Overlapping Community Recognition in Social Network using k-LEPSO Algorithm

Pratiksha V. Raut, Archana S. Vaidya

Abstract —In data mining domain searching the overlapping communities is a pivotal problem. Using particle swarm optimization it is expected to provide number of small community structures. The main LEPSO algorithm is play an important role in this whole system. LEPSO broadly justify with three methods. These three methods are particle swarm optimization, line graph theory and ensemble learning. Then it will generate the neighbour node list for each and every node. Optimization potentiality is improved by using the clustering. It will also helpful for detecting the undetectable communities in social network. After that we can apply post processing strategy. Using synthetic database and real world data overlapping communities are detected. Here the main performance measures are time, accuracy, precision and recall. As we compare the existing system with the proposed system produced the overlapping communities in less time.

Index Terms —Social network, line graph, overlapping Communities detection, particle swarm optimization (PSO), Ensemble learning, k means clustering.

I. Introduction

Remarkable growth is taking place in social network from few years. Community is known as group of users. Community mining technique is mostly applied, to the multimedia related application features such as commenting on photos, recommendation, user modelling, and advertising. A social network is basically represented with the help of graphs. Graph relating the nodes and edges. Node basically represent the entity. Edges represents the communication or interaction between nodes. Here we get the subgraph of overlapping community.

The LEPSO algorithm is used in this system. LEPSO algorithm is meta huristic in nature. These three methods are first is ensemble learning, second is line graph theory and third is particle swarm optimization. Optimization strategy plays an important role in this system for enhancing this the ensemble learning is needed. Here we give the social networking graph as an input to the system so it uses theory of line graph. The overlapping community optimization is the key objective of this paper. The community detection is the NP hard kind of problem.

Overlapping community recognition is very difficult task when we handle the large network. So this give the better solution for this problem. This system also provide the output within less time. So this system require less computing time and this is the main advantage of this system.

II. Literature Review

Multiple methods for detecting the overlapping communities are implemented. Some of them are discussed here.

Ming Cheung et al. [3] presented the uses and steps that how we can generate the graph like tags and all. We know that recommendation use for the extract information of data and show the best results using optimization path in between the network. Using images also we can get the information e.g. social media the most popular example as we know.

Shangrong Huang et al. [4] presented the concept of the network correlation i.e. use in graph generation. In every data mining concept, we know that the help of feature selection and feature extraction is useful for recommendation system for our self and other people also. If any user is satisfied with particular system, then he frankly recommends that or pass the comments to other person.

M. E. J. Newman et al. [5] proposed the how one network will get divide into sub graphs for simplicity of the overall system. And in that process it removes the repeated part for optimization purpose like extra nodes, edges will get removed because of unnessiarily.

Qing Cai et al. [6] presented the how graph using network will get formed and how we can see the generation of problem detection. For that developed the methods which are useful for that for better result. And also for network communities of generation.

Marek Lipczak et al. [7] Proposed the idea of detection of communities using some clustering algorithm e.g. ACGA algorithm. Using that cluster will get generated and we will get one perfect solution of the problem. We know the concept of actual working of the clustering algorithm that is make a cluster using neighbouring.

Maoguo Gong et al. [8] proposed the how we can improve the efficiency of any product while mining the data and extract the information from it. K means clustering is also the best example for this regarding to forms the cluster. All this is based upon the real world data.

Clara Pizzuti [9] proposed the genetic algorithm working for the partition purposed. Using the real world data it is helpful to solve the complexity of the inner side problem. And also check the interconnection network problems.

Ronghua Shang et al. [10] presented the modularity detection problem using improvement of genetic concepts. Using adjustment of the parameter MIGA concept will get used for annealing search for improve the stability and accuracy.

Jianwu Li et al. [11] presented the ECGA concept for statistical learning to solve the complex problems coming in the network. May be that network is linear or not bust still it is helpful to overcome it and solve

the problem.

Ahmed A. A. Esmin et al. [12] proposed the clustering data information i.e. how will data classified form labelled and unlabelled data into their respective groups. PSO algorithm will get use over here for handling the high dimensions.

Mattia Broilo et al. [13] presented the feedback system using their relevance and also useful to solve the gap issues in between the graph solutions avoid the spacing some complexities solve that problem using better and better solution e.g. high dimensionalities of the process while generating the network.

Hsien-Hsin Chou et al. [14] proposed the PSO concepts for optimizing parameters when fuzzy logic comes highly up. Here the point is clear and that is noise rate must be above 0.5.TPSO is highly use to improve the parameters.

James Kennedy et al. [15] presented the how to resolve the binary variable generation problem to optimize techniques using reviews.

Ashish Sureka et al. [16] proposed the observation of information to solve the difficulties in between the task. And also it depends upon the social networking analysis for mining the information their also some hidden task.

Duan xiaodong et al. [17] presented the network community detection problem using wide range area network information in it. It mainly operates on the dolphin and kerbs dataset for the better quality results.

III. System Overview

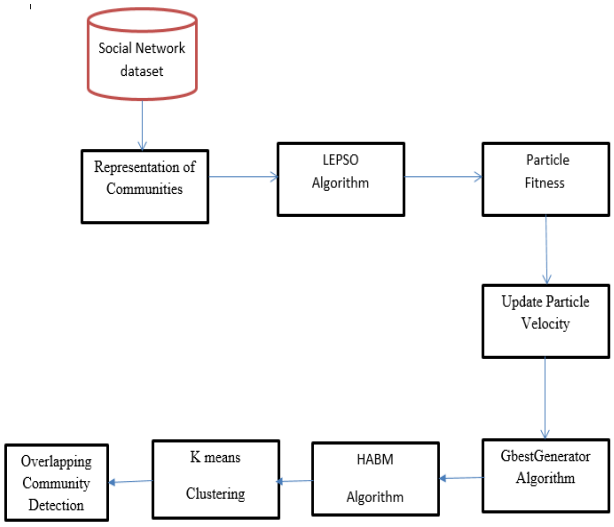

Figure.1. Block Diagram of Proposed System

Social network data set is given as an input the system. After uploading the dataset the proposed system find outs the neighbours of each and every node. The output is the neighbour node list. This is done at representation of communities block. Subsequently LEPSO algorithm is used for finding overlapping communities. This algorithm contains three subparts. First is finding particle fitness, second is Gbest generator and third is HABM algorithm.

Particle swarm optimization algorithm is implemented in first phase. Here we have to find the particle fitness value for all the nodes. Then algorithm will update the particle velocity. Then PSO algorithm gives the best min and best max values. These min and max value are passed to the gbest algorithm. This algorithm provides the value global best. In the last phase the HABM algorithm is applied. It provides the overlapping communities. This algorithm is proposed by authors. We have modified this algorithm by applying clustering in subsequent phase of above given algorithm. We have added the hybrid clustering for getting the overlapping communities in less time. This hybrid clustering is based on the k means clustering.

A. Algorithms

Here we have to use the three concepts are as follows.

- ✓ LEPSO Algorithm
- ✓ Gbest generator algorithm
- ✓ HABM algorithm
- ✓ k means clustering

- LEPSO Algorithm:
  LEPSO broadly justify with three methods. These three methods are particle swarm optimization, line graph theory and ensemble learning. The initial stage is to give the data of each node with their neighbour list. Then it will apply the particle swarm optimization algorithm on it. Due to this algorithm it simply generates the pbest and gbest values.

- Gbest generator algorithm:
  These algorithm is the second phase of this system. The values which we get in the above algorithm that we will use as an input to this algorithm. This gbest generator algorithm will produce the gbest value among the all values. Here each and every particle have different fitness values. Here every particle have different velocity.

- HABM algorithm:
  This is the third phase of the system. HABM stands for hierarchical agglometric bottom up merging strategy. So this algorithm will give us the overlapping communities among the social network.

- k means clustering :
  Clustering is basically simplest form of clustering. k means clustering uses an iterative refinement method. It require the initialization of k value. Then find the clustures according to mean value.

IV. Results

A. Performance Measures

Here basically I take the four datasets for the overview. This datasets are dolphin, Zachery, karate and email respectively. So this performance measures are observe with respect to this four datasets.

There are four performance measures used:

- Time efficiency
- Precision
- Recall
- F1 measures

B. Results

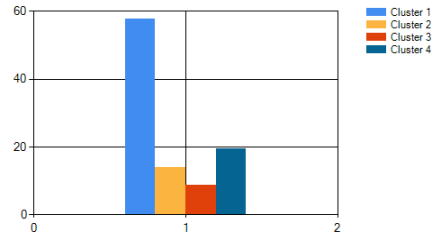

Figure 2. Percent of Clusters.

In this above graph shows the values of clusters which are formed after applying k means clustering on it.

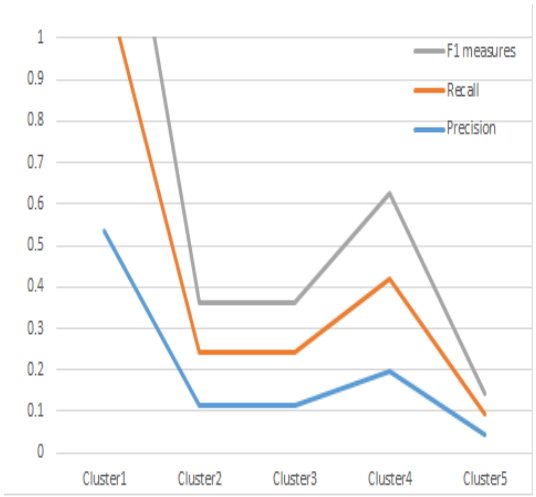

Figure.3. Precision Recall Graph

In this above graph shows the values of precision, recall and F1 measure. These three parameters are very essential for measuring the accuracy of the proposed system.

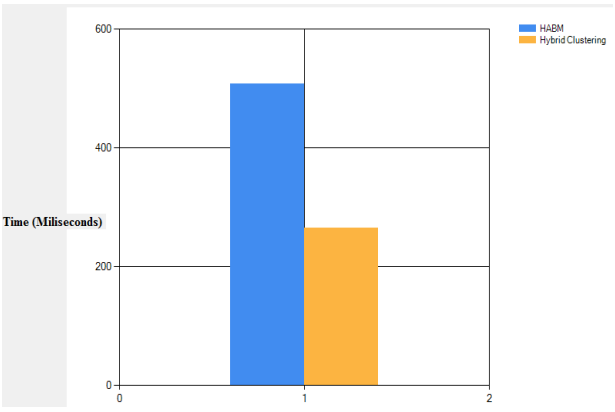

Figure.4. Time Graph

In this above fig.4 shows the time graph. In this time graph we compare existing system as HABM and the existing system with contribution as Hybrid clustering. So here the existing system i.e. HABM required more time to detect the overlapping communities. The new system with contribution i.e. detect overlapping communities in less time than existing system.

Conclusion

In this system it will effectively generates the overlapping communities. Here mainly HABM algorithm is used for community detection. Here we added the contribution is Hybrid clustering. HABM algorithm takes more time as compare to hybrid clustering. Hybrid clustering require less computation time as compare to HABM algorithm. It will also provide the more accuracy than the existing system. The accuracy is measure with respect to precision, recall and F1 measures. In future system can recommend the advertisement or other stuff to user according to their point of interest with the help of clusters. In future system will improve efficiency for handling large-scale network for detecting communities.

Acknowledgment

I have a tremendous pleasure in presenting the project "Overlapping Community Recognition in Social Network using k-LEPSO Algorithm" under the guidance of Prof. A. S. Vaidya and PG coordinator Prof. A. S. Vaidya. I am really obligated and appreciative to Head of the Department Dr. D. V. Patil for their significant direction and consolation. I might likewise want to thank the Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies Research, Nashik-5 for giving the required offices, Web get to and vital books. At last I must express my sincere heartfelt gratitude to all the Teaching Non-teaching Staff members of Computer Department of GESRHSCOE who helped me for their important time, support, remark, thoughts.

References

[1] Faliang Huang, Xuelong Li, Shichao Zhang, Jilian Zhang, Jinhui Chen and Zhinian Zhai "Overlapping Community Detection for Multimedia Social Networks" 2017.

[2] Pratiksha V. Raut, Archana S. Vaidya "A Review on Overlapping Community Detection in Multimedia Social Network" ISSN 2319 –1953, 2018.

[3] M. Cheung, J. She, and Z. Jie, "Connection discovery using big data of user shared images in social media," *IEEE Trans. Multimedia*, vol. 17, no. 9 pp. 1417–1428, Sep. 2015.

[4] S. Huang, J. Zhang, L. Wang, and X.-S. Hua, "Social friend recommendation based on multiple network correlation," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 287–299, Feb. 2016.

[5] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, 2004, Art. no. 026113.

[6] Q. Cai *et al.*, "A survey on network community detection based on evolutionary computation," *Int. J. Bio-Inspired Comput.*, vol. 8, no. 2, pp. 84–98,2016.

[7] M. Lipczak and E.Milios, "Agglomerative genetic algorithm for clustering in social networks," in *Proc.Genetic Evol. Comput.*, 2009, pp. 1243–1250.

[8] M. Gong *et al.*, "A non-dominated neighbor immune algorithm for community detection in networks," in *Proc. Conf. Genetic Evol. Comput.*, 2011, pp. 1627–1634.

[9] C. Pizzuti, "Boosting the detection of modular community structure with genetic algorithms and local search," in *Proc. Symp. Appl. Comput.* 2012, pp. 226–231.

[10] R. Shang *et al.*, "Community detection based on modularity and an improved genetic algorithm," *Physica A, Statist. Mech. Appl.*, vol. 392, no. 5, pp. 1215–1231, 2013.

[11] J. Li and Y. Song, "Community detection in complex networks using extended compact genetic algorithm," *Soft Comput.*, vol. 17, no. 6, pp. 925–937, 2013.

[12] A. A. A. Esmin, R. A. Coelho, and S.Matwin, "Areviewon particle swarm optimization algorithm and its variants to clustering high-dimensional data," *Artif. Intell. Rev.*, vol. 44, no. 1, pp. 23–45, 2013.

[13] M. Broilo and F.G.B.DeNatale, "A stochastic approach to image retrieval using relevance feedback and particle swarm optimization," *IEEE Trans. Multimedia*, vol. 12, no. 4, pp. 267–277, Jun. 2010.

[14] H. H. Chou, L. Y. Hsu, and H. T. Hu, "Turbulent -PSO-based fuzzy image filter with no-reference measures for high-density impulse noise," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 296–307, Feb. 2013.

[15] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct.1997, vol. 5, pp. 4104–4108.

[16] A. Sureka *et al.*, "Mining YouTube to discover extremist videos, users and hidden communities," in *Proc. Asia Inf. Retrieval Symp.*, 2010, pp. 13–24.
[17] X. D. Duan, C. R. Wang, X. D. Liu, and Y. P. Lin, "Web community detection model using particle swarm optimization," in *Proc. Congr. Evol. Comput.*, 2008, pp. 1074–1079

AUTHORS' DETAIL

Pratiksha Vishwas Raut
Student,
Gokhale Education Society's R. H. Sapat college of engineering /Computer Department, Nashik, India
Email: pratiksharaut8888@gmail.com

Archana Suhas Vaidya
Assistant Professor
Gokhale Education Society's R. H. Sapat college of engineering /Computer Department, Nashik, India
Email: Archana.s.vaidya@gmail.com

CITE THIS ARTICLE AS :

Pratiksha V. Raut, Archana S. Vaidya "Overlapping Community Recognition in Social Network using k-LEPSO Algorithm," International Journal of Technology and Science, vol. 6, Issue. 1, pp. 21-24, 2019