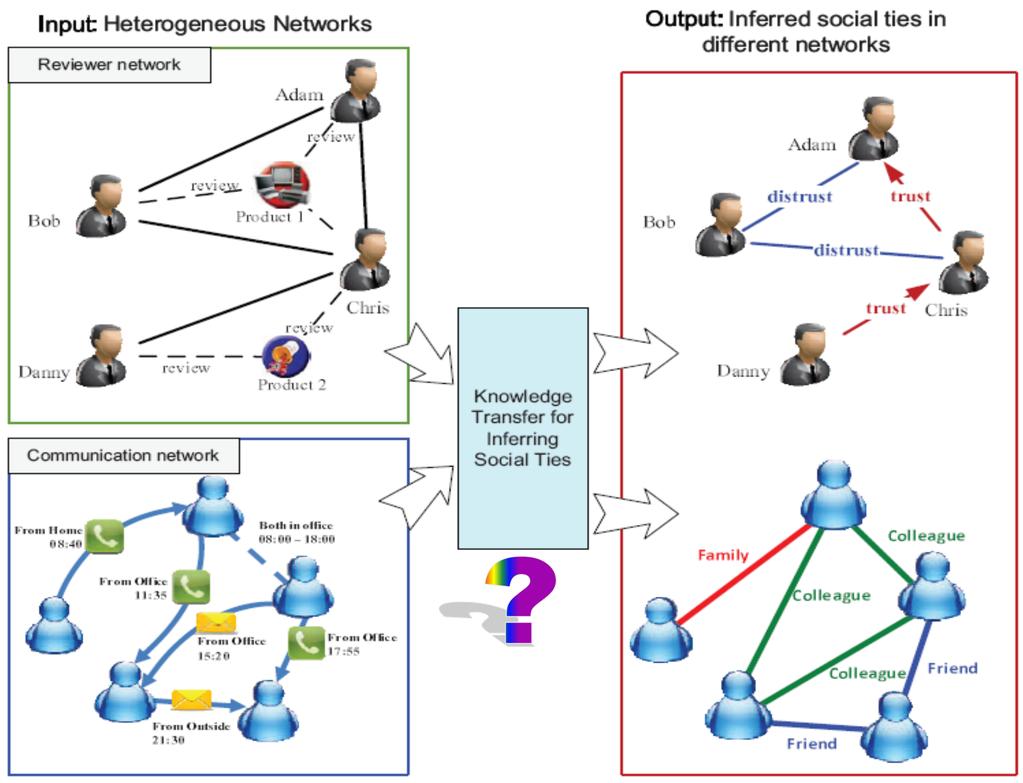
Inferring Social Ties across Heterogeneous Networks

Jie Tang*, Tiancheng Lou*, Jon Kleinberg+

*Tsinghua University +Cornell University

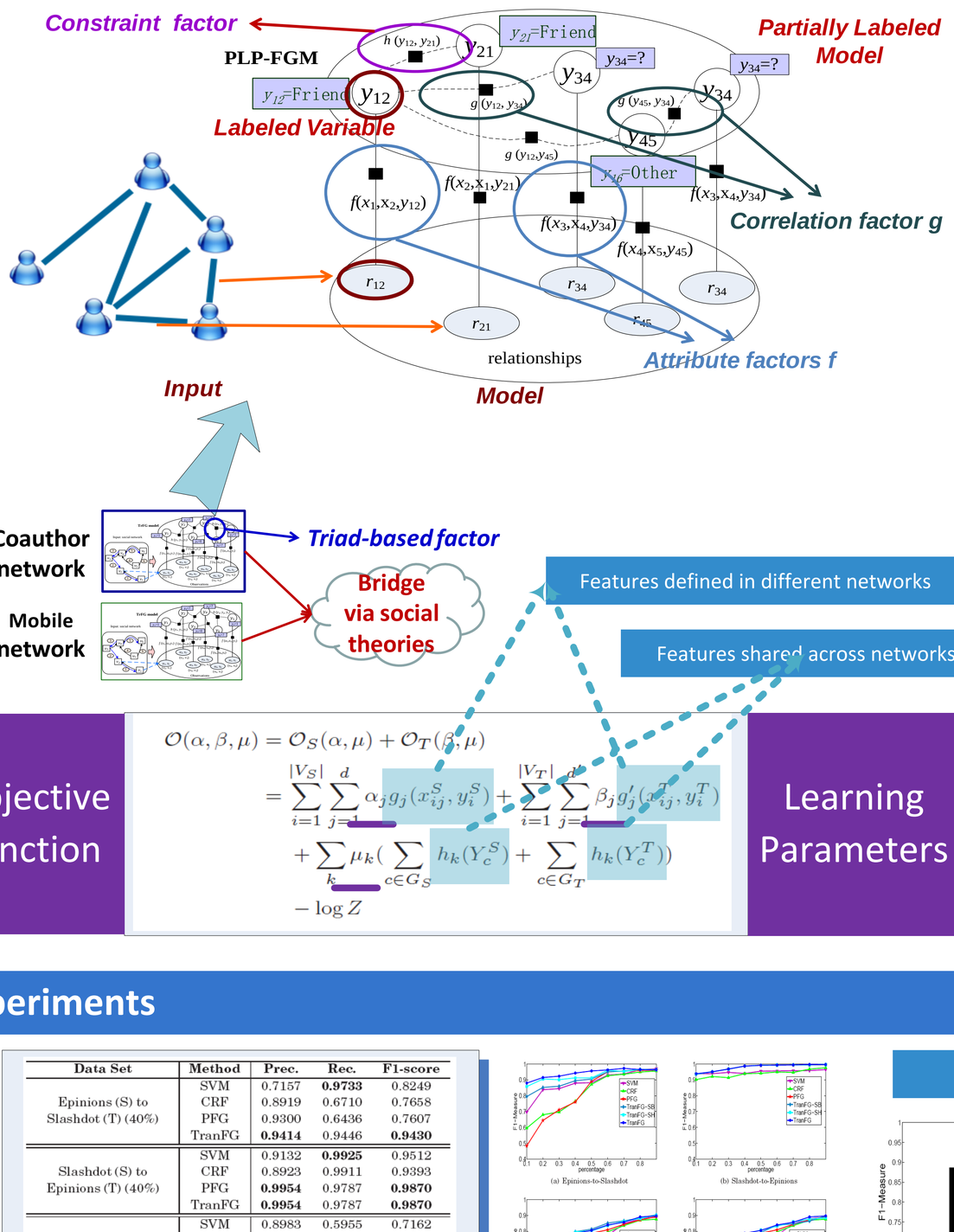
Motivation



Problem Definition

INPUT: a source networks G_S with abundantly labeled relationships and a target network G_T with limited labeled relationships.
OUTPUT: a predictive function f : (G_T | G_S) -> Y_T for inferring the type of relationships.

Modeling : Transfer Factor Graph Model



Objective function

$$\mathcal{O}(\alpha,\beta,\mu) = \mathcal{O}_S(\alpha,\mu) + \mathcal{O}_T(\beta,\mu)$$
$$= \sum_{i=1}^{|V_S|}\sum_{j=1}^{d}\alpha_j g_j(x_{ij}^S, y_i^S) + \sum_{i=1}^{|V_T|}\sum_{j=1}^{d'}\beta_j g'_j(x_{ij}^T, y_i^T)$$
$$+ \sum_k \mu_k\Big(\sum_{c\in G_S} h_k(Y_c^S) + \sum_{c\in G_T} h_k(Y_c^T)\Big)$$
$$- \log Z$$

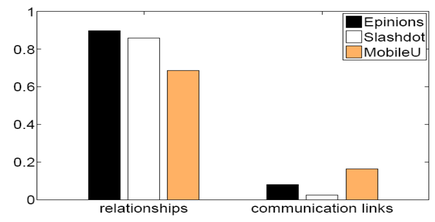
Learning Parameters

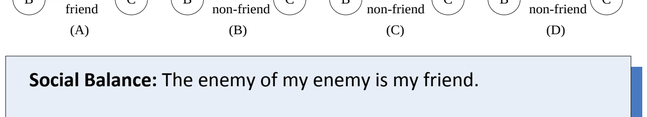
Bridge: Social Theories

Data Collection

Relationship	Dataset	#Nodes	#Edges
Trust	Epinions	131,828	841,372
Friendship	Slashdot	77,357	516,575
Friendship	MobileU	107	5,436
Manager-subordinate	MobileD	232	3,567
Advisor-advisee	Coauthor	815,946	2,792,833
Manager-subordinate	Enron	151	3,572

Social Balance

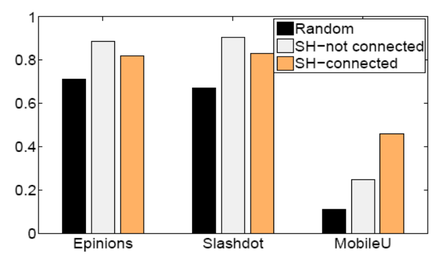


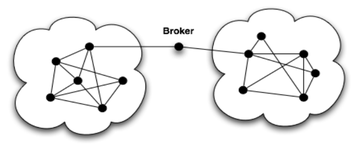


Social Balance: The enemy of my enemy is my friend.

Observation: Different networks have very different structural balance probabilities. While based on friendships three networks have a relatively similar probabilities.

Structural Hole

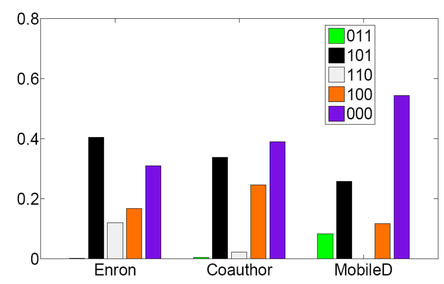


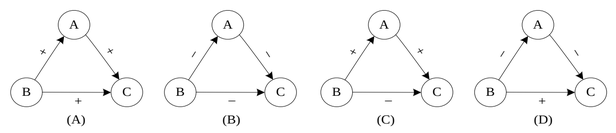


Structural Hole: A person is said to span a structural hole if she connects people who are otherwise not well connected to one another.

Observation: Users are more likely (+70%) to have the same type of relationship with structural hole nodes.

Social Status

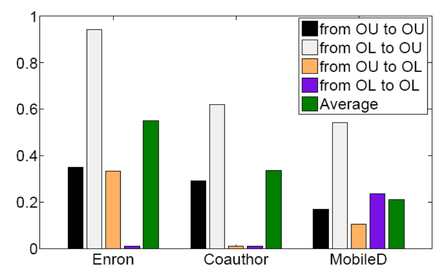


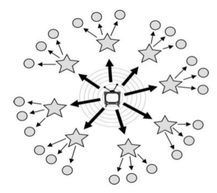


Social Status: In a triangle of three nodes, We take each negative edge, reverse its direction, and flip its sign to positive, then the resulting triangle should be acyclic.

Observation: Nearly 99% of Triads in The three networks satisfy the social status theory, and they share a similar distribution on the five frequent forms of triads.

Two Step Flow

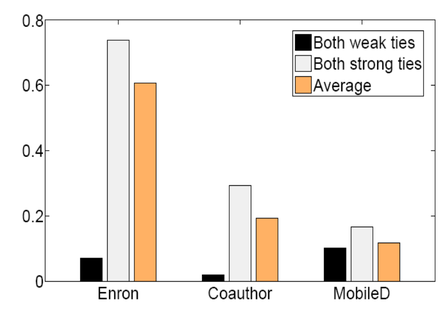


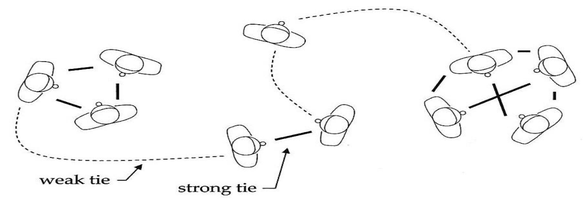


Two Step Flow: Ideas (innovations) usually flow first to opinion leaders, and then from them to a wider population.

Observation: Opinion leaders are more likely to have a higher social-status than ordinary users.

Strong/Weak Tie





Strong/Weak Tie: Strong tie implies that one's close friends, and a user may have the similar types of relationships with friends of strong ties and more diverse relationships with friends of weak ties.

Observation: Two strong ties result in a higher likelihood to share the same type, while two weak ties are much more uncertain.

Experiments

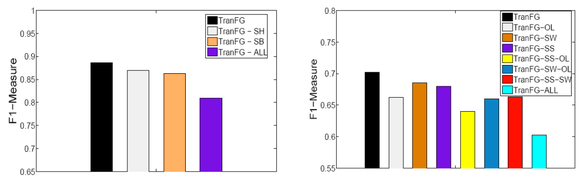
Friendship

Data Set	Method	Prec.	Rec.	F1-score
Epinions (S) to Slashdot (T) (40%)	SVM	0.7157	**0.9733**	0.8249
	CRF	0.8919	0.6710	0.7658
	PFG	0.9300	0.6436	0.7607
	TranFG	**0.9414**	0.9446	**0.9430**
Slashdot (S) to Epinions (T) (40%)	SVM	0.9132	**0.9925**	0.9512
	CRF	0.8923	0.9911	0.9393
	PFG	**0.9954**	0.9787	**0.9870**
	TranFG	**0.9954**	0.9787	**0.9870**
Epinions (S) to MobileU (T) (40%)	SVM	0.8983	0.5955	0.7162
	CRF	0.9455	0.5417	0.6887
	PFG	**1.0000**	0.5924	0.7440
	TranFG	0.8239	**0.8344**	**0.8291**
Slashdot (S) to MobileU (T) (40%)	SVM	0.8983	0.5955	0.7162
	CRF	0.9455	0.5417	0.6887
	PFG	**1.0000**	0.5924	0.7440
	TranFG	0.7258	**0.8599**	**0.7872**

Directed

Data Set	Method	Prec.	Rec.	F1-score
Enron (T) (40%)	SVM	**0.9524**	0.5556	0.7018
	CRF	0.7778	0.7673	0.7725
	PFG	0.9130	0.7241	0.8077
MobileD (S) to Enron (T)	TranFG (M)	0.8438	0.7941	0.8182
Coauthor (S) to Enron (T)	TranFG (C)	0.9091	**0.8824**	**0.8955**
Coauthor (T) (40%)	SVM	0.6910	0.3727	0.4842
	CRF	**0.8472**	0.2937	0.4362
	PFG	0.8189	0.3377	0.4782
	TPFG	0.5936	**0.7611**	0.6669
MobileD (S) to Coauthor (T)	TranFG (M)	0.8235	0.3889	0.5283
Enron (S) to Coauthor (T)	TranFG (E)	0.8193	0.6415	**0.7196**
MobileD (T) (40%)	SVM	0.5249	0.3725	0.4358
	CRF	0.4454	**0.5763**	0.5025
	PFG	**0.8739**	0.3731	0.5229
Enron (S) to MobileD (T)	TranFG (E)	0.8013	0.4808	0.6010
Coauthor (S) to MobileD (T)	TranFG (C)	0.8323	0.5154	**0.6366**

Factor Contribution



Homogeneous Networks

Data Set	Method	Prec.	Rec.	F1-score
Slashdot (S) to Slashdot (T) (40%)	PFG	0.9300	0.6436	0.7607
	TranFG	0.9948	0.9185	0.9551
	TranFG-Heter	0.9414	0.9446	0.9430
Epinions (S) to Epinions (T) (40%)	PFG	0.9954	0.9787	0.9870
	TranFG	0.9954	1.0000	0.9977
	TranFG-Heter	0.9954	0.9787	0.9870
MobileU (S) to MobileU (T) (40%)	PFG	1.0000	0.5924	0.7440
	TranFG	0.9259	0.7895	0.8523
	TranFG-Heter	0.8239	0.8344	0.8291
Enron (S) to Enron (T) (40%)	PFG	0.9130	0.7241	0.8077
	TranFG	0.9394	0.9688	0.9538
	TranFG-Heter	0.9091	0.8824	0.8955
Coauthor (S) to Coauthor (T) (40%)	PFG	0.8189	0.3377	0.4782
	TranFG	0.8321	0.7433	0.7852
	TranFG-Heter	0.8193	0.6415	0.7196
MobileD (S) to MobileD (T) (40%)	PFG	0.8739	0.3731	0.5229
	TranFG	0.8843	0.6115	0.7230
	TranFG-Heter	0.8323	0.5154	0.6366

Case Study