International Journal of Engineering & Technology, 7 (3.24) (2018) 706-712

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Study of Identifying Structural Hole Spanners in Large Scale Networs

K.Shruthi, Dr. Y. Sri Lalitha

[1] *M.Tech, Department of IT, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana,India.*
[2] *Professor, Department of IT, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India.*
**Corresponding author E-mail: srilalitham.y@gmail.com,*

Abstract

The expansion of internet and increasingly evolving Social Network Sites (SNS) has facilitated internet-users to access a common stage for communication. With the increase in communications between a group of internet-users or between a pair of individuals, keeping track of their interactions is becoming a complex task. Social Networks datasets contain valuable information that can be examined to extract useful patterns. Social Network Analysis (SNA) is a study of patterns and relations in data. This involves identifying Communities or sub-communities from a large network of nodes. The individuals or entities in such Networks are partitioned into closely connected groups called Communities. There can be network structures that connect two or more communities. The Individuals that connect communities in such networks are termed as Structural Hole Spanners (SHS). Closely related communities and Structural Hole Spanners play an important role in various business applications. Therefore the study of identifying the dense communities and highly qualified structural holes in a large network of nodes has become study of interest among researchers. The evolution of Big Data Analytics is making it feasible. This paper studies the techniques of community detection and structural holes identification in literature. It made a preliminary study on SHS and presented two popular Structural Holes Identification Methods with their results.

***Keywords**: Filter Techniques, Lower upper bound estimation, all pairs shortest paths, top-k structural hole, social network*

1. Introduction

A **Social Network** is a collection of People or Institutes that have a common interest and wants to communicate their interest connected through a set of links called interactions thereby forming a structure. The structure and the interactions represents the relationships among these users. With unprecedented improvement in network range, there is a tremendous enthusiasm for developing capable and expandable models to research some unique kind of properties of such large networks. Most public nets show the arrangement of network as groups. That is, the vertices of a system be able to be assembled keen on various collection of group, wherever the vertices of a same system share relative properties, interests and resources. Systems in network perform a fundamental role in the dispersal of information inside the network; Information inside a system streams rapidly and is spread to various systems across over the system breaking points or links. social structures of various groups and has introduced the possibility of structural holes as positions that can join various organizations and bring profits to the beholder. A Community/Social Network is represented by graph ,which consists of nodes as individuals/organizations and edges by interactions.

The social network point of view provides a set of strategies to examine the structure of entire social entities as well as a variety of concepts describing the patterns noticed in these structures. Social network analysis (SNA) is the technique of performing analytics on data found in social environmental circles to understand the relationships and structures of underlying network. With the alarming rise of social media, many different concepts have been brought about through the behavior of people within these networks. Of great confusion was how to identify and award credits to members who greatly influence the social network.

A **Community** is a group of entities that has something in common, such as areas of research - Research Community, common stands on politics – political community, Student Community, Teaching Community, Business Community, Genealogical Links, etc. They often share a subject of common interest, Sense of Place (geographical) or Virtual Space through various communication platforms. Usually Communities are categorized as small/large communities. The Similarity/distance Measures the distance between communities and the categories of communities.

The entities in such arrangements can be alienated into a unlike collections of closely related social groups and these entities who join particular systems, mentioned as hole spanners. Information within these clusters tends to be rather homogeneous and redundant. It shows that information got from people in a comparative system incline to be consistent, while information through contacts with people from different systems is significantly more repetitive, there is said to be a structural hole between them.

Structural holes Spanners are the nodes that bridge different communities. Structural holes have capability to accumulate resources/information and thus, get advantage from access. Structural holes appears in social networks when there is no direct connection among the entities of communities. The concept of structural holes explicate the advantage to be derived from innovation or improvements in various business/research opportunities in social networks and their interconnecting groups. The concept can be applied to the relationships between individuals, organizations, or other entities, that exhibits social networks.

Motivated by previous works this paper is a study of two major aspects of Social Networks Community detection and Structural hole spanner identification in large scale Social Networks.

Community detection techniques are discussed in section 2, section 3 details on Structural hole spanners identification methods, conclusions and future works are discussed in section 4.

2. Community Detection

Learning of organized groups or clusters in Social Network Sites is termed as society detection. People with common interests, similar areas of works get connected in Social Networking Sites to form virtual gathering or Communities. Detection of such collaborations have significant claims such as determining a set of compatible users for recommending a product in Marketing, finding a research group, identifying Protein Networks in Bio-logical Networks and so on. Recent studies proposed a variety of Community Detection Methods and applied them in different fields of Research. Community detection methods are classified based on the strategies they have been developed on. The Learnings used are based on Clustering, Graph Partition, Genetic Algorithms, Label Propagation, Semantics of Networks, Overlapping Networks and Dynamic Networks [2].

Approaches

Graph Partition Approach

The Social network is treated as a graph and these methods partitions graph into sub-graphs such that there exists very less connection between the partitioned sub-graphs. [3] is the earliest method proposed using this approach. It divides the nodes of the graph by repeatedly swapping nodes between partitions based on the edge cost and there by removing the edge with minimum cost. In [4], the maximum likelihood method is explored to determine a set of candidates, where in each of the derived candidate is itself a result of minimum cut graph partition algorithm.

Clustering Approach

The intension of Community Detection is to identify groups, interrelated subgroups which relates to the concept of clustering a Machine Learning Technique. Primary study of Community detection methods can be seen in [36]. This work uses divisive method on edge-between-ness of an edge as the number of shortest paths between pairs of nodes that contain the edge for a graph with undirected and un-weighted edges. Communities are constructed by identifying the high edge-between-ness edges, and discarding those from original graph. This approach grabbed the attention of researchers and many extensions were proposed [5-12]. In [6], algorithm [36] is extended in Bio-Informatics to determine functional modules in the yeast proteome network by assigning weights to graphs. In [5] ,it reduced the computational complexity by using indexing methods. In [8], he proposed network partitioning using centrality and between-ness properties of graph theory. [13] extended [36] by proposing strong and weak collaborations. The edge to be removed is identified using edge Clustering Coefficient and thereby forming portioned community. To address the Large Scale networks, Parallel Version of GN algorithm with Map Reduce Model and GraphChi [14]. Optimized partitioning to identify community using Greedy method for scalable networks is used in [15]. In [16], an iterative two phase algorithm is deigned, wherein Initially, all nodes are assigned to diverse groups and then the gain of partitioning is computed to move a node from one group to another. If the partition gain value is optimistic, the node is moved to a new group. Next step the groups found in previous step are considered as nodes and the weights of links are determined. The Simulation for network partition optimization is used in [17]. An improved performance of [17] explores the concept of intra and inter edges[18] and [19] uses empirical search methods to optimize the modularity function using extreme optimization technique. Determining groups from complex networks with improved accuracy is addressed in [20]. Each vertex in these networks is a self-directed representative and exhibits a gathering behavior and collects the group information and directs itself to the proper community. It is believed in [21] that sub-groups exists in communities and determining sub-groups is essential to determine a community, hence they applied Hierarchical Fuzzy Spectral Clustering approach to determine a Community. To address the ever changing networks (Dynamic) and communities in social network sites, the density-based incremental clustering approach is Presented in [22] for large dynamic datasets with noise. To find groups in a network, in [23] a graph flow simulation algorithm is proposed and simulated that a graph resembles network of communities. This approach has two different processes of ‘expansion’ and ‘inflation’. In [24] Markov based Community Detection method is specified. It used a visit of vertices in a network to find the groups accessible in the network construction. In [25] an improved approach to identify the communities in the network of nodes is suggested by performing the short random walks many a times. The aggregate similar nodes are grouped together to form clusters by considering the consensus of the communities.

Genetic Approach

Genetic Algorithm (GA) is an empirical search method based on the theory of natural evolution and exhibits best performance in a given circumstance. It proceeds by a set of solution called chromosome and calculates the fitness of these solutions.

The GA-Net algorithm presented in [26] introduced an effective fitness function that cumulates the score of community of nodes represented as a graph. The vertices of these networks are showed by genes and alleles. The algorithm produced quality communities by optimizing the scores. It formed dense communities by selecting the promising search path for optimal results without knowing in advance the number of communities. A multi-objective community identification is addressed in [27], to generate inter-related communities or sub-communities in hierarchical model. It generated a set of communities at varying hierarchies and the communities at higher level of hierarchy contains lower number of communities and the ones deeper in hierarchy contain fine cohesive communities. Single and Multi Objective optimization approaches are addressed in [28]. Roulette Selection based Genetic Algorithm is employed in the former approach while NSGA-II algorithm was used in the later approach. Single objective groups identification in a network is presented in [29], where in undirected graphs with the concept of group-score is used as a fitness function.

In [30], the combination of hierarchal clustering and fitness functions applied to identify the community structures in a network.

It employed repeated divisions of network represented as a graph. Then a nested fitness functions are applied to them. In [31] also optimized the network modularity using GA. A multi-cultural algorithm presented in [32] to identify a group of similar taste used the GA-Net based fitness function to determine the communities. In [33] the best fitness value is employed in the state space search of representing network to automatically signal the direction of search by identifying the valid search paths for the individuals to form communities.

Label Propagation Approaches

In a Social Network, the Label is circulated among the nodes of network and each node receives the labels attained by a maximum number of its neighboring nodes.

The nodes are assigned a label which is most accepted among its neighbors. For each node the labels of neighbor nodes are examined, and the label that is most common to its neighbors is the label assigned to the node. Label Propagation Algorithm(LPA) was proposed in [33] where each node tries to achieve a label possessed by maximum number of its neighboring nodes and stops the process till all nodes acquire such label. The LabelRankT algorithm by [35] detects communities by considering the edge weight and direction in dynamic networks and exhibits efficiency in detecting evolving communities. In [34] Label propagation in overlapping communities is suggested, the algorithm Balanced Multi Label Propagation Algorithm (BMPLA) removed restriction on membership of a node. belonging to the largest number of communities membership of nodes in community belonging the limited number of communities, thereby allowing nodes to belong to any number of communities.

Semantics Based Approaches

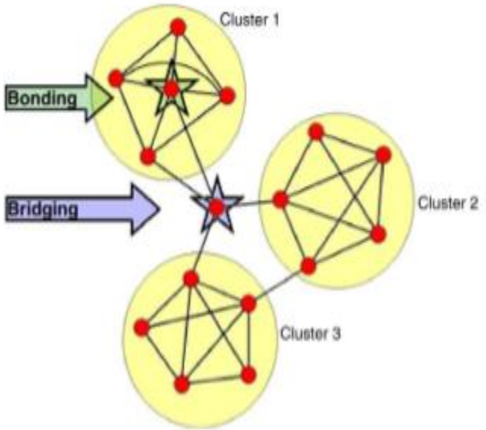
Usually, to identify the communities the underlying network topology is applied. In Semantic based approaches, the content specified by the users with the underlying network can determine communities. The messages shared among the users, the attributes of nodes can be used to determine the partitions in the social networks. The most prominent context identification in Text data called LDA is applied in different semantic based Community partitions [37]. The overlapping communities are identified establishing relations based on the topic of interaction called Link field topic (LFT)[38] modelling and Semantic Link Weight (SLW). LFT overrules the priori indication of the number of communities. SLW based on LFT, evaluates the weights of Semantic Links. In [39] a semantic system is developed utilizing data from the remark content extracated from the underlying HTML source documents. A normal score is secured for two clients for each connection expecting remarks to be implicit connections between individuals.. A community detection algorithm, identifies and in addition labels communities with the help of tags utilized by group during the social labeling procedure and the semantic affiliations derived between labels. In [40], a topic oriented methodology comprising of an amalgam of social items bunching and link investigation has been used. The K-Means clustering algorithm is modified and it is called as 'Entropy Weighting K-Means (EWKM) algorithm' has been utilized to group the social objects. A subspace clustering algorithm is applied to gather all the social objects into subjects. In [41] and [36] , the two strategies are broadly utilized in the process to distinguish topical networks. In [42], LDA is used to discover hyper bunches in the blog substance and after that sentiment analysis is done to additionally discover the meta-clusters in these units. A Link-Content model is proposed in [43] for finding theme based networks in social communities. Network has been displayed as a distribution employing Gibbs testing. A great part of the concentration inside community recognition has been on recognizing disjoint networks. This sort of detection expects that the system can be divided into strongly connected regions in which there are more connections with each other compared to the other part of network. It is sure that the people in social network are the members in multiple communities. For instance, a person can have a contact with his or her friends, family, colleagues, for ex. Consider a scientist – he/she might be strong in certain areas of scientific research whose knowledge is required in many research projects, consider another example, an individual may have accounts in different social media sites, this count of logins for an individual may belong is unlimited and have an access to these sites at the same time. In [44] and [45], the authors demonstrated that the overlapping is necessary in many networks. Thus, there is developing an enthusiasm in overlapping community algorithms that recognize an arrangement of groups that are not really disjoint. There could be nodes that are present in more than one group. Here, we provide the latest research developments in Overlapping Community Detection methods. The term 'Cover' is a group of communities found in public nets, wherein a node may belong to more than one cluster or community, according to a membership factor. Some of overlapping community detection methods are discussed here.

1. **Clique Percolation Method (CPM)** :-- The method in [46] assumes that there exists overlapping groups of completely associated sub-graphs and recognizes networks via scanning for nearby cliques.
2. **Line Graph and Link Partitioning** ::-- To identify the community structures, process of dividing links rather than nodes is examined in graph networks. A node in the original graph is considered as overlapping node, if its associated links are assigned to multiple clusters of communities. [47].
3. **Fuzzy Detection** :- Since a user can belong to multiple communities, their inherently exists vagueness in the individual belonging-ness to a community. Therefore fuzzy methods are introduced to handle such criteria. The quality of belonging-ness among all the sets of nodes and communities is determined using Fuzzy community Detection method. Initially, a delicate participation vector, or having a fitting factor[48], is determined for every node. The disadvantage of such systems is the process of discovering the dimensionality k of the partnership features. This value can be passed as an argument or automatically calculated from the data
4. **Local Expansion and Optimization::-** Algorithms using local extension and improvement depend on increasing a natural community [49] or a partial community. The vast majority of them depend on a local benefit work that describes the nature of a closely associated group of nodes.
5. **Agent-Based and Dynamical Algorithms::-** In Label base community detection methods, the nodes with common Labels are grouped to form communities [50],[51], these methods are stretched out to overlapping community identification by enabling a node to have different labels. In [48], each node replaces its fitting factors by taking an average of the coefficients of every node in its neighborhood in a synchronous manner in every iteration. To keep track of the number of communities to which a node can be related is recorded using a parameter v.

3. Structural Holes Detection:

Structural holes are a thought of social network research, at first made by Ronald Stuart Burt. The examination of structural holes spreads to the fields of human science, financial issues and informatics. Burt [52]exhibited this concept, to try to make it to clear about the beginning stage of differentiations in social capital. Burt's theory prescribes that individuals have certain point of benefits/drawbacks of how they are fixed into neighborhoods or other social structures. A structural hole indicates the important individuals in the group who have equivalent and distinguished sources of information.

Most social structures tend to be depicted by dense group of strong relationships, generally called network terminations. The approach relies upon a principal belief that the homogeneity of information, new concepts and behaviour is generally more higher in any social group of people than that of two groups of people. A man who behaves as a broker between something like two social groups of solidly related individuals could gain basic advantages. In particular, the circumstance of an expansion between different categories of groups allows the trade or limit of gainful information beginning with one group then onto the following. Besides, the individual can combine each one of the considerations he gets from different sources and get the best information, An innovative idea among all. Meanwhile, broker has a precarious position, as associates with heterogeneous groups can be weak and take a long time.



The thought of structural hole theory is slightly similar to the concept of weak ties, by Mark Granovetter[53]. According to the opinion of weak ties, the more powerful the association between two people, the more possible their contacts can overlap ,so that they are going to have common ties with an equivalent third party. This recommends tying the distinguishrd social groups is a potential point of supplying new concepts. Consequently, Granovetter says that strong ties are not going to exchange new information. The two thoughts rely upon the identical fundamental model, anyway it is possible to distinguish a couple of complexities between them. While Granovetter says that if a contact would fill in as an expansion depends upon the nature of a tie, Burt considers the other method of causality. In this way, he relies on the proximal reason, while Granovetter chooses for the distal reason(strength of ties).

Structural holes have a collection of genuine applications, includes:

The Structural holes are used to generate a community in an academic system that produces the addresses of people with relative research interests, and the people who can join unlike research systems are simply more Powerful in uniting ideas from different research people and develop integrative works. In community identification, recognizing center points that associates particular groups can help to separate and discover systems. In the wide spread occurrence of infections and gossips spreading, quarantining structural holes can halt the outspread of infections and rumours in various systems. In viral displaying, the most convincing structural holes can dispatch the promotion of new things in different social media sites. In graph density, structural holes are extraordinary contender for k-shattering since they interface diverse parts of a network and by eliminating them will influence the network to be disconnected. In the Organizations, structural holes are utilized to examine both their current internal and external social networks. If there is a requirement for more thoughts and development in the organization, they can search for where structural holes may exist. In new startup companies, understanding the effect of structural holes in one's social communities can be key for business people included while settling on risk taking choices, both to discover profit by the structural holes and also to keep away from misuse by others' utilization of them [54][55]. In the entrepreneurial decision-making process, Structural holes can make a difference directly. An entrepreneur can traverse structural holes to potentially increase some extra knowledge and successfully

Expansion danger bringing trademark to entrepreneurial choice making. Clinched alongside [54], discovered that the existence from claiming structural gaps and spreading In these gaps provoked as much entrepreneurial subjectexamination— basically farmers, will make higher dangers Toward building bigger hydroelectric micro-power plants.

In an organization's advertising, By distinguishing potential sites for standard notices and banner advertisements that are least controlled, strongly connected, and connecting structural holes in their social network structure, an organization can accomplish their most elevated ROI in their publicizing endeavors by augmenting their navigate to impression rate as anticipated and found in [56] (an investigation on 25 Twitter-related sites).

Hole spanners are the nodes in the network which connects different communities. Structural hole spanners can be identified using many algorithms. Some of the are listed below.

1. PageRank [61] Algorithm Calculates the significance of every node and afterwards chooses those nodes with the most highest scores as structural hole spanners.
2. PathCount Algorithm[59] is like between-ness centrality Also allots each vertebrate fossil science a score that is the amount for briefest approaches Around all-pairs most brief paths, ahead which those vertebrate fossil science lies, at that purpose decides those top- k vertices with the the vast majority raised scores. Constraint Algorithm [52] measures the constraint by its neighbors on each vertex and selects those vertices with the minimum constraint values.
3. 2-Step Algorithm [60] ,this method a score or value is allocated to every node based on the quantity of pairs of its neighbors without edges between them, then chooses the maximum top-k scores.
4. HIS [63] Algorithm allots every node a value that recreates the Probability of node as a structural gap spanner again those provided for subset for system communities, acknowledging that groups would provided for.
5. MaxD [63] Algorithm is to find an arrangement of k vertices to such an extent that the minimum cut of social groups in a networks will be decreased significantly, after deleting these vertices, considering that social groups are given. For any pair of networks groups, this method chooses vertices as structural hole spanners using greedy approach. In each round, it picks the vertex whose elimination will result in a maximum decrease of the minimum cut.
6. Central Algorithm search for k vertices with the minimum average shortest distances separations to different vertices [57], [58], i.e., the k vertices with the greatest closeness centrality.

   The quickest randomized algorithm recommended in [58] will be connected with transform the Normal briefest separation with distinctive vertices to each vertebrate fossil science.
7. AP_BICC calculation, this strategy employments explanation focuses and the limited opposite closeness centrality will identify those top-k structural gap spanners [57]. HAM Algorithm [62] proposed a harmonic modularity strategy to handle the identifications of the two networks and structural hole spanners together.
8. Greedy Algorithm finds a set of *k* hole spanners with in *k* iteration and in each iteration one vertex is removed from network.[57].

4. Our Approach

This paper experimented with Greedy and AP_Greedy methods for the top-k structural hole spanner problem. A greedy algorithm is an algorithmic framework that makes best choices at each input stage with the intention that it derives an optimal solution at the final stage. Algorithm picks the best result at the moment without regard for consequences. At every input greedy picks the best answer with its loacl scope and repeates this process till all inputs are considered. In this process it pays little attention to derive the best solution after all inputs considered. Hence, this approach will not give the best solution always.

The block diagram respresents how the dataset is used to identiy the Structural holes using greedy and AP-greedy(Articulation Point) algorithms.

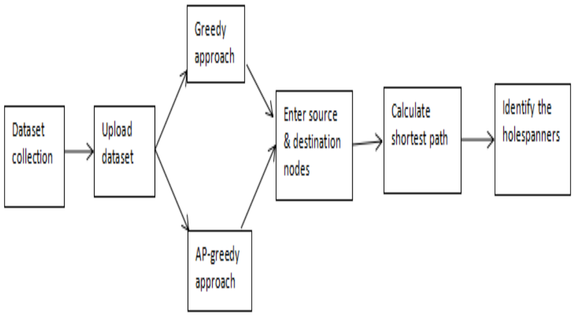


Fig: system Architecture

The approach used two metrics to measures the quality of the hole spanners (S) identified based on i.) the no.of distinct communities that the individuals in a set of hole spanners spans ii.)the sum of the sizes of communities spanned by the hole spanners. We assess those effects of diverse calculations utilizing these two measurements so as to figure out with which level our model maps those true structural hole spanners. It makes use of general relativity and quantum cosmology collaboration network dataset. It consists of Co-author information constituting 5242 vertices and 28,980 edges. It describes the collaborations between authors papers in GR-QC category. If one author *i* presented a paper with another authored *j*, an edge (i,j) is represented in the graph pointing from node i to node j. And the graph is considered as an un-directed graph. If the paper is presented by *k* authors this generates a completely connected sub-graph of *k* nodes.

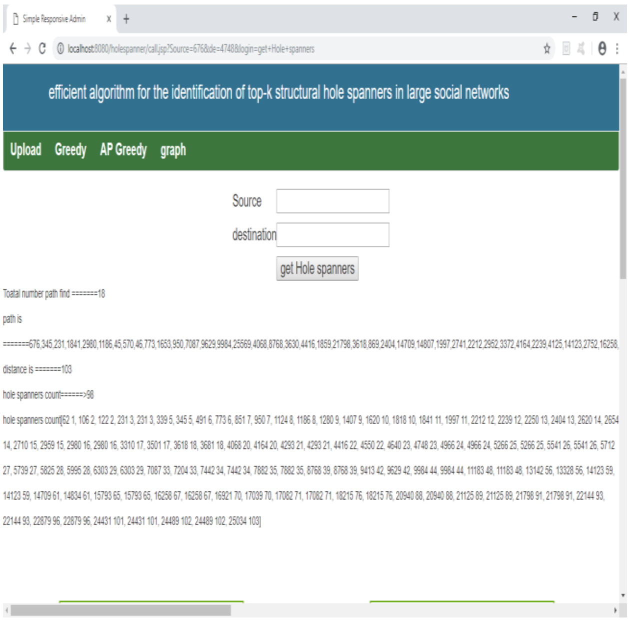


Fig: Result of Greedy algorithm

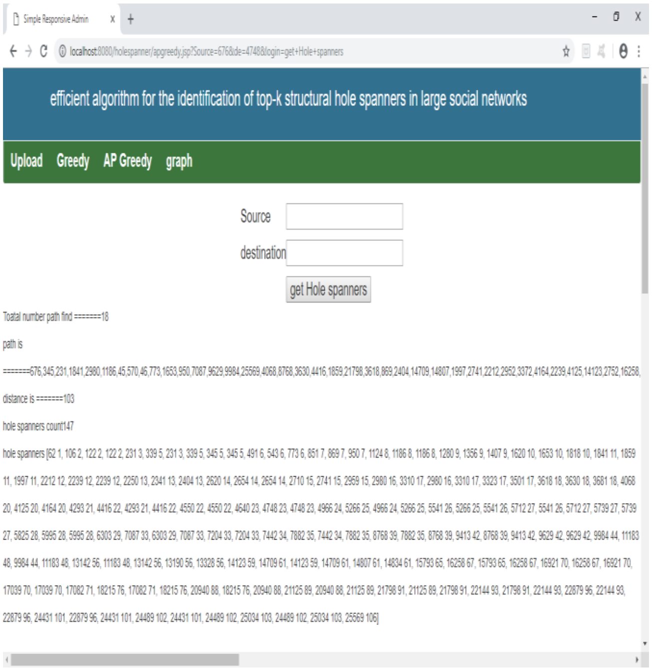


Fig: Result of Ap greedy algorithm

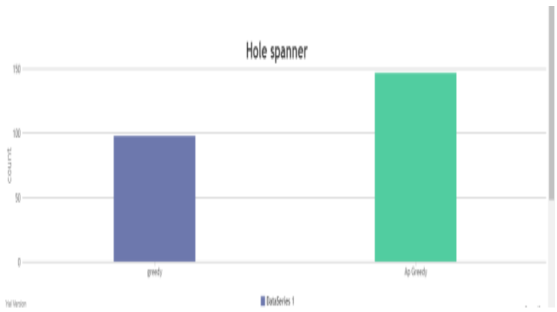


Fig: Results comparison

From the Results it is observed that AP_Greedy approach derived more number of qualitied holespanners for the given dataset.

5. Conclusion and Future work

This paper addressed to give broad concept of Community Detection Techniques and Structural Hole Spanner methods in a huge range of social network. This work implemented two algorithms named Greedy and AP(articulation point) based greedy algorithm. It is observed that the above two algorithms identified the best structural hole spanners and also observed that the AP based greedy approach has derived more no.of hole spanners than greedy approach.

Future Work: - In future we want to extend this study by applying advanced algorithms to determine accurate structural holes and study on various datasets. The focus of this work is to identify different research communities and try to build a community among different research areas and find the best structural hole spanners of such network. Such structural hole spanners are people who connects different research communities and are more effective to join ideas from diverse study group, thereby enabling the inter-disciplinary research works.

References

[1] James, Paul (2006). Gloabalism, Nationalism, Tribalism:Bringing Theory Back In--Volume 2 of Towards a Theory of Abstract Community. London:Sage Publications.

[2] Punam Bedi, Chavvi Sharma,"Community Detection in Social Networks", in : Data Mining and Knowledge Discovery · February 2016.
[3] Kernighan BW, Lin S. An efficient heuristic procedure for partitioning graphs. Bell system technical journal 1970, 49(2):291-307.
[4] Newman M. Community detection and graph partitioning. EPL (Europhysics Letters) 2013, 103 (2):28003. doi:10.1209/0295-5075/103/28003.
[5] Rattigan MJ, Maier M, Jensen D. Graph clustering with network structure indices. In: Proceedings of the 24th International conference on Machine learning(ICML): ACM; 2007 : 783790.doi:10.1145/1273496.1273595.
[6] Chen J, Yuan B. Detecting functional modules in the yeast protein-protein interaction network. Bioinformatics 2006, 22 (18):2283-2290. doi:10.1093/bioinformatics/btl370.
[7] Holme P, Huss M, Jeong H. Subnetwork hierarchies of biochemical pathways. Bioinformatics 2003, 19 (4):532-538.
[8] Pinney JW, Westhead DR. Betweenness-based decomposition methods for social and biological networks. Interdisciplinary Statistics and Bioinformatics 2006:87-90.
[9] Gregory S. An algorithm to find overlapping community structure in networks. In: Knowledge discovery in databases: PKDD Springer; 2007, 91-102.doi:10.1007/978-3-540-74976-9_12.
[10] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A. Self-similar community structure in a network of human interactions. Physical review E 2003, 68 (6):065103. doi:10.1103/PhysRevE.68.065103.
[11] Arenas A, Danon L, Diaz-Guilera A, Gleiser PM, Guimera R. Community analysis in social networks. The European Physical Journal B-Condensed Matter and Complex Systems 2004, 38 (2):373-380. doi:10.1140/epjb/e2004-00130-1.
[12] Tyler JR, Wilkinson DM, Huberman BA. E-mail as spectroscopy: Automated discovery of community structure within organizations. The Information Society 2005, 21 (2):143-153.
[13] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America 2004, 101 (9):2658-2663. doi:10.1073/pnas.0400054101.
[14] Moon S, Lee J-G, Kang M, Choy M, Lee J-w. Parallel community detection on large graphs with MapReduce and GraphChi. Data & Knowledge Engineering 2015, Article in Press. doi:10.1016/j.datak.2015.05.001.
[15] Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Physical review E 2004, 70 (6):066111. doi:10.1103/PhysRevE.70.066111.
[16] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008. doi:10.1088/17425468/2008/10/P10008.
[17] Guimera R, Sales-Pardo M, Amaral LAN. Modularity from fluctuations in random graphs and complex networks. Physical Review E 2004, 70 (2):025101. doi:10.1103/PhysRevE.70.025101.
[18] Zhou Z, Wang W, Wang L. Community Detection Based on an Improved Modularity. Pattern Recognition 2012:638-645. doi:10.1007/978-3-642-33506-8_78.
[19] Duch J, Arenas A. Community detection in complex networks using extremal optimization. Physical review E 2005, 72 (2):027104. doi:10.1103/PhysRevE.72.027104.
[20] Ye Z, Hu S, Yu J. Adaptive clustering algorithm for community detection in complex networks. Physical Review E 2008, 78 (4):046115. doi:10.1103/PhysRevE.78.046115.
[21] Wahl S, Sheppard J. Hierarchical Fuzzy Spectral Clustering in Social Networks Using Spectral Characterization. In: The Twenty-Eighth International Flairs Conference; 2015 : 305-310
[22] Falkowski T, Barth A, Spiliopoulou M. DENGRAPH: A density-based community detection algorithm. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI); 2007: 112115.doi:10.1109/WI.2007.74.
[23] Dongen SV. Graph Clustering by Flow Simulation, PhD thesis, University of Utrecht. 2000.
[24] Nikolaev AG, Razib R, Kucheriya A. On efficient use of entropy centrality for social network analysis and community detection. Social Networks 2015, 40:154-162. doi:10.1016/ j.socnet.2014.10.002.
[25] Steinhaeuser K, Chawla NV. Identifying and evaluating community structure in complex networks. Pattern Recognition Letters 2010, 31 (5):413-421. doi:10.1016/j.patrec.2009.11.001.
[26] Pizzuti C. GA-Net: A genetic algorithm for community detection in social networks. In: Parallel Problem Solving from Nature–PPSN X: Springer; 2008, 1081-1090.doi:10.1007/978-3-54087700-4_107.
[27] Pizzuti C. A multiobjective genetic algorithm to find communities in complex networks. IEEE Transactions on Evolutionary Computation 2012, 16 (3):418-430. doi:10.1109/TEVC.2011.2161090.
[28] Hafez AI, Ghali NI, Hassanien AE, Fahmy AA. Genetic algorithms for community detection in social networks. In: 12th International Conference on Intelligent Systems Design and Applications (ISDA): IEEE; 2012 : 460-465.doi:10.1109/ISDA.2012.6416582.
[29] Mazur P, Zmarzlowski K, Orlowski AJ. A Genetic Algorithms Approach to Community Detection. Acta Physica Polonica Series A-General Physics 2010, 117(4).
[30] Liu X, Li D, Wang S, Tao Z. Effective algorithm for detecting community structure in complex networks based on GA and clustering. In: International Conference on Computational Science (ICCS 07): Springer; 2007:657-664.
[31] Tasgin M, Herdagdelen A, Bingol H. Community detection in complex networks using genetic algorithms. arXiv preprint arXiv: 0711.0491 2007.
[32] Zadeh PM, Kobti Z. A Multi-Population Cultural Algorithm for Community Detection in Social Networks. Procedia Computer Science 2015, 52:342-349. doi:10.1016/j.procs.2015.05.105.
[33] Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 2007, 76 (3):036106. doi:10.1103/PhysRevE.76.036106.
[34] Wu Z-H, Lin Y-F, Gregory S, Wan H-Y, Tian S-F. Balanced multi-label propagation for overlapping community detection in social networks. Journal of Computer Science and Technology 2012, 27(3):468-479.
[35] Xie J, Chen M, Szymanski BK. LabelrankT: Incremental community detection in dynamic networks via label propagation. In: Proceedings of the Workshop on Dynamic Networks Management and Mining: ACM; 2013:25-32.
[36] Girvan M, Newman M. Community structure in social and biological networks. Proceedings of the national academy of sciences 2002, 99 (12):7821-7826. doi:10.1073/pnas.122653799.
[37] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research 2003, 3:993-1022.
[38] Xin Y, Yang J, Xie Z-Q. A semantic overlapping community detection algorithm based on field sampling. Expert Systems with Applications 2015, 42 (1)
[39] Xia Z, Bu Z. Community detection based on a semantic network. Knowledge-Based Systems 2012, 26. doi:10.1016/j.knosys.2011.06.014.
[40] Zhao Z, Feng S, Wang Q, Huang JZ, Williams GJ, Fan J. Topic oriented community detection through social objects and link analysis in social networks. Knowledge-Based Systems 2012, 26:164-173. doi:10.1016/j.knosys.2011.07.017.
[41] Deerwester SC, Dumais ST, Landauer TK, Furnas GW, Harshman RA. Indexing by latent semantic analysis. JAsIs 1990, 41 (6):391-407.
[42] Nguyen T, Phung D, Adams B, Tran T, Venkatesh S. Hyper-community detection in the blogosphere. In: Proceedings of second ACM SIGMM workshop on Social media: ACM; 2010.
[43] Natarajan N, Sen P, Chaoji V. Community detection in content-sharing social networks. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining: ACM; 2013.doi:10.1145/2492517.2492546.
[44] KELLEY, S., GOLDBERG, M., MAGDON-ISMAIL, M., MERTSALOV, K.,AND WALLACE, A. 2011. Defining and discovering communities in social networks. In Handbook of Optimization in Complex Networks, Springer, 139–168.
[45] REID, F., MCDAID, A. F.,AND HURLEY, N. J. 2011. Partitioning breaks communities. In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining(ASONAM'11). 102–109.
[46] FARKAS, I., ABEL, D., PALLA, G.,AND VICSEK, T. 2007. Weighted network modules. New J. Phys. 9, 6, 180.
[47] AHN, Y.-Y., BAGROW, J. P.,AND LEHMANN, S. 2010. Link communities reveal multiscale complexity in networks. Nature 466, 761–764.
[48] GREGORY, S. 2010. Finding overlapping communities in networks by label propagation. New J. Phys. 12, 10.
[49] LANCICHINETTI,A.,FORTUNATO,S,AND KERTESZ, J.2009. Detecting the overlapping and hierarchical community structure of complex networks. New J. Phys. 11, 3.

[50] RAGHAVAN,U.N.,ALBERT,R.,AND KUMARA,S.2007.Near line artime algorithm to detect community structures in large-scale networks. Phys. Rev. E 76, 3.
[51] XIE, J.AND SZYMANSKI, B. K. 2011. Community detection using a neighborhood strength driven label propagation algorithm. In Proceedings of the IEEE Network Science Workshop (NSW'11). 188–195.
[52] R. S. Burt, Structural holes: The social structure of competition. Harvard university press, 2009.
[53] Granovetter, M. S. (1973). "The Strength of Weak Ties" (PDF). The American Journal of Sociology. **78** (6): 1360–1380. doi:10.1086/225469. JSTOR 2776392.
[54] Aarstad, J. (2014). Structural holes and entrepreneurial decision making. Entrepreneurship Research Journal, 4(3), 261–276.
[55] Adams, M., Makramalla, M., & Miron, W. (2014). Down the rabbit hole: How structural holes in entrepreneurs' social networks impact early venture growth. Technology Innovation Management Review, 4(9), 19–27.
[56] Hunter III, D. & Chinta, R. (2013). Structural holes and banner ad click-throughs. Technology & Investment, 4, 30.44.
[57] M. Rezvani, W. Liang, W. Xu, and C. Liu, "Identifying top-k structural hole spanners in large-scale social networks," in Proc. 24th ACM Int. Conf. Inform. Knowl. Manage., 2015, pp. 263–272.
[58] S. Chechik, E. Cohen, and H. Kaplan, "Average distance queries through weighted samples in graphs and metric spaces: high scalability with tight statistical guarantees," in Proc. 18th Int. Workshop Approximation Algorithms Combinatorial Optimization Problems, 19th Int. Workshop Randomization Comput., 2015, pp. 659–679.
[59] S. Goyal and F. Vega-Redondo, "Structural holes in social networks," J. Econ. Theory, vol. 137, no. 1, pp. 460–492, Nov. 2007.
[60] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogeneous networks," in Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM), 2012, pp. 743–752.
[61] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." 1999.
[62] L. He, C. T. Lu, J. Ma, J. Cao, L. Shen, and P. S. Yu, "Joint community and structural hole spanner detection via harmonic modularity," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 875–884.
[63] T. Lou and J. Tang, "Mining structural hole spanners through information diffusion in social networks," in Proc. ACM 22nd Int. Conf. World Wide Web (WWW), 2013, pp. 825–836.