Week 13-14: Matchings in Bipartite Graphs

May 3, 2018

Problem 1. Consider an m-by-n chessboard in which certain squares are forbidden. What is the largest number of non-attacking rooks that can be placed on the board?

Problem 2. Consider again an m-by-n chessboard where certain squares are forbidden. What is the largest number of dominoes that can be placed on the board so that each domino covers two allowed squares and no two dominoes overlap?

Problem 3. A company has n jobs available, with each job demanding certain qualifications. There are m people who apply for the n jobs. What is the largest number of jobs that can be filled from the available m applicants if a job can be filled only by a person who meets its qualifications?

1 Matchings

Definition 1.1. A (simple) graph is a system of ordered pair $G = (V, E)$ consisting of a nonempty vertex set V of **vertices** and an edge set E of **edges**, such that each edge $e \in E$ is assigned an unordered pair $\{u, v\}$ of two vertices (called **endpoints** of e), and no two edges are assigned to the same unordered pair of vertices; we usually write $e = uv$ and say that the edge e is **incident with** the vertices u, v.

A graph $G = (V, E)$ is said to be **bipartite** if the vertex set V can be partitioned into two disjoint parts V_1, V_2 such that each edge $e \in E$ has its two endpoints $u \in V_1$ and $v \in V_2$.

Definition 1.2. A subset M of edges of a graph $G = (V, E)$ is called a **matching** if no two edges of M incident with a common vertex; the two endpoints of an edge in M are said to be **matched under** M. A matching M is said to be **maximum** if there is no matching M' such that $|M'| > |M|$. The **matching number** of G is the cardinality of a maximum matching of G, i.e.,

$$m(G) = \max\{|M| : M \text{ is a matching of } G\}. \tag{1}$$

A matching M is said to be **perfect** if every vertex of G is matched under M.

Example 1.1. Let $X = \{x_1, x_2, x_3, x_4\}$ and $Y = \{y_1, y_2, y_3, y_4, y_5\}$. Consider the 4-by-5 board, whose rows are indexed by elements of X and whose columns are indexed by elements of Y, having forbidden positions shown below

	y_1	y_2	y_3	y_4	y_5
x_1		$\times$			
x_2				$\times$	$\times$
x_3			$\times$		$\times$
x_4	$\times$				

We associate with this board a bipartite graph $G = (X \cup Y, E)$ whose vertex set is $X \cup Y$ and the edge set E is given by

$$x_i y_j \in E \;\Leftrightarrow\; (i, j)\text{-sqaure is allowed.}$$

The graph $G = (X \cup Y, E)$ is called a **rook-bipartite graph**. In this example, the edge set E is the complement of the set $\{x_1y_2, x_2y_4, x_2y_5, x_3y_3, x_3y_5, x_4y_1\}$ in $\{x_iy_j : 1 \le i \le 4, 1 \le j \le 5\}$.

Example 1.2. Consider a 4-by-5 board whose squares are alternately colored black and white, with some forbidden squares. For identification we label the non-forbidden black squares $b_1, b_2, \ldots, b_6$ and non-forbidden white squares $w_1, w_2, \ldots, w_7$; see below.

w_1	$\times$	w_2	b_1	w_3
b_2	w_4	$\times$	$\times$	$\times$
$\times$	b_3	w_5	b_4	$\times$
$\times$	w_6	b_5	w_7	b_6

We define a bipartite graph $G = (X \cup Y, E)$, where

$$X = \{b_1, b_2, \ldots, b_6\}, \quad Y = \{w_1, w_2, \ldots, w_7\},$$

and $b_i w_j \in E$ iff the squares b_i and w_j have a common side.

Example 1.3. Four people x_1, x_2, x_3, x_4 apply for five jobs y_1, y_2, y_3, y_4, y_5. Assume that (1) the person x_1 is qualified for the jobs y_1, y_2, y_4, y_5; (2) the person x_2 is qualified for the jobs y_2, y_3, y_4; (3) the person x_3 is qualified for the jobs y_2, y_5; and (4) the person x_4 is qualified for the jobs y_1, y_2, y_4, y_5. We can construct a bipartite graph $G = (X \cup Y, E)$, where $X = \{x_1, x_2, x_3, x_4\}$, $Y = \{y_1, y_2, y_3, y_4, y_5\}$, and

$$x_i y_j \in E \iff \text{the person } x_i \text{ is qualified for the job } y_j.$$

2 Covering

Definition 2.1. A vertex subset C of a graph $G = (V, E)$ is called a **covering** of G if every edge of G has an endpoint in C. A covering C is said to be **minimum** if there is no covering C' such that $|C'| < |C|$. The **covering number** of G is the cardinality of a minimum covering of G, i.e.,

$$c(G) = \min\big\{|C| : C \text{ is a covering of } G\big\}. \tag{2}$$

Let M be a matching and C be a covering of a graph G. Since each edge of M is covered by a vertex in C and distinct edges of M must be covered by distinct vertices of C, so $|M| \leq |C|$. We thus have

$$m(G) \leq c(G).$$

It may be speculated that $m(G) = c(G)$. Unfortunately, this is not true in general. However, whenever G is bipartite, the equality $m(G) = c(G)$ holds.

Lemma 2.2. *Let M be a matching and C be a covering of a graph G. If $|M| = |C|$, then M is a maximum matching and C is a minimum covering.*

Proof. Trivial by contrapositive argument. □

Theorem 2.3 (König, 1931)**.** *If G is a bipartite graph, then $m(G) = c(G)$. In words it states that the number of edges in a maximum matching is equal to the number of vertices in a minimum covering.*

Proof. It is known that $m(G) \leq c(G)$. We only need to show that $m(G) \geq c(G)$. Let M be a matching and C a covering of G.

Let G be a bipartite graph with the bipartition (X, Y), and let M^* be a maximum matching of G. Denote by U the set of all □

3 Labeling Algorithm

Definition 3.1. Let M be a matching of a graph G. A vertex v of G is said to be **saturated** by M (or M**-saturated**) if v is an endpoint of an edge of M; otherwise it is said to be **unsaturated** by M (or M**-unsaturated**).

A path P (self-intersection is allowed) of G is said to be M**-alternating** if the edges of P alternate between M and its complement $\bar{M} = E \smallsetminus M$. An M-alternating path is further said to be M**-augmenting** if its initial and terminal vertices are not matched under M.

For each M-augmenting path P of a graph G, since the initial and terminal vertices are not matched by M, the first and the last edges can not be edges of M. So the length of P must be odd. The 1st, 3rd, 5th, ..., edges of P are contained in the complement $\bar{M}$, i.e., the odd edges of P belong to $\bar{M}$. The 2nd, 4th, 6th, ..., edges of P are contained in M, i.e., the even edges of P belong to M. The edges with either the initial vertex or the terminal vertex as an endpoint are edges of $\bar{M}$.

Proposition 3.2. *Let M be a matching of a graph G and P a path in G.*

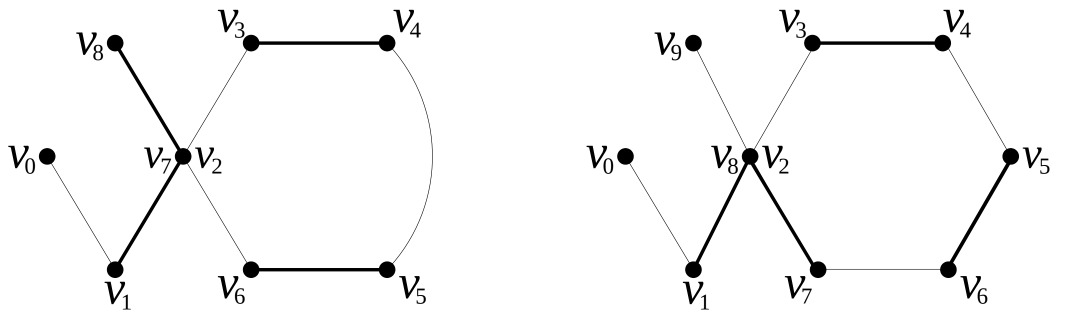
(a) *If P is M-alternating, then P has no self-intersect vertices.*

(b) *If P is an M-augmenting path, then the symmetric difference*

$$M' = M \Delta P := (M \smallsetminus P) \cup (P \smallsetminus M)$$

is a matching of G and $|M'| = |M| + 1$.

Proof. (a) Suppose that the path $P = v_0 v_1 \ldots v_{2m+1}$ has self-intersection, i.e., two of the vertices $v_0, v_1, \ldots, v_{2m+1}$ are the same, say, $v_i = v_j$ with $i < j$. There are two possibilities: $j-i$ is odd and $j-i$ is even. In the former case, we see that either the edges $v_{i-1}v_i, v_jv_{j+1}$ belong to M or the edges $v_iv_{i+1}, v_{j-1}v_j$ belong to M. This is a contradiction since two edges of M share the common vertex $v_i(v_j)$. In the latter case, we see that either the edges $v_{i-1}v_i, v_{j-1}v_j$ belong to M or the edges v_iv_{i+1}, v_jv_{j+1} belong to M, so two edges of M share the common vertex $v_i(v_j)$.



(b) Since M is a matching, no two edges of $M \smallsetminus P$ share a common vertex. Since P has no self-intersection, no two edges of $P \smallsetminus M$ share a common vertex.

Note that the vertices of $M \cap P$ are internal vertices of P, and neither the initial vertex nor the terminal vertex of P is an endpoint of M. We see that the endpoints of $M \smallsetminus P$ are disjoint from P, of course, disjoint from $P \smallsetminus M$. Thus the symmetric difference $M' = M \Delta P$ is a matching. Clearly, $|M'| = |M| + 1$. $\square$

Theorem 3.3 (Berge)**.** *Let M be a matching of a graph $G = (V, E)$. If there is no M-augmenting path, then M is a maximum matching of G.*

Proof. Suppose that M is not a maximum matching. Then there exists a matching M' such that $|M'| > |M|$. Consider the graph $G^* = (V, E^*)$, where

$$E^* = (M - M') \cup (M' - M).$$

Since $|M'| > |M|$, we have

$$|M' - M| > |M - M'|.$$

The graph G^* has the property that each vertex is incident with at most one edge in $M - M'$ and at most one edge in $M' - M$, i.e., at most two edges of E^*. This means that the degree of every vertex of G^* is at most two. Thus each component of G^* is either a simple path (without self-intersection) or a cycle. The paths and cycles must be M-alternating and can be classified into four types:

Type 1. A simple path whose first and last edges are in $M' - M$.

Type 2. A simple path whose first and last edges are in $M - M'$.

Type 3. A simple path whose first edge is in $M - M'$ and whose last edges is in $M' - M$, or whose first edge is in $M' - M$ and whose last edge is in $M - M'$.

Type 4. A cycle.

Note that a Type 1 path has more edges in M' than the edges in M. A Type 2 path has more edges in M than the edges in M'. A Type 3 path has equal number of edges in both M and M'. And a Type 4 cycle has the same number of edges in both M and M'. Since $|M' - M| > |M - M'|$, there exists at least one path $P = v_0 v_1 \ldots v_{2k+1}$ of Type 1, whose first and last edges are in $M' - M$. Since P is a component of G^*, the initial and terminal vertices v_0, v_{2k+1} are not incident with edges in $M - M'$. We claim that both v_0 and v_{2k+1} are *not* incident with edges of M.

Suppose that v_0 (v_{2k+1}) is incident with an edge e in M. The edge e cannot be in M' (since the vertex is already incident with an edge in $M' - M$). So e belongs to $M - M'$, i.e., v_0 (v_{2k+1}) is incident with at least two edges in G^*, which is a contradictory to the fact that v_0 (v_{2k+1}) is an initial (terminal) vertex of the path P, and that P is a component of G^*. Now we see that P is an M-augmenting path, since its initial and terminal vertices are not incident with edges of M. The symmetric difference $M \Delta P$ gives a larger matching of G, this is contradictory to that M is a maximum matching of G. □

Algorithm 3.4 (Labeling Algorithm)**. Input:** bipartite graph $G = (X \cup Y, E)$ with a bipartition $\{X, Y\}$, $|X| \leq |Y|$, and a matching M of G. **Output:** Showing that M is a maximum matching or constructing an M-augmenting path.

STEP 0: All vertices of G are **unscanned**. Begin by labeling with $(*)$ all vertices of X that are M-unsaturated. In case there is no M-unsaturated vertex in X, STOP. The matching M is already a maximum matching, and X is a minimum covering.

If there exist vertices of X that are unsaturated by M, then all such vertices are labeled with $(*)$, go to STEP 1.

STEP 1: While there exists a labeled, but unscanned, vertex x in X, we label with (x) for all vertices of Y that are adjacent to x by edges not in M but not labeled previously. The vertex x is called a **scanned vertex**. In case that no new label can be assigned to a vertex of Y, i.e., x is not adjacent with any unlabeled vertex of Y by edges of $\bar{M}$, STOP. The matching M is a maximum matching, and $S = X^{\mathrm{un}} \cup Y^{\mathrm{lab}}$ is a minimum covering, where X^{un} is the set of unlabeled vertices of X and Y^{lab} is the set of labeled vertices of Y.

Otherwise, go to STEP 2. (There are no labeled but unscanned vertices in X, i.e., the set of both labeled and unscanned vertices of X is empty, equivalently, all labeled vertices of X are scanned vertices.)

STEP 2: (Notice that Y certainly has some labeled vertices, some are M-saturated, some are M-unsaturated.) If all labeled but unscanned vertices of Y are M-saturated, select one of such vertices, say y, and label with (y) for all vertices of X that are adjacent to y by edges of M but not labeled previously. The vertex y is called a **scanned vertex**. Then return to STEP 1.

If there exists a labeled but unscanned vertex y in Y that is M-unsaturated, STOP. An M-augmenting path P can be constructed as follows: Chase back from y through labels in reversing order to construct $P = yx_1y_1x_2y_2\ldots x_ky_kx_{k+1}$, where

- (x_1) is the label of the vertex y, and $(y, x_1) \in \bar{M}$;
- (y_i) are the labels of x_i, and $(x_i, y_i) \in M$, $1 \leq i \leq k$;
- (x_{i+1}) are the labels of y_i, and $(y_i, x_{i+1}) \in \bar{M}$, $1 \leq i \leq k$.
- The label of x_{k+1} is $(*)$.

Proof. Since each vertex receives at most one label and is scanned at most once, the Labeling Algorithm halts after a finite number of steps. When the algorithm stops in STEP 0, it is trivial that M is a maximum matching.

Case 1. Algorithm 3.4 stops in STEP 1.

We first show that the set $S = X^{\mathrm{un}} \cup Y^{\mathrm{lab}}$ is a covering. Suppose that S is not a covering, that is, there is an edge $e = (x, y)$, where $x \in X$ and $y \in Y$, such that $x \notin X^{\mathrm{un}}$ and $y \notin Y^{\mathrm{lab}}$, i.e., x is labeled and y is unlabeled. The edge e is either in M or not in M. If e is not in M, then by STEP 1 the vertex y receives the label (x); this is a contradiction. If e is in M, then x is M-saturated. Thus the label of x can not be $(*)$ by STEP 0; the label of x must be (y') by STEP 2 from a vertex y' of Y, where y' has been labeled and $e' = (x, y')$ is an edge of M. Since M is a matching, we see that $y = y'$. This is contradictory to that y is unlabeled.

Next we show that $|M| = |S|$. This equality implies that M is a maximum matching and S is a minimum covering. To see that $|M| = |S|$, we establish a bijection from S to M.

For each $y \in Y^{\mathrm{lab}}$, the vertex y is labeled, we claim that y is M-saturated. In fact, if y is scanned, then y is already M-saturated by definition of scanned vertex. If y is unscanned, then y must be M-saturated; otherwise, the M-unsaturation of y implies that the Algorithm 3.4 stops in STEP 2, which is a contradictory to that the Algorithm stops in STEP 1.

Let y correspond to the unique edge (x', y) of M incident with y. The vertex x' receives the label (y) in STEP 2. So $x' \notin X^{\mathrm{un}}$. Hence each vertex of Y^{lab} is incident with one edge of M whose other endpoint belongs to $X \smallsetminus X^{\mathrm{un}}$.

For each vertex $x \in X^{\mathrm{un}}$, since x is not labeled, the vertex x must be M-saturated, otherwise, x receives the label $(*)$. Let x correspond to the unique edge (x, y') of M incident with x. This edge (x, y') is a distinct from each edge (x, y) of M incident with a vertex $y \in Y^{\mathrm{lab}}$.

We have established an injection from S to M. Then $|S| \leq |M|$. Since $|M| \leq |S|$, we conclude that $|M| = |S|$.

Case 2. Algorithm 3.4 stops in STEP 2. The vertex y is labeled and is unscanned.

- Let (x_1) be the label of y. Then $(y, x_1) \in \bar{M}$ by STEP 1, the vertex x_1 must be labeled before y, and x_1 is scanned by definition.
- Let (y_1) be the label of x_1. Then $(x_1, y_1) \in M$ by STEP 2, the vertex y_1

must be labeled before x_1, and y_1 is scanned by definition.

- Let (x_2) be the label of y_1. Then $(y_1, x_2) \in \bar{M}$ by STEP 1, the vertex x_2 must be labeled before y_1, and x_2 is scanned by definition.
- Let (y_2) be the label of x_2. Then $(x_2, y_2) \in M$ by STEP 2, the vertex y_2 must be labeled before x_2, and y_2 is scanned by definition.

Continue this procedure, we obtain a sequence of vertices

$$y,\ x_1,\ y_1,\ x_2,\ y_2,\ \ldots,\ x_k,\ y_k,\ x_{k+1},$$

where x_{k+1} has label $(*)$, and the edge $(y_k, x_{k+1}) \in \bar{M}$. Clearly,

$$P = yx_1y_1x_2y_2 \ldots y_kx_{k+1}$$

is an M-alternating path, and the initial and terminal vertices y, x_{k+1} are M-unsaturated. Hence P is an M-augmenting path. Thus

$$M' := (M - P) \cup (P - M)$$

is a new matching and $|M'| = |M| + 1$. $\square$

The Algorithm 3.4 can be easily modified to give an algorithm to find a maximum matching and a minimum covering for a bipartite graph. We give such an algorithm in the following.

Algorithm 3.5 (Matching-Covering Algorithm)**. Input:** a bipartite graph $G = (X \cup Y, E)$ and a matching M consisting of a single edge. **Output:** a maximum matching M.

Step 0. All vertices are **unscanned**. Label with the symbol $(*)$ all vertices of X unsaturated by M. If there is no M-unsaturated vertex in X, STOP. The matching M is a maximum matching, and X is a minimum covering. Otherwise, go to STEP 1.

Step 1. While there exists a labeled, but unscanned, vertex x in X, we label with (x) for all vertices of Y that are adjacent to x by edges not in M but not labeled previously. The vertex x is called a **scanned vertex**. In case that no new label can be assigned to a vertex of Y, i.e., x is not adjacent with any unlabeled vertex of Y by edges of $\bar{M}$, STOP. The matching M is a maximum matching, and $S = X^{\mathrm{un}} \cup Y^{\mathrm{lab}}$ is a minimum covering, where X^{un} is the set of unlabeled vertices of X and Y^{lab} is the set of labeled vertices of Y. Otherwise, go to STEP 2.

Step 2. If all labeled but unscanned vertices of Y are M-saturated, select one of such vertices, say y, and label with (y) for all vertices of X that are adjacent to y by edges of M but not labeled previously. The vertex y is called a **scanned vertex**. Then return to STEP 1.

If there exists a labeled but unscanned vertex y in Y that is M-unsaturated, an M-augmenting path P can be constructed, set $M := (M - P) \cup (P - M)$, return to STEP 0.

4 Systems of Distinct Representatives

Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of subsets of a set X. A family $S = (a_1, a_2, \ldots, a_n)$ of elements of X is called a **system of representatives** for $\mathcal{A}$ if

$$a_1 \in A_1, \quad a_2 \in A_2, \quad \ldots, \quad a_n \in A_n;$$

and is further called a **system of distinct representatives** (SDR) if, in addition to that S is a system of representatives, all the elements $a_1, a_2, \ldots, a_n$ are distinct.

Example 4.1. Let (A_1, A_2, A_3, A_4) be a family of subsets of the set

$$Y = \{a, b, c, d, e\}$$

defined by

$$A_1 = \{a, b, c\}, \quad A_2 = \{b, d\}, \quad A_3 = \{a, b, d, e\}, \quad A_4 = \{a, d, e\}.$$

Then (c, d, d, e) is a system of representatives, and (b, d, e, a) is an SDR.

Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of sets. If $\mathcal{A}$ has an SDR

$$S = (a_1, a_2, \ldots, a_n),$$

then for any selection $\{i_1, i_2, \ldots, i_k\}$ of indices, the union

$$A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}$$

contains the subset $\{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\}$. Thus we have

$$\text{(MC)}: \quad |A_{i_1} \cup \cdots \cup A_{i_k}| \geq k, \quad \forall\, i_1 < \cdots < i_k. \tag{3}$$

The condition (3) is known as the **Marriage Condition (MC)** because of the following Marriage Problem. The condition (3) is also known as **Hall's Condition** because of Theorem 4.1.

Example 4.2 (Marriage Problem)**.** There are n women in a society, all men are eager to marry. If there were no restriction on who marries whom, in order to marry off all the women, we need only require that the number of men be at least as large as the number n of women. But we would expect that each woman and each man would insist some compatibility with a spouse, thereby eliminating some of the men as potential spouses for each woman. Thus each woman would arrive at a certain set of compatible men from the set of available men. Let $(A_1, A_2, \ldots, A_n)$ be the family of subsets of the men, where A_i denotes the set of compatible men for the ith woman, $i = 1, \ldots, n$. Then *marrying off all women* corresponds to an SDR $(y_1, \ldots, y_n)$ of $(A_1, \ldots, A_n)$. The marriage is that the ith woman marries the man $y_i \in A_i$, $i = 1, \ldots, n$. No two women marry the same man.

Theorem 4.1 (Hall, 1935)**.** *A family $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ of sets has an* SDR *if and only if the Marriage Condition* (MC) *holds.*

Proof. **Proof by Contradiction.** We have seen that MC is necessary. We only need to prove that MC is sufficient. Let G be the bipartite graph whose vertex set is the union $X \cup Y$, where

$$X = \{x_1, x_2, \ldots, x_n\}, \quad Y = A_1 \cup A_2 \cup \cdots \cup A_n,$$

and whose edge set is

$$E = \{(x_i, y_j) \in X \times Y : y_j \in A_i\}.$$

To show that $\mathcal{A}$ has an SDR, it is equivalent to show that the bipartite graph G has the matching number $m(G) = n$. It then suffices to show that the covering number $c(G) = n$.

Suppose this is not true, i.e., there exits a covering C such that $|C| < n$. Let $C_1 = C \cap X$ and $C_2 = C \cap Y$. Then

$$|C_1| + |C_2| < n.$$

Since $|C_1| < n$ and $|X| = n$, then $X \smallsetminus C_1$ is nonempty. We write

$$X \smallsetminus C_1 = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}.$$

Since C is a covering, there is no edge of G from a vertex in $X \smallsetminus C_1$ to a vertex in $Y \smallsetminus C_2$. This means that all the sets A_{i_1}, A_{i_2}, ..., A_{i_k} must be contained in C_2. Hence $A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_k}$ is a subset of C_2. Thus

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| \leq |C_2|.$$

Since $|C_1| + |C_2| < n$ and $|C_1| = n - k$, we have

$$|C_2| < n - |C_1| = k.$$

Therefore

$$|A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}| < k,$$

which is contradictory to the Marriage Condition.

Proof by Induction. For $n = 1$, it is trivially true.

Assume $n \geq 2$. There are two cases to be considered: the **tight case** and the **case with room to spare**.

The Tight Case. There is a proper subfamily of $\mathcal{A}$ of k sets whose union contains exactly k elements, where $1 \leq k \leq n - 1$. (By MC the union cannot contain fewer than k elements, so we are tight.) Without loss of generality, we may assume that the subfamily is the first k sets $A_1, \ldots, A_k$. Set

$$A := A_1 \cup \cdots \cup A_k.$$

Clearly, $|A| = k$. Since $\mathcal{A}$ satisfies MC, so does the subfamily $(A_1, \ldots, A_k)$. Since $k < n$, it follows by induction that $(A_1, \ldots, A_k)$ has an SDR $(y_1, \ldots, y_k)$. Since $|A| = k$ and $y_1, \ldots, y_k$ are distinct, we have $A = \{y_1, \ldots, y_k\}$. It forces that $A_1 = \{y_1\}, \ldots, A_k = \{y_k\}$.

Now consider the family $\mathcal{A}^* := (A_{k+1} \smallsetminus A, \ldots, \mathcal{A}_n \smallsetminus A)$ of $n-k$ sets. We claim that $\mathcal{A}^*$ satisfies MC. In fact, for each choice of indices $k+1 \leq i_1 < \cdots < i_j \leq n$, consider the subfamily $(A_1, \ldots, A_k, A_{i_1}, \ldots, A_{i_j})$ of $\mathcal{A}$. Then by MC property of $\mathcal{A}$,

$$|A \cup A_{i_1} \cup \cdots \cup A_{i_j}| = |A_1 \cup \cdots \cup A_k \cup A_{i_1} \cup \cdots \cup A_{i_j}| \geq k + j.$$

Since $|A| = k$, it follows that

$$|(A_{i_1} \smallsetminus A) \cup \cdots \cup (A_{i_j} \smallsetminus A)| \geq j.$$

We haver seen that the family $\mathcal{A}^*$ satisfies MC.

Since $n - k \leq n - 1$, by induction $\mathcal{A}^*$ has an SDR $(y^*_{k+1}, \ldots, y^*_n)$. Then $(y_1, \ldots, y_k, y^*_{k+1}, \ldots, y^*_n)$ is an SDR for $\mathcal{A}$.

The Case with Room to Spare. For every k with $1 \leq k \leq n - 1$ and every subfamily of $\mathcal{A}$ with k sets, its union contains at least $k + 1$ elements. (The union contains more elements that needed for MC, so we have room to spare.) Then each set A_i contains at least one element, actually two because of room to spare. Fix an element $y_n \in A_n$, set $A'_i := A_i \smallsetminus \{y_n\}$ for $1 \leq i \leq n - 1$. We claim that the family $\mathcal{A}' := (A'_1, \ldots, A'_{n-1})$ satisfies MC. In fact, for each choice $(i_1, \ldots, i_k)$ of indices with $1 \leq i_1 < \cdots < i_k \leq n - 1$, we have

$$|A'_{i_1} \cup \cdots \cup A'_{i_k}| \geq |A_{i_1} \cup \cdots \cup A_{i_k}| - 1 \geq (k + 1) - 1 = k.$$

This means that $\mathcal{A}'$ satisfies MC. By induction, $\mathcal{A}'$ has an SDR $(y_1, \ldots, y_{n-1})$. Hence $(y_1, \ldots, y_{n-1}, y_n)$ is an SDR for $\mathcal{A}$. $\square$

The Hall theorem can be stated in terms of bipartite graphs.

Theorem 4.2. *Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ be a family of sets. Then the size $m(\mathcal{A})$ of a largest subfamily of $\mathcal{A}$ that has an SDR is given by*

$$m(\mathcal{A}) \;=\; \min\left\{n, n - k + |A_{i_1} \cup \cdots \cup A_{i_k}| : i_1 < \cdots < i_k\right\}.$$

Proof. Let $G = (X \cup Y, E)$ be the bipartite graph associated with the family $\mathcal{A}$, where $X = \{1, 2, \ldots, n\}$, $Y = A_1 \cup \cdots \cup A_n$, and

$$E = \big\{(i, y) \in X \times Y : y \in A_i\big\}.$$

To find the largest size of a subfamily of $\mathcal{A}$ that has an SDR, it is equivalent to finding the matching number of the bipartite graph G; and by Theorem 2.3 it is equivalent to finding the covering number of G. Thus $m(\mathcal{A}) = c(G)$.

Let S be a covering of G. Let $|S \cap X| = n - k$ and let $\{i_1, i_2, \ldots, i_k\}$ be the complement of $S \cap X$ in X. Since the edges (i_j, a), where $a \in A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}$, can not be covered by $S \cap X$, such edges must be covered by $S \cap Y$. It follows that

$$S \cap Y = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}.$$

Thus

$$|S| = |S \cap X| + |S \cap Y| = n - k + |A_{i_1} \cup \cdots \cup A_{i_k}|.$$

Therefore, the covering number $c(G)$ of G is

$$c(G) = \min\big\{n, n - k + |A_{i_1} \cup \cdots \cup A_{i_k}| : i_1 < \cdots < i_k\big\}.$$

□

5 Stable Marriages

In the Marriage Problem of n men and n women, we assume that each woman ranks each man in accordance with her preference for that man as a spouse, no tie is allowed. So each woman has a total ordering about the n men. Similarly, each man has a total ordering about the n women. It is clear that there are $n!$ possible ways of complete marriage. A complete marriage is called **unstable** if there exist two women A, B and two men a, b such that

- A and a get married,
- B and b get married,
- A prefers b rather than a,
- b prefers A rather than B.

A complete marriage is called **stable** if it is not unstable. Does there always exist a stable marriage? If it does, how to find a stable marriage?

Let $G = (X \cup Y, E)$ be a complete bipartite graph, where X is the set of the n women and Y is the set of the n men. We write

$$X = \{x_1, x_2, \ldots, x_n\},$$

$$Y = \{y_1, y_2, \ldots, y_n\},$$

$$E = \big\{(x_i, y_j) : x_i \in X,\ y_j \in Y\big\}.$$

Each woman x_i ranks each man y_j by assigning integer r_{ij} between 1 and n. Then $(r_{i1}, r_{i2}, \ldots, r_{in})$ is a permutation of $\{1, 2, \ldots, n\}$. Similarly, each man y_j ranks each woman x_i by assigning integer s_{ji} so that $(s_{j1}, s_{j2}, \ldots, s_{jn})$ is a permutation of $\{1, 2, \ldots, n\}$. We thus have the **preference ranking matrix**

$$M = \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array} \\ \left[\begin{array}{cccc} (r_{11}, s_{11}) & (r_{12}, s_{21}) & \cdots & (r_{1n}, s_{n1}) \\ (r_{21}, s_{12}) & (r_{22}, s_{22}) & \cdots & (r_{2n}, s_{n2}) \\ \vdots & \vdots & & \vdots \\ (r_{n1}, s_{1n}) & (r_{n2}, s_{2n}) & \cdots & (r_{nn}, s_{nn}) \end{array}\right] \end{array}.$$

Example 5.1. For $n = 2$, let the preference ranking matrix be

$$M = \left[\begin{array}{cc} (1,2) & (2,2) \\ (2,1) & (1,1) \end{array}\right].$$

This means that the woman x_1 prefers the man y_1 than y_2, the woman x_2 prefers the man y_2 than y_1. The two women x_1 and x_2 have opposite taste. However, both y_1 and y_2 prefer x_2 than x_1. The men y_1 and y_2 have the same taste. There are two possible complete marriages:

$$x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow y_2 \quad \text{and} \quad x_1 \leftrightarrow y_2,\ x_2 \leftrightarrow y_1.$$

The first one is stable because the three guys x_1, x_2, y_2 are happy, only y_1 is upset, but he has no choice. However, the second one is unstable because the wife x_2 prefers the man y_2 than her current husband y_1 and the husband y_2 prefers the woman x_2 than his current wife x_1.

Example 5.2. For $n = 3$, let the preference ranking matrix be

$$M = \begin{bmatrix} (1,3) & (3,2) & (2,1) \\ (2,1) & (1,3) & (3,2) \\ (3,2) & (2,1) & (1,3) \end{bmatrix}$$

There are $3! = 6$ possible complete marriages. There are two complete stable marriages:

$$\begin{array}{ll} x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow y_2,\ x_3 \leftrightarrow y_3 & \text{(each woman gets her first choice)} \\ x_1 \leftrightarrow y_3,\ x_2 \leftrightarrow y_1,\ x_3 \leftrightarrow y_2 & \text{(each man gets his first choice)} \end{array}$$

The first marriage is stable even each man gets his last choice. The second marriage is also stable even each woman gets her second choice. In general letting each woman get her first choice does not necessarily result a complete marriage. For instance, if two women have the same first choice. There are four unstable complete marriages:

$$\begin{array}{ll} x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow y_3,\ x_3 \leftrightarrow y_2 & (x_2 \text{ and } y_1 \text{ prefer each other}) \\ x_1 \leftrightarrow y_3,\ x_2 \leftrightarrow y_2,\ x_3 \leftrightarrow y_1 & (x_3 \text{ and } y_2 \text{ prefer each other}) \\ x_1 \leftrightarrow y_2,\ x_2 \leftrightarrow y_1,\ x_3 \leftrightarrow y_3 & (x_1 \text{ and } y_3 \text{ prefer each other}) \\ x_1 \leftrightarrow y_2,\ x_2 \leftrightarrow y_3,\ x_3 \leftrightarrow y_1 & (x_1 \text{ and } y_3 \text{ prefer each other}) \end{array}$$

We now show that a stable marriage always exists and give an algorithm to find a stable marriage. Thus complete chaos can be avoided.

Algorithm 5.1 (Gale-Shapley, 1962)**. Deferred Acceptance Algorithm** for stable marriage of women and men. Begin with all women marked as **rejected**, and no man rejected any woman.

STEP 0. If there is no rejected woman, then STOP; a stable marriage is obtained. Otherwise, go to STEP 1.

STEP 1. Let each woman, who is marked as rejected, choose the man whom she ranks highest among all those men who have not yet rejected her. Then go to STEP 2.

STEP 2. Let each man pick out the woman he ranks highest among all those women who have chosen him and whom he has not yet rejected, defer to her decision, and reject the others. Then return to STEP 0.

Proof. During the execution of the algorithm, the women propose to men in STEP 1, some men and some women become engaged in STEP 2. However, the men engaged are able to break engagements if they receive a better offer in the next run. It is clear that no two men can engage a same woman because of STEP 1, and no two women can be engaged to a same man because of STEP 2. We have the following properties:

- Once a man becomes engaged he remains engaged throughout the execution of the algorithm, but his fiancee may be changed. A change of fiancee is an improvement in the eyes of the man who breaks engagement.

- A woman may be engaged or disengaged several times during the execution of the algorithm. However, each new engagement results in a less desirable fiance.

Since each woman cannot be rejected twice by a same man, the number of times to be rejected for each woman is at most the number of men. This indicates that eventually there is no rejected woman, so the algorithm stops after a finite number of runs of STEP 1 and STEP 2. Whenever there is no rejected woman, then no two women had proposed to a same man in STEP 1 in the previous run. Thus each man was chosen by one woman, so each man was engaged in STEP 2. Now we obtain a complete marriage. Next we show that this marriage is stable.

Given two women x, x' and two men y, y' such that x marries y and x' marries y':

$$x \leftrightarrow y, \quad x' \leftrightarrow y'.$$

If x prefers y' rather than y, then x chose y' at some early stage of the algorithm in STEP 1, and y' may accept or reject x right after in STEP 2, but eventually y' rejected x at some stage later of the algorithm in STEP 2. This means that y' preferred someone else rather than x when x chose him. Since every man gets his fiancee improved during execution of the algorithm, thus y' prefers x' rather than x. Thus, it can not be happen that y' prefers x rather x'. □

Example 5.3. For $n = 4$, apply the Deferred Acceptance Algorithm to find a stable marriage for the preference ranking matrix

$$M = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \left[\begin{array}{cccc} (3,3) & (1,4) & (4,1) & (2,4) \\ (4,2) & (1,2) & (2,3) & (3,2) \\ (4,1) & (2,1) & (3,4) & (1,3) \\ (4,4) & (1,3) & (3,2) & (2,1) \end{array}\right] \end{array}$$

Applying Step 1,

$$x_1 \to y_2,\ x_2 \to y_2,\ x_3 \to y_4,\ x_4 \to y_2.$$

Applying Step 2, we obtain a partial marriage

$$x_1 \leftrightarrow \emptyset,\ x_2 \leftrightarrow y_2,\ x_3 \leftrightarrow y_4,\ x_4 \leftrightarrow \emptyset.$$

Continuing to apply Step 1 and Step 2,

$$\begin{array}{l} x_1 \to y_4,\ x_2 \leftrightarrow y_2,\ x_3 \leftrightarrow y_4,\ x_4 \to y_4; \\ x_1 \leftrightarrow \emptyset,\ x_2 \leftrightarrow y_2,\ x_3 \leftrightarrow \emptyset,\ x_4 \leftrightarrow y_4; \\ x_1 \to y_1,\ x_2 \leftrightarrow y_2,\ x_3 \to y_2,\ x_4 \leftrightarrow y_4; \\ x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow \emptyset,\ x_3 \leftrightarrow y_2,\ x_4 \leftrightarrow y_4; \\ x_1 \leftrightarrow y_1,\ x_2 \to y_3,\ x_3 \leftrightarrow y_2,\ x_4 \leftrightarrow y_4; \\ x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow y_3,\ x_3 \leftrightarrow y_2,\ x_4 \leftrightarrow y_4. \end{array}$$

We thus obtain a stable marriage

$$x_1 \leftrightarrow y_1,\ x_2 \leftrightarrow y_3,\ x_3 \leftrightarrow y_2,\ x_4 \leftrightarrow y_4.$$

Among the stable complete marriages we may wish to compare one stable complete marriage with certain other stable complete marriages. A man y is called **feasible** for a woman x provided there is a stable complete marriage in which y is the spouse of x. A man that is not feasible for a woman is called **infeasible** for the woman. Obviously, feasibility is symmetric in the sense that if y is feasible to x, then x is also feasible to y. A stable complete marriage is called **optimal for a woman** x provided that x ranks her spouse in the marriage higher than or equal to all her feasible spouses. A stable complete marriage is called **women-optimal** provided that it is optimal for each woman. In a similar way we can define a **men-optimal** stable marriage. It is not obvious whether there exist women-optimal and men-optimal stable complete marriages. In fact, it is not even clear whether it results a complete marriage when each woman is independently given her best feasible spouse. The following theorem clarifies these questions.

Theorem 5.2. *The stable complete marriage obtained from the Deferred Acceptance Algorithm, with the woman choosing the men, is women-optimal.*

Proof. The idea is to show that, if a woman x is rejected by a man y, then the man y is infeasible for x. If so, at the end of the Deferred Acceptance Algorithm, each woman obtains as her spouse the man she ranks highest among all the men that are feasible for her, and hence the complete marriage is women-optimal. To this end, we show that at the end of the kth round of Algorithm 5.1, all the men who have rejected a woman x are infeasible for x. We proceed induction on k.

Recall that a man y is infeasible for a woman x if and only if there is no stable complete marriage in which y is the spouse of x; in other words, if there is a complete marriage in which y is the spouse of x, then the complete marriage is unstable.

For $k = 1$, let x be a woman who is rejected by a man y at the end of the first round. Then x ranks y highest among all men, but y prefers another woman x' who also ranks y highest among all men. If there is a complete marriage such that $x \leftrightarrow y$ and $x' \leftrightarrow y'$, then x' prefers y rather than y' and y prefers x' rather than x. This means that the complete marriage is unstable. Now assume it is true for k, that is, all men are infeasible for those women whom they reject at or before the end of the kth round. Consider the case of $k + 1$.

Let y be a man who rejects a woman x at the end of the $(k + 1)$th round. We need to show that y is infeasible for x. Suppose y is feasible for x, this is, there is a stable complete marriage $\mathcal{M}$ in which y is the spouse of x. Since the men who did reject x before the $(k+1)$th round are infeasible for x, the woman x ranks y highest among all the men feasible for x. Since y rejects x in Step 2 of the $(k + 1)$th round, the man y chooses another woman x' rather than x in the same Step 2. Note that x' ranks y highest among all the men who did not reject her before the $(k + 1)$th round, and by the induction hypothesis, the men who did reject x' before the $(k+1)$th round are infeasible for x'. So x' ranks y highest among all the men feasible for x' and some possibly infeasible for x'. Let y' be the spouse of x' in $\mathcal{M}$. Then in the complete marriage $\mathcal{M}$, we have

$$x \leftrightarrow y, \quad x' \leftrightarrow y',$$

but x' prefers y rather than y' (since y' is feasible and y is ranked the highest by x' among all the feasible men and some possibly infeasible men for x'), y prefers x' rather than x (since y rejects x and accept x'). This means that the complete marriage $\mathcal{M}$ is unstable, contrary to that $\mathcal{M}$ is stable. □

Corollary 5.3. *In any women-optimal stable complete marriage, each man is paired with the woman he ranks lowest among all the women feasible for him.*

Proof. Suppose the theorem is not true, that is, there exist a women-optimal stable marriage $\mathcal{M}$ such that $x \leftrightarrow y$, and a stable complete marriage $\mathcal{M}'$ such tat $x' \leftrightarrow y$, but y ranks x' lower than x. Let y' be the spouse of x in the complete marriage $\mathcal{M}'$. The in the stable complete marriage $\mathcal{M}'$, we have

$$x \leftrightarrow y', \quad x' \leftrightarrow y.$$

Since both y, y' are feasible for x (because $\mathcal{M}$ and $\mathcal{M}'$ are stable marriages) and x ranks y highest among all the men feasible for x (because $\mathcal{M}$ is women-optimal), then x prefers y rather than y'. Note that y prefers x rather x' by assumption. This means that the complete marriage $\mathcal{M}'$ is unstable, contrary to that $\mathcal{M}'$ is stable. □

Corollary 5.4. *The women-optimal and men-optimal stable complete marriages are identical if and only if there is exactly one stable complete marriage.*

Proof. If the women-optimal and men-optimal stable complete marriages are identical, then by Corollary 5.3 each woman (man) gets her (his) best and worst partner taken over all stable complete marriages. It follows that there is exactly one stable complete marriage. Conversely, if there is exactly one stable complete marriage, then the women-optimal and men-optimal stable complete marriages must be identically the same. □

Example 5.4. Suppose an even number $2n$ of girls wish to pair up as roommates. Each girl ranks the other girls in the order $1, 2, \ldots, 2n-1$ of preference. A **complete marriage** in this situation is a pairing of the girls into n pairs. A complete marriage is **unstable** provided there exist two girls who are not roommates such that each of them prefers the other person's roommate to her current roommate. A complete marriage is **stable** if it is not unstable. Does there always exists a stable complete marriage?

Consider the case of four girls a, b, c, d; each of the four girls ranks the others as the following

$$\begin{aligned} a:&\ \ b, c, d \\ b:&\ \ c, a, d \\ c:&\ \ a, b, d \\ d:&\ \ \text{any order} \end{aligned}$$

We shall see that there is no stable complete marriage. Suppose a and d are roommates, then b and c would be roommates. However, a prefers c rather d and c prefers a rather b. So the complete marriage

$$a \leftrightarrow d,\ b \leftrightarrow c$$

is unstable. Similar, the complete marriages

$$b \leftrightarrow d,\ a \leftrightarrow c \quad \text{and} \quad c \leftrightarrow d,\ a \leftrightarrow b$$

are also unstable. So there is no stable complete marriage.