

## ◆ 师道之声

文章编号: 1672-5913(2019)02-0001-04

中图分类号: G642

# 有向图的增强 ——一个适合以问题求解为导向教学的例子

李晓明

(北京大学 计算机系, 北京 100871)

DOI:10.16512/j.cnki.jsjy.2019.02.002

过去几年,在北大有一门通选课,叫“社会科学中的计算思维方法”。每次上课的时候,我总会给出一个小小的有向图的例子,让同学们观察,看最少添加几条边,可以让那个图变成强连通的,这个看起来像是一个趣味数学游戏的活动常常会引起学生们的积极参与。由于例子很小,大家通常在两三分钟里就能给出正确的结果。然后,我就留下这个游戏推广到一般的问题让有兴趣的同学思考。

问题:给定一有向图 $G$ ,最少添加几条边,使之成为一个强连通图。

看起来这个问题对那些喜欢数学和逻辑思维的学生的确是有吸引力的:几乎每一次,都会有同学尝试给出一个完整的解。于是,我也做了一下文献搜索,发现这个问题的最早解是由Eswaran和Tarjan在1976年给出的<sup>[1]</sup>。1995年, Frank将问题推广到 $k$ -强连通,并给出了解<sup>[2]</sup>。有意思的是,2005年还有人发表论文,指出文献[1]中的解有点错误,并给予了纠正<sup>[3]</sup>。

读那些文献,第一感觉是,那是一个已经被深入研究并解决了的问题,从学术上讲不太会有什么进一步的新意,但几年的课堂教学,却让我有了另外一方面的认识。虽然那些文献中提出了多种解法,但往往都比较复杂,不适合教学;既然这问题那么容易引起学生的兴趣,是不是可以找到一个比较适合教学的方法(算法)呢?

于是,基于对那几篇文献的理解和近几年来断断续续和几个选课学生讨论的结果,特别是2016年的学生张浩千和2018年的学生姜百翼提出的想法,一个适合作为以问题求解为导向的教学方案浮现出来。在这里,我们不妨先介绍这个方案本身,然后再讨论它适合什么样的学生和课程。

## 1) 问题的导入。

实践表明,用某种“真实的数据”来引出这个问题是比较有效的。我的做法是在前面的一次课上给每个学生发一张纸,让每人除了写上自己的名字外,还写出班里其他3个同学的名字,于是,我就有了一个有向图的数据。经过课后的整理,利用一个画图软件,就得到如图1所示的一个有向图的图像。

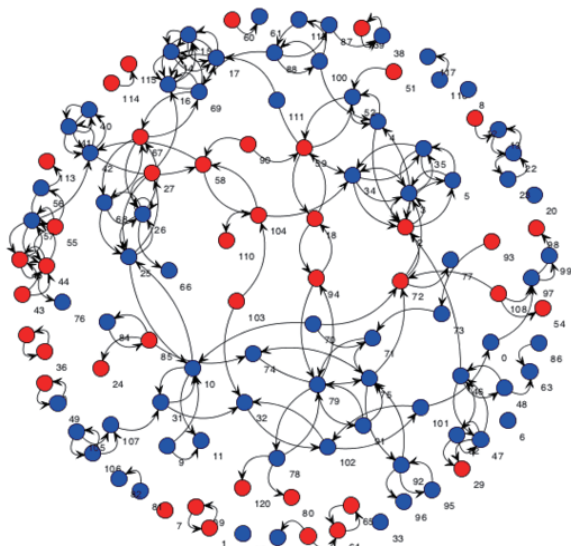


图1 班级社会网络图

在下次课上展示这个图像,让同学们意识到这是他们上次提供数据的结果,有益于建立起课堂学习的内容和他们相关的体验。举例来说,可以提问如下。

基于这样一个图所表达的关系,假设有向边意味着“可以直接传话”,如果要让每个同学都能(直接或间接地)传话给另一个同学(即两个节点之间存在有向路径),最少需要添加几条边?

如果班里的人数较多,如图中节点数超过

20, 就会让大家感到无从下手, 这时候, 可以单独给出一个小例子, 让他们尝试一下如何添加最少的边使其变成强连通, 这样就可以体会到这个问题的挑战性所在, 从而产生寻求一般解的愿望。

此时, 有向图、有向路径、强连通、强连通分量、弱连通 (即不考虑边的方向性, 是连通的) 等概念, 作为下面展开讨论的基本“语汇”, 就都可以自然带入了。所谓“以问题求解为导向的教学” (我觉得类似于教育学中近年来不少人推崇的“Project-based learning” (PBL)), 就是让抽象的概念随着对问题理解的展开而出现在学生面前, 而不是一上来首先是定义、定理、例子, 然后再讲应用。

## 2) 问题的定义 (抽象) 与初步分析。

现在, 我们可以抽象出本文一开始给出的问题了, 即:

问题: 给定一有向图  $G$ , 如何添加最少的边, 将它增广为一个强连通图。

经过问题导入阶段的小例子, 学生们应该已经意识到入度为 0 的节点和出度为 0 的节点是求解这个问题需要关注的重点。以此为基础, 几个可以逐步展开的分析点如下。

(1) 为避免在一开始陷入不必要的枝节, 不妨先假设有向图  $G$  是弱连通的。令  $p$  为入度为 0 节点个数,  $q$  为出度为 0 节点个数。此时, 大多数同学都能够很快意识到至少需要  $\max\{p, q\}$  条边才能将  $G$  增广为强连通图, 即  $\max\{p, q\}$  是必要的。达成这样的认识对建立信心有益处。

(2)  $\max\{p, q\}$  也是充分的吗? 这是第一个具有挑战性的问题。学生们肯定是分两派了, 持肯定意见的只能是大致的感觉, 持否定意见的也许能给出反例, 从而证明了他们的正确。例如图 2 (实线边) 就是一个反例, 其中每一个节点都在一个环中, 于是  $p=q=0$ , 但还是需要 1 条边 (虚线) 才能让它强连通。

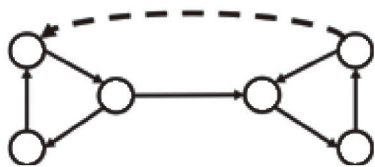


图2 关于一般情况下  $\max\{p, q\}$  充分性的一个反例

在什么样的情况下  $\max\{p, q\}$  不足以解决问题? 从上面这个小例子容易想到的一个一般情形, 即由非平凡连通分量 (即不是单个节点) 构成的有向图就可能产生这样的问题, 进而也容易

猜想将一个图 ( $G$ ) 的非平凡连通分量“收缩”成一个节点, 所导致的结果图应该和  $G$  在增广强连通上具有某种等价性。此时就可以引入有向无环图 (DAG) 的概念, 以及由  $G$  通过“收缩”其强连通分量为一个节点, 从而得到一个 DAG 的过程 (算法)。

必要的话, 可以让学生讨论将 DAG 增广为强连通与将  $G$  增广为强连通的等价性的证明。

(3) 到这里引出一个自然的问题: 若  $G$  是有向无环图 (DAG), 为了将它增广为强连通图, 用  $\max\{p, q\}$  条边是充分的吗?

这也就是经过前面的导入和提炼后形成的核心问题。接下来就是在课堂上可以一步步展开的讨论。

## 3) 问题的求解。

努力将看起来错综复杂的情形转变为简单清晰的情形, 不断将问题简化, 是值得我们在教学中贯彻的一个基本方法。这个将有向图增广为强连通的问题是体现这种方法的一个好例子。前面将一个一般的有向图  $G$  转变为 DAG 可以看成是其中一个步骤, 接下来的步骤如下。

(1) 我们现在考虑用  $\max\{p, q\}$  条边能否解决问题。容易想到的是那些添加的边应该首先用来减少出度为 0 和入度为 0 节点的个数, 即要跨在它们之间。如果  $p \neq q$ , 不妨假设  $p > q$  (反过来情形类似处理), 就要想多出的那些入度为 0 的节点怎么办? 此时, 若用  $p-q$  条边在入度为 0 的节点之间做连接, 只要不形成环, 我们就得到一个入度为 0 与出度为 0 个数相等 (记为  $n$ ) 的图, 记为  $D$ 。学生们不难认识到, 若能用  $n$  条边解决  $D$  的问题, 也就是用  $\max\{p, q\}$  条边解决了 DAG 的问题。于是我们实现了对问题的又一次简化, 即只要考虑  $p=q=n$  的情形即可。

那么, 我们讨论的就是一个相当“规范化”的有向无环图  $D$ , 如图 3 所示。

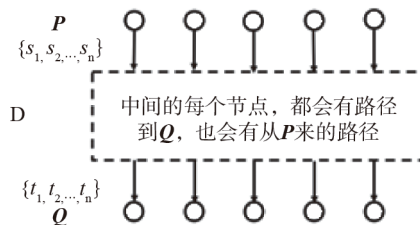


图3  $G \rightarrow \text{DAG} \rightarrow D$ , 一个“规范化”表示的有向图

其中, 我们将入度为 0 和出度为 0 的节点集合突出出来, 用  $P$  和  $Q$  分别指代。图中虚线框

中含有所有其他节点，其中文字所表达的断言也是适合让学生在课上证明的（其中要用到  $D$  是有向无环弱连通的性质）。

(2) 现在要追求的，就是看能否在  $D$  的基础上添加  $n$  条边，让它成为强连通。通过前面的讨论，大家基本能认识到，如果可能的话，那些边一定都是从  $Q$  到  $P$ 。那么，是不是随便怎么连，只要是一个无冗余、无遗漏的 1—1 映射就可以了呢？不是！这里有一个例子。

图 4 中， $n=2$ 。如果添加的两条边是  $2 \rightarrow 1$  和  $4 \rightarrow 3$ ，虽然不再有入度为 0 或出度为 0 的节点，但结果图显然还不是强连通的。不过，若添加的边是  $2 \rightarrow 3$  和  $4 \rightarrow 1$ ，就可解决问题。

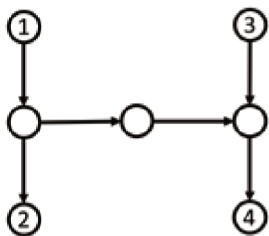
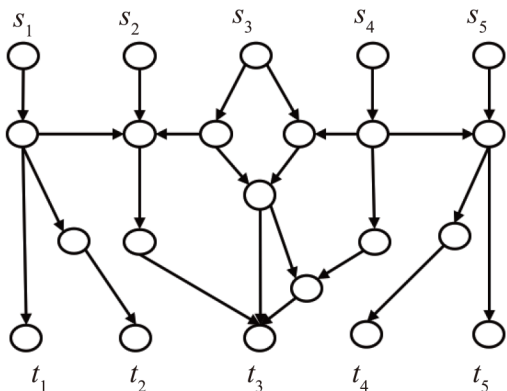


图 4 随意添加边不一定能解决问题的一个例子

通过这个例子，学生们对在  $P$  和  $Q$  之间添加边所面对的状况会有切实的体验，直觉上能感到应该争取和避免什么。此时，聪明的学生也许就能总结出具体的方法（课堂上不容易，但在课下进一步深入思考后有可能）。不过，从教学的角度，我想把“不断化简”的途径再往前开拓一步。

(3) 这里的关键是要启发出一个很有用的认识：若添加一些边能让  $D$  中包含  $P$  和  $Q$  的某个子图强连通，则那些边也就将  $D$  增广为强连通。以图 3 为例换句话说，如果添加一些从  $Q$  到  $P$  的边，加上虚线框中的某些从  $P$  到  $Q$  的路径，能保证  $P$  和  $Q$  中的节点两两之间都有双向路径，



则  $D$  中所有节点两两之间也就都存在双向路径，即成为强连通。

这样的认识一旦提出，直觉上是容易被接受的。如果教学目标中希望强调理论性，这里可以插入一个关于证明的讨论；而在强调较快的课堂节奏，让学生很快把握问题求解过程全貌的教学设计下，也可以直接往下进行，而将证明留在后面。

这个认识的意义在于，可以基于图 3 中的  $D$ ，得到一个两部分相等的二部图  $B_n$ ，其节点集合就是  $P=\{s_1, s_2, \dots, s_n\}$  和  $Q=\{t_1, t_2, \dots, t_n\}$ ，存在边  $s_i \rightarrow t_j$ ，当且仅当在  $D$  中有一条从  $s_i$  到  $t_j$  的路径。图 5 是一个例子。

(4) 于是，问题就变成如何添加  $n$  条边，让  $B_n$  成为强连通。这里，依然按照“化简”的思路，即考虑是否可以添加 1 条边，将  $B_n$  变成  $B_{n-1}$ ，一个保持关键的连通性质但出度为 0 和入度为 0 节点数分别减 1 的二部图。若这个操作能做到，就可以有  $B_n \rightarrow B_{n-1} \rightarrow \dots \rightarrow B_k$ ，直到  $B_k$  是一个“完全二部图”，即每一个  $s$  都有到每一个  $t$  的边。那样的图上有丰富的边，于是有许多种添  $k$  条边使之成为强连通的方式，其中一种就是对所有  $i$ ，添上  $t_i \rightarrow s_i$ 。在如此得到的结果图中，可以看到一个将所有节点串起来的环（也就是强连通）：

$$s_1 \rightarrow t_2 \rightarrow s_2 \rightarrow t_3 \rightarrow \dots \rightarrow t_k \rightarrow s_k \rightarrow t_1 \rightarrow s_1$$

(5) 如何添加 1 条边，将  $B_n$  变成  $B_{n-1}$ ？记  $O(s_i) \subseteq Q$  和  $I(t_j) \subseteq P$  分别为节点  $s_i$  和  $t_j$  的邻居集合。如果  $B_n$  不是完全二部图，就存在  $s_i$ ， $O(s_i) \subset Q$ ，也就是存在  $t_j \notin O(s_i)$ 。取一个这样的  $s_i$  和  $O(s_i)$  之外的  $t_j$ ，用  $t_j \rightarrow s_i$  作为添加边。

这样，图中入度为 0 和出度为 0 的节点都少一个，正是我们希望的，但它已经不是二部图。 $B_{n-1}$  如何形成？它除了没有  $s_i$  和  $t_j$ ，还应该保持在添加了边  $t_j \rightarrow s_i$  之后， $P$  和  $Q$  中剩余节点之间

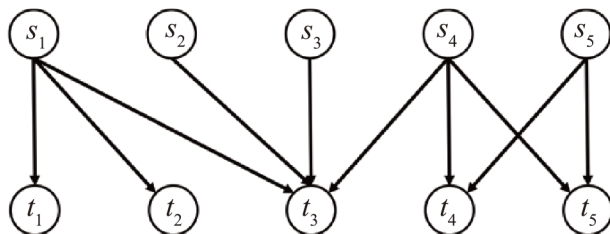


图 5  $D \rightarrow B_n$  的例子



的路径关系,也就是:

$B_{n-1} = B_n$  节点集  $\{s_i, t_j\}$  + 边集  $\{I(t_j) \rightarrow O(s_i)\}$

至此,遵循不断去繁化简的思路,就完成了一条路径的展示:

$G \rightarrow \text{DAG} \rightarrow D \rightarrow B_n \rightarrow B_{n-1} \rightarrow \dots \rightarrow B_k$

这条路径总会达到一个完全二部图  $B_k$  (极端情况就是两部各只有一个节点),于是就得到一个用最少的边,将有向图增广为强连通图的实施过程框架,可以看成是一个“高层的”算法。若  $\text{DAG}$  中有  $\max\{p, q\} = p \geq q = n$ , 可见这个过程用到的边数为  $(p-q) + (q-k) + k = p = \max\{p, q\}$ , 正是所希望的。当然,其中每一个环节的落实都会涉及某些具体算法,如在从  $D$  到  $B_n$  这个环节会用到广度优先搜索、从  $G$  到  $\text{DAG}$  的环节需要检测强连通分量等。因此,将这个问题求解过程完整实现,也可以成为一个不错的数据结构和程序设计综合练习,其输入为一个有向图  $G$  的边集合或邻接矩阵,输出为需要添加的边集合。

至此,还有什么遗留问题吗?答案是有的!那就是在前面的(1)中,有  $G$  是弱连通的假设。如果  $G$  本身包含几个没有任何关联的部分(弱连通分量),我们希望将它增广为强连通图,应该怎么做?

假设  $G$  有  $m$  个弱连通分量,对应入度为 0 和出度为 0 的节点个数分别为  $\{p_1, q_1\}, \{p_2, q_2\}, \dots, \{p_m, q_m\}$ 。一个迅速的想法会是:每个分量按照上述算法处理,然后用  $m$  条边将它们连成一个大“环”,于是总共要用  $m + \sum_{i=1}^m \max\{p_i, q_i\}$  条边,但这是不正确的!图 6 就是一个反例。

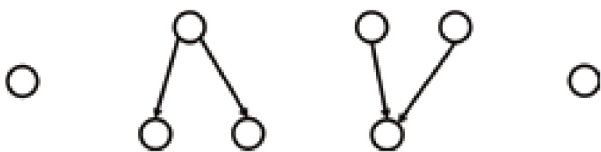


图 6  $G$  不是弱连通的,它由 4 个弱连通分量构成

按照刚才那个想法,将它增广为强连通图,需要  $2+2+4=8$  条边,但在图 7 所示的方案中,5 条边(虚线)就够了。

受这个例子的启发,新的(正确的)结论就是:假设  $G$  有  $m$  个弱连通分量,分别有入度为 0

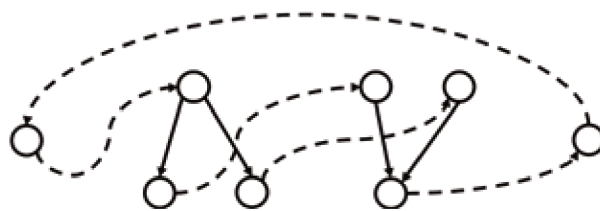


图 7 针对非弱连通图的一种优化实施方案

和出度为 0 节点的个数  $\{p_1, q_1\}, \{p_2, q_2\}, \dots, \{p_m, q_m\}$ , 其中,若某个弱连通分量为一个孤立点,则合理地认为它既是 1 个出度为 0 的节点,又是 1 个入度为 0 的节点(相当于是由两个节点和一条边构成的子图)。这样,记  $p = \sum_{i=1}^m p_i, q = \sum_{i=1}^m q_i$ , 将  $G$  增广为强连通图需要的最少边数就是  $\max\{p, q\}$ , 既必要也充分。具体实施层面,则是在  $G$  有孤立点的情况下,先将孤立点变成一个 2 节点 1 条边的子图,后面的过程则和前述完全相同,最后再将这些由孤立点衍生的两节点认定为同一个节点即可。

至此,我们完成了一个以问题求解为导向的教学方案的描述,其中涉及的知识大致分布在离散数学和算法设计课程中。如果要实现的话,还涉及程序设计和数据结构课程,那么安排在什么课程里更合适呢?在目前多数学校计算机专业的教学计划中,一二年级的课程基本比较传统,为了兼顾现状,可以将其考虑作为数据结构与算法课程的一个综合作业。这样的作业,不是老师简单地给学生布置一个任务,而是老师引导学生历经上述思考与推理的过程(大致需要 2~4 学时),然后由学生做程序实现。这样,学生体验的就不是急匆匆地完成作业,而是解决问题的思想方法,以及相关知识的综合与深入应用过程。其中,可以梳理出几条值得作为证明练习的断言,从而让学生不仅能在直觉上认同“ $G \rightarrow \text{DAG} \rightarrow D \rightarrow B_n \rightarrow B_{n-1} \rightarrow \dots \rightarrow B_k$ ”这样一个宏观的框架,还能得到对其正确性的严谨理解。

另外一种考虑,则是开发出 8~10 个类似分量的问题,形成一门以问题求解为导向的新课,不仅可以针对计算机专业开设,还可以供其他专业的学生选修。

参考文献:

- [1] Eswaran K P, Tarjan R E. Augmentation problems[J]. SIAM Journal on Computing, 1976, 5(5): 653-665.
- [2] Frank A, Jordan T. Minimal edge-covering of pairs of sets[J]. Journal of Combinatorial Theory, 1995, 65(1): 73-110.
- [3] Raghavan S. A note on Eswaran and Tarjan's algorithm for the strong connectivity augmentation problem[C]//Golden B, Raghavan S, Wasil E. The Next Wave in Computing, Optimization, and Decision Technologies. Berlin: Springer, 2005: 19-26.

(编辑:宋文婷)