TEAM FORMATION PROBLEMS: AN INTEGER LINEAR OPTIMIZATION APPROACH

Autora: Tatiane Fernandes Figueiredo
Orientador: Manoel Bezerra Campêlo Neto
tatianefernandes@lia.ufc.br,mcampelo@lia.ufc.br
Universidade Federal do Ceará
Campus do Pici, BL 910, CEP 60.440-900, Fortaleza, Ceará, Brazil.

RESUMO

Dado um conjunto de indivíduos, cada um com uma habilidade única, e uma rede social captando a afinidade mútua entre eles, o *Team Formation Problem* (TFP) visa encontrar uma única equipe que reúna as habilidades necessárias para realizar uma tarefa, enquanto busca otimizar os custos de comunicação entre os indivíduos envolvidos. Na primeira parte desta tese, estudamos uma generalização do TFP denominada como *Multiple Team Formation Problem* (MTFP), que permite múltiplas equipes com demandas distintas de trabalhadores para cada tipo de habilidade, assim como a possibilidade de fracionamento do tempo de dedicação de cada indivíduo entre as equipes. Na segunda parte, introduzimos uma nova variante do TFP, denominada *Competitive Teams Formation Problem* (CTFP). Utilizamos a teoria do equilíbrio social para representar a rede social por meio de um grafo de sinal e consideramos simultaneamente custos de comunicação intra-equipe e inter-equipes. Para o MTFP, propomos um modelo de Programação Linear Inteira (PLI) e famílias de desigualdades válidas. Experimentos computacionais atestam que o modelo PLI fortalecido por desigualdades válidas supera consistentemente o modelo quadrático apresentado na literatura para resolução do MTFP. Também consideramos uma versão generalizada do MTFP em que os indivíduos podem ter múltiplas habilidades. Para tal, adaptamos o modelo PLI inicial, gerando dois novos modelos, e apresentamos um conjunto de desigualdades válidas que os fortalecem. Para o CTFP, propomos um modelo PLI, além de desigualdades válidas que melhoram o desempenho computacional do mesmo. Por fim, encerramos a tese com as conclusões gerais e trabalhos futuros.
Palavras-chave: Problema de formação de equipes. Branch & Cut. Desigualdades Válidas.

ABSTRACT

Given a group of individuals, each one with a single skill, and a social network capturing the mutual affinity among them, the Team Formation Problem (TFP) aims to find a single team that meets the skills needed to perform a task while seeking to optimize the communication costs between the involved individuals. In the first part of this thesis, we study a generalized version of the TFP denominated as Multiple Team Formation Problem (MTFP), which allows multiple teams with distinct demands of workers per ability and fractions of dedication time per team for each individual. In the second part, we introduce a new variant of the TFP to be called Competitive Teams Formation Problem (CTFP). We use the theory of social balance to represent the social network as a signed graph and consider both intra-team and inter-teams communication costs. For the MTFP, we propose an Integer Linear Programming (ILP) model and sets of valid inequalities. Computational experiments attest that the ILP model strengthened by valid inequalities consistently outperforms the existing quadratic model for MTFP. We also consider a generalized version of the MTFP where individuals may have multiple skills. To handle this version, we adapt the initial ILP model into two new models and present other valid inequalities that strengthen them. For the CTFP, we propose an ILP model and valid inequalities that enhance the computational performance of the model. Finally, we close this thesis with general conclusions and future works.
Keywords: Team formation problem. Branch & Cut. Valid Inequalities.

https://proceedings.science/p/157568?lang=pt-br

1. Team Formation Problems

In the last decades, work teams formation has been shown as pivotal to the organizational success of companies [Cohen e Bailey, 1997; Barrick et al., 1998]. In the face of hard competition and technological challenges, an increasing number of companies has used work group strategies as an effective approach to resolve employee motivation issues and to accomplish organizational productivity goals [Neuman et al., 1999].

Although much scientific research has proven that the technical competences are fundamental attributes on teams design [Fitzpatrick et al., 2001; Tseng et al., 2004; Hlaoittinun et al., 2007], the success or failure of a work team is also strongly influenced by the health of the relationships amongst its members [Ballesteros-Pérez et al., 2012]. The social identity theory suggests that the more members identify with their respective teams, the more likely they are to actively contribute to the welfare of the group and work towards common goals [Gundlach et al., 2006]. Indeed, empirical research indicates that members of these teams display higher affective commitment and have higher performance when compared to teams that have low cohesion [Andrews et al., 2008].

Thus, the challenge of assembling successful work teams can be defined as the problem of finding sets of workers who have certain desired skills, while simultaneously considering the social compatibility among them [Lappas et al., 2009]. This kind of problem has received great attention from the OR community over the past years. In the literature, several ways to measure the quality of relationships among team members have been proposed, most of them based on psychology tools such as Myers-Briggs Type Indicator [Myers, 1962, 1997], Kolbe Conative Index [Kolbe, 1993], Sociometry [Moreno, 1941].

The first attempt to incorporate skills requirements and expectations of group synergy into a quantitative model was presented in 2001. Askin e Huang [2001] introduced a mixed-integer goal programming model to form teams for cellular manufacturing, where the analysis of social relationships was accomplished through the Kolbe Conative Index (KCI) [Kolbe, 1993]. KCI was used to study the behavior or instinctive drive of each individual as a way to evaluate the interpersonal style of workers. In the sequel, Chen e Lin [2004] presented a quadratic programming model that receives as input the personality profile of each worker. For such, the Myers–Briggs type indicator [Myers, 1997] was used to identify the personality and preferences of each team member.

With the advent of online social networks, Wi et al. [2009] and Lappas et al. [2009], in a pioneering way, introduced the use of Social Network Analysis (SNA) techniques to measure relationships between members of possible work teams. Using a graph that represents a social network, basic SNA concepts such as degree, centrality and closeness were employed in algorithms and mathematical models to identify a team of experts. Additionally, Lappas et al. [2009] formally defined the problem of finding a work team in a social network. Given a task that requires a set of skills and a pool of workers organized in a social network, each one possessing one of the skills, the *Team Formation Problem* (TFP) refers to finding a subset of the workers that cover the skill requirements and can communicate effectively with each other. The concept of effective communication can be measured by various *communication cost* metrics. In their work, Lappas et al. [2009] computed the communication cost based on the distances between team members in the social network. They proved that, similar to many other grouping optimization problems, the TFP is NP-Hard (even for single team instances).

Due to the current relevance of work teams formation in the business context and the popularization of online social networks, the seminal paper by Lappas et al. [2009] has triggered an increasing number of researches on the TFP or related problems. Many papers have been published

at a fast pace in literature [e.g. Li e Shan, 2010; Kargar e An, 2011; Majumder et al., 2012; Anagnostopoulos et al., 2012; Ballesteros-Pérez et al., 2012; Rangapuram et al., 2013; Gutiérrez et al., 2016; Campêlo et al., 2020; Kouvatis et al., 2020, and references therein]. Some of them present extensions for the TFP, for instance dealing with multiple team formation, part-time allocation and/or multiple skilled workers, as well as with different metrics to evaluate the communication costs.

In this scenario, problems involving the formation of multiple teams have been the focus of more recent works. Ballesteros-Pérez et al. [2012] introduced an optimization process carried out by means of matrix calculations that seeks both to maximize positive interactions and to minimize negative interactions between pairs of team members. For such, the authors assume that all employees have the same skills and knowledge to carry out the required tasks. By their turn, Gutiérrez et al. [2016] presented a generalization of the TFP which allows distinct demands of workers per skill, as well as fractions of dedication time per worker in each team. In this generalized version, called *Multiple Team Formation Problem* (MTFP), it is still assumed that each worker possesses a unique skill. The authors presented a quadratic optimization model, as well as constraint programming and heuristic solution approaches. In Campêlo et al. [2020], another variant of TFP was considered, where disjoint teams have to be formed from a set of individuals with multiple skills. The problem was show to be NP-hard even when a single team and unique skill are considered. An integer linear program (ILP) along with a polyhedral study were presented. Computational experiments with the ILP model and a simulated annealing heuristic are reported in that work.

In the three aforementioned papers, the communication costs between workers are measured using *sociometric matrices*. These matrices, derived from the sociometric theory developed by Moreno [1941], reveal the "social atoms" formed by each individual in his/her social network. Through this theory, the relationships between individuals can be modeled by an edge-weighted graph. By considering the workers as the vertices of a graph, the affinity of a team can be defined as a function of the weights of the edges linking the team members. The most common way to edge-weighting is the use of simple ordinal rating scales [Roistacher, 1974]. In particular, the rates may simply lie in the set $\{-1, +1\}$, for example to represent like/dislike or trust/distrust [Heider, 1946]. In this case, signs $+/-$ can be used instead so as to obtain a *signed graph*.

Also studied by social psychologist Heider [e.g Heider, 1946], the signed graphs were used to describe the relationships between individuals belonging to a same group, providing a systematic statement of the principle of *structural balance*. The structural balance theory is based on the notion of cognitive consistency between friendship and hostility. For example, an enemy of a friend is probably my enemy as well, while a friend of a friend is probably my friend or can become one. In simple terms, the interaction of individuals follows the tendency to create stable (albeit not certainly conflict-free) social groups [Levorato e Frota, 2017].

Although the structural balance analysis has proven to be an attractive tool for social network researchers [e.g. Yang et al., 2007; Doreian e Mrvar, 2009; Leskovec et al., 2010; Tang et al., 2016] and has motivated the definition of Integer Linear Programming (ILP) models to solve clustering problems [e.g. Bansal et al., 2004; Brusco e Steinley, 2011; Figueiredo e Frota, 2013, 2014], its use to ground communication costs functions in team formation problems is still recent. A single work (by Kouvatis et al. [2020]) provided some definitions of compatibility between pairs of users in a signed network, and algorithms for computing it. The authors also defined new versions of the TFP in signed networks and presented heuristic algorithms for solving them.

This thesis presents approaches for exact resolution of two problems derived from TFP. The thesis structure follows the format of scientific articles. This means that each chapter is self-contained. It includes its own introduction and bibliographic review of the approached problems,

https://proceedings.science/p/157568?lang=pt-br

in addition to the obtained results and specific conclusions. Part 1 is devoted to the Multiple Team Formation Problem (MTFP) defined by Gutiérrez et al. [2016] and its generalization to comprise the case of multiple skills individuals, denominated MMTFP. In Part 2, we introduce and mathematically define the Compatible Teams Formation Problem. It uses the concept of structural balance to form teams of skilled individuals. The following is a brief summary of our contributions and published papers.

1.1. Our contributions and published papers

As mentioned before, the thesis has two main parts. In the first one, we focus on the MTFP, as defined by Gutiérrez et al. [2016]. Besides, we consider a generalization of the problem, where individuals may have multiple skills, as in the extension of TFP, referred to as Sociotechnical Teams Formation Problem (STFP), proposed by Campêlo et al. [2020]. The characteristics of the mentioned problems are summarized in the Table 1.

Table 1: Summary of the characteristics of the central problems related to MTFP.

Reference	# teams		Individuals per skill		Skills per individual		Fractions of dedication time	
	$= 1$	≥ 1	$= 1$	≥ 1	$= 1$	≥ 1	No	Yes
TFP (Lappas et al. [2009])	✓		✓		✓		✓	
MTFP (Gutiérrez et al. [2016])		✓		✓	✓			✓
STFP (Campêlo et al. [2020])		✓		✓		✓	✓	
Part 1 of the thesis		✓		✓	✓	✓		✓

In the papers presented by Gutiérrez et al. [2016] and Campêlo et al. [2020], the communication costs are calculated by considering signed and weight sociometric matrices that measure the degree of relationship between each pair of individuals belonging to a social network. These matrices can also be represented by an edge-weighted graph, so that the affinity of a team is given as a function of the edge weights of each pair of individuals belonging to the team. More precisely, in the the MTFP the communication cost of a team is the summation of the weights of edges induced by its members, taking proportionally to their fractions of dedication times.

To solve the MTFP, we presented an Integer Linear Program formulation. Since the problem is NP-Hard, instances of practical interest may not be solved to optimality in a reasonable time. In order to improve resolution time, we proposed valid inequalities that are derived via reformulation-linearization techniques - RLT [Sherali e Adams, 1990]. The performance of our ILP formulation together with the inequalities was compared with the quadratic model by Gutiérrez et al. [2016], unique model presented in the literature. The comparative analysis took into account a set of benchmark instances of size and characteristics compatible with real applications as well as instances generated from real-world social networks. It showed that our approach consistently outperforms — in terms of effectiveness and efficiency — the quadratic model by Gutiérrez et al. [2016]. Besides, our numerical experiments indicated that the derived valid inequalities are very useful to strengthen the formulation and to significantly reduce its resolution time.

https://proceedings.science/p/157568?lang=pt-br

Still in the first part, we use extra variables to straightforwardly adapt our MTFP model to deal with the case where individuals may have multiple skills. Besides, we project out some of the variables to get a second ILP model with fewer variables but with an exponential number of constraints. We show how to separate these constraints in polynomial time. The correctness of this model and the separation procedure is demonstrated via max-flow/min-cut arguments. We also present valid inequalities for both formulations, again using RLT techniques. We apply the separation routine to solve the second program with the branch-and-cut method. We computationally compare the two formulations and show their potential to solve instances generated from the original MTFP instances, including randomly generated graphs as well as graphs obtained from real world social networks.

Finally, we try to draw general conclusions on which parameters or combinations of parameters most affect the hardness of the MTP and its version with multiple skills. In order to offer some insights into this issue, we have plotted some graphics to compare the variation in the average computational times versus some specific parameters. We analyze the integer and binary quadratic formulations, the pure ILP formulations and its best strengthened version.

Preliminary results that we have obtained for the MTFP were presented at CLAIO 2018. The complete version of this work was published in Computers & Operations Research [Campêlo e Figueiredo, 2021]. A second paper with the results for the MTFP with multiple skills, in this thesis denominated by MMTFP, was presented at SBPO 2021. An extended version of this paper is being finalized to be submitted to a journal. It is worth to mention that my approach to MTFP and MMTFP is similar to the one I have used in my Master's dissertation, but to another variant of the TFP where disjoint teams have to be formed from a set of multi-skilled individuals that can dedicate their working time to at most one project. Some results for this latter problem obtained during the Master's and the beginning of PhD have been published in Annals of Operations Research [Campêlo et al., 2020].

In the second part of the thesis, we introduce a new version of the TFP that considers the concepts presented by Kouvatis et al. [2020] about compatible groups of individuals in social networks represented by signed graphs. These notions are related to the formation of competitive work teams. In the introduced version of the TFP, to be called *Competitive Teams Formation Problem* (CTFP), we simultaneously consider skill requirements and structural balance constraints when forming multiple teams. It is required to exist only positive relationships between individuals of the same team and only negative relationships between individuals of different teams. This intends that workers within a same team have no disgust to each other whereas teams are considerably competitive with each other. The objective is to include as many individuals as possible in these teams in order to satisfy the skill requirements of each team.

To solve the CTFP, we present an integer linear program formulation and characterize valid inequalities that enhance its computational performance. The inequalities are derived from edge-contraction operations on the input signed graph which are defined using the principle of structural balance. This favors the description and generation of the inequalities, as the size of the graph is iteratively reduced with the operations. Since each group of valid inequalities presented is potentially exponential in size, and the separation problems are usually NP-hard, we propose a set of polynomial-time algorithms to generate the inequalities used to strengthen the model. We carry out some computational experiments to evaluate the performance of the ILP formulation and the inequalities. The benchmark instances comprise a set of synthetic instances of size and characteristics compatible with real applications as well a set instance adapted from Gülpinar et al. [2004]. The computational results indicate that the derived valid inequalities are very useful to strengthen

https://proceedings.science/p/157568?lang=pt-br

the formulation and to significantly reduce its resolution time.

As future work, we would suggest the design of efficient separation heuristics, which would allow the use of the valid inequalities as cuts. A deeper study of the polytope associated with the CTFP would be another proposal for the continuation of this work. Results on these two points could be used to improve the efficiency of the branch-and-cut. Similarly to the MTFP, we could also think of several possible variants of the CTFP. For instance, the time dimension could be introduced as well. Particularly, one could consider other communication cost metrics in signed graphs, such as those presented by Kouvatis et al. [2020], where negative edge parity in the shortest paths between a pair of individuals is used to define their compatibility. As the authors provide only heuristic algorithms for the generation of the teams, an interesting track of research would be the generation of ILP models using such metrics. It is worth noting the valid inequalities presented here would also be valid for these new models.

Regarding this second problem, there is a paper submitted to European Journal of Operational Research (EJOR) under revision. Part of the results was obtained during a scientific visit to Univertité d'Avignon et des Pays de Vaucluse, in collaboration of professor and researcher Rosa Figueiredo.

https://proceedings.science/p/157568?lang=pt-br

References

Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., e Leonardi, S. (2012). Online team formation in social networks. In *Proceedings of the 21st International Conference on World Wide Web*, p. 839–848.

Andrews, M. C., Kacmar, K. M., Blakely, G. L., e Bucklew, N. S. (2008). Group cohesion as an enhancement to the justice—affective commitment relationship. *Group & Organization Management*, 33(6):736–755.

Askin, R. G. e Huang, Y. (2001). Forming effective worker teams for cellular manufacturing. *International Journal of Production Research*, 39(11):2431–2451.

Ballesteros-Pérez, P., González-Cruz, M. C., e Fernández-Diego, M. (2012). Human resource allocation management in multiple projects using sociometric techniques. *International Journal of Project Management*, 30(8):901–913.

Bansal, N., Blum, A., e Chawla, S. (2004). Correlation clustering. *Machine learning*, 56(1-3): 89–113.

Barrick, M. R., Stewart, G. L., Neubert, M. J., e Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83 (3):377–391.

Brusco, M. e Steinley, D. (2011). Exact and approximate algorithms for variable selection in linear discriminant analysis. *Computational Statistics & Data Analysis*, 55(1):123–131.

Campêlo, M. e Figueiredo, T. (2021). Integer programming approaches to the multiple team formation problem. *Computers & Operations Research*. to appear.

Campêlo, M., Figueiredo, T., e Silva, A. (2020). The sociotechnical teams formation problem: a mathematical optimization approach. *Annals of Operations Research*, 286:201–216.

Chen, S.-J. e Lin, L. (2004). Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *IEEE Transactions on Engineering Management*, 51 (2):111–124.

Cohen, S. G. e Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3):239–290.

Doreian, P. e Mrvar, A. (2009). Partitioning signed social networks. *Social Networks*, 31(1):1–11.

Figueiredo, R. e Frota, Y. (2013). An improved branch-and-cut code for the maximum balanced subgraph of a signed graph. Technical report, arXiv preprint arXiv:1312.4345.

Figueiredo, R. e Frota, Y. (2014). The maximum balanced subgraph of a signed graph: Applications and solution approaches. *European Journal of Operational Research*, 236(2):473–487.

Fitzpatrick, E., Askin, R., e Goldberg, J. (2001). Using student conative behaviors and technical skills to form effective project teams. In *31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education. Conference Proceedings*, volume 3, p. S2G–8. IEEE.

https://proceedings.science/p/157568?lang=pt-br

Gülpinar, N., Gutin, G., Mitra, G., e Zverovitch, A. (2004). Extracting pure network submatrices in linear programs using signed graphs. *Discrete Applied Mathematics*, 137(3):359–372.

Gundlach, M., Zivnuska, S., e Stoner, J. (2006). Understanding the relationship between individualism–collectivism and team performance through an integration of social identity theory and the social relations model. *Human Relations*, 59(12):1603–1632.

Gutiérrez, J. H., Astudillo, C. A., Ballesteros-Pérez, P., Mora-Melià, D., e Candia-Véjar, A. (2016). The multiple team formation problem using sociometry. *Computers & Operations Research*, 75: 150–162.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, 21(1):107–112.

Hlaoittinun, O., Bonjour, E., e Dulmet, M. (2007). A team building approach for competency development. In *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, p. 1004–1008. IEEE.

Kargar, M. e An, A. (2011). Discovering top-k teams of experts with/without a leader in social networks. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, p. 985–994.

Kolbe, K. (1993). *Pure instinct*. Random House, New York.

Kouvatis, I., Semertzidis, K., Zerva, M., Pitoura, E., e Tsaparas, P. (2020). Forming compatible teams in signed networks. In *Proceedings of the 23rd International Conference on Extending Database Technology (EDBT)*, p. 363–366. OpenProceedings.org.

Lappas, T., Liu, K., e Terzi, E. (2009). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 467–476. ACM.

Leskovec, J., Huttenlocher, D., e Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1361–1370.

Levorato, M. e Frota, Y. (2017). Brazilian congress structural balance analysis. *Journal of Interdisciplinary Methodologies and Issues in Science*, 2:1–27.

Li, C.-T. e Shan, M.-K. (2010). Team formation for generalized tasks in expertise social networks. In *2010 IEEE Second International Conference on Social Computing*, p. 9–16. IEEE.

Majumder, A., Datta, S., e Naidu, K. (2012). Capacitated team formation problem on social networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 1005–1013.

Moreno, J. L. (1941). Foundations of sociometry: An introduction. *Sociometry*, 4(1):15–35.

Myers, I. B. (1962). *The Myers-Briggs type indicator manual*. Educational Testing Service, Princeton, NJ.

Myers, I. B. (1997). *Introduction to type: A description of the theory and applications of the Myers-Briggs Type Indicator*. Vision Australia Student Support.

https://proceedings.science/p/157568?lang=pt-br

Neuman, G. A., Wagner, S. H., e Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, 24(1):28–45.

Rangapuram, S. S., Bühler, T., e Hein, M. (2013). Towards realistic team formation in social networks based on densest subgraphs. In *Proceedings of the 22nd international conference on World Wide Web*, p. 1077–1088.

Roistacher, R. C. (1974). A review of mathematical methods in sociometry. *Sociological Methods & Research*, 3(2):123–171.

Sherali, H. e Adams, W. (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3 (3):411–430.

Tang, J., Chang, Y., Aggarwal, C., e Liu, H. (2016). A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, 49(3):1–37.

Tseng, T.-L. B., Huang, C.-C., Chu, H.-W., e Gung, R. R. (2004). Novel approach to multi-functional project team formation. *International Journal of Project Management*, 22(2):147–159.

Wi, H., Oh, S., Mun, J., e Jung, M. (2009). A team formation model based on knowledge and collaboration. *Expert Systems with Applications*, 36(5):9121–9134.

Yang, B., Cheung, W., e Liu, J. (2007). Community mining from signed social networks. *IEEE Transactions on Knowledge and Data Engineering*, 19(10):1333–1348.

https://proceedings.science/p/157568?lang=pt-br