# CoupledLP: Link Prediction in Coupled Networks

Yuxiao Dong[#], Jing Zhang[+], Jie Tang[+], Nitesh V. Chawla[#], Bai Wang[*]

[#]University of Notre Dame

[+]Tsinghua University

[*]Beijing University of Posts and Telecommunications

# Mobile Networks

at&t

2015.08.08 10:30

2015.08.08 10:48

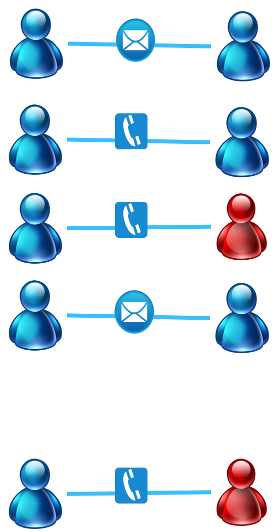
2015.08.08 11:01

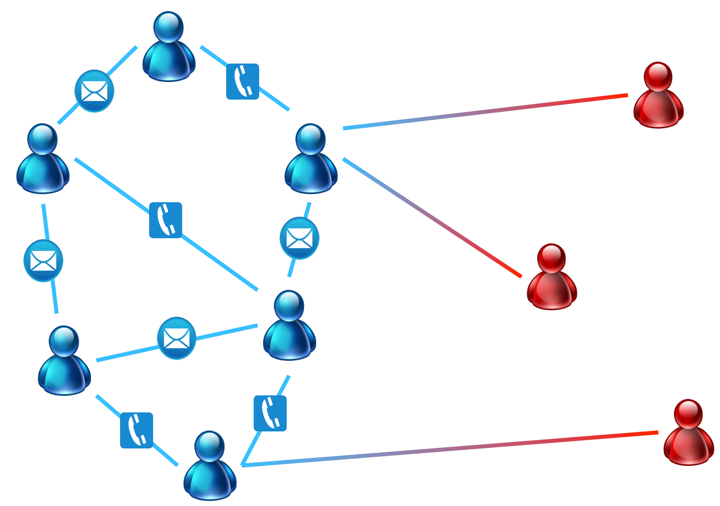
2015.08.08 11:29

……

2016.01.01 00:00

……





# Mobile Networks

at&t

2015.08.08 10:30

2015.08.08 10:48

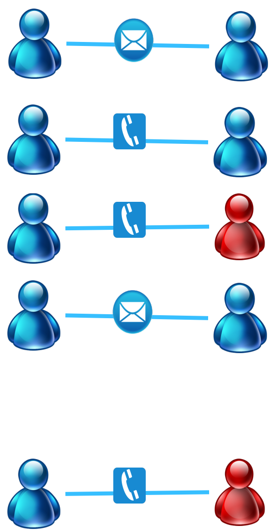
2015.08.08 11:01

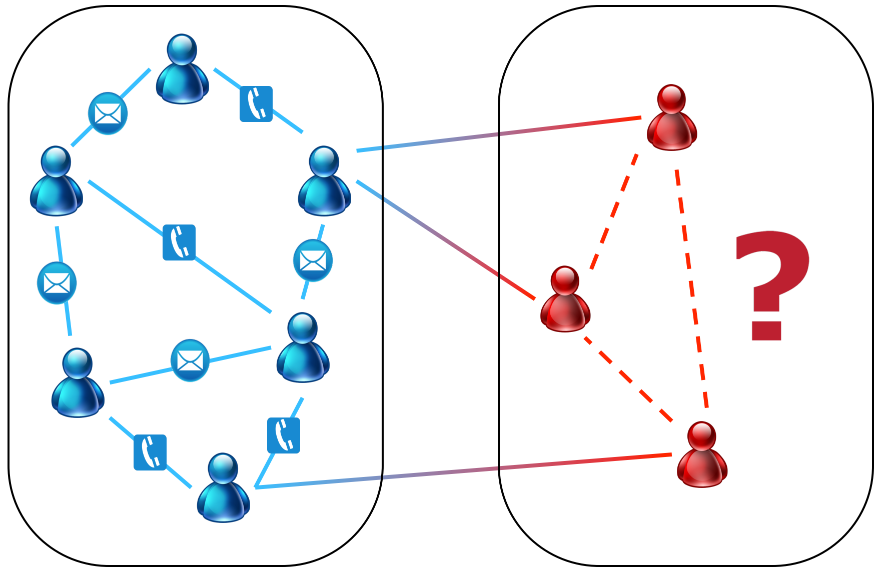
2015.08.08 11:29

……

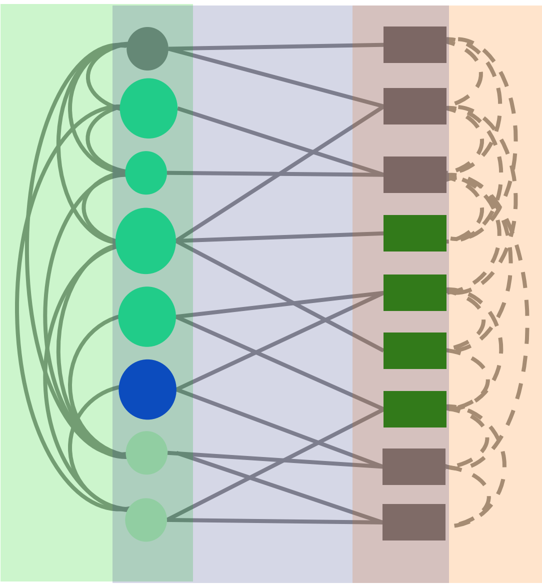
2016.01.01 00:00

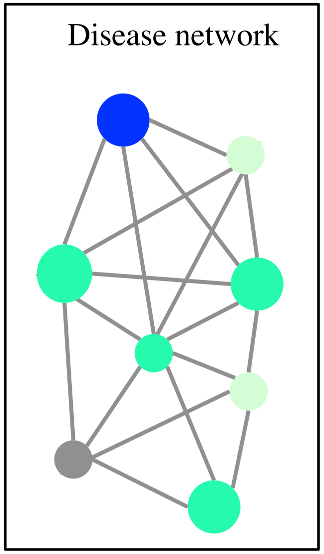
……

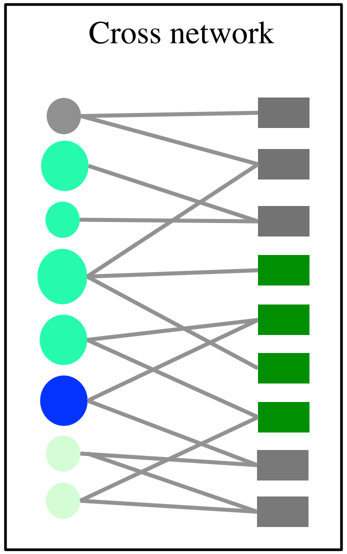


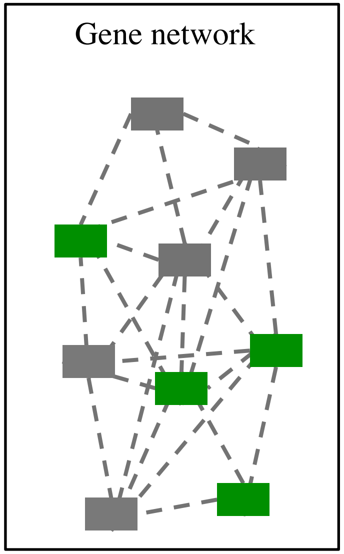


# Disease-Gene Networks





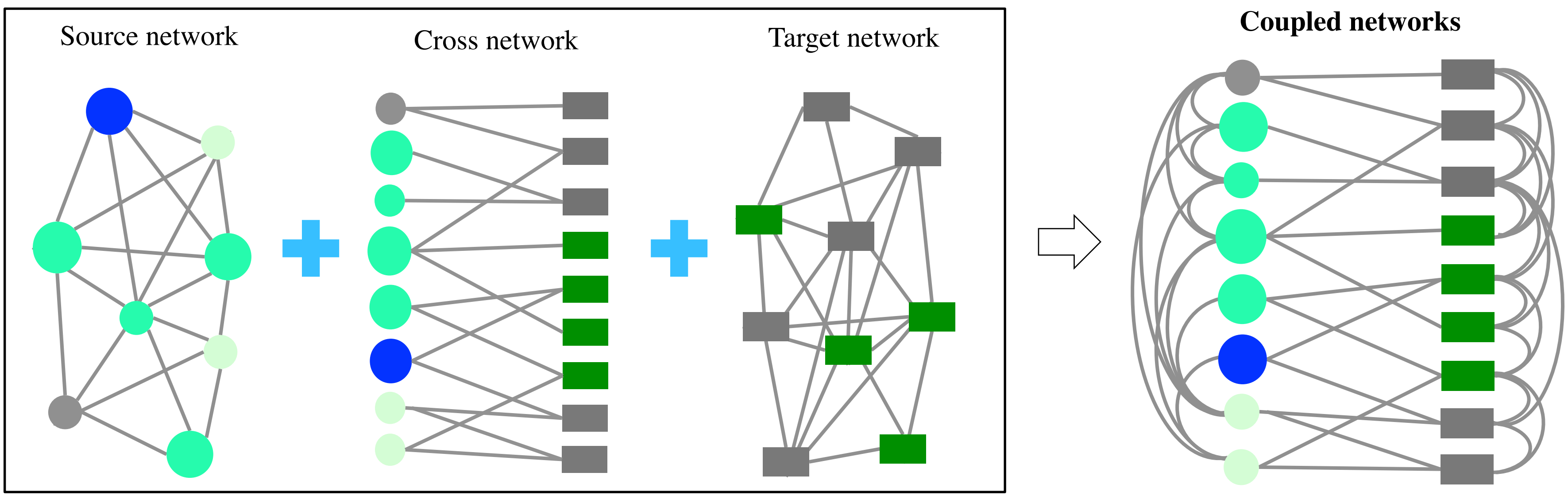




1. K.I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. **PNAS 2007.**
2. D. Davis, N. V. Chawla. Exploring and Exploiting Disease Interactions from Multi-Relational Gene and Phenotype Networks. **PLoS One 2011.**
3. J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási. Uncovering disease-disease relationships through the incomplete interactome**. Science 2015.**

# Coupled Networks

Given a source network $\boldsymbol{G^S = (V^S, E^S)}$ and a target network $\boldsymbol{G^T = (V^T, E^T)}$, they compose coupled networks if there exists a cross link $\boldsymbol{e_{ij}}$ with one node $\boldsymbol{v_i} \in \boldsymbol{V^S}$ and the other node $\boldsymbol{v_j} \in \boldsymbol{V^T}$. The cross network $\boldsymbol{G^C = (V^C, E^C)}$ is a bipartite network containing all the cross links in the coupled networks.

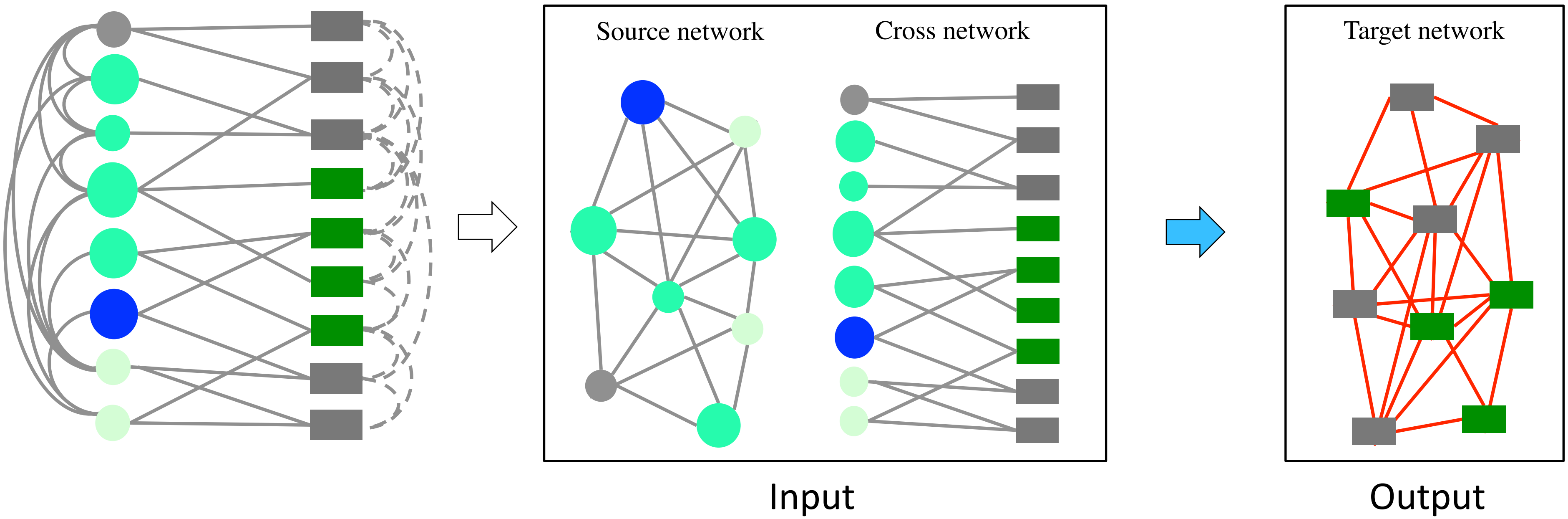


# Coupled Link Prediction

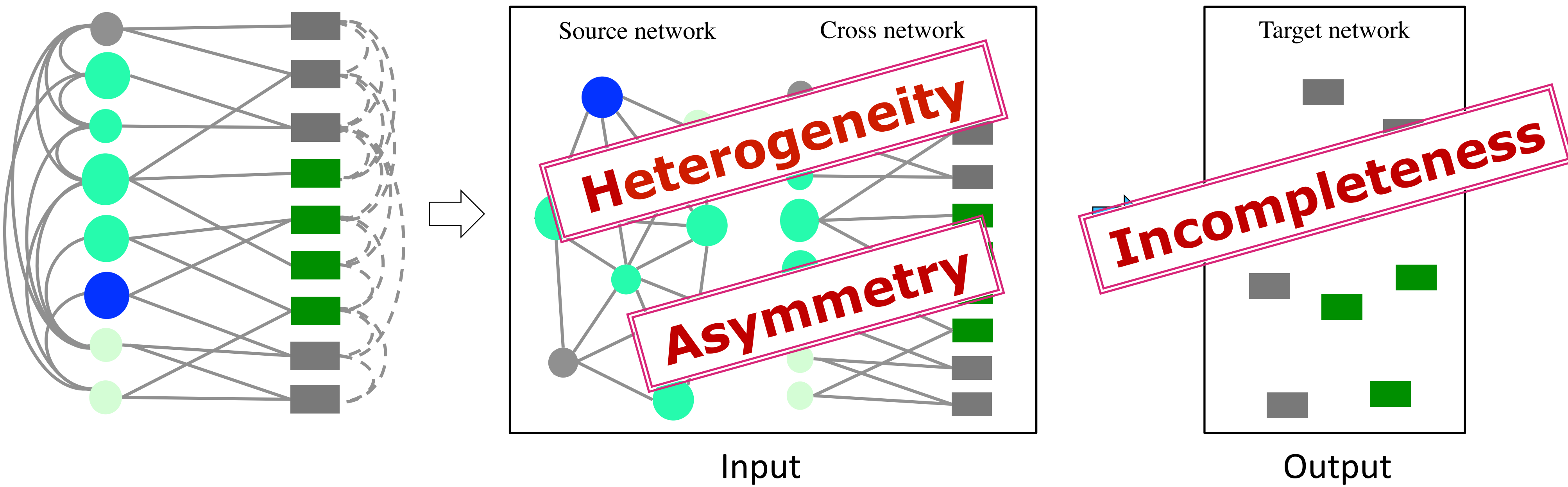
Given the source network $\boldsymbol{G^S}$ and the cross network $\boldsymbol{G^C}$ in coupled networks $\boldsymbol{G = (G^S, G^T, G^C)}$, the task is to find a predictive function:

$$\boldsymbol{f : (G^S, G^C) \rightarrow Y^T}$$

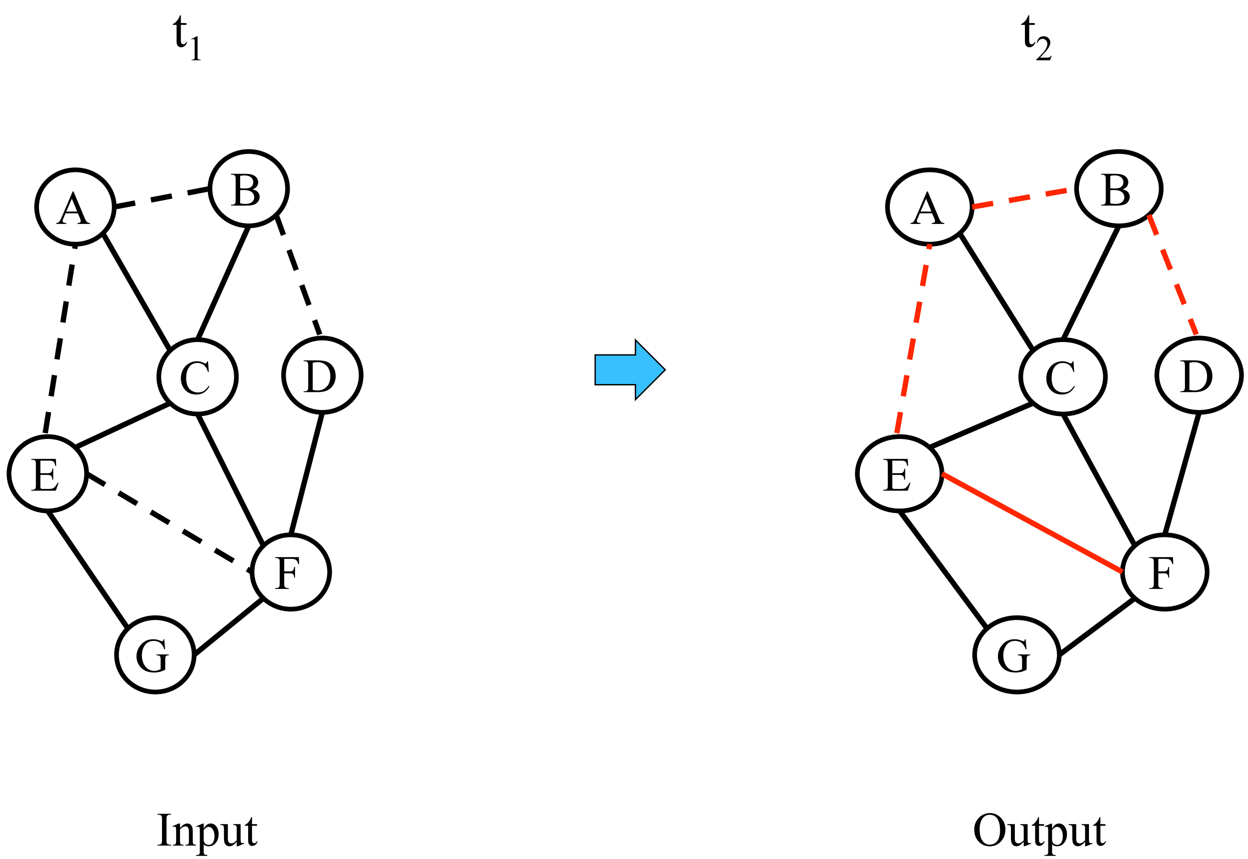
where $\boldsymbol{Y^T}$ is the set of labels for the potential links in the target network $\boldsymbol{G^T}$.



# Challenges

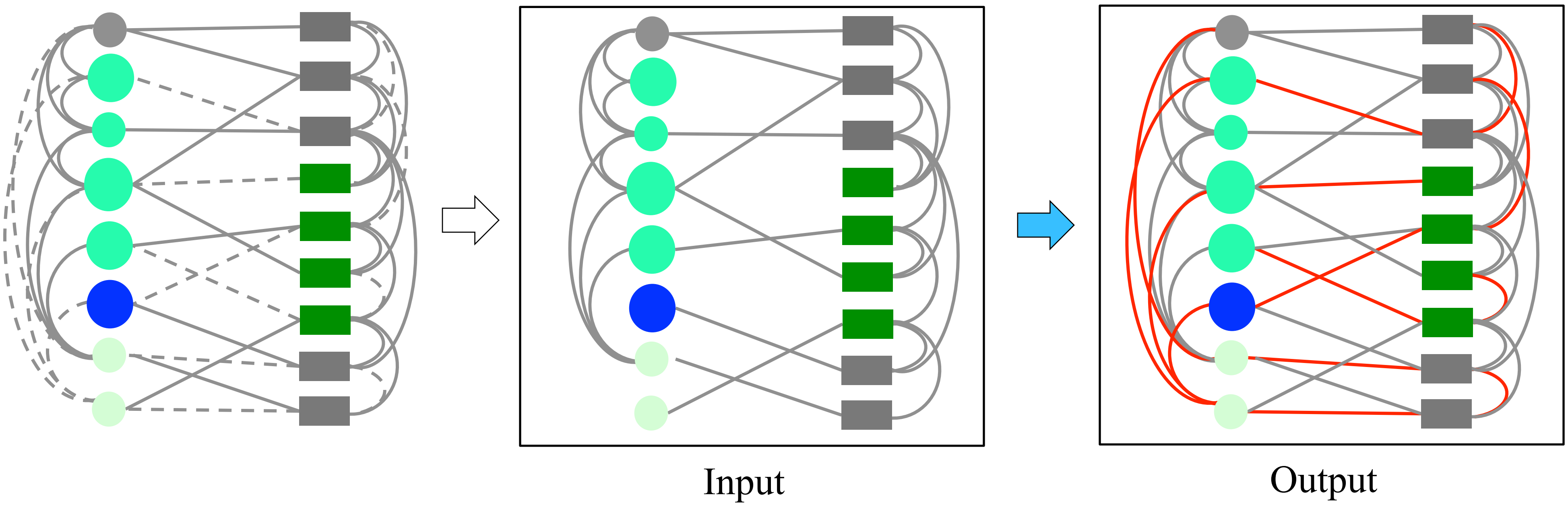

Source network
Cross network
Heterogeneity
Asymmetry
Input
Target network
Incompleteness
Output

# Related Work: Traditional Link Prediction



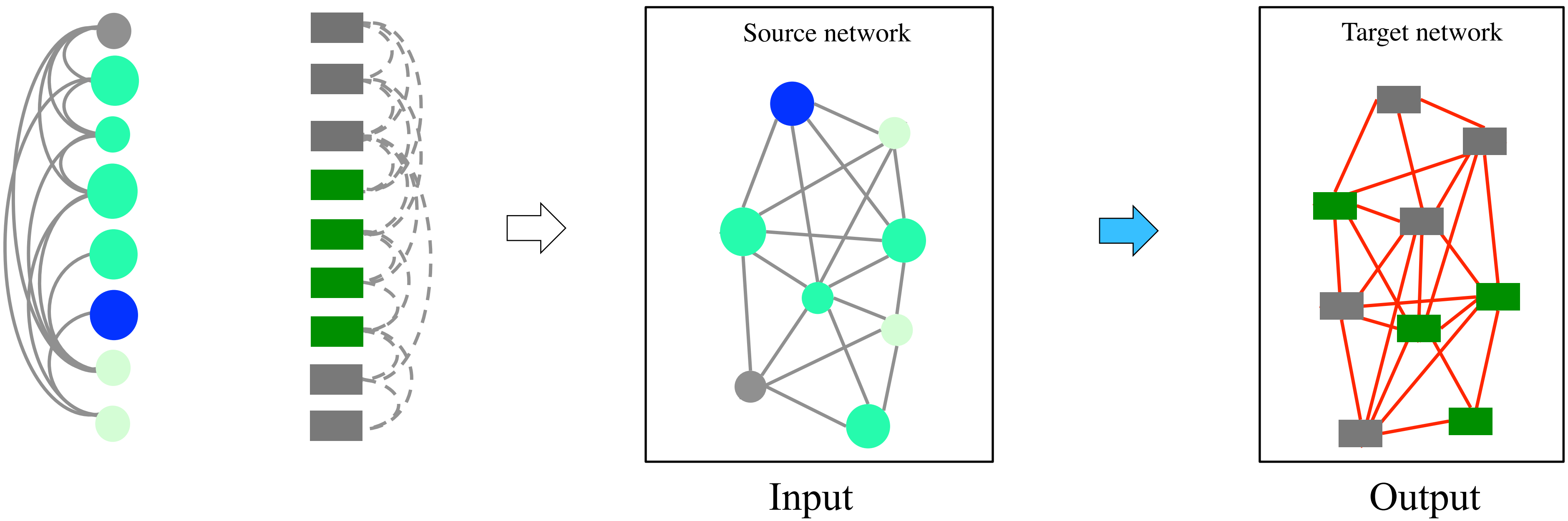

1. D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. **CIKM'03.**

# Related Work: Heterogeneous Link Prediction



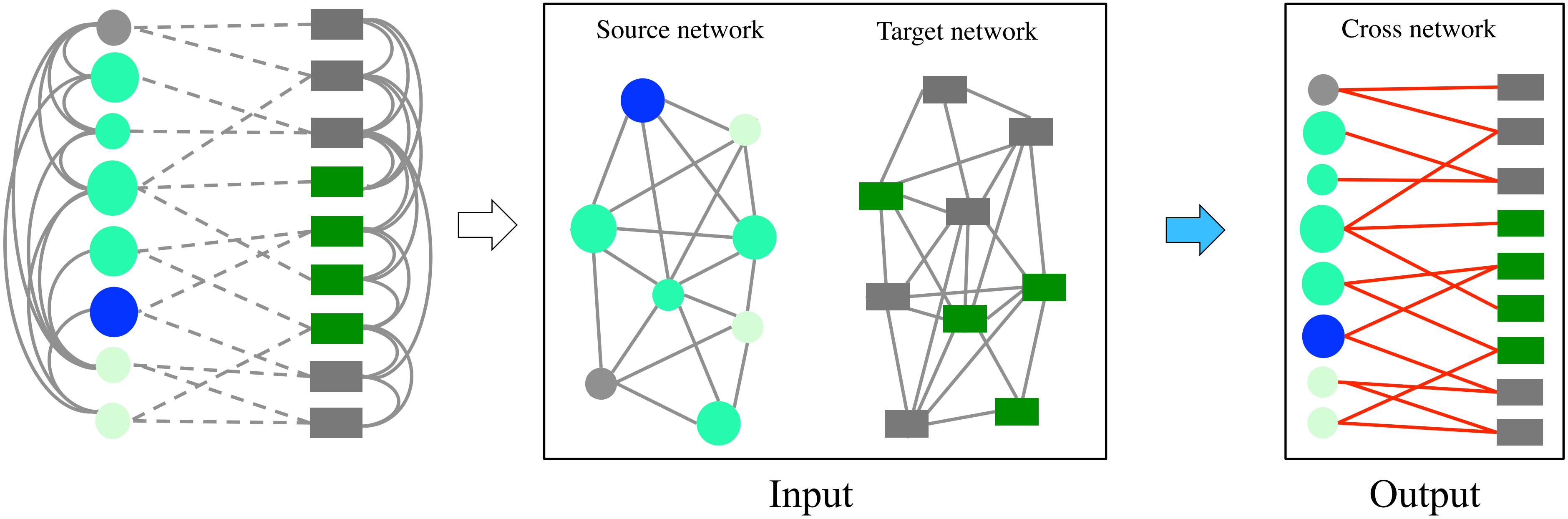

1. Y. Sun, J. Han, C. C. Aggarwal, N. V. Chawla. Will Will This Happen? Relationship Prediction in Heterogeneous Information Networks. **WSDM'12.**

# Related Work: Transfer Link Prediction



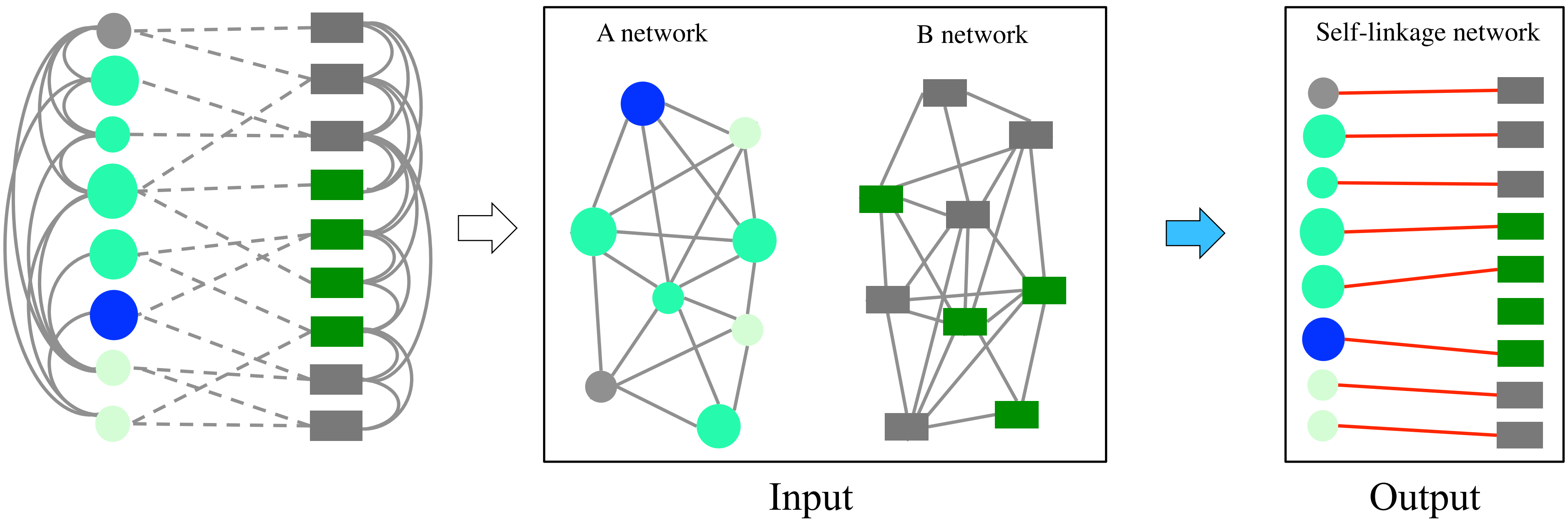

1. Y. Dong, J. Tang, S. Wu, J. Tian, N. V. Chawla. J. Rao, H. Cao. Link Prediction and Recommendation across Heterogeneous Networks. **ICDM'12**
2. J. Tang, T. Lou, J. Kleinberg, S. Wu. Transfer Link Prediction across Heterogeneous Networks. **TOIS 2015.**

# Related Work: Cross-Domain Link Prediction



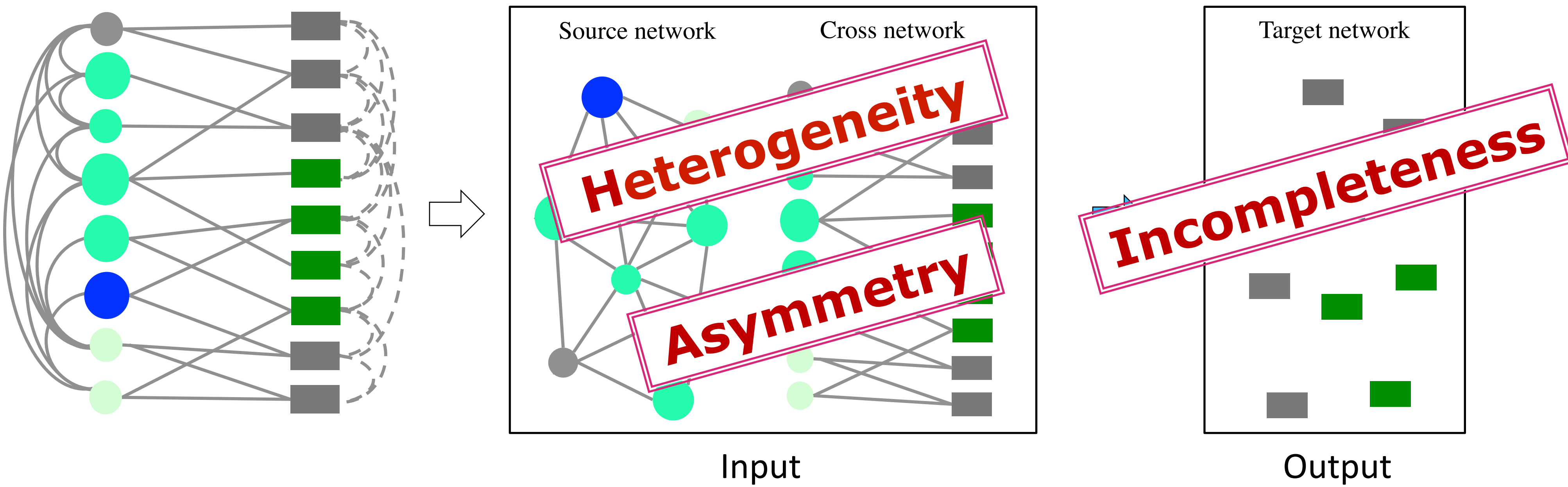

1. J. Tang, S. Wu, J. Sun, H. Su. Cross-Domain Collaboration Recommendation. **KDD'12.**

# Related Work: Anchor Link Prediction




1. X. Kong, J. Zhang, P. S. Yu. Inferring anchor links across multiple heterogeneous social networks. **CIKM'13.**
2. Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. S. Yu. COSNET: Connecting Heterogeneous Social Networks with Local and Global Consistency. **KDD'15.**

# Challenges


Source network
Cross network
Target network
Heterogeneity
Asymmetry
Incompleteness
Input
Output

# CoupledLP Framework

**1. Implicit Target Network Construction**

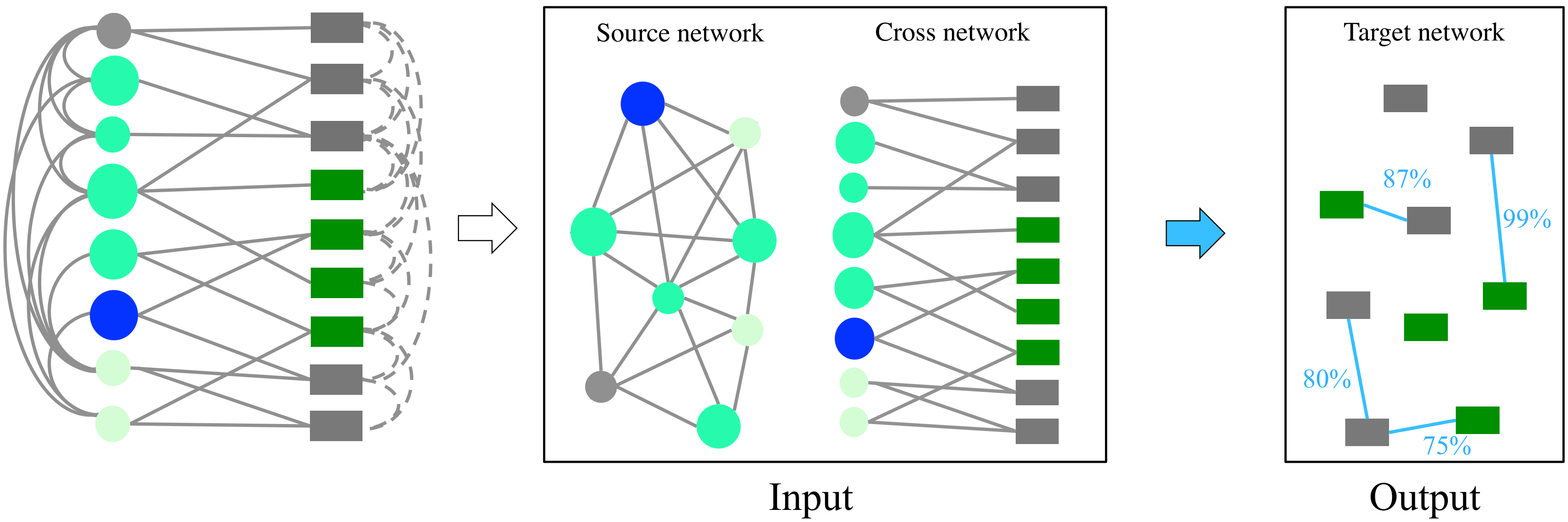
- Solve Incompleteness

**2. Coupled Factor Graph Model**

- Solve Asymmetry
- Solve Heterogeneity

# CoupledLP Framework

**Incompleteness**

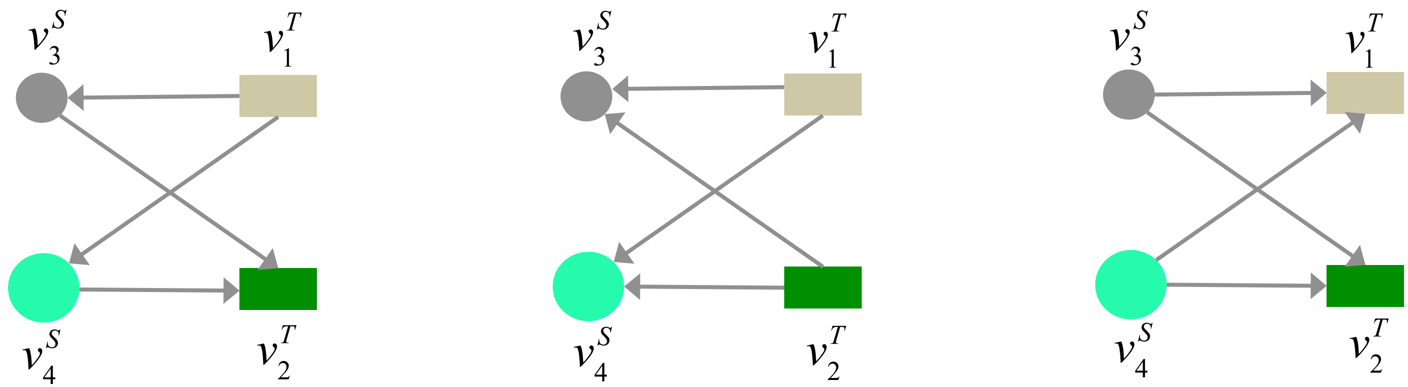



**Implicit Target Network Construction**

1. V. Leroy, B. B. Cambazoglu, and F. Bonchi. Cold start link prediction. In **KDD ’10.**

# CoupledLP: Implicit Target Network

Atomic Propagations for constructing an implicit target network



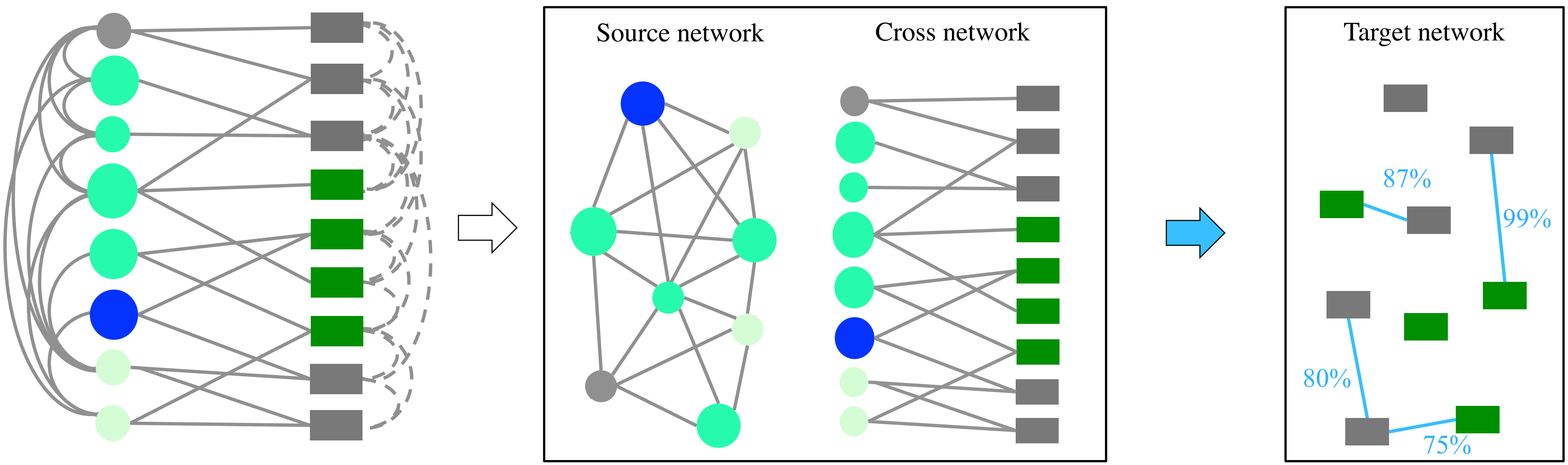

Direct　　Coupling　　Co-citation

$$\boxed{MM \quad + \quad MM^T \quad + \quad M^TM} \quad \textbf{top } z\%$$

1. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In **WWW ’04**

# CoupledLP Framework

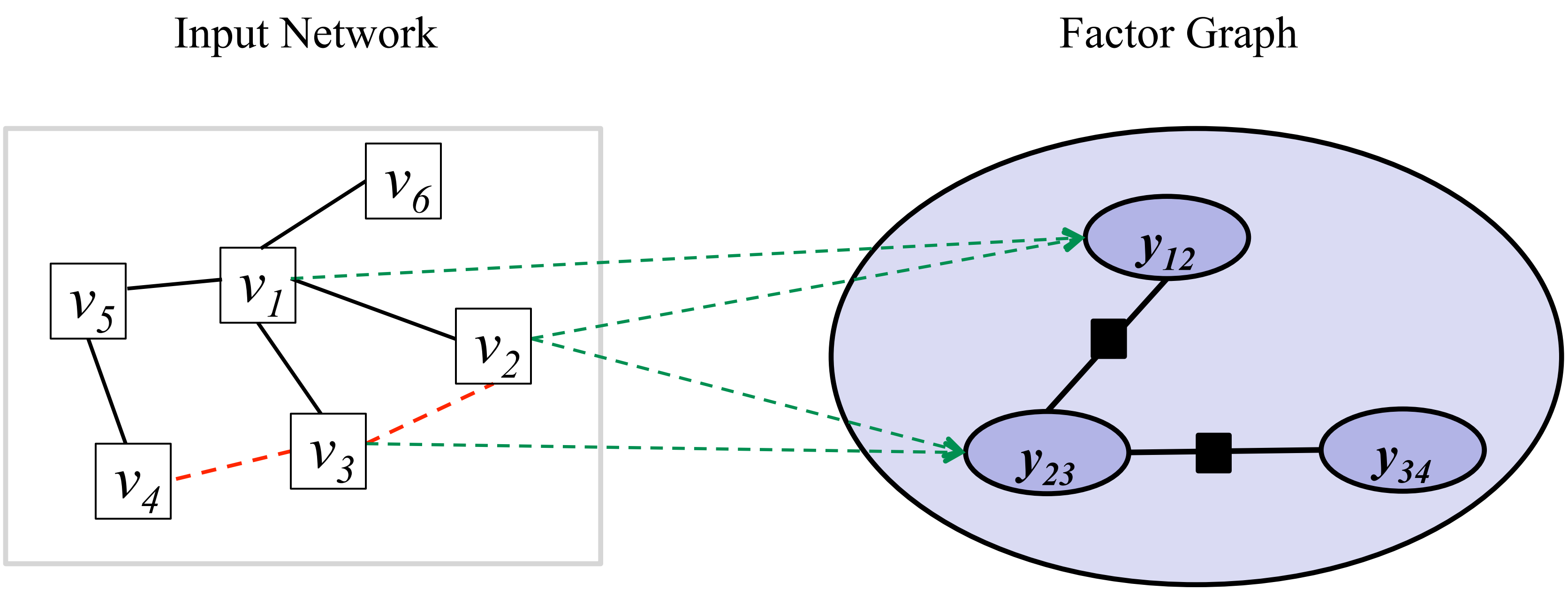
**Asymmetry** **Heterogeneity**




Input Output

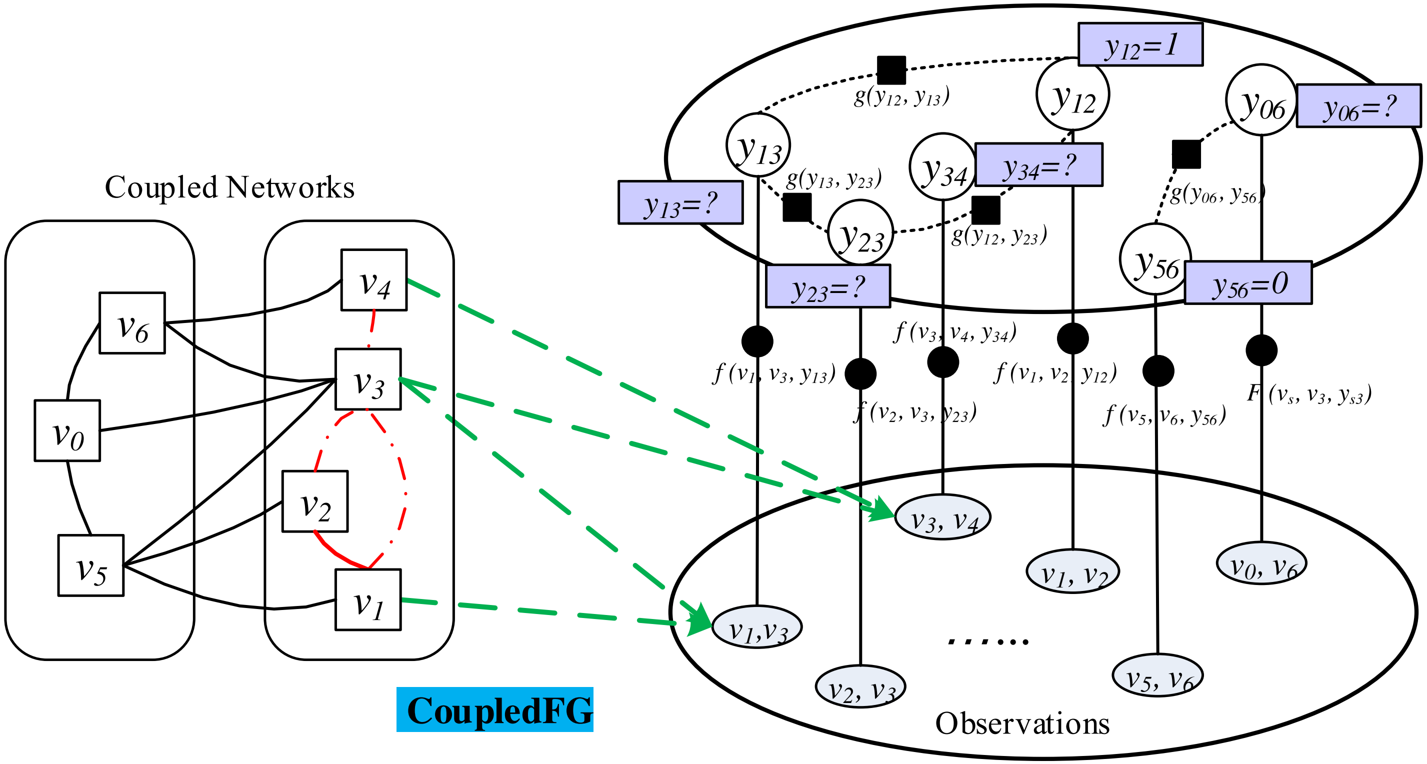
**Coupled Factor Graph**

# Basic Idea



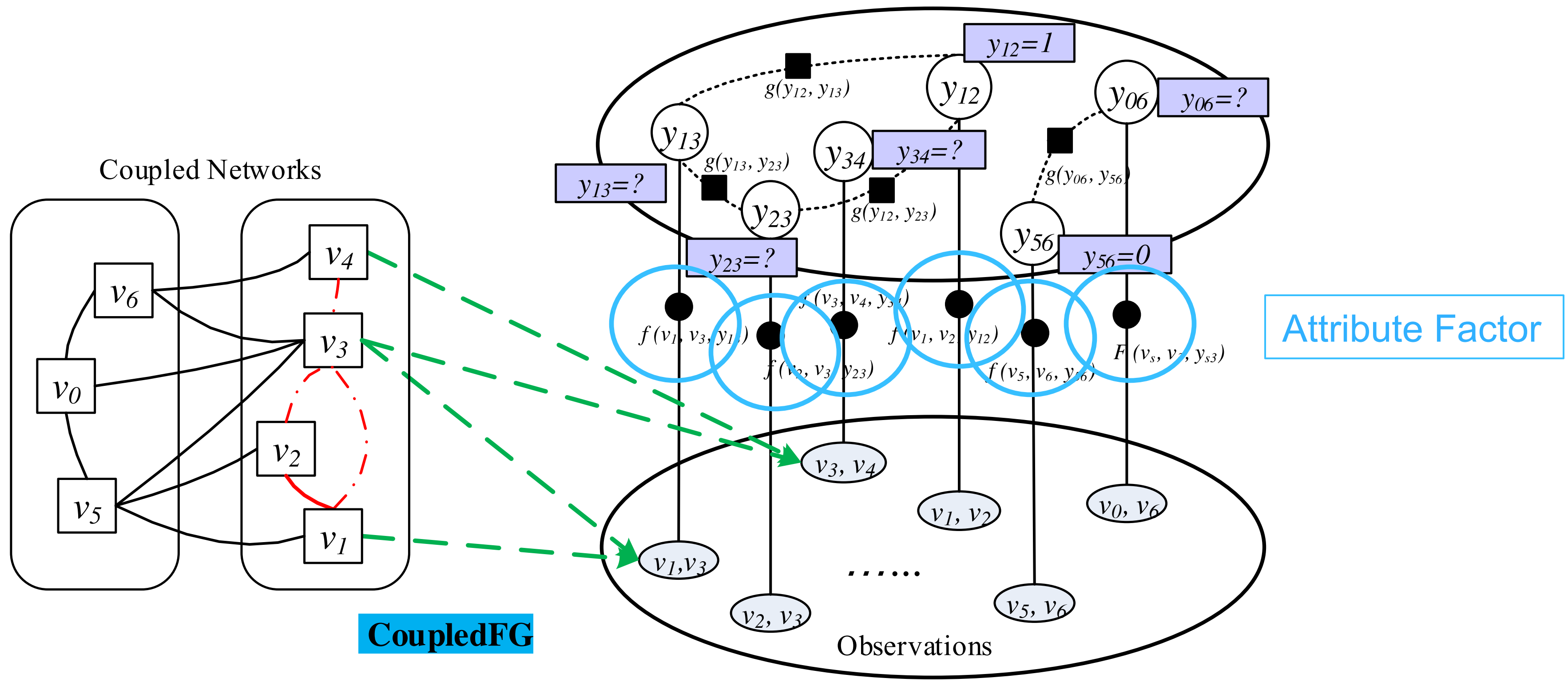

1. F. R. Kschischang, B. J. Frey, and H. andrea Loeliger. Factor graphs and the sum-product algorithm. In **IEEE TOIT 2001**.
2. Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring social ties across heterogeneous networks. In **WSDM '12**

# CoupledLP: Coupled Factor Graph




$$P(Y|\mathbf{X}, G) \propto P(\mathbf{X}|Y) \cdot P(Y|G)$$

# CoupledLP: Coupled Factor Graph




$$P(Y|\mathbf{X}, G) \propto P(\mathbf{X}|Y) \cdot P(Y|G)$$

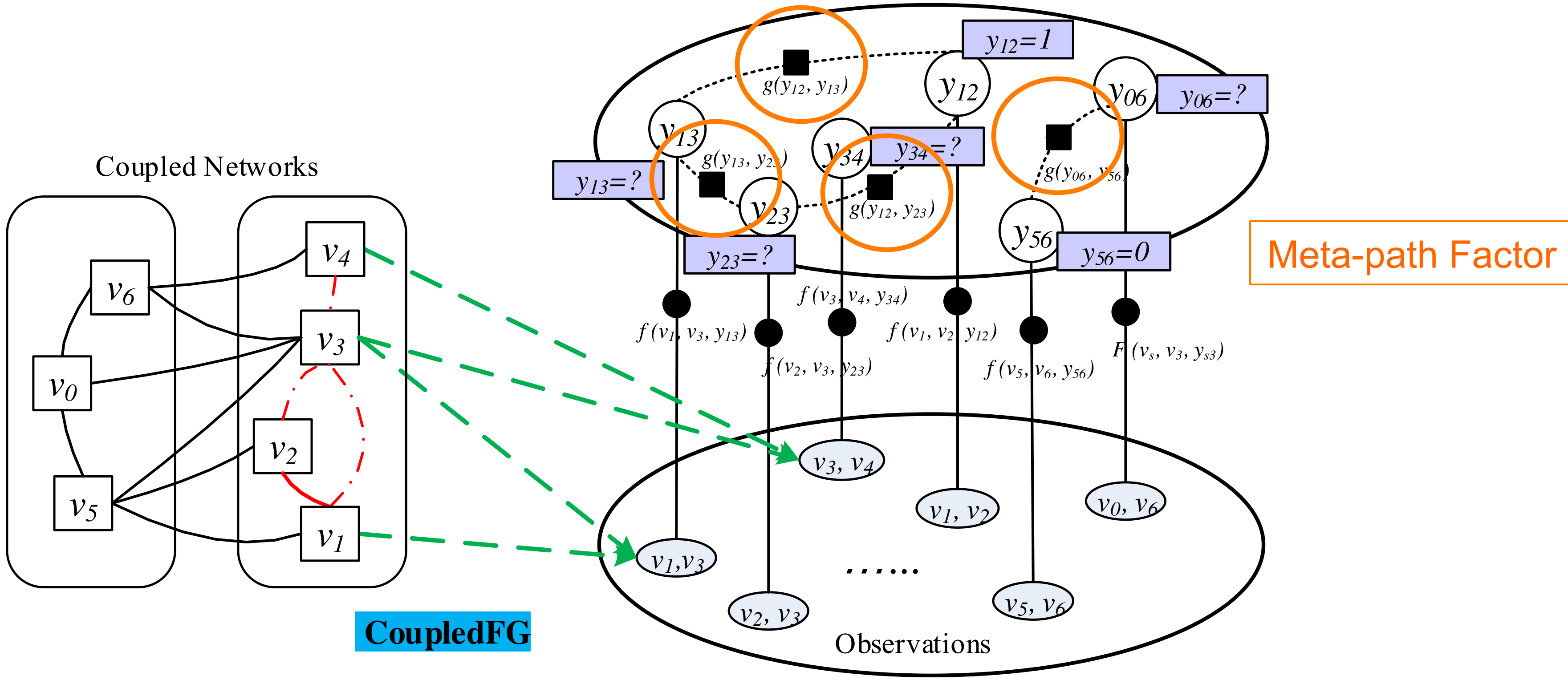
**Asymmetry**

$$\propto \prod_{e \in E^S} \prod_{k=1}^{K} P(x_{ek}^S | y_e^S) \prod_{e \in E^T} \prod_{k=1}^{K} P(x_{ek}^T | y_e^T)$$

model source and target network separately

1. Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring social ties across heterogeneous networks. In **WSDM '12.**

# CoupledLP: Coupled Factor Graph




$$P(Y|\mathbf{X},G) \propto P(\mathbf{X}|Y)\cdot P(Y|G)$$

**Asymmetry**

**Heterogeneity**

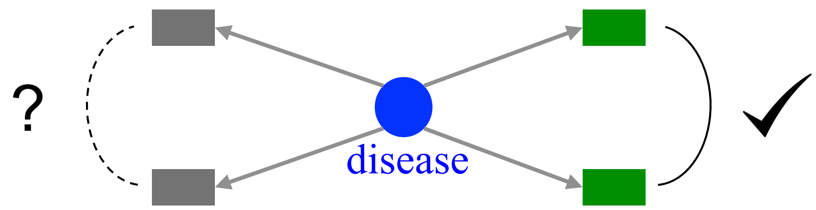
$$\propto \prod_{e\in E^S}\prod_{k=1}^{K} P(x^S_{ek}|y^S_e) \prod_{e\in E^T}\prod_{k=1}^{K} P(x^T_{ek}|y^T_e) \prod_{\pi\in\Pi} P(Y_\pi)$$

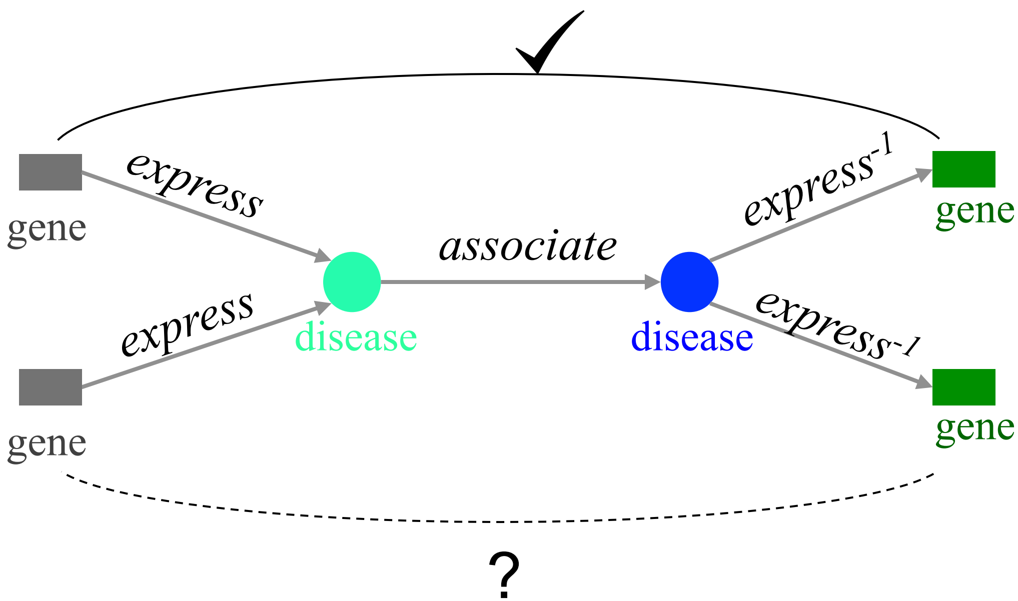
model source and target network separately

meta-path

1. Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring social ties across heterogeneous networks. In **WSDM '12.**

# CoupledLP: Coupled Factor Graph




……

**meta-path**

1. Y. Sun, J. Han, C. C. Aggarwal, N. V. Chawla. Will Will This Happen? Relationship Prediction in Heterogeneous Information Networks. In **WSDM'12.**

# CoupledLP: Coupled Factor Graph

- Factor Initialization: exponential-linear

$$P(x_{ek}^S | y_e^S) = \frac{1}{Z_\alpha} \exp\{\alpha_k f_k(x_{ek}^S, y_e^S)\}$$

$$P(x_{ek}^T | y_e^T) = \frac{1}{Z_\beta} \exp\{\beta_k g_k(x_{ek}^T, y_e^T)\}$$

$$P(Y_\pi) = \frac{1}{Z_\gamma} \exp\{\gamma_\pi h_\pi(Y_\pi)\}$$

- Objective Function:

model source & target network separately

$$\mathcal{O}(\theta) = \boxed{\sum_{e \in E^S} \left(\sum_{k=1}^{K} \alpha_k f_k(x_{ek}^S, y_e^S)\right)} + \boxed{\sum_{e \in E^T} \left(\sum_{k=1}^{K} \beta_k g_k(x_{ek}^T, y_e^T)\right)}$$
$$+ \boxed{\sum_{\pi \in \Pi} \gamma_\pi h_\pi(Y_\pi) - \log Z}$$

meta-path

bridge source & target networks

# CoupledLP: Coupled Factor Graph

**Input**: a source network $G_S$, a target network $G_T$, and the learning rate $\eta$
**Output**: estimated parameters $\theta = (\{\alpha\}, \{\beta\}, \{\mu\})$

Initialize $\theta \leftarrow 0$;
Perform statistics according to social theories;
Construct social theories based features $h_k(Y_c)$;
**repeat**
**Step 1:** Perform LBP to calculate marginal distribution of unknown variables in the source network $P(y_i|x_i, G_S)$;
**Step 2:** Perform LBP to calculate marginal distribution of unknown variables in the target network $P(y_i|x_i, G_T)$;
**Step 3:** Perform LBP to calculate the marginal distribution of clique $c$, i.e., $P(y_c|\mathbf{X}_c^S, \mathbf{X}_c^T, G_S, G_T)$;
**Step 4:** Calculate the gradient of $\mu_k$ according to Eq. 8 (for $\alpha_j$ and $\beta_j$ with a similar formula);
**Step 5:** Update parameter $\theta$ with the learning rate $\eta$:

$$\theta_{\text{new}} = \theta_{\text{old}} + \eta \cdot \frac{\mathcal{O}(\theta)}{\theta}$$

**until** *Convergence*;

Learning: Gradient Decent method

1. Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring social ties across heterogeneous networks. In **WSDM '12.**

# CoupledLP Framework

**1. Implicit Target Network Construction**

- Solve Incompleteness

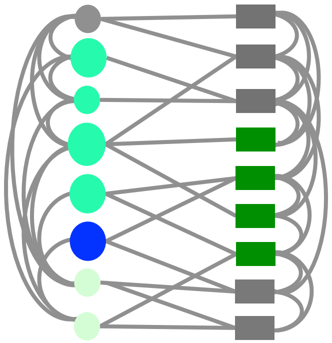
**2. Coupled Factor Graph Model**

- Solve Asymmetry
- Solve Heterogeneity

# Experiments: Data

***k***: average degree; ***cc***: clustering coefficient; ***ac***: associative coefficient

| | $D$ | $G$ | $D \leftrightarrow G$ | $A_a$ | $A_b$ | $A_a \leftrightarrow A_b$ | $E_a$ | $E_b$ | $E_c$ | $E_a \leftrightarrow E_b$ | $E_a \leftrightarrow E_c$ | $E_b \leftrightarrow E_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #Nodes | 703 | 1132 | 1835 | 348,640 | 63,687 | 235,715 | 2,531,187 | 655,755 | 354,166 | 1,912,933 | 1,255,046 | 625,379 |
| #Links | 74523 | 2450 | 10483 | 613,614 | 96,325 | 306,213 | 3,355,197 | 649,322 | 311,432 | 1,844,342 | 1,131,593 | 507,894 |
| $k$ | 212.01 | 4.33 | 11.43 | 3.52 | 3.02 | 2.59 | 2.65 | 1.98 | 1.75 | 1.92 | 1.80 | 1.62 |
| $cc$ | 0.2639 | 0.0377 | 0 | 0.0237 | 0.0225 | 0 | 0.0457 | 0.0366 | 0.0317 | 0 | 0 | 0 |
| $ac$ | -0.0256 | 0.1761 | -0.2556 | 0.2011 | 0.1671 | 0.0654 | 0.2848 | 0.2693 | 0.2806 | 0.0231 | -0.0305 | 0.1113 |



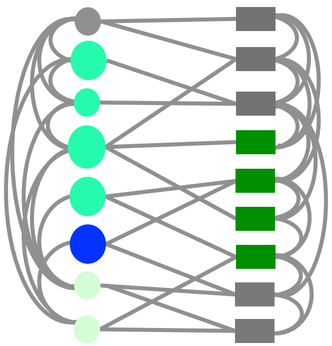
Healthcare Networks
Disease (**D**)---Gene (**G**)


1. D. Davis, N. V. Chawla. Exploring and Exploiting Disease Interactions from Multi-Relational Gene and Phenotype Networks. **PLoS One 2011.**
2. N. Du, C. Faloutsos, B. Wang, and L. Akoglu. Large human communication networks: patterns and a utility-driven generator. In **KDD '09.**
3. Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In **KDD'14.**

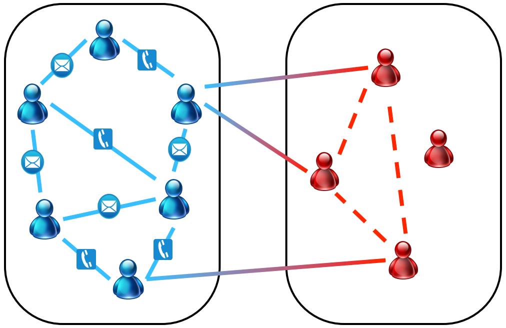
# Experiments: Data

***k***: average degree; ***cc***: clustering coefficient; ***ac***: associative coefficient

| | $D$ | $G$ | $D \leftrightarrow G$ | $A_a$ | $A_b$ | $A_a \leftrightarrow A_b$ | $E_a$ | $E_b$ | $E_c$ | $E_a \leftrightarrow E_b$ | $E_a \leftrightarrow E_c$ | $E_b \leftrightarrow E_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #Nodes | 703 | 1132 | 1835 | 348,640 | 63,687 | 235,715 | 2,531,187 | 655,755 | 354,166 | 1,912,933 | 1,255,046 | 625,379 |
| #Links | 74523 | 2450 | 10483 | 613,614 | 96,325 | 306,213 | 3,355,197 | 649,322 | 311,432 | 1,844,342 | 1,131,593 | 507,894 |
| $k$ | 212.01 | 4.33 | 11.43 | 3.52 | 3.02 | 2.59 | 2.65 | 1.98 | 1.75 | 1.92 | 1.80 | 1.62 |
| $cc$ | 0.2639 | 0.0377 | 0 | 0.0237 | 0.0225 | 0 | 0.0457 | 0.0366 | 0.0317 | 0 | 0 | 0 |
| $ac$ | -0.0256 | 0.1761 | -0.2556 | 0.2011 | 0.1671 | 0.0654 | 0.2848 | 0.2693 | 0.2806 | 0.0231 | -0.0305 | 0.1113 |



Healthcare Networks
Disease (**D**)---Gene (**G**)



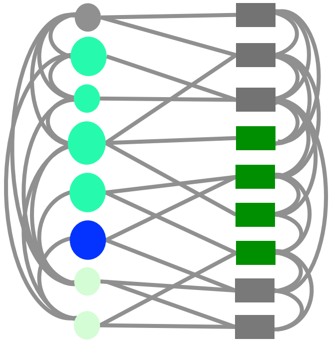
Mobile Phone Call Networks
**Two** Operators: ***Aa---Ab***

1. D. Davis, N. V. Chawla. Exploring and Exploiting Disease Interactions from Multi-Relational Gene and Phenotype Networks. **PLoS One 2011.**
2. N. Du, C. Faloutsos, B. Wang, and L. Akoglu. Large human communication networks: patterns and a utility-driven generator. In **KDD '09.**
3. Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In **KDD'14.**

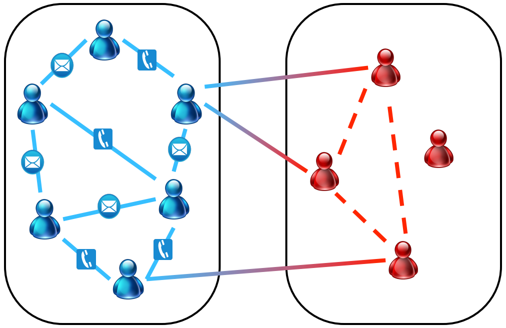
# Experiments: Data

$k$: average degree; $cc$: clustering coefficient; $ac$: associative coefficient

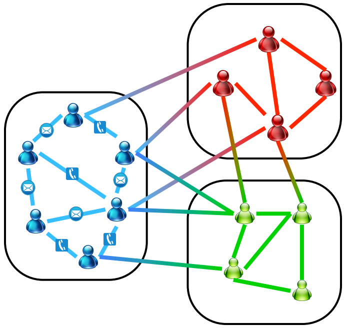
| | $D$ | $G$ | $D \leftrightarrow G$ | $A_a$ | $A_b$ | $A_a \leftrightarrow A_b$ | $E_a$ | $E_b$ | $E_c$ | $E_a \leftrightarrow E_b$ | $E_a \leftrightarrow E_c$ | $E_b \leftrightarrow E_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #Nodes | 703 | 1132 | 1835 | 348,640 | 63,687 | 235,715 | 2,531,187 | 655,755 | 354,166 | 1,912,933 | 1,255,046 | 625,379 |
| #Links | 74523 | 2450 | 10483 | 613,614 | 96,325 | 306,213 | 3,355,197 | 649,322 | 311,432 | 1,844,342 | 1,131,593 | 507,894 |
| $k$ | 212.01 | 4.33 | 11.43 | 3.52 | 3.02 | 2.59 | 2.65 | 1.98 | 1.75 | 1.92 | 1.80 | 1.62 |
| $cc$ | 0.2639 | 0.0377 | 0 | 0.0237 | 0.0225 | 0 | 0.0457 | 0.0366 | 0.0317 | 0 | 0 | 0 |
| $ac$ | -0.0256 | 0.1761 | -0.2556 | 0.2011 | 0.1671 | 0.0654 | 0.2848 | 0.2693 | 0.2806 | 0.0231 | -0.0305 | 0.1113 |



Healthcare Networks
Disease (**D**)---Gene (**G**)



Mobile Phone Call Networks
**Two** Operators: ***Aa---Ab***



Mobile Phone Call Networks
**Three** Operators: ***Ea---Eb---Ec***


1. D. Davis, N. V. Chawla. Exploring and Exploiting Disease Interactions from Multi-Relational Gene and Phenotype Networks. **PLoS One 2011.**
2. N. Du, C. Faloutsos, B. Wang, and L. Akoglu. Large human communication networks: patterns and a utility-driven generator. In **KDD ’09.**
3. Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In **KDD’14.**

# Experiments: Data

***k***: average degree; ***cc***: clustering coefficient; ***ac***: associative coefficient

| | $D$ | $G$ | $D \leftrightarrow G$ | $A_a$ | $A_b$ | $A_a \leftrightarrow A_b$ | $E_a$ | $E_b$ | $E_c$ | $E_a \leftrightarrow E_b$ | $E_a \leftrightarrow E_c$ | $E_b \leftrightarrow E_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #Nodes | 703 | 1132 | 1835 | 348,640 | 63,687 | 235,715 | 2,531,187 | 655,755 | 354,166 | 1,912,933 | 1,255,046 | 625,379 |
| #Links | 74523 | 2450 | 10483 | 613,614 | 96,325 | 306,213 | 3,355,197 | 649,322 | 311,432 | 1,844,342 | 1,131,593 | 507,894 |
| $k$ | 212.01 | 4.33 | 11.43 | 3.52 | 3.02 | 2.59 | 2.65 | 1.98 | 1.75 | 1.92 | 1.80 | 1.62 |
| $cc$ | 0.2639 | 0.0377 | 0 | 0.0237 | 0.0225 | 0 | 0.0457 | 0.0366 | 0.0317 | 0 | 0 | 0 |
| $ac$ | -0.0256 | 0.1761 | -0.2556 | 0.2011 | 0.1671 | 0.0654 | 0.2848 | 0.2693 | 0.2806 | 0.0231 | -0.0305 | 0.1113 |

**Asymmetry**

**Heterogeneity**

# Experiments: Coupled Networks

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Statistics | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
| #Candidate links | 243,393 | 19,014 | 376,416 | 1,280,959 | 972,808 | 2,594,169 | 424,793 | 1,655,878 | 252,471 | 372,421 |
| #Positive links | 1,582 | 11,015 | 25,694 | 57,138 | 179,265 | 373,511 | 83,657 | 232,814 | 46,954 | 63,544 |
| %Positive links | 0.65% | 57.93% | 6.83% | 4.46% | 18.43% | 14.40% | 19.69% | 14.06% | 18.60% | 17.06% |

10 Coupled Prediction Cases

# Experiments

**AUPR or AUROC or Precision @ k**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | | | | | | | | | | |
| AA | | | | | | | | | | |
| JC | | | | | | | | | | |
| PA | | | | | | | | | | |
| PF | | | | | | | | | | |
| IT | | | | | | | | | | |
| LRC-IT | | | | | | | | | | |
| LRC | | | | | | | | | | |
| DT-IT | | | | | | | | | | |
| DT | | | | | | | | | | |
| CoupledLP-IT | | | | | | | | | | |
| CoupledLP | | | | | | | | | | |

10 Coupled Prediction Cases

# Baselines

**AUPR or AUROC or Precision @ k**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | | | | | | | | | | |
| AA | | | | | | | | | | |
| JC | | | | | | | | | | |
| PA | | | | | | | | | | |
| PF | | | | | | | | | | |
| IT | | | | | | | | | | |
| LRC-IT | | | | | | | | | | |
| LRC | | | | | | | | | | |
| DT-IT | | | | | | | | | | |
| DT | | | | | | | | | | |
| CoupledLP-IT | | | | | | | | | | |
| CoupledLP | | | | | | | | | | |

**Unsupervised Methods:**

- ✓Common Neighbors (CN)
- ✓Adamic Adar (AA)
- ✓Jaccard Coefficient (JC)
- ✓Preferential Attachment (PA)
- ✓PropFlow (PF)
- ✓Implicit Target Network (IT)

features →

**Supervised Methods:**

- ✓Logistic Regression (LRC)
  - LRC-IT
- ✓Decision Tree (DT)
  - DT-IT
- ✓CoupledLP
  - ✓CoupledLP-IT

LRC-IT, DT-IT, CoupledLP-IT: **NO** **I**mplicit **T**arget network construction


1. R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In **KDD '10.**
2. L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In **WSDM'11.**

# Experiments

**AUPR or AUROC or Precision @ k**

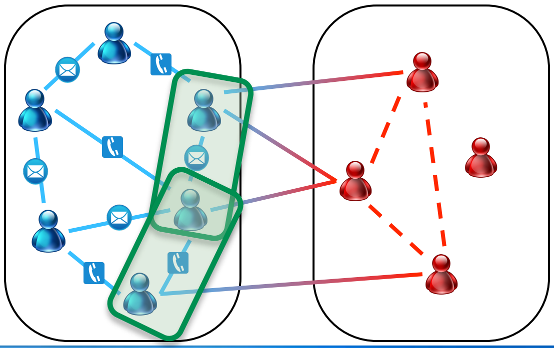
| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | | | | | | | | | | |
| AA | | | | | | | | | | |
| JC | | | | | | | | | | |
| PA | | | | | | | | | | |
| PF | | | | | | | | | | |
| IT | | | | | | | | | | |
| LRC-IT | | | | | | | | | | |
| LRC | | | | | | | | | | |
| DT-IT | | | | | | | | | | |
| DT | | | | | | | | | | |
| CoupledLP-IT | | | | | | | | | | |
| CoupledLP | | | | | | | | | | |

Training Links:

- ✓source links between nodes with cross links
- ✓1% target links

✓Test Links:

- ✓99% target links



# Evaluation Metrics

**AUPR or AUROC or Precision @ k**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | | | | | | | | | | |
| AA | | | | | | | | | | |
| JC | | | | | | | | | | |
| PA | | | | | | | | | | |
| PF | | | | | | | | | | |
| IT | | | | | | | | | | |
| LRC-IT | | | | | | | | | | |
| LRC | | | | | | | | | | |
| DT-IT | | | | | | | | | | |
| DT | | | | | | | | | | |
| CoupledLP-IT | | | | | | | | | | |
| CoupledLP | | | | | | | | | | |

- ✓ Area Under Precision Recall Curve (AUPR)
- ✓ Area Under Receiver Operating Characteristic Curve (AUROC)
- ✓ Precision at Top k


1. R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In **KDD '10.**
2. L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In **WSDM'11.**

# AUPR Results

**AUPR**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | 0.0155 | 0.6011 | 0.3017 | 0.1348 | 0.3598 | 0.2319 | 0.3817 | 0.2079 | 0.3145 | 0.2654 |
| AA | 0.0167 | 0.5912 | 0.3344 | 0.1596 | 0.4541 | 0.2800 | 0.4838 | 0.2562 | 0.3802 | 0.3180 |
| JC | 0.0803 | 0.4812 | 0.0835 | 0.0903 | 0.3848 | 0.3082 | 0.4140 | 0.3429 | 0.3628 | 0.3579 |
| PA | 0.0083 | 0.7566 | 0.0820 | 0.0599 | 0.1446 | 0.1287 | 0.1525 | 0.1250 | 0.1560 | 0.1471 |
| PF | 0.0233 | 0.5501 | 0.1455 | 0.0989 | 0.3504 | 0.2248 | 0.3722 | 0.2138 | 0.2833 | 0.2446 |
| IT | 0.0155 | 0.6011 | 0.3715 | 0.2059 | 0.4344 | 0.3157 | 0.4568 | 0.2940 | 0.4008 | 0.3559 |
| LRC-IT | 0.0140 | 0.7830 | 0.3610 | 0.1880 | 0.4580 | 0.3140 | 0.5240 | 0.2870 | 0.4230 | 0.3500 |
| LRC | 0.0190 | 0.7930 | 0.3820 | 0.2030 | 0.4920 | 0.3160 | 0.5190 | 0.2910 | 0.4270 | 0.3590 |
| DT-IT | 0.0070 | 0.6270 | 0.2760 | 0.1050 | 0.3440 | 0.1620 | 0.3810 | 0.1550 | 0.2900 | 0.2260 |
| DT | 0.0080 | 0.6310 | 0.2530 | 0.1030 | 0.3580 | 0.1640 | 0.3470 | 0.1557 | 0.3060 | 0.2420 |
| CoupledLP-IT | 0.0303 | 0.8249 | 0.4291 | 0.2483 | 0.5088 | 0.3484 | 0.5257 | 0.3240 | 0.4537 | 0.3855 |
| CoupledLP | 0.0249 | 0.8432 | 0.4305 | 0.2776 | 0.5481 | 0.3591 | 0.5420 | 0.3399 | 0.4692 | 0.4133 |

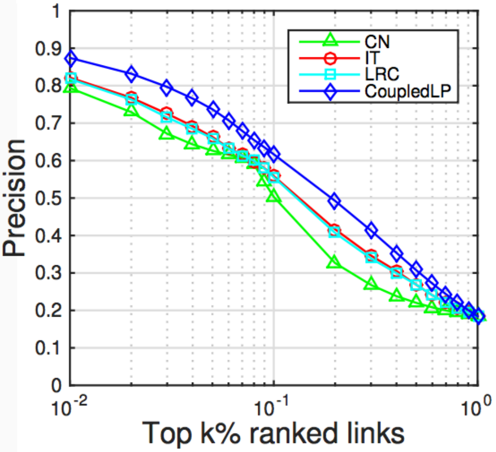
# AUROC Results

**AUROC**

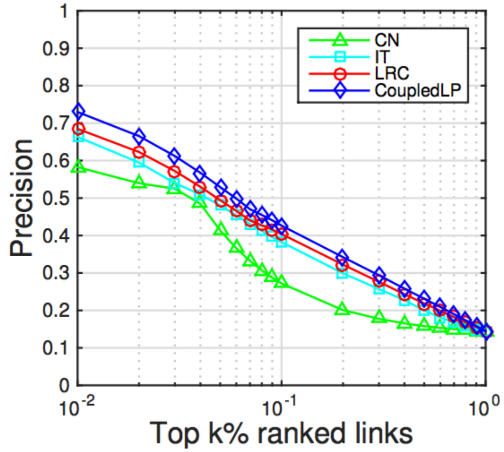
| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | 0.6384 | 0.5330 | 0.6754 | 0.5896 | 0.6090 | 0.5556 | 0.6133 | 0.5418 | 0.5736 | 0.5552 |
| AA | 0.6544 | 0.5289 | 0.7658 | 0.6933 | 0.7408 | 0.6664 | 0.7486 | 0.6357 | 0.6826 | 0.6543 |
| JC | 0.6507 | 0.3666 | 0.5974 | 0.5220 | 0.7186 | 0.6116 | 0.7280 | 0.5977 | 0.6652 | 0.6327 |
| PA | 0.4850 | 0.7073 | 0.5802 | 0.5615 | 0.3835 | 0.4460 | 0.3746 | 0.4462 | 0.4131 | 0.4270 |
| PF | 0.6426 | 0.4890 | 0.7275 | 0.7006 | 0.7339 | 0.6649 | 0.7389 | 0.6554 | 0.6736 | 0.5552 |
| IT | 0.6384 | 0.5330 | 0.7735 | 0.7273 | 0.6867 | 0.6435 | 0.6969 | 0.6335 | 0.6756 | 0.6618 |
| LRC-IT | 0.5450 | 0.7160 | 0.7590 | 0.7280 | 0.7580 | 0.6930 | 0.7750 | 0.6840 | 0.7200 | 0.6890 |
| LRC | 0.6230 | 0.7320 | 0.8210 | 0.7750 | 0.7670 | 0.7070 | 0.7730 | 0.6950 | 0.7290 | 0.7030 |
| DT-IT | 0.5010 | 0.5830 | 0.7190 | 0.6260 | 0.6690 | 0.5480 | 0.6930 | 0.5410 | 0.6340 | 0.5920 |
| DT | 0.5140 | 0.5930 | 0.7460 | 0.6530 | 0.6750 | 0.5510 | 0.6730 | 0.5440 | 0.6450 | 0.6040 |
| CoupledLP-IT | 0.6825 | 0.7586 | 0.8052 | 0.7424 | 0.7597 | 0.7017 | 0.7664 | 0.6885 | 0.7314 | 0.7004 |
| CoupledLP | 0.6790 | 0.7865 | 0.8336 | 0.7807 | 0.7779 | 0.7127 | 0.7769 | 0.7016 | 0.7405 | 0.7157 |

# Precision@k Results

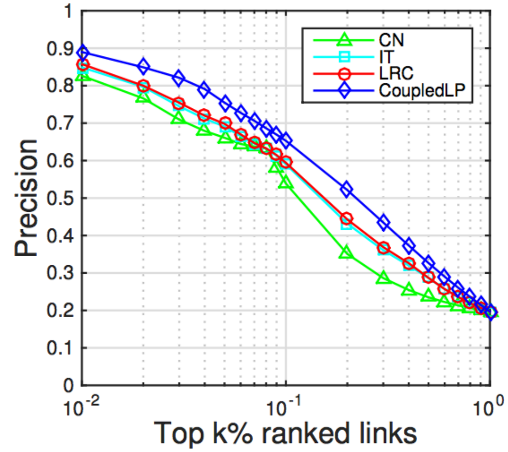
**Precision @ k**



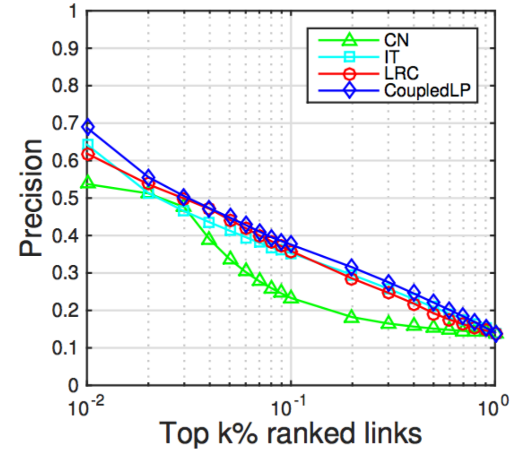

(a) European $E_a$ to $E_b$



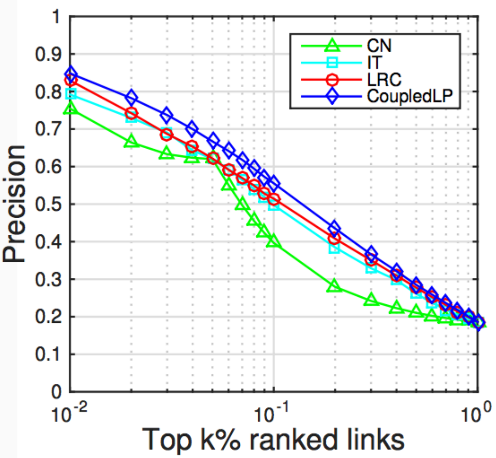

(b) European $E_b$ to $E_a$



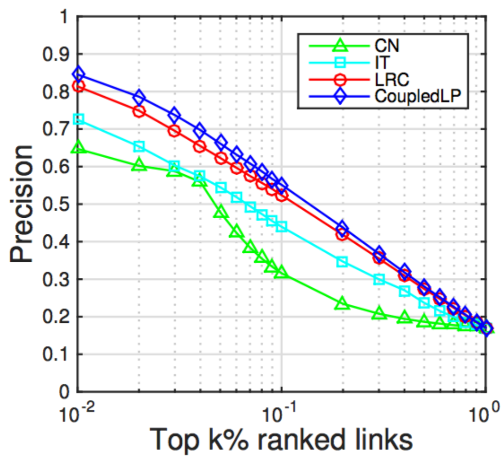

(c) European $E_a$ to $E_c$



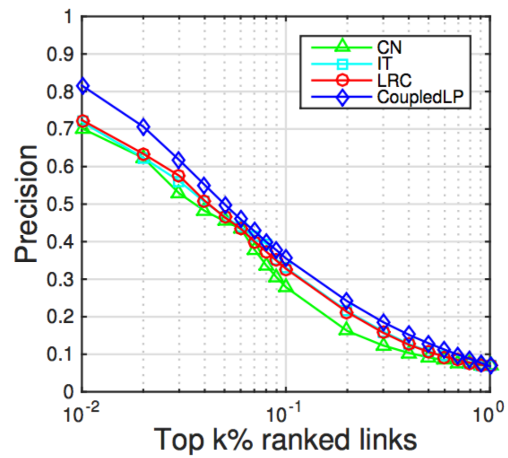

(d) European $E_c$ to $E_a$



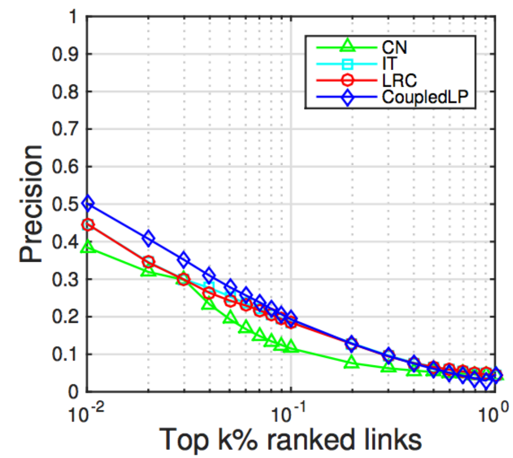

(e) European $E_b$ to $E_c$




(f) European $E_c$ to $E_b$




(g) Asian $A_a$ to $A_b$




(h) Asian $A_b$ to $A_a$

# Effects of Implicit Target Network

**AUPR**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | 0.0155 | 0.6011 | 0.3017 | 0.1348 | 0.3598 | 0.2319 | 0.3817 | 0.2079 | 0.3145 | 0.2654 |
| AA | 0.0167 | 0.5912 | 0.3344 | 0.1596 | 0.4541 | 0.2800 | 0.4838 | 0.2562 | 0.3802 | 0.3180 |
| JC | 0.0803 | 0.4812 | 0.0835 | 0.0903 | 0.3848 | 0.3082 | 0.4140 | 0.3429 | 0.3628 | 0.3579 |
| PA | 0.0083 | 0.7566 | 0.0820 | 0.0599 | 0.1446 | 0.1287 | 0.1525 | 0.1250 | 0.1560 | 0.1471 |
| PF | 0.0233 | 0.5501 | 0.1455 | 0.0989 | 0.3504 | 0.2248 | 0.3722 | 0.2138 | 0.2833 | 0.2446 |
| IT | 0.0155 | 0.6011 | 0.3715 | 0.2059 | 0.4344 | 0.3157 | 0.4568 | 0.2940 | 0.4008 | 0.3559 |
| LRC-IT | 0.0140 | 0.7830 | 0.3610 | 0.1880 | 0.4580 | 0.3140 | 0.5240 | 0.2870 | 0.4230 | 0.3500 |
| LRC | 0.0190 | 0.7930 | 0.3820 | 0.2030 | 0.4920 | 0.3160 | 0.5190 | 0.2910 | 0.4270 | 0.3590 |
| DT-IT | 0.0070 | 0.6270 | 0.2760 | 0.1050 | 0.3440 | 0.1620 | 0.3810 | 0.1550 | 0.2900 | 0.2260 |
| DT | 0.0080 | 0.6310 | 0.2530 | 0.1030 | 0.3580 | 0.1640 | 0.3470 | 0.1557 | 0.3060 | 0.2420 |
| CoupledLP-IT | 0.0303 | 0.8249 | 0.4291 | 0.2483 | 0.5088 | 0.3484 | 0.5257 | 0.3240 | 0.4537 | 0.3855 |
| CoupledLP | 0.0249 | 0.8432 | 0.4305 | 0.2776 | 0.5481 | 0.3591 | 0.5420 | 0.3399 | 0.4692 | 0.4133 |

# Effects of Implicit Target Network

**AUPR**

| Method | $D$ to $G$ | $G$ to $D$ | $A_a$ to $A_b$ | $A_b$ to $A_a$ | $E_a$ to $E_b$ | $E_b$ to $E_a$ | $E_a$ to $E_c$ | $E_c$ to $E_a$ | $E_b$ to $E_c$ | $E_c$ to $E_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| CN | 0.0155 | 0.6011 | 0.3017 | 0.1348 | 0.3598 | 0.2319 | 0.3817 | 0.2079 | 0.3145 | 0.2654 |
| AA | 0.0167 | 0.5912 | 0.3344 | 0.1596 | 0.4541 | 0.2800 | 0.4838 | 0.2562 | 0.3802 | 0.3180 |
| JC | 0.0803 | 0.4812 | 0.0835 | 0.0903 | 0.3848 | 0.3082 | 0.4140 | 0.3429 | 0.3628 | 0.3579 |
| PA | 0.0083 | 0.7566 | 0.0820 | 0.0599 | 0.1446 | 0.1287 | 0.1525 | 0.1250 | 0.1560 | 0.1471 |
| PF | 0.0233 | 0.5501 | 0.1455 | 0.0989 | 0.3504 | 0.2248 | 0.3722 | 0.2138 | 0.2833 | 0.2446 |
| IT | 0.0155 | 0.6011 | 0.3715 | 0.2059 | 0.4344 | 0.3157 | 0.4568 | 0.2940 | 0.4008 | 0.3559 |
| LRC-IT | 0.0140 | 0.7830 | 0.3610 | 0.1880 | 0.4580 | 0.3140 | 0.5240 | 0.2870 | 0.4230 | 0.3500 |
| LRC | 0.0190 | 0.7930 | 0.3820 | 0.2030 | 0.4920 | 0.3160 | 0.5190 | 0.2910 | 0.4270 | 0.3590 |
| DT-IT | 0.0070 | 0.6270 | 0.2760 | 0.1050 | 0.3440 | 0.1620 | 0.3810 | 0.1550 | 0.2900 | 0.2260 |
| DT | 0.0080 | 0.6310 | 0.2530 | 0.1030 | 0.3580 | 0.1640 | 0.3470 | 0.1557 | 0.3060 | 0.2420 |
| CoupledLP-IT | 0.0303 | 0.8249 | 0.4291 | 0.2483 | 0.5088 | 0.3484 | 0.5257 | 0.3240 | 0.4537 | 0.3855 |
| CoupledLP | 0.0249 | 0.8432 | 0.4305 | 0.2776 | 0.5481 | 0.3591 | 0.5420 | 0.3399 | 0.4692 | 0.4133 |

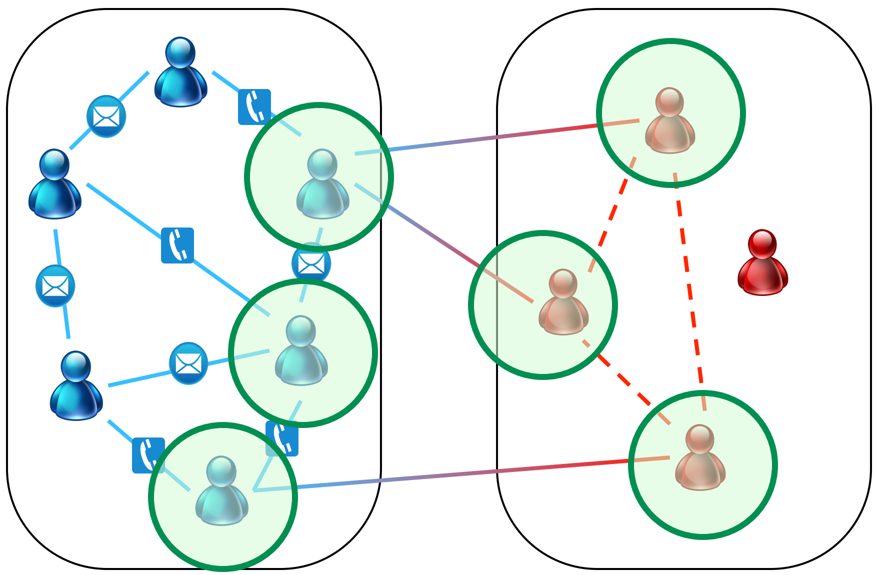
# Future Work

1. Efficiency of CoupledLP

2. One-step prediction framework

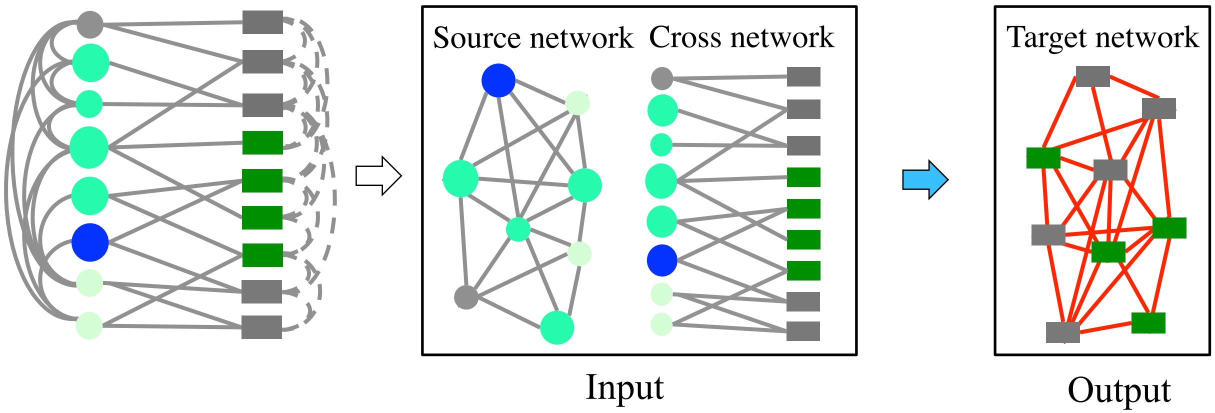
3. User behavior in coupled networks

…… ……



# Conclusion

❖ Coupled Link Prediction Problem




❖ CoupledLP Framework

**Implicit Target Network Construction**

- Solve Incompleteness

**Coupled Factor Graph Model**

- Solve Asymmetry
- Solve Heterogeneity

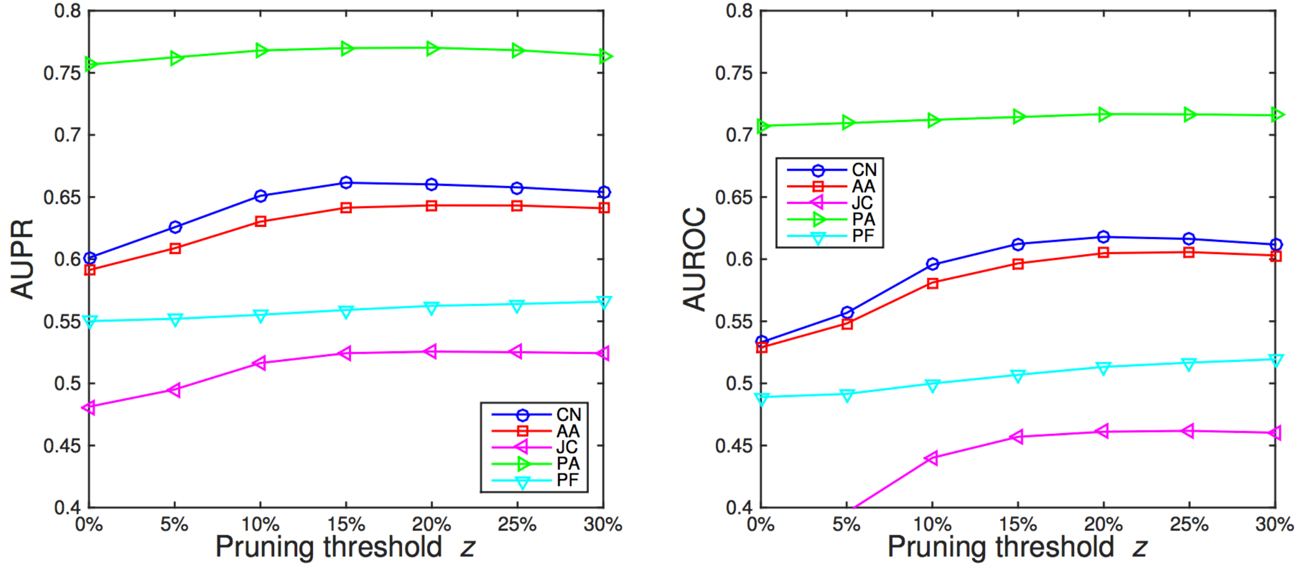
# Questions

**Thank You!**

Data & Code:
https://aminer.org/coupledlp

# Effects of Implicit Target Network

+5%
AUPR




(a) AUPR (b) AUROC

+8%
AUROC

*x-axis: pruning threshold z*
*y-axis: AUPR / AUROC*

Unsupervised methods on implicit target network