18．433 组合最优化

匹配算法

September 4, 9, 11 授课教师：*Santosh Vempala*

给定一个图$G=(V,E)$，若边子集 M 内任两边不相邻，则称其为*匹配*。若顶点$v\in V$与匹配中的边相关联，则称v是*匹配的*，否则称v是*未匹配的*。含边数最多的匹配称为图G的*最大匹配*。特别地，若在此最大匹配下图G的每个顶点都是匹配的，则称其为*完美匹配*。

若图G的顶点集存在一个拆分A和B，且G中每条边恰好都是一个端点属于A而另一个端点属于B，则称图G为*二部图*。下面的定理给出了二部图存在完美匹配的条件。对于$U\subseteq A$，记$N(U)$为所有与U中顶点相邻的顶点的集合。

定理 1（**Hall**）顶点集为A和B的二部图存在完美匹配当且仅当$|A|=|B|$且$|N(U)|\geq|U|\ (\forall U\subseteq A).$

*证明：*若二部图存在完美匹配，显然右边是一个必要条件。

对顶点集为A和B的二部图，若右边条件成立，下面我们对$|A|$利用归纳法来证明此图存在完美匹配。如果$|A|$等于 0 或 1，显然结论成立。下面分两种情况考虑：

1、假设$(\forall U\subseteq A, U\neq\varnothing, U\neq A)\ |N(U)|>|U|$。令$e=(u,v)$，$G'=G-\{u\}-\{v\}$.

在G'中，对$\forall U\subseteq A-\{u\}$均有

$$|N_{G'}(U)|\ \geq\ |N(U)|-1\ \geq\ |U|.$$

由归纳假设知G'存在一个$A-\{u\}$到$B-\{v\}$的完美匹配M，则$M\cup\{e\}$为G中A到B的一个完美匹配。

2、否则：必存在非空集$A'\subset A$满足$|N(A')|=|A'|$。令G_1表示$A'\cup N(A')$的导出子图，G_2表示$G-A'-N(A')$的导出子图。

在G_1中，$(\forall U\subseteq A')\ N_G(U)=N_{G_1}(U)$，且$|N_{G_1}(U)|\geq|U|$. 由归纳假设知$G_1$存在一个$A'$到$N(A')$的匹配$M_1$。

在G_2中，对$\forall U\subseteq A-A'$有

$$N_G(U\uplus A')=N_{G_2}(U)\uplus N_G(A').$$

也就是说，

$$\begin{aligned}|N_{G_2}(U)| &= |N_G(U \uplus A'| - |N_G(A')| \\ &\geq |U \uplus A'| - |A'| \\ &= |U|.\end{aligned}$$

则归纳假设知 G_2 存在一个由 $A-A'$ 到 $B-N(A')$ 的匹配 M_2。从而，$M_1 \cup M_2$ 是图 G 的一个完美匹配。

□

在这几次讲座中我们的目的是提出一个找出给定图的最大匹配的快速算法。在整个课程中，“快速”意味着多项式时间，即算法的运行时间可以用输入图形规模的一个固定的多项式作为上界。其中，图形的规模决定于图的顶点的个数 n 和边的条数 m 。

取图 G 的一个匹配 M：如果 G 的每一个顶点都被 M 匹配，则 M 是一个完美匹配从而也是一个基数为 $\frac{n}{2}$ 的最大匹配。如果 M 不是完美匹配，那要么可以找到一个比 M 基数更大的匹配，也就是*增广*M，要么断定 M 已经是最大匹配。增广 M 的一种方法如下：找一条路 P，从未匹配顶点开始，交替的经过 M 中的边和不在 M 中的边（即匹配边和未匹配边），最后结束于未匹配顶点。删去 P 上在 M 中的边而增加路上其余的边，得到边的集合 M'，也就是 M 和 P 的对称差，记为 $M \oplus P$. 容易证 M' 也是一个匹配，而且所含边数比 M 多一条。这样的交错路 P 称为一条*可增广*路。上述分析就产生了下面的“算法”。

匹配算法：

{

1、 从任意匹配开始。

2、 找当前匹配的一条可增广路。

3、 增广当前匹配。

4、 尽可能地重复上面两步。

}

算法终止的时候，得到一个没有可增广路的匹配 M。接下来要做什么？下面的引理告诉我们此时 M 一定是最大匹配。

引理 2. M 是最大匹配当且仅当 M 没有可增广路。

证明. 我们已经知道 M 有可增广路时一定不是最大匹配。

相反地，设 M 没有可增广路，下面我们将说明 M 是一个最大匹配。思路是取最大匹配 M^* 来证明 $|M| = |M^*|$。考虑 M 和 M^*的对称差 $M \oplus M^*$，由前面可知它是在 M 中不在 M^*和在 M^*中而不在 M 中的边的集合，即 $M \oplus M^* = (M - M^*) \cup (M^* - M)$。

因为 M 和 M^* 的导出子图最大度数均为 1，故 $M \oplus M^*$ 的导出子图最大度数为 2。注意到这样一个子图仅由不交路和/或圈组成。另外，注意到 M 和 M^* 都是匹配，所以，这些路和圈上的边交错地包含于 M 和 M^*。

首先考虑导出图中的圈。所有圈的边数一定为偶数，否则一定有一个顶点与 M 或 M^* 中的两条边相邻，这与匹配的定义矛盾。所以，这些圈中含有的 M 和 M^* 的边数相等。

接下来考虑导出子图中的路。假设有一条路 P 含有奇数条边，那么要么来自 M 的边比来自 M^* 的边多一条，要么相反。在前一种情况下，P 是 M^* 的一条可增广路，与 M^* 是最大匹配矛盾。后一种情况下，P 是 M 的一条可增广路，这与我们初始的假设矛盾。因此，所有的路都含有偶数条边，也就是来自 M 和 M^* 的边数相等。

所以，在 $M \oplus M^*$ 的生成子图中路和圈中来自 M 和 M^* 的边数都相等。最后来考虑那些不在 $M \oplus M^*$ 的生成子图中的边。这些边要么同时在 M 和 M^* 中要么既不在 M 中也不在 M^* 中。综上所述，M 和 M^* 的基数相等，从而 M 是一个最大匹配。 □

那我们的算法要用多长时间呢？在第 2 和 3 步中，每做一次迭代匹配的基数增加 1，那我们最多做 $\frac{n}{2}$ 次迭代。接下来的问题就是找到一条可增广路需要多长时间。实际上我们必须弄清楚如何寻找可增广路。我们先来看*二部图*的情况，这相对容易一些。

1 二部图

给定一个二部图，M 是它的一个匹配，令 $A^U \subseteq A$，$B^U \subseteq B$ 为未被 M 匹配的顶点。我们的目标是找到 M 的一条可增广路。令 S 表示从 A^U 出发通过交错路可达的顶点的集合，如果 S 包含 B^U 中的一个顶点，那么到这个顶点的这条交错路就是 M 的一条可增广路。

集合 S 可以通过如下方法构造一个交错*森林* F 得到：

1、将 A^U 中的所有顶点加入 F，每个均为 F 的独立分支。
2、加入从 $A \cap V(F)$ 中顶点到 B 中顶点的边，注意，不要合并任何两个连通分支。也就是说，如果 B 中的一个顶点不止邻接于一个分支，那么仅将其添加到一个分支中。
3、加入 M 中与 $B \cap V(F)$ 中顶点相邻接的边。
4、重复上面两步直到再没有边被加入 F。

如果森林中存在 B^U 中的一个顶点，就给出了一条可增广路。否则，由下面的引理可知，M 即为最大匹配。

引理 3. M 是最大匹配当且仅当 F 不含有 B^U 中的顶点。

证明. 如果 F 包含 B^U 中的一个顶点 v，那么从 v 到 A^U 中顶点的这条路在顶点 v 所在的分支中是一条以未匹配的顶点为端点的交错路，也就是一条可增广路。因此 M 不是最大匹配。

相反地，如果F不含B^U中的顶点。为了证明我们的结论引入一个“*顶点覆盖*”的概念。顶点覆盖是指一个顶点子集，满足图中任一边至少有一个端点含于此子集中。下面我们要证明G中存在一个顶点覆盖，其大小等于当前匹配。又因为任意顶点覆盖的大小不小于最大匹配（匹配中每条边必有一个顶点包含于任意顶点覆盖中），这样就可以证明M是最大匹配。

令$X = A - V(F)$，$Y = B \cap V(F)$。我们不妨先假设$X \cup Y$是一个顶点覆盖。那显然M与$X \cup Y$中的每个顶点相关联。既然M是一个匹配，所以M中不存在两端点都在$X \cup Y$中的边。因为，给定一个被匹配的顶点$a \in V(F)$，设(a, b)为匹配边，那么由F的构建可知b一定也在$V(F)$中，从而M中每条边至少与$X \cup Y$中的一个端点相交。综上，$|M| = |X \cup Y|$。

剩下的问题就是来证明$X \cup Y$是顶点覆盖。假设不是，那么必存在一条以$a \in A$和$b \in B$为两端点的边(a, b)没有被覆盖，从而有$a \in V(F)$而$b \notin V(F)$，可得(a, b)不是一条匹配边。另外，$b \notin B^U$，否则它已经被添加到$V(F)$中。因此b是被匹配的，即在另一条匹配边，比如说(a', b)上，这里$a' \neq a$。但这就意味着还可以添加路aba'来扩展F，与F已经是最大矛盾。 □

从引理的证明中可以得出下面的定理。

定理 4. *(König)* 在二部图中，最大匹配的大小等于最小顶点覆盖的大小。 □

我们称A有一个到B的匹配，如果最大匹配的大小为$|A|$。另外，任意顶点子集$X \subset V$，$\Gamma(X)$表示与其相邻的顶点集合。由上面的König定理可以得出经典的 Frobenius-Hall 定理（实际上两者等价）。

定理 5.*(Frobenius-Hall)* A有一个到B的匹配当且仅当A的任意子集X满足$X \leq |\Gamma(X)|$。

证明. 显然，如果存在子集X满足$X > |\Gamma(X)|$，则一定不存在以$|A|$为基数的匹配。反之，假设$X \leq |\Gamma(X)|$对任意$X \subset A$成立，我们想证明最小顶点覆盖的大小为$|A|$，从而就可以证明本定理。由假设知A中任一顶点至少与一条边相关联，且$|A| \leq |B|$。注意到顶点集A是一个大小为$|A|$的顶点覆盖。设$X \cup Y$是另一顶点覆盖，这里$X \subset A$，$Y \subseteq B$，可以看出$\Gamma(A - X) \subseteq Y$，从而$|A - X| \leq |\Gamma(A - X)| \leq |Y|$，所以$|X \cup Y| \geq |A|$，即证明了$A$是一个最小覆盖，大小为$|A|$。 □

定理 6. 在二部图中寻找一个最大匹配需要$O(m\sqrt{n})$。

证明. 易知找一条增广路需要 $O(m)$，而最多增广 $\frac{n}{2}$次，因此总时间为$O(mn)$。我们可以对上面的分析加以改进。注意，寻找增广路的算法可能找到不止一条增广路，这种情况下，我们可以对一个最大的不交增广路的集合进行增广。做这种改进后，可以证明总的阶段（交错森林的构造）数为 $O(\sqrt{n})$。其中关键的结论如下（证明留作习题）：

结论 7. 在每个阶段中最短增广路的长度递增。

有了这个结论，可知在经过$\sqrt{n}$ 个阶段后所有的增广路长度至少为 $2\sqrt{n}+1$。考虑最大匹配 M^*和对称差 $M \oplus M^*$，如果 M 不是最大匹配，则对称差 $M \oplus M^*$中必存在一条交错路是 M 的可增广路。既然每一条可增广路长度至少为 $2\sqrt{n}+1$，且总共仅有 $O(\sqrt{n})$条这样的路，那 $|M^*|-|M| < \sqrt{n}$，因此，最多再经过$\sqrt{n}$ 个阶段算法就会终止。 □

2 一般图

不难看出前面的算法不适用于一般图。主要是因为一般图中可能存在含有最多匹配边的奇圈，即长度为 $2k+1$ 含有 k 条匹配边的圈。这样的奇圈称为*花*，图 1 给出了一个例子，其中匹配边用粗体标明。

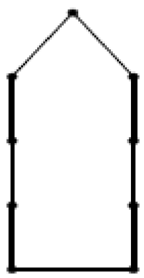

图 1：一个花

下面的引理给出了处理花的一个方法，它是求一般图中最大匹配的 Edmonds 算法的核心思想。

引理 8.*(圈收缩)* M是图 G 的一个匹配，B 是一个花，且与M中其余的边是顶点不交的（即没有共同顶点）。把B收缩为单个顶点后得到一个新图 G'，那么由M导出的匹配 M'是 G' 的最大匹配当且仅当M是图 G 的最大匹配。

证明. 首先，假设 M'不是 G' 的最大匹配。由引理 2 可知，G' 中含有 M'的一条可增广路 P'。假设 P' 在 G 中不与B相交，那 P' 一定也是 G 中的一条可增广路，从而M不是 G 的最大匹配。所以，P' 在 G 中一定与B相交。特别地，收缩后的B一定是 G' 中路 P' 上的一个端点，因为B与 M'是顶点不交的。不妨设 P' 与B相交于顶点v，u为B中未被匹配的顶点。将B中从v出发，先经过与之相关联的匹配边，最后到达 u的路称为P''，易知 $P = P' \cup P''$ 是G 中的一条可增广路，所以此时，M也不是G 的最大匹配。

接下来，假设M不是G 的最大匹配，我们要证明 M'也不是 G' 的最大匹配。取 G 中的一条增广路 P，假设其与B相交，否则 P 也是 G' 中的可增广路。注意到，B 仅包含一个未被匹配的顶点，所以 P 中至少有一个顶点，不妨设为 w，位于B外。用 P' 表示从 w 开始

沿着 P 直到与 B 相交的一段路，可知 P' 也是 G' 的一条可增广路。从而结论得证。 □

要在一般图中找一条可增广路，我们可以修改一下二部图中的算法，使之可以发现花。发现了，就对其进行收缩，然后在新的图上重新开始。在新图上找到的任意增广路都可以很容易的对应到原图中的增广路，而且，由上面的引理可知，如果在新图中匹配是最大的，那么原图中对应的匹配也是最大的。

下面是算法的正式描述。令 M 为图 G 的一个匹配，U 为未被匹配顶点的一个子集（如果每个顶点都是被匹配的，那么这个匹配就是最大匹配），我们要构造一个森林 F，使 U 中的每一个顶点在一个连通分支上。像以前那样交替的增加未匹配边和匹配边来扩展 F。那么被添加到 F 中的 M 的边与 U 的距离为奇数。而且，与 U 距离为奇数的顶点度数为 2（一条未匹配边，一条匹配边），我们把这样的顶点称为*内*顶点，而其余的称*外*顶点。U 中所有的顶点都是外顶点。

现在，考虑外顶点的相邻顶点。只可能出现下列四种情况之一：

1、 如果我们找到外顶点 x，与不在 F 中的顶点 y 相邻，那么把边 (x, y) 和 (y, z) 添加到 F 中，这里 (y, z) 是 M 的一条边。

2、 如果属于不同分支的两个外顶点是相邻的，那么这两个分支的根之间存在一条可增广路。

3、 如果同一个分支中的两个外顶点 x, y 是相邻的，那么边 (x, y) 和 F 中从 x 到 y 的路形成一个圈，记为 C。令 P 为 C 和这个分支的根之间的一条路。首先，我们可以转换 P 中的边而得到与 P 同样大小的一个匹配 M_1。然后，C 满足圈收缩引理的条件，我们可以把 C 收缩成一个顶点，从而得到一个新图 G'。接下来我们的目的就变为在 G' 中寻找一条可增广路。

4、 如果每个外顶点都仅与内顶点相邻，那么 M 已经是最大匹配。为了得出这个结论，不妨假设 F 有 p 个内顶点和 q 个外顶点，那么 $q - p = |U|$，因为每个被匹配的外顶点与一个内顶点相对应且反之亦然。如果我们从 G 中删去 F 所有的内顶点，那每个外顶点在一个独立分支上。而且这意味着 G 中的任意匹配至少要漏掉外顶点中的，也是图 G 中的 q－p 个顶点。而 M 恰好漏掉了 q－p 个顶点，所以它一定是最大匹配。

从算法的描述可以得出下面的引理：

引理 9. 在算法的每一步，我们或者增大 F，或者减小 G，或者找到一条可增广路，再或者就是终止于一个最大匹配。 □

定理 10. 找到一个最大匹配需要 $O(n^4)$。

证明. 显然算法的增广次数小于 n。另外，在找到一条增广路前最多收缩 n 个花。寻找一条增广路或花需时 $O(m)$，因为在森林的构建中每条边最多检查一次。所以算法总需时为 $O(mn^2) = O(n^4)$。 □

从 Edmonds 算法中可以得到下面的定理。

定理 11(Tutte). 图G存在完美匹配当且仅当对任意的顶点子集X, 从G中删去X得到的图$G \setminus X$中所含的奇分支个数最多为$|X|$。

证明. 右边条件的必要性是显然的：如果存在某个顶点子集X满足$G \setminus X$含有的奇分支个数大于$|X|$，那么X中没有足够的顶点去匹配所有的奇分支，因为奇分支需要一个外部顶点去匹配且只能用X中的顶点去匹配。

要证明右边条件的充分性，考虑 Edmonds 算法最后一步中的森林。用X表示内顶点集，$p = |X|$。因为收缩得到的顶点一定是未被匹配的，所以X中不含这样的顶点。考虑$G \setminus X$，其中外顶点分别在不连通的分支上（算法之所以结束，就是因为外顶点仅与内顶点相邻）。它们中有一些是原图中的花收缩得到的。跟前面算法的描述中一样，我们用q表示外顶点的个数，那么未被匹配的顶点个数为$q - p$。那么由假设（应用于集合X）可知，在原图中最多有$|X|$个奇分支，也就是说最多有 p 个外顶点。换句话说，$q = p$且所有的顶点都是被匹配的。

在他描述这个算法的论文（名为“Paths, Trees and Flowers”）中，Edmonds 还给出了多项式时间算法的定义，在以后的数十年中，这个定义在复杂性理论中扮演了一个重要的角色。

注：

Observation　观测结论：从观测得出的推论或判断　　　译为：“结论”