



Effects of Ego Networks and Communities on Self-Disclosure in an Online Social Network

Presenter: Young D. Kwon
ydkwon@cse.ust.hk

2019. 8. 28

*Young D. Kwon, Reza Hadi Mogavi, Ehsan Ul Haq, Youngjin Kwon,
Xiaojuan Ma, and Pan Hui
Department of Computer Science and Engineering
Hong Kong University of Science & Technology*

Contents



I. Introduction

II. Preliminaries

III. Key Findings

IV. Inferring Self-Disclosure

V. Conclusion

Introduction

❖ Proliferation of Online Social Networks (OSNs)

Facebook



***2.1 Billion
Active Users***

Twitter



***303 Million
Active Users***

Google+



***3.5 Million
Active Users***

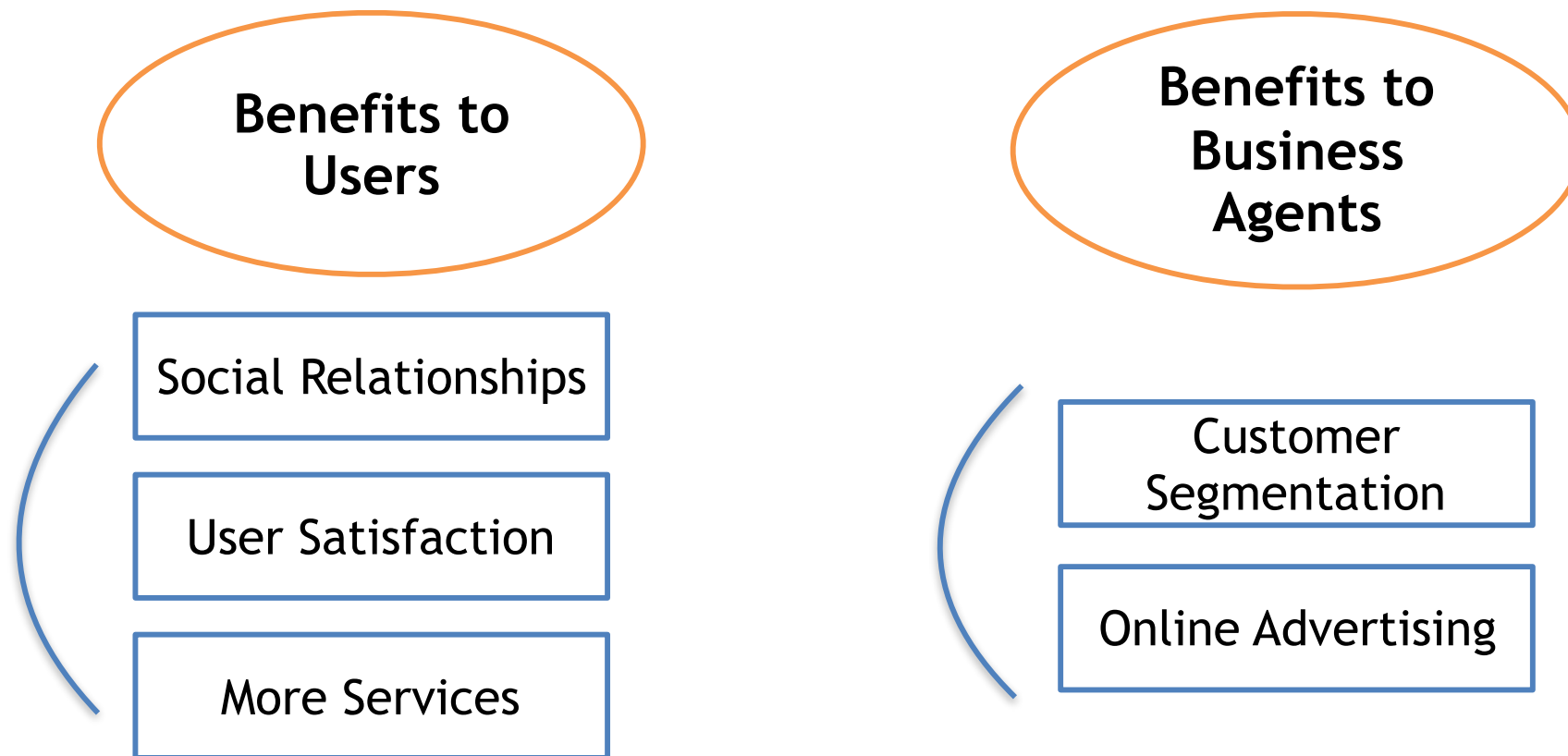
✓ Statista, January 2018. URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

✓ TechTimes, May 2015. URL: <http://www.techtimes.com/articles/51205/20150506/many-users-google-really.htm>

Motivation

❖ What is Self-Disclosure & Why is it important?

Self-disclosure: Act of revealing personal information to others



- ✓ Self-disclosure The self in social psychology. [R Archer, 1980]
- ✓ Friendship Maintenance: An Analysis of Individual and Dyad Behaviors. [D Oswald et al., 2004]
- ✓ Self-disclosure and liking: a meta-analytic review. [N. L. Collins and L. C. Miller, 1994]
- ✓ Online social networks: why we disclose. [H. Krasnova et al., 2010]

Motivation : Existing Solutions



❖ Survey datasets

- **Limited number** of survey participants

❖ Manually-annotated datasets

- A sampled subgraph may be biased
- Neglect user dynamics at **community level**

✓ Self-Disclosure Behavior on Social Networking Web Sites. [E Loiacono, 2015]

✓ The Impact of User Diversity on the Willingness to Disclose Personal Information in Social Network Services A Comparison of Private and Business Contexts. [A Schaar et al., 2013]

✓ Online social networks: why we disclose. [H Krasnova, 2010]

✓ Self-Disclosure Topic Model for Classifying and Analyzing Twitter Conversations. [Bak et al., 2014]

✓ Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. [Balani and Choudhury et al., 2016]

✓ Modeling Self-Disclosure in Social Networking Sites. [YC Wang et al., 2016]

Our Solutions



1. Conduct a quantitative study using **Large-scale data on Self-disclosure behaviors**

- 4.7 million users and 47 million relation links in an OSN
- More than 70% of all publicly known users were collected when the dataset was crawled
- Capture comprehensive and unbiased view of network structures on a large scale

2. Take into account the both **ego networks and communities**

- Analyze the influence of users' direct social networks and communities

Objective

❖ 3 Research Questions



Q1: What **ego network** properties can be derived and how much do those features influence the users' self-disclosure?

Q2: What **community** properties can be derived and how much do those features influence the users' self-disclosure?

Q3: To what extent is the **self-disclosure of users affected by** network properties at the individual and community levels?

Contents



I. Introduction

II. Preliminaries

III. Key Findings

IV. Inferring Self-Disclosure

V. Conclusion

Data & Terminologies

❖ Google+ Dataset

- **Large Scale** (4.7M users, 47M links)
- >70% of all publicly known users
- Capture **comprehensive and unbiased view** of network structures on a large scale

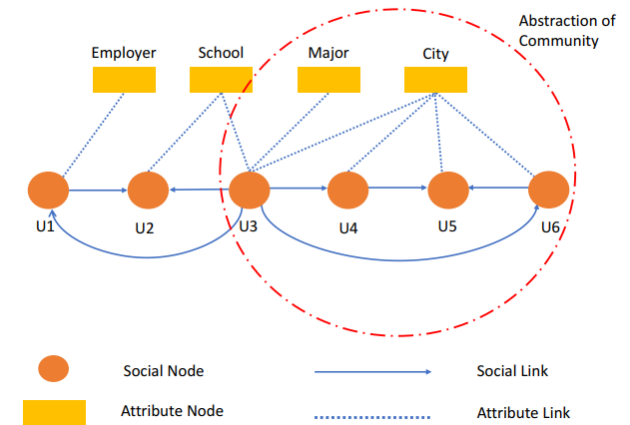
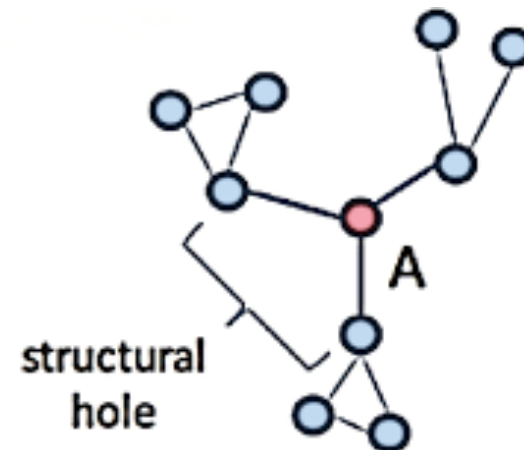
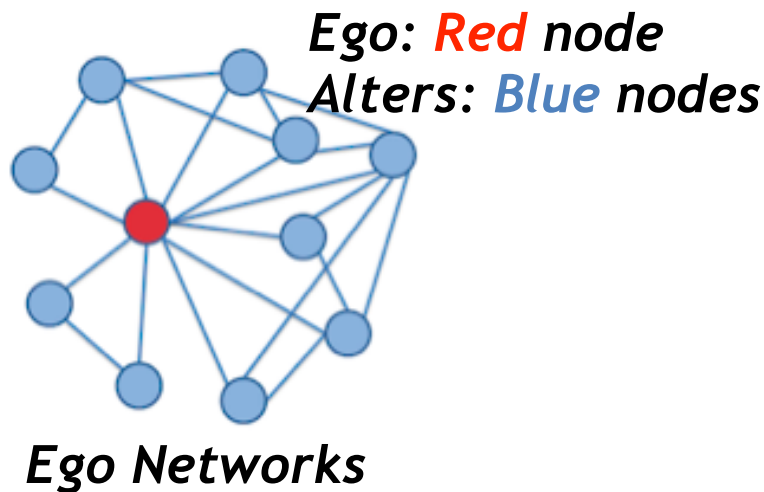


Illustration of an OSN with an abstract community

❖ Ego Networks & Structural Holes Theory



Characterization of Users



❖ Online Privacy Theory

- Inspired by the **Communication Privacy Management theory (CPM)**
- **Open Users:** disclose all optional information
- **Closed Users:** disclose none
- **Moderate Users:** rest of users who lie between open and closed users

Contents



I. Introduction

II. Preliminaries

III. Key Findings

IV. Inferring Self-Disclosure

V. Conclusion

Self-Disclosure in Ego Networks

- ❖ **Q1:** What **ego network** properties can be derived and how much do those features influence the users' self-disclosure?

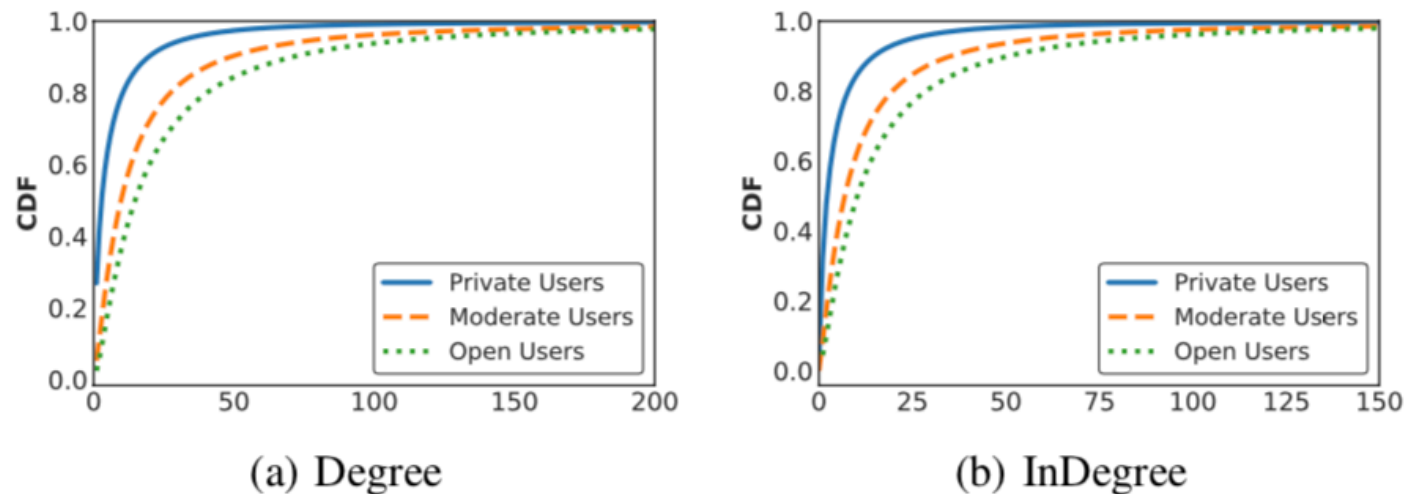


Figure: Structural differences in ego networks of closed, moderate, and open users

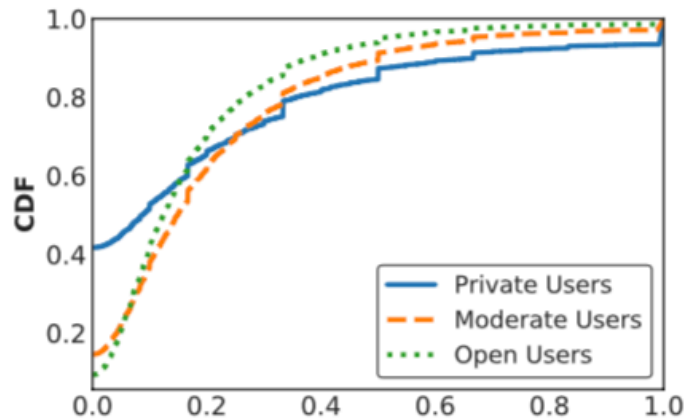
❖ Comparison of Medians

- Open vs. Moderate / Closed: 50% / 400% medians increase
- Moderate vs. Closed: 233% medians increase

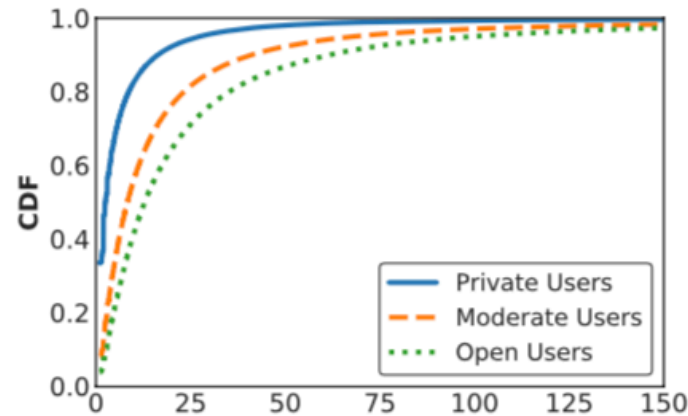
Kruskal-Wallis Test & Mann-Whitney's U Test

Self-Disclosure in Ego Networks

- ❖ **Q1:** What **ego network** properties can be derived and how much do those features influence the users' self-disclosure?



(c) Clustering Coefficient



(d) Effective Network Size

Figure: Structural differences in ego networks of closed, moderate, and open users

❖ Comparison of Medians

- (Figure C) Moderate vs. Open / Closed: 20% / 42% medians increase
- (Figure D) Open vs. Moderate / Closed: 54% / 415% medians increase

Summary & Discussion



❖ Ego Network Features

- **(Degree)** Ego network features are **positively correlated with self-disclosure**
- **(Clustering Coefficient)** Interestingly, moderate users tend to have more dense ego networks than open users
- **(Effective Network Size)** Users are more likely to reveal information when they are in bridge positions where they can utilize positional advantages
- Potential relations between the tendency of **self-disclosure** and the **sociological theory of structural holes**

Self-Disclosure in Communities

❖ **Q2:** What **community** properties can be derived and how much do those features influence the users' self-disclosure?

❖ Community Detection

- Use Louvain community detection algorithm

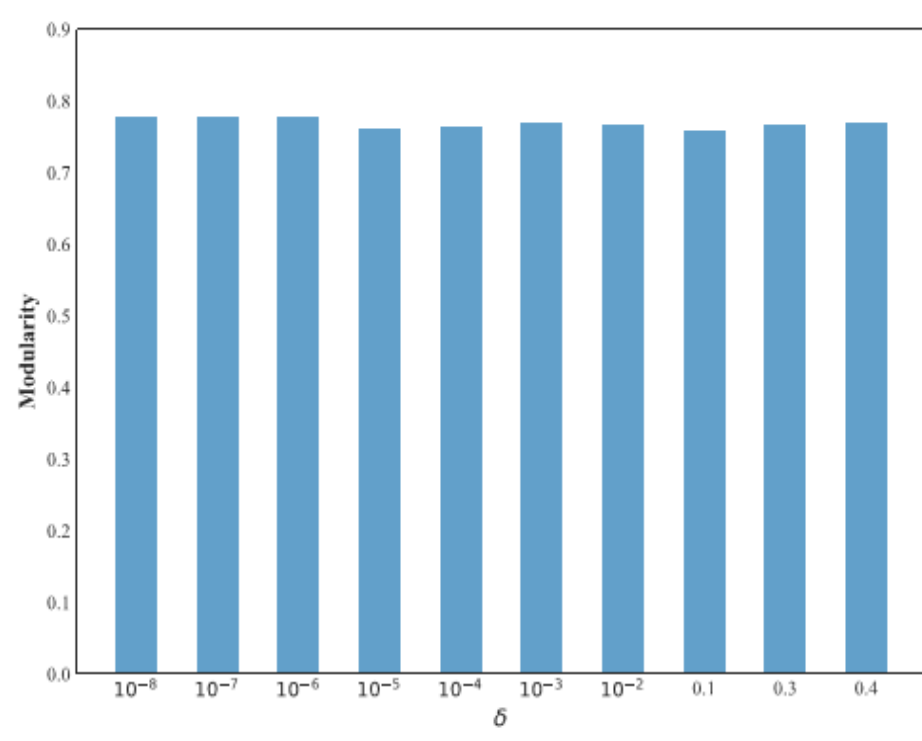
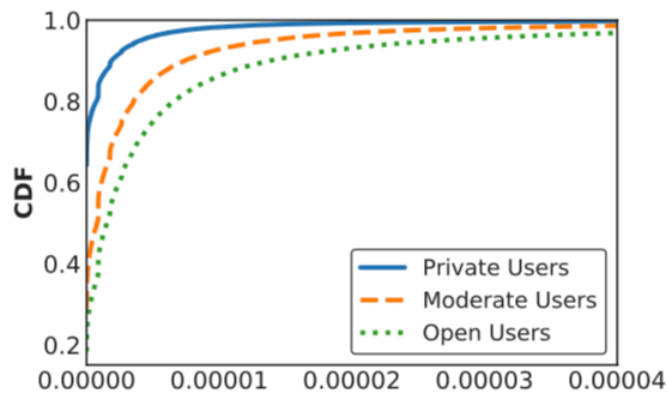


Figure: Modularity

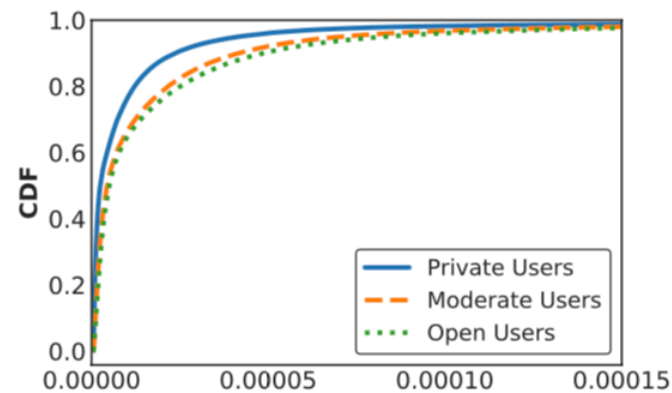
Self-Disclosure in Communities

❖ **Q2:** What **community** properties can be derived and how much do those features influence the users' self-disclosure?

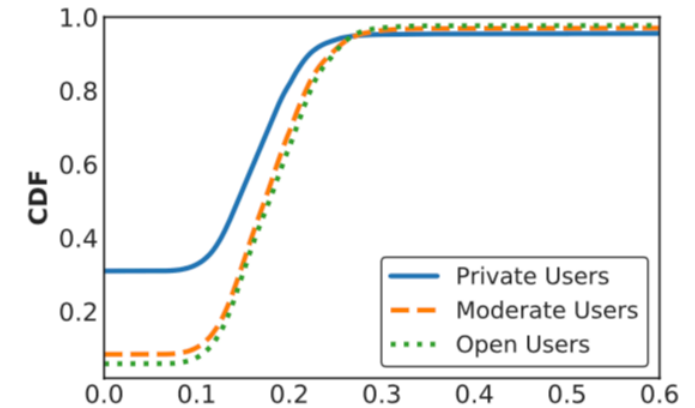
❖ Positional Properties in Communities



(a) Betweenness Centrality



(b) PageRank



(c) Closeness Centrality

Figure: Comparison of three user groups based on different positional properties in the contexts of communities

❖ Post-hoc Test

- (Figure A) **Open vs. Moderate:** Significant (87% median increases)
- (Figure B) **Open vs. Moderate:** Small Effect Size ($r = 0.03$)
- (Figure C) **Open vs. Moderate:** Small Effect Size ($r = 0.03$)

Self-Disclosure in Communities

❖ **Q2:** What **community** properties can be derived and how much do those features influence the users' self-disclosure?

❖ **Structural Properties of Communities**

- **Community Size:** KW Test $p < 0.001$
- **Network Average Clustering Coefficient:** KW Test $p < 0.001$
- **Average Degree:** KW Test $p < 0.001$
- **Distance Measure:** KW Test $p < 0.001$

Kruskal-Wallis Test (KW Test)

Summary & Discussion



❖ Community Features

- All properties of positional and structural properties of communities show significance
- **(Betweenness Centrality)** Being positioned as a bridge in a community shows significant differences according to Betweenness Centrality
- Further confirms the importance of **structural holes theory**

Contents



I. Introduction

II. Preliminaries

III. Key Findings

IV. Inferring Self-Disclosure

V. Conclusion

Predicting Self-Disclosure of Users

❖ Proposed Features

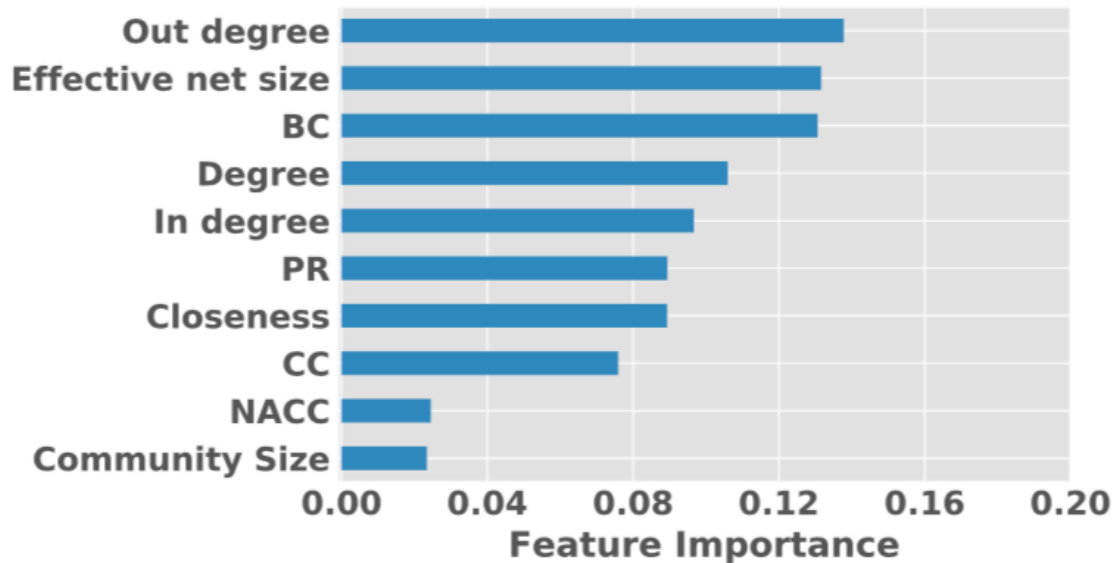
- 1) **Ego network properties:** Degree, in/out degree, CC, and effective network size (Section IV).
- 2) **Positional properties in communities:** BC, PR, and closeness centrality (Section V-B).
- 3) **Structural properties of communities:** Community size, NACC, average degree, average shortest path length, and diameter (Section V-C).

❖ Performance of models learned with different features

Feature Sets	Precision	Recall	F1
Degree & CC [11]	0.655	0.615	0.633
+ Ego Network Properties	0.656	0.745	0.697
+ Positional Properties	0.662	0.760	0.707
+ Structural Properties (Full)	0.666	0.761	0.710

Predicting Self-Disclosure of Users

❖ Feature Importance for distinguishing user types



- 2 of Top 3 important features support that users' roles as bridges in a community are important to self-disclosing behaviors

Contents



I. Introduction

II. Preliminaries

III. Key Findings

IV. Inferring Self-Disclosure

V. Conclusion

Conclusion



- ❖ **Extend** the analysis of users' self-disclosing behaviors to the large-scale dataset by characterizing them into three user types based on the **online privacy theory, CPM**
- ❖ **Study** self-disclosure of users concerning two different levels of granularity, ego networks and user communities, and present the possible explanation for users' self-disclosing behaviors using the **sociological theory of structural holes**
- ❖ **Explore** the possibility of inferring the self-disclosure levels of users given that we can only access the structural information of an OSN as well as confirm the importance of the features relevant to the structural holes theory

Future Works



- ❖ **Verify** our results with other data sources
- ❖ **Explore** possibility of predicting the future status of the self-disclosure of users dynamically
- ❖ **Investigate** the causality relation between the self-disclosure and the user's position in a network



Thank you!

Any questions?

You can find me at:

ydkwon@cse.ust.hk

<http://www.youngkwon.org/>

Young D. Kwon, Reza Hadi Mogavi, Ehsan Ul Haq, Youngjin Kwon,
Xiaojuan Ma, and Pan Hui
System and Media Lab, Dept. of CSE, HKUST