Thoughts on Human Emotions, Communication Breakthroughs, and the Next Generation of Data Mining

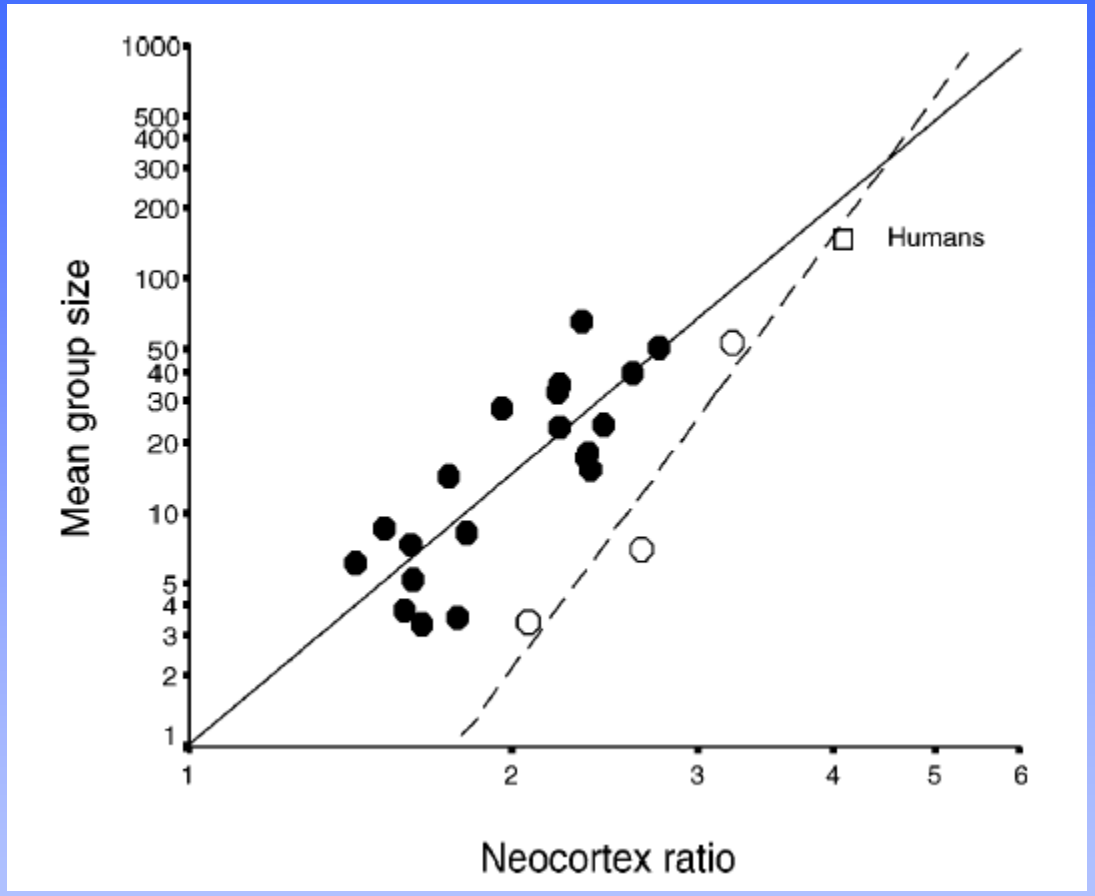
Hillol Kargupta

University of Maryland Baltimore County & Agnik

Roadmap

- Human emotions and communication
- Communication breakthroughs of the past
- What is missing?
- How data mining can help

Human Emotions and the Need for Interactions



- *R.I.M. Dunbar*, THE SOCIAL BRAIN: Mind, Language, and Society in Evolutionary Perspective, Annual Review of Anthropology, October 2003, Vol. 32, Pages 163-181

The First Breakthrough: Speech

- Early form of language, 200,000 years ago
- Local Communication
- Can communicate with only those who are nearby and can hear what you are saying.

Oracle of Apollo, Delphi

Extending the Range Over Time

- 30,000 BC
- Observe an event
- Document for posterior generations
- One to Some

African talking drum.

African talking drum.

Expanding the Reach

A Scandinavian fire beacon.

African talking drum.

19th century postal system in Eastern Europe.

18th century stamp in India.

Evolution of Communication Structure

- One to Some
- One to One

Technology in 19th-21st Century

Siemens Telex

Telephone from 1896.

Radio from 1959

Cell-phone in 2007

Further Evolution of Communication Structure

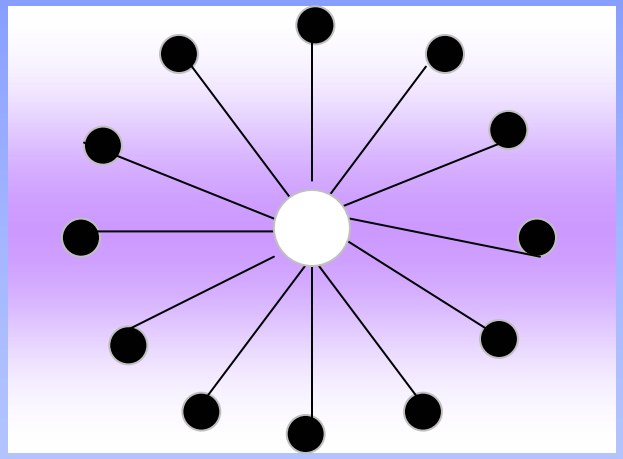
- One to One
- Many to One
- Mostly Address-based

That is Changing

- Spams
- Social networking sites
- Search engines
- Citizen Journalism

Problems of Current Client-Server Models

- Economics of Mass Communication
- Privacy and Intellectual Property Issues
- Not Scalable



Reliance on a central server.

Current Approach

- Taking your TV remote away and letting someone else to find the right content for you…Hmmm…

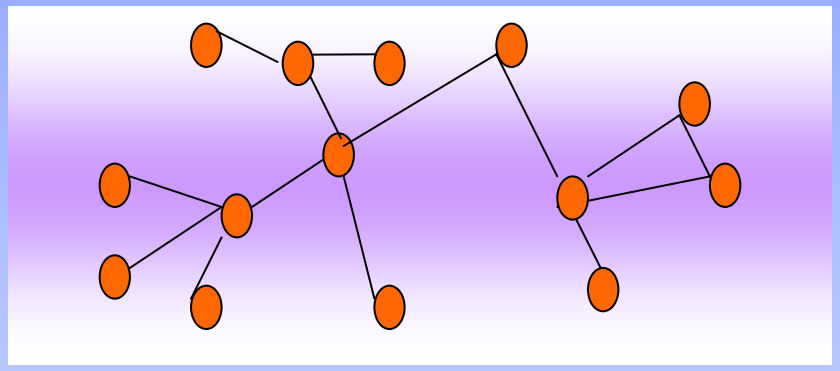
Note the Remote

Channel 1
Channel 2
.
,
,
,

Channel 150

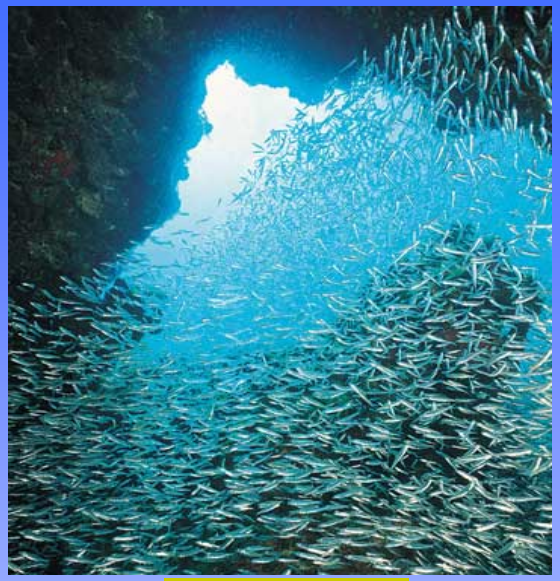
A Local Approach

- Local control in distributed systems
- Efficient global communication through local interactions
- Bounding the cost at every node



Examples in Natural Systems

- Human societies
- Swarm behavior in fish schools
- Insect colonies



Fish school

Termite colonies

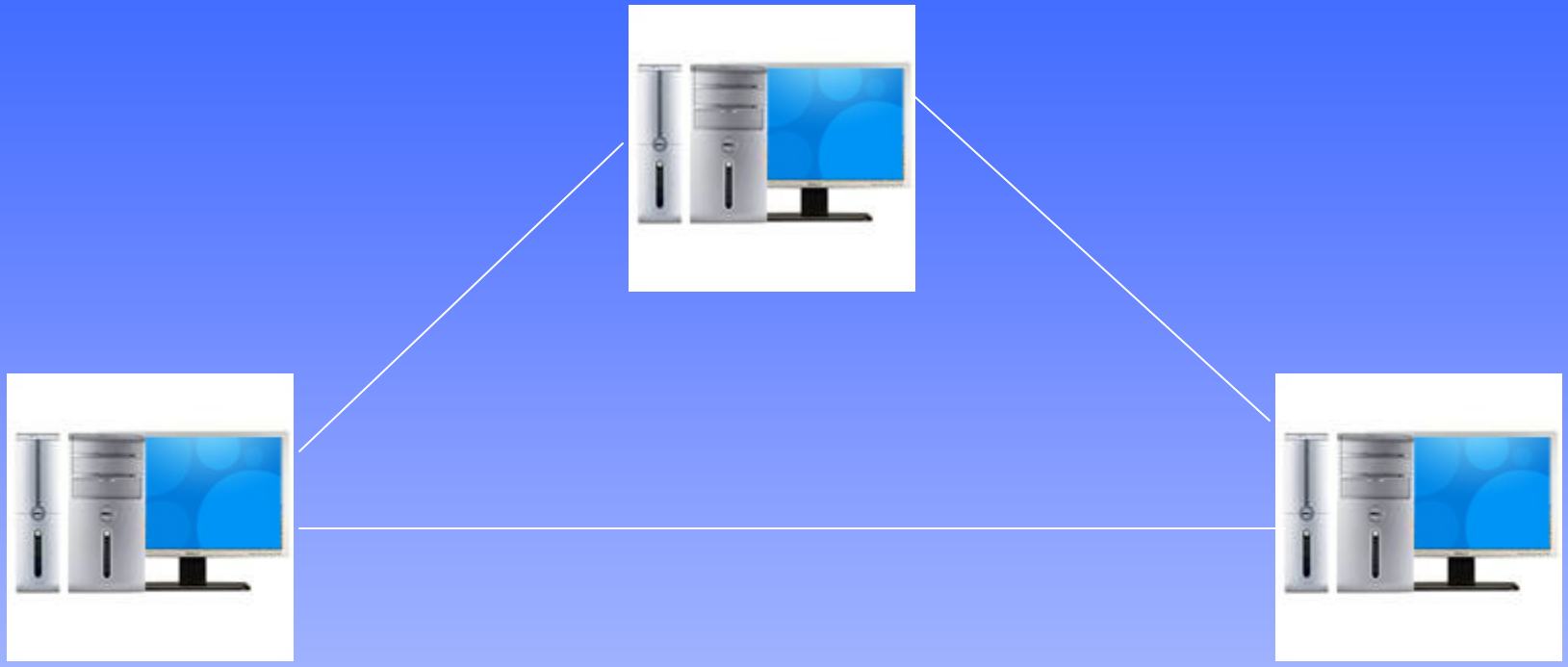
Peer-to-peer (P2P) Networks

- Relies primarily on the computing resources of the participants in the network rather than a relatively low number of servers.
- P2P networks are typically used for connecting nodes via largely ad hoc connections.
- No central administrator/coordinator
- Peers simultaneously function as both "clients" and "servers"
- Privacy is an important issue in most P2P applications

Where do we find P2P Networks?

- Applications:
  - File-sharing networks: KaZAa, Napster, Gnutella
  - P2P network storage, web caching,
  - P2P bio-informatics,
  - P2P astronomy,
  - P2P Information retrieval
- P2P Sensor Networks?
- P2P Mobile Ad-hoc NETwork (MANET)?
- Next Generation:
  - P2P Search Engines, Social Networking, Digital libraries, P2P “YouTube”?

P2P Web Mining



- Web mining in a sever-less environment

Useful Browser Data

- Web-browser history
- Browser cache
- Click-stream data stored at browser (browsing pattern)
- Search queries typed in the search engine
- User profile
- Bookmarks

- Challenges
  - Indexing, clustering, data analysis in a decentralized asynchronous manner
  - Scalability
  - Privacy

References on P2P Web Mining

- K. Das, K. Bhaduri, K. Liu, H. Kargupta. (2006). Identifying Significant Inner Product Elements in a Peer-to-Peer Network. *IEEE Transactions on Knowledge and Data Engineering*. (Accepted, in press)

- K. Liu, K Bhaduri, K. Das, P. Nguyen, H. Kargupta (2006). Client-side Web Mining for Community Formation in Peer-to-Peer Environments. *ACM SIGKDD Explorations*. Volume 8, Issue 2, Pages 11 - 20.

P2P NASA Astronomy Data Mining

- Virtual Observatories
  - Client-server architecture
  - Consider Sloan Digital Sky Survey:
    - 2M hits per month
    - traffic is doubling every 15 months
  - Need better scalability
- MyDB: Download and locally manage your data
- Network of such databases
- Searching, clustering, and outlier detection in P2P virtual observatory data network.

- NASA AIST Project at UMBC

Some References

- D. Peleg. (2000) Distributed Computing: A Locality-Sensitive Approach, SIAM,Philadelphia.
- M. Naor and L. Stockmeyer. (1995). What can be computed locally? SIAM Journal on Computing, Volume 24 , Issue 6, Pages: 1259 - 1277
- H. Kargupta and K. Sivakumar, (2004) Existential Pleasures of Distributed Data Mining. Data Mining: Next Generation Challenges and Future Directions. Editors: H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha. AAAI/MIT Press.
- S. Datta, K. Bhaduri, C. Giannella, R. Wolff, and H. Kargupta. (2006). Distributed Data Mining in Peer-to-Peer Networks. IEEE Internet Computing special issue on Distributed Data Mining, Volume 10, Number 4, Pages 18 - 26.
- Assaf Schuster and Ran Wolff. (2003) Association Rule Mining in Peer-to-Peer Systems Special Issue on Distributed and Mobile Data Mining, IEEE Transactions on System, Man, Cybernetics, Part B.

Recommendations and a Question

- Think computing from a truly interdisciplinary perspective
- Technology does not matter unless it can “sync” with human needs
- Does the current client-server model for connecting with others “sync” with our basic needs?