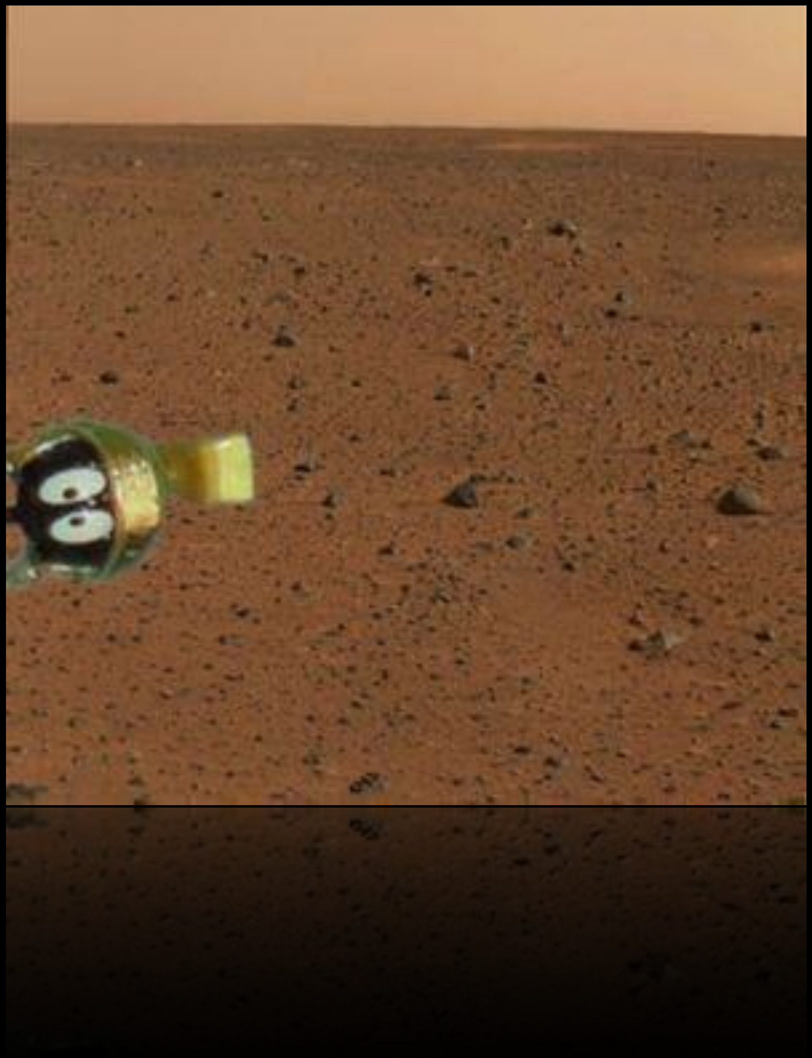
Why does Alice follow Bob?

Filippo Menczer
Center for Complex Networks and Systems Research
School of Informatics and Computing
Indiana University, Bloomington

The Role of Information Diffusion in the Evolution of Social Networks

Lilian Weng[1], Jacob Ratkiewicz[2], Nicola Perra[3], Bruno Gonçalves[4], Carlos Castillo[5], Francesco Bonchi[6], Rossano Schifanella[7], Filippo Menczer[1], Alessandro Flammini[1]
[1]School of Informatics and Computing, Indiana University Bloomington, USA [2]Google Inc.
[3]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, USA
[4]Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France
[5]Qatar Computing Research Institute [6]Yahoo! Research Barcelona
[7]Department of Computer Science, University of Torino, Italy



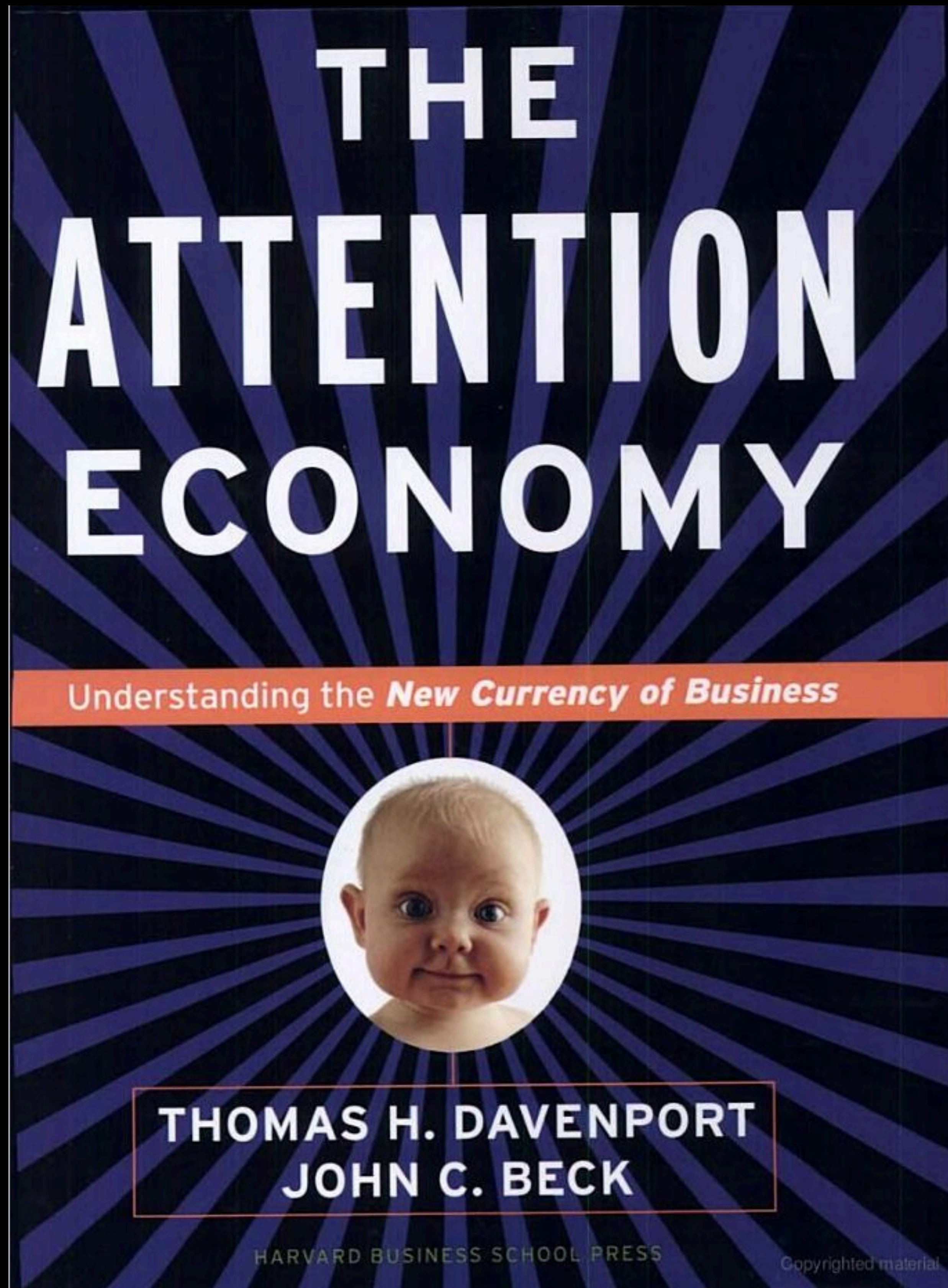
Marvin

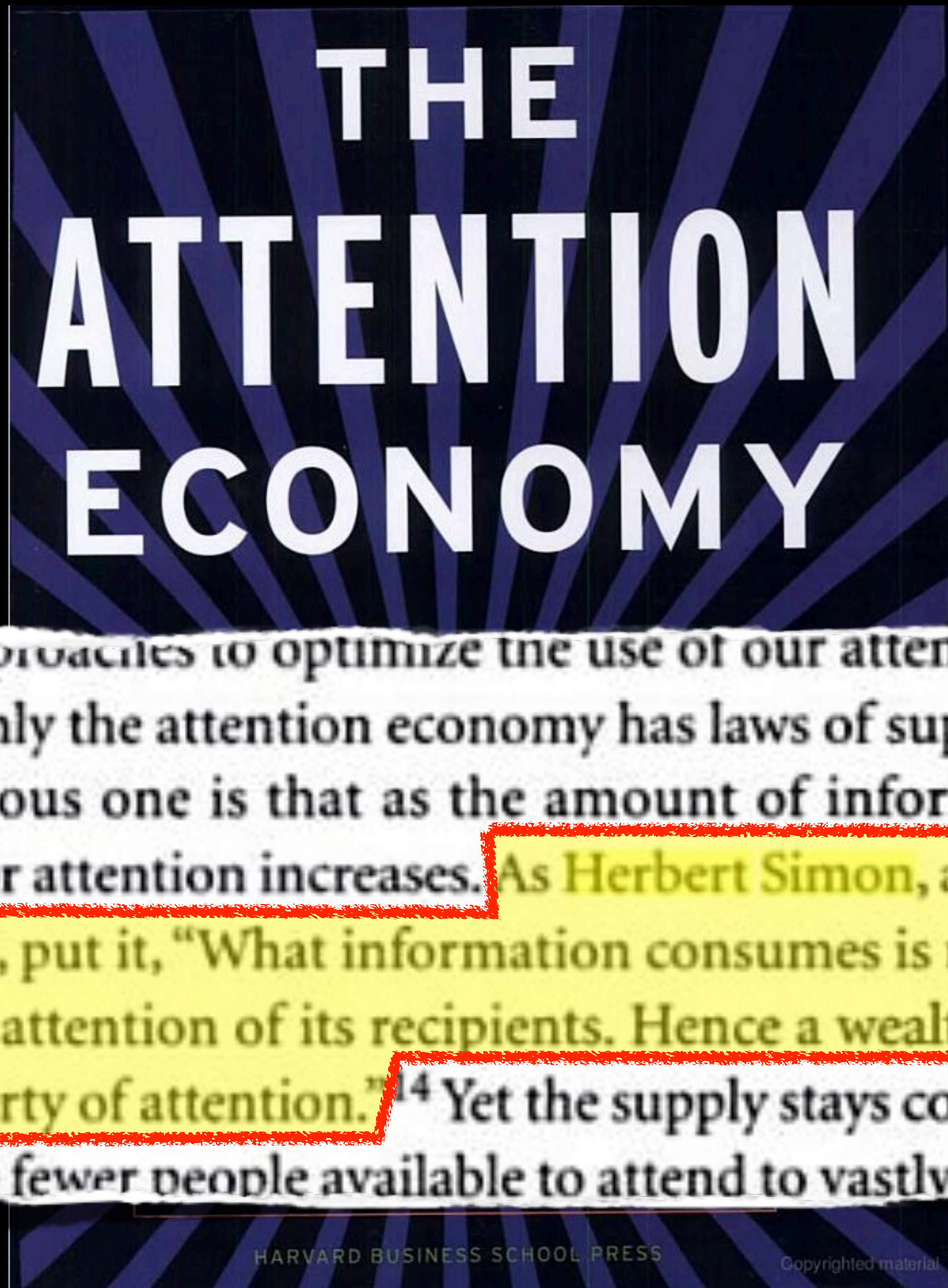
2012.08.05 23:05

Stamen


Mousy Smell, Mousy Do | New Pig Flu | Evolution in Action
www.sciencenews.org
Science News
MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC ■ OCTOBER 20, 2012
#Politicalspam
Tracking truthiness in the Twitterverse
Reckless on Ritalin
Great Plate Breakup
Voyager 1 Nears Solar System's Outer Limit

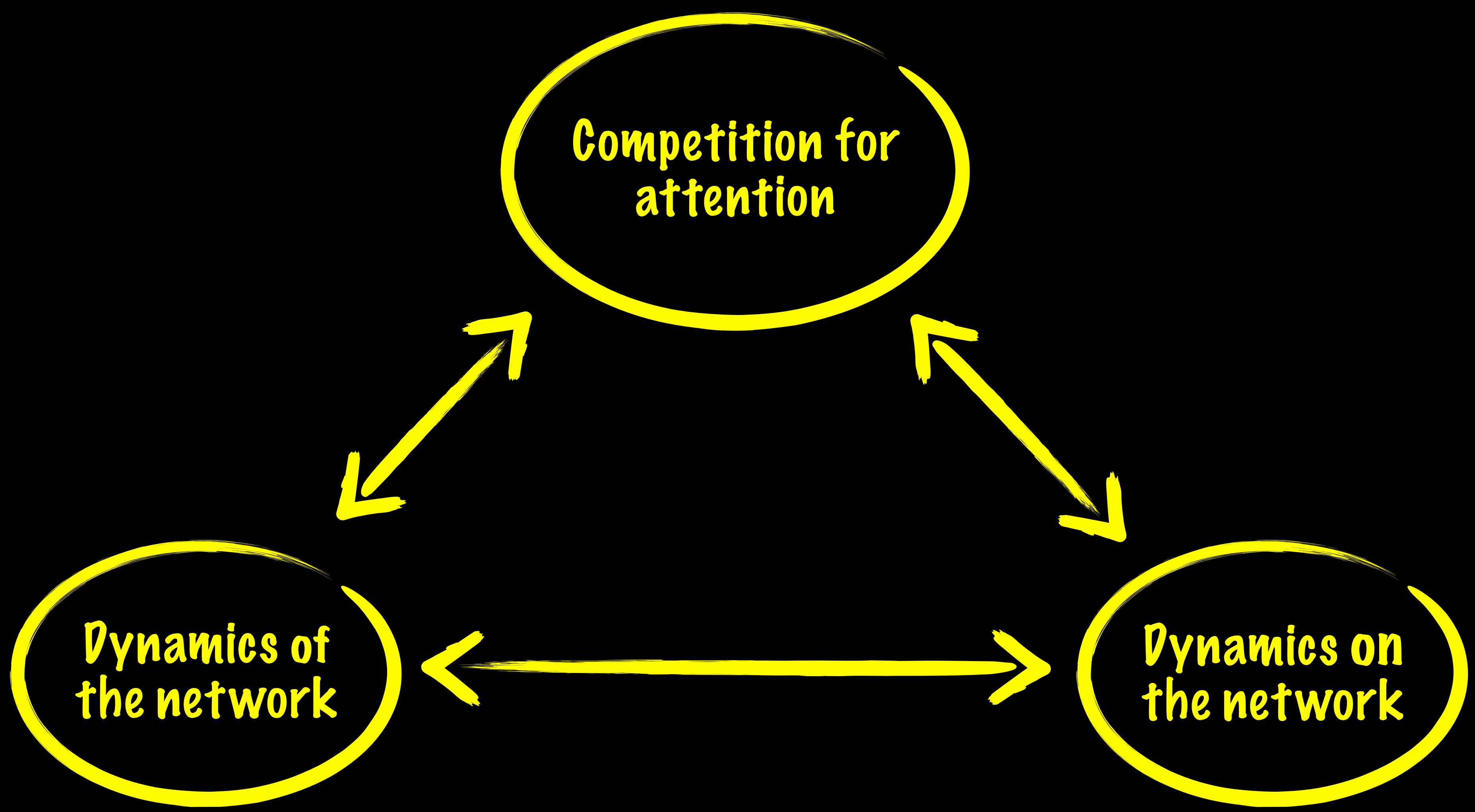
truthy.indiana.edu

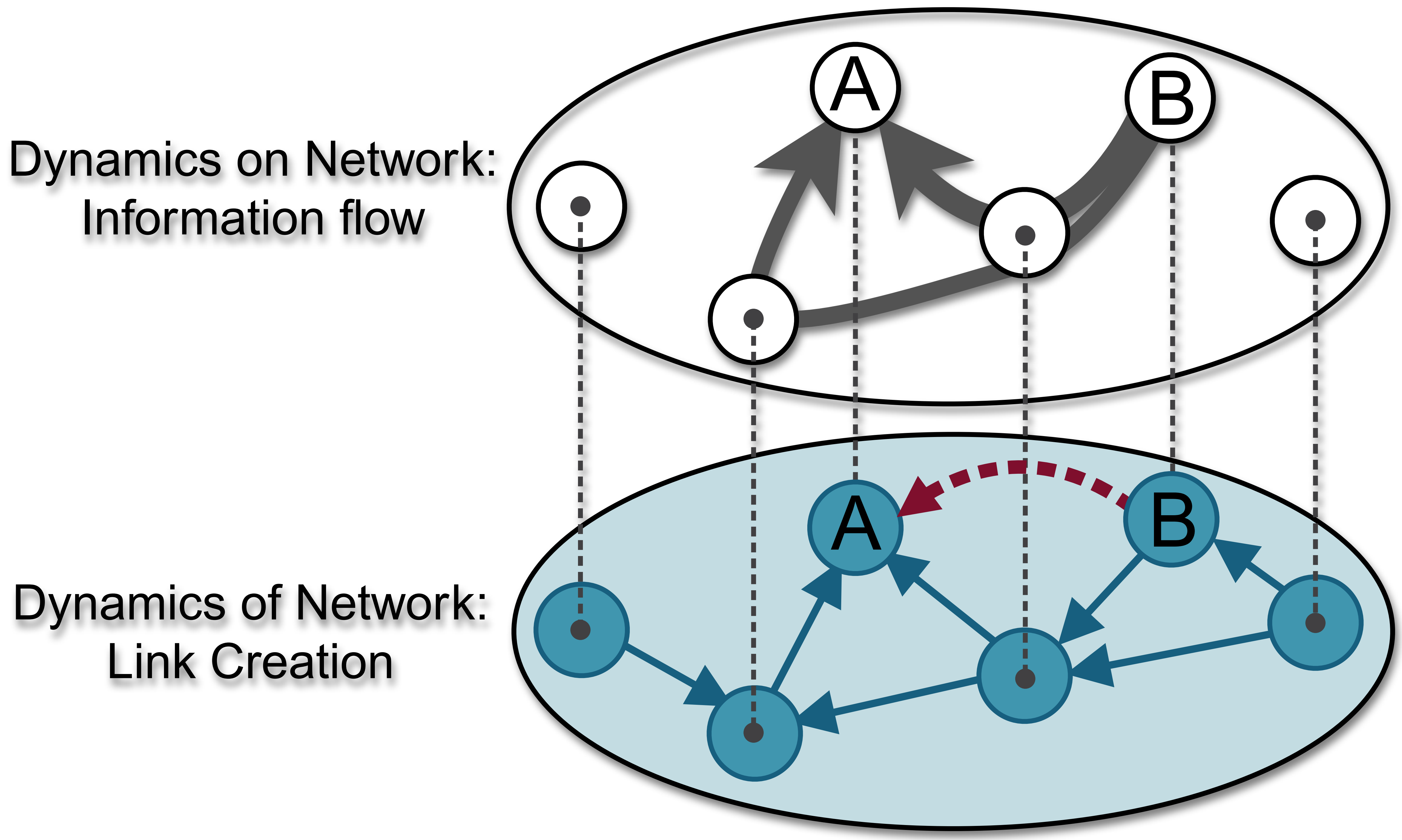

THE
ATTENTION
ECONOMY
Understanding the ***New Currency of Business***
THOMAS H. DAVENPORT
JOHN C. BECK
HARVARD BUSINESS SCHOOL PRESS
Copyrighted material

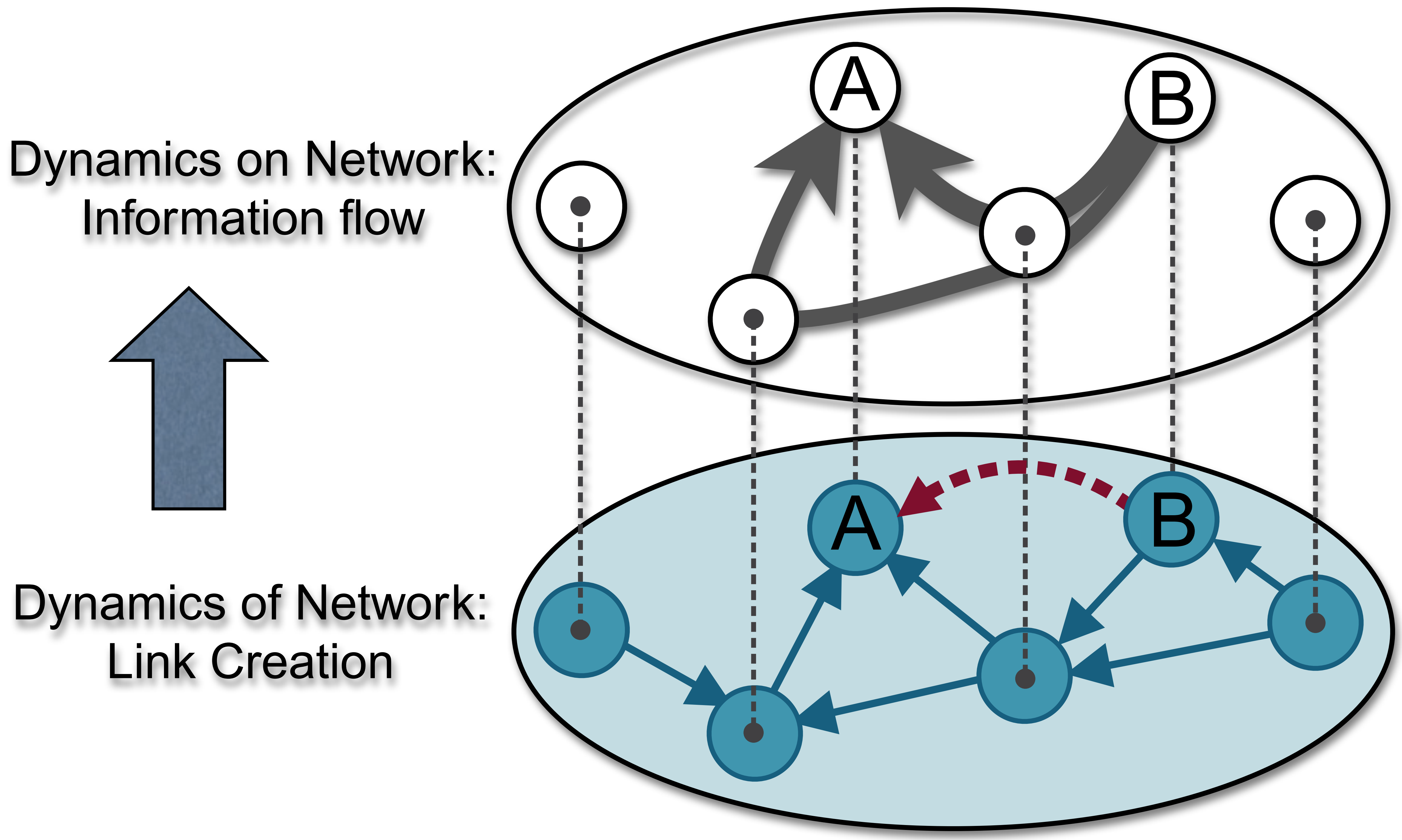


similar approaches to optimize the use of our attention.

Certainly the attention economy has laws of supply and demand. The most obvious one is that as the amount of information increases, the demand for attention increases. As Herbert Simon, a Nobel prize–winning economist, put it, "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention."[14] Yet the supply stays constant or even shrinks if there are fewer people available to attend to vastly more information. As

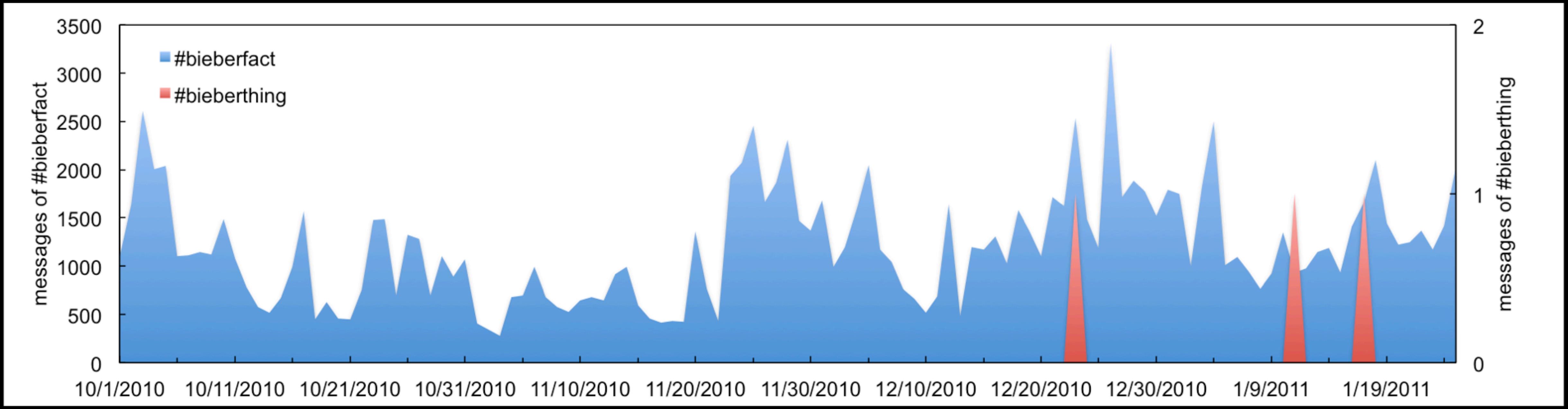

Competition for attention
Dynamics of the network
Dynamics on the network


Dynamics on Network:
Information flow
A
B
Dynamics of Network:
Link Creation
A
B

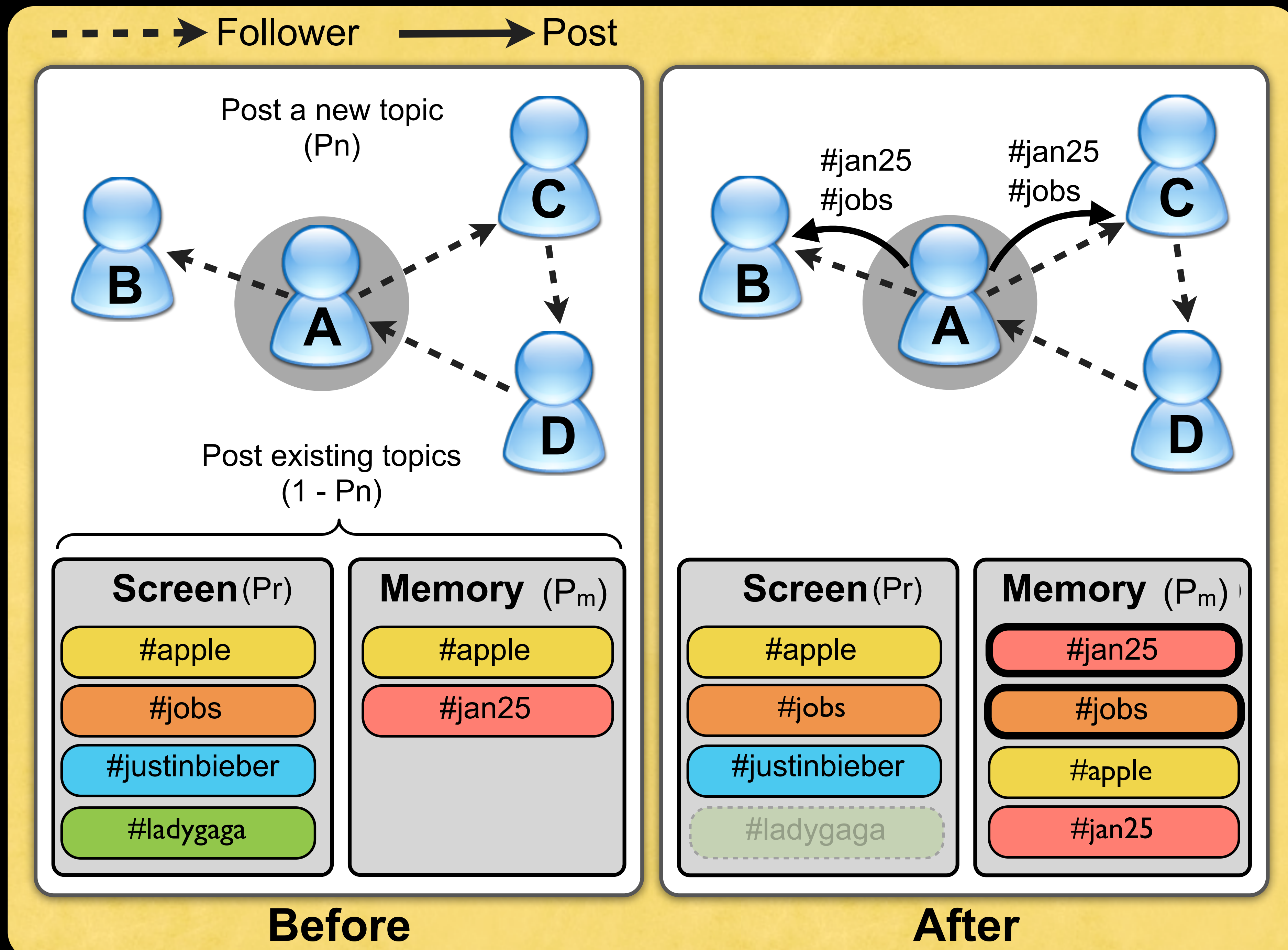


Dynamics on Network:
Information flow
A
B
Dynamics of Network:
Link Creation
A
B

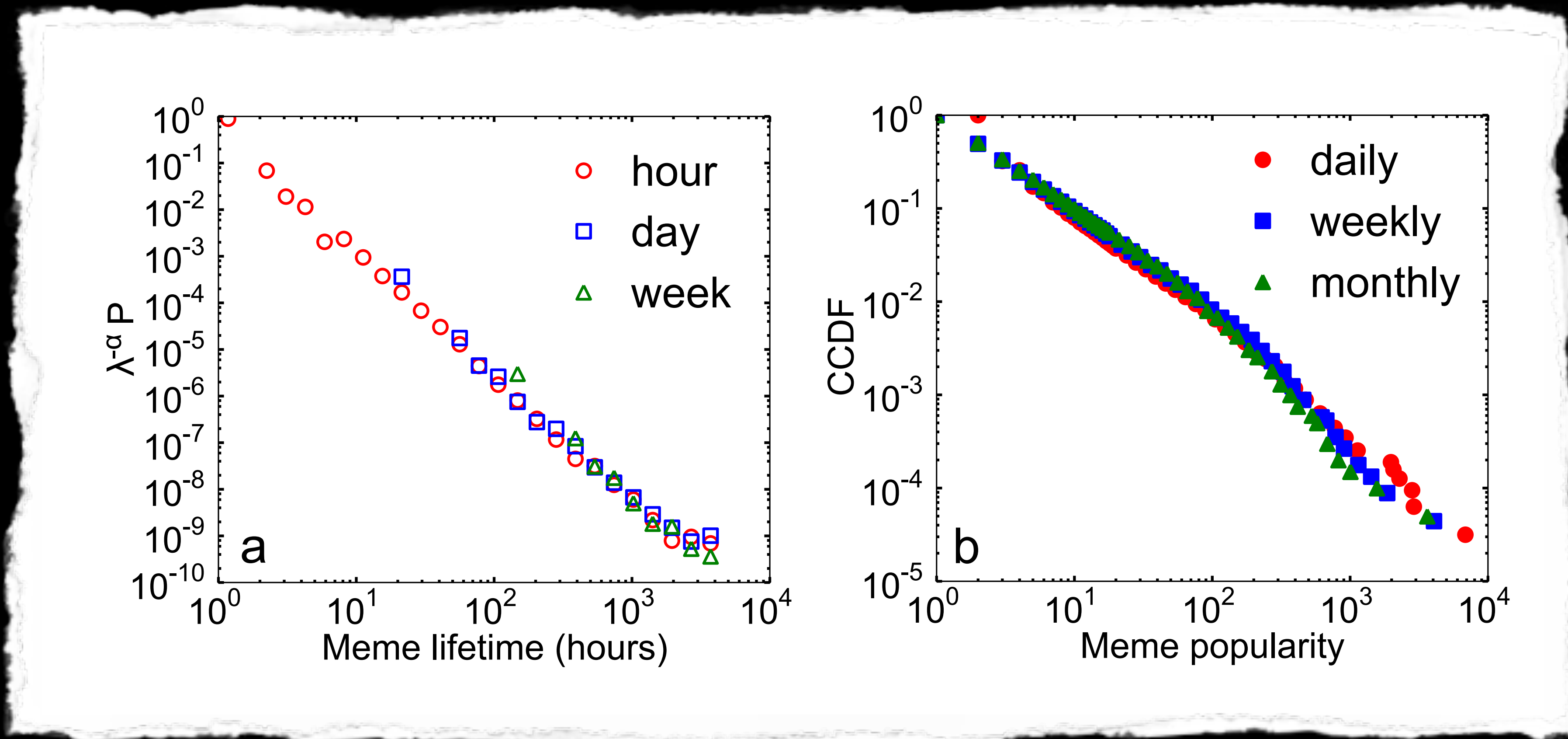
meme	popularity (number of messages)	lifetime (longest consecutive number of days)
#bieberfact	139760	145
#bieberthing	3	1



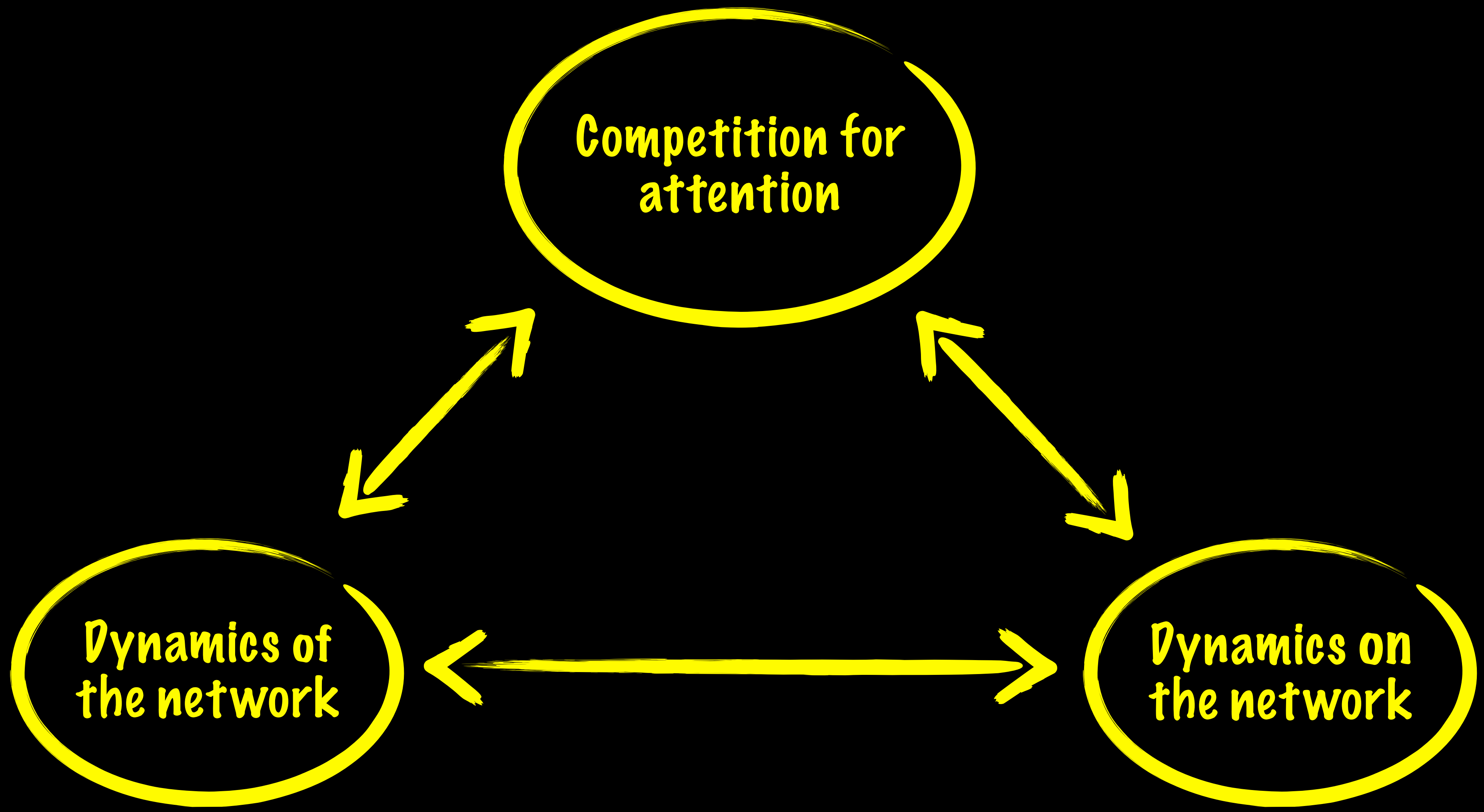
Two key ingredients

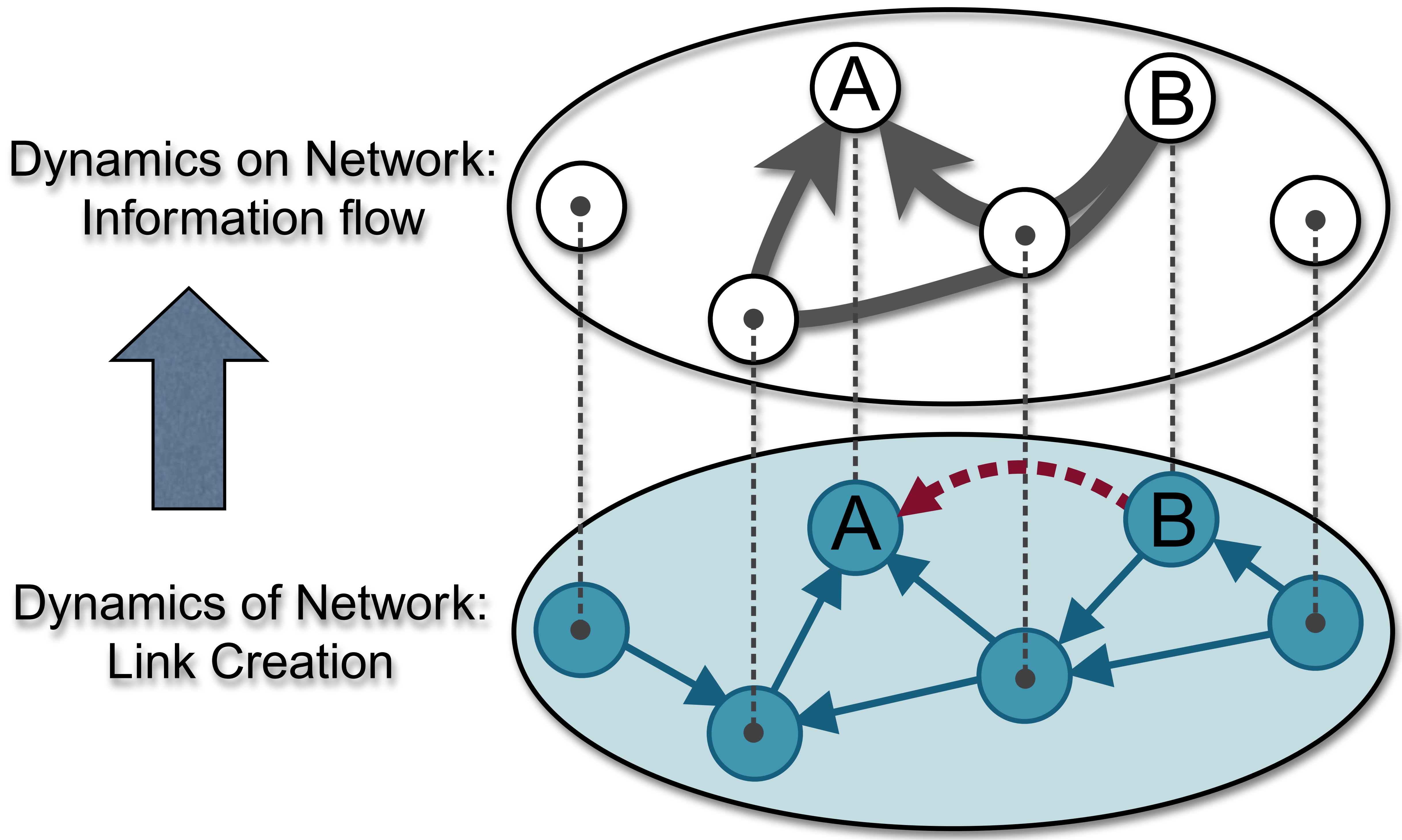


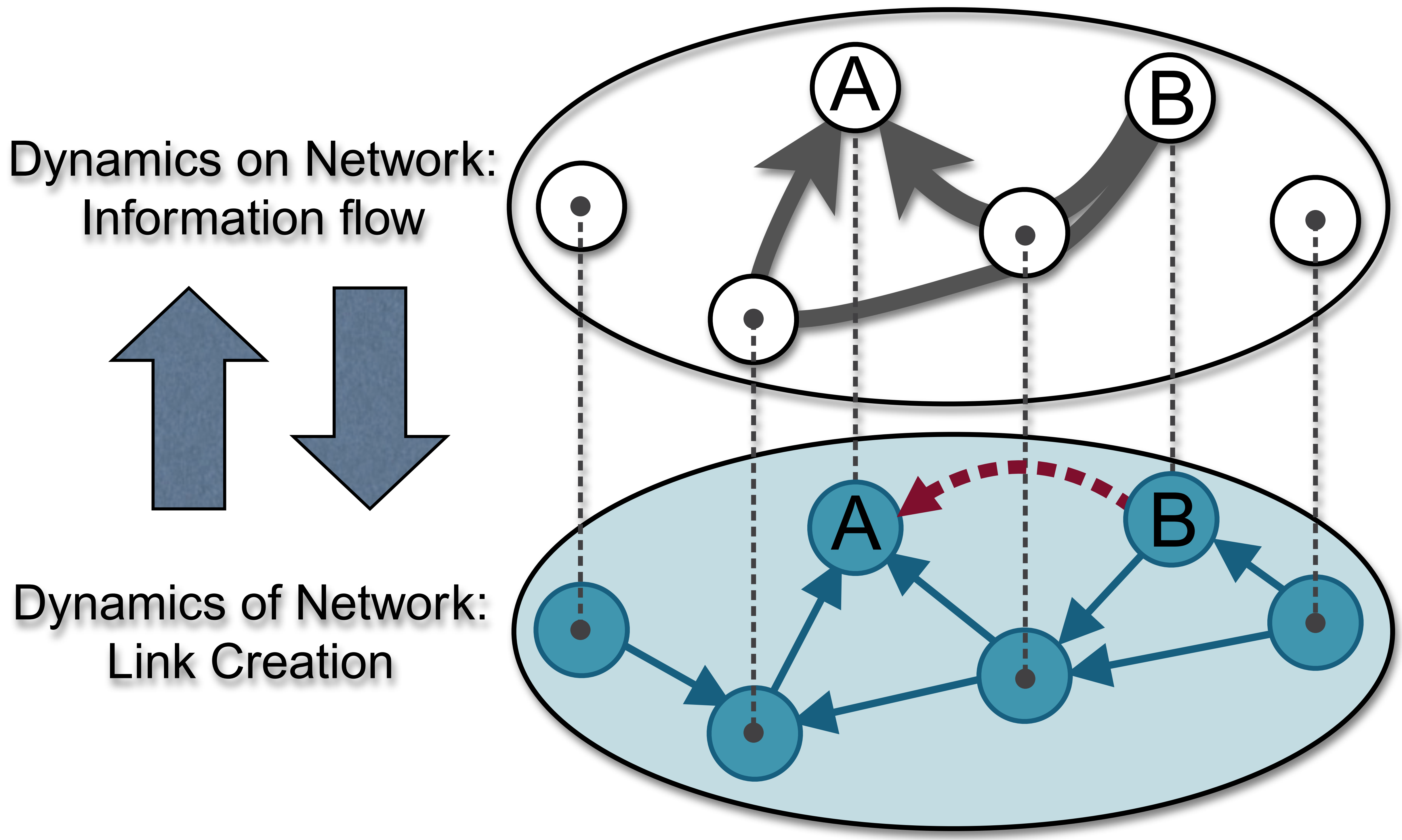
Weng *et al.* *Nature Sci. Rep.* 2012

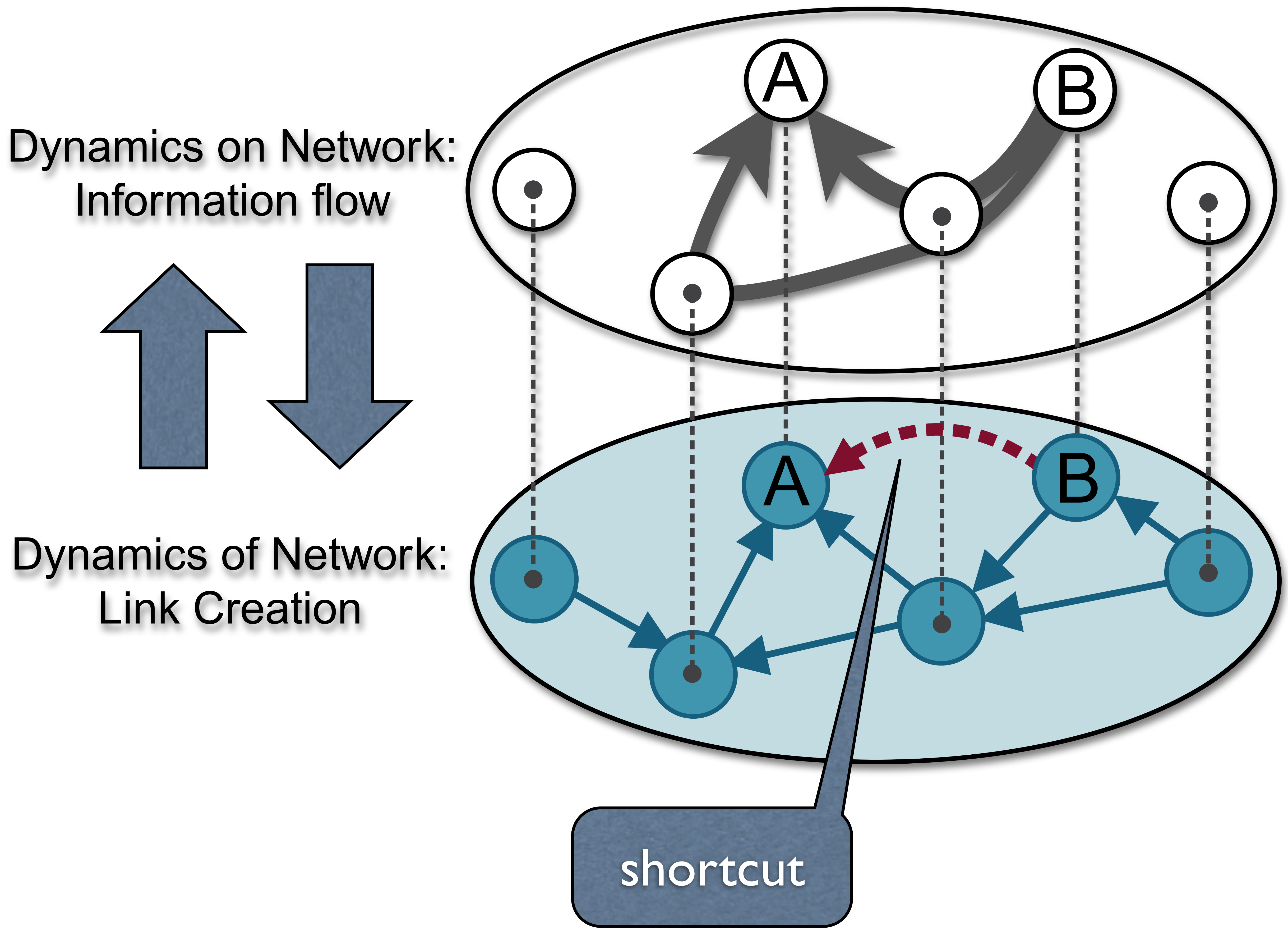


Dataset: Twitter 10% sample
October 2010 – January 2011
~12.5M users, ~1.3M hashtags

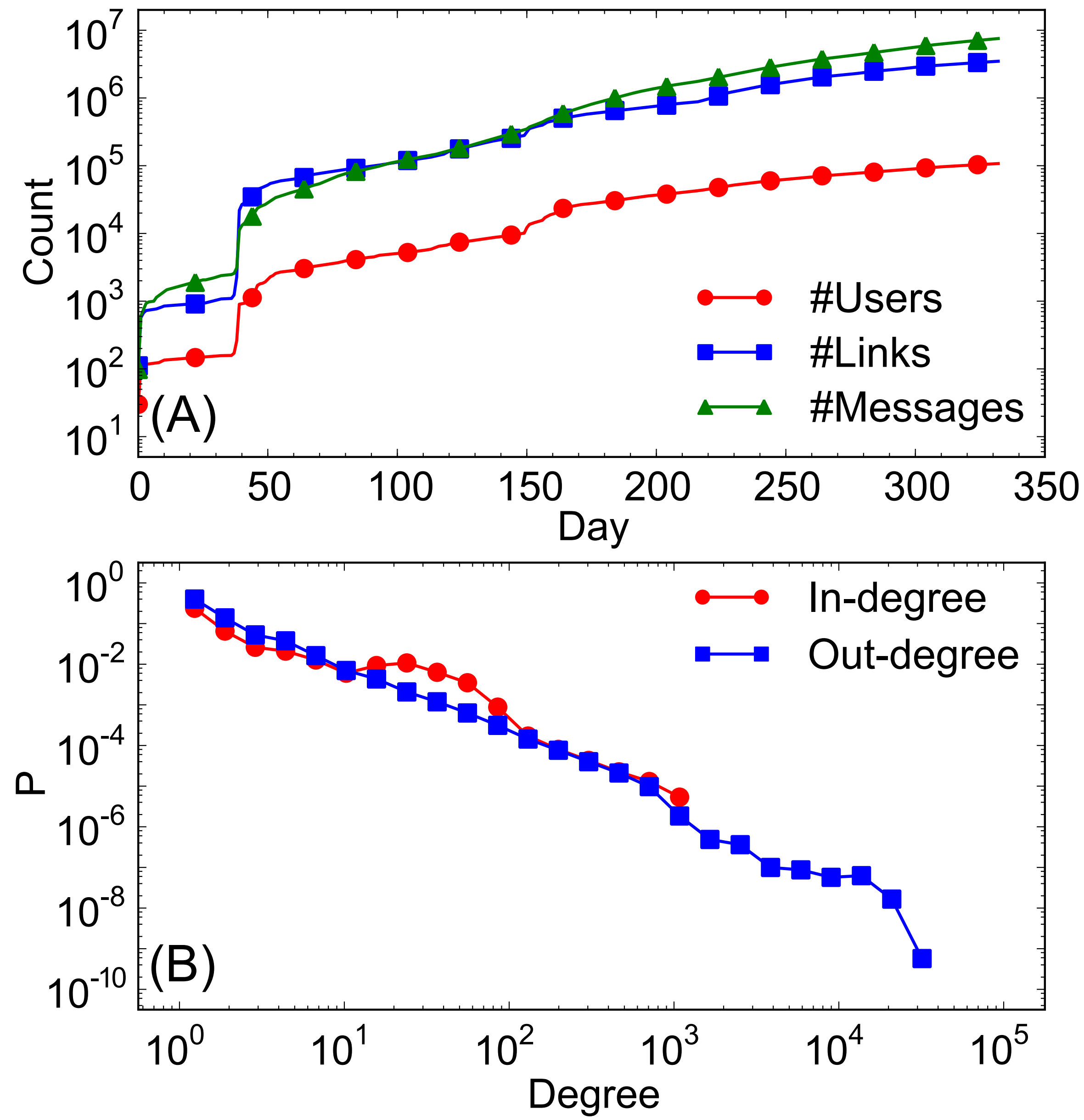

Competition for attention
Dynamics of the network
Dynamics on the network


Dynamics on Network:
Information flow
Dynamics of Network:
Link Creation
A
B
A
B


Dynamics on Network:
Information flow
Dynamics of Network:
Link Creation
A
B
A
B

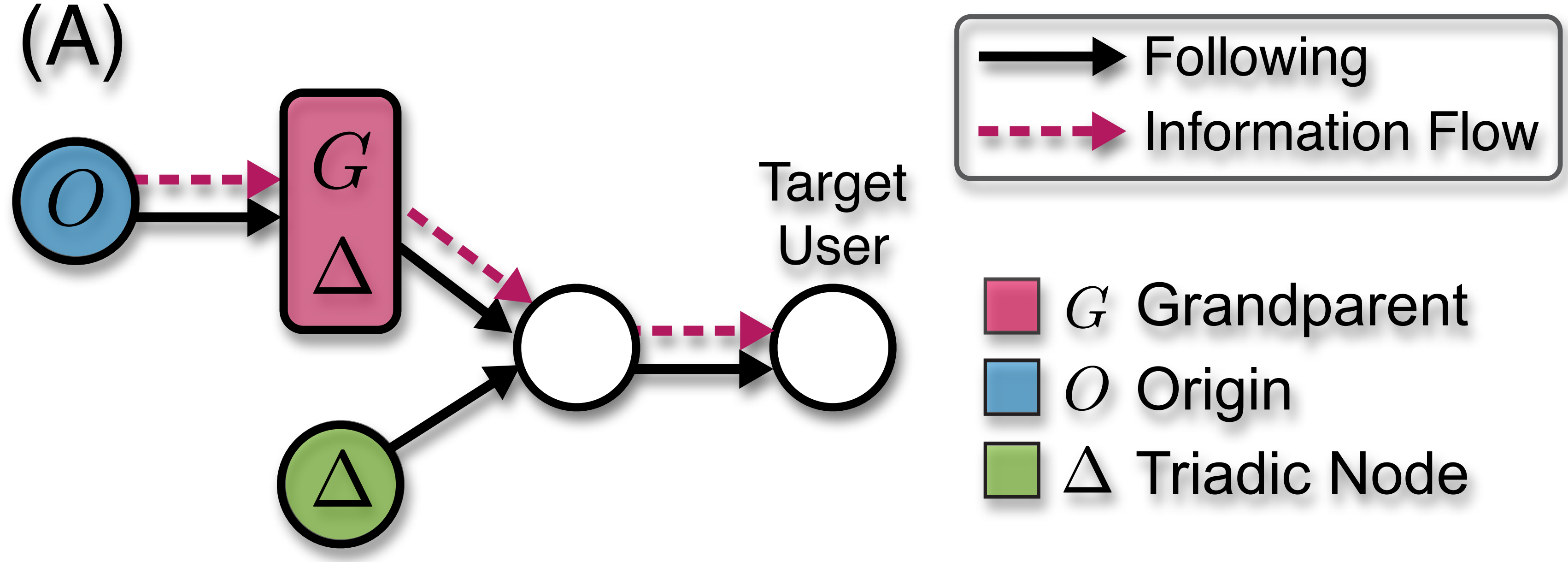

Dynamics on Network:
Information flow
A
B
Dynamics of Network:
Link Creation
A
B
shortcut

Dataset: Yahoo! Meme
April 2009 – March 2010

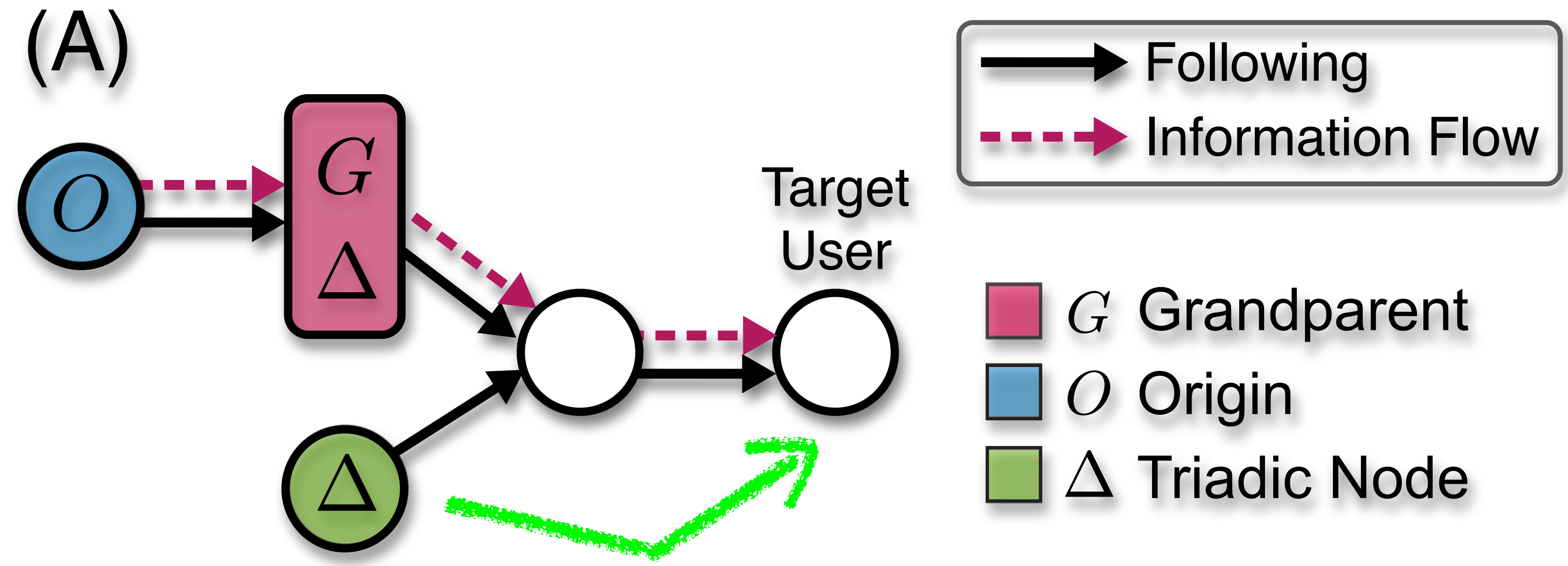


~128k users, ~3.5M links, ~7M posts

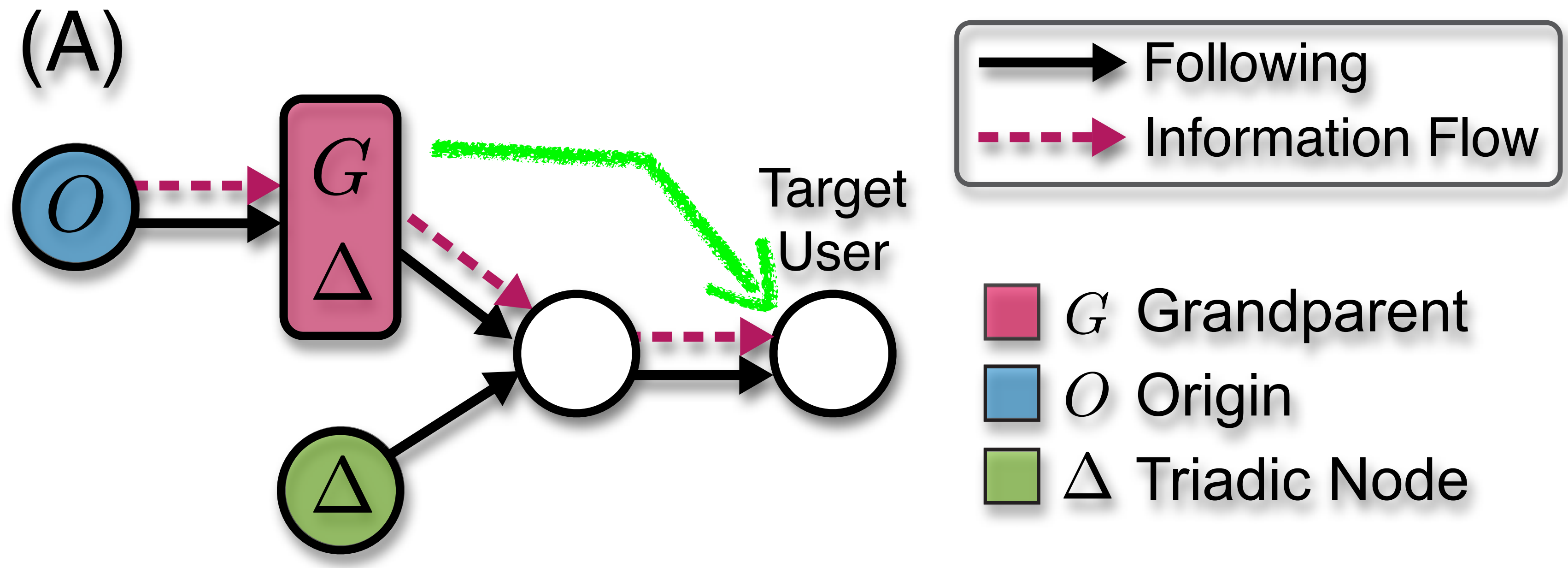
(A)

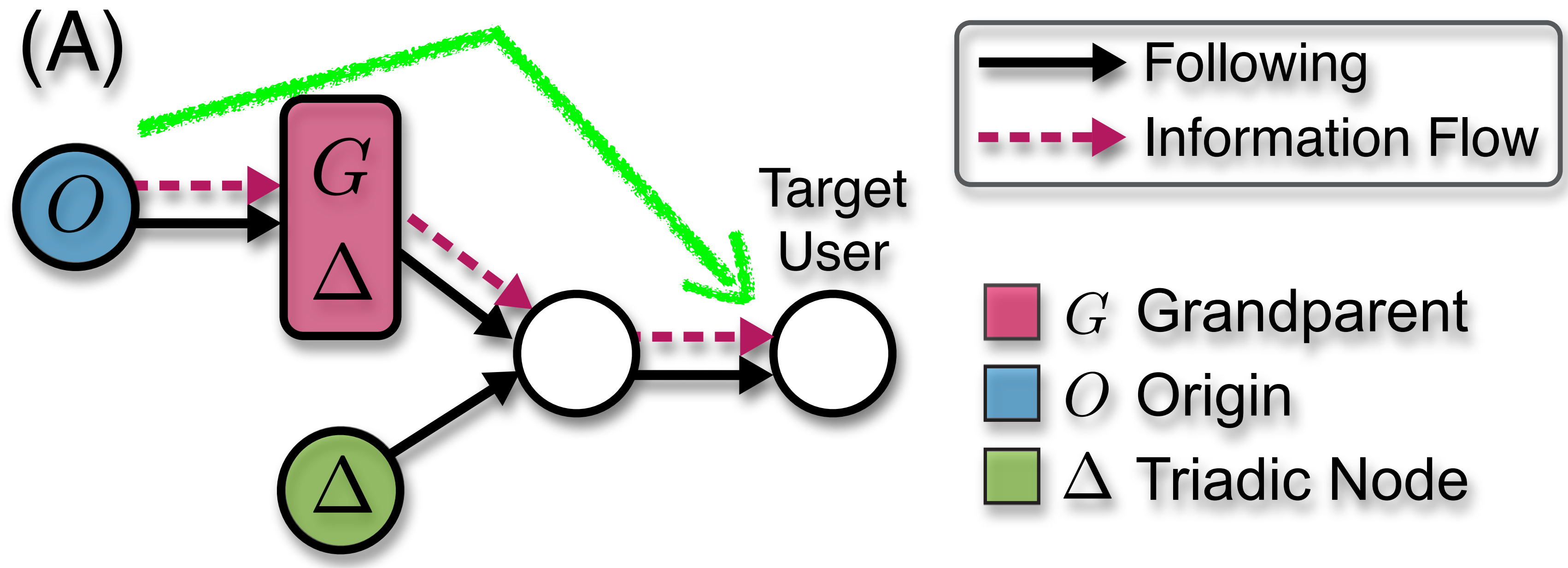


G
Δ
O
Δ
Target User
Following
Information Flow
G Grandparent
O Origin
Δ Triadic Node

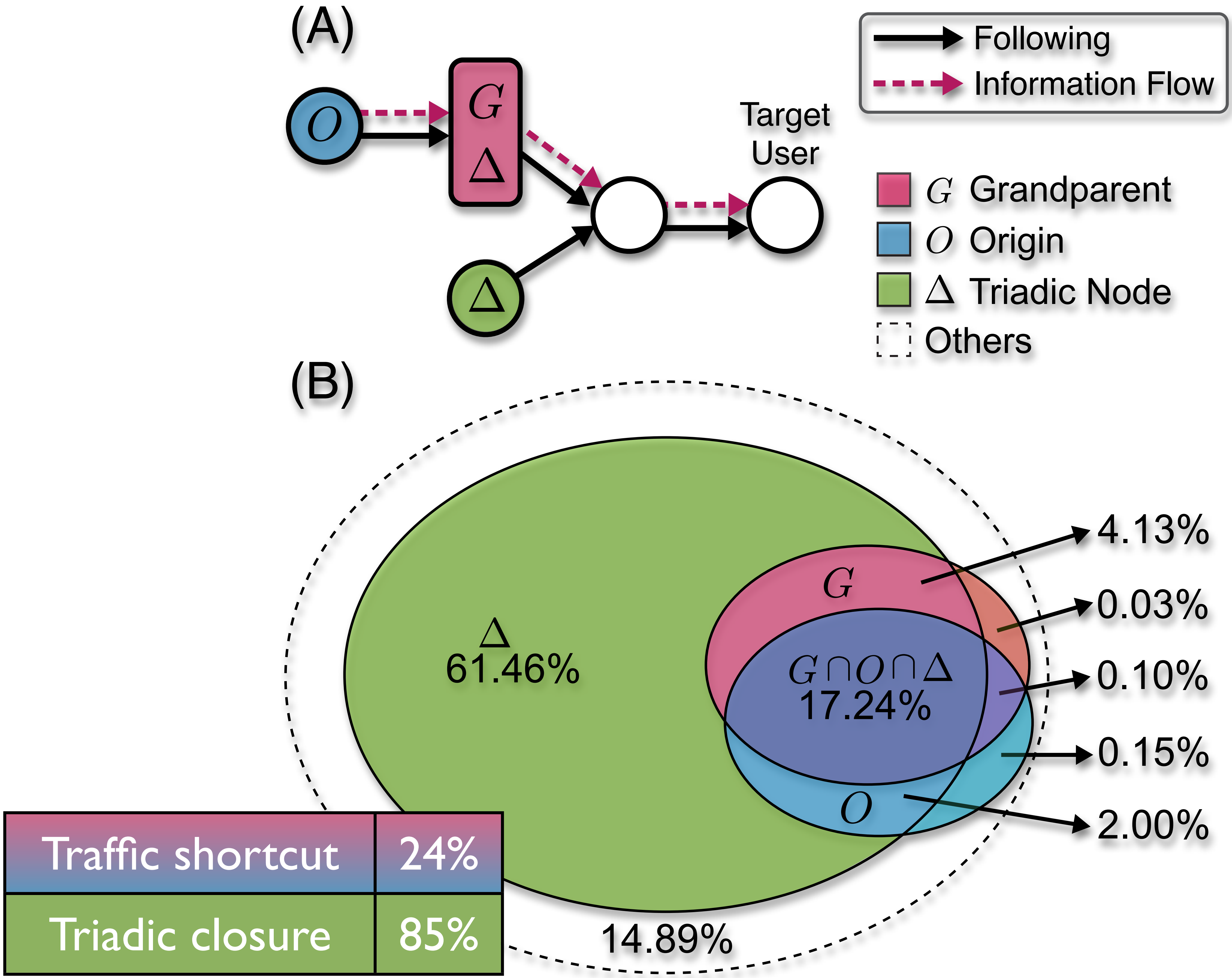


(A)
O
G
Δ
Δ
Target
User
Following
Information Flow
G Grandparent
O Origin
Δ Triadic Node


(A)
G
Δ
O
Δ
Target User
Following
Information Flow
G Grandparent
O Origin
Δ Triadic Node



(A)
O
G
Δ
Δ
Target User
Following
Information Flow
G Grandparent
O Origin
Δ Triadic Node


(A)
O
G
Δ
Δ
Target User
Following
Information Flow
G Grandparent
O Origin
Δ Triadic Node
Others
(B)
Δ
61.46%
G
G∩O∩Δ
17.24%
O
4.13%
0.03%
0.10%
0.15%
2.00%
14.89%
Traffic shortcut
24%
Triadic closure
85%

Could this happen by chance?

$$z = \frac{S - E}{\sigma} \in \aleph(0,1)$$

Could this happen by chance?

Actual number of links of that type in the data

Expected number of links of a certain type according to the null hypothesis (by chance). E.g., links to grandparents:

$$E_G = \sum_{\ell}^{L} \frac{N_G(\ell)}{N(\ell) - k(\ell) - 1}$$

$$z = \frac{S - E}{\sigma} \in \aleph(0,1)$$

Could this happen by chance?

Actual number of links of that type in the data

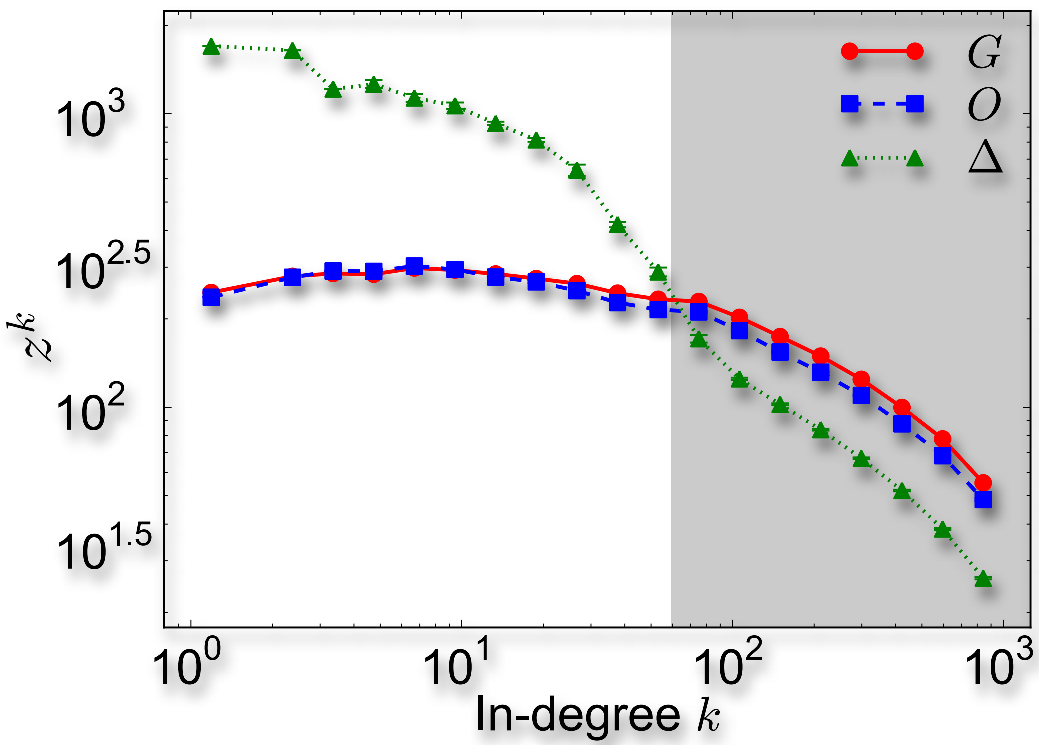
Expected number of links of a certain type according to the null hypothesis (by chance). E.g., links to grandparents:

$$E_G = \sum_{\ell}^{L} \frac{N_G(\ell)}{N(\ell) - k(\ell) - 1}$$

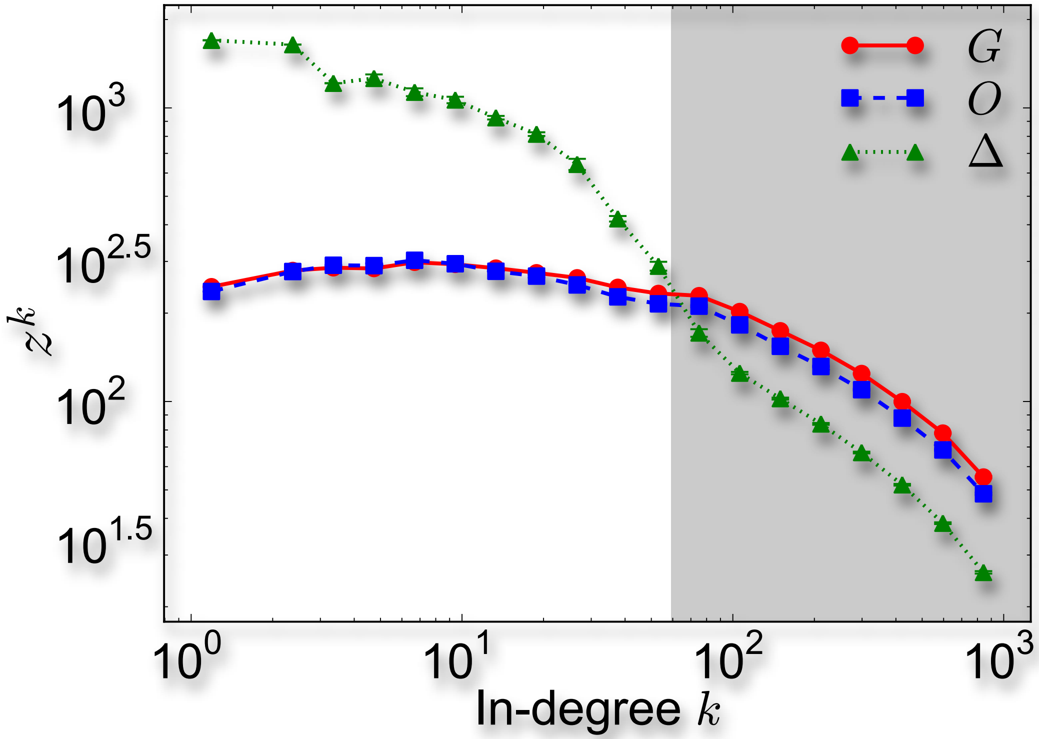
$$z = \frac{S - E}{\sigma} \in \aleph(0,1)$$

z_{Δ}, z_G, z_O very large ⇒ reject null hypothesis:

links are <u>not</u> created randomly

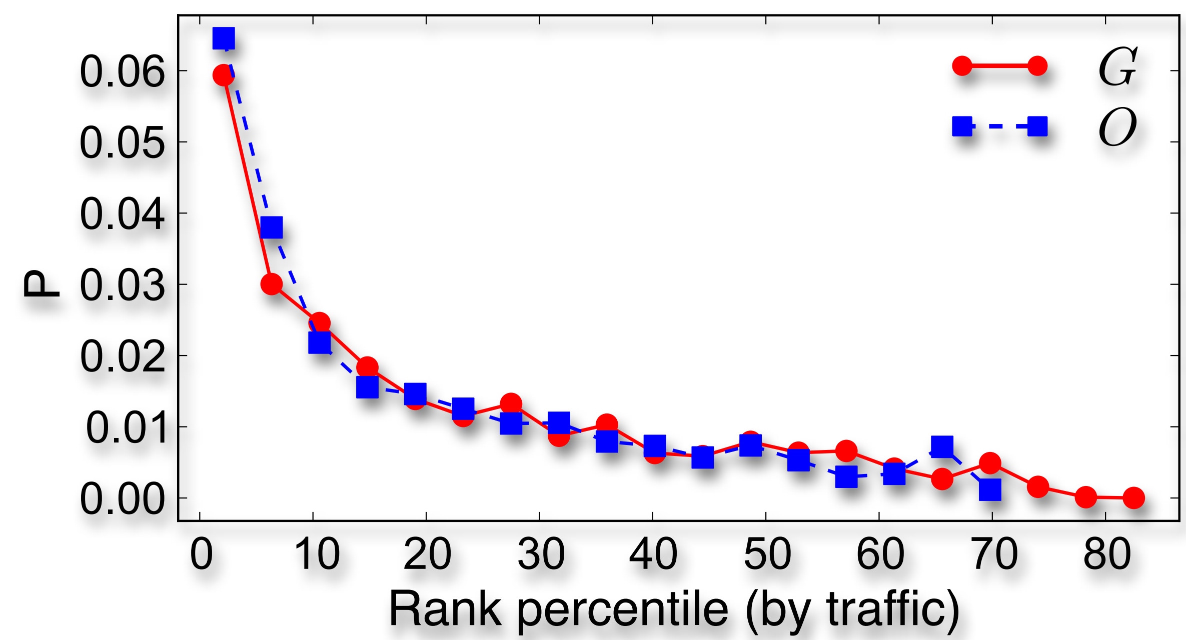


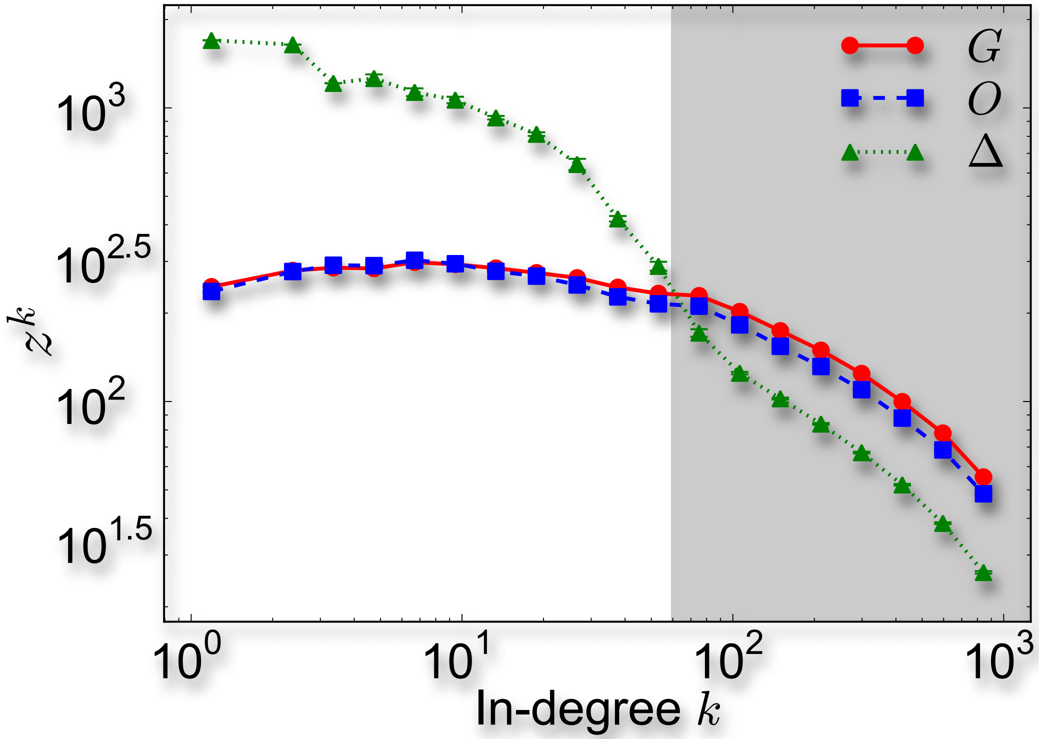
Preference for traffic-based shortcuts as users become more active



Preference for traffic-based shortcuts as users become more active

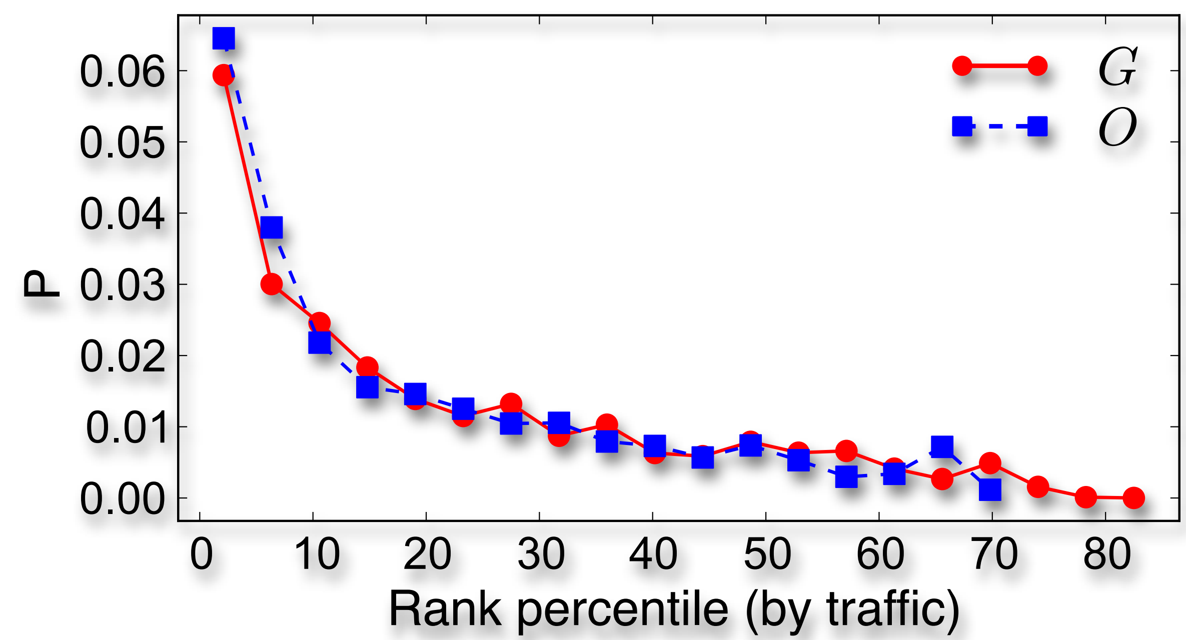
The more posts we see from someone, the more we are likely to follow them

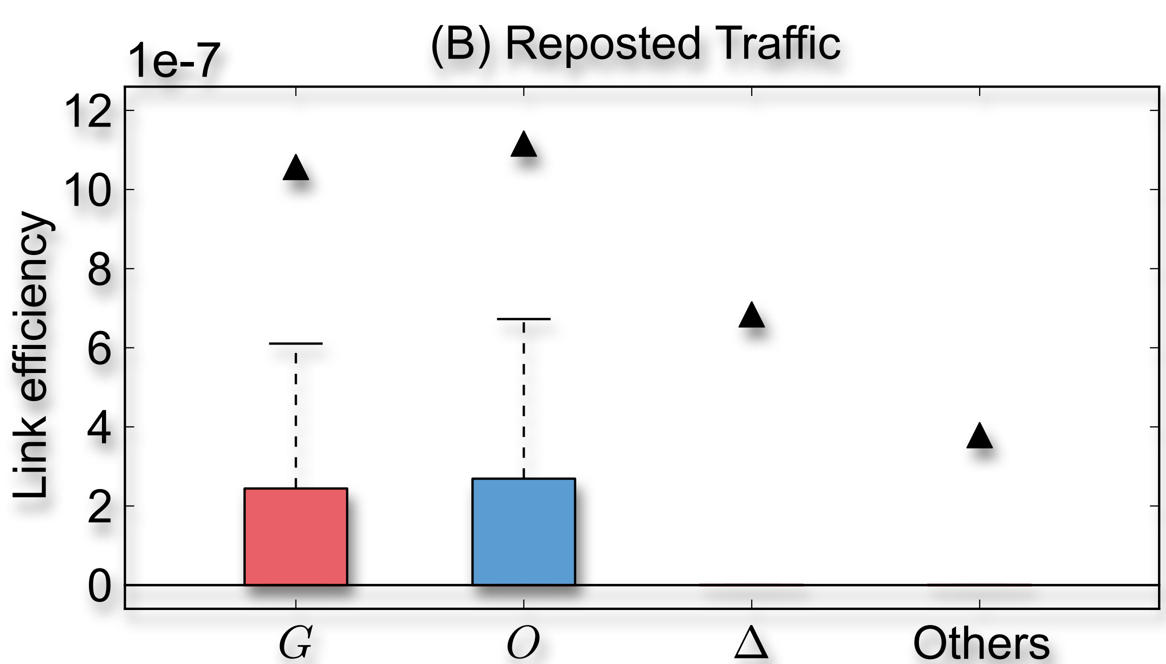




Preference for traffic-based shortcuts as users become more active

The more posts we see from someone, the more we are likely to follow them





Shortcuts are more efficient at carrying messages we see and report

Maximum Likelihood Estimation

Maximum Likelihood Estimation

$\forall$ link ℓ, compute:

$f(\ell \mid \Gamma, \Theta)$ = likelihood of the target being followed by the creator according to a particular strategy Γ, given the network configuration Θ at the time when ℓ is created.

Maximum Likelihood Estimation

$\forall$ link ℓ, compute:

$f(\ell \mid \Gamma, \Theta)$ = likelihood of the target being followed by the creator according to a particular strategy Γ, given the network configuration Θ at the time when ℓ is created.

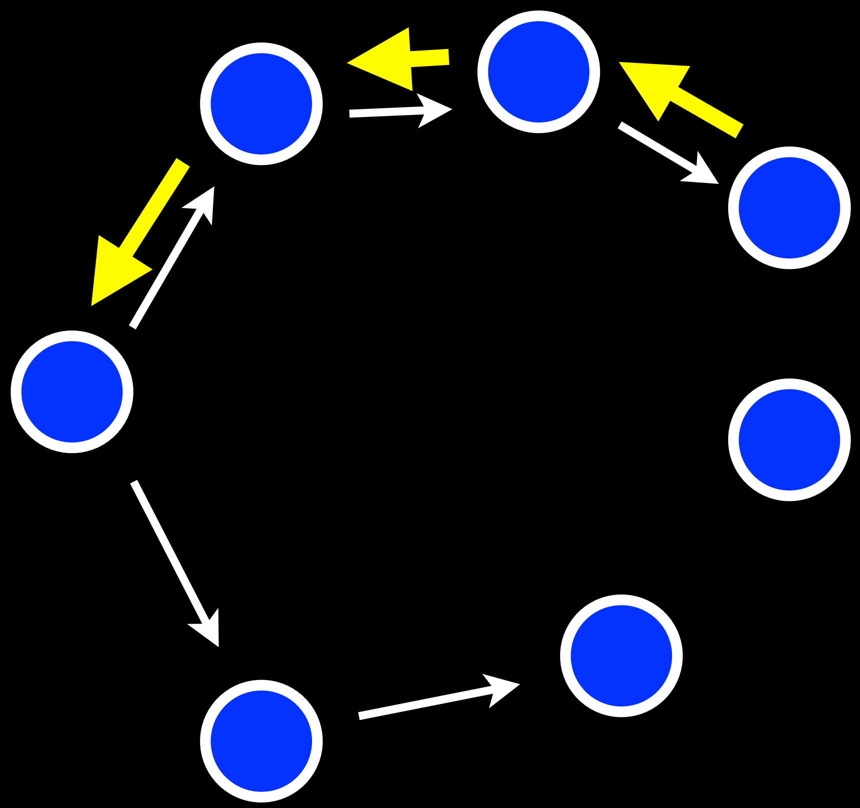
Single strategy

Combined strategy

Individual strategy

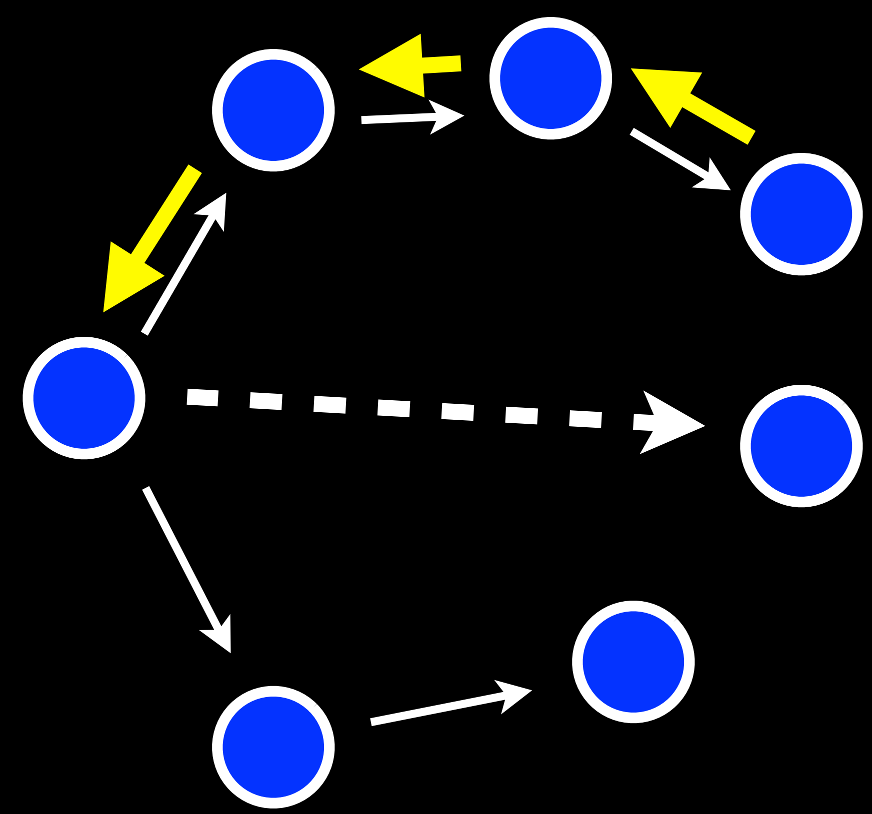
MLE single strategies

- Random
- Triadic closure (**Δ**)
- Grandparent (**G**)
- Origin (**O**)
- Traffic shortcut (**G** ∪ **O**)



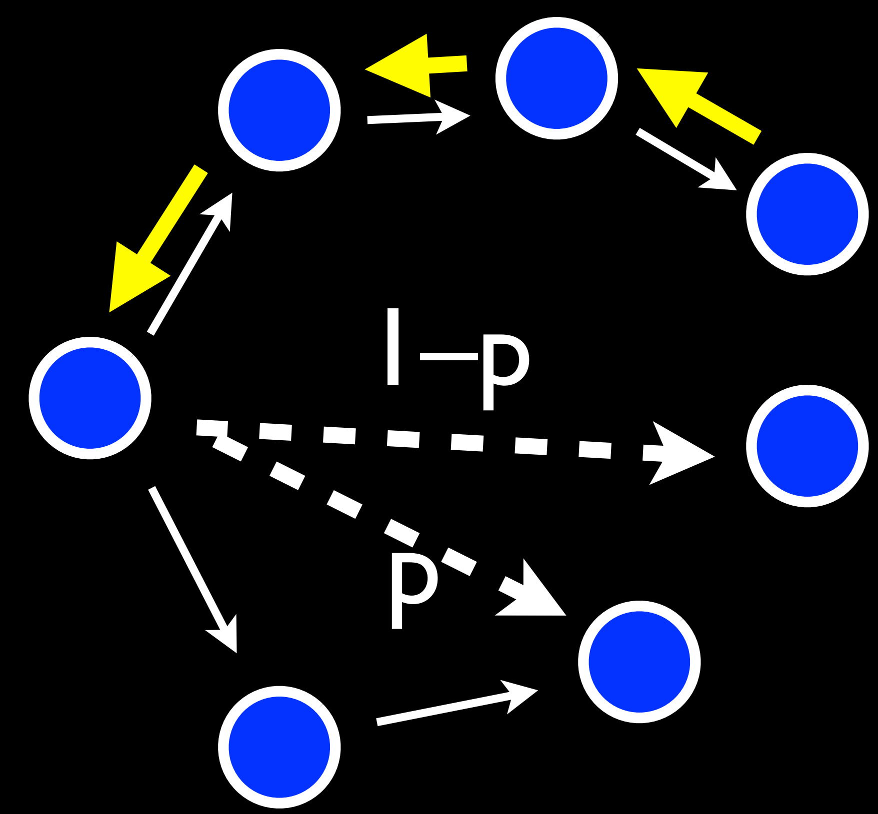
MLE single strategies

- Random
- Triadic closure (**Δ**)
- Grandparent (**G**)
- Origin (**O**)
- Traffic shortcut (**G** ∪ **O**)



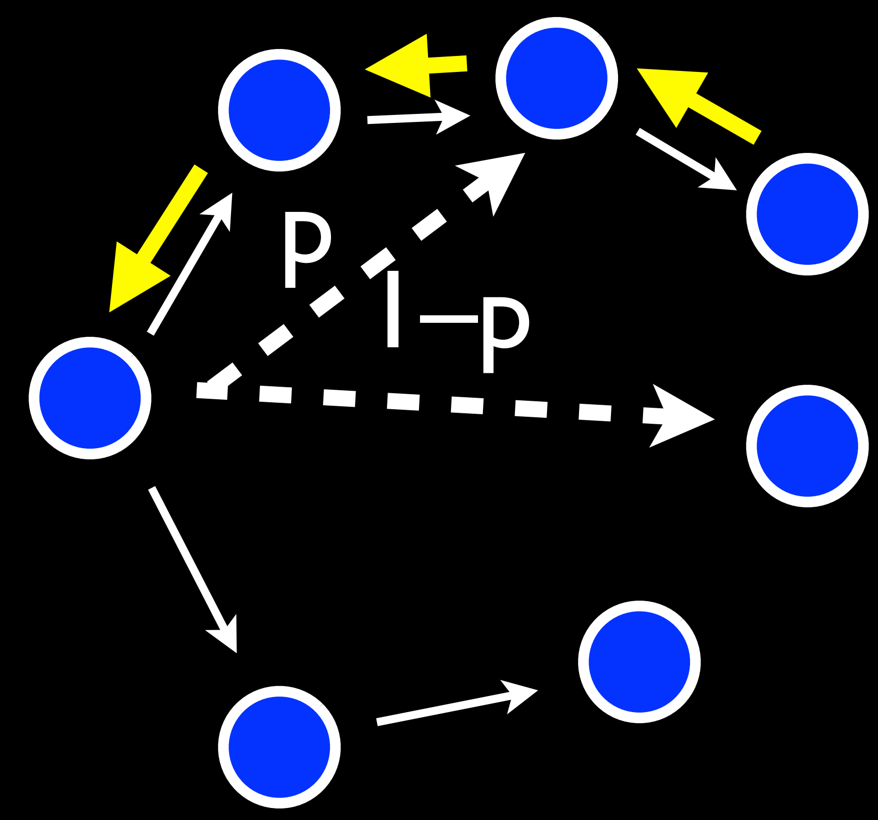
MLE single strategies

- Random
- Triadic closure (Δ)



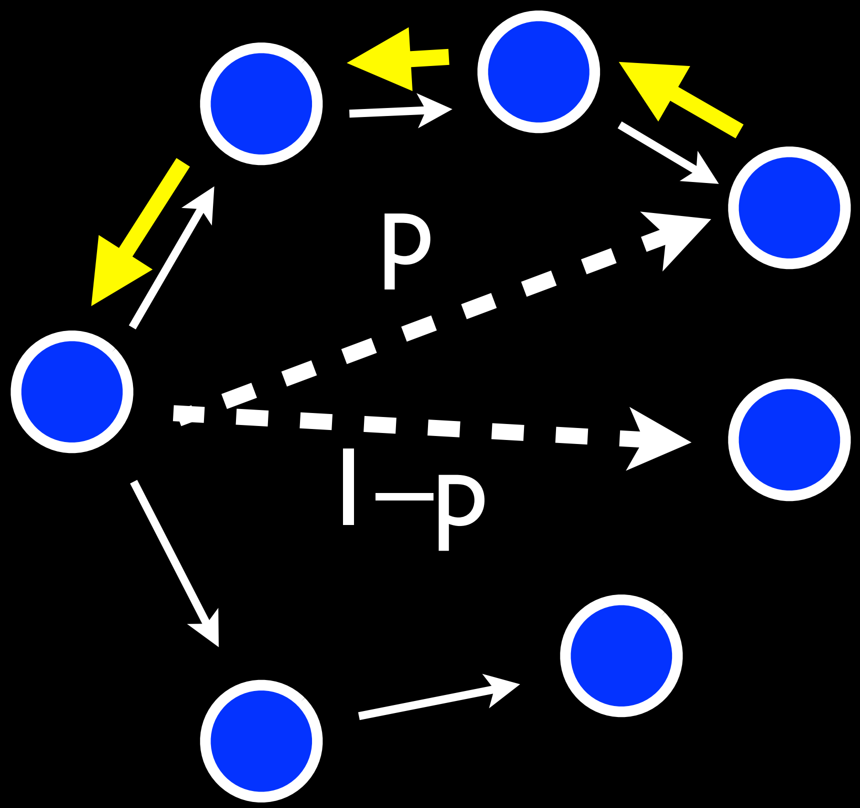
MLE single strategies

- Random
- Triadic closure (**Δ**)
- Grandparent (**G**)



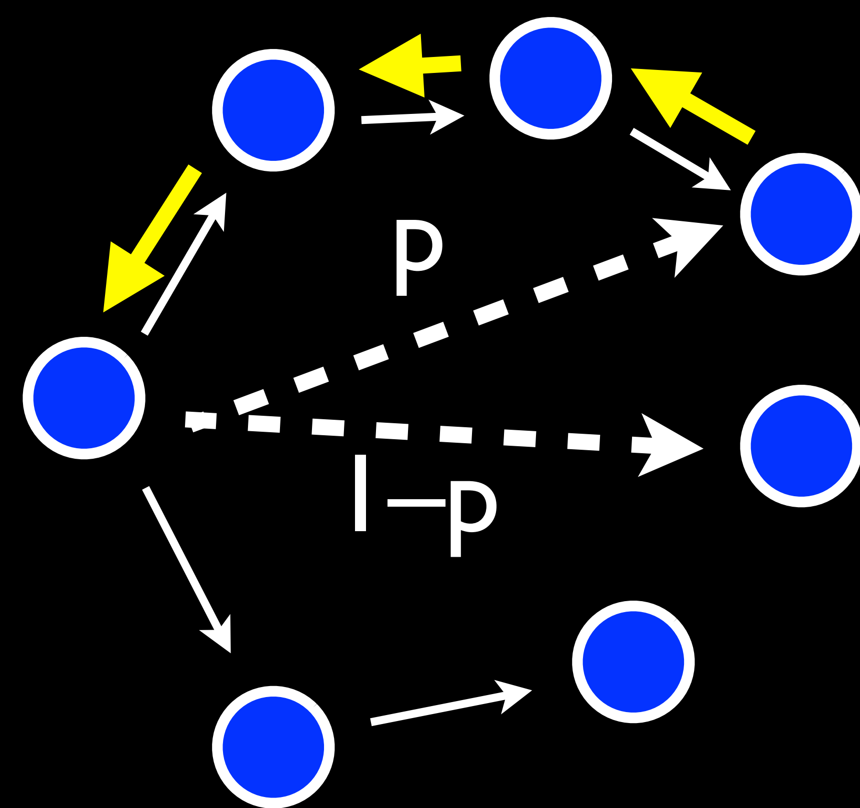
MLE single strategies

- Random
- Triadic closure (**Δ**)
- Grandparent (**G**)
- Origin (**O**)
- Traffic shortcut (**G** ∪ **O**)



MLE single strategies

- Random
- Triadic closure (**Δ**)
- Grandparent (**G**)
- Origin (**O**)
- Traffic shortcut (**G** ∪ **O**)

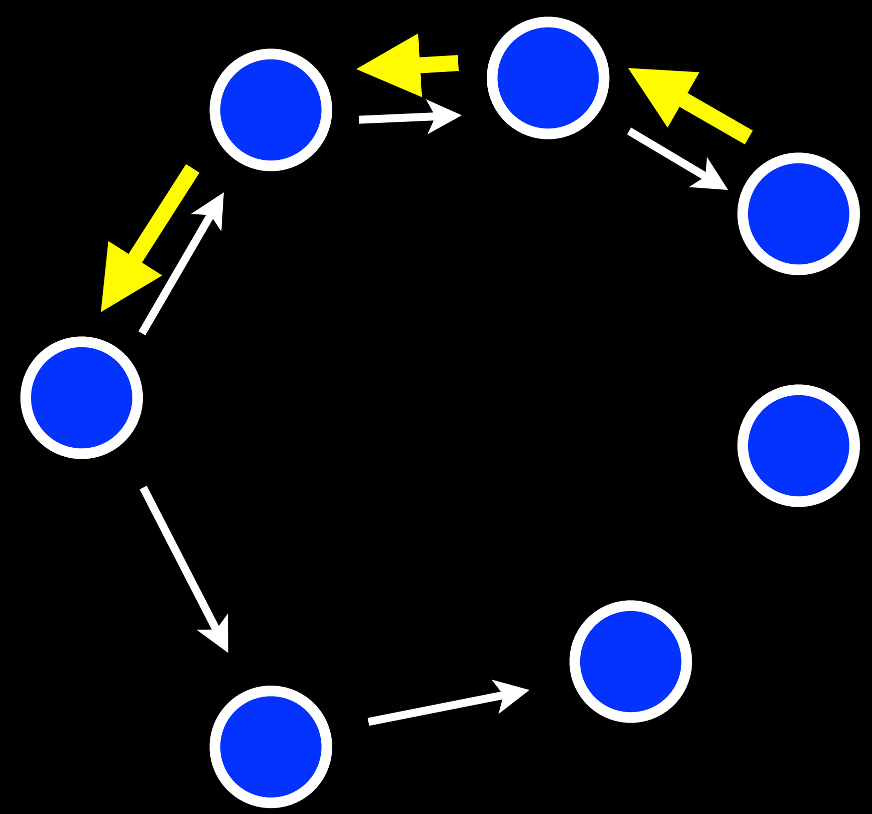


Example:

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

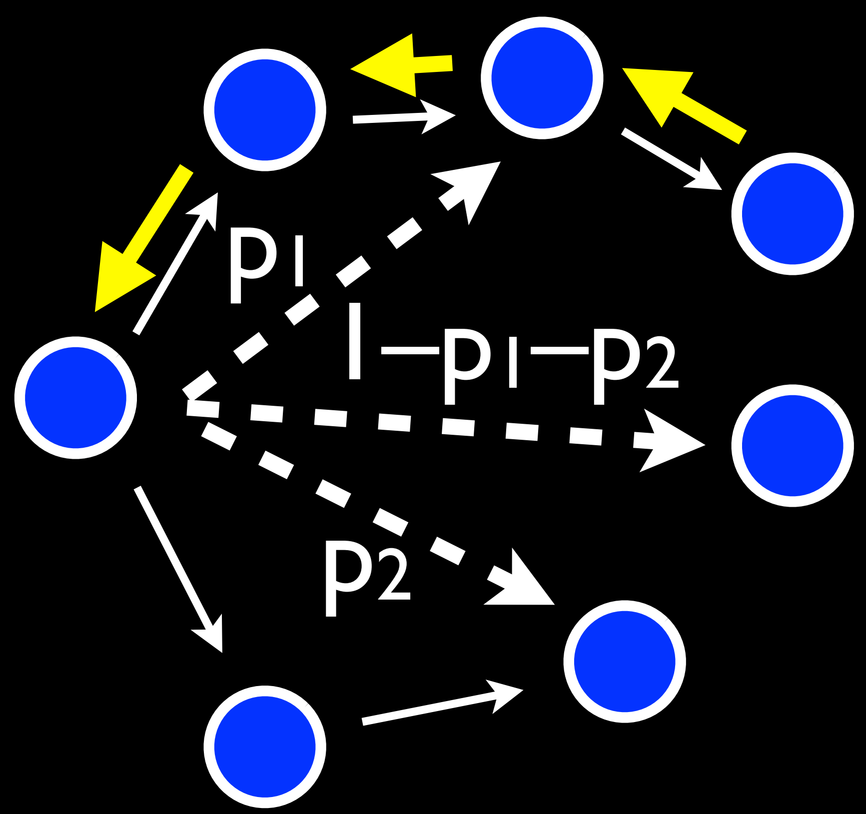
MLE combined strategies

- Grandparent or triadic closure (**G + Δ**)
- Origin or triadic closure (**O + Δ**)
- Traffic shortcut or triadic closure (**G** ∪ **O + Δ**)



MLE combined strategies

- Grandparent or triadic closure (**G + Δ**)
- Origin or triadic closure (**O + Δ**)
- Traffic shortcut or triadic closure (**G ∪ O + Δ**)

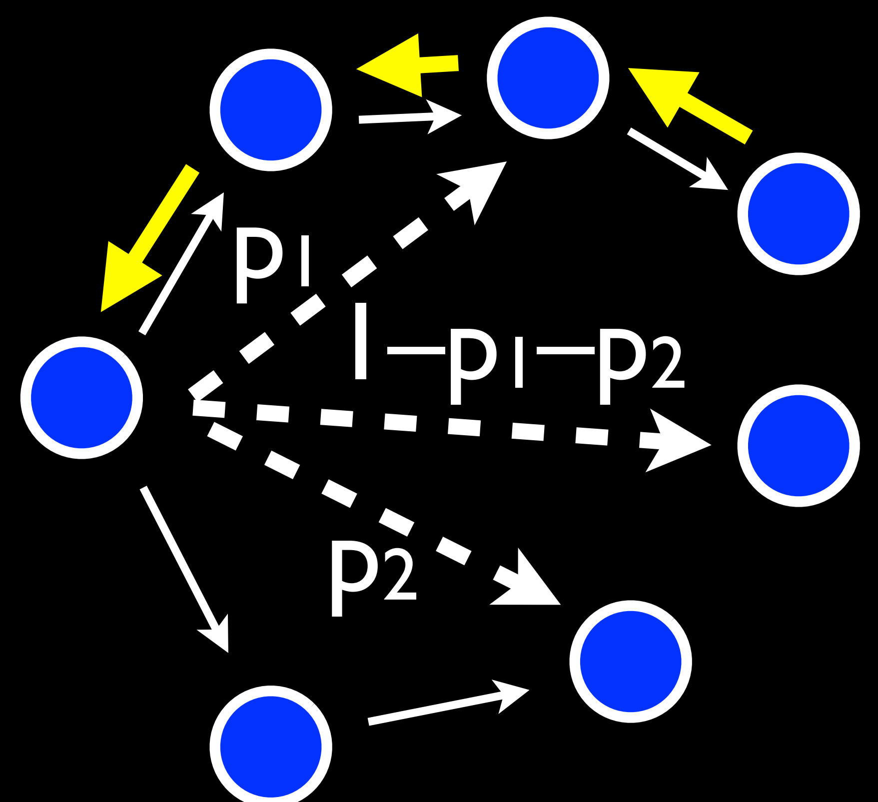


MLE combined strategies

Example:

$$
\begin{aligned}
&\log \mathcal{L}_{G+\Delta}(p_1, p_2) \\
&= \log \prod_{\ell=1}^{L} [p_1 f(\ell|G,\Theta) + p_2 f(\ell|\Delta,\Theta) \\
&\quad + (1 - p_1 - p_2) f(\ell|\mathrm{Rand},\Theta)] \\
&= \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=0}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=0}} \log \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}.
\end{aligned}
$$

- Grandparent or triadic closure (**G + Δ**)
- Origin or triadic closure (**O + Δ**)
- Traffic shortcut or triadic closure (**G ∪ O + Δ**)

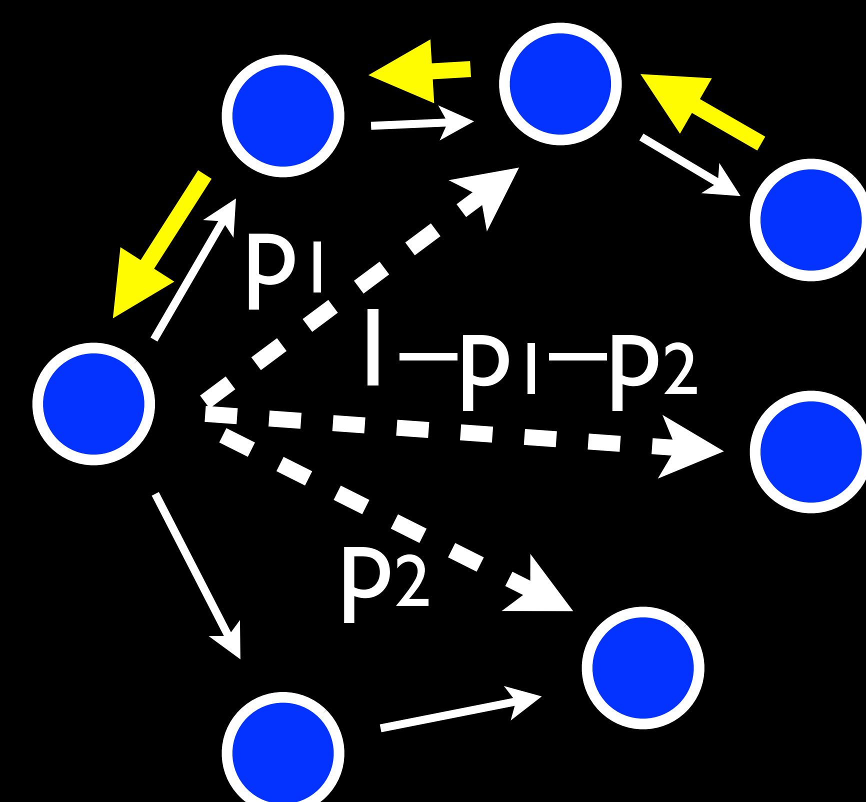


MLE combined strategies

Example:

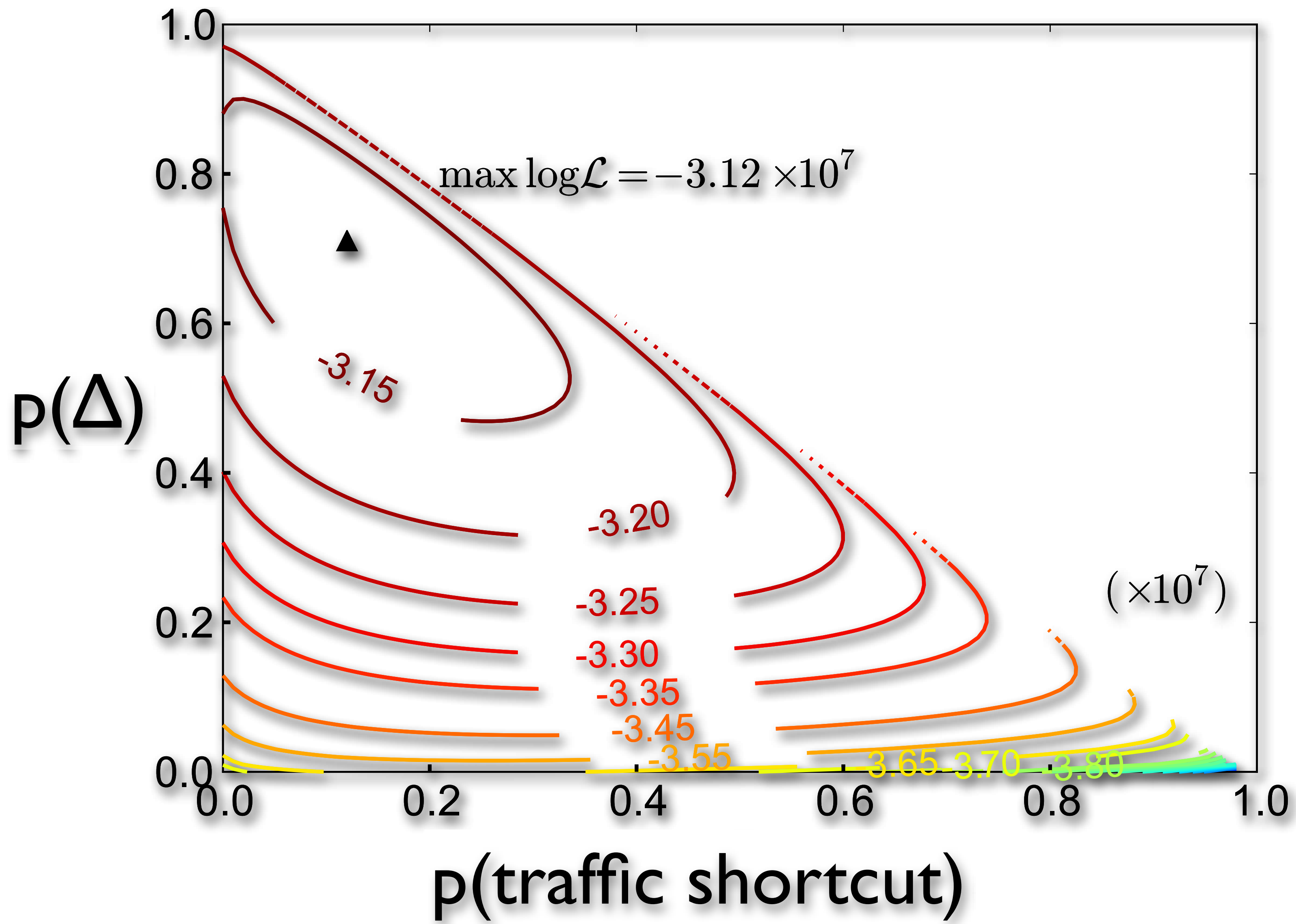
$$
\begin{aligned}
&\log \mathcal{L}_{G+\Delta}(p_1, p_2) \\
&= \log \prod_{\ell=1}^{L} [p_1 f(\ell|G,\Theta) + p_2 f(\ell|\Delta,\Theta) \\
&\quad + (1-p_1-p_2) f(\ell|\mathrm{Rand},\Theta)] \\
&= \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=1 \\ \mathbf{1}_\Delta(\ell)=0}} \log\left(\frac{p_1}{N_G(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=1}} \log\left(\frac{p_2}{N_\Delta(\ell)} + \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}\right) \\
&\quad + \sum_{\substack{\mathbf{1}_G(\ell)=0 \\ \mathbf{1}_\Delta(\ell)=0}} \log \frac{1-p_1-p_2}{N(\ell)-k(\ell)-1}.
\end{aligned}
$$

- Grandparent or triadic closure (**G + Δ**)
- Origin or triadic closure (**O + Δ**)
- Traffic shortcut or triadic closure (**G ∪ O + Δ**)



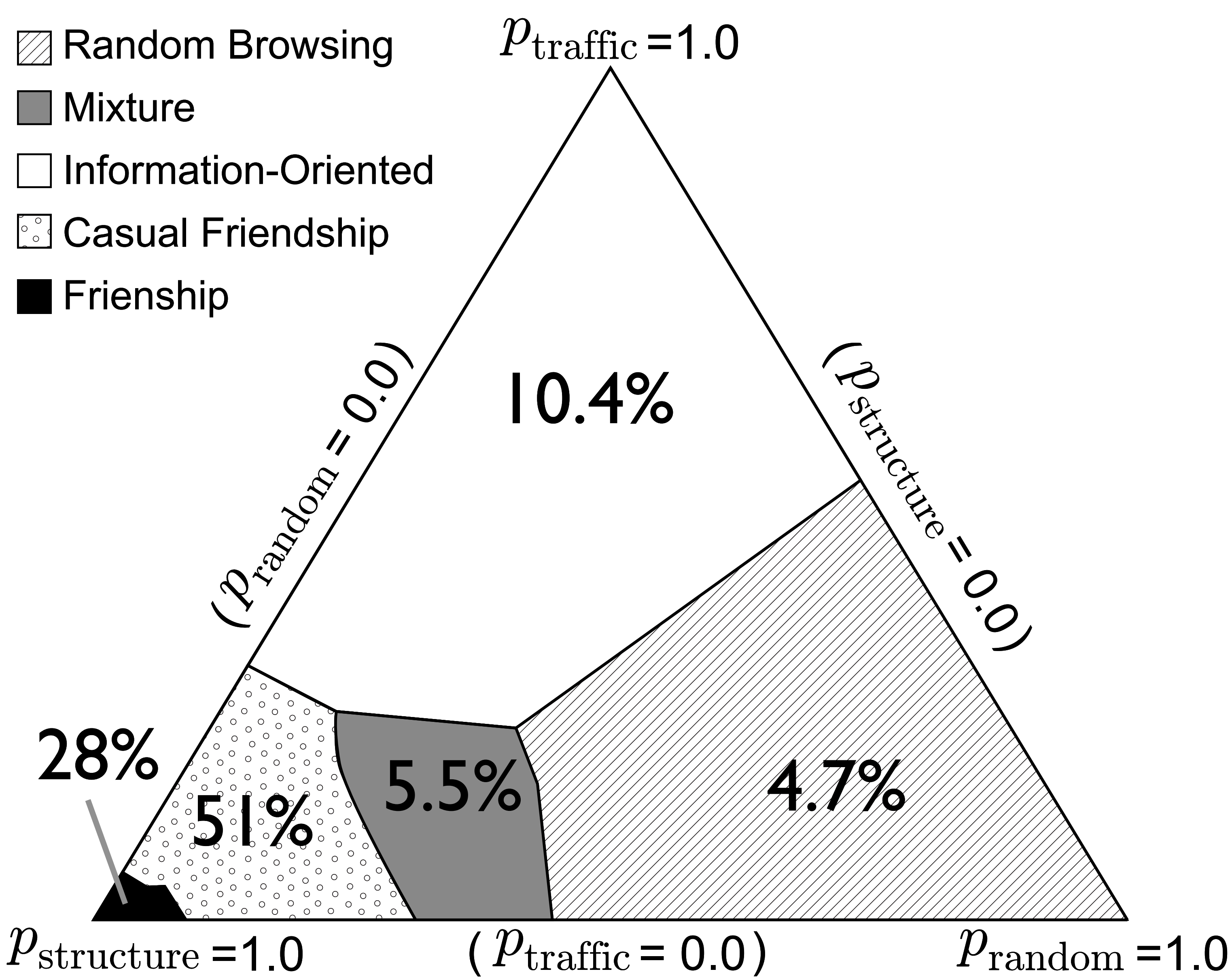
Maximum Likelihood

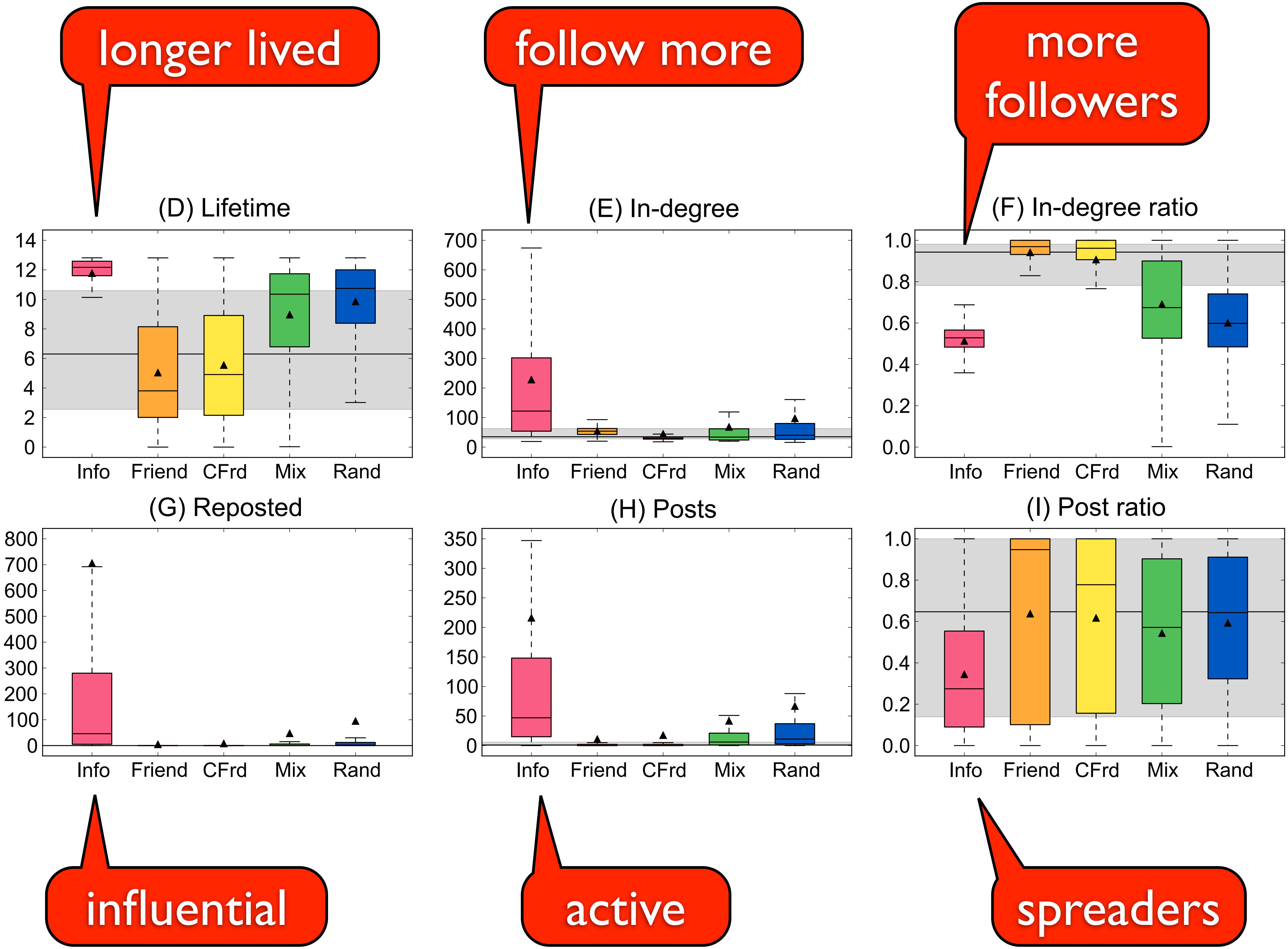
(**G** ∪ **O** + **Δ**)



Maximum Likelihood

($\mathbf{G} \cup \mathbf{O} + \boldsymbol{\Delta}$)




longer lived
follow more
more followers
(D) Lifetime
(E) In-degree
(F) In-degree ratio
Info
Friend
CFrd
Mix
Rand
(G) Reposted
(H) Posts
(I) Post ratio
influential
active
spreaders

As users become more active, more popular, and more influential, they make the network more "efficient" by shortening the distance between producers and consumers of information.

Lilian Weng[1], Jacob Ratkiewicz[2], Nicola Perra[3], Bruno Gonçalves[4], Carlos Castillo[5], Francesco Bonchi[6], Rossano Schifanella[7], Filippo Menczer[1], Alessandro Flammini[1]

[1]School of Informatics and Computing, Indiana University Bloomington, USA [2]Google Inc.

[3]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, USA

[4]Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France

[5]Qatar Computing Research Institute [6]Yahoo! Research Barcelona

[7]Department of Computer Science, University of Torino, Italy

Papers: cnets.indiana.edu/groups/nan/truthy

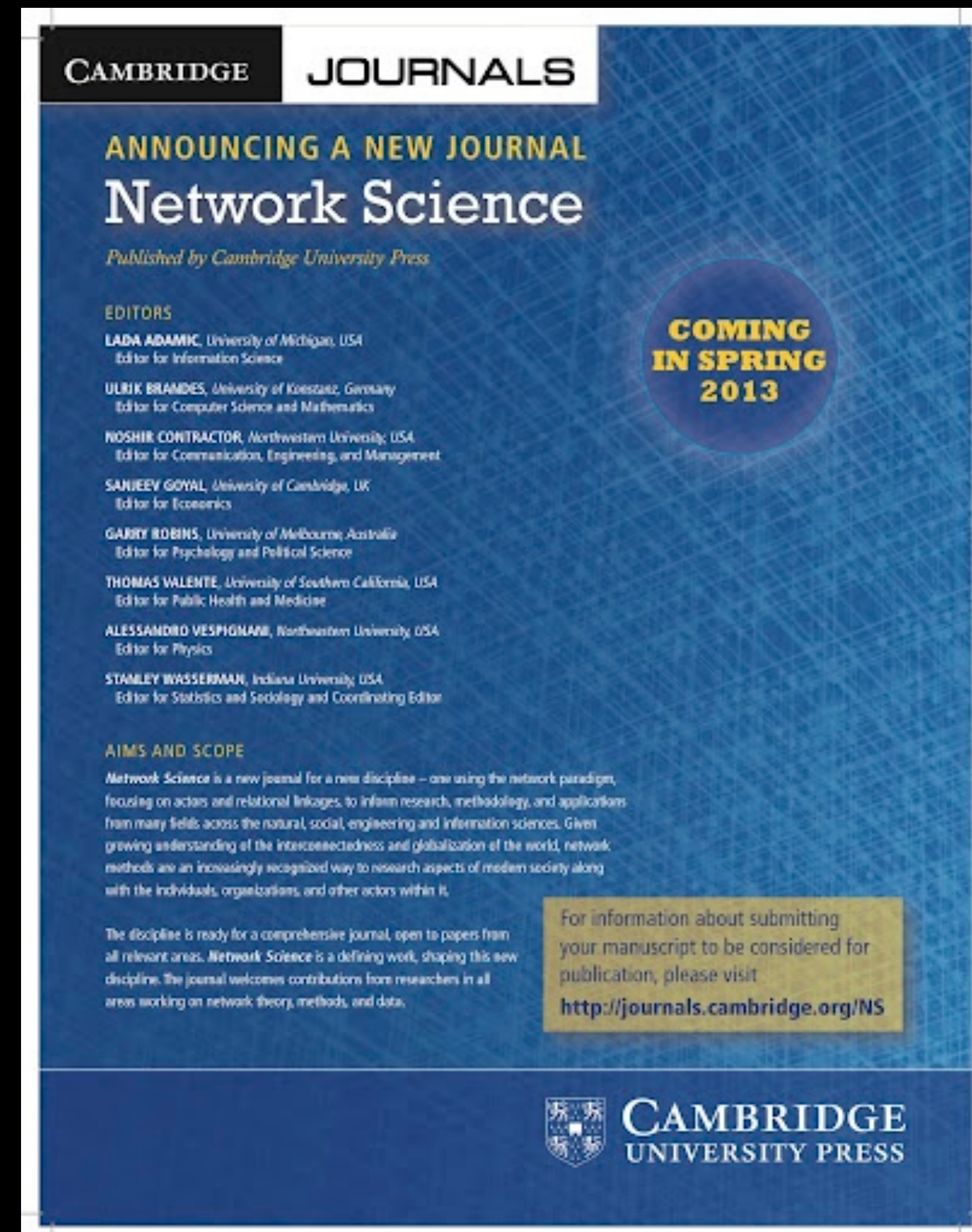
PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

NSF

JAMES S. McDONNELL FOUNDATION

DARPA


SpringerOpen
EPJ Data Science
Group
Profile
Connection
Social
Community
Friends
WWW
online
web
Virtual
computer
Network
Context
internet
Wall
Viral
Search
visitors
Like
messages
account
media
edp sciences
Springer


CAMBRIDGE JOURNALS
ANNOUNCING A NEW JOURNAL
Network Science
Published by Cambridge University Press
COMING IN SPRING 2013
EDITORS
LADA ADAMIC, University of Michigan, USA
Editor for Information Science
ULRIK BRANDES, University of Konstanz, Germany
Editor for Computer Science and Mathematics
NOSHIR CONTRACTOR, Northwestern University, USA
Editor for Communication, Engineering, and Management
SANJEEV GOYAL, University of Cambridge, UK
Editor for Economics
GARRY ROBINS, University of Melbourne, Australia
Editor for Psychology and Political Science
THOMAS VALENTE, University of Southern California, USA
Editor for Public Health and Medicine
ALESSANDRO VESPIGNANI, Northeastern University, USA
Editor for Physics
STANLEY WASSERMAN, Indiana University, USA
Editor for Statistics and Sociology and Coordinating Editor
AIMS AND SCOPE
Network Science is a new journal for a new discipline – one using the network paradigm, focusing on actors and relational linkages, to inform research, methodology, and applications from many fields across the natural, social, engineering and information sciences. Given growing understanding of the interconnectedness and globalization of the world, network methods are an increasingly recognized way to research aspects of modern society along with the individuals, organizations, and other actors within it.
The discipline is ready for a comprehensive journal, open to papers from all relevant areas. Network Science is a defining work, shaping this new discipline. The journal welcomes contributions from researchers in all areas working on network theory, methods, and data.
For information about submitting your manuscript to be considered for publication, please visit
http://journals.cambridge.org/NS
CAMBRIDGE UNIVERSITY PRESS