MATCHING ESTABLE [GALE Y SHAPLEY, 1962]

JUAN DE MAIRENA (?)

Supongamos que hay N varones $\{a_1, \ldots, a_N\}$ y N mujeres $\{b_1, \ldots, b_N\}$. Cada varón tiene un orden de preferencias de las N mujeres, y a su vez, cada mujer tiene un orden de preferencia de los N varones. El objetivo es formar N parejas (matemáticamente, hallar una biyección φ entre ambos conjuntos) estables.

Decimos que el matching φ es *inestable* si existe un par a_i, b_j tal que $\varphi(a_i) = b_h$ pero a_i prefiere a b_j antes que b_h, y $\varphi(a_k) = b_j$ pero b_j prefiere a a_i antes que a su pareja actual a_k. Si no, será *estable*.

Hay muchas preguntas naturales: ¿existe tal matching estable?, ¿es único?, ¿cómo encontrarlo(s)? Si hay más de uno, ¿cómo compararlos? ¿existe alguno óptimo (y en qué sentido lo es)?

El resultado es extremadamente sencillo:

Teorema 1. *Existe al menos un matching estable, dado por el siguiente algoritmo:*

1. *Cada hombre invita a bailar a su primera opción.*
2. *Cada mujer evalúa las propuestas, se queda con la mejor en su orden de preferencias, y rechaza al resto.*
3. *Cada hombre rechazado invita a bailar a su siguiente opción (aunque en el momento ella esté con otro).*
4. *Se iteran 2 y 3 hasta que ninguna mujer reciba dos o más invitaciones.*

Demostración. Dividamos la demostración en distintas etapas.

Paso 1: el algoritmo termina a lo sumo en N^2 iteraciones, pues en cada una, al menos un hombre es rechazado por una mujer, y como la tacha de su lista, no volverá a invitarla.

Paso 2: una mujer que fue invitada una vez, ya no queda libre aunque puede cambiar de pareja. En particular, en cada ciclo el número de parejas que están bailando permanece igual o aumenta.

Paso 3: si un hombre fue rechazado por $N-1$ mujeres, esto quiere decir que todas ellas están ya bailando con $N-1$ hombres, con lo cual la restante estará sola y lo aceptará.

Paso 4: el número de parejas no se puede estancar en un valor $k < N$, hay $N-k$ hombres que aunque se limiten a invitar a algunas de esas k mujeres, tarde o temprano resultará rechazado y deberán invitar a una de las $N-k$ mujeres restantes.

Paso 5: el matching φ_H dado por este algoritmo es estable: si $\varphi(a_i) = b_h$, pero a_i prefiere a b_j antes que a b_k, esto quiere decir que en algún momento invitó a b_j y ella lo rechazó por otro a_l. Entonces, b_j prefiere a a_l antes que a a_i (si b_j volvió a cambiar de pareja, es por un a_m a quien prefiere antes que a a_l, y por transitividad, a a_i). Luego, φ_H no puede ser inestable. □

Observacion 1. *No debe considerarse que el algoritmo es sexista, a lo sumo, es costumbrista. Por simetría, intercambiando hombres y mujeres, se tiene un nuevo algoritmo, y por lo tanto una nueva asignación que podemos llamar φ_M. No es difícil construir un ejemplo donde φ_H y φ_M son diferentes para el mismo conjunto de hombres y mujeres.*

Considerando que hay más de un matching φ posible, podemos definir para cada hombre a_i y cada mujer b_j los conjuntos

$$A_i = \{b_k : \text{ existe un matching estable } \varphi \text{ tal que } \varphi(a_i) = b_k\},$$
$$B_j = \{a_k : \text{ existe un matching estable } \varphi \text{ tal que } \varphi(a_k) = b_j\},$$

que podríamos llamar las mujeres y hombres *alcanzables.* Podemos definir un orden en los hombres de B_j y las mujeres de A_i de acuerdo a las preferencias de b_j y a_i, y tenemos el siguiente resultado sorprendente:

Teorema 2. *Sea φ_H el matching estable dado por el Teorema de Gale y Shapley. Entonces, para todo a_i, $\varphi_H(a_i)$ es la mejor opción de a_i en A_i, esto es,*

$$\varphi_H(a_i) = \text{máx}\{b_k \in A_i\}.$$

Además, para toda b_j, $\varphi_H^{-1}(b_j)$ es la peor opción de b_j en B_j,

$$\varphi_H^{-1}(b_j) = \text{mín}\{a_k \in B_j\}.$$

La demostración no es complicada, y la veremos más adelante.

El teorema tiene aplicaciones prácticas en muchos modelo de asignación, y ha sido generalizado a diferentes situaciones. En principio, una aplicación importante fue la asignación de residentes en hospitales; cada residente rankea los hospitales de acuerdo a distintos factores (horarios, sueldos, distancias, calidad, especialidad, etc.), y a su vez cada hospital rankea a los alumnos (de acuerdo a un examen, sus notas, experiencia, etc.). Observemos que cada hospital cuenta por tantas personas como plazas ofrece. El mismo ejemplo vale para estudiantes y universidades con cupos, o sistemas de becas.

La existencia de distintas asignaciones estables posibles genera toda clase de problemas combinatorios interesantes. Por ejemplo, dado el conjunto de preferencias posibles, cuántas asignaciones estables tiene una preferencia en promedio, cuál es el número máximo de asignaciones que puede tener una preferencia, etc.

Si se elimina la restricción de que sean hombres y mujeres y queremos formar parejas con un grupo de $2N$ personas, donde cada uno tiene un orden de preferencias de las demás, tenemos el *roommates problem* (por la asignación de dormitorios en colegios). En este caso, tenemos el siguiente resultado, también de Gale y Shapley [**?**]:

Teorema 3. *No siempre existe una asignación estable en el roommates problem.*

Demostración. Sean cuatro personas, α, β, γ y δ, y las preferencias:

- Para α: $\beta > \gamma > \delta$.
- Para β: $\gamma > \alpha > \delta$.
- Para γ: $\alpha > \beta > \delta$.
- Para δ: $\alpha > \beta > \gamma$.

Por simetría, cualquiera que esté asignado a δ, prefiriría estar con el que lo tiene como primera opción. Por ejemplo, si asignamos α con δ, tenemos que α y γ prefirían estar juntos antes que con sus respectivas parejas. □