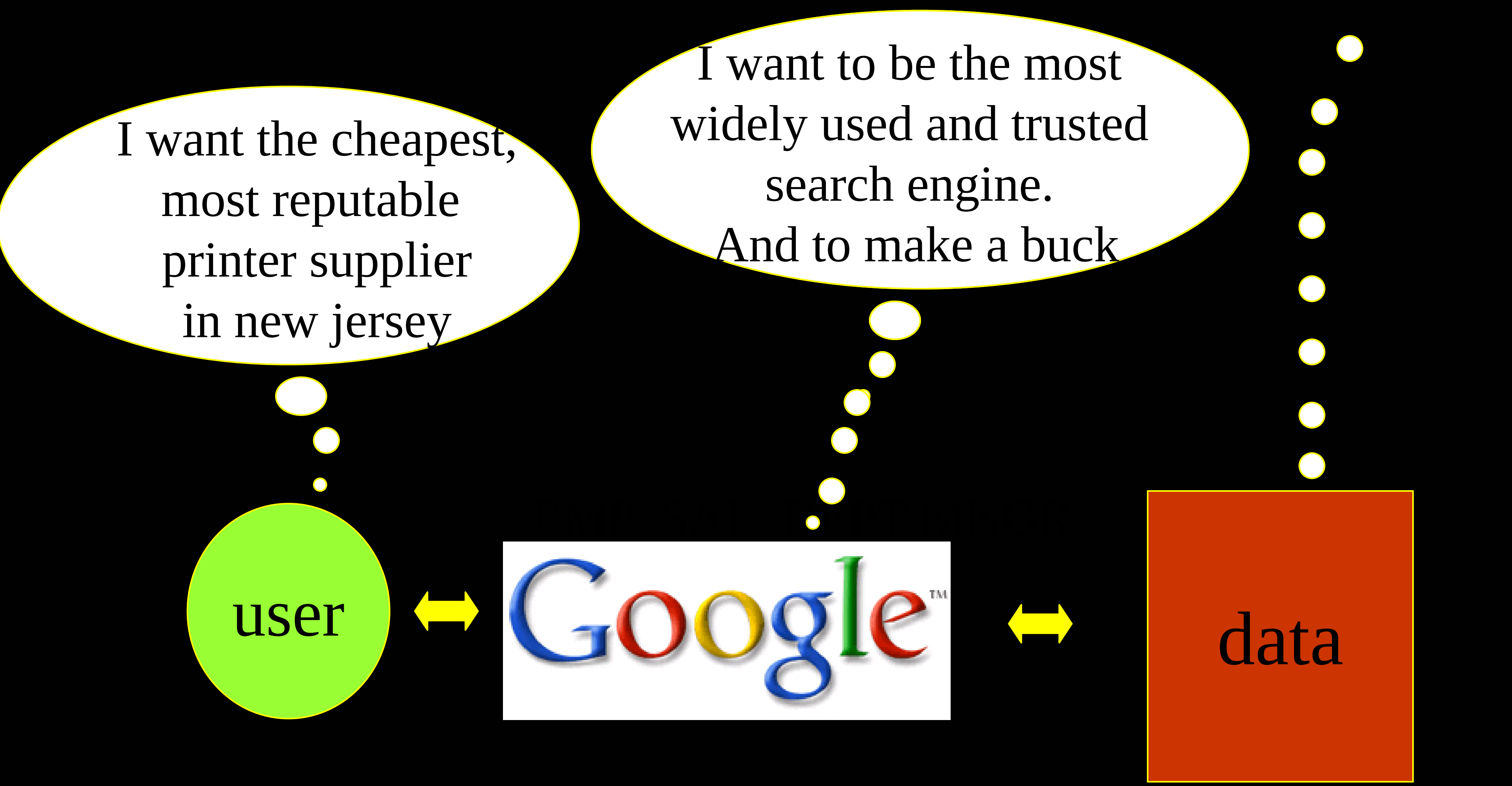
Economic Aspects of Information Systems

Christos H. Papadimitriou
UC Berkeley
Google christos

The old information systems

The new information systems



Logic databases

Canfly(x,y) :- Canfly(x,z), Flight(z,y).

Canfly(x,y) :- Flight(x,y).

- Some of the most elegant work in Database Theory during the 1980s
- Never picked up by practitioners
- Main criticism: "Good for poorly designed databases"

The return of logic

- The Internet is a deliberately badly designed database.
- e.g.:

```
spam(p,n,s) :- spam(p',n',s),address(p,n,s), n= n ± 1.
spam(p,n,s) :- spam(p',n',s'),carpool(p,p'),
                                    address(p,n,s).
   spam(christos, 303, badlands).
```

What is privacy?

- one of society’s most vital concerns
- central for e-commerce
- arguably the most crucial and far-reaching current challenge and mission of CS
- least understood scientifically (e.g., is it rational?)
- see, e.g., www.sims.berkeley.edu/~hal, ~/pam,
- [Stanford Law Review, June 2000]

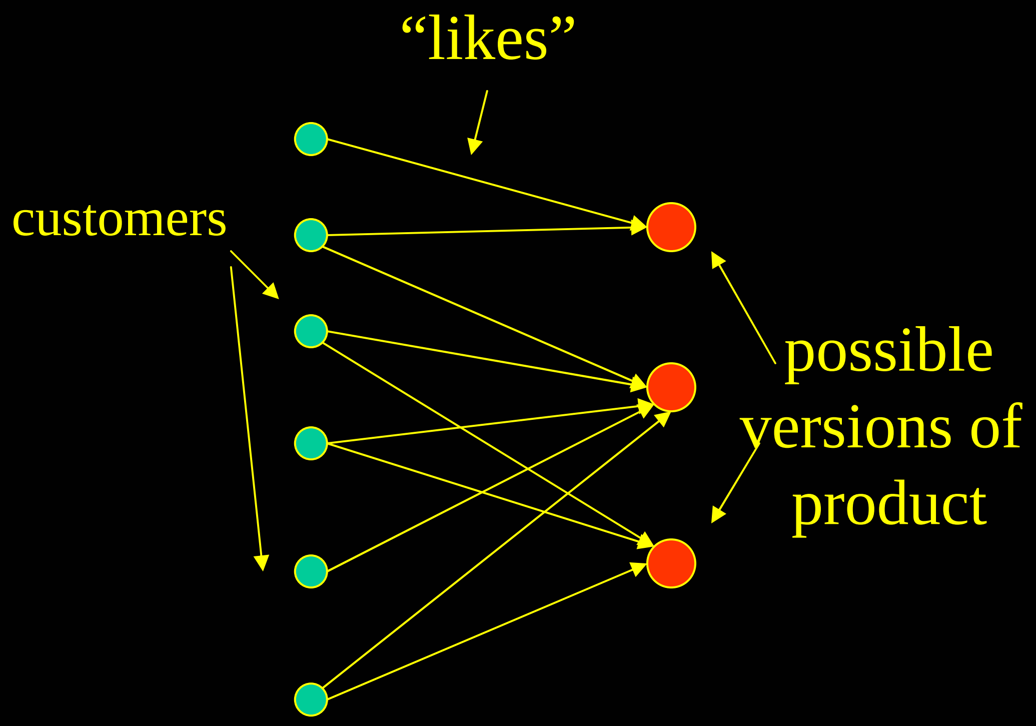
some thoughts on privacy

- also an economic problem
- surrendering private information is either good or bad for you
- example: privacy vs. search costs in computer purchasing

thoughts on privacy (cont.)

- *personal information is intellectual property controlled by others, often bearing negative royalty*
- selling mailing lists vs. selling aggregate information: false dilemma
- Proposal: *Take into account the individual's utility when using personal data for decision-making*

e.g., marketing survey



e.g. total revenue: $2m = 10$

- company's utility is proportional to the majority
- customer's utility is 1 if in the majority

- *how should all participants be compensated?*

Collaborative Game Theory

- How should A, B, C split the loot (=20)?
- We are given what each subset can achieve by itself as a function v from the powerset of {A,B,C} to the reals
- $v(\{\}) = 0$

Values of v

- A: 10
- B: 0
- C: 6
- AB: 14
- BC: 9
- AC: 16
- ABC: 20

first idea (notion of "fairness"): the core

A vector $(x_1, x_2, \ldots, x_n)$ with $\Sigma_i\, x_i = v([n])$ (= 20)
is in the core if for all S we have

$$x[S] \geq v(S)$$

In our example: A gets 11, B gets 3, C gets 6

Problem: Core is often empty (e.g., AB = 15)

second idea: the Shapley value

$$x_i = E_\pi(v[\{j: \pi(j) \leq \pi(i)\}] - v[\{j: \pi(j) < \pi(i)\}])$$

(Meaning: Assume that the agents arrive at random. Pay each one his/her contribution. Average over all possible orders of arrival.)

Theorem [Shapley]: The Shapley value is the only allocation that satisfies Shapley's axioms.

In our example…

- A gets:

10/3 + 14/6 + 10/6 + 11/3 = 11

- B gets:

0/3 + 4/6+ 3/6 +4/3 = 2.5

- C gets the rest = 6.5
- NB: Split the cost of a trip among hosts…

Values of *v*

- A: 10
- B: 0
- C: 6
- AB: 14
- BC: 9
- AC: 16
- ABC: 20

e.g., the UN security council

- 5 permanent, 10 non-permanent
- A resolution passes if voted by a majority of the 15, including all 5 P
- $v[S] = 1$ if $|S| > 7$ *and* S contains 1,2,3,4,5; otherwise 0
- What is the Shapley value (~power) of each P member? Of each NP member?

e.g., the UN security council

- What is the probability, when you are the 8th arrival, that all of 1,…,5 have arrived?
- Ans: Choose(10,2)/Choose(15,7) ~ .7%

Permanent members: ~ 18%

Therefore, P ≠ NP

third idea: bargaining set
fourth idea: nucleolus

.

.

.

seventeenth idea: the von Neumann-Morgenstern solution

[Deng and P. 1990] complexity-theoretic critique of solution concepts

Applying to the market survey problem

- Suppose largest *minority* is *r*
- An allocation is in the core as long as losers get 0, vendor gets > 2*r*, *winners split an amount up to twice their victory margin*
- (plus another technical condition saying that split must not be too skewed)

market survey problem: Shapley value

- Suppose margin of victory is at least ε > 0%
- (realistic, close elections never happen in real life…)
- Vendor gets *m*(1+ ε)
- Winners get 1+ ε
- Losers get ε
- (and so, no compensation is necessary)

e.g., recommendation system

- Each participant *i* knows a set of items B_i
- Each benefits 1 from every new item
- Core: empty, unless the sets are disjoint!
- Shapley value: For each item you know, you are owed an amount equal to
  1 / (#people who know about it)

  --i.e., *novelty pays*

e.g., collaborative filtering

- Each participant likes/dislikes a set of items (participant is a vector of 0, ±1)
- The "similarity" of two agents is the inner product of their vectors
- There are *k* "well separated types" (vectors of ± 1), and each agent is a *random perturbation* and *random masking* of a type

collaborative filtering (cont.)

- An agent gets advice on a 0 by asking the most similar other agent who has a ± 1 in that position
- Value of this advice is the product of the agent’s true value and the advice.
- How should agents be compensated (or charged) for their participation?

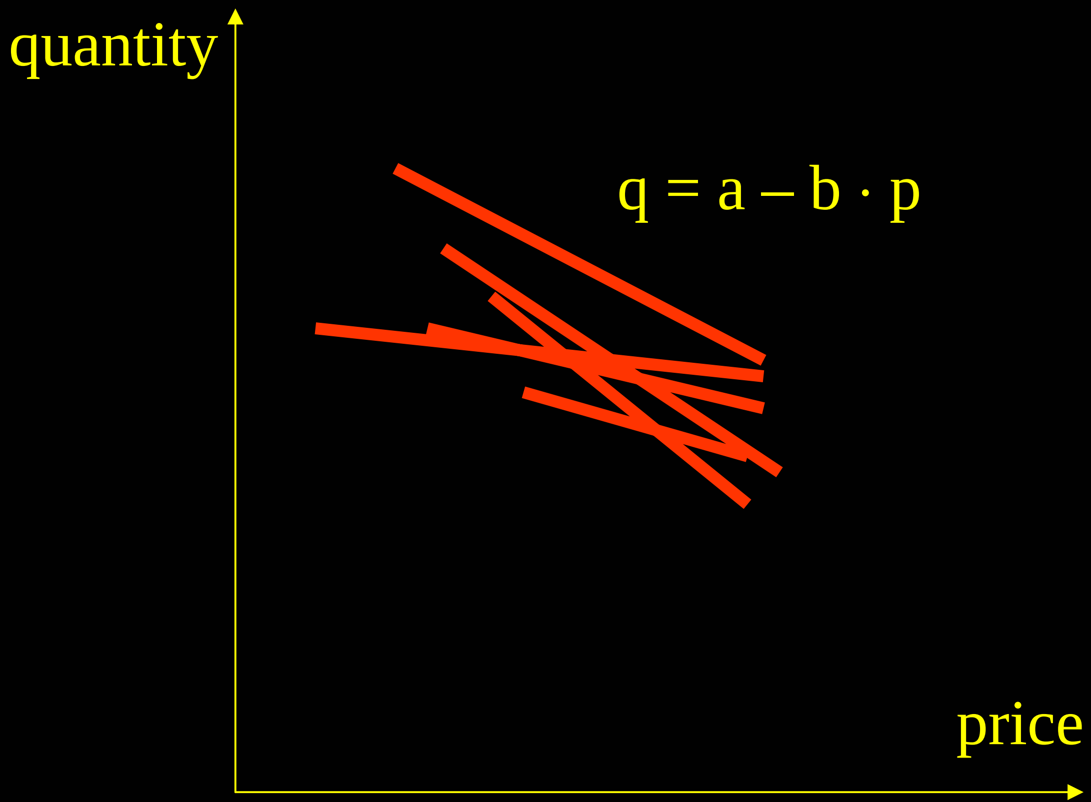
collaborative filtering (result)

Theorem: An agent's compensation (= value to the community) *is an increasing function of how typical* (close to his/her type) *the agent is.*

The economics of clustering

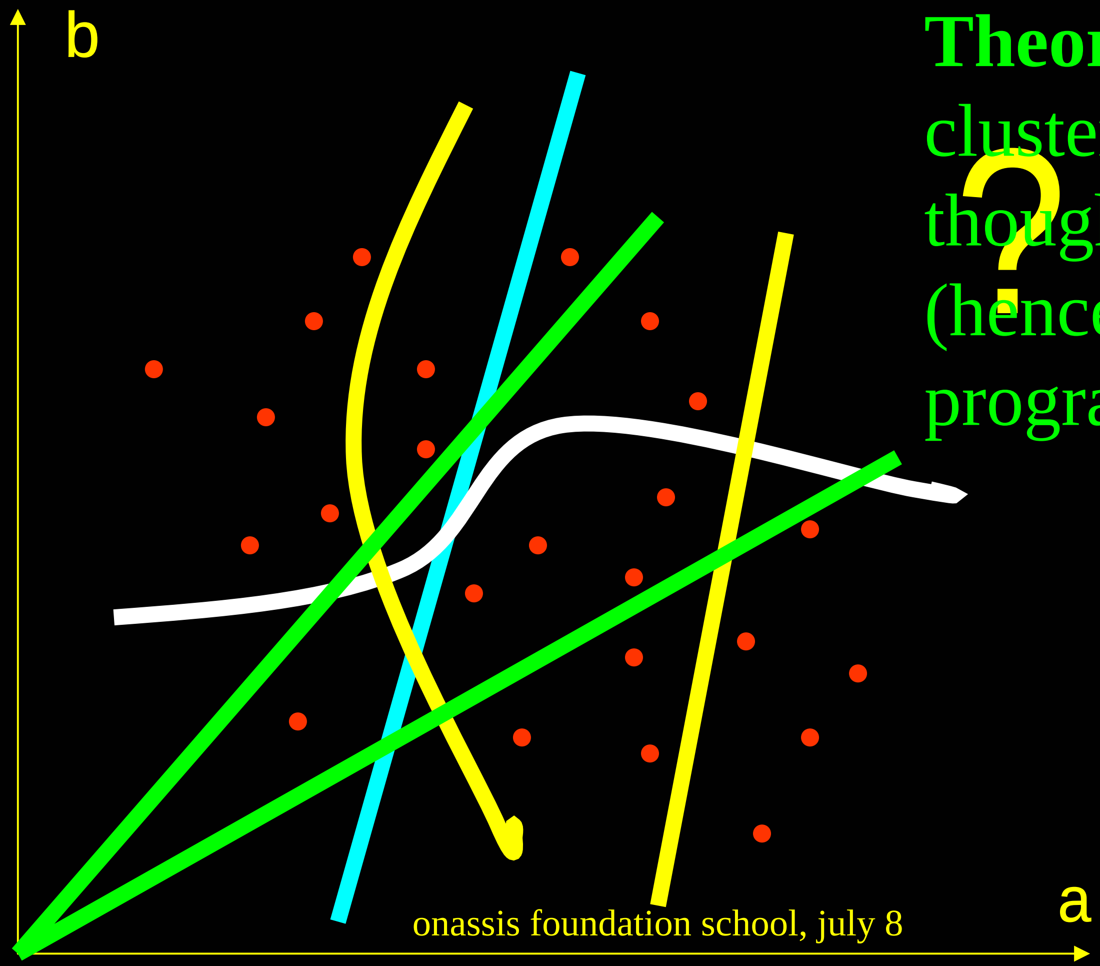
- The practice of clustering: Confusion, too many criteria and heuristics, no guidelines
- The theory of clustering: ditto!
- “It’s the economy, stupid!”
  [Kleinberg, P., Raghavan STOC 98, JDKD 99]

Example: market segmentation



Segment monopolistic market to maximize revenue

or, in the a – b plane:



Theorem: Optimum clustering is by lines though the origin (hence: $O(n^2)$ dynamic programing)

So…

- The new databases are selfish agents, and you understand them through economics
- Privacy has an interesting and central economic aspect
- Which gives rise to neat math/algorithmic problems
- And clustering is a meaningful problem only in a well-defined economic context