

El disseny de mercats sense diners

Caterina Calsamiglia

ICREA y GL de Cincies Socials Computacionals
Barcelona Supercomputing

Bojos per l'economia, 2025-26

El món està ple de plataformes

Avui moltes decisions passen a través de plataformes digitals.

Exemples:

- ▶ Google → quina informació veiem
- ▶ Uber → quin conductor ens toca
- ▶ Airbnb → quins pisos apareixen
- ▶ LinkedIn → quines feines veiem

Les plataformes decideixen qui veu què i amb qui pot interactuar.

Cada interacció deixa dades

Cada vegada que interactuem amb aquests sistemes generem dades.

- ▶ clics
- ▶ eleccions
- ▶ preferències
- ▶ temps de resposta

Això genera el que anomenem **Big Data socioeconòmic**.

Les dades reflecteixen el comportament de les persones dins d'un sistema amb regles.

Què ens diu realment el comportament?

El comportament depèn de molts factors:

- ▶ la informació disponible
- ▶ les opcions que tenim
- ▶ les regles del sistema
- ▶ els incentius

Si canviem les regles, canviem el comportament i també les dades que observem.

L'economia ajuda a entendre això

L'economia estudia:

- ▶ com prenen decisions les persones
- ▶ com les regles afecten aquestes decisions
- ▶ com podem dissenyar sistemes que promoguin les decisions desitjables

Aquest camp sovint s'anomena:

Market Design o Disseny de Mercats

Idees clau

- ▶ Pensar en la presa de decisions individuals i collectives amb cura és clau per entendre qui accedeix a què i per què.
- ▶ En les properes hores treballarem sobre dos problemes concrets, però moltes de les idees són molt transversals al disseny de política pública en general:
 - ▶ Assignació de nens a escoles
 - ▶ Assignació de joves a la universitat

Elecció d'escola

- ▶ 2/3 dels països de l'OCDE han incrementat la capacitat de tria de les famílies
- ▶ Abans, un anava a l'escola del barri...
 - ▶ Fack i Grenet (2010): preus de l'habitatge i qualitat de l'escola
 - ▶ Excepte en barris amb alta presència d'escoles privades
- ▶ Triaven aquells que podien triar barri o optar per opcions privades
- ▶ Així doncs, l'increment en la tria:
 - + Evita capitalitzar el valor de l'escola a través del preu de l'habitatge
 - + Igualtat d'oportunitats
 - Preocupacions:
 - Només les famílies més afavorides aprofiten l'opció?
 - Incrementa la segregació?
- ▶ Com s'implementa aquesta tria és clau

Quin és el problema principal a resoldre?

- ▶ La capacitat de cada escola és limitada, algunes poden tenir més demanda que places disponibles
- ▶ En moltes ciutats el problema es resol de manera centralitzada
 - ▶ Les famílies entreguen una llista amb un rànquing d'escoles, per si no entren a la seva opció preferida
 - ▶ Si hi ha més demanda que places en una escola, es prioritzen les sol·licituds segons punts que defineixen la prioritat (barri, familiars, loteria)
 - ▶ Entren les sol·licituds amb més punts fins a omplir la capacitat
 - ▶ Què passa amb els que no entren a la seva opció preferida?
- ▶ **Mecanisme de Boston (BM)** d'Abdulkadiroglu i Sönmez (2003): l'assignació en cada ronda és definitiva

	2005	2006	2007	2008	2009	2010
Enter in 1st choice	88	85	83	85	87	89
Enter in 2nd choice	5	6	6	6	6	6
Enter in 3rd-10th choice	2	3	3	3	2	3
Does not enter in any of their choice	5	6	8	6	5	2
1st choice is neighborhood	73	80	82	83	82	82

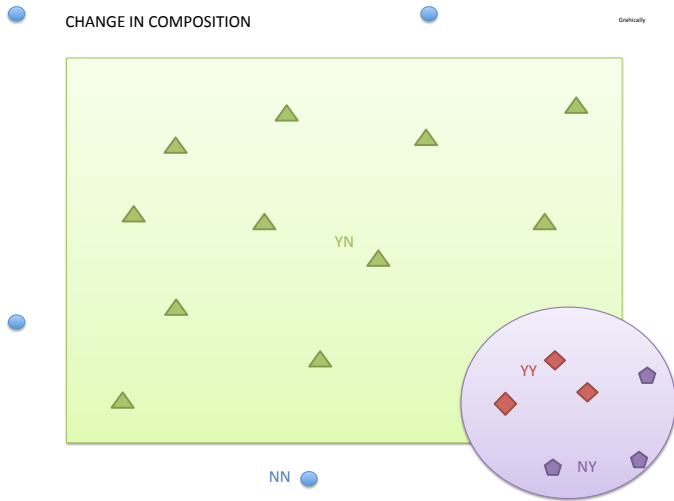
Per què és rellevant que es pugui demanar el que es vol?

- ▶ Que les famílies descartin algunes opcions, si al final no hi ha lloc per a tothom?
 - ▶ Abdulkadiroglu, Che i Yasuda (2010) i Miralles (2008): auto-selecció eficient
 - ▶ Calsamiglia i Miralles (2017): les famílies demanen l'escola per la qual tenen ms punts
 - ▶ Calsamiglia, Martinez-Mora i Miralles (2020): les escoles privades poden generar segregació
- ▶ Entendre quin d'aquests fenòmens teòrics s'aplica és fonamental

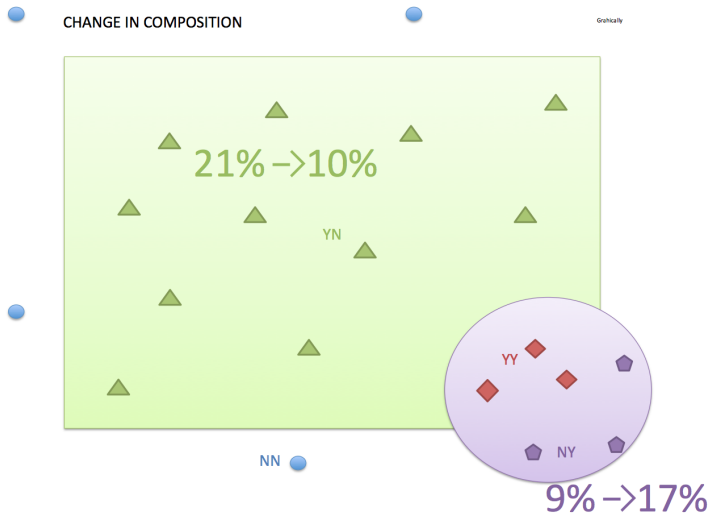
Aversió al risc o valor de la proximitat?

	2005	2006	2007	2008	2009	2010
Enter in 1st choice	88	85	83	85	87	89
Enter in 2nd choice	5	6	6	6	6	6
Enter in 3rd-10th choice	2	3	3	3	2	3
Does not enter in any of their choice	5	6	8	6	5	2
1st choice is neighborhood	73	80	82	83	82	82

Canvi en el disseny de barris a Barcelona el 2007



Calsamiglia and Güell (2018)



Més resultats

- ▶ Algunes famílies s'arriesquen molt
 - ▶ Acaben no entrant a cap de les escoles sol·licitades
 - ▶ Acaben a una escola privada (aunque en Barcelona només un 4% de les escoles són privades i el 14% dels que s'arriesquen, i no entren, acaben a la privada).

Conclusions

- ▶ Amb el mecanisme de Boston, les prioritats tenen un paper fonamental: les famílies demanen, en gran mesura, escoles per les quals tenen més punts.
- ▶ Les escoles privades poden facilitar la segregació entre escoles inicialment similars.

Alternativa: Gale Shapley Acceptació Diferida

1. En cada ronda, noms es retenen les sollicituds
 2. A la segent ronda es consideren les sollicituds retingudes juntament amb les noves
 3. Les assignacions són definitives quan cap sollicitud és rebutjada
- ▶ Demanar una escola molt sollicitada no redueix les possibilitats d'entrar en qualsevol altra
 - ▶ Aquestes normes permeten a les famílies demanar el que volen sense haver d'assumir riscos
 - ▶ **Calsamiglia i Miralles (2017) demostren que, si hi ha prioritats per barri i hi ha escoles que ningú vol, l'assignació acaba sent la del barri**

Calsamiglia, Fu, Güell (2020)

- ▶ Creem una base de dades anonimitzada ajuntant 5 bases de dades diferents a través de l'IDESCAT (!)
 - ▶ Elecció d'escoles
 - ▶ Cens i padró de famílies i escoles
 - ▶ Assignació i matrícula tant en el sistema públic com en el privat
 - ▶ Avaluacions externes realitzades a les escoles
 - ▶ Enquesta a famílies sobre imports pagats en els diferents centres concertats
- ▶ Creem un model que replica el problema de decisió de les famílies
- ▶ Inferim les preferències
- ▶ Simulem què hauria passat a Barcelona si canviéssim de Boston a Gale-Shapley

Resultats

- ▶ Canviar només el mecanisme (com van fer a Boston i en moltes ciutats d'Anglaterra):
 - ▶ **beneficia el 10%** dels individus i **perjudica el 28%**
- ▶ Per a aquells per als quals l'escola del barri NO és la preferida:
 - ▶ Amb GS, el 30% acaba al barri i el 52% en la seva preferida
 - ▶ Amb BM, el 20% acaba al barri i el 59% en la seva preferida

Conclusions

- ▶ BM fora les famílies a demanar escola per la qual tenen prioritat, afavorint aquells que es poden arriscar
 - ▶ perquè tenen punts per una escola dolenta
 - ▶ perquè tenen una alternativa privada
- ▶ GS permet sol·licitar les escoles desitjades, però acaba sobreassignant al barri!
- ▶ **Cal clarament plantejar-se el sistema de punts i si volem que tenir opcions privades afecti l'elecció!**

Accés a la universitat

- ▶ A la universitat pública, (com a les escoles) la capacitat d'un grau ve prefixada
- ▶ Idealment cada individu hauria de poder escollir la carrera que vol estudiar, però no sempre és possible
- ▶ En molts països existeix un sistema centralitzat on s'entrega un rànding de preferències
- ▶ Si hi ha sobredemanda, utilitzem les notes del batxillerat i la de la Selectivitat-PAU per ordenar els candidats
 - ▶ Bagrut (Israel), Maturità (Itàlia), Bac (França)
 - ▶ Vestibular (Brasil), Gaokao (Xina), OSS (Turquia)
- ▶ El sistema aplicat és el d'Acceptació Diferida de Gale Shapley

- ▶ ¿Per què utilitzem un examen com la selectivitat tenint tantes avaluacions a les escoles?
 - ▶ Per estandarditzar els continguts temàtics
 - ▶ Estandarditzar l'avaluació
 - ▶ Transparència i objectivitat
- ▶ Riscos:
 - ▶ Un examen únic avalua dos anys de treball
 - ▶ La incapacitat de fer l'examen pot afectar l'equitat del sistema
 - ▶ El format de l'examen condiciona la manera d'aprendre durant el batxillerat
- ▶ Selectivitat i la Covid-19
 - ▶ Al març del 2020 tanquen les escoles, tothom a casa...
 - ▶ ¿Què farem amb la selectivitat?
 - ▶ ¿Es podria NO FER?

La predictibilitat de la selectivitat

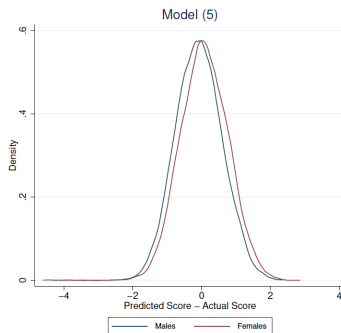
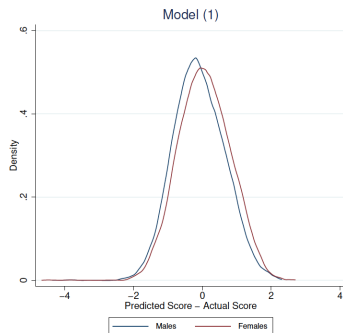
- ▶ Si tenim dades històriques de:
 - ▶ selectivitat
 - ▶ notes a l'escola
 - ▶ escola assistida
 - ▶ nivell socioeconòmic
 - ▶ gènere
 - ▶ ...
- ▶ i sabem les notes dels estudiants a dia d'avui, ¿podem predir la nota que trauràs a la selectivitat?

Table 2: Selectivitat. Regressions and predictions

(a) In sample (cohorts 2010 and 2011)

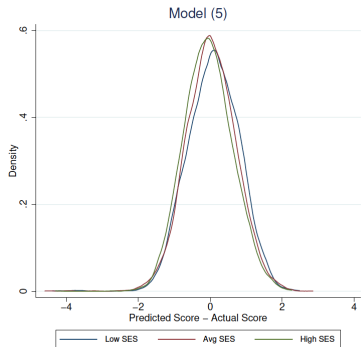
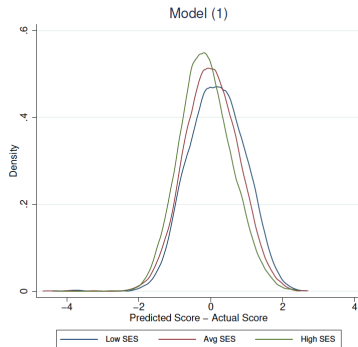
	(1)	(2)	(3)	(4)	(5)
High School GPA	0.670** (0.0099)	0.669** (0.0090)	0.681** (0.0111)	0.174** (0.0235)	0.0995** (0.0192)
Art Track		-0.106* (0.0495)	-0.0794 (0.0684)	-0.0959* (0.0488)	-0.0335 (0.0568)
Scientific Track		0.00536 (0.0149)	-0.0109 (0.0152)	-0.0396** (0.0140)	-0.0599** (0.0140)
Peer average GPA		-0.00101 (0.0522)			
Catalan				0.0645** (0.0150)	0.108** (0.0112)
Spanish				0.0632** (0.0154)	0.0858** (0.0109)
History/Philosophy				0.0952** (0.0132)	0.118** (0.0112)
Foreign Language				0.351** (0.0109)	0.354** (0.0094)
Track subjects				0.0773** (0.0132)	0.0911** (0.0106)
School FE	No	No	Yes	No	Yes
N	20425	20425	20425	20425	20425
R^2	0.442	0.442	0.518	0.507	0.585
RMSE	0.747	0.746	0.694	0.701	0.644
MAE	0.598	0.597	0.552	0.559	0.510
abs rank deviation	0.187	0.187	0.169	0.171	0.154

Per genere



Note. Differences between predicted and actual rank, cohort 2012

Per nivell socioeconòmic



Note. Differences between predicted and actual rank, cohort 2012

Reacció asimètrica a la pressió

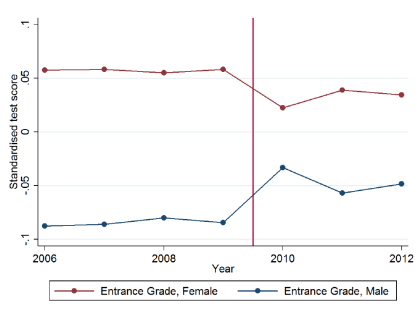
- ▶ Azmat, Calsamiglia i Iriberry (2016)
 - ▶ Panell de dades d'una escola a Barcelona: ESO, batxillerat, selectivitat
 - ▶ Exàmens de diferents pesos
 - ▶ Les dones tenen millor rendiment en general, però aquesta diferència es redueix quan el pes dels exàmens és més gran
 - ▶ La diferència desapareix a la selectivitat
- ▶ Calsamiglia Fawaz (2020)
 - ▶ Dades de Corea d'examen de prova vs examen real d'accés a la universitat
 - ▶ Exàmens de diferents pesos
 - ▶ La diferència de gènere desapareix a les escoles només de noies(!)
 - ▶ L'assignació d'escoles a Corea és aleatòria (!)
 - ▶ La diferència en el comportament entre homes i dones sembla ser *nurture*, no *nature*

Reforma de la selectivitat el 2010 a Catalunya

- ▶ Abans del 2010 la nota d'accés era 60% nota de batxillerat i 40% selectivitat
- ▶ A partir del 2010: la nota va del 0 al 14 i s'afegeix la nota de dues assignatures específiques (2 punts cadascuna)
- ▶ Les notes de selectivitat de les assignatures específiques tenen un pes molt més gran!
- ▶ ¿Impacte sobre el gènere?

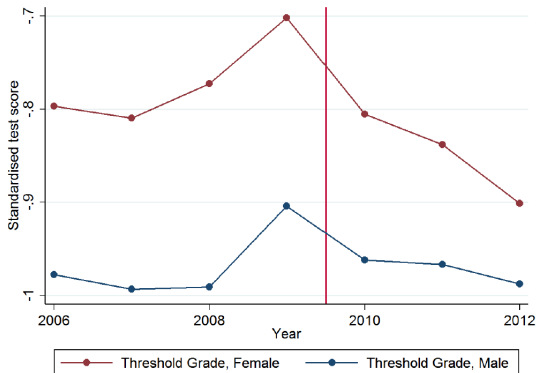
Impacte en els resultats

Figure 1: Effect of the reform on the gender gap in Entrance Grades



Notes de tall de les carreres a les quals accedeixen

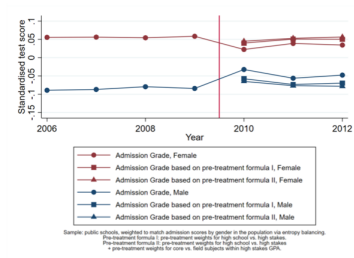
Threshold Grades



Possibles motius

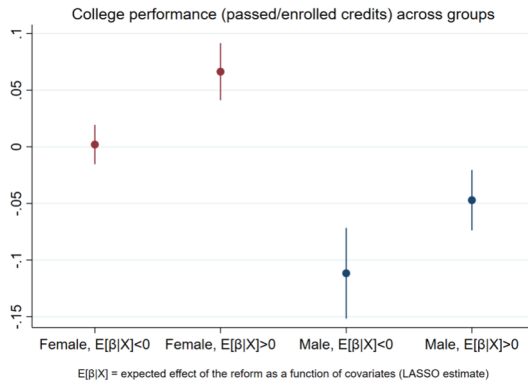
- ▶ Mecànic:
 - ▶ més pes a la selectivitat
 - ▶ més pes en algunes assignatures concretes
- ▶ Comportament:
 - ▶ En comptar més, més diferència a la selectivitat
 - ▶ Anticipant això, més diferència a l'escola

Machanical vs Behavioral



- ▶ 70% of the effect is mechanical
- ▶ Does not imply absence of effects on performance or effort reallocation, but suggests they would mostly cancel out

Seleccionem millor?



Discussió i conclusions

- ▶ ¿Seleccióem millor ara?
- ▶ L'error de predicció, ¿captura alguna cosa interessant?
- ▶ ¿En base a què, sobretot, per a què seleccióem?
- ▶ Problema "identificat" a Xile, EUA, etc.
- ▶ ¿De quina altra manera podem definir qui entra i qui no?

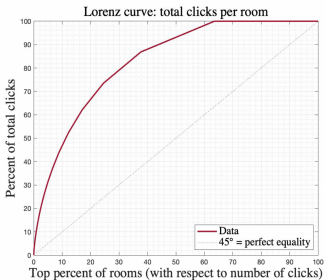
Exemple 1: Lloguer d'habitacions en línia

Details

- Current **ranking algorithm** is unpersonalized
 - **Inputs:** unrelated to customer preferences, such as date room was posted, landlord is "registered" (linked bank account)
 - **Ranking invariant across users**
- **Statistics**
 - 1,202 users vs. 45,462 rooms
 - 173,223 searches (~144 per user on avg.)
 - ≥ 2 million search results (~12-20 per search)
 - $P(\text{click}) = 4.3\%$ (1 out of ~23)
 - $P(\text{request}) = 0.41\%$ (1 out of ~244)
 - $P(\text{request} \mid \text{clicked}) = 9.5\%$ (1 out of ~11)



Measuring congestion



- To quantify congestion we use “Lorenz curve” (underlies Gini coefficient)
 - 45° line: no congestion
- Bowed shape indicates congestion
 - Top 20% of rooms get 70% of clicks;
 - ~40% of rooms never get clicked

Figure: Concentració en les habitacions observades

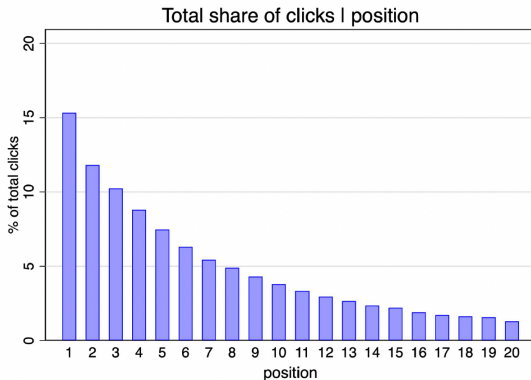
Congestió: preferències o disseny de la plataforma?

- ▶ Observem una congestió extrema:
 - ▶ molts usuaris es concentren en un conjunt petit d'habitacions,
 - ▶ moltes habitacions gairebé no reben atenció.
- ▶ Són possibles dues explicacions molt diferents.

Dues hipòtesis

- ▶ **Preferències:** els usuaris realment prefereixen les mateixes poques habitacions.
- ▶ **Disseny:** els usuaris són dirigits cap a les mateixes habitacions segons com es mostren les opcions.

Congestion and position



- Request probabilities **invariant to position, conditioning on having clicked**
- Suggests preferences are distinct from position: informs empirical approach

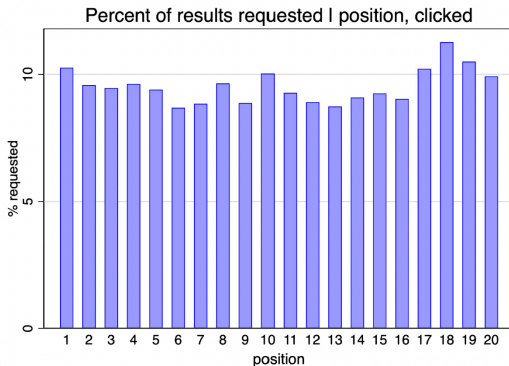


Figure: Les sol.licituds no estan correlacionades amb el rànkning

Què ens permet observar la plataforma

- ▶ Per a cada cerca observem dues etapes del comportament:
 - ▶ **Visualitzacions / clics:** quines habitacions decideixen inspeccionar els usuaris.
 - ▶ **Sol.licituds:** a quines habitacions els usuaris expressen interès activament.
- ▶ Crucialment, la informació disponible per als usuaris difereix entre aquestes etapes.

Estructura d'informació

- ▶ Abans de fer clic: ordre + un petit conjunt de característiques visibles.
- ▶ Després de fer clic: moltes característiques addicionals i detallades de l'habitació.

Què ens diuen realment les visualitzacions i les sol.licituds

- ▶ **Visualitzacions / clics** reflecteixen:
 - ▶ el rànquing i la visibilitat,
 - ▶ característiques observables limitades,
 - ▶ restriccions d'atenció.
- ▶ **Sol.licituds (condicionades a la visualització)** reflecteixen:
 - ▶ com valoren els usuaris el conjunt complet de característiques de l'habitació,
 - ▶ els compromisos entre característiques quan s'eliminen els efectes de posició.

Recuperant preferències a gran escala

- ▶ Modelant les sol·licituds condicionades a les visualitzacions podem:
 - ▶ separar els efectes de posició de les preferències,
 - ▶ estimar com influeixen les diferents característiques en les decisions dels usuaris.
- ▶ Les dades de comportament a gran escala són essencials perquè:
 - ▶ les preferències són heterogènies,
 - ▶ la identificació depèn de decisions repetides en contextos diferents.

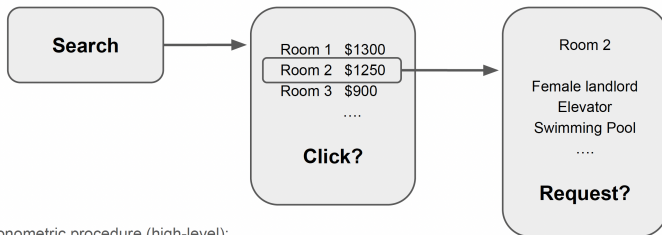
Per què el Big Data és important aquí

L'estimació granular i heterogènia de preferències només és possible amb grans quantitats de dades de comportament.

Empirical approach

Is congestion problem due to faulty recommendation algorithm?

- What if we introduce **personalization**? Make algorithm preference-aware?
- To do this, we need to estimate customer preferences.



Econometric procedure (high-level):

- Estimate customer/room utilities from request propensity conditional on click
- Estimate click propensity taking utilities as given.

Figure: Aprenent preferències

De les preferències als contrafactuals

- ▶ Un cop recuperades les preferències, podem preguntar-nos:
 - ▶ Què passaria si les habitacions es mostressin de manera diferent?
 - ▶ Què passaria si els rànquings fossin personalitzats?
- ▶ Canviar l'algoritme canvia el comportament, de manera que els contrafactuals ho han de tenir en compte.

Idea clau

Entendre el procés que genera les dades és essencial per fer anàlisis contrafactuals amb sentit.

Trade-off: congestion in clicks vs. utility

- Summarize via **Efficiency frontier**: plotting utility and congestion on different axes
- Tradeoff: achieve less congestion only by sacrificing utility
 - +vice versa
- Black dot** (data)
- (congestion, utility) point observed in the data is **inefficient**.

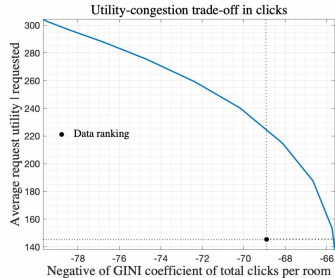


Figure: Podem millorar simultàniament en les dues dimensions!

Què hem après amb aquest exemple?

- ▶ Les dades de plataformes no reflecteixen noms preferncies.
- ▶ Reflecteixen també:
 - ▶ com es mostren les opcions,
 - ▶ què veuen els usuaris,
 - ▶ quines decisions prenen en cada etapa.
- ▶ Per entendre les dades hem d'entendre el comportament.

Per què és important entendre el comportament

- ▶ Les plataformes generen enormes quantitats de dades.
- ▶ Però aquestes dades depenen de:
 - ▶ el disseny de la plataforma
 - ▶ l'algoritme de recomanació
 - ▶ el que els usuaris poden veure
- ▶ Sense entendre això podem interpretar malament les dades.

Què podem fer amb aquestes dades

- ▶ Recuperar preferències reals dels usuaris
- ▶ Entendre quins factors impulsen el comportament
- ▶ Simular què passaria si canviem els algoritmes

Objectiu

Utilitzar dades i models per dissenyar sistemes que funcionin millor.