Uncertainty in Ontology Mapping:
A Bayesian Perspective

Yun Peng, Zhongli Ding, Rong Pan

Department of Computer Science and Electrical engineering
University of Maryland Baltimore County
ypeng@umbc.edu

Outline

- Motivations
  - Uncertainty in ontology representation, reasoning and mapping
  - Why Bayesian networks (BN)
- Overview of the approach
- Translating OWL ontology to BN
  - Representing probabilistic information in ontology
  - Structural translation
  - Constructing conditional probability tables (CPT)
- Ontology mapping
  - Formalizing the notion of "mapping"
  - Mapping reduction
  - Mapping as evidential reasoning
- Conclusions

Motivations

- Uncertainty in ontology engineering
  - In representing/modeling the domain
    - Besides ***A*** *subclasOf* ***B***, also ***A*** is a *small* subset of ***B***
    - Besides ***A*** *hasProperty* ***P***, also *most* objects with ***P*** are in ***A***
    - ***A*** and ***B*** overlap, but none is a subclass of the other
  - In reasoning
    - How close a description ***D*** is to its most specific subsumer and most general subsumee?
    - Noisy data: leads to over generalization in subsumptions
    - Uncertain input: the object is *very likely* an instance of class ***A***

Motivations

– In mapping concepts from one ontology to another
  - Similarity between concepts in two ontologies often cannot be adequately represented by logical relations
    – Overlap rather than inclusion
  - Mappings are hardly 1-to-1
    – If ***A*** in onto1 is similar to ***B*** in onto2, ***A*** would also be similar to the sub and super classes of ***B*** (with different degree of similarity)

- Uncertainty becomes more prevalent in web environment
  – One ontology may import other ontologies
  – Competing ontologies for the same or overlapped domain

Bayesian Networks

- Why Bayesian networks (BN)
  - Existing approaches
    - Logic based approaches are inadequate
    - Others often based on heuristic rules
    - Uncertainty is resolved during mapping, and not considered in subsequent reasoning
      - Loss of information
  - BN is a graphic model of dependencies among variables:
    - Structural similarity with OWL graph
    - BN semantics is compatible with that of OWL
    - Rich set of efficient algorithms for reasoning and learning

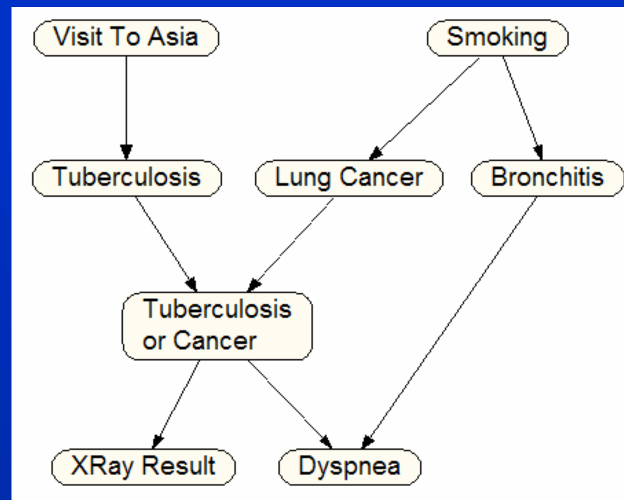
Bayesian Networks

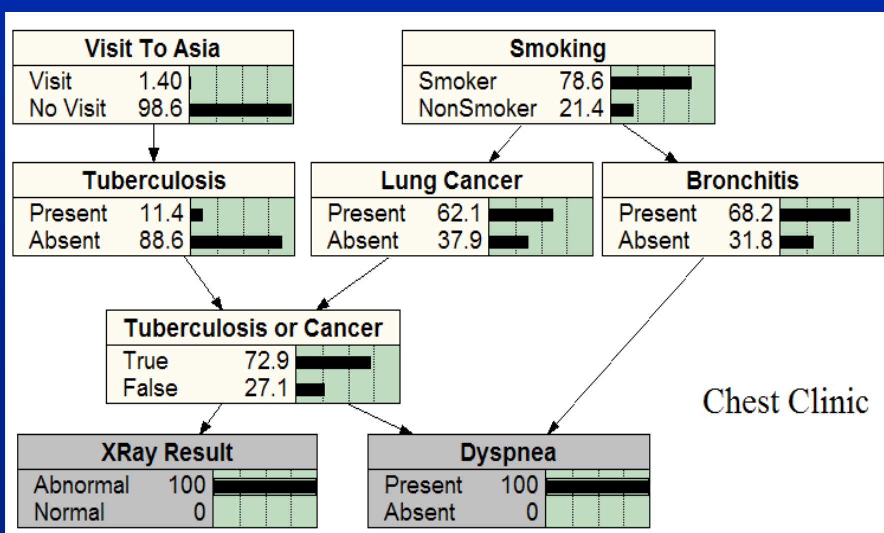
- Directed acyclic graph (DAG)
  - Nodes: (discrete) random variables
  - Arcs: causal/influential relations
  - A variable is independent of all other non-descendent variables, given its parents
- Conditional prob. tables (CPT)
  - To each node: $P(x_i \mid \pi_i)$ where π_i is the parent set of x_i
- Chain rule:
  - $P(x_1,\ldots x_n) = \Pi_i P(x_i \mid \pi_i)$
  - Joint probability as product of CPT

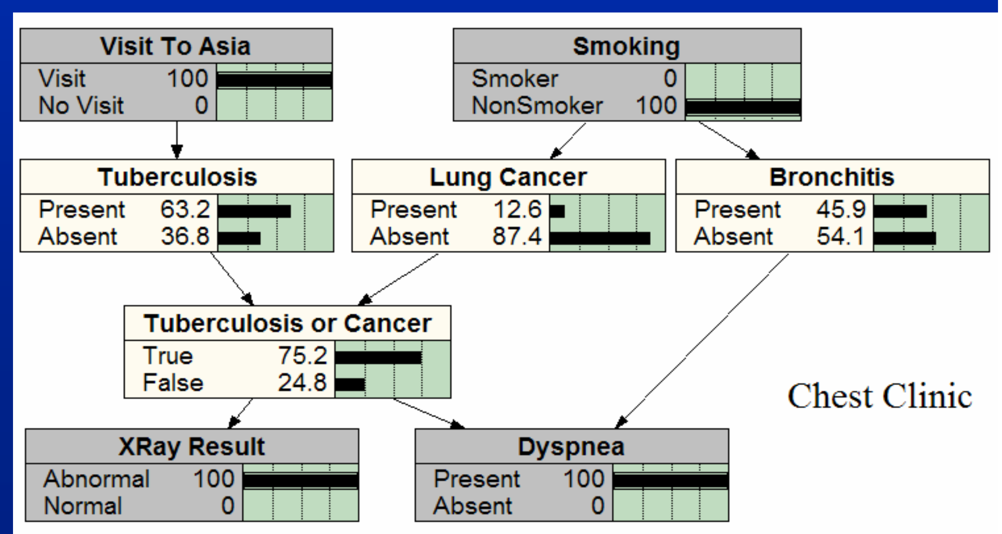
Bayesian Networks

Visit	No_Visit
1.000	99.000

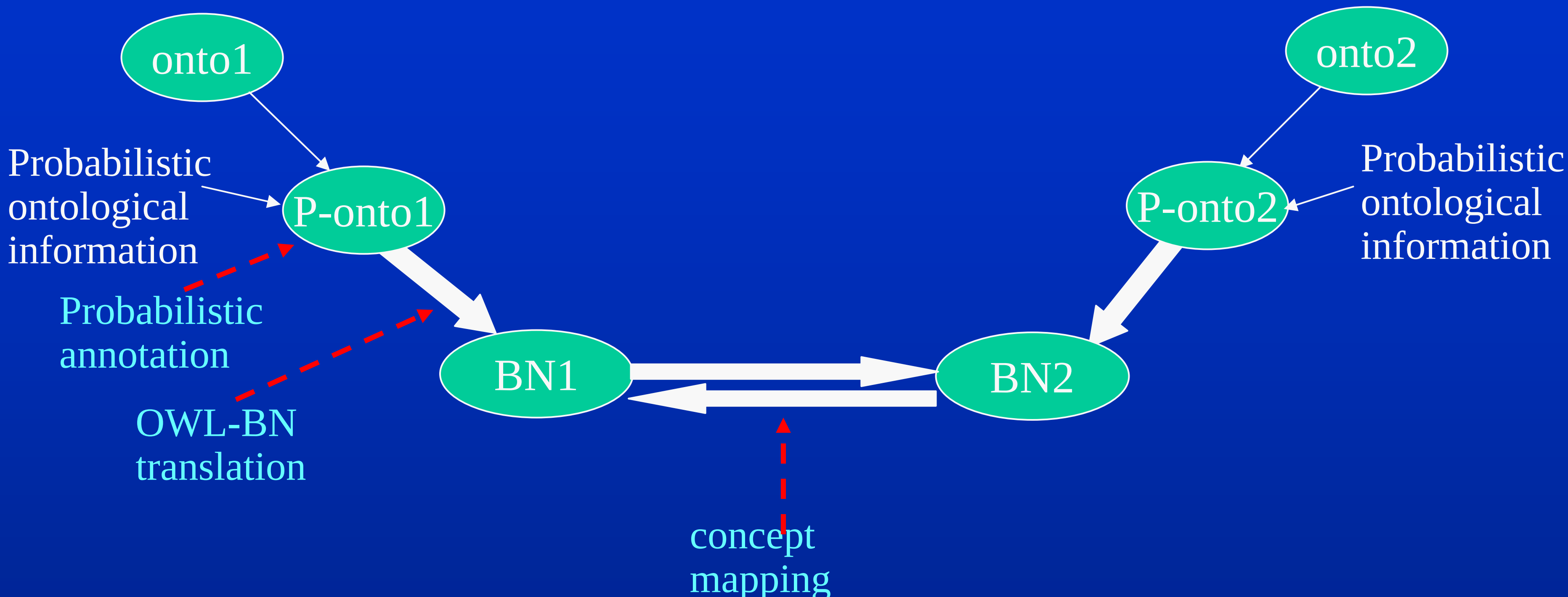
Tuberculosis	Cancer	Tb...
Present	Present	True
Present	Absent	True
Absent	Present	True
Absent	Absent	False







Overview of The Approach



- OWL-BN translation
  - By a set of translation rules and procedures
  - Maintain OWL semantics
  - Ontology reasoning by probabilistic inference in BN
- Ontology mapping
  - A parsimonious set of links
  - Capture similarity between concepts by joint distribution
  - Mapping as evidential reasoning

OWL-BN Translation

- Encoding probabilities in OWL ontologies
  - Not supported by current OWL
  - Define new classes for prior and conditional probabilities
- Structural translation: a set of rules
  - Class hierarchy: set theoretic approach
  - Logical relations (equivalence, disjoint, union, intersection...)
  - Properties
- Constructing CPT for each node:
  - Iterative Proportional Fitting Procedure (IPFP)
- Translated BN will preserve
  - Semantics of the original ontology
  - Encoded probability distributions among relevant variables

Encoding Probabilities

- Allow user to specify prior and conditional Probabilities.
  - Two new OWL classes: “PriorProbObj” and “CondProbObj”
  - A probability is defined as an instance of one of these classes.
- P(A): e.g., P(Animal) = 0.5

```
<prob:PriorProbObj rdf:ID="P(Animal)">
  <prob:hasVariable><rdf:value>&ont;Animal</rdf:value></prob:hasVariable>
  <prob:hasProbValue>0.5</prob:hasProbValue>
</prob:PriorProbObj>
```

- P(A|B): e.g., P(Male|Animal) = 0.48

```
<prob:CondProbObjT rdf:ID="P(Male|Animal)">
  <prob:hasCondition><rdf:value>&ont;Animal</rdf:value></prob:hasCondition>
  <prob:hasVariable><rdf:value>&ont;Male</rdf:value></prob:hasVariable>
  <prob:hasProbValue>0.5</prob:hasProbValue>
</prob:CondProbObjT>
```

Structural Translation

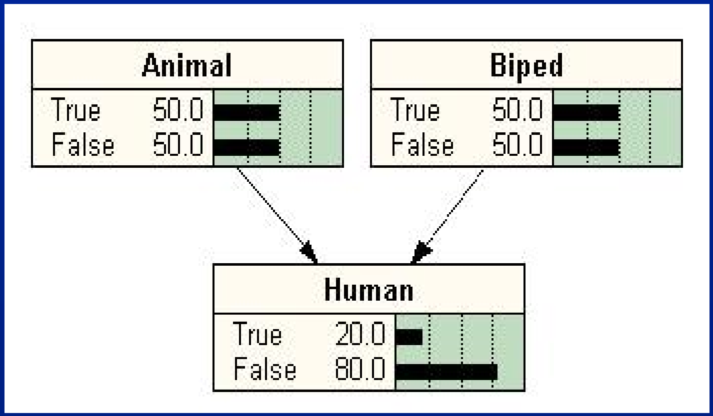
- Set theoretic approach
  - Each OWL class is considered a **set** of objects/instances
  - Each class is defined as a node in BN
  - An arc in BN goes from a superset to a subset
  - Consistent with OWL semantics

```
<owl:Class rdf:ID="Human">
  <rdfs:subclassOf rdf:resource="#Animal">
  <rdfs:subclassOf rdf:resource="#Biped">
 </owl:Class>
```

RDF Triples:

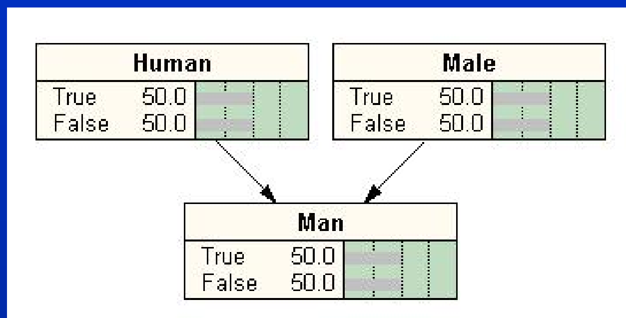
(Human rdf:type owl:Class)
(Human rdfs:subClassOf Animal)
(Human rdfs:subClassOf Biped)

Translated to BN



Structural Translation

- Logical relations
  - Some can be encoded by CPT (e.g.. Man = Human ∩ Male)

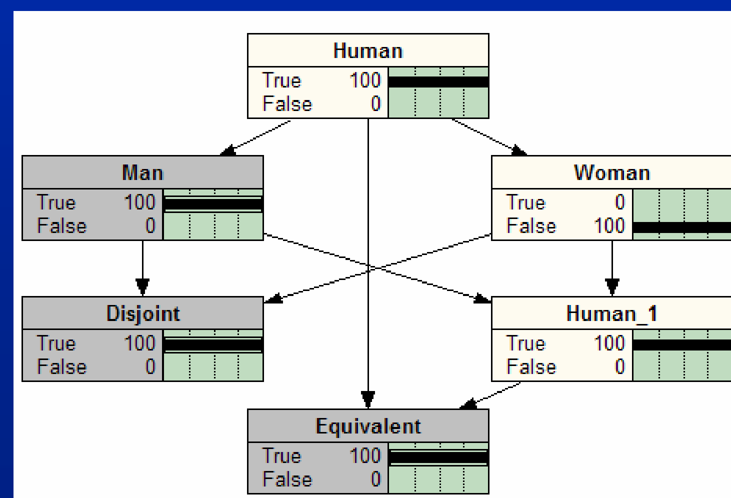


Human	Male	True	False
True	True	100.00	0.000
True	False	0.000	100.00
False	True	0.000	100.00
False	False	0.000	100.00

  - Others can be realized by adding control nodes

    Man ⊂ Human
    Woman ⊂ Human
    Human = Man ∪ Woman
    Man ∩ Woman = ∅

auxiliary node: Human_1
Control nodes: Disjoint, Equivalent

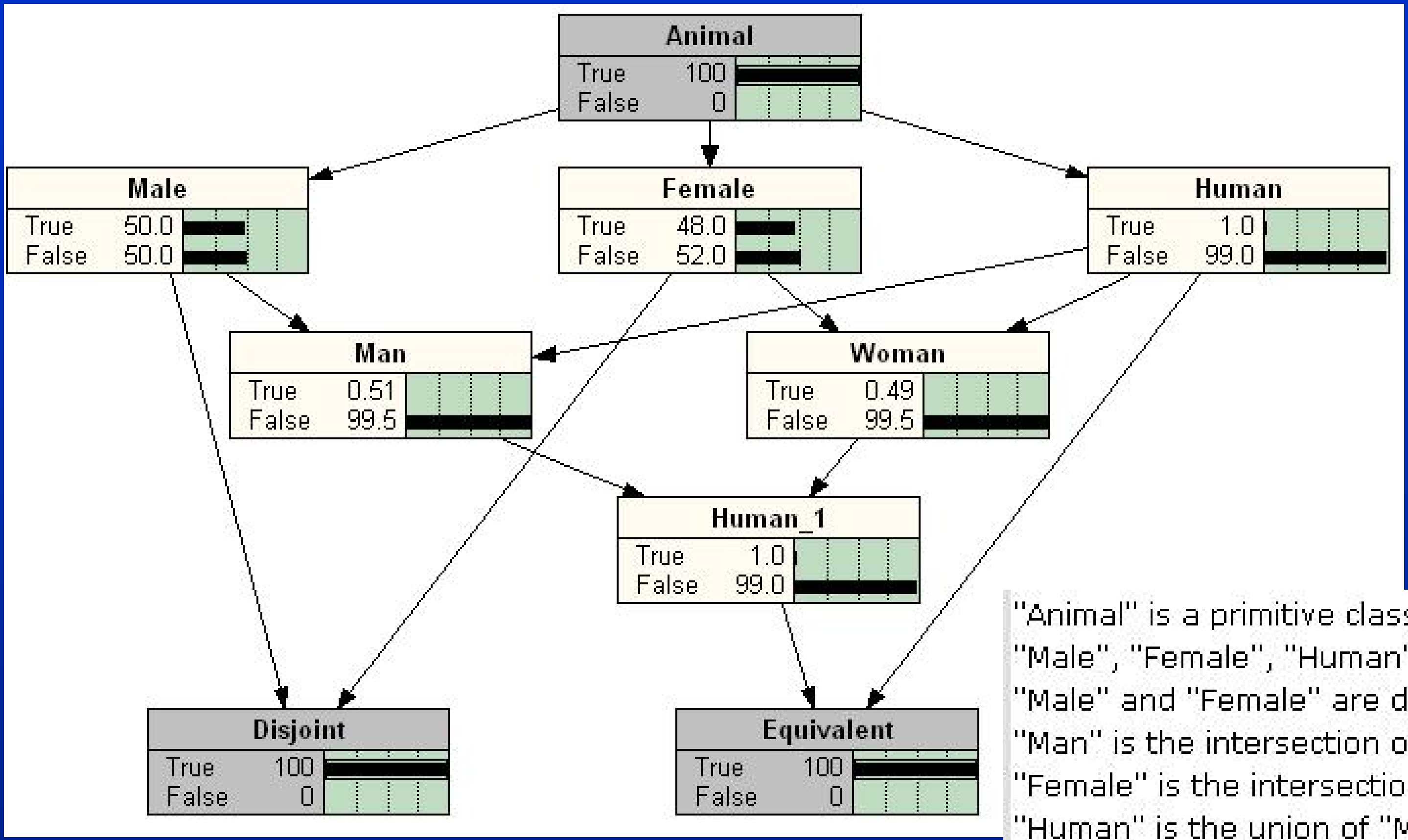


Constructing CPT

- Imported Probability information is not in the form of CPT
- Assign initial CPT to the translated structure by some default rules
- Iteratively modify CPT to fit imported probabilities while setting control nodes to *true*.
  - IPFP (Iterative Proportional Fitting Procedure)

    To find $Q(x)$ that fit $Q(E_1)$, … $Q(E_k)$ to the given $P(x)$
    - $Q_0(x) = P(x)$; then repeat $Q_i(x) = Q_{i-1}(x)\ Q(E_j)/\ Q_{i-1}(E_j)$ until converging
    - $Q_\infty(x)$ is an I-projection of $P(x)$ on $Q(E_1)$, … $Q(E_k)$ (minimizing Kullback-Leibler distance to P)
  - Modified IPFP for BN

Example



"Animal" is a primitive class
"Male", "Female", "Human" are subclasses of "Animal"
"Male" and "Female" are disjoint with each other
"Man" is the intersection of "Male" and "Human"
"Female" is the intersection of "Female" and "Human"
"Human" is the union of "Man" and "Woman"

Probability information:

- P(Animal) = 0.5
- P(Male|Animal) = 0.5
- P(Female|Animal) = 0.48
- P(Human|Animal) = 0.01

Ontology Mapping

- Formalize the notion of *mapping*
- Mapping involving multiple concepts
- Reasoning under ontology mapping
- Assumption: ontologies have been translated to BN

Formalize The Notion of Mapping

- Simplest case: Map concept E^1 in Onto^1 to E^2 in Onto^2
  - How similar between E^1 and E^2
  - How to impose belief (distribution) of E^1 to Onto^2
- Cannot do it by simple Bayesian conditioning

  $$P(\text{x}|\, E^1) = \sum_{E^2} P(\text{x}|\, E^2) P(E^2 \,|\, E^1)\ \text{similarity}(E^1, E^2)$$

  - Onto^1 and Onto^2 have different probability space (Q and P)
    - $Q(E^1) \neq P(E^1)$
    - New distribution, given E^1 in Onto^1: $P^*(x) \neq \sum P(x|E^1)P(E^1)$
  - similarity(E^1, E^2) also needs to be formalized

UMBC

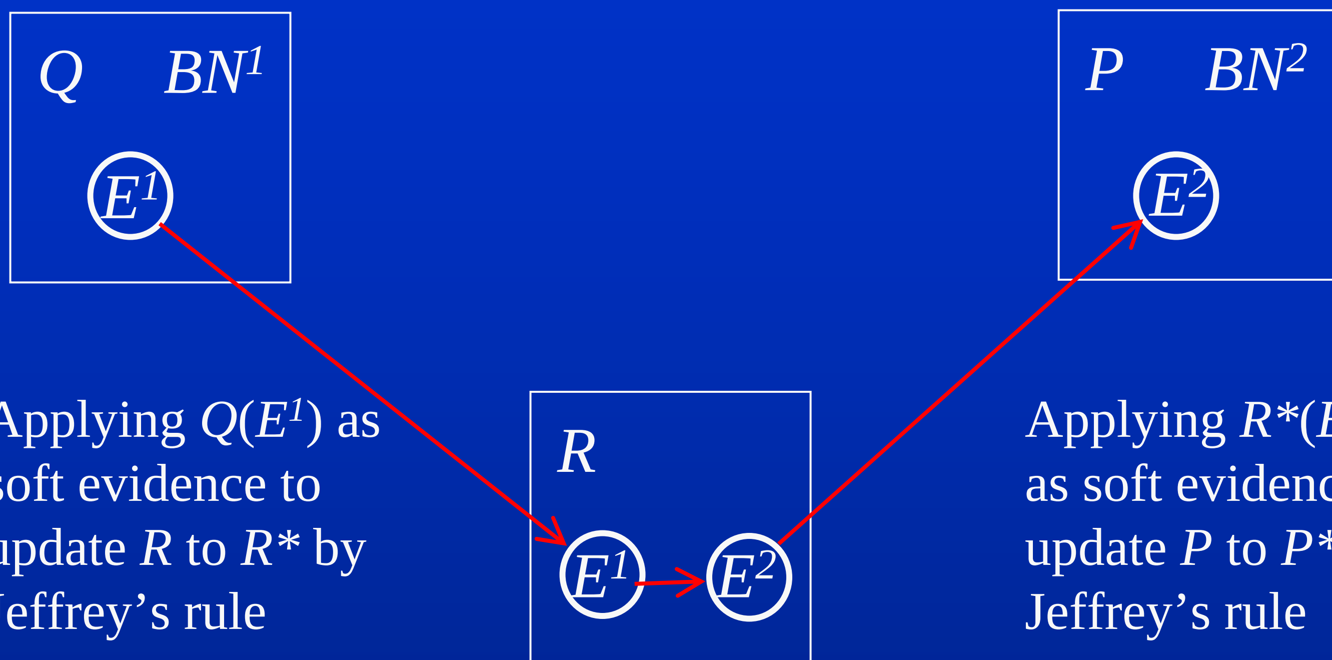
an Honors University in Maryland

Formalize The Notion of Mapping

- Jeffrey's rule
  - Conditioning cross prob. spaces
  - $P^*(x) = \Sigma P(x|E^1)Q(E^1)$
  - P^* is an I-projection of $P(x)$ on $Q(E^1)$ (minimizing Kullback-Leibler distance to P)
  - Update P to P^* by applying $Q(E^1)$ as soft evidence in BN
- similarity(E^1, E^2)
  - Represented as joint prob. $R(E^1, E^2)$ in another space R
  - Can be obtained by learning or from user
- Define

map(E^1, E^2) = <E^1, E^2, BN^1, BN^2, $R(E^1, E^2)$>

Reasoning With map(E^1, E^2)



Applying $Q(E^1)$ as soft evidence to update R to R^* by Jeffrey's rule

Applying $R^*(E^2)$ as soft evidence to update P to P^* by Jeffrey's rule

Using similarity(E^1, E^2):

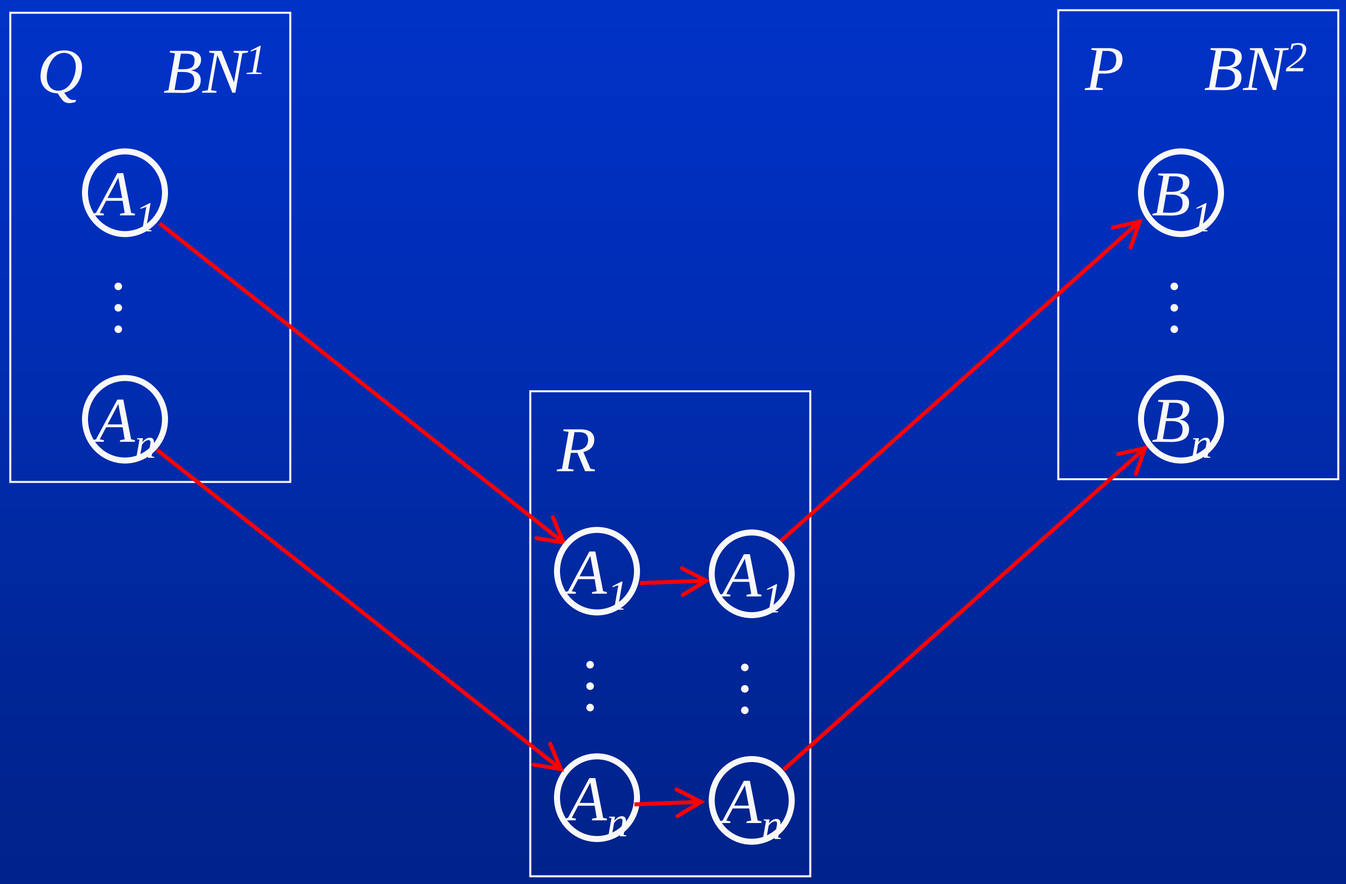
$R^*(E^2)$

$= R^*(E^1, E^2)/R^*(E^1)$

UMBC

an Honors University in Maryland

Reasoning With Multiple map(E^1, E^2)



Multiple pair-wise mappings: map(A_k, B_k):
Realizing Jeffrey's rule by IPFP

Mapping Reduction

- Multiple mappings
  - One node in BN1 can map to all nodes in BN2
  - Most mappings with little similarity
  - Which of them can be removed without affecting the overall
- Similarity measure:
  - *Jaccard-coefficient*: $\mathrm{sim}(E^1, E^2) = P(E^1 \cap E^2)/R(E^1 \cup E^2)$
  - A generalization of subsumption
  - Remove those mappings with very small sim value
- Question: can we further remove other mappings
  - Utilizing knowledge in BN

Conclusions

- Summary
  - A principled approach to uncertainty in ontology representation, reasoning and mapping
- Current focuses:
  - OWL-BN translation: properties
  - Ontology mapping: mapping reduction
- Prototyping and experiments
- Issues
  - Complexity
  - How to get these probabilities