

清华大学

# 综合论文训练

题目：基于期刊引用网络有向特征的  
学术期刊评价指标研究

系 别：电机工程与应用电子技术系

专 业：电气工程及其自动化

姓 名：谢睿

指导教师：梁恒 副教授

2017年6月9日

## 中文摘要

近年来期刊评价指标是一个研究热点，许多新指标被提出。本文将几个有向网络层级结构特性指标引入期刊评价，分别针对期刊引用网络和文章引用网络，指出应用层级结构特性指标的步骤，选用或设计计算方法。之后，将新的期刊评价指标和几个常用的现有指标同时应用于一些实际引用网络，并对比分析评价结果。结果显示，层级结构特性指标可用于期刊评价，且与现有指标反映的期刊质量特征不完全相同。

**关键词：**期刊评价；有向网络；层级结构

## ABSTRACT

In recent years, indices for journal evaluation has become a hot topic in bibliometrics research and a number of new indices have been put forward. In this thesis, some indices based on hierarchy structure in directed networks are introduced into the field of journal evaluation. The methods used to apply these indices are described, and the calculation process is selected or designed carefully. The new indices and multiple frequently-used existing indices are applied to some real citation networks at the same time, with the aim of assessing the journal evaluation result by comparison. It has been observed in experiments that it is reasonable to use hierarchy based indicators in the journal evaluation process, but these indices may reveal some different journal quality characteristics compared with existing indices.

**Keywords:** journal evaluation; directed network; hierarchy structure

# 目 录

第1章 引言 .....	1
1.1 研究意义 .....	1
1.2 研究现状 .....	1
1.2.1 学术期刊评价指标 .....	1
1.2.2 网络的有向特性与层级结构 .....	3
1.3 本文主要内容 .....	4
第2章 有向网络层级结构指标简介 .....	5
2.1 符号 .....	5
2.2 $\beta$ -measure .....	5
2.2.1 定义 .....	6
2.2.2 性质 .....	6
2.3 LRC 与 GRC 指标 .....	7
2.3.1 不加权情况 .....	7
2.3.2 加权情况 .....	8
2.3.3 案例 .....	8
2.4 Agony 指标 .....	9
2.4.1 定义 .....	9
2.4.2 性质 .....	10
2.4.3 案例 .....	10
2.4.4 小结 .....	11
第3章 数据收集与指标计算 .....	12
3.1 实验数据收集与预处理 .....	12
3.1.1 数据集 .....	12
3.1.2 数据预处理 .....	12
3.2 指标的计算方法 .....	13

3.2.1	影响因子 .....	13
3.2.2	期刊的 h-指数 .....	13
3.2.3	PageRank .....	14
3.2.4	$\beta$ -measure .....	14
3.2.5	LRC .....	15
3.2.6	Agony .....	16
<b>第4章</b>	<b>计算方法推导 .....</b>	<b>19</b>
4.1	LRC 指标的算法 .....	19
4.2	Agony 最小解的计算方法 .....	20
4.2.1	定义 .....	20
4.2.2	等价线性规划问题 .....	21
4.2.3	层级的上界 .....	26
4.2.4	无圈有向图与 agony 最小解 .....	32
<b>第5章</b>	<b>实验结果与分析 .....</b>	<b>43</b>
5.1	数学期刊引用网络 .....	43
5.2	DP 引用网络 .....	49
5.3	AAV8 引用网络 .....	56
5.4	小结 .....	62
<b>第6章</b>	<b>结论 .....</b>	<b>63</b>
6.1	本文工作总结 .....	63
6.2	存在的问题与改进方向 .....	63
	插图索引 .....	65
	表格索引 .....	67
	参考文献 .....	68
	致 谢 .....	70
	声 明 .....	71
附录 A	外文资料的调研阅读报告或书面翻译 .....	72

## 主要符号对照表

$IF$	影响因子
$\mathbb{R}_+$	正实数集合
$\mathbb{N}$	自然数集合
$\mathbb{Z}$	整数集合
$I$	单位阵
$D$	有向图
$V$	顶点集合
$E$	有向边集合
$D_\omega$	加权有向图
$\omega$	加权有向图权函数
$n$	有向图顶点数
$m$	有向图边数
$S_D(i)$	图 $D$ 中顶点 $i$ 的后继点集合
$s_D(i)$	图 $D$ 中顶点 $i$ 的后继点数量
$P_D(i)$	图 $D$ 中顶点 $i$ 的前驱点集合
$p_D(i)$	图 $D$ 中顶点 $i$ 的前驱点数量
$d_D^+(i)$	图 $D$ 中顶点 $i$ 的出度
$d_D^-(i)$	图 $D$ 中顶点 $i$ 的入度
$\beta_i(D)$	图 $D$ 中顶点 $i$ 的 $\beta$ -measure
$\beta_i(D_\omega)$	图 $D_\omega$ 中顶点 $i$ 的 $\beta$ -measure
LRC	Local reaching centrality
GRC	Global reaching centrality
$A_i$	有向图中顶点 $i$ 可达的点的集合（不包括顶点 $i$ ）
$C_R(i)$	有向图中顶点 $i$ 的 LRC
$C_R^{max}$	有向图中最大 LRC
$d^{out}(i, j)$	有向图中从 $i$ 到 $j$ 的最少边数路的边数
$\omega_i^{(k)}(j)$	加权有向图中从 $i$ 到 $j$ 的总权最大的最少边数路中第 $k$ 条边的权
$r$	有向图顶点的层级分配方案

$R$	有向图的层级分配方案集合
$Agony(D)$	图 $D$ 的 agony
$h(D)$	不加权有向图 $D$ 的 hierarchy
$Agony(D_\omega)$	图 $D_\omega$ 的 agony
DAG	无圈有向图 (directed acyclic graph)
DP	数据集 Data for Pagerank
AAV8	数据集 AMiner ACM-Citation-network V8
$PY$	文章年份矩阵
$PJ$	文章期刊矩阵
$PC$	文章引用矩阵
$JC$	期刊引用信息
$NPY$	期刊发表文章总数矩阵
$SJC$	按年份截取 $JC$ 并求和后的矩阵
$SNP$	按年份截取 $NPY$ 并求和后的矩阵
$n_p$	文章数量
$n_j$	期刊数量
$IF(i)$	期刊 $i$ 的影响因子
$h(i)$	期刊 $i$ 的 h-指数
$R^p$	PageRank 向量
$A_D$	图 $D$ 的邻接矩阵
$P_D$	图 $D$ 的可达矩阵
$r_0$	Agony 最小解
$S_A$	最优层级分配方案的集合
$M^D$	图 $D$ 的关联矩阵
$V(C)$	图 $C$ 的顶点集合
$E(C)$	图 $C$ 的边集合
$q(e, r)$	层级分配方案 $r$ 在边 $e$ 的惩罚 (不考虑加权)
$Q(r)$	层级分配方案 $r$ 的总惩罚
$JC_n$	去除期刊自引之后的期刊引用信息
$JC_i$	通过阈值 $i$ 去除加权后的期刊引用信息
WoS	Web of Science
$IF_n$	使用 $JC_n$ 计算的影响因子

$P$	使用 $JC_n$ 计算的加权 PageRank 指标
$\beta_i$	使用 $JC_i$ 计算的 $\beta$ -measure
$\beta_\omega$	使用 $JC$ 计算的加权 $\beta$ -measure
$C_{Ri}$	使用 $JC_i$ 计算的 LRC 指标
$C_{R\omega}$	使用 $JC$ 计算的加权 LRC 指标
$r_i$	使用 $JC_i$ 计算的 agony 指标
$r_\omega$	使用 $JC$ 计算的加权 agony 指标
$P^P$	基于文章引用网络并对所有文章求平均值的 PageRank 指标
$\beta^P$	基于文章引用网络并对所有文章求平均值的 $\beta$ -measure
$C_R^P$	基于文章引用网络并对所有文章求平均值的 LRC 指标
$r^P$	基于文章引用网络并对所有文章求平均值的 agony 指标

## 第1章 引言

本章首先指明期刊评价指标的研究意义，然后从现有期刊评价指标、基于网络有向特性的层级结构指标这两个方面总结研究现状，最后介绍本文的主要内容和组织结构。

### 1.1 研究意义

研究期刊评价指标有重要的实际意义，原因在于以下几个方面<sup>[1]</sup>：

第一，目前世界上出版的期刊总量非常大，每个学科都同时有很多期刊，因此存在选择的问题。

第二，许多人需要选择期刊。例如，研究者希望能在最好的期刊上发表文章，以提高研究成果的影响力；读者限于时间、资金等限制，只能阅读有限的期刊，因此希望选择文章最具价值的期刊；评价学者的学术能力时，也常常采用发表文章所在期刊的质量作为一种衡量标准。

第三，目前最常用的期刊评价指标是影响因子，但它是一个有广泛争议的指标，一些文献总结了它的不足，而且认为它反映期刊质量的能力有限<sup>[1,2]</sup>。

### 1.2 研究现状

本节介绍现有期刊评价指标和有向网络层级结构指标，复杂网络相关背景可参考文献 [3-5]。

#### 1.2.1 学术期刊评价指标

评价学术期刊的方法可以分为两大类，一类是根据文章的使用情况计算，另一类则是在学术界进行调查和统计。第二类方法的一个例子是，文献 [6] 通过对北美、欧洲与澳大利亚的一些相关科学家进行网上调查，评价并排序信息系统领域期刊，进一步分析地域对于排序结果的影响。由于引用、访问量等信息可以被准确记录，因此第一类易确保数据的有效性，避免调查统计中潜在的问题，所以通常认为第一类方法较客观。

Garfield 在 20 世纪中期提出了影响因子 (impact factor, IF) 作为学术期刊评价指标<sup>[1]</sup>。某期刊在  $y$  年的影响因子, 是该期刊  $(y - 1)$  和  $(y - 2)$  年发表的文章在  $y$  年获得的引用数, 除以该期刊  $(y - 1)$  和  $(y - 2)$  年发表的文章总数, 即

$$IF_y = \frac{(y - 1) \text{ 年文章在 } y \text{ 年被引用数} + (y - 2) \text{ 年文章在 } y \text{ 年被引用数}}{(y - 1) \text{ 年文章数} + (y - 2) \text{ 年文章数}}$$

影响因子应用广泛, 比如 Web of Science 提供各领域世界顶尖期刊影响因子的计算结果; 一些期刊编辑用影响因子和其他相关指标 (立即指数、被引半衰期) 分析期刊在所处领域中的地位<sup>[7]</sup>。文献 [8] 采用计算影响因子与调查统计两种方法分析信息系统领域的期刊, 比较结果的差距, 给出综合评价。然而, 影响因子也引起许多争论。影响因子定义的局限性主要体现在这几个方面: 高影响因子期刊中的典型文章被引用数不一定高; 影响因子反映的是学术界对期刊的关注度, 而不是重要性或效用; 影响因子可能被多种因素干扰而产生偏差, 这些因素包括研究领域、出版周期、引用时间分布、负面引用等等<sup>[1]</sup>。除此以外, 文献 [1] 还介绍了针对这些因素的修改版影响因子, 最后讨论应该怎样应用影响因子。

近十余年来, 许多其他学术期刊指标被提出。2005 年 J. E. Hirsch 引入 h-指数用于评价科学家。Tibor Braun 等人推广 h-指数至期刊评价<sup>[9]</sup>。它的定义是, 如果一个期刊发表了  $a$  篇被引用数不少于  $a$  的文章, 则这个期刊的 h-指数等于  $a$ 。h-指数的主要优点是, 第一, 它比较健壮, 受异常值影响较小; 第二, 它同时考虑了质量与数量; 第三, 由文章被引用数目可以直接计算 h-指数。他们将影响因子和 h-指数同时应用于实际数据中, 并将结果进行对比, 发现两个指标有明显差异。例如, 文章总数少的期刊 h-指数不可能高, 但影响因子可以很高。然而, 与影响因子相似, h-指数与学科有很大关系。文献 [9] 认为, h-指数是对影响因子的有效补充。

学术引用网络和万维网都是有向复杂网络。Google 的 PageRank 算法根据万维网中的超链接进行网页排序。在这个过程中, 某个网页的得分与两方面因素有关: 指向这个网页的超链接的数量、那些含有指向这个网页超链接的网页的得分<sup>[10]</sup>。Johan Bollen 等人将加权的 PageRank 算法应用于学术期刊评价, 并指出影响因子是对期刊受关注程度的衡量, 而加权 PageRank 算法得到的分数是

对期刊声望的衡量<sup>[11]</sup>。文献 [11] 进一步定义了 Y 因数来同时反映影响因子与加权 PageRank，然后基于实际数据对比这些学术期刊评价指标，分析了跨学科和不同学科内部的情况，发现这两种指标线性相关但有差异，而它们的综合指标 Y 因数的表现与 h-指数有相似之处。除了 PageRank，还可以利用网络结构得到其他指标，例如接近中心性（closeness centrality，反映某点到其余点的距离）和中介中心性（betweenness centrality，根据经过某点的最短路数目定义）<sup>[2]</sup>。

随着互联网的发展，科学家越来越多地通过互联网访问数据库来获得学术论文。基于论文引用数据的学术期刊评价指标可以类比至论文使用日志数据，并由此得到一族新的指标<sup>[2]</sup>。

一些文献通过实验对比分析多种学术期刊评价指标的效果。文献 [12] 计算了 7 种反映引用情况的导出量：平均被引用次数、不被引用的比例、被引用超过 20 次的比例、年平均被引用次数、被引用数和引用数之比、自引量、即时引用次数。文献 [12] 将期刊分别按这些量排名，再由 7 种排名算出综合排名。Johan Bollen 等人对 39 种学术期刊指标做主成分分析<sup>[2]</sup>。这些指标包括影响因子、被引半衰期、h-指数、PageRank 等基于网络结构的指标，以及它们各种类型的变化形式。结论是，学术影响力具有多维的结构，不能被一个指标充分体现。而与其他指标相比，影响因子的表现不是非常好，应用它时必须小心。

### 1.2.2 网络的有向特性与层级结构

一些文献研究基于网络有向特性的节点评价方法。文献 [10] 针对网页超链接组成的有向网络，提出 PageRank 指标，给出算法，并实验验证。如上文所述，PageRank 指标已被应用于学术期刊评价<sup>[11]</sup>。文献 [13] 为了衡量有向网络中节点之间的支配关系，研究  $\beta$ -measure。文献 [14] 提出 local reaching centrality (LRC)，想法是对每个节点计算能直接或间接支配的节点总数，以获得层级信息。之后，文献 [14] 在多种网络中实验所提指标，研究这些网络的层级特性。文献 [15] 提出 agony 用于衡量社会网络中的层级关系，想法是级别较高的节点应该不被级别较低的节点支配。文献 [16,17] 推广 agony 至加权有向网络。

另一些文献针对具体问题研究网络的层级特性。文献 [18] 提出一种在电子邮件网络中得出社会层级结构的方法。文献 [19] 研究怎样在有向网络中找出有代表性的传播路径，并且指出可由此得到网络层级结构信息。

### 1.3 本文主要内容

本文研究学术期刊的评价指标。首先总结有向网络层级结构特性指标，然后将其引入期刊引用网络和文章引用网络，作为学术期刊质量的新评价指标。之后针对引用网络的特点，设计或选用指标计算方法，并利用多个实际引用数据集实验。

本文余下章节的安排是：第 2 章介绍本文采用的几种层级结构指标的定义、原理和基本信息；第 3 章介绍实验数据来源，叙述实验涉及的各种指标的计算方法；第 4 章是指标计算方法的推导，针对本文所提计算方法的合理性给出证明；第 5 章对 3 个实验数据集的引用网络，分别使用现有指标与基于层级结构特性的指标，给出期刊评价，并对比分析结果；第 6 章总结全文，说明研究结论及其意义，指出本文的不足与改进方向。

## 第2章 有向网络层级结构指标简介

本章首先说明本文使用的基本符号及其意义，限制本文讨论的有向图范围，然后根据已有研究内容<sup>[14,15,17,20]</sup>，介绍一些有向网络层级结构指标的定义、原理和基本性质。

### 2.1 符号

令  $\mathbb{R}_+$  表示正实数集合， $\mathbb{N}$  表示自然数集合， $\mathbb{Z}$  表示整数集合，集合左侧加符号  $\#$  表示集合的元素个数。令  $I$  表示单位阵，即对角线上是 1，其他元素是 0 的方阵。

不加权的有向图用  $D = (V, E)$  表示，其中  $V$  是顶点集合， $E$  是有向边集合。有时也用  $(i, j)$  指代从  $i$  出发指向  $j$  的有向边。本文考虑有限无自环的有向图，即  $V$  和  $E$  有限且对任意  $i \in V$  有  $(i, i) \notin E$ 。

用  $D_\omega = (V, E, \omega)$  表示加权有向图，其中  $V$  是顶点集合， $E$  是有向边集合， $\omega : E \rightarrow \mathbb{R}_+$  是从每条有向边  $e \in E$  到正实数的函数。这里仍然要求  $V$  和  $E$  有限且对任意  $i \in V$  有  $(i, i) \notin E$ 。

令  $n = \#V$  为有向图顶点总数， $m = \#E$  为边总数。如果  $(i, j) \in E$ ，则将  $j$  称为  $i$  的后继点，将  $i$  称为  $j$  的前驱点。对任意  $i \in V$ ，定义  $i$  的后继点集合为  $S_D(i) = \{j \in V | (i, j) \in E\}$ ，前驱点集合为  $P_D(i) = \{j \in V | (j, i) \in E\}$ 。并且，记  $s_D(i) = \#S_D(i)$ ， $p_D(i) = \#P_D(i)$ ，即点  $i$  的后继点数量与前驱点数量。分别记顶点  $i$  在图  $D$  中的入度和出度为  $d_D^-(i)$  和  $d_D^+(i)$ 。

### 2.2 $\beta$ -measure

$\beta$ -measure 由文献 [20] 引进。文献 [13] 研究了  $\beta$ -measure，推广至加权有向图，并给出了定义这个指标的等价公理。

$\beta$ -measure 用于衡量支配关系，将从顶点  $i$  到顶点  $j$  的有向边理解为顶点  $i$  支配顶点  $j$ 。

### 2.2.1 定义

定义 2.1 (不加权有向图的  $\beta$ -measure): 设  $D = (V, E)$  是不加权的有向图, 将  $D$  的  $\beta$ -measure 定义为

$$\beta_i(D) = \sum_{j \in S_D(i)} \frac{1}{p_D(j)}, \forall i \in V \quad (2-1)$$

由以上定义, 一个顶点的  $\beta$ -measure 不仅与该点的后继点数量有关, 还与这些后继点的前驱点数量有关。一个顶点的后继点越多, 且这些后继点的前驱点越少, 则  $\beta$ -measure 越高。

在加权有向图  $D_\omega = (V, E, \omega)$  中, 对任意  $i \in V$ , 定义  $\lambda_\omega(i) = \sum_{j: (j,i) \in E} \omega(j, i)$ , 即顶点  $i$  的前驱点给  $i$  的总权。

定义 2.2 (加权有向图的  $\beta$ -measure): 设  $D_\omega = (V, E, \omega)$  是加权有向图, 将  $D_\omega$  的  $\beta$ -measure 定义为

$$\beta_i(D_\omega) = \sum_{(i,j) \in E} \frac{\omega(i, j)}{\lambda_\omega(j)} = \sum_{(i,j) \in E} \frac{\omega(i, j)}{\sum_{k: (k,j) \in E} \omega(k, j)}, \forall i \in V \quad (2-2)$$

不加权有向图  $D = (V, E)$  可以看作加权有向图  $D_\omega = (V, E, \omega)$  的特例, 只要令  $\omega : E \rightarrow \mathbb{R}_+$  定义为  $\omega(i, j) = 1, \forall (i, j) \in E$ 。通过这种方法, 从加权有向图的  $\beta$ -measure 定义也可以得到不加权有向图的  $\beta$ -measure 定义。

### 2.2.2 性质

$\beta$ -measure 有一些基本性质, 以下讨论针对不加权有向图, 但都可以推广到加权有向图。

命题 2.1: 对任意不加权有向图  $D = (V, E)$ , 有

$$\sum_{i \in V} \beta_i(D) = \#\{j \in N \mid p_D(j) > 0\} \quad (2-3)$$

命题 2.2: 对任意不加权有向图  $D = (V, E)$  和任意  $i \in V$ , 如果  $S_D(i) = \emptyset$ , 则  $\beta_i(D) = 0$ 。

**命题 2.3:** 对任意不加权有向图  $D = (V, E)$  和任意  $i, j \in V$ ，如果  $S_D(i) = S_D(j)$  且  $P_D(i) = P_D(j)$ ，则  $\beta_i(D) = \beta_j(D)$ 。

命题 2.1 表示，不加权有向图的  $\beta$ -measure 总和等于被支配顶点的总数。命题 2.2 的意义是，对于没有后继点的顶点， $\beta$ -measure 为 0。命题 2.3 说明，如果两个顶点的后继点集合与前驱点集合是相同的，那么它们的  $\beta$ -measure 相等。可以看出，命题 2.2 和命题 2.3 指出的性质是对层级结构指标的自然要求。

$\beta$ -measure 在一定情况下对于图的并具有可加性，且文献 [13] 证明这个性质连同上述的 3 个性质可作为  $\beta$ -measure 的等价定义。

## 2.3 LRC 与 GRC 指标

文献 [14] 基于 m-reach centrality，提出 local reaching centrality (LRC)，作为有向图中衡量某点层级的指标，并且进一步定义 global reaching centrality (GRC)，作为衡量有向图层级特性强度的指标。

与  $\beta$ -measure 相似，此处一条从顶点  $i$  到顶点  $j$  的有向边表示顶点  $i$  对顶点  $j$  有优势。

### 2.3.1 不加权情况

**定义 2.3 (不加权有向图的 LRC):** 设  $D = (V, E)$  是不加权的有向图， $n = \#V$ ，某顶点  $i \in V$ 。令集合  $A_i = \{j \in V - \{i\} | D \text{ 中存在从 } i \text{ 到 } j \text{ 的路}\}$ ，定义点  $i$  的 LRC 为

$$C_R(i) = \frac{\#A_i}{n-1} \quad (2-4)$$

LRC ( $C_R(i)$ ) 的意义为从  $i$  沿有向边能够到达顶点的比例，因此取值范围是  $[0, 1]$ 。在具有层级结构的有向网络中，一个顶点的级别越高，一般来说它沿网络能到达的层级数与顶点数也就越多。

**定义 2.4 (GRC):** 令  $C_R^{max} = \max\{C_R(i) | i \in V\}$ ，则图  $D$  的 GRC 定义为

$$GRC = \frac{\sum_{i \in V} [C_R^{max} - C_R(i)]}{n-1} \quad (2-5)$$

GRC 衡量有向图层级特性强度，取值范围也是  $[0, 1]$ 。对于星型网络（一个顶点指向所有其他顶点，且除此以外没有其他边的网络）， $GRC = 1$ ，表明这种网络有很强的层级特性。而有向完全图（任意不同顶点  $i$  和  $j$  之间有  $i$  到  $j$  的边） $GRC = 0$ 。

### 2.3.2 加权情况

设  $D_\omega = (V, E, \omega)$  是加权的有向图， $n = \#V$ ，某顶点  $i \in V$ 。对任意  $j \in V$ ，令  $d^{out}(i, j)$  为  $D_\omega$  中从  $i$  到  $j$  的最少边数路的边数，取值范围是  $[0, +\infty]$ ，其中  $+\infty$  表示没有从  $i$  到  $j$  的路。令  $\omega_i^{(k)}(j)$  为  $i$  到  $j$  的最少边数路上第  $k$  条边的权（如果有多条最少边数路，取总权最大的最少边数路）。

定义 2.5 (加权有向图的 LRC): 在加权有向图  $D_\omega = (V, E, \omega)$  中，定义顶点  $i$  的 LRC 为

$$C_R(i) = \frac{1}{n-1} \sum_{j: 0 < d^{out}(i, j) < +\infty} \frac{\sum_{k=1}^{d^{out}(i, j)} \omega_i^{(k)}(j)}{d^{out}(i, j)} \quad (2-6)$$

与  $\beta$ -measure 类似，可以把不加权有向图看作是加权有向图的特例，导出不加权情况下的 LRC。GRC 的定义在加权或不加权时不变。

容易验证，LRC 具有命题 2.2 和命题 2.3 指出的性质。

### 2.3.3 案例

表 2.1 一些网络的 GRC 指标（摘自文献 [14]）

类型	$i \rightarrow j$ 意义	顶点数	平均顶点度	GRC
食物网	$i$ 吃 $j$	135	4.452	0.814
新陈代谢网	$j$ 是 $i$ 的产物	1173	2.442	0.048
法语网络	$j$ 跟随 $i$	9424	2.578	0.657

文献 [14] 计算了许多实际网络的 LRC 与 GRC 指标，包括食物网、电气网络、新陈代谢网、神经网络、因特网、语言网络等等。一些结果见表 2.1。

## 2.4 Agony 指标

文献 [15] 提出针对不加权有向图的 agony 指标, 证明求 agony 的问题与求最大欧拉子图问题的关系, 并由此给出多项式时间的算法。文献 [16] 将 agony 推广到加权有向图, 指出相应问题与网络费用流问题之间的联系, 且进一步研究了改变罚函数形式后的算法。

### 2.4.1 定义

对有向图  $D = (V, E)$ , 函数  $r : V \rightarrow \mathbb{N} = \{0, 1, 2, \dots\}$  称为图  $D$  的一种层级分配方案。记  $R$  为所有分配方案的集合, 即  $R = \{r : V \rightarrow \mathbb{N}\}$ 。

与前面介绍的两种指标不同, agony 指标的定义认为有向边表示推荐关系, 即一条从顶点  $i$  到顶点  $j$  的有向边表示  $i$  的层级比  $j$  低。

如果对于一种层级分配, 存在  $i$  到  $j$  的有向边, 但  $i$  的层级不低于  $j$ , 则用罚函数来描述这种异常情况及其程度。在所有层级分配方案中, 使全图总惩罚最小的方案被认为是最好的, 并由此给出各点的层级 (有可能不唯一)。

**定义 2.6 (不加权有向图的 agony):** 设  $D = (V, E)$  是不加权的有向图, 则  $D$  的 agony 定义为

$$Agony(D) = \min_{r \in R} \left( \sum_{(i,j) \in E} \max\{r(i) - r(j) + 1, 0\} \right) \quad (2-7)$$

文献 [15] 证明了不加权有向图的 agony 不大于图的边数  $m$ , 即 agony 的取值范围是  $[0, m]$ 。根据这一点, 定义反映有向图层级特性强度的指标 hierarchy。它的取值范围是  $[0, 1]$ , 越高表示层级特性越强。

**定义 2.7 (不加权有向图的 hierarchy):** 设  $D = (V, E)$  是不加权的有向图, 则  $D$  的 hierarchy 定义为

$$h(D) = 1 - \frac{1}{m} Agony(D) \quad (2-8)$$

文献 [16] 给出了加权有向图的 agony 定义。

定义 2.8 (加权有向图的agony): 设  $D_\omega = (V, E, \omega)$  是加权的有向图, 则  $D_\omega$  的 agony 定义为

$$Agony(D_\omega) = \min_{r \in \mathbb{R}} \left( \sum_{(i,j) \in E} \omega(i,j) \max\{r(i) - r(j) + 1, 0\} \right) \quad (2-9)$$

### 2.4.2 性质

求 agony 的问题是一个整数线性规划问题, 例如在不加权有向图的情形中, 这个问题是

$$\begin{aligned} \min \quad & \sum_{(i,j) \in E} x(i,j) \\ \text{s. t.} \quad & x(i,j) \geq r(i) - r(j) + 1 \quad \forall (i,j) \in E \\ & x(i,j) \geq 0 \quad \forall (i,j) \in E \\ & r(i) \geq 0 \quad \forall i \in V \\ & r(i) \in \mathbb{Z} \quad \forall i \in V \end{aligned} \quad (2-10)$$

将这个问题松弛成线性规划, 写出对偶问题, 再将对偶变量加上取整数的条件, 则得到求最大边数欧拉子图 (不要求连通, 只要求子图中每个顶点出度等于入度) 的问题。文献 [15] 证明后者的最优值等于 agony, 并由此得出算法。

在不加权有向图的情形中, 有两种特殊情况的 agony 已知, 即无圈有向图 (directed acyclic graph, DAG) 和完全图, 并且它们分别代表层级特性最强与最弱的情况。

**命题 2.4:** 设  $D = (V, E)$  是不加权无圈有向图, 则  $Agony(D) = 0, h(D) = 1$ 。

**命题 2.5:** 设  $D = (V, E)$  是不加权的有向完全图, 则  $D$  是边数为  $m = \#E$  的不加权有向图中 agony 最大的, 且  $Agony(D) = m, h(D) = 0$ 。

文献 [17] 提出在 agony 解集中找到使各点层级最小的一个解的方法, 并证明了它的唯一性。这个解具有命题 2.2 和命题 2.3 指出的性质 (在改变有向边方向的含义之后)。

### 2.4.3 案例

文献 [16] 计算了 SNAP 数据集<sup>[21]</sup> 中许多实际网络的 agony 指标。一些网络的结果见表 2.2。

表 2.2 一些网络的 agony 计算结果（摘自文献 [16]）

名字	顶点数	边数	层级数
Amazon	403k	3m	17
Reef	258	4232	19
WikiVote	7k	104k	12

#### 2.4.4 小结

本章说明了本文采用的基本符号，介绍了几个基于网络有向特性的层级结构指标，即  $\beta$ -measure、LRC 指标和 agony 指标。它们都具有针对加权有向图和不加权有向图的不同形式，且不加权情况可看作加权情况的特例。

## 第3章 数据收集与指标计算

### 3.1 实验数据收集与预处理

#### 3.1.1 数据集

本文实验使用三个数据集：Data for PageRank、AMiner ACM-Citation-network V8 和数学期刊引用网络。

Data for PageRank (DP) 是生物学领域的文章引用网络。这些文章发表时间在 1985 年至 2007 年之间。该数据集包含来自 397 种期刊的 286419 篇文章，共计 867698 次内部引用。信息包括文章的 WoS 编号、学科类别、文献类型、发表年份、所属期刊、作者信息和引用信息。

AMiner ACM-Citation-network V8<sup>[22]</sup> (AAV8) 是计算机领域的文献引用网络。这些文献的发表时间在 2016 年及之前，共计 2381719 篇文献和 8650089 次引用。该数据集的信息包括文献的编号、标题、作者、发表年份、引用信息和摘要。

数学期刊引用网络包含数学领域 311 种数学期刊，信息来源是 Web of Science Journal Citation Reports，信息形式是各个期刊 2015 年以前各年文章在 2015 年的被引用次数（区分引用期刊），以及各个期刊各年发表的文章总数。

#### 3.1.2 数据预处理

本实验数据集是两个文章引用网络和一个期刊引用网络。这两种类型的网络采用不同的预处理步骤，以得到便于后续操作的矩阵形式数据。

对于文章引用网络，首先对文章、年份和期刊编号，并用表格记录编号与内容的对应关系。然后，根据文章信息得出文章年份矩阵  $PY$  和文章期刊矩阵  $PJ$ ，根据引用信息得出文章引用矩阵  $PC$ 。其中这三个矩阵都是稀疏的 0-1 矩阵。对任意矩阵  $A$ ，用  $A(i, j)$  表示  $A$  第  $i$  行第  $j$  列的元素。定义

$$PY(i, j) = \begin{cases} 1, & \text{如果文章 } i \text{ 在 } j \text{ 年发表} \\ 0, & \text{否则} \end{cases} \quad (3-1)$$

$$PJ(i, j) = \begin{cases} 1, & \text{如果文章 } i \text{ 属于期刊 } j \\ 0, & \text{否则} \end{cases} \quad (3-2)$$

$$PC(i, j) = \begin{cases} 1, & \text{如果文章 } i \text{ 被文章 } j \text{ 引用} \\ 0, & \text{否则} \end{cases} \quad (3-3)$$

利用这三个矩阵，可以得到期刊引用信息  $JC$  和期刊发表文章总数矩阵  $NPY$ ，其中  $JC(i, j, k, l)$  表示期刊  $i$  在  $k$  年发表的文章被期刊  $j$  在  $l$  年引用的次数， $NPY(i, j)$  表示期刊  $i$  在  $j$  年发表的文章总数。

对于期刊引用网络，将期刊和年份编号之后，可以直接得到期刊引用信息  $JC$  和期刊发表文章总数矩阵  $NPY$ 。

## 3.2 指标的计算方法

为了说明各指标的计算方法，假设要分析  $y$  年对  $y-2$  和  $y-1$  年的引用信息，截取  $JC$  和  $NPY$  的相应部分，并对  $y-2$  年与  $y-1$  年求和。记得到的矩阵分别为  $SJC$  和  $SNP$ ，其中  $SJC(i, j)$  表示期刊  $i$  在  $y-2$  和  $y-1$  年发表的文章被期刊  $j$  在  $y$  年引用的次数， $SNP(i)$  表示期刊  $i$  在  $y-2$  和  $y-1$  年发表的文章总数。

### 3.2.1 影响因子

设  $i$  是某期刊， $SNP(i) > 0$ ，则该期刊在  $y$  年的影响因子 (IF) [1] 是

$$\begin{aligned} IF(i) &= \frac{\text{期刊 } i \text{ 于 } y-2 \text{ 和 } y-1 \text{ 年发表的文章在 } y \text{ 年被引用总次数}}{\text{期刊 } i \text{ 于 } y-2 \text{ 和 } y-1 \text{ 年发表的文章总数}} \\ &= \frac{\sum_j SJC(i, j)}{SNP(i)} \end{aligned} \quad (3-4)$$

### 3.2.2 期刊的 h-指数

期刊引用网络不足以计算 h-指数，这里使用文章引用网络。h-指数的定义 [9] 是

$$h(i) = \max\{j \in \mathbb{N} \mid \text{期刊 } i \text{ 至少 } j \text{ 篇文章被引量不小于 } j\} \quad (3-5)$$

分别为每个期刊计算 h-指数，先计算该期刊所有文章的被引总次数，从高到低排列得  $L_{n_p \times 1}$ ，其中  $n_p$  表示该期刊文章总数。之后，从单篇文章最高引用次数开始，检查是否满足条件，满足则找到了 h-指数，不满足则减小引用次数阈值，如此循环。见算法 1。

---

**算法 1** 计算 h-指数

---

输入:  $L_{n_p \times 1}$

输出:  $h$

- 1:  $h \leftarrow L(1)$
  - 2:  $n \leftarrow \#\{i \in \mathbb{Z} | 1 \leq i \leq n_p, L(i) \geq h\}$
  - 3: **while**  $n < h$  **do**
  - 4:      $h \leftarrow \max\{L(i) | 1 \leq i \leq n_p, L(i) < h\}$
  - 5:      $n \leftarrow \#\{i \in \mathbb{Z} | 1 \leq i \leq n_p, L(i) \geq h\}$
  - 6: **end while**
  - 7: **return**  $h$
- 

### 3.2.3 PageRank

假设对引用矩阵（期刊引用矩阵或文章引用矩阵）计算 PageRank 指标。首先将其非零列正规化，使单列求和结果为 1，记得到的矩阵为  $A_{n_j \times n_j}^p$ ，其中  $n_j$  是期刊数。根据文献 [10]，PageRank 向量  $R_{n_j \times 1}^p$  是矩阵  $A^p + E \cdot \mathbf{1}_{1 \times n_j}$  的最大特征值对应的特征向量。这里取  $E_{n_j \times 1}$  为各分量相等的向量，且  $\|E\|_1 = 0.15, E(i) \geq 0, 1 \leq i \leq n_j$ 。采用文献 [10] 中的算法计算  $R^p$ ，即算法 2。

### 3.2.4 $\beta$ -measure

$\beta$ -measure 要求有向图没有自环，因此令引用矩阵的对角线元素为 0。

可以计算两种  $\beta$ -measure。第一种针对不加权有向图。例如文章引用网络。又如利用阈值将期刊引用网转换成不加权的有向图  $D$ ，规则是图  $D$  中有顶点  $i$  到顶点  $j$  的有向边当且仅当期刊  $j$  对期刊  $i$  的引用数大于指定阈值。根据定义 2.1 计算，可得图  $D$  的  $\beta$ -measure。可以改变引用数阈值，得到不同的不加权有向图。

---

**算法 2** 计算 PageRank<sup>[10]</sup>

---

输入:  $A_{n_j \times n_j}^p$ ,  $E_{n_j \times 1}$ ,  $\epsilon > 0$

输出:  $R_{n_j \times 1}^p$

```
1:  $R_0^p \leftarrow E$ 
2:  $R_1^p \leftarrow A^p R_0^p$ 
3:  $d \leftarrow \|R_0^p\|_1 - \|R_1^p\|_1$ 
4:  $R_1^p \leftarrow R_1^p + dE$ 
5:  $\delta \leftarrow \|R_1^p - R_0^p\|_1$ 
6:  $i \leftarrow 1$ 
7: while  $\delta > \epsilon$  do
8:    $R_{i+1}^p \leftarrow A^p R_i^p$ 
9:    $d \leftarrow \|R_i^p\|_1 - \|R_{i+1}^p\|_1$ 
10:   $R_{i+1}^p \leftarrow R_{i+1}^p + dE$ 
11:   $\delta \leftarrow \|R_{i+1}^p - R_i^p\|_1$ 
12:   $i \leftarrow i + 1$ 
13: end while
14: return  $R_i^p$ 
```

---

第二种针对加权有向图。例如，将期刊引用网络当作加权有向图  $D_\omega$ ，根据定义 2.2 计算加权有向图的  $\beta$ -measure。其中  $D_\omega$  中从顶点  $i$  到顶点  $j$  的边的权  $\omega(i, j) =$  期刊  $j$  引用期刊  $i$  的次数。

### 3.2.5 LRC

与小节 3.2.4 相同，可以对不加权有向图  $D$  和加权有向图  $D_\omega$  分别计算 LRC 指标  $C_R$ 。

根据定义 2.3，对  $D$  要计算的是每个顶点的可达点数，可以直接利用  $D$  的邻接矩阵计算可达矩阵<sup>[23]</sup>。具体做法如下：设图  $D$  的邻接矩阵  $A_D$  为  $n \times n$  的矩阵，且

$$A_D(i, j) = \begin{cases} 1, & \text{如果顶点 } i \text{ 到顶点 } j \text{ 有边} \\ 0, & \text{否则} \end{cases} \quad (3-6)$$

则  $A_D$  的  $l$  次幂  $A_D^l$  中元素  $A_D^l(i, j)$  是顶点  $i$  到顶点  $j$  边数为  $l$  的链数。令  $B_l = I + A_D + A_D^2 + \cdots + A_D^l$ ，则其中某一列的非零元素位置表示从相应顶点经过边数至多为  $l$  的链能到达的所有顶点。若  $B_l$  的非零元素位置和  $B_{l+1}$  的非零元素位置完全相同，则可达矩阵  $P_D$  为

$$P_D(i, j) = \begin{cases} 1, & \text{如果 } B_l(i, j) > 0 \\ 0, & \text{否则} \end{cases} \quad (3-7)$$

其中  $P_D(i, j) = 1$  等价于顶点  $i$  可达顶点  $j$ <sup>[23]</sup>。对每个顶点  $i$ ，计算可达顶点总数（不包括顶点  $i$  自身）与  $n - 1$  的比，即得不加权有向图中的  $C_R(i)$ 。

对于加权有向图  $D_\omega$ ，根据定义 2.5，需要求出任意两个顶点间最小边数路的最大权。Dijkstra 算法是一种求加非负权有向图中一个顶点到其他顶点最短路的算法<sup>[24]</sup>。修改 Dijkstra 算法，得出计算加权有向图中某点 LRC 的算法 3（见定理 4.1）。

### 3.2.6 Agony

与小节 3.2.4 相似，可以对不加权有向图  $D$  和加权有向图  $D_\omega$  计算最优层级分配方案。本小节的目标是计算使各顶点层级最低的最优层级分配方案  $r_0 : V \rightarrow \mathbb{N}$ 。期刊引用网络规模较小，直接求解线性规划问题（见推论 4.2）

$$\begin{aligned} \min \quad & \sum_{(i,j) \in E} \omega(i, j)x(i, j) + \frac{1}{M} \sum_{i \in V} r(i) \\ \text{s. t.} \quad & x(i, j) \geq r(i) - r(j) + 1 \quad \forall (i, j) \in E \\ & x(i, j) \geq 0 \quad \forall (i, j) \in E \\ & r(i) \geq 0 \quad \forall i \in V \end{aligned} \quad (3-8)$$

其中网络是加权有向图  $D_\omega = (V, E, \omega)$ （用小节 2.2.1 的方法将不加权有向图看成加权有向图）， $n = \#V$ ， $M = n^2$ 。

文章引用网络规模较大，但有圈的可能性很小，可以利用这一点计算最优层级分配方案。以下方法与加权与否无关。

第一步，检查该有向图是否无圈<sup>[25]</sup>。设  $D = (V, E)$ ，删除  $D$  中没有前驱点或没有后继点的顶点（即入度或出度为 0 的顶点）。对剩下的有向图，继续删除这样的顶点，直至没有可删的顶点，即算法 4。

---

**算法 3** 计算加权有向图中某点的 LRC

---

输入: 加权有向图  $D_\omega = (V, E, \omega)$  ( $V = \{v_1, v_2, \dots, v_n\}$ ) 及点  $i \in V$

输出:  $C_R(i)$

```
1:  $r_{n \times 1} \leftarrow -1_{n \times 1}$ ,  $R_0 \leftarrow \{i\}$ 
2:  $r(i) \leftarrow 0$ 
3:  $w_{n \times 1} \leftarrow 0_{n \times 1}$ 
4:  $j \leftarrow 1$ 
5: while  $j \leq n$  且  $R_{j-1} \neq \emptyset$  do
6:    $R_j \leftarrow \emptyset$ 
7:   for  $k = 1, 2, \dots, n$  do
8:     if  $r(v_k) < 0$  且  $\exists l \in V, s.t. r(l) = j - 1, (l, v_k) \in E$  then
9:        $R_j \leftarrow R_j \cup \{v_k\}$ 
10:       $w(v_k) \leftarrow \max\{w(l) + \omega(l, v_k) | r(l) = j - 1, (l, v_k) \in E\}$ 
11:       $r(v_k) \leftarrow j$ 
12:     end if
13:   end for
14:    $j \leftarrow j + 1$ 
15: end while
16:  $C_R(i) \leftarrow \frac{1}{n-1} \sum_{j:r(j)>0} \frac{w(j)}{r(j)}$ 
17: return  $C_R(i)$ 
```

---

此时, 如果  $V$  中所有的顶点均被删完, 则说明图  $D$  是无圈有向图 (见定理 4.9)。如果有剩余的顶点, 则  $D$  不是无圈有向图, 设此时剩下的边是  $E_R$ 。令  $E_n = E - E_R$ , 得到新的图  $D_n = (V, E_n)$ ,  $D_n$  无圈 (见定理 4.8)。应该指出, 这时  $D_n$  的最优层级分配方案不一定是  $D$  的最优层级分配方案, 但由于图  $D$  本身有圈的可能性很小, 结果应该相差不多 (见第 5 章实验结果)。

第二步, 采用针对无圈有向图的最优层级分配算法计算  $r : V \rightarrow \mathbb{N}$ , 即算法 5 (见定理 4.11)。

---

**算法 4** 删除出度或入度为 0 的顶点<sup>[25]</sup>

---

输入: 有向图  $D = (V, E)$

输出: 有向图  $D_R = (V_R, E_R)$ 。

- 1:  $V_R \leftarrow V, E_R \leftarrow E, D_R \leftarrow (V_R, E_R)$
  - 2: **while**  $\exists i_0 \in V_R, \text{ s. t. } d_{D_R}^+(i_0) \cdot d_{D_R}^-(i_0) = 0$  **do**
  - 3:      $V_R \leftarrow V - \{i_0\}, E_R \leftarrow \{e = (i, j) \in E_R \mid i, j \in V_R\}, D_R \leftarrow (V_R, E_R)$
  - 4: **end while**
  - 5: **return**  $D_R$
- 

---

**算法 5** 计算 DAG 中使各顶点层级最低的最优层级分配方案  $r_0 : V \rightarrow \mathbb{N}$ 

---

输入: 无圈有向图  $D = (V, E)$  ( $V = \{v_1, v_2, \dots, v_n\}$ )

输出:  $r_0 : V \rightarrow \mathbb{N}$

- 1:  $R_0 \leftarrow \emptyset$
  - 2: **for**  $k = 1, 2, \dots, n$  **do**
  - 3:     **if**  $d_D^-(v_k) = 0$  **then**
  - 4:          $R_0 \leftarrow R_0 \cup \{v_k\}, r_0(v_k) \leftarrow 0$
  - 5:     **end if**
  - 6: **end for**
  - 7:  $j \leftarrow 1$
  - 8: **while**  $R_{j-1} \neq \emptyset$  **do**
  - 9:      $R_j \leftarrow \emptyset$
  - 10:    **for**  $k = 1, 2, \dots, n$  **do**
  - 11:       **if**  $\exists l \in R_{j-1}, \text{ s.t. } (l, v_k) \in E$  **then**
  - 12:            $R_j \leftarrow R_j \cup \{v_k\}, r_0(v_k) \leftarrow j$
  - 13:       **end if**
  - 14:    **end for**
  - 15:     $j \leftarrow j + 1$
  - 16: **end while**
  - 17: **return**  $r_0$
-

## 第4章 计算方法推导

### 4.1 LRC 指标的算法

本节证明算法 3 能求出加权有向图的 LRC 指标。

**引理 4.1:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $p_0 : v_0 e_1 v_1 e_2 \dots v_{s-1} e_s v_s$  是  $D_\omega$  中从  $v_0$  到  $v_s$  最大总权的边数最少的路 (即  $p_0$  在从  $v_0$  到  $v_s$  的边数最少的路中具有最大总权), 则  $q_0 : v_0 e_1 v_1 e_2 \dots e_{s-1} v_{s-1}$  是从  $v_0$  到  $v_{s-1}$  最大总权的边数最少的路。

**证明** 对任意路  $p$ , 记路  $p$  的边数为  $e(p)$ , 总权为  $w(p)$ 。设  $q_1$  是从  $v_0$  到  $v_{s-1}$  最大总权的边数最少的路。则边数  $e(q_1) \leq e(q_0) = e(p_0) - 1 < e(p_0)$ , 由于  $p_0$  是从  $v_0$  到  $v_s$  的边数最少的路, 因此  $q_1$  不通过  $v_s$ 。令  $p_1$  为  $q_1$  连接  $e_s v_s$  产生的链, 则  $p_1$  是从  $v_0$  到  $v_s$  的路。

边数  $e(p_1) = e(q_1) + 1 \leq e(q_0) + 1 = e(p_0)$ , 又由  $p_0$  是从  $v_0$  到  $v_s$  的边数最少的路以及  $p_1$  是  $v_0$  到  $v_s$  的路知  $e(p_1) = e(p_0)$ , 从而  $e(q_1) = e(q_0)$ ,  $q_0$  是从  $v_0$  到  $v_{s-1}$  的边数最少的路。

总权  $w(p_1) = w(q_1) + \omega(e_s) \geq w(q_0) + \omega(e_s) = w(p_0)$ , 由  $p_0$  和  $p_1$  都是  $v_0$  到  $v_s$  的边数最少的路以及  $p_0$  是  $v_0$  到  $v_s$  的边数最少的路中具有最大权的, 因此总权  $w(p_1) \leq w(p_0)$ , 从而  $w(p_1) = w(p_0)$ ,  $w(q_1) = w(q_0)$ , 即  $q_0$  是  $v_0$  到  $v_{s-1}$  最大总权的边数最少的路。  $\square$

**定理 4.1:** 输入加权有向图  $D_\omega = (V, E, \omega)$  ( $V = \{v_1, v_2, \dots, v_n\}$ ) 及点  $i \in V$  时, 算法 3 的输出是 LRC 指标  $C_R(i)$ 。

**证明** 由算法 3 第 16 行知, 算法结束时

$$\text{输出} = \frac{1}{n-1} \sum_{j:r(j)>0} \frac{w(j)}{r(j)}$$

因此只需证  $r(j) > 0$  时,  $r(j) = d^{\text{out}}(i, j)$  (即  $i$  到  $j$  的边数最少路的边数), 且  $w(j)$  为  $i$  到  $j$  边数最少的路的最大总权。

考虑第 5 至 15 行的循环，对循环次数  $j$  归纳。第 1 次循环结束时， $k \in R_1$  当且仅当存在从  $i$  到  $k$  的边数为 1 的边数最少的路。当  $l \leq 1$  时，对任意  $k \in R_l$ ， $w(k)$  为  $i$  到  $k$  边数最少的路的最大总权。假设第  $j$  次循环结束时，若  $l \leq j$ ，则  $k \in R_l$  当且仅当存在从  $i$  到  $k$  的边数为  $l$  的边数最少的路，且对任意  $k \in R_l$ ， $w(k)$  为  $i$  到  $k$  边数最少的路的最大总权。由  $R_{j+1}$  的定义， $R_{j+1}$  是能从  $R_j$  经一条边到达，且不属于  $R_l, \forall l \leq j$  的点的集合。由引理 4.1 可得，在第  $j+1$  次循环结束时， $k \in R_{j+1}$  当且仅当存在从  $i$  到  $k$  的边数为  $j+1$  的边数最少的路，且对任意  $k \in R_{j+1}$ ， $w(k)$  为  $i$  到  $k$  边数最少的路的最大总权。由数学归纳法，循环全部结束时，若  $r(k) > 0$ ，则  $r(k)$  等于  $i$  到  $k$  的边数最少路的边数，且  $w(k)$  为  $i$  到  $k$  边数最少的路的最大总权。  $\square$

## 4.2 Agony 最小解的计算方法

本节为 agony 最小解的相关计算方法给出证明。首先定义 agony 最小解，然后证明一定条件下可通过问题 (3-8) 求出 agony 最小解。因为问题 (3-8) 中  $M$  的选取与层级上界有关，所以讨论了层级的一些上界。之后，证明算法 4 和算法 5 的有效性，讨论对非无圈有向图应用算法 4 和算法 5 能得出 agony 最小解的条件，并对一种特殊情况给出修改得到 agony 最小解的算法。

### 4.2.1 定义

由定义 2.8，对于加权有向图  $D_\omega = (V, E, \omega)$ ，求 agony 的问题是一个整数线性规划问题

$$\begin{aligned}
 \min \quad & \sum_{(i,j) \in E} \omega(i,j)x(i,j) \\
 \text{s. t.} \quad & x(i,j) \geq r(i) - r(j) + 1 \quad \forall (i,j) \in E \\
 & x(i,j) \geq 0 \quad \forall (i,j) \in E \\
 & r(i) \geq 0 \quad \forall i \in V \\
 & r(i) \in \mathbb{Z} \quad \forall i \in V
 \end{aligned} \tag{4-1}$$

记集合  $S_A = \{r : V \rightarrow \mathbb{Z} \mid \text{存在 } x \text{ 使 } (x, r) \text{ 是问题 (4-1) 的最优解}\}$ 。

**定理 4.2** (文献 [17] Proposition 5): 存在  $r_0 \in S_A$ ，使得对任意  $r \in S_A$ ， $i \in V$  有  $r_0(i) \leq r(i)$ 。

显然，对于给定的  $D_\omega = (V, E, \omega)$  而言， $r_0$  唯一，称  $r_0$  为  $D_\omega$  的 agony 最小解。

#### 4.2.2 等价线性规划问题

将整数线性规划 (4-1) 松弛成线性规划问题

$$\begin{aligned}
 \min \quad & \sum_{(i,j) \in E} \omega(i,j)x(i,j) \\
 \text{s. t.} \quad & x(i,j) \geq r(i) - r(j) + 1 \quad \forall (i,j) \in E \\
 & x(i,j) \geq 0 \quad \forall (i,j) \in E \\
 & r(i) \geq 0 \quad \forall i \in V
 \end{aligned} \tag{4-2}$$

令函数  $F(x, r) = \sum_{(i,j) \in E} \omega(i,j)x(i,j)$ ，函数  $G(x, r) = \sum_{(i,j) \in E} \omega(i,j)x(i,j) + \frac{1}{M} \sum_{i \in V} r(i)$ 。将问题 (4-2) 中的目标函数换成  $G(x, r)$ ，约束不变，得

$$\begin{aligned}
 \min \quad & \sum_{(i,j) \in E} \omega(i,j)x(i,j) + \frac{1}{M} \sum_{i \in V} r(i) \\
 \text{s. t.} \quad & x(i,j) \geq r(i) - r(j) + 1 \quad \forall (i,j) \in E \\
 & x(i,j) \geq 0 \quad \forall (i,j) \in E \\
 & r(i) \geq 0 \quad \forall i \in V
 \end{aligned} \tag{4-3}$$

本小节说明，一定条件下，问题 (4-3) 的最优解  $r = r_0$ ，且其对偶问题可写成网络最小费用流问题。

**定义 4.1:** 设  $A$  是一个矩阵，若任取  $A$  的子方阵  $B$ ， $B$  的行列式等于 0 或 1 或 -1，则称  $A$  是全单位模矩阵 (Totally unimodular matrix)。

**定义 4.2:** 设  $D = (V, E)$  是有向图， $n = \#V$ ， $m = \#E$ ，定义  $D$  的关联矩阵为  $M_{n \times m}^D$ ，其中

$$M^D(i, j) = \begin{cases} 1, & \text{如果 } j \text{ 是从 } i \text{ 出发的边} \\ -1, & \text{如果 } j \text{ 是指向 } i \text{ 的边} \\ 0, & \text{否则} \end{cases} \tag{4-4}$$

定理 4.3 (文献 [26] Theorem 13.9): 设  $D$  是有向图,  $M^D$  是  $D$  的关联矩阵, 则  $M^D$  是全单位模矩阵。

引理 4.2: 若  $A$  是全单位模矩阵, 则  $-A$ 、 $A^T$  和  $[A \ I]$  也是全单位模矩阵, 其中  $I$  是单位阵。

可直接用全单位模矩阵的定义验证引理 4.2。

定理 4.4 (文献 [27] Theorem 13.1): 设  $A$  是  $m \times n$  的全单位模矩阵,  $b \in \mathbb{Z}^m$ , 则多面体  $P := \{x | Ax = b, x \geq 0\}$  的所有顶点是整数点。

推论 4.1: 对于问题 (4-2) 和问题 (4-3), 若最优解存在, 则存在各分量为整数的最优解。

证明 问题 (4-2) 和问题 (4-3) 的可行域相同, 记为  $P$ 。则

$$P = \left\{ (x, r) \left| \begin{array}{l} x(i, j) \geq r(i) - r(j) + 1, \forall (i, j) \in E \\ x(i, j) \geq 0, \forall (i, j) \in E \\ r(i) \geq 0, \forall i \in V \end{array} \right. \right\} \quad (4-5)$$

$$= \left\{ (x, r) \left| \begin{array}{l} x(i, j) - s(i, j) - r(i) + r(j) = 1, \forall (i, j) \in E \\ x(i, j) \geq 0, \forall (i, j) \in E \\ r(i) \geq 0, \forall i \in V \\ s(i, j) \geq 0, \forall (i, j) \in E \end{array} \right. \right\} \quad (4-6)$$

令  $n = \#V$ ,  $m = \#E$ ,  $M_{n \times m}^D$  是  $D$  的关联矩阵,  $I$  表示单位阵,  $0$  表示零矩阵,  $1$  表示全一矩阵。令

$$y = \begin{bmatrix} x_{m \times 1} \\ s_{m \times 1} \\ r_{n \times 1} \end{bmatrix}, A = \begin{bmatrix} I_{m \times m} \\ -I_{m \times m} \\ -M_{n \times m}^D \end{bmatrix}^T, b = 1_{m \times 1}$$

则

$$P = \{(x, r) | Ay = b, y \geq 0, y = (x, s, r)\}$$

由定理 4.3 和引理 4.2, 可得  $A$  是全单位模矩阵。由定理 4.4,  $P$  的顶点均为整数点。

问题 (4-2) 和问题 (4-3) 是线性规划, 其标准形式若存在最优解, 则存在基本可行解达到最优。因此问题 (4-2) 和问题 (4-3) 存在整数的最优解。  $\square$

问题 (4-3) 作为线性规划的标准形式是

$$\begin{aligned}
 \min \quad & \sum_{(i,j) \in E} \omega(i,j)x(i,j) + \frac{1}{M} \sum_{i \in V} r(i) \\
 \text{s. t.} \quad & x(i,j) - s(i,j) - r(i) + r(j) = 1, \forall (i,j) \in E \\
 & x(i,j) \geq 0, \forall (i,j) \in E \\
 & s(i,j) \geq 0, \forall (i,j) \in E \\
 & r(i) \geq 0, \forall i \in V
 \end{aligned} \tag{4-7}$$

**定理 4.5:** 对于加权有向图  $D_\omega = (V, E, \omega)$ , 设  $r_0$  是 agony 最小解,  $n = \#V$ ,  $m = \#E$ 。若  $\omega(i,j) \in \mathbb{N}, \omega(i,j) > 0, \forall (i,j) \in E$ , 问题 (4-3) 中  $M > \sum_{i \in V} r_0(i)$ , 则问题 (4-3) 存在整数的最优解。此时若  $(x_B, s_B, r_B)$  是问题 (4-3) 标准形式 (4-7) 的一个基本可行解且达到最优, 则  $r_B = r_0$ 。

**证明** 第一步, 证明问题 (4-3) 存在整数的最优解。

因为  $\omega(i,j), x(i,j), r(k), M \geq 0, \forall (i,j) \in E, k \in V$ , 所以问题 (4-3) 的目标函数大于等于 0, 问题 (4-3) 有界。令  $x(i,j) = 1, r(k) = 0, \forall (i,j) \in E, k \in V$ , 则  $(x, r)$  是问题 (4-3) 的可行解。因此线性规划问题 (4-3) 可行且有界, 存在最优解。

由推论 4.1, 问题 (4-3) 存在整数的最优解, 设为  $(x_3, r_3)$ 。

第二步, 证明  $(x_3, r_3)$  是问题 (4-1) 的一个最优解。

令函数  $F(x, r) = \sum_{(i,j) \in E} \omega(i,j)x(i,j)$ , 函数  $G(x, r) = \sum_{(i,j) \in E} \omega(i,j)x(i,j) + \frac{1}{M} \sum_{i \in V} r(i)$ 。

由于  $r_0$  是 agony 最小解, 存在  $x_0$  使得  $(x_0, r_0)$  是问题 (4-1) 的最优解。由推论 4.1, 问题 (4-2) 存在整数的最优解  $(x_2, r_2)$ , 则  $(x_2, r_2)$  是问题 (4-1) 的可行解, 所以  $F(x_2, r_2) \geq F(x_0, r_0)$ 。问题 (4-2) 可行域包含问题 (4-1) 可行域, 因此  $F(x_0, r_0) \geq F(x_2, r_2)$ 。从而得到  $F(x_0, r_0) = F(x_2, r_2)$ ,  $(x_0, r_0)$  也是问题 (4-2) 的最优解。

记  $P$  为问题 (4-2) 和问题 (4-3) 的可行域。则

$$\begin{aligned}
 F(x_0, r_0) &= \min_{(x,r) \in P} F(x, r) \\
 &\leq F(x_3, r_3) \\
 &\leq G(x_3, r_3) \\
 &= \min_{(x,r) \in P} G(x, r) \\
 &\leq G(x_0, r_0) \\
 &= F(x_0, r_0) + \frac{1}{M} \sum_{i \in V} r_0(i) \\
 &< F(x_0, r_0) + 1
 \end{aligned}$$

因此  $F(x_0, r_0) \leq F(x_3, r_3) < F(x_0, r_0) + 1$ 。

又因为  $(x_0, r_0), (x_3, r_3)$  均为整数点, 且  $\omega(i, j) \in \mathbb{Z}, \forall (i, j) \in E$ , 所以  $F(x_0, r_0), F(x_3, r_3) \in \mathbb{Z}$ , 进而  $F(x_0, r_0) = F(x_3, r_3)$ 。又  $(x_3, r_3)$  是问题 (4-1) 的可行解, 所以  $(x_3, r_3)$  是问题 (4-1) 的一个最优解。

第三步, 证明  $r_3 = r_0$ 。

由于  $r_0$  是 agony 最小解,  $(x_3, r_3)$  是问题 (4-1) 的一个最优解, 因此  $r_0(i) \leq r_3(i), \forall i \in V$ 。假设  $r_3 \neq r_0$ , 则存在  $i \in V$  使  $r_3(i) > r_0(i)$ , 这时有

$$\begin{aligned}
 G(x_0, r_0) &= F(x_0, r_0) + \frac{1}{M} \sum_{i \in V} r_0(i) \\
 &= F(x_3, r_3) + \frac{1}{M} \sum_{i \in V} r_0(i) \\
 &< F(x_3, r_3) + \frac{1}{M} \sum_{i \in V} r_3(i) \\
 &= G(x_3, r_3)
 \end{aligned}$$

注意到  $(x_0, r_0)$  是问题 (4-3) 的可行解, 上式与  $(x_3, r_3)$  是问题 (4-3) 的一个最优解矛盾, 因此  $r_3 = r_0$ 。

第四步, 证明  $r_B = r_0$ 。

由推论 4.1 的证明过程, 问题 (4-3) 标准形式的基本可行解均为整数点。又由第三步知, 整数点中只有满足  $r = r_0$  才可能是问题 (4-3) 的最优解。因此

$r_B = r_0$ 。

□

问题 (4-3) 标准形式 (4-7) 的对偶问题是

$$\begin{aligned}
 & \max \quad \sum_{(i,j) \in E} f(i,j) \\
 & \text{s. t.} \quad f(i,j) + t_1(i,j) = \omega(i,j), \forall (i,j) \in E \\
 & \quad \quad -f(i,j) + t_2(i,j) = 0, \forall (i,j) \in E \\
 & \quad \quad -\sum_{j:(i,j) \in E} f(i,j) + \sum_{j:(j,i) \in E} f(j,i) + t_3(i) = \frac{1}{M}, \forall i \in V \\
 & \quad \quad t_1(i,j), t_2(i,j), t_3(k) \geq 0, \forall (i,j) \in E, k \in V
 \end{aligned} \tag{4-8}$$

可写成

$$\begin{aligned}
 & \max \quad \sum_{(i,j) \in E} f(i,j) \\
 & \text{s. t.} \quad 0 \leq f(i,j) \leq \omega(i,j), \forall (i,j) \in E \\
 & \quad \quad -\sum_{j:(i,j) \in E} f(i,j) + \sum_{j:(j,i) \in E} f(j,i) + t(i) = \frac{1}{M}, \forall i \in V \\
 & \quad \quad t(i) \geq 0, \forall i \in V
 \end{aligned} \tag{4-9}$$

问题 (4-9) 中,  $-\sum_{j:(i,j) \in E} f(i,j) + \sum_{j:(j,i) \in E} f(j,i) + t(i) = \frac{1}{M}, \forall i \in V$  对所有  $i \in V$  求和, 可得  $\sum_{i \in V} t(i) = \frac{n}{M}$ 。

根据  $D_\omega = (V, E, \omega)$  构造图  $\tilde{D}_\omega = (\tilde{V}, \tilde{E}, \tilde{\omega})$ 。设  $s$  是一个新顶点, 令  $\tilde{V} = V \cup \{s\}$ ,  $\tilde{E} = E \cup \{(s,i) | i \in V\}$ ,  $\tilde{\omega}(e) = \omega(e), \forall e \in E$ ,  $\tilde{\omega}(e) = 0, \forall e \in \tilde{E} - E$ 。

令  $g: \tilde{E} \rightarrow \mathbb{R}$  为  $g(e) = f(e), \forall e \in E$ ,  $g(s,i) = t(i), \forall i \in V$ 。令  $b: \tilde{V} \rightarrow \mathbb{R}$  为  $b(i) = \frac{1}{M}, \forall i \in V$ ,  $b(s) = -\frac{n}{M}$ 。令  $a: \tilde{E} \rightarrow \mathbb{R}$  为  $a(e) = -1, \forall e \in E$ ,  $a(e) = 0, \forall e \in \tilde{E} - E$ 。令  $c: \tilde{E} \rightarrow \mathbb{R}$  为  $c(e) = \omega(e), \forall e \in E$ ,  $c(e) = +\infty, \forall e \in \tilde{E} - E$ 。

则问题 (4-9) 可转化为  $\tilde{D}_\omega = (\tilde{V}, \tilde{E}, \tilde{\omega})$  中的网络最小费用流问题<sup>[28]</sup>

$$\begin{aligned}
 & \min \quad \sum_{(i,j) \in \tilde{E}} a(i,j)g(i,j) \\
 & \text{s. t.} \quad 0 \leq g(i,j) \leq c(i,j), \forall (i,j) \in \tilde{E} \\
 & \quad \quad -\sum_{j:(i,j) \in \tilde{E}} g(i,j) + \sum_{j:(j,i) \in \tilde{E}} g(j,i) = b(i), \forall i \in \tilde{V}
 \end{aligned} \tag{4-10}$$

利用一个例子说明以上过程。

**例 4.1:** 设图  $D_\omega$  如图 4.1 所示, 其中  $v_1$ 、 $v_2$ 、 $v_3$  是节点,  $(v_1, v_2)$ 、 $(v_2, v_3)$ 、 $(v_3, v_2)$  是边。则图  $\tilde{D}_\omega$  应该如图 4.2 所示, 即增加了节点  $s$  和边  $(s, v_1)$ 、 $(s, v_2)$ 、

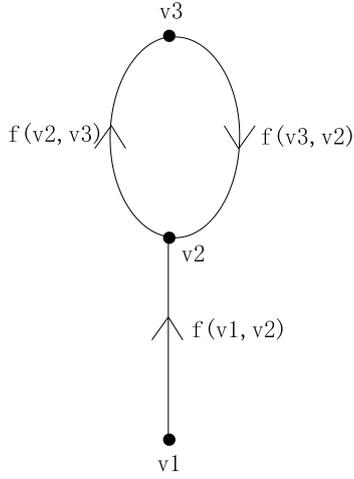


图 4.1 例 4.1 图  $D_\omega$

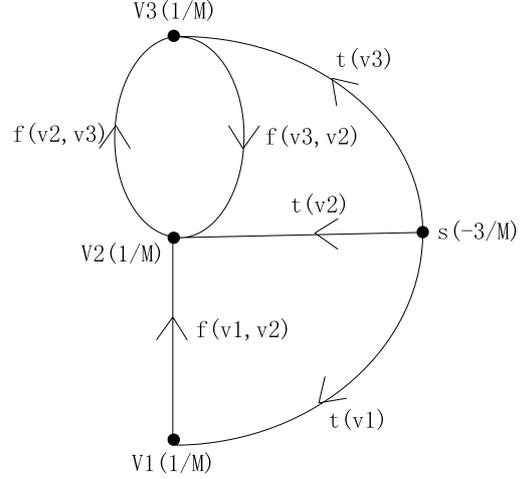


图 4.2 例 4.1 图  $\tilde{D}_\omega$

$(s, v3)$ 。在各边附近标出了新的网络流  $g$  的值，在节点旁括号内标出了节点向图外的流出量。

### 4.2.3 层级的上界

本小节指出 agony 最小解中层级的几种上界，结果可用于在问题 (4-3) 中取  $M$  使得  $M > \sum_{i \in V} r_0(i)$ 。

对加权有向图  $D_\omega = (V, E, \omega)$  的层级分配方案  $r : V \rightarrow \mathbb{N}$ ，定义函数  $q(e, r)$  为

$$q(e, r) = \max\{r(i) - r(j) + 1, 0\} \quad (4-11)$$

其中  $e \in E$  且从  $i$  出发指向  $j$ 。定义函数  $Q(r)$  为

$$Q(r) = \sum_{e \in E} \omega(e)q(e, r) \quad (4-12)$$

层级分配方案  $r_1$  是 agony 解（即  $r_1 \in S_A$ ）当且仅当  $Q(r_1) = \min_{r \in R} Q(r)$ 。

**命题 4.1:** 设  $D_\omega = (V, E, \omega)$  是加权有向图， $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解。若  $i, j \in V$ ，满足  $r_0(i) - r_0(j) > 1$ ，令  $R_{ij} = \{r | r \in \mathbb{N}, r_0(j) < r < r_0(i)\}$ ，则对任意  $\bar{r} \in R_{ij}$ ，存在  $k \in V$  使  $r_0(k) = \bar{r}$ 。

**证明** 因为  $r_0(i), r_0(j) \in \mathbb{Z}$  且  $r_0(i) - r_0(j) > 1$ , 所以  $r_0(j) + 1 \in R_{ij}$ ,  $R_{ij} \neq \emptyset$ 。

反证, 设  $\bar{r} \in R_{ij}$ , 但对任意  $k \in V$  都有  $r_0(k) \neq \bar{r}$ 。

令  $V_1 = \{k \in V | r_0(k) < \bar{r}\}$ ,  $V_2 = \{k \in V | r_0(k) > \bar{r}\}$ , 则  $V_1 \cup V_2 = V$ ,  $V_1 \cap V_2 = \emptyset$ 。

定义函数  $r_1 : V \rightarrow \mathbb{Z}$  为

$$r_1(k) = \begin{cases} r_0(k), & k \in V_1 \\ r_0(k) - 1, & k \in V_2 \end{cases}$$

对任意  $k_1 \in V_1$ ,  $r_1(k_1) = r_0(k_1) \geq 0$ 。对任意  $k_2 \in V_2$ ,  $r_1(k_2) = r_0(k_2) - 1 > \bar{r} - 1 > r_0(j) - 1 \geq -1$ , 由  $r_1(k_2)$  是整数知  $r_1(k_2) \geq 0$ 。因此  $r_1(k) \geq 0, \forall k \in V$ ,  $r_1$  是一种层级分配方案。

对任意  $k_1 \in V_1, k_2 \in V_2$  有  $r_0(k_1) < \bar{r} < r_0(k_2)$ 。由于  $r_0(k_1), \bar{r}, r_0(k_2) \in \mathbb{Z}$ ,  $r_1(k_1) = r_0(k_1) < \bar{r} \leq r_1(k_2) = r_0(k_2) - 1$ , 所以  $r_1(k_1) < r_1(k_2), \forall k_1 \in V_1, k_2 \in V_2$ 。据此容易验证, 将  $r_0$  换成  $r_1$  后, 不改变顶点间的层级大小关系, 但  $|r_1(k_x) - r_1(k_y)| \leq |r_0(k_x) - r_0(k_y)|, \forall k_x, k_y \in V$ 。因此,  $q(e, r_1) \leq q(e, r_0), \forall e \in E$ , 进而  $Q(r_1) \leq Q(r_0)$ , 所以  $r_1$  是 agony 解。但  $j \in V_2, r_1(j) < r_0(j)$ , 与  $r_0$  是 agony 最小解矛盾。

因此, 对任意  $\bar{r} \in R_{ij}$ , 存在  $k \in V$  使  $r_0(k) = \bar{r}$ 。 □

**命题 4.2:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解。则存在  $i_0 \in V$ , 使  $r_0(i_0) = 0$ 。

**证明** 反证, 假设对任意  $i \in V$  有  $r_0(i) \neq 0$ , 那么  $r_0(i) \geq 1$ 。

令  $r_1 : V \rightarrow \mathbb{Z}, r_1(i) = r_0(i) - 1, \forall i \in V$ 。则  $r_1(i) \geq 0$ , 因此  $r_1$  是一种层级分配方案。

将  $r_0$  换成  $r_1$  后, 不改变顶点间的层级相对关系, 即  $r_0(i) - r_0(j) = r_1(i) - r_1(j), \forall i, j \in V$ 。所以  $q(e, r_0) = q(e, r_1), \forall e \in E$ , 从而  $Q(r_0) = Q(r_1)$ ,  $r_1$  是  $D_\omega$  的 agony 解。然而  $r_0(i) > r_1(i), \forall i \in V$ , 与  $r_0$  是 agony 最小解矛盾。

因此, 存在  $i_0 \in V$ , 使  $r_0(i_0) = 0$ 。 □

**命题 4.3:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解,  $n = \#V$ , 则  $\max_{i \in V} r_0(i) \leq n - 1$ 。

**证明** 令  $S_0 = \{r_0(i) : i \in V\}$  是  $r_0 : V \rightarrow \mathbb{N}$  的取值集合。则由层级分配方案的定义，任取  $s \in S_0$  有  $s \in \mathbb{N}$ 。根据  $S_0$  的定义，有  $\#S_0 \leq \#V = n$ 。

由命题 4.2 知，存在  $i_0 \in V$  使  $r_0(i_0) = 0$ ，因此  $0 \in S_0$ 。又  $\max_{i \in V} r_0(i) \in S_0$ ，根据命题 4.1， $S_0 = \{0, 1, 2, \dots, \max_{i \in V} r_0(i)\}$ ， $\#S_0 = \max_{i \in V} r_0(i) + 1$ 。

因为  $\#S_0 \leq n$ ，所以  $\max_{i \in V} r_0(i) \leq n - 1$ 。 □

**命题 4.4:** 设  $D_\omega = (V, E, \omega)$  是加权有向图， $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解，且  $r_0^{max} = \max_{i \in V} r_0(i)$ ，则

$$\sum_{i \in V} r_0(i) \leq -\frac{1}{2}(r_0^{max})^2 + (n - \frac{1}{2})r_0^{max} \quad (4-13)$$

**证明** 令  $S_0 = \{r_0(i) : i \in V\}$ ， $n = \#V$ ，由命题 4.3 及其证明过程， $r_0^{max} \leq n - 1$  且  $S_0 = \{0, 1, 2, \dots, r_0^{max}\}$ 。因此

$$\begin{aligned} \sum_{i \in V} r_0(i) &\leq (0 + 1 + 2 + \dots + r_0^{max}) + (n - r_0^{max} - 1)r_0^{max} \\ &= -\frac{1}{2}(r_0^{max})^2 + (n - \frac{1}{2})r_0^{max} \end{aligned} \quad \square$$

**定理 4.6:** 设  $D_\omega = (V, E, \omega)$  是加权有向图， $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解， $n = \#V$ ，则

$$\sum_{i \in V} r_0(i) \leq \frac{1}{2}n^2 - \frac{1}{2}n \quad (4-14)$$

**证明** 设  $r_0^{max} = \max_{i \in V} r_0(i)$ ，则由命题 4.3 和层级分配方案定义有  $0 \leq r_0^{max} \leq n - 1, r_0^{max} \in \mathbb{Z}$ 。根据命题 4.4 有

$$\begin{aligned} \sum_{i \in V} r_0(i) &\leq \max \left\{ -\frac{1}{2}(r_0^{max})^2 + (n - \frac{1}{2})r_0^{max} \mid 0 \leq r_0^{max} \leq n - 1, r_0^{max} \in \mathbb{Z} \right\} \\ &= -\frac{1}{2}(r_0^{max})^2 + (n - \frac{1}{2})r_0^{max} \Big|_{r_0^{max}=n-1} \quad \square \\ &= \frac{1}{2}n^2 - \frac{1}{2}n \end{aligned}$$

**推论 4.2:** 对于加权有向图  $D_\omega = (V, E, \omega)$ , 设  $r_0$  是 agony 最小解,  $n = \#V$ ,  $\omega(i, j) \in \mathbb{N}, \omega(i, j) > 0, \forall (i, j) \in E$ 。设线性规划问题 (3-8) 标准形式的最优基本解对应  $(x, r)$ , 则  $r(i) = r_0(i), \forall i \in V$ 。

**证明** 由定理 4.6,  $M = n^2 > \frac{1}{2}n^2 - \frac{1}{2}n \geq \sum_{i \in V} r_0(i)$ 。利用定理 4.5 即得证。  $\square$

**命题 4.5:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解,  $\bar{E} = \{e = (i, j) \in E \mid r_0(i) \geq r_0(j) - 1\}$ ,  $i_0 \in V$ 。令  $\bar{D} = (V, \bar{E})$ ,  $I_0 = \{j \in V \mid \text{在 } \bar{D} \text{ 中 } j \text{ 可达 } i_0\}$  (包含  $i_0$ ), 则  $\min_{i \in I_0} r_0(i) = 0$ 。

**证明** 反证, 设  $\min_{i \in I_0} r_0(i) \neq 0$ , 则  $\min_{i \in I_0} r_0(i) \geq 1$ 。

定义  $r_1 : V \rightarrow \mathbb{Z}$  为

$$r_1(k) = \begin{cases} r_0(k), & k \in V - I_0 \\ r_0(k) - 1, & k \in I_0 \end{cases}$$

因为  $\min_{i \in I_0} r_0(i) \geq 1$ , 所以  $r_1(k) \geq 0, \forall k \in V$ ,  $r_1$  是一种层级分配方案。

设  $e = (i, j) \in E$ , 有以下 4 种情况:

(1)  $i, j \in I_0$ ,

$$\begin{aligned} q(e, r_1) &= \max\{r_1(i) - r_1(j) + 1, 0\} \\ &= \max\{(r_0(i) - 1) - (r_0(j) - 1) + 1, 0\} \\ &= \max\{r_0(i) - r_0(j) + 1, 0\} \\ &= q(e, r_0) \end{aligned}$$

(2)  $i, j \in V - I_0$ ,

$$\begin{aligned} q(e, r_1) &= \max\{r_1(i) - r_1(j) + 1, 0\} \\ &= \max\{r_0(i) - r_0(j) + 1, 0\} \\ &= q(e, r_0) \end{aligned}$$

(3)  $i \in I_0, j \in V - I_0$ ,

$$\begin{aligned} q(e, r_1) &= \max\{r_1(i) - r_1(j) + 1, 0\} \\ &= \max\{(r_0(i) - 1) - r_0(j) + 1, 0\} \\ &\leq \max\{r_0(i) - r_0(j) + 1, 0\} \\ &= q(e, r_0) \end{aligned}$$

(4)  $i \in V - I_0, j \in I_0$ 。由  $j \in I_0$  知在  $\bar{D}$  中  $j$  可达  $i_0$ 。如果  $(i, j) \in \bar{E}$ ，则  $i$  可达  $i_0$ ，进而  $i \in I_0$ ，与  $i \in V - I_0$  矛盾。因此  $(i, j) \in E - \bar{E}$ 。由  $\bar{E}$  的定义知， $r_0(i) < r_0(j) - 1$ ， $q(e, r_0) = 0$ 。因为  $r_0(i), r_0(j) \in \mathbb{Z}$ ，所以  $r_0(i) \leq r_0(j) - 2$ 。从而

$$\begin{aligned} q(e, r_1) &= \max\{r_1(i) - r_1(j) + 1, 0\} \\ &= \max\{r_0(i) - (r_0(j) - 1) + 1, 0\} \\ &\leq \max\{(r_0(j) - 2) - (r_0(j) - 1) + 1, 0\} \\ &= 0 \\ &= q(e, r_0) \end{aligned}$$

所以，对任意  $e \in E$  有  $q(e, r_1) \leq q(e, r_0)$ 。因此  $Q(r_1) \leq Q(r_0)$ ， $r_1$  是  $D_\omega$  的 agony 解。由  $I_0$  定义有  $i_0 \in I_0$ ，所以  $r_1(i_0) = r_0(i_0) - 1 < r_0(i_0)$ ，与  $r_0$  是 agony 最小解矛盾。

所以， $\min_{i \in I_0} r_0(i) = 0$ 。 □

**定理 4.7:** 设  $D_\omega = (V, E, \omega)$  是加权有向图， $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解， $l$  是  $D_\omega$  中边数最多的路的路的边数，则  $\max_{i \in V} r_0(i) \leq l$ 。

**证明** 只需证明，对任意  $i_0 \in V$  有  $r_0(i_0) \leq l$ 。

令  $\bar{E} = \{e = (i, j) \in E \mid r_0(i) \geq r_0(j) - 1\}$ ， $\bar{D} = (V, \bar{E})$ ， $I_0 = \{j \in V \mid \text{在 } \bar{D} \text{ 中 } j \text{ 可达 } i_0\}$ 。由命题 4.5 知， $\min_{i \in I_0} r_0(i) = 0$ ，因此存在  $i_1 \in I_0$  使  $r_0(i_1) = 0$ 。根据  $I_0$  的定义，存在  $\bar{D}$  中从  $i_1$  到  $i_0$  的路，设为  $p = v_0 e_1 v_1 \cdots v_{s-1} e_s v_s$ ，其中  $v_0 = i_1$ ， $v_s = i_0$ ， $e_1, e_2, \cdots, e_s \in \bar{E}$ 。因此  $r_0(v_{k-1}) \geq r_0(v_k) - 1, k = 1, 2, \cdots, s$ ， $r_0(i_0) = r_0(v_s) \leq r_0(v_{s-1}) + 1 \leq \cdots \leq r_0(v_0) + s = r_0(i_1) + s = s$ 。

因为  $\bar{E} \subseteq E$ ，所以  $p$  也是  $D_\omega$  中的路。由  $l$  定义有  $l \geq s \geq r_0(i_0)$ 。

所以,  $\max_{i \in V} r_0(i) \leq l$ . □

**推论 4.3:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解,  $n = \#V$ , 则  $\max_{i \in V} r_0(i) \leq n - 1$ .

**证明** 设  $l$  是  $D_\omega$  中边数最多的路的边数, 则  $l \leq n - 1$ . 由定理 4.7,  $\max_{i \in V} r_0(i) \leq l \leq n - 1$ . □

**推论 4.4:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $r_0 : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解,  $l$  是  $D_\omega$  中边数最多的路的边数, 则

$$\sum_{i \in V} r_0(i) \leq -\frac{1}{2}l^2 + (n - \frac{1}{2})l \quad (4-15)$$

**证明** 设  $r_0^{max} = \max_{i \in V} r_0(i)$ . 由定理 4.7,  $0 \leq r_0^{max} \leq l$ . 由命题 4.4,

$$\begin{aligned} \sum_{i \in V} r_0(i) &\leq -\frac{1}{2}(r_0^{max})^2 + (n - \frac{1}{2})r_0^{max} \\ &\leq \max \left\{ -\frac{1}{2}k^2 + (n - \frac{1}{2})k \mid 0 \leq k \leq l, k \in \mathbb{Z} \right\} \quad \square \\ &= -\frac{1}{2}l^2 + (n - \frac{1}{2})l \end{aligned}$$

**例 4.2:** 设图  $D = (V, E)$  如图 4.2 所示, agony 最小解在各顶点的值标在节点旁。图  $D$  的顶点数为  $n = 3$ , 最长路边数  $l = 2$ ,  $\max_{i \in V} r_0(i) = 2 = n - 1 = l$ ,  $\sum_{i \in V} r_0(i) = 3 = \frac{1}{2}n^2 - \frac{1}{2}n = -\frac{1}{2}l^2 + (n - \frac{1}{2})l$ 。

**例 4.3:** 设图  $D = (V, E)$  如图 4.3 所示, 节点旁是 agony 最小解的取值。图  $D$  的顶点数为  $n = 2$ , 最长路边数  $l = 1$ ,  $\max_{i \in V} r_0(i) = 0 < n - 1 = l$ ,  $\sum_{i \in V} r_0(i) = 0 < \frac{1}{2}n^2 - \frac{1}{2}n = -\frac{1}{2}l^2 + (n - \frac{1}{2})l$ 。

**例 4.4:** 设图  $D = (V, E)$  和 agony 最小解如图 4.4 所示。图  $D$  的顶点数为  $n = 5$ , 最长路边数为  $l = 3$ ,  $\max_{i \in V} r_0(i) = 1 < l < n - 1$ ,  $\sum_{i \in V} r_0(i) = 2 < -\frac{1}{2}l^2 + (n - \frac{1}{2})l < \frac{1}{2}n^2 - \frac{1}{2}n$ 。

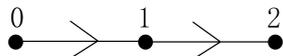


图 4.3 例 4.2

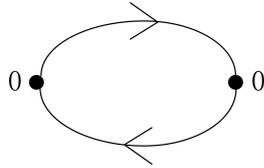


图 4.4 例 4.3

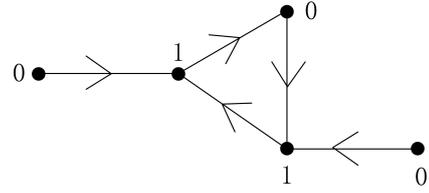


图 4.5 例 4.4

#### 4.2.4 无圈有向图与 agony 最小解

本小节为算法 4 和算法 5 提供证明，并且讨论应用这两个算法能得到非无圈有向图 agony 最小解的条件，对于不满足上述条件的一种特殊情况，给出修改算法。

**命题 4.6 (文献 [25] Proposition 1.4.2):** 设  $D = (V, E)$  是无圈有向图，则图  $D$  中存在入度为 0 的顶点和出度为 0 的顶点（即存在  $i, j \in V$  使  $d_D^-(i) = 0$ ,  $d_D^+(j) = 0$ ）。

**命题 4.7:** 在算法 4 中输入有向图  $D = (V, E)$ ，设输出是  $D_R = (V_R, E_R)$ 。若  $C$  是  $D$  中的一个圈，则  $C$  中的顶点和边满足  $V(C) \subseteq V_R$ ,  $E(C) \subseteq E_R$ 。

**证明** 算法 4 每次循环删除一个点和与之连接的边。因此，只需证明  $V(C) \subseteq V_R$ 。

反证，设  $C = v_1 e_1 v_2 \cdots v_s e_s v_1$ ,  $v_1 \notin V_R$ 。由于  $e_s = (v_s, v_1), e_1 = (v_1, v_2) \in E$ ，删  $v_1$  时已删  $e_s$  或  $e_1$ ，所以删  $v_1$  前已删  $v_s$  或  $v_2$ 。同理可以依次得到，删  $v_s$  前已删  $v_{s-1}$  或  $v_1$ ，删  $v_{s-1}$  前已删  $v_{s-2}$  或  $v_s$  等等，一直到删  $v_2$  前已删  $v_1$  或  $v_3$ 。若删  $v_1$  前未删  $v_s$ ，则可知已删  $v_2$ ，进而已删  $v_3, \cdots, v_{s-1}, v_s$ ，矛盾。因此删  $v_1$  前已删  $v_s$ ，同理可得删  $v_s$  前已删  $v_{s-1}$ ，进而已删  $v_{s-2}, \cdots, v_2, v_1$ ，矛盾。

所以  $C$  中的顶点和边满足  $V(C) \subseteq V_R$ ,  $E(C) \subseteq E_R$ 。 □

**定理 4.8:** 在算法 4 中输入有向图  $D = (V, E)$ ，设输出是  $D_R = (V_R, E_R)$ 。令  $E_n = E - E_R$ ,  $D_n = (V, E_n)$ ，则  $D_n$  是无圈有向图。

**证明** 反证，设  $D_n$  中有圈  $C$ ，则  $C$  也是  $D$  中的圈。根据命题 4.7， $E(C) \subseteq E_R$ ，因此  $e \notin E_n = E - E_R, \forall e \in E(C)$ ，与  $C$  是  $D_n$  中的圈矛盾。

因此， $D_n$  是无圈有向图。 □

**定理 4.9:** 在算法 4 中输入有向图  $D = (V, E)$ , 设输出是  $D_R = (V_R, E_R)$ 。若  $V_R = \emptyset$ , 则  $D$  是无圈有向图; 否则,  $D$  不是无圈有向图。

**证明** 算法 4 的每次循环删一个点, 及与之相连的边。因为  $D$  有限, 所以算法 4 会结束。

若  $V_R = \emptyset$ , 则  $E_R = \emptyset$ 。令  $E_n = E - E_R$ ,  $D_n = (V, E_n)$ , 则  $D = D_n$ 。由定理 4.8,  $D_n$  是无圈有向图, 因此  $D$  是无圈有向图。

若  $V_R \neq \emptyset$ , 断言  $D_R$  不是无圈有向图, 否则由命题 4.6,  $D_R$  中有入度为 0 的点, 算法 4 没有终止, 所以  $D_R$  不是无圈有向图。由于无圈有向图删任意数量顶点后仍然是无圈的, 因此  $D$  不是无圈有向图。

所以, 可以用算法 4 的输出是否空图判断输入有向图是否无圈。  $\square$

**定理 4.10:** 设  $D_\omega = (V, E, \omega)$  是一个无圈有向图,  $r_0$  是  $D_\omega$  的 agony 最小解, 则对任意  $i \in V$  有  $r_0(i)$  等于以  $i$  为终点的路的最大边数 (若  $i$  的入度为 0, 则  $r_0(i) = 0$ )。

**证明** 定义  $r : V \rightarrow \mathbb{N}$  为,  $r(i)$  等于以  $i$  为终点的路的最大边数 (若  $i$  的入度为 0, 则  $r(i) = 0$ )。则  $r$  是一种层级分配方案。

首先证明,  $Q(r) = 0$ 。

任取  $e = (i, j) \in E$ , 设  $p$  是一条以  $i$  为终点的路, 且在以  $i$  为终点的路中达到最大边数。由于  $D_\omega$  无圈, 因此  $p$  不经过  $j$ 。令  $q$  为  $p$  连接  $e$  产生的链, 则  $q$  是一条以  $j$  为终点的路。由  $r$  的定义,  $r(j) \geq q$  的边数 =  $p$  的边数 + 1 =  $r(i) + 1$ 。因此,  $q(e, r) = \max\{r(i) - r(j) + 1, 0\} = 0$ , 进而有  $Q(r) = 0$ 。

因为对任意层级分配方案,  $Q \geq 0$ , 所以  $r$  是 agony 解。

由于  $r_0$  是 agony 最小解, 因此对任意  $i \in V$  有  $r_0(i) \leq r(i)$ , 且  $Q(r_0) = 0$ 。因为  $\omega(e) > 0, \forall e \in E$ , 所以  $q(e, r_0) = 0, \forall e \in E$ 。

然后证明, 对任意  $i \in V$  有  $r_0(i) \geq r(i)$ 。设  $p = v_0 e_1 v_1 \cdots e_s v_s$  是一条以  $i$  为终点的路。因为  $q(e, r_0) = 0, \forall e \in E$ , 所以  $0 \leq r_0(v_0) < r_0(v_1) < \cdots < r_0(v_s) = r_0(i)$ 。又  $r_0(v_0), r_0(v_1), \cdots, r_0(v_s) \in \mathbb{Z}$ , 所以  $r_0(i) \geq s$ 。由  $p$  的任意性与  $r$  的定义知  $r_0(i) \geq r(i)$ 。

所以,  $r = r_0$ , 即对任意  $i \in V$  有  $r_0(i)$  等于以  $i$  为终点的路的最大边数 (若  $i$  的入度为 0, 则  $r_0(i) = 0$ )。  $\square$

**推论 4.5:** 设  $D_\omega = (V, E, \omega)$  是一个无圈有向图,  $r_0$  是  $D_\omega$  的 agony 最小解,  $l$  是  $D_\omega$  中边数最多的路的边数, 则  $\max_{i \in V} r_0(i) = l$ 。

**证明** 根据定理 4.10, 对任意  $i \in V$  有  $r_0(i)$  等于以  $i$  为终点的路的最大边数, 在两侧对  $i \in V$  取最大值即得  $\max_{i \in V} r_0(i) = l$ 。  $\square$

**推论 4.6:** 设  $D_\omega = (V, E, \omega)$  是一个无圈有向图, 则  $D_\omega$  的 agony 最小解与权函数  $\omega : E \rightarrow \mathbb{R}_+$  无关。

**证明** 利用定理 4.10 即得。  $\square$

**命题 4.8:** 设有向图  $D = (V, E)$  无圈,  $i \in V$  满足入度  $d_D^-(i) > 0$ , 则存在  $i_0 \in V$  和  $D$  中从  $i_0$  到  $i$  的路使  $d_D^-(i_0) = 0$ 。

**证明** 设  $p$  是图  $D$  中以  $i$  为终点的最长的路,  $p = v_0 e_1 v_1 \cdots e_s v_s$ , 其中  $v_s = i$ 。

假设  $d_D^-(v_0) > 0$ , 则存在  $v_{-1} \in V$  使  $e_0 = (v_{-1}, v_0) \in E$ 。因为  $D$  无圈, 所以  $p$  不经过  $v_{-1}$ ,  $p_1 = v_{-1} e_0 v_0 e_1 \cdots e_s v_s$  是从  $v_{-1}$  到  $v_s$  的路。路  $p_1$  的边数大于  $p$ , 且以  $i$  为终点, 与  $p$  的定义矛盾。

因此,  $d_D^-(v_0) = 0$ , 取  $i_0 = v_0$  即得证。  $\square$

**定理 4.11:** 输入无圈有向图  $D = (V, E)$  时, 算法 5 的输出是  $D$  的 agony 最小解。

**证明** 第一步, 由算法 5 第 1 至 6 行, 以及之后的步骤没有改变  $R_0$  的事实知,  $R_0 = \{i \in V | d_D^-(i) = 0\}$ 。

第二步, 对任意  $i \in V$ , 证明算法 5 进行有限步之后  $r_0(i)$  有定义。

如果  $d_D^-(i) = 0$ , 那么算法 5 进行到第 7 行时,  $r_0(i)$  已经有定义。显然, 算法 5 的第 1 至 7 行只有有限步。

如果  $d_D^-(i) > 0$ , 由命题 4.8, 存在  $i_0 \in V, d_D^-(i_0) = 0$  和  $D$  中从  $i_0$  到  $i$  的路, 设为  $p = v_0 e_1 v_1 \cdots e_s v_s$ , 其中  $v_0 = i_0, v_s = i$ 。当  $j \geq 1$  时, 由算法 5 第 8 至 16 行知,  $R_j = \{k \in V | \exists l \in R_{j-1}, s. t. (l, k) \in E\}$ 。因为  $v_0 = i_0 \in R_0, e_1 \in E$ , 所以第 1 次 while 循环后,  $v_1 \in R_1$ 。由于  $R_1 \neq \emptyset$ , while 循环继续。类似地可以得到第 2 次 while 循环后,  $v_2 \in R_2$ 。以此类推, 可得第  $s$  次 while 循环后,  $i = v_s \in R_s$ , 此时  $r_0(i)$  已有定义。

因此，对任意  $i \in V$ ，算法 5 进行有限步之后  $r_0(i)$  有定义。

第三步，证明若  $i_N \in R_N, N \geq 1$ ，则存在图  $D$  中以  $i_N$  为终点的路  $p$ ，使其边数为  $N$ 。

因为  $i_N \in R_N = \{k \in V | \exists l \in R_{N-1}, \text{s. t. } (l, k) \in E\}$ ，所以存在  $i_{N-1} \in R_{N-1}, e_N = (i_{N-1}, i_N) \in E$ 。再由  $R_{N-1}$  的定义可知，存在  $i_{N-2} \in R_{N-2}, e_{N-1} = (i_{N-2}, i_{N-1}) \in E$ 。以此类推，可以得到  $i_{N-3} \in R_{N-3}, \dots, i_1 \in R_1, i_0 \in R_0$  以及  $e_{N-2} = (i_{N-3}, i_{N-2}), \dots, e_1 = (i_0, i_1) \in E$ 。令  $p = i_0 e_1 i_1 \dots i_{N-2} e_{N-1} i_{N-1} e_N i_N$ ，由于图  $D$  无圈，所以  $p$  是图  $D$  中的一条路，且边数为  $N$ 。

第四步，证明算法 5 能在有限步后结束。

设 while 循环进行到了第  $N$  次， $N \geq 1$ ，则  $R_{N-1} \neq \emptyset$ ，存在  $i_{N-1} \in R_{N-1}$ 。由第三步知，存在图  $D$  中以  $i_{N-1}$  为终点的路  $p$ ，使其边数为  $N-1$ 。因此， $N-1 \leq n$ ，while 循环最多进行有限次。所以算法 5 能在有限步后结束。

第五步，证明算法 5 的输出  $r_0 : V \rightarrow \mathbb{N}$  满足  $Q(r_0) = 0$ 。

只需证明，对任取  $e = (i_0, j_0) \in E$ ，有  $q(e, r_0) = 0$ 。根据算法 5 中对  $r_0$  的定义过程（即第 4 行和第 12 行）， $i_0 \in R_{r_0(i_0)}$ 。又因为  $e = (i_0, j_0) \in E$ ，所以  $j_0 \in R_{r_0(i_0)+1}$ 。由于算法 5 第 12 行的赋值要么首次定义  $r_0(v_k)$ ，要么增加  $r_0(v_k)$ ，所以  $r_0(j_0) \geq r_0(i_0) + 1$ ，从而  $q(e, r_0) = 0$ 。

第六步，设  $\bar{r} \rightarrow \mathbb{N}$  是图  $D$  的 agony 最小解，证明对任意  $i \in V$ ，有  $r_0(i) \leq \bar{r}(i)$ 。

当  $r_0(i) > 0$  时， $d_D^-(i) > 0$ ，根据定理 4.10， $\bar{r}(i)$  等于以  $i$  为终点的路的最大边数。因为  $i \in R_{r_0(i)}$ ，由第三步知，存在图  $D$  中以  $i$  为终点的路  $p$ ，使其边数为  $r_0(i)$ 。因此， $r_0(i)$  小于等于以  $i$  为终点的路的最大边数，即  $r_0(i) \leq \bar{r}(i)$ 。

当  $r_0(i) = 0$  时， $d_D^-(i) = 0$ ，根据定理 4.10， $\bar{r}(i) = 0 = r_0(i)$ 。

所以，对任意  $i \in V$ ，有  $r_0(i) \leq \bar{r}(i)$ 。

第七步，因为第五步得出  $Q(r_0) = 0$ ，所以  $r_0$  是图  $D$  的 agony 解。又由第六步知，对任意  $i \in V$ ，有  $r_0(i) \leq \bar{r}(i)$ 。由于  $\bar{r}$  是图  $D$  的 agony 最小解，所以  $r_0(i) = \bar{r}(i), \forall i \in V$ ，即算法 5 的输出是  $D$  的 agony 最小解。□

考虑加权有向图  $D_\omega = (V, E, \omega)$ ，其中  $\omega : E \rightarrow \mathbb{R}_+$ 。设  $n = \#V$ ， $m = \#E$ ， $D = (V, E)$ ， $M_{n \times m}^D$  是  $D$  的关联矩阵（见定义 4.2）。由推论 4.1 知， $D_\omega$  的 agony 解是问题 (4-2) 的最优解。

问题 (4-2) 作为线性规划的标准形式是

$$\begin{aligned}
 \min \quad & \sum_{(i,j) \in E} \omega(i,j)x(i,j) \\
 \text{s. t.} \quad & x(i,j) - s(i,j) - r(i) + r(j) = 1, \forall (i,j) \in E \\
 & x(i,j) \geq 0, \forall (i,j) \in E \\
 & s(i,j) \geq 0, \forall (i,j) \in E \\
 & r(i) \geq 0, \forall i \in V
 \end{aligned} \tag{4-16}$$

写成矩阵形式是

$$\begin{aligned}
 \min \quad & c^T y \\
 \text{s. t.} \quad & Ay = b \\
 & y \geq 0
 \end{aligned} \tag{4-17}$$

其中

$$y = \begin{bmatrix} x_{m \times 1} \\ s_{m \times 1} \\ r_{n \times 1} \end{bmatrix}, A = \begin{bmatrix} I_{m \times m} \\ -I_{m \times m} \\ -M_{n \times m}^D \end{bmatrix}^T, b = 1_{m \times 1}, c = \begin{bmatrix} \omega_{m \times 1} \\ 0_{m \times 1} \\ 0_{n \times 1} \end{bmatrix}$$

问题 (4-17) 的对偶是

$$\begin{aligned}
 \max \quad & b^T f \\
 \text{s. t.} \quad & A^T f + t = c \\
 & t \geq 0
 \end{aligned} \tag{4-18}$$

等价于

$$\begin{aligned}
 \max \quad & \sum_{(i,j) \in E} f(i,j) \\
 \text{s. t.} \quad & f(i,j) + t_1(i,j) = \omega(i,j), \forall (i,j) \in E \\
 & -f(i,j) + t_2(i,j) = 0, \forall (i,j) \in E \\
 & -\sum_{j:(i,j) \in E} f(i,j) + \sum_{j:(j,i) \in E} f(j,i) + t_3(i) = 0, \forall i \in V \\
 & t_1(i,j), t_2(i,j), t_3(k) \geq 0, \forall (i,j) \in E, k \in V
 \end{aligned} \tag{4-19}$$

设  $y$  是问题 (4-17) 的可行解,  $(f, t)$  是问题 (4-18) 的可行解。则  $c^T y - b^T f = 0$  当且仅当  $y$  和  $(f, t)$  分别是问题 (4-17) 和问题 (4-18) 的最优解。

因为  $c^T y - b^T f = c^T y - (Ay)^T f = y^T(c - A^T f)$ , 所以  $c^T y - b^T f = 0$  等价于  $y^T(c - A^T f) = 0$ , 即

$$x(i, j)(\omega(i, j) - f(i, j)) = 0, \forall (i, j) \in E \quad (4-20a)$$

$$s(i, j)f(i, j) = 0, \forall (i, j) \in E \quad (4-20b)$$

$$r(i) \left( \sum_{j:(i,j) \in E} f(i, j) - \sum_{j:(j,i) \in E} f(j, i) \right) = 0, \forall i \in V \quad (4-20c)$$

**定义 4.3 (循环<sup>[25]</sup>):** 设  $D = (V, E)$  是一个有向图, 函数  $g : E \rightarrow \{0\} \cup \mathbb{R}_+$  满足  $\sum_{j:(i,j) \in E} g(i, j) - \sum_{j:(j,i) \in E} g(j, i) = 0, \forall i \in V$ 。则称  $g$  是  $D$  中的一个循环。

**定义 4.4 (圈流<sup>[25]</sup>):** 设  $D = (V, E)$  是一个有向图, 函数  $g : E \rightarrow \{0\} \cup \mathbb{R}_+$  满足存在  $D$  中圈  $C$  和  $k \in \{0\} \cup \mathbb{R}_+$  使  $g(e) = k, \forall e \in E(C)$  并且  $g(e) = 0, \forall e \in E - E(C)$ 。则称  $g$  是  $D$  中的一个圈流。

**定理 4.12 (文献 [25] Corollary 3.3.3):** 设  $D = (V, E)$  是一个有向图, 则  $D$  中的循环能分解为有限个圈流在每条边上的和。

**命题 4.9:** 设  $D_\omega = (V, E, \omega)$  是加权有向图, 其中  $\omega : E \rightarrow \mathbb{R}_+$ 。在算法 4 中输入  $D_\omega$ , 设输出是  $D_R = (V_R, E_R)$ 。设对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i, j) = \sum_{j:(j,i) \in E_R} \omega(j, i)$ 。定义  $f_0 : E \rightarrow \mathbb{R}$  为

$$f_0(e) = \begin{cases} \omega(e), \forall e \in E_R \\ 0, \forall e \in E - E_R \end{cases} \quad (4-21)$$

则存在  $t_0$ , 使得  $(f_0, t_0)$  是问题 (4-19) 的最优解。

**证明** 对任意  $e \in E$ , 有  $0 \leq f_0(e) \leq \omega(e)$ ; 对任意  $i \in V$ , 有

$$\sum_{j:(i,j) \in E} f_0(i, j) - \sum_{j:(j,i) \in E} f_0(j, i) = \sum_{j:(i,j) \in E_R} \omega(i, j) - \sum_{j:(j,i) \in E_R} \omega(j, i) = 0$$

由此易知, 存在  $t_0$  使  $(f_0, t_0)$  是问题 (4-19) 的可行解。

设  $(f, t)$  是问题 (4-19) 的最优解。由  $f(e) + t_1(e) = \omega(e), -f(e) + t_2(e) = 0, t_1(e), t_2(e) \geq 0, \forall e \in E$ , 知  $0 \leq f(e) \leq \omega(e), \forall e \in E$ 。

在等式  $-\sum_{j:(i,j) \in E} f(i, j) + \sum_{j:(j,i) \in E} f(j, i) + t_3(i) = 0, \forall i \in V$  中对  $i \in V$  求和得  $\sum_{i \in V} t_3(i) = 0$ 。因为  $t_3(i) \geq 0, \forall i \in V$ , 所以  $t_3(i) = 0, -\sum_{j:(i,j) \in E} f(i, j) + \sum_{j:(j,i) \in E} f(j, i) = 0, \forall i \in V$ ,  $f$  是  $D_\omega$  中的循环。

由定理 4.12, 存在分别在圈  $C_1, C_2, \dots, C_s$  中的圈流  $g_1, g_2, \dots, g_s$  使得  $f(e) = g_1(e) + g_2(e) + \dots + g_s(e), \forall e \in E$ 。

根据命题 4.7,  $E(C_1) \cup E(C_2) \cup \dots \cup E(C_s) \subseteq E_R$ 。因此,  $f(e) = 0, \forall e \in E - E_R$ ,

$$\sum_{e \in E} f(e) = \sum_{e \in E_R} f(e) \leq \sum_{e \in E_R} \omega(e) = \sum_{e \in E_R} f_0(e) = \sum_{e \in E} f_0(e)$$

所以,  $(f_0, t_0)$  是问题 (4-19) 的一个最优解。  $\square$

**命题 4.10:** 设  $D_\omega = (V, E, \omega)$  是加权有向图, 其中  $\omega : E \rightarrow \mathbb{R}_+$ 。在算法 4 中输入  $D_\omega$ , 设输出是  $D_R = (V_R, E_R)$ 。令  $D_n = (V, E - E_R)$ ,  $r_0$  是  $D_n$  的 agony 最小解,  $\bar{r}$  是  $D_\omega$  的 agony 最小解。设对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i, j) = \sum_{j:(j,i) \in E_R} \omega(j, i)$ , 则  $q(e, \bar{r}) = 0, \forall e \in E - E_R, r_0(i) \leq \bar{r}(i), \forall i \in V$ , 且  $\bar{r}(i) \geq \bar{r}(j) - 1, \forall e = (i, j) \in E_R$ 。

**证明** 令  $f_0 : E \rightarrow \mathbb{R}$  如命题 4.9 中定义, 则存在  $t_0$ , 使得  $(f_0, t_0)$  是问题 (4-19) 的最优解。

因为  $\bar{r}$  是  $D_\omega$  的 agony 最小解, 所以存在  $\bar{x}, \bar{s}$  使  $(\bar{x}, \bar{s}, \bar{r})$  是问题 (4-16) 的最优解。因此,  $(f_0, t_0)$  和  $(\bar{x}, \bar{s}, \bar{r})$  满足公式 (4-20)。

由公式 (4-20a) 与  $\omega(e) > 0, \forall e \in E$  可知,  $\bar{x}(e) = 0, \forall e \in E - E_R$ , 因此  $q(e, \bar{r}) = 0, \forall e \in E - E_R, \bar{r}$  是  $D_n$  的 agony 解。因为  $r_0$  是  $D_n$  的 agony 最小解, 所以  $r_0(i) \leq \bar{r}(i), \forall i \in V$ 。

由公式 (4-20b) 与  $f_0(i, j) = \omega(i, j) > 0, \forall e = (i, j) \in E_R$  可知,  $\bar{s}(i, j) = 0, \forall e = (i, j) \in E_R$ , 因此  $\bar{r}(i) \geq \bar{r}(j) - 1, \forall e = (i, j) \in E_R$ 。  $\square$

**命题 4.11:** 设  $D_\omega = (V, E, \omega)$  是加权有向图,  $\omega : E \rightarrow \mathbb{R}_+$ 。在算法 4 中输入  $D_\omega$ , 设输出是  $D_R = (V_R, E_R)$ 。设对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i, j) = \sum_{j:(j,i) \in E_R} \omega(j, i)$ 。设  $r : V \rightarrow \mathbb{N}$ , 满足  $q(e, r) = 0, \forall e \in E - E_R$ , 并且  $r(i) \geq r(j) - 1, \forall e = (i, j) \in E_R$ 。则  $r$  是  $D_\omega$  的 agony 解。

**证明** 令  $f_0 : E \rightarrow \mathbb{R}$  如命题 4.9 中定义, 则存在  $t_0$  使得  $(f_0, t_0)$  是问题 (4-19) 的最优解。令  $x(e) = q(e, r), s(e) = q(e, r) - (r(i) - r(j) + 1), \forall e = (i, j) \in E$ 。

对  $(f_0, t_0)$  和  $(x, s, r)$  验证公式 (4-20):

因为  $\sum_{j:(i,j) \in E} f_0(i, j) - \sum_{j:(j,i) \in E} f_0(j, i) = 0, \forall i \in V$ , 所以公式 (4-20c) 成立。

当  $e = (i, j) \in E - E_R$  时,  $x(e) = q(e, r_0) = 0, f_0(e) = 0$ , 所以公式 (4-20a) 和公式 (4-20b) 对任意  $e \in E - E_R$  成立。

当  $e = (i, j) \in E_R$  时,  $\omega(e) - f_0(e) = 0$ , 所以公式 (4-20a) 成立。因为  $r(i) - r(j) + 1 \geq 0$ , 所以  $s(e) = 0$ , 公式 (4-20b) 成立。

因此,  $(x, s, r)$  是问题 (4-16) 的一个最优解,  $r$  是  $D_\omega$  的 agony 解。  $\square$

**定理 4.13:**  $D_\omega = (V, E, \omega)$  是加权有向图, 其中  $\omega : E \rightarrow \mathbb{R}_+$ 。在算法 4 中输入  $D_\omega$ , 设输出是  $D_R = (V_R, E_R)$ 。设  $D_n = (V, E - E_R)$ , 则  $D_n$  是无圈有向图。令  $D_n$  的 agony 最小解是  $r_0 : V \rightarrow \mathbb{N}$ 。如果对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i, j) = \sum_{j:(j,i) \in E_R} \omega(j, i)$ , 对任意  $e = (i_1, i_2) \in E_R$ , 有  $r_0(i_1) - r_0(i_2) + 1 \geq 0$ 。则  $r_0$  是  $D_\omega$  的 agony 最小解。

**证明** 由定理 4.8 知,  $D_n$  是无圈有向图。又因为  $r_0$  是  $D_n$  的 agony 最小解, 所以  $q(e, r_0) = 0, \forall e \in E - E_R$  (由定理 4.10 的证明过程可知)。

又因为  $r_0$  满足  $r_0(i) \geq r_0(j) - 1, \forall e = (i, j) \in E_R$ , 根据命题 4.11,  $r_0$  是  $D_\omega$  的 agony 解。

设  $\bar{r} : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解。根据命题 4.10,  $r_0(i) \leq \bar{r}(i), \forall i \in V$ 。

因为  $\bar{r}$  是  $D_\omega$  的 agony 最小解,  $r_0$  是  $D_\omega$  的 agony 解, 所以  $r_0(i) \geq \bar{r}(i), \forall i \in V$ , 进而  $r_0(i) = \bar{r}(i), \forall i \in V$ , 即  $r_0$  是  $D_\omega$  的 agony 最小解。  $\square$

如果定理 4.13 的条件中, 仅  $r_0(i_1) - r_0(i_2) + 1 \geq 0, \forall e = (i_1, i_2) \in E_R$  不成立, 则可以通过算法 6, 将  $r_0$  修改为  $D_\omega$  的 agony 最小解。

**定理 4.14:** 设  $D_\omega = (V, E, \omega)$  是加权有向图, 其中  $\omega : E \rightarrow \mathbb{R}_+$ 。在算法 4 中输入  $D_\omega$ , 输出是  $D_R = (V_R, E_R)$ 。设  $D_n = (V, E - E_R)$ , 则  $D_n$  是无圈有向图。在算法 5 中输入  $D_n$ , 输出是  $r_0 : V \rightarrow \mathbb{N}$ 。设对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i, j) = \sum_{j:(j,i) \in E_R} \omega(j, i)$ 。在算法 6 中输入  $D_\omega, D_R, r_0$ , 则算法 6 在有限步后结束, 输出  $r : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解。

---

**算法 6** 修改基于无圈有向图得出的层级分配方案

---

输入: 加权有向图  $D_\omega = (V, E, \omega)$ 、算法 4 的输出  $D_R = (V_R, E_R)$ 、算法 5 的输出

$r_0 : V \rightarrow \mathbb{N}$  (要求满足对任意  $i \in V_R$ , 有  $\sum_{j:(i,j) \in E_R} \omega(i,j) = \sum_{j:(j,i) \in E_R} \omega(j,i)$ )

输出:  $D_\omega$  的 agony 最小解  $r : V \rightarrow \mathbb{N}$

```
1:  $r \leftarrow r_0$ 
2: while  $\exists e_0 = (i_0, j_0) \in E_R, \text{ s. t. } r(i_0) < r(j_0) - 1$  do
3:    $E_S \leftarrow \{e = (i, j) \in E - E_R \mid r(i) \geq r(j) - 1\}$ 
4:    $D_S \leftarrow (V, E_S)$ 
5:    $V_{i_0} \leftarrow \{k \in V \mid \text{在 } D_S \text{ 中 } i_0 \text{ 可达 } k\}$  (包含  $i_0$ )
6:   对所有  $k \in V_{i_0}$ ,  $r(k) \leftarrow r(k) + 1$ 
7: end while
8: return  $r$ 
```

---

**证明** 定理 4.8 说明  $D_n$  是无圈有向图。定理 4.11 及其证明过程说明  $r_0$  是  $D_n$  的 agony 最小解, 并且  $q(e, r_0) = 0, \forall e \in E - E_R$ 。

设  $\bar{r}$  是  $D_\omega$  的 agony 最小解。命题 4.10 说明,  $q(e, \bar{r}) = 0, \forall e \in E - E_R$ ,  $r_0(i) \leq \bar{r}(i), \forall i \in V$  并且  $\bar{r}(i) \geq \bar{r}(j) - 1, \forall e = (i, j) \in E_R$ 。

由于  $r_0(i) \in \mathbb{Z}, \forall i \in V$ , 算法 6 中对  $r$  的操作只有赋值  $r_0$  和增加 1, 因此  $r(i) \in \mathbb{Z}, \forall i \in V$ 。

第一步, 证明每次循环结束时,  $r$  满足  $q(e, r) = 0, \forall e \in E - E_R$ 。

记第  $l$  次循环结束时的  $r$  为  $r^{(l)}$ ,  $r^{(0)} = r_0$ 。对  $l$  归纳。  $l = 0$  时,  $q(e, r^{(l)}) = q(e, r_0) = 0, \forall e \in E - E_R$ 。假设  $r^{(l-1)}$  满足  $q(e, r^{(l-1)}) = 0, \forall e \in E - E_R$ 。按照算法 6, 有

$$r^{(l)}(k) = \begin{cases} r^{(l-1)}(k), & \forall k \in V - V_{i_0} \\ r^{(l-1)}(k) + 1, & \forall k \in V_{i_0} \end{cases}$$

对任意  $e = (i, j) \in E - E_R$  分以下几种情况讨论:

(1)  $i, j \in V - V_{i_0}$ , 则  $q(e, r^{(l)}) = \max\{r^{(l)}(i) - r^{(l)}(j) + 1, 0\} = \max\{r^{(l-1)}(i) - r^{(l-1)}(j) + 1, 0\} = q(e, r^{(l-1)}) = 0$ 。

(2)  $i, j \in V_{i_0}$ , 则  $q(e, r^{(l)}) = \max\{r^{(l)}(i) - r^{(l)}(j) + 1, 0\} = \max\{(r^{(l-1)}(i) + 1) -$

$$(r^{(l-1)}(j) + 1) + 1, 0\} = \max\{r^{(l-1)}(i) - r^{(l-1)}(j) + 1, 0\} = q(e, r^{(l-1)}) = 0。$$

(3)  $i \in V - V_{i_0}, j \in V_{i_0}$ , 则  $q(e, r^{(l)}) = \max\{r^{(l)}(i) - r^{(l)}(j) + 1, 0\} = \max\{r^{(l-1)}(i) - (r^{(l-1)}(j) + 1) + 1, 0\} \leq \max\{r^{(l-1)}(i) - r^{(l-1)}(j) + 1, 0\} = q(e, r^{(l-1)}) = 0。$

(4)  $i \in V_{i_0}, j \in V - V_{i_0}$ 。因为  $j \in V - V_{i_0}$ , 所以在  $D_S$  中  $i_0$  不可达  $j$ 。又由于在  $D_S$  中  $i_0$  可达  $i$ , 因此  $e = (i, j) \notin E_S$ 。因为  $e = (i, j) \in E - E_R$ , 所以由  $E_S$  的定义知  $r^{(l-1)}(i) < r^{(l-1)}(j) - 1$ , 因此  $r^{(l-1)}(i) \leq r^{(l-1)}(j) - 2$ 。  $q(e, r^{(l)}) = \max\{r^{(l)}(i) - r^{(l)}(j) + 1, 0\} = \max\{(r^{(l-1)}(i) + 1) - r^{(l-1)}(j) + 1, 0\} = 0。$

所以,  $q(e, r^{(l)}) = 0, \forall e \in E - E_R$ 。根据数学归纳法, 每次循环结束时,  $r$  满足  $q(e, r) = 0, \forall e \in E - E_R$ 。

第二步, 设  $\bar{r} : V \rightarrow \mathbb{N}$  是  $D_\omega$  的 agony 最小解, 证明每次循环结束时,  $r(i) \leq \bar{r}(i), \forall i \in V$ 。

同样对循环次数  $l$  归纳。  $l = 0$  时,  $r^{(l)}(i) = r_0(i) \leq \bar{r}(i), \forall i \in V$ 。

假设  $r^{(l-1)}$  满足  $r^{(l-1)}(i) \leq \bar{r}(i), \forall i \in V$ 。因为算法 6 循环中  $e_0 = (i_0, j_0) \in E_R$ , 所以由命题 4.10 有  $\bar{r}(i_0) \geq \bar{r}(j_0) - 1$ 。又因为  $r^{(l-1)}(i_0) < r^{(l-1)}(j_0) - 1$ , 所以  $r^{(l)}(i_0) = r^{(l-1)}(i_0) + 1 < r^{(l-1)}(j_0) \leq \bar{r}(j_0) \leq \bar{r}(i_0) + 1$ , 即  $r^{(l)}(i_0) < \bar{r}(i_0) + 1$ 。因为  $r^{(l)}(i_0), \bar{r}(i_0) \in \mathbb{Z}$ , 所以  $r^{(l)}(i_0) \leq \bar{r}(i_0)$ 。

对任意  $k \in V_{i_0}$ , 存在  $D_S$  中路  $p = v_0 e_1 v_1 \cdots e_s v_s$  使  $v_0 = i_0, v_s = k$ 。因为由命题 4.10 有  $q(e, \bar{r}) = 0, \forall e \in E - E_R, E_S \subseteq E - E_R$ , 所以  $\bar{r}(v_i) \geq \bar{r}(v_{i-1}) + 1, i = 1, 2, \dots, s$ 。因此,  $\bar{r}(k) = \bar{r}(v_s) \geq \bar{r}(v_{s-1}) + 1 \geq \cdots \geq \bar{r}(v_0) + s = \bar{r}(i_0) + s \geq r^{(l)}(i_0) + s$ 。因为  $r^{(l-1)}(i) \geq r^{(l-1)}(j) - 1, \forall e = (i, j) \in E_S$ , 所以  $r^{(l)}(k) = r^{(l)}(v_s) = r^{(l-1)}(v_s) + 1 \leq r^{(l-1)}(v_{s-1}) + 2 \leq \cdots \leq r^{(l-1)}(v_0) + s + 1 = r^{(l)}(v_0) + s = r^{(l)}(i_0) + s \leq \bar{r}(k)$ 。

根据数学归纳法, 每次循环结束时,  $r(i) \leq \bar{r}(i), \forall i \in V$ 。

第三步, 证明有限次循环后, 算法 6 结束。

对于每次循环, 存在  $e_0 = (i_0, j_0) \in E_R$  使  $r(i_0) < r(j_0) - 1$ 。根据  $V_{i_0}$  的定义,  $i_0 \in V_{i_0}$ , 因此每次循环后  $r$  至少有一个分量增加。又因为  $r$  各分量不减, 且每次循环结束时  $r(i) \leq \bar{r}(i), \forall i \in V$ , 所以有限次循环后, 算法 6 结束。

第四步, 证明算法 6 的输出  $r = \bar{r}$ 。

由算法 6 中循环的结束条件知, 输出的  $r$  满足  $r(i) \geq r(j) - 1, \forall e = (i, j) \in E_R$ 。由第一步可知,  $q(e, r) = 0, \forall e \in E - E_R$ 。根据命题 4.11,  $r$  是  $D_\omega$  的 agony 解。

由第二步知,  $r(i) \leq \bar{r}(i), \forall i \in V$ 。因为  $\bar{r}$  是  $D_\omega$  的 agony 最小解, 所以算法 6 输出  $r$  是  $D_\omega$  的 agony 最小解。  $\square$

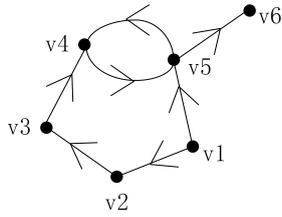


图 4.6 例 4.5 图  $D$

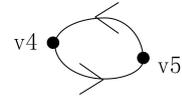


图 4.7 例 4.5 图  $D_R$

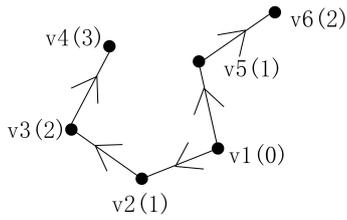


图 4.8 例 4.5 图  $D_n$  的 agony 最小解

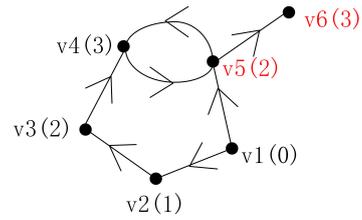


图 4.9 例 4.5 图  $D$  的 agony 最小解

例 4.5: 设有向图  $D = (V, E)$  的顶点和边如图 4.6 所示, 每条边的权都是 1。将图  $D$  输入算法 4, 输出图  $D_R = (V_R, E_R)$  如图 4.7 所示, 因此图  $D$  不是无圈有向图。设  $D_n = (V, E - E_R)$ , 将  $D_n$  输入算法 5, 结果如图 4.8 所示 (图  $D_n$  的 agony 最小解  $r_0$  标在顶点旁的括号内)。因为  $D_R$  各顶点出度等于入度,  $D_n$  中边  $(v5, v4)$  不满足  $r_0(v5) \geq r_0(v4) - 1$ , 所以  $r_0$  不是图  $D$  的 agony 最小解。将图  $D$ 、 $D_R$  和  $r_0$  输入算法 6, 在第一次 while 循环中, 将  $v5$  和  $v6$  的等级提高 1, 发现满足循环结束条件, 输出  $r$  如图 4.9 所示。此时  $r$  是图  $D$  的 agony 最小解。

## 第5章 实验结果与分析

第3章已经使用3个数据集得到多个引用网络。本章对这些网络分别应用现有指标和基于层级结构特性的指标，得到相应的期刊评价结果并对比分析。

### 5.1 数学期刊引用网络

本小节实验采用从 Web of Science 数据库 (WoS) 中收集的期刊引用网络信息，即数学类别中 311 种期刊在 2015 年获得的对其 2013 年和 2014 年出版文章的引用信息。

为获得对该网络的直观感受，首先将部分信息可视化。

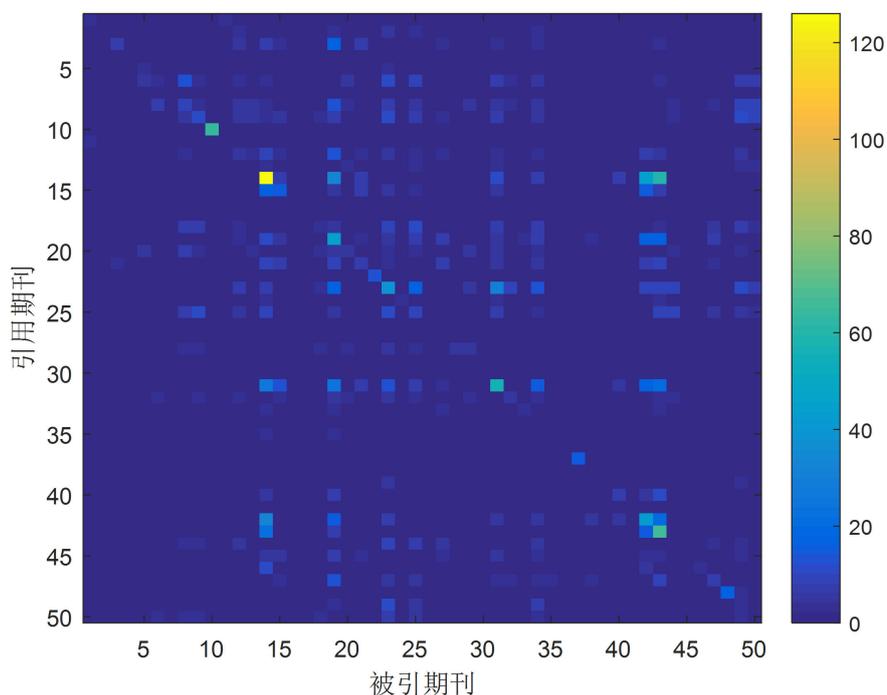


图 5.1 数学期刊引用网络 (部分)

前 50 种期刊之间引用次数关系如图 5.1 所示，其中各种期刊按照 WoS 给出的影响因子排序。图中对角线表示自引现象。

图 5.2 和图 5.3 分别表示该网络中各期刊发表文章总数和被引量总数。

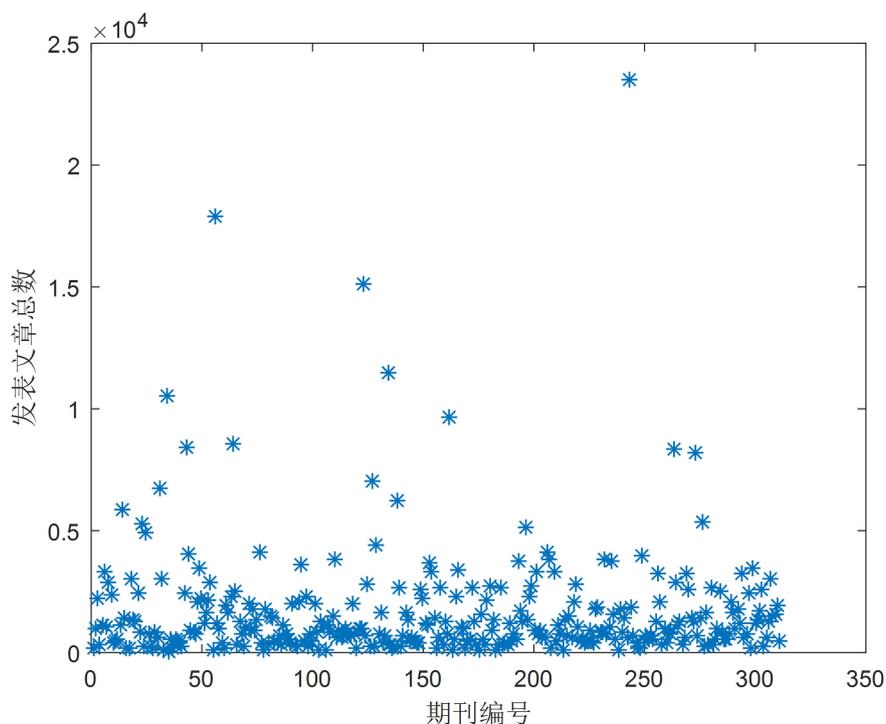


图 5.2 数学期刊引用网络发表文章总数

图 5.2 表明发表文章总数很分散，最少的是 57 篇，最多的是 23500 篇，并且 WoS 的影响因子排序与发表文章总数没有明显关系，实际上相关系数是  $-0.0272$ 。图 5.3 中标出了被引量总数和 WoS 影响因子排序的线性拟合曲线，二者有较弱的相关性，相关系数是  $-0.2885$ 。对比图 5.2 和图 5.3，总体而言被引量总数小于发表文章总数，这有两大主要原因：第一，数学领域期刊的引用半衰期较长，文章发表后第一二年获得的引用量总数在长时段中占比很小。第二，因为数学是基础学科，对数学领域文献的引用很大一部分来源于其他学科，比如物理学、计算机科学、经济学等等，而这些引用信息没有被包含在此处使用的期刊引用网络内。因此，本节计算的各种指标只可能反映各个文献在数学学科内的影响力。

设  $JC$  是表示期刊引用信息的矩阵，行数和列数均为期刊总数，行代表被引期刊，列代表引用期刊，元素的数值是引用次数。设  $JC_n$  是将  $JC$  的对角线元素设为 0 得到的矩阵，表示去除期刊自引之后的期刊引用信息。设  $JC_i$  是与  $JC_n$  维数相同的 0-1 矩阵，且  $JC_i$  中元素等于 1 当且仅当  $JC_n$  中对应元素大于

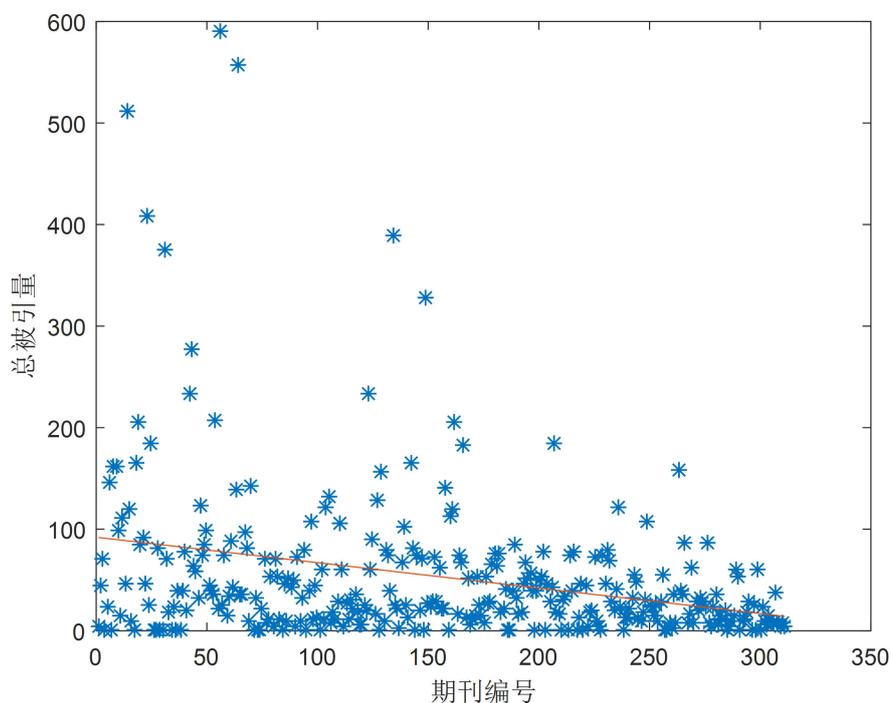


图 5.3 数学期刊引用网络被引量总数

等于  $i$ 。其中  $i$  是一个正整数，将其称为  $JC_i$  的阈值。则从  $JC_n$  到  $JC_i$  是将加权的期刊引用网络通过阈值变成不加权有向网络的过程。

对该期刊引用网络计算影响因子、PageRank 指标、 $\beta$ -measure、LRC 指标、agony 指标及其相关形式，如表 5.1 中所示。

表 5.2 和表 5.3 展示了在各指标中排名前 10 的期刊的缩写表示。

从表 5.2 中可以看出，两个影响因子指标  $IF$  和  $IF_n$  得到的前 10 名期刊相似度很高。 $IF$  前 10 名的 11 种期刊中有 8 种在  $IF_n$  的前 10 名中出现。并且，Bulletin of Mathematical Sciences (Bull. Math. Sci.)、Journal of Topology (J. Topol.) 和 Advances in Nonlinear Analysis (Adv. Nonlinear Anal.) 在  $IF$  和  $IF_n$  中的排序均为 1、2、3。但是，它们在 WoS 中按影响因子排序分别是 35、28、55。这又一次表明本小节使用的数学期刊引用网络只能反映期刊在数学领域内的影响力。注意到 J. Topol. 和 Adv. Nonlinear Anal. 是针对数学学科中某些专业领域的期刊，说明专业性强的期刊也能在这两个指标下取得很好的成绩。将 PageRank 指标  $P$  和两个影响因子指标的前 10 名期刊结果对比，可以看出结果不重合。说明 PageRank 指标反映的期刊特性与影响因子指标有一定差异。加入几个基

表 5.1 基于期刊引用网络计算的指标

指标类型	符号	计算方法
影响因子	$IF$	对 $JC$ 用公式 (3-4)
影响因子 (除自引)	$IF_n$	对 $JC_n$ 用公式 (3-4)
加权 PageRank 指标	$P$	对 $JC_n$ 用算法 2
$\beta$ -measure (不加权)	$\beta_1$	对 $JC_1$ 用定义 2.1
$\beta$ -measure (不加权, 阈值 $i$ )	$\beta_i$	对 $JC_i$ 用定义 2.1
$\beta$ -measure (加权)	$\beta_\omega$	对 $JC_n$ 用定义 2.2
LRC 指标 (不加权)	$C_{R1}$	对 $JC_1$ 用小节 3.2.5 的不加权 LRC 计算方法
LRC 指标 (不加权, 阈值 $i$ )	$C_{Ri}$	对 $JC_i$ 用小节 3.2.5 的不加权 LRC 计算方法
LRC 指标 (加权)	$C_{R\omega}$	对 $JC_n$ 用算法 3
Agony 指标 (不加权)	$r_1$	对 $JC_1$ 求解 (3-8)
Agony 指标 (不加权, 阈值 $i$ )	$r_i$	对 $JC_i$ 求解 (3-8)
Agony 指标 (加权)	$r_\omega$	对 $JC_n$ 求解 (3-8)

表 5.2 数学期刊引用网络现有指标前 10 期刊

排名	$IF$	$IF_n$	$P$
1	Bull. Math. Sci.	Bull. Math. Sci.	Comput. Complex.
2	J. Topol.	J. Topol.	Theor. Comput. Syst.
3	Adv. Nonlinear Anal.	Adv. Nonlinear Anal.	Adv. Math.
4	Fract. Calc. Appl. Anal.	J. Eur. Math. Soc.	Discret. Math.
5	J. Nonlinear Sci. Appl.	Adv. Nonlinear Stud.	Electron. J. Comb.
6	J. Funct. space	Adv. Calc. Var.	J. Comb. Theory Ser. A
7	ARS Math. Contemp.	Evol. Equ. Control Theory	J. Algebra
8	Adv. Nonlinear Stud.	J. Geom. Anal.	Invent. Math.
9	J. Eur. Math. Soc.	J. Funct. space	Math. Ann.
10	Adv. Calc. Var., Evol. Equ. Control Theory	Complex Anal. Theory	Oper. Eur. J. Comb.

表 5.3 数学期刊引用网络层级结构指标前 10 期刊

排名	$\beta_\omega$	$C_{R\omega}$	$r_\omega$
1	J. Algebra	Linear Alg. Appl.	Ann. Funct. Anal.
2	J. Inequal. Appl.	J. Differ. Equ.	Z. Anal. ihre. Anwend.
3	Linear Alg. Appl.	J. Math. Anal. Appl.	J. Inverse Ill-Posed Probl., Funkc. Ekvacioj-Ser. Int.
4	J. Funct. Anal.	J. Algebra	
5	Ergod. Theory Dyn. Syst.	J. Funct. Anal.	Mem. Am. Math. Soc., Publ. Res. Inst. Math. Sci., Mich. Math. J., J. Symplectic Geom., Rep- resent. Theory, Anal. Math. Phys., Nagoya Math. J., Kyushu J. Math., Comput. Complex.
6	Discret. Math.	Nonlinear Anal.-Theory Methods Appl.	
7	Discret. Comput. Geom.	Adv. Math.	
8	Int. J. Number Theory	Duke Math. J.	
9	Commun. Pure Appl. Anal	Qual. Theor. Dyn. Syst.	
10	Calc. Var. Partial Differ. Equ.	Bound. Value Probl.	

于加权引用网络的层级结构指标做对比，可以发现这些指标之间存在联系。例如，JOURNAL OF ALGEBRA (J. Algebra) 同时在  $P$ 、 $\beta_\omega$  和  $C_{R\omega}$  的前 10 名中出现，并且在  $\beta_\omega$  的排序中排第一；LINEAR ALGEBRA AND ITS APPLICATIONS (Linear Alg. Appl.) 同时在  $\beta_\omega$  和  $C_{R\omega}$  的前 10 名中出现，并且在  $C_{R\omega}$  的排序中排第一；COMPUTATIONAL COMPLEXITY (Comput. Complex.) 同时在  $P$  和  $r_\omega$  的前 10 名中出现，并且在  $P$  的排序中排第一。然而，总体而言不同指标之间前 10 名排序重合度比较低。 $IF$ 、 $IF_n$  和  $C_{R\omega}$  选出的前 10 名在 WoS 的影响因子排序中也比较靠前，但其他指标与 WoS 影响因子联系较不明显。

对已计算的所有指标，采用斯皮尔曼等级相关系数 (Spearman rank-order correlations) 反映它们之间的联系<sup>[2]</sup>。做法为，由各指标取得期刊的排序 (获得同样分数的期刊取可能排序的平均值)，计算各指标排序向量的线性相关系数矩阵。其中指标  $IF$ 、 $IF_n$ 、 $P$ 、 $\beta_\omega$ 、 $C_{R\omega}$ 、 $r_\omega$  的相关系数如矩阵 (5-1) 所示，并且在矩阵 (5-1) 右侧已标记对应指标。

$$\begin{bmatrix}
 1.0000 & 0.9703 & 0.4870 & 0.5779 & 0.4321 & 0.2795 \\
 0.9703 & 1.0000 & 0.5105 & 0.5750 & 0.4268 & 0.3555 \\
 0.4870 & 0.5105 & 1.0000 & 0.8280 & 0.3660 & 0.4949 \\
 0.5779 & 0.5750 & 0.8280 & 1.0000 & 0.4202 & 0.3197 \\
 0.4321 & 0.4268 & 0.3660 & 0.4202 & 1.0000 & 0.1714 \\
 0.2795 & 0.3555 & 0.4949 & 0.3197 & 0.1714 & 1.0000
 \end{bmatrix}
 \begin{matrix}
 IF \\
 IF_n \\
 P \\
 \beta_\omega \\
 C_{R\omega} \\
 r_\omega
 \end{matrix}
 \quad (5-1)$$

其中，除指标  $C_{R\omega}$  和  $r_\omega$  以外 (它们之间相关系数是 0.1714，如果它们实际非线性相关，则观察到不小于 0.1714 的相关系数的概率是 0.0024)，其它两指标之间的线性相关性都很显著 (出现概率小于  $10^{-5}$ )。这说明以上指标之间都有一定关联，在某指标中排名较高，则在另一指标中也倾向于取得较高的排名。因为这些指标中的影响因子和 PageRank 指标的有效性已经过许多相关研究的检验，所以以上基于层级结构的指标都能在某种程度上反映期刊的质量。但是它们的结果与现有指标并不完全相同。

矩阵 (5-1) 中相关系数较高的有  $IF$  和  $IF_n$ 、 $P$  和  $\beta_\omega$ 。 $IF$  和  $IF_n$  都是影响因子指标，因此排序结果比较接近。而  $P$  和  $\beta_\omega$  定义有类似性，都考虑了每种期刊的被引总量，与引用者的引用量。

利用矩阵 (5-1)，对各指标的排序结果做主成分分析，各主成分对应方差比例如表 5.4 所示。

表 5.4 数学引用期刊网络各指标排序主成分对应方差比例

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>	<i>PC5</i>	<i>PC6</i>
对应方差比例	58.2%	15.9%	12.4%	10.7%	2.3%	0.4%
累积方差比例	58.2%	74.2%	86.6%	97.3%	99.6%	100%

投影所有的指标至前两个主成分 *PC1* 和 *PC2*，结果如表 5.5 所示。之后，在图 5.4 中标出各指标。

表 5.5 数学引用期刊网络各指标投影至主成分空间

投影	<i>IF</i>	<i>IF<sub>n</sub></i>	<i>P</i>	$\beta_\omega$	$C_{R\omega}$	$r_\omega$
<i>PC1</i>	7.961	6.853	-5.802	-1.495	4.214	-11.731
<i>PC2</i>	-5.579	-6.071	1.697	2.540	11.704	-4.291

图 5.4 进一步验证了 *IF* 和 *IF<sub>n</sub>* 很接近，*P* 和  $\beta_\omega$  较接近， $r_\omega$  和  $C_{R\omega}$  相差较大。

## 5.2 DP 引用网络

本节实验采用 Data for PageRank (DP) 网络，截取 2007 年对 2005 年和 2006 年的引用信息。该网络包括 397 种期刊，但是在 2005 年和 2006 年只有其中的 226 种期刊的文章在该数据集中有记录，因此仅对这 226 种期刊计算各项指标以及排名。

首先，处理文章引用数据得到期刊引用网络。类似小节 5.1，可以得到矩阵  $J_C$ 、 $J_{C_n}$  和  $J_{C_i}$ 。图 5.5 展示矩阵  $J_C$  各元素取自然对数后的情况。可以清楚地看到图 5.5 中的对角线，表示这些期刊的自引现象。

图 5.6 和图 5.7 分别是表示发表文章总数和被引量总数分布情况的直方图。可以看出，大部分期刊在 2005 年至 2006 年发表文章总数在 250 篇以下，大部分期刊的这些文章在 2007 年的总被引量在 250 次以下。然而，存在期刊发表文章总数或者被引量总数超过 2000（实际上，期刊 Applied and environmental

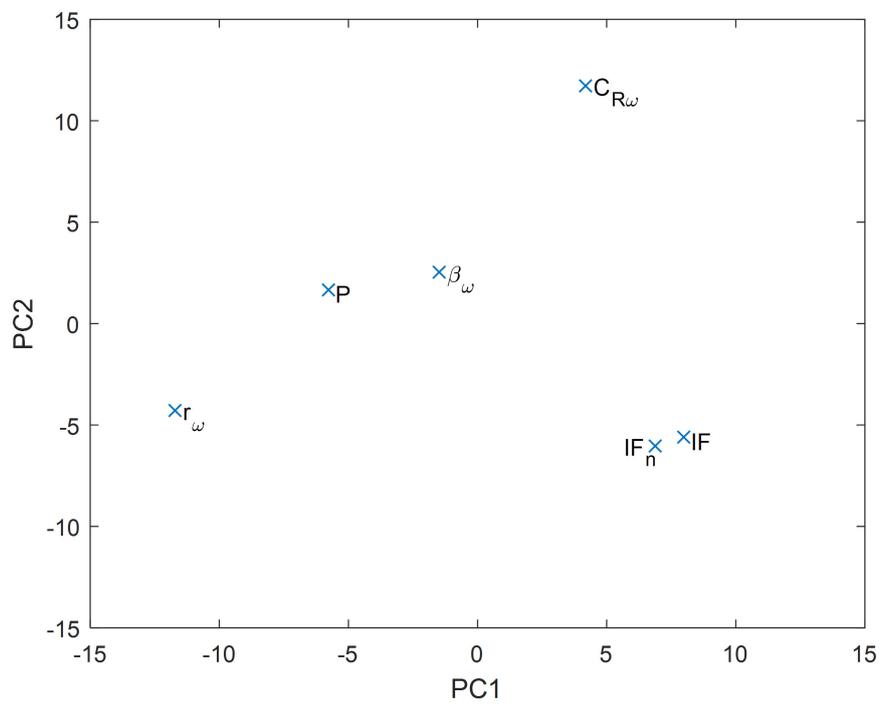


图 5.4 数学期刊引用网络各指标在主成分空间的投影

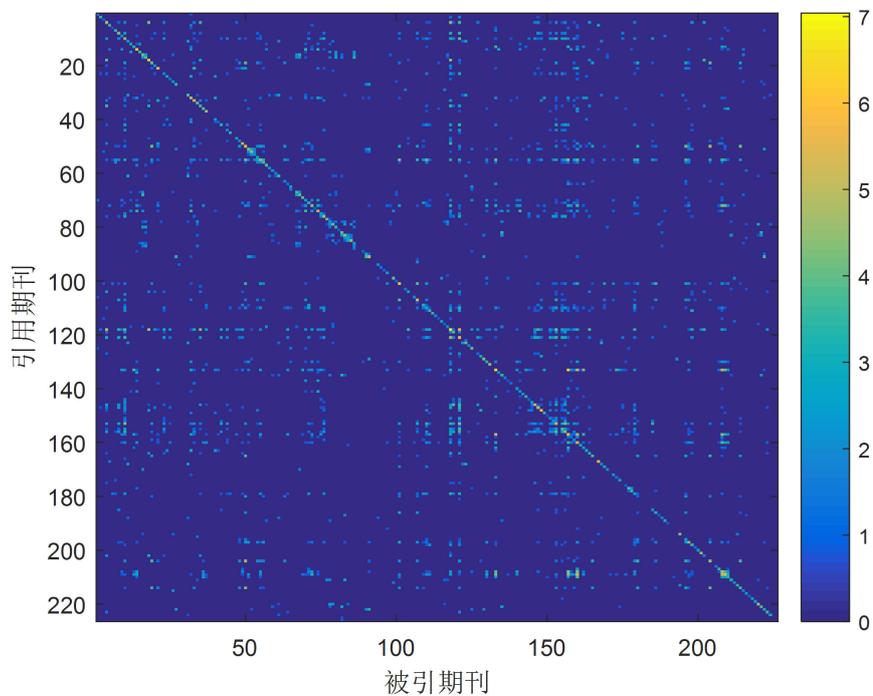


图 5.5 DP 期刊引用网络

microbiology 的发表文章总数和被引量总数同时大于 2000)，说明在发表文章总数和被引量总数两个方面，这些期刊具有明显的多样性。

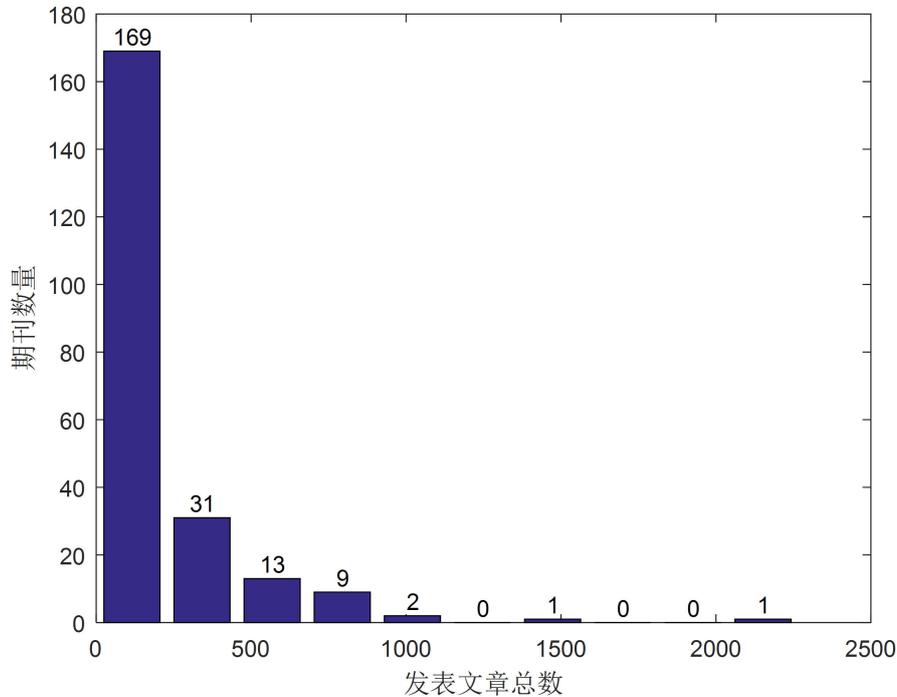


图 5.6 DP 网络期刊发表文章总数直方图

与小节 5.1 类似，对 DP 期刊引用网络计算表 5.1 中列出的各种指标。

DP 网络本身是文章引用网络，因此可以同时计算基于文章引用网络的其他指标，如表 5.6 所示，表中各个指标均利用  $PC$  计算，其中  $PC$  是 2005、2006 和 2007 年所有文章之间的引用矩阵，元素只取 0 或 1。实验中， $PC$  是 70730 阶方阵。

计算  $C_R^p$  时发现，从该网络中某一点出发，沿被引用关系可到达的点都可以在 7 步以内到达。该网络中共有 52605 条边，计算  $r^p$  时，采用算法 4 后剩下的图中有 596 条边，421 个点，且其中有 123 个点出度和入度不相等。采用求解问题 (4-2) 的方法计算，可得 agony 最小解等级数总和不超过 38562。令  $M = 39000$ ，求解问题 (4-3) 计算 agony 最小解，得出另一种排名，与  $r^p$  排名各分量差距总和是 205，平均每个期刊差 0.91，而  $C_R^p$  和  $r^p$  的排名平均每个期刊差 10.3。

挑出影响因子排名前 10 的期刊，将它们在各项指标下的排名列在表 5.7 和

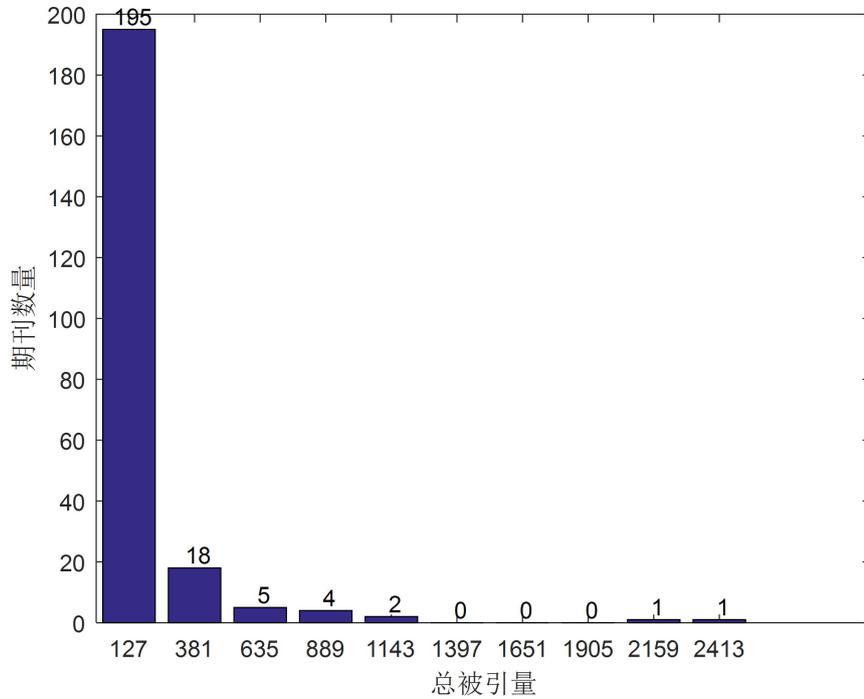


图 5.7 DP 网络期刊被引量总数直方图

表 5.6 基于文章引用网络计算的指标

指标类型	符号	计算方法
h-指数	$h$	用算法 1 计算 2007 年对 2005 和 2006 年的引用情况
PageRank 指标 (文章网)	$P^P$	用算法 2 后对该期刊所有文章得分求平均值
$\beta$ -measure (文章网)	$\beta^P$	用定义 2.1 后对该期刊所有文章得分求平均值
LRC 指标 (文章网)	$C_R^P$	用小节 3.2.5 的不加权 LRC 计算方法后对该期刊所有文章得分求平均值
Agony 指标 (文章网)	$r^P$	用小节 3.2.6 的文章引用网络 agony 计算方法后对该期刊所有文章得分求平均值

表 5.8 中进行比较。

表 5.7 DP 网络部分期刊各项指标排名 (一)

期刊	$IF$	$P$	$\beta_1$	$\beta_\omega$	$C_{R1}$	$C_{R\omega}$	$r_1$	$r_\omega$
GENOME RES	1	5	5	4	21	4	13	5
ADV BIOCHEM ENG BIOT 1	2	144	176	186	1	131	1	1
CHRONOBIOL INT	3	129	120	104	21	203	77	97
ADV BIOCHEM ENG BIOT 2	4	132	152	143	1	129	1	5
METAB ENG	5	34	72	83	21	56	13	5
MOL THER	6	6	15	5	21	5	13	33
STEM CELLS	7	13	23	11	21	40	77	33
ADV BIOCHEM ENG BIOT 3	8	172	184	195	1	139	1	5
BIOFOULING	9	165	198	189	21	22	170	97
BIOINFORMATICS	10	1	3	2	21	1	13	5

表 5.8 DP 网络部分期刊各项指标排名 (二)

期刊	$IF_n$	$h$	$P^P$	$\beta^P$	$C_R^P$	$r^P$
GENOME RES	2	3	35	3	2	4
ADV BIOCHEM ENG BIOT 1	1	122	97	2	3	3
CHRONOBIOL INT	181	19	22	8	24	11
ADV BIOCHEM ENG BIOT 2	3	122	29	1	1	1
METAB ENG	4	19	60	4	9	7
MOL THER	12	4	33	6	7	9
STEM CELLS	29	4	72	15	18	23
ADV BIOCHEM ENG BIOT 3	5	179	143	11	29	8
BIOFOULING	160	19	65	12	15	10
BIOINFORMATICS	17	2	56	9	11	15

可以从表 5.7 和表 5.8 中看出, 使用不同评价指标时, 得出的期刊排名相差很大。

在几个现有指标中,  $IF$  和  $IF_n$  比较接近, 例如  $IF$  的前 5 名均出现在  $IF_n$  的前 10 名中。但是在某些期刊上, 它们的表现完全不同, 例如 CHRONOBIOL-OGY INTERNATIONAL (CHRONOBIOL INT) 的  $IF$  排名是 3, 但  $IF_n$  排名是

181。这说明该期刊自引数量较多而其他期刊对它的引用数量较少。指标  $P$  和  $h$  的表现与两个影响因子指标相比有较大差别。例如，3 个 *Advances in Biochemical Engineering-Biotechnology* (ADV BIOCHEM ENG BIOT) 的  $P$  和  $h$  排名均在 100 名外。DP 数据集将 ADV BIOCHEM ENG BIOT 的不同栏目分别作为出版物编号，因此表中出现多次 ADV BIOCHEM ENG BIOT。每一个栏目文章总数都不多，由于基于 PageRank 的  $P$  和基于  $h$ -指数的  $h$  都可以反映期刊出版高质量文献的数量，因此发表文章总数较少的 ADV BIOCHEM ENG BIOT 栏目在这两个指标下的名次都远大于影响因子指标下的名次。

在基于期刊引用网络层级结构的指标  $\beta_1$ 、 $\beta_\omega$ 、 $C_{R1}$ 、 $C_{R\omega}$ 、 $r_1$  和  $r_\omega$  中， $\beta_1$  和  $\beta_\omega$  的结果较接近，而且与期刊引用网络 PageRank 指标  $P$  也很相似。这又一次验证了  $\beta$ -measure 和 PageRank 定义的相似性。观察  $C_{R1}$  发现，有大量期刊得到相同的排名。实际上由期刊引用网络导出的不加权网络  $JC_1$  中包含一个约 200 顶点的强连通子图，计算出的  $C_{R1}$  在全部顶点中只有 3 种不同取值，且绝大部分期刊集中于 1 种取值，它们有共同的排名 21。因此， $C_{R1}$  不适合用于衡量期刊质量。相比之下， $C_{R\omega}$  表现得比  $C_{R1}$  更有意义，而且与  $\beta_1$ 、 $\beta_\omega$  和  $P$  指标对比显示出类似的变化趋势。基于 agony 方法的期刊引用网络指标  $r_1$  和  $r_\omega$  均只包含 5 个层级，因此这两个指标存在大量相同排名的期刊。所以它们不适合比较质量相近的期刊，而只能给期刊划分大致的层次。相对而言， $r_\omega$  似乎比  $r_1$  表现好一些，这可以归结于  $r_\omega$  使用了更多的信息。

由于除指标  $h$  外，基于文章引用网络的指标  $P^p$ 、 $\beta_\omega^p$ 、 $C_R^p$ 、 $r^p$  均采用对期刊中所有文章求平均值的方法，因此它们反映期刊发表文章的平均质量，而不能反映高质量文章的数量。指标  $P^p$  得出的这 10 个期刊的排名均偏大。而  $\beta_\omega^p$ 、 $C_R^p$  和  $r^p$  的结果均与  $IF$  排名重合度很高，例如  $IF$  排名前 10 的期刊分别有 7、5 和 7 种在  $\beta_\omega^p$ 、 $C_R^p$  和  $r^p$  中排名前 10。

对上述共计 14 个指标，类似小节 5.1 计算斯皮尔曼等级相关系数。图 5.8 展示相关系数，图 5.9 展示在二者不相关时观察到这种结果的概率，两图中各指标均按照  $IF$ 、 $IF_n$ 、 $P$ 、 $\beta_1$ 、 $\beta_\omega$ 、 $C_{R1}$ 、 $C_{R\omega}$ 、 $r_1$ 、 $r_\omega$ 、 $h$ 、 $P^p$ 、 $\beta^p$ 、 $C_R^p$ 、 $r^p$  的顺序排列。

从图 5.8 和图 5.9 中可以看出，除了第 6 个指标  $C_{R1}$ ，其余任意两指标之间都有显著的线性相关。指标  $IF$ 、 $IF_n$ 、 $\beta^p$ 、 $C_R^p$  和  $r^p$  之间的斯皮尔曼等级相关系数均超过 0.7，说明它们相互之间线性相关程度很高。指标  $P$ 、 $\beta_1$ 、 $\beta_\omega$  和  $h$  之

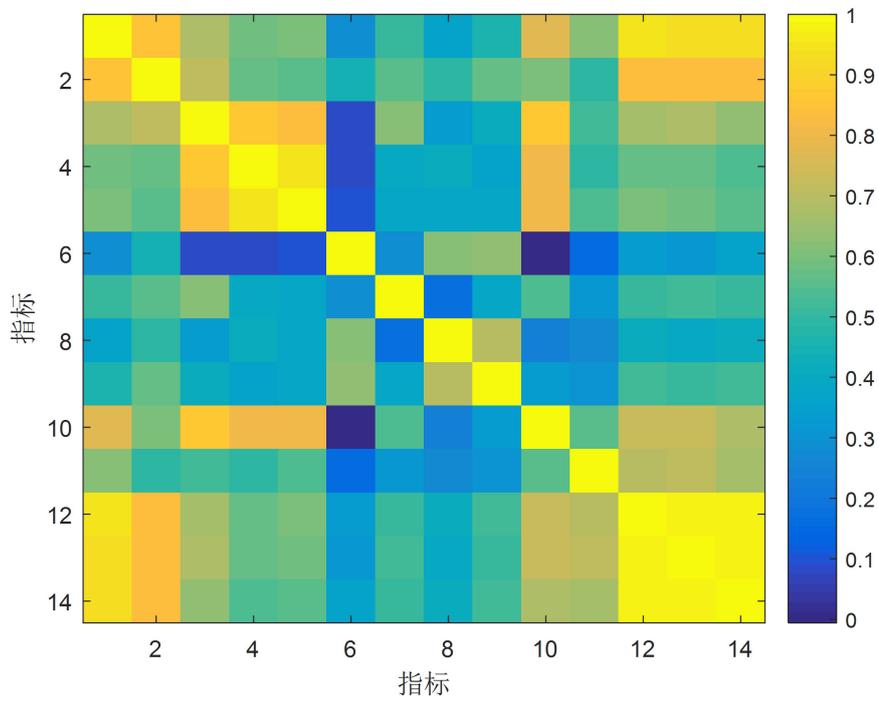


图 5.8 DP 网络各指标等级相关系数

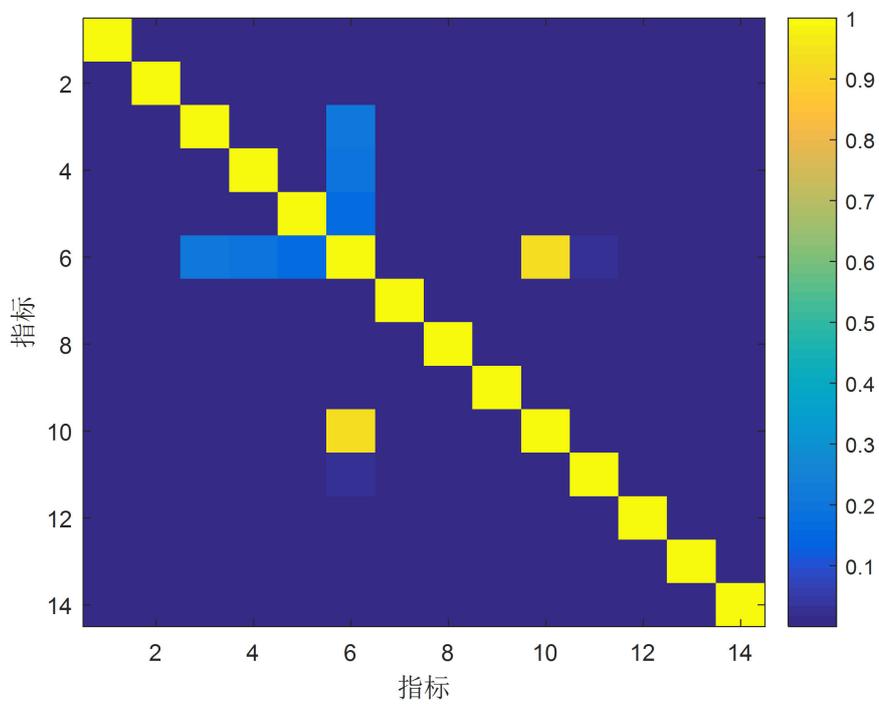


图 5.9 DP 网络相关系数对应概率

间也是这样。这些观察均与对  $IF$  排名前 10 期刊的分析相符合。

去掉指标  $C_{R1}$ ，对剩下 13 个指标的排序结果做主成分分析，各主成分对应方差比例如表 5.9 所示。

表 5.9 DP 网络各指标排序主成分对应方差比例

	$PC1$	$PC2$	$PC3$	$PC4$	$PC5$	$PC6$
对应方差比例	64.0%	10.6%	9.1%	6.2%	3.7%	2.1%
累积方差比例	64.0%	74.6%	83.7%	89.9%	93.6%	95.7%

投影所有的指标至前两个主成分  $PC1$  和  $PC2$ ，结果如表 5.10 和表 5.11 所示，并在图 5.10 中标出各指标。

表 5.10 DP 网络各指标投影至主成分空间（一）

投影	$IF$	$IF_n$	$P$	$\beta_1$	$\beta_\omega$	$C_{R\omega}$
$PC1$	0.357	-2.305	5.137	5.273	5.298	0.522
$PC2$	-5.104	-2.562	3.402	6.538	5.908	0.142

表 5.11 DP 网络各指标投影至主成分空间（二）

投影	$r_1$	$r_\omega$	$h$	$P^p$	$\beta^p$	$C_R^p$	$r^p$
$PC1$	-9.854	-9.415	6.143	1.285	-0.713	-0.457	-1.271
$PC2$	-6.634	4.077	1.047	-3.101	-5.563	-5.611	-5.806

由图 5.10 可以看出，这 13 个指标大致可以分为 3 类：第一类是指标 1、2、7、11、12、13、14，即  $IF$ 、 $IF_n$ 、 $C_{R\omega}$ 、 $P^p$ 、 $\beta^p$ 、 $C_R^p$ 、 $r^p$ ，反映期刊发表文章的平均质量；第二类是指标 3、4、5、10，即  $P$ 、 $\beta_1$ 、 $\beta_\omega$ 、 $h$ ，同时反映期刊发表文章的质量与高质量文章的数量；第三类是指标 8 和 9，即  $r_1$  和  $r_\omega$ 。以上三类指标中均含有基于层级结构特性的指标。

### 5.3 AAV8 引用网络

本节实验采用数据集 AMiner ACM-Citation-network V8。对于期刊引用网

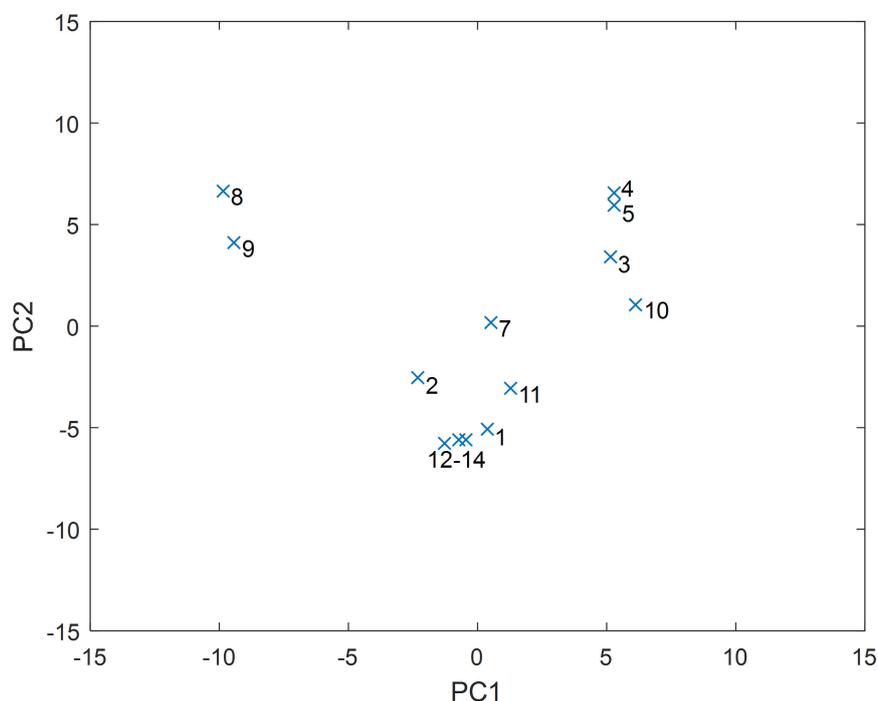


图 5.10 DP 网络各指标在主成分空间的投影

络，使用 2010 年对 2008 和 2009 年的引用信息。由于该数据集中出版物种类数量很多（273275），因此事先进行筛选，选取其中在 2008 和 2009 年发表文章总数不少于 50 篇的期刊，共有 1715 种期刊，42947 次引用。

处理文章引用数据得到相应的期刊引用网络和矩阵  $JC$ 、 $JC_n$  和  $JC_i$ 。图 5.11 和图 5.12 展示发表文章总数和被引量总数分布情况。

在 AAV8 数据集中，大部分期刊该时间段发表文章总数在 400 以下，被引量总数在 50 以下。说明该数据集可能有明显的遗漏现象。

对于文章引用网络，使用 2008 年至 2010 年三年间的文章引用信息，并限制在上述的 1715 种期刊内。

与小节 5.2 类似，对 AAV8 期刊引用网络计算表 5.1 中所列指标，对文章引用网络计算表 5.6 中所列指标，共计 14 个指标，并分别按照各个指标计算期刊排名。

计算  $C_R^p$  的过程表明，从该网络中某一点出发，沿着被引用关系可达点均可在 9 步内到达。该网络含 293187 篇文章和 106858 次引用信息，计算  $r^p$  时，将该网络输入算法 4，输出是含 328 个顶点和 2322 条边的子图，其中有 328 个点

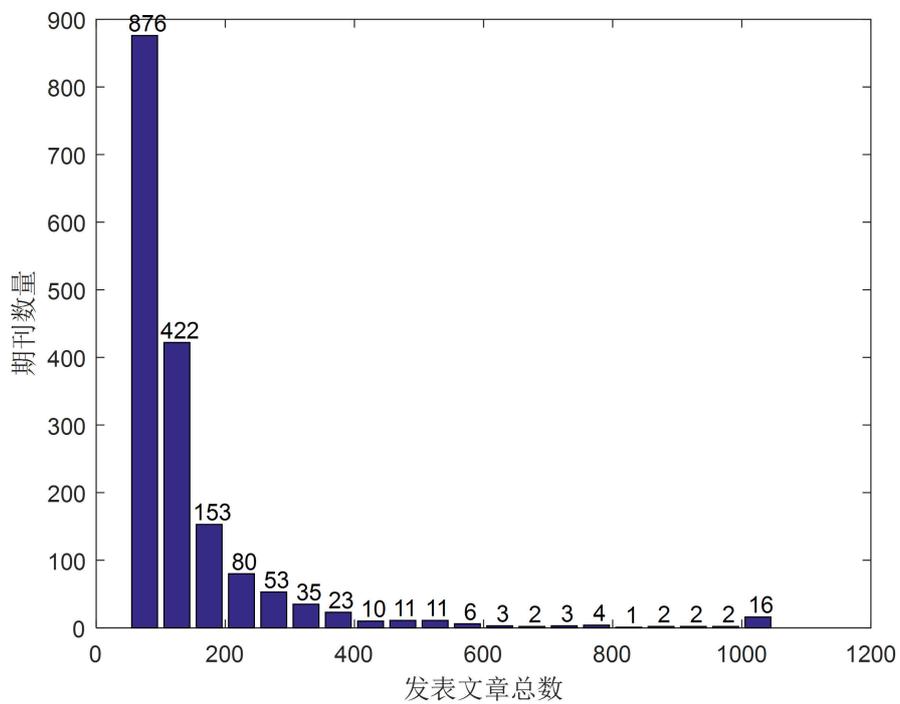


图 5.11 AAV8 网络期刊发表文章总数

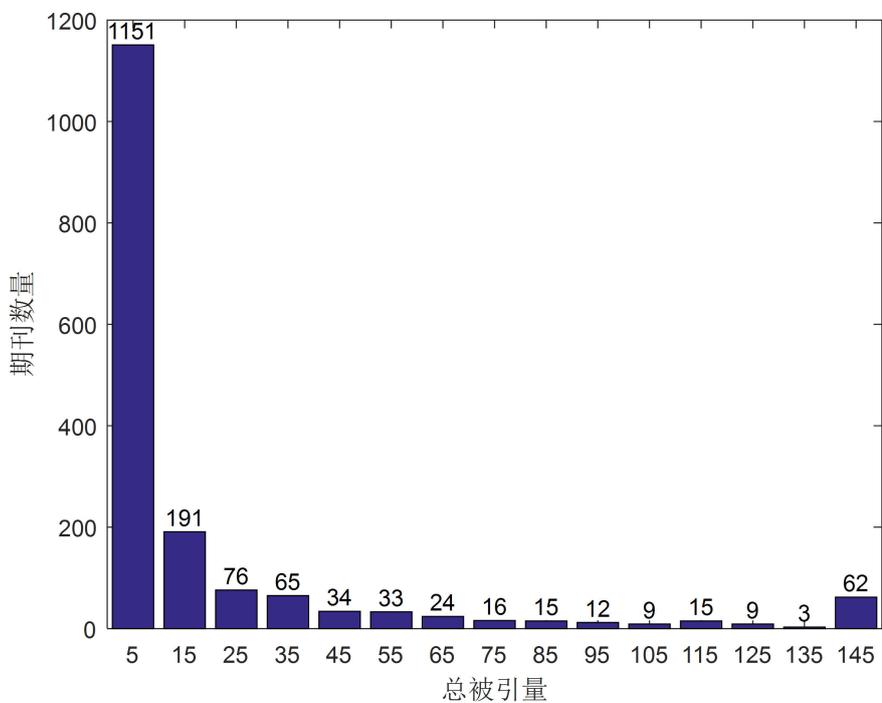


图 5.12 AAV8 网络期刊被引量总数

出度和入度不等。采用求解问题 (4-2) 的方法计算, 可得 agony 最小解等级数总和不超过 74443。令  $M = 75000$ , 求解问题 (4-3) 计算 agony 最小解, 得出另一种排名, 求出的两种期刊排名分量差距总和是 7146, 平均每个期刊差 4.17, 而  $C_R^p$  和  $r^p$  的排名平均每个期刊差 48.1。

与小节 5.2 类似, 找出  $IF$  指标排名前 10 的期刊, 观察它们在其他指标下的排名, 如表 5.12 和表 5.13 所示。

表 5.12 AAV8 网络部分期刊各项指标排名 (一)

期刊	$IF$	$P$	$\beta_1$	$\beta_\omega$	$C_{R1}$	$C_{R\omega}$	$r_1$	$r_\omega$
ACM Comput. Surv.	1	69	21	65	1094	910	147	38
Int. J. Mob. Commun.	2	574	247	213	1094	582	1343	1337
Inf. Sci.	3	16	3	6	1094	30	1125	1169
SIAM J. Imaging Sci.	4	64	201	151	1094	137	644	390
IEEE Trans. Signal Process. 2	5	131	301	307	31	15	147	38
IEEE J. Sel. Areas Commun.	6	99	161	94	1094	53	1125	390
IEEE Trans. Signal Process. 1	7	96	159	164	31	8	147	38
J. Manage. Inform. Syst.	8	24	74	31	1094	190	644	837
ACM SIGMOD	9	126	84	43	31	4	9	1
Int. J. Elec. Fin.	10	528	982	1045	1094	268	1343	1337

表 5.13 AAV8 网络部分期刊各项指标排名 (二)

期刊	$IF_n$	$h$	$P^p$	$\beta_\omega^p$	$C_R^p$	$r^p$
ACM Comput. Surv.	1	19	79	3	4	53
Int. J. Mob. Commun.	356	5	34	46	69	80
Inf. Sci.	69	1	345	31	24	31
SIAM J. Imaging Sci.	8	19	506	76	77	182
IEEE Trans. Signal Process. 2	2	19	299	15	29	18
IEEE J. Sel. Areas Commun.	4	11	284	16	26	61
IEEE Trans. Signal Process. 1	3	19	138	18	12	11
J. Manage. Inform. Syst.	17	19	314	60	22	84
ACM SIGMOD	5	19	22	2	3	2
Int. J. Elec. Fin.	656	46	400	39	34	37

一方面，指标  $IF$  排名前 10 的期刊在  $IF_n$  前 10 中出现 6 种，在指标  $P$ 、 $h$ 、 $\beta_\omega^p$ 、 $C_R^p$ 、 $r^p$  中普遍较靠前，说明这几种指标能反映期刊的质量。另一方面，这些期刊在指标  $r_1$ 、 $r_\omega$ 、 $P^p$  的排名中变化很大，说明这几种指标反映的期刊质量信息与  $IF$  有所差异。在这个例子中， $C_{R1}$  仍然几乎不能反映任何与期刊质量有关的信息。相比之下， $C_{R\omega}$  的表现稍好一些。指标  $P$ 、 $\beta_1$  和  $\beta_\omega$  的变化趋势很相似，与之前两个网络的结果相符。

类似小节 5.2，将上述 14 个指标的斯皮尔曼等级相关系数绘制在图 5.13 中，同时在图 5.14 中展示对应的概率。

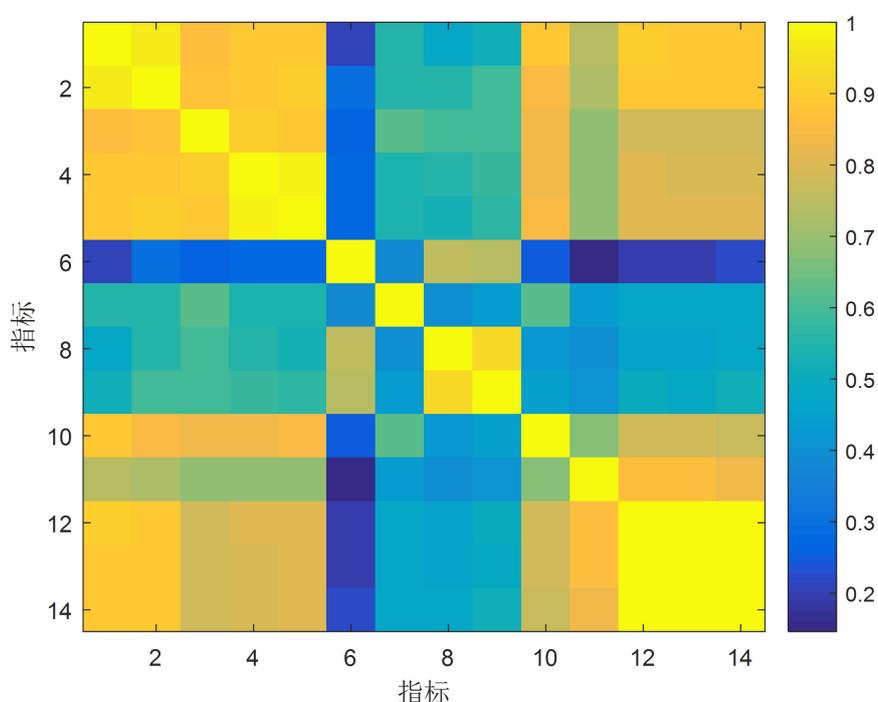


图 5.13 AAV8 网络各指标等级相关系数

从图 5.13 中可以看出， $IF$ 、 $IF_n$ 、 $P$ 、 $\beta_1$ 、 $\beta_\omega$ 、 $h$ 、 $P^p$ 、 $\beta^p$ 、 $C_R^p$ 、 $r^p$  这 10 个指标相互之间相关系数较高。指标  $C_{R1}$  与大多数指标的相关系数都较低。指标  $r_1$  和  $r_\omega$  两者比较接近。

类似小节 5.2，去掉  $C_{R1}$  并对其他 13 个指标的排序结果做主成分分析，并将各指标投影至前 2 个主成分，相关结果如表 5.14、表 5.15、表 5.16 和图 5.15 所示。

由图 5.15，指标 8、9（即  $r_1$  和  $r_\omega$ ）很接近且离其他指标都很远。指标 11-14

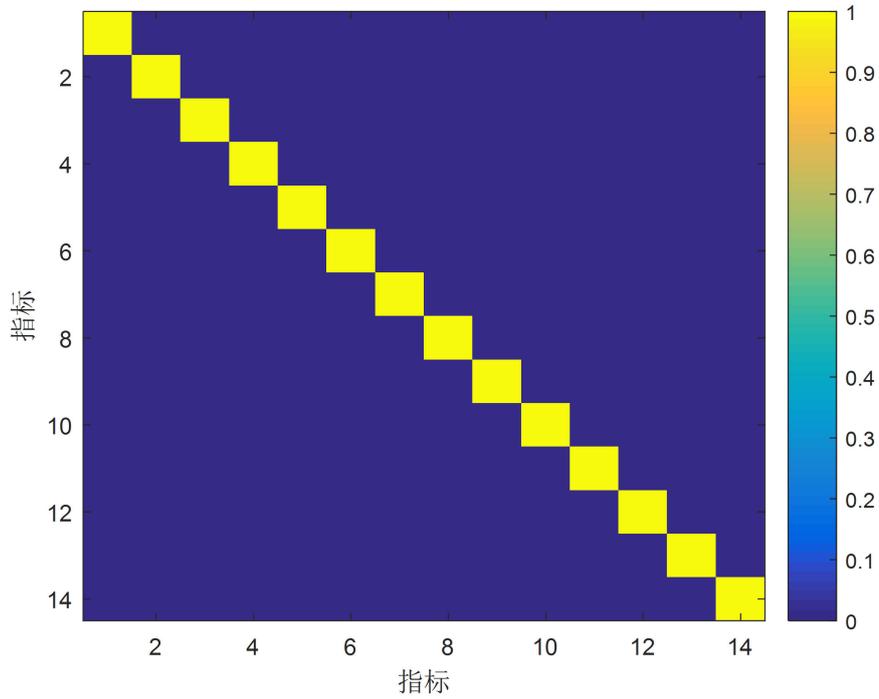


图 5.14 AAV8 网络相关系数对应概率

表 5.14 AAV8 网络各指标排序主成分对应方差比例

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>	<i>PC5</i>	<i>PC6</i>
对应方差比例	74.1%	10.1%	5.9%	4.0%	2.0%	1.3%
累积方差比例	74.1%	84.2%	90.1%	94.1%	96.1%	97.4%

表 5.15 AAV8 网络各指标投影至主成分空间（一）

投影	<i>IF</i>	<i>IF<sub>n</sub></i>	<i>P</i>	$\beta_1$	$\beta_\omega$	<i>C<sub>R</sub><math>\omega</math></i>
<i>PC1</i>	-9.018	-4.603	0.383	-2.653	-3.770	9.905
<i>PC2</i>	1.553	-0.346	6.301	3.497	3.562	26.391

表 5.16 AAV8 网络各指标投影至主成分空间（二）

投影	<i>r<sub>1</sub></i>	<i>r<sub><math>\omega</math></sub></i>	<i>h</i>	<i>P<sup>P</sup></i>	$\beta^P$	<i>C<sub>R</sub><sup>P</sup></i>	<i>r<sup>P</sup></i>
<i>PC1</i>	30.096	28.778	-6.630	-9.748	-11.230	-11.251	-10.259
<i>PC2</i>	-9.265	-8.320	10.414	-8.487	-8.322	-8.298	-8.680

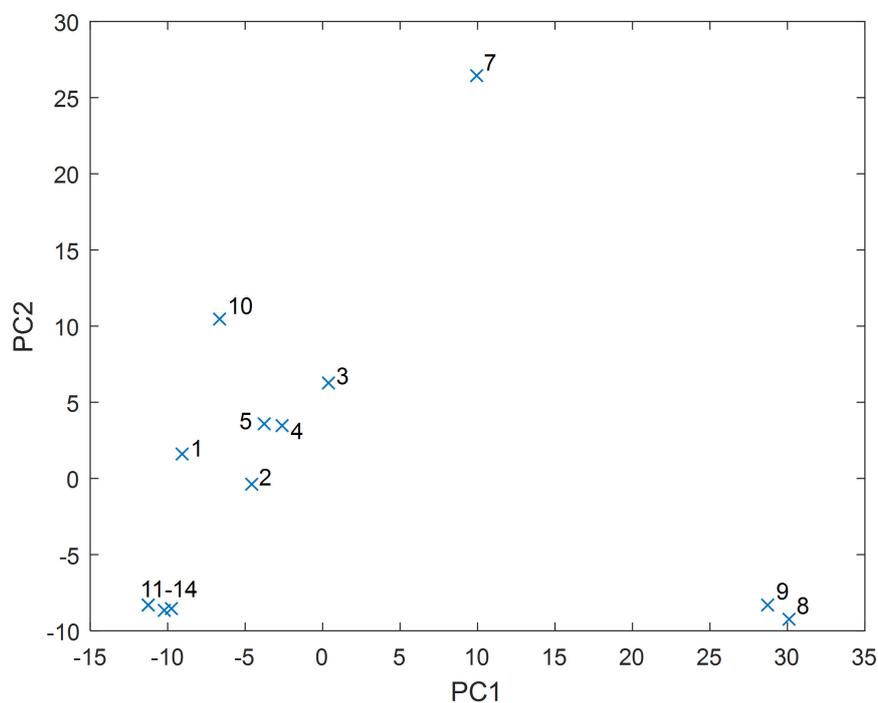


图 5.15 AAV8 网络各指标在主成分空间的投影

相互之间非常接近，它们都是基于文章引用网络的指标。指标 7（即  $C_{R\omega}$ ）比较孤立。

## 5.4 小结

本章分别在数学、生物学、计算机科学 3 个学科的引用网络中，利用现有指标和基于层级结构的指标进行期刊排序，并对不同指标的排序结果对比分析，得出以下几个结论：

第一，基于层级结构特性的指标能够反映期刊的质量，但是与现有指标影响因子、PageRank 指标、h-指数等均有差异。并且，本文采用的 3 种层级结构指标相互之间也有差异。

第二，对于期刊引用网络，利用加权的层级结构指标一般比不加权的表现更好，例如 LRC 指标中  $C_{R\omega}$  与  $C_{R1}$  相比在体现期刊质量方面好一些。

第三，基于文章引用网络的层级结构指标比基于期刊引用网络的更接近一些现有指标（即影响因子、PageRank 指标、h-指标）的结果。

## 第6章 结论

### 6.1 本文工作总结

本文将 3 种有向网络层级结构特性指标引入基于引用网络的期刊评价，它们分别是  $\beta$ -measure、LRC 指标和 agony 指标。本文分别针对期刊引用网络和文章引用网络的特点，研究了以上指标的计算方法。本文使用 3 个实际引用数据集进行实验，分别是数学期刊引用网络、生物学文章引用网络 DP 和计算机科学文章引用网络 AAV8。对于其中的文章引用网络，数据处理得到相应的期刊引用网络，并分别计算层级结构特性指标，得出各指标下期刊的评价。为了检验结果，利用相同的数据，计算几个常用的现有期刊评价指标及其变形，分别是影响因子、基于 PageRank 原理的指标和 h-指数。对比结果显示，层级结构特性指标可以应用于期刊引用网络或者文章引用网络作为期刊评价指标，它们与现有指标之间有一定联系，但反映的期刊特性不完全相同。相比之下，基于文章引用网络的层级结构特性指标评价结果与现有指标结果更接近。

本文工作的主要贡献有以下几点：第一，将一些有向网络层级结构特性指标引入期刊评价，为期刊评价提供了新的指标；第二，分别对期刊引用网络和文章引用网络，设计或选用了相关指标的计算方法；第三，在多个引用数据集中，将层级结构特性指标和现有指标评价结果对比分析。

### 6.2 存在的问题与改进方向

本文是有向网络层级结构特性指标应用于期刊评价的初步研究，没有研究的问题还有许多，可以作为之后的研究方向：第一，与现有指标相比，层级结构特性指标反映的期刊质量特性有何不同，以及如何从实际意义解释；第二，是否还有其他种类的有向网络层级结构特性指标可以用于期刊评价；第三，层级结构特性指标有多种变形，例如期刊引用网络或者文章引用网络、加权或者利用阈值不加权、时间窗口的选取等因素导致的变形，其中哪些是最适合期刊评价的形式；第四，层级结构特性指标在不同学科领域中表现是否有差异，如何将其应用于跨领域的引用网络，以及如何在评价某一领域期刊时考虑跨领域

的引用；第五，如何修改文章引用网络层级结构特性指标的定义，使其不仅考虑期刊发表文章的平均质量，而且计及高质量文章的数量；第六，层级结构特性指标和现有指标及调查方法相比，在期刊评价方面有哪些优缺点，何时适宜采用层级结构特性指标。

## 插图索引

图 4.1	例 4.1 图 $D_\omega$ .....	26
图 4.2	例 4.1 图 $\tilde{D}_\omega$ .....	26
图 4.3	例 4.2 .....	32
图 4.4	例 4.3 .....	32
图 4.5	例 4.4 .....	32
图 4.6	例 4.5 图 $D$ .....	42
图 4.7	例 4.5 图 $D_R$ .....	42
图 4.8	例 4.5 图 $D_n$ 的 agony 最小解 .....	42
图 4.9	例 4.5 图 $D$ 的 agony 最小解 .....	42
图 5.1	数学期刊引用网络（部分） .....	43
图 5.2	数学期刊引用网络发表文章总数 .....	44
图 5.3	数学期刊引用网络被引量总数 .....	45
图 5.4	数学期刊引用网络各指标在主成分空间的投影 .....	50
图 5.5	DP 期刊引用网络.....	50
图 5.6	DP 网络期刊发表文章总数直方图 .....	51
图 5.7	DP 网络期刊被引量总数直方图 .....	52
图 5.8	DP 网络各指标等级相关系数 .....	55
图 5.9	DP 网络相关系数对应概率 .....	55
图 5.10	DP 网络各指标在主成分空间的投影 .....	57
图 5.11	AAV8 网络期刊发表文章总数 .....	58
图 5.12	AAV8 网络期刊被引量总数 .....	58

图 5.13	AAV8 网络各指标等级相关系数 .....	60
图 5.14	AAV8 网络相关系数对应概率 .....	61
图 5.15	AAV8 网络各指标在主成分空间的投影 .....	62

## 表格索引

表 2.1	一些网络的 GRC 指标（摘自文献 [14]） .....	8
表 2.2	一些网络的 agony 计算结果（摘自文献 [16]） .....	11
表 5.1	基于期刊引用网络计算的指标 .....	46
表 5.2	数学期刊引用网络现有指标前 10 期刊 .....	46
表 5.3	数学期刊引用网络层级结构指标前 10 期刊 .....	47
表 5.4	数学引用期刊网络各指标排序主成分对应方差比例 .....	49
表 5.5	数学引用期刊网络各指标投影至主成分空间 .....	49
表 5.6	基于文章引用网络计算的指标 .....	52
表 5.7	DP 网络部分期刊各项指标排名（一） .....	53
表 5.8	DP 网络部分期刊各项指标排名（二） .....	53
表 5.9	DP 网络各指标排序主成分对应方差比例 .....	56
表 5.10	DP 网络各指标投影至主成分空间（一） .....	56
表 5.11	DP 网络各指标投影至主成分空间（二） .....	56
表 5.12	AAV8 网络部分期刊各项指标排名（一） .....	59
表 5.13	AAV8 网络部分期刊各项指标排名（二） .....	59
表 5.14	AAV8 网络各指标排序主成分对应方差比例 .....	61
表 5.15	AAV8 网络各指标投影至主成分空间（一） .....	61
表 5.16	AAV8 网络各指标投影至主成分空间（二） .....	61

## 参考文献

- [1] Dong P, Loh M, Mondry A. The "impact factor" revisited[J]. *Biomedical digital libraries*, 2005, 2(1):7.
- [2] Bollen J, Sompel H, Hagberg A, et al. A principal component analysis of 39 scientific impact measures[J]. *PloS one*, 2009, 4(6):e6022.
- [3] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439):509–512.
- [4] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks[J]. *Nature*, 1998, 393(6684):440–442.
- [5] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.
- [6] Mylonopoulos N A, Theoharakis V. On site: global perceptions of IS journals[J]. *Communications of the ACM*, 2001, 44(9):29–33.
- [7] Zucker K J, Cantor J M. The numbers game: The impact factor and all that jazz[J]. *Archives of Sexual Behavior*, 2003, 32(1):3–5.
- [8] Barnes S J. Assessing the value of IS journals[J]. *Communications of the ACM*, 2005, 48(1):110–112.
- [9] Braun T, Glänzel W, Schubert A. A Hirsch-type index for journals[J]. *Scientometrics*, 2006, 69(1):169–173.
- [10] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the Web[R]. Technical report, Stanford InfoLab, 1999.
- [11] Bollen J, Rodriguez M A, Sompel H. Journal status[J]. *Scientometrics*, 2006, 69(3):669–687.
- [12] Katerattanakul P, Han B, Hong S. Objective quality ranking of computing journals[J]. *Communications of the ACM*, 2003, 46(10):111–114.
- [13] Van Den Brink R, Gilles R P. Measuring domination in directed networks[J]. *Social networks*, 2000, 22(2):141–157.
- [14] Mones E, Vicsek L, Vicsek T. Hierarchy measure for complex networks[J]. *PloS one*, 2012, 7(3):e33799.
- [15] Gupte M, Shankar P, Li J, et al. Finding hierarchy in directed online social networks[C]. *Proceedings of the 20th international conference on World Wide Web*. ACM, 2011. 557–566.
- [16] Tatti N. Hierarchies in directed networks[C]. *2015 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2015. 991–996.
- [17] Tatti N. Tiers for peers: a practical algorithm for discovering hierarchy in weighted networks[J]. *Data Mining and Knowledge Discovery*, 2016. 1–37.

- [18] Creamer G, Rowe R, Herskop S, et al. Segmentation and automated social hierarchy detection through email network analysis[J]. *Lecture Notes in Computer Science*, 2009, 5439:40–58.
- [19] Macchia L, Bonchi F, Gullo F, et al. Mining summaries of propagations[C]. *Data Mining (ICDM)*, 2013 IEEE 13th International Conference on. IEEE, 2013. 498–507.
- [20] Brink R, Gilles R P. A social power index for hierarchically structured populations of economic agents[C]. *Imperfections and Behavior in Economic Organizations*. Berlin: Springer, 1994: 279–318.
- [21] Leskovec J, Krevl A. SNAP Datasets: Stanford large network dataset collection[EB/OL]. [2017-04-10]. <http://snap.stanford.edu/data>.
- [22] Tang J, Zhang J, Yao L, et al. Arnetminer: Extraction and mining of academic social networks[C]. *KDD'08*, 2008. 990–998.
- [23] 耿素云, 屈婉玲, 王捍贫. 离散数学教程[M]. 北京: 北京大学出版社, 2002.
- [24] Dijkstra E W. A note on two problems in connexion with graphs[J]. *Numerische mathematik*, 1959, 1(1):269–271.
- [25] Bang-Jensen J, Gutin G Z. *Digraphs: theory, algorithms and applications*[M]. Berlin: Springer Science & Business Media, 2008.
- [26] Schrijver A. *Combinatorial optimization: polyhedra and efficiency*[M]. Berlin: Springer Science & Business Media, 2002.
- [27] Papadimitriou C H, Steiglitz K. *Combinatorial optimization: algorithms and complexity*[M]. New Jersey: Courier Corporation, 1982.
- [28] Orlin J B. A faster strongly polynomial minimum cost flow algorithm[J]. *Operations research*, 1993, 41(2):338–350.