



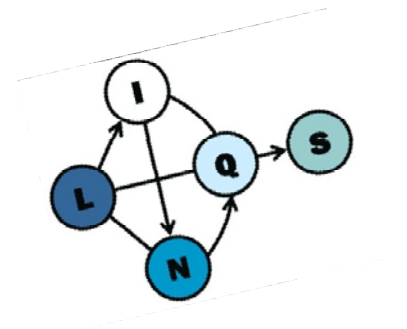
Graph Identification & Privacy

Lise Getoor

University of Maryland, College Park

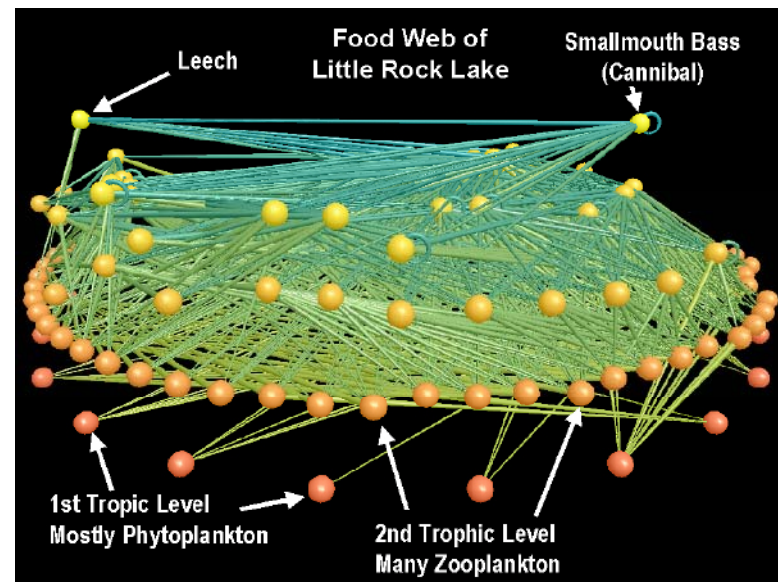
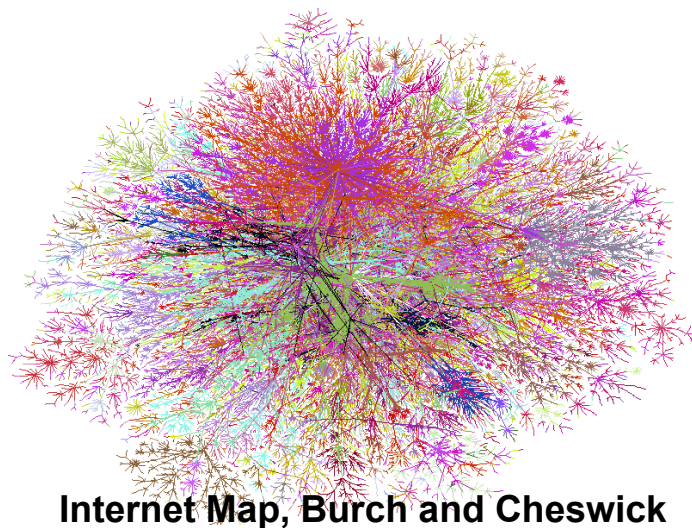


Statistical Inference in Complex Graphs
December 5, 2008



Graphs and Networks everywhere...

- o The Web, social networks, communication networks, financial transaction networks, biological networks, etc.



Food Web, Martinez et al.

Wealth of Data

- Inundated with data describing networks
- But much of the data is noisy and incomplete and at WRONG level of abstraction for analysis

Identification

- On the other hand, the data can be joined and sensitive information can be inferred

Privacy

Overview: Identification

- Many real world datasets are relational in nature
 - Social Networks – people related to each other by relationships like friendship, family, enemy, boss_of, etc.
 - Biological Networks – proteins are related to each other based on if they physically interact
 - Communication Networks – email addresses related by who emailed whom
 - Citation Networks – papers linked by which other papers they cite, as well as who the authors are
- However, the observations describing the data are noisy and incomplete
- **graph identification problem** is to **infer** the appropriate **information graph** from the **data graph**

Example: Organizational Hierarchy

Ideally:

- Know who are the criminals
- Know where the criminals stand in the organization
- Know friends and social groups belong to



In Reality:

- Annotated only a handful of individuals
- Don't have social structure, have an email communication network which reflects that structure

Enron Investigators



Information Graph



Data Graph

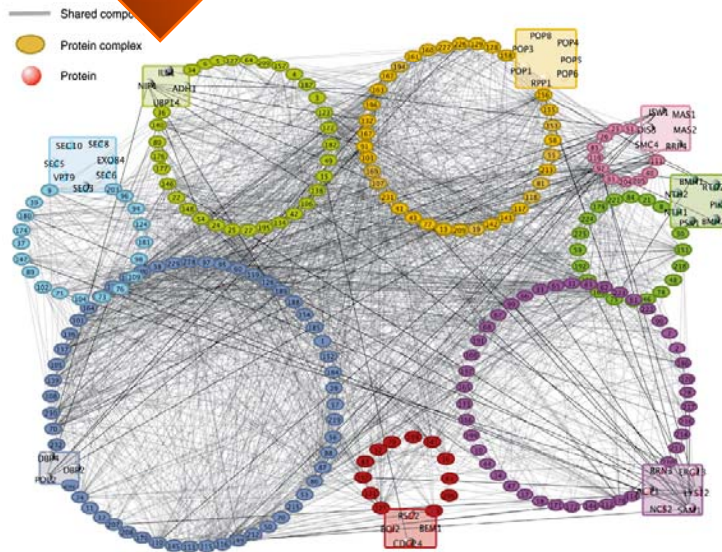
Example: Protein Interaction Network

Ideally:

- Know which proteins interact
- Know functions of proteins
- Known complexes of proteins



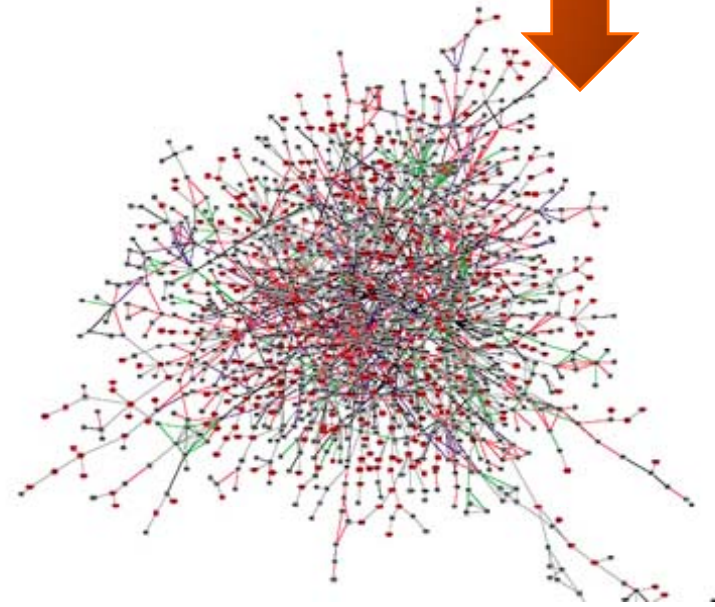
Network Research Group



Information Graph

In Reality:

- Accurate and Complete Information is expensive
- Available information is noisy and incomplete (i.e., high throughput)



Data Graph

Example: Internet Security

Ideally:

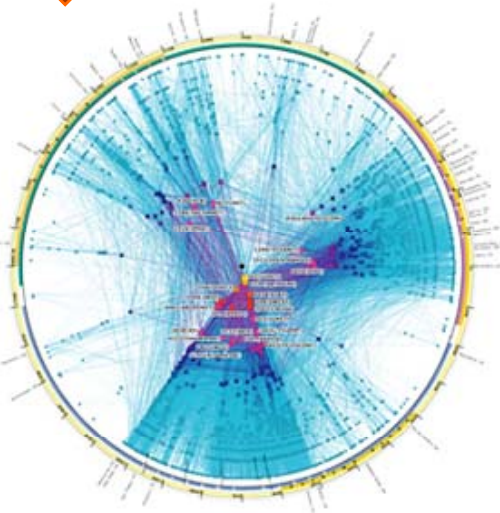
- Know the network from an AS and ISP level
- Know which computers are malicious and launching a DDOS attack



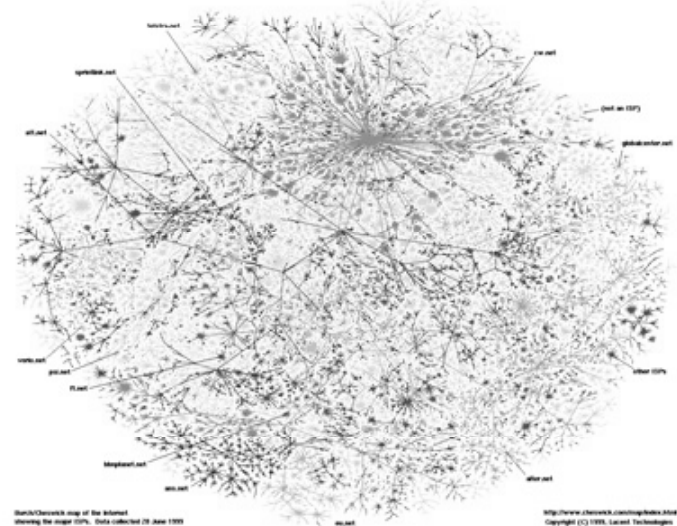
Network Operator

In Reality:

- Only have trace route information in IP address level
- Do not know legitimate traffic vs. malicious traffic



Information Graph

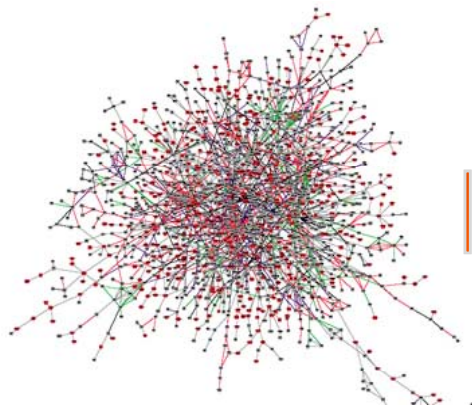


Data Graph

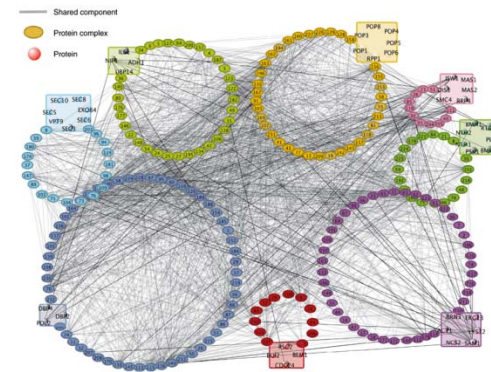
Solution

o Graph Identification:

- Infer the information graph that we want from the data graph that we have
- Assumption:
 - Dependencies exist such that knowledge of the nodes, edges, and attributes of the data graph can let us infer the nodes, edges, and attributes of the information graph



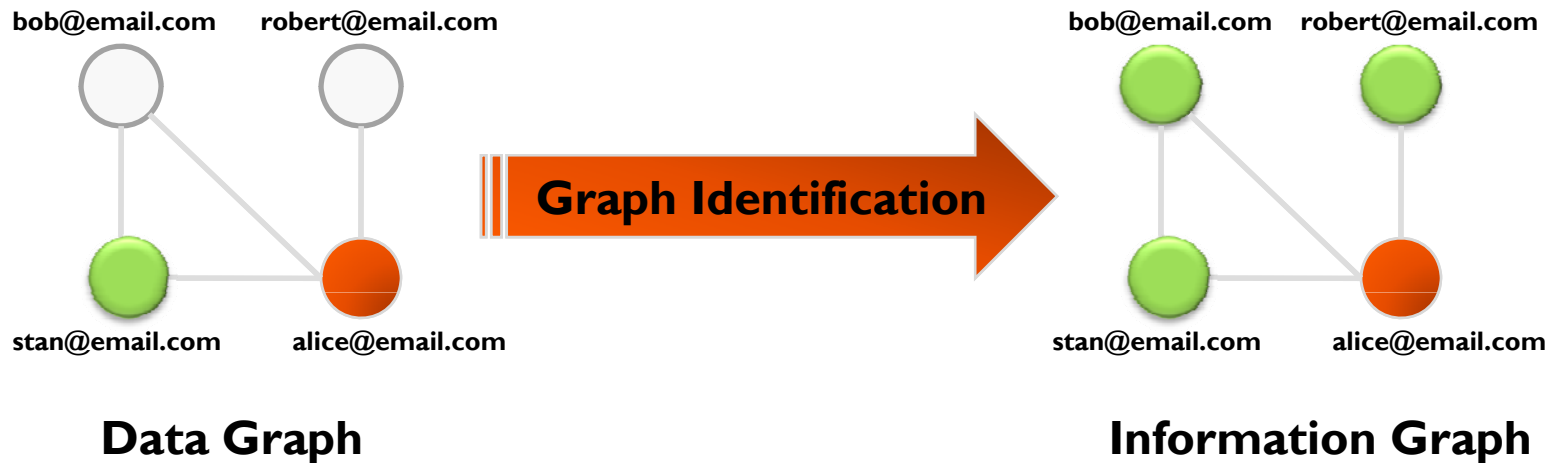
Data Graph



**Information
Graph**

Collective Classification

Collective Classification (CC): Given a set of labels (orange and green), label the objects whose label is unknown with the correct label

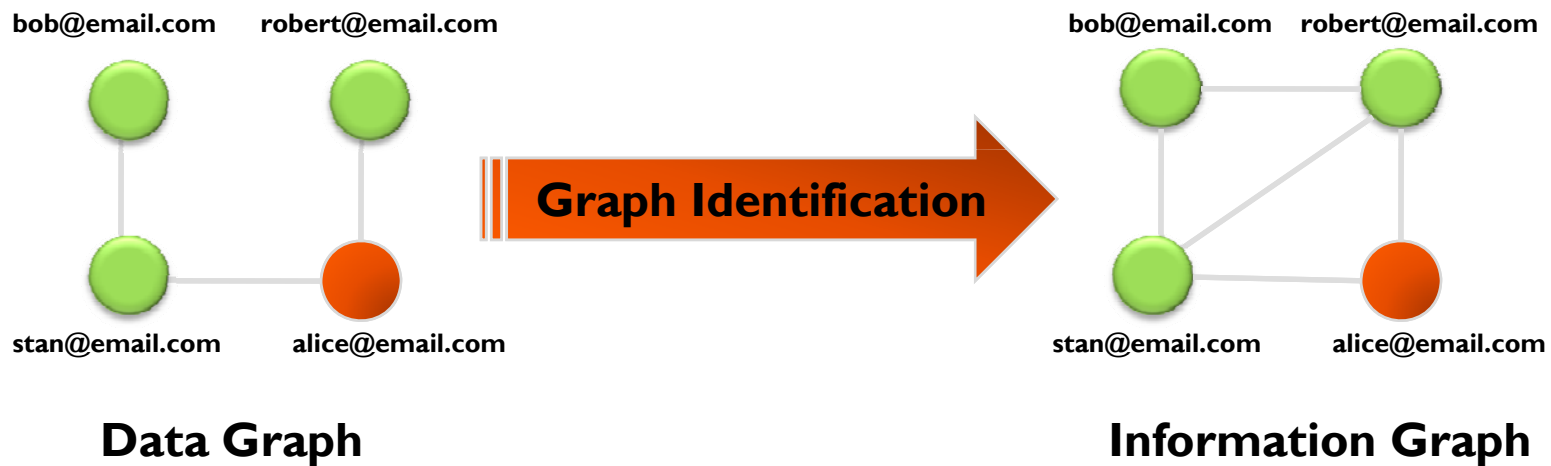


Assumptions:

- Set of nodes and edges in data and information graphs are the same
- Inference depends on known labels and attributes of the nodes and edges

Link Prediction

Link Prediction (LP): Predict the existence of edges

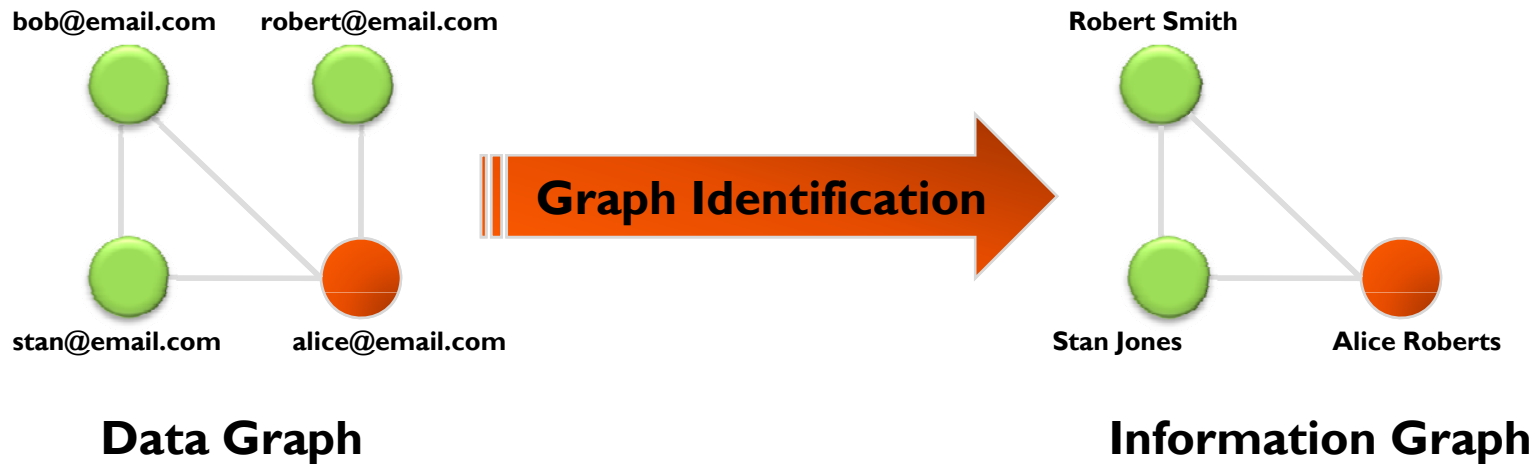


Assumptions:

- Set of nodes and attributes in data and information graphs are the same
- Inference depends on known labels and attributes of the nodes and edges

Entity Resolution

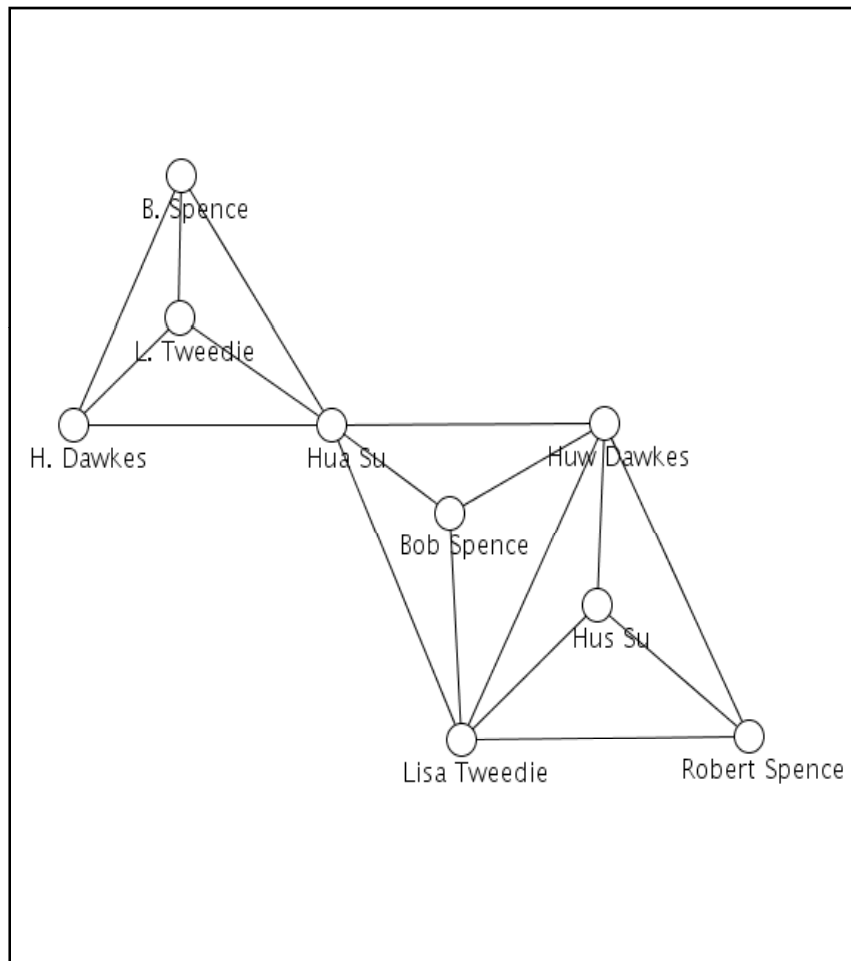
Entity Resolution (ER): Identify the the underlying entity represented by the references



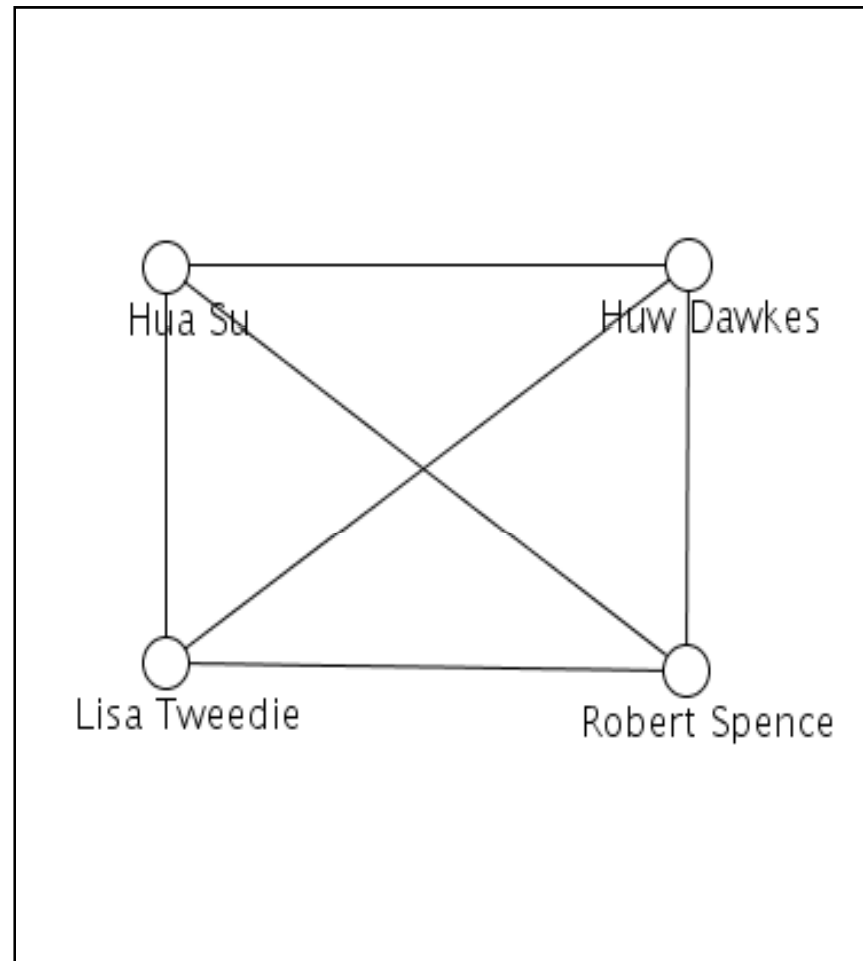
Assumptions:

- Edges and attributes of entities based on the edges and attributes of the merged references (if known)
- Inference only depends on known labels, nodes, and edges

InfoVis Co-Author Network Fragment

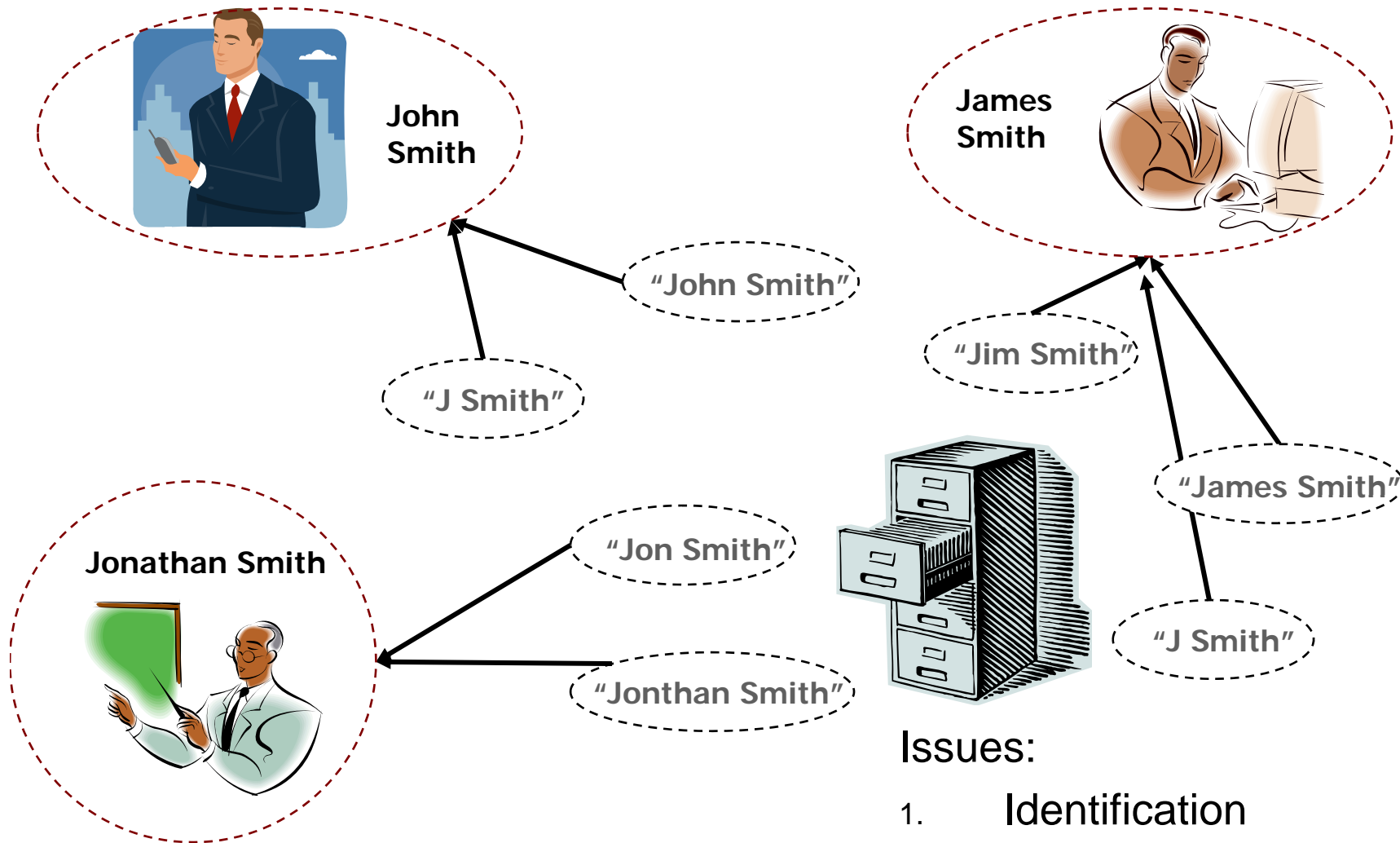


before



after

The Entity Resolution Problem



Issues:

1. Identification
2. Disambiguation

Attribute-based Entity Resolution

Pair-wise classification

"J Smith"	"James Smith"	?
"Jim Smith"	"James Smith"	0.8
"J Smith"	"James Smith"	?
"John Smith"	"James Smith"	0.1
"Jon Smith"	"James Smith"	0.7
"Jonthan Smith"	"James Smith"	0.05

1. Choosing threshold: precision/recall tradeoff
2. Inability to disambiguate
3. Perform transitive closure?

Entity Resolution

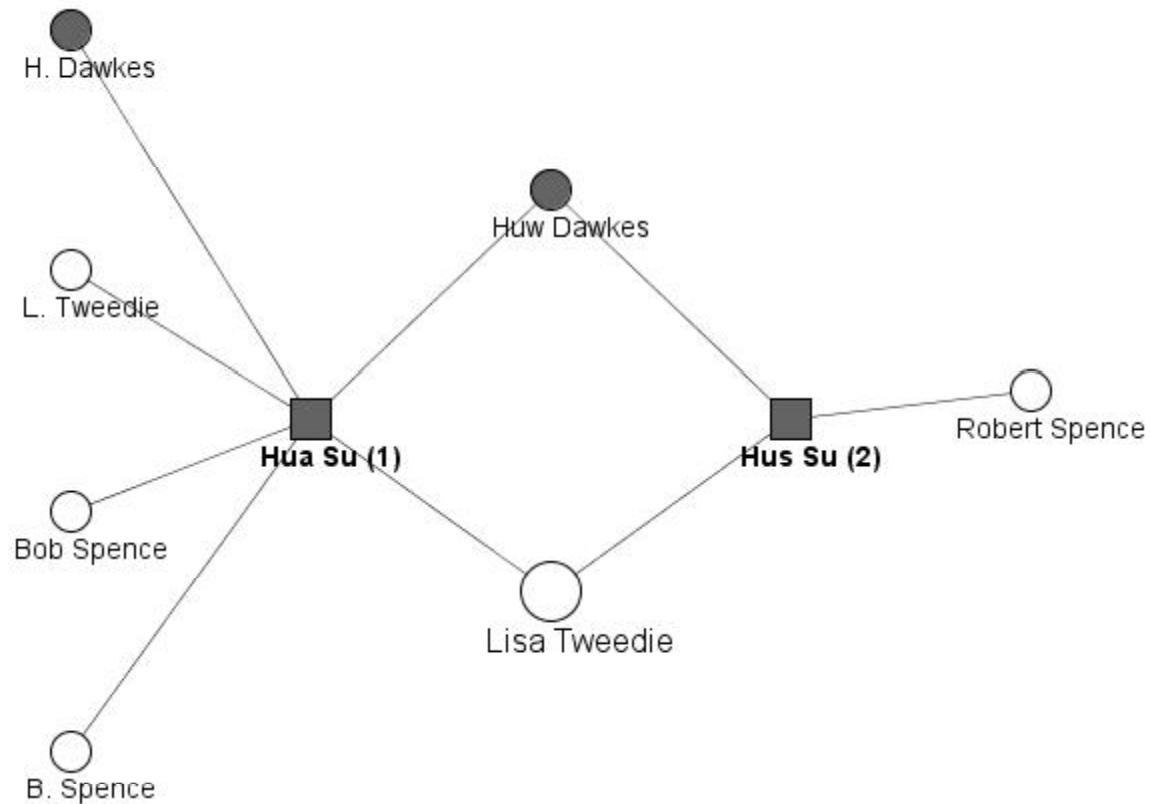
- o The Problem
- o **Relational Entity Resolution**
- o Algorithms

Relational Entity Resolution

- References not observed independently
 - Links between references indicate relations between the entities
 - Co-author relations for bibliographic data
 - To, cc: lists for email
- Use relations to improve identification and disambiguation

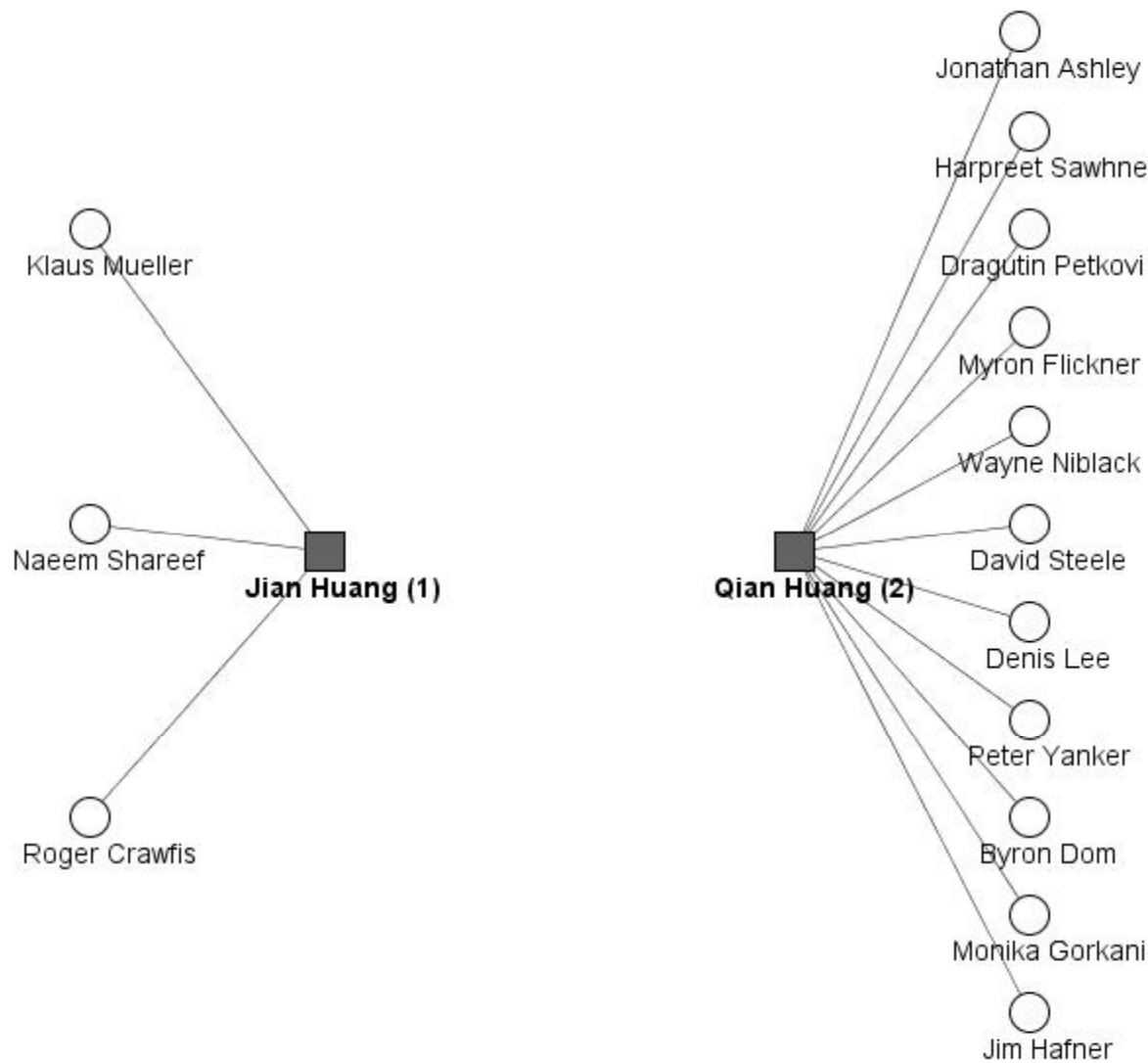
Pasula et al. 03, Ananthakrishna et al. 02, Bhattacharya & Getoor 04,06,07, McCallum & Wellner 04, Li, Morie & Roth 05, Culotta & McCallum 05, Kalashnikov et al. 05, Chen, Li, & Doan 05, Singla & Domingos 05, Dong et al. 05

Relational Identification



Very similar names.
Added evidence from
shared co-authors

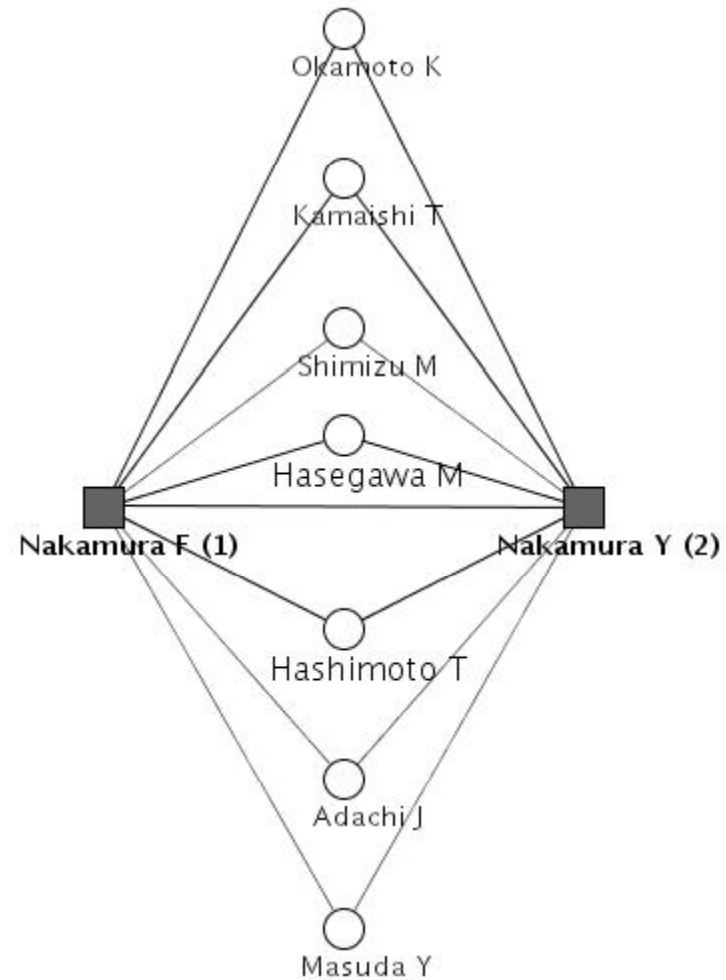
Relational Disambiguation



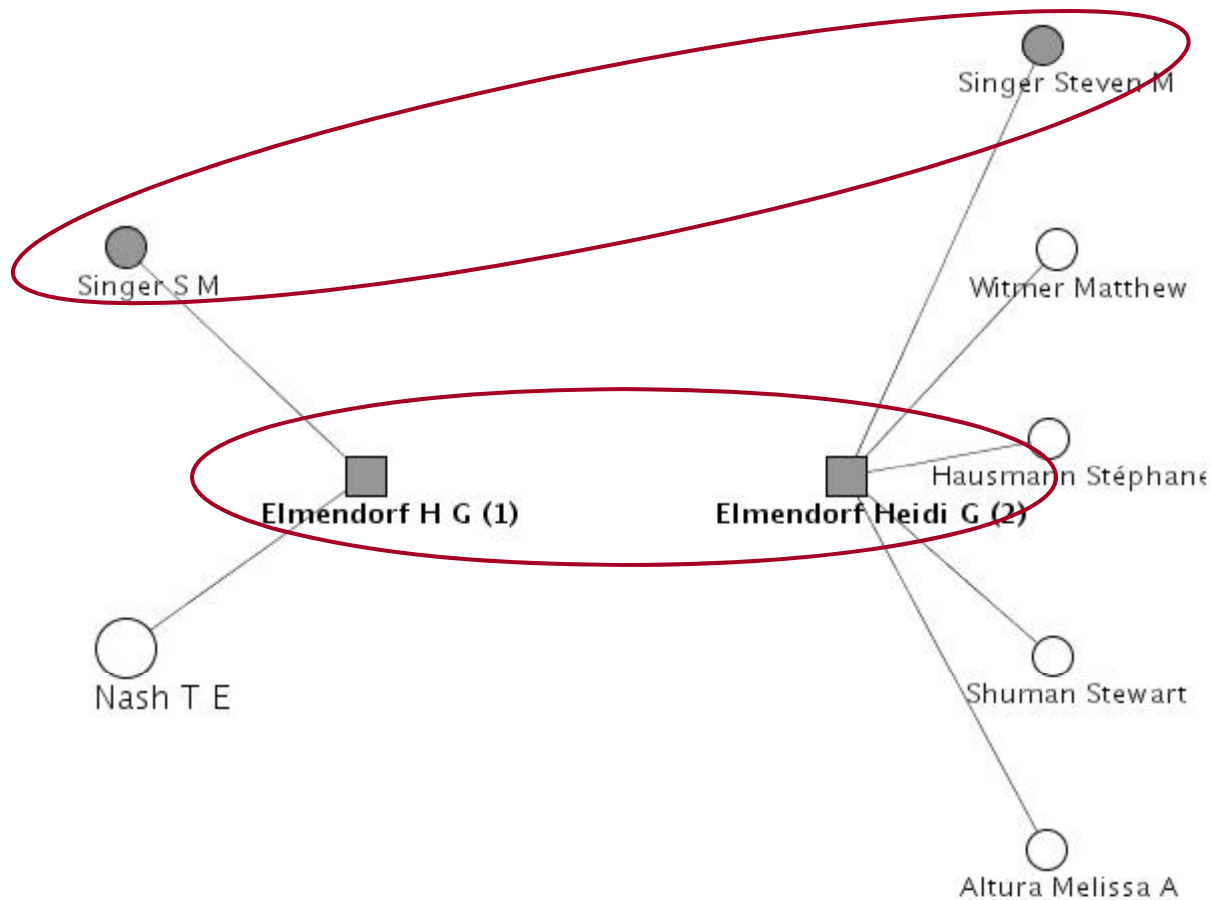
Very similar names
but no shared
collaborators

Relational Constraints

Co-authors are typically distinct

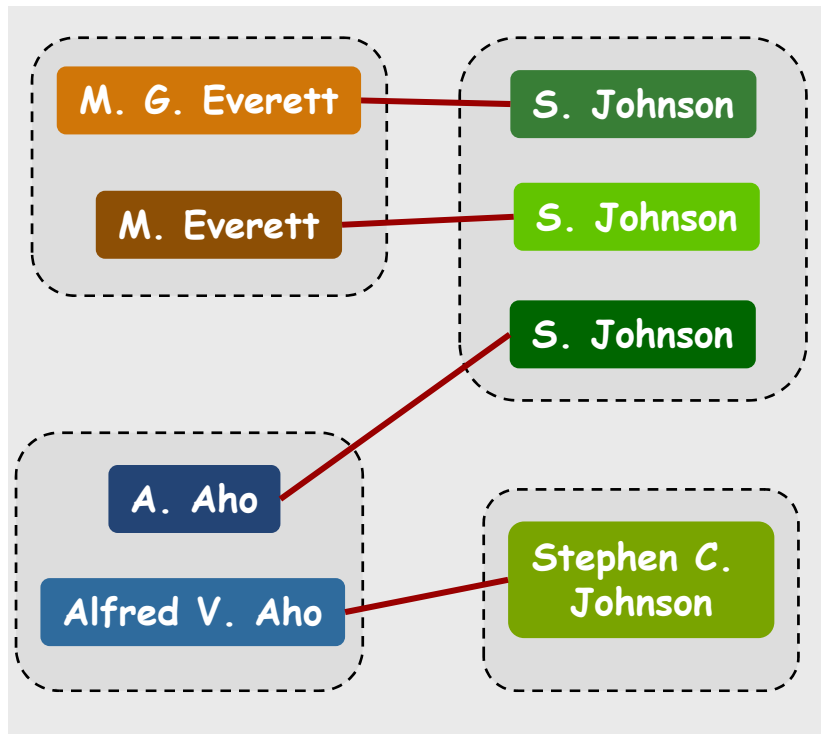


Collective Entity Resolution

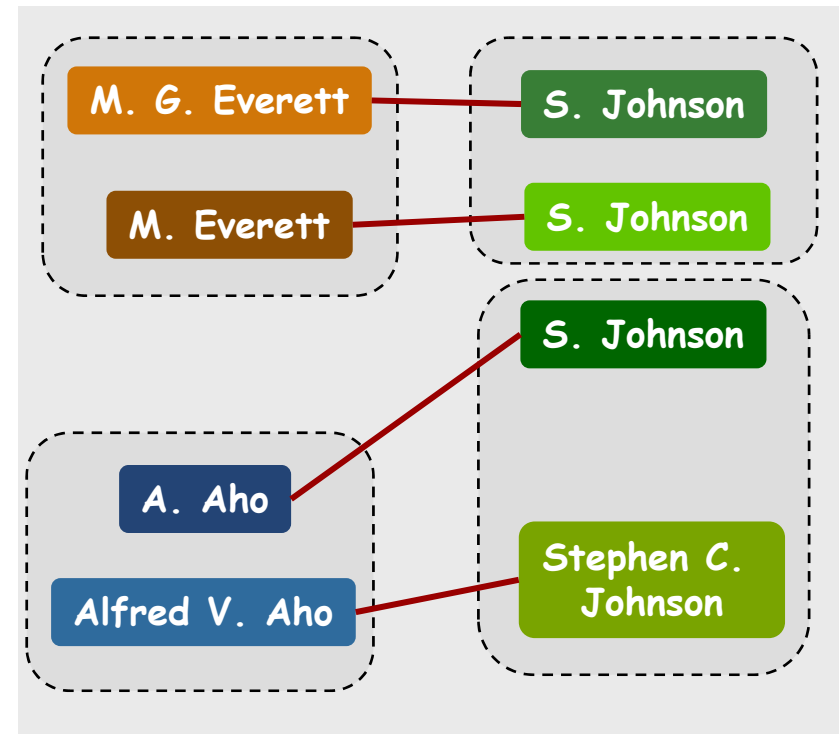


One resolution
provides evidence
for another => joint
resolution

Cut-based Formulation of RC-ER



Good separation of attributes
Many cluster-cluster relationships
➤ Aho-Johnson1, Aho-Johnson2,
Everett-Johnson1



Worse in terms of attributes
Fewer cluster-cluster relationships
➤ Aho-Johnson1, Everett-Johnson2

Objective Function

- Minimize:

$$\sum_i \sum_j w_A sim_A(c_i, c_j) + w_R sim_R(c_i, c_j)$$

weight for
attributes

similarity of
attributes

weight for
relations

Similarity based on relational
edges between c_i and c_j

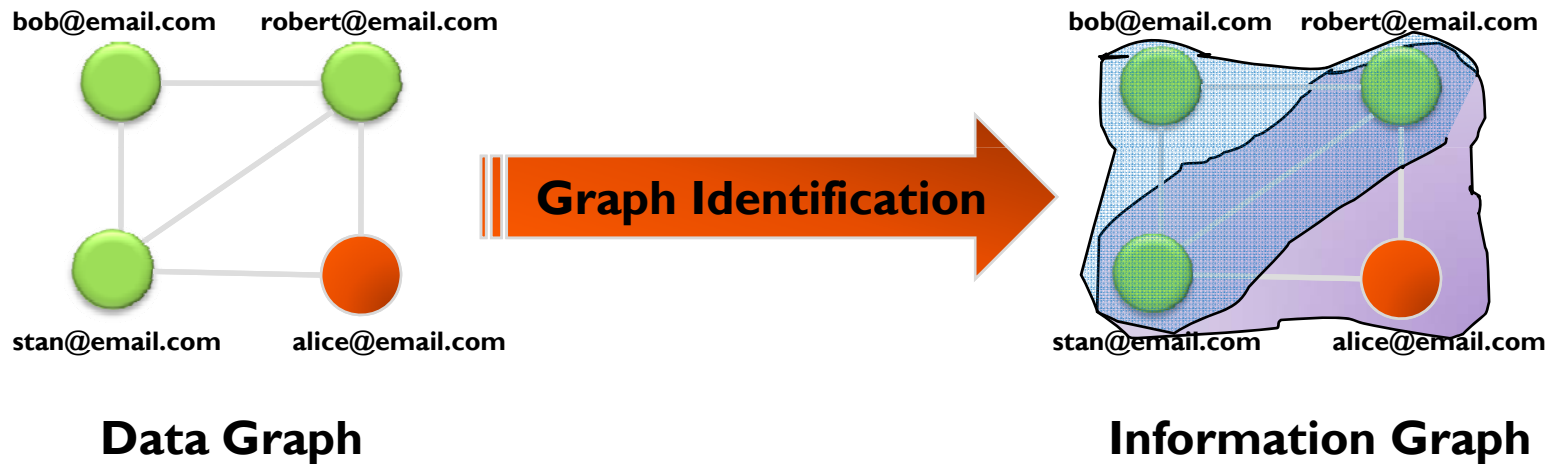
- **Greedy clustering algorithm:** merge cluster pair with max reduction in objective function

Relational Clustering Algorithm

1. Find similar references using 'blocking'
 2. Bootstrap clusters using attributes and relations
 3. Compute similarities for cluster pairs and insert into priority queue
 4. Repeat until priority queue is empty
 5. Find 'closest' cluster pair
 6. Stop if similarity below threshold
 7. Merge to create new cluster
 8. Update similarity for 'related' clusters
-
- $O(n k \log n)$ algorithm w/ efficient implementation

Group Detection

Group Detection (GD): Detect the underlying group(s) that the nodes and edges belong to



Assumptions:

- Set of nodes, edges, and attributes in data and information graphs are the same
- Inference only depends on known labels, nodes, and edges

Inference from Email Communications



Data Graph



Graph Identification



Information Graph

- No direct mapping from the nodes, edges, and attributes of data to information graph
- Need to infer existence of all the nodes and edges
- Need to infer the values of attributes based on data graph, as well as the nodes, edges, and other attributes of the information graph

Flipside....

Privacy

Privacy in social networks

- Identity disclosure
 - Entity resolution
- Attribute disclosure
 - Collective classification
- Link re-identification
 - Link prediction
- Group membership disclosure
 - Group detection

A public profile on Facebook



Add Emily as a Friend

View Photos of Emily (5)

Send Emily a Message

Poke Emily

Information

Networks:
The World Bank
Washington, DC
Birthday:
February 2

Friends

78 friends

See All



Julia
Bucknall



Roi
Weitz



David
Pollak

Emily Schneeweis got up at 5 and cleaned the house (laundry, floors, fridge, sheets, recycling, bills..). 6 hours ago

Wall

Info

Photos

Boxes

Basic Information

Networks:
The World Bank
Washington, DC
Sex:
Female
Birthday:
February 2
Hometown:
Washington, DC
Political Views:
Liberal

attributes

Personal Information

Favorite Movies:
400 Blows, Being John Malkovich, Breakfast at Tiffany's, Casablanca, The Devil Wears Prada, Diva, The Diving Bell and the Butterfly, Eternal Sunshine of the Spotless Mind, Lost in Translation, Manhattan, sex, lies and videotape, Volver
Favorite Books:
Divisadero, Emma's War, Kafka on the Shore, The Interpreter of Maladies, Love in the Time of Cholera, Remains of the Day
Favorite Quotations:
Normal people are people you don't know well.

Groups

See All (12)

Member of:
Bryn Mawr College Class of 1991, Dogs at the Astoria, The Trews, Sarah Palin is NOT Hillary Clinton, I have more Foreign Policy Experience than Sarah Palin, DC Foodies, Bryn Mawr College Alumna, PeaceCorpsConnect - Returned Peace Corps Volunteers, IDS Alumni: George Washington University, Thailand will always be the Kingdom of Thailand not the republic, International Finance Corporation / The World Bank Group, Peace Corps Thailand

groups

Pages

See All (5)



World Bank Publications
Non-Profit

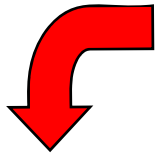


Bryn Mawr College
Education

friends

Emily's friends and groups

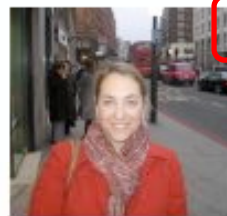
friends



group affiliation



Emily has 78 friends.



Elise Labott

private profile

Networks: Turner Broadcasting
CNN



Paul Barry

public profile

Message | View Friends

Networks: Washington, DC



Daniela Araujo

Message | View Friends

Networks: The World Bank

Displaying members of Sarah Palin is NOT Hillary Clinton.

500+ Members

No Officers

5 Admins



Name:

Kim Hennessey

Network:

Washington, DC



Name:

Alx Healy

Network:

Washington, DC



Name:

Elise Labott

Network:

Turner Broadcasting
CNN





Identity disclosure

- Occurs when the adversary is able to determine the mapping from a record to a specific individual
- Privacy literature has concentrated on structural identification

[Search](#) [Profile Search](#) | [Friend Finder](#)

[Show results from](#) Washington, DC [Show More Filters](#)

Displaying 1 – 10 out of over 500 people results at Washington, DC for: **bob davis**

	Name: Robert Davis Networks: Mary Washington Washington, DC	Add as Friend Send a Message View Friends
	Name: Robert Davis Network: Washington, DC	Add as Friend Send a Message
	Name: Robert Davis Network: Washington, DC	Add as Friend Send a Message View Friends
	Name: Bob Davis Network: Washington, DC	Add as Friend Send a Message

Attribute disclosure

- Occurs when an adversary is able to determine the value of a user attribute that the user intended to stay private
 - Example: is someone liberal?

private profile



public profile



[Add Paul as a Friend](#)
[View Photos of Paul \(1\)](#)
[Send Paul a Message](#)
[Poke Paul](#)

Paul Barry

[Wall](#) [Info](#) [Photos](#) [Boxes](#)

Basic Information

Networks: Washington, DC
Sex: Male
Birthday: June 2
Relationship Status: Married to Kacie Goddard

Education and Work

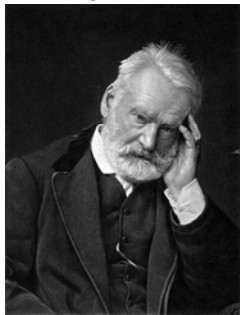
College: American '92
International Studies
High School: Cooper High School '85
Employer: WashingtonPost.Newsweek
Position: Network Engineer
Time Period: March 1999 – Present

Link re-identification

- Occurs when an adversary is able to infer that two entities participate in a particular type of sensitive relationship or communication

Disease data

has hypertension



father-of



Communication data

?



call



Robert Lady



Search data

Query 1:

“how to tell if your wife is cheating on you”

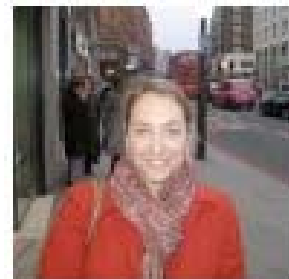
Query 2:

“myrtle beach golf course job listings”

same-user

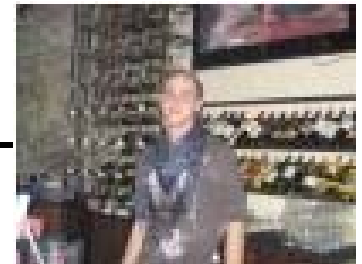
Social network data

Elise Labott



friends

Robert Davis



Group membership disclosure

- Occurs when an adversary is able to infer that a person affiliates with a group relevant to the classification of a sensitive attribute.
 - Example: is she liberal?

private profile



group affiliation?



Displaying members of Sarah Palin is NOT Hillary Clinton.

500+ Members

No Officers

5 Admins



Name: **Kim Hennessey**
Network: Washington, DC



Name: **Alx Healy**
Network: Washington, DC



Name: **Elise Labott**
Network: Turner Broadcasting
CNN

Anonymization Process

Valuable Data!

Test table 5-6
Graduate research assistants per 1,000 researchers,
by field, 1993

Field	Research assistants	Doctoral researchers	Ratio per 1,000 researchers
Science and engineering, total	89,729	140,843	638
Science, total	61,962	130,329	488
Physical sciences	12,344	20,026	611
Mathematics	1,396	9,517	147
Computer science	3,747	1,988	1,876
Environmental science	4,788	6,015	805
Life science	28,058	61,765	542
Psychology	4,500	16,821	381
Social sciences	7,240	25,886	288
Engineering	27,777	17,529	1,586

No = research assistant
See appendix tables 5-10 and 5-11
Source: U.S. Engineering Institute - 1993

**Data
Anonymization**

No privacy breaches!

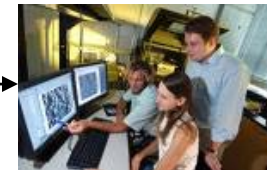
Test table 5-6
Graduate research assistants per 1,000 researchers,
by field, 1993

Field	Research assistants	Doctoral researchers	Ratio per 1,000 researchers
Science and engineering, total	89,729	140,843	638
Science, total	61,962	130,329	488
Physical sciences	12,344	20,026	611
Mathematics	1,396	9,517	147
Computer science	3,747	1,988	1,876
Environmental science	4,788	6,015	805
Life science	28,058	61,765	542
Psychology	4,500	16,821	381
Social sciences	7,240	25,886	288
Engineering	27,777	17,529	1,586

No = research assistant
See appendix tables 5-10 and 5-11
Source: U.S. Engineering Institute - 1993



Public



**Data representation?
Privacy breach?
Value of data?**



Anonymizing nodes

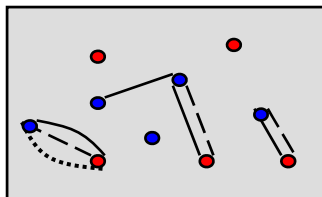
Ana	21	F	20740
Bob	25	M	83201
Chris	24	M	20742
Don	29	M	83209
Emma	28	F	83230
Fabio	31	M	83222
Gia	24	F	20640
Halle	29	F	83201
Ian	23	M	20760
John	24	M	20740

5-anonymity

applied to nodes

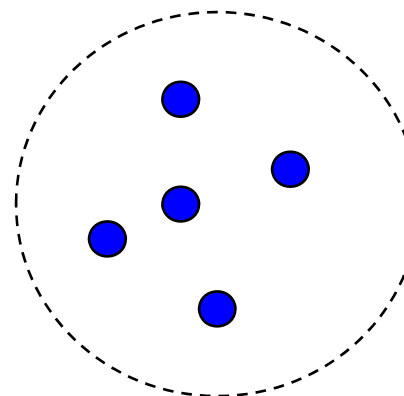
●	< 25	*	20***
●	≥ 25	*	832**
●	< 25	*	20***
●	≥ 25	*	832**
●	≥ 25	*	832**
●	≥ 25	*	832**
●	< 25	*	20***
●	≥ 25	*	832**
●	< 25	*	20***
●	< 25	*	20***

original data graph

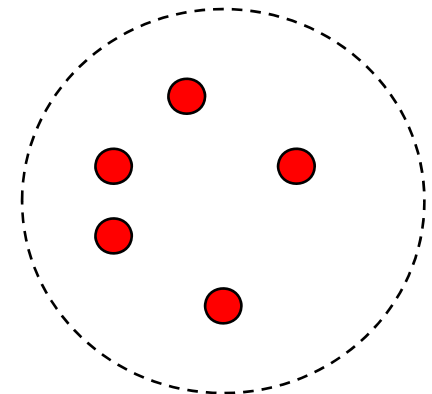


Equivalence

classes

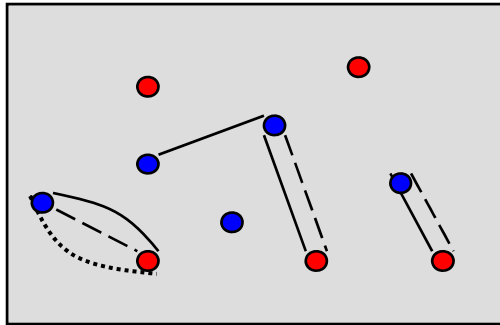


anonymized data graph

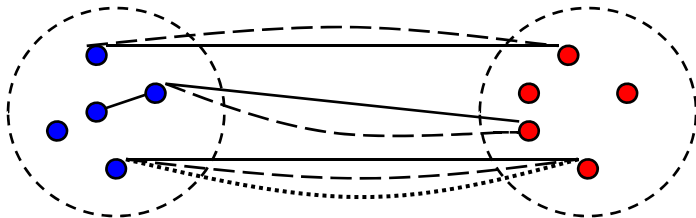


Anonymizing links

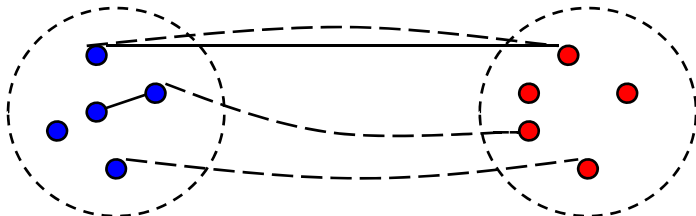
original graph



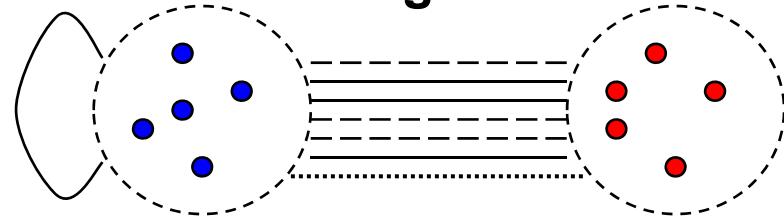
intact links



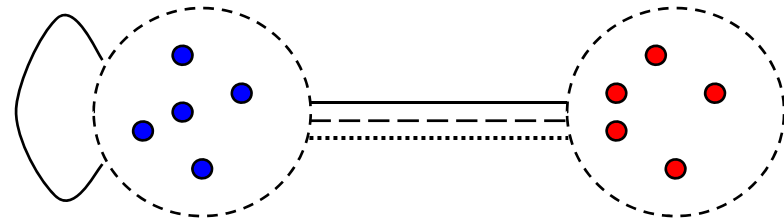
partial link removal



cluster-edge method



constrained cluster-edge method



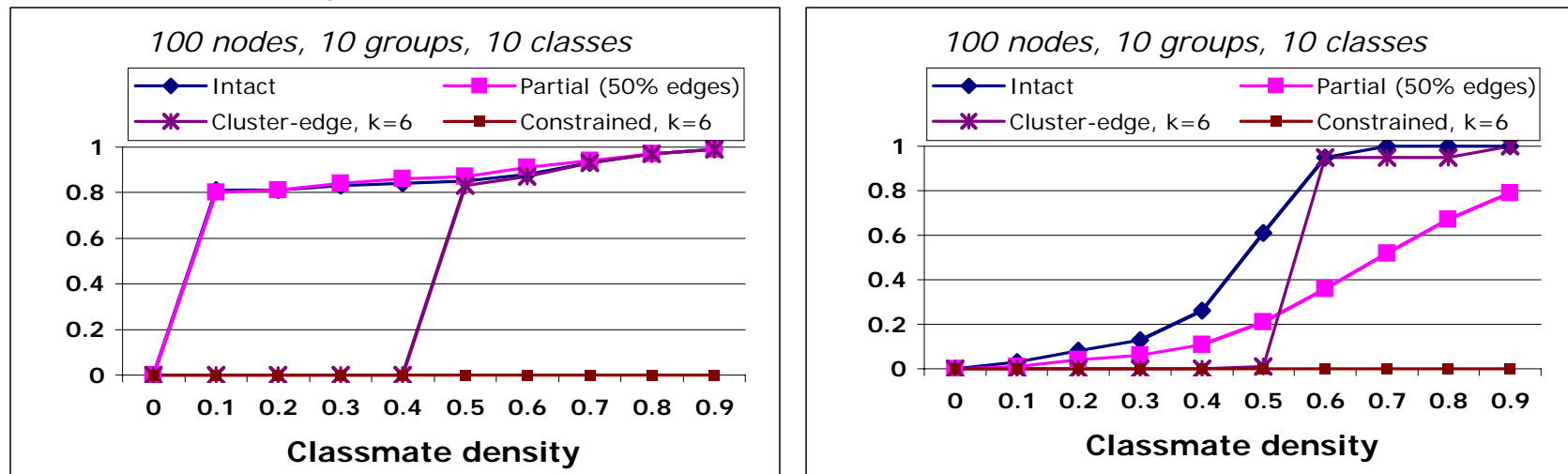
all links removed



Link re-identification results

- Synthetic dataset of students
 - Class enrollment and research group information
 - Observed links - classmates and groupmates
 - Sensitive link – friends
- Anonymize the data using the proposed methods
- Compute the existence prob of sensitive edge using a Noisy-Or model

Prediction precision and recall rates at various classmate densities



Reference: E. Zheleva, L. Getoor. Preserving the privacy of sensitive relationships in graph data. PinKDD 2007.

Attribute disclosure

- o In the context of online social networks



A screenshot of a Facebook profile for Emily Schneeweis. A red arrow points from the word "friends" to the "Friends" tab on the left. Another red arrow points from the word "group affiliation" to the "Groups" section on the right, which lists "The World Bank Publications" and "Bryn Mawr College".

Emily has 78 friends.



Profile of Elise Labott. The name "Elise Labott" is enclosed in a red box, with a red arrow pointing to the text "private profile". The profile shows a photo of a woman and lists networks: Turner Broadcasting, CNN.



Profile of Paul Barry. The name "Paul Barry" is enclosed in a red box, with a red arrow pointing to the text "public profile". The profile shows a photo of a family and lists networks: Washington, DC.



Profile of Daniela Araujo. The profile shows a photo of a woman and lists networks: The World Bank.

Displaying members of Sarah Palin is NOT Hillary Clinton.

500+ Members No Officers 5 Admins



Profile of Kim Hennessey. The profile shows a photo of two people and lists name: Kim Hennessey, network: Washington, DC.



Profile of Alx Healy. The profile shows a photo of a man and lists name: Alx Healy, network: Washington, DC.



Profile of Elise Labott. The profile shows a photo of a woman and lists name: Elise Labott, network: Turner Broadcasting, CNN.

Attribute disclosure results

- Given: public profiles (attribute label known), private profiles, groups, links

PROPERTY	flickr	facebook	dogster	BibSonomy
Number of possible labels	55	2/6	7	2
Sensitive attribute	location	gender/polviews	breed category	spammer



Emily Schneeweis got up at 5 and cleaned the house (laundry, floors, fridge, sheets, recycling, bills...). 6 hours ago

Wall Info Photos Boxes

Basic Information

Networks: The World Bank, Washington, DC

Sex: Female

Birthday: February 2

Hometown: Washington, DC

Political Views: Liberal

Personal Information

Favorite Movies: 400 Blows, Being John Malkovich, Breakfast at Tiffany's, Casablanca, The Devil Wears Prada, Drive, The Diving Bell and the Butterfly, Eternal Sunshine of the Spotless Mind, Translation, Manhattan, sex, lies and videotape, Vi Divasidero, Emma's War, Kafka on the Shore, The Maladies, Love in the Time of Cholera, Remains of the Day

Favorite Books: Divasidero, Emma's War, Kafka on the Shore, The Maladies, Love in the Time of Cholera, Remains of the Day

Favorite Quotations: Normal people are people you don't know well.

Groups

Member of: Bryn Mawr College Class of 1991, Dogs at the Arts Triennial, Sarah Palin is NOT Hillary Clinton, I have no Policy Experience than Sarah Palin, DC Foodies, Bryn Mawr College Alumna, PeaceCorpsConnect - Returned Volunteers, IDS Alumni, George Washington University, Thailand will always be the Kingdom of Thailand - republic, International Finance Corporation / The Group, Peace Corps Thailand

Pages

World Bank Publications Non-Profit

Bryn Mawr College Education

Find a Dogster

Find by name

Find by town

GO

Find Adoptable Dogs

Advanced Search

Search Photo Tags

Search Video Tags

Login

Enter Email

GO

My Account

Join Now

Forgot my password?

GO

My Account

Messages

See the Dogs!

Adoption

Community

Answers

Local Listings

Watch Videos

Resources

Read Diaries

DogsterPlus

Dogster Store

Dogster Info

Visit Catster

About caterina / Caterina Fake

← Photostream Send FlickrMail Buy caterina a Pro Account

Add Emily as a Friend

View Photos of Emily (5)

Send Emily a Message

Poke Emily

Flickr Workr, friend of small dogs, art fiend, book lover. I can't come to my personal Flickr Mail account (I delete them witho check out the [help page](#) or post in the forums.

<http://www.caterina.net/>

San Francisco, CA

Information

Networks: The World Bank, Washington, DC

Birthday: February 2

caterina's contacts (575)

kentgoldman heliosonnet clarkadial Paulo Coelho Indie Craft John Bucknall Rei Weitz David Polak

rust_sender etylabs Ety Labs

caterina's public groups

- 43 Things
- mappr
- One Letter
- Parking Lot Indicator
- NEW Horror
- Maitre Goulemier et Miladus
- Ransom Note Helper
- NEW a day in the life of ... [22nd September 2008]
- Web 2.0

Reference: E. Zheleva, L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. Under submission.

Attribute disclosure results

- Approaches to achieving attribute disclosure:
 - Using overall distribution - BASIC
 - Link-based - BLOCK, AGG, CC, LINK
 - Group-based - CLIQUE-LINK, GROUP

Table 2: Attack accuracy assuming 50% private profiles. The successful attacks are shown in bold.

	DEL	FLICKR	FACEBOOK (GENDER)	FACEBOOK (POLVIEWS)	DOGSTER	BIBSONOMY
BASIC		27.7%	50.0%	56.5%	28.6%	92.2%
Random guess		1.8%	50.0%	16.7%	14.3%	50%
BLOCK		8.8%	49.1%	6.1%	-	-
AGG		28.4%	50.2%	57.6%	-	-
CC		28.6%	50.4%	56.3%	-	-
LINK		56.5%	68.6%	58.1%	-	-
CLIQUE-LINK		46.3%	51.8%	57.1%	60.2%	-
GROUP		63.5%	73.4%	45.2%	65.5%	94.0%
GROUP (50% node coverage)		83.6%	77.2%	46.6%	82.0%	96.0%

Reference: E. Zheleva, L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. Tech Report.

What's the connection?

Inference \Rightarrow Identification

\neg Identification \Rightarrow Privacy

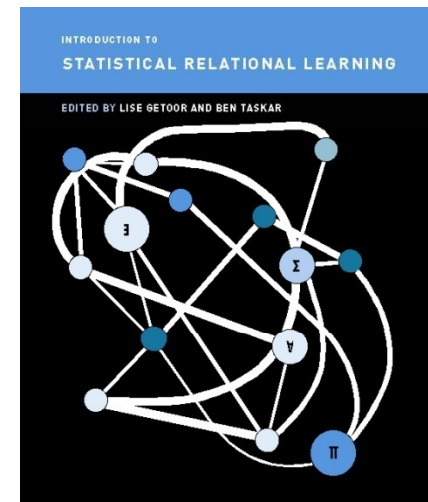
Bigger picture: What can tools can CS/ML offer for inference in complex networks?

Statistical Relational Learning (SRL)

- Methods that combine expressive knowledge representation formalisms such as relational and first-order logic with principled probabilistic and statistical approaches to inference and learning

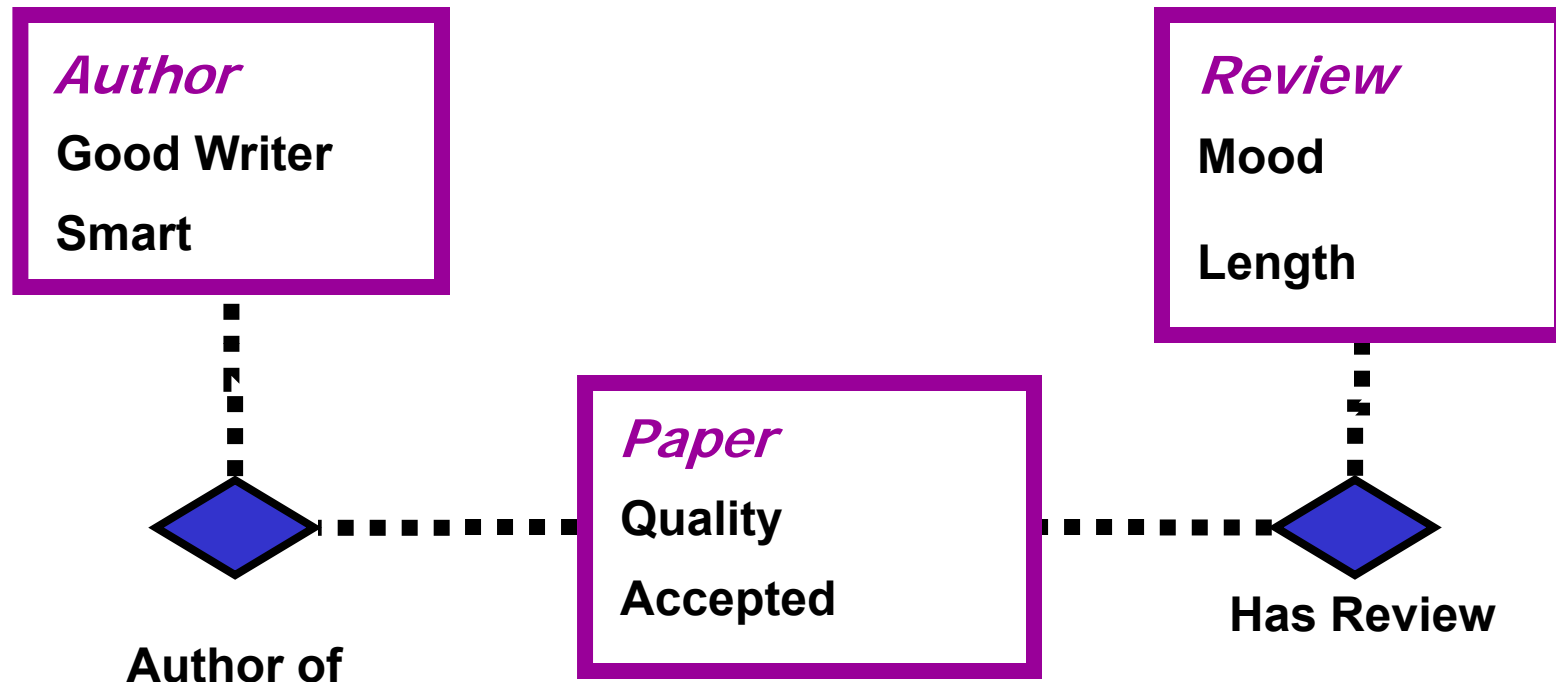


Dagstuhl April 2007



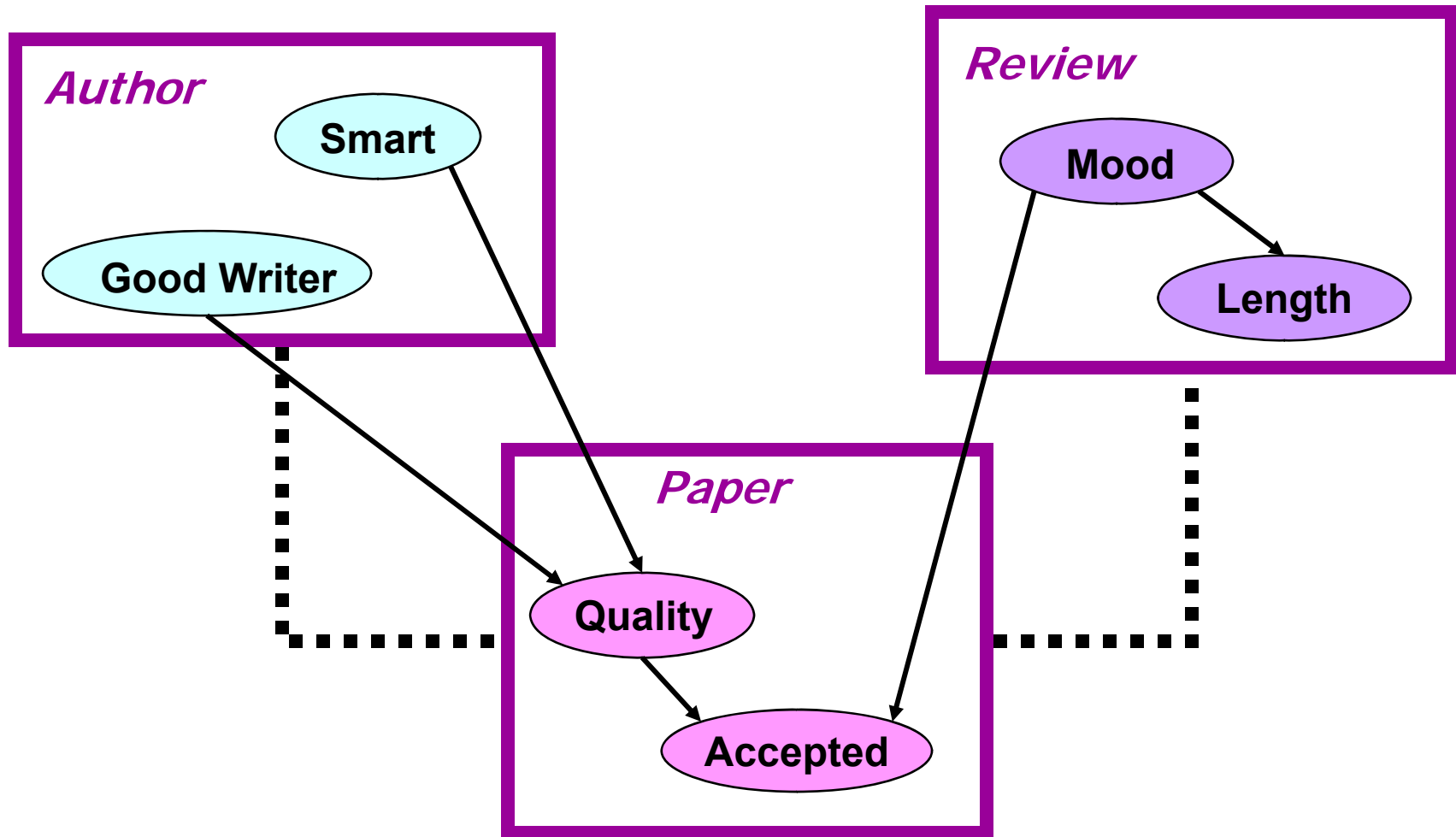
- Hendrik Blockeel, Mark Craven, James Cussens, Bruce D'Ambrosio, Luc De Raedt, Tom Dietterich, Pedro Domingos, Saso Dzeroski, Peter Flach, Rob Holte, Manfred Jaeger, David Jensen, Kristian Kersting, Heikki Mannila, Andrew McCallum, Tom Mitchell, Ray Mooney, Stephen Muggleton, Kevin Murphy, Jen Neville, David Page, Avi Pfeffer, Claudia Perlich, David Poole, Foster Provost, Dan Roth, Stuart Russell, Taisuke Sato, Jude Shavlik, Ben Taskar, Lyle Ungar and many others

Relational Schema

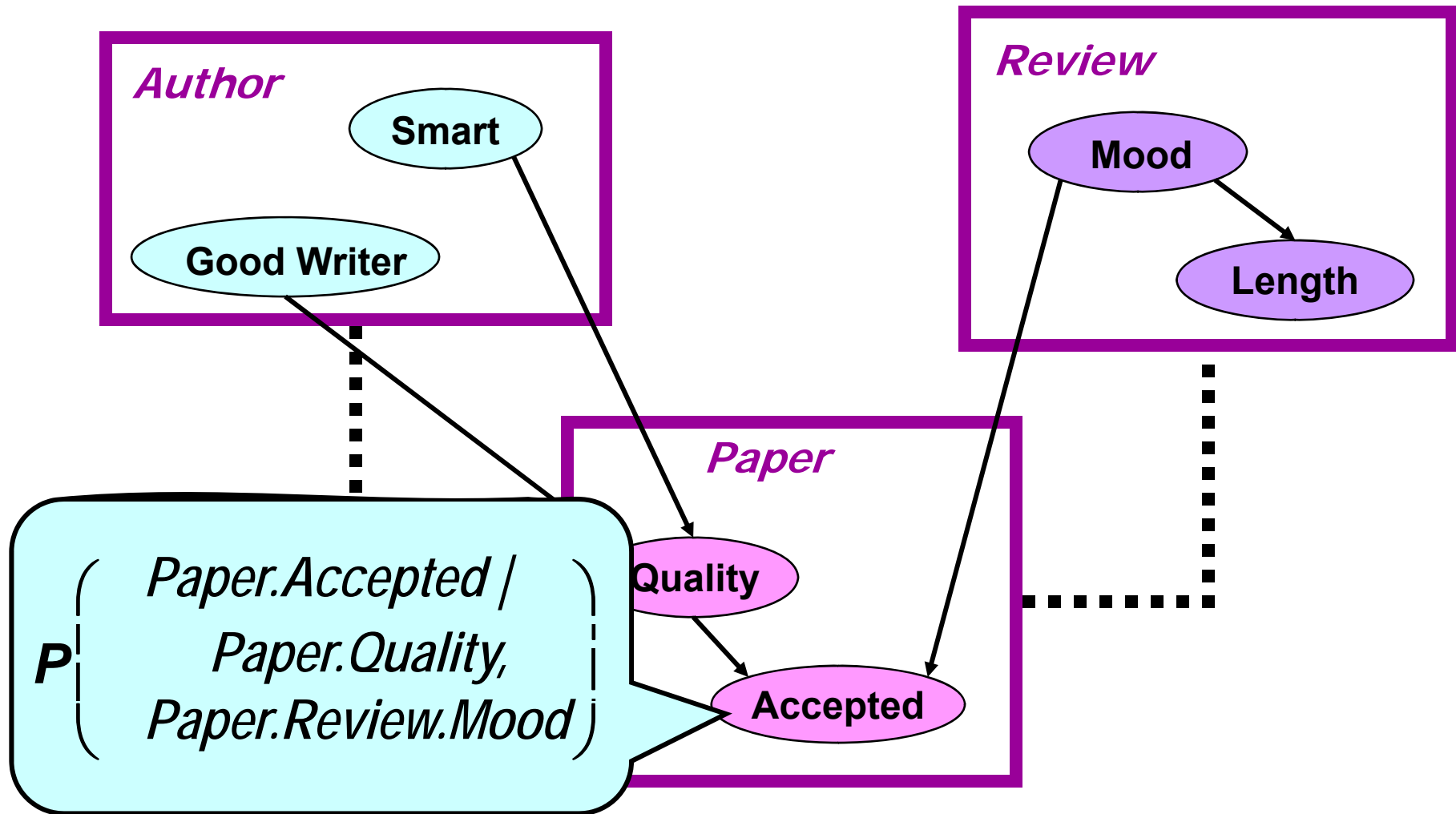


- o Describes the types of objects and relations in the database

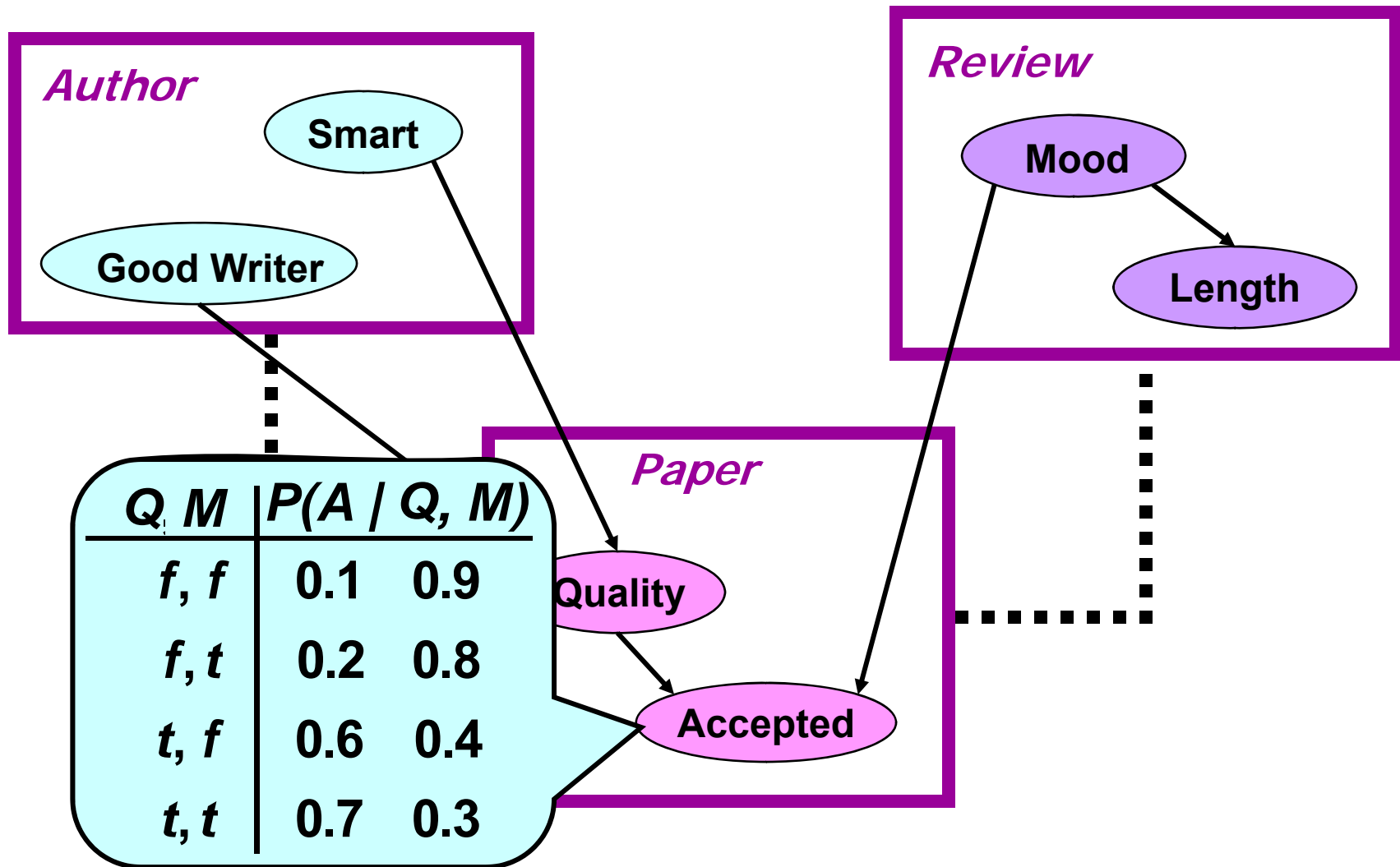
Probabilistic Relational Model



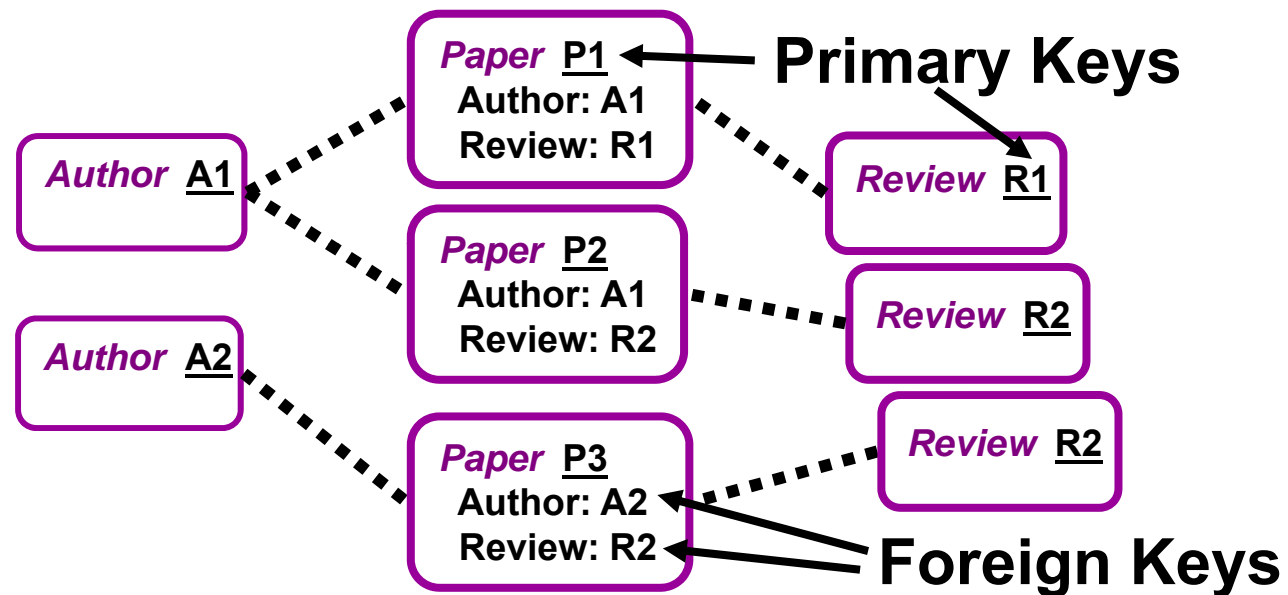
Probabilistic Relational Model



Probabilistic Relational Model



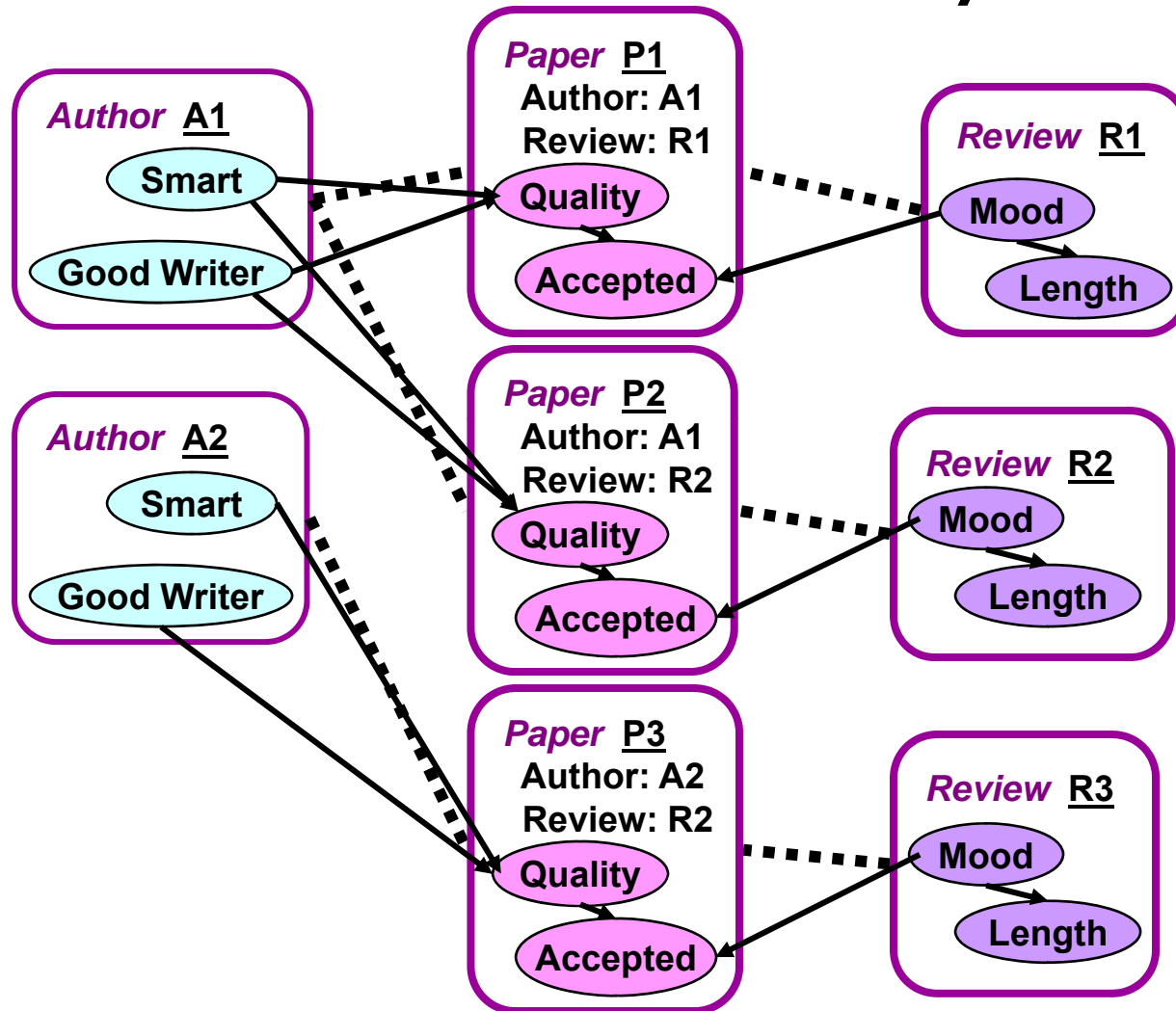
Relational Skeleton



Fixed relational skeleton σ :

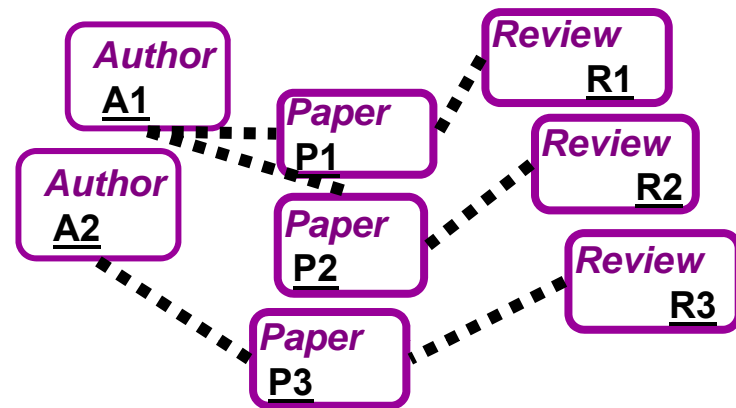
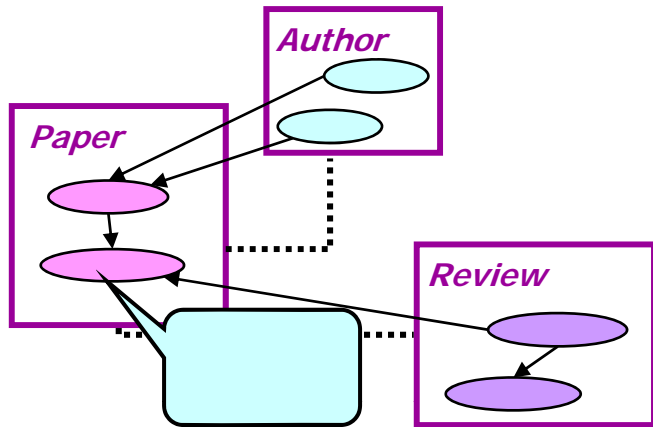
- set of objects in each class
- relations between them

PRM w/ Attribute Uncertainty



PRM defines distribution over instantiations of attributes

PRM with AU Semantics



PRM + relational skeleton σ =

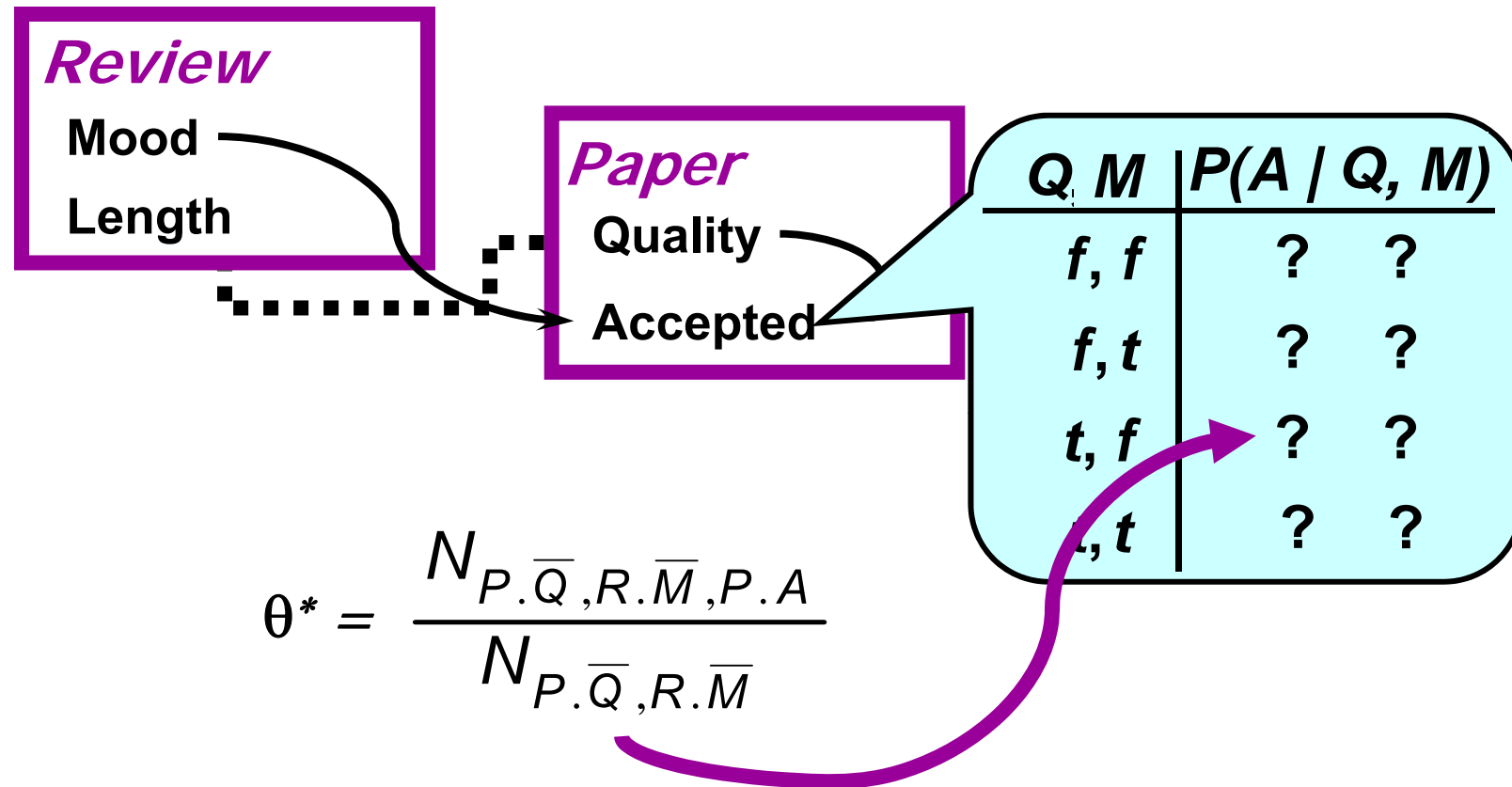
probability distribution over completions I :

$$P(I \mid \sigma, S, \theta) = \prod_{x \in \sigma} \prod_{x.A} P(x.A \mid \text{parents}_{S, \sigma}(x.A))$$

\nearrow
Objects

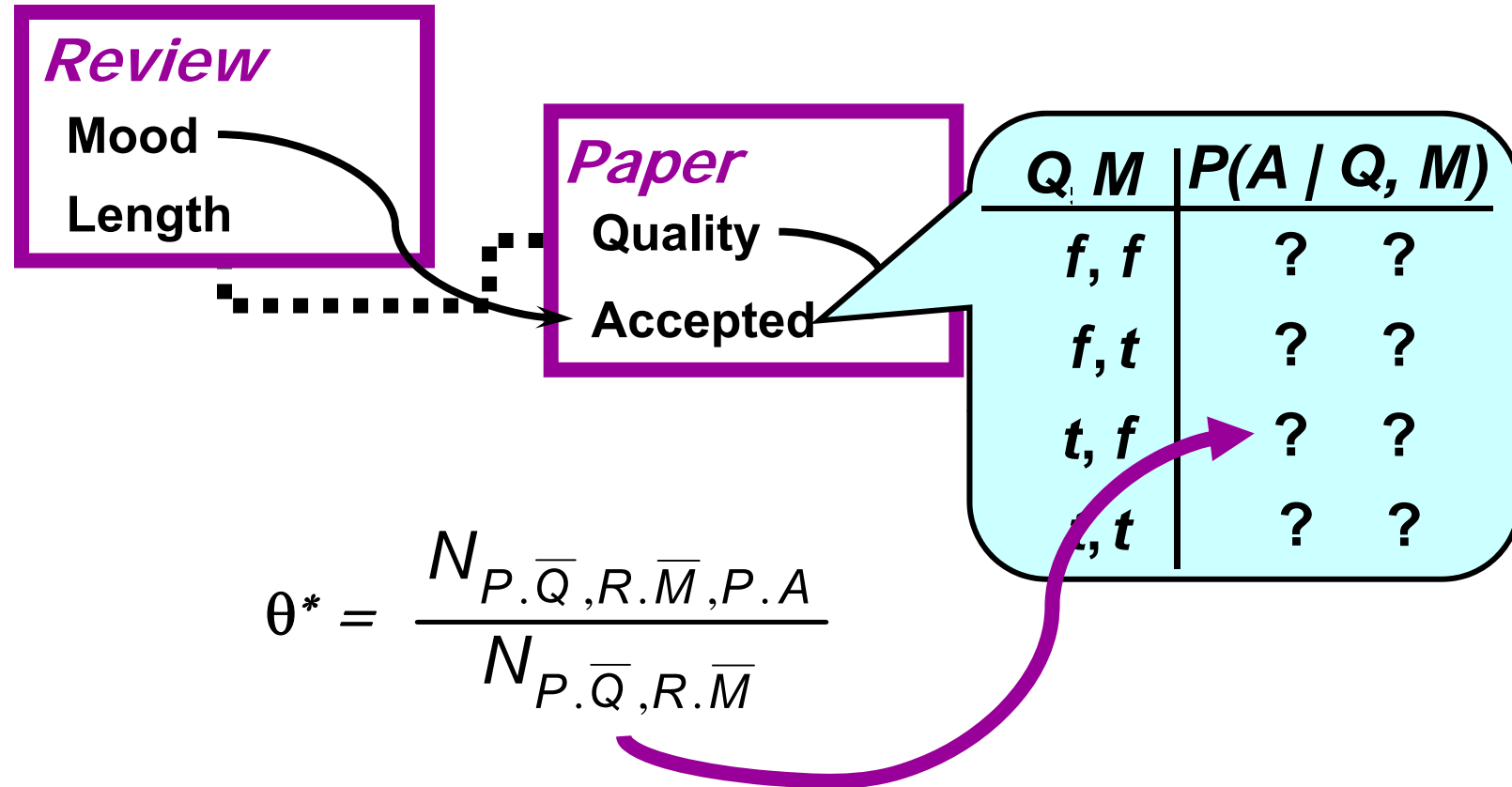
\nwarrow
Attributes

ML Parameter Estimation



where $N_{P.\bar{Q}, R.\bar{M}, P.A}$ is the number of accepted,
 low quality papers
 whose reviewer was in a poor mood

ML Parameter Estimation



Query for counts:

$$\text{Count} \left[\pi \begin{array}{l} P.Quality \\ R.Mood \\ P.Accepted \end{array} \left(\begin{array}{c} \text{Review table} \\ \bowtie \\ \text{Paper table} \end{array} \right) \right]$$

Structure Selection

- Idea:

- define scoring function
- do local search over legal structures

- Key Components:

- legal models
- scoring models
- searching model space

Structure Selection

- Idea:

- define scoring function
- do local search over legal structures

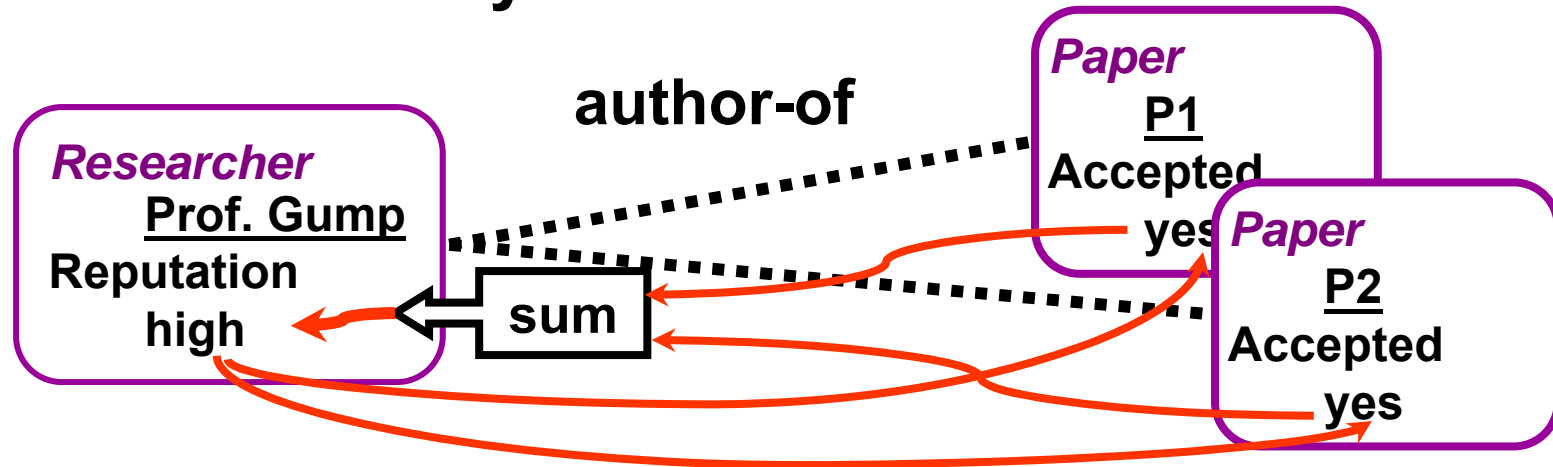
- Key Components:

- » legal models

- scoring models
- searching model space

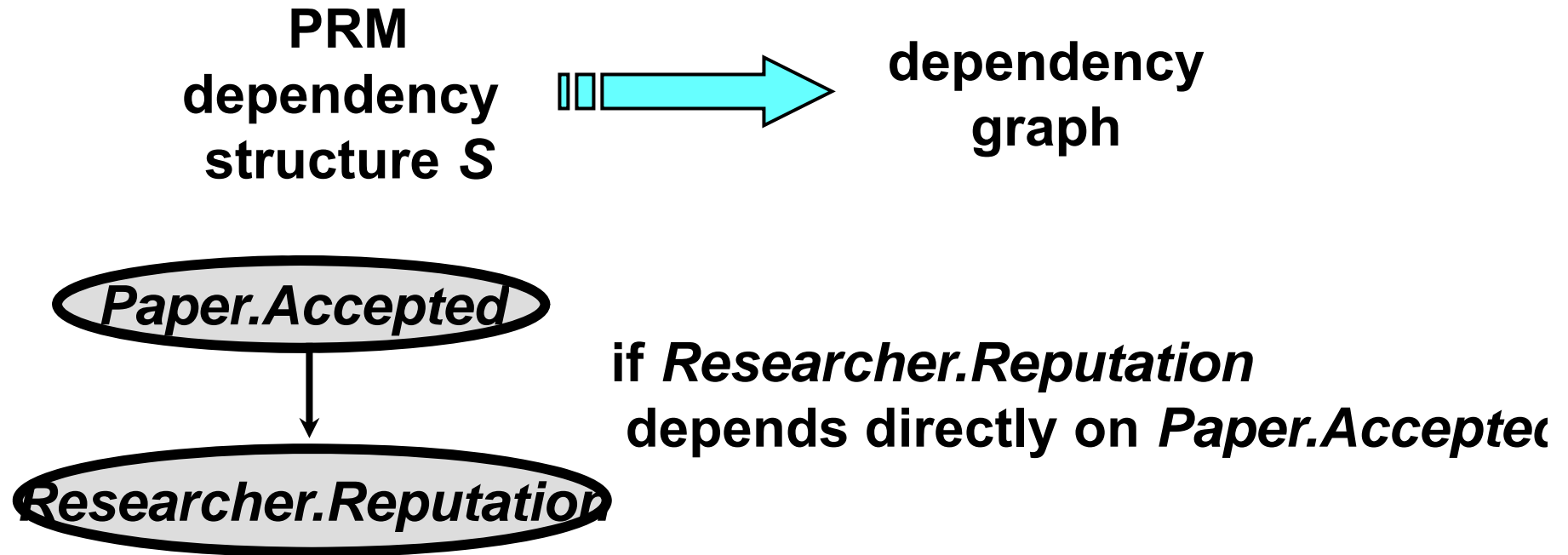
Legal Models

- PRM defines a coherent probability model over a skeleton σ if the dependencies between object attributes is acyclic



How do we guarantee that a PRM is acyclic for every skeleton?

Attribute Stratification



Attribute stratification:

dependency graph acyclic \Rightarrow acyclic for any σ

*Algorithm more flexible; allows certain
cycles along guaranteed acyclic
relations*

Structure Selection

- Idea:

- define scoring function
- do local search over legal structures

- Key Components:

- legal models
- » scoring models – same as BN
- searching model space

Structure Selection

- o Idea:

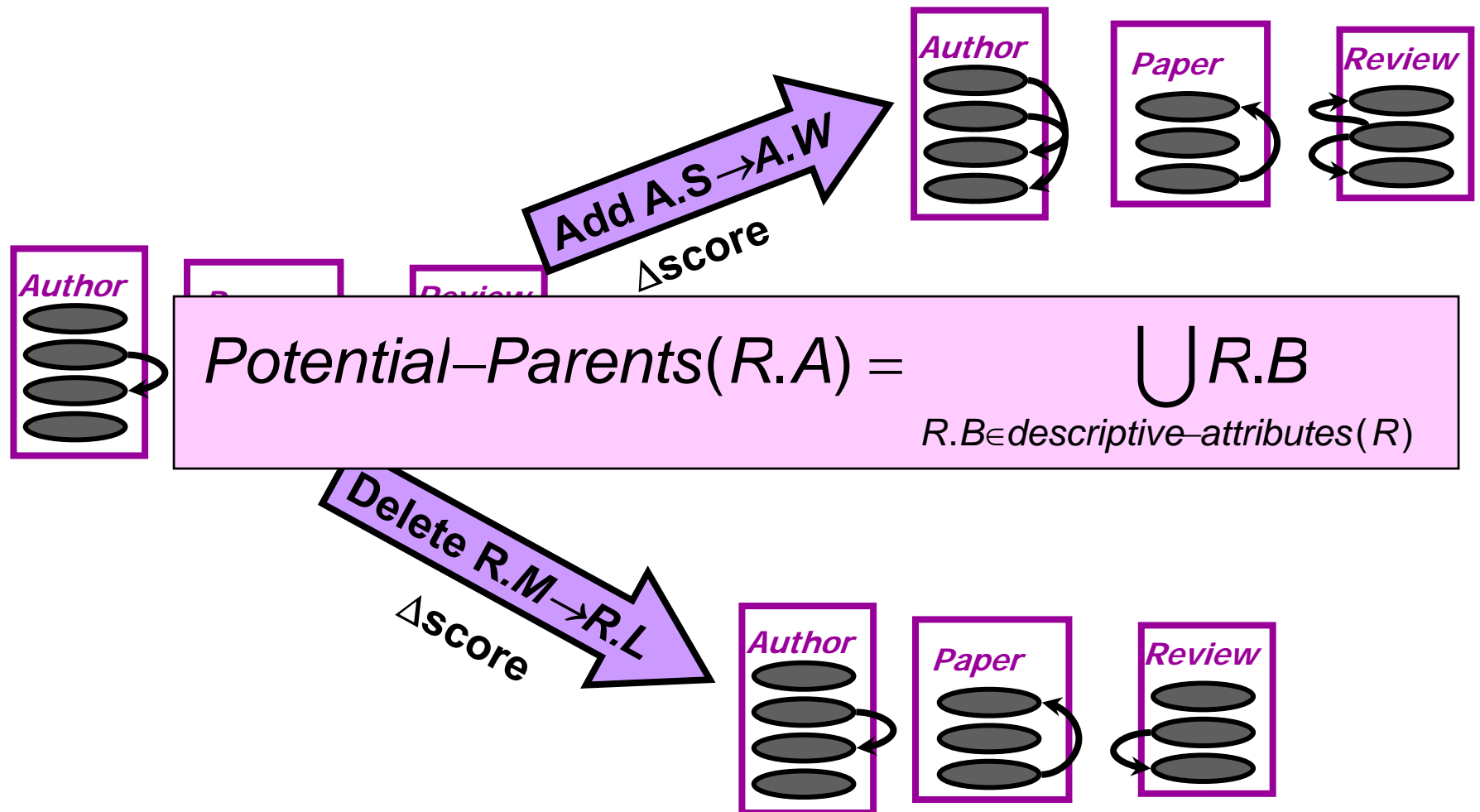
- define scoring function
- do local search over legal structures

- o Key Components:

- legal models
- scoring models
- » searching model space

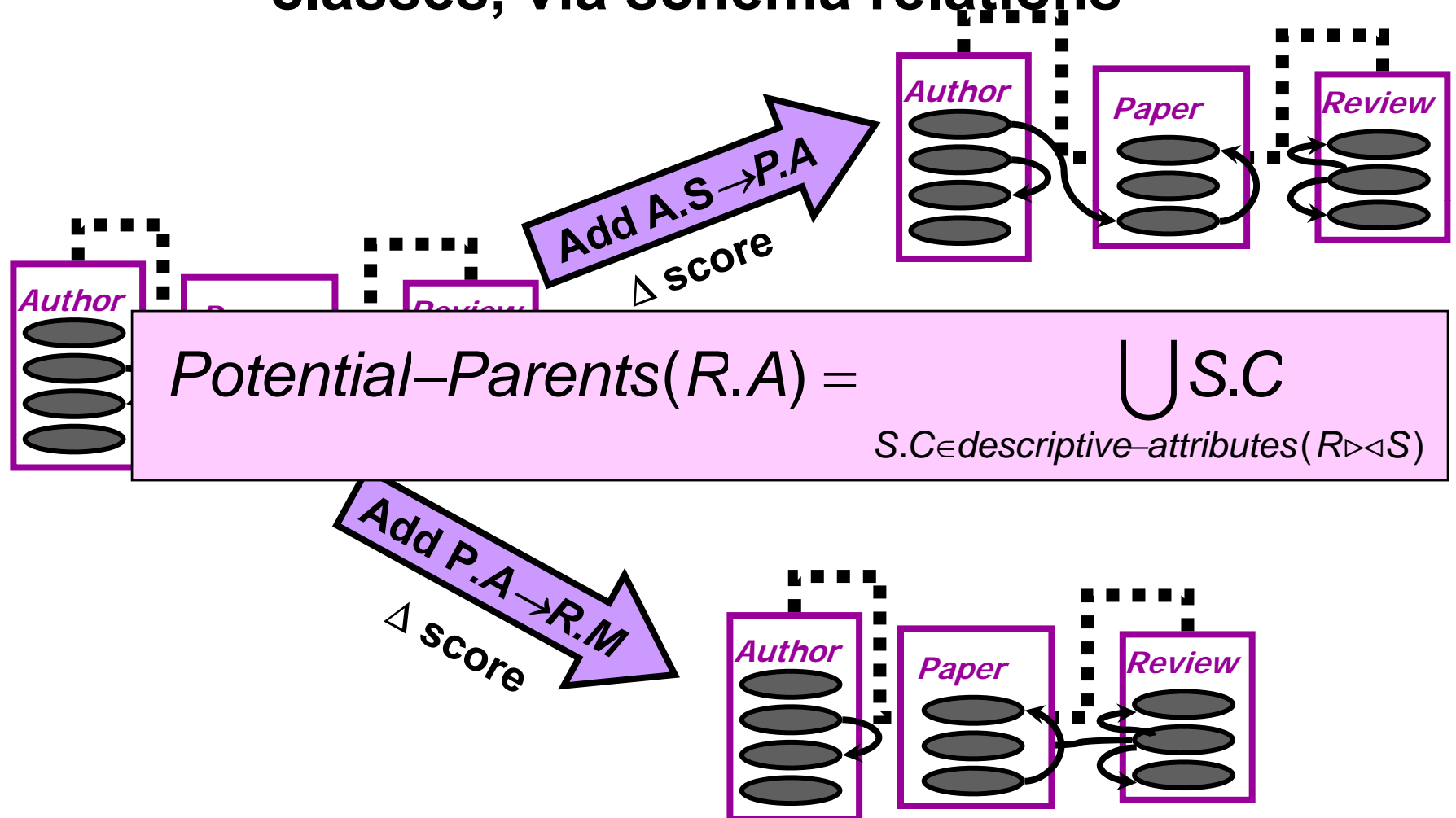
Searching Model Space

Phase 0: consider only dependencies within a class



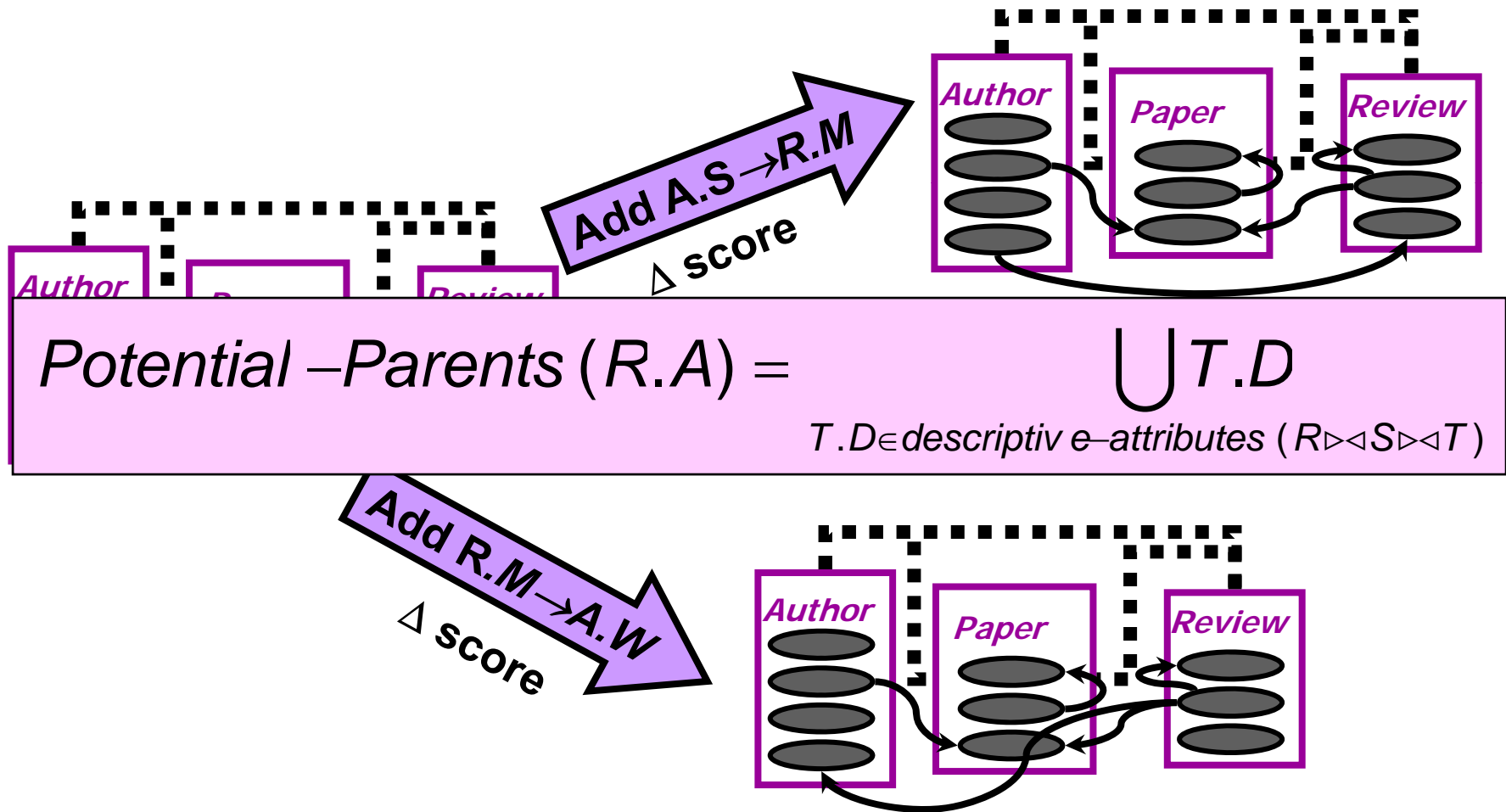
Phased Structure Search

Phase 1: consider dependencies from “neighboring classes, via schema relations



Phased Structure Search

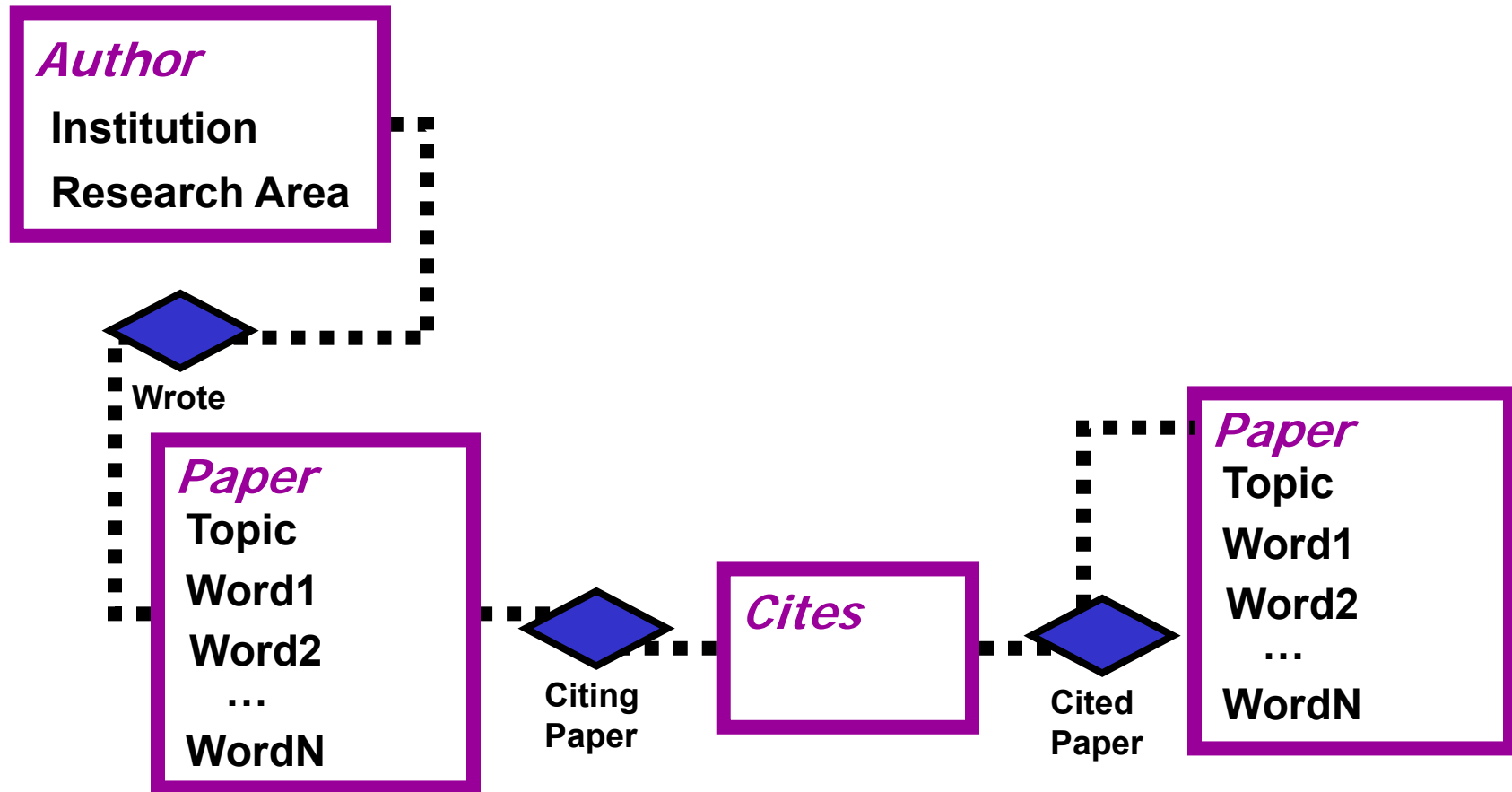
Phase 2: consider dependencies from “further” classes, via relation chains



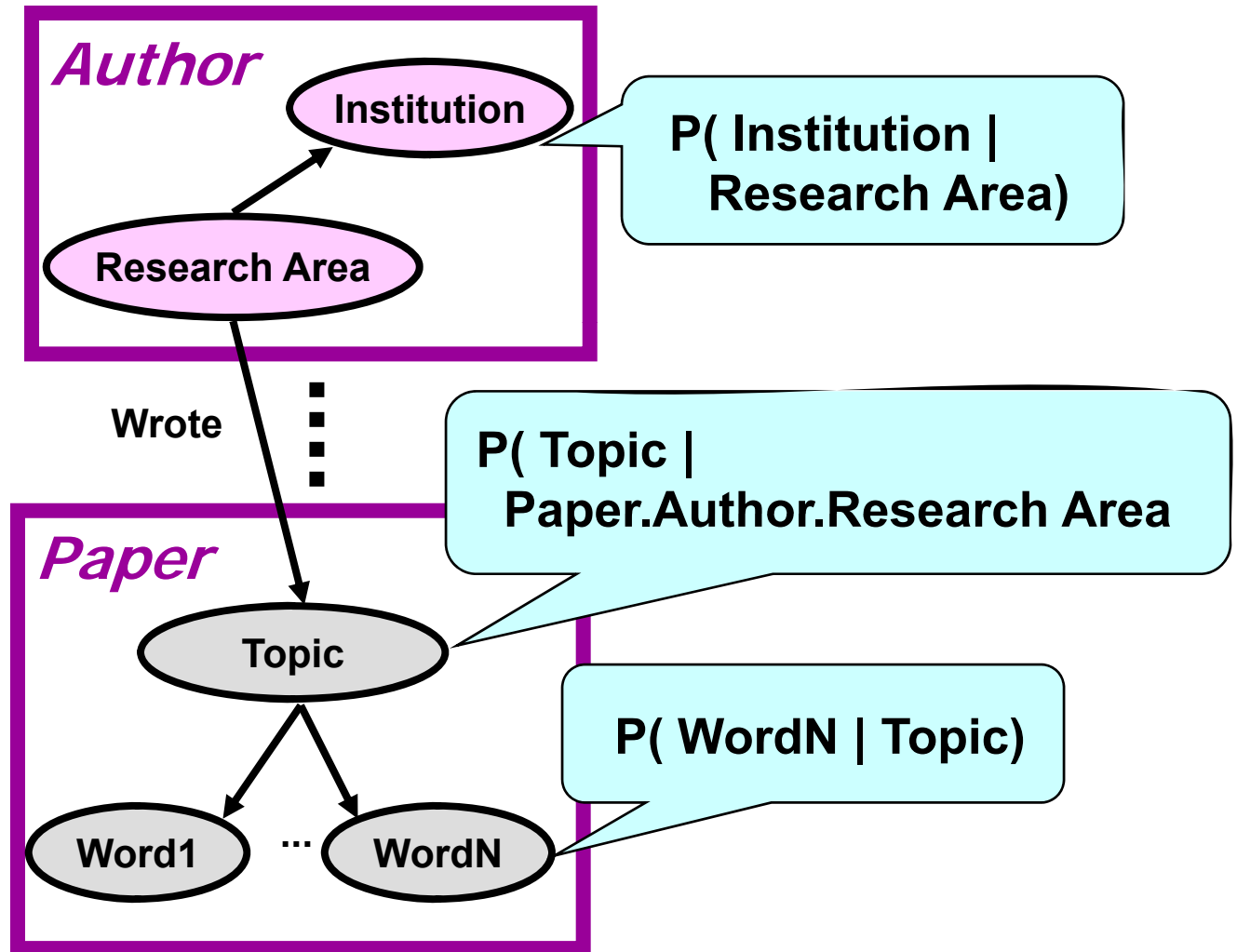
Structural Uncertainty

- Motivation: PRM with AU only well-defined when the skeleton structure is known
- May be uncertain about relational structure itself
- Construct probabilistic models of relational structure that capture **structural uncertainty**
- Mechanisms:
 - Reference uncertainty
 - Existence uncertainty
 - Number uncertainty
 - Type uncertainty
 - Identity uncertainty

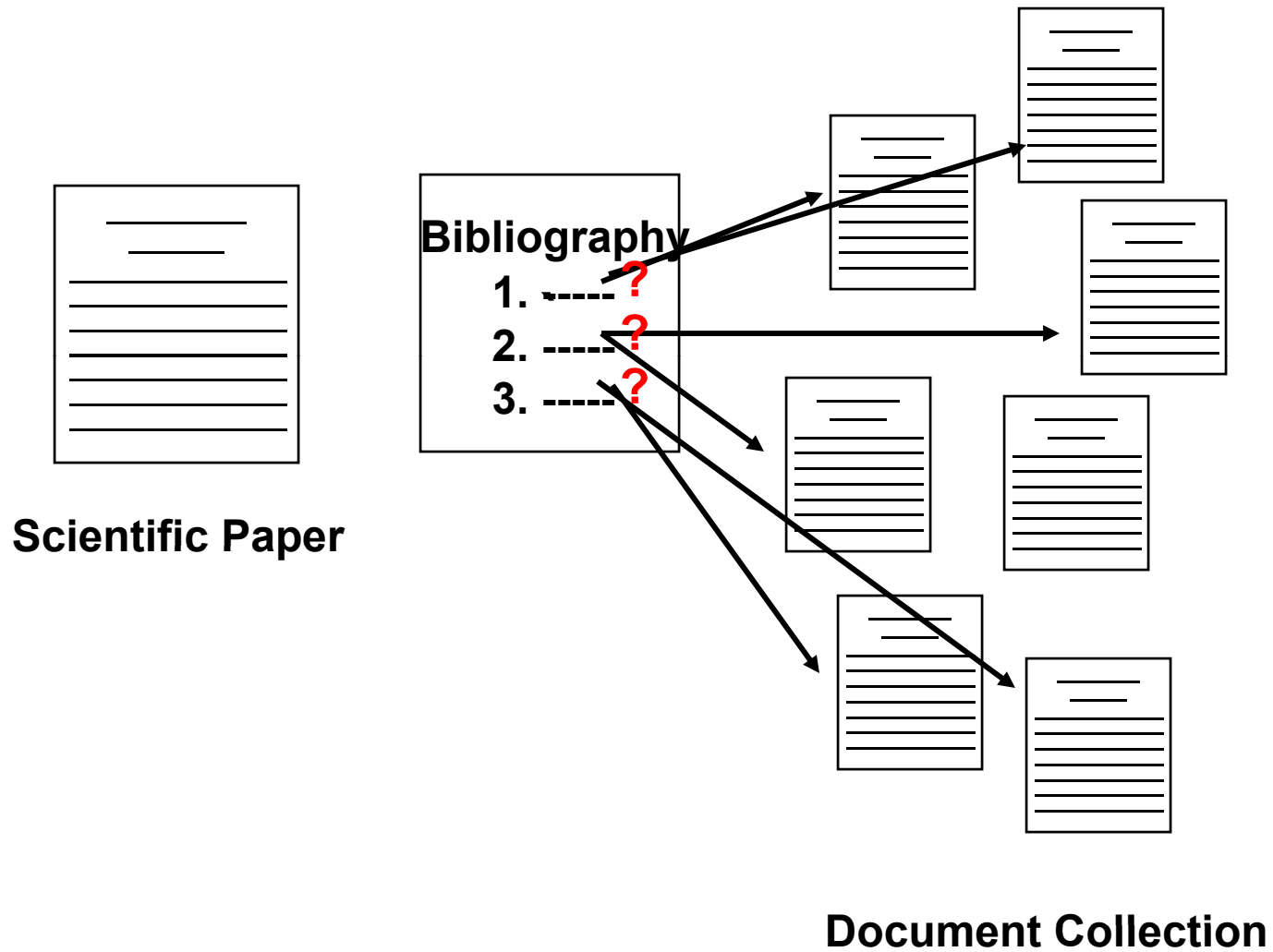
Citation Relational Schema



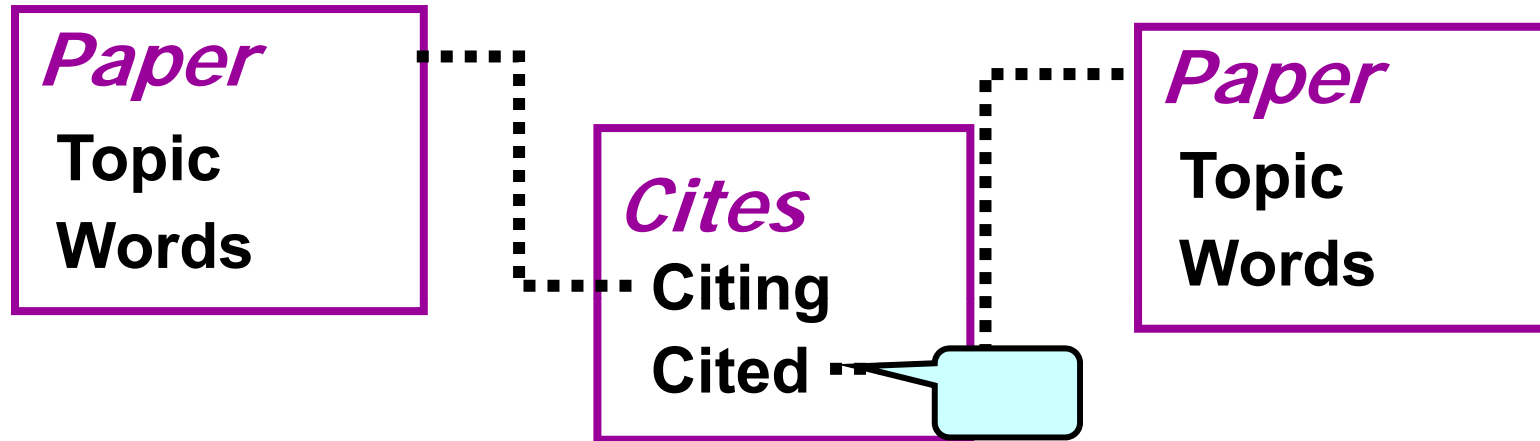
Attribute Uncertainty



Reference Uncertainty



PRM w/ Reference Uncertainty

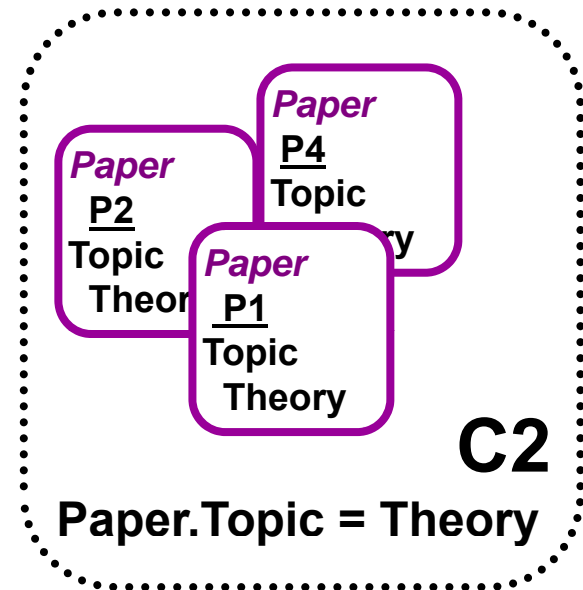
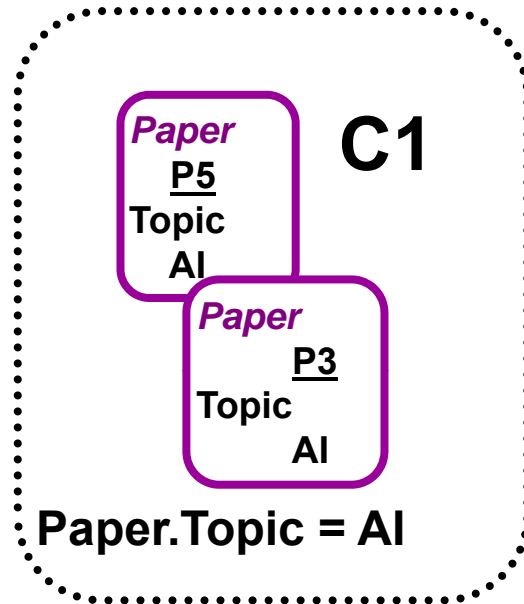
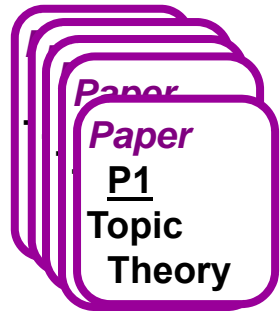


Dependency model for foreign keys

Naïve Approach: multinomial over primary key

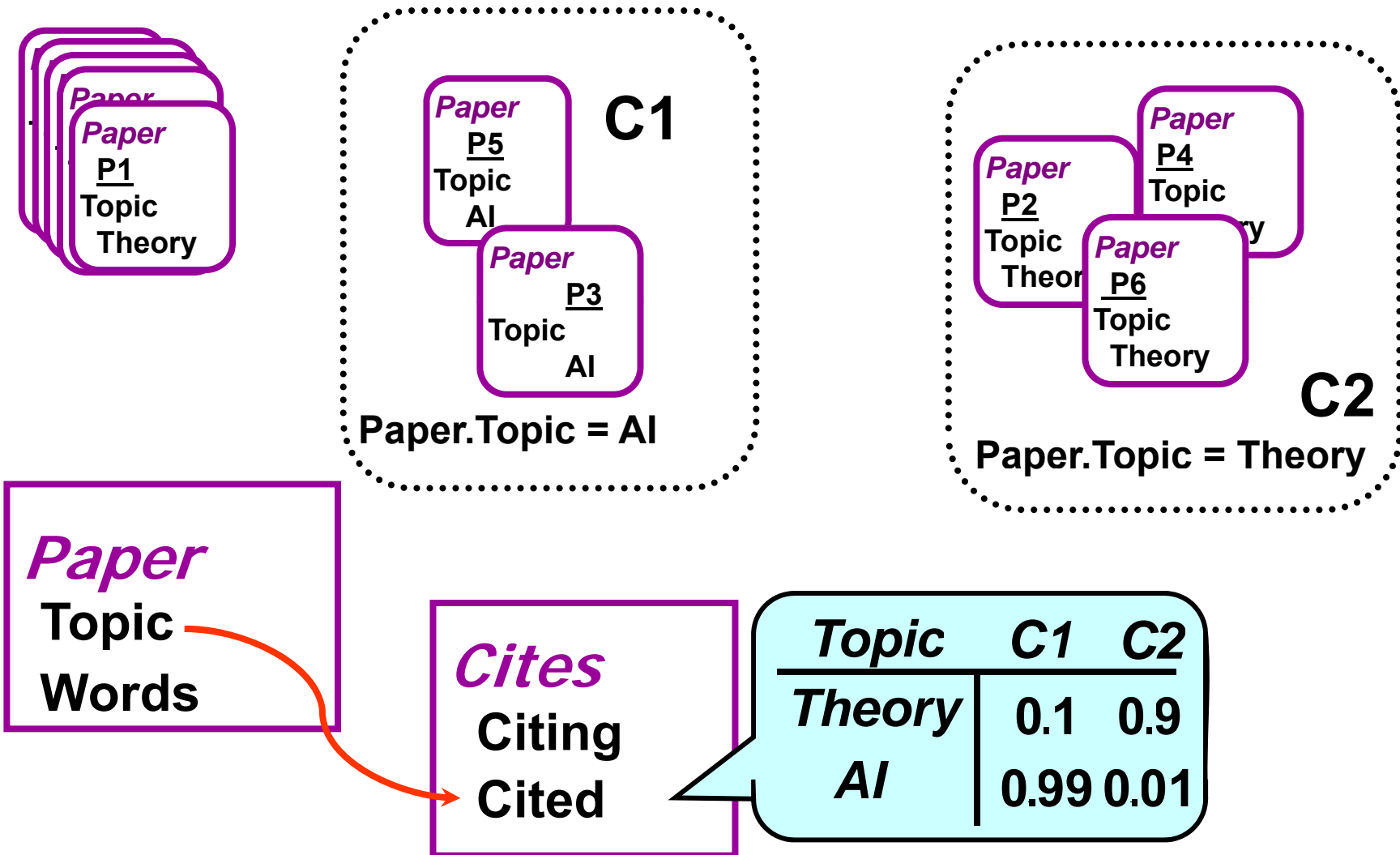
- noncompact
- limits ability to generalize

Reference Uncertainty Example

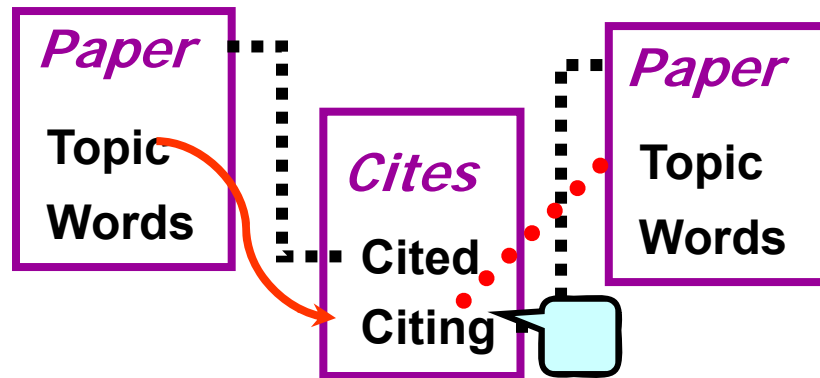


<i>Cites</i>	
Citing	Cited
C1	C2
0.3	0.7

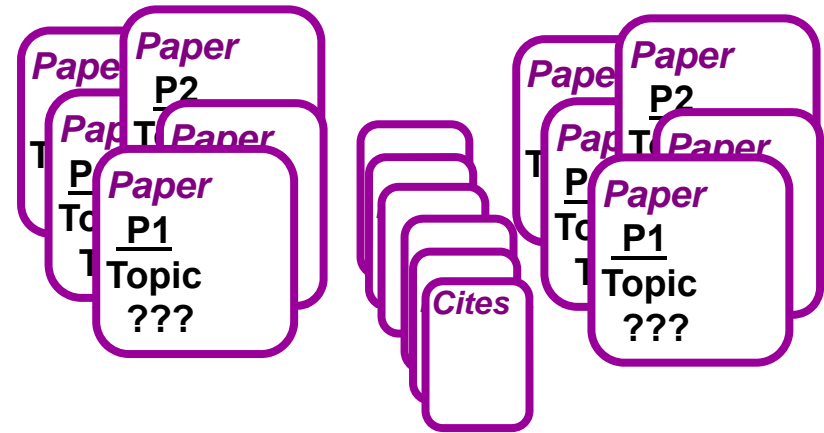
Reference Uncertainty Example



PRMs w/ RU Semantics



PRM RU



entity skeleton σ

PRM-RU + *entity skeleton* σ

\Rightarrow probability distribution over full instantiations I

Learning

PRMs w/ RU

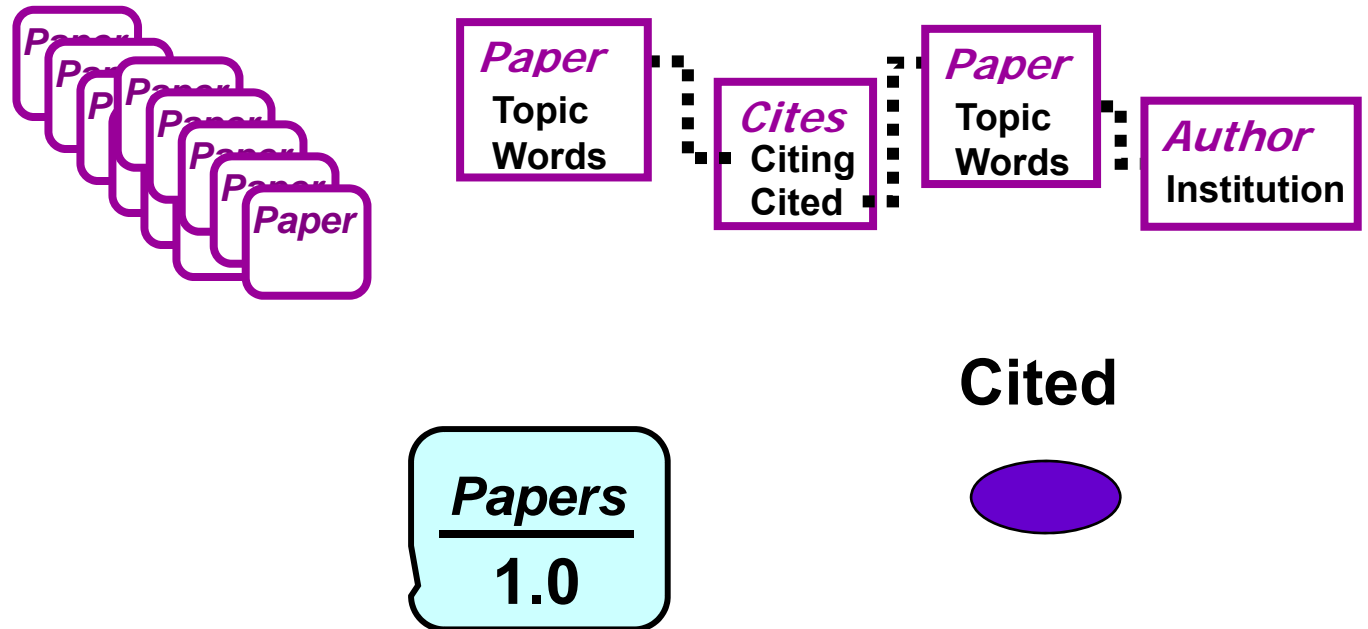
o **Idea:**

- define scoring function
- do phased local search over legal structures

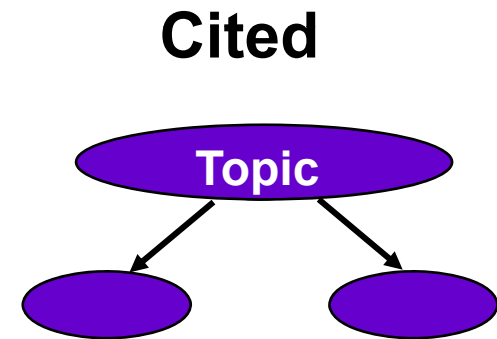
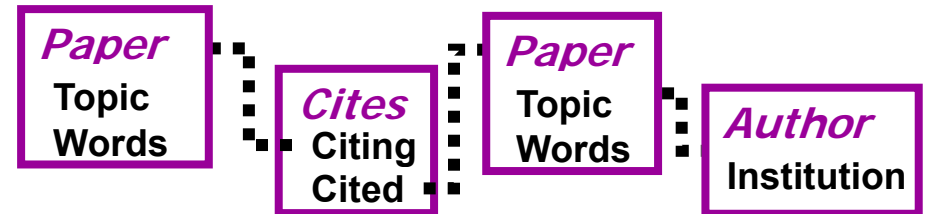
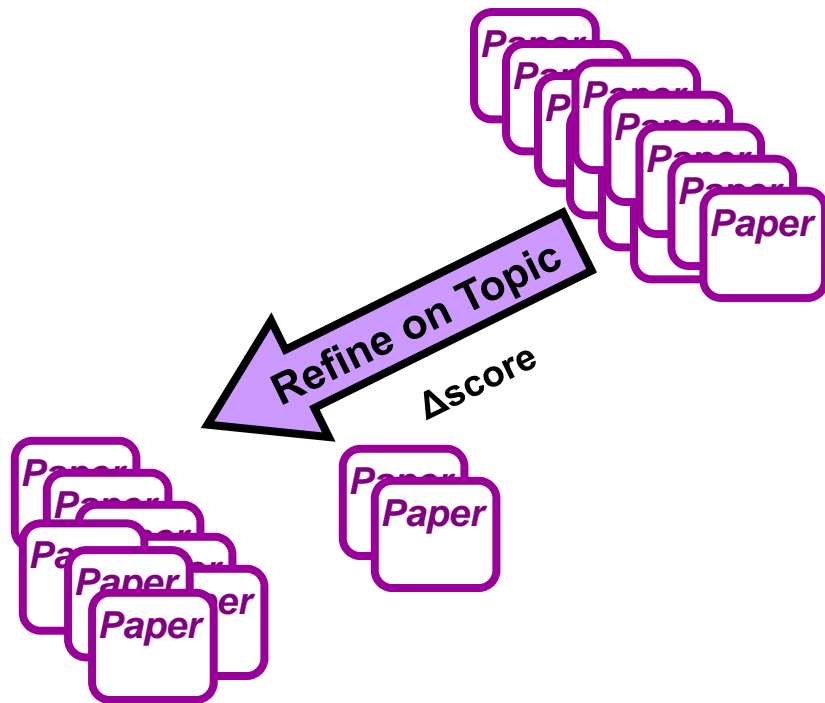
o **Key Components:**

- legal models
model new dependencies
- scoring models
unchanged
- searching model space
new operators

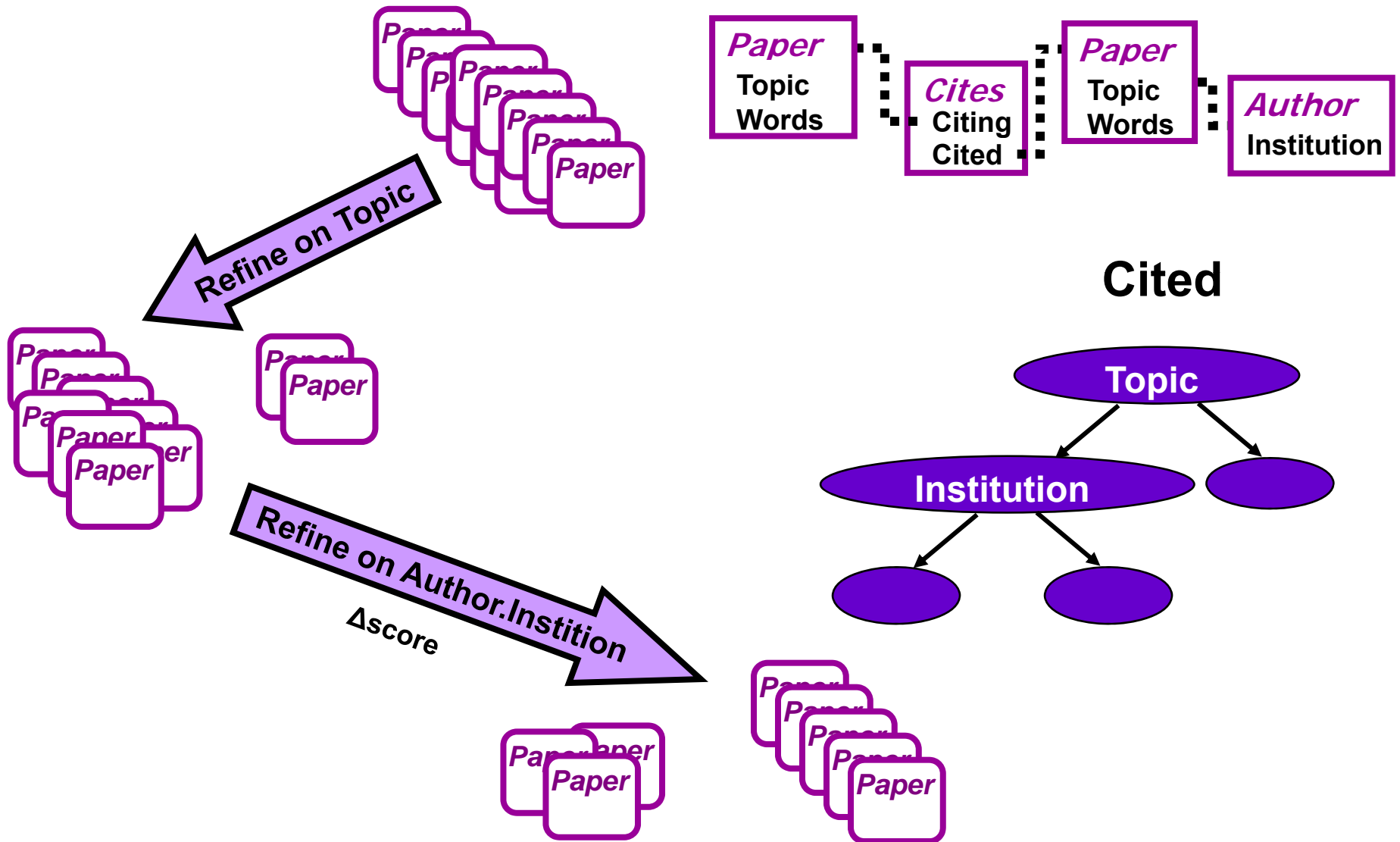
Structure Search



Structure Search: New Operators



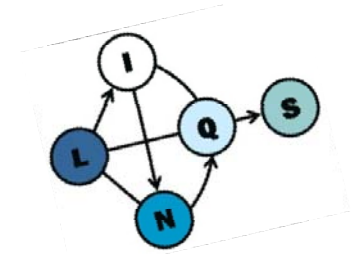
Structure Search: New Operators



Summary: Probabilistic Relational Models

- Focus on objects and relationships
 - what types of objects are there, and how are they related to each other?
 - how does a property of an object depend on other properties (of the same or other objects)?
- Representation support
 - Attribute uncertainty
 - Structural uncertainty
 - Class Hierarchies
- Efficient Inference and Learning Algorithms

LINQS Group @ UMD



- Members: myself, *Indrajit Bhattacharya*, Mustafa Bilgic, *Lei Guang*, Sam Huang, *Rezarta Islamaj*, *Hyunmo Kang*, Louis Licamele, *Qing Lu*, Walaa El-Din Mustafa, Galileo Namata, Barna Saha, Prithivaraj Sen, Vivek Sehgal, Hossam Sharara, Elena Zheleva



Thanks!

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

Work sponsored by the National Science Foundation,
Google, Microsoft, KDD program,
National Geospatial Agency, Army Research Office

