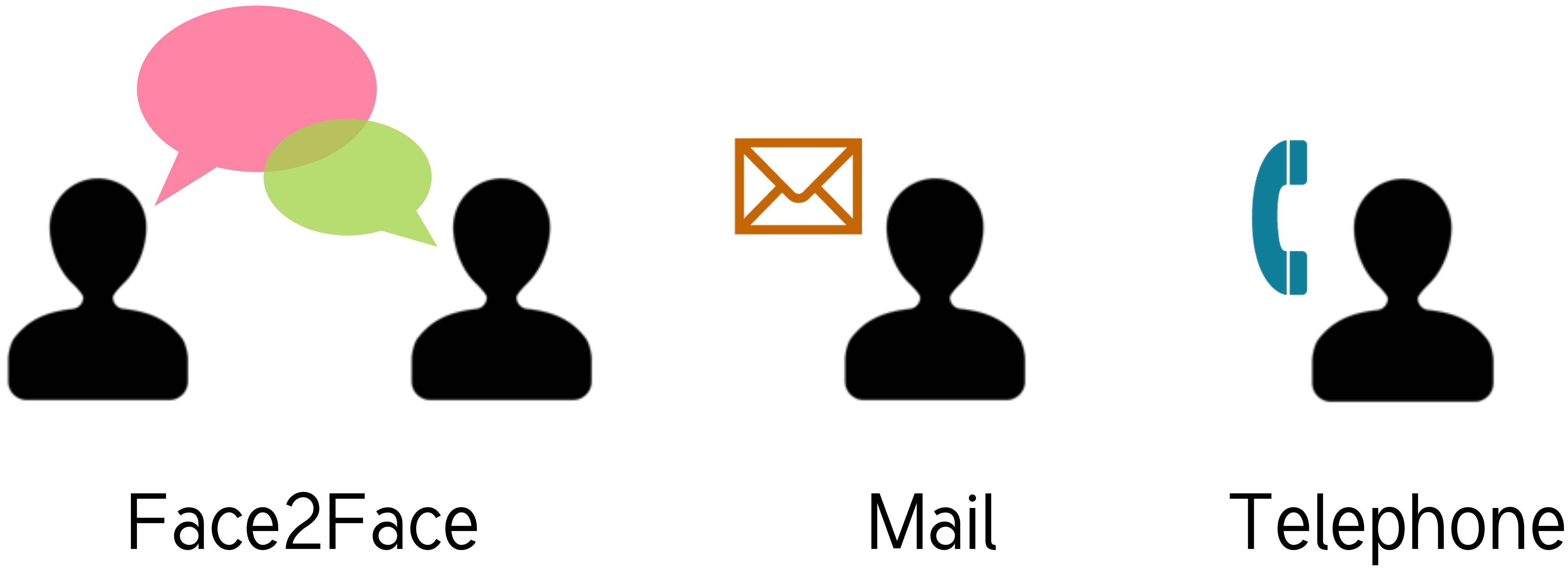
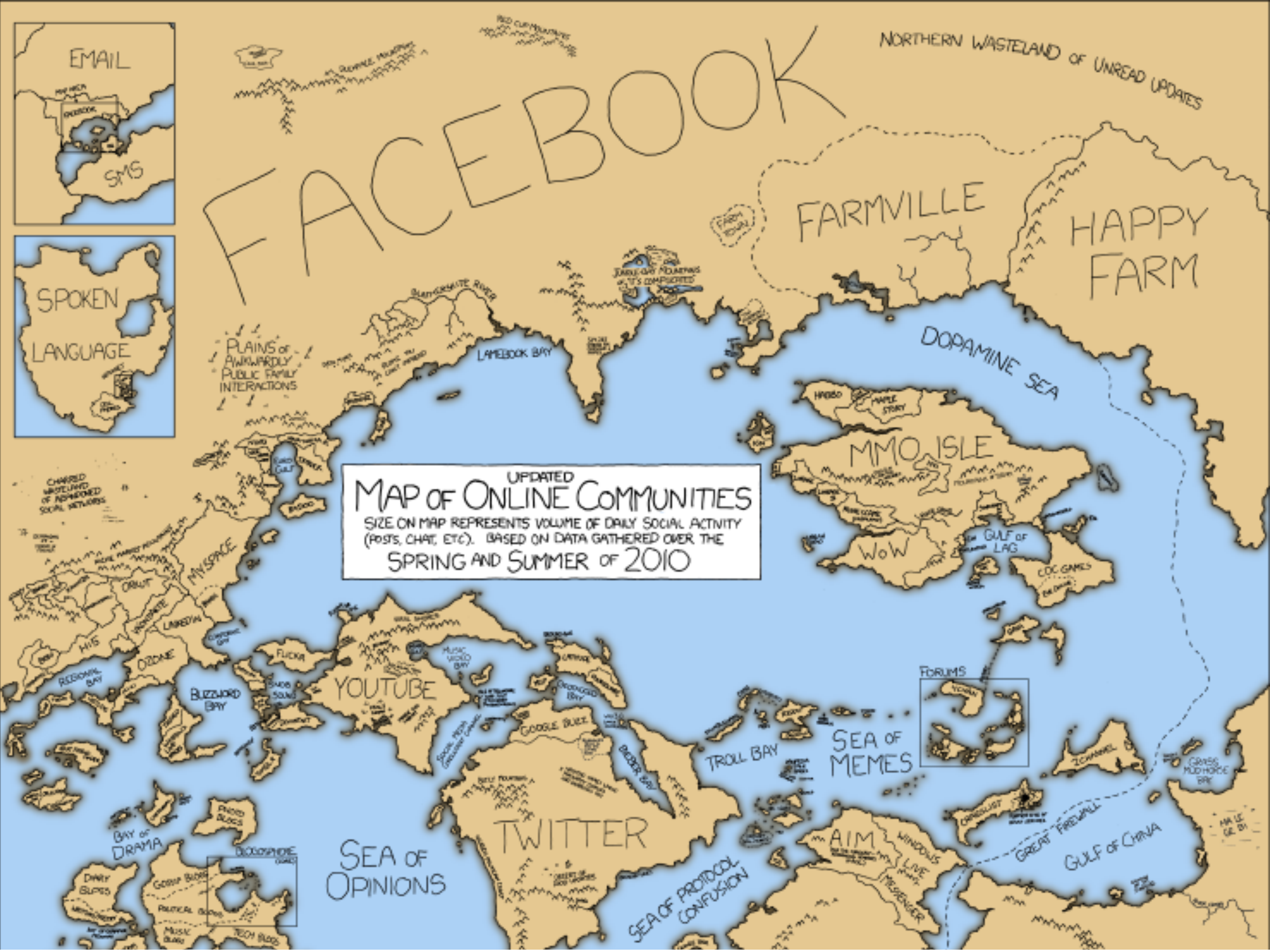
Information Diffusion on Online Social Networks

Lilian Weng
Center for Complex Networks and Systems Research
School of Informatics and Computing
Indiana University Bloomington


Face2Face
Mail
Telephone



UPDATED MAP OF ONLINE COMMUNITIES
SIZE ON MAP REPRESENTS VOLUME OF DAILY SOCIAL ACTIVITY (POSTS, CHAT, ETC). BASED ON DATA GATHERED OVER THE
SPRING AND SUMMER OF 2010
EMAIL
SMS
SPOKEN LANGUAGE
FACEBOOK
NORTHERN WASTELAND OF UNREAD UPDATES
FARM TOWN
FARMVILLE
HAPPY FARM
DOPAMINE SEA
LAMEBOOK BAY
PLAINS OF AWKWARDLY PUBLIC FAMILY INTERACTIONS
CHARRED WASTELAND OF ABANDONED SOCIAL NETWORKS
MYSPACE
ORKUT
LINKEDIN
HI5
REGIONAL BAY
BUZZWORD BAY
FLICKR
YOUTUBE
MUSIC VIDEO BAY
GOOGLE BUZZ
TWITTER
SEA OF OPINIONS
BAY OF DRAMA
PHOTO BLOGS
BLOGOSPHERE
DIARY BLOGS
GOSSIP BLOGS
POLITICAL BLOGS
MUSIC BLOGS
TECH BLOGS
SEA OF PROTOCOL CONFUSION
AIM
WINDOWS LIVE MESSENGER
TROLL BAY
SEA OF MEMES
FORUMS
2CHANNEL
CRAIGSLIST
GREAT FIREWALL
GULF OF CHINA
GRASS MUD HORSE BAY
HABBO
MAPLE STORY
MMO ISLE
WOW
GULF OF LAG
CDC GAMES

It becomes inexpensive and easy for people to produce, spread, and exchange information with each other.

24 PB data / Day

20 Hrs uploaded / Min

50 Mil tweets / Day

700 Bil min spent / Month

72.9 Items ordered / Sec

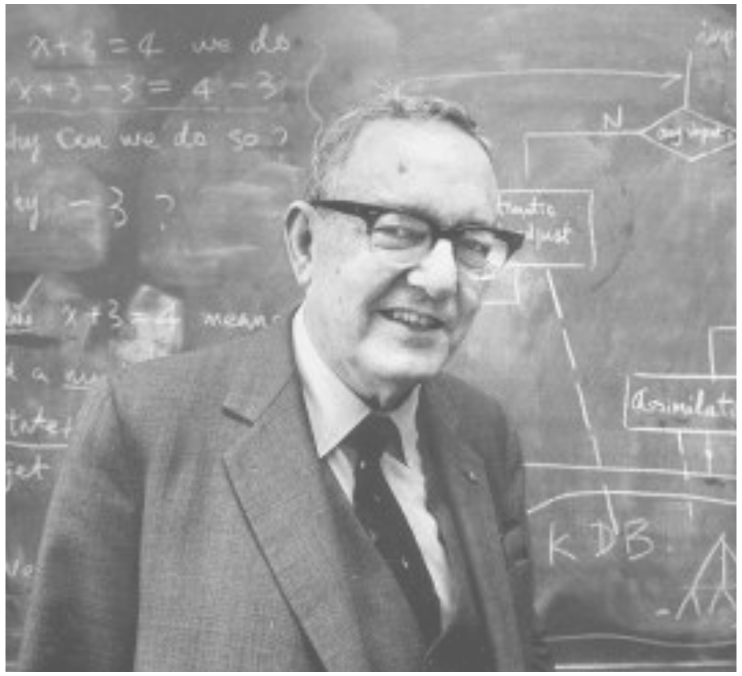
2.9 Mil emails / Sec

(IBM, 2012)

COMPUTATIONAL FRAMEWORK FOR BIG DATA


TRACK
OBSERVE
ANALYZE
MODEL
PREDICT

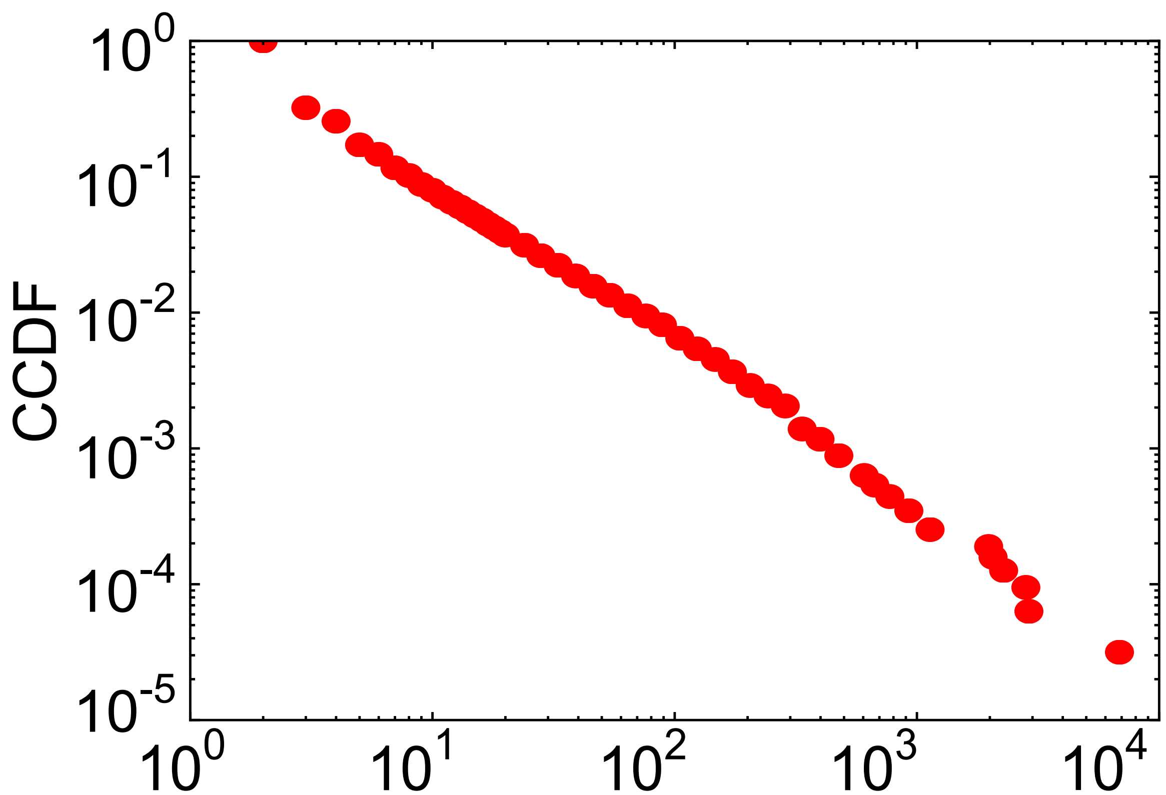
ATTENTION ECONOMY



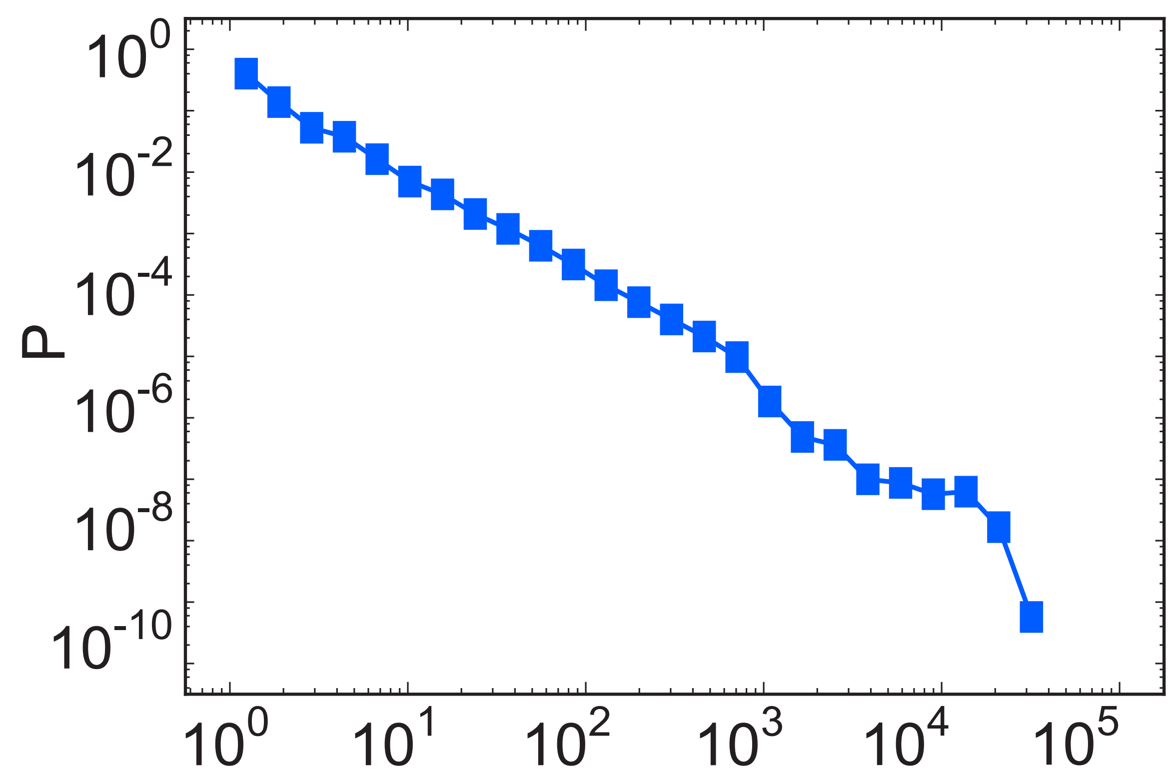
Herbert A. Simon, 1971

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

FIERCE COMPETITION



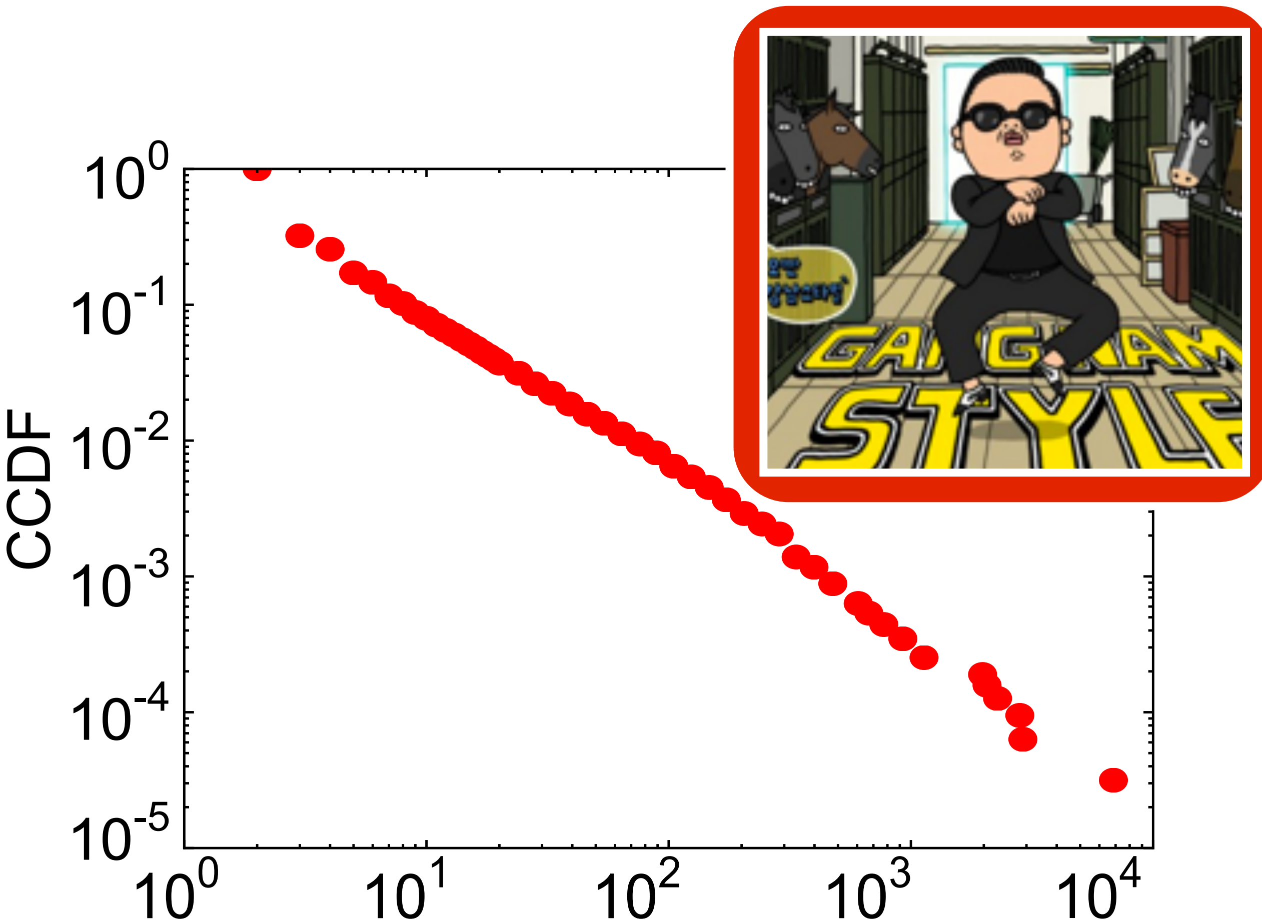
Hashtag Popularity
daily retweets
[Twitter]



User Popularity
followers
[Yahoo! Meme]

FIERCE COMPETITION

1.9 Bil Views

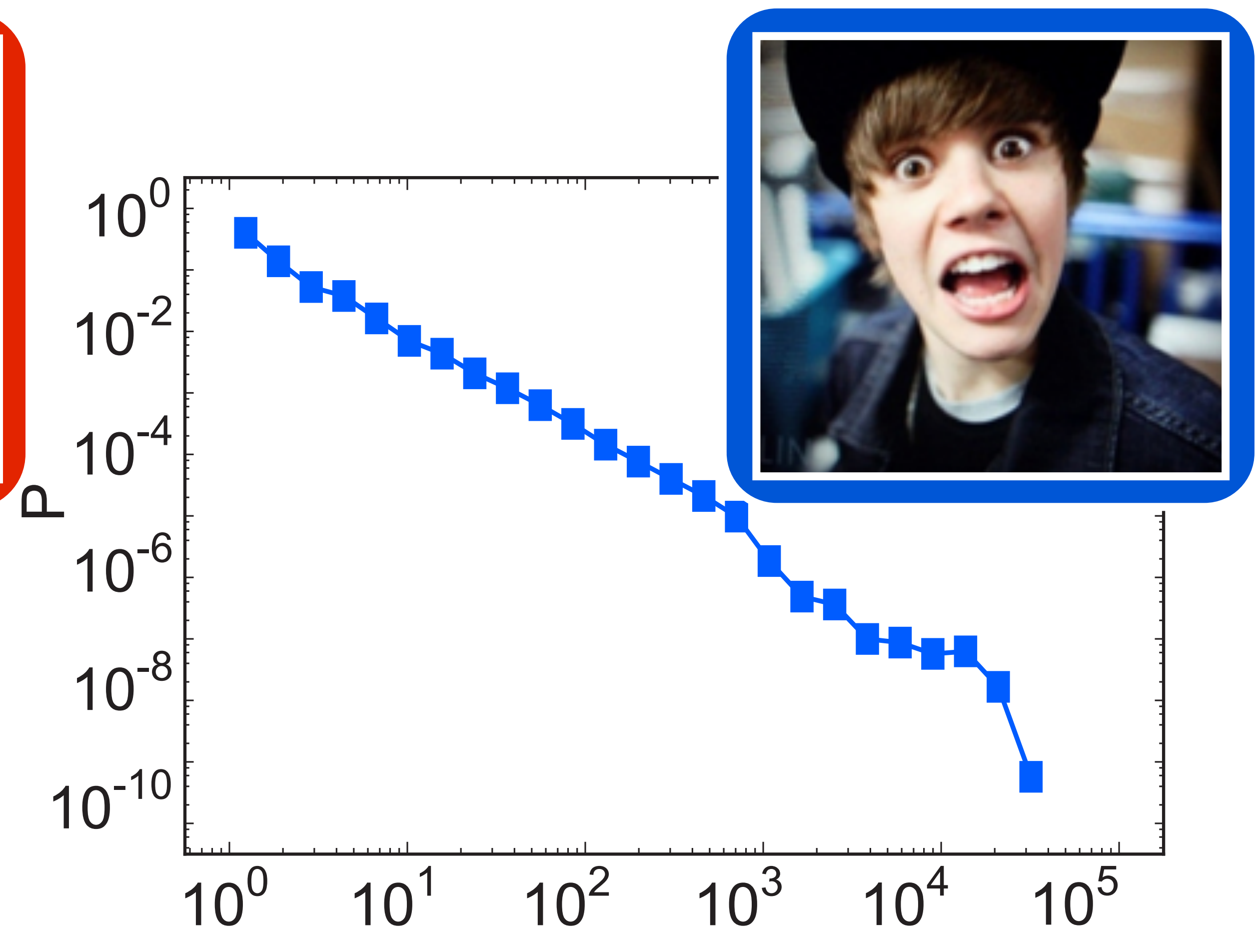


Hashtag Popularity

daily retweets

[Twitter]

50.7 Mil Followers

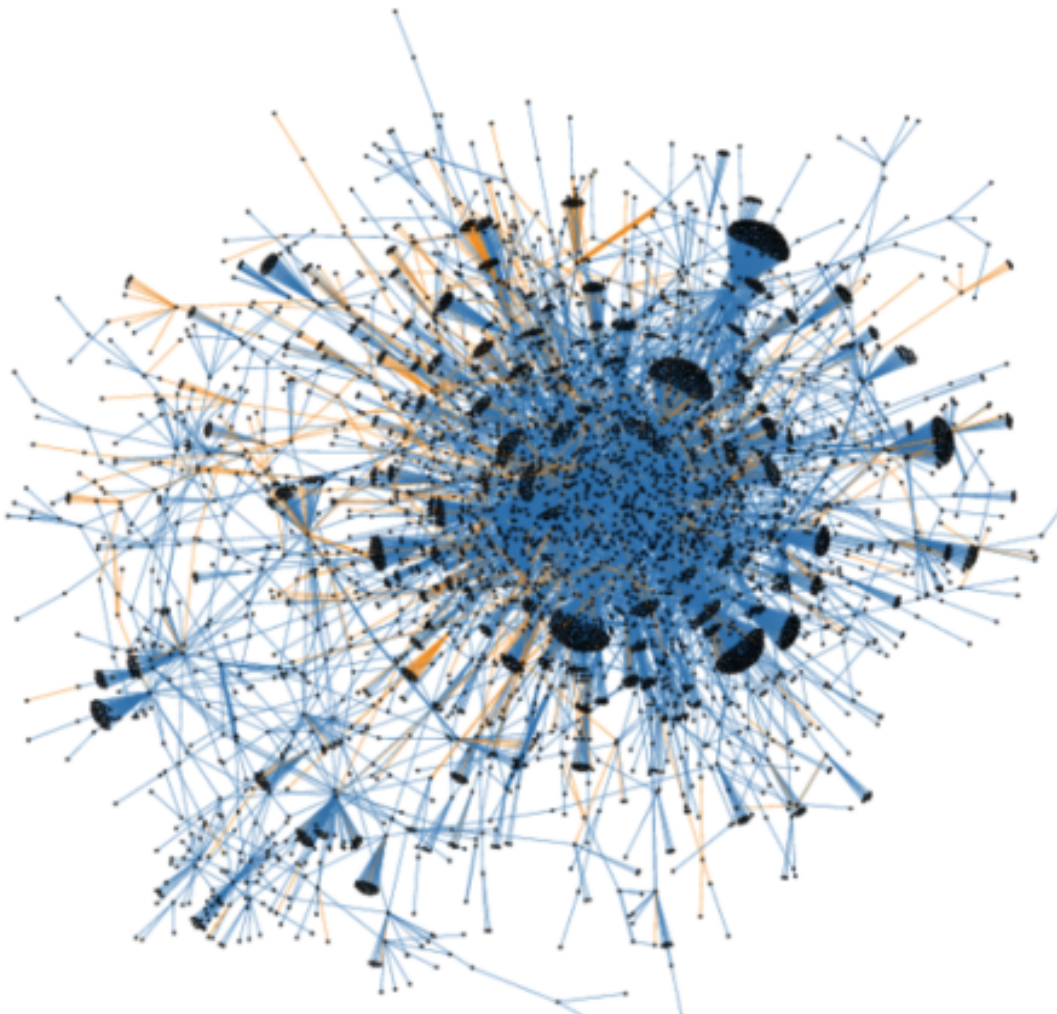


User Popularity

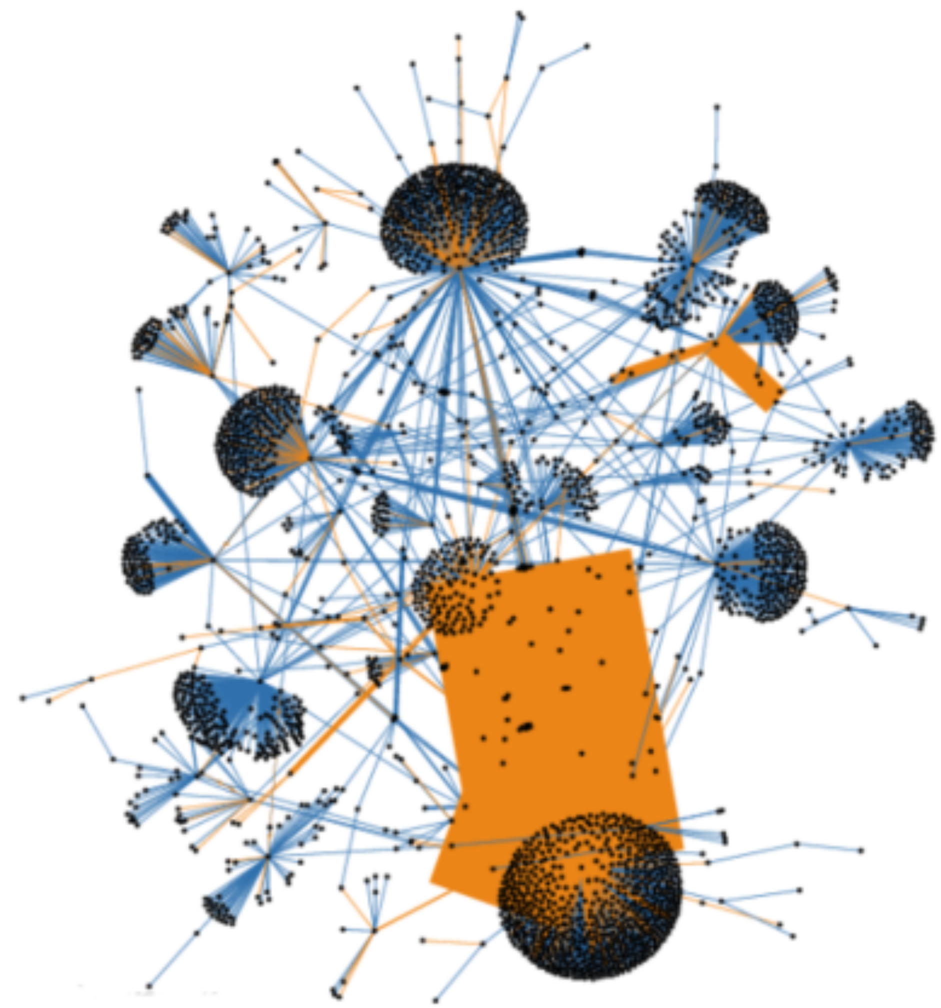
followers

[Yahoo! Meme]

INFORMATION DIFFUSION HAPPENS IN THE REAL WORLD.



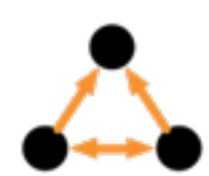
#tcot



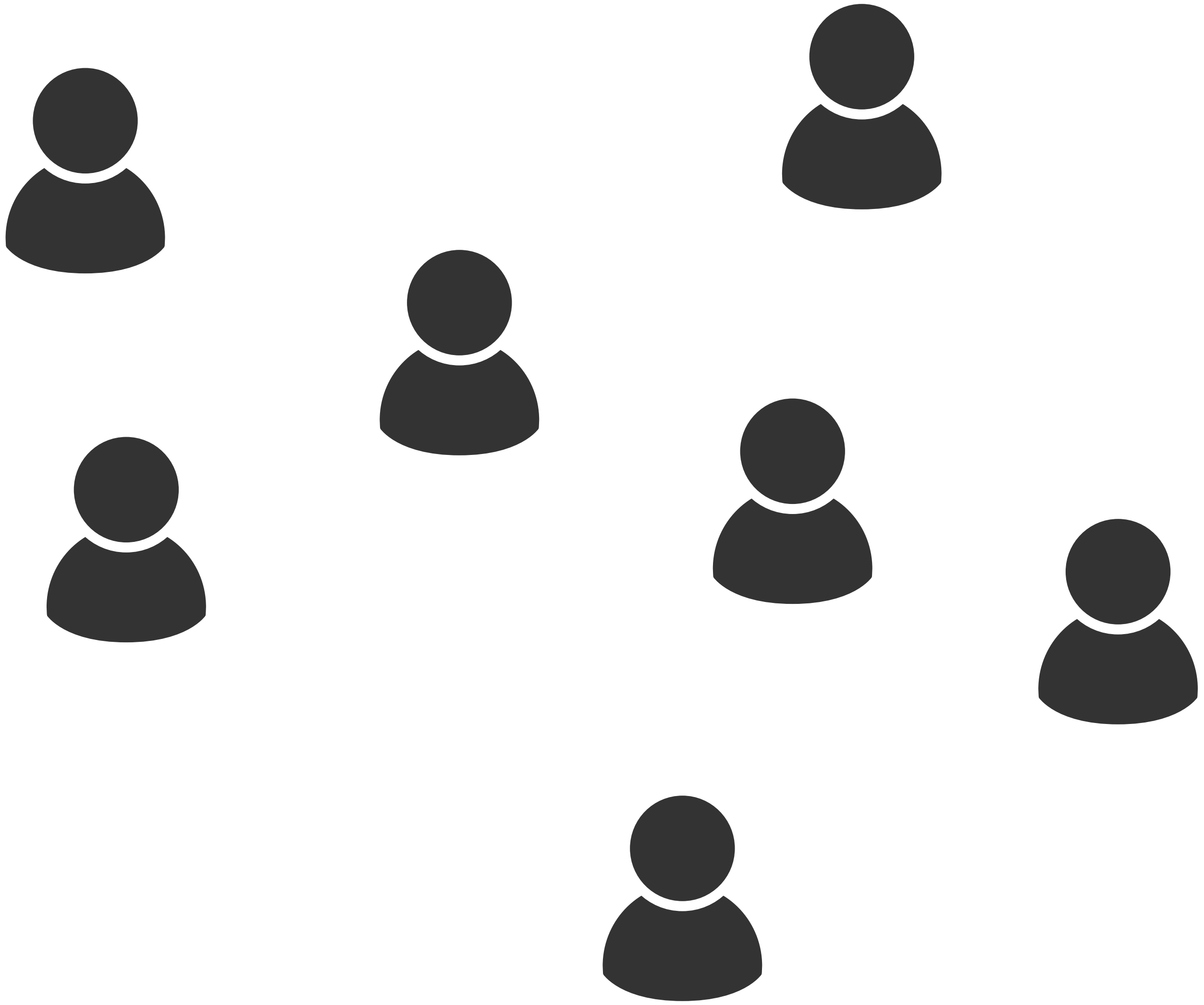
@ladygaga

Truthy

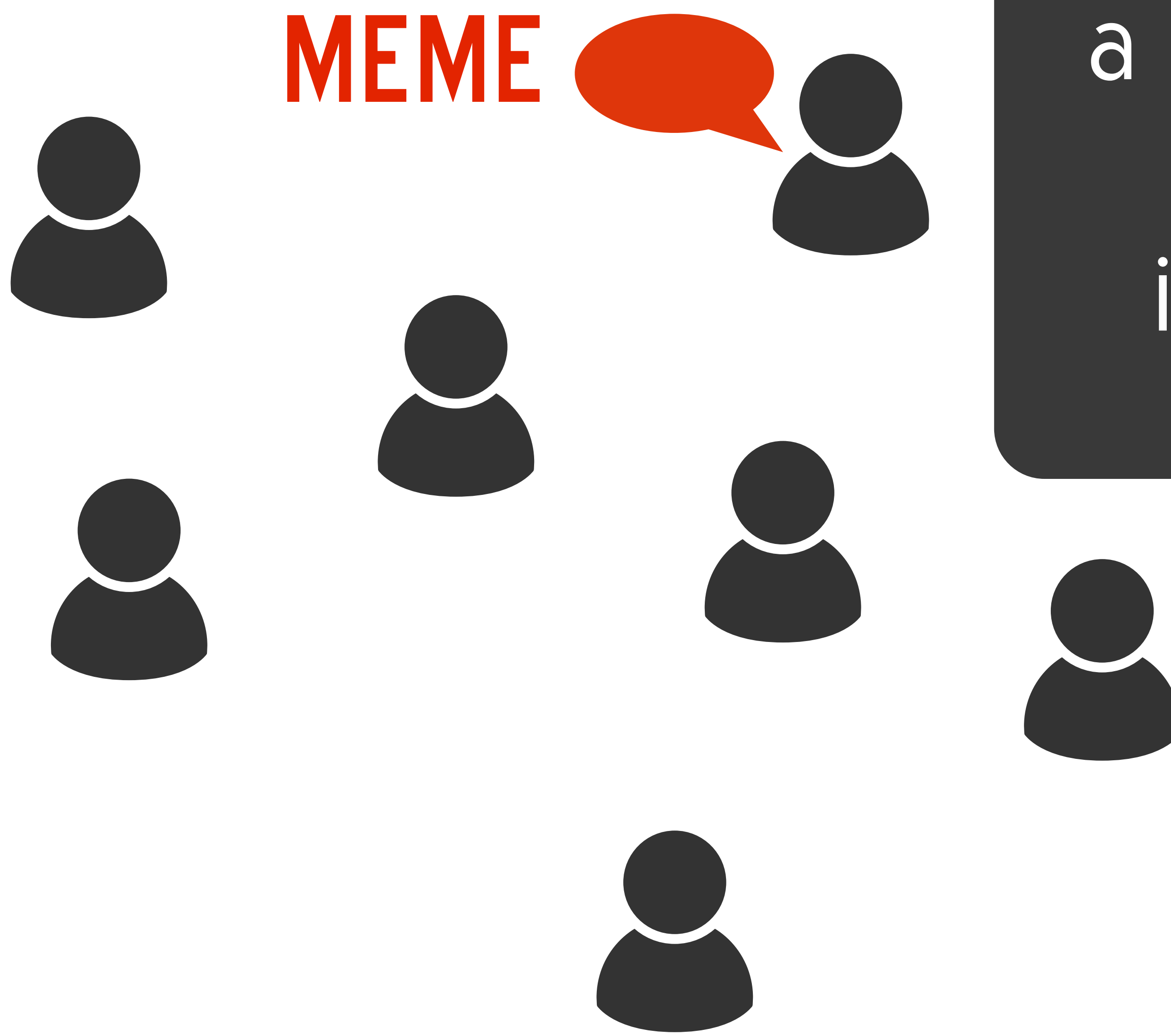
Retweet



Mention

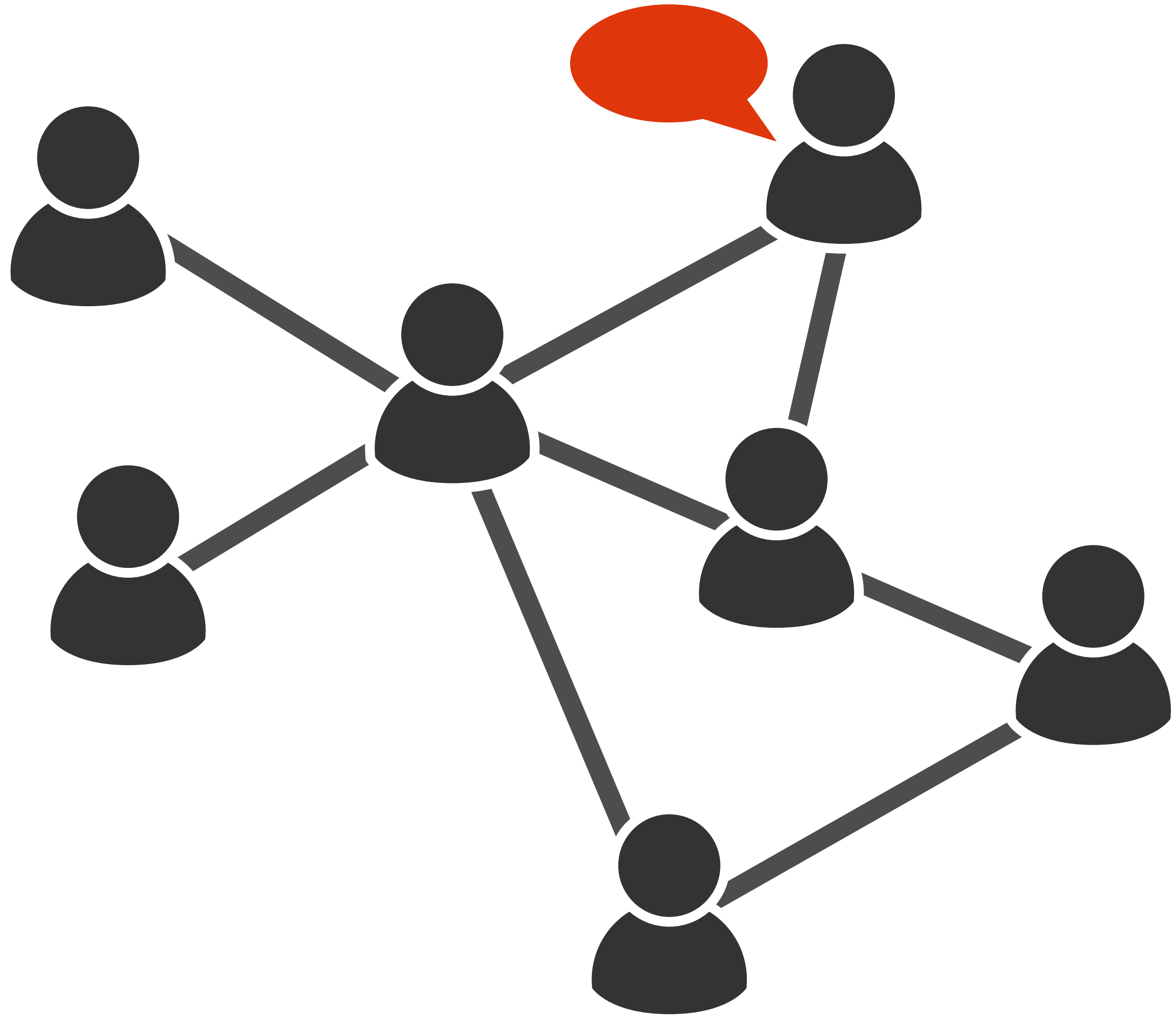


1. **People** who produce and share information

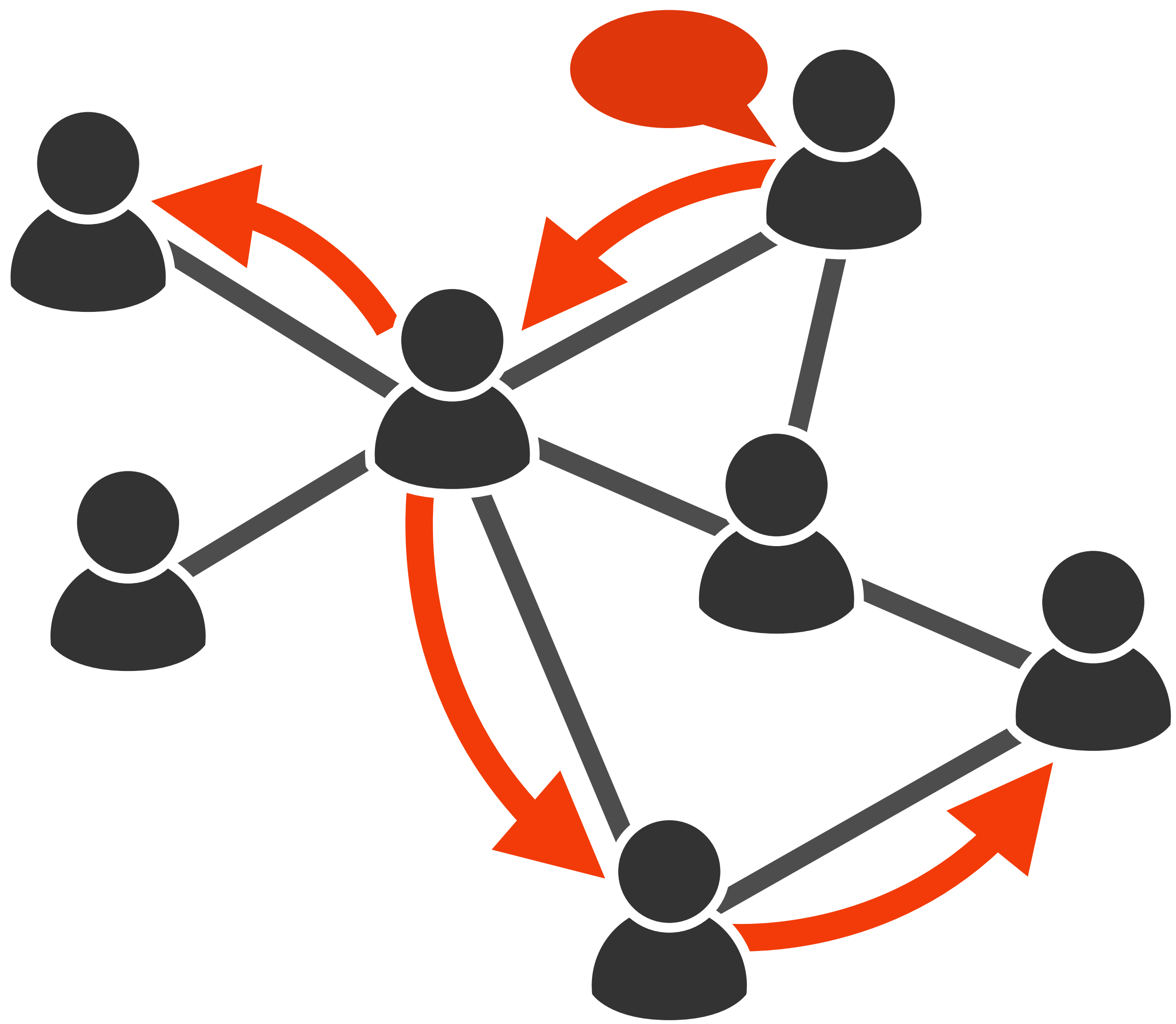


a transmissible unit of information.
(Dawkins, 1989)

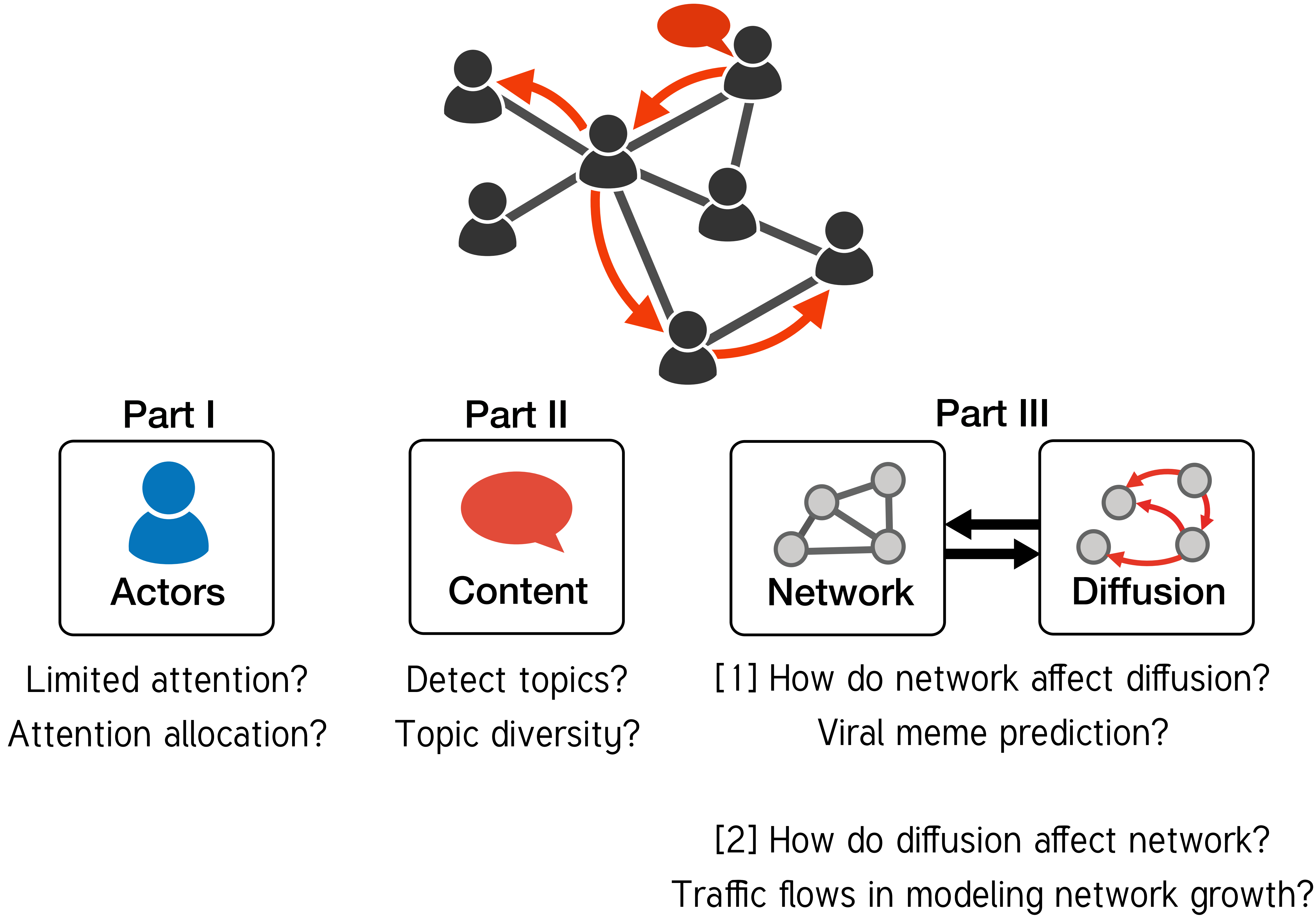
1. **People** who produce and share information
2. **Content** of transmissible messages



1. **People** who produce and share information
2. **Content** of transmissible messages
3. Social relationships forming the **network**

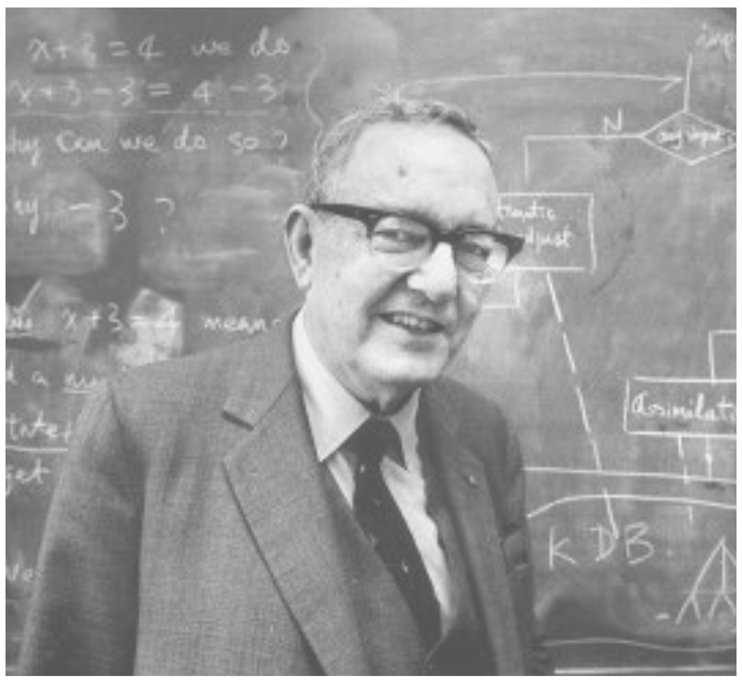


1. **People** who produce and share information
2. **Content** of transmissible messages
3. Social relationships forming the **network**
4. The mechanism of **diffusion** process


Part I
Actors
Limited attention?
Attention allocation?
Part II
Content
Detect topics?
Topic diversity?
Part III
Network
Diffusion
[1] How do network affect diffusion?
Viral meme prediction?
[2] How do diffusion affect network?
Traffic flows in modeling network growth?

ACTOR

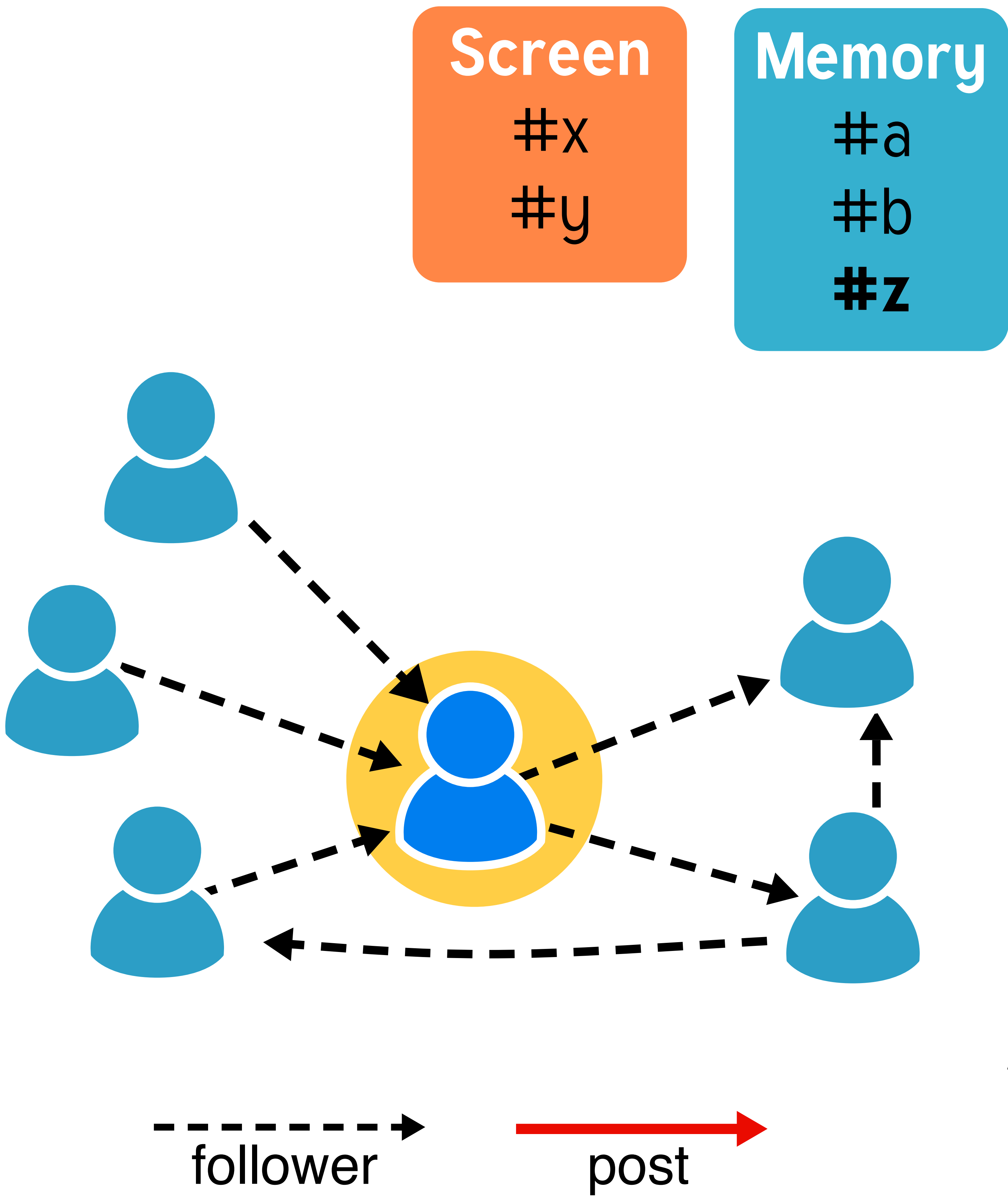
Part One



Herbert A. Simon, 1971

What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

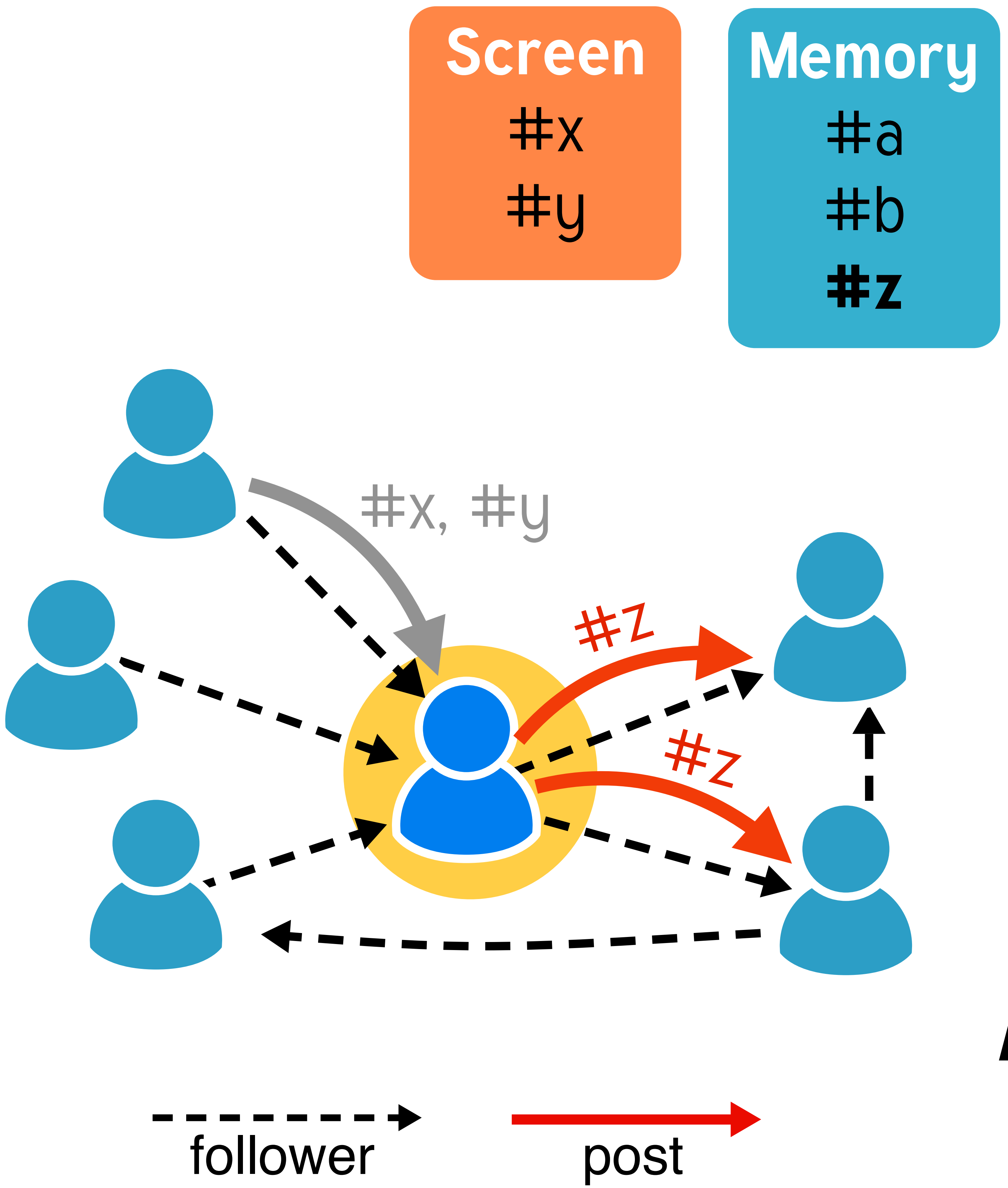
How does
limited individual attention
affect information diffusion?



Screen:
receiving posts
from neighbors

Memory:
storing sent posts

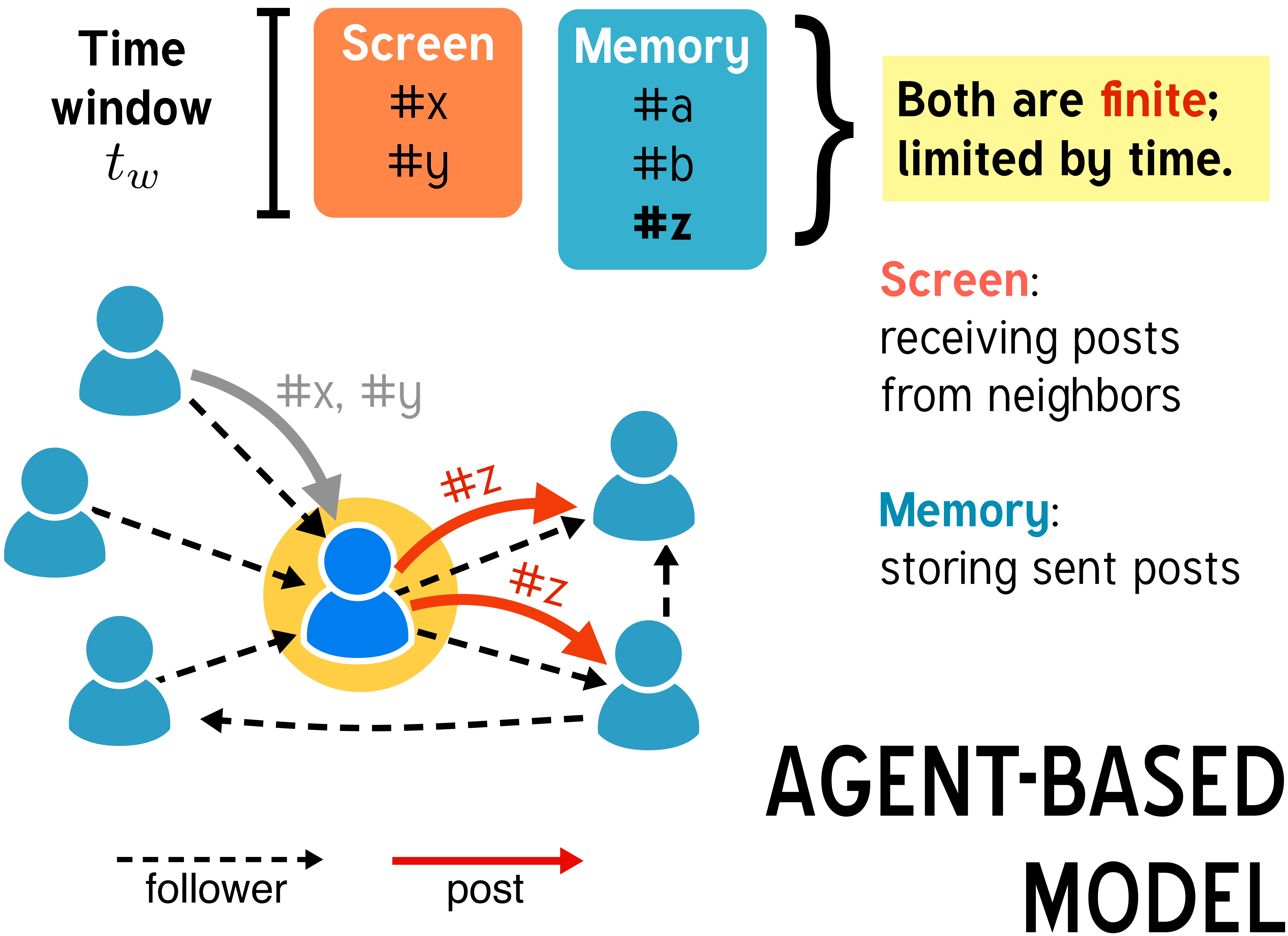
AGENT-BASED MODEL

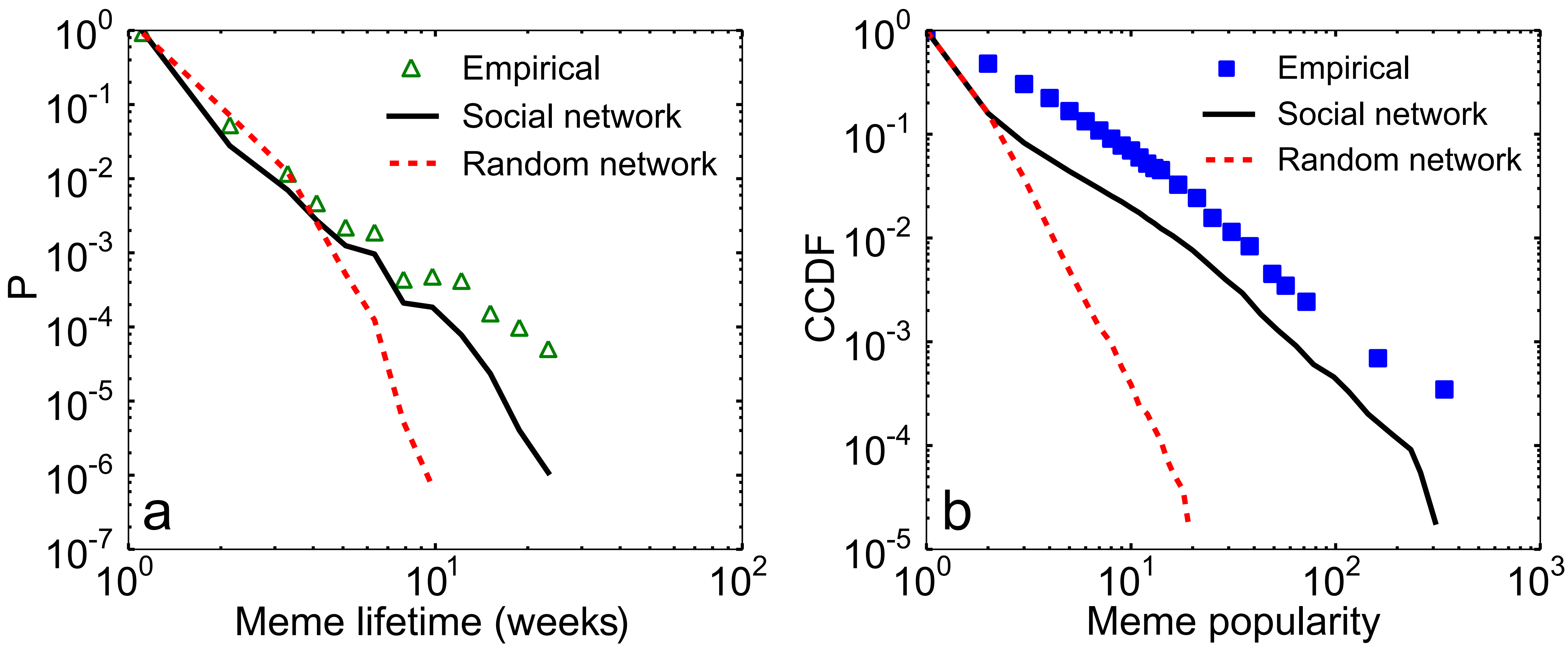


Screen:
receiving posts
from neighbors

Memory:
storing sent posts

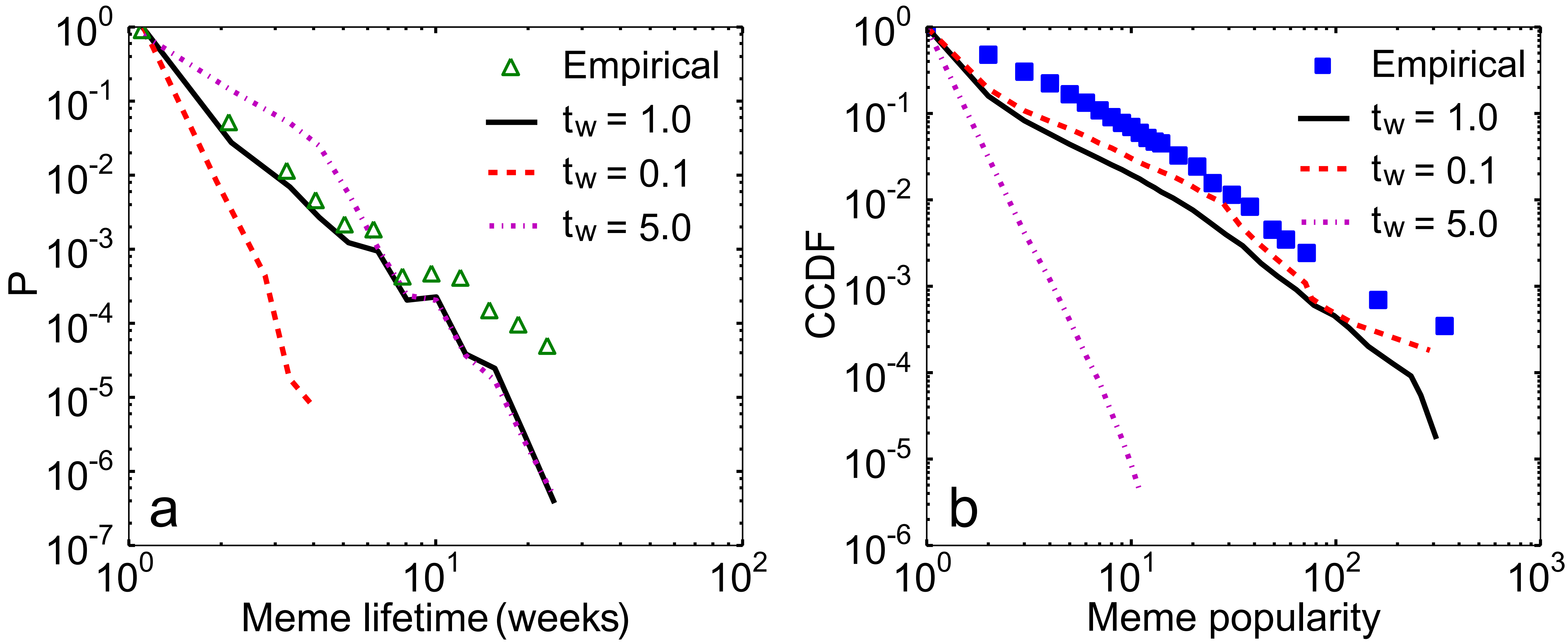
AGENT-BASED MODEL


Time window t_w
Screen
#x
#y
Memory
#a
#b
#z
Both are finite; limited by time.
Screen: receiving posts from neighbors
Memory: storing sent posts
#x, #y
#z
#z
follower
post
AGENT-BASED MODEL



Social network structure matters

(Weng et al. 2012)



Attention matters

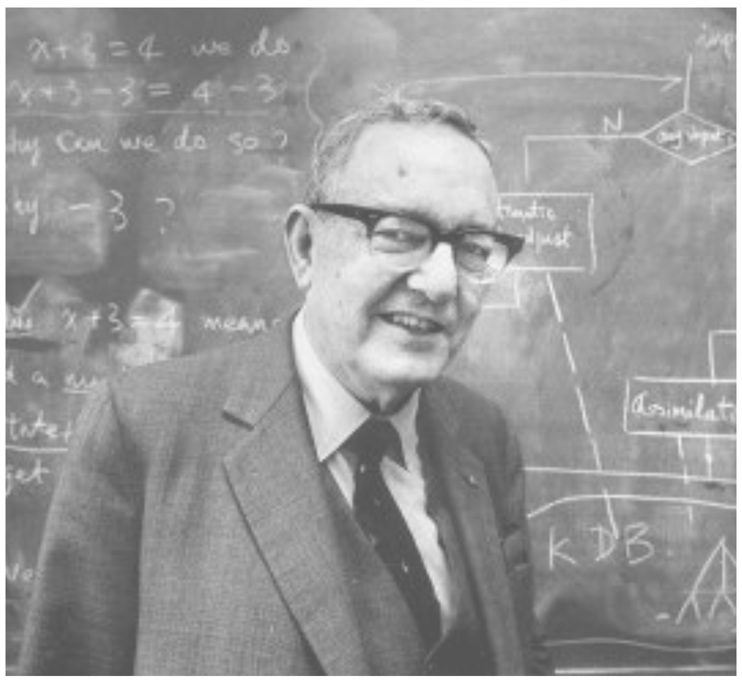
(Weng et al. 2012)

Social network structure

+

Competition for limited attention

→ **Heterogeneity** of meme dynamics

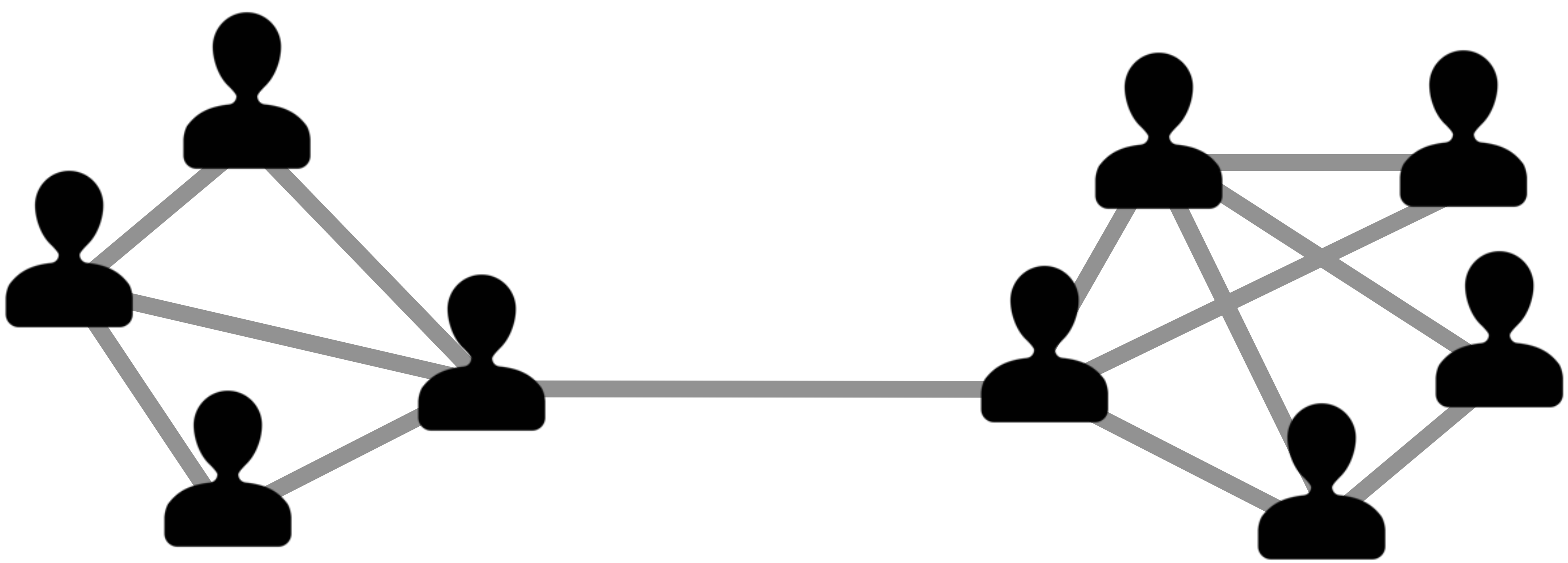


Herbert A. Simon, 1971

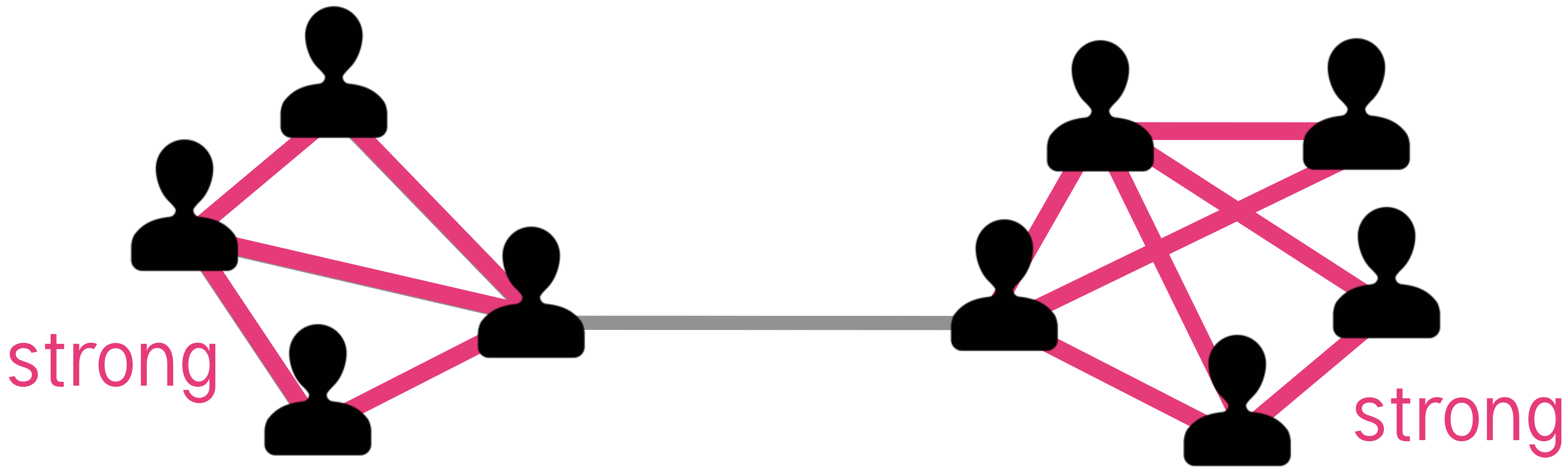
"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.

How does an individual allocate his/her limited attention?

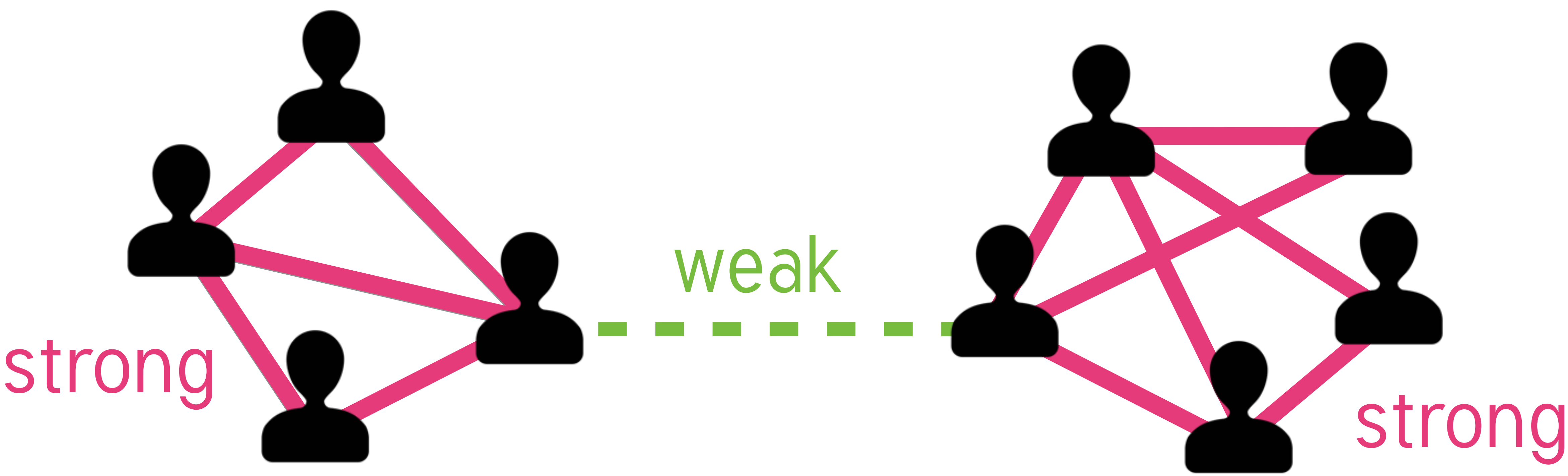
- Weak Tie Hypothesis (Granovetter,1973; Onnela et al.,2007)
- Limited Attention (Simon,1971; Gonçalves et al., 2011)



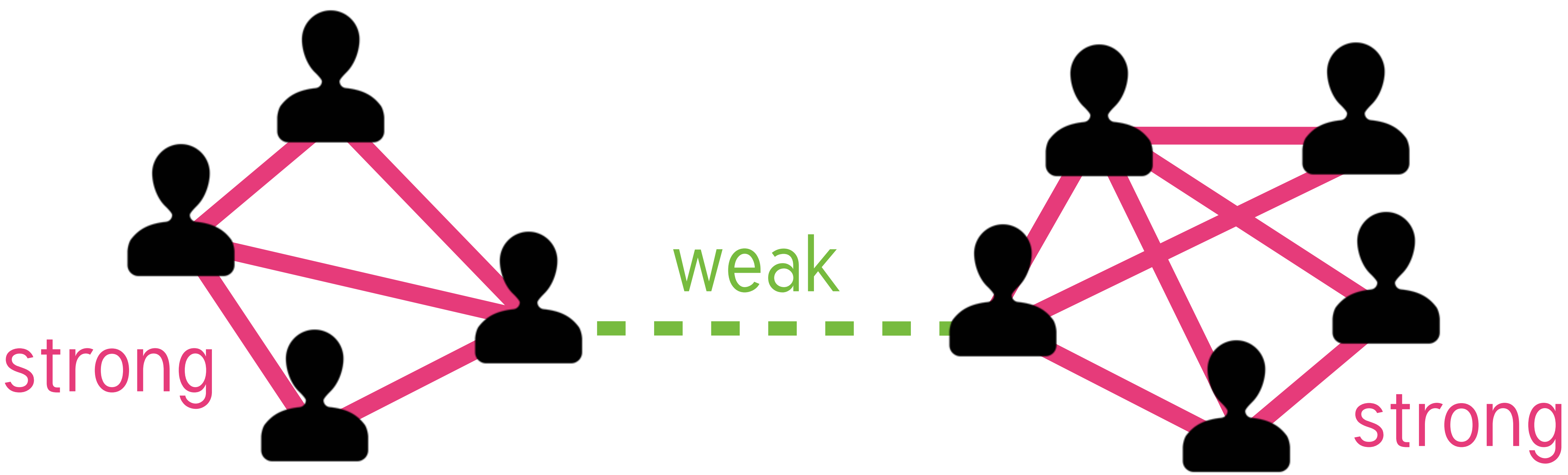
- Weak Tie Hypothesis (Granovetter, 1973; Onnela et al., 2007)
- Limited Attention (Simon, 1971; Gonçalves et al., 2011)



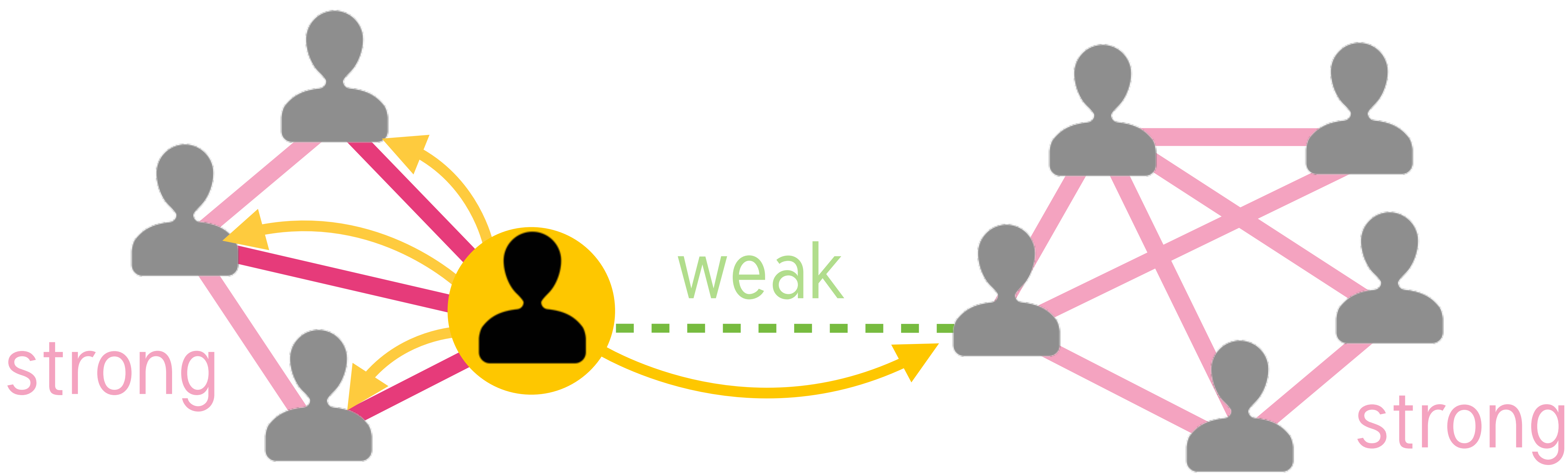
- Weak Tie Hypothesis (Granovetter, 1973; Onnela et al., 2007)
- Limited Attention (Simon, 1971; Gonçalves et al., 2011)



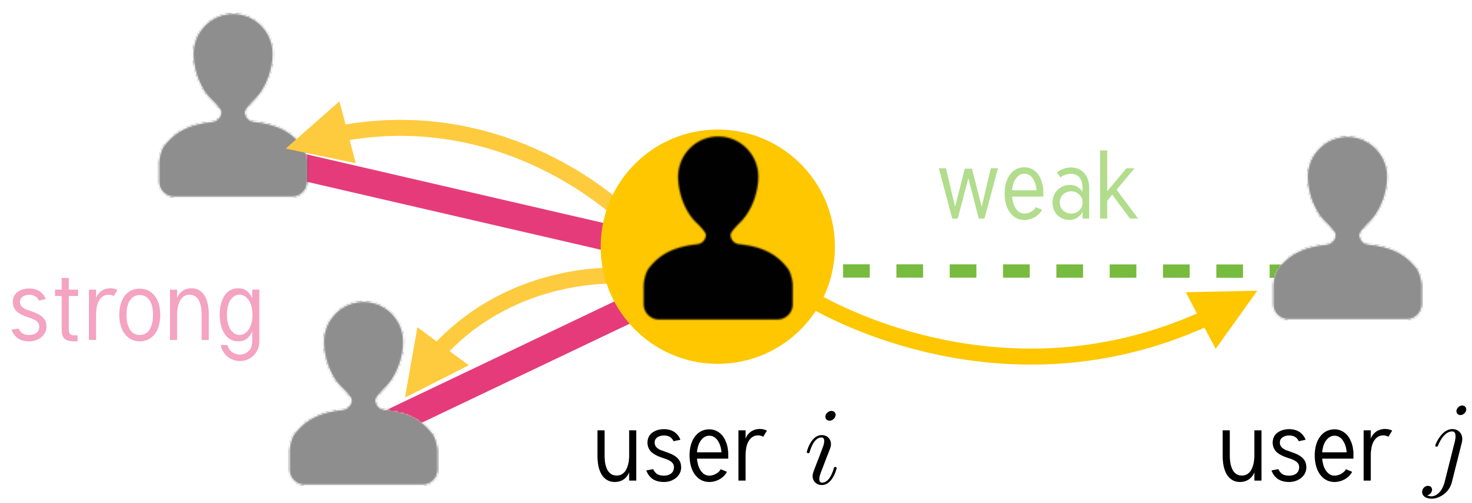
- Weak Tie Hypothesis (Granovetter, 1973; Onnela et al., 2007)
- Limited Attention (Simon, 1971; Gonçalves et al., 2011)



How does an individual allocate attention among strong and weak ties?

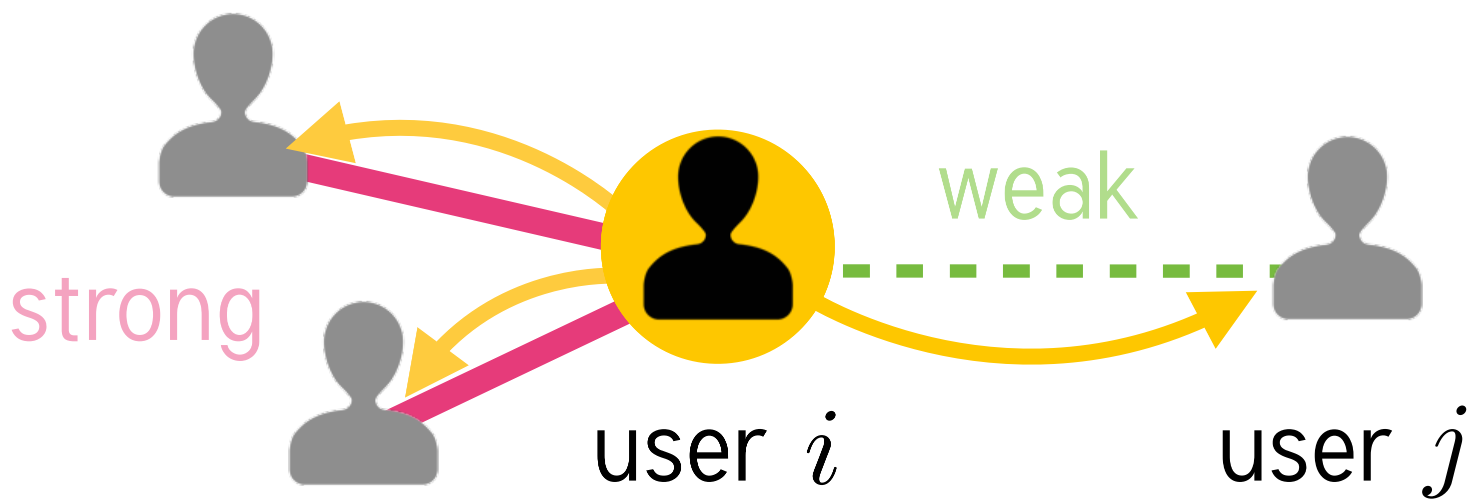


MEASURE OF ATTENTION



$$a(i \to j) = \boxed{\text{Total amount of attention of user } i} \times \boxed{\text{Frac. of activity of } i \text{ directing to } j}$$

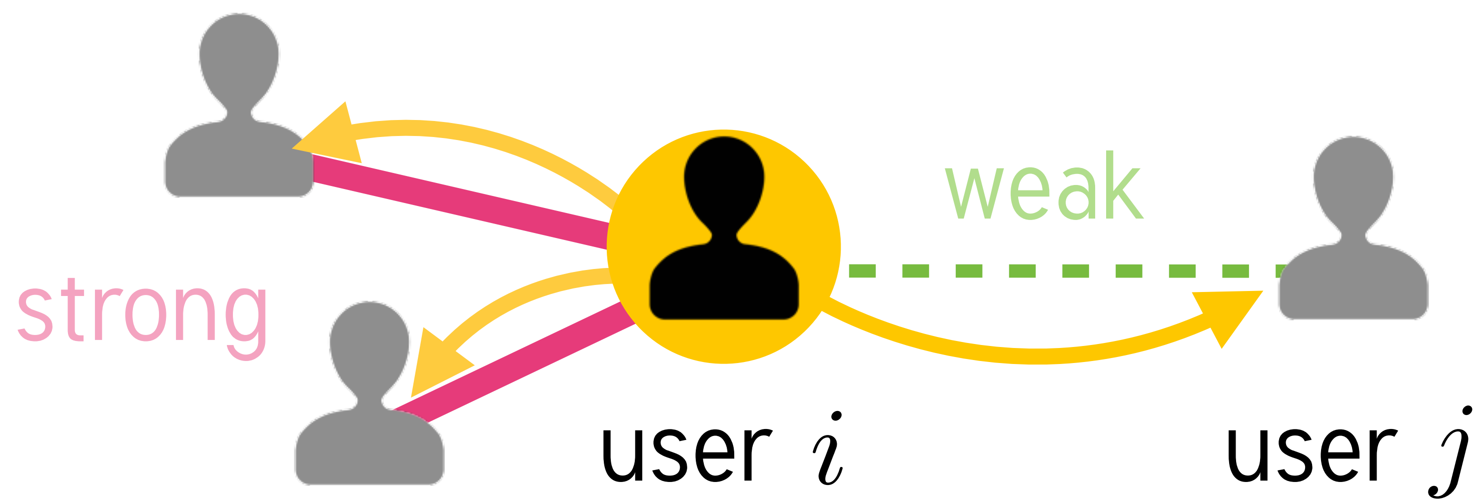
MEASURE OF ATTENTION



$a(i \to j) =$ Total amount of attention of user i $\times$ Frac. of activity of i directing to j

$$\frac{w(i,j)}{\sum_{u \in N_i} w(i,u)}$$

MEASURE OF ATTENTION

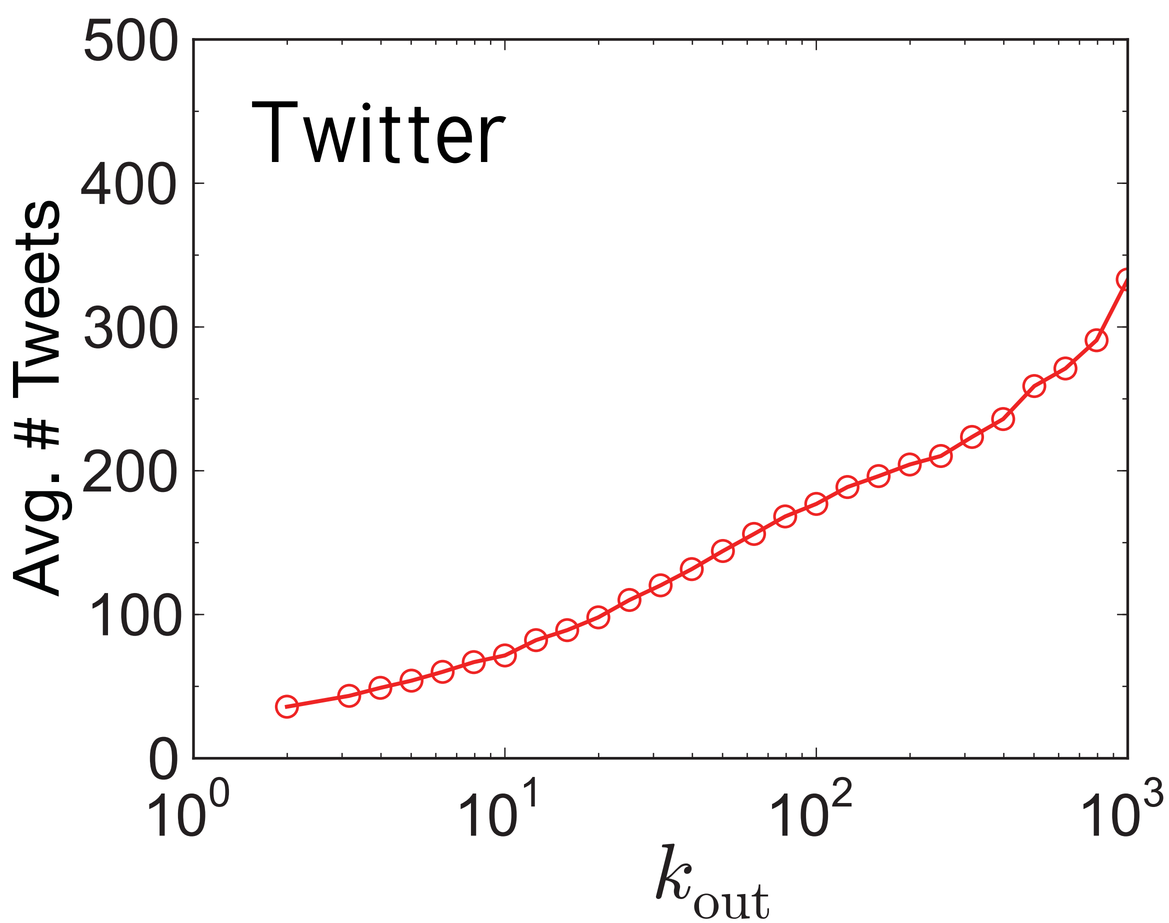


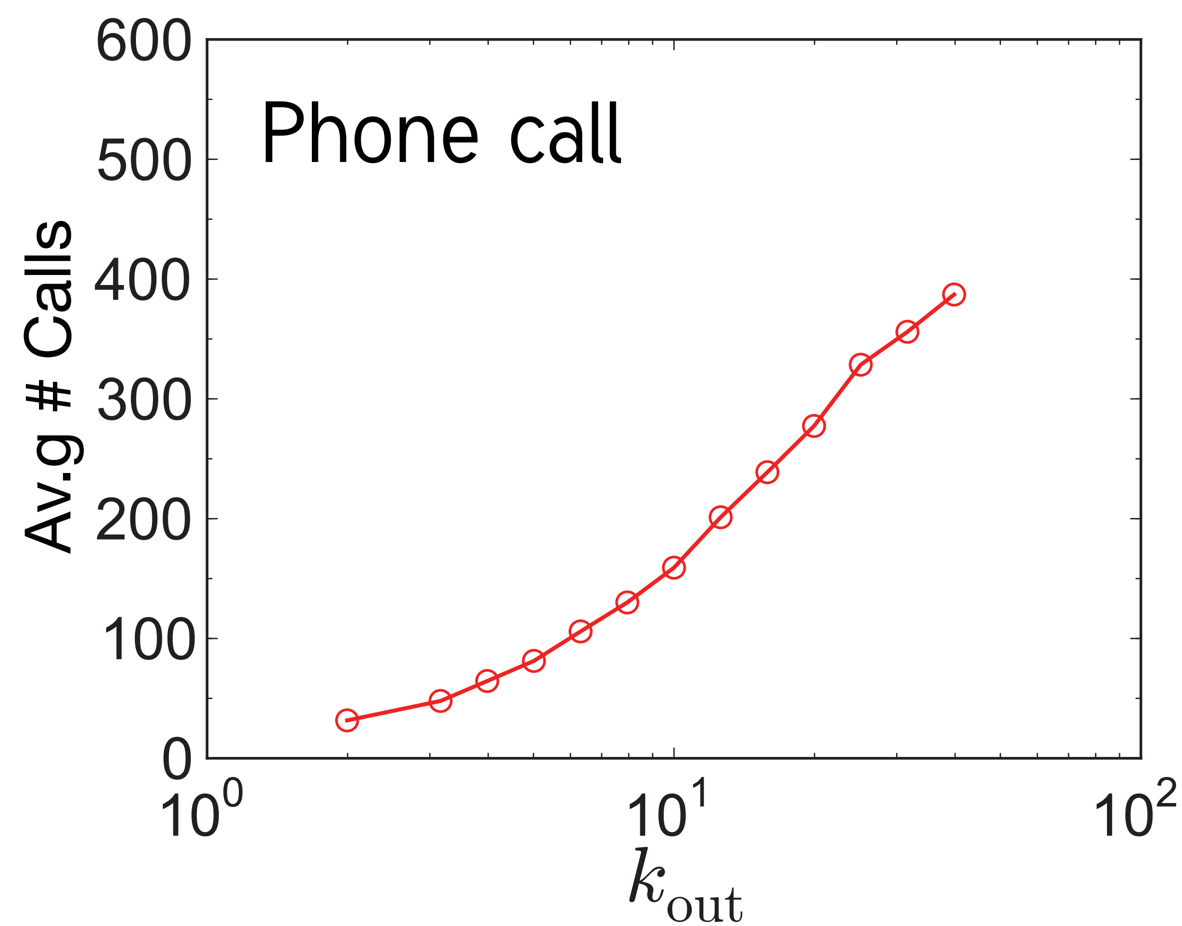
$$a(i \to j) = \boxed{\text{Total amount of attention of user } i} \times \boxed{\text{Frac. of activity of } i \text{ directing to } j}$$

tweets
calls

$$\frac{w(i,j)}{\sum_{u \in N_i} w(i,u)}$$

MEASURE OF ATTENTION





$a(i \to j) =$ Total amount of attention of user i $\times$ Frac. of activity of i directing to j

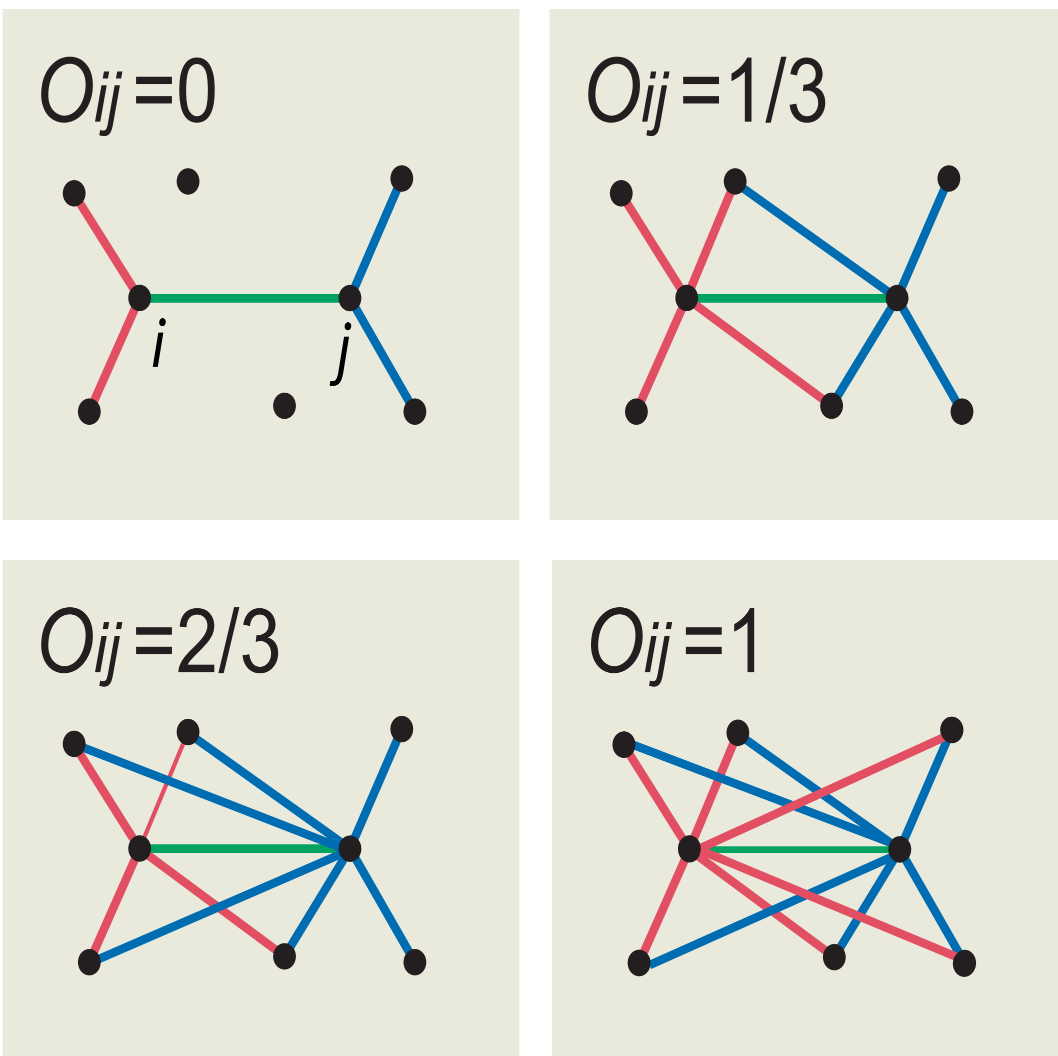
$\propto \log k_{\text{out}}(i)$

$$\frac{w(i,j)}{\sum_{u \in N_i} w(i,u)}$$

MEASURE OF ATTENTION

$$a(i,j) = \boxed{\log k_{\mathrm{out}}(i)} \cdot \boxed{\frac{w(i,j)}{\sum_{u \in N_i} w(i,u)}}$$

MEASURE OF TIE STRENGTH



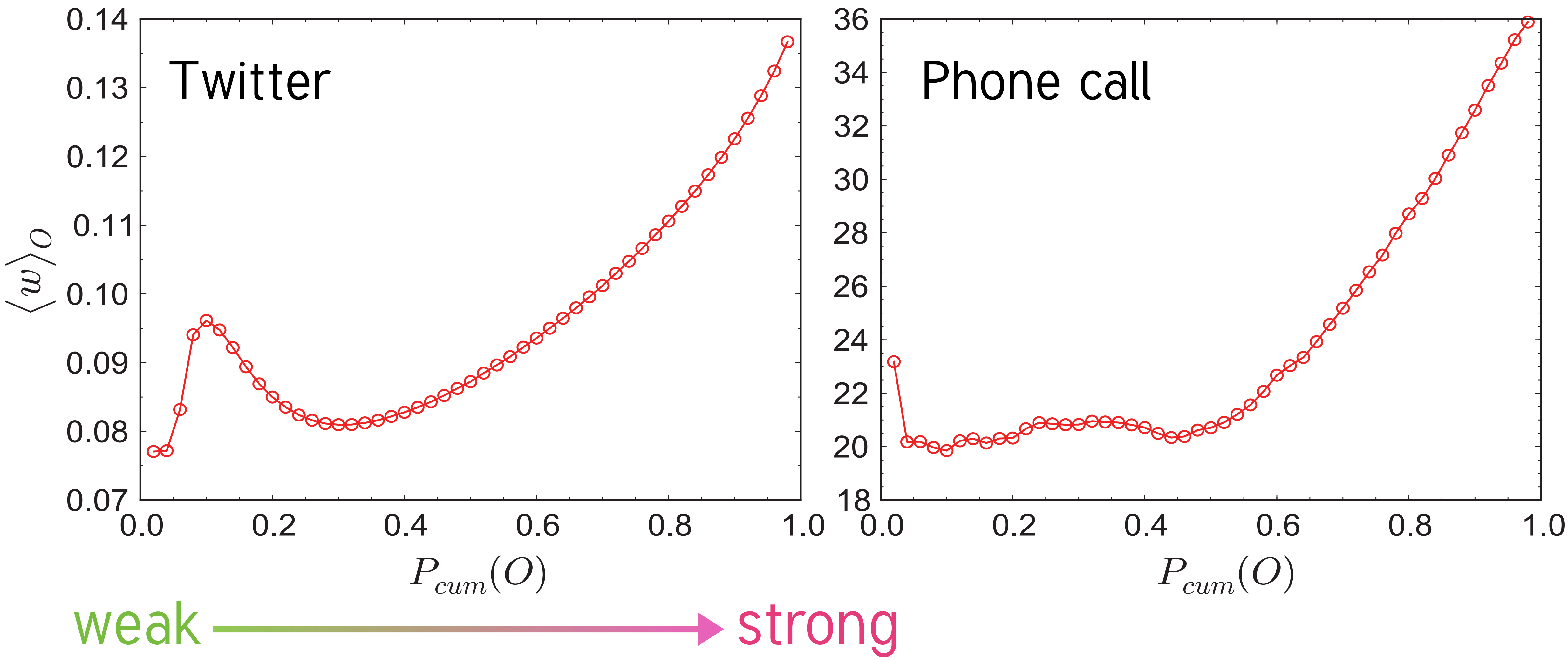
(Onnela et al.,2007)

Jaccard between neighbor sets:

$$O_{ij} = \frac{|N_i \cap N_j \setminus \{i, j\}|}{|N_i \cup N_j \setminus \{i, j\}|}$$

$$N_i = \{k \mid (i, k) \in E \wedge (k, i) \in E\}$$

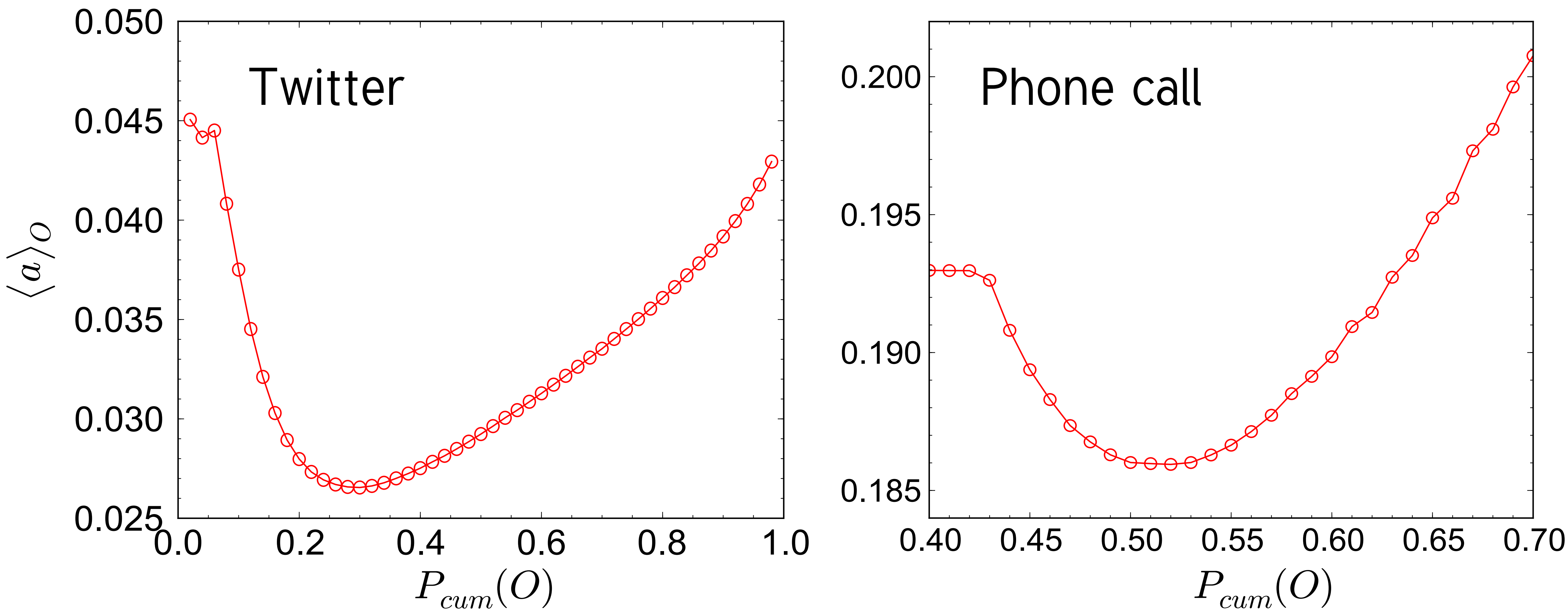
TRAFFIC ON STRONG TIES



People communicate more with strong ties

(Weng et al. 2014)

ATTENTION ON WEAK TIES

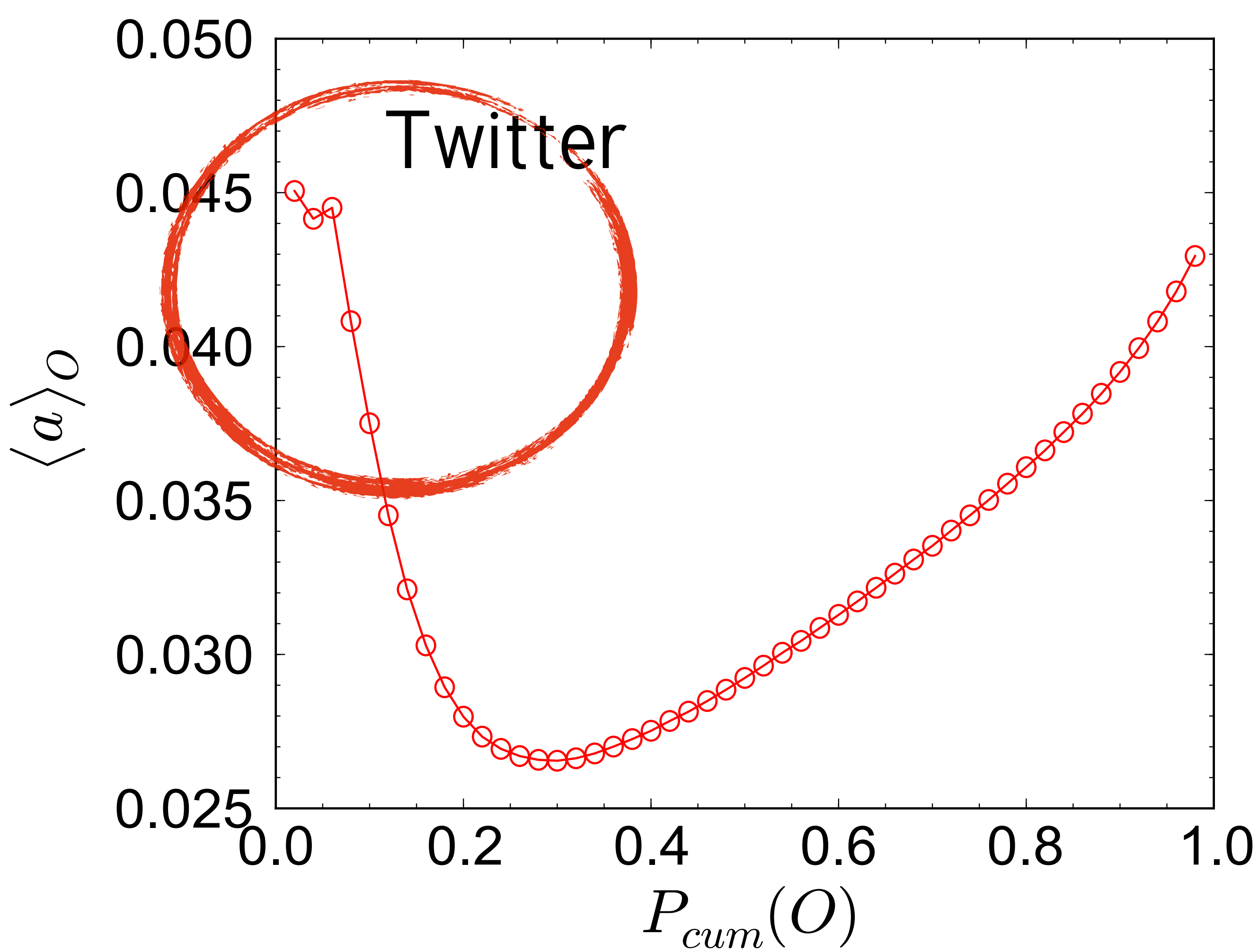


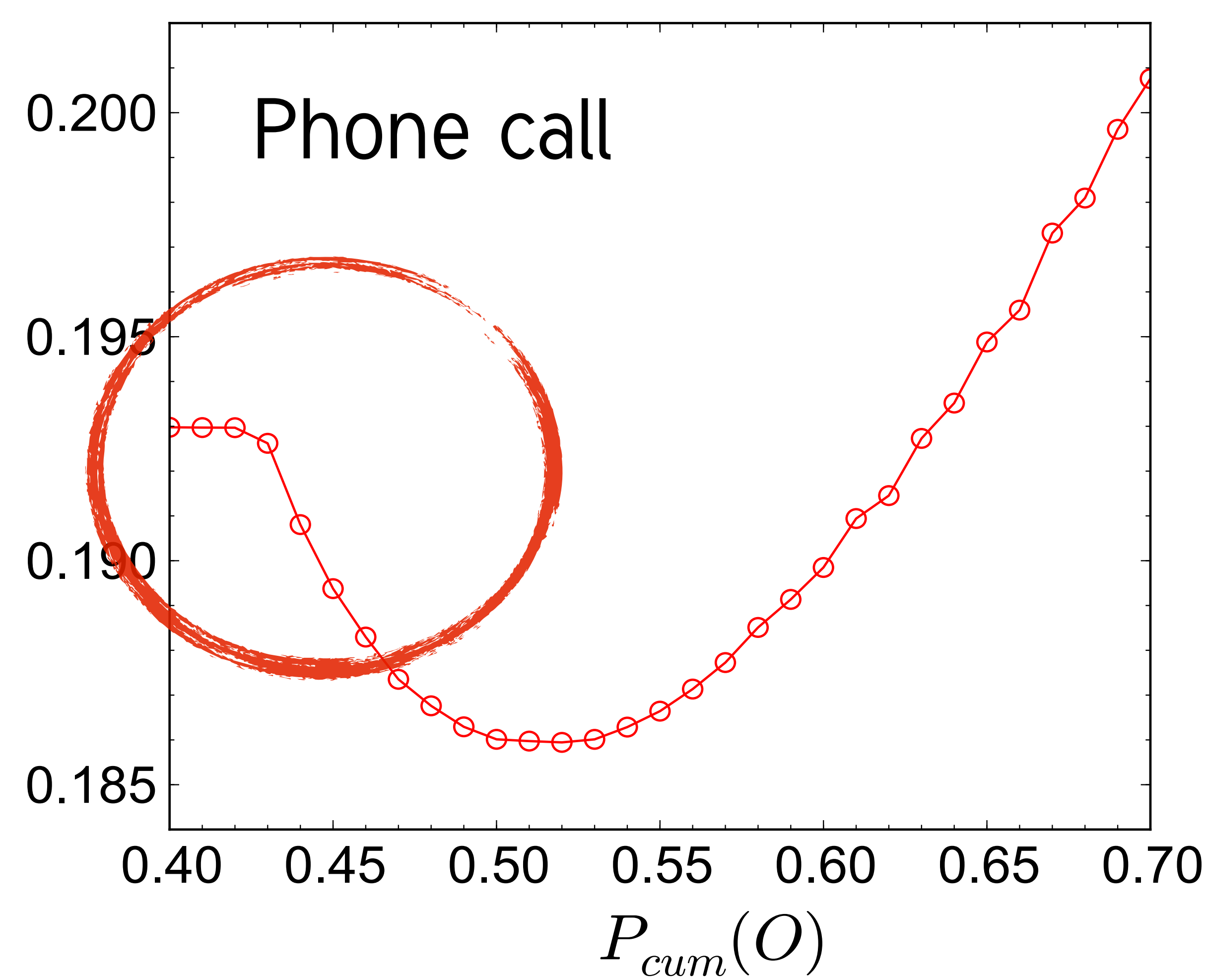
weak → strong

Very weak ties attract much attention

(Weng et al. 2014)

ATTENTION ON WEAK TIES





weak → strong

Very weak ties attract much attention

(Weng et al. 2014)

WHAT DO WE LEARN?

"*… a poverty of attention*"

- Heterogeneous dynamics of meme popularity

"*… a need to allocate that attention*"

- Weak ties attract much attention

CONTENT

Part Two

CONTENT

- **Innate** appeal of content
  - Emotion (Berger and Milkman,2009)
  - Text formation (Tsur and Rappoport,2012)
- **Topic** Identification
  - LDA (Weng et al., 2010)
  - Folksonomy (Michelson and Macskassy, 2010)
  - User groups (Java et al., 2007)
  - Topic locality (Davison, 2000)
- **Diversity** of user interests (An et al., 2011)

CONTENT

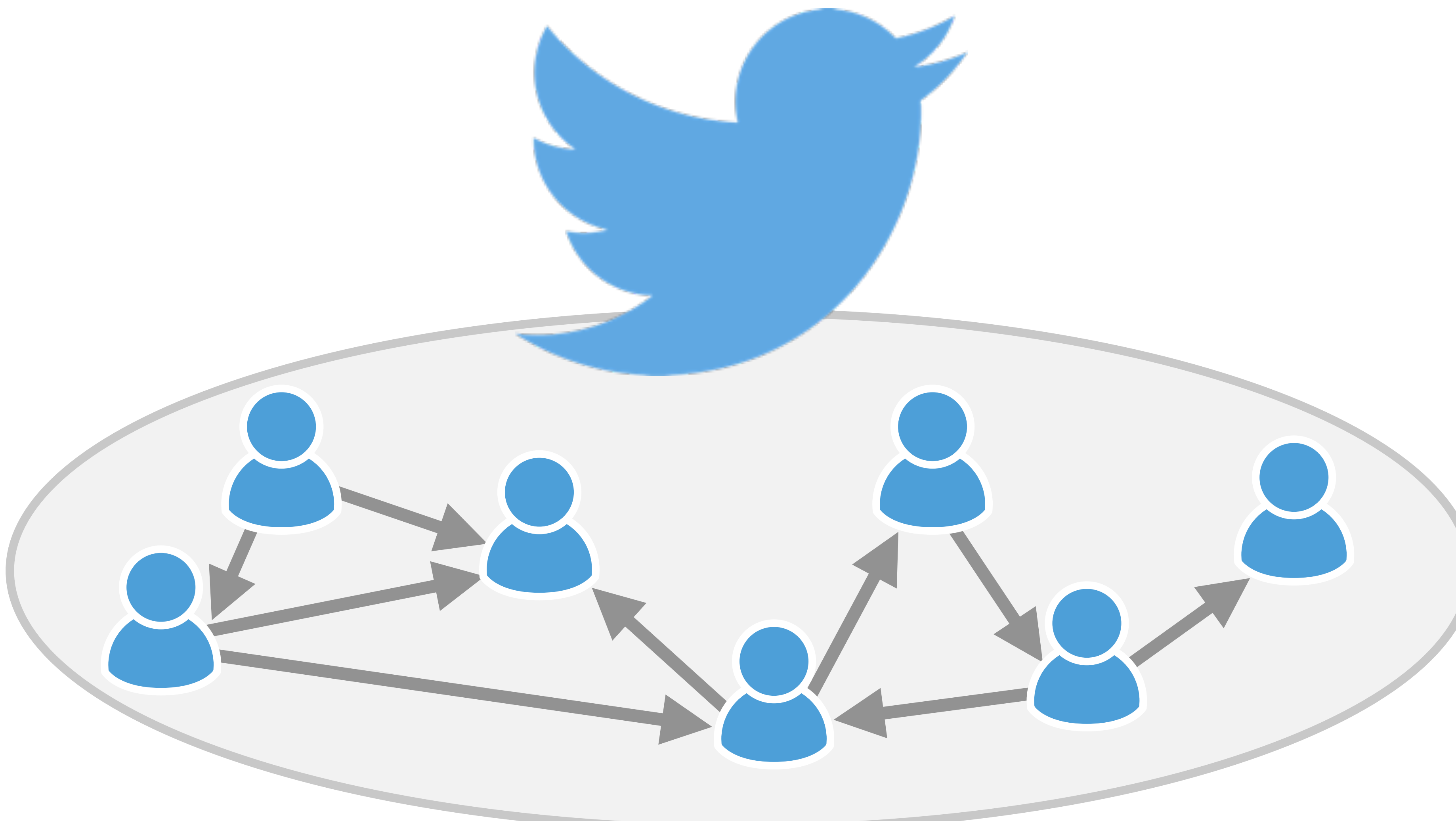
- Innate appeal of content
  - Emotion (Berger and Milkman,2009)
  - Text formation (Tsur and Rappoport,2012)
- Topic Identification
  - LDA (Weng et al., 2010)
  - Folksonomy (Michelson and Macskassy, 2010)
  - User groups (Java et al., 2007)
  - Topic locality (Davison, 2000)
- Diversity of user interests (An et al., 2011)

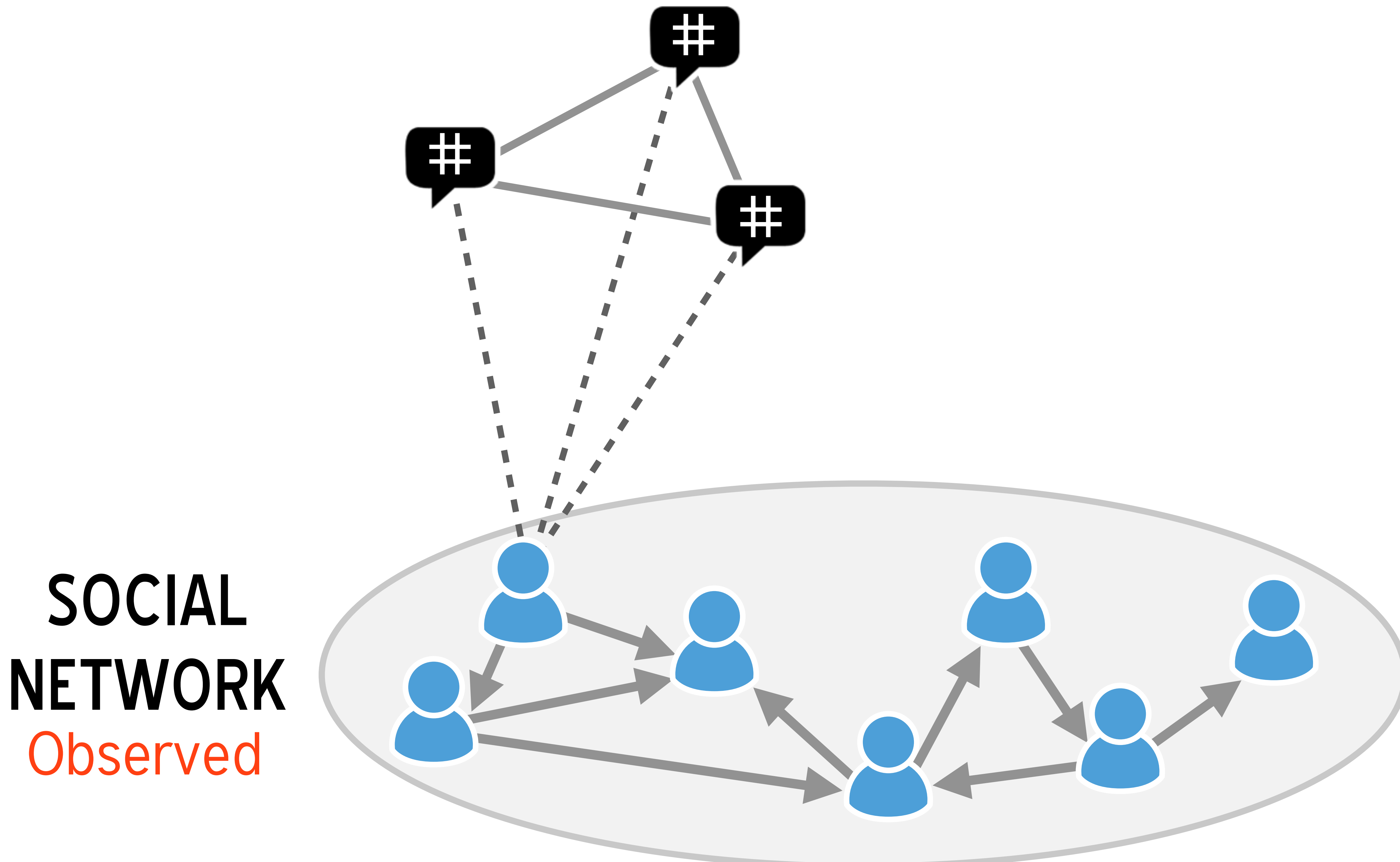
Can we detect **topics** in social media?

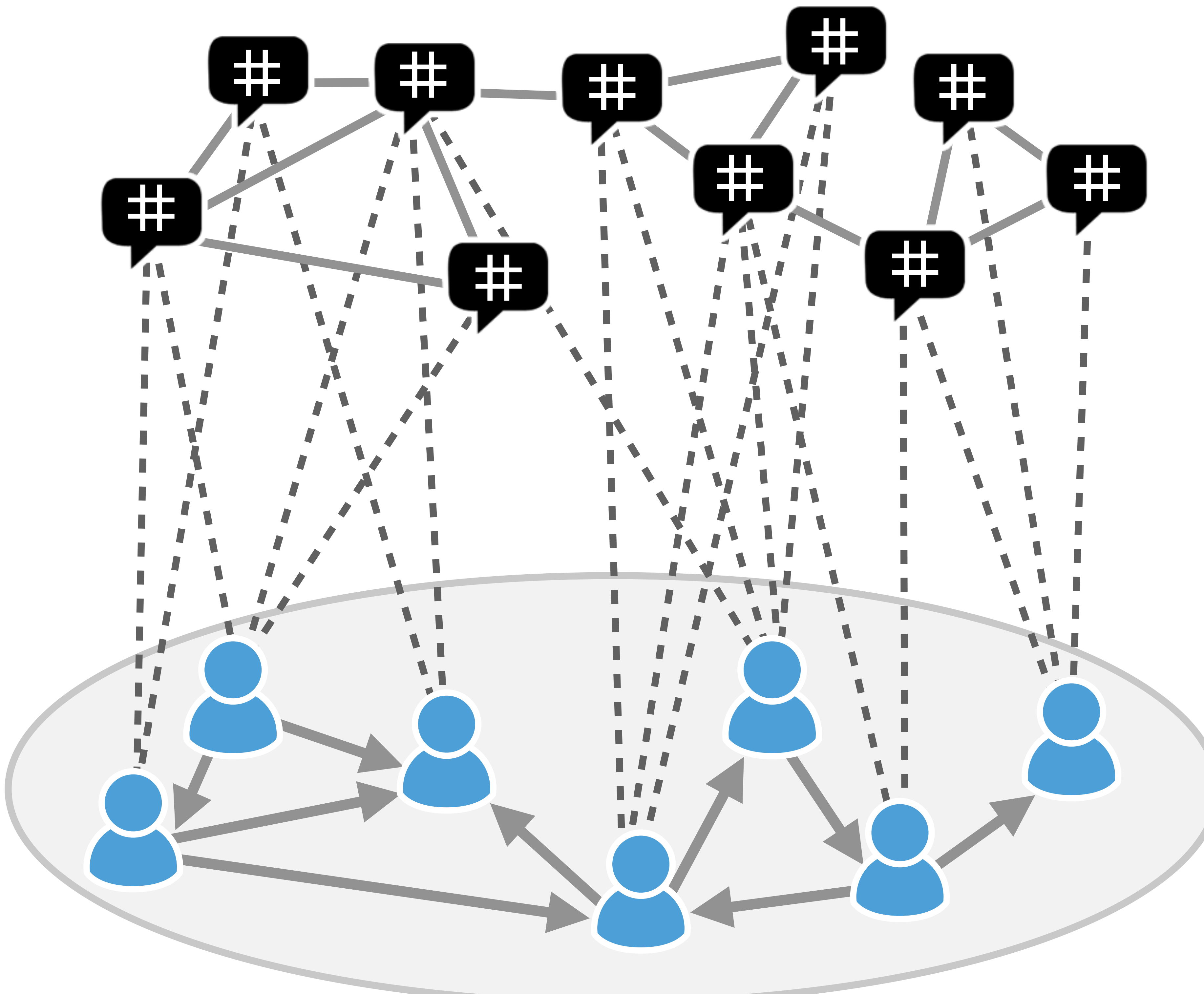
Topic locality

SOCIAL NETWORK

Observed

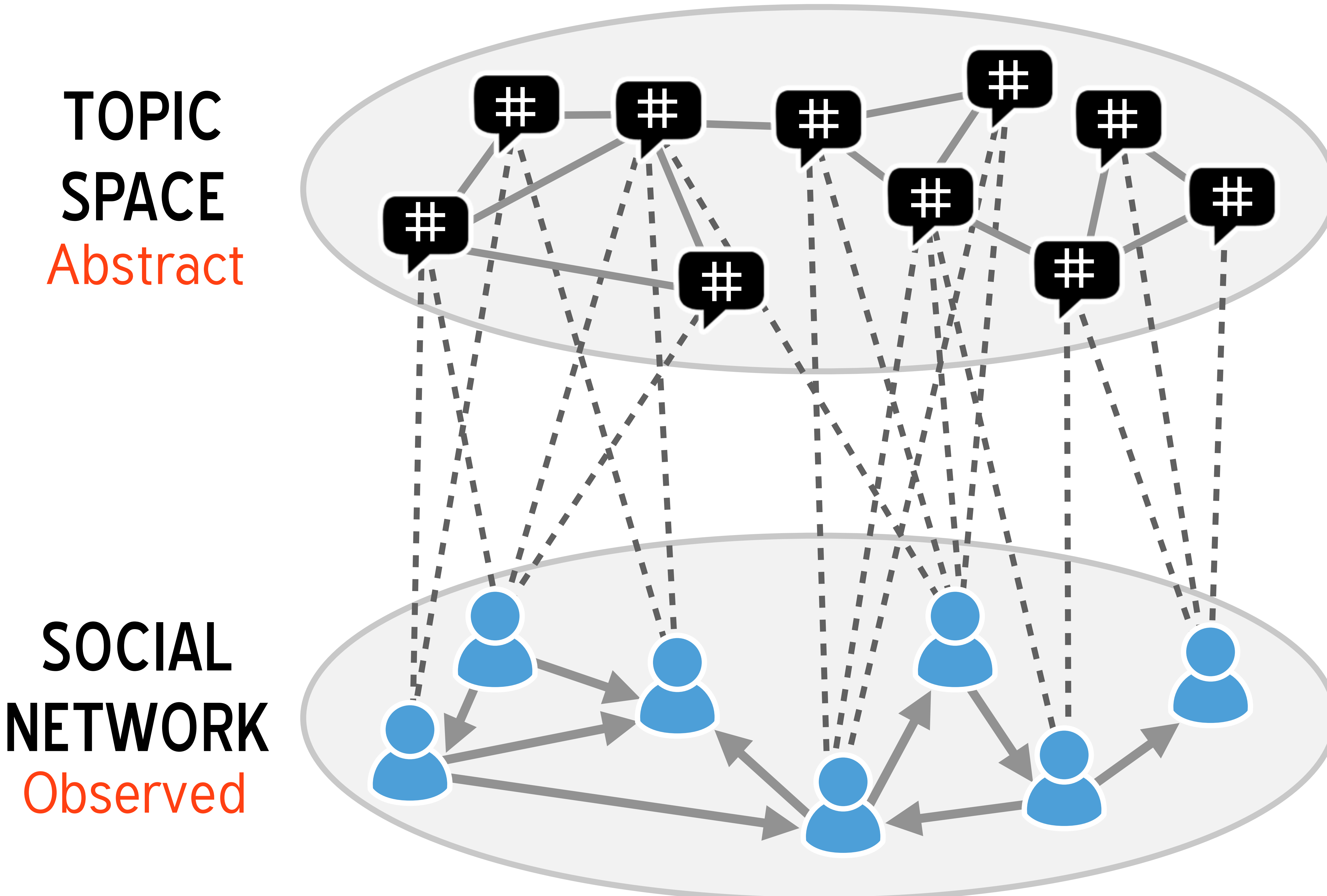


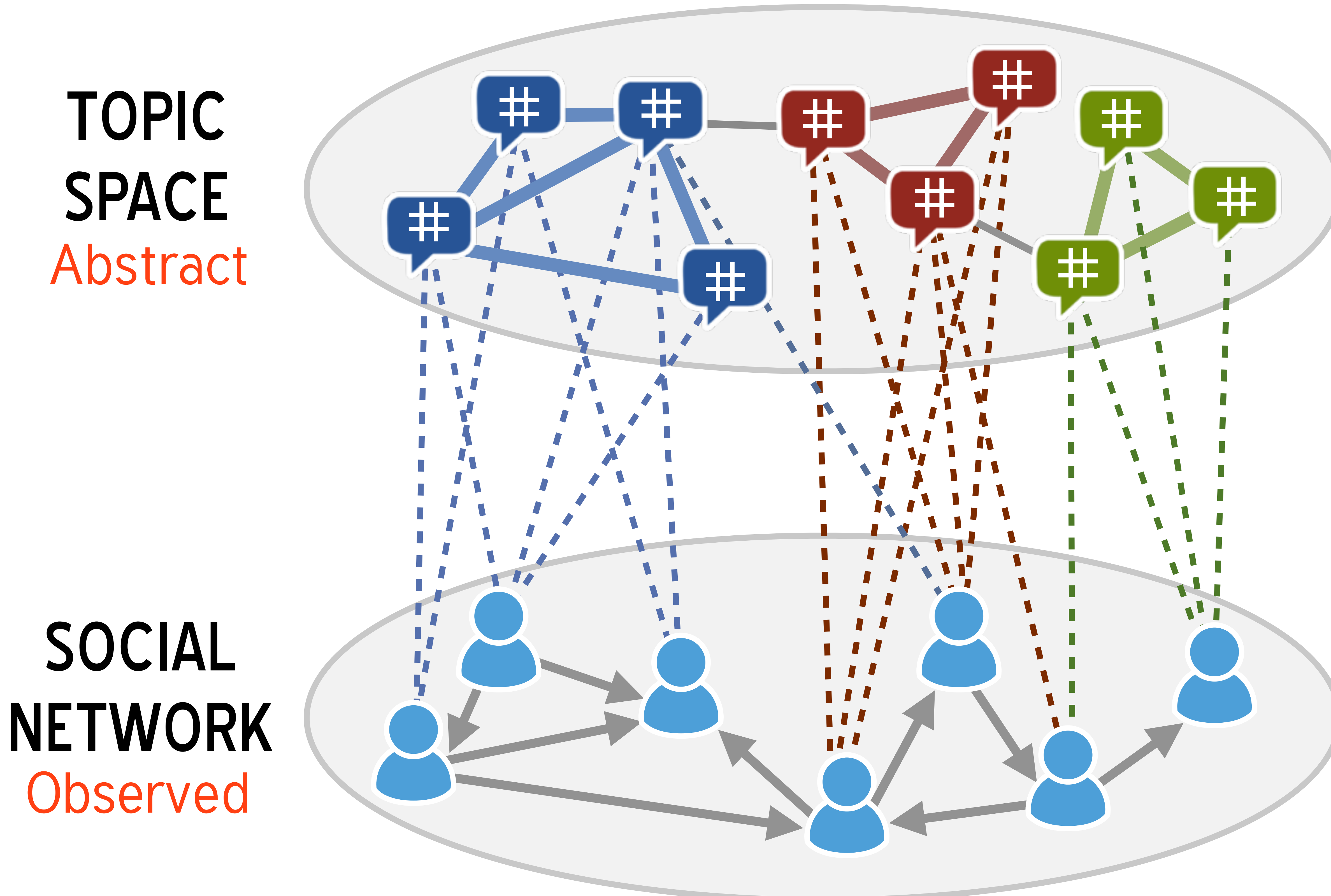

SOCIAL
NETWORK
Observed



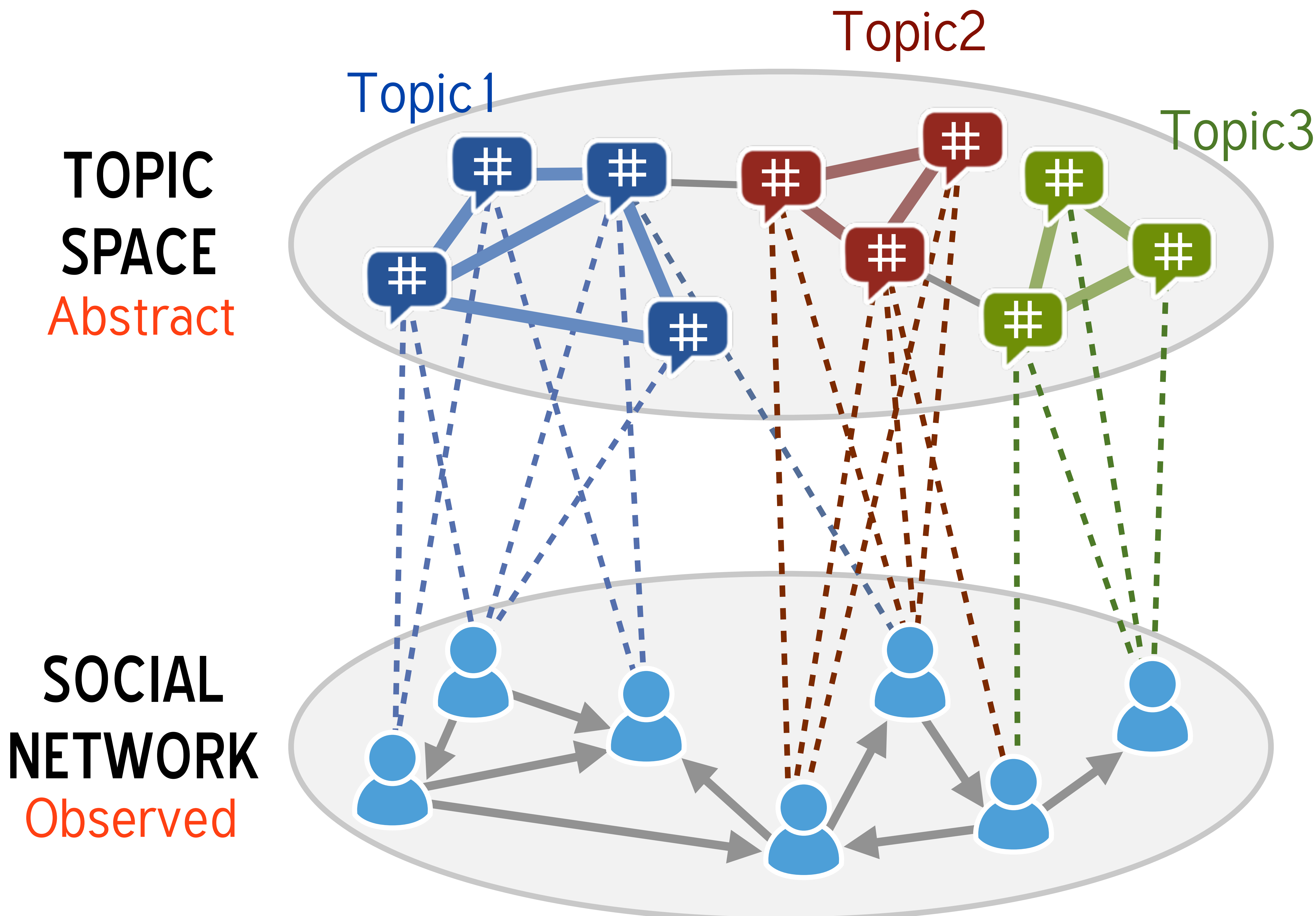
SOCIAL NETWORK

Observed


TOPIC
SPACE
Abstract
SOCIAL
NETWORK
Observed



TOPIC
SPACE
Abstract
SOCIAL
NETWORK
Observed


Topic2
Topic1
Topic3
TOPIC
SPACE
Abstract
SOCIAL
NETWORK
Observed

TECHNOLOGY

- #google, #microsoft, #supercomputers, #ibm, #wikipedia, #pinterest, #startuptip, #topworkplace

LIFESTYLE

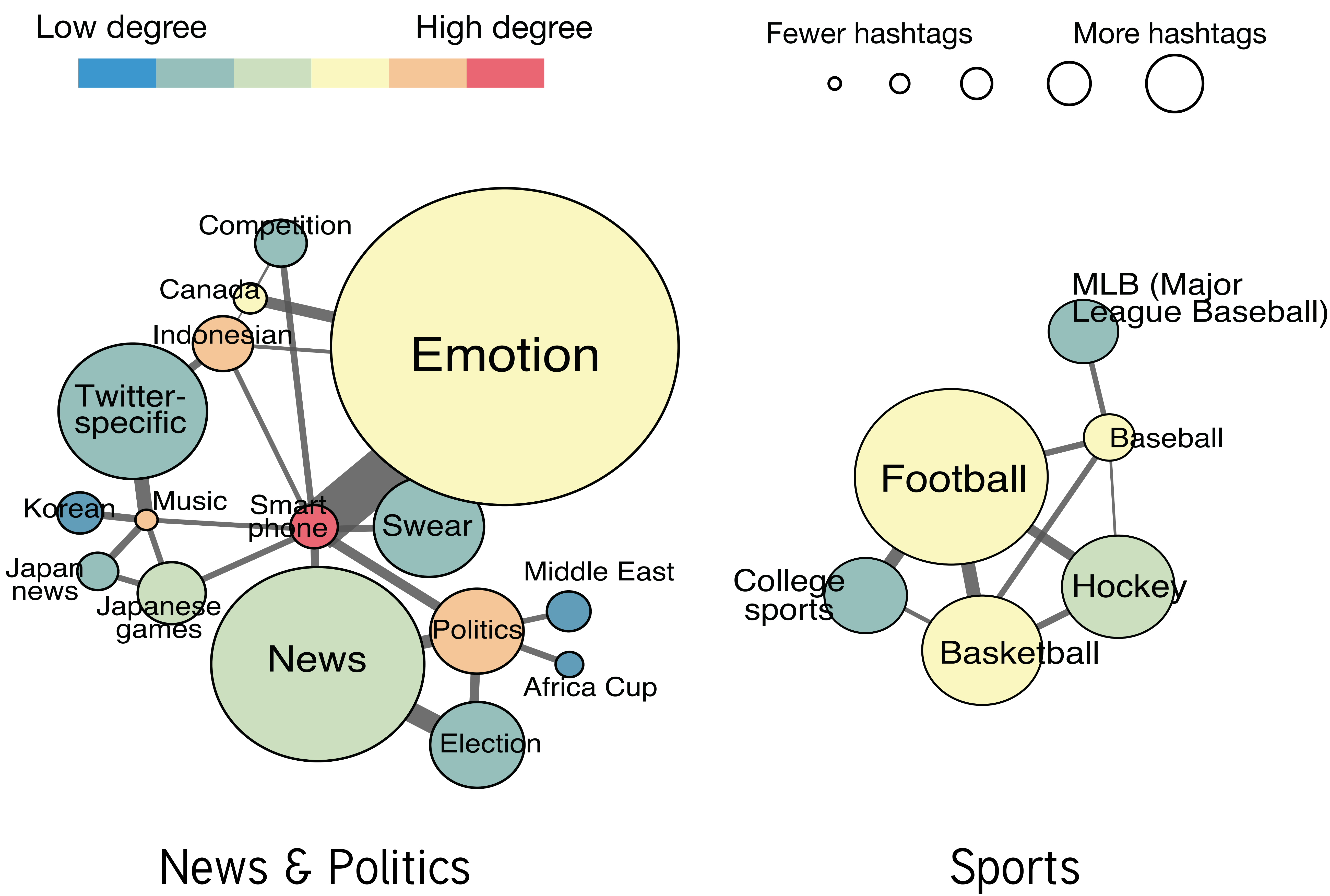
- #tcot, #p2, #top, #usgovernment, #dems, #owe, #politics, #teaparty #pizza, #pepsi, #cheese, #health, #vacation, #caribbean, #ford, #honda, #volkswagen, #hm

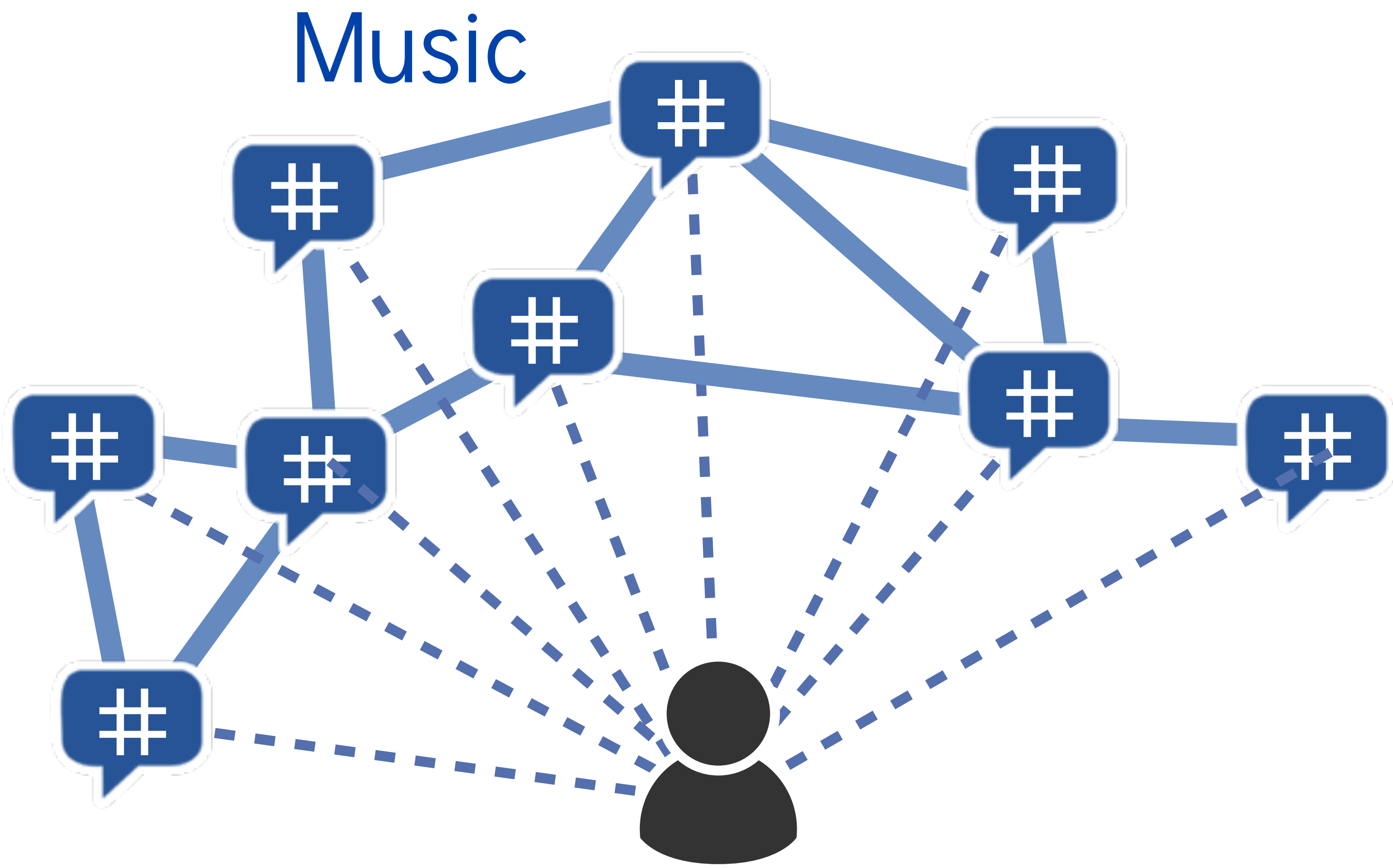
TWITTER-SPECIFIC

- #followme, #followback, #teamfollowback, #followfriday, #friday, #justretweet, #instantfollower, #rt

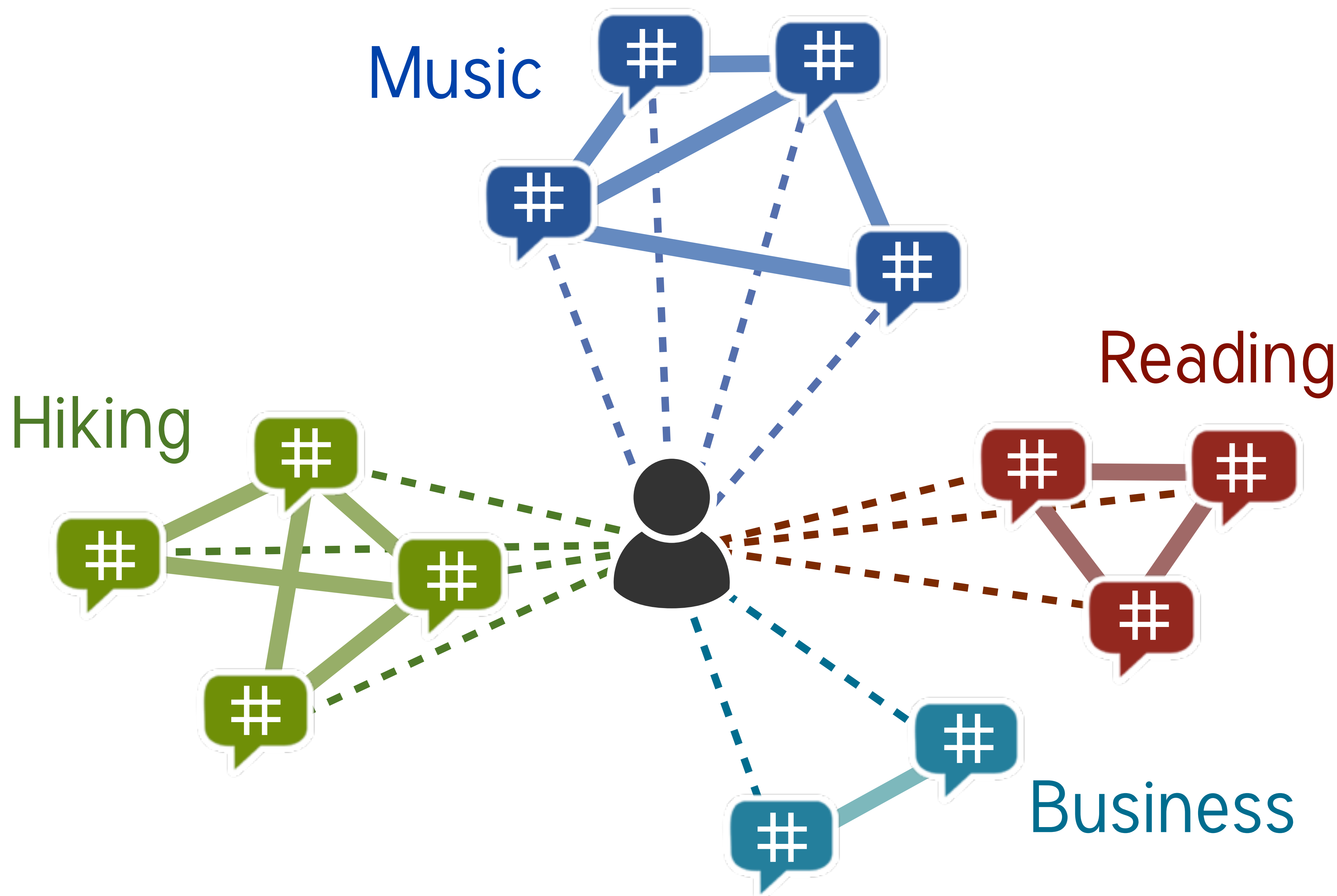
MOBILE DEVICE

- #apple, #galaxy3, #note2, #iosapp, #mp3player, #releases

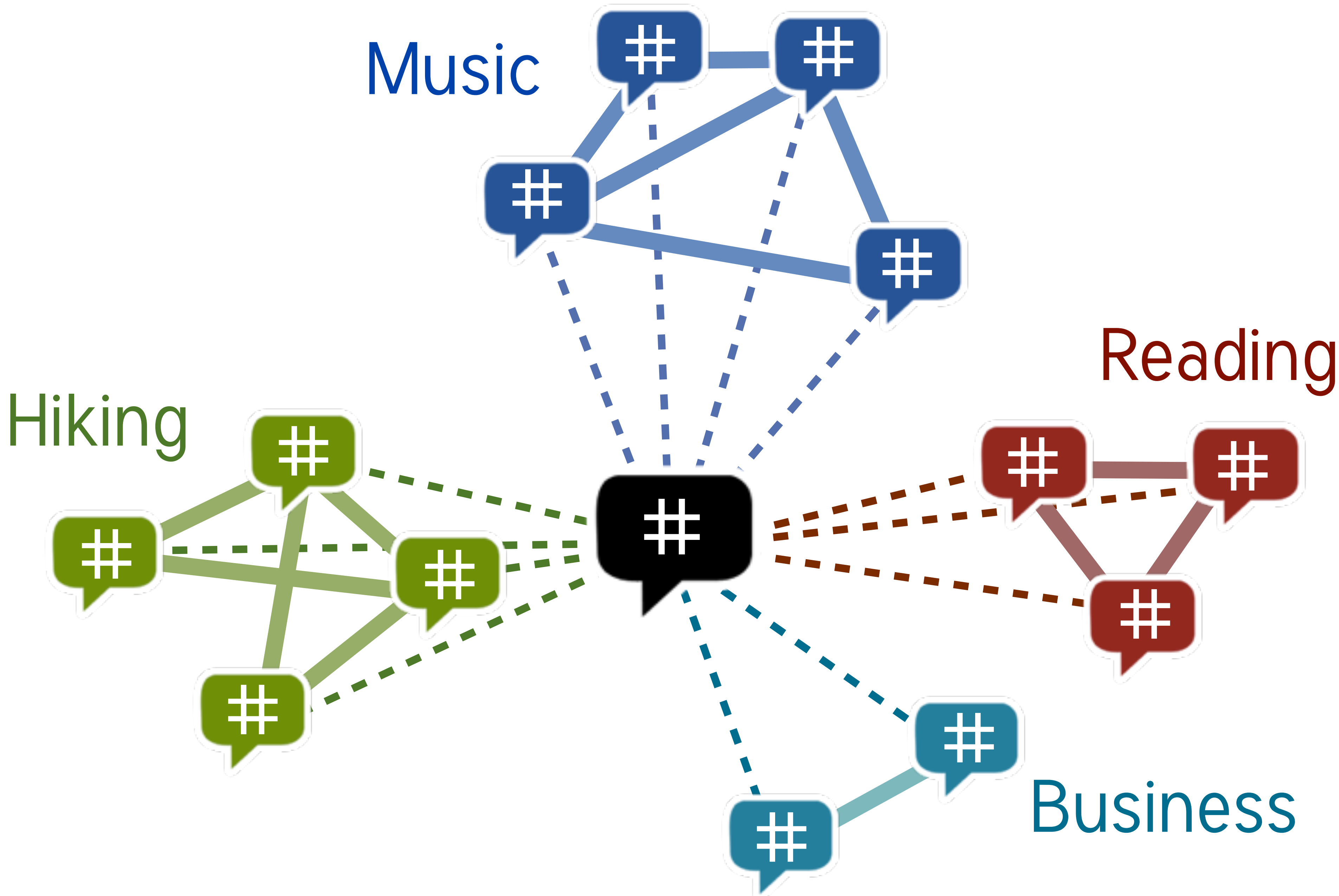

Low degree
High degree
Fewer hashtags
More hashtags
Competition
Canada
Indonesian
Emotion
Twitter-specific
Korean
Music
Smart phone
Swear
Japan news
Japanese games
News
Politics
Middle East
Africa Cup
Election
MLB (Major League Baseball)
Baseball
Football
College sports
Hockey
Basketball
News & Politics
Sports



Focused Interests



Diverse Interests



Diverse Interests

TOPICAL DIVERSITY

Entropy of topics associated with hashtags

used by a **user**

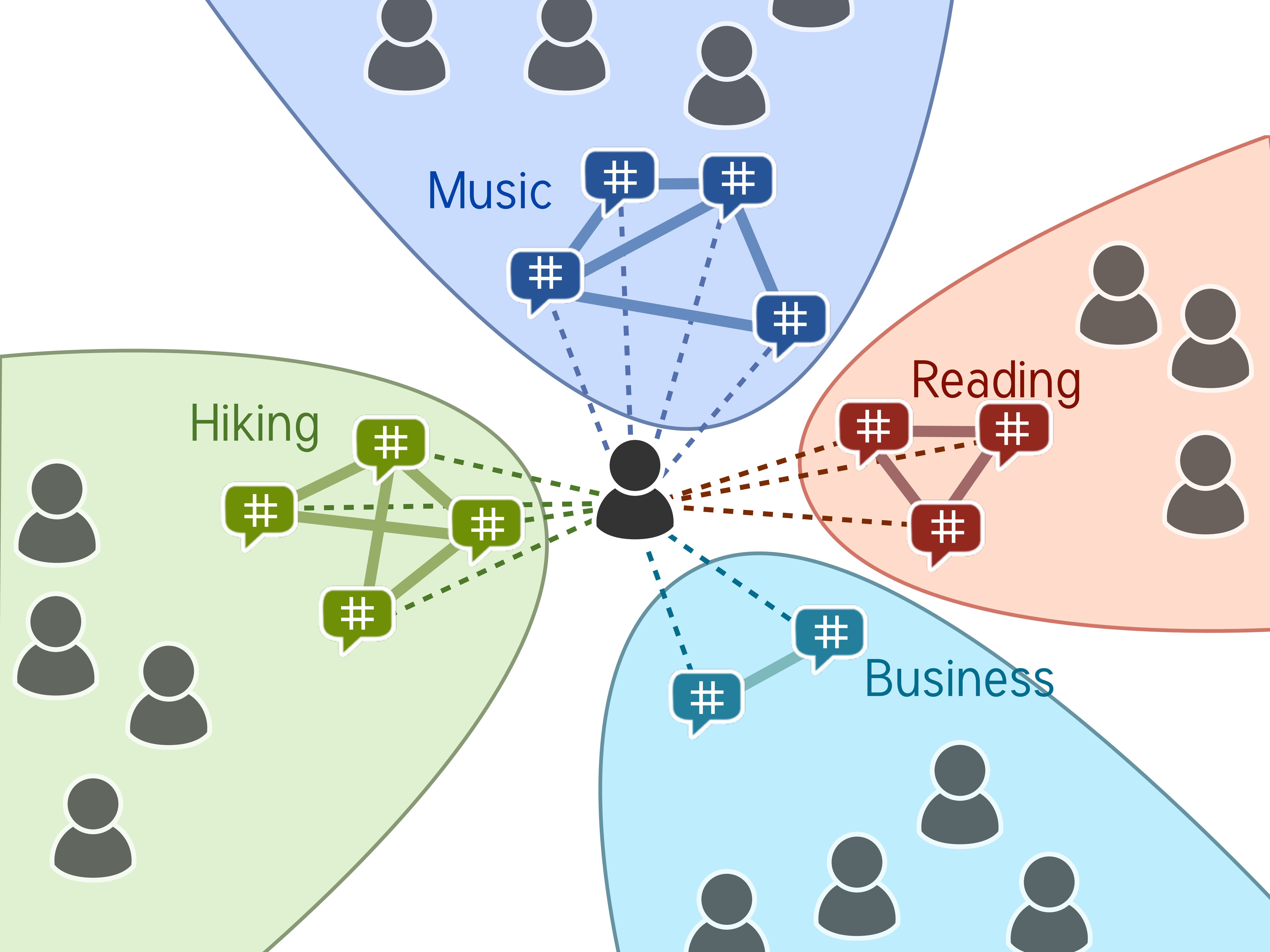
occurring with a **hashtag**

Small entropy
Focused interests

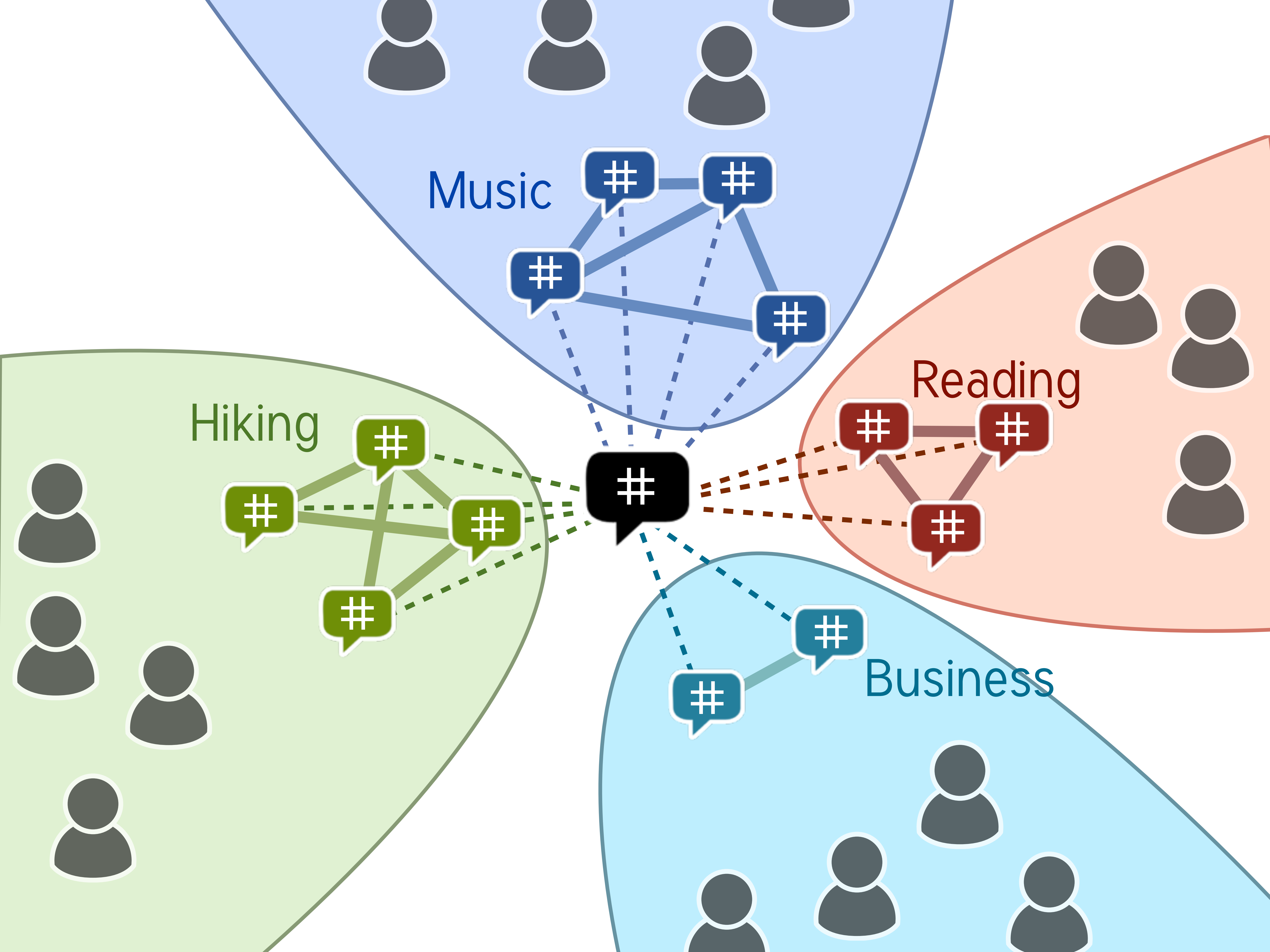
Large entropy
Broad interests

(Weng et al. 2014)

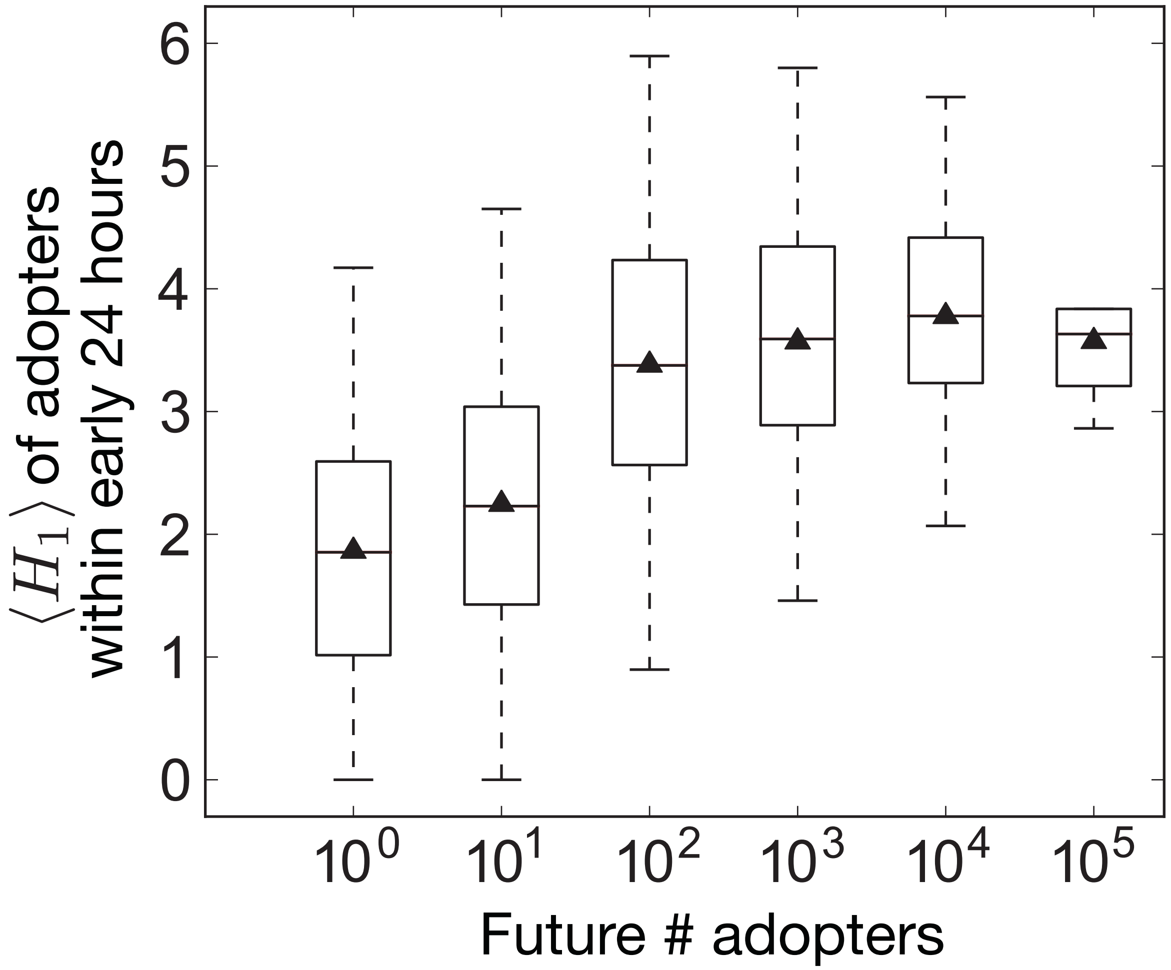
How does topical diversity affect user and content popularity?



Music
Reading
Hiking
Business

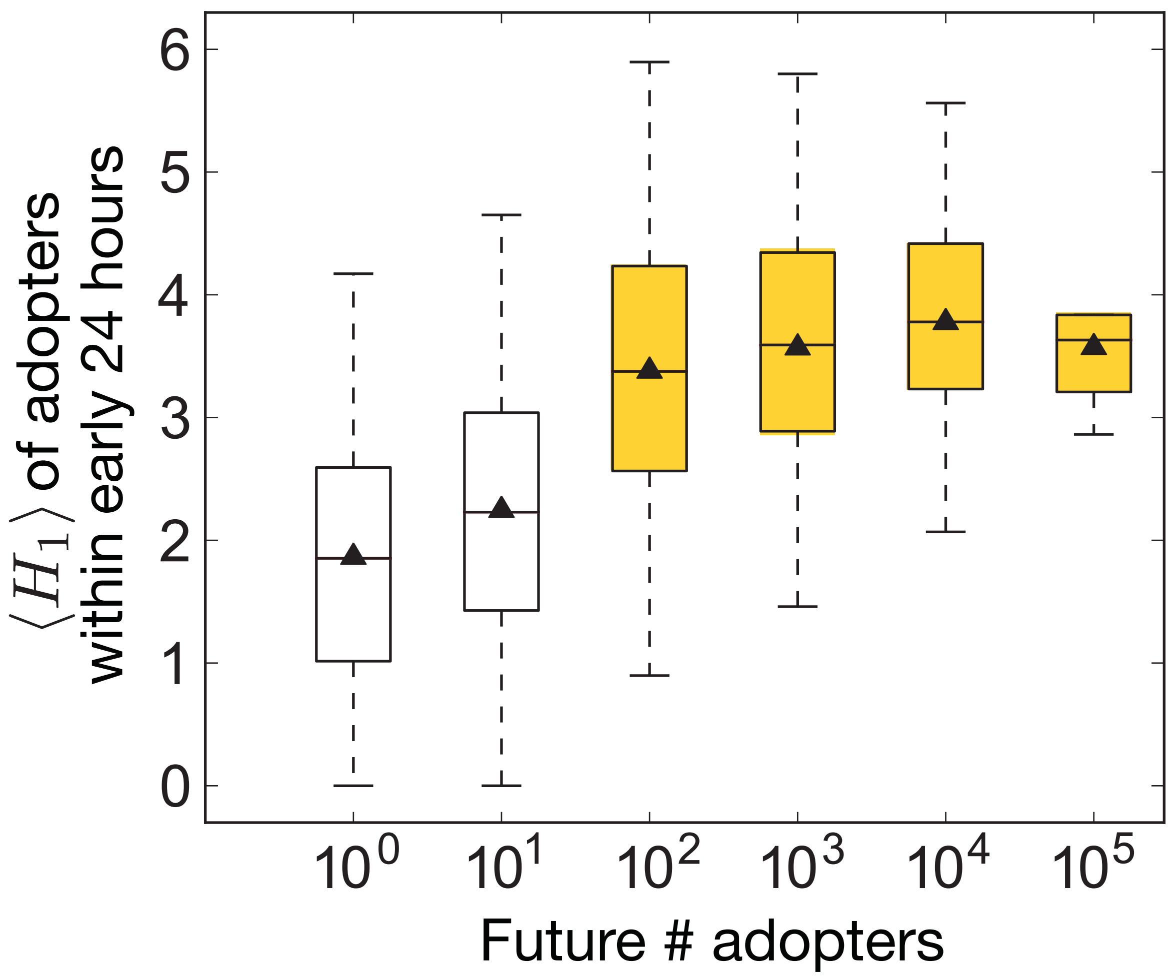

Music
Reading
Hiking
Business

VIRAL HASHTAGS



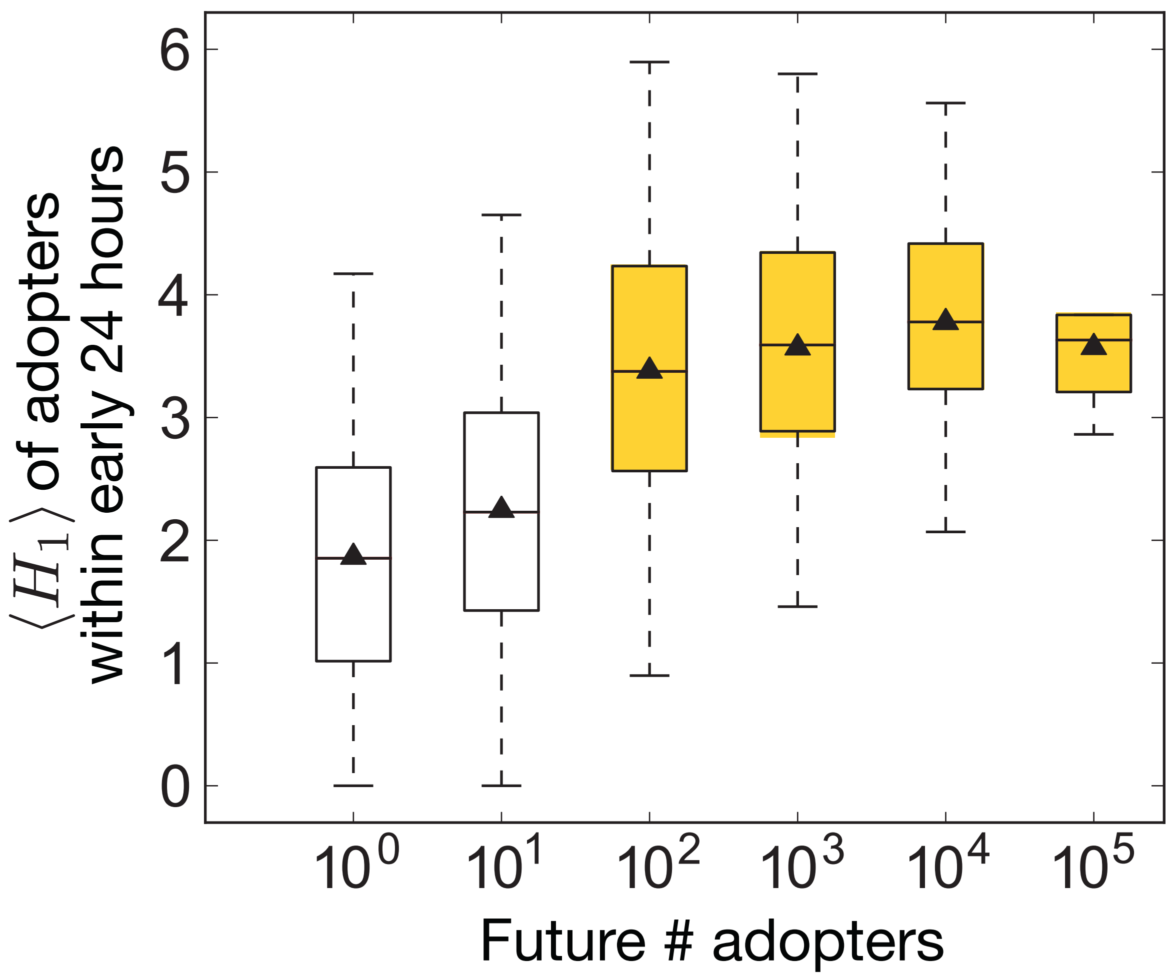
- Early adopters have diverse topical interests

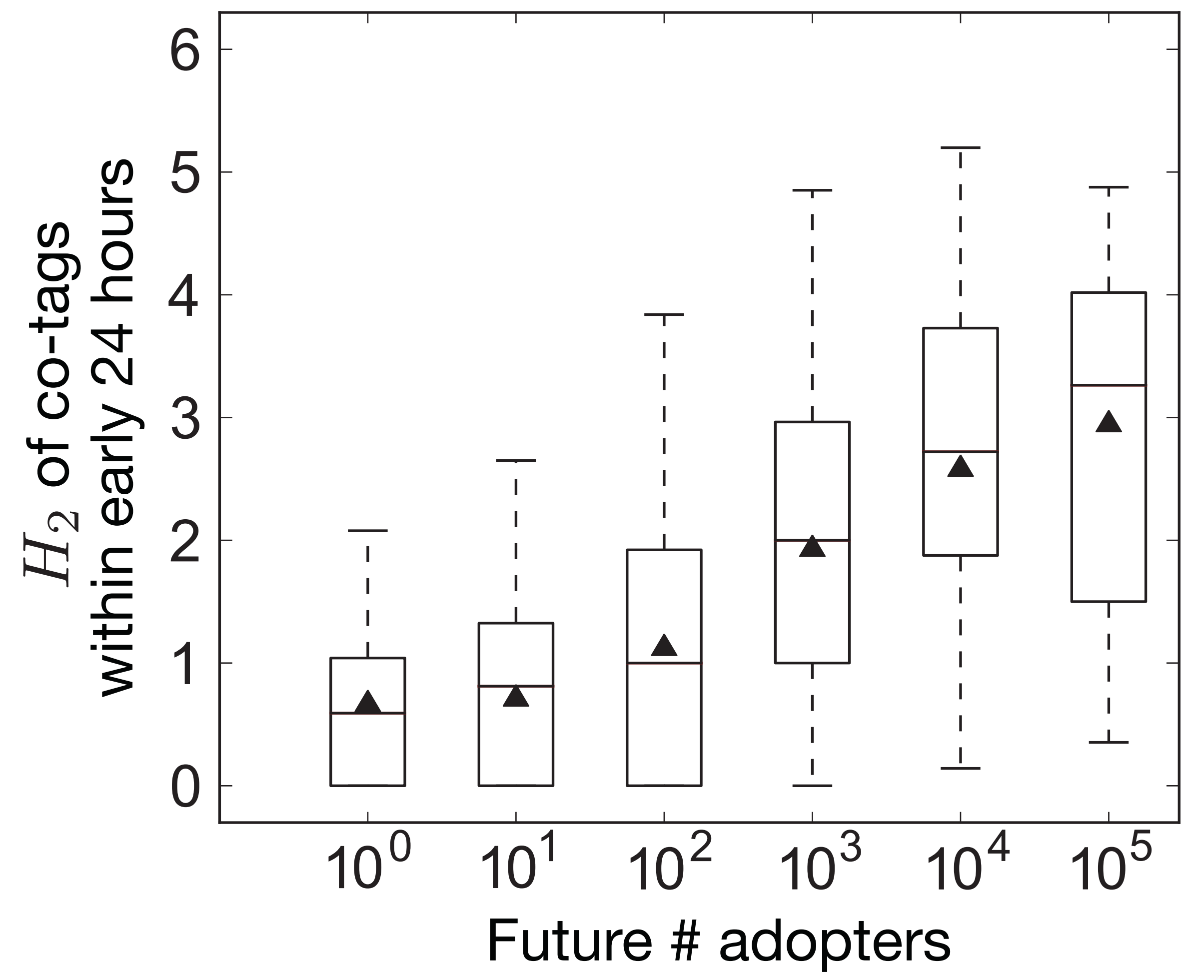
VIRAL HASHTAGS



- Early adopters have diverse topical interests

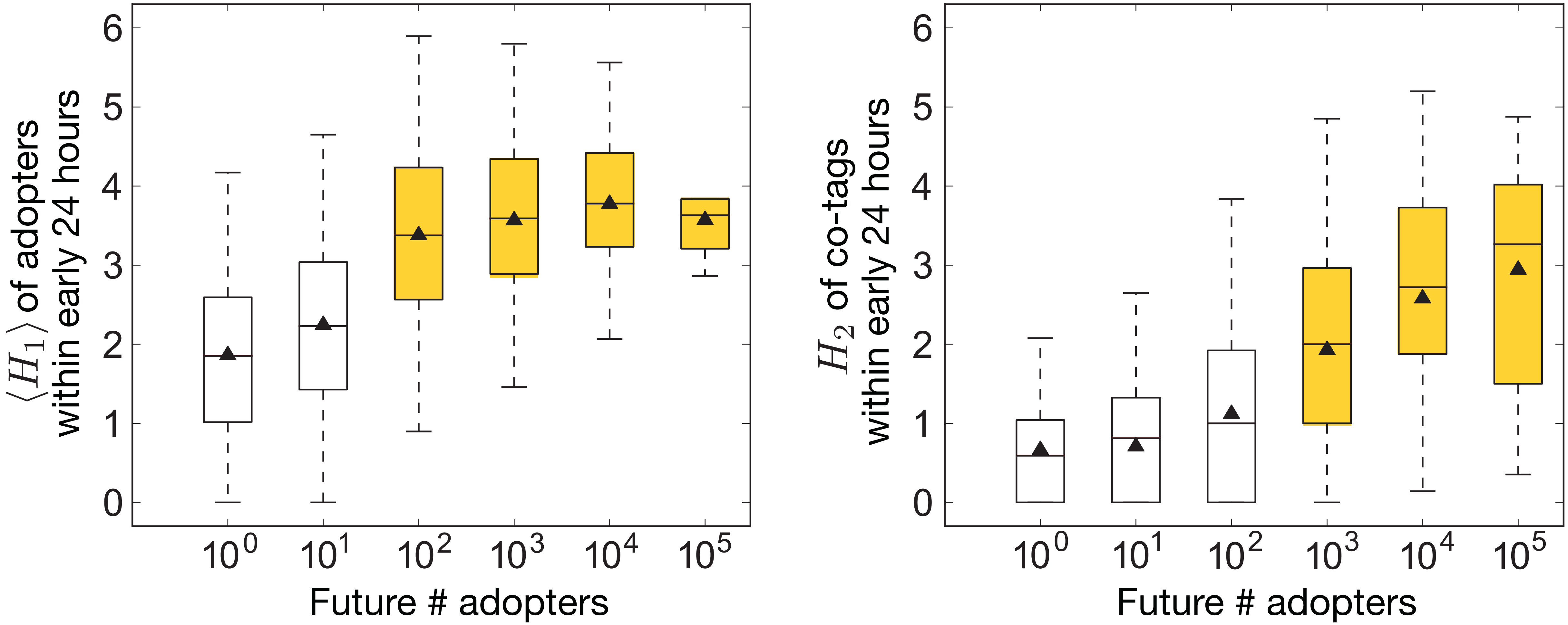
VIRAL HASHTAGS





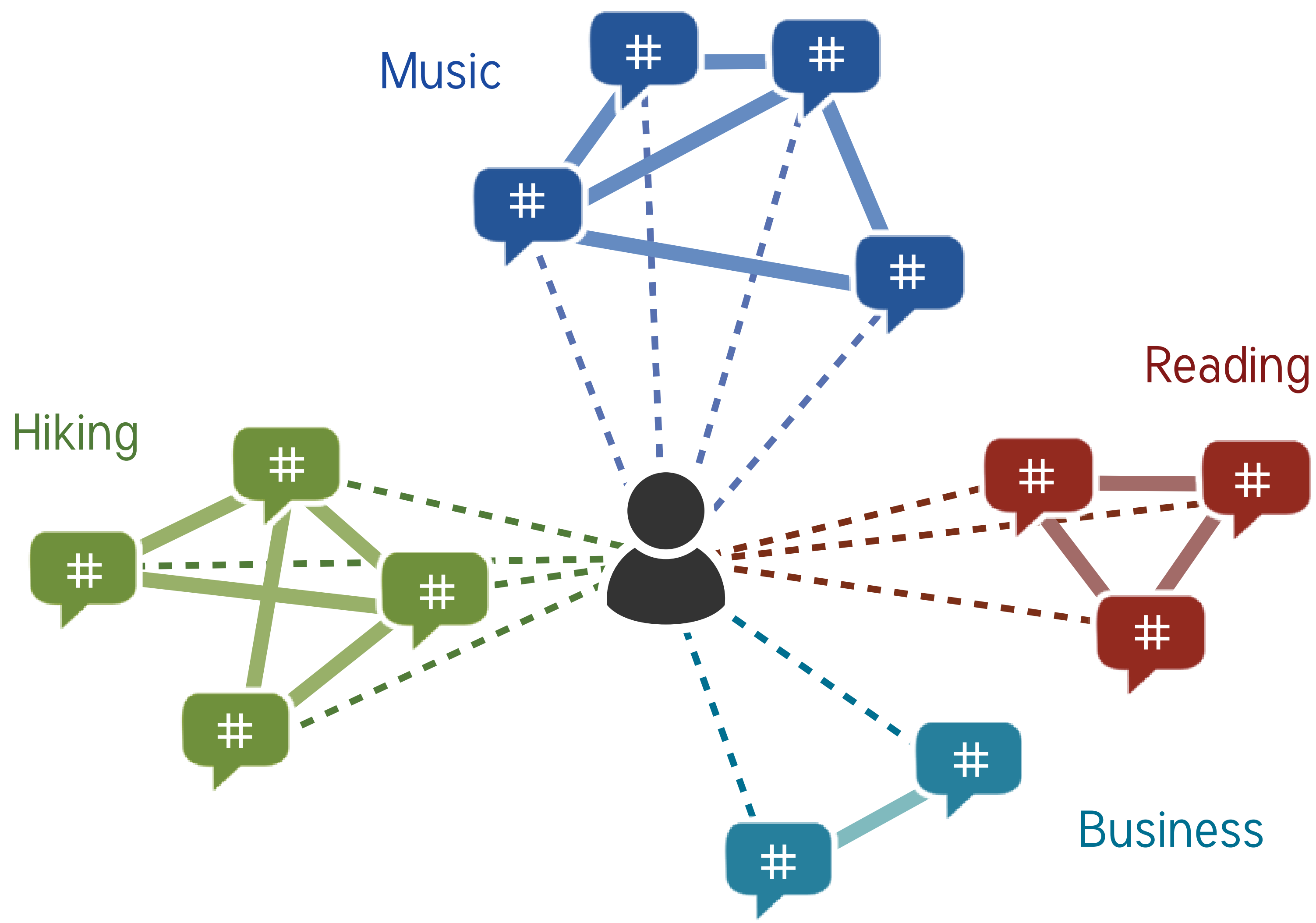
- Early adopters have diverse topical interests
- Co-occur with other tags about diverse topics

VIRAL HASHTAGS

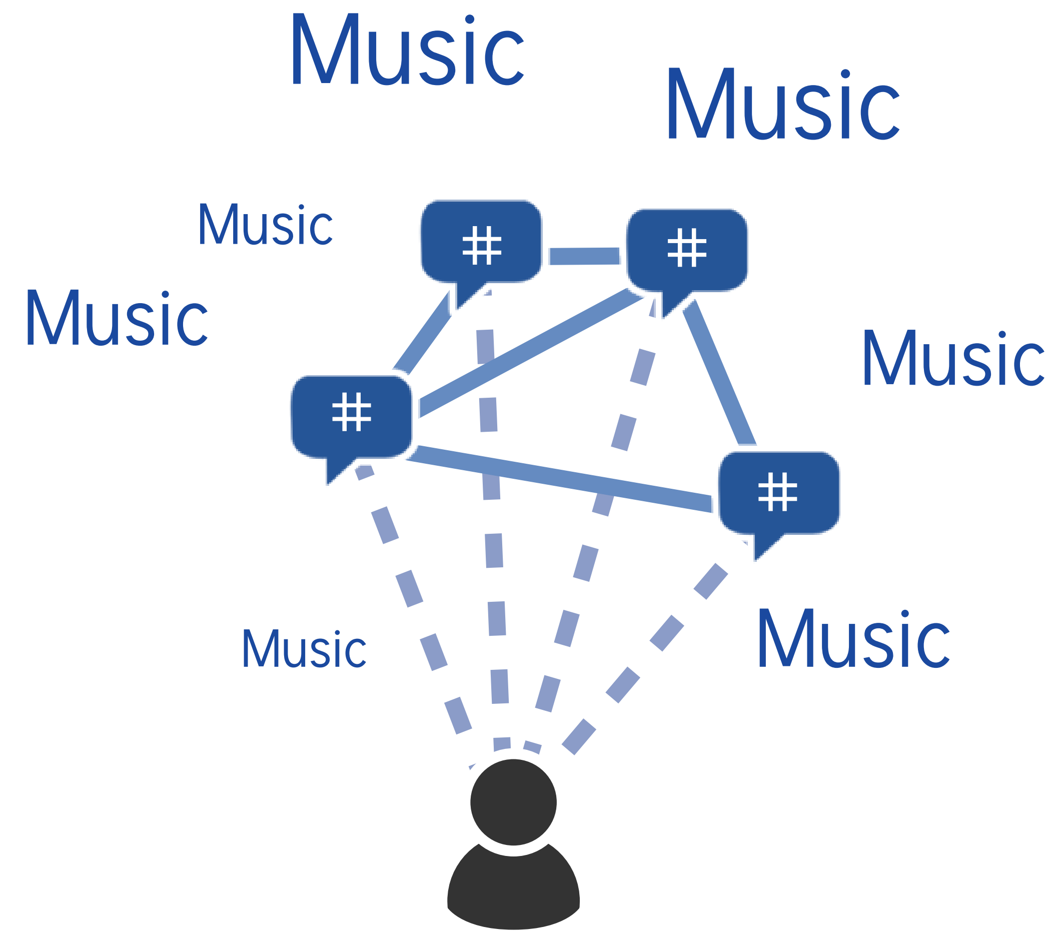


- Early adopters have diverse topical interests

- Co-occur with other tags about diverse topics



Hypothesis 1



Hypothesis 2

POPULAR USERS

Direct measure of social influence

How many times a user is retweeted

	Coefficient	SE
(Intercept)	20.9	0.5
Num. followers (fol)†	193.0 ***	0.5
Num. tweets (twt)†	51.1 ***	0.5
Content interestingness (β)†	3.9 ***	0.5
Diversity of interests (H_1)†	-9.1 ***	0.5

† Variables are normalized by Z-score. *** $p < 0.001$

Staying **Focused** helps increase social impact

(Weng et al. 2014)

POPULAR USERS

Direct measure of social influence

How many times a user is retweeted

	Coefficient	SE
(Intercept)	20.9	0.5
Num. followers $(fol)^{\dagger}$	have a big audience group	
Num. tweets $(twt)^{\dagger}$	be productive	
Content interestingness $(\beta)^{\dagger}$	create interesting content	
Diversity of interests $(H_1)^{\dagger}$	stay focused	

† Variables are normalized by Z-score. *** $p < 0.001$

No simple recipe of success

(Weng et al. 2014)

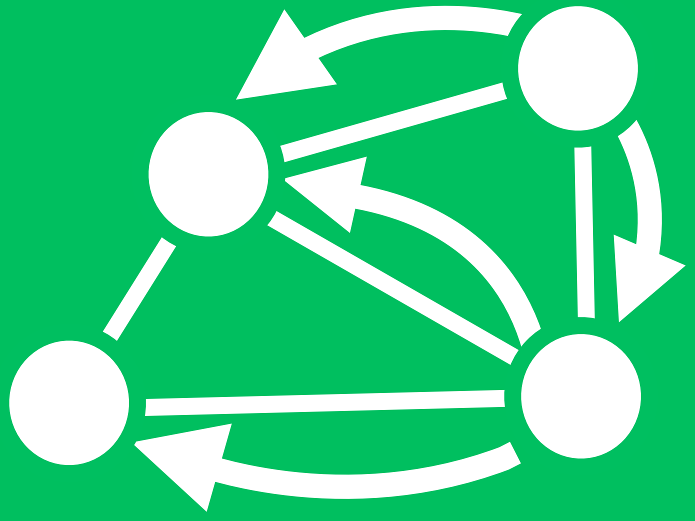
Not me!

(Weng et al. 2010)

(Weng et al. 2014)

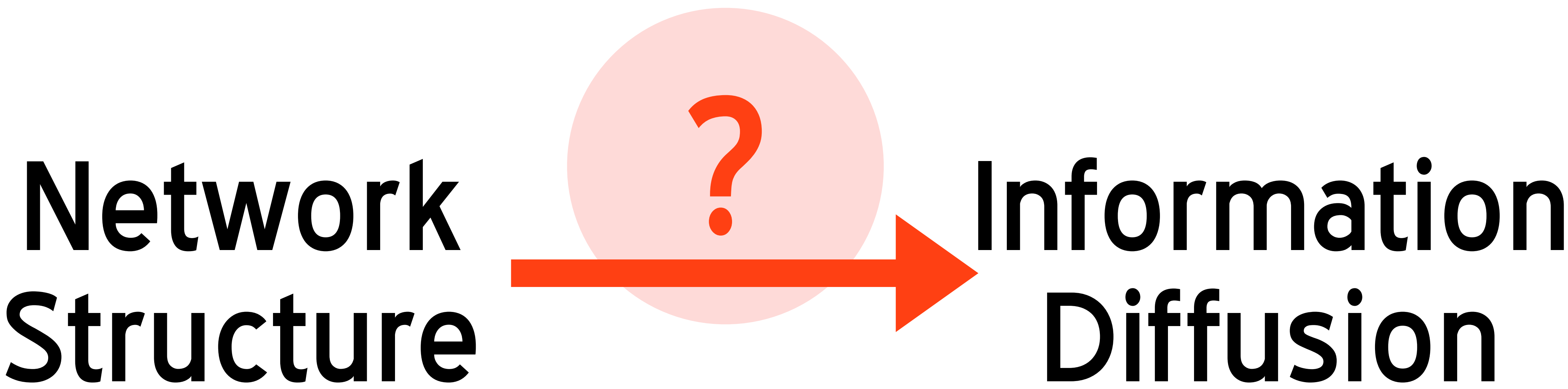
WHAT DO WE LEARN?

- Detect topics by finding communities
- Broad topical interests help content go viral
- Focused interests make users impactful



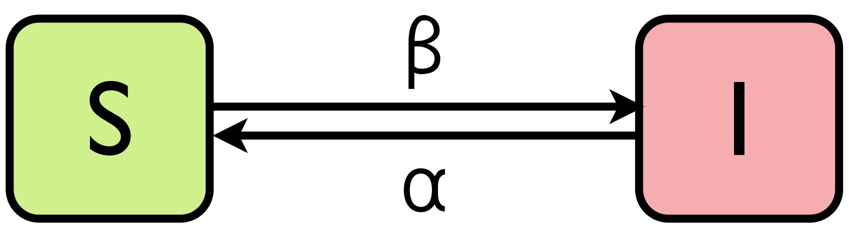
DIFFUSION on NETWORK

Part Three

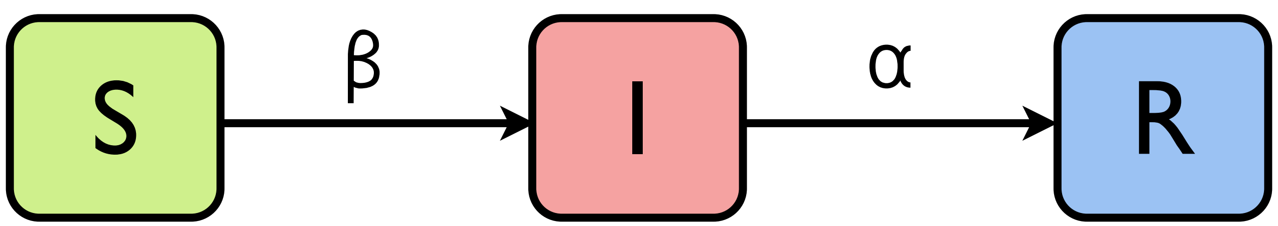

Network
Structure
?
Information
Diffusion

INFORMATION DIFFUSION

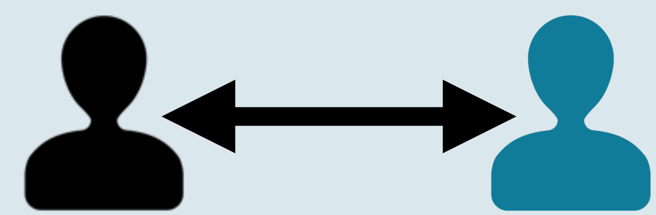
- The SIS Model (Bailey, 1975)



- The SIR Model (Anderson & May, 1992)

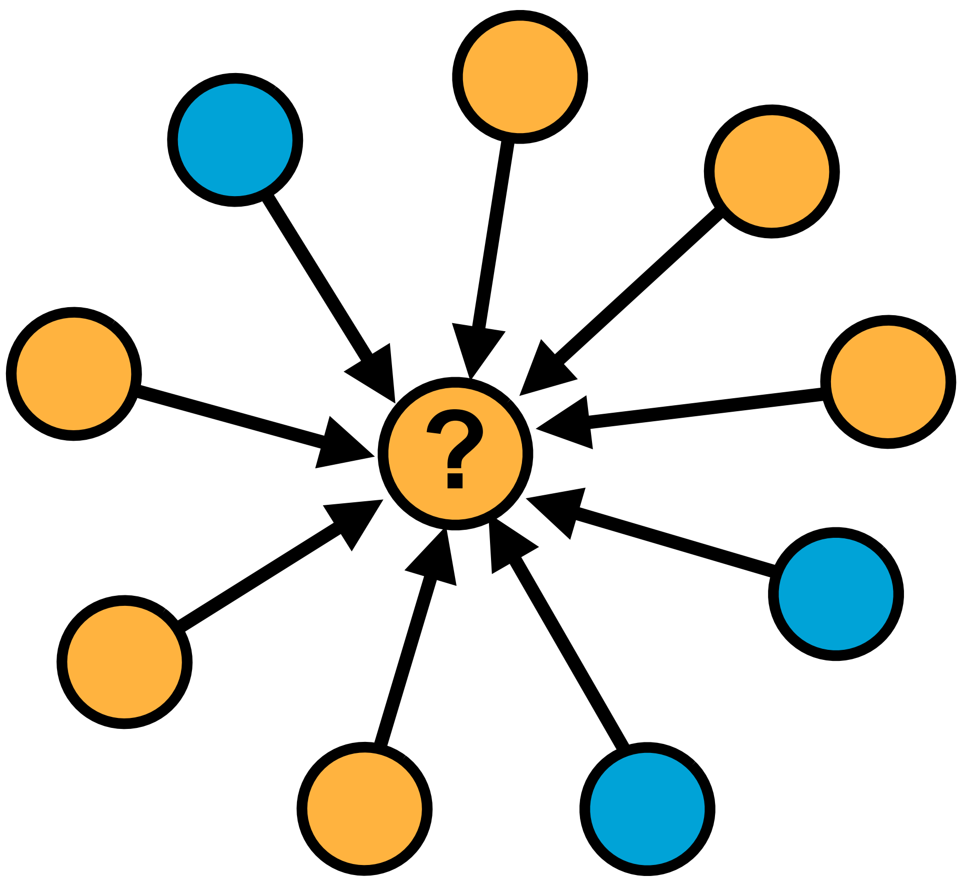


EPIDEMIC MODELS

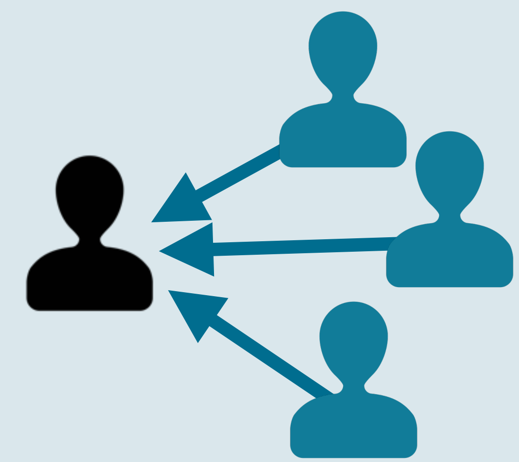


Diseases
Simple contagion

INFORMATION DIFFUSION



THRESHOLD MODEL
(Granovetter, 1978)

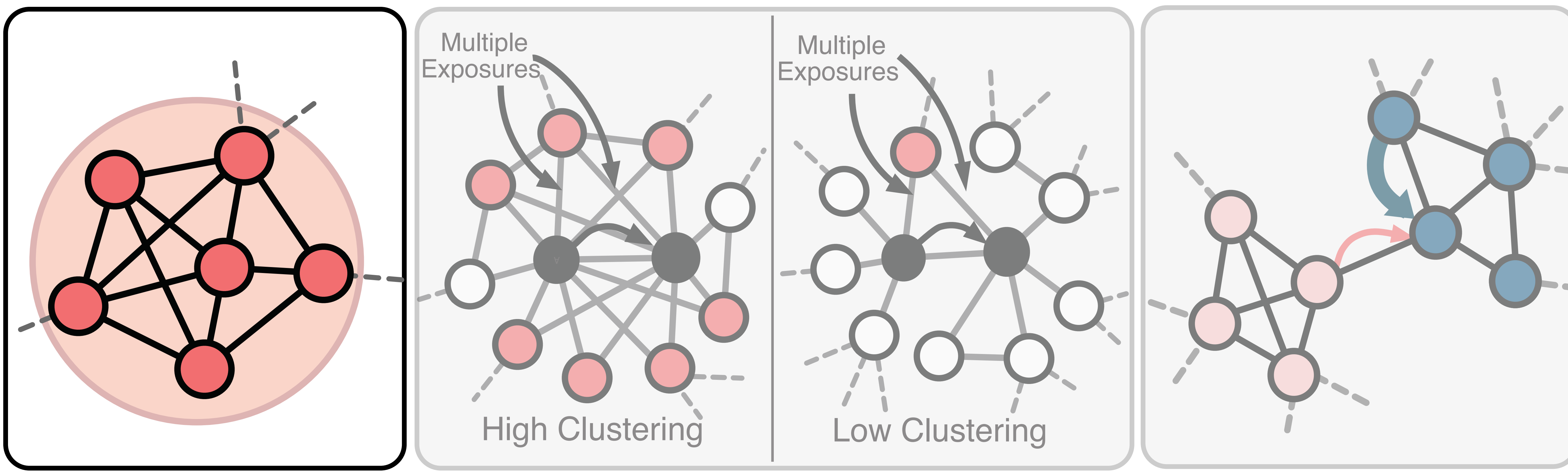


Ideas or behavior:
Complex contagion

- DBLP (Backstrom et al., 2006)
- Twitter (Huberman et al., 2008; Romero et al., 2011)
- Wikipedia (Cosley et al., 2010)
- Facebook (Ugander et al., 2012)

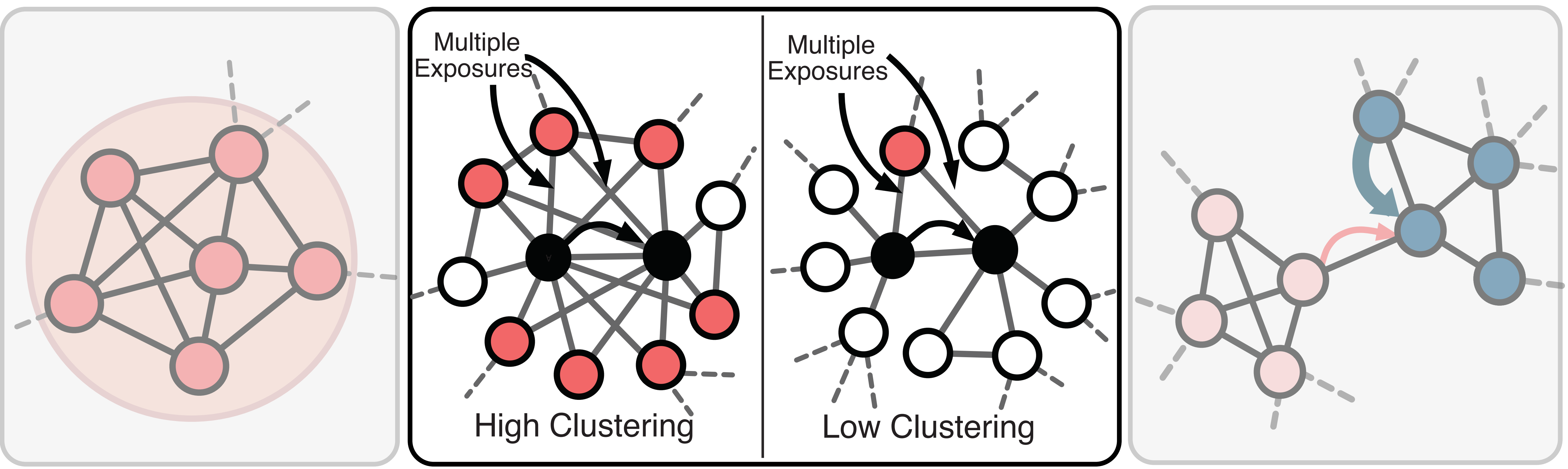
COMMUNITY TRAPPING EFFECT

Structural
Trapping



COMMUNITY TRAPPING EFFECT

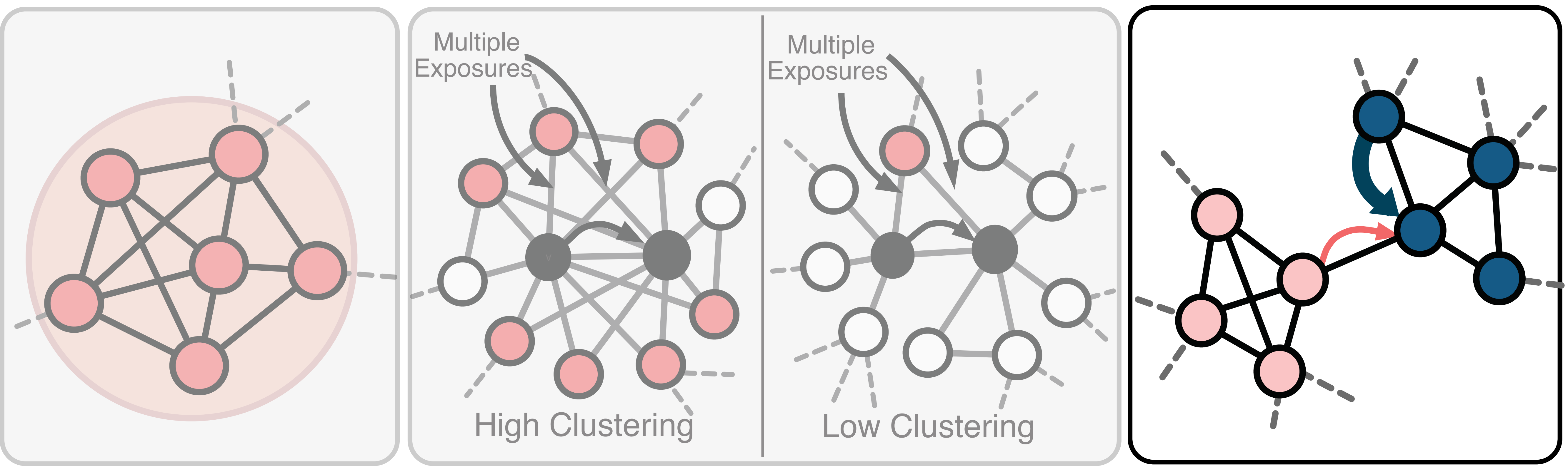
Social
Reinforcement
(Centola, 2010)



COMMUNITY TRAPPING EFFECT

Homophily

(McPherson et al.,2001)



Is communication and meme adoption concentrated inside communities?

NULL MODELS

	Community trapping effects		
	Network	Reinforcement	Homophily
M1: Random distribution			
M2: Random diffusion	√		
M3: Social reinforcement	√	√	
M4: Homophily	√		√

NULL MODELS

	Community trapping effects		
	Network	Reinforcement	Homophily
M1: Random distribution			
M2: Random diffusion	√		Simple contagion
M3: Social reinforcement	√	√	
M4: Homophily	√		√

NULL MODELS

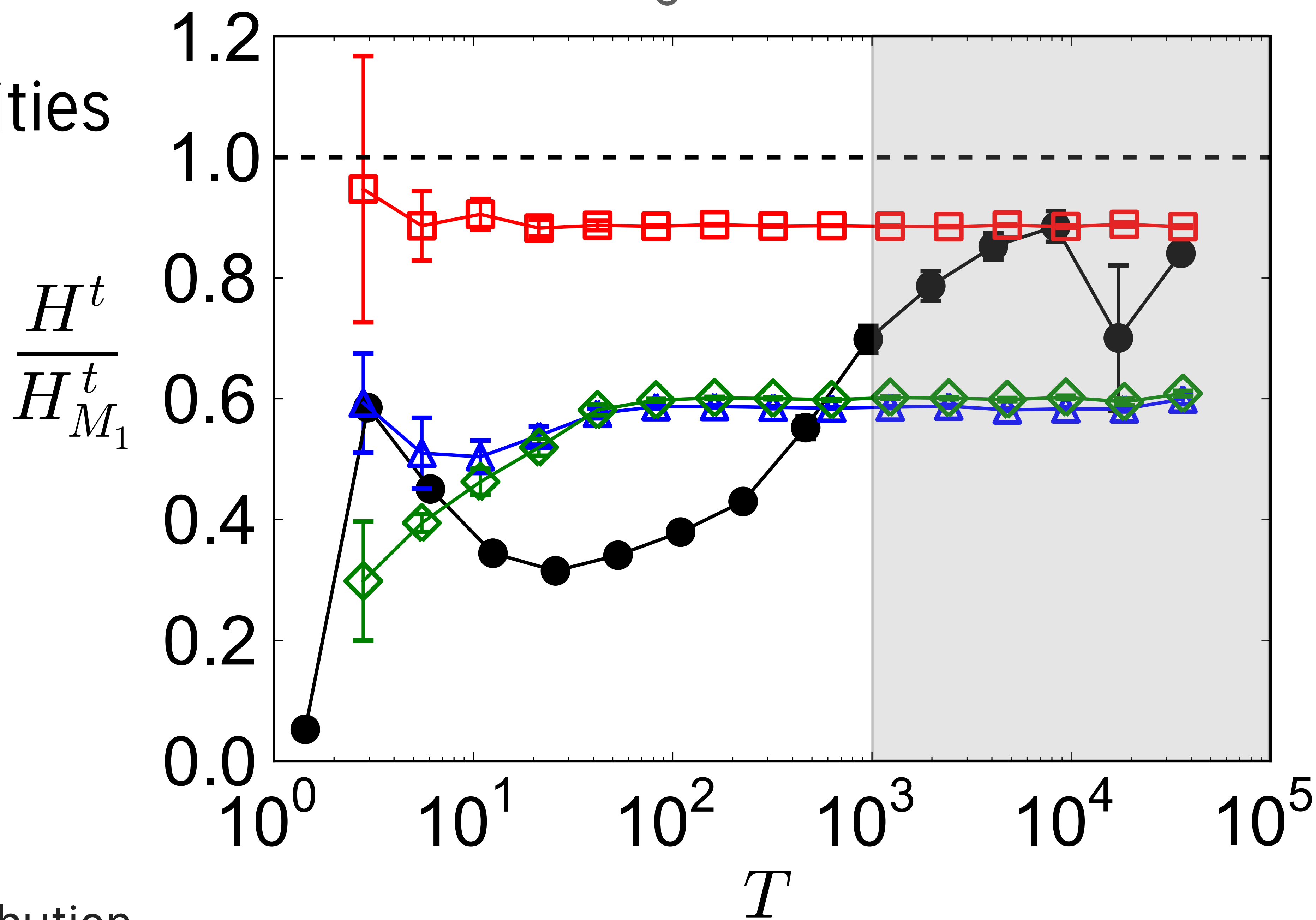
	Community trapping effects		
	Network	Reinforcement	Homophily
M1: Random distribution			
M2: Random diffusion	√		
M3: Social reinforcement	√	√	
M4: Homophily	√		√

Complex contagion

RELATIVE USAGE ENTROPY

(Weng et al. 2013)

Entropy of # tweets distributed in different communities



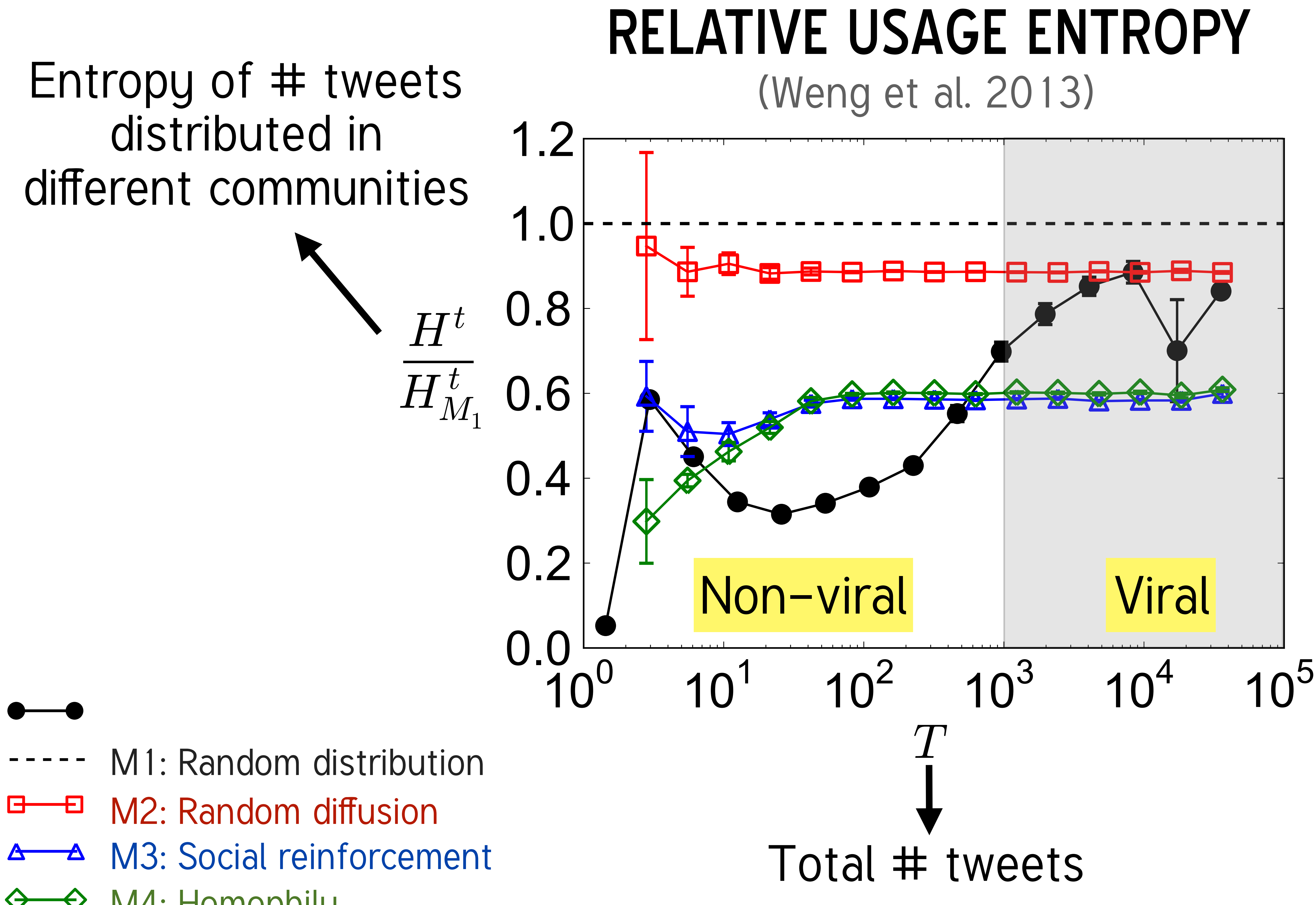
Total # tweets

M1: Random distribution

M2: Random diffusion

M3: Social reinforcement

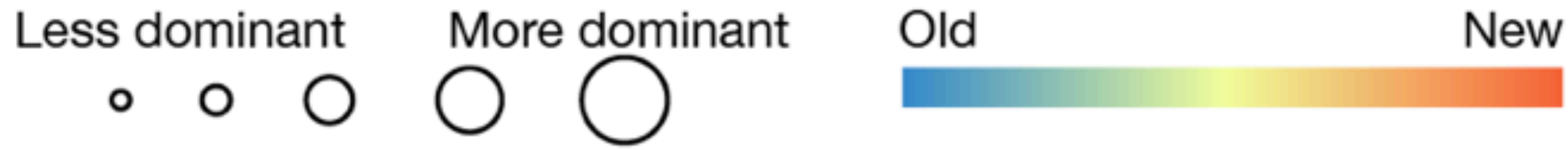
M4: Homophily

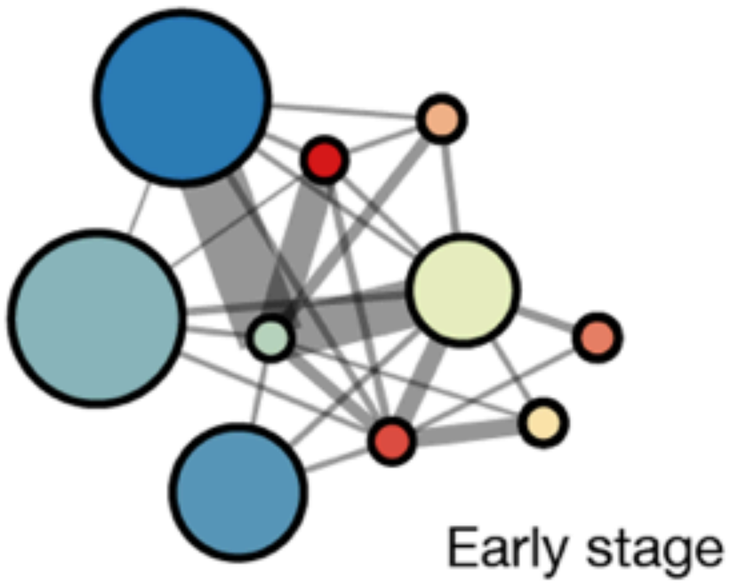


RELATIVE USAGE ENTROPY
(Weng et al. 2013)
Entropy of # tweets distributed in different communities
$\frac{H^t}{H^t_{M_1}}$
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10^0
10^1
10^2
10^3
10^4
10^5
T
Total # tweets
Non-viral
Viral
M1: Random distribution
M2: Random diffusion
M3: Social reinforcement
M4: Homophily

Viral memes are less trapped by communities, spreading like diseases

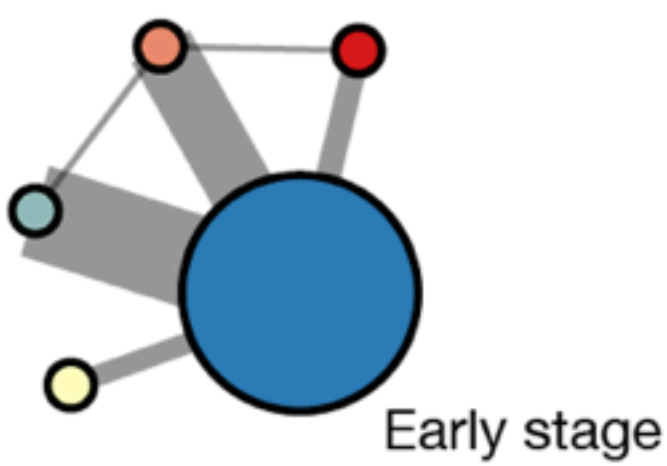
Can we predict the future meme virality by qualifying concentration across communities?





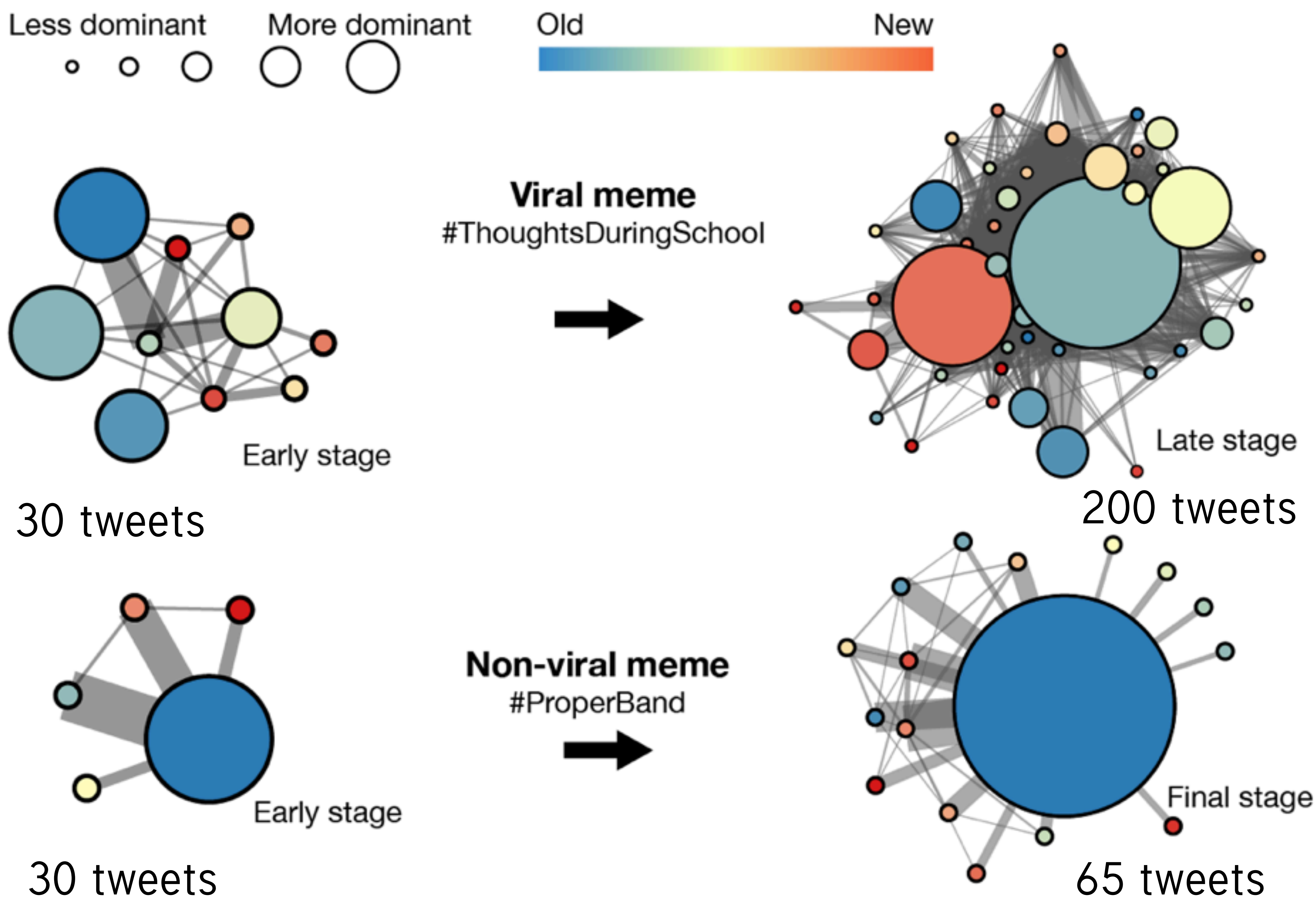
#ThoughtsDuringSchool

30 tweets



#ProperBand

30 tweets


Less dominant
More dominant
Old
New
Viral meme
#ThoughtsDuringSchool
Early stage
Late stage
30 tweets
200 tweets
Non-viral meme
#ProperBand
Early stage
Final stage
30 tweets
65 tweets

VIRALITY PREDICTION

Binary classification
Predict whether a meme is viral (>1000 tweets)

ΔF_1

(Weng et al. 2013)

VIRALITY PREDICTION

1 Community-blind features
- # Early adopters
- Size of infection frontier

Binary classification
Predict whether a meme is viral (>1000 tweets)

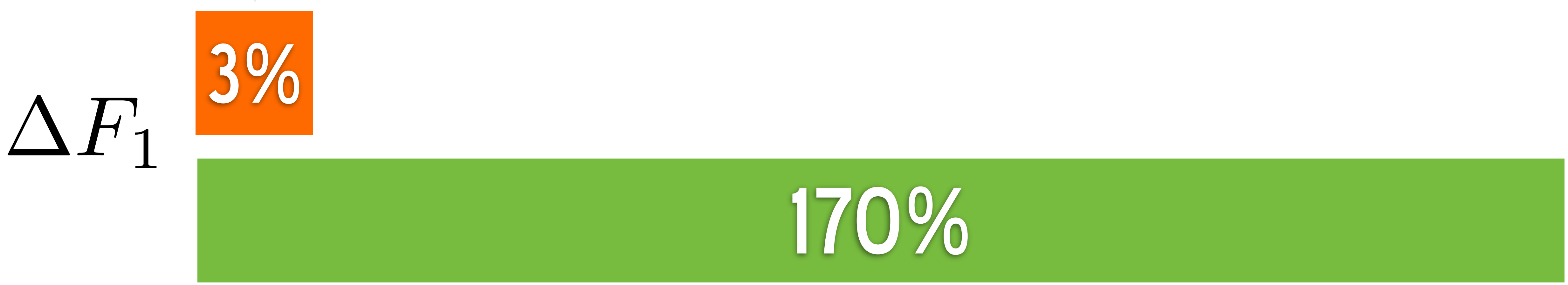
ΔF_1 3%

(Weng et al. 2013)

VIRALITY PREDICTION

1 Community-blind features
- # Early adopters
- Size of infection frontier

Binary classification
Predict whether a meme is viral (>1000 tweets)



2 Community-based features
- # Infected communities
- Entropy
- Frac. intra-community RT/@

(Weng et al. 2013)

A STRONGER CLASSIFIER

More features for *multi-label classification*:

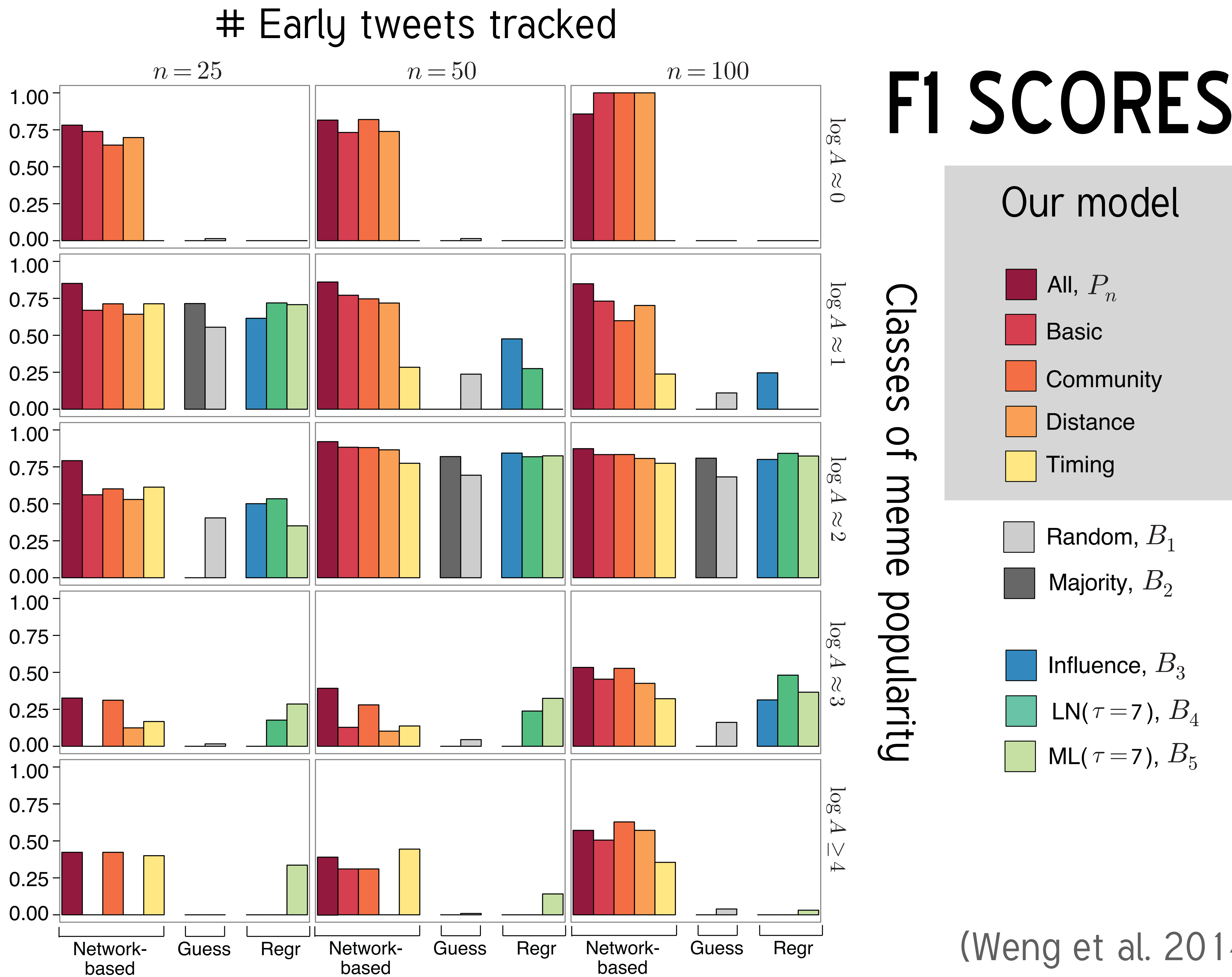
1) Community-based features
2) Network topology
3) Time sequence

Stronger baselines:

1) Early popularity
2) Time sequence of early popularity
3) PageRank & # followers of early adopters

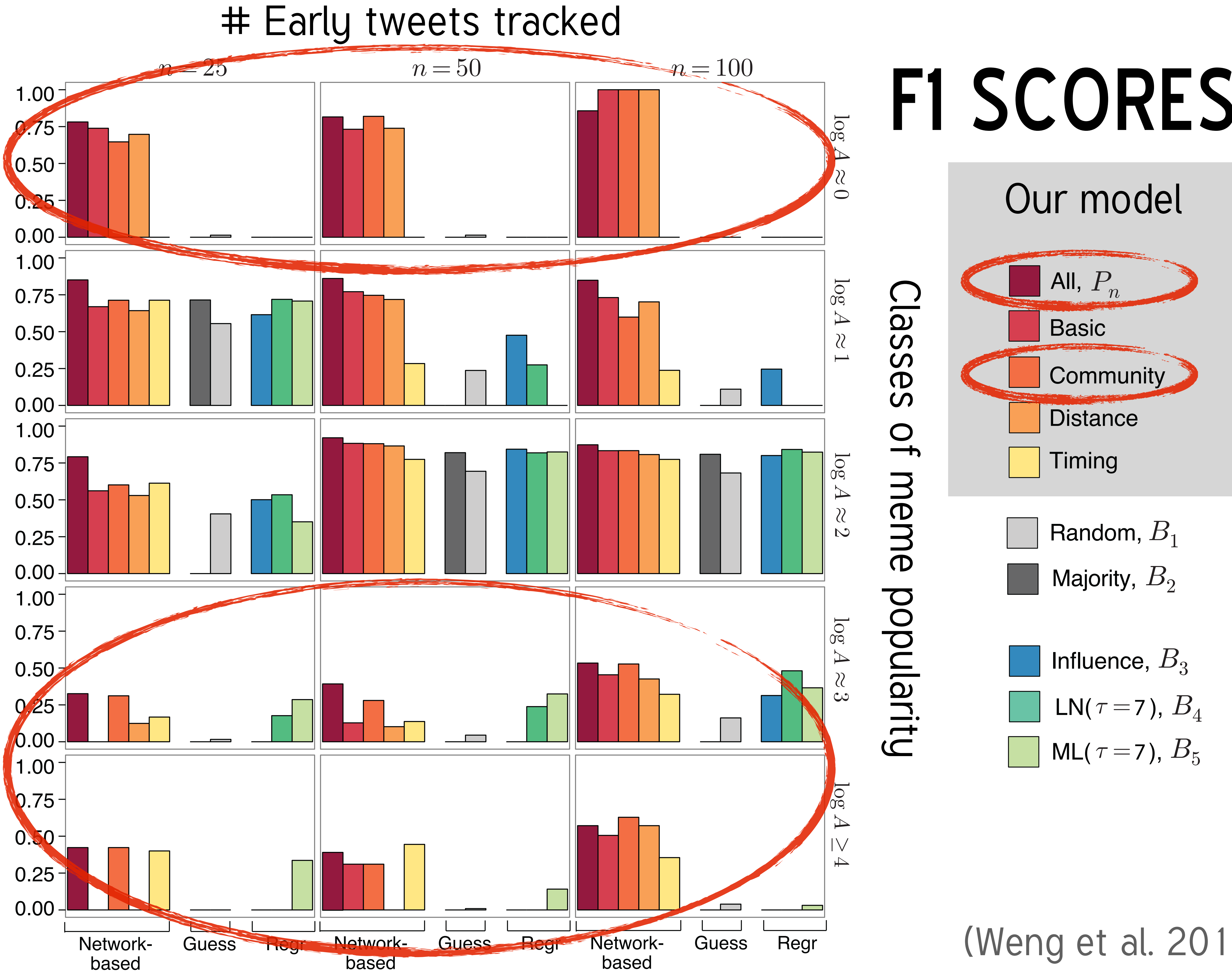
(Weng et al. 2014)

F1 SCORES



(Weng et al. 2014)

F1 SCORES



(Weng et al. 2014)


?
Network Structure
Information Diffusion

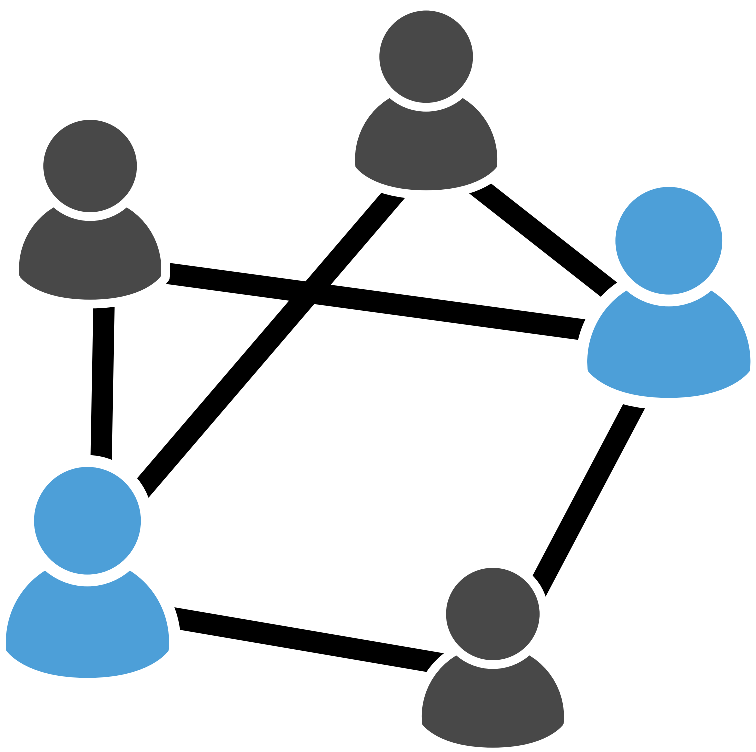
NETWORK GROWTH

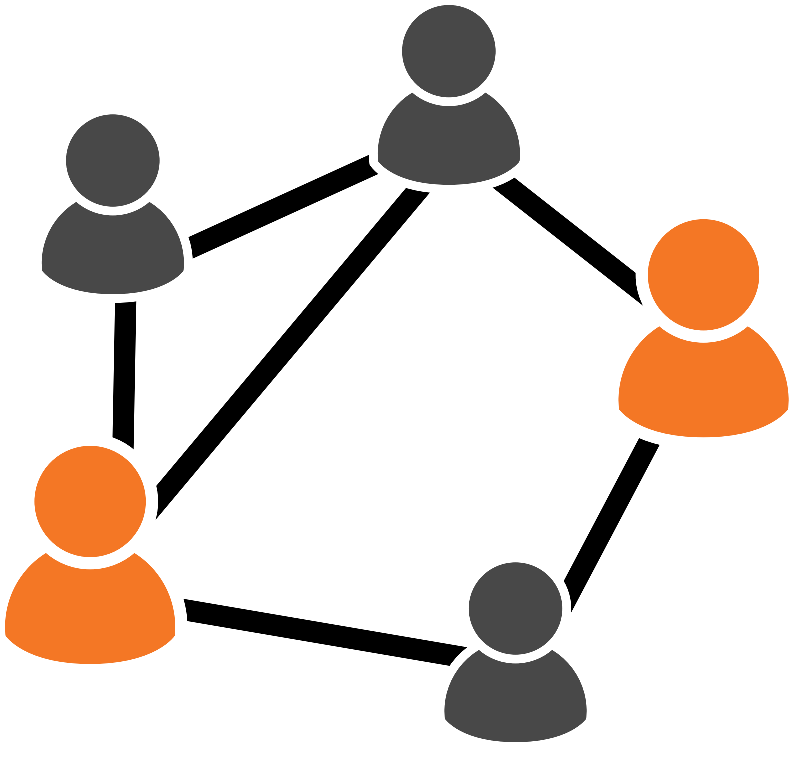
- **Random** network (Erdos & Renyi, 1959)
- **Small world** network (Watts and Strogatz, 1998)
- **Scale-free** network
  - Barabasi-Albert model (Barabasi & Albert, 1999)
  - Copying model (Kumar et al., 2000)
  - Ranking model (Fortunato et al., 2006)
- **Social** network features
  - Homophily (Holme and Newman, 2006)
  - Triadic closure (Leskovec et al., 2008)
  - Community structure (Clauset et al., 2008)

NETWORK GROWTH

- **Random** network (Erdos & Renyi, 1959)
- **Small world** network (Watts and Strogatz, 1998)
- **Scale-free** network
  - Barabasi-Albert model (Barabasi & Albert, 1999)
  - Copying model (Kumar et al., 2000)
  - Ranking model (Fortunato et al., 2006)
- **Social** network features
  - Homophily (Holme and Newman, 2006)
  - Triadic closure (Leskovec et al., 2008)
  - Community structure (Clauset et al., 2008)
  - **Information diffusion!** (Weng et al.,2013)

NEW LINKAGE

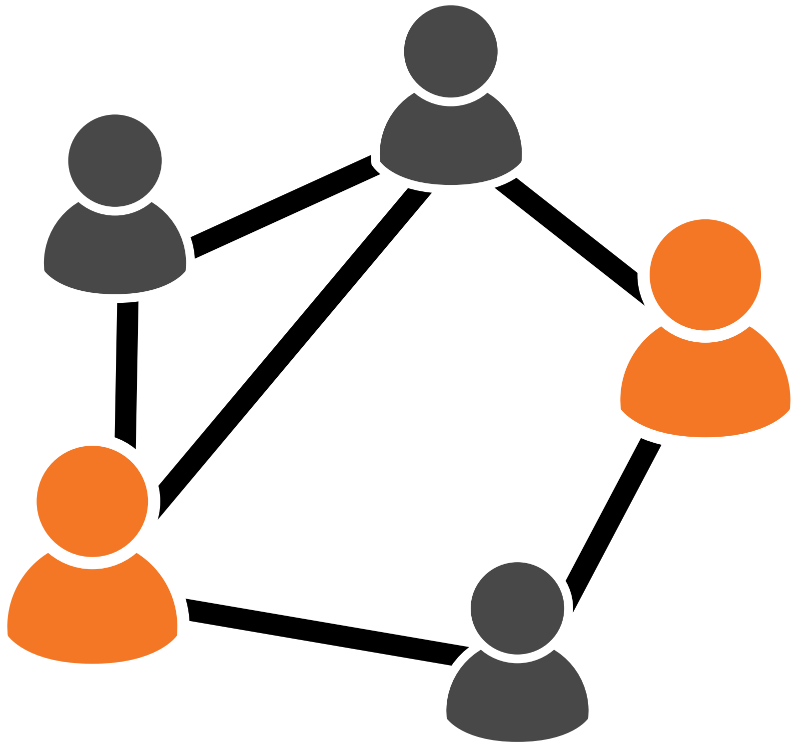




NEW LINKAGE

Social circle

- Many common friends
- **Triadic closure**



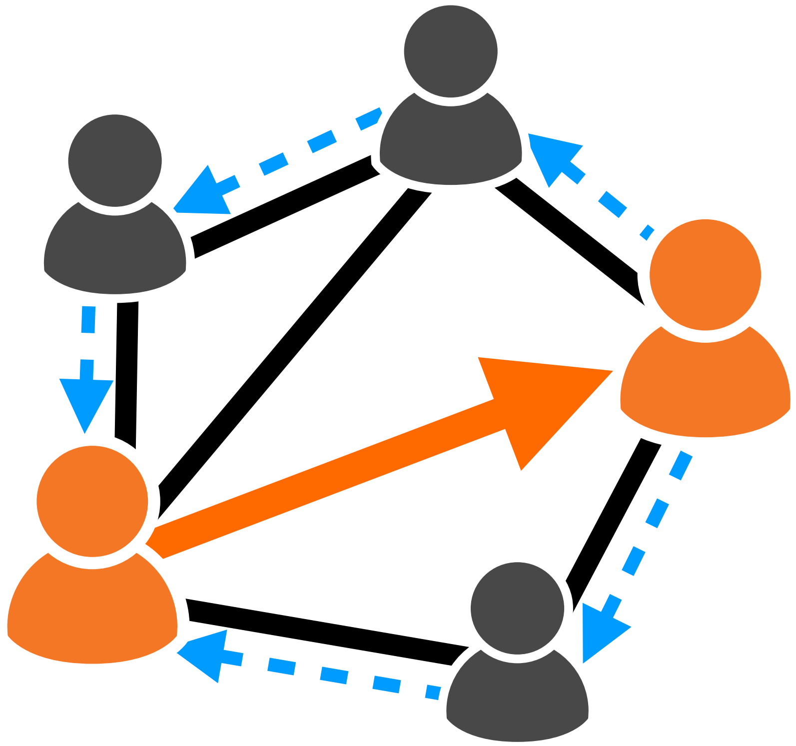
NEW LINKAGE

Social circle

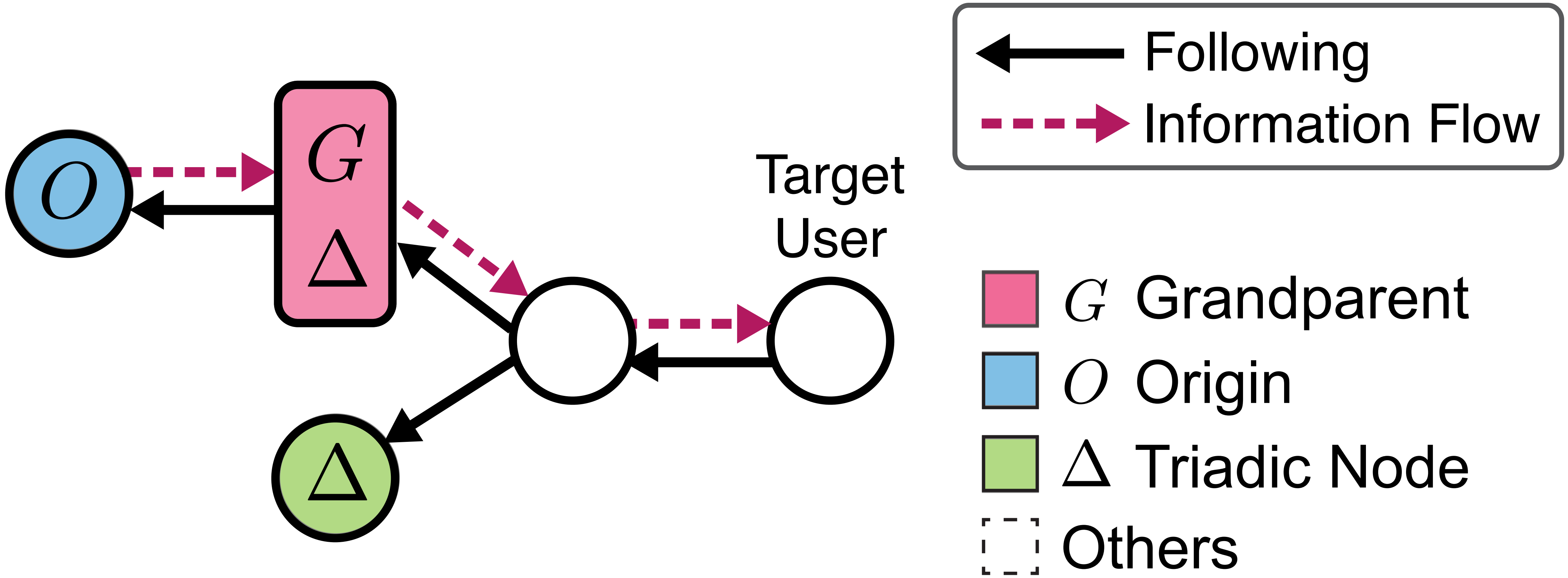
- Many common friends
- **Triadic closure**

Information

- Obtain valuable information
- **Efficient information flows**



YAHOO! meme



Rand Follow a random user

Δ Follow a triadic node

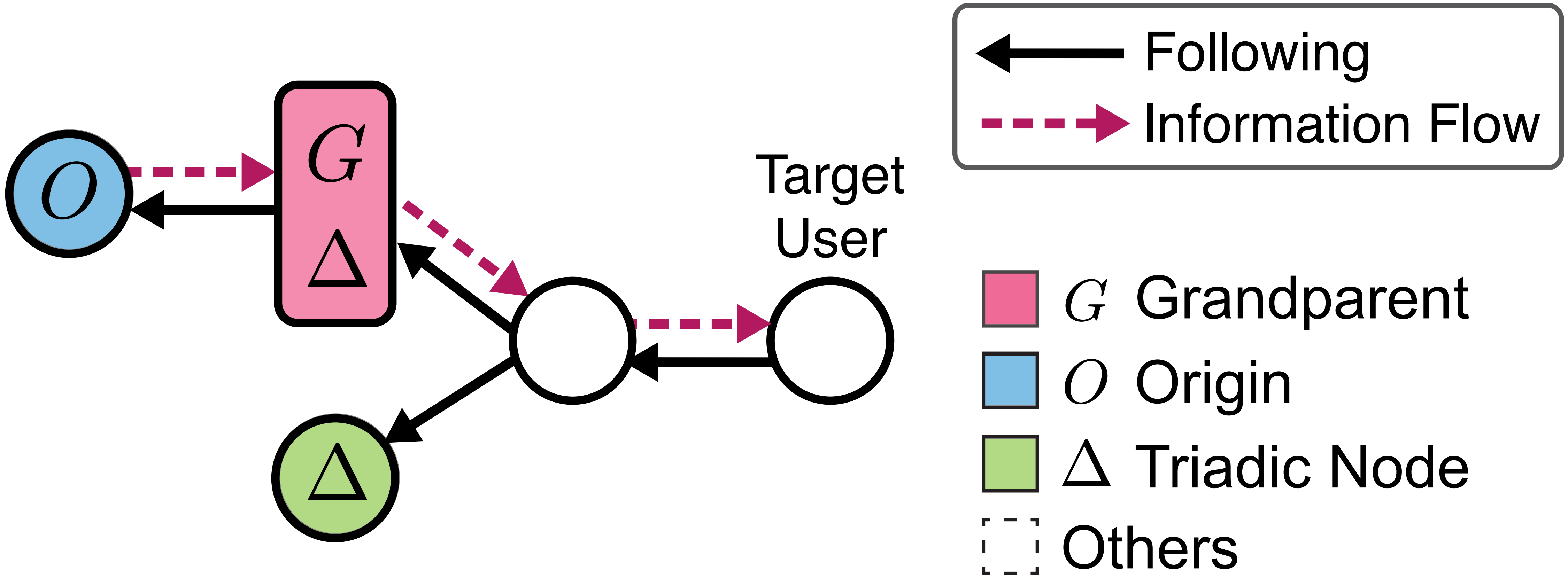
G Follow a grandfather

O Follow an origin

$G \cup O$ **Traffic shortcuts**

T

STRATEGIES



Rand	Follow a random user
Δ	Follow a triadic node
G	Follow a grandfather
O	Follow an origin
$G \cup O$	**Traffic shortcuts**

T

STRATEGIES

RULES OF EVOLUTION

Maximum Likelihood Estimation, MLE

Γ – Strategy for creating every link ℓ

Θ – Network configuration

$f(\ell|\Gamma,\Theta)$ – Likelihood of the target being followed by the creator according to a particular strategy

$$\prod f(\ell|\Gamma,\Theta) \quad \to \log \prod f(\ell|\Gamma,\Theta) = \sum \log f(\ell|\Gamma,\Theta)$$

RULES OF EVOLUTION

Maximum Likelihood Estimation, MLE

Γ – Strategy for creating every link ℓ

Θ – Network configuration

$f(\ell|\Gamma,\Theta)$ – Likelihood of the target being followed by the creator according to a particular strategy

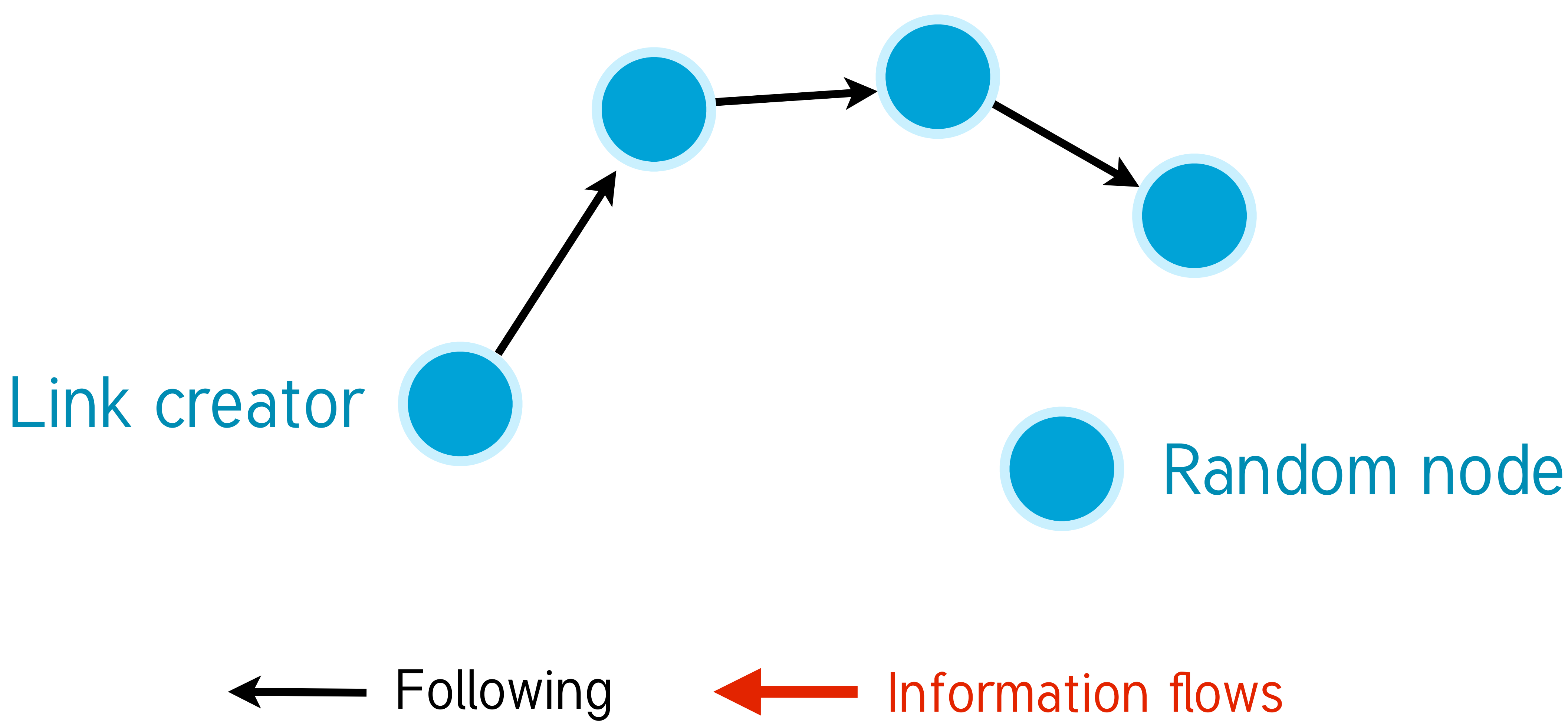
MAXIMIZE IT!

$$\prod f(\ell|\Gamma,\Theta) \quad \rightarrow \log \prod f(\ell|\Gamma,\Theta) = \sum \log f(\ell|\Gamma,\Theta)$$

SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

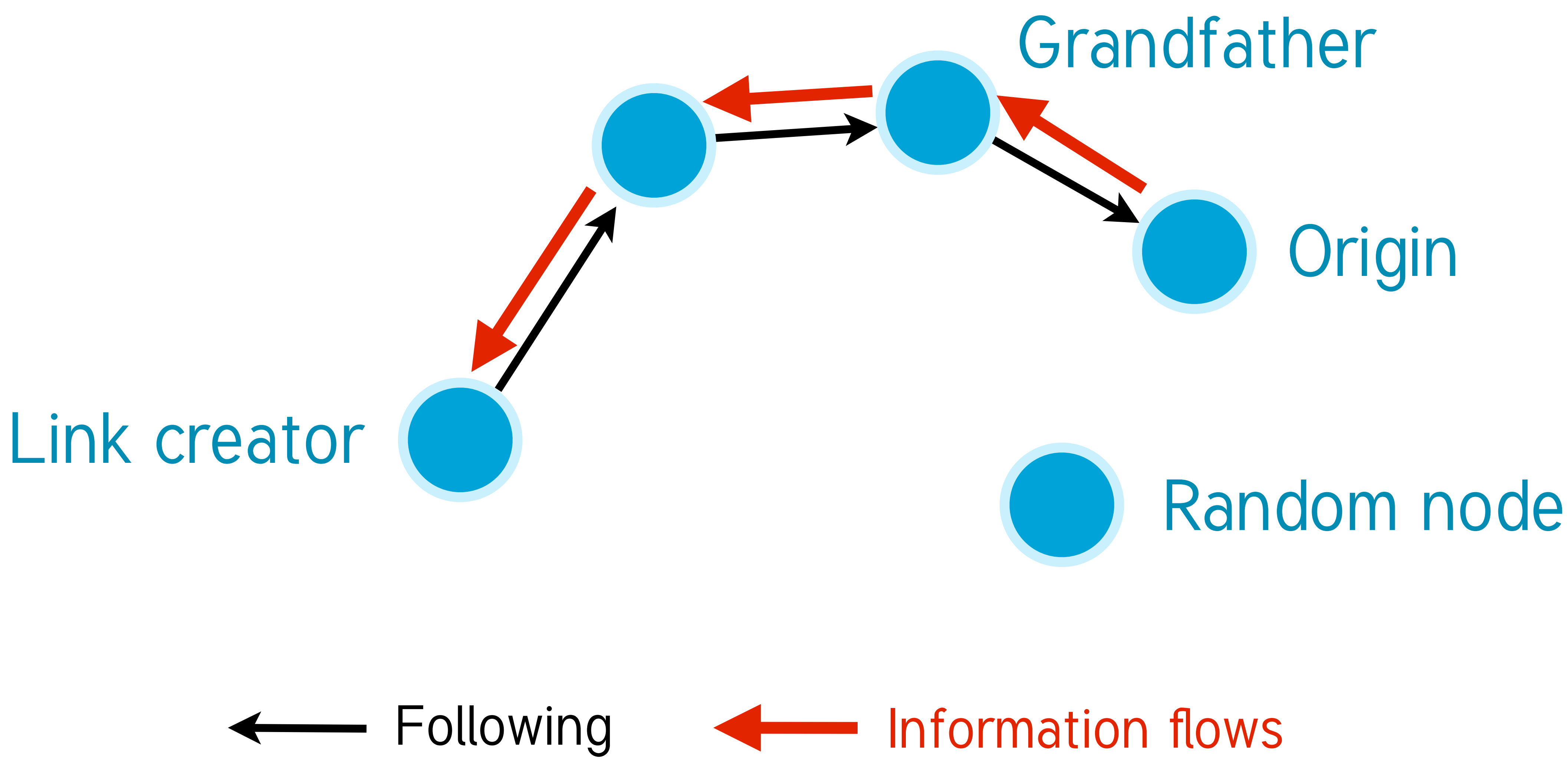
$$\mathcal{L}_G(p) =$$



SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

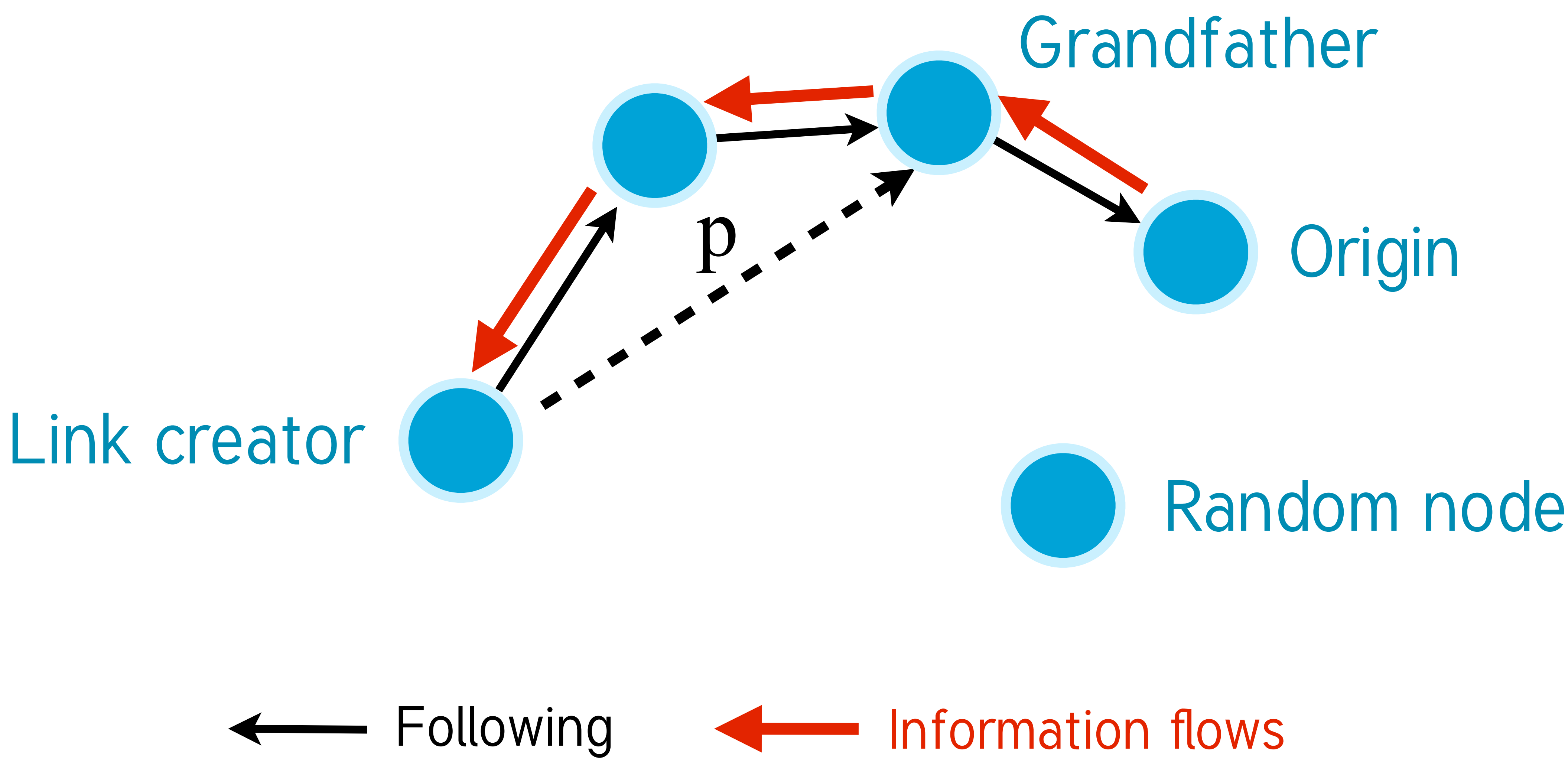
$$\mathcal{L}_G(p) =$$



SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

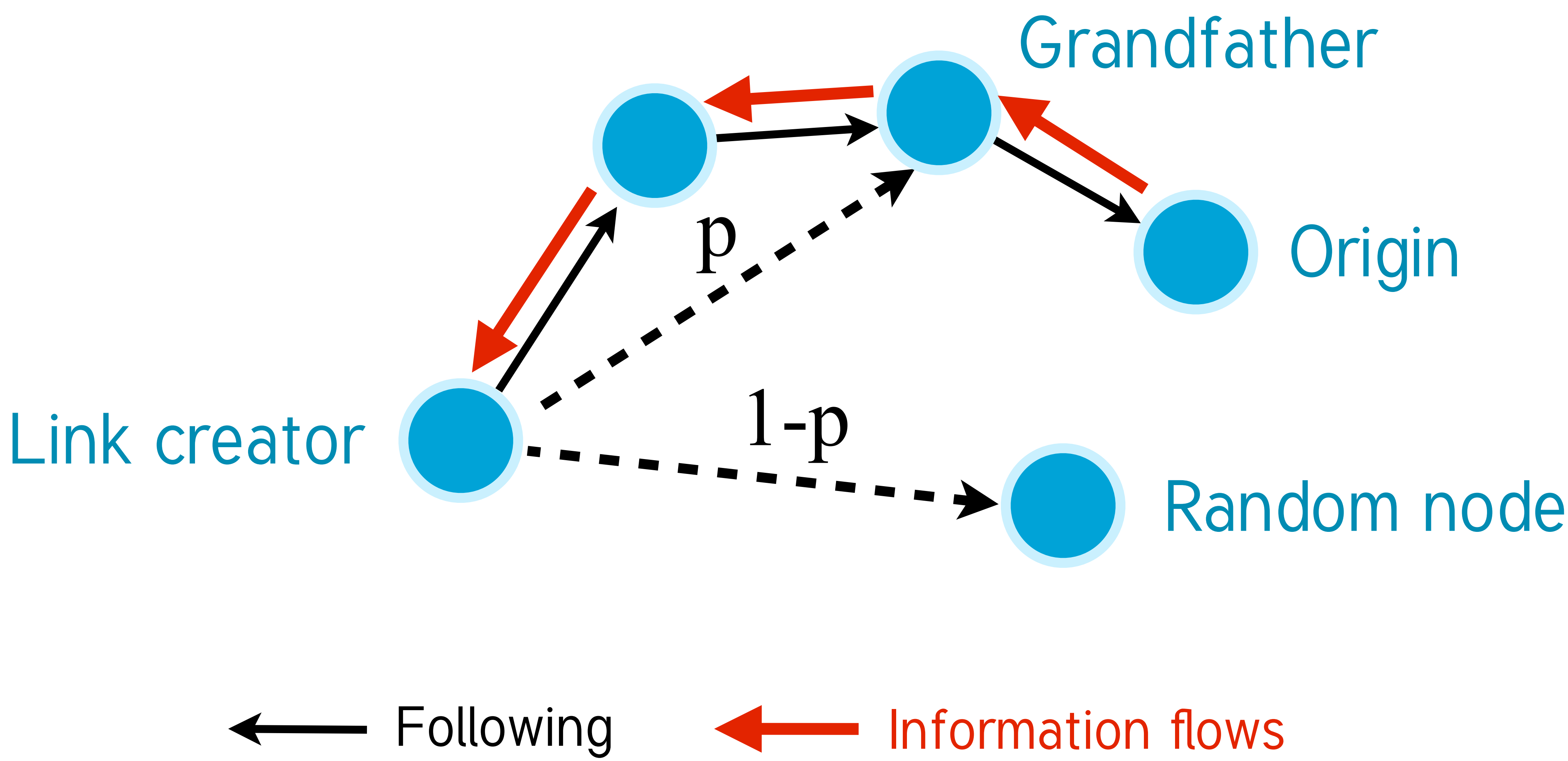
$$\mathcal{L}_G(p) = \qquad p f(\ell | G, \Theta)$$



SINGLE STRATEGY

Example: when $\Gamma = G + \text{Rand}$

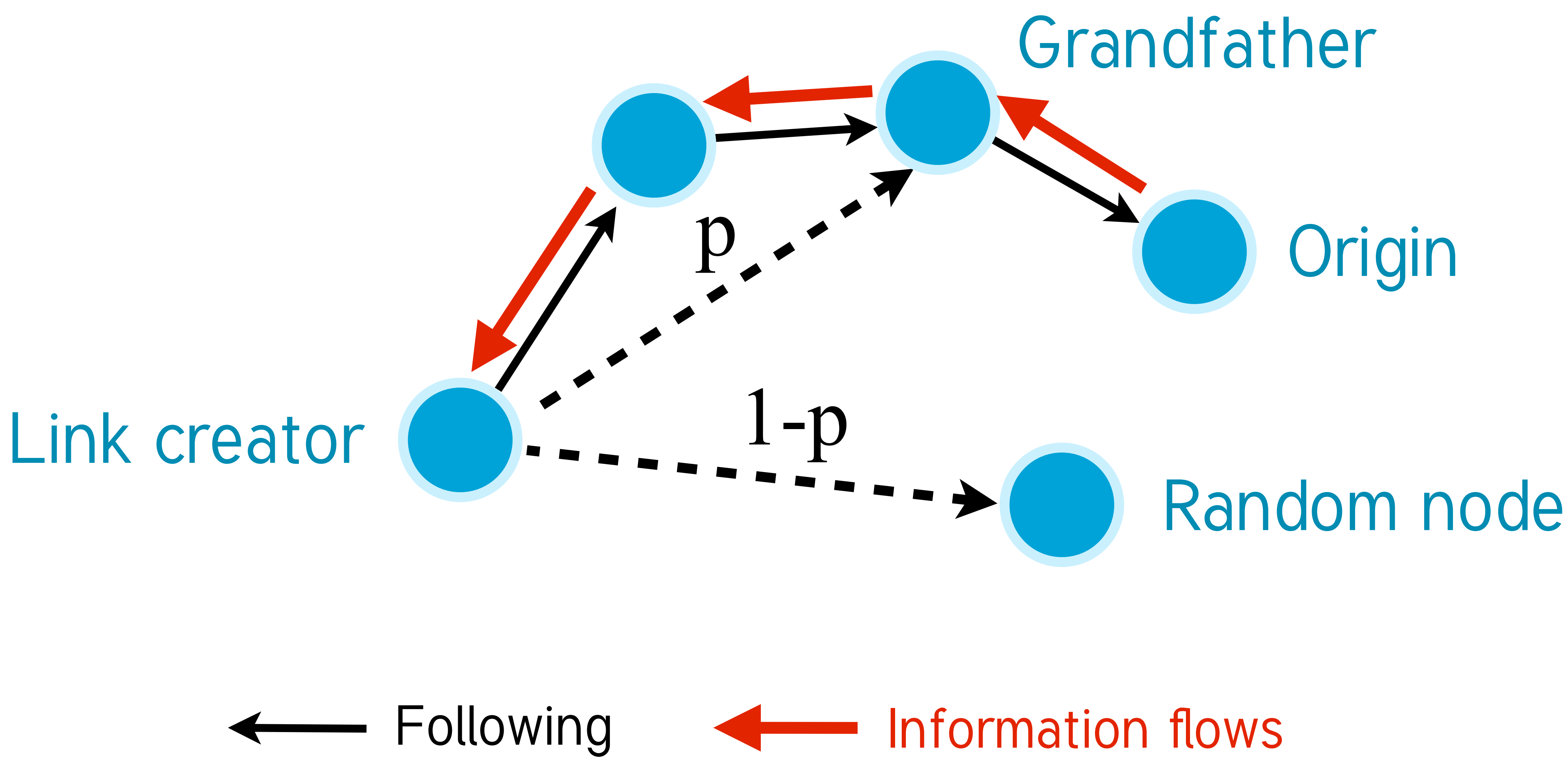
$$\mathcal{L}_G(p) = \quad p f(\ell|G,\Theta) + (1-p) f(\ell|\text{Rand},\Theta)$$



SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

$$\mathcal{L}_G(p) = \prod_{\ell=1}^{L} \left(p f(\ell | G, \Theta) + (1-p) f(\ell | \mathrm{Rand}, \Theta) \right)$$



SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

$$\mathcal{L}_G(p) = \prod_{\ell=1}^{L} (pf(\ell|G,\Theta) + (1-p)f(\ell|\mathrm{Rand},\Theta))$$

$$= \prod_{\ell=1}^{L} \left(p\frac{\mathbf{1}_G(\ell)}{N_G(\ell)} + (1-p)\frac{1}{\ell - k(\ell) - 1} \right)$$

$$= \prod_{\mathbf{1}_G(\ell)=1} \left(\frac{p}{N_G(\ell)} + \frac{1-p}{l - k(\ell) - 1} \right) \prod_{\mathbf{1}_G(\ell)=0} \frac{1-p}{\ell - k(\ell) - 1}$$

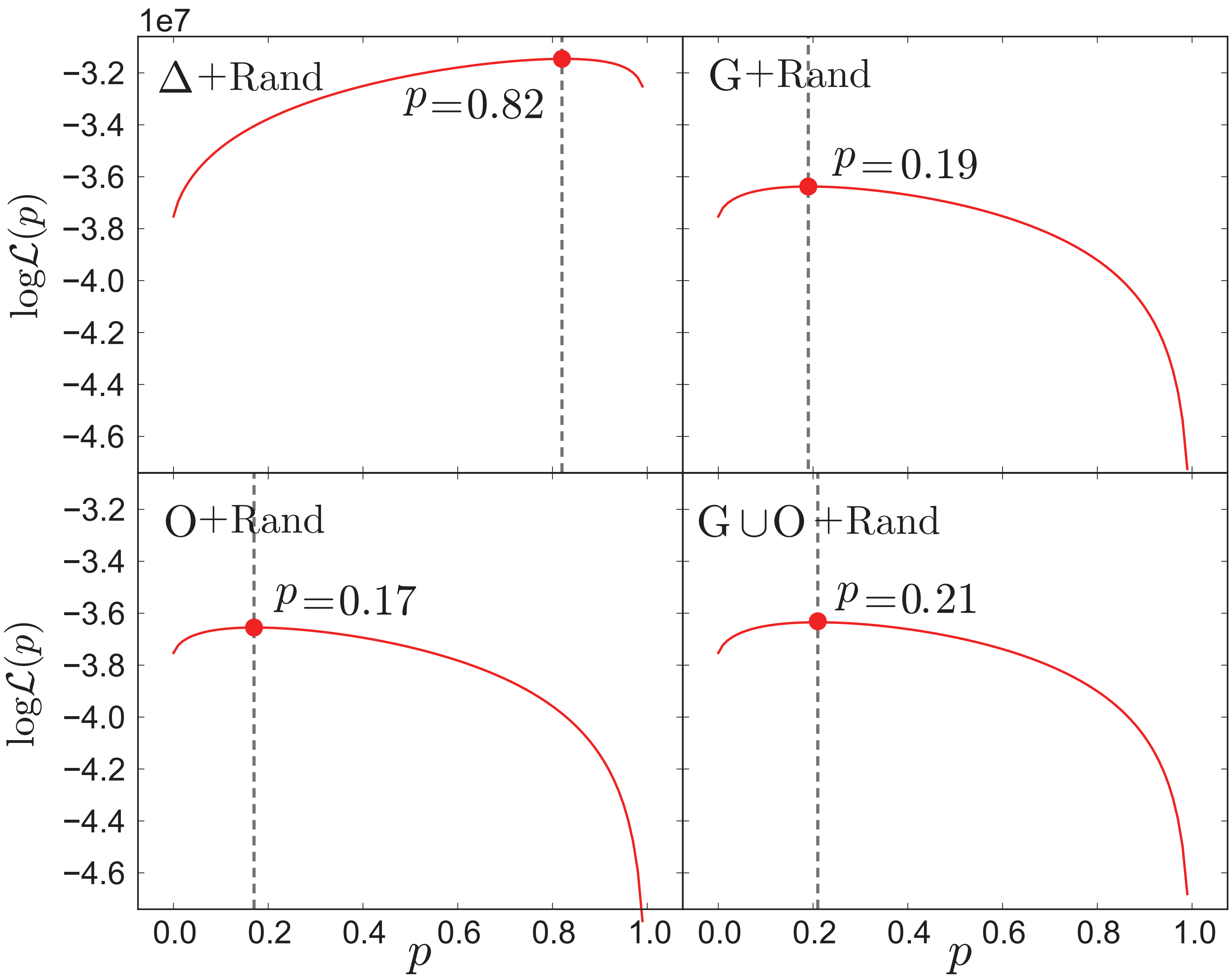
SINGLE STRATEGY

Example: when $\Gamma = G + \mathrm{Rand}$

$$\log \mathcal{L}_G(p) = \sum_{\mathbf{1}_G(\ell)=1} \ln\left(\frac{p}{N_G(\ell)} + \frac{1-p}{\ell - k(\ell) - 1}\right) + \sum_{\mathbf{1}_G(\ell)=0} \ln \frac{1-p}{\ell - k(\ell) - 1}$$

Find the best $p \in (0,1)$ that **MAXIMIZES** $\log \mathcal{L}_G(p)$

SINGLE STRATEGY

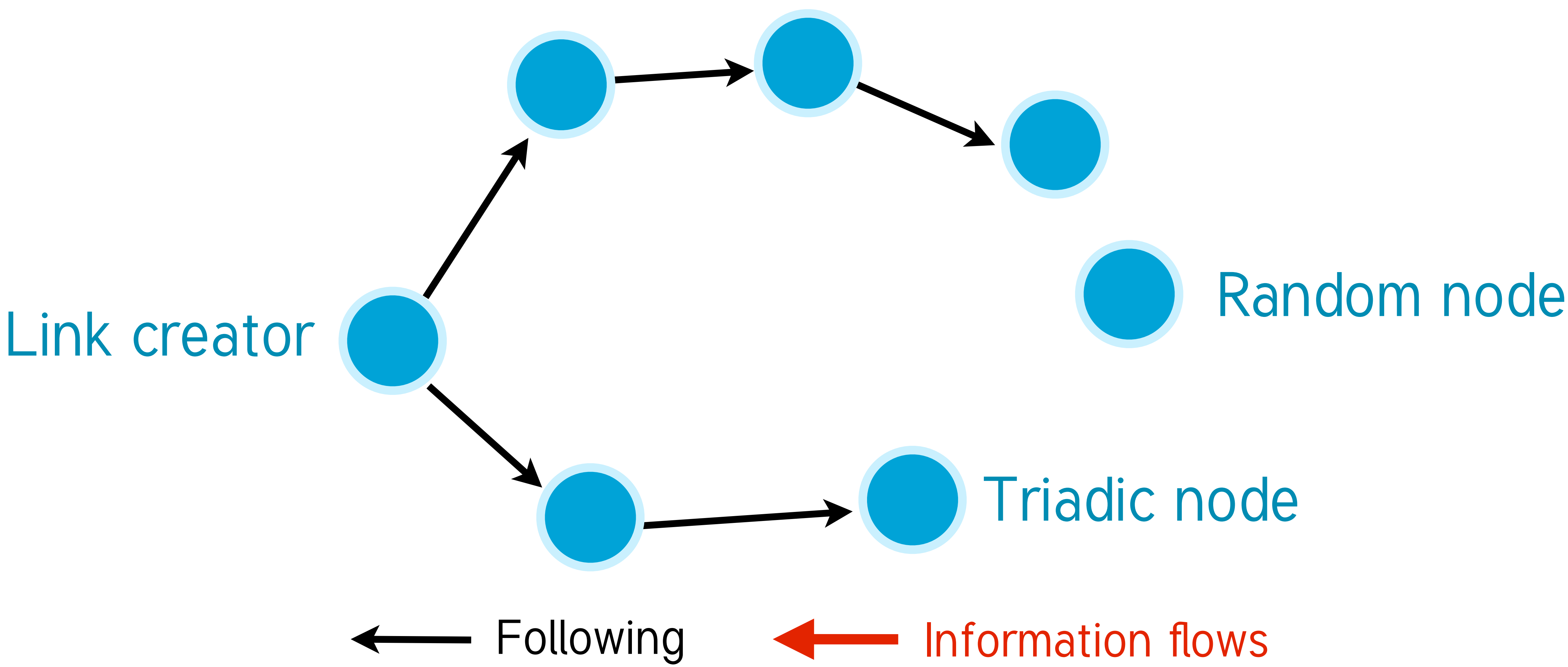


(Weng et al. 2013)

COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

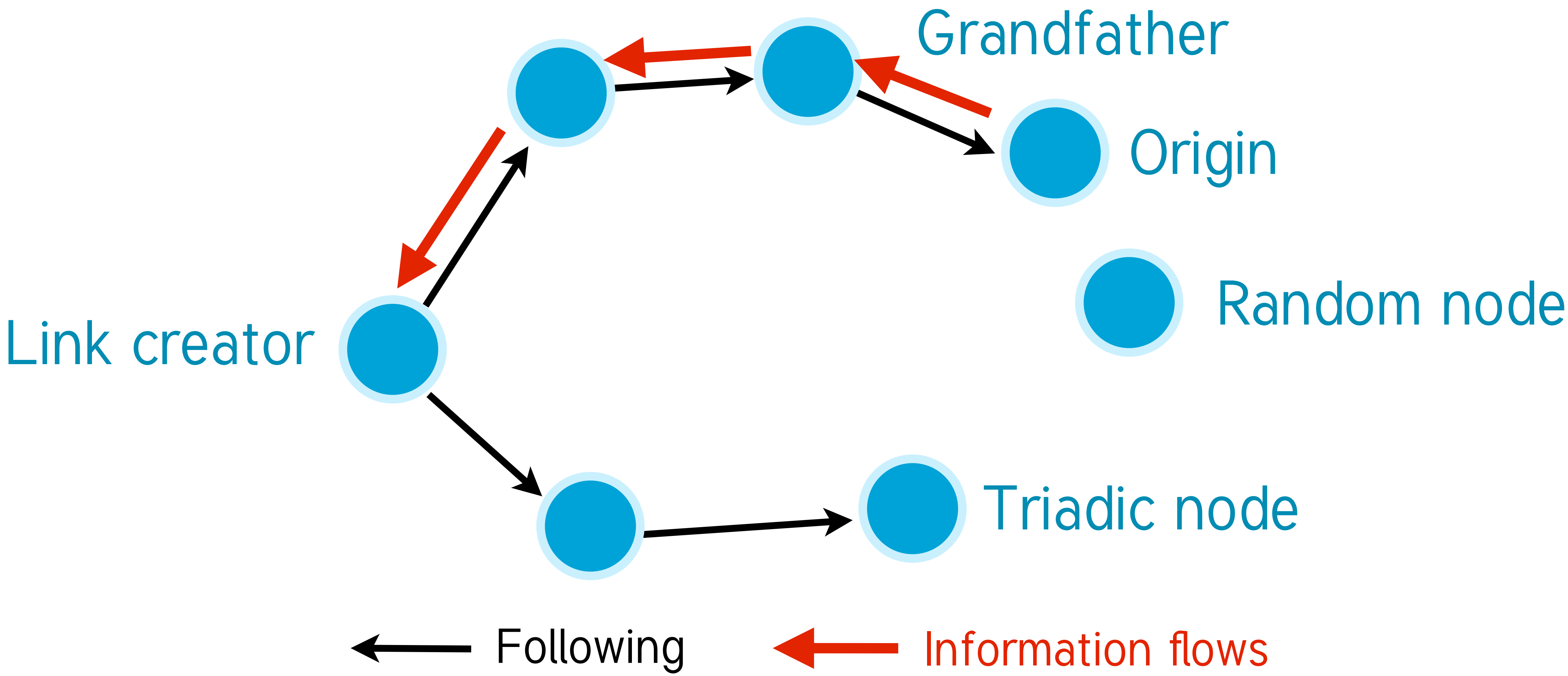
$$\mathcal{L}_{G+\Delta}(p_1, p_2) =$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

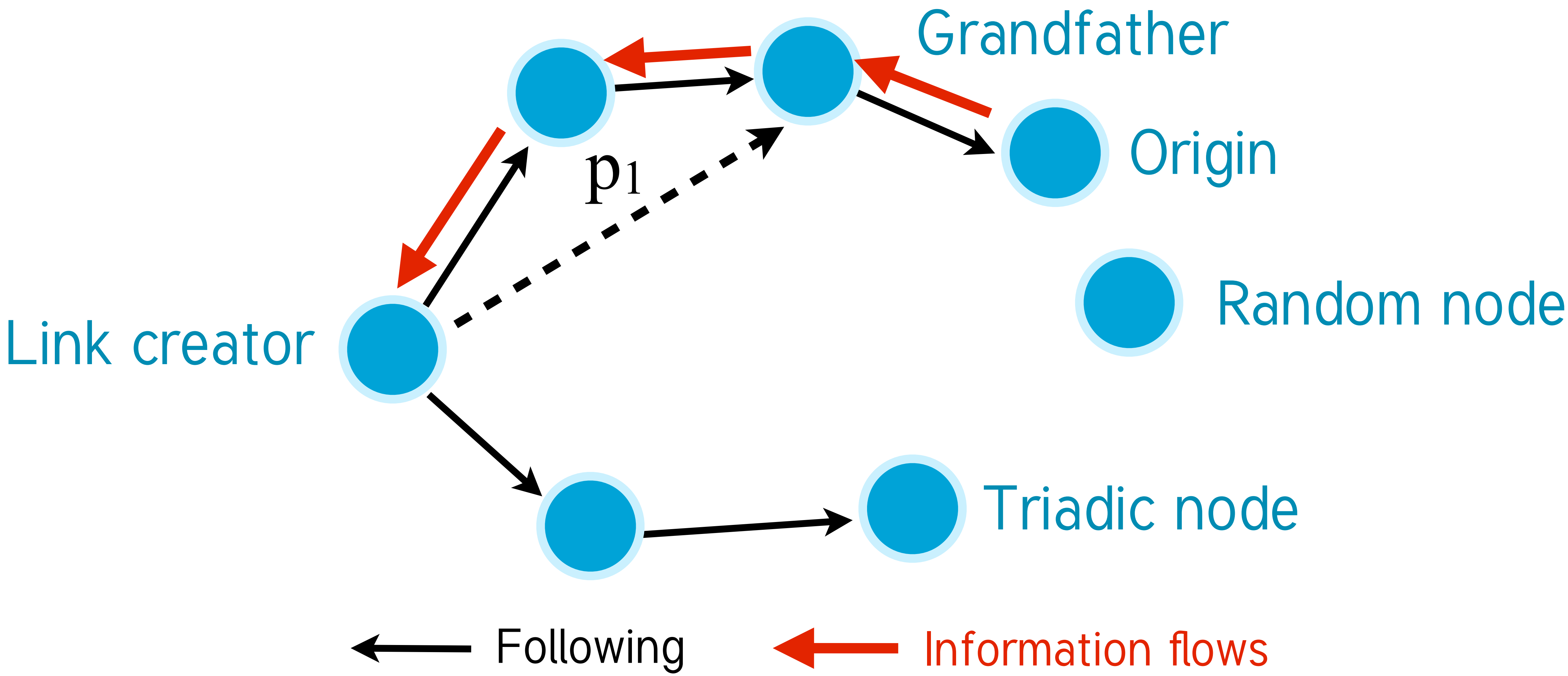
$$\mathcal{L}_{G+\Delta}(p_1, p_2) =$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

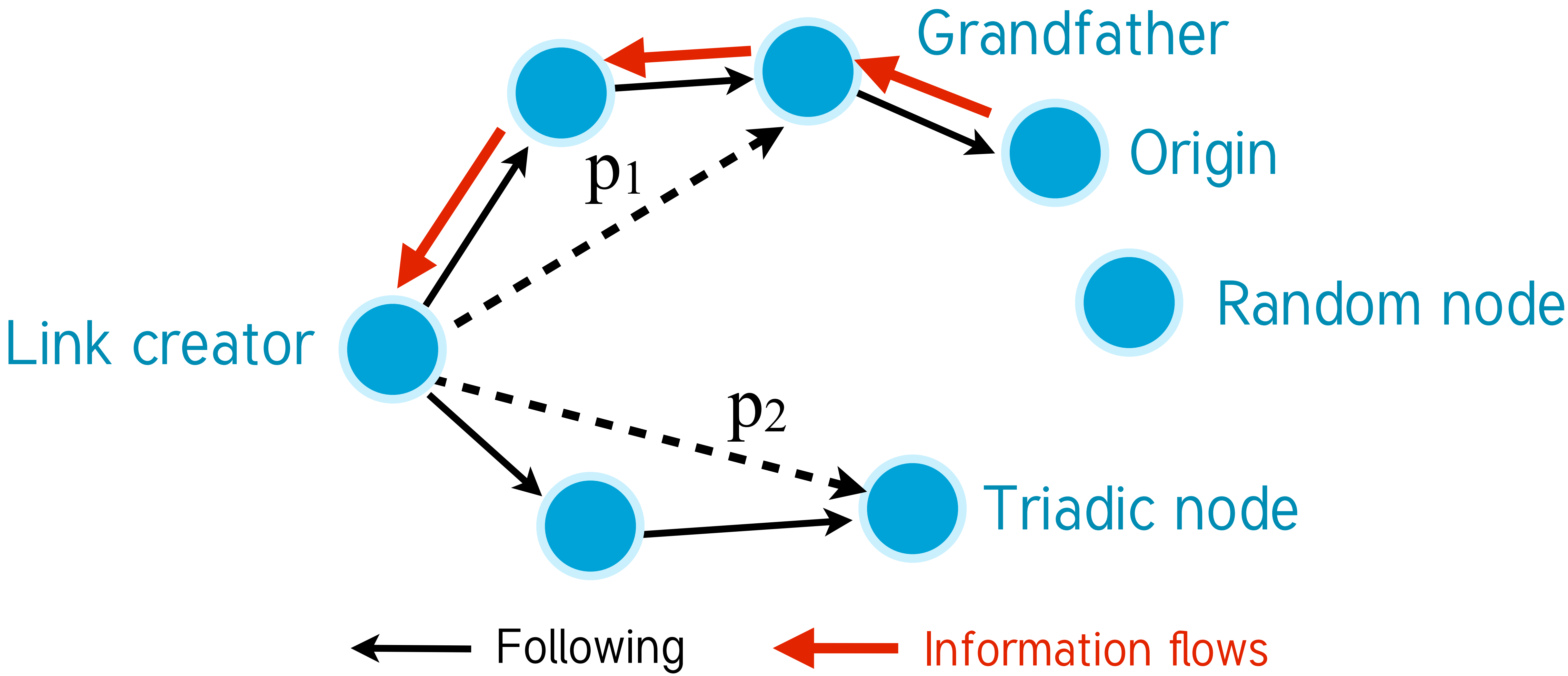
$$\mathcal{L}_{G+\Delta}(p_1, p_2) = \quad p_1 f(\ell | G, \Theta)$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

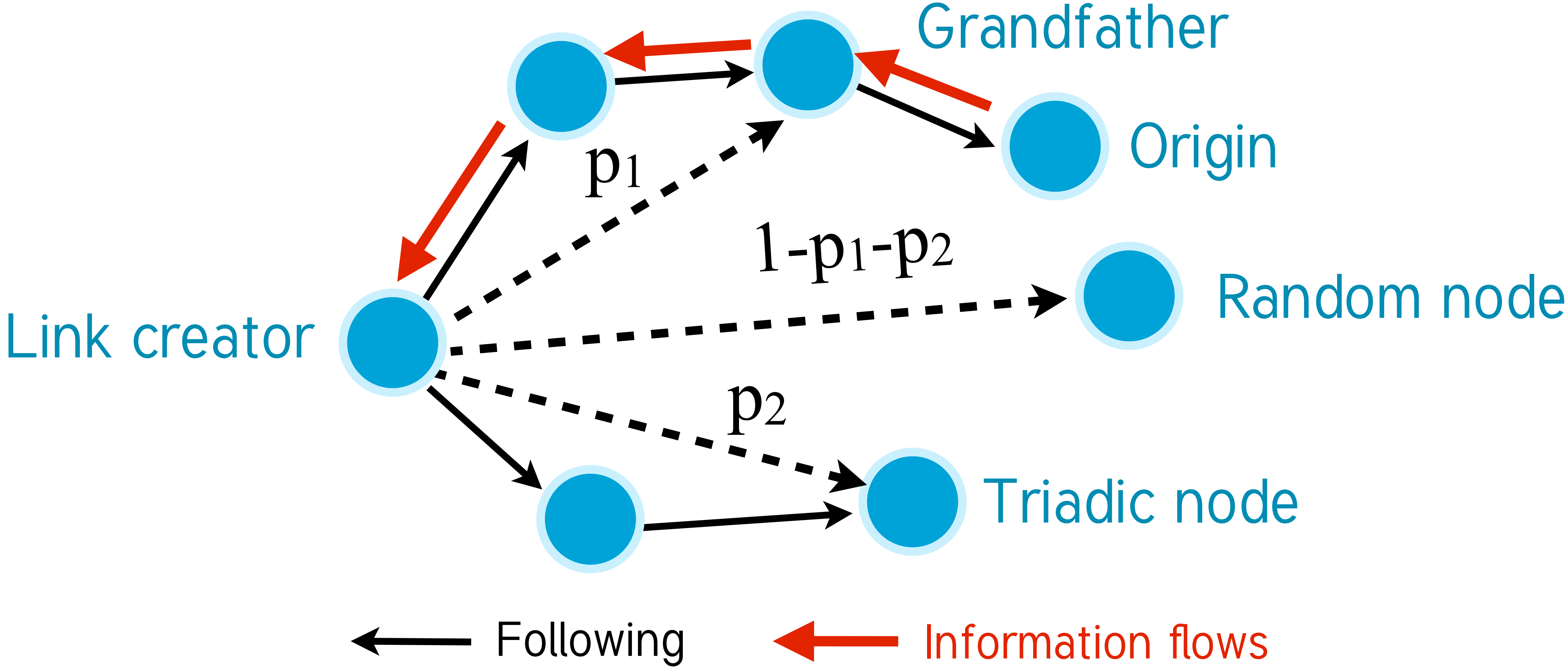
$$\mathcal{L}_{G+\Delta}(p_1, p_2) = \quad p_1 f(\ell | G, \Theta) + p_2 f(\ell | \Delta, \Theta)$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

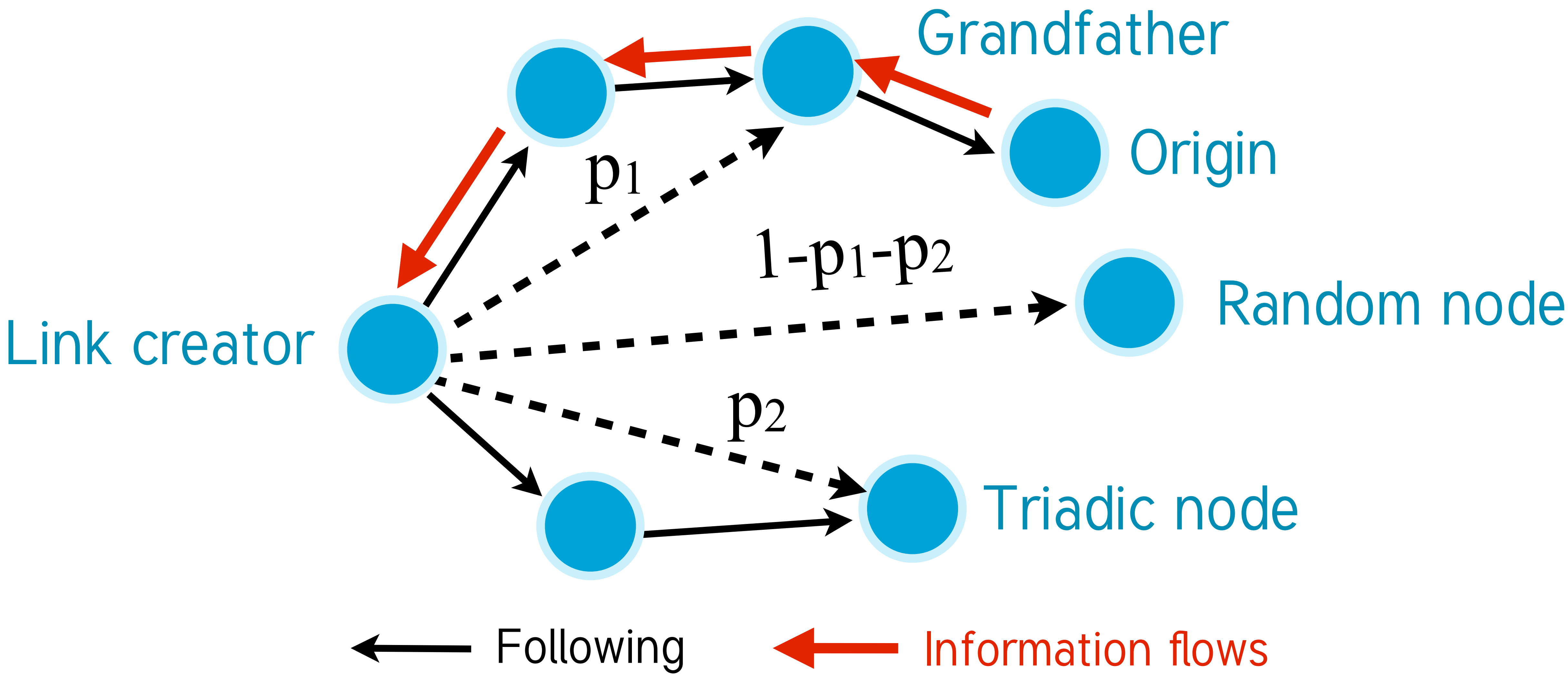
$$\mathcal{L}_{G+\Delta}(p_1, p_2) = \quad p_1 f(\ell|G,\Theta) + p_2 f(\ell|\Delta,\Theta) + (1-p_1-p_2) f(\ell|\mathrm{Rand},\Theta)$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

$$\mathcal{L}_{G+\Delta}(p_1, p_2) = \prod_{\ell=1}^{L} \Big(p_1 f(\ell|G,\Theta) + p_2 f(\ell|\Delta,\Theta) + (1-p_1-p_2) f(\ell|\mathrm{Rand},\Theta)\Big)$$



COMBINED STRATEGY

Example: when $\Gamma = G + \Delta + \mathrm{Rand}$

$$\log \mathcal{L}_{G+\Delta}(p_1, p_2)$$

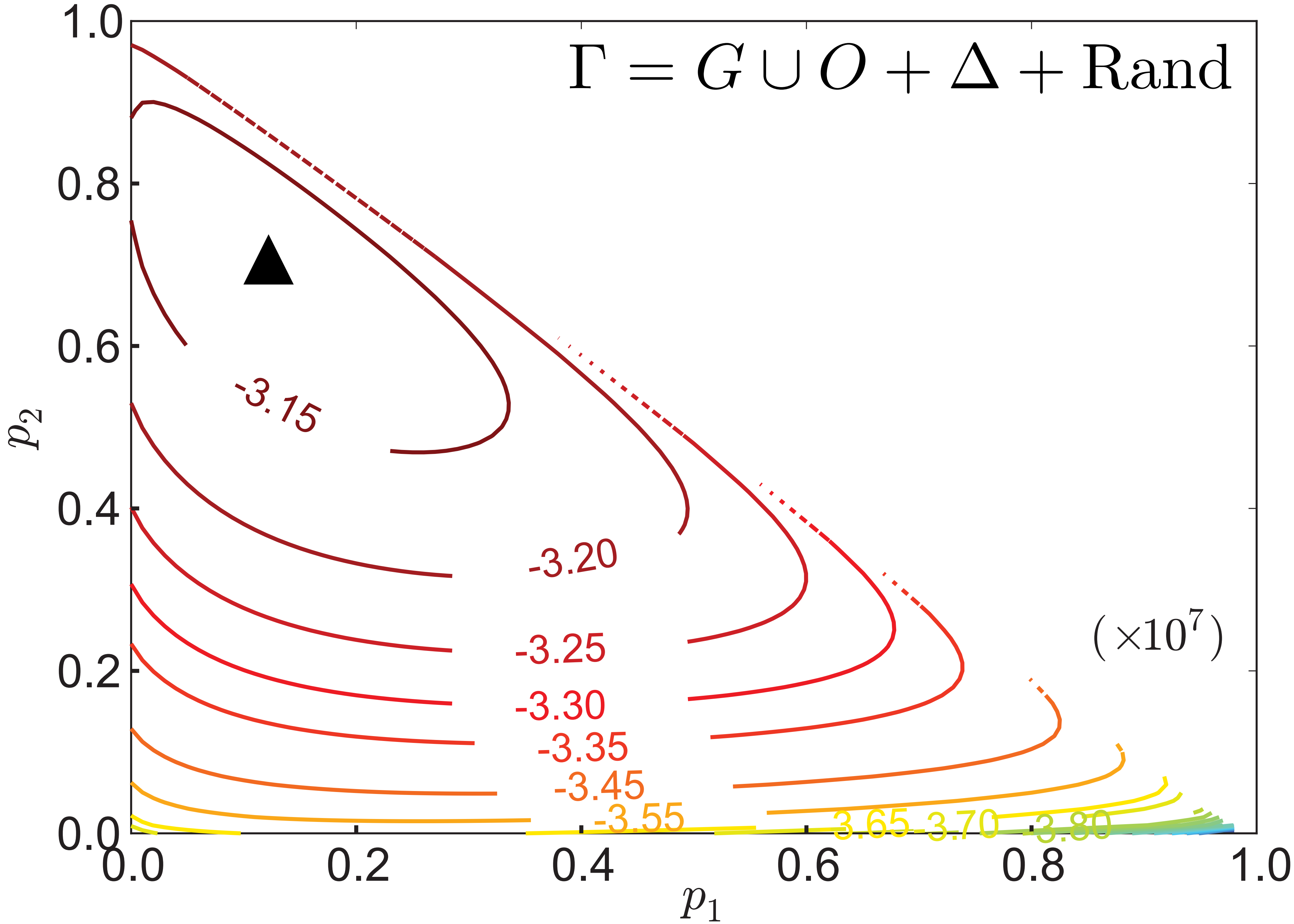
$$= \log \prod_{\ell=1}^{L} \left(p_1 f(\ell|G,\Theta) + p_2 f(\ell|\Delta,\Theta) + (1-p_1-p_2) f(\ell|\mathrm{Rand},\Theta)\right)$$

$$= \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{\ell - k(\ell) - 1}\right) + \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=0}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{1-p_1-p_2}{\ell - k(\ell) - 1}\right)$$

$$+ \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{\ell - k(\ell) - 1}\right) + \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=0}} \log \frac{1-p_1-p_2}{\ell - k(\ell) - 1}$$

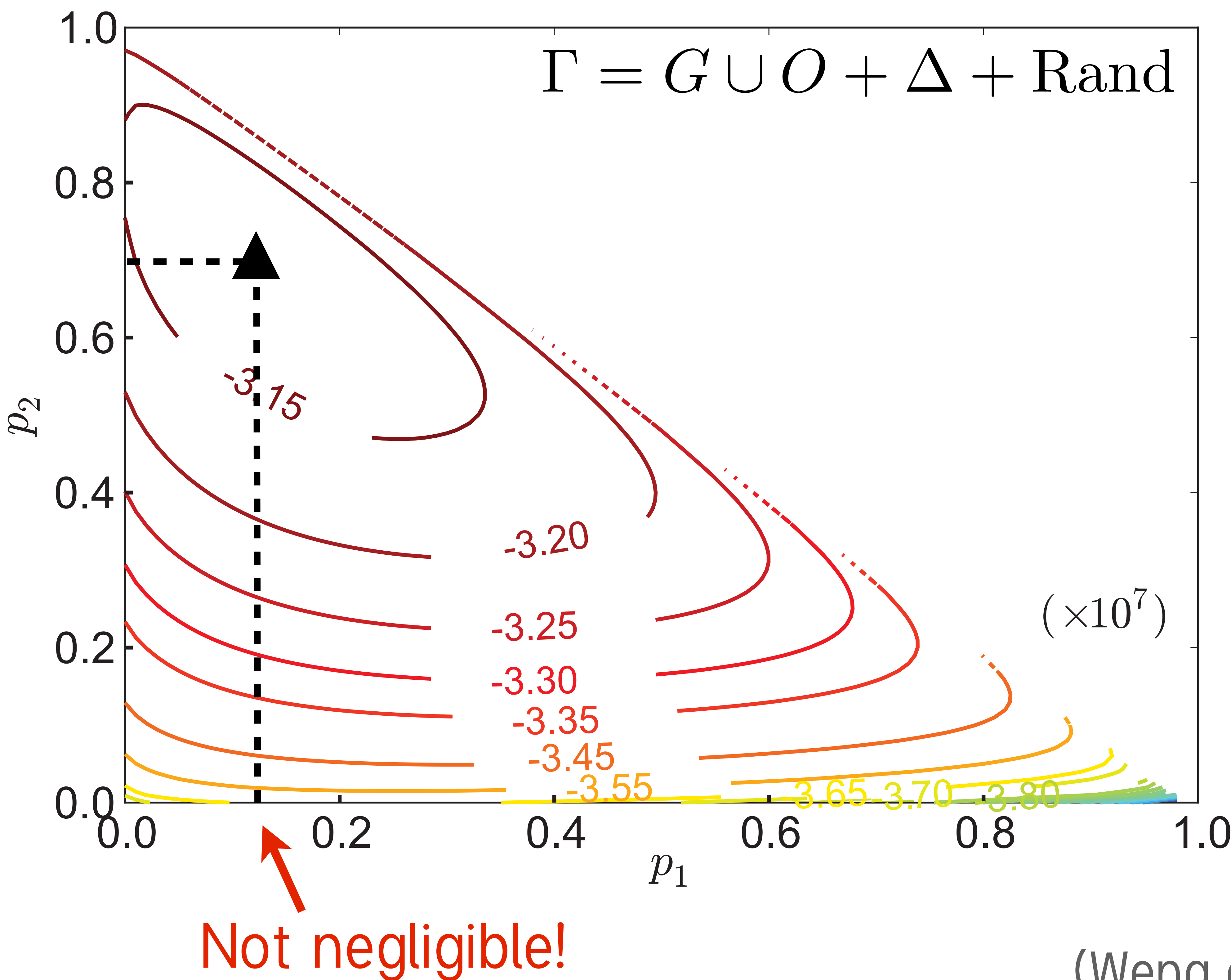
Find the best $p_1, p_2 \in (0,1)$ that **MAXIMIZES** $\log \mathcal{L}_{G+\Delta}(p_1, p_2)$

COMBINED STRATEGY



(Weng et al. 2013)

COMBINED STRATEGY



(Weng et al. 2013)

What is individual link creation strategies in accord with information flows?

CLASSIFY USERS

- Run MLE model for each individual

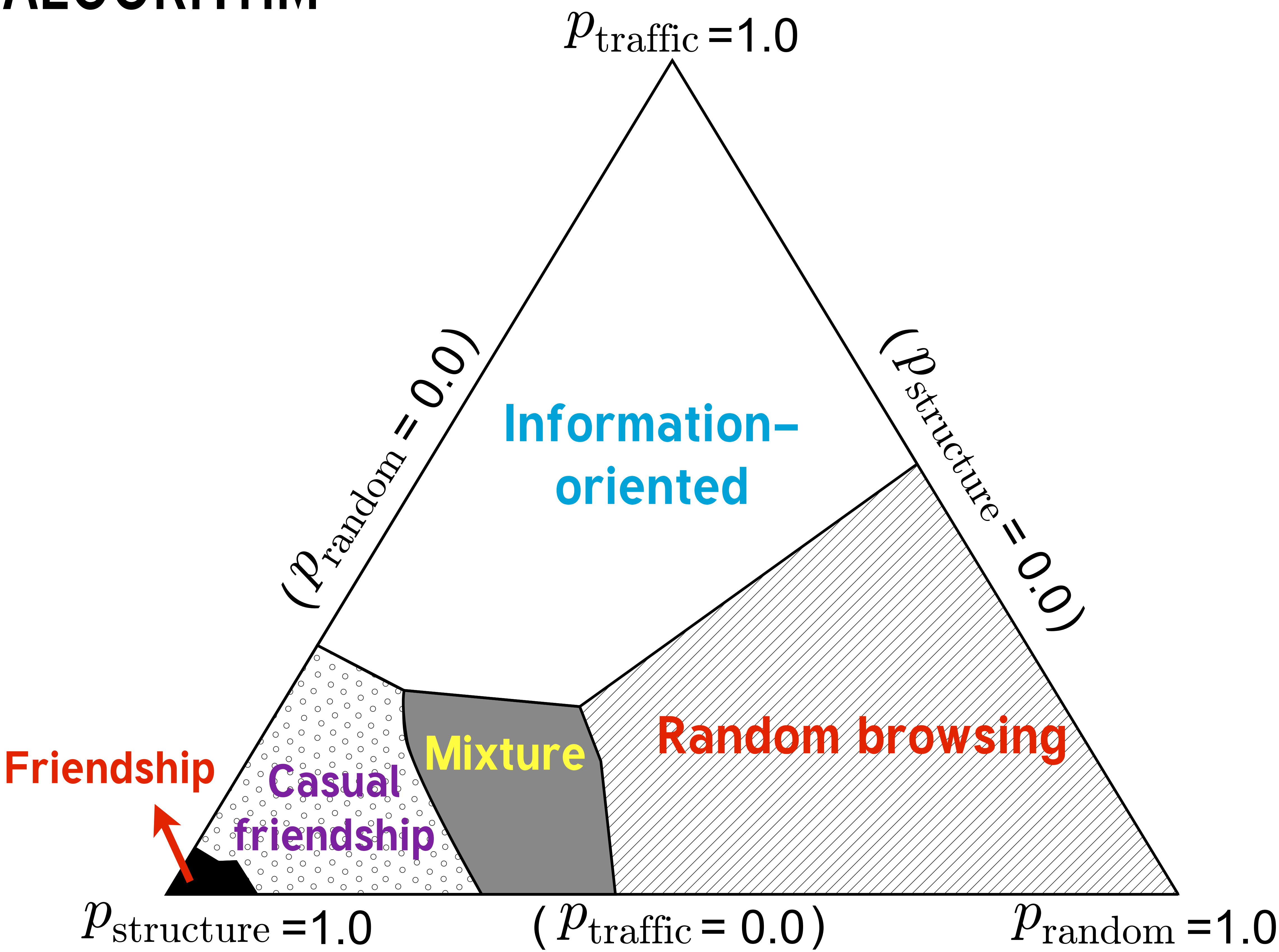
$$G \cup O + \Delta + \mathrm{Rand}$$

$$p_{\mathrm{traffic}} = p_1$$

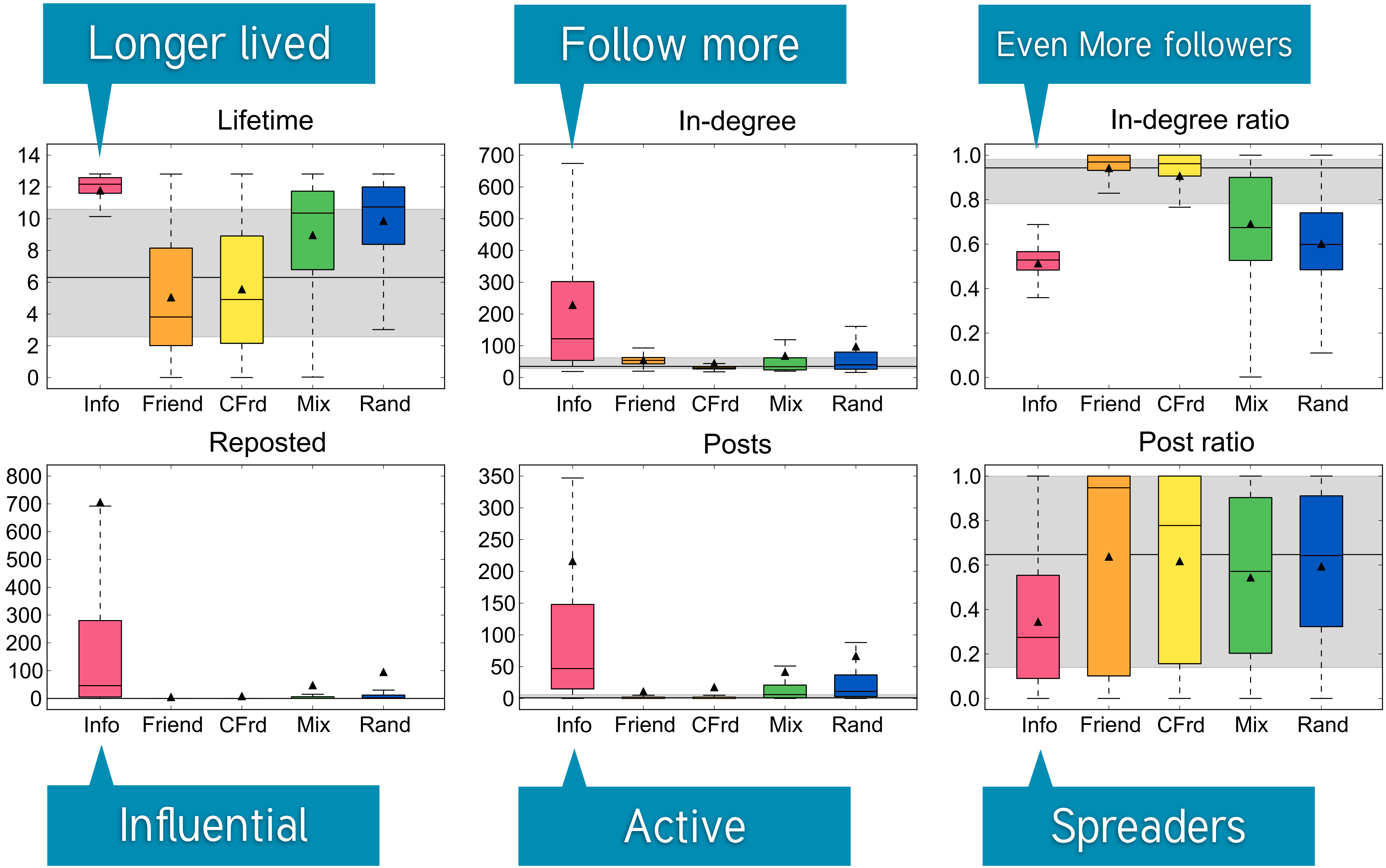
$$p_{\mathrm{structure}} = p_2$$

$$p_{\mathrm{random}} = 1 - p_1 - p_2$$

EM ALGORITHM

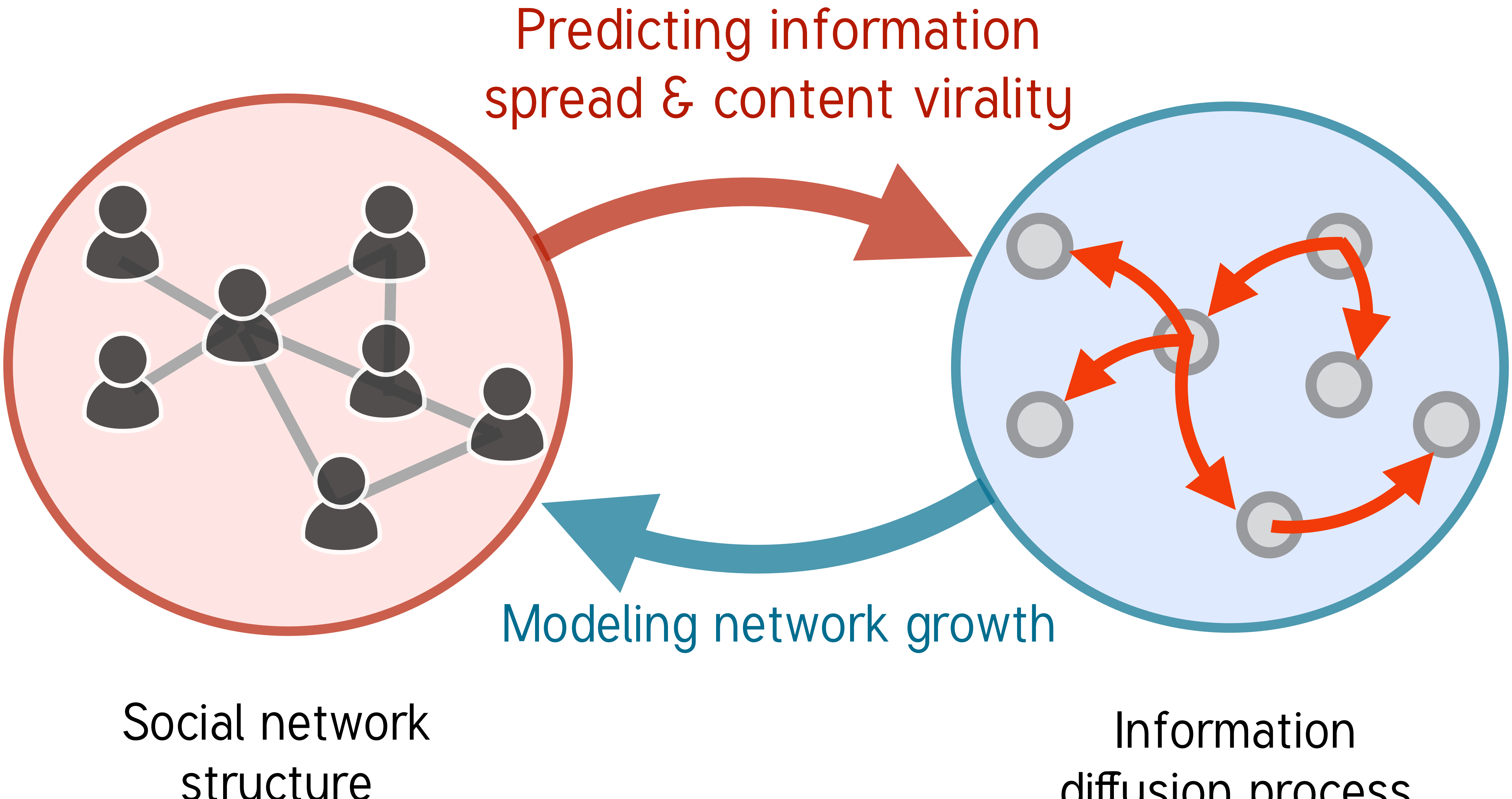


(Weng et al. 2013)



(Weng et al. 2013)

WHAT DID WE LEARN?



FUTURE WORK

FUTURE PROJECTS

1. Evaluation tool of missing link prediction algorithms with **longitudinal** data

2. Shrinkage of human **attention span**

3. Gap between **online** and **offline** social networks

CHALLENGES

1. Data sampling
2. Universality
3. Privacy
4. Open access

CHALLENGES

1. Data sampling
2. Universality
3. Privacy
4. Open access

CHALLENGES

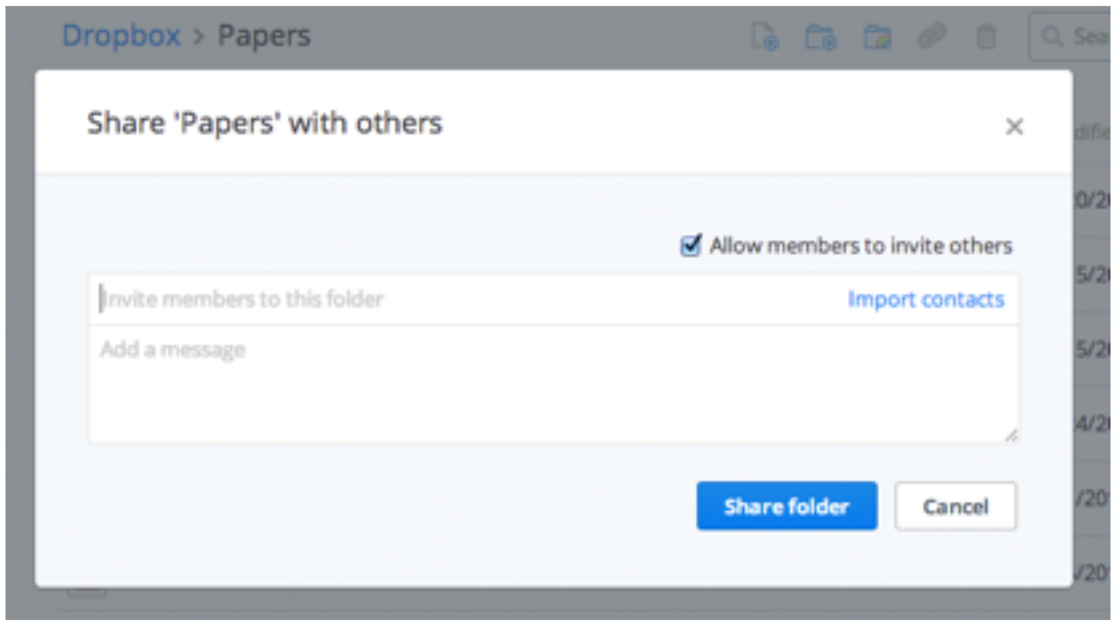
1. Data sampling
2. Universality
3. Privacy
4. Open access

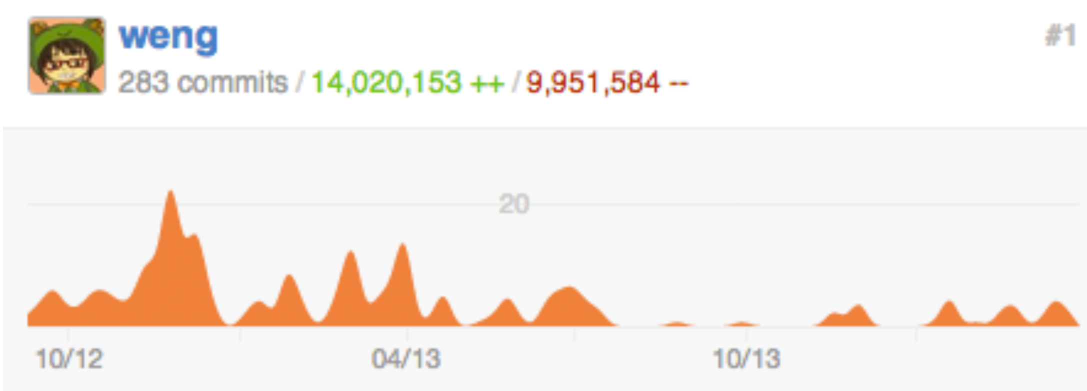
CHALLENGES

1. Data sampling
2. Universality
3. Privacy
4. Open access

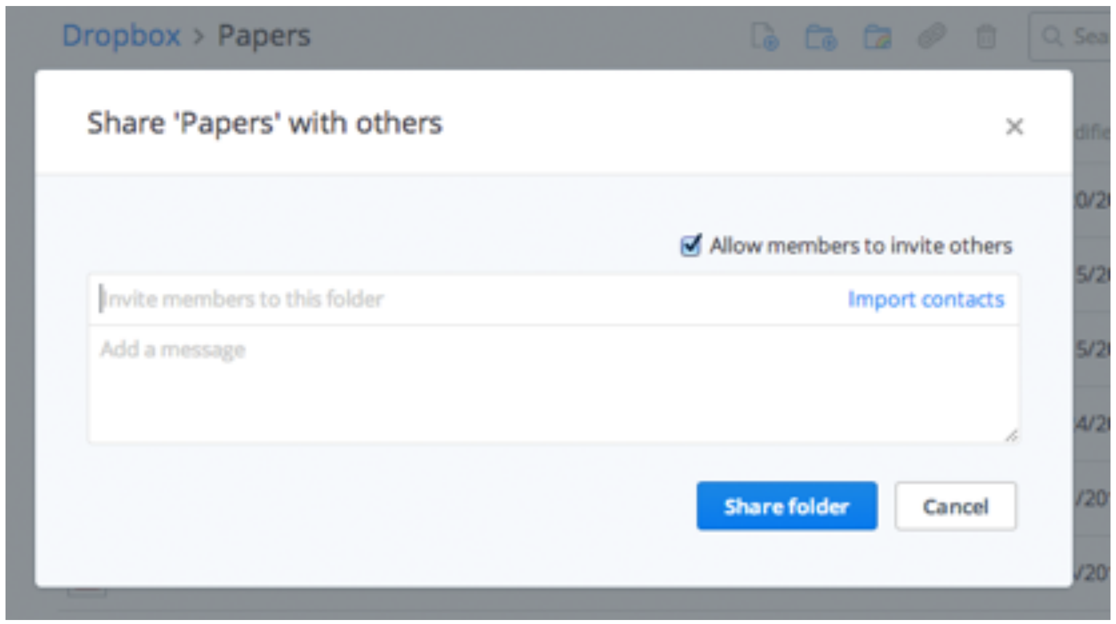
MY FUTURE WORK

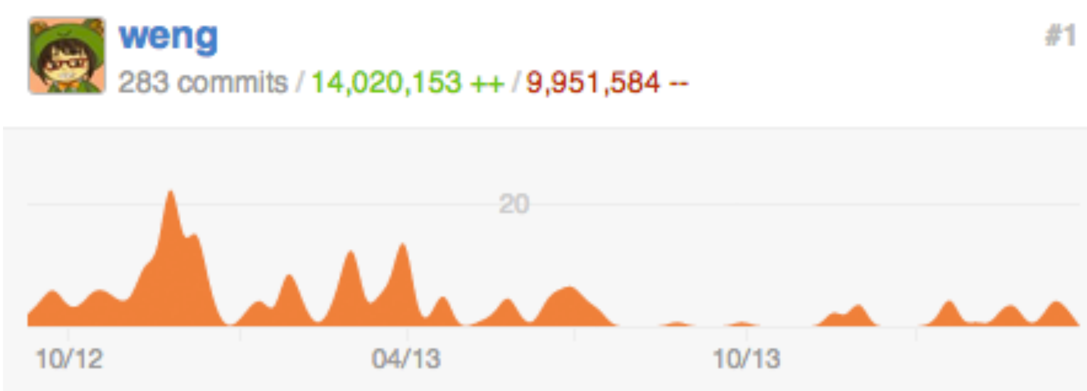
MY FUTURE WORK





MY FUTURE WORK





Casual Business Fun Geek

A BETTER UNDERSTANDING
OF ONLINE INFORMATION
DIFFUSION

NETWORK IS POWERFUL
IN STUDYING
COMPLEX DYNAMICS OF
HUMAN SOCIETY

Selected Papers

- L. Weng, A. Flammini, A. Vespignani, & F. Menczer. **Competitions among topics in a world with limited attention**. *Nature **Sci. Rep.***, (2)335, 2012.
- L. Weng, et al. **The Role of Information Diffusion in the Evolution of Social Networks**. In: ***KDD***. 2013.
- L. Weng, F. Menczer, & Y.-Y. Ahn. **Virality Prediction and Community Structure in Social Networks**. *Nature **Sci. Rep.***, (3)2522, 2013.
- L. Weng, F. Menczer, & Y.-Y. Ahn. **Predicting Meme Virality in Social Networks using Network and Community Structure.** In: ***ICWSM***. 2014.
- L. Weng & T. Lento. **Topic-based Clusters in Egocentric Networks on Facebook**. In: ***ICWSM***. 2014.
- L. Weng & F. Menczer. **Topicality and Social Impact: Diverse Messages but Focused Messengers**. *Under review*. 2014.

Check http://lilianweng.github.io for more and details.

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

SIGILLUM UNIVERSITATIS INDIANENSIS
LUX ET VERITAS
MDCCCXX

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington

CNetS

Truthy

givealink.org
Share your links!

THANK YOU!
QUESTIONS?

Sincerely, Lilian