IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. II (May-Jun. 2016), PP 23-35
www.iosrjournals.org

Fmfc: Fuzzy Based Improved Mutual Friend Crawling

Varnica
(M.Tech(CSE), Department of Computer Science and Engineering, GNDU Regional Campus, Gurdaspur, Punjab, India.)

Abstract: *Online Social networks are becoming an important part of research in every field. Crawling such networks and extracting data from them is a very difficult task. Till date, many crawling techniques have been designed for this purpose. The properties of complex networks have been reviewed in this paper. This paper has focused on mutual friend crawling and applied it on a dataset. The proposed technique makes the use of fuzzy logic and a comparison has been made between the existing and proposed technique. The effectiveness of the proposed technique is evaluated using different parameters. This paper also includes a survey of the existing methodologies designed and implemented by different authors. It also provides future directions in this area of research.*

Keywords: *ACC, ADD, APL, BFS, Clustering Coefficient, Complex Networks, DFS, MFC, Overheads.*

I. Introduction

The time taken to crawl the entire network using the existing web crawling techniques, introduces a bias which is required to be minimized. The most common ways used for solving this problem are sampling based on the nodes (users) ids in the network or crawling the network until one feels that an adequate amount of data has been generated. In this paper, a new technique is proposed in which all the users which belong to the same community are crawled first before moving on to the next community. This reduces the bias introduced by the time taken for crawling the whole network. This technique enables a researcher to selectively obtain users belonging to the same community and begin with the evaluation of the network.

1.1 Complex Network

Complex systems are basically networks which are organized into different compartments such that each compartment has its own role and function to perform. All the compartments consist of nodes. The links between nodes are of high density whereas the links between compartments are of low density [40].

A complex network represents the interactions in the real world in the form of a mathematical model. The major problem that occurs in complex networks is the identification of the communities which are hidden in the structure of these networks.

1.2 Communities in Complex Network

A complex network consists of a large number of communities connected to each other. A community is a collection of a number of nodes connected to each other. The nodes in a particular community perform same function and exhibit similar properties. The connectivity among the nodes in a network is more than the connectivity between different communities. This means that more links exist among the nodes of a community than between the nodes of different communities. There exist some networks which consist of overlapping communities. Overlapping communities are the one in which nodes have participation in more than one community.

For each type of network, communities play a very important role as each community depicts a particular function of the network. In a network like internet, there exist communities consisting of web pages (nodes) which are related to a particular topic. Another example can be a network of a city consisting of communities formed by people which belong to the same society. Also, nodes in a community are placed according to the role they perform like nodes which interact with the nodes of other communities are placed at the boundary and nodes which control the activities of all the other nodes in the community are placed in the centre.

Till date, no specific definition has been used for community in a network. Different authors use different definitions of a community for their research work. Several generalizations have been used for different definitions of communities like considering the properties of the network, considering the edge weights of the weights of the network, considering the structure of the network etc. Detecting the communities of a complex network is a very important task for studying the properties of the network. For this purpose, several community detection algorithms have been designed by different authors.

1.3 Properties of Complex Network

Complex networks are very popular and is an emerging topic for research. These networks exhibit many properties which have been studied in the past. Some of these properties are:

1.3.1 Clustering Coefficient

Clustering coefficient is defined as the ratio of number of directed links that exist between the neighbours of a node to the number of possible links that could exist between the neighbours of that node. The clustering coefficient of a network is the average clustering coefficient of all the nodes in the network. [3]. For ith node of a network, clustering coefficient can be calculated as,

$$\mathrm{C} = \frac{2\mathrm{e}}{\mathrm{k}(\mathrm{k}-1)}$$

Where, K is the number of neighbours of ith node and e is the number of edges between these neighbours.

The value of clustering coefficient lies between 0 and 1. The higher value of clustering coefficient indicates that there is higher degree of “cliquishness” between the nodes of the network. The ‘0’ value of clustering coefficient for a graph indicates that it has no “triangles” of connected nodes and if the value of clustering coefficient for a graph is ‘1’ then it is a perfect clique.

1.3.2 Degree Distribution

A network consists of large number of nodes and all the nodes have varying degree. Degree distribution P(k) for a graph gives the probability of a randomly selected node to have a degree ‘k’ in the network. It is used to describe the distribution of the links among the nodes in the graph [28].

1.3.3 Assortativity

The assortativity coefficient ‘r’ is a measure of the connectivity among the nodes in a network which are common in some way like nodes with similar degree. Networks like social networks show assortativity as highly connected nodes tend to be connected to nodes with high degree whereas networks like biological networks show dissortativity as the nodes with high degree tend to be connected with nodes of low degree. The assortativity coefficient r varies from 1 to -1. If r=1, it means that that the nodes tend to be connected to nodes with similar degree but if r=-1, it means that the nodes tend to be connected to nodes with varying degree [3].

1.3.4 Density

The size of the network can be known from the total number of nodes in the network. Density is used to define the level of linkage between the nodes in a network. It is generally calculated as the ratio of the number of existing links between the nodes to the number of possible links. Mathematically, it can be written as,

$$\text{Density}=\frac{2\mathrm{E}}{\mathrm{N}(\mathrm{N}\text{-}1)}$$

Where, E is the edges of the network an N is the number of nodes in the network. For a complete network i.e., for a network in which all the nodes are connected with each other, the value of density is 1 [3].

1.3.5 Community Structure

A community is a set of entities which are linked to all the other entities in the network. The entities in one community perform the same function and share some common properties. A community structure reveals the internal organization of the nodes. Different communities combine to form a complex network.

Networks like biological networks and social networks reveal modular structure [22]. These structures exhibit more connections within a community than between different communities. The model proposed by Girvan and Newman was the first model which generated networks with the property of community structure.

1.4 Mutual Friend Crawling

Mutual friend crawling (MFC) was introduced by Blenn N. et al. . This approach crawls all the nodes of a network in such a way that all the communities are visited one after another. This algorithm assumes the knowledge of the degree of the neighbouring nodes which is a very difficult task in online social networks. The communities found using mutual friend crawling are smaller as compared to those found by other methods but the properties are same[16]. MFC is generally based on breadth first search algorithm with two differences. The first difference is that a map is used to store all the discovered nodes. The second difference is in the way the next node is chosen to crawl the network. For this, a reference score is calculated[13]. Reference score of a node is given by the ratio of number of discovered nodes to all the nodes which are linked to it. The list of all the discovered nodes is prepared and the one having highest value of reference score is being processed next [13].

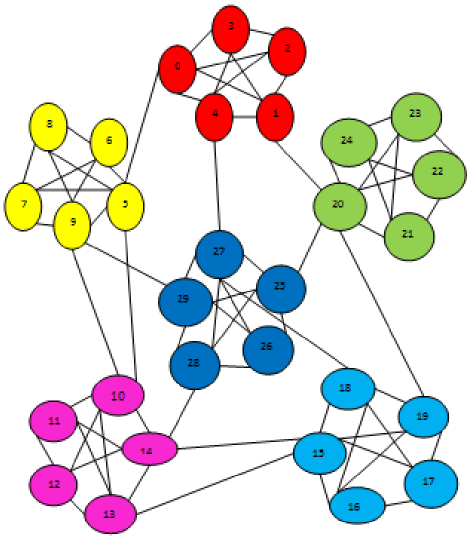


Figure 1: Mutual Friend Crawling

The Fig. 1 shows that an entire community is crawled before crawling other connected communities. Nodes are labelled based on the order of traversal during the crawling process. Different colours of the nodes denote different communities.

This technique first crawls the nodes on that side of the network where more links have been detected. If the community structure is based on the definition that links in a community are more than the links in between communities then nodes in a network are visited one after another. This technique is also applicable on weighted graphs. The only difference is how reference score is calculated. For weighted graphs, reference score is given as the ratio of weights of found references to a node and the strength of that node [13].

II. Literature Survey

Blenn N. et al. (2012) [13], introduced a new way of crawling large online social networks. This technique was named mutual friend crawling. It is compared with standard methods of crawling using breadth first search and depth first search. This was the first analysis of crawling toward community structure. In this method, the communities can be analyzed by the researchers even when the crawling process is running. In the proposed algorithm, the communities are crawled one after another. As compared to breadth first crawling and depth first crawling, this algorithm considers the degree of nodes for the crawling purpose. This algorithm uses a parameter called reference score for selecting the nodes to be crawled. In the results it is found that the performance of mutual friend crawling is better than BFS and DFS. The comparison of the proposed algorithm is done with fast and greedy modularity maximizing method, random walk method and the Louvain method. For this, three datasets have been used which are Zachary's karate club, Girvan and Newman's American college football games and the Digg network. These methods are compared in terms of modularity and a metric which defines the community structure representation of the network in graphical form and Jaccard similarity index. Future work is required in terms of existence of overlapping communities in the network. Also more realistic results are required to get accurate performance of the algorithm.

Gjoka M. et al. (2011) [17], introduced a practical framework for obtaining users of any online social network using its social graph. Various crawling techniques have been studied like breadth first sampling and random walk sampling. Both of these techniques result in a bias towards high degree nodes. They hay also contributed for the use of formal convergence diagnostics which help us to determine when to stop sampling even in the absence of ground truth. A comparison has been done between re-weighted random walk and metropolis-hastings random walk. This comparison is made in terms of bias and efficiency. The results show that re-weighted random walk gives better performance than metropolis-hastings random walk in terms of efficiency. The implementation of high performance distributed crawler faced many challenges when applied to an online social network because of which the data collection time has been kept very small. An offline comparison is made between the ground truth and different crawling methods. Facebook has been used as the case study and the implementation of these methods on it shows the structural properties of the users. A uniform sample of facebook users is obtained which can be used for further research on crawling. The privacy awareness of facebook has also been studied and the results show that about 84% of the users remain their settings unchanged.

Orman G. et al. (2011) [15], reviews the properties which describe the community structure of the complex networks. The LFR benchmark has been studied and steps have been taken to improve the realism of the networks generated by this benchmark. After the generation of networks, the community detection algorithms have been applied and a study of community structure generated by these algorithms is done to compare the results. The properties of the generated network has been analyzed out of which community size and hub dominace are realistic properties, embeddedness and average distance are partially realistic properties and scaled density is non-realistic property. The study is done on undirected and un-weighted networks and has no overlapping communities. The observations show that the communities generated do not posses all the properties as a real network does. The small communities are very dense but they should be star-shaped and the large communities are not dense which shows the opposite characteristic of real world network. Different algorithms like Louvain method, fast and greedy, Markov cluster, infomap etc. have been applied to the generated network and their performance has been analyzed in terms of normalized mutual information. Some of these algorithms have generated communities wth very small size and even having only single node. Infomap has shown the best results while fast and greedy shows the least results.

Bas Van Kester (2011) [16], focuses on the problems of crawling online social network i.e., the crawling process is very time consuming and also analyzing the networks through different community detection methods is very expensive. A new crawling method called mutual friend crawling has been proposed which allows the researchers to analyze the networks even before the crawling process is complete. The proposed method is compared with the existing crawling methods like breadth first crawling and depth first crawling. Two ways of obtaining social networks have been discussed which are networks which are constructed manually by the researchers and online social networks. A brief overview of commonly used community detection techniques is given like edge betweenness clustering used for small online social networks, label propagation used for large online social networks, spinglass and fast and greedy community detection used for networks with different sizes. Different measures like modularity, jacard index etc. have been used for comparison of community detection algorithms. The datasets used are football network by Girvan and Newman, Enron email network and computer generated cluster networks. The results of the algorithms discussed are compared with the ground truth and it is seen that the spinglass community detection algorithm and the ground truth are most similar than other algorithms. But this algorithm is more expensive than the others. The results show that the proposed mutual friend crawling algorithm performs better than the exiting breadth first and depth first crawling techniques. Future work is required for overlapping communities and performance of very large online social networks needs to be analyzed.

Kumar S. et al. (2010) [23], has collaborated the concept of web crawler and web mining. The internet is increasing at a very fast rate and so the process of web mining has become more important. The different types of web mining have been discussed in detail. Web usage mining is a very recent topic of research which estimates the behavior of the user when it is interacting with the web. A brief explanation has been given about web content mining. Research has been done to find the type of data on which different mining techniques are applicable. The use of different mining techniques like web structure mining, web content mining etc. has also been detailed. The use of web crawlers in search engines is also given along with architecture of web crawler which can be used by different users to extract web pages related to their topics of interest. The different types of sources from which data is collected for the purpose of web mining are also explained.

Saramaki J. et al. (2007) [30], present the different generalizations on clustering coefficient which is a very important property of complex networks. The different equations used for calculating clustering coefficient, by different authors have been studied in detail. Also, their advantages and disadvantages have been discussed which provide an idea of which method is appropriate according to requirement. The main focus has been laid on the weighted clustering coefficient. According to the study, it is required that the weighted clustering coefficient should also consider the weight present in the neighborhood of a particular vertex. The different definitions have been studied on two networks i.e., international trade network and scientific collaboration network. The results conclude that the topological nature of the network can be analyzed from weighted clustering coefficient. The number of edges in the neighborhood can be measured but to calculate the amount of weight in the neighborhood is still a question left for future research.

Latapy M. and Magnien C. (2008) [28], has introduced the first practical way to check the properties of the complex networks with the growth in its graph. The changes in the properties have been observed till stability is reached. Their approach has been used on real world networks which contradicts the previous assumption used by different authors for measuring the properties of complex networks i.e., to collect data from complex networks in proportion to the measuring ability and time limit of the methodology used. The amount of data which was extracted was considered as representing whole of the complex network. Several datasets have been used in this study like the inet dataset, the web dataset, the ip dataset and the p2p dataset. The properties studied are size of the network, average degree, average density, clustering, degree distribution and transitivity. In the results for the four datasets used, it has been observed that some properties are stable and some are unstable as a result of which there is no inter-dependency among the properties. The graphical representation for all these properties has been shown for the four datasets used in the paper. They depict clear variation among the values of the properties like for inet dataset the average degree graph shows that the values increase with the increase in the size of the network whereas for the web dataset, stability is observed for average degree. Therefore it has been concluded that for evaluation on small part of the network, if the property is stable then same behavior can be assumed for the whole of the network.

Pavalam S. M. et al. (2011) [18], has reviewed the different web crawling algorithms being used for crawling data from different networks. The type of web crawling algorithm should be chosen according to the requirement of the search procedure. The process of web crawling has been described which includes the selection of target web pages, selection of a seed node from where the crawling process is to start, priorities of different web pages, revisited web pages etc. different algorithms like breadth first search, depth first search, random walk, best first search, genetic algorithm etc. have been detailed. The use of these algorithms and the strategy followed by them is reviewed so that the readers can easily identify the algorithm which best suits its requirements. The advantages and disadvantages of the algorithms are also mentioned in the survey. The use of these algorithms by different authors is also reviewed. It has been concluded that the genetic algorithm gives best results among all the other algorithms discussed.

Rosvall M. et al. (2007) [31], have presented an article related to the community structure of complex networks. A basic framework has been designed for identifying the communities of a complex network. The graphical working of this framework has also been described. A signaler has been used in the framework which has complete knowledge about the network and divides the network into different modules so that more and more information can be extracted from the network. Many social networks have been used to apply their approach and evaluate the results. The results are then compared with the previous approaches that have been adopted by many authors. The designed approach is very useful for the networks in which the communities are of same size and degree. To check the performance different benchmark tests have been applied.

Reichardt J. and Bornholdt S. (2006) [33], have used weighted and directed networks for their work. They have designed a general framework which can be used for solving community detection problems. This framework has been particularly developed for finding community of a desired node of the network rather than finding all the communities of the network. Work has also being done on studying the overlap in the community structure of the network. The framework has been very helpful for large networks as finding communities of such a network is a very time consuming task. Two measures have been used by the framework which can be used for comparison with other algorithms implemented by different authors for community detection in complex networks. The performance of the framework has been evaluated by using benchmarks. The relationship between graph portioning and community detection has also been focused. Different definitions of community have been found by studying the literature by different authors. The properties of the networks which can be evaluated from these definitions have also been discussed. The co-authorship network has also been studied.

Danon L. et al. (2005) [36], have focused on different methodologies which are used for community detection in complex networks. Identifying the communities of a network and then characterizing them is a very important task in community detection. Much progress has been achieved in this field. But still, many issues exist such as the implementation of such techniques remains very costly. The method of community detection should be chosen based on the requirements of the authors. Some authors consider accuracy as the basic requirement and some consider the concept of overlapping communities as the basic requirement etc. Different methods have been compared considering the cost that is incurred on the application of such methods on complex methods. Emphasis has been laid on the different factors which should be considered for choosing an algorithm. Some criteria like accuracy, sensitivity, computational effort, speed etc. are very important while selecting a method for community detection. Speed has been mostly considered as a factor for when applying a method on large networks. It has been concluded that more efforts are required to find a method which is fast enough in detecting communities in complex network.

III. Proposed Technique

The proposed technique will use fuzzy based clustering to obtain best communities therefore, is also detecting communities during runtime. The proposed Fuzzy Based Improved Mutual Friend Crawling (FMFC***)*** has been designed and tested using MATLAB tool to evaluate the effectiveness of the proposed technique over the existing one.

Algorithm:-

1. Load football dataset.
2. Evaluate the size of data i.e., M for number of rows and N for number of columns.
3. Define complex network nodes i.e., N=$\sqrt{M+N}$

4. Define edges for nodes

K=N/3

Therefore, network is 33% connected.

5. Evaluate triangular membership function from football dataset and store them in y_1, y_2 and y_3 respectively.
6. Evaluate fuzzy relationship value

$$\mathrm{R} = \frac{(\mathrm{N}+1)(\sum y2 \; + \; \sum y3)}{(1+\sum y1)}$$

7. Construct a queue Q.
8. Insert the starting node in Q.
9. Store the initial node and 0 as the number of found references in R
10. While Q is not empty do
11. For all elements in R calculate the reference score
12. reference_score = value in R/degree of the node
13. max_score = max(max_score, reference_score)
14. end for
15. select the element having maximum score
16. next_node ← dequeue element having max_score from Q
17. delete next_node from R
18. if next_node has not been visited yet then
19. add all the neighbors of next_node to queue
20. increment number of found references to neighbor by 1and store it in R
21. remember that node (next_node) was visited
22. end if
23. end while

IV. Experimental Evaluation

4.1 Results of Existing Technique

In this section, i present the results obtained by implementing the technique called mutual friend crawling in MATLAB. For this, the dataset used is “American College Football Games” of the year 2000 and is gathered from the internet. The nodes in the network represent the different teams and the edges show the link between the teams. There exists a link between the teams if they have played a match against each other.

To check the performance, three parameters have been considered i.e., execution time, overheads and number of clusters. Also, the properties of complex networks i.e., average path length, average clustering coefficient and average degree distribution has been studied.

Table 1: Evaluation of the properties of complex network from a well known dataset

Average Path Length (APL)	Average Clustering Coefficient (ACC)	Average Degree Distribution (ADD)	Execution Time
1	1	14	1.5358
1.2857	0.68889	10	2.0030
1.7143	0.6	6	2.3252
1.7143	0.6	6	2.6526
2.2857	0.5	4	2.9919

The Table 1 shows the different values of the properties along with the execution time. The properties derived are average path length, average clustering coefficient and average degree distribution. It can be seen very clearly that the value of average path length is increasing with the decrease in the density of the network. Also, the value of average clustering coefficient and average degree distribution is decreasing with the network becoming less dense.

In the results, the complex network with different values of APL, ACC and ADD is shown in figures 2, 3, 4 and 5. With the changing values of the properties, the density of the network also changes. The nodes are represented with pink colour showing different teams in the dataset and the links between them is represented with blue colour edges.

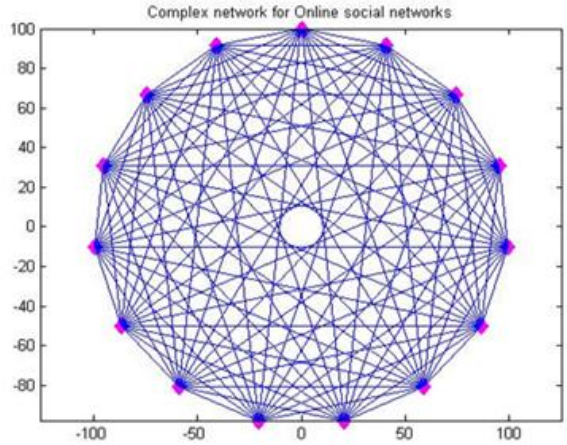


Figure 2: Complex Network with APL=1, ACC=1 and ADD=14

Fig. 2 is the most dense network. The nodes in pink are the different teams of the dataset which are linked to each other edges in blue colour. The average path length for this graph is 1, average clustering coefficient is 1 and the average degree distribution is 14.

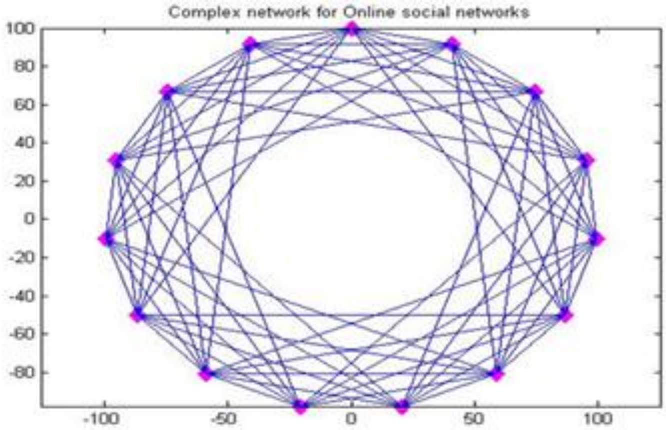
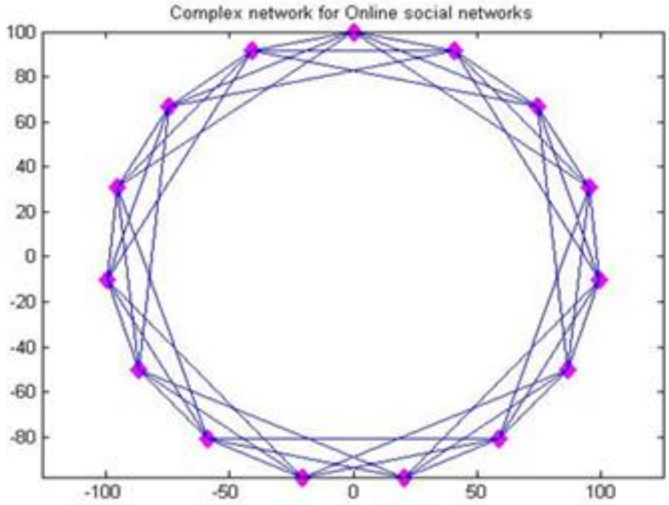


Figure 3: Complex Network with APL=1.2857, ACC=0.68889 and ADD=10

Fig. 3 is less dense as compared to Fig. 8 with the varying values of the properties. Also the execution time (0.8134) is observed to be more than the execution time for the previous graph (0.4806) as shown in the Table 1



.

Figure 4: Complex Network with APL=1.7143, ACC=0.6 and ADD=6

Fig. 4 is less dense than previous graphs Two such graphs are observed.. The value of the properties for both the graphs is observed to be same but the execution time varies as shown in the Table.

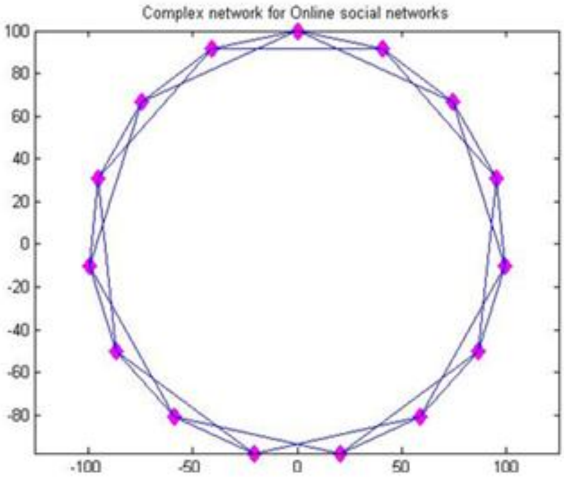


Figure 5: Complex Network with APL=2.2857, ACC=0.5 and ADD=4

Fig. 5 is the least dense graph. It has been observed that the value of degree distribution is least and the value of average path length is the highest for this graph. The execution time also comes up to be the highest for this graph.

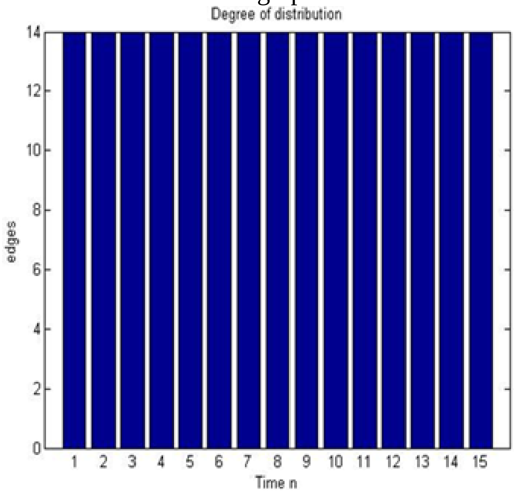


Figure 6: Degree of Distribution=14

Fig. 6 shows the resultant graph of degree distribution for the network in Fig. 2. The degree of distribution is observed to be the highest for this network.

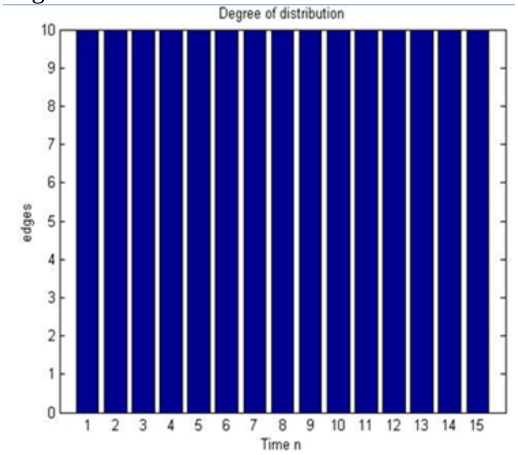


Figure 7: Degree of Distribution=10

Fig. 7 shows the graph of degree distribution for the network in Fig. 3.

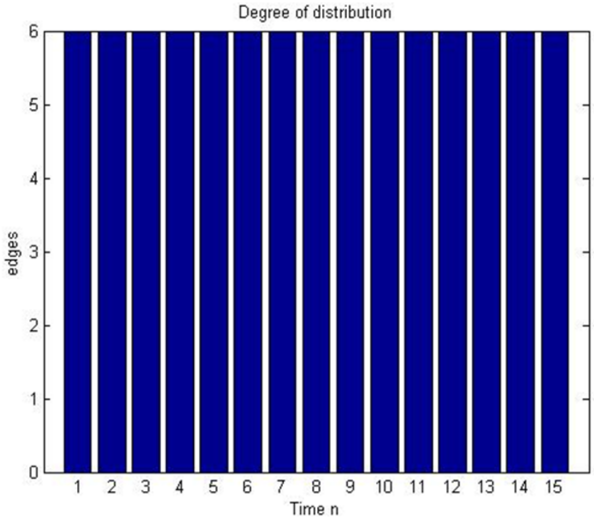


Figure 8: Degree of Distribution=6

Fig. 8 shows the graph of degree distribution for the network in Fig. 4. Here again two similar graphs are observed for the same value of degree of distribution.

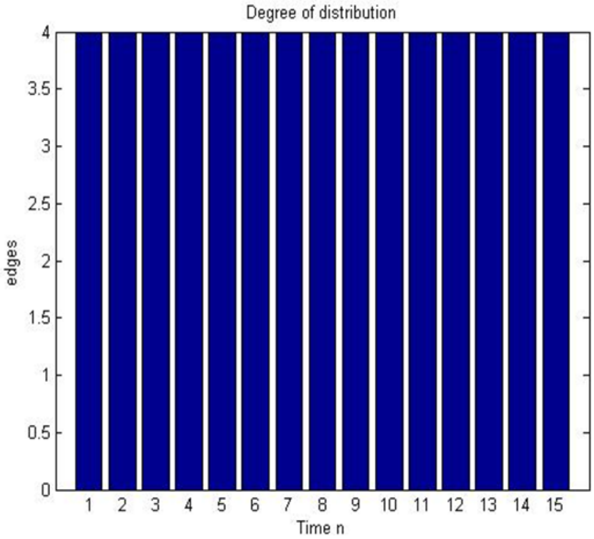


Figure 9: Degree of Distribution=4

Fig. 9 shows the graph of degree distribution for the network in Fig. 5. The degree of distribution is observed to be the least for this network and the execution time is the highest.

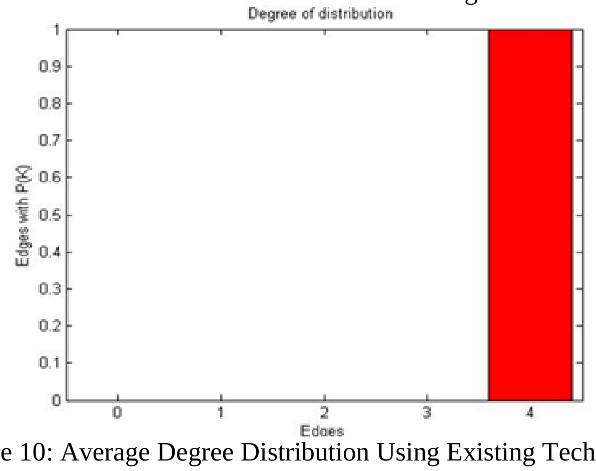
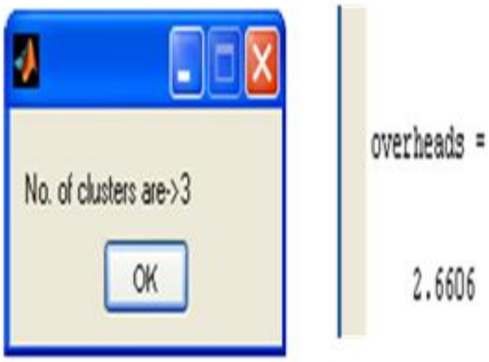


Figure 10: Average Degree Distribution Using Existing Technique

Fig. 10 shows the graph of average degree distribution for the whole network. P(k) is the probability of the number of nodes having degree ‘k’.



No. of clusters are->3

```
overheads =

    2.6606
```

Figure 11: Number of Clusters and Overheads Using Existing Technique

4.2 Results of Proposed Technique

The following figures are the results of the proposed technique i.e., fuzzy based improved mutual friend crawling.

Table 2: Evaluation of the Properties of Complex Network by Applying Fuzzy Logic in Mutual Friend Crawling

Average Path Length (APL)	Average Clustering Coefficient (ACC)	Average Degree Distribution (ADD)	Execution Time
0.85	0.85	2.15	0.6090

The Table 2 shows the values of average path length, average clustering coefficient, average degree distribution and execution time found using fuzzy logic in the existing technique. The proposed technique i.e., FMFC makes use of fuzzy membership function to generate the network of the football dataset. The different values derived are shown in this Table.

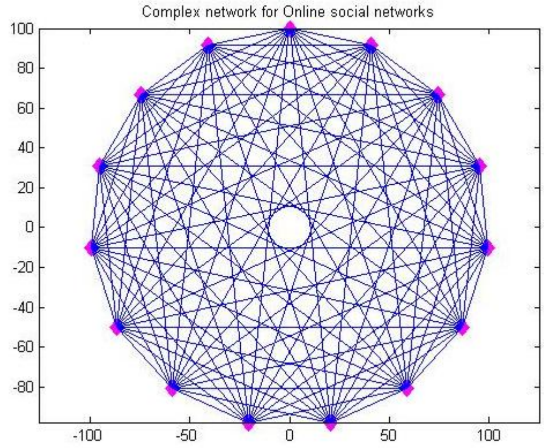


Figure 12: Complex Network of Football Dataset

This is the graph of complex network for the football dataset which is similar to the previous graph generated using the existing technique. After applying fuzzy logic in mutual friend crawling, no change is found in the representation of the network. It consists of 15 nodes which represent the communities of the network.

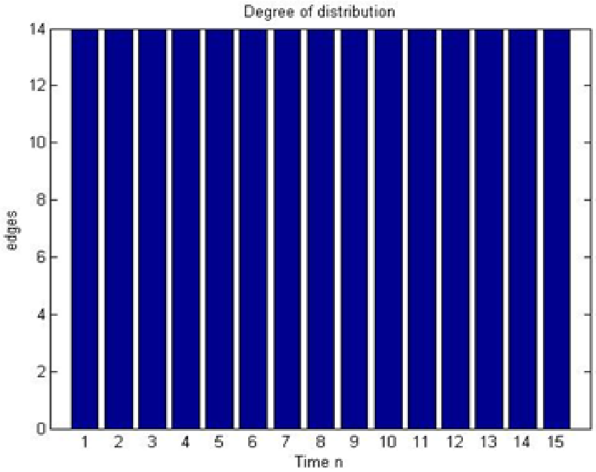


Figure 13: Degree of Distribution of Football Dataset

The Fig. 13 shows the distribution of the network having degree 14. The x-axis represents the time of distribution of the edges and the y-axis shows the edges.

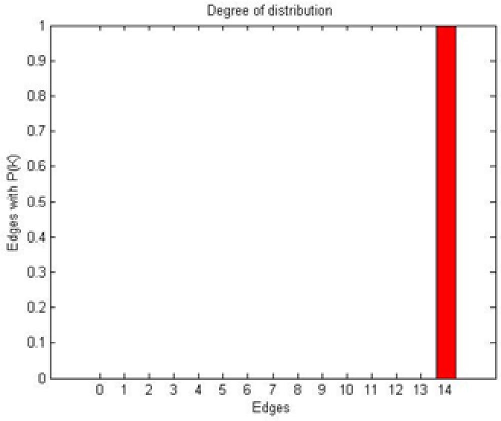


Figure 14: Average Degree Distribution Using Proposed Technique

`Fig. 14 shows the graph of probability of degree distribution of the network. This graph shows better distribution by using fuzzy logic as compared to the previous graph of existing technique. The existing technique gives distribution at degree 4, which is the least degree for the network, whereas the proposed technique gives distribution at degree 14 which is the maximum degree of the network.

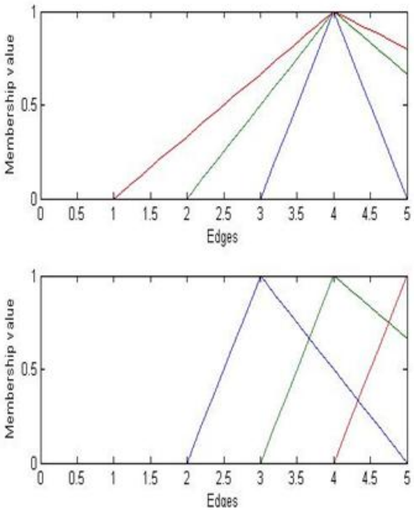
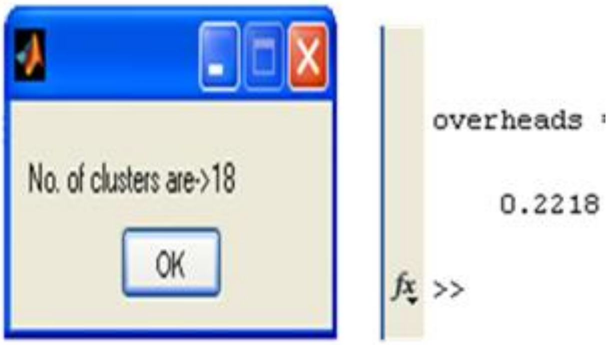


Figure 15: Membership Value of Edges



No. of clusters are->18

```
overheads =

    0.2218

>>
```

Figure 16: Number of Clusters and Overheads Using Proposed Technique

Table 3 Comparison of Existing and Proposed MFC

Sr. No.	Parameter	Existing Technique	Proposed Technique
1.	Number of Clusters	3	18
2.	Overheads	2.6606	0.2218
3.	Execution Time	2.9919	0.6090

The Table 3 shows the comparison between the existing mutual friend crawling technique and fuzzy based improved mutual friend crawling technique. The comparison is based on three parameters i.e., number of clusters, overheads and execution time. The number of clusters found using existing MFC are 3. This is a very small number. Compared to this, the number of clusters using FMFC is 18, which is a more profitable value. The value of overheads for existing MFC is very high i.e., 2,6606 whereas when fuzzy logic is applied in MFC, the overheads are reduced to a value 0.2218. This means we get better performance in fuzzy based MFC. The execution time of existing technique is found to be very high i.e., 2.9919. This means that more time is consumed in crawling the network of the dataset used. Using fuzzy logic, the execution time is reduced to 0.6090. Therefore, less time is used in crawling the network. Hence, using fuzzy logic in MFC proves to be beneficial in terms of number of clusters, overheads and execution time.

V. Conclusion

The different web crawling algorithms are used for crawling data from different networks. The type of web crawling algorithm should be chosen according to the requirement of the search procedure. In this paper the process of web crawling has been described which includes the selection of target web pages, selection of a seed node from where the crawling process is to start, priorities of different web pages, revisited web pages etc. Different algorithms like breadth first search, depth first search, random walk, best first search, genetic algorithm etc. are used in the literature. The use of these algorithms and the strategy followed by them is reviewed so that the readers can easily identify the algorithm which best suits its requirements. Also, the different properties of complex network like assortativity, community structure, degree distribution, clustering coefficient, density etc have been studied and explained in detail.

VI. Future Scope

This work has not considered the use of genetic algorithm to enhance the results further, but many researchers has shown that the genetic based web crawling technique outperforms over the available techniques. Therefore in near future a genetic algorithm based mutual friend crawling will be proposed to enhance the best community detection rate.

References

[1] Yadav A. and Singh P., "Web Crawl Detection and Analysis of Semantic Data", International Journal of Computer Trends and Technology, Volume 21, Number 1, ISSN: 2231-2803 (2015).

[2] Ahuja M., Singh J. and Varnica, "Web Crawler: Extracting the Web Data", International Journal of Computer Trends and Technology, Volume 13, number 3, ISSN: 2231-2803(2014).

[3] Mini Singh Ahuja and Jatinder Singh, "Future Prospects in Community Detection", International Journal of Computer Science Engineering and Information Technology Research, Vol. 4, Issue 5, ISSN: 2249-6831 (2014).

[4] Mcphail M., "The Statistical Properties of Complex Networks", Australian Mathematical Sciences Institute (2014).

[5] Janbandhu R., Dahiwale P. and Raghuwanshi M., "Analysis of Web Crawling Algorithms", International Journal on Recent and Innovation Trends in Computing and Communication, Volume 2, Issue 3, ISSN: 488-492 (2014).

[6] Kumar R., Jain A. and Agrawal C., "Survey of Web Crawling Algorithms", Advances in Vision Computing: An International Journal, Volume 1 (2014).

[7] Yan H. et al., "Community detection using global and local structural information", Pramana – J. Phys., Vol. 80, No. 1 (2013).

[8] Iswary R. and Nath K., “Web Crawler”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, ISSN: 2278-1021 (2013).

[9] Yang J. and Leskovec J., “Defining and Evaluating Network Communities Based on Ground-Truth”, Springer, DOI 10.1007/s10115-013-0693-z (2013).

[10] Orman G., Labatut V. and Cherifi H., “Towards Realistic Artificial Benchmark for Community Detection Algorithms”, International Journal of Web Based Communities, pp. 349-370 (2013).

[11] Nath R. and Chopra K., “Web Crawlers: Taxonomy, Issues and Challenges”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, pp. 944-948 (2013).

[12] Kausar M., V. S. Dhaka and Singh S., “Web Crawler: A Review”, International Journal of Computer Applications (0975-8887), Volume 63, Number 2 (2013).

[13] Blenn N., Doerr C., Kester B. and Mieghem P., “Crawling and Detecting Community Structure in Online Social Networks using Local Information”, Springer, LNCS 7289, pp. 56-67 (2012).

[14] Khurana D. and Kumar S., “Web Crawler: A Review”, International Journal of Computer Science & Management Studies, Vol. 12, Issue 01, ISSN: 2231 –5268 (2012).

[15] Orman G., Labatut V. and Cherifi H., “Qualitative Comparison of Community Detection Algorithms”. International DICTAP 2011, DOI 10.1007/978-3-642-22027-2_23 (2011).

[16] Bas Van Kester, “Efficient crawling of community structures in online social networks”, PVM 2011-071, Tu Delft (2011).

[17] Gjoka M., Kurant M., Butts C. and Markopoulou A., “Practical Recommendations on Crawling Online Social Networks”, IEEE Journal on Selected Areas in Communications, Vol. 29, No. 9 (2011).

[18] Pavalam S. M., Raja S., Akorli F. and Jawahar M., “A Survey of Web Crawler Algorithms”, International Journal of Computer Science Issues, Volume 8, Issue 6, ISSN: 1694-0814 (2011).

[19] Meo P., Nocera A., Terracina G. and Ursino D., “Recommendation of Similar Users, Resources and Social Networks in a Social Internetworking Scenario”, Information Sciences, Volume 181, Number 7, pp. 1285-1305 (2011).

[20] Olston C. and Najork M., “Web Crawling”, NOW The Essence of Knowledge, Vol. 4, No. 3 (2010) 175–246 (2010).

[21] Coscia M., Giannotti F. and Pedreschi D., “A Classification for Community Discovery Methods in Complex Networks”, Wiley Online Library, Volume 4, DOI:10.1002 (2010).

[22] Fortunato S., “Community Detection in Graphs”, Physics Reports, 486(3-5):75 (2010).

[23] Kumar Pani S., Mohapatra D. and Keshari Ratha B., “Integration of Web Mining and Web Crawler: Relevance and State of Art”, International Journal on Computer Science and Engineering, Volume 2, No. 3, 772-776 (2010).

[24] Lancichinetti A. and Fortunato S., “Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities”, Physical Review E80, 016118 (2009).

[25] Lancichinetti A. and Fortunato S., “Community detection algorithms: A comparative analysis”, Physical Review E 80, 056117 (2009).

[26] Orman G. and Labatut V., “A Comparison of Community Detection Algorithms on Artificial Networks”, Discovery Science, Porto : Portugal, DOI : 10.1007/978-3-642-04747-3_20 (2009).

[27] Lancichinetti A., Fortunato S. and Radicchi F., “Benchmark graphs for testing community detection algorithms”, Physical Review E 78, 046110 9 (2008).

[28] Latapy M. and Magnien C., “Complex Network Measurements: Estimating the Relevance of Observed Properties”, IEEE INFOCOM, Phoenix, USA, pp. 2333-2341 (2008).

[29] Blondel V., Guillaume J. and Lambiotte R., “Fast Unfolding of Communities in Large Networks”, Journal of Statistical Mechanics: Theory and Experiment, P10008 (2008).

[30] Saramaki J., Kivela M., Onnela J., Kaski K. and Kertesz J., “Generalizations of the Clustering Coefficient to Weighted Complex Networks”, Physical Review E 75, 027107 (2007).

[31] Rosvall M. and Carl T. Bergstrom, “An Information-Theoretic Framework for Resolving Community Structure in Complex Networks”, The National Academy of Sciences of the USA, Volume 104, No. 18, 7327-7331 (2007).

[32] Boccaletti S., Latora V., Moreno Y., Chavez M. and D.-U. Hwang, “Complex Networks: Structure and Dynamics”, Elsevier, Physics Reports 424, DOI: 10.1016, 175-308 (2006).

[33] Reichardt J. and Bornholdt S., “Statistical Mechanics of Community Detection”, Physical Review E74, 016110 (2006).

[34] Trusina A., “Complex Networks: Structure, Function, Evolution”, Umea University, ISBN: 91-7305-924-2 (2005).

[35] Cotta C. and Merelo J., “The Complex Network of Evolutionary Computation Authors: an Initial Study”, arXiv: physics/0507196v2 (2005).

[36] Danon L., Diaz-Guilera A., Duch J. and Arenas A., “Comparing Community Structure Identification”, Journal of Statistical Mechanics: Theory and Experiment, DOI: 10.1088/1742-5468/2005/09/P09008 (2005).

[37] Pant G., Srinivasan P. and Menczer F., “Crawling the Web”, Springer Berlin Heidberg, ISSN: 978-3-662-10874-1 (2004).

[38] M.E.J. Newman, “The Structure and Function of Complex Networks”, SIAM Review 45, 167-256 (2003).

[39] Girvan M. and M. E. J. Newman, “Community Structure in Social and Biological Networks”, National Academy of Sciences of the United States of America, Volume 99, Number 12, pp. 7821-7826 (2002).

[40] Wang Z., “Community Detection Approaches In Complex Networks: A Review”, Department of Applied Mathematics, Fudan University.

[41] Varnica and Mini Singh Ahuja, “Studying the Properties of Complex Network Crawled Using MFC”, International Research Journal of Engineering and Technology, Volume 2, Issue 4, e-ISSN: 2395-0056 *(2015).*