CLASSIFICATION FOR COMMUNITY STRUCTURE IN SOCIAL MEDIA NETWORKS

[1]Kavita Sawant, [2]Sayali Dhokare, [3]Gayatri Zalwar, [4]Komal Kumari, [5]Mrunal Bhavsar

Abstract: This system presents a novel hybrid method based on both optimization and hybrid hierarchical clustering for community structure in social networks. Because the hybrid hierarchical model usually generates the problem of convergence to a locally optimal detected community, we proposed a new modularity based opinion leader function and we introduced a met heuristic namely genetic algorithm to optimize this introduced quality unction. We introduce a genetic hybrid formulation of hierarchical clustering for community detection in social networks, where hierarchical bottom-up and top-down methods are combined to produce the same community structure. Performances of the proposed approach are evaluated using both real and artificial networks. Experiments show the efficiency of the introduced method in improving the execution time and enhancing the quality of the clustering results.

Keywords: Social Media, Machine Learning, Natural Language processing.

Introduction:

The hierarchy structure of networks describes the organization of communities joined to form the entire network. Thus, hierarchical clustering for community discovering present a crucial task.

The hierarchical organization of social networks is an important issue to discover the hierarchical relationships between communities. We are interested in the detection of communities according to the hierarchical structure of the network. The whole network constitutes a single community. Then, the network is split into several sub-partitions. Single clustering technique and complete clustering technique are the most well-known approaches.

This system presents a novel hybrid method based on both optimization and hybrid hierarchical clustering for community structure in social networks. Single clustering technique and complete clustering technique are the most well-known approaches. Contrary to the first method which uses the minimum distance between elements in different partitions, complete linkage clustering technique aims at merging element with maximum distance. Hence, single linkage clustering tends to produce larger communities, while complete linkage clustering tends to produce compact and spherical clusters. Moreover, the third most popular method of distance measuring is the mean linkage clustering which focuses on the average of all distance between elements in two partitions. Indeed, average clustering technique represents compromises between the two extremes distances.

Related work:

In this study, we tried to benefit from both hierarch0ical method (agglomerative and divisive) and introduced a genetic hybrid hierarchical community detection method. Indeed, we propose a genetic hybrid formulation of hierarchical clustering for community detection in social network, where hierarchical bottom up and top-down methods are combined in order to produce the same community structure. Therefore, the proposed method assumes the existence of an initial community structure originally composed of n partitions.

1. Optimal initial partitioning for high quality Hybrid hierarchical community detection in social Networks

Description: In this paper, we presented a community detection method that uncovers simultaneously both the hierarchical process for community structure of complex networks.

2. CORE CLUSTER: A degeneracy based graph clustering framework

Description: In this article, we present CORECLUSTER, an efficient graph clustering framework based on the concept of graph degeneracy that can be used along with any known graph clustering algorithm. Our approach capitalizes on processing the graph in a hierarchical manner provided by its core expansion sequence, an ordered partition of the graph into different levels according to the k-core decomposition

3. An introduction to social network data Analytics

Description: A social network is a de□need as a network of interactions or relationships, where the nodes consist of actors, and the edges consist of the relationships or interactions between these actors.

4. Learning similarity metrics for Event identification in social media

Description: In this paper, we can enable event browsing and search in state-of-the-art search engines. To address this problem, we exploit the rich context associated with social media content, including user-provided annotations and automatically generated information.

5. Link communities reveal multiscale complexity in networks

Description: In contrast to the existing literature, which has entirely focused on grouping nodes, link communities naturally incorporate overlap while revealing hierarchical organization. We find relevant link communities in many networks, including major biological networks such as protein–protein interaction and metabolic

Motivation:

Nowadays social media is used to introduce new issues and discussion on social media.

The number of users participates in the discussion via social media. Different users belong to different kinds of groups. Positive and negative comments will be posted by the user and they will participate in the discussion. Here we proposed a system to group different kind of users and

system specifies from which category they belong to. For example a film industry, politician, etc.

System Architecture:

In this study, we tried to benefit from both hierarchical method (agglomerative and divisive) and introduced a genetic hybrid hierarchical community detection method. Indeed, we propose a genetic hybrid formulation of hierarchical clustering for community detection in social network, where hierarchical bottom up and top-down methods are combined in order to produce the same community structure. Therefore, the proposed method assumes the existence of an initial community structure originally composed of n partitions.

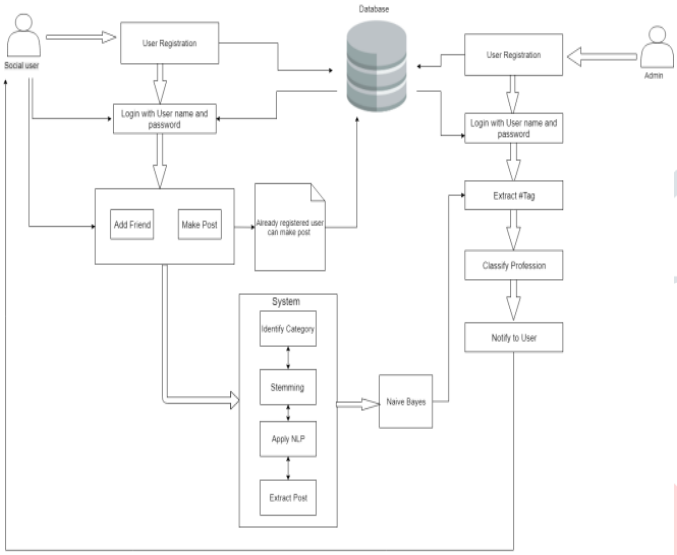


Fig. The System Architechture

Mathematical Model

Input= {K,T}

K= Input post .

S={s,i,F,o}

S represents our proposed system.

s represents start state of the system.

i represents input of the system i.e. post Details.

o represents output of the system i.e. predict professions.

F = {f1, f2, f3}

Represents Functions of the system.

f1= User Details.

f2= User Posts.

f3= User professions.

Algo = {n, K, s}

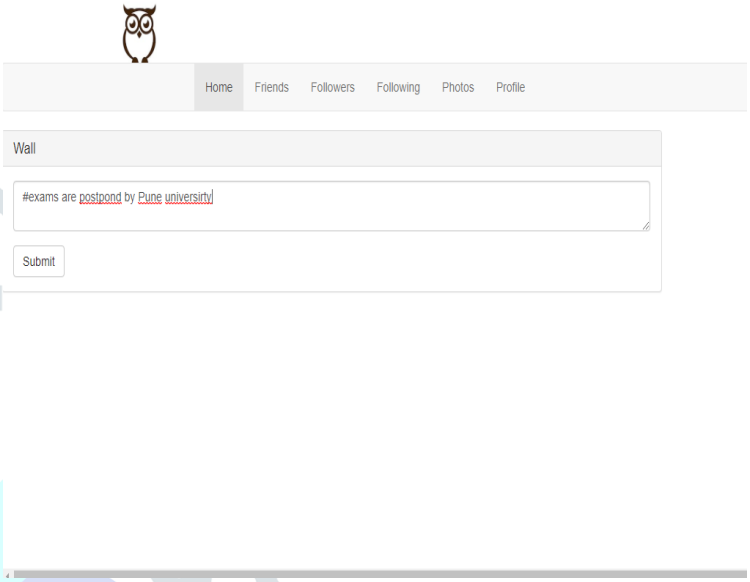
Apply algorithm on input

n=Naïve Bayes

Output S = {GD}

GD = Predicted professions.

Output



Conclusion:

In this paper, we tried to optimize the efficiency of the introduced hybrid hierarchical clustering for community structure in social network. Our main contribution is the genetic hybrid hierarchical clustering technique relying on an objective function which measures a novel modularity function based on influential user's degree.

References:

[1] Toujani.R and Akaichi.J. Optimal initial partitionning for high quality hybrid hierarchical community detection in social networks. In Proceedings of the International Conference on Control, Decision and Information Technologies (CoDIT17), 2017.

[2]. J. P. Ahn, Y.-Y. Bagrow and Lehmann.S. Link communities reveal multi-scale complexity in networks. Nature, (446):761764, 2010.

[3]. Hila Becker, Mor Naaman, and Luis Gravano. Learning similarity metrics for event identification in social media. In Proceedings of the third ACM international conference on Web search and data mining,pages 291–300. ACM, 2010.

[4]. Fragkiskos D.M.-Dimitrios M.T. Christos, G. and V. Michalis. Corecluster: A degeneracy based graph clustering framework. In Proceedings of the Twenty- ighth AAAI Conference on Artificial Intelligence, 2014.

[5] Toujani.R and Akaichi.J. Machine learning and metaheuristic for sentiment anaysis in social networks. In Proceedings of the Metaheuristic Internatianal Conference (MIC'15), 2015. 3. Charu C Aggarwal. An introduction to social network data analytics. In *Social network data analytics*, pages 1–15. Springer, 011.