Online Social Networks - Connecting People Like Never Before

SIAM Annual Meeting 2008
July 8\textsuperscript{th}, 2008

Sue Moon
\textit{Joint Work with}
Hyunwoo Chun, Seungyeop Han, Haewoon Kwak,
Young-Ho Eom, Yong-Yeol Ahn, Hawoong Jeong

KAIST
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
Degree Distribution

Two scaling regions

Figure 1-(a): degree distribution, CCDF
Clustering Coefficient Distribution

![Clustering Coefficient Distribution Graph](image_url)
Degree Correlation

(c) Not assortative
Average Path Length

< 5 is about 90%
Evolution of Degree Distributions

Two kinds of driving force

\( \varphi(k) \) vs. \( k \) for different years.
Evolution of Path Length

Start of densification?
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

- 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

\[\varphi(k)\]

\(k\) vs. \(\varphi(k)\)

- Cyworld
- orkut
- MySpace

Exponents:
- Cyworld: -2.7
- MySpace: -2.1
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

CCDF

Degree $k$

CCDF

Degree $k$
Clustering Coefficients

![Graph](image)

- $C(k)$ vs Degree $k$ for different categories (bi, w bi, friend)
Degree Correlations

\[ K_{nn}(k) \]

\[ \text{Degree } k \]

\[ 10^0 \quad 10^1 \quad 10^2 \quad 10^3 \quad 10^4 \quad 10^5 \quad 10^6 \]

\[ 10^0 \quad 10^1 \quad 10^2 \quad 10^3 \quad 10^4 \quad 10^5 \quad 10^6 \]
K-Core

![Graph showing the number of nodes vs. k for different network types: activity, bidirection, friend. The graph plots on a log-log scale, with k on the x-axis and the number of nodes on the y-axis. The graph shows the number of nodes decreases as k increases, with different network types exhibiting distinct patterns.]
Activity Network

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

The graph visualizes the normalized Z-score for different motifs over various months:
- July 2003
- January 2004
- April 2004
- June 2004
- January 2005

The motifs are labeled from 1 to 13 at the bottom of the graph, and the Y-axis represents the normalized Z-score.
Capacity Cap

![Graph showing the relationship between number of friends and median node strength.](image)

- **X-axis**: Number of friends
- **Y-axis**: Median node strength
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