International Journal of Engineering & Technology, 7 (4.36) (2018) 802-805

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Effective Utilization of Shared Nearest Node for Message Diffusion in Social Network Using Dbscan

P. Apoorva[1*], S. Akshay[2] , R. Priyanka[3], N. Nayana[4]

[1]Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, India.
[2]Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, India.
E-mail:akshayshantharam26@gmail.com
[3]Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, India.
E-mail:priyajcob33@gmail.com
[4]Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, India.
E-mail:nayananagendra95@gmail.com
**Corresponding author E-mail: apoorvaap7@gmail.com*

Abstract

The social networking service has been enormously used among various people to share information or to build social relationship between acquaintances and other people as well. This term is used to describe a social structure where many users can bring forth their perspective on certain global information or imbalances that has been occurred over centuries. The goal of Information diffusion is to spread messages over a network with a lesser time complexity and efficient accessibility. Here, to ease the process of message diffusion in social networking, we are finding overlapping nodes between commonly Shared Nearest nodes and aid in spreading the information more appropriately by reducing the complexity in the existing system and promoting an efficient level of performance. Density-based clustering is a relevant method we have used to trace shared nearest neighbor node. Also, we provide security for the data that is being diffused by implementing the RSA security algorithm and providing the security key along with the information and hence the group of people who are eligible to access the data with the security key can only access the data. Hence the information is being diffused evenly to each part in the cluster with less time complexity and efficiency.

Keywords: *Shared Nearest Neighbor Clustering (SNN), information diffusion, overlapping nodes, density based clustering, Social Network Analysis (SNA), compels networks, community structure*.

1. Introduction

The scope of social network is wide and evolving around different set of people. It creates an awareness about the various loop holes in society, provides a platform to have or share perspectives about several issues, share photos, videos and shed light on different interests within their network.

Information diffusion or message diffusion is a vital part of a social network. Diffusion helps people to share their views on the current trends and create an awareness irrespective of their location.It may be surprised to see how fast the information can spread in a minimum span of time through a social network. Message diffusion would pass through several people at a very short span of time and spread among different networks. Different entities connected to their network, can analyze the similar interests, views and make the best use of social networks. It makes connecting with like-minded processionals easy and, through quality interactions, can significantly aid in expanding their contact lists. The process by which an individual can locate likeminded people having the same sort of interests and the process of passing information across various platforms have been under tremendous changes in today's social networking world. As such, to perform accurate diffusion of information, we are likely to concentrate more on the shared neighbor nodes. That is, we must find the overlapping nodes between two entities rather than working with non-overlapping nodes.

Clustering is the core part of overlapping nodes. The goal of this technique is to find similarities between several objects of data that are expressed with the available similarity method. Clustering becomes tedious, as the dimensionality of the nodes or density increases. Clustering technique has reached its level of importance and usage at the dawn of Social network.

Researchers like Fergal Reid predicted the diffusion in community models by examining the structure of common networks and by spreading the information on them using an appropriate model of SIR. They showed that by increasing the structure of overlapping groups of network, we can have a faster way of spreading the information, whereas increasing the non-overlapping community structure doesn't allow for any of these advantages. The results showed that the role of weak nodes in an information diffusion was over-stated [1].

A varied form of clustering has been used and implemented at many fields. In here, we use density-based clustering which help to cluster the medium level data and reduces the time complexity as well.

Our project also aims at providing security for the information diffused over the network by using the algorithm used for security in computing. Mainly, we use the RSA algorithm to ensure security with the provision of security keys with the information that is being diffused to the different clusters. Hence, the provided information will be safe and can be viewed only by authorized users.

We implement data analysis and cluster analysis to find out the overlapping nodes using the DBSCAN algorithm and diffuse the information with security key to all over the clustersand hence effectively spread the information to all the nodes.

By comparing the diffusion by increasing the overlapping community structure and diffusion in shared nearest neighbor clustering between overlapping nodes, the proposed method provides the most accurate way for information diffusion.

2. Related work

A systematic research and explorations in social networks has been conducted by many researchers. EytanBakshy discussed the fundamental task of social networks in a diffusion of information. By exposing signals about friends that has been shared among 250 subjects in situ and proving that the information spreads faster when users who are exposed to the signals diffuse that information rather than those who aren’t and concluded that weak ties might play a major role in information diffusion than the strong ones.

[1]Gayathri et al, gave an ensemble of clustering which gave a combined version of both DBSCAN and Proclus algorithm to gain a cluster of even smallest of data points with a better quality. These combined algorithms were tested using synthetic datasets and they showed that the methods can also be used on every sort of datasets with high dimensions, regardless of their shapes.

[2] M.E.J Newman, argues that social networks vary from other types of networks including technological and biological networks with their significant clustering and the ability to show positive correlations, and demonstrating the group of structure in networks correlations by using a simple model that they expected as sortative mixing in such networks whenever there is a dissimilarity in the sizes of the groups and that the predicted level of separated mixing compared well with real world networks.

[3] Muhammad U. Ilyas observed that the most common problem in the analysis of social networks, is caused by the nodes that are always influenced and recognized the hubs in social network, and identified the influenced nodes at the centrality of neighbors with the help of PCC (principal component centrality).The graph of 70,000 different users who are friends in common on online social networking sites, were processed to propose a PCC’S performance. By comparing the different hubs that are formed by PCC, they could show that the adding of more such features in PCC, also gives a new set of hubs without having to replace the previous ones that were identified.

[4] Many types of experiments and algorithms were developed based on the approach of density based clustering. The immediate apprehension of promoting the definition given by DB scan clustering, was done by Hans-peter kriegal, with the method of statistical approach and an appropriate algorithm that can be applied on huge databases. Finding reliable density threshold is a tedious task.

Hierarchical methods were adapted and discussed to overcome this problem.”

[5] Comparative study of Density based Clustering Algorithms are done by Pooja BatraNagpal. Parameters of size 6 were taken to compare the algorithms. This experiment helps to choose an appropriate density based algorithms to cluster in various conditions.”

[6] Clustering depends basically on density and distance, but these concepts become increasingly tougher to define as the dimensionality increases.

LeventErtoz offered the definitions for density and the suitable measure for similarity to analyze the data in high dimensionality.

The work was done with the help of different measures in similarities that rely on the amount of neighborhood that the two points share and predicted those points that belong to the density as the sum of similarity between different points of the neighborhood that are near.

The experiment gave a new algorithm for clustering which is been focused on these ideas by eliminating number of noise and form a cluster by considering them. This approach was used for different clusters of different shapes and sizes.

This proposed study gave an example which involved very few datasets of high dimensionality.

[7] Social networking sites enable multiple usersacross the globe to provide and access contents of several news and topics. AdreinGuille presented a method oriented survey by representing the numerous issues faced by diffusion process and proposed a system that summarizes the actual status of the current network. The proposed system gave an absolute inclusive of social analysis and guided efforts that have already been made around diffusion of information in networking. The proposed work was mainly intended to provide researchers a tool to understand the existing works in a quicker way and possible improvements to ponder on. It is supervised with the user-friendly interfaces and gave an actual structure and visualizations.

[8]Clustering high dimensional data is often of interest. Clustering becomes difficult as the dimensionality of the data increases. It critically depends on density and similarity. Jian Yin described a shared nearest neighbor clustering algorithm with high dimensionality and evaluated it on the spatiotemporal data set of multidimensionality.

The proposed work examined the traditional SNN algorithm in the region of spatiotemporal cluster analysis of data and proposed a high dimensional shared nearest neighbor clustering algorithm step by step to solve shortcomings of SNN approach.

The proposed algorithm could overcome the spatial temporal complexity in an effective way and provided several performances including core points and results of clusters. The proposed work proved that DSNN can reduce computation effectively and it can also judge core points correctly and outliers as well, and gain better clustering performance than SNN algorithm with better clustering methods.”

3. Existing system

Information diffusion is an integral part in an online social networking service. The weight of spreading of information depends mainly on the number of people concerned with the diffusion and the appropriate amount of time consumed for the rapid spreading of news. Though there are several existing algorithms discussed and experimental status performed to help in diffusion process, none of them are effective enough for the rapid spreading and imperceptible usage of information in an accurate way. Nonetheless, even though we can spread information to all the nodes in the network, there are important cases in which usage of nearest neighbor nodes shared between overlapping nodes are more relevant.

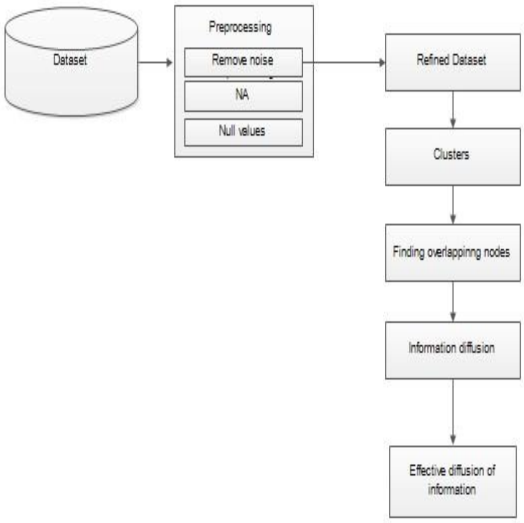
In an experiment done by LeventErtoz, it is found that the basic similarity among two points can beanalyzed by considering their commonly shared nearest neighbor node and the points that isn’t belong to any of the clusters can be taken near the closest point by assigning them to the cluster [6].

Shared nearest neighbor clustering algorithm with high dimensionality was predicted by researcher Jian Yin using a traffic dataset. In their work, they also observed that DSNN (dimensional nearest neighbor clustering algorithm) cannot deal with the flow of data [10].

4. Proposed system

Clustering of commonly shared nodes has a vital role in information diffusion. Nonetheless, even though we can spread information to all the nodes in a network, there are important cases in which usage of overlapping nodes is more accurate. Our objective is to effectively utilize the information in shared nearest neighbor nodes between overlapping nodes.

5. Architecture diagram



In our current proposed system, we have collected the student dataset and that dataset will undergo proposing process which involves removal of noise, null values, and other inconsistent data. After these processes, we obtain a complete refined data we form a cluster using DBSCAN. We then found that the overlapping node between the clusters can effectively diffuse the information throughout the network. This is what the above architecture diagram suggests.

6. Methodology

In our proposed system, finding the overlapping node or a SNN for which the information must be diffused for effective spread of information for both the clusters is performed. To do that, we first collected the student data that got preprocessed and being clustered then finding the SNN node. The brief description about the methods used is described below.

6.1. Data Collection

We have collected information from students of reputed colleges by creating a Google form with certain queries, thus collecting the required materials.

6.2. Prepossessing

We analysed the collected information and it will be prepossessed by eliminating the noise, null value, and not applicable data by which we will get the refined dataset.

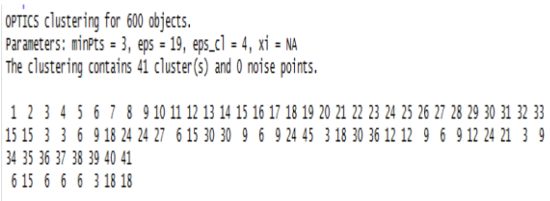
6.3. Clustering

The nodes with more than or equal to 80% of similarity is clustered with the elimination of noise and the null values using DBSCAN. Density Based Clustering Algorithm is a clustering algorithm in which the data is being clustered based on their density and the similarity between the groups. This makes the cluster by eliminating the noise and the non-noise data will be clustered based on their density.

6.4. Finding between anes of different clusters (SNN)

Finding the node which is being shared to two different clusters is also called as the between anes of two different clusters.

7. Experimental results and result analysis



```
OPTICS clustering for 600 objects.
Parameters: minPts = 3, eps = 19, eps_cl = 4, xi = NA
The clustering contains 41 cluster(s) and 0 noise points.

 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
15 15  3  3  6  9 18 24 24 27  6 15 30 30  9  6  9 24 45  3 18 30 36 12 12  9  6  9 12 24 21  3  9
34 35 36 37 38 39 40 41
 6 15  6  6  6  3 18 18
```

Figure 1: Shows display of the Clusters For The Dataset

The above picture represents the cluster of the complete dataset for 600 objects. There are 41 clusters being done based on the density and the similarity.

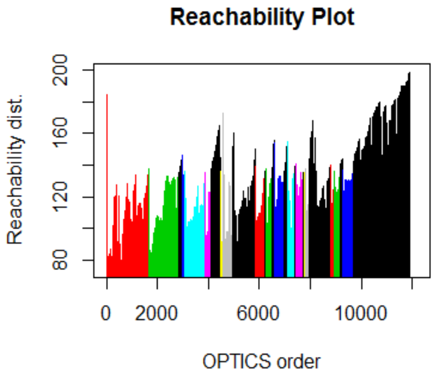


Figure 2: Chart shows result of OPTICS used in clustering with presence of noise.

The above graph represents the data that are with the presence of noise. In the above graph plot with the black colour represents the presence of noise.

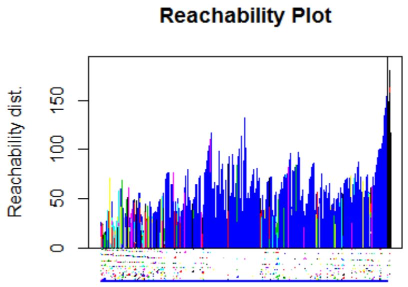


Figure 3: Result of DBSCAN clustering with the elimination of noise.

The above graph represents the non-noise data. That is, all the plots in black colour is being completely removed. The plot of the data with non-noise is ready for implementing Shared Nearest Neighbour Node.

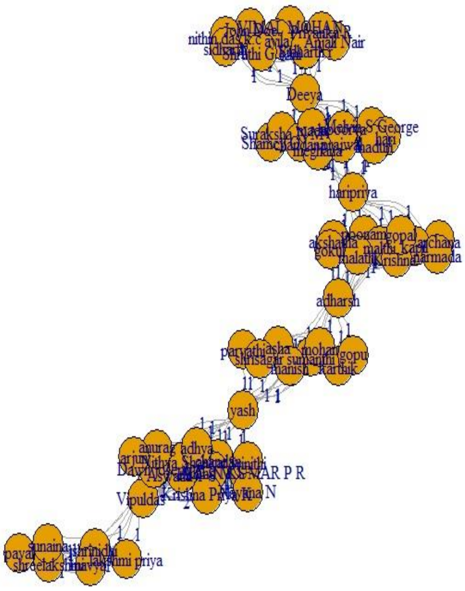


Figure 4: Result of finding shared nearest neighbour node

In the above graph representation of the overlapping nodes found, the nodes called "vipuldas", "yash", "adharsh", "haripriya", "deeya" are the nodes which have shared to both the clusters.If the information is passed to those shared nodes, then the effective spread of information will take place.

8. Conclusion

In this paper, we have proposed a DBSCAN clustering algorithm which combines several ideas to overcome many of the challenges faced in social networking. e.g., finding clusters in the presence of noise and outliers and data that has clusters of different shapes, density, and sizes. We first examined the process of information diffusion on various social networking services. After analyzing several densities based approaches, we adapted the DBSCAN clustering algorithm which suits well for medium-level data. We prove that information diffusion between overlapping nodes based on shared nearest neighbor node clustering has quicker benefit than those of overlapping nodes.

There are still some drawbacks in SNN approach. For example, we cannot always assure the security for the spreading of an information and the details of the user. How to provide security keys between the shared nodes and the authorization for various users are our future work.

References

[1] Agrawal R, Gehrke J, Gunopulos D &Raghavan P, *Automatic subspace clustering of high dimensional data for data mining applications*, Vol.27, No.2, (1998), pp.94-105.

[2] Bakshy E, Rosenn I, Marlow C &Adamic L, "The role of social networks in information diffusion", *Proceedings of the 21st international conference on World Wide Web*, (2012), pp.519-528.

[3] Ertoz L, Steinbach M & Kumar V, "A new shared nearest neighbor clustering algorithm and its applications", *Workshop on Clustering High Dimensional Data and its Applications at 2nd SIAM International Conference on Data Mining*, (2002), pp.105-115.

[4] Gayathri S, Metilda MM &Babu SS, "A Shared Nearest Neighbour Density based Clustering Approach on a Proclus Method to Cluster High Dimensional Data", *Indian Journal of Science and Technology*, Vol.8, No. 22, (2015), pp.1-6.

[5] Ilyas MU &Radha H, "Identifying influential nodes in online social networks using principal component centrality", *IEEE International Conference on Communications (ICC)*, (2011), pp.1-5.

[6] Nagpal PB & Mann PA, "Comparative study of density based clustering algorithms", *International Journal of Computer Applications*, Vol.27, No.11, (2011), pp.421-435.

[7] Reid F & Hurley N, "Diffusion in networks with overlapping community structure", *IEEE 11th International Conference on Data Mining Workshops*, (2011), pp.969-978.

[8] Yadav PS, Sharma P & Yadav DK, "Implementation of RSA algorithm using Elliptic curve algorithm for security and performance enhancement", *International Journal of Scientific & Technology Research*, Vol.1, No.4, (2012), pp.102-105.

[9] Milgram, S & Fergal R, "The small world problem", *Psychology Today*, Vol.2, No.1, (1967), pp.60–67.

[10] Watts, D & Strogatz, S, "Collective dynamics of 'small-world' networks", *Nature*, Vol.393, No.6684, (1998), pp.440–442.

[11] Akshay S. and Apoorva P, "Bandwidth optimized multicast routing algorithm based on hybrid mesh and tree structure with collision control in MANET using lempel-ziv-oberhumer method", *International Conference on Communication and Signal Processing (ICCSP)*, (2017), pp.0495-0500.

[12] Malvika DMK, Sandhya S & Akshay S, "Control of the Locomotion of Temperature Sensor", *International Journal of Applied Engineering Research*, Vol.10, No.6, (2015), pp.14405–14419.