Community Detection Algorithms for Cultural and Natural Heritage Data in Social Networks

Andreas Kanavos[1,2(✉)], Maria Trigka[2], Elias Dritsas[2], Gerasimos Vonitsanos[2], and Phivos Mylonas[3,4]

[1] Department of Regional Development, Ionian University, Lefkada, Greece
akanavos@ionio.gr
[2] Computer Engineering and Informatics Department, University of Patras, Patras, Greece
{trigka,dritsase,mvonitsanos}@ceid.upatras.gr
[3] Department of Informatics, Ionian University, Corfu, Greece
fmylonas@ionio.gr
[4] Institute of Entrepreneurship Development, Larissa, Greece

Abstract. In social network analysis, it is crucial to discover a community through the retrospective decomposition of a large social graph into easily interpretable subgraphs. Four major community discovery algorithms, namely the Breadth-First Search, the Louvain, the MaxToMin, and the Propinquity Dynamics, are implemented. Their correctness was functionally evaluated in the four most widely used graphs with vastly different characteristics and a dataset retrieved from Twitter regarding cultural and natural heritage data because this platform reflects public perception about historical events through means such as advanced storytelling in users timelines. The primary finding was that the Propinquity Dynamics algorithm outperforms the other algorithms in terms of NMI for most graphs. In contrast, this algorithm with the Louvain performs almost the same regarding modularity.

Keywords: Community detection · Cultural and natural heritage management · Graph mining · Modularity · NMI · Social networks

1 Introduction

To extract essential knowledge, researchers are led to process and analyze the excessively growing abundance of data [36]. The management of cultural information related to cultural and natural resources has become a crucial driver of the industry. The demand for collections of services that include guided visits to historical monuments, is continuously high worldwide in physical and digital markets. Today, the digital world plays a vital role in promoting cultural and natural content. Through its continuous expansion, the contribution to the preservation of the cultural and natural heritage and the discovery of new elements is achieved.

Published by Springer Nature Switzerland AG 2021
I. Maglogiannis et al. (Eds.): AIAI 2021 Workshops, IFIP AICT 628, pp. 395–406, 2021.
https://doi.org/10.1007/978-3-030-79157-5_32

In general, the term network can refer to any interconnected group or system that interacts in a complex way to serve a purpose. The needs of modern society and the development of technology, have contributed to creating various types of networks. The main categories of networks are business networks, financial networks, hotel networks, and social networks. Besides, the widespread use of web applications is responsible for creating even more social networks [29]. The particular interest of this study is networking, originating from Social Media. In social networks, there are differences between the entities and the types of relationships that are created. Nevertheless, we can gain insights into several social phenomena. Also, extracted knowledge such as detecting communities, can be used supportively in several other tasks.

The last decade, social networks are an integral part of individuals in every aspect of daily life. Through these networks, users share data, such as images, videos, music, movies, experiences, beliefs, and interests. Moreover, they have the opportunity to exchange views and be informed about issues that concern them. As the number of social networks and their users' increases, scientists in turn struggle to provide users with high-quality services. Users grouping can help to highlight interaction patterns and identify common attributes amongst individuals in real-world activities.

Social networks are usually made up of people who communicate with each other and belong to connected communities. The identification of these communities constitutes an elementary task in social network analysis. As the nature of these networks is complex and dynamic, communities are not easily identified. This is an open and often challenging issue and can be considered an optimization problem. The aim is to identify sets of nodes with more interconnections and simultaneously fewer intra-links with other nodes. Since community detection is usually categorized as an NP-Hard problem, these evolutionary algorithms have attracted massive attention in this field in recent years.

Studies indicate that in social networks, the distribution of the clustering coefficient follows the power law, and it decreases when the degree of nodes increases. The clustering coefficient is an essential factor in a graph that measures the tendency of the nodes to cluster together. This characteristic implies that nodes in social networks tend to form sub-graphs. In other words, social networks are formed by connected communities. Discovering these communities is essential in order to understand the structure of the network. Therefore, a lot of research has been conducted to render them efficient. Generally, a community is defined as a group of nodes with more links between themselves and fewer external links to other nodes.

In this paper, we aim to identify the types of relations between Twitter network users while also detect communities in data related to DIMOLEON project. Initially, four major community discovery algorithms, namely the Breadth-First Search, the Louvain, the MaxToMin, and the Propinquity Dynamics are evaluated in four most widely used graphs with vastly different characteristics in terms of two popular metrics, namely the Normalized Mutual Information (NMI) and Modularity. In following, a dataset retrieved from Twitter regarding cultural and natural heritage data was employed, where users evaluate the extracted communities from each algorithm. One major contribution of this work is the employment of Twitter derived content, as well as the proposal of a methodology that identifies users with similar behavior and features in terms of the platform towards its efficient utilization.

The rest of the paper is organized as follows. In Sect. 2 the main research in this area is reviewed. In Sect. 3, some preliminaries regarding metrics are introduced whereas in Sect. 4, the algorithms are described in detail. Implementation details are presented in Sect. 5 and evaluation with results are discussed in Sect. 6. Finally Sect. 7 concludes the paper.

2 Related Work

Social networks analysis is strongly related to graph clustering algorithms, and web searching algorithms; for a complete review of this area, one should consult [9,10,21,23]. A group of network nodes, where the links between the nodes are dense, is defined as a community [37]. It refers to groups of nodes in a network or graph with common properties in system operations. The field is related to the HITS algorithm [20], as well as the link analysis in the web with cornerstone the analysis of the significance of web pages in Google using the PageRank citation metric [28], and other numerous variants proposed in [22]. In HITS, two metrics are used: a web page as an information authority and a hub. In PageRank, a metric based on the level of the importance of the incoming links is employed.

Various algorithms have been introduced in the literature on community detection, [10,29,31,32]. HITS can be used to compute communities if used to explore non-principal eigenvectors. In bibliography, we come across the problem related with communities regarding graph partitioning. The proposed algorithms are mainly related to spectral distribution approaches that achieve the partition of objects through the eigenvectors of the matrices [27,35]. The technique of spectral partitioning was proposed in [8,33]. In [34], the utilization of hierarchical clustering for graph partitioning is brought forward.

An important method by which the initiative for further research was made was advocated in [11], where modularity was introduced along with a divisive method for the problem of community detection. Besides, some works [1,25,26] suggest an algorithm that selects the partition that will maximize modularity using it as a measure of partition quality. The modularity-based criterion is an important way of identifying community structures in networks, as the quality of identified communities is quantified. The criterion for determining the partition of communities is the dense internal connections within the communities and the few connections between them. Researchers have considered algorithms with different approaches based on the concept of modularity. In complex networks, some of these algorithms show low performance while regarding other algorithms, prior knowledge of the network is required [3,12,14,30].

Also, in [13–15], the concept of influence from the side of users to the side of networks is expanded and personality has been utilized as the key characteristic for identifying influential networks. The result is to create this type of communities in Twitter graphs using a modularity-based community detection algorithm, taking into account users' personalities. Additionally, the edges of the graph based on the user personality, are eliminated by the insertion of a pre-processing step. Moreover in [17–19], the behavior of users on an emotional level is enhanced by introducing a new methodology that effectively aids in community detection.

Similarly, there are several ways to assess the clustering quality, namely community coherence [24]. However, most of the existing coherence metrics are either prohibitively expensive, such as the maximum distance between vertices, or are prone to outliers, such as the diameter-based metrics [6,7]. In [5,16] are described implementations of established community discovery algorithms, namely the CNM over Neo4j, the Walktrap, the Louvain, and the Edge Betweeness or Newman-Girvan. To evaluate these algorithms efficiently, we rely mainly on how the graph partitioning obtained by such an algorithm translates into the functional Twitter domain and not on other structural criteria.

3 Preliminaries

In this section, some details regarding the structure of the community along with the metrics of centrality and modularity will be introduced.

3.1 Community Structure

Typically, a network is considered as an undirected graph $G = (V, E)$, where $V = \{v_i | i \in [1, 2, \ldots, N]\}$ is the set of vertices and $E = \{e_{ij} | i, j \in [1, 2, \ldots, N]\}$ is the set of edges. Let us assume that $deg(v)$ is the degree of v in the graph G.

3.2 Basic Metrics

Centrality. One indicator that is widely used for network data is centrality measures. They reflect the prominence of a unit in different substantive settings, such as status, visibility, structural power, or prestige [4]. These measures can be distinguished in the following:

1. The *Degree Centrality* of a node v indicates the number of nodes, which are directly linked to this node. It is defined as

$$C_{\mathrm{D}}(v) = deg(v) \tag{1}$$

2. The *Closeness Centrality* is related to the closeness of a vertex v and indicates how close a node is to all other nodes in the network. It is defined as

$$C_C(u) = \sum_{v \in V} d(u, v) \tag{2}$$

where $d(u, v)$ is the number of edges in a shortest path connecting the vertices u and v, known as geodesic distance.

3. The *Betweenness Centrality* indicates how much a vertex is in-between others. This metric is computed as the number of shortest paths between any couple of nodes in the graph containing target node v. It is defined as

$$C_{\mathrm{B}}(v) = \sum_{y \neq z \in N} \frac{p_{st}(v)}{p_{st}} \tag{3}$$

where $p_{st}(v)$ denotes the number of shortest paths between s and t containing v, and p_{st} denotes the number of all shortest paths between s and t in the network.

Modularity. It is a measure that captures the structure of the network and is used to evaluate the strength of a network division into communities [40]. In general, it is defined as

$$Q(V) = \frac{1}{2M} \sum_{v_i, v_j \in V} \left(A_{i,j} - \frac{deg(v_i)deg(v_j)}{2M} \right) \delta_{c_i, c_j}, \tag{4}$$

where $M = \frac{1}{2} \sum_i deg(v_i)$ and $\boldsymbol{A} = [A_{ij}] \in \mathbb{N}^{N \times N}$ is the adjacency matrix of given graph having values equal to 0 and 1; $A_{ij} = 1$ when two nodes are connected with an edge with $e_{ij} \in E$. Additionally, $\frac{deg(v_i)deg(v_j)}{2M}$ captures the expected number of edges between nodes v_i and v_j when edges are randomly distributed. If $c_i = c_j$ then $\delta_{c_i, c_j} = 1$, or if $c_i \neq c_j$ then $\delta_{c_i, c_j} = 0$, where c_i denotes that v_i belongs to the community c. The higher values of modularity indicate better quality of community detection.

It should be noted that although Louvain algorithm is based on the modularity optimization for detecting communities in networks, a limitation of this type of methods is their inability to detect small communities.

4 Community Detection Algorithms

In this section, the community detection algorithms considered in the experimental evaluation, are properly analyzed.

4.1 Louvain Algorithm

This algorithm constitutes a greedy approach of modularity maximization. As mentioned above, the strength by which a network is divided into communities is measured through modularity. Connections between nodes within communities on social networks with high modularity are dense, while between nodes of different communities are sparse.

The algorithm starts by assigning each node to a separate community and evolves with the movement of nodes to neighbouring communities to improve modularity. Precisely, it consists of two phases, the *Modularity Optimization* and the *Folding*. In the first phase, the first step assumes that on the original graph G, each node $v \in V$ is assigned to a separate community, whereas in the second step, each node v is placed in the community where the network modularity is maximized; this step is repeated until there are no moves. In following, in the second phase, the algorithm checks whether the new modularity is higher than the previous one and if so, the graph G is replaced by the one formed between the communities and returns to second-last step, otherwise it terminates.

During the folding phase, a new graph G' is created, where each node of this new graph represents a community of G. Each pair of nodes $v'_i, v'_j \in G'$ are linked if the corresponding communities of G have edges between them. The weight of the edge (v'_i, v'_j) is equal to the sum of the weights between communities. In addition, the sum of the inner edges of the communities of G is represented by a self-loop edge at node v' of G'.

Finally, as this method generates a hierarchical structure of communities, it has complexity equal to $O(Nlog_2N)$, where N is the number of nodes in the network.

4.2 Propinquity Dynamics

The term propinquity comes from sociology and refers to the physical or psychological proximity among individuals. In community detection, this specific term captures the probability of two nodes to be involved in a coherent community. The Propinquity Dynamics (PD) algorithm does not require prior knowledge of the communities' structure and obtains the propinquity information from the topology of the graph from a self-organized dynamic process [39].

Through several rounds of mutual reinforcement between the topology and propinquity, the community structures are naturally emerged. To achieve the highest efficiency, the propinquity is calculated in an incremental way. Nodes that belong to more than one communities (e.g., there are overlaps) can be determined by post-processing.

The PD algorithm can efficiently discover communities from very large scale graph data with complexity equal to $O(k|V|)$ in sparse graphs, where V and k are the number of graph nodes and the iteration count, respectively. Another advantage of the algorithm is that it emphasizes on scaling without loss of community quality.

Coherent Neighborhood Propinquity: Given that the diameter in a coherent graph is not greater than 2, the propinquity considers only the local 2-hop neighborhood, assuming the resulting communities are cohesive. Based on this, the number of common neighbours of a node pair is an important criterion for defining their neighborhood. Therefore, the total connectivity of the local neighborhood must be taken into account for the evaluation of the neighborhood.

Propinquity Calculation: The propinquity calculation can be implemented by finding for each pair of nodes the intersection of their neighbours and then counting the edges that connect the common neighbours. The complexity of this calculation is approximately $O((|V| + |E|)|E|)$, where E is the number of edges.

4.3 MaxToMin Algorithm

PD creates a graph topology with many communities, whereas the Breadth-First Search (BFS) identifies communities with the limitation that there may exist nodes that do not belong exclusively to one community. To address the above weaknesses, the algorithm MaxToMin has been proposed.

Initially, MaxToMin randomly starts from the node that is connected to the "strongest" edge (where the neighborhood size is considered as the corresponding weight) in the graph, i.e. the edge having the highest weight. MaxToMin will then try to access the nodes with the strongest edges that are neighbours of that node and attach them in the same community.

The algorithm moves from strong to less strong edges, without being able to move from a weak connection to a sounder one. An iteration of the algorithm achieves the finding of a community, and terminates when no other weak edges associated with the accessed are considered. In case a node can be accessed by its L independent executions of the algorithm, then this node is assigned to the corresponding L communities and is considered as an overlap to these communities.

5 Implementation

Our aim is to evaluate the performance of the Breadth-First Search, the Louvain, the MaxToMin and the Propinquity Dynamics algorithms in terms of Normalized Mutual Information (NMI) and Modularity. NMI is considered a normalization of the Mutual Information (MI) score; it takes values from 0, which refers to no mutual information, to 1, which denotes perfect correlation. Connections between nodes within communities with high modularity are dense, while on the other hand, connections between nodes of different communities are sparse.

5.1 Graphs

Initially, we chose four most widely used graphs in order to utilize our experimental evaluation, namely Zachary Karate Club, Dolphin Network, Polbooks and American College Football [2,38]. A synopsis of these networks is presented in Table 1 in ascending order, according to their number of vertices.

Initially, Zachary Karate Club dataset is considered a social network of friendships between 34 members of a karate club at a US university in the 1970s. A conflict between president and instructor led to a partition of the club into two organizations of nearly equal size. Furthermore, Dolphin Network is an undirected social network of frequent relationships between 62 dolphins in a community living off Doubtful Sound in New Zealand. The Polbooks dataset consists of a directed network of hyperlinks between weblogs on US politics, recorded in 2005. This network is divided according to blog political orientation, that is either conservative or liberal. Finally, American College Football dataset is considered a network of American football games between division colleges during the regular season fall of 2000; it has 115 teams separated into 12 divisions, where each division consists of 8 to 12 teams.

Table 1. Graphs synopsis

Name	Description	Vertices	Edges
Karate	Zachary's karate club	34	78
Dolphins	Dolphin social network	62	159
Polbooks	Books about US Politics	105	441
Football	American college football	115	613

5.2 Twitter Dataset

An approach based on specific topics was used for collecting tweets via a keyword search query for the generation of our test dataset. Keywords that are relevant to cultural and natural heritage in the domain of Greece and in conjunction to DIMOLEON project were downloaded. These keywords are related to different heritages, specific tourist destinations and activities. The filtered dataset resulted in $5,000$ tweets from 01/02/2021 to 28/02/2021 as we have only kept tweets posted in English language.

6 Evaluation

In this section, the performance of the four community detection algorithms in terms of the four graphs and the retrieved dataset from Twitter, were evaluated.

6.1 Graphs

Table 2 presents the results of the performance of MaxToMin and Breadth-First Search in terms of NMI metric, the number of iterations, and vertices without community for the four different graphs. MaxToMin algorithm outperforms Breadth-First algorithm regarding the NMI in all graphs while the number of iterations is relatively low. Also, there are no vertices that do not belong to any community regarding the Karate and Dolphins graphs.

Table 2. Normalized mutual information for MaxToMin and Breadth-First Search algorithms

	MaxToMin			Breadth-First Search		
Graph	NMI	Iterations	Vertices without community	NMI	Iterations	Vertices without community
Karate	0.9240	1	0	0.3098	3	22
Dolphins	0.5989	1	0	0.4687	22	2
Polbooks	0.5778	4	15	0.4947	4	15
Football	0.9268	4	2	0.9099	4	8

The results of the other two algorithms in terms of NMI and Modularity metrics are depicted in Table 3. PD algorithm has higher values of NMI, and in some graphs, the difference is over 20%; specifically, in Dolphins graph, the difference is higher than 30%. The two algorithms achieve almost the same performance regarding modularity, except the Karate graph where Louvain outperforms PD by almost 3%.

Table 3. Normalized mutual information and modularity for Louvain and propinquity dynamics algorithms

	Normalized mutual information		Modularity	
Graph	Louvain	Propinquity dynamics	Louvain	Propinquity dynamics
Karate	0.6994	0.9240	0.4020	0.3714
Dolphins	0.6324	0.9428	0.5171	0.5143
Polbooks	0.5537	0.6383	0.5220	0.5124
Football	0.9111	0.9268	0.5811	0.6010

6.2 User Evaluation

For user evaluation of the downloaded Twitter dataset, we organized an online survey and asked students associated with the University of Patras to evaluate the extracted communities from each algorithm.

Users were presented with the communities wherein each community, the corresponding user with their tweets, was considered. After browsing through the dataset, users were asked to choose whether each community contains users with similar features. According to their personal beliefs, we examine three options: dense community, sparse community, and in-between.

The results are presented in the following Table 4 where users evaluate the communities discovered from the four algorithms, and the percentages of communities are depicted. As in previous experiments utilized for the four graphs, Louvain and Propinquity Dynamics have the highest number of dense communities. All four algorithms perform almost the same regarding the number of sparse communities. We have to consider that more features can further improve the performance of the algorithms, e.g., Twitter metrics, like the number of followers and the number of tweets per user.

It is worth mentioning the fact that the majority of hashtags regarding this dataset consists of terms like #Acropolis, #Ancientathens, #Ancientgreece, #Epidaurus, #HerodesAtticus, #Knossos, #Mycenae, #Parthenon, #Thermopylae as well as #Vergina.

Table 4. Percentages of Communities with similar Nodes

Algorithm	Dense	Sparse	In-between
Breadth-First Search	22	23	55
Louvain	32	23	45
MaxToMin	27	22	51
Propinquity dynamics	33	20	47

7 Conclusions and Future Work

In this paper, we aim to understand the community detection problem in terms of cultural and natural heritage data. More to the point, four major community discovery algorithms, namely the Breadth-First Search, the Louvain, the MaxToMin, and the Propinquity Dynamics are evaluated in several popular graphs having different characteristics in terms of two well-known metrics, namely the Normalized Mutual Information (NMI) and Modularity. Furthermore, a dataset retrieved from Twitter regarding cultural heritage data was taken into consideration. In following, users evaluated the extracted communities from each algorithm by rating the density of each community. A methodology, tailored to the prerequisites of the digital culture domain and generic for summarizing the plethora of cultural items, is proposed as it supports sufficient multi-modal clustering and semantic annotations for such cultural items.

As future work, it is in our keen interest to investigate the scalability problems that are considered when dealing with more extensive graphs. Withal, we aim to conduct an additional series of experiments using other subjects in order to identify the parameters that influence the outcomes of the algorithms at a more refined granularity level. The dimension of time in social network analysis can gain potential due to the dynamic nature of these networks; that is, the communities evolution over time should be measured in terms of functionality and size. Ultimately, the incorporation of advanced data structures that could offer more efficient solutions can be considered a future aspect to be tackled within the current DIMOLEON project research activities.

Acknowledgement. This research has been co-financed by the European Union and Greek national funds through the Competitiveness, Entrepreneurship and Innovation Operational Programme, under the Call “Research - Create - Innovate”, project title: “Using Digital Tools and Applications for Outdoor Alternative Tourism Operators - DIMOLEON”, project code: T2EDK-03168, MIS code: 5069920.

References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exper. **2008**(10), P10008 (2008)

2. Chen, Y., Zhao, P., Li, P., Zhang, K., Zhang, J.: Finding communities by their centers. Sci. Rep. **6**(1), 1–8 (2016)
3. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**(6), 066111 (2004)
4. Das, K., Samanta, S., Pal, M.: Study on centrality measures in social networks: a survey. Soc. Netw. Analy. Min. **8**(1), 13 (2018)
5. Drakopoulos, G., Gourgaris, P., Kanavos, A.: Graph communities in Neo4j. Evolving Syst. **11**(3), 397–407 (2020)
6. Drakopoulos, G., Kanavos, A., Makris, C., Megalooikonomou, V.: On converting community detection algorithms for fuzzy graphs in Neo4j. In: International Workshop on Combinations of Intelligent Methods and Applications (2015)
7. Drakopoulos, G., Kanavos, A., Makris, C., Megalooikonomou, V.: Comparing algorithmic principles for fuzzy graph communities over Neo4j. In: Hatzilygeroudis, I., Palade, V., Prentzas, J. (eds.) Advances in Combining Intelligent Methods. ISRL, vol. 116, pp. 47–73. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-46200-4_3
8. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Math. J. **23**(2), 298–305 (1973)
9. Flake, G.W., Lawrence, S., Giles, C.L.: Efficient identification of web communities. In: 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 150–160 (2000)
10. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
11. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
12. Jia, G., et al.: Community detection in social and biological networks using differential evolution. In: Hamadi, Y., Schoenauer, M. (eds.) LION 2012. LNCS, pp. 71–85. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34413-8_6
13. Kafeza, E., Kanavos, A., Makris, C., Chiu, D.K.W.: Identifying personality-based communities in social networks. Adv. Conceptual Model. **8697**, 7–13 (2013)
14. Kafeza, E., Kanavos, A., Makris, C., Pispirigos, G., Vikatos, P.: T-PCCE: twitter personality based communicative communities extraction system for big data. IEEE Trans. Knowl. Data Eng. **32**(8), 1625–1638 (2020)
15. Kafeza, E., Kanavos, A., Makris, C., Vikatos, P.: T-PICE: twitter personality based influential communities extraction system. In: IEEE International Congress on Big Data, pp. 212–219 (2014)
16. Kanavos, A., Drakopoulos, G., Tsakalidis, A.K.: Graph community discovery algorithms in neo4j with a regularization-based evaluation metric. In: 13th International Conference on Web Information Systems and Technologies (WEBIST), pp. 403–410 (2017)
17. Kanavos, A., Perikos, I.: Towards detecting emotional communities in twitter. In: 9th IEEE International Conference on Research Challenges in Information Science (RCIS), pp. 524–525 (2015)
18. Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.K.: Integrating user's emotional behavior for community detection in social networks. In: 12th International Conference on Web Information Systems and Technologies (WEBIST), pp. 355–362 (2016)
19. Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.K.: Emotional community detection in social networks. Comput. Electr. Eng. **65**, 449–460 (2018)
20. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5), 604–632 (1999)

21. Lancichinetti, A., Fortunato, S.: Community detection algorithms: a comparative analysis. Phys. Rev. E **80**(5), 049902 (2009)
22. Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton (2006)
23. Leskovec, J., Lang, K.J., Mahoney, M.W.: Empirical comparison of algorithms for network community detection. In: 19th International Conference on World Wide Web (WWW), pp. 631–640 (2010)
24. Mylonas, P., Wallace, M., Kollias, S.: Using k-nearest neighbor and feature selection as an improvement to hierarchical clustering. In: Vouros, G.A., Panayiotopoulos, T. (eds.) SETN 2004. LNCS (LNAI), vol. 3025, pp. 191–200. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24674-9_21
25. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**(6), 066133 (2004)
26. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. **103**(23), 8577–8582 (2006)
27. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems (NIPS), pp. 849–856 (2001)
28. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report (1999)
29. Papadopoulos, S., Kompatsiaris, Y., Vakali, A., Spyridonos, P.: Community detection in social media. Data Min. Knowl. Discov. **24**(3), 515–554 (2012)
30. Pizzuti, C.: GA-Net: a genetic algorithm for community detection in social networks. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) PPSN 2008. LNCS, vol. 5199, pp. 1081–1090. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87700-4_107
31. Plantié, M., Crampes, M.: Survey on social community detection. In: Social Media Retrieval, pp. 65–85. Computer Communications and Networks (2013)
32. Porter, M.A., Onnela, J., Mucha, P.J.: Communities in networks. CoRR (2009)
33. Pothen, A., Simon, H.D., Liu, K.P.P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. **11**(3), 430–452 (1990)
34. Scott, J.: Social network analysis. Sociology **22**(1), 109–127 (1988)
35. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) **22**(8), 888–905 (2000)
36. Vallet, D., Fernández, M., Castells, P., Mylonas, P., Avrithis, Y.: A contextual personalization approach based on ontological knowledge. In: 2nd International Workshop on Contexts and Ontologies: Theory, Practice and Applications (C&O-2006) Collocated with the 17th European Conference on Artificial Intelligence (ECAI-2006), vol. 210 (2006)
37. Yang, B., Liu, D., Liu, J.: Discovering communities from social networks: methodologies and applications. In: Handbook of Social Network Technologies and Applications, pp. 331–346 (2010)
38. Yin, C., Zhu, S., Chen, H., Zhang, B., David, B.: A method for community detection of complex networks based on hierarchical clustering. Int. J. Distrib. Sens. Netw. **11**, 849140:1–849140:9 (2015)
39. Zhang, Y., Wang, J., Wang, Y., Zhou, L.: Parallel community detection on large networks with propinquity dynamics. In: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 997–1006 (2009)
40. Zhu, J., Chen, B., Zeng, Y.: Community detection based on modularity and k-plexes. Inf. Sci. **513**, 127–142 (2020)