Cyberlearners and Learning Resources

Leyla Zhuhadar  
Dept. of Computer Engineering and Computer Science  
Knowledge Discovery and Web Mining Lab  
University of Louisville, USA  
leyla.zhuhadar@wku.edu

Rong Yang  
Dept. of Mathematics and Computer Science  
Western Kentucky University, USA  
rong.yang@wku.edu

ABSTRACT

The discovery of community structure in real world networks has transformed the way we explore large systems. We propose a visual method to extract communities of cyberlearners in a large interconnected network consisting of cyberlearners and learning resources. The method used is a heuristic and is based on visual clustering and a modularity measure. Each cluster of users is considered as a subset of the community of learners sharing a similar domain of interest. Accordingly, a recommender system is proposed to predict and recommend learning resources to cyberlearners within the same community. Experiments on real, dynamic data reveal the structure of community in the network. Our approach used the optimal discovered structure based on the modularity value to design a recommender system.

Keywords


1. INTRODUCTION

Nearly 20 years ago, it seemed that the Worldwide Web (WWW) played a considerable role in facilitating the way people share information. Today, it is obvious that the Web is not only about sharing information, but it is a place where people create, share, interact and learn. If we view the Web purely, like a natural phenomenon, and if we study, say, the behaviors of learners using the social media networks to learn how people learn, this could answer fundamental questions about human behaviors and the impact of the Internet on the social process.

Using analytics to discover hidden patterns in big data

2. BACKGROUND AND RELATED WORK

The term Learning Analytics occurred the first time in Berk [3], but it was more related to business intelligence as stated in Shum et al.,[22]. Noticeably, over the last two years, more research related to learning analytics occurred. The following summary does not exhaust the vast amount of research done in this area, but it gives a short summary. For example, Gašević et al.[11] experimented with LOCO-Analyst on two master’s level courses for learning analytics with feedback; their goal was to discover the semantic knowledge based on community and accordingly update the learning process of this community. Johnson et al.[14] mentioned some learning analytics tools that have emerged in higher education: i) Northern Arizona University uses an academic early alert and retention system to improve students’ retention and success; ii) Purdue University utilizes Signals, an analytical data mining tool that identify students who need help; iii) Ball State University designed a visualized collaborative writing system to help better evaluate student performance; iv) the University of Wollongong in Australia uses SNAP[3] to visualize data from discussion boards to find patterns in students’ behaviors. The last example is the closest to our system. The similarity between both systems is twofold: goal and method. The goal is to collect data about students to better understand their behaviors and accordingly provide a better understanding on how to help these students. There is a major difference, however between our approach and theirs; in our system this help comes as a recommendation based on similar a community of learners. We consider the similarity in both context and in communities in social networks; whereas, SNAP is based only on context. Both systems use visualization as a method to discover patterns. As we know, our cognitive
system is not able to deal with vast amounts of information [6]. The importance of communities in social networks has long been recognized. In 1999, Kleinberg et al. [16] used the concept of a bipartite core to identify Web communities. Flake et al. [8] introduced one of the most attractive definitions for a community, both because of its intuitive appeal and its computational simplicity. There, a community is defined as a set of web pages in which each member page has more links (in either direction) inside the community than outside. The exact proportion of inside to outside links can be varied as required. Girvan and Newman [12, 19] devised a method for community determination based on betweenness centrality by generalizing it to edges and finding communities by deleting edges from the network in order of decreasing betweenness (and recalculating the betweenness between deletions). Modularity, introduced in Newman [6], of a community decomposition of a network. This measure was turned into a fast and effective community identification mechanism by using a greedy algorithm to approximately optimize the modularity values. Even this fast algorithm was improved in Clauset et al. [5]. For a recent and much more comprehensive survey see Fortunato [9].

3. METHODOLOGY

3.1 Data Collection

The data set used in this study is part of HyperMany-Media's (HMM) Logfile. We extracted and used only the last six months from the Logfile. In our previous research Zhuhadar et al. [23], we found that i) the profiles of our users are evolving (users’ interests change over time; i.e., a user might register for courses in chemistry, but after three months, the same users switch to courses in biology) and ii) our platform is an evolving domain (new courses are added to the platform each semester). Accordingly, we provide recommendations based on this dynamic change of students’ interests. Also, we argue that building a dynamic recommender system based on a social network needs to be scalable to accommodate current and new users. If we considered using the whole Logfile which consists of activities of 750,000 users so far, the time needed to extract recommendations from the best candidates in the Logfile, on the fly, would be impractical. The data set used in this research consists of users’ logs during the following period (2/1/2011-8/1/2011). Each entry has the following fields: user name, visited resource, number of visits. The number of visits is used as a Weighted Degree in the graph. The more the user (learner) visited a learning resource, the closer the learner is considered to the hub (learning resource); therefore, users who are close to hubs are considered as authorities in that specific domain. Our assumption is built upon the theory of reinforced learning. A very old concept that was introduced in 1913, by Ebbinghaus, in [7]. Ebbinghaus found that if learners are introduced to a problem over many trials, an exponential learning curve is produced. Finally, we visualized our Logfile using a graph analysis tool called Gephi [2].

3.2 Evaluation Method (Categorization Criteria and Determining the Energy Levels)

As we discussed in section 2, there is a variety of measures and methods for finding communities in social networks. In this research, a modified version of the modularity measure proposed by Blondel et al. [4] is utilized to compare the quality of clusters (Equation 1) for measuring the success of a community decomposition of a network. This measure is considered fast and effective to identify communities by using a greedy algorithm to approximately optimize the modularity values.

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j),$$  

where $A_{ij}$ represents the weight of the edge between $i$ and $j$, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex $i$, $c_i$ is the community to which vertex $i$ is assigned, the $\delta$-function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise. This measure is considered fast and effective to identify communities by using a greedy algorithm to approximately optimize the modularity values.

3.3 Detecting Communities

In this section, we identify the methods used to detect communities of similar users from our extracted Logfile. First, we define a set of various force laws to recognize communities in the network structure. Once the communities have been recognized, we ensure that each community has its own energy state which determines the relevance level of that particular community within its range of proximity. We used force directed methods to discover the similarity between users. Three types of methods were used: (i) the Yifan Hu Algorithm [13], (ii) the Fruchterman and Reingold Algorithm [10], and finally, the Force Atlas 2 Algorithm [18].

4. EXPERIMENTAL ANALYSIS

We deployed three algorithms (i) Yi Fan Hu Algorithm, (ii) Fruchterman and Reingold Algorithm, and (iii) Force Atlas 2 Algorithm on HyperManyMedia’s Logfile extracted during the period of (2/1/2011-8/1/2011). After filtering out some data based on the conditions (user’s visits>= 10 & length of accumulated sessions>= 30 minutes) and deleting outliers, our network consisted of 8,510 Nodes (# of users) and 23,079 Edges (# of edges between users and learning resources). Each edge connects a user (learner) to a learning resource. In this small portion of the Logfile, the number of learning resources is ~ 10,000 learning objects.

First, we noticed that the Yifan Hu Algorithm [10] proved to be efficient. It seems to overcome the localized nature of the Kernighan-Lin Algorithm [15] and also the local minima of the Fruchterman and Reingold’s Algorithm [10]. We also deployed the Force Atlas 2 Algorithm, calculated the modularity for each force directed method, and visualized the network.

4.1 Force Directed Method (Fruchterman Reingold Algorithm)

We used three parameters in the Fruchterman Reingold's [10] force directed method: (i) Area (which defines the number of nodes in the graph); (ii) Gravity (it works to attract all nodes to the center to avoid dispersion of disconnected components); and (iii) Speed (convergence speed). We ran 20 trials and the best results obtained in trial 6 had a modularity of 0.606 and number of communities (clusters) of 14. Accordingly, we present the social network structure in Figure 1.
4.2 Force Directed Method (Force Atlas 2 Algorithm)

The Force Atlas 2 Algorithm [18] uses a classic force-vector, similar to the Fruchterman Rheingold. This algorithm benefits from Barnes-Hut optimization techniques [1] and its own repulsive and tolerance levels [1]. We ran 20 trials and the best results obtained in trial 18 had a modularity of 0.610 and number of communities (clusters) of 14. Accordingly, we present the social network structure in Figure 2. We noticed that we got the best results in this method when there is (i) a little repulsive force given by Scaling and (ii) a higher attractive force given by Gravity.

4.3 Force Directed Method (Yi Fan Hu Algorithm)

The Yi Fan Hu Algorithm [13] overcomes local minima by using Barnes and Hut’s [1] octree technique which approximates the short-and-long range force efficiently. It uses the adaptive cooling schemes and general repulsive force models to develop the set of forces to be applied on the data set for formation of the communities. We ran 20 trials and the best results obtained in trial 12, had a modularity of 0.607 and the number of communities (clusters) of 15. Accordingly, we present the social network structure in Figure 3.

4.4 Discovering the Best Communities Structure in HMM’s Logfile Social Network

To summarize, by running multi-trials, we discovered the best combinations of parameters for each algorithm, as shown in Table 1. However, we noticed a slight difference in the results that could be inferred as follows: We found more clusters using the Yi fan Hu Algorithm; whereas, we obtained a better modularity measure using the Force Atlas 2 algorithm. Therefore, we decided to use the Force Atlas 2 algorithm for clustering. Figure 2 presents the notion of detecting communities of users in the Social Web.

4.5 Designing a Social Recommender System

We considered that providing recommendations to a learner based on similarity metrics between the users and him/her and extracted from the social network would answer this question. We propose adding a social recommender system to HMM repository where recommendations provided to a user (learner) is based on detecting triangles in the community [21], refer to Figure .

Table 1: Evaluation of the three Force-directed methods

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Modularity</th>
<th># of Communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruchterman and Rheingold</td>
<td>0.606</td>
<td>14</td>
</tr>
<tr>
<td>Force Atlas 2</td>
<td>0.610</td>
<td>14</td>
</tr>
<tr>
<td>Yi Fan Hu</td>
<td>0.607</td>
<td>15</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

We consider our research is different than any other research that has been done on detecting community using graph-based methods for the following reasons: i) our research is an applied study on a real platform visited by thousands of users on a daily basis; ii) we used data collected from HyperManyMedia’s Logfile to discover communities in social networks; iii) finally, we proposed the triads concept for recommendations, keeping in mind that our current experiments are based on triangles of nodes. In our future work, we plan to experiment and compare the results based on learners’ feedback. In addition, we plan to complete our evaluation of the visual recommender system, using objective metrics as well as user testing.

6. REFERENCES

[22] Shum, S. B., and Ferguson, R. Social Learning Analytics.