18.997 Topics in Combinatorial Optimization	February 5, 2004

Lecture 2

Lecturer: Michel X. Goemans *Scribe: Robert Kleinberg*

In this lecture, we will:

- Present Edmonds' algorithm for computing a maximum matching in a (not necessarily bipartite) graph G.
- Use the analysis of the algorithm to derive the Edmonds-Gallai Decomposition Theorem stated in the last lecture.

1 Recapitulation

Recall the following essential definitions and facts from the last lecture. A *matching* in an undirected graph G is a set of edges, no two of which share a common endpoint. Given a graph G and a matching M, a vertex is *matched* if it is the endpoint of an edge in M, *unmatched* otherwise; we will often designate the set of unmatched vertices by X. Given a graph G with matching M, an *M-alternating path* is a path whose edges are alternately in M and not in M. (Here we use *path* to mean a simple path, i.e. one with no repeated vertices. We'll refer to a non-simple path as a *walk*.) If both endpoints of an M-alternating path belong to the set X of unmatched vertices, it is called an *M-augmenting path*. Recall the following theorem from last time.

Theorem 1 *A matching M is of maximum size if and only if G contains no M-augmenting path.*

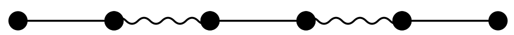

Figure 1: An M-augmenting path

2 Flowers, Stems, and Blossoms

The following construction is useful for finding M-augmenting paths. Given a graph $G = (V, E)$ with matching M; construct a directed graph $\hat{G} = (V, A)$ with the same vertex set as G, and with edge set determined by the rule that $(u, w) \in A$ if and only if there exists v with $(u, v) \in E \setminus M$, $(v, w) \in M$. Observe that every M-augmenting path in G corresponds to a path in $\hat{G}$ which begins at a vertex in X and ends at a neighbor of X. However, the converse is not true, because an M-alternating walk may begin at a vertex in X and end at a neighbor of X, without being an M-augmenting path, if it contains an odd cycle. Figure 2 illustrates an example of such a walk. This motivates the following definition.

Definition 1 *An M-flower is an M-alternating walk $v_0, v_1, v_2, \ldots, v_t$ (numbered so that $(v_{2k-1}, v_{2k}) \in M$, $(v_{2k}, v_{2k+1}) \notin M$) satisfying:*

1. $v_0 \in X$.

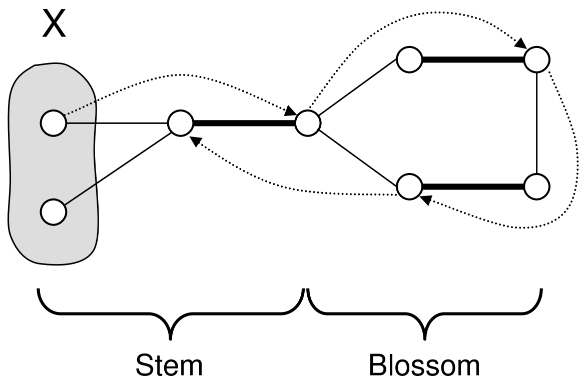

Figure 2: An M-flower

2. *$v_0, v_1, v_2, \ldots, v_{t-1}$ are distinct.*

3. *t is odd.*

4. *$v_t = v_i$, i even.*

The portion of the flower from v_0 to v_i is called the stem*, while the portion from v_i to v_t is called the* blossom*.*

Lemma 2 *Let M be a matching in G, and let $P = (v_0, v_1, \ldots, v_t)$ be a shortest alternating walk from X to X. Then either P is an M-augmenting path, or $v_0, v_1, \ldots, v_j$ is an M-flower for some $j < t$.*

Proof: If $v_0, v_1, \ldots, v_t$ are all distinct, P is an M-augmenting path. Otherwise, assume $v_i = v_j$, $i < j$, and let j be as small as possible, so that $v_0, v_1, \ldots, v_{j-1}$ are all distinct. We shall prove that $v_0, v_1, \ldots, v_j$ is an M-flower. Properties 1 and 2 of a flower are automatic, by construction. It cannot be the case that j is even, since then $(v_{j-1}, v_j) \in M$, which gives a contradiction in both of the following cases.

- $i = 0$: $(v_{j-1}, v_j) \in M$ contradicts $v_0 \in X$.
- $0 < i < j - 1$: $(v_{j-1}, v_j) \in M$ contradicts the fact that M is a matching, since v_i is already matched to a vertex other than v_{j-1}.

This proves that j is odd. It remains to show that i is even. Assume, by contradiction, that i is odd. Then $v_{j+1} = v_{i+1}$ (since both are equal to the other endpoint of the matching edge containing $v_j = v_i$), and we may delete the cycle from P to obtain a shorter alternating walk from X to X. (See Figure 3.) $\square$

Given a flower $F = (v_0, v_1, \ldots, v_t)$ with blossom B, observe that for any vertex $v_j \in B$ it is possible to modify M to a matching M' satisfying:

1. Every vertex of F belongs to an edge of M' except v_j.
2. M' agrees with M outside of F, i.e. $M \triangle M' \subseteq F$.
3. $|M'| = |M|$.

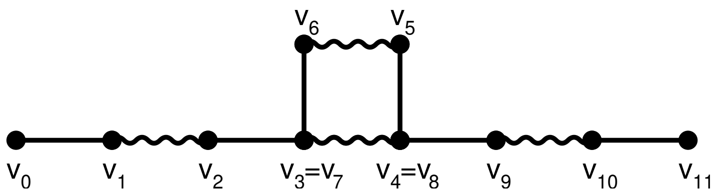

Figure 3: An alternating walk from X to X which can be shortened.

To do so, we take M' to consist of all the edges of the stem which do not belong to M, together with a matching in the blossom which covers every vertex except v_j.

Whenever a graph G with matching M contains a blossom B, we may simplify the graph by *shrinking* B, a process which we now define.

Definition 2 (Shrinking a blossom) *Given a graph $G = (V, E)$ with a matching M and a blossom B, the* shrunk graph G/B *with matching M/B is defined as follows:*

- $V(G/B) = (V \setminus B) \cup \{b\}$
- $E(G/B) = E \setminus E[B]$
- $M/B = M \setminus E[B]$

where $E[B]$ denotes the set of edges within B, and b is a new vertex disjoint from V.

Observe that M/B is a matching in G, because the definition of a blossom precludes the possibility that M contains more than one edge with one but not both endpoints in B. Observe also that G/B may contain parallel edges between vertices, if G contains a vertex which is joined to B by more than one edge.

The relation between matchings in G and matchings in G/B is summarized by the following theorem.

Theorem 3 *M is a maximum size matching in G if and only if M/B is a maximum size matching in G/B.*

Proof: ($\Longrightarrow$) Suppose N is a matching in G/B larger than M/B. Pulling N back to a set of edges in G, it is incident to at most one vertex of B. Expand this to a matching in N^+ in G by adjoining $\frac{1}{2}(|B| - 1)$ edges to match every other vertex in B. Then $|N^+|$ exceeds $|M|$ by the same amount that $|N|$ exceeds $|M/B|$.

($\Longleftarrow$) If M is not of maximum size, then change it to another matching M', of equal cardinality, in which B is an entire flower. (If S is the stem of the flower whose blossom is B, then we may take $M' = M \triangle S$.) Note that M'/B is of the same cardinality as M/B, and b is an unmatched vertex of M'/B. Since M' is not a maximum size matching in G, there exists an M'-augmenting path P. At least one of the endpoints of P is not in B. So number the vertices of P $u_0, u_1, \ldots, u_t$ with $u_0 \notin B$, and let u_i be the first node on P which is in B. (If there is no such node, then $u_i = u_t$.) This sub-path $u_0, u_1, \ldots, u_i$ is an (M'/B)-augmenting path in G/B. $\square$

```
M := ∅
X := {unmatched vertices}
Form the directed graph Ĝ.
while Ĝ contains a directed path P̂ from X to N(X)
    Find such a path P̂ of minimum length.
    P := the alternating path in G corresponding to P̂
    if P is an M-augmenting path,
        modify M by augmenting along P.
    else
        P contains a blossom B.
        Recursively find a maximum size matching M′ in G/B.
        if |M′| = |M/B|    /* M is already a max matching. */
            return M       /* Done! */
        else               /* M can be enlarged */
            Unshrink M′ as in the proof of Theorem 3,
             to obtain a matching in G of size > |M|.
end
```

Figure 4: Algorithm for computing a maximum matching

3 A polynomial-time maximum matching algorithm

The algorithm for computing a maximum matching is specified in Figure 4.

The correctness of the algorithm is established by Lemma 2 and Theorem 3. The running time may be analyzed as follows. We can compute X and $\hat{G}$ in linear time, we can find $\hat{P}$ in linear time (by breadth-first search), and we can shrink a blossom in linear time. We can only perform $O(n)$ such shrinkings before terminating or increasing $|M|$. The number of times we increase $|M|$ is $O(n)$. Therefore the algorithm's running time is $O(mn^2)$. With a little more work, this can be improved to $O(n^3)$. (See Schrijver's book.) The fastest known algorithm, due to Micali and Vazirani, runs in time $O(\sqrt{n}\,m)$.

4 Combinatorial consequences of the algorithm

Our aim now is to use the analysis of the algorithm to derive the following two combinatorial theorems, both of which were asserted in the last lecture.

Theorem 4 (Tutte-Berge Formula) *For a graph G and a set of vertices $U \subseteq V(G)$, let $c_o(G \setminus U)$ denote the number of odd components of the graph $G \setminus U$, i.e. the number of components with an odd number of vertices. Then the cardinality of a maximum size matching, $\nu(G)$, satisfies:*

$$\nu(G) = \min_{U \subseteq V} \frac{1}{2} \left[|V| + |U| - c_o(G \setminus U) \right]. \tag{1}$$

Theorem 5 (Edmonds-Gallai Decomposition) *Given a graph G, let*

$$\begin{aligned} D(G) &:= \{v : \exists \text{ a maximum size matching missing } v\} \\ A(G) &:= N(D(G)) \\ C(G) &:= V(G) \setminus (D(G) \cup A(G)) \end{aligned}$$

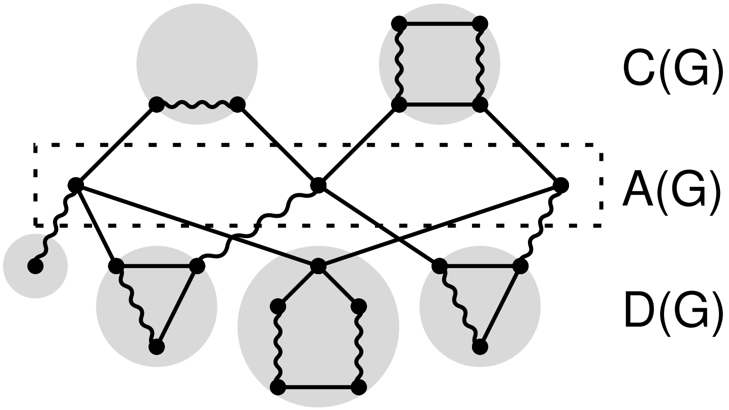

Figure 5: Edmonds-Gallai Decomposition

Then $U = A(G)$ achieves the minimum on the right side of the Tutte-Berge formula, $D(G)$ is the union of the odd components of $G \setminus A(G)$, and $C(G)$ is the union of the even components of $G \setminus A(G)$. Moreover, every odd component of $G \setminus A(G)$ is factor-critical. (A graph H is factor-critical *if for every vertex v, there is a matching in H whose only unmatched vertex is v.)*

To prove these theorems, consider a maximum size matching M in G, take an unmatched vertex $x \in X$, and consider all the vertices which can be reached by an alternating path from x. The first edge on such a path must lie outside of M, the second edge must lie in M, and so on, leading to a picture as in Figure 6.

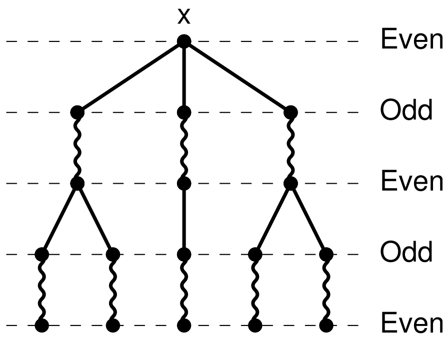

Figure 6: Vertices reachable by alternating paths from x.

Motivated by this picture, we make the define the following three subsets of $V(G)$:

$$\begin{aligned}
\mathsf{Even} &:= \{v : \exists \text{ an alternating path of even length from } X \text{ to } v\} \\
\mathsf{Odd} &:= \{v : \exists \text{ an alternating path from } X \text{ to } v\} \setminus \mathsf{Even} \\
\mathsf{Free} &:= \{v : \nexists \text{ an alternating path from } X \text{ to } v\}
\end{aligned}$$

We will sometimes refer to a vertex as being even, odd, or free, according to which of these sets it belongs to.

Claim 6 *If there is an edge from* Even *to v, then there is an alternating walk of odd length from X to v, and there is an alternating path from X to v.*

Proof: If $e = (u, v)$ is an edge between Even and v, and P is an alternating path of even length from X to u, then an alternating walk of odd length from X to v is constructed as follows. If $e \in M$, then we take P and delete the final edge, which is necessarily e. If $e \notin M$, then we append e to P. If this alternating walk is not a path, it can only be because v lies on P, in which case P contains a sub-path which is an alternating path from X to v. □

Corollary 7 *In G there is no edge between* Even *and* Free*.*

Define the *shrunk graph* G_0 to be the graph obtained at the innermost level of the recursion in the matching algorithm given above, on the final iteration of the **while** loop. Let M_0 be the maximum size matching in G_0 computed by the algorithm. Since G_0 has no flowers, and M_0 is a maximum matching, it follows that G_0 has no alternating walk from X to X.

Claim 8 *In G_0, there is no edge between two vertices in* Even*.*

Proof: If such an edge $e = (u, v)$ exists, then by Claim 6, G_0 contains an alternating walk P of odd length from X to v. But v is even, so there is also an alternating path P' of even length from X to v. Concatenating P with the reverse of P', we obtain an alternating walk from X to X, contradicting the definition of G_0. □

It is worth noting that Claim 8 doesn't necessarily hold in G. This is because all the vertices of a blossom are even. (The stem is an even-length alternating path from X to one vertex v of the blossom, and all other vertices of the blossom are reachable from v by an even-length alternating path which goes around the blossom either clockwise or counter-clockwise.)

Claim 9 Even $= D(G) = \{v : \exists$ *a maximum-size matching missing* $v\}$.

Proof: Certainly if v is even then there a maximum-size matching M' missing v. Such a matching is obtained by taking an even-length alternating path P from X to v and putting $M' = M \triangle P$. Conversely, if there exists a maximum-size matching M' missing v, then $M \triangle M'$ is a union of even-length cycles and paths, and v is an endpoint of one of these paths, because it does not belong to an edge of M'. The other endpoint of this path P does not belong to an edge of M, i.e. it is an element of X. This confirms that P is an even-length alternating path from X to v. □

Claim 10 Odd $= A(G) = N(D(G))$.

Proof: If v is odd, then there is an alternating path of odd length from X to v. The vertex preceding v on this path must be even, which confirms that Odd $\subseteq N(\text{Even}) = N(D(G))$. The reverse inclusion follows from Claim 6, which ensures that every vertex adjacent to Even belongs to Even $\cup$ Odd. □

Claim 11 Free $= C(G) = V(G) \setminus (D(G) \cup A(G))$.

Proof: Immediate from the definition of Free, and from the preceding two claims which identify Even, Odd with $D(G), A(G)$, respectively. □

Claim 12 *In G_0, every free vertex is matched to another free vertex by M, and every odd vertex is matched to an even vertex by M.*

Proof: Every vertex which is free or odd is incident to an edge of M, because no such vertex may belong to X. If $e = (u, v)$ is an edge of M with u odd, and if P is an odd-length alternating path from $x \in X$ to u, then $P \cup \{e\}$ is an even-length alternating path from X to v. (It is not possible that $v \in P$ because $v \notin X$ and every vertex in $P \setminus \{x\}$ is already matched to a vertex of P other than u.) This proves that every odd vertex is matched to an even vertex. That means a free vertex may not be matched to an odd vertex, but it also may not be matched to an even vertex (by Corollary 7), so every free vertex is matched to another free vertex. $\square$

Claim 13 *Every component of $G \setminus A(G)$ is a subset of either $D(G)$ or $C(G)$. The even-cardinality components are subsets of $C(G)$, while the odd-cardinality components are subsets of $D(G)$. Moreover, if M is a maximum-size matching in G, then every component H of $D(G)$ satisfies one of the following:*

- *$|X \cap H| = 1$, and $M \cap \delta(H) = \emptyset$. (The* coboundary *of a vertex set U, denoted by $\delta(U)$, is the set of edges with exactly one endpoint in U.)*
- *$X \cap H = \emptyset$, $M \cap \delta(H)$ contains exactly one edge, and this edge joins H to $A(G)$.*

Proof: The proof is by induction on the number of blossoms which are shrunk during the execution of the maximum matching algorithm. If no blossoms are shrunk, then $G = G_0$, and the claim is a consequence of the following observations:

1. If u is an even vertex of G_0, then every neighbor of u is in $\mathsf{Odd} = A(G_0)$. (By Corollary 7 and Claim 8.)
2. Therefore every vertex u in $\mathsf{Even} = D(G_0)$ is an isolated vertex of $G_0 \setminus A(G_0)$. Moreover, u either belongs to X, or is joined to $A(G_0)$ by an edge of M.
3. For every component H of $\mathsf{Free} = C(G_0)$, the edge set $M \cap E[H]$ is a perfect matching in H. (By Claim 12.)

Now for the induction step, suppose B is a blossom in G and that the claim holds for G/B. Then B corresponds to a vertex $b \in G/B$ which is an even vertex in some component H_b of $D(G/B)$. (The stem of the flower containing B corresponds to an even-length alternating path from X to b in G/B.) When we inflate b to B, we claim that:

1. Except for b, all even vertices of G/B remain even. All vertices of B are also even in G.
2. All odd vertices of G/B remain odd.
3. All free vertices of G/B remain free.

If these are true, then we'll be done, because this says that inflating b to B doesn't change the set $A(G)$, and it doesn't change the components of $G \setminus A(G)$ except that a vertex of H_b inflates into an odd cycle. Note that this doesn't change the parity of $|V(H_b)|$. Also, inflating b to B doesn't change the number of unmatched vertices in H_b, nor does it change the number of matching edges in $\delta(H)$.

It remains to prove (1)-(3). For (1), let $b^* \in B$ denote the vertex where the stem joins the blossom. If $P = (v_0, v_1, \ldots, v_t)$ is an even-length path in G/B with $v_0 \in X$ and $v_t \neq b$, then one of the following cases applies.

- P avoids b. In this case, P is also a path in G, and there is nothing to prove.
- $b = v_s$, s even. In this case, the sub-path $(v_0, \ldots, v_s)$ lifts to a path $P_0 = (v_0, \ldots, v_{s-1}, b^*)$ in G ending with the last edge of the stem. The next edge (v_s, v_{s+1}) corresponds to an edge (w, v_{s+1}) in G, with $w \in B$. Let P_1 be an even-length alternating path in B from b^* to w. We can splice together P_0, P_1, and the path $P_2 = (w, v_{s+1}, v_{s+2}, \ldots, v_t)$ to obtain an even-length alternating path in G from v_0 to v_t.

- $b = v_s$, s odd. In this case, the sub-path $(v_s, \ldots, v_t)$ lifts to a path $P_2 = (b^*, v_{s+1}, v_{s+2}, \ldots, v_t)$ in G beginning with the last edge of the stem. The desired even-length path in G from v_0 to v_t is constructed by a splicing process as before, but this time in reverse.

Finally, every vertex w in B is an even vertex of G, because we may obtain an even-length alternating path from X to w by taking the stem and appending an even-length path in B from b^* to w.

To prove (2) and (3), let $A^*(G)$ denote the set of vertices in G which are odd vertices of G/B. (We wish to eventually prove that $A^*(G) = A(G)$, but for now we will not assume it.) From the induction hypothesis, we have the following characterization of components of $G \setminus A^*(G)$: each such component H satisfies

1. $|V(H)|$ is even. $X \cap H = \emptyset$, and $M \cap \delta(H) = \emptyset$.
2. $|V(H)|$ is odd. $|X \cap H| = 1$, and $M \cap \delta(H) = \emptyset$, or
3. $|V(H)|$ is odd. $X \cap H = \emptyset$, and $M \cap \delta H$ consists of a single edge joining H to $A^*(G)$.

Components of the first two types will be called *inaccessible*. Components of the third type will be called *accessible*, and the edge $M \cap \delta(H)$ will be called the *gateway* to such a component. The terminology is justified by the following characterization of alternating paths in G which begin at a vertex $x \in X$: such a path P does not visit any inaccessible component except for the one containing x, and if P visits an accessible component H, then it reaches H by traversing the gateway edge. The proof is by contradiction: if not, let H_0 be the first component of $G \setminus A^*(G)$ not containing x which is reached by traversing an edge $e = (v, w) \notin M$. We must have $u \in A^*(G)$ since there are no edges between distinct components of $G \setminus A^*(G)$. The edge preceding e in P is an edge $e' = (u, v) \in M$. Since $u \in A^*(G)$, e' is a gateway edge and v belongs to some other component H_1 of $G \setminus A^*(G)$. P could not have reached H_1 by traversing e' (since it is a simple path, and it exits H_1 by traversing e'), so it must have reached H_1 via a non-gateway edge, contradicting the fact that H_0 was the first such component.

This characterization of alternating paths in G immediately proves (3), since components of $G \setminus A^*(G)$ corresponding to free vertices of G/B satisfy (3) and are inaccessible. To see that it also proves (2), consider any $v \in A^*(G)$. In G/B there is an alternating path of odd length from X to v; this is also an alternating path in G, so all that remains is to show that G contains no alternating path of even length from X to v. Let $e = (u, v)$ be the edge of M containing v. If P is an alternating path of even length from X to v, then e must be the last edge of P. But u belongs to an accessible component H, and e is its gateway edge. But this means the only way for P to reach H is to traverse e (because P starts in X, and $X \cap H = \emptyset$), and this contradicts the fact that P is a simple path. $\square$

Claim 14 $|M| = \frac{1}{2}\left[|V| + |A(G)| - c_o(G \setminus A(G))\right]$.

Proof: Every vertex in $V \setminus X$ belongs to one and only one edge in M, so

$$|M| = \frac{1}{2}\left(|V| - |X|\right). \tag{2}$$

Now, Claim 13, establishes that each odd component H of $G \setminus A(G)$ satisfies one of the following two criteria:

- $|X \cap H| = 1$, $M \cap \delta(H) = \emptyset$.
- $X \cap H = \emptyset$, and $M \cap \delta(H)$ consists of a single edge joining H to $A(G)$. (Moreover, every vertex in $A(G)$ is an endpoint of exactly one such edge.)

Hence,

$$c_o(G) = |X| + |A(G)|. \tag{3}$$

Combining (2) and (3) we obtain the desired formula. $\square$

This claim establishes that

$$\nu(G) \geq \min_{U \subseteq V} \frac{1}{2} \left[|V| + |U| - c_o(G \setminus U) \right].$$

The reverse inequality is trivial, so we have proved the Tutte-Berge formula. Note that we have also established all of the claims in the Edmonds-Gallai Decomposition Theorem, except for the assertion that every component of $D(G)$ is factor-critical. This part of the theorem will be addressed in the next lecture.