

# Aplicaciones de Redes Profundas a la Teoría de Emparejamiento

Alvaro J. Riascos Villegas  
Universidad de los Andes y Quantil

14 de mayo de 2026

# Contenido

- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

# Introducción

- Mercados de dos lados: Trabajadores y empresas, médicos y hospitales, estudiantes, escuelas públicas, donantes y receptores de riñones, vendedores y compradores de acciones, plataformas digitales.
- Mercados de un lado: estudiantes y dormitorios universitarios, subastas o licitaciones.
- En muchos casos no existe un mercado organizado y usar transferencias monetarias no es ético.
- Comenzamos estudiando los mercados en los cuales no hay una transferencia de dinero.
- Es un ejemplo exitoso de la aplicación de la teoría de juegos.

# Mercado de matrimonios

- Sea  $M = \{m_1, \dots, m_n\}$  y  $W = \{w_1, \dots, w_p\}$  el conjunto de hombres y mujeres.
- Cada hombre (o mujer) tienen unas preferencias por el otro grupo incluyendo una preferencia por estar solo (i.e., preferencias racionales: completas y transitivas). Denotamos por  $P(m) = w_{i_1}, \dots, m, \dots, w_{i_n}$  las preferencias de  $m$  de mayor a menor.
- Cuando hay indiferencia, por ejemplo entre  $w_{i_2}$  y  $w_{i_3}$  lo denotamos por  $P(m) = w_{i_1}, [w_{i_2}, w_{i_3}], \dots, m, \dots, w_{i_n}$ .
- También usamos  $>_m$  y  $\geq_m$  como preferencias estrictas o débiles.
- El conjunto de preferencias se denota por  $P = \{P(m_1), \dots, P(m_n), P(w_1), \dots, P(w_p)\}$
- Un mercado de matrimonios es  $(M, W, P)$ .
- Una mujer  $w$  es aceptable para  $m$  si  $w \geq_m m$ .

## Definition (Emparejamiento y Racionalidad Individual)

Una función 1-1  $\mu : M \cup W \rightarrow M \cup W$  tal que  $\mu^2 = id$  y si  $\mu(m) \neq m$ , entonces  $\mu(m) \in W$  (similar para las mujeres).

- Decimos que  $\mu(m)$  es la pareja de  $m$ .
- Ejemplo:

$$\mu = \begin{array}{ccccc} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{array}$$

## Definition (Racionalidad Individual)

$\mu$  es individualmente racional si cada agente es aceptable para su pareja (i.e., si ningún agente bloquea de forma unilateral el emparejamiento).

## Definition (Estabilidad)

Un emparejamiento  $\mu$  estable si no es bloqueado por ningún agente o pareja de agentes (i.e.,  $\mu$  es bloqueado por una pareja  $(m, w)$  si ambos prefieren estar emparejados a estar con la pareja que les corresponde en el emparejamiento  $\mu$ :  $m >_w \mu(w)$ ,  $w >_m \mu(m)$ ).

## Example (Estabilidad)

$$P(m_1) = w_2, w_1, w_3$$

$$P(m_2) = w_1, w_3, w_2$$

$$P(m_3) = w_1, w_2, w_3$$

$$P(w_1) = m_1, m_3, m_2$$

$$P(w_2) = m_3, m_1, m_2$$

$$P(w_3) = m_1, m_3, m_2$$

El emparejamiento:

$$\mu = \begin{array}{ccc} w_1 & w_2 & w_3 \\ m_1 & m_2 & m_3 \end{array}$$

no es estable,  $(w_2, m_1)$  lo bloquean, y:

$$\mu = \begin{array}{ccc} w_1 & w_2 & w_3 \\ m_1 & m_3 & m_2 \end{array}$$

es estable. Corresponde al algoritmo de aceptación diferida en el que la mujer propone.

# Contenido

- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

## Variaciones simples del modelo

- El anterior modelo nos permite desarrollar una teoría (i.e., para el mercado de matrimonios) sobre emparejamientos estables.
- Pero antes veamos algunos ejemplos de lo que puede salir mal en situaciones muy similares al modelo de emparejamiento propuesto.

# El problema de los compañeros de habitación - Gale y Shapley

- Mercado de un solo lado, muchos a uno.
- Tenemos  $n$  (par) personas que hay que emparejar para que ocupen  $\frac{n}{2}$  habitaciones.
- Cada persona tiene preferencias sobre los otros  $n - 1$ .
- En este caso un resultado (i.e., emparejamiento) es estable si no existen dos personas que no estén emparejadas que ambas prefieran estar juntas a estar con las pareja que les toco.

# El problema de los compañeros de habitación

- Suponga que  $n = 4$  y las preferencias son de la forma:

$$p(a) = b, c, d \quad (1)$$

$$p(b) = c, a, d \quad (2)$$

$$p(c) = a, b, d \quad (3)$$

$$p(d) = \text{arbitrario} \quad (4)$$

- No puede existir un emparejamiento estable porque alguien debe quedar con  $d$  y esa persona es el preferido de alguno de los otros dos. Entonces ese par bloquearían el emparejamiento.

# El problema de conformar una familia - Alkan

- Existen tres tipos de personas: padres, madres e hijos.
- Cada tipo de personas tiene preferencias por el conjunto de parejas de las otras dos.
- Un emparejamiento es la conformación de una familia: padre, madre y un solo hijo.
- Un conjunto  $(m, w, c)$  bloquea un emparejamiento si cada uno de los miembros de este conjunto prefiere conformar esta familia, a la familia con la que están en el emparejados.

# El problema de conformar una familia - Alkan

- Supongamos un caso en el que tenemos tres hombres, tres mujeres y tres hijos.
- Las preferencias son de la forma:

$$P(m_1) = (w_1, c_3), (w_2, c_3), (w_1, c_1), \dots \text{ (arb.)}$$

$$P(m_2) = (w_2, c_3), (w_2, c_2), (w_3, c_3), \dots \text{ (arb.)}$$

$$P(m_3) = (w_3, c_3), \dots \quad P(w_1) = (m_1, c_1), \dots \text{ (arb.)}$$

$$P(w_2) = (m_2, c_3), (m_1, c_3), (m_2, c_2), \dots \text{ (arb.)}$$

$$P(w_3) = (m_2, c_3), (m_3, c_3), \dots \text{ (arb.)}$$

$$P(c_1) = (m_1, w_1), \dots \quad P(c_2) = (m_2, w_2), \dots \text{ (arb.)}$$

$$P(c_3) = (m_1, w_3), (m_2, w_3), (m_1, w_2), (m_3, w_3), \dots \text{ (arb.)}$$

# El problema de conformar una familia - Alkan

- Las preferencias son de la forma:

$$P(m_1) = (w_1, c_3), (w_2, c_3), (w_1, c_1), \dots \text{ (arb.)}$$

$$P(m_2) = (w_2, c_3), (w_2, c_2), (w_3, c_3), \dots \text{ (arb.)}$$

$$P(m_3) = (w_3, c_3), \dots \quad P(w_1) = (m_1, c_1), \dots \text{ (arb.)}$$

$$P(w_2) = (m_2, c_3), (m_1, c_3), (m_2, c_2), \dots \text{ (arb.)}$$

$$P(w_3) = (m_2, c_3), (m_3, c_3), \dots \text{ (arb.)}$$

$$P(c_1) = (m_1, w_1), \dots \quad P(c_2) = (m_2, w_2), \dots \text{ (arb.)}$$

$$P(c_3) = (m_1, w_3), (m_2, w_3), (m_1, w_2), (m_3, w_3), \dots \text{ (arb.)}$$

- Obsérvese que si en un emparejamiento se conforma la familia  $(m_1, w_1, c_3)$  o  $(m_2, w_2, c_3)$  este emparejamiento es bloqueado por  $(m_3, w_3, c_3)$  y  $(m_1, w_2, c_3)$  es bloqueado por  $(m_2, w_3, c_3)$ .

# Muchos a uno: Trabajadores a firmas

- Mercado de dos lados, muchos a uno.
- Un conjunto de trabajadores tiene preferencias sobre las empresas en las que pueden trabajar.
- Las empresas tiene preferencias por subconjuntos de trabajadores que desean contratar.
- Un emparejamiento en este caso es lo obvio.
- Una firma  $F$  y un conjunto de trabajadores  $C$  bloquea un emparejamiento si  $F$  prefiere el conjunto de trabajadores  $C$  al que esta emparejado y cada trabajador de  $C$  que no se encuentre empleado por  $F$  prefiere  $F$  a la firma con la que le toco.

# Muchos a uno: Trabajadores a firmas

- Supongamos que tenemos dos firmas y tres trabajadores.
- Las preferencias son de la forma:

$$P(F_1) = \{w_1, w_3\}, \{w_1, w_2\}, \{w_2, w_3\}, \{w_1\}, \{w_2\}$$

$$P(F_2) = \{w_1, w_3\}, \{w_2, w_3\}, \{w_1, w_2\}, \{w_3\}, \{w_1\}, \{w_2\}$$

$$P(w_1) = F_2, F_1 \quad P(w_2) = F_2, F_1 \quad P(w_3) = F_1, F_2$$

# Muchos a uno: Trabajadores a firmas

$$\mu_4 = \left( \begin{array}{cc} F_1 & F_2 \\ \{w_2\} & \{w_1, w_3\} \end{array} \right), \text{ bloqueado por } (F_1, \{w_2, w_3\})$$

$$\mu_5 = \left( \begin{array}{cc} F_1 & F_2 \\ \{w_1\} & \{w_2, w_3\} \end{array} \right), \text{ bloqueado por } (F_2, \{w_1, w_3\})$$

Todo emparejamiento individualmente racional es bloqueado.

# ¿Cómo identificar un emparejamiento estable?

- Una estrategia sería: como el conjunto de emparejamientos es finito, listarlos todos y uno por uno ir probando (esto pues ser muy ineficiente).
- Ver ejemplo 2.4 de Knuth en Roth y Sotomayor.
- Sin embargo, sí existe un algoritmo que siempre encuentra un emparejamiento estable en el modelo básico de emparejamiento (i.e., mercado de matrimonios). El célebre algoritmo de aceptación diferida de Gale y Shapley.

# Contenido

- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida**
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

# Mercado de Matrimonios y Admisiones a la Universidad

- Vamos a estudiar dos problemas: el mercado de matrimonios y el de admisiones a la universidad.
- El algoritmo de aceptación se debe a Gale y Shapley.
- Comenzamos primero con el mercado de matrimonios que es un caso particular del problema de admisiones.
- Este mercado es uno a uno.

## Theorem (Gale y Shapley)

*Existe un emparejamiento estable para todo mercado de matrimonios.*

# Algoritmo de aceptación diferida

- 1 Cada hombre le propone matrimonio a la mujer que más prefiere entre las aceptables para el (véase las observaciones para el caso en el que hay indiferencia).
- 2 Las mujeres rechazan las ofertas de todos los hombre que no son aceptables o que no son los preferidos entre las ofertas que recibieron: se comprometen temporalmente con la mejor oferta.
- 3 Los hombres que fueron rechazados, en el anterior paso o más adelante en el proceso, hacen ofertas a la siguiente mujer más preferida entre las que no lo han rechazado anteriormente y sus aceptables.
- 4 Se repite hasta que ningun hombre es rechazado.
- 5 El resultado es un emparejamiento estable.

## Demostración.

Si no fuera estable hay una pareja que bloquea. Como en esa pareja el hombre prefiere a esa mujer más que a la que le tocó, le debió ofrecer y fue rechazado por ella. Pero si ella lo rechazó debió terminar en el algoritmo con un hombre mejor (las que reciben ofertas en el algoritmo solo pueden mejorar en las rondas). Una contradicción. □

# Algoritmo de aceptación diferida: Observaciones

- En cualquier paso del anterior algoritmo si hay indiferencia entre alternativas usar cualquier regla predefinida para elegir.
- El algoritmo siempre termina porque ningun hombre le ofrece más de una vez matrimonio a una mujer.
- El resultado final es un emparejamiento: en cada etapa cada hombre esta emparejado máximo una mujer y cada mujer máximo con un hombre.
- Es individualmente racional. En todos los pasos, hombres y mujeres está emparejados con mujeres y hombres aceptables.

# Algoritmo de aceptación diferida: Ejemplo

## Example

Considere un mercado con las siguientes preferencias.

$$P(m_1) = w_1, w_2, w_3, w_4$$

$$P(w_1) = m_2, m_3, m_1, m_4, m_5$$

$$P(m_2) = w_4, w_2, w_3, w_1$$

$$P(w_2) = m_3, m_1, m_2, m_4, m_5$$

$$P(m_3) = w_4, w_3, w_1, w_2$$

$$P(w_3) = m_5, m_4, m_1, m_2, m_3$$

$$P(m_4) = w_1, w_4, w_3, w_2$$

$$P(w_4) = m_1, m_4, m_5, m_2, m_3$$

$$P(m_5) = w_1, w_2, w_4$$

Cuando el hombre propone el emparejamiento estable es:

$$\mu_M = \begin{matrix} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{matrix}$$

Si la mujer propone:

$$\mu_W = \begin{matrix} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{matrix}$$

# Algoritmo de aceptación diferida: Ejemplo

## Definition

Decimos que  $\mu$  estable es óptimo para los hombres si ningún hombre prefiere otro emparejamiento estable.

- Obsérvese que en este ejemplo todos los hombres prefieren débilmente el emparejamiento en el que ellos proponen y de forma análoga las mujeres.
- Este resultado es importante porque no es completamente obvio.
- Consideremos el siguiente ejemplo donde existe mucha competencia de hombres por una mujer y mujeres por un hombre.

## Example

Consideremos un mercado con las siguientes preferencias:

$$P(m_1) = w_1, w_2, w_3 \quad P(w_1) = m_1, m_2, m_3 \quad (5)$$

$$P(m_2) = w_1, w_2, w_3 \quad P(w_2) = m_1, m_3, m_2 \quad (6)$$

$$P(m_3) = w_1, w_3, w_2 \quad P(w_3) = m_1, m_2, m_3 \quad (7)$$

Los dos únicos emparejamiento son aquellos en el que el hombre propone y la mujer propone.

# Algoritmo de aceptación diferida: Optimalidad

- Gale y Shapley demostraron que en todo mercado de emparejamiento con preferencias estrictas existe uno, y solo un emparejamiento óptimo para el hombre (y la mujer), y se obtiene usando el algoritmo de preferencias diferidas.
- La prueba es por inducción en los pasos del algoritmo de aceptación diferida y es muy parecida a la prueba que daremos al final de optimalidad para los estudiantes del algoritmo de aceptación diferida en el caso de admisiones a la universidad.

## Theorem (Gale - Shapley Optimalidad)

*Si las preferencias son estrictas, siempre existe un emparejamiento óptimo para los hombres (i.e., el emparejamiento en el que los hombres proponen).*

# Algoritmo de aceptación diferida: No existencia de optimalidad

## Example

Cuando las preferencias no son estrictas pueden no existir un emparejamiento estable óptimo para ningún lado del mercado.

$$P(m_1) = [w_2, w_3], w_1, P(w_1) = m_1, m_2, m_3 \quad (8)$$

$$P(m_2) = w_2, w_1, P(w_2) = m_1, m_2 \quad (9)$$

$$P(m_3) = w_3, w_1, P(w_3) = m_1, m_3 \quad (10)$$

Los emparejamientos estables son:

$$\mu_1 = \begin{array}{ccc} w_1 & w_2 & w_3 \\ m_2 & m_1 & m_3 \end{array}$$

$$\mu_2 = \begin{array}{ccc} w_1 & w_2 & w_3 \\ m_3 & m_2 & m_1 \end{array}$$

que no son óptimos.

# Optimalidad de Pareto

- Se puede demostrar que la asignación óptima para los hombres es óptima de Pareto en un sentido débil para los hombres (un resultados análogo para las mujeres).
- Obsérvese que el concepto de estabilidad de Pareto no esta restringido a emparejamientos estables.

# Comportamiento Estratégico en el Mercado de Matrimonios

- Consideremos la revelación de preferencias como la estrategia de cada jugador.
- El juego es, dadas las preferencias reveladas, el resultado del juego es el emparejamiento del algoritmo de aceptación diferida de los hombres.
- Vamos a mostrar que puede suceder que una mujer, que asume que todos los demás reportan de forma sincera sus preferencias, prefiera mentir antes de reportar sus preferencias.

# Ejemplo manipulación estratégica

- Considere un mercado con estas preferencias (verdaderas):

$m_1$	$m_2$	$m_3$	$w_1$	$w_2$	$w_3$
$w_1$	$w_3$	$w_2$	$m_3$	$m_1$	$m_3$
$w_2$	$w_1$	$w_1$	$m_2$	$m_3$	$m_2$
$w_3$	$w_2$	$w_3$	$m_1$	$m_2$	$m_1$
$m_1$	$m_2$	$m_3$	$w_1$	$w_2$	$w_3$

- Hombres ofrecen, el resultado final del algoritmo de aceptación diferida es:

$$\mu = \begin{array}{cccccc} m_1 & m_2 & m_3 & w_1 & w_2 & w_3 \\ w_1 & w_3 & w_2 & m_1 & m_3 & m_2 \end{array}$$

# Ejemplo manipulación estratégica

- Desviación unilateral de  $w_1$ :

$m_1$   $m_2$   $m_3$   $w_1$   $w_2$   $w_3$

$w_1$   $w_3$   $w_2$   $m_3$   $m_1$   $m_3$

$w_2$   $w_1$   $w_1$   $m_2$   $m_3$   $m_2$

$w_3$   $w_2$   $w_3$   $w_1$   $m_2$   $m_1$

$m_1$   $m_2$   $m_3$   $m_1$   $w_2$   $w_3$

- Hombres ofrecen, el resultado final del algoritmo de aceptación diferida es:

$$\mu = \begin{array}{cccccc} m_1 & m_2 & m_3 & w_1 & w_2 & w_3 \\ w_2 & w_3 & w_1 & m_3 & m_1 & m_2 \end{array}$$

- Obsérvese que en este resultado  $w_1$  mejora.

# No manipulación estratégica

- Con preferencias estrictas, si quien oferta son los hombres, ellos no tienen incentivos a manipular (lo mismo para las mujeres).

# Admisiones a la Universidad: Muchos a uno

- Este problema es muy similar excepto que varios estudiantes (análogo a mujeres) pueden ser asignados a una misma universidad (análogo a hombres). Cada universidad tiene una cuota de estudiantes que puede ingresar.
- Formalmente tenemos un conjunto de estudiantes  $S = \{s_1, \dots, s_n\}$  y universidades  $U = \{u_1, \dots, u_p\}$ . Cada universidad  $i$  tiene una cuota  $q_i$  de estudiantes que pueden ser ingresados.
- Los conceptos de estudiante (universidad) aceptable y emparejamiento estable son los naturales.
- Una diferencia importante es que este es un mercado de muchos a uno.

# Algoritmo de aceptación diferida: Extensión

- El algoritmo funciona exactamente de la misma forma que el caso de los matrimonios. Excepto que las universidades (i.e., mujeres) pueden aceptar temporalmente estudiantes hasta su cuota.
- En cada ronda las universidades rechazan los estudiantes que no son aceptables y aceptan temporalmente hasta la cuota de ingresos que tienen.
- La asignación resultante es estable (la prueba es idéntica a la anterior).
- Ahora con preferencias estrictas, la asignación es óptima para los estudiantes.

# Optimalidad para los estudiantes

- Llamemos a una universidad posible para un estudiante si este puede ser admitido a esa universidad en algún emparejamiento estable.
- Para ningun estudiante el algoritmo de aceptación diferida rechaza una universidad posible. Si esto es verdad, la asignación final es óptima para los estudiantes pues ninguna universidad mejor o igual que con la quedó emparejado sería posible.
- Para entender por qué el algoritmo no rechaza universidades posibles, supongamos por inducción que hasta cierto punto no se han rechazado universidades posibles para ningun estudiante.

# Optimalidad para los estudiantes

- Consideremos una ronda adicional y supongamos que un estudiante es rechazado por una universidad  $u$  que temporalmente prefiere a  $s_1, \dots, s_m$ . Tenemos que mostrar que es imposible para el.
- Si fuera posible entonces existiría un emparejamiento estable  $\mu$  en el que el estudiante va a la universidad  $u$ .
- Ahora en este emparejamiento  $\mu$ , algún  $s_i$  tuvo que quedar emparejado con una universidad que le gusta menos que  $u$ . Si le gustara más quiere decir que en el algoritmo de aceptación diferida fue rechazado por una universidad que era posible en contradicción con hipótesis de inducción.
- Pero si este es el caso  $s_i$  y  $u$  bloquearían a  $\mu$  y  $\mu$  no sería estable.

# No Optimalidad para las Universidades

- Ahora, contrario a lo que sucede en el caso del mercado de matrimonios (uno a uno), en este caso el algoritmo de aceptación diferida (Universidades proponen) puede no ser óptimo de Pareto en un sentido débil.
- Para ver esto veamos el siguiente ejemplo.

# No Optimalidad para las Universidades

## Example

En este ejemplo supongamos que las cuotas son  $q_{m_1} = 2, q_{m_2} = 1, q_{m_3} = 1$ , entonces el emparejamiento estable en el que las universidades proponen es dominado estrictamente por el emparejamiento  $\mu(m_1) = (w_2, w_4), \mu(m_2) = (w_1), \mu(m_3) = (w_3)$  y  $\mu(w_1) = (m_2), \mu(w_2) = (m_1), \mu(w_3) = (m_3), \mu(w_4) = (m_1)$

$R^{m_1}$	$R^{m_2}$	$R^{m_3}$	$R^{w_1}$	$R^{w_2}$	
$w_1$	$w_1$	$w_3$	$m_3$	$m_2$	
$w_2$	$w_2$	$w_1$	$m_1$	$m_1$	
$w_3$	$w_3$	$w_2$	$m_2$	$m_3$	
$w_4$	$w_4$	$w_4$	$m_1$	$m_3$	

# Contenido

- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

# ¿Qué son las matemáticas?

*What, then, to raise the old question once more, is mathematics?  
The answer, it appears, is that any argument which is carried with  
sufficiente precision is mathematical.*

Gale and Shapley. 1962. College Admissions and the Satability of  
Marriage

# Contenido

- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

## Weighted Polytope Rules

- El objetivo es encontrar mecanismos que aproximen una función objetivo  $g$  que puede no satisfacer propiedades deseables, con un mecanismo  $f$  que sí satisfaga restricciones de diseño (i.e., estabilidad).
- Por ejemplo  $g$  puede ser un mecanismo en uso que se quiere cambiar por uno nuevo.
- Automated mechanism design without money via machine learning. 2016. Narasimhan, Agarwal, Parkes.

- $N = \{1, \dots, n\}$  conjunto de agentes.
- $\Omega$  conjunto de resultados.
- $\succ$  relación de preferencia estricta (se puede extender al caso con indiferencia). El conjunto de todas las relaciones de preferencia  $\mathcal{P}$ .
- Un *mecanismo* es una función  $f : \mathcal{P} \rightarrow \Omega$ .
- Los agentes reporta sus preferencias (posiblemente de forma manipulada) y el mecanismo determina el resultado.
- Las preferencias de los agentes salen con una distribución  $\mathcal{D}$ .

- Sea  $g : \mathcal{P} \rightarrow \Omega$  un objetivo (regla) que puede no satisfacer las restricciones de diseño.
- El problema a resolver es cómo encontrar un mecanismo  $f : \mathcal{P} \rightarrow \Omega$  que si satisfaga las restricciones de diseño y sea lo más aproximado a  $g$ .
- La cercanía se mide con una función de distancia  $D : \Omega \times \Omega \rightarrow R$ .
- Sea  $\mathcal{F}$  los mecanismos que satisfacen un conjunto de restricciones de diseño.

- Dada una función objetivo  $g$  (i.e., función de elección social), diseñar un mecanismo  $f$  que arroje resultados similares al deseado socialmente y que satisfaga un conjunto de restricciones (i.e., racionalidad individual, estabilidad):

$$\min_{f \in \mathcal{F}} E_{\succ \sim \mathcal{D}} [D(g(\succ), f(\succ))] \quad (11)$$

- Ahora resolvemos la versión empírica. Generamos preferencias y valores de la función objetivo:  $(\succ^i, g(\succ^i))$ :

$$\min_{f \in \mathcal{F}} \frac{1}{m} \sum_{i=1}^m D(g(\succ^i), f(\succ^i)) \quad (12)$$

# Emparejamientos

- Supongamos el mismo número de hombres y de mujeres  $n$  (se puede extender al caso mucho a uno). Una función emparejamiento es  $y : H \cup M \rightarrow H \cup M \cup \{\phi\}$ , donde  $\phi$  denota quedarse solo.
- Dado  $\succ$  sea  $A(\succ)$  el conjunto de parejas hombre y mujer que prefieren estar emparejadas entre ellos que quedarse solos.

# Emparejamientos: Representación Lineal

- Una función de emparejamiento se puede representar como una matriz Booleana  $y \in \{0, 1\}^{nm}$  donde  $y_{hm} = 1$  si y solo si  $(h, m)$  están emparejados y la columnas (y filas) suman menos o igual a uno.
- El conjunto de funciones de emparejamiento estables  $\Omega(\succ)$  son aquellas que satisfacen las restricciones:

$$y_{hm} = 0, (h, m) \notin A(\succ) \quad (13)$$

$$y_{hm} + \sum_{h' \succ_m h} y_{h'm} + \sum_{m' \succ_h m} y_{hm'} \geq 1 \quad (14)$$

Dado un emparejamiento, la primera dice que es individualmente racional. La segunda se viola solo cuando una pareja bloquea: si  $(h, m)$  no satisfacen la restricción, bloquean (obsérvese que cada suma tiene a los sumo un uno).

- Ahora extendamos el conjunto de matrices a valores entre cero y uno. Denotamos este conjunto por  $\hat{\Omega}(\succ)$ .

## Theorem (Roth)

*Un emparejamiento es estable si y solo si es un punto extremo de  $\hat{\Omega}(\succ)$ .*

# Emparejamientos: Weighted Polytope Rule

- Definamos una función de pesos  $\lambda : \mathcal{P} \rightarrow R^{n \times n}$ .  $\lambda(\succ)_{mh}$  representa el peso que se le da a ese emparejamiento  $(m, h)$ .

## Definition (WP-Rule)

Dado  $\lambda$ :

$$f_{WP}(\succ; \lambda) = \operatorname{argmax}_{y \in \hat{\Omega}(\succ)} \sum_m \sum_h \lambda_{mh}(\succ) y_{mh} \quad (15)$$

- El algoritmo de aceptación diferida es un caso particular. Dependiendo de la función de pesos se obtiene la asignación en la que los hombres proponen o las mujeres proponen.

# WP Rules: Función de Pesos Parametrizada

- Definimos la **función de rango**: número de agentes del lado opuesto que el agente prefiere *menos* que al candidato en cuestión:

$$\text{rank}(\succ_h, m) = |\{m' \in M : m \succ_h m'\}| \quad (16)$$

$$\text{rank}(\succ_m, h) = |\{h' \in H : h \succ_m h'\}| \quad (17)$$

- Casos especiales: la WP-Rule reduce a las reglas DA para pesos específicos:
  - $\lambda_{hm}(\succ) = \text{rank}(\succ_h, m) \Rightarrow$  emparejamiento óptimo para hombres
  - $\lambda_{hm}(\succ) = \text{rank}(\succ_m, h) \Rightarrow$  emparejamiento óptimo para mujeres

- Forma parametrizada** que interpola entre ambos extremos:

$$\lambda_{hm}(\succ; W) = a_{hm} \text{rank}(\succ_h, m) + b_{hm} \text{rank}(\succ_m, h) + c_{hm} \quad (18)$$

con  $W = [a, b, c]$ ,  $a, b, c \in R^{n \times n}$ .

## Theorem (Weighted Polytope Rules)

*$f_{WP}(\gamma, \lambda)$  son estables y contienen las asignaciones que se obtienen del algoritmo de aceptación diferida.*

- Usando como conjunto de restricciones de diseño  $\mathcal{F}$  todas las WPR, los autores resuelven el problema empírico usando como distancia la distancia de Hamming (i.e., el número de veces que el emparejamiento del target es distinto al observado).
- El problema lo resuelven usando SVM. En principio pueden utilizarse Redes Neuronales Profundas.

# Contenido

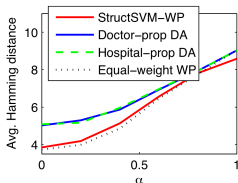
- 1 Introducción
- 2 ¿Cómo pueden salir mal las cosas?
- 3 Mercados de dos lados: Algoritmo de aceptación diferida
  - Mercado de Matrimonios: Uno a uno
  - Comportamiento Estratégico
  - Admisiones a la Universidad: Muchos a uno
- 4 ¿Qué son las matemáticas?
- 5 Automated Mechanism Design
- 6 Resultados

# Generación de Datos Sintéticos

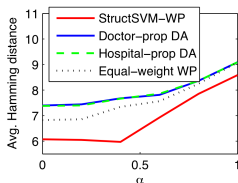
- Problema uno-a-uno con **10 hospitales y 10 doctores**; las preferencias pueden incluir indiferencia (ties).
- Las preferencias se generan combinando un **puntaje base** (común a todos) y un **puntaje específico** del agente, con parámetro  $\alpha \in [0, 1]$  que controla el peso del puntaje base:
  - $\alpha$  alto  $\Rightarrow$  preferencias muy correlacionadas  $\Rightarrow$  **menos emparejamientos estables**
  - $\alpha$  bajo  $\Rightarrow$  preferencias heterogéneas  $\Rightarrow$  mayor flexibilidad para aproximar el target
- Dos reglas objetivo **no estables** (targets):
  - **Equal-weighted Hungarian**: maximiza la suma de los rangos de cada pareja  $\sum[\text{rank}(h, m) + \text{rank}(m, h)]$
  - **Diversity-inducing Hungarian**: igual, más un bono por emparejar grupos específicos de hospitales con grupos específicos de doctores
- 1000 preferencias generadas, divididas en entrenamiento, validación y prueba (1/3 cada uno).

# Datos Sintéticos

- Se generan preferencias para 10 hombres y 10 mujeres y con un parámetro  $\alpha$  se controla la correlación entre estas.
- Se utilizan dos formas de emparejar que **no son estables**: equal-weighted y diversity-inducing.



(a) Equal-weighted Target



(b) Diversity-inducing Target

# Interpretación: Datos Sintéticos

- A mayor  $\alpha$ , el error crece para **todos los métodos**: preferencias muy correlacionadas reducen el conjunto de emparejamientos estables y la flexibilidad para aproximar el target.
- Para la regla **equal-weighted**: StructSVM-WP recupera automáticamente una WP-Rule con pesos iguales, superando a las reglas DA.
- Para la regla **diversity-inducing**: la WP-Rule entrenada supera a los tres baselines incluso en el régimen de alta correlación ( $\alpha$  grande).
- StructSVM-WP domina consistentemente a los tres baselines: DA doctor propone, DA hospital propone, y WP pesos iguales sin entrenar.

# Datos Reales: Elección de Escuelas (Wake County, NC)

- **Datos:** 37 escuelas y 5504 estudiantes del sistema escolar público de Wake County, Carolina del Norte.
- Se estima un **modelo Plackett-Luce** de preferencias y se generan preferencias para **5 escuelas, 100 estudiantes** (cuota de 25 por escuela).
- **Indiferencias:** estudiantes con 1–2 opciones reportadas; escuelas con rangos agrupados en 20 categorías.
- Cuatro reglas objetivo **no estables**:
  - **BM:** Mecanismo de Boston (propuesto por estudiantes)
  - **EH:** Hungarian con pesos iguales (*equal-weighted*)
  - **MH:** Hungarian con pesos mayores a  $1/3$  de estudiantes con estatus de minoría
  - **DH:** Hungarian *diversity-inducing*
- 200 ejemplos generados, divididos en entrenamiento y prueba (100 cada uno).

- Error de Hamming en el conjunto de prueba (menor es mejor):

Método	BM	EH	MH	DH
StructSVM-WP	<b>27.2</b>	36.0	<b>22.1</b>	<b>47.7</b>
DA (estudiante propone)	28.4	46.5	33.4	57.3
WP pesos iguales	29.7	<b>34.3</b>	23.0	53.2

- StructSVM-WP supera a los baselines en **3 de 4** reglas objetivo.
- Para EH, la WP con pesos iguales ya aproxima bien el target (el propio target asigna pesos iguales).