Probabilistic Models of Relational Domains

Daphne Koller

Stanford University

Overview

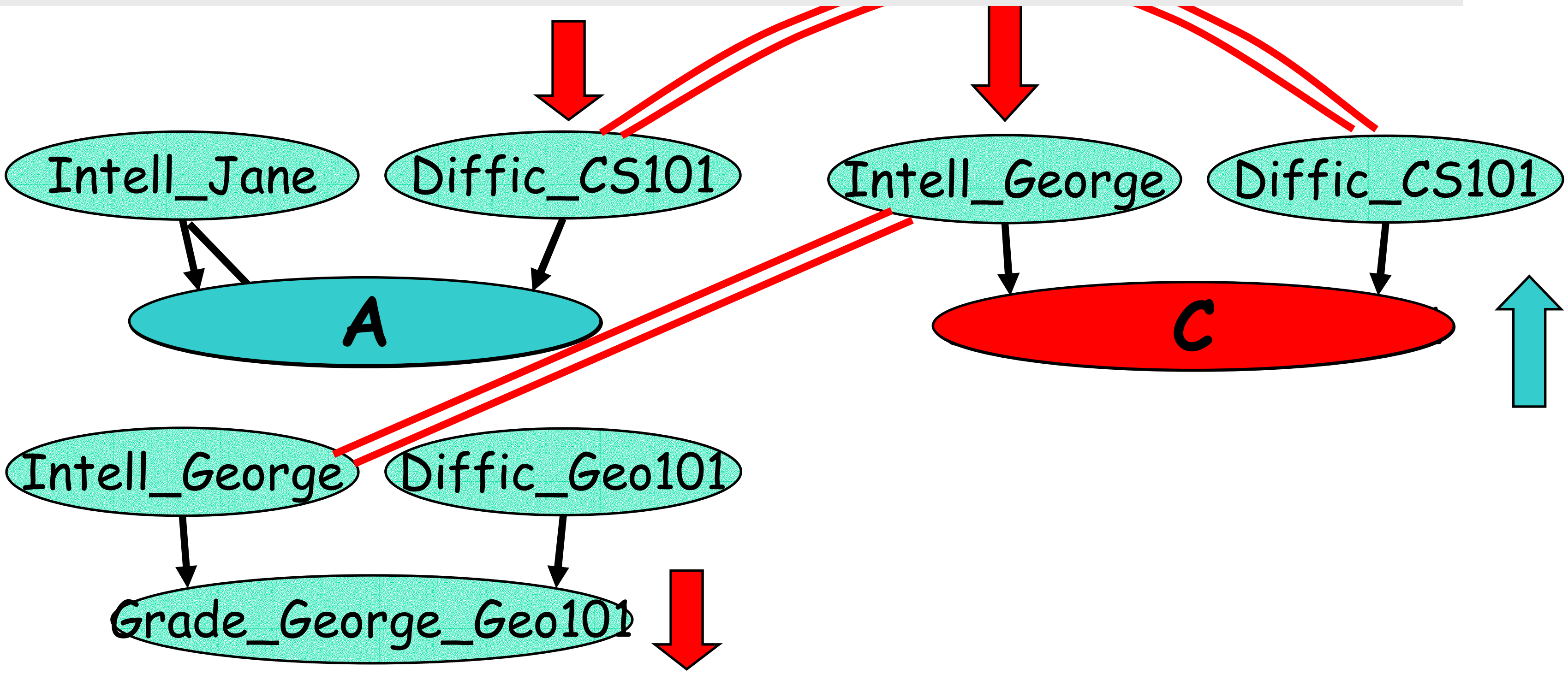
- Relational logic
- Probabilistic relational models
- Inference in PRMs
- Learning PRMs
- Uncertainty about domain structure
- Summary

Attribute-Based Models & Relational Models

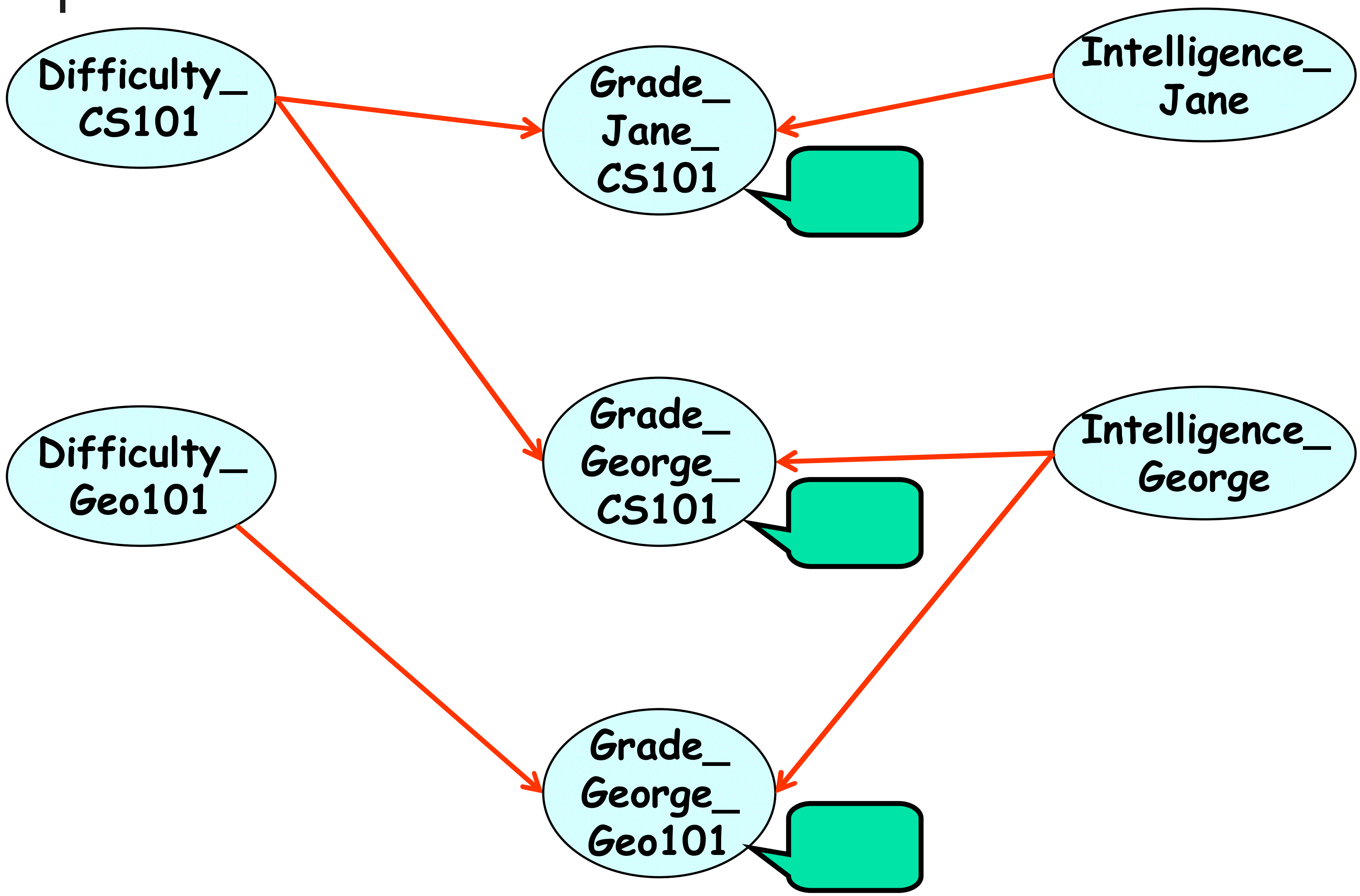
Bayesian Networks: Problem

- Bayes nets use attribute-based representation
- Real world has objects, related to each other

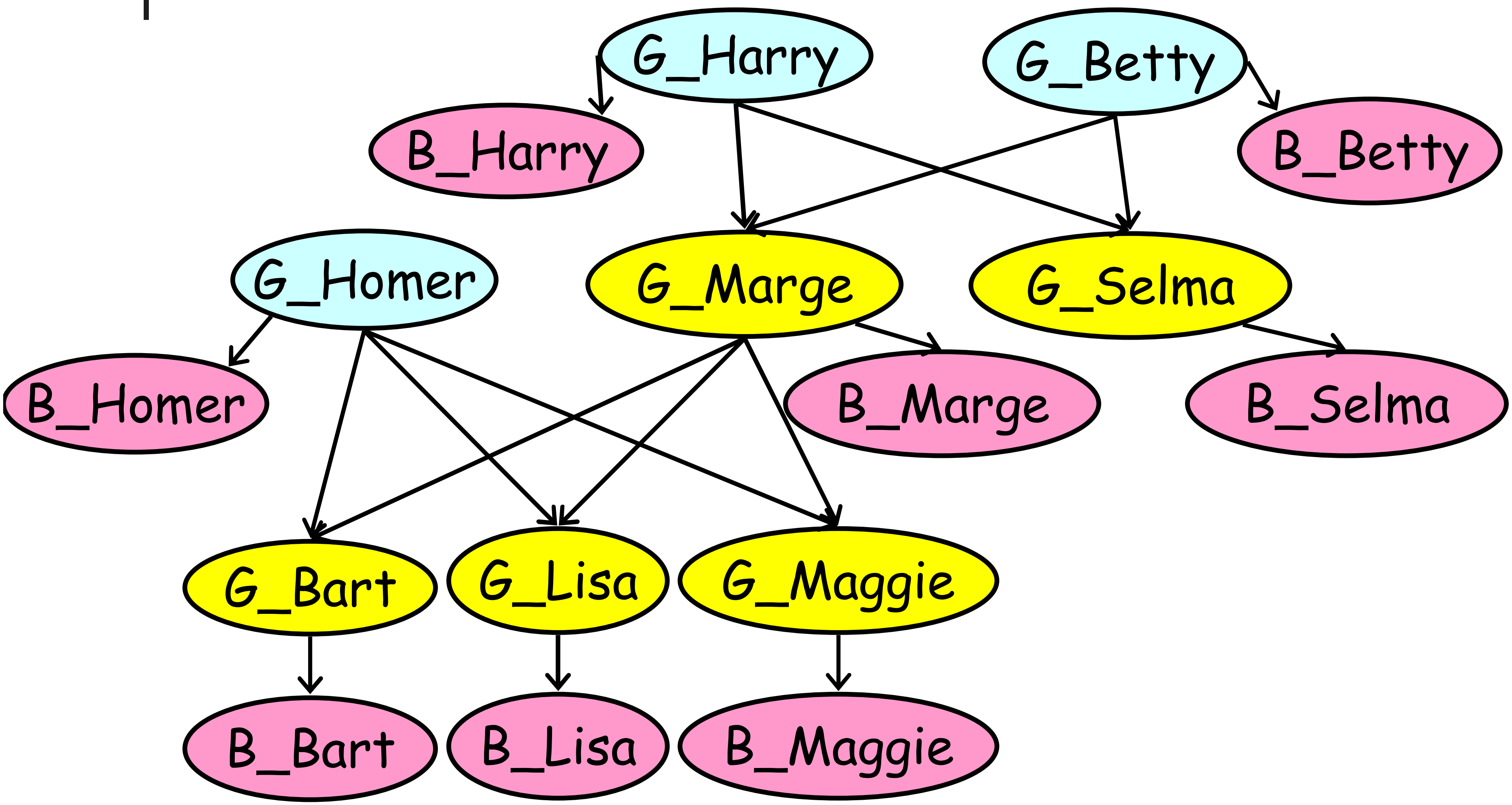
These "instances" are *not* independent



The University BN



The Genetics BN



G = genotype
B = bloodtype

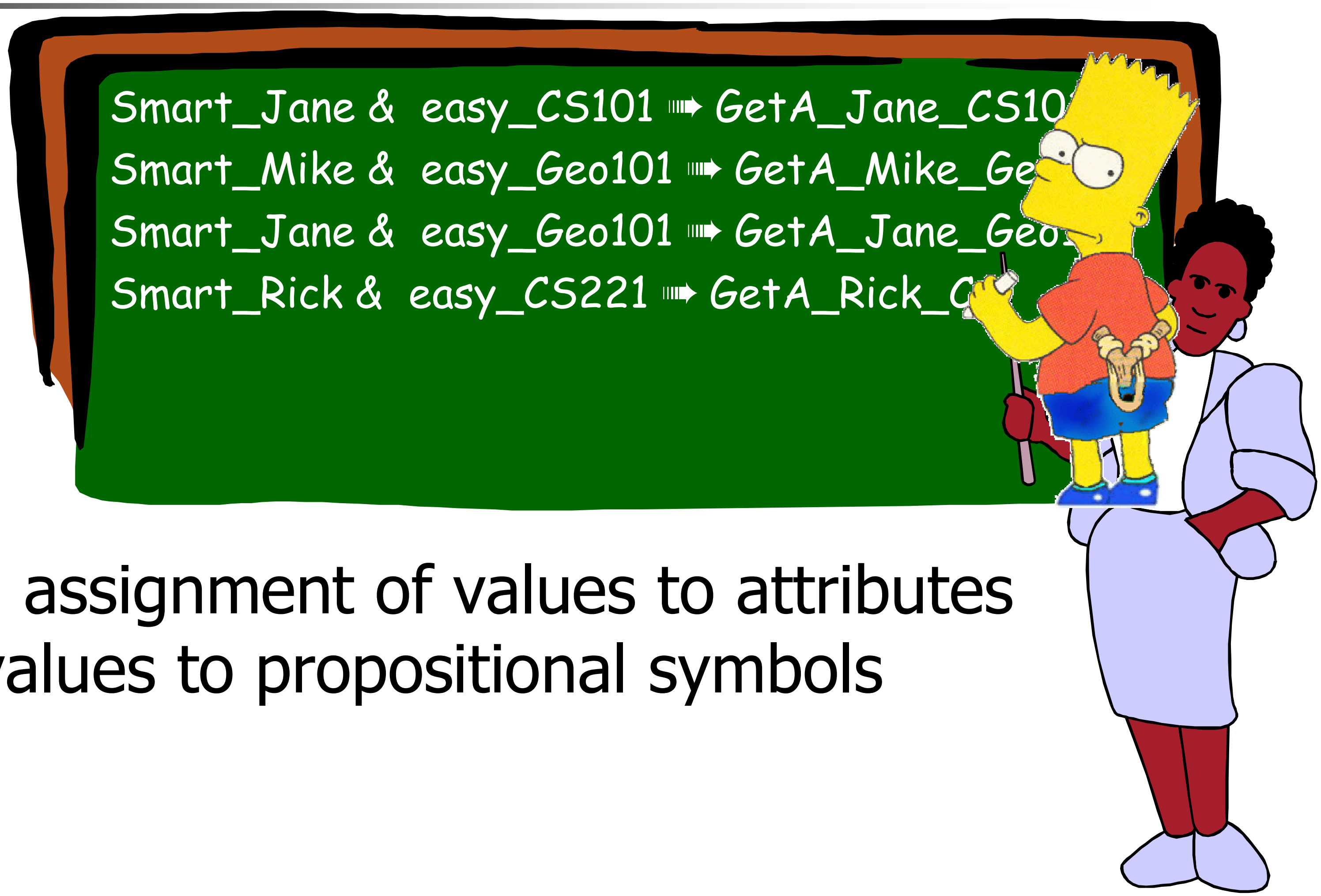
Simple Approach

- Obvious solution:
  - Use a graphical model with shared parameters
- Nodes share not only parameters, but also local dependency structure
- Want to enforce this constraint:
  - For human knowledge engineer
  - For network learning algorithm

Simple Approach II

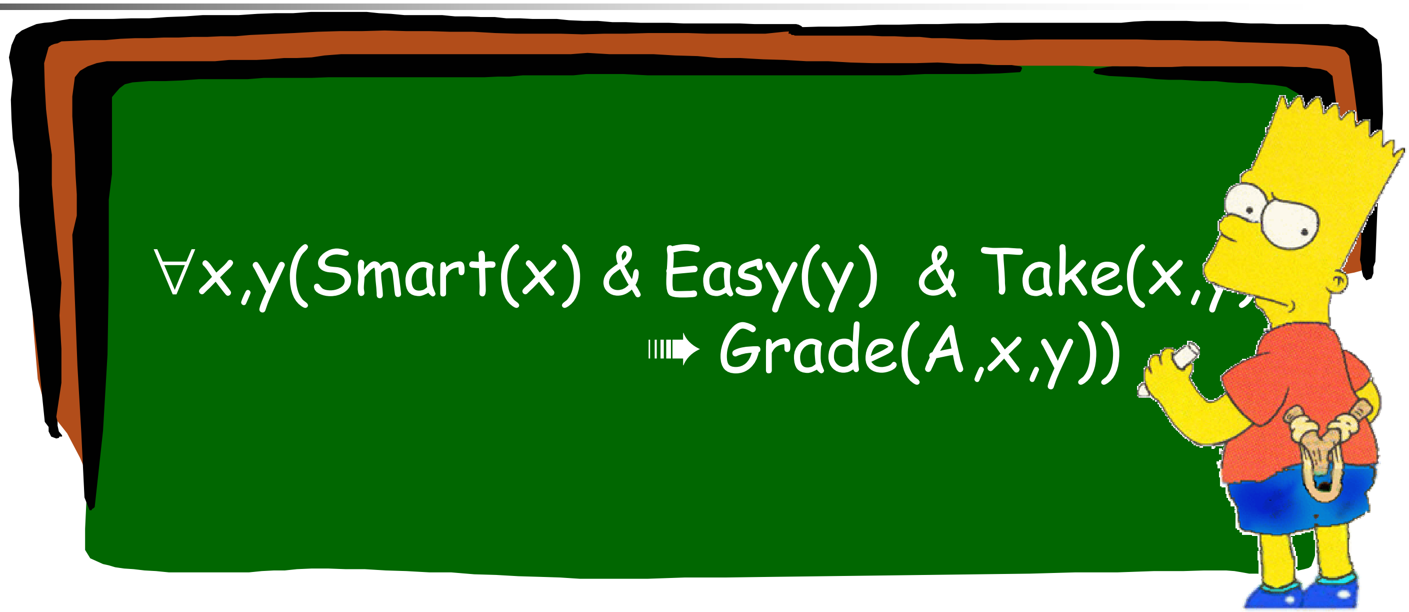
- How do we specify shared structure across different nodes?
  - Each person depends on his mother
  - But different people have different mothers
  - How do we specify the mapping
- We can write a special-purpose program for each domain:
  - genetic inheritance (family tree imposes constraints)
  - university (course registrations impose constraints)
- Is there something more general?

Attribute-Based Worlds



- World = assignment of values to attributes / truth values to propositional symbols

Object-Relational Worlds



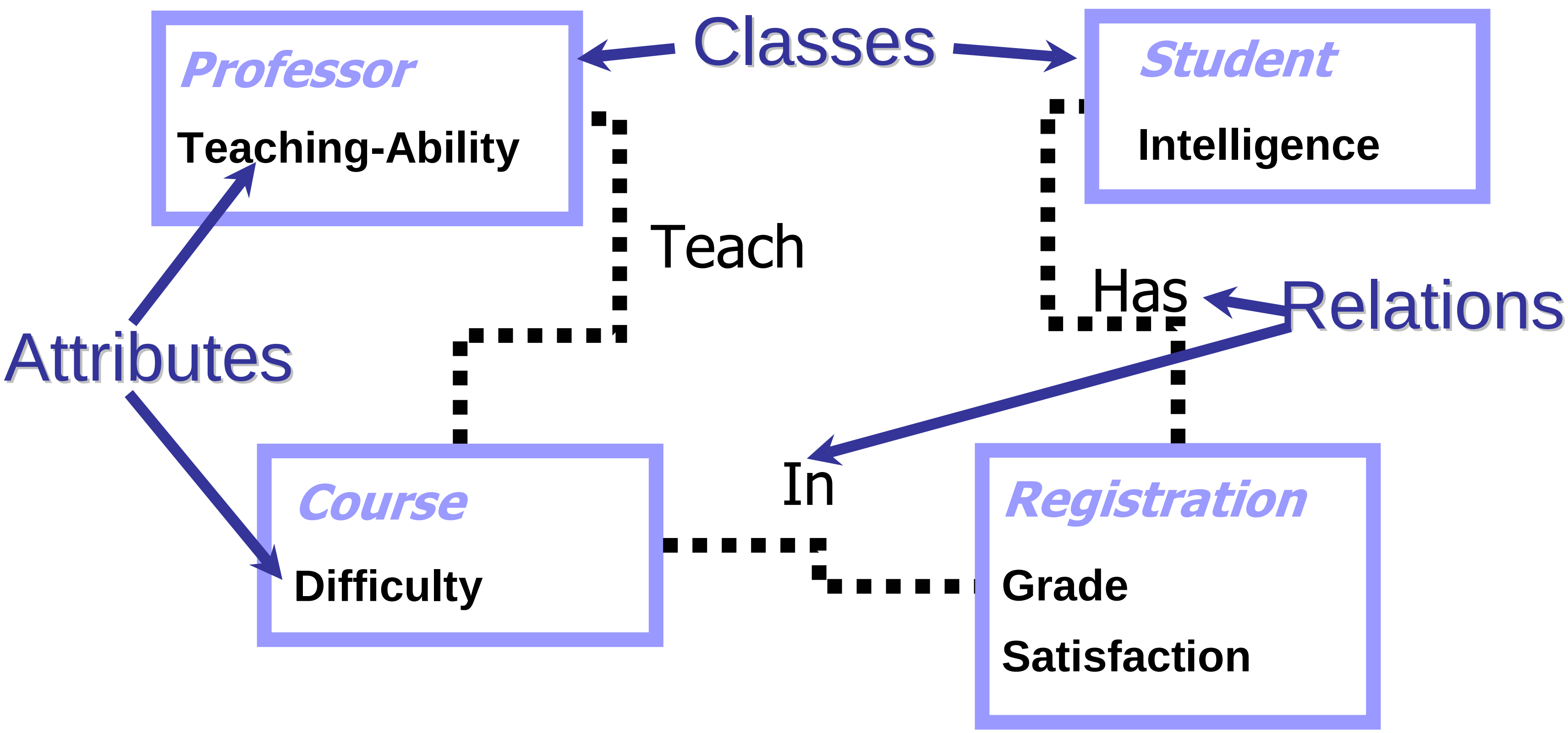
- World = relational interpretation:
  - Objects in the domain
  - Properties of these objects
  - Relations (links) between objects

Relational Logic

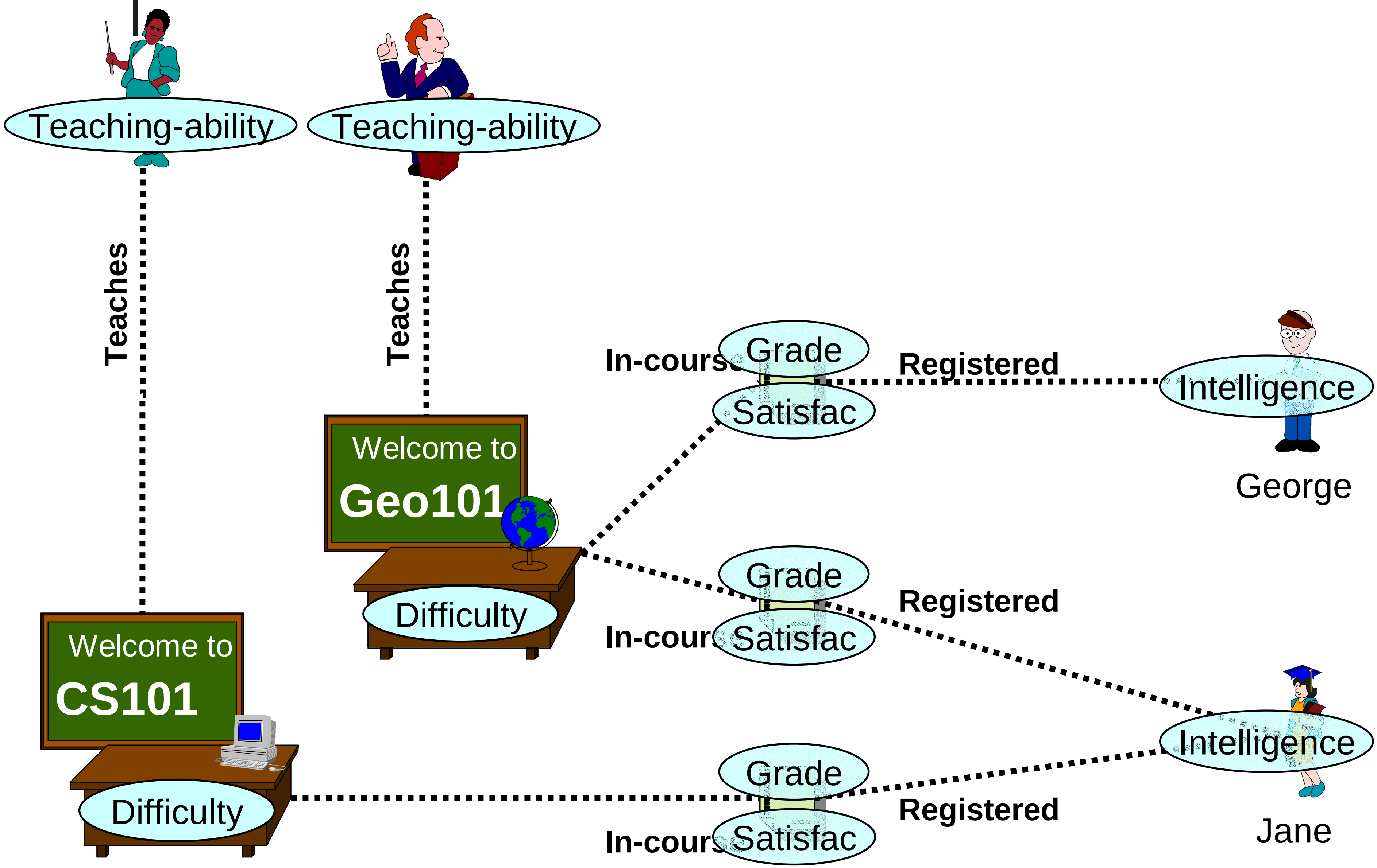
- General framework for representing:
  - objects & their properties
  - classes of objects with same model
  - relations between objects
- Represent a model at the template level, and apply it to an infinite set of domains
- Given finite domain, each instantiation of the model is propositional, but the template is not

Relational Schema

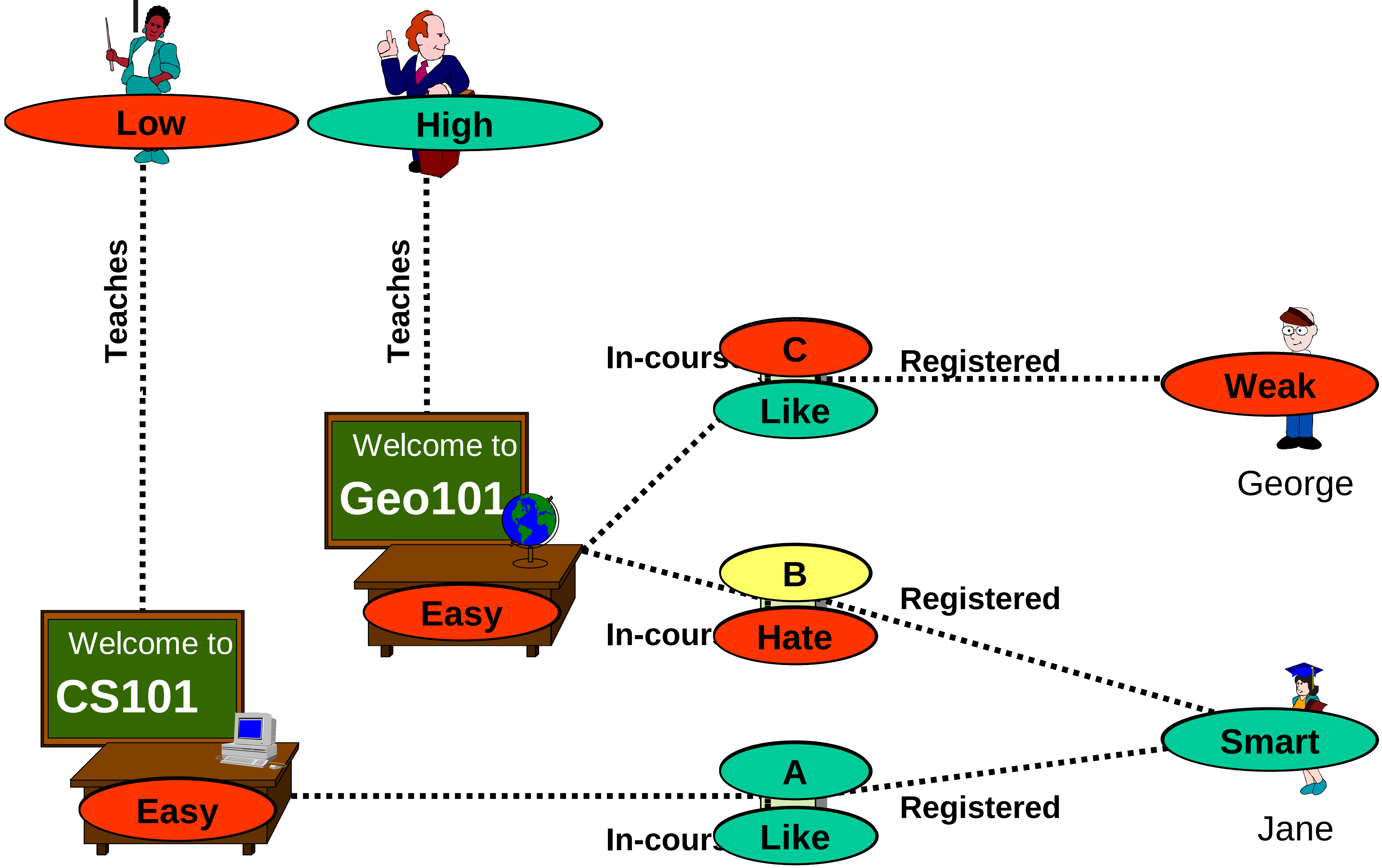
- Specifies types of objects in domain, attributes of each type of object & types of relations between objects



St. Nordaf University



Possible Worlds



Relational Logic: Summary

- Vocabulary:
  - Classes of objects:
    - Person, Course, Registration, …
  - Individual objects in a class:
    - George, Jane, …
  - Attributes of these objects:
    - George.Intelligence, Reg1.Grade
  - Relationships between these objects
    - Of(Reg1,George), Teaches(CS101,Smith)
- A *world* specifies:
  - A set of objects, each in a class
  - The values of the attributes of all objects
  - The relations that hold between the objects

Binary Relations

- Any relation can be converted into an object:
  - $R(x_1, x_2, \ldots, x_k) \rightarrow$
    new "relation" object y,
    $R_1(x_1, y), R_2(x_2, y), \ldots, R_k(x_k, y)$
  - E.g., registrations are "relation objects"

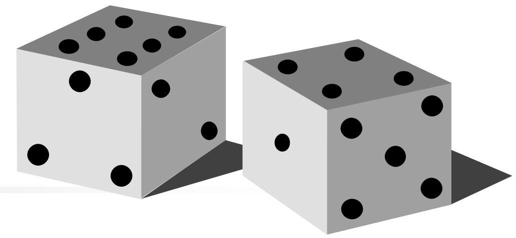
$\Rightarrow$ Can restrict attention to binary relations R(x,y)

Relations & Links

- Binary relations can also be viewed as links:
- Specify the set of objects related to x via R
- $R(x,y) \rightarrow y \in x.R^1, x \in y.R^2$
- E.g., Teaches(p,c) $\rightarrow$
  - p.Courses = {courses c : Teaches(p,c)}
  - c.Instructor = {professors p : Teaches(p,c)}

Probabilistic Relational Models: Relational Bayesian Networks

Probabilistic Models



- Uncertainty model:
  - space of "possible worlds";
  - probability distribution over this space.
- In attribute-based models, world specifies
  - assignment of values to fixed set of random variables
- In relational models, world specifies
  - Set of domain elements
  - Their properties
  - Relations between them

PRM Scope

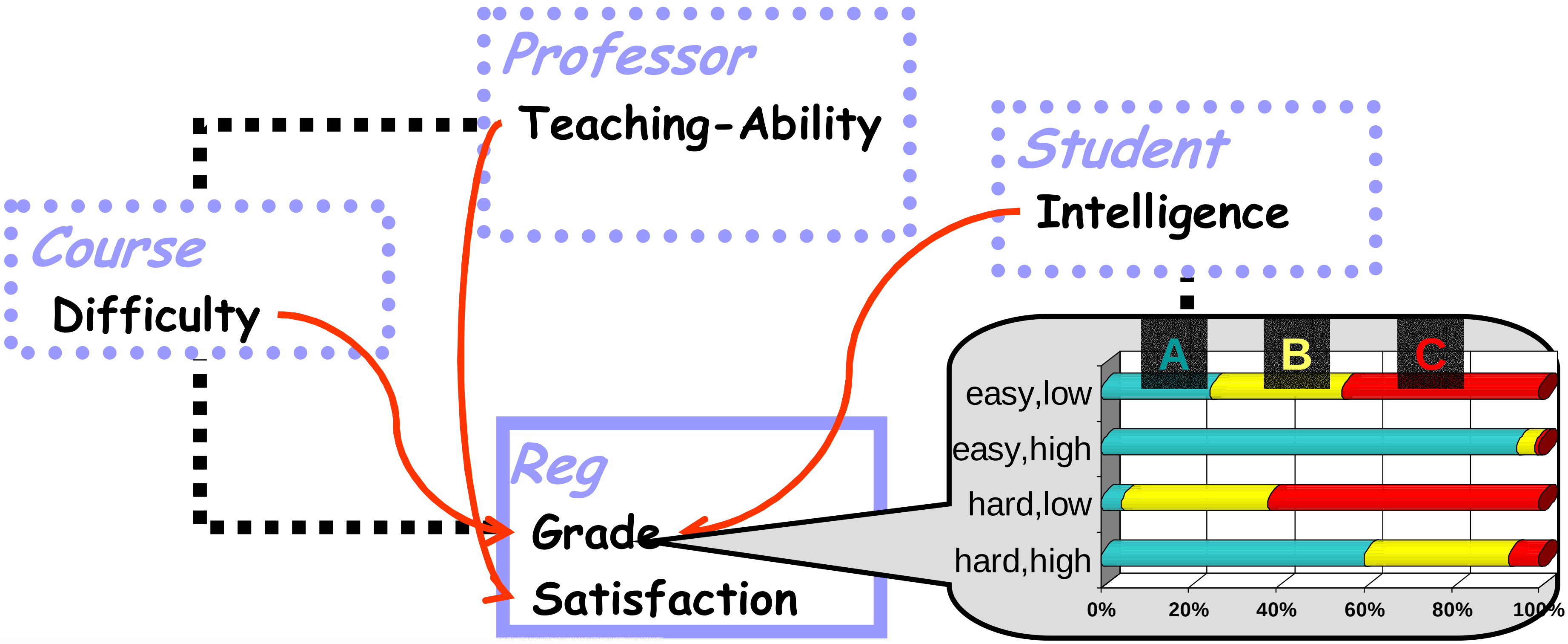
- Entire set of relational worlds is infinite and too broad
- Assume circumscribed class of sets of worlds Ω_ξ consistent with some type of background knowledge ξ
- PRM Π is a template defining $P_\Pi(\Omega_\xi)$ for any such ξ

- Simplest class — attribute-based PRMs:
  - ξ = *relational skeleton:*
    - finite set of entities E and relations between them
  - Ω_ξ = all assignments of values to all attributes of entities in E
  - PRM template defines distribution over Ω_ξ for any such ξ

[K. & Pfeffer '98; Friedman, Getoor, K. Pfeffer '99]

Relational Bayesian Network

- Universals: Probabilistic patterns hold for all objects in class
- Locality: Represent direct probabilistic dependencies
  - Links define potential interactions



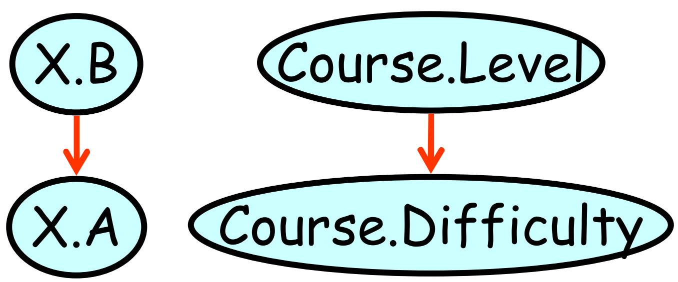
[K. & Pfeffer; Poole; Ngo & Haddawy]

RBN: Semantics

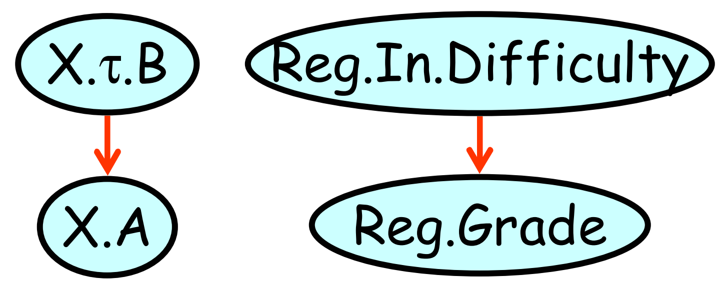
- ξ: set of objects & relations between them
- Ω_ξ: the set of all assignments of values to all attributes of all objects in ξ
- $P_\Pi(\Omega_\xi)$ is defined by a ground Bayesian network:
  - variables: attributes of all objects
  - dependencies: determined by
    - relational links in ξ
    - dependency structure of RBN model Π

RBN Structure

- For each class X and attribute A, structure specifies parents for X.A

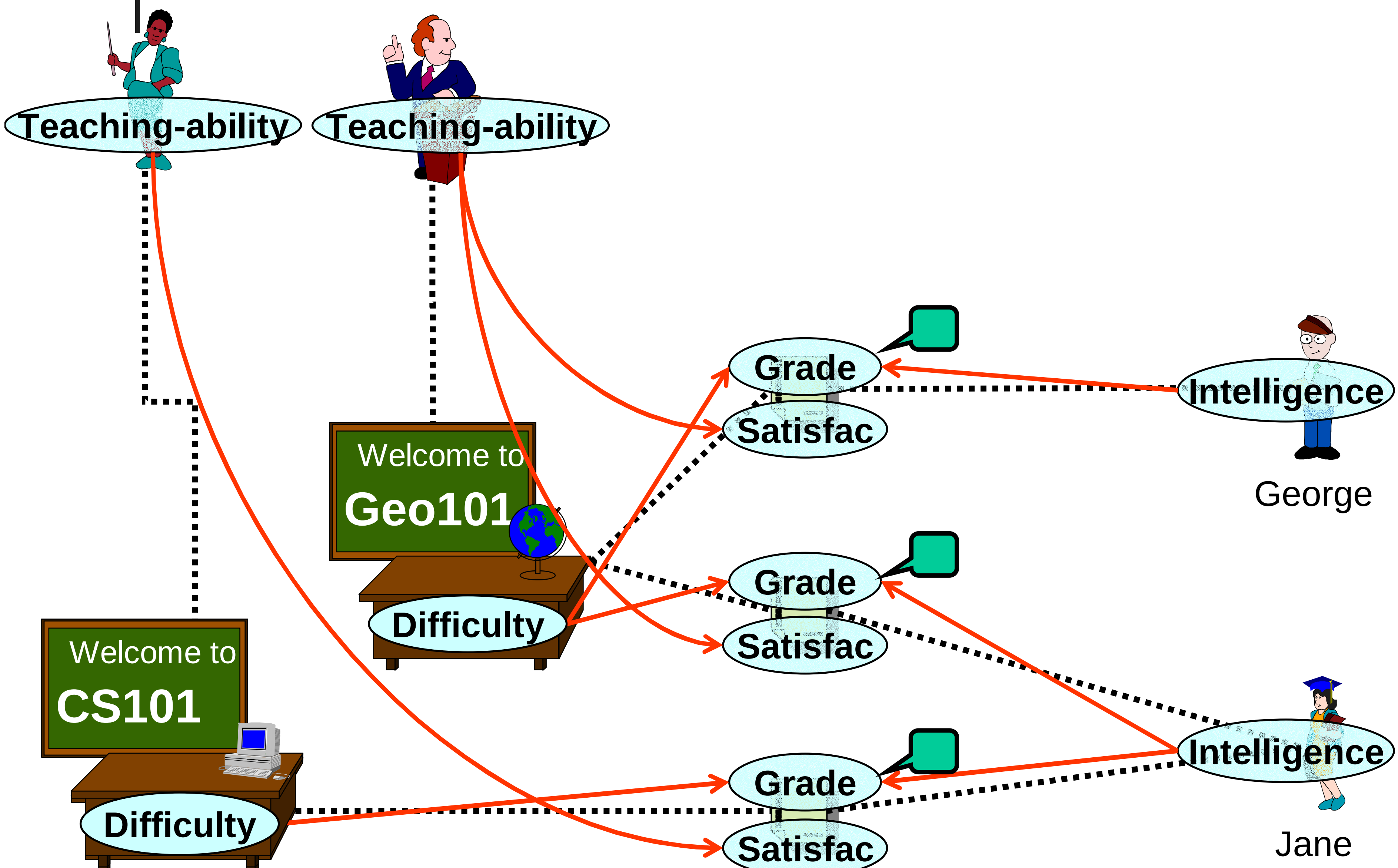


- For any object x in class X, x.B is parent of x.A

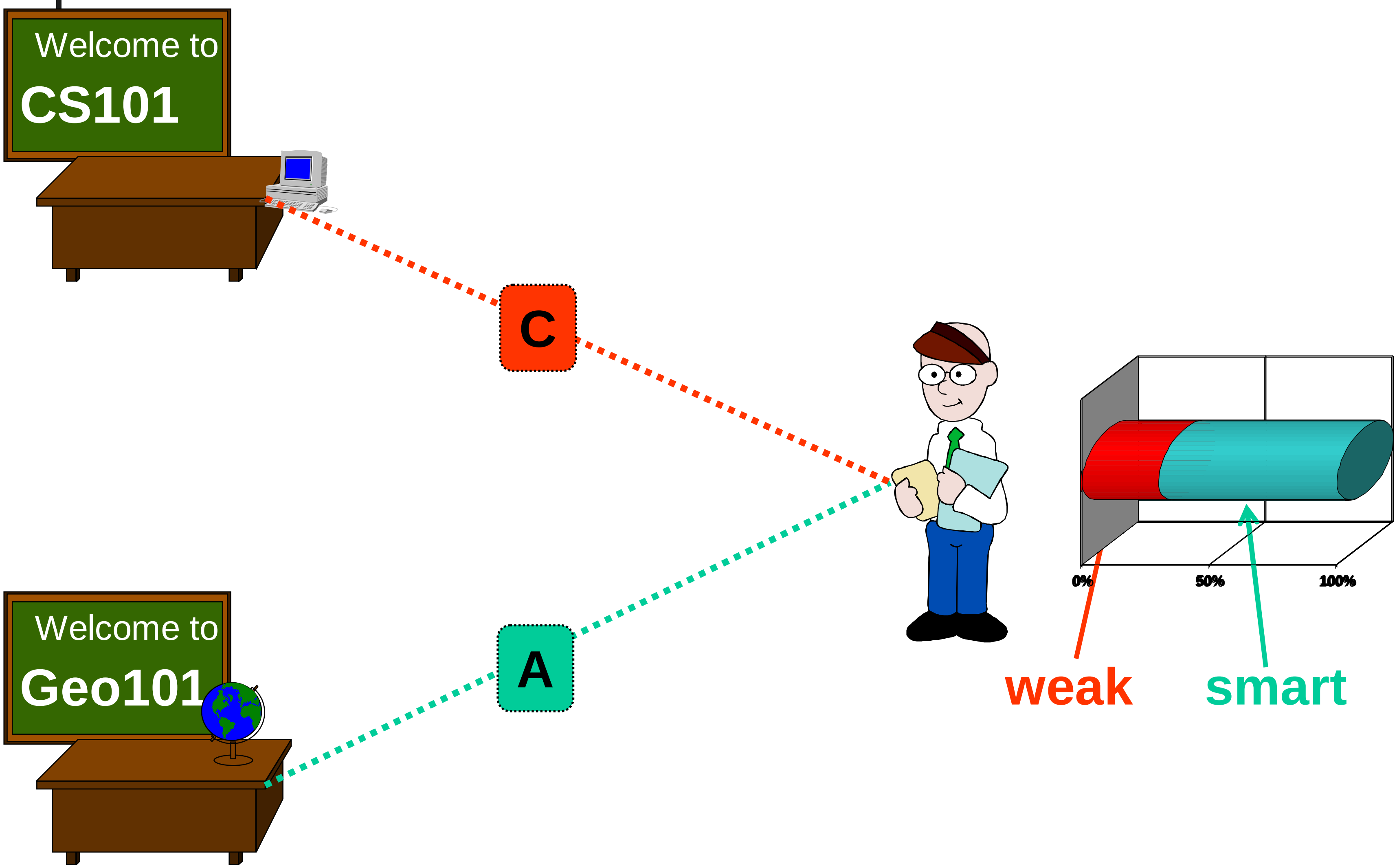


- τ: link or chain of links
- For any object x in class X, x.τ.B is parent of x

RBN Semantics



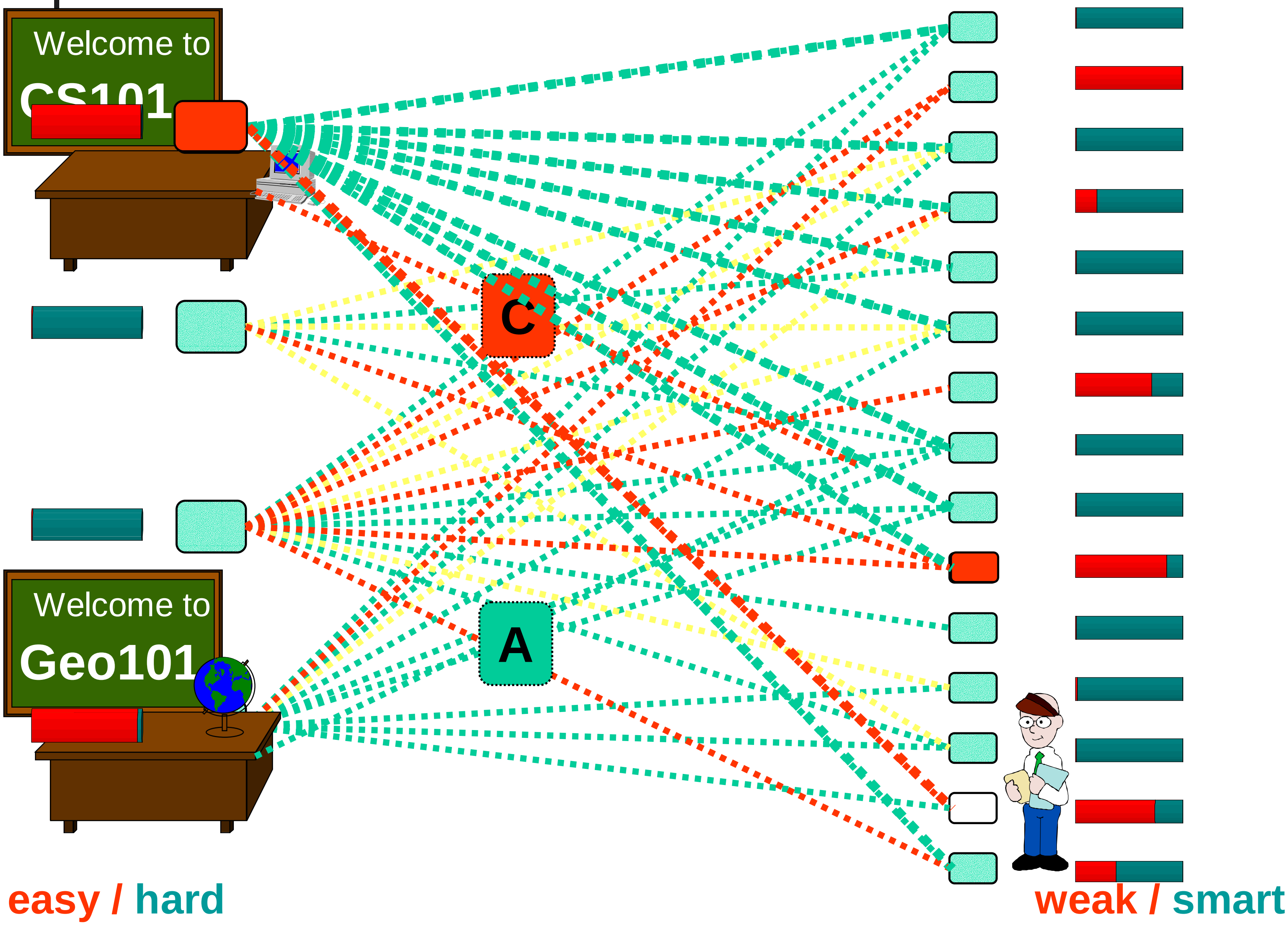
The Web of Influence



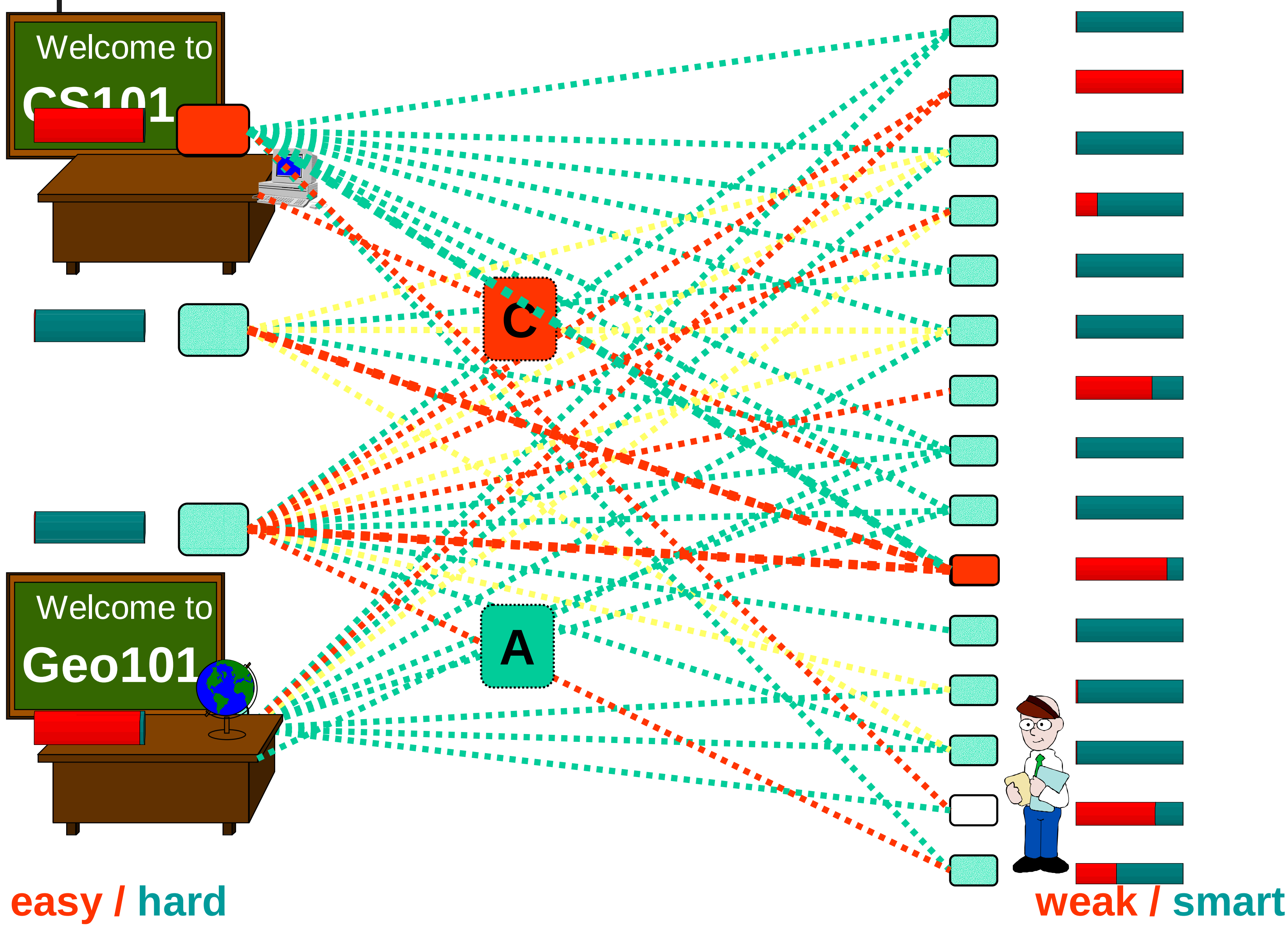
easy / hard

weak / smart

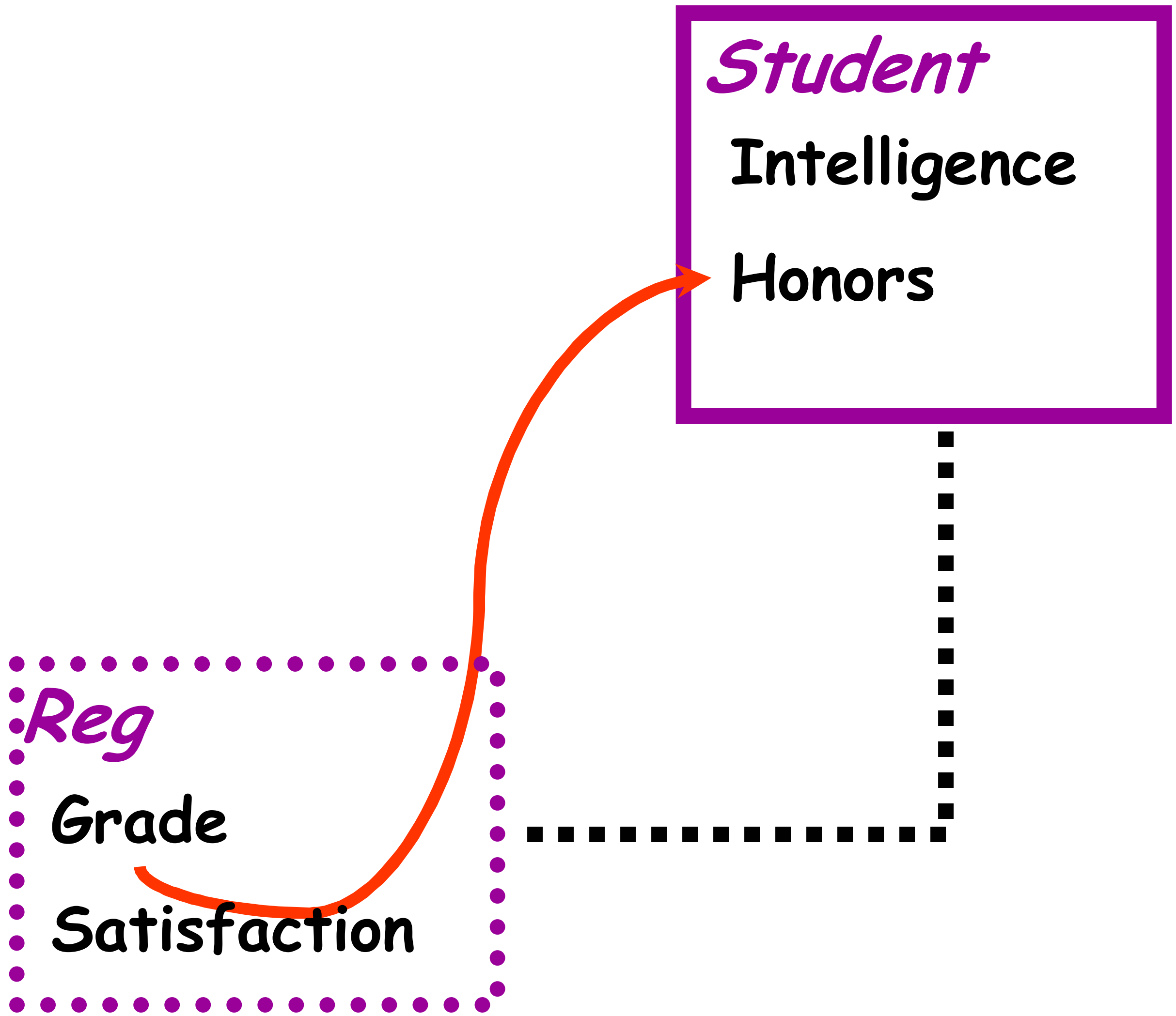
The Web of Influence



The Web of Influence

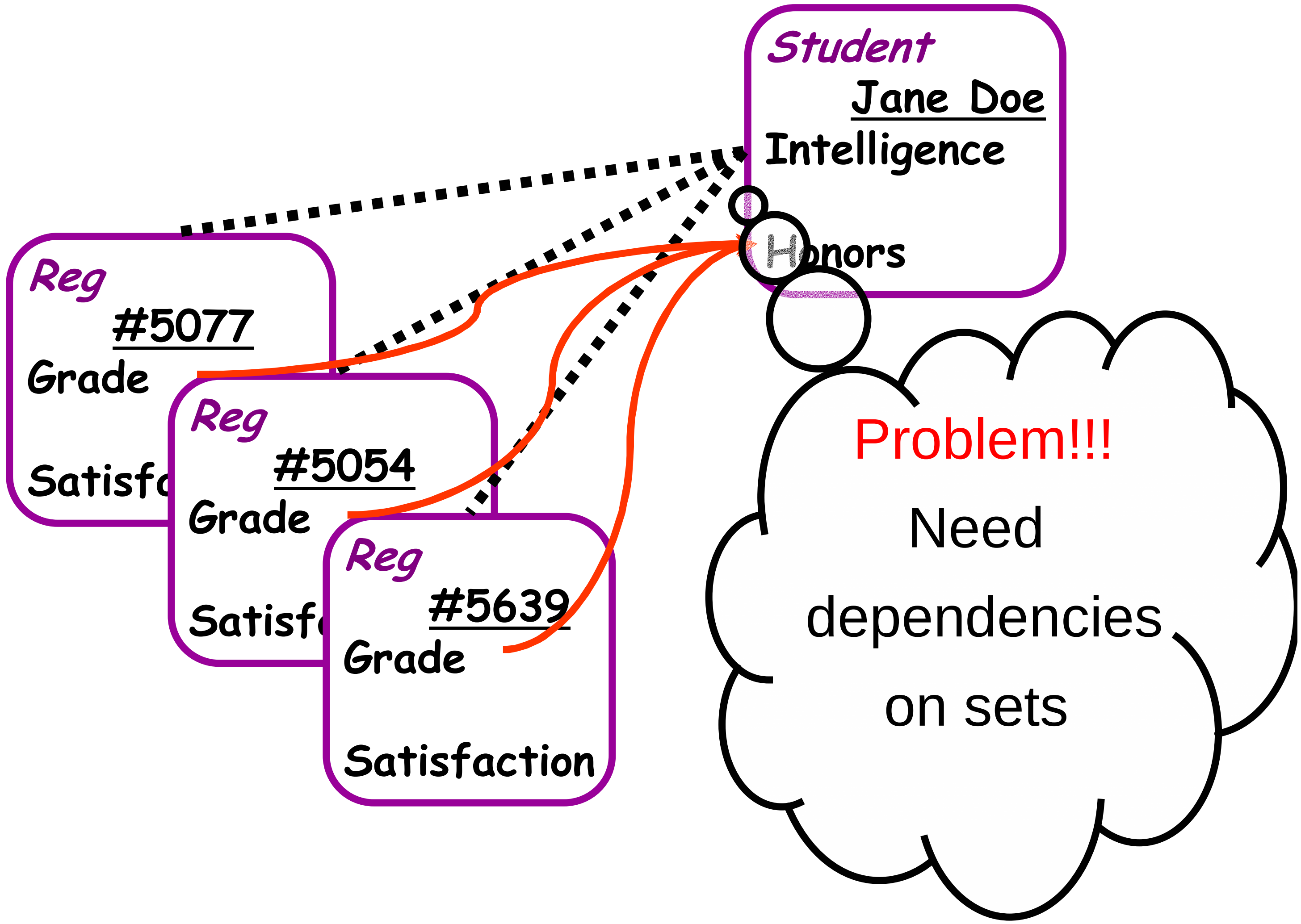


Aggregate Dependencies

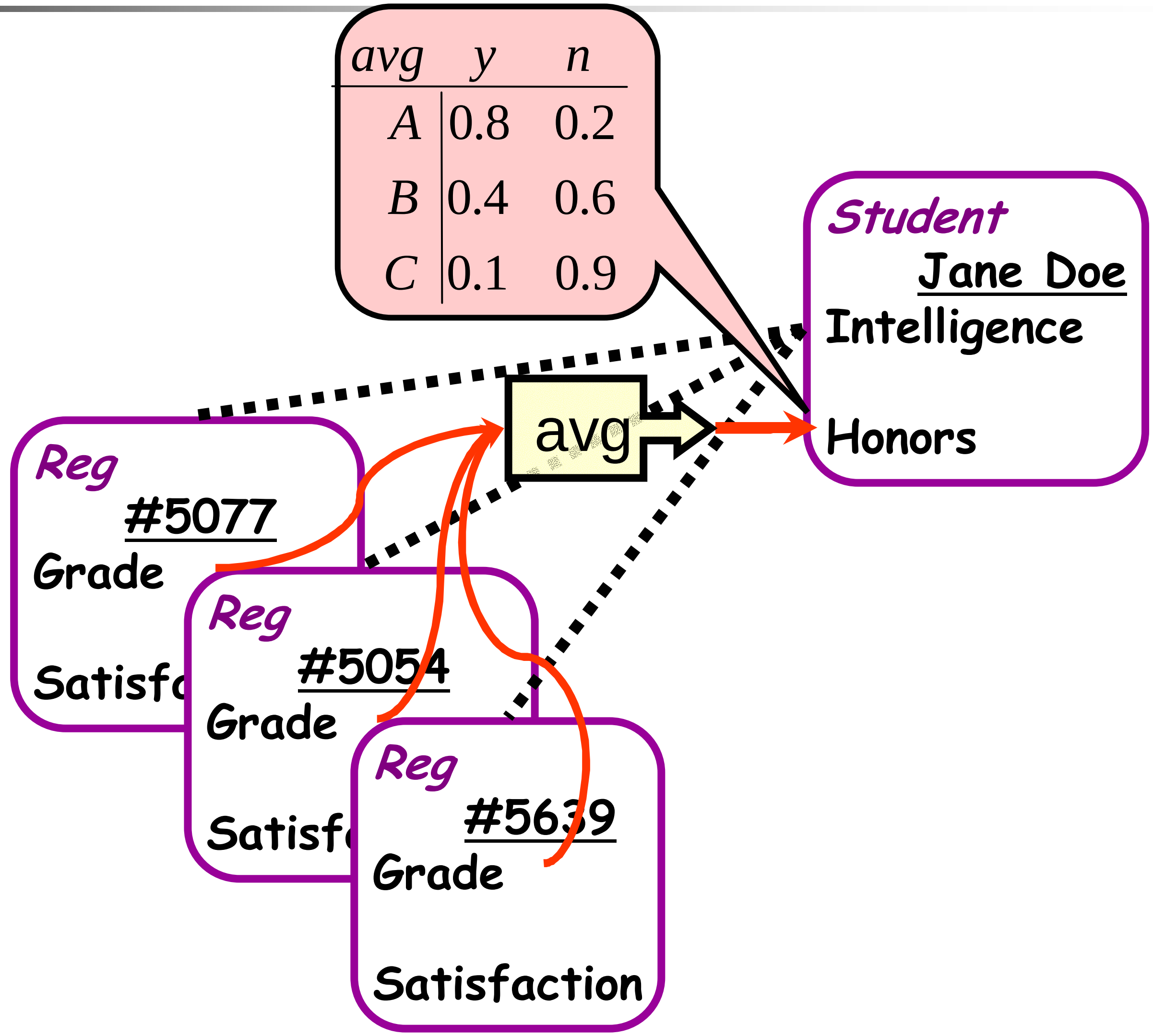


[K. & Pfeffer '98; Friedman, Getoor, K. Pfeffer '99]

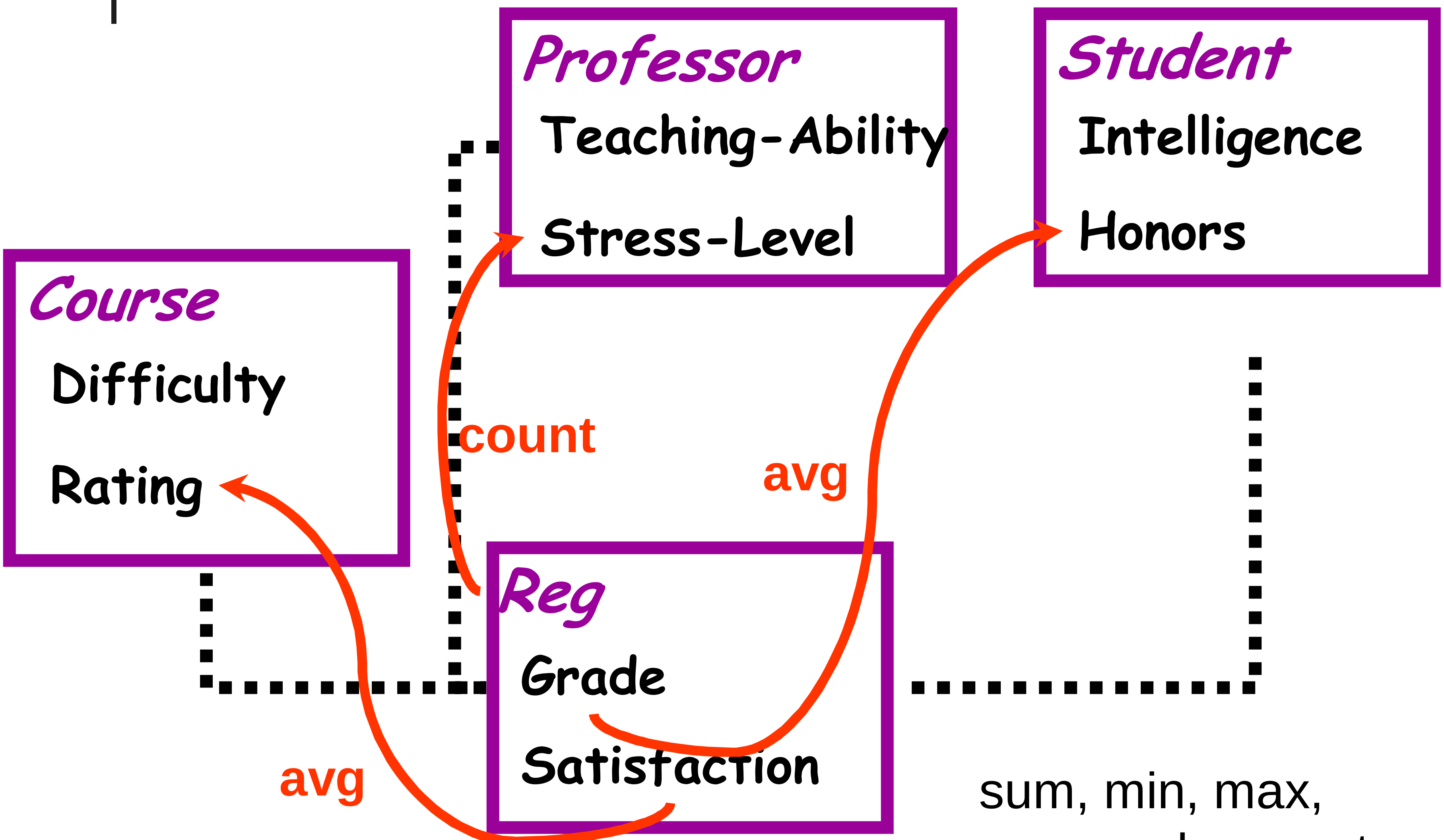
Aggregate Dependencies



Aggregate Dependencies



Aggregate Dependencies



sum, min, max,
avg, mode, count

Basic RBN: Summary

- RBN specifies
  - A probabilistic dependency structure *S*:
    - A set of parents X.τ.B for each class attribute *X.A*
  - A set of *local probability models:*
    - Aggregator to use for each multi-valued dependency
    - Set of CPD parameters $\Theta_{X.A}$
- Given relational skeleton structure ξ, RBN induces a probability distribution over worlds ω
  - Distribution defined via ground BN over attributes x.A

$$P(\omega \mid \xi, S, \Theta) = \prod_{\underbrace{X.A}_{\text{Attributes}}} \prod_{\underbrace{x \in X_\xi}_{\text{Objects}}} P(x.A \mid parents_{S,\xi}(x.A), \Theta_{X.A})$$

Extension: Class Hierarchy

- Subclasses inherit all attributes of parents, but may have additional ones
- For inherited attribute X.A, subclass can:
  - inherit parent's probabilistic model
  - overwrite with local probabilistic model
- Example:
  - Professor has subclasses assistant, associate, full
  - Inherit distribution over *Stress-Level*
  - Modify distribution over *Salary*

[K. & Pfeffer '98]

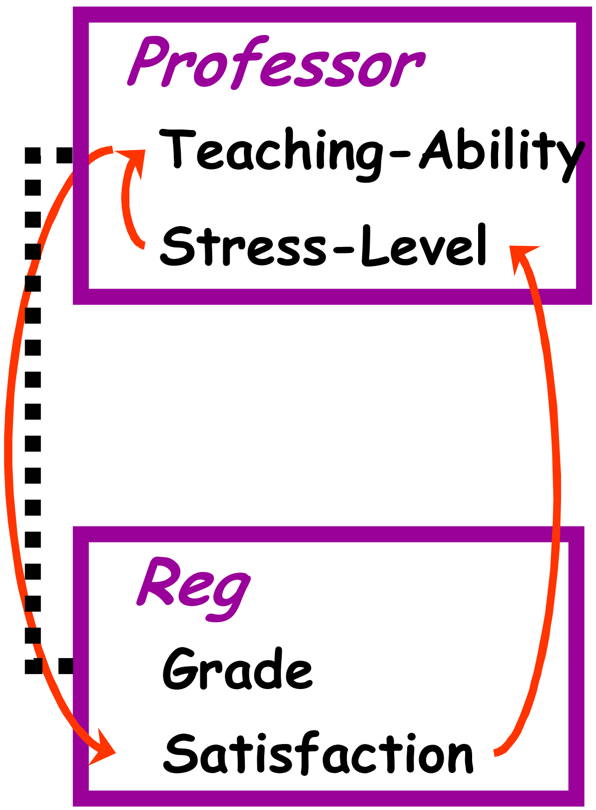
Extension: Class Hierarchies

- Hierarchies allow reuse in knowledge engineering and in learning
  - Parameters and dependency models shared across more objects

- If class assignments specified in ξ, class hierarchy does not introduce complications

[K. & Pfeffer '98]

Coherence & Acyclicity



Smith.Stress-level depends probabilistically on itself
[Friedman, Getoor, K. Pfeffer '99]

- For given skeleton ξ, PRM Π asserts dependencies between attributes of objects:

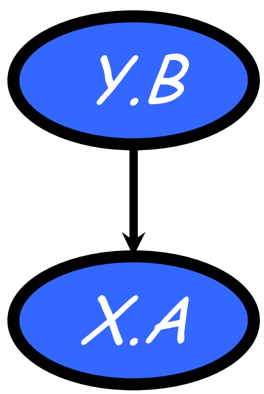
$$y.B \rightarrow_{\xi,\Pi} x.A$$

- Π defines coherent probability model over σ if $\rightarrow_{\xi,\Pi}$ is acyclic

Guaranteeing Acyclicity

How do we guarantee that a PRM Π is acyclic for ***every*** object skeleton ξ?

PRM dependency structure S ⟹ template-level dependency graph



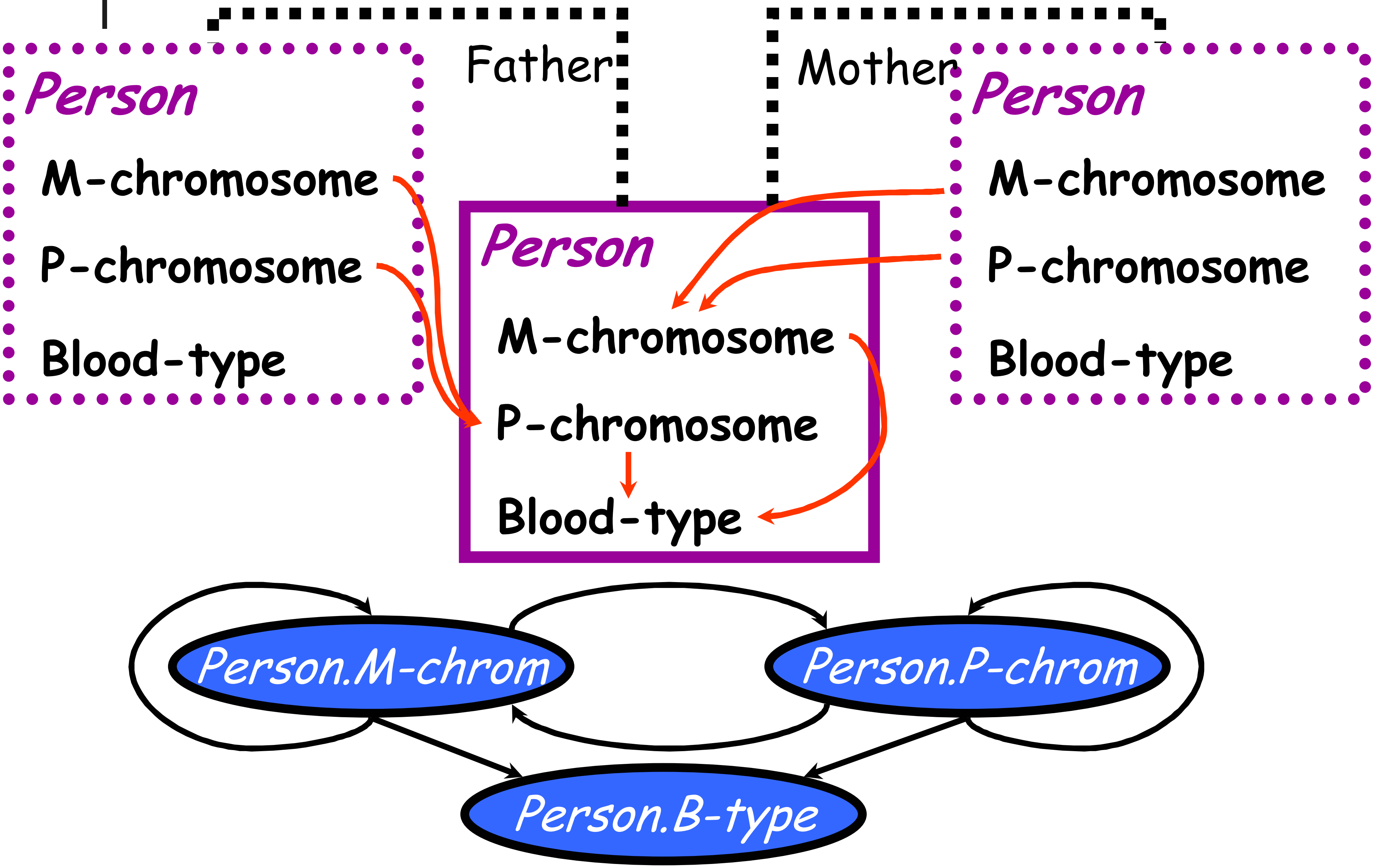
if $X.\tau.B \in$ Parents(X.A), and class of $X.\tau$ is Y

Attribute stratification:

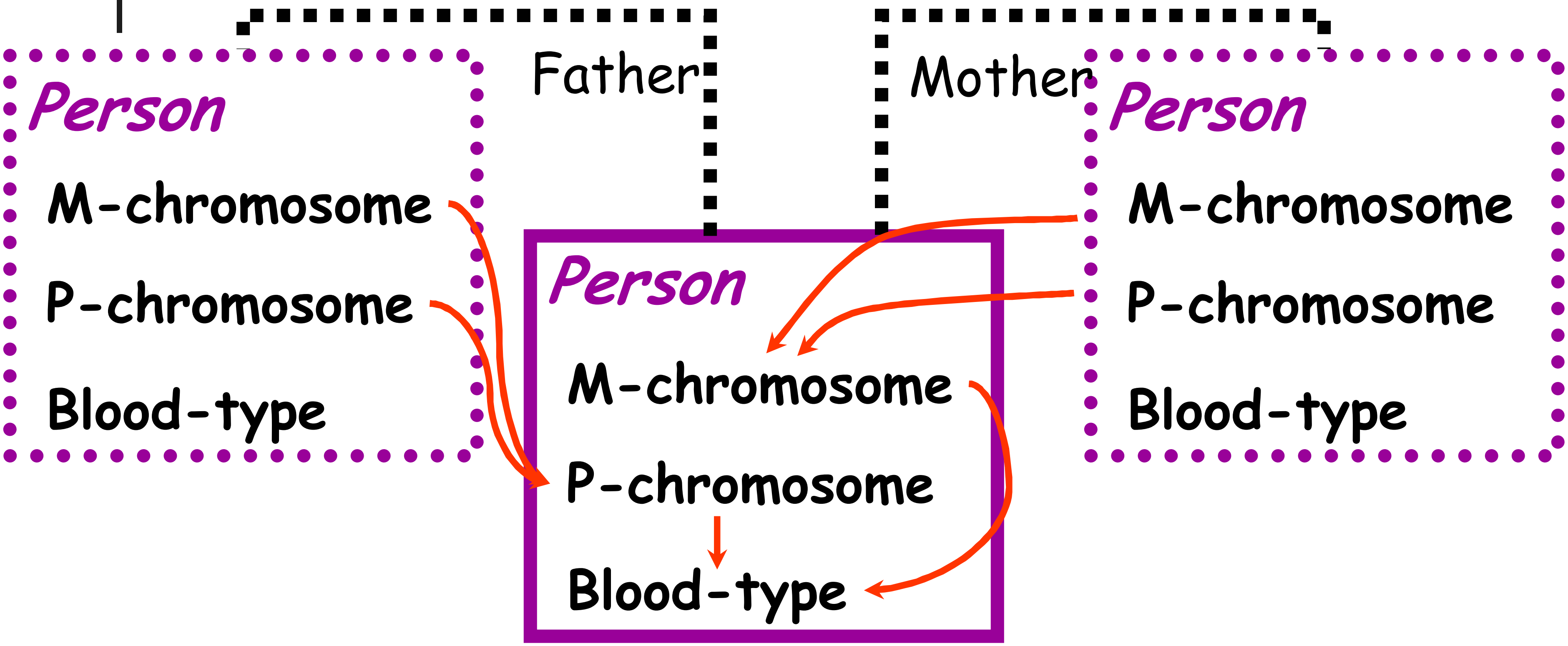
class dependency graph acyclic $\Rightarrow$

$\rightarrow_{\xi,\Pi}$ acyclic for all ξ

Limitation of Stratification



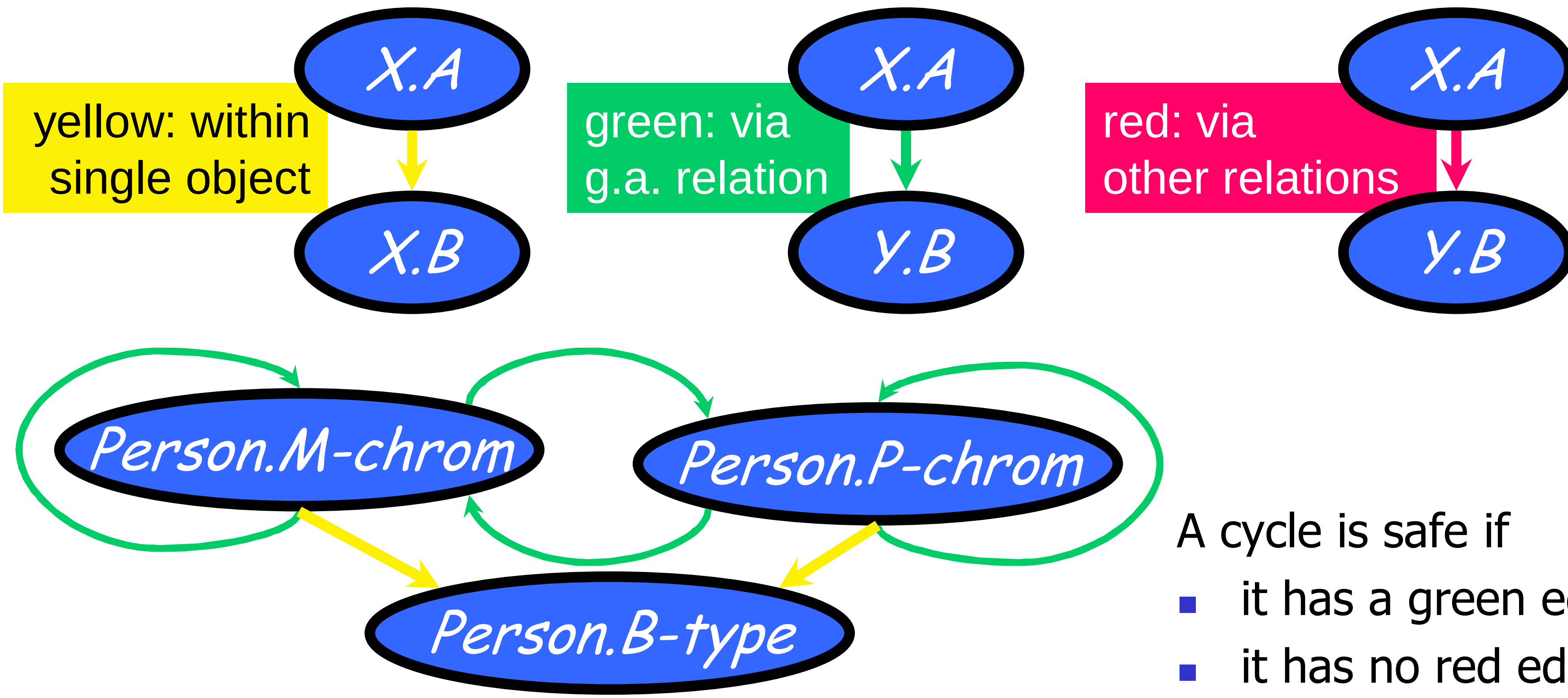
Limitation of Stratification



- Prior knowledge: the Father-of relation is acyclic
  - Dependence of *Person.A* on *Person.Father.B* cannot induce cycles

Guaranteeing Acyclicity

- With guaranteed acyclic relations, some cycles in the dependency graph are guaranteed to be safe.
- We color the edges in the dependency graph



A cycle is safe if

- it has a green edge
- it has no red edge

Object-Oriented Bayesian Nets

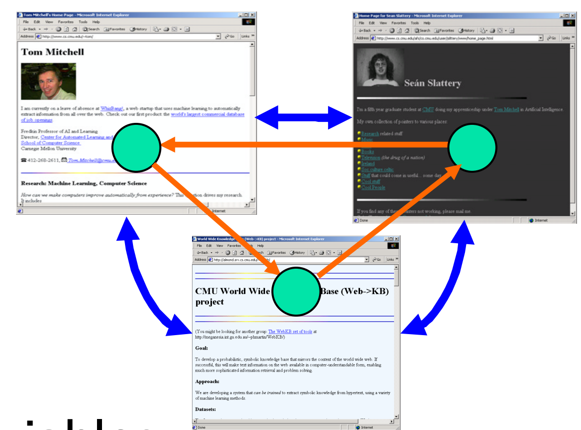
- OOBNs are RBN with only one type of relation
  - One object can be a "part-of" another
  - Objects can only interact with component parts
  - Other types of relationships must be embedded into the "part-of" framework
- Defines "neat" hierarchy of related objects
- Provides clearly defined object interface between object x and its enclosing object y

[K. & Pfeffer '97]

Probabilistic Relational Models: Relational Markov Networks

Why Undirected Models?

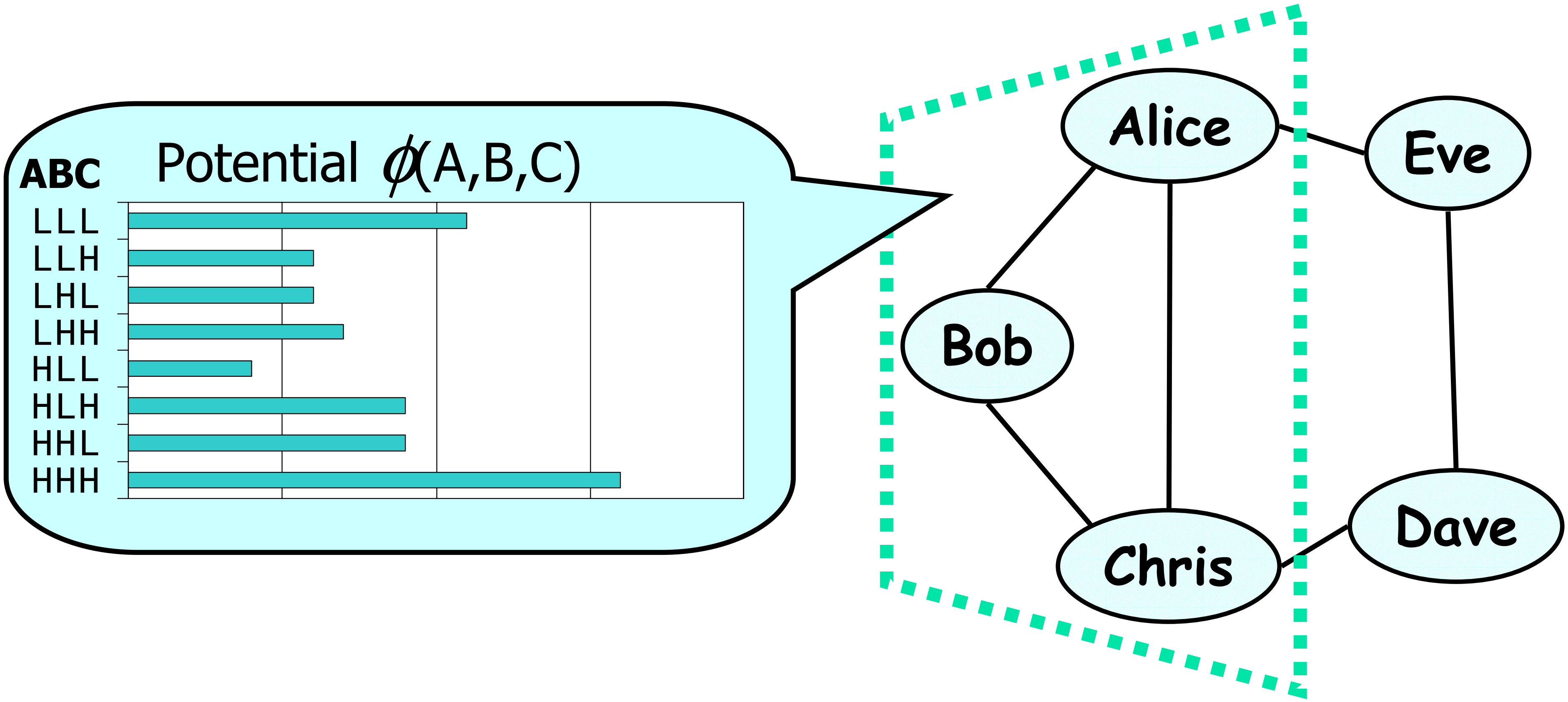
- Symmetric, non-causal interactions
  - E.g., web: categories of linked pages are correlated
  - Cannot introduce direct edges because of cycles
- Patterns involving multiple entities
  - E.g., web: “triangle” patterns
  - Directed edges not appropriate



- “Solution”: Impose arbitrary direction
  - Not clear how to parameterize CPD for variables involved in multiple interactions
  - Impossible to do within a class-based parameterization

[Taskar, Abbeel, K. 2001]

Markov Networks: Review



$$P(A,B,C,D,E) = \frac{1}{Z}\phi(A,B,C)\phi(C,D)\phi(D,E)\phi(E,A)$$

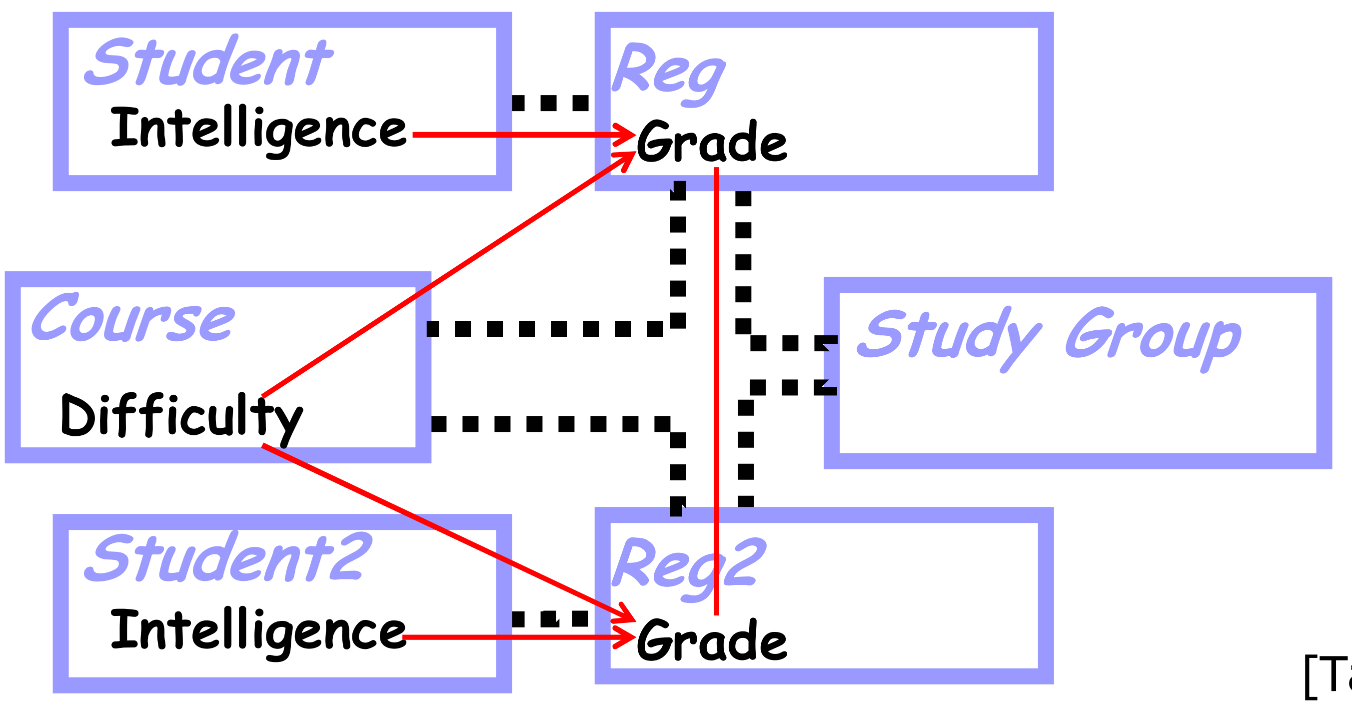
Markov Networks: Review

- A Markov network is an undirected graph over some set of variables **V**
- Graph associated with a set of *potentials* ϕ_i
  - Each potential is factor over subset $\mathbf{V}_i$
  - Variables in $\mathbf{V}_i$ must be a (sub)clique in network

$$P(\mathbf{V}) = \frac{1}{Z}\prod_i \phi_i(\mathbf{V}_i)$$

Relational Markov Networks

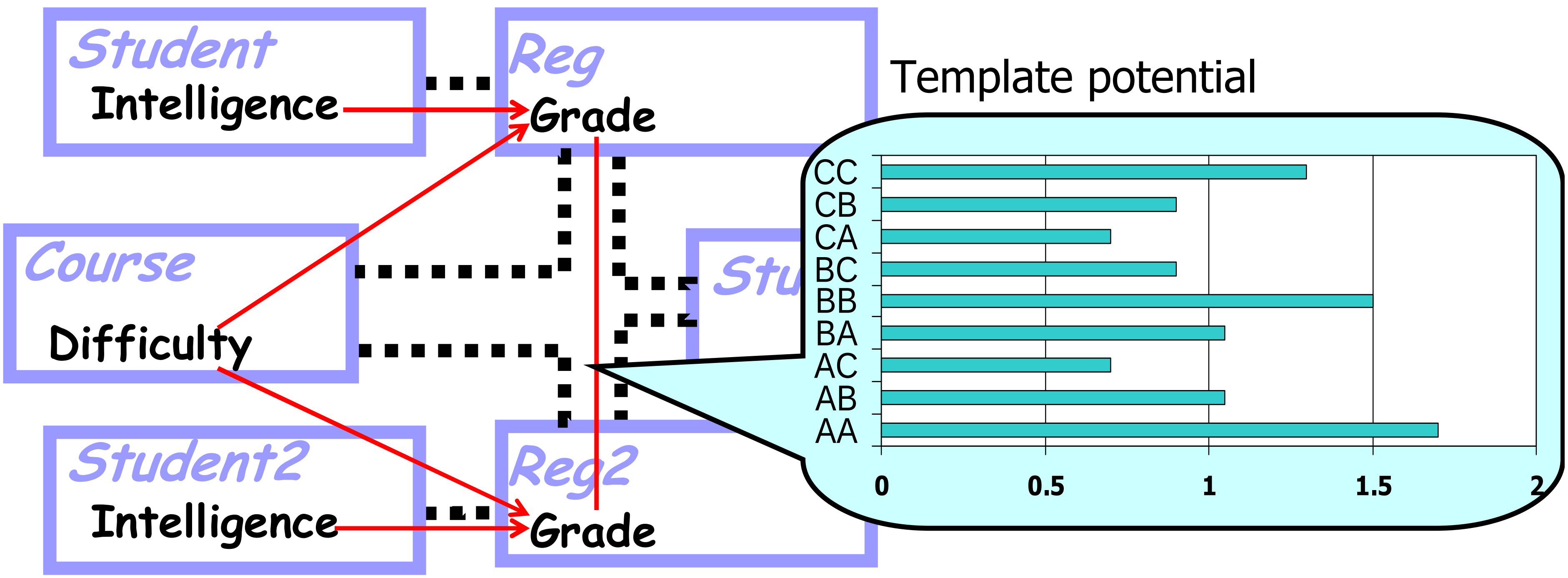
- Probabilistic patterns hold for groups of objects
- Groups defined as sets of (typed) elements linked in particular ways



[Taskar, Abbeel, K. 2002]

Relational Markov Networks

- Probabilistic patterns hold for groups of objects
- Groups defined as sets of (typed) elements linked in particular ways

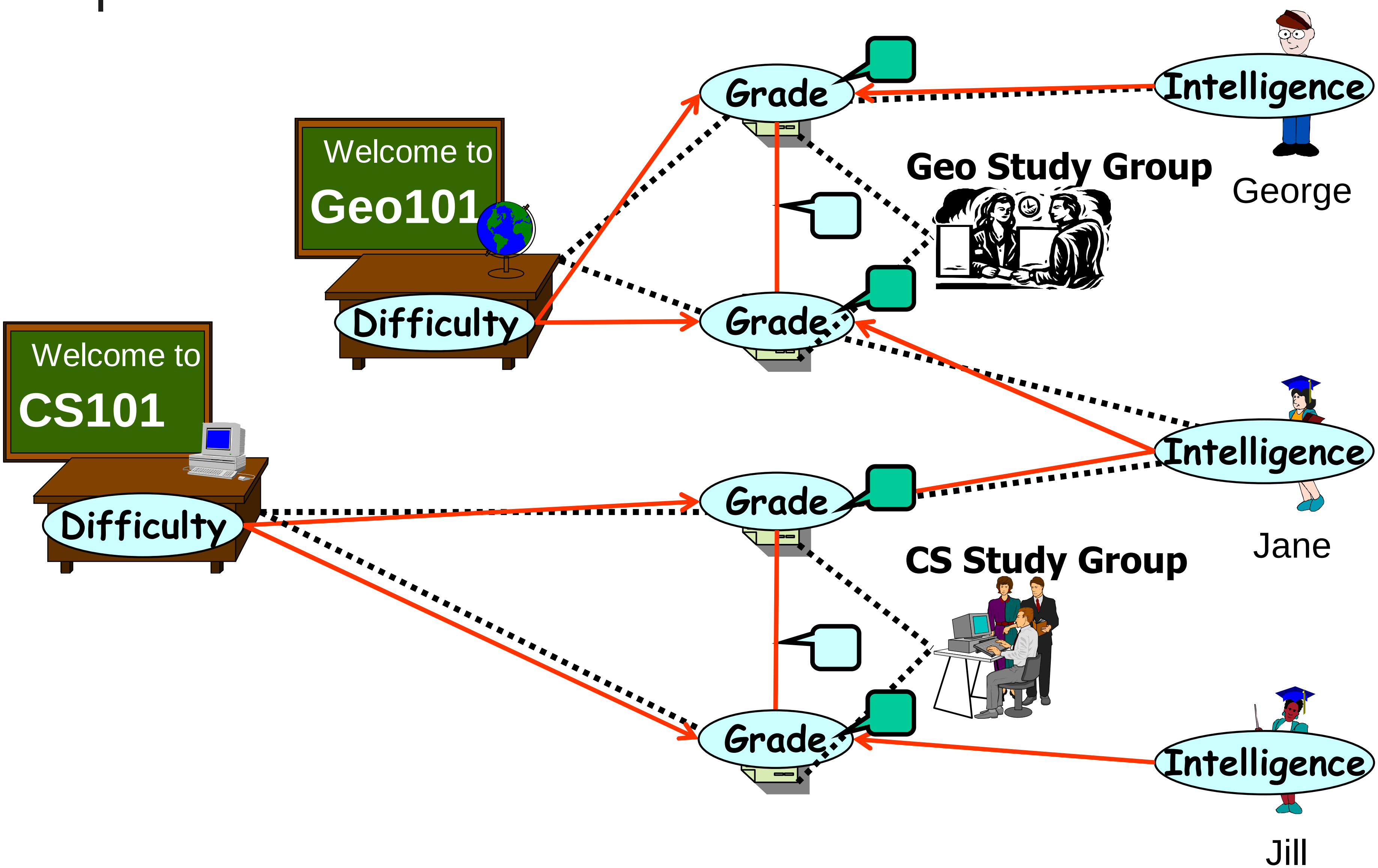


RMN Language

- Define *clique templates*
  - All tuples {reg R_1, reg R_2, group G}
    s.t. In(G, R_1), In(G, R_2)
  - Compatibility potential $\phi(R_1.\text{Grade}, R_2.\text{Grade})$

- Ground Markov network contains potential $\phi(r_1.\text{Grade}, r_2.\text{Grade})$ for all appropriate $r_{1,}$ r_2

Ground MN (or Chain Graph)



PRM Inference

Inference: Simple Method

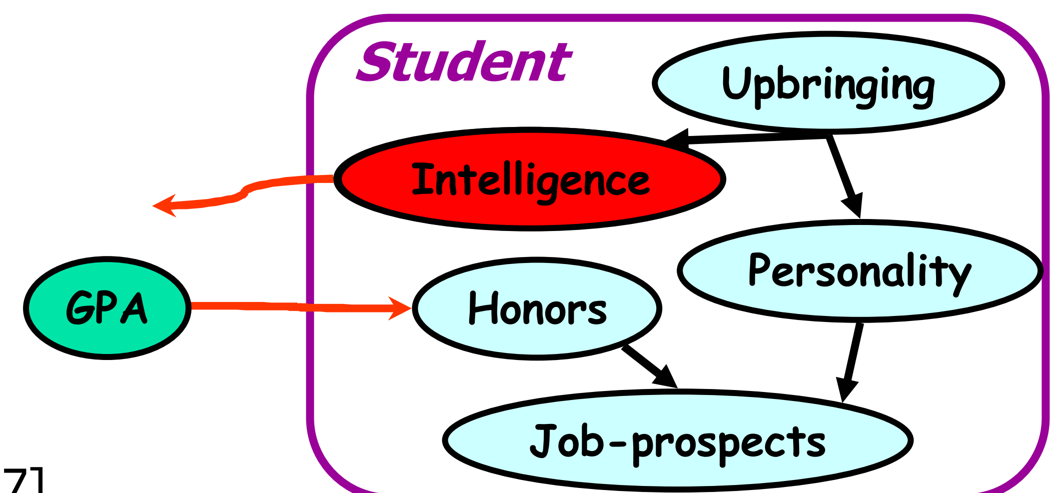
- Define ground network as in semantics
- Apply standard inference methods

- Problem:
  - Very large models can be specified very easily
  - Resulting ground network often highly connected
  - Exact inference is typically intractable

- In practice, often must resort to approximate methods such as belief propagation

Exploiting Structure: Encapsulation

- Objects interact only in limited ways
- Can define *object interface*:
  - Outputs: Object attributes influencing other objects
  - Inputs: External attributes influencing object
- Object is independent of everything given interface
- Inference can be encapsulated within objects, with "communication" limited to interfaces



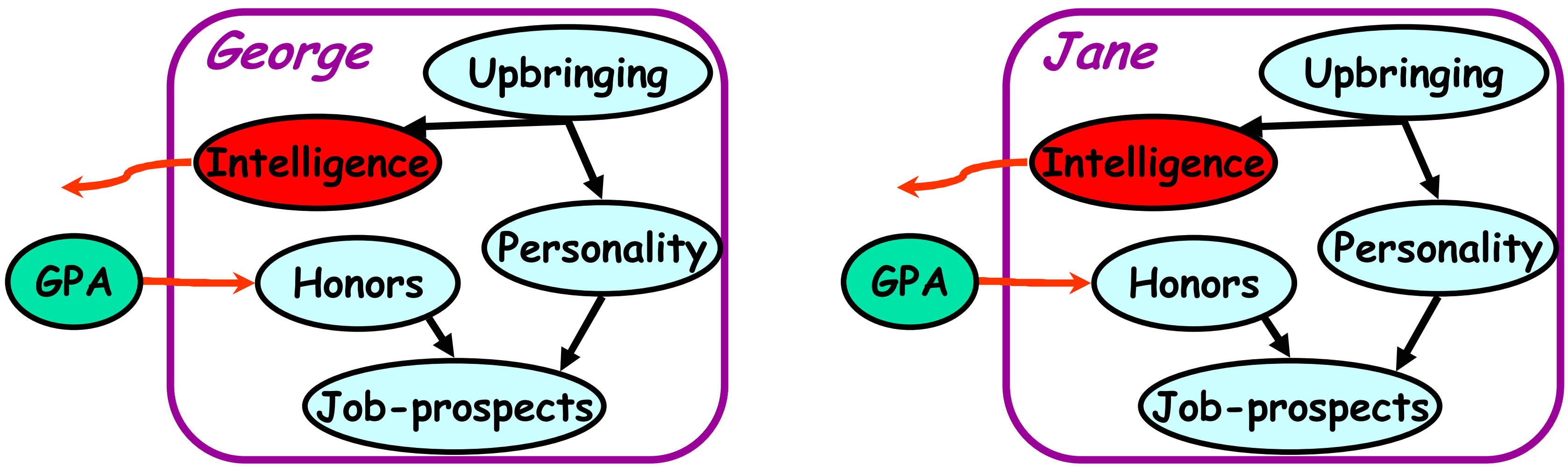
[K. & Pfeffer, 1997]

Exploiting Structure: Encapsulation

- Marginalize object distribution onto interface
- Dependency graph over interfaces induced by
  - Inter-object dependencies
  - And hence by the relational structure
- Perform inference over interfaces
  - If interaction graph has low tree-width, can use exact inference
    - E.g., part-of hierarchy in OOBNs
  - If relational structure is more complex, can use BP
    - A form of Kikuchi BP, where cluster selection is guided by relational structure

Exploiting Structure: Reuse

- Objects from same class have same model
- For *generic* objects – no internal evidence – marginalize interface is the same
- Can reuse inference – a form of "lifting"



[Pfeffer & K. 1998]

Exploiting Structure: Reuse

- Generic objects often play same role in model
  - Multiple students that all take the same class
- Reuse: compute interface once
- Combinatorics: compute total contribution to probability in closed form

P(*k* students like the class | teaching-ability = low) =

P(generic student likes the class | teaching ability = low)

[Pfeffer & K. 1998]

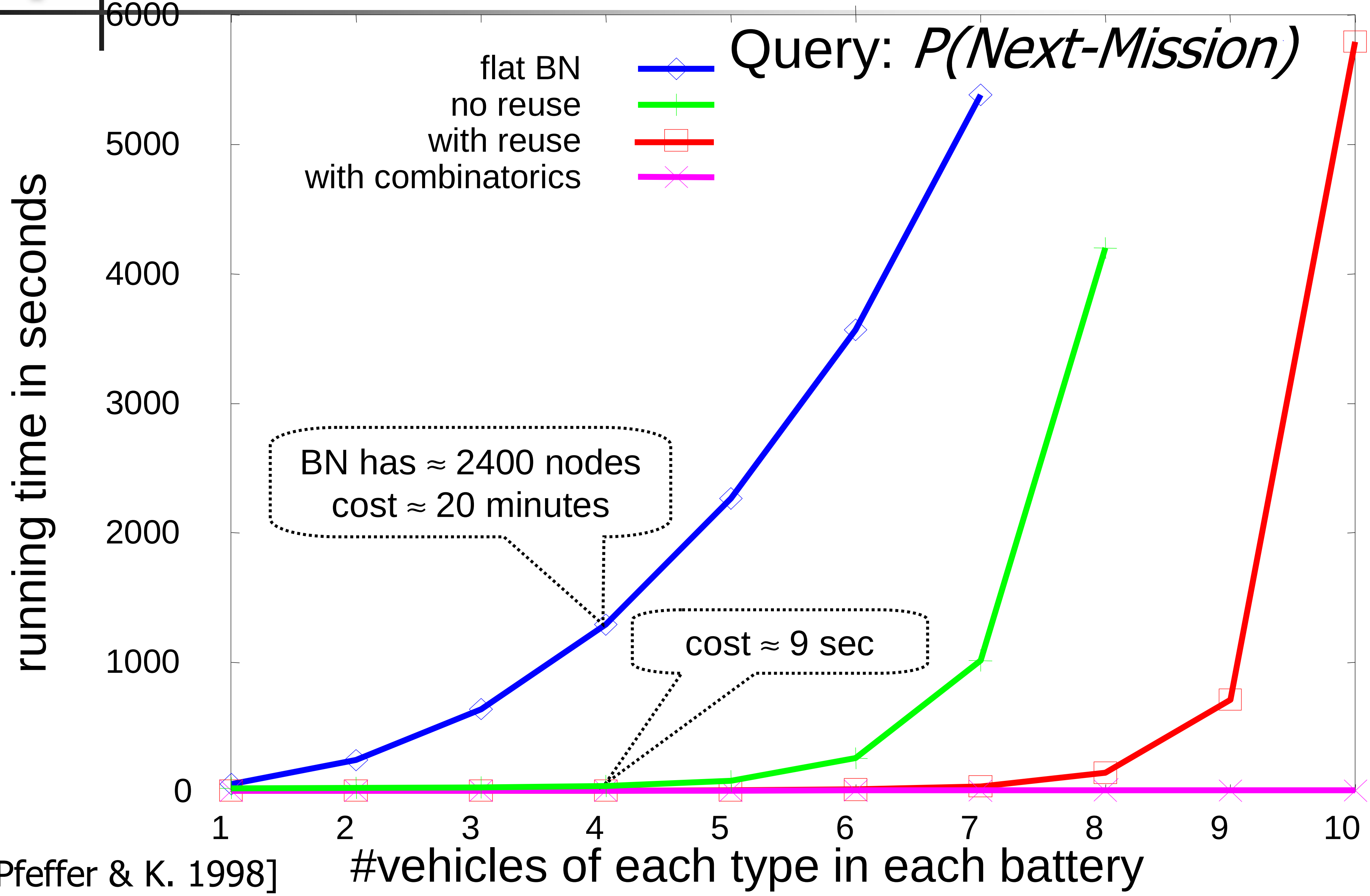
Case study

Battlefield situation assessment for missile units

- several locations
- many units
- each has detailed model

- Example object classes:
  - Battalion
  - Battery
  - Vehicle
  - Location
  - Weather.
- Example relations:
  - At-Location
  - Has-Weather
  - Sub-battery/In-battalion
  - Sub-vehicle/In-battery

Effect of Exploiting Structure



[Pfeffer & K. 1998]

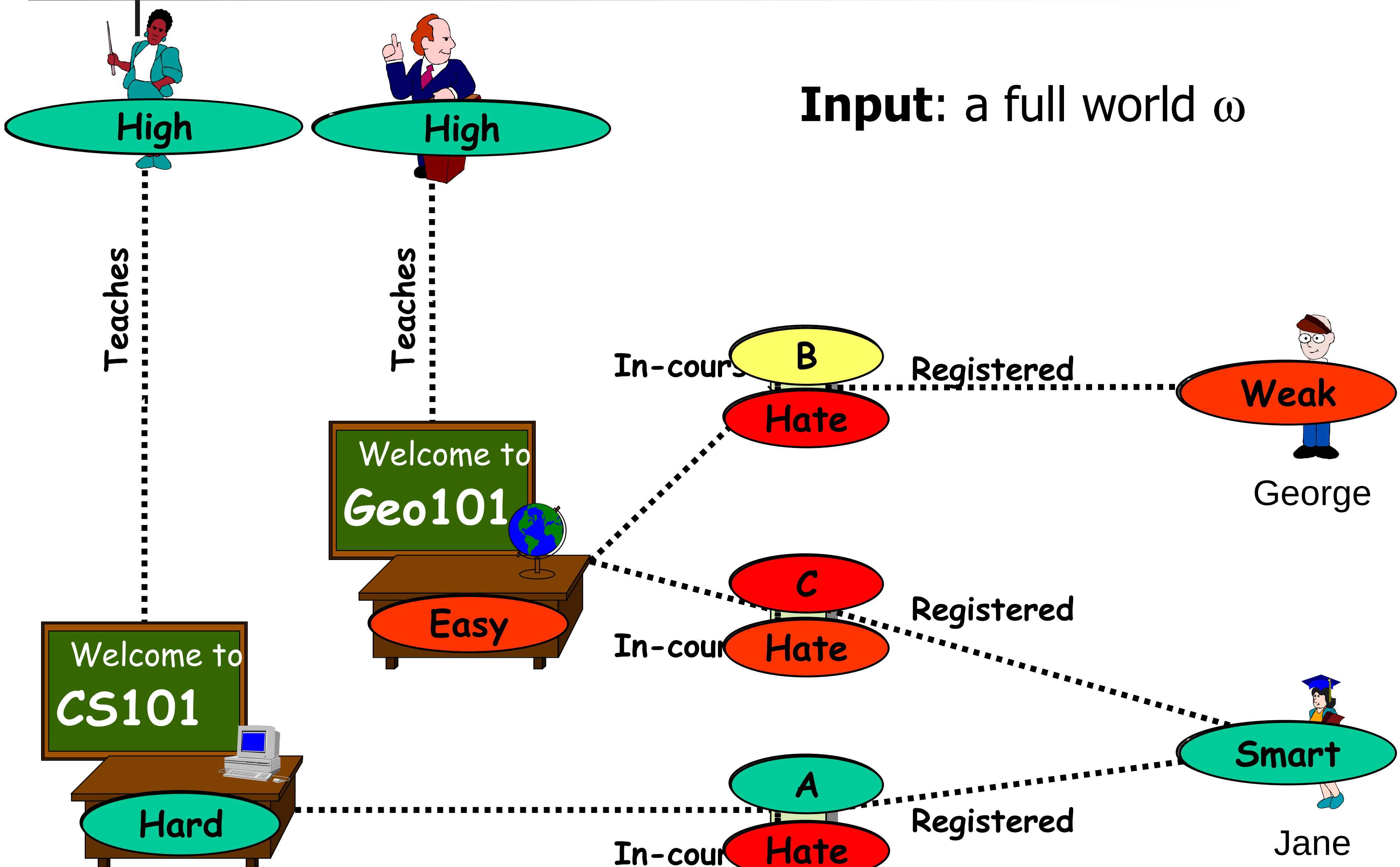
PRM Learning

Outline

- Relational Bayesian networks
  - Likelihood function
  - ML parameter estimation
  - EM
  - Structure learning
- Relational Markov networks
  - Parameter estimation
- Applications:
  - Collective classification – web data
  - Relational clustering – biological data

PRM Learning: Input

Input: a full world ω



Likelihood Function

$$L(S,\Theta : \omega) = P(\omega \mid \xi, S, \Theta)$$

$$= \underbrace{\prod_{X.A}}_{\text{Attributes}} \underbrace{\prod_{x \in X_\xi}}_{\text{Objects}} P(x.A \mid parents_{S,\sigma}(x.A), \Theta_{X.A})$$

- Likelihood of a BN with shared parameters
- Joint likelihood is a product of likelihood terms
  - One for each attribute X.A and its family
- For each X.A, the likelihood function aggregates counts from all occurrences x.A in world ω

[Friedman, Getoor, K., Pfeffer, 1999]

Likelihood Function: Multinomials

Log-likelihood:

$$\log P(\omega \mid \xi, S, \Theta) =$$

$$\sum_{X.A} \sum_{\mathbf{u} \in Val(Pa(X.A))} \sum_{a \in Val(X.A)} M(a, \mathbf{u}) \log \theta_{x|\mathbf{u}}$$

Sufficient statistics:

$$M(a, \mathbf{u}) =$$

$$\left| \{ x \in X_{\omega} : x.A = a, parents_{S,\omega}(x.A) = \mathbf{u} \} \right|$$

RBN Parameter Estimation

- MLE parameters:

$$\hat{P}(\mathrm{Reg}.Grade = A \mid Student.Intell = hi, Course.Diff = lo)$$
$$= \frac{M(\mathrm{Reg}.Grade = A, Student.Intell = hi, Course.Diff = lo)}{M(\mathrm{Reg}.Grade = *, Student.Intell = hi, Course.Diff = lo)}$$

- Bayesian estimation:
  - Prior for each attribute X.A
  - Posterior uses aggregated sufficient statistics

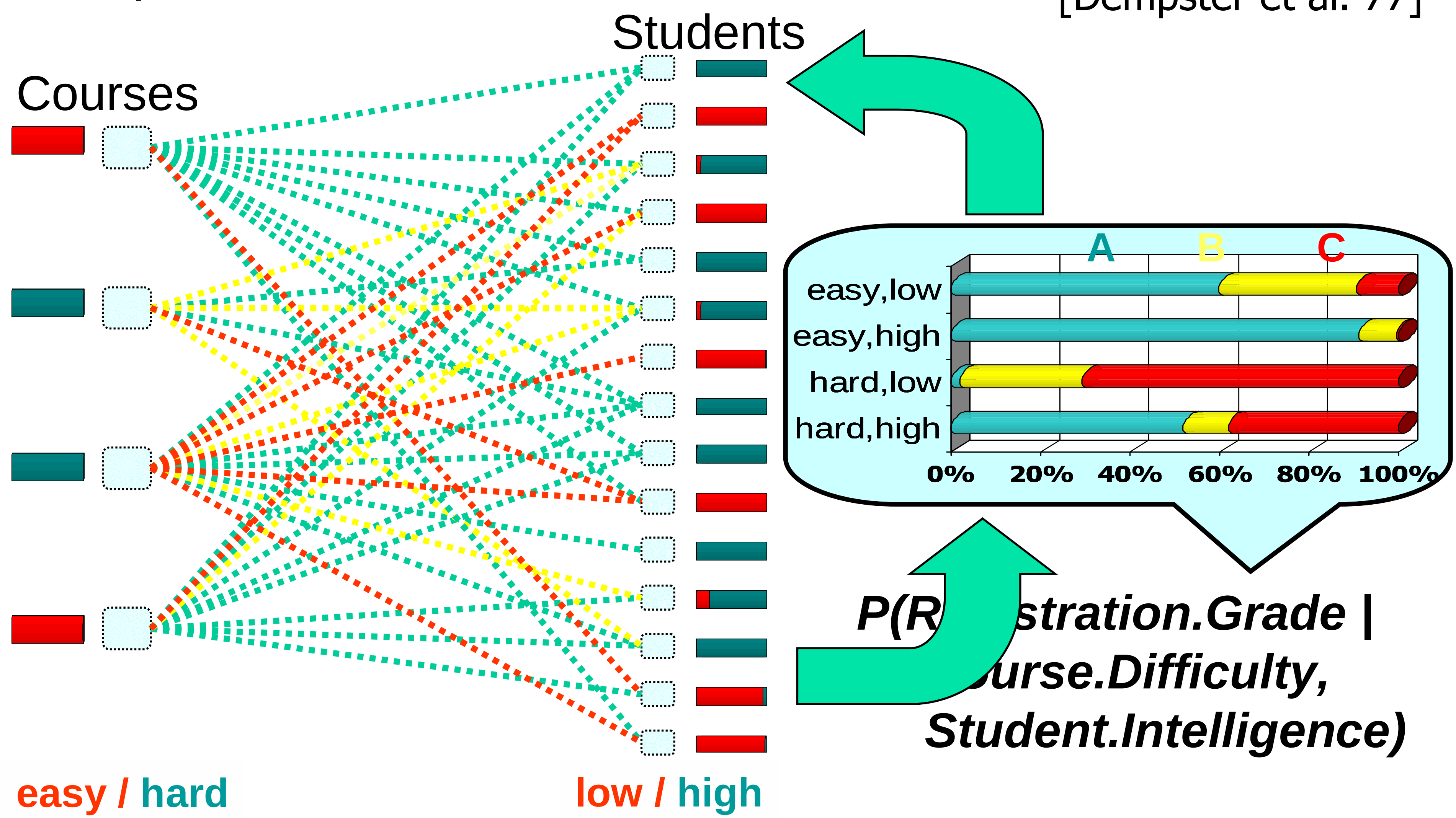
Learning w. Missing Data

- EM Algorithm applies essentially unchanged
  - E-step computes expected sufficient statistics, aggregated over all objects in class
  - M-step uses ML (or MAP) parameter estimation

- Key difference:
  - In general, the hidden variables are **not** independent
  - Computation of expected sufficient statistics requires inference over entire network
  - [Same reasoning as for forward-backward algorithm in temporal models]

Learning w. Missing Data: EM

[Dempster et al. 77]



Learning RBN Structure

- Define set of legal RBN structures
  - Ones with legal class dependency graphs
- Define scoring function — Bayesian score

$$Score(S:\omega) = \log[\underbrace{P(\omega \mid S,\xi)}_{\text{marginal likelihood}}\underbrace{P(S)}_{\text{prior}}]$$

$$P(\omega \mid S,\xi) = \int L(S,\Theta_S:\omega)P(\Theta_S)d\Theta_S$$

  - Product of family scores:
    - One for each X.A
    - Uses aggregated sufficient statistics
- Search for high-scoring legal structure

[Friedman, Getoor, K., Pfeffer, 1999]

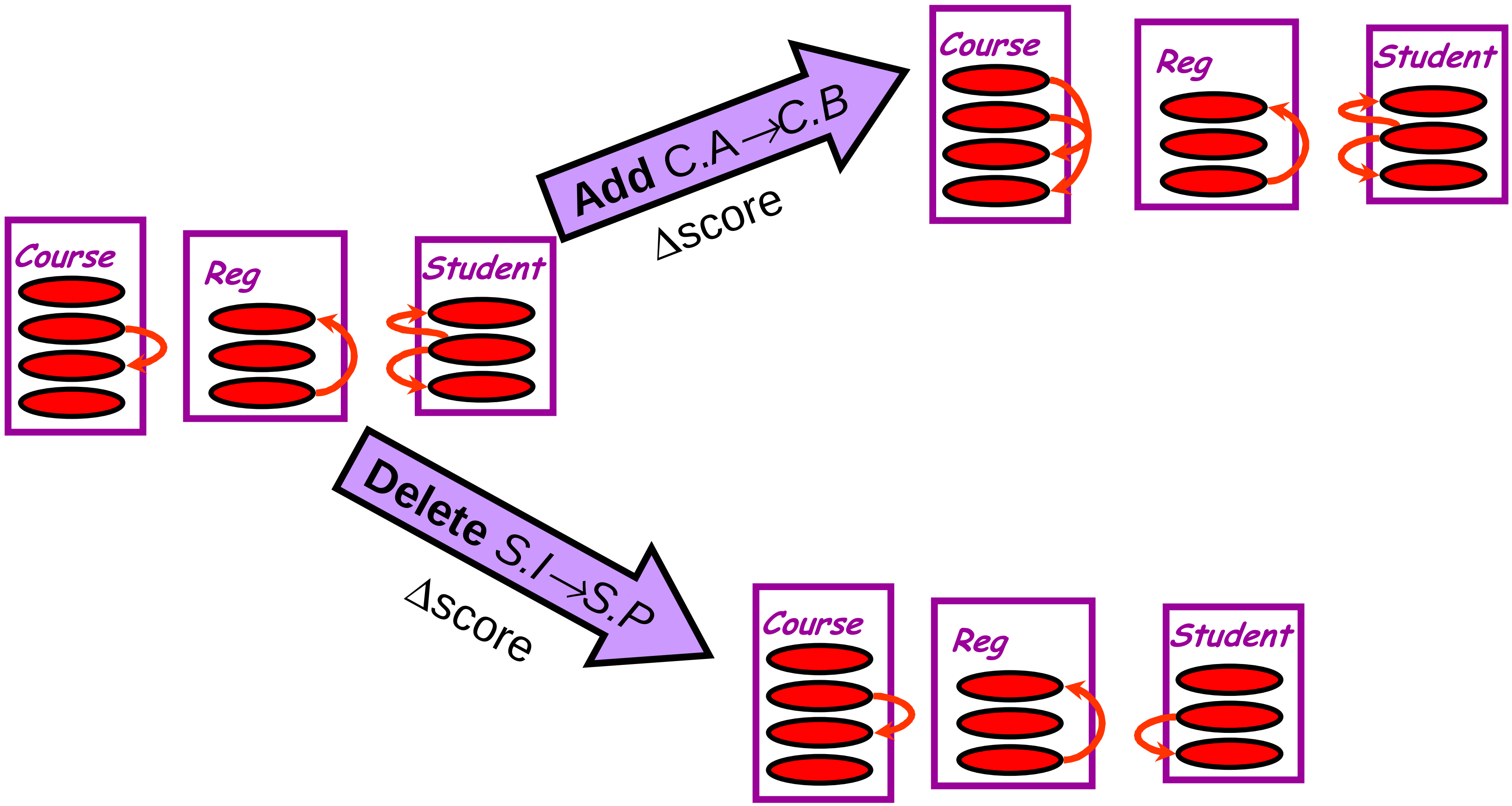
Learning RBN Structure

- All operations done at class level
  - Dependency structure = parents for X.A
  - Acyclicity checked using class dependency graph
  - Score computed at class level

- Individual objects only contribute to sufficient statistics
  - Can be obtained efficiently using standard DB queries

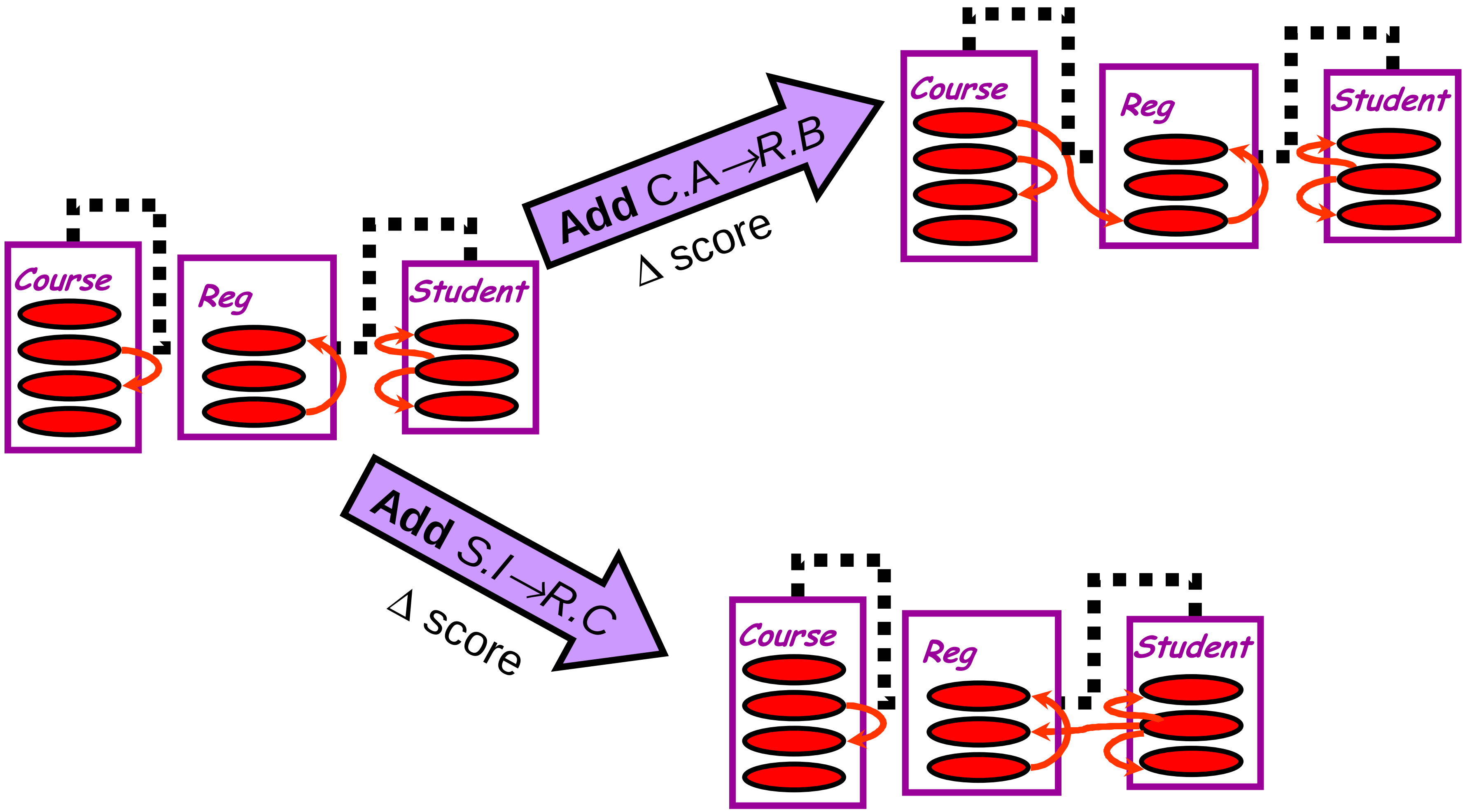
Exploiting Locality: Phased Search

Phase 0: consider only dependencies within a class



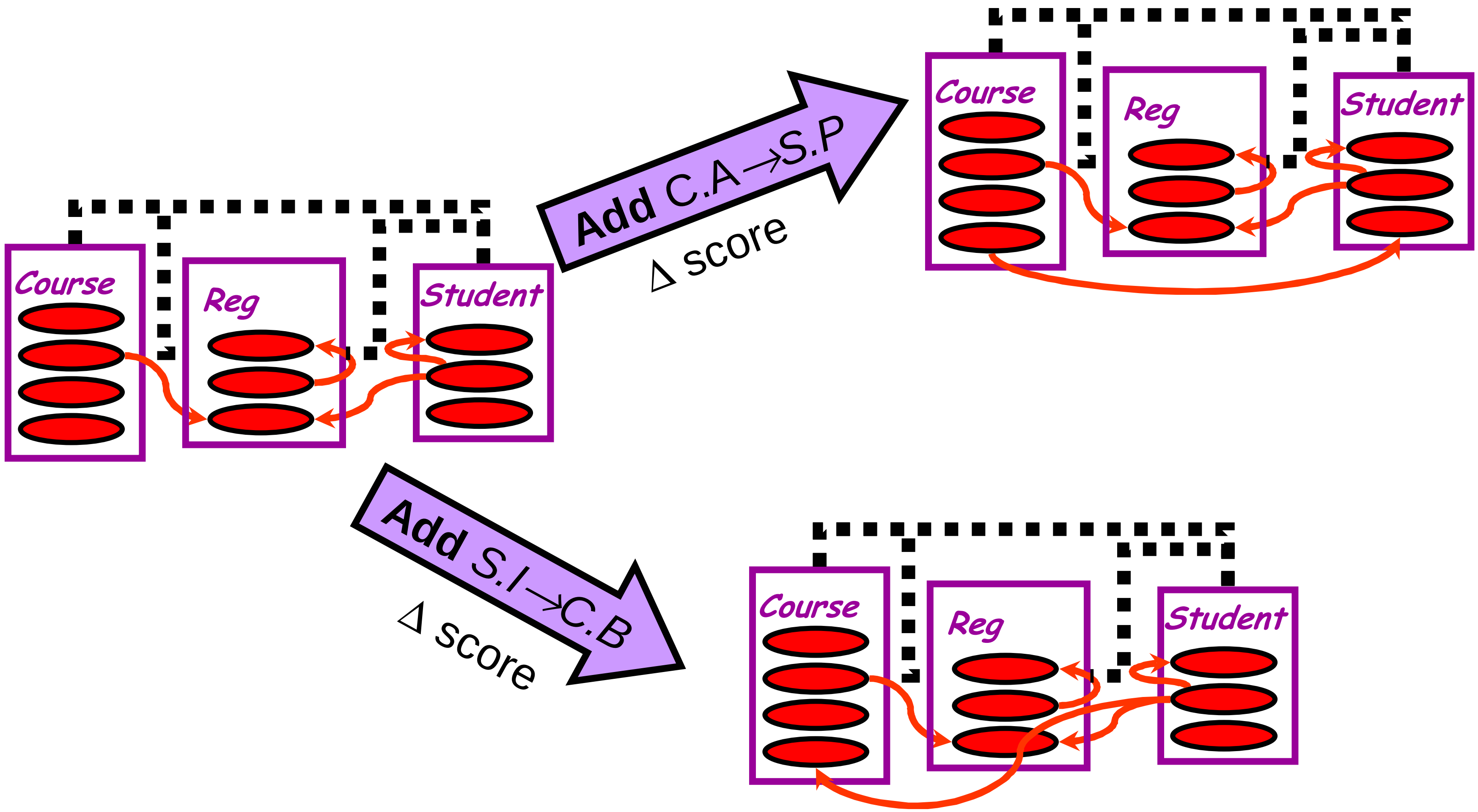
Exploiting Locality: Phased Search

Phase 1: consider dependencies one link away

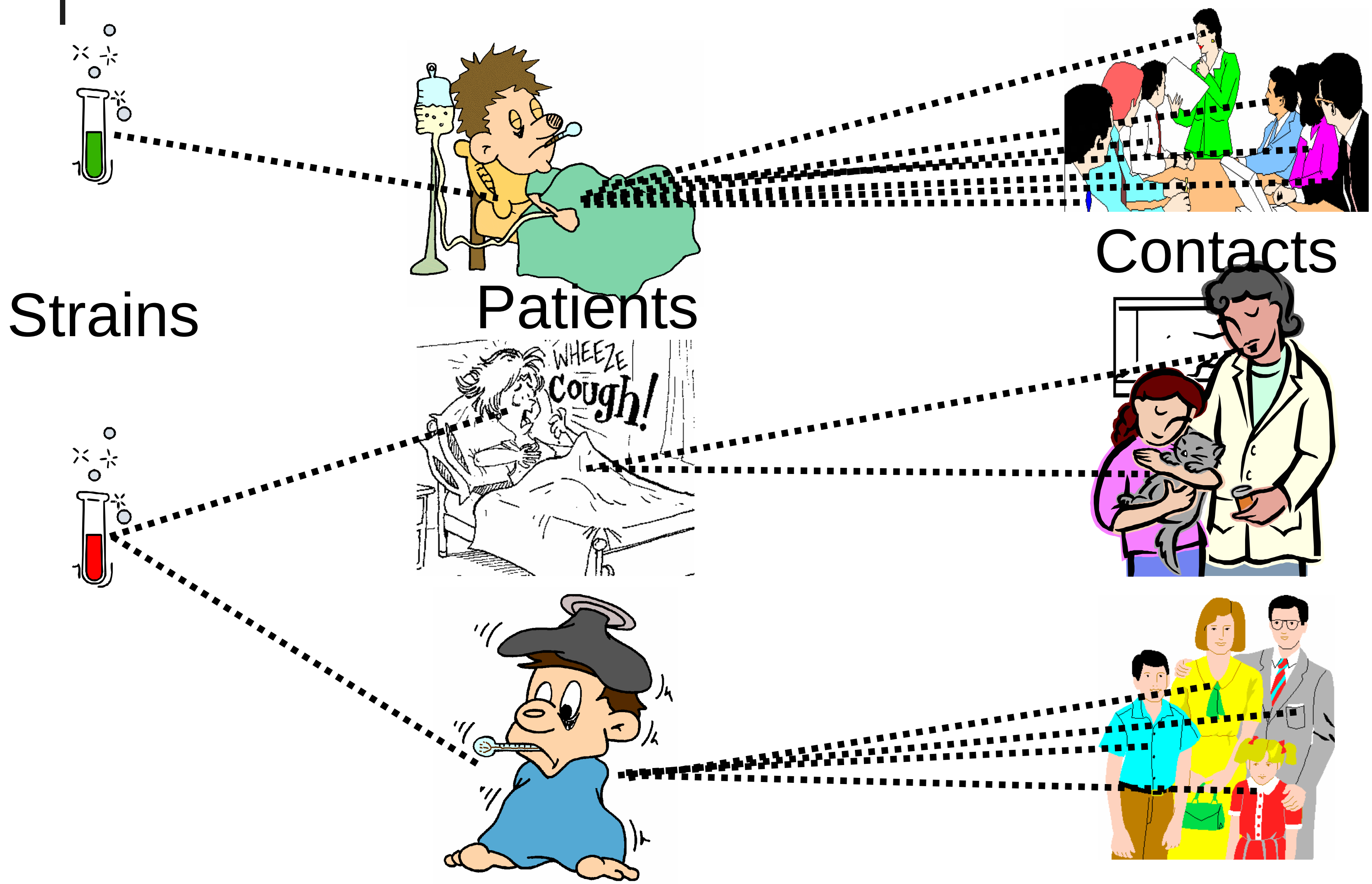


Exploiting Locality: Phased Search

Phase *k*: consider dependencies *k* links away

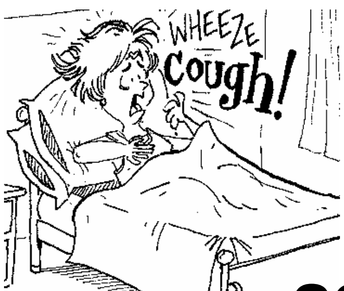


TB Patients in San Francisco



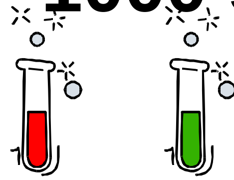
[Getoor, Rhee, K., Small, 2001]

TB Patients in San Francisco

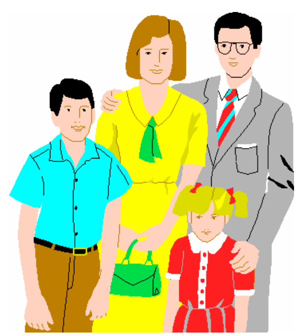


2300 patients

1000 strains

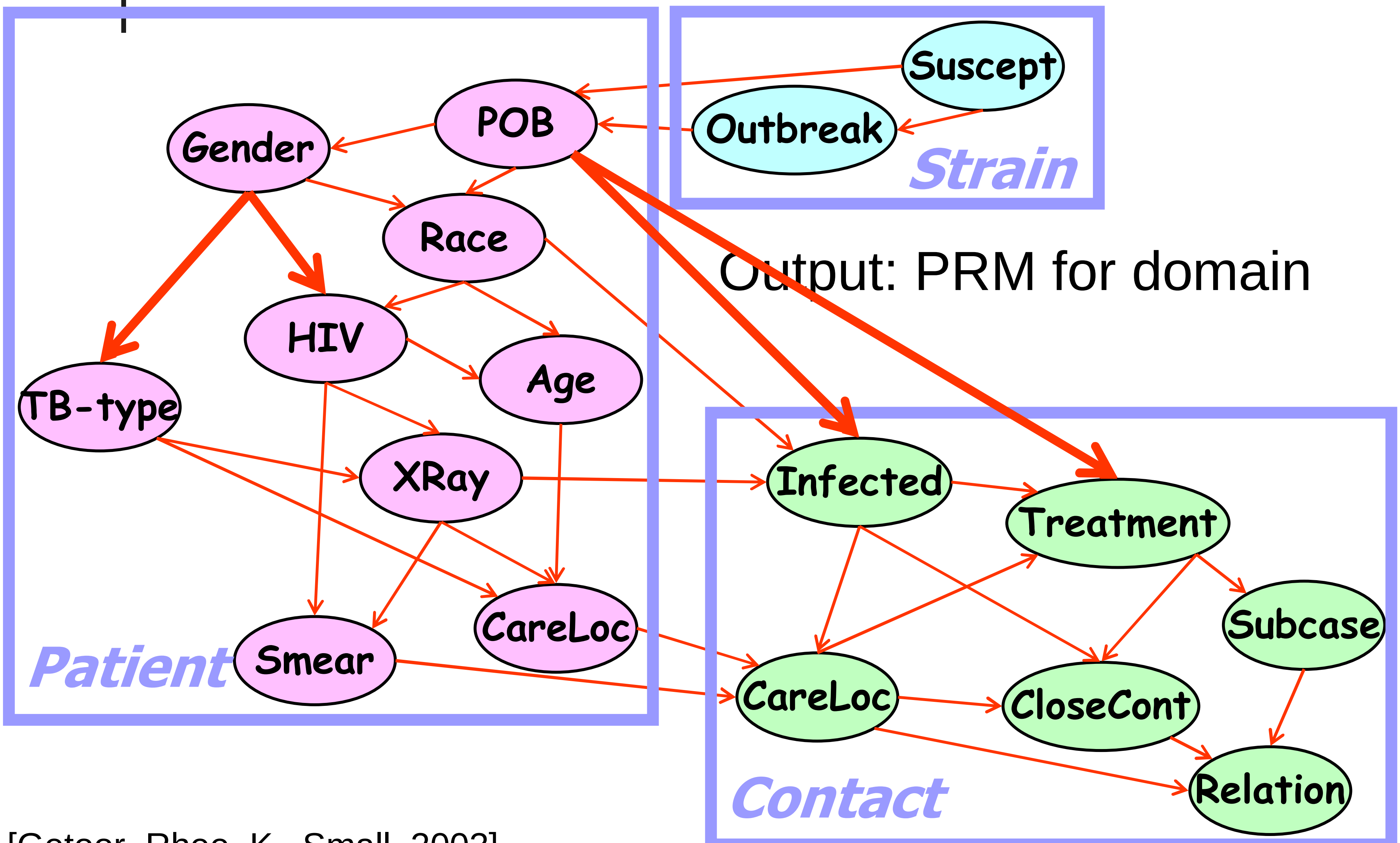


Input: Relational database



20000 contacts

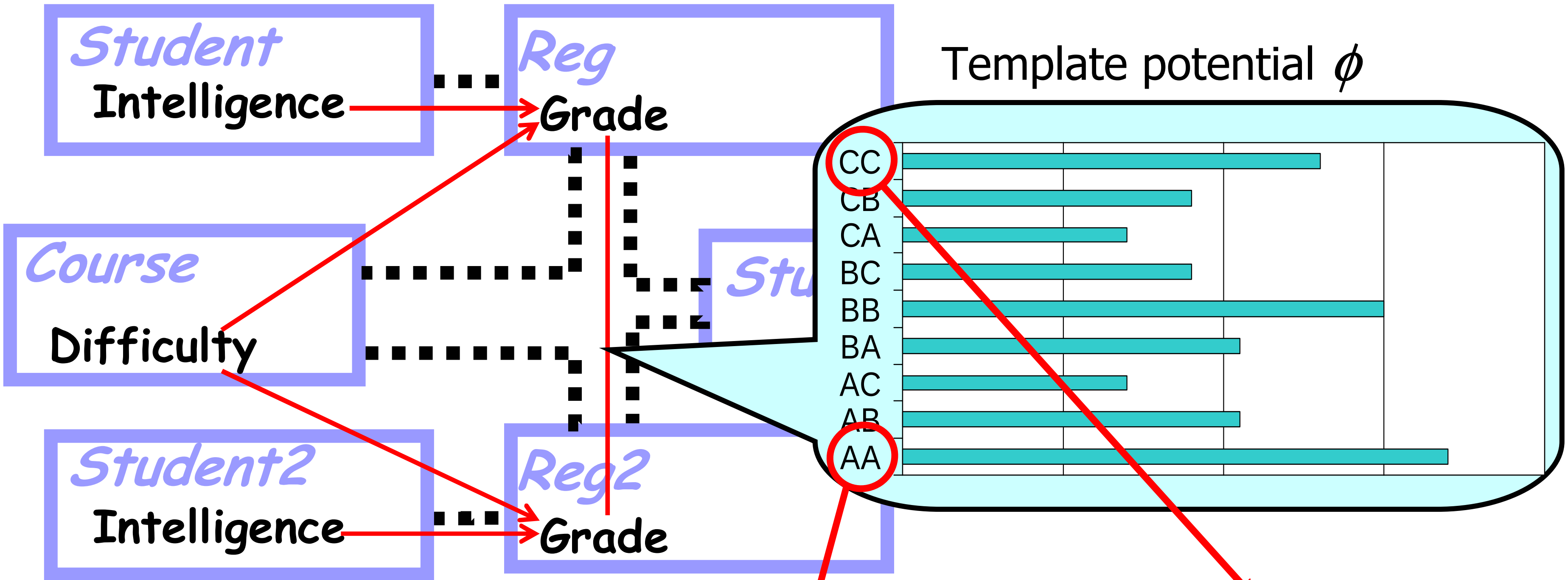
TB Patients in San Francisco



Output: PRM for domain

[Getoor, Rhee, K., Small, 2003]

Learning RMN Parameters



$$\phi = \exp\{w_{AA} \cdot f_{AA} + \ldots + w_{CC} \cdot f_{CC}\}$$

Parameterize potentials as log-linear model

[Taskar, Abbeel, K., 2002]

Learning RMN Parameters

$$\ell(\boldsymbol{w}:\omega) = \log P_{\boldsymbol{w}}(\omega \mid \boldsymbol{w}) = \sum_i w_i \cdot \underbrace{f_i(\omega)}_{\text{Counts in } \omega} - \log Z$$

For example:

$f_{AA}(\omega)$ = # of tuples {reg r_1, reg r_2, group g} s.t.

In(g, r_1), In(g, r_2); r_1.Grade=A, r_2.Grade=A

Learning RMN Parameters

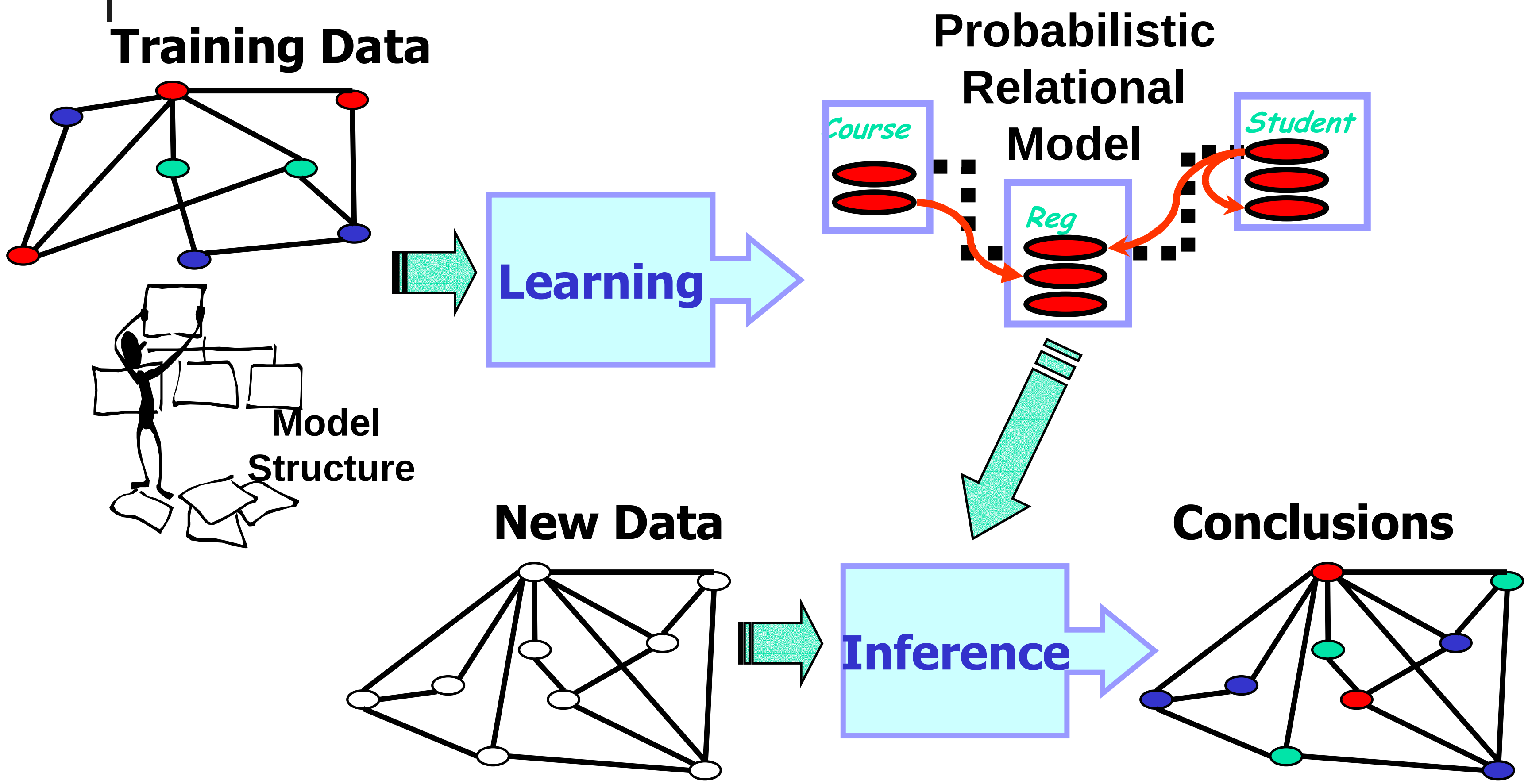
- Parameter estimation is not closed form
- Convex problem $\Rightarrow$ unique global maximum
- Can use methods such as conjugate gradient

$$\frac{\partial \ell}{\partial w_{AA}} = \#(Grade = A, Grade = A) \quad \text{actual count}$$

$$- \sum P(Grade = A, Grade = A) \quad \text{- expected count}$$

- Gradient process tries to find parameters s.t. expected counts = actual counts
- Computing expected counts requires inference over ground Markov network

Collective Classification



Example:

- Train on one year of student intelligence, course difficulty, and grades
- Given only grades in following year, predict all students’ intelligence

Discriminative Training

- Goal: Given values of observed variables ω.O=o, predict desired target values ω.T=t*
- Do not necessarily want the model to fit the joint distribution P(ω.O=o, ω.T=t*)
- To maximize classification accuracy, we can consider other optimization criteria
  - Maximize conditional log likelihood

$$P(\omega.T = t^* \mid \omega.O = o)$$

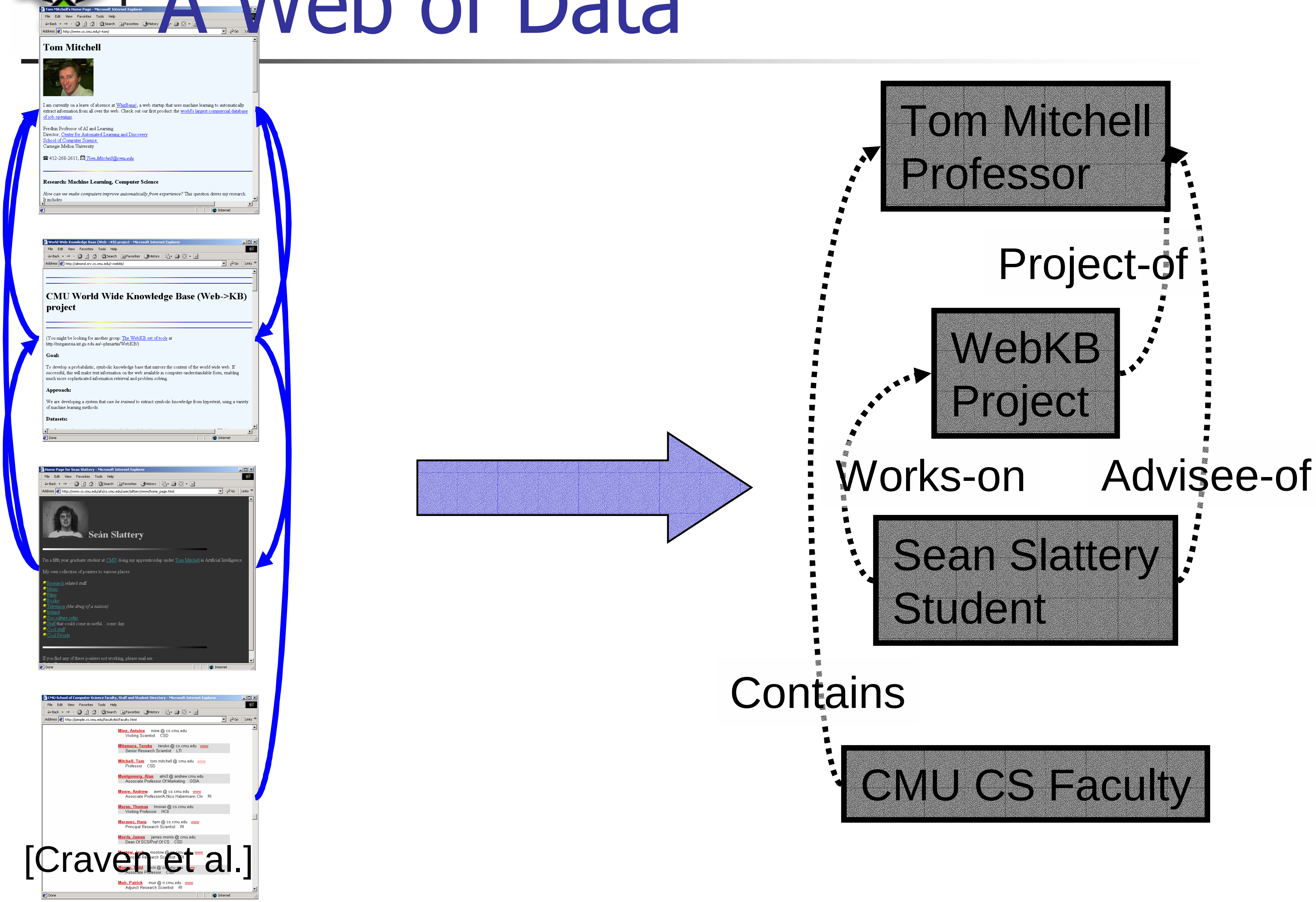
  - Maximize margin

$$\log P(\omega.T = t^* \mid \omega.O = o)$$

$$-\log[\underbrace{\max_{t \neq t^*} P(\omega.T = t \mid \omega.O = o)}_{\text{P(second highest probability label)}}]$$

[Taskar, Abbeel, K., 2002;
Taskar, Guestrin, K. 2003]

A Web of Data



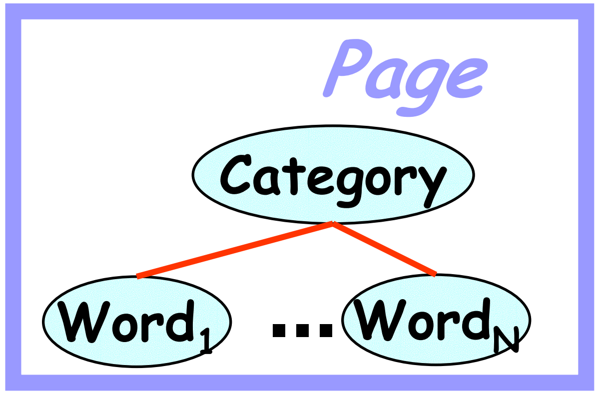
[Craven et al.]

Web Classification Experiments

- WebKB dataset
  - Four CS department websites
  - Bag of words on each page
  - Links between pages
  - Anchor text for links
- Experimental setup
  - Trained on three universities
  - Tested on fourth
  - Repeated for all four combinations

Standard Classification

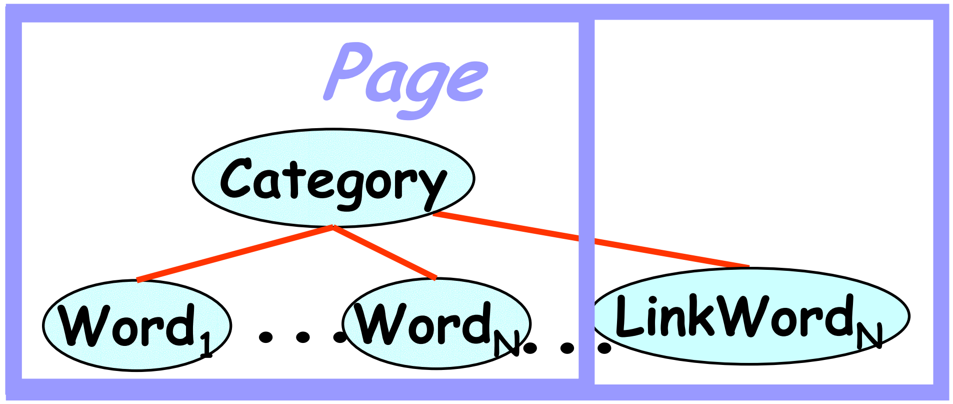
Professor department extract information computer science machine learning ...

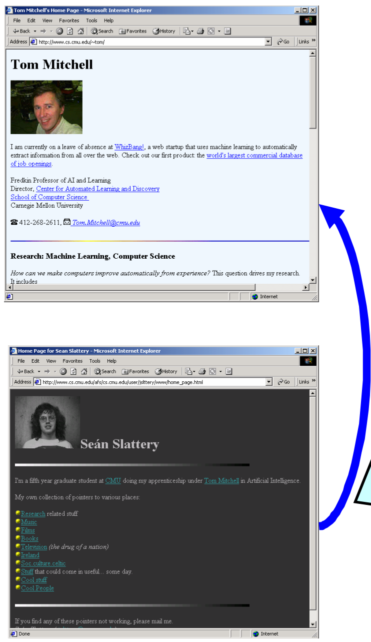


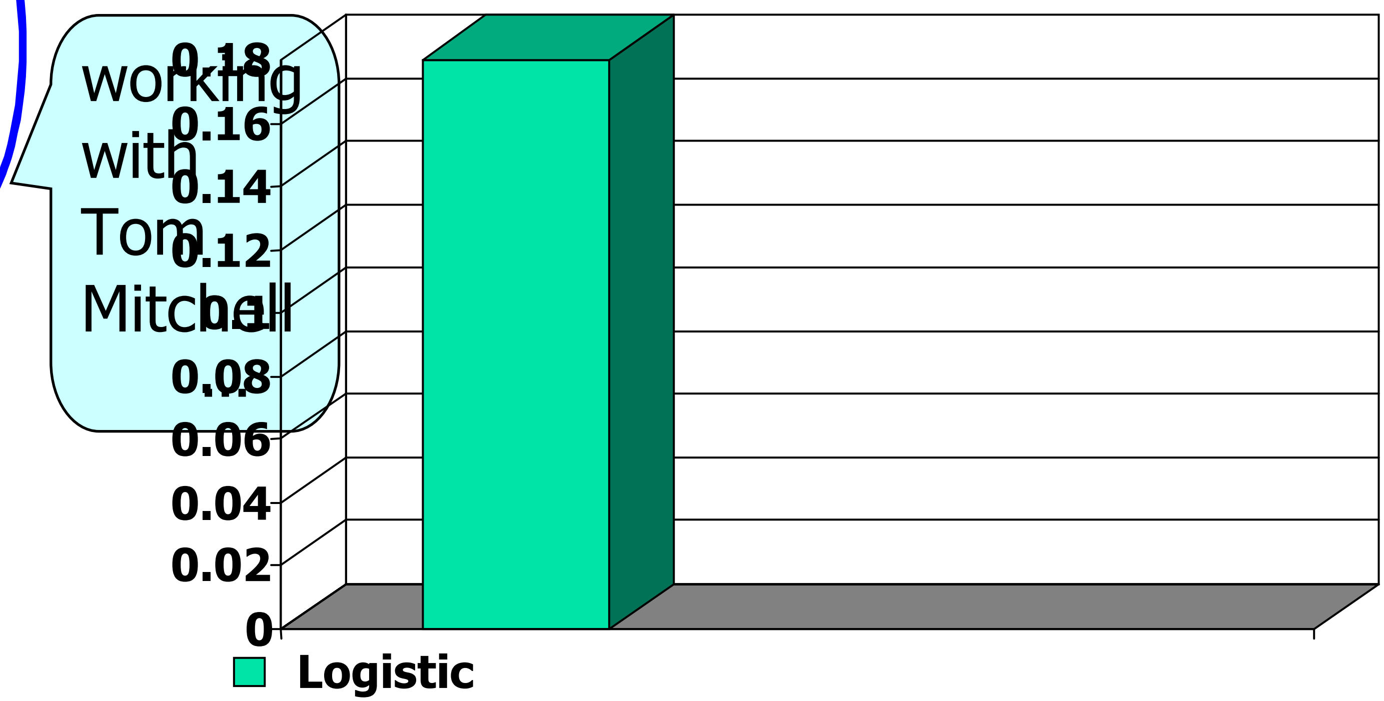
Logistic Regression

Categories:
faculty
course
project
student
other

Standard Classification

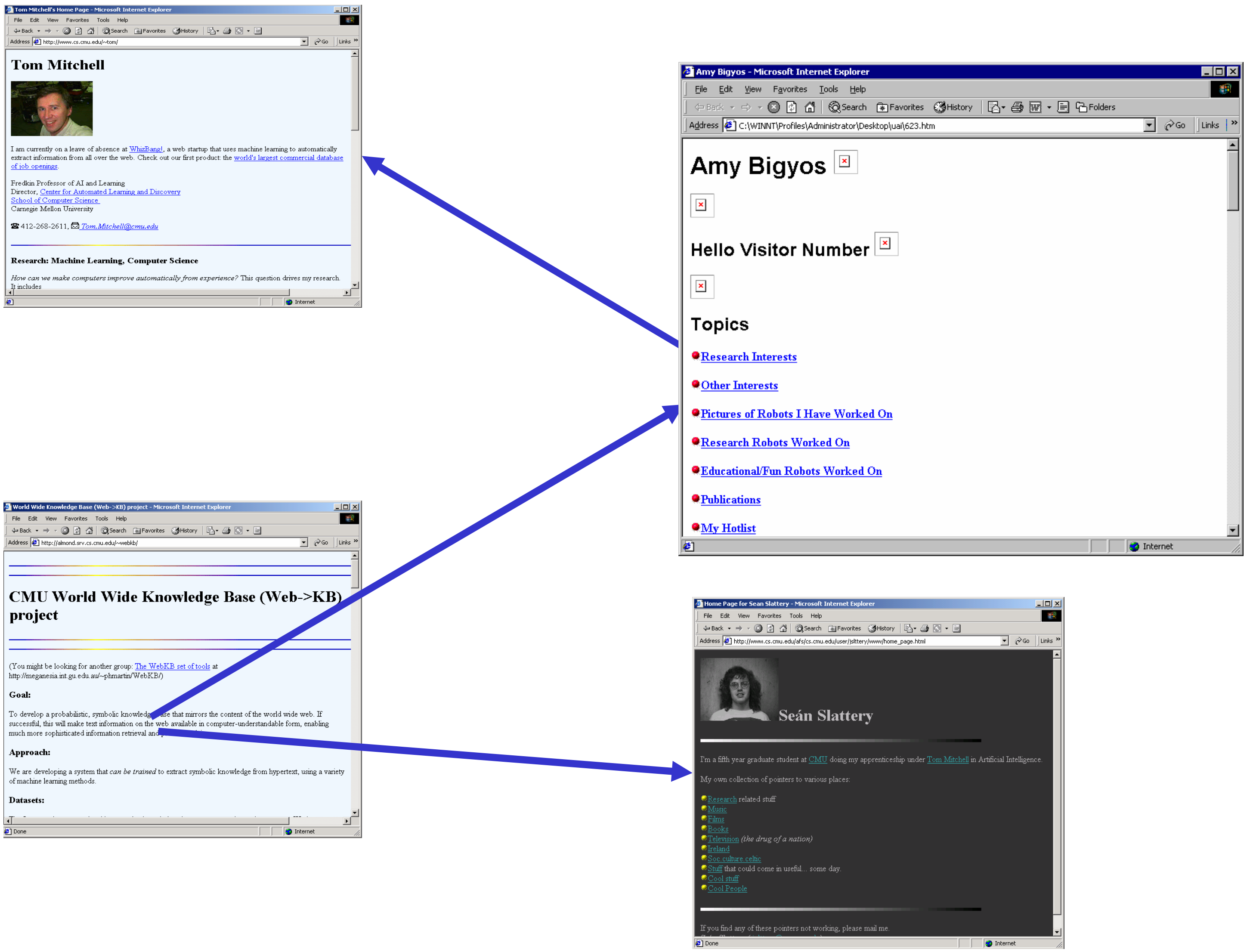




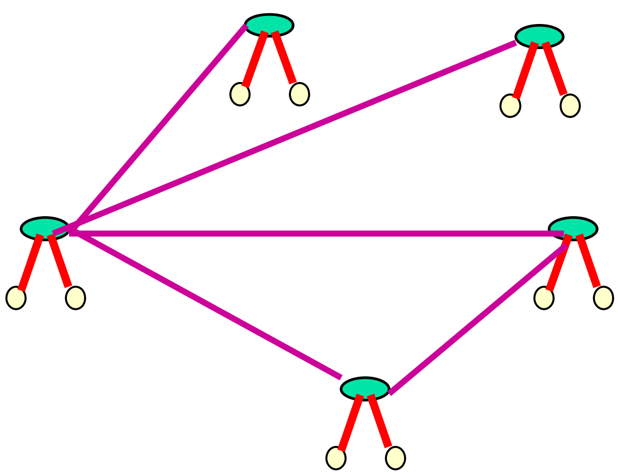


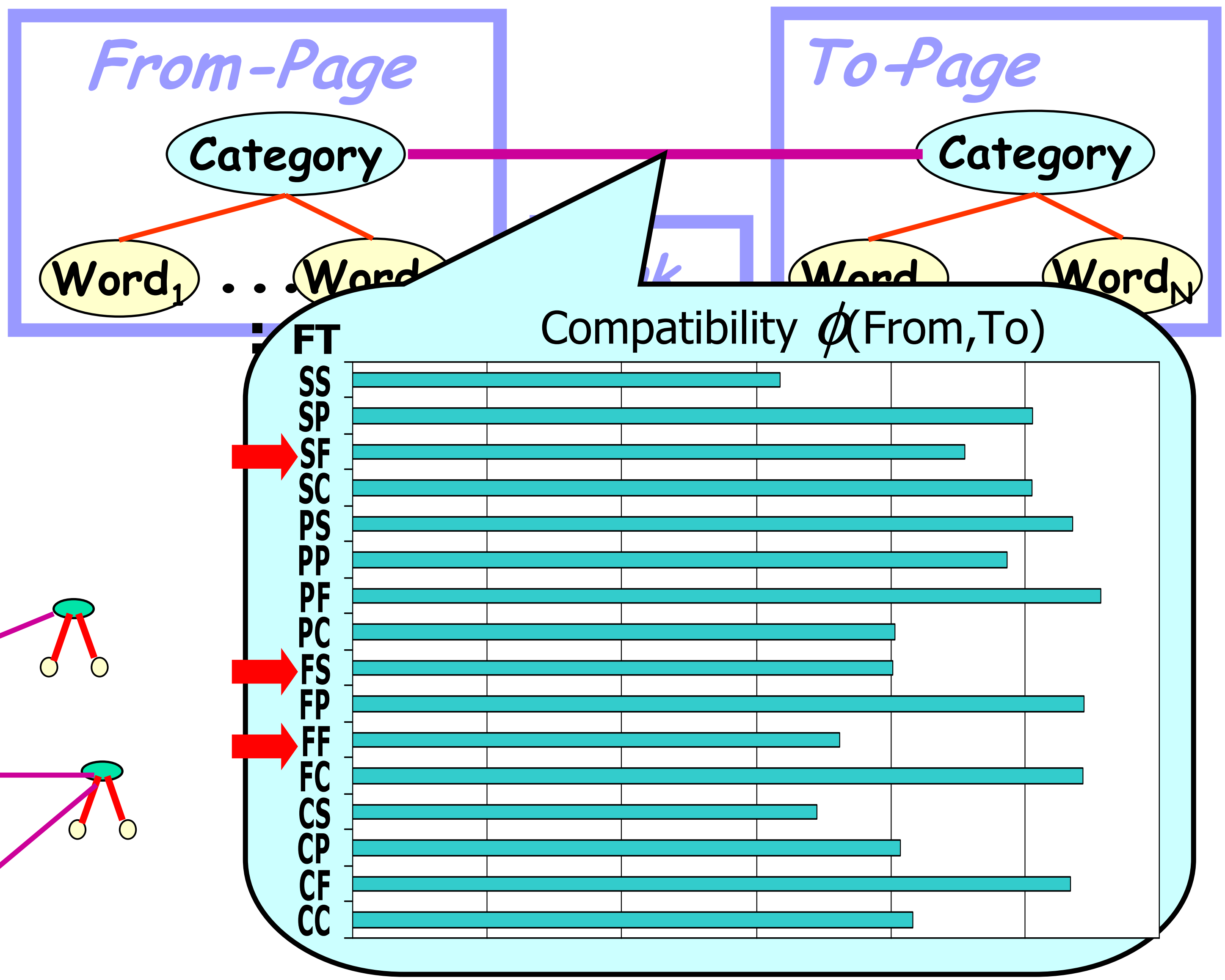
Power of Context

Professor? ***Student?*** ***Post-doc?***

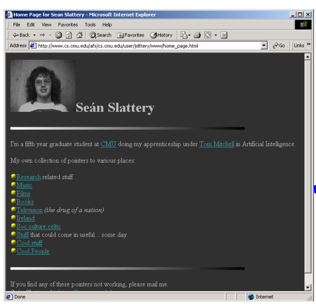


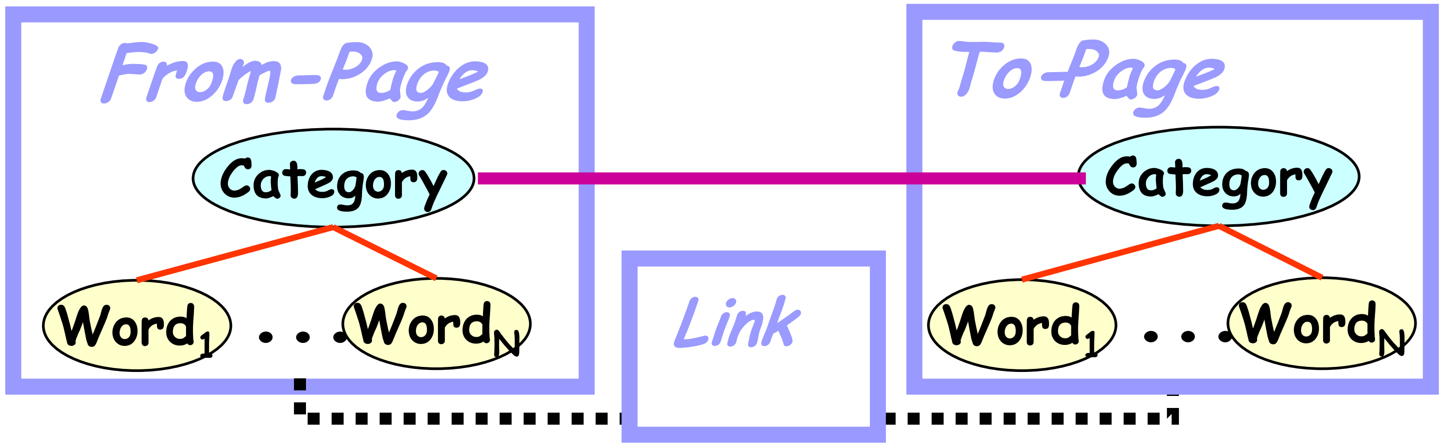
Collective Classification

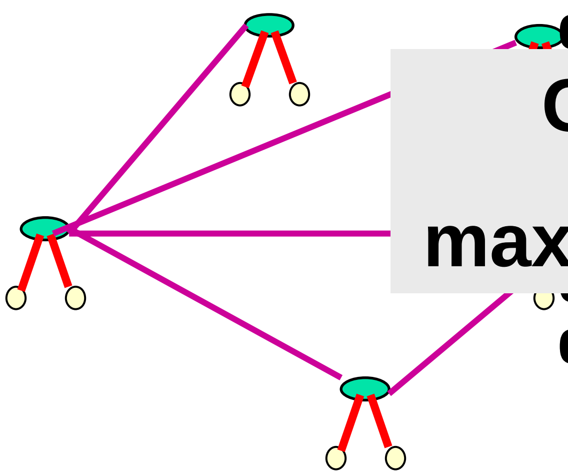


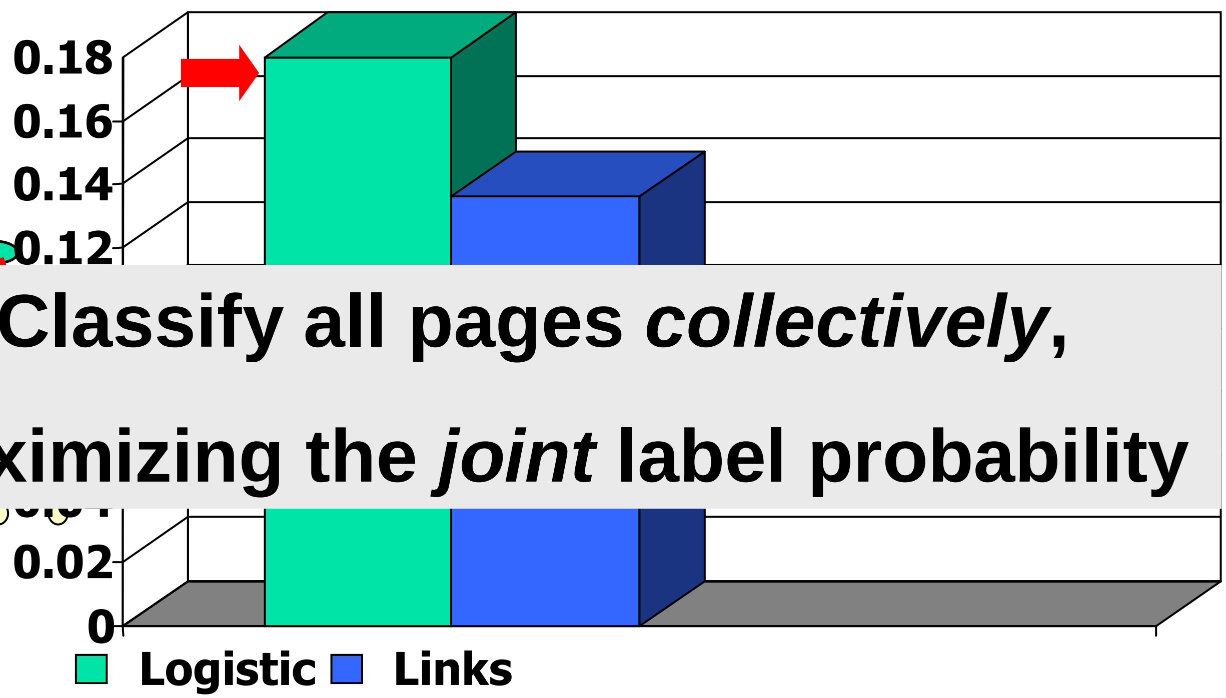


Collective Classification



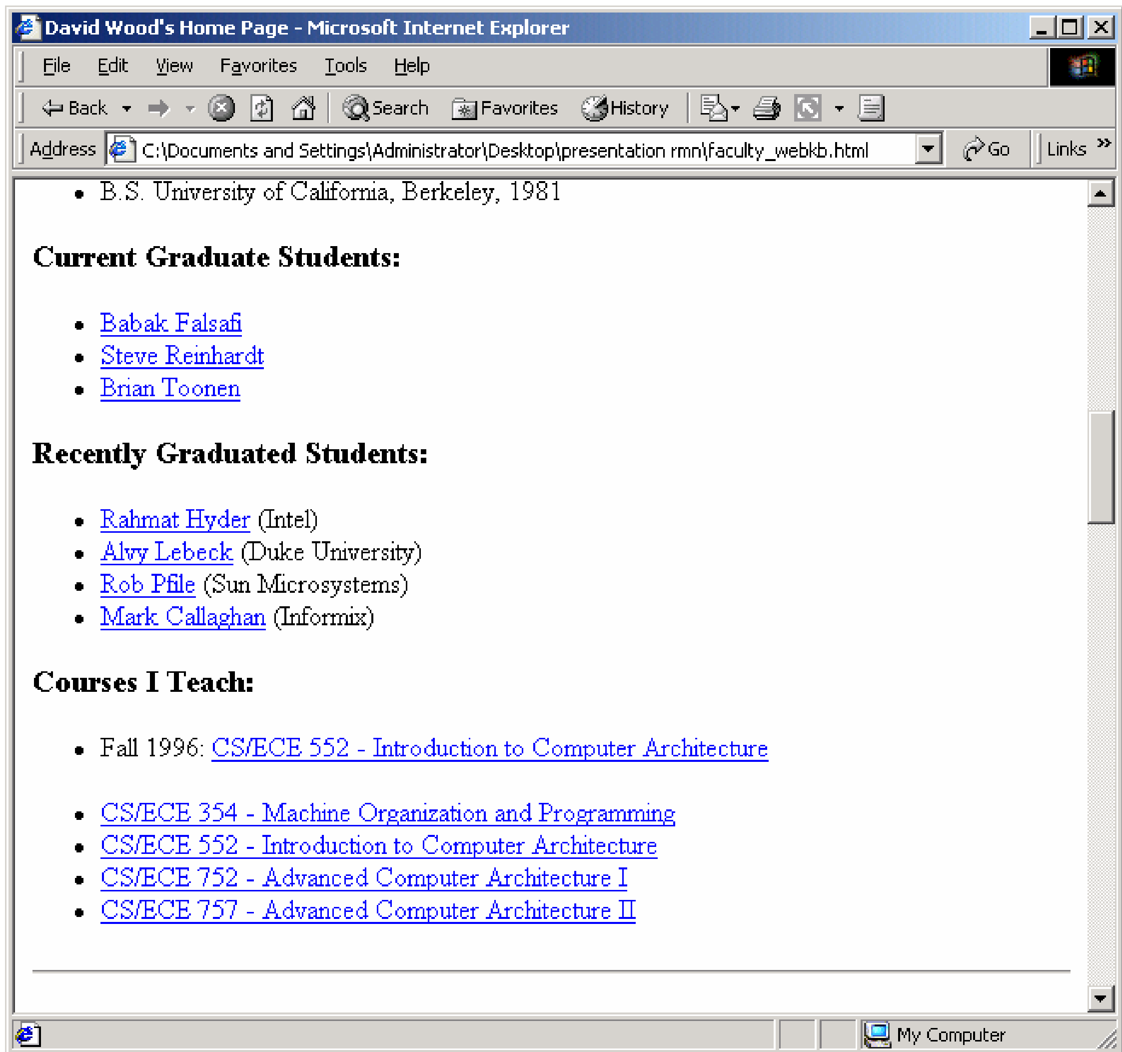




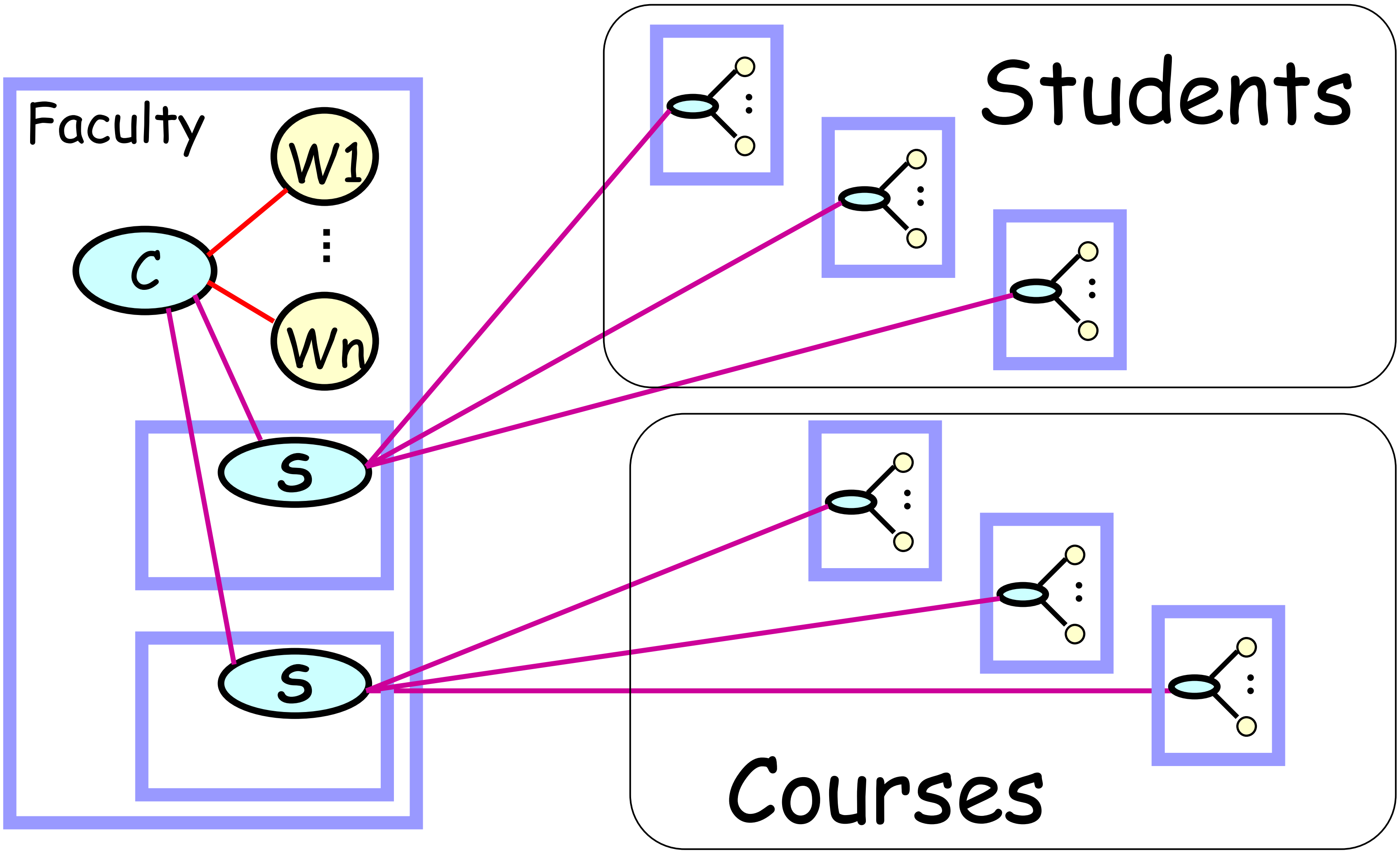


Classify all pages *collectively*, maximizing the *joint* label probability

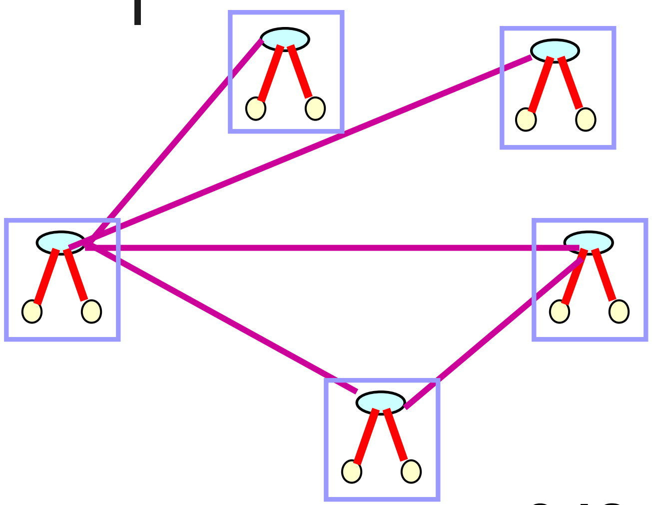
More Complex Structure

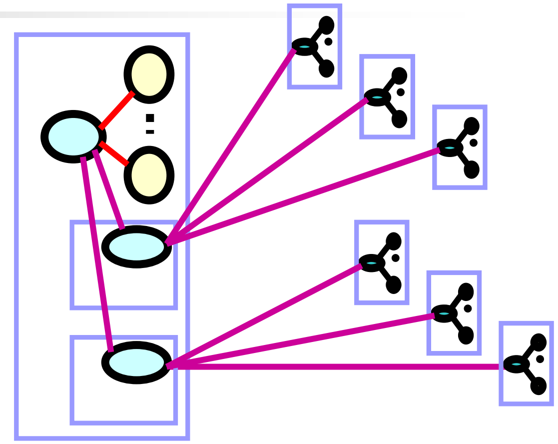


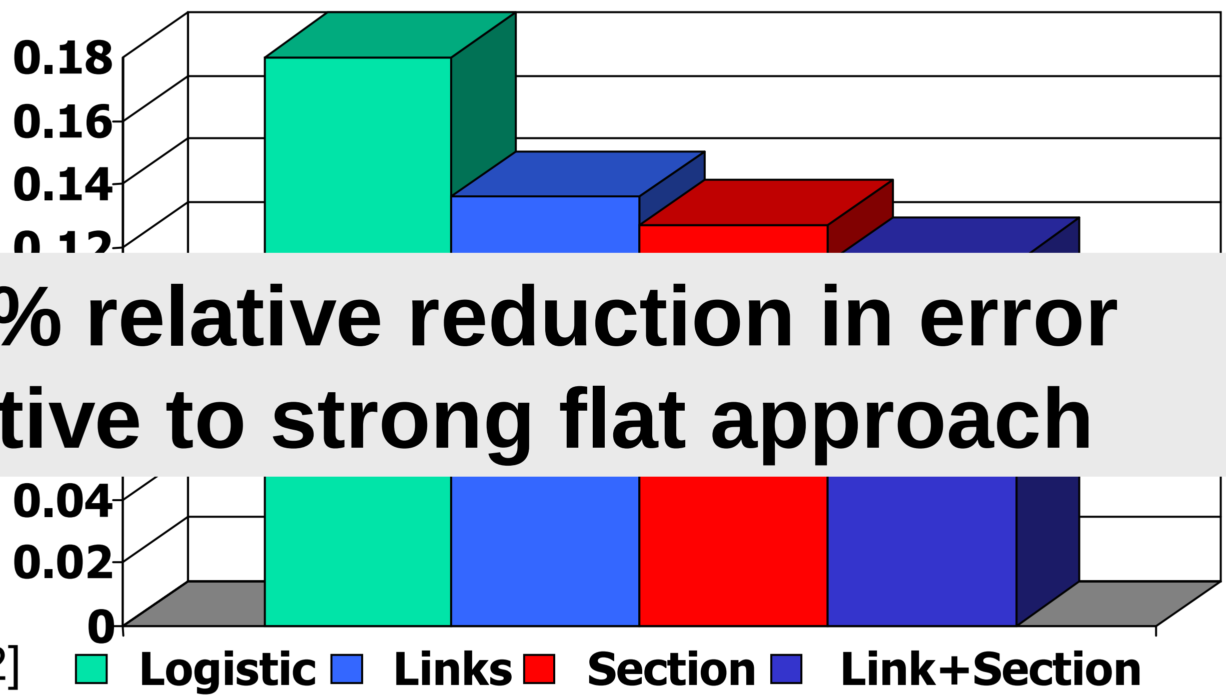
More Complex Structure



Collective Classification: Results



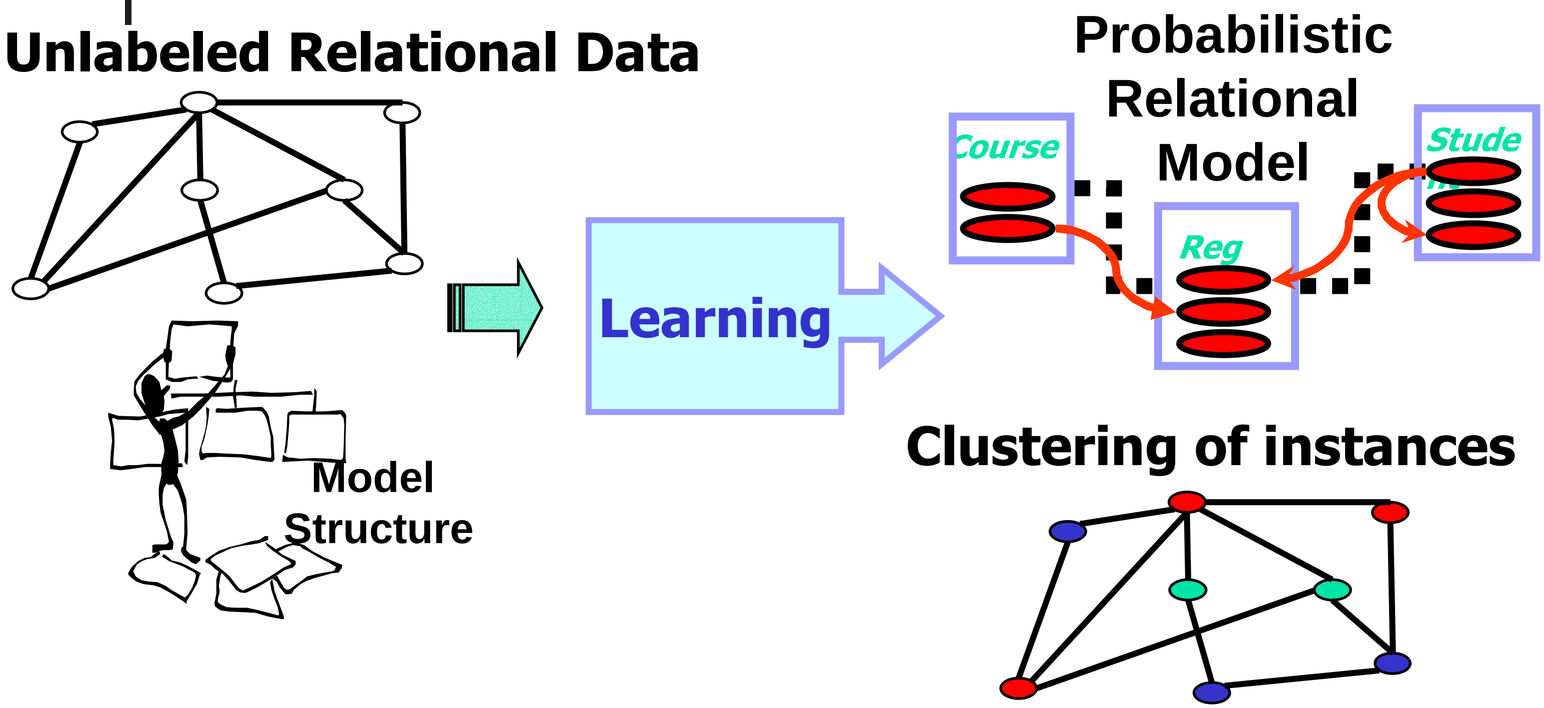




35.4% relative reduction in error relative to strong flat approach

[Taskar, Abbeel, K., 2002]

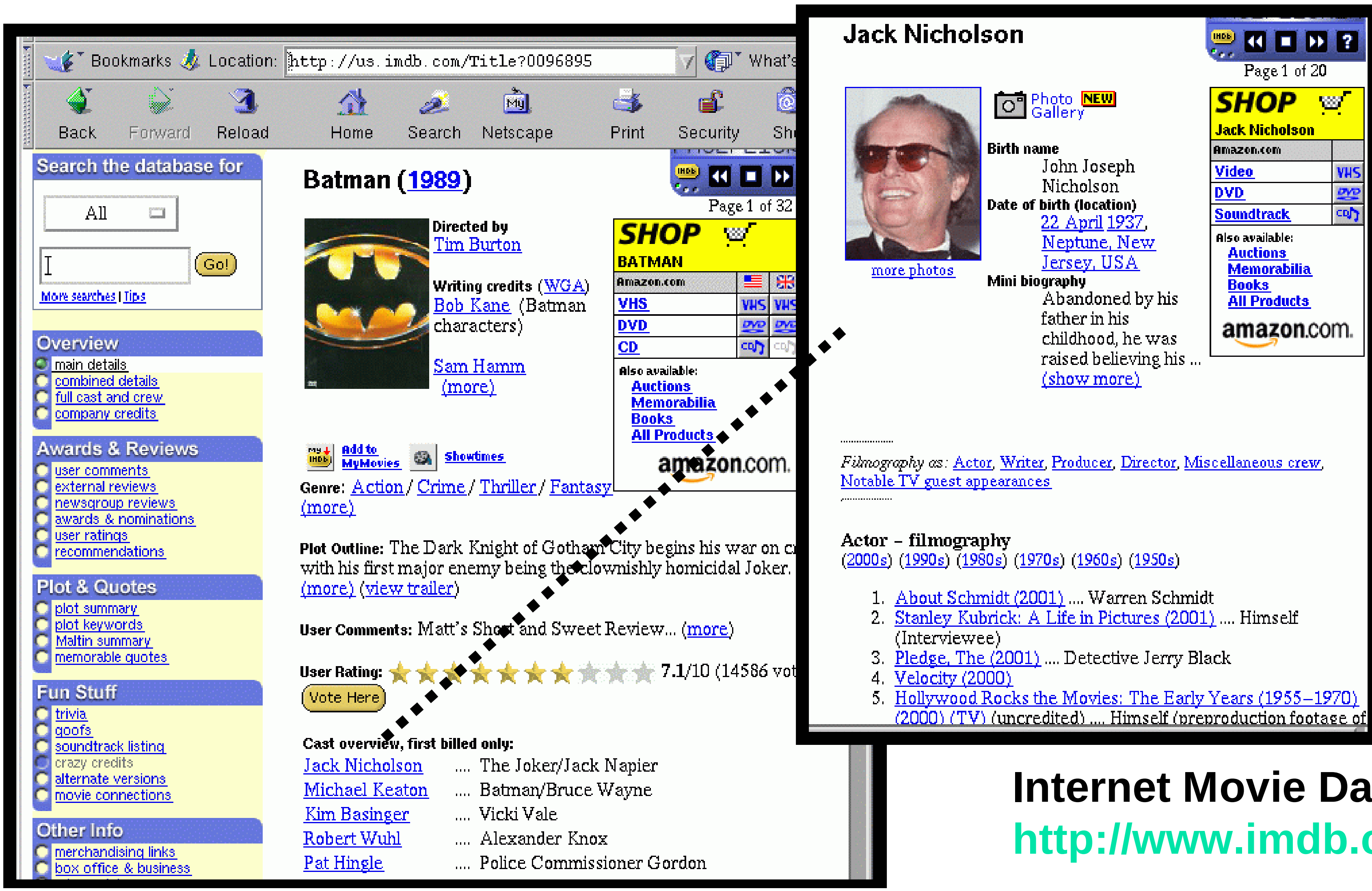
Relational Clustering



Example:

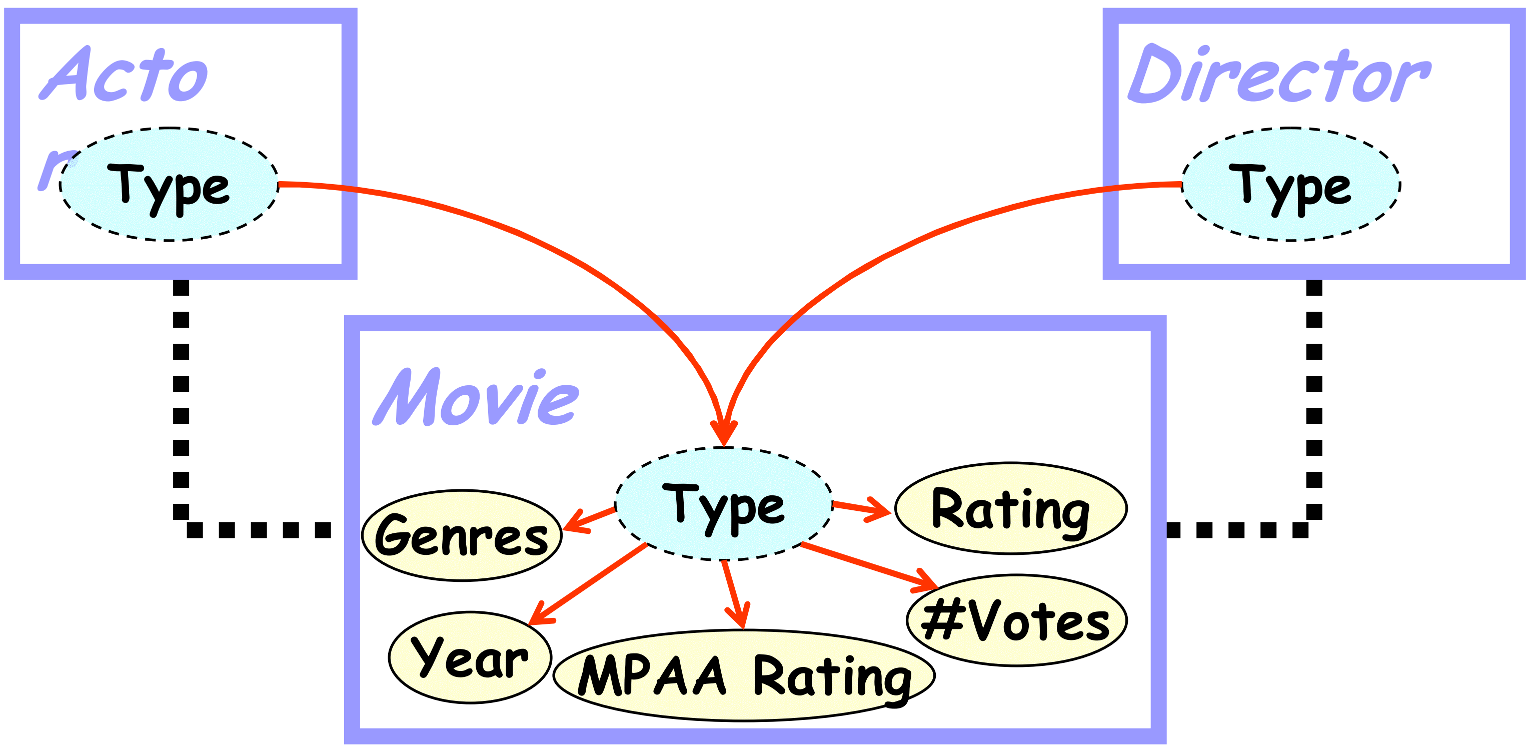
- Given only students' grades, cluster similar students

Movie Data



Internet Movie Database
http://www.imdb.com

Discovering Hidden Types



[Taskar, Segal, K., 2001]

Discovering Hidden Types

Movies

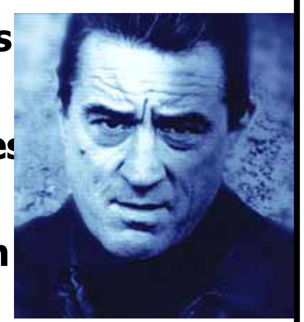
Wizard of Oz
Cinderella
Sound of Music
The Love Bug
Pollyanna
The Parent Trap
Mary Poppins
Swiss Family Robinson

Terminator 2
Batman
Batman Forever
GoldenEye
Starship Troopers
Mission: Impossible
Hunt for Red October

...

Actors

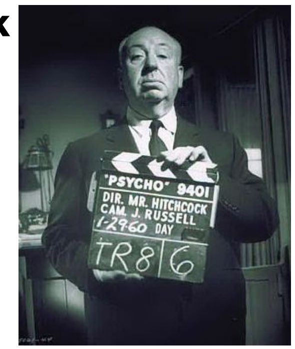
Sylvester Stallone
Bruce Willis
Harrison Ford
Steven Seagal
Kurt Russell
Kevin Costner
Jean-Claude Van Damme
Arnold Schwarzenegger



Anthony Hopkins
Robert De Niro
Tommy Lee Jones
Harvey Keitel
Morgan Freeman
Gary Oldman

...

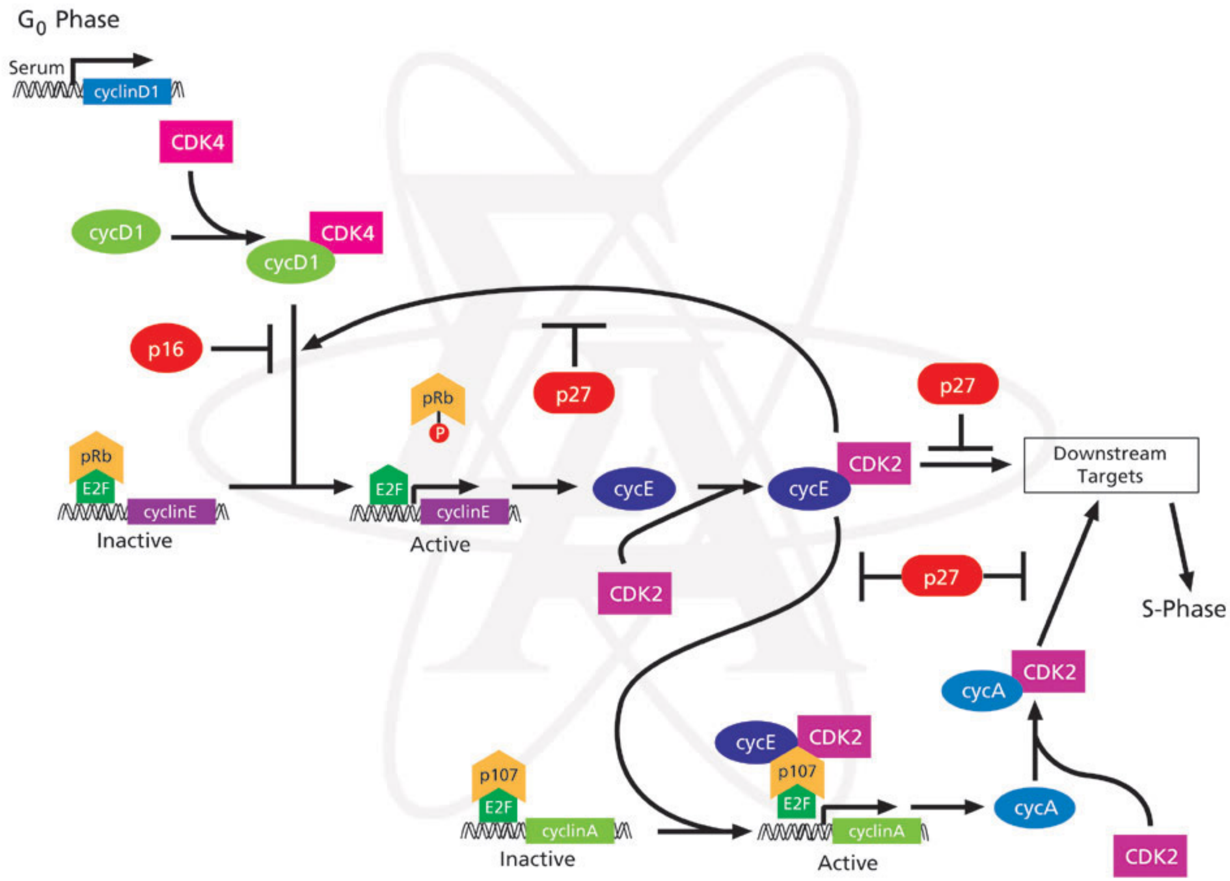
Directors



Alfred Hitchcock
Stanley Kubrick
David Lean
Milos Forman
Terry Gilliam
Francis Coppola

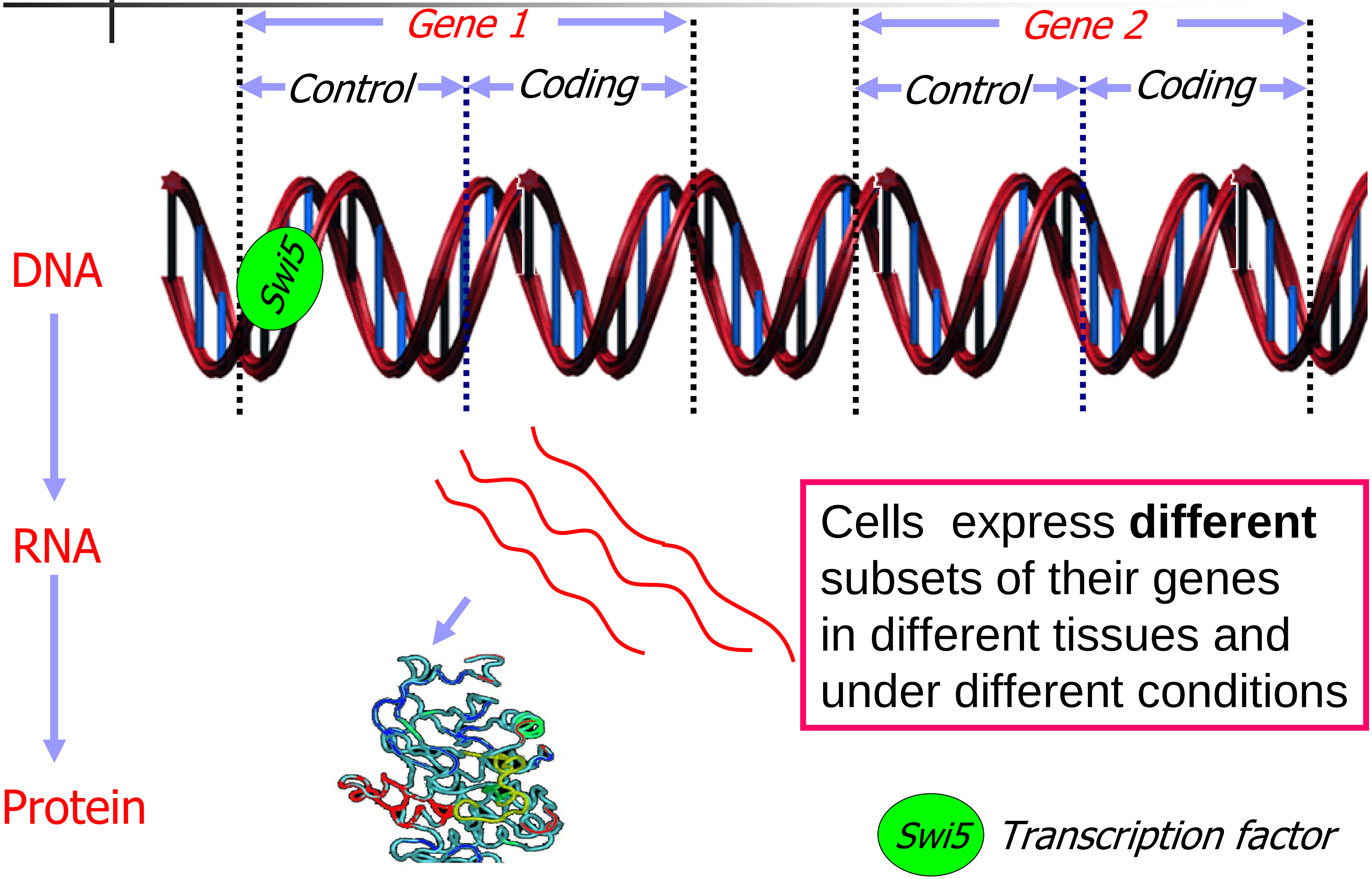
Steven Spielberg
Tim Burton
Tony Scott
James Cameron
John McTiernan
Joel Schumacher

Biology 101: Pathways



- Pathways are sets of genes that act together to achieve a common function

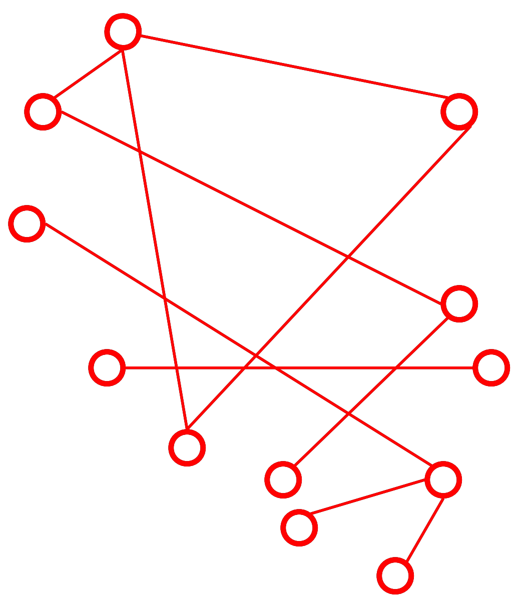
Biology 101: Expression



Cells express **different** subsets of their genes in different tissues and under different conditions

Finding Pathways: Attempt I

- **Use protein-protein interaction data**



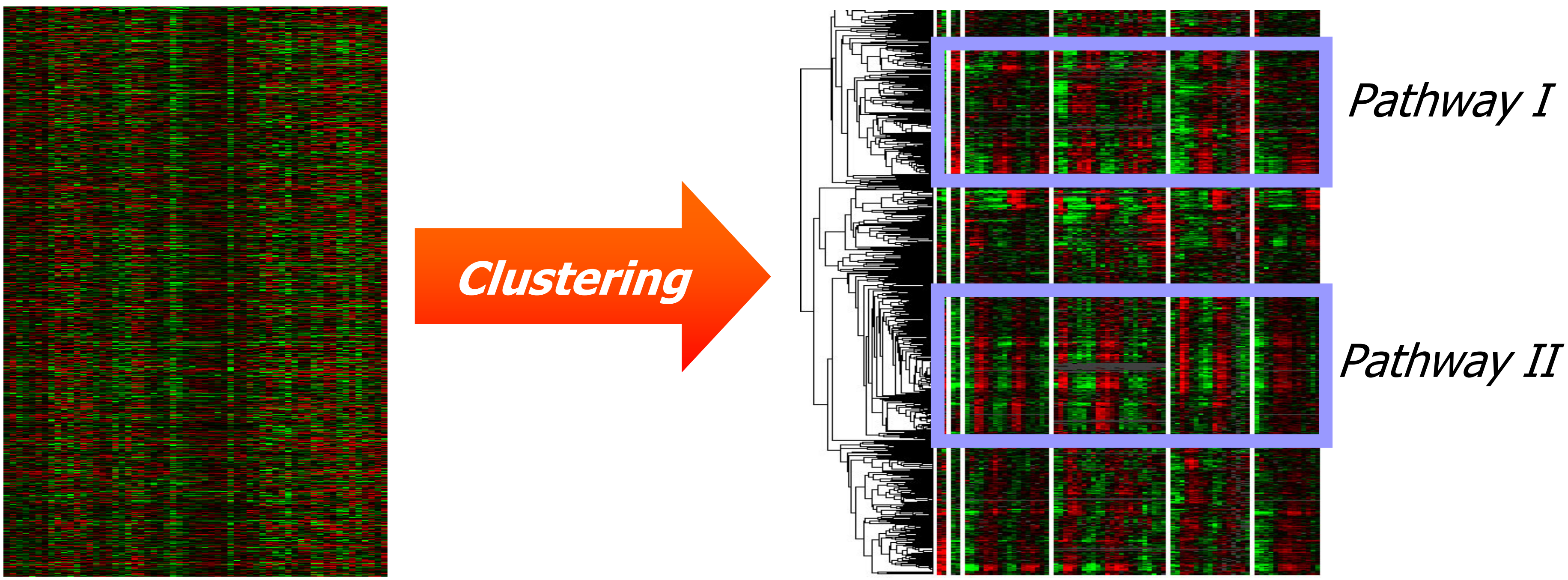
Pathway III

Finding Pathways: Attempt I

- **Use protein-protein interaction data**
- **Problems:**
  - Data is very noisy
  - Structure is lost:
    - Large connected component (3527/3589 genes) in interaction graph

Finding Pathways: Attempt II

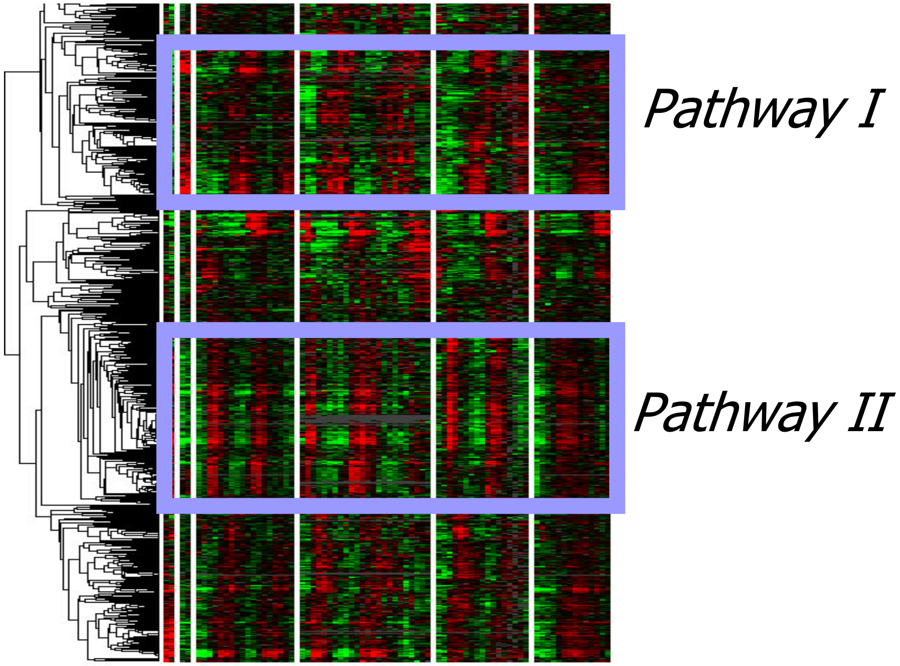
- **Use gene expression data**
  - Thousands of arrays available under different conditions



Finding Pathways: Attempt II

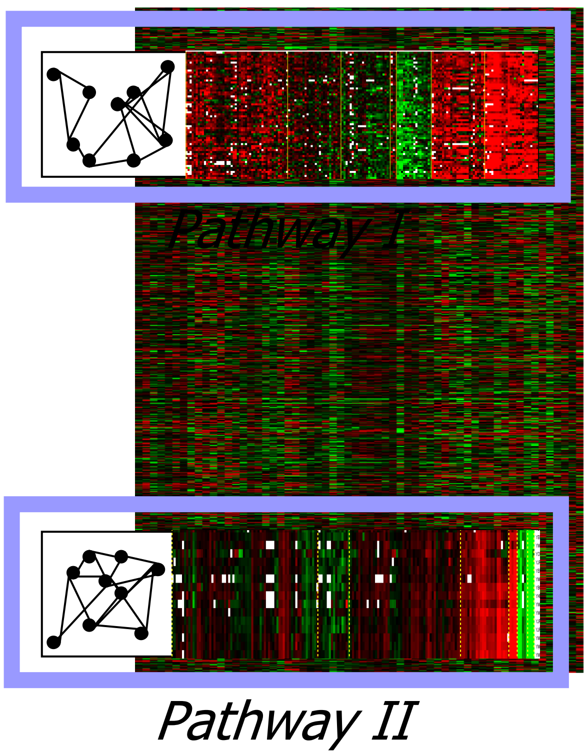
- **Use gene expression data**
  - Thousands of arrays available under different conditions

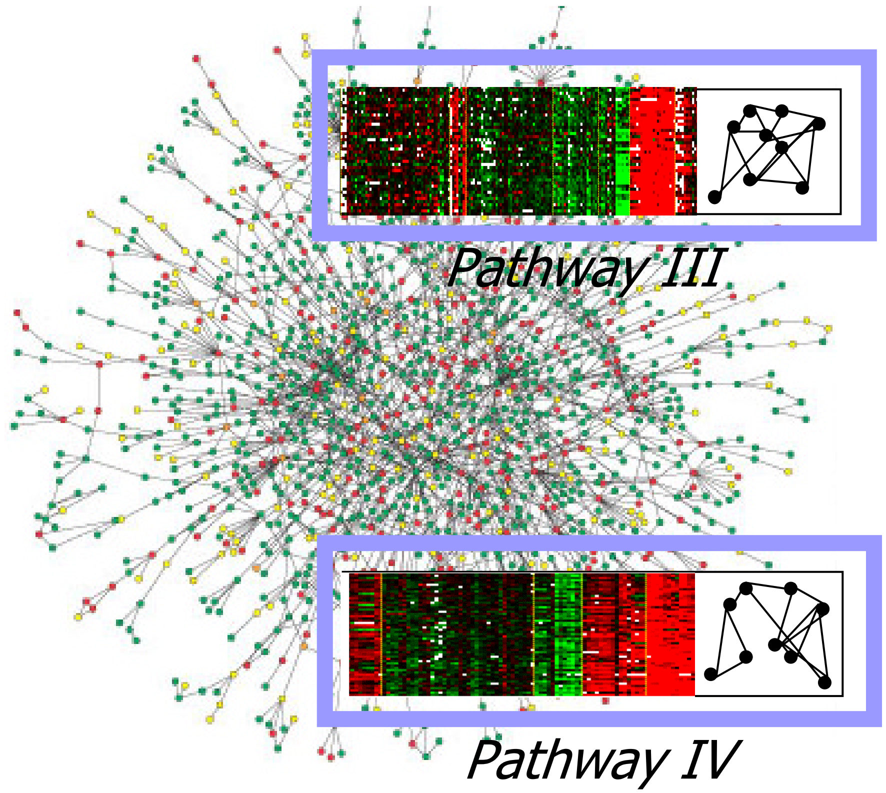
- **Problems:**
  - Expression is only 'weak' indicator of interaction
  - Data is noisy
  - Interacting pathways are not separable



Finding Pathways: Our Approach

- **Use both types of data to find pathways**
  - Find "active" interactions using gene expression
  - Find pathway-related co-expression using interactions





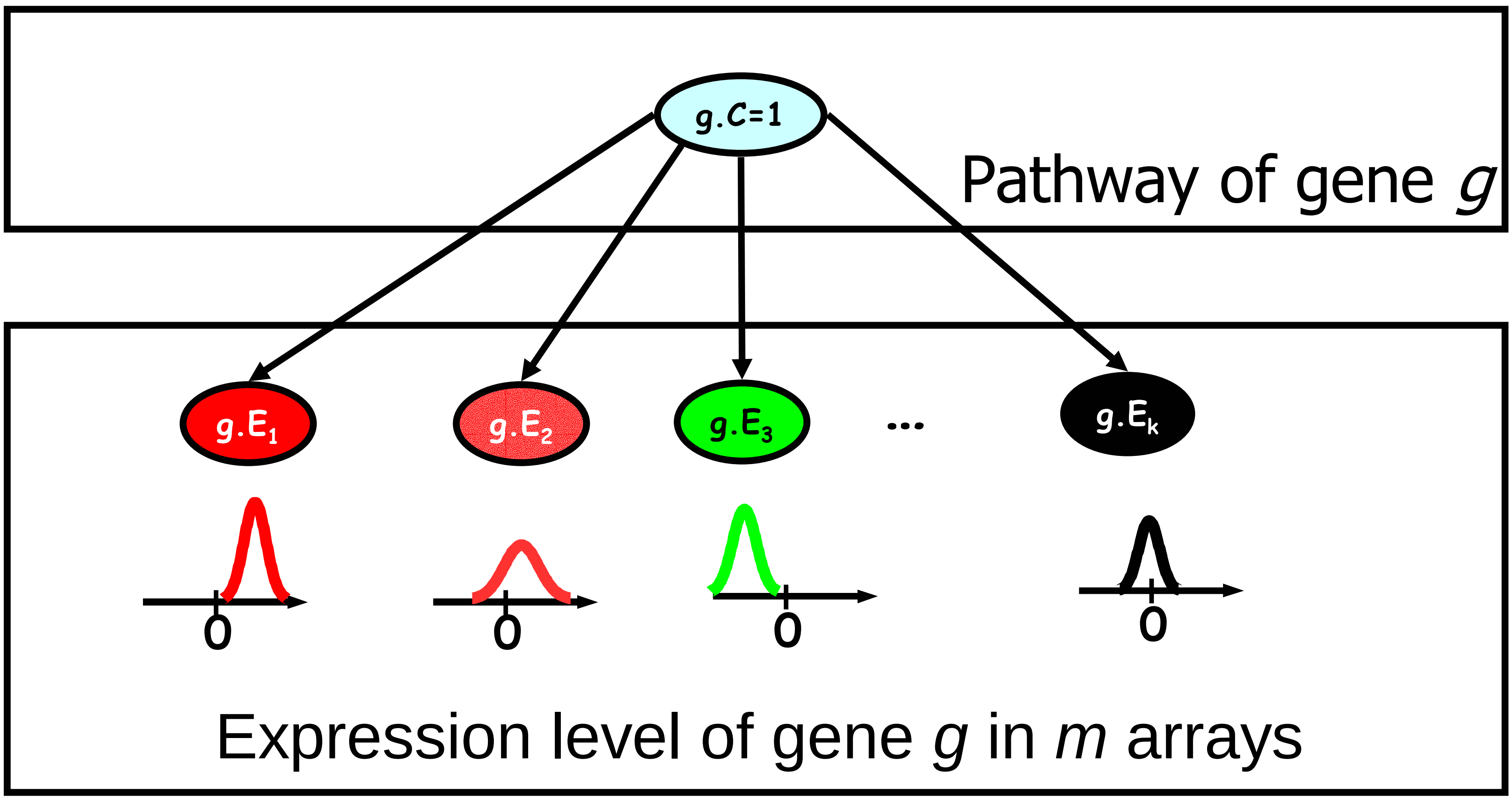
Probabilistic Model

- Genes are partitioned into "pathways":
  - Every gene is assigned to one of 'k' pathways
  - Random variable for each gene with domain {1,…,k}
- Expression component:
  - Model likelihood is higher when genes in the same pathway have similar expression profiles
- Interaction component:
  - Model likelihood is higher when genes in the same pathway interact

[Segal, Wang, K., 2003]

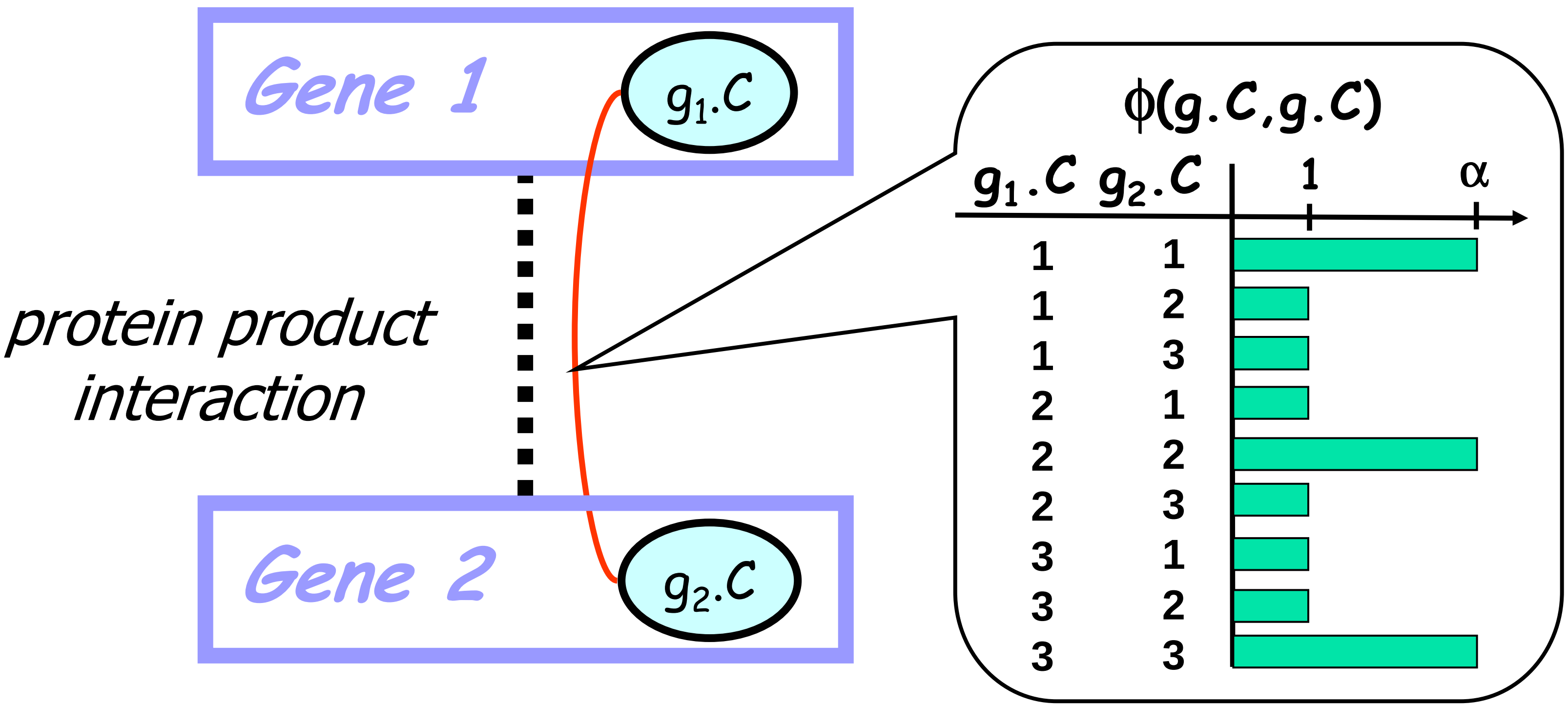
Expression Component

Naïve Bayes



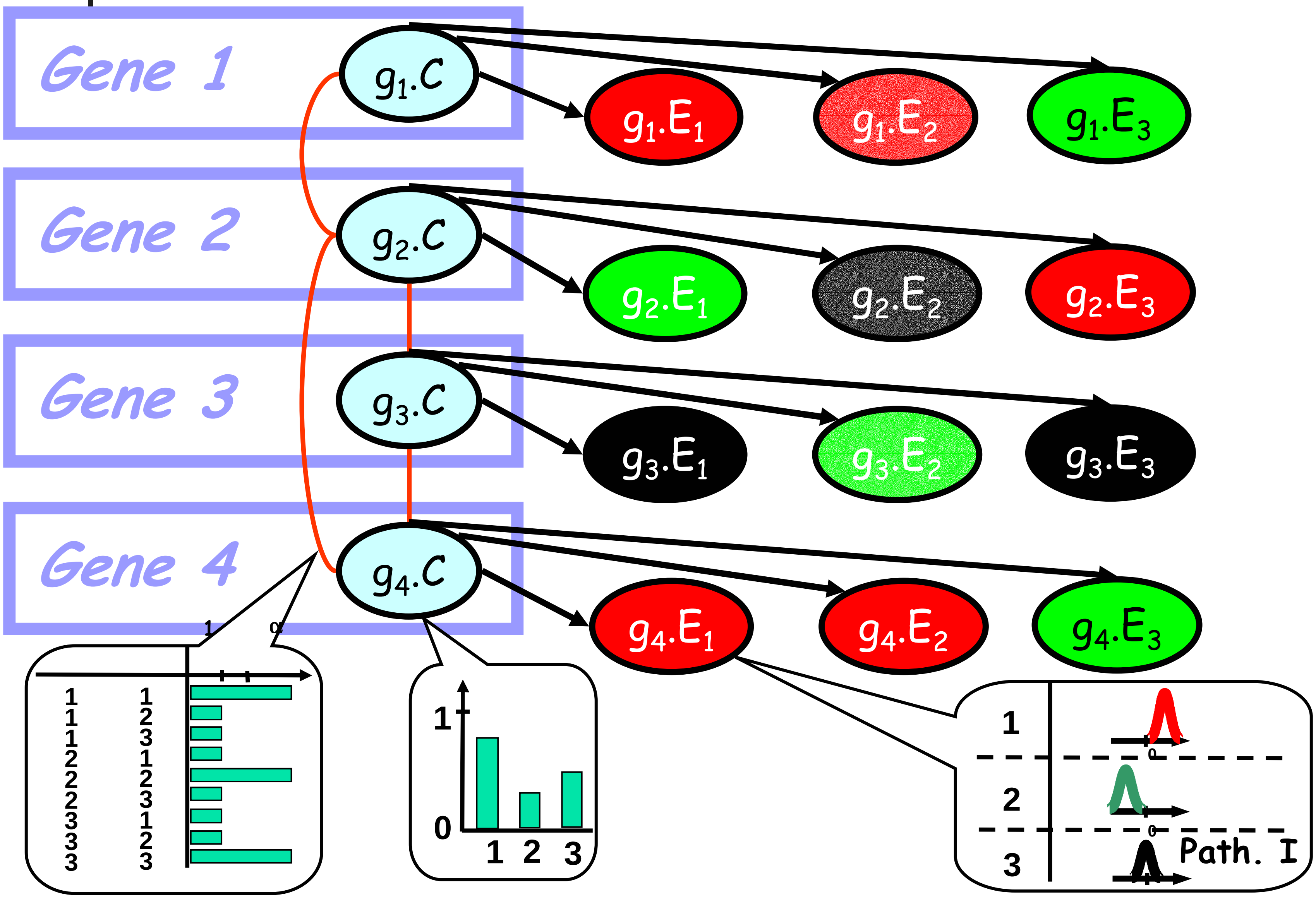
Protein Interaction Component

- Interacting genes are more likely to be in the same pathway

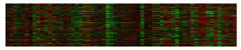


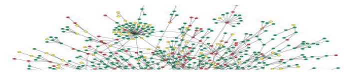
Compatibility potential

Joint Probabilistic Model



Learning Task

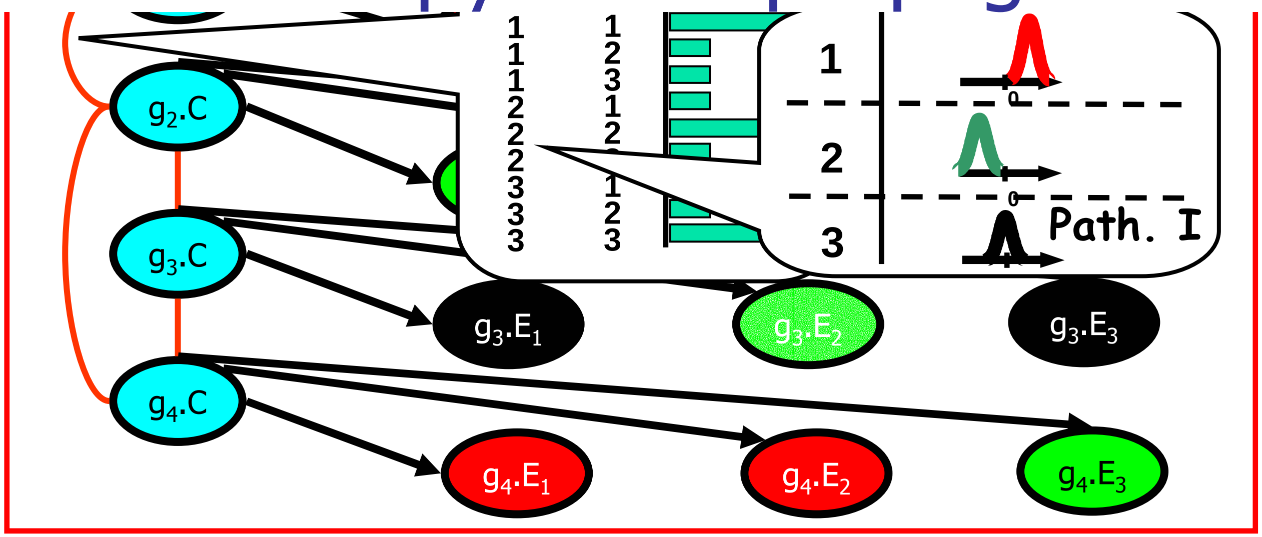




Large Markov network with high connectivity

$\Downarrow$

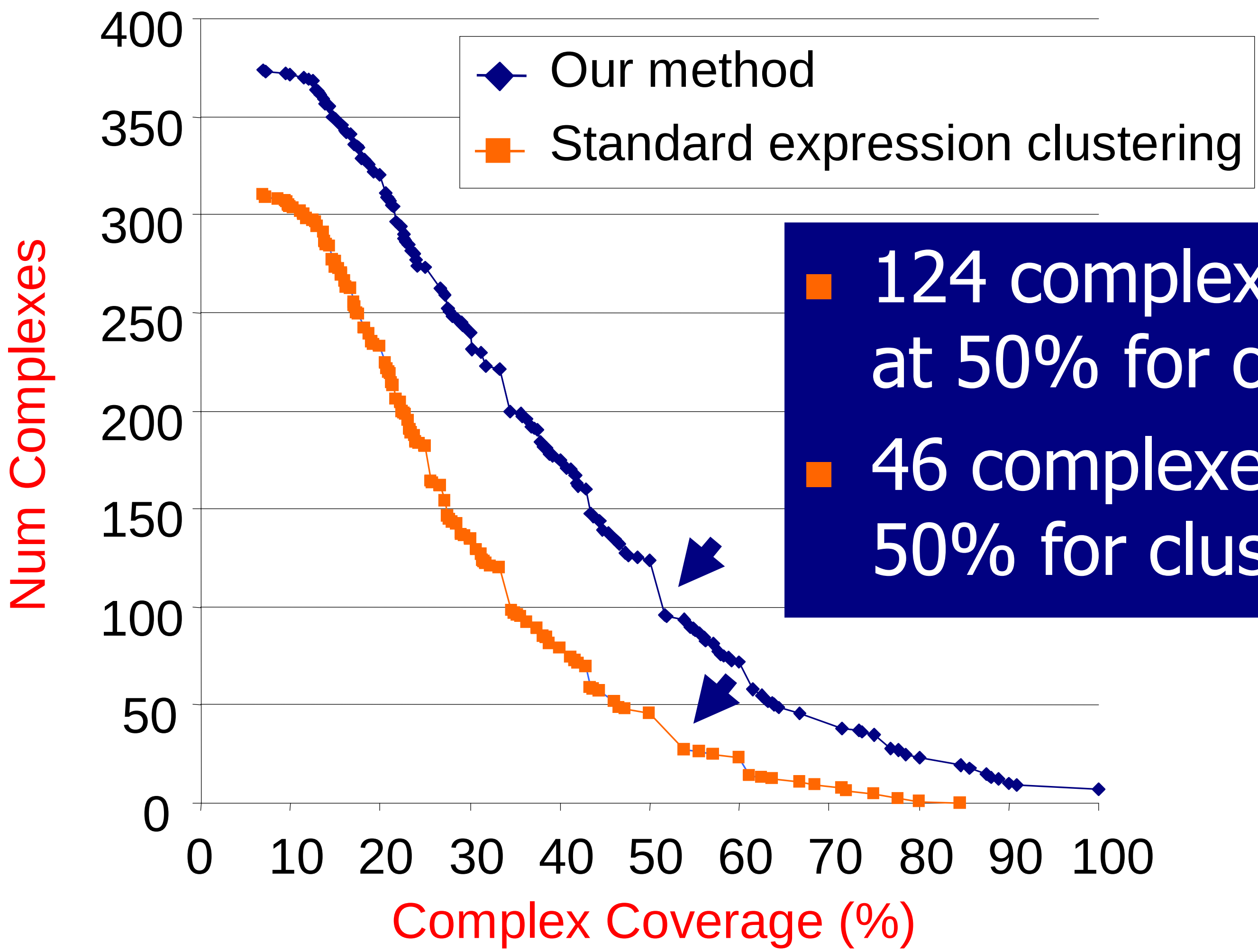
Use loopy belief propagation



- E-step: compute pathway assignments
- M-step: Estimate Gaussian distribution parameters
- Estimate compatibility potentials using cross-validation

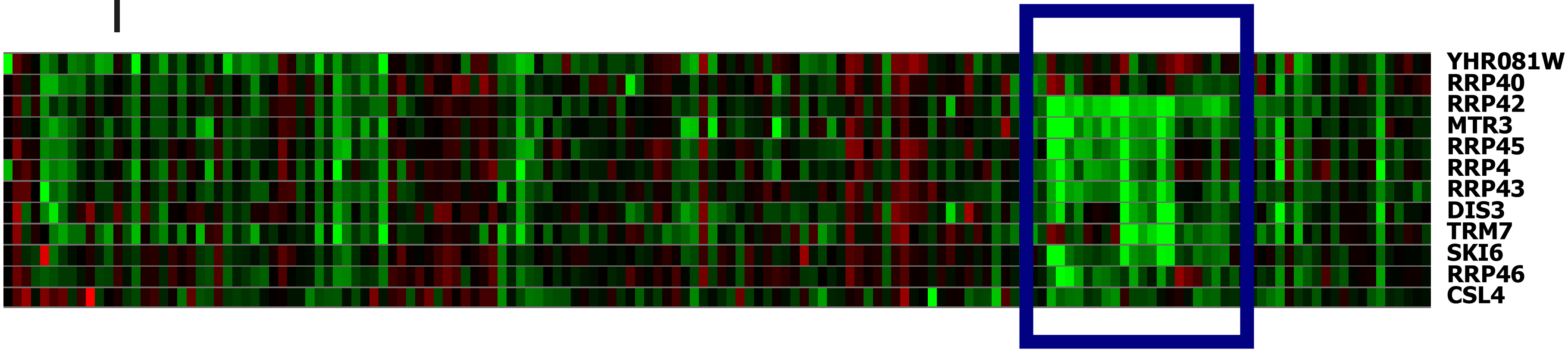
Capturing Protein Complexes

- Independent data set of interacting proteins

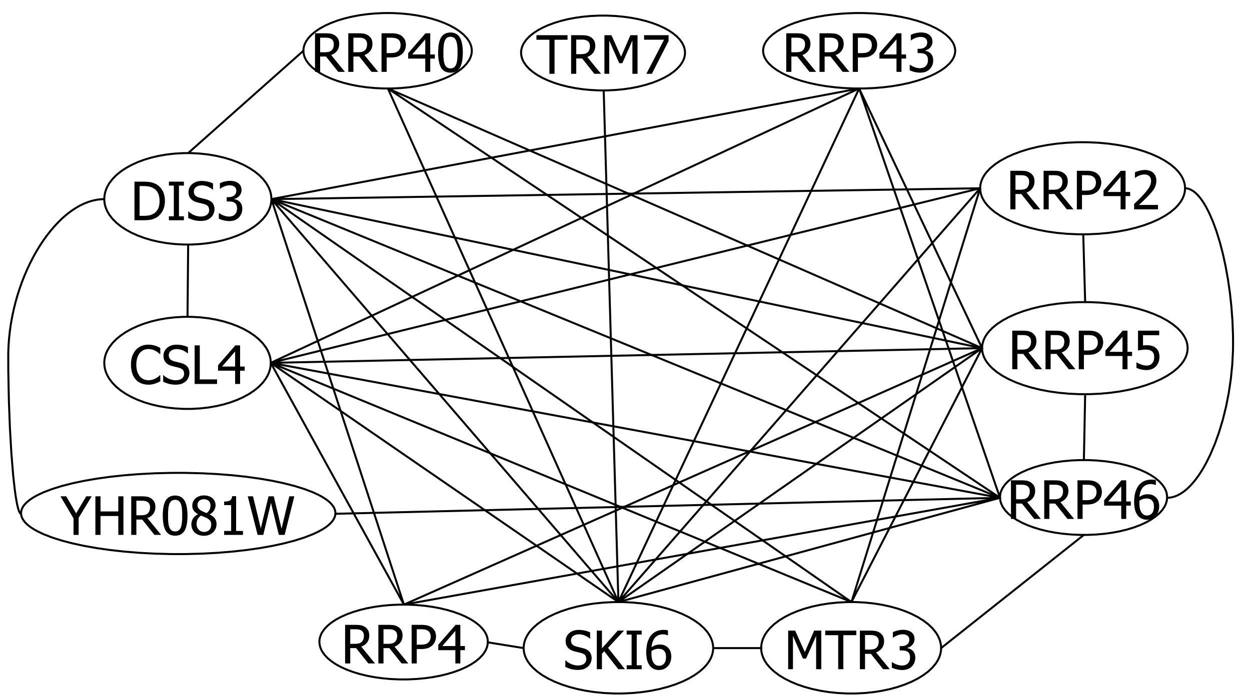


- 124 complexes covered at 50% for our method
- 46 complexes covered at 50% for clustering

RNAse Complex Pathway

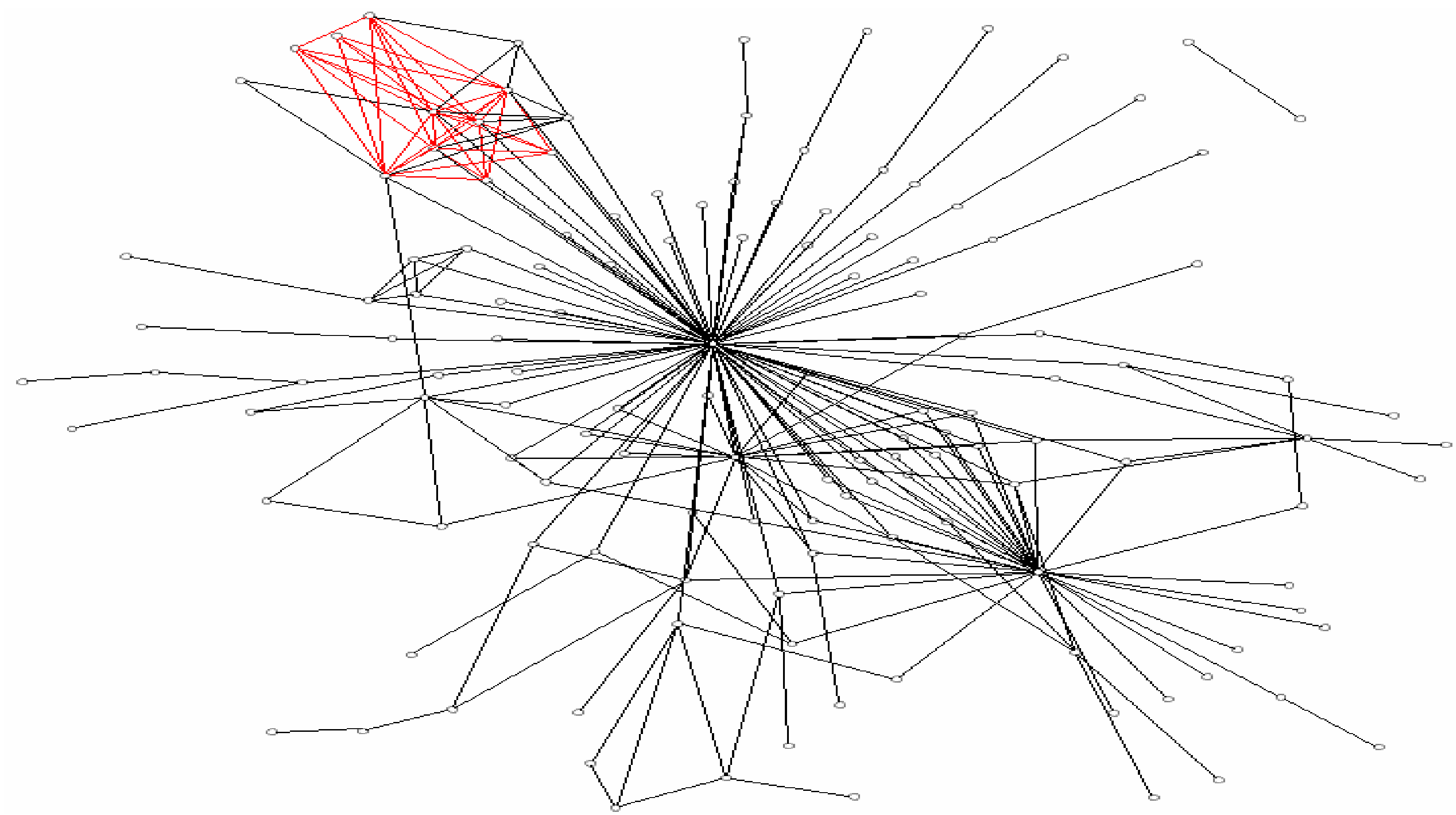


- Includes all 10 known pathway genes
- Only 5 genes found by clustering



Interaction Clustering

- RNAse complex found by interaction clustering as part of cluster with **138** genes



Uncertainty about Domain Structure

or

PRMs are not just template BNs/MNs

Structural Uncertainty

- Class uncertainty:
  - To which class does an object belong
- Relational uncertainty:
  - What is the relational (link) structure
- Identity uncertainty:
  - Which “names” refer to the same objects
  - Also covers data association

Relational Uncertainty

Probability distribution over graph structures

- Link existence model
  - E.g., hyperlinks between webpages
  - Each potential link is a separate event
  - Its existence is a random variable
- Link reference model
  - E.g., instructors teaching a course
  - Fix set of outgoing links per object
  - Distribution over # of endpoints for outgoing link
  - Each link has distribution over link endpoint
    - e.g., instructor link for CS course likely to point to CS prof
- Many other models possible

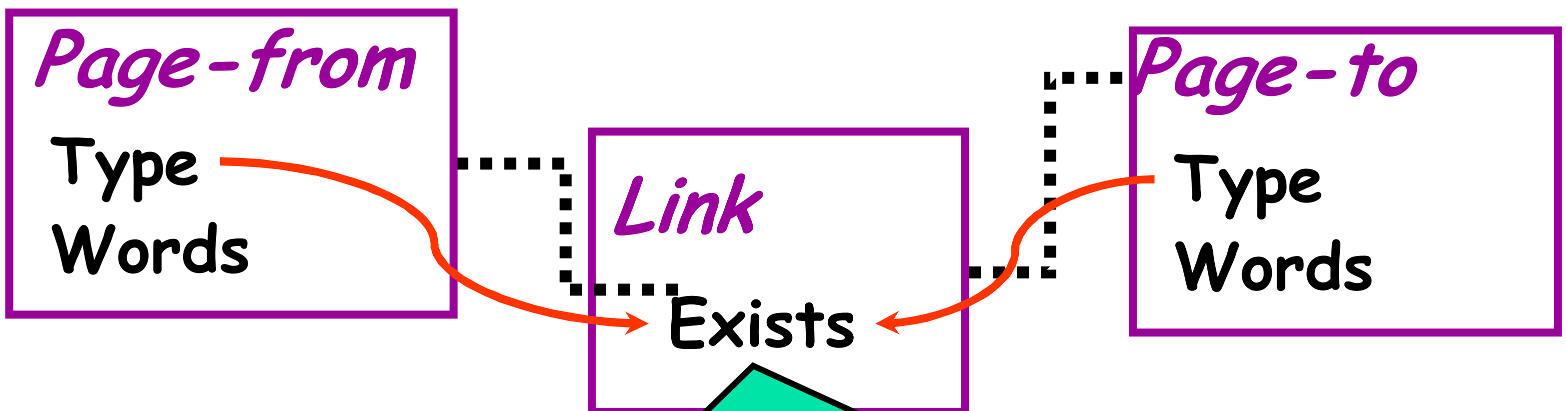
[Getoor, Friedman, K. Taskar, 2002]

Link Existence Model

- Background knowledge ξ is an *object skeleton*
  - A set of entity objects
- PRM defines distribution over worlds ω
  - Assignments of values to all attributes
  - Existence of links between objects
- Define objects for any potential links
  - E.g., a potential link object for any pair of webpages w1, w2
- Each potential link object has *link existence* attribute, denoting whether the link exists or not
- Link existence variables have probabilistic model

[Getoor, Friedman, K. Taskar, 2002]

Exists Uncertainty Example

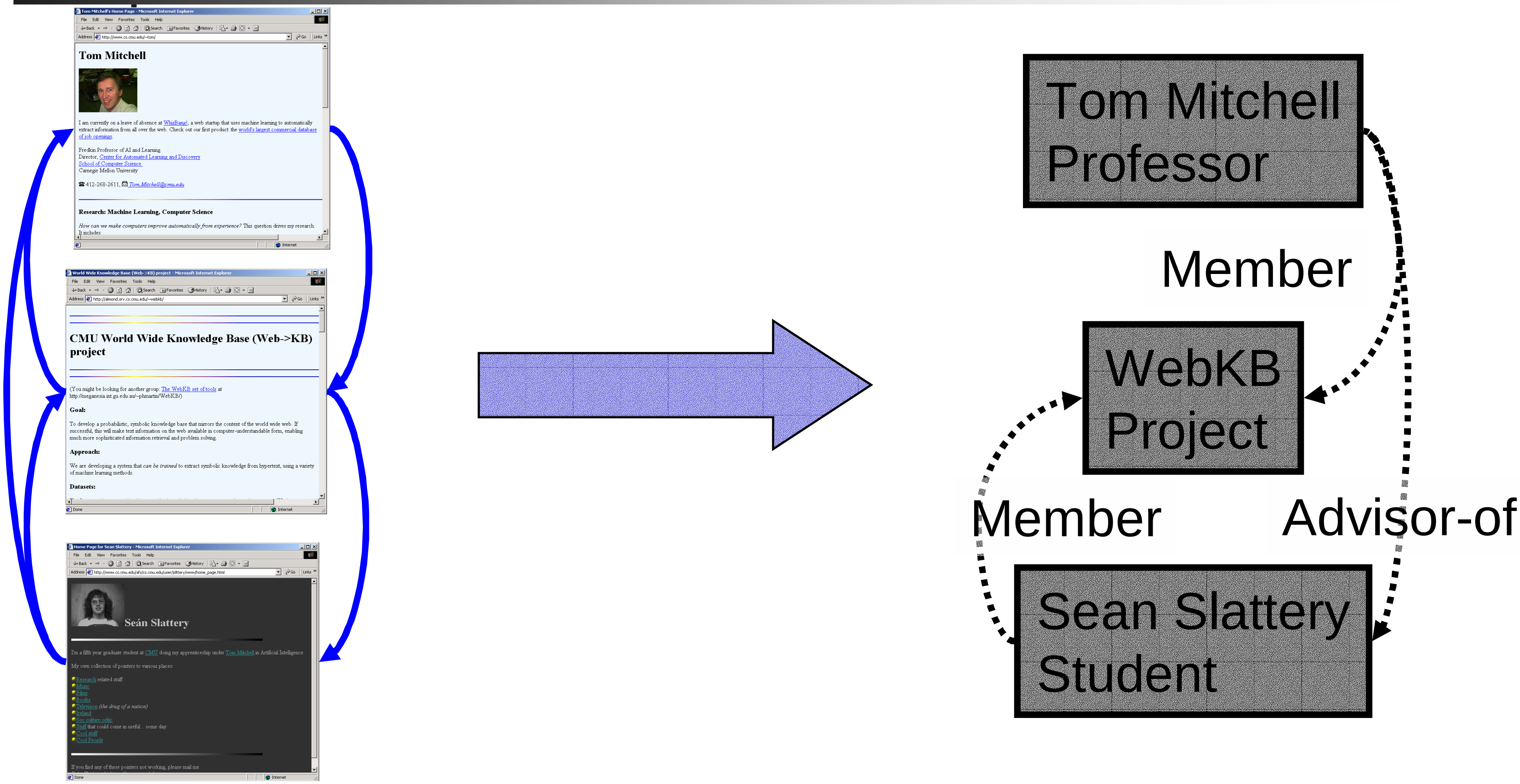


From.Type	To.Type	False	True
Student	Student	0.999	0001
Student	Professor	0.995	0005
Professor	Student	0.998	0002
Professor	Professor	0.999	0001

Why Are Link Models Useful?

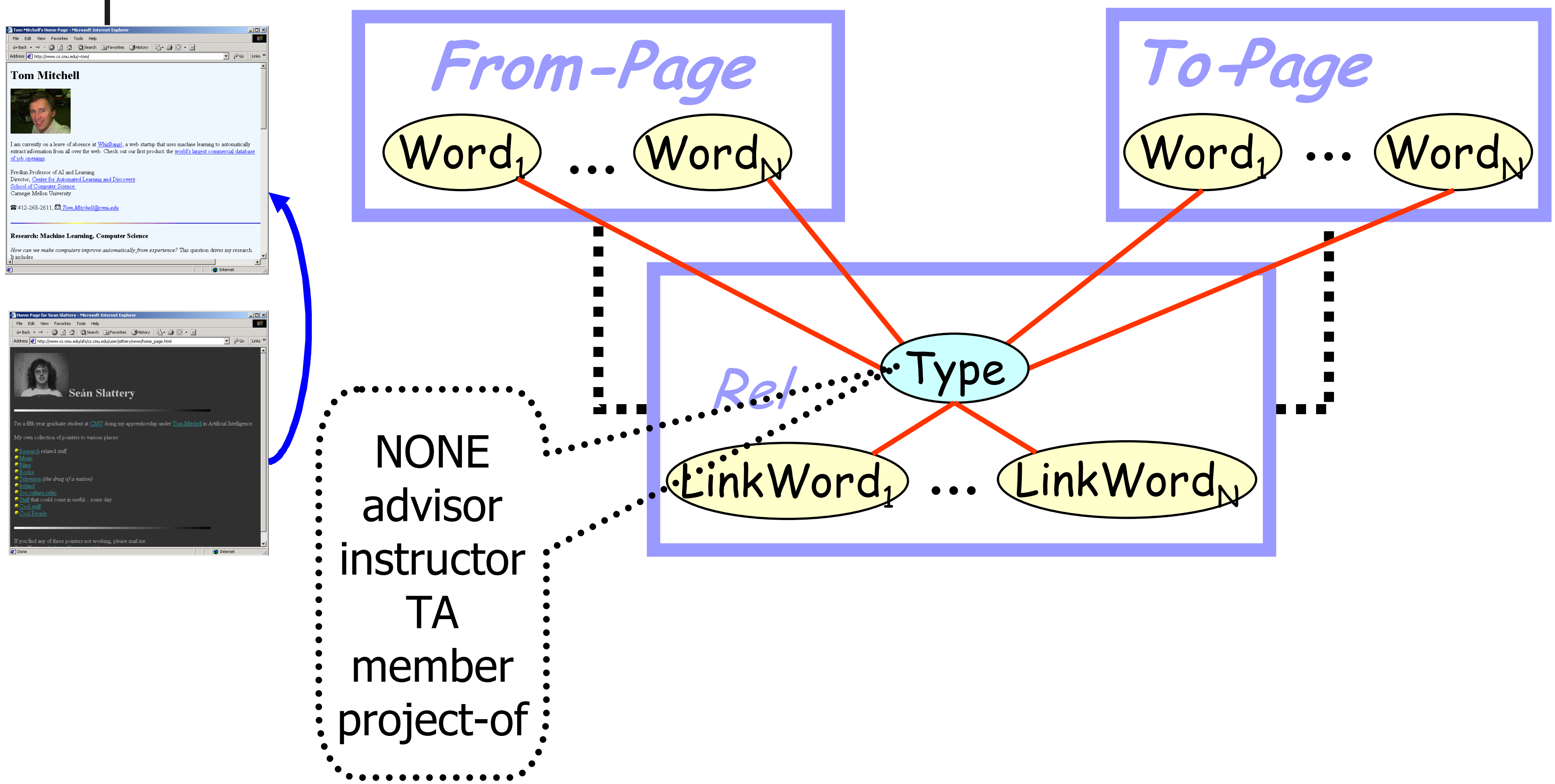
- Predict which links exist in the world
  - Which professor teaches each course
  - Which student will register for which course
- Use known links to infer values of attributes
  - Given that student registered for a hard class, is she more likely to be a graduate student
  - Given that one page points to another, is it more likely to be a faculty page?

Predicting Relationships

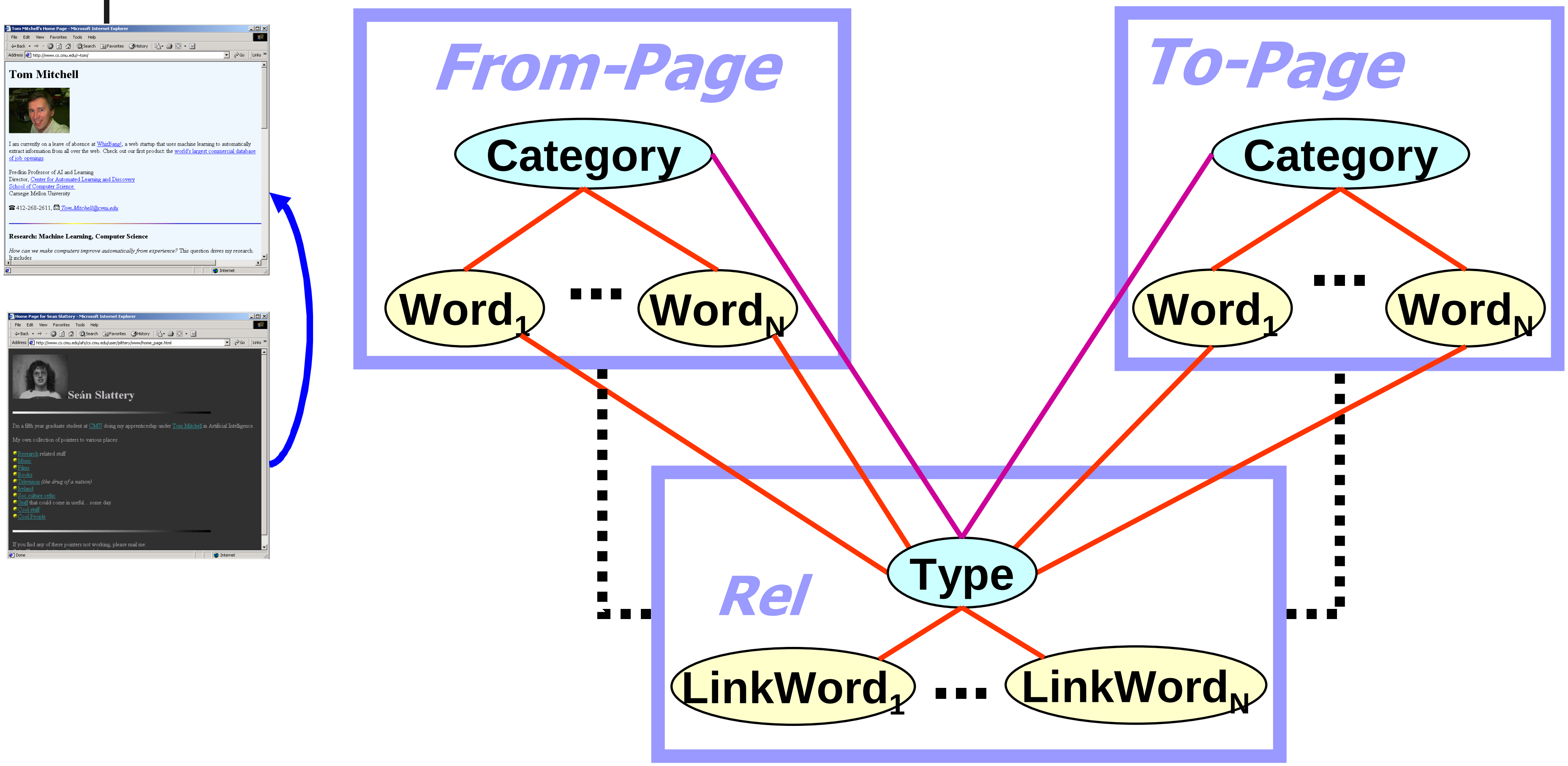


Predict and classify relationships between objects

Flat Model

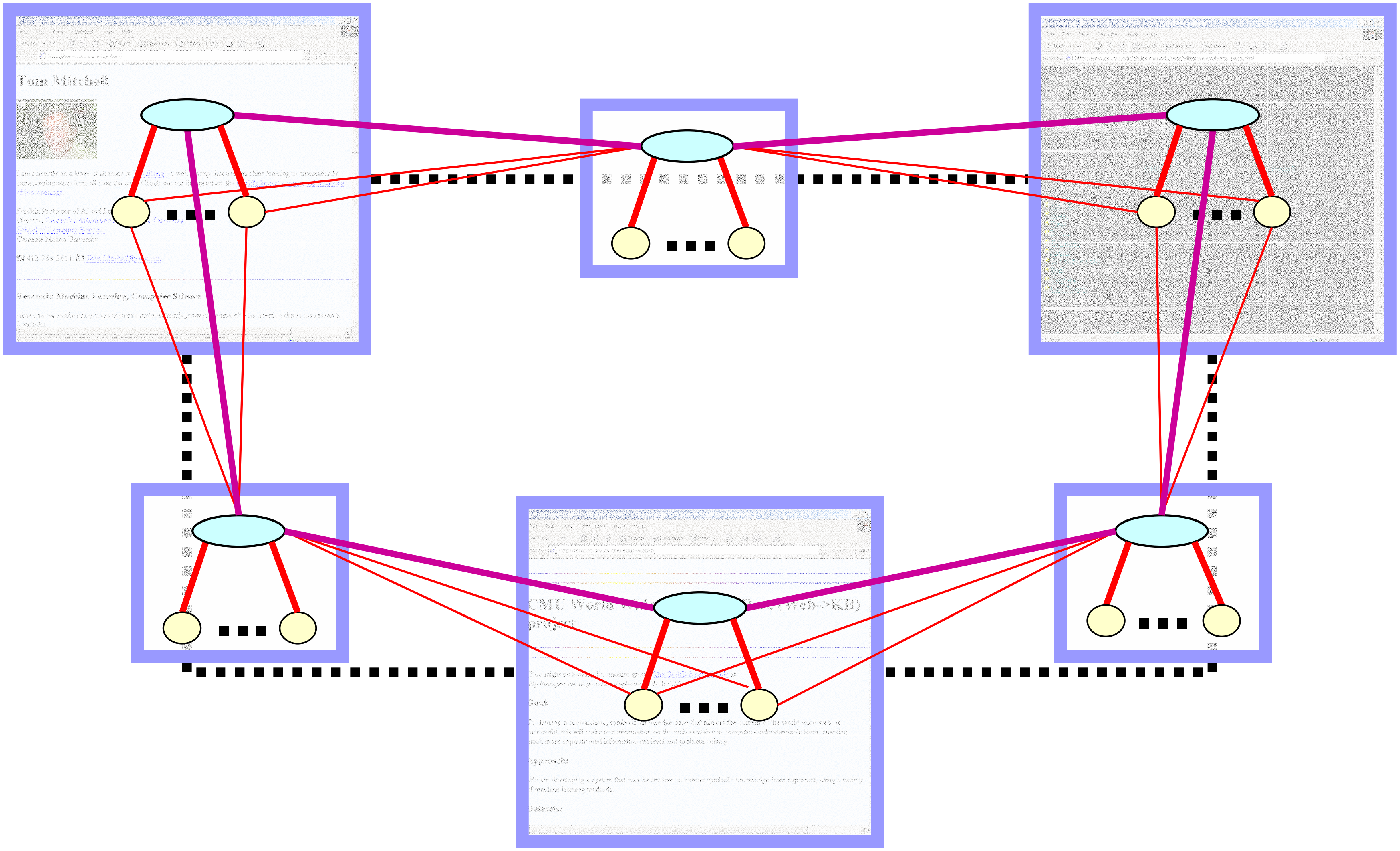

Tom Mitchell
Seán Slattery
From-Page
$Word_1$ … $Word_N$
To-Page
$Word_1$ … $Word_N$
Rel
Type
$LinkWord_1$ … $LinkWord_N$
NONE
advisor
instructor
TA
member
project-of

Collective Classification: Links

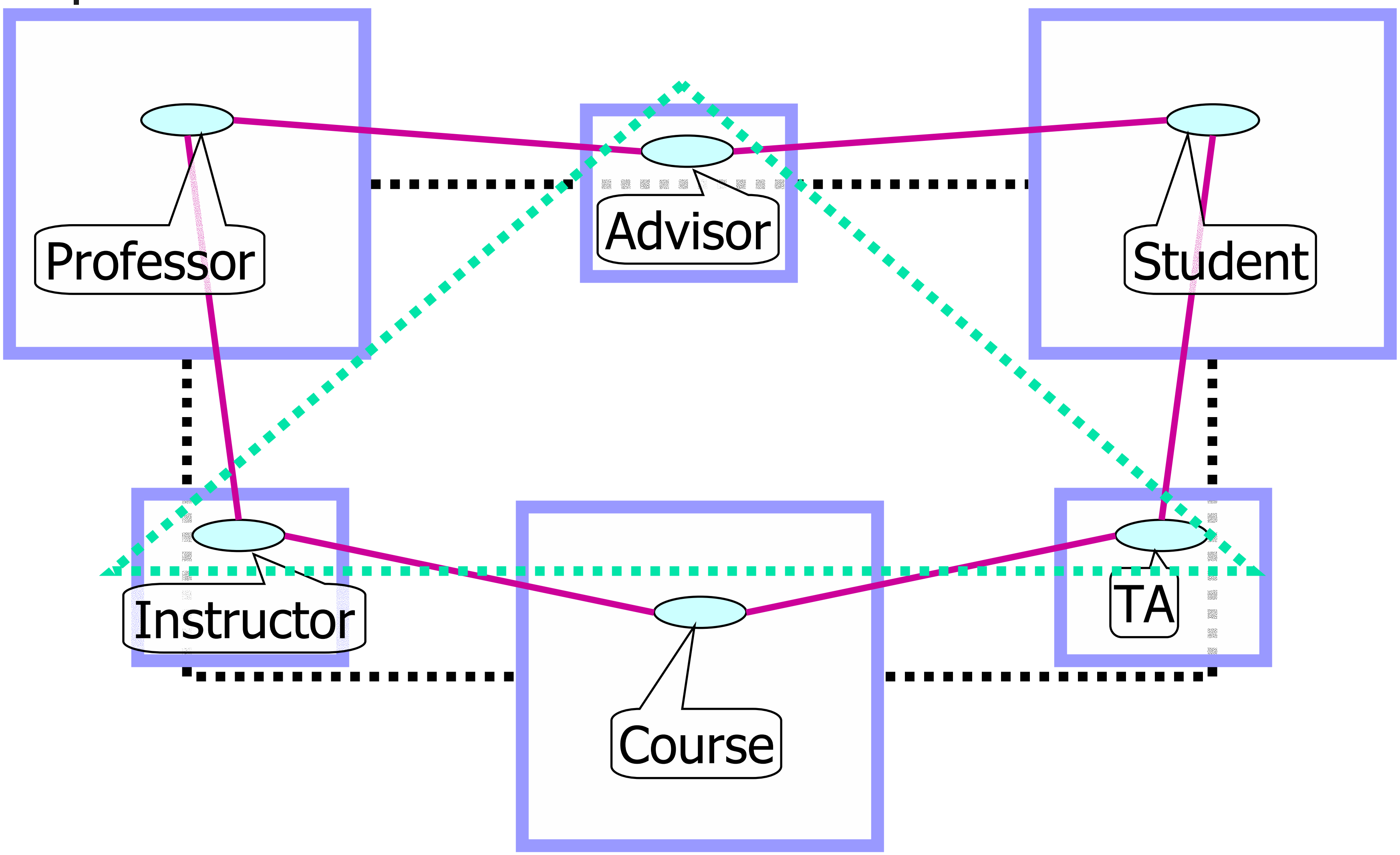


[Taskar, Wong, Abbeel, K., 2002]

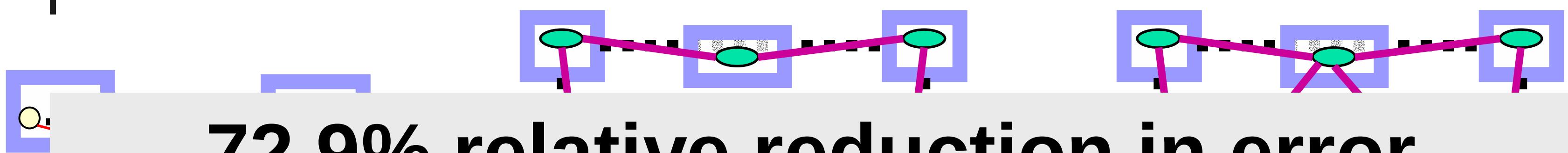
Link Model


Tom Mitchell
CMU World Wide Web Knowledge Base (Web->KB) project

Triad Model

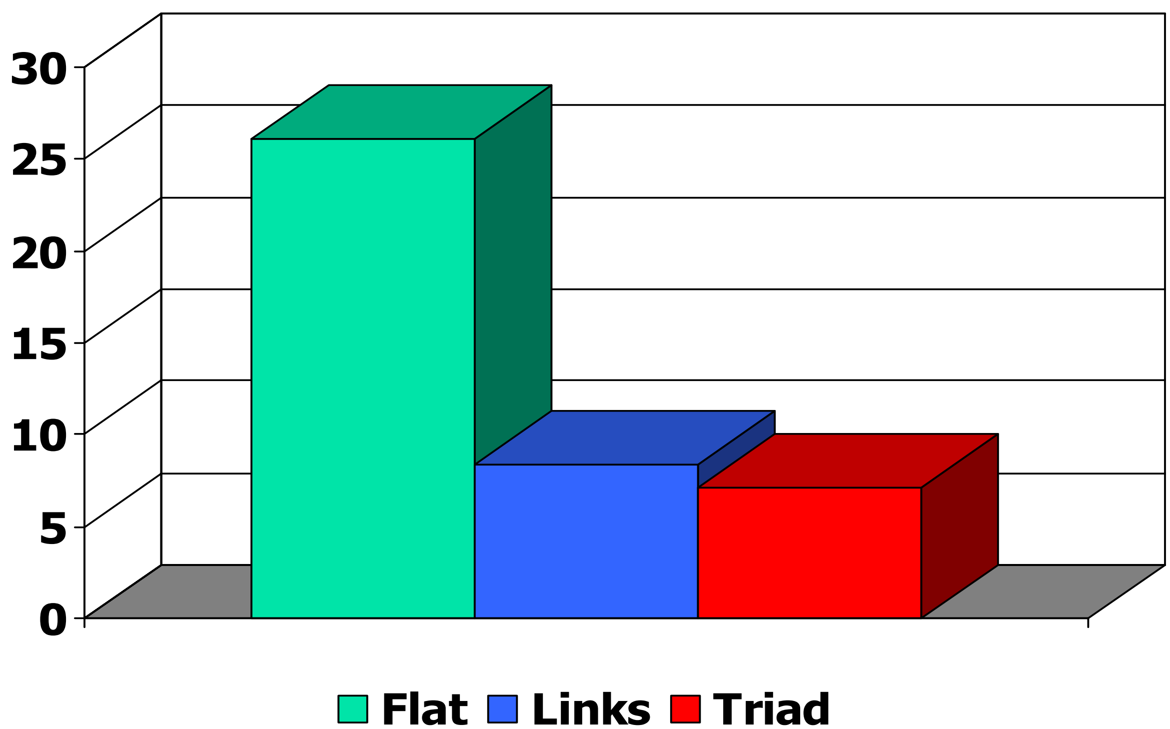


Link Prediction: Results



72.9% relative reduction in error relative to strong flat approach

- *Error measured over links predicted to be present*
- *Link presence cutoff is at precision/recall break-even point (≈30% for all models)*



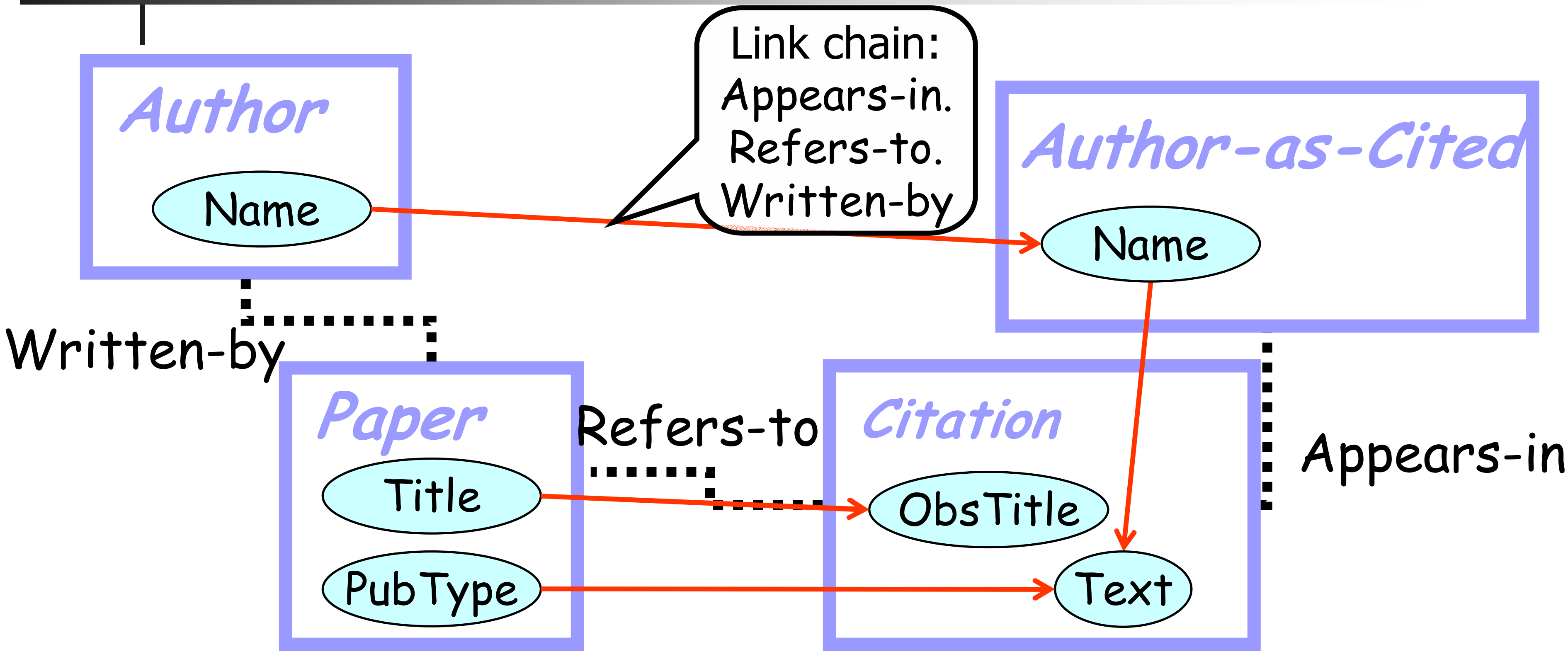
[Taskar, Wong, Abbeel, K., 2002]

Identity Uncertainty Model

- Background knowledge ξ is an *object universe*
  - A set of potential objects
- PRM defines distribution over worlds ω
  - Assignments of values to object attributes
  - Partition of objects into equivalence classes
  - Objects in same class have same attribute values

[Pasula, Marthi, Milch, Russell, Shpitser, 2002]

Citation Matching Model*



- Each citation object associated with paper object
- Uncertainty over equivalence classes for papers
- If $P_1=P_2$, have same attributes & links

Title, PubType

Authors

* Simplified

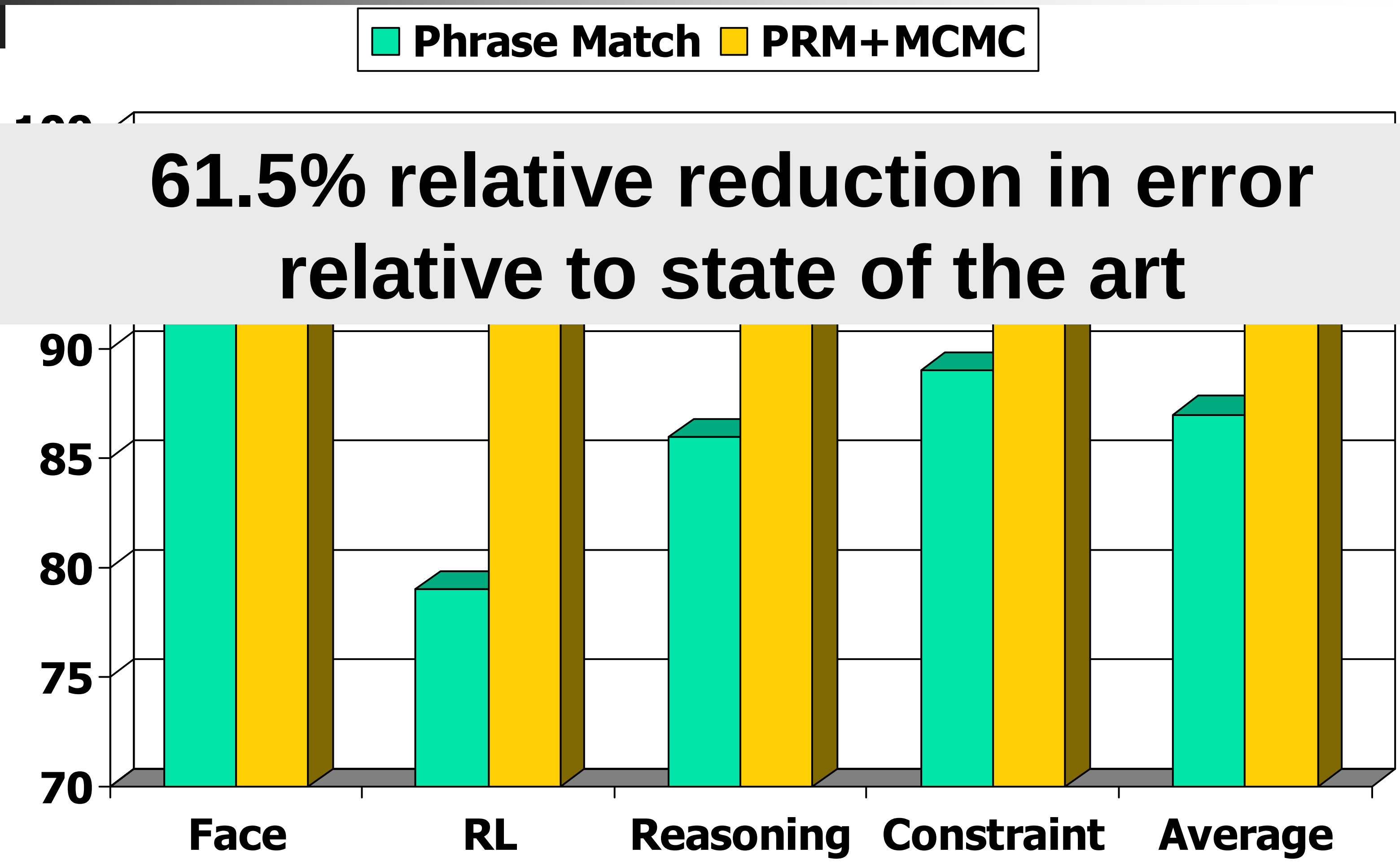
Identity Uncertainty

- Depending on choice of equivalence classes:
  - Number of objects changes
  - Dependency structure changes
- No “nice” corresponding ground BN

- Algorithm:
  - Each partition hypothesis defines simple BN
  - Use MCMC over equivalence class partition
  - Exact inference over resulting BN defines acceptance probability for Markov chain

[Pasula, Marthi, Milch, Russell, Shpitser, 2002]

Identity Uncertainty Results



61.5% relative reduction in error relative to state of the art

Accuracy of citation recovery:
% of actual citation clusters recovered perfectly

[Pasula, Marthi, Milch, Russell, Shpitser, 2002]

Summary: PRMs …

- Inherit the advantages of graphical models:
  - Coherent probabilistic semantics
  - Exploit structure of local interactions

- Allow us to represent the world in terms of:
  - Objects
  - Classes of objects
  - Properties of objects
  - Relations

So What Do We Gain?

- Convenient language for specifying complex models
- “Web of influence”: subtle & intuitive reasoning
- A mechanism for tying parameters and structure
  - within models
  - across models
- Framework for learning from relational and heterogeneous data

So What Do We Gain?

New way of thinking about models & problems

- Incorporating heterogeneous data by connecting related entities

- New problems:
  - Collective classification
  - Relational clustering

- Uncertainty about richer structures:
  - Link graph structure
  - Identity

But What Do We *Really* Gain?

Are PRMs just a convenient language for specifying attribute-based graphical models?

- Simple PRMs ≈ relational logic w. fixed domain and ∀ only
  - Induce a "propositional" BN
- Can augment language with additional expressivity
  - Existential quantifiers & functions
  - Equality
- Resulting language is inherently more expressive, allowing us to represent distributions over
  - worlds where dependencies vary significantly [Getoor et al., Pasula et al.]
  - worlds with different numbers of objects [Pfeffer et al., Pasula et al.]
  - worlds with infinitely many objects [Pfeffer & K.]
- **Big questions: Inference & Learning**

Thank You!

http://robotics.stanford.edu/~koller/