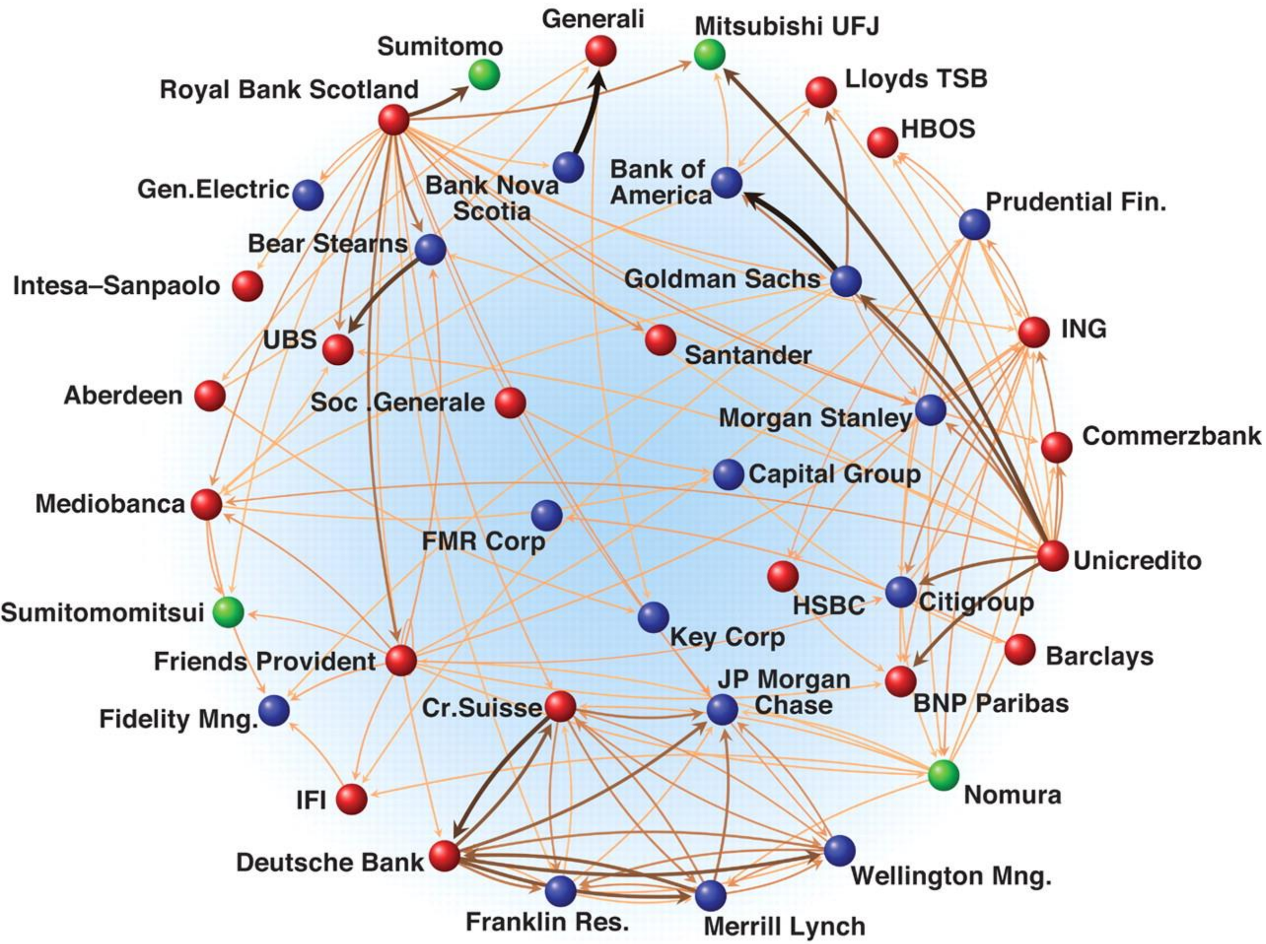
Business Network Analytics



F Schweitzer et al. Science 2009

Sep, 2017

Daning Hu
Department of Informatics
University of Zurich
Business Intelligence Research Group

Research Methods and Goals

What

- Social network analysis (Metrics)
- **Describe** the changes in network evolution
  - Temporal changes in network topological measures
- Dynamic network recovery
- (Relational) data mining

Why

- Econometric **identification** of casual Social and Economic influence
  - Distinguishing homophily
  - Confounding factors
  - PSM, DID, RD, etc.
  - Explanations

How

- **Combine** social science methods, data mining, machine learning with econometric analysis
- **Predict** link formation
- **Simulate** the evolution of networks

Causal Effects in Networks and Social Interactions

- The settings of interest may be in the magnitude of social interactions, or peer effects, that is
  - the effects of changing treatments for one unit on the outcomes of other linked units.
  - or all units in a subpopulation are linked and influence each other's outcome, e.g., classroom setting, (Manski 1993), roommates (Sacerdote 2001).

Identifying Social Influence in Online Social Networks*

- Previous research identifies “clustering” of behaviors in social networks and infers social influence from it.
  - Correlation of Observed Behaviors and Network Structure
  - Friends adoption of the behavior is correlated in time

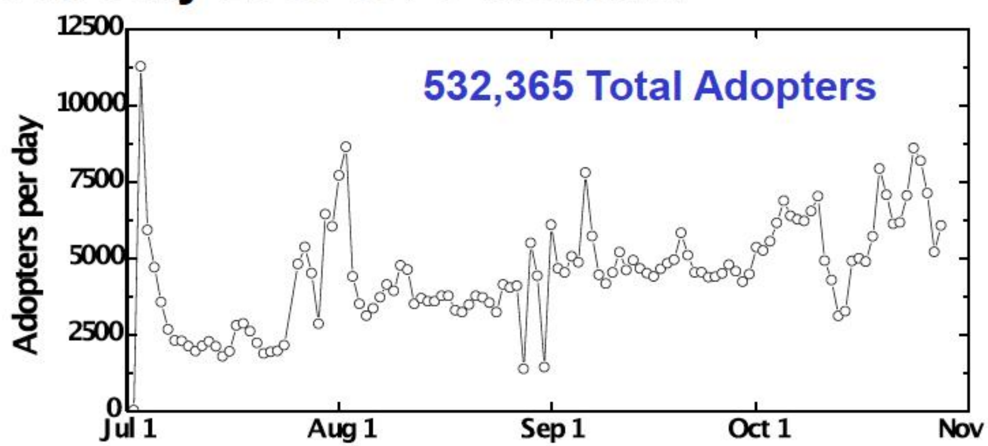
- How to identify social (peer) influence in social networks
  - A large stream of studies focused on distinguishing Influence Based contagion From Homophily driven diffusion in social networks
  - Science, Marketing Science, PNAS
  - Competing theory: Homophily - Birds of a feather, flock together.

- * Some of the contents are from Prof. Sinan Aral’s previous presentations.

Case I: Yahoo Study (Sinan Aral, PNAS)

- Global **IM Network of 27 Million Users** from Yahoo! (Daily Traffic)
- Detailed **demographics** and **geographic** data.
- Comprehensive, detailed and precise data on **online behaviors/activities**.

- Day by Day **adoption** and **usage** of a mobile service application (Yahoo Go) launched in July 2007 for 5 months.



Defining Social (Peer) Influence and Homophily

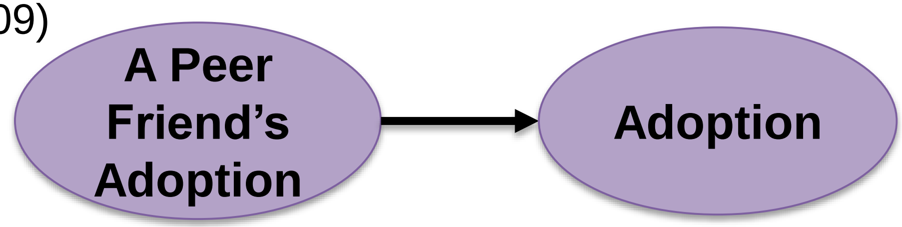
- **Peer Influence:**

Aral (2011) conceptualized peer influence based on the utility theory as "*how the behaviors of one's peers change the utility one expects to receive from engaging in a certain behavior and thus the likelihood that one will engage in that behavior.*"

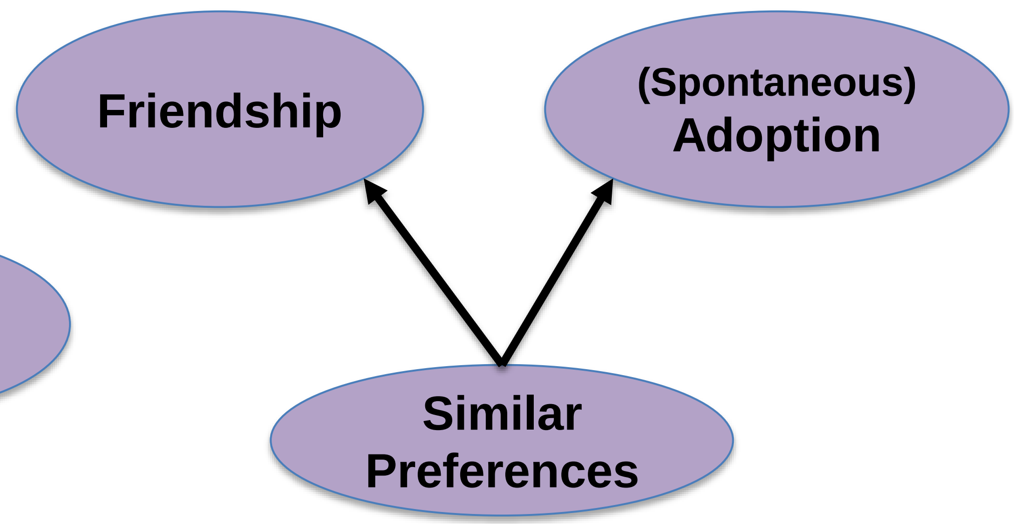
- **Homophily**
  - People who are alike tend to form social relationships with each others
  - Their shared characteristics may shape similar preferences and adoption behaviors

Social Mechanisms behind Correlated Adoptions

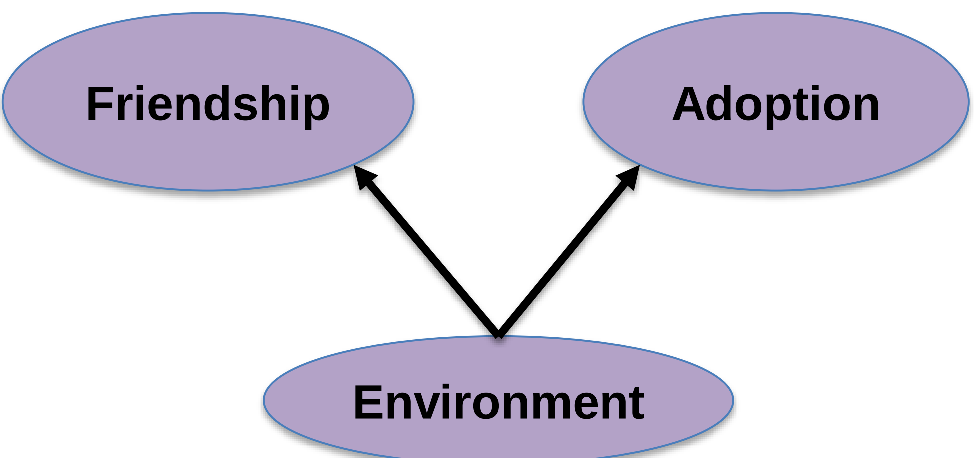
- **Social influence**-driven (correlated) adoption (Aral et al. 2009)
  - Software downloads (Duan et al. 2009)
  - Dinning choices (Cai et al. 2009)
  - Movie sales (Moretti 2011)
  - Facebook app (Aral and Walker 2011)



- **Homopihly (Preference)**-driven adoption
  - Like-minded people tend to become friends and choose similarly.



- Other Confounding factors



- How to **distinguish** them?

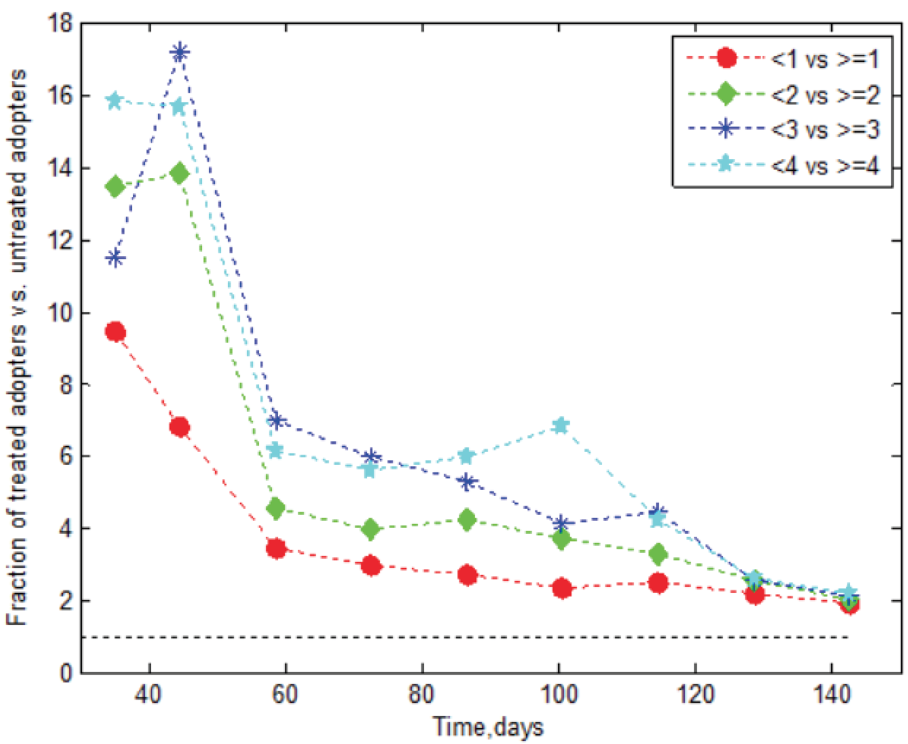
Demographic Data and User Online Activity

Demographic data	
Gender	Self-reported gender of users.
Age	Self-reported age of users. Users below the age of 18 were excluded from the sample due to IRB requirements.
Primary country*	Observed daily. Refers to the country from which users accessed the portal most often.
Secondary country*	Observed daily. Refers to the country from which users accessed the portal second most often.
Mobile device†	Observed daily. The type of device most frequently used by the user to access Yahoo! services from a mobile platform. Includes 2,030 unique devices.
Go device‡	Observed daily. The type of device most frequently used by the user to operate Yahoo! Go software. Includes 111 unique devices.
IM network data	
Number of messages	Observed daily. Number of messages sent to and received from each Yahoo! Messenger contact.
Online activity and browsing behavior§	
Total page views (PV)	Total number of Web pages viewed on Yahoo! websites.
Front page PV	Total number of front page Web pages viewed on Yahoo! websites.
News PV	Total number of news-related Web pages viewed on Yahoo! websites.
Finance PV	Total number of finance-related Web pages viewed on Yahoo! websites.
Sports PV	Total number of sports-related Web pages viewed on Yahoo! websites.
Weather PV	Total number of weather-related Web pages viewed on Yahoo! websites.
Search PV	Total number of search-related Web pages viewed on Yahoo! websites.
Flickr (Photo-sharing) PV	Total number of Flickr (photo-sharing) Web pages viewed on Yahoo! websites.
e-mail PV	Total number of e-mail-related Web pages viewed on Yahoo! websites.

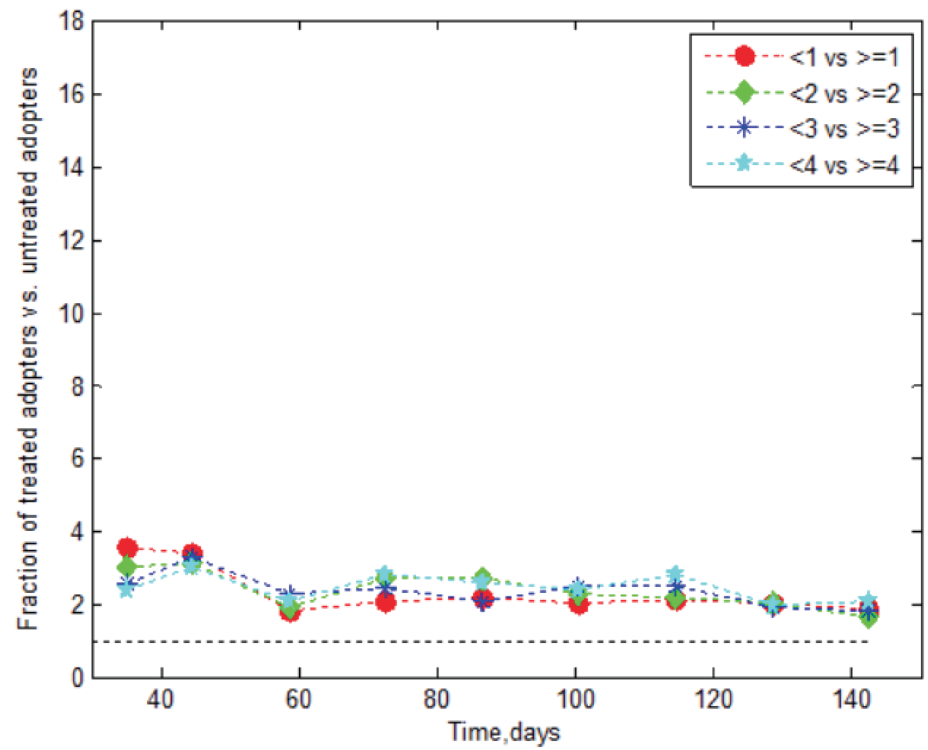
- * Some of the contents are from Prof. Sinan Aral’s previous presentations.

Distinguish Peer Influence from Homphily: Matching

"Influence" Estimates Comparing Adoption in Treated and Untreated Cases Under ***Randomized Matching*** Over Time (Methods used by those who take AM as evidence of influence)



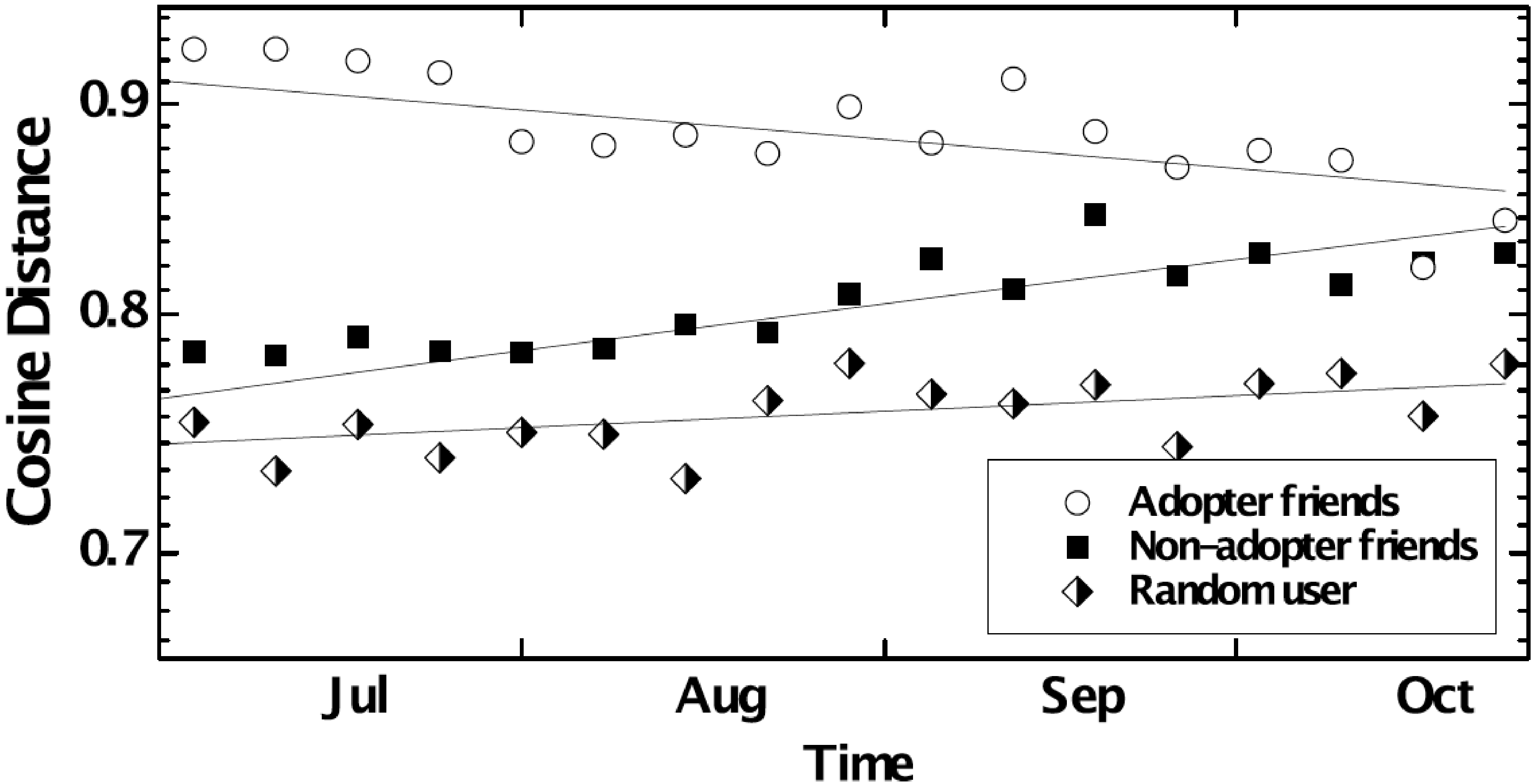
"Influence" Estimates Comparing Adoption in Treated and Untreated Cases In Our ***Dynamic Matched Sampling Framework*** Over Time



- Much of the estimated influence is really observable homophily

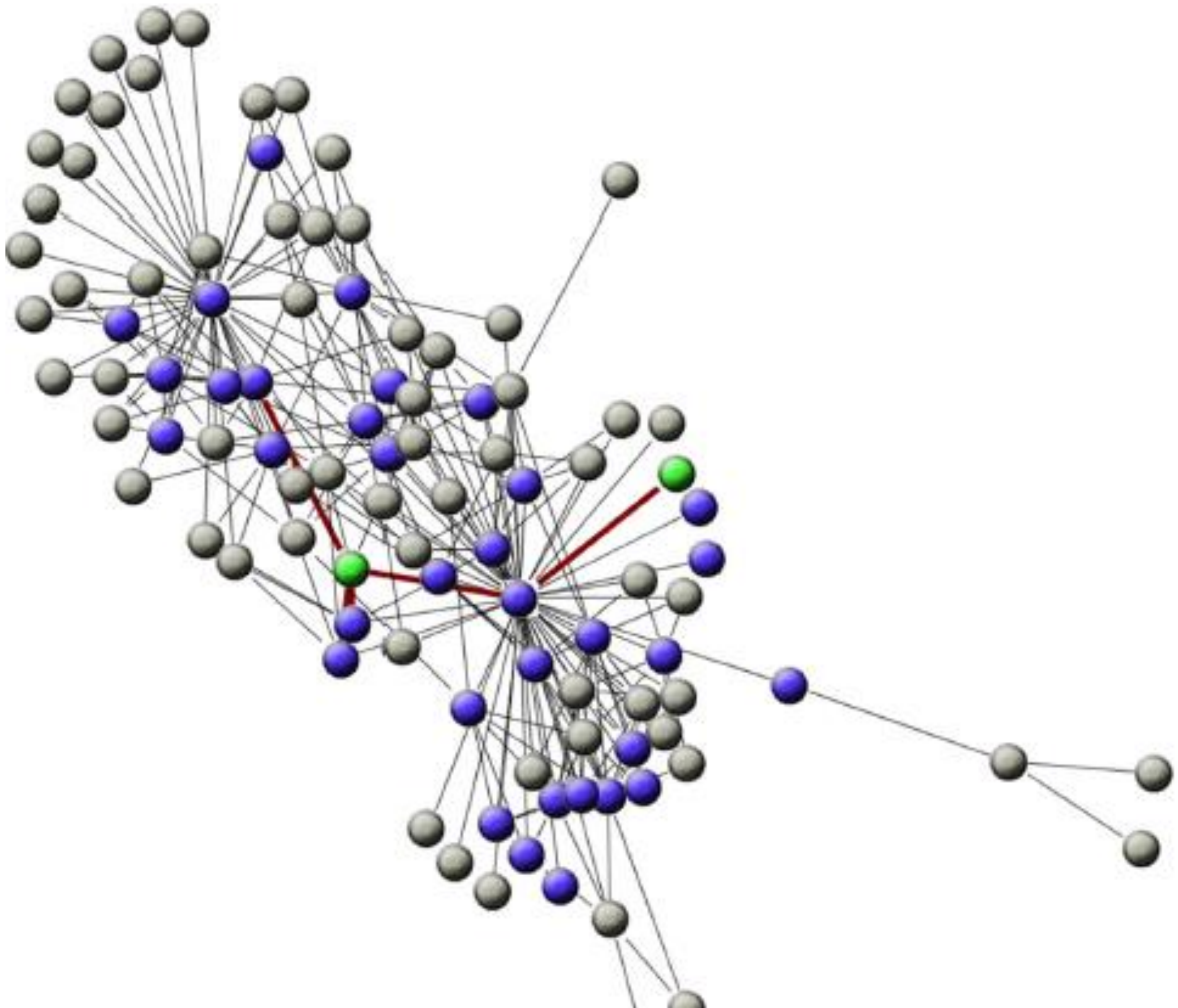
Exaggerated Homophily Amongst Early Adopters*

Cosine Distances Of Vectors of Observable Demographic, Geographic and Behavioral Data



The iPhone Effect

Dynamic Network Recover: “Snowball” Sampling of Yahoo GO Service Users



http://www.pnas.org/content/suppl/2009/12/10/0908800106.DCSupplemental/SM1.avi

Findings

- Decisions tend to cluster in network space and in time.

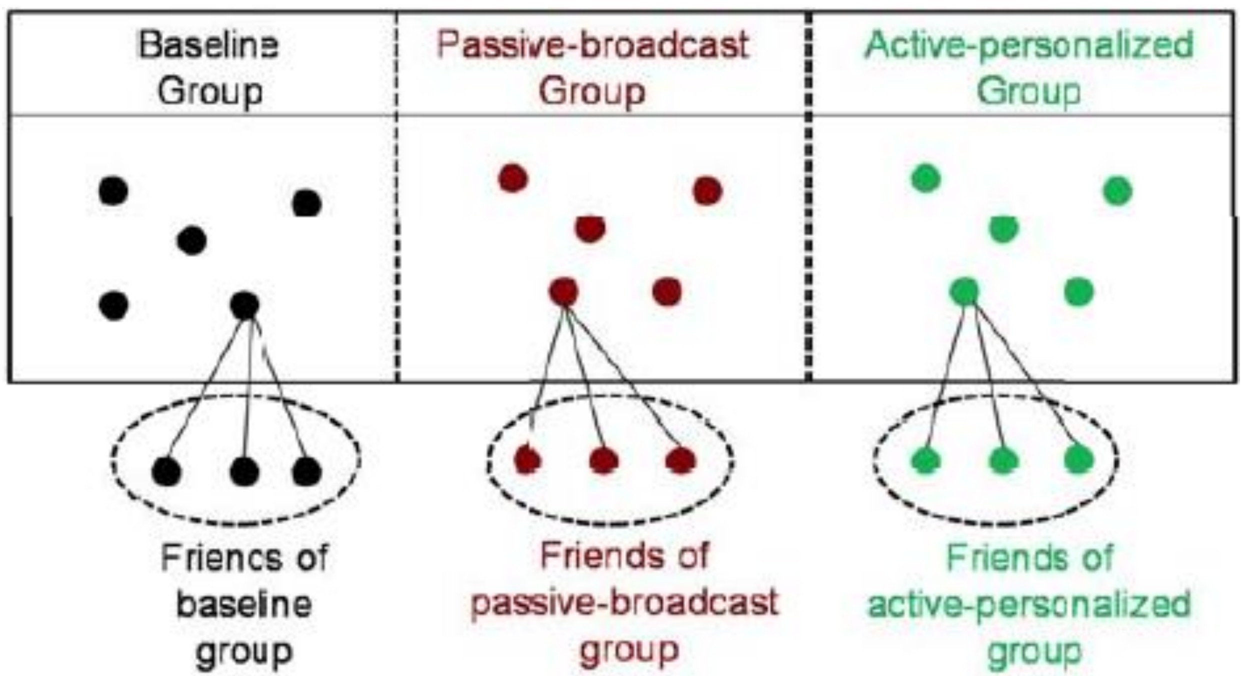
- Clustering may be caused by: Influence, Homophily, & Confounding Factors.

- Homophily is to a large extent responsible for what seems at first to be a contagious process (peer influence at work).
  - Implications for Policies (Marketing, Organizations, Social Policy)

- **But how about heterogeneity in products?**

Case II: Facebook Study (Sinan Aral, Management Science 2010)

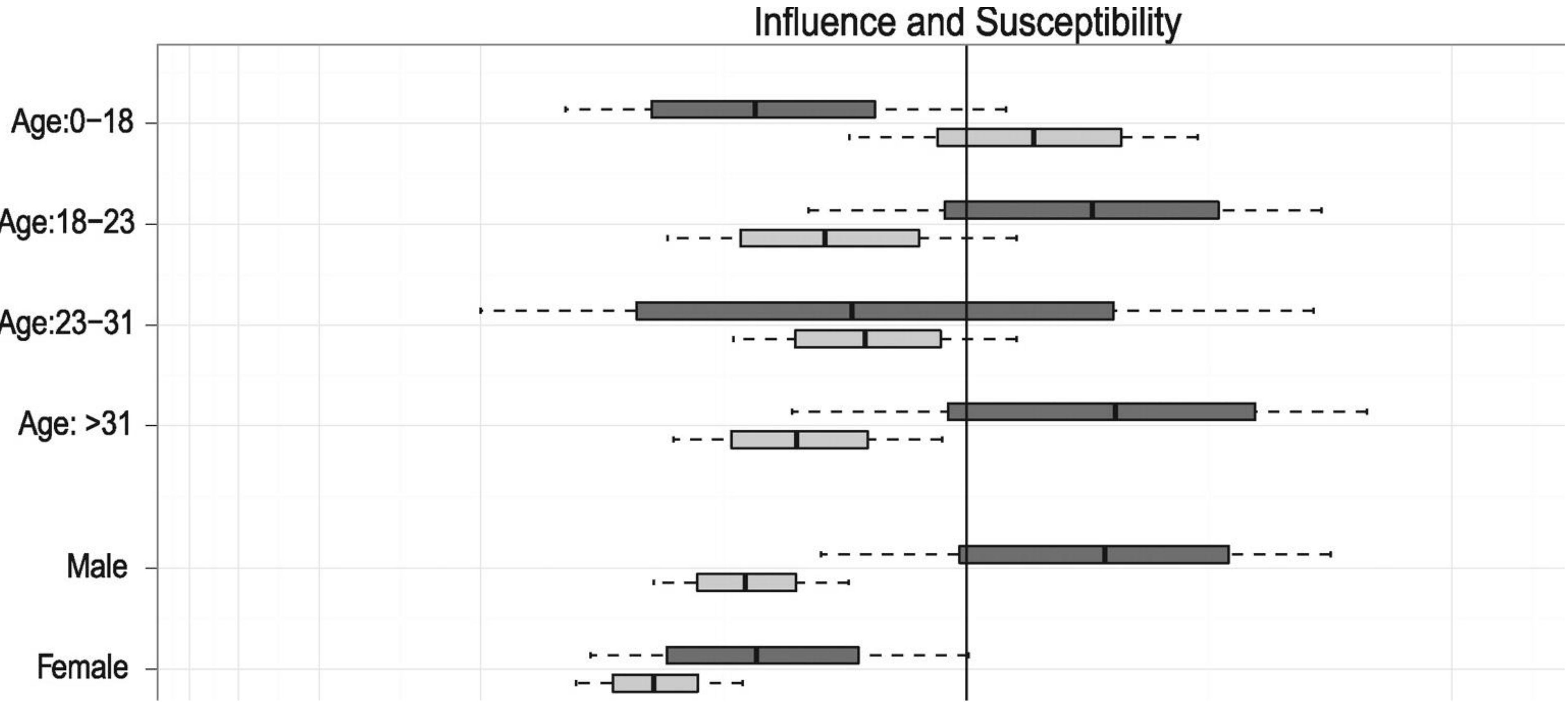
- 10 K Experimental Users
- 1.4 M Friends of Experimental Users
- They Observe application diffusion over this network
  - Facebook profiles
  - Adoption
  - Use



Which Features Spread Influence/Contagion Best?

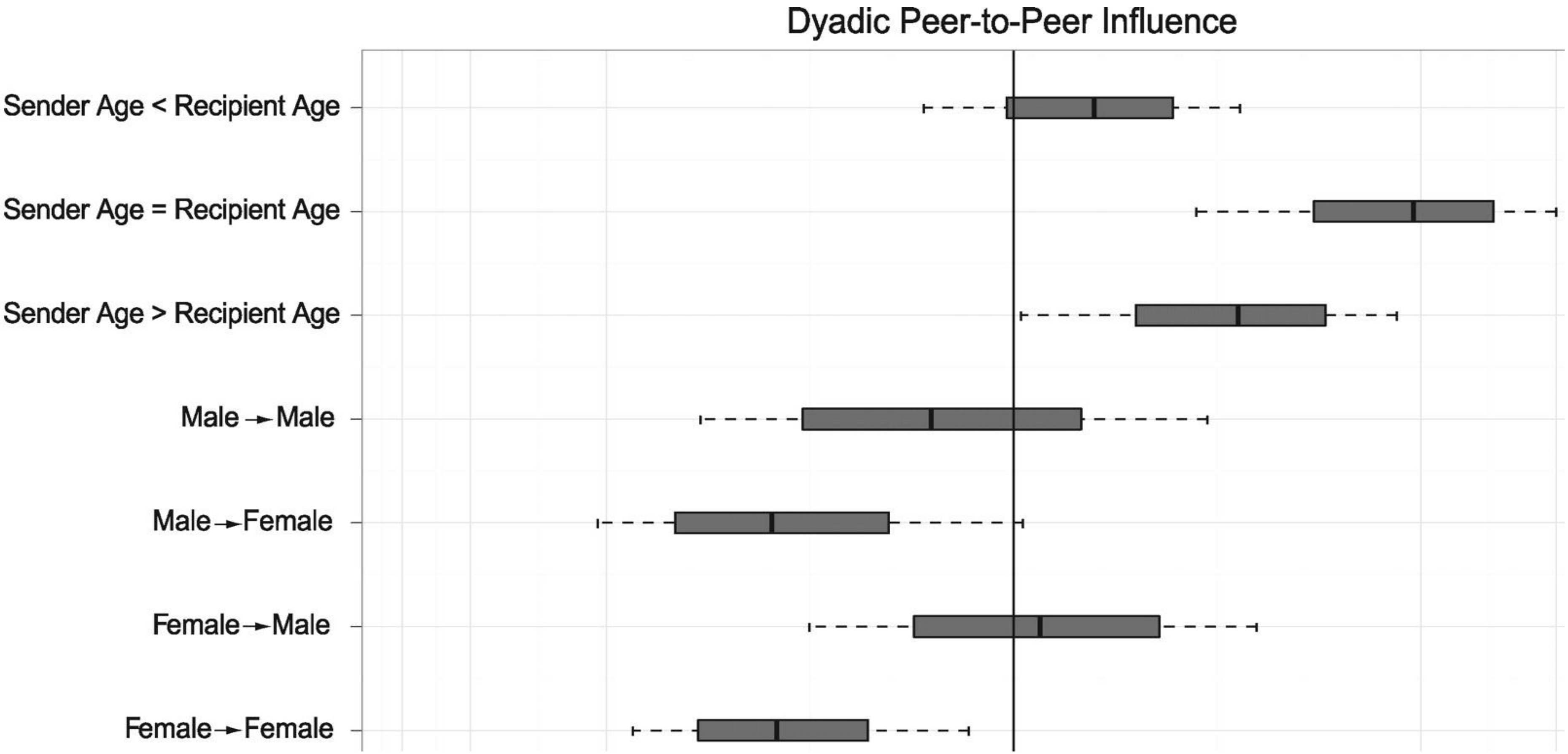
	Personal Invitations	Passive Awareness
Influence Per Message	↑ 6%	↑ 2%
Global Diffusion	↑ 98%	↑ 246%
Stickiness	↑ 17%	↑ N/A

Case III: Identifying Influential and Susceptible Members of Social Networks (Sinan Science 2012)



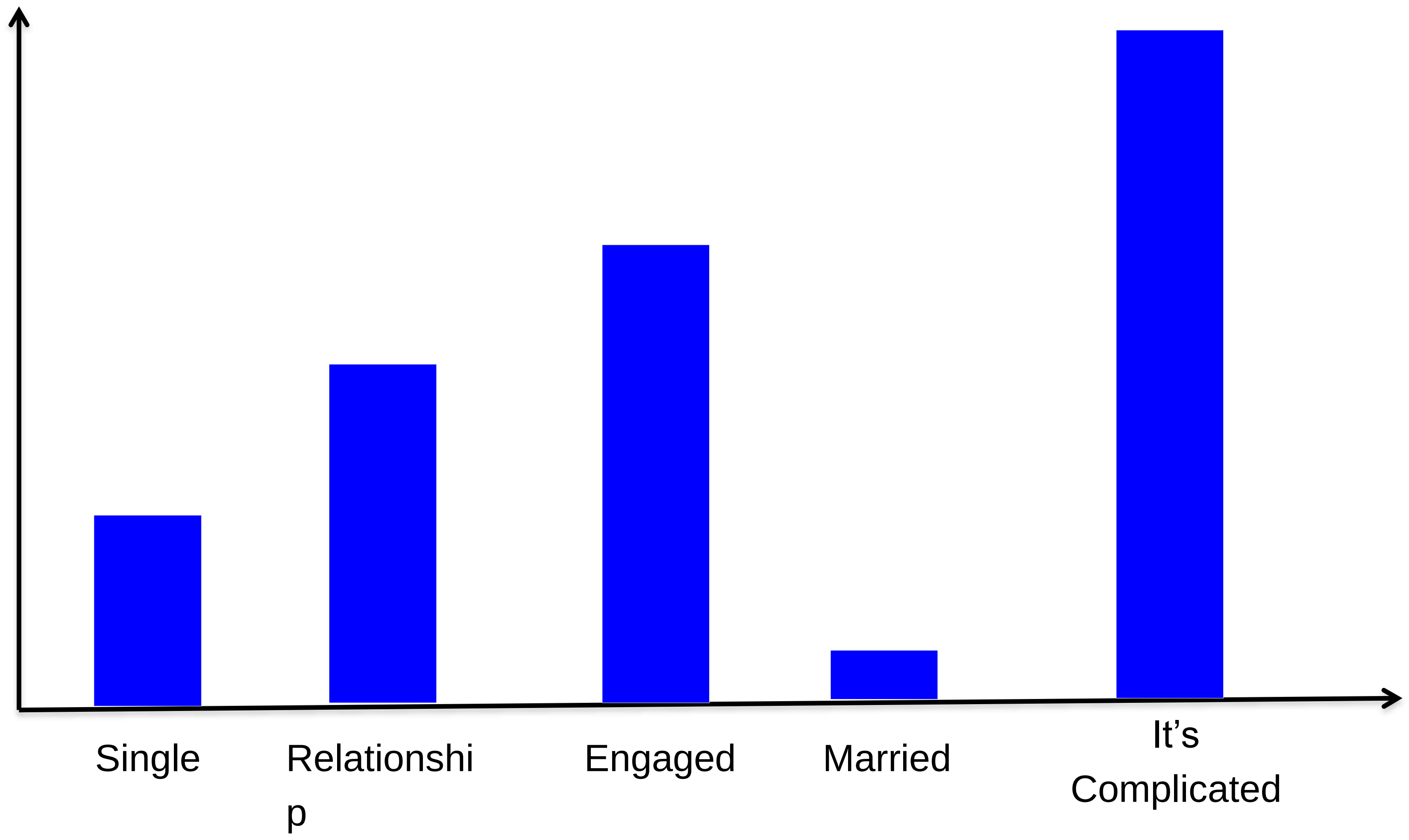
- Influence increases with age.
- Susceptibility decreases with age.
- Women are less susceptible to influence than men.

Dyadic Influence Models involving Age, Gender, and Relationship Status



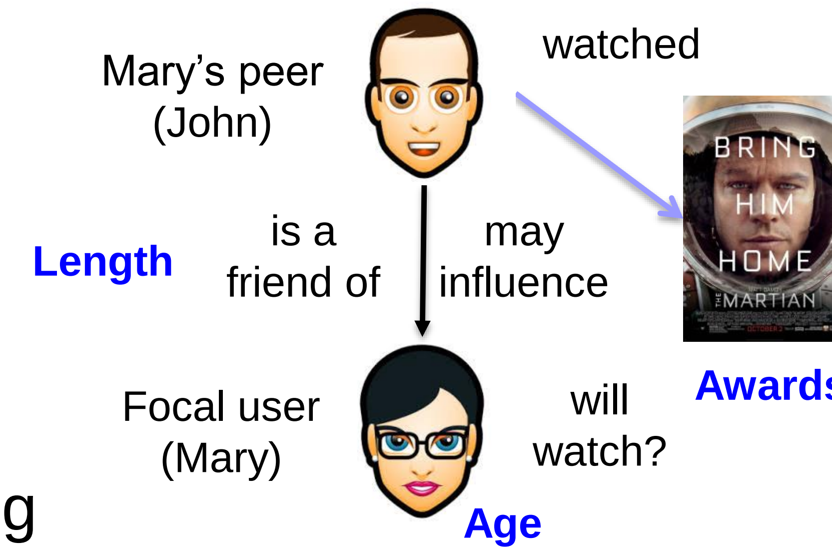
- Influence transmits over relationship pairs of the same age
- Order people influence younger people more than the other way around.
- Married to Single, Relation to Single

Facebook Users' Susceptibility to Influence*



Case IV: Heterogeneity in Product Diffusion (Daning Hu, ICIS 2016)

- How to study the heterogeneity in product diffusions trough peer influence in social networks?



- In the Big Data era, firms are having
  - Population-scale, micro-level digitized data
  - **Influencer Marketing:** "The Rise of Niche and Micro-Influencers"

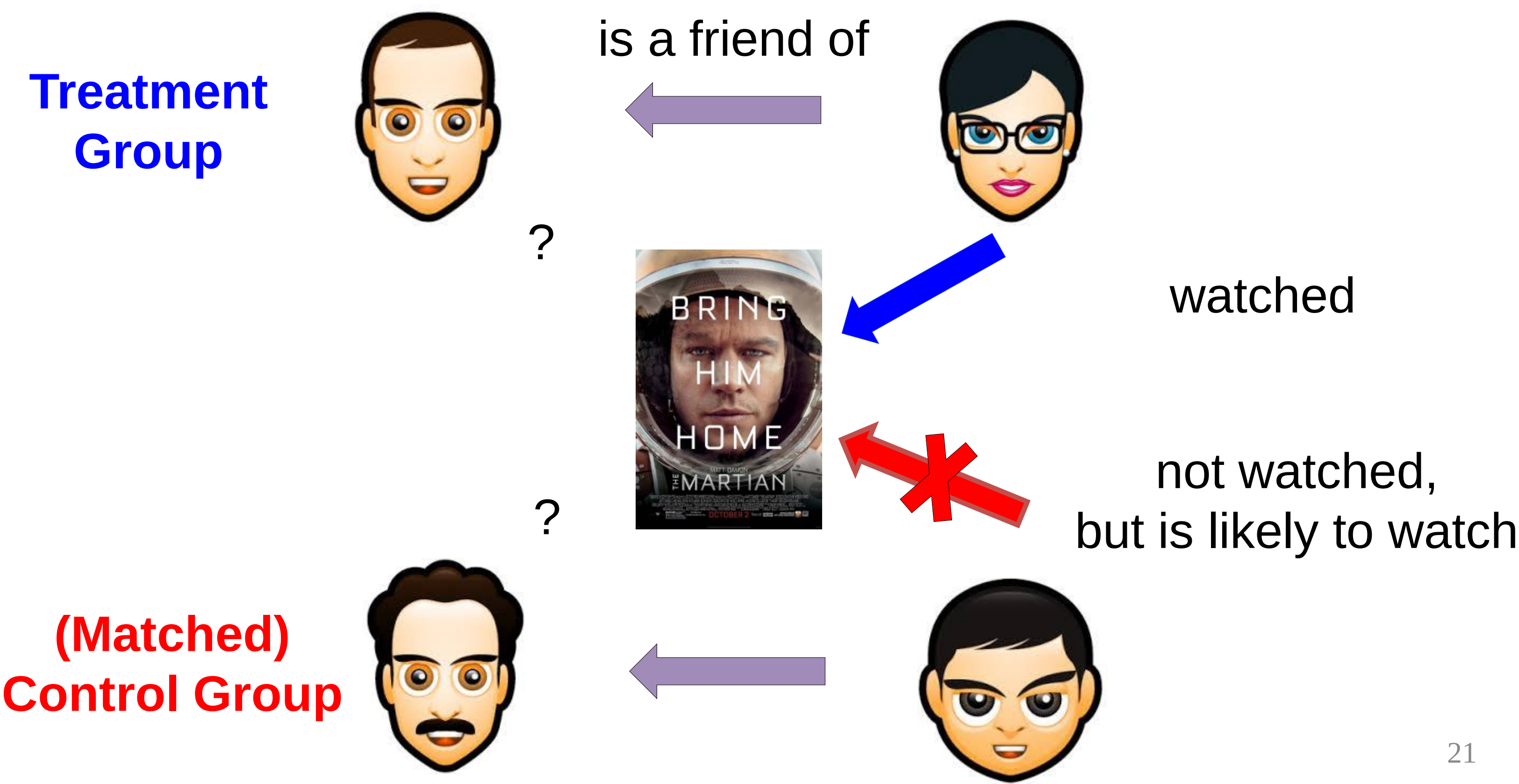
- It's critical to understand how to help promote the diffusion of various types of products besides popular ones like iPhone.

Dataset

- Our raw data is from a major Swiss broadband cable company which provides, phone, Internet, TV, and VOD services, etc.
  - Demographics for more than **480,000** customers:
    - Gender, age, primary language, and anonymized account ID.
  - **360 million** customers' phone calls (2011-13):
    - Anonymized phone numbers, call time, duration, costs, etc.
  - **3.9 million** Video-on-Demand purchases:
    - movie/music/Pvideos title, purchase time, costs, etc.
  - Crawl information for more than **13,000** movies from IMDB
    - Genre, rating, awards, production, sales, etc.

Social (Peer) Influence Identification: Matching

- Identify Peer Influence (**Control Homophily: Matching**)
  - **Treatment group**: VOD users who have at least one user friend that watched a selected movie M.
  - **Control group**: VOD users who do **NOT** have a user friend that watched M but is very likely to in terms of observable characteristics.



Propensity Score Matching (PSM)

- To control for homophily, we use PSM to match customers' likelihood to have one or more friends who watched a selected movie.
  - For each selected time period t, we calculated p_{it}, the propensity for one to be treated, using a logistic regression with 33 covariates (Table 1):

$$p_{it} = P(T_{it} = 1 \mid X_{it}) = \frac{exp[\alpha_{it} + \beta_{it}X_{it} + \varepsilon_{it}]}{1 + exp[\alpha_{it} + \beta_{it}X_{it} + \varepsilon_{it}]}$$

Cid34, Age = 33, gender = male, …# of friends = 5, # of vod = 4, …

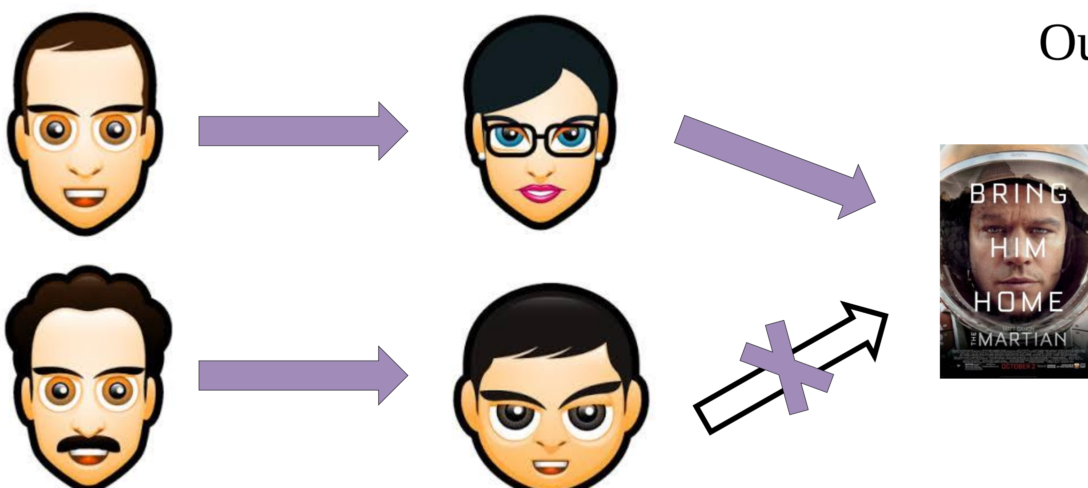
Cid98, Age = 32, gender = male, …# of friends = 4, # of vod = 4, …

Cid34

Cid98

Treatment Group

Control Group



Outcomes:

Treated Adopters N^+

Untreated Adopters N^-

Homopihly in Observable Demographic and Behaviors (1)

	Characteristic	**Detail (for each customer)**
Demographic	Gender	Self-reported gender (Coded as male 1, female 0)
	Age	Self-reported birth year of customer (from 1912 to 2012)
	Preferred contact language	Population-wide probability of purchasing the focal video for the corresponding preferred contact language
Phone call behavior	Number of friends (Degree)	Total number of the customers' friends (nodes that have at least one phone call relationship)
	Average number of outgoing calls per month	$= \frac{\text{Total number of outgoing calls}}{\text{Number of active months}}$
	Average number of incoming calls per month	$= \frac{\text{Total number of incoming calls}}{\text{Number of active months}}$
	Percentage (frequency) of outgoing calls	$= \frac{\text{Total number of outgoing calls}}{\text{Total number of outgoing and incoming calls}}$
	Percentage (duration) of outgoing calls	$= \frac{\text{Total number of outgoing calls}}{\text{Total duration of outgoing and incoming calls}}$
	Average duration per outgoing call	$= \frac{\text{Total duration of outgoing calls}}{\text{Total number of outgoing calls}}$
	Average duration per incoming call	$= \frac{\text{Total duration of incoming calls}}{\text{Total number of incoming calls}}$
	Average minutes to outgoing calls per friend	$= \frac{\text{Total minutes of outgoing calls in 2 years}}{\text{Total number of his friends}}$
VOD-related behavior	Number of purchased videos	Total number of videos this customer has purchased
	Average price per purchased video	$= \frac{\text{Total cost of the purchased videos}}{\text{Total number of his friends}}$

Homopihly in Observable Demographic and Behaviors (2)

	Number of videos watched per month	$= \frac{\text{Number of videos}}{\text{Number of active months}}$
	Average age of the purchased video	$= \frac{\text{Sum of the ages for all his purchased videos}}{\text{Number of videos}}$; the age is calculated as the number of days between the video release day and the day it was being watched
Average friends' demographic	Percentage of German speaking friends	$= \frac{\text{Number of friends with German contact language}}{\text{Number of friends}}$
	Percentage of French speaking friends	$= \frac{\text{Number of friends with French contact language}}{\text{Number of friends}}$
	Percentage of English speaking friends	$= \frac{\text{Number of friends with English contact language}}{\text{Number of friends}}$
	Percentage of Italian speaking friends	$= \frac{\text{Number of friends with Italian contact language}}{\text{Number of friends}}$
	Percentage of friends of same gender	$= \frac{\text{Number of male friends}}{\text{Number of friends}}$
	Average friends' gender	Average friends' gender
	Average friends' age	Average friends' age

Selected Results for PSM Analysis

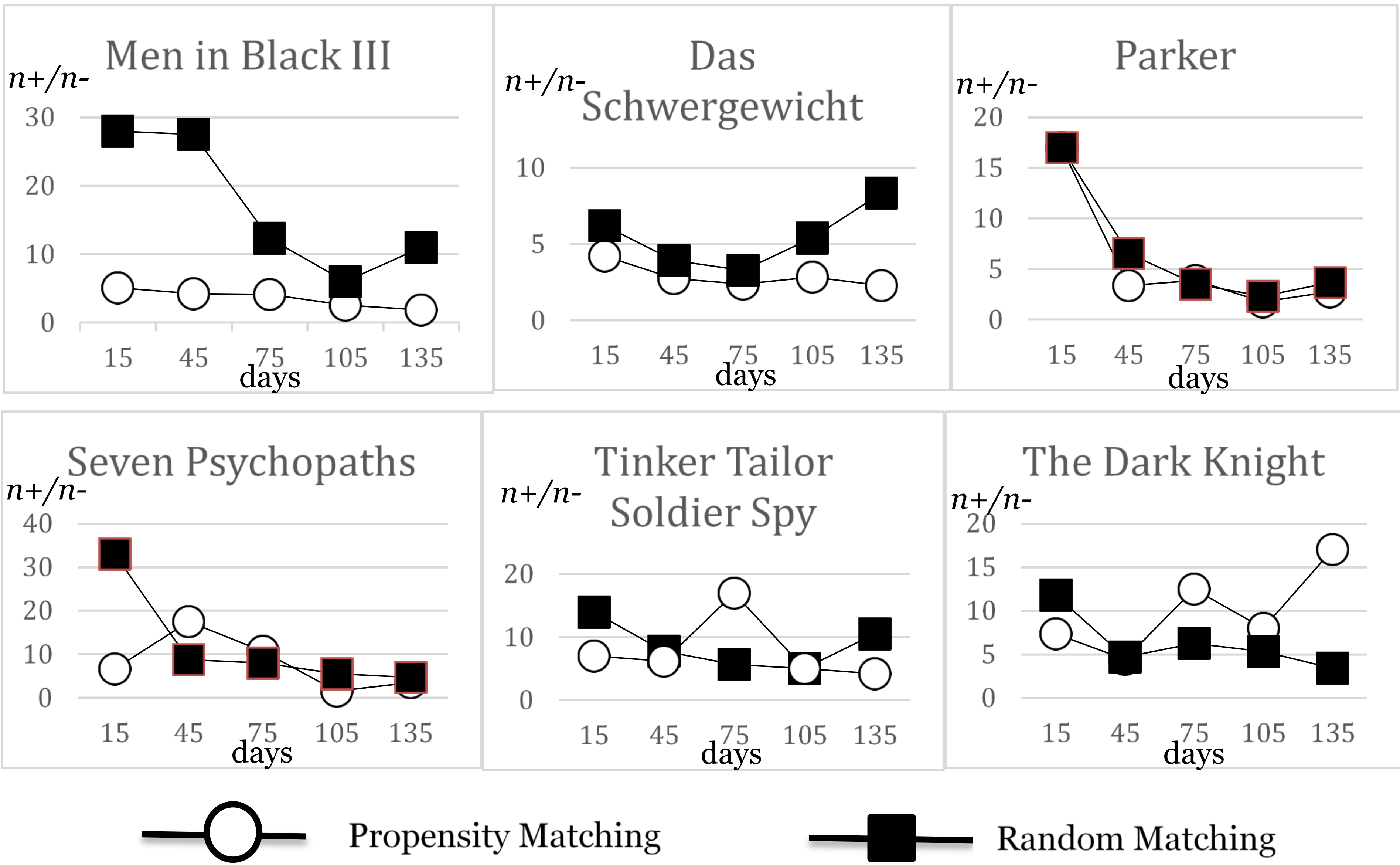


Figure 2. HDPSM Analysis Results for Selected VOD Movies

Research Methods and Goals

What

- Social network analysis (Metrics)
- **Describe** the changes in network evolution
  - Temporal changes in network topological measures
- Dynamic network recovery
- (Relational) data mining

Why

- Econometric **identification** of casual Social and Economic influence
  - Distinguishing homophily
  - Confounding factors
  - PSM, DID, RD, etc.
  - Explanations

How

- **Combine** social science methods, data mining, machine learning with econometric analysis
- **Predict** link formation
- **Simulate** the evolution of networks

Statistical Analysis of Determinants for Link Formation

- Proportional hazards model (Cox Regression Analysis)
  - $h(t, x_1, x_2, x_3 \ldots) = h_0(t) \exp(b_1 x_1 + b_2 x_2 + b_3 x_3 \ldots)$
  - Homophily in ***age*** *(group)* and ***race***
  - Shared affiliations:
    - ***Mutual acquaintances*** (through crimes)
    - ***Vehicle affiliation*** (same vehicle used by two in different crimes)

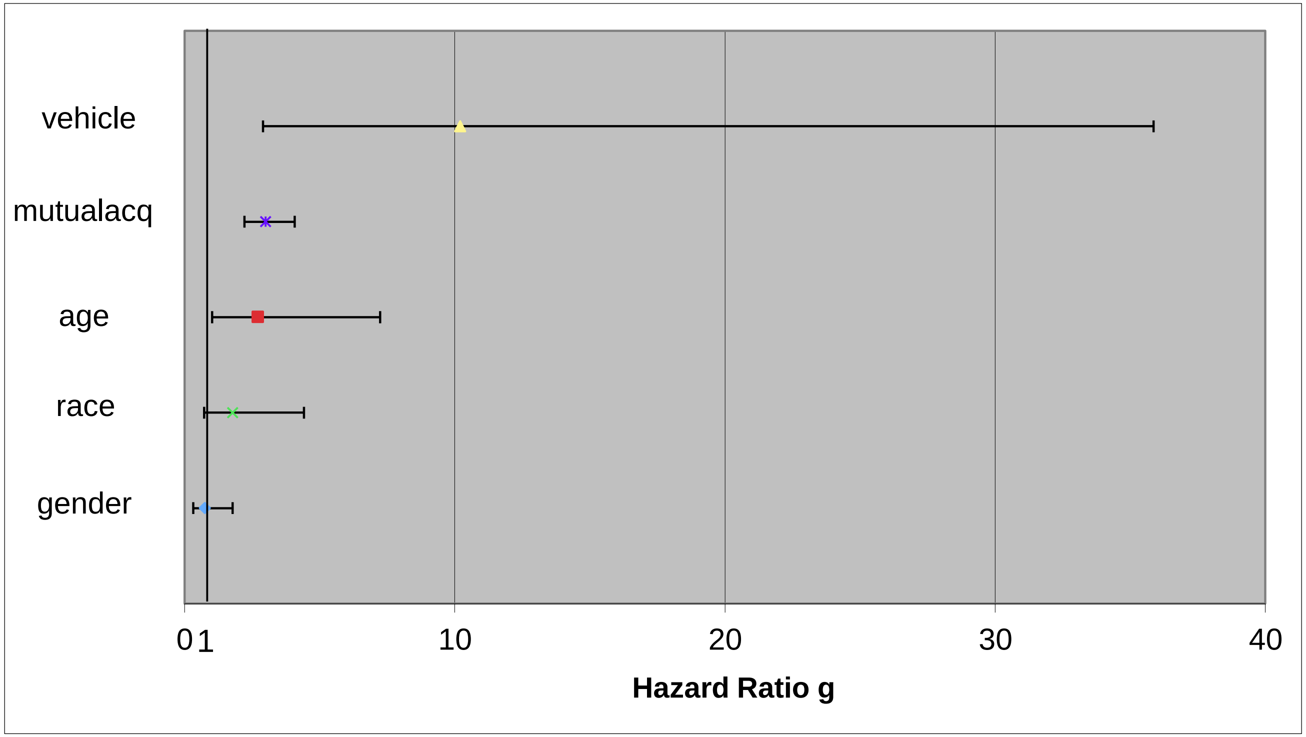
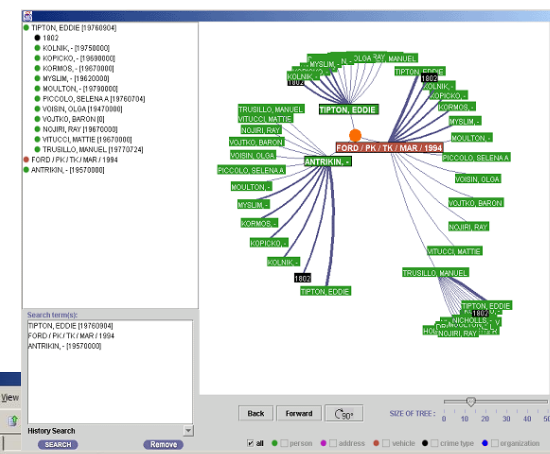


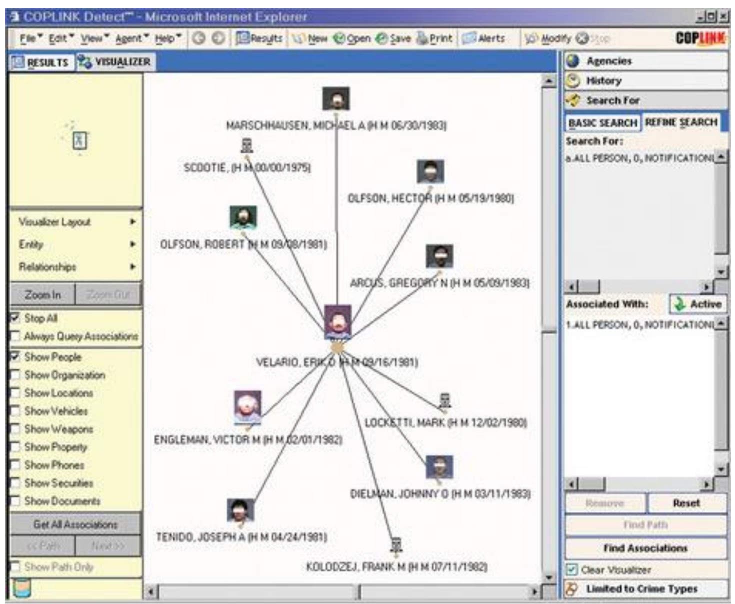
Fig.3. Results of multivariate survival (Cox regression) analysis of triadic closure (link formation).

BI Application: Co-offending Prediction in COPLINK

- IBM's COPLINK is an intelligent police information system aims to to help speed up the crime detection process.

- COPLINK calculates the co-offending likelihood score based on the *proportional hazards model* .
    - A ranked list of individuals based on their predicted likelihood of co-offending with the suspect under investigation.





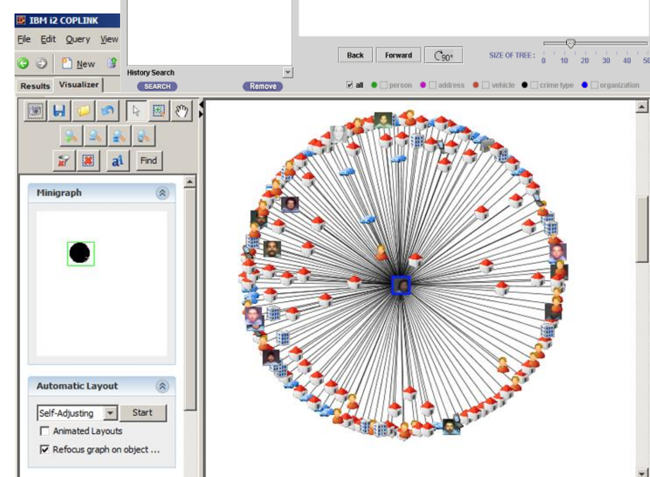


Fig.4. Screenshots of the COPLINK system

Research Methods and Goals

What

- Social network analysis (Metrics)
- **Describe** the changes in network evolution
  - Temporal changes in network topological measures
- Dynamic network recovery
- (Relational) data mining

Why

- Econometric **identification** of casual Social and Economic influence
  - Distinguishing homophily
  - Confounding factors
  - PSM, DID, RD, etc.
  - Explanations

How

- **Combine** social science methods, data mining, machine learning with econometric analysis
- **Predict** link formation
- **Simulate** the evolution of networks

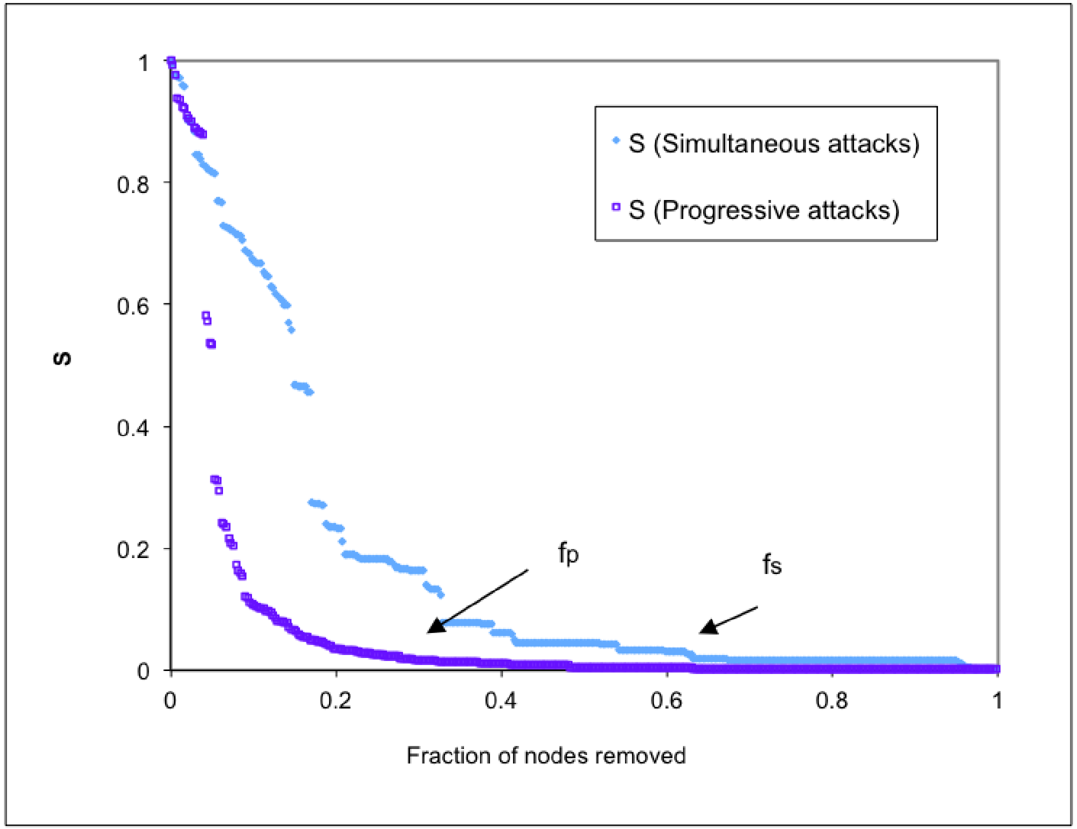
Simulate Attacks on Dark Networks

- Three attack (i.e. node removals) strategies:
  - Attack on hubs (highest degrees)
  - Attack on bridge (highest betweenness)
  - Real-world Attack (Attack order based on real-world data)

- Simulate two types of attacks to examine the robustness of the Dark networks
  - Simultaneous attacks (the degree/betweenness of nodes are **NOT** updated after each removal) – **Static**
  - Progressive attacks (the degree/betweenness of nodes are updated after each removal) – **Dynamic**

Simultaneous Vs. Progressive Attacks

- Both Dark networks are more vulnerable to ***progressive*** attacks than ***simultaneous*** attacks.
  - Dynamic updates are more effective



*** The relative size of the largest cluster** that remains connected: **S**

Hub Vs. Bridge Attacks

- Both hub and bridge attacks are far more effective than ***real-world*** arrests – Policy implications?

- Both Dark networks are more vulnerable to ***Bridge*** attacks than ***Hub*** attacks.
  - Bridge (highest beweenness): Field lieutenants, operational leaders, etc.
  - Hub (highest degree) : e.g., Bin Laden