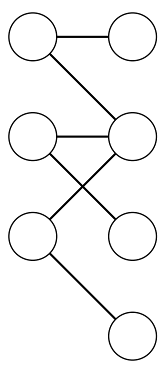
Complementi di Algoritmi e Strutture Dati

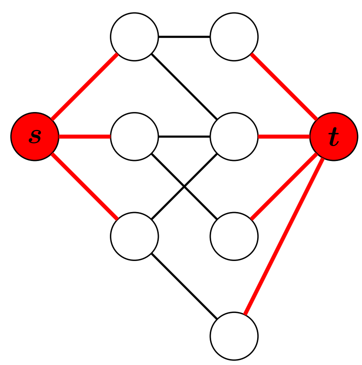
The matching theorem

Instructor: *Nicolò Cesa-Bianchi* version of June 8, 2021

A set of edges in a graph $G = (V, E)$ is independent if no two edges have an incident vertex in common. Independent sets of edges are called **matchings**. M is a matching of $U \subseteq V$ if every vertex in U is incident with some edge in M. The vertices in U are then called matched (by M); vertices not incident with any edge of M are unmatched. A **perfect matching** in $G = (V, E)$ is a matching of all vertices in V. We want to find conditions ensuring the existence of large or perfect matchings in arbitrary graphs.

Matchings in bipartite graphs. An important special case of matching considers bipartite graphs $G = (V, E)$ where $V = X \cup Y$ and $X \cap Y = \emptyset$. We study this special case through the lens of the max-flow min-cut theorem. First, we transform G in a flow network $G' = (V', E')$ where $V' \equiv V \cup \{s, t\}$ and $E' = E \cup \{(s, x) : (x \in X\} \cup \{(y, t) : y \in Y\}$, see figure below here.





The **max flow** problem in G' is to find an admissible flow of maximum value between s and t under a capacity constraint $c : E \to \mathbb{R}$, where $c(e) \geq 0$ is a nonnegative capacity assigned to each edge.

A **flow** is a function $f : E \to \mathbb{R}$ assigning $f(e) \geq 0$ to each e of G'. A flow f is **admissible** when the two following set of constraints are satisfied:

1. **capacity** constraints, $f(e) \leq c(e)$ for all $e \in E$.
2. **flow conservation** constraints, for all $v \in V$,

$$f(s, x) = \sum_{y \in Y \,:\, (x,y) \in E} f(x, y) \qquad x \in X$$

$$f(y, t) = \sum_{x \in X \,:\, (x,y) \in E} f(x, y) \qquad y \in Y$$

The **value** of an admissible flow f is

$$V_f = \sum_{x \in X} f(s, x) = \sum_{y \in Y} f(y, t)$$

A **cut** of G' is a partition S, T of V' such that $s \in S$ and $t \in T$. The **cost** of a cut is $c(\Gamma) = \sum_{e \in \Gamma} c(e)$, where $\Gamma = \Gamma(S,T) = \{(u,v) \in E : u \in S,\, v \in T\}$. The **min-cut** problem is to find a cut of minimum cost. The next result, which we do not prove here, is a fundamental consequence of linear programming duality.

Theorem 1 (Max-flow min-cut) *In any flow network, the maximum value of an admissible flow equals the minimum cost of a cut.*

We also use (without proof) this important fact.

Theorem 2 (Integral flow) *If each edge in a flow network has integral capacity, then there exists an integral admissible flow of maximum value.*

Fact 3 *Let $G = (V, E)$ be a bipartite graph and let G'_∞ be the flow network such that $c(s,x) = c(y,t) = 1$ for all $x \in X$ and $y \in Y$, and $c(x,y) = \infty$ for all $(x,y) \in E$. Then the value of the maximum flow in G'_∞ equals the size of a maximum matching in E.*

Proof. Consider first the flow network G' derived by G by setting $c(e) = 1$ for all $e \in E'$. Due to the integral flow theorem, and recalling that $c(e) = 1$ for all $e \in E'$, there exists a maximum flow f^* such that $f^*(e) \in \{0,1\}$ for all $e \in E'$. Due to the flow conservation constraints, if $f^*(x,y) = 1$ for some $(x,y) \in E$, then it must be $f^*(x,y') = 0$ for all $y' \in Y \setminus \{y\}$ and $f^*(x',y) = 0$ for all $x' \in X \setminus \{x\}$. Hence f^* defines a set of V_f edge-disjoint paths from s to t of the form $P = \{(s,x),(x,y),(y,t)\}$ with $f(e) = 1$ if and only if $e \in P$.

$$M = \{(x,y) \in E \,:\, f^*(x,y) = 1\}$$

is a matching in G of size $|M| = V_{f^*}$. Hence the maximum matching M^* satisfies $|M^*| \geq V_{f^*}$. Now assume that by assigning an arbitrarily higher integer capacity to the edges in E, V_{f^*} also increases. Then there must be a pair of edges (s,x) and (y,t) whose assigned flow went from 0 to 1. However, because of the flow conservation constraints, the corresponding flow on (x,y) cannot be larger than 1. But this contradicts the hypothesis that V_{f^*} was maximum for G'. Hence V_{f^*} is also maximum for G'_∞.

For the other direction, let M^* be a maximum matching in G. Then there exists an edge-disjoint path $P = \{(s,x),(x,y),(y,t)\}$ in G'_∞ for each $(x,y) \in M^*$. Let f be the flow such that $f(e) = 1$ if and only if e belongs to one of these paths. This f is admissible and has value $V_f = |M^*|$. This implies that the maximum flow f^* satisfies $V_{f^*} \geq |M^*|$, and the proof is concluded. □

The next theorem gives a characterization of bipartite graphs that contain a perfect matching. If $G = (V,E)$ with $V = X \cup Y$ is bipartite, then it can contain a perfect matching only if $|X| = |Y|$. For all $W \subseteq V$, let $N(W) \equiv \bigcup_{w \in W} N(w)$.

Theorem 4 (Hall, 1935) *A bipartite graph G such that $|X| = |Y|$ contains a perfect matching of if and only if $|N(W)| \geq |W|$ for all $W \subseteq X$.*

Hall's theorem is also known as the marriage theorem, where the vertices are viewed as individuals in two disjoint groups and edges represent a potential relationship between two individuals.

PROOF. Assume G has a perfect matching and fix any $W \subseteq X$. Then each $w \in W$ is uniquely matched to a $y \in N(w)$. But this is impossible unless $|N(W)| \geq |W|$.

Vice versa, assume $|N(W)| \geq |W|$ for all $W \subseteq X$ holds. We consider two cases.

Case 1. There exists a flow f in G'_∞ with value $|X|$. Then Fact 3 implies that there exists a matching of size $|X|$ which must then be perfect.

Case 2. Any flow f in G'_∞ has value $V_f < |X|$. Then the max-flow min-cut theorem implies that the minimum cut S, T has cost $k = c(S,T) < |X|$. Now, the minimum cut S, T must be such that

$$\Gamma(S,T) \equiv \{(y,t) \,:\, y \in S\} \cup \{(x,s) \,:\, x \in T\}$$

because all edges between X and Y have infinite capacity. So we can write

$$c(S,T) = \big|(X \cap T) \cup (Y \cap S)\big| = \big|X \cap T\big| + \big|Y \cap S\big| = k < |X| = \big|X \cap T\big| + \big|X \cap S\big|$$

Therefore, $\big|Y \cap S\big| < \big|X \cap S\big|$. Now set $W \equiv X \cap S$ and note that $N(W) \subseteq Y \cap S$ must hold. Otherwise, there exists $y \in T$ such that $(x,y) \in E$ for some $x \in S$. But this implies $c(S,T) = \infty$ and we have a contradiction. Hence, $|N(W)| \leq |Y \cap S| < |W|$, which contradicts our initial assumption. $\square$