CoupledLP: Link Prediction in Coupled Networks

Yuxiao Dong — University of Notre Dame — ydong1@nd.edu

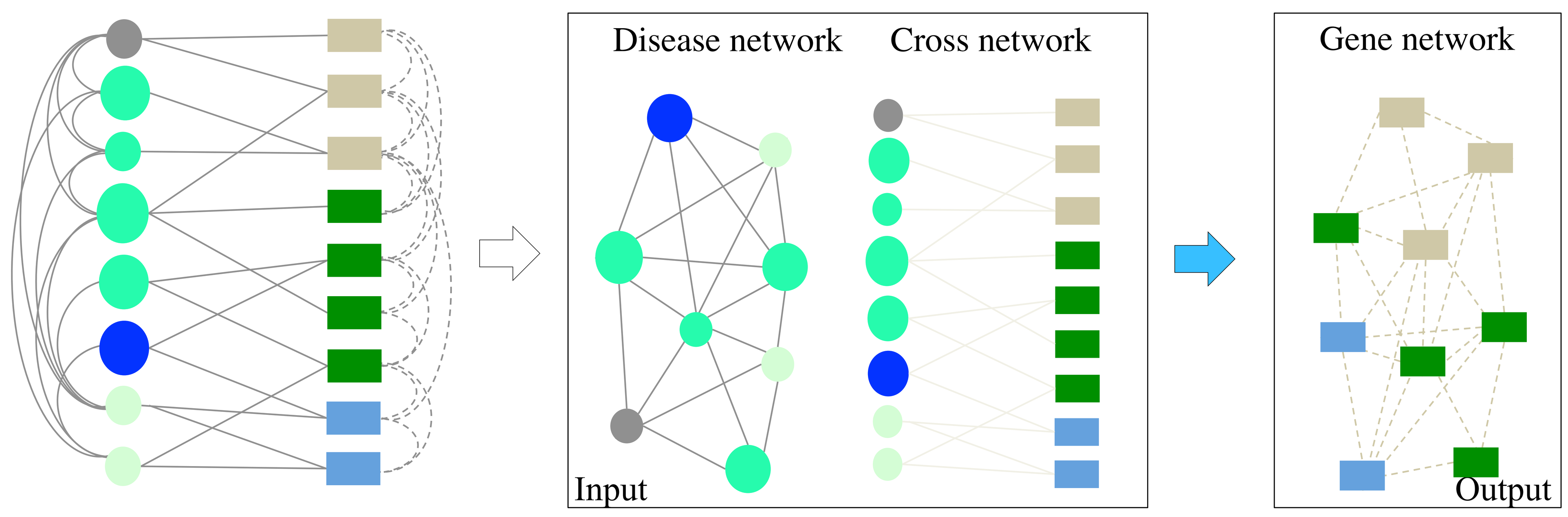
Jing Zhang — Tsinghua University — zhangjing12@tsinghua.edu.cn

Jie Tang — Tsinghua University — jietang@tsinghua.edu.cn

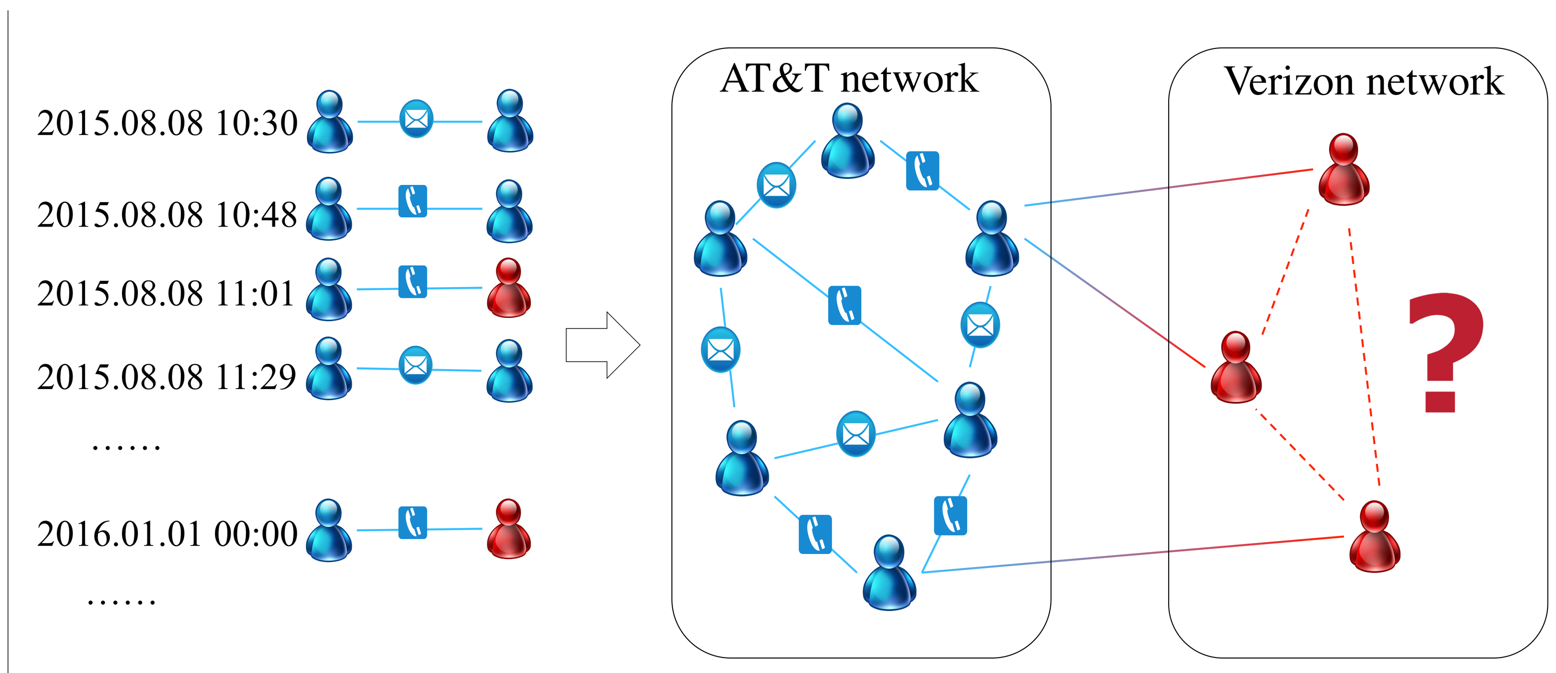
Nitesh V. Chawla — University of Notre Dame — nchawla@nd.edu

Bai Wang — Beijing U. of Posts and Telecoms — wangbai@bupt.edu.cn

Motivating Examples



Coupled Disease-Gene Networks



Coupled Mobile Social Networks

Problem Definition

❖ **Coupled Networks**

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

❖ **Coupled Link Prediction**

Given the source network G^S and the cross network G^C in coupled networks $G = (G^S, G^T, G^C)$, the task is to find a predictive function: $f : (G^S, G^C) \rightarrow Y^T$,
where Y^T is the set of labels for the potential links in the target network G^T, with $y_{ij} = 1$ indicating a link exists between v_i and v_j, and $y_{ij} = 0$ indicating no link exists between them.

❖ **Challenges**

Incompleteness. We do not have structure information between two users in target network—there is a visibility of links that go from source network to target network but not beyond that.
Heterogeneity. The source and target networks with multi-typed objects are twisted and coupled with one another. This makes it difficult to directly use a supervised learning approach.
Asymmetry. Following the heterogeneity, the two coupled networks usually present different network properties—such as the average degree or clustering coefficient.

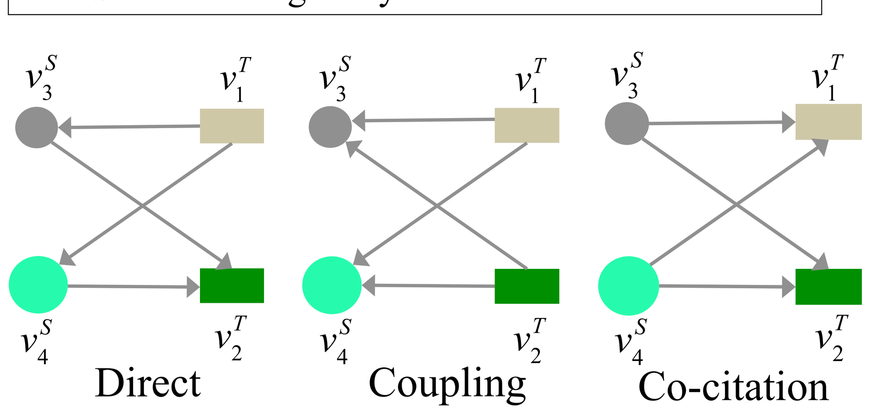
❖ **Related Work**

Problem	Transfer Link Prediction	Cross-Domain Link Prediction	Heterogeneous Link Prediction	Coupled Link Prediction
Input	disease network + (part of gene network)	disease network + gene network	part of coupled networks	disease network + cross network + (part of gene network)
Output	gene network	remaining links in cross network	remaining single- or multi-typed links	links in gene network

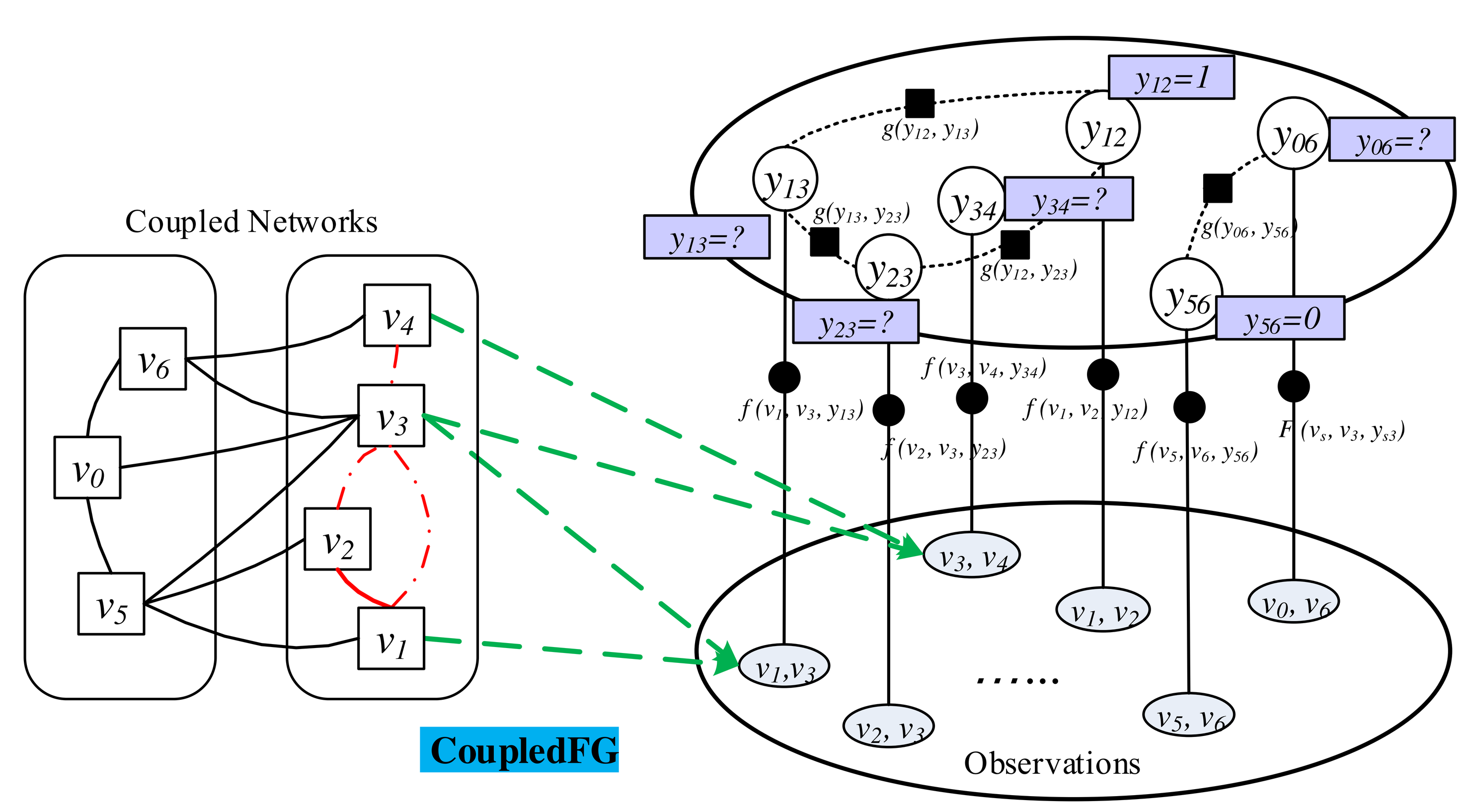
CoupledLP Framework

Implicit Target Network Construction
- Solve Incompleteness

Coupled Factor Graph Model
- Solve Asymmetry
- Solve Heterogeneity

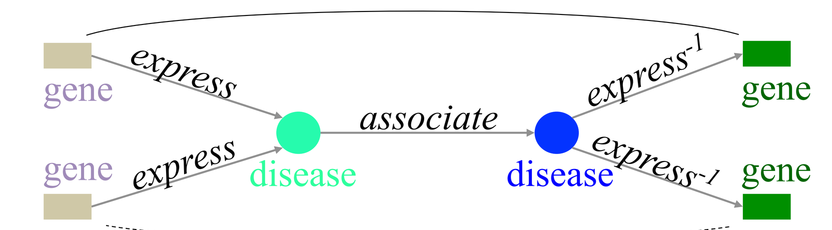


Implicit Target Network Construction



Coupled Factor Graph

❖ **Coupled Meta-Path**



❖ **Joint Distribution**

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

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

model source & target networks separately — meta-path

❖ **Objective Function**

$$\mathcal{O}(\theta) = \sum_{e \in E^S} \Big(\sum_{k=1}^{K} \alpha_k f_k(x^S_{ek}, y^S_e)\Big) + \sum_{e \in E^T} \Big(\sum_{k=1}^{K} \beta_k g_k(x^T_{ek}, y^T_e)\Big) + \sum_{\pi \in \Pi} \gamma_\pi h_\pi(Y_\pi) - \log Z$$

Experiments

Coupled Network Data

❖ Disease−Gene Networks (D, G)
❖ Asian Mobile Networks (A)
❖ European Mobile Networks (E)

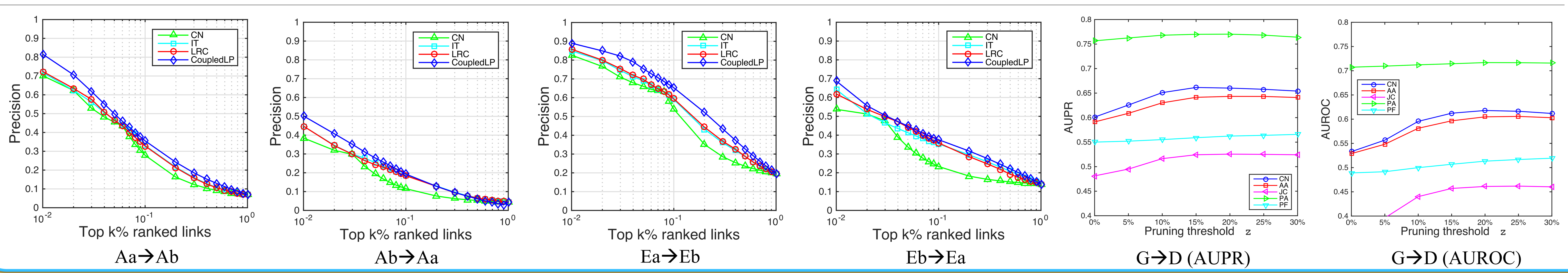
attributes	D	G	D ↔ G	Aa	Ab	Aa↔Ab	Ea	Eb	Ec	Ea↔Eb	Ea↔Ec	Eb↔Ec
#nodes	703	1,132	1835	348,640	63,687	235,715	2,531,187	655,755	354,166	1,912,933	1,255,046	625,379
#edges	74,523	2,450	10483	613,614	96,325	306,213	3,355,197	649,322	311,432	1,844,342	1,131,593	507,894
average degree	212.01	4.33	11.43	3.52	3.02	2.59	2.65	1.98	1.75	1.92	1.80	1.62
clustering coefficient	0.2639	0.0377	0	0.0237	0.0225	0	0.0457	0.0366	0.0317	0	0	0
associative coefficient	-0.0256	0.1761	-0.2556	0.2011	0.1671	0.0654	0.2848	0.2693	0.2806	0.0231	-0.0305	0.1113

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

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



Aa→Ab | Ab→Aa | Ea→Eb | Eb→Ea | G→D (AUPR) | G→D (AUROC)

Reference

1. D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In **CIKM'03.**
2. K.I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. In **PNAS 2007.**
3. Y. Dong, J. Tang, S. Wu, J. Tian, N. V. Chawla. J. Rao, H. Cao. Link Prediction and Recommendation across Heterogeneous Networks. In **ICDM'12**
4. Y. Sun, J. Han, C. C. Aggarwal, N. V. Chawla. Will Will This Happen? Relationship Prediction in Heterogeneous Information Networks. In **WSDM'12.**
5. J. Tang, T. Lou, J. Kleinberg. Inferring social ties across heterogeneous networks. In **WSDM'12.**

ID: 417

Yuxiao Dong, Jing Zhang, Jie Tang, Nitesh V. Chawla, Bai Wang. CoupledLP: Link Prediction in Coupled Networks. In ACM SIGKDD'15.

iCeNSA | KEG

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