

Online Social Networks - Connecting People Like Never Before

SIAM Annual Meeting 2008
July 8th, 2008

Sue Moon

Joint Work with

Hyunwoo Chun, Seungyeop Han, Haewoon Kwak,
Young-Ho Eom, Yong-Yeol Ahn, Hawoong Jeong



Motivation

- So much of our “social” interaction captured online
- Information (bad or good) flows thru our social network
- Online social network = information pipeline

Cyworld

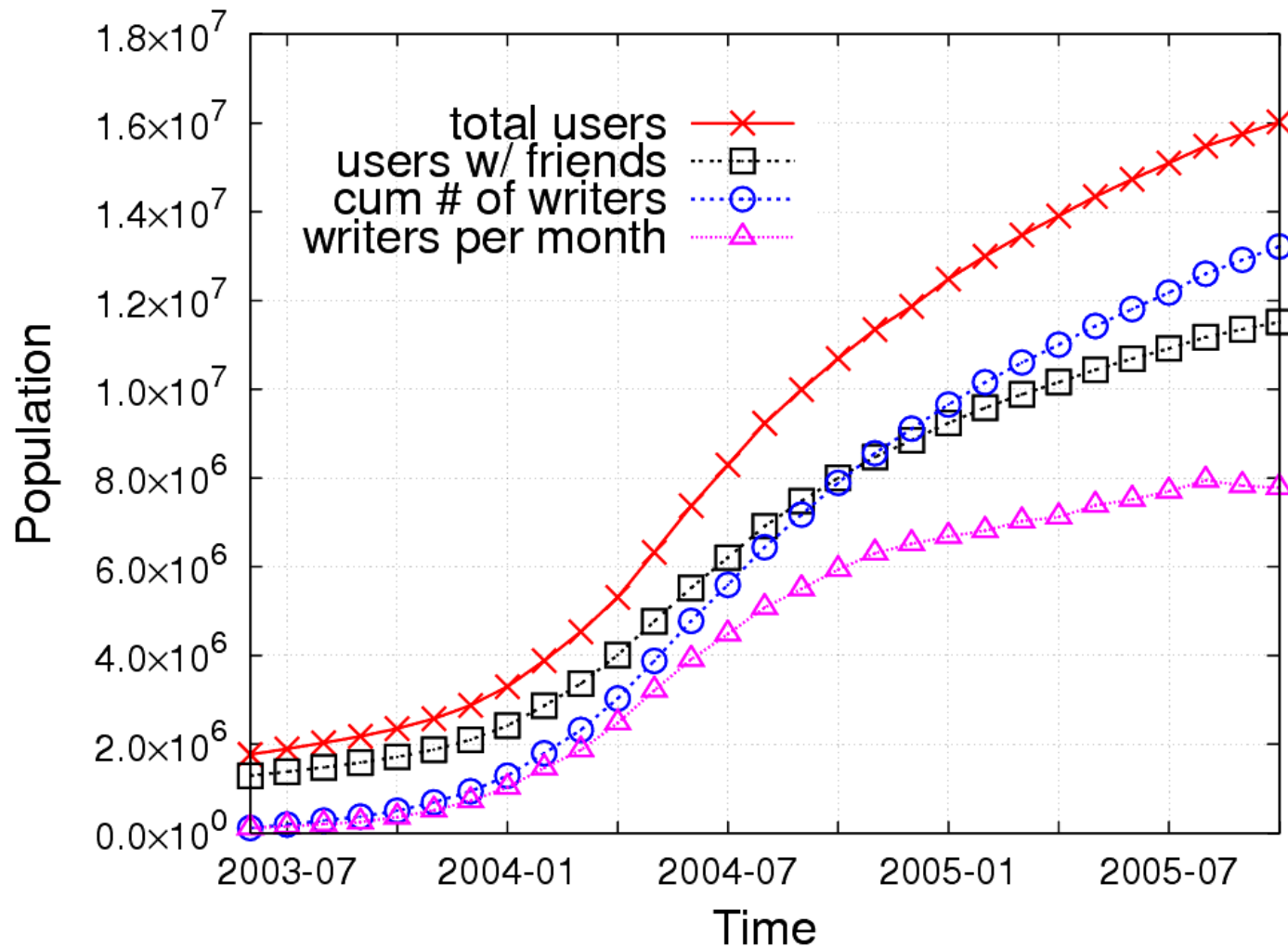


- Largest SNS in South Korea
 - Started in September 2001
 - 10 million users in 2004
 - 20 million users out of 48 million population
- Front runner of many features
 - Friend (*il-chon*) relationship
 - Guestbook
 - Testimonial (*il-chon-pyung*)
 - Photos - scraps
 - Avatar in cyber home

Uniqueness of Cyworld Data

- Complete snapshots
 - Apr/Sep/Nov of 2005
- Complete dump of guestbook logs
 - Jun 2003 to Oct 2005

Cyworld Growth in Numbers



Degree Distribution

Two scaling regions

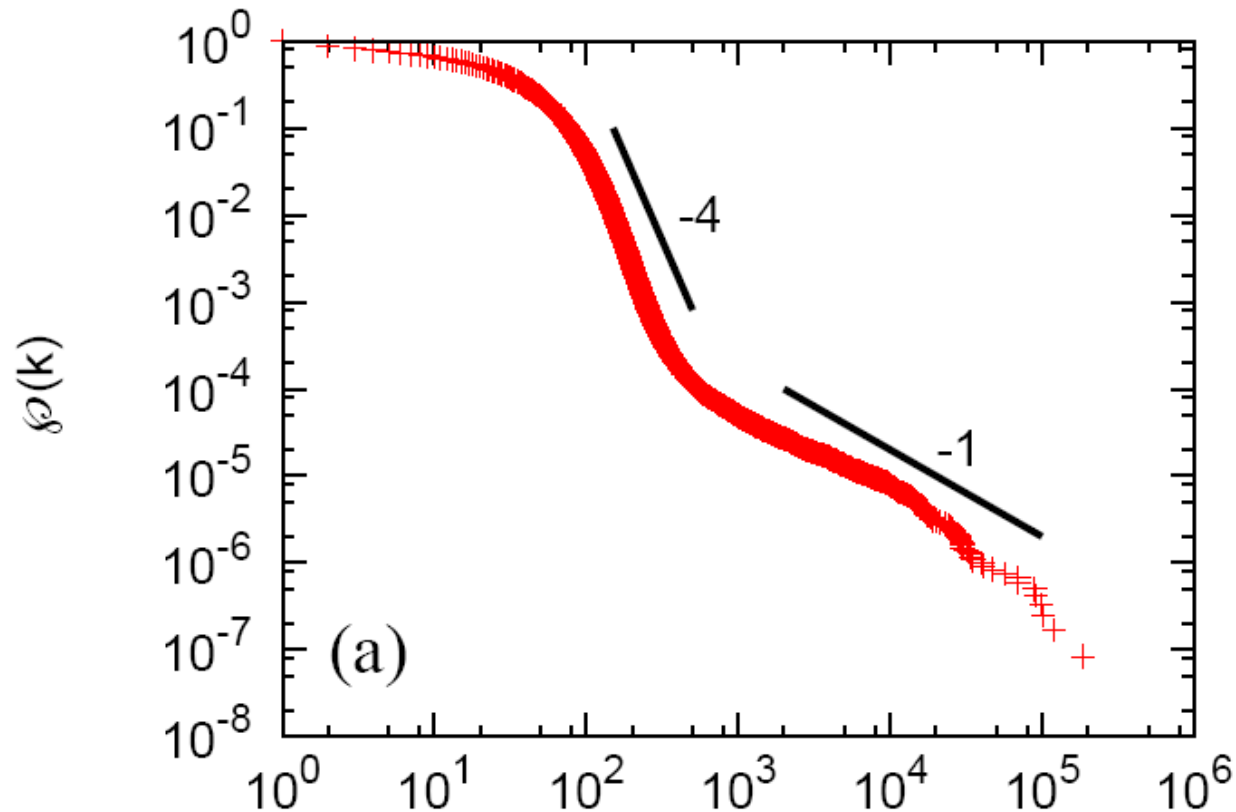
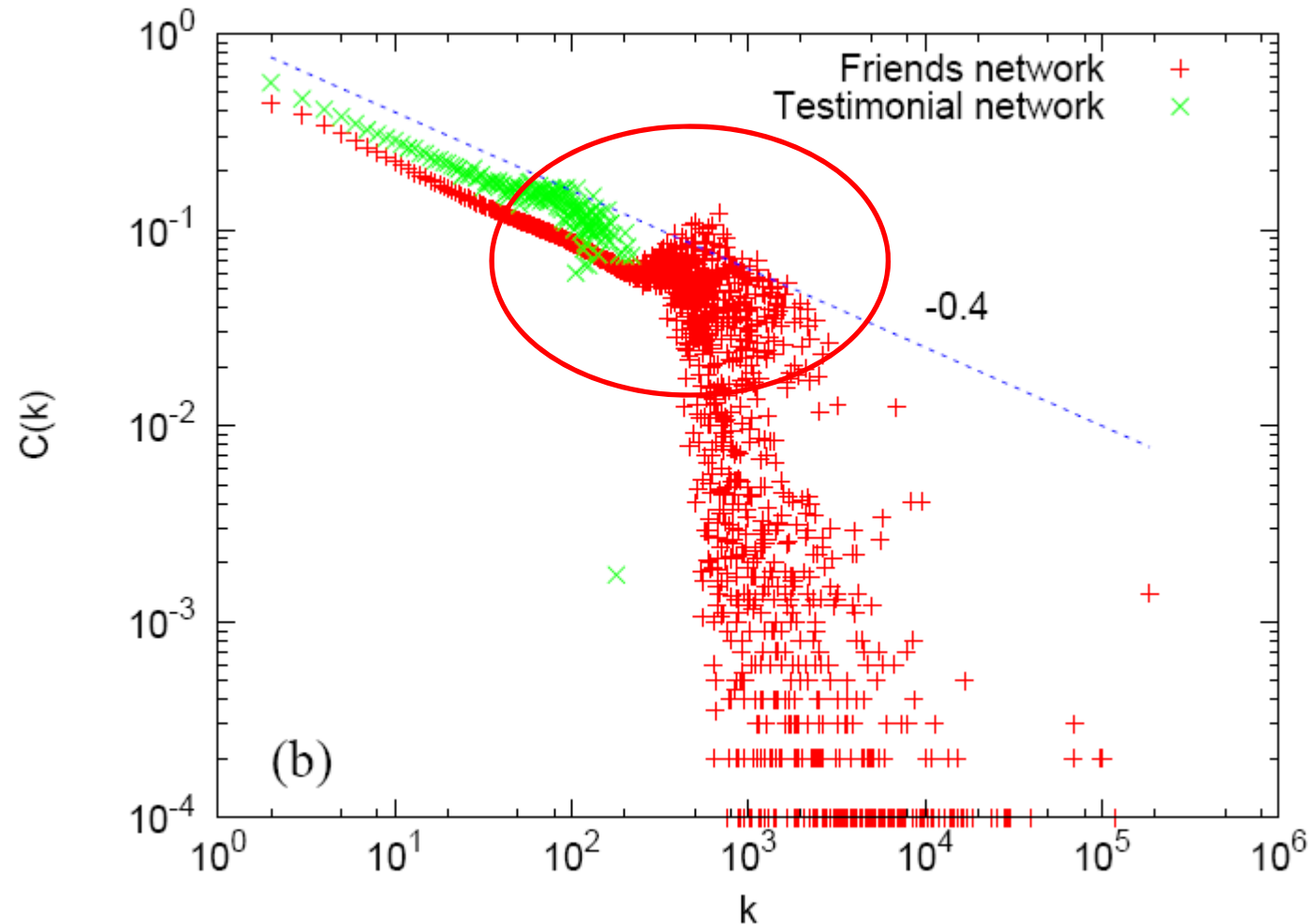
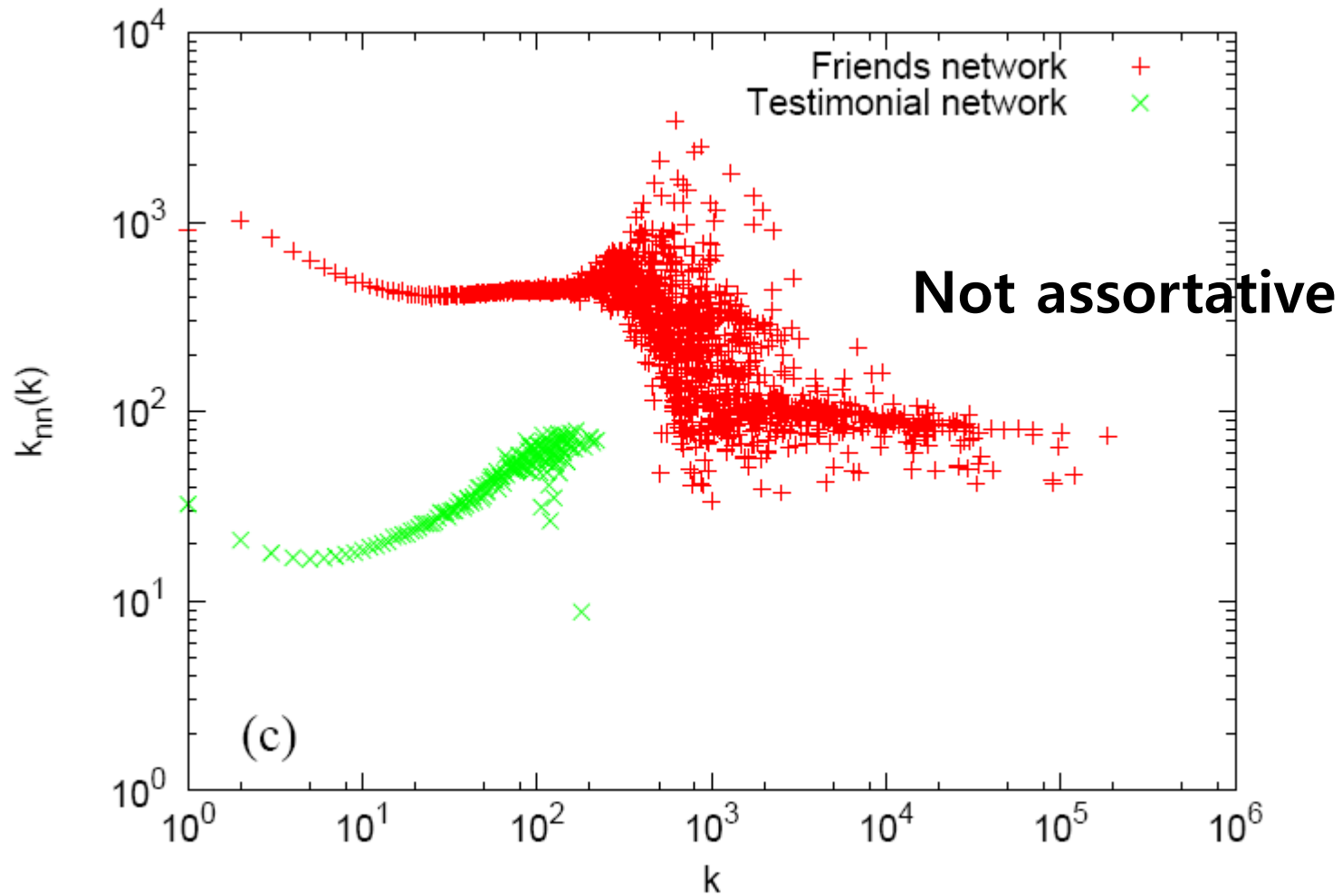


Figure 1-(a): degree distribution, CCDF

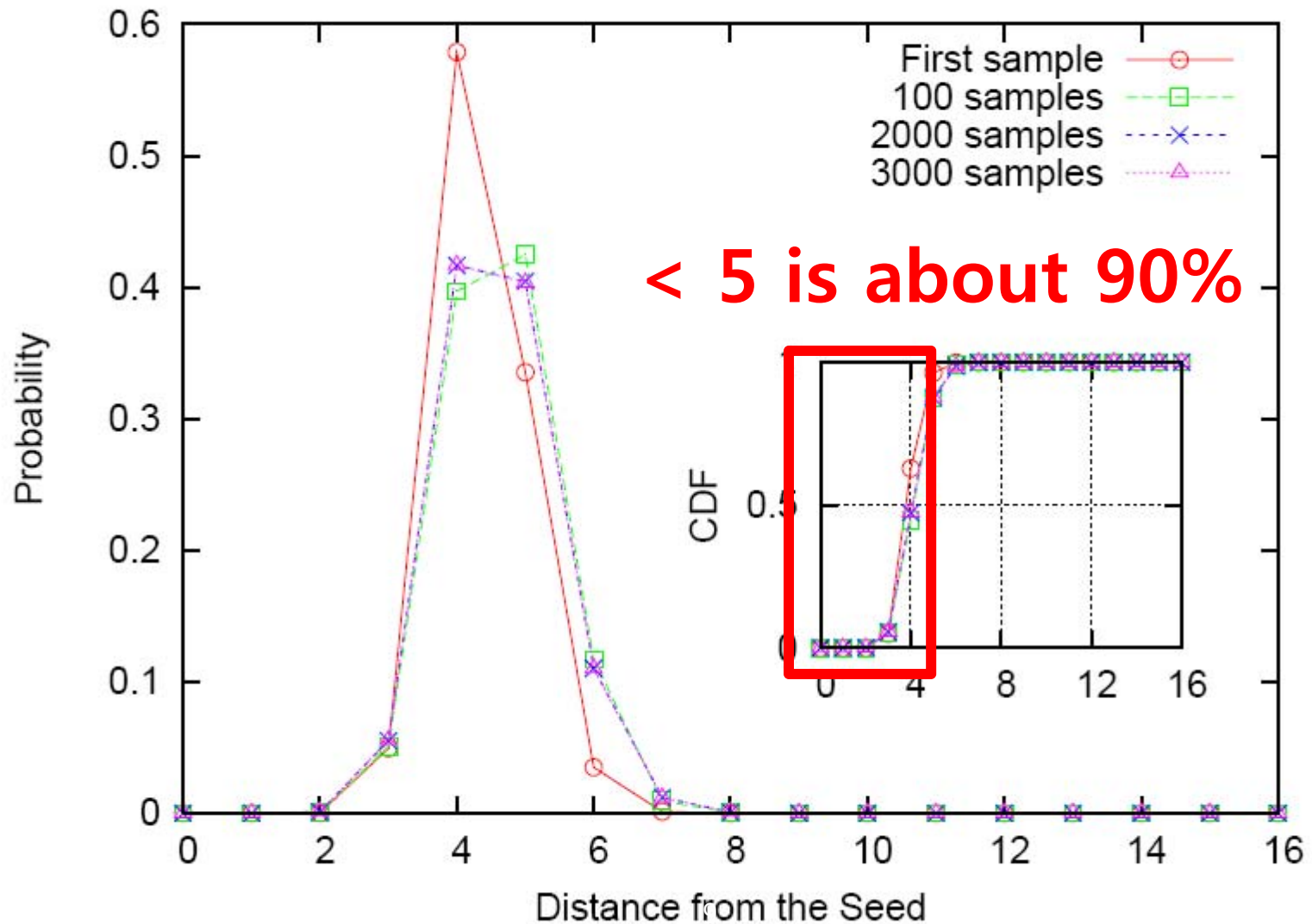
Clustering Coefficient Distribution



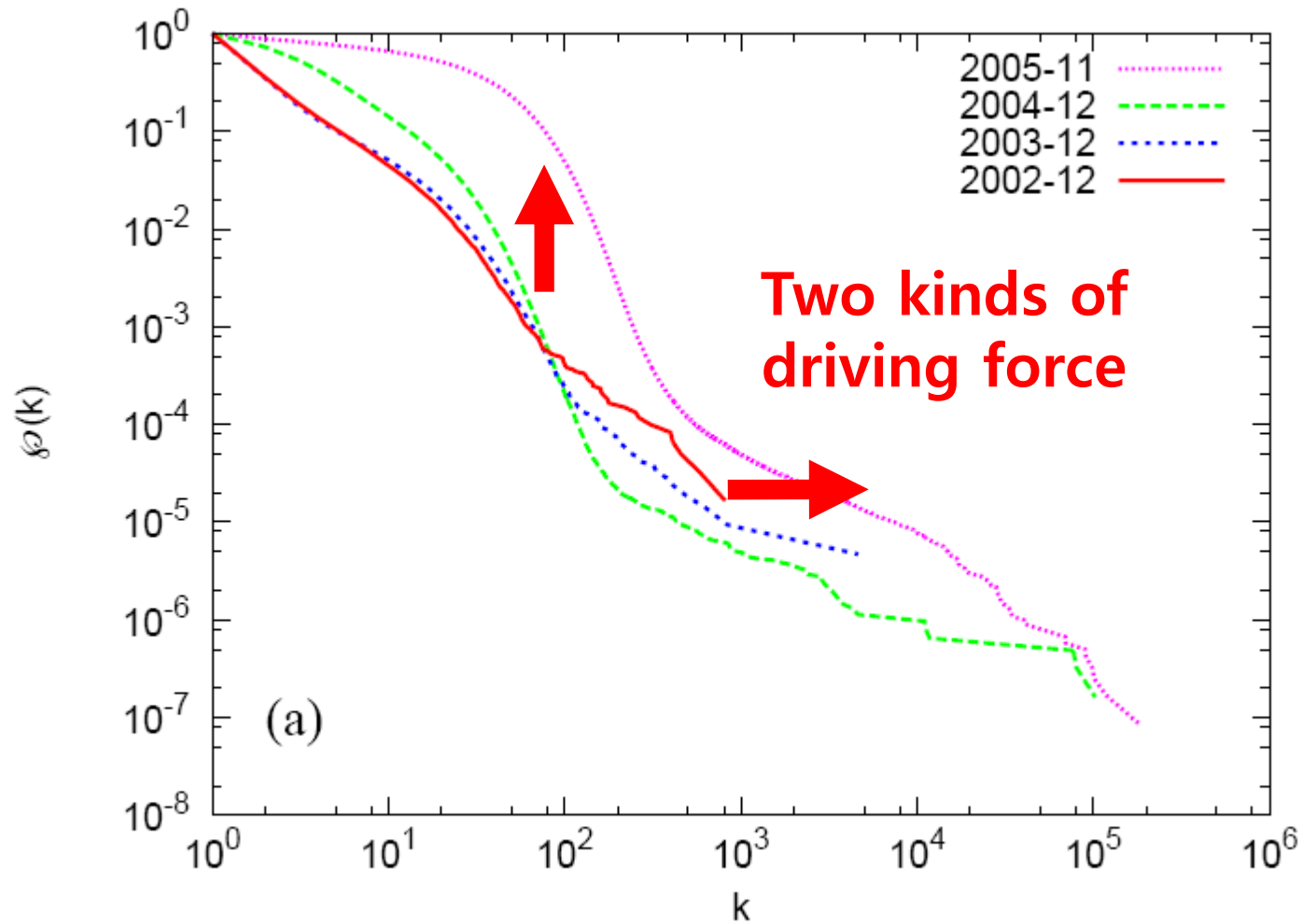
Degree Correlation



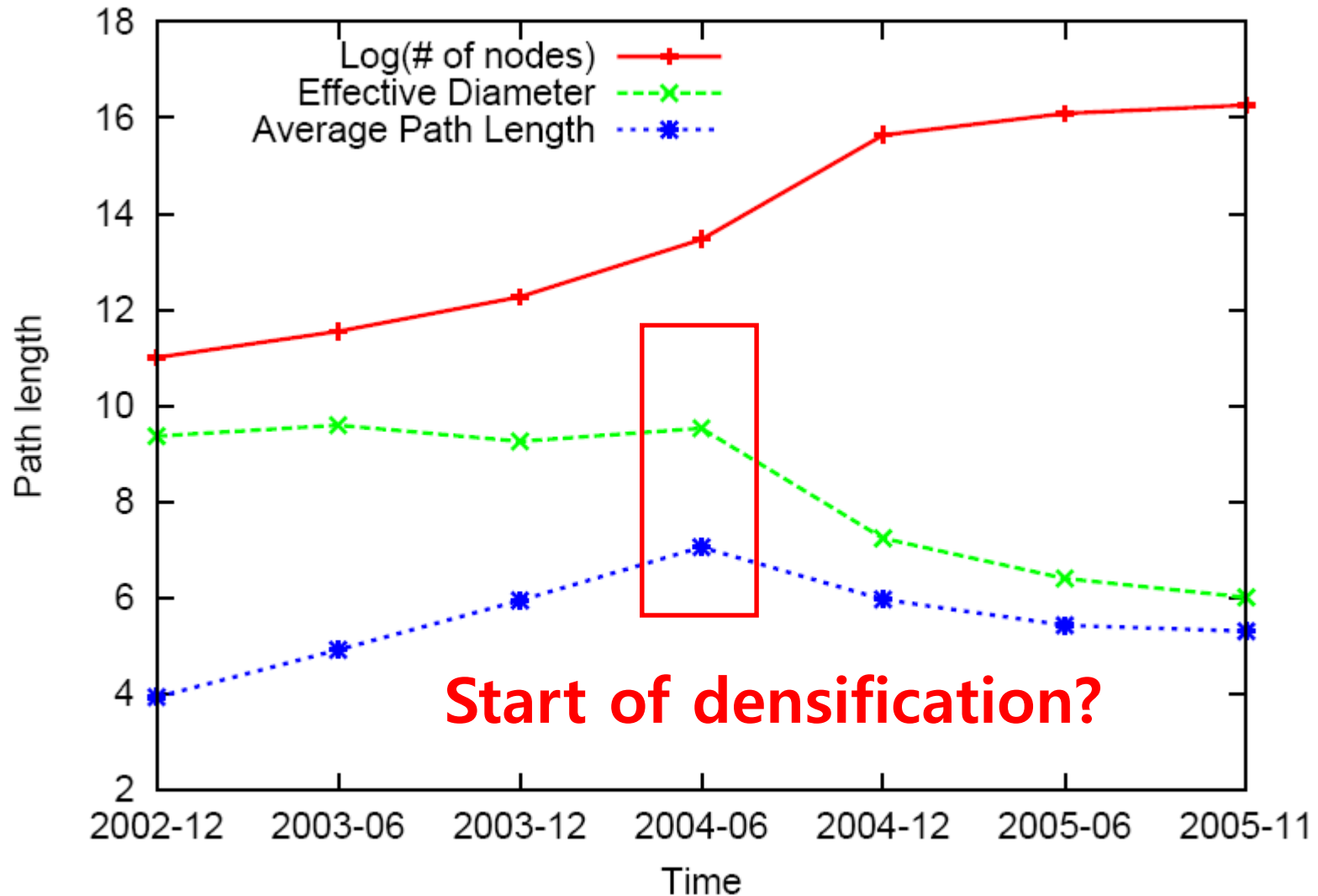
Average Path Length



Evolution of Degree Distributions



Evolution of Path Length



MySpace Data Set



- Snowball sampled
 - During Sep/Oct 2006
 - Random seed to 100,000 users
 - About 23% of users had friend list hidden

Orkut Data Set

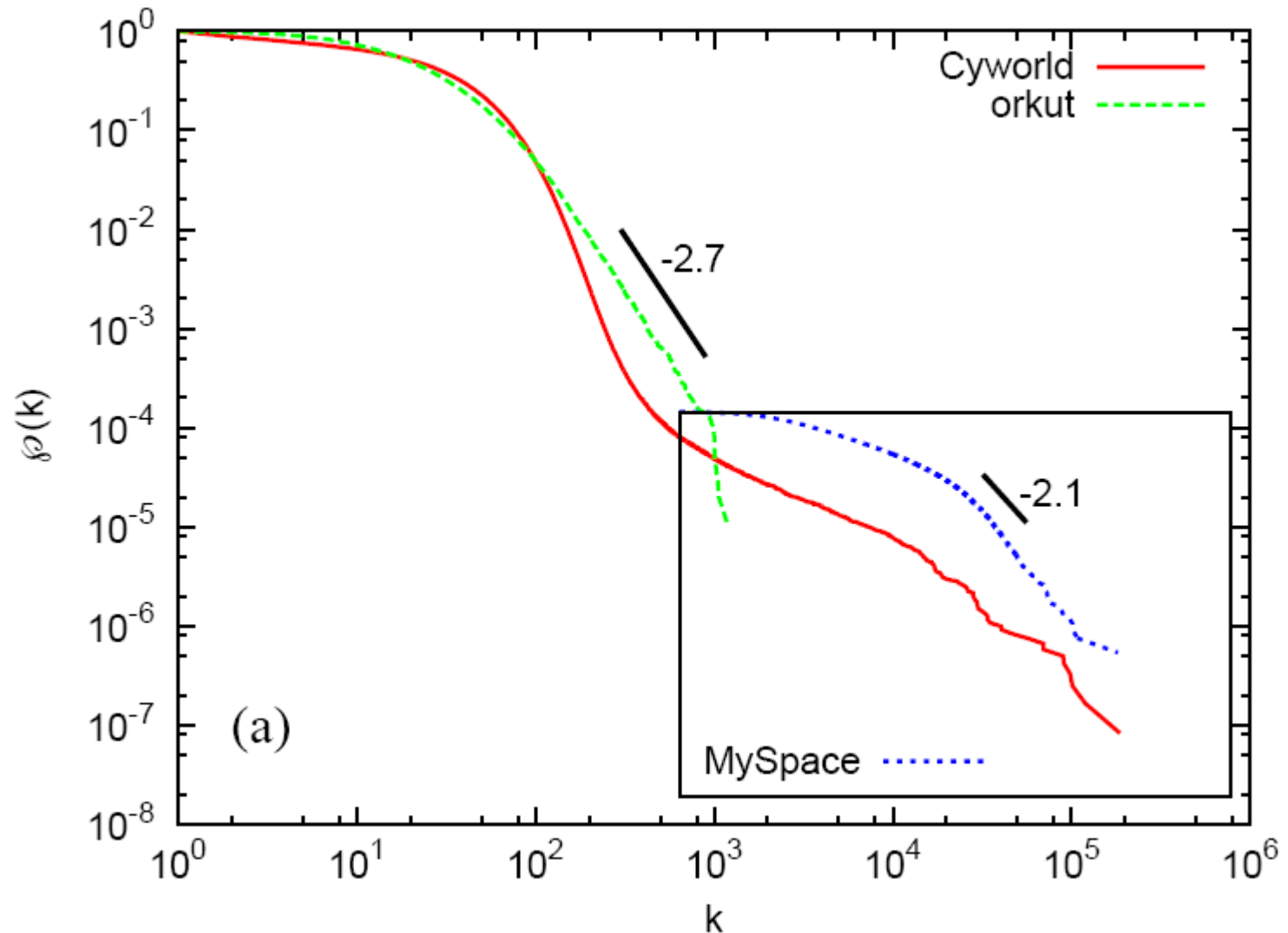
orkut

- Google SNS
- Snowball sampled
 - During Jun to Sep 2006
 - 100,000 users

Thoughts on Sampling

- Challenges in crawling the web for data
 - Random access?
 - Direction?
- Sampling large graphs
 - Leskovec & Faloutsos, KDD '06
 - Stutzbach et al., GI '06, IMC '06
- Not aware of samples collected by
Metropolized Random Walk

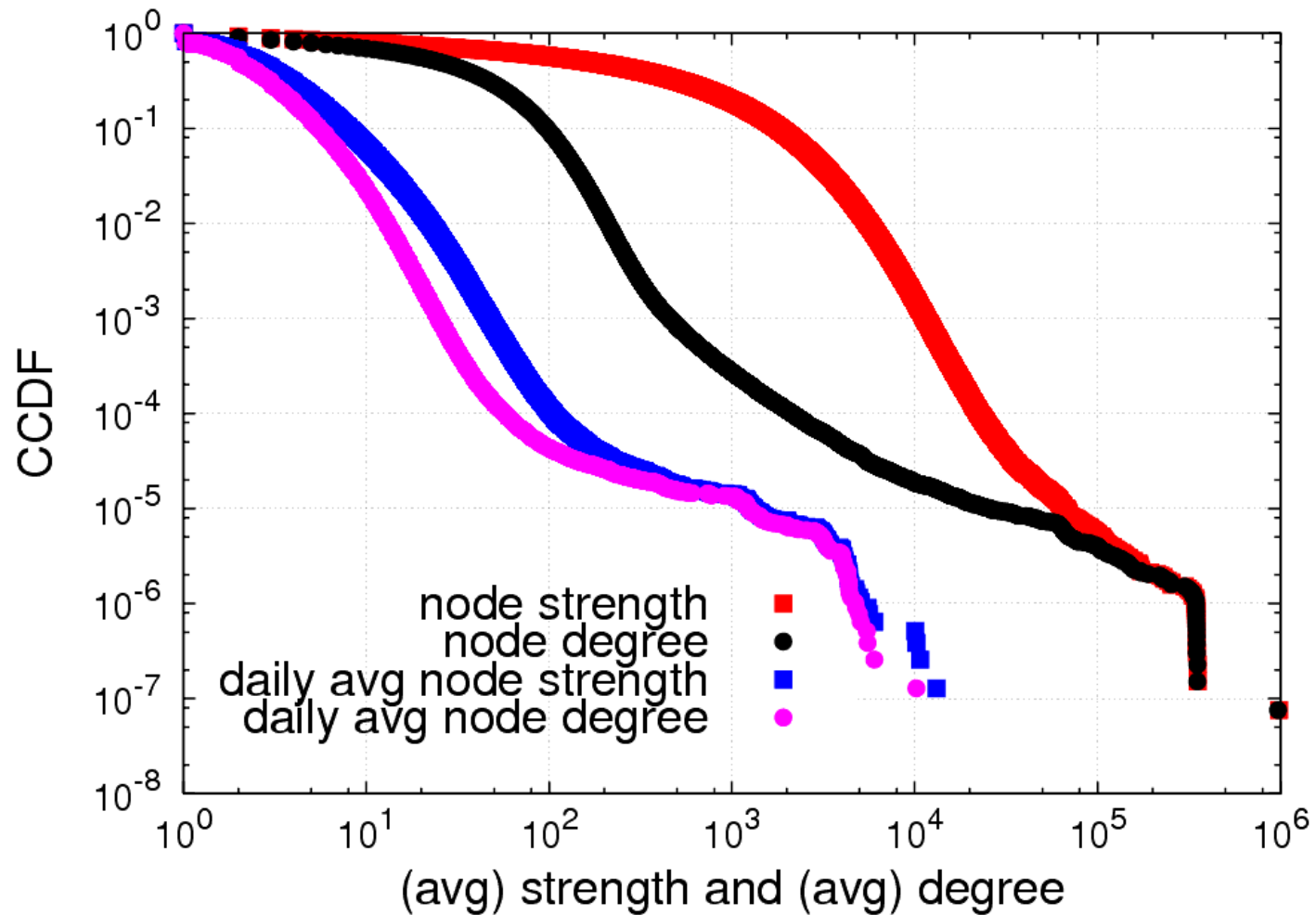
Degree Distributions



Activity Network

- Graph Representation of User Activity
 - Node = Users
 - Link = A guestbook entry from a user to another
 - Weight = # of entries
 - Strength = Sum of weights on outgoing edges

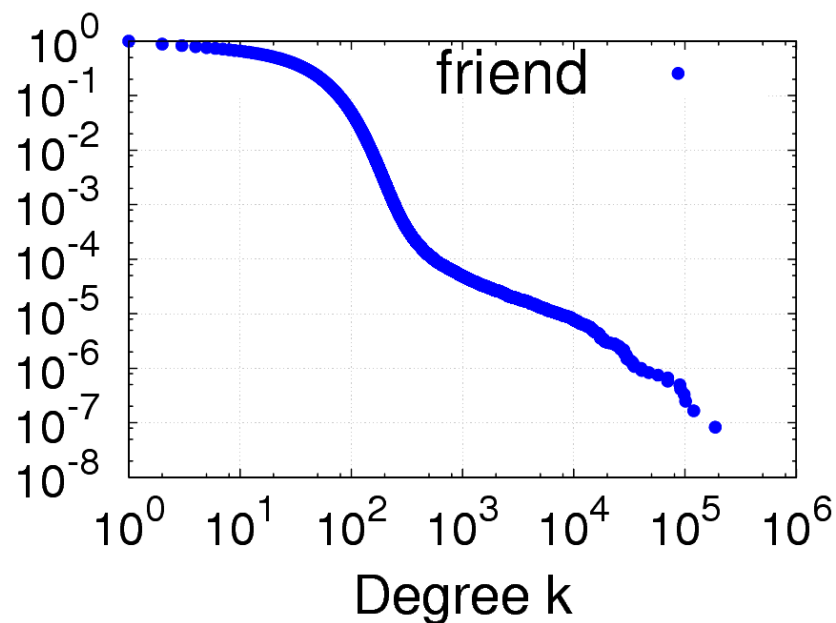
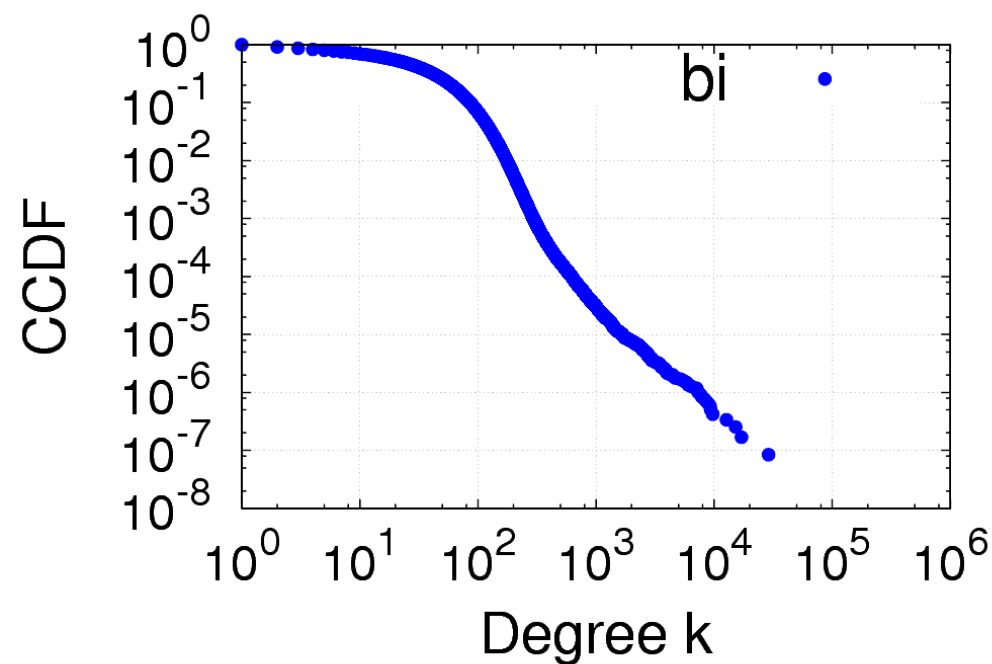
CCDF of Strength and Degrees



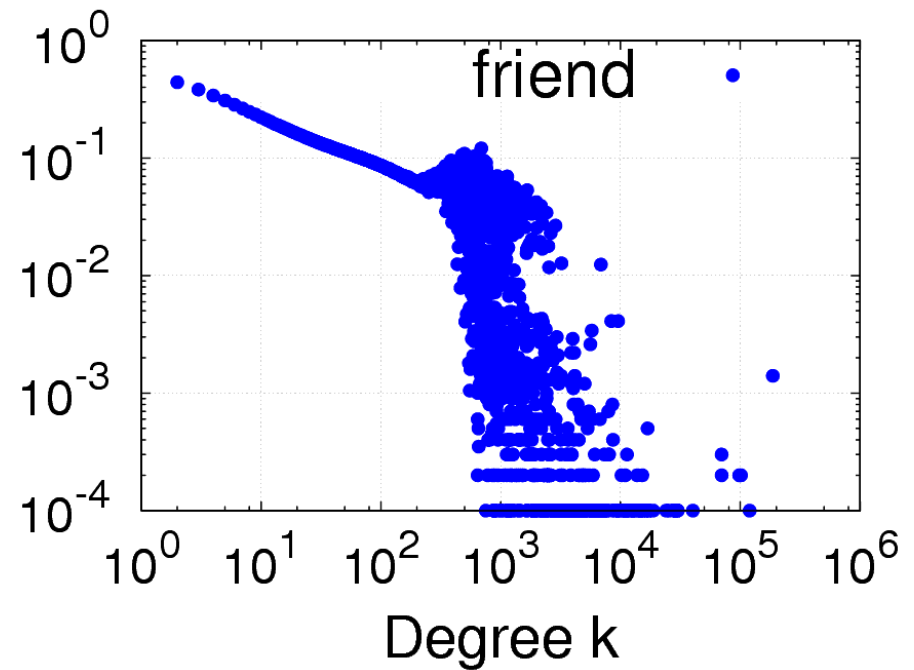
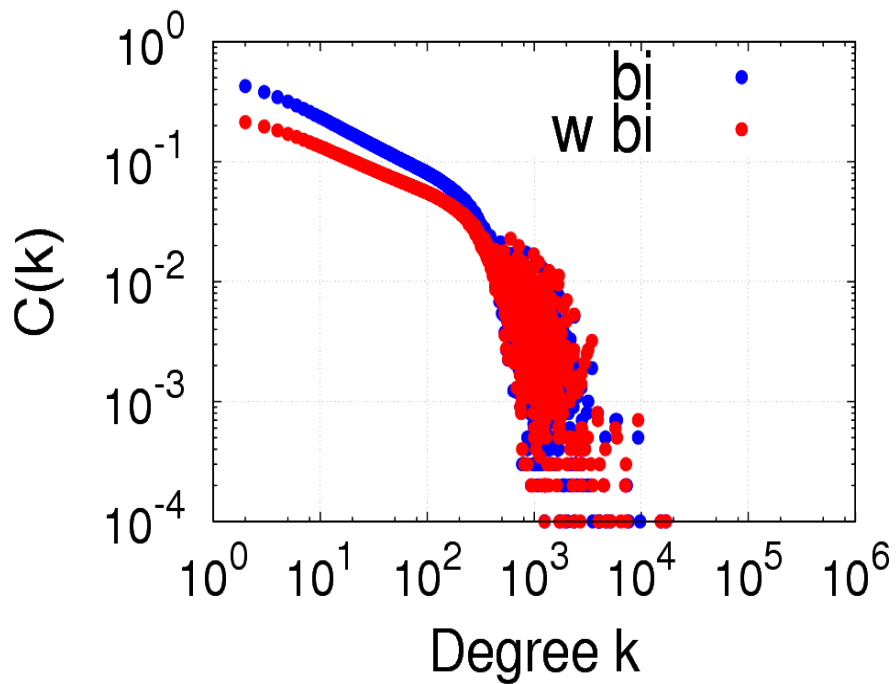
Friend vs Activity

- Metrics of Comparison
 - Degree distribution
 - Clustering coefficients
 - Degree correlation
 - K-core

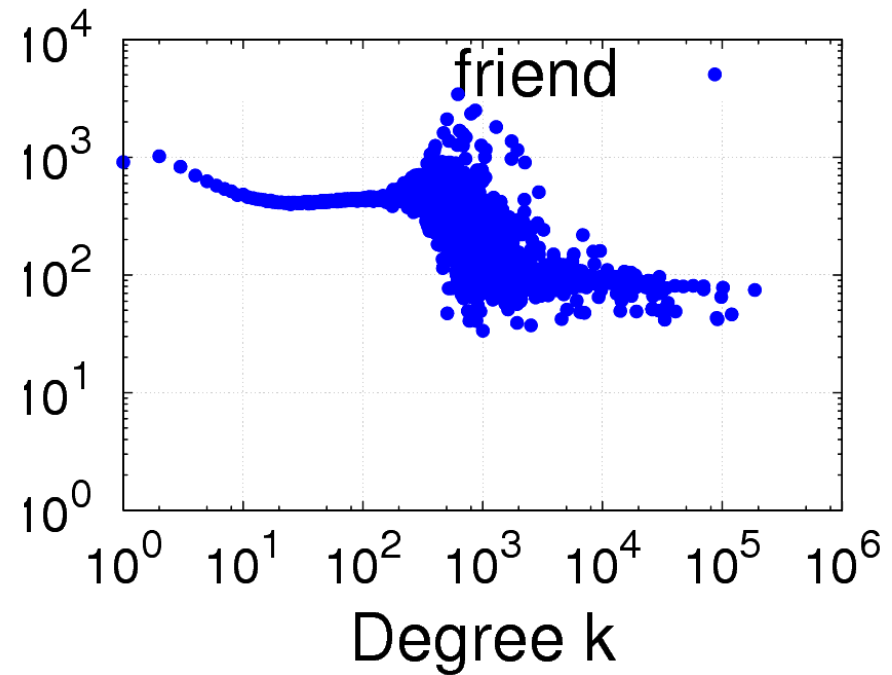
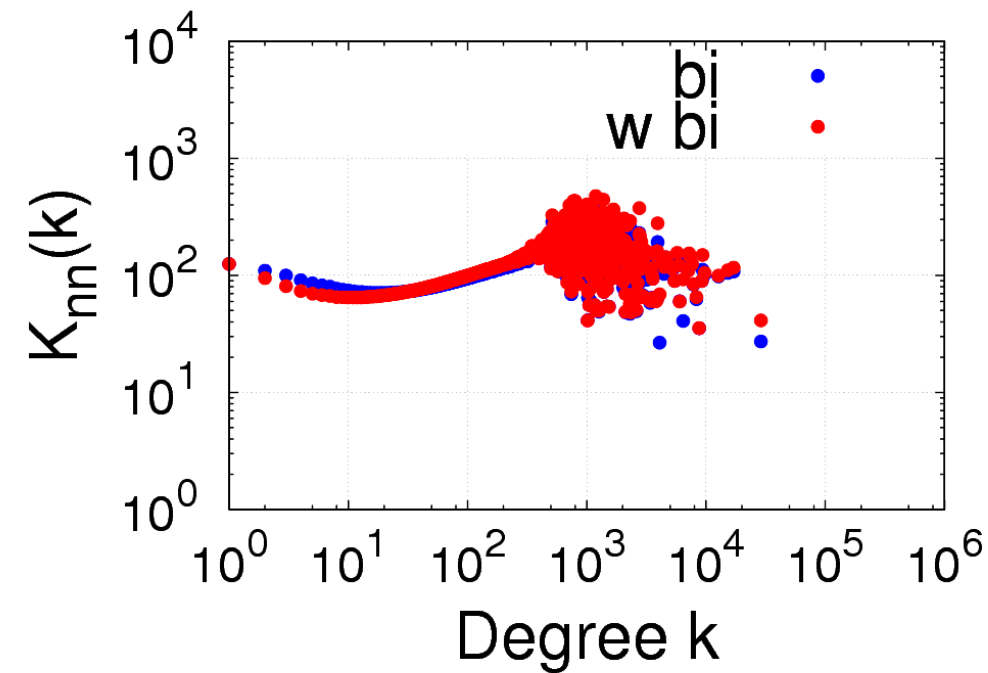
Degree Distribution



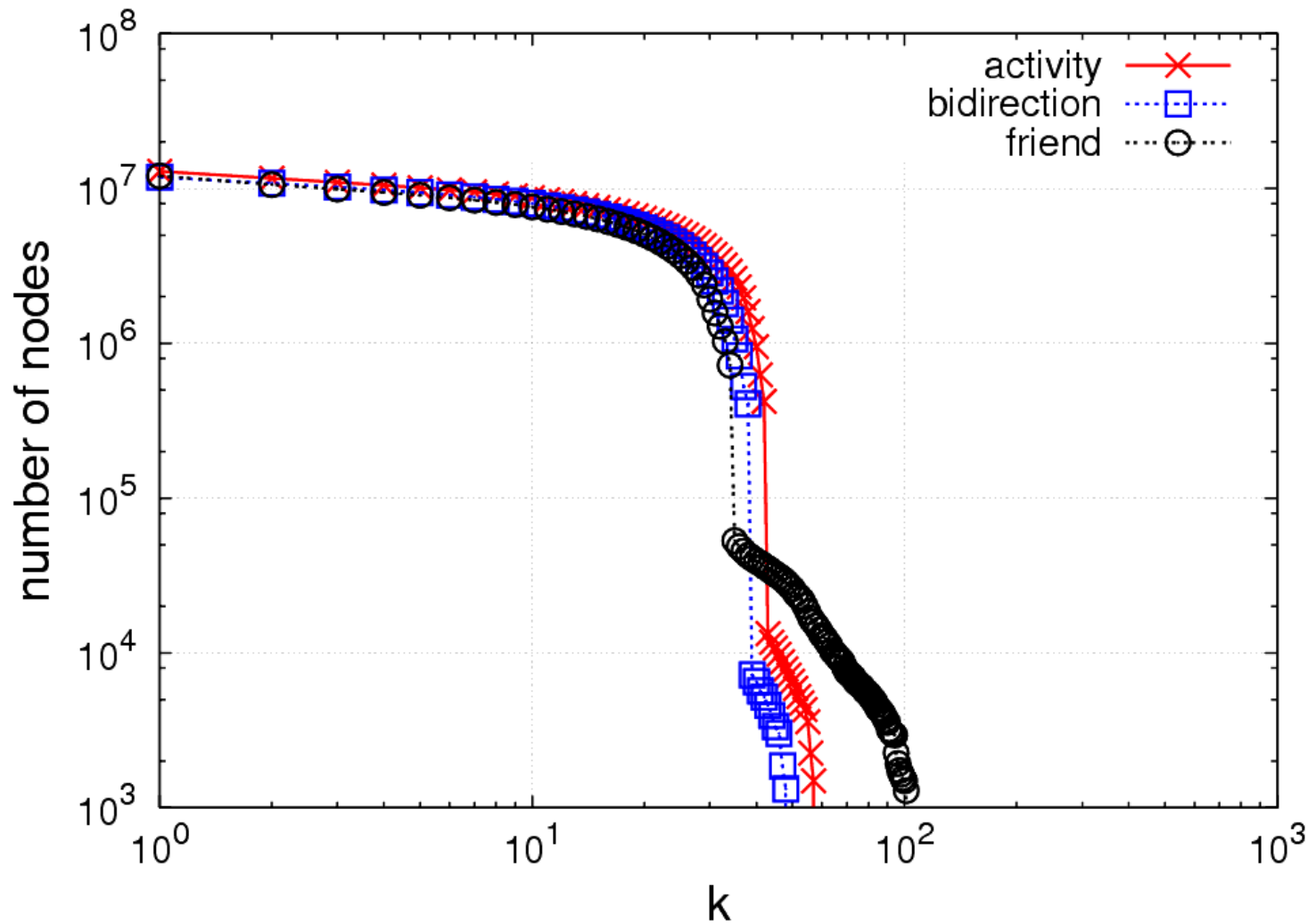
Clustering Coefficients



Degree Correlations



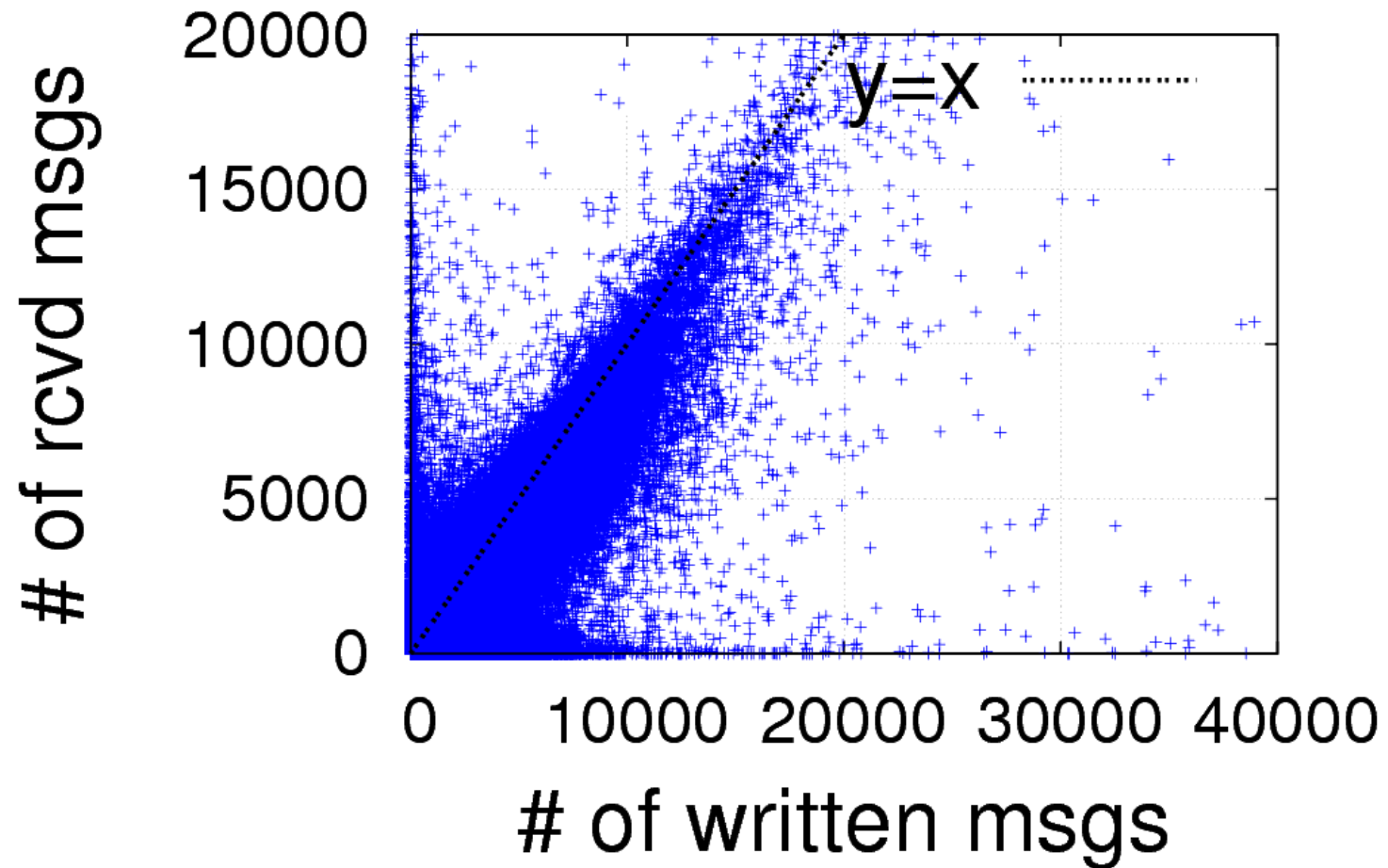
K-Core



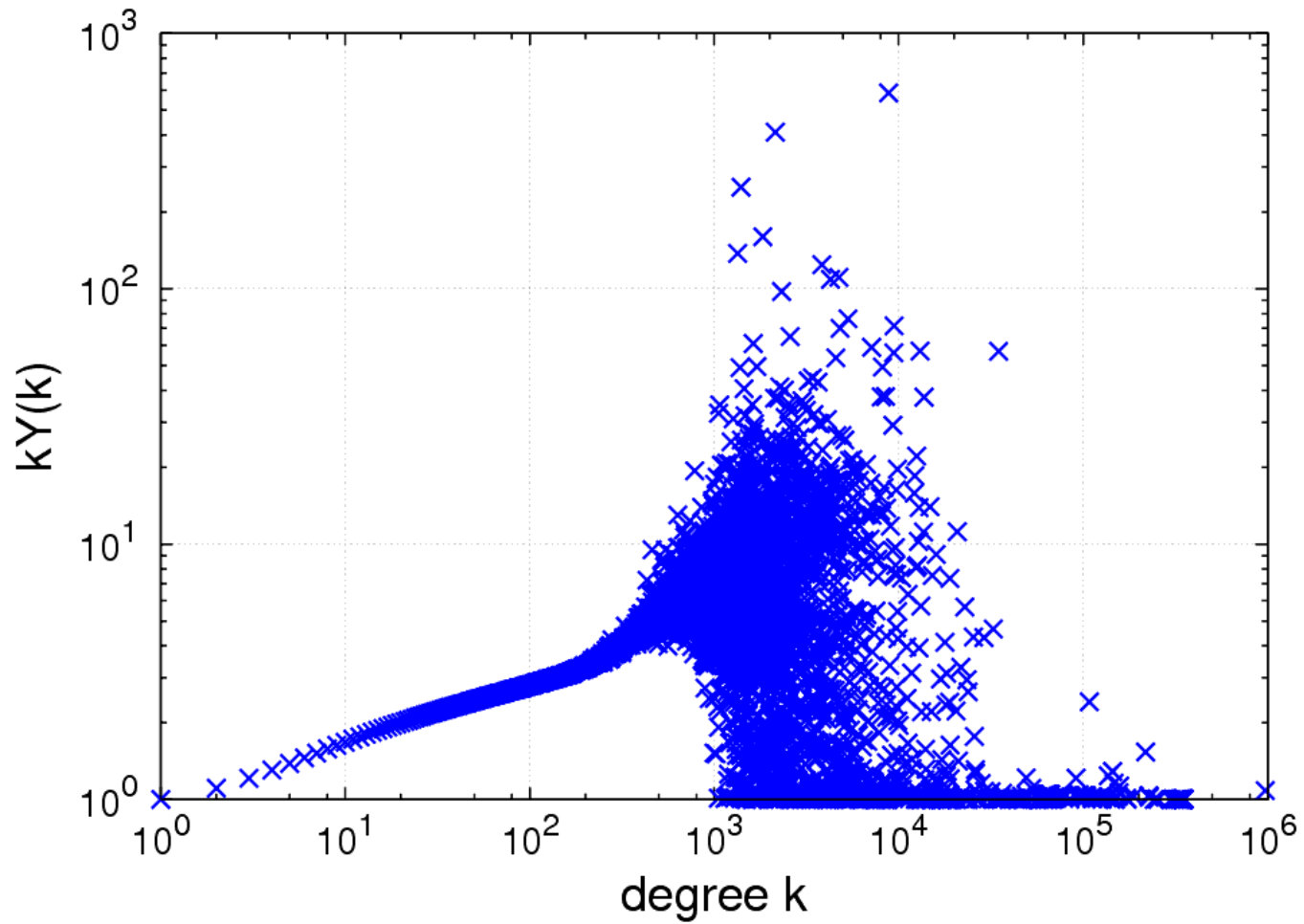
Activity Network

- As a directed and weighted graph
 - Link reciprocity
 - Disparity
 - Network motifs
- Capacity Cap
- Activity Time Interval

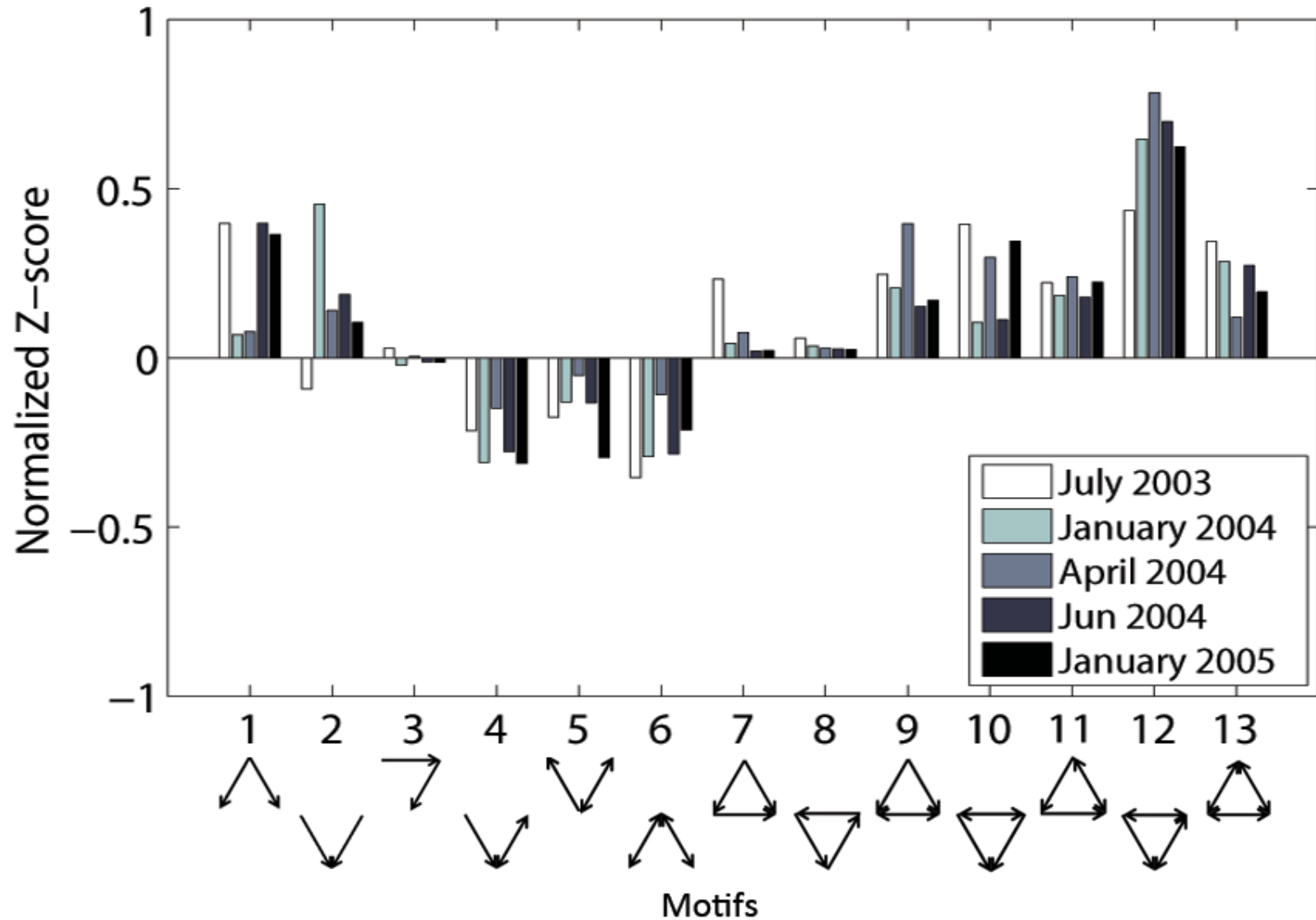
Link Reciprocity



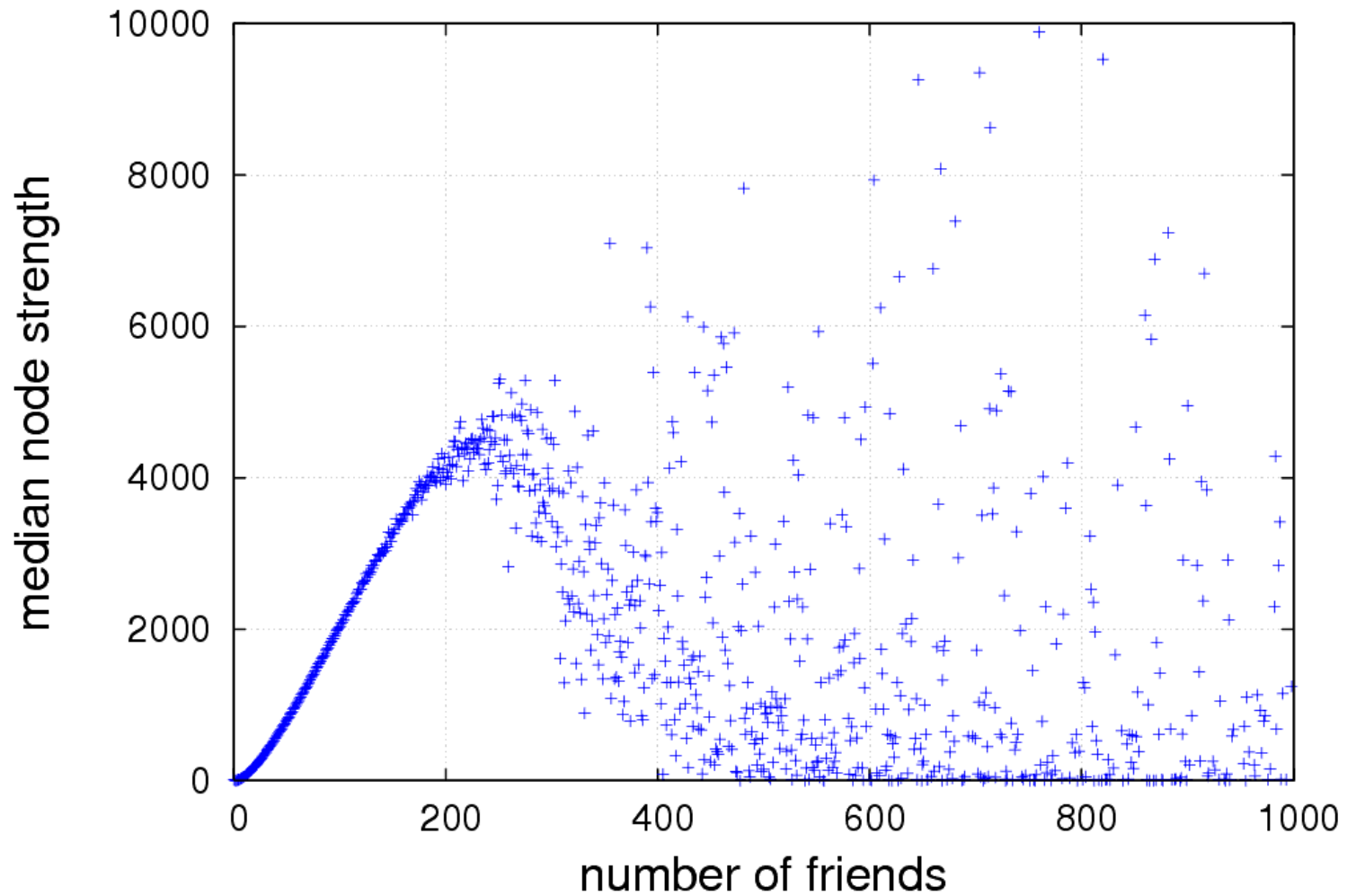
Disparity



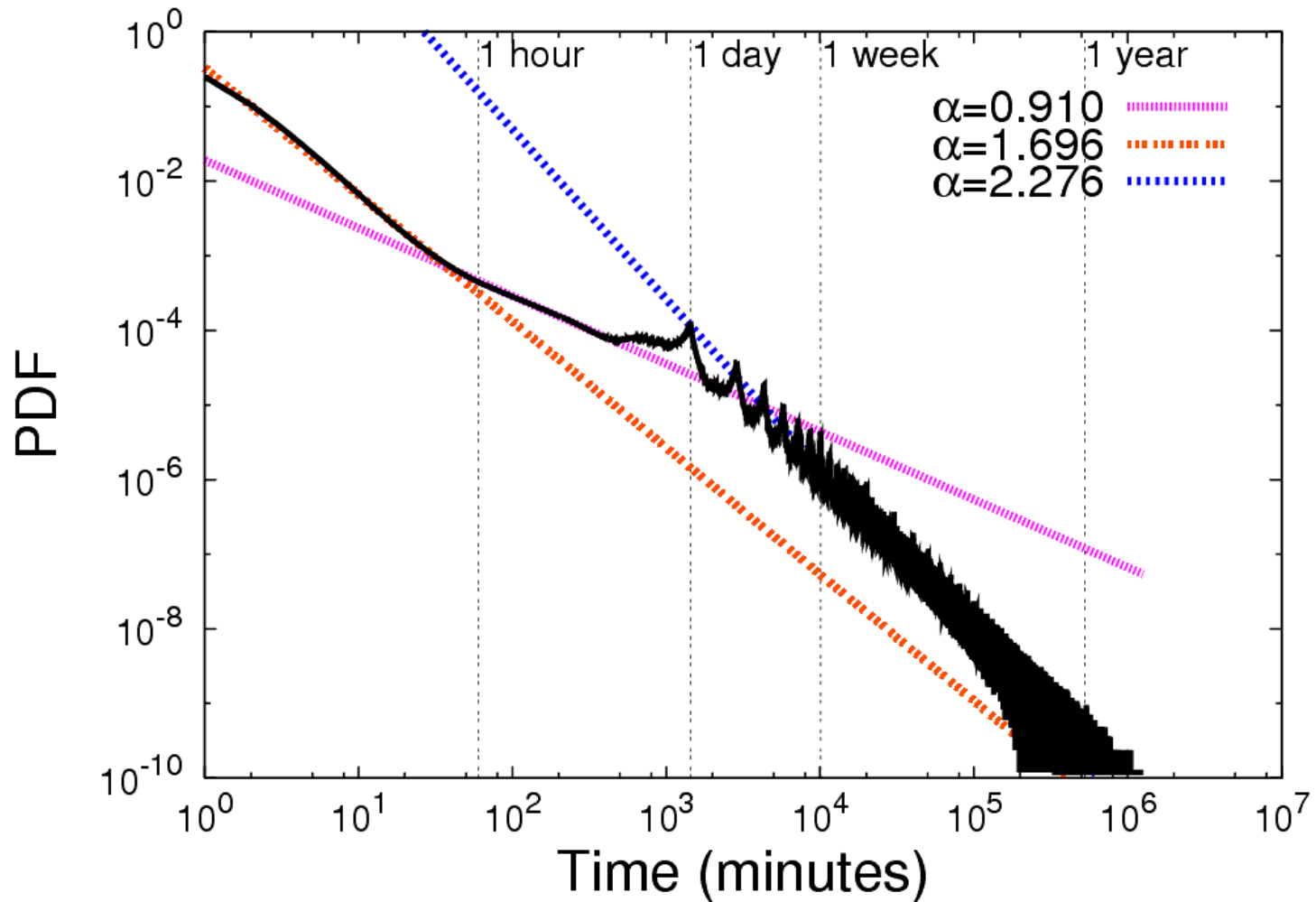
Network Motifs



Capacity Cap



Activity Time Interval



Summary

- Comparison of friend and activity networks
 - Activity closer to human social networks
- Activity as a weighted and directed graph
 - Link reciprocity shows equality
 - Disparity shows the importance of power users
- Future Work
 - Group dynamics
 - Microscopic dynamics