Development of Valuable Users Identification model in Online Education Communities Based on Optimal Removal Strategy

Yu Bai[1,*], Gang Yin[2], Huaimin Wang[3] ,Ye Deng[4]

[1]College of Computer, National University of Defense Technology, Changsha, 410073 China

[4]College of Information Systems and Management, National University of Defense Technology, Changsha, 410073 China

Keywords: Valuable users, Online education communities, Removal strategy, Tabu search

Abstract. The valuable users in online education communities identification model has broad applications and recently has received growing attention. This paper first presents an optimized removal strategy model for online education communities, and the network performance is quantitatively measured by the size of the largest connected component and " R ". Here, we introduce the tabu search into the optimal removal strategy problem to identify the valuable users. The efficiency of the proposed solution was verified by comparing it with other removal strategies used in real-world networks. Numerical experiments suggest that our solution can find the most valuable users for node failure removals, and it can be identified through global searches. Our understanding of the optimal removal strategy may also shed light on an incentive mechanism for active users and deserves additional study.

Introduction

One recent trend in education is the application of social computing and Web 2.0, as evidenced by several initiatives in the U.S. and Europe[1-3]. Therefore, the application of online educational community has been popular in recent years by integrating Web 2.0 activities into the classroom. And thanks to these platforms, the past can only be done in the classroom can now be completed online, what is more is that within educational community the use of computers has mostly been focused on enhancing learning in formal settings[4]. Nowadays, the concept of online community is constantly innovating, and it is interesting to notice that the platform not only can be used to share the knowledge and opinions[5], but also can evaluate students' learning ability and discover which concept dose student has not well mastered. The data collected from the case[6]shows that most of the students who took part in the educational community feel that the approach of study online is exciting and rewarding.

Based on Web 2.0, online educational community's users can be active contributors rather than just passive observers. In addition, the contribution of intelligence can be used for both instructors and students to create education materials that can have great value. As more data have been generated, the intelligence will become more accurate, and it attracting a lot of researchers to analyze.

At present, most of the researches focus on the analysis of the behavior of students and the evaluation of learning ability, and the problem of identifying valuable user in online education communities is not obtained more attention, although some related works have been devoted to the study of identifying valuable user in social networking sites. For example, the series presents approaches of measuring node importance in social network analysis (SNA)[7], and some studies focus on explore the spread of influence in social networks[8]. Many enterprises have begun to identify certain users of SNS to conduct online marketing and reputation management[9]. However, this study emphasized the importance of each node, they assume there existed the influence relationships among users in order to identify the important user. Moreover, some works have recently focused on measuring the central degree, defined as the number of links incident upon a node which people who have more ties are more likely to affect other users[10]. It draws lessons from the theory of Page Rank which is a measure originally designed to classify web pages[11].

However, as far as we know, most studies could not identify valuable user in online education. Thus, we try efforts to solve the problem through the theory of complex network. In fact, many systems in the real world can be described as complex networks [12-14]. Examples include the Internet, metabolic networks, electric power grids, supply chains, urban road networks, the world trade web, and online community among many others. All over the online education communities, databases are collecting data concerning user's communication. This information is usually transformed into network structures in which the nodes represent the individuals in the data set and the links possible communication between these individuals. In this work, we will address the problem of identifying valuable user in online education communities by tabu search.

The rest of this paper is organized as follows. In Sec. 2, we first present optimization model for identifying valuable user in online education communities. In Sec. 3, we propose optimal removal strategy based on tabu search. The experiments on are shown in Sec. 4. Finally, the conclusions and discussions are presented in Sec. 5.

Optimization Model for Identifying Valuable User in Online Education Communities

Consider online education communities formalized in terms of a simple undirected graph $G(V,E)$, where V is the set of nodes, and $E \subseteq V \times V$ is the set of edges. Let $N = |V|$ and $W = |E|$ be the number of nodes and the number of edges, respectively. Let $A(G) = (a_{ij})_{N \times N}$ be the adjacency matrix of G, where $a_{ij} = a_{ji} = 1$ if nodes v_i and v_j are adjacent.

We only consider node removal approaches in this study and assume that the removed edges are removed if one node is removed, i.e., if node v_i is removed, then

$$A(i,:) \leftarrow [\ ] \ and \ A(:,i) \leftarrow [\ ]. \qquad (1)$$

Denote by $\hat{V} \subseteq V$ the set of nodes that are removed and denote by $n = |\hat{V}|$ the removal strength. Denote by $C = (x_1, x_2, \ldots x_N)$ a removal strategy, where $x_i = 0$ if $v_i \in \hat{V}$, otherwise $x_i = 1$. Thus we obtain $n = \sum_{i=1}^{N} x_i$. Our goal is to find the optimal removal strategy C^*, which maximizes the removal effect $\Phi(C)$. The most valuable users must be the corresponding nodes which maximize the removal effect $\Phi(C)$ after their removal.

There are various alternative measures of removal effect. We first consider the critical removal fraction of nodes that is required to characterize the network damage due to a removal strategy. This measure emerged from random graph theory [15] and was stimulated by Albert et al. [16]. Instead of a strict extreme property, the measure statistically considers how the removal of nodes leads to a deterioration of network performance and eventually to the disintegration of the network at a given critical removal fraction f_c. The disintegration of network performance is measured in terms of the size of the giant component, where the lower the value of f_c, the more destructive the removal strategy.

Furthermore, because the measure of network damage f_c is the critical fraction of removals at which the network completely collapses, it ignores situations in which the network suffers large damage without completely collapsing. To demonstrate the impact of a removal strategy on the networks in depth, we also consider the measure of network damage R, which is defined as [17, 18]

$$R = \frac{1}{N+1}\sum_{Q=0}^{N} S(Q), \quad (2)$$

where $S(Q)$ is the fraction of nodes in the giant component after removing $Q = Nf$ nodes. The normalization factor $1/(N+1)$ ensures that the network damage with different sizes can be compared. The measure of network damage R, which corresponds to the integral of the curves $S(Q)$, not only measures after how many removals are required for the network to collapse but also considers the size of the giant component for each number of removed nodes. The range of possible R values is between $1/(N+1)$ and 0.5, where $R = 0$ corresponds to an empty network of isolated nodes, and $R = 0.5$ corresponds a fully connected network. In particular, lower 'R' values correspond to more destructive removal strategies.

Consequently, the optimization model of identifying valuable user in online education communities can be described as follows:

$$\begin{aligned} &\max\ \Phi(C = (x_1, x_2, \ldots x_N)) \\ &s.t. \begin{cases} \sum_{i=1}^{N} x_i = n \\ x_i = 0 \ or \ 1 \end{cases} \end{aligned} \quad (3)$$

Optimal Removal Strategy Based on Tabu Search

Consider a simple example firstly. Let G be an undirected graph with 100 vertices and removal strength of 10. Although the network contains few nodes, the total number of removal strategy combinations is astronomically large, corresponding to $C_{100}^{10} = \frac{100!}{10! \times 90!} \approx 1.73 \times 10^{13}$. Therefore, we consider utilizing an intelligent optimization algorithm for the combinatorial optimization problem.

In most of cases, tabu search is regarded as an efficient tool for solving global optimization problems. Tabu searches use a local search or neighborhood search procedure to iteratively move from one potential solution x to an improved solution x' in the neighborhood of x, until some a stop criterion has been satisfied. Local search procedures often stuck in poor-scoring areas or areas where scores plateau. To avoid these pitfalls and explore regions of the search space that would be left unexplored by other local search procedures, tabu search carefully explores the neighborhood of each solution as the search progresses. The solutions admitted to the new neighborhood, $N^*(x)$, are determined through the use of memory structures. Using memory structures, the search progresses by iteratively moving from the current solution x to an improved solution x' in $N^*(x)$. These memory structures form what is known as a "tabu list", a set of rules and banned solutions that are used to filter the solutions admitted to the neighborhood $N^*(x)$ to be explored by the search [19, 20]. Our implementation is based on a tabu search and coding of nodes. In the following section, we provide a detailed description of the optimal removal strategy algorithm.

In the following section, we introduce the optimal removal strategy. For each iteration of OAS, the neighborhood of the current solution is obtained by swap the state of two nodes randomly.

Each step of the OAS is described below:

Procedure 1:（initialization of the algorithm） Set length L according to the tabu list, define the scale of the candidate solutions $n_{candidate}$ and the maximum iteration steps.

Procedure 2: Generate the initial solution C_0: Generate a vector $C = (\overbrace{1,\ 1,\ 1,\ ......1,\ 1}^{N})$, which means that every node is coded to be an element x_i of the vector and set at a value of 1. Then choose n nodes to randomly remove to obtain the initial solution C_0. Therefore, let the present optimal solution be $C^* = C_0$ and the present solution $C_{now} = C_0$.Then utilize decoding method to transform C_0 to G_0 so as to calculate the $S(Q)_{now}$ of the present solution and set $S(Q)_{best} = S(Q)_{now}$.

Procedure 3: Determine whether it satisfies the termination condition. If it satisfied, output the result. If not satisfied, continue to the next step.

Procedure 4: Generate candidate solutions: Generate $n_{candidate}$ new solutions by swapping the state of two nodes randomly and utilize decoding method to calculating their $S(Q)$ (the fraction of nodes in the giant component).

Procedure 5: If the candidate solution with a minimum $S(Q)$ has not been tabooed, then set the candidate solution to C_{now}. Or it satisfies the aspiration criterion that tabu can be disregarded if the better solution is found. In detail, the aspiration criterion means that if a swap is tabooed, however, the corresponding $S(Q)$ is better than $S(Q)_{now}$. Therefore, the swap satisfies the aspiration criterion and we will condone the swap. Then let the condoned solution be C_{now}. Otherwise, if the candidate solution which contains swap (move) on the tabu list does not satisfy the aspiration criterion, we would choose the candidate solution with sub-optimal $S(Q)$ which has not been tabooed to be C_{now} and add the 'best' swap about the two corresponding nodes into the tabu list. If $S(Q)_{now} < S(Q)_{C^*}$, then let $C^* = C_{now}$. Next, if the tabu list is full, element will be allowed to expire, which follows the same order they are added.

Procedure 6: Turn to step 3.

The algorithm diagram of the OAS is presented in Fig. 1.

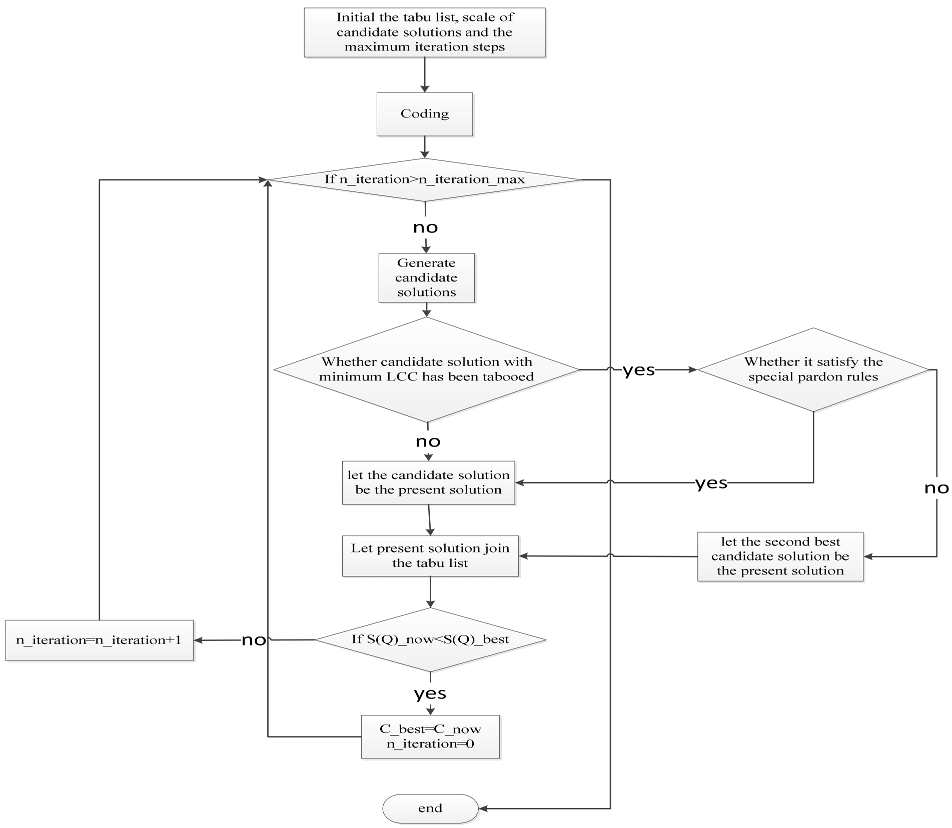

Fig. 1.The algorithm diagram of the OAS.

Experiments

To explore the efficiency of the OAS on the effect in depth, we performed it using our solution in real online education community named Trustie (www.trustie.net). The nodes were removed according to our optimal removal strategy, and we used f_c and R as measures of network damage. The technique for computing the R is just as equation 3. As for the critical threshold f_c, we choose $\kappa \equiv < k^2 > / < k > < 2$ as the criterion for the disintegration of networks. After each node is removed, we calculate κ. When κ becomes less than 2, we record the number of nodes t removed up to that point. The threshold f_c is defined as $f_c = < t > / N$. We show the relative sizes of giant component S as a function of f.

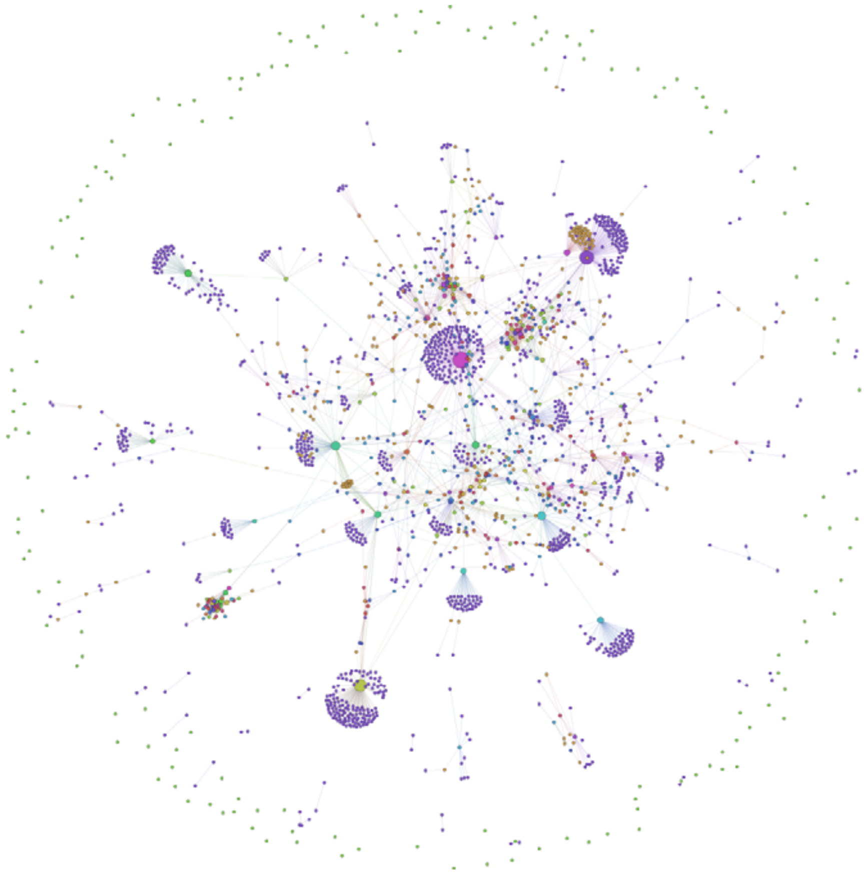

Fig. 1. The real online education communities derived from the Trustie.

As shown in fig.1, the data corresponding to the real online education communities was derived from the Trustie, which includes posters and their replied users. It contains 3687 nodes and exhibits a power-law distribution. We test the efficiency of the removal strategy on such network. Besides OAS, the most traditional strategy is to select the vertices based on descending order of degrees in the initial network and then to remove the vertices one by one beginning with the vertex with the highest degree. As Holme summarized, this strategy is known as the 'ID' attack strategy, which is simply a local attack strategy. Researchers have also used the initial distribution of the betweenness to in global strategies, which we refer to as 'IB'. As more vertices are removed, the network structure changes, which leads to different distributions of the degree and the betweenness compared to the initial structure. The third attack strategy, known as ''RD removal'', uses the recalculated degree distribution at every removal step, and the fourth strategy, known as ''RB removal,'' is based on the recalculated betweenness at every step.

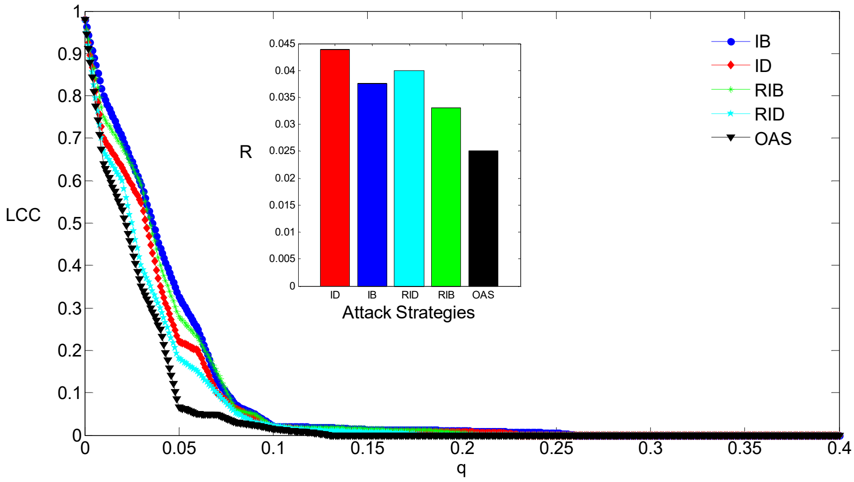

Fig. 2.Value of LCC, f_c and R of the OAS removal strategy and other removal strategies versus the fraction q of nodes removed in the real online education communities The points are plotted every node removed for the model networks.

As seen in Fig. 2, the OAS is much more efficiency than the other removal strategies because the 'R' for the OAS is invariably the lowest. In addition, it is significantly that our solution can find the most valuable users for node failure removals and its correspondingrankingis shown in table 1 through global searches.

Table 1. The ranking of valuable users

Ranking	Name
1	Haifang Zhou
2	TulongZuoshou
3	Jingtao Tang
4	Juan Chen
5	Gang Yin
6	Letong Feng
7	Ting Wang
8	Dou Dou
9	Yangbin Tang
10	Wei Zheng

Summary

In this paper, we first presents an optimized removal strategy model for online education communities, and the network performance is quantitatively measured by the size of the largest connected component and " R ". Regardless as to whether calculation or recalculate method are used, such as in degree-based and betweenness-based attacks, our solution yielded a greater degree of efficiency than other attack strategies. As presented here, we can approximate the 'best' choice for nodes failure removal through a global search.

We believe that our understanding of the optimal removal strategy may also shed light on an incentive mechanism for active users and deserves additional study.

Acknowledgement

The research is supported by the National Natural Science Foundation of China (Grant No.61432020, 61472430, 61502512, 61432020).

References

[1]. Association CR. Cyberinfrastructure for education and learning for the future: A vision and research agenda [M]. Computing Research Association, 2005

[2]. Borgman CL. Fostering learning in the networked world: The cyberlearning opportunity and challenge [M]. DIANE Publishing, 2011

[3]. Redecker C. Review of learning 2.0 practices: study on the impact of web 2.0 innovations of education and training in Europe [J]. 2009,

[4]. Mifsud L. Wireless and Mobile Technologies in Education, 2002 Proceedings IEEE International Workshop on [C]. IEEE, 2002.

[5]. Korica P, Maurer H, Schinagl W. THE GROWING IMPORTANCE OF E-COMMUNITIES ON THE WEB1 [J]. 2006,

[6]. Tsai W-T, Li W, Elston J, Chen Y. Collaborative learning using wiki web sites for computer science undergraduate education: A case study [J]. IEEE Transactions on Education, 2011, 54 (1):114-24

[7]. Wasserman S, Faust K. Social network analysis: Methods and applications [M]. Cambridge university press, 1994

[8]. Kempe D, Kleinberg J, Tardos É. International Colloquium on Automata, Languages, and Programming [C]. Springer, 2005.

[9]. Jones B, Temperley J, Lima A. Corporate reputation in the era of Web 2.0: the case of Primark [J]. Journal of Marketing Management, 2009, 25 (9-10):927-39

[10]. Masuda N, Kurahashi I, Onari H. Suicide ideation of individuals in online social networks [J]. PloS one, 2013, 8 (4):e62262

[11]. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web [J]. 1999,

[12]. Albert R, Barabási AL. Statistical mechanics of complex networks [J]. Rev Mod Phys, 2002, 74 (1):47-51

[13]. Newman MEJ. The structure and function of complex networks [J]. SIAM Rev, 2003, 45 (2):167-256

[14]. Wang XF. Complex networks: Topology, dynamics and synchronization [J]. Int J Bifurcation &Chaos, 2002, 12 (5):885-916

[15]. Bollobás B. Random Graphs [M]. New York: Academic Press, 1985

[16]. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks [J]. Nature, 2000, 406 (6794):378-82

[17]. Herrmann HJ, Schneider CM, Moreira AA, Andrade JS, Havlin S. Onion-like network topology enhances robustness against malicious attacks [J]. J Stat Mech, 2011, P01027

[18]. Schneider CM, Moreira AA, Andrade JS, Havlin S, Herrmann HJ. Mitigation of malicious attacks on networks [J]. Proc Natl Acad Sci U S A, 2011, 108 (10):3838-41

[19]. Glover F. Tabu search-part I [J]. ORSA Journal on computing, 1989, 1 (3):190-206

[20]. Glover F. Tabu search—part II [J]. ORSA Journal on computing, 1990, 2 (1):4-32

