Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 189-202
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.5024

Measuring the community value in online social networks

Reza Mohamaddoust[a], Javad Mohammadzadeh [a*], Majid Khalilian[a], Alireza Nikravanshalmani[a]

[a] *Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran.*

(Communicated by Dr. Ehsan Kozegar)

Abstract

Communities in social networks form with different purposes and play a significant role in interpersonal interactions. Analysis of virtual communities indicates a more precise understanding of the behaviours and desires of individuals in social networks. In this paper, new measures have been proposed for analyzing implicit and explicit communities in Online Social Networks (OSNs). The measures of “potential value of the community members” and “value of the community messages”, which are used for calculating the measure of “community value” are among the most important measures introduced in this paper. Another measure introduced is “user influence rate” in a community, which represents the contribution of a person in creating value in a community. To provide a sound dataset, we collected the information from several real implicit communities in Twitter based on different hashtags. Finally, the suggested measures have been analyzed and compared statistically and behaviourally across different communities. The results of this research well indicate the importance and practicality of the measures introduced in Community analysis of Twitter.

Keywords: Community Value, data analysis, social computing, implicit community, Twitter

1. Introduction

Use of Online Social Networks (OSNs) has grown extensively, and its influence on the politics, culture, and economy of counties is undeniable. Many phenomena and concepts in psychology, social science and marketing are also observed in the interactions among individuals in OSNs. Social capital [13], opinion leadership [5, 1], information diffusion [20, 14] , influence [21, 9, 19, 3] and homophily [18], the formation and detection of communities [28, 10, 11, 8], Customer lifetime value (CLV) [26], etc. are among the concepts which can be observed and investigated in OSNs.

*Corresponding Author: Javad Mohammadzadeh

Email address: r.mohamaddoust@pnu.ac.ir, j.mohammadzadeh@kiau.ac.ir, Khalilian@kiau.ac.ir, Nikravan@kiau.ac.ir (Reza Mohamaddoust[a], Javad Mohammadzadeh [a*], Majid Khalilian[a], Alireza Nikravanshalmani[a])

Received: June 2020 *Accepted:* January 2021

Extensive research has been performed to analyze the communities and their associated issues in social networks [7, 25, 17, 6, 23, 16].

In this paper, by introducing new measures, the communities in Twitter OSN have been analyzed (through its method and results can be generalized to other OSNs as well). Regarding the analysis of communities, this paper covers the following:

- The possibility of analyzing communities whose dataset information is not sufficient for graph extraction (which are reasons for issues such as insufficient information required for graph representation or huge cost and complexity for graph extraction from the network
- Introducing new measures for valuation on different communities in OSNs
- Statistical and behavioural analysis of the data collected from implicit communities in which the relations among the members are not transparent

The methods proposed in this research can be employed for different uses including detection of influential messages and influential users, comparing communities, helping in the improvement of the accuracy of recommender systems, and effective sampling of OSNs.

A real community is a group of individuals with common economic, social, or political features or interests who mostly live in relatively proximate places. A virtual community develops when like-minded users join together in social media and initiate interaction with each other. In other words, the creation of any community needs (I) a set of at least two members with common desires and interests and (II) interactions associated with those desires [27].

Communities form either implicitly or explicitly. An explicit community meets the three following conditions and criteria:

- The community members know that they are its members.
- Nonmembers know who the members of this community are.
- The community members have often more interactions with each other rather than with non-members.

In contrast to explicit communities, in implicit communities the individuals relate to others as an unknown community implicitly. For example, individuals who contact Germany from Iran are not necessarily friend with each other and do not consider each other as the members of an explicit community. However, in the telephone operator's view, they create an implicit community, for whom similar advertisements should be considered for marketing purposes. In OSNs analysis can consider the individuals who compose messages about the same or similar subjects as the members of an implicit community. Finding implicit communities is of great interest [27]. However, generally, community detection is an ill-defined problem [15].

If the different features of a community which consist of individuals and messages could be extracted as quantitative values, then one could better compare different communities. In this paper, to achieve this quantity, a measure called "community value" has been proposed, which can be used across various applications including the comparison of communities, community evolution, understanding the community behaviour, and so on. For example, in digital marketing, different explicit and implicit communities are formed for brands in social networks. The questions are: 'What is the effect of different marketing activities such as advertisements, variations in the sale methods, and other issues?' 'How companies can enhance the value of their communities to keep the interests of their brand, and influence their customers (potential and actual) more deeply and extensively ?' Accordingly, the existence of an efficient and accurate measure for measuring the value of communities can have extensive uses. Not only does propose this paper a measure for community evaluation, it also measures the contribution of each message and person in the value obtained for the community. In this paper, to examine the introduced measures, first two datasets have been presented for the implicit community in Twitter social network. Finally, to analyze more communities with five datasets, the measures are then compared with each other.

An important issue for analyzing social networks is the different limitations for creating a suitable dataset. It specifically means a proper dataset with various metrics capable of measuring extensive and comprehensive measures. In this paper, a measure has been introduced for valuation of communities, which uses different statistical metrics of Twitter. This measure has various applications such as finding influential tweets, identifying influential users, finding opinion leaders, comparing communities with each other, creating effective recommender systems, sampling [4] the communities, etc. are among the important uses of the proposed measures. In this paper, Twitter's statistical metrics have been used including number of likes and retweet for each message, the number of tweets, number of followers, number of likes, number of replies, and number of retweets for each user.

In most papers such as the one by Riquelme and Cantergiani [21] and many other cases researched by them, the community structure is presented by graph [22]. To draw the graph, certain information is required. For example, for graphs type $G1$, the follower's information is used to draw the edge between the nodes [21]. In $G2$ graphs, it should be exactly specified which aspect or aspects are of interest to the researcher. For example, the relationships between liking or retweeting messages are used to draw the edge between the graph vertices. In practice, the more extensive this information, the more complete the graph can be. However, due to the limitations that Twitter applies, obtaining this information is sometimes very difficult or even impossible. For example, to know exactly who have liked a certain message, there is no complete answer. Alternatively, finding all users that have retweeted the message of a certain user is very difficult. Accordingly, obtaining a graph for a community to analyze it is sometimes impossible or very costly. In this paper, the community is analyzed without drawing its graph. The measures proposed in this paper have used the statistical information of the messages and Twitter account of the users instead of precise information among the users. The aim is to calculate important measures such as the value of a community and the influence rate of each user in the community.

Riquelme and Cantergiani [21] made a distinction between metric and measure. According to them, metric is a simple mathematical statement which helps in providing essential information about the social network as a numerical value. Accordingly, one can combine metrics to define a (ranking) measure. In other words, a formula or an algorithm can be defined which presents a criterion for ranking each user in a network. Definitely, there are also more complicated measures than mere combination of metrics.

Riquelme and Cantergiani [21] introduced a measure for measuring the user activity, which deals with four metrics including the number of tweets sent by a user, the number of retweets others have had for a user, the number of likes of a user's messages, and the number of a user's replies. However, the like of tweets has not been taken into account in the dataset they have used for measuring these measures.

Domenico et al.[12] collected a dataset involving dissemination of scientific information in Twitter, before, during, and after the declaration of the discovery of a novel particle with the features of the elusive Higgs boson on 4th July 2012. This dataset includes mentions, retweets, and reply of users to each other. However, the like or statistical values including the number of each user's followers or the number of likes of each message are not specified in it. Its information is suitable for drawing a graph, but it lacks other points required in the measures proposed in this paper.

2. Methods

A community on Twitter can consist of individuals and messages related to a specific period of time. For each individual in Twitter, there is certain statistical information (metrics) on their account page, which can be observed by anyone (provided that the user page is public). This statistical information includes the number of messages liked by the user, the number of his or her statuses (including the tweets they have written, the number of replies and retweets), the number of followers, and the number of their friends. Meanwhile, there is also statistical information for each message including the number of likes, number of retweets, and number of replies to that message. From now on, we call this statistical information (metrics) as the features of individuals and messages, respectively.

The reply to a tweet can be considered a message. The replies themselves can have other new replies, where these new replies are not necessarily a reply to the initial tweet. Thus, we do not consider the number of replies as message features. Accordingly, to enhance the accuracy and to elucidate the issue, the number of likes and number of retweets are considered as message features. With regards to the individual features, all metrics except for the number of friends (followings) are taken into account (since the number of followings has no effect on the proposed measures). That metrics and notations used in this paper are provided in Table 1.

Table 1: List of metrics and proposed measures

Notation	Description
CLV	The community Life Value or simply community value
CMV	Community Messages Value
CUV	Community Users Value
LM_i	The number of a message likes
RTM_i	The number of a message retweets
UIR	User Influence Rate
MV_i^j	The i_{th} message whose writer is the jth individual in the community
UMV_j	The value of messages written by the jth individual
UPV_i	The i_{th} user potential value
UIR_i	The i_{th} user influence rate
M	The number of messages in a community
N	The number of unique users in a community

The community value and each individual value in the community are new measures, which to the best of our knowledge, have been introduced for the first time in this paper. The community value refers to the total value of the members of a community along with the value of messages in that community. The value of individuals is equal to the sum of the values of the features of the community individuals (metrics mentioned above). Similarly, the value of the community messages is equal to the sum of the values of the features of that community messages.

As shown in Table 1, CMV is the value of the community messages and CUV represents the value of the community users, which are obtained by Equations as follows:

$$CMV = \sum_{j=1}^{m} (RTM_i + LM_i) \quad (1)$$

$$CUV = \sum_{j=1}^{n} (RTU_i + OTU_i + LU_i + FU_i) \quad (2)$$

The number of messages in a community is equal to m (the number of investigated messages which can be related to a certain time range or a specific number of messages), while the number of unique users in the community is equal to N. Hence, via dividing CMV by the number of messages in the community, the average value of the message in the community ($\overline{CMV}$) is obtained. Furthermore, through dividing CUV by the total number of unique users in the community, the average value of each individual in the community ($\overline{CUV}$) is also obtained. The value of each message is equal to the following equation:

$$MV_i = RTM_i + LM_i \quad (3)$$

Hence, the value of the messages created by a special user is as follows:

$$UMV_j = \sum_{j=1}^{m} MV_i^j \quad (4)$$

The value of any individual for the communities he or she is a member of can be potentially valuable. This value has considered all activities of the user on Twitter up to this moment. Accordingly, the value of each user can be regarded as their potential value for that community, so that in the future with the actions of the same individual (such as creating a message or performing different activities), it directly causes the creation of actual value for the community. In this way, these activities can influence CMV value (its effect becomes actual). Furthermore, the presence of more individuals with potential value helps their followers to have more attention and motivation to be present in these communities. The potential value of each user is calculated as follows:

$$UPV_i = RTU_i + OTU_i + LU_i + FU_i \quad (5)$$

As each message is created by only one individual, thus the magnitude of influence or contribution of an individual in the value of the community with the number of messages they have created in the community (UMV_i) along with the potential value it has for the community is defined as follows.

$$UIR_i = \alpha \times \frac{UMV_i}{CMV} + \beta \times \frac{UPV_i}{CUV}, \alpha + \beta = 1 \quad (6)$$

UIR_i value can be zero which is minimum and one which is maximum. The measure proposed for the value of a community is also calculated as follows:

$$CLV = \alpha \times m \times \log_2 CMV + \beta \times n \times \log_2 CUV, \alpha + \beta = 1 \quad (7)$$

As CUV and CMV values are mostly large values, by calculating the algorithm, they change into lower values. α and β can have different values. Since in a community, CMV suggests the extent of reactions to the messages in the community, thus it can be more important than the potential of individuals (which is adapted from their activities and the number of their followers in the past and even in other communities). Indeed, when an individual changes his or her potential value to actual value by creating a message, liking, or retweeting in the community of interest, this means that CMV value increases. Accordingly, it is suggested that value be considered larger than β value. Based on experience, $\alpha = 0.80$ and $\beta = 0.20$ are considered logical values for calculating the CLV of most of the investigated communities.

Since the value of a community can be a large value (such as the value of many real-world goods such as the house, jewelry, car, etc.), determination of the upper bound for this value is difficult. Thus, it seems that there is no need to normalize CLV. With this in mind, the large distance of the CLV of different communities suggests their large value distance (as with the real world).

3. Dataset

As was observed, the metrics required for calculation of the introduced measures are various. As far as we know, no dataset covering all of these metrics has been created so far. Accordingly, the stages of creating suitable datasets are explained as follows.

Twitter is an OSN and a microbloging service provider, which allows users to send text messages known as tweets up to 280 characters to each other. Any user in Twitter can perform actions upon the messages. These actions include composing the message (tweet), retweeting the message of others (demonstrating the tweet of others in the user's personal page so that his or her followers can see that tweet on their personal page), liking which represents expression of liking the message of others, and replying which involves replying to the others' messages.

To collect the dataset information, API REST [24] of Twitter was used. The collected tweets were related to the event of the football match between Barcelona and Real Madrid teams on 23 December 2017. The time range for data collection included seven days, three days of which were before the match, while three other days belonged to the post-match period. The number of messages and the number of unique individuals in both communities (Barcelona community and Real Madrid community) along with the searched hashtags are indicated in Table 2. The existence of hashtag in messages is crucial, as hashtags in Twitter are used for classifying messages, developing ideas, and promoting special issues or individuals. The individuals who promote a special hashtag in their messages are indeed considered the members of an implicit community. The collected tweets were all in English. All of the community messages were created by the public users of Twitter. All of the public messages of Twitter which had the hashtags mentioned in Table 2 were collected completely within the stated time period (as all of the messages of both communities were less than 18000 tweets, where '18000 tweets' is one of the limitations applied for API REST of Twitter. This limitation gives the permission of 180 requests in every 15 minutes, where at most 100 tweets in each request can be received).

Dataset extraction stages involve the following:

1. collecting information through search using $APIREST$ of Twitter and based on the intended hashtags

2. converting and storing the received information from $JSON$ to CSV format (comma-separated values)

3. removing the extra information and data preprocessing to be incorporated into the dataset

4. De-identification of the received information (without altering the statistical nature of the information)

Table 2: The hashtags searched to collect the tweets of the two implicit communities of Barcelona and Real Madrid

Community Name	The searched hashtags	Number of messages	The number of unique users
Barcelona (#FCBarcelona)	#FCBarcelona, #Messi	6734	4932
Real Madrid (#Realmadrid)	#Realmadrid, #Ronaldo	5328	3209

The explanations related to dataset variables (features) are indicated in Table 3. Each message and individual on Twitter has a unique identification number known as Tweet ID and User ID. The value of user statuses_count is equal to the total sum of the number of replies, tweets, and retweets of each individual on Twitter.

Table 3: Dataset variables description

Type of feature	Description of
Created_at	The date and time of sending the message as coordinated universal time (UTC)
Screen_name	The name with which the individual is known in Twitter
user id	The unique identification number of the individuals in Twitter
Tweet id	The unique identification number of the message in Twitter
Favorit_count	The number of likes received by a message
Retweet_count	The number of retweets of a message by the individuals
User follower_count	The number of followers of an individual who has written a message
User statuses_count	The total sum of the number of replies, tweets, and retweets of an individual
User like_count	The total number of likes of an individual in Twitter
Text	The message text

4. Results

In this section, the proposed measures are calculated using the values of the collected datasets. Thereafter, the communities are analyzed statistically and behaviourally based on the values of the different measures. Table 4 provides the values calculated for the metrics and measures for Barcelona and Real Madrid communities. The number of messages and unique users of Barcelona community are larger than those of Real Madrid (note that the data of both communities have been collected within the same time frame). It is one of the reasons behind the better values of the metrics and measures calculated in Table 4 in favor of Barcelona community (the final result of the match, the performance of the teams in the standings and other points can also be influential). Except for the average statuses count of the individuals, in the rest of the cases, Real Madrid community has lower values compared to Barcelona community.

Table 4: The values of metrics and measures of Barcelona and Real Madrid communities

Metric/measure	Barcelona community	Real Madrid community
The total number of tweets	6,734	5,328
The number of unique users	4,932	3,209
statuses count	58,216,421	48,908,465
Average statuses	11803.8	15241
Like Count	22,859,547	14,631,229
Average Like	4634.9	4559.4
Follower Count	68,039,203	32,703,494
Average Follower Count	13795.5	10191.2
Number of tweets' likes	122,873	61,692
Average number of tweets' likes	18.2	11.6
Number of tweets' retweets	35,775	17,555
Average number of tweets' retweets	5.3	3.3
CMV	158,648	79,247
Average CMV	23.6	14.9
CUV	149,115,171	96,243,188
AverageCUV	30234.2	29991.6
CLV ($\alpha = 0.80 and \beta = 0.20$)	119,849	86,387.24

CLV values have been shown with $\alpha = 0.80$ and $\beta = 0.20$ in Fig. 1 along with the values of the two main parts of its formula. As expected,CLV value is closer to CMV value based on the values of the utilized coefficients of $\alpha = 0.80$ and $\beta = 0.20$. As was emphasized previously, reaction to the messages created by the individuals in the community has an actual value, while CUV has a potential effect on the community value. If it is actualized, it directly affects CMV.

Fig. 2 compares three other measures related to 1% of individuals in both communities to analyze the behaviour of the data of the communities. UMV related to 1% of the community individuals have claimed 87 and 84% of the total UMV of Barcelona and Real Madrid communities, respectively. Furthermore, 1% of the community individuals has accounted for 48 and 38% of the total UMV of the individuals in Barcelona and Real Madrid communities, respectively. Regarding UIR, this value has been 77 and 75 % for Barcelona and Real Madrid communities, respectively. As can be seen, there is a close behavioural similarity in 1% of both communities. On the other hand, it can be stated that this similarity also exists in the other 99% of the community individuals. Thus, in behaviour analysis of these three measures, it can be stated that both communities have indicated similar behaviours.

The maximum UIR of Barcelona and Real Madrid communities is 49% and 54%, respectively, which belongs to the official user account of both clubs on Twitter. Furthermore, the first three Twitter user accounts of Barcelona in terms of UIR belong to Barcelona club itself, though in different languages. Thus, Barcelona has also a UIR of 53% if they are included.

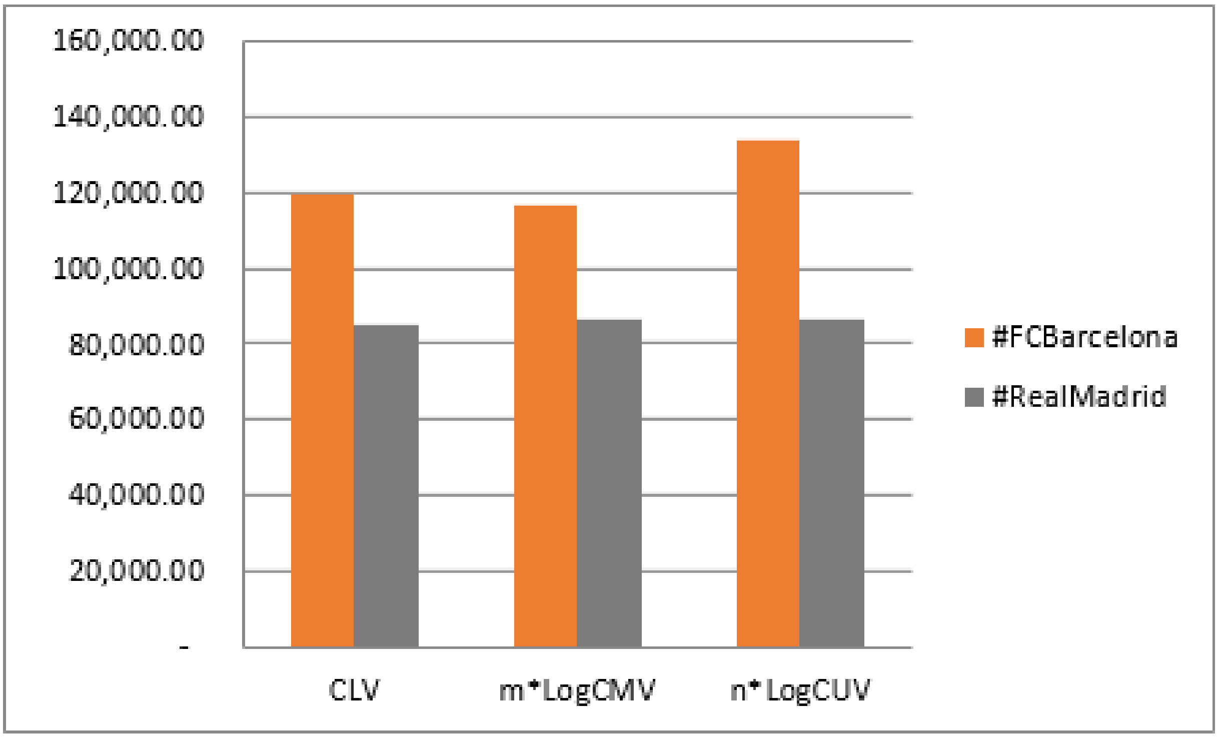

Figure 1: Comparing the community value of Barcelona and Real Madrid.

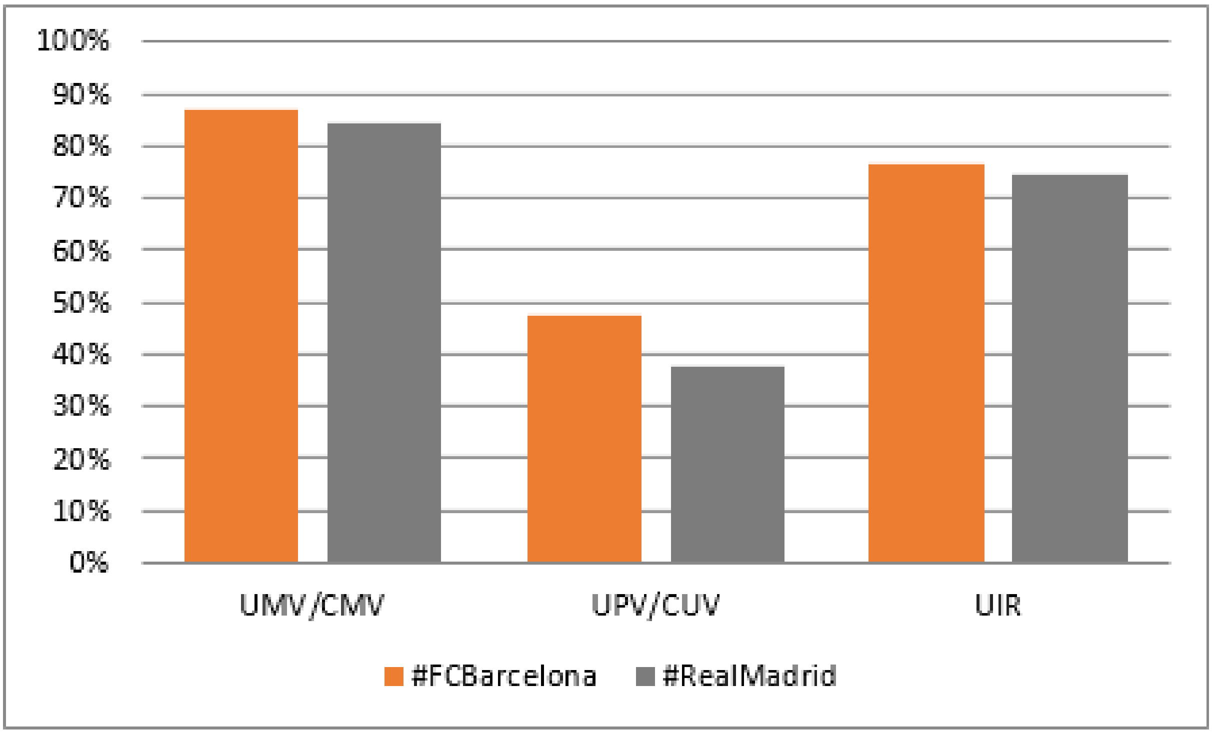

Figure 2: The contribution percentage related to the top 1% of each community across the three measures

The value of 1% of the top messages of Barcelona and Real Madrid messages to the total value of the messages of each community is 82% and 83%, respectively. For Barcelona and Real Madrid communities, 1% of tweets includes 67 and 53 messages, respectively. The maximum MV values in

Barcelona and Real Madrid alone account for 14 and 11% of CMV measure, respectively. Again, behavioural similarity can be observed in both communities.

Figs. 3 and 4 illustrate UMV values for the top 100 individuals except for the first rank individual (due to his or her large distance). Also, Figs. 5 and 6 display UIR values for the top 100 individuals except for the first rank (due to his or her large distance) for Barcelona and Real Madrid communities, respectively. As UIR and UMV value of the rest of the users is close to zero, they have not been demonstrated.

One of the well-known models in the area of complex networks which also captures social networks is Scale-Free network model[?]. Networks with a power_law degree distribution are called Scale_Free networks. Among the important features of this type of networks is the existence of a large number of data with low values (low_degree nodes) and a very trivial number of data with large values (high_degree nodes).

Distribution of exponential law is evident in the diagrams of Figs. 3, 4, 5 and 6. For UIR measure, the nodes in the graph can be considered to be equivalent to the community individuals. For UMV, the nodes can be considered as messages. Similarly, the node degree (the number of edges connected to a node) can be considered as UIR value for each individual (node), while UMV can also be assigned to each individual. Therefore, it can be stated that the studied communities follow Scale-free network model. As mentioned previously, in Scale-Free networks, the number of data with low values (nodes with a low degree) is very high, while a very trivial number of the data have large values (nodes with a high degree), as with what can be observed in the behaviour of the values of the mentioned measures.

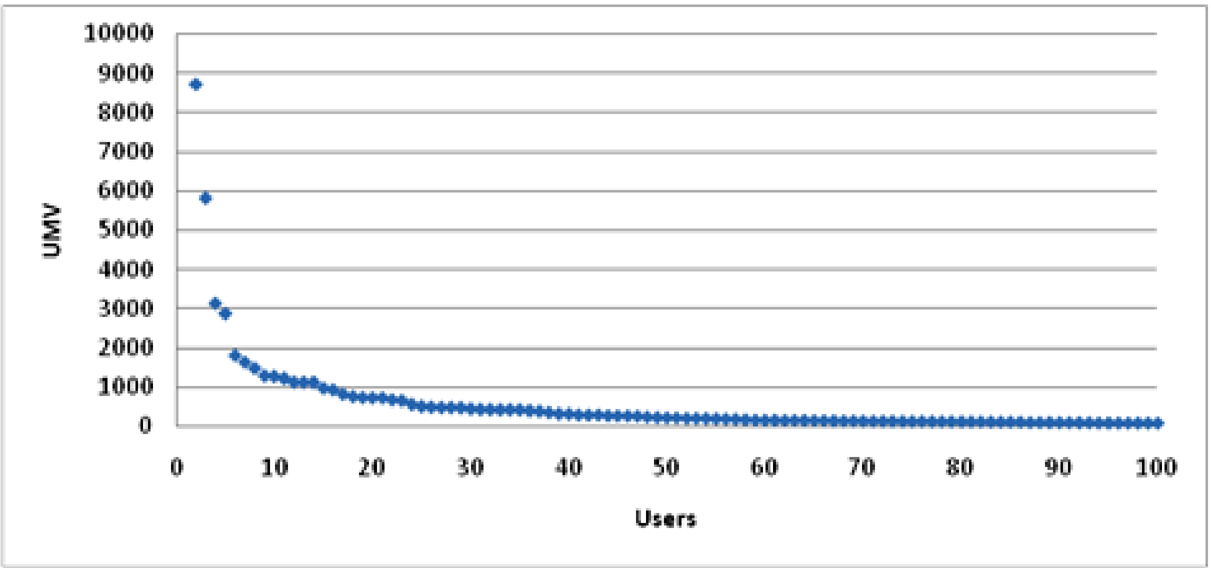

Figure 3: UMV value of the first 100 individuals in Barcelona community (except for the first rank).

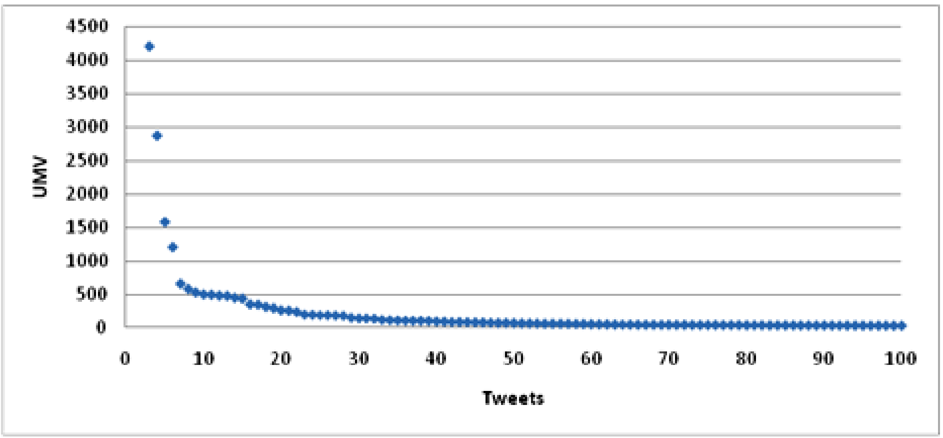

Figure 4: UMV value of the first 100 individuals in Real Madrid community (except for the first rank).

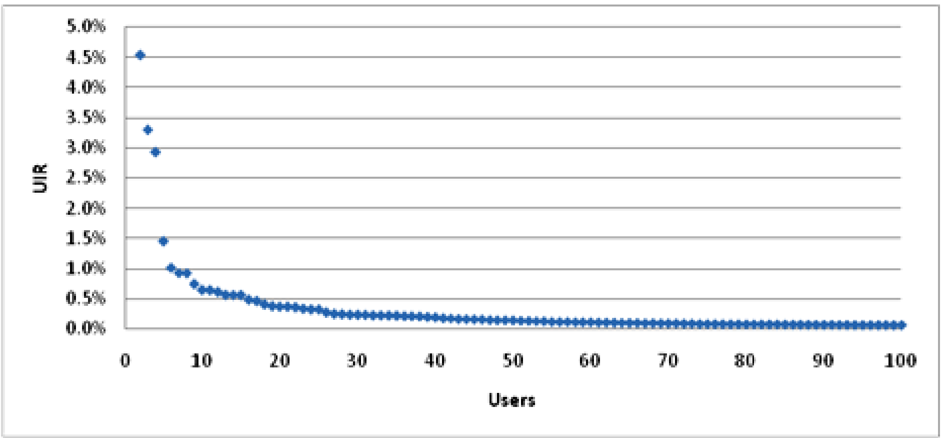

Figure 5: The contribution percentage of the first 100 individuals in terms of UIR in Barcelona community (except for the first rank).

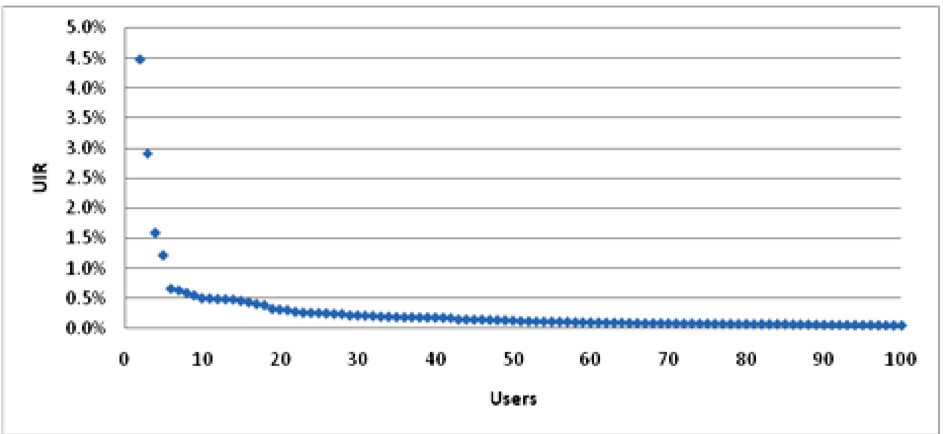

Figure 6: The contribution percentage of the first 100 individuals in terms of UIR in Real Madrid community (except for the first rank).

Various behavioural similarities were observed in the behaviour analysis of the data of both communities, while in the real world, these two teams are complete competitors. As was observed in the diagrams, in distribution of CMV value, similar behaviours are governing across the individuals and messages as well as the distribution of UIR value in both communities. In distribution of CUV values, this similarity slightly declines across the community individuals. However, since eventually CUV involves β coefficient (which practically is very lower than α) in calculation of CLV and UIR, this distance is not very sensible.

Note that Barcelona won the game against Real Madrid in this match, which affects the values of metrics and measures calculated for Barcelona community positively, as mentioned in Table 4. Nevertheless, this case does not significantly affect the general behaviour of the data of the communities either.

As the value of the messages of a community significantly affects the value of the community (CMV is more influential than CUV), thus investigation of CMV is crucial for most communities. On the other hand, to better generalize the obtained results, other communities can also be examined. Accordingly, three datasets were collected for three other communities, and based on distribution ofCMV values, we compared the five datasets with each other (two previous communities and three new communities). These hashtags included #WeAreTheArsenal introduced by the official page of Arsenal football team, #FCBayern introduced by the official page of Bayern Munich club's twitter page, and #Diet. As mentioned previously, the dataset of the intended communities was also created from 27 January 2018 until 2 February (seven days). #diet hashtag was also chosen as an irrelevant hashtag compared to the others so that its behaviour would be compared with the behaviour of club teams in terms of CMV measure. Fig. 7 demonstrates the MV share related to 10% of the top messages in all of the five communities. diet with 73% has the lowest value. Arsenal and Bayern Munich with 77 and 81% MV have a similar behaviour. Overall, in all of the five communities, the major part of CMV value belongs only to 10% of the messages of that community.

Fig. 8 demonstrates the percentage of messages whose MV is zero (they have no like and retweet) across all of the five communities. The percentages of messages from the communities of Bayern Munich, Arsenal, Barcelona, Real Madrid and #Diet, whose MV value is zero are 30, 46, 48, 56, and 56%, respectively. As can be observed, except for Bayern Munich, almost half of the messages of the other communities have not received any like or retweet. Nevertheless, this number does not affect CMV value, yet when calculating CLV, the number of messages even if MV is zero has been taken into account.

As predicted, all of the five communities follow the power_law distribution model in the distribution of CMV values, with the above_mentioned behavioural similarities also confirming this point.

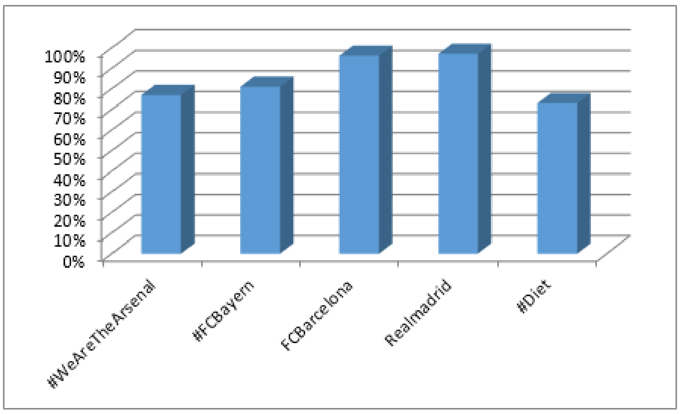

Figure 7: The percentage of contribution of 10% of the messages with the maximumMV in each community.

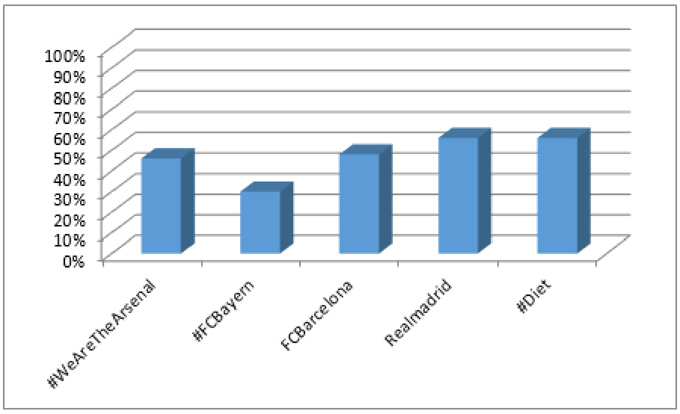

Figure 8: The percentage of the messages of the five communities, whose number of likes and number of retweets is zero.

5. Conclusion and future work

In this paper, an important measure was introduced for valuation of communities in Twitter. This measure called community value was obtained based on measuring the value of the current messages (actual value) and individuals (emphasizing their potential value) of a community. On the other hand, a dataset was collected based on the hashtags in Twitter messages throughout special stages. These datasets were used as implicit communities in this research. Among the advantages of these datasets has been extensive use of different metrics in Twitter such as the number of likes and retweets for each message as well as the number of tweets, replies, retweets, followers, and number of likes of each individual. In the next stage, the presented measures were calculated using the values of the collected datasets. In the end, statistical and behavioural analysis of the measures introduced across different communities was performed. Among the important results of behaviour analysis of the communities is that the distribution of UIR and UMV measured values follows power_law distribution across different datasets. Indeed, Scale_free model can be observed based on the distribution of the values of these measures. This model causes the results obtained to both apply to the dataset of interest and be generalized to all similar communities in social networks. The behavioural similarity of the five studied communities while being different with each other in terms of their values of initial metrics is one of the research results. In this paper, to analyze the implicit communities, instead of forming a graph and then analyzing it (which has its own special complexities), one can use the presented methods (based on statistical analysis of metrics and then behavioural analysis of measures), which have various usages.

For future works, use of further metrics such as the number of reply to the messages, number of mentions, use of image processing methods for better analysis of multimedia messages, and predicting the behaviour of communities across different social networks are issues that require further research.

References

[1] A. Aleahmad, P. Karisani, M. Rahgozar and F. Oroumchian, OLFinder: Finding opinion leaders in online social networks. Journal of Information Science, (2015) 1-16.

[2] A.L. Barabasi, and E. Bonabeau, Scale-free netwroks. Scientific American. 288, (2003) 60-69 .

[3] P. Basaras, D. Katsaros and L. Tassiulas, Detecting influential spreaders in complex, dynamic networks. Computer 46 (4), (2013) 24–29.

[4] V. Bhatia and R. Rani, An efficient algorithm for sampling of a single large graph. Tenth International Conference on Contemporary Computing (IC3), (Noida, 2017) 1-6.

[5] F. Bodendorf and C. Kaiser, Detecting opinion leaders and trends in online social networks. Proceedings of the 2nd ACM workshop on Social web search and mining, (ACM, 2009) 65-68.
[6] O. Boutemine and M. Bouguessa,Mining Community Structures in Multidimensional Networks. Acm Transactions on Knowledge Discovery from Data. 11(4), (2017) 1–36 .
[7] R. Cazabet, P. Borgnat and P. Jensen, Enhancing Space-Aware Community Detection Using Degree Constrained Spatial Null Model.Workshop on Complex Networks Comple Net. Springer, (2017) 47–55.
[8] J. H. Chin and K. Ratnavelu, Detecting community structure by using a constrained label propagation algorithm. PLOS ONE 11, (2016) 1–21 .
[9] G. F. de Arruda,et al. Role of centrality for the identification of influential spreaders in complex networks. Physical Review E 90, 032812 (2014).
[10] P. DeMeo, E. Ferrara, G. Fiumara and A. Provetti, Mixing local and global information for community detection in large networks. J. Comput. Syst. Sci. 80, 72–87 (2014).
[11] J. Ding, X. He, J. Yuan, Y. Chen and B. Jiang, Community detection by propagating the label of center. Physica A: Statistical Mechanics and its Applications 503, (2018) 675–686 .
[12] M.D. Domenico, A. Lima, P. Mougel and M. Musolesi, The Anatomy of a Scientific Rumor. Scientific Reports 3, 2980 (2013).
[13] N.B. Ellison, C. Steinfield and C. Lampe, The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. Journal of Computer-Mediated Communication, 12 (4), (2007) 1143–1168.
[14] A. Guille, H. Hacid, C. Favre and D.A. Zighed, Information diffusion in online social networks: A survey. ACM SIGMOD Record. 42 (2), (2013) 17-28.
[15] D. Hric and S. Fortunato, Community detection in networks: A user guide. Physics Reports, (2016) 659, 1-44 .
[16] A. Khushnood, et al. Popularity and Novelty Dynamics in Evolving Networks. Scientific reports 8, 6332 (2018).
[17] C. W. Loe and H. J. Jensen, Comparison of communities detection algorithms for multiplex. Physica A Statistical Mechanics & Its Applications. 431, (2015) 29–45.
[18] M., McPherson, L. Smith-Lovin and J. M. Cook, Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27, (2001) 415–444.
[19] F. Morone and H. A. Makse, Influence maximization in complex networks through optimal percolation. Nature 524, (2015) 65–68 .
[20] D. M. Romero, B. Meeder and J. Kleinberg, Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In: Proc. 20th Int. Conf. on World Wide Web, WWW' 11, (ACM, New York, NY, USA, 2011) 695–704 .
[21] F. Riquelme and P.C. Cantergiani, Measuring user influence on Twitter: A survey. Information Processing and Management. Information Processing & Management. 52(5), (2016) 1-27 .
[22] L. Tang and H. Liu Graph Mining Applications to Social Network Analysis. Managing and Mining Graph Data. 40, (2010) 487-513.
[23] O. Tsur and A. Rappoport, What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. Proceedings of the fifth ACM international conference on Web search and data mining, (ACM, 2012) 643–652.
[24] Twitter, Inc. Twitter API Developer. http://dev.twitter.com/ (2017).
[25] W. Ulrich and N. J. Gotelli, Pattern detection in null model analysis. Oikos 122(1), (2013) 2–18.
[26] R. Venkatesan and V. A Kumar, customer lifetime value framework for customer selection and resource allocation strategy. Journal of Marketing. 68, (2004) 106-125.
[27] R. Zafarani, M.A. Abbasi and H. Liu, Social media mining, an introduction (Cambridge, 2014).
[28] F. D. Zarandi and M. K. Rafsanjani, Community detection in complex networks using structural similarity. Physica A: Statistical Mechanics and its Applications 503, (2018) 882–891 .

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.