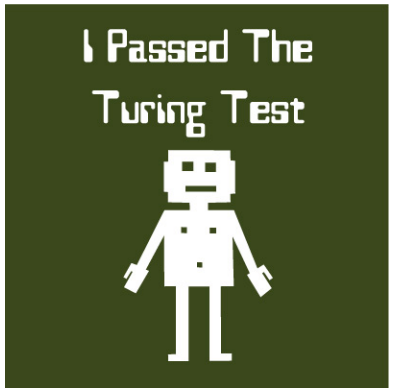
Global Learning of Textual Entailment Graphs

Supervisors: Eytan Ruppin, Ido dagan, Shimon Edelman

15/8/12

Natural Language Understanding

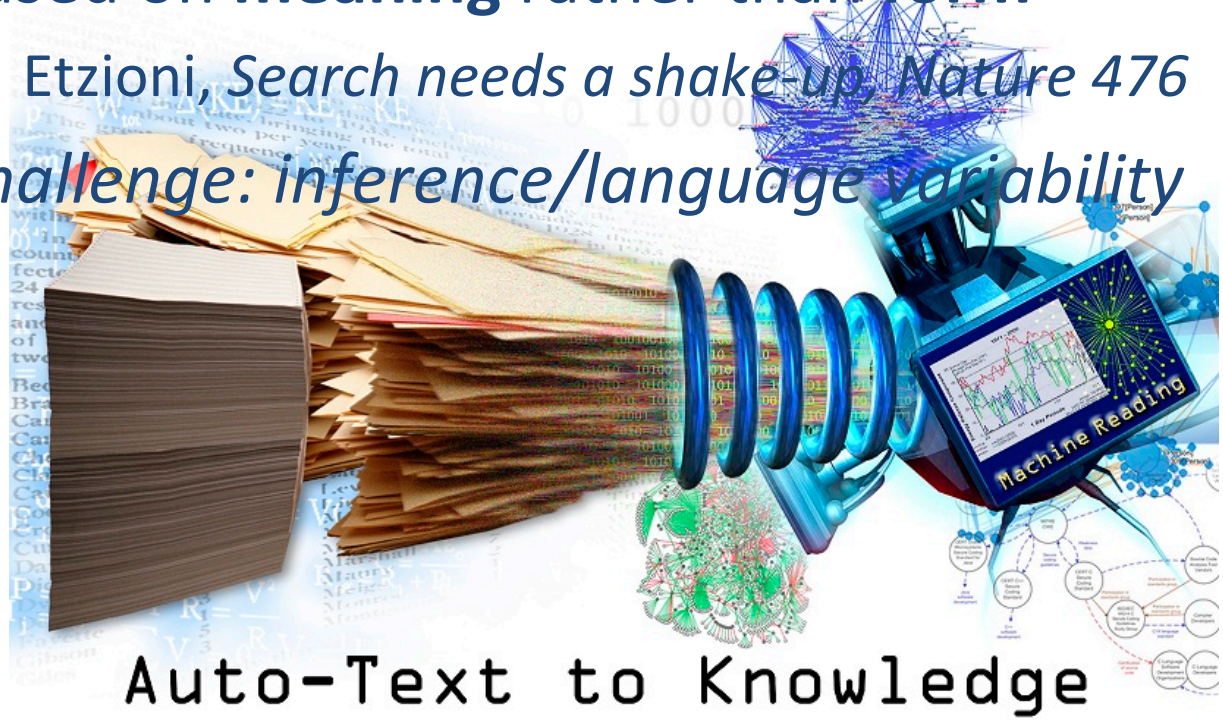
- Roots in early AI (50's)



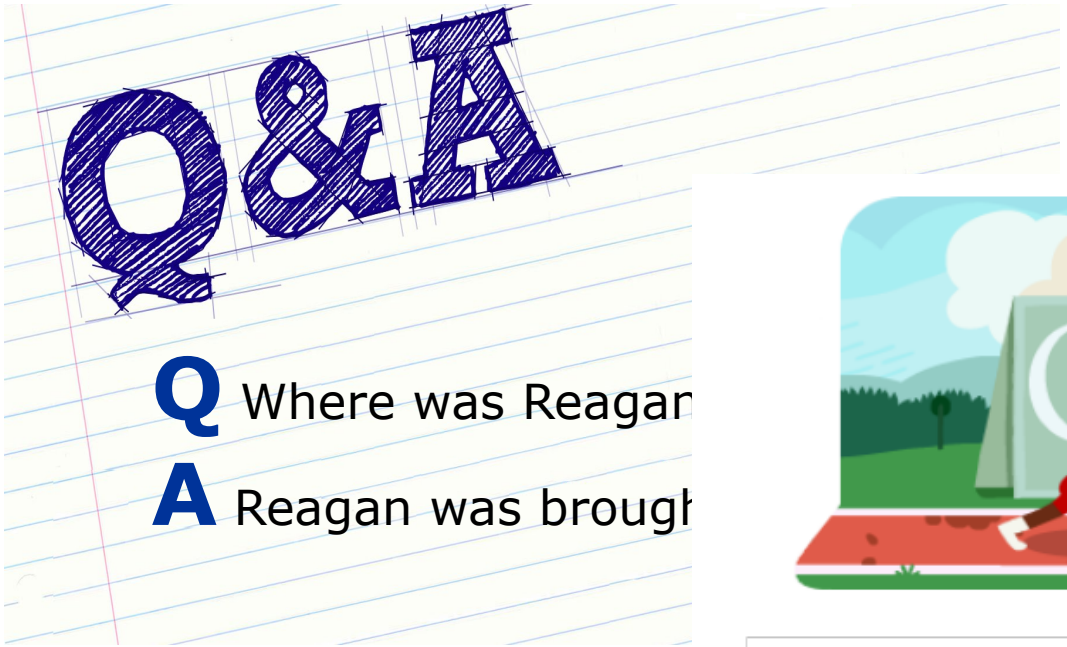
- Information explosion
  - 2008: 1 trillion unique URLs found by Google

Organizing Textual Information

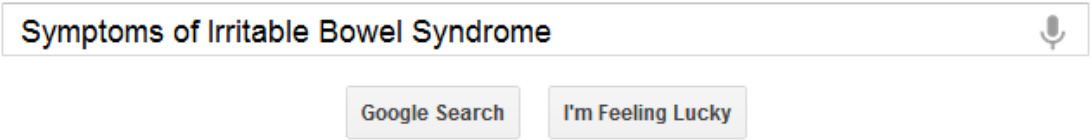
- Acute need to organize textual information
  - Based on **meaning** rather than **form**
    - Etzioni, *Search needs a shake-up, Nature 476*
  - *Challenge: inference/language variability*



NLU Applications



IR



… IBS is characterized by lower abdominal pain …

NLU Applications

IE

… Google acquires mobile E-mail utility reMail …

Purchase events:

IBM	Coremetrics
Yahoo!	Overture
Google	reMail
…	…

Sacramento Kings

The Sacramento Kings are a professional basketball team based in Sacramento, California, United States. They are members of the Western Conference of the National Basketball Association. Wikipedia

en.wikipedi…

Founded: 1945

Arena: Power Balance Pavilion

Head coach: Keith Smart

Division: Pacific Division

Location: Sacramento

Conference: Western Conference

Player	Num	Position
Jimmer Fredette	7	Point guard
DeMarcus Cousins	15	Center
Tyreke Evans	13	Shooting guard
Isaiah Thomas	22	Point guard
Chuck Hayes	42	Power forward

NLU Applications

- Multi-document Summarization

Obama gave a speech last night in the Israeli lobby conference

In his speech at the American Israel Public Affairs Committee yesterday, the president challenged …

Barack Obama's AIPAC address yesterday …

Textual Entailment

Textual Entailment (TE) - Definition

- A directional relation between two text fragments: Text (*t*) and Hypothesis (*h*).

t textually entails *h* ($t \Rightarrow h$) if humans reading *t* will infer that *h* is most likely true

t: Usain Bolt celebrates 100m gold with Swedish women's handball team

h: Usain Bolt won a gold medal

- <u>Recognising Textual Entailment (RTE)</u>: given *t* and *h*, recognize whether *t* textually entails *h*

Appeal of Textual Entailment

- Inferences required by semantic applications can be reduced to recognizing textual entailment:
  - QA: (Harabagiu and Hickl, 2006)
    - *Reagan was brought up in Dixon* $\Rightarrow$ *Reagan was raised in X*
  - IE: (Romano et al., 2006)
    - *Google acquired e-mail utility reMail* $\Rightarrow X_{company}$ *purchase* $Y_{company}$
  - Summarization: omit sentences entailed by current summary
- Led to seven challenges since 2005 for semantic inference systems

Entailment rules

- TE systems require **knowledge**
  - encoded as entailment rules:

$$giraffe \Rightarrow mammal$$
$$X \text{ was brought up in } Y \Rightarrow X \text{ was raised in } Y$$
$$X_{city} \text{ capital of } Y_{country} \Rightarrow X_{city} \text{ be part of } Y_{country}$$
$$X\ [VERB_{trans}]\ Y \Rightarrow Y \text{ be } [VERB_{participle}] \text{ by } X$$

- Automatic methods for learning large-scale knowledge resources is imperative

Predicative Entailment Rules

- <u>Proposition:</u> NL expression containing one predicate and at least one argument:

  Andy Murray jumped up and down
  Ryan Lochte beat Michael Phelps
  Yohan Blake passed the baton to Ussain Bolt

- <u>Propositional template:</u> proposition where at least one argument is replaced by a variable:

  X jumped up and down
  X_{player} *beat* Y_{player}
  X passed the baton to Y

Predicative Entailment Rules

Y is a symptom of X $\Rightarrow$ X cause Y

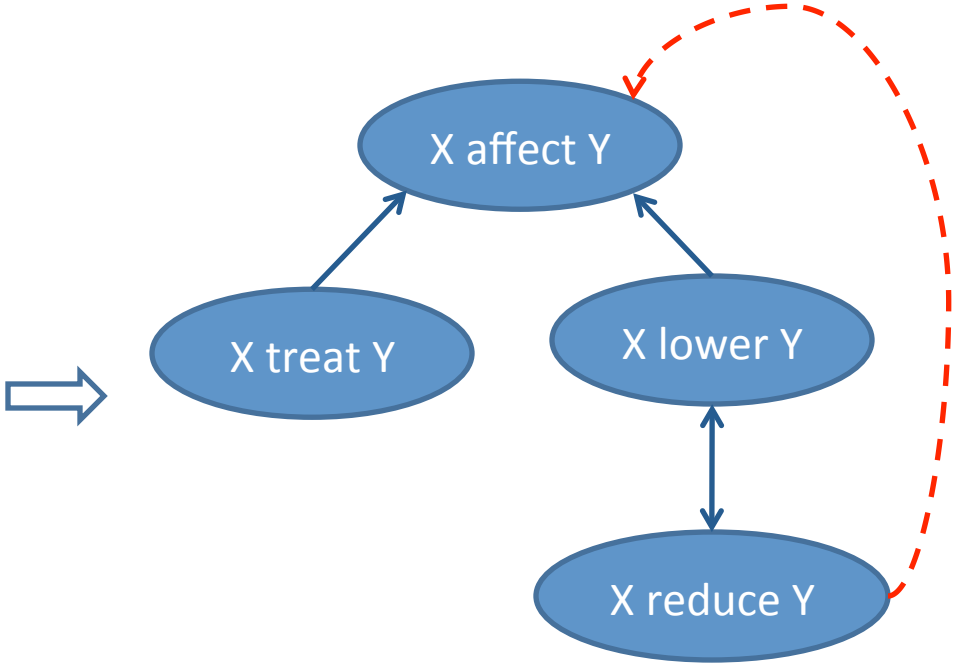
X cause an increase in Y $\Rightarrow$ X affect Y

X's treatment of Y $\Rightarrow$ X treat Y

- Information is conveyed by propositions

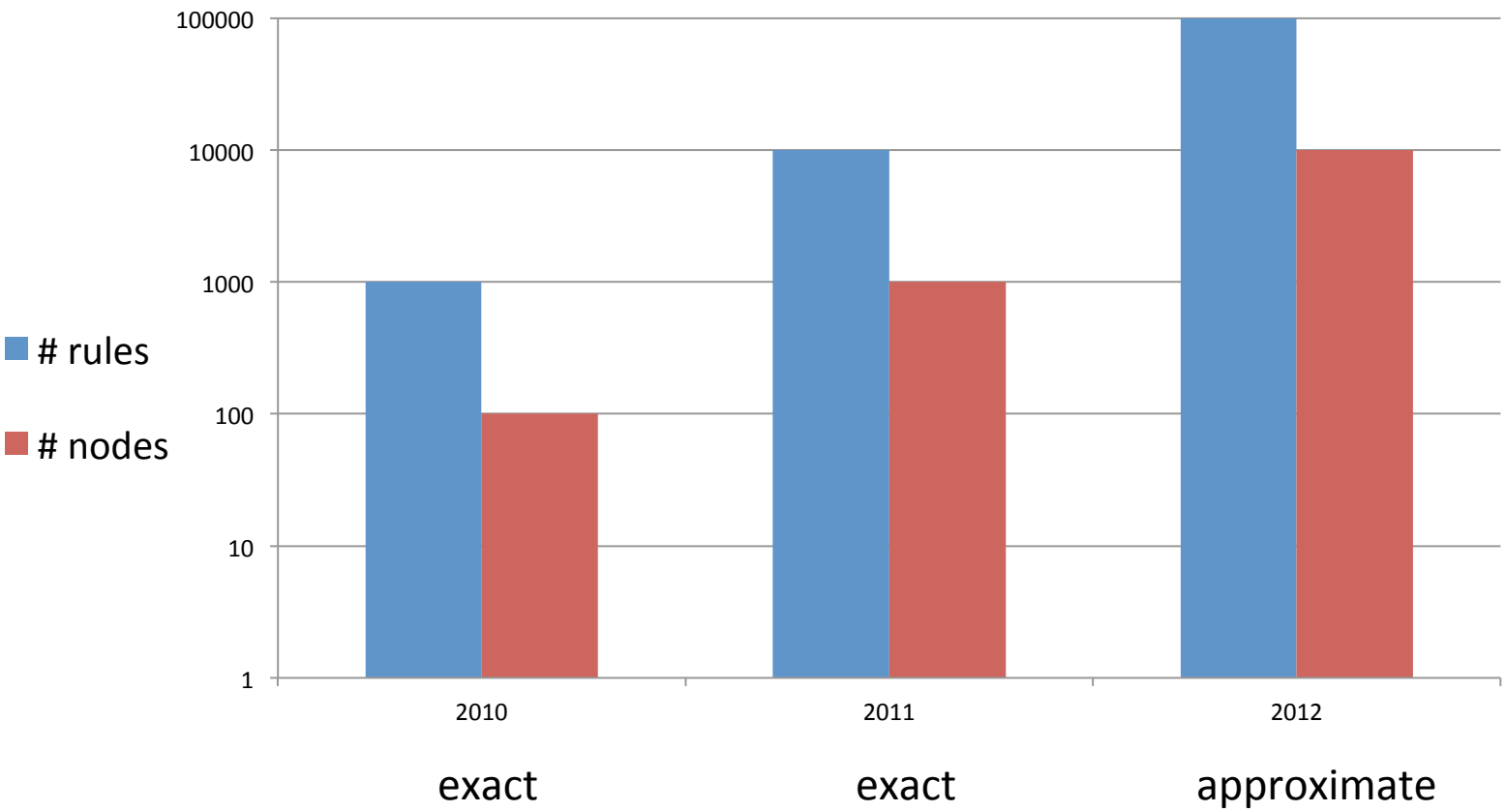
Contribution 1: Entailment Graphs

X affect Y ⇒ X treat Y	✗
X treat Y ⇒ X affect Y	✓
X affect Y ⇒ X lower Y	✗
X lower Y ⇒ X affect Y	✓
…	
…	
X lower Y ⇒ X reduce Y	✓
X reduce Y ⇒ X lower Y	✓



- <u>Advantage</u>: employ structural constraints

Contribution 2: Scaling



Contribution 3: Applications

- Public release of learned rules:
  1. High precision resource: 100,000 rules
  2. High coverage resource: 10,000,000 rules
- Textual exploration application

irsrv2.cs.biu.ac.il:8080/exploration/

BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: *alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs,*

---

headache [Explore]

- associate _ with headache | associate headache with _ (287)
- _ experience headache | _ have headache | _ suffer from headache (82)
- headache accompany _ (59)
- _ treat symptom of headache (18)
  - _ treat headache (16)
    - _ relieve headache (5)
      - _ reduce headache (1)
    - _ reduce headache (1)
- symptom of _ poisoning include headache (23)
- _ accompany headache (20)
- headache common in _ (8)
- _ prevent headache (7)

Drug, Chemical or Other Substance (7)

Test or Procedure (3)

Occupation or Discipline (2)

Behavior or Activities (1)

Disease or Natural Phenomenon or Process (1)

Food (1)

high blood pressure (1)

tea (1)

Mayo Clinic - Tips for Women (and Men) in Search of a Good Night's Sleep

In addition to the usual sources coffee **tea** soda be aware of caffeine in chocolate in medications used to **treat headaches** colds sinus_congestion

Outline

- Background
- Basic algorithm and ILP formulation
- Scalability 1: Graph decomposition
- Scalability 2: Tree-based approximation
- Text exploration application

Background

Local Learning

Input: pair of predicates (p_1,p_2)
Question: $p_1 \rightarrow p_2$?

- Sources of information:
  1. Lexicographic: WordNet (Szpektor and Dagan, 2009), FrameNet (Ben-aharon et al, 2010)
  2. Pattern-based (Chklovsky and Pantel, 2004)
  3. Distributional similarity (Lin and Pantel, 2001; Szpektor and Dagan, 2008; Bhagat et al, 2007;Yates and Etzioni, 2009; Poon and Domingos 2010; Schoenmackers et al., 2010)

Properties of Information Sources

1. Lexicographic:
   - WordNet relations: hyponym, derivation, entailment
   - Limited coverage: *"X cause a reduction in Y"*
2. Pattern-based:
   - *"he scared and even startled me"*
   - Requires very large corpus (web-scale)
   - Distinguishes different semantic relations:
     - "to X and even Y" vs. "Either X or Y"
3. Distributional similarity

Distributional Similarity

X affect Y		
X	Y	#
insulin	metabolism	7
Zantac	BP	4
…	…	…

X treat Y		
X	Y	#
Zantac	BP	9
diet	diabetes	2
…	…	…

- Vary w.r.t to representation and similarity computation
  - Symmetric vs. directional
- Good coverage
- Discerning exact semantic relation is difficult

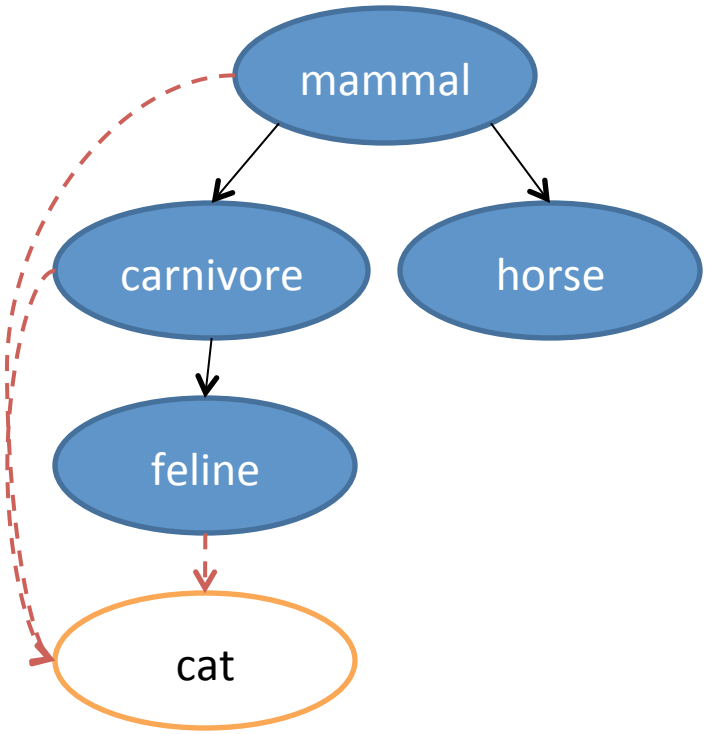
Global Learning

Input: Set of predicates P
Question: Find $E = \{ (p_1,p_2) \mid p_1 \rightarrow p_2 \}$

- Applying transitivity constraints:
  - Undirected graphs:
    - Paraphrasing (Yates and Etzioni, 2009)
    - Co-reference resolution (Finkel and manning, 2008)
  - Directed graph
    - Taxonomy induction (Snow et al., 2006)
    - Ontology induction (Poon and Domingos, 2010)
    - Temporal information extraction (Ling and Weld, 2010)

Global Learning

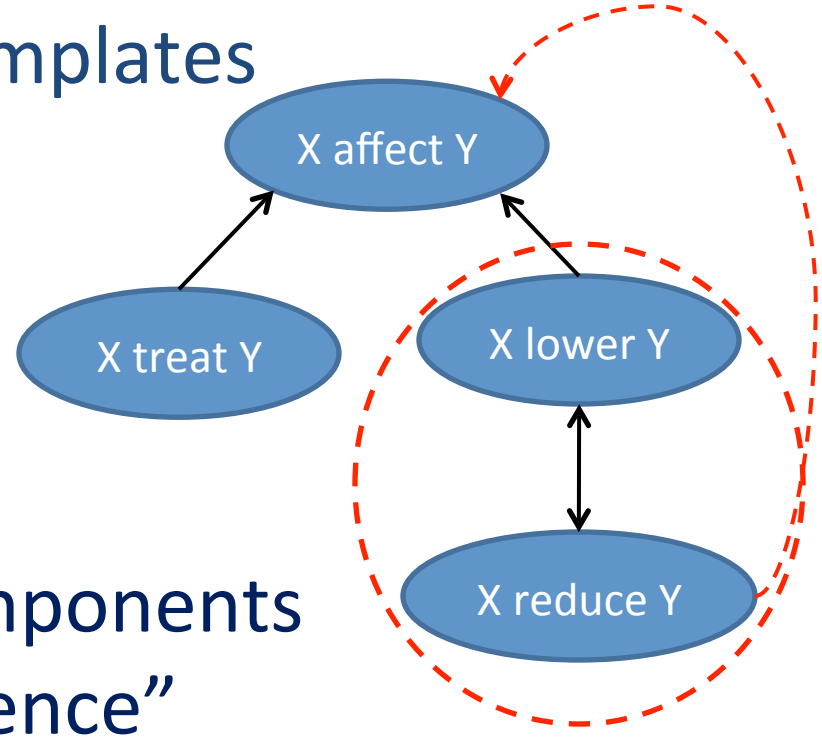
- Snow et al. (2006) presented an algorithm for taxonomy induction.
- At each step they add the concept that maximizes the likelihood of the taxonomy given the transitivity constraint.



Basic Algorithm and ILP Formulation

Entailment Graphs

- Nodes: propositional templates
- Edges: entailment rules



Properties

- Entailment is **transitive**
- Strong connectivity components correspond to "equivalence"
  - Caveat: ambiguity

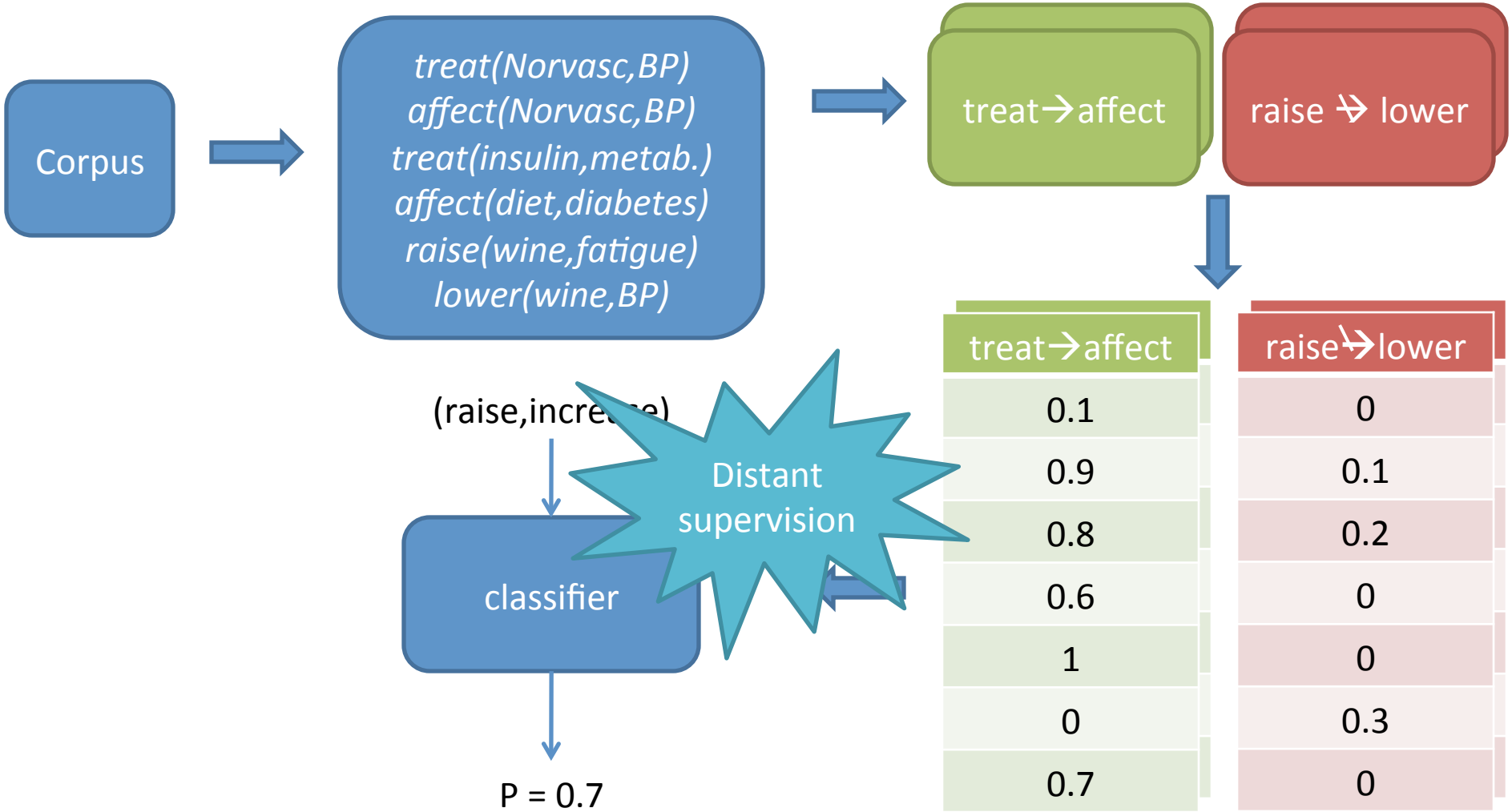
Learning Entailment Graph Edges

Input: Corpus C

Output: Entailment graph $G = (P,E)$

1. Extract propositional templates P from C
2. Train a **local** entailment classifier: given (p_1,p_2), estimate whether $p_1 \rightarrow p_2$
3. Decoding: Find the edges of the graph using the local probabilities and a transitivity constraint

Local Entailment Classifier – Step 1+2



Graph Objective Function

- Use local classifier probabilities p_{ij} to express the graph probability:

$$\hat{G} = \arg\max_{X} \sum_{i \neq j} w_{ij} \cdot x_{ij}$$

$$\begin{cases} 1 & i \rightarrow j \\ 0 & \text{else} \end{cases}$$

$$w_{ij} = \log \frac{p_{ij} \cdot \theta}{(1 - p_{ij}) \cdot (1 - \theta)}$$

"density" prior

Global Learning of Edges – Step 3

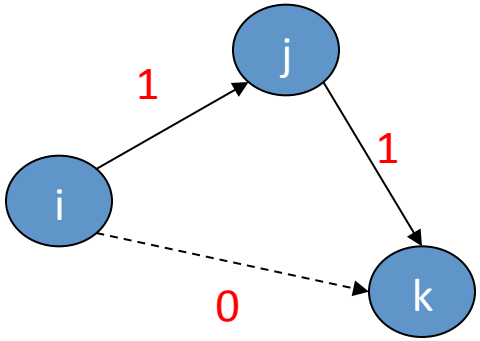
Input: Set of nodes V, weighting function $w: V \times V \rightarrow R$
Output: Set of directed edges E that maximizes the objective function under a **global transitivity constraint**

- Problem is NP-hard:
  - Reduction from "Transitive Subgraph" (Yannakakis, 1978)

Input: Directed graph $G = (V,E)$
Output: Maximal set of edges $A \subseteq E$ such that $G' = (V,A)$ is transitive

- Integer Linear Programming Formulation

Integer Linear Program



$$\hat{G} = \arg\max \sum_{i \neq j} w_{ij} \cdot \boxed{x_{ij}}$$

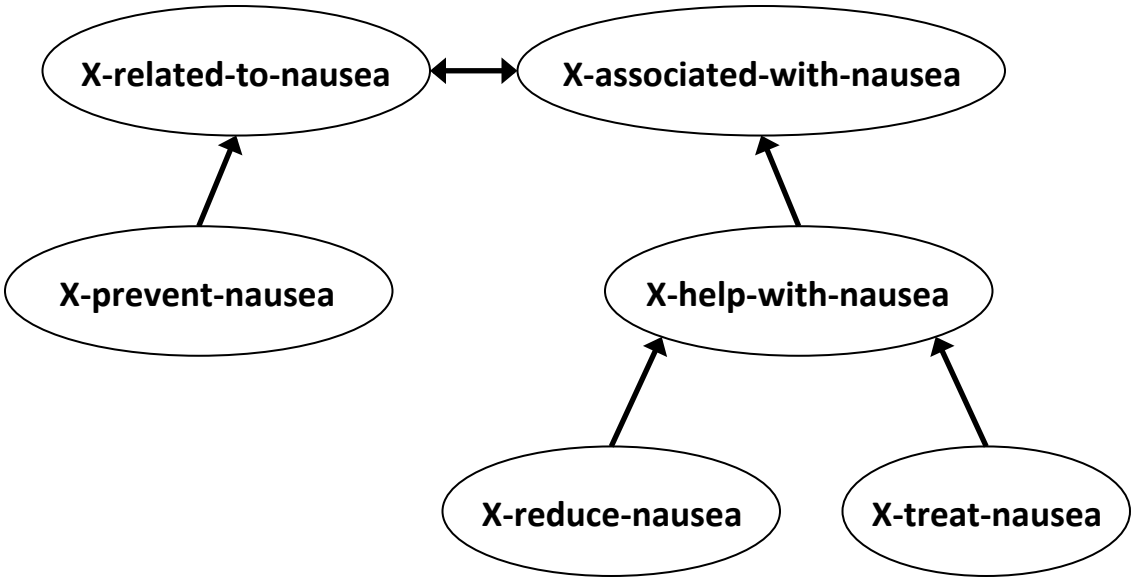
$$\forall i, j, k \in V, x_{ij} + x_{jk} - x_{ik} \leq 1$$

$$x_{ij} \in \{0,1\}$$

1+1-0 = 2 > 1

- Variables: x_{ij}
- Objective function: maximizes P(G)
- Global transitivity constraint: - $O(|V|^3)$ constraints

Evaluation: Focused Entailment Graphs



- Argument is instantiated by a target concept *(nausea)*
- Instantiating an argument reduces ambiguity

irsrv2.cs.biu.ac.il:8080/exploration/

BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: *alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs,*

---

headache [Explore]

- associate _ with headache | associate headache with _ (287)
- _ experience headache | _ have headache | _ suffer from headache (82)
- headache accompany _ (59)
- _ treat symptom of headache (18)
  - _ treat headache (16)
    - _ relieve headache (5)
      - _ reduce headache (1)
    - _ reduce headache (1)
- symptom of _ poisoning include headache (23)
- _ accompany headache (20)
- headache common in _ (8)
- _ prevent headache (7)

Drug, Chemical or Other Substance (7)

Test or Procedure (3)

Occupation or Discipline (2)

Behavior or Activities (1)

Disease or Natural Phenomenon or Process (1)

Food (1)

high blood pressure (1)

tea (1)

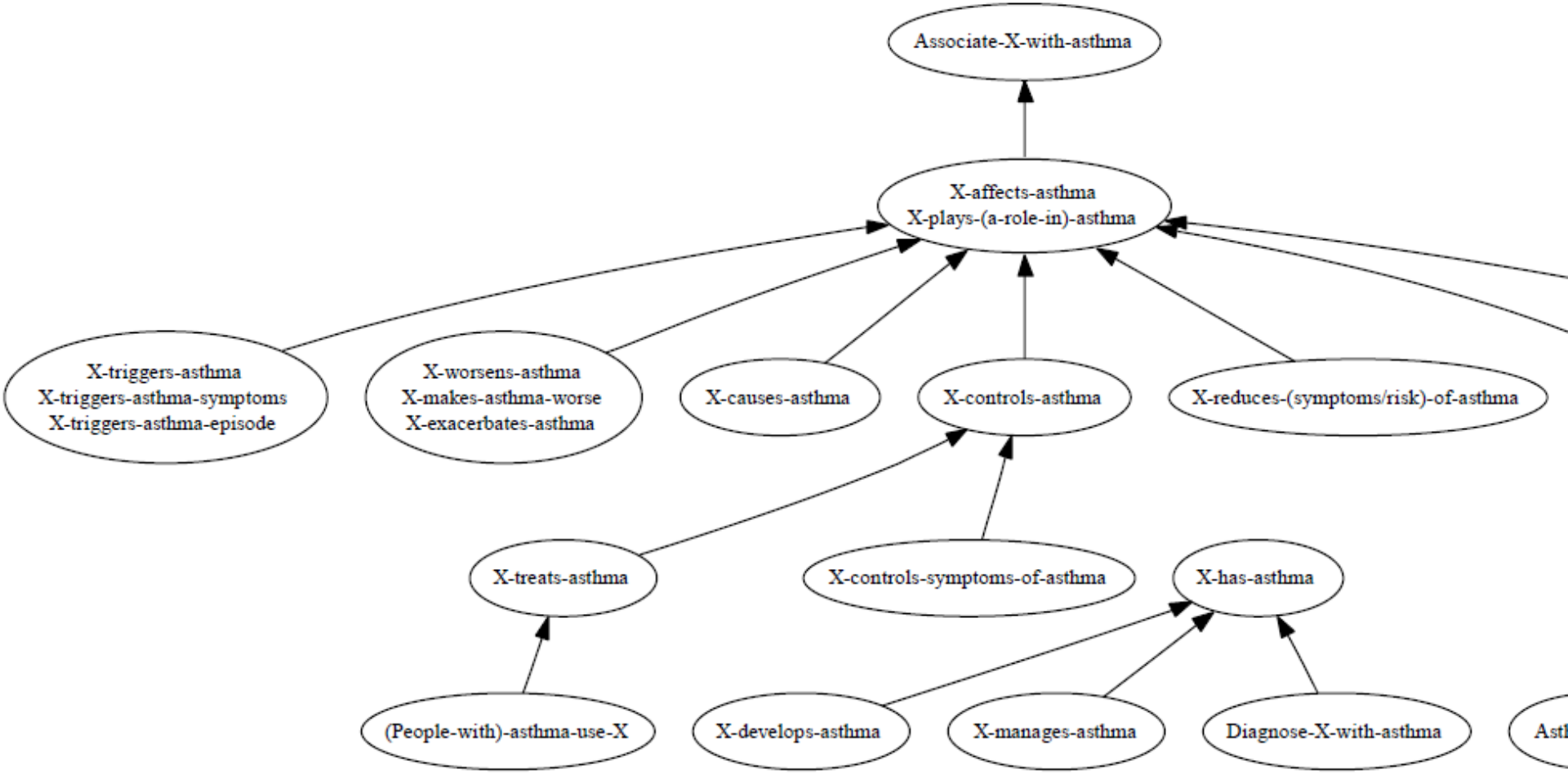
Mayo Clinic - Tips for Women (and Men) in Search of a Good Night's Sleep

In addition to the usual sources coffee **tea** soda be aware of caffeine in chocolate in medications used to **treat headaches** colds sinus_congestion

Experimental Evaluation

- 50 million word tokens **healthcare** corpus
- Ten medical students prepared gold standard graphs for 23 medical concepts:
  - Smoking, seizure, headache, lungs, diarrhea, chemotherapy, HPV, Salmonella, Asthma, etc.
- Evaluation:
  - F_1 on set of learned edges vs. gold standard

Gold Standard Graph - Asthma



Evaluated algorithms

- Local algorithms
  - Distributional similarity
  - WordNet
  - Local classifier
- Global algorithms
  - ILP/Snow et al. (greedy optimization)

Results

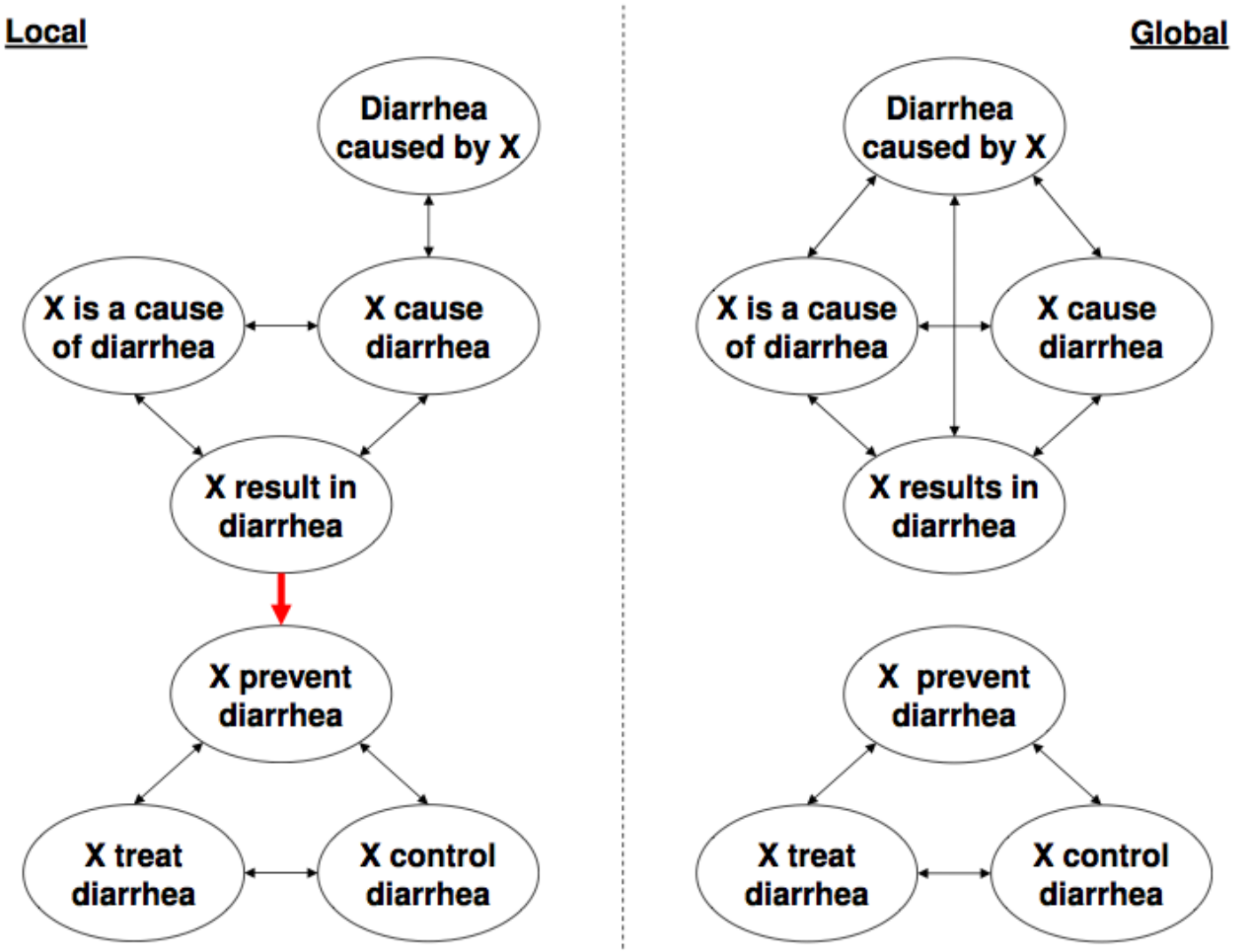
	recall	Precision	F_1
ILP-global	46.0	50.1	**43.8***
Greedy	45.7	37.1	36.6
ILP-local	44.5	45.3	38.1
$Local_1$	53.5	34.9	37.5
$Local_2$	52.5	31.6	37.7
$Local^*_1$	53.5	38.0	39.8
$Local^*_2$	52.5	32.1	38.1
WordNet	10.8	44.1	13.2

- The algorithm significantly outperforms all other baselines.

Results

	Global=true/Local=false	Global=false/Local=true
Gold standard = true	48	42
Gold standard = false	78	494

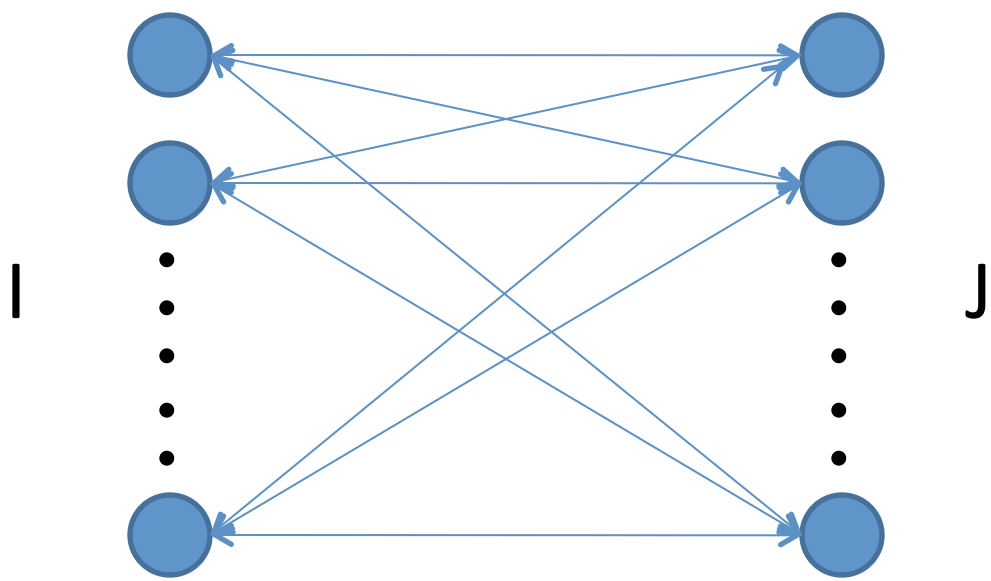
- Global algorithm avoids false positives.


Local
Diarrhea caused by X
X is a cause of diarrhea
X cause diarrhea
X result in diarrhea
X prevent diarrhea
X treat diarrhea
X control diarrhea
Global
Diarrhea caused by X
X is a cause of diarrhea
X cause diarrhea
X results in diarrhea
X prevent diarrhea
X treat diarrhea
X control diarrhea

Graph Decomposition

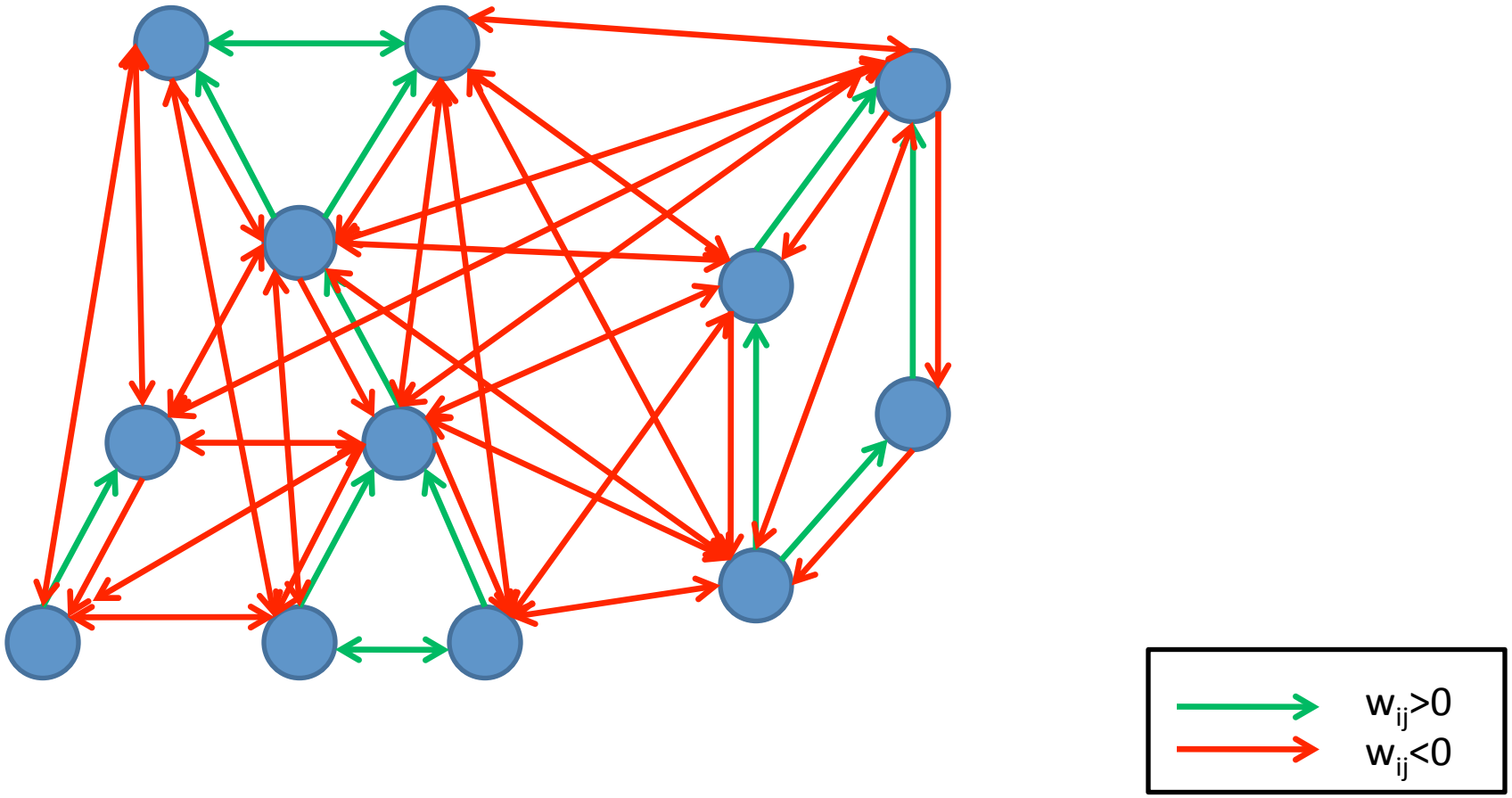
Graph Decomposition

- Proposition: If we can partition the nodes to *I,J* such that $w_{ij}<0$ for every *i,j* then any *(i,j)* is not an edge in the optimal solution

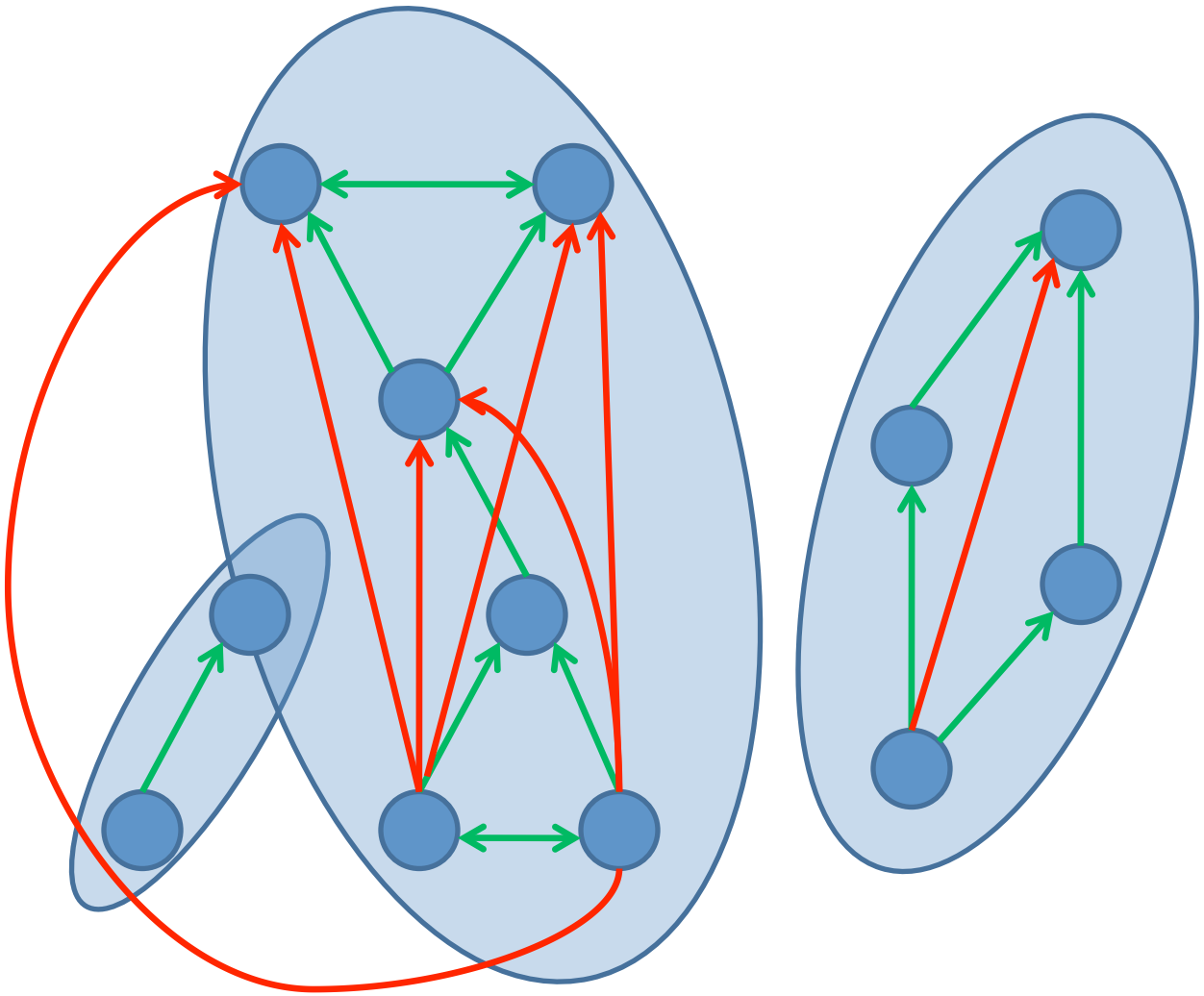


Graph Decomposition

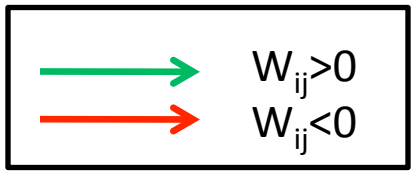
- Entailment graphs are sparse



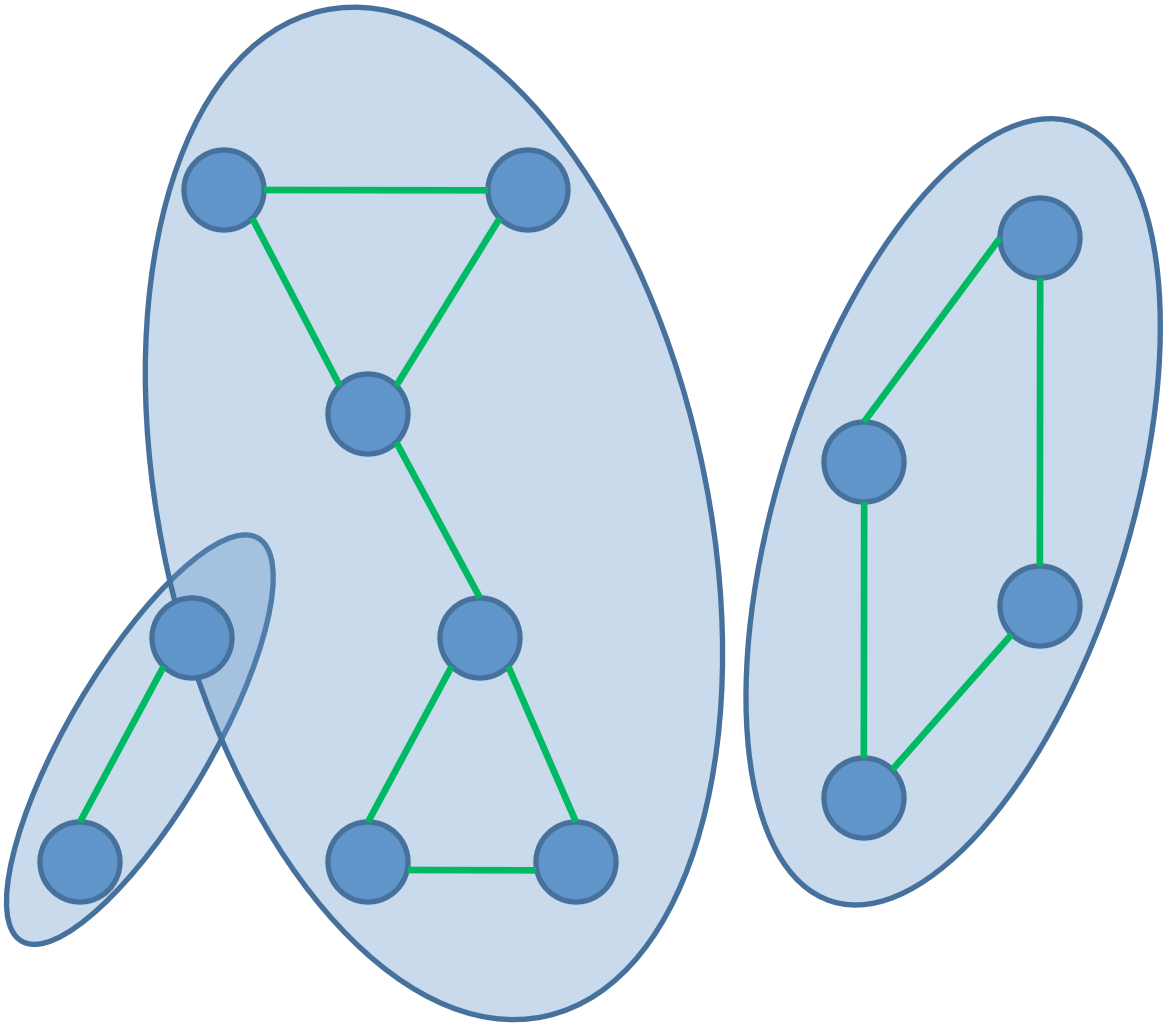
Scalability – Graph Decomposition



- Local solution
- Trans. violations
- Local solution
- Global Solution
- Local Solution
- Conn. components

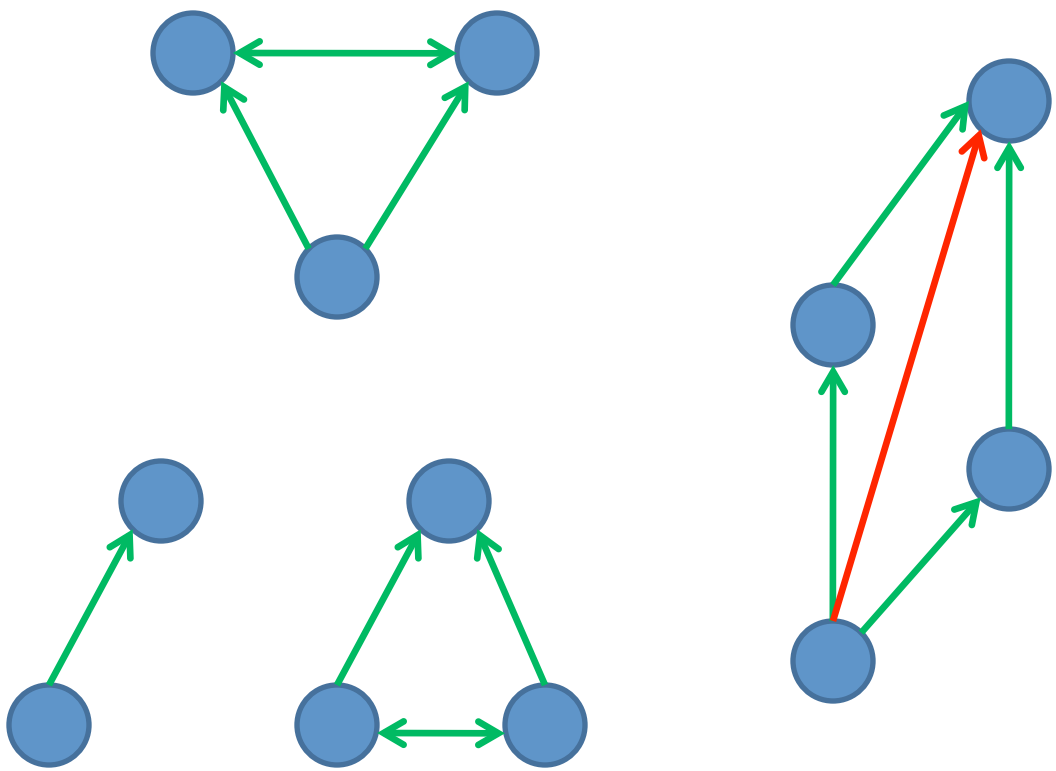


Scalability – Graph Decomposition



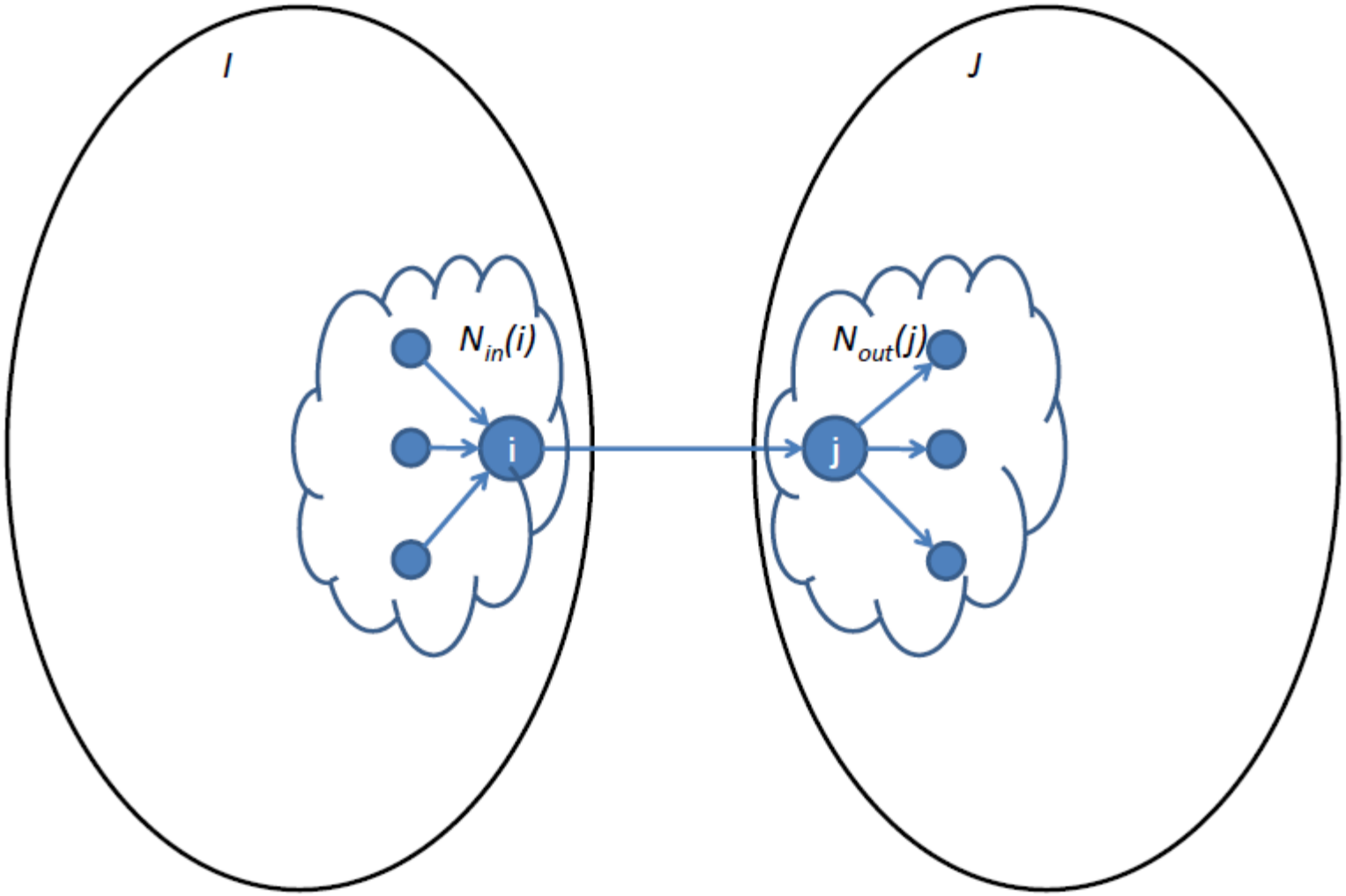
1. Input
2. Undirected positive edges
3. Conn. Components
4. Apply ILP

Scalability – Graph Decomposition



5. Global solution

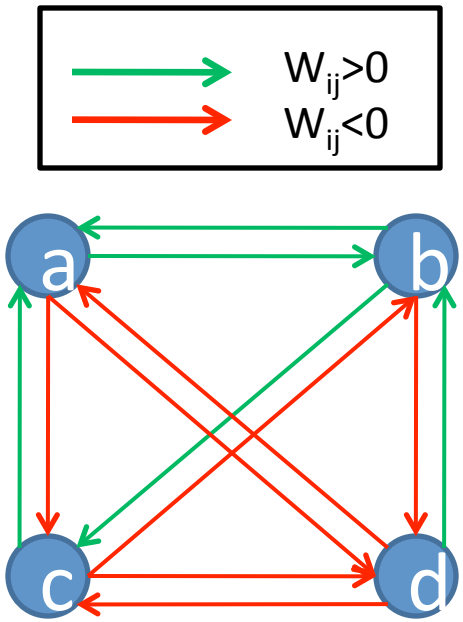
Decomposition Extension



Cutting-plane Method

- Applied in parsing (Riedel and Clarke, 2006)
- Number of constraints: $O(|V|^3)$
- A reasonable local classifier does not violate most of the constraints
- Add constraints only if they are violated

Cutting-plane Method



Constraints:

1. (c,a,b)
2. (d,b,a)
3. (d,b,c)

- Empirically converges in 6 iterations and reduces number of constraints from 10^6 to 10^3-10^4

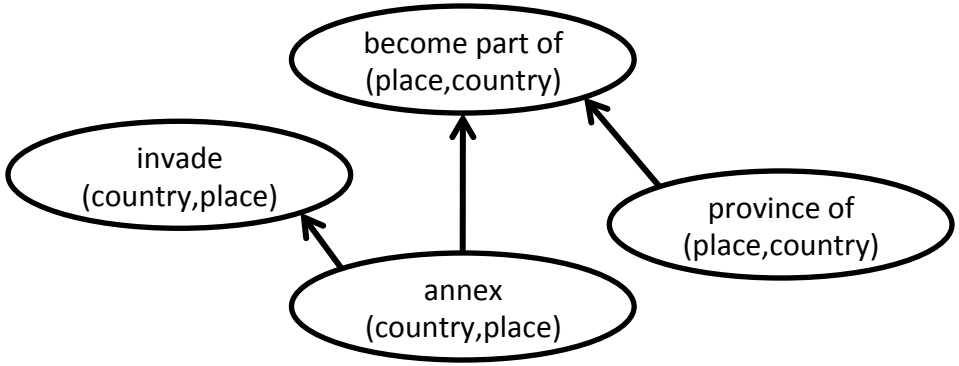
Final Algorithm

- Given a set of predicates:
  - Decompose into components
  - Apply a cutting-plane method on each component

Evaluation: Typed Entailment Graphs

- Predicate variables are typed:
  - *$X_{company}$ acquire $Y_{company}$ → $Y_{company}$ is sold to $X_{company}$*
- Rules of wide-applicability but less ambiguous
- Schoenmackers et al. (EMNLP 2010) used a <u>local</u> algorithm to learn 30,000 entailment rules
- We want to use a <u>global</u> algorithm

Typed Entailment Graphs



- A graph is defined for every pair of types
- Nodes are termed *typed predicates*
- Types alleviate the ambiguity problem

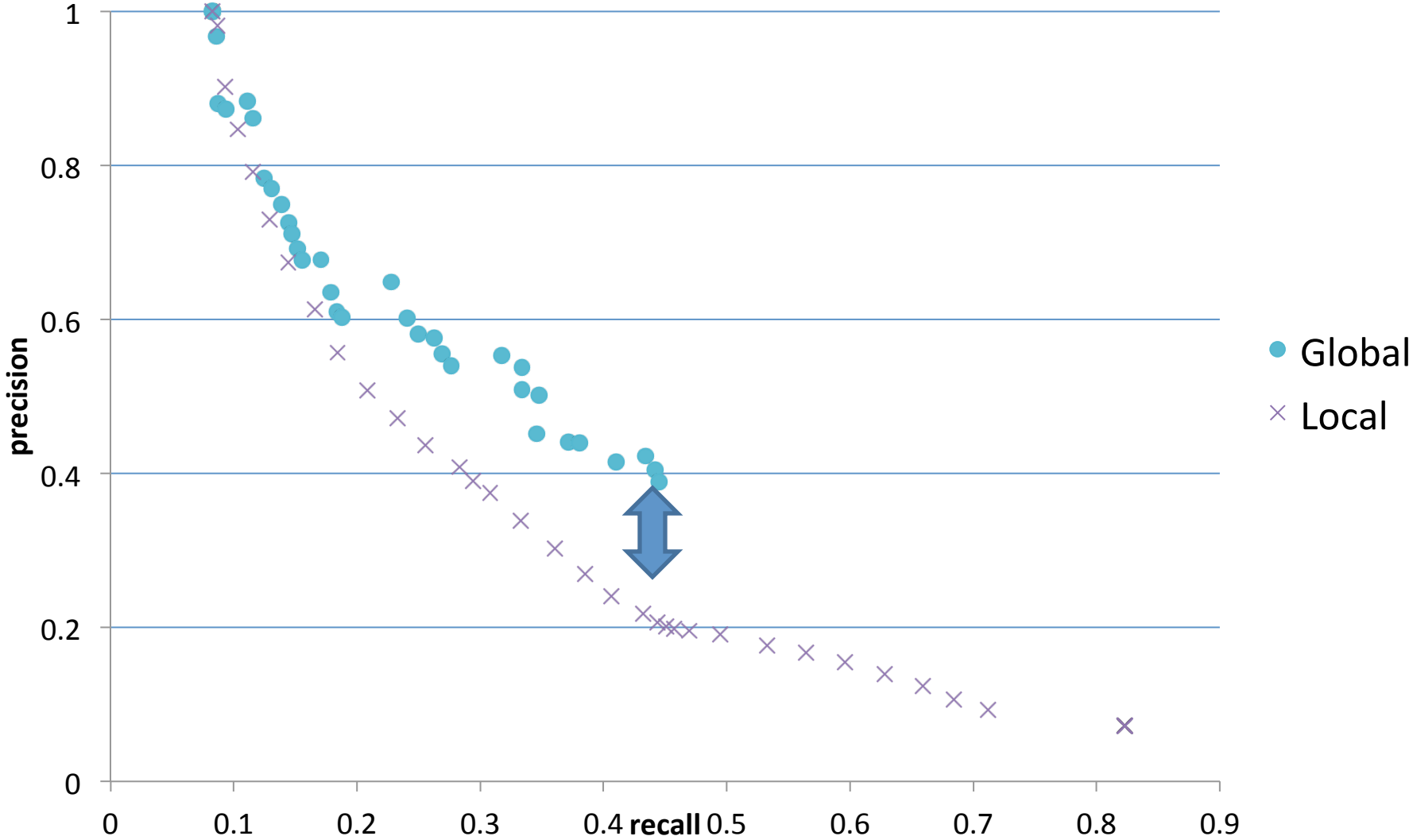
Experiment 1 - Transitivity

- 1 million TextRunner tuples over 10,672 typed predicates and 156 types
- Consist ~2,000 typed entailment graphs
- 10 gold standard graphs of sizes: 7, 14, 22, 30, 38, 53, 59, 62, 86 and 118
- Evaluation:
  - F_1 on set of edges vs. gold standard
  - Area Under the Curve (AUC)

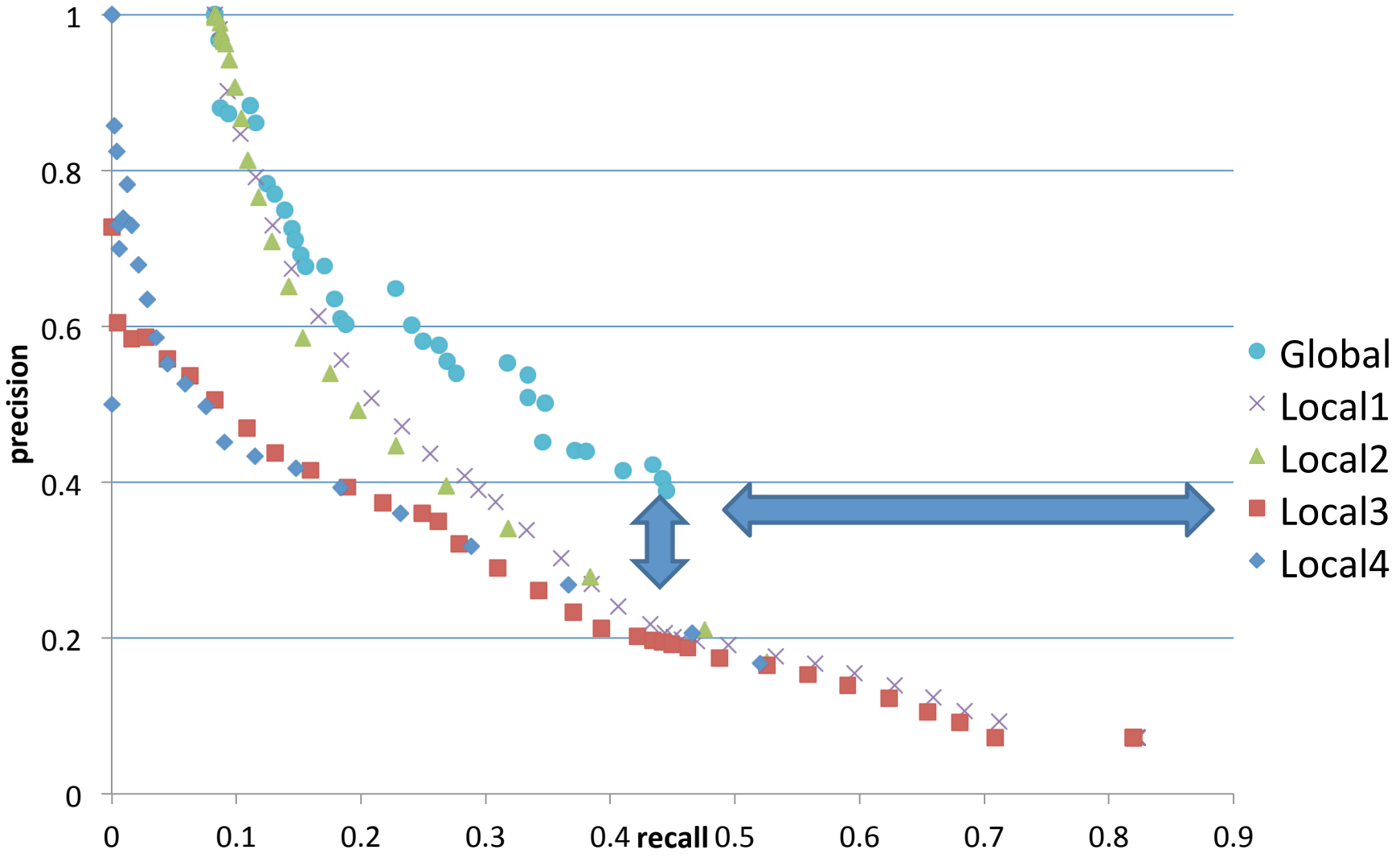
Evaluated algorithms

- Local algorithms
  - Distributional similarity (DIRT, BInc, etc.)
  - Distributional similarity + background knowledge
  - ILP with No transitivity constraints
  - Sherlock rule resource
- Global algorithm: ILP

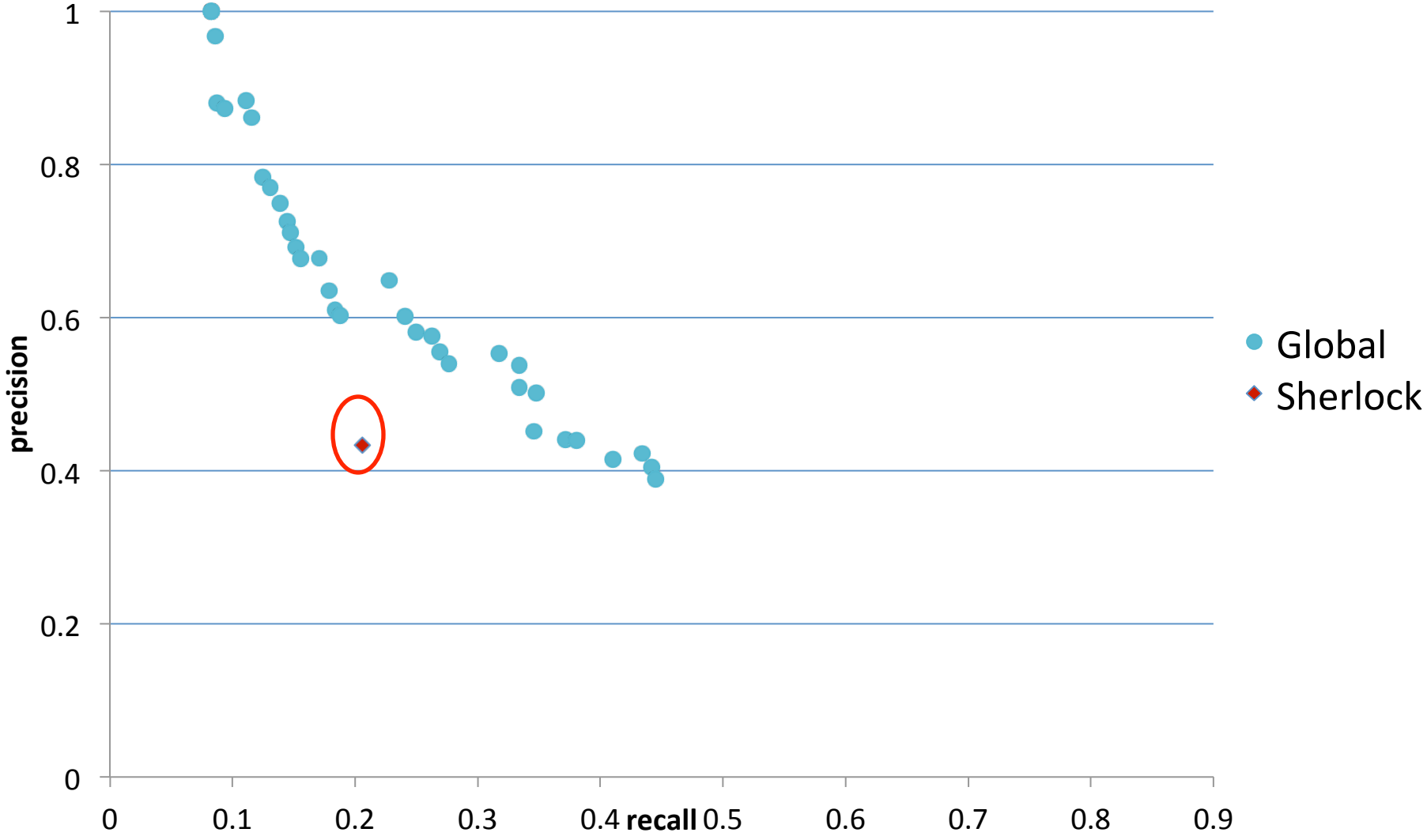
Precision-Recall: Global vs. Local



Precision-Recall: Global vs. Local



Precision-Recall: Global vs. Local



Results

	Micro-average			AUC
	R	P	F_1	
ILP	**43.4**	42.2	**42.8**	**0.22**
Clsf	30.8	37.5	33.8	0.17
Sherlock	20.6	**43.3**	27.9	N/A
SR	38.4	23.2	28.9	0.14
DIRT	25.7	31.0	28.1	0.13
BINC	31.8	34.1	32.9	0.17

- R/P/F_1 at point of maximal micro-F_1
- Transitivity improves rule learning over typed predicates

Experiment 2 - Scalability

- Run ILP with and without graph decomposition over ~2,000 graphs
- Compare number of learned rules for various sparseness parameters

Results

Sparsity	# learned rules	Δ(%)
-1.75	6,242/7,466	+20
-1	16,790/19,396	+16
-0.6	26,330/29,732	+13

- Scaling techniques add 3,500 new rules to best configuration
- Corresponds to 13% increase in relative recall

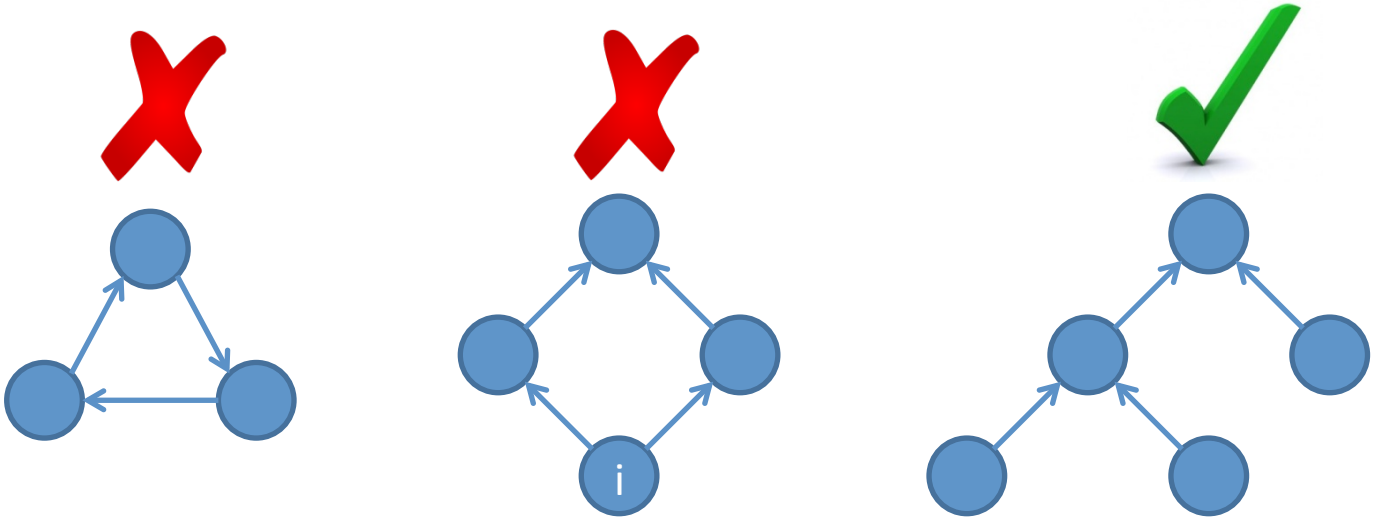
Tree-based Approximation

Tree-based Approximation - Outline

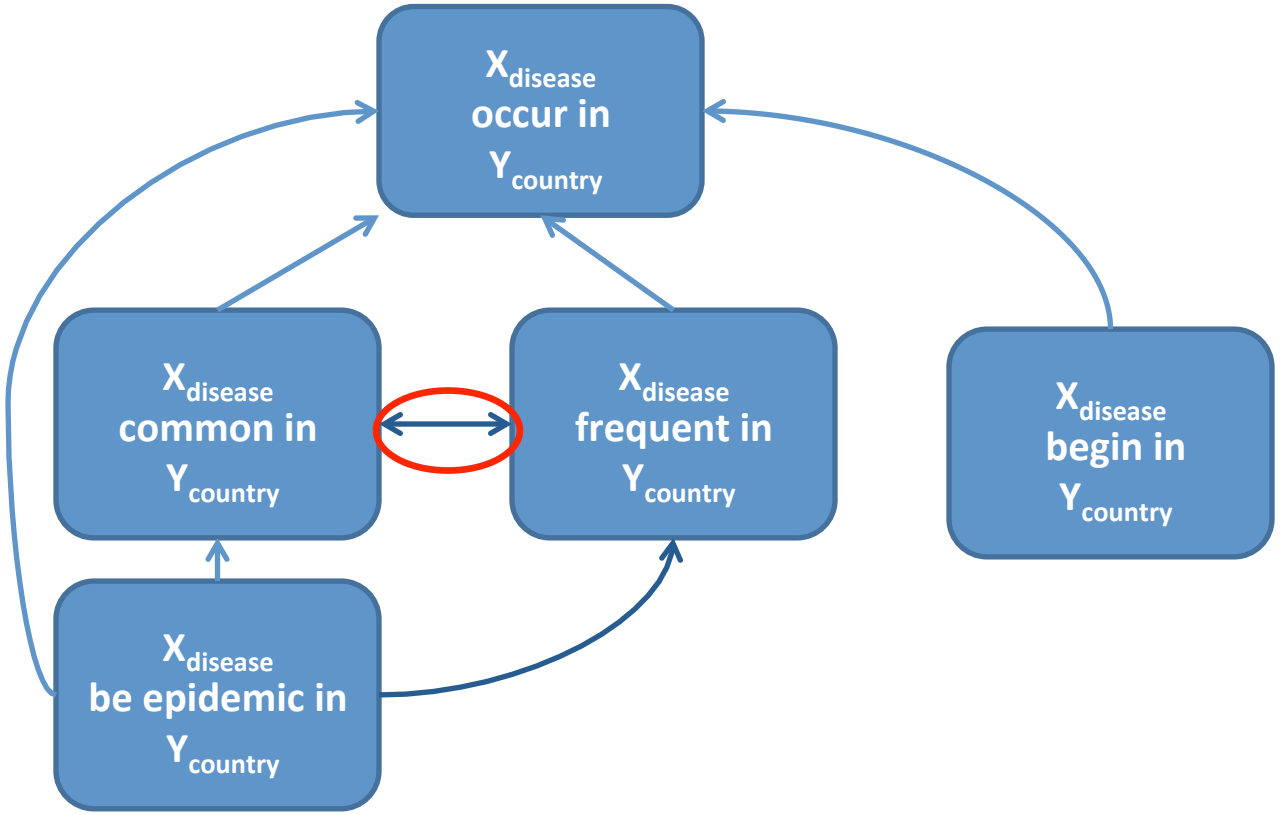
1. Forest-reducible graphs (FRG)
2. Tree-Node-Fix algorithm (TNF)
3. Experiments:
    1. Typed entailment graph
    2. Untyped entailment graphs

Directed Forest

- For an edge *i*→*j* in a DAG we say that *i* is a *child* of *j* and *j* a *parent* of *i*
- A directed forest is a DAG where all nodes have no more than one parent

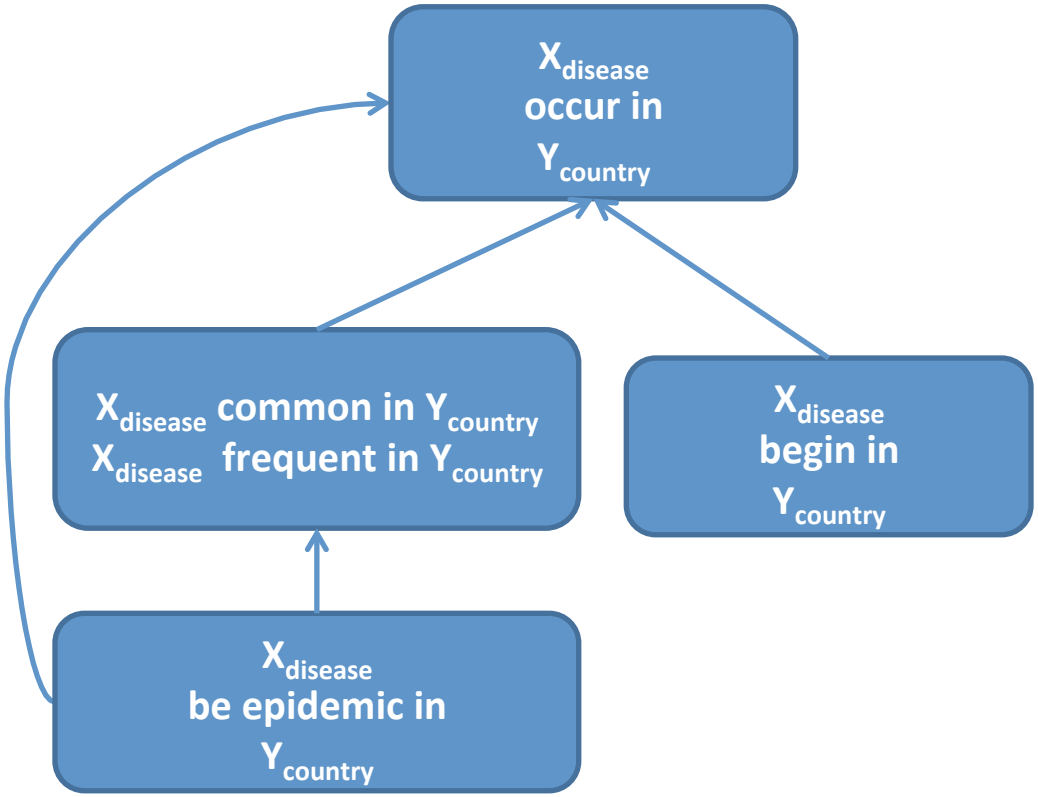


Entailment Graphs are not Forests



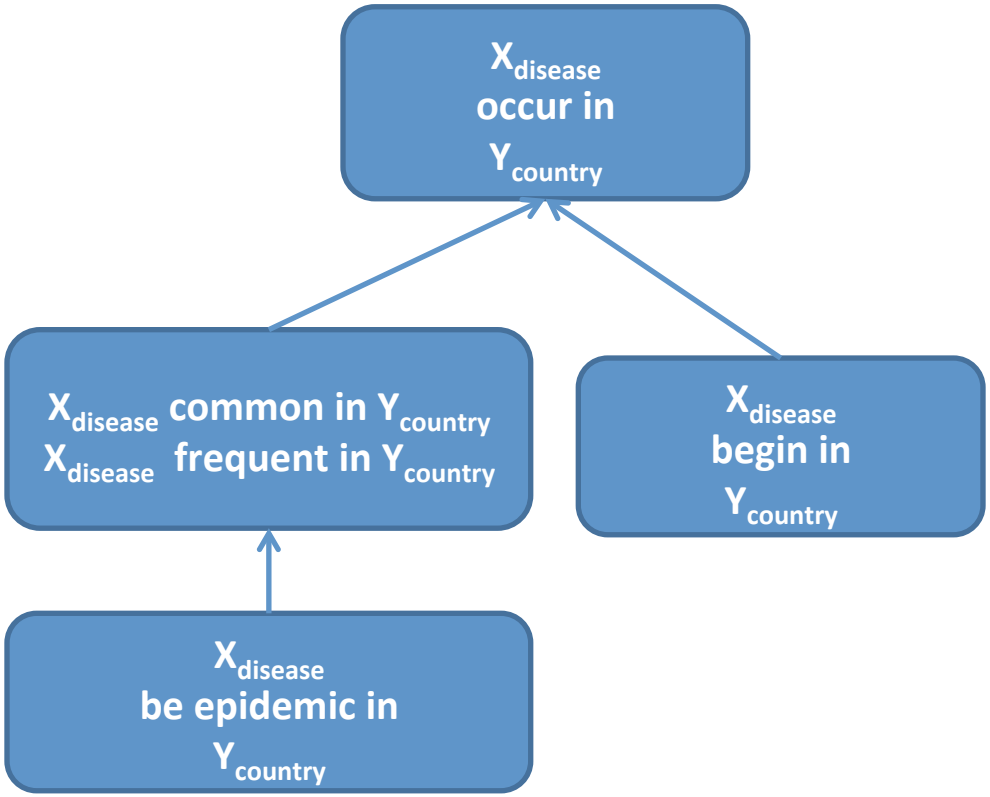
- But generally entailment is from "specific" predicates to more "general" ones

Computing SCC Graph



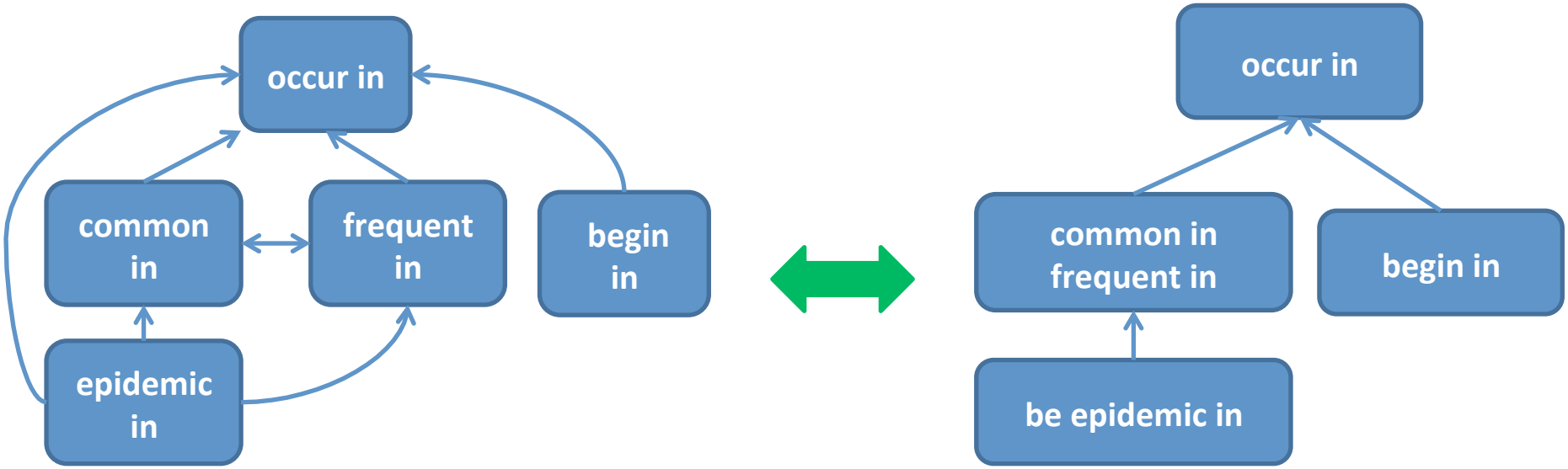
- Contract strongly connected components.

Transitive Reduction



- Delete edges inferred by transitivity

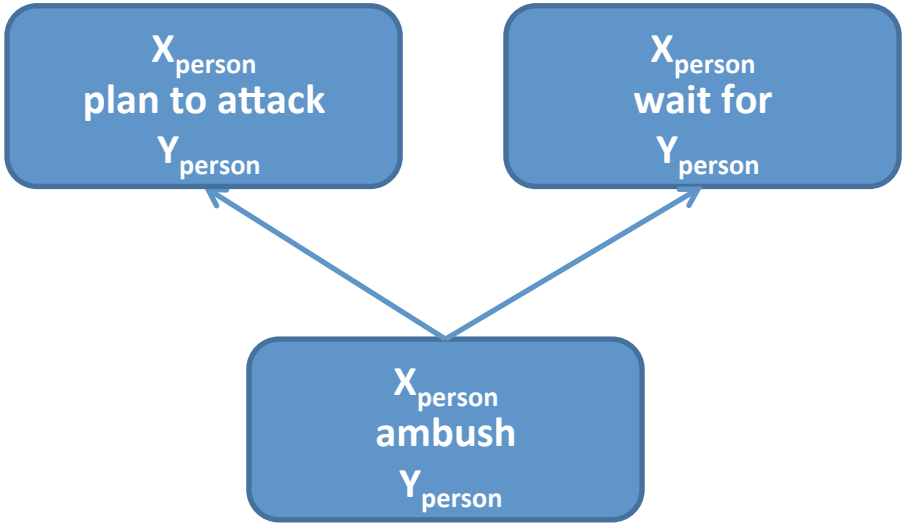
Reduced Graphs



- *Reduced graphs* are an equivalent representation for transitive graphs.
- The reduced graph in this case is a directed forest

Forest-reducible Graphs (FRG)

- Hypothesis: The reduced graph of entailment graphs is usually a directed forest, i.e., entailment graphs are *FRGs*
- Not necessarily true:

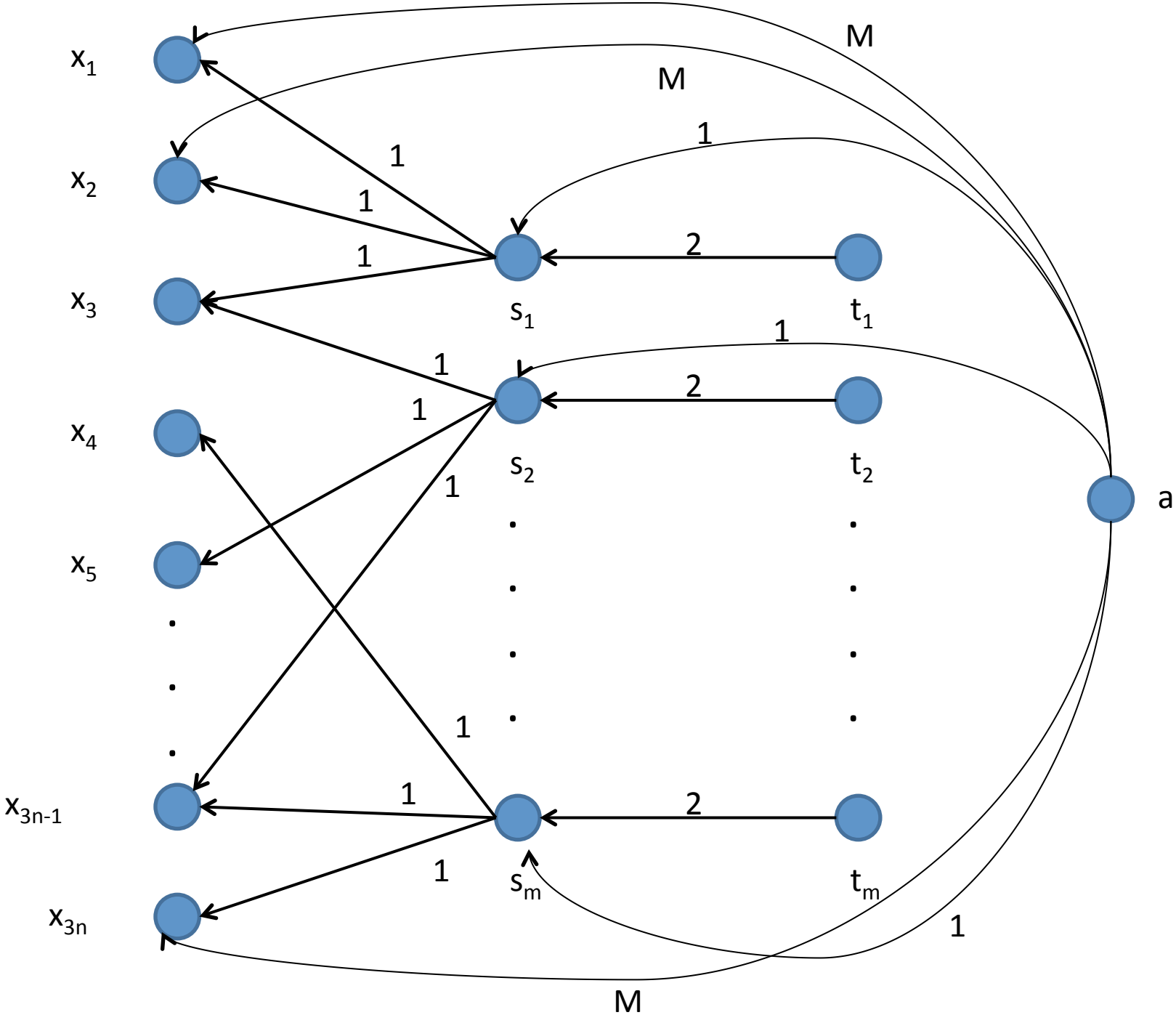


- But exceptions are rare…

FRG Assumption is reasonable

- Over our Typed entailment graph data set:
  - We manually constructed FRGs >= .95 recall and 1.0 precision on 7 gold standard graphs
- Finding the best FRG is still NP-hard, by a polynomial reduction from *X3C*.

Input: Set X of size $3n$, subsets $S_1, S_2, \ldots, S_m$ of size 3
Output: Subsets $S_{i1}, \ldots, S_{in}$ covering X


x_1
x_2
x_3
x_4
x_5
x_{3n-1}
x_{3n}
s_1
s_2
s_m
t_1
t_2
t_m
a
M
M
M
1
1
1
1
1
1
1
1
1
1
1
2
2
2

FRG Summary

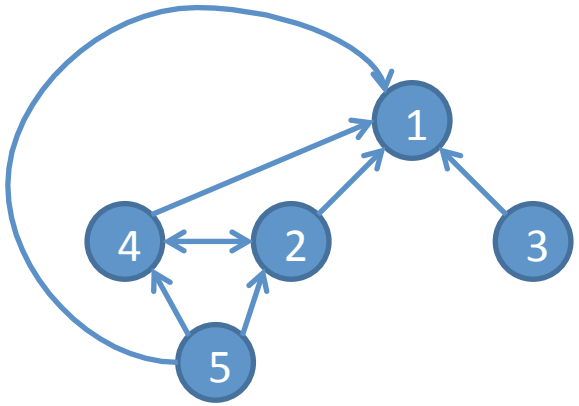
- Entailment graphs are often FRGs
  - Interesting linguistic observation
  - Leads to an efficient method for learning edges

Tree-based Approximation - Outline

- Forest-reducible graphs (FRG)
- Tree-Node-Fix algorithm (TNF)
- Experiments:
  - Typed entailment graph
  - Large untyped entailment graphs + resource

Sequential Approximation

```
Input: Nodes V and weighting function w
1. Initialize graph (e.g., empty graph)
2. For every node v in V:
   1. Re-attach v in a locally-optimal way
3. Until convergence
```

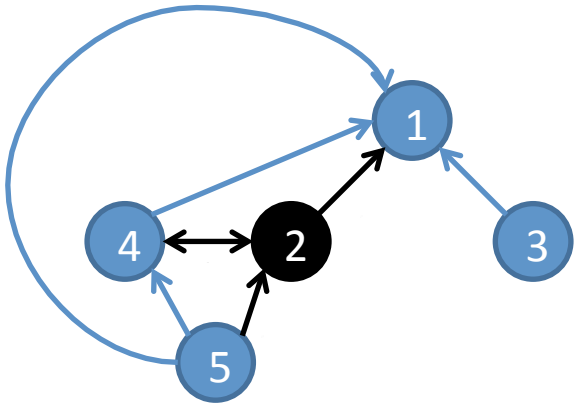


Sequential Approximation

```
Input: Nodes V and weighting function w
1. Initialize graph (e.g., empty graph)
2. For every node v in V:
   1. Re-attach v in a locally-optimal way
3. Until convergence
```

Candidate edges:

(1,**2**)
(**2**,1)
(**2**,3)
(3,**2**)
(**2**,4)
(4,**2**)
(**2**,5)
(5,**2**)



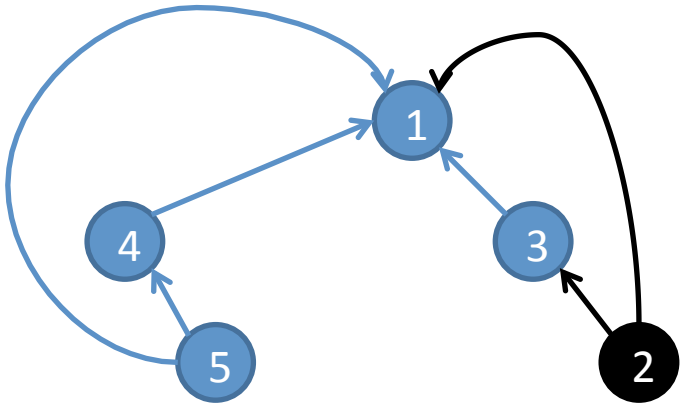
Reattach (2)

Sequential Approximation

```
Input: Nodes V and weighting function w
1. Initialize graph (e.g., empty graph)
2. For every node v in V:
   1. Re-attach v in a locally-optimal way
3. Until convergence
```

Candidate edges:

(1,**2**)
(**2**,1)
(**2**,3)
(3,**2**)
(**2**,4)
(4,**2**)
(**2**,5)
(5,**2**)

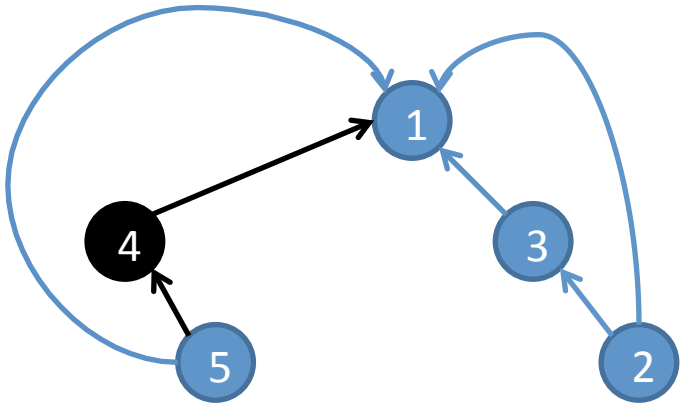


Sequential Approximation

```
Input: Nodes V and weighting function w
1. Initialize graph (e.g., empty graph)
2. For every node v in V:
   1. Re-attach v in a locally-optimal way
3. Until convergence
```

Candidate edges:

(1,4)
(4,1)
(2,4)
(4,2)
(3,4)
(4,3)
(4,5)
(5,4)



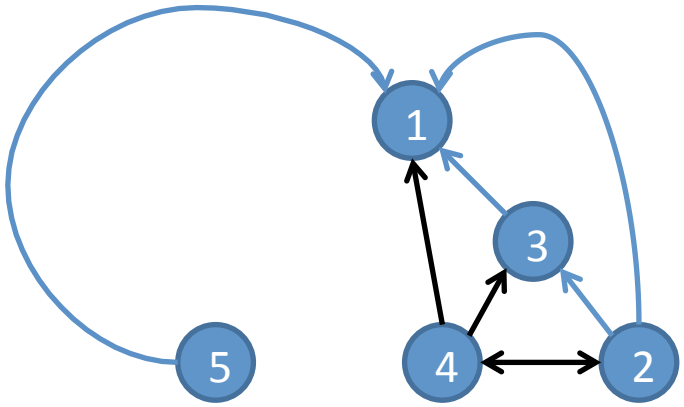
Reattach (4)

Sequential Approximation

```
Input: Nodes V and weighting function w
1. Initialize graph (e.g., empty graph)
2. For every node v in V:
   1. Re-attach v in a locally-optimal way
3. Until convergence
```

Candidate edges:

(1,4)
(4,1)
(2,4)
(4,2)
(3,4)
(4,3)
(4,5)
(5,4)



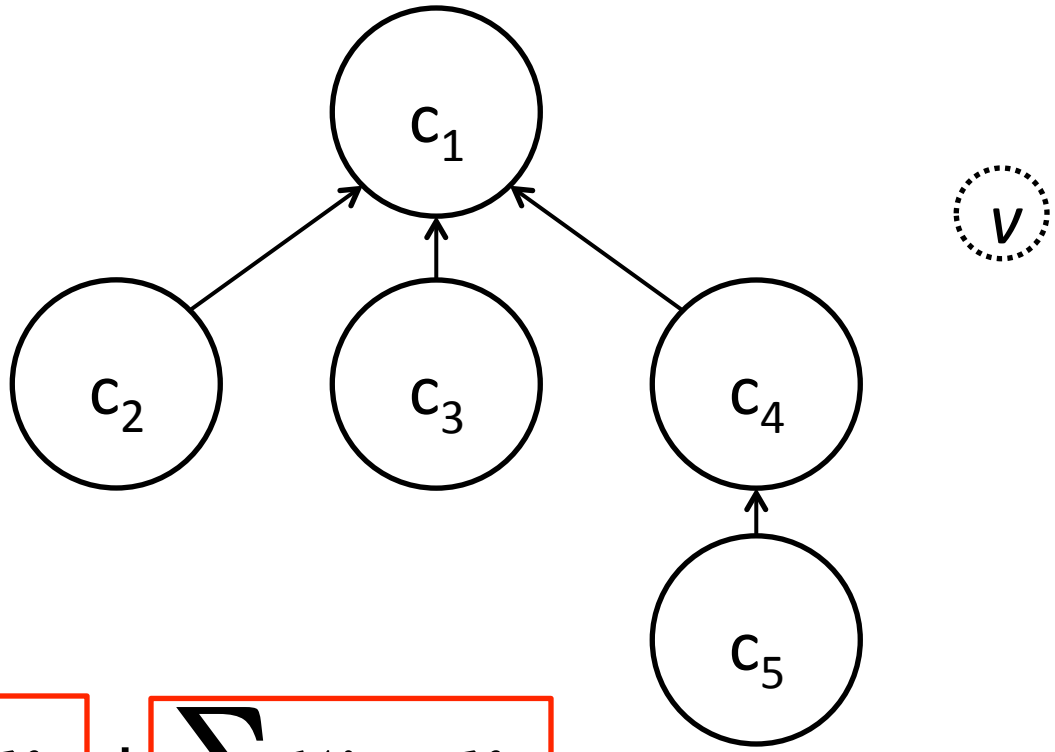
Tree-Node-Fix: Re-attach *v*

Input: Graph *G*, node *v* in *V*, function *w*

1. *Delete v*
2. Compute reduced graph G_{red} (directed forest)
3. Re-attach *v* in a **locally-optimal way**.

Node re-attachment in FRGs takes linear time!

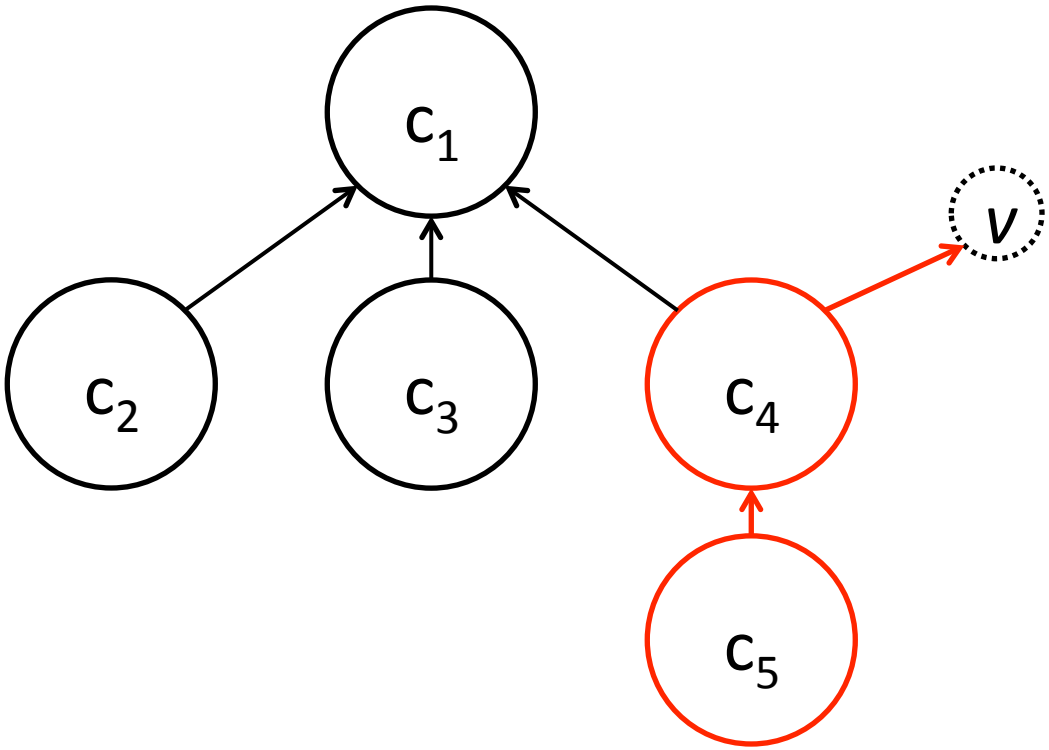
Objective Function



$$\arg\max \boxed{\sum_{i \neq v} w_{iv} \cdot x_{iv}} + \boxed{\sum_{k \neq v} w_{vk} \cdot x_{vk}}$$

Incoming edges Outgoing edges

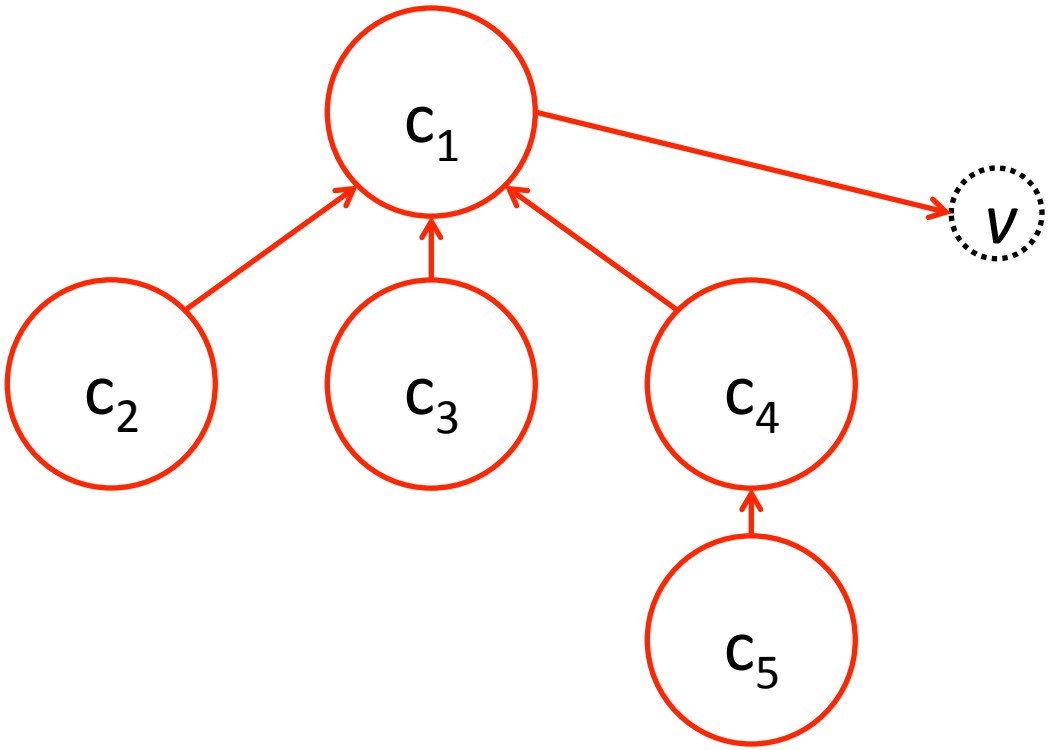
Scores for Components



$$S_{v-in}(c) = \sum_{i \in c} w_{iv} + \sum_{d \in child(c)} S_{v-in}(d)$$

$S_{v\text{-}in}(c_4)$

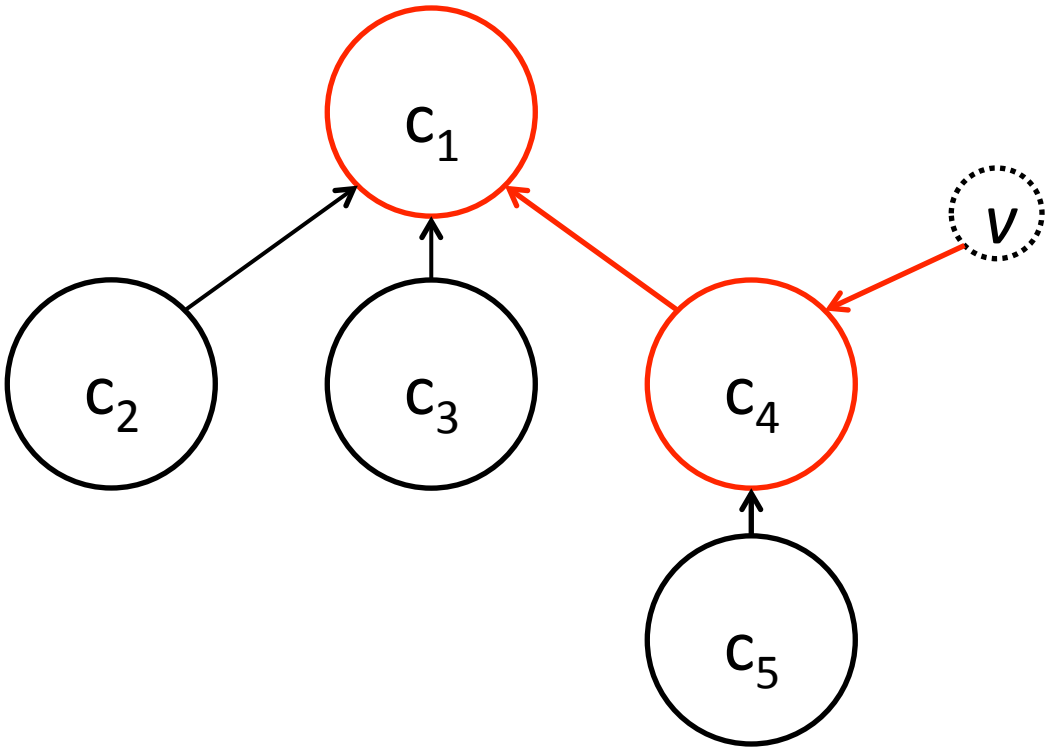
Definitions



$$S_{v-in}(c) = \sum_{i \in c} w_{iv} + \sum_{d \in child(c)} S_{v-in}(d)$$

$S_{v\text{-}in}(c_1)$

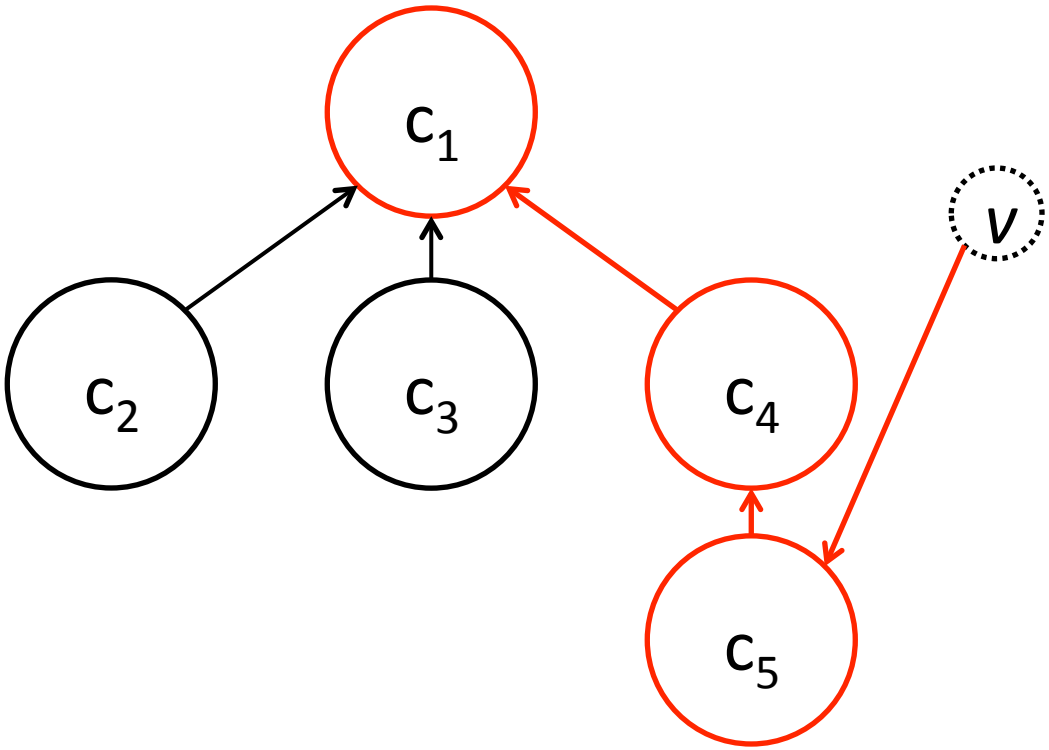
Definitions



$$S_{v-out}(c) = \sum_{i \in c} w_{vi} + S_{v-out}(parent(c))$$

$S_{v\text{-}out}(c_4)$

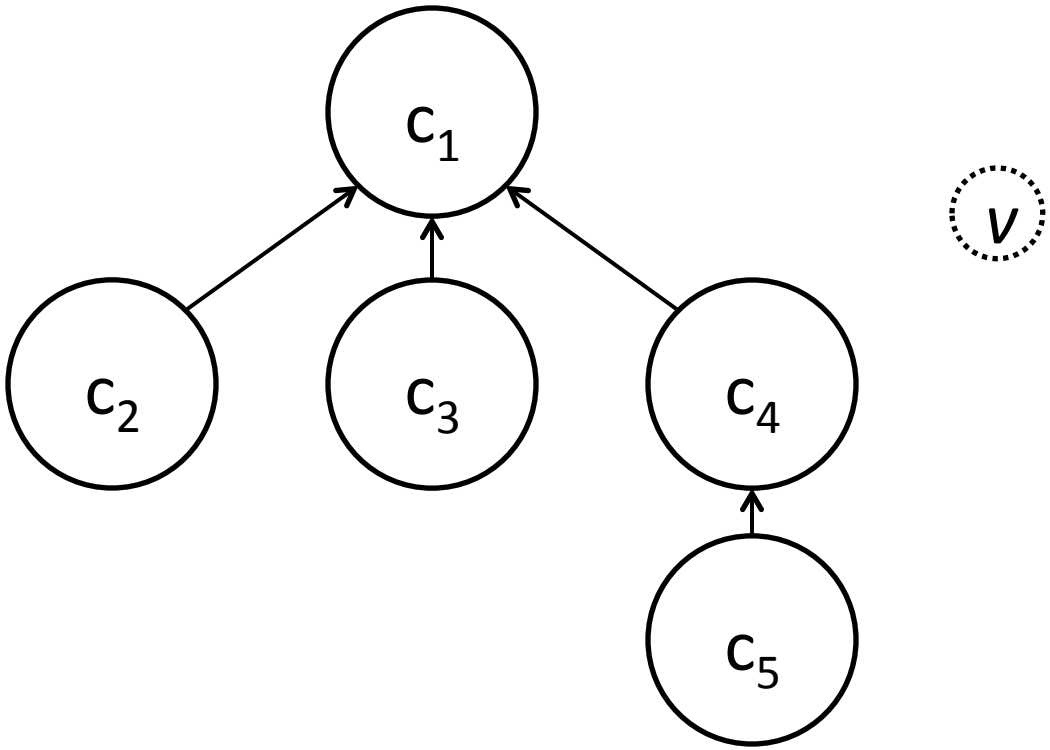
Definitions



$$S_{v-out}(c) = \sum_{i \in c} w_{vi} + S_{v-out}(parent(c))$$

$S_{v\text{-}out}(c_5)$

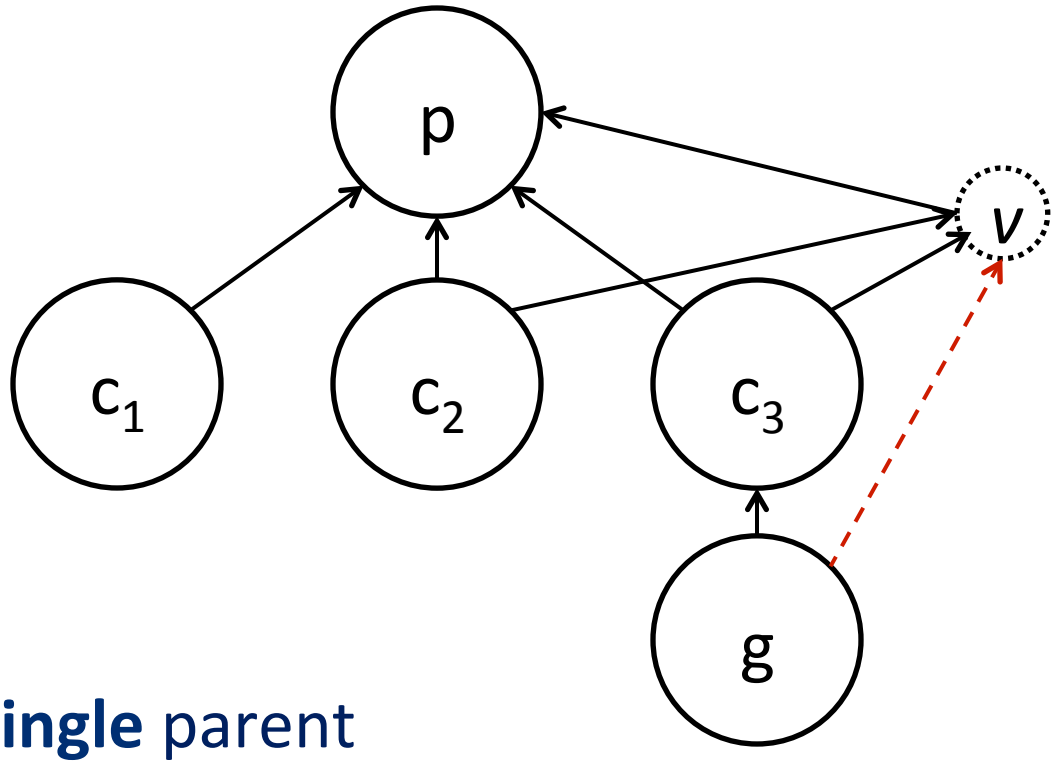
TNF – case 1



- Insertion into one of the components

$$S_{v-in}(c) + S_{v-out}(c)$$

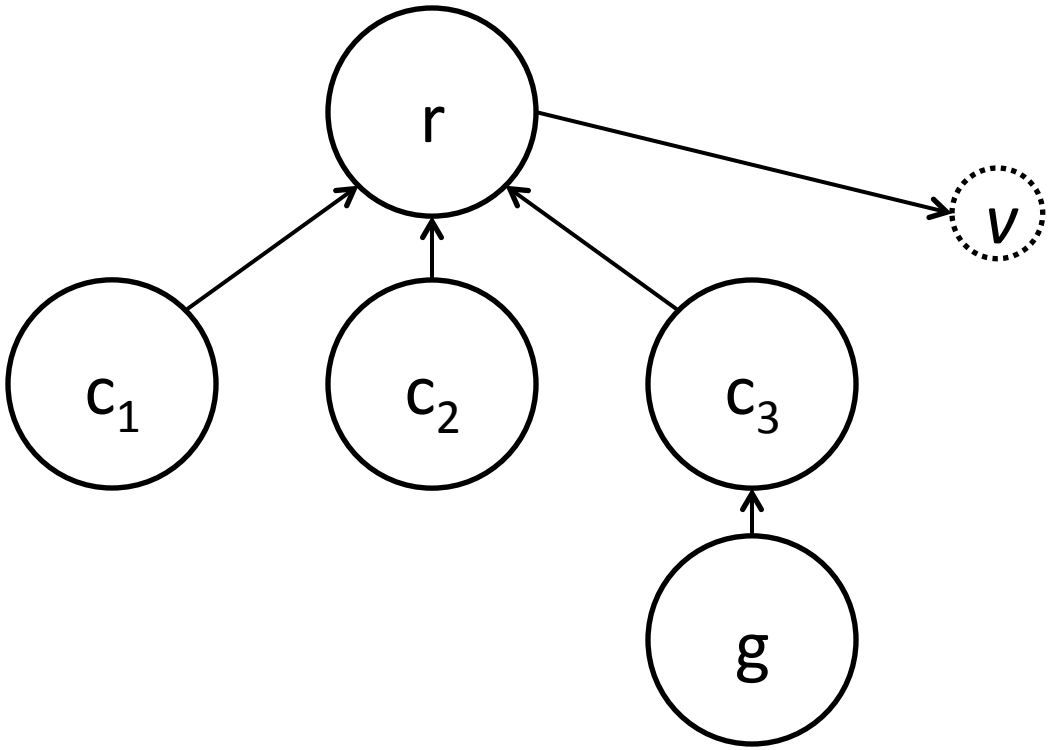
TNF – case 2



- Choose a **single** parent
- Only direct children of *p* can become children of *v*
- Once a parent is chosen, choose children **independently**:

$$S_{v-out}(p) + \sum_{c} \max(0, S_{v-in}(c))$$

TNF – case 3



- Choose no parent (forest root)
- Choose children **independently** from roots.

$$\sum_{r \in roots} \max(0, S_{v-in}(r))$$

TNF - Summary

- Computing $S_{v\text{-}in}(c)$ and $S_{v\text{-}out}(c)$ is linear (DP)
- Choosing best case is linear
- Re-attachment is linear
- TNF is quadratic
- Extension: Tree-Node-Component-Fix (TNCF)
  - Allow re-attachment of reduced graph components.

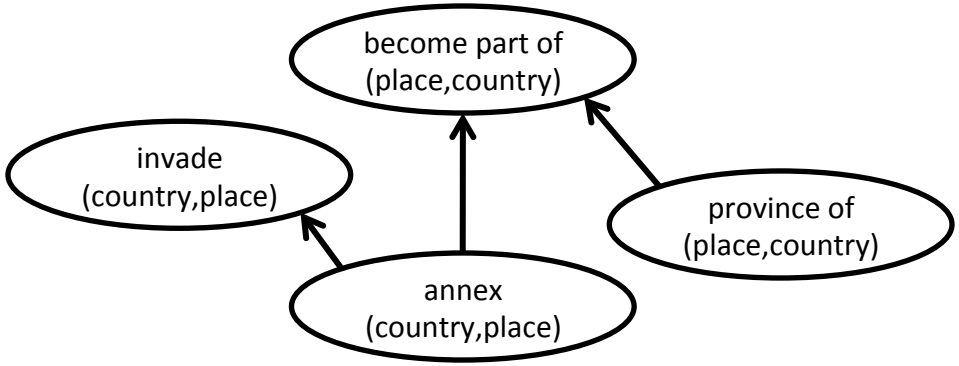
Complexity Results

	Transitive Graph	FRG
Entire graph	NP-hard	NP-hard
Node re-attachment	?	Linear

Outline

- Forest-reducible graphs (FRG)
- Tree-Node-Fix algorithm (TNF)
- Experiments:
  - Typed entailment graph
  - Large untyped entailment graphs + resource

Experiment 1: Typed Graphs

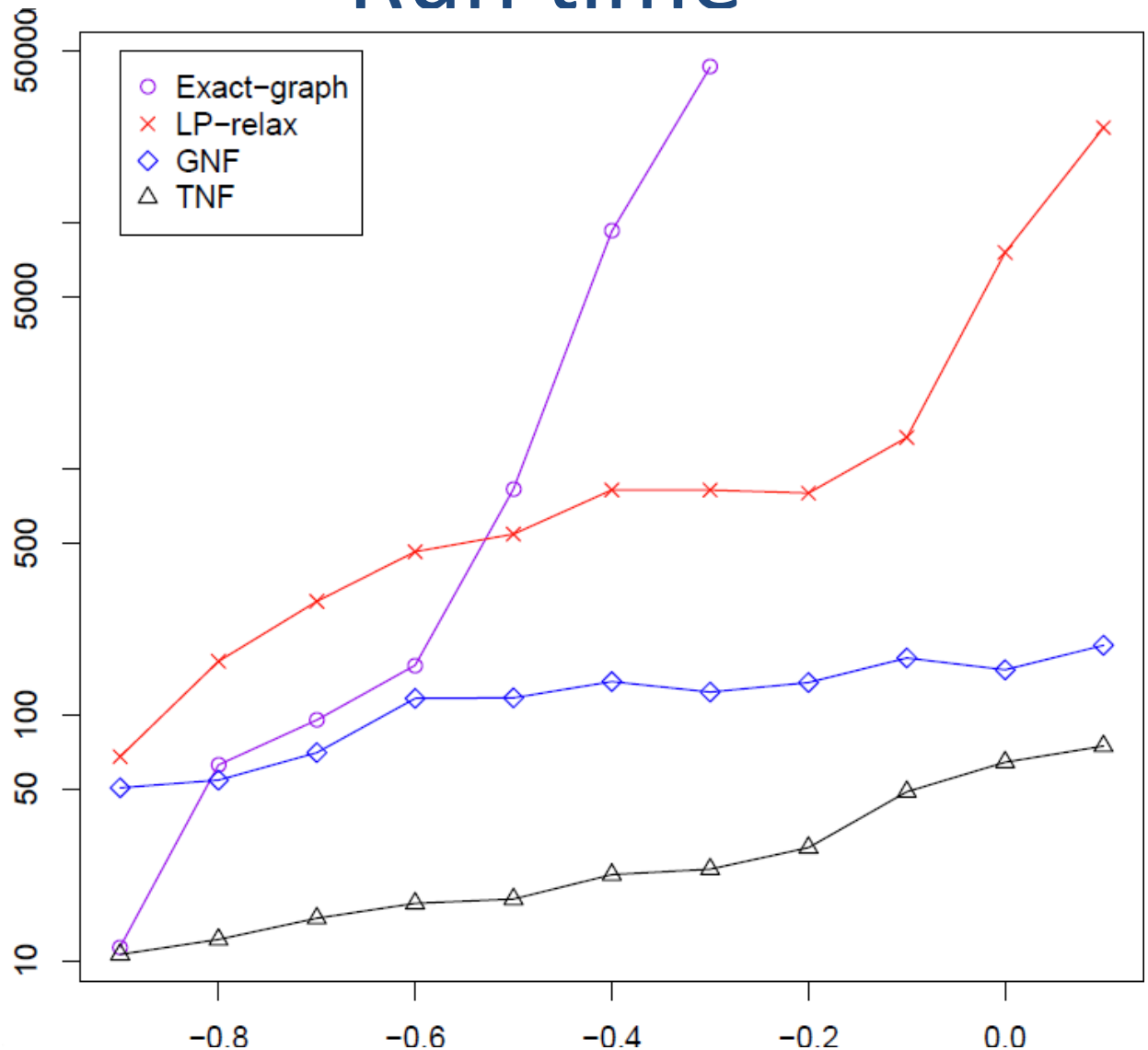


- ACL 11' experimental setting
  - Medium-size graphs
  - Predicates/nodes are disambiguated

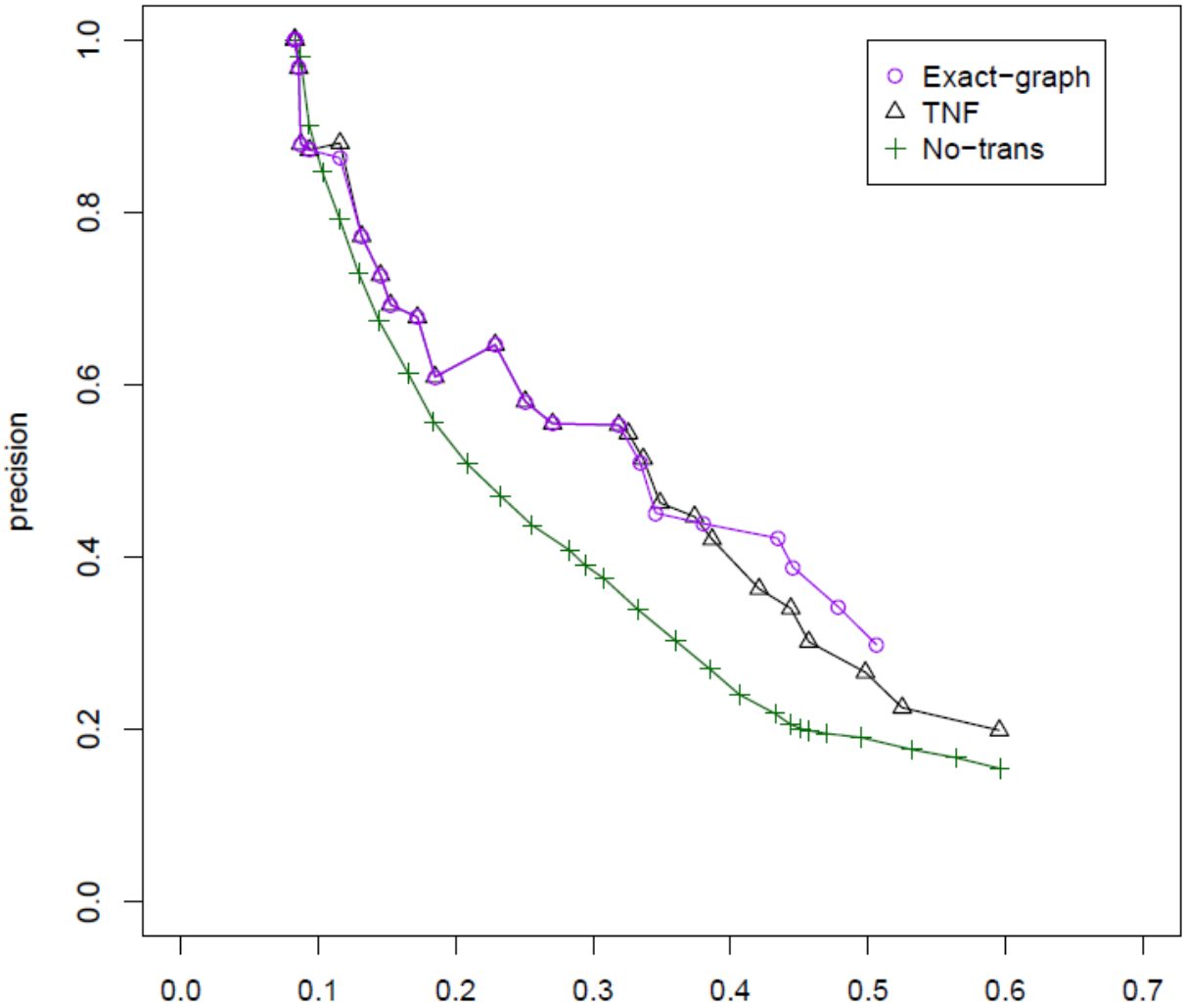
Evaluated Algorithms

- No-trans: no transitivity constraints (w_{ij})
- Exact-graph: ACL 11' exact method
- Exact-forest: Find optimal FRG with ILP solver
- LP-relax (Martins et al., 09')
- Tree-Node-Fix (TNF)
- Graph-Node-Fix (GNF)

Run time

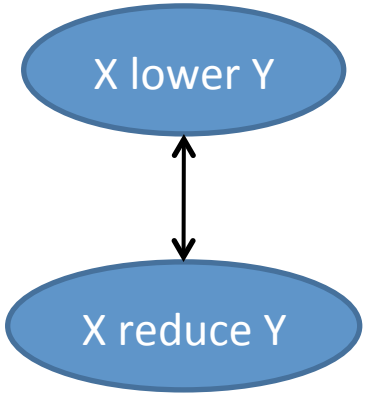

Exact−graph
LP−relax
GNF
TNF
50000
5000
500
100
50
10
−0.8
−0.6
−0.4
−0.2
0.0

Precision-recall Curve



Expriment 2: Untyped Graphs

- Large graph: 10,000 predicates
- Based on REVERB (fader et al., 11')
- Predicates are ambiguous
  - Transitivity and FRG assumption
- Features: distributional similarity, lexicographic and string-similarity features for every pair of predicates



Training and Test Set Generation

- Gold standard annotation:
  - Instance-based evaluation with crowdsourcing (Zeichner et al., 2012)
  - Training set: ~3,300 pairs of predicates
  - Test set: ~1,750 pairs of predicates
- Trained classifier provides w_{ij}
- **We managed to run only No-trans, TNF, and TNCF**

Experiment 2 - Results

Method	Recall	Precision	F1
No-trans	0.14	0.7	0.24
TNF	0.15	0.7	0.25
TNCF	0.16	0.8	0.27

- FRG and Transitivity assumptions limit recall
- TNCF – improved precision for recall ~ 0.15
- ***High precision resource of > 100,000 entailment rules***

Resource

- High-precision resource of > 100,000 entailment rules
- In addition, 15 million rules learned over 100,000 predicates
  - Outperforms DIRT (Lin and Pantel, 2001)
- Syntactic and lexical representations
- http://u.cs.biu.ac.il/~nlp/downloads/

irsrv2.cs.biu.ac.il:8080/exploration/

BIU Health-care Exploration System

Explore your search results by drilling down and up the medical concepts, their semantic relations, and their mentions in the text.

The following 22 concepts are currently supported: *alcohol, asthma, biopsy, brain, cancer, CDC, chemotherapy, chest, cough, diarrhea, FDA, headache, HIV, HPV, lungs,*

---

headache [Explore]

- associate _ with headache | associate headache with _ (287)
- _ experience headache | _ have headache | _ suffer from headache (82)
- headache accompany _ (59)
- _ treat symptom of headache (18)
  - _ treat headache (16)
    - _ relieve headache (5)
      - _ reduce headache (1)
    - _ reduce headache (1)
- symptom of _ poisoning include headache (23)
- _ accompany headache (20)
- headache common in _ (8)
- _ prevent headache (7)

Drug, Chemical or Other Substance (7)

Test or Procedure (3)

Occupation or Discipline (2)

Behavior or Activities (1)

Disease or Natural Phenomenon or Process (1)

Food (1)

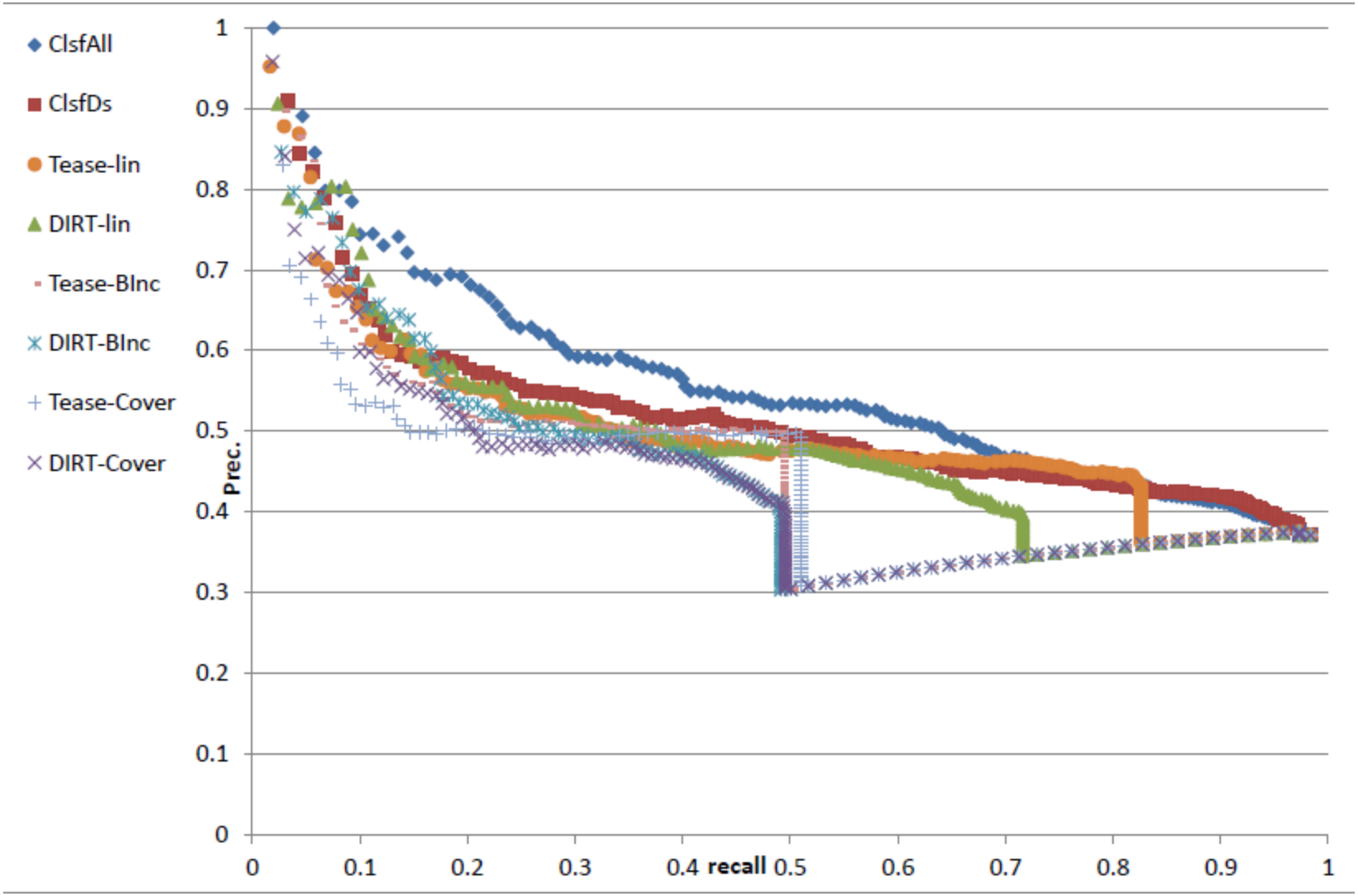
high blood pressure (1)

tea (1)

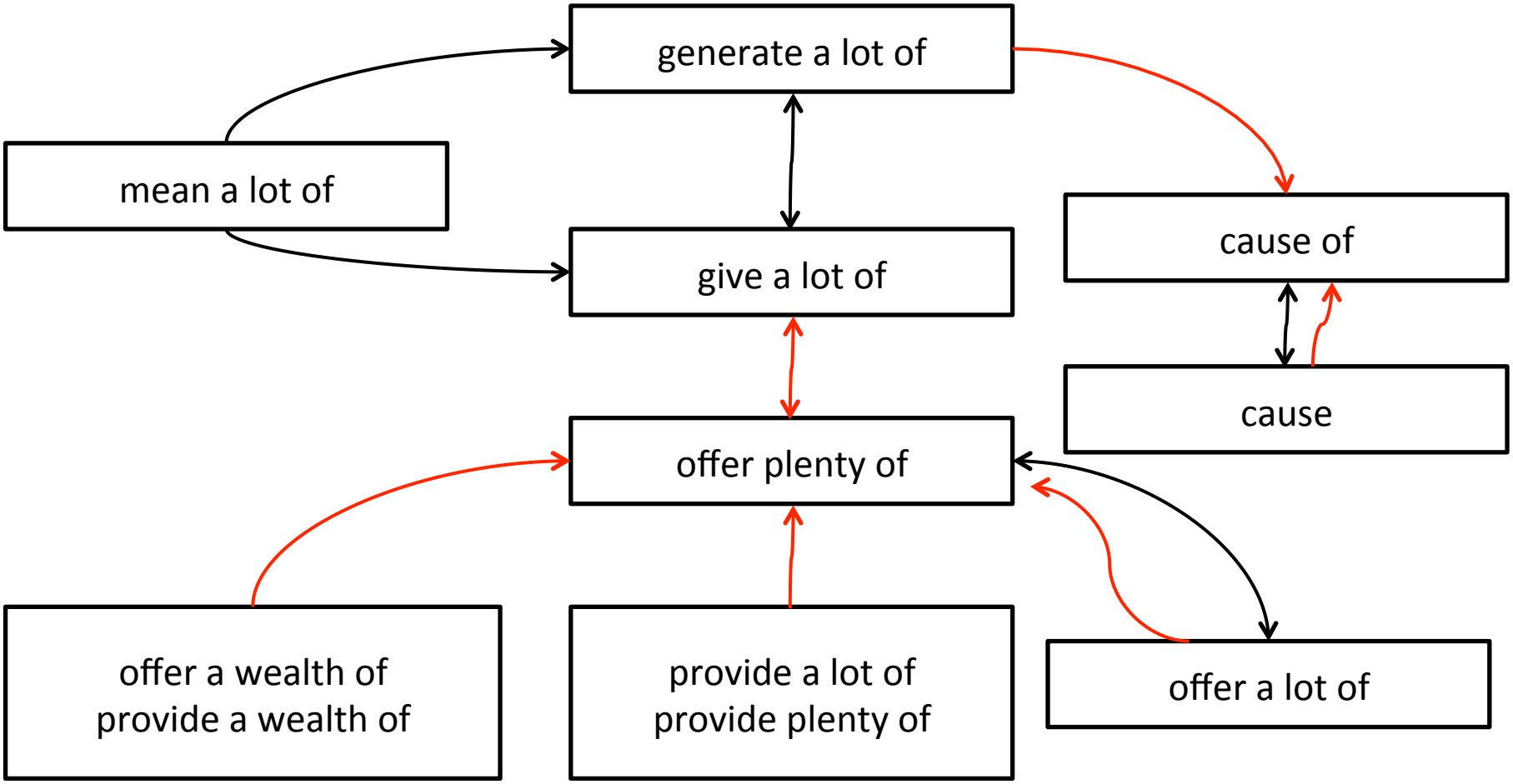
Mayo Clinic - Tips for Women (and Men) in Search of a Good Night's Sleep

In addition to the usual sources coffee **tea** soda be aware of caffeine in chocolate in medications used to **treat headaches** colds sinus_congestion

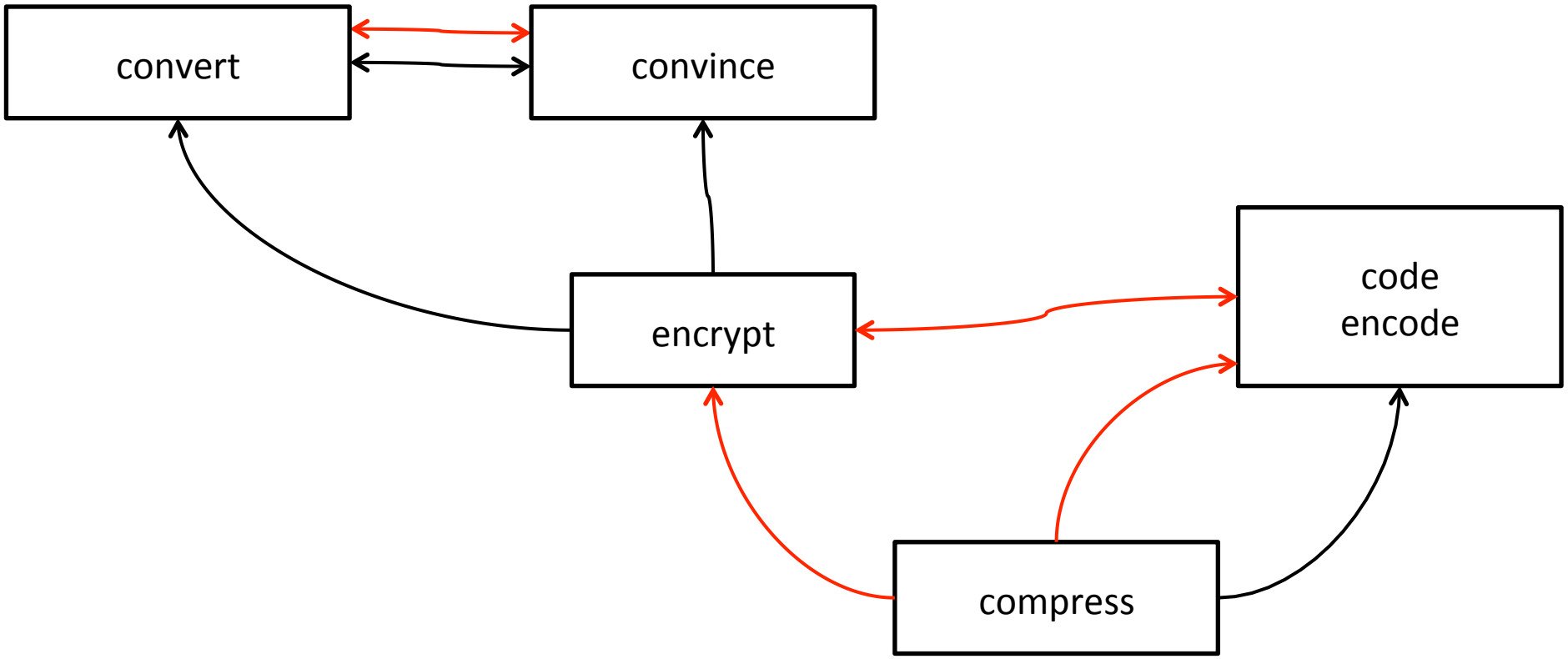
Local Results



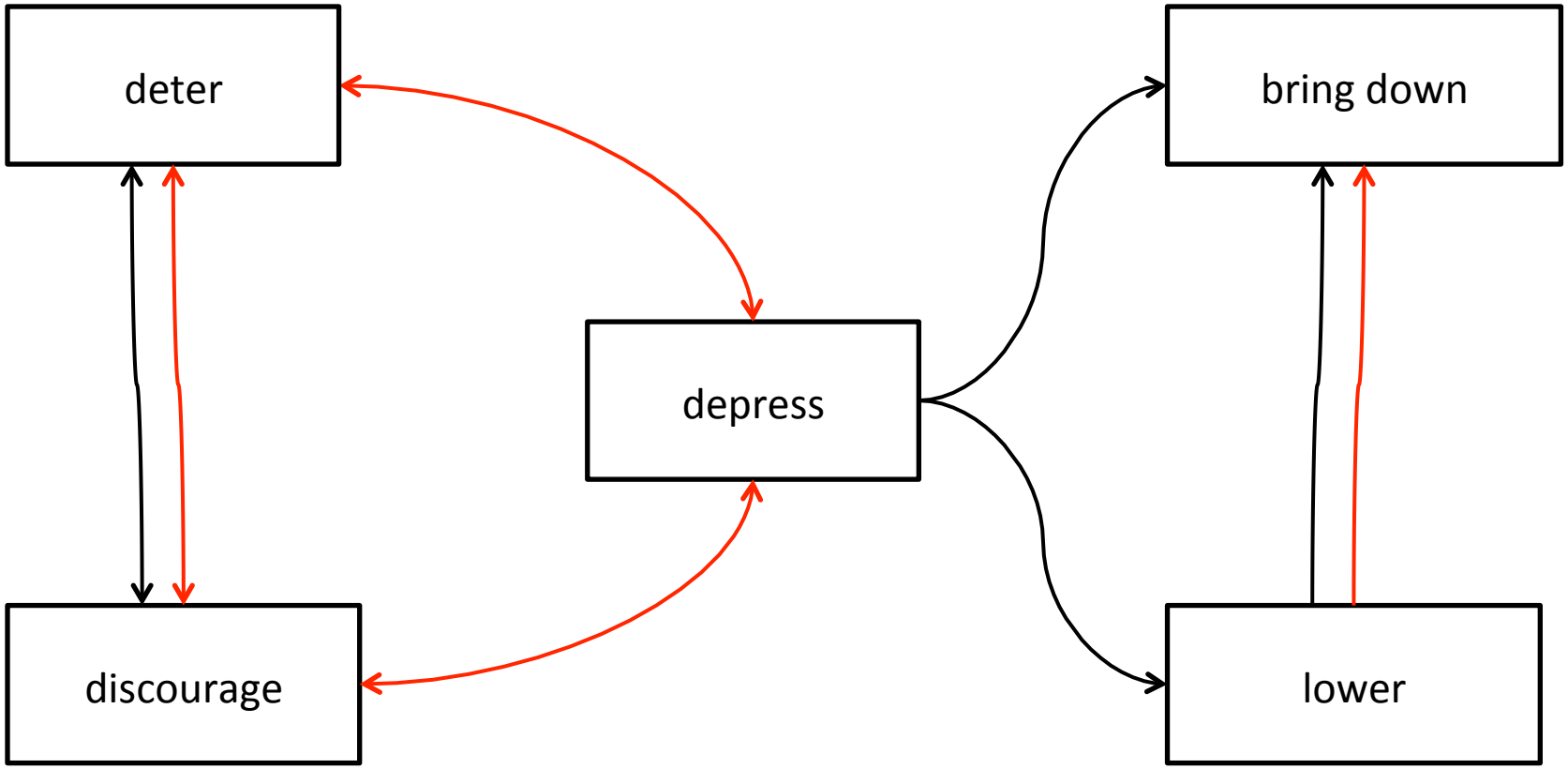
Re-attachment Examples



Re-attachment Examples



Re-attachment Examples



Conclusions

- NLU Applications require large knowledge resources of predicative entailment rules
- In this dissertation we presented
  - A model that utilizes graph structure to improve rule learning
  - Algorithms that scale for learning entailment graphs
  - A large knowledge resource of rules
  - An application for text exploration
- Future work: ambiguity

Thanks for coming!