this feature, there may be very active Twitter readers, who cannot be identified by current metrics, because they leave no trace on the network. Therefore, 'participation' in a social network is specified as the metrics that can be measured like doing tweets, re-tweets, mentions, replies, etc. In this sense, Yin and Zhang (2012) define the activity of a user as the probability of seeing a tweet by a user. Determining if a user has seen a tweet is impossible, but if a user re-tweets a tweet, we would consider it as being read by the user; so re-tweeting is one of the measures of influence.

Re-tweets and likes also involve an implicit interaction between the user's action and the author of the original tweet (Riquelme and González-Cantergiani, 2016). Perhaps the simplest activity measure is the tweet rank (Nagmoti et al., 2010), which is just a metric that counts the number of tweets of a user. Tweet count score (Noro et al., 2012) is another way of measuring the influence. It counts the number of original tweets plus the number of re-tweets. Therefore, visible activities of a user can be measured as influence metrics.

Due to the discuss rank (Jabeur et al., 2012), a user is considered active if he initiates conversations around a topic. It is based on multigraphs or graphs with multiple edges between two nodes. In this algorithm, nodes are the users, and the edge is based on followers, re-tweets and mentions among users.

The IP influence (Romero et al., 2011) measures both the users' influence and their passivity. The passivity of a user is defined as the difficulty for the user to be influenced by another in some period of time, which means they do not cooperate with the spread of content over the network. This algorithm is defined by considering metrics of re-tweets, followers and followings.

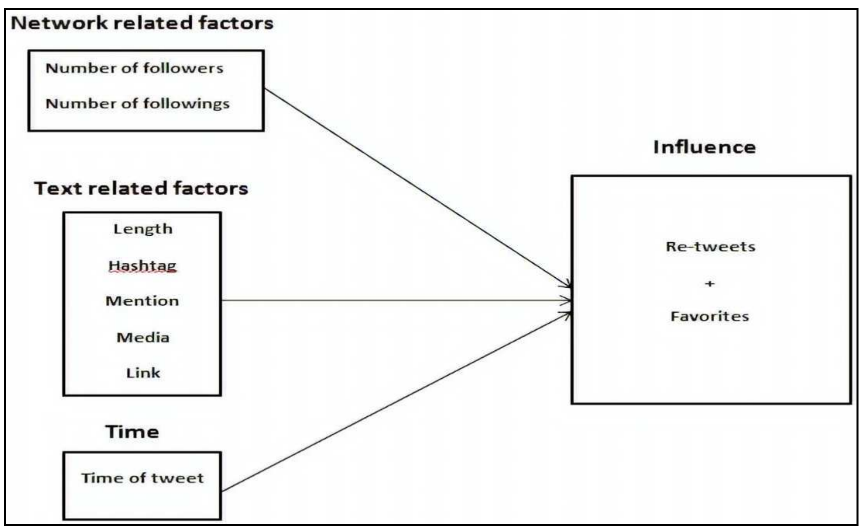
Considering previous research findings, the current research limits the scope of the study to e-retailing market as a highly competitive market and relies on original user-generated data to understand consumer response toward e-WOM about e-retailers in Twitter. This study also applies different data mining and text mining techniques so as to investigate if different data analyses tools can make more validity in results and eliminate potential errors.

3 Research conceptual model and research questions

Considering all measures and factors mentioned above leads us to some possible important influential factors that might determine content effectiveness in Twitter. This research considers the total number of re-tweets plus favourites for each content as a measure of influence and accordingly as content effectiveness measure. We also assumed that user's network specifications including the number of followers and the number of followings are related to the content effectiveness, which he/she posts on the network.

Based on an understanding of the content marketing literature, this survey considers that some of tweets' features might also contribute to the effectiveness of content in terms of being re-tweeted or liked (that is called favourite in Twitter social media). These important features are using media, URLs, mentions, and hashtags in a tweet and the length of it. Finally, we assume that the time of posting might have an effect on being spread in the network. The conceptual model comprising of all potential influential factors is represented in Figure 1.

Figure 1 Research conceptual model



3.1 Followers

The number of followers is an indicator of influence in social media. Since users try to make social capital by their networks to promote their tweets (Ghosh et al., 2012), their followers are an indicator of being influential. Here we consider that the high number of followers can cause a user to have a stronger network, and therefore leads to receiving more tweets from other users and also more followings as a result of being an influential user. The first question is: does the number of followers affect content influence (made up of the sum of re-tweets and favourites of a post) in Twitter social media?

3.2 Followings

The high number of followings causes a user's posts to appear on their Twitter newsfeed and leads to being more re-tweeted and liked, resulting in spreading an e-WOM content (Jin and Phua, 2014; Lahuerta-Otero and Cordero-Gutiérrez, 2016). In the second question, this research intends to investigate if the number of followings is effective on the tweet influence, in Twitter social media.

3.3 Length

Length, measured by the number of characters of a tweet, is another independent variable. Previous research works suggest different results for the effect of length. Bennet (2014) found that a tweet of a length of 71–100 characters is ideal. Enge (2014) also proved that the larger the tweet is, the more effective it will be on influence. Lahuerta-Otero and Cordero-Gutiérrez (2016) declared that longer tweets are less influential. Some researchers accept this reverse relation and believe that in long tweets, other users cannot add other characters (Abolqami et al., 2015). In the third question, this research investigates if the length of a tweet has any effect on content influence in Twitter social media.

3.4 Media

Media includes images and videos. Media can make content more interesting. Hubspot's social-media scientist, Zarrella (2009) found that tweets with images are 94% more likely to be re-tweeted than tweets without it. Consequently, the fourth question is: does including the media in a tweet affect its influence in Twitter social media?

3.5 Link

More than half of re-tweets contain a link (Zhang et al., 2014). Using link can cause a tweet to have a detailed description and compensate the restrictions for the number of characters included in a tweet. There are different and even contradictory beliefs about link effect on tweet influence. For example, according to Malhotra et al. (2012), in a study on 47 well-known companies, and based on Lahuerta-Otero and Cordero-Gutiérrez (2016), using links would not help posts to be re-tweeted. In line with that, Lee (2015) and Ross (2014) proved that using links has no effect on influence. While Bongwon et al. (2010) and also Zarrella (2009) found that tweets including link are more likely to be re-tweeted. Although, some authors believe that using links does not help tweets to get more engagement such as re-tweets, favourites, and mentions (Lee, 2015), this research intends to investigate if including a link in a tweet affects the influence in Twitter social media.

3.6 Hashtag

The number of hashtags is a way of measuring engagement (Lee, 2015; Pal and Counts, 2011). Using hashtag enables users to identify trending topics and have reactions to it. So, by using hashtag the e-WOM contents about a special company can be found more conveniently, and consequently the probability of being re-tweeted or liked will increase.

Due to Boyd et al. (2010) almost one-fifth of tweets have a hashtag. So, we investigate the effect of hashtag on influence in Twitter social media.

3.7 Mention

Mentions are of contextual characteristics of tweets (Boyd et al., 2010). According to Boyd et al. (2010), more than nearly one-tenth of re-tweets have a mention inside. Using mention makes this chance for users to get alert if their name is included in a tweet, so they can engage in it by re-tweeting or liking it (Enge, 2014). So by using mention, more users would be called to see a tweet leading to more favourites and re-tweets. The next question might address the effect of using mentions on influence in Twitter social media.

3.8 Time

According to Lee (2014) and based on his survey on 4.8 million tweets across 10,000 profiles, early morning hours appear to be the time in which tweets receive more clicks on average. Late at night and evenings are also the times, effective on increasing re-tweets and favourites. This research also considers the potential effect of time of posting a tweet on its effectiveness as it has been less investigated in previous studies.

4 Research methodology

4.1 Sample selection

In order to investigate the research questions in e-retailing sector we have chosen two of the most famous American e-retailing companies, eBay and Amazon, based on the high speed of effect that e-WOM has on online businesses and can instantly cause a transaction and the worldwide reputation and success of these two companies in this sector. In 2016 Amazon became 11 among the world's most innovative companies on Forbes lists, 237 in Global 2000 including rank 44 in sales, 998 in profit and 387 in assets, 8 in market value, and 12 in world's most valuable brands. In 2016, on Forbes lists, eBay also became 466 in Global 2000, including 686 in sales, 347 in profit, 1,213 in assets, and 371 in market value, it also took ranking 166 among America's best employers.

Data is collected from an online dataset website, http://followthehashtag.com, which collects data about different companies which are active on Twitter. The data is retrieved by following the hashtags. The primary data is 62,246 tweets that are collected from 28th March 2016 to 15th June 2016. The final data after cleaning the dataset and cutting out the repetitive data consists of 48,129 tweets on both companies. The data includes all tweets from ordinary users that were creating posts, using either 'eBay' or 'Amazon' as keywords.

4.2 Measures

Based on research conceptual model depicted in Figure 1, the independent variables of this research are the number of followers, the number of followings, hashtag, mention, media, link, length and time. The contents of tweets are also used for doing some text mining analysis for supplementary findings.

In order to improve the results of the analysis, the best measure for each of the variables is chosen such as using binary measures for hashtag and mention because of their low dispersion. Binary measure for the link is also used, because it is important whether the content includes a link or not. This also applies to the media. Influence, the dependent variable of this study, is measured by the sum of re-tweets and favourites of each tweet based on literature and because in our research, the spread of a tweet and being seen by other users matter for investigating the success of e-WOM. So, here influence is considered as the sum of re-tweets and favourites. Given that the posting users are usually ordinary people, the number of re-tweets and favourites are mostly in the range of 1 to 10 and data includes lots of posts with zero Influence. For balancing the dispersion of the data, two categories have been generated, i.e., zero influence and non-zero influence, which would be called categories A and B. In the second step, three categories are introduced, i.e., zero influence, influence from 1 to 10, and influence more than 10, which would be called categories A, B and C. To apply each algorithm, the data is separated into training data, including 70% of the data for doing analysis and the other 30% as test data for verifying the algorithm that is done by software.

4.3 Methods

As the goal of the current research is to find patterns and correlations within a large dataset including a large number of variables and samples, data mining and text mining techniques are reliable and appropriate tools and the large variety of their techniques helps in verifying the results and choosing the best among them; these methods are also an innovation introduced in this study and make it distinct from previous studies in the field.

Among the existing classification methods in data mining, decision tree is chosen for analysis because the most relevant input variables should be considered to form decision tree models. Variable importance is computed in this method based on the reduction of model accuracy when the variable is removed and by using the tree model derived from historical data, it would be easy to predict the result for future records (Song and Lu, 2015). Different algorithms are used to find the best answers that are able to verify each other. Decision tree algorithms such as regression tree, CART, and random forest are applied to data. To perform text mining, tm package is used in R software in order to identify the frequent items, associations of special words, topics and count the most commonly used words in e-retailing sector. Finally, some sentiment analysis has been done by using sentiment lexicons of Stanford University. R programming language, version 3.1.2, is used for analysing data by using relative packages and data mining algorithms.

Regression tree is formed by rpart function from rpart package in R software. By this function, the decision tree is drawn and variable importance is calculated to identify the most important variables of the tree. CART decision tree is drawn by applying Ctree function from the party package, in depth 3 and 4. The algorithm is tested by the test data with prediction function. Tree function from Tree package is also used to draw a decision tree in order to verify the results of other decision trees. By applying random forest function from random forest package on the data, a multitude of decision trees are constructed, including 100 to 500 trees at training time and output the class that is the mode of the classes or mean prediction of the individual trees. The mean decrease accuracy and mean decrease Gini of variables are calculated by applying random forest algorithm to investigate variables' importance. The algorithm is tested by predicting the influence of the test data.

5 Research findings

In order to distinguish the variables when abbreviated in decision trees, 'following' is replaced with 'tracing', link with URL, mention is written with a lowercase letter, and media is abbreviated with a capital letter. First of all, Ctree function is applied on the data after cleaning it and categorising it into two classes, A and B. 'A' refers to the data with zero influence and 'B' refers to the data with non-zero influence. Ctree function is applied in depth 3 and as Figure 2 represents, length is the most important variable due to being the root of the tree. Based on the results, the lengthier a tweet is, the more probable it will be for it to be categorised into class B. Having no media falls into the class A, but URL does not make fundamental changes. The more the number of followers, the more the probability of being categorised in class B. Length of more than 139 characters, neutralises the effect of media and the data either with media or without it, stay mostly on class B. Having a length of more than 139 characters along with media, increases the percentage of class B. By applying Ctree in depth 4, link, followers, media and length appear to be important variables, but they do not make much difference between branches. It verifies that length is one of the effective variables as it increases the percentage of class B by getting more then 119 or 102 characters. By testing the algorithm with the test data in confusion matrix, 67% of the data is found to verify the algorithm. By testing it with training data, 68% confidence is also found in Ctree, applied in depth 3. Refer to Tables 1 and 2.

Figure 2 Ctree, depth = 3, two-class categorised data

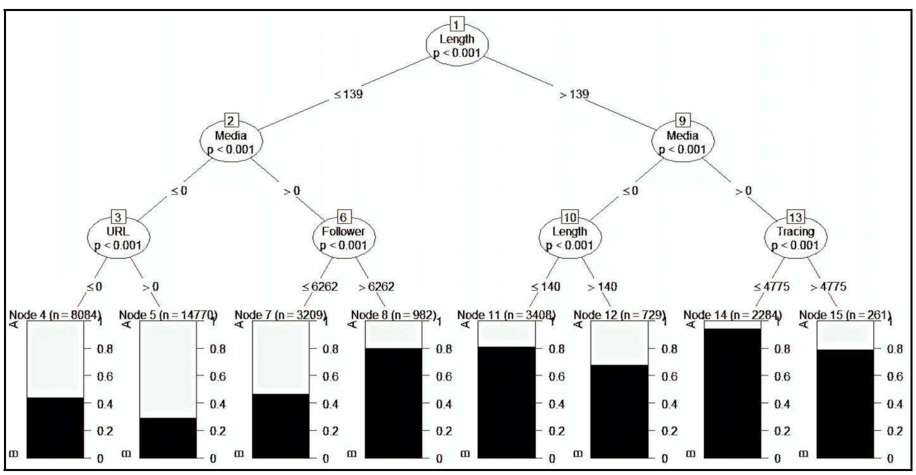


Table 1 Confusion matrix for testing Ctree on train data with depth 3 on two-class categorised data

Categorisation	*A*	*B*
A	16,740	9,323
B	1,277	6,387

Table 2 Confusion matrix for testing Ctree on test data with depth 3 on two-class categorised data

Categorisation	*A*	*B*
A	7,129	4,069
B	603	2,601

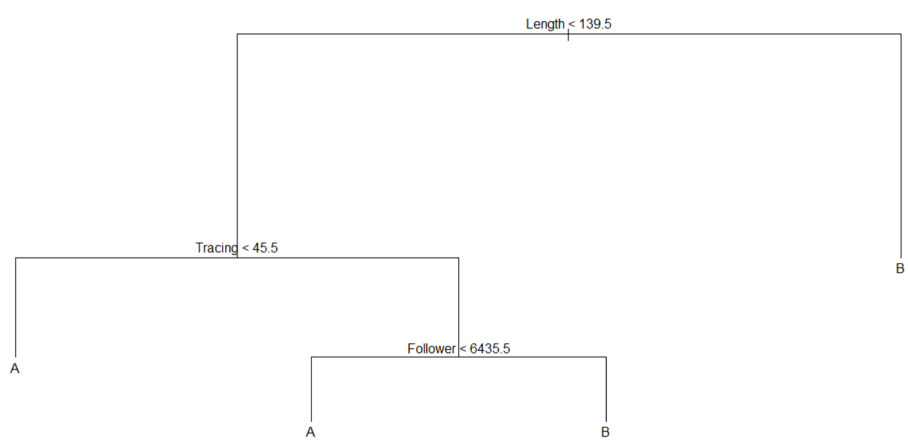
In depth 4, 69% and 68% accuracy is also gained in a confusion matrix for train and test data, respectively. To verify the above results, other decision tree algorithms such as tree classification, regression tree and random forest are tested. In regression tree as shown in Figure 3, length of the tweet and the number of followers are the most important variables, which have a direct relationship with influence. According to Figure 3, length of more than 139.5 characters together with the number of followers of more than 6,411, leads to being included in class B. length of more than 139.5 characters also leads to class B regardless of the number of followers. By applying importance function as for Table 3, in this algorithm, we find length, followers and followings as the most important variables respectively.

Table 3 Variable importance identified in regression tree, two-class categorised data

Variable	*Length*	*Follower*	*Tracing (following)*
Importance	2,318.8179	972.5409	228.8587

In tree function, the number of followers, followings and length are among the fundamental variables in the decision tree with a direct relationship. You can see the results in Figure 3.

Figure 3 Tree classification, two-class categorised data

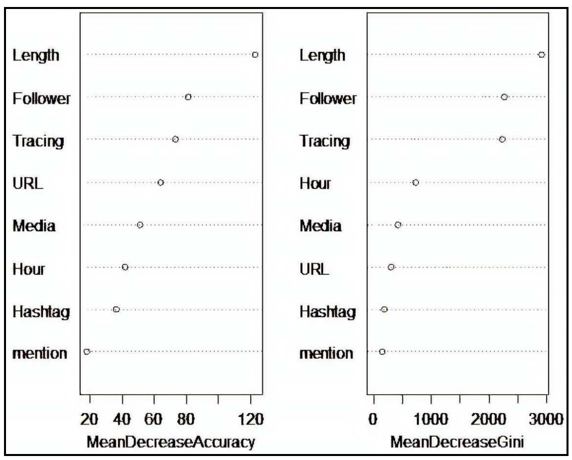


In random forest, 500 trees with OOB (out of bag error) have been plotted equal to 23.56% and 76% of the data verifying the algorithm; as Table 4 demonstrates by testing training data. By drawing the diagram of mean decrease Gini and mean decrease accuracy for random forest algorithm, length, follower and following become the most influential variables based on having bigger mean decrease Gini and mean decrease accuracy. Please see Figure 4. The mentioned variables have more importance based on importance function in random forest algorithm. By testing the algorithm with the test data, we find 75% accuracy in the model.

Table 4 Confusion matrix, testing by training data prediction, number of trees: 500, OOB estimate of error rate: 27.88%

Categorisation	*A*	*B*	*Class error*
A	15,056	3,051	0.1684984
B	4,908	10,772	0.3130102

Figure 4 Random forest algorithm, two-class categorised data



To conduct a deeper investigation and decreasing the human errors in classification of influence, an analysis considering three classes in data has been deployed. The three classes considered as A, B, and C are defined as: A: influence = 0, B: influence = < 10, C: influence > 10.

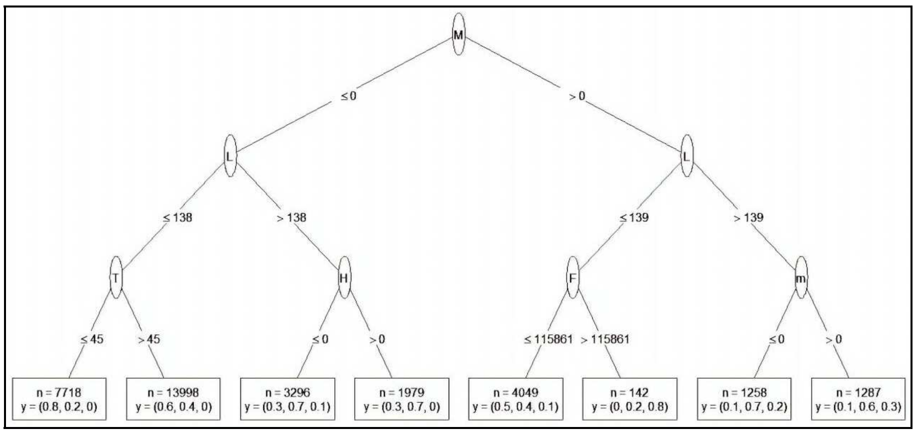
By applying Ctree function, in depth 3, on training data, based on Figure 5, media is the most important variable in a decision tree. In the second depth, there is a length limit, defined by being less than 138 characters, the C class goes to zero percent and by increasing the length, influence goes up. In the third depth, we have the number of followings that has a direct relation with influence. Another influencing variable is the number of followers that highly increases the percent of class C. By testing the algorithm with training data, 64% of validity is derived and by examining the test data, the precision of 63.56% is obtained.

Exerting depth 4 by Ctree function, in the fourth depth, again follower, following and length make considerable differences. In this depth the validity of 71% and 67% for training and test data is obtained by testing the model. In regression tree and on three-class categorisation, length, the number of followers, and the number of followings are the directive variables, which have a direct relationship with influence. Applying importance function on this function, again length, following and follower become the most important variables. You can see results in Table 5.

Table 5 Variable importance identified in regression tree, three-class categorised data

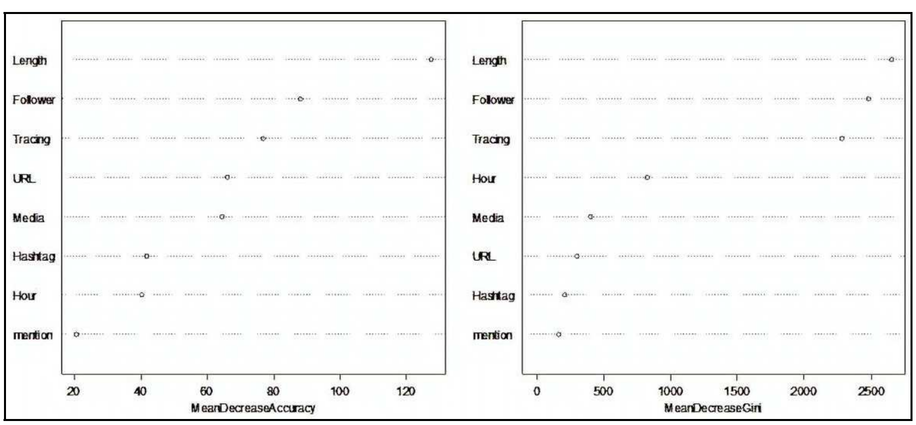
Variable	*Length*	*Follower*	*Tracing (following)*
Importance	2,318.8179	972.5409	228.8587

Figure 5 Ctree, depth = 3, three-class categorised data



In tree function, the number of followers, the number of followings and length are found as fundamental variables that have a direct relationship with influence. In random forest algorithm, with 72% and 71% of validity based on confusion matrix on training and test data respectively, the variables of length, follower and following are the most important variables as they have remarkably bigger mean decrease Gini and mean decrease accuracy rather than the other variables. See Figure 6.

Figure 6 Random forest algorithm, three-class categorised data



Along with the mentioned algorithms analyses, text mining is applied on the contents of all tweets by identifying frequent items, associations, frequent topics, wordcloud and sentiment analysis. As a result, no special difference is recognised between zero influence contents and non-zero ones, but some general results are driven for the whole contents; By finding frequent items and wordcloud, results show that the names of the companies and the word 'stock' are the most frequent items and the names of companies like Apple, Google, Tesla and Netflix are among frequent items too. By finding the top 15 topics out of the most frequent ones, we find that the most used topics are about comparing the stocks of these two companies with the above-mentioned ones. Words like 'buy' and 'sell' are also used in topics frequently. You can see five topics of fifteen frequent topics in Tables 6 and 7 that mostly include the names of famous companies and the word 'stock'.

Table 6 Top topics for zero influence tweets

Top topics' rank	*Special words of topic*
1	'amzn, buy, aapl, amazon, stock, spi, ebay, short, twtr, call'
2	'amzn, amazon, inc, buy, read, stock, ebay, googl, share, nflx'
3	'amzn, amazon, inc, googl, ebay, msft, read, goog, sell, day'
4	'amzn, ebay, read, amazon, stock, price, trade, see, high, goog'
5	'amzn, ebay, read, googl, price, nflx, earn, msft, new, target'

Table 7 Top topics for non-zero influence tweets

Top topics' rank	*Special words of topic*
1	'googl, ebay, amazon, tsla, amzn, twtr, nflx, https, spi, will'
2	'amzn, stock, spx, amazon, nflx, market, ebay, qqq, twtr, top'
3	'amzn, aapl, spi, trade, qqq, amazon, option, tsla, get, twtr'
4	'amzn, aapl, nflx, amazon, spx, goog, earn, googl, trade, tsla'
5	'amzn, aapl, stock, amazon, googl, qqq, tsla, spi, lnkd, sell'

Deploying sentiment analysis also does not retrieve any meaningful differences, but as a whole the words are mostly neutral and the others are mostly positive rather than negative. Results are demonstrated in Tables 8 and 9 and show that most of words in both, zero influence tweets and non-zero ones are neutral.

Table 8 Sentiment analysis on tweets with zero influence

Level of sentiment	*–5*	*–4*	*–3*	*–2*	*–1*	*0*	*1*	*2*	*3*	*4*	*6*	*13*
Frequency	1	3	19	82	5,622	280,930	7,965	160	21	3	1	1

Table 9 Sentiment analysis on tweets with non-zero influence

Level of sentiment	*–4*	*–3*	*–2*	*–1*	*0*	*1*	*2*	*3*	*4*	*6*	*16*
Frequency	1	8	116	5,258	281,453	7,794	206	19	4	1	1

6 Discussion

E-WOM, especially in social media, has become one of the most important marketing instruments for companies in the current competitive market. Social media including Twitter is one of the most important sources of e-WOM, in which people get involved in spreading contents and engaging in viral marketing process by creating a network of friends. By investigating e-WOM contents and the features of the users, marketers would be able to distinguish between influential and non-influential users. As a result, companies can make these users involved in their advertising campaigns and in spreading e-WOM through different platforms. They can even know the features of an effective content leading to notions that can be considered in viral marketing and even marketing campaigns (Kumar and Mirchandani, 2012). Consequently, knowing the effective features can help firms create more effective contents. Data mining is of the most practical techniques for investigating data and finding patterns in them. Data mining and other analytical techniques are widely used by firms to extract knowledge from marketing data and might be easily applied for marketing purposes in social networks as well. Current research identifies the features of an effective content along with the features of the user spreading it that includes the number of followings and followers. In this research, being effective is called influence that is the number of re-tweets plus the times a post is liked. The effective factors, which can help companies spread positive contents by viral marketing and get to know the users with effective contents are analysed. These influential characteristics would be considered in order to encourage social network users to spread positive e-WOM (Chu and Kim, 2011).

Accordingly, in this research some assumptions mentioned in literature about the effectiveness of factors, and some additional ones in social media are evaluated. Considering network related factors as well as text related specifications to investigate the most influential factors role play in tweet diffusion, exclusively in e-retailing market along with data mining and text mining techniques, make the current research exceptional.

Deploying data mining analysis process as mentioned in the latter part, we find that the number of followers, followings, and text length are the most effective factors that have a direct relation with influence. Based on e-WOM purpose, the message must be seen or re-tweeted in Twitter social media; this is why favourite is considered as a factor of being seen. The high number of followers and followings cause a post to be seen more and proportionally, the possibility of being re-tweeted increases if seen more. As the results prove, in very high numbers of followers and followings, influence increases to more than 10. In conformance to literature, also the number of individuals following a user indicates popularity (Bigonha et al., 2010). Consequently, users that follow others indirectly increase their popularity, so more users can see tweets and some of them might eventually engage with tweets, as they are perceived as a source of credibility (Jin and Phua, 2014).

The results also demonstrate that the length of a post that is retrieved from the number of characters is effective on influence. Considering the text length, we find a direct relation between length and influence that is also proved in some previous studies such as Tao et al. (2012). Enge (2014) in line with the current research, supposes that the longer the tweet, the better the chances of engagement (in the form of re-tweets and favourites).

By applying Ctree function on both, two-class categorised and three-class categorised data in depth 3 and 4, media is reflected as an impressive factor, which means having media possibly increases the effectiveness of a post. Including media, the content is expected to be more interesting and have more fun for users, leading to more favourites and re-tweets. Based on the natural structure of tweets, it is easy to distinguish quickly, especially when they include figures or videos. According to Rogers (2014), users who are looking at a Twitter page with very little time, will more likely see those tweets that contain images or videos and this would increase the possibility of sharing the content.

Time of tweeting a post has not been deducted in decision trees, but according to Lee (2014), early morning hours appear to be the time in which tweets receive more attention. Time is seen in logistic regression as supplementary analysis. In this algorithm, time has indirect relation with influence and this is in line with the research done by Lee (2014) but because of having large null and residual deviance, the result is not remarkable.

Link, mention, and hashtag are seen in the latter depths of the trees drawn by Ctree function and have a direct relation with influence but their presence makes no specific difference between branches. Also, regarding the larger mean decrease Gini and mean decrease accuracy in random forest function, link, hashtag, and mention are the least important variables, respectively. Literature shows that hashtag and mention have a direct relation with influence because they help users to find subjects based on their interests more conveniently (Pal and Counts, 2011; Enge, 2014; Lee, 2015). Nevertheless, data analysis in the current dataset does not exhibit any significant relation. In some previous studies, link is also proved not to help the content to be liked or re-tweeted more (Lee, 2015; Ross, 2014). The reason maybe that clicking on the link leads the users to another page and again makes getting information harder. It might also look like advertising and not an e-WOM content from an ordinary user.

7 Conclusions

This study suggests that practitioners and marketers can explore the tweets with the specifications mentioned as effective factors on influence and find their influential users in Twitter social media. They can cooperate with these users in social media marketing and their marketing campaigns. They can also make effective contents, regarding the effective factors such as media and the length of contents. Another effective factor based on the current research is the network of a user. Accordingly, marketers must extend their networks in social media by increasing the number of followers and followings in order to help contents broadcast more easily, leading to better viral marketing.

For further research, it is proposed to conduct a data analysis on a larger number of tweets, in a longer period of time, or different time periods to determine the robustness of results. It could also be done on more than two e-retailing companies (eBay and Amazon) so that they involve different countries and not just the USA. Also, other competitive industries such as telecommunications and Hi-tech companies could be studied. Other software programs such as Python and RapidMiner could be applied to verify the results of the R software. This research is only focused on Twitter social media; and further research works could be planned on other social media platforms or other social networks to analyse the network effect.

In text mining, words in special categorisations (i.e., goods, stocks, other news, special events, and so on) can be searched in tweets. In this way, the difference between the contents of two categorisations of influence (zero influence and non-zero influence) can be better analysed. The main filter in this survey is two keywords, eBay and Amazon, but we have not explored if the contents are positive or negative, if they are directly related to the company or not, and if they are advertisements or just news or other features. Basically, we have assumed that all tweets of these companies are simply contents about them that can attract users' attention, even just as a reminder of their names. Still, in further research works the content of tweets can be analysed and maybe only some certain contents could be considered as e-WOM.

References

Abolqami, H., Karaghouli, W.A., Baeshen, Y., Erkan, I., Evans, C. and Ghoneim, A. (2015) 'Electronic word of mouth in social media: the common characteristics of retweeted and favourited marketer-generated content posted on Twitter', *International Journal of Internet Marketing and Advertising*, Vol. 9, No. 4, p.338, DOI: 10.1504/ijima.2015.072886.

Asur, S. and Huberman, B. (2010) 'Predicting the future with social media', in *WI-IAT*, Vol. 1, pp.492–499.

Barreto, A.M. (2014) 'The word-of-mouth phenomenon in the social media era', *International Journal of Market Research*, Vol. 56, No. 5, pp.631–654.

Bennet, S. (2014) *The Optimal Length of a Tweet, Hashtag, Podcast, Slideshare and (Almost) Everything Online* [online] http://www.adweek.com/socialtimes/ideal-length-social-media-podcast-blog/502513 (accessed 12 December 2017).

Bigonha, C., Cardoso, T.N., Moro, M.M., Almeida, V.A. and Gonçalves, M.A. (2010) 'Detecting evangelists and detractors on Twitter', in *18th Brazilian Symposium on Multimedia and the Web*, pp.107–114.

Bongwon, S., Lichan, H., Peter, P. and Chi, E.H. (2010) 'Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network', in *IEEE International Conference on Privacy, Security, Risk and Trust*, pp.177–184.

Boyd, D., Golder, S. and Lotan, G. (2010) 'Tweet, tweet, retweet: conversational aspects of retweeting on Twitter', in *2010 43rd Hawaii International Conference on System Sciences (HICSS)*, IEEE, pp.1–10.

Boyd, D.M. and Ellison, N.B. (2007) 'Social network sites: definition, history, and scholarship', *Journal of Computer-Mediated Communication*, Vol. 13, No. 1, pp.210–230.

BusinessDictionary (2015) *Definition of Influencer* [online] http://www.businessdictionary.com /definition/influencers.html (accessed 10 December 2017).

Chaney, P. (2012) *Word of Mouth Still Most Trusted Source Says Nielson; Implications for Social Commerce* [online] https://digitalwellbeing.org/word-of-mouth-still-most-trusted-resource-says-nielsen-implications-for-social-commerce/ (accessed 16 April 2012).

Chen, Y. and Xie, J. (2008) 'Online consumer review: word-of-mouth as a new element of marketing communication mix', *Management Science*, Vol. 54, No. 3, pp.477–491.

Cheung, C.M.K. and Thadani, D.R. (2012) 'The impact of electronic word-of-mouth communication: a literature analysis and integrative model', *Decision Support Systems*, Vol. 54, No. 1, pp.461–470.

Chevalier, J.A. and Mayzlin, D. (2006) 'The effect of word of mouth on sales: online book reviews', *Journal of Marketing Research*, Vol. 43, No. 3, pp.345–354.

Chu, S.C. and Kim, Y. (2011) 'Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites', *International Journal of Advertising*, Vol. 30, No. 1, pp.47–75.

comScore (2007) *Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior* [online] https://www.comscore.com/Insights/Press-Releases/2007/11/Online-Consumer-Reviews-Impact-Offline-Purchasing-Behavior (accessed 10 December 2017).

Dellarocas, C., Zhang, X.M. and Awad, N.F. (2007) 'Exploring the value of online product reviews in forecasting sales: the case of motion figures', *Journal of Interactive Marketing*, Vol. 21, No. 4, pp.23–45.

Dey, L., Haque, S.M., Khurdiya, A. and Shroff, G. (2011) 'Acquiring competitive intelligence from social media', in *Proceedings of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*, ACM, p.3.

Duan, W., Gu, B. and Whinston, A.B. (2008a) 'Do online reviews matter? – An empirical investigation of panel data', *Decision Support Systems*, Vol. 45, No. 4, pp.1007–1016.

Duan, W., Gu, B. and Whinston, A.B. (2008b) 'The dynamics of online word-of-mouth and product sales – an empirical investigation of the movie industry', *Journal of Retailing*, Vol. 84, No. 2, pp.233–242.

Dwyer, R. (2007) 'Measuring the value of electronic word of mouth and its impact in consumer communities', *Journal of Interactive Marketing*, Vol. 21, No. 2, pp.63–79.

Enge, E. (2014) *Twitter Engagement Unmasked: a Study of More Than 4M Tweets* [online] https://www.stonetemple.com/twitter-engagement-umasked/ (accessed 11 December 2014).

French, A.M. and Read, A. (2013) 'My mom's on Facebook: an evaluation of information sharing depth in social networking', *Behaviour and Information Technology*, Vol. 32, No. 10, pp.1049–1059.

Ghosh, S., Viswanath, B., Kooti, F., Sharma, N.K., Korlam, G., Benevenuto, F. et al. (2012) 'Understanding and combating link farming in the Twitter social network', in *Proceedings of the 21st International Conference on World Wide Web*, ACM, pp.61–70.

Golbeck, J. and Hendler, J.A. (2006) 'Inferring binary trust relationships in web-based social networks', *ACM Trans, Internet. Techn.*, Vol. 6, No. 4, pp.497–529.

Henning-Thurau, T., Gwinner, K.P., Walsh, G. and Gremler, D.D. (2004) 'Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet?', *Journal of Interactive Marketing*, Vol. 18, No. 1, pp.38–52.

Ho-Dac, N., Carson, S.J. and Moore, W.L. (2013) 'The effects of positive and negative online customer reviews: do brand strength and category maturity matter?', *Journal of Marketing*, Vol. 77, No. 6, pp.37–53.

Jabeur, L.B, Tamine, L, and Boughanem, M. (2012) 'Active microbloggers: Identifying inuencers, leaders and discussers in microblogging networks', *Proc. Int. Symp. String Process. Inf. Retr.*, LNCS 7608, pp.111–117.

Jin, S.A. and Phua, J.J. (2014) 'Following celebrities' tweets about brands: the impact of Twitter-based eWoM on consumers' source credibility perception, buying intention, and social identification', *Journal of Advertising*, Vol. 43, No. 2, pp.181–195.

Joeckel, S. (2007) 'The influence of user and online review ratings on the performance of video games in the US market', *Proceedings of DiGRA 2007 Conference*, pp.629–638.

King, R.A., Racherla, P. and Bush, V.D. (2014) 'What we know and don't know about online word-of-mouth: a review and synthesis of the literature', *Journal of Interactive Marketing*, Vol. 28, No. 3, pp.167–183.

Kleina, A., Ahlfb, H. and Sharmac, V. (2015) 'Social activity and structural centrality in online social networks', *Telematics and Informatics*, Vol. 32, No. 2, pp.321–332.

Kozinets, R.V., de Valck, K., Wojnicki, A.C. and Wilner, S. (2010) 'Networked narratives: understanding word-of-mouth marketing in online communities', *Journal of Marketing*, Vol. 74, No. 3, pp.71–86.

Kumar, V. and Mirchandani, R. (2012) 'Increasing the ROI of social media marketing', *MIT Sloan Management Review*, Vol. 54, No. 1, p.55.

Lada, A. (2005) 'How to search a social network', *Social Networks*, Vol. 27, No. 3, pp.187–203.

Lahuerta-Otero, E. and Cordero-Gutiérrez, R. (2016) 'Looking for the perfect tweet. The use of data mining techniques to find influencers on Twitter', *Computers in Human Behavior*, Vol. 64, pp.575–583 [online] http://www.sciencedirect.com/science/article/pii/S0747563216305258, DOI: 10.1016/j.chb.2016.07.035.

Laroche, M., Habibi, M.R. and Richard, M.O. (2013) 'To be or not to be in social media: how brand loyalty is affected by social media?', *International Journal of Information Management*, Vol. 33, No. 1, pp.76–82.

Lee, K. (2014) *The Optimal Length for Every Social Media Update and More* [online] https://blog.bufferapp.com/optimal-length-social-media (accessed 20 October 2017).

Lee, K. (2015) *What Analyzing 1 Million Tweets Taught Us* [online] https://thenextweb.com/socialmedia/2015/11/03/what-analyzing-1-million-tweets-taught-us/ (accessed 29 July 2015).

Lee, S.H., Noh, S.E. and Kim, H.W. (2013) 'A mixed methods approach to electronic word-of-mouth in the open-market context', *International Journal of Information Management*, Vol. 33, No. 4, pp.687–696.

Lehto X., Park J.K., Park O. and Lehto M.R. (2007) 'Text analysis of consumer reviews: the case of virtual travel firms', in Smith, M.J. and Salvendy, G. (Eds.): *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design, Human Interface 2007, Lecture Notes in Computer Science*, LNCS 4557, pp.490–499, Springer, Berlin, Heidelberg.

Li, J., Peng, W., Li, T., Sun, T., Li, Q. and Xu, J. (2014) 'Social network user influence sense-making and dynamics prediction', *Expert Syst. Appl.*, Vol. 41, No. 11, pp.5115–5124.

Liu, Y. (2006) 'Word of mouth for movies: its dynamics and impact on box office revenue', *Journal of Marketing*, Vol. 70, No. 3, pp.74–89.

Malhotra, A., Malhotra, C.K. and See, A. (2012) 'How to get your messages retweeted', *MIT Sloan Management Review*, Vol. 53, No. 2, pp.61–66.

Messias, J., Schmidt, L., Oliveira, R. and Benevenuto, F. (2013) 'You followed my bot! Transforming robots into influential users in Twitter', *First Monday*, Vol. 18, No. 7.

Morone, F. and Makse, H.A. (2015) 'Influence maximization in complex networks through optimal percolation', *Nature*, Vol. 524, No. 7563, pp.65–68.

Nagmoti, R., Teredesai, A. and De Cock, M. (2010) 'Ranking approaches for microblog search', *2010 IEEE/WIC/ACM International Conference on Web Intelligence, Main Conference Proceedings*, IEEE Computer Society, Toronto, Canada, August 31–September 3, pp.153–157.

Noro, T., Ru, F., Xiao, F. and Tokuda, T. (2012) 'Twitter user rank using keyword search', in *22nd European-Japanese Conference on Information Modelling*, IOS Press, Prague, pp.31–48.

Pal, A. and Counts, S. (2011) 'Identifying topical authorities in microblogs', in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp.45–54.

Phelps, J.E., Lewis, R., Mobilio, L., Perry, D. and Raman, N. (2004) 'Viral marketing of electronic word-of-mouth advertising: examining consumer responses and motivations to pass along email', *Journal of Advertising Research*, Vol. 45, No. 4, pp.333–348.

Qin, L. (2011) 'Word-of-mouth for movies: a predictor and an outcome of box office revenue?', *J. Electron. Commerce Res.*, Vol. 12, No. 3, pp.187–198.

Riquelme, F. and González-Cantergiani, P. (2016) 'Measuring user influence on Twitter: a survey', *Journal of Information Processing and Management*, Vol. 52, No. 5, pp.949–975.

Rogers, S. (2014) *What Fuels a Tweet's Engagement?* [online] https://blog.twitter.com/2014/whatfuels-a-tweets-engagement (accessed 16 April 2015).

Romero, D.M., Galuba, W., Asur, S. and Huberman, B.A. (2011) 'Influence and passivity in social media', in Gunopulos, D., Hofmann, T., Malerba, D. and Vazirgiannis, M. (Eds.): *Machine Learning and Knowledge Discovery in Databases – European Conference, ECML PKDD 2011, Proceedings, Part III*, in Lecture Notes in Computer Science, Athens, Greece, 5–9 September 2011, Springer, Vol. 6913, pp.18–33.

Ross, P. (2014) *Photos Get the Most Engagement on Twitter* [online] http://www.socialbakers.com/blog/2306-photos-get-the-most-engagement-on-twitter (accessed 9 December 2015).

Rui, H., Liu, Y. and Whinston, A. (2011) 'Whose and what chatter matters?', in *The Impact of Tweets on Movie Sales Framework*, Working Papers, University of Texas at Austin, pp.1–29.

Song, X., Tseng, B.L., Lin, C-Y. and Sun, M-T. (2006) 'Personalized recommendation driven by information flow', *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Seattle, Washington, USA, 6–11 August, pp.509–516.

Song, Y. and Lu, Y. (2015) 'Decision tree methods: applications for classification and prediction', *Shanghai Arch Psychiatry*, Vol. 27, No. 2, pp.130–135.

Sparks, B.A., Perkins, H.E. and Buckley, R. (2013) 'Online travel reviews as persuasive communication: the effects of content type, source, and certification logos on consumer behavior', *Tourism Management*, Vol. 39, No. 9, pp.1–9.

Tao, K., Abel, F., Hauff, C. and Houben, G.J. (2012) 'What makes a tweet relevant for a topic', *Making Sense of Microposts*, #MSM2012, pp.49–56

Verhoef, P.C., Beckers, S.F. and Van Doorn, J. (2013) 'Understand the perils of co-creation', *Harvard Business Review*, Vol. 91, No. 9, p.28.

Yin, Z. and Zhang, Y. (2012) 'Measuring pair-wise social influence in microblog', in *2012 International Conference on Privacy, Security, Risk and Trust, PASSAT 2012, and 2012 International Conference on Social Computing, Socialcom 2012*, IEEE, Amsterdam, Netherlands, 3–5 September, pp.502–507.

Zarrella, D. (2009) *The Science of Retweets* [online] http://danzarrella.com/scienceof-retweets.pdf (accessed 15 March 2016).

Zhang, L., Peng, T.Q., Zhang, Y.P., Wang, X.H. and Zhu, J.J. (2014) 'Content or context: which matters more in information processing on microblogging sites', *Computers in Human Behavior*, Vol. 31, No. 1, pp.242–249.

Zmuda, N. (2012) *Pepsi Debuts First Global Campaign* [online] http://adage.com/article/cmo-strategy/pepsi-debuts-global-campaign-live/234379/ (accessed 30 April 2012).