IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

CLCLP: Overlapping Community Detection in Networks Using a Combination of Link Community and Label Propagation Algorithm

To cite this article: R Jayaprakash 2022 *IOP Conf. Ser.: Mater. Sci. Eng.* **1258** 012048

CLCLP: Overlapping Community Detection in Networks Using a Combination of Link Community and Label Propagation Algorithm

Jayaprakash R[1]

[1]Department of Computer Science,
Nallamuthu Gounder Mahalingam College, Pollachi, Tamil Nadu, India

jpinfosoft@gmail.com

Abstract. Overlapping Community Detection is a tremendous method for accepting a crucial shape and complicated network software. As of overlapping communities descriptions of real-world networks must subsist quantified for community identification, this research suggested a method for discovering overlapping communities termed the Combination of Link Community with Label Propagation (CLCLP) algorithm. CLCLP converts a node separation issue into a link division issue because of its simplicity and competency of the label propagation technique, an improved label propagation algorithm that prioritizes connections over nodes to perceive communities has been developed. The proposed CLCLP then employs post-processing to develop discovered overlapping communities by removing over-overlapping and inaccurate fragile secures partitions. When tested on a large scale of synthetic and real-world networks, the proposed CLCLP approach attains good precision (i.e., accuracy) and NMI (Normalized Mutual Information) in recognizing overlapping communities in networks

Keywords: Propagation, Link, Detection, Cluster, Networks, Label.

1. Introduction

Nodes in many real-world networks could logically fit in to more than on communitiy, and hence the communities may overlap. In public networks, an individual be a part of a community of unit members, a society of associates, or a neighbourhood of coworkers. An internet page can wrap subjects that are related with atypical communities in an information network. Community detection methods that have been used in the past have not been successful in revealing community overlaps. Individuals who are unable to discern community be related in networks with logically overlapping communities are missing out on critical network structural information [1]. As a result, overlapping community detection techniques have a larger set of people paying attention to them. Overlapping communities can be identified using unique methods. Instead of partitioning the nodes, the followig ways is accroding to screening the transformation of a network's edges into communities [2, 3]. To recognise the communities in a network, a range of detection of a community techniques, commonly referred to clustering algorithms, have been suggested. Different communities are given up by

IOP Conf. Series: Materials Science and Engineering 1258 (2022) 012048 doi:10.1088/1757-899X/1258/1/012048

particular community detection algorithms because they use various definitions of a community. On an actual network, Figure 1 shows that how two fundamentally different community recognition algorithms recognise the same communities.

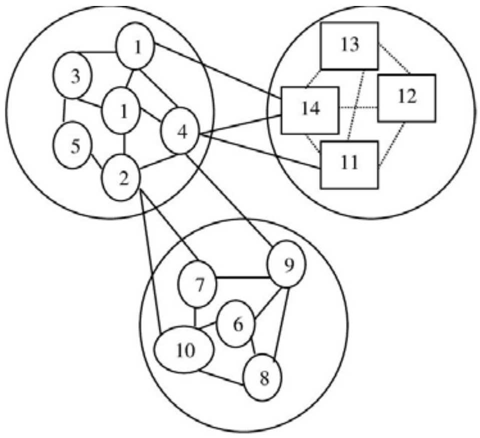

Figure 1. Zachary karate Club Network

Figure 1: The square and circular nodes in the Zachary karate club network represent two clusters of members. [4]

Previous community discovery algorithms assume that the entire network topology is known and accessible. To submit these algorithms as universal algorithms, because they requirement universal information from the whole network so as to categories of the communities therein networks. Because huge networks cannot have such information power, limiting algorithms are gaining popularity [5, 6]. By only looking at a small area of the network, local algorithms can extend a small number of known seed nodes into overlapping communities. Recognizing high-performing a large-scale real-world network communities are often computationally expensive and do not scale well. Parallelism is one way for processing the scalability of community detection. Parallelism can significantly boost up community discovery and furthermore required for dealing with massive amounts of data from the real world.

Local community identification methods are a quick and scalable tool for detecting communities. By just inspecting the neighborhood of 'seed nodes' in the network, local algorithms can finish calculations in parallel by starting with seed nodes and growing into communities. The enlargement of each node in the network into a community is a local method to limited community exposure. However, this method is computationally expensive and results in multiple replica communities. As a result, the goal of the test is to discover the ideal number of seeds to disperse into communities in order to cover the most popular nodes in a network.

The major information of community detection algorithms in difficult networks has been enlarged. Several of them can notice overlapping communities; a few can notice non-overlapping communities. Furthers can observe together overlapping and non-overlapping communities. To overcome the above exposed challenges, here an improved technique to recognize overlapping communities in a large scale network is proposed. The proposed method showed that by significant communities as partition of a network's links, and thus permitting entity nodes to appear in several communities, to calculate the extent to which every pair of communities in a network overlaps. This paper defines two actions of community overlap and applies them to the community formation of networks from special disciplines.

2. Related Works

(Wang, et.al., 2013) proposed the idea of local random walk and a novel similarity metric are introduced. Based on the novel similarity measurement, the dissimilarity index among every node of a network is considered firstly [9]. Then in order to maintain the novel distance node (similarity) to the greatest extent possible, the network formation is linked into low-dimensional space by the multidimensional scaling (MDS). Lastly, fuzzy c-means clustering is utilized to discover fuzzy communities in a network.

(C. Liu, J. Liu, and Z. Jiang, 2014) extended the unique similarity to the noticed similarity based on the social stability assumption [10]. Then, based on the signed similarity and the usual contradiction among negative and positive links, two goal functions are considered to form the problem of detecting communities in SNs as a multi-objective problem. Subsequently, the authors proposed a multi-objective evolutionary algorithm, called MEAs-SN. In MEAs-SN, to conquer faults of indirect and direct illustrations for communities, an indirect and direct pooled representation is considered.

(R. Aldecoa and I. Marin, 2014) reviewed that communities can be exactly distinguished by exploiting Surprise, a global network parameter [11]. Now, the authors presented “SurpriseMe”, a tool that incorporating the results of seven of the finest algorithms obtainable to approximate the highest Surprise value. SurpriseMe also creates similarity matrices that permit which allow you to visualize the relationships among the algorithms’ solutions.

(H.-J. Li and J. J. Daniels, 2015) discussed about combining the particular characteristics of actual society, the authors presented a framework to examine the importance of a social community [12]. The dynamics of social connections are formed by recognizing social leaders and matching hierarchical structures. Instead of a direct assessment with the standard outcome of a random model, they computed the similarity of a specified node with the leader by the amount of frequent neighbors.

(X. Yu, J. Yang, and Z.-Q. Xie, 2015) provided clustering algorithm for 'community discovery' based on the “Link-Field-Topic (LFT) model” recommended to handle the problem of a fixed number of communities [13]. Because the clustering technique is self-contained in terms of context sampling, the number of communities does not need to be predefined. They developed “Semantic Link Weight (SLW)” based on the LFT study to assess “semantic weight” of links for every sampling field in order to tackle the problem of “overlapping community detection.”

(X. Wen, W. Chen, Y. Lin, and T. Gu, 2016) developed a “multi-objective evolutionary algorithm (MOEA) for overlapping community detection” based on maximal cliques (X. Wen, W. Chen, Y. Lin, and T. Gu, 2016) This programmed proposes a novel depiction scheme based on the well-known maximal-clique graph [14]. Because the maximal-group graph is constructed by employing the positions of maximum groups of a unique graph as nodes, and two maximal groups are allowed to assign the novel graph's identical nodes, overlap is a necessary attribute. Because of this attribute, the new representation method allows MOEAs to handle the overlapping community detection problem in a manner similar to divided community detection, hence simplifying optimization problems.

(Chakraborty, et.al., 2016) Presented a comprehensive structure could recognize together overlapping and non-overlapping communities, not including some previous contribution about the network or else its community division [15]. For the community division authors introduced a vertex-based metric, GenPerm, that computes how much a vertex belongs to every of its element communities.

3. Proposed Methodology

The proposed methodology is termed CLCLP and it works with a mix of “overlapping-community detection techniques” of Link Community and Label Propagation. The CLCLP demonstrates that link separation is theoretically common for difficulty of overlapping community recognition. The CLCLP converts the nodes separation problem into a link separation issue, before identifying communities using a label propagation technique that prefers links over nodes, despite the inefficiency of the label propagation algorithm. The CLCLP then does post-processing to advance found overlapping communities by avoiding overlapping and erroneous weak secure partitioning.

IOP Conf. Series: Materials Science and Engineering 1258 (2022) 012048 doi:10.1088/1757-899X/1258/1/012048

4. Network Model Creation

A "Graph G = (V, E), where V is the group of vertices and E is the group of edges" can be used to describe a given input of a network model. Overlapping community detection job is discovering densely connected overlapping sub-graphs in *SG*. The given network G= (*V*, *E*) problem is to expose vertex pair association within the local communities and their equivalent communities. While a vertex has the same multiple relationships to the local communities, the proposed CLCLP algorithm should connect it to only one of its corresponding local communities or communities.

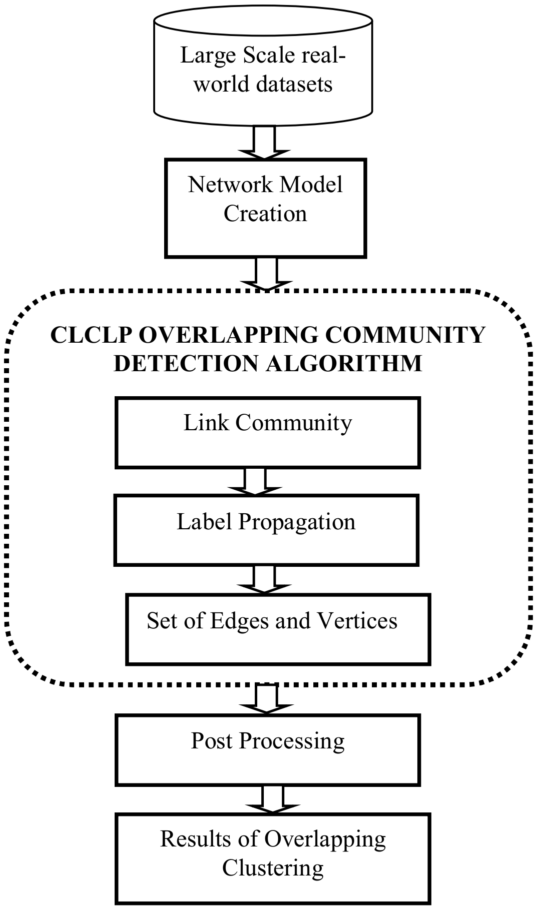

Figure 2: Improved CLCLP process Flow

5. Local community detection

The proposed CLCLP algorithm identifies local communities as follows: The counts of all vertices accessible in the network and arbitrarily chosen an initial vertex (v_i) and calculates its closeness. Then for both neighboring vertex pair in the (v_i), the y-close relative function is calculated. The y-close connected vertices are then clustered into their equivalent local communities. At every edge in the local community is labeled with its equivalent highest vertex close relative.

IOP Conf. Series: Materials Science and Engineering 1258 (2022) 012048 doi:10.1088/1757-899X/1258/1/012048

6. Improved CLCLP Overlapping Community Detection Algorithm

The suggested CLCLP (Combination of Link Community with Label Propagation) innovative overlapping community discovery technique consists of two stages to locate overlapping communities, as shown in Procedure 1. In the paper, an enhanced community discovery technique named CLCLP for large-scale networks to recognize overlapping communities in hard networks. Instead of considering all neighbours' in the present community, the suggested strategy only considers the neighbours of nodes introduced in the most recent expansion. CLCLP has linear time complexity with respect to network size, which is superior to most known overlapping community discovery techniques, according to theoretical study. CLCLP's efficacy and efficiency were demonstrated on large-scale synthetic and real-world networks, demonstrating the proposed algorithm's superiority over existing state-of-the-arts in terms of both effectiveness and efficiency. The proposed CLCLP method achieves good accuracy and NMI on finding communities that overlap in networks when tested on a vast scale of real-world and synthetic networks.

Procedure 1: CLCLP
Input: Given Network formation G = (V, E); *m* is the number of clusters to combine.
Output: Overlapping cluster SC of G as an output
Step 1: *LCLP* ← LinkCommunityLabelPropagation (G)
Step 2: SC ← Post Processing (*LCLP*)
Step 3: return S*C*

The CLCLP algorithm accepts a novel link Community with label propagation algorithm with favourite on links to separation into disjoint communities of nodes event on the connection fit in to the community which the connection members to.

Procedure 2: Link Community Label Propagation
Input: Given Network formation G = (V, E)
Output: ***SC***
Step 1: Set every communication with a single label
Step 2: Iterations = 100
Step 3: While Iterations > 0 **do**
Iterations --;
Node ← $\{ed|ed \in E\}$
For each ed ∈ Node **do**
if *ed* has an identical highest amount of labels of neighbouring edges in $\text{Node}_{\text{nighbour}}$ (*ed*) **then**
linkChoice(ed) = $\{f \in \text{Node}_{\text{nighbour}}$ (*ed*) | max linkSimilarity $LS(ed, f)\}$
if isExist(linkChoice(*ed*) && nodeLabels.includes (linkChoice(ed).label **then**
ed.label = *linkChoice*(*ed*).label;
else
ed.label = randomly choose form Node labels;
end if
else
label(ed) = label with maximum occurrence with labels of neighbouring edges;
end if
end for
if each edge has a label that is shared by a high number of nearby edges **then**

quit;
end if
end while
Step 4: Keep edges with similar label into the similar cluster C_l;
Step 5: Into *C*, keep vertices that appear near the margins of the similar community in C_l;
Step 6: return C

The *LS* among two links ed_{ak} and ed_{bk} is follwed as,

$$LS(ed_{ak}, ed_{bk}) = \frac{|N(a) \cap N(b)|}{|N(a) \cup N(b)|} \quad (1)$$

Where *N(u)* is the formation neighbourhood of vertex u.

Procedure 1 and Procedure 2 are interred linked. After a link portions, the node be apt noticed as overlapping if its occurrence links fit in to unusual neighbourhoods. To conquer this issue noticed the overlapping community compositions improved, CLCP adopted a post-processing process encloses two stages: managing over-overlapping and combining related clusters. The process of post-processing was demonstrated in procedure 3.

Procedure 3: Post Processing

Input: *SC;* combine factor *m*
Output: ***OC*** (Overlapping Cluster)
Step 1: for every overlapping node *n* **do**
Neighbour (n) $= N(v) \setminus \{n\}$
for every Cluster $\in$ SC $\wedge$ n $\in$ Cluster **do**
IntraCluster (n) = $\{n_x | n_x \in$ Neighbour(n) $\wedge\, n_x \in$ Cluster$\}$
end for
end for
Step 2: Combine Clusters
for every Cluster $\in C$ **do**
if $C_l \in C \wedge C_1$.dimension() $\leq$ C.dimension() **then**
u = OC $\cup$ C_1
OC-Cluster;
OC $-C_l$;
OC $\cup$ u;
end if
end for
Step 3: return OC

7. Experimental Results

The experimental performance results of the proposed CLCLP algorithm perform on large-scale hard networks by evaluating with “overlapping-community detection algorithms”. In the following experimental setting carried out through MATLAB, this comprises network parameters and evaluation metrics, and the test datasets.

IOP Conf. Series: Materials Science and Engineering 1258 (2022) 012048 doi:10.1088/1757-899X/1258/1/012048

Table 1. Simulation Parameters

Parameters	Symbol
Number of Links	L
Quantity of Nodes	*N*
Quantity of Communities	J
Quantity of links in community	*Lc*
Node Energy	E_{node}
Number of Links between the nodes	N_a *and* N_b

Figure 3 shows the Normalized Mutual Information (NMI) [16] rates of four “overlapping community detection algorithms” on the Lancichinetti–Fortunato–Radicchi (LFR)-N sets. From Figure 3, it can be seen that CLCLP performs significantly superior than OCLN (Overlapping-Community Detection Algorithm using Local-Neighbourhood Information), NISE (Neighbourhood-Inflated Seed Expansion), and LFM (Local Fitness Method) on the LFR-N set. The proposed CLCLP is superior than OCLN expressions of NMI.

Table 2. NMI values of the four methods compared

Nodes	**0.9M**	**1M**	**1.1M**	**1.2M**
NISE	0.5	0.62	0.68	0.72
LFM	0.75	0.77	0.78	0.79
OCLN	0.79	0.8	0.82	0.83
CLCLP	**0.79**	**0.83**	**0.86**	**0.91**

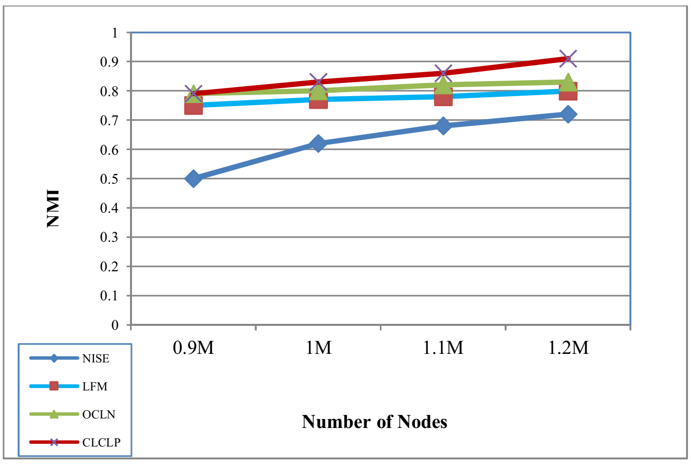

Figure 3. Chart of NMI Ratio

Figure 4 demonstrates the accuracy [17] measures of special "overlapping-community detection algorithms" on the LFR-N group. The LFR-N set consists of 12 LFR networks. Their node size n varies from 100000 (0.1 M) to 1200000 (1.2 M) with an interval of 100000.It can be practical of proposed CLCLP takes O(m) time forever attains a finest performance an expressions of accuracy using MATLAB.

Table 3. Comparison of Accuracy measures of the four algorithms.

Nodes	0.9M	1M	1.1M	1.2M
NISE	0.61	0.65	0.68	0.71
LFM	0.78	0.79	0.79	0.79
OCLN	0.87	0.88	0.89	0.89
CLCLP	**0.89**	**0.90**	**0.912**	**0.93**

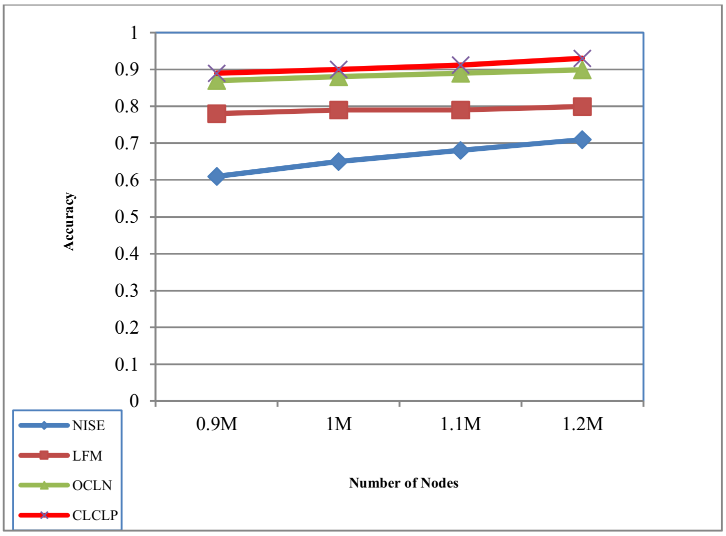

Figure 4. Chart of Accuracy Ratio

8. Conclusion

In this paper, an enhanced community discovery technique named CCLP was given to recognise overlapping communities in hard networks. Overlapping node partitioning is a problem was changed into overlapping link partition by using Link community based label propagation. The CLCLCP has innovatively utilised a new combination of CLCLP algorithm with favourite on connections and lastly locate the overlapping communities during post-processing technique to circumvent over-overlapping. The proposed CLCLP method achieves good accuracy and NMI on finding communities that overlap in networks when tested on a vast scale of real-world and synthetic networks.

References

[1] Ullas Gargi and Wenjun Lu. 2011 Large-Scale Community Detection on YouTube for Topic Discovery and Exploration *Proceedings of the Fifth International Conference on Weblogs and Social Media The AAAI Press*

[2] T. Evans and R. Lambiotte 2009 Line graphs, link partitions, and overlapping communities.

Physical Review E 80 1–8
[3] Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. 2010 Link communities reveal multiscale complexity in networks. *Nature* 466 7307 761–764
[4] W.W. Zachary 1977 An information flow model for conflict and fission in small groups *Journal of Anthropological Research* 33 452–473
[5] Michele Coscia, Giulio Rossetti, Fosca Giannotti, and Dino Pedreschi 2012 DEMON: a local-first discovery method for overlapping communities. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12* ACM Press, 615
[6] Aaron Clauset 2005 Finding local community structure in networks. *Physical Review E*, 72 2 026132
[7] Sucheta Soundarajan and John E Hopcroft 2014 Use of Local Group Information to Identify Communities in Networks *ACM Transactions on Knowledge Discovery from Data*.
[8] Joyce Jiyoung Whang, David F Gleich, and Inderjit S Dhillon 2013 overlapping community detection using seed set expansion. *Proceedings of the 22nd ACM international Conference on information & knowledge management - CIKM '13* ACM Press 2099–2108
[9] W. Wang, D. Liu, X. Liu, and L. Pan 2013 Fuzzy overlapping community detection based on local random walk and multidimensional scaling. *Phys. A, Stat. Mech. Appl*., 392 24 6578–6586
[10] C. Liu, J. Liu, and Z. Jiang 2014 A multi objective evolutionary algorithm based on similarity for community detection from signed social networks. *IEEE Trans. Cybern*., 44 12 2274
[11] R. Aldecoa and I. Marin 2014 SurpriseMe: An integrated tool for network community structure characterization using surprise maximization. *Bioinformatics*, 30 7 1041–1042
[12] H.-J. Li and J. J. Daniels 2015 Social significance of community structure: Statistical view. *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip Top*., 91 1 012801
[13] X. Yu, J. Yang, and Z.-Q. Xie 2015 A semantic overlapping community detection algorithm based on field sampling. *Expert Syst. Appl.,* 42 1 366–375
[14] X. Wen, W. Chen, Y. Lin, and T. Gu 2016 A maximal clique based multiobjective evolutionary algorithm for overlapping community detection. *IEEE Trans. Evol. Comput*., 1–14
[15] T. Chakraborty, S. Kumar, N. Ganguly, A. Mukherjee, and S. Bhowmick 2016 GenPerm: A unified method for detecting non-overlapping and overlapping communities. *IEEE Trans. Knowl. Data Eng*., 28 8 2101–2114
[16] A Lancichinetti, S. Fortunato, and J. Kertész 2009 Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys*., 11 3 19–44