

考虑社会关系的评委选择方法

背景动机简介

随着收到的期刊和会议数量的发展。分配适当的审稿人来审查它们对编辑来说变得很繁重。已经提出了许多方法来帮助解决这个问题，但是几乎所有方法都关注于评论者和论文之间的专业适应性，而忽略了论文作者和被分配审阅它的专家之间的社会关系。但是，这可能会严重影响审查结果的公平性。因此，我们尝试设计一种算法，根据专业适应性，他们与作者的社会关系和能力限制，即评论者可以检查的最大论文数量，自动分配最合适的评论者。

问题定义

给定论文集合和评审人集合，我们的目标是为每一篇论文分配一定数目的评审人，满足

1. 每一篇论文都要被分配足够数量的评审人
2. 每位评审人不能被分配超过能力上限的论文
3. 论文的研究问题与评审人的研究领域尽量相近
4. 评审人之间和评审人与论文作者之间应当没有直接联系，同时社会关系应当尽量弱

我们统一使用 P 代表论文集合， R 代表评审人集合， A 代表所有作者的集合， I 代表所有研究领域和研究问题的集合。对于 P 中的每一篇论文 $p \in P$ ，我们指定其研究领域为 $I_p = \{I_1^p, I_2^p, \dots, I_k^p\}$ （也可以将 I_p 理解为一个 $|I|$ 维的向量，其中 $I_i^p = 1$ 代表论文 p 研究问题涉及领域 I_i ）。同时我们将 p 的作者集合表示为 $A^p = \{A_1^p, A_2^p, \dots, A_k^p\}$ 。同样的，对于 R 中的每个评审人 $r \in R$ ，我们将其研究领域表示为 $I^r = \{I_1^r, I_2^r, \dots, I_k^r\}$ ，对评审人 j ，我们使用 l_j 表示其可以评审的文章数目上限。

我们可以将最后的评委分配结果表示成矩阵 $M \in \mathbb{R}^{|P| \times |R|}$ ，其中 $|P|$ 是论文总数， $|R|$ 是评委总数。对矩阵 M 中的每一个 m_{ij} 元素， $m_{ij} = 1$ 表示将评审 j 分配给论文 i 。我们使用 M^i 表示矩阵的第 i 行， M_j 表示矩阵的第 j 列。那么目标1和2可以分别表示为

- $\forall i \in [1, |P|], \sum_j m_{ij} = k$
- $\forall j \in [1, |R|], \sum_i m_{ij} \leq l_j$

同时，所有作者和评审人（这两个集合交集可能非空）可以构成一个包含其社会关系的网络 $G = (V_G, E_G)$ 。该网络可以根据几个原则构建。例如，我们可以使用顶点来表示作者或评审人，并使用边缘来说明他们之间存在合著关系，或者我们可以基于他们在Twitter等社交媒体中的关注-被关注关系来构建网络。

在网络 G 的基础上，我们可以衡量论文作者及其评审人之间的社交关系的紧密程度。每篇论文 p 的作者集合 A^p 和评审集合 $B^p = \{j | m_{ij} \in M^i \ \& \ m_{ij} = 1\}$ 的并集可以构成一个 G 的子图 $F = (V_F, E_F)$ ，其中 $V_F = \{v \in V | v \in (A^p \cup B^p)\}$ ， $|E_F| = \{(v_i, v_j) | v_i, v_j \in V_F, (v_i, v_j) \in E_G\}$ 。然后我们可以利用任意有效指标（ k 密度）度量子图的结构稀疏程度。

$$k - density(F) = \frac{\sum_{u \in V_F} |N_G^k(u) \cap V_F|}{|V_F|(|V_F| - 1)}$$

，其中 $N_G^k(u)$ 表示节点 u 在原网络 G 中与其距离不大于 k 的节点的集合。

我们可以使用一个简单的相似度度量方式来衡量任意评审人 r 的研究领域与文章 p 的研究问题的相关性

$$w(p, r) = \frac{|I^p \cap I^r|}{|I^p|}$$

， w 越高说明评审人的研究方向与论文的研究问题越相近，也就越能给出专业的修改意见。

在以上工作的基础上我们可以将目标3和4分别表示为

- $\max(\frac{\sum_{b \in B^p} w(p, b)}{|B^p|})$
- $\min(\sum_{p \in P} k - density(F_p))$
- $\forall p \in P, |E_{F_p}| = 0$

即要尽量提升评审人对论文研究问题的熟悉程度，同时在保证没有直接的联系的前提下降低评审与作者及评审之间的联系。

针对任意分配方案 M ，目标函数可以定为 $f(M) = \max(\frac{\sum_{b \in B^p} w(p, b)}{|B^p|}) / \min(\sum_{p \in P} k - density(F_p))$ ，然后将问题表示为寻找最合适的矩阵 M ，使 M 在满足

1. $\forall i \in [1, |P|], \sum_j m_{ij} = k$
2. $\forall j \in [1, |R|], \sum_i m_{ij} \leq l_j$
3. $\forall p \in P, |E_{F_p}| = 0$

的前提下最大化 $f(M)$ 。

（个人感觉）这个问题可以扩展为两个部分，第一部分是获得评委和论文的研究问题，这也是目前 **reviewer assignment** 问题研究的一个主要部分，属于信息抽取领域。第二部分是给每个评审安排合适的论文，这个问题应当属于社交事件组织邻域。