Computers & Operations Research 74 (2016) 127–142

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Social structure optimization in team formation

Alireza Farasat*, Alexander G. Nikolaev

Department of Industrial and Systems Engineering, University at Buffalo (SUNY), Buffalo, NY, USA

A R T I C L E I N F O

Article history:
Received 23 April 2015
Received in revised form
24 April 2016
Accepted 26 April 2016
Available online 28 April 2016

Keywords:
Team formation problem
Social network analysis
Combinatorial optimization
Discrete optimization applications
Lin–Kernighan heuristic

A B S T R A C T

This paper presents a mathematical framework for treating the Team Formation Problem explicitly incorporating Social Structure (TFP-SS), the formulation of which relies on modern social network analysis theories and metrics. While recent research qualitatively establishes the dependence of team performance on team social structure, the presented framework introduces models that quantitatively exploit such dependence. Given a pool of individuals, the TFP-SS objective is to assign them to teams so as to achieve an optimal structure of individual attributes and social relations within the teams. The paper explores TFP-SS instances with measures based on such network structures as edges, full dyads, triplets, k-stars, etc., in undirected and directed networks. For an NP-Hard instance of TFP-SS, an integer program is presented, followed by a powerful Lin–Kernighan-TFP (LK-TFP) heuristic that performs variable-depth neighborhood search. The idea of such λ-opt sequential search was first employed by Lin and Kernighan, and refined by Helsgaun, for successfully treating large Traveling Salesman Problem instances but has seen limited use in other applications. This paper describes LK-TFP as a tree search procedure and discusses the reasons of its effectiveness. Computational results for small, medium and large TFP-SS instances are reported using LK-TFP and compared with those of an exact algorithm (CPLEX) and a Standard Genetic Algorithm (SGA). Finally, the insights generated by the presented framework and directions for future research are discussed.

1. Introduction

The success of a project as well as the productivity of a whole organization often depends on the effectiveness and efficiency of work of participating teams [3]. The challenge of assembling successful teams can be addressed by formulating a problem of grouping individuals or assigning them to (sub)sets so as to optimize some outcome-related objectives [57]. Team Formation Problem (TFP) has received attention from the operations research community over the past years [17,26,23,56]. However, despite the common understanding that the social structure between members of the same team plays an important role in the team's output, such consideration has not been explicitly taken into account in mathematical modeling, primarily due to the lack of quantitative means to do so [36,59].

This paper addresses the challenge of developing a mathematical framework for incorporating social structure measures into TFP. It identifies the means to quantify social structure by assessing the impact of each individual's local network on their work-related outcome. For example, such outcome can be the amount of goods produced, the number of errors committed (self-reported or observed), it can be some job satisfaction indicator, the frequency of conflicts at workplace, etc. Rooted in social science theories, the presented framework allows one to build models for TFP explicitly incorporating Social Structure (TFP-SS). The class of TFP-SS models sheds light on team building strategies and also advances the emerging quantitative research of social theories and team outcomes [40,15].

The presented framework elucidates the connection between work environment, social network theories and measurable team outcomes: see Fig. 1. Social network theories motivate the use of graph-based constructs, called network structures, for representing social relations: such network structures include edges, full dyads, k-stars, and (un)directed triplets, among others. Theories of social exchange, structural holes, homophily, reciprocity, transitivity and network evolution described later in Section 3 support the design of interpretable network structure measures as functions of network structures in TFP-SS (e.g., the number of transitive triplets in a given graph). The resulting models are useful for both descriptive and prescriptive purposes. Given historical work-related outcome data for differently structured teams, the researcher can quantify the impact of each theory on team performance, by estimating the weight of each respective network structure measure in a model of the outcome. Then, by adjusting team roster decision variables, the outcome can be driven in the

* Corresponding author.
E-mail addresses: afarasat@buffalo.edu (A. Farasat), anikolae@buffalo.edu (A.G. Nikolaev).

http://dx.doi.org/10.1016/j.cor.2016.04.028
0305-0548/

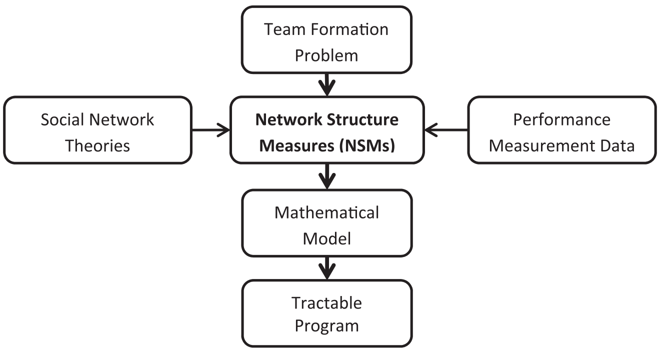

Fig. 1. The proposed framework for the TFP.

desired direction.

This optimization-permitting ability together with the reliance of the TFP-SS models on social theories distinguishes the presented framework from the existing network clustering, community detection and clique problems literature [50,44,51]. Importantly, the presented framework allows one to more closely control individual team members' local networks, which play a big role in information transmission, according to the structural holes theory. This paper also presents an extremely efficient Lin–Kernighan TFP (LK-TFP) algorithm for solving TFP-SS, based on variable depth first neighborhood search.

To summarize, this paper presents a framework for formulating and solving team formation problems that employ information provided by the local social network of individuals. This framework addresses problems at a meeting point of social science and operations research that have significant practical appeal. The efforts in building and treating models for TFP-SS lead to the creation of a methodological toolbox that quantifies social and behaviorial aspects of working in teams, particularly in professional nursing, rescue, police operations, sport teams and academic research. The paper's key contributions are three-fold.

(1) It motivates and justifies the use of mathematical programming and optimization techniques in the area of social science, where most problems have been previously qualitatively addressed by observations, experiments and basic statistical methods.
(2) It presents a prescriptive, quantitative approach to a real-world application of social network analysis, as opposed to the existing descriptive studies. It also introduces a framework to operations research, computer and social scientists for modeling more complex problems in the area of social science from an operations research perspective.
(3) It identifies the relation between established social science findings and team outcomes. It presents explicit, rigorous functions of social structures to evaluate the outcomes. It also describes how social structures and individual attributes can be incorporated into mathematical models of the outcome regardless of the network type (e.g., directed, undirected, weighted or unweighed).
(4) Designing and testing methods for solving the TFP where the optimal social structure is sought within the teams. The paper presents both an exact method and an efficient heuristic exploiting the Lin–Kernighan-inspired variable depth neighborhood search.

The rest of the paper is organized as follows. Section 2 offers a review on existing models based on Social Network Analysis (SNA) and motivates a call for prescriptive models in this field. Section 3 provides an overview of social network theories and defines relevant network structure measures that quantify social structure. Section 4 discusses the relation between work-related outcomes and social structure measures. Section 5 gives a formal statement of a special-case non-trivial instance of TFP-SS and studies this instance in greater detail. Section 6 presents LK-TFP algorithm. Section 7 reports experimental results on TFP-SS instances of varied sizes with undirected and directed networks. Section 8 concludes the paper and discusses future research directions.

2. Emerging prescriptive research in SNA

The science of SNA encompasses a set of techniques for modeling network-based systems (see Wasserman and Faust [55] for SNA motivation and position statement). These techniques range from studying centrality measures [13] to building complex probabilistic models describing network structure and formation [4,46,5]. More recently, the domain of SNA has attracted the attention of exact science professionals whose expertise allowed for advances in modeling interactions between agents [18,42,53].

The main deficiency of the existing SNA tools is that they mostly offer descriptive insights [41], rather than prescriptive capabilities. The dearth of models that could allow a decision-maker to optimally change/influence a social network structure accentuates the difficulty in handling such tasks, and at the same time, calls for filling this gap. The existing works in the area of optimization and prediction are notable [54,37,12], however, they have focused on small, highly constrained tasks as opposed to introducing broad classes of problems and general methodologies for addressing them. Of such prescriptive efforts, the models for finding subsets of influential individuals in networks are the most studied [33,28,6].

There exist models that incorporate such graph-based measures as network diameter, density, and centrality, into TFP. However, again, most of these studies are descriptive and focus on impacts of social relations, expressed by SNA measures, on team performance [10,40,1,15]. Existing prescriptive models considering a team's social network use little information captured in the social network structure. Basic SNA concepts such as closeness, diameter, and minimum spanning tree have been employed in identifying a team of experts so as to minimize intra-team communication costs [36,20,52], and in some cases, individual member costs [32,31].

In 2003, in a study of 816 organization founding teams, Ruef et al. showed that homophily and network constraints are the key factors defining team composition [48]. In a more recent study of 2349 open-source software (OSS) development teams, Hahn et al. reported positive correlations between the developers' decisions to join project teams, the collaborative ties with project initiators and the perceived status of other (non-initiator) members [29].

Zhu et al. investigated the impacts of personal and dyadic motives on team formation [60]. They used Exponential Random Graph modeling to find that individuals first get interested in a project due to personal motives such as self-interest, mutual interest, collective action and coordination cost. The typical secondary considerations include dyad-based considerations explained by the social theories of homophily, swift trust, social exchange and co-evolution.

TFPs have been formulated in multiple fields of practice and attacked by researchers in diverse disciplines. For instance, the problem of assigning students of a class to different projects can be framed as a TFP [22]. Kim et al. presented a Parallel Balanced Team Formation (PBTF) problem and employed MapReduce to solve PBTF variations considering the diversity of team members' skills including task-handling and communication abilities [34]. Agrawal et al. presented *MaxTeam* and *MaxPartition* to model problems in a similar setting [2]. An integer programming approach was