Harvesting Social Knowledge

Bernardo A. Huberman

Information Dynamics Laboratory
HP Labs

motivation

a key differentiator of great organizations is their ability to extract, aggregate, analyze, and properly act on information quickly

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)

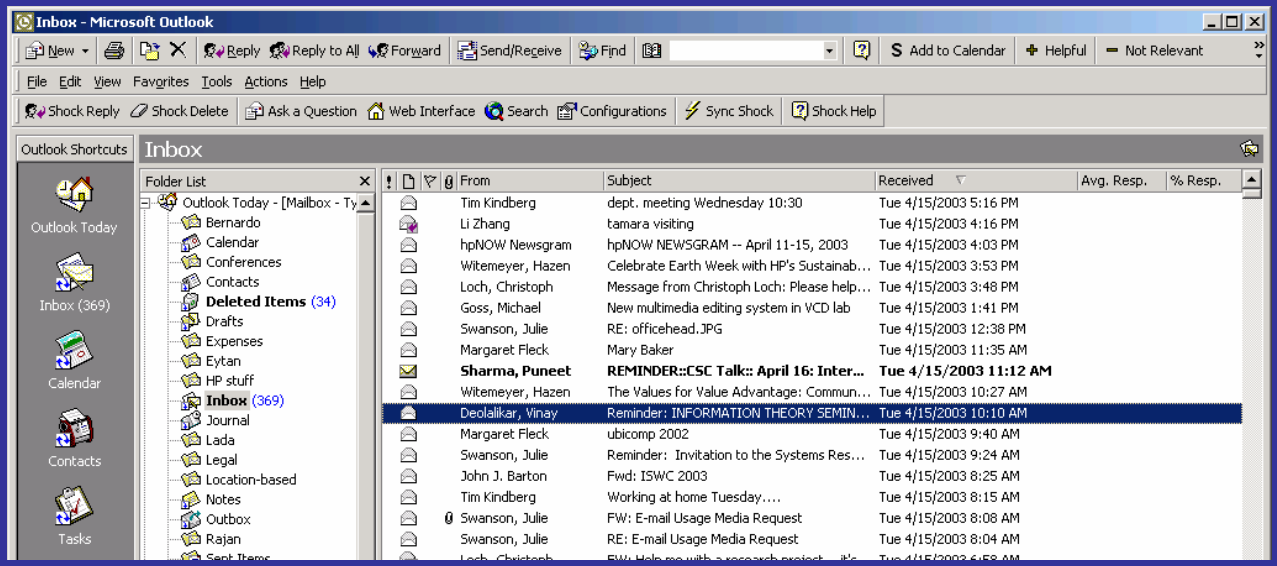
- how information flows within organizations (allen)

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

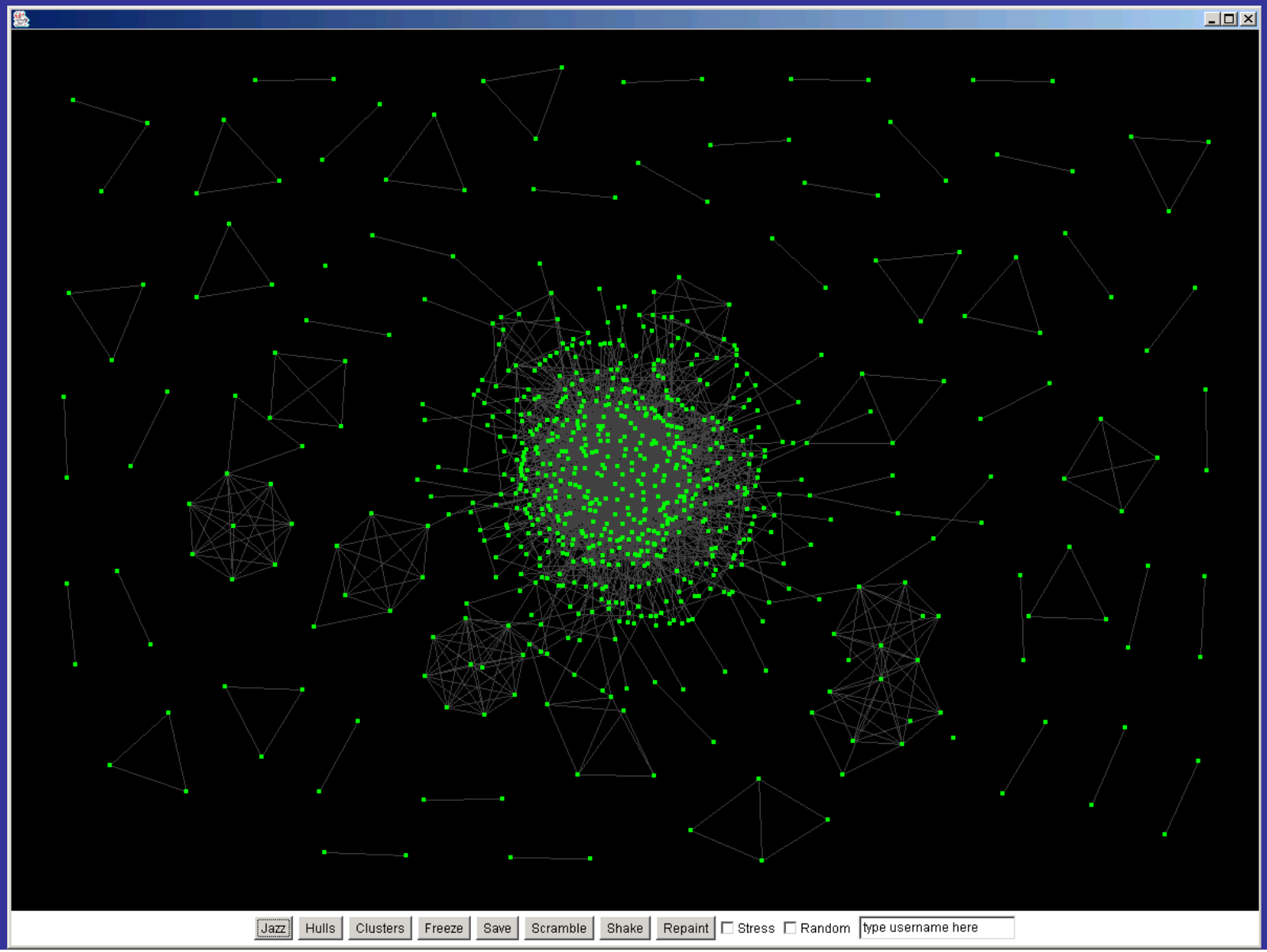
uncovering communities with e-mail

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

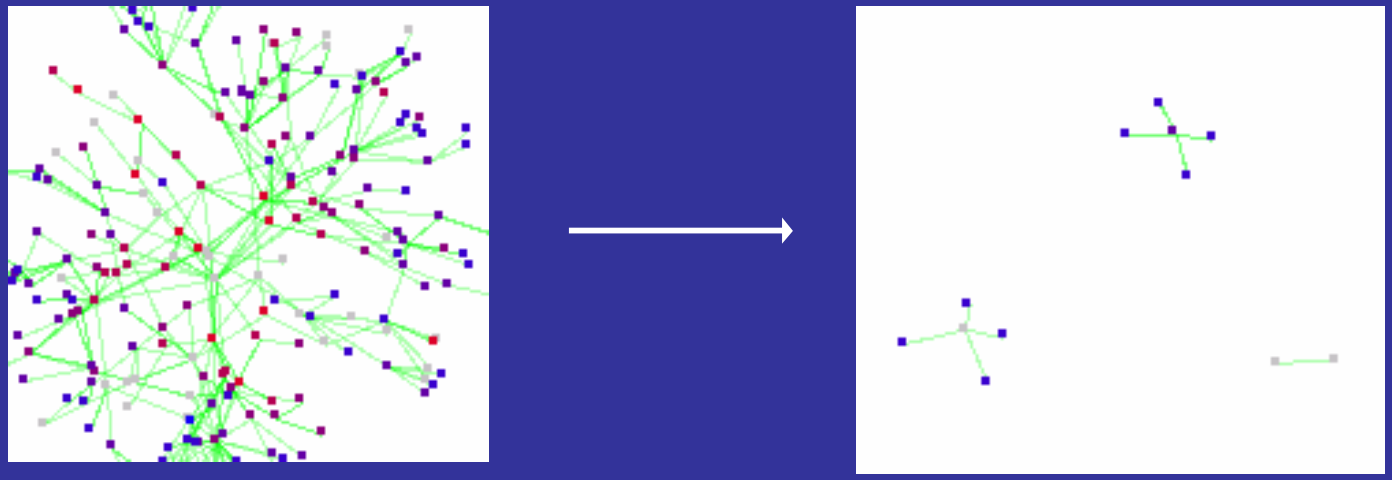


hp labs email network



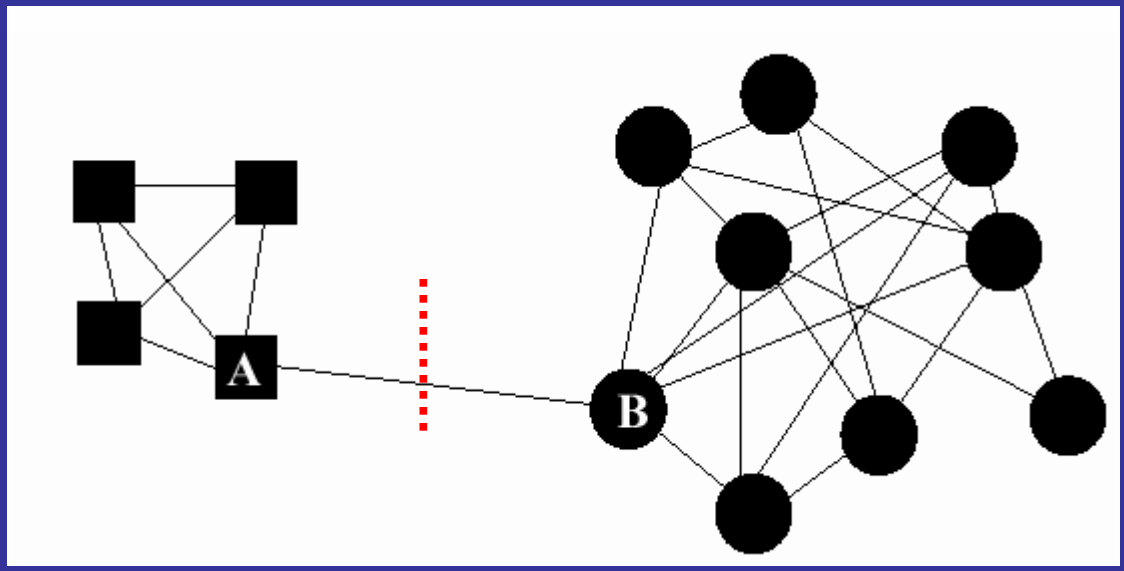
our goal

- 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

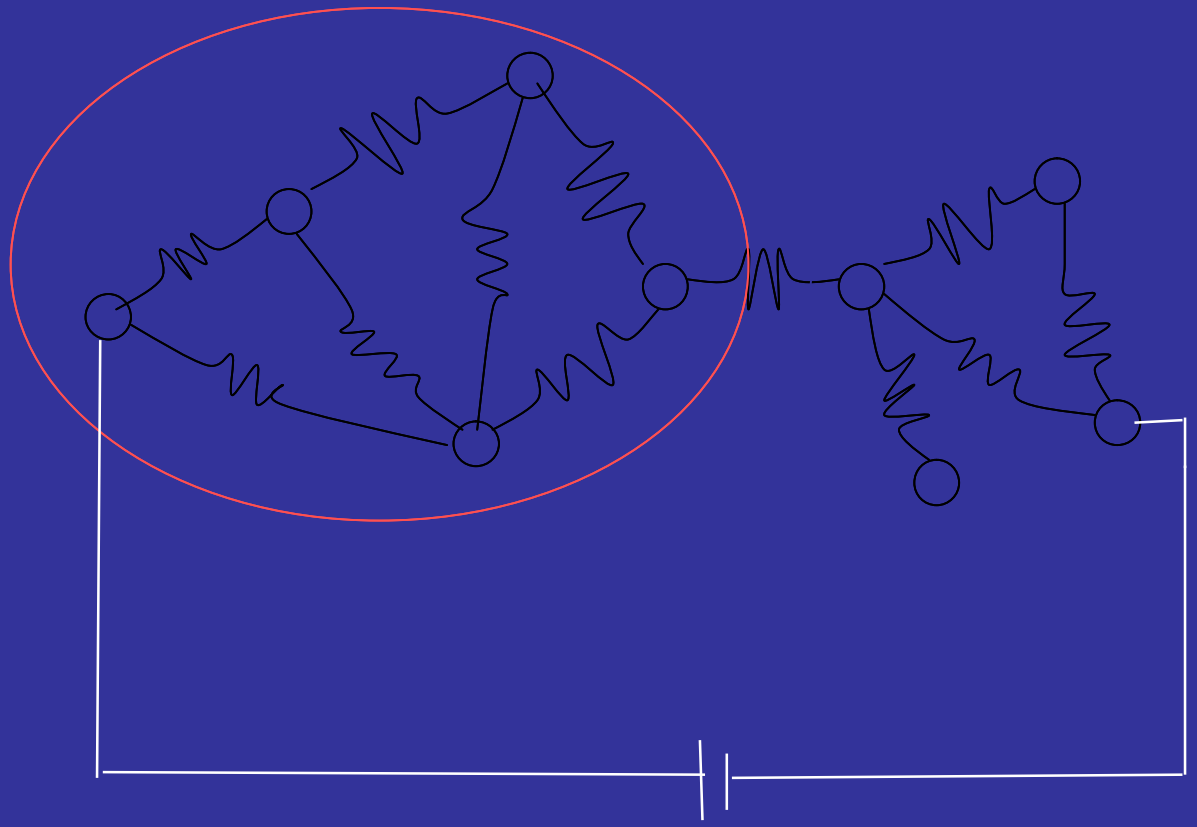
a problem

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

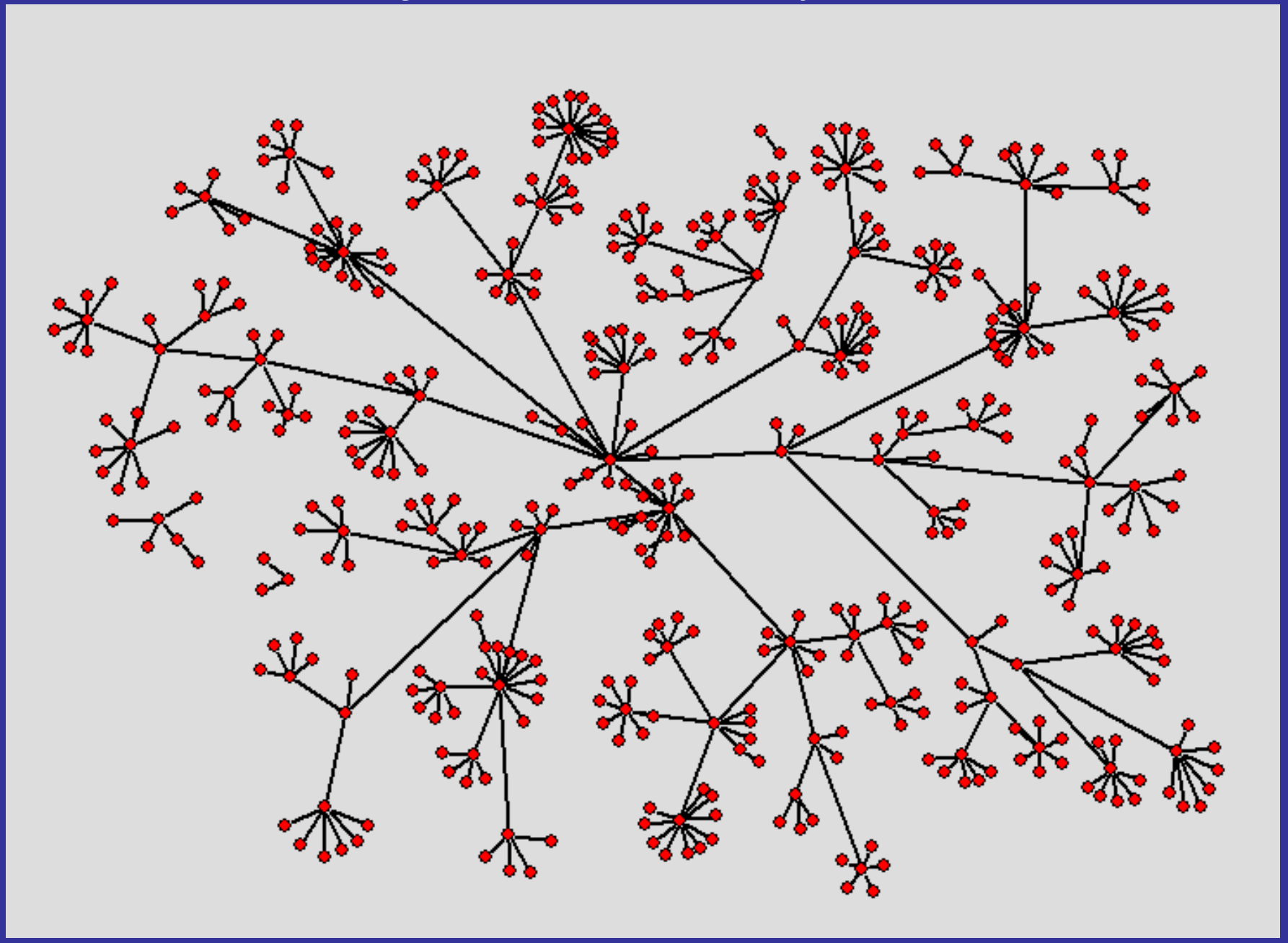
penrose	Mobile & Media Systems Lab
mistyr	HPL Advanced Studies
vinayd	HPL Advanced Studies
seroussi	HPL Advanced Studies
tsachyw	HPL Advanced Studies

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

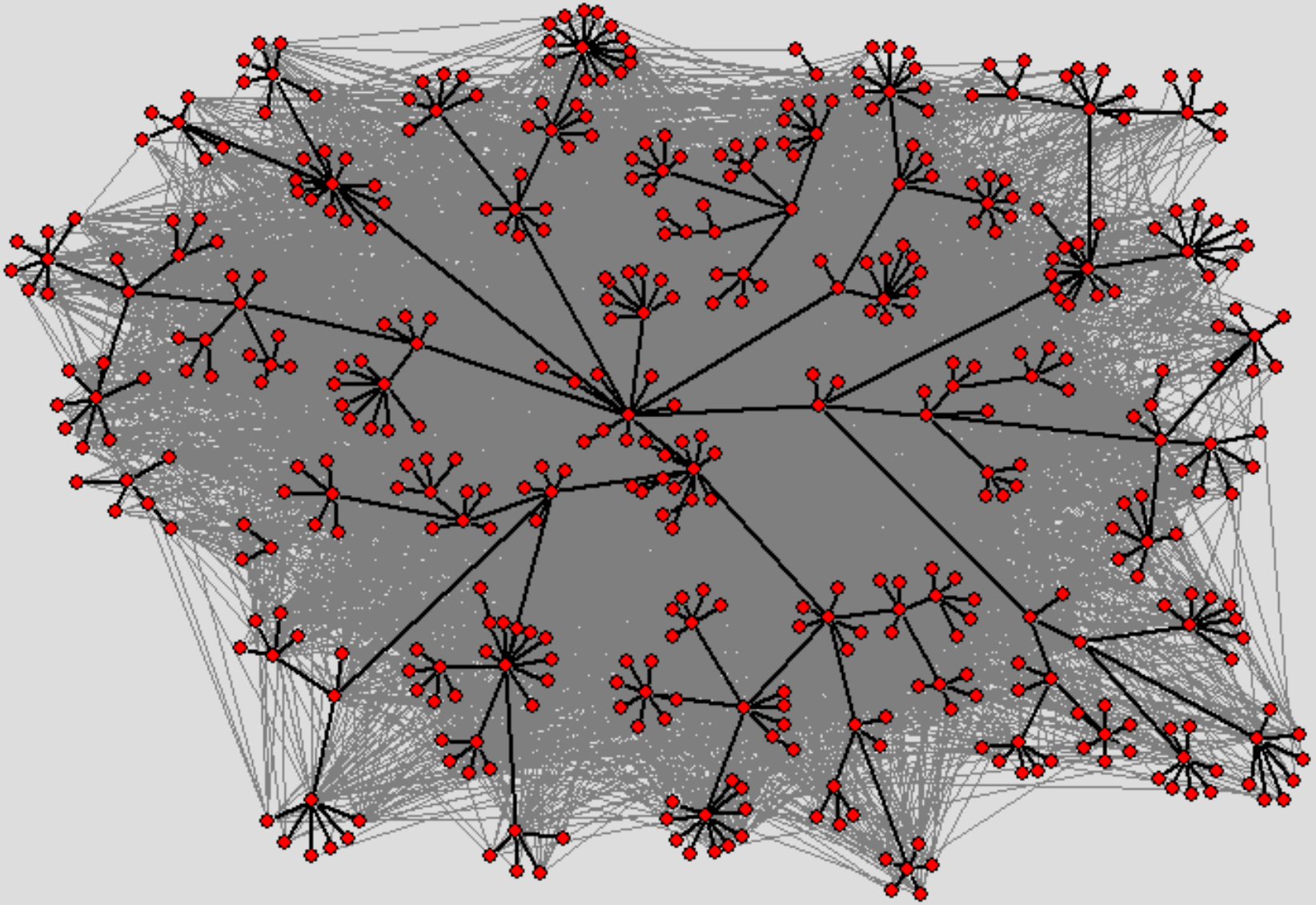
venky	Mobile & Media Systems Lab
dohlberg	HPL Advanced Studies
kvincent	Hardcopy Tech Lab
pmcc	University Relations
trangvu	HPL Communications
markstei	HPL Advanced Studies
hollerb	HPL Research Operations
krishnav	Handheld HQ
babcock	REWS Americas
gita	Solutions & Services Tech Cntr
bgee	HPL - Research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

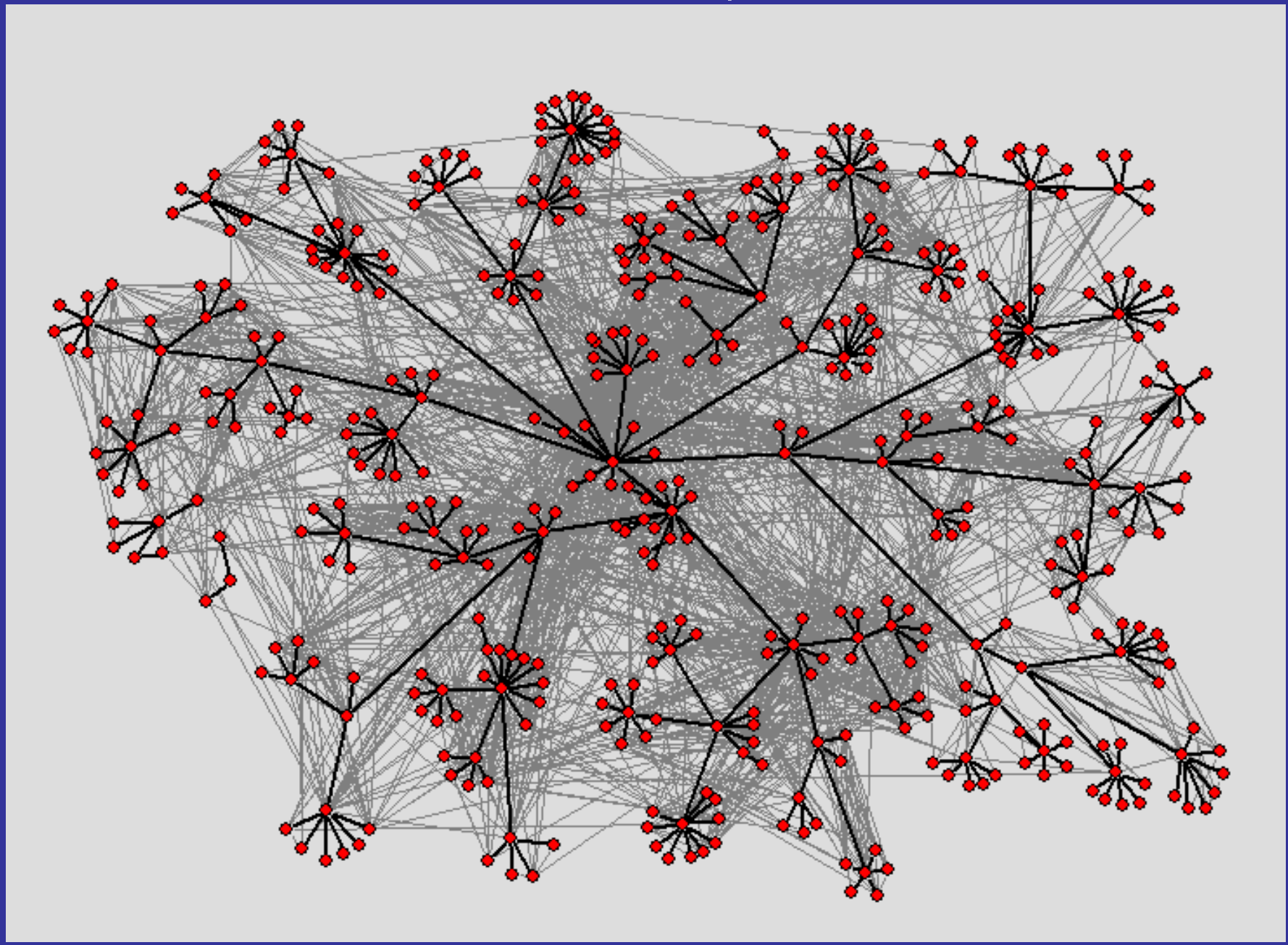
organizational hierarchy



email correspondents scrambled

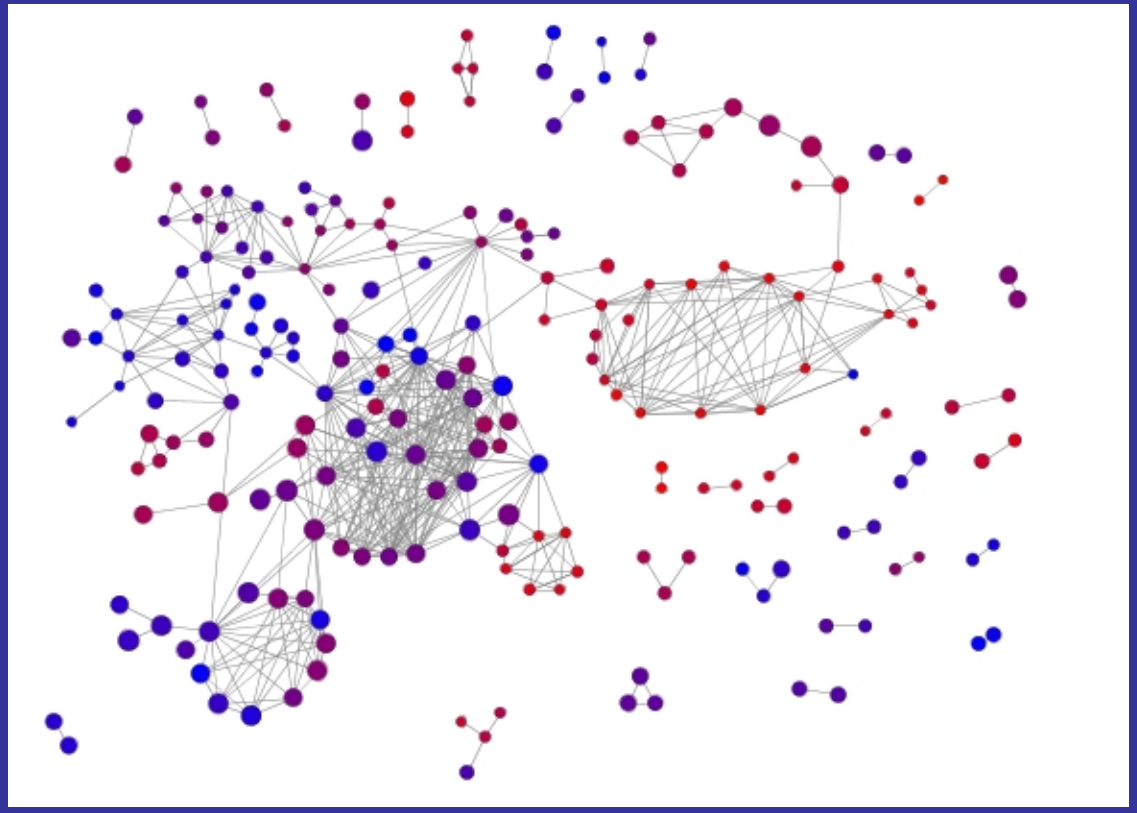


actual email correspondence



document similarity by usage

similarity: overlap in users accessing documents

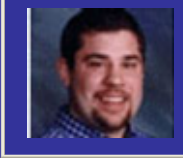


earlier documents are blue, later ones are red.
size of node reflects the number of users accessing the document.

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 Cen tral 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	**John R Bugarin**	SAM	AMCI	Solution Architech	Systems Integration	30
0.35		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 fo r .N MCAD for .NET, MCSD for Visual Studio 6.0, MCSE for Windows 2000, and MCDBA for MSSQL 20				
DOCS	**Tom Kern**	SAM	AMCI	Tech Consulting	Systems Integration	236
0.29		Tom Kern is a consultant for the Enterprise Microsoft Services .Net Solutions practice. Tom has wor on a variety of custom software projects based on Microsoft technologies.				
DOCS	**Martyn Dowsett**	SEM	EMCI	Tech Consulting	Systems Integration	46
0.26		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 ro und all things .Net.				

a new people finder

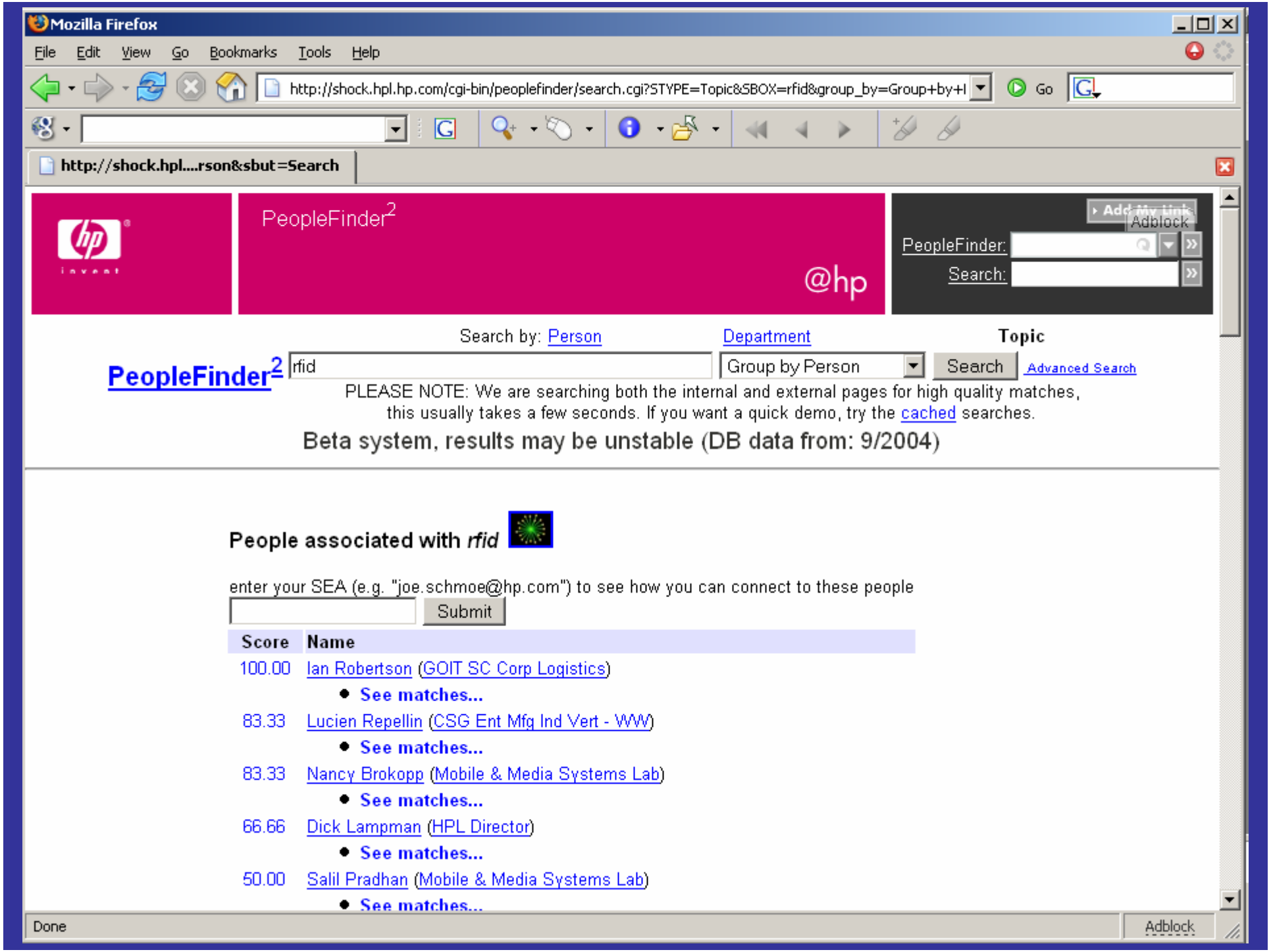
there is a trove of information in power point presentations, public repositories within the organization, and the internal website of the enterprise

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

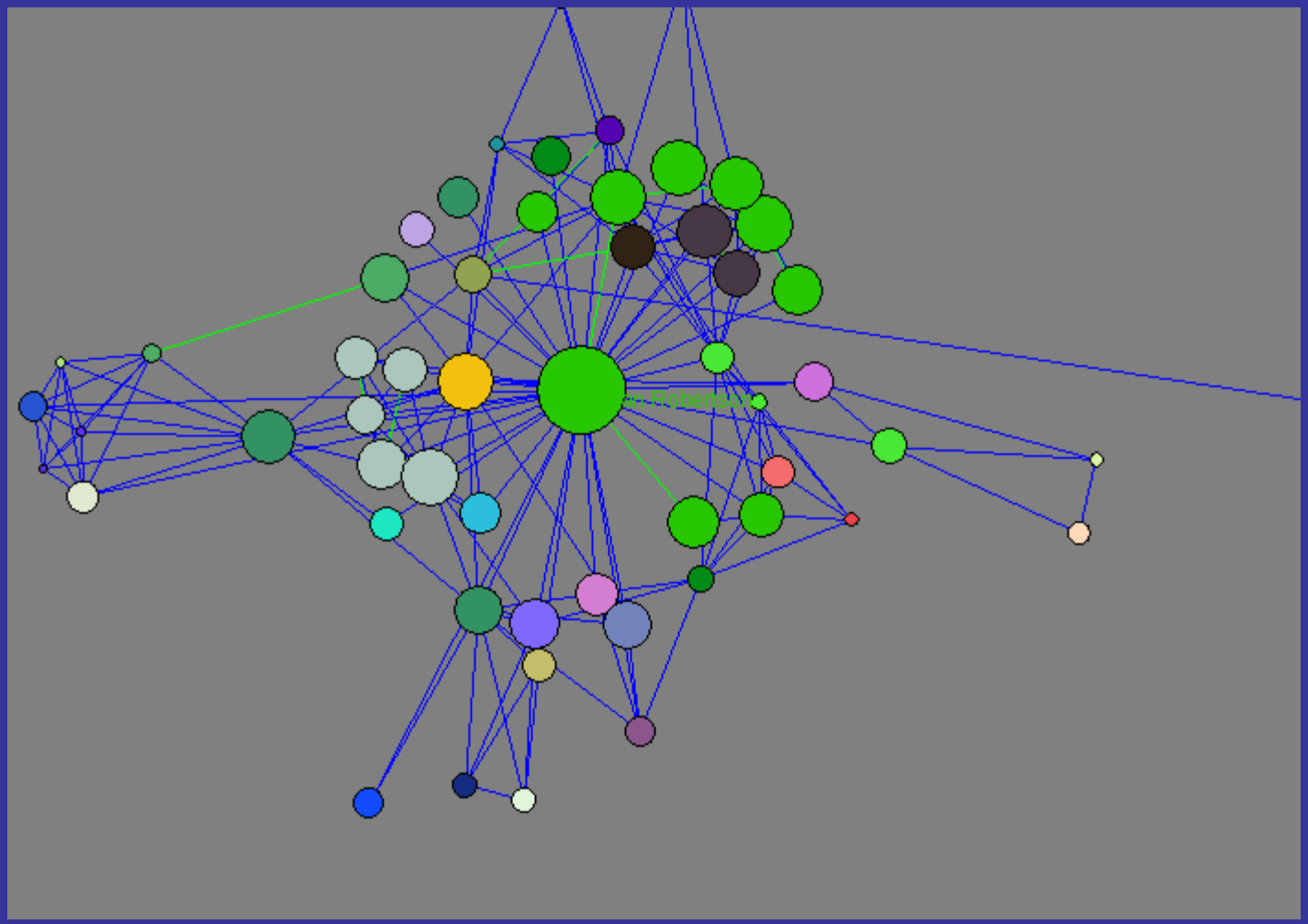
it also discovers who is working on what

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
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

related individuals to ian robertson



information flow

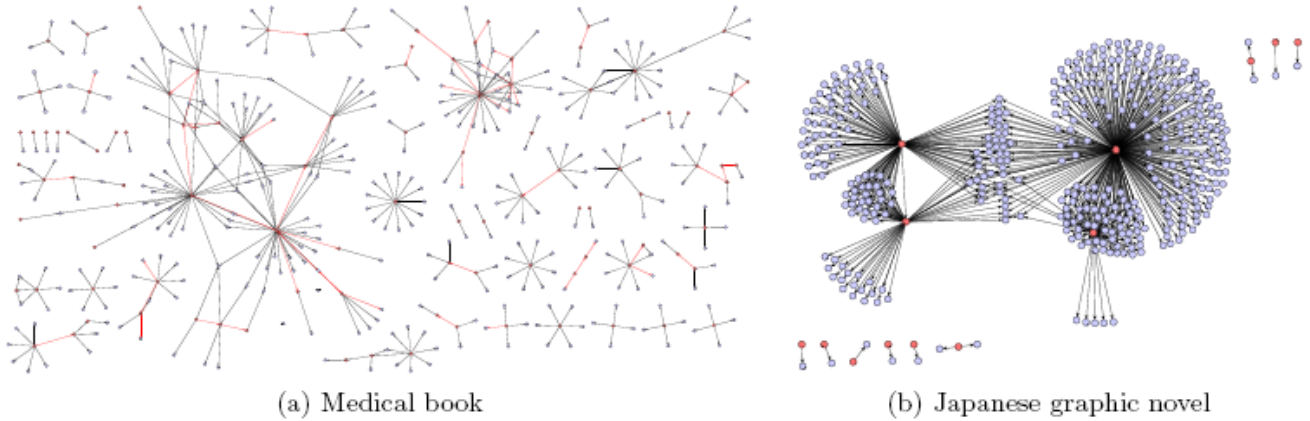
how does information flow in a community or organization?

does the structure of the social network affect it?

how far does it spread?

Wu, Adamic and Huberman

recommendation networks

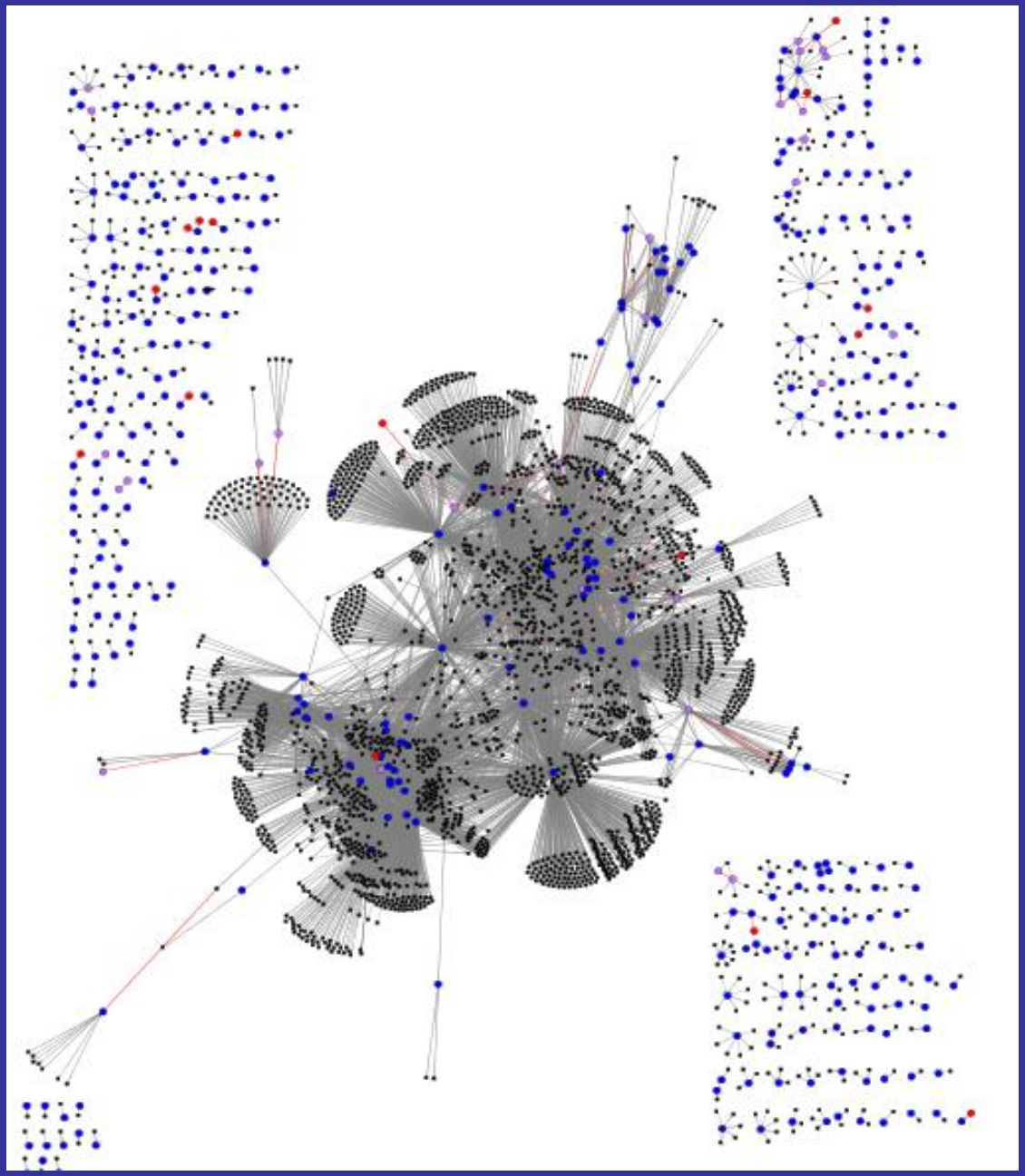


(a) Medical book (b) Japanese graphic novel

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



product recommendation network

- purchase following a recommendation

customer recommending a product

- customer not buying a recommended product

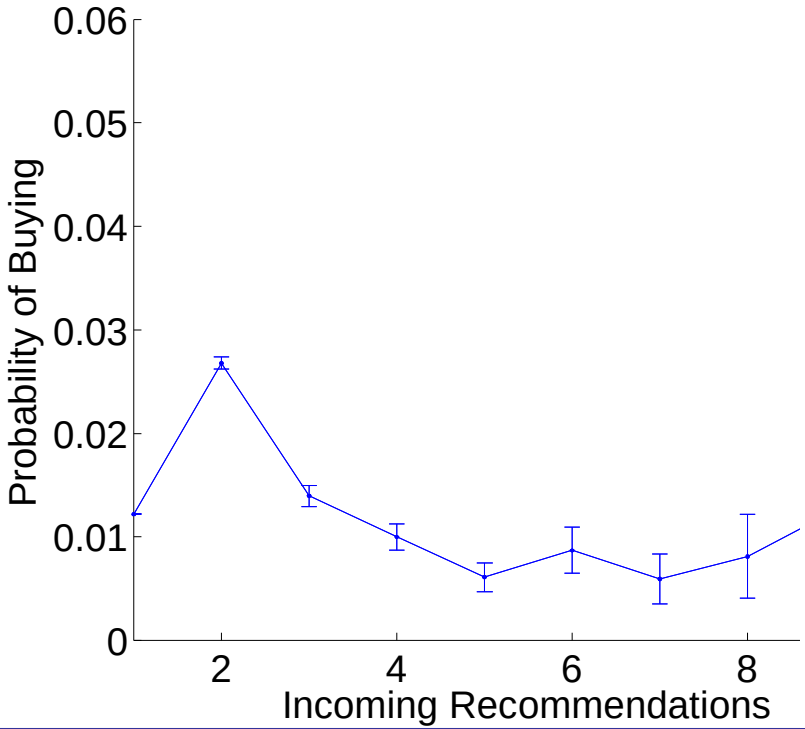
j. leskovec, l.adamic and b.a. huberman

observations on product groups

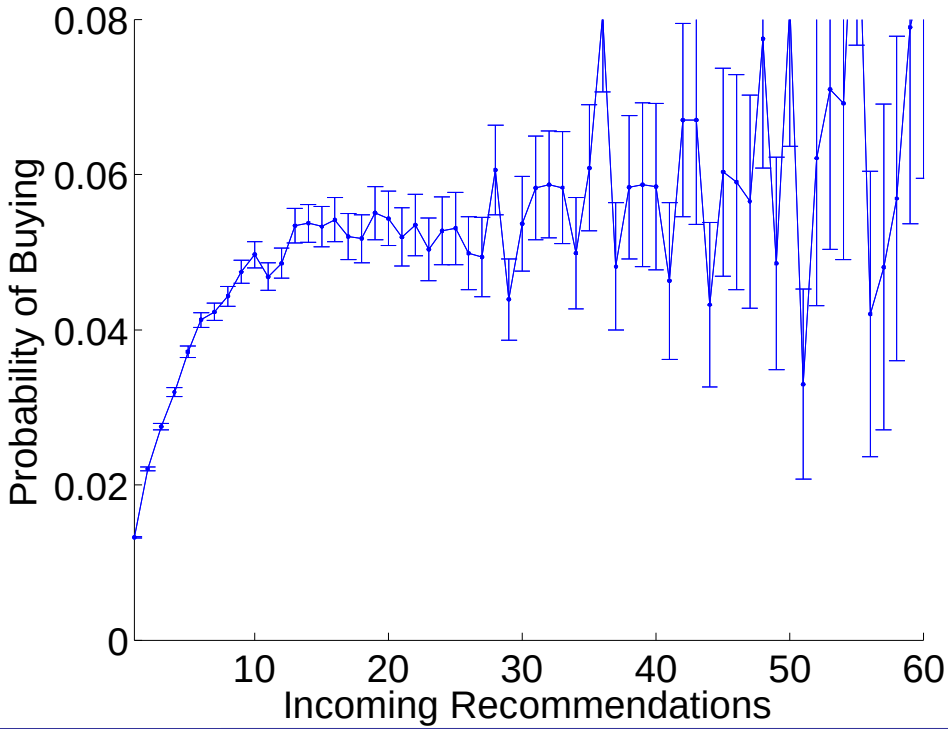
- there are relatively few DVD titles, but DVDs account for ~ 50% of recommendations.
- recommendations per person
  - DVD: 10
  - books and music: 2
  - VHS: 1
- recommendations per purchase
  - books: 69
  - DVDs: 108
  - music: 136
  - VHS: 203
- Overall there are 3.69 recommendations per node on 3.85 different products.
- music recommendations reached about the same number of people as DVDs but used only 1/5 as many recommendations
- book recommendations reached by far the most people – 2.8 million.
- networks are highly disconnected

does receiving more recommendations increase the likelihood of buying?

BOOKS



DVDs

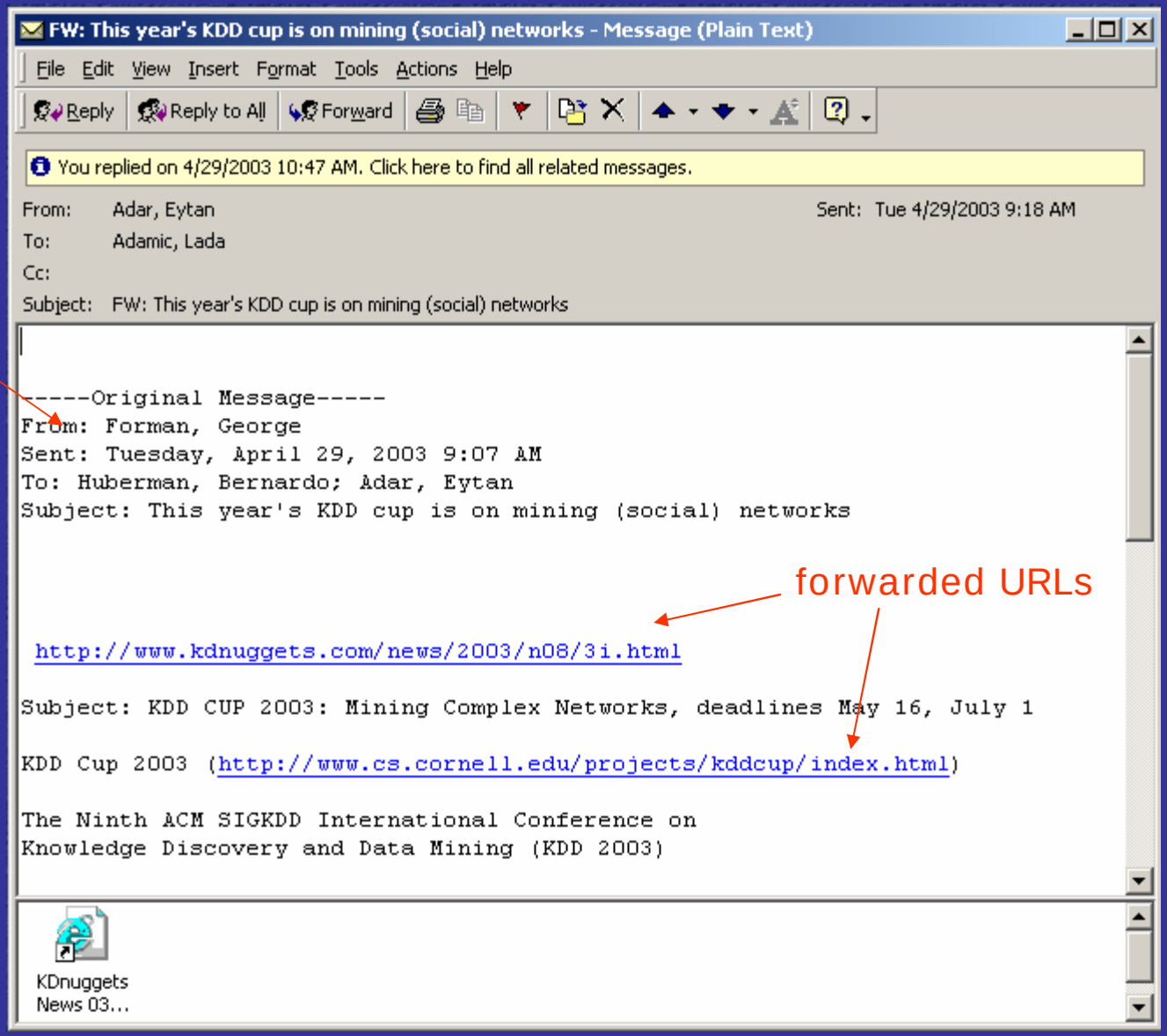


so, how effective is viral marketing?

- 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

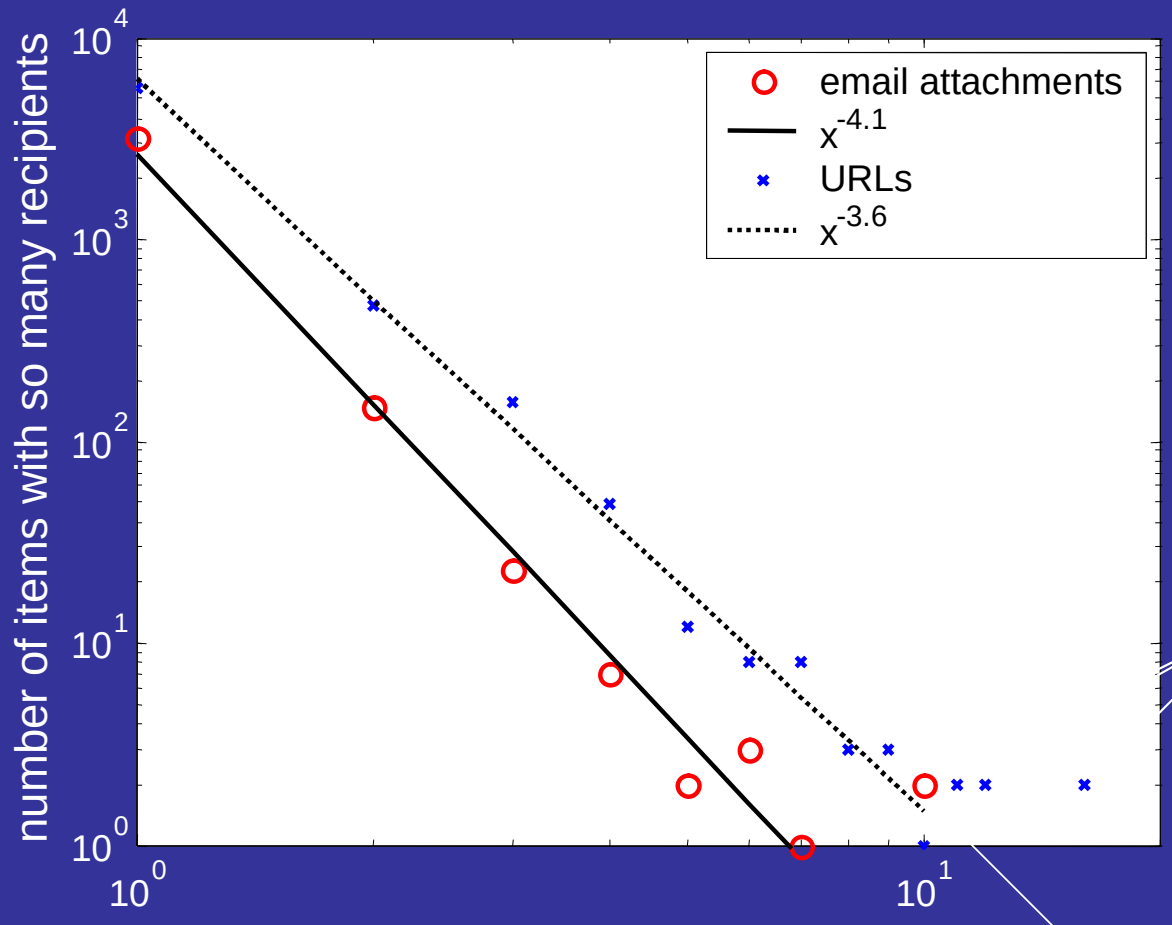
an experiment with forwarding messages

forwarded message



results

average = 1.1 for attachments, and 1.2 for URLs



ads at the bottom of hotmail & yahoo messages

short term expense control

collaborative tagging

a new, fast growing trend

in science: nature's *connotea*, *citeulike*

social: *delicious*, yahoo's *myweb*

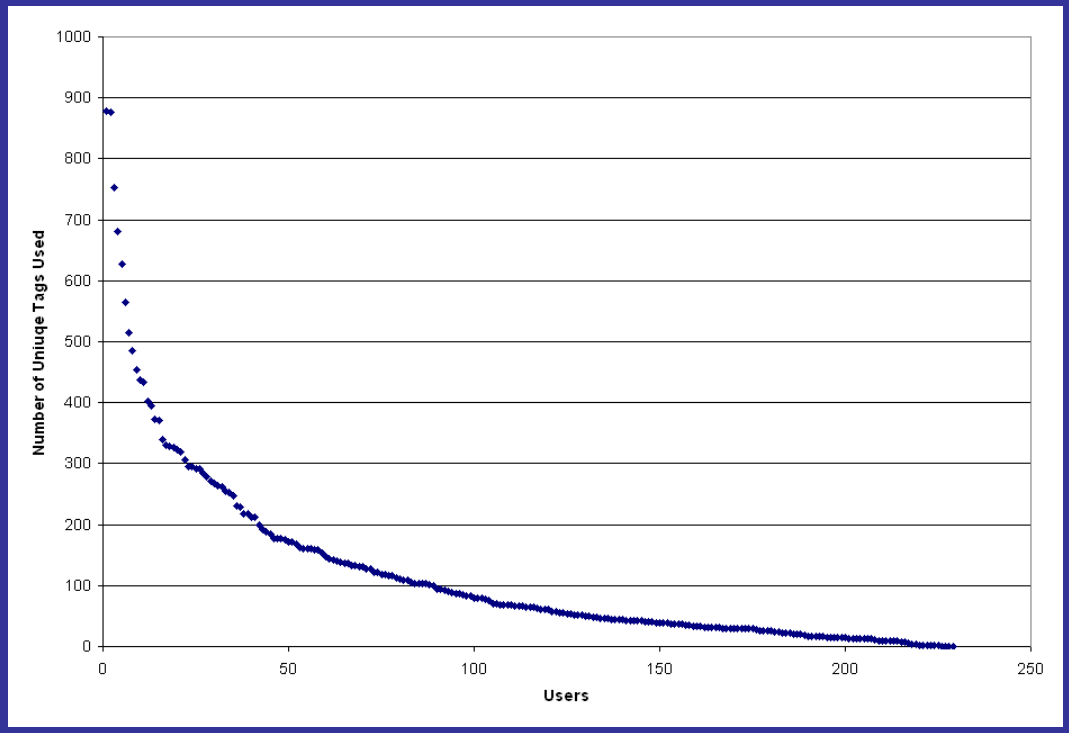
and not collaborative but tagging nonetheless: *flickr*, *technorati*

delicious dynamics

a social tagging system: not only can one see one’s own bookmarks, one can also see all of every other user’s bookmarks.

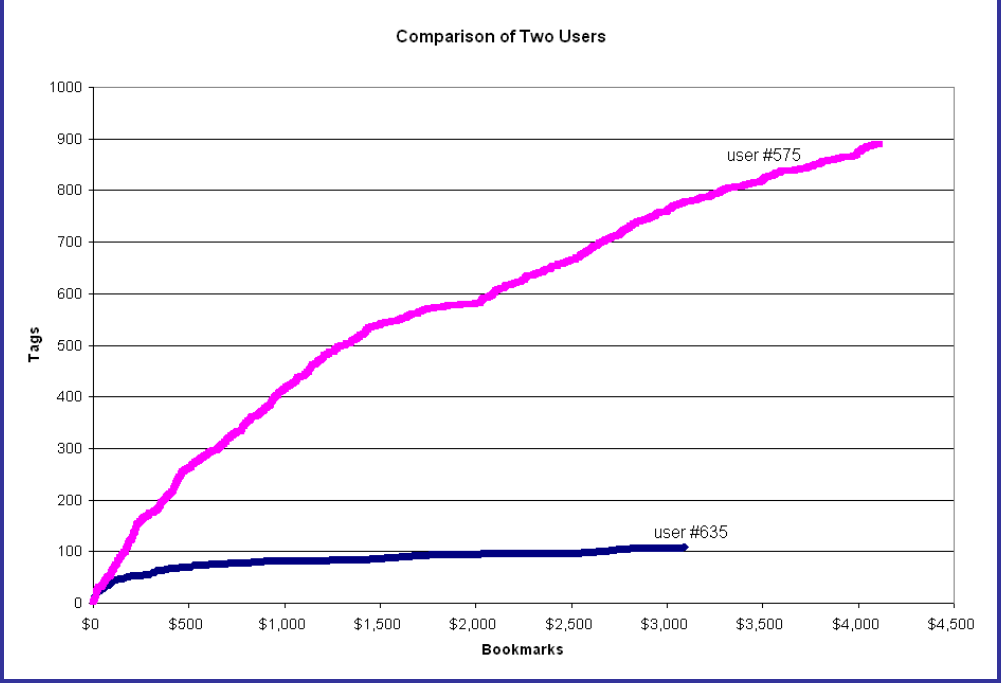
s. golder and b. huberman

number of tags in each user's tag list, in decreasing order.

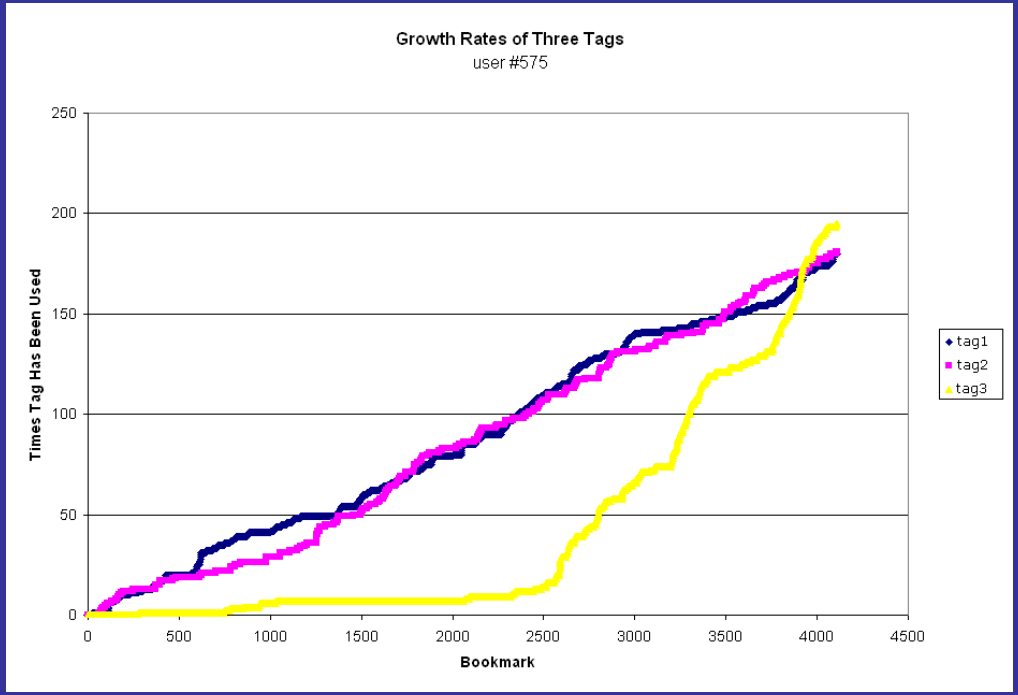


power law graph

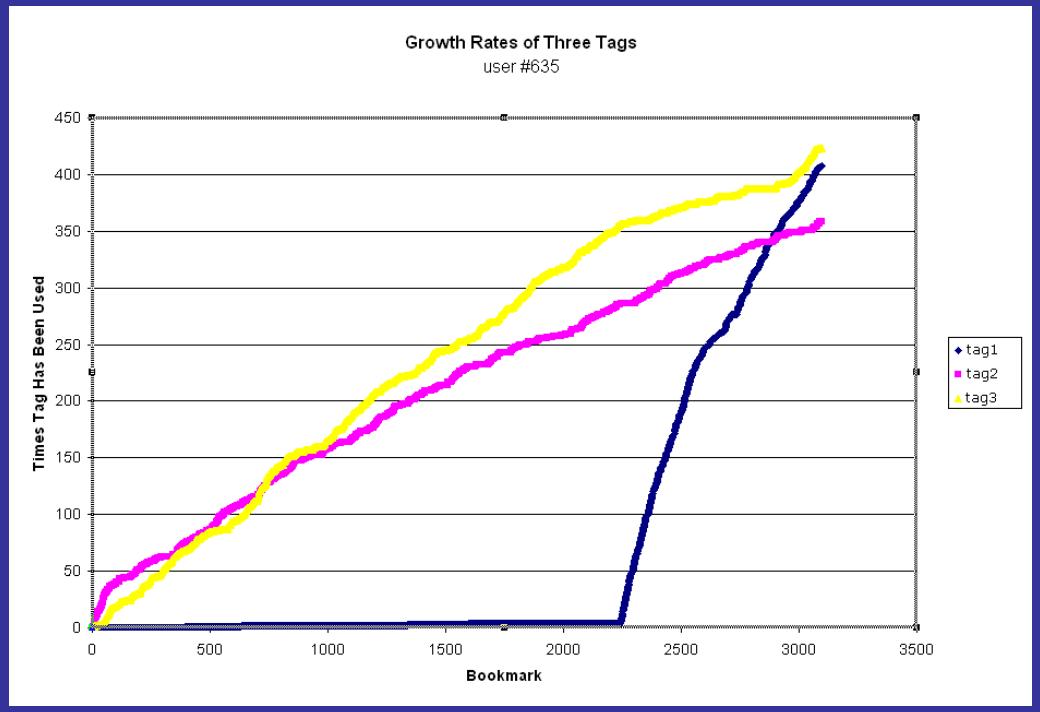
user behavior



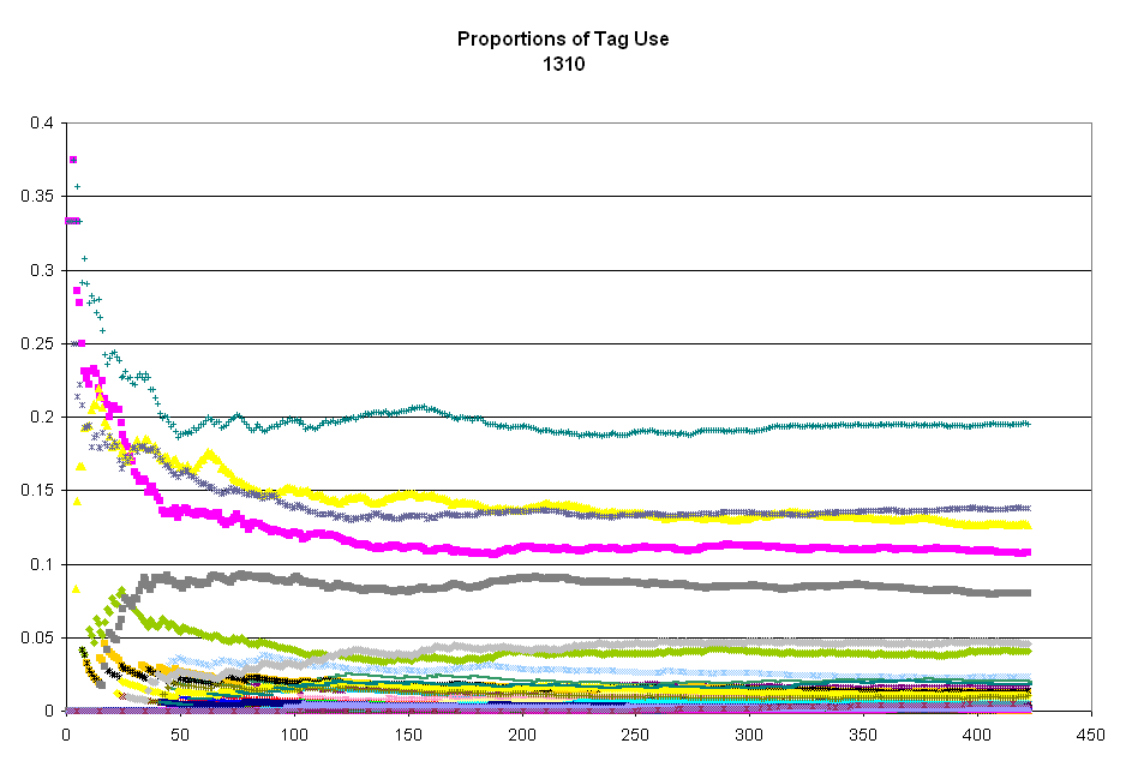
user behavior



user behavior

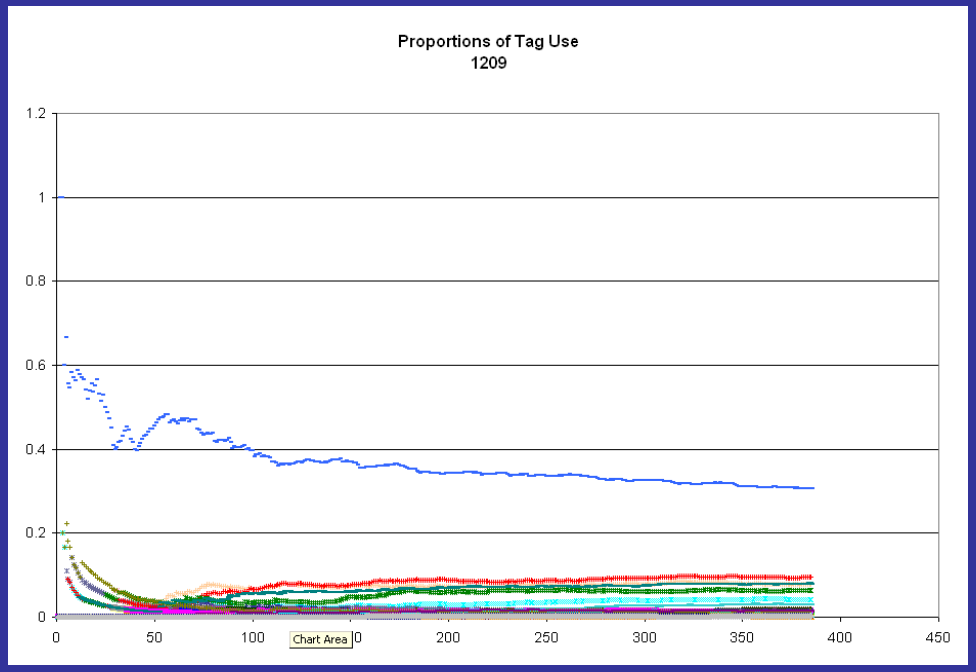


patterns in tagging dynamics



relative fractions of tags for a given, fixed, URL (vertical axis) as a function of time (horizontal axis) as measured in units of bookmarks added

patterns in tagging dynamics

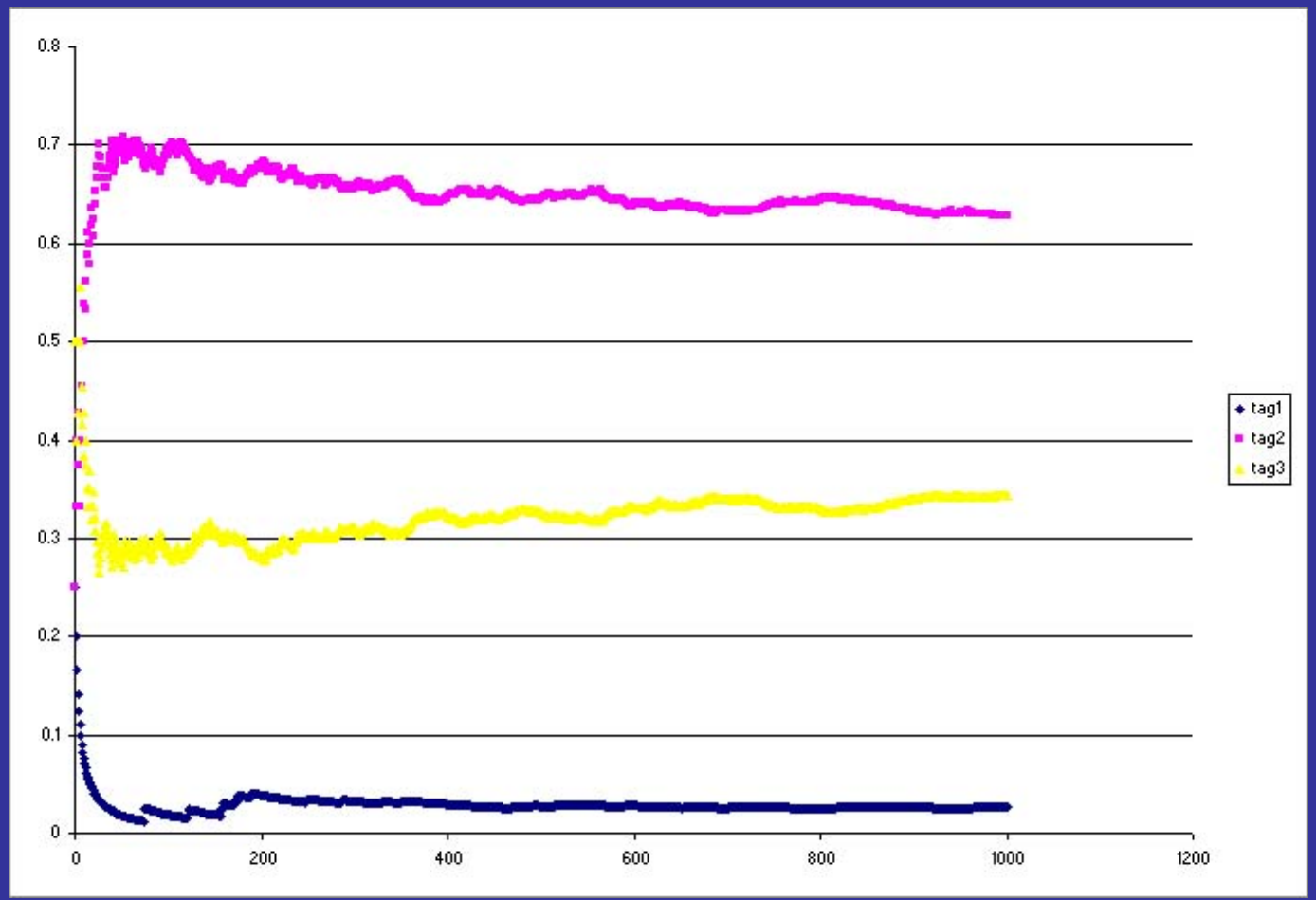


tagging dynamics

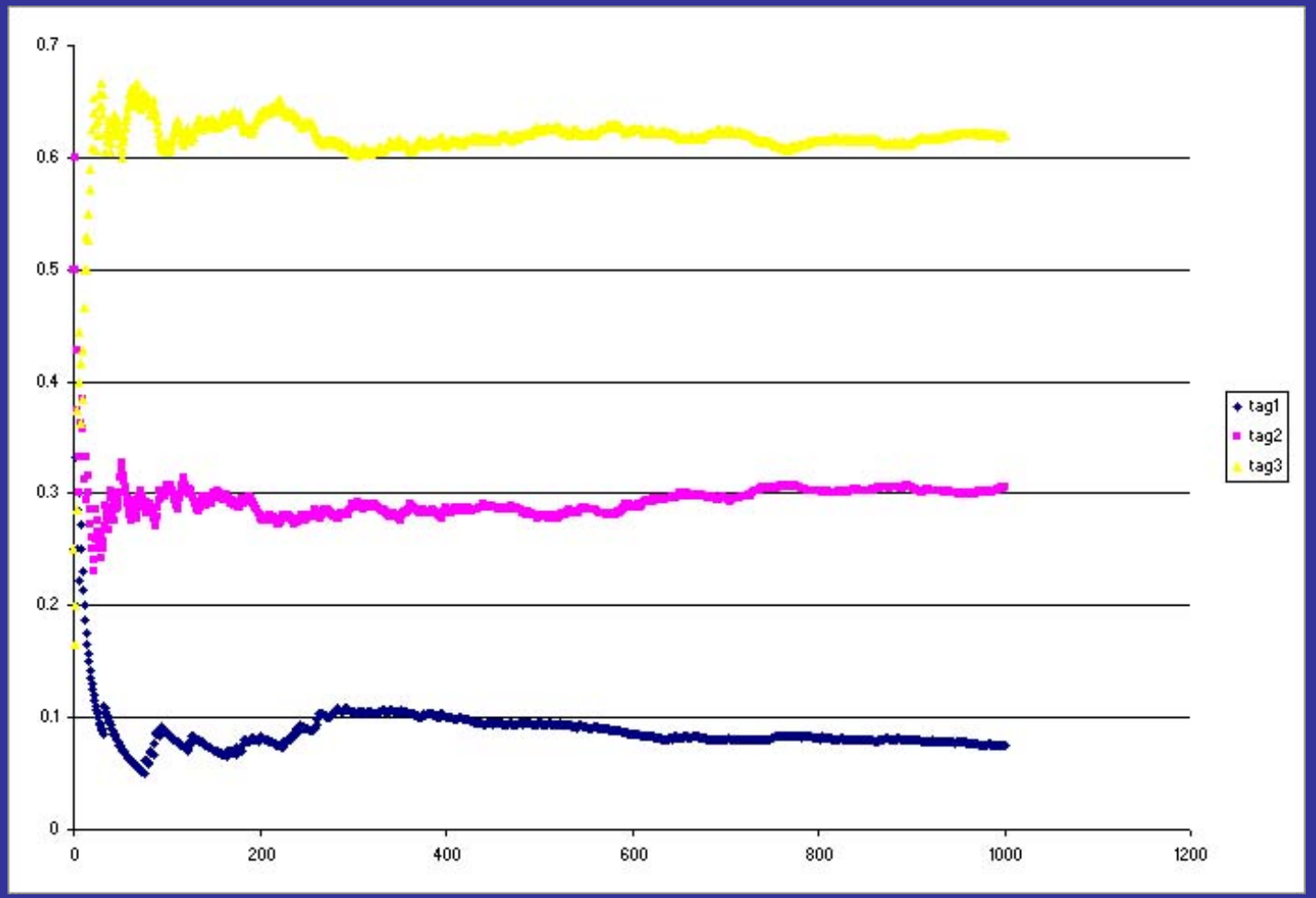
after the first 100 or so bookmarks, each tag's frequency is a nearly fixed proportion of the total frequency of all tags used

a nascent consensus seems to form, not affected by the addition of further tags.

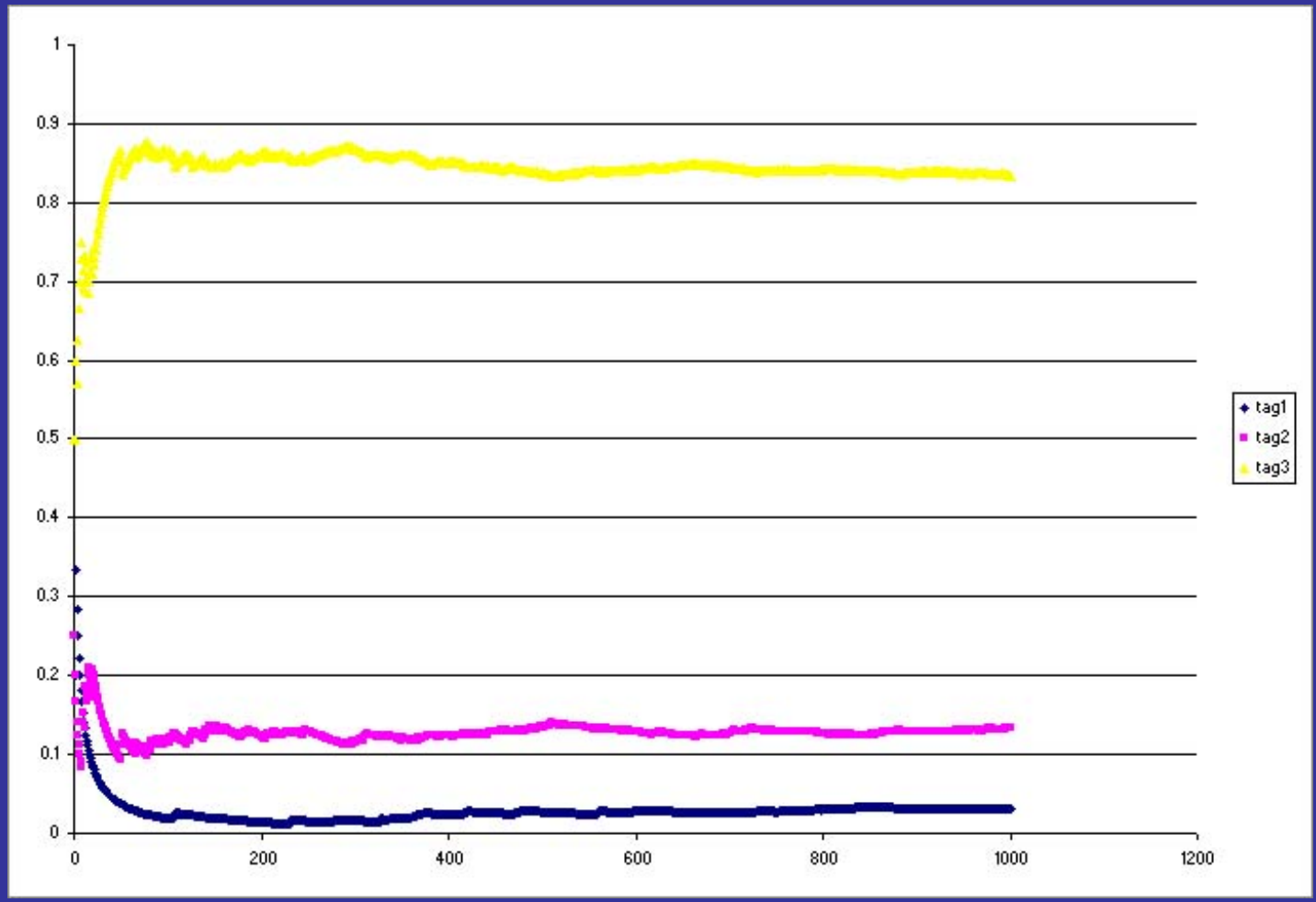
eggenberger-polya process



eggenberger-polya process



eggenberger-polya process



it is all about the power of the implicit

for more information go to:

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