Content and Causality in Influence Networks

ICWSM

July 20, 2011

@sinanaral

invitation to give a keynote at ICWSM in Barcelona July 2011 Inbox | X

Lada Adamic to Sinan — show details 9/30/10 — Reply

Dear Sinan,

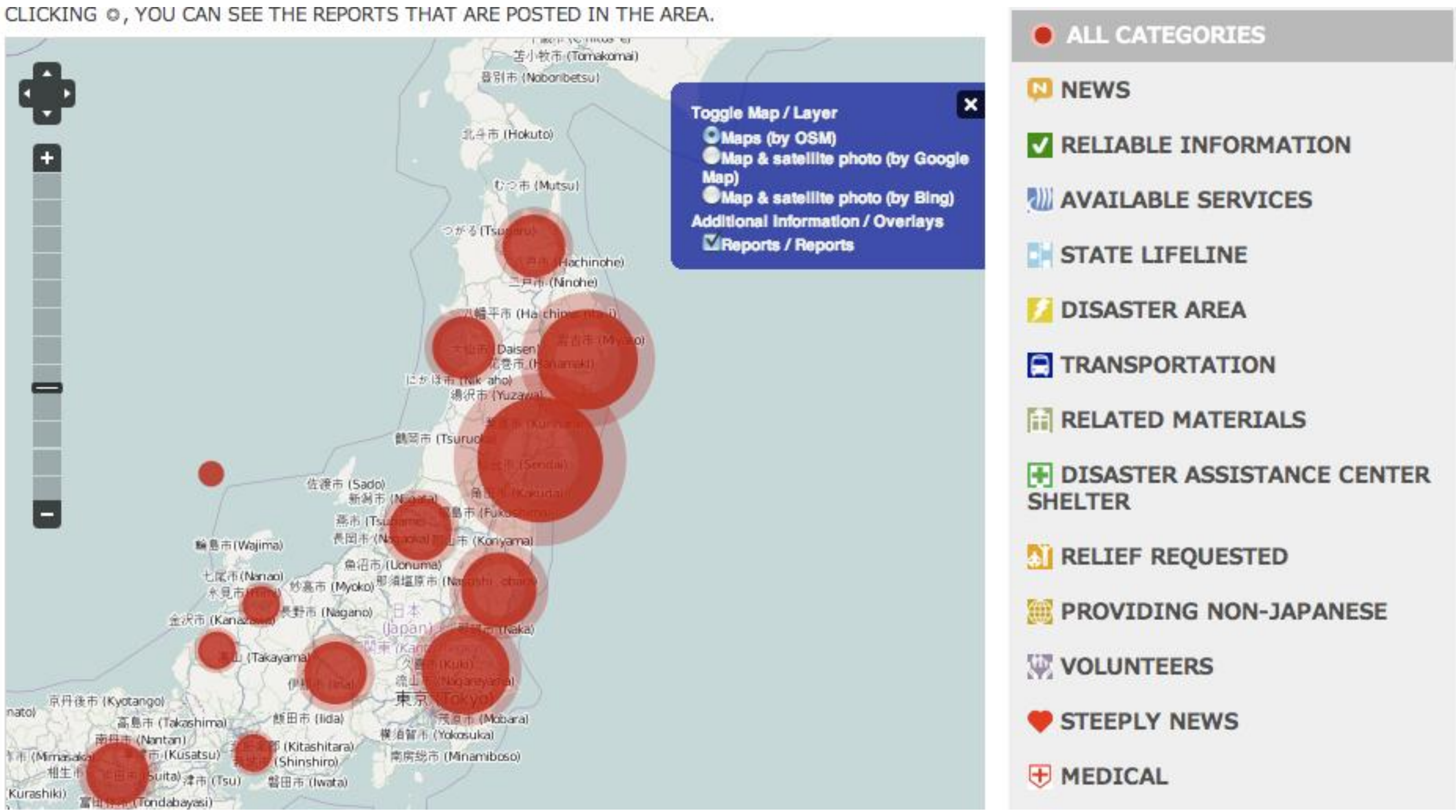
We would like to invite you to give a plenary keynote talk for 50-to-60 minutes at the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011). ICWSM 2011 will be held on July 17-20, 2011 in Barcelona (Spain) and will be collocated with IJCAI 2011. For more details see http://www.icwsm.org.

Sinan Aral wrote:

> pff - are you kidding? I'm in.
>
> :))!
>
> *Sinan Aral*
> NYU, Stern School of Business
> Affiliated Faculty, MIT.
> http://pages.stern.nyu.edu/~saral


BUY
TELL
COMMERCE
GETS
SOCIAL
SOCIAL
DRIVEN
HOW YOUR NETWORKS ARE DRIVING WHAT YOU BUY
DAVID ROWAN & TOM CHESHIRE
TIMBA SMITS
FEATURING YOUR FAVOURITE ONLINE SOCIAL NETWORKS
£
FACEBOOK • TWITTER • YOUTUBE • BLIPPY • FOURSQUARE • AMAZON • LINKEDIN • GO TRY IT ON • SHOPKICK • GROUPON


“This Revolution Will Be Tweeted.”


CLICKING ◎, YOU CAN SEE THE REPORTS THAT ARE POSTED IN THE AREA.
Toggle Map / Layer
Maps (by OSM)
Map & satellite photo (by Google Map)
Map & satellite photo (by Bing)
Additional Information / Overlays
Reports / Reports
苫小牧市 (Tomakomai)
登別市 (Noboribetsu)
北斗市 (Hokuto)
むつ市 (Mutsu)
二戸市 (Ninohe)
佐渡市 (Sado)
輪島市 (Wajima)
七尾市 (Nanao)
長野市 (Nagano)
飯田市 (Iida)
京丹後市 (Kyotango)
高島市 (Takashima)
南丹市 (Nantan)
横須賀市 (Yokosuka)
南房総市 (Minamiboso)
磐田市 (Iwata)
ALL CATEGORIES
NEWS
RELIABLE INFORMATION
AVAILABLE SERVICES
STATE LIFELINE
DISASTER AREA
TRANSPORTATION
RELATED MATERIALS
DISASTER ASSISTANCE CENTER SHELTER
RELIEF REQUESTED
PROVIDING NON-JAPANESE
VOLUNTEERS
STEEPLY NEWS
MEDICAL

KLOUT

hunch

Social Amp

rankur

kontagent

wefollow

Causality

Content

Causality

Content

Influence - Current Formalizations

@sinanaral

(e.g. in Studies of Twitter)

1. Influence = In Degree (Number of Followers)
2. Influence = f(Network Position) (e.g. Centrality, Betweenness)
3. Influence = Volume of Information Broadcast (Number of Tweets)
4. Influence = Breadth of Information Broadcasts (Tweets*Followers)
5. Influence = Breadth and Novelty of Information Broadcasts (Novel Tweets*Followers)
6. Influence = Breadth of Information Cascades (Tweets*Followers + Retweets*Followers of Followers and so on)
7. Influence = Triggers of Broad Novel Information Cascades (Novel Tweets*Followers + Novel Retweets*Followers of Followers)
8. Influence = Peer Behavior Adoption at time t + 1

NEW YORK UNIVERSITY NYU STERN LEONARD N. STERN SCHOOL OF BUSINESS

A Stricter Definition of "Influence"

@sinanaral

Social Influence: *how the behaviors of one's peers change the likelihood that (or extent to which) one engages in a behavior.*

Aral, S. (2011) "Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science*; 30(2): 217-223.

Implications of this Definition

1. Social Influence is about causal behavior change:

   How peer behaviors change the likelihood that or extent to which one will engage in the behavior.

2. Room for multiple social processes:

   Change in Utility Function or Change in Perception of the Behavior or Product… Persuasion or Awareness; Imitation or Social Learning.

3. Enables Consideration of Systems of Behavior

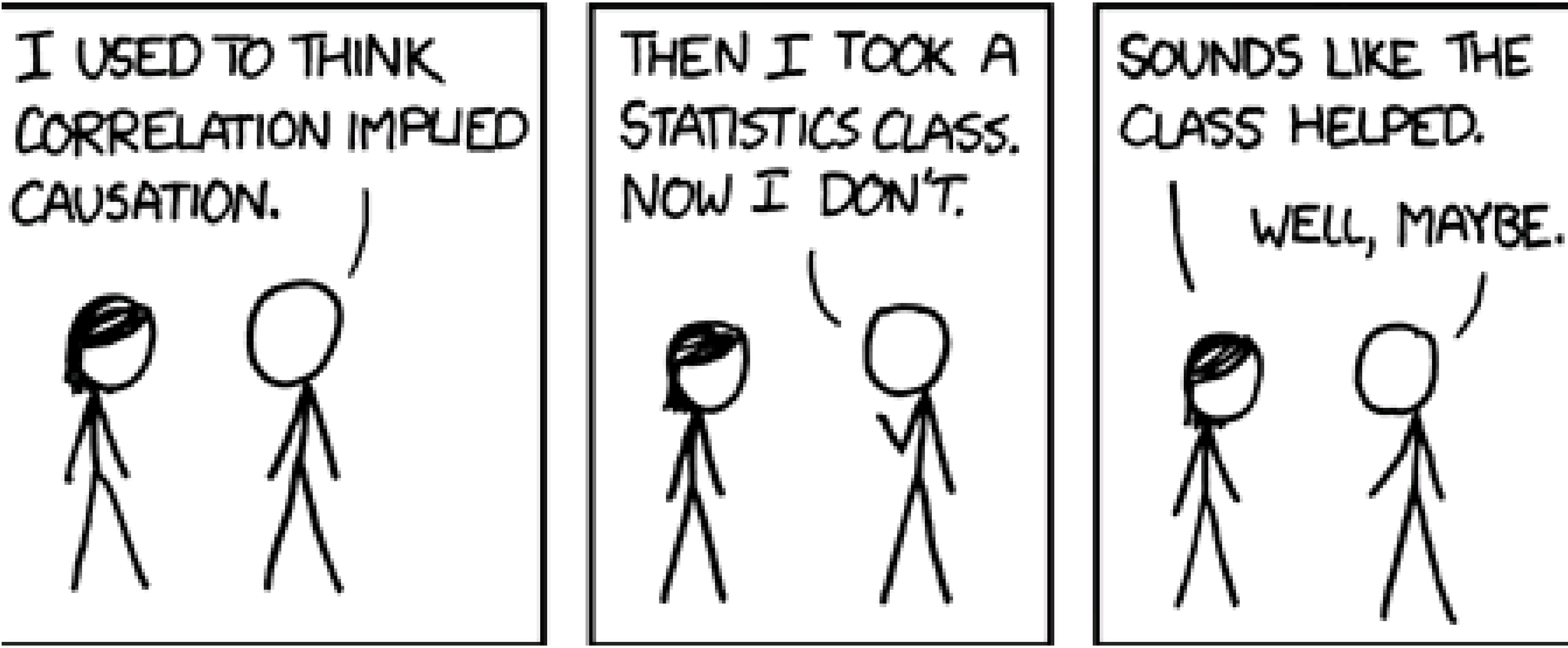
   Don't need activation on the focal behavior in question. Contrasts assumptions in marketing and innovation diffusion literature. Enables considerations of complementary behaviors.

4. Markovian assumptions can be strict or relaxed.

   Should influence processes be memoryless or favor recency? Cumulative effects vs. enthusiasm or excitement effects.

Causal Statistical Estimation

@sinanaral



Reflection and Endogeneity

Manski (1993)

@sinanaral

Identifying Causal Peer Effects in Networks is Notoriously Difficult

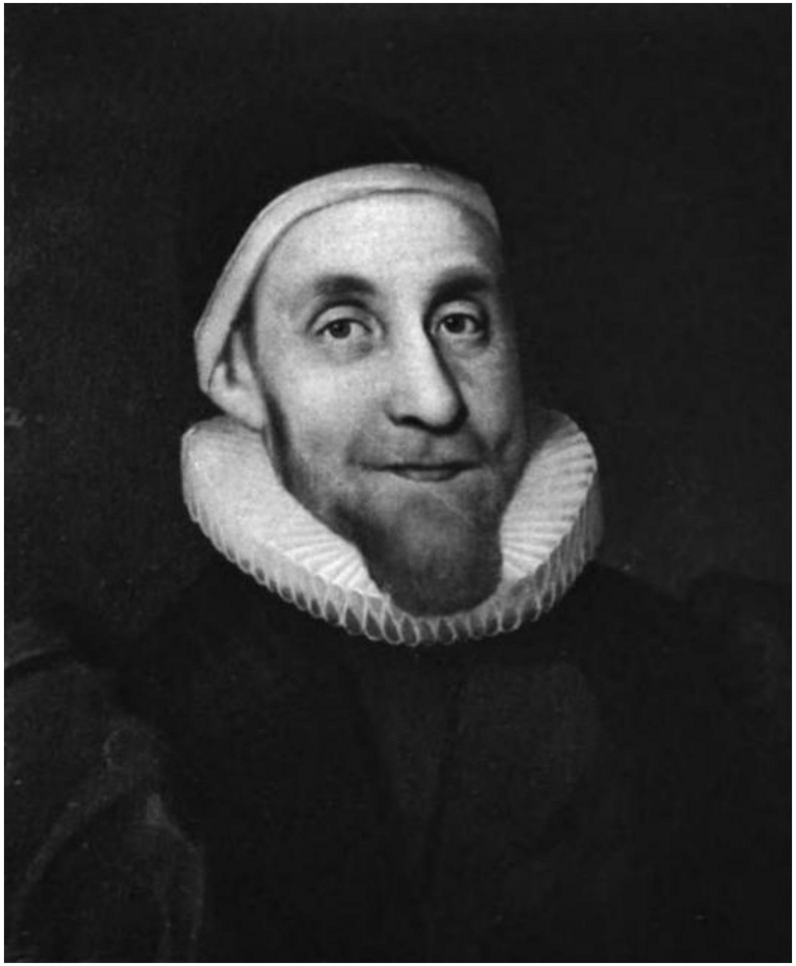
Now lots of empirical evidence that human behaviors tend to cluster in network space and time,…

but is this because of peer influence or alternate explanations?

“Obesity is Contagious”


The NEW ENGLAND JOURNAL of MEDICINE

Christakis & Fowler (2007)



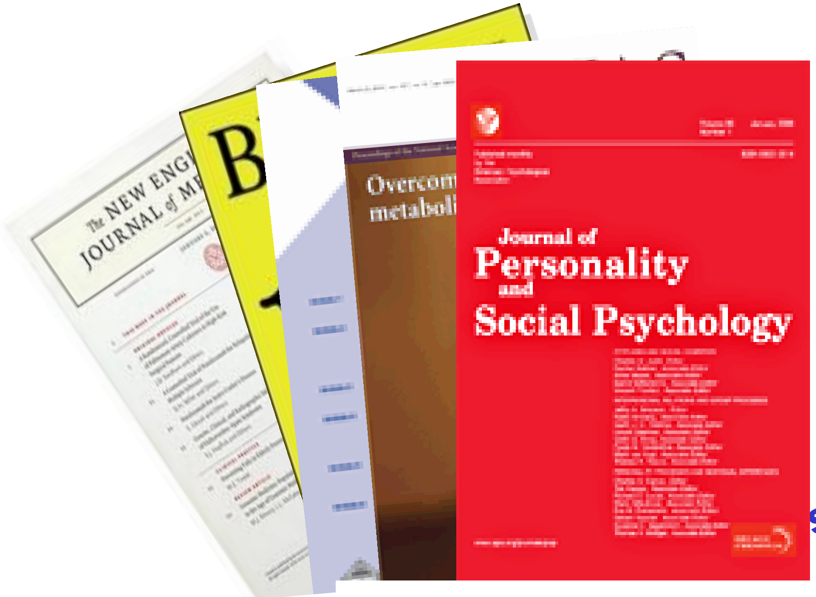
Robert Burton (1577-1640)

Homophily

"Birds of a feather, flock together"

“Everything is Contagious”

@sinanaral



“Loneliness is Contagious”

Alternative explanations:

- Reflection
- Homophily
- Confounding

Causal Structure of the underlying dynamic process implies:

1. ***Different Diffusion Properties*** and

2. ***Different Optimal Containment or Promotion Policies***.

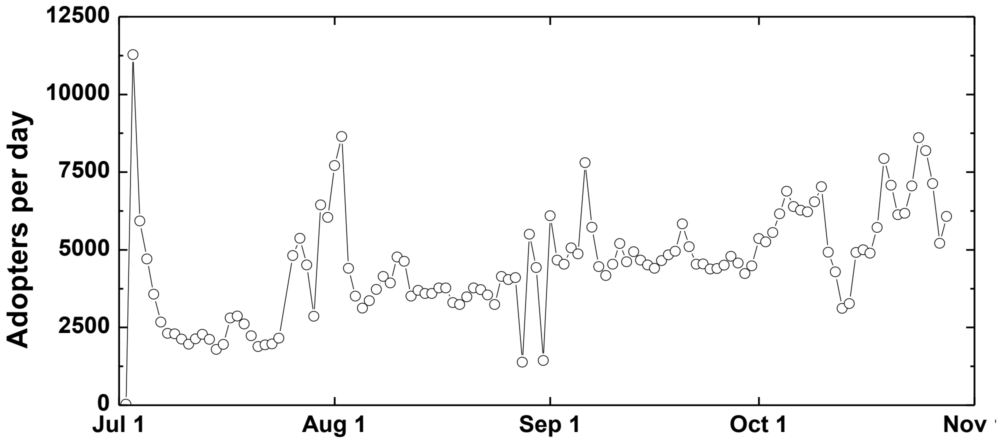
Causal Estimation in Networks

1. **Peer Effects Models (Extended Spatial Autoregressive Models)** (e.g. Manski 1993, Frank & Strauss 1986, Bramoulle et al 2009, Kelejian & Prucha 1998, Moffitt 2001, Lee 2006, Oestreicher-Singer and Sundararajan 2008)
    - Use variation in group size or structure as instrumental variables to identify deviations from group means.
2. **Actor Oriented Models (Co-Evolution of Networks and Behavior)** (e.g. Snijders 2001, Steglich, Snijders and Pearson 2004, Aral 2010)
    - Model micro decisions that maximize Behavioral and Network utility.
    - Apply Continuous Time Markov Models to Panel Network Data.
    - Estimate with MCMC or other simulated method of moments techniques
3. **Natural Experiments, Instrumental Variables** (Sacredote 2001, Tucker 2008)
4. **Structural Models** (Ghose and Han 2010)
5. **Automated Discovery of QED** (Jensen et al 2008)
6. **Dynamic Matched Sample Estimation – Yahoo IM** (Aral et al 2009)
    - Treatment – those with *n* friends who adopted at or before time *t.*
    - Control – those as likely to have *n* friends …"…as the treated, but who do not.
7. **Randomized Trials in Massive Networks** (Aral & Walker 2010a, b; Aral 2011)

Dynamic Matched Sampling

- 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.

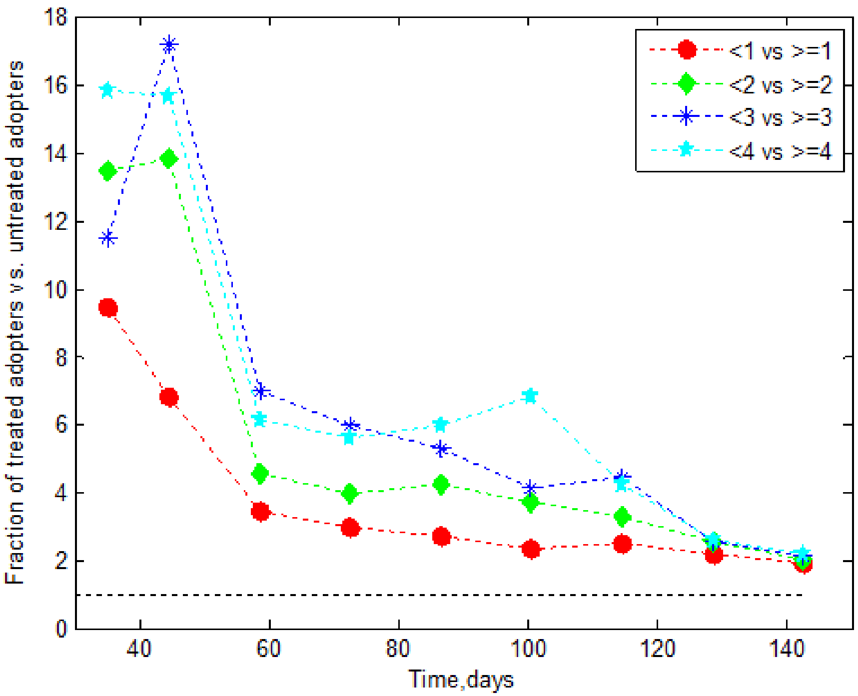
- Day by Day adoption and usage of a mobile service application (Yahoo Go) launched in July 2007. (532,365 Adopters)



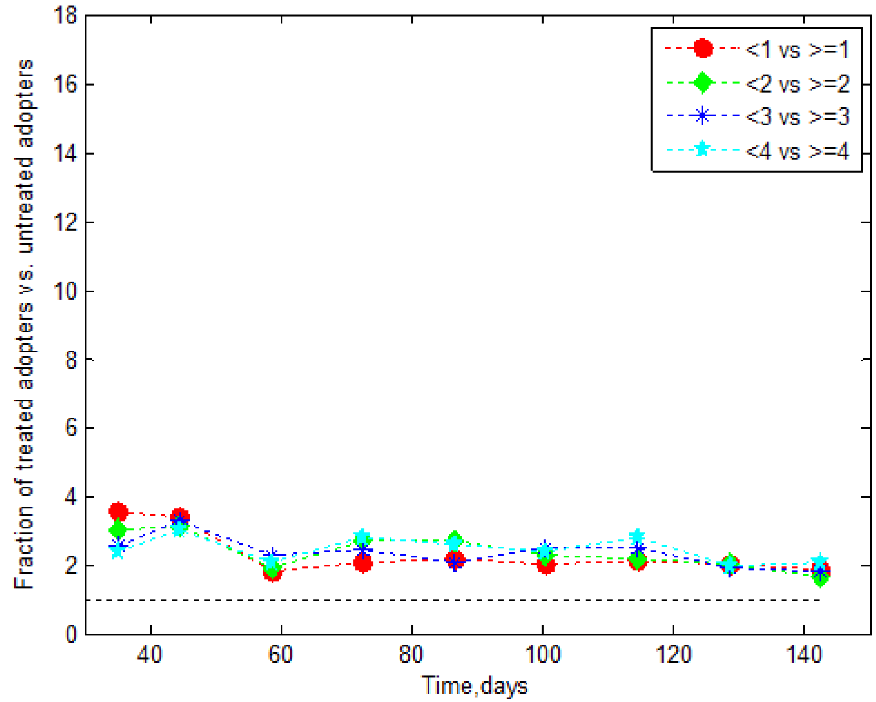
Aral, Muchnik, Sundararajan (2009) "Distinguishing Influence Based Contagion from Homophily Driven Diffusion in Dynamic Networks," *PNAS*, 106(51).

Distinguishing Influence from Homophily

"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)



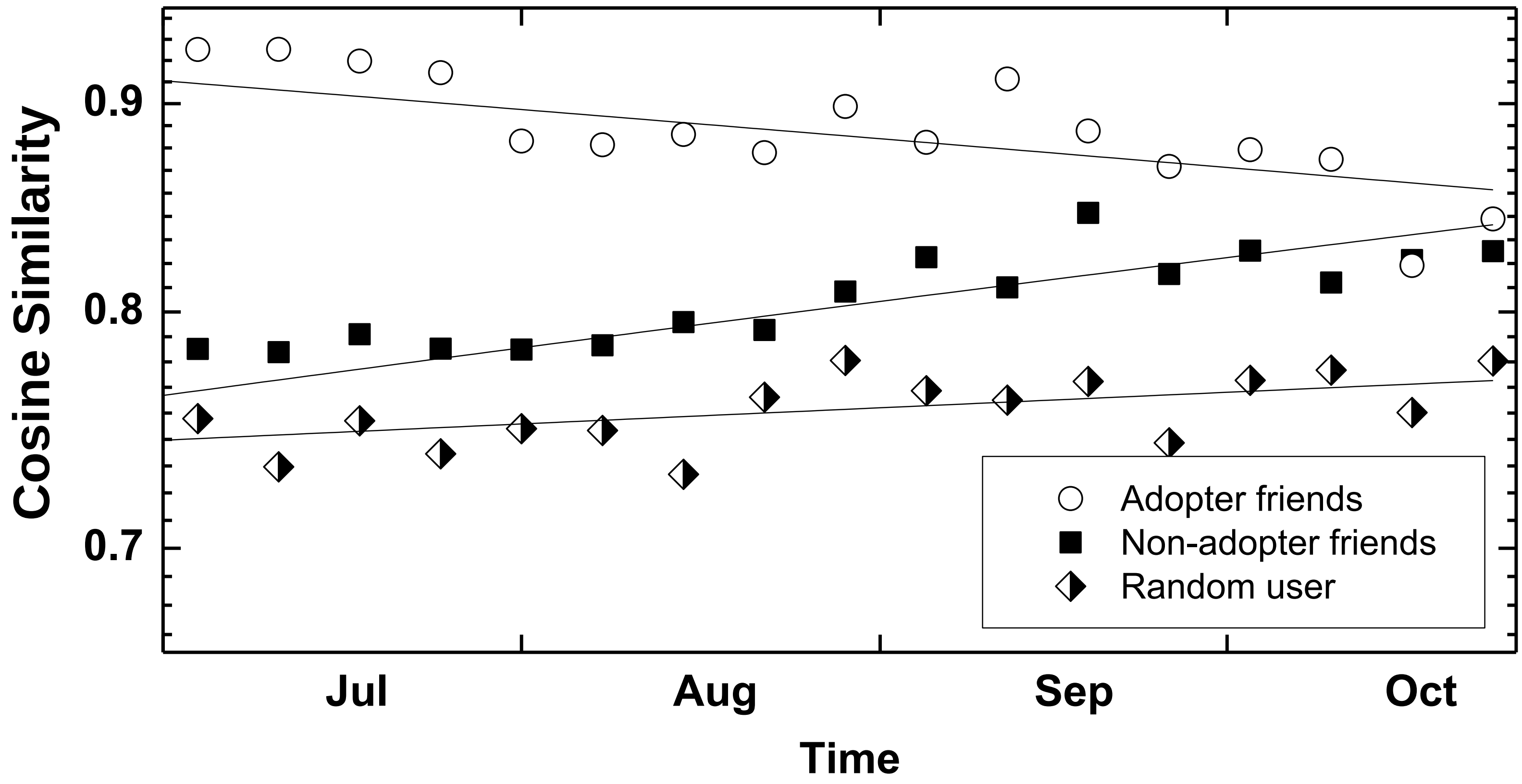
"Influence" Estimates Comparing Adoption in Treated and Untreated Cases In Our ***Dynamic Matched Sampling Framework*** Over Time



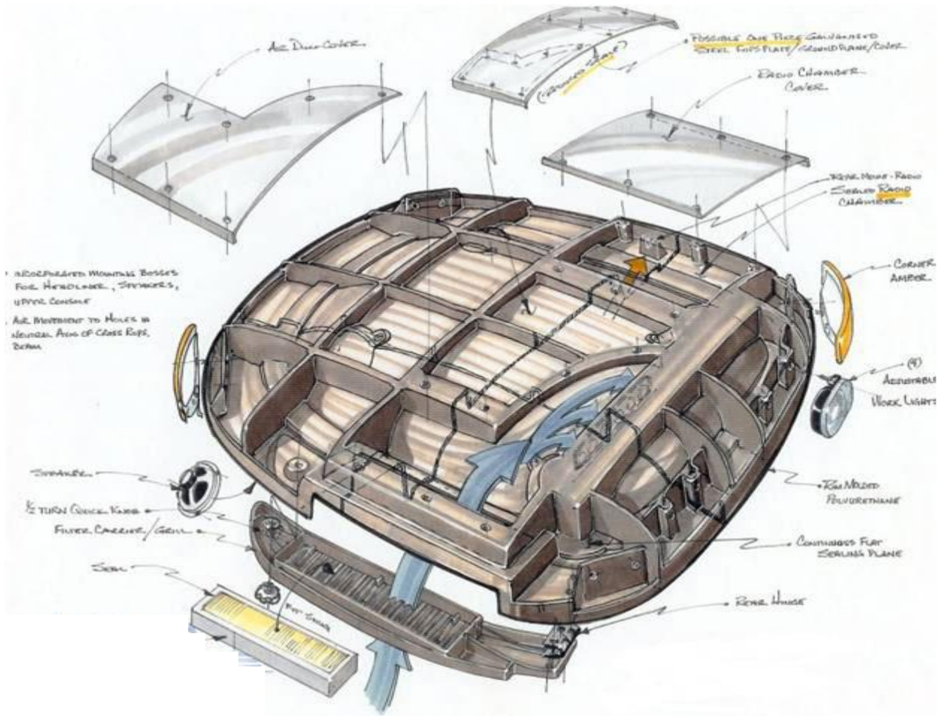
Much of the estimated influence is really observable homophily.

Exaggerated Homophily Amongst Early Adopters

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



Viral Product Design



Can we engineer products so they are more likely to be viral shared?

Aral & Walker (2011) “Creating Social Contagion through Viral Product Design: A Randomized Trial of Influence in Networks,” <u>*Management Science*</u>, Forthcoming

Viral Product Design

The process of explicitly engineering products so they are more likely to be shared amongst peers.

Viral Product Characteristics

Content likely to inspire viral sharing

Usefulness, topicality, prominence, positive valence and unexpectedness (Berger and Milkman 2009, Stephen and Berger 2009, Berger and Heath 2005, Phelps et al 2004, Heath, Bell and Sternberg 2001).

Viral Product Features

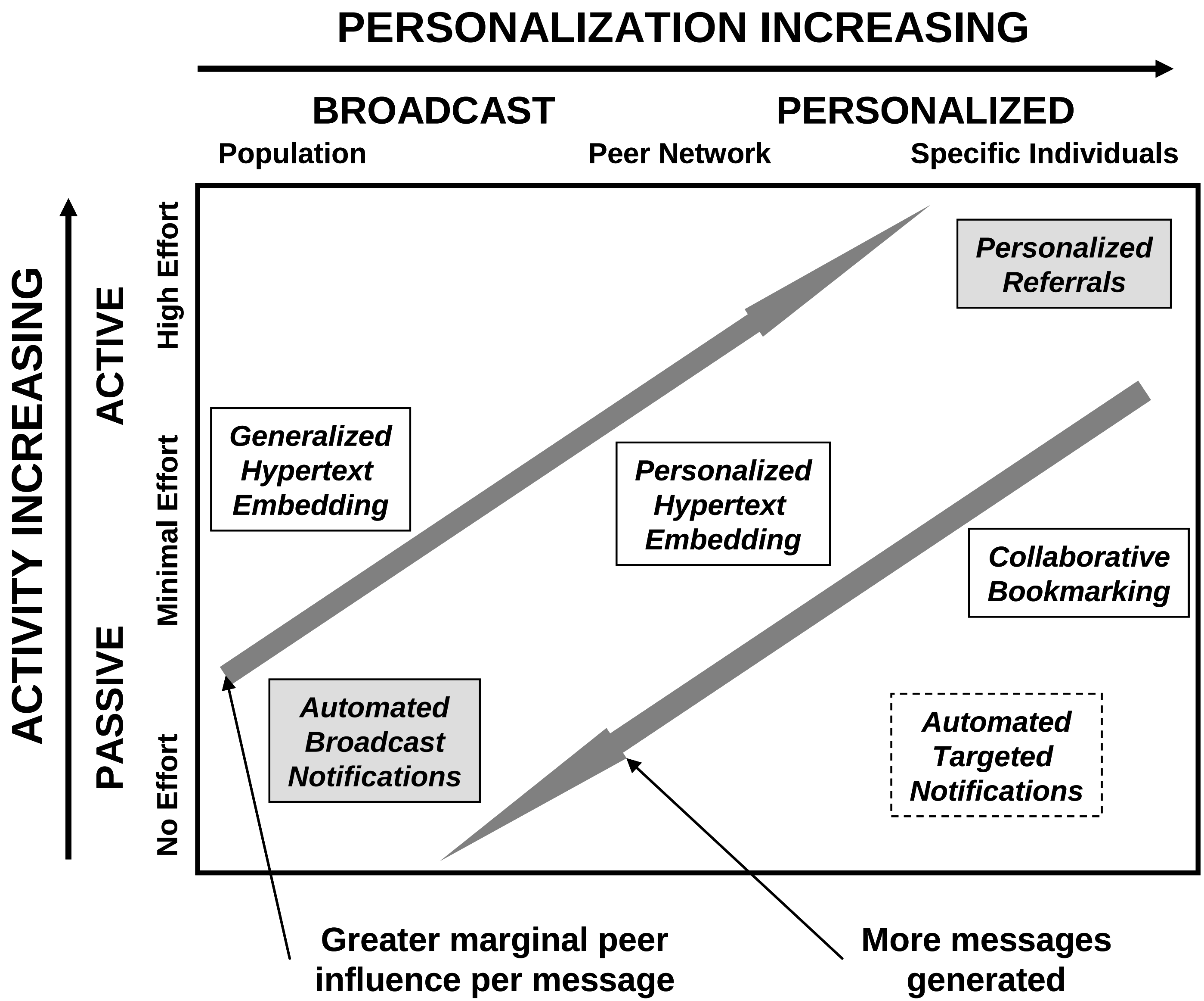
Modalities of use

Invites
Notifications
Hypertext Embedding

(No literature)

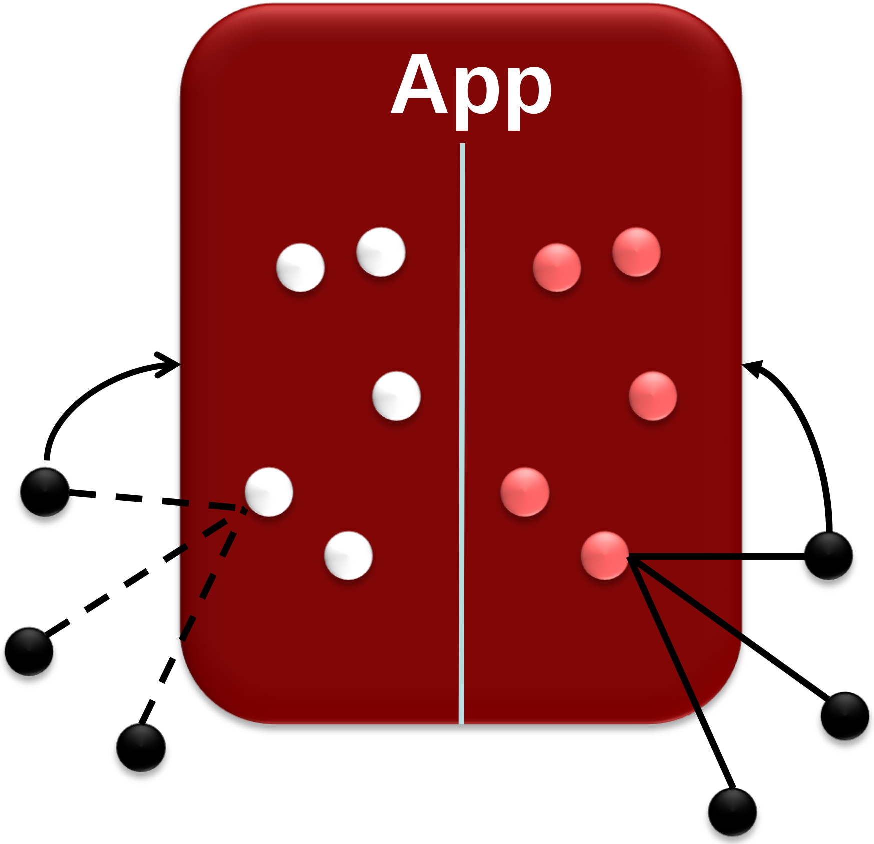
Viral Feature Space

@sinanaral



The Setup

@sinanaral



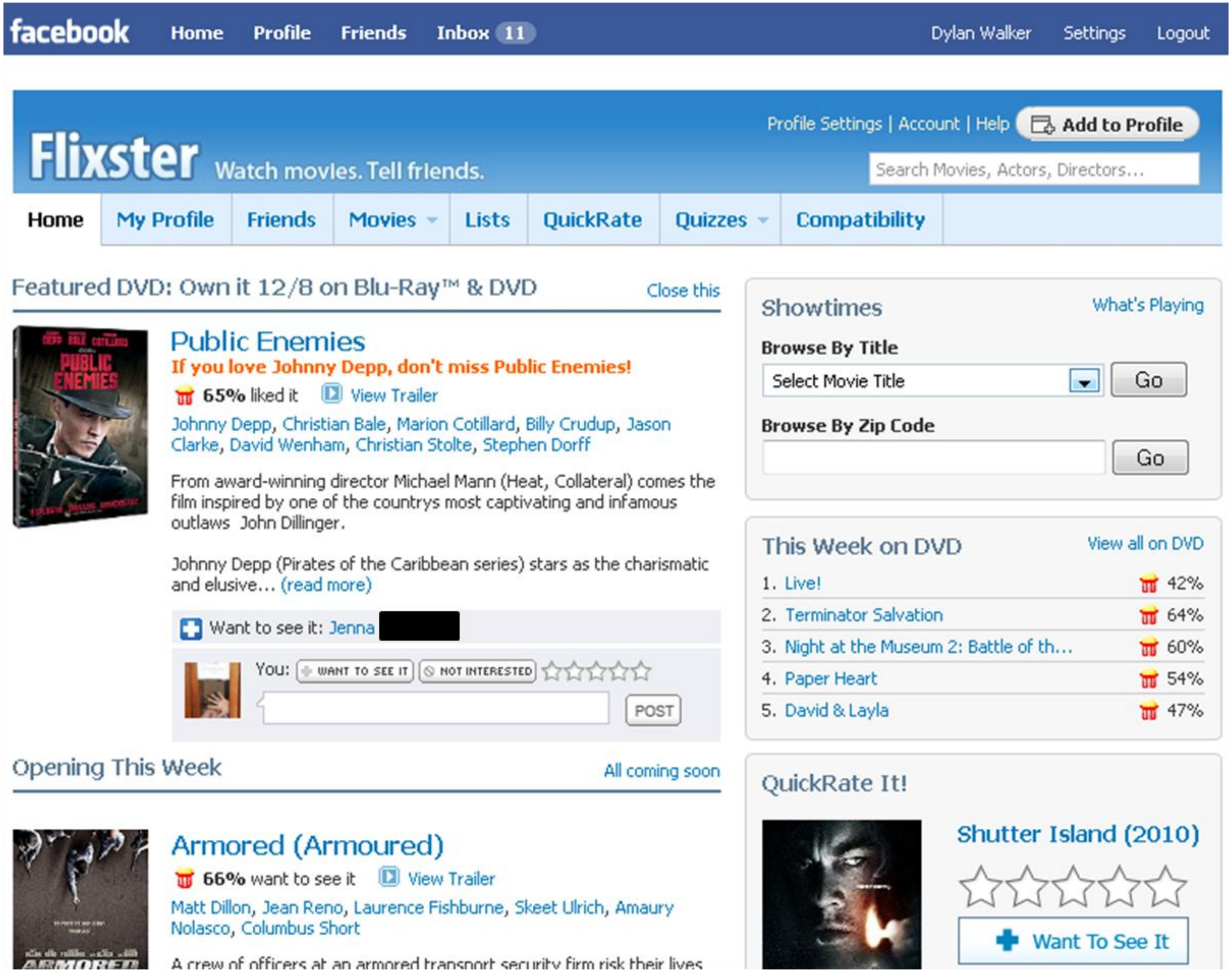
- Randomly Enabled Viral Messaging.
- Observed the Adoption and Use of the App by Friends of Control and Experimental Group Users.

Data

- ~ 10K Experimental Users
- ~ 1.4M Friends of Experimental Users
- We Observe Application Diffusion Over this Network
  - Adoption
  - Use

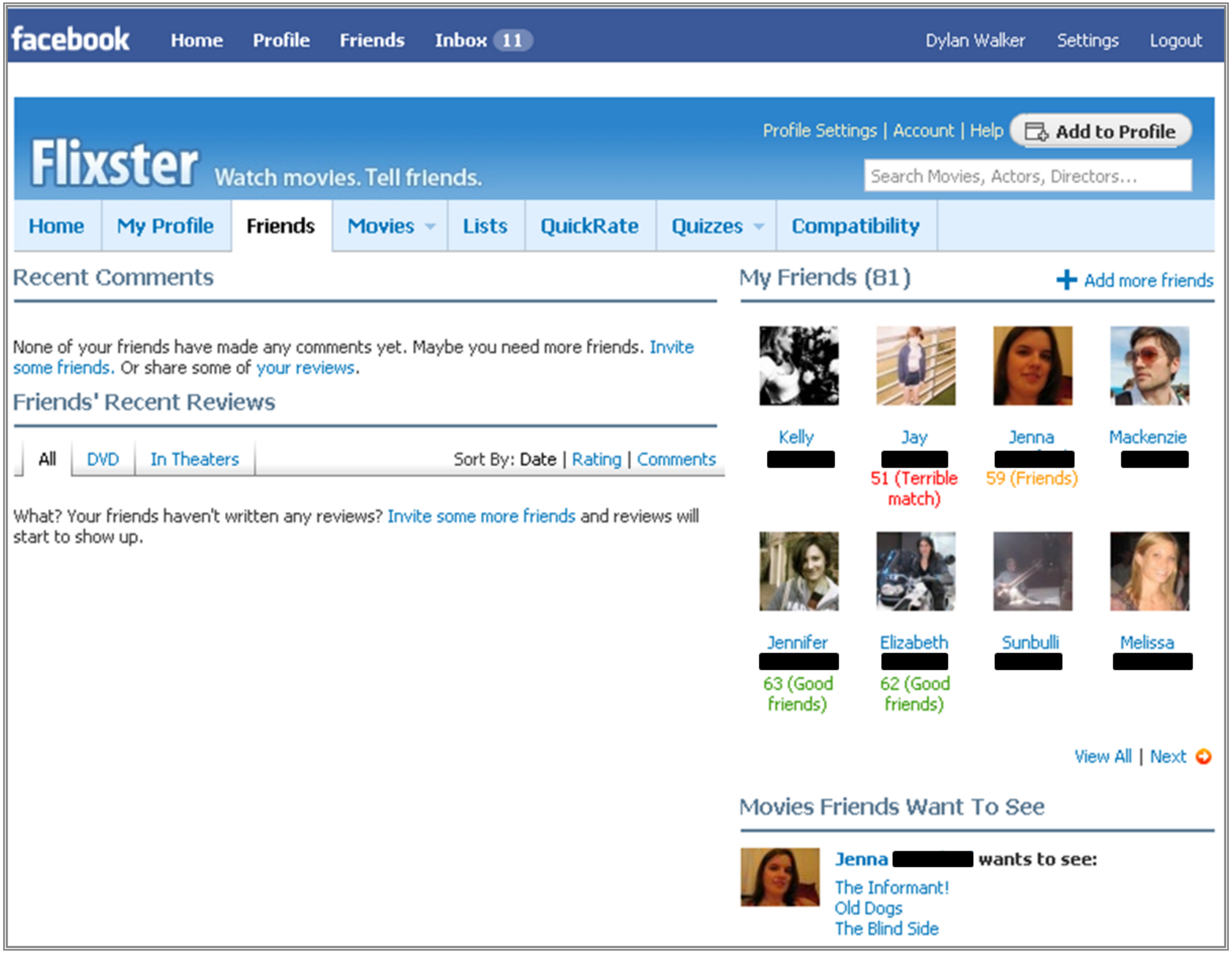
Flixster - An Example Facebook Application

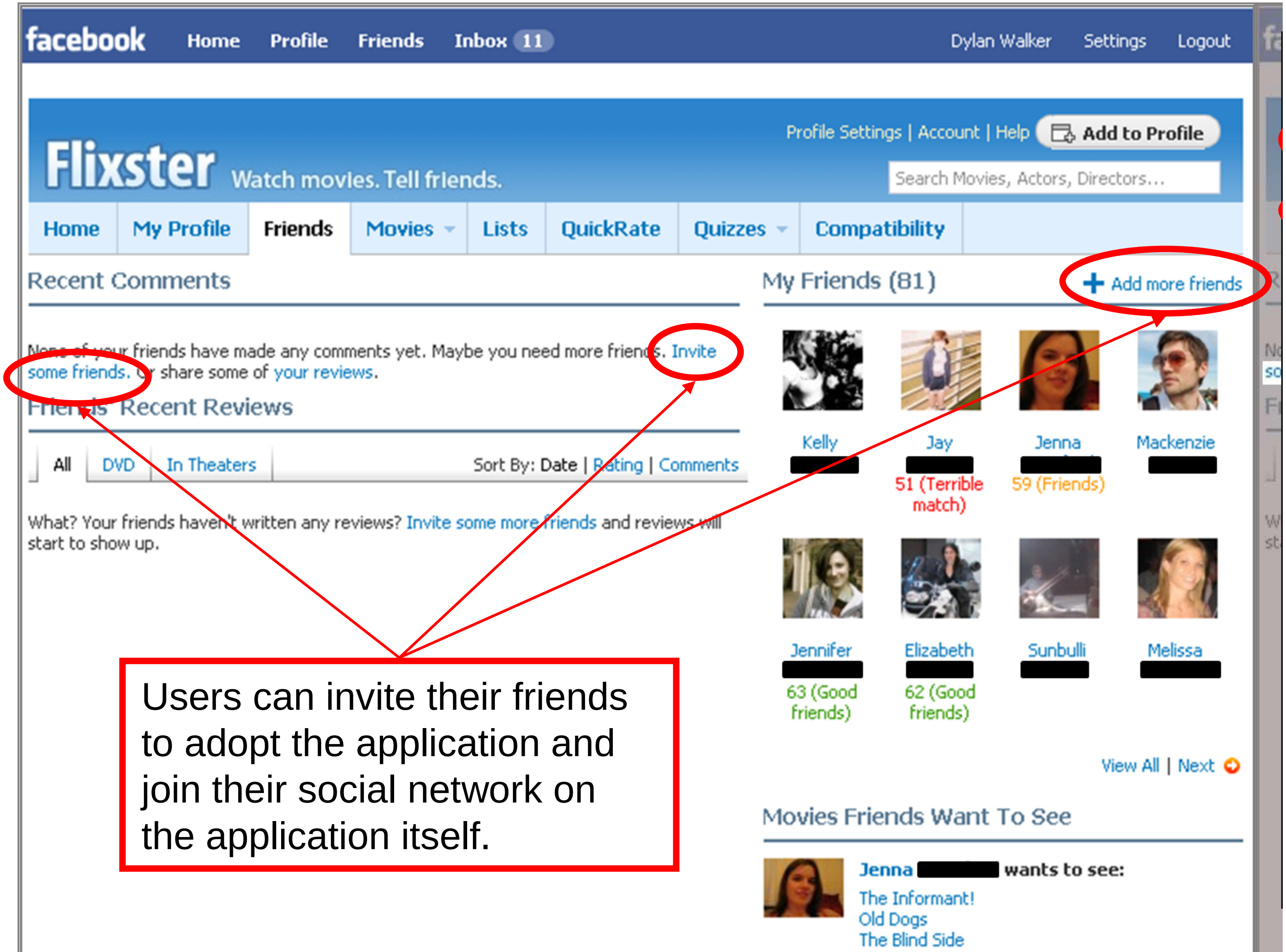
@sinanaral

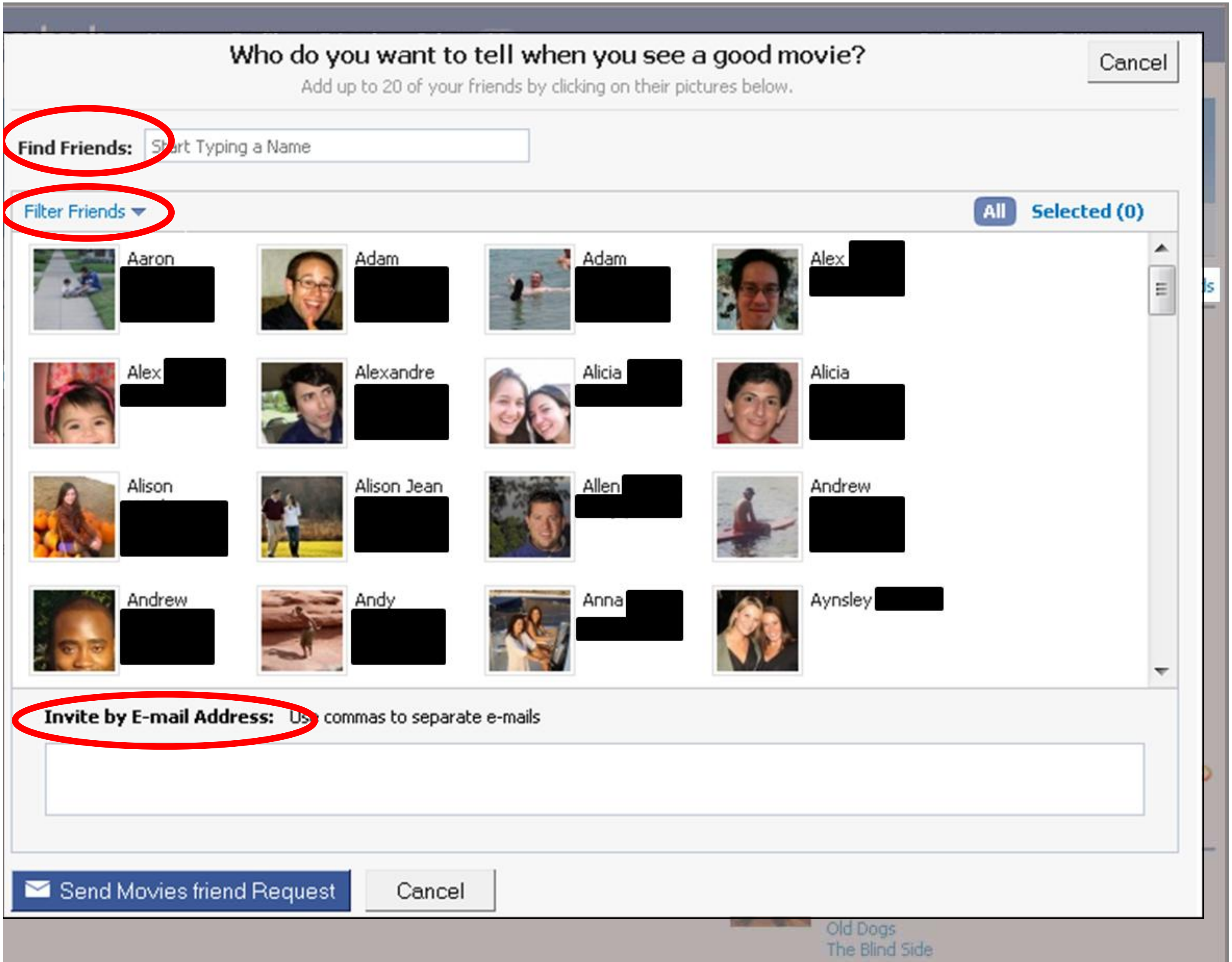


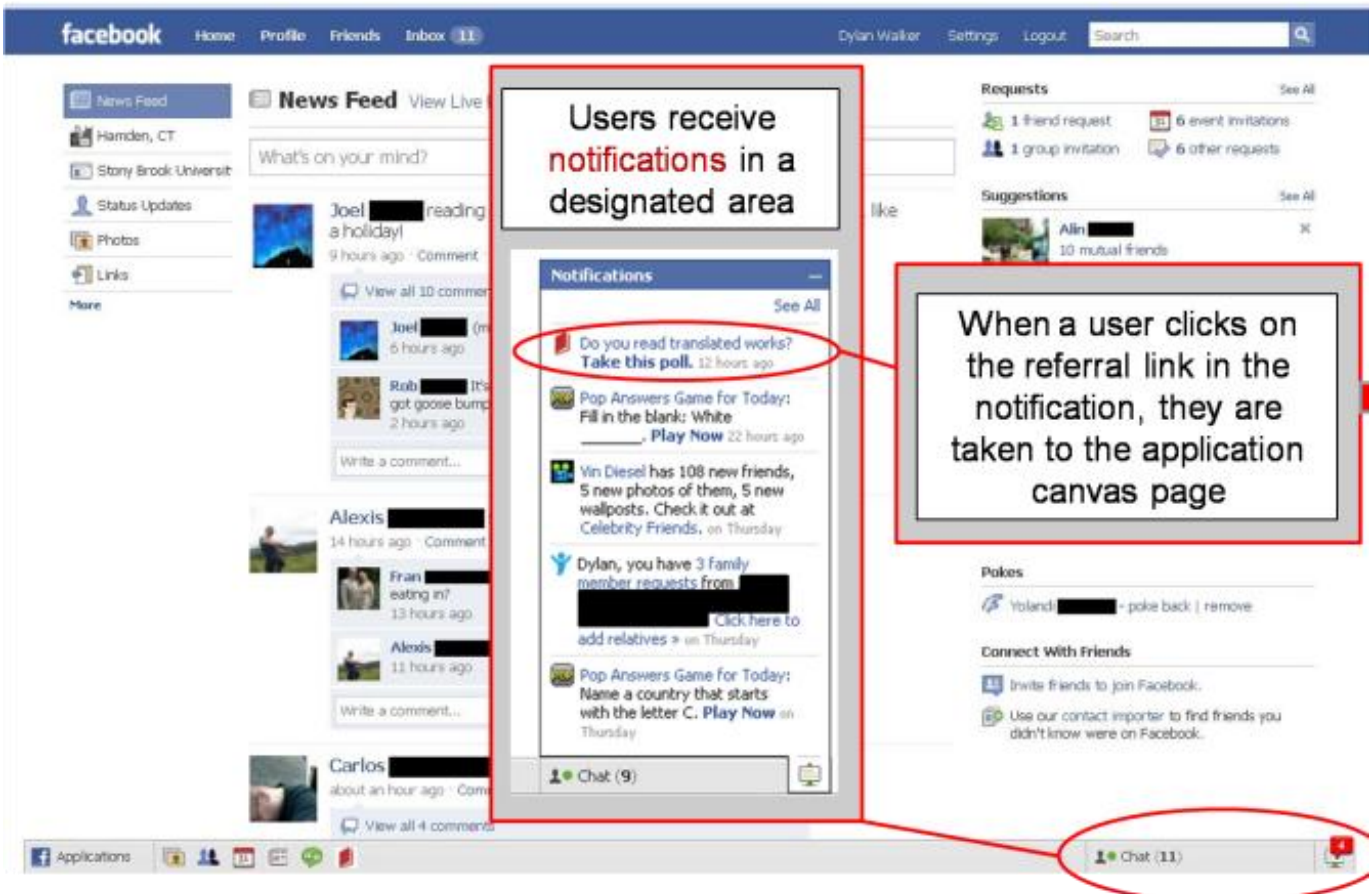
Flixster - An Example Facebook Application

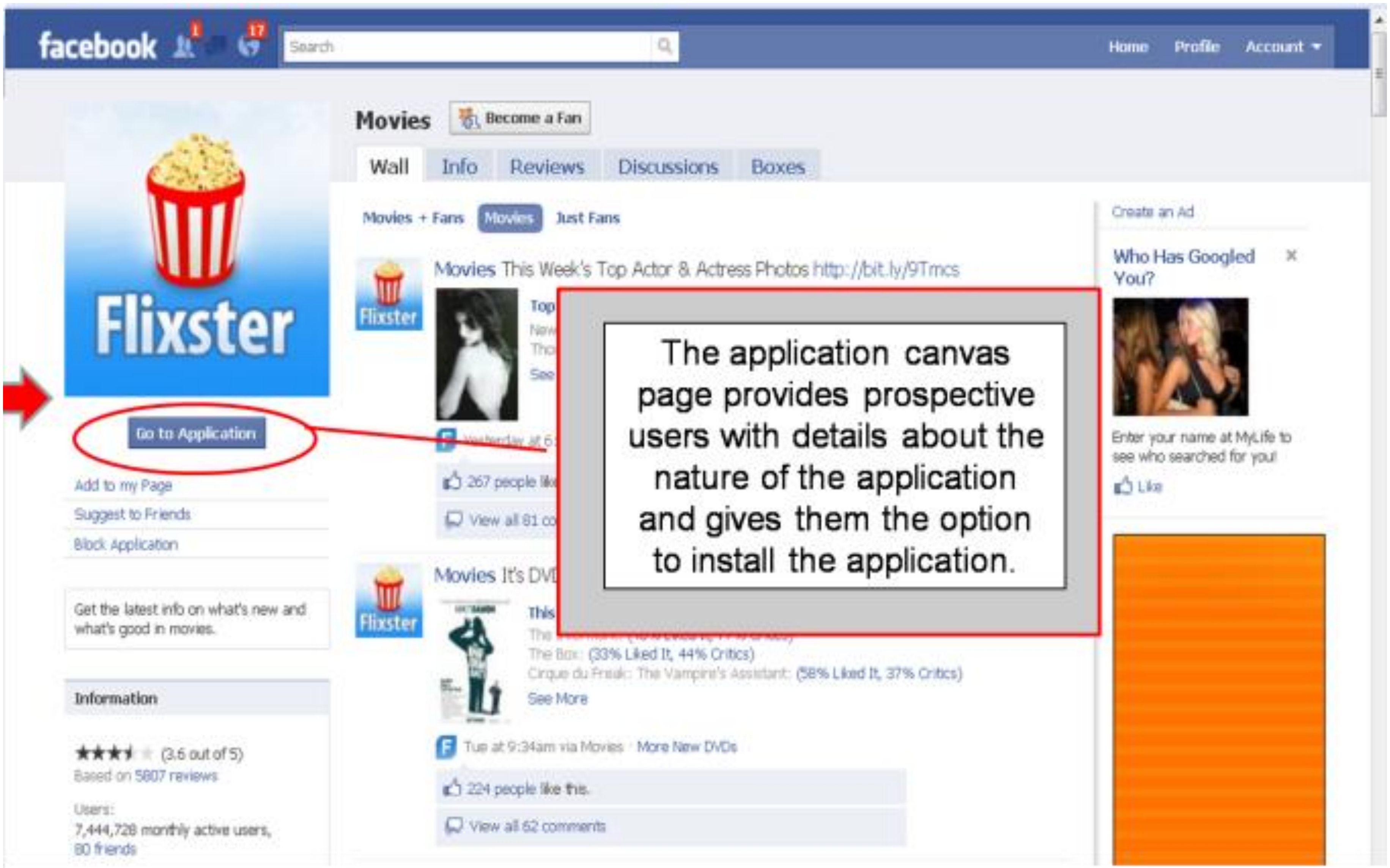
@sinanaral



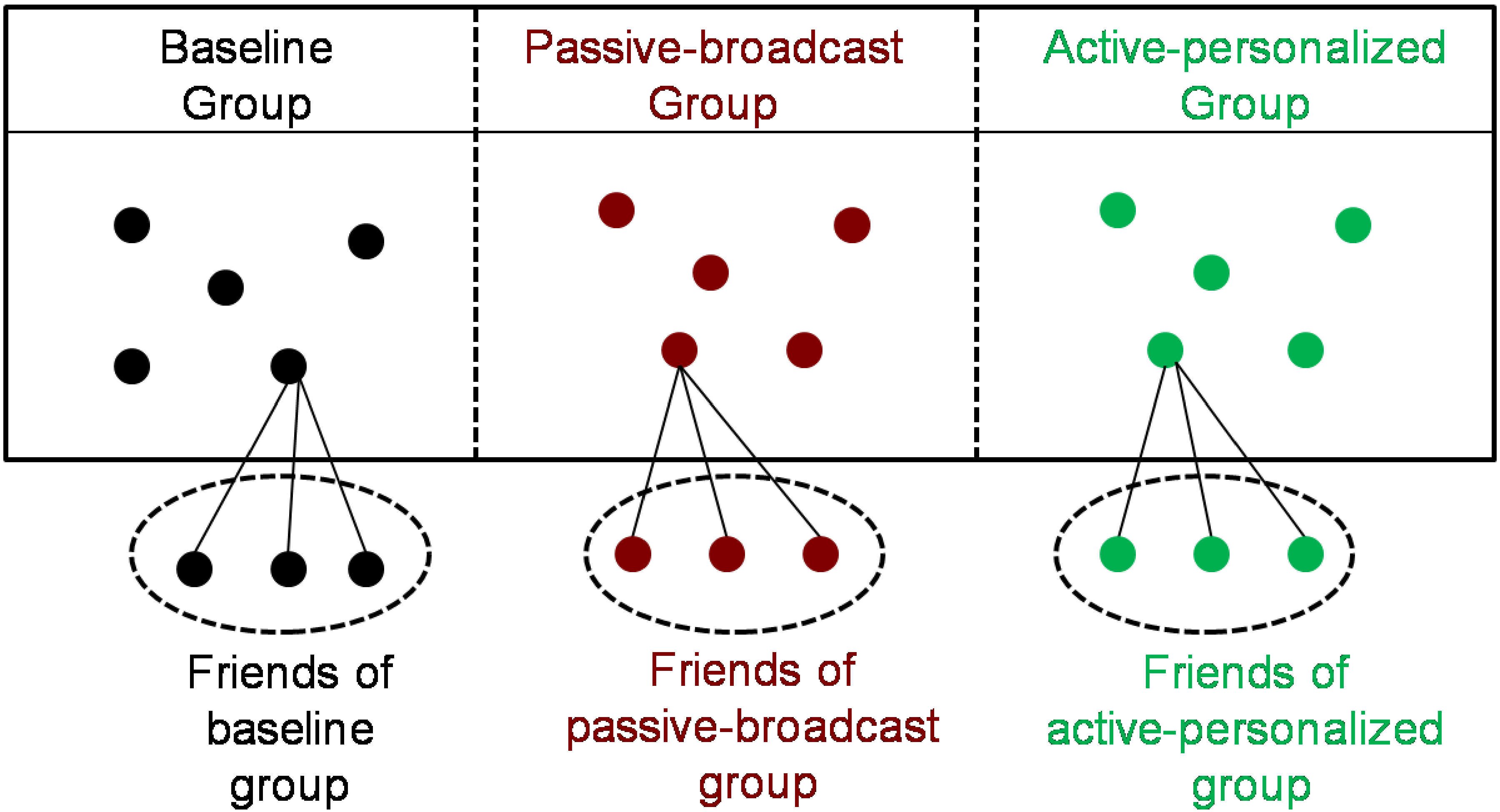

facebook
Home
Profile
Friends
Inbox 11
Dylan Walker
Settings
Logout
Profile Settings | Account | Help
Add to Profile
Flixster
Watch movies. Tell friends.
Search Movies, Actors, Directors...
Home
My Profile
Friends
Movies
Lists
QuickRate
Quizzes
Compatibility
Recent Comments
None of your friends have made any comments yet. Maybe you need more friends. Invite some friends. Or share some of your reviews.
Friends' Recent Reviews
All
DVD
In Theaters
Sort By: Date | Rating | Comments
What? Your friends haven't written any reviews? Invite some more friends and reviews will start to show up.
My Friends (81)
Add more friends
Kelly [redacted]
Jay [redacted]
51 (Terrible match)
Jenna [redacted]
59 (Friends)
Mackenzie [redacted]
Jennifer [redacted]
63 (Good friends)
Elizabeth [redacted]
62 (Good friends)
Sunbulli [redacted]
Melissa [redacted]
View All | Next
Movies Friends Want To See
Jenna [redacted] wants to see:
The Informant!
Old Dogs
The Blind Side
Users can invite their friends to adopt the application and join their social network on the application itself.


Who do you want to tell when you see a good movie?
Cancel
Add up to 20 of your friends by clicking on their pictures below.
Find Friends:
Start Typing a Name
Filter Friends
All
Selected (0)
Aaron [redacted]
Adam [redacted]
Adam [redacted]
Alex [redacted]
Alex [redacted]
Alexandre [redacted]
Alicia [redacted]
Alicia [redacted]
Alison [redacted]
Alison Jean [redacted]
Allen [redacted]
Andrew [redacted]
Andrew [redacted]
Andy [redacted]
Anna [redacted]
Aynsley [redacted]
Invite by E-mail Address: Use commas to separate e-mails
Send Movies friend Request
Cancel
Old Dogs
The Blind Side


Users receive notifications in a designated area
When a user clicks on the referral link in the notification, they are taken to the application canvas page
facebook
Home
Profile
Friends
Inbox 11
Dylan Walker
Settings
Logout
Search
News Feed
Notifications
See All
Do you read translated works? Take this poll. 12 hours ago
Pop Answers Game for Today: Fill in the blank: White _______. Play Now 22 hours ago
Vin Diesel has 108 new friends, 5 new photos of them, 5 new wallposts. Check it out at Celebrity Friends. on Thursday
Dylan, you have 3 family member requests from [redacted] Click here to add relatives » on Thursday
Pop Answers Game for Today: Name a country that starts with the letter C. Play Now on Thursday
Chat (9)
Requests
1 friend request
6 event invitations
1 group invitation
6 other requests
Suggestions
Pokes
Connect With Friends
Invite friends to join Facebook.
Applications
Chat (11)



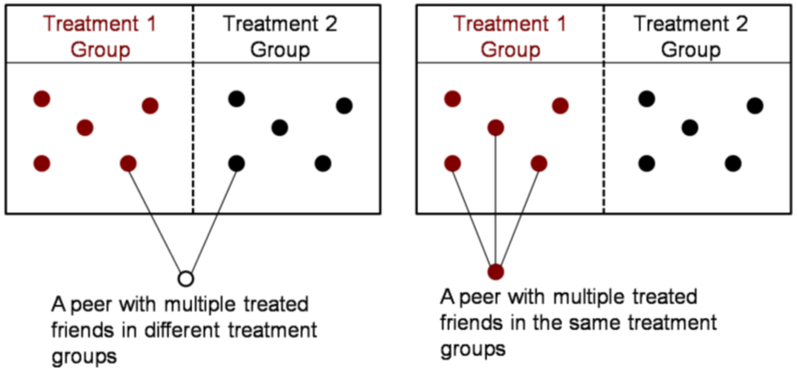
The application canvas page provides prospective users with details about the nature of the application and gives them the option to install the application.

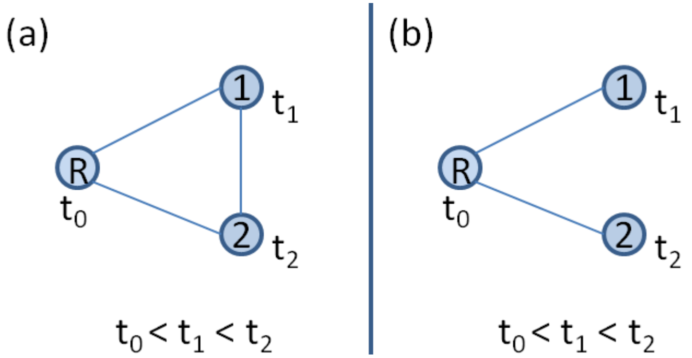

Baseline Group
Passive-broadcast Group
Active-personalized Group
Friends of baseline group
Friends of passive-broadcast group
Friends of active-personalized group

Preventing Contamination and Leakage

@sinanaral

- ***Leakage and contamination could occur if peers are***
  - ***a) connected through indirect pathways,***
  - ***b) connected to multiple treated peers in different treatment groups or***
  - ***c) connected to multiple treated peers in the same treatment group.***



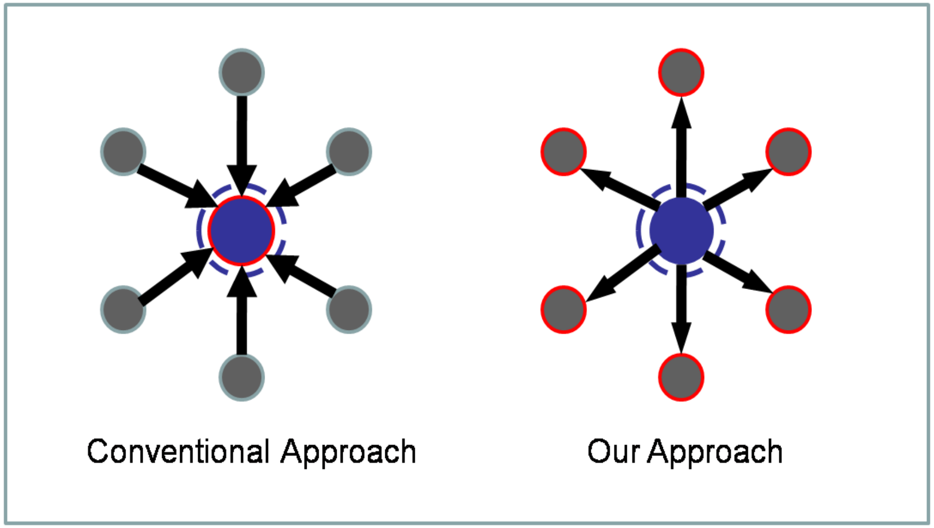


- ***These scenarios are rare in our data***
- ***We control for leakage and peers of multiple treated users by only evaluating recruited users and right censoring contaminated peers.***
- ***Contaminated: Any peer with multiple treated peers after time t at which they have multiple treated peers.***
- ***This may make our results more conservative but also minimizes leakage and contamination***

Conventional Approach in Observational Data

$$P(y_{it}=1 \mid y_{it-1}=0)=F(x_{it}\gamma, \beta \sum_j w_{ij} y_{jt})$$

"Inside-Out" Estimation



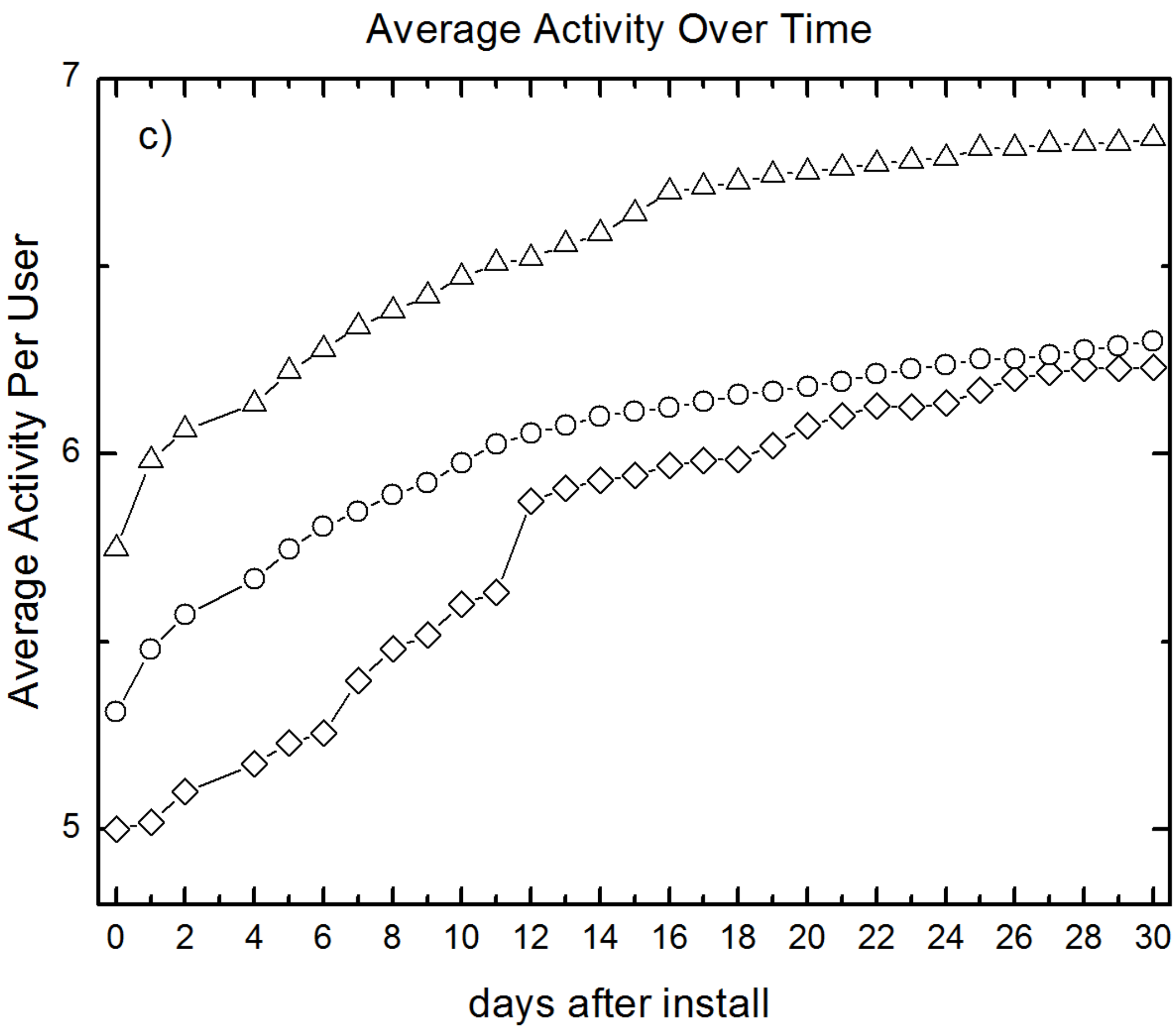
A Variance Corrected Stratified Proportional Hazards Model

$$\lambda_k(t, X_{ki}) = \lambda_{0k}(t) e^{X_{ki}\beta}$$

Which Features Spread Contagion Best?

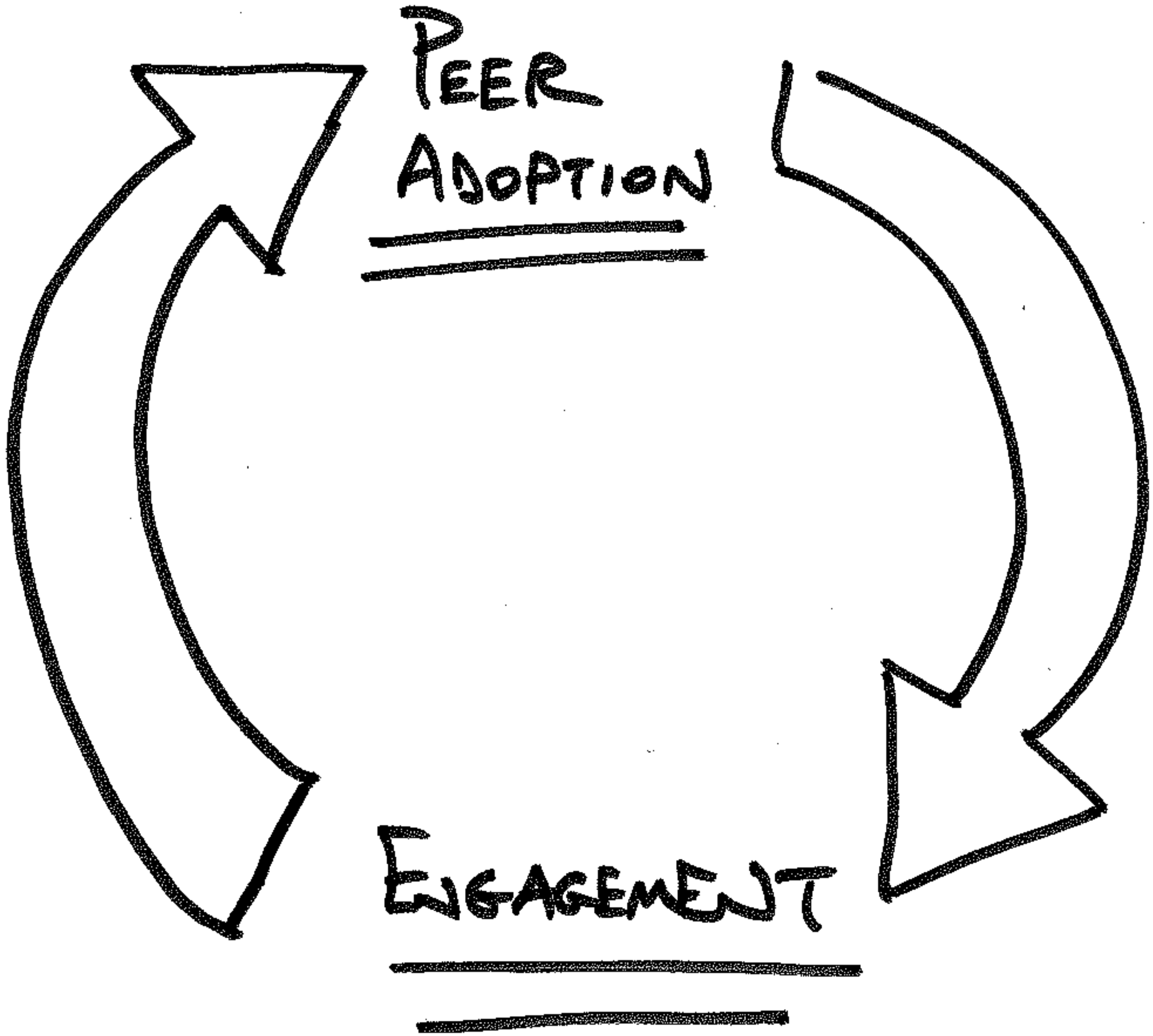
	Personal Invitations	Passive Awareness
Influence Per Message	↑6%	↑2%
Global Diffusion	↑98%	↑246%
Stickiness		

Sustained Engagement



Which Features Spread Contagion Best?

	Personal Invitations	Passive Awareness
Influence Per Message	↑6%	↑2%
Global Diffusion	↑98%	↑246%
Stickiness	↑17%	0%


PEER ADOPTION
ENGAGEMENT

Causality

Content

Causality

Content

NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN
SCHOOL OF BUSINESS

Content Gives…

@sinanaral

Meaning

Intension

Relevance

Persuasion

Reaction

Context

Some Examples of Research

@sinanaral

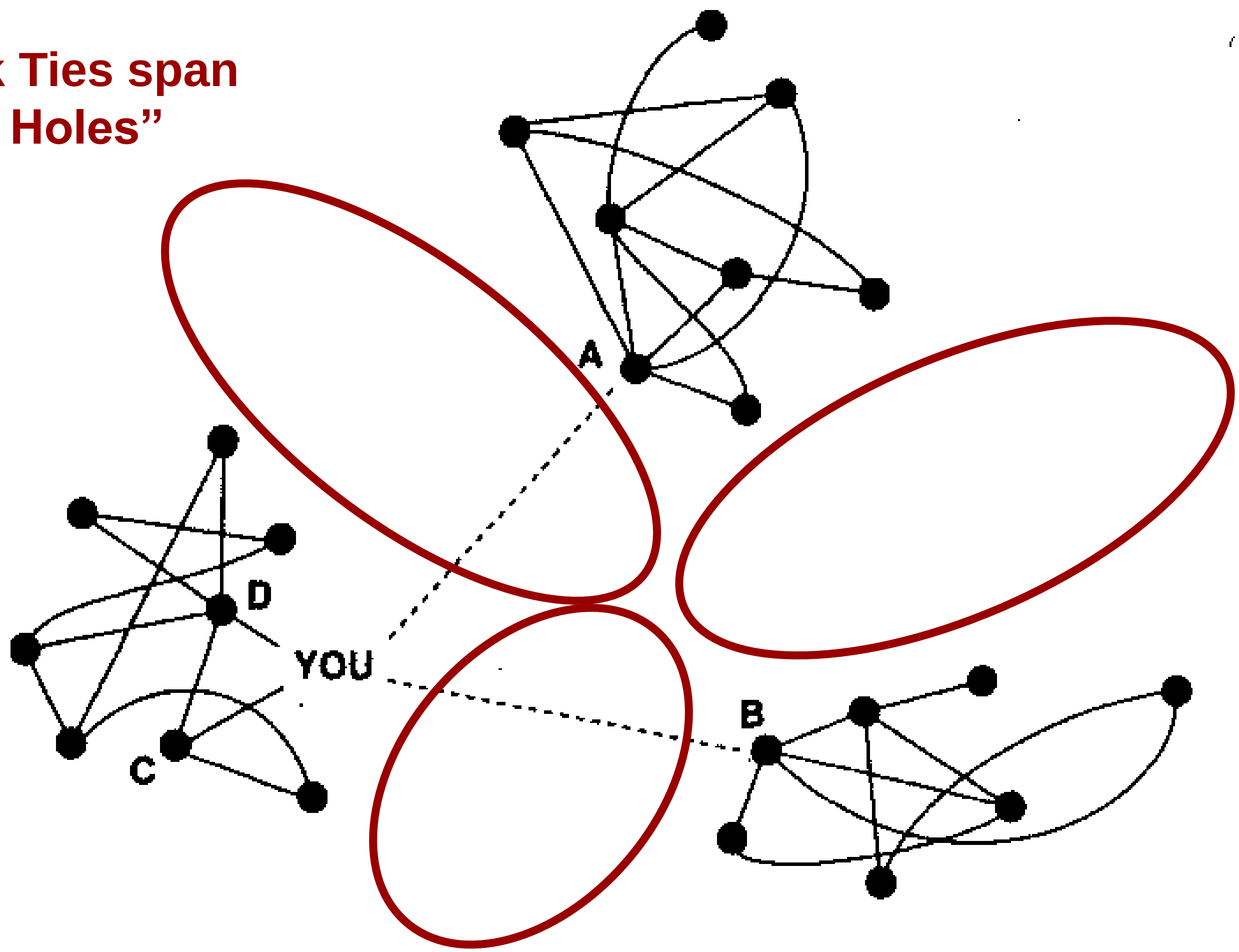
- Leskovec, Backstrom and Kleinberg (2009) “Meme-Tracking and the Dynamics of the News Cycle.”
- Bakshy, Hofman, Mason, Watts (2011) “Everyone’s an Influencer: Quantifying Influence on Twitter.”
- Wu, Hofmann, Mason, Watts (2011) “Who Says What to Whom on Twitter.”
- McCallum, Wang and Corrada-Emmanuel (2007) "Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email."
- Blei, Ng, Jordan (2003) “Latent Dirichlet Allocation.”
- And tons of really interesting research at ICWSM!

Strengths of Weak Ties & Structural Holes

@sinanaral

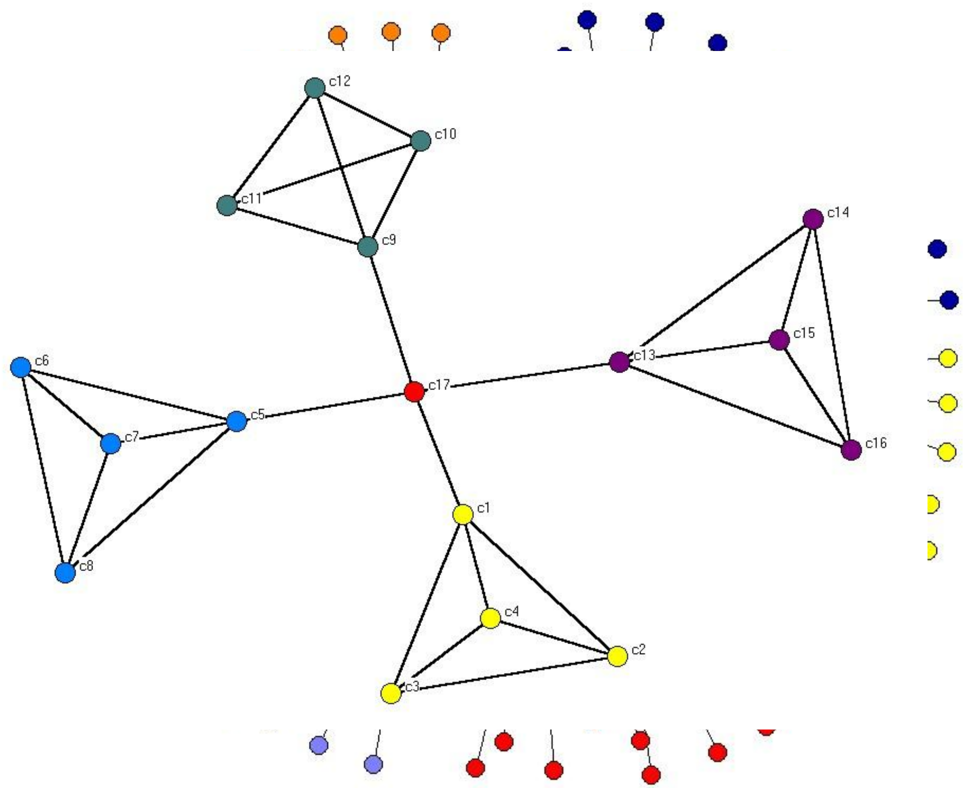
Opportunities for "Brokerage"…
are typically enabled by weak ties

Bridging Weak Ties span "Structural Holes"



Information Advantage

@sinanaral



Value of information comes from from its uneven distribution across local network neighborhoods.

Connection to diverse neighborhoods gives access to novel pools of information.

Novel information is valuable due to its local scarcity.

Actors with scarce, novel information can

- ✓ **broker opportunities, engage in information arbitrage**
- ✓ **use information as a commodity, or**
- ✓ **apply information to problems that are intractable given local information (innovation).**

A 40 year old assumption…

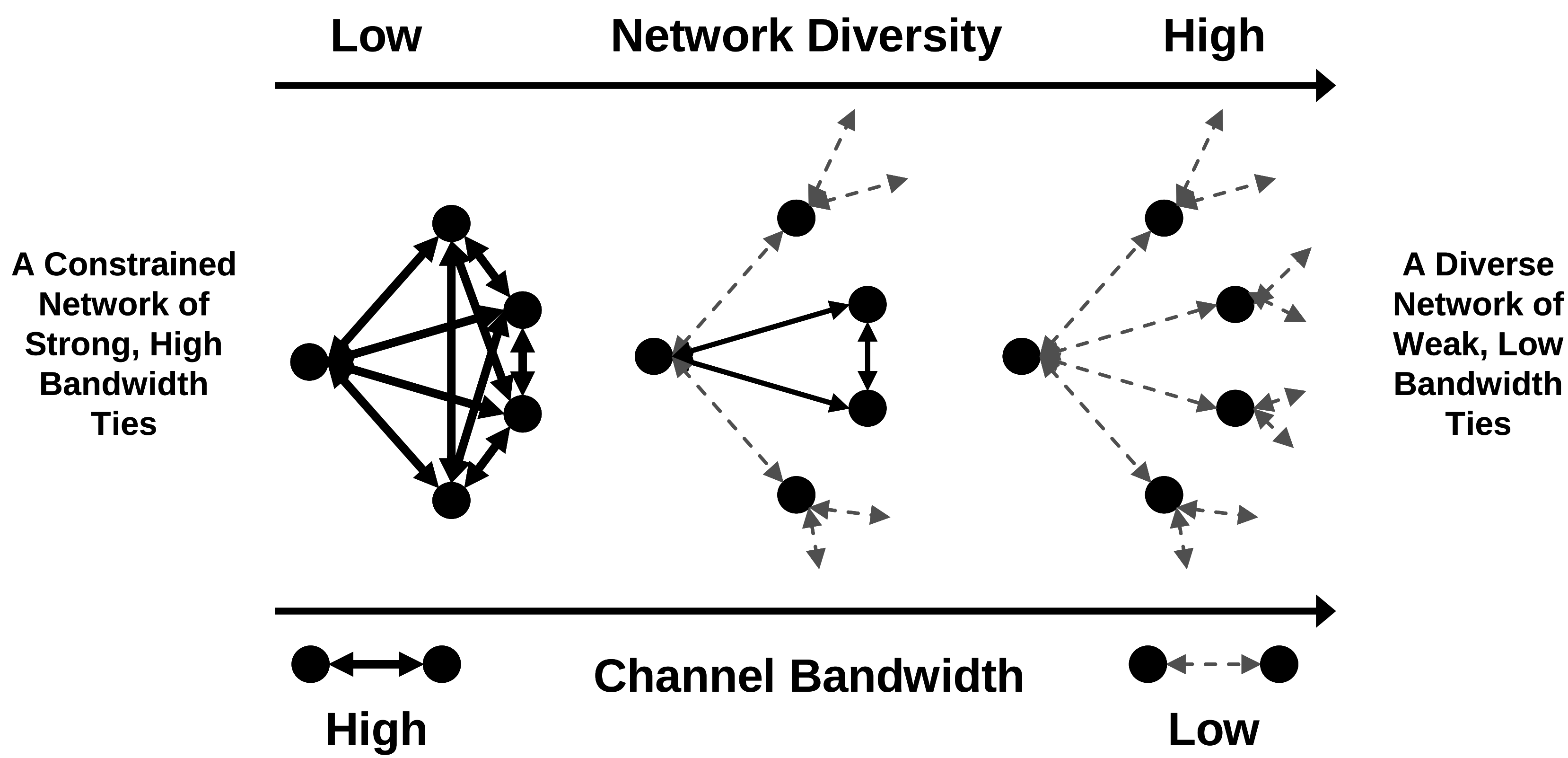
- **Network structure is associated with performance.**
  - **Productivity of R&D teams (Reagans & Zuckerman 2001)**
  - **Labor Market Outcomes (Montgomery 1991, 1992)**
  - **Wages, Promotion, Job Placement (Burt 1992)**
  - **Innovation (Burt 2004)**
  - **Productivity of information workers (Aral et al 2006)**
- **Key theoretical mechanism: access to information.**

DIVERSE NETWORKS → **DIVERSE, NOVEL INFORMATION** → **PRODUCTIVITY, PERFORMANCE, INNOVATION**

"The Diversity-Bandwidth Tradeoff"

@sinanaral

The Theory is Problematic because Structural Diversity is likely to be associated with weak ties, creating A Tradeoff Between Network Diversity and Channel Bandwidth.



Diversity ⇔ weak ties => lower bandwidth, frequency, topical dimension, detail, complexity.

Aral & Van Alstyne (2011) "The Diversity-Bandwidth Tradeoff." *American Journal of Sociology*; Forthcoming.

Information Environment Mediates Tradeoff

1. ***Information Overlap*** – The degree to which topics are uniformly or heterogeneously distributed over nodes.

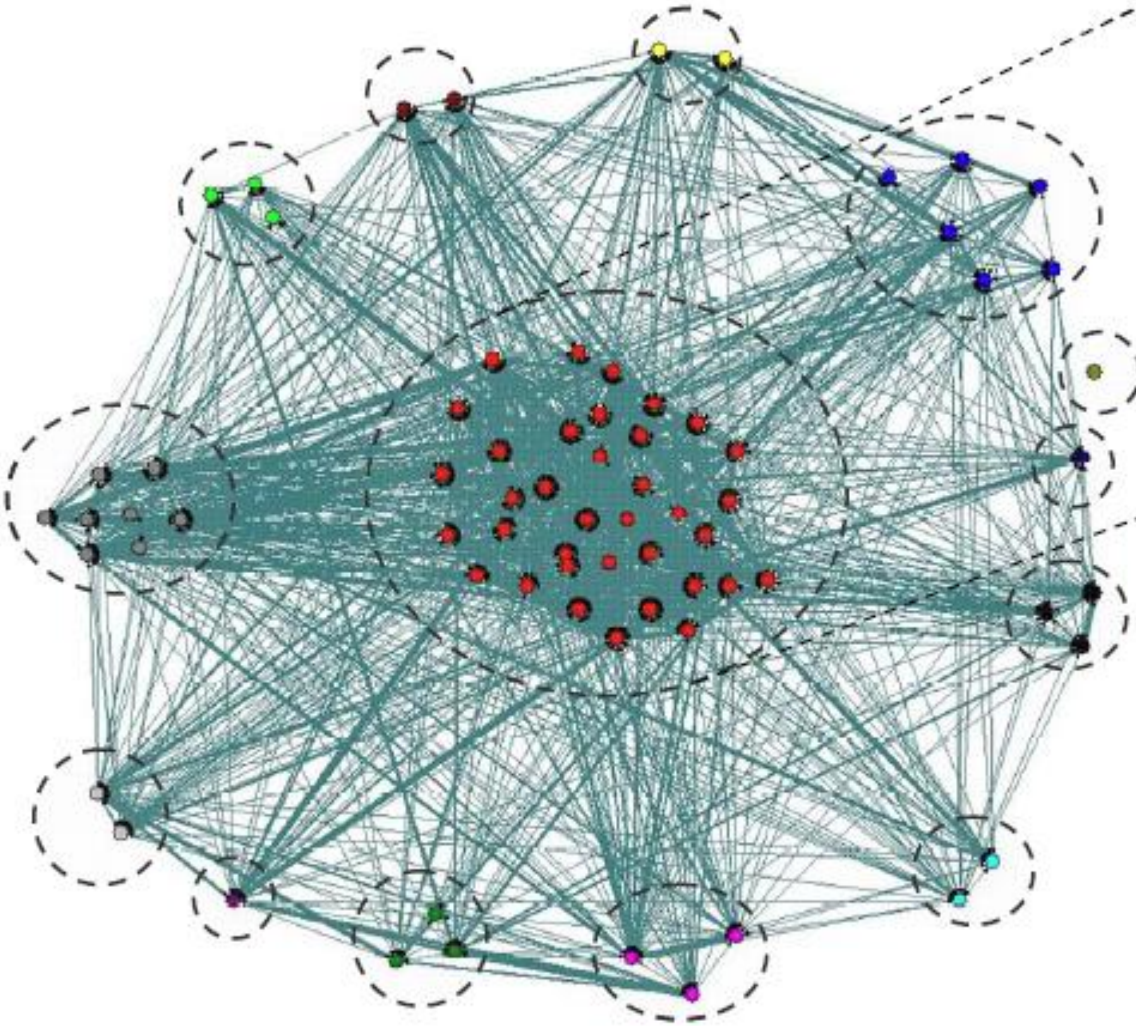
2. ***Size of the Topic Space*** – How many distinct topics exist in the network.

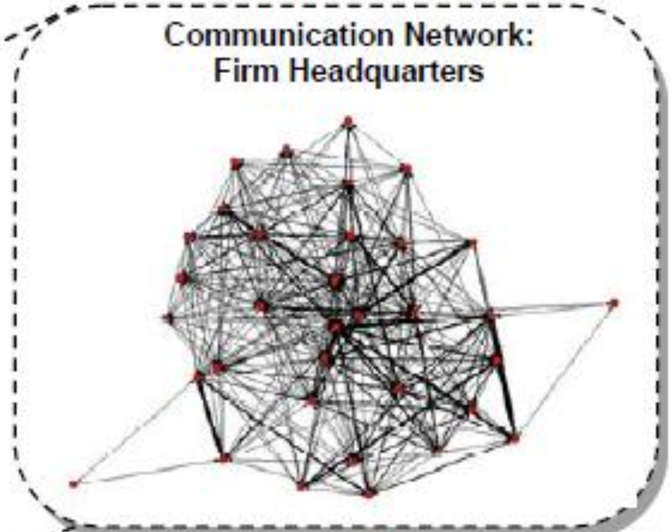
3. ***Refresh Rate*** – How often actors' information is changing per unit time.

Study Context - Executive Search

@sinanaral

Firm Communication Network

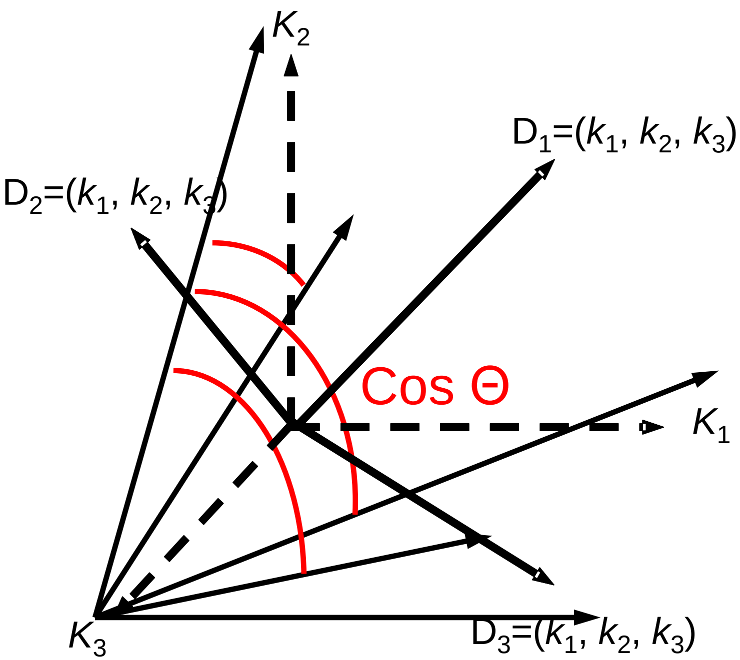


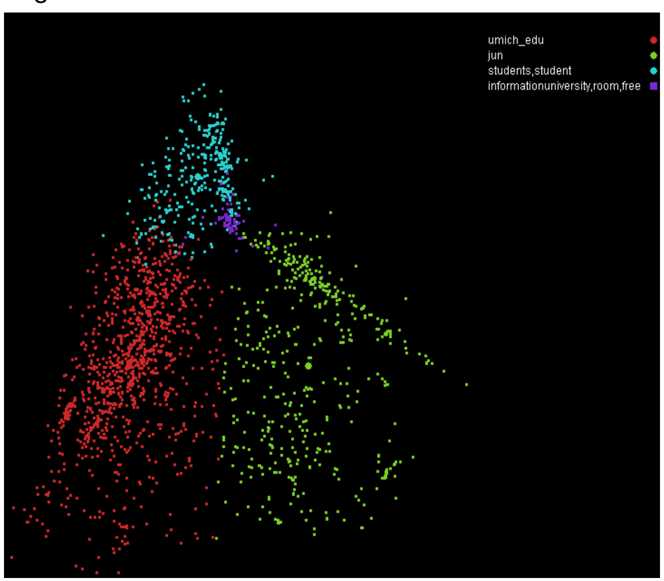


	FIRM	HQ
Recruiters	73	34
Average Density	5.41 (19.08)	11.02 (32.44)
	Network Constraint	**Information Diversity**
Partner	.24 (.14)	.59 (.12)
Consultant	.30 (.18)	.55 (.14)
Researcher	.32 (.20)	.55 (.15)
Mean	.29 (.17)	.57 (.14)

Measuring Information Diversity

@sinanaral





Vector Space Model

Represent each email as a multidimensional vector of term frequencies

Represent In-boxes and Out-boxes as collections of email vectors

Measure 'variance' of the vectors in someone's in-box or out-box

E.g. Variance of cosine distances between email vectors and information theoretic measures

Validate diversity measurement using public data from Wikipedia.org

Results

1. Diversity and Bandwidth Tradeoff… And both predict access to novel information.

2. Whether Brokers receive more novel information per unit time depends on the information environment in which the broker is situated.

    a) Greater Information Overlap – Bandwidth more valuable than diversity.

    b) Larger Topic Space – Bandwidth more valuable than diversity.

    c) Faster the Refresh Rate – Bandwidth more valuable than diversity.

Implications

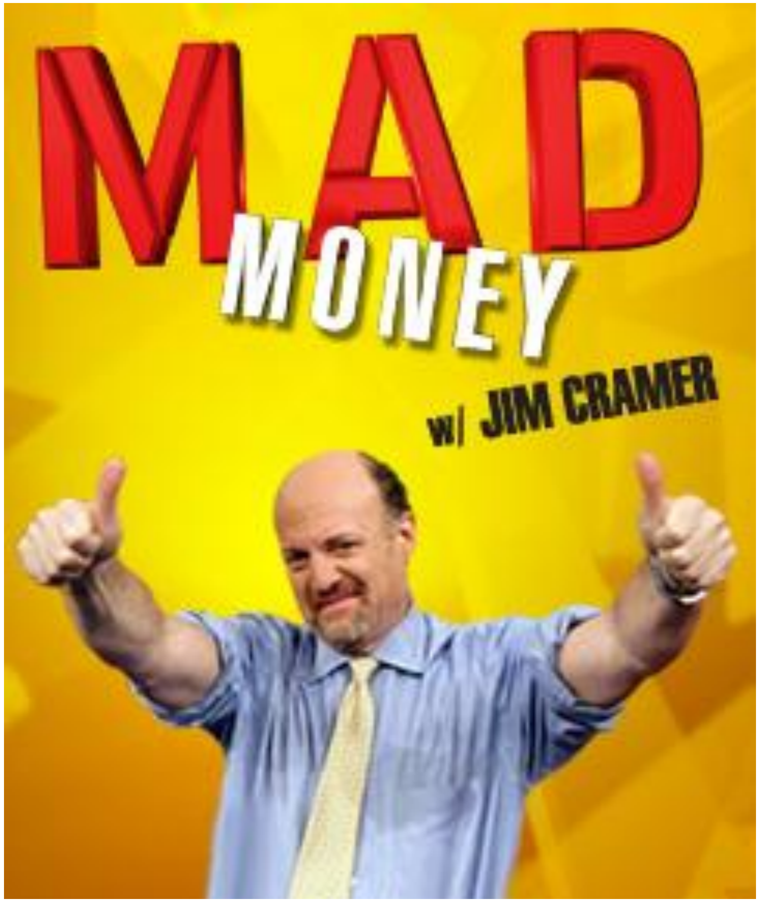
1. Strength of Weak Ties and Structural Hole Theory is *context dependent*.

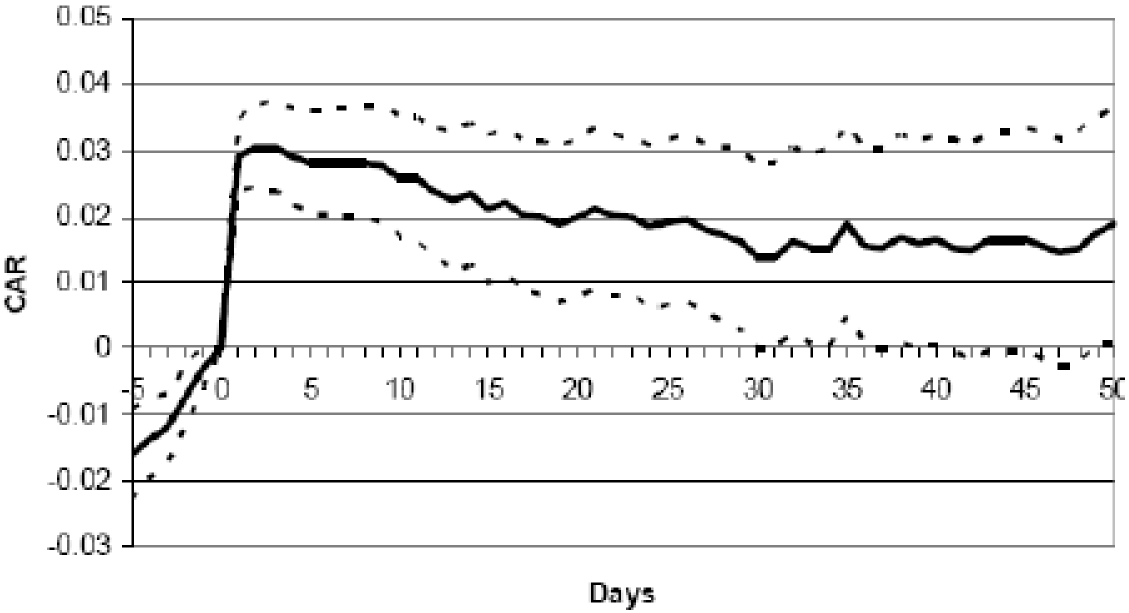
2. Wide Bridges vs. Thick Bridges – which is more important for Complex Contagions (Centola and Macy 2007)?

   - ✓ Wide Bridge – many ties – reinforcement from multiple parties.
   - ✓ Thick Bridges – few strong ties – reinforcement from a *trusted* party with *detailed and comprehensive* information about the behavior.
   - ✓ Which is more important? Under what conditions?
   - ✓ More research is needed.

"The Cramer Effect"

@sinanaral





Engelberg et al (2010), Neumann and Kenny (2007), Karniouchina et al (2009) Lim and Rosario (2010)

Aral, Ipeirotis & Taylor (2011) "Content and Context: Identifying the Impact of Qualitative Information on Consumer Choice." *NYU Stern Working Paper*.

"The Cramer Effect"

- LDA on of show transcripts – was Cramer "more persuasive" when he was making particular arguments.
- How prior knowledge and attention impacted his "influence"
- Collected Google Search, News, Trading history data for each stock in the two weeks before and after each recommendation (and ran LDA on the news as well).
- Found:

  1. *A Selection Effect* – Cramer chooses to recommend stocks trending in attention (news, search, trade volume) *just prior* to his recommendation.
  2. *Importance of Novelty* – Boosts his influence *and* this effect more important for some arguments than others.

Causality

Content

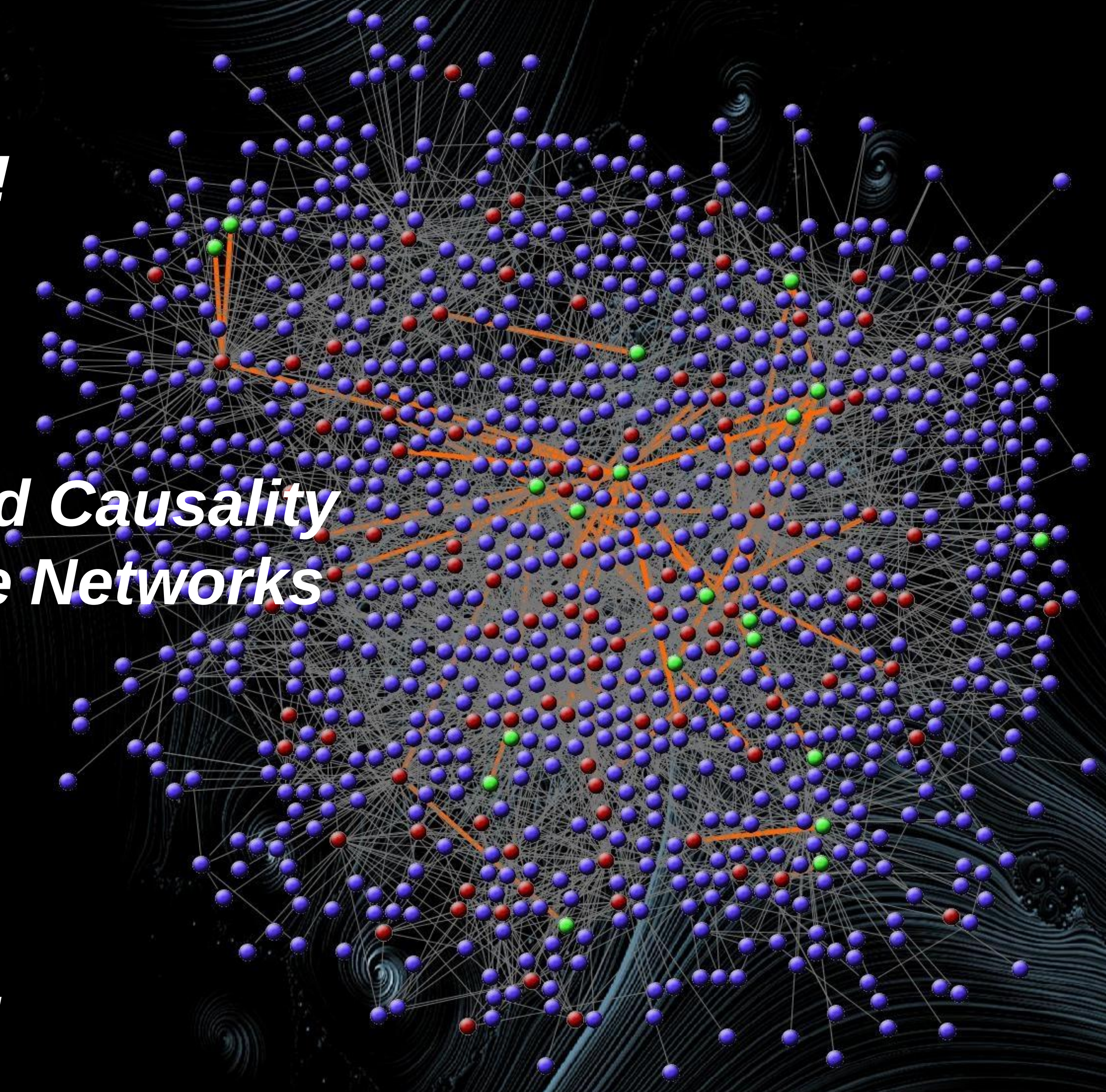
Collaborators:

Erik Brynjolfsson, Marshall Van Alstyne, Panos Ipeirotis, Arun Sundararajan, Lev Muchnik, Dylan Walker, Sean Taylor

Generous Funding and Data:

NSF, Microsoft, Yahoo, IBM, Facebook, Cisco, SAP, France Telecom, the Marketing Science Institute, the Institute for Information Innovation and Productivity

Thank You!



Content and Causality in Influence Networks

Sinan Aral
NYU
@sinanaral

	7	8	9	10
	Application Activity	Application Activity	Application Activity	Application Activity
	Beta (SE)	*Beta (SE)*	*Beta (SE)*	*Beta (SE)*
Viral State = Passive	.129* (.074)	.112 (.079)	.062 (.076)	-.037 (.074)
Viral State = Active	.190*** (.074)	.171** (.079)	.091 (.076)	-.006 (.074)
Degree	-.0001 (.0001)	-.0001 (.00	-.0002**	-.0002** (.0001)
Facebook		.054 (.0		.026* (.014)
				—
				.022*** (.001)
Invites				.055** (.024)
Number of Adopters			.607*** (.030)	.360*** (.031)
F Value (d.f.)	3.51*** (3)	4.87*** (4)		
R^2	.002	.003		
Observations	6310	5766		

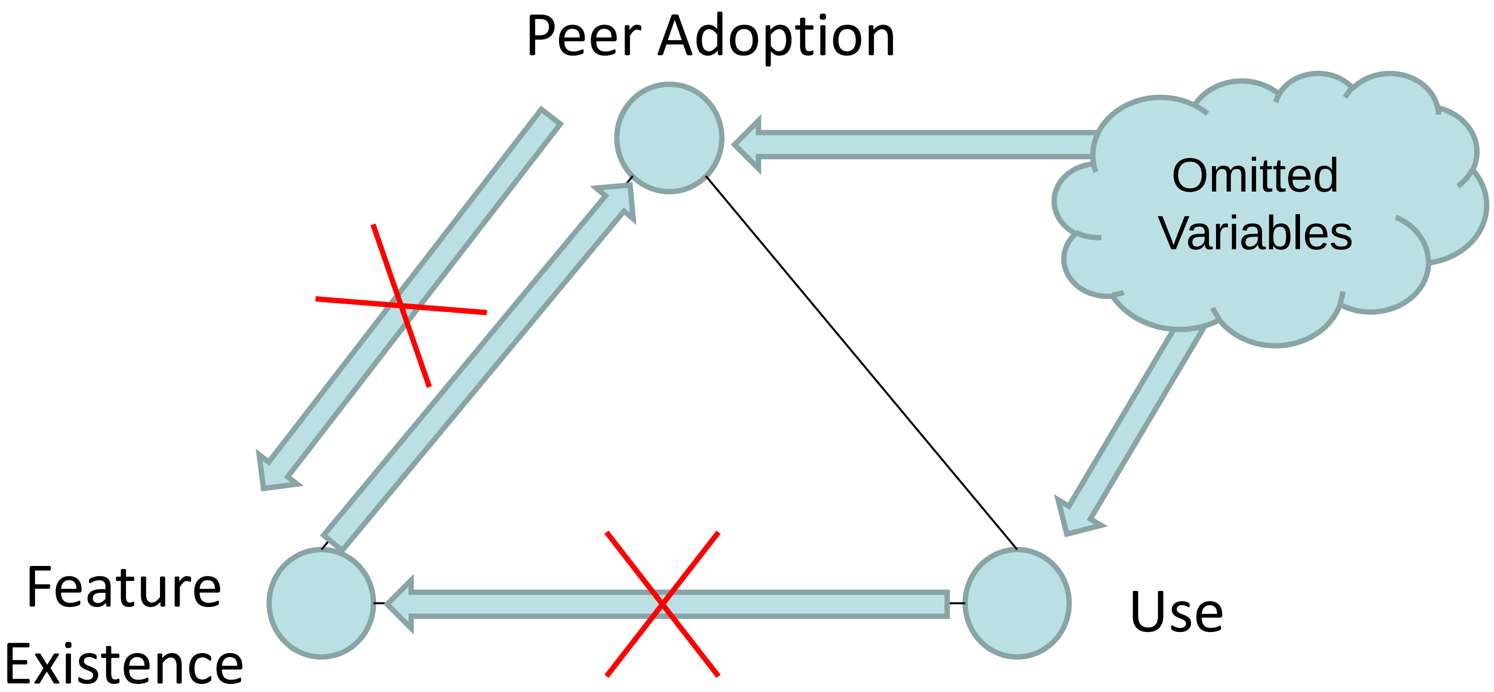
Notes: ***p<.(

Viral state is correlated with application activity

However, this relationship disappears when you control for number of adopters in your local network

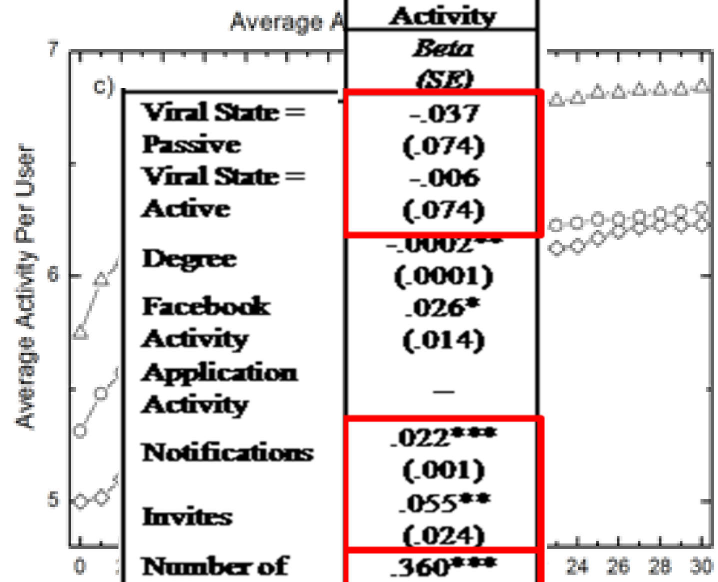
Number of adopters is still correlated with activity even when you control for use of viral features

Viral Features, Adoption and Use

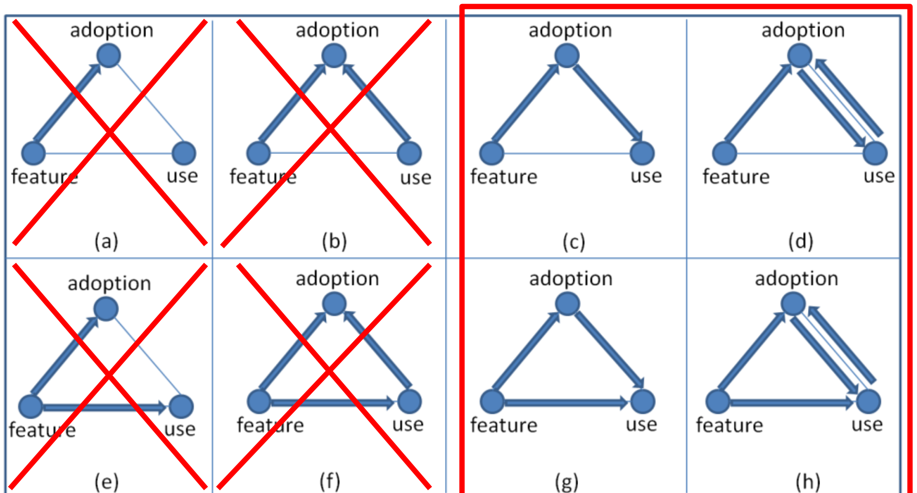


Adjudicating Alternate Explanations

- *Correlation between Peer Adoption & Application U… ld be evidence of Network Effects.*
- *It could also be explained by:*
  - *Unobserved Heterogeneity / Omitted Var…*
  - *Demand Effects – Existence of Features … e.*
- *But, treatment is randomized so omi… the variation we see in adoption anc… s.*
- *There could be an interaction effect … and a feature itself, but we observe a… adoption and application use contro…*



	Application Activity
	Beta (*SE*)
Viral State = Passive	-.037 (.074)
Viral State = Active	-.006 (.074)
Degree	-.0002** (.0001)
Facebook Activity	.026* (.014)
Application Activity	—
Notifications	.022*** (.001)
Invites	.055** (.024)
Number of Adopters	.360*** (.031)



1. (a) and (b) inconsistent with the discrepancy in app use between treatment groups.
2. If Demand Effects: Controlling for them should remove any spurious correlation between peer adoption and use.
3. Controlling for feature use, peer adoption still highly correlated with app use.

Date: Sun, 01 Feb 2009 10:02:10 -0500
From: xxx@yyy
To: abc@123
CC: zzz@yyy,
abc@zzz
Subject: ~~Re:~~ IWP 1 Extended Thoughts

Actually,

~~To be~~ even more succinct about 3 main take aways:

1. Given ~~our~~ ability ~~to~~ make connections between abstract concepts, ~~our~~ productivity ~~is~~ determined more ~~by our~~ ability ~~to~~ multitask, than ~~by our~~ ability ~~to~~ conduct sequential work faster.

~~So~~, lets explore ~~the~~ mechanisms behind multitasking ~~a~~ bit more:

2. ~~The~~ relationship between output ~~and~~ multitasking ~~is~~ convex ~~at~~ low levels ~~of~~ multitasking ~~and~~ concave ~~at~~ high levels ~~of~~ multitasking (~~Because~~ information inputs ~~are~~ non-rival ~~and~~ complementary, unlike physical inputs, ~~their~~ use enables convexity ~~in the~~ relationship between multitasking ~~and~~ output ~~at~~ low levels ~~of~~ multitasking. Because human information processing ~~is~~ constrained ~~by~~ bounded rationality, ~~and~~ limited cognitive capacity ~~the~~ relationship between multitasking ~~and~~ output ~~is~~ concave ~~at~~ higher levels ~~of~~ multitasking).

~~So~~, how ~~do we~~ acquire ~~the~~ inputs ~~we~~ use?... Socially:

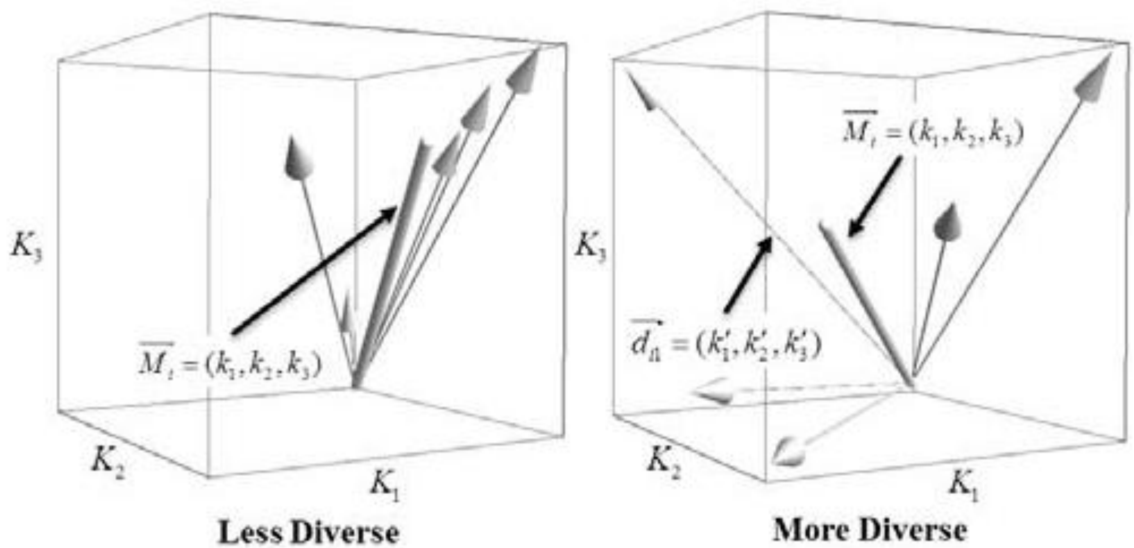
3. Efficient positioning ~~in the~~ social network creates efficient means ~~to~~ gather ~~and~~ use information ~~and is~~ correlated ~~with~~ higher productivity. [~~Because we~~ require ~~a~~ social support system ~~of~~ information acquisition (embodied ~~in our~~ social networks) which ~~we~~ rely ~~on to~~ extend ~~our own~~ individual mental capacity. ~~We~~ gather information inputs socially (~~and~~ through IT ~~which we use as a~~ control variable)]

Dr. XXX

Construction of E-mail Vectors

1. Header information is extracted to create the social network. Names are matched and identities are validated by hand.
2. The subject and body of the e-mail message are analyzed to extract frequencies of use of keywords (Steps 3-6).
3. Stop words (e.g. "a, "an," "the," "and," and other common words) with high frequency across all e-mails are removed as shown by words that have been struck.
4. Keywords are extracted based on the three principles outlined on pages 30-31 of the manuscript.
5. Keywords are root-stemmed, such that for example "multi-task," "multitasking," become "multitask*."
6. The frequency of each key word is counted and recorded.
7. A vector representing the e-mail is created which logs the e-mail ID, the ID number of each keyword used and the frequency of use of each keyword noted inside brackets as follows: <E-mailID7842B|748821<9>; … ; 849247<2>|>
   A vector representing the example e-mail to the left is shown in truncated form below.
8. The content similarity of e-mail vectors is then compared using several standard distance metrics such as the Cosine distance.

$$\overrightarrow{d_i} = (IWP<1>; connection^*<1>; productivity<2>; multitask^*<9>; sequential<1>; output<3>; ...; input^*<4>; social^*<5>; IT<1>; control<1>; variable<1>)$$



The diversity of the information in an email inbox or outbox is measured by calculating the Cosine Distance of each email vector d_i to the mean vector of that inbox or outbox M_i, and then averaging across all emails:

$$CosDist = 1 - Cos\left(d_{ij}^{I}, M_i^{I}\right)$$

$$ID_i^I = \frac{\sum_{j=1}^{N}\left(1 - Cos\left(d_{ij}^{I}, M_i^{I}\right)\right)^2}{N}$$

Identifying Social Influence in Networks

- "Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science*
- "Distinguishing Influence Based Contagion from Homophily Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences, Dec. 22, 2009, vol. 106, no.51.*
- "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks." Forthcoming in *Management Science*
- "The Diversity-Bandwidth Tradeoff" Forthcoming in *American Journal of Sociology*

Five Broad Research Questions

1. What is peer influence (formally)?
2. How do characteristics of the product or behavior affect peer influence and contagion?
   - What characteristics makes a viral video go viral?
   - How do we design viral products?
   - How do network externalities inherent in a product affect its diffusion?
3. What is the role of sustained use in creating sustainable contagions?
   - Do users need to adopt and maintain interest or use or just adopt once?
   - What is the role of churn? Engagement?
4. How do distributions of individual characteristics over network nodes affect contagion?
   - How does assortativity / homophily affect propagation of influence?
   - Do persuasive individuals tend to be of high degree?
   - Do influentials tends to be surrounded by susceptibles or do influentials cluster?
5. Are their 'systems' of complementary contagion management strategies?
   - How do referral incentives and targeting interact? Are they complements or substitutes?

Viral Feature Randomization

We specify three types of treatments that we refer to as the Viral Feature State of a user:

- Non-Viral User (Vi = 0): Both passive and active viral features are disabled (baseline)

  [5% of total user population]

- Passive Viral User (Vi = 1): Only passive viral features (e.g., notifications) are enabled

  [47.5% of total user population]

- Active Viral User (Vi = 2): Both passive viral features and active viral features (e.g., invites) are enabled

  [47.5% of total user population]

Findings

1. ***Viral Product Design Features produce econometrically identifiable peer influence and social contagion effects.***

2. ***Active-Personalized Viral Features are more effective per message but are used much less often and so produce less total peer adoption in the network than Passive-Broadcast features.***

3. ***Our data are consistent with the existence of positive network externalities that drive a feedback loop of peer adoption and sustained product use.***

4. ***Viral Feature outperform traditional banner ads and email campaigns in generating adoption.***

NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN SCHOOL OF BUSINESS

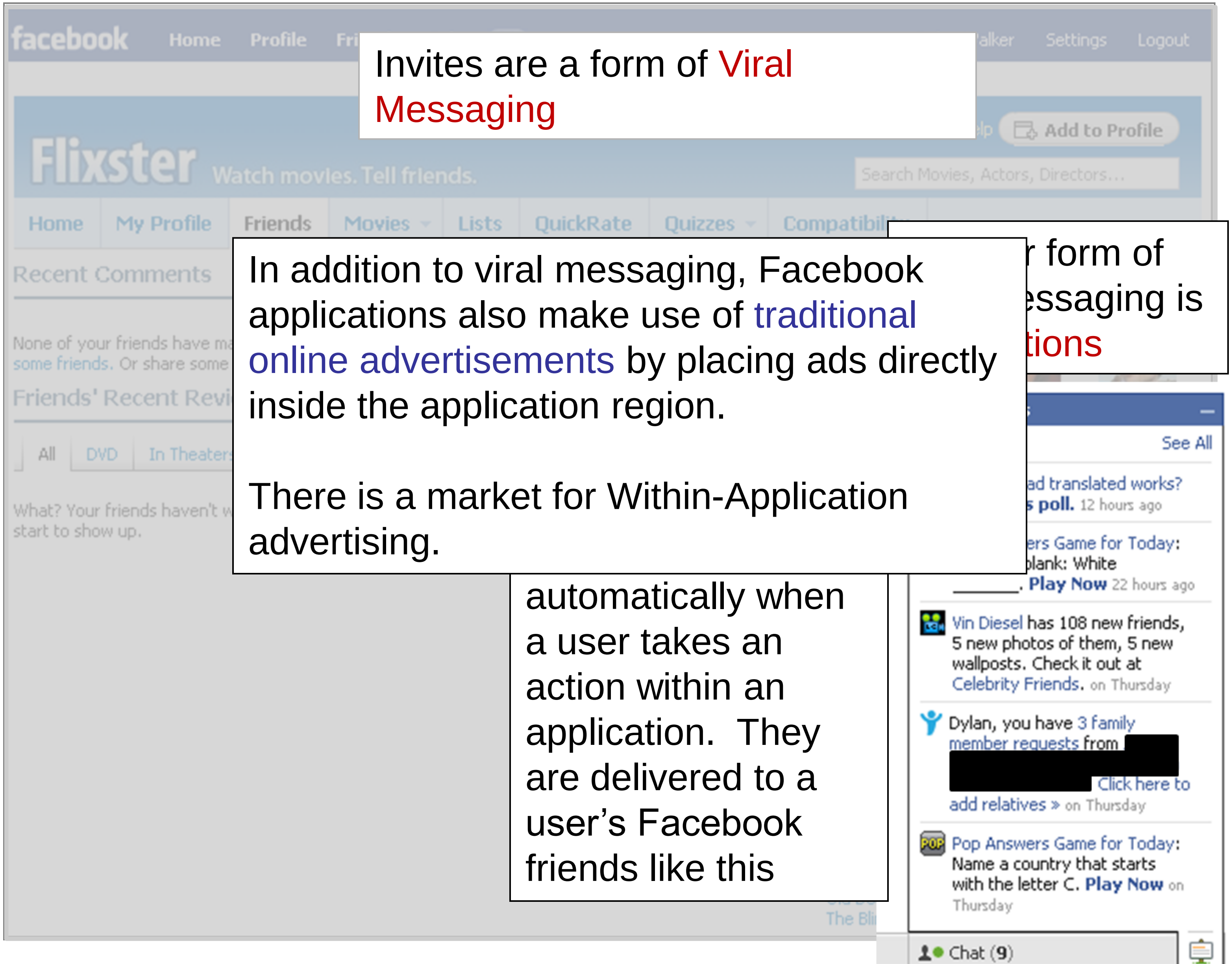
Table 4: Variance-Corrected Proportional Hazards of Contagion in Networks of Baseline, Passive and Active Treatment Groups

	1	2	3	4
	Hazard Ratio (SE)	*Hazard Ratio (SE)*	*Hazard Ratio (SE)*	*Hazard Ratio (SE)*
Viral State = Passive	3.46*** (1.18)	3.35*** (1.15)	2.50** (.86)	2.51** (.86)
Viral State = Active	4.44*** (1.64)	4.21*** (1.56)	3.33*** (1.24)	3.31*** (1.24)
Application Activity		1.02*** (.004)	1.02*** (.003)	1.02*** (.003)
Notifications			1.02*** (.002)	1.02*** (.002)
Invites				1.06** (.028)
Log Likelihood	-4694.359	-4631.795	-4544.845	-4542.577
X^2 (d.f)	19.34*** (2)	57.41*** (3)	298.78*** (4)	307.47*** (5)
Observations	3929	3929	3929	3929

Notes: ***p<.001; **p<.05; *p<.10;

Click Stream Data Corroborate Results

Table 5: Click Stream Analysis of Responses to Viral Messages and Adoption			
	1	2	3
	Messages Sent	*Adoptions via Click Through Installation*	*Adoption Rate (Marginal Impact)*
Invitations	160	16	.10
Notifications	69980	666	.01



Invites are a form of Viral Messaging

In addition to viral messaging, Facebook applications also make use of traditional online advertisements by placing ads directly inside the application region.

There is a market for Within-Application advertising.

r form of
essaging is
tions

automatically when a user takes an action within an application. They are delivered to a user's Facebook friends like this

Which Features Spread Contagion Best?

@sinanaral

	Personal Invitations	Passive Awareness
Influence Per Message	↑6%	↑2%
Global Diffusion	↑98%	↑246%
Stickiness	↑17%	0%

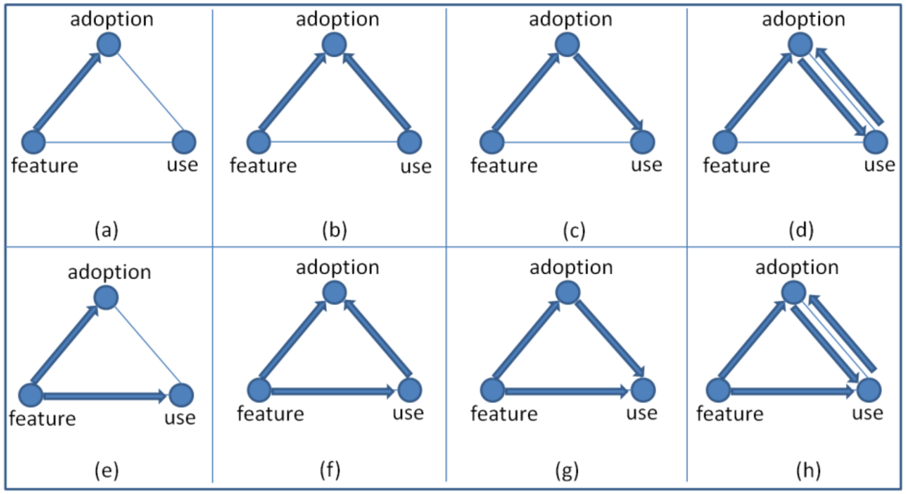
NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN SCHOOL OF BUSINESS

	7	8	9	10
	Application Activity	Application Activity	Application Activity	Application Activity
	Beta (SE)	*Beta (SE)*	*Beta (SE)*	*Beta (SE)*
Viral State = Passive	.129* (.074)	.112 (.079)	.062 (.076)	-.037 (.074)
Viral State = Active	.190*** (.074)	.171** (.079)	.091 (.076)	-.006 (.074)
Degree	-.0001 (.0001)	-.0001 (.0001)	-.0002** (.0001)	-.0002** (.0001)
Facebook Activity		.054*** (.016)	.042*** (.015)	.026* (.014)
Application Activity				—
Notifications				.022*** (.001)
Invites				.055** (.024)
Number of Adopters			.607*** (.030)	.360*** (.031)
F Value (d.f.)	3.51*** (3)	4.87*** (4)	83.54*** (5)	128.92*** (7)
R^2	.002	.003	.07	.14
Observations	6310	5766	5766	5766

Notes: ***p<(

Adjudicating Alternate Explanations

- ***Correlation between Peer Adoption & Application Use could be evidence of Network Effects.***
- ***It could also be explained by:***
  - ***Unobserved Heterogeneity / Omitted Variables***
  - ***Demand Effects – Existence of Features make Product more interesting to use.***
- ***But, treatment is randomized so omitted variables cannot explain the variation we see in adoption and use across treatment groups.***
- ***There could be an interaction effect between an omitted variable and a feature itself, but we observe a correlation between peer adoption and application use controlling for feature use.***



(a) (a) and (b) inconsistent with the discrepancy in app use between treatment groups.

(b) If Demand Effects: Controlling for them should remove any spurious correlation between peer adoption and use.

(c) Controlling for feature use, peer adoption still highly correlated with app use.

λ_{0k}

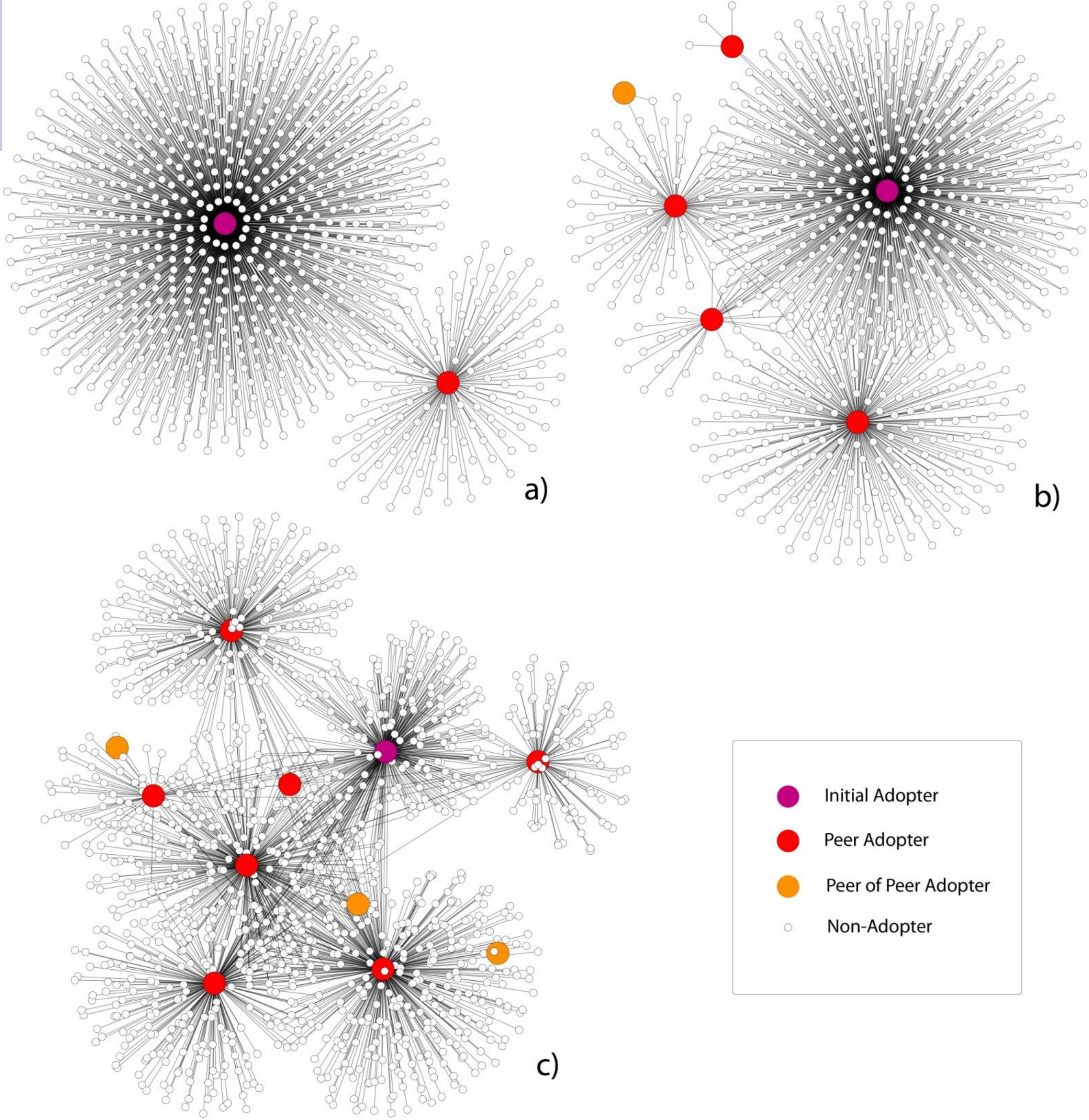
Table 6: Baseline Hazards Over *k* Events (*k* = 1…6)

	1	2	3	4
	Mean (SD)	*Min*	*Max*	*N*
	.0002 (.0001)	.0001	.001	523
	.002 (.001)	.001	.013	128
	.015 (.024)	.005	.14	42
	.034 (.010)	.021	.054	20
	.046 (.008)	.037	.067	15
	.099 (.044)	.053	.14	7

WIN Workshop <www.winworkshop.net>

Jon Kleinberg (CS)
Matt Jackson (Economics)
Ron Burt (Sociology)
James Fowler (Pol Sci)
Lazslo Barabasi (Physics)
Michael Kearns (CS)
Sanjeev Goyal (Economics)
Michael Macy (Sociology)
Duncan Watts (Physics)
Sandy Pentland (CS)
David Lazer (Pol Sci)
Bernardo Huberman (Physics)
Christos Faloutsos (CS)

The purpose of WIN: "to *bring together leading researchers studying 'information in networks'* to build a *lasting multidisciplinary research community*."


a)
b)
c)
Initial Adopter
Peer Adopter
Peer of Peer Adopter
Non-Adopter

Influence Prompted by Viral Messaging

@sinanaral

Viral Feature Randomization

@sinanaral

User in **Experimental Group**

User in **Control Group**

	Baseline (N=405)	Passive (N=4600)	Active (N=4682)	*t*-statistic (B-P)	*t*-statistic (B-A)	*t*-statistic (P-A)
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*t-statistic (SE)*	*t-statistic (SE)*	*t-statistic (SE)*
Age	31.51 (13.80)	30.81 (13.31)	29.94 (13.27)	.46 (13.35)	1.03 (13.31)	1.45 (13.24)
Gender (1=Male)	.25 (.44)	.33 (.47)	.32 (.47)	-1.57 (.47)	-1.42 (.46)	.40 (.47)
Degree†	171.79 (223.88)	170.25 (278.64)	166.97 (248.77)	.09 (275.13)	.32 (247.15)	.55 (263.82)
Number of Facebook Wall Posts	40.52 (79.89)	36.45 (94.16)	37.07 (246.76)	.46 (93.11)	.15 (238.20)	-.09 (188.31)

Table 3. Summary Statistics and Mean Comparisons of Active, Passive and Baseline Users						
	1	2	3	4	5	6
	Baseline (N = 405)	Passive (N = 4600)	Active (N = 4682)	*t*-statistic (B-P)	*t*-statistic (B-A)	*t*-statistic (P-A)
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*t-statistic (SE)*	*t-statistic (SE)*	*t-statistic (SE)*
Number of Adopters in User's Local Network	.01 (.12)	0.07 (.35)	0.10 (.44)	-2.84*** (.34)	-3.60*** (.43)	-3.64*** (.40)
Percentage of Adopters in User's Local Network	.02 (.002)	.09 (.01)	.15 (.01)	-1.92* (.01)	-2.35** (.01)	-2.83*** (.01)
Maximum Diffusion Depth	.01 (.11)	.04 (.22)	.05 (.24)	-2.53* (.21)	-3.01*** (.24)	-1.98*** (.23)

NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN

Creating Social Contagion through Viral Product Design

A Randomized Field Experiment

Sinan Aral,
NYU Stern & MIT

Dylan Walker,
NYU Stern

NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN

Randomized Peer Influence Experiments in Online Social Networks

WISE 2009
Phoenix, AZ

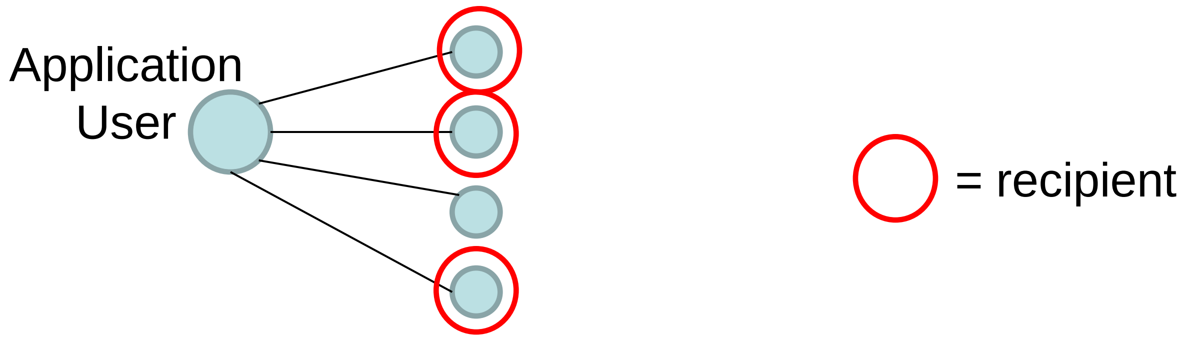
Sinan Aral,
NYU Stern & MIT

Dylan Walker,
NYU Stern

Randomization of Passive Viral Messages

@sinanaral

- Notifications are typically one-to-many messages
- The recipients of notification are randomly selected from the set of the sender's social network peers.



A new set of random recipients are selected for each notification

- We record when notifications are sent and to whom
- We record recipient responses to notifications
  - If/when the user clicks on the notification, leading to an application installation page
  - If/when the user proceeds to install the application

Hypothesis 1 (H1). *Enabling viral product design features increases the likelihood of adoption among peers of current users.*

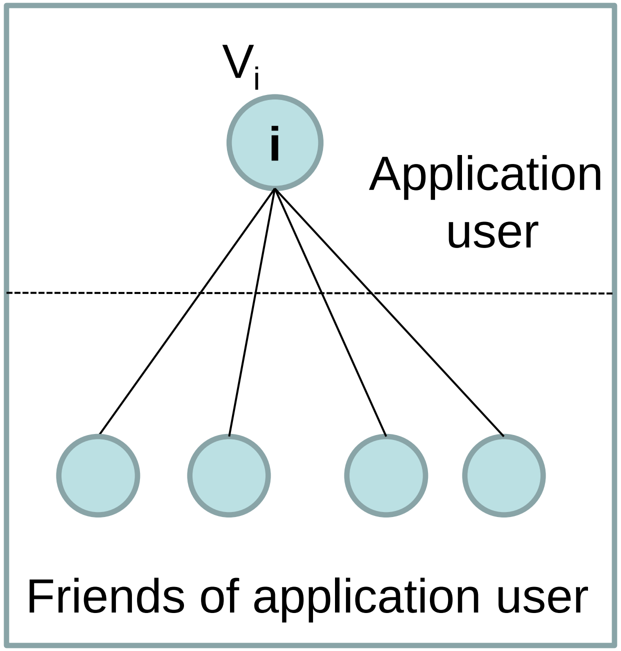
Hypothesis 2 (H2). V*iral product design features that require more activity on the part of the user and are more personalized to recipients create greater marginal increases in the likelihood of adoption per message.*

Hypothesis 3 (H3). V*iral product design features that require more activity on the part of the user and are more personalized to recipients generate fewer total viral messages.*

Enables Examination of Numerous Hypotheses

- Influence conditional on Viral Features
  - Active Viral Messages
  - Passive Viral Messages
  - Traditional Advertizing
- Influence conditional on Viral Features and
  - Characteristics of Sender
  - Characteristics of Recipient
  - Characteristics of Dyad
  - Network Properties of Sender
  - Network Properties of Recipient
  - Network Properties of Dyad

Influence conditional on Sender Attributes and Viral Features



$\{X_i\}$ Attributes of user **i**

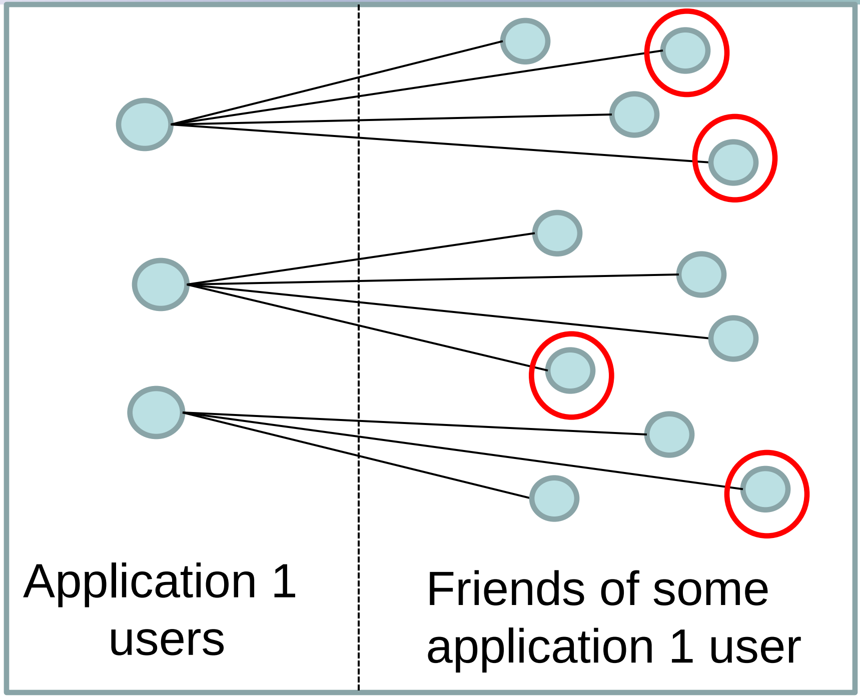
V_i Viral Feature State of user **i**

$$F(\,friends\;\;of\;\;i\;|V_i,\{X_i\})$$

Fraction of friends that adopt Application 2 conditional on the Viral Feature State of user **i** and the attributes of user **i**

Notice, this measure will pick up if some attributes of user **i** make him more or less influential

Susceptibility to Influence via Viral Messaging



$\{X_j\}$ Attributes of friend **j**

$R_j = 1$ iff friend **j** received a viral message

$$F(\,friends\ \ j \mid R_j = 1, \{X_j\})$$

Fraction of all friends of some Application 1 user that adopt Application 1 conditional on the receipt of a viral message and the attributes of friend

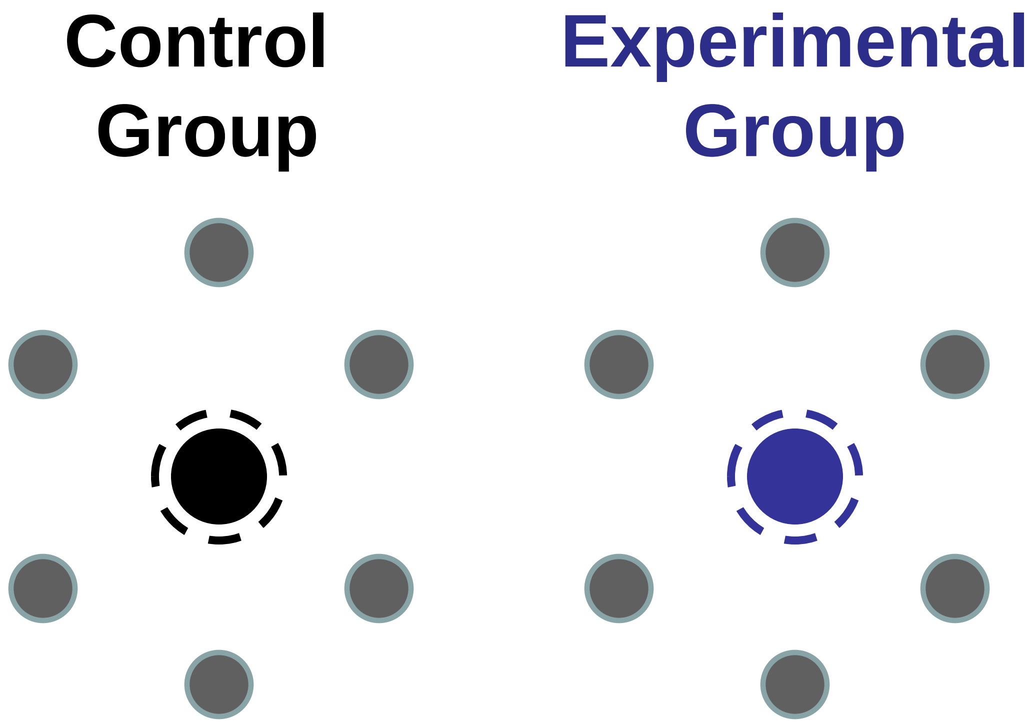
Notice, this measure will pick up if some attributes of a friend **j** make him more or less easily influenced

Experiment 2: Randomized Traditional Advertizing

**Randomly Showed Traditional Banner Ads to
Some (Experimental) Application Users**

Experiment 2: Randomized Traditional Ads

@sinanaral



Experimental Group

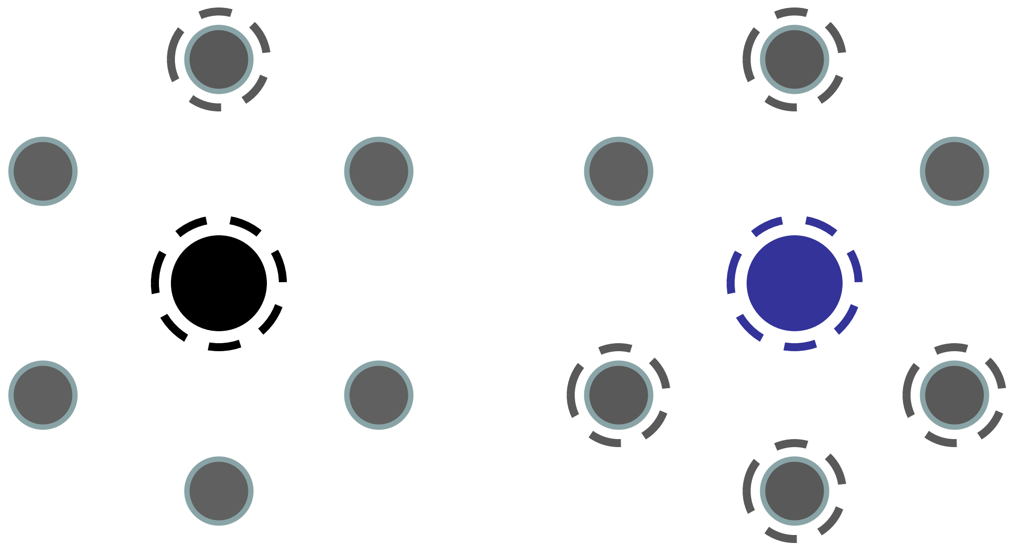
Sees a Banner Ad for Application 2

Control Group

Sees No Banner Ad

Experiment 2: Randomized Traditional Ads

Control Group | **Experimental Group**



We then compare:

✓click throughs
✓adoption
✓usage
data

of neighbors of:

✓Experimental Group
✓Control Group

Allows us to test:

1. Average Treatment Effect of Traditional Advertizing on Peer Adoption and Network Propagation
2. A focus on latent communication channels

II. Influence Prompted by Traditional Advertising

@sinanaral

Hey Dylan, have you heard of this “Application 2”? It gets you citations.

I like citations. I’ll try it out.

As in this example, some channels of communication between friends are latent.

Facebook Applications

Facebook applications

- are independently developed web applications
- run within the framework of Facebook
- are integrated with the social functions of Facebook:
  - can utilize the inter-user communication channels and knowledge of the social network provided by Facebook,

- When a users elects to adopt or install a Facebook application
  - grants the application access to:
    - the user's immediate social network
    - the user's profile data and the profile data of their immediate social network peers (according to their privacy settings)

A two-application approach to experimentation

@sinanaral

Notice, that when Prof. Mendelson visits Application 1, it **doesn't show him an advertisement**, because he's not in the treatment group

This idea can be generalized.

You can imagine using Application 1 as a delivery system for **ultra-targeted** advertisements, or to select treatment populations based on specific user attributes.

I won't discuss that here.

We use Application 1 to deliver advertisements for a second application to randomly selected users.

We then observe the Application 2 adoption amongst the peers of Application 1 users

Simultaneous Experiments

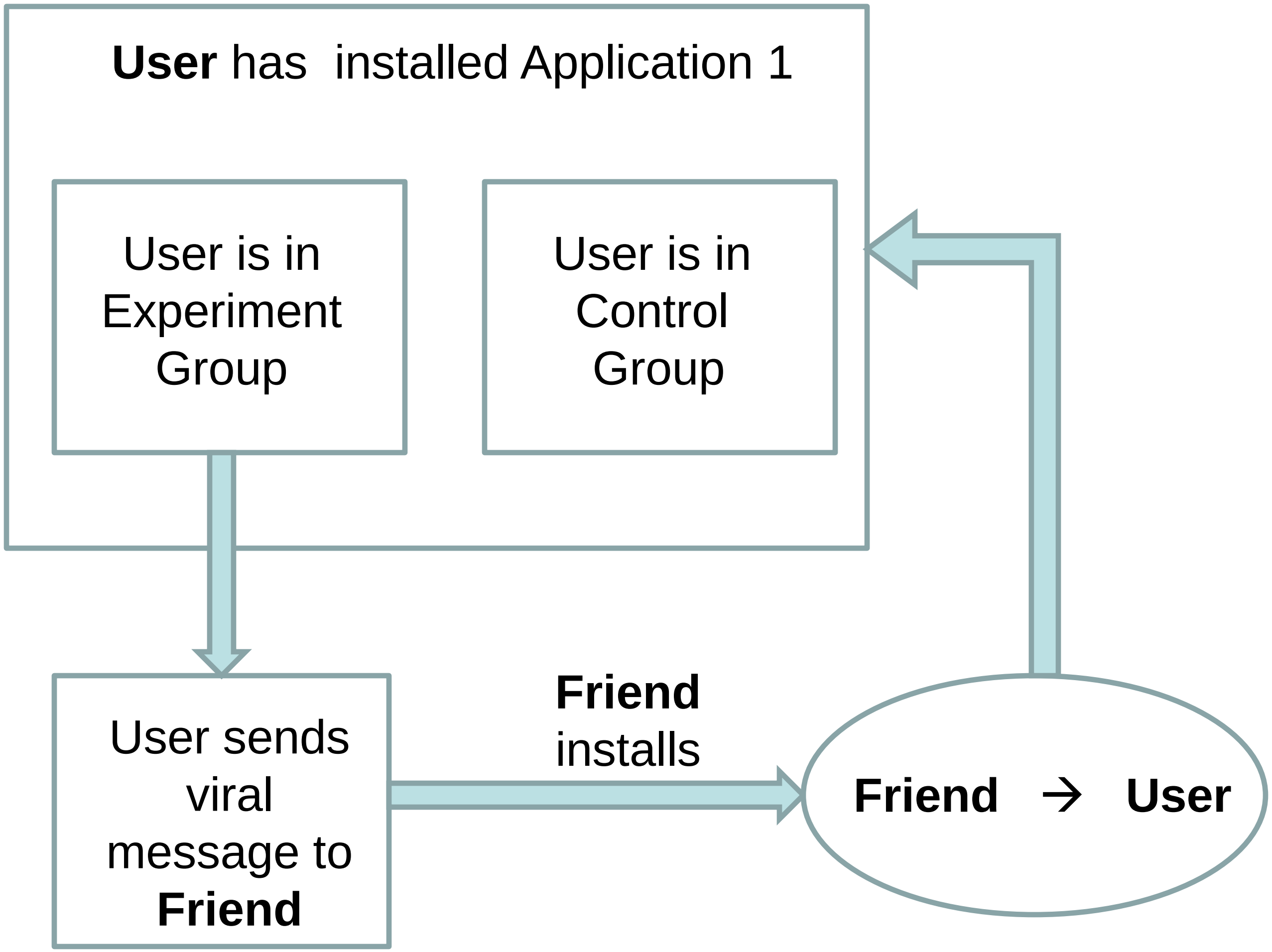
We perform both experiment I and II simultaneously, because application users are a precious resource.

The possible treatment an Application 1 user receives are given by:

	Non Viral	Passive Viral	Active Viral
Ad	2.5%	23.75%	23.75%
No Ad	2.5%	23.75%	23.75%

Potential Issues: Viral Adoption Chains

@sinanaral



Are there differences between:

- Users who became users through "initial advertising"
- Users who became users through viral messaging

Potential Issues – Selection Bias

At least two levels of potential selection bias:

1. Selection into experiment – Adoption of App 1.
   - Here we are careful about how representative our sample is of the population of Facebook users.
   - We can test observable characteristics.
   - But there may be unobserved differences.

2. Selection into the treatment – sending of viral messages (choosing to actively invite friends)
   - Here we are careful about whether those who choose to message are systematically different.
   - We test the average treatment effect of having messaging turned on – that's random.
   - This is analogous to traditional selection bias in randomized trials. There are known tests and methods

Potential Issues – The Application Matters

- Whether users choose to invite and whether friends of users respond depend on the applications we use.
  - More engaging, more naturally "viral" applications may induce different usage behavior and different responses from recipients of viral messages.
  - Some applications have network externalities.
  - We use representative, engaging, real world applications (not developed solely for the experiment).
  - We carefully consider how our applications compare to others e.g. in how viral they are, if they have network externalities etc.

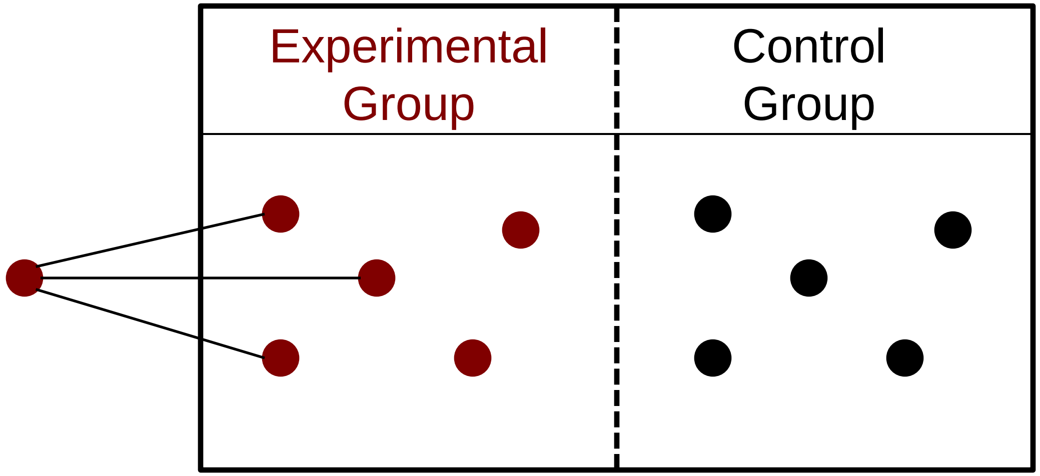
Potential Issues

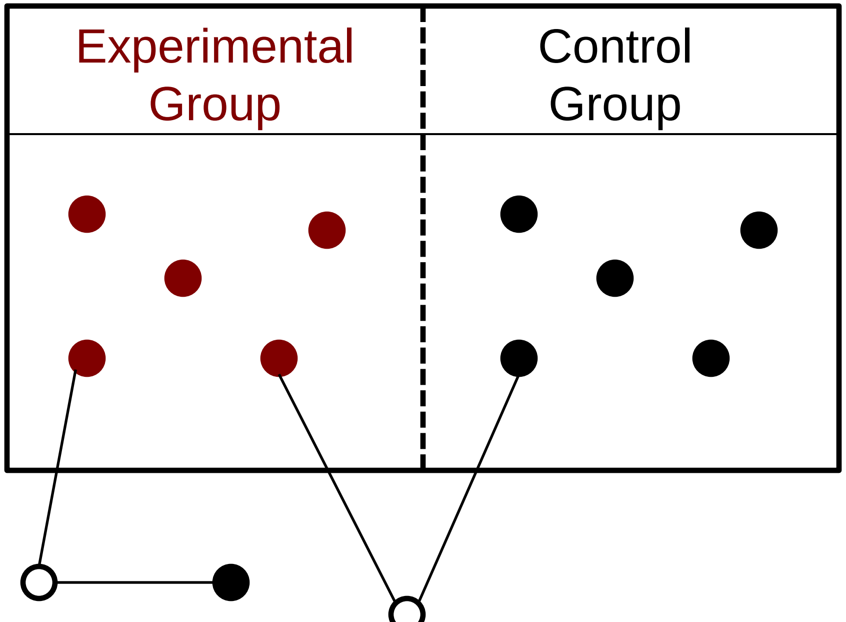
@sinanaral

Friends of multiple treated users

We group peers with n treated friends together for modeling.

Peers are not double counted





Friends of users in experiment and control

Depending on the nature of the treatment, simultaneous peers of users in the experiment and control group may be unclearly classified.

In some treatment cases, simultaneous relationships control group users is of no consequence.

Potential Issues – Treatment Echo Effect

- Which viral message are users responding to?

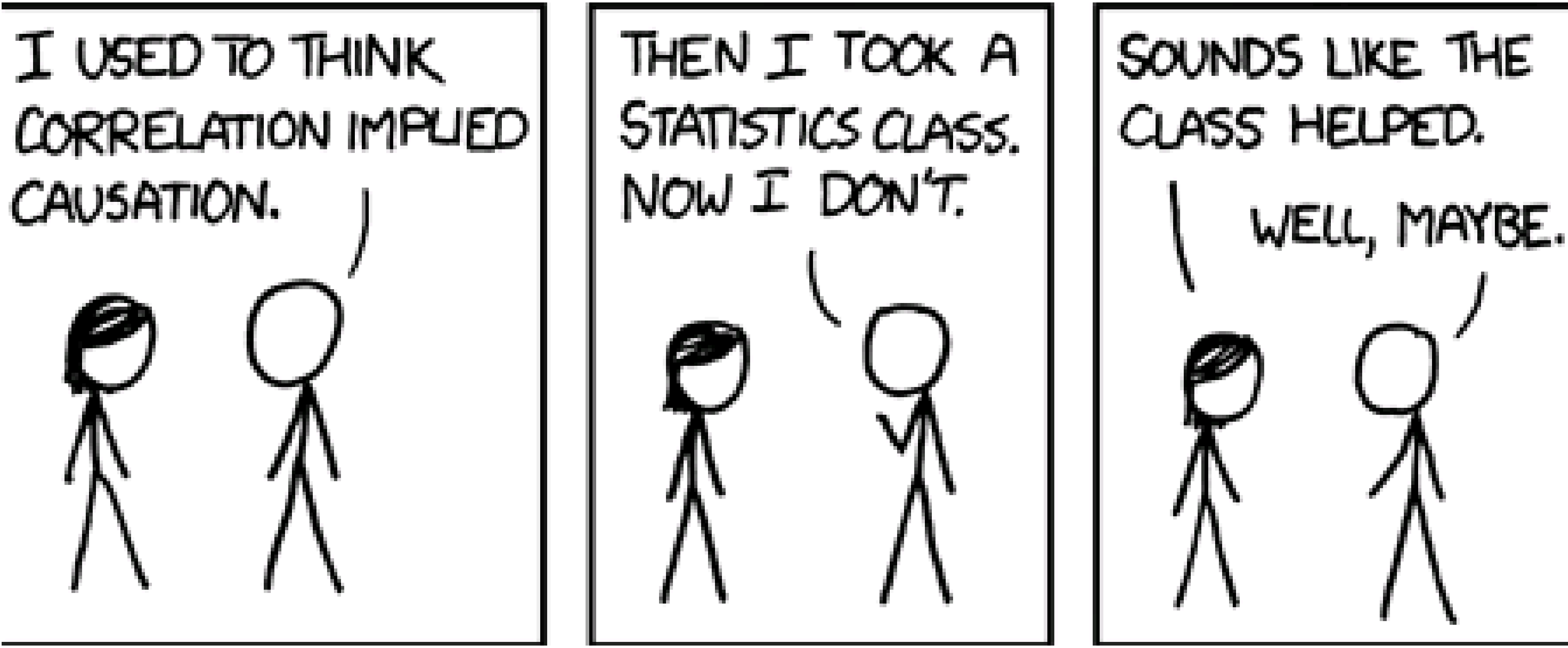
    - It could be the case that some peers respond positively to a viral message only after they have received multiple viral messages from the same or many different friends.

    - Even when a user doesn't click on a viral message, they may be affected by it.
        - Complex Contagion Model (Centola and Macy 2005)

  – Because we observe when viral messages are received, we can distinguish 1st, 2nd, … , Nth time recipients receive a viral message (and from whom) and classify them accordingly.

Causal Statistical Estimation

@sinanaral



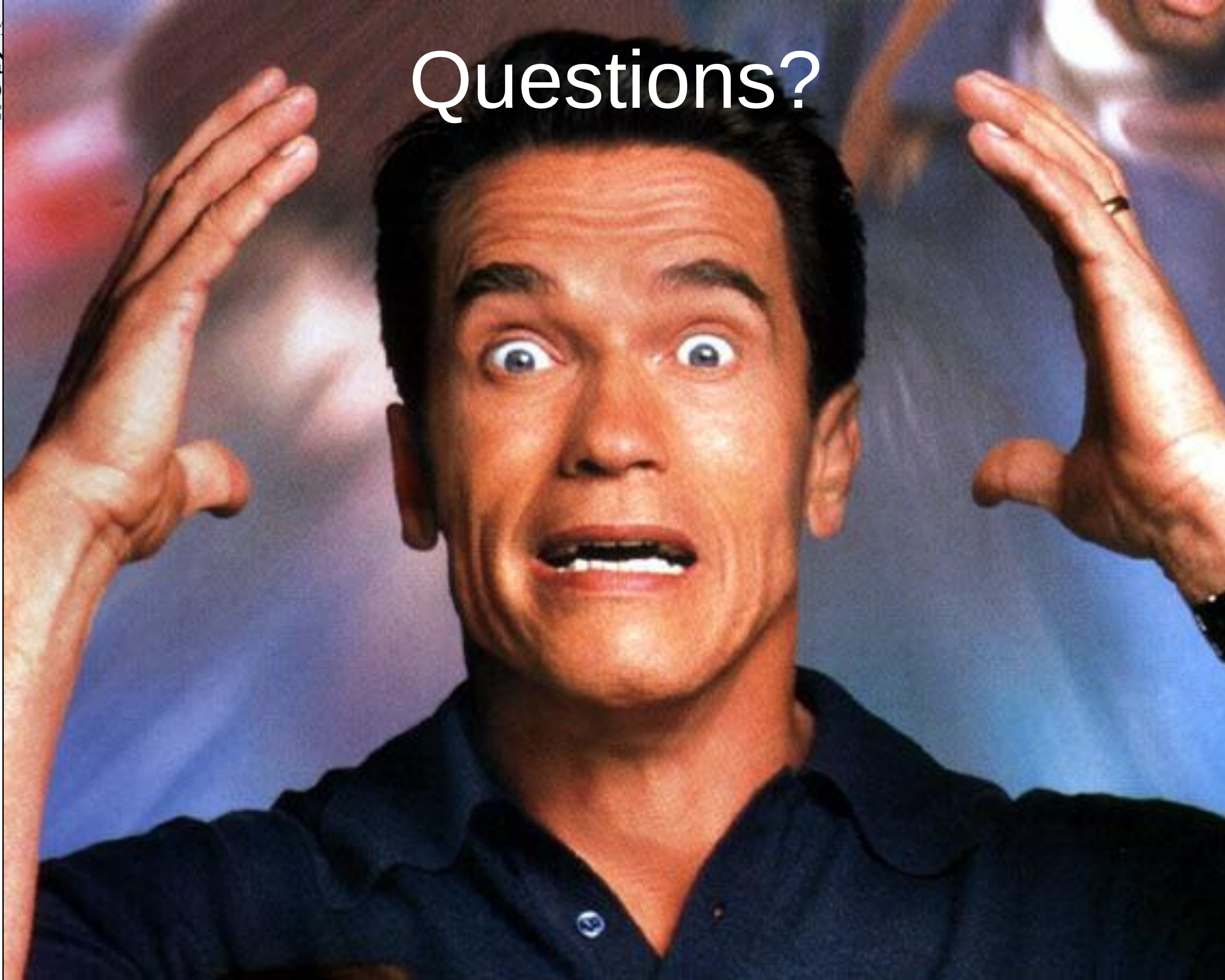
NEW YORK UNIVERSITY
NYU STERN
LEONARD N. STERN SCHOOL OF BUSINESS

Experiments are Running As I Speak

Stay Tuned!

Thank You!!

Questions?



Potential Issues – Exact Conditional Probabilities

- We have to be very aware of exactly what the response is conditional on and characterize results accordingly…
- Examples:
  - Profile Box – Influence conditional on installing the box and 'recipients' surfing the home page of the user.

Potential Issues – Exact Conditional Probabilities

- We have to be very aware of exactly what response is conditional on …
- Examples:
  - Profile Box – Influence conditional on installing the box and ‘recipients’ surfing the home page of the user.

Potential Issues – Selection Effect on Peer Response

- We have to be very aware of exactly what the response is conditional on and characterize results accordingly…
- Examples:
  - Profile Box – Influence conditional on installing the box and 'recipients' surfing the home page of the user.

Current Status of Experiments

- We have data on millions of Facebook users
- experiment has been running for 1 week
- we have 3500 application users, hundreds of thousands of peers
- Thousands of viral messages have been sent
- we plan to continue the experiment for 5-6 weeks
- stay tuned

Acknowledgements

- Collecting data of this scale involves significant technical challenges
  - Hundreds of Gigabytes of data being collected live
  - Weeks of data processing
  - This would not have been possible without the “Stern Cloud” created and operated by the Stern Center for Research Computing
  - We are indebted to Norman White for his extremely timely assistance

Privacy and Human Subjects Safeguards

@sinanaral

1) Privacy safeguards
   - Facebook users are given a large number of privacy options that control the visibility of their data
   - Facebook users explicitly grant permission to applications
   - All user data that we collect are de-identified
2) Experimental data on human subjects
   - Our experimental application features involve **minimal harm**
   - Our experiments are analogous to A/B tests involving application features. Standard tests such as these are performed regularly by online service vendors to improve user experience.
3) Facebook Terms of Service
   - In accordance with Facebook TOS, direct profile data is not retained
   - We retain and work with **derived** data only

return

Extra Slide: Profile Data

- Demographics
  - age,sex, current and hometown location, education and employment history,
- Relationships and Social Interactions
  - Facebook friends, family members, relationship status to significant other, social group participation, photo co-appearance
- Preferences
  - religious views, political views, activities, interests,
- Product Tastes
  - movies, music, book, tv shows, Facebook application adoption

return

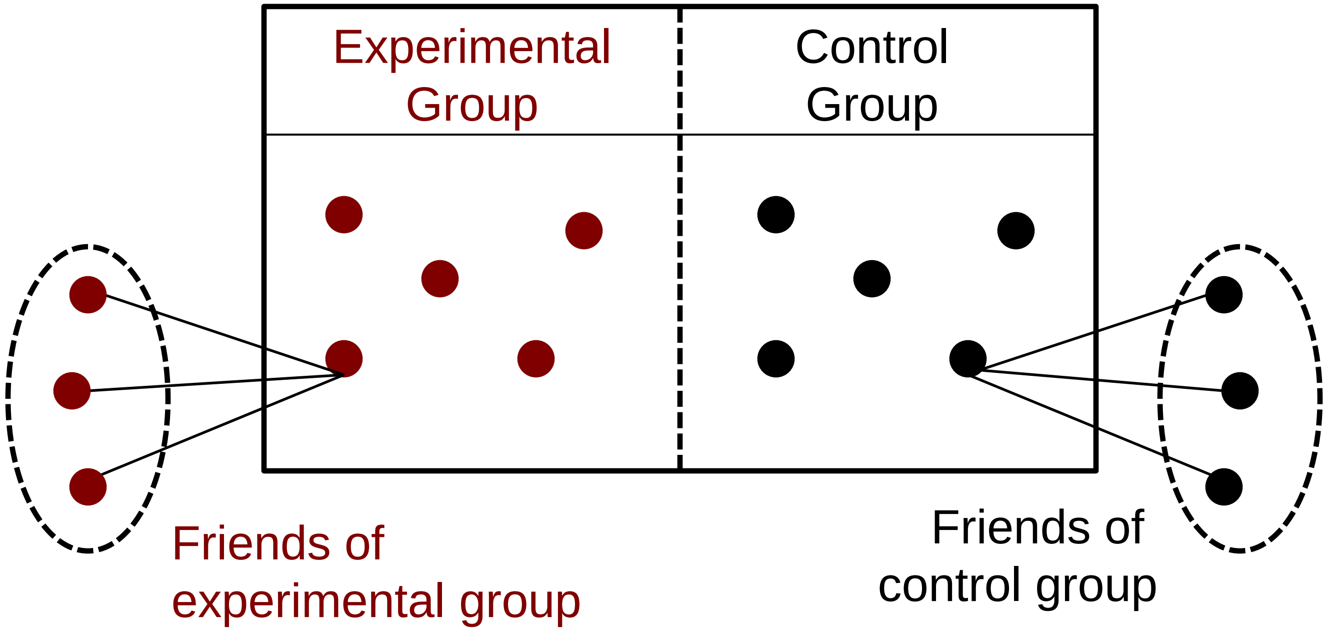
Anatomy of a Facebook Page

@sinanaral

Experimental Design: Treatment Randomization

@sinanaral

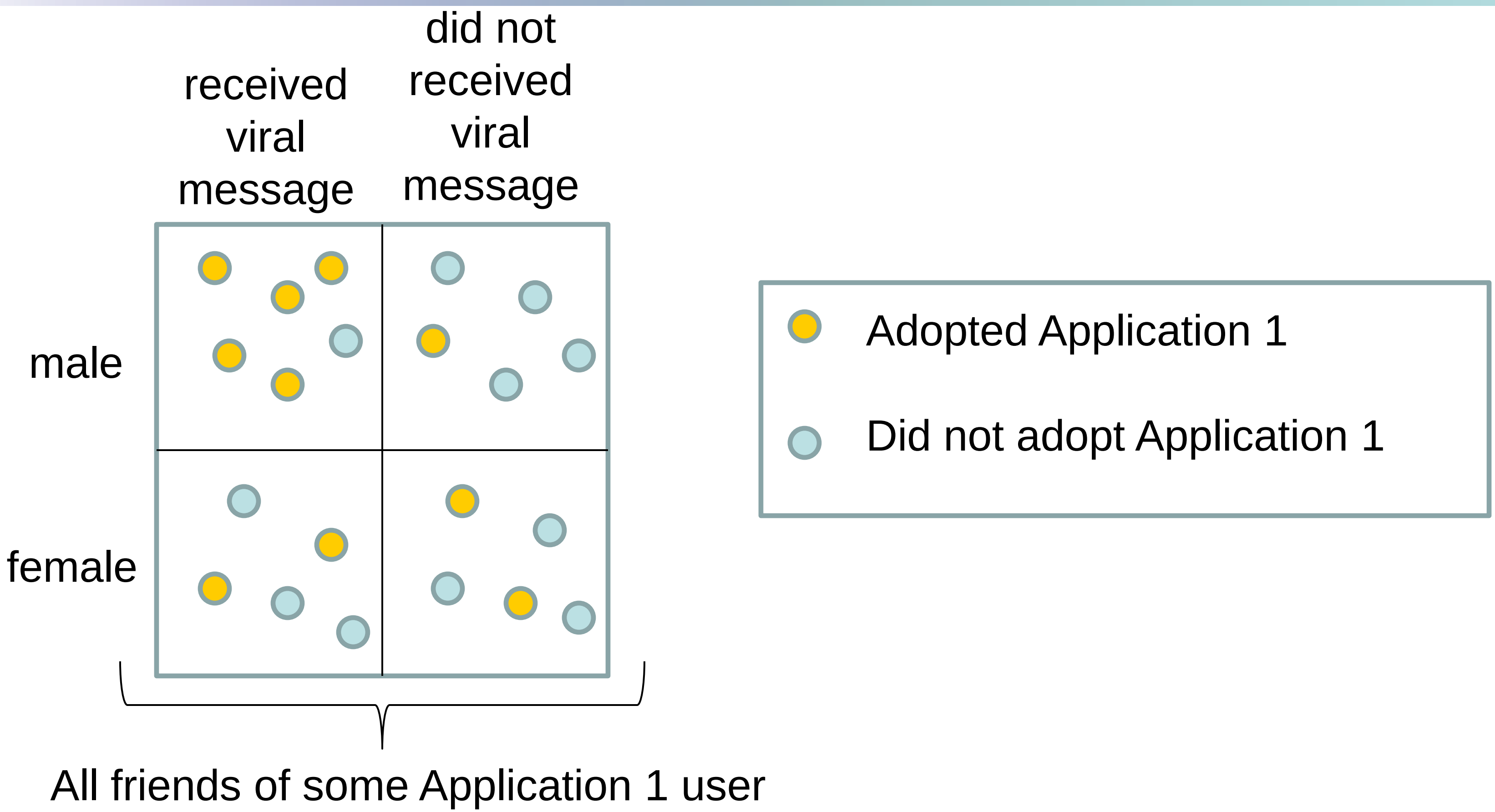
- To study the influence that a Facebook user exerts on his peers, we randomly split all users of Application 1 into two groups



- Users in the **Experimental Group** are subjected to the treatment while users in the **Control Group** are not.

- We then examine the behavior of the friends of users in both groups

Example: Effect of Gender on Susceptibility of Recipient



In this example it appears that females are relatively unaffected by viral messages, whereas males exhibit a greater likelihood to adopt when they receive a viral message

Randomized Trials of Peer Influence

@sinanaral

<u>Experiment 1</u>: Randomized Viral Messaging

Randomly Enabled Some (Experimental) Application Users to send Viral Messages to their Friends

<u>Experiment 2</u>: Randomized Traditional Advertizing

Randomly Showed Traditional Banner Ads to Some (Experimental) Application Users

Compared Adoption and Usage Behavior of the <u>Neighbors</u> of Experimental Groups 1 and 2 to the <u>Neighbors</u> of:

<u>Control Groups</u>: No Messaging; No Ads

Two Experiments

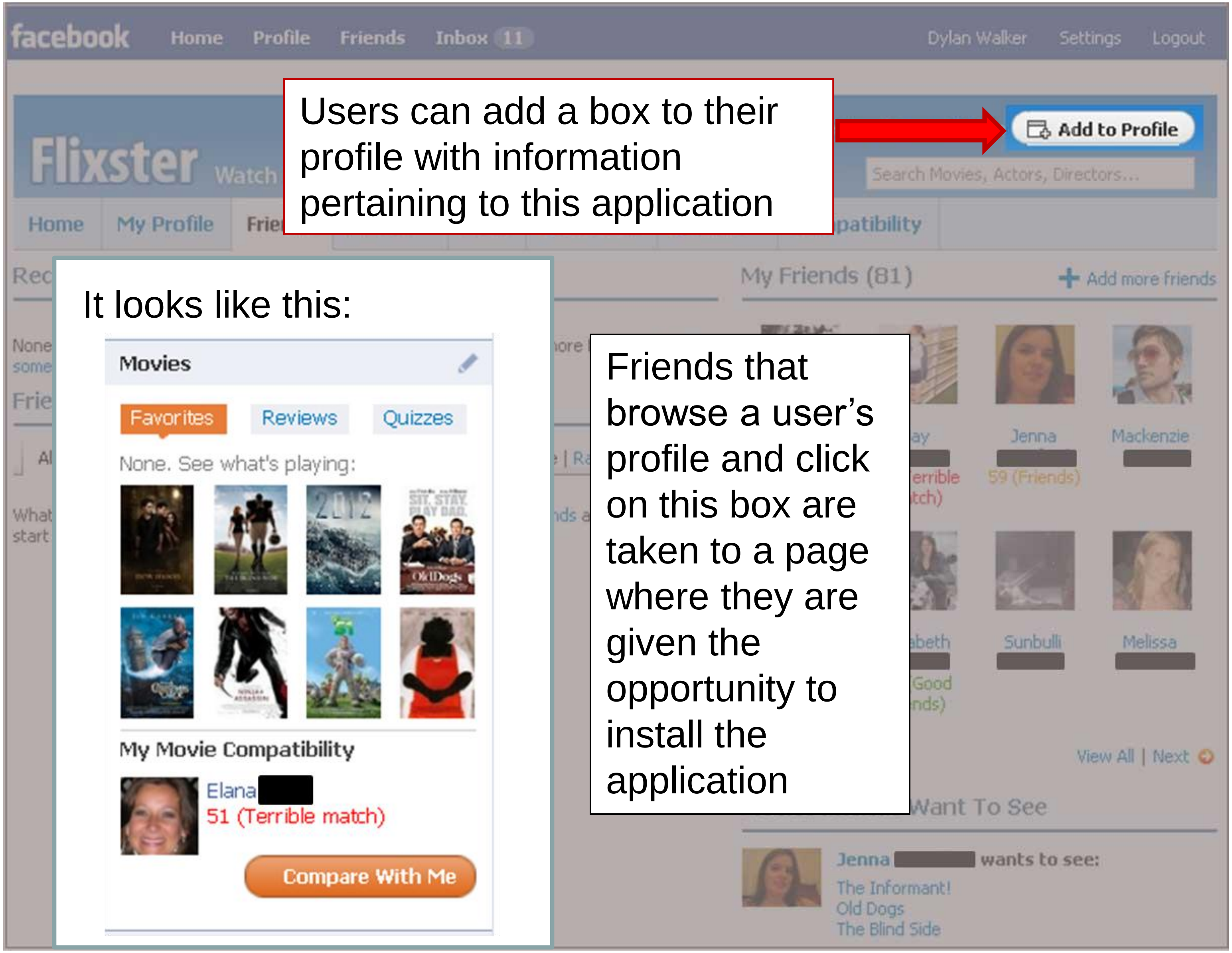
@sinanaral

I. Influence arising from exchange of viral messages
   - Treated users are given the ability to send viral messages within Application 1
   - These viral messages encourage the recipient to install Application 1
   - Observe peer adoption of Application 1

II. Influence arising from exposure to traditional advertisement
   - Treated users receive an advertisement for Application 2 that is displayed within Application 1
   - Observe peer adoption of Application 2

Viral Messaging Channels

@sinanaral



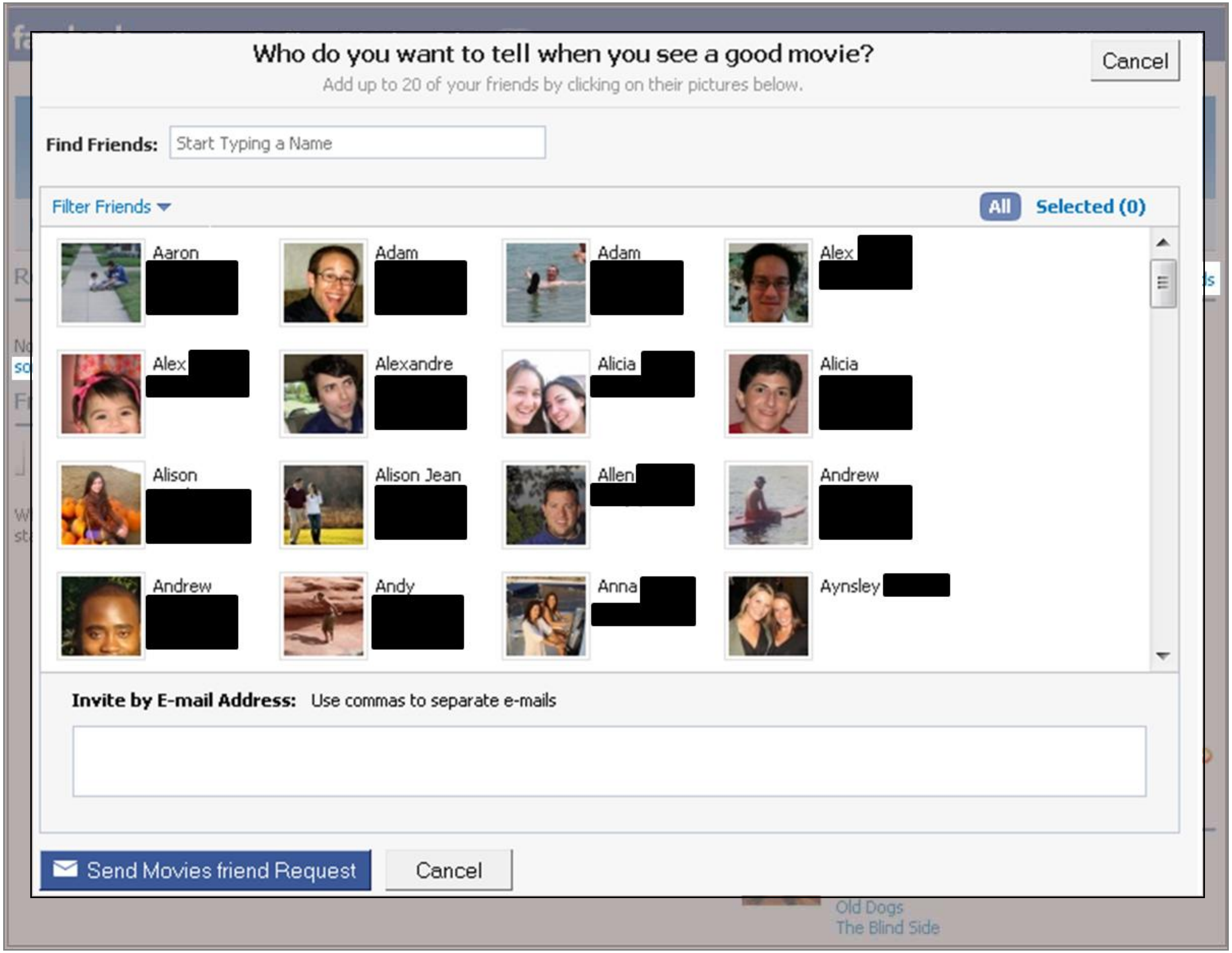
Users can add a box to their profile with information pertaining to this application

It looks like this:

Friends that browse a user's profile and click on this box are taken to a page where they are given the opportunity to install the application

Viral Messaging Channels

@sinanaral



Viral Communications

- We classify viral communications into two categories

Passive viral messages

Impersonal messages that are generated automatically by an application as a consequence of an application user's activity

e.g., profile box, notifications

Active viral messages:

Direct solicitations by a user to his friends to adopt an application, join a group or engage in an application-specific behavior

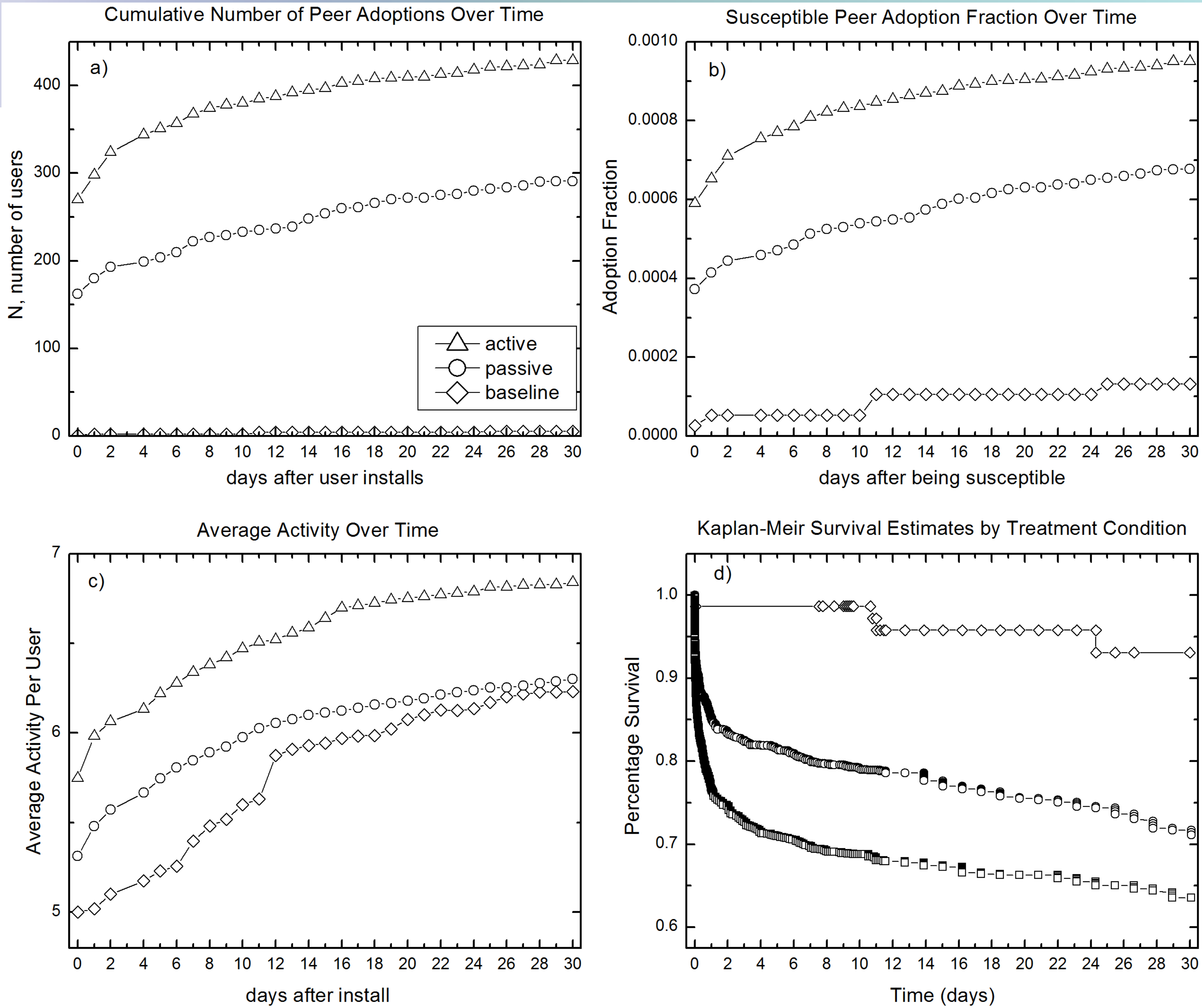
e.g., invites

“Causal Knowledge Requires Controlled Variation.”

– Armin Falk and James Heckman, Science (2009)

Comparing treated to untreated without random assignment creates bias in the potential untreated outcomes of treatment and comparison groups.

Randomization: treatment and control groups differ in expectation only through their exposure to the treatment.



Cumulative Number of Peer Adoptions Over Time
a)
N, number of users
days after user installs
active
passive
baseline
Susceptible Peer Adoption Fraction Over Time
b)
Adoption Fraction
days after being susceptible
Average Activity Over Time
c)
Average Activity Per User
days after install
Kaplan-Meir Survival Estimates by Treatment Condition
d)
Percentage Survival
Time (days)

	7	8	9	10
	Application Activity	Application Activity	Application Activity	Application Activity
	Beta *(SE)*	*Beta* *(SE)*	*Beta* *(SE)*	*Beta* *(SE)*
Viral State = Passive	.129* (.074)	.112 (.079)	.062 (.076)	-.037 (.074)
Viral State = Active	.190*** (.074)	.171** (.079)	.091 (.076)	-.006 (.074)
Degree	-.0001 (.0001)	-.0001 (.00	-.0002**	-.0002** (.0001)
Facebook		.054 (.0		.026* (.014)
				—
Notifications				.022*** (.001)
Invites				.055** (.024)
Number of Adopters			.607*** (.030)	.360*** (.031)
F Value (d.f.)	3.51*** (3)	4.87*** (4)		
R^2	.002	.003		
Observations	6310	5766		

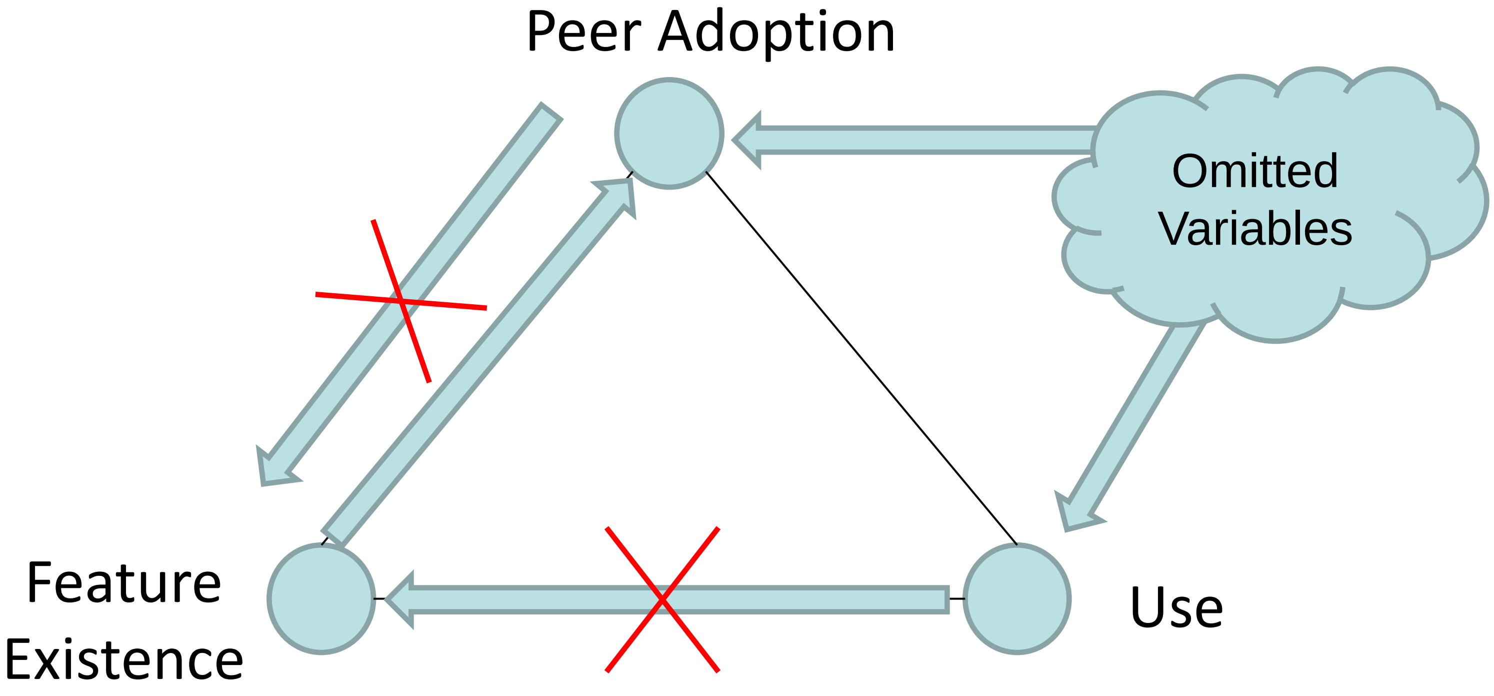
Notes: ***p<.(

Viral state is correlated with application activity

However, this relationship disappears when you control for number of adopters in your local network

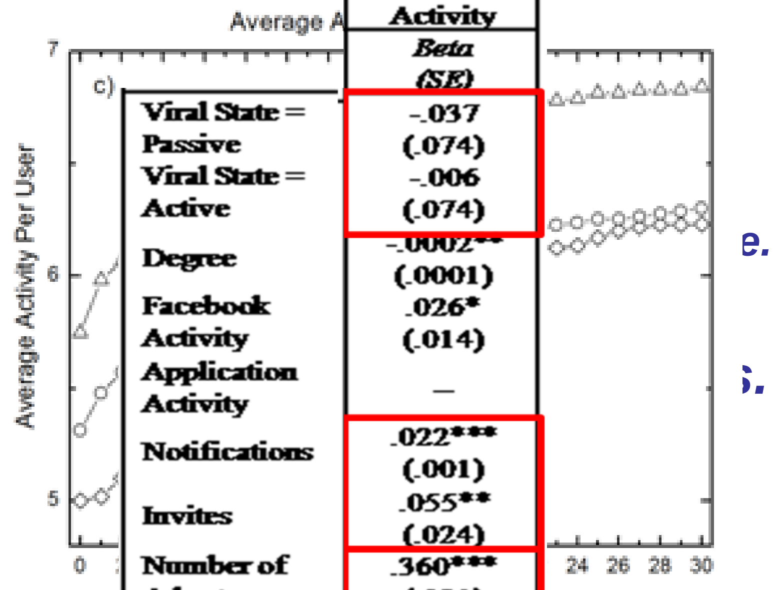
Number of adopters is still correlated with activity even when you control for use of viral features

Viral Features, Adoption and Use

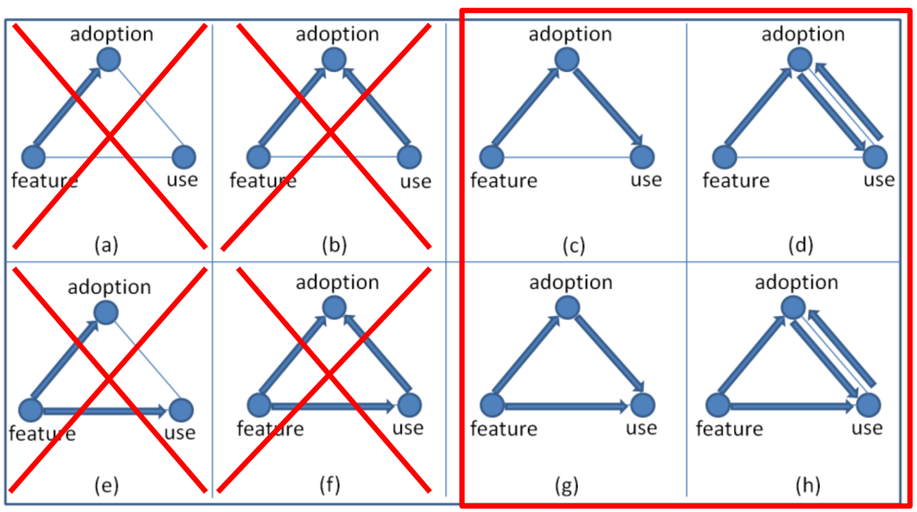


Adjudicating Alternate Explanations

- ***Correlation between Peer Adoption & Application U... ld be evidence of Network Effects.***
- ***It could also be explained by:***
  - ***Unobserved Heterogeneity / Omitted Var***
  - ***Demand Effects – Existence of Features ... e.***
- ***But, treatment is randomized so omi... the variation we see in adoption anc... s.***
- ***There could be an interaction effect ... and a feature itself, but we observe a... adoption and application use contro...***



	Application Activity
	Beta *(SE)*
Viral State = Passive	-.037 (.074)
Viral State = Active	-.006 (.074)
Degree	-.0002** (.0001)
Facebook Activity	.026* (.014)
Application Activity	—
Notifications	.022*** (.001)
Invites	.055** (.024)
Number of Adopters	.360*** (.031)



1. (a) and (b) inconsistent with the discrepancy in app use between treatment groups.
2. If Demand Effects: Controlling for them should remove any spurious correlation between peer adoption and use.
3. Controlling for feature use, peer adoption still highly correlated with app use.

Baseline Hazards Increasing in *k* adoption events