A Survey on Overlapping Community Detection for Multimedia Social Networks Using LEPSO and W-CPM

SunitaArunChavan[1], MansiBhonsle[2], Diksha Bhoyar[3]
[1]M.tech Student, [2]Assistant Professor (Guide), [3]Assistant Professor(Co-guide),
Department of Computer Engineering, G.H.R.C.E.M, Wagholi, Pune, India
Savitribai Phule Pune University, Maharashtra, India

Abstract- Traditional community detection, it is more challenging to find overlapping communities in complex networks, especially when the multimedia social networks. However, extant overlapping community discovery with swarm intelligence frequently generates overlapping community structures with extreme small communities. To deal with the issue, in this paper, an efficient algorithm (LEPSO) is proposed for overlapping communities' discovery, which is based on line graph theory, ensemble learning, and particle swarm optimization (PSO). Specifically, a discrete PSO, including an encoding scheme with ordered neighbors and a particle updating strategy with ensemble clustering, is devised for improving the optimization capability to search communities hidden in social networks. Then, a post-processing approach is provided for combining the finer-grained and suboptimal overlapping communities. The contribution work is Weak Cliques Percolation Method (W-PCM) for overlapping community detection, which is a new measure for characterizing the similarity between weak cliques, is also suggested to check whether the weak cliques can be merged into a community.

Keywords- Social Network, Particle Swarm Optimization (PSO), DPSO, Ensemble learning, Line Graph, Overlapping Communities Detection, Clique percolation, Weak Clique.

I. INTRODUCTION

Social media websites, such as Twitter and Facebook, have billions of users sharing opinions, photos and videos every day. Usually, users are provided with various features like reply, comment, subscribe and connect to in order to interact, engage and share information with each other. Such interactions lead to formation of closely knit user groups or densely connected clusters of users around specific topics within the social network; these groups are called communities. Communities discovery is of great importance for understanding the organization and function of social networks, and the extracted communities can be used in various applications such as topic discovery, targeted advertisement, recommendation of multimedia resources such as photos and videos. In reality, social network users are naturally characterized by multiple community memberships. A social network can be modeled as a graph by mapping entities to nodes, and interactions between the entities to edges. Actually, overlapping communities' detection problem can be modeled as computing the optimal cover of graph nodes through optimizing some given objective function. This optimization problem leads to a class of community detection algorithms based on swarm intelligence techniques.

Motivation:

1. Discovering overlapping communities from a social network is equivalent to detecting disjoint communities in the corresponding line graph of the social network.
2. Boosting search ability to find high-quality and finer-grained overlapping communities from social networks.
3. Merge finer-grained and suboptimal overlapping communities for post-processing strategy.

II. RELATED WORK

In this paper investigate the structure of social networks and develop an algorithm for Network Correlation-based Social Friend Recommendation (NC-based SFR). To accomplish this goal, correlate different "social role" networks, find their relationships and make friend recommendations. NC-based SFR is characterized by two key components: 1) Related networks are aligned by selecting important features from each network. 2) The network structure should be maximally preserved before and after network alignment. Advantages are: The NC-based SFR outperforms other methods in friend recommendation. Achieves the highest precision in friend prediction. Disadvantages are: Need to work on network correlation for applications other than friend recommendation. A proposed method for a system [2] to discover and recommend user connections in follower/followee relationships using user shared images directly. Methods using bag-of-features tagging (BoFT) are proposed as a recommendation system to discover user connections and recommend follower/followee relationships by their shared images. Advantages are: The discovered connections are also proven to be able to identify user gender. The follower/followee recommendation using discovered connections by user shared images is possible, and the recommendation is 60% better than UserT. Disadvantages are: SGs is getting more difficult and costly in today's online social networks.

The paper [3] presents a surveyon evolutionary algorithms for network community detection. The evolutionary algorithms inthis survey cover both single objective and multi-objective optimizations. The network models involve weighted/unweighted, signed/unsigned, overlapping/non-overlapping and static/dynamic ones. Advantages are: An EA optimizes the problem by having a population of initialized solutions and then apply stochastic components to generate new solutions in the decision space. High Efficiency. Disadvantages are: Need to analysis on complex networks. In [4] paper, propose an efficient overlapping community detection algorithm using a seed expansion approach. In particular, develop new seeding strategies for a personalized PageRank clustering scheme that optimizes the conductance community score. An important step in method is the neighborhood inflation step where seeds are modified to represent their entire vertex neighborhood. Advantages are: Proposed method is faster than other state-of-the-art overlapping community detection methods. Effective in finding good overlapping communities in real-world networks. Disadvantages are: Need to specify the number of communities for NISE.

In [5] paper, by leveraging the user-venue check-in network and user/venue attributes, proposed a multimode multi-attribute edge-centric coclustering framework to detect overlapping communities for LBSNs users. In this paper, propose a two-step hierarchical clustering algorithm to detect overlapping communities of LBSNs users, where a variant of *k*-means is used as the baseline method. Advantages are: The proposed framework was able to discover high quality overlapping communities. Easy to understand. The company is able to provide better services and/or achieve better benefits. Disadvantages are: Need handle user's several interesting problems. In [6] paper, apply static and dynamic community and node level features to the case of community evolution prediction problem. First, significant features are identified for each separate event happening to the community. Size ratio is the most important feature to predict events happening to communities. Second, results indicate that community fate prediction depends on how to detect communities and dynamics of OCD algorithm. Advantages are: The significance of features for CEP highly depends on how to define and detect communities. More realistic. Disadvantages are: Needs to implement fine-grained algorithms.

The paper [7] proposes a framework for evaluating algorithms' ability to detect overlapping nodes, which helps to assess overdetection and underdetection. After considering community-level detection performance measured by normalized mutual information, the Omega index, and node-level detection performance measured by F-score. Advantages are: For low overlapping density networks, SLPA, OSLOM, Game, and COPRA offer better performance. The overlapping detection in networks with high overlapping density and high overlapping diversity. Disadvantages are: Expensive computation. In [8] paper, a community detection method based on modularity and an improved genetic algorithm (MIGA) is put forward. MIGA takes the modularity Q as the objective function, which can simplify the algorithm, and uses prior information (the number of community structures), which makes the algorithm more targeted and improves the stability and accuracy of community detection. Advantages are: Good local searching ability of the network. Lower computational complexity. Disadvantages are: Quality is less. Need the local optimal solution of large-scale problems for community detection.

Extended compact genetic algorithm (ECGA) use statistical learning mechanism [9] to build a probability distribution model of all individuals in a population, and then create new population by sampling individuals according to their probability distribution instead of using traditional crossover and mutation operations. ECGA used to find more complex network communities such as overlapping communities, hierarchical communities, and dynamic communities. Advantages are: ECGA is more effective than several other algorithms including GATHB, GA-Net, GN and CNM. ECGA solves complex networks optimization problems. Disadvantages are: ECGA is slower than CNM. The paper [10] has presented a review of previous research in the area of Particle Swarm Optimization, PSO Variants and its application to Clustering High-Dimensional Data. Some of the reviewed algorithms based on PSO can be also applied for clustering complex and linearly non-separable datasets, without any prior knowledge of the number of clusters. Advantages are: PSO has fewer parameters to adjust. PSO used to clustering high-dimensional data. The PSO algorithm is very fast, simple and easy to understand and implement. Disadvantages are: When the search space is high its convergence speed becomes very slow. It also shows poor quality results when it deals with large and complex dataset.

III. EXISTING SYSTEM

Most existing work focuses on disjoint communities discovery from social networks, i.e., each network node, representing a multimedia resource or a user, belongs to one community only. In reality, social network users are naturally characterized by multiple community memberships. For instance, on the popular photo-sharing website Flickr, a user maybe active in subscribing to users from a tourism group in order to view landmark photos, and she may also become a fan of other users from a sport group who publish photos related to football and hockey. Similar observations can be obtained on the video-sharing website YouTube. Therefore, for a social network depicted in disjoint communities resulting from hard-partitioning techniques are less reasonable, compared to overlapping communities. Traditional algorithms based on PSO cannot capture true social relation between

members. Existence of superfluous small communities may lead to unsatisfactory community structure.

Existing system faces some problems given below:

- **Scalability, dynamic networks**
  Currently, the size of social networks is in scale of billions of nodes and connections. As the network is expanding, both the space requirement to store the network and time complexity to process the network would increase exponentially.
- **Heterogeneity**
  Raw social media networks comprise multiple types of edges and vertices. Usually, they are represented as hypergraphs or k-partite graphs. Majority of community detection algorithms are not applicable to hypergraphs or k-partite graphs.
- **Evolution**
  Due to highly dynamic nature of social media data, the evolving nature of network should be taken into account for network analysis applications. So far, the discussion on community detection has progressed under the silent assumption that the network under consideration is static.
- **Evaluation**
  The lack of reliable ground-truth makes the evaluation extremely difficult.
- **Privacy**
  Privacy is a big concern in social media. Facebook, Google often appear in debates about privacy.

IV. SYSTEM OVERVIEW

In this paper, proposes LEPSO, a meta-heuristic approach that combines together line graph theory, ensemble learning and particle swarm optimization techniques for overlapping communities detection. Specifically, transform the overlapping communities detection problem into a disjoint communities detection problem on the corresponding line graph, and represent a community in a social network by a particle that is encoded based on ordered-neighbor-list. Then use ensemble clustering techniques to improve the optimization strategy, so as to effectively optimize modularity of the line graph. After that, convert the disjoint communities generated by DPSO into overlapping communities. Finally, obtain the result by merging overlapping communities according to community overlapping rate. Experiments on real-world and synthetic networks indicate that the proposed method is able to find meaningful community structures from networks with satisfactory convergence rate.

The contribution work is a W-CPM to address the high computational cost of CPM for overlapping community detection in large-scale complex networks. In W-CPM, do not exactly identify the *k*-cliques but only find some weak cliques determined by two nodes in the network. In identifying a weak clique between two nodes, only need to check the common neighbors of the two nodes. Therefore, the identification of weak cliques developed in W-CPM is much more efficient than that of *k*-cliques adopted in a variety of existing methods based on clique percolation theory.

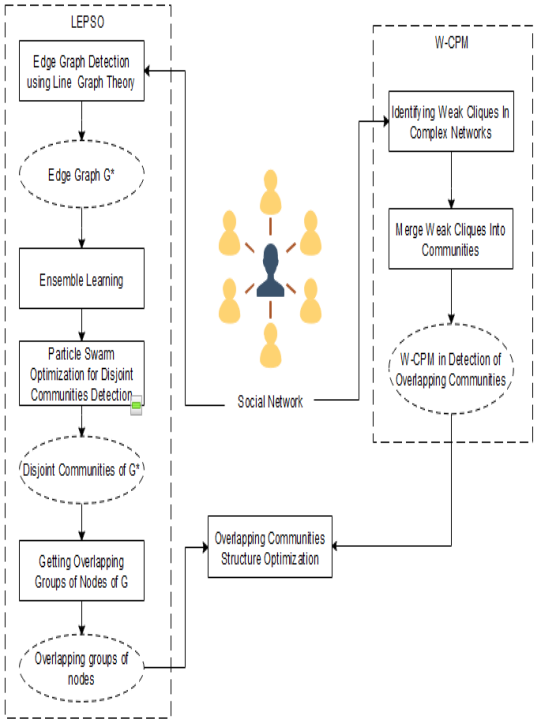

Fig.1: Proposed System Architecture

Advantages are:

1. Utilize distribution information of the neighbors of each vertex, so as to guarantee legality of newborn particles produced during initialization or moving.
2. Elimination of illegal particles completely.
3. LEPSO shows better stability than the two randomized algorithms, and scales up well with network size.
4. Post-processing finer-grained overlappingComm-unities is beneficial to enhance community quality.
5. Adoption of ensemble learning technique can help LEPSO to avoid trapping in local optimal partition of the line graph.
6. Merging strategy is also beneficial to boosting the quality of overlapping communities in the original graph.
7. LEPSO is more robust than EPM and MCMOEA with respect to network modularity
8. LEPSO is more suitable for community detection tasks on networks with complex community structure.
9. W-CPM algorithm based on weak cliques cannot only reduce the computational time, but also improve the quality of found communities.

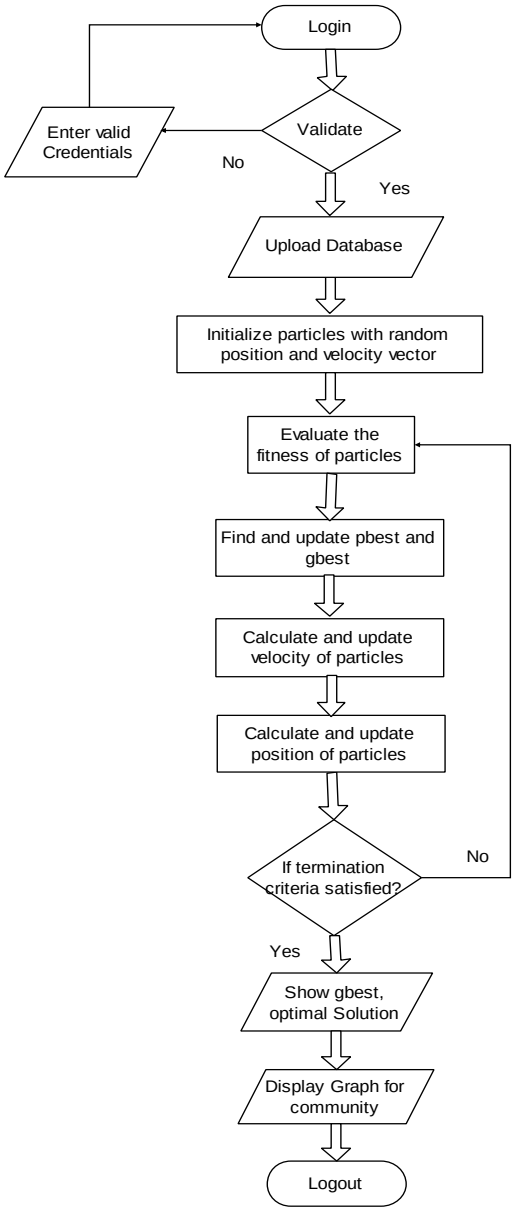

Fig.2: Flowchart of proposed algorithm

V. CONCLUSION

A meta-heuristic algorithm, LEPSO used for overlapping communities'discovery from social networks. Specifically, a particle representation scheme based on ordered neighbor list and a particle update strategy are proposed. Also, a hierarchical agglomerative and bottom-up merging strategy is designed to post-process the generated fine-grained overlapping communities. This conducted extensive experiments and the results show that 1) compared with the non-randomized and randomized algorithms, our LEPSO is superior in terms of validity and robustness, and 2) the proposed hierarchical agglomerative and bottom-up merging strategy can improve quality of the generated overlapping communities. The W-CPM algorithm based on weak cliques cannot only reduce the computational time, but also improve the quality of found communities.

VI. REFERENCES

[1]. 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.

[2]. 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.

[3]. 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.

[4]. J. J. Whang, D. F. Gleich, and I. S. Dhillon, "Overlapping community detection using neighborhood-inflated seed expansion," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 5, pp. 1272–1284, May 2016.

[5]. Z.Wang, D. Zhang, and X. Zhou, "Discovering and profiling overlappingcommunities in location-based social networks," *IEEE Trans. Syst. Man,Cybern., Syst.*, vol. 44, no. 4, pp. 499–509, Apr. 2014.

[6]. M. Shahriari*et al.*, "Predictive analysis of temporal and overlapping community structures in social media," in *Proc. 5th Int. Conf. Companion World Wide Web*, 2016, pp. 855–860.

[7]. J. Xie, S.Kelley, and B.K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," *ACM Comput. Surv.*, vol. 45, no. 4, 2013, Art. no. 43.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. 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.

[12]. Y. Zhang and D. Y. Yeung, "Overlapping community detection via bounded nonnegative matrix tri-factorization," in *Proc. Knowl. Discovery Data*, 2012, pp. 606–614.

[13]. M. Zhou, "Infinite edge partition models for overlapping community detection and link prediction," *Comput. Sci.*, pp. 1135–1143, 2015.

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

[15]. 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.