Social Network Analysis and Mining (2021) 11:115
https://doi.org/10.1007/s13278-021-00798-0

ORIGINAL ARTICLE

Mining social applications network from business perspective using modularity maximization for community detection

Zeeshan Akbar[1] · Jun Liu[1] · Zahida Latif[2]

Received: 3 August 2020 / Revised: 27 August 2021 / Accepted: 14 September 2021 / Published online: 2 November 2021

Abstract
There are different social applications available for different purposes. A lot of information about different fields including politics, sports, business, movie industry, etc., pass by and people are not well informed about most important happenings taking place in the world. Social applications usage varies among people in different parts of the world. A social application in a community may be popular for a particular purpose such as Twitter that may be used as a core application for political use among people in one part of the world, whereas other people may use Facebook, WeChat or YouTube for entertainment and other purposes and may not be aware of the important political changes taking place in the world. Social media usage by businesses can be improved by knowing the maximum usage of particular social applications among different communities of people so that targeted contents including information, advertisements, services and recommendations can be forwarded to them. In this paper, we mine social applications network by extracting knowledge according to the popularity of social applications. r-neighborhood technique is used for removal of edges from social applications network. Users are assigned to different communities based on the modularity scores. Optimal communities are found using divisive clustering approach that partitions the graph until maximum modularity score is achieved. Community detection method is also performed in gephi tool and using k-nearest neighbors graph. The trends of the social applications are analyzed among different communities, and it is seen that r-neighborhood, k-nearest neighbors and gephi tool result in Twitter, YouTube and Facebook as the most popular applications among other social applications. Related contents can be forwarded to the respective communities as well as people of a community defined by popularity of a social application can also be well informed about other happenings in the world such as Twitter and YouTube communities that may advertise about different products, whereas Facebook and YouTube communities are advertised with political news. The modularity function of k-nearest neighbors has the highest value and gives better interpretation of communities than other two techniques.

Keywords Community detection · Modularity · Social network · Divisive clustering · Similarity

1 Introduction

The online interaction of users has tremendously increased with the availability of different social applications and networking sites like Facebook, Twitter, WeChat, YouTube, Skype and many others. People with similar behaviors using social applications are linked with each other. The community structure in different networks like Internet, email, transportation, biochemical, citation and social networks shows a set of nodes with dense connections within community and sparse links out of community (Newman and Girvan 2004). The detection of such community structures in network systems is one of the key issues, known as community detection. The structures revealed by detecting communities in different networks are meaningful such as online and contact-based groups in social networks, customers' groups with similar interests in purchasing from online social networks, clusters of scientists in interdisciplinary collaboration networks. (Fortunato 2010).

Modularity maximization has been one of the most popular methods for community detection over partitioning a

✉ Zeeshan Akbar
zeeshan.mohmand@gmail.com

Jun Liu
liukeen@mail.xjtu.edu.cn

Zahida Latif
lateef.zahida@gmail.com

[1] Xian Jiaotong University China, Xi'an, China

[2] Center for Advanced Studies in Engineering (CASE), Islamabad, Pakistan

network (Newman and Girvan 2004; Newman 2006, 2004; Leicht and Newman 2008). Algorithms for modularity optimization including greedy algorithms such as Fast general hierarchical method, greedy optimization-based agglomeration algorithm, three forms of CNM algorithm by integrating consolidation ratio metrics, heuristic method by optimizing modularity (Newman 2004; Clauset et al. 2004; Wakita and Tsurumi 2007; Blondel et al. 2008), sampling technique using unsupervised method comprising of the proximity estimation and validation of hierarchical group of networks (Sales-Pardo et al. 2007), Eigen spectrum, spectral graph tri-partitioning algorithm, objective function maximization by proposing two new spectral methods, heuristic algorithm Qcut and recursive algorithm HQcut, Kcut spectral methods (Newman 2006; Newman 2006; Richardson et al. 2009; White and Smyth 2005; Ruan and Zhang 2008; Ruan and Zhang 2007; Newman 2013), extremal optimization algorithm (Duch and Arenas 2005), mathematical programming by proposing two unique linear programming and vector programming algorithms (Agarwal and Kempe 2008) and simulated annealing such as cartographic method and Monte Carlo methods (Guimera et al. , 2005a; b; Massen and Doye 2005; Medus. et al. 2005) have been proposed. This quality metric of network has been used as measurement of strength of community structure and is the difference between actual edges within community and expected edges in a randomized graph of same nodes and degrees. The degree is the number of edges connected to a node. This paper focuses on divisive clustering by maximizing graph modularity that add scores of every pair of nodes placed together in a single community.

Divisive clustering algorithms are 'top-down' in which all nodes are initially in a single cluster. The cluster splits recursively until each node forms its own cluster. Girvan–Newman algorithm (Girvan and Newman 2002) is a common divisive method that uses edge betweenness, the sparse connections between vertices of different communities, to determine the strength of edges and delete those edges whose has biggest betweenness until algorithm finds no edge for deletion. Another algorithm called Fast-Newman (Newman 2004) takes modularity as an objective function and gives optimal outcome when objective function indicated by Q has the highest value.

When dense clusters are selected which have sparse connections to the rest of the graph, this process is called community identification. In social networks, many overlapping of these communities are present with each node participating in many communities, which reveals the network features. Many approaches exist for community detection. However, the coupled-seed expansion method is effective as compared to many other existing algorithms such as Bigclam, OSLOM, SE, Demon, OMSTMO, LC, Ego-Splitting (Asmi et al. 2021). The modularity-based local community detection methods are widely used but also have some limitations to seed node selection and community instability. Considering the local modularity density and using Jaccard coefficient, the local communities can be formed by core area detection stage and the extension stage of the local communities which also provides efficiency and precision (Guo et al. 2021). A more generalized modularity measure called f-modularity when applied to simulated networks and also to the real-world market networks quantifies the community structure estimating the information existing between discrete random samples and big amount of value space (Guo et al. 2021). A more recent new algorithm which is slightly different from the graph neural network of unsupervised network community detection using modularity optimization has been proposed which is more efficient than fast Louvain method (Sobolevsky 2021). (v) Social networks and their analysis combine many techniques such as K-means clustering algorithm for many novel predictions such as drug target interactions using Bayes network, Naïve Bayes and SVM (Aghakhani et al. 2018).

Social applications network comprises different communication applications that facilitates different purposes including news sharing, marketing, entertainment, relationships, education, merchandising. Users of different social applications have more than one account and use these accounts for different purposes depending on the situation. In this case, Twitter is used for politics, YouTube for videos, WeChat for transactions, Facebook for profile information of products, WhatsApp for personal communications, Skype for meeting/interviews and Instagram for pictures.

In this research, we derive insights from a social applications network by creating a cosine similarity weighted graph of users. The cosine similarity is defined as counting same applications used by users divided by square root of the total applications used by one user multiplied by square root of the total applications used by other user. r-neighborhood technique is used for pruning edges of a network in which edges for a particular value of r are maintained while removing all other edges. It is hard to group a web of customers together present in r-neighborhood graph. Also, to determine whether a customer is present in a single or various communities, we use graph modularity maximization to make decisions about community assignments. Knowledge is extracted by analyzing the trends of social applications in order to forward advertisements, information, services, and recommendations to users. k-nearest neighbors' technique is also implemented for deletion of edges from social applications network of users. Communities are detected using modularity maximization by divisive clustering approach from r-neighborhood graph and k-nearest neighbors' graph. Gephi tool is also used to perform modularity maximization. All the three techniques indicate Twitter, YouTube and Facebook that are the most popular applications among other applications. However, modularity function of k-nearest neighbors has the highest value of 0.581 as compared to r-neighborhood and gephi tool which have values of 0.554 and 0.555.

2 Related work

2.1 Community detection

A review of various community detection metrics is presented and an efficient algorithm has been proposed that maximizes modularity density (Q_{ds}) (Chen et al. 2014). In another study, ten algorithms are re-implemented and evaluated on real-world datasets for community detection in a proposed framework (Wang et al. 2015). A new paradigm called HICODE is proposed to detect hidden community structures in many domains of real world. Experiments show that hidden communities exist in network (He et al. 2018). Community detection acts as a tool for analyzing network data, for example communities in social network defines the nature of social interactions among people.

There are natural divisions that exist in many complex systems and social networks that can be grouped into clusters having strong connections within the clusters and sparse links between them, known as community structure. In context of social applications, web has evolved and became a source of information helpful in analysis of web information using different models and brought intelligence through automation of web services (Cena et al. 2011). Improving recommender systems by describing different approaches used for recommendation and suggesting possible extensions for the limitations of mentioned approaches that can enhance the performance of the recommendation systems by forwarding different services and contents through web automatically (Adomavicius and Tuzhilin 2005). The hidden community structures in a social network that have to be explored in any social network are of great significance. To resolve this problem, graph compression-based community detection algorithms exist (Zhao et al. 2021) where the number of communities in a compressed social network with their initial community seeds is found out simultaneously. Addressing the heterogeneous properties of a vertex and using new probabilistic c-means model that uses attribute and structural similarities. This new model serves like fuzzy community detection that resolves the overlapping community detection problem (Naderipour et al. 2021). For stream graph, the local overlapping communities are detected at the end points of a newly found edge with common communities (Panchal 2021).

Based on the review of different empirical studies about the functionality and structure of a variety of networks, the task of community detection gives an insight into the core structure of networks. Developments in the statistical characteristics of different networks such as clustering, path lengths, degree distributions were mainly focused (Newman 2003). Due to complexity of the internal structure, these networks are defined as complex networks. Mathematical models, used to represent networks, are called graphs. In modern graph theory, the problem of partitioning a graph is also known as community detection (Diestel 2012; Bollobás 1998). Typically, there are two types of graph clustering algorithms with the first type having condensed regions of nodes and second type cluster different graphs using edges and structural characteristics (Aggarwal and Wang 2010). Different solutions include a new efficient, scalable algorithm based on recursive shingling and clustering steps that specifies huge dense subnetworks. A label distribution algorithm that assigns unique label to each community requires linear time for computations and is therefore less expensive (Gibson et al. 2005; Raghavan et al. 2007).

2.2 Modularity optimization

A new method related to the community structure is developed in many social and biological networks for the detection of communities. This new technique is based on the centrality indices to find the boundaries of the communities (Girvan and Newman 2002). This quality function having certain drawbacks like it may be unable to specify modules below a scale depending on the network size and degree. This drawback is validated in different real and artificial biological, technological and social networks (Fortunato and Barthelemy 2007; Wakita and Tsurumi 2007). Modularity is widely used because of the capability of auto-detection of optimal number of clusters by utilizing k-nearest neighbor graph construction and applying distance modularity by modifying Louvain algorithm (Ruan 2009; Shakarian et al. 2013). A graph with high modularity value indicates quality partitions and a good community structure. There are many modularity maximization methods introduced. One of the hierarchical method that maximizes modularity is Louvain algorithm (Adomavicius and Tuzhilin 2005). On large-scale networks, this algorithm runs very fast besides its ease of implementation and also avoids the resolution limit of modularity. A famous scholar Fortunato recommended it as best performance modularity optimization algorithm for community detection (Fortunato 2010).

2.3 Nearest neighbors

Neighborhood graphs model relationships among data points in various fields of machine learning including clustering, semi-supervised learning or dimensionality reduction. The two popular techniques are the r-neighborhood graph in which a specific point is connected to other points for a particular value of r and k-nearest neighbor graph (kNN) in which a point is connected to k-nearest neighbors. kNN is a popular classification technique (Samanthula et al. 2014; Xu et al. 2018; Wu et al. 2008; Cover and Hart 1967) that is used in different fields such as novel Voronoi-based kNN approach in spatial databases that outperforms online distance-based methods (Kolahdouzan and Shahabi 2004), gene classification by combining genetic algorithm and kNN method GA/kNN for assessment (Li al. 2001), and fault detection using kNN method (FD-kNN) in semiconductors is developed to handle nonlinearity in operation data (He and Wang 2007).

3 Community detection from business perspective in social networks

Social network analysis is based on community detection with nodes and edges representing the actors and their social connections, respectively, in a social graph which are commonly web in a dense manner with highly related and yet separated groups from each other. A lot of work has been done in this field of social network analysis, and many methods have been proposed in this regard (Chunaev 2020). The businesses around the world are growing due to social media boom as their target audience join and use these social networks in a regular manner and businesses have to take advantage of these social media platforms like Facebook, Twitter or Instagram to reach their highly targeted potential customers. Social media users and customers log into their accounts regularly with 70 percent of users logging into at least one per day (Pew Research Center 2021) which is the best source of staying on the top of customers' minds with effective digital marketing strategy.

With Facebook having almost over 2.7 billion active users around 180 countries and Twitter having 1 billion active users per month worldwide, the business owners should embed and understand the relevance of social networks and should design their communication strategies. The rapid growth of personal communities to business communities in online social networks proves it to be a highly cost-effective way of engaging with the customers with a significant value. Targeting the right customers on right social media platforms should be the integral part of any business plan with customer behaviors, demographics and trend analysis being properly worked upon in social media marketing strategy.

3.1 Contributions

Social media applications usage has changed the business dynamics in a tremendous manner, making it the only way forward to the future. This research serves to be a part of the new wave of making smarter business decisions by keeping near to the customers as much as possible. Both internal and external communications are crucial for the survival and progress of the businesses. Following are the contributions of the research:

- r-neighborhood, k-nearest neighbors' methods are used for removal of edges from network.
- Modularity maximization using divisive clustering approach is used for the detection of communities.
- Gephi tool is also used for detection of communities.
- The modularity score using r-neighborhood, k-nearest neighbors, and gephi tool is compared determining which technique results in better detection of communities.
- Knowledge is extracted according to popularity of social applications used in each community.
- The aim is to improve the scope, quality, richness, depth, interactivity and reach of the targeted contents using social applications popularity in a particular community. The effective decisions can also be taken among different fields such as improvement in business, i.e., forwarding product contents through particular social application maximum usage in a community. Community detection is performed by maximizing modularity using r-neighborhood, kNN, gephi and results are compared.

4 Methodology

4.1 Research framework

This research presents different social applications with different functionalities such as transactions, politics, video and profile information accessed through different mediums including mobile, tablet, computer and iPad for particular purpose. A set of users using those social applications is considered. The similarity between users is determined using cosine similarity, and a network of similar users is constructed. r-neighborhood and k-nearest neighbor's graphs are constructed by removing unnecessary edges from user similarity network. Communities are detected in r-neighborhood and kNN graphs using modularity maximization by divisive clustering approach. Gephi tool is also used for communities' detection using modularity maximization. Knowledge is extracted by determining which technique gives better and clear interpretation of communities.

The metadata consist of two types of data. Each social application used consists of functionality, purpose, application number and medium of access. It is also known that which user used the particular application by specifying the application number. In this paper, we have 32 instances representing different social applications accessed more than

Table 1 Social applications

App usage	App name	Functionality	Purpose	Medium of access
1	WeChat	Transactions	Merchandise	Mobile App
2	Twitter	Politics	News Sharing	Mobile App
3	YouTube	Videos	Entertainment	Mobile App
4	Facebook	Profile Info	Marketing	Mobile App
5	Instagram	Pictures	Brands Info	Mobile App
6	WhatsApp	Friends/Family	Greetings/Personal	Mobile App
7	YouTube	Videos	Entertainment	Tablet Web Browser
8	YouTube	Videos	Entertainment	Tablet App
9	Skype	Meetings/Interviews	Educational/Professional	Mobile App
10	WhatsApp	Friends/Family	Greetings/Personal	Computer Web Browser
11	Facebook	Profile Info	Marketing	Tablet App
12	WhatsApp	Friends/Family	Greetings/Personal	Tablet App
13	WeChat	Transactions	Merchandise	Tablet App
14	WeChat	Transactions	Merchandise	Computer App
15	Instagram	Pictures	Brands Info	Tablet App
16	WeChat	Transactions	Merchandise	iPad App
17	Twitter	Politics	News Sharing	Tablet App
18	YouTube	Videos	Entertainment	Mobile Web Browser
19	Facebook	Profile Info	Marketing	Computer Web Browser
20	Instagram	Pictures	Brands Info	Computer Web Browser
21	Facebook	Profile Info	Marketing	Tablet Web Browser
22	Facebook	Profile Info	Marketing	Mobile Web Browser
23	Skype	Meetings/Interviews	Educational/Professional	Tablet App
24	Twitter	Politics	News Sharing	Computer Web Bowser
25	Instagram	Pictures	Brands Info	Tablet Web Browser
26	Twitter	Politics	News Sharing	Tablet Web Browser
27	Facebook	Profile Info	Marketing	Computer App
28	Instagram	Pictures	Brands Info	Mobile Web Browser
29	Facebook	Profile Info	Marketing	iPad App
30	YouTube	Videos	Entertainment	Computer Web Browser
31	Facebook	Profile Info	Marketing	iPad Web Browser
32	Instagram	Pictures	Brands Info	iPad App

Table 2 List of users

User Number	App #					
User1	2	24				
User2	17	24	26			
User10	8	30				
User 100	1	2	11	22	28	30	31
User11	12	25	28			
User12	24	26				
User13	7	18	29	30		
User14	1	4	9	11	14	26
User15	28	29				
User16	17	24				
User17	2	11	28			
User18	1	2	11	15	22	
User19	9	16	25	30		
User20	14	22	25	30		

Table 3 Cosine similarity weighted matrix

	User1	User10	User100	User11	User12	User13	User14	User15	User16	User17	User18	User19	User2	User20
User1	0.000000	0.000000	0.267261	0.000000	0.500000	0.000000	0.000000	0.000000	0.500000	0.408248	0.316228	0.000000	0.408248	0.000000
User10	0.000000	0.000000	0.267261	0.000000	0.000000	0.353553	0.000000	0.000000	0.000000	0.000000	0.000000	0.353553	0.000000	0.353553
User100	0.267261	0.267261	0.000000	0.218218	0.000000	0.188982	0.308607	0.267261	0.000000	0.654654	0.676123	0.188982	0.000000	0.377964
User11	0.000000	0.000000	0.218218	0.000000	0.000000	0.000000	0.000000	0.408248	0.000000	0.333333	0.000000	0.288675	0.000000	0.288675
User12	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.288675	0.000000	0.500000	0.000000	0.000000	0.000000	0.816497	0.000000
User13	0.000000	0.353553	0.188982	0.000000	0.000000	0.000000	0.000000	0.353553	0.000000	0.000000	0.000000	0.250000	0.000000	0.250000
User14	0.000000	0.000000	0.308607	0.000000	0.288675	0.000000	0.000000	0.000000	0.000000	0.235702	0.365148	0.204124	0.235702	0.204124
User15	0.000000	0.000000	0.267261	0.408248	0.000000	0.353553	0.000000	0.000000	0.000000	0.408248	0.000000	0.000000	0.000000	0.000000
User16	0.500000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.816497	0.000000
User17	0.408248	0.000000	0.654654	0.333333	0.000000	0.000000	0.235702	0.408248	0.000000	0.000000	0.516398	0.000000	0.000000	0.000000
User18	0.316228	0.000000	0.676123	0.000000	0.000000	0.000000	0.365148	0.000000	0.000000	0.516398	0.000000	0.000000	0.000000	0.223607
User19	0.000000	0.353553	0.188982	0.288675	0.000000	0.250000	0.204124	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000
User2	0.408248	0.000000	0.000000	0.000000	0.816497	0.000000	0.235702	0.000000	0.816497	0.000000	0.000000	0.000000	0.000000	0.000000
User20	0.000000	0.353553	0.377964	0.288675	0.000000	0.250000	0.204124	0.000000	0.000000	0.000000	0.223607	0.500000	0.000000	0.000000

once through different mediums for different purposes and a list of about 324 usages of these applications by 100 users.

4.2 Cosine similarity weighted graph construction

A user-to-user graph is constructed using cosine similarity matrix that shows how much users are similar to each other in terms of usage of social applications. Consider two vectors (1, 1) and (1, 0) where 1 represents usage of application by a user. The cosine similarity between users is calculated as (Foreman 2013):

Matching common applications usage between the two users divided by square root of total applications used by first user multiplied by square root of total applications used by second user.

Cosine (45) = 1 common application/SQRT {total applications used by first user} * SQRT {total applications used by second user} = 0.707.

This weighted graph using cosine similarity shows each pair of users having either a zero or nonzero value showing the strength of an edge, an affinity matrix.

4.3 r-Neighborhood graph construction

An r-neighborhood graph for set of nodes with vertex set V and edge v, such that the edge $v \in V$ to its similar nodes in V for a given similarity, i.e., cosine similarity is constructed. To create adjacency matrix that comprises edges of certain strength for a given set of points $x_1, x_2, x_3 \ldots\ldots x_n$, the r-neighborhood graph is $G_{n,\ r}$: For an edge from point x_i to x_j, A_{ij} is 1, if $Simil\ (x_i, x_j) \geq r$, for all $1 \leq i, j \leq n, i \neq j$. In this case, r-neighborhood graph is produced for $r = 0.5$, in which edges are removed that has strength between users with similarity less than 0.5.

4.4 k-Nearest neighbors graph construction

In k-nearest neighbors graph, each node is connected to its nearest neighbors for a k value. Given a set of nodes P, the kNN graph is $G\ (P, E)$, whereas $E = \{(u, v\ Simil\ (u, v)), v \in NN(u)_{simil}\}$ where $NN(u)_{simil}$ is the nearest neighbor for each $u \in P$. In this case $k = 5$, we construct 5NN graph from the affinity matrix where five edges that have highest affinities are coming out of each node. Adjacency matrix is generated from affinity matrix, l represents the fifth highest affinity of each user, so A_{uv} is 1, if $Simil\ (u, v) \geq l$, for all $1 \leq u, v \leq n$, $u \neq v$.

4.5 Modularity maximization using divisive clustering

Modularity maximization using divisive clustering is used for community detection. This method assigns scores to

Fig. 1 Social applications similarity network of users

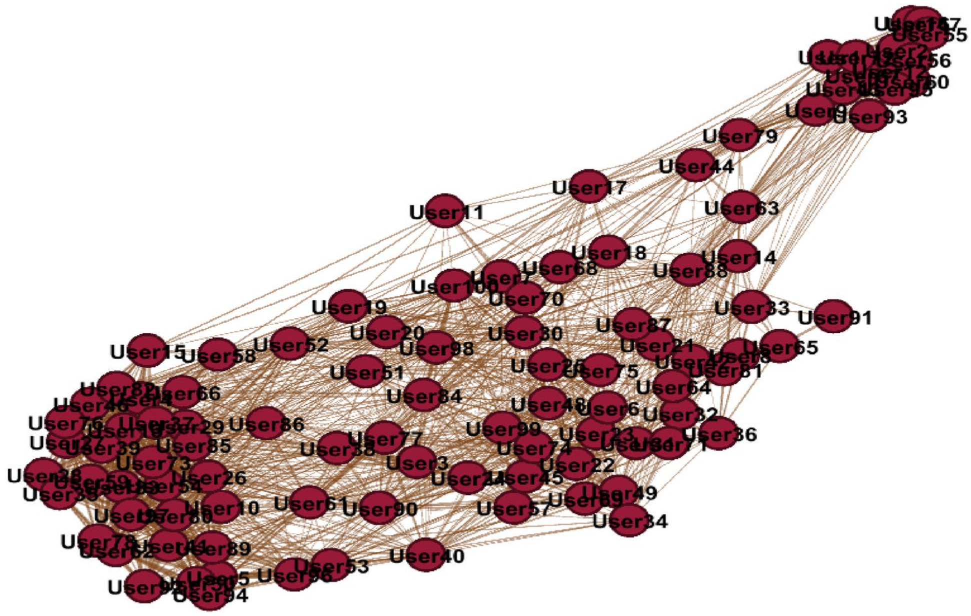


Table 4 Adjacency matrix

	User1	User10	User100	User11	User12	User13	User14	User15	User16	User17	User18	User19	User2	User20
User1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
User10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User100	0	0	0	0	0	0	0	0	0	1	1	0	0	0
User11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User12	1	0	0	0	0	0	0	0	1	0	0	0	1	0
User13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User16	1	0	0	0	1	0	0	0	0	0	0	0	1	0
User17	0	0	1	0	0	0	0	0	0	0	1	0	0	0
User18	0	0	1	0	0	0	0	0	0	1	0	0	0	0
User19	0	0	0	0	0	0	0	0	0	0	0	0	0	1
User2	0	0	0	0	1	0	0	0	1	0	0	0	0	0
User20	0	0	0	0	0	0	0	0	0	0	0	1	0	0

each pair of nodes in the r-neighborhood network. Divisive clustering splits the graph into two communities and uses an optimization algorithm for different community assignments in order to get maximum modularity score. The two communities are further divided into four and so on, until modularity maximization stops and gives optimal communities. Mathematically,

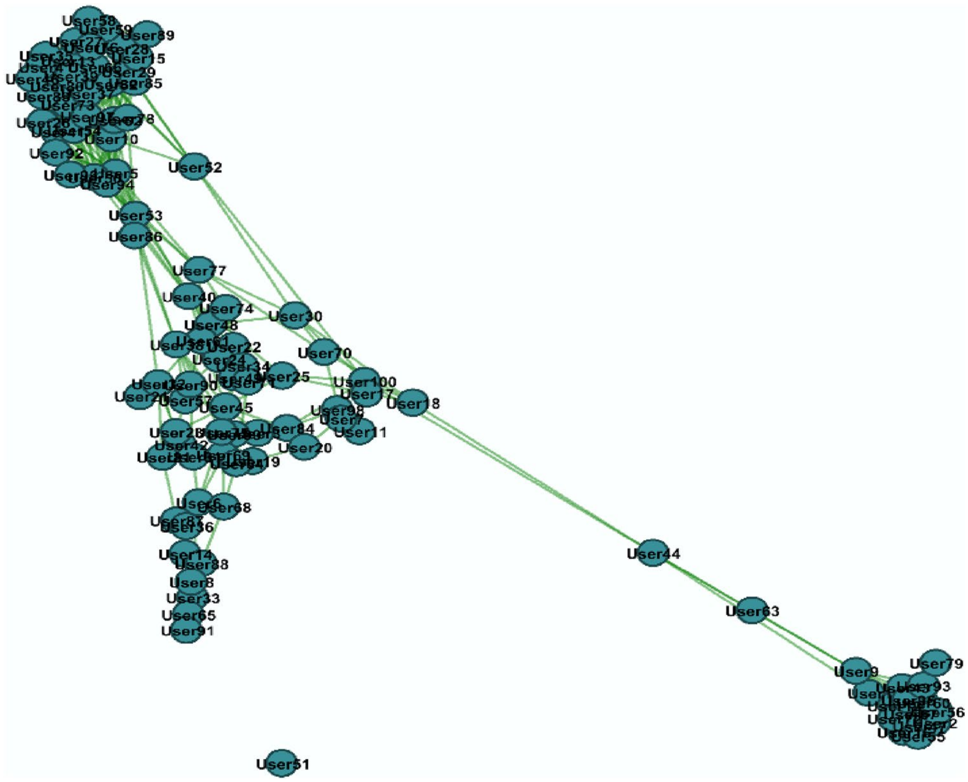
$$Q = \sum_{c_i \in C} \left[\frac{\left|E^{\text{in}}_{c_i}\right|}{|E|} - \left(\frac{2\left|E^{\text{in}}_{c_i}\right| + \left|E^{\text{out}}_{c_i}\right|}{2|E|} \right)^2 \right] \tag{1}$$

In the above equation, C represents all communities, where c_i refers to a particular community, $\left|E^{in}_{c_i}\right|$ shows edges of nodes inside community c_i, $\left|E^{out}_{c_i}\right|$ are the links to nodes of other community and $|E|$ represents total edge count in a network.

4.6 Knowledge extraction

Knowledge is extracted by determining the maximum usage of social applications in a particular community so that targeted contents can be forwarded to a community using those popular social applications. The modularity maximization

Fig. 2 *r*-neighborhood graph



for community detection is also performed using gephi tool. Knowledge extracted is compared with *r*-neighborhood and *k*-nearest neighbors' techniques for the same purpose of application popularity.

5 Results and discussion

This section presents the results and analysis. In Table 1, 32 instances of social applications including Twitter, WeChat, YouTube, Facebook, Instagram, WhatsApp and Skype are accessed for different purposes such as news sharing, merchandise, entertainment, marketing, brands information, educational/professional, greetings/personal.

In Table 2, for simplicity and specificity, metadata of only 13 random users from a list of 100 users accessing social applications are presented. User1 and User2 use Twitter through mobile, tablet and computer, User10 accesses YouTube. The User100 uses WeChat, Twitter, Facebook, YouTube and Instagram like other users.

In Table 3, we present user-to-user cosine similarity matrix that shows user similarity within range 0–1, with 1 having highest similarity in the context of application usage. As the user has maximum similarity to himself but our interest is to construct graph of users that are similar to one another in terms of applications usage and not to himself, so those values are made 0. Other values show how much similar applications are used by the two users. A value of 0.5 or above shows that more than 50% of the applications used by the two users are same.

A social applications network of users is shown in Fig. 1, with users similar to each other from the affinity matrix. Nodes representing users are connected to each other with value above zero in the affinity matrix and no node is connected to itself because those values are made zero representing no edge. The usage of applications by different users is converted into thousands of edges. Even if a single application is common between two users representing a very small value of cosine similarity, it is shown by an edge in the network.

5.1 r-Neighborhood graph construction

To produce *r*-neighborhood graph, adjacency matrix is created, as shown in Table 4 from user similarity matrix that comprises only those edges that are of certain strength and not having small cosine similarity value that may appear due to random usage of a single application. In this case, $r = 0.5$, comprises 20 percent of relationships among users that have highest affinities. The values above $r = 0.5$ are made 1, whereas the values below are made 0. The value of 1 represents an edge between users, whereas 0 means that the edge does not exist in the adjacency matrix.

Table 5 Scores for pair of nodes

	User1	User10	User100	User11	User12	User13	User14	User15	User16	User17	User18	User19	User2	User20
User1	− 0.14102	− 0.23076	− 0.10256	− 0.01282	0.83333	− 0.25641	− 0.02564	− 0.07692	0.87179	− 0.03846	− 0.06410	− 0.02564	− 0.14102	− 0.03846
User10	− 0.23076	− 0.37762	− 0.16783	− 0.02097	− 0.27272	− 0.41958	− 0.04195	− 0.12587	− 0.20979	− 0.06293	− 0.10489	− 0.04195	− 0.23076	− 0.06293
User100	− 0.10256	− 0.16783	− 0.07459	− 0.00932	− 0.12121	− 0.18648	− 0.01864	− 0.05594	− 0.09324	0.97202	0.95337	− 0.01864	− 0.10256	− 0.02797
User11	− 0.01282	− 0.02097	− 0.00932	− 0.00116	− 0.01515	− 0.02331	− 0.00233	− 0.00699	− 0.01165	− 0.00349	− 0.00582	− 0.00233	− 0.01282	− 0.00349
User12	0.83333	− 0.27272	− 0.12121	− 0.01515	− 0.19696	− 0.30303	− 0.03030	− 0.09090	0.84848	− 0.04545	− 0.07575	− 0.03030	0.83333	− 0.04545
User13	− 0.25641	− 0.41958	− 0.18648	− 0.02331	− 0.30303	− 0.46620	− 0.04662	− 0.13986	− 0.23310	− 0.06993	− 0.11655	− 0.04662	− 0.25641	− 0.06993
User14	− 0.02564	− 0.04195	− 0.01864	− 0.00233	− 0.03030	− 0.04662	− 0.00466	− 0.01398	− 0.02331	− 0.00699	− 0.01165	− 0.00466	− 0.02564	-0.00699
User15	− 0.07692	− 0.12587	− 0.05594	− 0.00699	− 0.09090	− 0.13986	− 0.01398	− 0.04195	− 0.06993	− 0.02097	− 0.03496	− 0.01398	− 0.07692	− 0.02097
User16	0.87179	− 0.20979	− 0.09324	− 0.01165	0.84848	− 0.23310	− 0.02331	− 0.06993	− 0.11655	− 0.03496	− 0.05827	− 0.02331	0.87179	− 0.03496
User17	− 0.0384	− 0.06293	0.97202	− 0.00349	− 0.04545	− 0.06993	− 0.00699	− 0.02097	− 0.03496	− 0.01048	0.98251	− 0.00699	− 0.03846	− 0.01048
User18	− 0.0641	− 0.10489	0.95337	− 0.00582	− 0.07575	− 0.11655	− 0.01165	− 0.03496	− 0.05827	0.98251	− 0.02913	− 0.01165	− 0.06410	− 0.01748
User19	− 0.0256	− 0.04195	− 0.01864	− 0.00233	− 0.03030	− 0.04662	− 0.00466	− 0.01398	− 0.02331	− 0.00699	− 0.01165	− 0.00466	− 0.02564	0.99300
User2	− 0.1410	− 0.23076	− 0.10256	− 0.01282	0.83333	− 0.25641	− 0.02564	− 0.07692	0.87179	− 0.03846	− 0.06410	− 0.02564	− 0.14102	− 0.03846
User20	− 0.0384	− 0.06293	− 0.02797	− 0.00349	− 0.04545	− 0.06993	− 0.00699	− 0.02097	− 0.03496	− 0.01048	− 0.01748	0.99300	− 0.03846	− 0.01048

In Fig. 2, after the removal of unnecessary edges from the similarity network of users for $r = 0.5$, the number of edges is reduced and it can be observed that one user is alone and has value less than $r = 0.5$ showing no edges to other users.

Communities are detected using modularity maximization by divisive clustering. We assign scores to each pair of nodes using r-neighborhood graph as shown in Table 5 and then perform divisive clustering until modularity maximization stops. The negative score indicates that they do not share an edge and placing them in a community will give a negative modularity score; however, positive score between users indicates that they share an edge and are similar and placing them in the same community gives maximum modularity score.

The modularity maximization problem is approached using divisive clustering by partitioning the graph. The modularity score of every user from Table 5 is calculated in such a way that if a user is assigned to community 1, scores of all those users from the respective row will be added that are also assigned to community 1. The total score is calculated by summing all the modularity scores in Table 6 and normalizing it by the total stub count of the network. The upper limits determine the modularity score of each community where upper limit 1 is the modularity score of community 1 and upper limit 2 is the modularity score of community 2. The first division of graph results into community 1 and community 0 with a total normalized modularity score of 0.464 as shown in Table 6.

To maximize modularity score, we further split the communities. In Table 7, further division of users is shown in different communities; however, the first partition of users is presented under the last assignment column, and total modularity score increased from 0.464 to 0.554. The modularity score after second division is calculated in the same way by adding scores of users if they are placed in the same community, since the two users will be in the same community only if their last and current community assignment is same such as User100 and User11 in Table 7. However, User 10 and User 100 are currently assigned to community 1, whereas their last community assignment is different so they will be in different communities. The modularity score after third division is calculated in the same way.

In Table 8, the users are further divided into communities but the total modularity score remains the same and communities are not further divided. The second partition is shown as last Assignment1 in Table 8. To analyze the user assignment to different communities, they are first encoded into decimal shown in Table 9. Div1 and Div2 represent the first and second community assignments.

Table 6 Community assignment on first division

	Community	Modularity Score	Upper Limit 1	Upper Limit 2
User1	1	6.1794872	6.179487179	12.74358974
User10	0	7.8881119	16.23776224	7.888111888
User100	1	3.4941725	3.494172494	11.74125874
User11	1	0.5617715	0.561771562	1.431235431
User12	1	7.3030303	7.303030303	14.3030303
User13	0	8.7645688	16.13053613	8.764568765
User14	1	1.1235431	1.123543124	2.843822844
User15	0	2.6293706	8.041958042	2.629370629
User16	1	5.6177156	5.617715618	11.63170163
User17	1	1.6853147	1.685314685	4.181818182
User18	1	2.8088578	2.808857809	6.888111888
User19	1	1.1235431	1.123543124	2.843822844
User2	1	6.1794872	6.179487179	12.58974359
User20	1	1.6853147	1.685314685	4.230769231
Total Score = 0.464				

Table 7 Community assignment on second division

	Community	Last assignment	Modularity Score	Upper Limit 1	Upper Limit 2
User1	0	1	8.87179490	10.05128205	8.871794872
User10	1	0	7.88811190	7.888111888	16.23776224
User100	1	1	4.04195800	4.041958042	9.324009324
User11	1	1	0.75524476	0.755244755	1.237762238
User12	0	1	10.48484800	11.12121212	10.48484848
User13	1	0	8.76456880	8.764568765	16.13053613
User14	1	1	1.51048950	1.51048951	2.456876457
User15	1	0	2.62937060	2.629370629	8.041958042
User16	0	1	8.06526810	9.184149184	8.065268065
User17	1	1	2.26573430	2.265734266	3.601398601
User18	1	1	2.77622380	2.776223776	6.920745921
User19	1	1	1.51048950	1.51048951	2.456876457
User2	0	1	8.87179490	9.897435897	8.871794872
User20	1	1	2.26573430	2.265734266	3.65034965
Total Score = 0.554					

5.2 Knowledge extraction

To analyze the trends of social applications in different communities, we determine the popularity of social applications by counting the number of times each application is accessed by users in every community. The partitioning of the graph results in a total of four communities represented by C1-C4 in Table 10. The first community C1 uses Facebook as the most popular social application; however, Twitter in the second community under C2 has highest values where YouTube is the popular application in C3 community. Community four is not clear but accesses Facebook in large number along with other social applications.

Table 8 Community assignment on third division

	Community	Last Assign-ment1	Last Assign-ment	Modularity Score	Upper Limit 1	Upper Limit 2
User1	0	0	1	8.871795	10.05128205	8.871794872
User10	1	1	0	7.8881119	7.8881119	16.23776224
User100	0	1	1	4.142191	7.317016317	4.1421911
User11	0	1	1	0.8927739	1.1002331	0.89277389
User12	0	0	1	9.342343	11.12121212	9.342343
User13	1	1	0	8.7645688	8.76456876	16.13053613
User14	1	1	1	1.7249417	1.7249417	2.242424242
User15	1	1	0	2.6293706	2.62937063	8.041958042
User16	0	0	1	8.06527	9.184149184	8.065268065
User17	0	1	1	2.6783217	3.188811189	2.678321678
User18	0	1	1	3.4638695	4.291375291	3.46386946
User19	1	1	1	1.7249417	1.724941725	2.242424242
User2	0	0	1	7.324167	9.897435897	7.324167
User20	1	1	1	1.5874126	1.587412587	4.328671329
Total Score = 0.554						

In Figs. 3, 4, 5 and 6, as compared to Table 10, it can be observed that Facebook is used in community one to which only a single member is assigned as shown in *r*-neighborhood graph, whereas community two is a Twitter community, which can be interpreted as political community. Community three members are fond of YouTube and are not interested in politics, which can be named as entertainment community. The last community which is not quite clear but most members use Facebook and Instagram than the rest of the applications. The members of this community are found to be socially interactive.

Table 9 Coding communities

	Div2	Div1	Communities
User1	0	1	1
User10	1	0	2
User100	1	1	3
User11	1	1	3
User12	0	1	1
User13	1	0	2
User14	1	1	3
User15	1	0	2
User16	0	1	1
User17	1	1	3
User18	1	1	3
User19	1	1	3
User2	0	1	1
User20	1	1	3

5.3 Community detection using gephi

In the following Table 11, communities are detected using *r*-neighborhood adjacency matrix and modularity maximization is performed using gephi tool. Five different communities are detected from 0 to 4 but for simplicity, the number of users presented in Table 11 is assigned from community zero to community three. The total modularity score using gephi is 0.555, whereas modularity score was 0.554 using divisive clustering.

The community assignment of users is shown in Fig. 7, where users are divided into five different communities. Nodes in the red show community one members, whereas the orange mesh represents community two. Community three users are the pink ones. Gray nodes are for community four users, and a single member is represented by blue node in community five which can be compared against Table 12, where only few applications are used.

5.4 Knowledge extraction using Gephi

Knowledge is extracted by determining the popularity of social applications in the assigned communities, presented in Table 12. There are five communities showing social applications usage. Twitter is accessed in community one. You Tube and Facebook are the popular social applications in community two, three and four along with other social applications. Community five is assigned a single member who uses Facebook, Instagram and WeChat.

In Fig. 8, Facebook, YouTube and Twitter are the obvious social applications used in all five communities which can be compared against Table 12. The results are almost similar

Table 10 Social applications popularity using *r*-neighborhood

App Usage	App name	Functionality	Purpose	Medium of Access	C1	C2	C3	C4
1	WeChat	Transactions	Merchandise	Mobile App	0	2	0	8
2	Twitter	Politics	News Sharing	Mobile App	0	7	0	3
3	YouTube	Videos	Entertainment	Mobile App	0	0	0	6
4	Facebook	Profile Info	Marketing	Mobile App	0	0	0	12
5	Instagram	Pictures	Brands Info	Mobile App	0	0	0	4
6	WhatsApp	Friends/Family	Greetings/Personal	Mobile App	0	0	1	11
7	YouTube	Videos	Entertainment	Tablet Web Browser	0	0	14	5
8	YouTube	Videos	Entertainment	Tablet App	0	0	16	4
9	Skype	Meetings/Interviews	Educational/Professional	Mobile App	0	0	0	10
10	WhatsApp	Friends/Family	Greetings/Personal	Computer Web Browser	0	1	1	5
11	Facebook	Profile Info	Marketing	Tablet App	1	0	1	11
12	WhatsApp	Friends/Family	Greetings/Personal	Tablet App	0	1	1	3
13	WeChat	Transactions	Merchandise	Tablet App	0	0	6	0
14	WeChat	Transactions	Merchandise	Computer App	0	0	0	9
15	Instagram	Pictures	Brands Info	Tablet App	0	0	0	6
16	WeChat	Transactions	Merchandise	iPad App	1	1	0	3
17	Twitter	Politics	News Sharing	Tablet App	0	7	0	0
18	YouTube	Videos	Entertainment	Tablet App	0	0	13	1
19	Facebook	Profile Info	Marketing	Computer Web Browser	0	0	0	5
20	Instagram	Pictures	Brands Info	Computer Web Browser	1	0	0	5
21	Facebook	Profile Info	Marketing	Tablet Web Browser	0	0	1	3
22	Facebook	Profile Info	Marketing	Mobile Web Browser	0	0	0	21
23	Skype	Meetings/Interviews	Educational/Professional	Tablet App	0	1	0	4
24	Twitter	Politics	News Sharing	Computer Web Bowser	0	12	0	0
25	Instagram	Pictures	Brands Info	Tablet Web Browser	0	0	0	6
26	Twitter	Politics	News Sharing	Tablet Web Browser	0	12	0	3
27	Facebook	Profile Info	Marketing	Computer App	0	1	1	7
28	Instagram	Pictures	Brands Info	Mobile Web Browser	0	0	1	5
29	Facebook	Profile Info	Marketing	iPad App	1	0	16	0
30	YouTube	Videos	Entertainment	Computer Web Browser	0	0	17	5
31	Facebook	Profile Info	Marketing	iPad Web Browser	1	0	0	16
32	Instagram	Pictures	Brands Info	iPad App	0	0	0	4

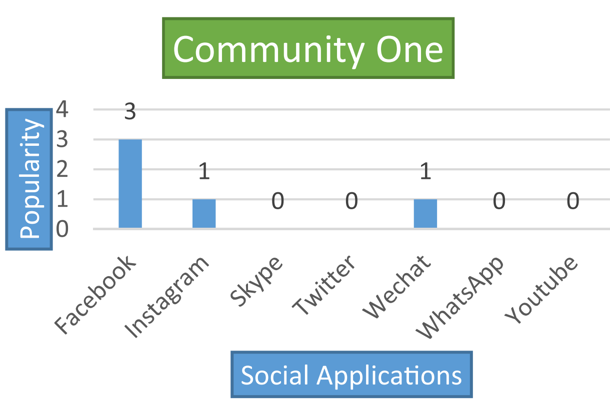


Fig. 3 Social application popularity

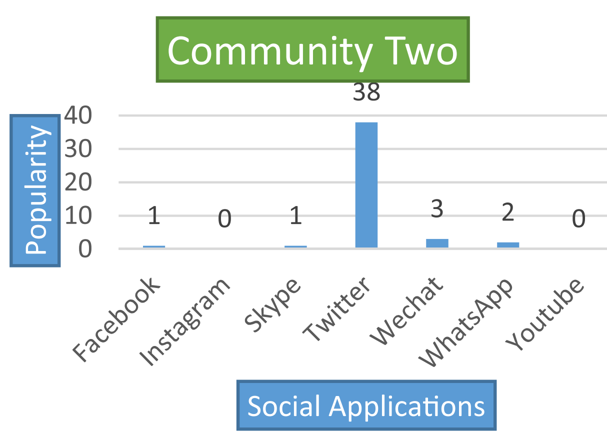


Fig. 4 Social applications popularity

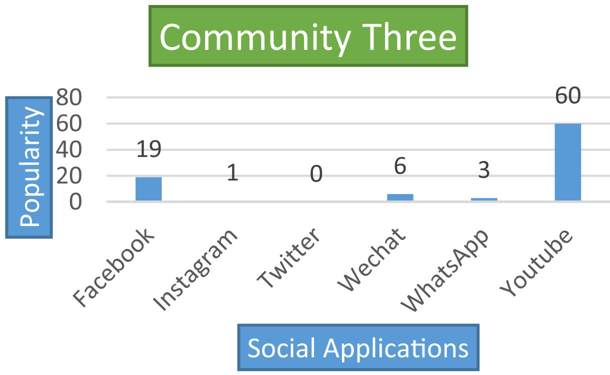


Fig. 5 Social applications popularity

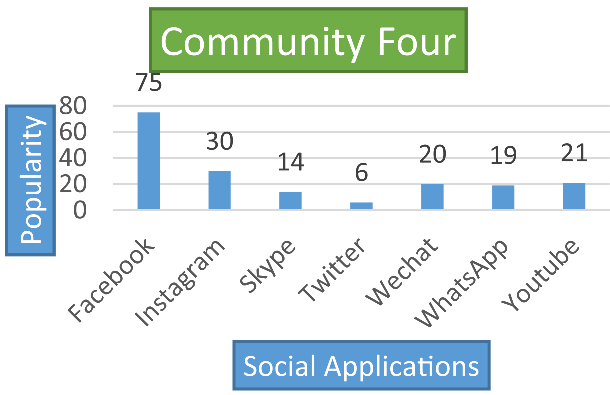


Fig. 6 Social applications popularity

Table 11 Community assignment Using Gephi

Label	Class	Score
User1	0	7.961538
User10	1	7.888112
User100	3	4.030303
User11	2	0.868298
User12	0	10.59091
User13	1	8.764569
User14	2	1.736597
User15	1	2.629371
User16	0	8.146853
User17	3	2.636364
User18	3	4.393939
User19	2	1.736597
User2	0	8.961538
User20	2	2.604895
Total Score = 0.555		

such as community one can be regarded as Twitter community following politics, community two is the same as You Tube community. In community three and four, Facebook is the most common social application used apart from Skype and WhatsApp.

5.5 kNN graph construction

Different values of k can be evaluated but in this research, we perform comparison of k-nearest neighbors with r-neighborhood and gephi tool. Since r-neighborhood and gephi tool resulted in a maximum of five different communities, therefore a graph for $k=5$ is constructed. Table 13 shows 5NN adjacency matrix where each user has a value of 1 in the respective row to five different users that have highest affinities in the similarity matrix than other users of the same row and are considered the nearest neighbors of that particular user.

The kNN graph for $k=5$, called 5NN graph, using adjacency matrix is constructed from user similarity graph in which all edges are removed from each node leaving it with five nearest neighbors that have highest affinities, as shown in Fig. 9.

In Table 14, we show community assignment of users to five different communities from 0 to 4 using kNN graph with a modularity score of 0.581.

The visualizations for the community assignment of 5NN graph are shown in Fig. 10, where users are divided into five different communities. Nodes in the red color present community zero, where the users of community one are pink in color. The nodes in yellow present users of community two. Community three is presented by gray color and finally nodes in blue belong to community four.

5.6 Knowledge extraction using 5NN graph

The five different communities are presented in Table 15 for knowledge extraction where each community represents the popular applications among other applications. Community one is the twitter community. Facebook has the highest values of usage in community two and three, whereas You Tube is popular in community four and five.

The Table 15 shows different social applications used in different communities. The knowledge extracted is shown below with the help of figures that can be compared against Table 15.

It can be seen in Figs. 11, 12, 13, 14 and 15 that community one again is a Twitter community. In community two and three, most users are Facebook followers. Community four and five are clearly an entertainment and social communities with You Tube and Facebook accessed more than other applications.

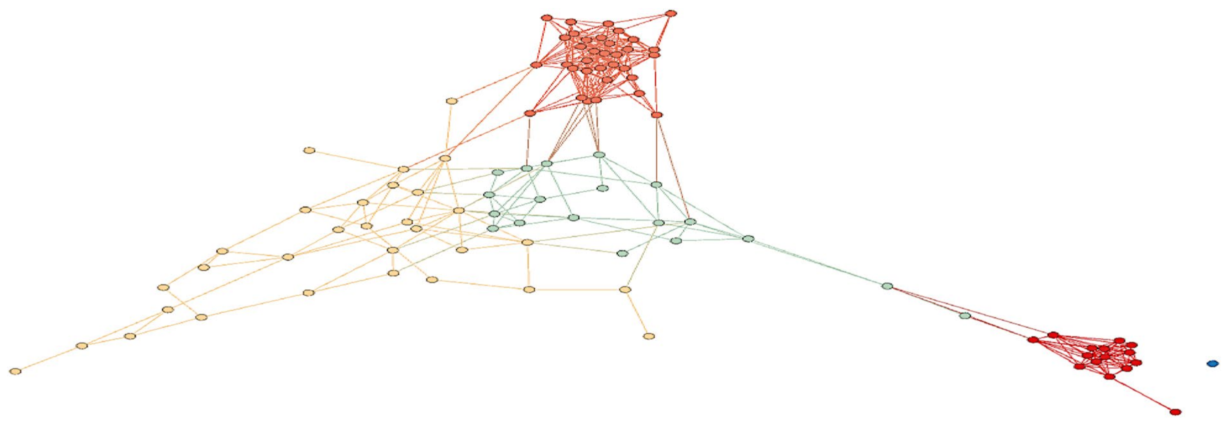


Fig. 7 Community assignment of users

Table 12 Social applications popularity using gephi

App Usage	App name	Functionality	Purpose	Medium of Access	C1	C2	C3	C4	C5
1	WeChat	Transactions	Merchandise	Mobile App	0	0	6	4	0
2	Twitter	Politics	News Sharing	Mobile App	5	0	0	5	0
3	YouTube	Videos	Entertainment	Mobile App	0	0	1	5	0
4	Facebook	Profile Info	Marketing	Mobile App	0	0	4	8	0
5	Instagram	Pictures	Brands Info	Mobile App	0	0	4	0	0
6	WhatsApp	Friends/Family	Greetings/Personal	Mobile App	0	1	5	6	0
7	YouTube	Videos	Entertainment	Tablet Web Browser	0	14	3	2	0
8	YouTube	Videos	Entertainment	Tablet App	0	16	1	3	0
9	Skype	Meetings/Interviews	Educational/Professional	Mobile App	0	0	10	0	0
10	WhatsApp	Friends/Family	Greetings/Personal	Computer Web Browser	0	1	5	1	0
11	Facebook	Profile Info	Marketing	Tablet App	0	1	3	8	1
12	WhatsApp	Friends/Family	Greetings/Personal	Tablet App	1	1	2	1	0
13	WeChat	Transactions	Merchandise	Tablet App	0	6	0	0	0
14	WeChat	Transactions	Merchandise	Computer App	0	0	9	0	0
15	Instagram	Pictures	Brands Info	Tablet App	0	0	5	1	0
16	WeChat	Transactions	Merchandise	iPad App	1	0	2	1	1
17	Twitter	Politics	News Sharing	Tablet App	7	0	0	0	0
18	YouTube	Videos	Entertainment	Tablet App	0	13	1	0	0
19	Facebook	Profile Info	Marketing	Computer Web Browser	0	0	3	2	0
20	Instagram	Pictures	Brands Info	Computer Web Browser	0	0	2	3	1
21	Facebook	Profile Info	Marketing	Tablet Web Browser	0	1	1	2	0
22	Facebook	Profile Info	Marketing	Mobile Web Browser	0	0	11	10	0
23	Skype	Meetings/Interviews	Educational/Professional	Tablet App	0	0	4	1	0
24	Twitter	Politics	News Sharing	Computer Web Bowser	12	0	0	0	0
25	Instagram	Pictures	Brands Info	Tablet Web Browser	0	0	5	1	0
26	Twitter	Politics	News Sharing	Tablet Web Browser	11	0	3	1	0
27	Facebook	Profile Info	Marketing	Computer App	0	1	2	6	0
28	Instagram	Pictures	Brands Info	Mobile Web Browser	0	1	2	3	0
29	Facebook	Profile Info	Marketing	iPad App	0	16	0	0	1
30	YouTube	Videos	Entertainment	Computer Web Browser	0	17	4	1	0
31	Facebook	Profile Info	Marketing	iPad Web Browser	0	0	11	5	1
32	Instagram	Pictures	Brands Info	iPad App	0	0	1	3	0

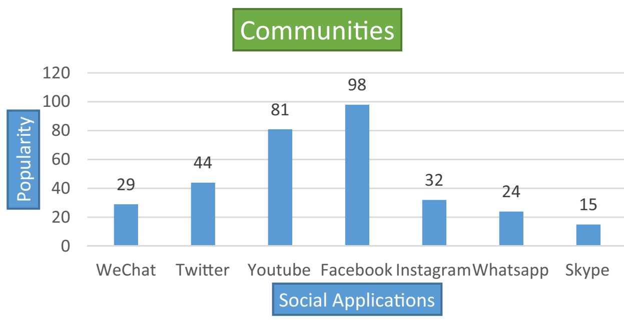


Fig. 8 Social applications popularity

The results of social applications popularity in different communities are almost the same that strengthen the knowledge extracted. All the three methods define communities with maximum usage of social applications that can help in forwarding information, advertisements, services and recommendations to a particular community through that

Table 14 Community assignment of 5NN graph

Label	Class
User1	0
User10	4
User100	1
User11	1
User12	0
User13	3
User14	2
User15	3
User16	0
User17	1
User18	1
User19	2
User2	0
User20	2
Total Score = 0.581	

Table 13. 5NN Adjacency matrix

	User1	User10	User100	User11	User12	User13	User14	User15	User16	User17	User18	User19	User2	User20
User1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
User10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User100	0	0	0	0	0	0	0	0	0	1	1	0	0	0
User11	0	0	0	0	0	0	0	1	0	1	0	0	0	0
User12	0	0	0	0	0	0	0	0	0	0	0	0	1	0
User13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
User16	0	0	0	0	0	0	0	0	0	0	0	0	1	0
User17	1	0	1	0	0	0	0	1	0	0	1	0	0	0
User18	0	0	1	0	0	0	0	0	0	1	0	0	0	0
User19	0	1	0	0	0	0	0	0	0	0	0	0	0	1
User2	0	0	0	0	1	0	0	0	1	0	0	0	0	0
User20	0	0	0	0	0	0	0	0	0	0	0	1	0	0

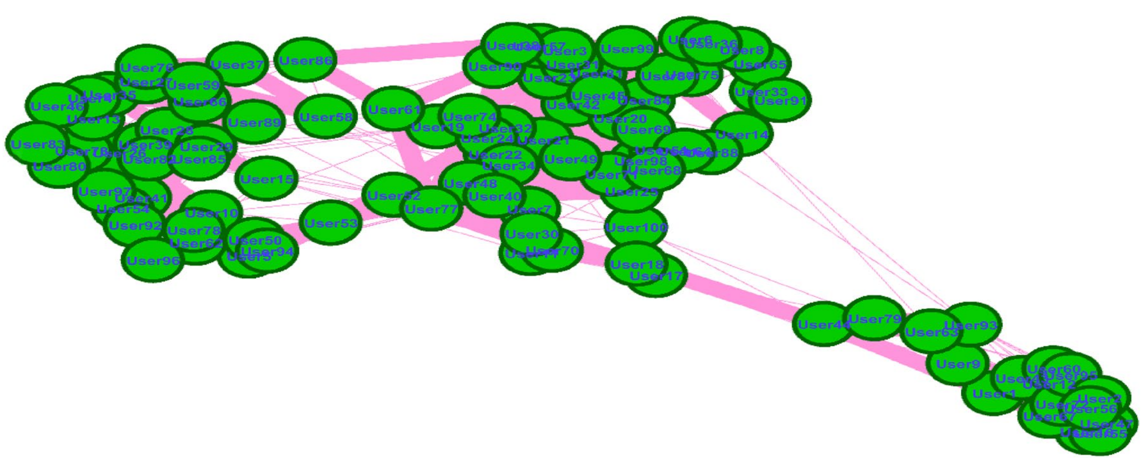


Fig. 9. 5NN graph

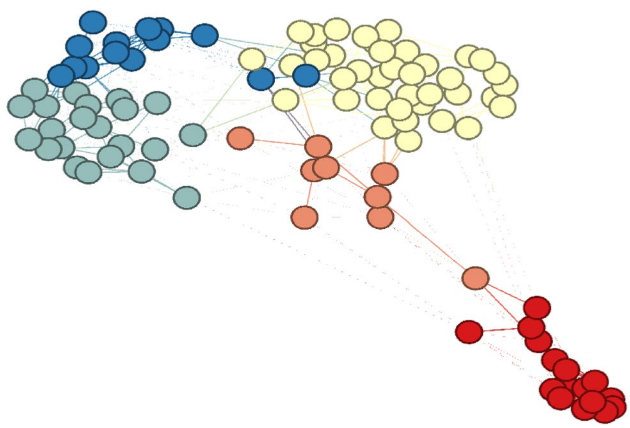


Fig. 10 Community assignment of users

application. In the context of comparison of *r*-neighborhood, kNN and gephi tool, the modularity score of *k*NN method has maximum value of 0.581 and gives better and clearer interpretation of communities detected in comparison with *r*-neighborhood and gephi tool. The *k*NN technique results in community one, three, four and five as Twitter, Facebook and YouTube communities; however, only community two is blur where the Facebook is observed as being most widely used application. The modularity score using gephi tool is 0.555 which is less compared to *k*NN as a result of which communities detected are not clear. The users of community one and two use Twitter and YouTube as major applications; however, the other three communities are not clear with any particular social application popularity which can be seen in Table 12. The modularity score using *r*-neighborhood is 0.554 and results in four different communities detected. Among the four communities, community two is defined as

Table 15 Social applications popularity using *k*NN

App Usage	App name	Functionality	Purpose	Medium of Access	0	1	2	3	4
1	WeChat	Transactions	Merchandise	Mobile App	1	3	6	0	0
2	Twitter	Politics	News Sharing	Mobile App	6	4	0	0	0
3	YouTube	Videos	Entertainment	Mobile App	0	0	3	1	2
4	Facebook	Profile Info	Marketing	Mobile App	0	0	11	0	1
5	Instagram	Pictures	Brands Info	Mobile App	0	0	4	0	0
6	WhatsApp	Friends/Family	Greetings/Personal	Mobile App	0	0	11	0	1
7	YouTube	Videos	Entertainment	Tablet Web Browser	0	0	4	15	0
8	YouTube	Videos	Entertainment	Tablet App	0	0	2	4	14
9	Skype	Meetings/Interviews	Educational/Professional	Mobile App	0	0	10	0	0
10	WhatsApp	Friends/Family	Greetings/Personal	Computer Web Browser	1	0	4	1	1
11	Facebook	Profile Info	Marketing	Tablet App	0	6	6	0	1
12	WhatsApp	Friends/Family	Greetings/Personal	Tablet App	1	3	0	1	0
13	WeChat	Transactions	Merchandise	Tablet App	0	0	0	4	2
14	WeChat	Transactions	Merchandise	Computer App	0	0	9	0	0
15	Instagram	Pictures	Brands Info	Tablet App	0	1	5	0	0
16	WeChat	Transactions	Merchandise	iPad App	1	0	4	0	0
17	Twitter	Politics	News Sharing	Tablet App	7	0	0	0	0
18	YouTube	Videos	Entertainment	Tablet App	0	0	1	9	4
19	Facebook	Profile Info	Marketing	Computer Web Browser	0	0	5	0	0
20	Instagram	Pictures	Brands Info	Computer Web Browser	0	0	6	0	0
21	Facebook	Profile Info	Marketing	Tablet Web Browser	0	0	3	1	0
22	Facebook	Profile Info	Marketing	Mobile Web Browser	0	5	15	0	1
23	Skype	Meetings/Interviews	Educational/Professional	Tablet App	1	0	4	0	0
24	Twitter	Politics	News Sharing	Computer Web Bowser	12	0	0	0	0
25	Instagram	Pictures	Brands Info	Tablet Web Browser	0	2	4	0	0
26	Twitter	Politics	News Sharing	Tablet Web Browser	12	0	3	0	0
27	Facebook	Profile Info	Marketing	Computer App	1	0	7	0	1
28	Instagram	Pictures	Brands Info	Mobile Web Browser	0	3	2	1	0
29	Facebook	Profile Info	Marketing	iPad App	0	0	1	12	4
30	YouTube	Videos	Entertainment	Computer Web Browser	0	2	4	12	4
31	Facebook	Profile Info	Marketing	iPad Web Browser	0	1	15	0	1
32	Instagram	Pictures	Brands Info	iPad App	0	0	4	0	0

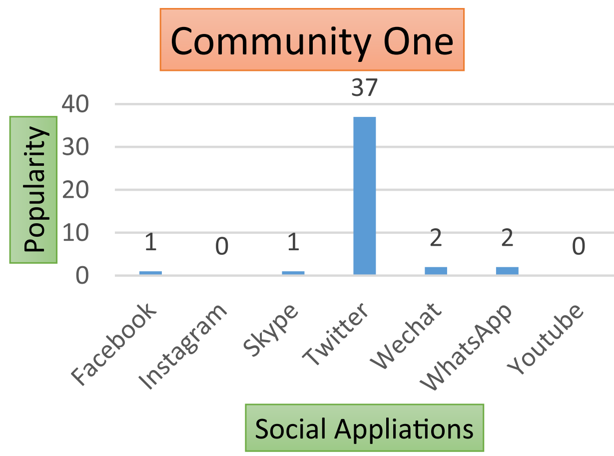


Fig. 11 Social applications popularity

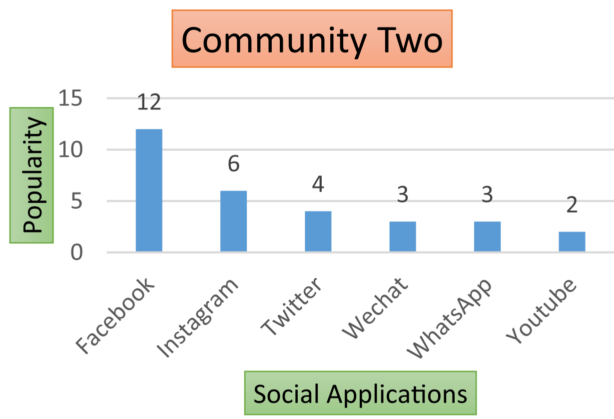


Fig. 12 Social applications popularity

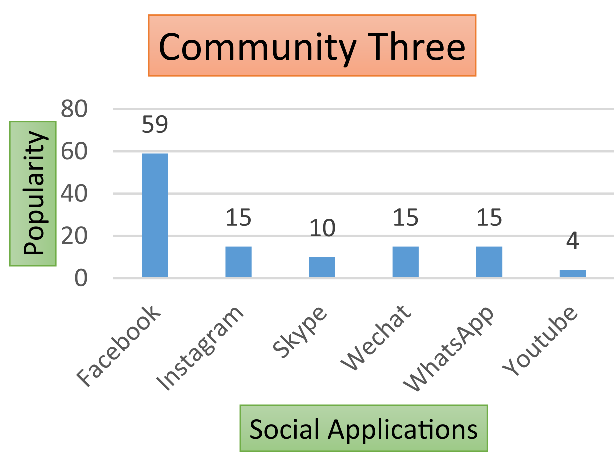


Fig. 13 Social applications popularity

Twitter community with other three communities not clearly defining the social application popularity in comparison with *k*NN.

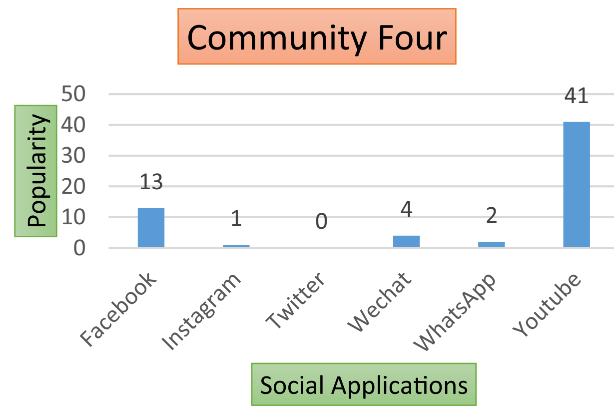


Fig. 14 Social applications popularity

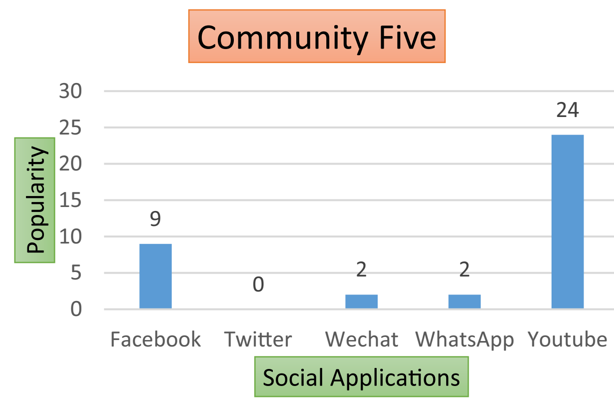


Fig. 15 Social applications popularity

6 Conclusions

This paper is related to extracting the knowledge by detecting communities in social applications network for social applications and using popularity of those applications in communities; different contents related to politics, business products, sports, movies and other services can be forwarded. Communities are detected using modularity maximization by divisive clustering. Once users are assigned to different communities, the popularity of social applications is determined by counting the number of times each social application is accessed by users in that community. It is observed that performing community detection using *r*-neighborhood, *k*-nearest neighbors and gephi gives similar results such as Twitter, YouTube and Facebook were the most common social applications used among different communities that reflects that people in such communities are mostly political, entertaining and social in nature. Targeted contents can be forwarded to these communities using those social applications. Also, one community may not be well aware of the other because of the difference in core functionality, for example, YouTube followers may not be much active about the political changes taking place in the world. So, by knowing the social application popularity in a community, the people can be informed like sharing news on YouTube, Facebook and advertising business products on Twitter, YouTube and forwarding the sports, movies contents on Facebook and Twitter.

References

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749

Agarwal G, Kempe D (2008) Modularity-maximizing graph communities via mathematical programming. Eur Phys J B 66(3):409–418

Aggarwal CC, Wang H (2010) A survey of clustering algorithms for graph data. In: Managing and mining graph data, pp 275–301. Springer, Boston, MA

Aghakhani S, Qabaja A, Alhajj R (2018) Integration of k-means clustering algorithm with network analysis for drug-target interactions network prediction. Int J Data Min Bioinform 20(3):185–212

Asmi K, Lotfi D, Abarda A (2021) The greedy coupled-seeds expansion method for the overlapping community detection in social networks. Computing, 1–19

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):P10008

Bollobás B (1998) Modern graph theory, volume 184 of. *Graduate texts in mathematics*

Cena F, Dattolo A, De Luca EW, Lops P, Plumbaum T, Vassileva J (2011) Semantic adaptive social web. In: International conference on user modeling, adaptation, and personalization, pp 176–180. Springer, Berlin, Heidelberg

Chen M, Kuzmin K, Szymanski BK (2014) Community detection via maximization of modularity and its variants. IEEE Trans Comput Soc Syst 1(1):46–65

Clauset A, Mark EJN, Cristopher M (2004) Finding community structure in very large networks. Phys Rev E 70(6):066111

Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27

Diestel R (2012) Graph theory, volume 173 of. *Graduate texts in mathematics*, 7

Duch J, Arenas A (2005) Community detection in complex networks using extremal optimization. Phys Rev E 72(2):027104

Foreman JW (2013) Data smart: Using data science to transform information into insight. Wiley, Hoboken

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Fortunato S, Barthelemy M (2007) Resolution limit in community detection. Proc Natl Acad Sci 104(1):36–41

Gibson D, Kumar R, Tomkins A (2005) Discovering large dense subgraphs in massive graphs. In: Proceedings of the 31st international conference on very large data bases, pp 721–732

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826

Good BH, De Montjoye YA, Clauset A (2010) Performance of modularity maximization in practical contexts. Phys Rev E 81(4):046106

Guimera R, Amaral LAN (2005a) Functional cartography of complex metabolic networks. Nature 433(7028):895–900

Guimera R, Amaral LAN (2005b) Cartography of complex networks: modules and universal roles. J Stat Mech Theory Exp 2005(02):P02001

Guo K, Huang X, Wu L, Chen Y (2021) Local community detection algorithm based on local modularity density. Appl Intell 1–16

Guo Y, Huang Z, Kong Y, Wang Q (2021) Modularity and mutual information in networks: two sides of the same coin. arXiv:2103.02542

He K, Li Y, Soundarajan S, Hopcroft JE (2018) Hidden community detection in social networks. Inf Sci 425:92–106

He QP, Wang J (2007) Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. IEEE Trans Semicond Manuf 20(4):345–354

Kolahdouzan M, Shahabi C (2004) Voronoi-based k nearest neighbor search for spatial network databases. In: Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, 840–851

Leicht EA, Newman ME (2008) Community structure in directed networks. Phys Rev Lett 100(11):118703

Li L, Darden TA, Weingberg CR, Levine AJ, Pedersen LG (2001) Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method. Comb Chem High Throughput Screen 4(8):727–739

Massen CP, Doye JP (2005) Identifying communities within energy landscapes. Phys Rev E 71(4):046101

Medus A, Acuña G, Dorso CO (2005) Detection of community structures in networks via global optimization. Physica A 358(2–4):593–604

Naderipour M, Zarandi MHF, Bastani S (2021) Fuzzy community detection on the basis of similarities in structural/attribute in large-scale social networks. Artif Intell Rev 1–35

Newman, M. E. (2013). Spectral methods for community detection and graph partitioning. *Physical Review E*, *88*(4), 042822.

Newman ME (2006a) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582

Newman ME (2006b) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74(3):036104

Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113

Newman ME (2004a) Analysis of weighted networks. Phys Rev E 70(5):056131

Newman ME (2004b) Fast algorithm for detecting community structure in networks. Phys Rev E 69(6):066133

Newman ME (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256

Panchal A (2021) Overlapping community detection in social networks

Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E 76(3):036106

Ruan J (2009) A fully automated method for discovering community structures in high dimensional data. In 2009 ninth IEEE international conference on data mining, 968–973. IEEE

Ruan J, Zhang W (2008) Identifying network communities with a high resolution. Phys Rev E 77(1):016104

Ruan J, Zhang W (2007) An efficient spectral algorithm for network community discovery and its applications to biological and social networks. In: Seventh IEEE international conference on data mining (ICDM 2007), 643–648. IEEE

Richardson T, Mucha PJ, Porter MA (2009) Spectral tripartitioning of networks. Phys Rev E 80(3):036111

Sales-Pardo M, Guimera R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. Proc Natl Acad Sci 104(39):15224–15229

Samanthula BK, Elmehdwi Y, Jiang W (2014) K-nearest neighbor classification over semantically secure encrypted relational data. IEEE Trans Knowl Data Eng 27(5):1261–1273

Shakarian P, Roos P, Callahan D, Kirk C (2013) Mining for geographically disperse communities in social networks by leveraging distance modularity. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 1402–1409

Sobolevsky S (2021) Recurrent graph neural network algorithm for unsupervised network community detection. arXiv:2103.02520

Wang M, Wang C, Yu JX, Zhang J (2015) Community detection in social networks: an in-depth benchmarking study with a procedure-oriented framework. Proc VLDB Endowm 8(10):998–1009

Wakita K, Tsurumi T (2007) Finding community structure in megascale social networks. In: Proceedings of the 16th international conference on World Wide Web, 1275–1276

Xu J, Luo L, Deng C, Huang H (2018) Bi-level distance metric learning for robust image recognition. In: Advances in neural information processing systems, 4198–4207

White S, Smyth P (2005) A spectral clustering approach to finding communities in graphs. In: Proceedings of the 2005 SIAM international conference on data mining, 274–285. Society for Industrial and Applied Mathematics

Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, Zhou ZH (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37

Zhao X, Liang J, Wang J (2021) A community detection algorithm based on graph compression for large-scale social networks. Inf Sci 551:358–372