A Graph Mining Method for Characterizing and Measuring User Engagement in Twitter

Ioannis Karamitsos*, Alaa Mohasseb† and Andreas Kanavos‡
*Department of Computing
Rochester Institute of Technology, Dubai, UAE
ixkcad1@rit.edu
†School of Computing
University of Portsmouth, Portsmouth, UK
alaa.mohasseb@port.ac.uk
‡Department of Informatics
Ionian University, Corfu, Greece
akanavos@ionio.gr

***Abstract*—In the modern world, social media plays a crucial role in the interchange of information and socialization with users. Twitter is a known social media platform that allows users to make relationship with others and express their opinions. The current work aims to identify the level of user engagement on Twitter with the use of graph mining. User engagement concerns the number of user connections with a tweet and can be measured using different tweet attributes including retweets, replies, etc. Specifically, t his s tudy i nvestigates t he v ariety o f e dges strength that user connections can implement in a Twitter networks. Next, we employed various weights in the graph mining models to evaluate the score of each connection. These tasks were followed by statistical analysis to measure the similarity between the two user profiles a s w ell a s a ttributes l ike f riendship, following and interaction in the Twitter social network. Results indicate that closely linked groups can be revealed and thus, a need for examining both group and individual behavior, will arise.**

***Index Terms*—Community Analysis, Graph Mining, Hemophilia, Social Media, Twitter, Twitter Analytics, User Engagement**

I. Introduction

Social networks are a platform that play an important role in creating social relationships between people with common interests or activities. This platform provides all kinds of connections between people and the degree of strength of these connections. Social networks constitute also an important component in society and are used as a measure of social connectivity to identify and evaluate the quality and quantity of information flow b etween p eople a nd w ithin g roups. To understand why people tend to form social networks and how these networks function, it is necessary to understand the connections between members [9].

In recent years, social networks have gained popularity and are easily accessible to people. As a result, user-created content and volume are exponentially increasing every day. With the introduction of graph mining, we can capture these huge amounts of data and analyze them for the desired purposes. Graph mining is a basic feature for studying the social network characteristics and user engagement between interactions [11]. The main goal of graph mining is to extract information or interaction patterns between social entities or between social groups [20].

Furthermore, a social network composed of social actors becomes active when relationships are formed in the course of regular interactions in everyday life and living, in cultural activities such as family celebrations, annual celebrations of various communities, engagements, etc. Among the many examples of regular interactions is when one family asks another for help, support, or advice, when a new friendship is formed or when individuals spend leisure time together, etc. Sometimes a relationship can be negative, alienation as opposed to reciprocity or integration, with even the security aspect being an important factor [18], [24].

Twitter is one of the most popular social networking sites and is used by almost everyone for updating news, sharing information, and marketing using 280 characters [1]. Twitter has many functions and features that dynamically change in its development. The main components and features of the Twitter platform that are used in this paper are the following:

- *Hashtag*, which are identifiers that start with the character "#" followed by a word or words not separated by spaces. They are used for searching posts on desired topics.
- *Mentions*, which are identified by the character "@". These allow a user to refer to another user in a tweet. The notification of a user A is done so that the mentioned user B sees the tweet and engages with it by giving a like or retweet or starts a discussion about the tweet by replying.
- *Retweet*, which makes possible the republish of a tweet of a user. The retweet also allows the user to add a personal comment and is also an expression of the strong interaction between the two users.
- *Reply*, where a user can reply by posting something additional or making a comment under a post. The replies begin with the character "@" followed by the screen name of the person writing the reply.

- *Follow*, where a user chooses who to "follow" on the social network and consequently who to connect with and build a social relationship with.
- *Friendship*, which is different from following as when a user follows a profile, it appears on his friends list, while he himself appears in the list of followers of that profile.

This paper aims to categorize user relationships in the social network according to their strength. The concept of strength refers to how closely two users are connected to the network, appreciating certain characteristics with users captured by Twitter, as well as the degree of interaction. Power characteristics include strong similar characteristics from their profiles on the social network [23]. The similarity characteristics take into account elements such as the common or very close location, the similar scale in the number of friends, the similar frequency of posts as well as the interaction criteria, such as the relationship of friendship and follow-up, the mentions and retweets of users, live tracking and the ability to exchange messages. Based on the above data which contribute to a different percentage, a score is extracted for each edge that expresses the connection of two members of the network and is therefore categorized according to this calculated score. This graph mining framework is applied into Twitter data at different interaction levels [2], [10].

The rest of the text is outlined as follows: In Section II, we provide the related work of the corresponding problem, while Section III discusses the proposed method, the metrics utilized as well as the details of the scores assigned to the edges of the social network. Section IV introduces the implementation aspects along with the dataset used for our experimental study, whereas Section V presents the research results. In Section VI, we summarize our contributions and draw future directions.

II. Related Work

Social networks are a point of attraction for the scientific community, as the need to analyze the social relationships and behavior of people in the network becomes more intense from different perspectives and for a variety of reasons. This section relates to some of these perspectives and collects data that contribute to the implementation of the idea of this study.

An efficient and fast technique for clustering graphs is presented in [4] that can handle graphs with a large number of nodes and a very large number of edges and is based on multilevel methods that use a weighted Kernel K-means objective function as a refinement algorithm. In addition, authors in [21] proposed a new method for mining subgraphs in graph-structured databases. The algorithm proposed is based on frequent hyper-clique patterns that try to identify the dependencies between graphs in a large database. The efficient pruning methods implemented in this study were based on depth-first and breadth-depth search methods.

Furthermore, a new method for clustering bipartite graphs, namely coring technique, was introduced in [16]. The proposed technique can solve the problem of partitioning a large graph into small subgraphs. The nodes of the clustered subgraphs are strongly connected to each other within the graph and weakly connected to the nodes of other graphs. The technique proposed by the authors can be used to compute clusters that have a very dense core region surrounded by regions of lower density.

A novel graph clustering mechanism, namely semi-supervised divisive (DIANA) hierarchical graph clustering algorithm was employed in [14]. This procedure can effectively solve the clustering problem without being aware of the structure of the underlying dataset. Concretely, the proposed algorithm increases the weight of the edge if two nodes are similar, otherwise the weight of the edge is decreased. Then, nodes with a small neighborhood are removed to form clusters, while on the other hand, nodes with similar neighborhood values are clustered in the same group.

Another important area is the relationship and interaction of users on social networks. The presence of a gap between two social network users is a strong criterion that reflects their actual relationship and interaction [25]. By creating a network of interactions between network members can provide more valuable information about their connection, even for people who are not directly connected in a friendship. Authors in [13] analyzed social network relationships in real time and capture a dynamic visualization of the network. The process of characterizing user relationships can start with a binary characterization $(0, 1)$ depending on whether they are present or not.

Moreover, the relationships between users in order to export other information such as geographic location, is utilized in [8]. Since location information is often chosen to be hidden from many users, this will be predicted using the data of user's connections to others as well as other alternative data that may implicitly include geographic location. The above user relationship and location information is valuable information, especially for the recent development of correlating web text with geographic location [22].

Another model based on user identity and interaction activity has been developed in [26]. What is more, it is an unsupervised method that aims to repeatedly understand how strong is the connection between two nodes. The model detects the interaction of user profiles and implies the strength of their connection depending on how similar are the two profiles. The key component on which user identity is based on, is the sociological concept of homophily, in which people with common or similar characteristics tend to form relationships with each other [19]. The results of a similar study depicts that pure homophily plays an important role in building connections in social networks [3].

The problem of predicting connections is addressed in [17] as authors also aim to find the parameters that lead to the most accurate prediction of future connections between social network users. In addition, the accuracy of edge predictions is proven to be increased by developing an algorithm that implies user profiles when creating a partial graph [5]. The utilized method does not include training data, but it is upon user's control to determine the amount of information accessed by the social network graph and the accuracy of the predictions.

III. Proposed Method

In this paper, data from the profiles of active social network users in order to draw conclusions regarding the strength of the relationships are analyzed, thus gaining knowledge for the degree among them [12], [27]. Specifically, data are extracted from the Twitter that contains a number of features from their profiles, in order to study the strength between the relationship, taking into consideration their connections as well as the similarity of these profiles [6]. The similarity between two user profiles can be considered as a major criterion for characterizing an edge as powerful ones.

In a nutshell, this paper examines the problem of identifying the degree of user engagement given the strength of their connection. In summary, the steps are presented below:

1) Initially, starting from an initial user, features from user profiles were considered and subsequently, a graph for representing the network is created.
2) The profiles are evaluated based on the degree of their contribution to the user relationship.
3) Finally, the relationships are categorized according to their strength and related statistical results are presented.

Similarity features take into account common or very close location, similar scale in number of friends, similar frequency of posts and interaction criteria, friendliness and follow relationship, user mentions and retweets. Based on the above data which contribute to a different percentage, a score is extracted for each edge that expresses the connection of two members and is therefore categorized according to this calculated score.

A. Metrics

In this subsection, the metrics that contribute to the calculation of the score of each edge will be recorded. There are two different metric categories utilized in our proposed method; the first one concerns similarity whereas the second deals with the interaction.

Specifically, the first attributes are related to the popularity of an account, which are the number of followers and friends. Another feature constitutes the number of tweets that each user has posted as well as the location that plays an important role. The location can be thought as a separate object of study and is considered a criterion of similarity since the closer geographically two users are, the greater the probability that they are also friends [5], [22].

Subsequently, metrics that are mostly associated with interaction include the mutual friendship condition. This is perhaps the most important feature as the aspect of bidirectionality is taking place. In addition to friendship, the “following” is also taken into account, since this means that a profile that follows another, is interested in following all of its activity on the social network. What is more, mentions and replies can be considered as interaction features. The last metric is the exchange of messages and specifically the authorization from both sides to be able to send and receive private messages between each other.

B. Calculation of Connection Scores

The score of each edge is calculated by taking into account the similarity presented in the above features and the user interaction. This process includes processing the set of edges collected and further checking for each metric if a specific condition is met. The sum of the scores is in the range $[0, 10]$ with 0 meaning no similarity and interaction and 10 meaning complete similarity and interaction between the profiles.

The above metrics are not all considered equal in their contribution to the strength of a connection; that’s why different weights are assigned to each one of them. The weights are given in the following Table I.

TABLE I
WEIGHTS PER METRIC

Metric Category	Metric	Symbol	Weight
Similarity	Friends Count	u_1	1
	Location	u_2	2
	Statuses Count	u_3	1
Interaction	Direct Message	u_4	1
	Following	u_5	1
	Mention	u_6	1.5
	Mutual Friendship	u_7	3
	Reply	u_8	1.5

Initially, features such as friends count and statuses count are assigned with weight equals to 1, while the location metric has weight equals to 2. That is, co-location can also imply not only online but also daily and face-to-face interaction. What is more, the existence of mutual friendship between two users is chosen as the most powerful metric as this denotes that they are strongly connected than when the connection is one-sided; the weight of this edge attribute is set to 3. Reciprocity in following indicates that there is interest on both sides to connect and therefore should be the most powerful parameter in this system as it is a measure of connection strength. Then, the mentions and replies have values equal to 1.5, whereas the weight of both “following” and direct message is 1, Concretely, mentions and replies reveal a significant interaction between users, which indicates an intimacy and negates the neutrality that exists in web social relationships.

Given the fact that each metric is considered an element of a vector W that represents the weights, then the resulting vector is $W = [w_1\ w_2\ w_3\ w_4\ w_5\ w_6\ w_7\ w_8]$.

Similarly, each vector V that contains for each edge all the features that are included, is denoted as $V = [v_1\ v_2\ v_3\ v_4\ v_5\ v_6\ v_7\ v_8]$.

Then, the total score for an edge connecting a user A to a user B within a particular Twitter subgraph is calculated as:

$$Score(A, B) = \sum_{i=1}^{8} w_i \tag{1}$$

This score will be calculated for each pair of edges giving therefore as output of our study a table with three columns, i.e. the source and the destination node as well as the calculated score from Equation 1.

In order for the edges categorization to take place, the score of each edge will be normalized with use of the following Equation 2:

$$Norm\ Sc(A,B) = \frac{Sc(A,B) - min(Sc(A,B))}{max(Sc(A,B) - min(Sc(A,B)} \cdot 10 \tag{2}$$

The resulting values are within the range $[0, 10]$.

C. Edges Categorization

The process of categorizing the edges is based on the strength of them. So five classes are chosen where the edges are categorized according to the obtained score from the previous calculations. The classes created according to the interval in which the score is located are as follows:

1) Indifferent [0, 2]
2) Weak (3, 4]
3) Medium (5, 6]
4) Strong (7, 8]
5) Very Strong (9, 10]

Of course, the higher scores indicate a stronger class. The maximum value of the score can be 10 and only appears if all metrics contribute optimal to a score indicating thus maximum similarity and interaction of the connections. The number of classes selected in this categorization is opted by several researches from different scientific disciplines and is also sufficient to obtain a good distribution of Twitter users' connections.

IV. Implementation

For the construction of our Twitter subgraph, we had to respect the current limitations of the Twitter API. It was collected in a time interval of one month, that is $(01/06/2022 - 30/06/2022)$. A topic-based sampling approach was used where tweets are collected via a keyword search query, namely #bigdata. This hashtag reflects a discussion topic with mostly scientific and business interest and quite sparse but also quite linear activity in time. The construction of the dataset was a two-step process by harvesting the related tweets per topic and then querying Twitter to get the followers and friends for each active user. That process resulted in a total number of 13.128 tweets, 1.354 user accounts from 115 different locations and 253.655 followers.

The following Figure 1 illustrates a social media graph and specifically the minimum cohesive graph of the particular users network where the nodes without edges were removed. There were such edgeless nodes as the access to the lists of some user profiles was prohibited. The blue dots represent Twitter profiles and the edges represent the "follow" connection between two users. As expected, the network appears to have much denser relationships internally, i.e. in its center, since the starting node constitutes the core of the network.

Table II presents the percentages of users based on the number of friends. Specifically, 22%, 35% and 23% of users have 0 to 100, 101 to 500 and 501 to $1,000$ friends respectively, whereas the remaining 20% have a bigger number of friends.

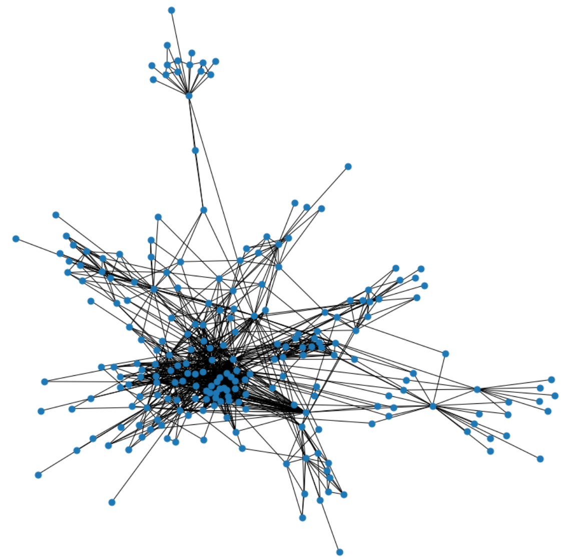

Fig. 1. Initial Network

TABLE II
PERCENTAGE OF USERS BASED ON NUMBER OF FRIENDS

Number of Friends	Percentage of Users
0 - 100	22
101 - 500	35
501 - 1000	23
1001 - 5000	11
over 5000	9

Table III depicts the percentages of users based on the number of followers. Concretely, 12% and 11% of users have 0 to 100 and 101 to 500 followers. Also, 9% and 13% of users have 501 to $1,000$ and $1,001$ to $5,000$ followers respectively, whereas the remaining 55% has over $5,000$ followers.

TABLE III
PERCENTAGE OF USERS BASED ON NUMBER OF FOLLOWERS

Number of Followers	Percentage of Users
0 - 100	12
101 - 500	11
501 - 1000	9
1001 - 5000	13
over 5000	55

Table IV presents the percentages of users based on the number of tweets. The percentages are very close to the ones in Table III as 6% and 9% of users have posted 0 to 100 and 101 to 500 tweets. Also, 10% and 25% of users have 501 to $1,000$ and $1,001$ to $5,000$ tweets respectively, whereas the remaining 50% has over $5,000$ number of tweets.

It is clear from the above tables, that there are many popular users within this particular dataset since the number of friends and followers is high. Also, users are quite active as large percentages of them have shared thousands of tweets.

TABLE IV
PERCENTAGE OF USERS BASED ON NUMBER OF TWEETS

Number of Tweets	Percentage of Users
0 - 100	6
101 - 500	9
501 - 1000	10
1001 - 5000	25
over 5000	50

V. EVALUATION

The major evaluation of our work concerns, except of the nodes, the edges. Specifically, we analyze the percentage that each metric contributes to the edge score, in following how the edges are distributed in the different classes and finally which are the maximum scores.

Table V presents the percentage of the edges overall score contribution. The metrics achieving the highest scores are the Direct Message and the Mutual Friendship with percentages equal to 35.5% and 33%, respectively. In following, the three similarity scores are following, i.e. Friends Count, Location and Statuses Count with percentage values equal to 9.5%, 9% and 7.5%, respectively. Finally, the lowest scores are achieved by the remaining three interaction metrics, namely Reply, Mention and Following with values lower than 3%.

TABLE V
EDGES OVERALL SCORE CONTRIBUTION PERCENTAGE

Metric Category	Metric	Overall Score Contribution
Similarity	Friends Count	9.5
	Location	9
	Statuses Count	7.5
Interaction	Direct Message	35.5
	Following	1
	Mention	1.5
	Mutual Friendship	33
	Reply	3

In the final analysis of the results, the categorization of the edges into five classes is implemented. The five classes are Very Strong, Strong, Medium, Weak and Indifferent as displayed in Table VI. This categorization is employed with use of the metrics and their respective scores. Indifferent, Weak and Medium classes achieve the highest percentages with values equal to 50.7%, 29.65% and 15.35% of edges, respectively. The small percentage of Strong and Very Strong edges in the social network is expected as strong ties indicate more close friendship between connecting users that is logically formed among a limited number of users.

TABLE VI
PERCENTAGES OF EDGES

Classes	Edges
Very Strong	0.55
Strong	3.75
Medium	15.35
Weak	29.65
Indifferent	50.7

Figure 2 is the same with Figure 1 but with additional color on the edges according to the above categories. Specifically, Very Strong and Strong edges are illustrated in red color, while Medium and Weak edges with yellow and Indifferent edges are presented in gray. In this graph, some colored areas can be distinguished in contrast to others; this may imply the existence of user groups that are connected to each other. In other words, smaller sub-networks are formed having strongly connected users; these users interact in a more detailed way.

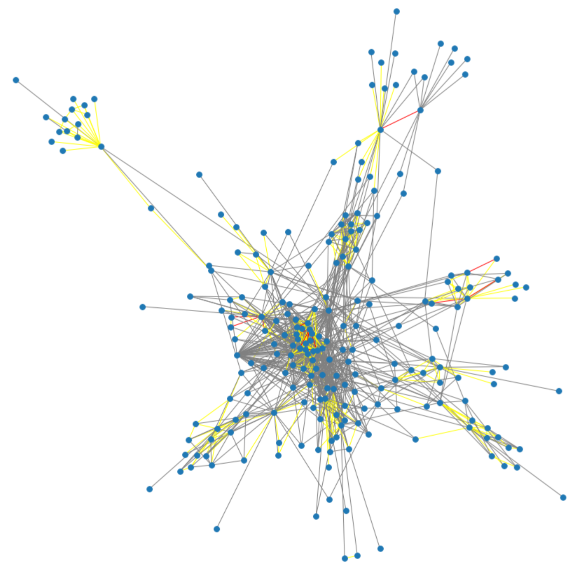

Fig. 2. Final Network

The aforementioned graph has density value equals to 0.04, which is considered an indicator of the number of connections given the maximum possible number of all pairs of nodes; this maximum number corresponds to value of density equal to 1. Additionally, the diameter of the graph is equal to 4; the diameter of a graph is the largest number of vertices that must be traversed in order to get from one vertex to another when looping paths are excluded from consideration.

A. Discussion

The results of this experimental study reveal several interesting facts. The metrics generated a diversity in the distribution of edges. Accordingly, a percentage of about 10% was contributed by the similarities in number of friends and posts. This is satisfactory as in contrast to Tables II and IV, there is a diversity within the dataset, e.g. there is a large percentage of users with less than 100 friends, etc.

The location similarity metric achieved a contribution percentage close to 10%, which could be even higher considering that many users don't put their location on their profile. Furthermore, we have to take into consideration the fact that many Twitter users may simultaneously use another social network platform, like Facebook or LinkedIn, and hence their interaction is shared among these platforms.

Finally, the most interesting result is derived from Figure 2 and pertains the smallest strong sub-networks that are formed within the original network. These colored edge regions could be clipped in a larger dataset along with their nodes and separately studied. This can be implemented by re-utilizing the present methodology to yield results regarding the robustness within different user groups.

VI. Conclusions and Future Work

The proposed study is based on real-time data but also on historical data. The combination of these contributes to the effective calculation of the strength of an edge connecting two users in a given social network. Of course, this information could be used to calculate the probability of a future connection between two users. Overall, the work offers knowledge about a rather important feature of social networks; the user relationships. This knowledge can contribute to additional conclusions regarding user engagement in social networks.

Furthermore, the research on user relationship in social networks exploits, except computer science, the fields of psychology and sociology as these establish the role of user behavior in the social network. The results show how strong an edge is and how much the individual involved in this calculation contribute.

Regarding future work, variations and combinations of the proposed methods presented in this work are worth trying, in order to study whether it is possible to further improve the performance [7], [15]. As work to come, we are interested in parallelizing the methods presented in the proposed paper for the creation of a nearly real-time user relationship analysis system. The implementation of streaming analytics technologies in combination with the present method could identify the alternations of relationships over time and further, will be able to learn and even predict changes in the network. Such an extension could also be combined with algorithms to predict the possibility of connection between two users in the social network.

References

[1] P. Candon. Twitter: Social communication in the twitter era. *New Media & Society*, 21(7), 2019.

[2] N. A. Christakis and J. H. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown and Company, 2009.

[3] M. Dehghani, K. Johnson, J. Hoover, E. Sagi, J. Garten, N. J. Parmar, S. Vaisey, R. Iliev, and J. Graham. Purity homophily in social networks. *Journal of Experimental Psychology: General*, 145(3):366, 2016.

[4] I. S. Dhillon, Y. Guan, and B. Kulis. A fast kernel-based multilevel algorithm for graph clustering. In *11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 629–634, 2005.

[5] R. Y. Dougnon, P. Fournier-Viger, and R. Nkambou. Inferring user profiles in online social networks using a partial social graph. In *28th Canadian Conference on Artificial Intelligence (AI)*, volume 9091, pages 84–99, 2015.

[6] G. Drakopoulos, A. Kanavos, K. Paximadis, A. Ilias, C. Makris, and P. Mylonas. Computing massive trust analytics for twitter using apache spark with account self-assessment. In *16th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 403–414, 2020.

[7] G. Drakopoulos, A. Kanavos, and A. K. Tsakalidis. Evaluating twitter influence ranking with system theory. In *12th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 113–120, 2016.

[8] C. A. D. Jr., G. L. Pappa, D. R. R. de Oliveira, and F. de Lima Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6):735–751, 2011.

[9] E. Kafeza, A. Kanavos, C. Makris, G. Pispirigos, and P. Vikatos. T-PCCE: twitter personality based communicative communities extraction system for big data. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1625–1638, 2020.

[10] N.-R. Kalogeropoulos, I. Doukas, C. Makris, and A. Kanavos. A graph-based extension for the set-based model implementing algorithms based on important nodes. In *16th International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, pages 143–154, 2020.

[11] A. Kanavos, G. Drakopoulos, and A. K. Tsakalidis. Graph community discovery algorithms in neo4j with a regularization-based evaluation metric. In *13th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 403–410, 2017.

[12] A. Kanavos and I. E. Livieris. Fuzzy information diffusion in twitter by considering user's influence. *International Journal on Artificial Intelligence Tools*, 29(2):2040003:1–2040003:22, 2020.

[13] J. Kim, E. Lee, J. Choi, Y. Bae, M. Ko, and P. Kim. Monitoring social relationship among twitter users by using nodexl. In *Research in Adaptive and Convergent Systems (RACS)*, pages 107–110, 2013.

[14] J. M. Kraus, G. Palm, and H. Kestler. On the robustness of semi-supervised hierarchical graph clustering in functional genomics. In *5th International Workshop on Mining and Learning with Graphs*, pages 147–150, 2007.

[15] I. Kyriazidou, G. Drakopoulos, A. Kanavos, C. Makris, and P. Mylonas. Towards predicting mentions to verified twitter accounts: Building prediction models over mongodb with keras. In *15th International Conference on Web Information Systems and Technologies (WEBIST)*, pages 25–33, 2019.

[16] T. V. Le, C. A. Kulikowski, and I. B. Muchnik. Coring method for clustering a graph. In *19th International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[17] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology (JASIST)*, 58(7):1019–1031, 2007.

[18] M.-F. G. Lin, E. S. Hoffman, and C. Borengasser. Is social media too social for class? a case study of twitter use. *TechTrends*, 57(2):39–45, 2013.

[19] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[20] P. Noordhuis, M. Heijkoop, and A. Lazovik. Mining twitter in the cloud: A case study. In *IEEE International Conference on Cloud Computing (CLOUD)*, pages 107–114, 2010.

[21] T. Ozaki and T. Ohkawa. Mining correlated subgraphs in graph databases. In *12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*, pages 272–283, 2008.

[22] R. Priedhorsky, A. Culotta, and S. Y. D. Valle. Inferring the origin locations of tweets with quantitative confidence. In *Computer Supported Cooperative Work (CSCW)*, pages 1523–1536, 2014.

[23] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *3rd International IEEE Conference on Privacy, Security, Risk and Trust (PASSAT) and 3rd International IEEE Conference on Social Computing (SocialCom)*, pages 180–185, 2011.

[24] B. K. Tripathy and A. Mitra. An algorithm to achieve k-anonymity and l-diversity anonymisation in social networks. In *4th International Conference on Computational Aspects of Social Networks (CASoN)*, pages 126–131. IEEE, 2012.

[25] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys*, pages 205–218, 2009.

[26] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *19th International Conference on World Wide Web (WWW)*, pages 981–990, 2010.

[27] V. Zamparas, A. Kanavos, and C. Makris. Real time analytics for measuring user influence on twitter. In *27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 591–597, 2015.