

社会网络中的团队形成问题研究综述*

黄健斌¹, 孙晓晶^{1,2}, 周瑜^{1,2}, 吕泽^{1,2}, 孙鹤立³, 贾晓琳³



¹(西安电子科技大学 软件学院, 陕西 西安 710071)

²(西安电子科技大学 计算机学院, 陕西 西安 710071)

³(西安交通大学 计算机科学与技术系, 陕西 西安 710049)

通讯作者: 黄健斌, E-mail: jbhuan@xidian.edu.cn

摘要: 团队形成问题作为一个出自于运筹学中的问题, 已经得到了深入的研究. 然而, 随着各种社交平台的流行以及网络通信的迅速发展, 社会化网络中的团队形成再次调动起了众多学者的研究热情. 社会化网络中的团队形成问题与传统的团队形成有很大的不同, 因此, 它不能再简单地借助集合覆盖、任务分配或者最大化匹配等经典问题来解决. 在充分调研和分析的基础上, 对社会化网络下的团队形成问题的研究现状进行了阐述. 综述了社会化网络下的团队形成问题的各种变形问题及其优化方法, 同时还归纳介绍了该研究中使用的实验数据集和评价指标. 最后, 对该问题今后可以开展的方向进行了展望.

关键词: 团队形成; 社会化网络; 任务分配; 变形问题; 优化

中图法分类号: TP311

中文引用格式: 黄健斌, 孙晓晶, 周瑜, 吕泽, 孙鹤立, 贾晓琳. 社会网络中的团队形成问题研究综述. 软件学报, 2017, 28(4): 993-1009. <http://www.jos.org.cn/1000-9825/5152.htm>

英文引用格式: Huang JB, Sun XJ, Zhou Y, Lü Z, Sun HL, Jia XL. Research survey on team formation in social networks. Ruan Jian Xue Bao/Journal of Software, 2017, 28(4): 993-1009 (in Chinese). <http://www.jos.org.cn/1000-9825/5152.htm>

Research Survey on Team Formation in Social Networks

HUANG Jian-Bin¹, SUN Xiao-Jing^{1,2}, ZHOU Yu^{1,2}, LÜ Ze^{1,2}, SUN He-Li³, JIA Xiao-Lin³

¹(School of Software, Xidian University, Xi'an 710071, China)

²(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

³(Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Team formation problem has received extensive study in operational research (OR). Recently, along with the popularity of various social sites and the rapid development of Internet communication, team formation in social networks has attracted great study enthusiasm of scholars once again. Team formation problem in social networks differs much from the traditional team formation which takes no account of the social relations and effective communication in a team. Thus it cannot be solved simply by means of set-cover, task assignment or maximum matching any more. Based on sufficient research and analysis, this work surveys the state of the art of the social-based team formation problem. The variants of this problem and their optimization methods are reviews as well. In the meantime, the data-sets and evaluation criteria of experiments in this study are introduced. Finally, the prospect of future work is presented.

Key words: team formation; social network; task assignment; variants; optimization

对于一个大规模任务, 比如大型软件开发、大型科研项目等, 常常需要构建一个团队进行分工完成. 然而, 团

* 基金项目: 国家自然科学基金(61472299, 61540008, 61672417); 陕西省自然科学基金基础研究计划(2014JQ8359)

Foundation item: National Natural Science Foundation of China (61472299, 61540008, 61672417); Basic Research Foundation of Shaanxi Province (2014JQ8359)

收稿时间: 2016-03-03; 修改时间: 2016-06-14, 2016-09-02; 采用时间: 2016-10-22; jos 在线出版时间: 2016-11-24

CNKI 网络优先出版: 2016-11-24 13:41:15, <http://www.cnki.net/kcms/detail/11.2560.TP.20161124.1341.005.html>

队内部的高效协作是影响团队效率的一个重要因素.随着网络信息技术的发展与普及,各种社交平台不断兴起,网络社交用户越来越多.我们可以通过在线社交网络及其他社会化网络发掘到人与人之间的社交关系,从而度量团队成员之间的通信代价,进而通过优化整个团队的通信开销来组建一个高效的项目团队.这就是近年来在社会网络分析和数据挖掘领域受到广泛研究的社会化网络中的团队形成问题.

团队形成问题较早时候已经在运筹学领域得到了深入研究,然而随着互联网技术的发展,社会化网络下的团队形成问题再次掀起了研究热潮.尤其是随着一些在线劳动力市场平台(如 Freelancer,Guru)、社交事件组织网站(如 Meetup,Plancast)等在线活动应用的出现,社会化网络中的团队形成问题在新的应用背景下变得更具趣味性和挑战性.过去的研究中,由于没有网络关系的背景,很多组合优化以及集合覆盖、二部图的最大匹配等方法都可以对该问题进行求解.社会化网络背景下的团队形成问题不同于运筹学中的经典任务分配问题,不能再简单地用集合覆盖和最大化匹配等方法来解决.

本文在对近年来团队形成问题的研究进程进行跟踪分析的基础上,对该问题的相关工作进行了深入的总结归纳.第 1 节介绍团队形成问题的背景知识及其应用价值.第 2 节简单介绍传统的无网络下的团队形成问题的研究.第 3 节对社会化网络下团队形成的相关变形问题进行归纳分类,主要包括基于运筹学组合优化和集合覆盖的无关系网络的研究和网络关系背景下的研究.第 4 节从优化目标出发,对团队形成问题的优化方法进行介绍.第 5 节介绍在该问题的研究中所使用的实验数据集和实验构造方法.第 6 节对该问题的研究进行总结,并展望未来的研究方向.

1 引言

随着科学研究和生产生活中的学科相互融合以及多领域知识的互相渗透,越来越多的任务和活动需要不同知识背景的人员跨学科合作方能完成.另一方面,随着网络通信方式的普及以及各种社交平台的广泛应用,人们的社交活动与网络产生了紧密的结合,在线合作沟通已成为新的趋势.近些年来,利用社交网站进行移动办公也变得流行起来.通过社交网络,大家可以分享信息,在线交流参与活动,在线视频会议,远程讨论决策,不仅可以交流互动,还可以合作共事.以上两种因素的结合,使得团队形成问题再次被聚焦,成为研究热点.团队形成的概念除了在社会化网络中有重要影响之外^[1-10],在多代理系统^[11]以及分布式人工智能^[12,13]等领域也都有着潜在的应用背景.

团队形成,不仅存在于普通人的日常生活或专业人员的工作业务中,如今更是有许多应用系统可自动形成团队.对于一些企业来说,通过不断增加员工人数来满足日益扩展的业务需求是不合算的.面对便捷的网络信息渠道,他们更倾向于在网上随时动态地获取能够满足当前需求的各种技能的专家.当有新的业务来临时,再去寻找新的业务需要的专家.近年来,国内外很多在线劳动力市场平台,如 Freelancer(www.freelancer.com),eLance(www.elance.com),Getfriday(www.getfriday.com),Guru(www.guru.com)以及 oDesk(www.odesk.com)等新兴招聘网站应运而生,并且引起了许多研究者的密切关注.他们对这一新型雇佣方式所带动的产业和出现的经济现象进行多维度挖掘与探索^[14-17].在这些平台上,用户主要分为两类角色:买方公司(buyer company)和服务提供者(service provider),而网站本身只是一个提供信息资源和保证交易的中间媒介^[14].买方公司可以把各种 IT 服务外包给网站注册的专业服务提供人员(专家),同时附上项目描述、需要的技能以及“招标”截止时间;而服务提供者如果愿意应聘并且能够完成该项目,则可通过“投标”竞价的方式来提交申请.在海量的投标专家中挑选出最终的应聘者来签约,是一项比较庞杂的任务,需要慎重选择和多方面考量.而随着现实业务复杂程度的增加,经常会有公司发布比较大型的项目,如软件系统开发等.这些项目由于需求的技能较多,因此往往需要挑选一组人员来合力协作,这就对买方公司的决策大大增加了难度.文献[18]模拟了真实的 Freelancer 中的团队形成场景,提出 ClusterHire 问题,对如何挑选专家组建团队来完成给定项目进行了研究.

在一个项目中,团队成员的业务能力固然是重要的,但同时,在合作中团队的协作效率也很重要,和谐融洽的团队合作会使得整个工作进展顺利,并在很大可能性上获得最终的成功.比如在一些大型软件开发系统中,整个团队的进度是各个模块共同作用的结果,而各个模块之间也会相互影响,因此,我们既要关注各个模块的进

度,也要关注模块之间的相互影响.一个软件项目可能需要多种不同技能,所以需要选一组人员来共同合作完成.在组建团队时,首先要考虑所选团队成员是否能够完成项目(即成员所拥有的技能是否能覆盖掉项目所需技能),其次还必须考虑成员之间的关系(即成员之间的关系是否亲近、沟通交流是否便利高效).除了软件项目开发外,实际生活中的其他一些活动也处处可见合作的场景,比如一篇论文的合著、一部影视剧中演员的合作,甚至一台外科手术中主刀医师和麻醉师以及护士助手的密切配合等,都是合作的例子.文献[19,20]研究发现,人与人之间的合作效率主要受到熟悉度(familiarity)、地理邻近度(geographical proximity)以及学科相似性(disciplinary similarity)这3大主要因素的影响.然而,能够相对完全地对这3大因素进行反映的一个载体就是社会网络.通过社会网络,可以挖掘到人与人之间的互动交流情况,进而能够判断他们之间能否有效合作,从而为组建团队提供参考.因此,社会化网络中的团队形成问题有着重大的研究价值和应用前景.

团队形成问题是指:给定一组候选专家(每位专家有自己的技能集合表示其拥有的技能)及一个项目(该项目依赖于一组技能),在候选专家中寻找一个团队,使得团队的技能集合能覆盖项目所需要的技能.该问题(或者变形问题)和集合覆盖(set-cover)问题、任务指派(task assignment)问题、甚至社团搜索(community search)问题都有相似之处,因此有时会在研究中进行互比较和转化.下面对这3类相关问题进行简单介绍.

- 集合覆盖问题

集合覆盖问题是一个典型的最优化问题,在组合优化、计算机科学以及复杂性理论中都有相关研究.它可以用来模拟现实场景中的资源选择问题,在人员调度、资源分配、网络安全、电路设计、运输车辆路径安排等领域有着广泛的应用,因而吸引了众多科学家的研究兴趣.

给定一个有限集合 X 和 X 的子集集合 \mathcal{F} ,且 X 的每一个元素属于 \mathcal{F} 中的至少一个子集,即 $X = \bigcup_{S \in \mathcal{F}} S$. 如果 \mathcal{F} 的一个子集 $C \subseteq \mathcal{F}$, 满足 $\bigcup_{S \in C} S = X$, 则称 C 是集合 X 的一个覆盖. 常规的集合覆盖问题定义为寻找 X 的一个最小规模覆盖 C , 即 $\bigcup_{S \in C} S = X$, 且 $|C|$ 最小. 这是一个 NP 难的最优化问题, 当前无法找到多项式时间的解法. 集合覆盖的经典解法是使用贪心的近似算法 Greedy-Set-Cover 来求得近似解, 每次选出 \mathcal{F} 中的一个能够最多地覆盖掉 X 中未被覆盖元素的子集 S 加入到 C 中. 集合覆盖在现实应用中有两种重要的变形问题, 分别是带权集合覆盖和 k -Set Cover 判定问题. 若对于任意的 $S \in \mathcal{F}$ 给定一个权重 $w(S)$, 求 X 的一个覆盖 C , 使得 $\sum_{S \in C} w(S)$ 最小, 这就是加权集合覆盖问题. 而判定版本的 k -Set Cover 则是回答 X 是否存在一个覆盖 C 使得 $|C| \leq k$ 的判定问题. 这两个问题的复杂度都是 NP 完全的.

集合覆盖问题之所以备受关注,是因为它能够很多生活中常见的组合问题进行抽象.考虑这样一个简单的例子:假设 X 表示解决某一问题所需的各种技能的集合,此外,有一个给定的可用来解决该问题的人员集合.现在希望组建一个人数尽可能少的委员会来解决该问题,使得对于 X 中的每一种技能,委员会中都有一位成员掌握该技能.这个问题就属于一个简单的团队形成问题,完全可以建模成一个集合覆盖来求解.只需要把专家集合映射到 \mathcal{F} ,同时把项目需要的技能集合映射到 X ,这样一个团队形成问题就完全转化成了集合覆盖问题.

- 任务指派问题^[21]

在实际生产、生活中,常常会遇到需要指派不同的工作人员去完成不同工作的场景.由于个人的专业特长不同,各人去完成各项任务需要花费的时间以及完成之后可以创造的价值一般也不同.因此,任务指派问题研究的是如何建立人员与任务之间的对应关系,使得目标(利润或者时间)最优.在基本的指派问题中,任务的个数和人员数目是相同的,这样的分派问题又称为平衡指派问题,运筹学研究中已有了经典的解法,如匈牙利算法、削高排除法、缩阵分析法等.对于任务数与人数不相等的情况,通常采用假想虚设数目少的量来转化到平衡指派问题,然后使用上述算法进行求解.这种任务指派问题和文献[18]中提出的 ClusterHire 问题很类似.ClusterHire 中员工的薪酬支出就相当于该问题中的完成任务需要花费的时间,而完成每个项目的预计收入则相当于该问题中完成每项任务后创造的利益.经过这样的映射,ClusterHire 问题就可以转化成一个任务指派问题.

- 社团搜索问题^[22,23]

在真实的社交网络或者生物网络中存在着大量社团结构,因此,社交网络中的社团检测(*community detection*)作为图挖掘领域的一个重要问题,一直备受瞩目.而社团搜索(*community search*)问题则是在社团检测问题上衍生出来的一个基于查询的变形问题.具体来说,社团检测问题侧重于寻找网络中所有的连接紧密的子图,而社团搜索问题是寻找包含给定查询节点的链接紧密的子图.文献[22]中提出的社团搜索问题的一个很有用的场景是活动组织和团队形成,比如几位科学家想要组织一场研讨会,他们几位是确定要参与的组织者,但是需要再邀请几位跟他们相互熟识或之前合作过的专家成立一个专家小组一起举办.除此之外,社团搜索在现实世界中有很多其他的应用场景.如文献[22]所述,由 *leader* 主导的团队形成问题(见第 3.2.5 节)可以转换成社团搜索问题来解决:只需把 *leader* 映射到查询节点上,就可以套用文献[22]中的社团搜索算法求解该问题.

2 无社交网络的团队形成问题

假设有一个项目 P , 其所需要的技能集合为 $X = \{s_1, s_2, \dots, s_m\}$. 此外, 存在一组候选专家集合, 每个专家均关联一个技能集合 X_i . 团队形成问题就是在候选专家集合中寻找一组专家, 使得这些专家的技能集合包含项目所需要的所有技能. 这种传统的团队形成是一种不考虑人员社会关系网络的团队形成问题, 它与社交网络下的团队形成最大的区别就是忽略了团队成员之间的社交关系. 社交化网络下的团队形成问题会考虑团队成员之间的社交关系, 并要求所形成的团队具有较小的通信开销. 传统的团队形成只需要满足项目所需的技能条件, 对合作效率不作要求, 因此就不需要考虑成员间的社会关系.

在运筹学方法的应用中, 已经出现了大量关于团队形成的研究^[24-27]. 在这样一个不考虑人员社交关系网络的模型下, 该问题的目标就变成了所需技能向候选人员的分派问题^[7]. 在这个方向上的研究工作基本上都倾向于将团队形成问题表达成一个整数线性规划的模型, 然后在其上用各种方法寻找出一个从人员到技能需求的最优匹配. 组合优化问题的求解已经有了很多经典的方法可以采用. 例如, 文献[25]把该问题建模成一个整数规划问题, 采用了分支剪枝^[28]的方法进行求解; 文献[26,27]采用了遗传算法求解该分派问题, 形成最终的团队.

在这种基于运筹学组合优化和集合覆盖的研究中, 虽然不考虑社交关系网络, 但是也有好几种不同的实例. 文献[29]把团队形成问题建模成一个任务分配问题, 提出了一个总体框架来为多个项目形成组建团队. 在文献[29]中, Anagnostopoulos 等人考虑到了单人任务负载不均和团队大小之间的矛盾, 提出一个负载平衡的任务覆盖问题. 他们指出: 一个有效的项目分配方案应该不仅要满足人员与项目的契合度, 而且必须考虑分配方案对所有团队成员的公平性. 也就是说, 不应该出现某些人参与的项目数目过多, 而相反地, 有些人的任务却很少的情况. 但是同时, 这种工作量的均衡分配必然是要以增加团队人员数目为代价来实现的. 文献[29]研究的是多个项目的团队形成问题. 该问题是为每一个项目分配一个团队, 使得专家所参与的项目个数最少.

文献[18]在不考虑社交网络关系的情况下, 增加了两个已知条件: 团队人员薪酬支出和每个项目的预计收益, 提出一种以团队所得利润最大化为优化目标的 *ClusterHire* 问题. 该文与文献[29]中的问题最为接近, 都研究的是为多个项目分配专家. 文献[18]与文献[29]的两点最大不同之处在于: 首先, 文献[18]的研究目标是寻找一个团队能够完成多个项目, 使得整个团队所获利润最大化, 不同于文献[29]每一个项目建一个团队; 其次, 文献[18]没有像文献[29]那样将项目设置为随时间在线不断到达的, 而是假定所有的项目都是预先确定的. 传统的团队形成问题完全可以建模成集合覆盖和任务分派问题, 甚至文献[18]中提出的收益最大化的 *ClusterHire* 问题也可以用加权集合覆盖来求解. 该文借助集合覆盖的近似解法 *Greedy-Set-Cover* 提出了 *ExpertGreedy* 和 *Project Greedy* 两种算法, 分别使用“覆盖的技能最多”和“项目收益最大”两个贪心策略来求解该问题.

3 社交网络中的团队形成研究

当今, 人们每时每刻都在参与某个社交圈子, 人与人之间的紧密关系成为两个人有效合作的保障. 有过合作经历的人一起共事, 必然比新认识的人一起合作更加和谐、高效, 因为关系亲近的人之间的相互信任与能力认同感在很大程度上保障了合作的成功^[30]. 近年的研究工作中提出了一些估计社交成员间关系强度的度量方法^[31-35]. 在团队合作中, 团队成员的关系亲疏可以表现为同一公司员工所属组织层次的距离. 例如, 所属同一部

门的员工之间由于相距较近,容易及时协调沟通,所以他们这些人就容易合作.当然,地理距离不是关系强度的绝对衡量,特别是在现代虚拟网络时代,项目工作中的接洽交接都可以通过网络在线处理.所以,人与人之间关系的亲密度更多地可以通过社会网络中的关系网络图来发掘.换句话说,一个团队项目的成功不但取决于项目成员的专业技能,同时,彼此之间的有效合作、沟通也是很重要的.很多时候,团队成员的专业能力强并不能代表团队项目的成功与否,如果这些成员之间不能高效、融洽地沟通交流,项目的进度,甚至整个项目最终的成果都会受到影响.

例如,假定一个 IT 项目经理需要组建一个工程师团队完成一个需要以下技能要求的项目:算法、软件工程、Web 编程、分布式系统、Java 程序设计、数据库、软件项目管理.同时,我们假定有一个候选专家集 $\{a,b,c,d,e,f,g,h\}$,即,有 8 位候选专家有意加入该项目.用 S_i 表示专家 i 拥有的技能集合,则已知 $S_a=\{\text{分布式系统}\}$, $S_b=\{\text{Web 编程}\}$, $S_c=\{\text{数据库,Java 程序设计}\}$, $S_d=\{\text{算法}\}$, $S_e=\{\text{软件工程}\}$, $S_f=\{\text{算法,Java 程序设计}\}$, $S_g=\{\text{分布式系统,数据库}\}$, $S_h=\{\text{软件项目管理}\}$.而这 8 位专家之间的社交关系网络图如图 1 所示,其中,每个节点代表一位专家,每一对节点之间的一条边代表其对应的两个专家之间可以高效合作.

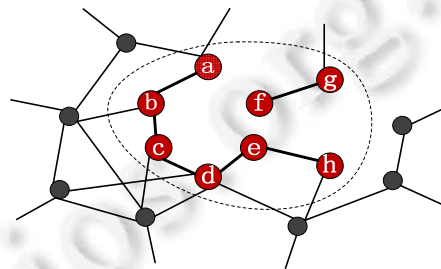


Fig.1 Cooperative relationship among experts $\{a,b,c,d,e,f,g,h\}$ in a social network

图 1 专家 $\{a,b,c,d,e,f,g,h\}$ 及其所在社会网络中的合作关系

如果不考虑这 8 位候选专家在其所处的社会网络中的交际关系,根据该项目的技能需求和每位专家所拥有的技能,可以看出, $\{a,b,c,d,e,h\}$ 和 $\{b,e,f,g,h\}$ 这两个专家子集都可以成为一个可选团队来完成该项目,因为这两个子集的专家所拥有的技能都可以完全覆盖项目要求的所有技能.但是,如果加上对这 8 位专家之间的社交关系的考虑,则由图 1 可以看出,只有子集 $\{a,b,c,d,e,h\}$ 是可以形成团队来完成该项目的.因为如图 1 所示,子集 $\{b,e,f,g,h\}$ 中的专家 b,e,h 和 f,g 之间是不连通的,即,这两部分专家不能有效地沟通合作.而子集 $\{a,b,c,d,e,h\}$ 中的 6 位专家之间的关系网络可以形成一个连通子图,所以, $\{a,b,c,d,e,h\}$ 组成的专家团队会是该项目的最优选择.从这个例子可以看出:社会关系网络的引入使得团队形成问题更加完善和合理,同时也增加了该问题的实际应用价值.

社会网络(social network)是指社会个体成员之间因为互动而形成的一种基于“网络”的相对稳定的关系体系.社会网络所构建的社会组织形式代表了现实生活中各种社会关系,包括朋友关系、同学关系、业务合作关系、种族信仰关系等.我们所讨论的社会网络中的团队形成问题,正是基于以上所描述的社会化关系网络,具体网络如:科学家合作关系网络、社交朋友圈、邮件网络等以个人为节点构成的刻画其所在的组织结构中的社会关系的网络.

3.1 问题形式化定义

给定一个项目(任务) P 以及其所需的所有技能集合 $X=\{s_1,s_2,\dots,s_m\}$.再给定一个候选专家集合 $C=\{c_1,\dots,c_n\}$,其中每个专家 c_i 拥有一种或多种不同的技能组成的集合记为 X_i ,显然有 $X_i \subseteq X$.同时,给定一个所有候选专家之间的社会关系网络 $G(C,E)$,用来刻画这些候选专家之间的交际关系以及关系的强弱. $G(C,E)$ 是一个无向带权图,图中的每个节点代表一位专家,每条边代表相应的两位专家之间有一定的社会关系.边上的权值则表达了两位专家的关系强度:权值越大,表示关系越强,则二者的通信代价越小;反之,表示关系越弱,通信代价也就越大.为描述方便起见,定义 $A_{s_i} \subseteq C$ 来表示 C 中拥有技能 s_i 的专家.同时定义一组专家 $T' \subseteq C$ 能覆盖的技能集合 $C(T',X) =$

$X \cap \left(\bigcup_{c_i \in T'} X_i \right)$. 社会网络中团队形成问题的基本形式就是在以上给定的条件下,寻找图 G 的一个子图 T ,使得 T 中

的所有节点所代表的专家拥有的技能合集能够覆盖项目所需技能,即 $C(T, X) = X$,并且要使得整个团队的通信代价最小.至于一个团队的通信代价,没有唯一的标准定义.根据问题的不同,可以定义不同的团队通信代价作为优化目标函数.

传统的无网络下的团队形成只需要选择一组专家,使得他们的技能集的并集满足项目所需的所有技能即可.而社会网络中的团队形成问题需要考虑团队成员间的合作效率,期望产生一个高效的团队,因此需要借助人员关系网络(如一个公司里员工之间的部门组织关系网络、社交平台里的朋友关系网络、计算机科学文献库 DBLP 中专家之间的合著关系网络等)来挖掘他们之间交际关系的强弱.该问题与传统无网络的团队形成问题相比,增加了很多挑战.

(1) 由于引入了网络,该问题蕴含了连通子图挖掘的相关子问题,使得复杂度大增;

(2) 社会网络下的团队形成是一个优化问题,它通过优化团队的通信代价函数来寻找一个合作高效的团队;

(3) 实际应用背景下的变形问题,会在基本问题的基础上增加更多的约束条件,进一步增加了问题复杂度.

3.2 各种变形问题

在近 20 年的研究中,团队形成问题从应用背景到问题定义以及研究方法都经历了一系列的演变.最初的基本团队形成问题被赋予了不同背景和条件,产生了一系列不同的变种问题,使得该研究更加充分全面.文献[4]中,Lappas 等人首次提出在团队形成问题中加入对候选专家社会网络关系的考虑,即社会网络中的团队形成.之后,相继出现了一系列对社会网络中的团队形成问题的研究^[1-9,36].这些工作中,研究的问题基于不同的场景约束,故而产生了一系列变形问题,包括单项目(SP)、多项目(MP)、指定技能所需人数的(GP)、技能分级的(SG)、Leader 主导的(leader)以及任务限量(capacitated)等种类的团队形成问题.下面将对这 6 种变形问题分别进行阐述.表 1 给出了网络关系下的团队形成问题在当前研究中的各种变形分类.

Table 1 A classification of variant problems in six aspects

表 1 6 种变形问题的分类

文献	社会化网络下的变形问题					
	SP	MP	GP	SG	Leader	Capacitated
[1]		■				
[2]					■	
[3]	■				■	
[4]	■					
[5]	■		■			
[6]	■					
[7]	■				■	
[8]	■					■
[9]			■	■		
[36]			■			

- 单个项目^[3-8].单个项目的团队形成问题指的是寻找一个最优的团队来完成一个单一项目,不需要考虑多个项目之间的人员交叉参与;
- 多项目^[1].多项目的团队形成指的是面向同一批候选专家,为多个项目寻找各自的团队;
- 指定技能所需人数^[5,9,36]的团队形成问题.主要思想是:在为项目组建团队时,对项目要求的每个技能所需的人员数目加以约束,要求每个技能至少需要一定数目的人员来完成;
- 技能分级的团队形成^[9].指的是对每位候选专家所拥有的技能,根据其水平高低给定一个等级值进行区分;然后,在选择团队成员的时候优先选择技能水平较高,即等级值较大的专家;
- Leader 主导的项目^[2,3,7]团队形成问题.指的是在一个项目中由一个主导人或者项目发起人来挑选其他团队成员,负责整个项目进行过程中各个模块间的所有沟通交流以及协调接洽;

- 任务限量的团队形成^[7].所谓任务限量的团队形成问题,是指限制每个专家的技能所能参与的最大项目个数.

3.2.1 单项目团队形成

单项目的团队形成只考虑完成一个项目,团队只需要满足这一个项目的要求.文献[3-8]都是针对单项目的团队形成问题进行的一系列研究.文献[4]定义了两种衡量整个团队的通信开销的实例,分别为基于子图半径的 Cc-R 和子图最小斯坦纳树(minimum Steiner tree)的边权之和 Cc-MST.寻找使得这两个通信开销最小的团队就对应地形成了两个问题:Diameter-TF 和 MST-TF 问题.在文献[4]中,作者指出这两个问题都是 NP 完全的,并给出了证明.针对 Diameter-TF 问题,文献[4]提出了基于文献[37]中的 Multichoice 算法的变形算法 RarestFirst.同时,针对 MST-TF 问题,文献[4]根据 MST-TF 问题与斯坦纳树问题的相似之处,提出了 CoverSteiner 和 EnhancedSteiner 两种算法.文献[3]提出了一种在社会网络中寻找 top- k 的最优团队的问题,指出:在很多情境下,用户往往是不满足于一个解决方案的,他们会比较倾向于希望得到前几个最优的答案,并将它们进行比较,然后根据个人偏好选择一个最满意的解.于是,在文献[3]中,作者提出了寻找 top- k 的最优团队的近似算法来对该问题进行研究.在文献[5]中,作者提出了一种基于集群的方法,把同一技能的所有专家集成为一个超节点,把所有专家之间的关系网络图压缩成一个集群图(group-graph),利用集群的模型解决该问题.文献[6]针对一个项目的完成,定义了一个“可负担的并可高效合作的”团队形成问题.该文在考虑项目的人员薪酬成本的同时,还考虑到了团队成员内部的通信开销,建立了一个双目标优化模型,并提出了几种 α - β 近似算法来求得 Pareto 最优解.值得一提的是,文献[8]与文献[6]研究的问题非常相近,也是对整个团队的雇佣支付成本和通信代价进行了双目标优化.文献[7]提出:在为一个项目组建团队的时候,应该考虑每个成员的最大工作容量,不应该出现某一成员承担的任务超出他的承受限额的情况.因此,在该文中定义了一种任务限量的团队形成问题.

3.2.2 多项目团队形成

在团队形成的实际场景中,有时需要为多个项目组建团队.多项目的团队形成问题的主要思想是:面向同一批候选专家,为多个项目寻找各自的团队.由于这些负责不同任务的团队是同一批专家集合,所以不可避免地会出现一个人参与多个项目的情况.而在一个项目里,每一个人也可能会负责多个技能.因此,多项目的团队形成问题往往会产生个别人工作负载过大、而有的人则负载很小的分配失衡的结果.文献[1]就对社会网络关系下的多项目团队形成进行了研究,并对上述矛盾进行了调和.

文献[1]定义了这样一个情境:多个任务以在线的流形式到来,形成一个任务序列.要求每到达一个任务,就要即时为之分配一个团队负责完成它.在解决该问题时,必须保证任务分配相对均衡,而且团队成员之间合作沟通的通信开销最小.Anagnostopoulos 等人在文献[1]之前的工作^[29]中,同样以平衡工作负载为优化目标,但是没有对通信开销加以考虑.

3.2.3 指定技能所需人数的团队形成

考虑这样一个实例:某公司手头有一个计划的项目,需要挑一些人员来完成.而该项目的实现需要这个团队包含至少 3 个数据研究员、2 个理论研究员以及 1 个数据挖掘研究员.在一个展示了公司内部所有人员的合作关系的网络关系图之下,公司该如何选拔人员组建这个项目团队?

这样的条件设定使得该问题在文献[4]的基础上得到了推广,这种广义的项目 P 定义为

$$P=(S,N)=\{(s_i,n_i)|\forall i,1\leq i\leq m,s_i\in S,n_i\in N^+\},$$

其中, m 是项目所需技能的总数目.在实际情况中我们可以发现,单项任务需要多人完成的项目是普遍存在的.文献[5,9,10,36]就是考虑了这样的实际需求,对每个技能需要的最少人数提出了限定,以组建一个能够覆盖所有项目要求的技能并且满足每个技能所要求人数的团队.经证明:求解这样一个广义的团队形成问题,其难度大大增加,但却更有意思了^[9].比如,即使假设候选集中的每个人最多只拥有一种技能,不考虑每个人拥有多种不同技能的情况,这个问题也是一个非平凡问题.

文献[5,9,10,36]均对这个变形问题进行了相应的研究.它们之间的最大区别在于各自的出发点不同,采用了不同的优化目标.在文献[9]中,社交关系网络中的边权值代表两个团队成员之间的合作协调度(collaborative

compatibility),采用的优化度量是团队成员之间的合作协调性.而文献[5,11]把团队成员之间合作的通信开销 (communication cost)作为优化目标.在文献[36]中,Amita 等人提出了一个基于密度的团队形成问题,采用最大化团队成员之间的关系密度作为优化目标,使用寻找最大化密度子图的方法来求解这个问题.Amita 等人在文献[36]中提到:在这个变形问题下,如果将问题退化到项目只需要一个技能的特例情况,它仍然是一个 NP 难问题.因此,文中提出了一种 3-近似算法,并对近似度进行了证明.在文献[5,10]中,Chen 等人也考虑到了项目包含的每个技能要求一定数目的人员来完成的实例,定义了一个一般化的团队形成问题(generalized team formation problem).Chen 等人以 Lappas 的文献[4]为参考,提出了两种改进的近似算法:首先,他们对文献[4]中已有的 Enhanced-Steiner 算法进行修改,以满足新问题的单个技能的多人员需要的条件;同时,不像文献[4]中的 Enhanced-Steiner 算法在初始时随机地选择一个节点作为种子开始寻找其他团队成员,改进后的算法使用一种基于密度的种子节点选择策略.此外,Chen 等人还提出了另一种基于群的算法,把拥有相同技能的人合并成一个群(group),然后把原始的社交关系网络图中不同群之间人员的连边合并成两个群之间的连边,形成一个群图(group graph),再在群图上使用改进后的 Enhanced-Steiner 算法.文献[9]中研究的也是这样的广义项目,然而文献[9]中探究的问题不仅仅对项目进行了一般化的拓展,同时还提出对每个人的每项技能进行量化分级.

3.2.4 技能分级的团队形成

技能分级是指每位专家所拥有的技能具有不同的等级,以此来显示该专家在某项技能上的水平高低或掌握程度.这样在选择人员时,在考虑通信开销之余,便会优先选择技能水平高的专家加入团队,最终产生的团队在业务能力上必然是高效的.

每位专家用 $e_i = (s_{e_i}, l_{e_i})$ 表示,其中, s_{e_i} 表示专家 e_i 所拥有的技能集合,向量 l_{e_i} 的每个分量表示专家在 s_{e_i} 上对应每种技能的水平等级.文献[9]使用了这种为技能水平进行等级量化的策略,在文献[4]的 RarestFirst 算法基础上,提出了一种针对广义项目的团队形成算法.在文献[9]中,每项技能都有一个给定的可接受等级 l_i ,即:

$$P=(S,N,L)=\{(s_i,n_i,l_i)|\forall i,1\leq i\leq m,s_i\in S,n_i\in N^+\}.$$

每位拥有技能 s_i 的专家,其对技能 s_i 的等级必须不低于 l_i 才可以参与该项目.专家 i 在技能 s_i 上的等级用下面的公式(1)计算:

$$skillgrading(i,s_i)=\frac{\sum_{j\in N_i}\left(\frac{|P_i^{s_i}\cap P_j^{s_i}|}{|P_i^{s_i}\cup P_j^{s_i}|}\times|P_i^{s_i}-\bar{P}_i|\times|P_j^{s_i}-\bar{P}_j|\right)}{\sqrt{(P_i^{s_i}-\bar{P}_i)^2\times\left(\sum_{j\in N_i}(P_j^{s_i}-\bar{P}_j)^2\right)}} \quad (1)$$

其中, $P_i^{s_i}$ 表示专家 i 在以往的项目经历中参与技能 s_i 的次数, \bar{P}_i 定义为

$$\bar{P}_i=\frac{\sum_{s\in SkillSet(i)}P_i^s}{|SkillSet(i)|} \quad (2)$$

3.2.5 Leader 主导的项目团队形成

在很多现实场景中,总会有一个项目发起人,或者一个项目总负责人作为 leader.然后,由该 leader 聘选若干符合该项目技能要求的专业人员组成一个团队来完成该项目.所以,所谓的 leader 主导的团队形成就是指整个团队由一个负责人来主导项目的开展.Leader 可以参与项目中的技术工作,也可以不参与,他的主要功能是负责团队成员之间的协调组织与信息传递.文献[2,3]研究了在 leader 主导的场景中的团队形成问题.在该场景下,整个团队的通信开销就体现为每个项目成员与 leader 之间的通信代价之和,如图 2 所示.文献[3]针对带 leader 的情景提出了一种近似算法,把每一个候选专家作为试选 leader,为项目中除去 leader 完成后所剩的其他技能选出其余成员,然后计算整个团队的通信开销.通过不断筛选,最后选取其中通信开销最小的 leader 以及由该 leader 选出的其他成员构成一个团队.而在文献[2]中,作者采用了一种基于寻找最大密度子图的方法,结合大量的组合优化的算法寻找最佳团队.同时,文献[2]在普通的团队形成问题中加入了一些更为通用的限制条件,比如预先指定几个既定成员,在这些已有成员的基础上寻找到其他所需的人员来完成所有技能.文献[7]中也采用了这种由一个指定的负责人来挑选其他成员的模式,只不过在算法中需要指定一个最大跳数 h ,在网络图中, h 度以内的

好友才会被选择, h 度以外的人员则认为是关系不够亲密, 不利于合作, 不予加入团队。

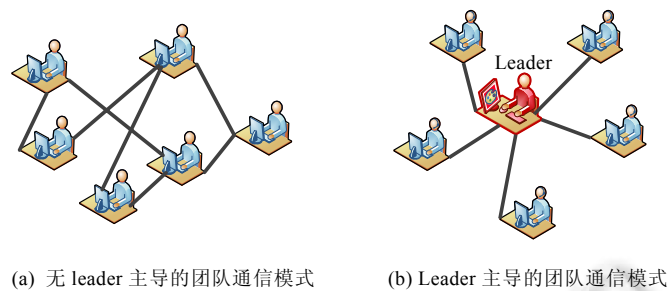


Fig.2 Communication schemas among a team without/with a leader
图 2 无 leader 和有 leader 主导的团队通信示意图

3.2.6 任务限量的团队形成

所谓任务限量的团队形成,指的是为所有的候选专家指定一个整数值,用来限制该专家每个技能所能参加的最大项目个数.这实际上是一种从工作负载合理化出发的一个变形问题.文献[7]首次在社会网络背景下定义了对人员工作负载容量进行限制的团队形成问题,并对该 NP 难问题提出了近似解法.文献[7]中,一个项目由一个用户 v 发起并定义,然后该用户在他的社会网络 $G(V,E,w)$ 中寻找到其他一组成员 $U \subseteq V$, 以使得 $\{v\} \cup U$ 组成的团队能够满足项目的技能需求,同时要满足每个人在项目中承担的任务数目不超过其最大负载量,而且整个团队的通信代价要尽可能地小.

除了以上 6 种类型的变种问题以外,还有其他文献也从不同侧面做了一些工作,为该问题的研究提供了一定的参考和借鉴价值.文献[19,20,24,26]从心理学角度出发,考虑了在合作中人的心理因素影响.其中,文献[19]结合团队组织的多方面因素,提出了一种多层次的团队组织形式(MTSs),文献[38]利用 Q-强化学习算法来研究社会关系网络下的任务流分派问题,文献[12,32]考虑了团队形成中关系网络的动态演变及其产生的影响,都为社会化网络中的团队形成问题提供了启发和思路.

4 优化目标的分类

在团队形成问题中,诸多文献从不同方面进行了探究,产生了多种变形问题.在这些问题中,已有的工作对整个团队的合作效率有不同的定义和度量,相应地,不同的定义和衡量标准就会产生出不同的优化目标.本节对当前工作中已有的优化目标进行归纳,分别对无网络关系和网络关系下的团队效率优化方法进行分类介绍.具体分类见表 2.

Table 2 A classification of optimization methods
表 2 优化方法的分类

无网络关系的团队效率优化	项目收益		文献[18]
	任务平衡		文献[29]
社会网络关系下的团队效率优化	单目标优化	通信开销	文献[3-5,9,36,39]
	双目标优化	通信开销和任务平衡	文献[1,7]
		通信开销和人员支付成本	文献[6,8]

4.1 无网络关系的团队效率优化

对无网络关系的团队形成问题的研究,主要有两类优化目标来对团队的合作效率进行优化,分别是以文献[18]为代表的项目收益优化和以文献[29]为代表的单人负载优化.

- 项目收益优化

在文献[18]中定义的团队形成问题是面向一组项目 \mathcal{P} 选出一个专家团队 \mathcal{T} , 并实现从该团队所有成员到这

多个项目的分配.在分配过程中,每个专家成员 $X \in \mathcal{X}$ 都有自己的薪酬要求 $C(X)$,每个项目完成之后有各自的固定收益 $F(P)$.同时,在组建团队时还有对雇员薪水的支出总预算 B .文献[18]定义的基于团队收益的 ClusterHire 问题即为:在以上条件的设定下,从候选专家集 \mathcal{X} 中寻找一个团队 \mathcal{T} ,使得 $F(\text{Cov}(\mathcal{T}))$ 最大且 $C(\mathcal{T}) \leq B$,其中,

$$F(\text{Cov}(\mathcal{T})) = \sum_{P \in \text{Cov}(\mathcal{T})} F(P) \quad (3)$$

$$\text{Cov}(\mathcal{T}) = \{P \in \mathcal{P} | \mathcal{T} \text{ covers } P\} \quad (4)$$

$$C(\mathcal{T}) = \sum_{X \in \mathcal{T}} C(X) \quad (5)$$

在文献[18]中,这种以团队收益为团队效率优化目标的方法不考虑团队专家之间的社会关系网络,而是着眼于团队在一定资金预算内可以完成的项目所带来的收益.

- 负载均衡

在文献[29]中,作者充分考虑了一个团队中各个人员的任务分配的公平性,定义了一种多项目的任务平衡覆盖(balanced task covering)模型.所谓的任务平衡覆盖,是指一个团队在进行任务分配时,不仅要保证团队成员所具备的技能要能覆盖项目所需的所有技能,同时必须考虑到任务分配的公平性,即:不能让一些人工作量过少而另一些人工作量过大,要尽可能地均衡分配.

用 $P^j = (P_1^j, P_2^j, \dots, P_m^j)$ 表示第 j 个人所具有的技能,其中,该向量的每一维的值都是一个二元变量. $P_i^j = 1$ 表示第 j 个人拥有第 i 项技能, $P_i^j = 0$ 则表示不拥有;用 $J^j = (J_1^j, J_2^j, \dots, J_m^j)$ 表示第 j 个项目是否需要技能 $i(i=1, 2, \dots, m)$,同样地,每一维都是一个二元变量1或0,分别表示项目 j 是否需要第 i 个技能.同时,如果人员 P^j 被分配到了第 i 个项目 J^i ,则用 $X_{ji}=1$ 标示,否则用 $X_{ji}=0$ 标示.那么,上述的任务平衡覆盖问题则可以用如下的线性规划形式来表示.

$$\min L \quad (6)$$

$$\sum_{j=1}^n P_\ell^j X_{ji} \geq J_\ell^i, \forall i=1, \dots, k, \ell=1, \dots, m \quad (7)$$

$$\sum_{i=1}^k X_{ji} \leq L, \forall j=1, \dots, n \quad (8)$$

$$L \geq 0, X_{ji} \in \{0, 1\} \quad (9)$$

其中, k 是项目个数, m 是技能的总个数, n 是人员的总数目.约束(7)保证了所有项目的技能都被满足;约束(8)则限制的是每一个团队成员参与的项目个数,即,工作负载不超过限值 L ; L 非负(如公式(9)所示).

4.2 社会关系网络中的团队效率优化

与无网络关系的团队形成不同,社会网络中的团队形成由于有关系网络的存在,因此在研究中产生了关于团队通信协作效率的优化.本小节对社会关系网络中的团队效率优化方法进行介绍,主要包括团队通信开销的单目标优化方法、同时对通信开销和负载均衡以及通信开销和人员支付成本进行优化的两种双目标优化方法.

4.2.1 通信开销优化

如果把一个团队所有成员放到一个关系网络图中,那么这两个成员能否有效合作的一个决定性因素就是他们的通信开销.因此,在对网络关系下的团队形成的研究中,对通信开销的优化占到了大多数^[3-5,9,36,39].文献[3-5,9]中,所有的候选专家之间形成一个无向带权的社交关系网络图.图中每个节点代表一位专家,两个节点之间有连边即表示两位专家之前有过合作关系,边上的权值代表对应的两位专家沟通的通信代价,权值越小表示越容易合作,反之亦然.两位专家之间的通信代价需要用两个人之间的距离来衡量.这个距离只是一个宽泛的概念,在实际问题中可以根据具体问题选择合适的标准来定义两个专家之间的距离.比如,可以定义为两个人之间以往的互动频率、地理位置的距离、协作的融洽性以及公司同事之间的组织层次关系等.在已有文献中,大都采用模糊的“距离”定义来描述团队形成中的通信代价,而并不明确限定用互动频率、地理位置距离还是公司组织层次关系来度量.这个距离标准的选择要根据具体应用场景和具体问题来确定,有时可能需要综合多个因素

来确定.但是在构造实验数据时,往往会根据专家们以往的合作情况来确定两个人之间的通信代价.

在定义好专家之间的沟通代价之后,需要建模一个团队的通信模型.文献[39]基于组织行为学中3种基本的组织沟通模型——链式沟通模型、轮式沟通模型和全通道式沟通模型,提出3种组内沟通代价和2种组间沟通代价计算方法.

在文献[4]中,作者提出了两种衡量团队通信开销的实例:子图直径 Cc-R 和最小生成树 MST.Cc-R 把构成团队的专家成员所形成的关系子图的直径,即,任意两个节点间的最短路径的最大值作为整个团队通信代价的度量.最小斯坦纳树则是通过构建子图的最小生成树,用这棵树上所有边的权值之和作为团队的通信开销.

然而文献[3]指出:文献[4]中的这两种通信代价度量方法在一些场合存在着局限性,不能很好地评价整个团队协作过程中的通信开销.文献[3]认为,文献[4]中的子图直径只是把任意成对节点间的最短路径的最大值作为团队的通信衡量.然而,如果项目足够复杂,可能会需要任意两种技能之间都要有“对话”.这时候,子图直径和最小生成树都不能包含所有可能的通信.基于这个考虑,文献[3]提出了一种通信代价函数 $sumDistance$.

$$sumDistance = \sum_{i=1}^m \sum_{j=i+1}^m d(c_{s_i}, c_{s_j}) \quad (10)$$

即,任意一组技能负责人之间的最短距离之和.公式(10)中, c_{s_i} 代表技能 s_i 的负责人(即:在团队最后的分工中,技能 s_i 由 c_{s_i} 来负责完成).同时,文献[3]还提出了一种 leader 主导的团队形成模式,其对应的通信代价函数为 $leaderDistance$.

$$leaderDistance = \sum_{i=1}^p d(c_{s_i}, L) \quad (11)$$

其中, L 是团队的 leader.

在文献[5]中,作者使用文献[4]中的斯坦纳树的方法优化通信开销.文献[9]则是对子图直径的度量方法进行改进,增加了对技能等级因子的考虑.需要指出的是,文献[39]中提出的基于3种基本组织沟通模型的组内通信代价计算方法分别就是上述的 MST, $leaderDistance$ 和 $sumDistance$.

在文献[36]中,候选专家的社交关系网络图采用两个节点对应的专家的合作内聚力(cohesiveness)作为边上的权值.权值越大,合作越容易.对应地,团队质量的优化目标就是最大化整个团队的内聚力.文献[36]采用寻找基于最大化密度子图的方法来为一个项目构建最优团队.一个子图 G 的密度采用 $d(G) = \frac{W(G)}{|V(G)|}$ 的方法来计算.

4.2.2 通信开销和负载均衡优化

除了众多优化通信开销的研究工作,也有不少文献试图加入其他优化目标来完善团队形成问题的研究.其中,文献[1,7]在优化通信开销的基础上还实现了对团队负载均衡的优化.然而,解决双目标优化问题是比较困难的.通常情况下,有一种解决方法是把一个目标函数作为约束条件,去优化另一个目标函数.文献[1]是为多项目的在线 workflow 分派团队.由于所要完成的项目任务是在线不断实时到来的,所以预先无法确定项目的数目,因此无法确定一个合理的单人工作负载限额.在文献[1]中,作者采用的方法是:给定一个通信代价的限额 B ,把通信开销函数作为约束条件,去优化最大工作负载.上述优化目标用公式表示如下:

$$\left. \begin{aligned} \min \max_{p^i \in P} L(p^i), i = 1, \dots, n, \\ cov(J^j, q^j) = 1, j = 1, \dots, m, \\ c(Q^j) \leq B, j = 1, \dots, m \end{aligned} \right\} \quad (12)$$

其中, m 是项目的总个数, n 是候选人员的总数目. $L(p^i)$ 表示第 i 个候选专家的工作负载,即参与的项目个数.约束条件 $cov(J^j, q^j) = 1$ 保证每个项目 J^j 所需技能都能被分派给团队 Q^j 所覆盖. $c(Q^j)$ 表示分派给第 j 个项目的团队 Q^j 的通信开销.

文献[7]提出一种 capacitated 的团队形成模式.所谓 capacitated 团队形成,指的是为所有的候选专家指定一个整数,用来限制该专家的技能最多能参与的项目个数.这实际上也是一种从工作负载合理化出发的变形问题.文献[7]首次提出该问题,并进行了探究.该文在问题的定义中就考虑了负载均衡,对人员负载进行隐式优化.

然后只需把通信开销作为显式优化目标,就可以实现二者兼顾的双目标优化.

4.2.3 通信开销和人员支付成本

文献[6,8]实现的是对通信开销和人员支付成本的双目标优化.给定一个候选专家间的社会关系网络图,每位专家由一个带权的节点表示,权值代表雇佣该专家所要支付的薪酬金额.节点间的边表示对应的两位专家可以进行合作,且通信成本为边上的权值.文献[6]提出了多种近似算法,分别把通信开销和人员薪酬成本作为约束条件,去优化另一个目标.算法中分别采用了子图直径和公式(10)中的 $sumDistance$ 两种通信开销函数.如果用 $t(c_i)$ 表示专家 c_i 的人力成本,则团队的人员支付成本函数为

$$PCost(T) = \sum_{i=1}^p t(c_i) \quad (13)$$

Kargar 等人在继文献[3]的工作之后,在文献[8]中提出一种既要满足通信开销最小化,同时又考虑人员支付成本最低的双目标优化的团队形成问题.他们采用的方法是:把两个目标函数组合成一个目标函数,进而转化成单目标优化问题.通信开销函数使用他们之前在文献[3]中提出的 $sumDistance$,即团队 T 的通信代价 $SD_G(T) = \sum_{i=1}^m \sum_{j=i+1}^m d(c_{s_i}, c_{s_j})$. 团队的人员支付成本为 $PC_G(T) = \sum_{i=1}^m t(c_{s_i})$, 其中, $t(c_{s_i})$ 为完成技能 s_i 的专家所需的薪酬成本.在确定了通信开销和人力成本函数之后,文献[8]将二者组合成一个新的函数作为优化目标,这种综合代价用公式(14)进行计算.

$$ComCost_G(T) = (m-1)(1-\lambda) \times PC_G(T) + 2\lambda \times SD_G(T) \quad (14)$$

其中, m 为技能总数目; λ 是一个 0~1 之间的实数,代表通信代价 $SD_G(T)$ 和人力成本 $PC_G(T)$ 之间的一个侧重权值.值得注意的是:文献[6]和文献[8]的团队人力成本函数的计算方式是有差异的:由于有的专家可能会在项目中负责不止一种技能,在文献[6]中,这些专家的人力成本只会被算作一份;而在文献[8]中,由于是按技能逐项计算成本的,因此这些专家就会得到多份酬金.

5 实验数据集与评价标准

5.1 数据集

在社会化网络的团队形成问题的研究中,学者们使用了不同的数据集来对算法进行实验验证与分析.主要的数据集和其相关论文见表 3.

Table 3 Experimental data sets

表 3 实验数据集

数据集	文献
DBLP	[2-9,36]
IMDb	[1,3,6,8,29]
Bibsonomy	[1,29]
GitHub	[7]
Flickr	[29]

DBLP 数据是被采用最多的数据集,文献[2-9,36]用 DBLP 数据库(<http://dblp.uni-trier.de/xml>)中的合著关系构造合作关系网络,网络中的专家节点由那些至少合作发表过一定数目论文的作者组成,边则根据专家之间的合著记录形成.在实验中,只收集发表在数据库(DB)、数据挖掘(DM)、人工智能(AI)和理论研究(T)这 4 个领域的 19 项会议中的论文,其分类目录如下:DB={SIGMOD,VLDB,ICDE,ICDT,EDBT,PODS},DM={WWW,KDD,SDM,PKDD,ICDM},AI={ICML,ECML,COLT,UAI},T={SODA,FOCS,STOC,STACS}.数据库中的每篇文献都有其作者姓名、标题、发表的会议名称和发表的日期等信息.团队形成问题中的一个重要的输入数据就是专家 i 的技能集合 S_i .在实验中,专家 i 的技能集合 S_i 由该专家的论文发表的会议所属的领域来确定.很多文献的实验设置中都把专家 i 发表过至少一定数目的领域集作为该专家的技能集合.网络图中连接任意两个节点 i_1, i_2 的边上的权值用 Jaccard 距离来计算.

$$w(i_1, i_2) = 1 - \frac{|P_{i_1} \cap P_{i_2}|}{|P_{i_1} \cup P_{i_2}|},$$

其中, P_i 表示专家 i 合著发表的论文集合.

除 DBLP 之外, IMDb 数据库也被广泛使用^[1,3,6,8,29]. IMDb 即互联网电影数据库(the Internet movie database), 是一个关于电影、电影演员、电影导演、电视节目、电视明星、电子游戏和电影制作小组的在线数据库. 它是目前全球互联网中最大的一个电影资料库, 里面包括了几乎所有的电影以及 1982 年以后的电视剧集. IMDb (<http://www.imdb.com/interfaces>) 数据资料中包括了影片的诸多信息, 如演员、片长、内容介绍、分级、评论等. 文献[3,6,8]在实验中使用 IMDb 数据库中的信息来构建演员之间的合作关系网络图. 这些文献采用的方法是在特定年限内参演过一定数目影片的演员作为专家节点, 对有过一定数目的合作记录的两个节点构建一条边, 边上的权值与 DBLP 一样, 按照 Jaccard 距离计算. 与文献[3,6,8]不同的是, 文献[1,29]把数据库中的影视工作人员分为导演和演员两大类, 通过 IMDb 的数据来构建导演之间的关系网络图: 如果两位导演指导(合作)过至少两位共同的演员, 则这两位导演节点之间建立一条边. 其边上的权值设为 e^{-rD} , 其中, D 代表与这两位导演都合作过的演员数目. 在实验中, 用每位演员或者导演参演或执导过的电影的流派作为每位专家的技能集合, 比如喜剧片、爱情片、犯罪片、传记片、宗教片等, 即为技能.

文献[1,29]使用 Bibsonomy 数据进行实验. Bibsonomy 是一个面向公众的书签和文献标记系统, 用户可以对书籍和出版文献资料打标签, 然后进行分享推荐. 该数据库中包含了大量的与计算机科学相关的书籍和出版资料, 包括论文、指南教程、博客帖子等, 数据库中记录了每份刊物的作者集合^[40,41]. 文献[1,29]将数据库中的作者作为专家节点, 边则通过作者之间的合著关系进行构建, 即: 如果两位作者有过合著记录, 则构建一条边. 在文献[1]中, 构建的网络图中的边上的权值设置为 e^{-rD} , 其中, D 代表两位作者共同发表的刊物的数目, r 设为 0.1. 每位作者的技能集合由 Bibsonomy 系统中用户对其发表的刊物所给出的标签集来确定.

社会关系网络中的团队形成问题主要使用以上 3 个数据库进行实验数据集的构造. 除此之外, 文献[7]还使用了 GitHub 数据集, 文献[29]使用 Flickr 数据集进行实验分析.

5.2 评价标准

实验中对算法的评价指标主要有以下 4 种.

- 通信代价: 各种通信代价函数, 包括直径、最小斯坦纳树、*sumDistance*、*leaderDistance* 等, 在很多论文^[2-9,36]中被用来衡量算法得到的团队的性能. 越小的通信代价, 意味着所得到的团队合作效率越高;
- 团队规模: 即团队成员的数目. 整个团队的通信代价正比于团队规模^[4], 较大的团队由于成员人数较多, 则意味着各个技能内部沟通较多, 技能之间的工作交涉也较繁杂, 因此会产生较大的通信代价. 同理, 较小的团队对应的通信代价则较小. 同时, 从人力成本方面考虑, 团队人数越少, 支付成本越低. 文献[2-5,9,36]在实验评估中均采用了这一评价标准;
- 团队合著量: 指的是在 DBLP 数据集下, 由团队成员中至少两人合作完成的文献的数量. 显然, 一个团队有越多的以往合作的经历, 其在当前项目中有效合作的系数就越大, 该团队的质量也就越优^[3,5];
- 技能频数(skill count): 在文献[3,5]中提出, 用来衡量一个项目里所有技能在其团队专家所著文献中出现的平均频率. 技能 s_i 的频数 $SC_{s_i} = \frac{1}{|\mathcal{T}|} \sum_{c_i \in \mathcal{T}} N_{c_i}(s_i)$, 其中, $N_{c_i}(s_i)$ 表示团队中专家 c_i 发表的文献中出现的频数.

技能 s_i 的频数. 项目对应的团队 \mathcal{T} 的技能频数定义为 $SC = \frac{1}{m} \sum_{i=1}^m SC_{s_i}$, m 为项目所需的技能总数. 该指标

的值越高, 就意味着形成的团队对项目要求的技能比较擅长, 团队整体业务能力与该项目比较对口.

实验中的项目是按照以下方法构造的: 每个项目指定所需的技能数目, 对每个技能数目 t 随机生成 n 个需要 t 项技能的项目. 变量 t 依情况而定, 基本上在 2~20 之间取值. n 的取值在不同论文中也不尽相同, 有的取 100^[3-6], 有的取 10^[2], 也有的取适中值 50^[8]. 在实验分析中, 使用不同的算法计算 t 对应的 n 个项目的某个评价指标的平均值, 在不同的 t 值、不同的算法之间进行比较.

6 结论与展望

本文在认真分析归纳的基础上,从团队形成问题的多种变形和优化方法以及研究中的主要实验数据集与实验设计策略 3 大方面对当前社会化网络中的团队形成问题进行了细致的总结综述.在当前已有的研究中,该问题的变形问题主要包括单项目的、多项目的、指定每个技能需要的人员数目的团队形成、技能分级的团队形成、Leader 主导的项目团队形成和任务限量的团队形成.在这些问题中,优化目标主要分为以下 4 类:(a) 通信开销;(b) 任务负载;(c) 人力成本;(d) 项目收益.

社会网络中的团队形成问题虽然已经有了比较深入的研究,取得了一定的成果,但是仍然面临着不少困难与挑战.

- (1) 专家技能信息的获取.在已有研究中,大多数都是通过 DBLP 数据库进行实验验证的,因此,每位专家的技能集合可以通过其文献发表记录来抽取.然而,在其他场景中,如何获取一个人的技能集是一个很有挑战的信息获取问题;
- (2) 如何衡量两个人之间的通信代价.两个人之间的关系远近受到很多方面因素的影响,如地理位置远近、人际交往中的熟识度以及学科领域相似性等.如何选取一个合理的度量来描述两个人之间的通信代价;
- (3) 社会关系网络的重构.在庞大的全局社会网络中,我们需要除去非候选人节点,只留下候选专家节点.然后从原全局网络中判断这些候选专家之间是否直接可达,或是间接几步可达来得到他们之间的关系强度值,从而建立他们之间的带权边.该方法只是理想状态下的解决方法.但是如果不存在一个包含了所有候选专家的全局网络,而这些专家分散于多个关系网络,那又该如何重构出候选专家之间的局部关系网络?
- (4) 一些变形问题的研究还很不成熟,比如技能分级的团队形成,进而引出对专家进行评级打分机制的讨论^[14,15],因此,需要进一步对问题和方法进行完善.

从当前的研究现状来看,未来的团队形成问题研究工作还有继续开展的空间.由于社会化网络是会随时间演化的,因此以后的工作中,可以考虑动态网络中的团队形成.同时我们考虑到,一个项目团队如果发生突发状况需要临时替换专家,或者由于需求的变动需要临时增加技能补进人员.在处理这些状况时,可能会引起当前已经构建好的团队的人员变动.所以,这一情况也是值得未来加以研究的.此外,我们在研究中还发现,近年来,随着 Meetup,Plancast 等平台而出现的社交事件组织(social event organization)问题^[42-48]与团队形成问题颇为相似:前者是为一组人安排活动,而后者是为一组技能分配人;还有本文第 3.2.5 节介绍的 leader 主导的团队形成问题也可以通过简单转化,建模成一个社团搜索问题^[49-55](见第 1 节).故而在未来的研究中,这 3 类问题可以互相借鉴,甚至合并研究.最后值得一提的是:借助国内外有影响力的 Freelancer,oDesk 等在线服务外包网站的强劲发展势头,上述挑战将会在相关技术的带动下得到相应的解决.

References:

- [1] Anagnostopoulos A, Becchetti L, Castillo C, Gionis A, Leonardi S. Online team formation in social networks. In: Mille A, Gandon F, Misselis J, eds. Proc. of the 21st Int'l Conf. on World Wide Web. New York: ACM Press, 2012. 839-848. [doi: 10.1145/2187836.2187950]
- [2] Rangapuram SS, Bühler T, Hein M. Towards realistic team formation in social networks based on densest subgraphs. In: Schwabe D, Almeida V, Glaser H, eds. Proc. of the 22nd Int'l Conf. on World Wide Web. New York: ACM Press, 2013. 1077-1088. [doi: 10.1145/2488388.2488482]
- [3] Kargar M, An A. Discovering top- k teams of experts with/without a leader in social networks. In: Berendt B, Vries AD, Fan W, eds. Proc. of the 20th ACM Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2011. 985-994. [doi: 10.1145/2063576.2063718]
- [4] Lappas T, Liu K, Terzi E. Finding a team of experts in social networks. In: Mille A, Gandon F, Misselis J, eds. Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2009. 467-476. [doi: 10.1145/1557019.1557074]

- [5] Li CT, Shan MK, Lin SD. On team formation with expertise query in collaborative social networks. *Knowledge and Information Systems*, 2015,42(2):441–463. [doi: 10.1007/s10115-013-0695-x]
- [6] Kargar M, Zihayat M, An A. Finding affordable and collaborative teams from a network of experts. In: *Proc. of the SIAM Int'l Conf. on Data Mining (SDM)*. 2013. 587–595. [doi: 10.1137/1.9781611972832.65]
- [7] Majumder A, Datta S, Naidu KVM. Capacitated team formation problem on social networks. In: Schaik BV, ed. *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2012. 1005–1013. [doi: 10.1145/2339530.2339690]
- [8] Kargar M, An A, Zihayat M. Efficient bi-objective team formation in social networks. In: Flach PA, Bie T, Cristianini N, eds. *Proc. of the 2012 European Conf. on Machine Learning & Knowledge Discovery in Databases*. Berlin: Springer-Verlag, 2012. 483–498. [doi: 10.1007/978-3-642-33486-3_31]
- [9] Farhadi F, Sorkhi M, Hashemi S, Hamzeh A. An effective expert team formation in social networks based on skill grading. In: Tchicaya A, Lorentz N, eds. *Proc. of the 2011 IEEE 11th Int'l Conf. on Data Mining Workshops (ICDMW)*. New York: IEEE, 2011. 366–372. [doi: 10.1109/ICDMW.2011.28]
- [10] Li CT, Shan MK. Team formation for generalized tasks in expertise social networks. In: Adlarson P, Amarian M, Bashkanov M, eds. *Proc. of the 2010 IEEE 2nd Int'l Conf. on Social Computing (SocialCom)*. New York: IEEE, 2010. 9–16. [doi: 10.1109/SocialCom.2010.12]
- [11] Liemhetcharat S, Veloso M. Weighted synergy graphs for effective team formation with heterogeneous ad hoc agents. *Artificial Intelligence*, 2014,208:41–65. [doi: 10.1016/j.artint.2013.12.002]
- [12] Awal GK, Bharadwaj KK. Team formation in social networks based on collective intelligence—An evolutionary approach. *Applied Intelligence*, 2014,41(2):627–648. [doi: 10.1007/s10489-014-0528-y]
- [13] Durfee EH, Lesser VR. Partial global planning: A coordination framework for distributed hypothesis formation. *IEEE Trans. on Systems Man & Cybernetics*, 1991,21(5):1167–1183. [doi: 10.1109/21.120067]
- [14] Hong Y, Pavlou PA. Online labor markets: An informal freelancer economy. *IBIT Report, Online Labor Markets: An Informal Economy*, 2013. [doi: 10.2139/ssrn.2132869]
- [15] Kokkodis M, Papadimitriou P, Ipeirotis PG. Hiring behavior models for online labor markets. In: Cheng X, Li H, Tang J, eds. *Proc. of the 8th ACM Int'l Conf. on Web Search and Data Mining*. New York: ACM Press, 2015. 223–232. [doi: 10.1145/2684822.2685299]
- [16] Mason W, Watts DJ. Financial incentives and the performance of crowds. *ACM SIGKDD Explorations Newsletter*, 2010,11(2):100–108. [doi: 10.1145/1809400.1809422]
- [17] Pallais A. Inefficient hiring in entry-level labor markets. *American Economic Review*, 2012,104(11):3565–3599. [doi: 10.2139/ssrn.2012131]
- [18] Golshan B, Lappas T, Terzi E. Profit-Maximizing cluster hires. In: *Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2014. 1196–1205. [doi: 10.1145/2623330.2623690]
- [19] Contractor N. Some assembly required: Leveraging Web science to understand and enable team assembly. *Philosophical Trans. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 2013,371(1987):20120385. [doi: 10.1098/rsta.2012.0385]
- [20] Cummings JN, Kiesler S. Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 2007,36(10):1620–1634. [doi: 10.1016/j.respol.2007.09.001]
- [21] Krumke SO, Thielen C. The generalized assignment problem with minimum quantities. *European Journal of Operational Research*, 2013,228(1):46–55. [doi: 10.1016/j.ejor.2013.01.027]
- [22] Sozio M, Gionis A. The community-search problem and how to plan a successful cocktail party. In: Rao B, Krishnapuram B, Tomkins A, eds. *Proc. of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2010. 939–948. [doi: 10.1145/1835804.1835923]
- [23] Li RH, Qin L, Yu JX, Mao R. Influential community search in large networks. *Proc. of the VLDB Endowment*, 2015,8(5):509–520. [doi: 10.14778/2735479.2735484]
- [24] Chen SJG, Lin L. Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *IEEE Trans. on Engineering Management*, 2004,51(2):111–124. [doi: 10.1109/TEM.2004.826011]
- [25] Zzkarian A, Kusiak A. Forming teams: An analytical approach. *IIE Trans.*, 1999,31(1):85–97. [doi: 10.1023/A:1007580823003]
- [26] Fitzpatrick EL, Askin RG. Forming effective worker teams with multi-functional skill requirements. *Computers & Industrial Engineering*, 2005,48(3):593–608. [doi: 10.1016/j.cie.2004.12.014]

- [27] Agustin-Blas LE, Salcedo-Sanz S, Ortiz-Garcia EG, Portilla-Figueras A, Perez-Bellido AM, Jimenez-Fernandez S. Team formation based on group technology: A hybrid grouping genetic algorithm approach. *Computers & Operations Research*, 2011,38(2): 484–495. [doi: 10.1016/j.cor.2010.07.006]
- [28] Joshi MV, Han EHS, Karypis G, Kumar V. Efficient parallel algorithms for mining associations. In: Zaki MJ, Ho CT, eds. *Proc. of the Large-Scale Parallel Data Mining 2002*. New York: Springer-Verlag, 2002. 83–126. [doi: 10.1007/3-540-46502-2_5]
- [29] Anagnostopoulos A, Becchetti L, Castillo C, Gionis A, Leonardi S. Power in unity: Forming teams in large-scale community systems. In: Huang J, ed. *Proc. of the 19th ACM Int'l Conf. on Information and Knowledge Management*. New York: ACM Press, 2010. 599–608. [doi: 10.1145/1871437.1871515]
- [30] Skopik F, Truong HL, Dustdar S. *Trust and Reputation Mining in Professional Virtual Communities*. Berlin, Heidelberg: Springer-Verlag, 2009. 76–90. [doi: 10.1007/978-3-642-02818-2_6]
- [31] Granovetter M. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1983,1(6):201–233. [doi: 10.2307/202051]
- [32] Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group formation in large social networks: Membership, growth, and evolution. In: Chairungar G, Lyle, Chaircraven P, eds. *Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2006. 44–54. [doi: 10.1145/1150402.1150412]
- [33] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks. In: Rappa M, Jones P, Freire J, eds. *Proc. of the 19th Int'l Conf. on World Wide Web*. New York: ACM Press, 2010. 641–650. [doi: 10.1145/1772690.1772756]
- [34] Xiang R, Neville J, Rogati M. Modeling relationship strength in online social networks. In: Rappa M, Jones P, Freire J, eds. *Proc. of the 19th Int'l Conf. on World Wide Web*. New York: ACM Press, 2010. 981–990. [doi: 10.1145/1772690.1772790]
- [35] Agarwal V, Bharadwaj KK. A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. *Social Network Analysis and Mining*, 2013,3(3):359–379. [doi: 10.1007/s13278-012-0083-7]
- [36] Gajewar A, Sarma AD. Multi-Skill collaborative teams based on densest subgraphs. In: *Proc. of the Computer Science*. 2011. 1077–1088. [doi: 10.1137/1.9781611972825.15]
- [37] Arkin EM, Hassinz R. Minimum diameter covering problems. *Networks*, 2000,36(3):147–155. [doi: 10.1002/1097-0037(200010)36:3<147::AID-NET1>3.0.CO;2-M]
- [38] Yu Y, Wang Y, Liu XM, Chen J. Workflow task assignment strategy based on social context. *Ruan Jian Xue Bao/Journal of Software*, 2015,26(3):562–573 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4766.htm> [doi: 10.13328/j.cnki.jos.004766]
- [39] Sun H, Jin M, Liu J, Yu G. Methods for team formation problem with grouping task in social networks. *Journal of Computer Research and Development*, 2015,52(11):2535–2544 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2015.20148136]
- [40] Jäschke R, Hotho A, Schmitz C, Stumme G. Analysis of the Publication Sharing Behaviour in BibSonomy. In: *Conceptual Structures: Knowledge Architectures for Smart Applications*. 2007. 283–295. [doi: 10.1007/978-3-540-73681-3_21]
- [41] Benz D, Hotho A, Jäschke R, Krause B, Mitzlaff F. The social bookmark and publication management system bibsonomy. *The VLDB Journal—The Int'l Journal on Very Large Data Bases*, 2010,19(6):849–875. [doi: 10.1007/s00778-010-0208-4]
- [42] Li K, Lu W, Bhagat S, Bhagat S, Lakshmanan LVS. On social event organization. In: *Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2014. 1206–1215. [doi: 10.1145/2623330.2623724]
- [43] She J, Tong Y, Chen L. Utility-Aware social event-participant planning. In: Sellis TK, Davidson SB, eds. *Proc. of the 2015 ACM SIGMOD Int'l Conf. on Management of Data*. New York: ACM Press, 2015. 1629–1643. [doi: 10.1145/2723372.2749446]
- [44] Shuai HH, Yang DN, Yu PS, Chen MS. Willingness optimization for social group activity. *Proc. of the VLDB Endowment*, 2013, 7(4):253–264. [doi: 10.14778/2732240.2732244]
- [45] Shuai HH, Yang DN, Yu PS, Chen MS. Scale-Adaptive group optimization for social activity planning. In: *Proc. of the Computer Science*. 2015. 45–57. [doi: 10.1007/978-3-319-18038-0_4]
- [46] Feng K, Cong G, Bhowmick SS, Ma S. In search of influential event organizers in online social networks. In: Dyreson C, Li F, Özsü MT, eds. *Proc. of the 2014 ACM SIGMOD Int'l Conf. on Management of Data*. New York: ACM Press, 2014. 63–74. [doi: 10.1145/2588555.2612173]
- [47] Shuai HH, Yang DN, Yu PS, Chen MS. A comprehensive study on willingness maximization for social activity planning with quality guarantee. *IEEE Trans. on Knowledge and Data Engineering*, 2016,28(1):2–16. [doi: 10.1109/TKDE.2015.2468728]

- [48] Guo B, Yu Z, Chen L, Zhou X. MobiGroup: Enabling lifecycle support to social activity organization and suggestion with mobile crowd sensing. *IEEE Trans. on Human-Machine Systems*, 2015. 1–13. [doi: 10.1109/THMS.2015.2503290]
- [49] Seo J, Croft WB, Smith DA. Online community search using conversational structures. *Information Retrieval*, 2011,14(6):547–571. [doi: 10.1007/s10791-011-9166-8]
- [50] Huang X, Lakshmanan LVS, Yu JX, Cheng H. Approximate closest community search in networks. *Proc. of the VLDB Endowment*, 2015,9(4):276–287. [doi: 10.14778/2856318.2856323]
- [51] Barbieri N, Bonchi F, Galimberti E, Gullo F. Efficient and effective community search. *Data Mining and Knowledge Discovery*, 2015,29(5):1406–1433. [doi: 10.1007/s10618-015-0422-1]
- [52] Cui W, Xiao Y, Wang H, Wang W. Local search of communities in large graphs. In: Dyreson C, Li F, Özsu MT, eds. *Proc. of the 2014 ACM SIGMOD Int'l Conf. on Management of Data*. New York: ACM Press, 2014. 991–1002. [doi: 10.1145/2588555.2612179]
- [53] Tong H, Faloutsos C. Center-Piece subgraphs: Problem definition and fast solutions. In: Chairungar G, Chaircraven P, Chairgunopulos P, eds. *Proc. of the 12th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2006. 404–413. [doi: 10.1145/1150402.1150448]
- [54] Kim SY, Choi DW, Chung CW. Finding a friendly community in social networks considering bad relationships. *The Computer Journal*, 2015,58(6):1469–1481. [doi: 10.1093/comjnl/bxu092]
- [55] Wang L, Lou T, Tang J, Hopcroft JE. Detecting community kernels in large social networks. In: Tchicaya A, Lorentz N, eds. *Proc. of the 2011 IEEE 11th Int'l Conf. on Data Mining (ICDM)*. New York: IEEE, 2011. 784–793. [doi: 10.1109/ICDM.2011.48]

附中文参考文献:

- [38] 余阳,王颖,刘醒梅,陈健.基于社会关系的工作流任务分派策略研究. *软件学报*,2015,26(3):562–573. <http://www.jos.org.cn/1000-9825/4766.htm> [doi: 10.13328/j.cnki.jos.004766]
- [39] 孙焕良,金铭宇,刘俊岭,于戈.社会上支持任务分组的团队形成方法. *计算机研究与发展*,2015,52(11):2535–2544. [doi: 10.7544/issn1000-1239.2015.20148136]



黄健斌(1975—),男,湖北随州人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据挖掘与知识发现,大数据分析与管理.



吕泽(1992—),男,硕士生,主要研究领域为社会网络分析与团队形成.



孙晓晶(1992—),女,硕士生,主要研究领域为数据挖掘,社交网络分析.



孙鹤立(1983—),女,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为社会网络团队形成.



周瑜(1985—),男,博士生,主要研究领域为异质信息网络分析.



贾晓琳(1963—),女,博士,高级工程师,主要研究领域为数据挖掘,知识发现.