

# Social Network Analysis in HCI

Derek L. Hansen and Marc A. Smith



Marigold Bays-Muchmore (baysmuc2)

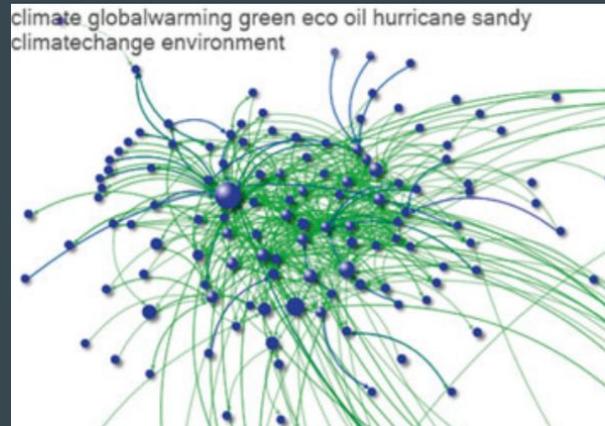
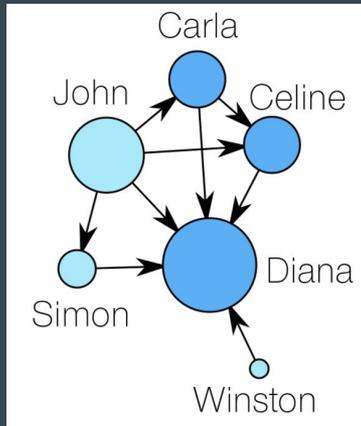
Hang Cui (hangcui2)

# Contents

Introduction	-----	What is Social Network Analysis? How does it relate to HCI?
History	-----	The emergence and evolution of SNA
HCI	-----	What are the goals of SNA in the context of HCI?
Social Network Analysis Questions	-----	What questions does SNA address?
Performing Social Network Analysis	-----	What steps and processes occur?
What Constitutes Good Work	-----	Expectations and Criteria for SNA
Example Research Applications	-----	How specifically has SNA been applied in HCI?
Looking to the Future	-	What does this mean for SNA/HCI? Where can this field head?

# Introduction

- “[SNA is the] systematic study of collections of social relationships”
- social actors implicitly or explicitly connected to one another
  - **entities** (e.g., people, organizations, nodes) joined together by *relationships* (e.g., ties, associations, links, edges)
  - more about “who you know” than “what you know” or “who you are.”



# Introduction

- HCI context
  - Provide theory and methods for better understanding/evaluating systems
  - Distinguishes between *simple population growth* and the *development of social structures* within that population
    - Success may depend on small population with a dense connections web as opposed to large population with sparser connections
    - Capture the social structure of a user population before, during, and after new technologies
  - Help identify potential influencers who can recruit new users

# History

- Social networks have formed for as long as people have interacted, traded, and engaged with one another
- Prior to the widespread use of digital information systems, generating records of social interactions was challenging
- Evolution of the methodology of social network analysis can be split into three phases:
  - Foundational Phase
  - Computational Phase
  - Network Data Deluge Phase

# History

- The Foundational Phase (Eighteenth century to the 1970's)
  - Defining and establishing the necessary mathematical graph theory foundation
    - Euler demonstrated value of using graph theory representation to solve mathematical puzzles
    - Erdős and Rényi provided formal mechanisms for generating random graphs that made statistical tests of network properties viable
  - Sociologists focused on patterns of social ties (as opposed to the study of individuals)
    - applied formal mathematical methods to describe, analyze, and visualize networks in what was then described using terms such as “sociometrics” and “sociograms”
    - Milgram’s famous “six degrees of separation” study involving chain letters
    - Mark Granovetter demonstrated the value of a social network approach by showing that “weak ties” were a much better source of new jobs than “strong ties”

# History

- The Computational Phase (1970's - mid 1990's)
  - Creation and systematic use of computational tools and methods
  - Leveraged the new capabilities of computers to analyze and visualize networks
  - Homans developed new techniques for identifying subgroups (i.e., clusters) in networks, while White developed techniques for finding people that occupy similar network positions (via “structural equivalence”)

Founding SNA Sociologist Barry Wellman argued “*SNA is not simply a method but is the core paradigm for explaining social action*”

# History

- Network Data Deluge Phase (Mid 1990's to Present)
  - wealth of real-time social network data is captured by our everyday use
  - SNA no longer purely academic
    - corporations, governments, and NPOs utilize SNA techniques to find criminals, rank Web sites, recommend books, identify influencers, restructure organizations.
    - analysis of social networks at a scale never before possible

# Human Computer Interaction: SNA Goals

1. Inform the design and implementation of new Computer-Supported Cooperative Work (CSCW) systems.
  - Characterize the social structure of a population of intended users
    - clarify requirements and challenges, better initial designs
    - Identify individuals with unique/important network positions
2. Understand and improve current CSCW systems.
  - Data from existing systems shows how current features are utilized by users
    - For example, “unfollowing” someone on Twitter partly explained by social network structures
  - Help community managers understand what is happening in large scale communities
    - Allow designers to develop tools that meet the particular needs of subpopulations

# Human Computer Interaction: SNA Goals

3. Evaluate the impact of CSCW system on social relationships.

- Many systems are designed to influence the social relationships of users
  - online exchange markets match buyers and sellers, corporate intranets help employees find internal experts
- Evaluation can be performed to assess the impact of a specific feature or social intervention.
  - online “icebreaker” activity assessed by looking at changes in the network

4. Design novel CSCW systems and features using SNA methods.

- Allows input to new systems and features
  - Tool that recommends potential friends on a social networking site uses SNA properties
  - Tools leverage network analysis and visualization to help gain insights into large datasets

# Human Computer Interaction: SNA Goals

5. Answer fundamental social science questions.

- Growing field of “computational social science”
  - Test hypotheses and theories at a much larger scale
    - Study of Facebook helped support and extend Granovetter’s original work that showed the importance of “weak ties”
    - Reducing the need for raw or self-reported data collection

# Social Network Analysis Questions

- Questions About Individual Social Actors
  - Identifying individuals who play an important or unique role within a particular social network
    - Who is most popular?
    - Who has the most influence?
    - Who is a bridge spanner?
- Questions About Overall Network Structure
  - Focus on overall distribution instead of focusing on the position of individuals
    - How interconnected are a group?
    - What is the distribution of individual network properties or social roles?
    - Are there subgroups of highly connected users ?
- Questions About Network Dynamics and Flows
  - How networks change over time
    - How do the structures of social relationship change?
    - How does the importance of specific individuals, social roles, or clusters change?
    - How does information spread?

# Performing Social Network Analysis

- Identify Goals and Research Questions

It is essential that analysts hone in on a few critical goals and turn them into specific research questions, lest they spend unreasonable amounts of time aimlessly meandering around the data.

# Performing Social Network Analysis

- Collect Data
  - Sources of Network Data

Data Source	Effort level
Raw data from system usage	Medium - High
Network survey	High
Application programming interfaces	Medium - High
Screen scraping	Medium - High
Network analysis importer tools	Easy
Existing datasets	Easy

# Performing Social Network Analysis

- Collect Data
  - Sources of Network Data
  - Types of Social Networks
    - Directed vs Undirected
    - Weighted vs Unweighted
    - Multiplex Networks
    - Unimodal vs Multimodal
    - Partial Networks

# Performing Social Network Analysis

- Collect Data
  - Sources of Network Data
  - Types of Social Networks
  - Representing Network Data

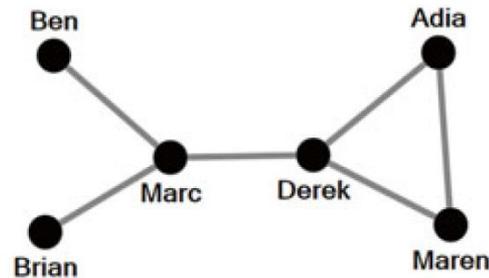
## Edge List

Node1	Node2
Brian	Marc
Derek	Marc
Derek	Adia
Marc	Ben
Maren	Derek
Maren	Adia

## Matrix

	Brian	Marc	Derek	Adia	Ben	Maren
Brian	0	1	0	0	0	0
Marc	1	0	1	0	1	0
Derek	0	1	0	1	0	1
Adia	0	0	1	0	0	1
Ben	0	1	0	0	0	0
Maren	0	0	1	1	0	0

## Graph



**Fig. 2** Three ways of representing network data

# Performing Social Network Analysis

- Analyze and Visualize Data
  - Node- Specific Metrics
    - Centrality
    - Identify structural significance of individual nodes
  - Aggregate Network Metrics
    - Density, Diameter, Average geodesic distance, Network centrality
    - Characterize the entire network
  - Network Clusters and Motifs
    - Detect small network component
  - Network Dynamics
    - Information flow
    - Network structure change
  - Network Visualization

# What Constitutes Good Work

- Use appropriate network metrics
- Do not claim more than what your data can support
- Use network virtualization that illustrates the core points
- Use appropriate statistical techniques to compare to baseline model
- Look at exemplary work for appropriate methods and techniques

# Example Research Applications

## The Life & Death of Online Gaming Communities: A Look at Guilds in World of Warcraft

- Study examined factors that could explain the success or failure of a game guild
  - Guilds vary in size and complexity but overall trend is that many are socially fragile and don't last long
  - Researchers interested in the structural properties of guilds specifically
- Methods
  - Use of quantitative data for social science research was backed by researchers experience with WoW
  - Used the open interface of game to directly collect data with “/who” command
    - Continuously collected data of which characters were playing, where, with who
    - Gave access to the following Max subgraph size analysis
      - Density
      - Centrality
      - Level
      - Average time spent together
      - Class
      - Balance

# Example Research Applications

Results:

Comparison of 2 month-long samples of guilds. Logistic regression model was fit with survival of guild as dependent variable, against predictors mentioned earlier.

Significant factors, ranked by Wald test:

1. Class Balance Ratio
2. Guild Size
3. Level Standard Deviation
4. Max Subgraph Size
5. Time in Instances
6. Density

	B	S.E.	Wald	df	Sig.	Exp(B)
@07_guildsize	.054	.011	24.481	1	.000	1.055
@07_max_subgraph_size	-.048	.012	15.845	1	.000	.953
@07_mass_count	.077	.109	.502	1	.479	1.081
@07_total_count	-.080	.072	1.237	1	.266	.923
@07_density	3.586	.916	15.339	1	.000	36.081
@07_centrality	-1.527	.889	2.948	1	.086	.217
@07_LvlAvg	.021	.006	13.935	1	.000	1.021
@07_LvlMedian	-.002	.003	.389	1	.533	.998
@07_LvlSD	.046	.010	23.283	1	.000	1.048
@07_Count60	-.022	.008	7.143	1	.008	.978
@07_AvgTogether	.090	.074	1.466	1	.226	1.094
@07TogetherRatio	-1.243	.377	10.887	1	.001	.288
@07_InstanceTimeRatio	7.011	1.782	15.481	1	.000	1109.076
@07_ClassBalance	.003	.005	.423	1	.515	1.003
@07_BalanceRatio	-.288	.054	28.135	1	.000	.750
Constant	-1.541	.195	62.413	1	.000	.214

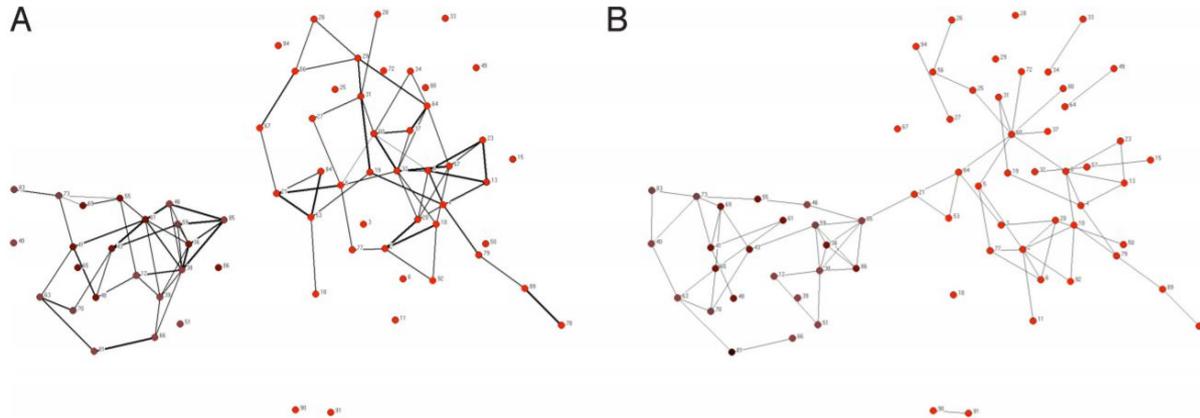
Regression Coefficients for Survival Model

# Example Research Applications

## Inferring friendship network structure by using mobile phone data

- Goal
  - demonstrate the power of collecting not only communication information but also location and proximity data from mobile phones over an extended period, and compare the resulting behavioral social network to self-reported relationships from the same group.
- Method
  - Observed 94 subjects using phones with software recording data over the period of nine months
  - Collected self-reported proximity data from participants about their relationships
  - Constructed social network graphs from observed data and self reported data

# Example Research Applications



**Fig. 5.** Inferred, weighted friendship network vs. reported, discrete friendship network. Frame *A* shows the inferred friendship network with edge weights corresponding to the factor scores for factor 2, extra-role communication. Frame *B* shows the reported friendship network. Node colors highlight the two groups of colleagues, first-year business school students (brown) and individuals working together in the same building (red).

The graphed networks overlap significantly, but are distinct in that long term relationships may not require constant proximity to exist, and self-reported data suffers from salience bias (how vivid an event was) and recency bias (how recent an event was)

# Example Research Applications - Twitter

## Goals

- How are people connected?
- What are the most influential people and topics
- How does information diffuse via retweet

## Data Collection

- Twitter API
  - User Profile
  - Trending topics
  - Tweets

# Example Research Applications -

## Analysis and Visualization

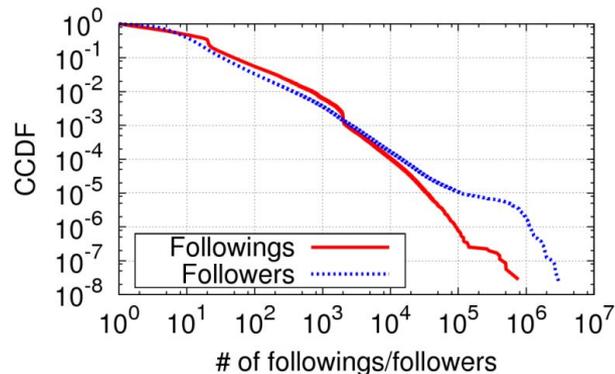


Figure 1: Number of followings and followers

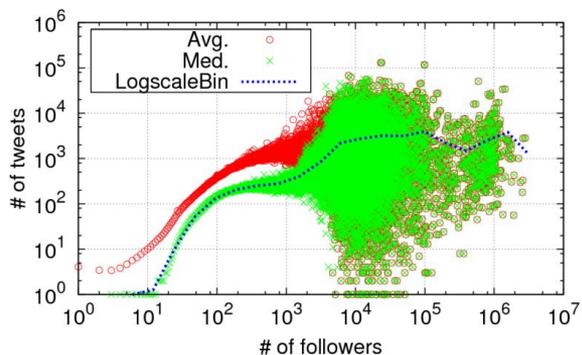


Figure 2: The number of followers and that of tweets per user

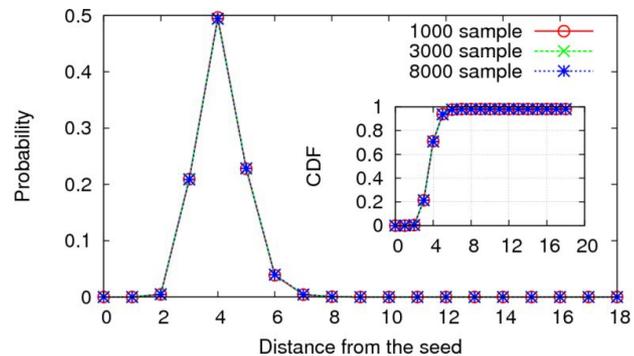


Figure 4: Degree of separation

# Looking Towards the Future

- With rise of Big Data as a field , SNA will draw from unprecedented amounts of information
  - Allows for more evidence for past SNA studies
  - Brings up new questions for SNA to explore, across research fields and disciplines
- SNA will continue to flourish as our social lives become increasingly mediated by technology

# References

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2007). The life and death of online gaming communities: a look at guilds in world of warcraft. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems , CHI'07 (pp. 839–848). New York, NY: ACM.

Eagle, N., Pentland, A. (. S.), & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences, 106 (36), 15274–15278.