Quantifying Reciprocity in Large Weighted Communication Networks

PAKDD 2012


Leman Akoglu


Pedro Vaz de Melo

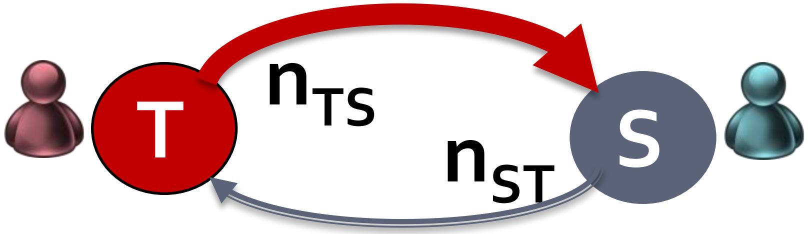

Christos Faloutsos

Carnegie Mellon University, USA

Universidade Federal de Minas Gerais , Brazil

Questions we answer:

- If T calls S n_{TS} times, what can we say about how many times S calls T?



- How can we quantify reciprocity between T and S?

- Does reciprocity depend on local topology? –e.g. degree similarity?

Why study reciprocity?

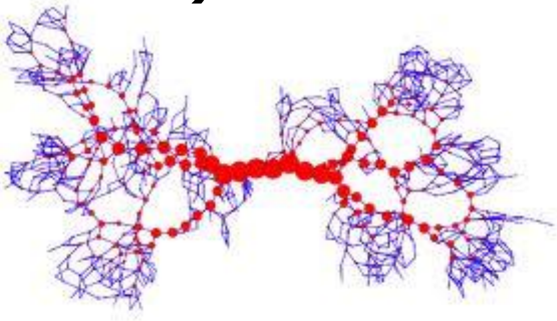
Reciprocity-based features are good for:

- trust prediction [Nguyen+ 2010]

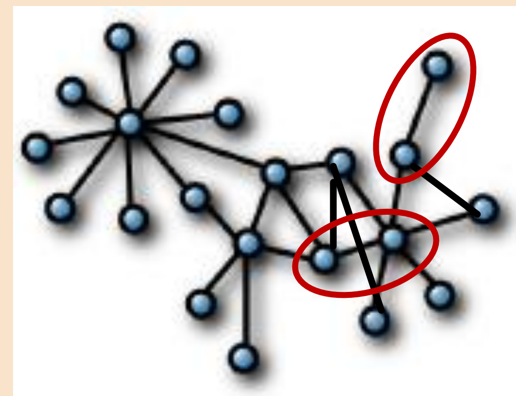
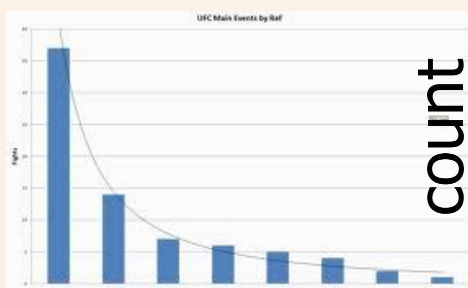
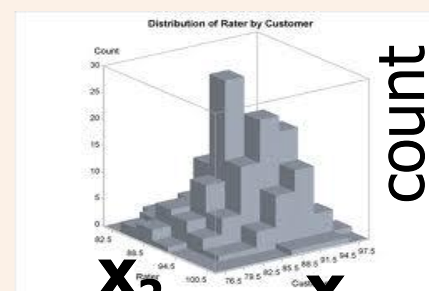
- spam detection

- network engagement/churn

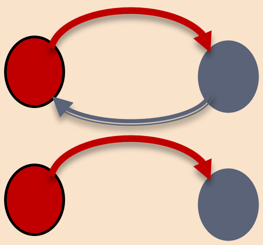
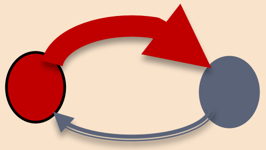
- propagation (rumor, ideas, viruses)
- link-persistence
  [Cesar-Hidalgo 2008]



Contributions –part 1

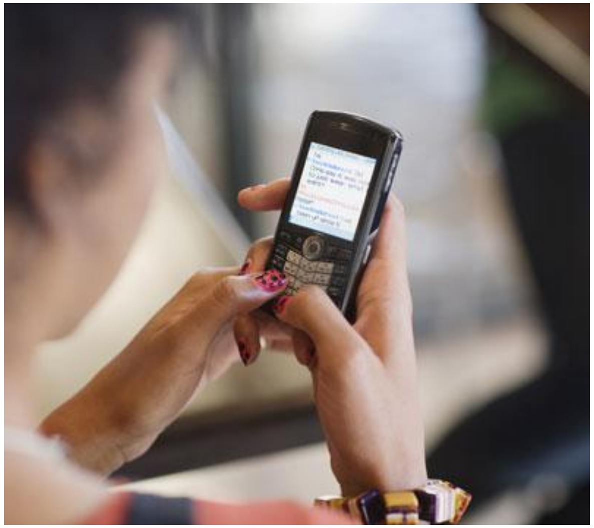
Previous Work	This paper
Node/triadic topology • degree dist., centrality, network value, influence • clustering coef., triangle closures, communities	Dyadic relations 
Univariate: Pr(x)  count x	Bivariate: $\Pr(x_1, x_2)$  count x_2 x_1

Contributions –part 2

Previous Work	This paper
Unweighted 	Weighted 
Global $r = \frac{L^{\leftrightarrow}}{L}$	Local $\frac{2\sqrt{w_{ij}w_{ji}}}{(w_{ij}+w_{ji})}$?
r=1 e.g. collaborations r=0 e.g. citations	Relations to local network topology

Data

- Phone call and SMS networks
  - ~2 million customers
  - ½ billion phone calls
  - 60 million SMS interactions
  - Dec. 2007 to May 2008



- edge-weights:
  - #SMS , #Calls , Duration

Data statistics

Network	N	E	W_N	$W_D(min)$
CALL	1,87M	49,50M	483,7M	$915\text{x}10^6$
CALL(mutual)	1,75M	41,84M	468,7M	$885\text{x}10^6$

Network	N	E	W_{SMS}
SMS	1,87M	8,80M	60,5M
SMS(mutual)	0,58M	2,10M	46,6M

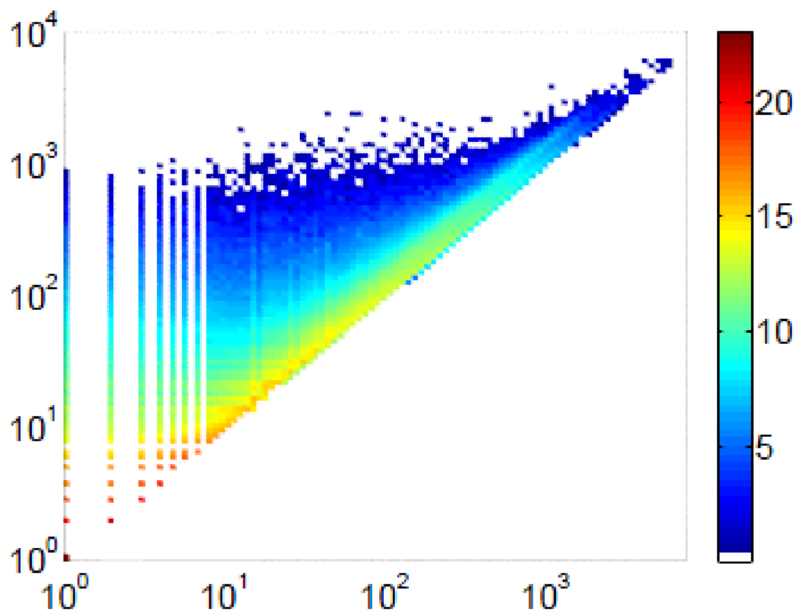
- CALL: r=0.84 SMS: r=0.24
- SMS-mutual shrinks (total weight ~remains)

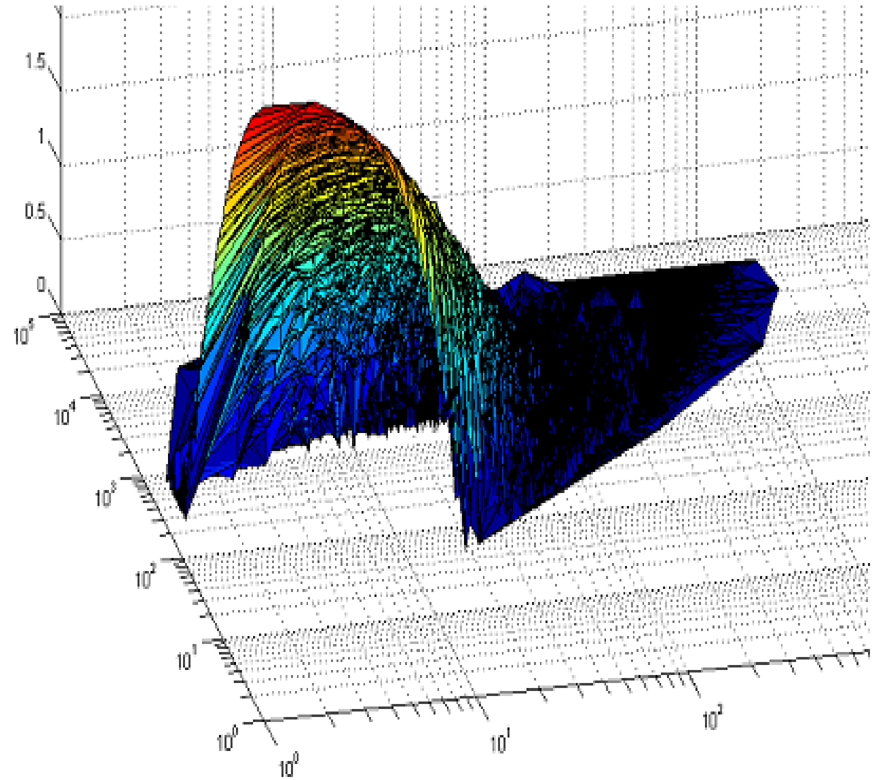
1) Bivariate Pattern of Reciprocity

- **Observed patterns**
- Our 3PL Model
- Competing Models
- Goodness of fit
- 3PL at work: anomalies

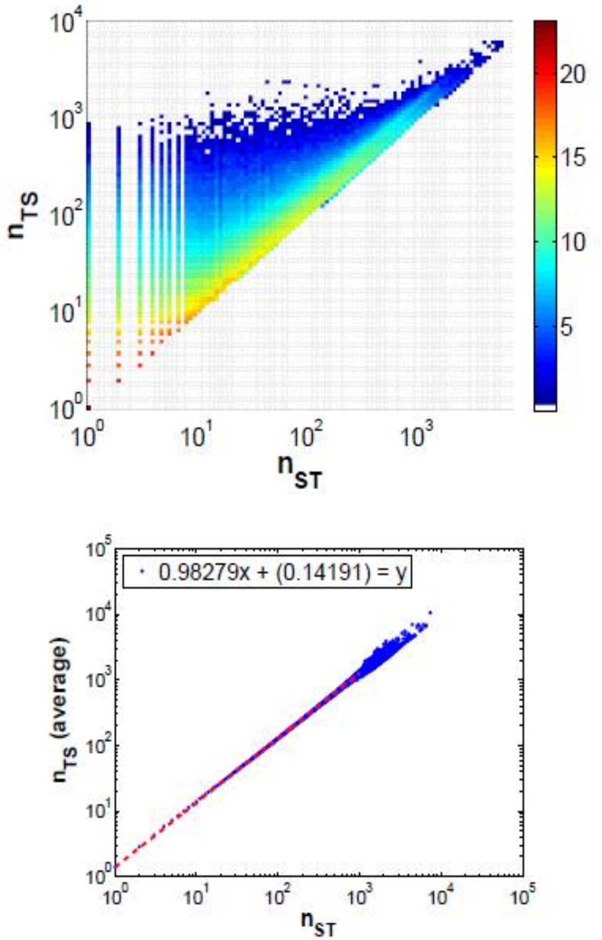
Observed bivariate patterns

- How can we model Prob(n_{ST}, n_{TS})? : Bivariate! 2D- Gaussian? Pareto? Yule? …

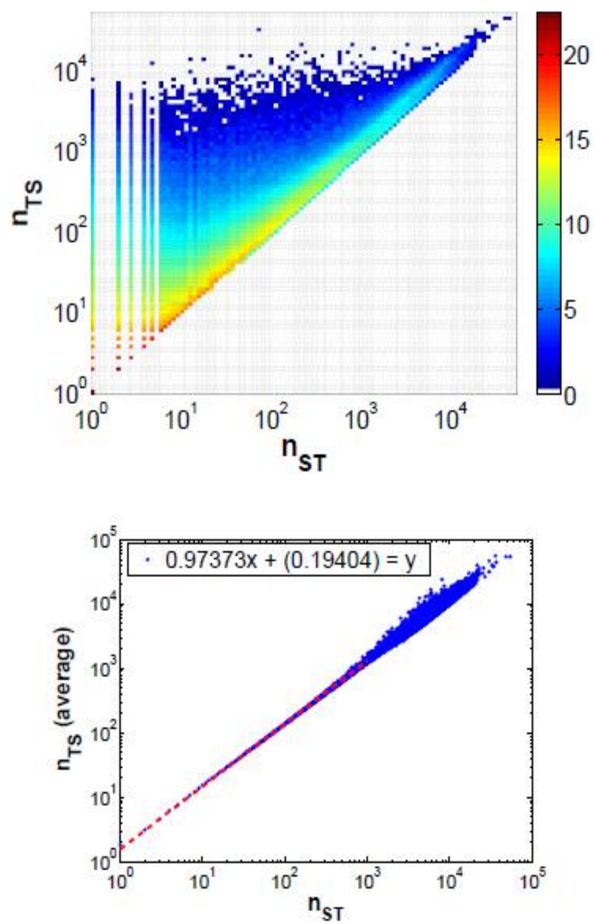




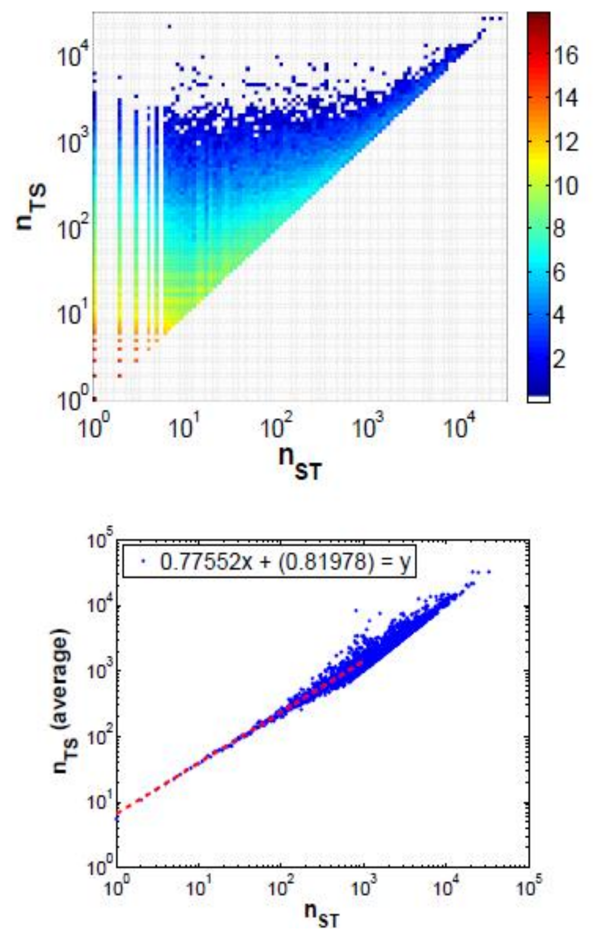
Observed bivariate patterns



CALL-N



CALL-D



SMS

1) Bivariate Pattern of Reciprocity

- Observed patterns
- Our 3PL Model
- Competing Models
- Goodness of fit
- 3PL at work: anomalies

Our 3PL Model

In human communication networks, distribution $\Pr(n_{ST}, n_{TS})$ of mutual edge weights follows a Triple Power Law (3PL)

$$\Pr(n_{ST}, n_{TS}) \propto \boxed{\frac{n_{ST}^{-\alpha} n_{TS}^{-\beta} (n_{TS} - n_{ST} + 1)^{-\gamma}}{Z(\alpha, \beta, \gamma)}}$$

$\alpha > 0, \beta > 0$ capture `rich-get-richer'

$\gamma > 0$ captures skewness in asymmetry

1) Bivariate Pattern of Reciprocity

- Observed patterns
- Our 3PL Model
- Competing Models
- Goodness of fit
- 3PL at work: anomalies

Competing Models

- Competitor 1: Bivariate Pareto

$$f_{X_1,X_2}(x_1,x_2) = k(k+1)(ab)^{k+1}(ax_1 + bx_2 + ab)^{-k-2}$$

$$a, b, k > 0$$

- Competitor 2: Bivariate Yule

$$f_{X_1,X_2}(x_1,x_2) = \frac{\rho_{(2)}(x_1+x_2)!}{(\rho+1)_{(x_1+x_2+2)}} \qquad \rho > 0$$

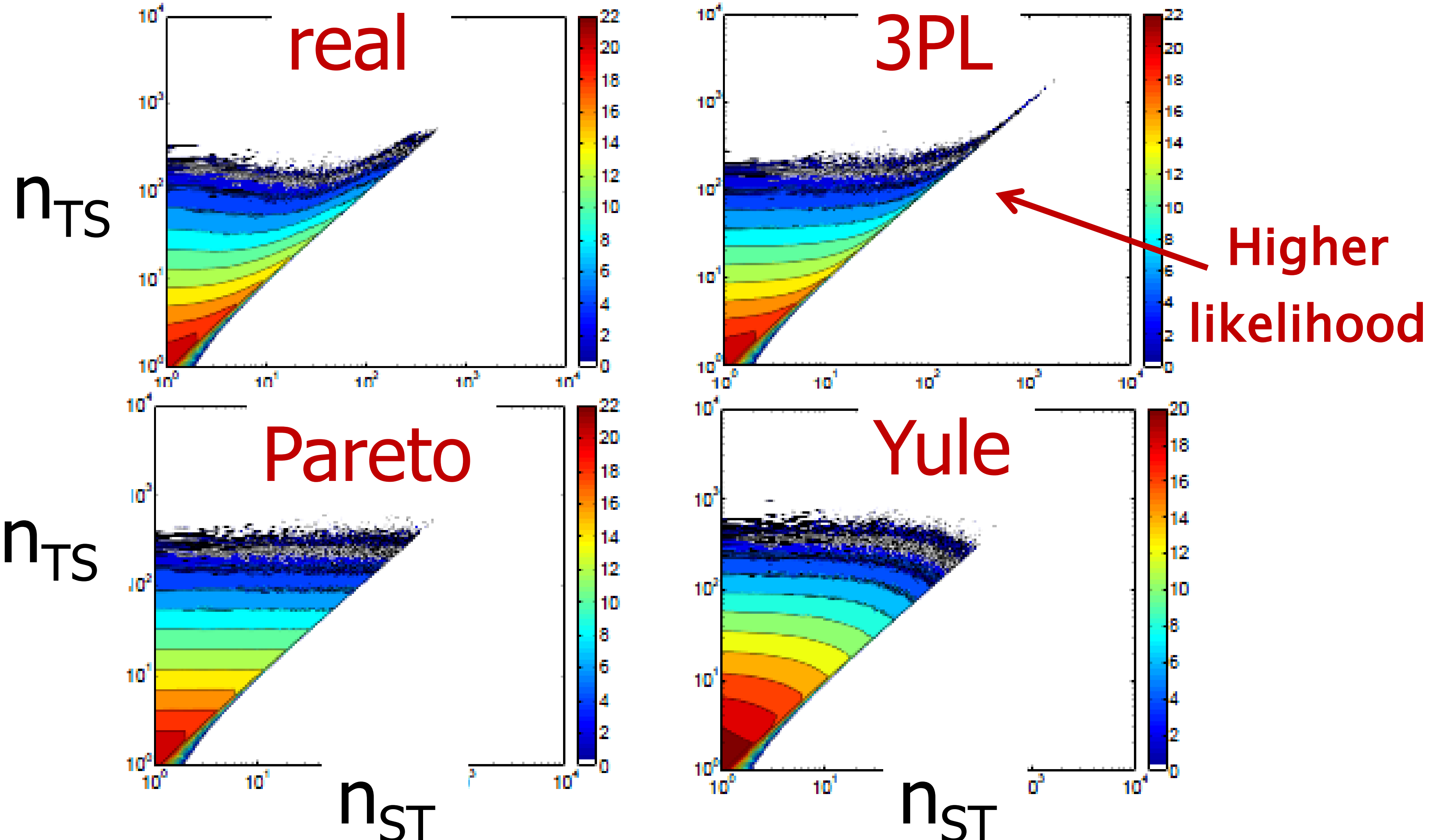
3PL vs Competing Models

$z = \mathcal{R}/\sqrt{2n}\sigma$

normalized log-likelihood ratio

	CALL-N	CALL-D	SMS
Triple Power Law (3PL)			
α	1e-06	1e-06	0.8120
β	2.0703	1.8670	1.5896
γ	0.8204	0.9650	0.3005
Loglikelihood	-7.55e+07	-8.88e+07	-5.41e+06
Bivariate Pareto			
k	0.7407	0.7657	0.7862
a	0.2119	0.5723	0.7097
b	10e+05	1.25e+04	0.7553
Loglikelihood	-7.77e+07	-9.26e+07	-5.39e+06
z	803.73	975.75	-41.06
p	0	0	0
Bivariate Yule			
ρ	1.11e-16	5.55e-17	1e-06
Loglikelihood	-8.59e+07	-10.00e+07	-5.41e+06
z	2.14e+03	1.93e+03	1.49
p	0	0	0.03

3PL vs Competing Models



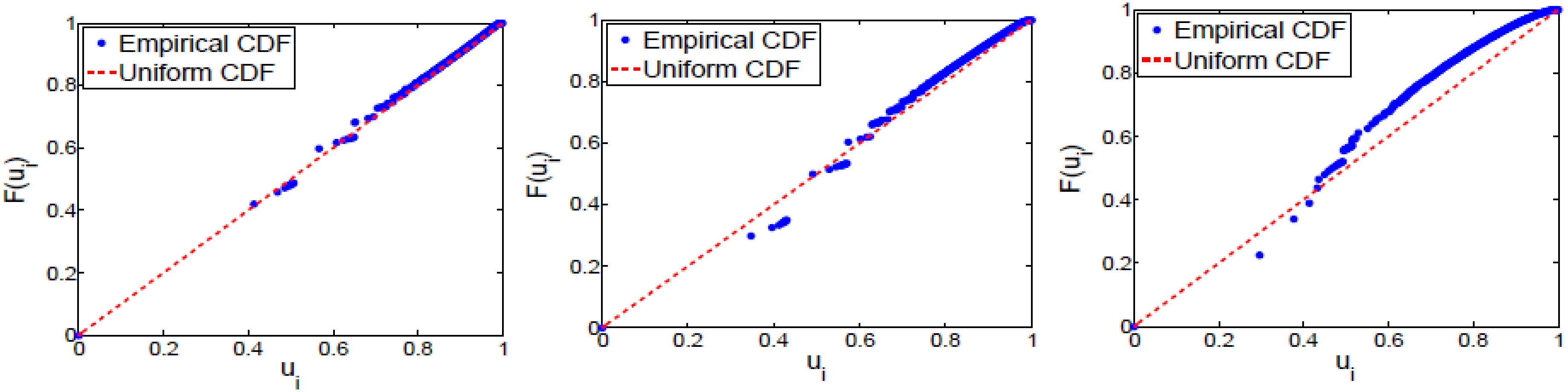
1) Bivariate Pattern of Reciprocity

- Observed patterns
- Our 3PL Model
- Competing Models
- Goodness of fit
- 3PL at work: anomalies

3PL Goodness of fit

$$\hat{u}_i = \hat{F}(x_{1,i}, x_{2,i})$$

If F is correct CDF, $\hat{u}_i$ is uniformly distributed.

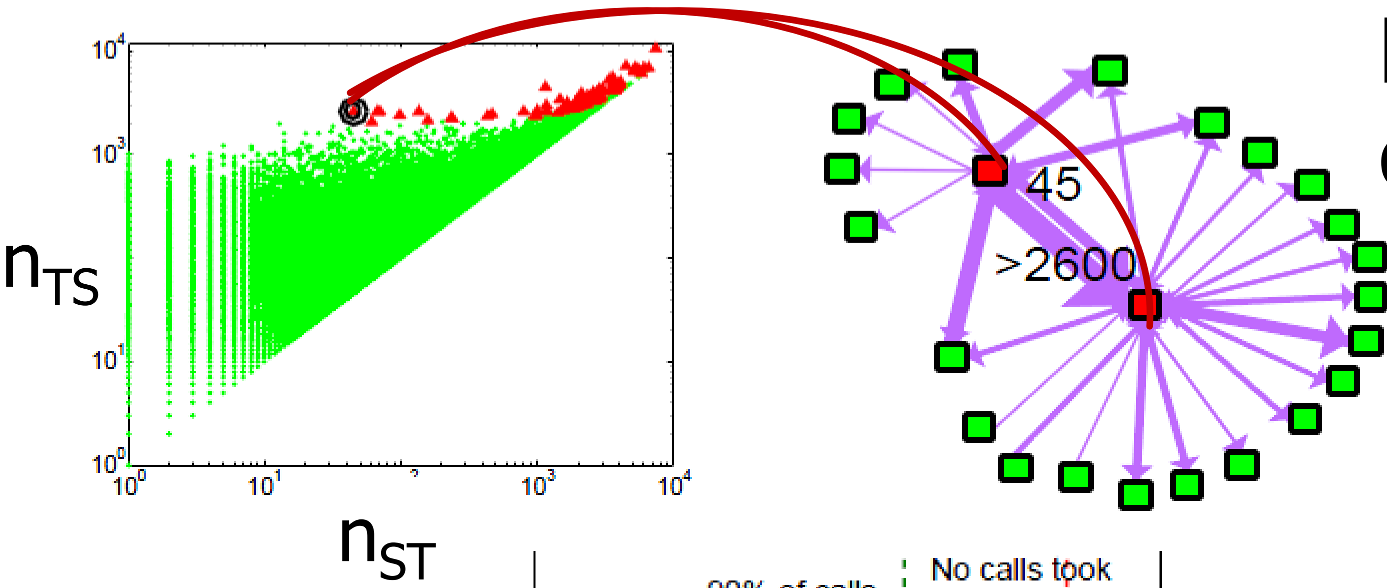


CALL-N CALL-D SMS

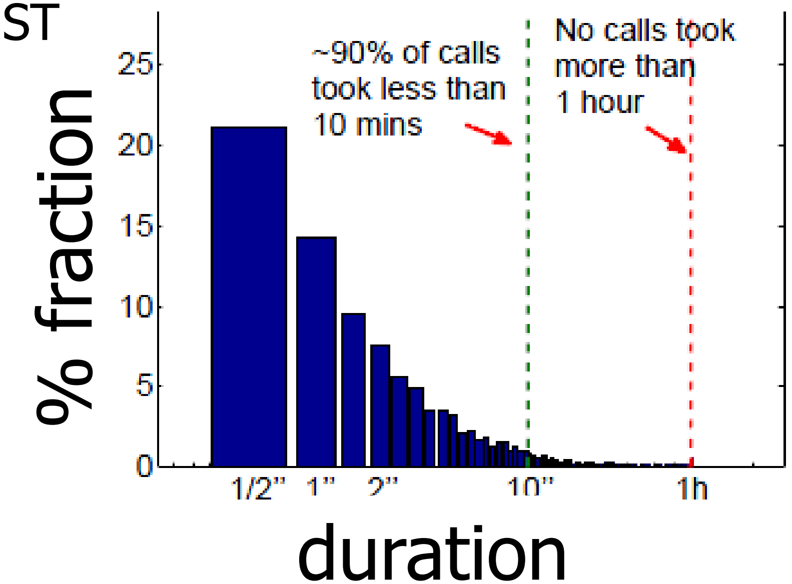
1) Bivariate Pattern of Reciprocity

- Observed patterns
- Our 3PL Model
- Competing Models
- Goodness of fit
- 3PL at work: anomalies

3PL at work: Anomalies



least likely dyads in red



2) Reciprocity and Local Network Topology

- Weighted metrics
- Local network overlap
- Assortativity

Weighted metrics

1) balance factor

0: non-mutual
1: fully mutual

$$Ratio\ r = \frac{\min(w_{ij}, w_{ji})}{\max(w_{ij}, w_{ji})} \in [0, 1]$$

$$Coherence\ c = \frac{2\sqrt{w_{ij} w_{ji}}}{(w_{ij} + w_{ji})} \in [0, 1]$$

$$Entropy\ e = -p_{ij} \log_2(p_{ij}) - p_{ji} \log_2(p_{ji}) \in [0, 1]$$

2) volume factor

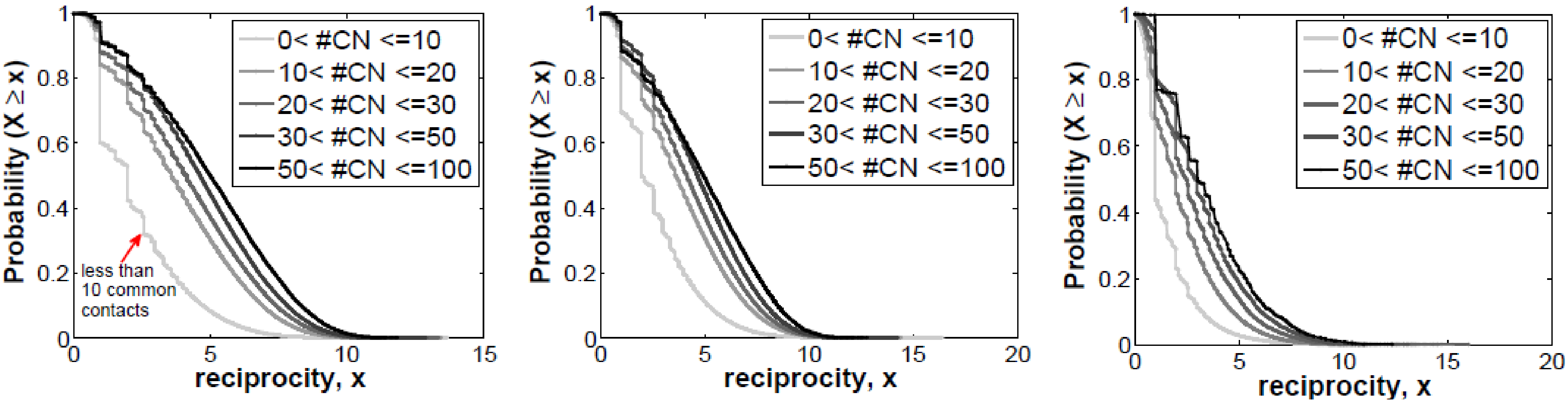
$$r_w = \frac{\min(w_{ij}, w_{ji})}{\max(w_{ij}, w_{ji})} \log(w_{ij} + w_{ji})$$

2) Reciprocity and Local Network Topology

- Weighted metrics
- Local network overlap
- Assortativity

Reciprocity and network overlap

Is there a relation between reciprocity and local network overlap (= #common neigh.s)?



CALL-N CALL-D SMS

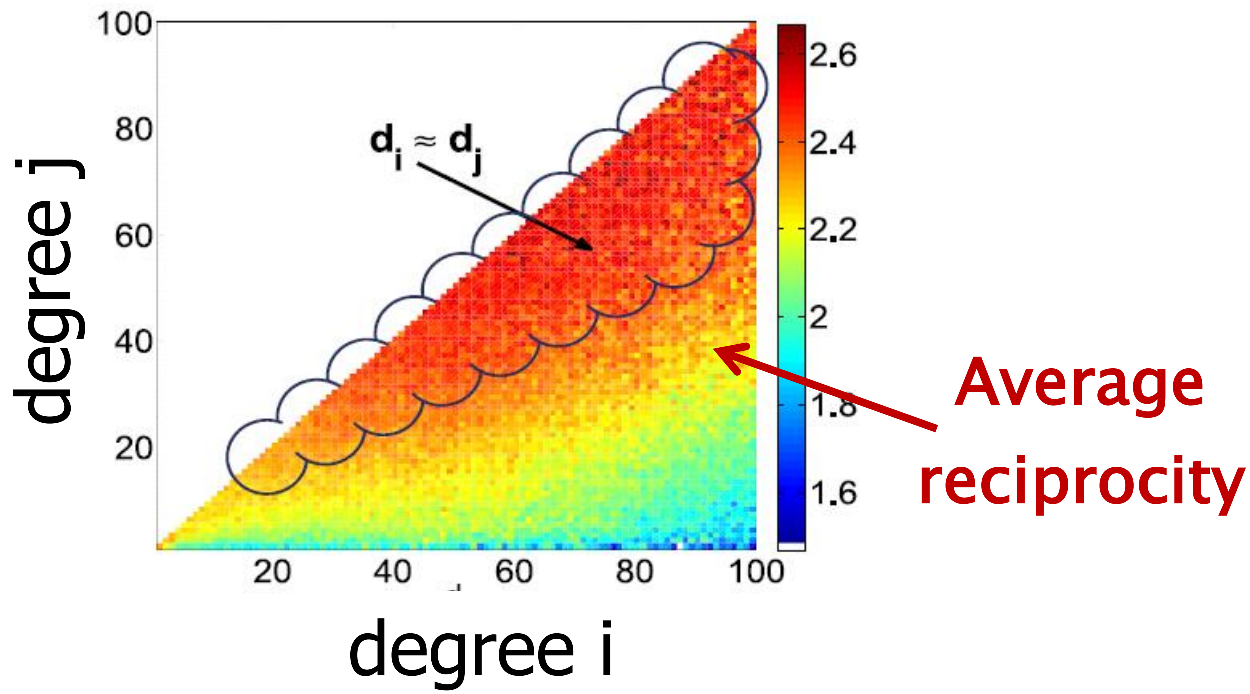
Larger network overlap → Higher reciprocity (i.e. more common friends)

2) Reciprocity and Local Network Topology

- Weighted metrics
- Local network overlap
- Assortativity

Reciprocity and assortativity

Is there a relation between reciprocity and degree assortativity (=similarity)?

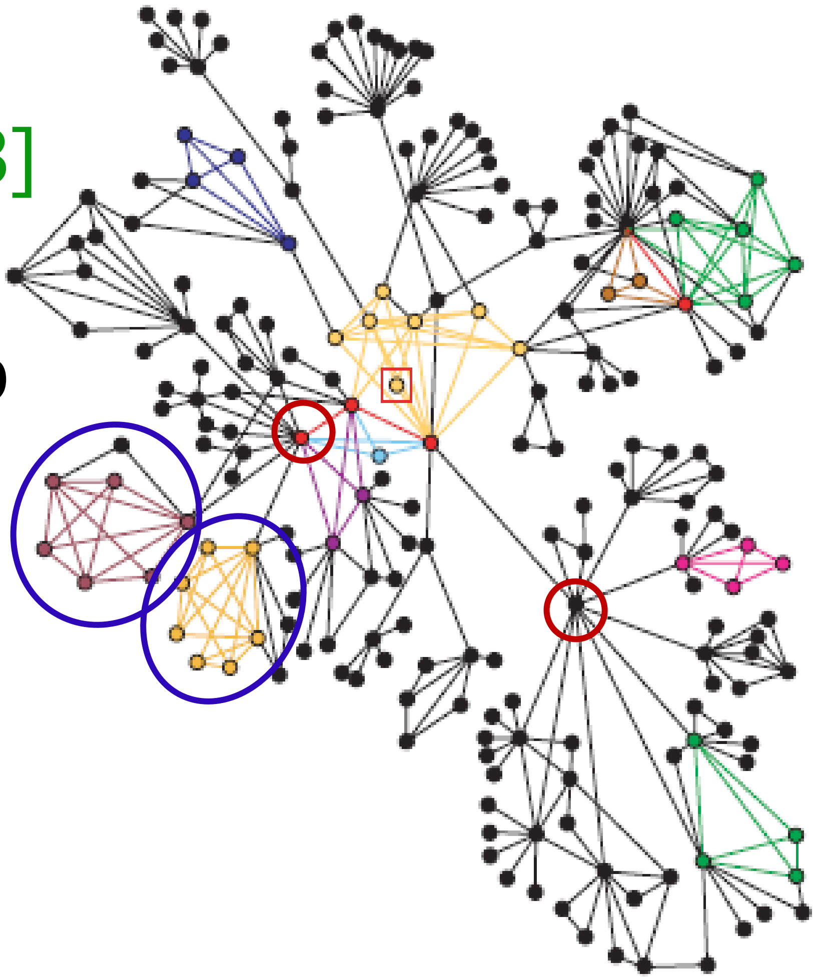


Larger degree similarity → Higher reciprocity (i.e. similar #contacts)

Discussion

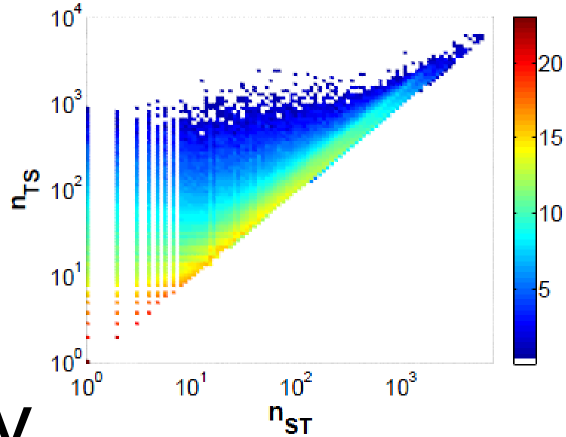
Our findings conform to:

- Clusters [Watts-Strogatz'98]
  - High degree similarity
  - High local network overlap
  - Reciprocity expected
- Hubs [Barabasi+'99]
  - Low degree similarity
  - Low local network overlap
  - Reciprocity not expected



Summary of contributions

- Patterns in dyad reciprocity
  - Mostly few/short calls & SMSs
  - Mostly reciprocal behavior



- New 3PL model for reciprocity
  - Good fit to >20M points
  - Better than competitors

$$\Pr(n_{ST}, n_{TS}) \propto \boxed{\frac{n_{ST}^{-\alpha} n_{TS}^{-\beta} (n_{TS} - n_{ST} + 1)^{-\gamma}}{Z(\alpha, \beta, \gamma)}}$$

- Study of local, weighted reciprocity
  - Higher for larger overlap and degree similarity

Future directions

- Better models for reciprocity
- Evolution of reciprocal behavior
- Caller prediction using reciprocal features
  - degree of reciprocity
  - inter-arrival time
  - avg. time passed since last call
  - #calls since last call
  - etc.

Thank you!

lakoglu@cs.cmu.edu
http://www.cs.cmu.edu/~lakoglu