

FINDING PROMINENT FEATURES IN COMMUNITIES IN SOCIAL NETWORKS USING ONTOLOGY

Vijay Nayak¹ and Bhaskar Biswas²

Department of Computer Engineering,
Indian Institute of Technology (BHU), Varanasi 221005 India

vijay.nayak.cse09@iitbhu.ac.in

bhaskar.cse@iitbhu.ac.in

ABSTRACT

Community detection is one of the major tasks in social networks. The success of any community depends upon the features that were selected to form the community. So it is important to have the knowledge of the main features that may affect the community. In this work we have proposed a method to find prominent features based on which community can be formed. Ontology has been used for the said purpose.

KEYWORDS

Social networks, community detection, Ontology, feature selection.

1. INTRODUCTION

Community Detection is one of the major tasks in social network analysis. Communities in social networks have a wide range of applications like viral marketing, sharing of information, sentiments, emotions etc. Communities are group of people/actors/ items who share some common topic(s) of interest. Most of the times, while detecting any community in a social network, many parameters are taken into account. For example, if anyone wants to find a community of interest, let say movies, in a social network he/she is connected to, may apply some parameter like actors, name of movie, genre of movie etc. to find such community. In general, more is the number of parameters, better is the community formed. At the same time, increasing the number of parameters increases the time complexity and computational complexity both. So, while adding any parameter or features for community detection, one must ensure that the addition should enhance the results of community detection algorithms in use. The major problem is that many times, it is difficult to find out the prominent features that can be used in community detection. There are two reasons behind this. First one, being the size of data to which the community detection algorithm is applied is huge. Second one is related to the first one, that is, due to this huge data many times the features selection cannot be done properly or some features are hidden or so.

One of the ways to reduce the size of data as stated in [14] is Ontology. In [14], the authors proposed Ontology as a means to represent the conceptual view of the data thereby reducing the size of the data. Ontology is explicit specification of conceptualization of a domain. In [14], the authors proposed a model to evaluate algorithms that can be applied to social networks using

Dhinaharan Nagamalai et al. (Eds) : ACITY, WiMoN, CSIA, AIAA, DPPR, NECO, InWeS - 2014

pp. 31–36, 2014. © CS & IT-CSCP 2014

DOI : 10.5121/csit.2014.4504

Ontology. Inspired by this we extend the work of [14] to find out the prominent features effecting the formation of communities in social networks.

Ontology was proposed for the semantic web for W3C.org. This inspired many researchers like [2],[4],[6],etc. for web search engine and web crawlers. In [1],[3],[8],[7],[12] and [11]the authors demonstrated the use of ontology in social networks in different ways. In 2013, Régine Lecocq et.al. [13] proposed a generalised prototype for analysing social networks through Ontology. The organisation of the rest of the paper is as follows: In section 2, the description of the proposed method/model is done. Section 3 describes the experiments we have performed and section 4 gives the results and conclusion which is followed by references in section 5.

2. METHOD

In the proposed method, the data from the social networks is extracted by web crawler/ apps/ any other tool and is stored in the database. Ontology was created capturing the features and properties of the extracted database. Since the ontology is used to represent different features of the dataset it can be used to represent the data itself for conducting any type of experiments. Any predefined or trivial communities based on the features of the dataset can be used as input for communities. Different community detection algorithms can be used to find communities from the dataset. As the community detection can be done using clustering algorithms, so in this work Markov Chain Clustering* algorithm is used to find communities. Further, communities and clusters are used interchangeably. Two nodes/actors in social networks are connected to each other by a specific relationship. This relationship is based on a set of properties or features of the nodes. These properties are represented in the form of ontology. Such a relationship is used to form communities. These communities are combined with previously stored ontology to form a modified ontology.

This modified ontology is then used to find the degree of overlap for every feature that is included in the ontology. The degree of overlap in this context is defined as intersection between communities. In an ideal community there should be distinctness in the nodes. So, lesser is the overlapping of the nodes, greater is the distinctness between them. The higher the degree of overlap the lesser is the intra-cluster distance between two communities or clusters and vice-versa. This observation resulted in concluding that if a certain feature had lesser degree of overlap than the other features over the specified communities then that feature had more influence in the formation of that specific cluster. The degree of overlap for every feature is calculated and compared. A feature which has minimum degree is observed to be influencing the formation of the community. For example, a person is a friend of some individual, this relationship of friendship can be used to form clusters. The properties/features which influence the friendship like location, movies, music and books, etc. are used in designing the ontology. A modified ontology is then created to find the degree of overlap for each property. These degrees of overlap are then compared to find which feature has minimum value thus influencing the formation of the clusters. The above method is described in the model shown in figure 1.

3. EXPERIMENT

3.1. Data preparation:

The ontology was created from a dataset extracted from Facebook (www.facebook.com). This dataset consisted of properties of each individual like location (city, state and country), likes of music and movies. The communities were formed by using MCL algorithm. The parameter used for clustering in MCL algorithm is friendship between each individual. The ontology and the clusters were combined to form a modified ontology. The degree of overlap for each feature/property was calculated using this modified ontology.

*www.Wikipedia.org

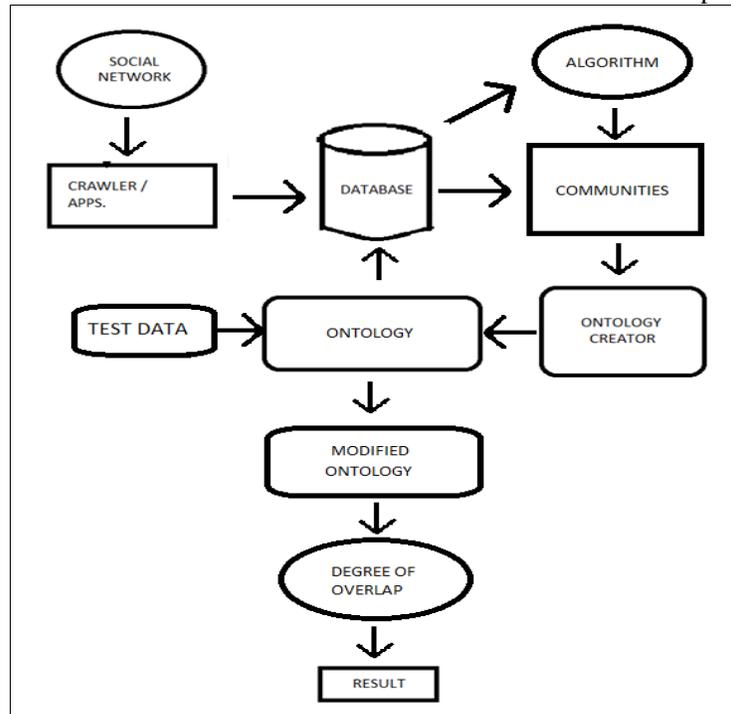


Figure 1. Model

3.2. Ontology Design

The tool Protégé was used for creation of Ontology [16] [18]. Ontology was designed with the features/properties as classes. The object properties in the ontology were defined so that it reflected the relation between the users and their features. This ontology was stored so that it could be reused for any further experiments using different algorithms. Here we used the MCL algorithm for finding clusters. The ontology was then combined with these clusters. The following is the class hierarchy of the modified ontology and its object properties:

Class hierarchy:

- Thing
 - Users
 - City
 - State
 - Country
 - Movies
 - Music
 - Cluster

Object properties:

- | | |
|-----------------|-----------------|
| • hasHomeCity | • isHomeCityOf |
| • hasHomeState | • isHomeStateof |
| • hasCountry | • isCountryOf |
| • hasLikedMovie | • MovieLikedBy |
| • hasLikedMusic | • MusicLikedBy |

The relations between users and its features were defined using the above object properties. For example, if the user U1 liked a movie M1 then the relation was defined as, U1 *hasLiked* Movie M1. The clusters obtained from the MCL algorithm were also mapped to their respective users in the same manner.

3.3. Observations:

Firstly, the degree of overlap was calculated. The degree of overlap for every feature with each cluster was calculated. It was defined as the ratio between the numbers of individuals of that feature belonging to cluster with the total number of individuals of that feature. For example, degree of overlap of feature ‘City’ for any cluster[i] is as follows:

$$\text{Degree of overlap (city, cluster[i])} = (\text{number of cities which have users in cluster[i]}) / (\text{total number of cities}) \quad \dots(1)$$

Then a total degree of overlap was calculated for every feature. It was defined as the ratio between the number of individuals in all the clusters with the product of total number of individuals of that feature and total number of clusters.

$$\text{Total Degree of overlap (city)} = (\text{total number of cities in all the clusters}) / (\text{total number of cities}) \times (\text{number of clusters}) \quad \dots(2)$$

The number of individuals in a certain cluster was found out using the DL Query in the tool Protégé. The tool Protégé used its inbuilt ‘reasoner’, in our case ‘Fact++’ to help DL Query. For example, the Table 1 below shows calculation of the values of degree of overlap for the feature home ‘City’ in the ontology using formula (1).

Table 1 Degree of Overlap (City)

	CITY		
	NUMBER OF CITIES	TOTAL	DEGREE OF OVERLAP
cluster 1	8	70	0.114286
cluster 2	8	70	0.114286
cluster 3	59	70	0.842857
cluster 4	4	70	0.057143
cluster 5	2	70	0.028571
TOTAL	81	350	0.231429

In the same manner, the rest of the values in Table 2 were calculated using formula (2).

Table 2 Total Degree of Overlap

Degree of overlap	City	State	Country	Movies	Music
cluster 1	0.114286	0.2	1	0.37415	0.617902
cluster 2	0.114286	0.333333	1	0.609524	0.331088
cluster 3	0.842857	0.933333	1	0.462585	0.342637
cluster 4	0.057143	0.2	1	0.160544	0.117421
cluster 5	0.028571	0.066667	1	0	0.00385
TOTAL	0.231429	0.346667	1	0.321361	0.282579

We observed the following:

1. For each cluster, some of the features had their degree of overlap greater than the other features. As shown in Table 2, the degree of overlap for 'cluster 1' for the feature 'city' i.e. 0.11 is lesser than that for music in the same cluster i.e. 0.61.
2. The total degree of overlap for some features was lesser than that for other features. For example, the total degree of overlap for 'City' is the lower than the other features, i.e.0.23 as shown in Table 2. Whereas that of the feature 'Country' is '1' in all the cases because every user had the same country. This showed the maximum overlap condition that was possible is all the data in cluster are similar and overlapping.

In the proposed method, the data from the social networks is extracted by web crawler/ apps/ any other tool and is stored in the database. Ontology was created capturing the features and properties of the extracted database. Since the ontology is used to represent different features of the dataset it can be used to represent the data itself for conducting any type of experiments.

4. RESULTS AND CONCLUSION:

The graph (Figure 2) below reflects the observations from Table 2. For each cluster, the bar with lowest height represents the feature which is more prominent than the others for that cluster. As shown in the graph, for cluster 1 the prominent feature is 'City' and in cluster 3 the prominent feature is 'Music'. The total degree of overlap shows the feature which is prominent for formation of all the communities.

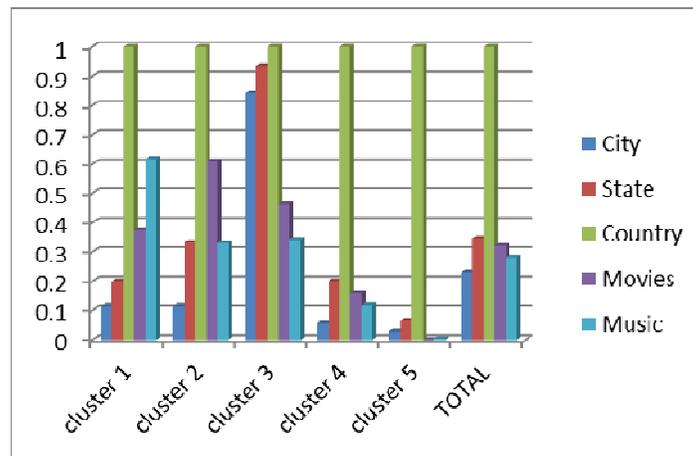


Figure 2. Degree of overlap for different clusters

Thus, the proposed method can be used to find which of the features are prominent in the formation of communities for different algorithms. Moreover, the ontology that was previously stored can be reused for different algorithms and also updated whenever needed for improving the results. Depending upon the prominent features we can use it for viral marketing, strategy planning, feature selection, etc. For future work, we plan to find sub-clusters among the already formed clusters. Then finding which features have affected the formation of these sub-clusters or sub-communities. This may give us knowledge as to how communities and their sub-communities are related with each other and on through which features.

REFERENCES

- [1] G.Aghila et.al, “Ontology-based Web Crawler”, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC’04), 2004.
- [2] Peter Mika, “Ontologies are us: A unified model of social networks and semantics”, Lecture Notes in Computer Science (Springer/ The Semantic Web – ISWC 2005) Volume 3729, pp 522-536, 2005.
- [3] Debajyoti Mukhopadhyay, Arup Biswas, Sukanta Sinha, “A New Approach to Design Domain Specific Ontology Based Web Crawler”, 10th International Conference on Information Technology (ICIT 2007), 289-291, 2007.
- [4] Jason , “J. Jung, Jérôme Euzenat, “Towards Semantic Social Networks”, Lecture Notes in Computer Science (Springer/The Semantic Web: Research and Applications), Volume 4519, pp 267-280, 2007.
- [5] Wu Chensheng, Hou Wei, Shi Yanqin, Liu Tong, “A Web Search Contextual Crawler Using Ontology Relation Mining”, International Conference on Computational Intelligence and Software Engineering (CiSE 2009) Page 1-4, 2009.
- [6] Wu Peng, Li SiKun, “Social Network Visualization via Domain Ontology”, International Conference on Information Engineering and Computer Science (ICIECS 2009) DOI: 10.1109/ICIECS.2009.5362898, 2009.
- [7] Liu Chen, Shan Wei, Zhang Qingpu, “Semantic Description of Social Network Based on Ontology”, Proceedings of the 2010 International Conference on E-Business and E-Government (ICEE ‘10), Pages 1936-1939, 2010.
- [8] Morteza Jamalzadeh, Navid Behravan, “Using Semantic Web Ontologies for better inter-operability on social network sites”, IEEE International Conference on Control System, Computing and Engineering, Page 103-108, 2011.
- [9] Reena Mishra, Shashwat Shukla, Deepak Arora, Mohit Kumar, “An Effective Comparison of Graph Clustering Algorithms via Random Graphs”, International Journal of Computer Applications (0975 – 8887) Volume 22– No.1, May 2011.
- [10] Sam K. M. and Chatwin C.R., "Ontology-Based Text-Mining Model for Social Network Analysis", Proceeding of the Sixth IEEE International Conference on Management of Innovation and Technology, pp. 226 - 231, 2012.
- [11] Moharram Challenger, “The Ontology and Architecture for an Academic Social Network”, International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 2, No 1, PP 22-27, SSN (Online): 1694-0814, 2012.
- [12] Régine Lecocq, Étienne Martineau, Maria Fernanda Caropreso, “An Ontology-based Social Network Analysis Prototype”, IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), DOI: 10.1109/CogSIMA.2013.6523839, Page(s): 149 – 154, 2013.
- [13] Bhaskar Biswas, Vijay Nayak and Harish Kumar Shakya, “Comparison of Algorithms for Social Networks using Ontology”, International Journal of Computer Applications, 85(13):31-34 January 2014.
- [14] <http://en.wikipedia.org/wiki/Ontology>.
- [15] http://en.wikipedia.org/wiki/Web_Ontology_Language.
- [16] <http://protege.stanford.edu>.

AUTHORS

Vijay Nayak is a post graduate student at Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, India. His research interests are mainly on social networks and data mining.



Dr. Bhaskar Biswas is Assistant Professor at Department of Computer Engineering, Indian Institute of Technology (BHU), Varanasi, India. Dr. Bhaskar had done his Ph.D from the Same Institute in Web Mining. His research interest includes Data mining, Social Networks Analysis, and Web Mining.

