



A Comment-Based Algorithm for Post-Ranking Rapprochement on Facebook

Mahdi Jemmali and Yousef Qawqzeh*

Department of Computer Science and Information, College of Science at Zulfi, Majmaah University, Majmaah 11952, Saudi Arabia

*Corresponding author: y.qawqzeh@mu.edu.sa

Abstract. This work investigates the effects of a comment, in an individual post, voted by a reputed person. The proposed algorithm utilized 10 variables for ranking comment's owner represented by the value of Cor variable. Then the model will analyze how such a vote will affect the rank of that post by increasing the upvotes or by increasing the downvotes. Eight variables are proposed to evaluate the rank of the post represented by the value of GW_p variable. At the end, the overall score of the post will be calculated represented by GS_p variable. Being simple and easy to implement, the proposed method is expected to measure the post-sensitive influence on participants on that given post. However, introducing user's weight (ranking) as a new parameter for the evaluation of post's weight, could highly correct the whole evaluation of post's ranking. As commenters vary in their weights (rankings), posts can be upvoted or downvoted because of commenter's opinion and thought on the given post. This work is novel and aimed at introduce a new method for post ranking that can be utilized for different purposes in different disciplinarians.

Keywords. Post ranking; Commenter's weight; Upvotes; Downvotes; Social media; Facebook

MSC. 65G20

Received: November 12, 2018

Accepted: December 22, 2018

Copyright © 2019 Mahdi Jemmali and Yousef Qawqzeh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The process by which communities come together, attract new members, and develop over time is a central research issue in the social sciences [1]. Social network sites such as Facebook, are becoming a prevalent form of communication between people [5]. Developing a model

that can be used for processing subjective-information effectively demands overcoming some challenges. Like rating users, the importance of an individual user comment, the importance of the topic being discussed, etc. [11]. These challenges greatly affect sentiment analysis or opinion mining process when mining knowledge about the importance of a given post in social media. Opinion mining or sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text [10]. An entropy measure has been used to analyses the behavioral characteristics of users [6]. In [4], a formal model was be presented and a new search algorithm for folksonomies is proposed. There is a need to develop an automated model that can rate users in Facebook, therefore, it can give weights to the post/comment for each individual user participating in social media activities. Influential users play an important role in online social networks since users tend to have an impact on one other [4]. One of the most explored topics in Facebook research is studies that examine personality and individual differences among users [2]. Therefore, the proposed work analyzes users and their behavior to identify influential users and predict user participation [3]. Some works in literature reviews treat the making decision and the ranking models to solve a real-life problems [7] and [9]. The proposed method tries to sound positively by suggesting a simple and easy to implement solution, thereby, the new proposed method could classify the importance of a given post by classifying the ranking of the commenters themselves. Twitter for example still measures a twitterer's influence as the number of followers she has [13]. That's means, the more followers she has, the more impact she appears to make in the Twitter context (she seems popular). Our method suggests that each user will get an updated-rank based on several factors solicited from his account in social media (e.g. Facebook, Twitter). The proposed factors are described in Table 1. The novel idea behind this work is to seek the approximated effect of each user in each post. Some social media users have strong effects on others (because they are famous, rich, political, academic, and so on.) while others may not have as such as effect on other user's opinions. Classifying a user to be likely more to foster a given post or to bias that post towards certain direction is highly needed. Due to the weight of user's comment on a given post, many other people have been influenced and biased their actual opinion. However, the used model needs to present the opinion information it has garnered in some reasonable rating fashion. The following graph represents the road map of the proposed ranking algorithm.

2. Classifier Modeling (Aggregating & Ranking)

Our discussion in this algorithm for classification and extraction of Facebook user rank shall introduce in-depth approach for the development of user rating classifier. Seeking a straightforward classification policy, our model focuses on classifying user's ratings as to their weights (Scale 1:100) as described in Table 2. The reason why we are looking for this important classification is that we want to formulate a ranking policy to a given user and a given post. In terms of opinion mining ("How strong is that user's opinion is?"), ("How many other users are influenced by that strong opinion"), and ("Is user 1 is more important than user 2 and so on. . ."). However, some notions with regards to this classifier should be defined, such as comments count and comment's owner rank.

There are several theoretical and graphical models for the analysis of user behavior in a social network, using different mechanisms [12] and [8]. The mentioned models operate in regimented time steps (nodes and edges). In our proposed mathematical model, we are trying to examine how such a node (user) may infect other nodes (users) based on its rate (rank). Should node x adopt his/ her friend's behavior, node x behavior will be biased. Such influences and infections may contribute badly towards having a “dislike” or “no share” for a given post. On the other hand, these influences and infections may contribute positively towards increasing the amounts of “likes” and “shares” for the same post. Utilizing 10 factors in our algorithm, it will be more accurate to achieve user rating. Based on that, the algorithm assigns certain weight to each user which in turn approaches the correction of user rating process and brings some insights about the importance of some posts and how people behave under the effect of weighted-user commenting on a certain post.

3. Methods

3.1 Comments count: C_c

For all comments written by different users, the comment-based algorithm will treat all comments as a statistical sampling of positive and negative comments voted by everyone in that given post.

3.2 Comment's owner rank: Cor

“Comment's owner rank” can be described generally as for example, if professor gives his/ her opinion by given some comments for a post in Facebook, this opinion can influence the students. So, professor comment has not the same degree of impact as the impact degree of his/her student. As Facebook users must have a rank of their comments, the following variables are included to calculate the interval-value of Cor index:

Table 1. Factors used for user rating in Facebook

#	Factor name	Description
1	NF	Number of friends
2	NC	Number of comments for each post
3	NP	Number of posts per week
4	NS	Number of sharing posts
5	NL	Number of times user connect to Facebook per week
6	NI	Number of waiting invitation
7	NMS	Number of sent message
8	NMR	Number of received message
9	NE	Number of created events
10	NT	Time spend connected in Facebook per week (min)

Every week Facebook updates user-weight impact

We denote by $Cor()$ the function which calculates the comment’s owner rank for each user. So, this function depends on the variables listed above: $NF, NC, NP, NS, NL, NI, NMS, NMR, NE$ and NT .

For each variable, the Facebook Analyzer (FA) fixes an appropriate threshold value denoted by $T()$. The FA is a proposed model that can be implemented by Facebook. As Facebook owns the databases for its users (F DB).

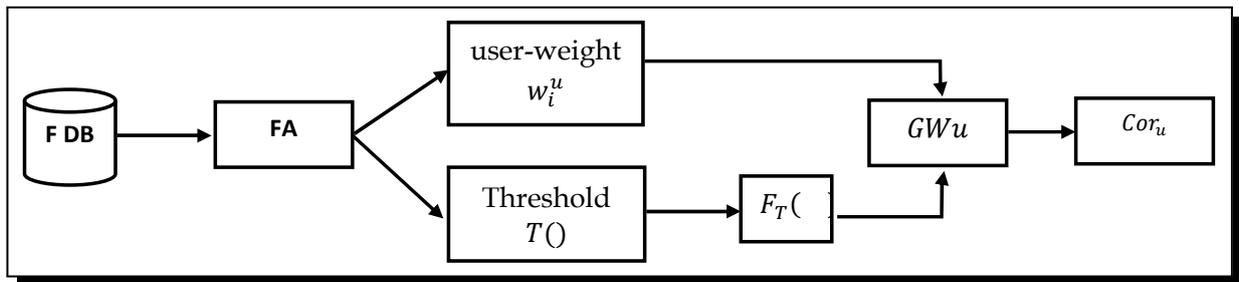


Figure 1. FA-based Cor_u calculation

- $NF : T(NF)$
- $NC : T(NC)$
- $NP : T(NP)$
- $NS : T(NS)$
- $NL : T(NL)$
- $NI : T(NI)$
- $NMS : T(NMS)$
- $NMR : T(NMR)$
- $NE : T(NE)$
- $NT : T(NE)$

For each variable, the (FA) gives an evaluation into interval-based rank [1-10]. This evaluation depends on distance between the value of variable and its $T()$. This evaluation is given by the function $F_T()$.

The FA modelling system is going to calculate the evaluation function $F_T()$. $F_T()$ can be calculated as follow:

$$F_T(VARvalue) = \begin{cases} \frac{VARvalue}{T(VAR)} \times 10; & \text{if } VARvalue \leq T(VAR), \\ 10; & \text{if } VARvalue > T(VAR). \end{cases} \tag{3.0}$$

Each variable has different impact. So, it is important to give for each variable a weight related to its impact. We denoted weight for the variable i by w_i^u . After assigning a weight for each variable by FA, the global user-weight denoted by GW_u can be calculated as:

$$GW_u = w_1^u \times F_T(NF) + w_2^u \times F_T(NC) + w_3^u \times F_T(NP) + w_4^u \times F_T(NS) + w_5^u \times F_T(NL) + w_6^u \times F_T(NI) + w_7^u \times F_T(NMS) + w_8^u \times F_T(NMR) + w_9^u \times F_T(NE) + w_{10}^u \times F_T(NT). \tag{3.1}$$

Now, we can calculate the global comment’s owner rank impact (Cor) for the user by equation (3.2):

$$Cor_u = \left(\frac{GW_u}{\sum_{i=1}^{10} w_i^u} \right) \times 10. \tag{3.2}$$

The output of Cor_u ($NF, NC, NP, NS, NL, NL, NI, NMS, NMR, NE, NT$) can vary as shown in Table 2 as the weighting impact:

Table 2. User weight impact

User	Weight interval
No known user	[1-10]
User rarely connect and activate	[11-20]
User can have little impact	[21-30]
User can have some popularity	[31-40]
User have good popularity	[41-50]
User have excellent popularity and active	[51-60]
Dynamic and not famous user	[61-70]
User has more friends and dynamic	[71-80]
User can be famous	[81-90]
Famous and activate user in domain	[91-100]

Example. If $NF = 100$ and $T(NF) = 200$ so the FA calculates the distance between NF and $T(NF)$ and gives score in [1-10]. For this example, FA will give $FT(NF)$ the value of 5. Thus, the evaluation of variable NF was done. The FA evaluates all remaining variables as the same evaluation of NF . So, let us evaluate the remaining variables in the same way:

$$F_T(NC) = 7, F_T(NP) = 4, F_T(NS) = 5, F_T(NL) = 8, F_T(NI) = 2, F_T(NMS) = 3, \\ F_T(NMR) = 4, F_T(NE) = 2, F_T(NT) = 7$$

Let the weight values:

$$w_1^u = 20, w_2^u = 10, w_3^u = 50, w_4^u = 5, w_5^u = 7, w_6^u = 70, w_7^u = 26, w_8^u = 17, w_9^u = 11, w_{10}^u = 23, \\ GW_u = 4 \times 20 + 7 \times 10 + 4 \times 50 + 5 \times 5 + 8 \times 7 + 2 \times 70 + 3 \times 26 + 4 \times 17 + 2 \times 11 + 7 \times 23, \\ Cor_u = \left(\frac{80 + 70 + 200 + 25 + 56 + 140 + 78 + 68 + 22 + 161}{239} \right) \times 10 = 37.6. \tag{3.3}$$

The weighting impact of user based in Table 1 is: “User can have some popularity”.

3.3 Variable presentation

To implement an intelligent component for the evaluation of a certain post in Facebook, we shall define some related variables as follow:

- L : Number of likes
- R : Rank given by user

- D : Number of dislikes
- S : Number of shares
- C : Number of comments
- W : Average number of words in all comments
- Sa : Satisfaction of comments
- U : Weight of users

For each variable, the Facebook analyzer (FA) fixes a propitiate threshold value denoted by TS .

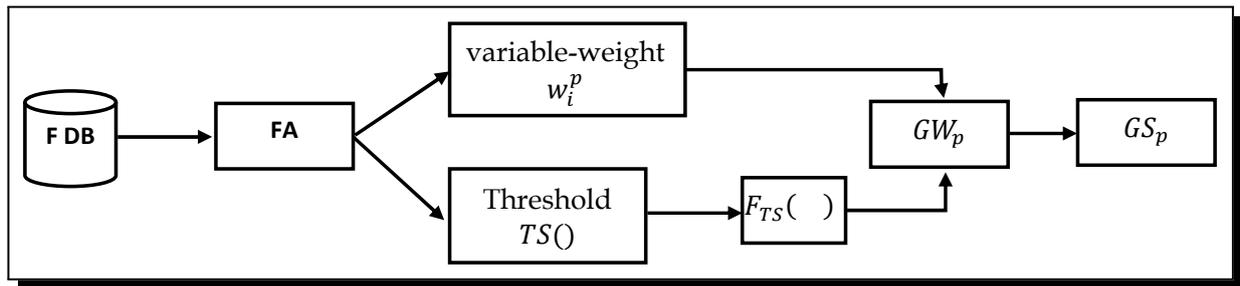


Figure 2. FA-based GS_p calculation

- L : $TS(L)$
- R : $TS(R)$
- D : $TS(D)$
- S : $TS(S)$
- C : $TS(C)$
- W : $TS(W)$
- Sa : $TS(Sa)$
- U : $TS(U)$

For each variable, the (FA) gives an evaluation in an interval-form [1-10]. This evaluation (denoted by F_{TS}) depends on the distance between the value of the variable and its TS .

The FA modelling system is going to calculate the evaluation function $F_{TS}()$. $F_{TS}()$ can be calculated as follow:

$$F_{TS}(VARvalue) = \begin{cases} \frac{VARvalue}{TS(VAR)} \times 10; & \text{if } VARvalue \leq TS(VAR), \\ 10; & \text{if } VARvalue > TS(VAR). \end{cases}$$

For the variable U , we do not calculate its TS , since it will be passed into from Cor_u .

$$F_{TS}(U) = \frac{\sum_{j=1}^n Cor_u^j}{10n}. \tag{3.4}$$

With n is the number of users participating in that post and Cor_u^j is the corresponding Cor_u .

Each variable has different impacts. So, it is important to give weight for each variable related to its impact. We denoted the weight of variable i by w_i^p . Now, to calculate the score of

post we can calculate the value of GW_p , where GW_p represents the sum multiplication of FTS and weight, as given by (3.5).

$$GW_p = w_1^p \times F_{TS}(L) + w_2^p \times F_{TS}(R) + w_3^p \times F_{TS}(D) + w_4^p \times F_{TS}(S) + w_5^p \times F_{TS}(C) + w_6^p \times F_{TS}(W) + w_7^p \times F_{TS}(Sa) + w_8^p \times F_{TS}(U). \tag{3.5}$$

Now, we calculate the global score of the post denoted by GS_p :

$$GS_p = \left(\frac{GW_p}{\sum_{i=1}^8 w_i^p} \right) \times 10. \tag{3.6}$$

Table 3. Global post score

Posts description	Global score GS_p
Post is not important	[1-15]
Post can make little views	[16-25]
Post has impact for some category of people	[26-35]
Post is important	[36-45]
Post is very important	[46-60]
High popularity of post	[61-70]
Post can make propaganda and buzz	[71-80]
Post becomesa phenomenon	[81-100]

Example. If $L = 100$ and $TS(NF) = 120$ so the FA calculate distance between NF and $TS(NF)$ and give score in [1-10]. For this example, FA, will give value of 8. Thus, the evaluation of variable L was done. Indeed, $F_{TS}(L) = 8$. The FA evaluate all remaining variables as the same evaluation of L .

So, let the flowing evaluation for remaining variables:

$$F_{TS}(R) = 6, F_{TS}(D) = 3, F_{TS}(S) = 2, F_{TS}(C) = 8, F_{TS}(W) = 9, F_{TS}(Sa) = 4, F_{TS}(U) = 5.$$

Let the weight values:

$$w_1^p = 15, w_2^p = 10, w_3^p = 45, w_4^p = 23, w_5^p = 65, w_6^p = 12, w_7^p = 8, w_8^p = 9, \tag{3.7}$$

$$GS_p = \left(\frac{120 + 60 + 135 + 46 + 520 + 108 + 32 + 45}{187} \right) \times 10 = 57.$$

The post description based in Table 3 is: “Post is very important”.

3.4 Posting process

3.4.1 Process overview

This section elaborates the proposed process for ranking user and weighting the given post. The whole process consists of several components linked together to accomplish the needed ranking and to calculate the overall weight of the global post score. The components of this process are listed below:

- Post

- User
- Variables weight (Post)
- Variables weight (User)
- Cor_u
- GS_p
- Range correspondence
- Post Rank

4. Results and Discussions

The proposed algorithm for post ranking stands on eight variables that once calculated can approach post ranking. The included indices for evaluation as listed consequently, number of likes (L), rank given by user (R), number of dislikes (D), number of shares (S), number of comments (C), average number of words in all comments (W), satisfaction of comment (Sa), and by user's weight (U). The first seven variables shall be calculated from the post, while the eighth variable (U) is calculated from user's weight. Aiming to approaching post ranking can help in marketing, politics, sciences, media and many more fields that may would like to investigate the effect of their post on audience opinion.

Number of likes and dislikes without doubt could give a fast evaluation of the given post. Still, number of comments and number of words per comment have something to add. The main important variable under study which is hypothesized in this work to deeply affect the upvotes or downvotes of the given post is user's weight. We meant by user's weight the weight that user could expose on other users which in turn may affect their opinion on that post. However, having eight variables and considering that each variable may contribute a certain value towards the rapprochement of post ranking in Facebook, surely has a big novelty.

The FA model is proposed to be utilized by Facebook to calculate user weight ranking based on different proposed variables introduced early. The results coming out of FA calculation process will be used to calculate user weight and variables thresholds. The previous calculations will be part of GW_u , Cor_u and GS_p calculations.

The algorithm starts by calculating user rank (weight) through the calculating of Cor index. And then the total Cor variable for all users participating in the given post, regardless it was (like, dislike, share, or comment), will be calculated via U variable. The U variable will be then passed into post variables to be part of post score calculation. The seven indices (L , R , D , S , C , W and Sa) are used to calculate the global post weight (GW_p). This is very important since the weight of the given post can be used to judge the importance of its contents and to judge its presentation.

Finally, the global score of the given post (GS_p) can be calculated by letting the U variable be passed to join the seven indices used previously to calculate the GW_p . The GS_p then can be interpreted per the description introduced in Table 3. In addition, (3.7) introduces the weight of a given comment. This weight was used to decide whether that user will influence the whole discussed-topic or not, and if it will influence that topic, does it affect the rest of commenters or not.

4.1 Experimental Results

This section examines the experimental results of this proposed algorithm. The collected data for user-variables are demonstrated into Table 4.

Table 4. User-variables values

	#	Variable Values									
		NF	NC	NP	NS	NL	NI	NMS	NMR	NE	NT
Users	1	1000	12	2	45	2	12	456	852	3	10800
	2	200	10	1	120	1	5	123	185	0	800
	3	500	25	3	130	5	6	56	65	0	1200
	4	1200	23	4	4	13	13	256	123	1	7200
	5	3500	412	5	510	128	560	895	1200	56	12896
	6	120	3	6	1	12	1	19	22	0	568
	7	100	2	3	1	56	0	32	30	0	452
	8	1800	89	8	1	120	120	96	23	2	9256
	9	1123	23	9	23	9	52	415	235	1	7589
	10	1936	20	12	46	61	28	632	561	1	11258

The data are collected from 10 different active users (friends), who agreed to participate in this study. Table 5 shows the weights and thresholds for each user-variable. Weights and thresholds values are given by FA analyzer as illustrated in Figure 2.

Table 5. User’s variables weights and thresholds

	NF	NC	NP	NS	NL	NI	NMS	NMR	NE	NT
Weights w_i^u	20	10	50	5	7	70	26	17	11	23
Threshold $T()$	1509	520	4	412	26	620	1254	2154	52	16200

As FA is responsible for calculating the threshold for each variable, the function $F_T()$ which represents user evaluation can be calculated as highlighted by eq. (3.0). Table 6 examines the $F_T()$ results for each user-variable. For example, user number 5 has an evaluation value of 10 for the variable NF, this means that particular user has the best evaluation for this variable (number of friends). In the other hand, user number 7 has an evaluation value of 0.7 for the same variable (NF) which means this user has very less number of friends compare to user 5.

Table 6. Evaluation function $F_T()$ for each variable

	#	$F_T()$									
		Ft(NF)	Ft(NC)	Ft(NP)	Ft(NS)	Ft(NL)	Ft(NI)	Ft(NMS)	Ft(NMR)	Ft(NE)	Ft(NT)
Users	1	6,6	0,2	5,0	1,1	0,8	0,2	3,6	4,0	0,6	6,7
	2	1,3	0,2	2,5	2,9	0,4	0,1	1,0	0,9	0,0	0,5
	3	3,3	0,5	7,5	3,2	1,9	0,1	0,4	0,3	0,0	0,7
	4	8,0	0,4	10,0	0,1	5,0	0,2	2,0	0,6	0,2	4,4
	5	10,0	7,9	10,0	10,0	10,0	9,0	7,1	5,6	10,0	8,0
	6	0,8	0,1	10,0	0,0	4,6	0,0	0,2	0,1	0,0	0,4
	7	0,7	0,0	7,5	0,0	10,0	0,0	0,3	0,1	0,0	0,3
	8	10,0	1,7	10,0	0,0	10,0	1,9	0,8	0,1	0,4	5,7
	9	7,4	0,4	10,0	0,6	3,5	0,8	3,3	1,1	0,2	4,7
	10	10,0	0,4	10,0	1,1	10,0	0,5	5,0	2,6	0,2	6,9

Now, as $F_T()$ and w_i^u are calculated using the help of FA, the global user-weight GW_u and global comment's owner rank Cor_u impacts can be calculated using (3.1) and (3.2). Table 7 describes the values of GW_u and Cor_u variables. Looking to user 5 again in Table 7, we notice that the user has a Cor_u value of 88. This value can be interpreted as based on the classification rules in Table 2 which demonstrates that "This user can be famous".

Table 7. Global user-weight and global comment's owner rank impact

	#	GW_u	Cor_u
Users	1	730,7	30,6
	2	227,8	9,5
	3	515,9	21,6
	4	880,8	36,9
	5	2104,9	88,1
	6	563,8	23,6
	7	474,2	19,8
	8	1080,1	45,2
	9	953,4	39,9
	10	1148,3	48,0

5. Conclusions

The proposed algorithm utilized 10 variables for ranking comment's owner (user rating). This work investigates the effects of a comment, in an individual post, voted by a reputed person. Then the model analyzes how such a vote affects the rank of that post by increasing the upvotes or by increasing the downvotes. Eight variables are proposed evaluating the rank of the post. Being simple and easy to implement, the proposed method is expected to measure the post-sensitive influence on participants on that given post. However, introducing user's weight (ranking) as a new parameter for the evaluation of post's weight, could highly correct the whole evaluation of post's ranking. As commenters vary in their weights (rankings), posts can be upvoted or downvoted because of commenter's opinion and thought on the given post. This work is novel and aimed at introducing a new method for post ranking that can be utilized for different purposes in different disciplinarians. The experimental results are shown in the results and discussions section. However, based on this proposed model, Facebook can utilize new FA analyzer to rank their users and to study the impact of their own users on other users in different topics of interest.

Acknowledgements

The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under project number No. 1440-68.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

- [1] L. Backstrom, D. Huttenlocher, J. Kleinberg and X. Lan, Group formation in large social networks: membership, growth, and evolution, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 44 – 54, DOI: 10.1145/1150402.1150412.
- [2] A. Błachnio, A. Przepiórka and P. Rudnicka, Psychological determinants of using Facebook: A research review, *International Journal of Human-Computer Interaction* **29** (2013), 775 – 787, DOI: 10.1080/10447318.2013.780868.
- [3] F. Erlandsson, P. Bródka, A. Borg and H. Johnson, Finding influential users in social media using association rule learning, *Entropy* **18** (2016), 164, DOI: 10.3390/e18050164.
- [4] A. Hotho, R. Jäschke, C. Schmitz and G. Stumme, Information retrieval in folksonomies: Search and ranking, in *European Semantic Web conference*, 2006, pp. 411 – 426, DOI: 10.1007/11762256_31.
- [5] K.-C. Hu, M. Lu, F.-Y. Huang and W. Jen, Click “like” on Facebook: The effect of customer-to-customer interaction on customer voluntary performance for social networking sites, *International Journal of Human-Computer Interaction* **33** (2017), 135 – 142, DOI: 10.1080/10447318.2016.1221203.
- [6] S. Jamali and H. Rangwala, Digging digg: Comment mining, popularity prediction, and social network analysis, in *Web Information Systems and Mining*, 2009 (WISM 2009), International Conference on, 2009, pp. 32 – 38, DOI: 10.1109/WISM.2009.15.
- [7] M. Jemmali, M. Alharbi and L.K.B. Melhim, Intelligent Decision-making algorithm for supplier evaluation based on multi-criteria preferences, in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, 2018, pp. 1 – 5, DOI: 10.1109/CAIS.2018.8441992.
- [8] D. Kempe, J. Kleinberg and É. Tardos, Maximizing the spread of influence through a social network, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137 – 146, DOI: 10.1145/956750.956769.
- [9] L.K.B. Melhim, M. Jemmali and M. Alharbi, Intelligent real-time intervention system applied in smart city, in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 2018, pp. 1 – 5, DOI: 10.1109/NCG.2018.8593047.
- [10] B. Pang and L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval* **2** (2008), 1 – 135, DOI: 10.1561/1500000011.
- [11] X. Tang and C.C. Yang, Identifying influential users in an online healthcare social network, in *Intelligence and Security Informatics (ISI)*, 2010 IEEE International Conference on, 2010, pp. 43 – 48, DOI: 10.1109/ISI.2010.5484779.
- [12] T.W. Valente, Network models of the diffusion of innovations, *Computational & Mathematical Organization Theory* **2** (1996), 163 – 164, DOI: 10.1007/BF00240425.
- [13] J. Weng, E.-P. Lim, J. Jiang and Q. He, Twiterrank: finding topic-sensitive influential twitterers, in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 2010, pp. 261 – 270, DOI: 10.1145/1718487.1718520.