18.997 Topics in Combinatorial Optimization	February 3rd, 2004

Lecture 1

Lecturer: Michel X. Goemans | *Scribe: Nick Harvey*

1 Nonbipartite Matching

Our first topic of study is matchings in graphs which are not necessarily bipartite. We begin with some relevant terminology and definitions. A *matching* is a set of edges that share no endvertices. A vertex v is *covered* by a matching if v is incident with an edge in the matching. A matching that covers every vertex is known as a *perfect matching* or a 1-*factor* (i.e., a spanning regular subgraph in which every vertex has degree 1). We will let $\nu(G)$ denote the cardinality of a maximum matching in graph G. A *vertex cover* is a set C of vertices such that every edge is incident with at least one vertex in C. The minimum cardinality of a vertex cover is denoted $\tau(G)$. The following simple proposition relates matchings and vertex covers.

Proposition 1 *If M is a matching and C is a vertex cover then $|M| \leq |C|$.*

Proof: For each edge in M, at least one of the endvertices must be in C, since C covers every edge. Since the edges in M do not share any endvertices, we must have $|M| \leq |C|$. □

This proposition implies that $\nu(G) = \max_M |M| \leq \min_C |C| = \tau(G)$, so $\nu(G) \leq \tau(G)$. Kőnig showed that in fact equality holds if G is a bipartite graph with no isolated vertices. Unfortunately if G is not bipartite then we may have $\nu(G) < \tau(G)$. For example, if G is the cycle on three vertices then $\nu(G) = 1$ but $\tau(G) = 2$. We will give another upper-bound for $\nu(G)$ after introducing some more definitions.

If $G = (V, E)$ is a graph and $U \subseteq V$, $G - U$ denotes the subgraph of G obtained by deleting the vertices of U and all edges incident with them. Let $o(G-U)$ denote the number of components of $G - U$ that contain an odd number of vertices. Let M be a matching in $G - U$ and consider a component of $G - U$ with an odd number of vertices. There must be at least one unmatched vertex v in this component, since any matching necessarily covers an even number of vertices. Treating M as a matching in G, it is possible that we could increase the size of M by matching v with some vertex in U. However, we can add at most $|U|$ edges to M in this manner, since the vertices in U will eventually all be matched. Thus any matching in G must have least $o(G - U) - |U|$ unmatched vertices. This argument shows that the maximum size of a matching is upper-bounded by $(|V| + |U| - o(G - U))/2$, for any subset U. The following theorem strengthens this result.

Theorem 2 (Tutte-Berge Formula) *Let $G = (V, E)$ be a graph. Then*

$$\nu(G) = \max_M |M| = \min_{U \subset V} (|V| + |U| - o(G - U))/2,$$

where the maximization is over all matchings M in G.

Proof: We will consider the case that G is connected. If G is not connected, the result follows by adding the formulas for the individual components. The proof proceeds by induction on the order of G. If G has at most one vertex then the result holds trivially. Otherwise, suppose that G has at least two vertices. We consider two cases.

Case 1: G contains a vertex v that is covered by *all* maximum matchings. The subgraph $G - v$ cannot have a matching of size $\nu(G)$, otherwise that would give a maximum matching for G that leaves v unmatched. Thus $\nu(G - v) = \nu(G) - 1$. By induction the result holds for the graph

$G - v$, so there exists a set $U' \subset V - v$ that achieves equality in the Tutte-Berge Formula. Defining $U = U' \cup \{v\}$, we see that

$$\begin{aligned}\nu(G) &= \nu(G - v) + 1 \\ &= (|V - v| + |U'| - o(G - v - U'))/2 + 1 \\ &= ((|V| - 1) + (|U| - 1) - o(G - U))/2 + 1 \\ &= (|V| + |U| - o(G - U))/2\end{aligned}$$

Case 2: For every vertex $v \in G$, there is a maximum matching that does not cover v. We will prove that each maximum matching leaves exactly one vertex uncovered. Suppose to the contrary, that is, each maximum matching leaves at least two vertices uncovered. We choose a maximum matching M and its two uncovered vertices u and v such that we minimize $d(u, v)$, the distance between vertices u and v. If $d(u, v) = 1$ then the edge uv can be added to M to obtain a larger matching, which is a contradiction.

Otherwise, $d(u, v) \geq 2$ so we may fix an intermediate vertex t on some shortest u-v path. By the assumption of the present case, there is a maximum matching N that does not cover t. Furthermore, we may choose N such that its symmetric difference with M is minimal. If N does not cover u then (N, u, t) contradicts our choice of (M, u, v). Thus N covers u and, by symmetry, v as well. Since N and M both leave at least two vertices uncovered, there exists a second vertex $x \neq t$ that is covered by M but not by N. Let xy be the edge in M that is incident with x. If y is also uncovered by N then $N + xy$ is a larger matching than N, a contradiction. So let yz be the edge in N that is incident with y, and note that $z \neq x$. Then $N + xy - yz$ is a maximum matching that does not cover t and has smaller symmetric difference with M than N does. This contradicts our choice of N, so each maximum matching must leave exactly one vertex uncovered. Then $\nu(G) = (|V| - 1)/2$. The Tutte-Berge Formula then follows by choosing $U = \emptyset$. □

A natural question to ask next is: Given a graph G, what is a set $U \subset V(G)$ giving equality in the Tutte-Berge Formula? Such a set is provided by the **Edmonds-Gallai Decomposition** of G. This decomposition partitions $V(G)$ into three sets: $D(G)$ is the set of all vertices v such that there is some maximum matching that leaves v uncovered, $A(G)$ is the neighbour set of $D(G)$, and $C(G)$ is the set of all remaining vertices.

Theorem 3 *The set $U = A(G)$ gives equality in the Tutte-Berge Formula. The set $D(G)$ contains all vertices in odd components of $G - U$, and $C(G)$ contains all vertices in even components of $G - U$.*

Let $G[\,D(G)\,]$ be the subgraph of G induced by $D(G)$. It turns out that every connected component H of $G[\,D(G)\,]$ is *factor critical*, meaning that $H - v$ has a perfect matching for every $v \in V(H)$. Thus for any odd component in $G[\,D(G)\,]$ we can actually choose any particular vertex to be left uncovered.

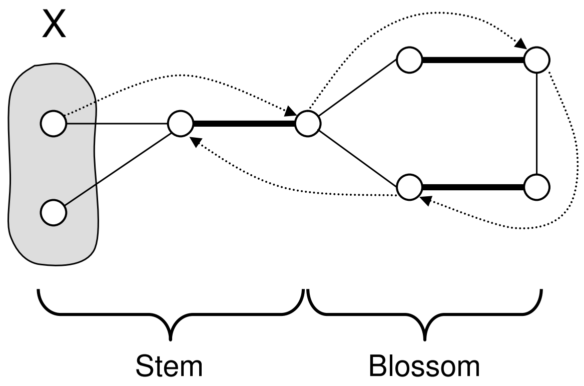
The Edmonds-Gallai Decomposition of a graph can be found as a byproduct of Edmonds' algorithm for finding a maximum matching. Before describing this algorithm, we need some more basic results. Let M be a matching in a graph G. An *alternating path* (relative to M) is a path P whose edges are alternately in M and not in M. An *augmenting path* for M is an alternating path with both endvertices uncovered by M. Let M' be the matching obtained by switching M-edges and non-M-edges along path P (i.e., $M' = M \triangle E(P)$). Then $|M'| = |M| + 1$, which explains why P is called an augmenting path.

Theorem 4 (Berge) *M is a maximum matching if and only if G contains no M-augmenting path.*

Proof: The "only if" direction is trivial, since any augmenting path can be used to increase the size of M. To prove the other direction, suppose that M is not maximum and let N be a maximum matching chosen with minimum symmetric difference with M. Consider the subgraph spanned by

$M \cup N$. Each vertex has degree at most 2, so the subgraph is a disjoint union of paths and cycles. There are no cycles or paths with equal number of edges from N and M, since $N \triangle M$ is minimum. There are no paths with more N-edges than M-edges otherwise N would not be maximum. It follows that every component is an augmenting path for M. $\square$

Theorem 4 implies the following approach for finding a maximum matching: start with an empty matching and repeatedly find augmenting paths to increase its size. **Edmonds' Algorithm** uses this approach and gives a specific method for finding augmenting paths. Consider a graph $G = (V, E)$ and a matching M in G. Let X be the set of uncovered vertices in G. To find an augmenting path for M, it will be helpful to define an auxiliary directed graph G' with vertex set V and arc set $A = \{uv \mid \exists x \in V \text{ such that } ux \in E \text{ and } xv \in M\}$. Observe that a directed path in G' corresponds to an (even length) alternating path in G. Furthermore, if there is an augmenting path for M then there is a directed path in G' starting at a vertex in X and ending at a neighbour of X. Unfortunately, the converse does not necessarily hold: G may contain a directed path in G' starting at a vertex in X and ending at a neighbour of X that does *not* correspond to an augmenting path. Such a path must necessarily have a prefix that is a *flower*, as shown in this figure.



The dotted arcs show a directed path in the auxiliary graph that starts at a vertex in X and ends at a neighbour of set X but does not correspond to an augmenting path. The graph contains flower, which consists of a *stem* and a *blossom*. The stem is simply an alternating path and the blossom is an odd-length cycle.