# Goal Model-Based Expert Finding

Matthieu Vergne
vergne@fbk.eu

Fondazione Bruno Kessler

Information and Communication Technology
University of Trento


April 17th, 2013

# Outline

# Expert Finding in RE

- RE $\Rightarrow$ Human- and knowledge-intensive process.

# Expert Finding in RE

- RE $\Rightarrow$ Human- and knowledge-intensive process.
- Broad mastering infeasible for a single person.
  - *Stakeholder or analyst.*

# Expert Finding in RE

- RE $\Rightarrow$ Human- and knowledge-intensive process.
- Broad mastering infeasible for a single person.
  - *Stakeholder or analyst.*
- Need to elicit new requirements or refine existing one.

# Expert Finding in RE

- RE $\Rightarrow$ Human- and knowledge-intensive process.
- Broad mastering infeasible for a single person.
  - *Stakeholder or analyst.*
- Need to elicit new requirements or refine existing one.
- Need to find "people who know": expert finding.

# Expert Finding in RE

- RE $\Rightarrow$ Human- and knowledge-intensive process.
- Broad mastering infeasible for a single person.
  - *Stakeholder or analyst.*
- Need to elicit new requirements or refine existing one.
- Need to find "people who know": expert finding.
  - *Actually "people who know **the best**".*

# Two Main Approaches (Mohebzada et al. [2012])

Lim et al. [2010, 2011], Lim and Finkelstein [2012]

- Based on stakeholders' recommendations.
- Relate stakeholders to *roles*.
- Weight regarding *salience* (i.e. "influence") and social network measures.
- *Role* interesting, but *salience* abstract.
- Superficial information on people's knowledge.

# Two Main Approaches (Mohebzada et al. [2012])

Lim et al. [2010, 2011], Lim and Finkelstein [2012]

- Based on stakeholders' recommendations.
- Relate stakeholders to *roles*.
- Weight regarding *salience* (i.e. "influence") and social network measures.
- *Role* interesting, but *salience* abstract.
- Superficial information on people's knowledge.

Castro-Herrera and Cleland-Huang [2009, 2010]

- Based on stakeholders' contributions in a forum.
- Identify *topics* through their common *terms*.
- Relate stakeholders to the *topics*.
- Deep information, but restricted to contributors.

# Research Problem

**Lack**

Poor identification of relevant experts for RE: too closed (contributors) or too abstract (role + salience).

# Research Problem

## Lack

Poor identification of relevant experts for RE: too closed (contributors) or too abstract (role + salience).

## Research Objective

Improve expert finding in RE by designing a more comprehensive approach, inspiring from existing ones.

# Research Problem

**Lack**

Poor identification of relevant experts for RE: too closed (contributors) or too abstract (role + salience).

**Research Objective**

Improve expert finding in RE by designing a more comprehensive approach, inspiring from existing ones.

**Principal inspiration:** consider *roles*, *topics* and *concepts* (prev. *terms*) as the basic blocks to describe the knowledge of the stakeholders.

# Research Problem

**Lack**

Poor identification of relevant experts for RE: too closed (contributors) or too abstract (role + salience).

**Research Objective**

Improve expert finding in RE by designing a more comprehensive approach, inspiring from existing ones.

**Principal inspiration:** consider *roles*, *topics* and *concepts* (prev. *terms*) as the basic blocks to describe the knowledge of the stakeholders.

**Actual focus:** identify the relevant blocks via the goal model and exploit other sources to relate stakeholders to these blocks.

# Global Process

1. Extract *roles*, *topics* and *concepts* from the GM.

# Global Process

1. Extract *roles*, *topics* and *concepts* from the GM.
2. Relate them considering their inter-dependencies in the GM.

# Global Process

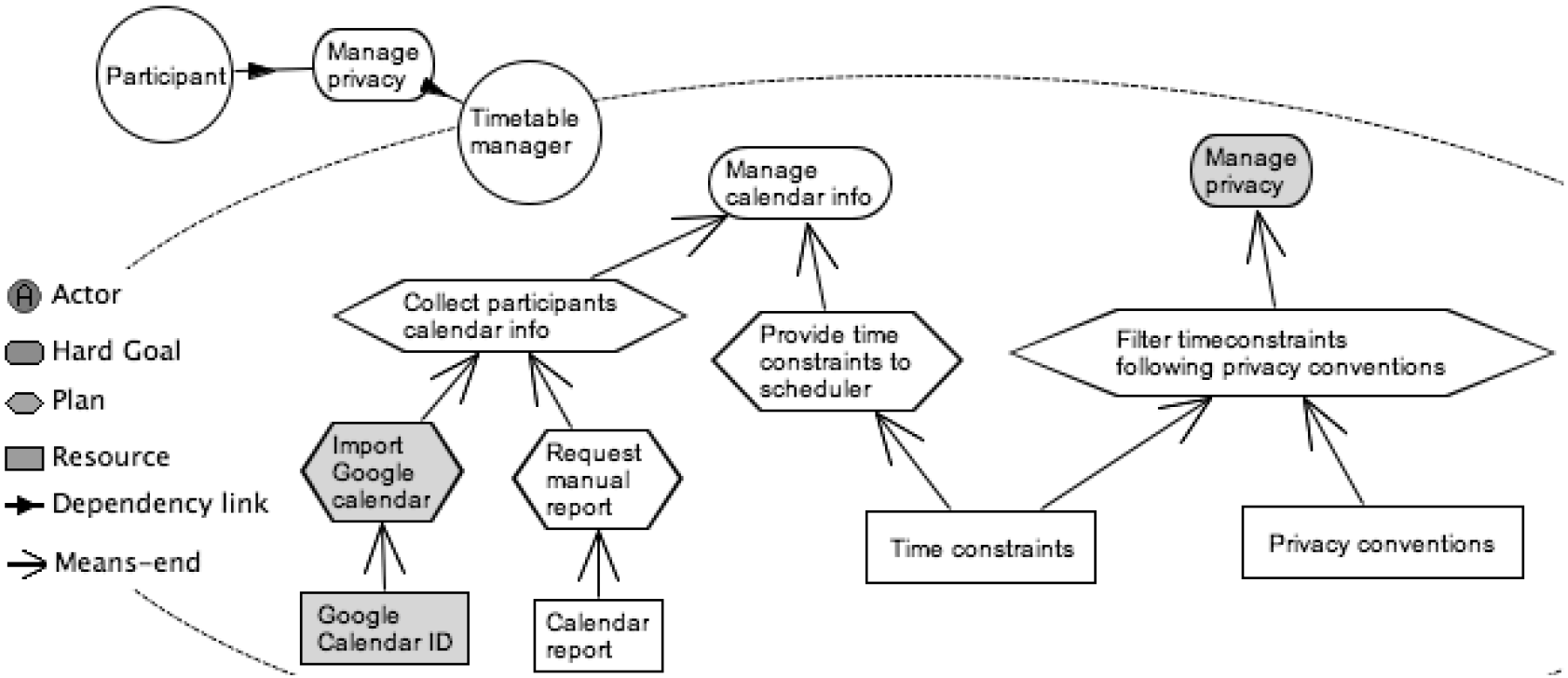
1. Extract *roles*, *topics* and *concepts* from the GM.
2. Relate them considering their inter-dependencies in the GM.
3. Relate stakeholders to them using other sources (e.g. forum, recommendations).

# Global Process

1. Extract *roles*, *topics* and *concepts* from the GM.
2. Relate them considering their inter-dependencies in the GM.
3. Relate stakeholders to them using other sources (e.g. forum, recommendations).
4. Rank the stakeholders depending on the portion of the GM that the analyst is focusing on.

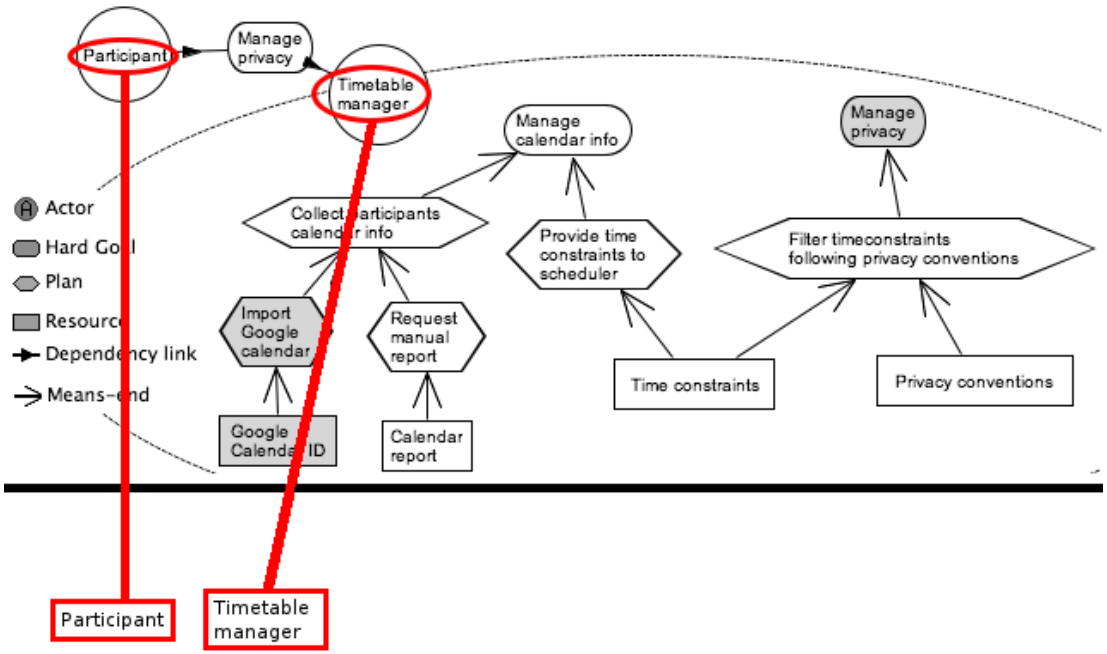
# Goal Model Analysis: Roles, Topics, Concepts

From the goal model...



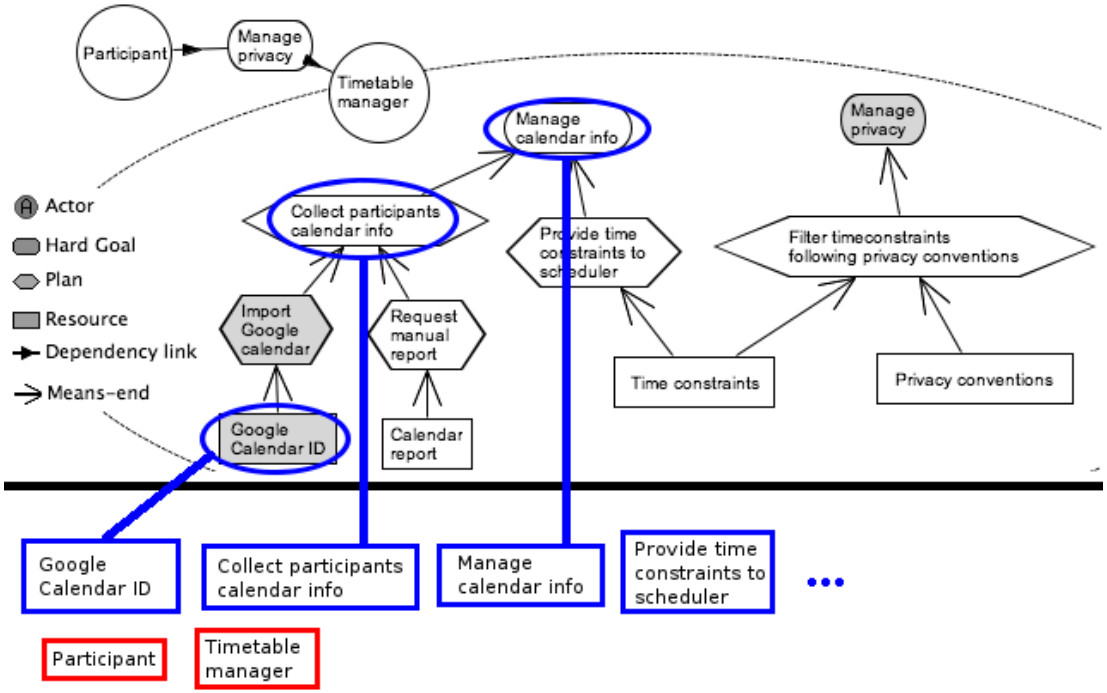
# Goal Model Analysis: Roles, Topics, Concepts

1) Extract roles:



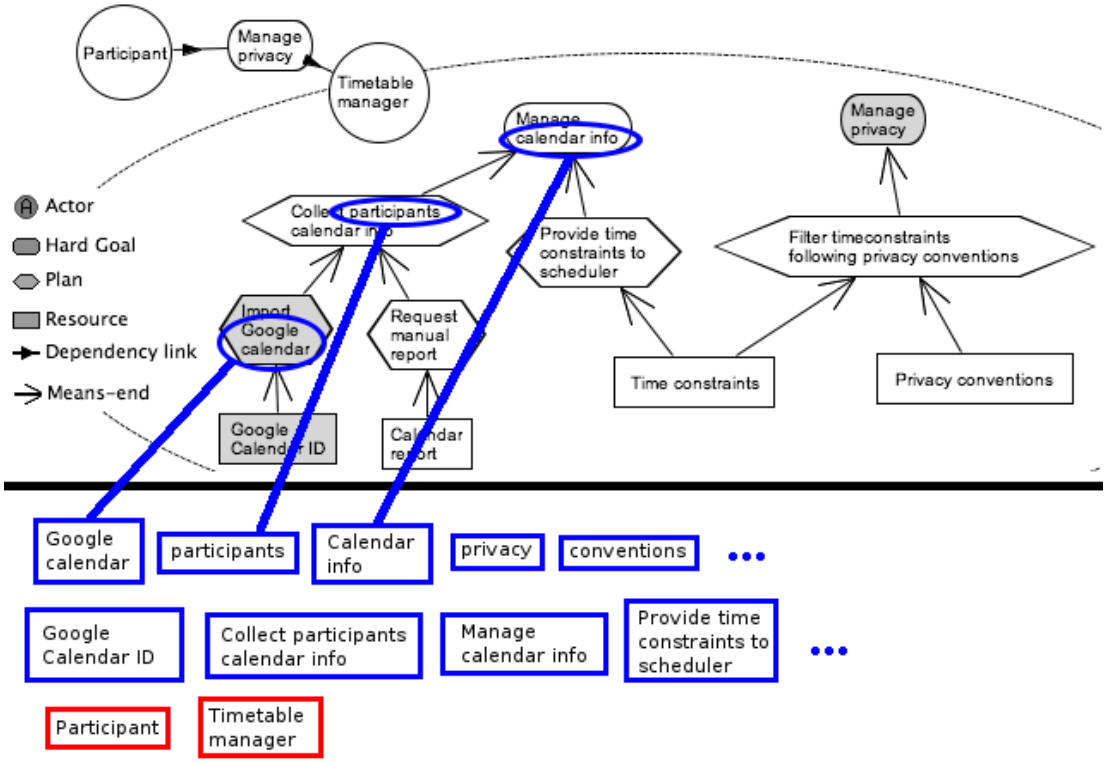
# Goal Model Analysis: Roles, Topics, Concepts

2) Extract concepts (1/2):



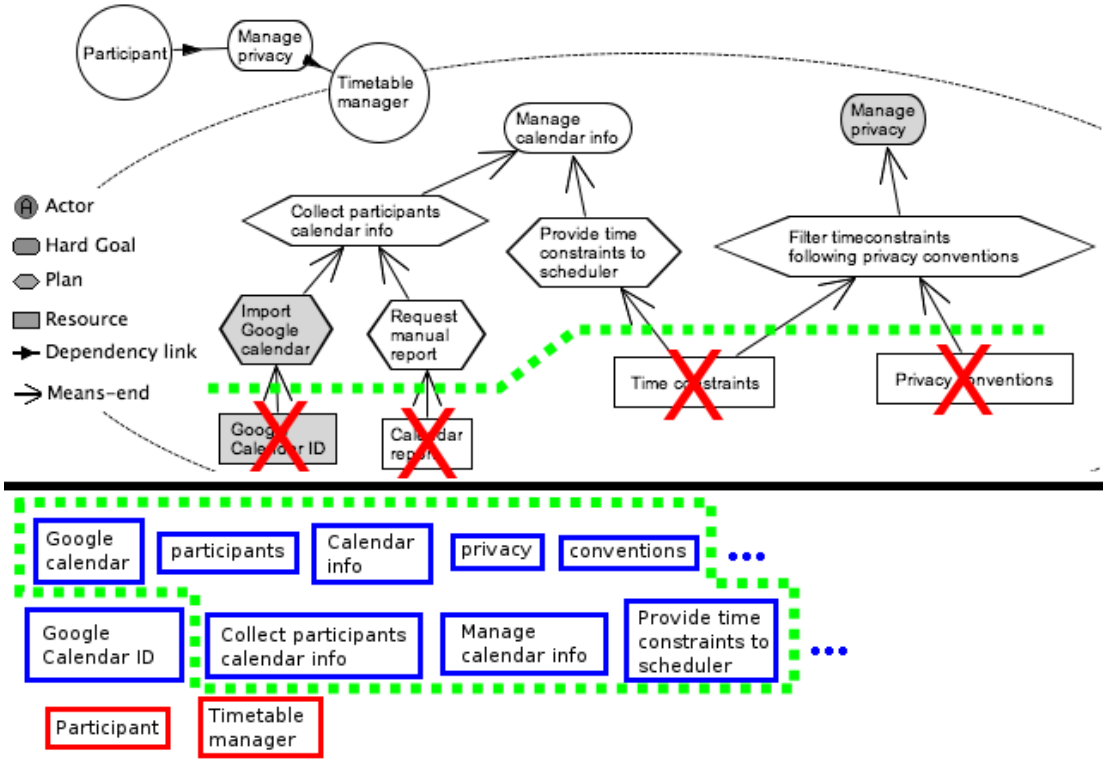
# Goal Model Analysis: Roles, Topics, Concepts

2) Extract concepts (2/2):



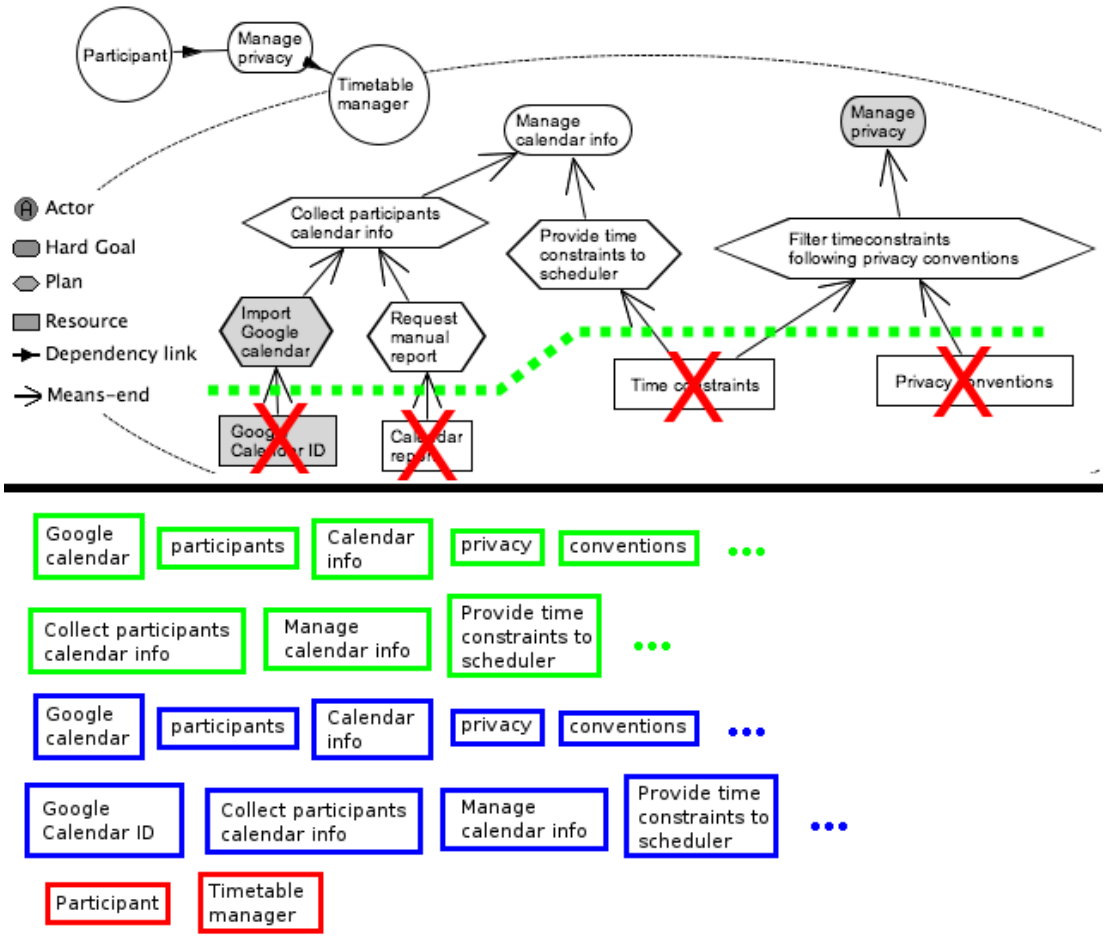
# Goal Model Analysis: Roles, Topics, Concepts

3) Extract topics (1/2):



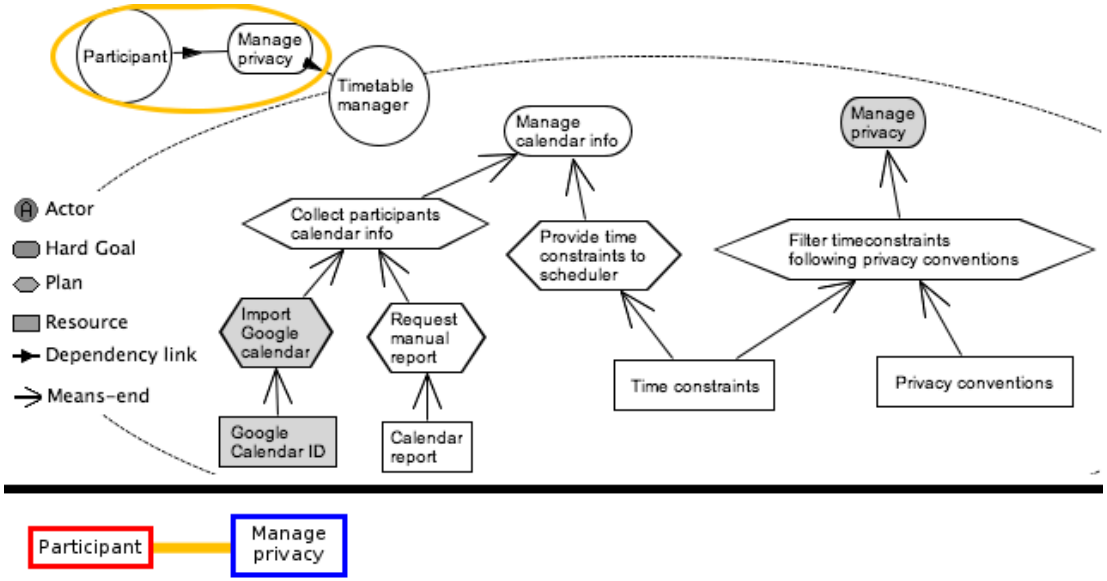
# Goal Model Analysis: Roles, Topics, Concepts

3) Extract topics (2/2):



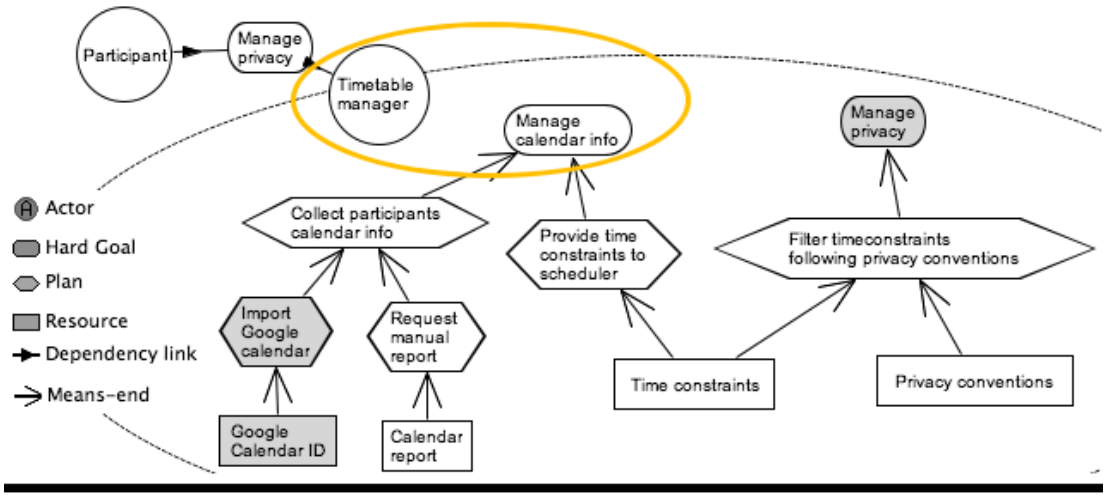
# Goal Model Analysis: Roles, Topics, Concepts

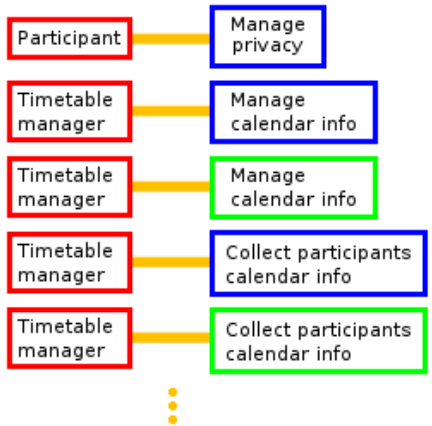
4) Extract relations (1/4):



# Goal Model Analysis: Roles, Topics, Concepts

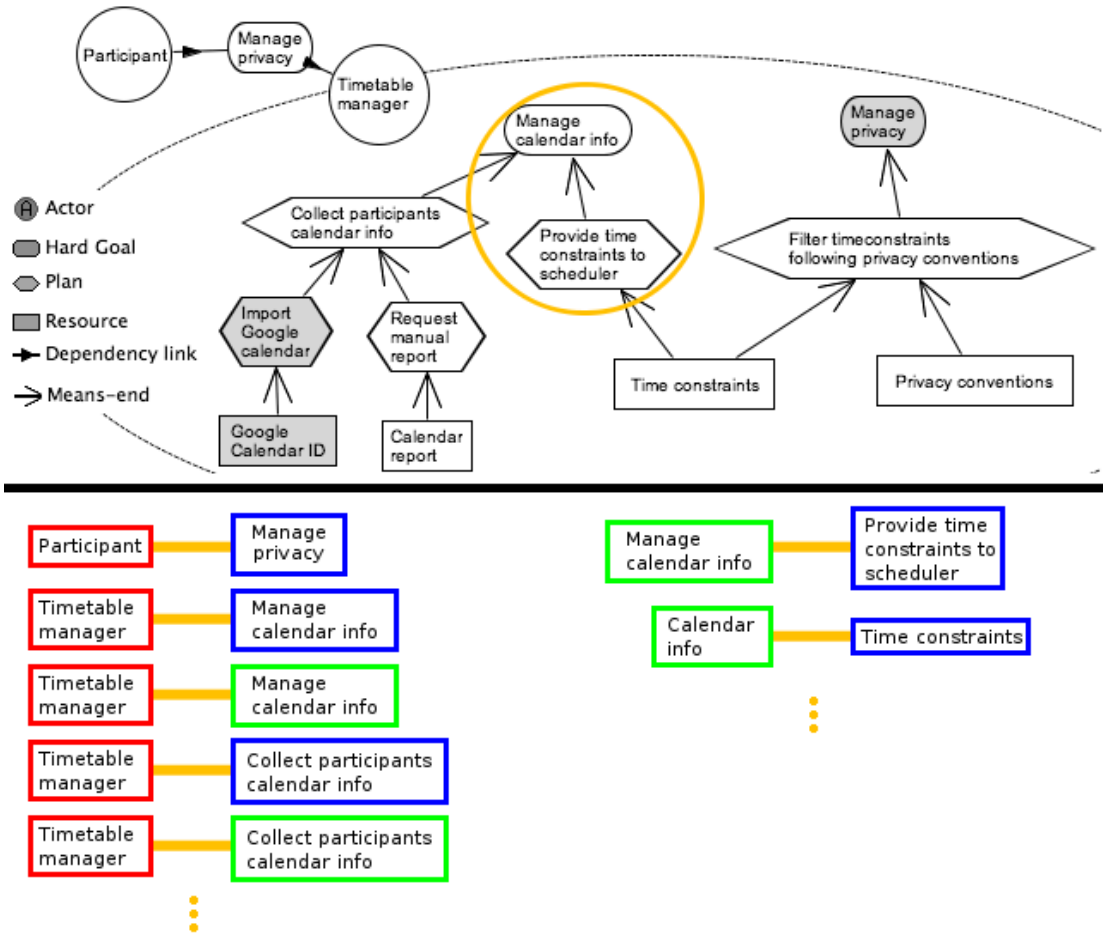
4) Extract relations (2/4):



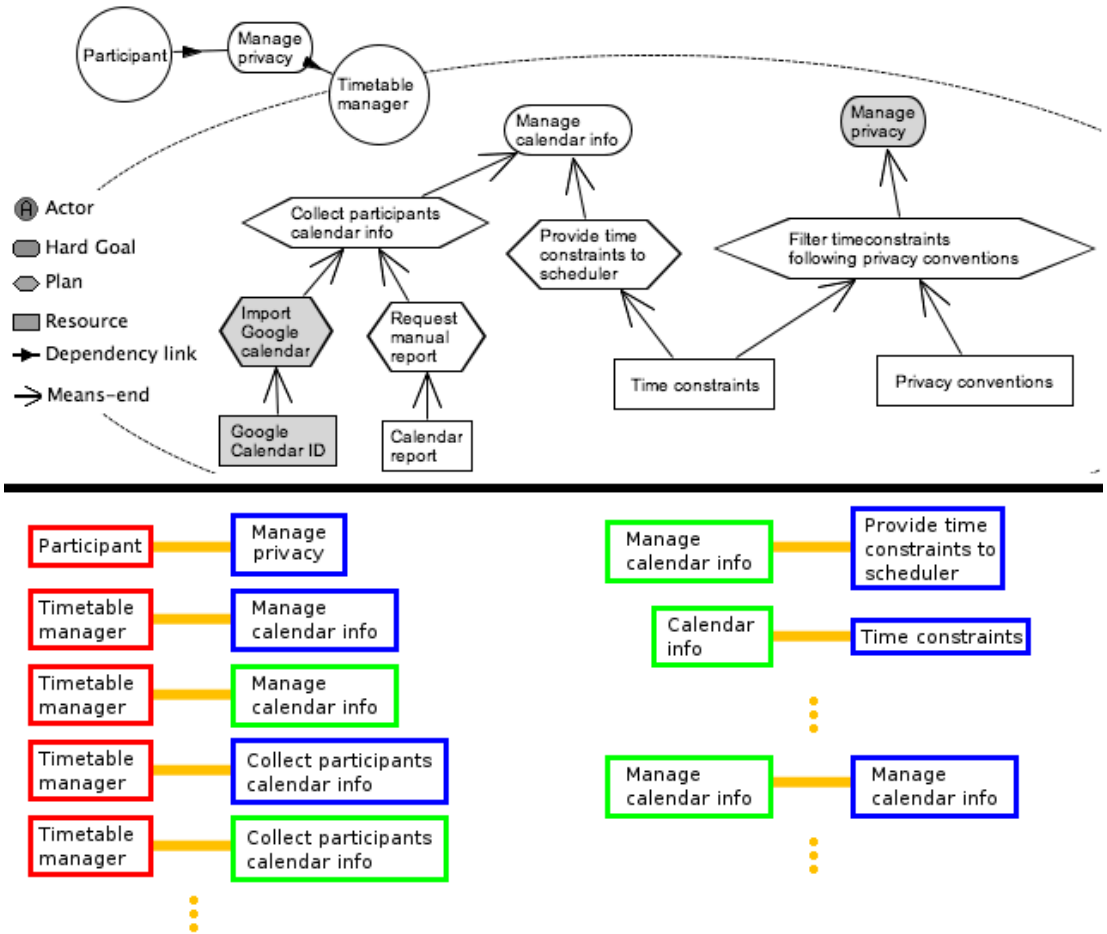
# Goal Model Analysis: Roles, Topics, Concepts

4) Extract relations (3/4):



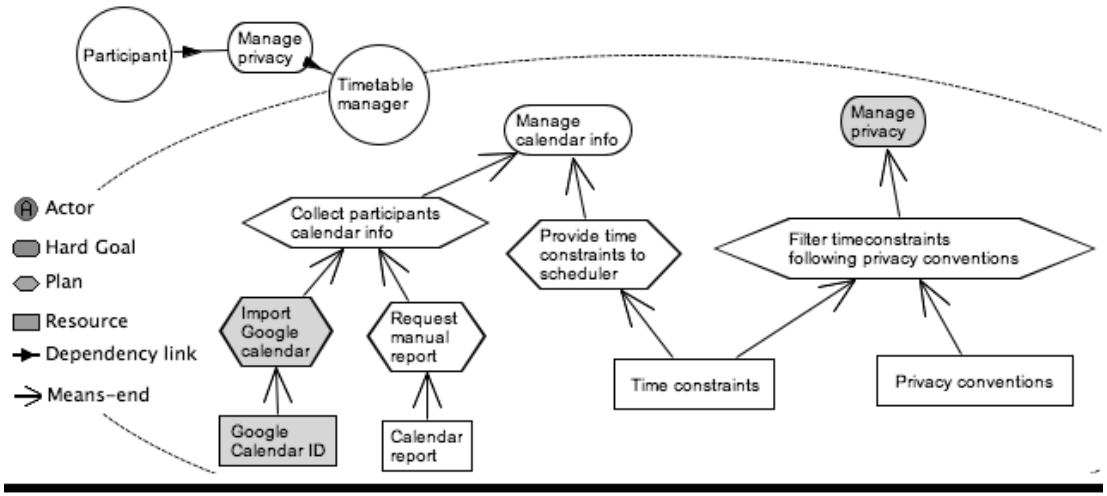
# Goal Model Analysis: Roles, Topics, Concepts

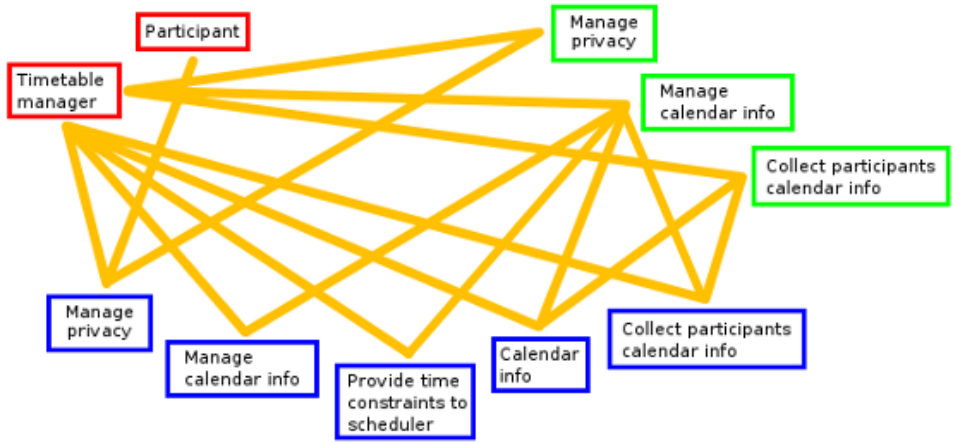
4) Extract relations (4/4):



# Goal Model Analysis: Roles, Topics, Concepts

5) Build the graph:



# Complete the Graph: Stakeholders

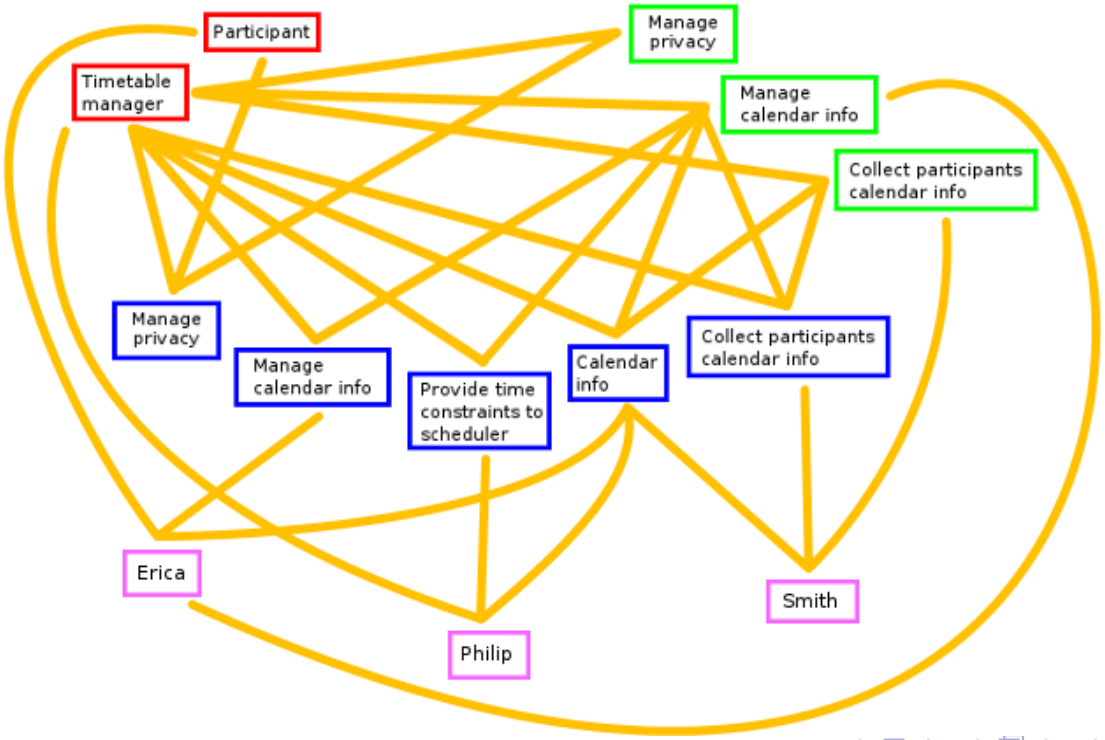
Exploit sources relating stakeholders to the blocks (nodes):

- Forum: stakeholder-concept, stakeholder-topic
- Recommendations: stakeholder-role
- List of employees: stakeholder-role

# Complete the Graph: Stakeholders

Exploit sources relating stakeholders to the blocks (nodes):

- Forum: stakeholder-concept, stakeholder-topic
- Recommendations: stakeholder-role
- List of employees: stakeholder-role



# Improve the Graph: Amount of Information

More a relation appear, stronger it is

- stakeholder represents +/- a role.
- stakeholder know +/- about a topic.
- stakeholder know +/- a concept.

# Improve the Graph: Amount of Information

More a relation appear, stronger it is

- stakeholder represents +/- a role.
- stakeholder know +/- about a topic.
- stakeholder know +/- a concept.

More a source is up-to-date, stronger its relations are

- Goal model, list of employees: updated vs. obsolete.
- Forum's post, recommendations: recent vs. old.

# Improve the Graph: Amount of Information

More a relation appear, stronger it is

- stakeholder represents +/- a role.
- stakeholder know +/- about a topic.
- stakeholder know +/- a concept.

More a source is up-to-date, stronger its relations are

- Goal model, list of employees: updated vs. obsolete.
- Forum's post, recommendations: recent vs. old.

Open question: How to quantify one source regarding the others?

# Improve the Graph: Amount of Information

More a relation appear, stronger it is

- stakeholder represents +/- a role.
- stakeholder know +/- about a topic.
- stakeholder know +/- a concept.

More a source is up-to-date, stronger its relations are

- Goal model, list of employees: updated vs. obsolete.
- Forum's post, recommendations: recent vs. old.

Open question: How to quantify one source regarding the others?
Let assume that we have a reliable weighted graph for now...

# Probabilistic Model: Markov Network

- Nodes $\rightarrow$ binary random variables $X_i \in \{\top, \bot\}$.
  - *Is the role/topic/concept/stakeholder wanted or rejected?*

# Probabilistic Model: Markov Network

- Nodes $\rightarrow$ binary random variables $X_i \in \{\top, \bot\}$.
  - *Is the role/topic/concept/stakeholder wanted or rejected?*
- Relations $\rightarrow$ potential functions $f_{i,j}(X_i, X_j) \in \mathbb{R}^+$.
  - *Ex:* $f_{i,j}(\top, \top) =$ *weight of the relation between* $X_i$ *and* $X_j$, $f_{i,j}(\top, \bot) = f_{i,j}(\bot, \top) = f_{i,j}(\bot, \bot) = 0$.

# Probabilistic Model: Markov Network

- Nodes $\rightarrow$ binary random variables $X_i \in \{\top, \bot\}$.
  - *Is the role/topic/concept/stakeholder wanted or rejected?*
- Relations $\rightarrow$ potential functions $f_{i,j}(X_i, X_j) \in \mathbb{R}^+$.
  - *Ex:* $f_{i,j}(\top, \top) =$ *weight of the relation between* $X_i$ *and* $X_j$, $f_{i,j}(\top, \bot) = f_{i,j}(\bot, \top) = f_{i,j}(\bot, \bot) = 0$.
- $P(X) = \frac{\prod_{i,j} f_{i,j}(x_i, x_j)}{Z}$ ($Z =$ normalization factor)

# Probabilistic Model: Markov Network

- Nodes $\rightarrow$ binary random variables $X_i \in \{\top, \bot\}$.
  - *Is the role/topic/concept/stakeholder wanted or rejected?*
- Relations $\rightarrow$ potential functions $f_{i,j}(X_i, X_j) \in \mathbb{R}^+$.
  - *Ex:* $f_{i,j}(\top, \top) =$ *weight of the relation between* $X_i$ *and* $X_j$, $f_{i,j}(\top, \bot) = f_{i,j}(\bot, \top) = f_{i,j}(\bot, \bot) = 0$.
- $P(X) = \frac{\prod_{i,j} f_{i,j}(x_i, x_j)}{Z}$ ($Z =$ normalization factor)
- Partial + conditional probabilities: $P(X' \in X)$, $P(X_1|X_2)$.
  - *Ex:* $P(Philip_S = \top | Manage\ privacy_T = \top, participant_R = \top)$

# Probabilistic Model: Markov Network

- Nodes $\rightarrow$ binary random variables $X_i \in \{\top, \bot\}$.
  - *Is the role/topic/concept/stakeholder wanted or rejected?*
- Relations $\rightarrow$ potential functions $f_{i,j}(X_i, X_j) \in \mathbb{R}^+$.
  - *Ex:* $f_{i,j}(\top, \top) =$ *weight of the relation between* $X_i$ *and* $X_j$, $f_{i,j}(\top, \bot) = f_{i,j}(\bot, \top) = f_{i,j}(\bot, \bot) = 0$.
- $P(X) = \frac{\prod_{i,j} f_{i,j}(x_i, x_j)}{Z}$ ($Z =$ normalization factor)
- Partial + conditional probabilities: $P(X' \in X)$, $P(X_1|X_2)$.
  - *Ex:* $P(Philip_S = \top | Manage\ privacy_T = \top, participant_R = \top)$

**Stakeholders ranking**

$P(S_1|Q) > P(S_2|Q) \Rightarrow S_1 > S_2$ for query $Q$.

# Markov Network Engine: libDai

- free and open source[1] C++ library.

[1] http://cs.ru.nl/~jorism/libDAI/
[2] http://www.cs.huji.ac.il/project/UAI10/

# Markov Network Engine: libDai

- free and open source[1] C++ library.
- Implements various inference methods for discrete graphical models.

[1] http://cs.ru.nl/~jorism/libDAI/
[2] http://www.cs.huji.ac.il/project/UAI10/

# Markov Network Engine: libDai

- free and open source[1] C++ library.
- Implements various inference methods for discrete graphical models.
- Manage approximation with timeout for heavy computations.

[1] http://cs.ru.nl/~jorism/libDAI/
[2] http://www.cs.huji.ac.il/project/UAI10/

# Markov Network Engine: libDai

- free and open source[1] C++ library.
- Implements various inference methods for discrete graphical models.
- Manage approximation with timeout for heavy computations.
- In the 3 winners of the UAI 2010 Approximate Inference Challenge[2]
  - *Program used for the experiments.*

[1] http://cs.ru.nl/~jorism/libDAI/

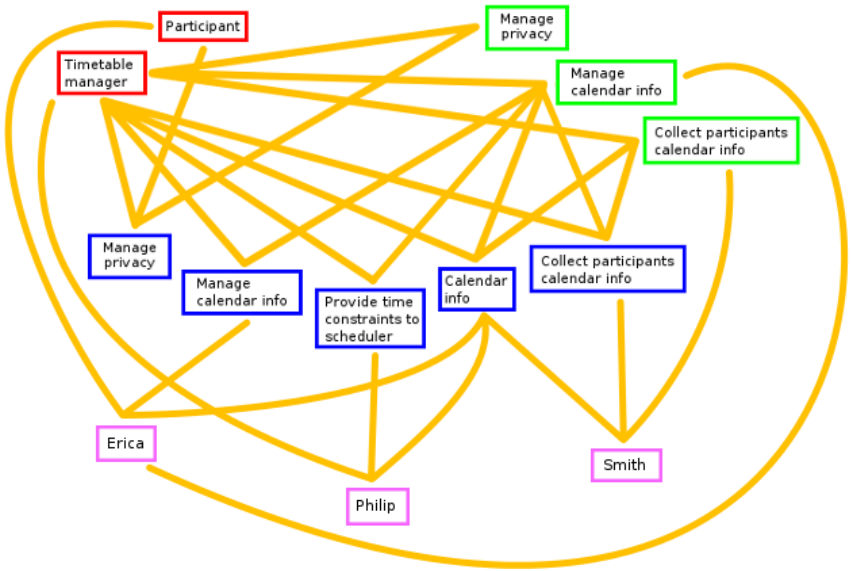
[2] http://www.cs.huji.ac.il/project/UAI10/

# Markov Network Engine: libDai

- free and open source[1] C++ library.
- Implements various inference methods for discrete graphical models.
- Manage approximation with timeout for heavy computations.
- In the 3 winners of the UAI 2010 Approximate Inference Challenge[2]
  - *Program used for the experiments.*
- **Main interest: compute loops!**

[1] http://cs.ru.nl/~jorism/libDAI/
[2] http://www.cs.huji.ac.il/project/UAI10/

# Queries




Queries tested in $\wp(R \cup T \cup C)$:

- $\|Q\| = 0$ $(Q = \emptyset)$.
- $\|Q\| = 1, 2, 3$.
- Exhaustive test (all combinations) $= 176$ queries.

# Parameters of the Network

- Different amount of info (including no info).

# Parameters of the Network

- Different amount of info (including no info).
- Different potential functions:
  - *identity (classic + normalised + semi-normalised)*
  - *identity + prior (classic + normalised + semi-normalised)*
  - *weight of evidence:* $w = \frac{p}{1-p}$ *(classic + symmetric).*

# Parameters of the Network

- Different amount of info (including no info).
- Different potential functions:
  - *identity (classic + normalised + semi-normalised)*
  - *identity + prior (classic + normalised + semi-normalised)*
  - *weight of evidence:* $w = \frac{p}{1-p}$ *(classic + symmetric).*
- Different weights for different relations (e.g. S-T > R-T).

# Parameters of the Network

- Different amount of info (including no info).
- Different potential functions:
  - *identity (classic + normalised + semi-normalised)*
  - *identity + prior (classic + normalised + semi-normalised)*
  - *weight of evidence:* $w = \frac{p}{1-p}$ *(classic + symmetric).*
- Different weights for different relations (e.g. S-T > R-T).
- High weight for trivial topic-concept relations.
  - *Ex: Manage privacy*$_T \rightarrow$ *Manage privacy*$_C$*.*

# Assumptions (Gold Standard)

- No data (network or query)
  $\Rightarrow$ all the elements have the same rank.

# Assumptions (Gold Standard)

- No data (network or query)
  $\Rightarrow$ all the elements have the same rank.
- $S_1$ is stronger related to $B \in R \cup T \cup C$ than $S_2$
  $\Rightarrow S_1 > S_2$ for $Q = \{B\}$.

# Assumptions (Gold Standard)

- No data (network or query)
  $\Rightarrow$ all the elements have the same rank.
- $S_1$ is stronger related to $B \in R \cup T \cup C$ than $S_2$
  $\Rightarrow S_1 > S_2$ for $Q = \{B\}$.
- $B_1$ related to $B_2$, $S_1$ related to $B_1$ and not $B_2$, $S_2$ related (equally) to $B_2$ but not $B_1$
  $\Rightarrow S_1 > S_2$ for $Q = \{B_1\}$, $S_1 < S_2$ for $Q = \{B_2\}$.

# Assumptions (Gold Standard)

- No data (network or query)
  $\Rightarrow$ all the elements have the same rank.
- $S_1$ is stronger related to $B \in R \cup T \cup C$ than $S_2$
  $\Rightarrow S_1 > S_2$ for $Q = \{B\}$.
- $B_1$ related to $B_2$, $S_1$ related to $B_1$ and not $B_2$, $S_2$ related (equally) to $B_2$ but not $B_1$
  $\Rightarrow S_1 > S_2$ for $Q = \{B_1\}$, $S_1 < S_2$ for $Q = \{B_2\}$.
- $Q_1 = \{B_1\}, ..., Q_n = \{B_n\}$ have common partial rankings
  $\Rightarrow Q = \{B_1, ..., B_n\}$ have the same partial rankings.

# Results... to Be Continued

Data generated, but not analyzed yet. However, a quick reading allows to notice some points:

# Results... to Be Continued

Data generated, but not analyzed yet. However, a quick reading allows to notice some points:

- ✓ Empty network = no ranking, regardless of the query and the parameters.

# Results... to Be Continued

Data generated, but not analyzed yet. However, a quick reading allows to notice some points:

- ✓ Empty network = no ranking, regardless of the query and the parameters.
- ✓ With data, the assumption on composed queries is generally satisfied.

# Results... to Be Continued

Data generated, but not analyzed yet. However, a quick reading allows to notice some points:

- ✓ Empty network = no ranking, regardless of the query and the parameters.
- ✓ With data, the assumption on composed queries is generally satisfied.
- × But the assumption on empty query is generally NOT satisfied. (close world representation?)

# Results... to Be Continued

Data generated, but not analyzed yet. However, a quick reading allows to notice some points:

- ✓ Empty network = no ranking, regardless of the query and the parameters.
- ✓ With data, the assumption on composed queries is generally satisfied.
- × But the assumption on empty query is generally NOT satisfied. (close world representation?)
- × With data, only 1 case satisfies all the applicable assumptions (*symmetric WoE* instead of classical *identity* + *prior*)

Summary:

- Extract roles, topics and concepts from a GM.
- Extract their relations from the GM.
- Relate stakeholders using other sources.
- Translate graph in a Markov network.
- Rank stakeholders based on probabilities.
- Experiments setting.

Summary:

- Extract roles, topics and concepts from a GM.
- Extract their relations from the GM.
- Relate stakeholders using other sources.
- Translate graph in a Markov network.
- Rank stakeholders based on probabilities.
- Experiments setting.

Future work:

- Analyse experiments' results.
- Fix and improve.
- Look at other techniques able to exploit the graph.
- Develop process to integrate specific techniques for concept extraction.

Thanks for your attention.

Questions?

C. Castro-Herrera and J. Cleland-Huang. A machine learning approach for identifying expert stakeholders. In 2009 Second International Workshop on Managing Requirements Knowledge (MARK), pages 45 –49, Sept. 2009. doi: 10.1109/MARK.2009.1.

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

S. L. Lim and A. Finkelstein. StakeRare: using social networks and collaborative filtering for large-scale requirements elicitation. IEEE Transactions on Software Engineering, 38(3):707–735, May 2012. ISSN 0098-5589. doi: 10.1109/TSE.2011.36. URL

http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5740931.

S. L. Lim, D. Quercia, and A. Finkelstein. StakeNet: using social networks to analyse the stakeholders of large-scale software projects. In Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 1, ICSE '10, pages 295–304, New York, NY, 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.

S. L. Lim, D. Damian, and A. Finkelstein. StakeSource2.0: using social networks of stakeholders to identify and prioritise requirements. In Proceedings of the 33rd International Conference on Software Engineering, ICSE '11, pages 1022–1024, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0445-0. doi:

http://doi.acm.org/10.1145/1985793.1985983. URL http://doi.acm.org/10.1145/1985793.1985983.

J. Mohebzada, G. Ruhe, and A. Eberlein. Systematic mapping of recommendation systems for requirements engineering. In 2012 International Conference on Software and System Process (ICSSP), pages 200 –209, June 2012. doi: 10.1109/ICSSP.2012.6225965.