



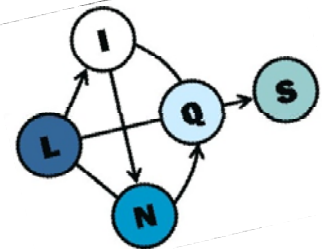
SRL: The Next Decade

Lise Getoor

University of Maryland, College Park



ILP07
June 21, 2007





Instructions

- o Dear panelists:

Thank you for agreeing to be on the panel on the long-term research agenda for inductive logic programming (ILP) and relational learning. With many different and interesting approaches to relational and structured machine learning being developed, we feel that this is an appropriate time to think about long-term strategic directions for the field. The panel is from 10:40 to 11:30 AM on the 21 of June in Austin Auditorium of Lasells Stewart Center in Oregon State University. It is going to be open to both ICML and ILP audiences.

To emphasize the long view, and the breadth of the research area, the title of the panel is

"Structured Machine Learning: The Next Ten Years"

Each of you will be given 5 minutes to make opening remarks. The rest of the time will be used to comments and questions from the audience and open discussion.

In your opening comments, we ask you to outline a 10-year research program in structured machine learning. Some relevant questions you might address are as follows:

- What are the important open problems in this area from AI/Computer Science point of view?
- What application problems have the most potential to make us confront these problems?
- How does solving these problems advance the state of the art and impact the world?
- How does it facilitate interaction between different "subareas" of this field including ILP, Statistical Relational Learning, Relational Reinforcement Learning, and others?
- What might be one or two focused Ph.D. thesis topics in this research program?

Please feel free to deviate from these if you think it is appropriate. We look forward to an insightful and thought-provoking discussion.

Thanks again for your time and effort in doing this.

Hendrik Blockeel Jude Shavlik
Prasad Tadepalli

ILP'07 Program Co-chairs

● ● ● Past: Focus on Representations



● ● ● Present: Focus on Tasks

○ **Collective Classification**

- Datasets and code available at <http://www.cs.umd.edu/linqs/projects/lbc>

○ Information Diffusion

○ **Entity Resolution**

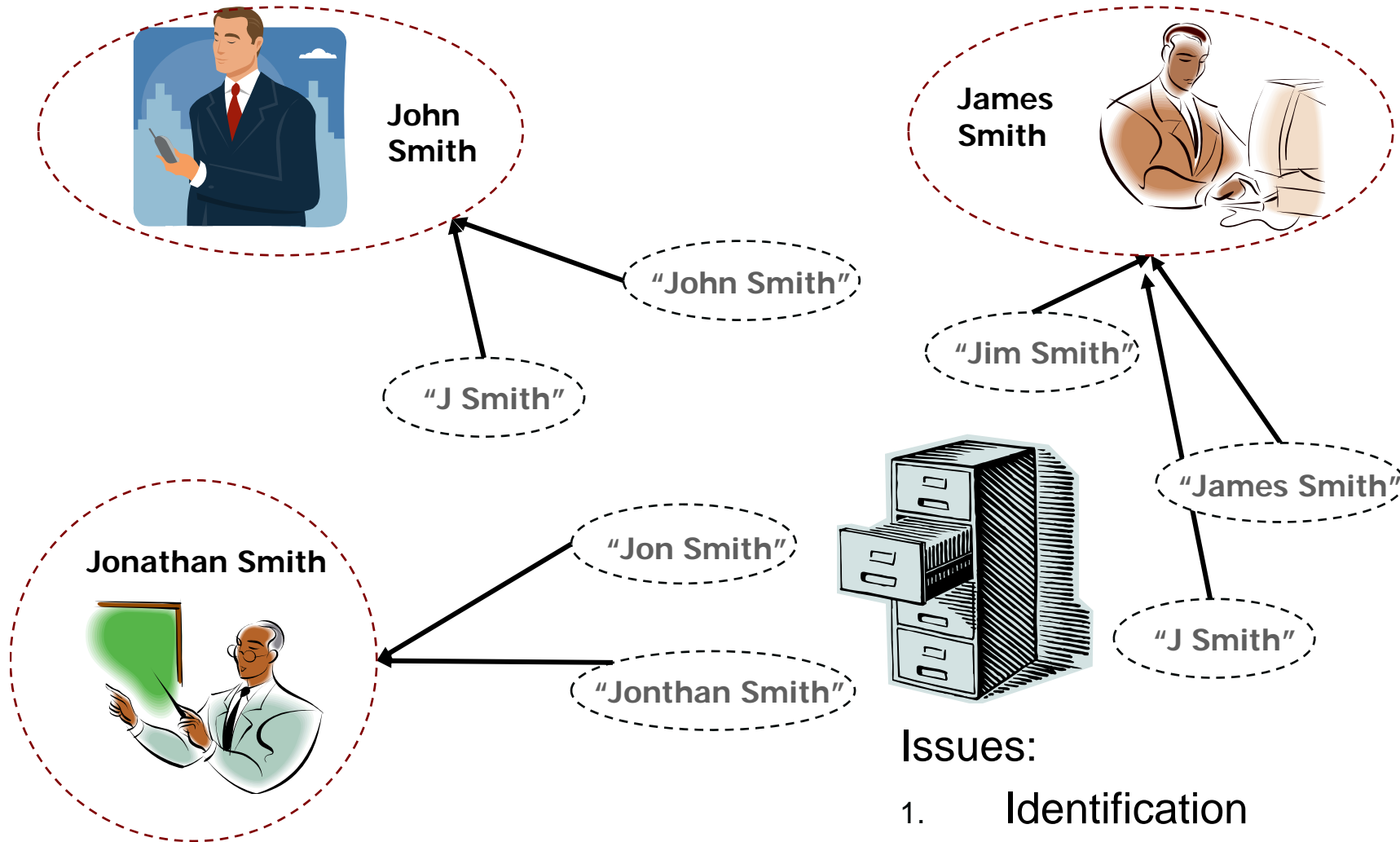
- Datasets and code available at <http://www.cs.umd.edu/linqs/projects/er>

○ Link Prediction

○ Community Discovery/Group Detection

○ Ontology Alignment

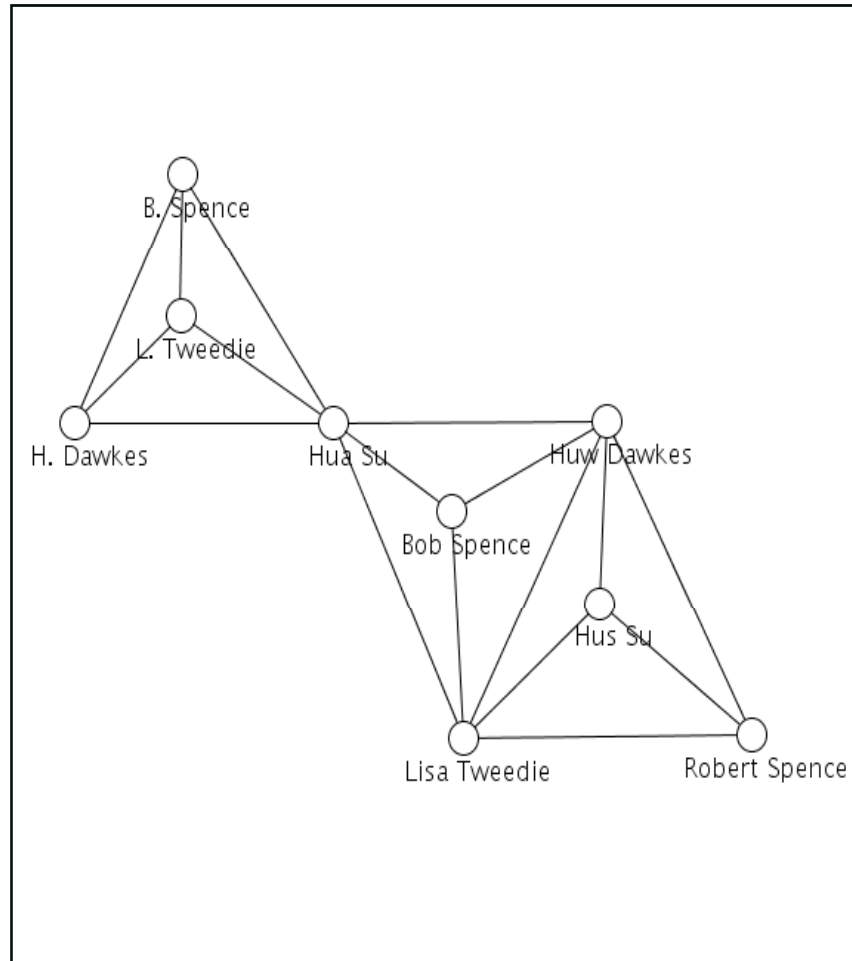
• • • The Entity Resolution Problem



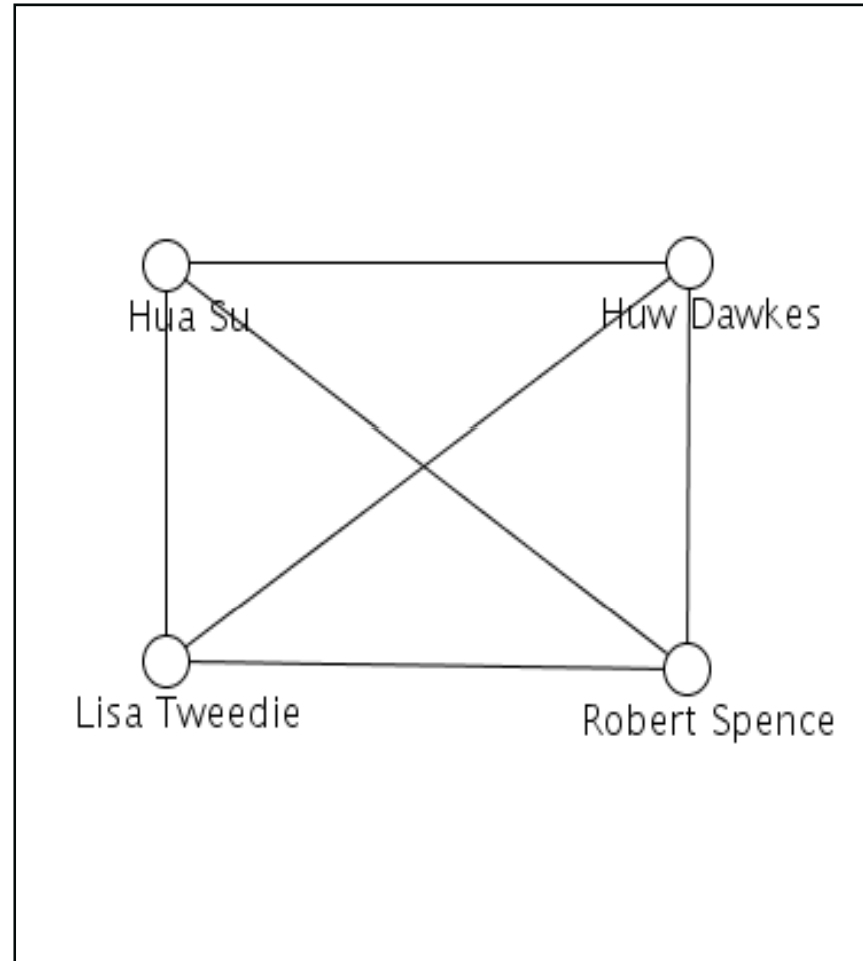
Issues:

1. Identification
2. Disambiguation

● ● ● InfoVis Co-Author Network Fragment



before



after

● ● ● Present: Focus on Tasks

- Collective Classification

- Datasets and data generator available at <http://www.cs.umd.edu/linqs/projects/lbc>

- Information Diffusion

- Entity Resolution

- Datasets and code available at <http://www.cs.umd.edu/linqs/projects/er>

- Link Prediction

- Community Discovery/Group Detection

- **Ontology Alignment**

● ● ● ILIADS

- **Goal:**

- Produce high-quality integration via a flexible method able to adapt to a wide variety of ontology sizes and structures

- **Method:**

- Combining statistical and logical inference
- Use schema (structure) and data (instances) effectively

- **Solution:**

- Integrated **L**earning **I**n **A**lignment of **D**ata and **S**chema (**ILIADS**)

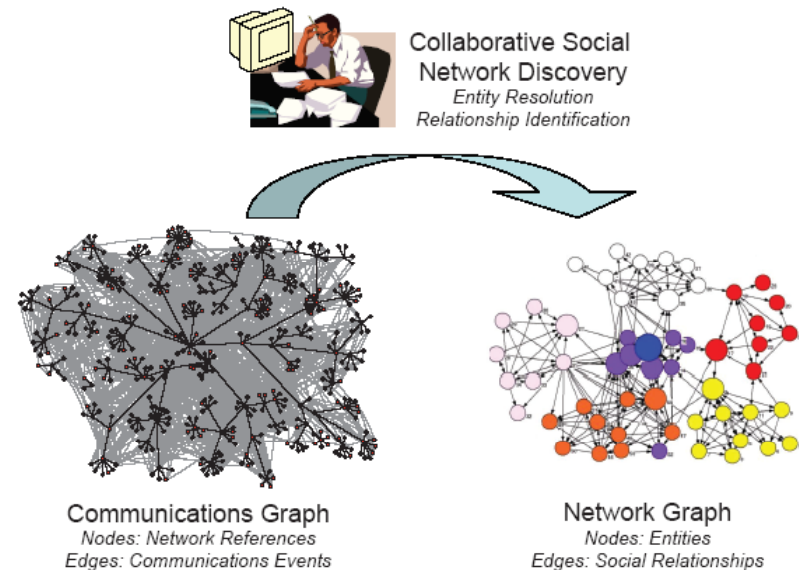
- Datasets and code available at:

<http://www.cs.umd.edu/lings/projects/iliads>

● ● ● Future: Focus on Integrated Tasks

○ Putting it all together...

- Bioinformatics
- Computer Vision
- Natural Language Processing
- Personal Information Management



● ● ● Research Agenda

- Visual Analytics
 - Complexity of the integrated SRL tasks require sophisticated user interfaces which allow user feedback and support explanation
- Query-time adaptive information gathering
 - Complexity of the integrated SRL tasks require flexible, adaptive algorithms which retrieve relevant information in real time
- Some related areas to keep in mind: resurgence of work in probabilistic databases (DB), social network analysis (social science), network science (physicists)

D-Dupe: An Interactive Tool for Entity Resolution

The screenshot displays the D-Dupe application window. On the left, a 'Find Duplicates' panel shows a table of similarity scores between nodes. The main area features a network graph with nodes like 'L. Tweedie', 'H. Dawkes', 'B. Spence', 'Lisa Tweedie', 'Hua Su', and 'Hus Su' connected by edges. Below the graph, a search results panel for 'hua' lists author IDs and names. Further down, 'Node Detail Viewer' and 'Edge Detail Viewer' provide specific information about selected nodes and edges.

Find Duplicates Table:

| Similarity | Node1 | Node2 |
|-------------------|--------|----------------|
| 0.888888888888889 | Hua Su | Hus Su |
| 0.746031746031746 | Hua Su | Alan Su |
| 0.650793650793651 | Hua Su | Stuart Shieber |
| 0.6 | | |
| 0.6 | | |
| 0.6 | | |
| 0.6 | | |
| 0.6 | | |
| 0.611111111111111 | Hua Su | Hank Hoek |
| 0.605555555555556 | Hua Su | Huw Dawkes |
| 0.6 | Hua Su | Allan Tuan |
| 0.6 | Hua Su | David Turo |
| 0.6 | Hua Su | Jianbo Shi |
| 0.6 | Hua Su | Jian Huang |
| 0.593434343434343 | Hua Su | Varun Saini |

Search Results (9 nodes found):

| AuthorID | AuthorName |
|----------|------------------|
| P573257 | M. C. Chuah |
| P507545 | Mei Chuah |
| P187165 | Mao Lin Huang |
| P470250 | Joshua Levasseur |
| P195636 | Mei C. Chuah |
| P112532 | Hua Su |
| P254127 | S. Huang |
| P74503 | Ed Huai-hsin Chi |
| P139655 | Jian Huang |

Node Detail Viewer (7 items):

| AuthorID | AuthorName |
|----------|---------------|
| P573115 | H. Dawkes |
| P572966 | B. Spence |
| P113087 | Huw Dawkes |
| P172581 | Lisa Tweedie |
| P573241 | L. Tweedie |
| P31332 | Bob Spence |
| P246545 | Robert Spence |

Edge Detail Viewer (3 items):

| ArticleId | Title | Source | Date |
|-----------|--|--|------------------------|
| acm857591 | Visualization for functional design | Proceedings of the 1995 IEEE Symposium Information Visualization | 10/30/1995 12:00:00 AM |
| acm223464 | The influence explorer | | |
| acm238587 | Externalising abstract mathematical models | | |

<http://www.cs.umd.edu/projects/lings/ddupe>

Novel combination of network visualization and statistical relational models well-suited to the visual analytic task at hand

Finding possible duplicates completed!





Thanks!

<http://www.cs.umd.edu/~getoor>

Work sponsored by the National Science Foundation,
Google, KDD program and National Geospatial Agency

