Online at: https://researchtrendsjournal.com ISSN No: 2584-282X

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 1; Issue 1; 2023; Page No. 250-254

Received: 02-10-2023
Accepted: 06-11-2023

Community detection and their performance, algorithms in social networks by using graph theory

[1]Satish Chandra Chouhan, [2]Dr. Savita Tiwari and [3]Dr. Ashfaq Ur Rahman

[1]Research Scholar, Madhyanchal Professional University, Bhopal, Madhya Pradesh, India
[2, 3]Department of Mathematics, Madhyanchal Professional University, Bhopal, Madhya Pradesh, India

DOI: https://doi.org/10.5281/zenodo.13235147

Corresponding Author: Satish Chandra Chouhan

Abstract

Researchers in the fields of social networks and graph theory are currently focused on deducing community structures from patterns of connections between people and websites, thanks to the widespread use of instant messaging systems and the dramatic change in the ease of content posting. In this paper, we introduce the ComTector (Com-munity Detector) algorithm, an improvement over previous methods for detecting communities in massive social networks. Evaluation of the spreading impact, description of the node's location, and identification of interaction centralities may all be accomplished using the identification methods of spreading influence nodes.

Keywords: Community detection, performance, algorithms, social networks, graph theory

Introduction

A social network is a kind of social structure that consists of a group of people or organizations and the relationships between them. With the social network approach comes a toolbox full of ideas and methodologies for dissecting the inner workings of whole social entities. Social network analysis is used to find important entities, analyze network dynamics, and find patterns both locally and globally in the study of these systems.

To connect with others who share your interests, hobbies, history, or real-life connections, you may join a social networking service, which is also known as a social networking site or SNS. A social network service includes a profile, which is a representation of the user, as well as their social connections and other related services. A person may sign up for a social network site, which is a web-based service that lets them make a public profile, add other users to their network, and see and interact with their connections. The majority of social media platforms allow users to communicate with one another using online mediums like email and instant chat.

In a social network, people or organizations are referred to as "nodes" and are bound together by various forms of interdependence, such friendship or family ties. Thus, technically speaking, we can say that social media is just a collection of things that are linked to each other in some way.

As an example, the set S= {social elements: social elements are connected}

Graph theory provides a solid foundation for representing social networks. One mathematical tool for representing the interconnections between various entities is the graph. The building blocks of a graph are the "nodes," or items, and the "edges," or connections, between them.

First, let's pretend that a finite number of persons have been given common or comparable qualities. The following is one possible representation of their social network:

Fig 1: An Example to Geometrical Representation of Social Network

Second Example: Let's pretend that a graph representing the communication among a countable society's members looks like this: each vertex represents a person, and each edge represents a connection in the communication chain between two of those vertices.

Everyone can access every other member of the network, either directly or via other members, since the graph is linked. It should be minimally linked as the graph is a tree.

Literature Review

Chandra Prakash, Chetan Agrawa, Pawan Meena (2022) [1] In light of the present situation, some observers are starting to pay more attention to online media. Online media platforms have the capability to create vast quantities of data on the client side. There are a plethora of mining tasks supplied by online media mining services, which helps businesses keep up with the data their users generate. The consumer may build their own local area at their leisure from among several long-distance interpersonal conversation places. The material found on the World Wide Web currently makes up a significant chunk of the whole virtual universe. The reason for this is the high number of users who actively engage in various online groups and keep their own accounts. One defining feature of digital media is the ease with which users may find new communities. This step is similar to the one in data mining called clustering. On the other hand, online media mining may make use of a different identification technique: local area by impact. Much effort has gone into local area recognition using this as a foundation, but it remains crucial. Recognizing locals who are making effective use of Leverage is the main goal of the event. Community detection also cares a lot about scalability and community quality. When compared against competing algorithms, several of these methods outperform the competition and scale well in large networks. With the use of Twitter's social data, we compared and contrasted the algorithms. Hence, the algorithms' scalability in the huge network as measured by the evaluation parameters has been shown. The fact that we have thoroughly tested the algorithm on a large social network is one factor that distinguishes our thesis from others.

Dong Wang *et al*. (2017) [2] In order to develop effective business strategies and online marketing campaigns, companies are increasingly turning to social networking sites (SNS) as a platform and channel for gathering user attitudes about goods and services. Capturing these feelings effectively for business decision support remains a huge problem on SNS due to the enormous number of users and complicated user connections. To describe a group of people who have strong feelings for the same product or service and are very close to one another, we use the term "Sentiment Community" in this research. Businesses would benefit greatly from finding these emotion groups in order to segment their customers and target them with marketing. We provide two approaches to find sentiment communities using optimization models of semi-definite programming (SDP), which take sentiments and connections into consideration. The suggested techniques performed well in our experimental testing. This study paves the way for further research on the impact of social media sentiment on corporate decision-making.

William P. Fox *et al*. (2014) [3] A social network analysis will provide a plethora of metrics and measurements. The result allows one to acquire a rank ordering of the nodes based on each of these metrics and measurements. Decisions on fooling or disrupting a certain network could be informed by this data. If every metric points to the same node as the important or key node, then everything is OK. When several metrics point to distinct critical nodes, what then? In this study, we will examine two approaches that may be used to pinpoint the important nodes or actors in any network. To determine which nodes had the most impact based on the inputs made by the decision makers, we use two methods to examine these outputs. We maximize the nodes' efficiency across all criteria using data envelopment analysis, and we take into account subjective and objective inputs using pairwise comparison matrices in the analytical hierarchy process (AHP). We use two popular networks found in the literature-the kite network and the information flow network-to demonstrate our findings. We go over the methodology' applicability to some elementary sensitivity analysis. When it comes to weighting the criteria according to the inputs of decision makers or the network structure, we found the AHP technique to be the most adaptable.

Shashank Sheshar Singh *et al*. (2022) [4] Examining the function and significance of nodes and edges is necessary for comprehending and contrasting various social networks. A number of social network analysis tasks benefit from centrality metrics, which quantify the significance of nodes and edges in a network. Analysts are able to do a plethora of related activities, such as investigating network structure, identifying prominent users, and predicting future connections, by using these centrality metrics. To foretell the network's future connections, this chapter details an investigation of centrality measurements. Additionally, we assess the efficacy of several centrality measures related to the link prediction issue on actual social networks using four distinct metrics: recall, precision, area under the curve (AUC), and AUPR.

From biological networks to social networks to recommender systems and beyond, link prediction is essential for analyzing the development of networks by detecting future connections inside complicated networks. In order to calculate similarity scores for anticipating relationships in these networks, researchers have suggested a number of centrality metrics, including degree, clustering coefficient, betweenness, and closeness centralities. These measurements of centrality make use of the network nodes' local and global knowledge. Using average centrality measures based on local and global centralities, such as Similarity based on Average Degree (SACD), Similarity based on Average Betweenness (SACB), Similarity based on Average Closeness (SACC), and Similarity based on Average Clustering Coefficient (SACCC), we offer a novel approach to link prediction using similarity score in this study. Our method included finding the average graph centrality, similarity scores via shared neighbors, and centrality ratings for each node individually. To find nodes with above-average centrality, we then applied centrality ratings to these shared neighbors. We evaluated our method by comparing the suggested metrics to some of the most popular and up-to-date local similarity-based link prediction metrics, such as common neighbor and the Centrality-based Parameterized Algorithm (CCPA), resource allocation, Adamic-Adar, common neighbors, and keyword network link prediction (KNLP). Four real-world datasets were used for our investigations. Significant gains are shown by the suggested similarity ratings that are based on average centralities. When compared to previous local similarity measures, we found an average improvement of 24% in terms of AUROC, and an improvement of 31% over more recent measures. The Area Under Precision-Recall (AUPR) also showed an improvement of 49% and 51%, respectively, when compared to the current and most recent measurements. The suggested method's improved performance is shown by our extensive trials.

Community detection in social networks

A directed graph, on the other hand, has connections that point in the same direction (relationships are bidirectional). Another option is to have the graph be either unweighted or weighted, where each connection has a cost or weight. The adjacency matrix and a basic undirected unweighted graph are shown in Figure 2. An adjacency matrix $A \in R\ n \times n$ may be used to describe graph G (V, E). In the case where i and j are integers from 1 to n, A_{ij} equals 1 if and only if v_i and v_j are connected, and 0 otherwise. N(i) is the set of all nearby nodes from vertex v_i, denoted as {j, v_j ∈V and {v_i, v_j} ∈ E}. The number of edges connecting node v_i to other nodes, denoted as ki= |N(i)|, is the degree of node v_i.

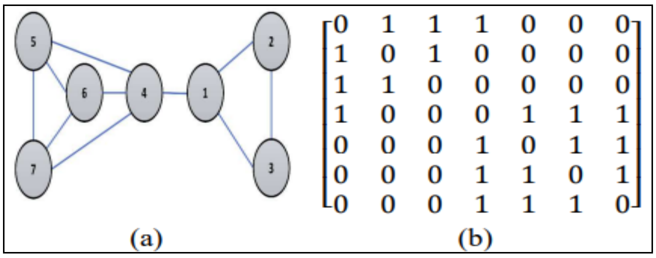


Fig 2: An Undirected Graph G = (V, E), with V= {1,2,3,4,5,6,7} and E= {(1,2), (1,3), (1,4), (2,3), (4,5), (4,6), (4,7), (5,6), (5,7), (6,7)}. (a)The Graphical Representation and (b) the Associated Adjacency Matrix. N (v4) = {v5, v6, v7, v1} and k4=4.

The goal of community discovery in a network G = (V, E) is to find k subsets of nodes that are represented by: C= {C1, C2…, Ck} and $V = \cup^k C i$. The community's utilized definition must be satisfied by this split. So, a community may be seen as a collection of nodes that are similar to one another and share certain properties, or as a highly induced subgraph where the number of internal linkages is larger than the number of exterior ones.

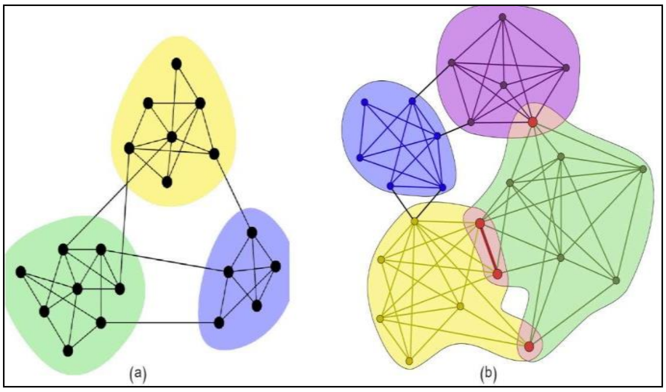
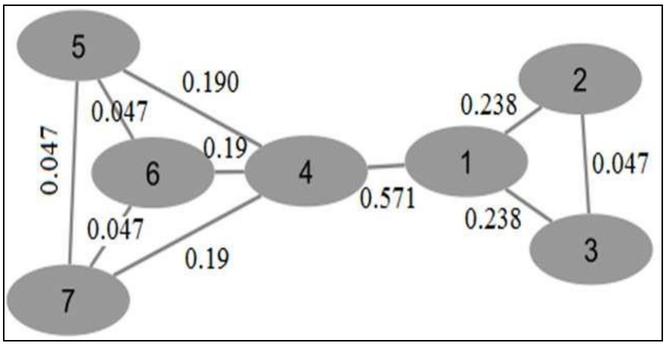


Fig 3: (a) Non-overlapping Communities. (b) Overlapping Communities

There are now two types of algorithms: disjoint and overlapping. Using separate modes, $C_i \cap C_j = \emptyset\ \forall\ (i, j) \in [1, k]$. Two communities, C_i and C_j, may be located in overlapping mode in such a way that $C_i \cap C_j \neq \emptyset$.

Algorithms For Community Detection

Radicchi *et al*. sought to further investigate the local measure of nodes and present a new algorithm based on removing the edge with the smallest edge clustering defined in the following, since the Girvan-Newman algorithm is computationally expensive and requires the repeated evaluation of edge betweenness for each edge in the entire graph:

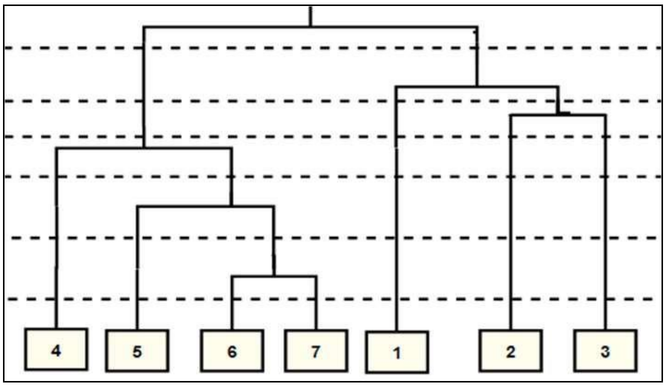


(a) After establishing the betweenness centralities of all edges in the graph, we may proceed to eliminate the edge (1,4) with the highest BC and then recalculate the BC for every other edge.

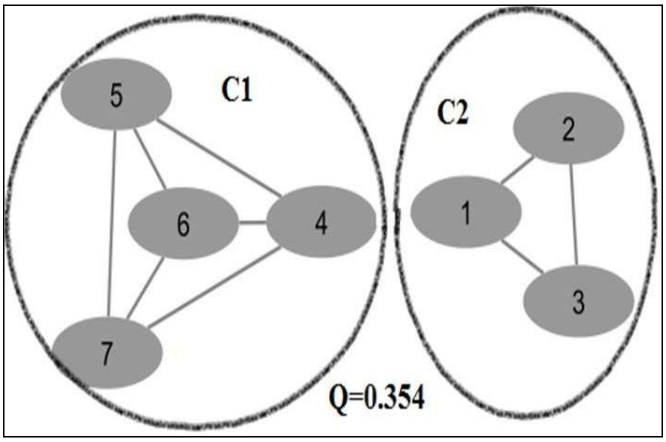
Table 1: Removed edge and Modularity

Removed edge	Modularity
(1,4)	0.354
(1,2)	0.354
(1,3)	0.214
(2,3)	0.134
(4,5)	0.134
(4,6)	0.134
(4,7)	0.0149
(5,6)	0.0149
(5,7)	-0.095
(6,7)	-0,15

(b) The modularity is achieved by carefully deleting each edge that has the highest BC.



(c) dendrogram representing community division, where each level is sliced to produce a collection of clusters.



(d) The most modular community structure is the one that results from removing edges (1,4)

The Girvan Newman Algorithm A model network with betweenness centralities at every edge; (b) a dendrogram showing the results; and (c) an immunity structure shown by each level. The most modular community structure (d)
Clustering of edges is defined as the ratio of the number of triangles to which an edge actually belongs to all the triangles that may include it. For the edge that goes from i to j, the edge-clustering coefficient is:

$$C_{i,j}^{3} = \frac{z_{i,j}^{3} + 1}{\min\left[(k_i - 1), (k_j - 1)\right]}$$

$z_{i,j}^{3}$ a total of all three-sided triangles constructed on edges e_{ij}

If the removal of edges does not divide the (sub-)graph, compute their clustering values and delete those with the lowest values. 2. Verify that the two resultant subgraphs satisfy the community-chosen definition if the removal splits the (sub)graph. In such case, hold on to that section of the dendrogram. In order to get a desirable community division outcome when the objective function reaches its maximum, several optimization techniques employ modularity as the optimized objective equation. The extremal optimization technique was suggested by Duch and Arenas as a divisive method for discovering communities in intricate networks. In this method, modularity is the global variable that needs optimizing, and λ_i, the fitness of node i, is the local variable that is defined as part of the extremal optimization process.

$$\lambda_i = \frac{k_{r(i)}}{k_i} - a_{r(i)}$$

$k_{r(i)}$ How many connections there are between nodes i and j inside a certain community r.

Conclusion

The final network split consists of the obtained community structures and other components. Through the use of several real-world instances, we have shown that ComTector is both efficient and useful. Using real-world networks as an example, we find that ComTector successfully extracts meaningful communities that align with both empirical facts and our gut feelings. Additionally, we use ComTector to examine scientific cooperation and telecommunication call networks, whose topologies are often opaque.

Reference

1. Chandra Prakash, Agrawa C, Meena P. Graph theory based data extraction method for community detection on social media. International Journal of Innovative Research in Technology and Management. 2022;6(5):16-23.
2. Wang D, Li J, Xu K, *et al*. Sentiment community detection: exploring sentiments and relationships in social networks. Electronic Commerce Research. 2017;17:103-132. doi:10.1007/s10660-016-9233-8.
3. Fox W, Everton S. Using data envelopment analysis and the analytical hierarchy process to find node influences in a social network. The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology. 2014;12:157-165. doi:10.1177/1548512913518273.
4. Singh SS, Mishra S, Kumar A, Biswas B. Link prediction on social networks based on centrality measures. Principles of Social Networking: The New Horizon and Emerging Challenges; c2022. p. 71-89.
5. Seo J, Kim MH. Finding influential communities in networks with multiple influence types. Information Sciences. 2021;548:254-274. doi:10.1016/j.ins.2020.10.011.
6. Vega L, Mendez-Vazquez A, López-Cuevas A. Probabilistic reasoning system for social influence analysis in online social networks. Social Network Analysis and Mining. 2021;11(1):1. doi:10.1007/s13278-020-00705-z.
7. Curado M, Tortosa L, Vicent JF. A novel measure to identify influential nodes: return random walk gravity centrality. Information Sciences. 2023;628:177-195.
8. Nagaraj R, Ramya GR, Yougesh Raj S. GitHub Users Recommendations Based on Repositories and User Profile. In International Conference on Emerging Research in Computing, Information, Communication

and Applications; c2023. p. 127-143. Singapore: Springer Nature Singapore.
9. Diaz-Garcia J. Text and opinion mining techniques in social media environments; c2023. p. 1-260. doi:10.13140/RG.2.2.22279.82086.

Creative Commons (CC) License
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.