The Dynamics of Information

Bernardo A. Huberman

Information Dynamics Laboratory
HP Labs

tapping tacit knowledge within social networks

- discover informal communities
- determine how information flows through these communities
- use that knowledge to discover what people are about and harvest their preferences and knowledge

discovering communities

Bruegel, Peter the Younger. Village Feast

traditional methods accurate but laborious

informal communities

communities that form around tasks or topics

– scientific and technical communities (ziman, crane)

– bureaucracies (crozier)

– how they grow and evolve to solve problems (huberman & hogg)

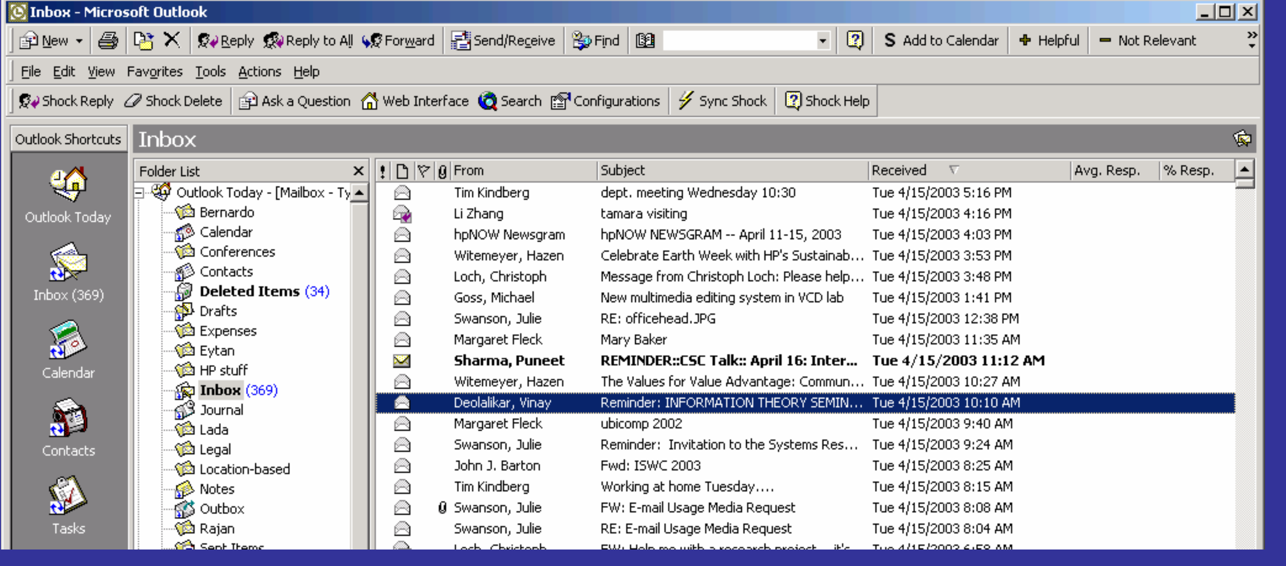
– how information flows within organizations (allen)

the measurement problem: interviews and surveys are accurate but time consuming. worse, they don’t scale

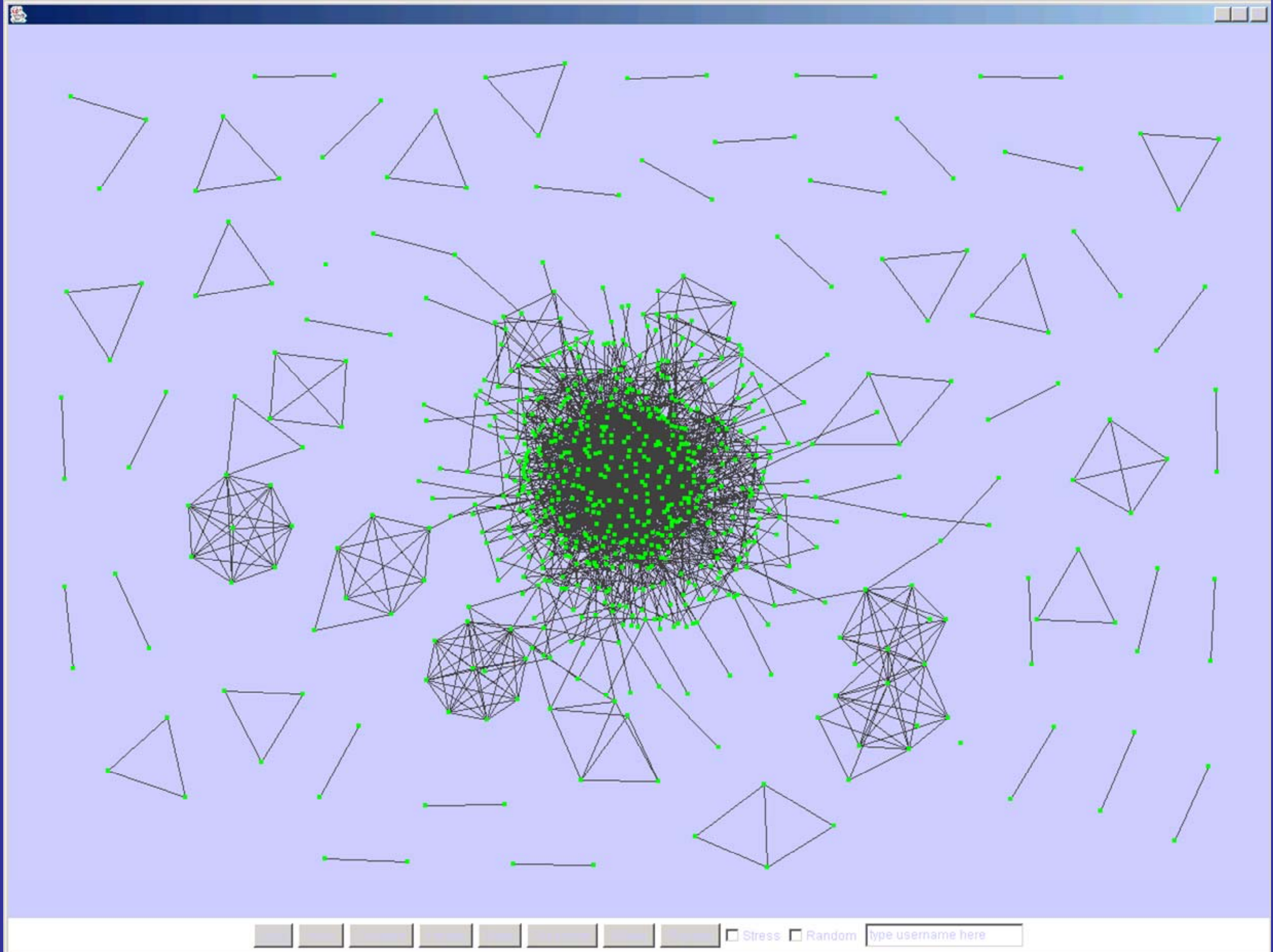
uncovering communities with e-mail

tyler,huberman and wilkinson, in *Communities and Technologies*, Kluwer Academic (2003)

- e-mail is a rich source of communication data
  - virtually everyone in the “knowledge economy” uses it
  - It provides data in a convenient format for research

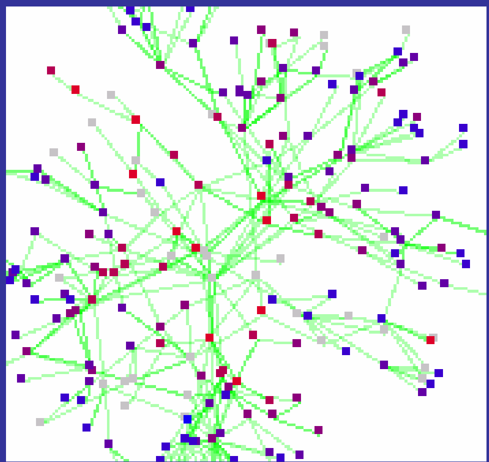


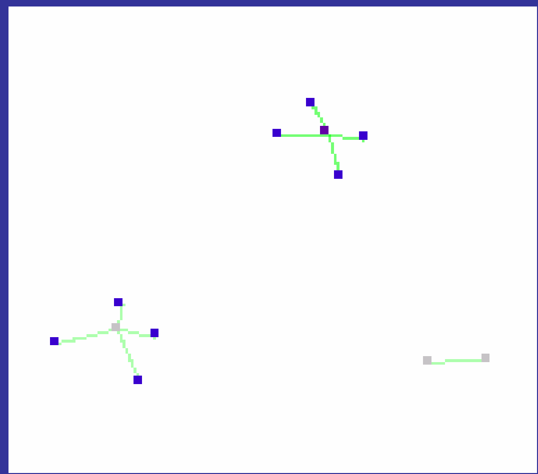
hp labs email network



our goal

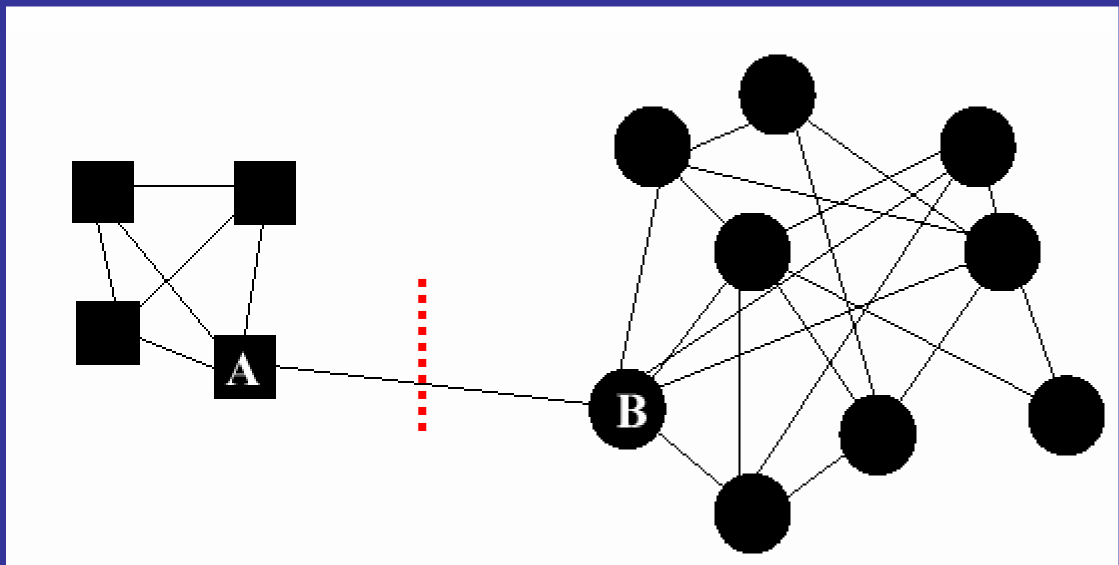
- decompose an organization's email network (dense and jumbled) into communities of practice (clean and distinct)





find communities using betweenness centrality

a graph has community structure if it consists of groups of nodes with many more links within each group than between different groups



betweeness of an edge: number of shortest paths that traverse it

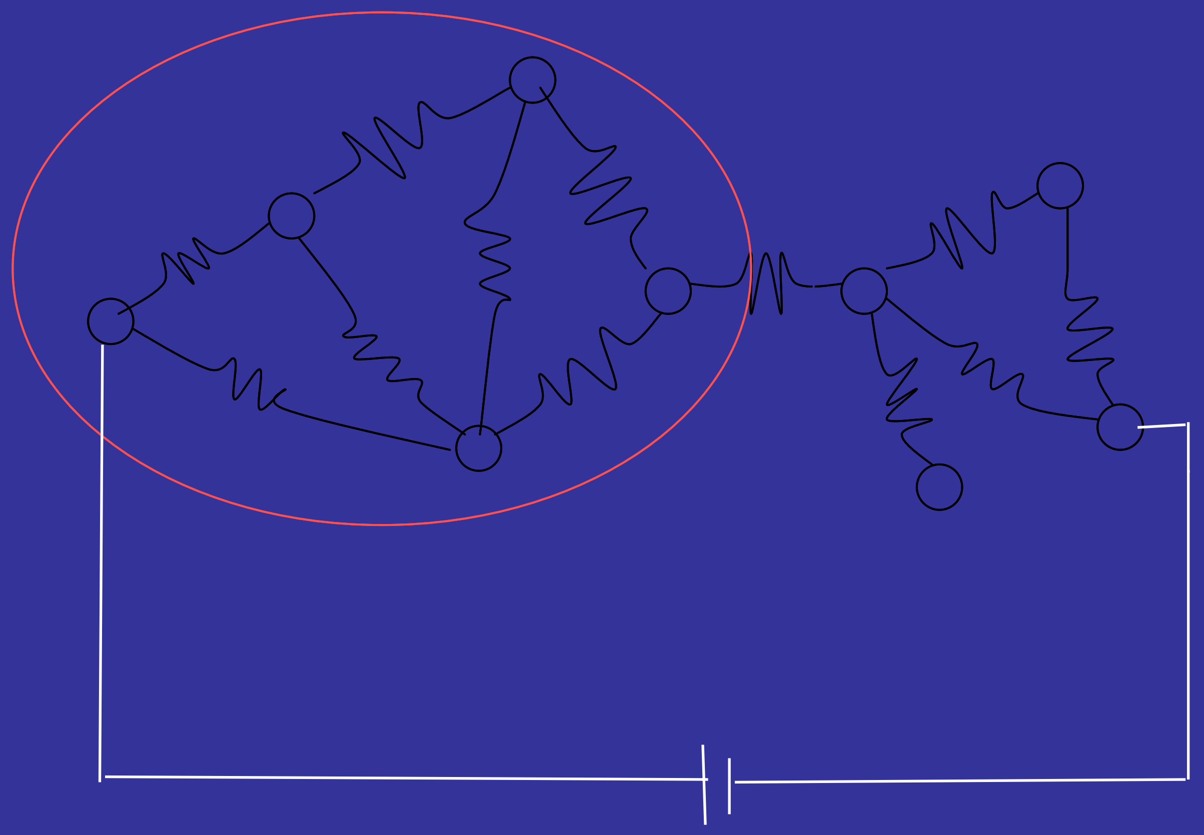
a problem

betweeness centrality is slow (scales as the cube of the number of nodes (Brandes, Girvan and Newman, Wilkinson and Huberman)

we have designed an algorithm that runs much faster (linearly in the number of nodes (*Wu and Huberman, Eur. Phys. Journal B38, 331-338 (2004*).

a different method

wu and huberman *Eur. Phys. Journal, B38, 331 (2004)*



examples

rragan	HPL Advanced Studies
olmos	HPL Advanced Studies
samuels	HPL Advanced Studies
saifi	HPL Advanced Studies
zhiyong	HPL Advanced Studies
gunyoung	HPL Advanced Studies
larade	HPL Advanced Studies

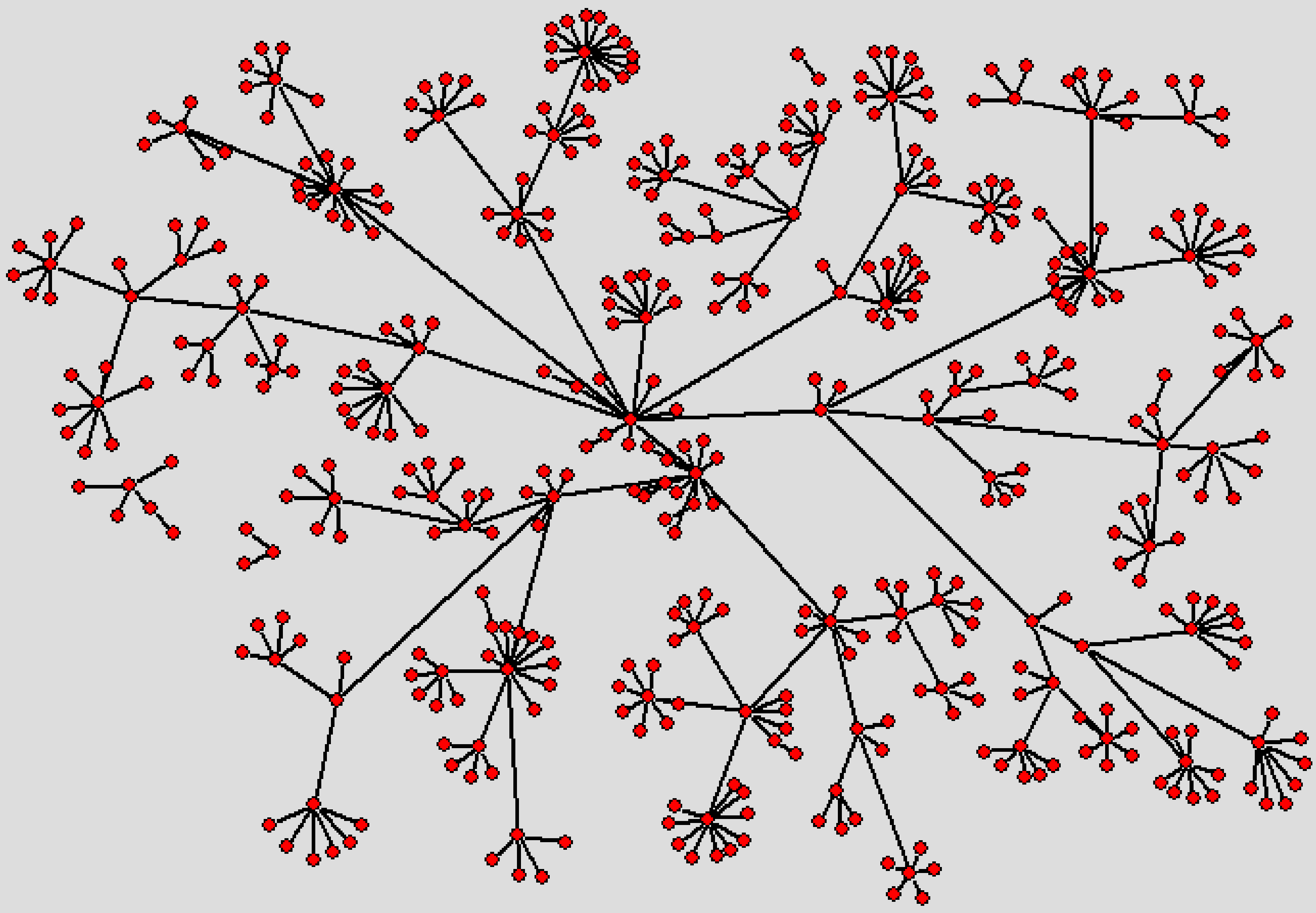
penrose	Mobile & Media Systems Lab
mistyr	HPL Advanced Studies
vinayd	HPL Advanced Studies
seroussi	HPL Advanced Studies
tsachyw	HPL Advanced Studies

reedrob	University Relations
carterpa	University Relations
sbrodeur	University Relations
pruyne	Internet Systems & Storage Lab
bouzon	University Relations
lmorell	University Relations
marcek	University Relations

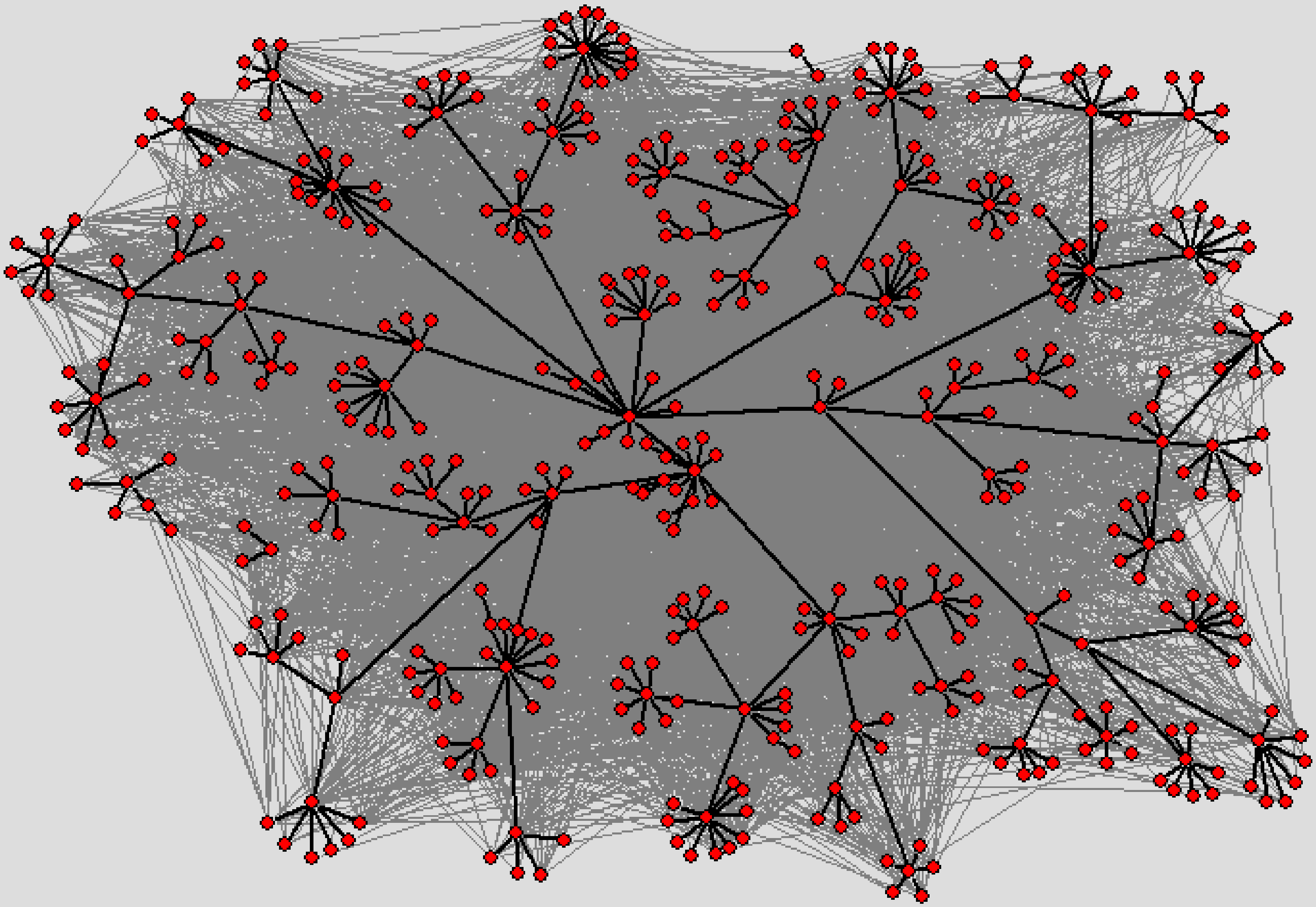
venky	Mobile & Media Systems Lab
dohlberg	HPL Advanced Studies
kvincent	Hardcopy Tech Lab
pmcc	University Relations
trangvu	HPL Communications
markstei	HPL Advanced Studies
hollerb	HPL Research Operations
krishnav	Handheld HQ
babcock	REWS Americas
gita	Solutions & Services Tech Cntr
bgee	HPL - Research Operations
meisi	HPL - Research Operations
henze	Information Access Lab

kuekes	HPL Advanced Studies
thogg	Systems Research Lab
kychen	Intelligent Enterprise Tech Lb
lfine	Systems Research Lab
akarp	Intelligent Enterprise Tech Lb

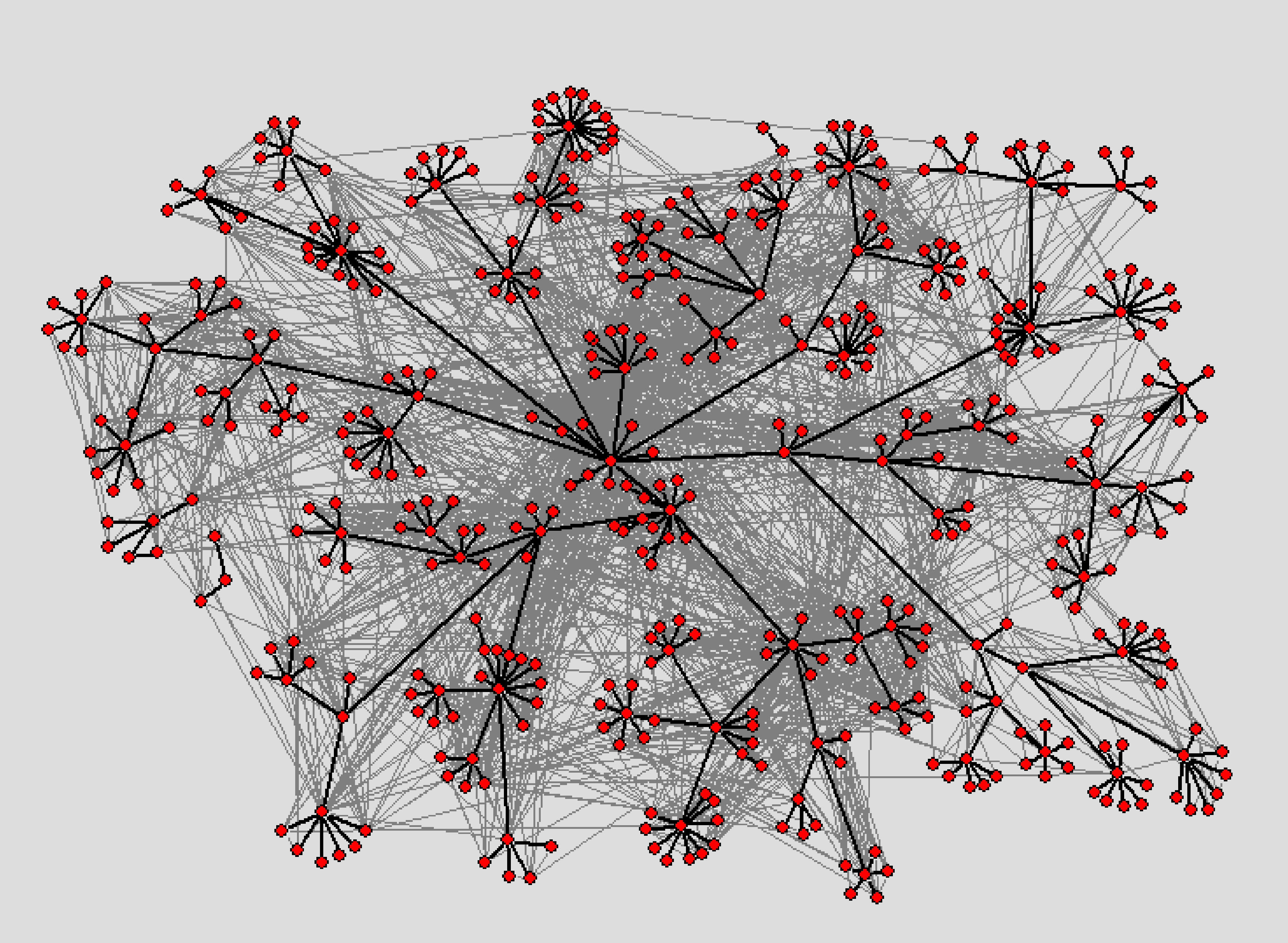
organizational hierarchy



email correspondents scrambled

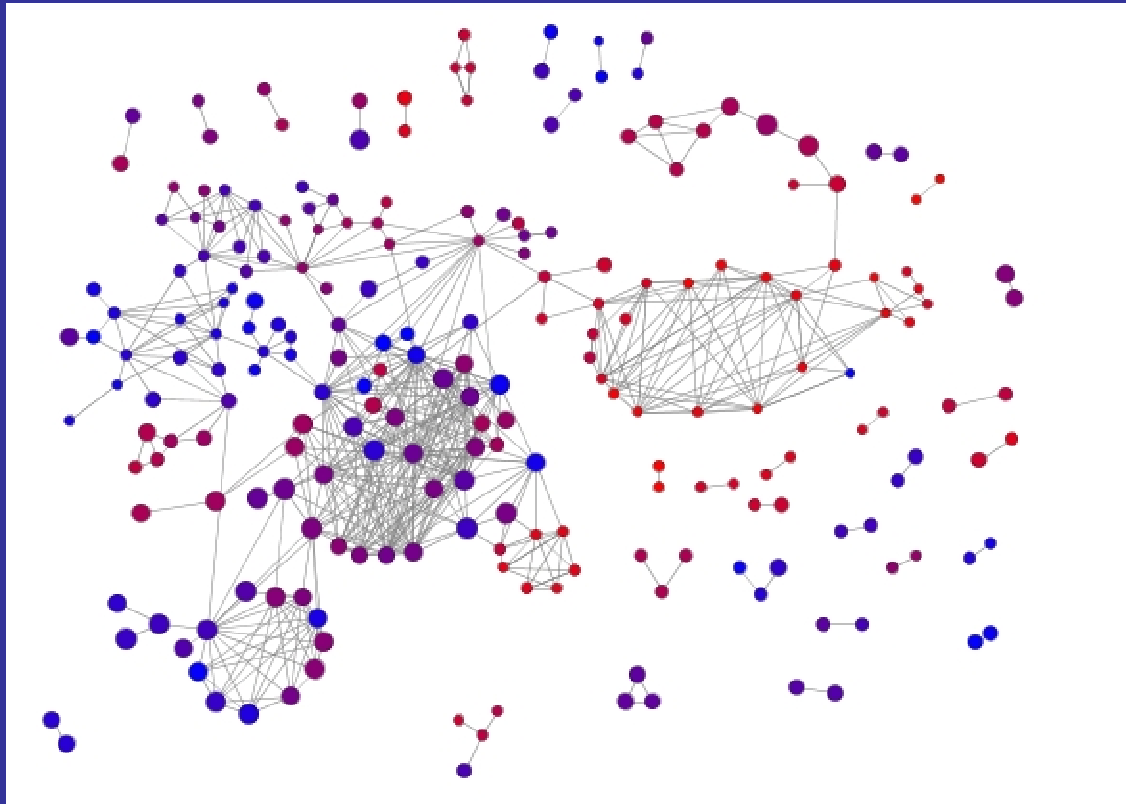


actual email correspondence



document similarity by usage

similarity: overlap in users accessing documents



earlier documents are blue, later ones are red.
size of node reflects the number of users accessing the document.
l. adamic

HPS-mining knowledge briefs

Paul Johansen | SAM | AMCI | Tech Consulting | Systems Integration | 32 docs viewed

Paul Johansen is a consultant with the .NET Solutions group within the Central EMS Practice in Minneapolis, Minnesota. Paul specializes in e-commerce UI and middle tier development and their related Microsoft technologies. In his spare time he enjoys the freezing Minnesota weather, cheering for the Vikings, Twins, Wolves and Wild and traveling the world.

users similar to Paul Johansen

sim	name	unit	group	function	family	#docs
DOCS 0.35	John R Bugarin 	SAM	AMCI	Solution Architech	Systems Integration	30
		John Bugarin is a member of the .NET Results North American Team. He has extensive experience developing customized solutions in Domino, Microsoft, and WebSphere. He is certified MCSD for .N MCAD for .NET, MCSD for Visual Studio 6.0, MCSE for Windows 2000, and MCDBA for MSSQL 20				
DOCS 0.29	Tom Kern 	SAM	AMCI	Tech Consulting	Systems Integration	236
		Tom Kern is a consultant for the Enterprise Microsoft Services .Net Solutions practice. Tom has wo on a variety of custom software projects based on Microsoft technologies.				
DOCS 0.26	Martyn Dowsett 	SEM	EMCI	Tech Consulting	Systems Integration	46
		Martyn Dowsett is a member of EMEA C&I currently working with Microsoft .NET. He has been designing, developing, and testing various kinds of software since 1979 and has experienced many examples of "how not to do things". He has worked on many projects and is experienced in the full project lifecycle. His current interests are round all things .Net.				

a new people finder

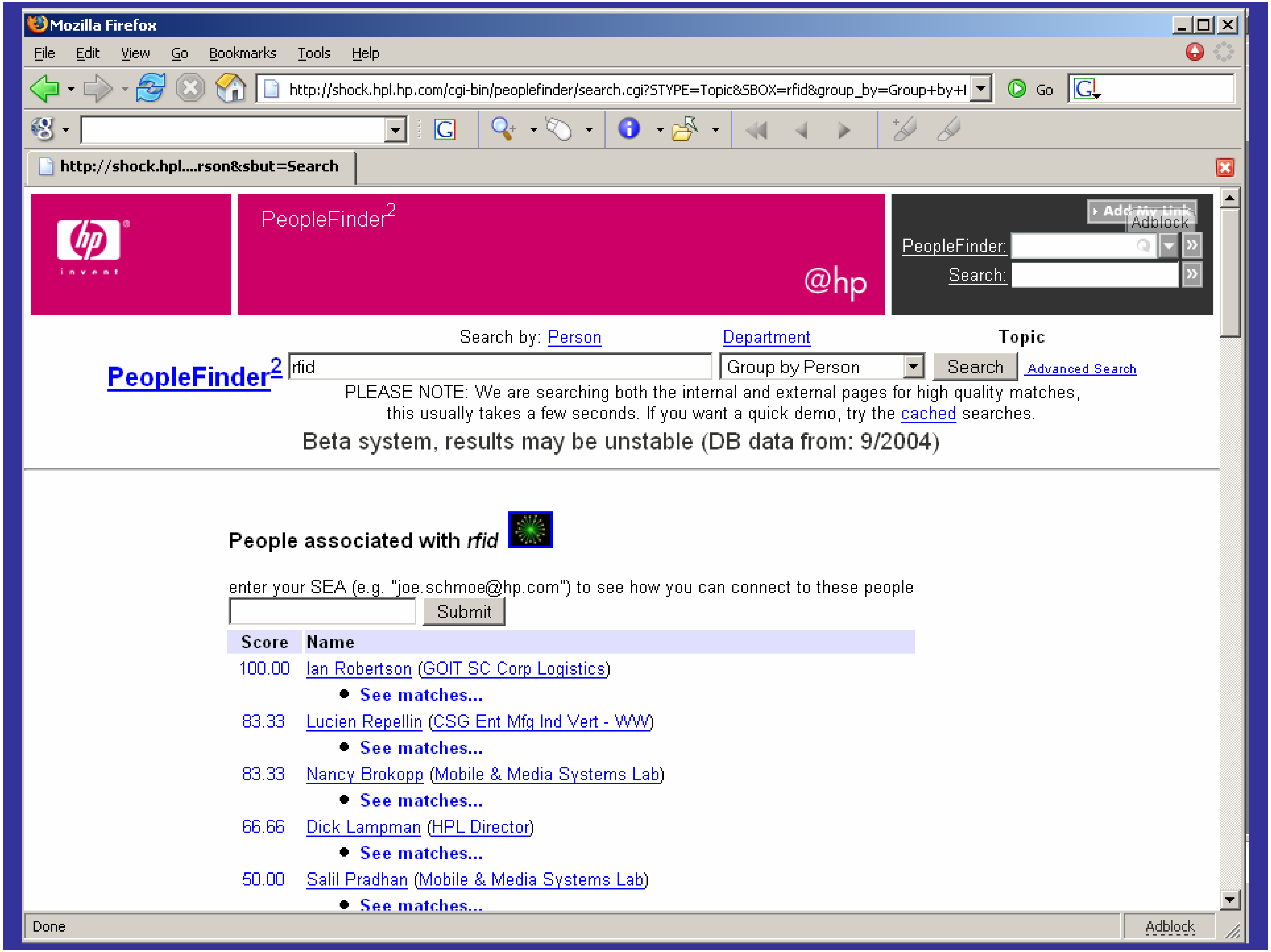
there is a trove of information in power point presentations, public repositories within the organization, and the internal website of the enterprise

peoplefinder^2 allows you to find out what people are *about,* as opposed to where in the organization they belong

it also discovers who is working on what

http://shock.hpl.hp.com/peoplefinder/

e. adar and l. adamic


Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://shock.hpl.hp.com/cgi-bin/peoplefinder/search.cgi?STYPE=Topic&SBOX=rfid&group_by=Group+by+I
Go
http://shock.hpl....rson&sbut=Search
invent
PeopleFinder2
@hp
Add My Link
Adblock
PeopleFinder:
Search:
Search by: Person
Department
Topic
PeopleFinder2
rfid
Group by Person
Search
Advanced Search
PLEASE NOTE: We are searching both the internal and external pages for high quality matches,
this usually takes a few seconds. If you want a quick demo, try the cached searches.
Beta system, results may be unstable (DB data from: 9/2004)
People associated with rfid
enter your SEA (e.g. "joe.schmoe@hp.com") to see how you can connect to these people
Submit
Score Name
100.00 Ian Robertson (GOIT SC Corp Logistics)
See matches...
83.33 Lucien Repellin (CSG Ent Mfg Ind Vert - WW)
See matches...
83.33 Nancy Brokopp (Mobile & Media Systems Lab)
See matches...
66.66 Dick Lampman (HPL Director)
See matches...
50.00 Salil Pradhan (Mobile & Media Systems Lab)
See matches...
Done
Adblock

information flow

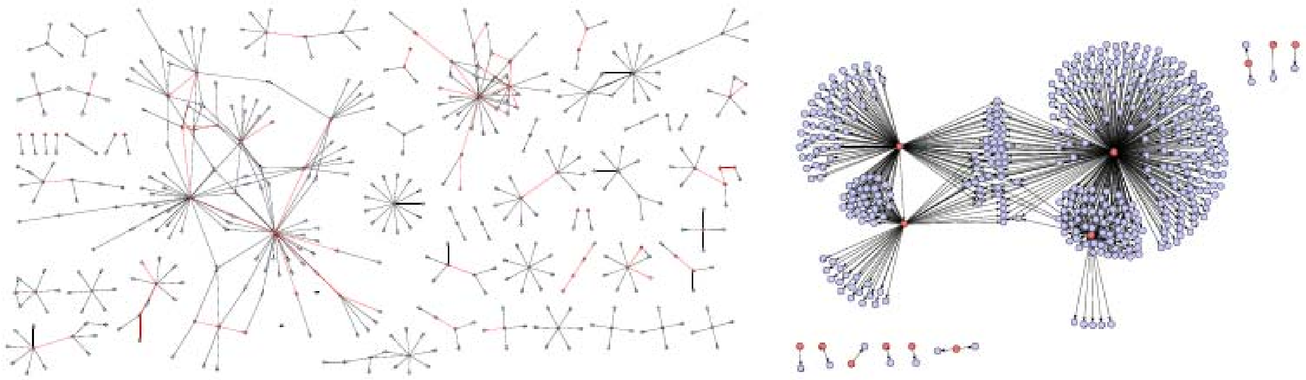
how does information flow in a community or organization?

does the structure of the social network affect it?

how far does it spread?

Wu, Adamic and Huberman

recommendation networks



(a) Medical book (b) Japanese graphic novel

Figure 1: Examples of two product recommendation networks: (a) First aid study guide *First Aid for the USMLE Step*, (b) Japanese graphic novel (manga) *Oh My Goddess!: Mara Strikes Back*.

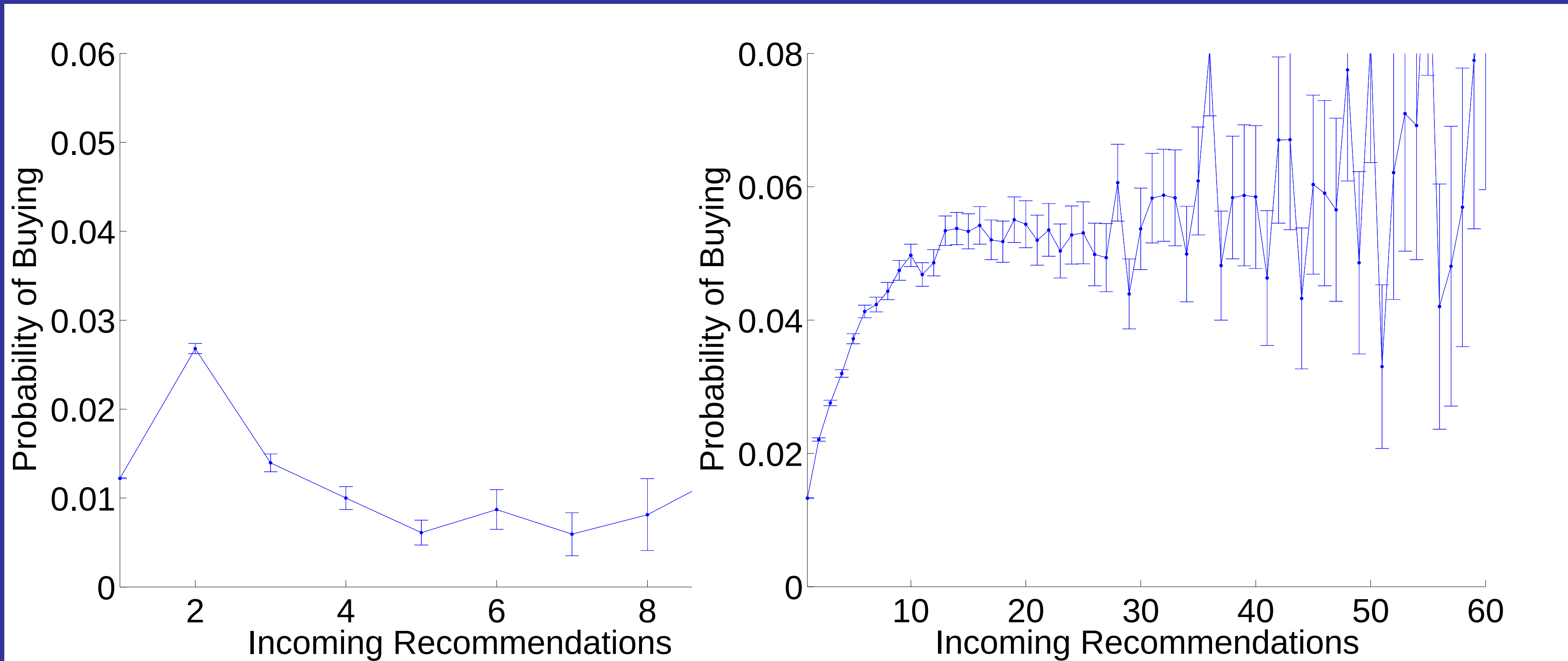
15 million recommendations and 4 million customers

j. leskovec, l.adamic and b.a. huberman

does receiving more recommendations increase the likelihood of buying?

BOOKS

DVDs



so, how effective is viral marketing?

- recommendations do not propagate very far (on average)
- but there are rare instances where the information chain is long
- they are not very effective at eliciting purchases

the future

we all care about it.

and invest resources in finding out about it.

Caravaggio ,The Fortune Teller, 1596-97 *“it is hard to predict anything, especially the future”*

Niels Bohr

how do organizations predict?

- they ask the experts (and consultants)
- have meetings (lots of them)
- designate someone as forecaster
- take a vote (not very good)

an alternative: markets

- markets aggregate and reveal information (hayek, lucas, etc.)
- to predict outcomes, use markets where the asset is information (rather than a physical good)
- example:
  - iowa electronic markets

markets within organizations

-problematic-

- low participation
- illiquidity
- information traps
- hard to motivate
- easily manipulated

a new mechanism

(with kay-yut chen and leslie fine)

- it identifies participants that have good predictive talents, and extracts their risk attitudes
- it induces them to be truthful
- while avoiding the pitfalls of small groups
- it aggregates information in nonlinear fashion

Information Systems Frontiers, Vol. 5, 47-61 (2003)

Management Science, Vol. 50, 983-994 (2004)

what is it based on?

people are not all the same

- –think of the information in peoples' heads as the assets and use portfolio theory
- –use a market mechanism to determine a individual's risk attitudes and performance

then, ask people to forecast and perform a nonlinear aggregation of their results taking into account their risk characteristics

the information gathering process is simple, decentralized in time, and inexpensive to implement

two stages

stage 1: a market for contingent securities.

it provides behavioral information, such as risk attitudes –synchronous-

stage 2: participants generate predictions on outcomes, which are then aggregated.

incorporates behavioral information

-asynchronous-

stage 2- forecasting

- participants are given 100 tickets
- to be allocated among 10 securities
- this determines probabilities
- true state pays according to the number of tickets allocated to it

aggregating predictions

the probability of event *S* occurring, conditioned on *I*, is given by

$$P(s \mid I) = \frac{p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \dots p_{s_N}^{\beta_N}}{\sum_{\forall s} p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \dots p_{s_N}^{\beta_N}}$$

with β an exponent that denotes behavioral attitudes

>1 risk averse

<1 risk seeking

=1 risk neutral

what determines the exponent?

$$P(s \mid I) = \frac{p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \dots p_{s_N}^{\beta_N}}{\sum_{\forall s} p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \dots p_{s_N}^{\beta_N}}$$

normalization constant

$$\beta_i = r(V_i / \sigma_i)c$$

~sum of prices/winning payoff
It measures market risk

holding value/risk
- measures relative risk of individuals

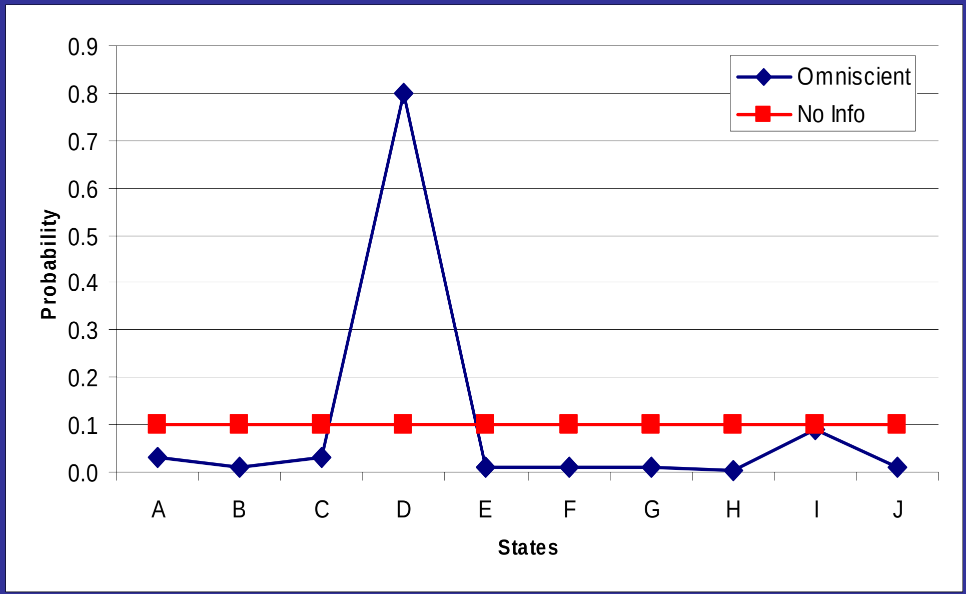
experiments

- human subjects in the laboratory (*hp* labs)
- each group receives diverse information
- run the two-stage mechanism
- and measure its performance

results

comparison to omniscient probability

Kullback-Leibler = 1.453

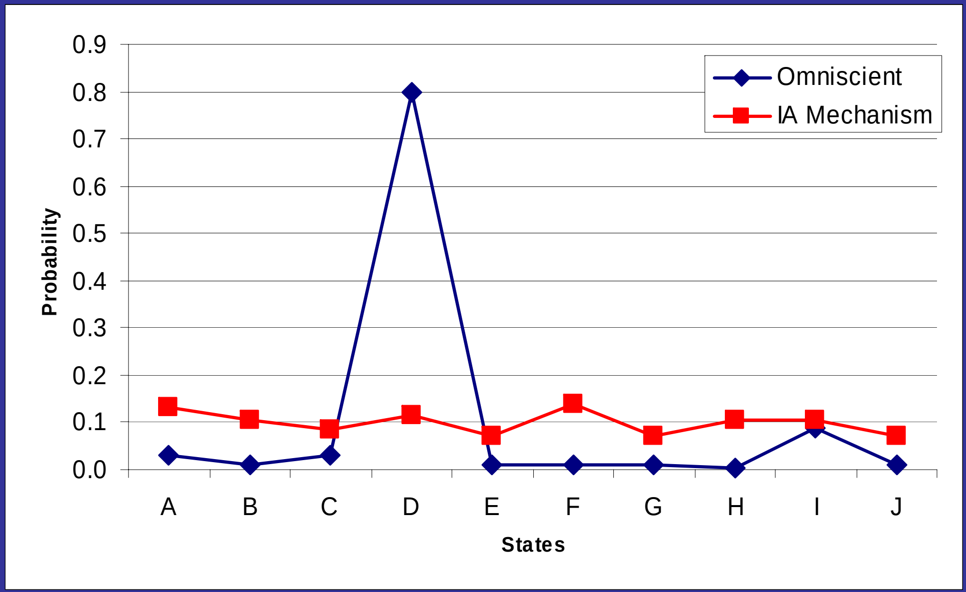


Experiment 4, Period 17
No Information

results

comparison to omniscient probability

Kullback-Leibler = 1.337

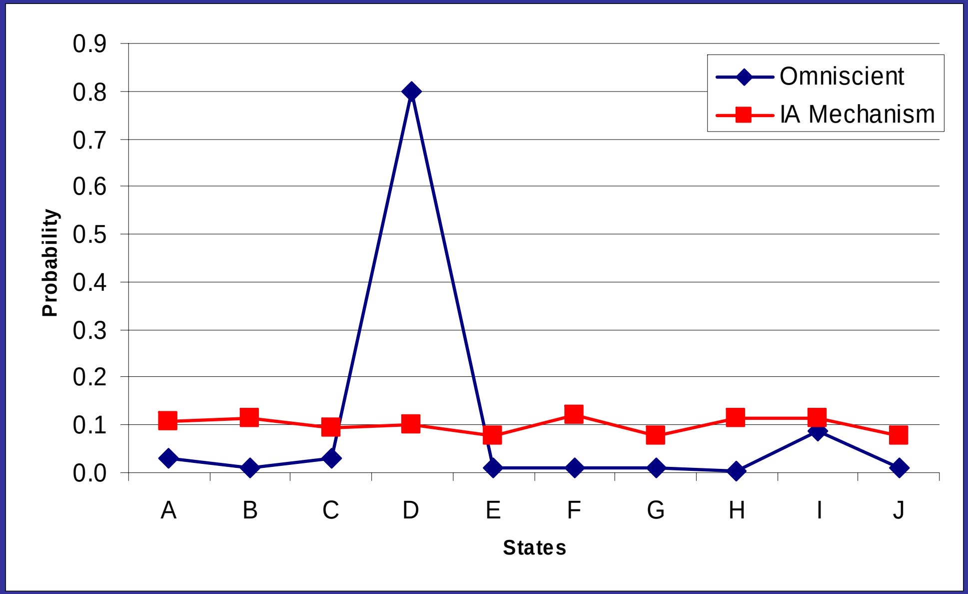


Experiment 4, Period 17
1 Player

results

comparison to omniscient probability

Kullback-Leibler = 1.448

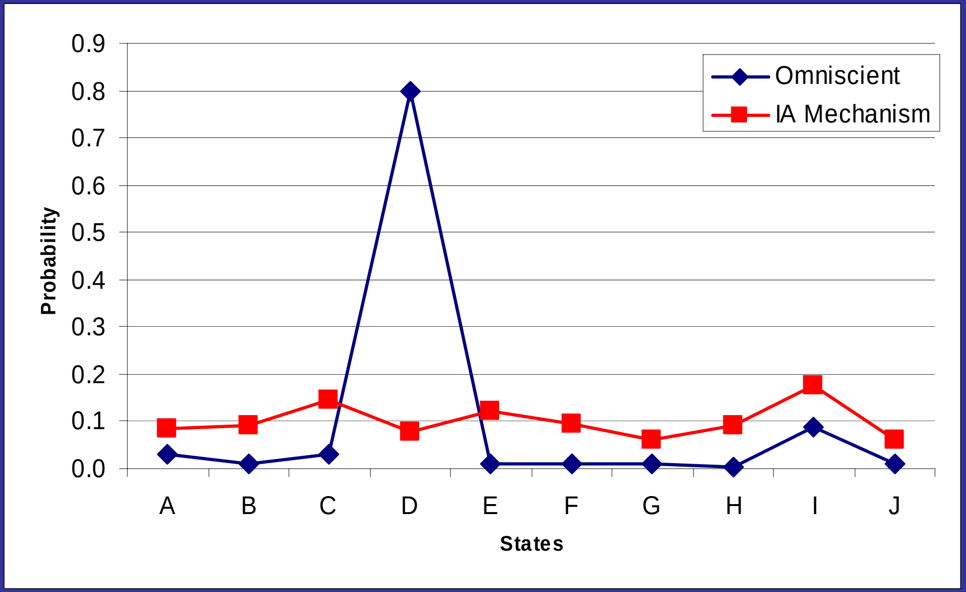


Experiment 4, Period 17
2 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.606

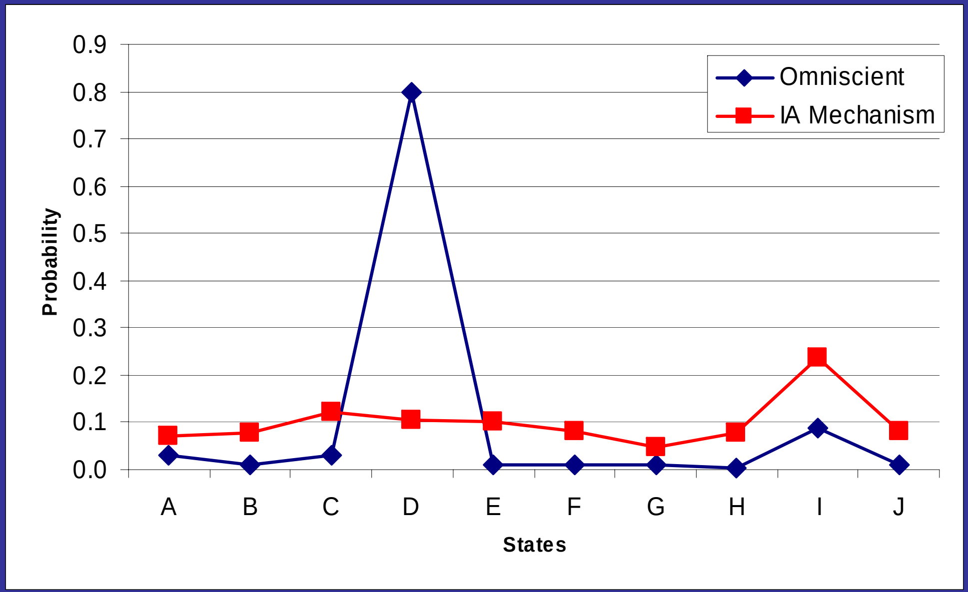


Experiment 4, Period 17
3 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.362

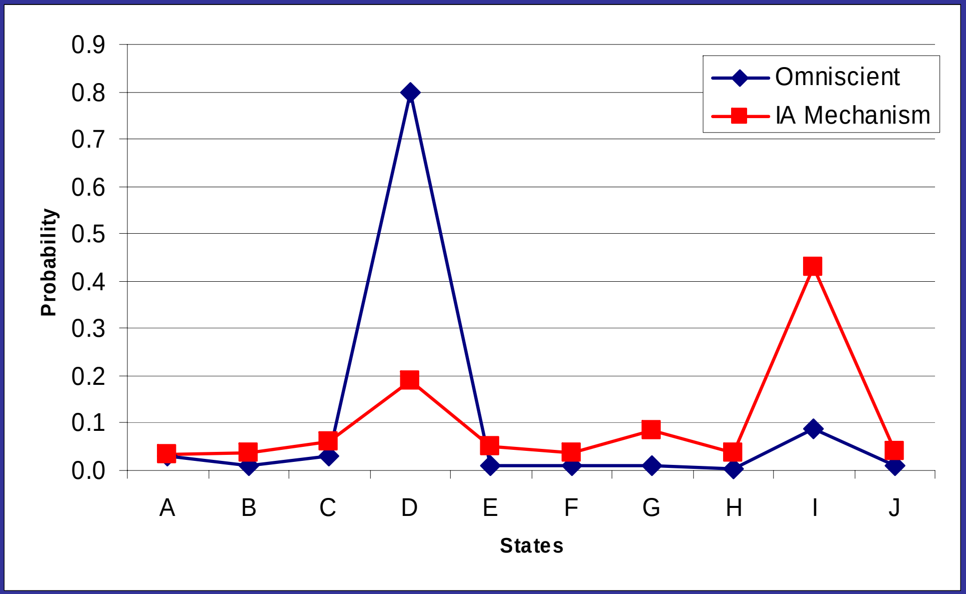


Experiment 4, Period 17
4 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.905

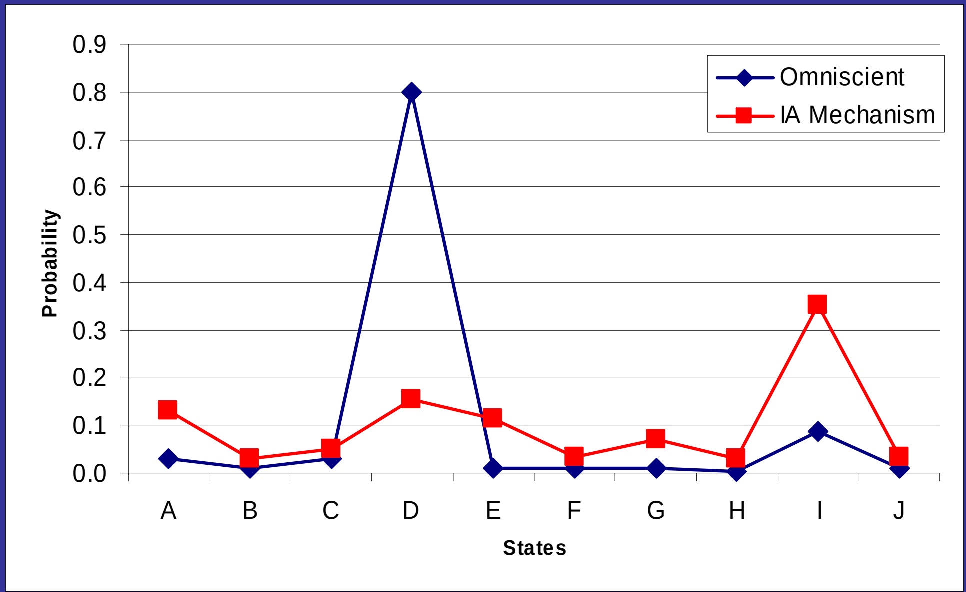


Experiment 4, Period 17
5 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.042

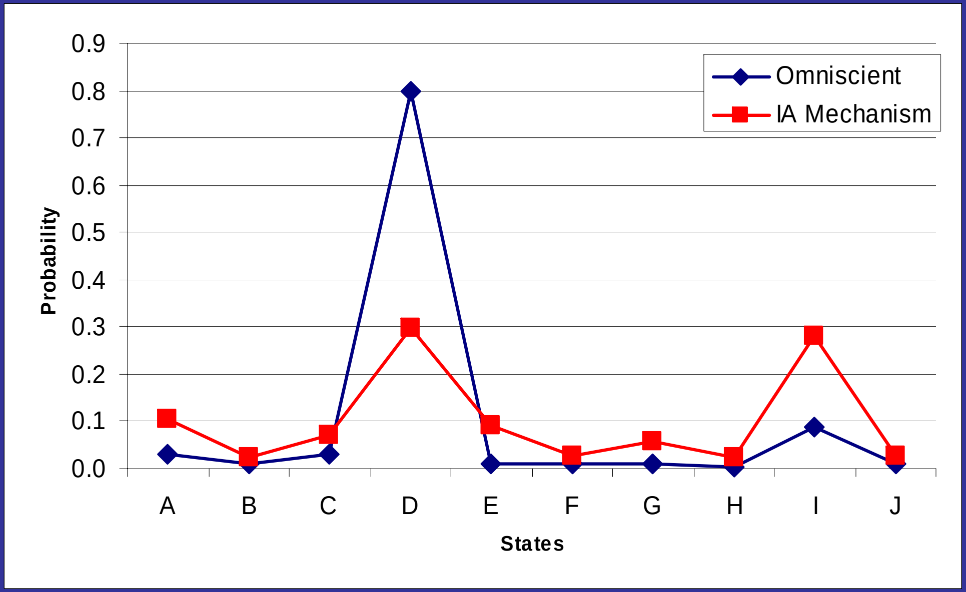


Experiment 4, Period 17
6 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.550

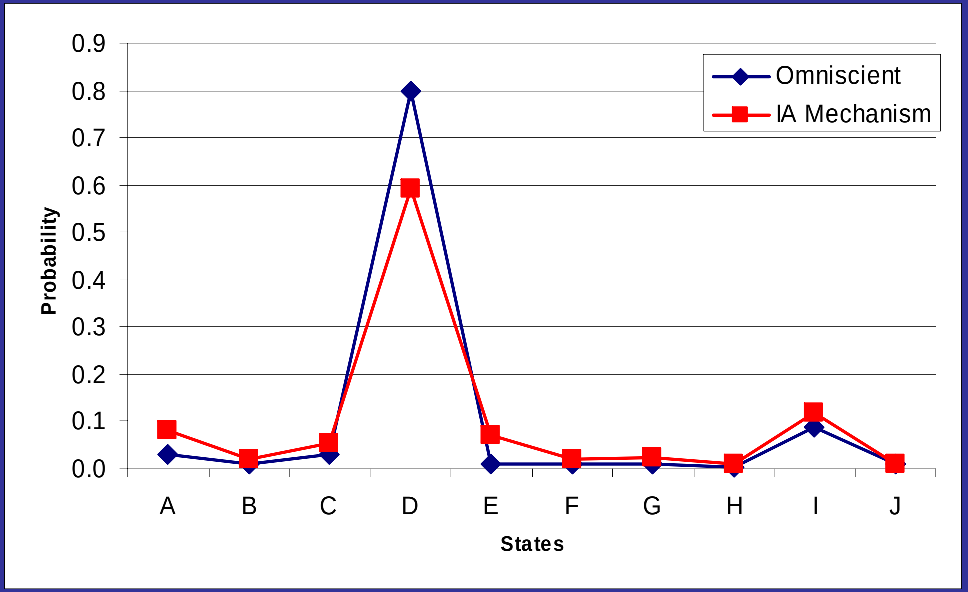


Experiment 4, Period 17
7 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.120

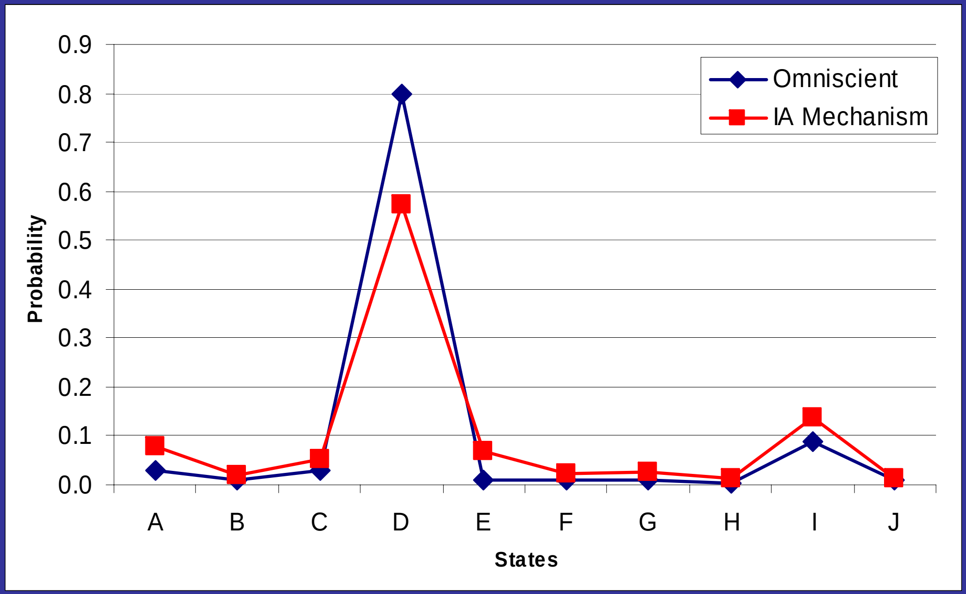


Experiment 4, Period 17
8 Players Aggregated

results

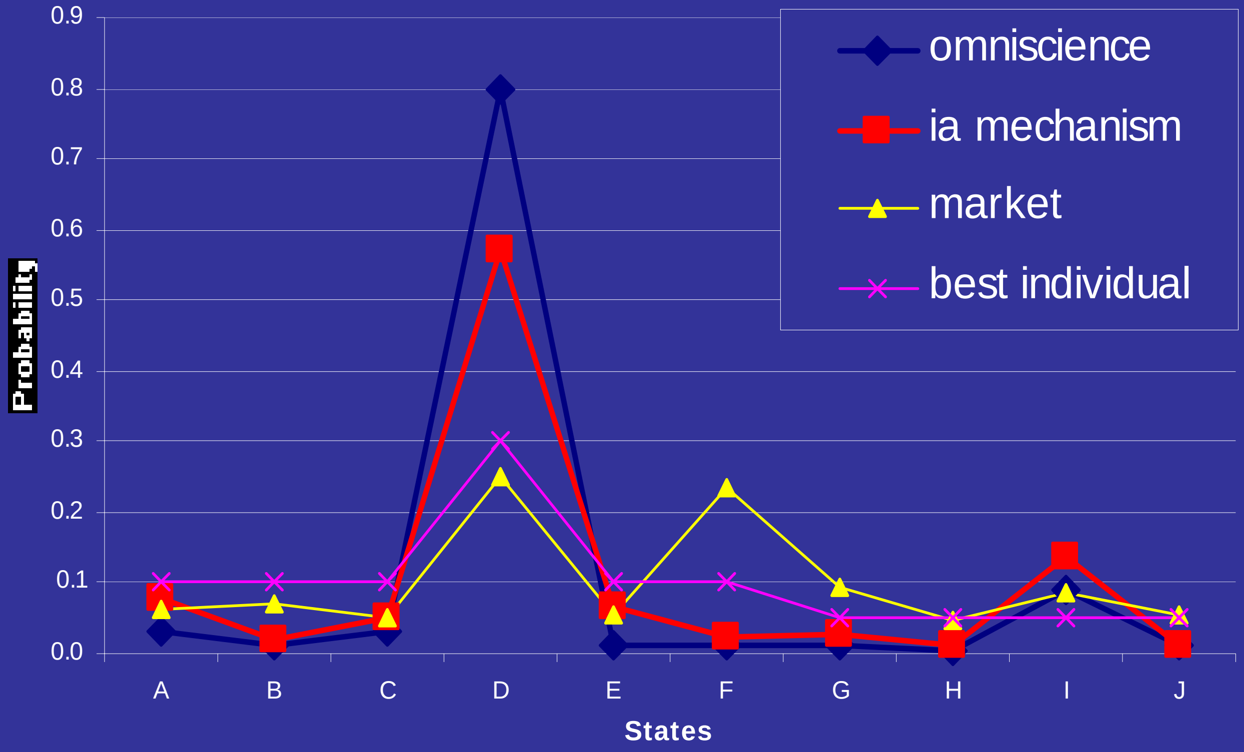
comparison to ominiscient probability

Kullback-Leibler = 0.133



Experiment 4, Period 17
9 Players Aggregated

overall performance



better than the best!

predicting in the real world

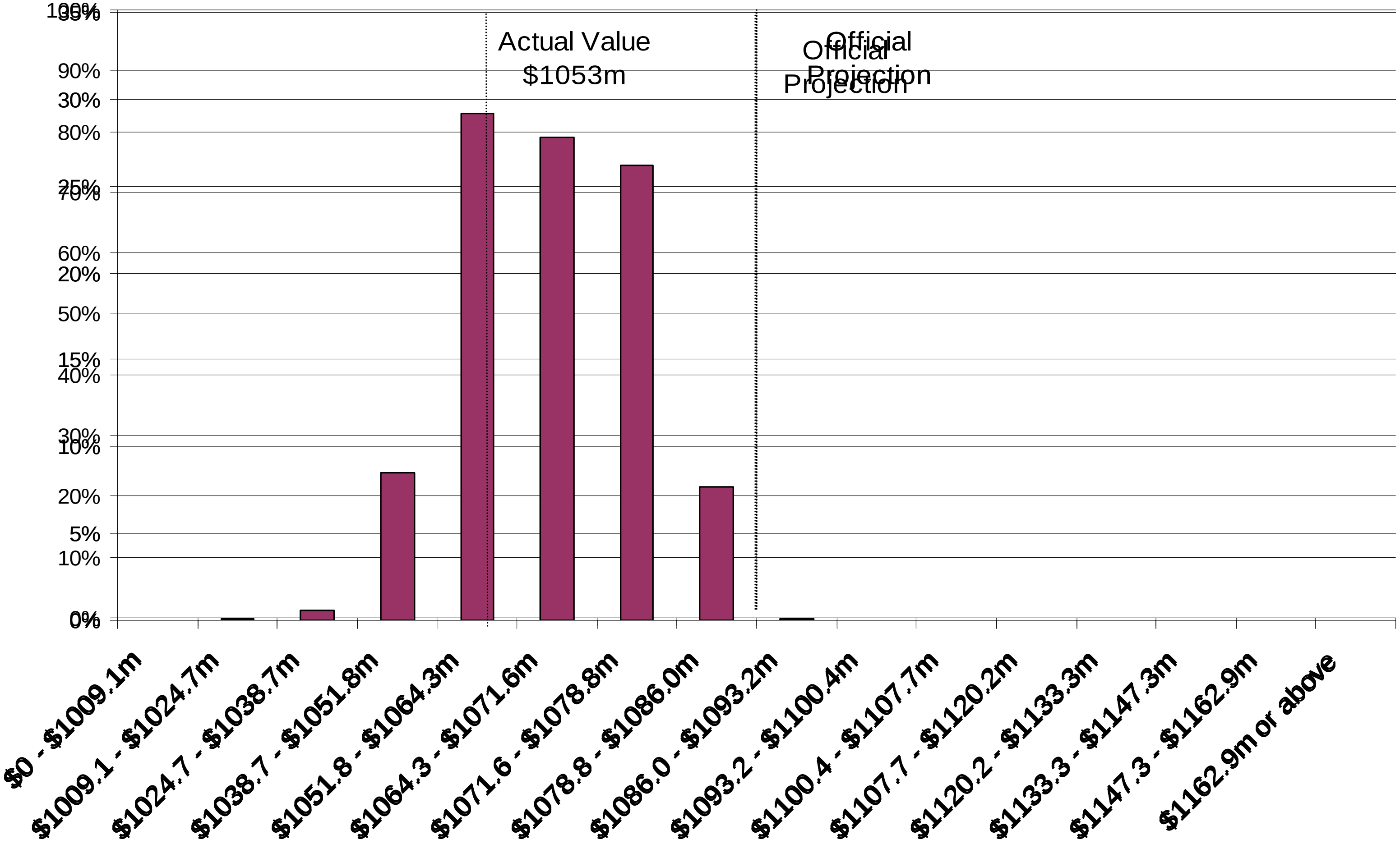
(as opposed to the laboratory)

we ran a pilot test with one of hp divisions

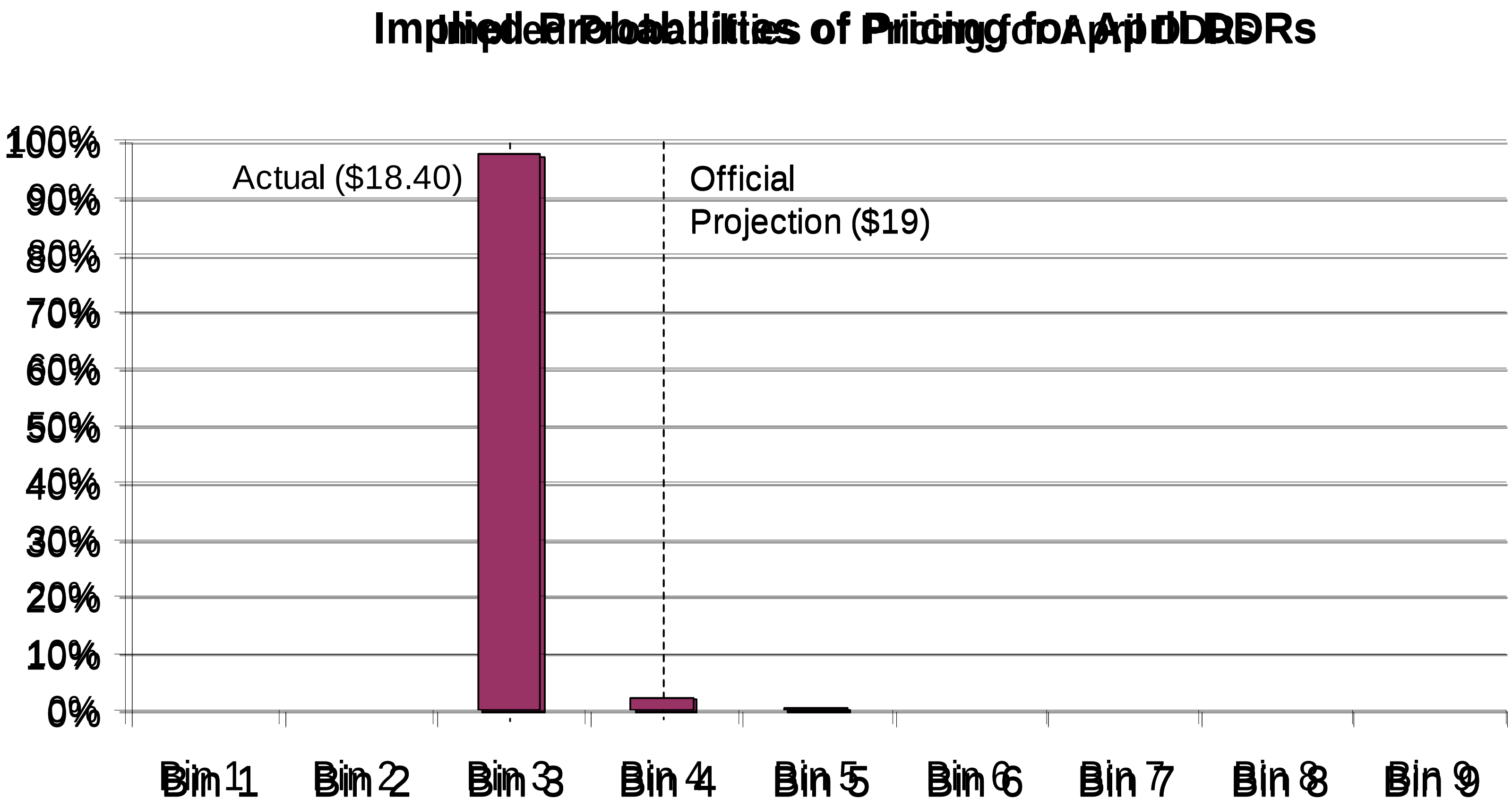
15 managers distributed worldwide

goal: to predict monthly revenues and profits

Implied Probabilities of Revenue Bins, September 2003



one more case: future component prices



it is all about the power of the implicit

for more information go to:

http://www.hpl.hp.com/research/idl