

Community Detection Using Node Attributes and Structural Patterns in Online Social Networks

Bikash C. Singh^{1,2}, Mohammad Muntasir Rahman³, Md Sipon Miah^{1,4} & Mrinal Kanti Baowaly⁵

¹ Department of Information & Communication Engineering, Islamic University, Kushtia, Bangladesh

² DiSTA, University of Insubria, Varese, Italy

³ Department of Computer Science & Engineering, Islamic University, Kushtia, Bangladesh, and School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China

⁴ Department of Information Technology, National University of Ireland, Galway, Ireland

⁵ Department of Computer Science & Engineering, Bangabandhu Sheikh Mujibur Rahman Science & Technology University, Bangladesh, and Institute of Information Science, Academia Sinica, Taiwan

Correspondence: Bikash C. Singh, Department of Information & Communication Engineering, Islamic University, Kushtia, Bangladesh, and DiSTA, University of Insubria, Varese, Italy. E-mail: bikash070@gmail.com

All authors have contributed equally.

Received: October 4, 2017

Accepted: October 15, 2017

Online Published: October 26, 2017

doi:10.5539/cis.v10n4p50

URL: <http://doi.org/10.5539/cis.v10n4p50>

Abstract

Community detection in online social networks is a difficult but important phenomenon in term of revealing hidden relationships patterns among people so that we can understand human behaviors in term of social-economics perspectives. Community detection algorithms allow us to discover these types of patterns in online social networks. Identifying and detecting communities are not only of particular importance but also have immediate applications. For this reason, researchers have been intensively investigated to implement efficient algorithms to detect community in recent years. In this paper, we introduce set theory to address the community detection problem considering node attributes and network structural patterns. We also formulate probability theory to detect the overlapping community in online social network. Furthermore, we extend our focus on the comparative analysis on some existing community detection methods, which basically consider node attributes and edge contents for detecting community. We conduct comprehensive analysis on our framework so that we justify the performance of our proposed model. The experimental results show the effectiveness of the proposed approach.

Keywords: online social community, structural cluster, nodes' attribute cluster, set theory, probability

1. Introduction

Online social media (e.g., Facebook, Twitter) are the platforms where individuals are involved in making relationships with others. Intuitively, these relationships express their social interactions (e.g., friendships, followings, followers) which commonly exhibit the properties of communities structures. Basically, community indicates the groups of individuals with common attributes make a dense connection inside each group and less connection between groups. Generally, in graphical representation, individual and relationship represent nodes and connections between individuals respectively in social networks. Typically, individuals in each social community may have different types of certain interests as such common attributes like as hometown, school mate, living city or common preferences such as photography, movies, music or travel, and hence, they tend to interact more frequently with each other than with users who are outside of their community.

Recently researchers have been interested to detect communities in social networks due to several reasons. One of the main reason is to analyze the human common behaviors. For example, to analyze a product whether it could be preferred by the consumers or not. In such case, we can focus on the communities and easily judge consumers' opinions regarding on this product. However, there are several applications for detecting communities in social networks. Numerous techniques have been developed for detecting community so far.

However, most of them consider the node attributes, edge contents or structural patterns of networks for proposing community detection algorithms (Zhou, Cheng, & Yu, 2009) (Qi, Aggarwal, & Huang, 2012, April) (Yang, McAuley, & Leskovec, 2013, December). Indeed these are the possible ways to detect communities in online social networks. With this assumption, we consider node attributes and structural patterns of the networks in together so that our community detection model can exploit these two views in term of detecting communities. With this motivation, we concentrate on the research paper in (Zhou, Cheng, & Yu, 2009) for exploiting tools to detect structural clusters and nodes' attributes clusters. After that, using nodes' clusters and structural clusters, we propose a community detection model that can detect overlapping communities along with distinguished communities based on the set theory and probability (see Section 4).

The contributions of this paper are two-fold. First, we build up an alternative approach that can detect overlapping communities along with distinguished communities, which implies to overcome the limitations in (Zhou, Cheng, & Yu, 2009). Our proposed algorithm that incrementally detects evolving communities in social networks. It delivers significant improvements over the existing solutions in both the quality of detected evolving communities and the speed of program execution.

Second, we generalized the algorithm introduced in (Zhou, Cheng, & Yu, 2009) to incorporate important network features such as node attributes and structural clusters and compare with another method proposed in (Qi, Aggarwal, & Huang, 2012, April).

2. Related Work

For instance, various existed community detection methods have considered network structures and node attributes in a combined manner to detect communities of individual in online social networks (Yang, McAuley, & Leskovec, 2013, December) (Akoglu, Tong, Meeder, & Faloutsos, 2012, April) (Moser, Ge, & Ester, 2007, August). Unfortunately, these methods cannot detect overlapping communities. Literatures have focused on topic models (Balasubramanyan & Cohen, 2011, April) (Xu, Ke, Wang, Cheng, & Cheng, 2012, May) (Sun, Aggarwal, & Han, 2012), which considered soft node community memberships that allow to detect overlapping communities. However, these methods are not appropriate for modeling communities because they do not allow a node to have high membership strength to multiple communities simultaneously (Yang & Leskovec, 2012). In paper (Zhou, Cheng, & Yu, 2009), authors have proposed method to cluster a large-scale graph associated with attributes based on both structural and attribute similarities. However, the limitation of their method is that it can not detect the overlapping community. To overcome this situation by considering the same assumption (e.g., structural and attributes similarity) we proposed another way to detect community properly in case of overlapping scenarios.

Indeed, there are some novel research works focused to detect overlapping communities in online social network. In paper (Xing, Meng, Zhou, Zhou, & Wang, 2015), authors proposed an overlapping community detection algorithm by considering local community expansion. This algorithm is composed of a three-phase process; in the first phase, it uses the local structural pattern of the social network to find the local communities and extend these communities with larger one according to overlapping score and finally, classifies the nodes, which does belong to any community. An alternative work (Nguyen, Dinh, Nguyen, & Thai, 2011, October) proposed DOCA which classify the nodes into local communities by considering the number of interactions among the nodes and then tries to combine highly overlapped communities if they share significant substructures. Other detection trends include methods based on nodes splitting (Gregory, 2007), modularity (Nicosia, Mangioni, Carchiolo, & Malgeri, 2009) and link-based detection methods (Ahn, 2009).

3. Motivation

In this paper, we study both nodes' attributes clusters and edge contents clusters to detect online social networks communities and make comparative analysis on these two models. Moreover we also propose an another alternative way to detect overlapping communities which is very simple and straightforward. As (Zhou, Cheng, & Yu, 2009) can not detect overlapping communities, our proposed model can overcome this problem. With this motivation, we describe the node attributes Similarities/Structural model stated in (Zhou, Cheng, & Yu, 2009) that supposes to detect communities in social network by considering the followings.

- Structure-based Clustering/Community. Figure 1b shows clusters obtained from node connectivity, i.e., coauthor relationship. Authors are closely connected within a cluster even if they might have quite different topics, e.g., half work on XML and the other half work on Skyline in one of the clusters.
- Attribute-based Clustering/Community. Figure 1c shows another clustering result based on attribute similarity, i.e., topics. Authors within clusters work on the same topics; however, the coauthor relationship may be lost due to the partitioning so that authors are quite isolated in one of the clusters.

- Structural/Attribute Clustering/Community. Figure 1d shows the clustering result based on both structure and attribute information. This clustering result balances the structural and attribute similarities: authors within one cluster are closely connected; meanwhile, they are homogeneous on research topics.

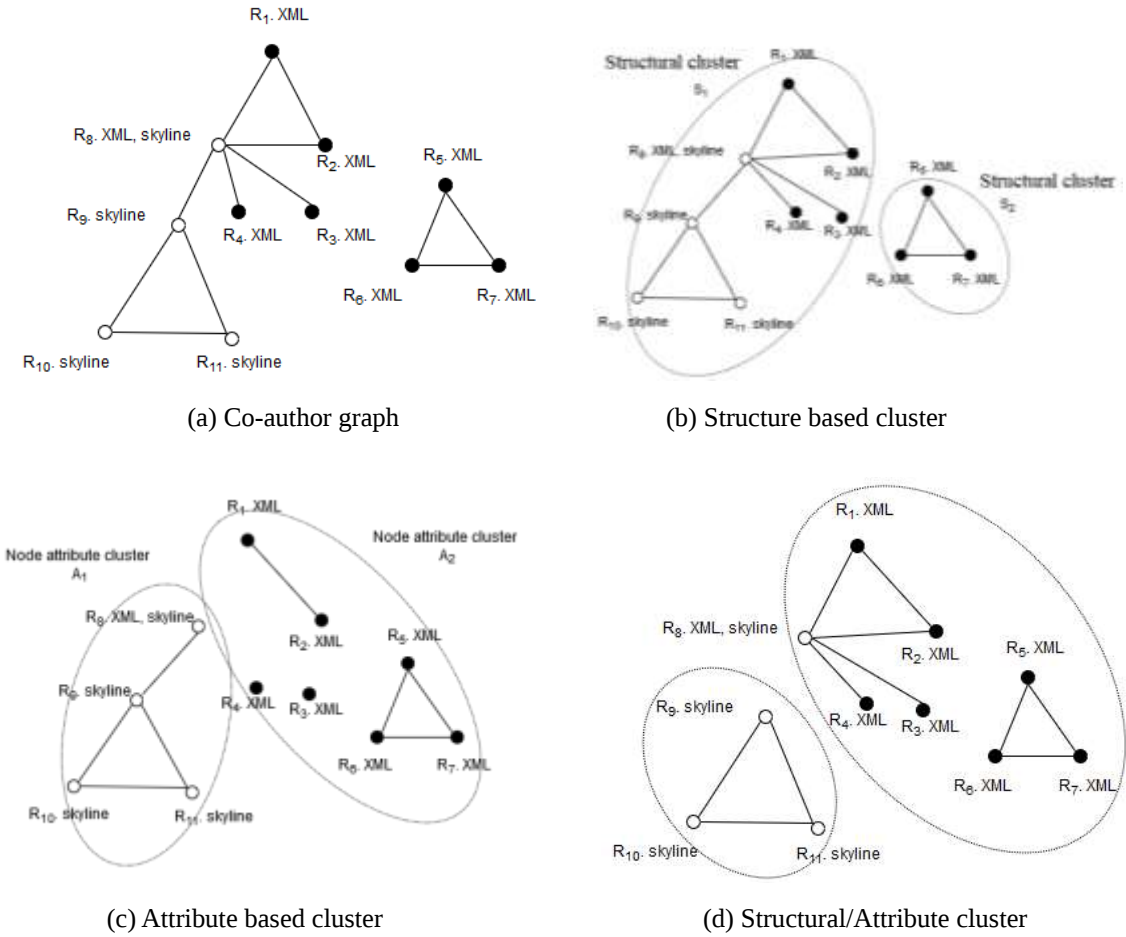


Figure 1. Coauthor network example with an attribute “topic”

The above model has some disadvantages as stated below:

- This model cannot identify the overlapping communities as shown in Figure 1d.
- It considers an augmented graph where they inserted attribute nodes and added extra edges among the attributed nodes and structural nodes. Therefore, the complexity of this model might be high.
- This model does not mention whether the nodes have no attributes then how it produces node attributes cluster.

Our motivation is to overcome these limitations as well as to improve the performance of the communities' detection algorithm. With this motivation, we are going to propose a community detection model for social networks so that our model can detect overlapping communities along with distinguished communities.

4. Proposed Model

The main aim of this paper is to propose an innovative community detection algorithm. With this intension, we mainly focus on the paper (Zhou, Cheng, & Yu, 2009) proposed SA-cluster method, and envision in different way so that our model can detect

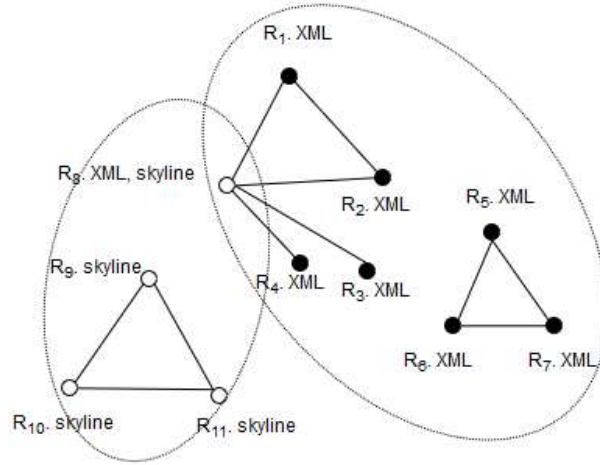


Figure 2. Overlapping communities

distinguished communities along with the overlapping communities. For doing so, we exploit the ideas proposed in SA-cluster for detecting the structural and node attributes communities individually. After successfully detecting the structural and node attributes communities, we use the following procedures to detect communities. For better explanation, we consider the Figure 1. Here, S_i represents the structural communities and A_j refers to the attribute communities, where $i=(1,2,3,...,n)$, and $j=(1,2,3,...,m)$.

In the first step of our proposed model, we find out the common nodes between structural clusters and nodes' attribute clusters for all combination. For example, structural based clusters as shown in Figure 1b and attribute based clusters as shown in Figure 1c, we proceed on to find out the common nodes for all combinations of structural communities and nodes' attributes clusters using the set theory as below:

$$S_1 \cap A_1 = \{R_8, R_9, R_{10}, R_{11}\}$$

$$S_1 \cap A_2 = \{R_1, R_2, R_3, R_4, R_8\}$$

$$S_2 \cap A_1 = \{\varnothing\}$$

$$S_2 \cap A_2 = \{R_5, R_6, R_7\}$$

More particularly, we are looking for the common nodes by exploiting the intersection among all combinations of structural clusters and nodes' attribute clusters. Those nodes are appeared both of structural clusters and nodes' attributes clusters, it means that these are the overlapping nodes involve in communities.

Secondly, for checking the overlapping communities in a given social network, we then exploit the probability theory. For this purpose, we compute the probabilities among all intersections of the combinations of structural clusters and attributes clusters, have been obtained in the first step. More precisely, at first, we progress with two sets of intersections obtained in the first step and consider same the attributes' cluster and different structural clusters. In this case, if a node belongs to both clusters then the probability might be non-zero. It means that the communities are overlapping with each other. Then we consider the next intersection obtained in first step with the previous overlapping communities, take into same consideration and so on until the final intersection. In this way, we can check all possible overlapping communities. Therefore, the probability can be computed as bellow:

$$P(d) = \begin{cases} \text{zero} & \text{No overlapping} \\ \text{greater than zero} & \text{Overlapping} \end{cases} \quad (1)$$

where, d is $(S_i \cap A_j) \cap (S_{i+1} \cap A_j)$, i is the number of structural communities and j is the number of attribute communities. If the probability is equal to zero, it seems that there is no overlapping nodes, otherwise nodes belongs to overlapping communities. In the final step, we find the communities. For doing so, we consider all combinations of intersections among attribute clusters and structural clusters obtained in the first step and do the same process as do in the second step. But here we consider the union operation among the set of intersections to discover communities. For instance, we have obtained 4 combination of intersections in the first step. Now we make union of these combinations by considering different structural clusters and identical attribute clusters. Thus, we can obtain a community as follow:

$$(S_1 \cap A_1) \cup (S_2 \cap A_1) = \{R_8, R_9, R_{10}, R_{11}\}$$

Similarly, we can detect another community by considering the identical nodes' attribute cluster and diverse structural clusters as given below:

$$(S_1 \cap A_2) \cup (S_2 \cap A_2) = \{R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8\}$$

If nodes are overlapping with other communities, our model can detect these overlapping communities as explained in the above and the resultant communities are shown in Figure 2.

Lemma 0.1. If n be the number of structural clusters and m be the number of nodes' attributes clusters then the total number of intersections is $n \times m$. Thus $(n \times m) - 1$ represents time complexity to find out all overlapping communities.

The proposed Algorithm 1 exhibits the whole idea. First, we consider the structural clusters and nodes' attribute clusters as input to our proposed algorithm, produced by the algorithm in (Zhou, Cheng, & Yu, 2009). Algorithm 1 depicts **Cn** as the final list of nodes appear in communities. Eventually, our proposed method computes all combinations of intersection **Ck** among structural clusters and nodes' attributes clusters (see line 2 in Algorithm 1). After then, it computes the intersection among the combinations of **Ck** for finding any node that appear in more

Algorithm 1 Community detection based on structural and nodes' attribute clusters

Input: S_i refers to structural clusters and A_j refers to nodes' attributes clusters
Output: Cn refers to nodes appeared in communities

```

1: Let  $S_i$  refers to structural clusters and  $A_j$  refers to nodes' attributes clusters
2: Calculate  $Ck = (S_i \cap A_j), k = 1, 2, \dots, n; i = 1, 2, \dots, n; j = 1, 2, \dots, m$ 
3: for  $x \in i$  do
4:   for  $y \in j$  do
5:     Calculate  $Probability = \cap_y^x P(S_{y+1} \cap A_x)$ 
6:     if ( $Probability == 0$ ) then
7:       No overlapping
8:       Calculate  $Cn = \cup_y^x (S_{y+1} \cap A_x)$ 
9:     else
10:      Overlapping exist
11:      Calculate  $Cn = \cup_y^x (S_{y+1} \cap A_x)$ 
12:    end if
13:    Return: Cn
14:  end for
15:  Return: Cn
16: end for

```

than one community and find the probability to check whether the nodes belongs to overlapping communities or not (see line 5 in Algorithm 1). In the next step, if the probability is greater than zero then it produces overlapping communities (see line 10 - 11). Finally, it provides the list of nodes **Cn** which appears in communities (as seen in lines 8, 11 in Algorithm 1). However the time complexity of the proposed Algorithm is $O(n \times m)$ where n is the number of structural clusters and m refers to the number of nodes' attribute clusters.

5. Comparative Analysis on Different Community Detection Models

5.1 Structural/Attribute Based Community Detection Model

In (Zhou, Cheng, & Yu, 2009), authors stated that structural and attribute similarities are two seemingly independent with each other. So it is difficult to incorporate these two communities in together. However, authors proposed a possible solution by considering distance function between two vertices v_i and v_j define as (Zhou, Cheng, & Yu, 2009):

$$d(v_i, v_j) = \alpha \times d_s(v_i, v_j) + \beta \times d_A(v_i, v_j) \quad (2)$$

where $d_S(v_i, v_j)$ and $d_A(v_i, v_j)$ measure the structural distance and attribute distance respectively, while α and β are the weighting factors. For doing so, authors insert a set of attribute vertices to a graph G . An augmented attributed Graph is shown in figure 3. An attributed graph is denoted as $G = (V, E, \Lambda)$, where V is the set of vertices, E is the set of edges, and $\Lambda = \{a_1, \dots, a_m\}$ is the set of m attributes associated with vertices in V for describing vertex properties. Each vertex $v_i \in V$ is associated with an attribute vector $[a_1(v_i), \dots, a_m(v_i)]$ where $a_j(v_i)$ is the attribute value of vertex v_i on attribute a_j . Researchers (Zhou, Cheng, & Yu, 2009) denote the size of the vertex set as $|V| = N$. Attributed graph clustering is to partition an attributed graph G into k disjoint sub graphs $G_i = (V_i, E_i, \Lambda)$, where $v = \bigcup_{i=1}^{\infty} F_i$ and $v_i \cap v_j = \emptyset$ for any $i \neq j$.

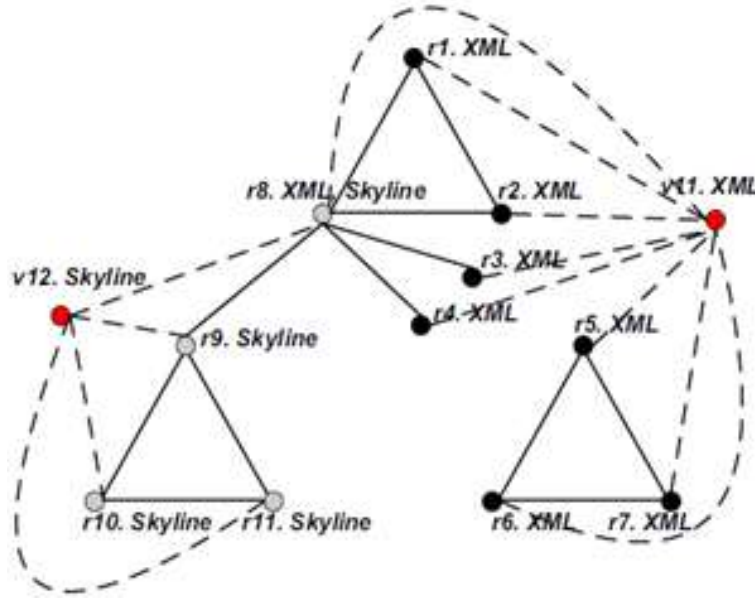


Figure 3. Augmented attribute graph

5.2 Edge Content Sharing/Structural Community Detection Model

There is another way to detect online social networks communities by considering both of edge content and structural similarities as shown in Figure 4. In this model (Qi, Aggarwal, & Huang, 2012, April), social network is denoted by the pair $G = (V, E)$. The vertex set V contains n nodes $\{v_1, v_n\}$ and the edge set E contains the m edges $\{e_1, \dots, e_m\}$. Let Γ denotes a $m \times n$ link matrix between edge and vertex set that encodes the link structure. For each edge e_i and vertex v_j , the value of $\Gamma_{i,j}$ is set to 1 if v_j is incident to e_i , otherwise $\Gamma_{i,j} = 0$. For detecting structural communities, they consider the edge set is E and latent representation of vertex set is V . Each column of V is the latent feature vector for the corresponding vertex. The researchers use the matrix factorization technique to design E and V such that the matrix product of E^T and V approximately represents the link matrix Γ . The optimum values are defined by minimizing the error of the approximation. $E^*, V^* = \argmin \|E^T \cdot V - \Gamma\|^2$. Δ is a $m \times n$ degree matrix. Replacing the value of V becomes, $O^1(E) = \|E^T \cdot E \cdot \Delta - \Gamma\|^2$. For detecting edge content based communities, the researchers assume that the content associated with each edge e_i , is denoted by c_i . They assume that the d -dimensional feature vector to represent the document c_i is denoted by $f_c(c_i)$, for notational purposes, they introduce a $d \times m$ matrix C to denote these extracted feature vectors.

In this matrix, the i th column contains the content feature vector $f_c(c_i)$ associated with the edge c_i , consider two edges e_i and e_j , with associated content denoted by c_i and c_j , and corresponding feature vectors denoted by $f_c(c_i)$ and $f_c(c_j)$ respectively. In this case, one can compute the cosine similarity $S_{i,j}$ between the two feature vectors as follows (Qi, Aggarwal, & Huang, 2012, April):

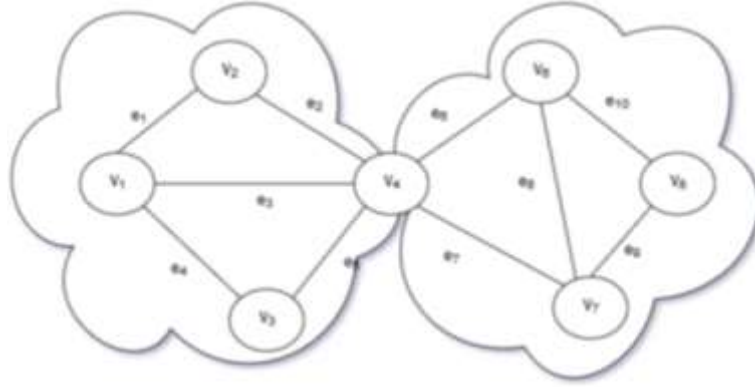


Figure 4. An example of edge content/structural community

$$S_{i,j} = \frac{f_c(c_i)^T \cdot f_c(c_j)}{\sqrt{f_c(c_i)^T \cdot f_c(c_i)} \cdot \sqrt{f_c(c_j)^T \cdot f_c(c_j)}} \quad (3)$$

The edge content based communities can be detected using the following equation (Qi, Aggarwal, & Huang, 2012, April):

$$O^c(E) = \min S_{i,j} \cdot \left\| \frac{f_e(e_i)}{\sqrt{s_i}} - \frac{f_e(e_j)}{\sqrt{s_j}} \right\|^2 \quad (4)$$

So, Edge content/structural based communities detection model is (Qi, Aggarwal, & Huang, 2012, April):

$$O(E) = O^1(E) + \alpha \cdot O^c(E) \quad (5)$$

Now, in term of comparing Equation 2 which is based on node attributes/structural similarities and Equation 5 based on edge content/structural similarities is differed only in the second part of the both equations since the first parts of these equations are related with the structural similarities.

In edge content/structural similarities model is indirect process to detect communities because it partitions the edges based on the content. However as we mentioned earlier in Section 3 that whether the nodes have no attributes then how the proposed method (Zhou, Cheng, & Yu, 2009) can proceed on. We believe that if we can merge the both of content/structural similarities and node attributes/structural similarities then we can use the edge contents to set up the value of attributes of nodes. We believe that this could be given more efficient result than others model. Because, It may be happened that every nodes may have no sufficient attributes values and reversely, the content of edge (message) may have no significant value to detect communities. By considering these of types matters into account, we plan to propose another community model in future.

6. Experimental Results

We carried out several experiments to measure the effectiveness of our proposed method in terms of detecting whether individuals involve in communities or not in online social networks.

6.1 Dataset

We use the DBLP (<http://dblp.uni-trier.de/xml/>) Bibliography dataset for evaluating the performance of our proposed method. We consider 1,000 authors and their coauthor relationships. Moreover, we consider research papers published in DBLP to make communities among authors. For this purpose, we use two relevant attributes namely prolific and primary topic. Authors are labeled as highly prolific if (s)he has ≥ 20 papers in DBLP and authors with <10 papers are labeled as low prolific. For attribute “primary topic”, we use a topic modeling approach (Zhou, Cheng, & Yu, 2009) to extract 50 research topics from a document collection composed of paper titles from the selected authors. The topic are most representative by considering the probability distribution of keywords.

6.2 Metrics

In order to measure the effectiveness of the proposed approach, we make use of conventional measures. In particular, since we have decisions in two categories such as whether the model can detect a node appeared in a community correctly or not. We have 2 labels (i.e., node belongs to community, node does not belongs to community), thus we exploit a 2X2 confusion matrix, see Table 1, representing the adopted notations. More

precisely, row of the matrix represents the actual value for a class, column represents a possible measured value, and an element identified by row and column specifies the type of error, if any, in measuring an item whose real value is specified in the row with the label corresponding to the column.

Table 1. Confusion matrix

	Measured value: Nodes belong to communities	Measured value: Nodes do not belongs to communities
Actual value: Nodes belong to communities	TP	FN
Actual value: Nodes do not belongs to communities	FP	TN

Table 2. Metrics definition

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{total number of samples}$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

6.3 Accuracy

In this experiment, we first check the accuracy of our proposed model in term of varying the number of communities. The obtained result has been shown in Figure 5. According to Figure 5, we show that the accuracy of our proposed model is deteriorated when number of communities increase and after a certain time it seems to be fixed at a level of 75% approximately. Moreover, we make a comparison between SA-cluster proposed in (Zhou, Cheng, & Yu, 2009) and our proposed model as shown in Figure 5. The experiment shows that SA-cluster produces better accuracy than our proposed model when we consider small number of communities. For further analysis, we increase the number of communities which may implies that nodes may have chance to be member of more communities. With this assumption, as depict in Figure 5, we see that the accuracy of SA-cluster going down and our proposed model perform better than SA-cluster since SA-cluster could not able to detect overlapping communities. AS our model consider the overlapping communities thus it produces good accuracy for detecting communities in case of the number of communities are increased.

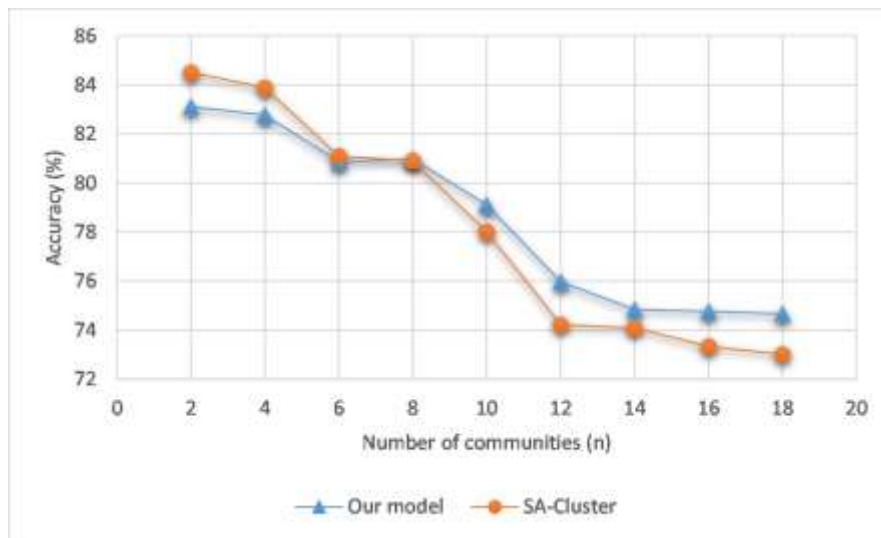


Figure 5. Accuracy for number of communities

Furthermore, we consider non-overlapping and overlapping communities separately and check the accuracy produced by our proposed model. The Figure 6 shows the obtained result that exhibits when we consider only

the non-overlapping dataset then the accuracy of our model is better than the overlapping dataset.

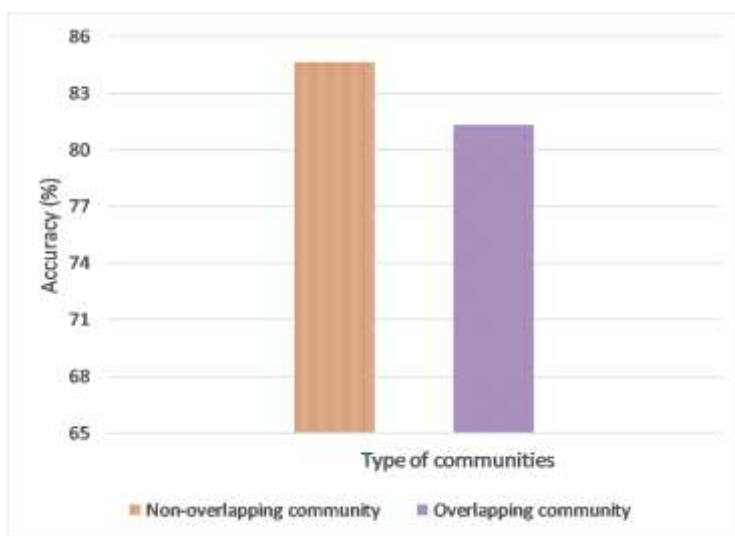


Figure 6. Accuracy for non-overlapping and overlapping communities

6.4 F1 Score

We measure the F1 score for evaluating the performance of our proposed approach. F1 score considers both the precision and the recall. The precision is the ratio of the number of correctly measured items (i.e., access requests) to the total number of measured items. Whereas, recall is the ratio of the number of correctly measured items to the total number of relevant items. Our analysis exhibits that our proposed approach works better when the number of communities is in small size and its' performance is degrading when number of communities is increased as shown in Figure 7. The obtained results also show that after a certain period (e.g., when the number of communities is 16) it produces constant accuracy.

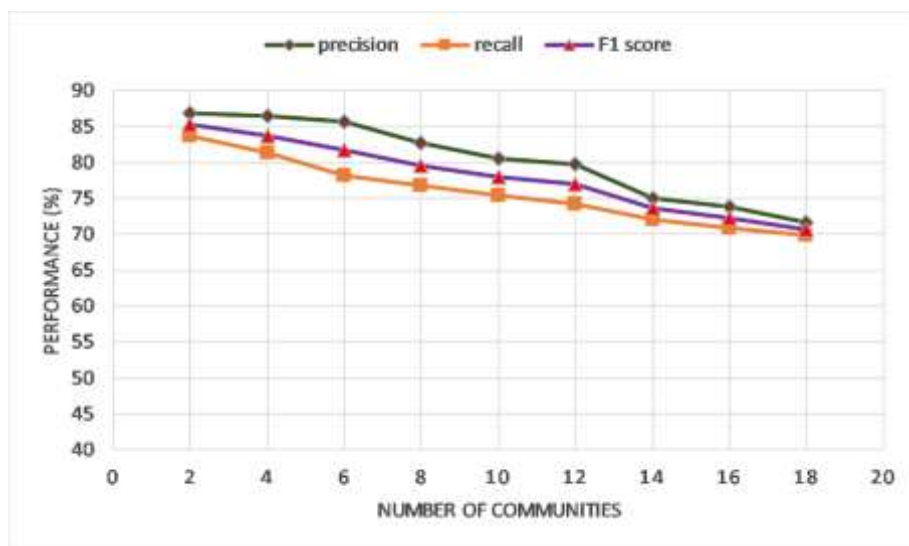


Figure 7. F1 score

7. Conclusion

In this paper, we have exploited set theory and probability theory in term of detecting online social communities. More precisely, we have used probability to check whether a node belongs to overlapping communities or not and then we have used set theory to enlist the number of nodes belongs to online social communities. In the future, we plan to define a mechanism that includes strategies on support of improving false positives/false negatives so as we can give more attention on misconfiguration cases. With this direction, we can improve the performance of the proposed algorithm.

References

- Ahn, Y. Y. (2009). Link communities reveal multiscale complexity in networks. *arXiv preprint arXiv:0903.3178*. <https://doi.org/10.1038/nature09182>
- Akoglu, L., Tong, H., Meeder, B., & Faloutsos, C. (2012, April). PICS: Parameter-free identification of cohesive subgroups in large attributed graphs. In *Proceedings of the 2012 SIAM international conference on data mining* (pp. 439-450). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611972825.38>
- Balasubramanyan, R., & Cohen, W. W. (2011, April). Block-LDA: Jointly modeling entity-annotated text and entity-entity links. In *Proceedings of the 2011 SIAM International Conference on Data Mining* (pp. 450-461). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611972818.39>
- Gregory, S. (2007). An algorithm to find overlapping community structure in networks. *Knowledge discovery in databases: PKDD 2007*, 91-102. Retrieved from http://dx.doi.org/10.1007/978-3-540-74976-9_12
- Moser, F., Ge, R., & Ester, M. (2007, August). Joint cluster analysis of attribute and relationship data without-a-priori specification of the number of clusters. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 510-519). ACM. <https://doi.org/10.1145/1281192.1281248>
- Nguyen, N. P., Dinh, T. N., Nguyen, D. T., & Thai, M. T. (2011, October). Overlapping community structures and their detection on social networks. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on* (pp. 35-40). IEEE. <https://doi.org/10.1109/PASSAT/SocialCom.2011.16>
- Nicosia, V., Mangioni, G., Carchiolo, V., & Malgeri, M. (2009). Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03), P03024. <https://doi.org/10.1088/1742-5468/2009/03/P03024>
- Qi, G. J., Aggarwal, C. C., & Huang, T. (2012, April). Community detection with edge content in social media networks. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on* (pp. 534-545). IEEE. <https://doi.org/10.1109/ICDE.2012.77>
- Sun, Y., Aggarwal, C. C., & Han, J. (2012). Relation strength-aware clustering of heterogeneous information networks with incomplete attributes. *Proceedings of the VLDB Endowment*, 5(5), 394-405. <https://doi.org/10.14778/2140436.2140437>
- Xing, Y., Meng, F., Zhou, Y., Zhou, R., & Wang, Z. (2015). Overlapping community detection by local community expansion. *Journal of Information Science and Engineering*, 31(4), 1213-1232. <https://doi.org/10.1109/ICTAI.2012.61>
- Xu, Z., Ke, Y., Wang, Y., Cheng, H., & Cheng, J. (2012, May). A model-based approach to attributed graph clustering. In *Proceedings of the 2012 ACM SIGMOD international conference on management of data* (pp. 505-516). ACM. <https://doi.org/10.1145/2213836.2213894>
- Yang, J., & Leskovec, J. (2012). Structure and overlaps of communities in networks. *arXiv preprint arXiv:1205.6228*. <https://doi.org/10.1145/2594454>
- Yang, J., McAuley, J., & Leskovec, J. (2013, December). Community detection in networks with node attributes. In *Data Mining (ICDM), 2013 IEEE 13th international conference on* (pp. 1151-1156). IEEE. <https://doi.org/10.1109/ICDM.2013.167>
- Zhou, Y., Cheng, H., & Yu, J. (2009). Graph Clustering Based on Structural/Attribute Similarities. *Journal Proceedings of the VLDB Endowment*, 2(1), 718-729. <https://doi.org/10.14778/1687627.1687709>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).