@IC^2S^2

Inferring Social Status and Rich Club Effects in Enterprise Communication Networks

Yuxiao Dong, Jie Tang, Nitesh V. Chawla
Tiancheng Lou, Yang Yang, Bai Wang

UNIVERSITY OF NOTRE DAME

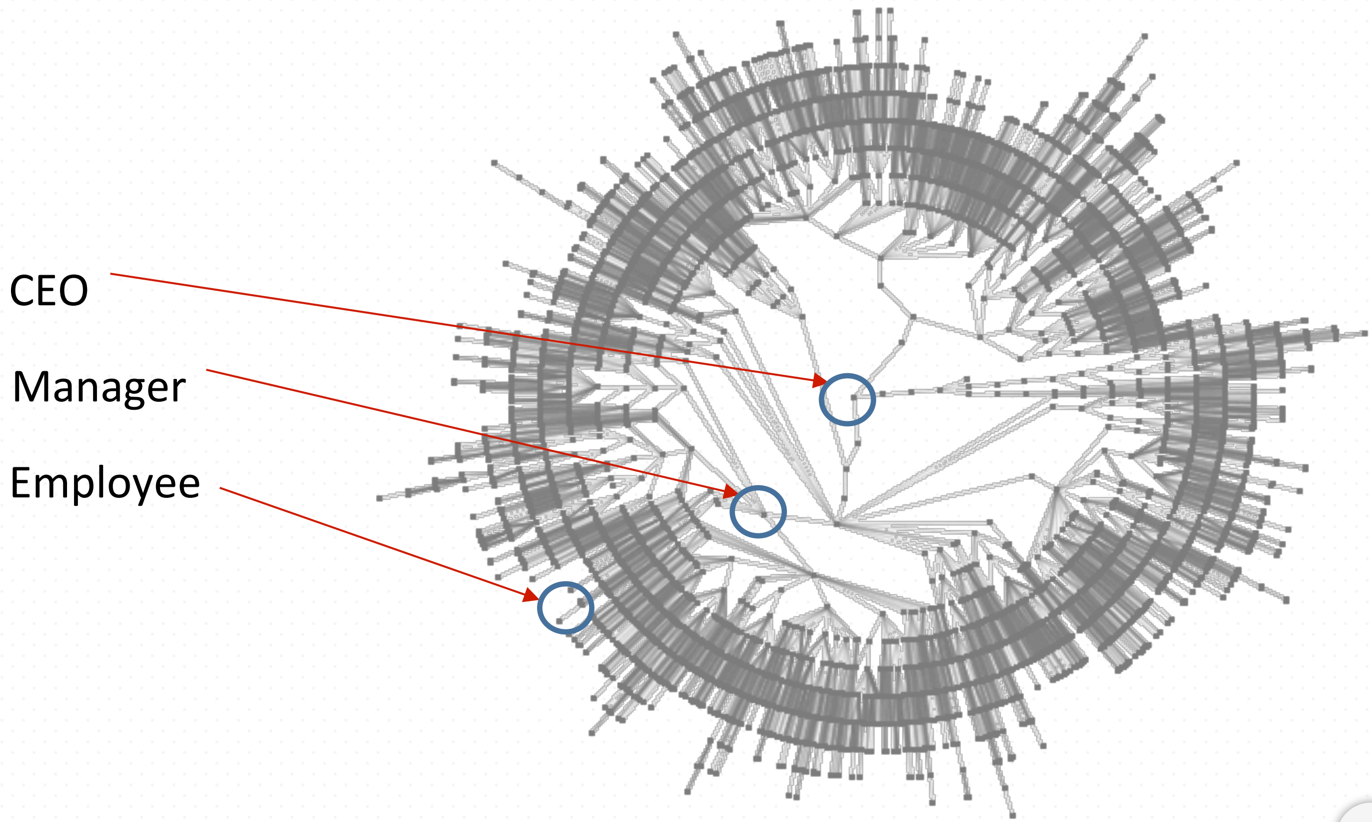
iCeNSA Interdisciplinary Center for Network Science & Applications

清華大學 Tsinghua University

Yuxiao Dong, Jie Tang, Nitesh V. Chawla, Tiancheng Lou, Yang Yang, Bai Wang.
Inferring Social Status and Rich Club Effects in Enterprise Communication Networks. PLOS ONE 2015.

Social Status

- Social status is defined as the relative rank or position that an individual holds in a social hierarchy.



Enterprise Communication Data

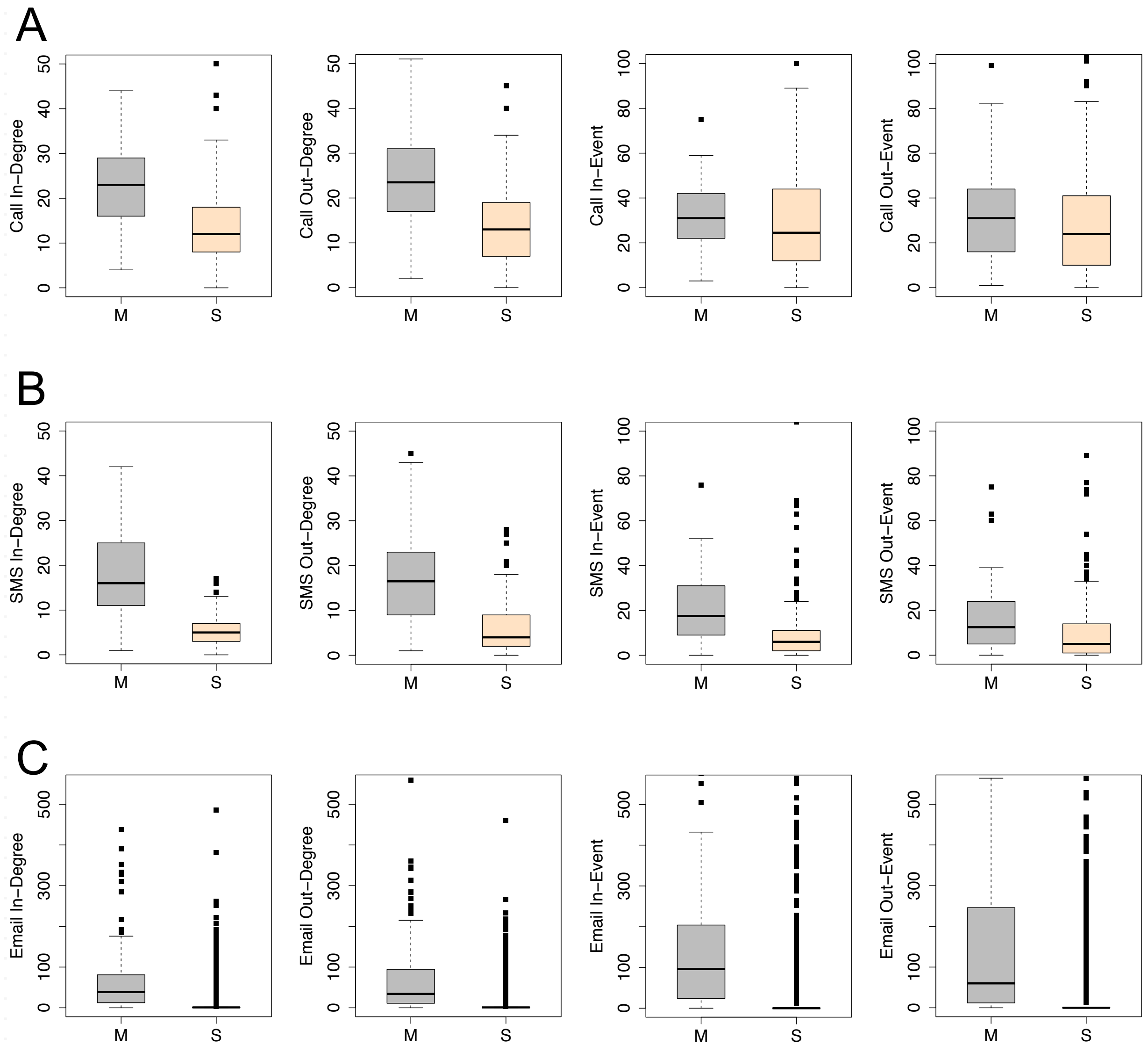
- Two Companies and Three Communication Channels
  - One Telecommunication Company
    - Phone call network (CALL)
    - Short-message network (SMS)
    - 50 managers and 182 subordinates
  - Enron Inc.
    - Email communication network (EMAIL)
    - 155 managers and 22232 subordinates

attributes	CALL	SMS	EMAIL
#nodes	232	232	22477
#edges	3340	3406	44728
clustering coefficient	0.3326	0.4761	0.1241
associative coefficient	0.1195	-0.0894	-0.2153

Problems

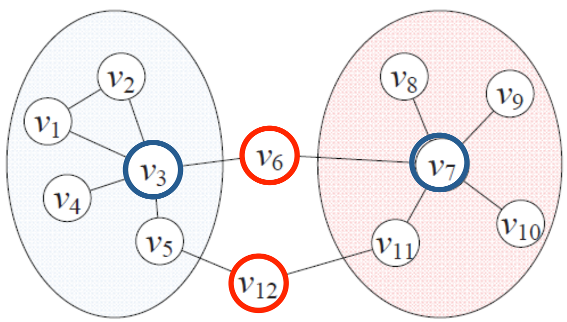
What is the interplay of social status with communication behavior?

High-status has high centralities?



M: Manager
S: Subordinate

Structural Hole vs. Status



- Structural holes serve as intermediaries between others who are not directly connected [1].
- We use HIS algorithm [2] to estimate the likelihood of each node in the network to span as structural hole

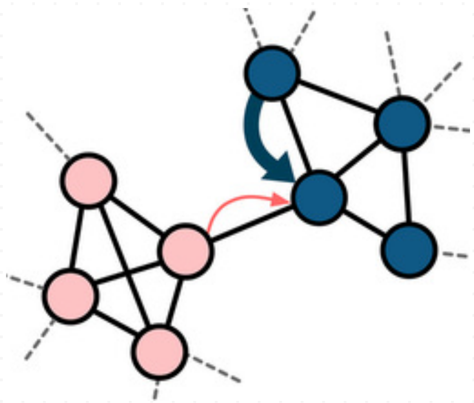
	CALL	SMS	EMAIL
M as SH	0.700 ****	0.550 ****	0.430 ****
M as SH (random)	0.207	0.207	0.007
S as SH	0.300 ****	0.450 ****	0.570 ****
S as SH (random)	0.793	0.793	0.993

*$p < 0.05$;
**$p < 0.01$;
***$p < 0.001$;
****$p < 0.0001$

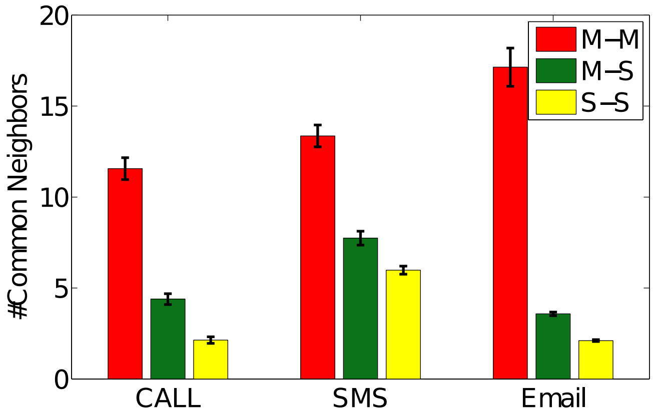
- Structural holes extracted from enterprise communication network structure reveal the social status of staff in the company.
- Managers usually need to operate the responsibility of correspondents and organizers within the company, especially for the experience for connecting different groups to cooperate.

[1] R. S. Burt. Structural Holes: The Social Structure of Competition. Harvard University Press, 1992.
[2] Tiancheng Lou, Jie Tang. Mining Structural Hole Spanners in Information Diffusion. WWW 2013.

Link Homophily vs. Status



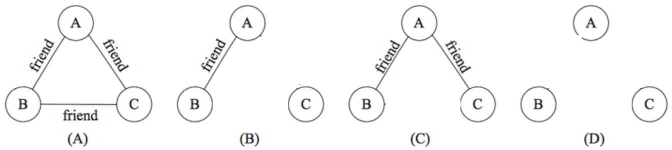
- Homophily is the tendency of individuals to associate and bond with similar others.
- Link homophily [15] tests whether two individuals who share more common neighbors will tend to have similar social status in the company.



- Two individuals are much more likely to be two managers in the company if they share more common contacts.
- Managers' ability of creating and maintaining social connections in enterprise networks is more prominent than subordinates'.

Lilian Weng, Fillippo Menczer, Yong-Yeol Ann. Virality Prediction and Community Structure in Social Network. Scientific Reports. Aug. 2013

Social Balance vs. Status



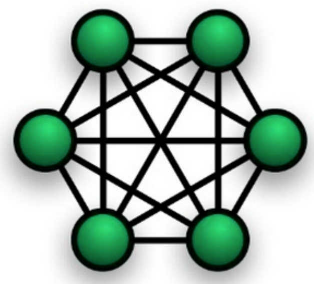
- A social triangle satisfies social balance theory, if all three users are friends or only one pair of them are friends.
- Assume two users are friends if they communicate each other at least once.

	M (CALL)	S (CALL)	M (SMS)	S (SMS)	M (EMAIL)	S (EMAIL)
M-sb	0.569****	0.348****	0.546****	0.468****	0.455****	0.047****
S-sb	0.174***	0.254*	0.289	0.299	0.066**	0.082****
sb	0.340	0.312	0.325	0.311	0.165	0.124

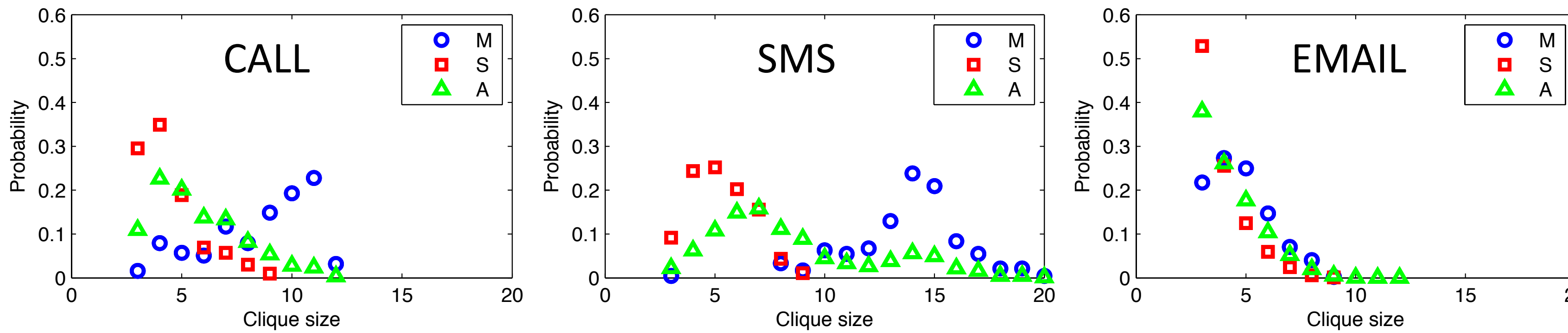
*$p < 0.05$;
**$p < 0.01$;
***$p < 0.001$;
****$p < 0.0001$

- Managers' overall balance ratios are larger than the subordinates' across all the three channels.
- Managers are more likely to form balanced structure among their manager-friends, and the subordinates with subordinates.

Social Clique vs. Status



- A clique is a subset of nodes such that for any two nodes, there exists an edge connecting them.
- In social sciences, clique is used to describe a group of persons who interact with each other more regularly and intensely than others.



- The size of the maximum clique in 50-manager CALL and SMS networks and 155-manager EMAIL network is 12, 20, and 9, respectively.
- The size of the maximum clique in 182-subordinate CALL and SMS networks and 22232-subordinate EMAIL network is 9, 10, and 9, respectively.

Social Status Prediction

Can social status be inferred from communications?

Social Status Prediction

Input:
$G = (V, E, Y^L), \mathbf{X}$

Output:
$f(G, \mathbf{X}) \rightarrow (Y^U)$

- *V:* node set
- *E:* edge set
- **X**: attribute matrix
- Y^L: nodes with labeled social status
- Y^U: nodes with unlabeled social status

Social Status Prediction

Network		Precision	Recall	F1	Accuracy
CALL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
SMS	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
EMAIL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				

Data:

CALL network
SMS network
EMAIL network

10% as training data
90% as test data

Social Status Prediction

Network		**Precision**	**Recall**	**F1**	**Accuracy**
CALL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
SMS	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
EMAIL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				

Computational Models:

Naïve Bayes
Bayes Network
Logistic Regression
Conditional Random Fields

Factor Graph Model

Social Status Prediction

Network		Precision	Recall	F1	Accuracy
CALL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
SMS	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
EMAIL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				

Evaluation Metrics:

Precision
Recall
F1 score
Accuracy

Social Status Prediction

Network		Precision	Recall	F1	Accuracy
CALL	*Naïve Bayes*	0.7334	0.7625	0.7416	0.7625
	Bayes Network	0.7409	0.6934	0.7110	0.6934
	Logistic Regression	0.7065	0.6795	0.6904	0.6795
	Conditional Random Fields	0.8078	0.8095	0.8086	0.8095
	Factor Graph Model	0.8514	0.8508	0.8511	0.8508
SMS	*Naïve Bayes*	0.8693	0.8734	0.8648	0.8734
	Bayes Network	0.8497	0.8512	0.8483	0.8512
	Logistic Regression	0.8129	0.7850	0.7935	0.7850
	Conditional Random Fields	0.8720	0.8761	0.8740	0.8760
	Factor Graph Model	0.9321	0.9276	0.9298	0.9276
EMAIL	*Naïve Bayes*	0.8847	0.8993	0.8847	0.8598
	Bayes Network	0.8936	0.9054	0.8164	0.8755
	Logistic Regression	0.8761	0.8772	0.7653	0.8483
	Conditional Random Fields	0.9033	0.8902	0.8967	0.8902
	Factor Graph Model	0.9319	0.9383	0.9373	0.9383

Social Status Prediction

Network		Precision	Recall	F1	Accuracy
CALL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
SMS	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				
EMAIL	*Naïve Bayes*				
	Bayes Network				
	Logistic Regression				
	Conditional Random Fields				
	Factor Graph Model				

Predictability:

85% of individuals’ Social Status can be revealed from mobile phone call network, and over 92% in both messaging and email networks.

Conclusion

- Unveil the social behavioral differences of individuals with different social status across three communication channels.
  - High-status individuals are more likely to be spanned as structural holes.
  - The principle of homophily, social balance and clique theory generally indicate a "rich club" maintained by high-status individuals.
- Propose ***computational models*** to demonstrate the predictability of social status in communication networks.
  - Over 85% of individuals' status can be revealed from call networks.
  - Over 92% in both mobile messaging and email networks.

Thanks

Q & A

Yuxiao Dong, Jie Tang, Nitesh V. Chawla, Tiancheng Lou, Yang Yang, Bai Wang.
Inferring Social Status and Rich Club Effects in Enterprise Communication Networks.
PLoS ONE 10 (3): e0119446. 2015