


*stuff*

Mining ~~All~~ About Jeff Ullman

*Svetlozar Nestorov*

University of Chicago

# Data Mining, Stanford, and Jeff

- Mining Data At Stanford (MIDAS)
- Extending Association-Rule Mining
  - Dynamic Itemset Counting
  - Causal Rules
  - Query Flocks
  - High Confidence Rules
- 

# What Data to Mine?

- Jeff's purchases at Safeway?
- When Jeff buys pretzels and chips he tend to buy beer!
- Unfortunately, this data is not available!
- What else?
- **Publications!**

# Publication Data

- Source: DBPL
- Jeff's pubs: 227 (he has more, of course!)
- Not enough data to mine
  - Plenty for the AI types
- Add all papers of Jeff's coauthors.
- So, we have:
  - 5297 papers
  - 2830 authors
  - 14279 author-paper pairs.

# Prolific Duos

- Authors who wrote more than 10 papers together.
  - Jeff and Alfred Aho
  - Christos Papadimitriou and Mihalis Yannakakis
  - Rajeev Rastogi and Avi Silberschatz
  - many more...
- No surprises!

# Prolific Trios and More

- Most prolific trios:
  - Stefano Paraboschi, Piero Fraternali and Stefano Ceri
  - Daniela Florescu, Dan Suciu and Alon Y. Levy
  - Divesh Srivastava, S. Sudarshan and Raghu Ramakrishnan
- No Jeff: Why?
- Prolific quartets:
  - S. Mehrotra, R. Rastogi, H. Korth, A. Silberschatz
  - D. Florescu, D. Suciu, A. Levy, M. Fernandez
  - M. Lenzerini, G. Giacomo, D. Calvanese, M. Vardi

# Let's Lower the Support

- Prolific = 5
- Jeff is part of 17 prolific duos!
- Jeff is part of several prolific trios:
  - H. Hunt and T. Szymanski,
  - J. Hopcroft and A. Aho
  - Y. Papakonstantinou & H. Garcia-Molina
  - Y. Sagiv and J. Widom
- Surprise: No Jeff, Jennifer, Hector?

# How about Confidence?

- A. Aho -> Jeff with 70%
- Jeff, Y. Papakonstantinou -> H. Garcia-Molina with 100%!
- But we have more info, such as conference and year, and more questions!
- Time to flock the query!!!

# Long-term Relationships

- Authors who wrote papers together in more than 10 different years:
  - A. Aho and Jeff
  - C. Papadimitriou and M. Yannakakis
  - S. Abiteboul and V. Vianu

# Here, There, Everywhere!

- Authors who wrote papers published in more than 10 different venues:
  - Y. Sagiv and Jeff
  - A. Aho and Jeff
- There are 6 trios!

# Different Worlds

- Authors who frequently publish papers in the same venue but never together.
  - J. Widom, R. Ramkrishnan, H. Jagadish
  - R. Snodgrass, A. Levy, P. Bernstein, J. Widom
  - J. Widom, R. Snodgrass, A. Mendelzon, P. Bernstein

# What's in a Title

- Most common words in titles:
  - Database
  - System
  - Query
  - Data
  - Distributed
  - Relational
  - Algorithms
- Algorithms for Querying Data in Distributed Relational Database Systems

# Disclaimer

- Standard data-mining caveat!
- The speaker assumes no responsibility whatsoever for the validity of these results. In fact, the results may be incorrect, bogus, false, and misleading.
- For entertainment purposes only!