AN OPTIMIZATION APPROACH FOR FORMING A TEAM OF EXPERTS IN SOCIAL NETWORKS

ALI BAGHERINIA[1] AND ELHAM AMINI[2]

[1]Department of Computer Engineering, Dehdasht Branch , Islamic Azad University, Dehdasht, Iran

ABSTRACT

It is important to organize an effective team of experts in social networks to cover the allocated task in many real-world projects. The cost of communication, and the good cooperation among the team members at the same time efficient, will lead us to make the project more successful. Then we propose an approach for team formation in social networks. Since this problem is NP-hard. We proposed a CSO based algorithm to solve this problem. Extensive experiments on real datasets shows the effectiveness of the proposed algorithm.

KEYWORDS

social network, team formation, expert team

1. INTRODUCTION

An expert network contains a group of professionals who a gathering of experts who can give specific information and service. Examples of social networks among professionals include Researchgate (www.researchgate.net), LinkedIn (www.linkedin.com), Xing (www.xing.com) and others. With the boundless utilization of the Web, online expert networks have turned out to be well known where an ever increasing number of organizations look for topic specialists to finish an assignment or venture. In such network, an expert is portrayed by their subject matters, training foundation, area, and so on. Also, a specialist can indicate his counselling rate.

Problem definition :

we study the he issue of finding a gathering of people who can work as a team to complete a particular task. We expect that there exists a pool of n person competitors P P= {1 . . . n}, where each candidate i has a set of skills P_i. Then we assume that these candidates are organized in a weighted and undirected social graph G(P,R). The weights on the edges of G should be interpreted as: a low-weight edge between nodes i, j implies that candidate i and j can collaborate more easily than candidates connected with a high-weight edge. These weights can be discovered in different ways in different application domains. For example, in a company, the weight between two employees may correspond to the length of the path from one employee to another through the organizational chart. In a scientific research community, the weight between two scientists in a country is related to the physical distance between two cities. Interpersonal relationships among individuals can also be used to calculate the weights. As formal, for given a task T that requires a set of skills. We must to find a set of individuals $P' \subseteq P$, such that every required skill in T is exhibited by at least one individual in P'. Additionally, the members of team P' should define a subgraph in G with low communication cost. The communication cost measures how effectively the team members can collaborate: the lower the communication cost, the better the quality of the team.

Example: a software project administrator wants to build a team of experts skilled in the T={C# programming, software designing, DBMS administrator, PHP programming}. Also assume there are five candidates, {c1, c2, c3, c4, c5}, with the following backgrounds: EXc1={C# programming}, EXc2= {PHP programming}, EXc3= {software designing, DBMS administrator}, EXc4= {software engineering} and EXc5={software designing, DBMS administrator, PHP programming}. The relationships among these candidates are represented by the social network shown in Figure 1, where the existence of an edge between two nodes in G indicates that the corresponding persons can collaborate effectively.

Without considering how effectively these people can collaborate, the manager can select either P0 = {c1, c2, c3} or p' = {c1, c4}, since both these teams have the required skillset.

However, the existence of graph G makes P0 = {c1, c2, c3} a superior solution, since the structure of G indicates that c1 and c4 cannot work together at all.

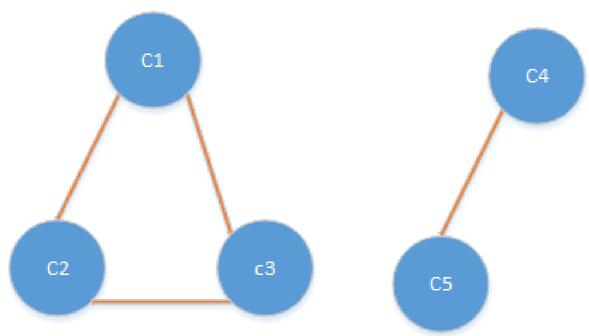

Figure 1. Connections between five individual candidates

2. Related Work

The need of viable cooperation among people in a team has been considered previously. Fitzpatrick and Askin [1] utilize the Kolbe Conative List (KCI) to gauge people's drive and disposition, which thus reflects the nature of the group. Chen and Lin [2] utilize the Myers Briggs test to quantify the applicants' identity and evaluate their relational connections as colleagues. Despite the fact that these methodologies are fascinating from the anthropological/mental perspective, they likewise overlook the existing diagram structure among people. Hence, these methodologies ought to be viewed as correlative to our own.

The flow of gathering development forms and their impact on the arrangement of groups in systems have been as of late tended to in [3]. The amusement theoretic parts of the same issue have been examined in [4]. These examinations are corresponding to our own and for the most part concentrate on giving valuable bits of knowledge about social procedures.Gaston et al. proposed a network structure among experts persons in a workforce pool [5].They provided an experimental study on how the performance of a team is affected by different graph structures among the expert persons. Although related, this work does not deal the complexity time of forming a team of experts in a given network. Some work in [6], [7] has also been devoted to the construction of the social network.

3. PROPOSED APPROACH

We assume a pool of n person competitors P = {1, . . . , n}, EX = {ex1, . . . , exm} to be a universe of m skills. Each person i is associated with a set of skills Pi ⊆ P. If exj∈ Pi we say that person i has skill exj. We often use the set of skills a person to refer to him. Also, we say that a subset of individuals P'⊆ P possesses skill exj if there exists at least one individual in P' that has exj .

A task T is simply a subset of skills required to perform a job. That is, T ⊆ A. If exj∈T we say that skill exj is required by task T. We can also define the cover of a set of persons P' with respect to task T, denoted by Cover (P', T), to be the set of skills that are required by T and for which there exists at least one individual in P' that has them. We define support set of ex denoted by Sup (ex), to be the set of persons in P that has ex skill; Sup (ex) = {i | i ∈P and ex∈ Pi }.

For every two nodes i, j ∈P we define the distance function d(i, j) to be the weight of the shortest path between i and j in G. Note that this distance function between the nodes is a metric and thus satisfies the triangle inequality. For every pair of nodes we also use Path(i, j) to represent the set of nodes that are along the shortest path from i to j. Apart from computing the distance between two nodes in G, we will often need the distance between a node i ∈ P and a set of nodes P' ⊆P. We define this to be $d(i,P') = \min_{j \in P'} d(i,j)$. In this case, we use Path(i, P') to represent the set of nodes that are along the shortest path from i to the node $k = \arg\min_{j \in P'} d(i,j)$. In Finally, given graph G and P'⊆ P, we use G(P') to denote the subgraph of G that contains only the nodes in P'. Then the problem is defined as finding P' so that P'⊆ P and Cover(P', T)=T and communication cost of P' edge (CC(P')) is minimized; we consider diameter communication as communication cost, that the diameter of a graph is the largest shortest path between any two nodes in the graph. Then this is a optimization problem that

$$\underset{P' \subseteq P, Cover(P',T)=T}{\text{minimize}} \ \alpha * CC(P') + (1-\alpha) * |P'| \ where \ \alpha \in [0,1], \ |P'| \ is \ the \ cardinality \ of \ P' \quad . (1)$$

In this paper we solve this problem by CSP (Cat Swarm Optimization).

3.1. CSO algorithm

Cat swarm optimization is an optimization heuristic algorithm based on social behaviour of cats. CSO is proposed by observing two behaviours of cats, i.e. seeking and tracing. So CSO consists of two modes: seeking mode and tracing mode. In seeking mode, cats are not moving anywhere. They only seek for next best position. On the other hand, in tracing mode they move to the next best position with high speed. This means that not all cats are moving at the same time in solution space. The seeking mode helps them to find next best positions quickly, that is why the solution is obtained in less number of iterations [8]. CSO algorithm is shown in the Figure 2. Using these steps causes cats eliminate unnecessary use of energy, leading to efficient convergence towards solution

Step 1. Initialize the swarm with N cats in D dimensions. Step 2. Assign random values as the velocity V_{id} to i^{th} cat which is in current position X_{id}, where *d* represents dimension index. Step 3. Randomly assign *PT* percentage of cats from the population to tracing mode and the rest to seeking mode, where *PT* represents the mixture ratio of cats in tracing mode to seeking mode. Step 4. Assess the fitness value of each cat and store the location of the best cat in memory. Step 5. Update position of cats as per the mode they are in. Step 6. Check stopping criteria, if true then exit, else goto step-3

Figure 2. general CSO algorithm

3.2. Team formation by CSO

In this paper we aim to efficiently form a team in social networks, focusing on minimizing the communication cost among persons that cover tasks in set T. The proposed algorithm uses an initial population of C cats among which some are in seeking mode while others are in tracing mode, according to PT. Each cat represents a team in social network graph, which is updated as per the mode that the cat is in. Assessing the fitness value (equation (1)) of the cats leads to finding the team having minimum cost. In each iteration, a new set of cats is chosen to be in tracing mode. The final solution, represented by the best position among the cats, gives the team that has the minimum communication cost among all team person (P') that cover all experts in T. Figure 3 shows general steps for algorithm.

1) Seeking mode:

This represents the majority of cats that search the global space while being in a resting state by intelligent position updating. Here the algorithm uses two basic factors SMP and CDC. SMP (seeking memory pool) represents the number of copies to be made for each cat. CDC (count of dimension to change) defines how many of the allocations are to be altered in a single copy. The general steps are as follows:

Step 1. Create *j* copies of the i^{th} cat as represented by SMP. **Step 2.** Modify CDC dimension of each copy randomly. **Step 3.** Evaluate fitness of every copy. **Step 4.** Find the best solutions among all copies that is the mappings having minimum cost. **Step 5.** Randomly choose a solution among them and replace it for the i^{th} cat.

Figure 3. Seeking mode of CSO algorithm

2) Tracing mode:

This mode shows the cats that are in a fast moving mode and search the local space by moving towards the next best position with high energy. The general steps are shown in Figure 4.

Step 1. Find velocity v_i^{t+1} for the i^{th} cat as: $v_i^{t+1} = w * v_i^t + r1 * c1 * (x_{best} - x_i^t)$ (2) *where w is the inertia weight, r1 is a random number such that* $0 \leq r1 \leq 1$ *and c1 is the acceleration constant.* v_i^t *is the previous velocity,* $x_{best}t$ *is the best location and* x_i^t *t is the current location.* **Step 2.** Update position of the cat as: $x_i^{t+1} = v_i^{t+1} + x_i^t$ (3) **Step 3.** Check if the position goes out of the defined range. If so, assign the boundary value to the position.

Step 4. evaluate the fitness value for the cats. **Step 5.** Update the solution set with the best positions of the current iteration.

Figure 4. Tracing mode of CSO algorithm

If for some reason, such as the dismantling of a team or parallel projects, there is a need for other teams. Therefore, the teams should form that some team members can shared. To do this, fuzzy clustering is used to make k cluster.

4. EXPERIMENT RESULT

We evaluate the proposed approach for the Team Formation problem using the scientific-collaboration graph extracted from the DBLP bibliography server. We show that our approach give the fairly good results in terms of the communication cost and the cardinality of the team. Examples of teams reported by our methods illustrate the effectiveness of our framework in real scenarios.

Task generation: Every generated task is characterized by two parameters: 1) t – the number of required skills in the task; and 2) s – the diversity of the required skills in terms of their corresponding areas. We use T(t, s) to refer to a task generated for a specific configuration of these parameters. For the results we report in this section we use t∈ {2, 4, . . . , 16} and s = 1. For every (s, t) configuration we generate 100 random tasks for this configuration and report the average results obtained by the different methods.

Communication cost: Figure 5 shows the average CC costs of the solutions achieved by proposed approach and Greedy-Diameter on tasks T(t, 1) with t∈{2, 4, . . . , 20}.

Cardinality of the team: Since the size of the team often has a positive correlation with the expenses of a project, we evaluate the cardinality of the teams formed by every Team Formation algorithm. The results in Figure 6 show that the proposed approach tends to report relatively large teams, especially for large values of t.

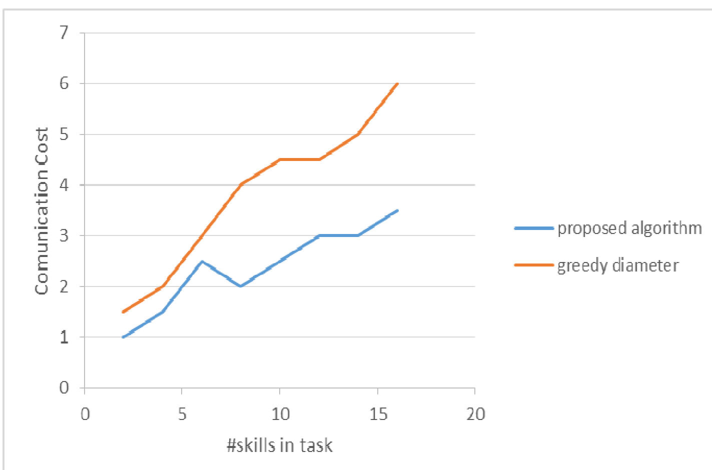

Figure 5. Average communication cost of the teams produced by each Team Formation algorithm for tasks T(t, 1) with t∈ {2, 4, . . . , 16}Of proposed and Greedy Diameter algorithms

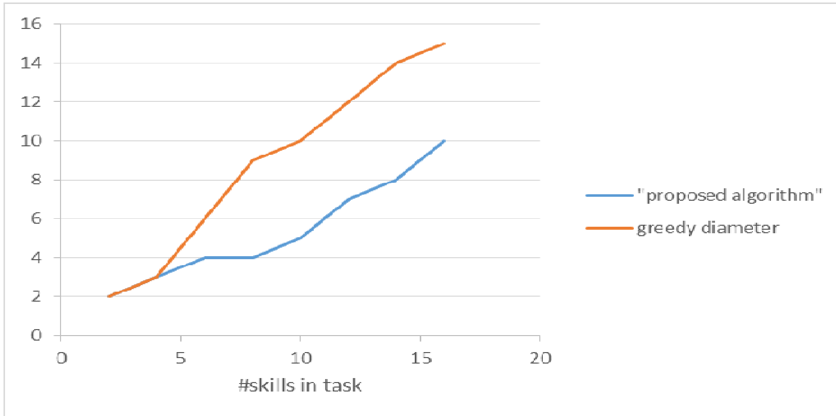

Figure 6. Average cardinality of the teams obtained by proposed and, Greedy Diameter algorithms

5. CONCLUSIONS

In this paper a swarm based method for team formation in social network is proposed. The evaluation of this method on DBLP bibliography server datasets with different scenario shows that our approach give the fairly good results in terms of the communication cost and the cardinality of the team. Examples of teams reported by our methods illustrate the effectiveness of our framework in real scenarios.

ACKNOWLEDGEMENTS

I would like to thank research deputy of Islamic Azad university, Dehdasht branch for supporting.

REFERENCES

[1] E. L. Fitzpatrick and R. G. Askin, "Forming effective worker teams with multi-functional skill requirements," Comput. Ind. Eng., vol. 48, no. 3, pp. 593–608, 2005.

[2] S.-J. Chen and L. Lin, "Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering," IEEE Trans. Eng. Manag., vol. 51, no. 2, pp. 111–124, 2004.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 44–54.

[4] M. Jackson, "Network Formation, The New Palgrave Dictionary of Economics and the Law." MacMillan Press, 2008.

[5] M. Gaston, J. Simmons, and M. DesJardins, "Adapting network structure for efficient team formation," in Proceedings of the AAAI 2004 fall symposium on artificial multi-agent learning, 2004.

[6] M. Cheatham and K. Cleereman, "Application of social network analysis to collaborative team formation," in Collaborative Technologies and Systems, 2006. CTS 2006. International Symposium on, 2006, pp. 306–311.

[7] H. Wi, S. Oh, J. Mun, and M. Jung, "A team formation model based on knowledge and collaboration," Expert Syst. Appl., vol. 36, no. 5, pp. 9121–9134, 2009.

[8] S.-C. Chu, P.-W. Tsai, and others, "Computational intelligence based on the behaviour of cats," Int. J. Innov. Comput. Inf. Control, vol. 3, no. 1, pp. 163–173, 2007