1 October 27: Bipartite matching Hopcroft-Karp König Hall

Throughout, $G = (X, Y, E)$ is a bipartite graph with $n = |X| + |Y|$ vertices and $m = |E|$ edges.

Notation. A *matching* $M \subseteq E$ is a set of edges no two of which share an endpoint. A vertex is *exposed* (or free) if no edge of M is incident to it. A path is *alternating* if its edges alternate between $E \setminus M$ and M, and *augmenting* if, in addition, both its endpoints are exposed. The *length* $|P|$ of a path is its number of edges; an augmenting path has odd length, and in a bipartite graph it goes from an exposed vertex of X to an exposed vertex of Y. For sets of edges, $A \oplus B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference. If P is augmenting for M, then $M \oplus P$ is a matching with $|M \oplus P| = |M| + 1$ (Figure **??**). Recall Berge's theorem: *M is maximum if and only if there is no augmenting path for M.*



Figure 1: Thick edges are matched. Flipping an augmenting path increases the matching by one.

fig:augment

We will use the following fact repeatedly.

Let M and N be matchings. Then $M \oplus N$ is a vertex-disjoint union of paths and even cycles whose edges alternate between M and N. If $|N| = |M| + s$, then $M \oplus N$ contains at least s vertex-disjoint augmenting paths for M.

Proof. Every vertex has at most one incident edge from each matching, hence degree at most 2 in $M \oplus N$; consecutive edges come from different matchings, so cycles are even. A cycle, or a path of even length, has as many M-edges as N-edges; any other path has one more of one kind. Since N has s more edges than M in total, at least s components are paths with one more N-edge than M-edges. Such a path starts and ends with N-edges, and its endpoints are exposed in M (otherwise the path would continue), so it is augmenting for M. □

1.1 The Hopcroft–Karp algorithm

The naive algorithm finds one augmenting path at a time, in $O(m)$ each, for $O(mn)$ total. Hopcroft and Karp find many shortest augmenting paths at once.

Algorithm. Start with $M \leftarrow \emptyset$. A *phase* does the following.

1. Let k be the length of a shortest augmenting path for M (stop if there is none).
2. Find a *maximal* set $\mathcal{P} = \{P_1, \ldots, P_r\}$ of vertex-disjoint augmenting paths of length k. Greedily: while possible, find an augmenting path of length k that avoids the vertices of the paths already in $\mathcal{P}$, and add it to $\mathcal{P}$.
3. Augment along all of them: $M \leftarrow M \oplus P_1 \oplus \cdots \oplus P_r$.

Since the paths are vertex-disjoint, each P_i is still augmenting after the others have been flipped, and $|M|$ grows by $r \geq 1$.

[Hopcroft–Karp] The algorithm computes a maximum matching in time $O(m\sqrt{n})$. This follows from Lemma **??** ($O(\sqrt{n})$ phases) and Lemma **??** ($O(m)$ time per phase).

Let M be a matching, P a shortest augmenting path for M, $M' = M \oplus P$, and P' an augmenting path for M'. Then $|P'| \geq |P| + 2|P \cap P'|$, where $|P \cap P'|$

- *P' shares no vertex with any P_i.* Then P' uses no edge of $M' \oplus M$, and its endpoints are exposed in M too, so P' is also an augmenting path for M, of length $\geq k$. It cannot have length exactly k: it would then be an augmenting path of length k avoiding all the P_i, contradicting the maximality of $\mathcal{P}$. So $|P'| > k$.

- *P' shares a vertex v with some P_i.* The paths $P_1, \ldots, P_r$ are vertex-disjoint, so they can be flipped in any order; renumber them so that this path is P_r. Consider the intermediate matchings

  $$M_0 = M, \qquad M_i = M_{i-1} \oplus P_i \quad (1 \leq i \leq r), \qquad \text{so that } M_r = M'.$$

  Each P_i is vertex-disjoint from $P_1, \ldots, P_{i-1}$, so it is still an augmenting path for M_{i-1}, of length k. By Lemma **??** applied at each step, the length of a shortest augmenting path never decreases along $M_0, M_1, \ldots, M_{r-1}$, so it is at least k for M_{r-1}. Hence P_r is a *shortest* augmenting path for M_{r-1}. Since $M' = M_{r-1} \oplus P_r$ and P' is augmenting for M', Lemma **??** gives

  $$|P'| \;\geq\; |P_r| + 2|P_r \cap P'| \;=\; k + 2|P_r \cap P'|.$$

  It remains to see that P' and P_r share an *edge*, not just the vertex v. Flipping P_r matches each of its vertices to a neighbour along P_r, so v is matched in M' by an edge of P_r. The endpoints of P' are exposed in M', so v is an interior vertex of P'. An alternating path through a matched vertex uses that vertex's matching edge, so P' contains this edge of P_r. Hence $|P_r \cap P'| \geq 1$ and $|P'| \geq k + 2$.

This proves the claim. (Note that Lemma **??** is applied to P_r with the matching M_{r-1} just before P_r is flipped, not with M: P' is augmenting for M', which equals $M \oplus P_r$ only when $r = 1$.)

Counting phases. Augmenting paths have odd length, so after $\lceil \sqrt{n} \rceil$ phases the current matching M has no augmenting path of length less than $\sqrt{n}$. Let M^* be a maximum matching. By Fact **??**, $M \oplus M^*$ contains $|M^*| - |M|$ vertex-disjoint augmenting paths for M, each with at least $\sqrt{n}$ vertices. They are disjoint, so $|M^*| - |M| \leq n/\sqrt{n} = \sqrt{n}$. Each further phase increases $|M|$ by at least one, so at most $\sqrt{n}$ more phases follow. □

Each phase can be implemented in time $O(m)$ (assuming no isolated vertices).

Proof. Orient the edges: non-matching edges from X to Y, matching edges from Y to X. Directed paths from an exposed vertex of X are then exactly the alternating paths.

Layered graph. Run a breadth-first search from all exposed vertices of X (level 0). Stop at the end of the first level k that contains an exposed vertex of Y. Keep only the edges going from a level ℓ to level $\ell + 1$, and give each vertex a counter equal to its number of remaining predecessors. Every augmenting path of length k is a path from level 0 to an exposed vertex of Y at level k in this layered graph (Figure **??**). This takes $O(m)$ time.

Extracting disjoint paths. Repeat:

1. Take an exposed vertex y of Y at level k with at least one predecessor.

2. Trace back from y, following any remaining predecessor, down to level 0. This gives an augmenting path P of length k; add it to $\mathcal{P}$.

3. Put all vertices of P in a deletion queue Q. While $Q \neq \emptyset$: remove a vertex u from Q, delete u and its outgoing edges, and decrement the counter of each successor v. If the counter of v reaches 0, put v in Q: no path from level 0 reaches it any more.

Every remaining vertex at level $\ell \geq 1$ always has a remaining predecessor, so the trace-back never gets stuck. Each vertex is deleted and each edge examined $O(1)$ times, so this part also takes $O(m)$. The paths found are vertex-disjoint, and the set is maximal: when no exposed vertex of level k has a predecessor left, every length-k augmenting path meets a deleted vertex. □

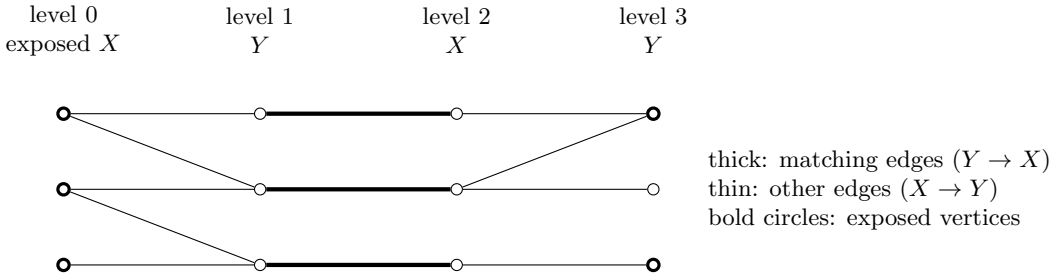


Figure 2: The layered graph of Lemma **??** with $k = 3$. The two augmenting paths $x_1y_1z_1w_1$ and $x_3y_3z_3w_3$ form a maximal vertex-disjoint set.
fig:layers

1.2 König's theorem

A *vertex cover* is a set C of vertices such that every edge has at least one endpoint in C.

[König] In a bipartite graph, the size of a maximum matching equals the size of a minimum vertex cover.

Proof. Let M^* be a maximum matching.

($\geq$) Every vertex cover contains an endpoint of each edge of M^*, and these endpoints are distinct. So every vertex cover has at least $|M^*|$ vertices.

($\leq$) We build a vertex cover of size $|M^*|$. Let L be the set of vertices reachable from an exposed vertex of X by an alternating path (including the exposed vertices themselves). Let

$$C = (X \setminus L) \cup (Y \cap L).$$

C is a vertex cover. Suppose an edge xy is not covered, i.e. $x \in X \cap L$ and $y \in Y \setminus L$. If $xy \notin M^*$, then an alternating path reaching x ends at x either at an exposed vertex or with a matching edge, so it can be extended by xy and $y \in L$. If $xy \in M^*$, then x is not exposed, so the alternating path reaching x arrives through its matching edge, i.e. through y, and again $y \in L$. Either way we get a contradiction: there is no edge between $X \cap L$ and $Y \setminus L$ (Figure **??**).

$|C| \leq |M^*|$. Every $y \in Y \cap L$ is matched, since otherwise the alternating path reaching y would be augmenting, contradicting the maximality of M^* (Berge). Every $x \in X \setminus L$ is matched, since exposed vertices of X are in L. Moreover, no edge of M^* has both endpoints in C. If $y \in Y \cap L$ is matched to x, then $x \in L$, by extending the path with the matching edge yx. So each vertex of C is matched, and no two of them to each other; hence $|C| \leq |M^*|$. □

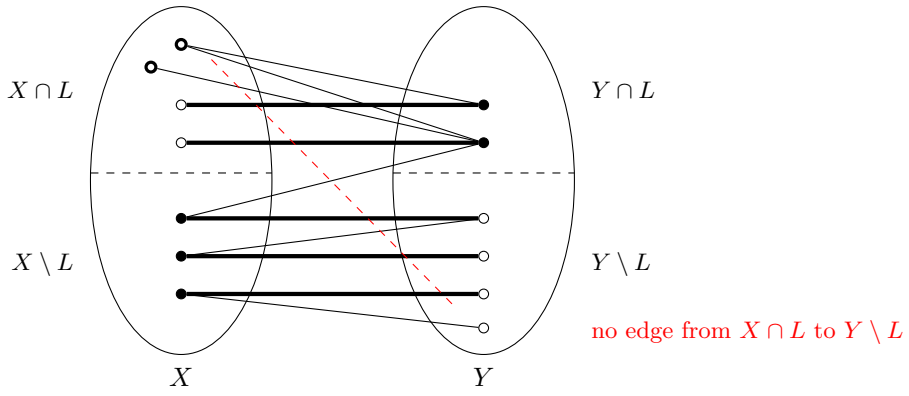


Figure 3: Proof of König's theorem. Thick edges belong to M^*; bold circles are exposed vertices of X. The black vertices, $(X \setminus L) \cup (Y \cap L)$, form a vertex cover with one vertex per edge of M^*.

fig:konig

König's theorem is false for non-bipartite graphs. In a triangle, a maximum matching has one edge, but every vertex cover needs two vertices. The proof is constructive: given M^*, the cover is found by a single alternating search, in time $O(m)$.

1.3 Hall's theorem

For $S \subseteq X$, let $N(S) \subseteq Y$ denote the set of neighbours of S.

[Hall] G has a matching of size $|X|$ if and only if $|N(S)| \geq |S|$ for every $S \subseteq X$.

Call the condition $|N(S)| \geq |S|$ for all $S \subseteq X$ *Hall's condition*. A nonempty $S \subsetneq X$ with $|N(S)| = |S|$ is a *tight set*.

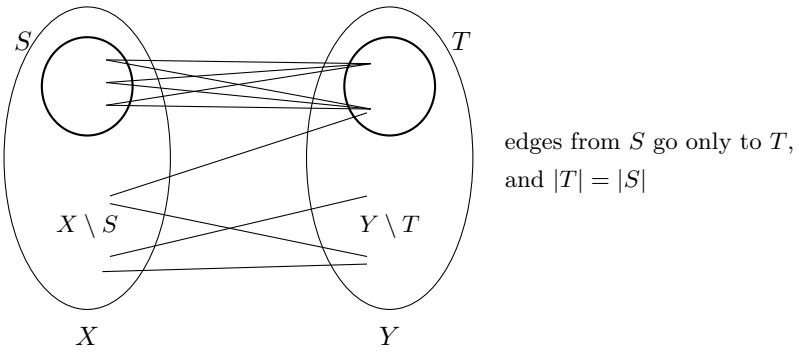


Figure 4: Case 2 of the proof of Hall's theorem: a tight set S and its neighbourhood T split the problem into two smaller ones.

fig:hall

Proof. *($\Rightarrow$)* Let M be a matching of size $|X|$ and $S \subseteq X$. The $|S|$ vertices matched by M to the vertices of S are distinct and belong to $N(S)$, so $|N(S)| \geq |S|$.

($\Leftarrow$) By induction on $|X|$. If $|X| = 0$ there is nothing to prove. Assume the statement holds for all bipartite graphs with fewer than $|X|$ left vertices, and that G satisfies Hall's condition.

Case 1: there is no tight set. Then every nonempty $S \subsetneq X$ satisfies $|N(S)| \geq |S| + 1$: the condition holds with a margin. Pick any $x \in X$; it has a neighbour y, since $|N(\{x\})| \geq 1$. Let G' be the graph obtained by deleting x and y. For nonempty $S \subseteq X \setminus \{x\}$ (a proper subset of X), deleting y loses at most one neighbour, so $|N_{G'}(S)| \geq |S| + 1 - 1 = |S|$. By induction, G' has a matching of $X \setminus \{x\}$; adding the edge xy matches all of X.

Case 2: there is a tight set S. Let $T = N(S)$, so $|T| = |S|$ (Figure **??**). We show that Hall's condition holds in the two subgraphs $G[S \cup T]$ and $G[(X \setminus S) \cup (Y \setminus T)]$. Both have fewer than $|X|$ left vertices, so by induction they have matchings of S into T and of $X \setminus S$ into $Y \setminus T$. Since T and $Y \setminus T$ are disjoint, the union is a matching of X.

- In $G[S \cup T]$: for $S_1 \subseteq S$, all neighbours of S_1 lie in $T = N(S)$, so its neighbourhood is the same as in G, and $|N(S_1)| \geq |S_1|$.
- In $G[(X \setminus S) \cup (Y \setminus T)]$: let $S_1 \subseteq X \setminus S$. By Hall's condition in G, $|N(S_1 \cup S)| \geq |S_1| + |S|$. Now $N(S_1 \cup S) = T \cup \big(N(S_1) \setminus T\big)$ and $|T| = |S|$, so $|N(S_1) \setminus T| \geq |S_1|$. Thus S_1 has at least $|S_1|$ neighbours outside T. □

Hall's theorem also follows from König's theorem. If there is no matching of size $|X|$, take a vertex cover C with $|C| < |X|$, and let $S = X \setminus C$. Every neighbour of S lies in $C \cap Y$, so $|N(S)| \leq |C \cap Y| = |C| - |C \cap X| < |X| - |C \cap X| = |S|$, which violates Hall's condition.