Expert Finding using Markov Networks in Open Source Communities

Matthieu Vergne[1,2], Angelo Susi[1]
vergne@fbk.eu, susi@fbk.eu

[1]Center for Information and Communication Technology
Fondazione Bruno Kessler

[2]Doctoral School in Information and Communication Technology
University of Trento

CAISE - June 2014

Outline

Expert Perspectives

Expert (Ericsson [2006])

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge
- Extended skills

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge
- Extended skills
- Recognised expert status

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge
- Extended skills
- Recognised expert status

2 perspectives:

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge (1)
- Extended skills (1)
- Recognised expert status

2 perspectives:

1. **content-based** : One has special skills or knowledge, whether people assess them or not

Expert Perspectives

Expert (Ericsson [2006])

- Extended knowledge (1)
- Extended skills (1)
- Recognised expert status (2)

2 perspectives:

1. **content-based** : One has special skills or knowledge, whether people assess them or not
2. **social-based** : One shows special skills or knowledge, whether he actually has them or not

OSS and Hybrid Contexts

In OSS projects, both users and companies are involved:

- Users' questions/answers from forums $\rightarrow$ content-based

OSS and Hybrid Contexts

In OSS projects, both users and companies are involved:

- Users' questions/answers from forums $\rightarrow$ content-based
- Companies structures identifying expertises $\rightarrow$ social-based

M. Vergne, A. Susi (vergne@fbk.eu, susi@fbk.eu) FBK - ICT Doctoral School Trento

OSS and Hybrid Contexts

In OSS projects, both users and companies are involved:

- Users' questions/answers from forums → content-based
- Companies structures identifying expertises → social-based

But lacks information due to:

- Too much forum participants to recommend → social-based

OSS and Hybrid Contexts

In OSS projects, both users and companies are involved:

- Users' questions/answers from forums → content-based
- Companies structures identifying expertises → social-based

But lacks information due to:

- Too much forum participants to recommend → social-based
- All employees do not participate in forums → content-based

OSS and Hybrid Contexts

In OSS projects, both users and companies are involved:

- Users' questions/answers from forums → content-based
- Companies structures identifying expertises → social-based

But lacks information due to:

- Too much forum participants to recommend → social-based
- All employees do not participate in forums → content-based

For huge OSS projects:

Perspectives need to be considered together.

General Idea

Our approach:

- Goal: improve expert finding in RE by combining content-based and social-based perspectives.

General Idea

Our approach:

- Goal: improve expert finding in RE by combining content-based and social-based perspectives.
- Concepts: stakeholders, roles, topics and terms
  - *Castro-Herrera and Cleland-Huang [2010], Lim et al. [2010]*

General Idea

Our approach:

- Goal: improve expert finding in RE by combining content-based and social-based perspectives.
- Concepts: stakeholders, roles, topics and terms
  - *Castro-Herrera and Cleland-Huang [2010], Lim et al. [2010]*
- Relations: evidences extracted from available sources of data

General Idea

Our approach:

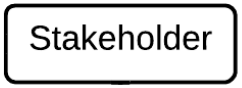
- Goal: improve expert finding in RE by combining content-based and social-based perspectives.
- Concepts: stakeholders, roles, topics and terms
  - *Castro-Herrera and Cleland-Huang [2010], Lim et al. [2010]*
- Relations: evidences extracted from available sources of data
- Computation: Markov networks (expertise probability)

General Idea

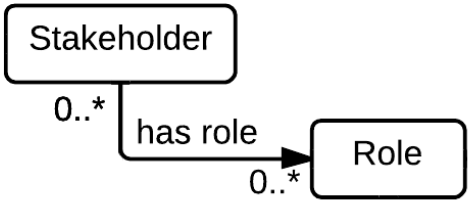
Our approach:

- Goal: improve expert finding in RE by combining content-based and social-based perspectives.
- Concepts: stakeholders, roles, topics and terms
  - *Castro-Herrera and Cleland-Huang [2010], Lim et al. [2010]*
- Relations: evidences extracted from available sources of data
- Computation: Markov networks (expertise probability)
- Outcome: ranking of potential experts to recommend

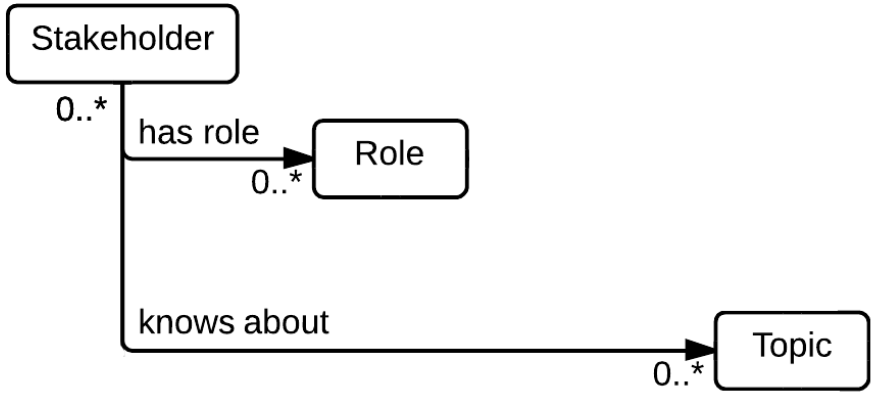
Concepts and Relations



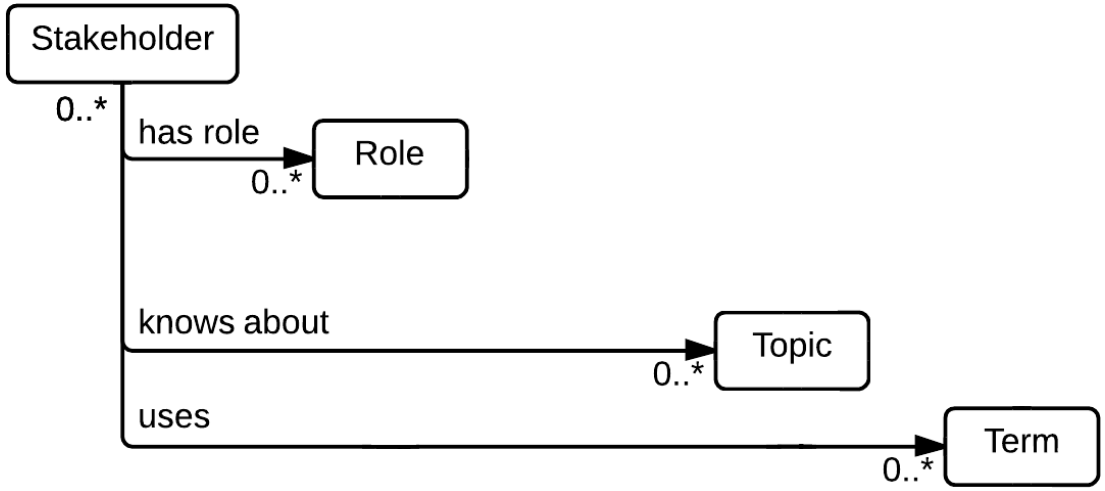
Concepts and Relations



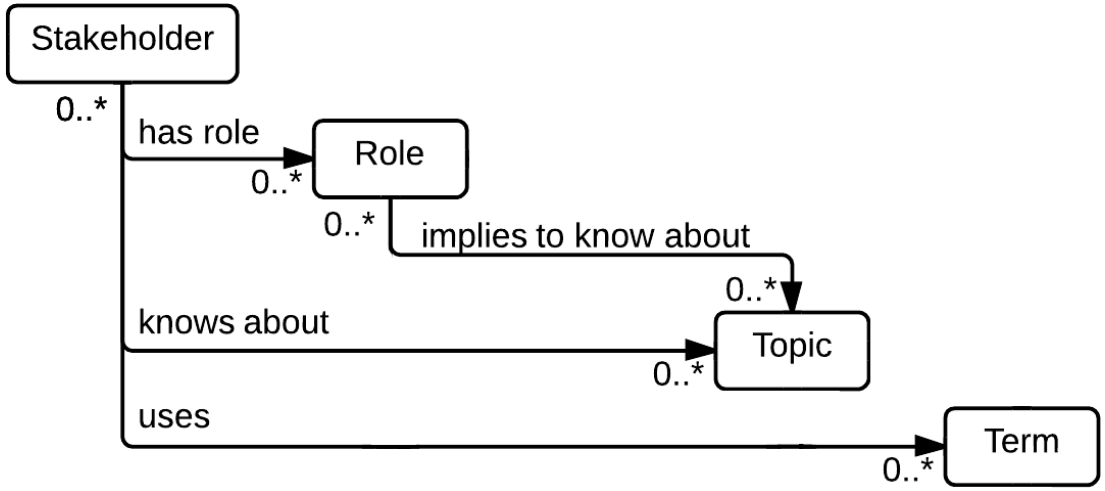
Concepts and Relations



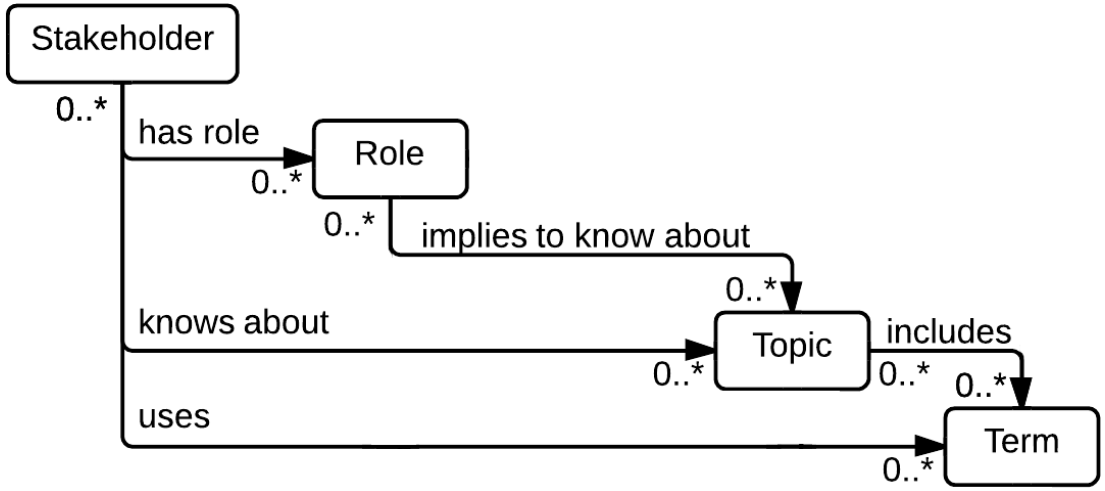
Concepts and Relations



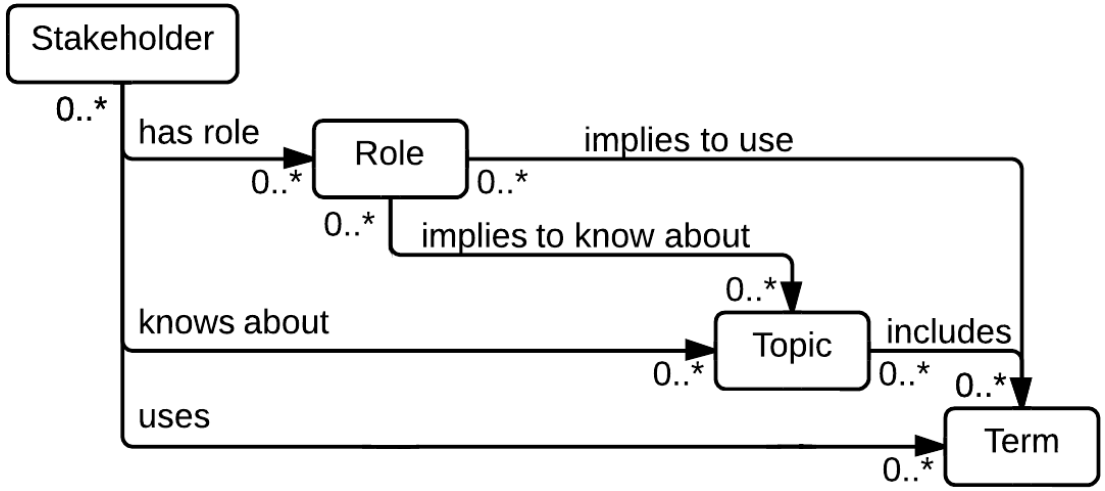
Concepts and Relations



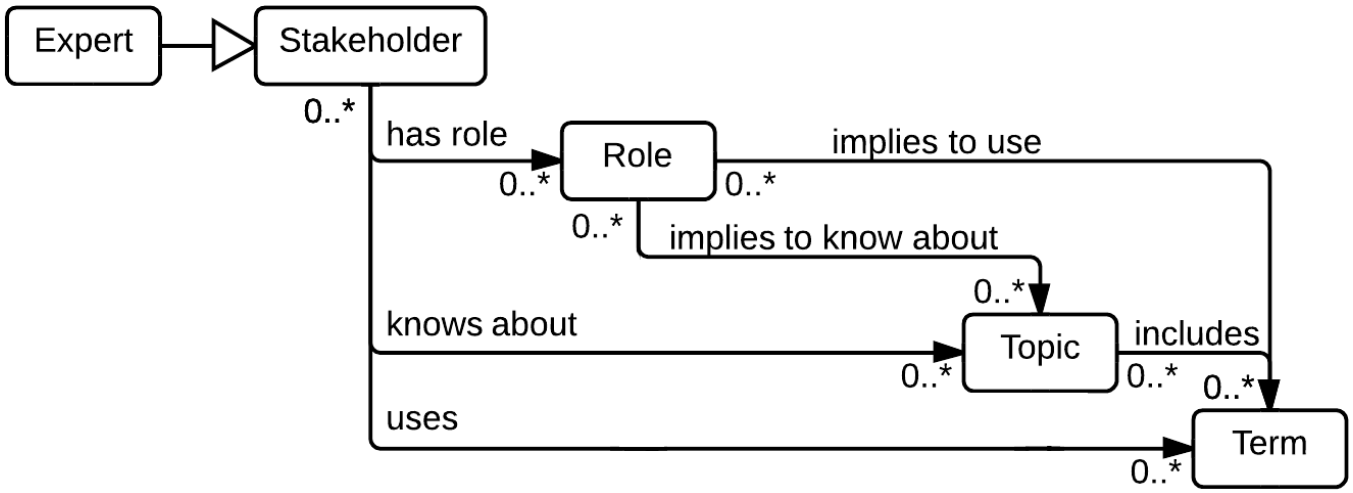
Concepts and Relations



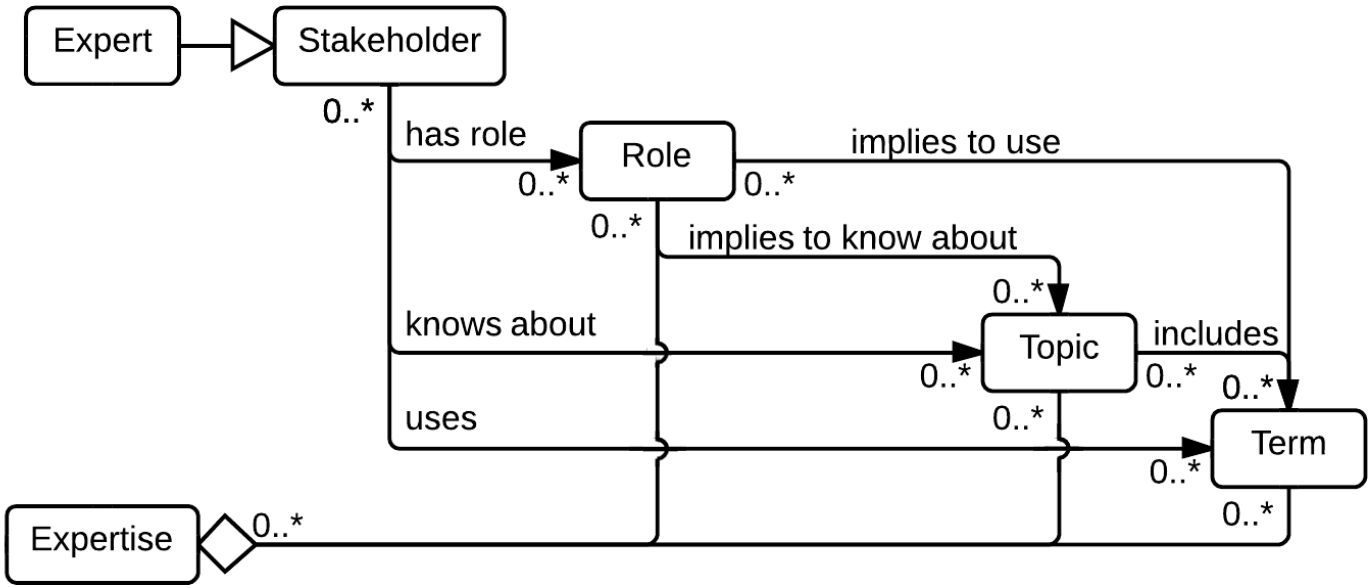
Concepts and Relations



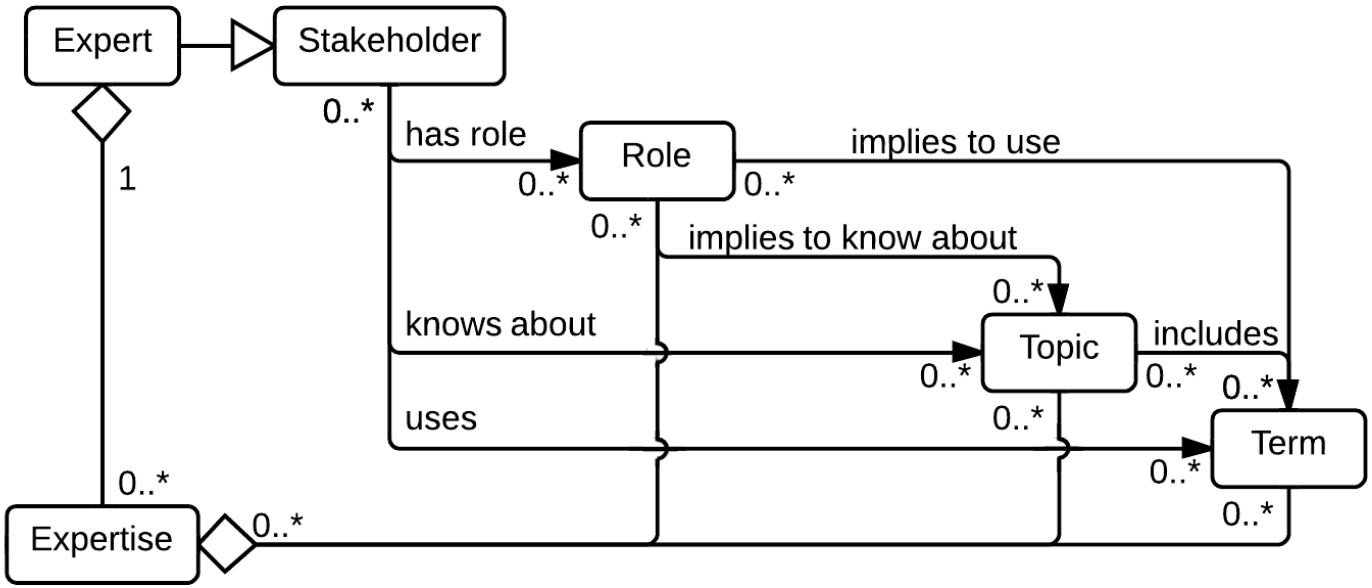
Concepts and Relations



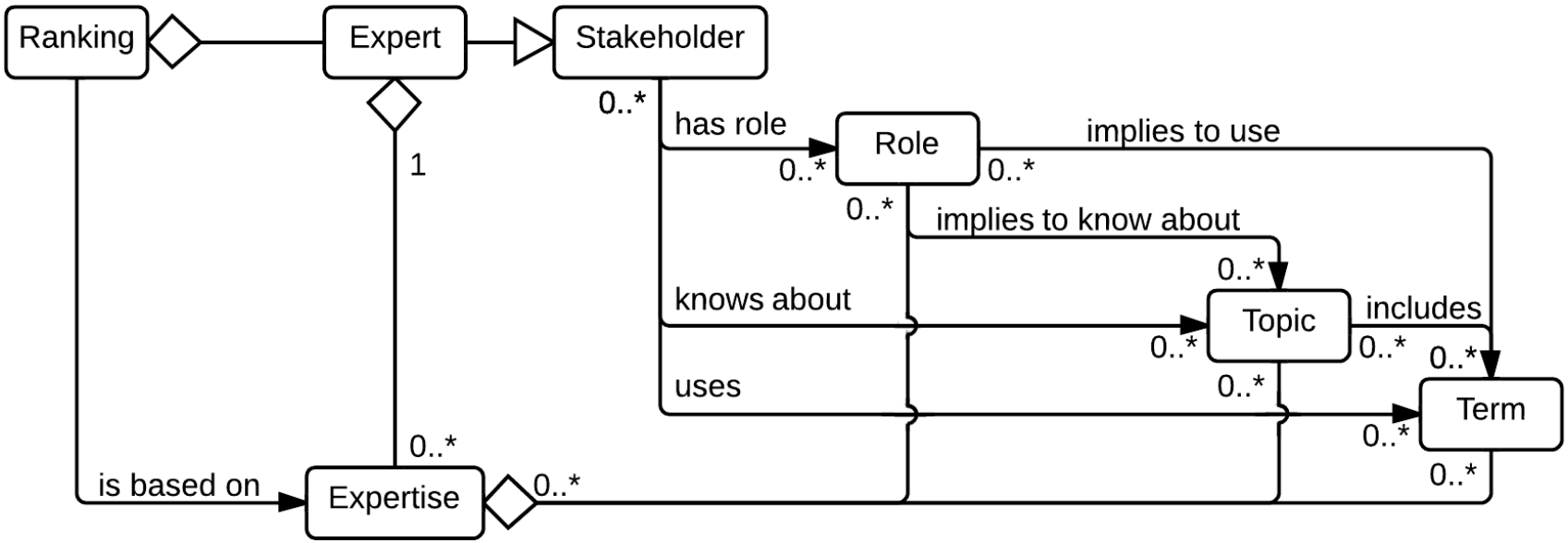
Concepts and Relations



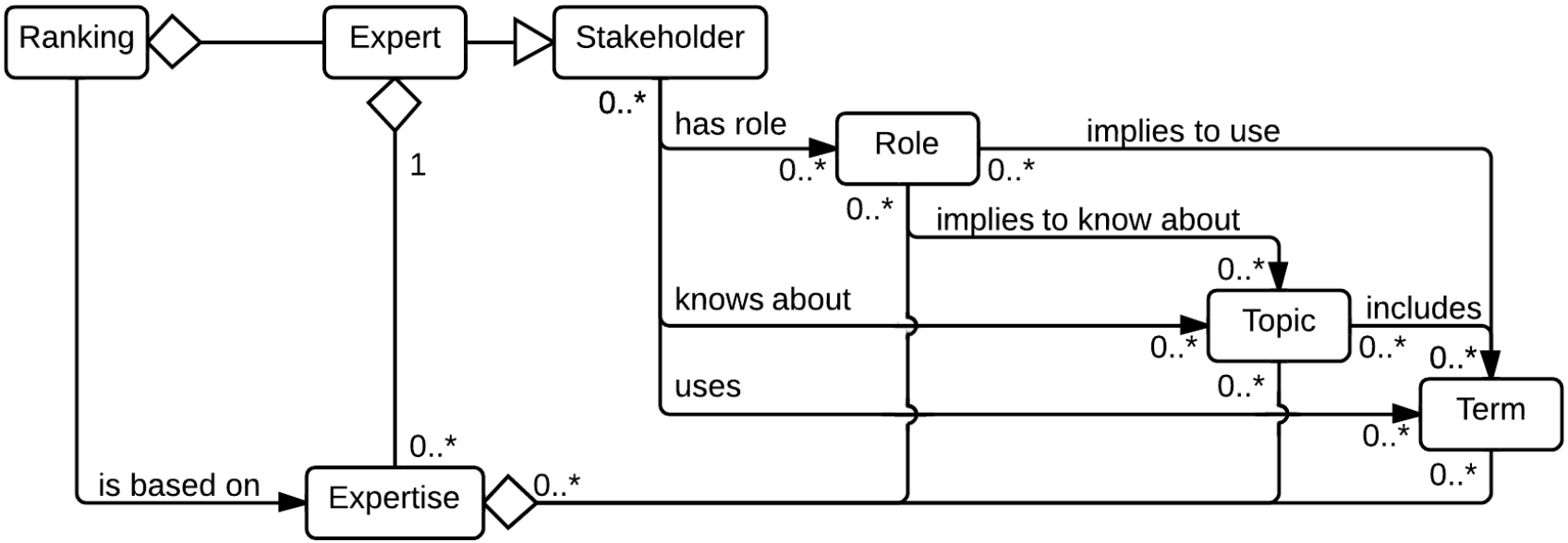
Concepts and Relations



Concepts and Relations



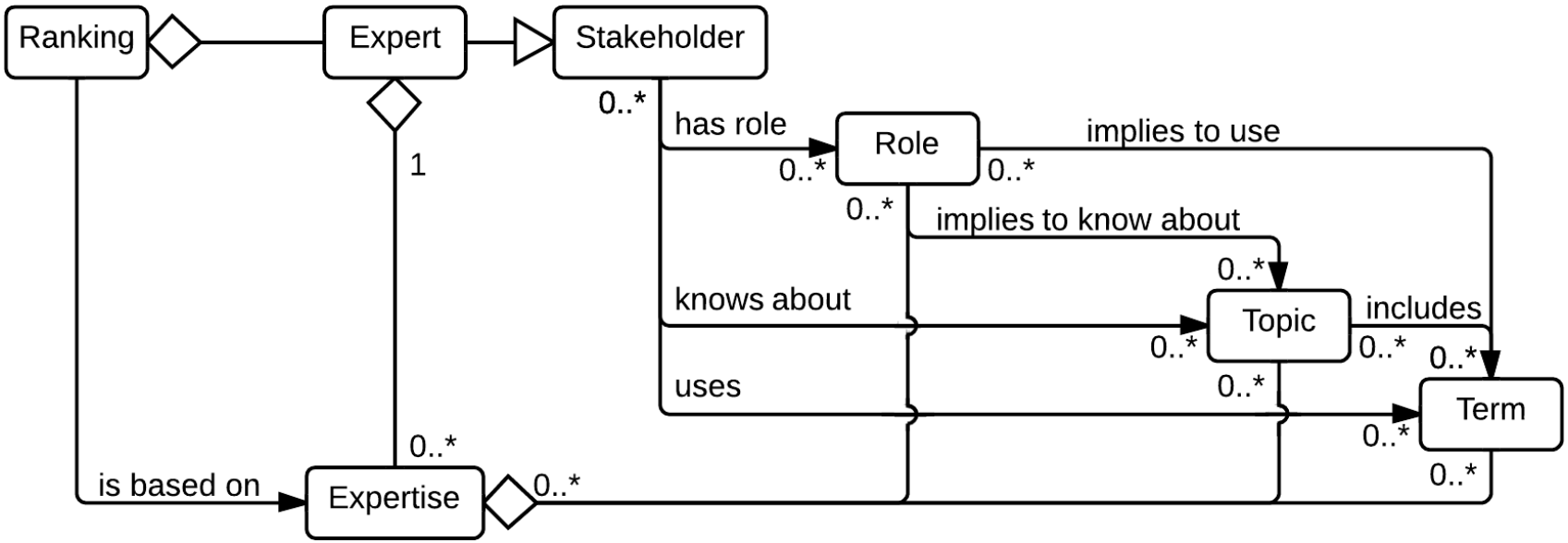
Concepts and Relations



Remarks:

- Directed relations for clarity: undirected correlations.

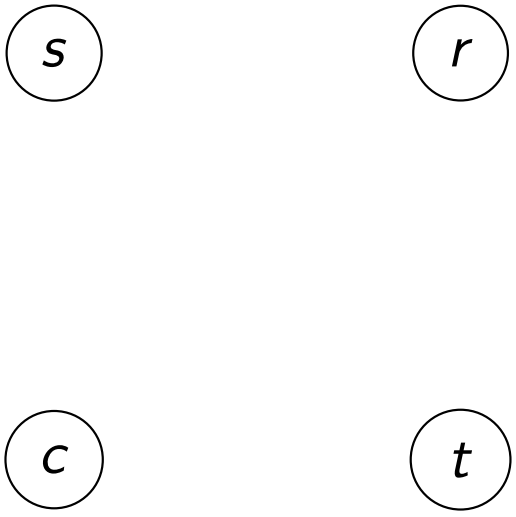
Concepts and Relations



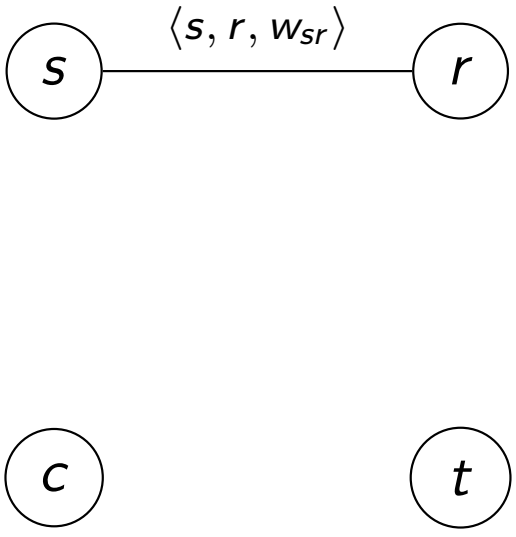
Remarks:

- Directed relations for clarity: undirected correlations.
- Relative expertise: $A > B$, not $A = \mathit{level}$.

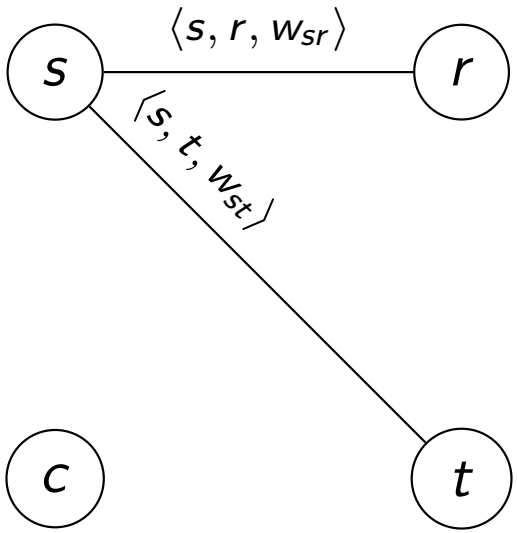
Model: Complete 4-Partite Weighted Graph



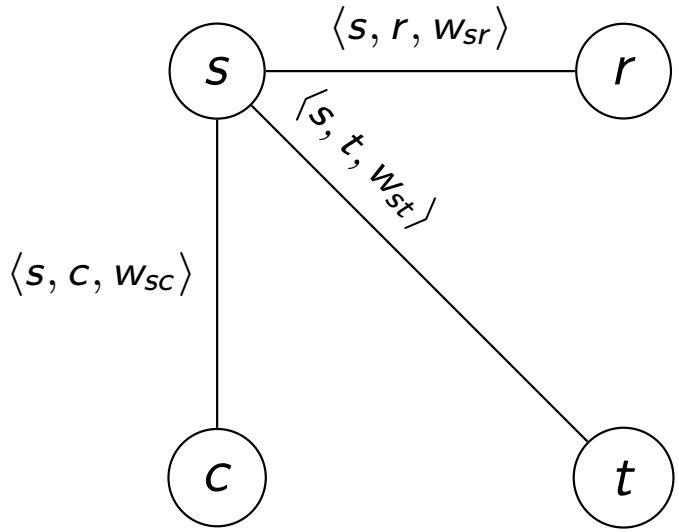
Model: Complete 4-Partite Weighted Graph



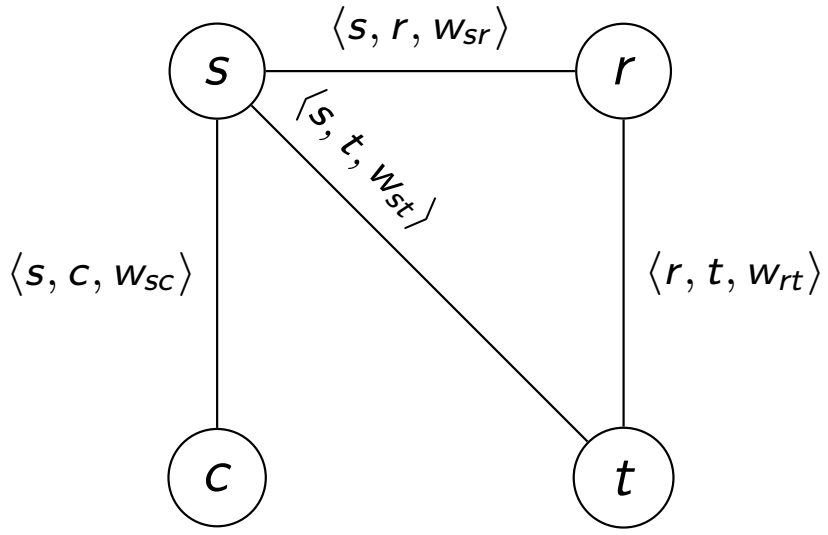
Model: Complete 4-Partite Weighted Graph



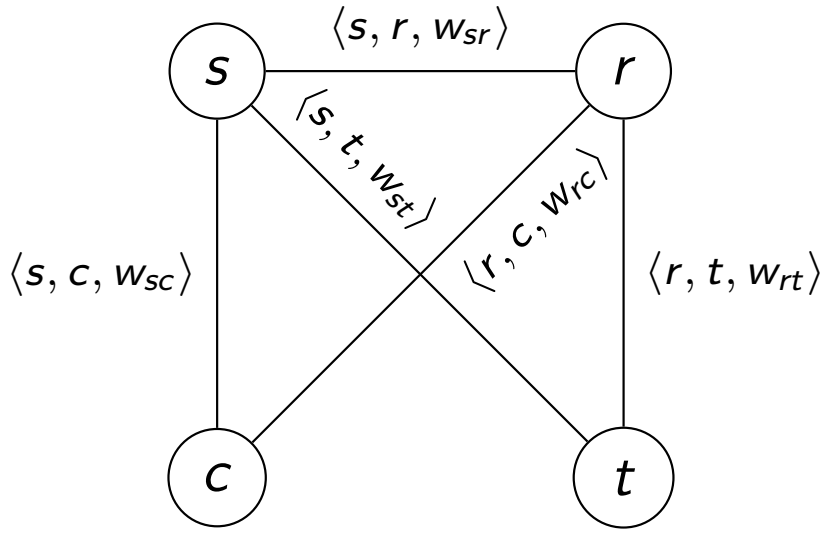
Model: Complete 4-Partite Weighted Graph



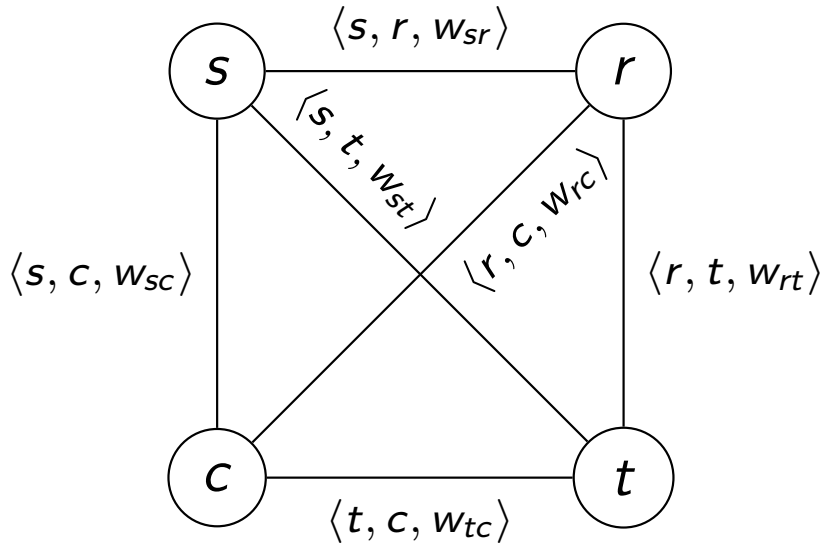
Model: Complete 4-Partite Weighted Graph



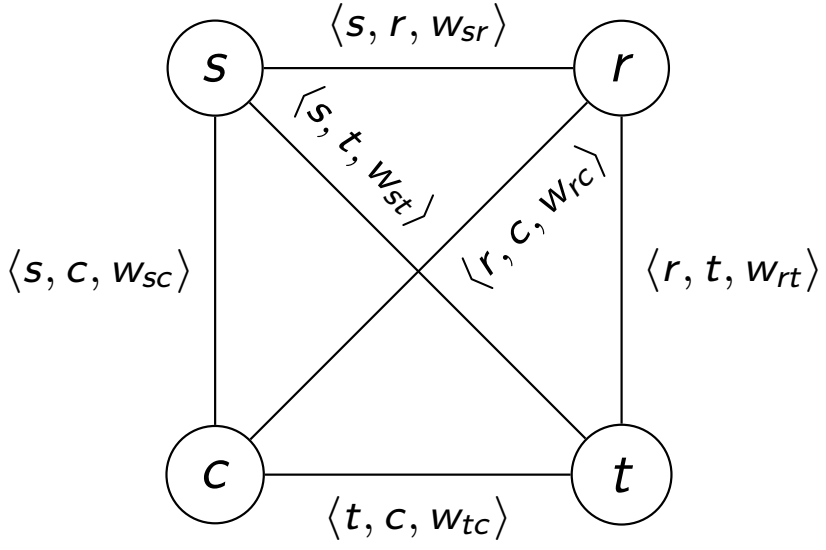
Model: Complete 4-Partite Weighted Graph

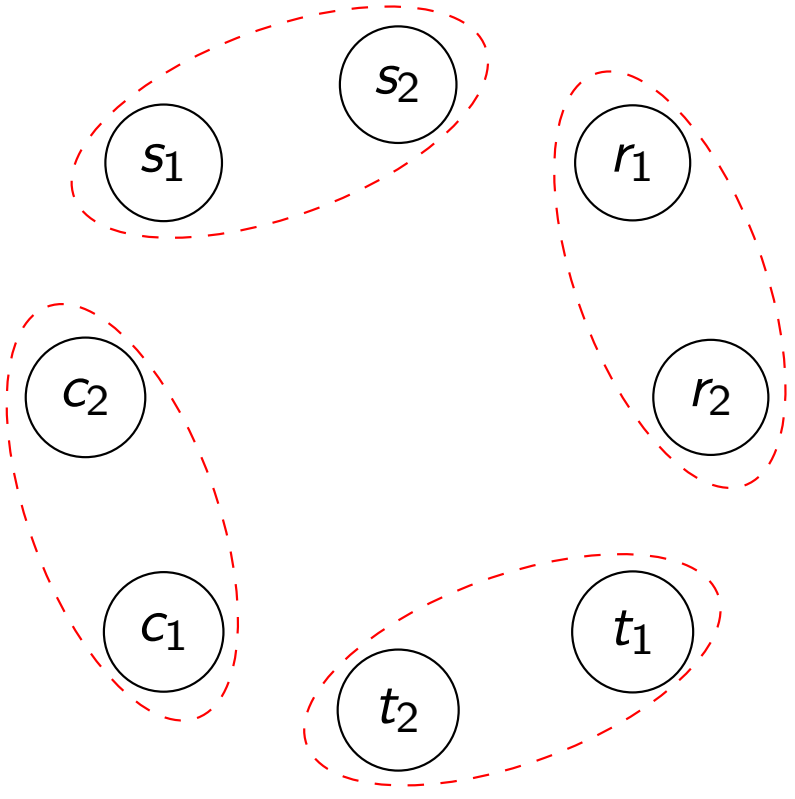


Model: Complete 4-Partite Weighted Graph

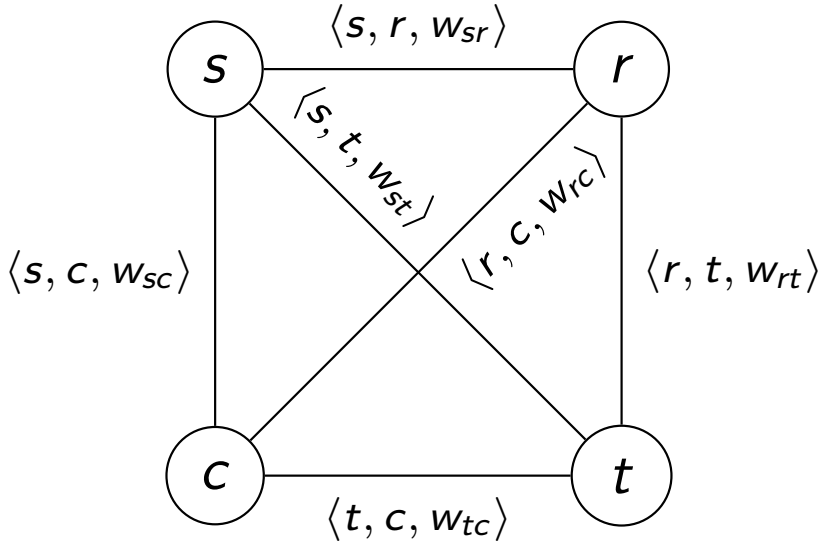


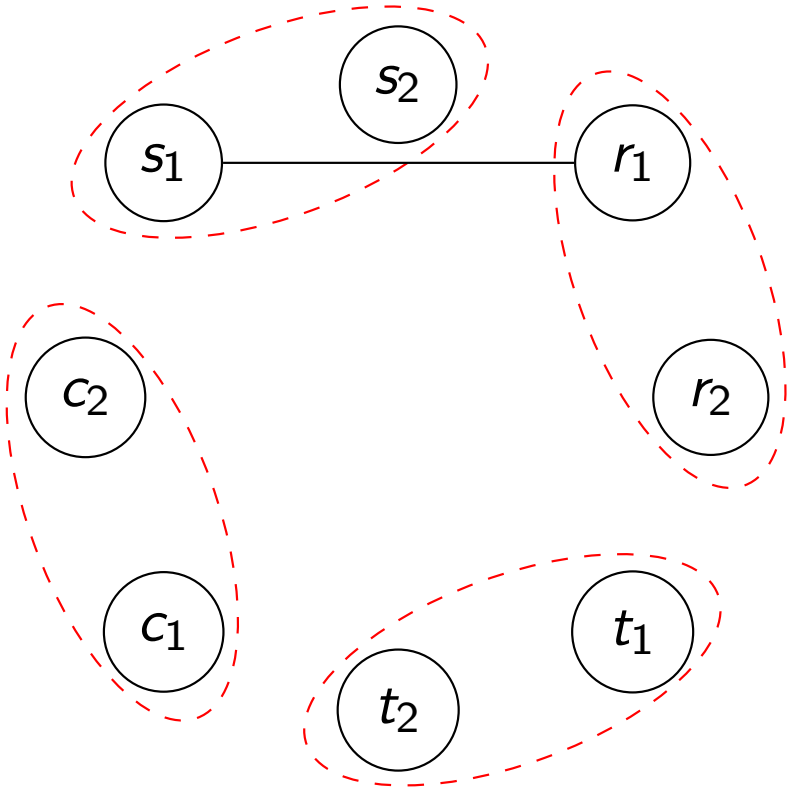
Model: Complete 4-Partite Weighted Graph



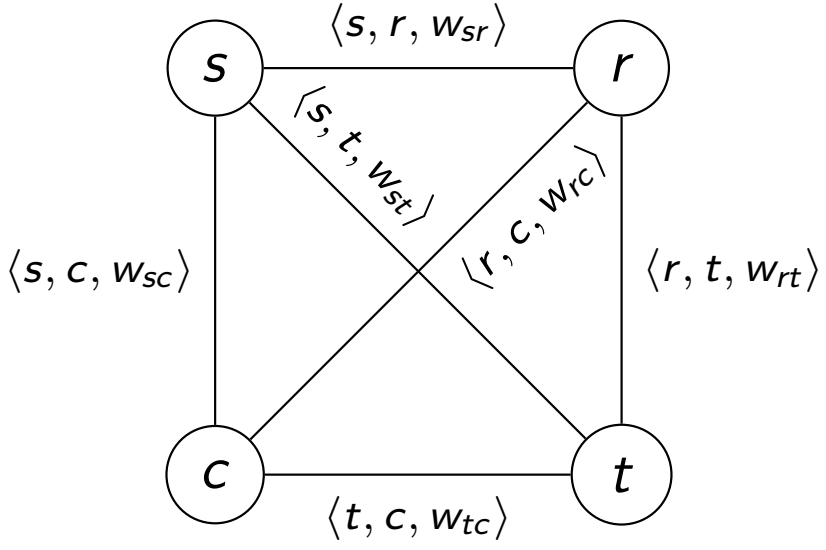


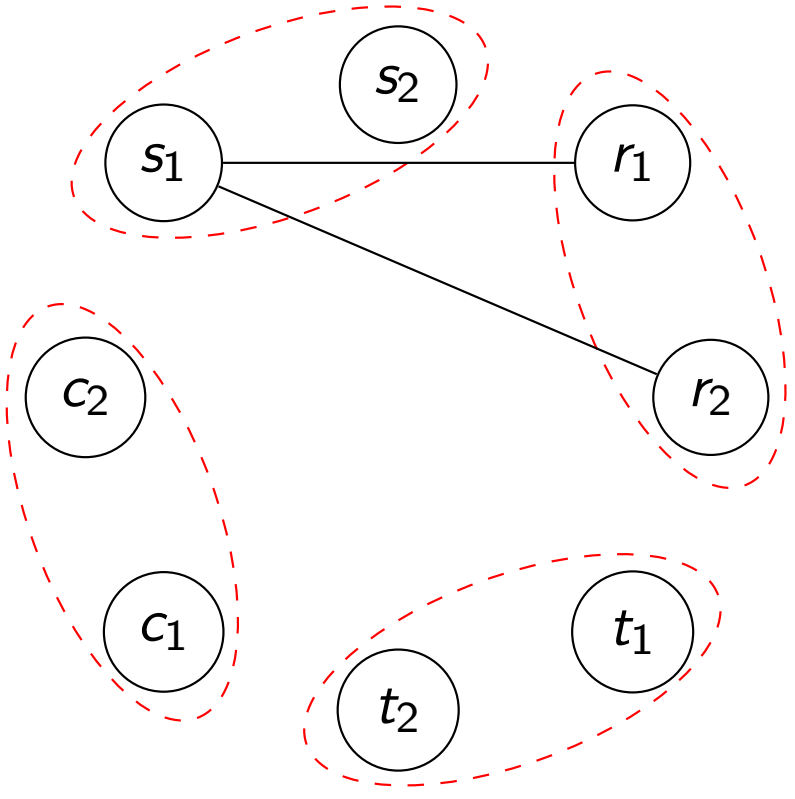
Model: Complete 4-Partite Weighted Graph



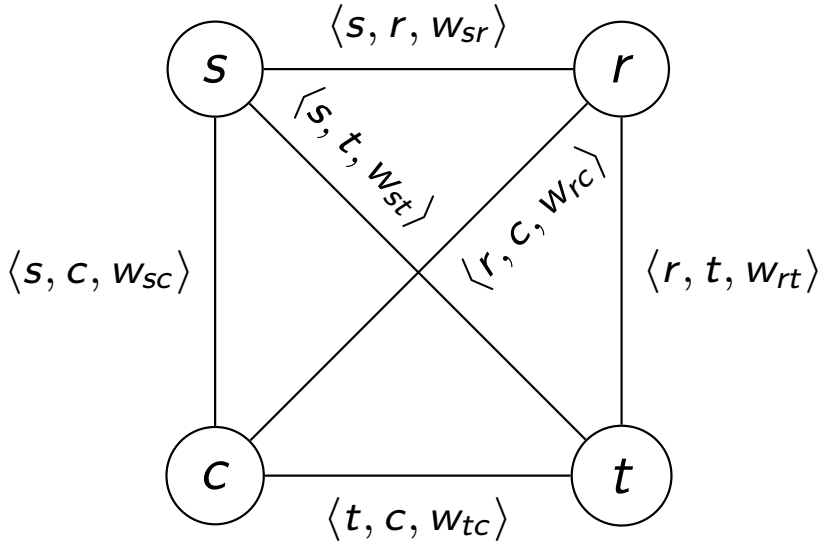


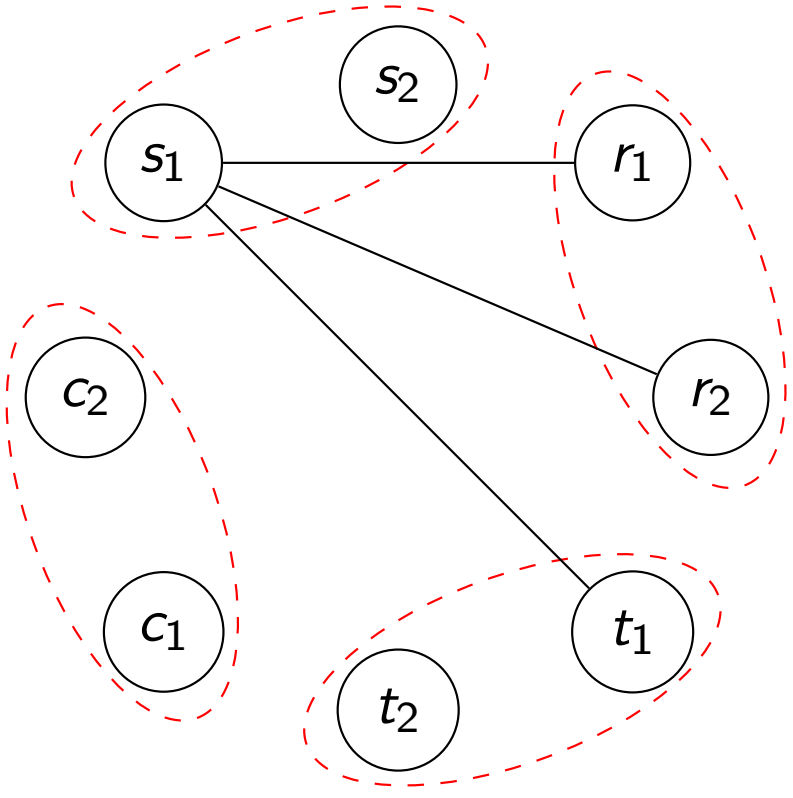
Model: Complete 4-Partite Weighted Graph



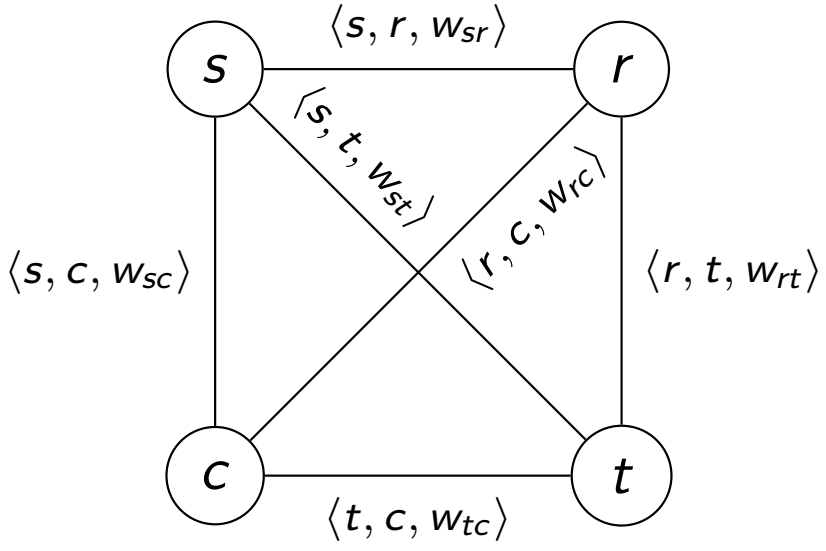


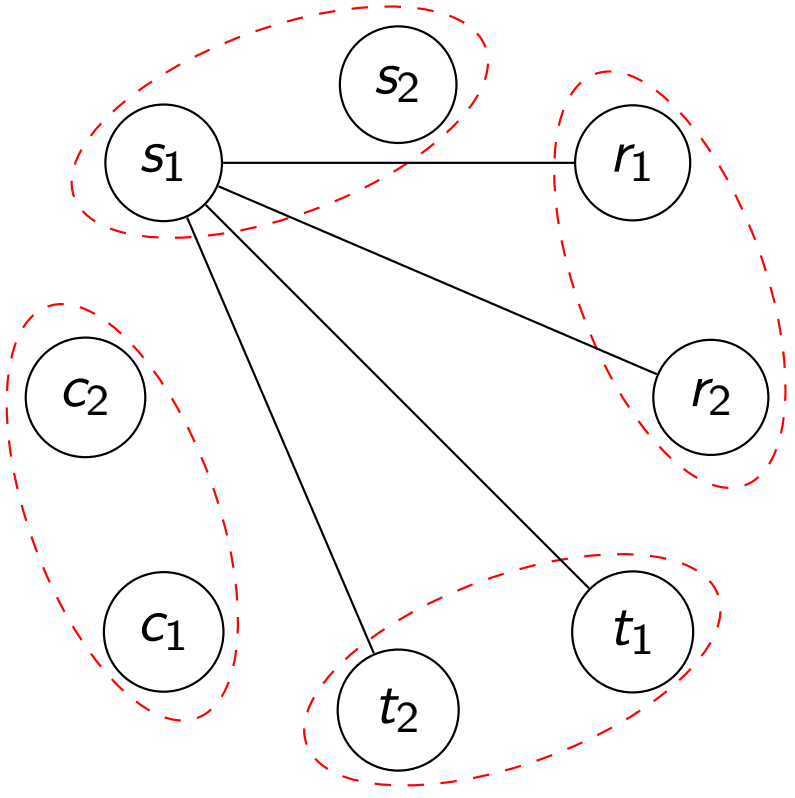
Model: Complete 4-Partite Weighted Graph



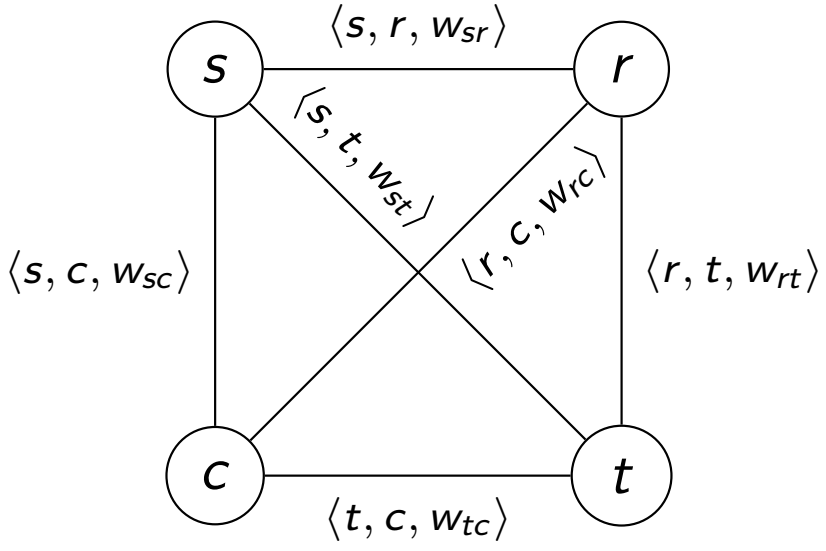


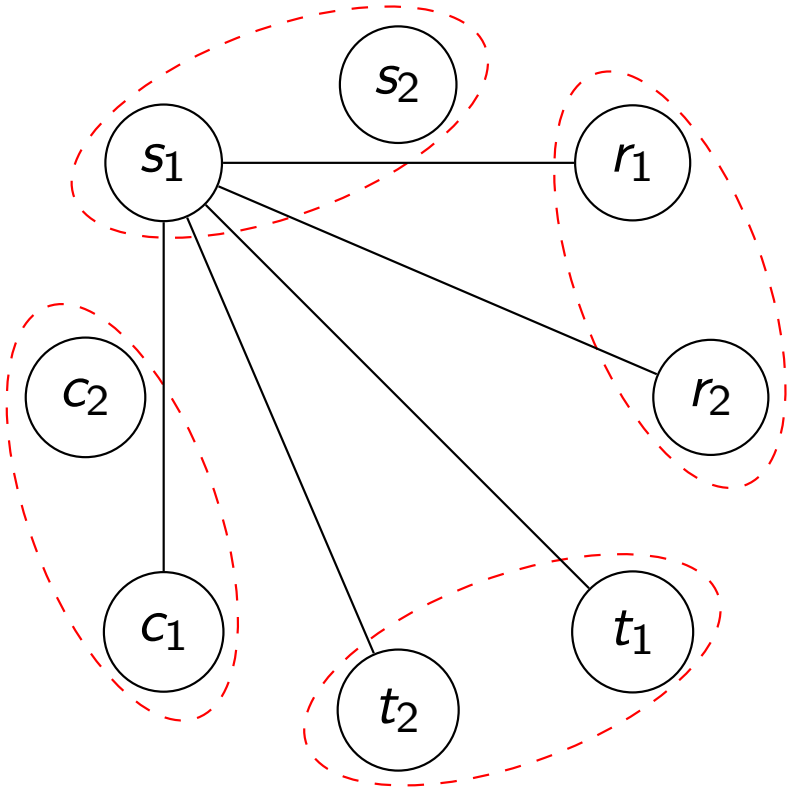
Model: Complete 4-Partite Weighted Graph



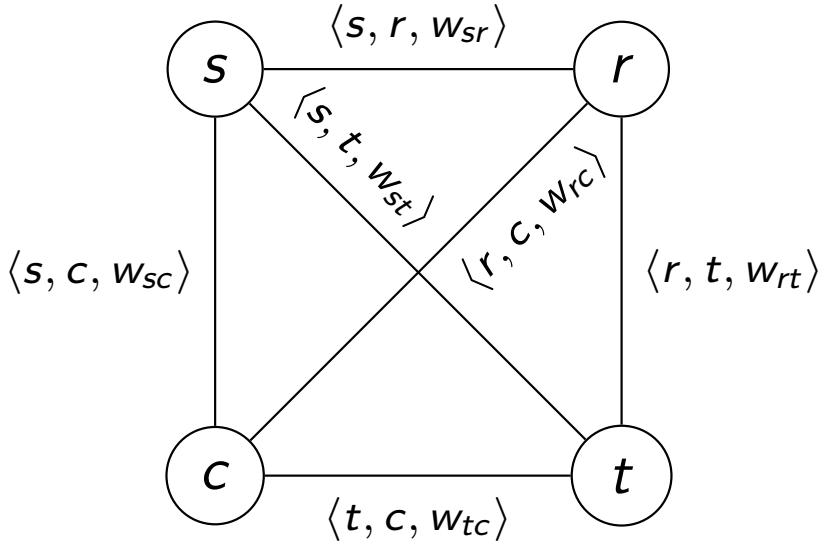


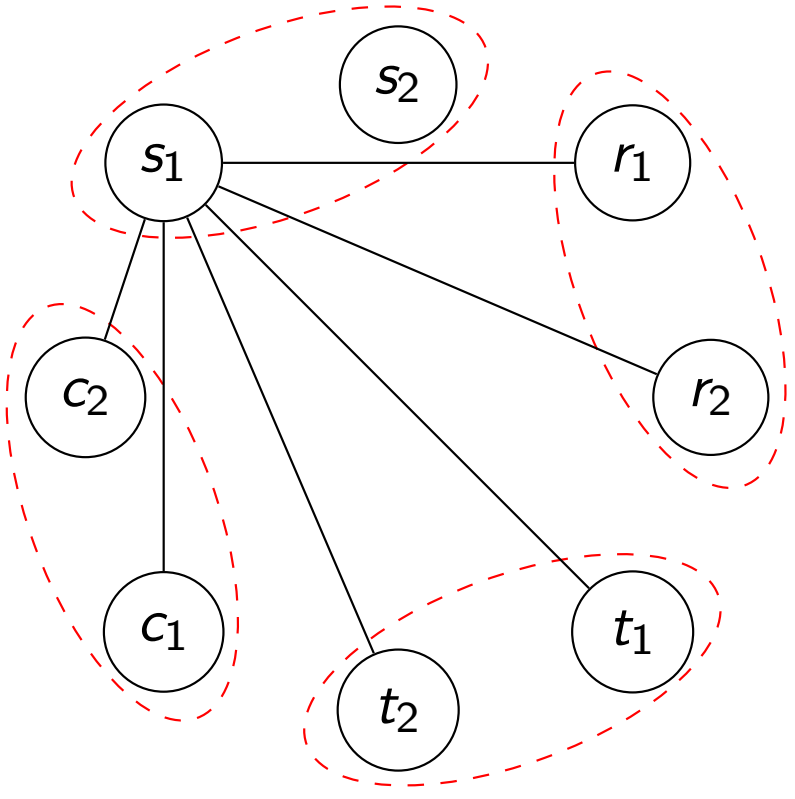
Model: Complete 4-Partite Weighted Graph



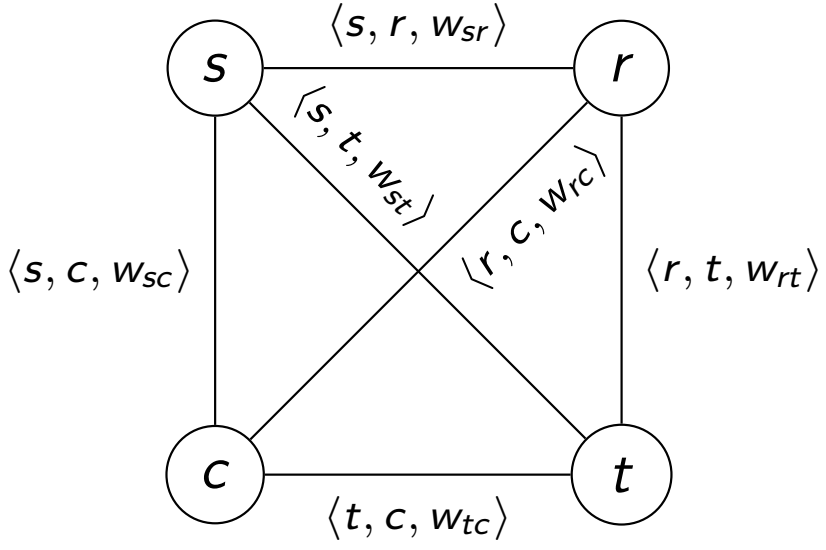


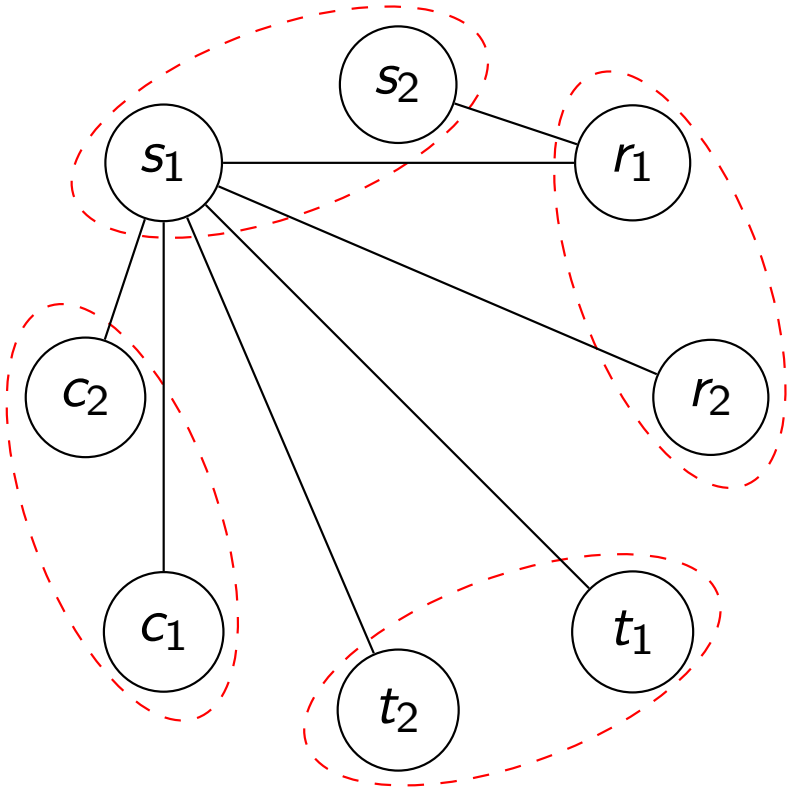
Model: Complete 4-Partite Weighted Graph



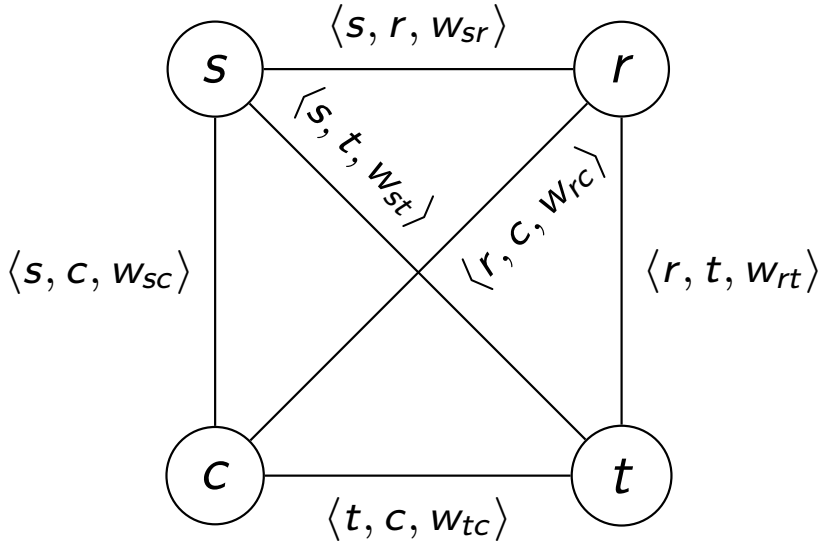


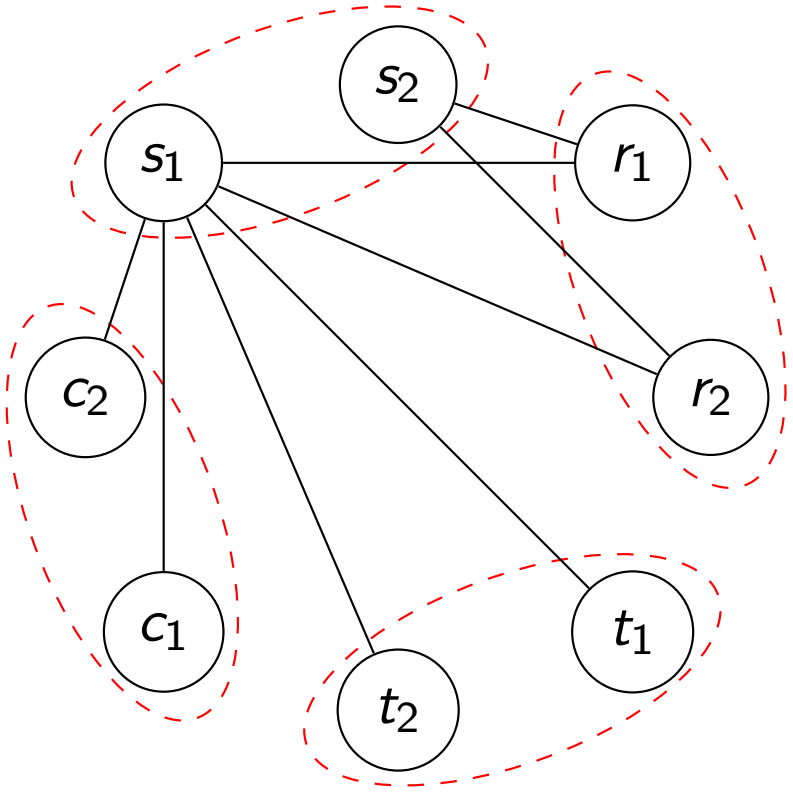
Model: Complete 4-Partite Weighted Graph



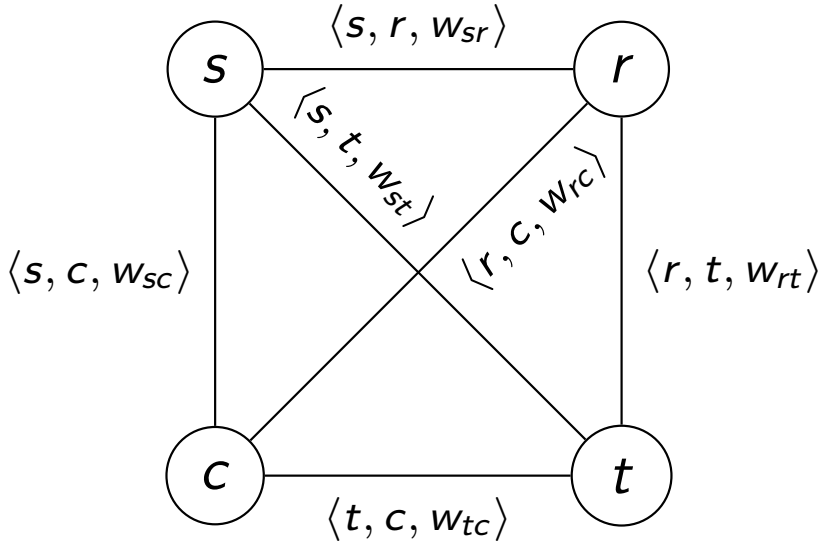


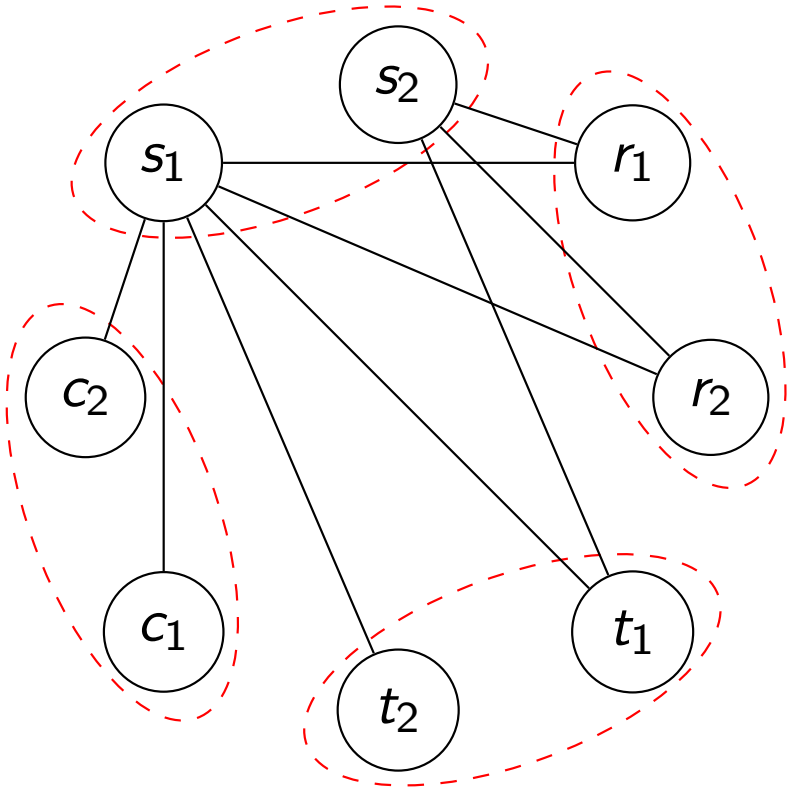
Model: Complete 4-Partite Weighted Graph



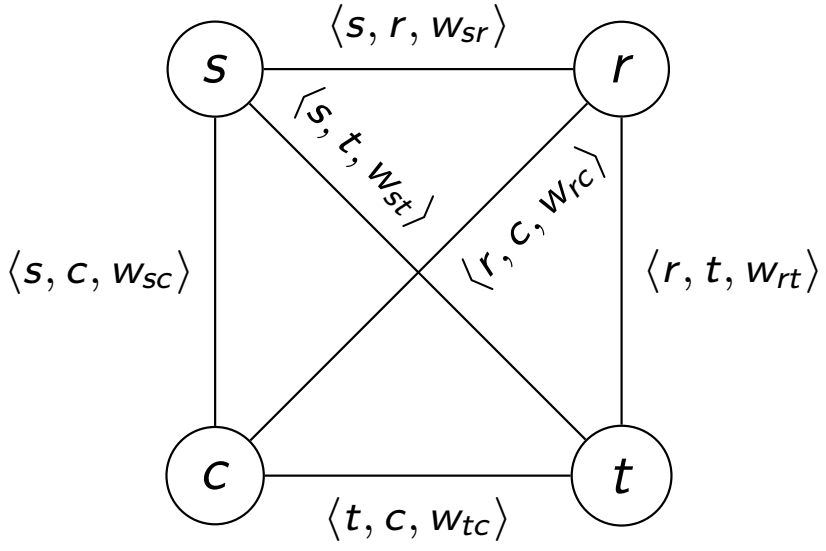


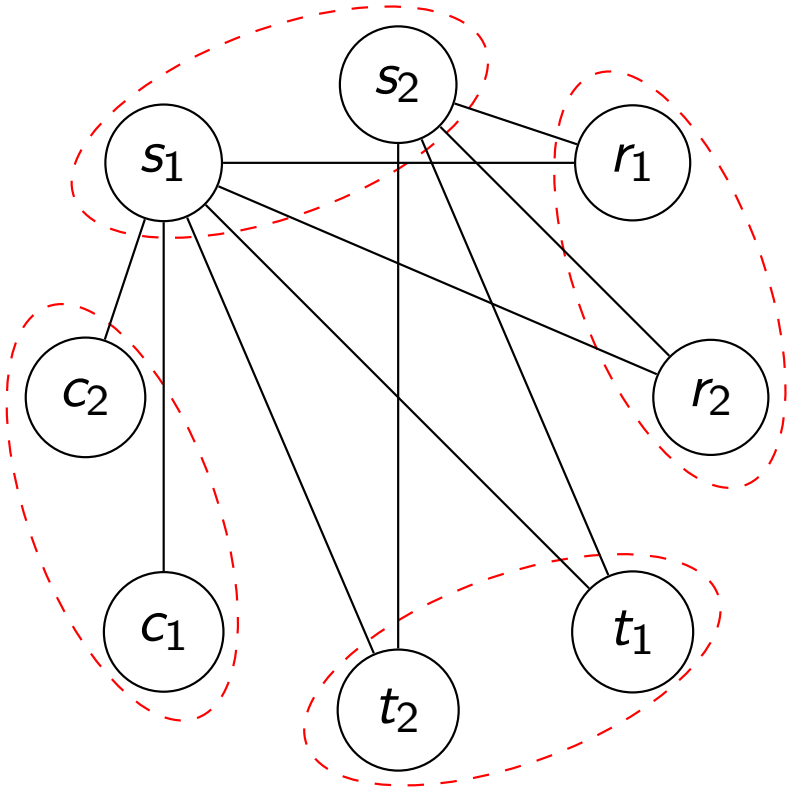
Model: Complete 4-Partite Weighted Graph



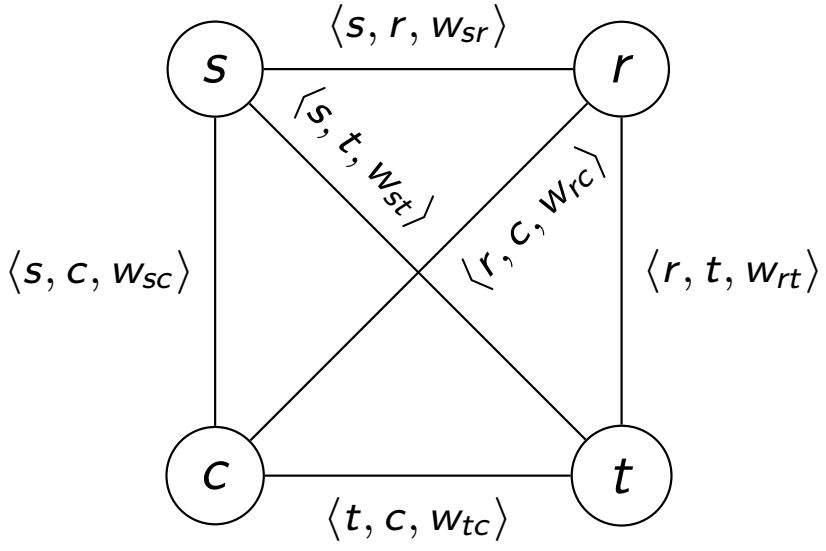


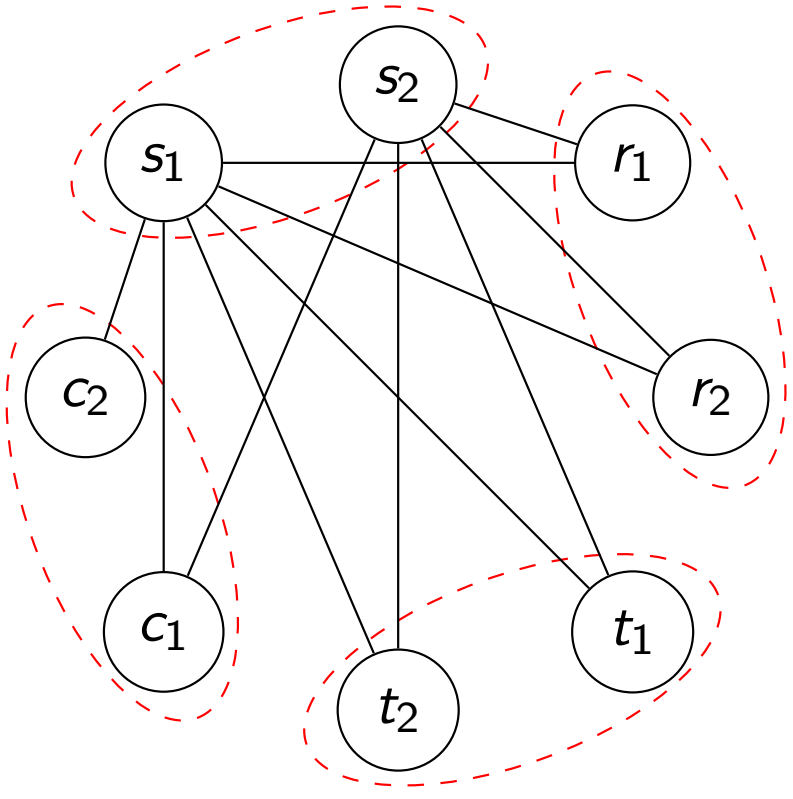
Model: Complete 4-Partite Weighted Graph



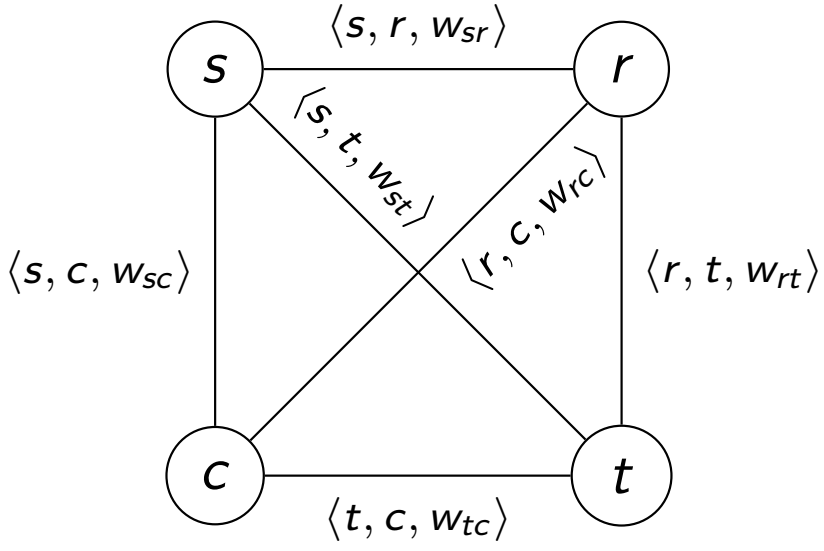


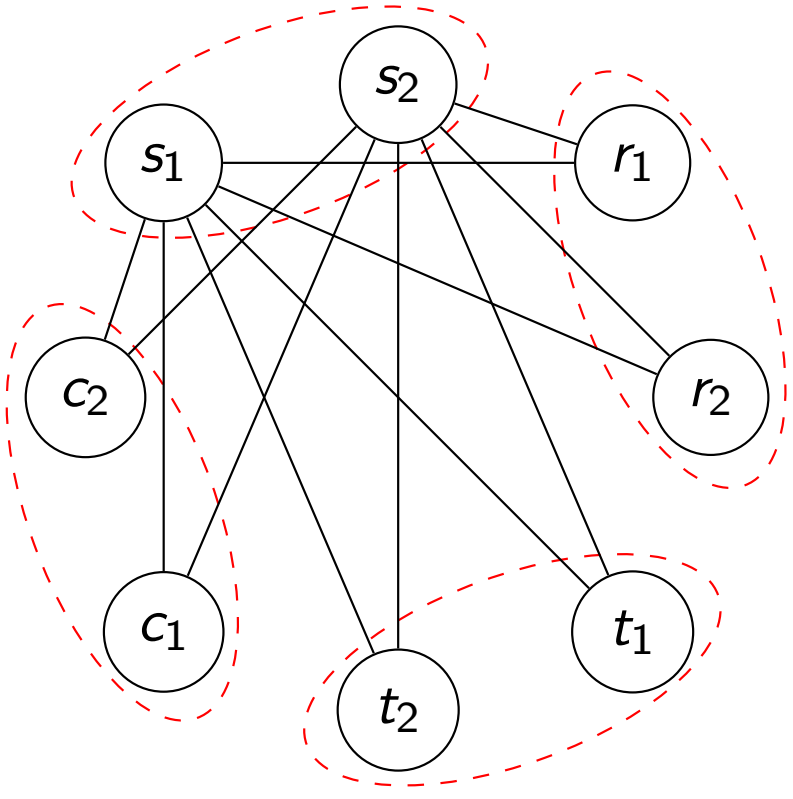
Model: Complete 4-Partite Weighted Graph



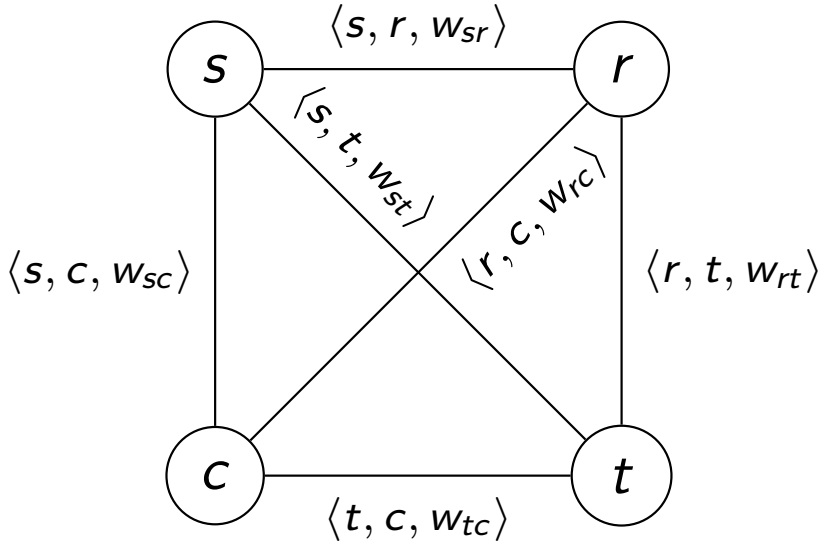


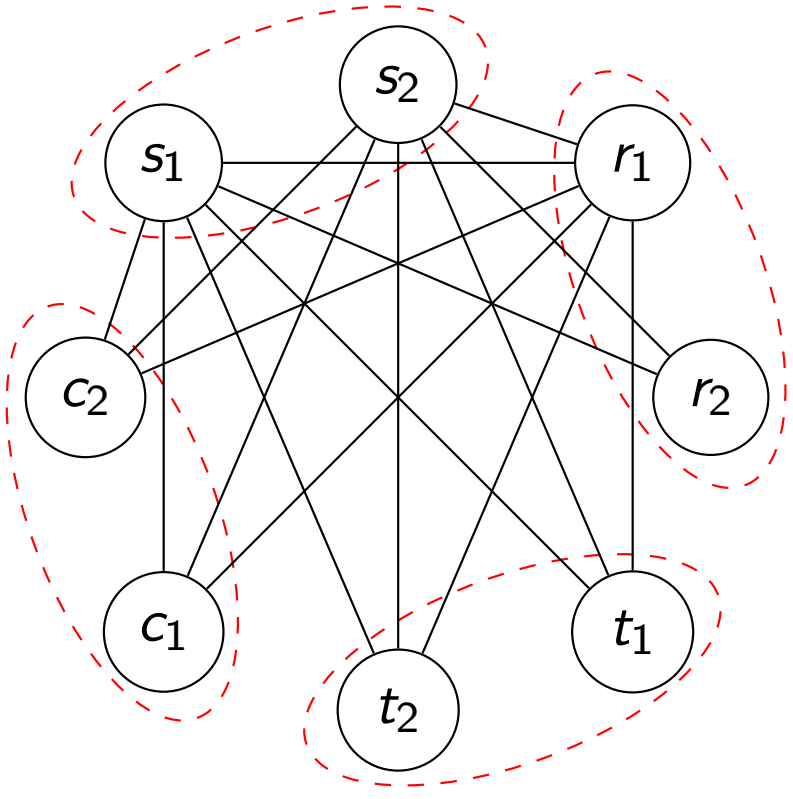
Model: Complete 4-Partite Weighted Graph



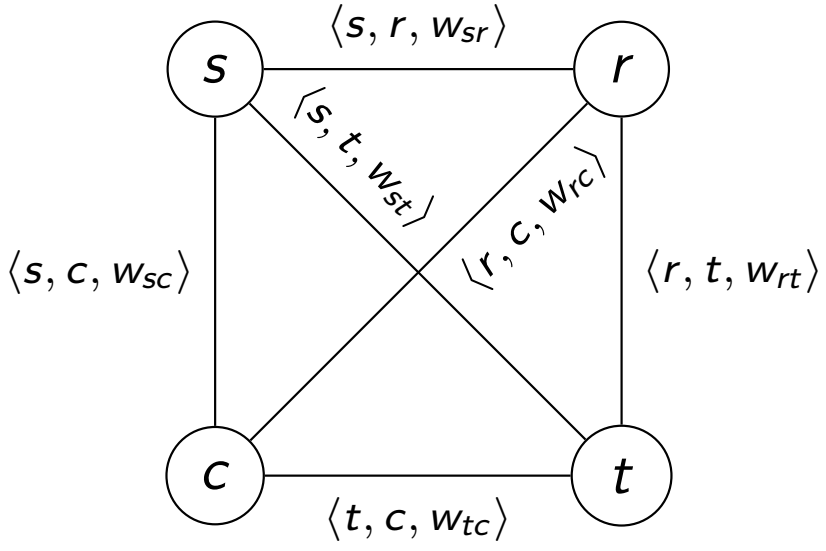


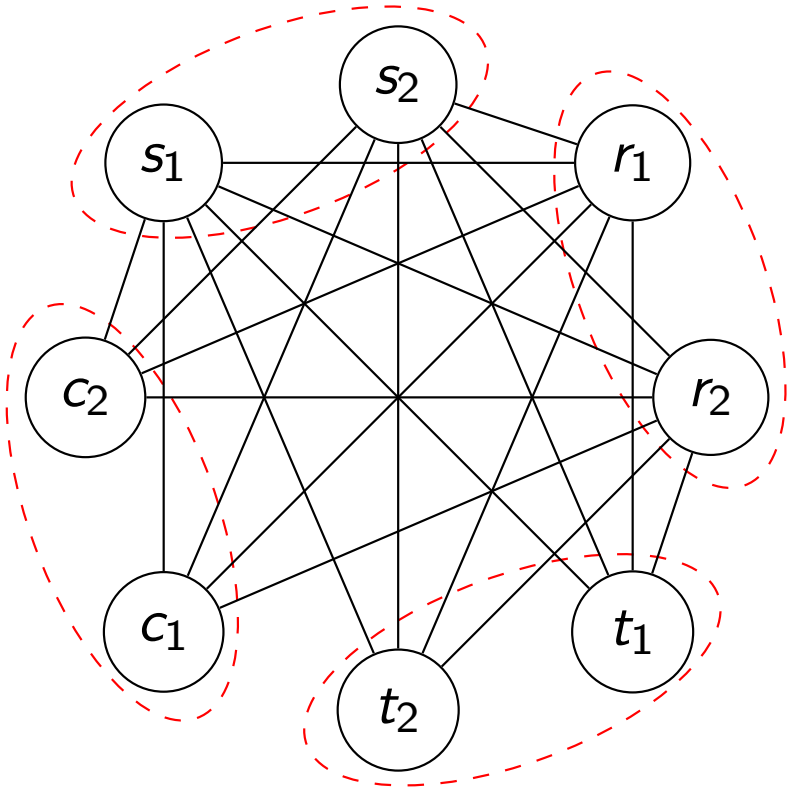
Model: Complete 4-Partite Weighted Graph



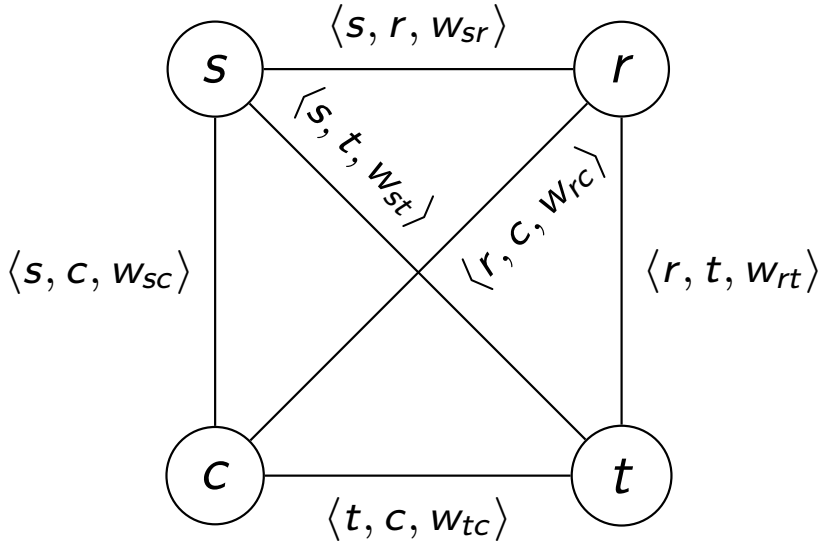


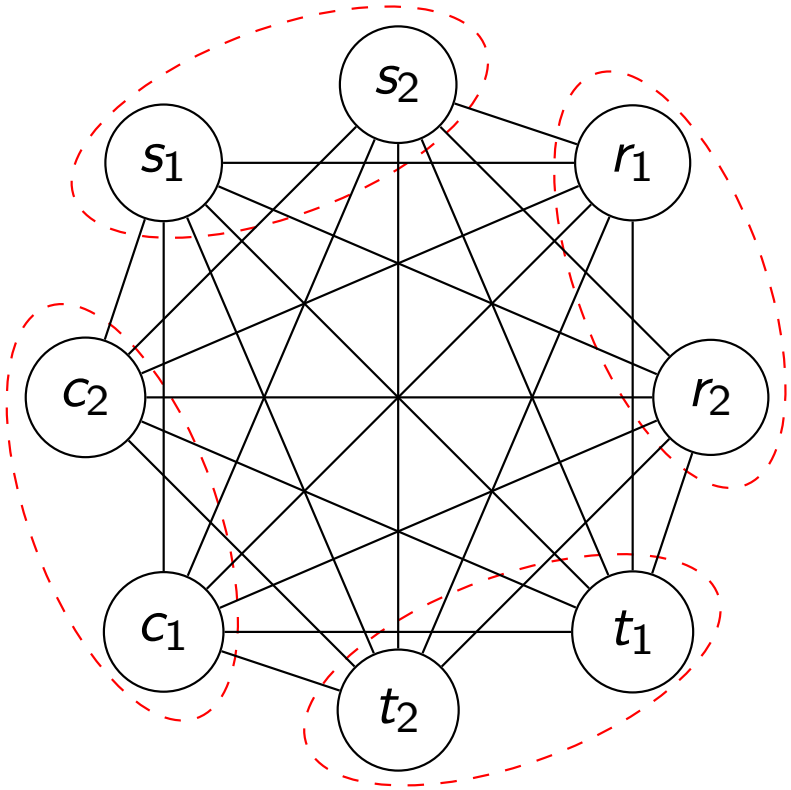
Model: Complete 4-Partite Weighted Graph





Model: Complete 4-Partite Weighted Graph





Weighting Policies

Which value for the weights?

- Amount of evidence: $w_{xy} \in \mathbb{R}^+$

Weighting Policies

Which value for the weights?

- Amount of evidence: $w_{xy} \in \mathbb{R}^+$
- $w_{xy} = 0 \Rightarrow$ no evidence

Weighting Policies

Which value for the weights?

- Amount of evidence: $w_{xy} \in \mathbb{R}^+$
- $w_{xy} = 0 \Rightarrow$ no evidence
- $(w_{xy}, w_{ab}) = (5, 10) \Rightarrow$ 2x evidence for $a - b$

Weighting Policies

Which value for the weights?

- Amount of evidence: $w_{xy} \in \mathbb{R}^+$
- $w_{xy} = 0 \Rightarrow$ no evidence
- $(w_{xy}, w_{ab}) = (5, 10) \Rightarrow$ 2x evidence for $a - b$
- Actual value depends on the interpretation of evidence
  - *Lim et al. [2010]: salience elicited from stakeholders*
  - *Castro-Herrera and Cleland-Huang [2010]: normalized term frequencies*

Weighting Policies

Which value for the weights?

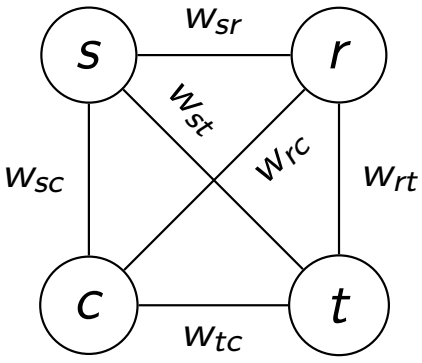
- Amount of evidence: $w_{xy} \in \mathbb{R}^+$
- $w_{xy} = 0 \Rightarrow$ no evidence
- $(w_{xy}, w_{ab}) = (5, 10) \Rightarrow$ 2x evidence for $a - b$
- Actual value depends on the interpretation of evidence
  - *Lim et al. [2010]: salience elicited from stakeholders*
  - *Castro-Herrera and Cleland-Huang [2010]: normalized term frequencies*
- Challenge: have meaningful weights

Weighting Policies

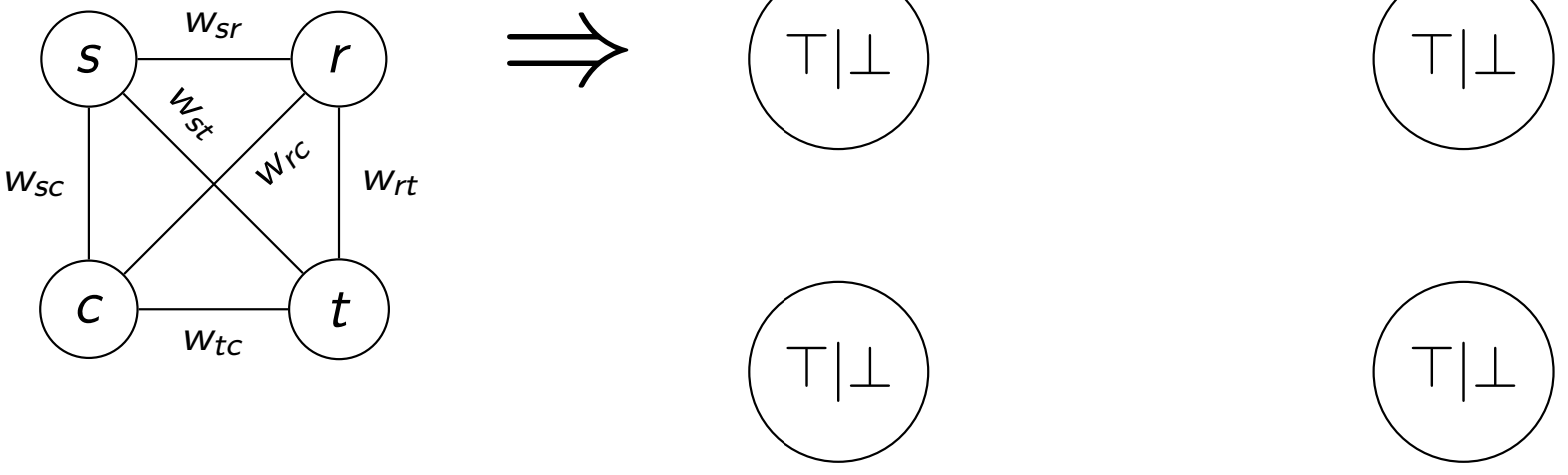
Which value for the weights?

- Amount of evidence: $w_{xy} \in \mathbb{R}^+$
- $w_{xy} = 0 \Rightarrow$ no evidence
- $(w_{xy}, w_{ab}) = (5, 10) \Rightarrow$ 2x evidence for $a - b$
- Actual value depends on the interpretation of evidence
  - *Lim et al. [2010]: salience elicited from stakeholders*
  - *Castro-Herrera and Cleland-Huang [2010]: normalized term frequencies*
- Challenge: have meaningful weights
- MN simplification: scale-independence (5 vs 10 = 1 vs 2)

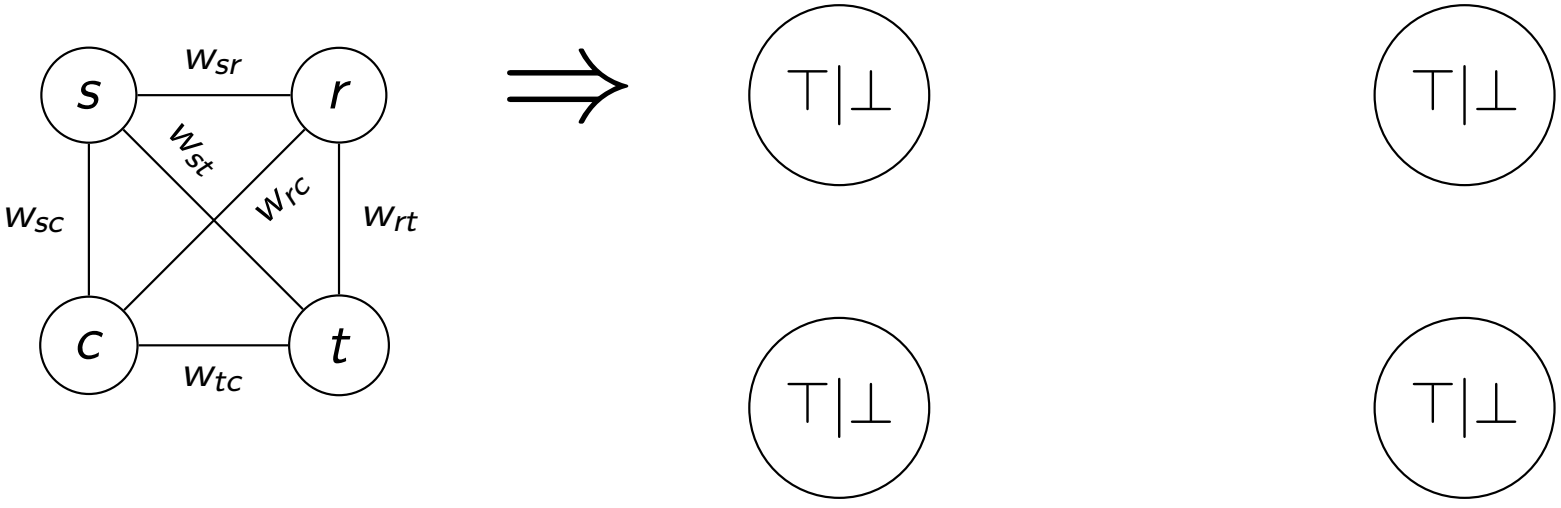
Markov Networks (Markov Random Field)



Markov Networks (Markov Random Field)



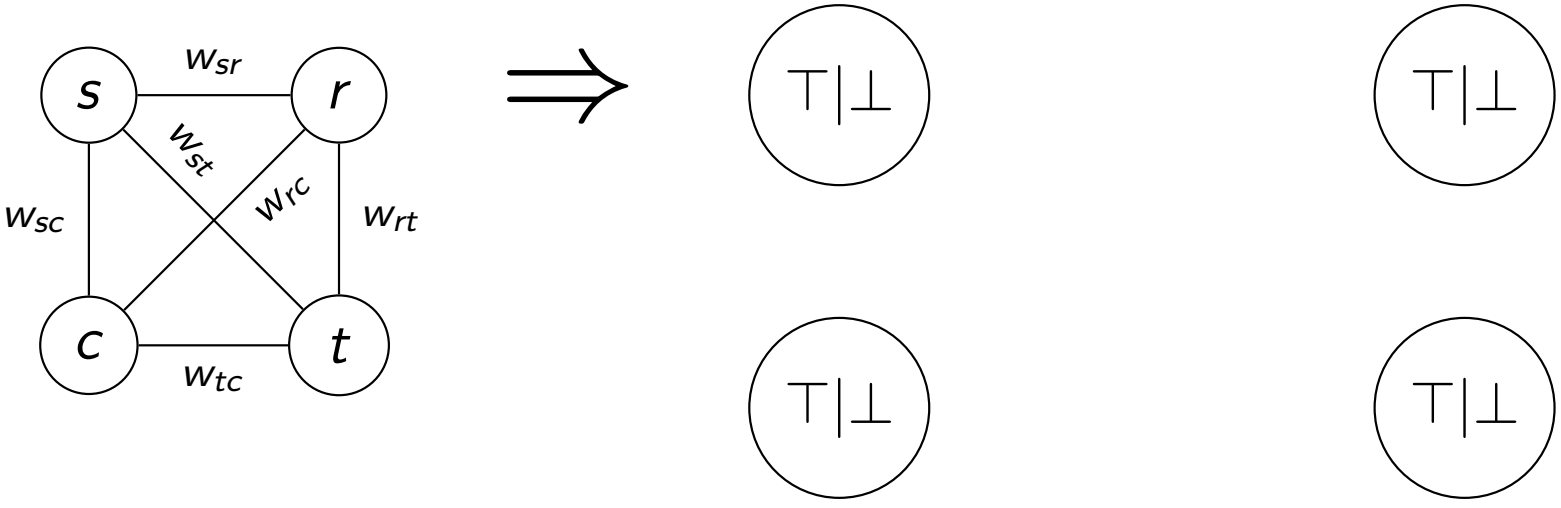
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert

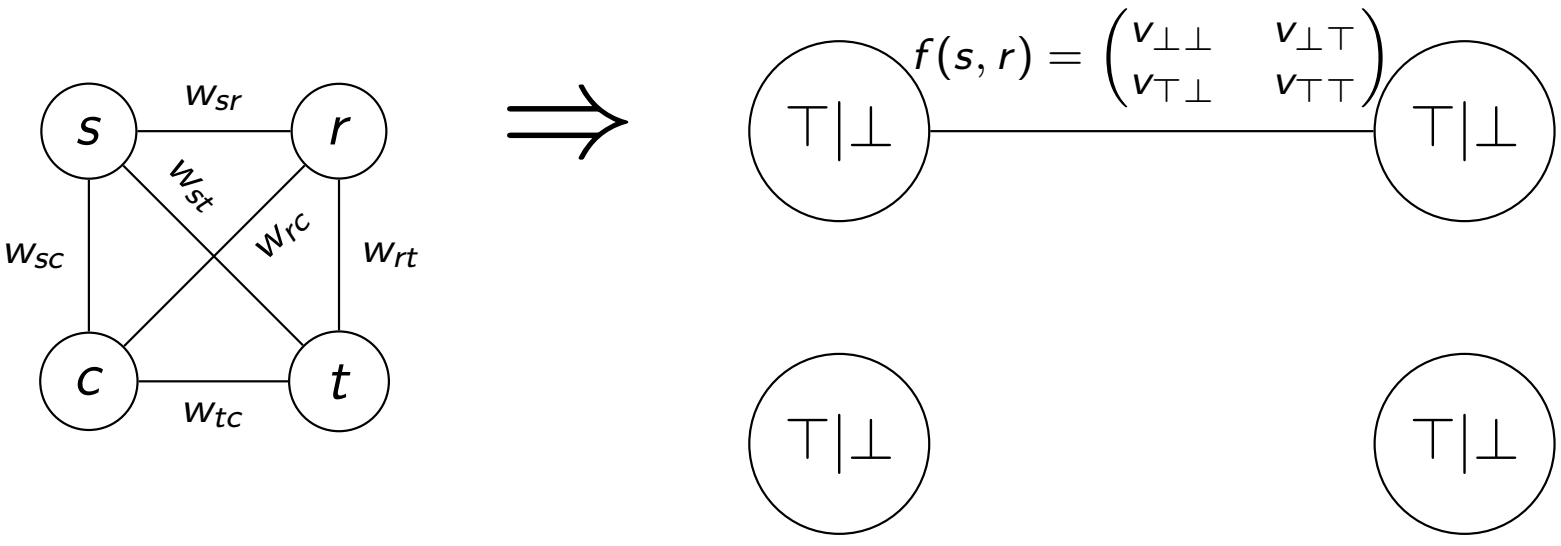
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$

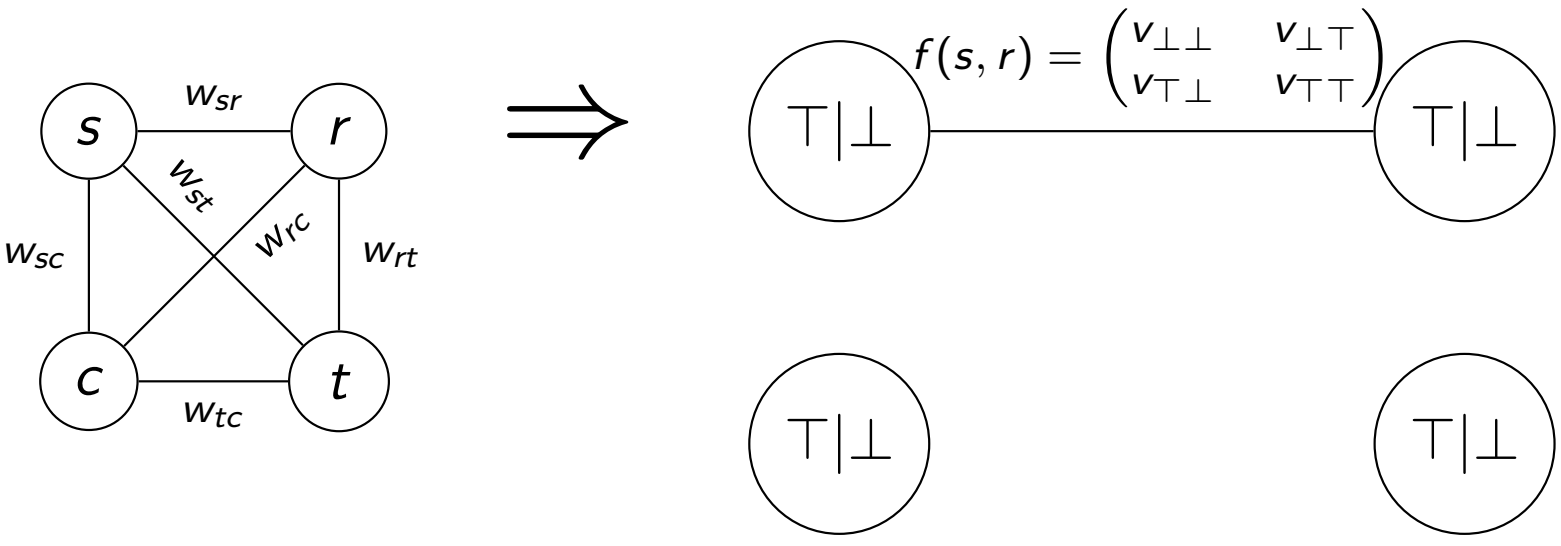
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$

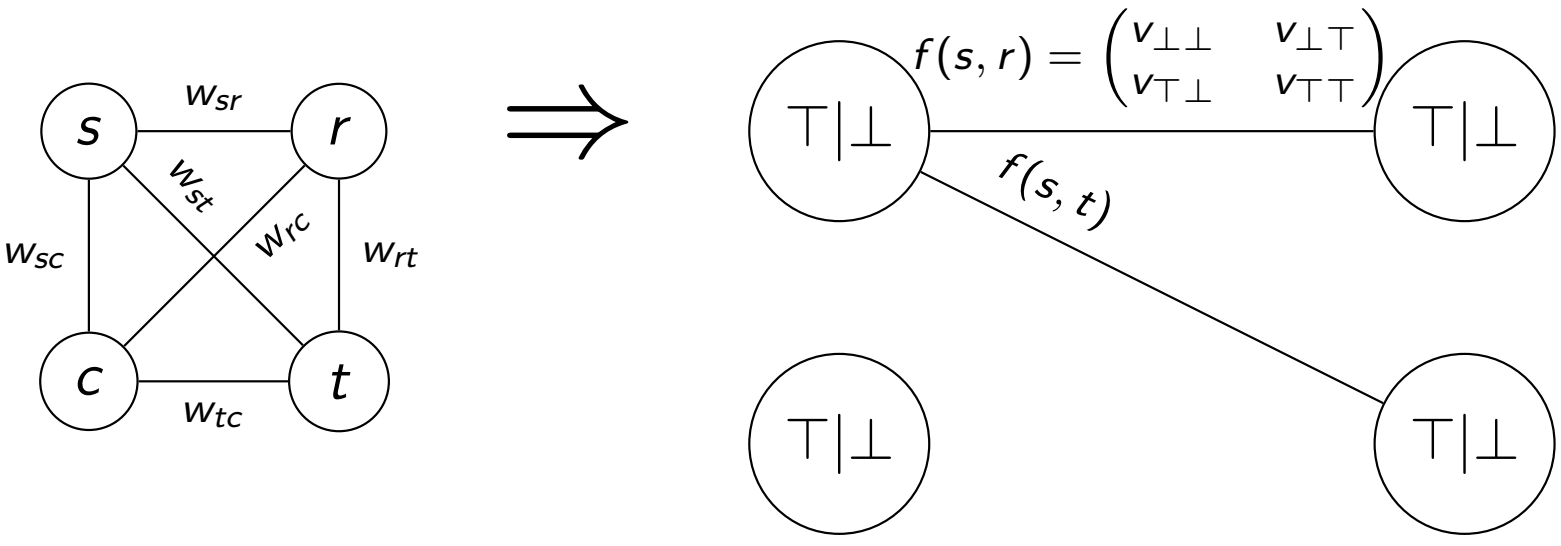
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\bot\bot}, v_{\bot\top}, v_{\top\bot} \\ w_{xy} & v_{\top\top} \end{cases}$

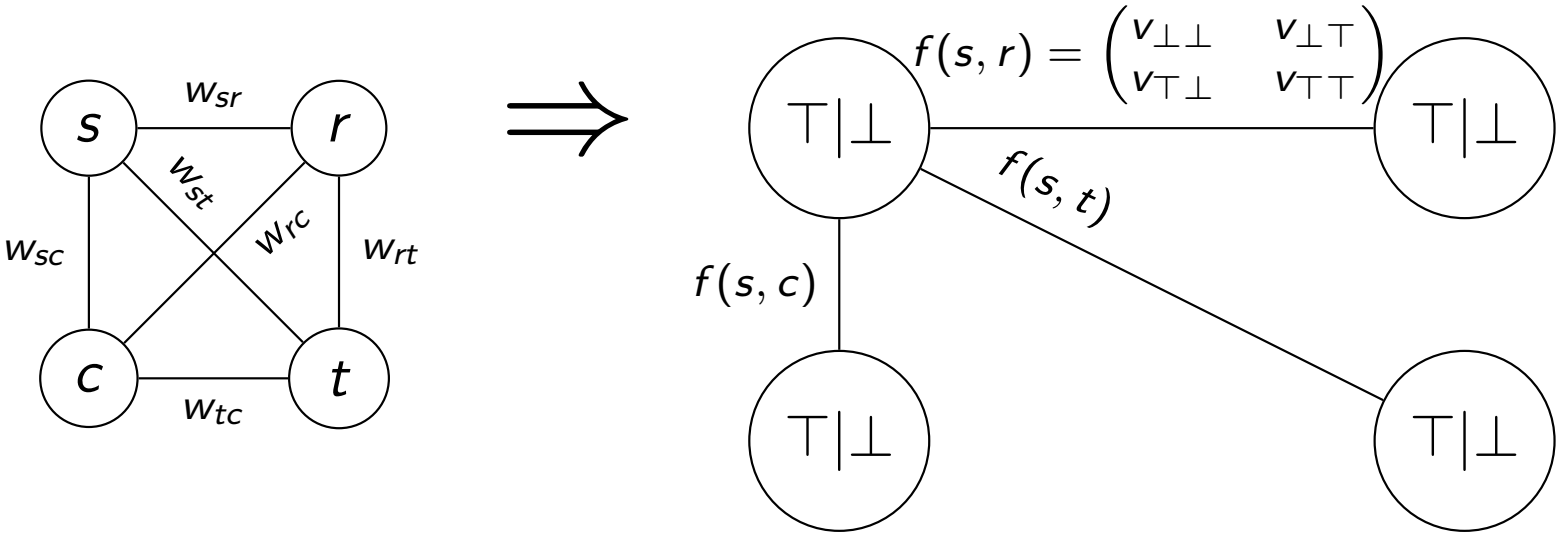
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\perp\perp}, v_{\perp\top}, v_{\top\perp} \\ w_{xy} & v_{\top\top} \end{cases}$

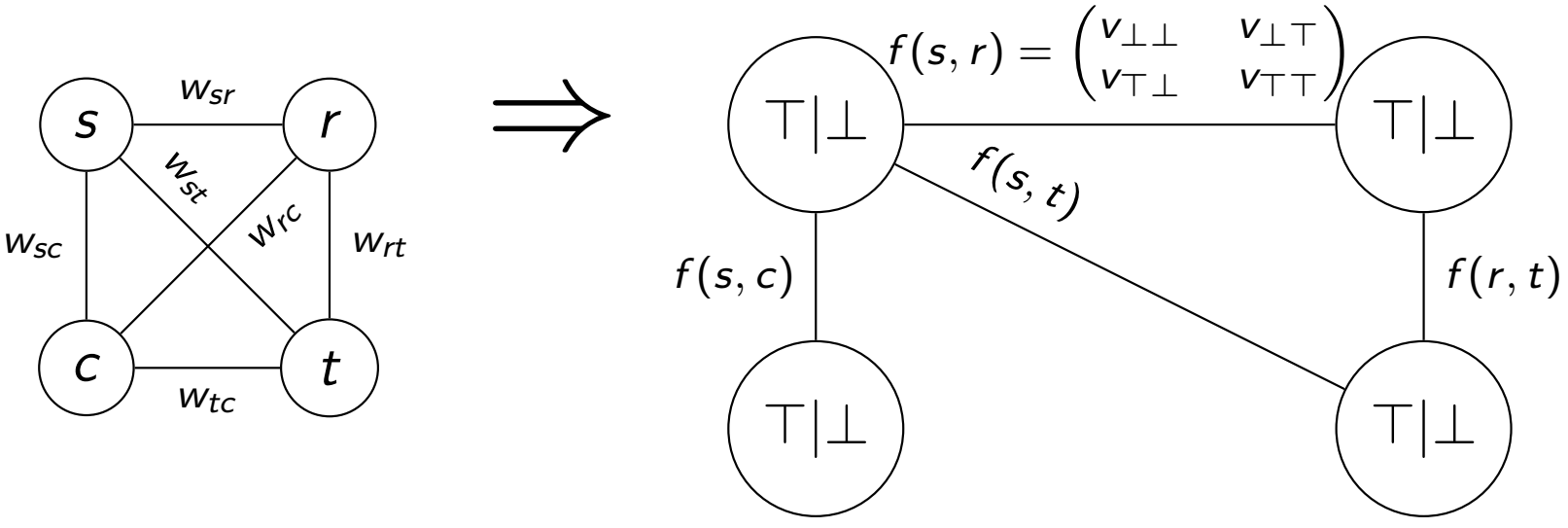
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\bot\bot}, v_{\bot\top}, v_{\top\bot} \\ w_{xy} & v_{\top\top} \end{cases}$

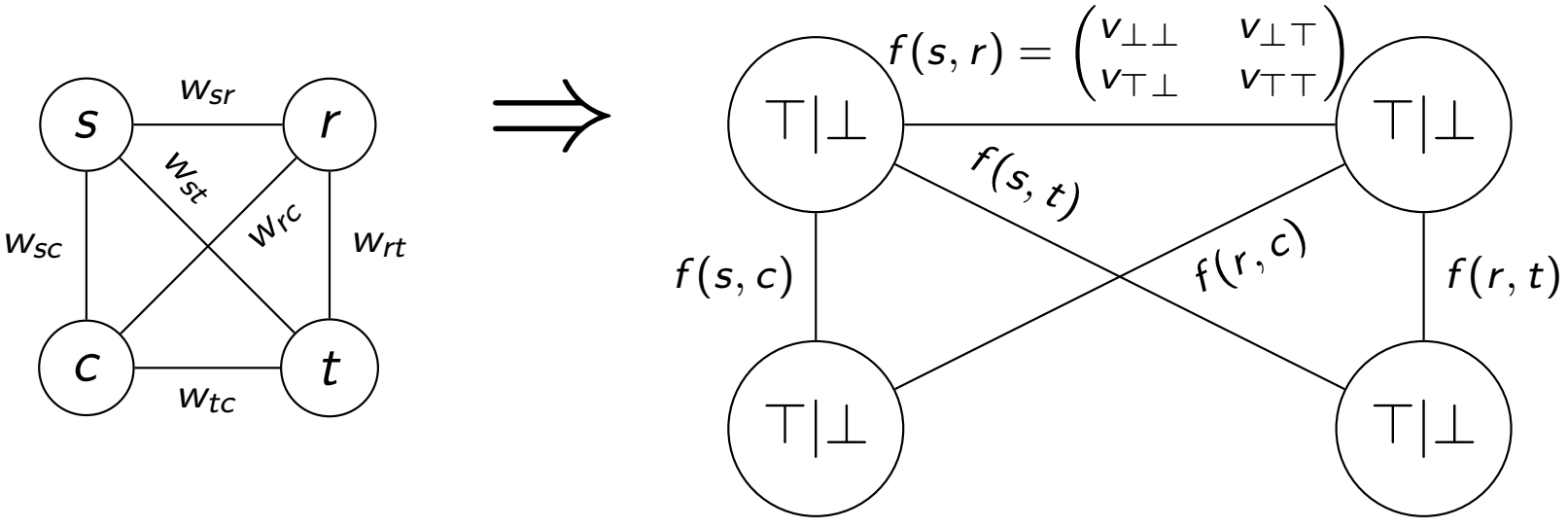
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\perp\perp}, v_{\perp\top}, v_{\top\perp} \\ w_{xy} & v_{\top\top} \end{cases}$

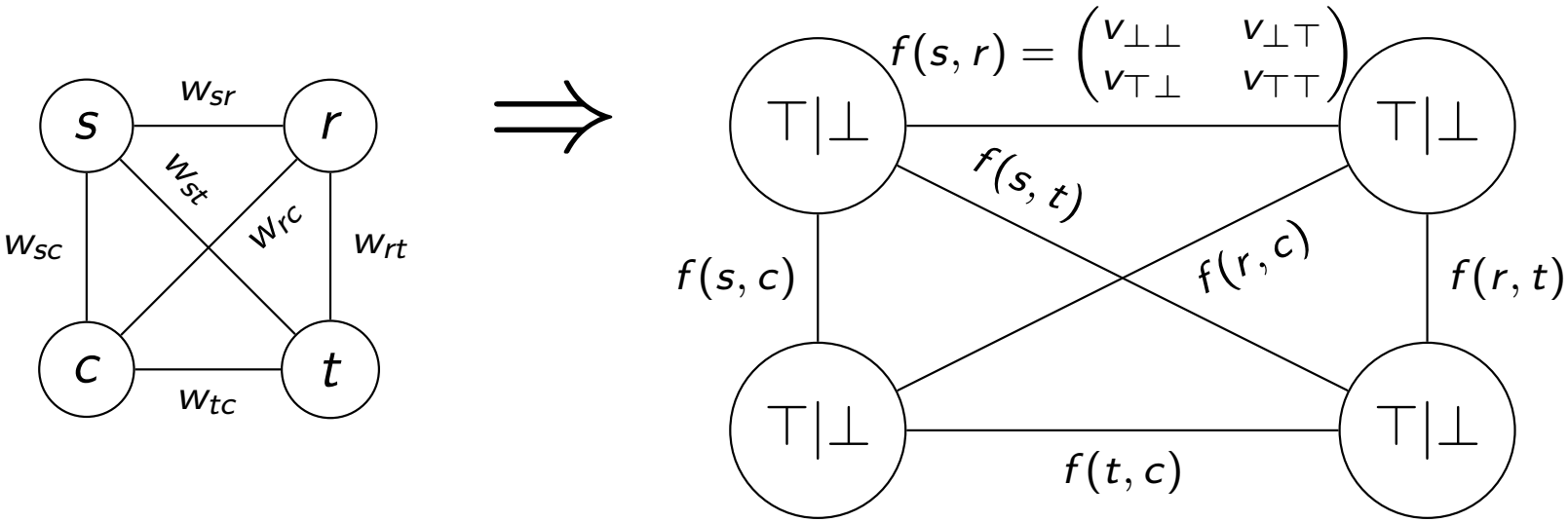
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\perp\perp}, v_{\perp\top}, v_{\top\perp} \\ w_{xy} & v_{\top\top} \end{cases}$

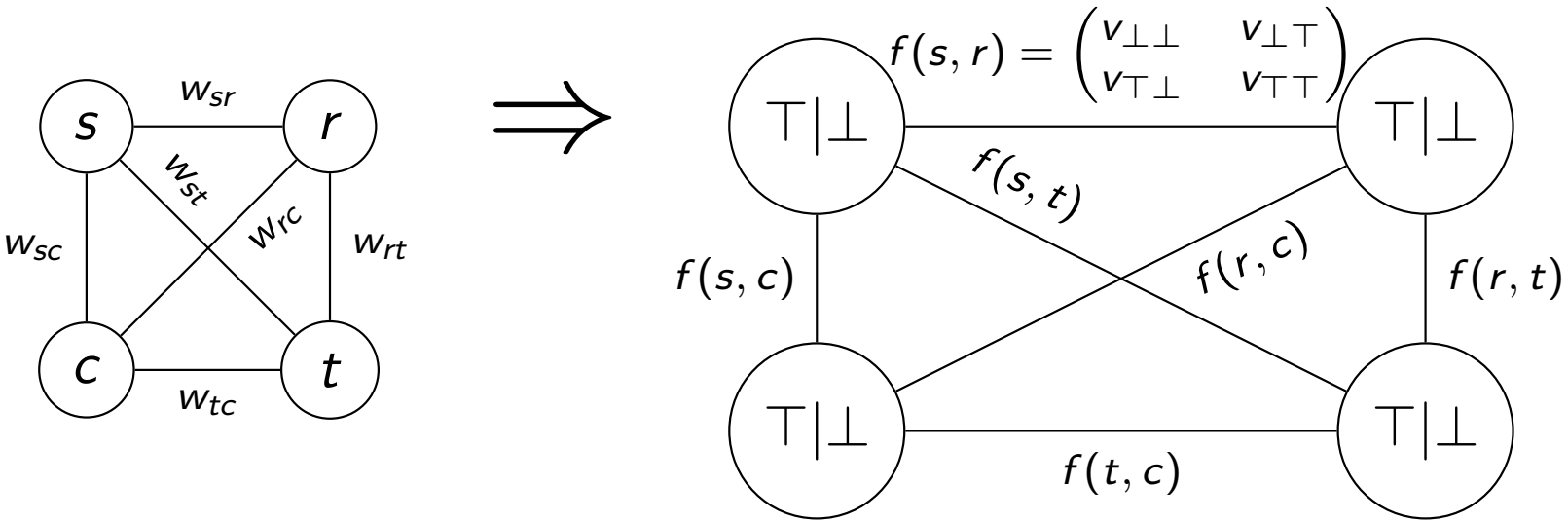
Markov Networks (Markov Random Field)



Interpretation:

- $s = \top \Rightarrow s$ is an expert
- $r/t/c = \top \Rightarrow$ looking for experts in $r/t/c$
- $f(x,y) = \begin{cases} 0 & v_{\perp\perp}, v_{\perp\top}, v_{\top\perp} \\ w_{xy} & v_{\top\top} \end{cases}$

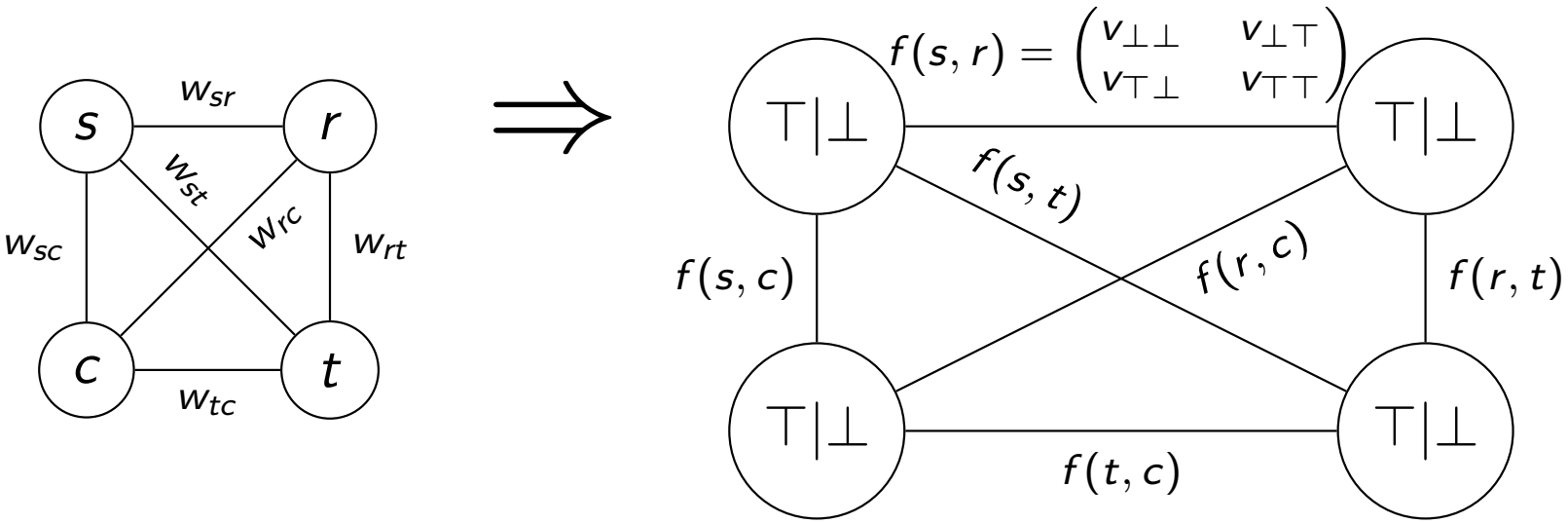
Markov Networks (Markov Random Field)



Probabilities:

- $MN = (N \text{ nodes}, M \text{ functions})$

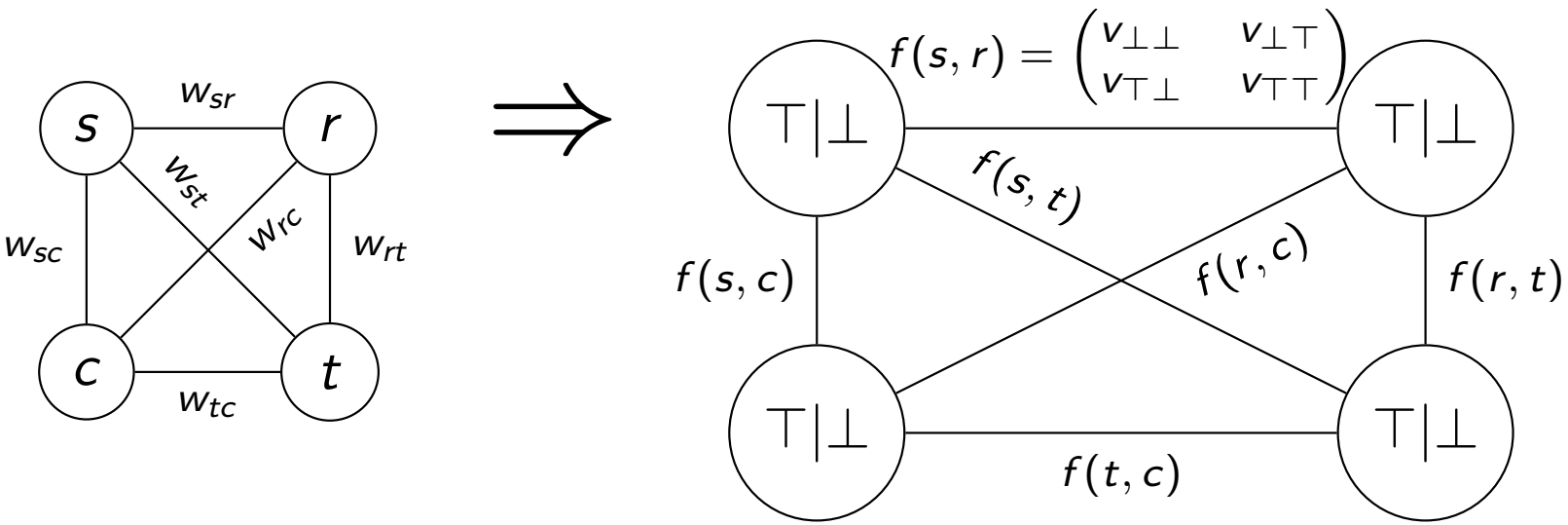
Markov Networks (Markov Random Field)



Probabilities:

- $MN = (N \text{ nodes}, M \text{ functions})$
- $P(n_1 = \top, n_2 = \perp, n_3 = \perp, ..., n_N = \top) = \frac{\prod_{i=1}^{M} f_i(n_\alpha, n_\beta)}{Z}$

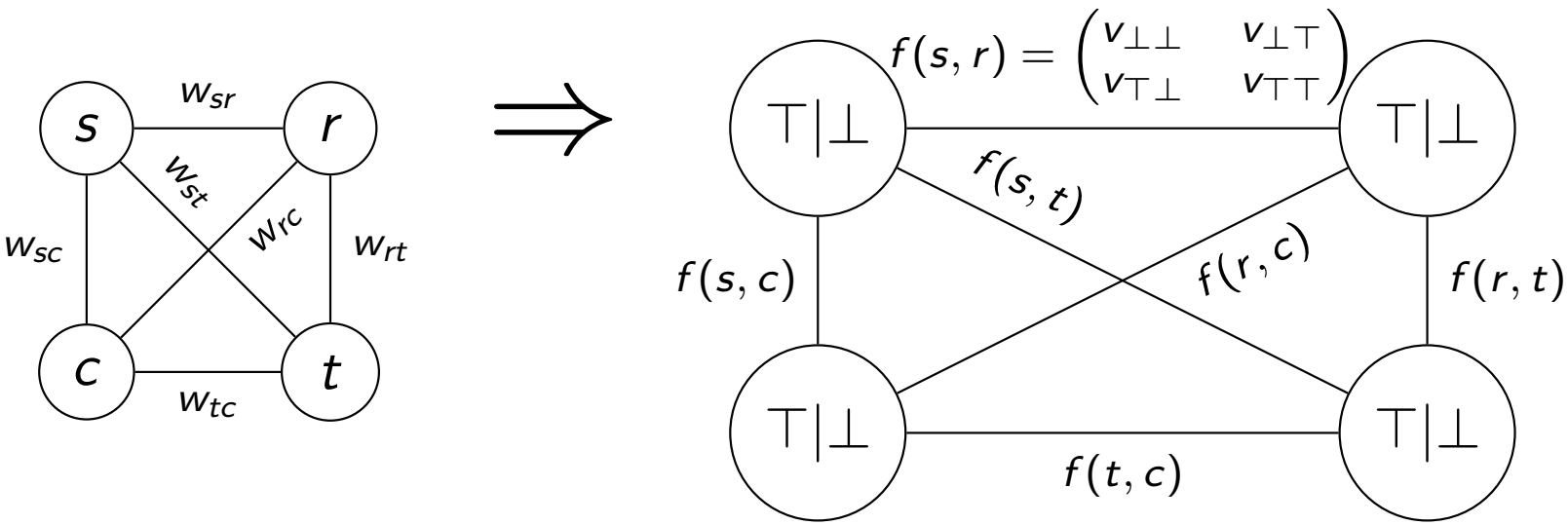
Markov Networks (Markov Random Field)



Probabilities:

- $MN = (N \text{ nodes}, M \text{ functions})$
- $P(n_1 = \top, n_2 = \perp, n_3 = \perp, ..., n_N = \top) = \frac{\prod_{i=1}^{M} f_i(n_\alpha, n_\beta)}{Z}$
- $P(n_1 = \top) = \sum_{\sigma_2,...,\sigma_N \in \{\perp,\top\}} P(n_1 = \top, n_2 = \sigma_2, ...)$

Markov Networks (Markov Random Field)



Probabilities:

- $MN = (N \text{ nodes}, M \text{ functions})$
- $P(n_1 = \top, n_2 = \perp, n_3 = \perp, ..., n_N = \top) = \frac{\prod_{i=1}^{M} f_i(n_\alpha, n_\beta)}{Z}$
- $P(n_1 = \top) = \sum_{\sigma_2,...,\sigma_N \in \{\perp,\top\}} P(n_1 = \top, n_2 = \sigma_2, ...)$
- $P(n_1 = \top | \{n_j = \sigma_j\}) = P(n_1 = \top)$ with Z reduced to sequences where $\{n_j = \sigma_j\}$

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$
- Ranking: sort from most to least probable experts.

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$
- Ranking: sort from most to least probable experts.

Interesting MN property: scale independence

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$
- Ranking: sort from most to least probable experts.

Interesting MN property: scale independence

- $f_i' = \alpha . f_i \Rightarrow P'(\{n_i = \sigma_i\}) = P(\{n_i = \sigma_i\})$

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$
- Ranking: sort from most to least probable experts.

Interesting MN property: scale independence

- $f_i' = \alpha . f_i \Rightarrow P'(\{n_i = \sigma_i\}) = P(\{n_i = \sigma_i\})$
- Can choose arbitrary scale as reference.

Experts Ranking with MN

Computation:

- Query: $P(s_i = \top | t_{cryptography} = \top)$
- Ranking: sort from most to least probable experts.

Interesting MN property: scale independence

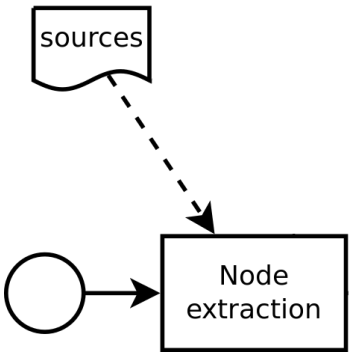
- $f_i' = \alpha . f_i \Rightarrow P'(\{n_i = \sigma_i\}) = P(\{n_i = \sigma_i\})$
- Can choose arbitrary scale as reference.
- Remaining problem: merging different semantics.

Full Process

Full Process

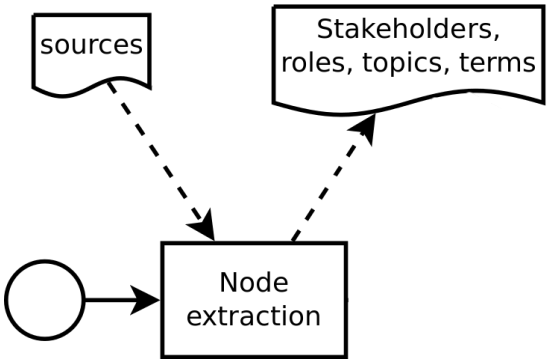
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.

Full Process



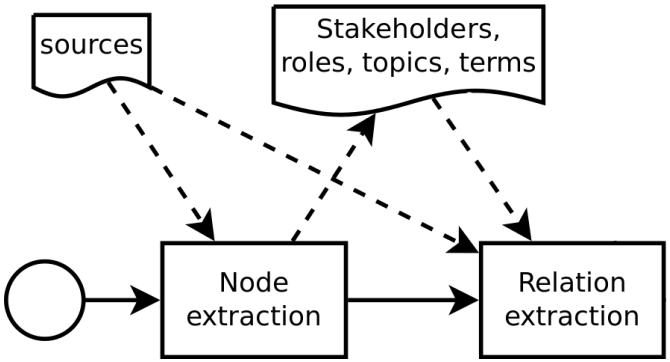
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.

Full Process



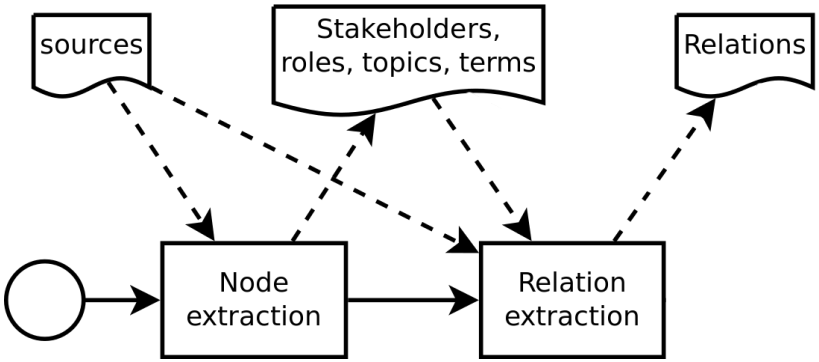
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.

Full Process



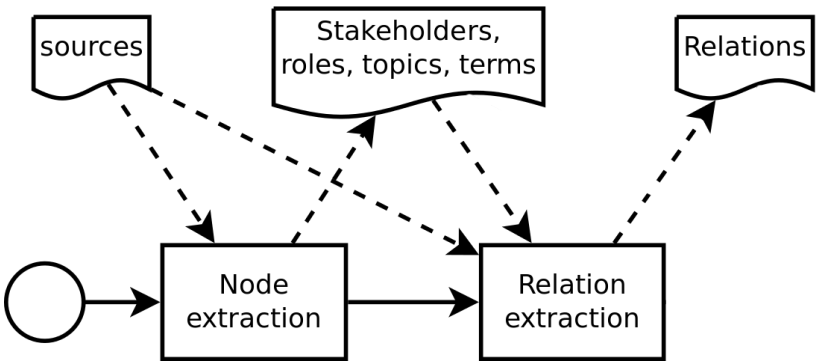
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.

Full Process



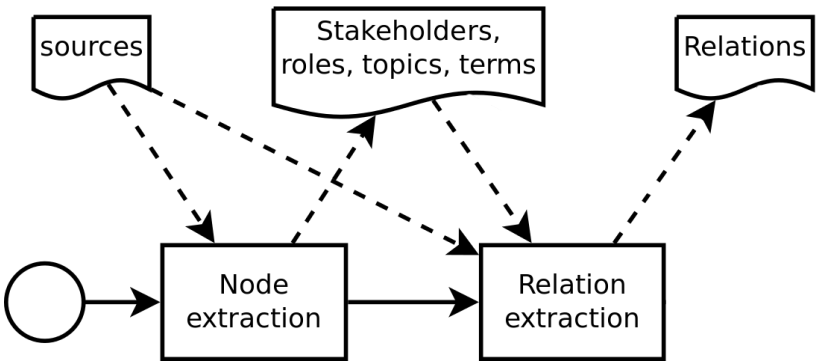
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.

Full Process



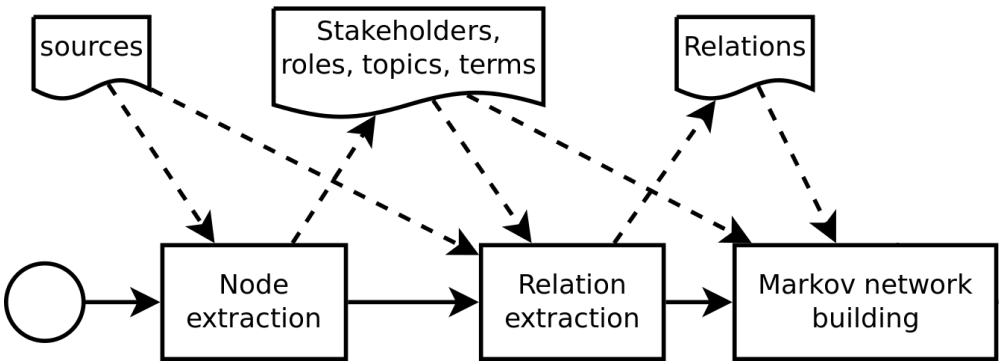
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors

Full Process



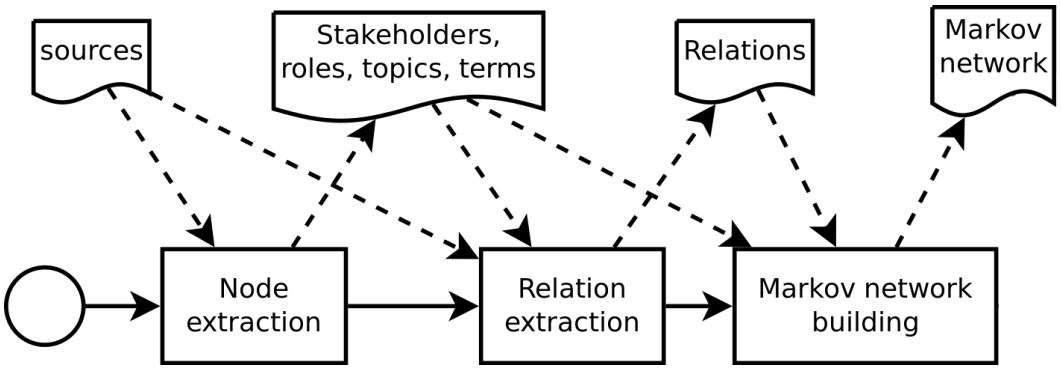
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



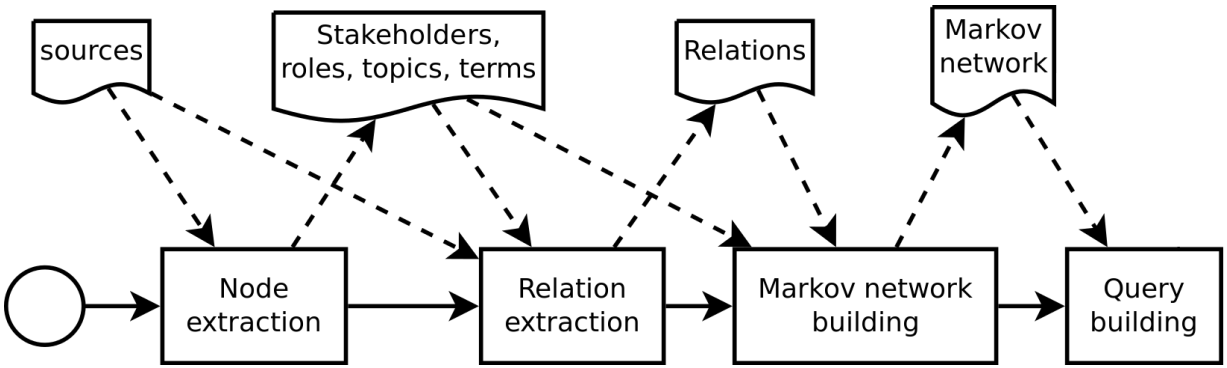
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



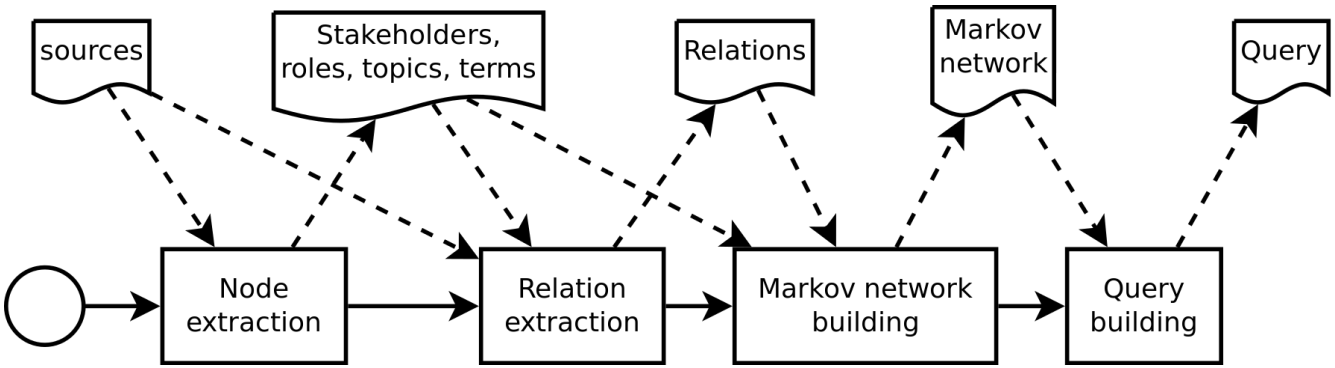
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



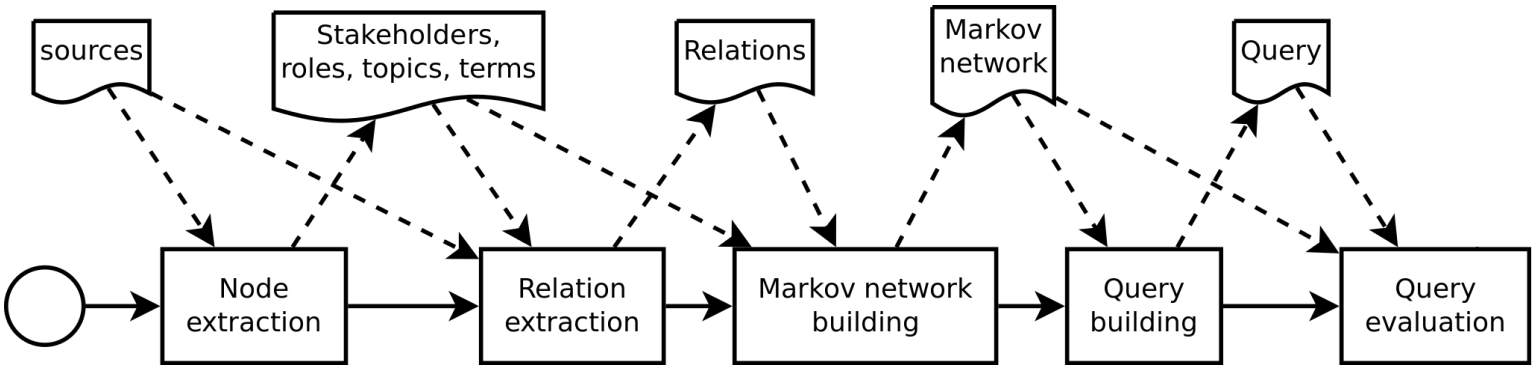
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



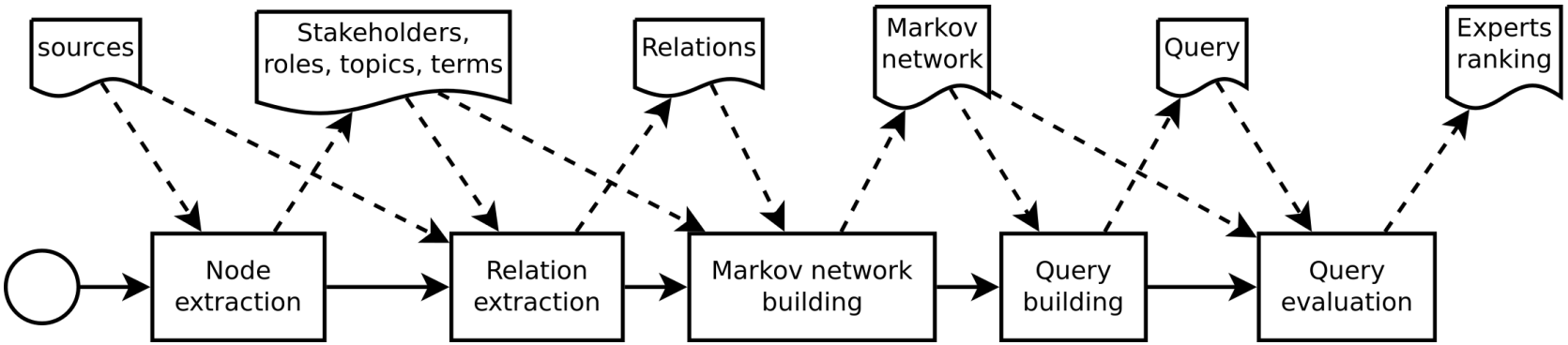
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



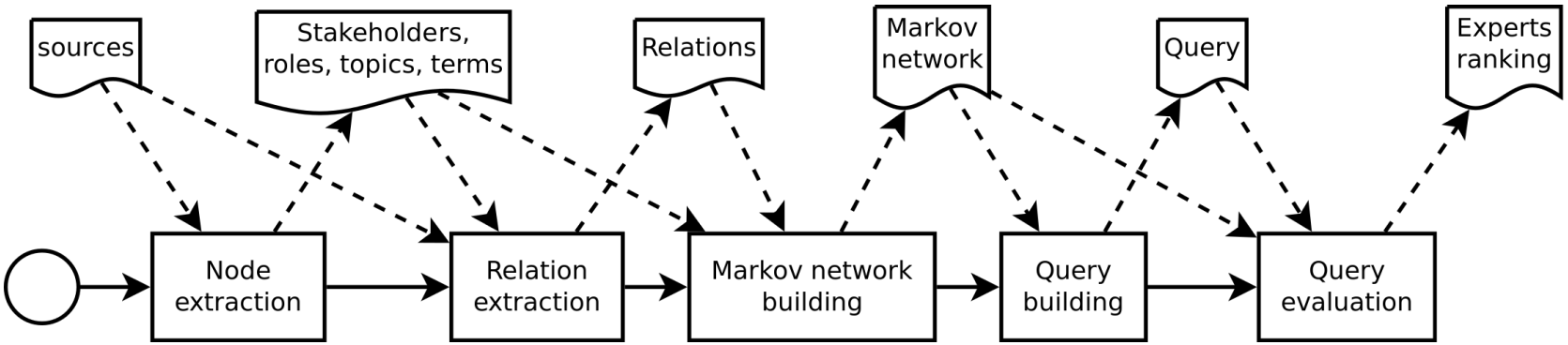
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



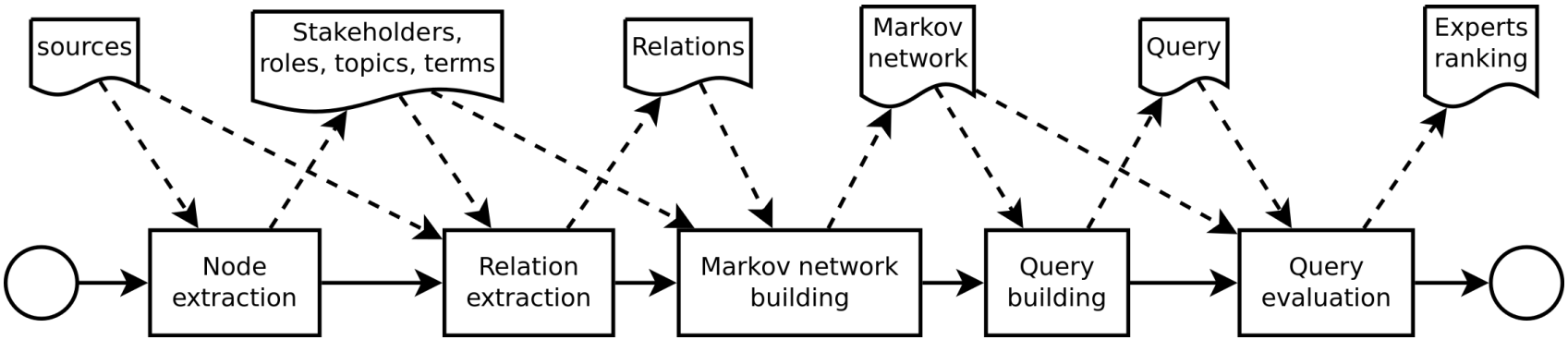
- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$

Full Process



- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$
- Querying $\neq$ filtering: $P(s = \top | r_{developer} = \top)$ does not rank only developers.

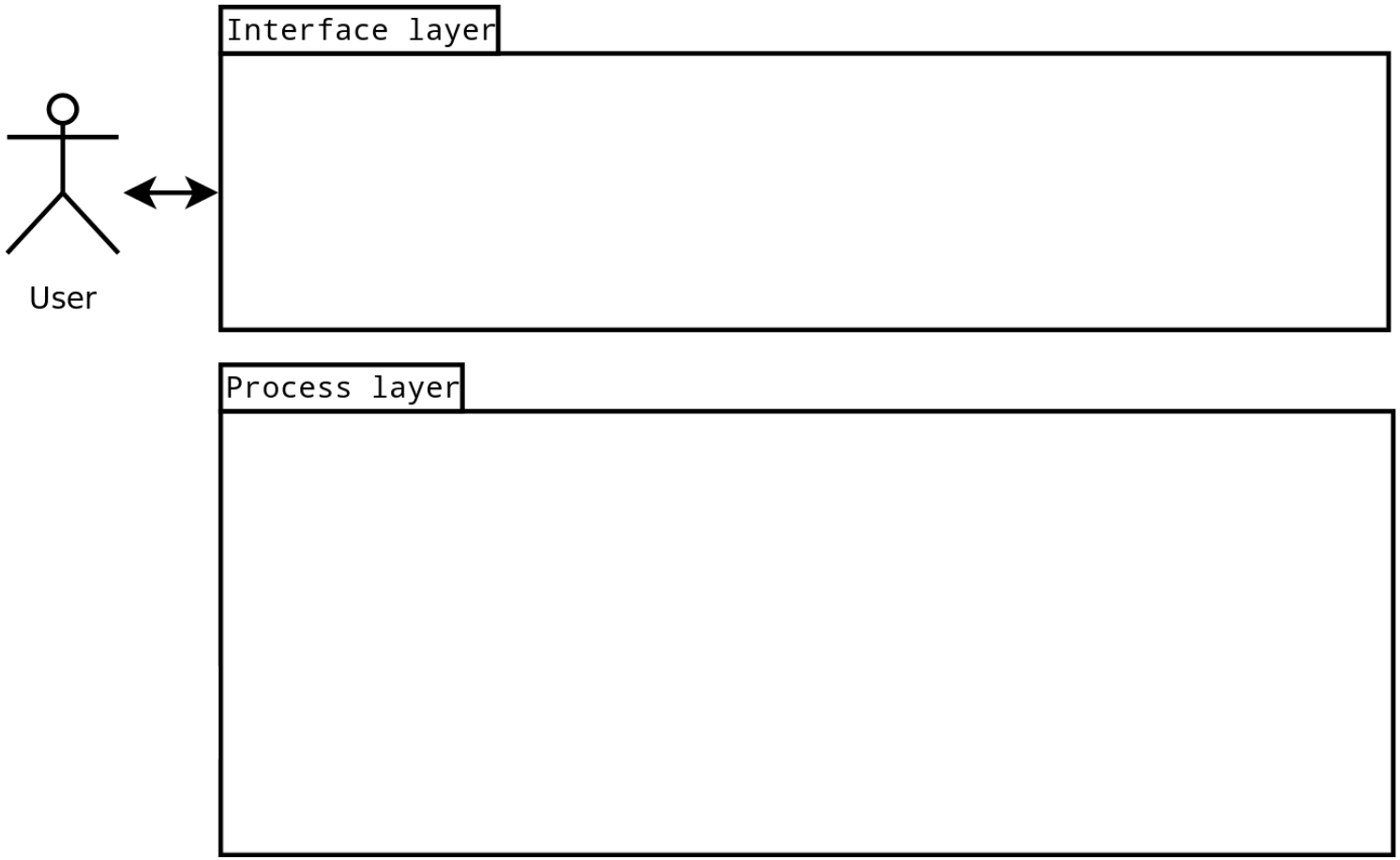
Full Process



- Sources: forum posts, e-mails, reports, goal-models, social networks, etc.
- Context-specific: sources, node/relation extractors
- Data merging: $\langle x, y, w_1 \rangle + \langle x, y, w_2 \rangle = \langle x, y, w_1 + w_2 \rangle$
- Querying $\neq$ filtering: $P(s = \top | r_{developer} = \top)$ does not rank only developers.

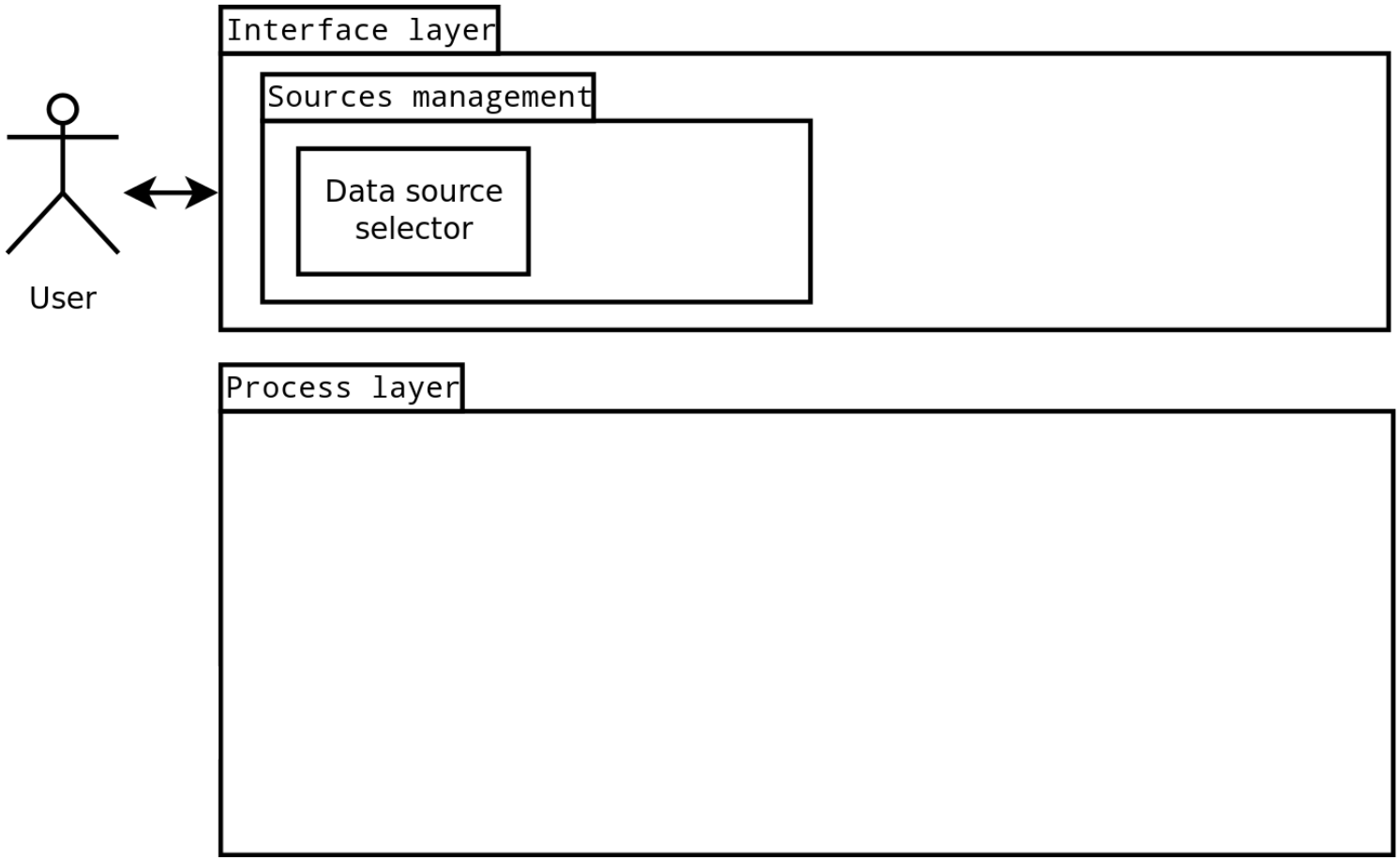
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



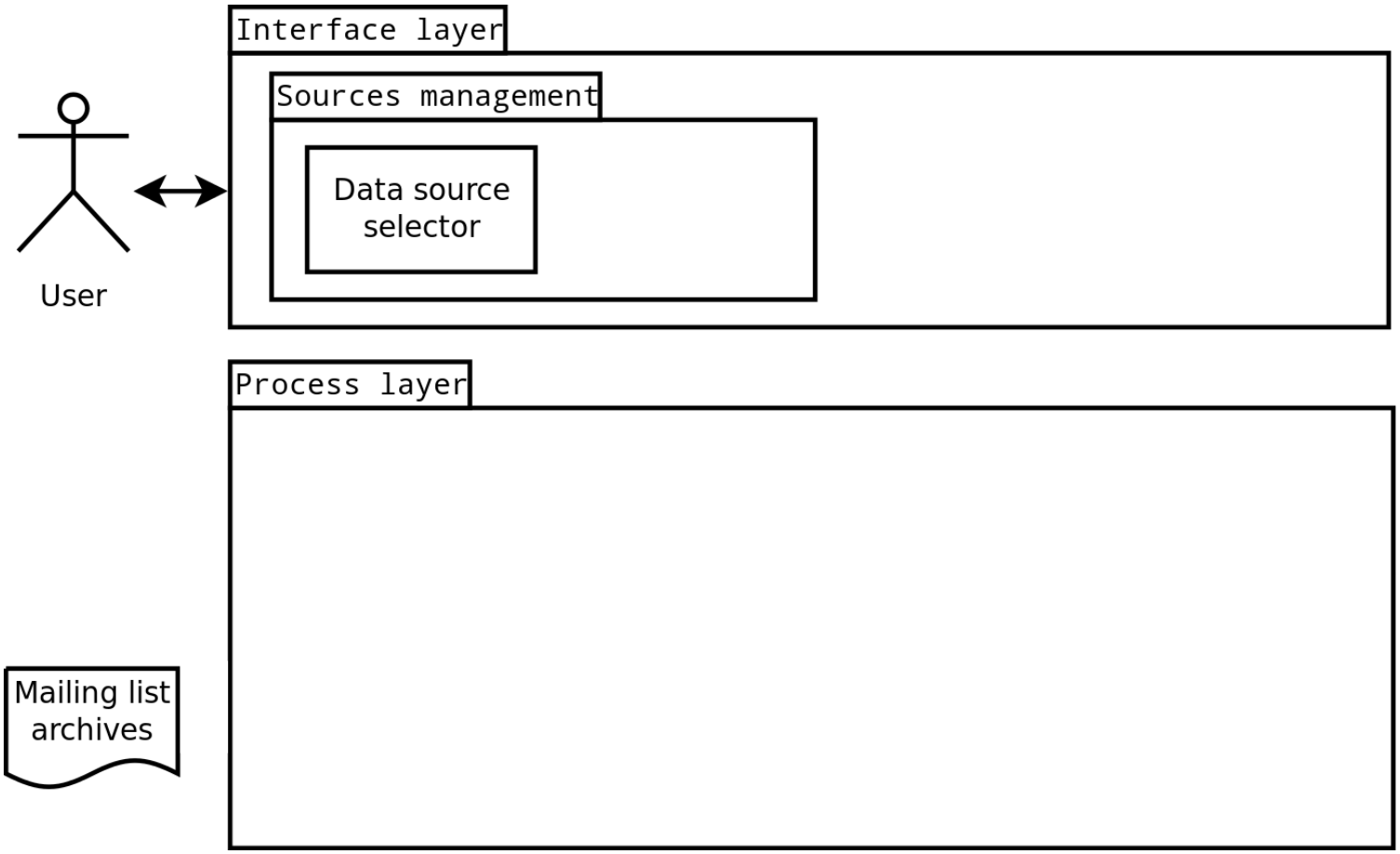
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



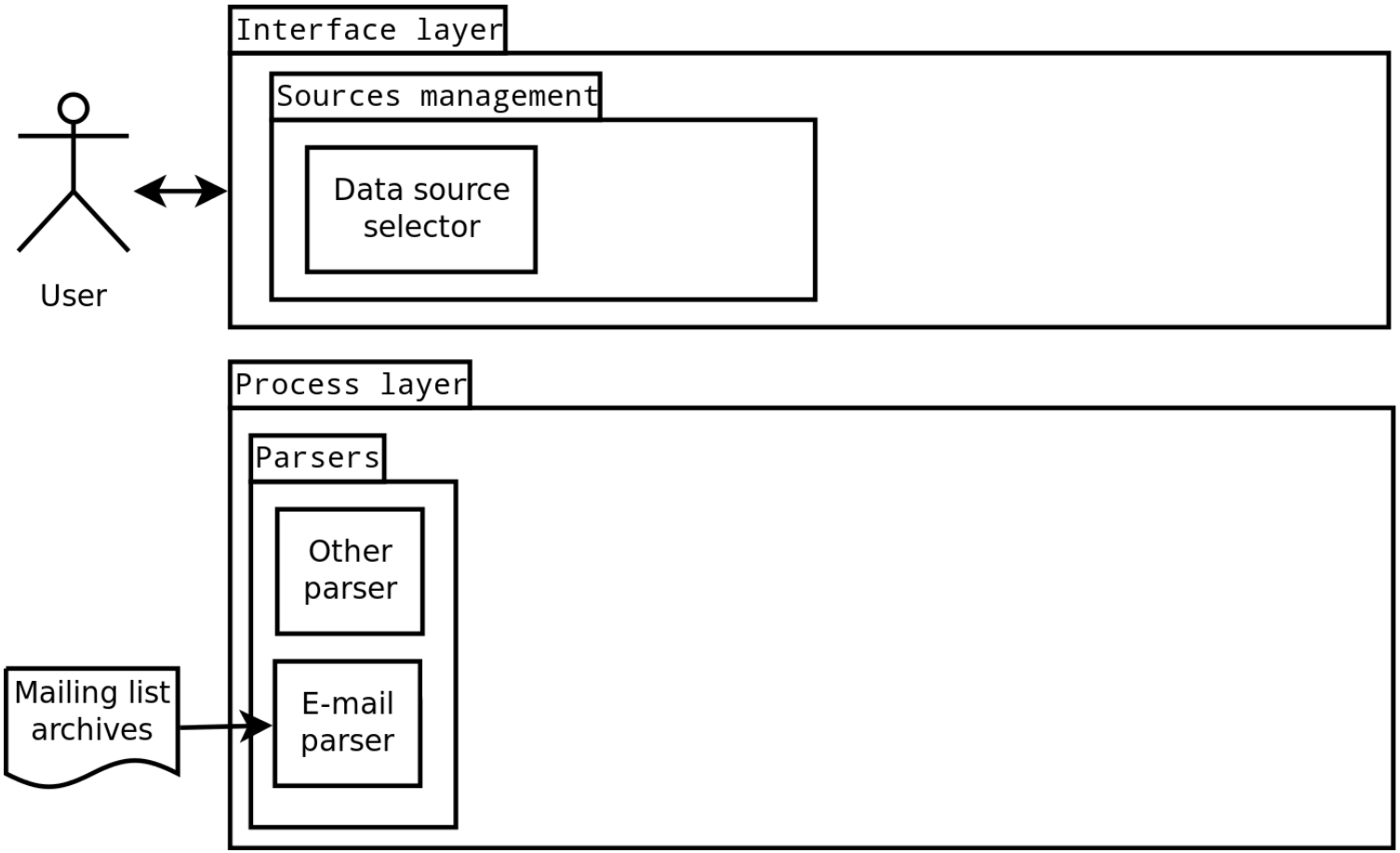
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



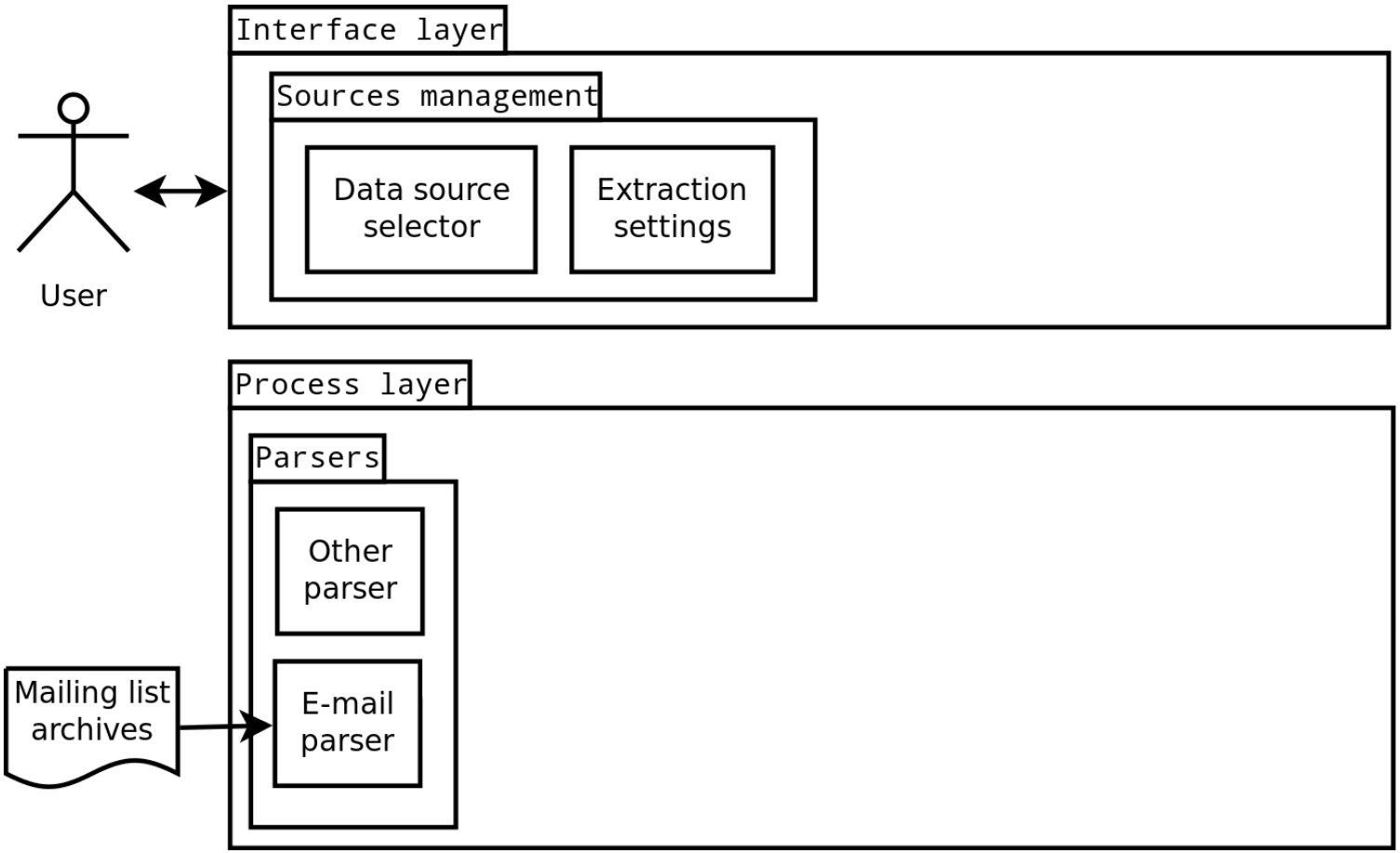
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



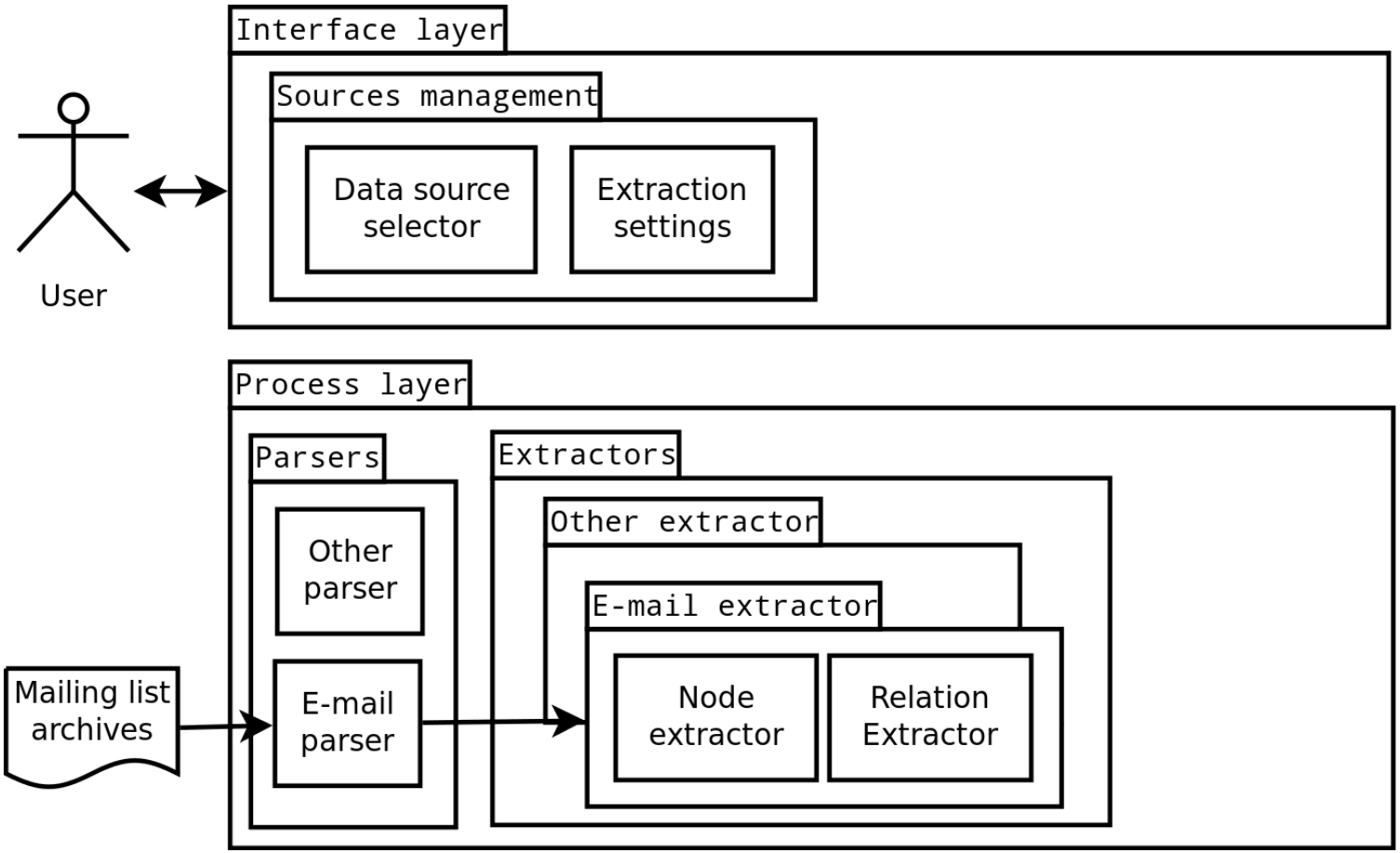
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



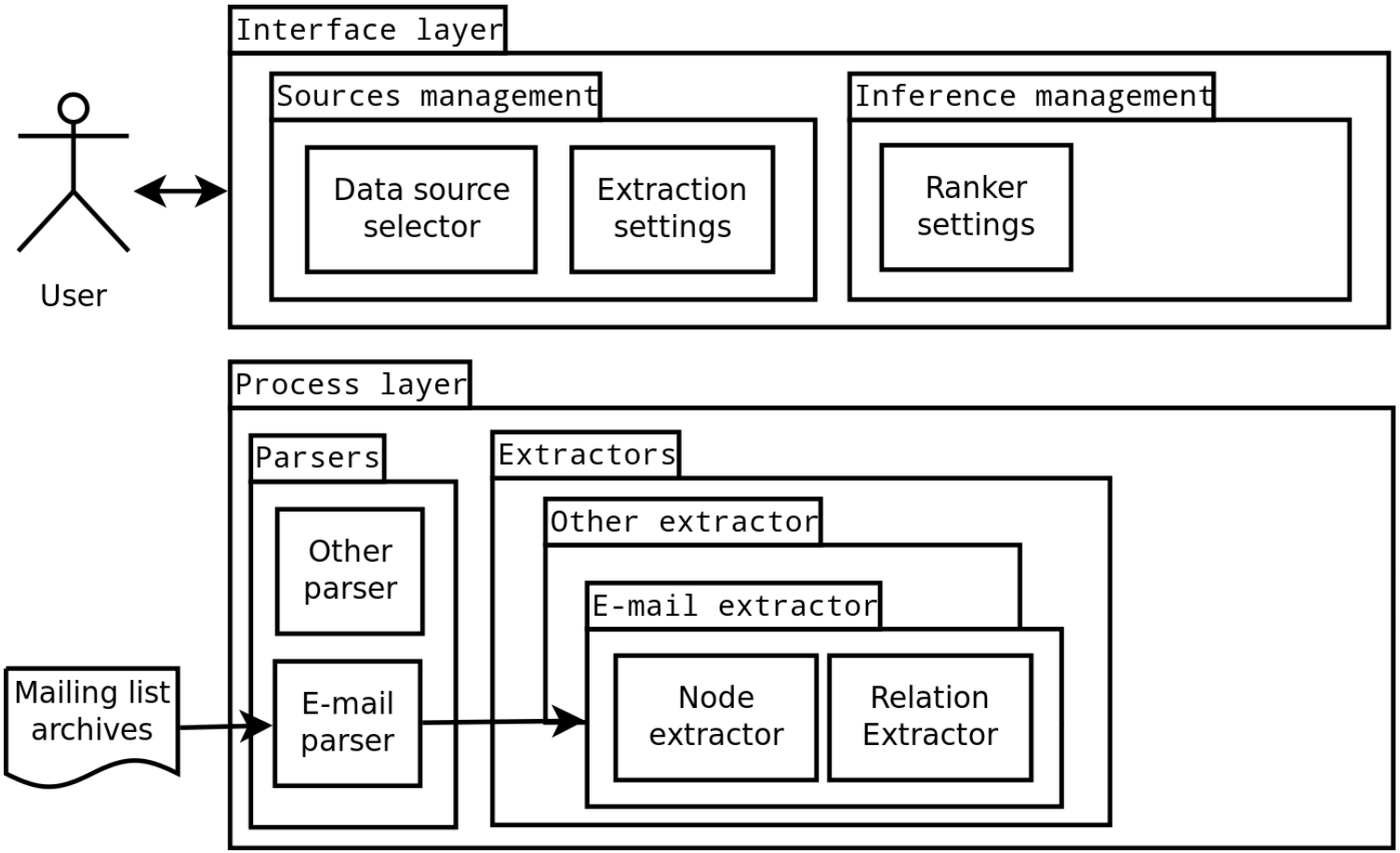
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



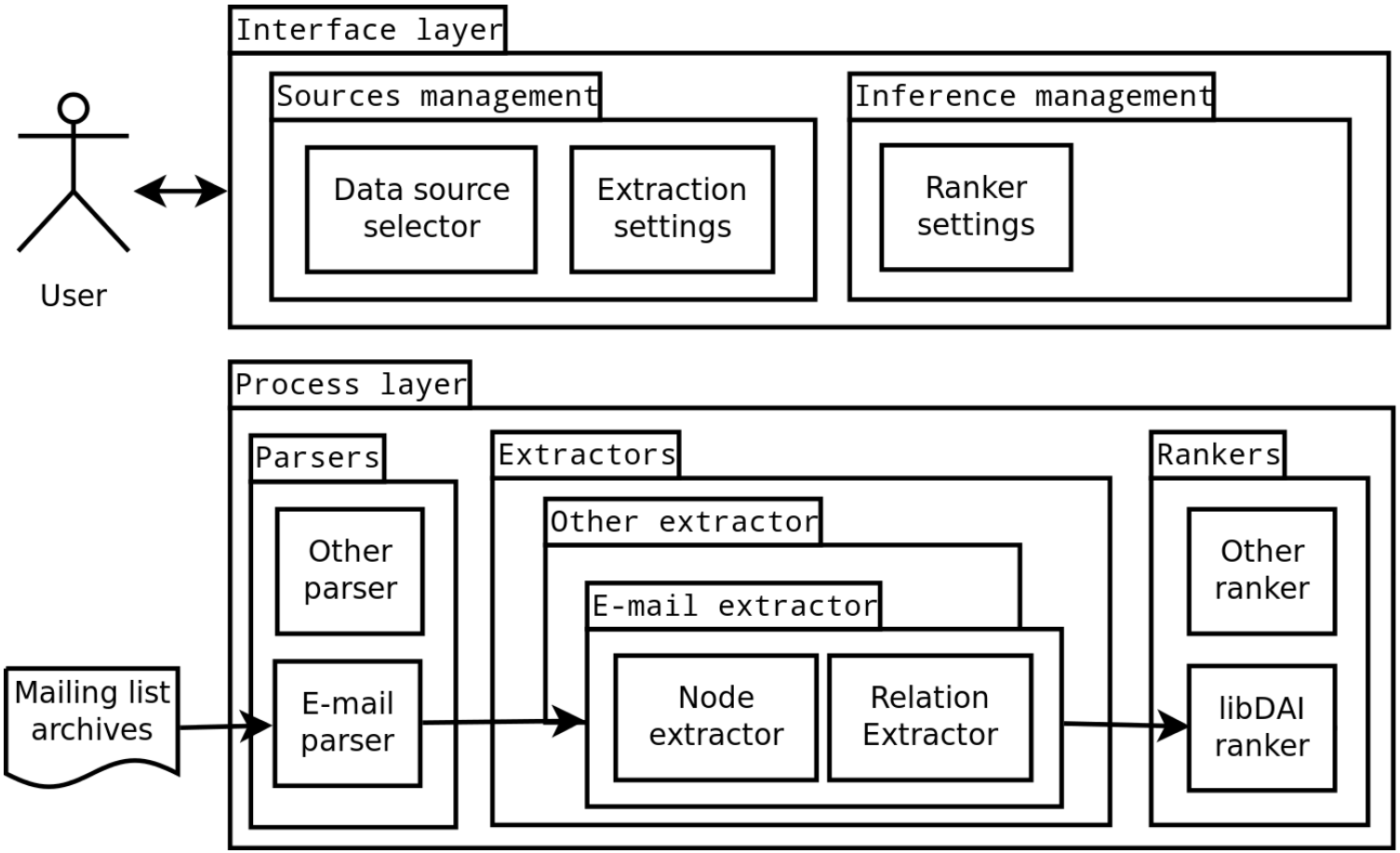
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



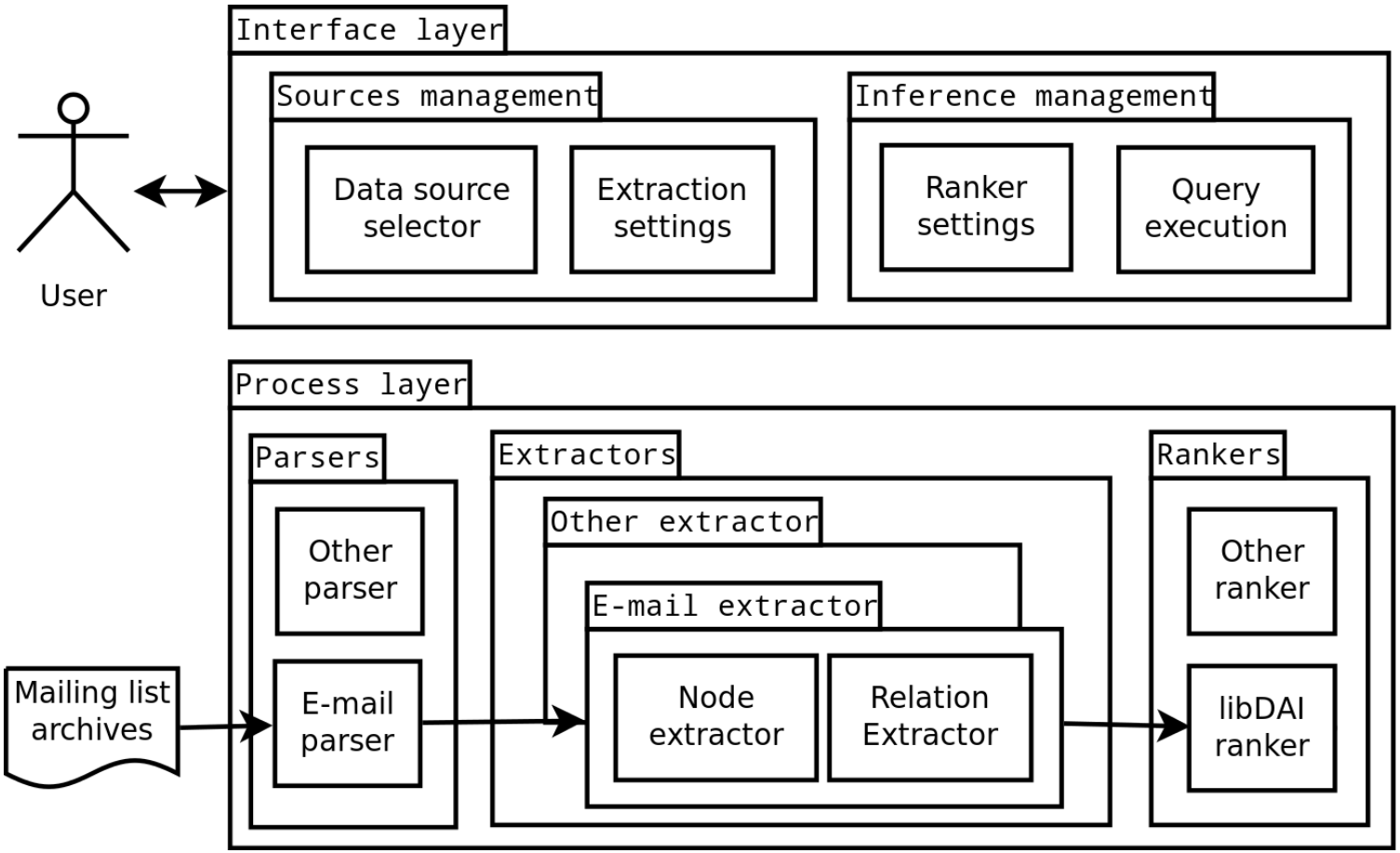
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



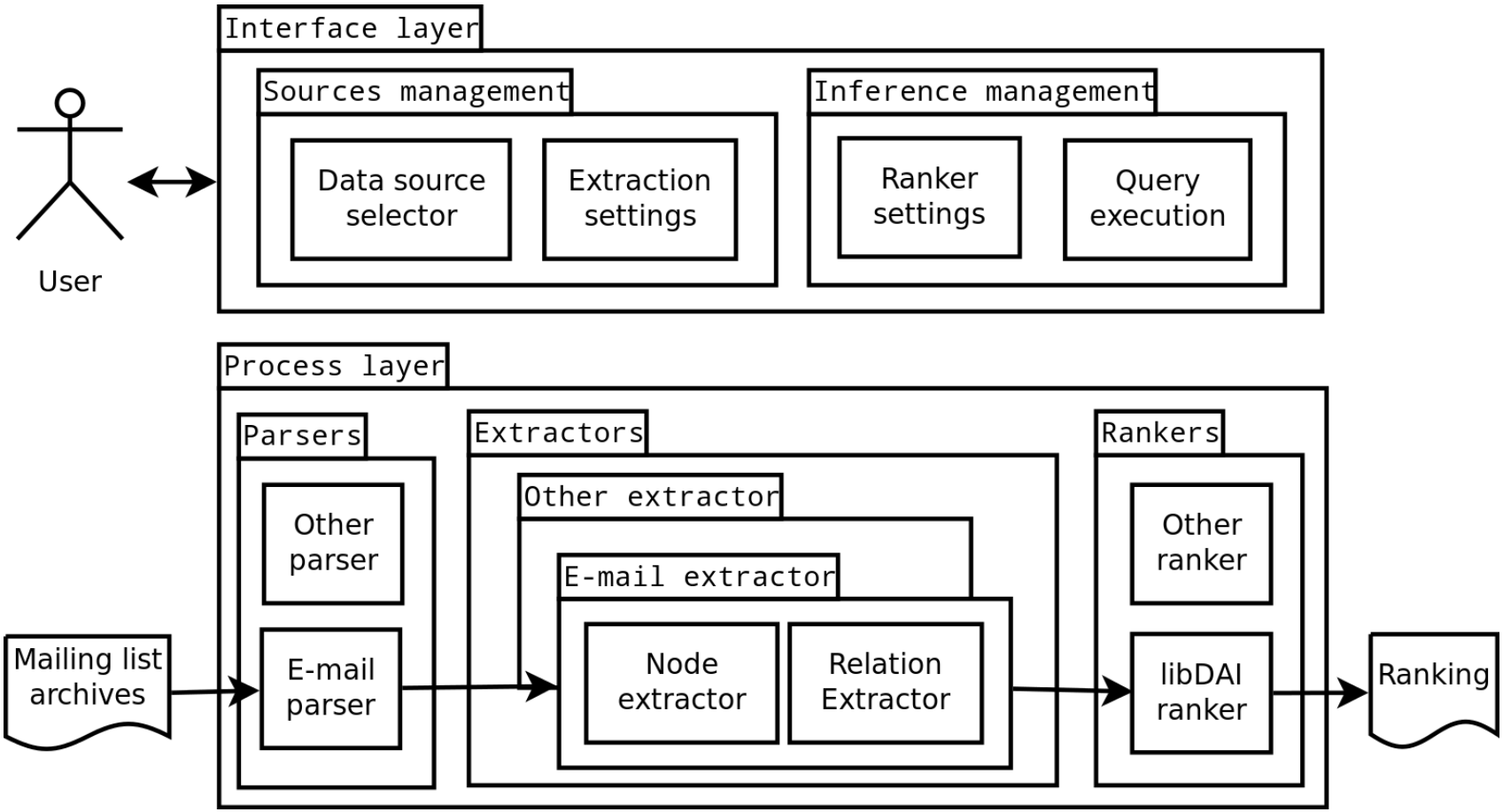
Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



Implementation

Coded in Java, uses GATE (NL) and libDai (MN).



Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion
- Nodes: S=author, C=message nouns, T=title nouns
  - *No role but for MN all the same (same weights ⇒ same results).*

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion
- Nodes: S=author, C=message nouns, T=title nouns
  - *No role but for MN all the same (same weights ⇒ same results).*
- Relations: count *stakeholder* using *term* for specific *topic*

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion
- Nodes: S=author, C=message nouns, T=title nouns
  - *No role but for MN all the same (same weights ⇒ same results).*
- Relations: count *stakeholder* using *term* for specific *topic*

Results:

- Discussions: 30 messages exchanged (balanced participation)

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion
- Nodes: S=author, C=message nouns, T=title nouns
  - *No role but for MN all the same (same weights ⇒ same results).*
- Relations: count *stakeholder* using *term* for specific *topic*

Results:

- Discussions: 30 messages exchanged (balanced participation)
- Extraction: 3 stakeholders, 4 topics, 293 terms, 2k relations

Simple Case: Cooking in Trento

Experiment

- Stakeholders: Alice, Bob and Carla
- Task: (1) Asian food + (2) Europ. dessert e-mail discussions
- Gold Standard: stakeholders' evaluation for each discussion
- Nodes: S=author, C=message nouns, T=title nouns
  - *No role but for MN all the same (same weights ⇒ same results).*
- Relations: count *stakeholder* using *term* for specific *topic*

Results:

- Discussions: 30 messages exchanged (balanced participation)
- Extraction: 3 stakeholders, 4 topics, 293 terms, 2k relations

S	$Q = \emptyset$	Rank	GS	Q_1	Rank	GS	Q_2	Rank	GS
Carla	0.501	1	-	0.499	3	2	0.49978	1	1
Bob	0.500	2	-	0.500	2	1	0.49946	3	2
Alice	0.499	3	-	0.501	1	1	0.49975	2	2

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Results:

- Discussions: 805 e-mails in 255 threads

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Results:

- Discussions: 805 e-mails in 255 threads
- Extraction: 120 S, 216 T, 5k C and 75k relations

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Results:

- Discussions: 805 e-mails in 255 threads
- Extraction: 120 S, 216 T, 5k C and 75k relations
- MN too heavy $\rightarrow$ reduced trials (5 S, 10 T, 100 C)

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Results:

- Discussions: 805 e-mails in 255 threads
- Extraction: 120 S, 216 T, 5k C and 75k relations
- MN too heavy $\rightarrow$ reduced trials (5 S, 10 T, 100 C)
- Different functions tried (Id, norm, prior, WoE, etc.)

Practical Case: XWiki OSS

Experiment

- Discussions: XWiki mailing list, Jan-May 2013
  - *OSS community (`http://dev.xwiki.org`) + company*
  - *Cleaning: unique authors + e-mails noise removal (quotations)*
- No Gold Standard, but obvious experts
  - *XWiki team, main topic-related participants*
- Nodes and Relations: same algos

Results:

- Discussions: 805 e-mails in 255 threads
- Extraction: 120 S, 216 T, 5k C and 75k relations
- MN too heavy $\rightarrow$ reduced trials (5 S, 10 T, 100 C)
- Different functions tried (Id, norm, prior, WoE, etc.)
- Obvious experts in top ranks for their topics

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited
- Querying improvements (e.g. weights)

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited
- Querying improvements (e.g. weights)
- Probabilities closeness: when differentiate?

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited
- Querying improvements (e.g. weights)
- Probabilities closeness: when differentiate?
- Different functions could improve difference

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited
- Querying improvements (e.g. weights)
- Probabilities closeness: when differentiate?
- Different functions could improve difference
- Merging policies for different sources (e.g. trust)

Results:

- Experiments results suffer from restricted context or preliminary state, but provide support
- XWiki suited for validation, but need improvements (cleaning, scalability)
- No use of roles, but XWiki's Hall of Fame + organizational models can be exploited

Approach:

- MN scalability $\rightarrow$ use approximation
- No relation for similar nodes, but could be exploited
- Querying improvements (e.g. weights)
- Probabilities closeness: when differentiate?
- Different functions could improve difference
- Merging policies for different sources (e.g. trust)
- Taxonomy and ontologies

Take Away Message

- Goal: expert ranking for RE

Take Away Message

- Goal: expert ranking for RE
- Idea: combine content-based and social-based perspectives

Take Away Message

- Goal: expert ranking for RE
- Idea: combine content-based and social-based perspectives
- Technique: use MN to compute expertise probability

Take Away Message

- Goal: expert ranking for RE
- Idea: combine content-based and social-based perspectives
- Technique: use MN to compute expertise probability
- Results: good support from preliminary experiment

M. Vergne, A. Susi (vergne@fbk.eu, susi@fbk.eu) FBK - ICT Doctoral School Trento

Thanks for your attention.

Questions?

M. Vergne, A. Susi (vergne@fbk.eu, susi@fbk.eu) FBK - ICT Doctoral School Trento

Expert Finding Examples

content-based:

- Mockus and Herbsleb [2002]: written code evidence to evaluate expertise in software pieces.
- Serdyukov and Hiemstra [2008]: authors' contributions in documents to infer expertise in related topics.

social-based:

- Zhang et al. [2007]: compare algos on social network built from askers/repliers identification in online forums.

Both:

- Karimzadehgan et al. [2009] exploit relationships + e-mails content between employees of a company.

Stakeholders Recommendation in RE

Literature review Mohebzada et al. [2012]

- Castro-Herrera and Cleland-Huang [2010]
  - *evaluate stakeholders knowledge through participation in forum*
  - *build abstract topics (term vectors) depending on messages common terms*
  - *relate stakeholders to topics they participate in*
  - *recommend other stakeholders to participate in new, similar topics*
  - *content-based: exploit data provided directly by stakeholders*
- StakeNet Lim et al. [2010]
  - *prioritise requirements depending on stakeholders rating*
  - *core stakeholders suggests others influencing the project*
  - *role and salience describe influence*
  - *built social network + apply measures to evaluate global influence*
  - *social-based: evaluate influence based on other stakeholders suggestions*

Node extractor for e-mails

Require: *mail*: Natural language e-mail
Ensure: S, R, T, C: Extracted stakeholders, roles, topics and terms

1: $S \leftarrow \{stakeholder(authorOf(mail))\}$
2: $R \leftarrow \emptyset$
3: $T \leftarrow \{topic(x) | x \in nounsOf(subjectOf(mail))\}$
4: $C \leftarrow \{term(x) | x \in nounsOf(bodyOf(mail))\}$

Relation extractor for e-mails

Require: $mail, S, R, T, C$: E-mail, stakeholders, roles, topics, terms
Ensure: L: Weighted relations

$L \leftarrow \emptyset$
$a \leftarrow author(mail)$
if $stakeholder(a) \in S$ **then**
  for all $t \in termsOf(bodyOf(mail))$ **do**
    if $term(t) \in C$ **then**
      $L \leftarrow merge(L, \{\langle stakeholder(a), term(t), 1\rangle\})$
    end if
  end for
end if
for all $topic \in T$ **do**
  if $nounOf(topic) \in nounsOf(subjectOf(mail))$ **then**
    $L \leftarrow merge(L, \{\langle stakeholder(a), topic, 1\rangle\})$
    for all $t \in nounsOf(bodyOf(mail))$ **do**
      if $term(t) \in C$ **then**
        $L \leftarrow merge(L, \{\langle topic, term(t), 1\rangle\})$
      end if
    end for
  end if
end for

Main differences with Karimzadehgan et al. [2009]

- social relationships not only for post-processing
- inference technique manage multiple topics

C. Castro-Herrera and J. Cleland-Huang. Utilizing recommender systems to support software requirements elicitation. In Proc. of the 2nd International Workshop on RSSE, pages 6–10, New York, USA, 2010. ACM. ISBN 978-1-60558-974-9. doi: 10.1145/1808920.1808922. URL http://doi.acm.org/10.1145/1808920.1808922.

K. A. Ericsson. The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, June 2006. ISBN 9781139456463.

M. Karimzadehgan, R. W. White, and M. Richardson. Enhancing expert finding using organizational hierarchies. In Advances in Information Retrieval, number 5478 in LNCS, pages 177–188. Springer Berlin Heidelberg, Jan. 2009. ISBN 978-3-642-00957-0, 978-3-642-00958-7. URL http://link.springer.com/chapter/10.1007/978-3-642-00958-7_18.

S. L. Lim, D. Quercia, and A. Finkelstein. StakeNet: using social

networks to analyse the stakeholders of large-scale software projects. In Proc. of the 32nd ACM/IEEE ICSE, volume 1, pages 295–304, New York, USA, 2010. ACM. ISBN 978-1-60558-719-6. doi: http://doi.acm.org/10.1145/1806799.1806844. URL http://doi.acm.org/10.1145/1806799.1806844.

A. Mockus and J. D. Herbsleb. Expertise browser: a quantitative approach to identifying expertise. In Proc. of the 24th ICSE, page 503–512, New York, USA, 2002. ACM. ISBN 1-58113-472-X. doi: 10.1145/581339.581401. URL http://doi.acm.org/10.1145/581339.581401.

J. Mohebzada, G. Ruhe, and A. Eberlein. Systematic mapping of recommendation systems for requirements engineering. In ICSSP, pages 200 –209, June 2012. doi: 10.1109/ICSSP.2012.6225965.

P. Serdyukov and D. Hiemstra. Modeling documents as mixtures

of persons for expert finding. In Proc. of the IR Research, 30th ECIR, page 309–320, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-78645-7, 978-3-540-78645-0. URL http://dl.acm.org/citation.cfm?id=1793274.1793313.

J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In Proc. of the 16th international conference on WWW, page 221–230, New York, USA, 2007. ACM. ISBN 978-1-59593-654-7. doi: 10.1145/1242572.1242603. URL http://doi.acm.org/10.1145/1242572.1242603.