Constrained Information-Theoretic Tripartite Graph Clustering to Identify Semantically Similar Relations

IJCAI'15, Buenos Aires, Argentina

Chenguang Wang (Peking Univ.), Yangqiu Song (UIUC), Dan Roth (UIUC),

Chi Wang (MSR), Jiawei Han (UIUC), Heng Ji (RPI), and Ming Zhang (Peking Univ.)

北京大學
PEKING UNIVERSITY

Outline

Problem: Relation Clustering

Approach: Constrained Tripartite Graph Clustering Model

Experiments

Open Information Extraction Relations

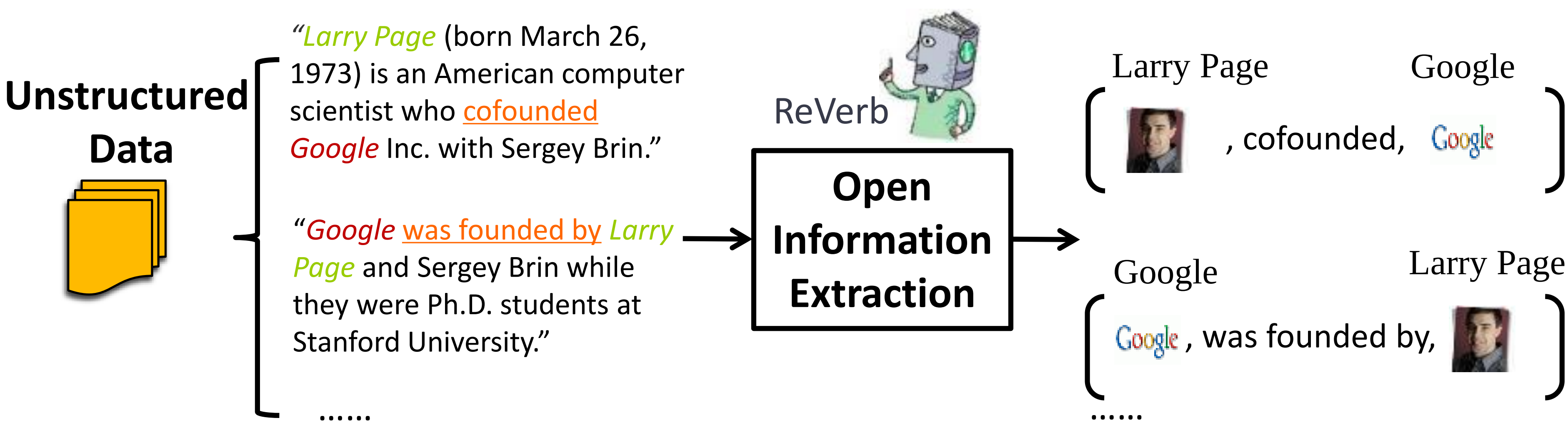
Open information extraction (IE) relations

Relations are not canonical:
Similar relations are expressed in different natural language ways.

Open Information Extraction Relations

Open information extraction (IE) relations

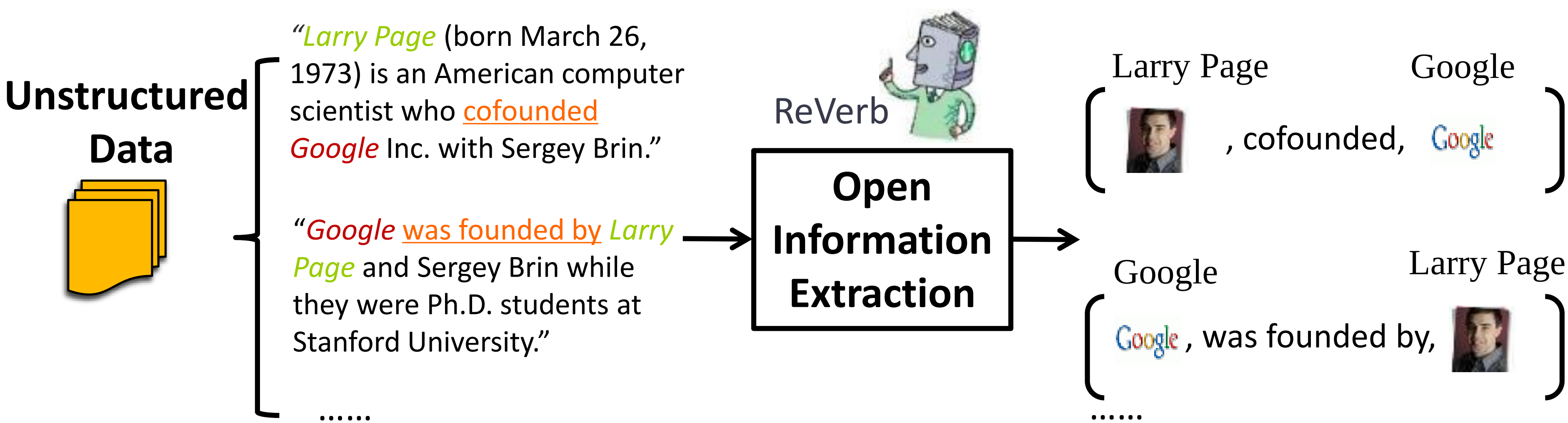
Relations are not canonical:
Similar relations are expressed in different natural language ways.



Open Information Extraction Relations

Open information extraction (IE) relations

Relations are not canonical:
Similar relations are expressed in different natural language ways.



Knowledge Base Relations

Knowledge base relations

Relations are not canonical:
Multi-hop relation and one-hop relation has the same meaning.

Knowledge Base Relations

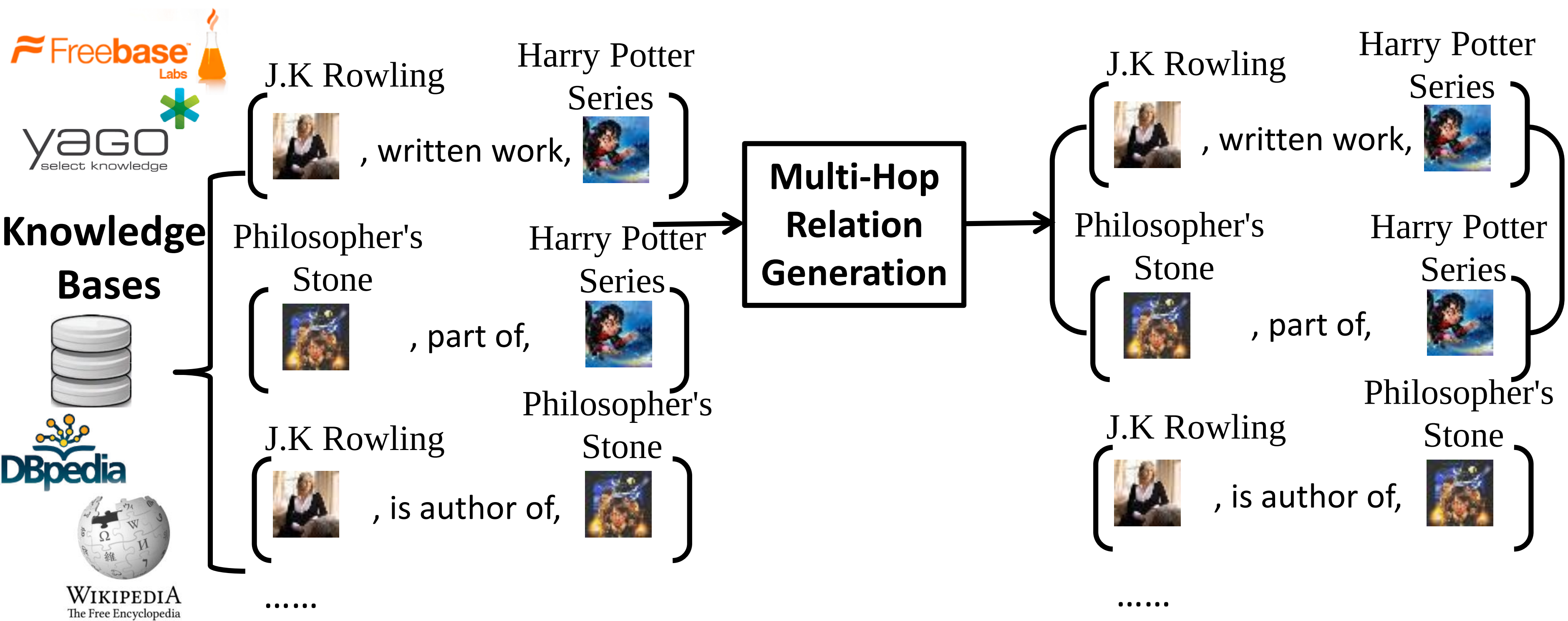
Knowledge base relations

Relations are not canonical:
Multi-hop relation and one-hop relation has the same meaning.

Knowledge Base Relations

Knowledge base relations

Relations are not canonical:
Multi-hop relation and one-hop relation has the same meaning.



Solution: Clustering Relations

Examples

(X, wrote, Y) and (X, ’s written work, Y)

(X, is founder of, Y) and (X, is CEO of, Y)

(X, written by, Y) and (X, part of, Z)^(Y, wrote, Z)

Solution: Clustering Relations

Examples

(X, wrote, Y) and (X, 's written work, Y)

(X, is founder of, Y) and (X, is CEO of, Y)

(X, written by, Y) and (X, part of, Z)^(Y, wrote, Z)

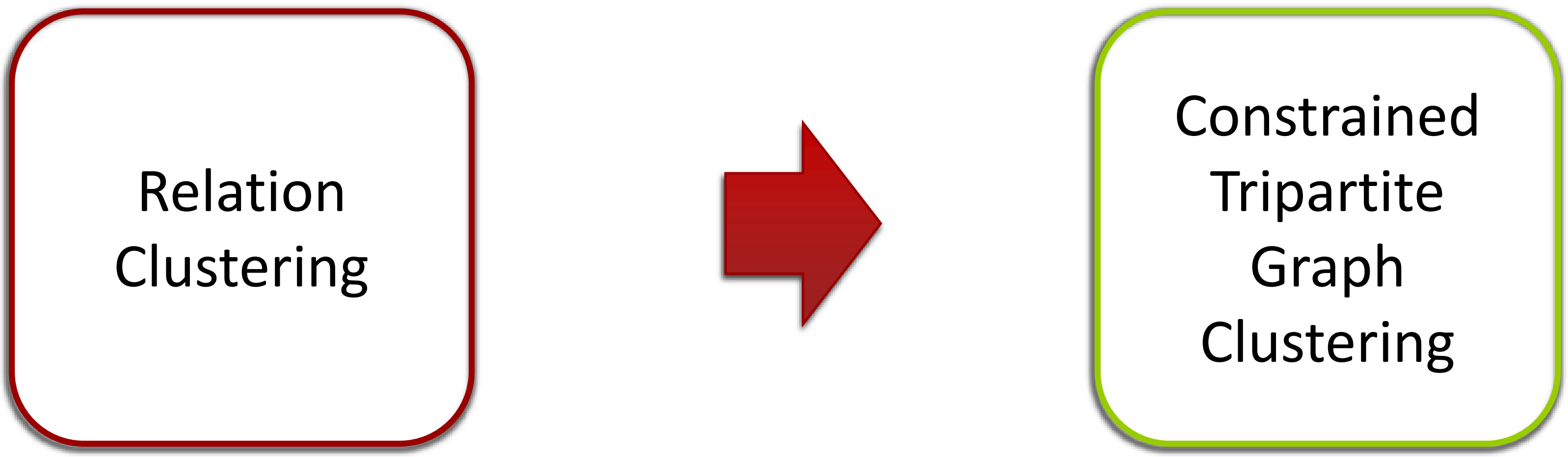
Applications

Knowledge base completion [Socher et al., 2013; West et al., 2014]

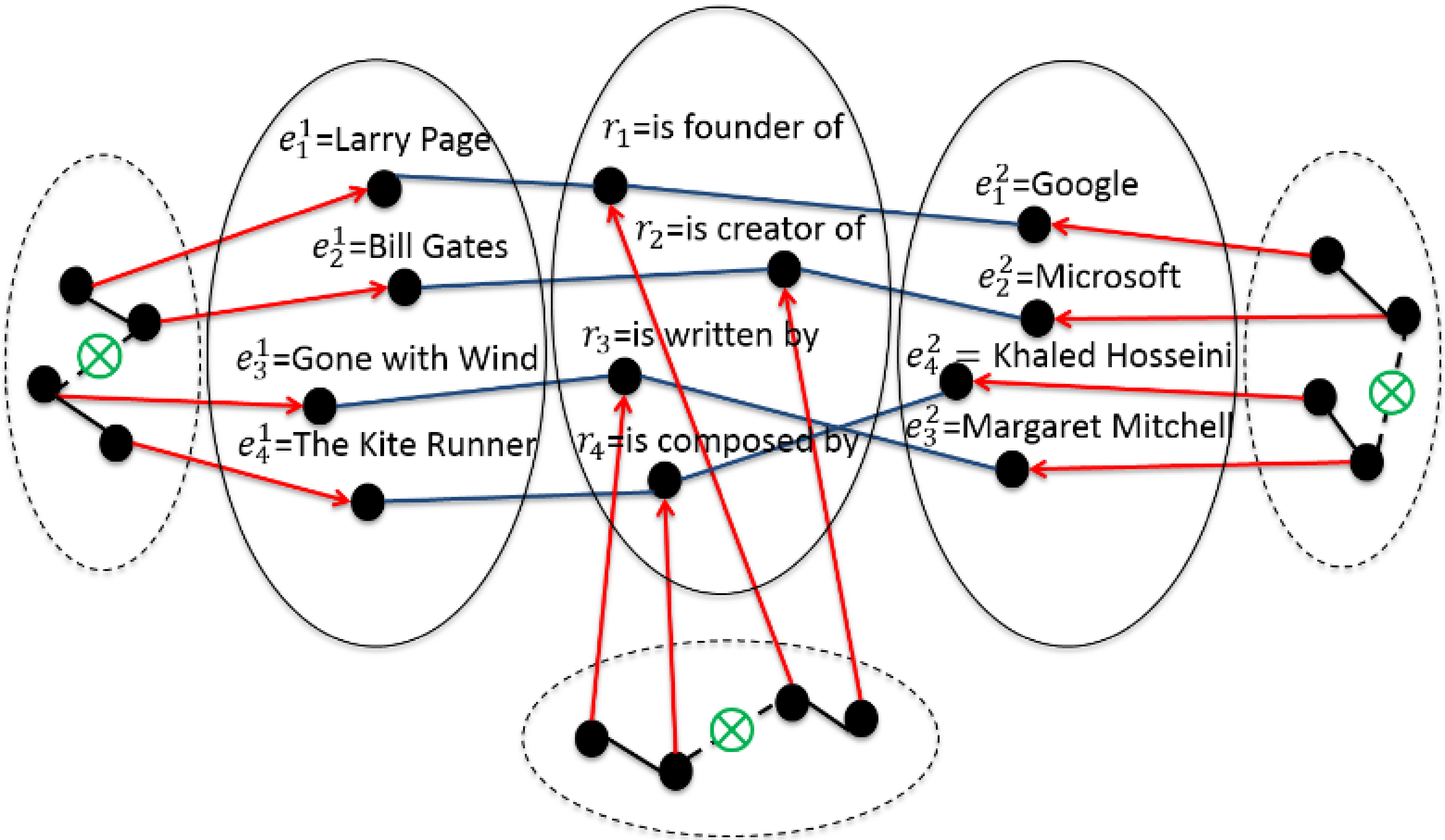
Information extraction [Chan and Roth, 2010; 2011; Li and Ji, 2014]

Knowledge inference [Richardson and Domingos, 2006]

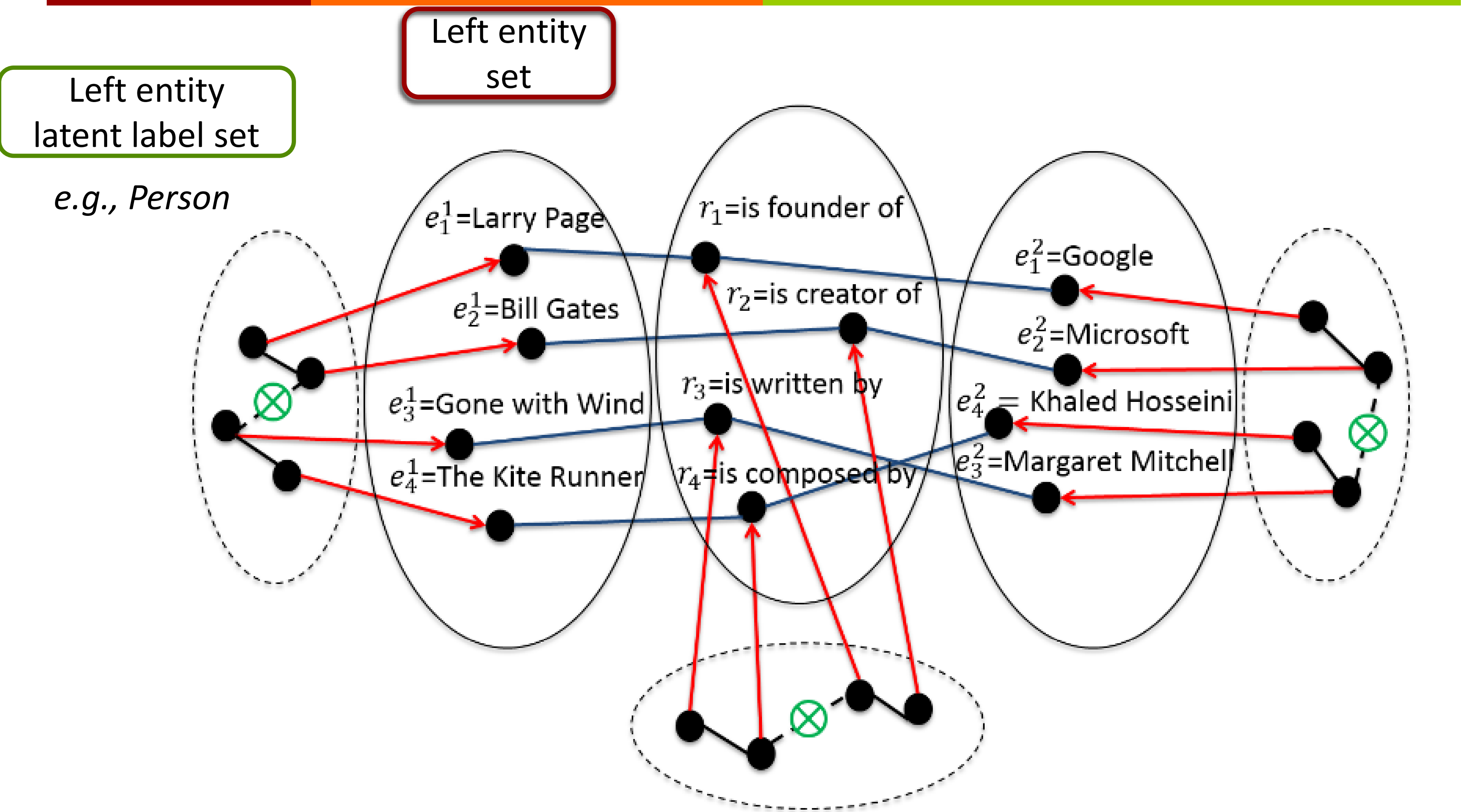
Relation Clustering


Relation Clustering
Constrained Tripartite Graph Clustering

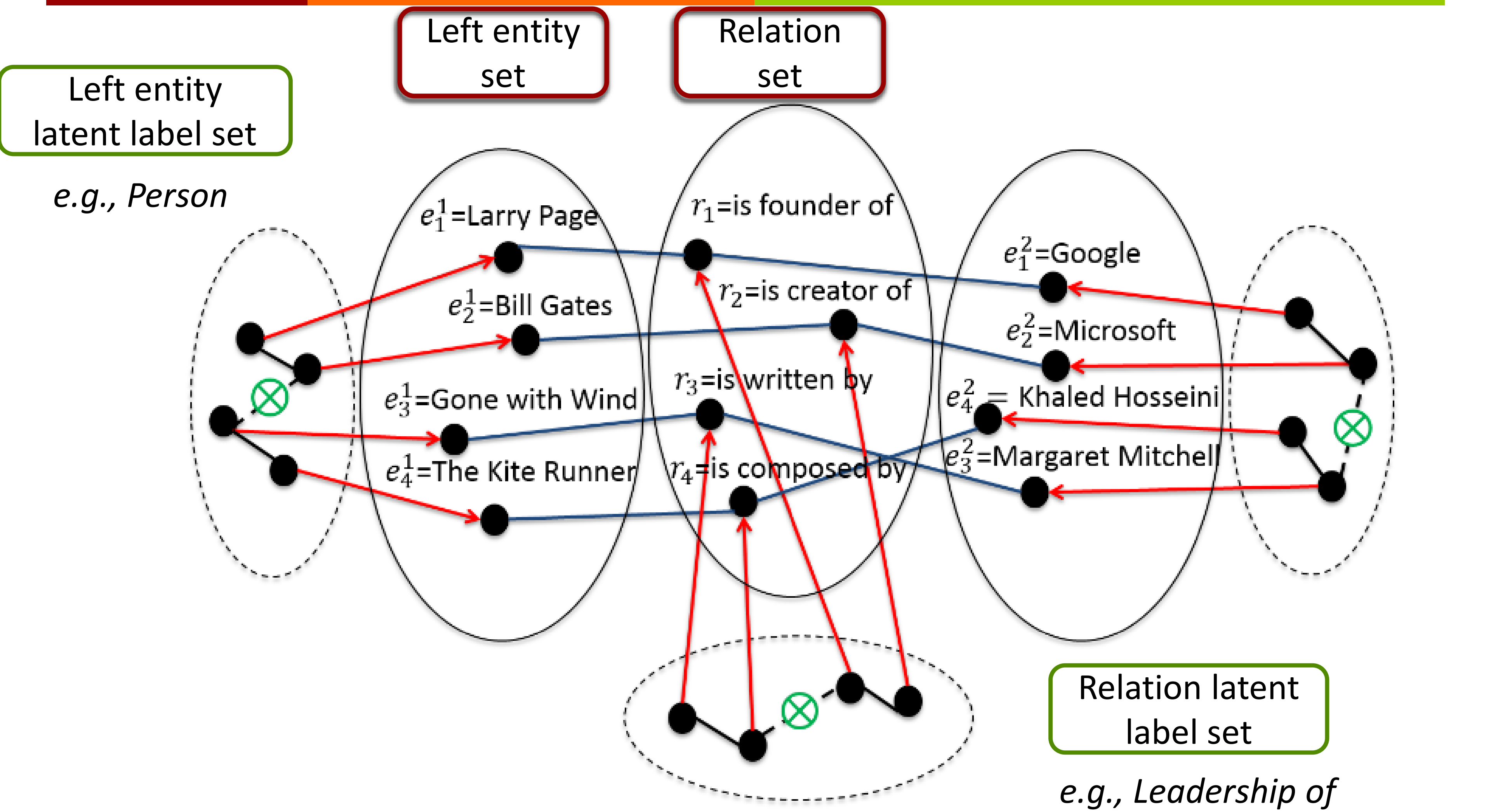
Problem Formulation: Constrained Tripartite Graph Clustering



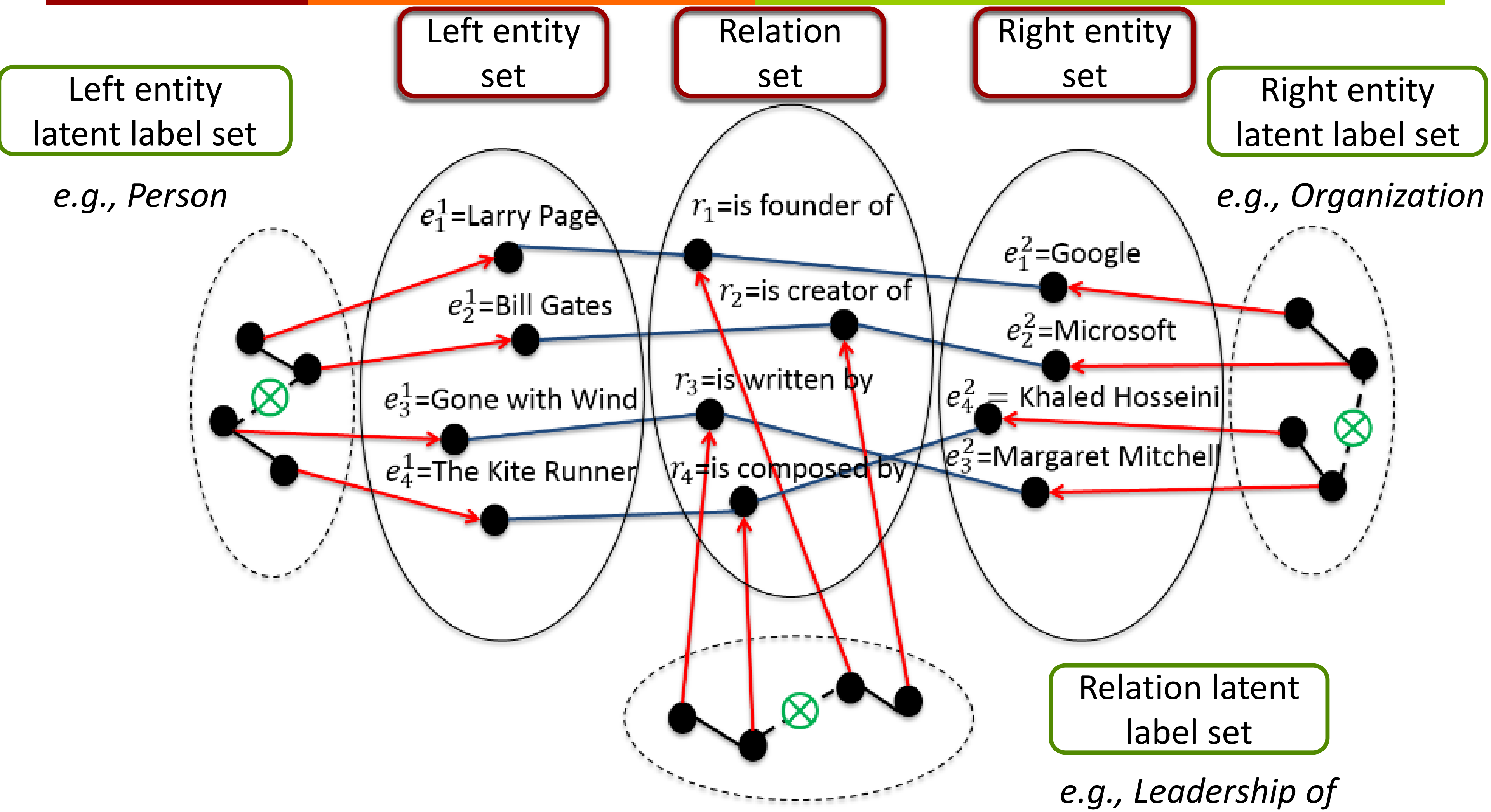
Problem Formulation: Constrained Tripartite Graph Clustering



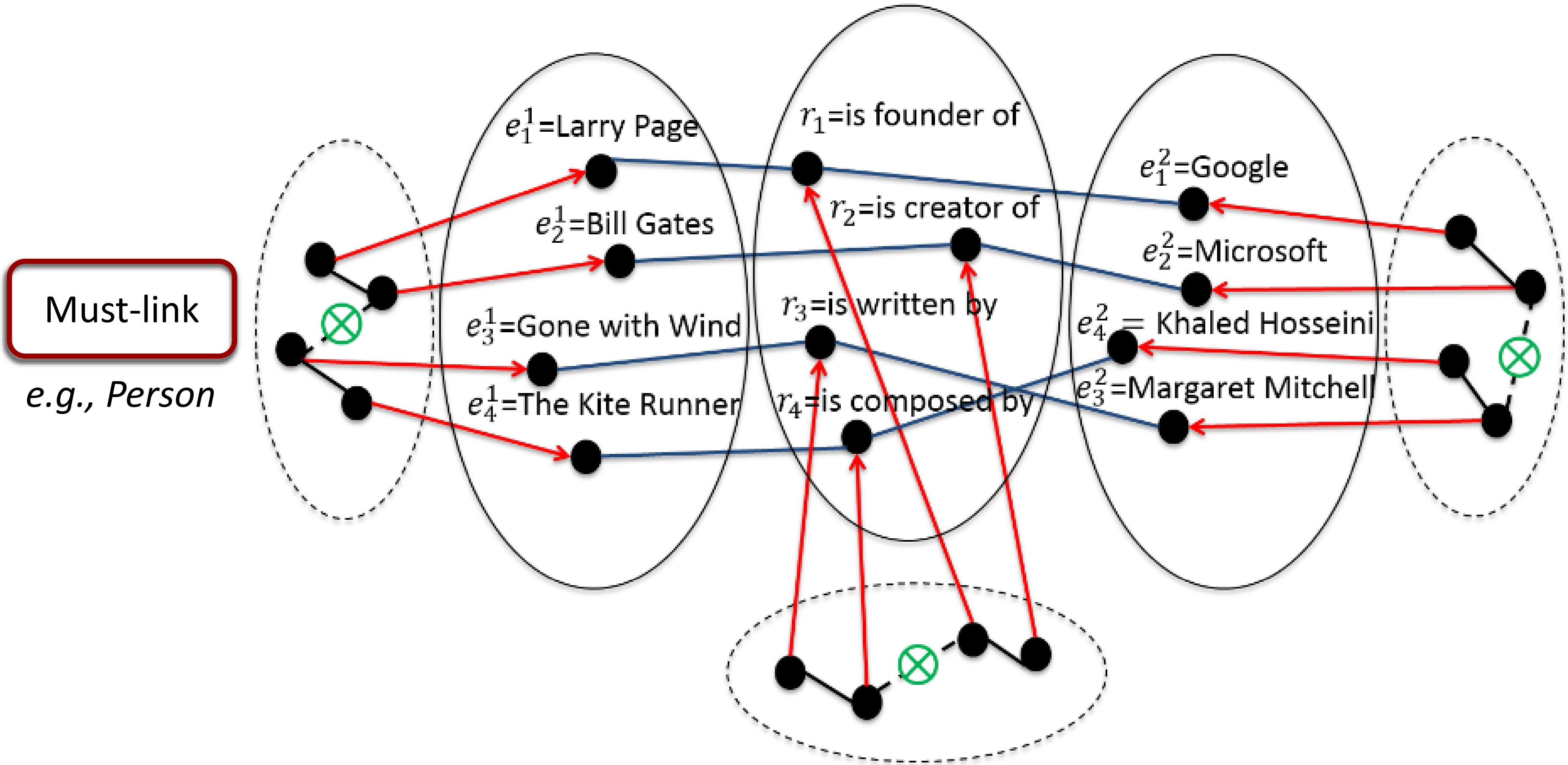
Problem Formulation: Constrained Tripartite Graph Clustering



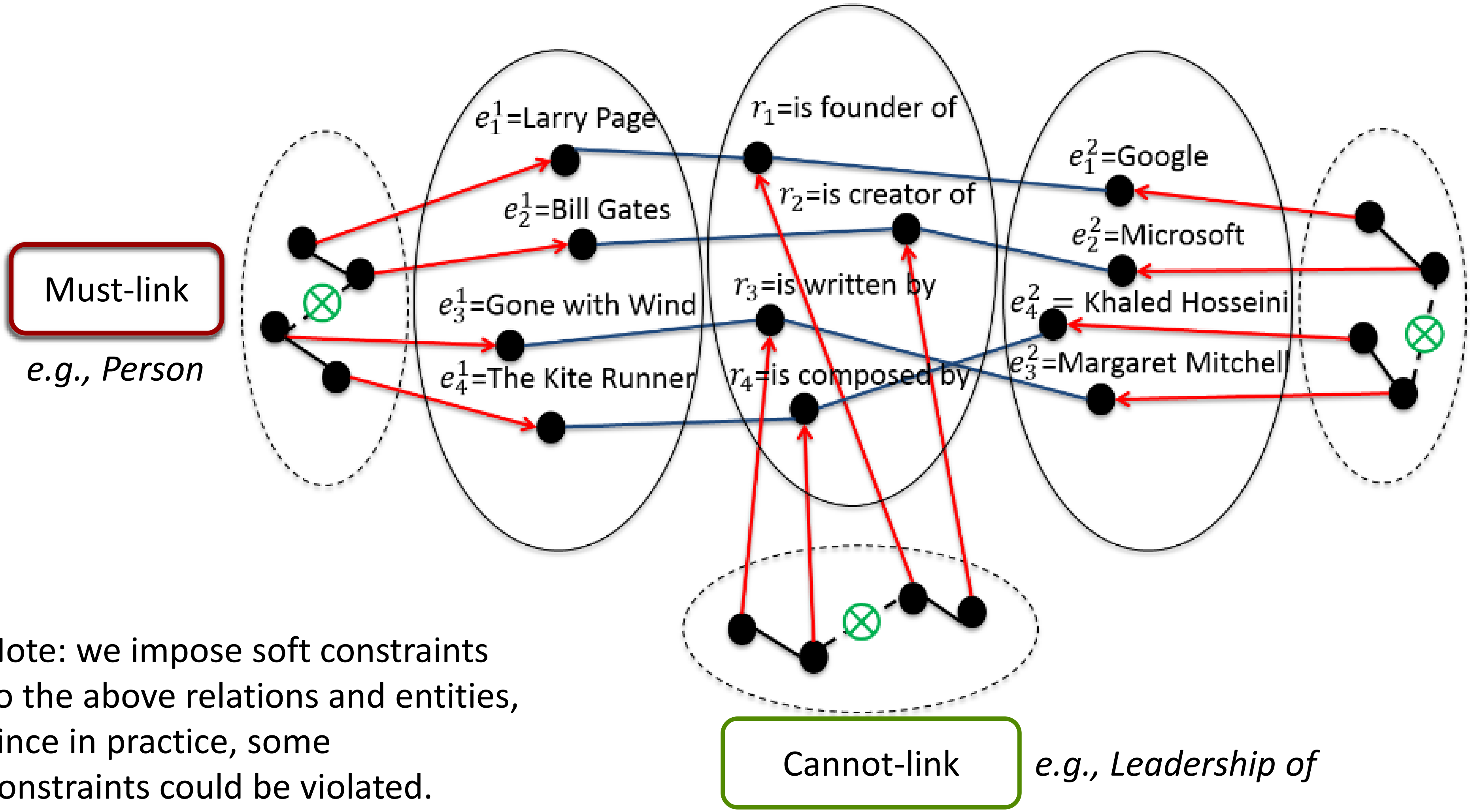
Problem Formulation: Constrained Tripartite Graph Clustering



Must-Link and Cannot-Link Constraints



Must-Link and Cannot-Link Constraints



Note: we impose soft constraints to the above relations and entities, since in practice, some constraints could be violated.

Model Description

Intuition

Relation triplet joint probability decomposition:

$$\mathrm{p}(e_i^1, r_m, e_j^2) \propto \mathrm{p}(r_m, e_i^1)\ \mathrm{p}(r_m, e_j^2) \qquad \textbf{Eq 1.}$$

Model Description

Intuition

Relation triplet joint probability decomposition:

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \quad \textbf{Eq 1.}$$

Calculated based on the co-occurrence count of r_m and e_i^I

Model Description

Intuition

Calculated based on the co-occurrence count of r_m and e_i^I

Relation triplet joint probability decomposition:

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \quad \textbf{Eq 1.}$$

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e^I_{k_{e^I}}})\, p(r_m | \widehat{r_{k_r}})\, p(e_i^I | \widehat{e^I_{k_{e^I}}}) \quad \textbf{Eq 2.}$$

Model Description

Intuition

Relation triplet joint probability decomposition:

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \qquad \textbf{Eq 1.}$$

Calculated based on the co-occurrence count of r_m and e_i^I

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e^I_{k_{e^I}}})\, p(r_m \mid \widehat{r_{k_r}})\, p(e_i^I \mid \widehat{e^I_{k_{e^I}}}) \qquad \textbf{Eq 2.}$$

$p(r_m, e_i^I)$ approximation Cluster indicators Cluster indices

Model Description

Intuition

Relation triplet joint probability decomposition:

Calculated based on the co-occurrence count of r_m and e_i^I

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2)$$ **Eq 1.**

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e_{k_{e^I}}^I})\, p(r_m | \widehat{r_{k_r}})\, p(e_i^I | \widehat{e_{k_{e^I}}^I})$$ **Eq 2.**

$p(r_m, e_i^I)$ approximation Cluster indicators Cluster indices

Objective Function

$$\{\mathcal{L}_{e^1}, \mathcal{L}_r, \mathcal{L}_{e^2}\} = argmin\, D_{KL}(p(R,\varepsilon^1)||q(R,\varepsilon^1)) + D_{KL}(p(R,\varepsilon^2)||q(R,\varepsilon^2))$$ **Eq 3.**

$$+\sum_{r_{m_1}=1}^{M} \sum_{r_{m_2}\in\mathcal{M}_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in \mathcal{M}_{r_{m_1}}) + \sum_{r_{m_1}=1}^{M} \sum_{r_{m_2}\in C_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in C_{r_{m_1}})$$

$$+\sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1\in\mathcal{M}_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}) + \sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1\in C_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in C_{e_{i_1}^1})$$

$$+\sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2\in\mathcal{M}_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}) + \sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2\in C_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in C_{e_{j_1}^2})$$

Model Description

Intuition

Relation triplet joint probability decomposition:

Calculated based on the co-occurrence count of r_m and e_i^I

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \qquad \textbf{Eq 1.}$$

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e^I_{k_{e^I}}})\, p(r_m | \widehat{r_{k_r}})\, p(e_i^I | \widehat{e^I_{k_{e^I}}}) \qquad \textbf{Eq 2.}$$

$p(r_m, e_i^I)$ approximation Cluster indicators Cluster indices

Objective Function

Multinomial distributions composed by $p(r_m, e_i^1)$

Multinomial distributions composed by $q(r_m, e_i^2)$

$$\{\mathcal{L}_{e^1}, \mathcal{L}_r, \mathcal{L}_{e^2}\} = argmin\, D_{KL}(p(R, \varepsilon^1) || q(R, \varepsilon^1)) + D_{KL}(p(R, \varepsilon^2) || q(R, \varepsilon^2)) \qquad \textbf{Eq 3.}$$

$$+\sum_{r_{m_1}=1}^{M} \sum_{r_{m_2} \in \mathcal{M}_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in \mathcal{M}_{r_{m_1}}) + \sum_{r_{m_1}=1}^{M} \sum_{r_{m_2} \in C_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in C_{r_{m_1}})$$

$$+\sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}) + \sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1 \in C_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in C_{e_{i_1}^1})$$

$$+\sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}) + \sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2 \in C_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in C_{e_{j_1}^2})$$

Model Description

Intuition

Relation triplet joint probability decomposition:

Calculated based on the co-occurrence count of r_m and e_i^I

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \quad \textbf{Eq 1.}$$

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e_{k_{e^I}}^I})\, p(r_m|\widehat{r_{k_r}})\, p(e_i^I|\widehat{e_{k_{e^I}}^I}) \quad \textbf{Eq 2.}$$

$p(r_m, e_i^I)$ approximation   Cluster indicators   Cluster indices

Objective Function

Multinomial distributions composed by $p(r_m, e_i^1)$

Multinomial distributions composed by $q(r_m, e_i^2)$

$$\{\mathcal{L}_{e^1}, \mathcal{L}_r, \mathcal{L}_{e^2}\} = argmin\, D_{KL}(p(R, \varepsilon^1)||q(R, \varepsilon^1)) + D_{KL}(p(R, \varepsilon^2)||q(R, \varepsilon^2)) \quad \textbf{Eq 3.}$$
$$+\sum_{r_{m_1}=1}^{M} \sum_{r_{m_2}\in \mathcal{M}_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in \mathcal{M}_{r_{m_1}}) + \sum_{r_{m_1}=1}^{M} \sum_{r_{m_2}\in C_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in C_{r_{m_1}})$$
$$+\sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1\in \mathcal{M}_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}) + \sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1\in C_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in C_{e_{i_1}^1})$$
$$+\sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2\in \mathcal{M}_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}) + \sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2\in C_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in C_{e_{j_1}^2})$$

Must-link set

Model Description

Intuition

Relation triplet joint probability decomposition:

Calculated based on the co-occurrence count of r_m and e_i^I

$$p(e_i^1, r_m, e_j^2) \propto p(r_m, e_i^1)\, p(r_m, e_j^2) \qquad \textbf{Eq 1.}$$

Motivated by Information-Theoretic Co-Clustering (ITCC) [I. S. Dhillon KDD'03] :

$$q(r_m, e_i^I) = p(\widehat{r_{k_r}}, \widehat{e^I_{k_{e^I}}})\, p(r_m \mid \widehat{r_{k_r}})\, p(e_i^I \mid \widehat{e^I_{k_{e^I}}}) \qquad \textbf{Eq 2.}$$

$p(r_m, e_i^I)$ approximation   Cluster indicators   Cluster indices

Objective Function

Multinomial distributions composed by $p(r_m, e_i^1)$

Multinomial distributions composed by $q(r_m, e_i^2)$

$$\{\mathcal{L}_{e^1}, \mathcal{L}_r, \mathcal{L}_{e^2}\} = argmin\, D_{KL}(p(R, \varepsilon^1) || q(R, \varepsilon^1)) + D_{KL}(p(R, \varepsilon^2) || q(R, \varepsilon^2)) \qquad \textbf{Eq 3.}$$
$$+ \sum_{r_{m_1}=1}^{M} \sum_{r_{m_2} \in \mathcal{M}_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in \mathcal{M}_{r_{m_1}}) + \sum_{r_{m_1}=1}^{M} \sum_{r_{m_2} \in \mathcal{C}_{r_{m_1}}} V(r_{m_1}, r_{m_2} \in \mathcal{C}_{r_{m_1}})$$
$$+ \sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in \mathcal{M}_{e_{i_1}^1}) + \sum_{e_{i_1}^1=1}^{V_1} \sum_{e_{i_2}^1 \in \mathcal{C}_{e_{i_1}^1}} V(e_{i_1}^1, e_{i_2}^1 \in \mathcal{C}_{e_{i_1}^1})$$
$$+ \sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in \mathcal{M}_{e_{j_1}^2}) + \sum_{e_{j_1}^2=1}^{V_2} \sum_{e_{j_2}^2 \in \mathcal{C}_{e_{j_1}^2}} V(e_{j_1}^2, e_{j_2}^2 \in \mathcal{C}_{e_{j_1}^2})$$

Must-link set   Cannot-link set

Experiments

Datasets		
Name		Description
Rel-KB	Freebase Labs	KB relations from Freebase, which particularly includes multi-hop relations
Rel-OIE	ReVerb	Open IE Relations extracted from Wikipedia using ReVerb

Experiments

Datasets		
Name		Description
Rel-KB	Freebase Labs	KB relations from Freebase, which particularly includes multi-hop relations
Rel-OIE	ReVerb	Open IE Relations extracted from Wikipedia using ReVerb

Experiments

Datasets	
Name	Description
Rel-KB Freebase Labs	KB relations from Freebase, which particularly includes multi-hop relations
Rel-OIE ReVerb	Open IE Relations extracted from Wikipedia using ReVerb

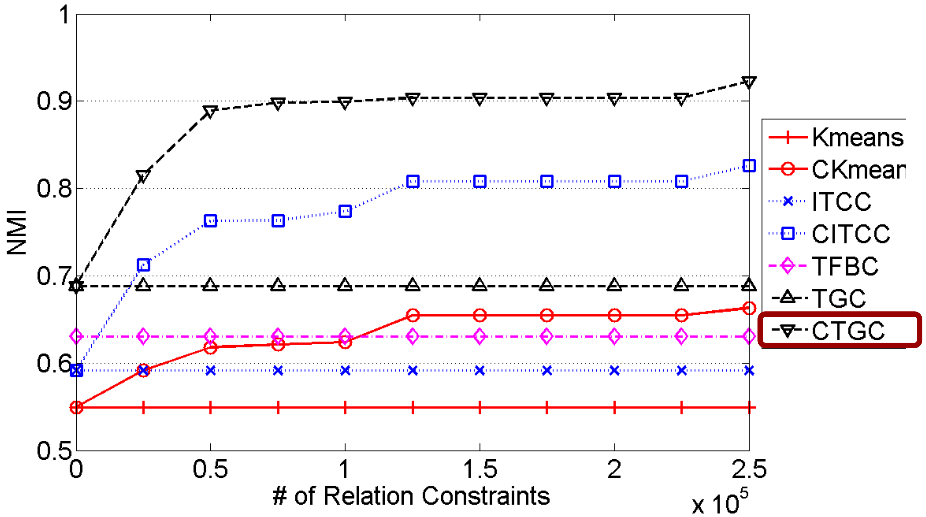
Relation Constraints for Rel-KB dataset (* Entity Constraints are similarly defined)	
Constraint Type	Description
Must-link	If two relations are generated from the same relation category, we add a must-link
Cannot-link	Otherwise

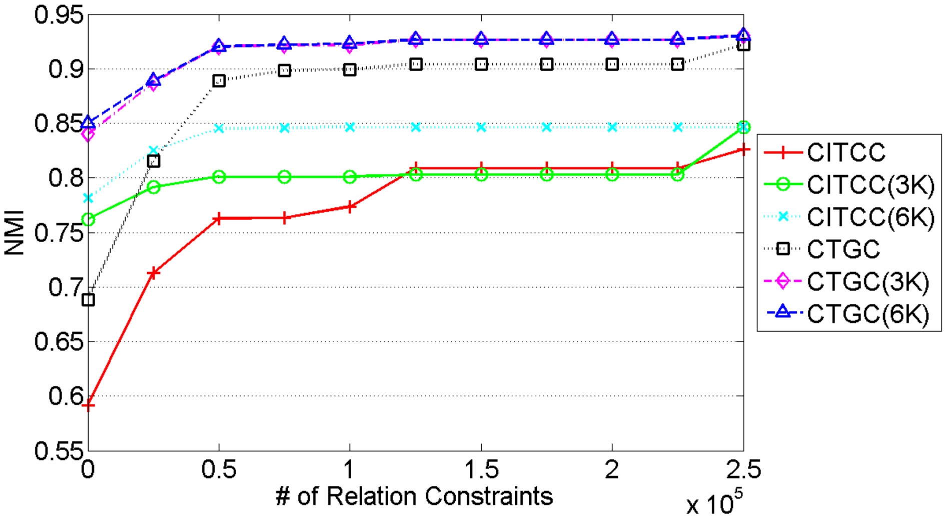
Relation Constraints for Rel-OIE dataset (* Entity Constraints are similarly defined)	
Constraint Type	Description
Must-link	If the similarity between two relation phrases is beyond a predefined threshold (experimentally, 0.5), we add a must-link to these relations
Cannot-link	Otherwise

Comparable Methods

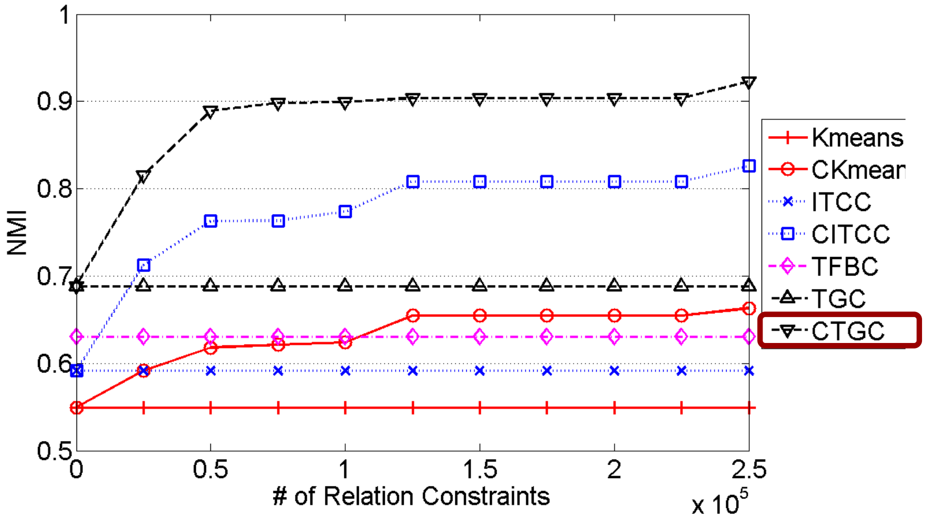
Methods	Description
Kmeans	One-dimensional clustering algorithm
CKmeans	Constrained Kmeans [S. Basu KDD’04]
ITCC	Information-theoretic co-clustering [I. S. Dhillon KDD’03]
CITCC	Constrained information-theoretic co-clustering [Y. Song TKDE’13]
TFBC	Tensor factorization based clustering [I. Sutskever NIPS’09]
TGC	Our method without constraints
CTGC	Our method

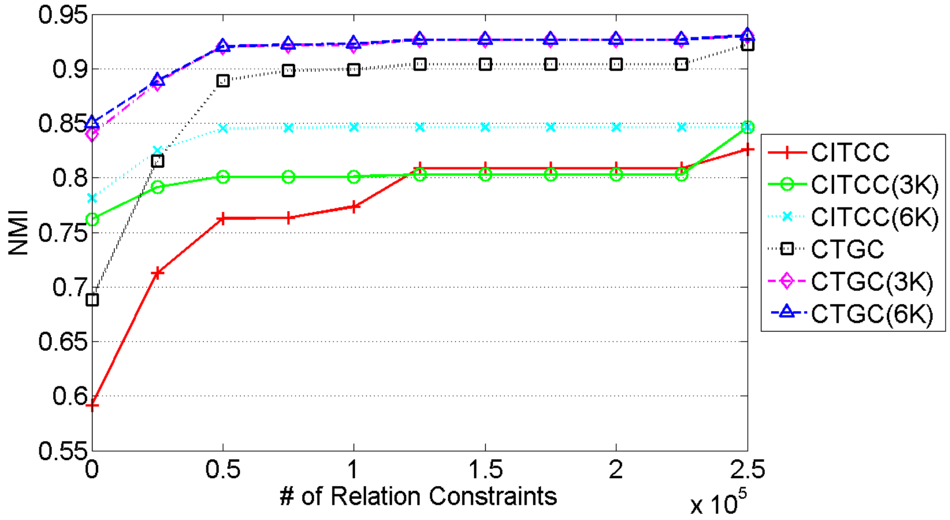
Analysis of Clustering Results





Analysis of Clustering Results



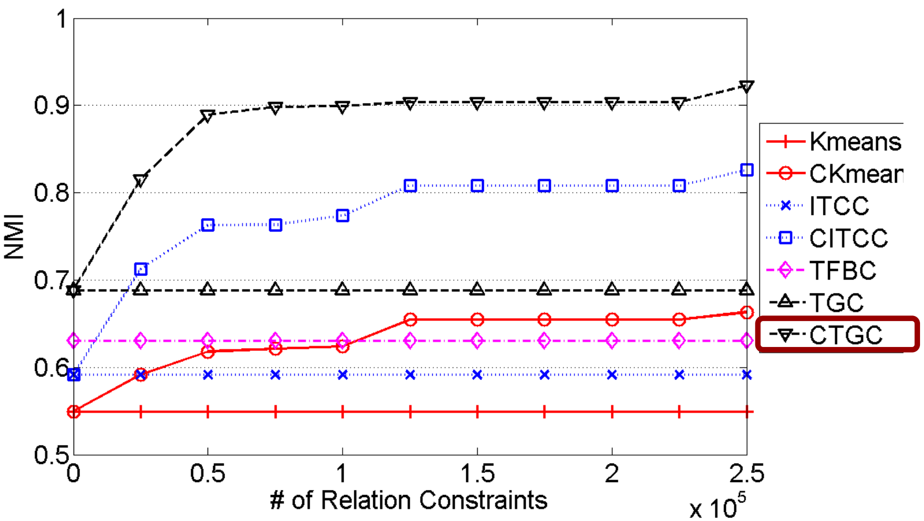


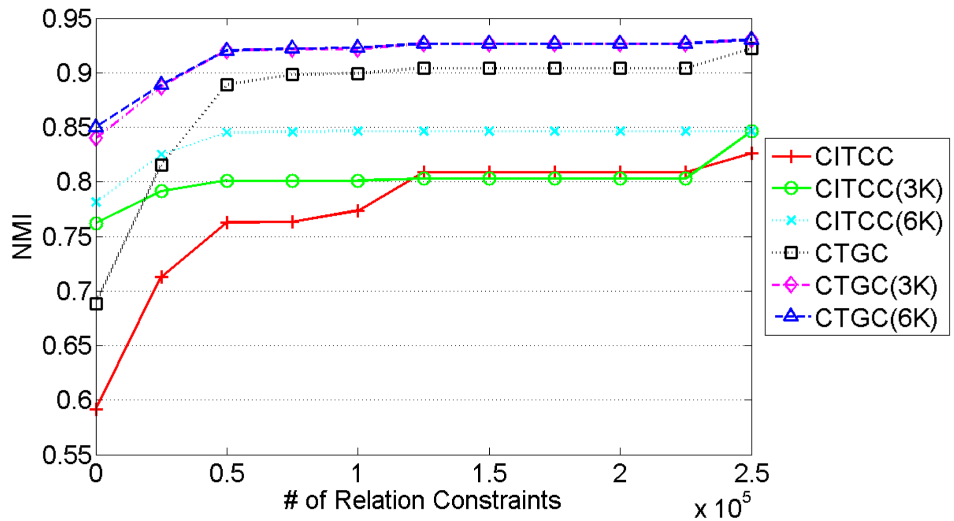
Finding #1:

Relation constraints are very effective:

CTGC and TGC perform better, with more relation constraints in CTGC, the improvement is more significant.

Analysis of Clustering Results





Finding #1:

Relation constraints are very effective:

CTGC and TGC perform better, with more relation constraints in CTGC, the improvement is more significant.

Finding #2:

*Entity constraints are also effective:

Even if we have little knowledge about relations, we can still expect better results if we know knowledge about entities.

Case Study of Clustering Results

Examples generated by CTGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, act in, Y); (X, , appears in, Y); (X, won best actor for, Y)

Examples generated by TGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, who played, Y); (X, starred in, Y); (X, ’s capital in, Y)

Case Study of Clustering Results

Examples generated by CTGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, act in, Y); (X, , appears in, Y); (X, won best actor for, Y)

Examples generated by TGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, who played, Y); (X, starred in, Y); (X, 's capital in, Y)

Finding #1:

Both CTGC and TGC generate reasonable results:
The tripartite graph structure enhances the clustering by using entity and relation together.

Case Study of Clustering Results

Examples generated by CTGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, act in, Y); (X, , appears in, Y); (X, won best actor for, Y)

Examples generated by TGC	
Category	Examples
Organization-Founder	(X, founded by, Y); (X, led by, Y); (Y, is the owner of, X); (X, , sold by, Y)
Actor-Film	(X, who played, Y); (X, starred in, Y); (X, 's capital in, Y)

Finding #1:

Both CTGC and TGC generate reasonable results:
The tripartite graph structure enhances the clustering by using entity and relation together.

Finding #2:

CTGC is better than TGC:
The must-link and cannot-link constraints help filter out illegitimate relations.

Recall

Problem

Relation clustering

CTGC

Constrained information-theoretic tripartite graph clustering model

Results

In both knowledge base and open information extraction, CTGC is effective

Recall

Problem

Relation clustering

CTGC

Constrained information-theoretic tripartite graph clustering model

Results

In both knowledge base and open information extraction, CTGC is effective

Thank You! ☺

If you have any problem,
please contact via wangchenguang@pku.edu.cn