International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

CSG Cluster: A Collaborative Similarity Based Graph Clustering for Community Detection in Complex Networks

Smita S Agrawal, Atul Patel

Abstract: Many real-world social networks exist in the form of the complex network, which includes very large scale structured or unstructured data. The large scale networks like brain graph, protein structure, food web, transportation system, WorldWide Web, online social network are sparsely connected globally anddensely connected locally. For detectingdensely connected clusters from complex networks, graph clustering methods are useful. Graph clustering performs through partition a graph based on edge cut, vertex cut, edge betweeness, vertex similarities, topological structure of graph. Most of the graph clustering methods predominantly emphasis on topological structure of graph and not bearing in mind the vertex properties/attributes or similarity based onindirectly connected vertices. In this paper, we propose a CSGCluster, a novel collaborative similarity based graph clustering methodfor community detection ina complex network. In this, we introduce concepts, Approachable Unitto find similaritiesfor directly connected vertices andintroduced shortest path strategy for indirectlyconnected vertices and based on that a graph clustering method,CSG-Cluster is presented. For this, a new collaborative similarity approach is adopted to computevertex similarities. In the CSG-Cluster method, weform a group of vertices based on distance measures based on calculated similarity with the help of K-Medoids framework. Performs experiment on two real datasets with other relevant methods in whichresults shows the effectiveness of CSG-Cluster.This idea is suitable for graph database to apply collaborative similarity during query processing.

Index Terms: attribute similarity, community detection, complex network, graph clustering, vertex similarity

I. INTRODUCTION

Nowadays, exponentially data generating and travels across the internet is not only significant in volume but also complicated. The amount of data travelling across the internet today is not only in the massive amount of size but is complex as well. Many real-world complex networks exist in the form of Big Data. Big Data of complex networks includes a large scale of data in the form of structured, semi-structured and unstructured[1–4].A various complex network is available in the form of human brain graph[5], protein structure[6,7], food web [8], transportation system [9], World Wide Web [10–12], etc. Most of the commercial and bioscience applications have complicated structure which required additional efforts to analyse it. Data visualisation of the complicated structure can be formed in trees, networks, graphs which provides the connectivity in a complex network.

In the complex network of IoT based application, data generation rate is exponential which generates heterogeneous / semi-structured data, and for such semi-structured data, different types of NoSQL databases are supports [13][14]

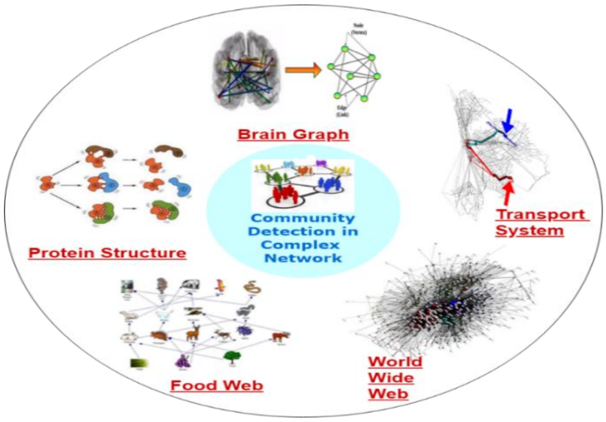

Figure 1: Applications of Community Detection in Complex Network

A graph data structure becomes popular nowadays for Big Data Analytics. Graph structure provides interaction between nodes, weights and their properties.In figure-1, various applications for community detection represents which form graph — for these applications, modelling and generating a graph based on their structure, attributes, weight and direction (whether directed or undirected) becomes an important task. Communities of proteins networks identify functionally related to protein-protein interaction and also supports earlier stages of diseases treatment related to drug discovery[6,7,15]. In a graph of a complex network, nodes are connected with the neighbour node and other nodes and form various kind of network. Community detection from complex network various measures like vertex similarity, edge betweenness, distance measures techniques used. Graph clustering based on densely connected networks used vertex similarity known as a structural similarity[16]. Most of the clustering performs similarity measures, either structure based or attribute based. There are few methods which considered both structures as well as attribute similarities.

Manuscript published on 30 June 2019.
* Correspondence Author (s)
Prof Smita S Agrawal, CSE Department, Institute of Technology, Nirma University, Ahmedabad , India.
Dr Atul Patel, Dean of CMPICA, CHARUSAT University, Changa, India

For example, publications of authors based on citation, co-author and research area different community may also form sometimes it forms overlapping of community in the different cluster for social network and publication network.

In proposed method,the relevantgraph clustering method for identifying structural and attribute similarity for community detection in a complex network is discussed.

Our method partitions a large scale graph basis on connectivity and attributes and partitioned into k number of clusters so that each cluster have densely connected subgraph with homogeneous attribute values as well.

In this, we propose a new graph clustering method, CSG-Cluster, based on Approachable Unit between two vertices for graph clustering. CSG clustering can achieve the graph clustering by identifying collaborative similarity (i.e. both structure and attribute) similarities. (1) A novel pair-wise topological similarity approach based on the approachable unit for direct vertex connectivity. (2) A novel path selection strategy based on the product of the maximum approachable unit is employed for indirect vertex connectivity.
The respite of this paper is organised as follow: Section 2 with the related work towards community detection methodology. The proposed work related to CSG-Cluster collaborative similarity for community detection in complex networks has been discussed in details in Section-3. In this, we introduce the concept of approachable Unit for directly connected vertices and shortest path strategy for indirectly connected vertices. Section -4 discussed experiments on two real datasets with other relevant methods in which results shows the effectiveness of CSG-Cluster and then, concludes the paper.

II. RELATED WORK

Many community detection method using graph clustering approach based on structural, attribute and collaborative similarity in a complex network. In a complex network, vertices are sparsely connected globally and densely connected locally[17]. In such a complex network, vertices of within a cluster have a higher density of edge and another way those vertices connected with other cluster have a lower density of edge.[18,19]. The vertex belongs to the same cluster/community is possible to have common parameters like connectivity, attributes, interests, follows and many more. For example, the Facebook network forms communities based on mutual friends, post follows by the user and based on that recommend new connection can part of the community. [20].W-Cluster[26], SA-Cluster[27] have been proposed graph clustering performs based on structural and attribute similarities. Unified neighbourhood random walk method is used to measure vertex closeness among the vertices. For attributed and structural based similarity weighted function is used where the weight for structure and attribute represents with α and β respectively.IGC-CSM[22], AR-Cluster[21] have been proposed for measures collaborative similarity by combines both structural and attribute similarities with consideration of vertex connectivity with another vertex whether it is connected, disconnected or indirectly connected. After finding of similarities between vertices, K-medoids framework is used for clustering. However AR-Cluster[21] have been adopted attracting factor and degree for directly connected nodes and recommending the degree for indirectly connected node which strengthen the structural similarities among the vertices.

Graph clustering measures the vertex closeness through vertexconnectivity with other vertices,structural similarity, random walk distance, etc. for community detection. Vertex closeness based on connectivity and structural similarity: In a graph clustering based on the number of the possible path between two vertices(i.e. connectivity)and the number of common vertices between two vertices(i.e. structural similarity), many algorithms have been proposed based on this approach for community detection.

The contribution of this paper summarised below:

1. Approachable Unit based on vertex degree and weight of connection for a directly connected node is presented
2. A new structural similarity for indirectly connected nodes measured throughshortest path calculation on Approachable Unit is used
3. A graph clustering method based on collaborative similarity is proposed.

Symbols	Description
V_a	Vertex **a** from a set of Vertices in graph G
d(Va)	No. of degrees of vertex a
K	Number of clusters
Va ↔Vb	Direct connectivity between two vertices
Va ΘVb	An indirect connection between vertices, i.e. through the intermediate node
Va⊗Vb	Disconnected two vertices
Web	the weight associated with the connection between vertices a and b
Wattri	weight of i^{th}attribute

Table 1: Frequently used Symbols

III. PROPOSED METHOD : CSG CLUSTER

In this section, we proposed Approachable Unit as AU and Approachable Path as AP which used to perform graph clustering on complex network. An undirected weighted attribute based graph exemplify as G(V,E,Λ,W) where V is number of vertices graph G, E is edges/connectivity of vertices with other vertices in which weight of edge of $V_{a \text{ and}}$ V_b defined as W_{eb} , Λ is attribute(s) of each vertex in which attributes of V_a represent as $attr_i(V_a)=\{attr_1(V_a), attr_2(V_a), attr_3(V_a), \ldots, attr_n(V_a)\}$. In a graph, if two vertices have direct link between them we represent as directly connected node and if any path possible for connectivity between two vertices we represent as indirectly connected otherwise those two vertices are disconnected.

A graph has topological structure, which represents the connection of vertices with other vertices with its due weight as well as vertex properties. Both topological structure and vertex properties have their importance; however, utmostpresent methodologies deliberate any one of them. In this context, we propose a CSG-cluster which is collaborative similarity based on graph clustering.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The proposed method is grounded on the Approachable Unit and Approachable Path. Some symbols which used frequently are described in Table 1.

A. Approachable Unit

AR-Cluster[21] analyzes Attracting factor and Attracting degree of graph clustering where for undirected graph they separately calculated Attracting Factor for directly connected node V_a and V_b, and consider its average in Attracting Factor , means calculated for $f(V_a,V_b)$ and $f(V_b,V_b)$ and its average is considered as Attracting Degree as $A(V_a,V_b)$. We proposed Approachable Unit of undirected graph for directly connected node V_a and V_b, consider the cumulative of all degrees of V_a and V_b and consider only once the weight. The Approachable Unit $AU(V_a,V_b)$ for vertex V_a and V_b is defined in equation (1).

$$\boldsymbol{AU(V_a,V_b)} \Downarrow = \ln\left(1 + \frac{d(V_a) + d(V_b) - 1}{\sum_{i=0}^{d(V_a)} W_{ai} + \sum_{j=0}^{d(V_b)} W_{bj} - W_{ab}} * W_{ab}\right), V_a \leftrightarrow V_b \quad (1)$$

B. Approachable Path

In a graph G, for indirectly connected vertex V_a and V_b find the shortest path from source node V_a to destination node V_b, which is based on the product of Approachable Unit of all intermediate nodes. The Approachable Path $AP(V_a,V_b)$ for vertex V_a and V_b is defined in equation (2).

$$\boldsymbol{AP(V_a,V_b)} \Downarrow = Max\left(\prod_{i=sourcenode}^{destnode} ApproachableUnit(V_a,V_b)\right), V_a \ominus V_b \quad (2)$$

C. Structural Similarity

For graph clustering, to identify the structural similarity between two directly or indirectly connected node, we adopt IGC-CSM [22]and Transitive node similarity[23]method mentioned in equation (3) and (4) which is only consider the weight of connected node only with the condition that there is no other possible path between them. To improve the density of cluster, we integrate the structure similarity discover in IGC-CSM with our new concept of Approachable Unit and Approachable Path. For a disconnected node, structural similarity defined as 0. The proposed structural similarity $Similarity(V_a,V_b)_{struct}$ for vertex V_a and V_b is defined in equation (5)

$$\boldsymbol{StructSimilarity(V_a,V_b)_{connected}} \Downarrow = \frac{W_{ab}}{\sum_{i=0}^{d(V_a)} W_{ai} + \sum_{j=0}^{d(V_b)} W_{bj} - W_{ab}}, V_a \leftrightarrow V_b \quad (3)$$

$$\boldsymbol{StructSimilarity(V_a,V_b)_{indirect}} \Downarrow = \prod_{i=sourcenode}^{destnode} StructSimilarity(V_a,V_b)_{connected}, V_a \ominus V_b \& V_i \in V \quad (4)$$

$$\boldsymbol{Similarity(V_a,V_b)_{struct}} \Downarrow = \begin{cases} StructSimilarity(V_a,V_b)_{connected} + AU(V_a,V_b), V_a \leftrightarrow V_b \\ StructSimilarity(V_a,V_b)_{connected} + AP(V_a,V_b), V_a \ominus V_b \text{ and } V_i \in V \\ 0, V_a \otimes V_b \end{cases} \quad (5)$$

D. Attribute Similarity

For graph clustering, to identify the attribute based similarity between two directly or indirectly connected node, we adopt IGC-CSM[22] and AR-Cluster[21] method mentioned in equation (6) and (7) The adopted attribute/context similarity $Similarity(Va,Vb)_{context}$ for vertex V_a and V_b is defined in equation (6) and (7).

$$\boldsymbol{Similarity(V_a,V_b)_{context}} \Downarrow = \begin{cases} \dfrac{\sum_{i=1}^{M} Common(V_a,V_b,attr_i) * Wattr_i}{\sum_{j=0}^{M} Wattr_j}, V_a \leftrightarrow V_b \text{ and } V_a \otimes V_b \\ \prod_{i=sourcenode}^{destnode} Similarity(V_a,V_b)_{context}, V_a \ominus V_b \text{ and } V_i \in V \end{cases} \quad (6)$$

$$\boldsymbol{Common(V_a,V_b,attr_i)} \Downarrow = \begin{cases} 1, & \text{if } V_a \text{ and } V_b \text{ have samethe attributes.} \\ 0, & otherwise \end{cases} \quad (7)$$

E. Collaborative Similarity

To perform the graph clustering based on the topological structure and attribute of the graph, we calculated structure and attribute similarity separately. Now, based on the application need, we can consider the due weight of structure and attribute, which is represented in the form of collaborative similarity. The proposed structural similarity $Similarity(V_a,V_b)_{collaborative}$ for vertex V_a and V_b is defined in equation (8). For collaborative similarity, we use influence parameter α which can defined in range 0 to 1. Value of $\alpha = 1$ represents the graph clustering only performs based on structural similarity. For all experiments w,e defined vathe lue of α as 0.5 which means its give equal weightage to structural and context similarity for graph clustering.

$$\boldsymbol{similarity(V_a,V_b)_{collaborative}} \Downarrow = \alpha * similarity(V_a,V_b)_{struct} + (1-\alpha) * similarity(V_a,V_b)_{context} \quad (8)$$

F. Distance Function

Based on similarity identified in the graph, a grouping of vertices in the respective cluster need to define. For the grouping of vertices in the cluster, we used K-medoid framework. For K-medoid framework, we adopted the distance function of AR-Cluster[21] and IGC-CSM[22]. Distance function Distance(Va,Vb)for vertex Va and Vb is defined in equation (9)

$$\boldsymbol{Distance(V_a,V_b)} \Downarrow$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

$$= \begin{cases} \dfrac{1}{similarity(V_a, V_b)_{collaborative}}, & V_a \leftrightarrow V_b \text{ and } V_a \leftrightarrow V_b \\ \infty, & V_a \otimes V_b \end{cases} \quad (9)$$

IV. EXPERIMENTAL RESULTS

A. Real Datasets

Two real datasets DBLP and Political blogs (PBlog)[24] are used for experiments of CSG-Cluster. The PBlog dataset is a network of US Politics 1490 number of web blogs with 19090 connections/hyperlinks among the US politics web blogs which recorded by Adamic and Glance in 2005. Each web blogs represent as vertices have attributes value with '0' or '1' which specifies with liberal or conservative, respectively. In US political web blogs, all directly connected vertices, the connection edge weight is 1. The US political blogs dataset can be downloaded from http://www-personal.umich.edu/~mejn/netdata/. The DBLP bibliographic dataset has citation information of publications which comprises information like paper title, authors (one or more than one), source title, and publishers. For experimental purpose, we used a subset of DBLP bibliography data which have four selected research areas as Database, Data Mining, Information Retrieval and Artificial Intelligence. Related to PBlog, in DBLP dataset each co-author as vertex have two attributes as 'primary topic' and 'prolific topic'. The basis on several paper publications by author value of prolific topic is defined. For a DBLP graph if author published number of papers >=20 then attribute value is to be considered as 'highly prolific, and if author published number of paper in the range of <20 to >=10 them attribute value is to be considered as 'prolific' otherwise is to be 'low prolific'. The primary topic is assigned based on the selection of one strongly associated topic out of a hundred topics through probability distributions.

B. Evaluation Measures

For comparing the quality of graph clustering method is one of the critical aspects. The quality of cluster can evaluate based on different aspect like verticesin a defined cluster, vertices are densely connected within a cluster, and the same attributes of vertices fall in the respective cluster. Evaluate a graph clustering algorithm, various methods and approaches have been proposed. We evaluate the proposed algorithm CSG-Cluster using the measurement of density and entropy.

For the identified cluster, the structural closeness measures through density. Denstiy[21] is defined in equation(10) as,

$$\boldsymbol{Density}(\{Vc\}) = \sum_{c=1}^{k} \frac{|\{(Vm,Vn)|Vm,Vn \in Vc,(Vm,Vn) \in E\}|}{|E|} \quad (10)$$

Where V_c is the vertex of cluster number from 1 to k. Where k represents a number of clusters. The higher value of density conveying that strongly connected vertices are defined in the same cluster.

To check the relevance of attribute towards the vertices in a defined cluster, entropy[25]value is used for measurement. During the evaluation, the low value of entropy guaranteed that vertices fall in the same cluster have similar attributes. Entropy defined in equation (11) as,

$$\boldsymbol{Entropy}(\{Vc\}c = 1 \text{ to } k)$$

$$\Downarrow$$

$$= \sum_{c=1}^{m} \frac{Wc}{\sum_{j=1}^{m} Wj} \sum_{l=1}^{k} \frac{|V_l|}{|V|} \boldsymbol{entropy}(a_c V_l) \quad (11)$$

where $entropy(a_c V_l) = -\sum_{n=1}^{n_c} P_{cln} \log_2 P_{cln}$ and P_{cin} is the percentage of vertices in respective cluster which havevalue a_{cn} on attribute a_c. It measures the weighted entropy from all attributes over k clusters.

C. Contrast Methods

We evaluated the proposed method with other relevant existing methods for community detection based on graph clustering.

CSG-Cluster Algorithm

Input:

a. **An undirected weighted attribute based graph exemplify as G(V,E,Λ,W) where V is number of vertices graph G, E is edges/connectivity of vertices with other vertices in which weight of edge of V_a and V_b defined as W_{eb}, Λ is attribute(s) of each vertex in which attributes of V_a represent as $attr_i(V_a)=\{ attr_1(V_a), attr_2(V_a), attr_3(V_a), \ldots, attr_n(V_a)\}$.**

b. **Influence parameter / weighted factor α**

c. **The number of cluster k**

Output:

a. **Forms K cluster based on value of α and number of cluster mentioned (i.e. $C_1, C_2, C_3, \ldots C_k$)**

1. Initialization

Distance$[V_a][V_b]$=0, CentroidOfCluster[]=0, $d(V_a)$

2. Similarity Calculation

For each vertex pair V_a and V_b in V where $V_a \leftrightarrow V_b$

Find $\boldsymbol{AU}(V_a, V_b)$ **//Approachable Unit**

End For

For each vertex pair V_a and V_b in V where $V_a \leftrightarrow V_b$

Find $\boldsymbol{AP}(V_a, V_b)$ **//Approachable Path**

End For

For each vertex pair V_a and V_b in V where $V_a \neq V_b$

$$\boldsymbol{similarity}(V_a, V_b)_{collaborative} = \alpha * \boldsymbol{similarity}(V_a, V_b)_{struct} + (1-\alpha) * \boldsymbol{similarity}(V_a, V_b)_{context}$$

$$\boldsymbol{Distance}(V_a, V_b) = \frac{1}{similarity(V_a, V_b)_{collaborative}}$$

End For

3. K-Medoids Clustering

Based on distance function identified the CentroidOfCluster[ ] for each cluster and forms the cluster using K-Medoid Framework

Algorithm:1

W-Cluster[26]:This algorithm for graph clustering performs based on structural and attribute similarities. Unified neighbourhood random walk method is used to measure vertex closeness among the vertices. For attributed and structural based similarity weighted function is used where the weight for structure and attribute represents with α and β respectively.

SA-Cluster[27]**:** This algorithm finds structural and attributes similarities based on random walk distance measures. Also, for understanding the different attributes contribution degree, it utilizes an adaptive weight strategy,which is time- consuming.

IGC-CSM[22]**:** The algorithm measures collaborative similarity by combines both structural and attribute similarities with consideration ofvertex connectivity with another vertex whether it is connected, disconnected or indirectly connected. After finding of similarities between vertices, K-medoids framework is used for clustering.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Also, uses a weighted shortest path strategy is adopted instead of all paths to reduce computational cost and a search space.

AR-Cluster[21]: This algorithm also combines the structural as well as attribute similarities for graph clustering. It is adopted attracting factor and attracting degree for directly connected node and recommending the degree for indirectly connected node which is strengthen the structural similarities among the vertices.**CSG-Cluster**(Proposed work) :This is our proposed algorithm. It also performs graph clustering that combines the structural and attribute similarities based on Approachable Unit and Shortest path strategy as Approachable Path. A novel clustering algorithm CSG Cluster with K-Medoids framework is presented for detecting communities using collaborative similarity measure is proposed. For understanding the process and steps of proposed CSG-Cluster methodology presented step by step through Algorithm-I in this paper.

D.Results

All of the experiments have a default value of influence parameter α set with 0.5 unless explicitly mentioned. Our proposed approach has the time complexity quadratic in nature, which is considerable for a medium scale graph with respect to the size of the graph. The alternate proposed method provides a consistent result without using a complex operation like matrix multiplication. We have measures the quality of experimental results by some state of the art measures as density and entropy. The range of a number of cluster in experiments is varying based on graph size. The results provide that strongly connected vertices defined in the same cluster. Density value is used to identify where dense connectivity among the original graph and in cluster exist or not.

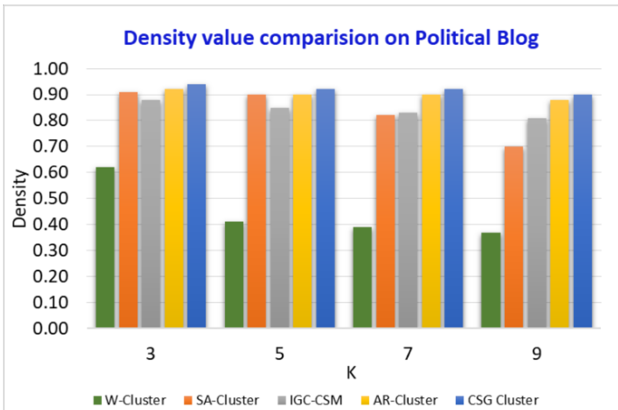

Figure 2 : Density - PBlog

In figure 2 and 3, the density values achieved by proposed and contrast methods are plotted for the different number of clusters. In Figure-2, the density value increase in comparison to other methods. Also, we have observed that the compared to Political Blogs dataset performance is gain in large scale dataset of DBLP. Moreover, our approach achieves better density value with compare to other contrast methods.

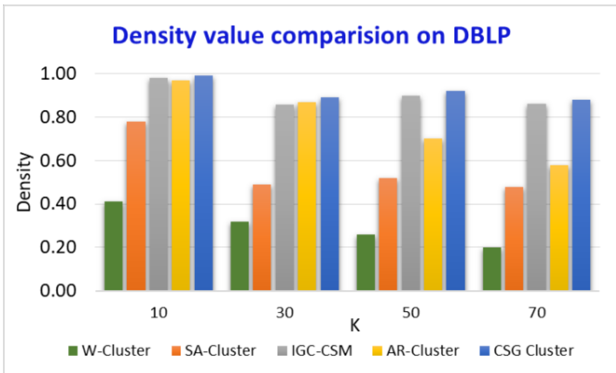

Figure 3: Density - DBLP

We give equal importance to the vertex properties as well as vertex connectivity. To check whether the ratio of vertex properties based on their similarity and dissimilarity, entropy is used. As mentioned in figure-4 and 5, for respective cluster-wise entropy value plotted for different algorithms.

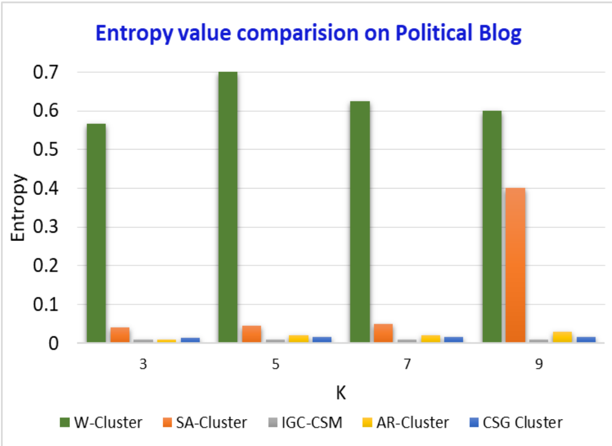

Figure 4: Entropy - PBlog

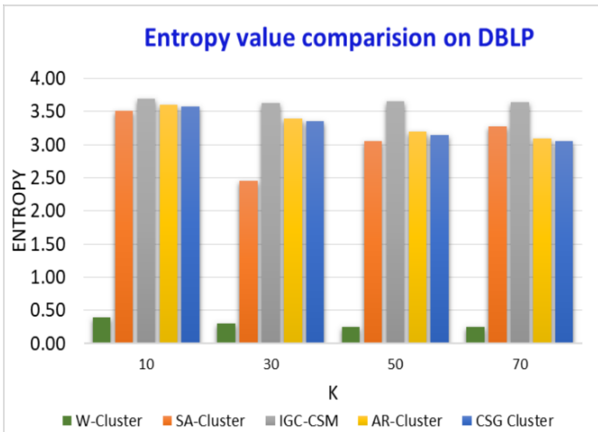

Figure 5: Entropy - DBLP

It shows that entropy value improved through the proposed algorithm, which represents that vertex with the same attributes defined in the same cluster.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

V.CONCLUSION

In this paper, we have presented Approachable Unit to find similarities for directly connected vertices and introduced Approachable Path to shortest path strategy for indirectly connected vertices, to partition the vertices by their topological structure and properties. For grouping vertices in, we have used the popular K-medoid framework. To measures the quality of a defined cluster, we have evaluated the clusters information through density and entropy measures. The proposed method achieved a satisfactory improvement in performance compared to other relevant methods. We perform experiments of the proposed algorithm on two real graph datasets in which achieves a good result with respect to cluster quality. The results provide that strongly connected vertices defined in the same cluster.

REFERENCES

1. M. Dehmer, F. Emmert-Streib, S. Pickl, A. Holzinger, Big Data of Complex Networks, Chapman & Hall/CRC, 2016.
2. Y. Wang, L.A. Kung, T.A. Byrd, Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, Technol. Forecast. Soc. Change. 126 (2018) 3–13. doi:10.1016/j.techfore.2015.12.019.
3. J.P. Verma, S. Agrawal, B. Patel, A. Patel, Big Data Analytics: Challenges and Applications for Text, Audio, Video, and Social media Data, Int. J. Soft Comput. Artif. Intell. Appl. 5 (2016) 41–51. doi:10.5121/ijscai.2016.5105.
4. S. Agrawal, A. Patel, a Study on Graph Storage Database of Nosql, Int. J. Soft Comput. Artif. Intell. Appl. 5 (2016) 33–39. doi:10.5121/ijscai.2016.5104.
5. J.O. Garcia, A. Ashourvan, S. Muldoon, J.M. Vettel, D.S. Bassett, Applications of Community Detection Techniques to Brain Graphs: Algorithmic Considerations and Implications for Neural Function, Proc. IEEE. 106 (2018) 846–867. doi:10.1109/JPROC.2017.2786710.
6. J. Han, M. Kamber, J. Pei, Chapter 9: Graph Mining, Social Network Analysis, and Multirelational Data Mining, Data Min. Concepts Tech. (2006) 535–589. http://books.google.com/books?hl=en&lr=&id=AfL0t-YzOrEC&oi=fnd&pg=PP2&dq=Data+Mining:+Concepts+and+Techniques&ots=Uv-WrRflz9&sig=E_20H417umQmqnf_jM9m9DyNRf0.
7. C. Giatsidis, F.D. Malliaros, M. Vazirgiannis, Advanced graph mining for community evaluation in social networks and the web, Proc. Sixth ACM Int. Conf. Web Search Data Min. - WSDM '13. (2013) 771. doi:10.1145/2433396.2433495.
8. E. Society, E. Monographs, The Seasonal Dynamics of The Chesapeake Bay Ecosystem Author (s): Daniel Baird and Robert E . Ulanowicz Published by : Wiley on behalf of the Ecological Society of America Stable URL : https://www.jstor.org/stable/1943071 Wiley , Ecological Society of America are collaborating with JSTOR to digitize , preserve and, 59 (2019) 329–364.
9. J. Celko, Graph Databases, 2014. doi:10.1016/B978-0-12-407192-6.00003-0.
10. S. Brin, L. Page, The anatomy of a large scale hypertextual Web search engine, Comput. Networks ISDN Syst. 30 (1998) 107–17. doi:10.1.1.109.4049.
11. T. Suel, J.Y.J. Yuan, Compressing the graph structure of the Web, Proc. DCC 2001. Data Compression Conf. (2001) 1–10. doi:10.1109/DCC.2001.917152.
12. K. Desai, V. Devulapalli, S. Agrawal, P. Kathiria, A. Patel, Web Crawler: Review of Different Types of Web Crawler, Its Issues, Applications and Research Opportunities, Int. J. Adv. Res. Comput. Sci. 8 (2017) 1199–1202. http://ijarcs.info/index.php/Ijarcs/article/view/3190.
13. S. Agrawal, A. Patel, Essentials of NoSQL Database in building applications based on Internet of Things (IoT), 10 (2017) 1829–1837.
14. S. Yadav, J. Verma, S. Agrawal, SUTRON: IoT-based Industrial/Home Security and Automation System to Compete the Smarter World, Int. J. Appl. Res. Inf. Technol. Comput. 8 (2017) 193–198.
15. H. Mahmoud, F. Masulli, S. Rovetta, G. Russo, Community detection in protein-protein interaction networks using spectral and graph approaches, in: Int. Meet. Comput. Intell. Methods Bioinforma. Biostat., 2013: pp. 62–75.
16. F. Dabaghi-zarandi, M.K. Rafsanjani, Community Detection in Social Networks, (n.d.).
17. N. Du, B. Wang, B. Wu, Community Detection in Large-Scale Social Networks ∗, (n.d.) 16–25.
18. J. Scott, Social network analysis, Sage, 2017.
19. S. Wasserman, K. Faust, Social network analysis: Methods and applications, Cambridge university press, 1994.
20. E. Baatarjav, S. Phithakkitnukoon, R. Dantu, Group Recommendation System for Facebook, (2008) 211–219.
21. H.F. Zhou, J. Li, J.H. Li, F.C. Zhang, Y.A. Cui, A graph clustering method for community detection in complex networks, Phys. A Stat. Mech. Its Appl. 469 (2017) 551–562. doi:10.1016/j.physa.2016.11.015.
22. W. Nawaz, K. Khan, S. Lee, Intra graph clustering using collaborative similarity measure Intra Graph Clustering using Collaborative Similarity, (2015). doi:10.1007/s10619-014-7170-x.
23. E. Tiakas, Graph Node Clustering via Transitive Node Similarity, 2010 14th Panhellenic Conf. Informatics. (2010) 72–77. doi:10.1109/PCI.2010.42.
24. L.A. Adamic, N. Glance, The political blogosphere and the 2004 US election: divided they blog, in: Proc. 3rd Int. Work. Link Discov., 2005: pp. 36–43.
25. Y. Zhou, H. Cheng, J.X. Yu, Clustering large attributed graphs: An efficient incremental approach, Proc. - IEEE Int. Conf. Data Mining, ICDM. (2010) 689–698. doi:10.1109/ICDM.2010.41.
26. H. Cheng, Y. Zhou, J.X. Yu, Clustering Large Attributed Graphs, ACM Trans. Knowl. Discov. Data. 5 (2011) 1–33. doi:10.1145/1921632.1921638.
27. Y. Zhou, H. Cheng, J.X. Yu, Graph Clustering Based on Structural / Attribute Similarities, Vldb. 2 (2009) 718–729. doi:10.14778/1687627.1687709.

AUTHORS PROFILE

Prof Smita Agrawal is working as Assistant Professor at Department of Computer Science and Engineering, Nirma University since January 2009. She has a professional experience of over 14 years that footholds in academia. She received her Master of Computer Applications degree from Gujarat Vidhyapith in 2004. She is currently pursuing her doctoral studies in the field of Big Data Analytics from CHARUSAT University, Gujarat. Prof. Agrawal has published several research papers in national and international conferences and journals with indexing in Scopus, ICI. Prof Agrawal has conducted ISTE approved STTP in the field of Web Services using PHP. She works in the area of Big Data Analytics, Parallel Processing, Web Development and IoT.

Atul Patel is Professor and Dean, Smt Chandaben Mohanbhai Patel Institute of Computer Applications – Charotar University of Science and Technology (CHARUSAT) Changa, India. Hereceived Bachelor degree in Science B.Sc (Electronics), M.C.A. Degree from Gujarat University, India. M.Phil. (Computer Science) Degree from Madurai Kamraj University, India. He has received his PhD degree from S. P. University. His main research areas are wireless communication and Network Security.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)