Math Overflow temporal network

***Abstract*—In this paper, we present our recent work on Graph algorithms, with a focus on algorithms and applications. We have analyzed mutual liking between pages on Mathoverflow online community which is a platform to ask and answer questions related to Mathematics.**

I. Introduction

This paper provides insights on knowledge sharing, collaboration practices, and authority identification within the mathematical domain through user interactions. Using techniques like PageRank, Connected Components, and Triangle Counting to analyze this social network dataset can reveal hidden links and structures that help us understand the MathOverflow community better.

II. Definition

facilitating the analysis of relationships and information flow within this mathematical platform [1].

A. Degree of connection

In order to discover influential contributors and patterns of involvement within the mathematical community, it assesses the degree of engagement and networking activity of individual users.

B. Chi-squared test

The MathOverflow dataset's categorical variables are compared using the Chi-squared test to see whether certain elements, such as user reputation levels, have a significant impact on engagement trends [3]. This test sheds light on the connection between user attributes and their interactions within the community by comparing observed and expected frequencies [4].

Formula:

$$X = \Sigma \frac{(O - E)}{E}$$

χ^2 is the Chi-squared test statistic.
Σ denotes the sum over all categories or cells.
O represents the observed frequency in each cell.
E represents the expected frequency in each cell,

C. MathOverflow online community

An online forum for mathematicians and math lovers is called MathOverflow. People can post queries about mathematics there and receive answers from other people. It draws eminent mathematicians who actively participate in discussions and problem-solving because of its high degree of knowledge. MathOverflow is an invaluable tool for gaining access to advanced mathematical information, facilitating cooperation, and participating in in-depth conversations on a variety of mathematical issues[4]. It places a heavy emphasis on high-quality content and has a worldwide user base.

III. Findings

Demo 1 - In this demonstration, we have loaded a graph from the higgs retweet network.edgelist dataset and applied two graph analytics algorithms [1].

PageRank Algorithm: We have identified the top 10 nodes with the highest PageRank scores. These nodes are considered the most influential in the network, based on their connections and interactions.

Connected Components Algorithm: We have also found the sizes of connected components within the graph. Connected components are groups of nodes that are strongly connected to each other, forming separate subgraphs. Here, we've listed the top 20 largest connected components along with their sizes [2].

Demo 2 - In this second demonstration, we have continued the analysis of the same graph by focusing on a different aspect:

Triangle Counting: We have calculated the number of triangles associated with each node in the graph. A triangle is a subgraph formed by three nodes and the edges connecting them. By counting triangles, we can identify nodes that participate in many local clustering patterns.The output lists the top 20 nodes with the highest triangle counts. These nodes are likely to be involved in numerous local interactions and can provide insights into the network's community structure and clustering behavior.

These two demonstrations showcase different aspects of graph analysis, including node importance, network structure, and local clustering, using the provided code and dataset. The actual results will depend on the specific data in your graph.

Demo 3 - In this demonstration, we combine PageRank analysis with triangle counting to gain a comprehensive understanding of the graph [3].

PageRank Algorithm: Similar to Demo 1, we've identified the top 10 nodes with the highest PageRank scores. These nodes represent the most influential entities in the network, based on their interactions and connections.

Triangle Counting: As in Demo 2, we've calculated the number of triangles associated with each node. Triangles are indicative of local clustering and community structure within the network [4].

A. About Dataset:

This is a temporary network of Math Overflow exchanges on the internet. A directed edge u, v, t. can represent one of three kinds of interactions:

(a) In the network sx mathoverflow a2q, user u responded to user v's query a.

(b)On user t time t.v's query at time t in the network responded to user v's query a

(c)On user v's response at time t n the graph sx mathoverflow c2a, user u left a comment.

Dataset statistics	*sx-mathoverflow-a2q)*
Nodes	21688
Temporal edge	107581
Edge in the graph	90489
Time span	2350 Days

Dataset statistics (sx-mathoverflow-a2q)

B. Data Format

where edges are separated by a new line and

- *SRC: id of the source node (a user)*
- *TGT: id of the target node (a user)*
- *UNIX TS: Unix timestamp (seconds since the epoch)*

About Algorithm:

A. PageRank

Google created the PageRank algorithm to evaluate the significance and relevance of websites in search engine results. Each webpage receives a numerical rating depending on the quantity and caliber of incoming links. Higher PageRank-scoring websites are regarded as having greater authority and show up higher in search results. The algorithm provides a rating of webpages that represents their importance within the web graph after iteratively redistributing scores until convergence.

B. Connected Components

A graph's clusters or groups of connected nodes can be found using the Connected Components approach. In order to identify nodes that are a member of the same connected component, iteratively traversing the network assigns each one of them a special label or identification. A direct or indirect path of connections is shared by all nodes within a connected component. This method helps with tasks like social network analysis and community recognition by evaluating the structure of networks and spotting different subgroups or communities inside a larger graph.

C. Triangle Counting

Triangles in a graph are found and counted is a using the Triangle Counting technique. It operates by looking at every node in the graph and determining whether any of its neighboring pairs form triangles. It maintains track of triangles that are verified. Understanding community structures and triadic interactions inside networks, such as social or cooperation networks, is largely dependent on this algorithm. It is useful for large-scale network research because of enhancements such as edge-list data structures and parallel processing, which frequently increase its performance.

V. Results And Discussion

- PageRank reveals key authorities and influential users in MathOverflow, aiding in the identification of prominent contributors to the community.
- Connected Components analysis unveils isolated subgroups and communities, shedding light on the network's underlying structure and clustering of users with shared interests.
- Triangle Counting highlights prevalent triadic interactions among users, providing insights into knowledge sharing patterns and collaborative relationships within the MathOverflow community.

Conclusion

PageRank, Connected Components, and Triangle Counting algorithms to the MathOverflow dataset has enriched our understanding of user interactions and community dynamics. These algorithms identified influential users, revealed community structures, and quantified triadic relationships, offering valuable insights for community management and collaboration within the MathOverflow platform.

Acknowledgement

We gratefully acknowledge Prof. Animesh Chaturvedi for helping us throughout.

Reference

[1] P. Holme and J. Saramäki. Temporal networks. Physics Reports, 519(3):97–125, 2012.

[2] Kossinets and D. J. Watts. Empirical analysis of an evolving social network. Science, 311(5757):88--90, 2006.

[3] Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In CHI, 2010.

[4] Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In CHI, 2003.