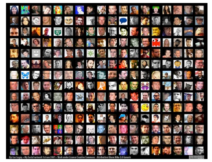


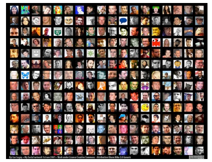
Social and Technological Network Analysis

Lecture 7: Information Cascades

(thanks to S. Gonzalez Bailon for some of the slides)

Dr. Cecilia Mascolo

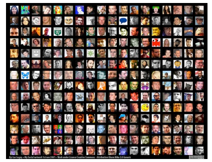
UNIVERSITY OF CAMBRIDGE



In This Lecture

- In this lecture we introduce the concept of "cascades" of information in networks and show examples and trade-offs for these to happen.
- We will discuss practical studies and applications

Decision Making and Behaviour Influence

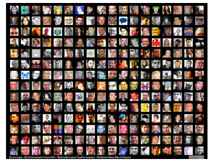


- How is new behaviour adopted?
- How does technology usage spread?
- People influence on ideas?

- The social network plays an important role in the decision making process
  - We study how.

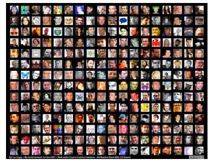
UNIVERSITY OF CAMBRIDGE

Early studies on Influence



- Ryan and Gross (1943) on adoption of hybrid corn in Iowa
  - Farmers learned of the corn from salesmen but were convinced on adoption by experience of neighbours in the community.
- Coleman, Katz and Menzel (1966) on adoption of tetracycline in US
  - Map of social connections among doctors.
  - Early adopter had higher socio-economical status and travelled more widely (also in corn case).
  - Decision on adoption was made in the context of the social structure (observing neighbours, friends and colleagues).

UNIVERSITY OF CAMBRIDGE



Model of Diffusion

- Nodes v and w and behaviours A and B
  - If both v and w adopt A, they each get payoff a>0
  - If both adopt B, they each get a payoff b>0
  - If they adopt opposite behaviour they both get a payoff of 0

		w: A	w: B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

UNIVERSITY OF CAMBRIDGE

Network Implications

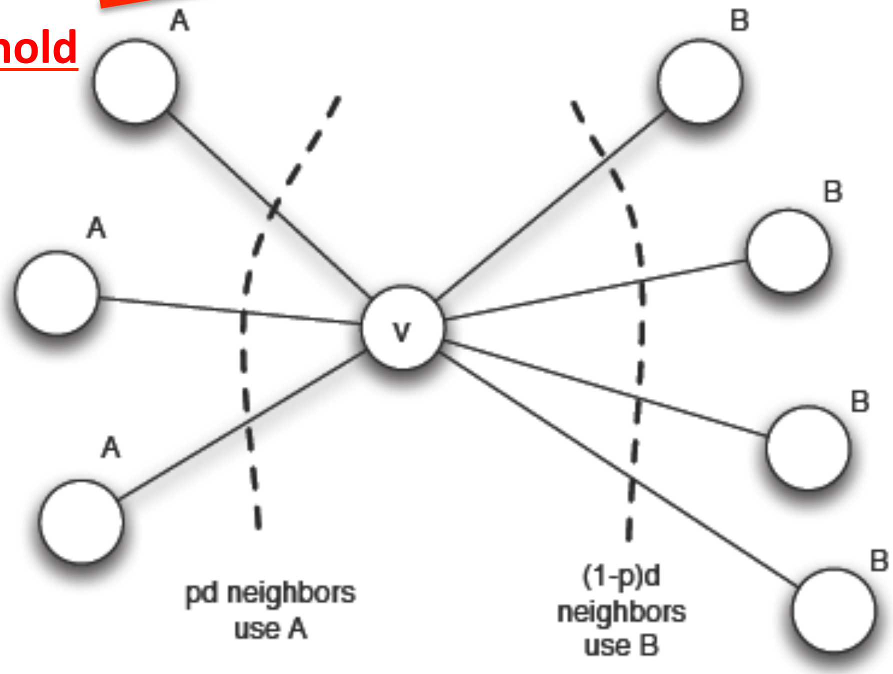
- p fraction of neighbours choose A
- (1-p) choose B
- d neighbours then: pd choose A and(1-p)d choose B

A better choice if :

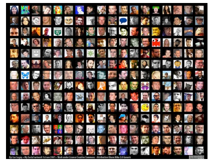
$$pda \geq (1-p)db$$

$$p \geq \frac{b}{a+b}$$

threshold

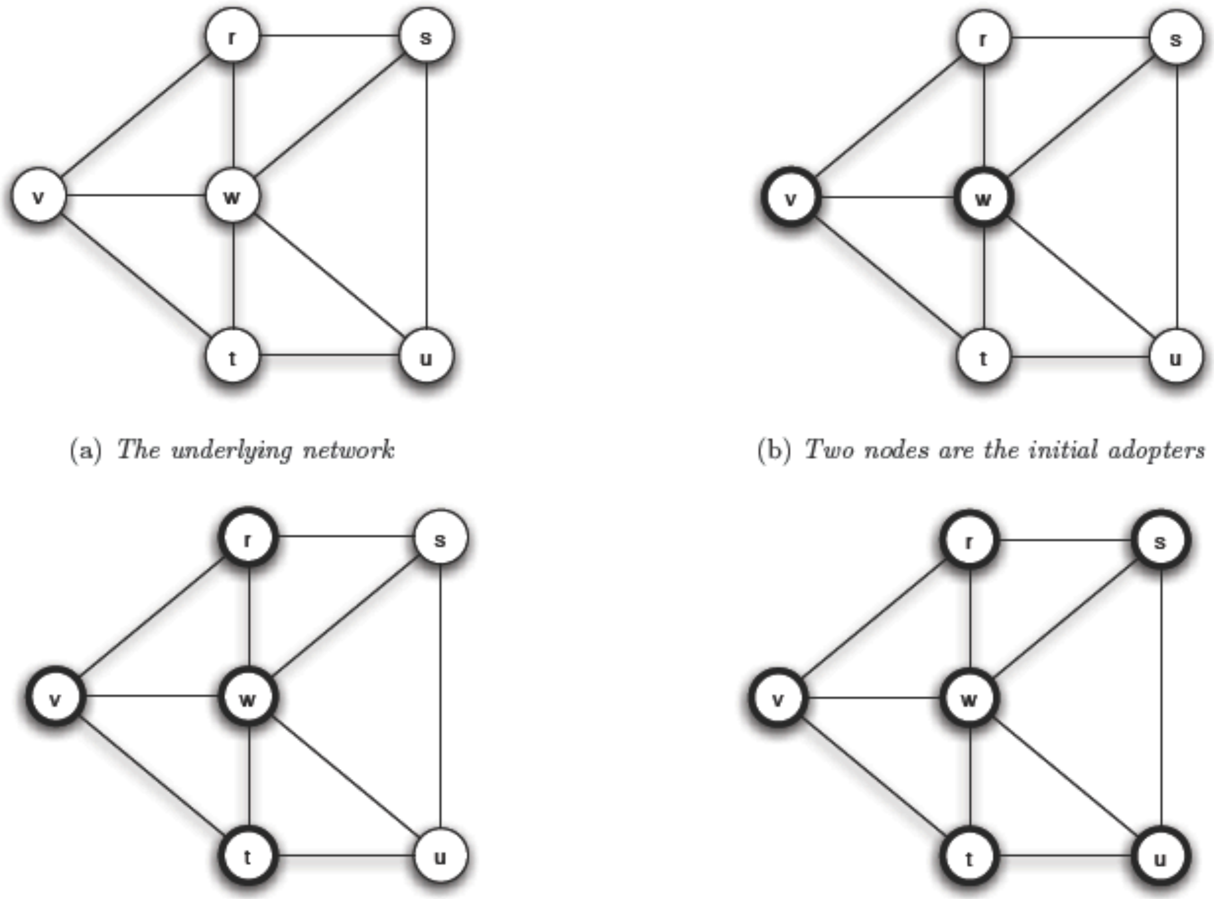


UNIVERSITY OF CAMBRIDGE



Larger Horizon

- A is new behaviour
- a=3, b=2
- b/a+b=2/5



(a) *The underlying network*

(b) *Two nodes are the initial adopters*

(c) *After one step, two more nodes have adopted*

(d) *After a second step, everyone has adopted*

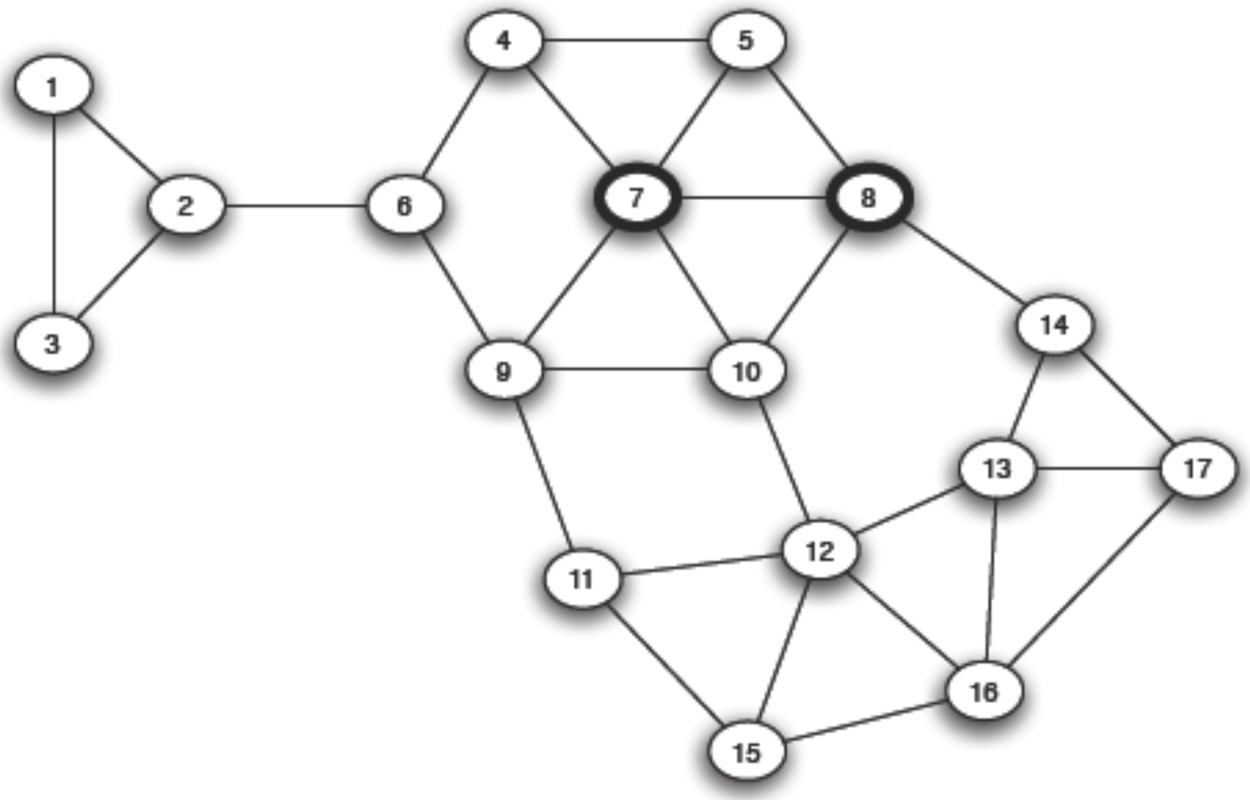
Light circles=B
Dark circles=A

UNIVERSITY OF CAMBRIDGE

Example explained

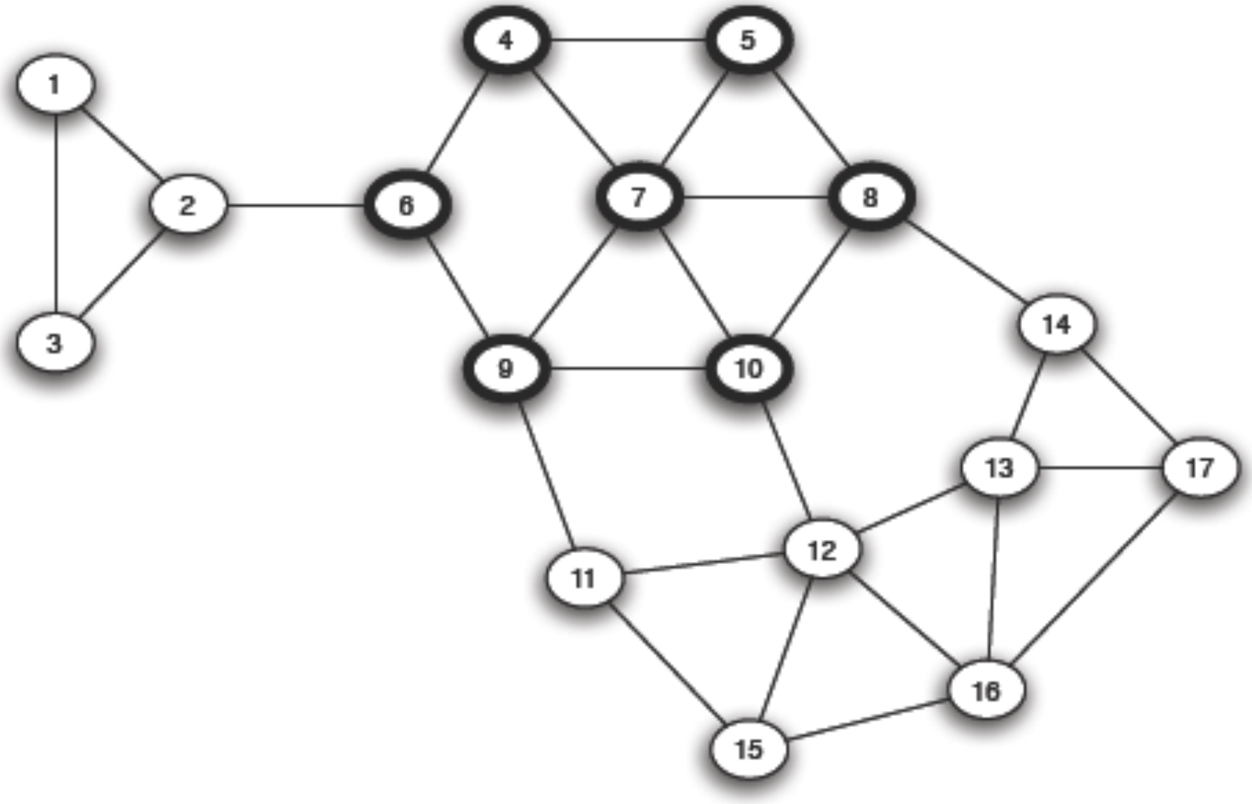
- 1st step: only v and w adopt A
- 2nd step: nodes r and t switch to A. 2/3 > 2/5 of neighbours choose A. u does not switch: 1/3<2/5 of neighbours chose A
- 3rd step: s and u switch to A

Chain Reactions



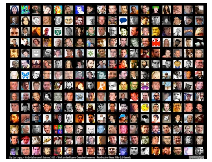
UNIVERSITY OF CAMBRIDGE

Cascade Stops!



UNIVERSITY OF CAMBRIDGE

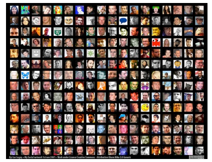
Cascades



- In some cases initial adoption by some nodes generate a **complete cascade** [for a specific threshold]

- Note that changing the threshold would change the behaviour in previous example
  - Threshold of 1/3 would generate a complete cascade

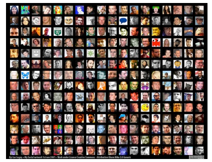
Viral Marketing



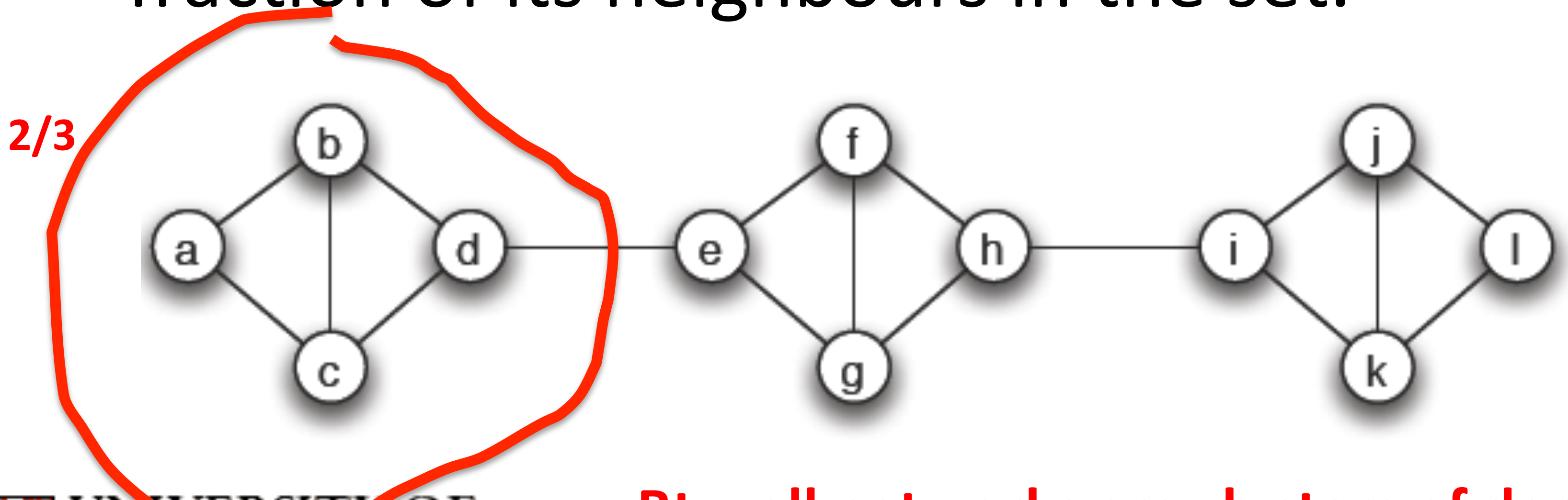
- How to penetrate new areas of the network
- Dissemination does not depend only on the network structure but also on this threshold!
  - Change the payoff! I.e., change the quality of the product [make a product slightly more attractive].
- When threshold cannot be changed
  - Convince key network nodes to switch (e.g nodes 12/13 good, but nodes 11 and 14 bad).

UNIVERSITY OF CAMBRIDGE

What Makes Cascades Stop?



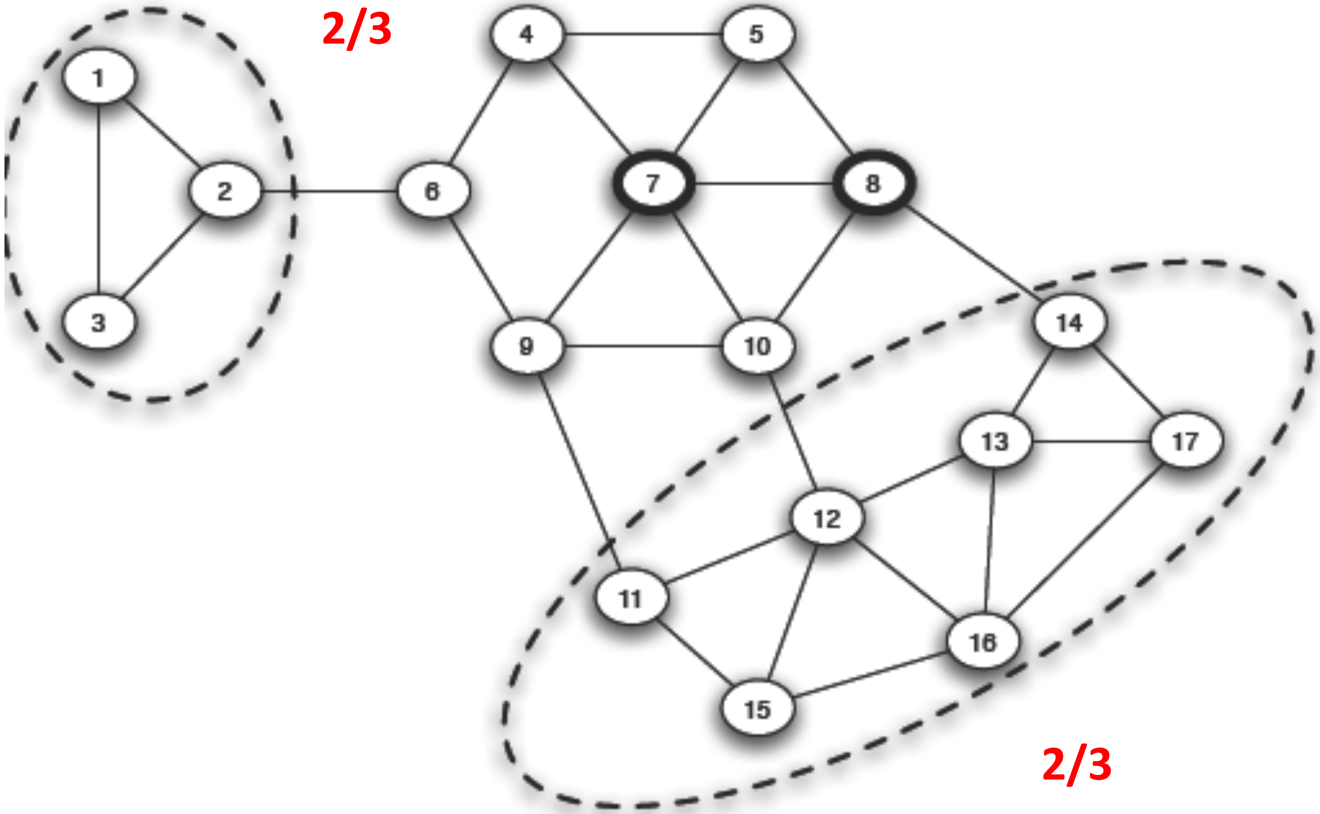
- Tightly knit communities sometimes cannot be penetrated.
- A cluster of density p is a set of nodes such that each node in the set has at least a p fraction of its neighbours in the set.



UNIVERSITY OF CAMBRIDGE

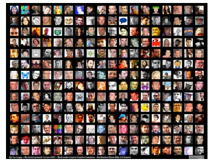
Btw all networks are clusters of density 1

Clusters as Obstacles to Cascades



UNIVERSITY OF CAMBRIDGE

Clusters and Cascades Relationship



- Set of initial adopters of A (S), threshold q

  1. If the remaining network contains a cluster of density greater than 1-q then set S will not cause a complete cascade.
  2. Whenever set S does not cause a complete cascade with threshold q the remaining network must contain cluster of density greater than 1-q

UNIVERSITY OF CAMBRIDGE

Cascade Capacity of Networks

- **Cascade Capacity** of a network is the largest value of the threshold q for which some finite set of early adopters can cause a complete cascade

- In the following case cascade capacity is ½
  - Even if the network is infinite

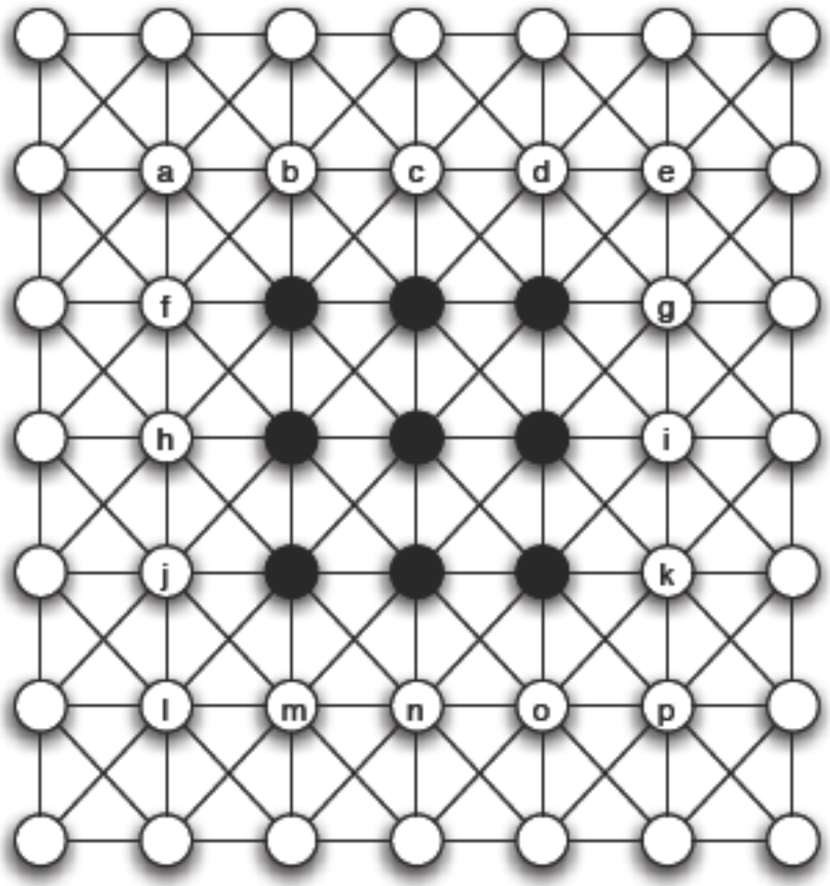
UNIVERSITY OF CAMBRIDGE

For q<=1/2 complete cascade

Cascade Capacity on a Grid

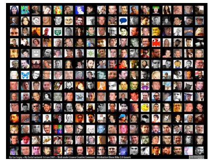
- If q <=3/8 there is a complete cascade
- If q is smaller (eg 2/8) cascade spreads even faster.
- Cascade Capacity is 3/8

- A network with a **large capacity** is one where cascades happen easily.



UNIVERSITY OF CAMBRIDGE

How large can Cascade Capacity be?

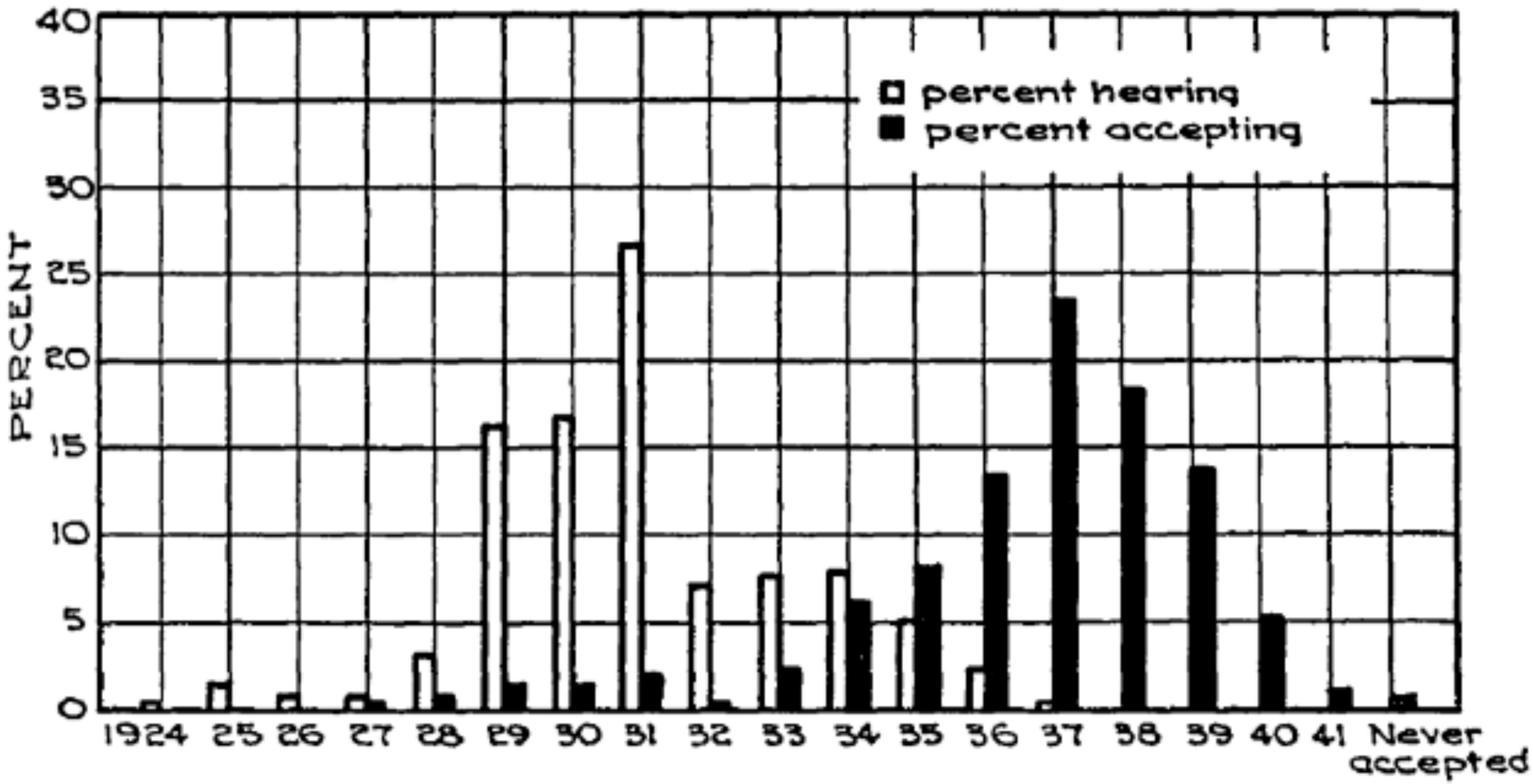


- Can it be higher than ½?
  - This would mean for instance that an inferior innovation can displace a superior one even when the inferior innovation starts with few initial adopters.

- No network has a capacity higher than ½.

UNIVERSITY OF CAMBRIDGE

Learning versus Adopting

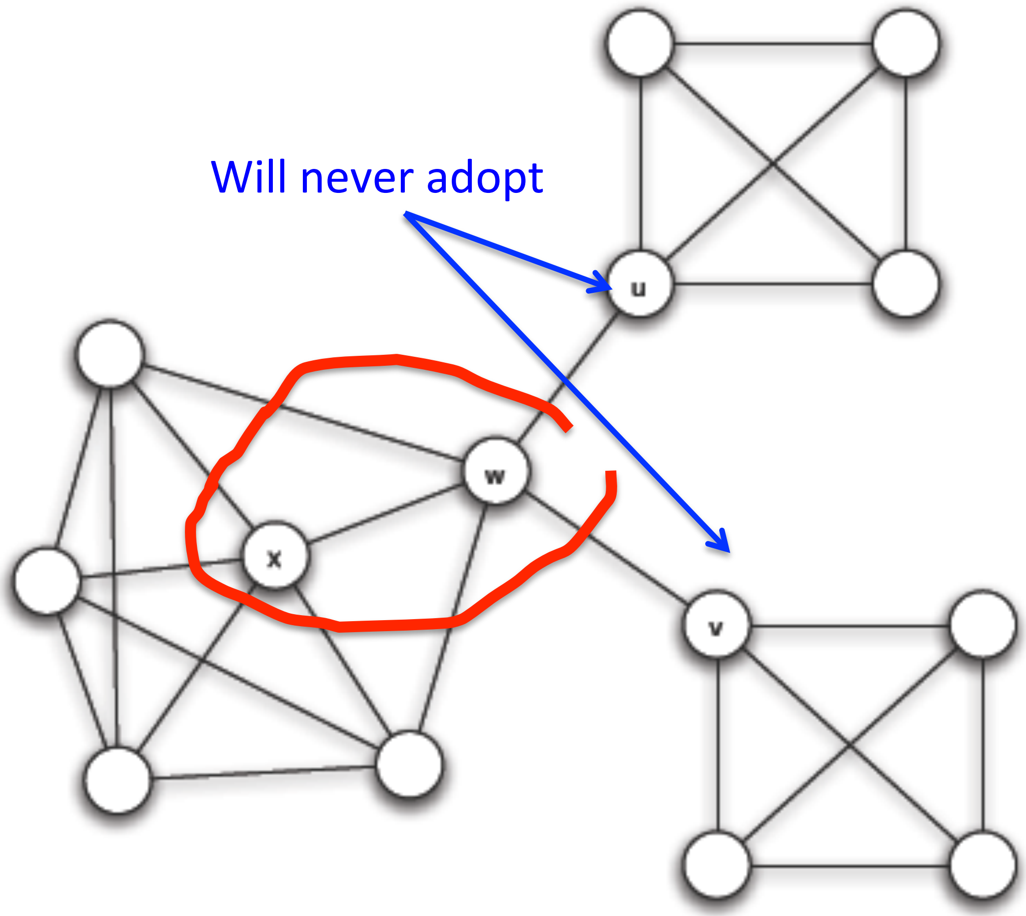


UNIVERSITY OF CAMBRIDGE

Adoption of hybrid corn

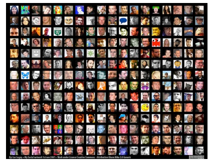
Role of Weak Ties

- v, w initial adopters
- q=1/2

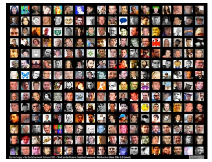


UNIVERSITY OF CAMBRIDGE

Weak Ties Role and Behaviour Adoption



- Weak ties are **very powerful in spreading** new information.
- **Weak ties are weak at transmitting behaviours that are somehow risky and costly to adopt.**

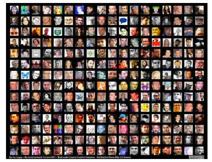


A Spreading Behaviour Experiment

- Controlled experiment
- Anonymous recruited participants
- Website on health behaviour
- Participants could decide if to adopt a behaviour based on adoption patterns of neighbours [assigned in the site]
  - "Adopt" means register for health forum
  - Participants assigned to either a random network or a clustered lattice [see Lecture 2 for example of the two networks] with same number of nodes and degree.

UNIVERSITY OF CAMBRIDGE

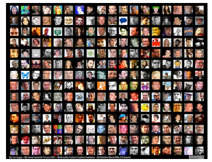
What do the experiment settings mean?



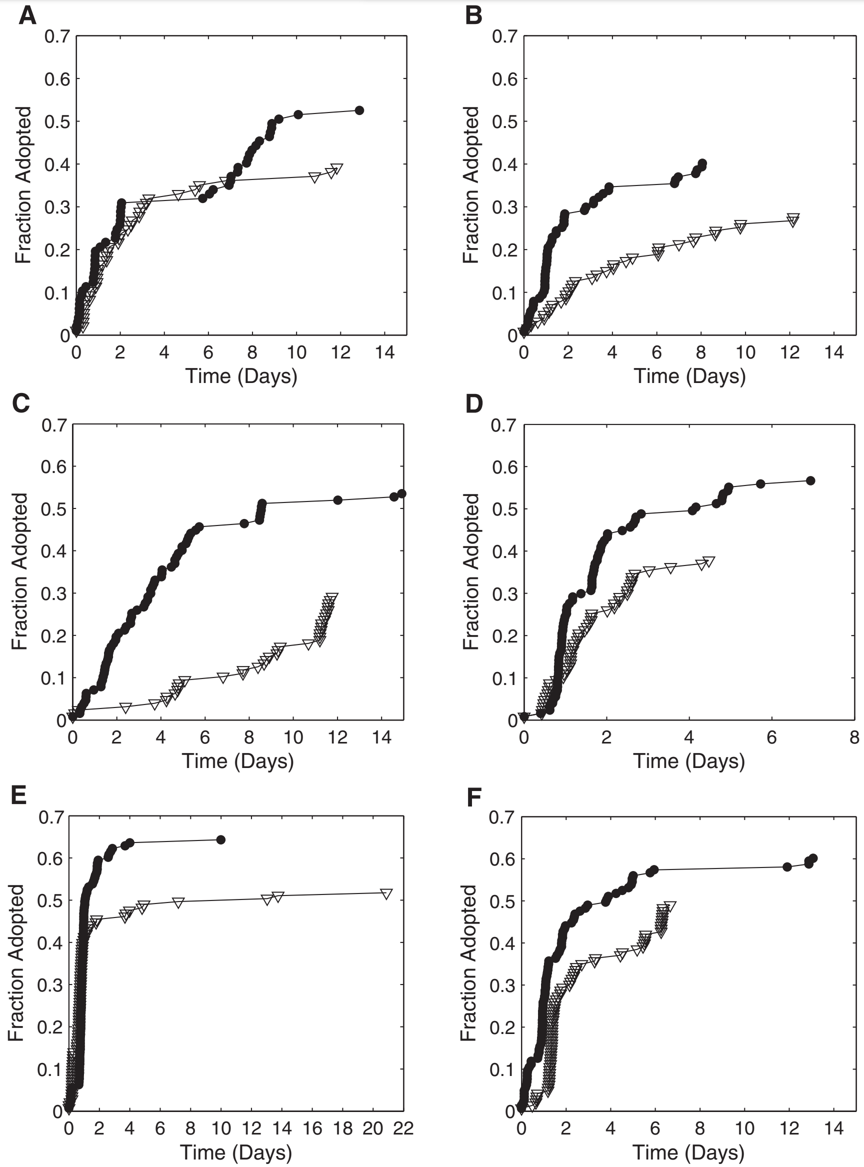
- The forum was known only to participants in the study
  - Influence could be studied
  - Only main difference was in the 2 network structures

UNIVERSITY OF CAMBRIDGE

Adoption of Behaviour: Results



Circles: Lattice
Triangles: Random
A-E various experiments with different N and degree

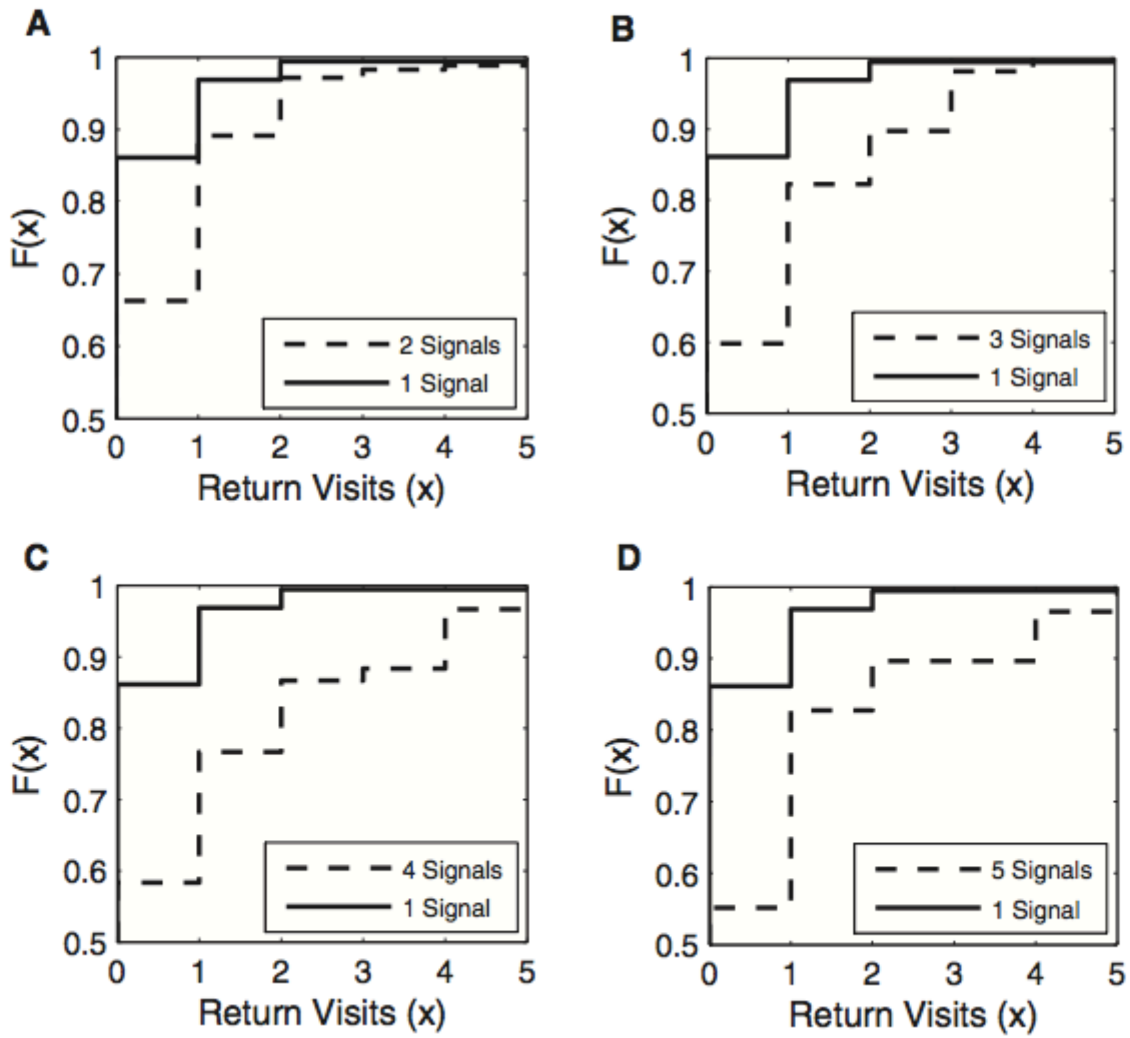


Topologies with greater cluster and diameter were better! Spreading was more than 4 times faster. Higher degree helped!

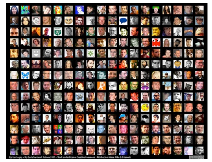
UNIVERSITY OF CAMBRIDGE

Social Reinforcement

Receiving more than 1 signal is good although beyond 4 there is no improvement



UNIVERSITY OF CAMBRIDGE



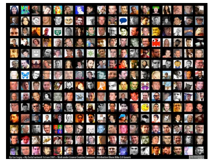
Nugget of the paper…

- Network structure plays in an important role on influencing behaviour
  - Structures containing more clusters were better at transmitting behaviour

- Reinforcement is very important in influencing a user

UNIVERSITY OF CAMBRIDGE

The Spanish ‘Indignados’ Movement



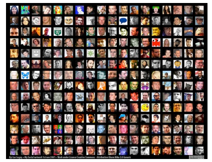
UNIVERSITY OF CAMBRIDGE

Data from Twitter (2011)

25-A	*Start of observation window*
5-M	First demonstration attempts.
9-M	Students occupy university.
15-M	**First mass demonstrations.**
16-M	Camps in squares of Madrid and Barcelona.
17-M	Police evictions. Protesters mobilise (and camp) in several cities.
18-M	Protests declared illegal.Thousands concentrate in squares.
19-M	Popular support and media coverage increase.
20-M	Thousands camped.
21-M	Tens of thousands protest, defying ban.
22-M	**Election day.** Protesters decide to stay camped in squares.
25-M	*End of observation window*

UNIVERSITY OF CAMBRIDGE

Timeline of hashtags



#hashtags

5 May:
nolesvotes
democraciarealya 15m
tomalacalle

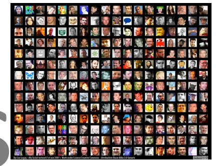
10 May:
15m nolesvotes
democraciarealya
tomalacalle

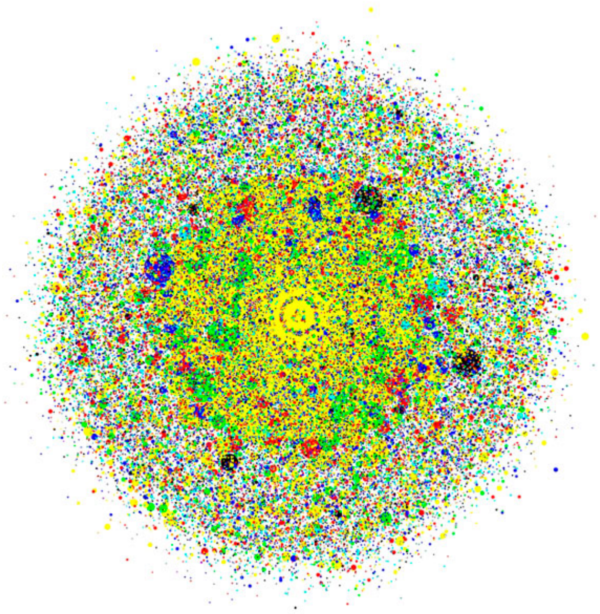
15 May:
15mani
15m
acampadasol
spanishrevolution
nolesvotes
democraciarealya
acampadabcn
tomalacalle
acampadasevilla
acampadavalencia

20 May:
spanishrevolution
acampadasol
15m
nonosvamos
nolesvotes
democraciarealya
acampadabcn
acampadasevilla
yeswecamp
acampadavalencia

25 May:
acampadasol
15m
spanishrevolution
consensodeminimos
democraciarealya
acampadabcn
nolesvotes
nonosvamos
acampadasevilla
acampadavalencia

The Twitter Network: Statistics

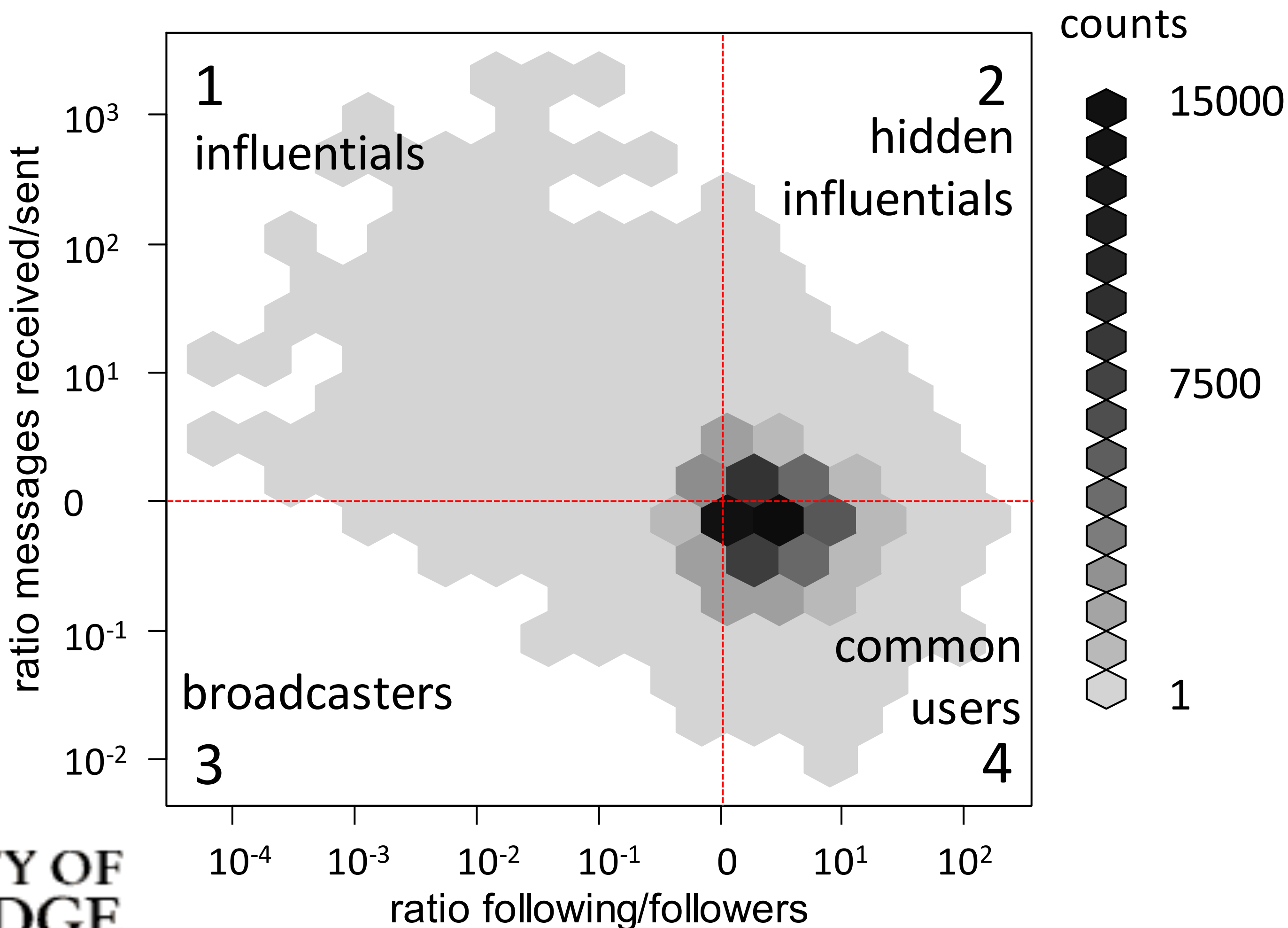




	Full Network	Symmetrical Network
N (# nodes)	87,569	80,715
M (# arcs)	6,030,459	2,644,367
$\langle k \rangle$ (avg degree)	69	33
C (clustering)	0.220	0.198
l (path length)	3.24	3.65
D (diameter)	11	11
r (assortativity)	-0.139	-0.0344
# strong components	5,249	139
N giant component	82,253	80,421
N 2^{nd} component	4	4
max(k_{in}) (# following)	5,773	5,082
max(k_{out}) (# followers)	31,798	5,082

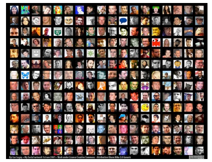
UNIVERSITY OF CAMBRIDGE

Distribution of Users in the Network by Activity

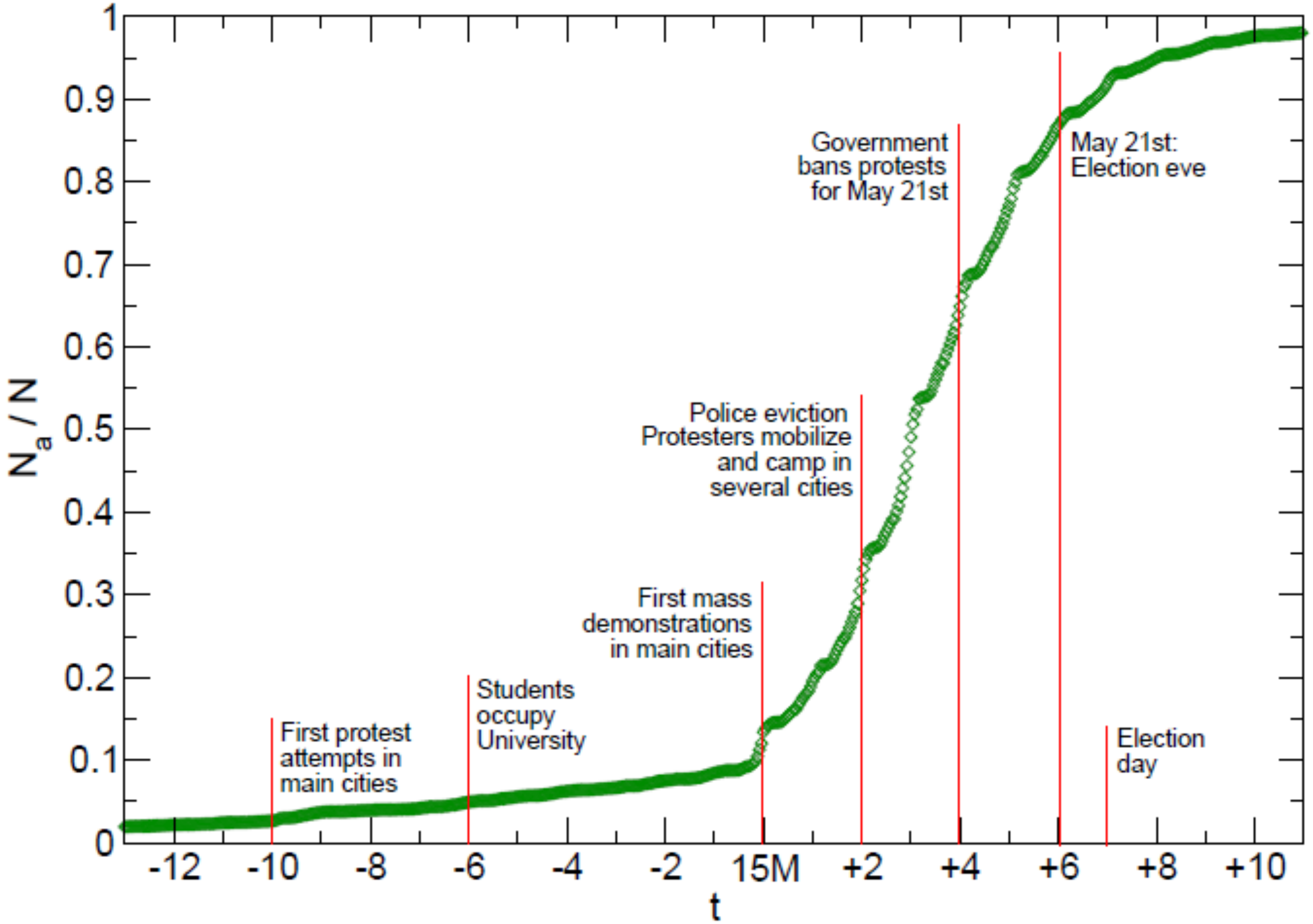


UNIVERSITY OF CAMBRIDGE

The Online Growth of the Movement



Na is the number of adopters ie those who sent at least 1 message related to the protest



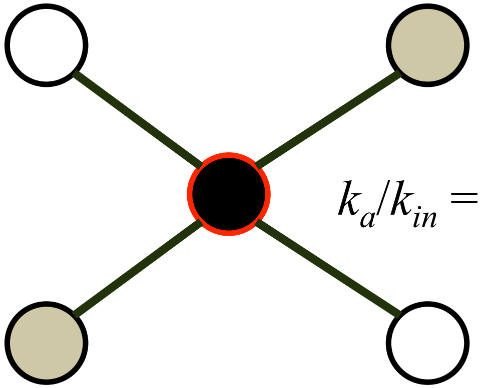
UNIVERSITY OF CAMBRIDGE

Recruitment and Activation Threshold

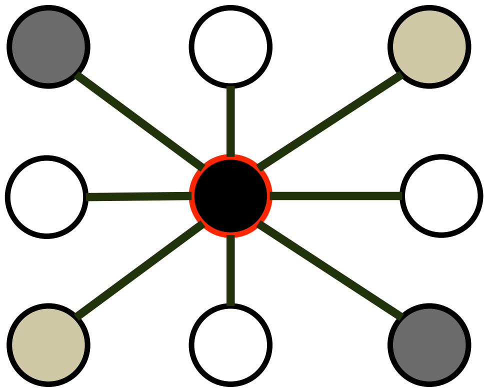
- Activation time: moment when users start emitting protest messages

- $k_a/k_{in} \approx 0 \rightarrow$ low threshold individuals (no need of 'local pressure')

- $k_a/k_{in} \approx 1 \rightarrow$ high threshold individuals (need high 'local pressure')



$k_a/k_{in} = 2/4 = 0.5$

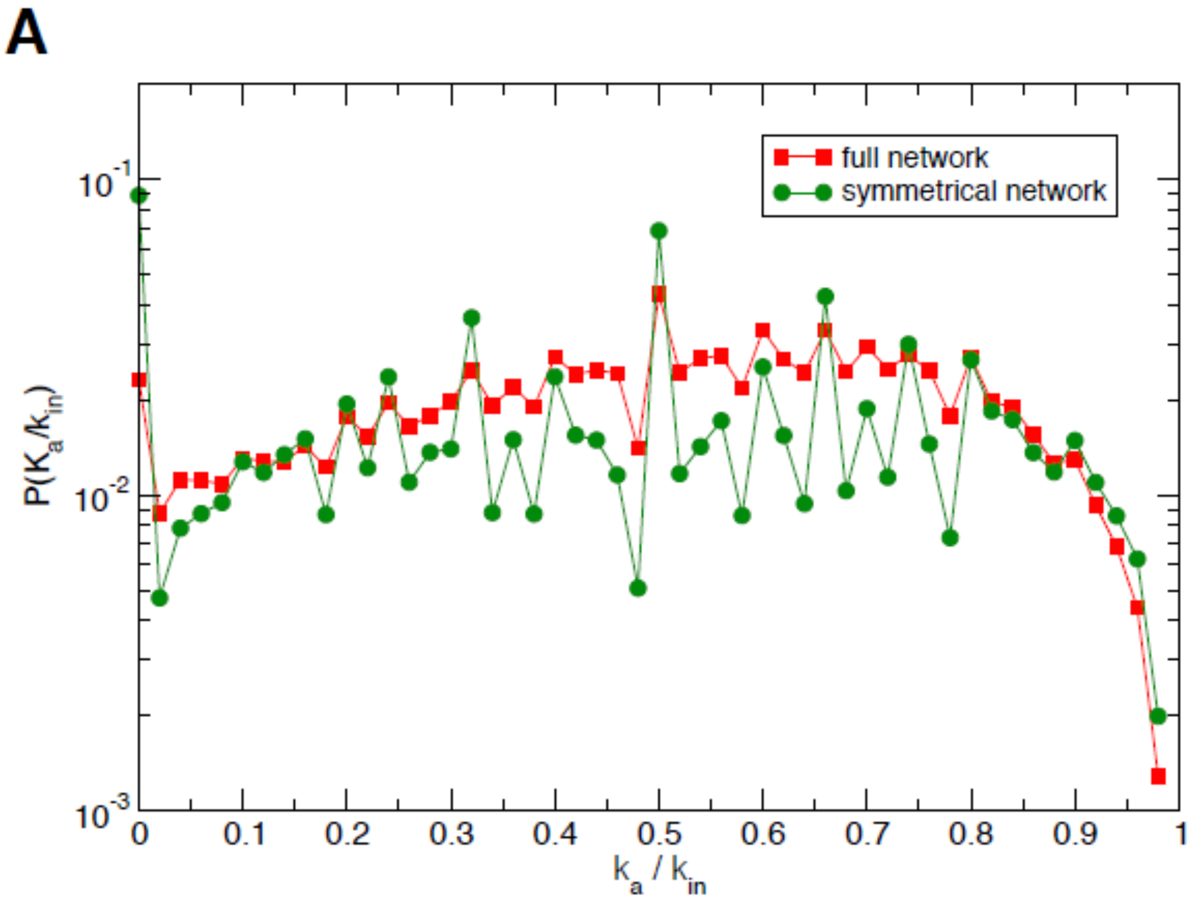


$k_a/k_{in} = 4/8 = 0.5$

UNIVERSITY OF CAMBRIDGE

Distribution of Thresholds

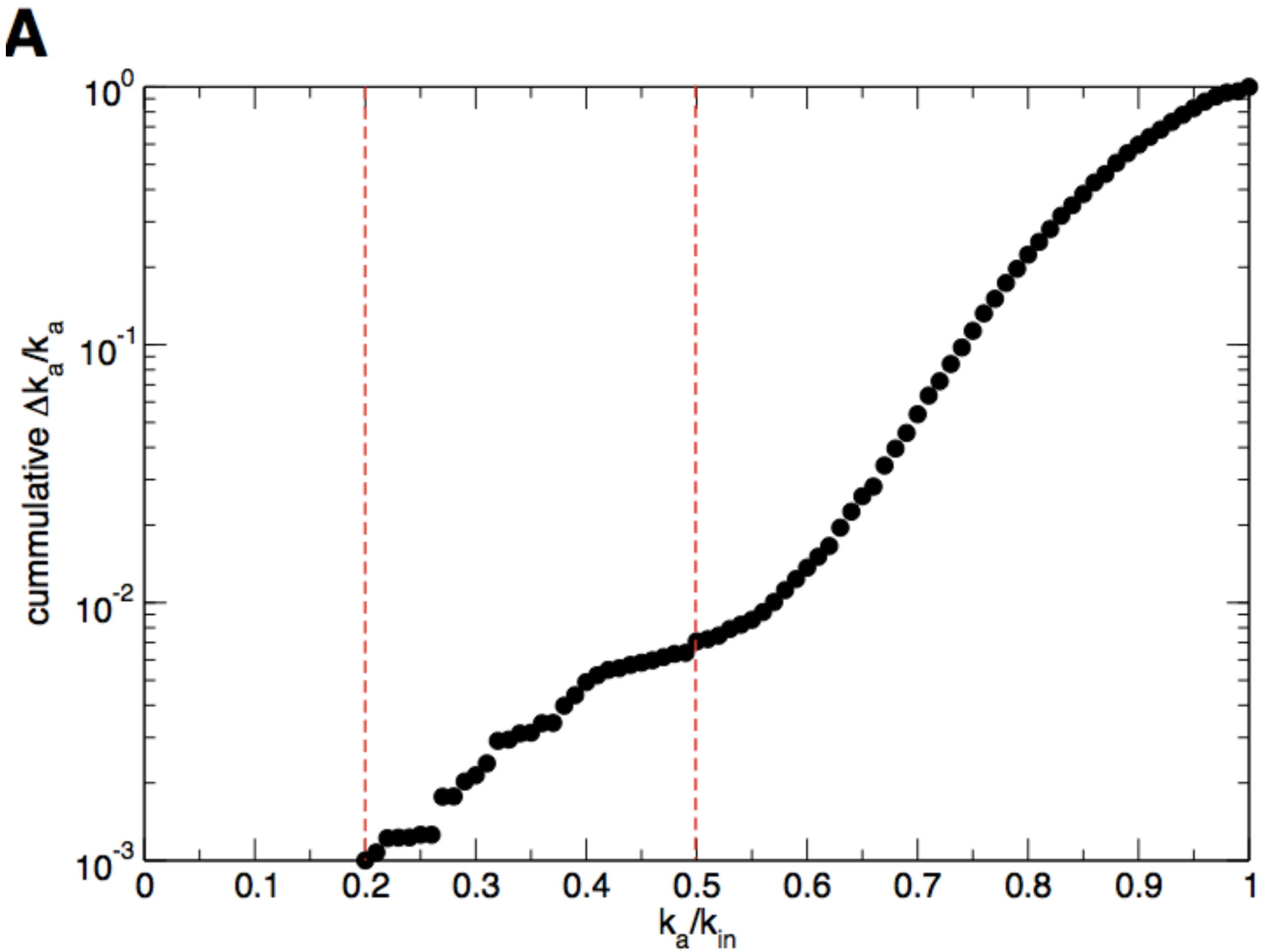
2 local max at 0 and at 0.5



UNIVERSITY OF CAMBRIDGE

Are activations affected by recruitment bursts?

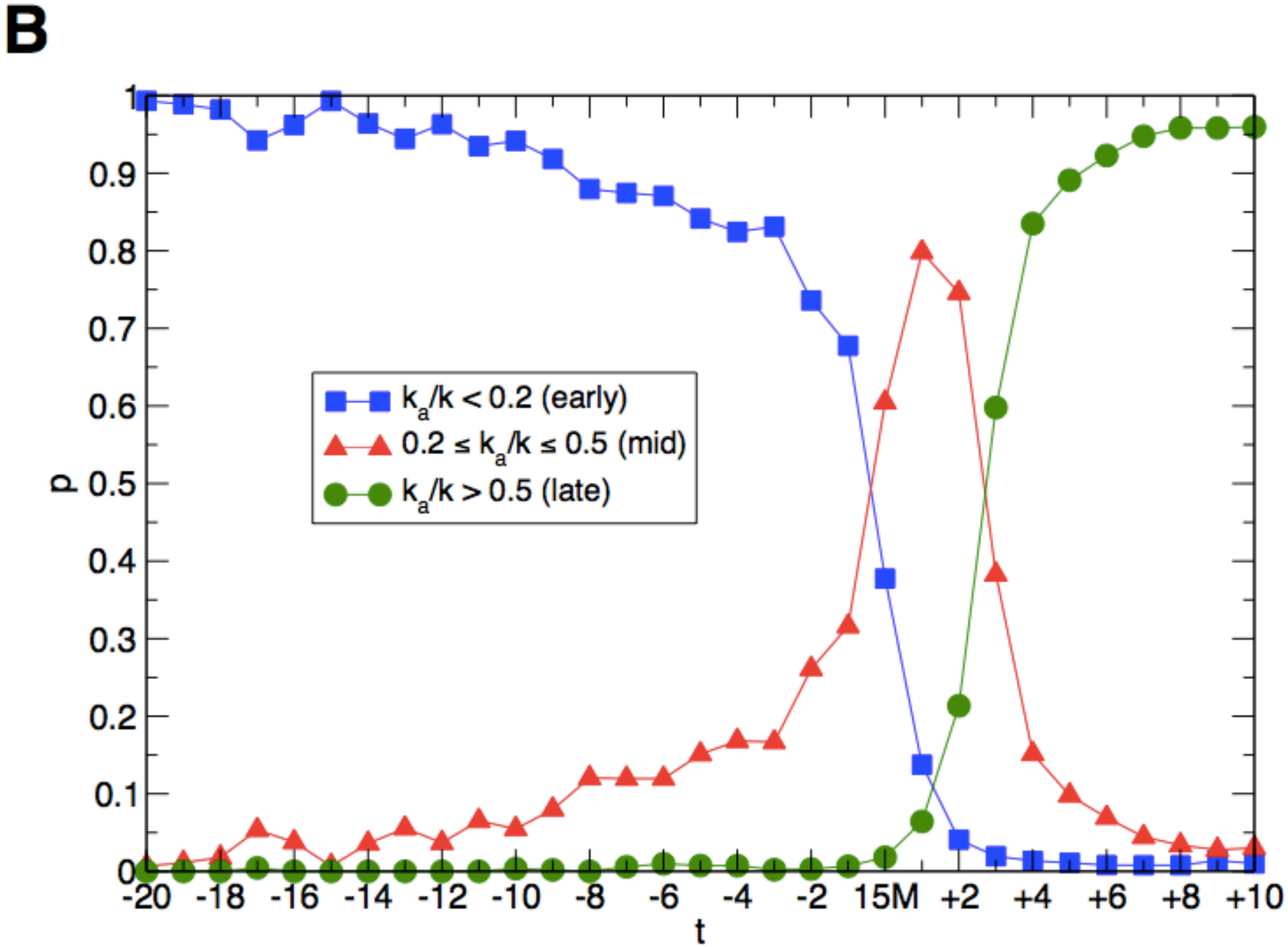
Y axis refers to the growth of activation. Early participants (low threshold) are not affected by bursts.



UNIVERSITY OF CAMBRIDGE

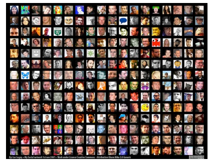
Joining time of different groups

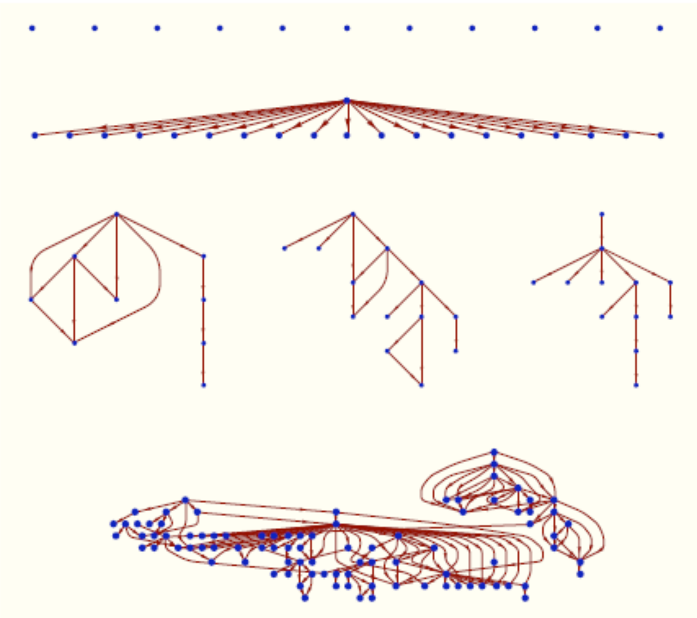
Y axis is percentage of joiners.



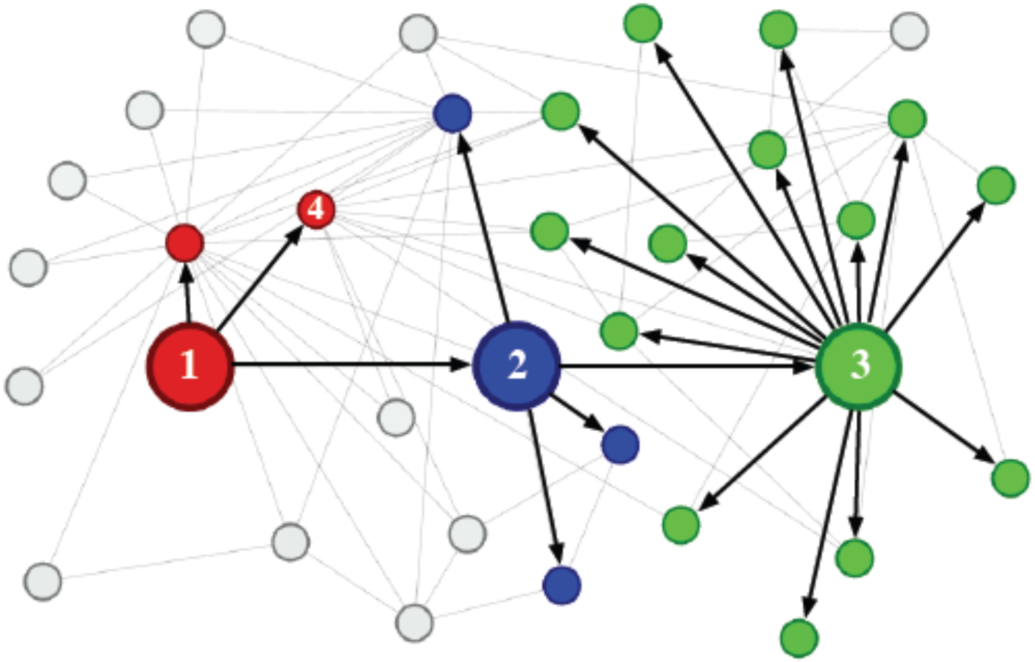
UNIVERSITY OF CAMBRIDGE

Information Cascades





bursts of activity

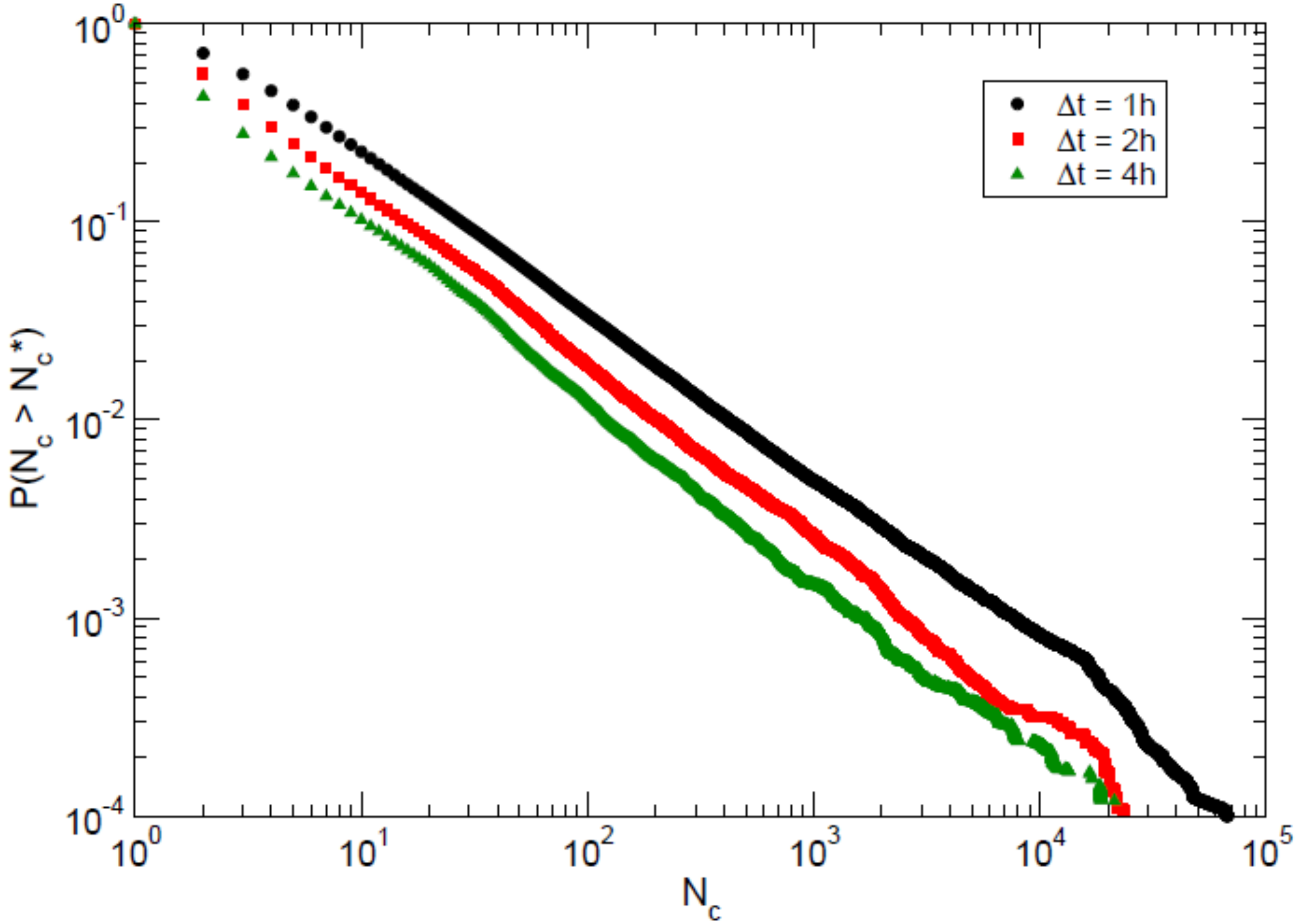


t $\qquad$ $t + \Delta t$ $\qquad$ $t + 2\Delta t$

UNIVERSITY OF CAMBRIDGE

Information Cascades

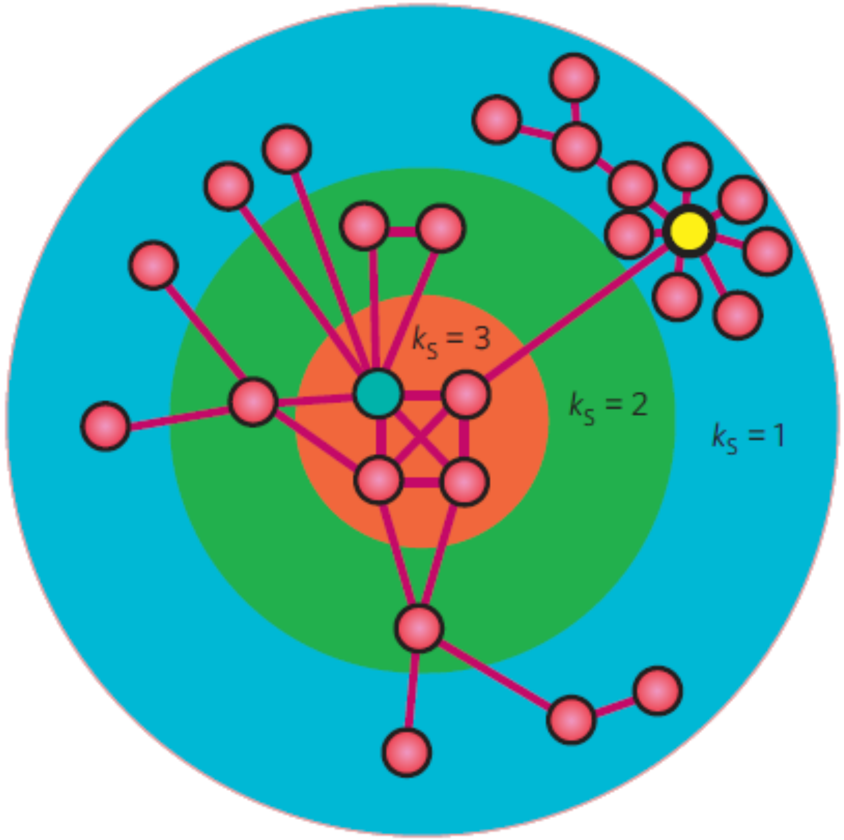
Cascades are quite short, in accordance with the literature



UNIVERSITY OF CAMBRIDGE

Where are Recruiters and Spreaders?

k-shell decomposition

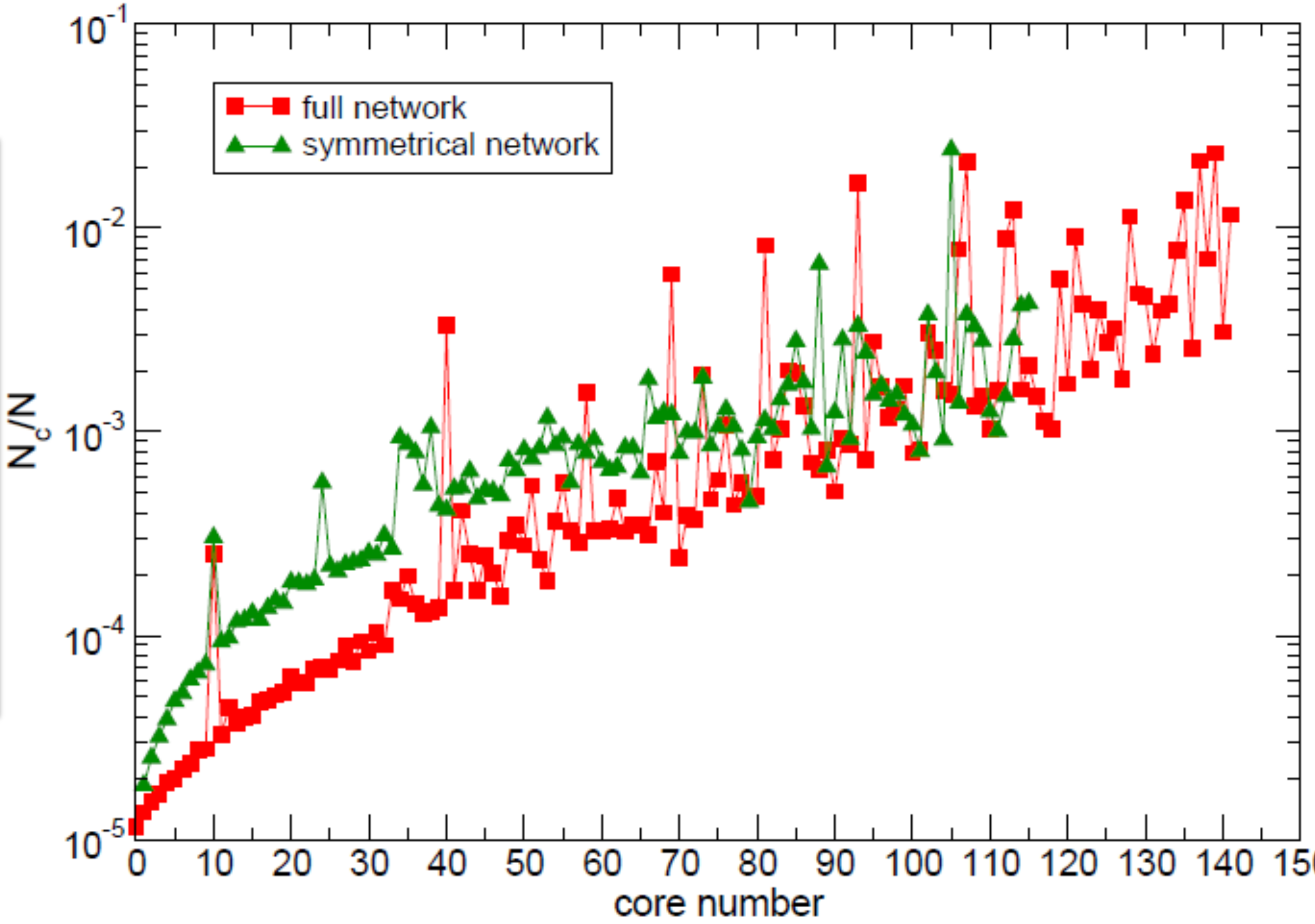


(Kitsak et al. 2010)

UNIVERSITY OF CAMBRIDGE

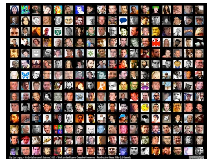
Where are the Spreaders?

Correlation with cascade size and core position (centrality) of node.
Additional result (not shown here) is that early adopers are spread in all cores (have different centrality)



UNIVERSITY OF CAMBRIDGE

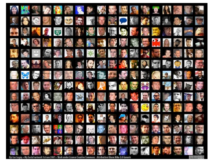
Nugget of the paper…



- Feedback between dynamics of recruitment and information diffusion
- **Being central is crucial for diffusion**, not so for recruitment
- Exogenous factors create random seeding in the network

UNIVERSITY OF CAMBRIDGE

Limitations of the approach

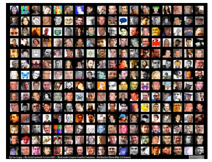


Two main limitations:

- no control for homophily
- no control for exposure to offline media

So it might be overestimating influence

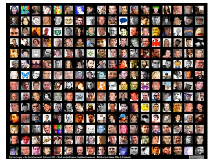
UNIVERSITY OF CAMBRIDGE



Summary

- We have introduced cascades and threshold models
- We have described two empirical example of behaviour spreading in online networks

UNIVERSITY OF CAMBRIDGE



References

- Chapter 19
- Damon Centola. The Spread of Behavior in an Online Social Network Experiment. *Science* 329, 1194 (2010)
- S. Gonzalez Bailon, J. Borge-Holthoefer, A. Rivero and Y. Moreno. The Dynamics of Protest Recruitment through an Online Network. Nature Scientific Reports, 1, 197 (2011).