无向图与有向图的连通性

清华算协 万成章

2024.8

- 对于一张图 $G = (V, E)$，V 是点集，E 是边集。默认 $n = |V|$，$m = |E|$。
- 无向图：边是无向的，即 E 的元素是 V 中元素形成的无序二元组。
- 有向图：边是有向的，即 E 的元素是 V 中元素形成的有序二元组。
- 若 $G' = (V', E')$ 是一张图且 $V' \subseteq V$，$E' \subseteq E$，则称之为 G 的子图。
- 给定一个点集 V 的子集 V'，令 E' 是 E 中所有两个端点都在 V' 里面的边的集合，则 (V', E') 称为集合 V' 的点导出子图，简称导出子图。
- 如果一张图没有重边和自环，称它是简单的。

- 我们不区分路径、迹、道路这些概念，统一用它们表示一个点边序列 $(v_1, e_1, v_2, \dots, v_{t-1}, e_{t-1}, v_t)$，满足 e_i 是从 v_i 到 v_{i+1} 的边。
- 但是，简单路径指的是不经过重复点的路径，即 $v_1, \dots, v_t$ 两两不同的路径。
- 我们不区分回路、环等概念，统一用它们表示 $v_1 = v_t$ 的路径。
- 简单环指的是不经过重复点的环，即 $v_1, \dots, v_{t-1}$ 两两不同的环。
- 如果存在 x 到 y 的路径，就称它们连通。如果任意一对点都连通，就称图是连通图。

深度优先搜索（DFS）

- 选定一个初始点，每当到达一个点时，遍历它所有没有到达过的邻居，递归访问。
- 令 $dfn(x)$ 表示 x 在这个过程中是第几个被访问到的点。那么 $dfn(1 \dots n)$ 是一个 $1 \dots n$ 的排列。
- 所有在 DFS 过程中走过的边形成了一棵以初始点为根的树（因为每个点都有唯一的 parent，就是访问到它的那个点），称为 DFS 树，它有一些不错的性质。

DFS 树

- 一棵子树中所有点的 dfn 构成了一段区间（因为它们是一起访问的，中间不可能插入访问过其他点）。
- 对于无向图，所有不在 DFS 树上的边一定都是祖先-后代的，而不是横叉的（因为假如有横叉的，就表示可以从一边走到另一边，那么在 DFS 过程中就会走过去，这样它就会出现在 DFS 树上了）。

如何判断一张图是否是连通的？

- 从任何一个点开始 DFS，如果成功访问到了所有点，说明图是连通的，否则图就是不连通的。
- 题外话：判连通 $\in \mathrm{L}^2$（可以只用 $O(\log^2 n)$ 的空间，显然 DFS 是做不到的）。
- 题外话：随机游走。

更多的关于“图有多连通”的指标？

- 组合方面：k-连通性。
- 代数方面：代数连通度。

流与割（复习一下离散数学）

- 无向图上，从 x 到 y 的最大流定义为：最多能找出几条从 x 到 y 的两两边不交的路径。
- 从 x 到 y 的最小割定义为：最少要删掉几条边，使得 x, y 不连通。
- 著名的最大流-最小割定理：x, y 之间，最大流等于最小割。
- 如果 x 和 y 之间的最大流 $\geq k$，就称它们是边 k-连通的，简称 k-连通（与之相对的还有点 k-连通）。k 越大，表示连通性越强（因为要删掉更多的边才能使得它们不连通）。

k-连通性

- k-连通性是一个等价关系：x, y 和 y, z 都 k-连通，推出 x, z 也 k-连通。
  证明：x, y 和 y, z 都 k-连通说明任意删掉 $k-1$ 条边后 x, y 和 y, z 还是连通，那么 x, z 自然也连通。
- 更细致的 k-连通结构：见 Gomory-Hu Tree。
  简单来说：存在一个带权树，使得任意两点的最大流等于它们树上路径边权最小值。

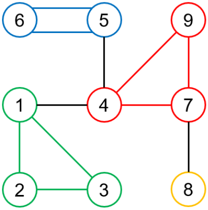
（边）双连通

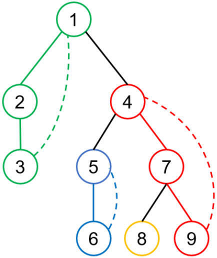
- 根据上面的定义可以知道：x, y 2-连通（或者叫双连通），当且仅当任意删掉一条边之后 x, y 还连通。
- 整张图双连通，当且仅当任意删掉一条边之后这张图还连通。
- 为此，我们定义割边（也可以叫桥）为删掉后使得图不连通的边，那么没有割边的图就是双连通的。

怎么判断图是不是双连通的？

- 尝试找到所有割边，如果有割边就不双连通。
- 事实上，找到所有割边后，我们就可以把这张图分成若干个双连通的（导出）子图，这些子图称作原图的**双连通分量**。

（边）双连通分量





左侧为原图，右侧为 DFS 树。

如何找到所有割边?

- 观察 DFS 树：首先，只有树边才有可能是割边。
- 令 x 是任意一个点，y 是 x 的父结点，如果除了 (x, y) 这条边外没有其他连接 x 子树内外的边了，那么 (x, y) 就是割边。
- 我们可以计算出 $low(x)$ 表示 x 子树内向子树外连的所有边中，端点 dfn 的最小值。那么 (x, y) 不是割边当且仅当 $low(x) < dfn(x)$（这表示存在一条从 x 子树内到子树外的边）。

这个算法称为 Tarjan 算法，时空复杂度都是 $O(n + m)$。

Remarks

- 边双连通图是连通性比较强的图，有一些更好的性质，例如耳分解（ear decomposition）。
- 一张图在忽略边双连通分量内部结构的前提下，可以看成一棵树，每个结点是一个边双连通分量，每条边是一个割边。
- 边双连通分量总是 DFS 树上的连通块，但是这一点对于一般的 k-连通分量未必成立。

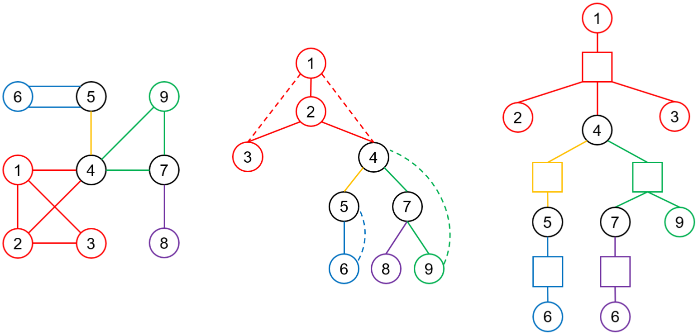
点连通性

- 仿照边连通性的定义，可以定义点连通性：
- 若 x, y 之间存在 k 条除端点外没有公共点的路径，则称 x, y 点 k-连通。
- 等价定义：若任意删掉 $k-1$ 个不同于 x, y 的点后，x, y 都一定连通，则 x, y 点 k-连通。
- 不太好的一点：点 k-连通不一定是等价关系。（考虑一条三个点的链，两两之间点连通性如何？）

好在，点双连通仍然有着比较简单的结构。

- 一张图同样可以划分为若干个子图，使得每个子图都是点双连通的，称为点双连通分量。
- 严格定义：极大的（也就是任意加入一个点都不行）点双连通导出子图。
- 不过，与边双连通划分时要求每个**点**都恰好在一个边双连通分量中不同，这里我们要求每条**边**都在恰好一个点双连通分量中。

点双连通分量



左侧为原图，中间为 DFS 树，右边是一个称为“圆方树”的抽象结构。

如何求点双连通分量？

- 和求边双连通分量类似，我们发现点双连通分量仍然是 DFS 树上的连通块。
- 设 y 是 x 的父结点，如果 $low(x) \geq dfn(y)$，这就表示 x 子树内怎么走都走不出 y 子树，那么 y 连同 x 子树内的点就构成了一个点双连通分量。在处理完这一切之后，就把 x 子树整个删掉，不参与后续算法。

这就是求点双连通分量的 Tarjan 算法，时空复杂度都是 $O(n+m)$。

Remarks

- 与割边一样，我们可以定义割点：删掉后会导致图不连通的点。
- 每个点双连通分量都有一些与其他点双连通分量连接的“端点”，这些就是割点，将点双连通分量也抽象成一个点（画成方点）与其中的割点（画成圆点）连接，就形成了上面的圆方树。
- 点双连通有与边双连通类似的耳分解。

在双连通之上，三连通也有一些独有的有趣性质，不过这里就不介绍了。

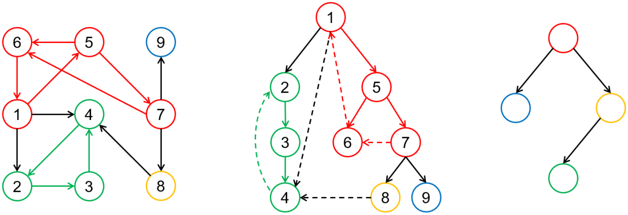
上面所说的内容都是针对无向图的，对于有向图，连通不再是一个无序的关系。

- 如果存在一条 x 到 y 的路径，则称 x 可达 y，通常这并不意味着 y 也可达 x。
- 如果将有向边看成无向边之后得到的无向图是连通的，则称原来的有向图是弱连通的。
- 而如果对于任意一对 x, y，都满足 x 可达 y，就称图是**强连通**的。这是我们主要感兴趣的部分。

强连通分量

- 一张有向图可以按点集划分成若干个导出子图，每个子图都是强连通的（同时还要是极大的，也就是加上任何其他的一些点都会变得不强连通），称为强连通分量。
- 类似于双连通分量的分解那样，在分解完成后，一张有向图就可以看成一张 DAG（有向无环图），其中每个点代表原图的一个强连通分量。

强连通分量



左侧为原图，中间为 DFS 树，右侧为抽象出来的 DAG。

求强连通分量

- 强连通分量同样是 DFS 树上的连通块。
- 对于以 x 为根的子树，如果 $low(x) \geq dfn(x)$，则表示不存在从其中到外部的边，因此这个子树内外不可能是同一个强连通分量。
- 找到所有这样的 x，将它们的子树和外部断开，剩下的每一块就是一个强连通分量了。
- 但是有向图的 DFS 树稍微复杂一些，可能包含一些不是祖先-后代边的边，这时上面这个做法的正确性不太显然。

一点算法细节的补充

- 我们递归地计算所有强连通分量，维护一个栈 DFS 目前访问到的点中，还没有被归入任何一个强连通分量的点集。
- 如果有一条非树边连到栈中的点，那么我们就要用它更新 low，否则就不更新（因为连到一个已经算完的强连通分量没有什么意义）。
- 一旦碰到 $low(x) = dfn(x)$ 的情况，就把 x 子树内剩下的所有点归为一个强连通分量。
- 在上面的过程中，我们找到强连通分量的顺序是从后到前的，也就是如果强连通分量 A 可以到强连通分量 B，那么 B 的编号会更小。

这就是求强连通分量的 Tarjan 算法，时空复杂度都为 $O(n+m)$。

应用：2-SAT

- 问题：有 n 个布尔变量 $x_1, \ldots, x_n$。给出 m 个条件，每个条件形如 $y_1 \vee y_2 \vee \ldots \vee y_k = 1$，其中每个 y_j 形如 x_i 或者 $\neg x_i$。问是否存在一种为 $x_1, \ldots, x_n$ 赋值的方法，使得所有条件都被满足。
- 这个叫 k-SAT 问题，当 $k = 2$ 时（也就是每个条件只包含两个变量），就是 2-SAT。
- 当 $k \geq 3$ 时，k-SAT 是 NP-Complete 的，但 $k = 2$ 时存在简单多项式算法，它需要用到强连通分量的相关知识。

2-SAT 算法

- 建立 $2n$ 个点 $a_1, \ldots, a_n, b_1, \ldots, b_n$，其中 a_i 表示 $x_i = 0$ 这一陈述，b_i 表示 $x_i = 1$ 这一陈述。
- 点 p 可达点 q 当且仅当根据条件能推出 $p \to q$。由于可达满足传递性，这个定义是合理的。
- 随便考虑一个条件，例如 $x_1 \vee x_2 = 1$，该如何建边？
- $a_1 \to b_2$ 和 $a_2 \to b_1$。

建模完成后，我们希望根据这张图判断 2-SAT 是否有解。如果有解，还希望能找到一组解。

2-SAT 算法

- 如果 $a_i \to b_i$ 且 $b_i \to a_i$，那么一定无解了，因为不管 x_i 取 0 还是 1，都会导致矛盾。
- 而如果上面两个命题不同时成立，则说明 a_i, b_i 不强连通，我们断言：如果每一对 a_i, b_i 都不强连通，则 2-SAT 一定有解。下面是构造性的证明。
- 找到 a_i, b_i 中所在强连通分量编号更小的那个，如果是 a_i 则令 $x_i = 0$，否则 $x_i = 1$。
- 可以验证，每个条件都能被满足，这是因为 Tarjan 算法满足：只有编号更大的强连通分量才可能到达编号更小的强连通分量。

时间复杂度为 $O(n+m)$。

耳分解

Ex. 给定一张无向图 G，问是否能够给每条无向边确定一个方向，使得得到的有向图强连通。

双极定向

Ex. 给定一张无向图 G。给出一个点的排列，使得每个前缀和后缀的导出子图连通（或报告无解）。

割集空间和环空间

Ex. 证明割集空间和环空间关于 $\mathbb{F}_2^m$ 互为正交补。

如何找到它们各自的一组基？

Ex. 证明割集空间和环空间关于 $\mathbb{F}_2^m$ 互为正交补。

如何找到它们各自的一组基？

Hint：DFS 树。

竞赛图

Ex. Laudau 定理。将点按出度从小到大排序，i 的出度为 d_i。则每个强连通分量都是一段连续区间，且右端点 r 满足 $\sum_{i=1}^{r} d_i = \binom{r}{2}$。

代数连通度：Laplacian 的第二小特征值。

Ex. 基于谱图论，对称马尔科夫链的 mixing time bound。