KINETICS OF COMPLEX SOCIAL CONTAGION

János Kertész
Central European University

Pohang, May 27, 2016

- Theory:

Zhongyuan Ruan, Gerardo Iniguez, Marton Karsai, JK:

Kinetics of social contagion

Phys. Rev. Lett. 115, 218702 (2015)

- Empirical study:

M. Karsai, G. Iniguez, Riivo Kiskas, Kimmo Kaski, JK:

Global contagion with local cascades: The anatomy of online adoption spreading

Submitted

Similarities and Differences

	Network	Transmission	External influence

Complex contagion process

(D. J. Daley, D. G. Kendall, Epidemics and rumours. Nature 204, 1118 (1964))

Social Contagion

Information, ideas and even behaviors can spread through networks of people reminiscent to how infectious diseases do – hence social contagion

There are important differences:

- Social pressure: The state of neighbors influence the transmission probability
- There is a flow of external influence due to media (like external field)

Diffusion of innovations is an example of complex social contagion.

Role of Innovation in Economy

Equilibrium theories: Static view. There are needs (demand), which can be satisfied by supply of goods and services at the price determined by their balance.

Change one parameter and assume smooth dependence.

Economic growth: Non-equilibrium. Increasing productivity, new products, new demand. (Schumpeter's "creative destruction").

Key element: **Innovation**

Innovation: creation of novel values through invention, ideas, technologies, processes.

Cascading Phenomena

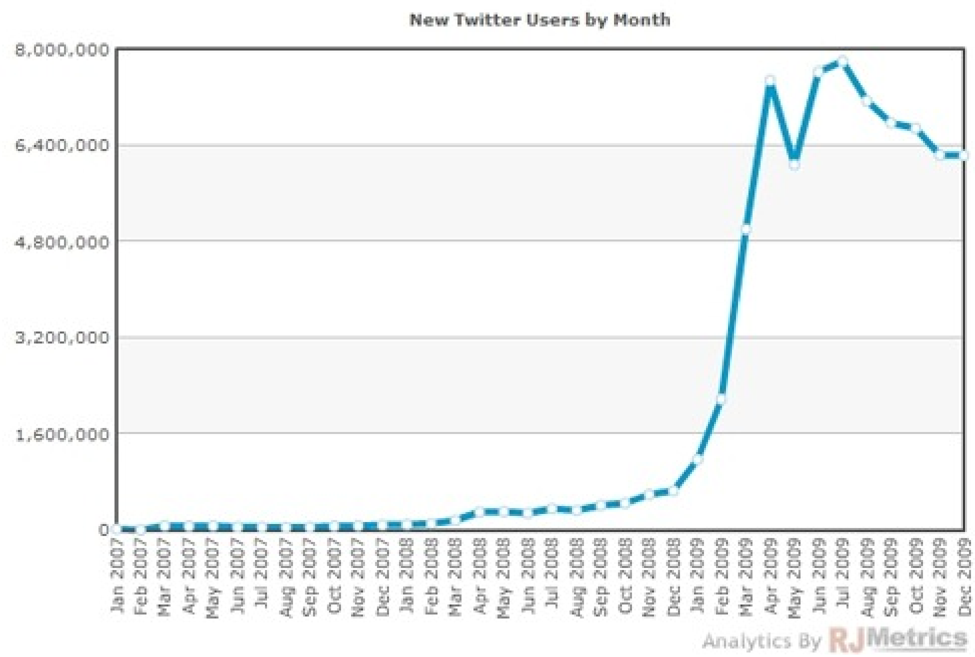
Complex social contagion can be surprisingly fast. A triggering perturbation may release rapid spreading.

Examples:

Rumor (false breakdown in nuclear power plant: Hungary, 2002)

Political movements (Arab spring 2011)

Innovation: Twitter (2009)



Threshold Model

Granovetter (Am. J. Sociology 1978) Threshold models
D. Watts (PNAS 2002) Mathematical form

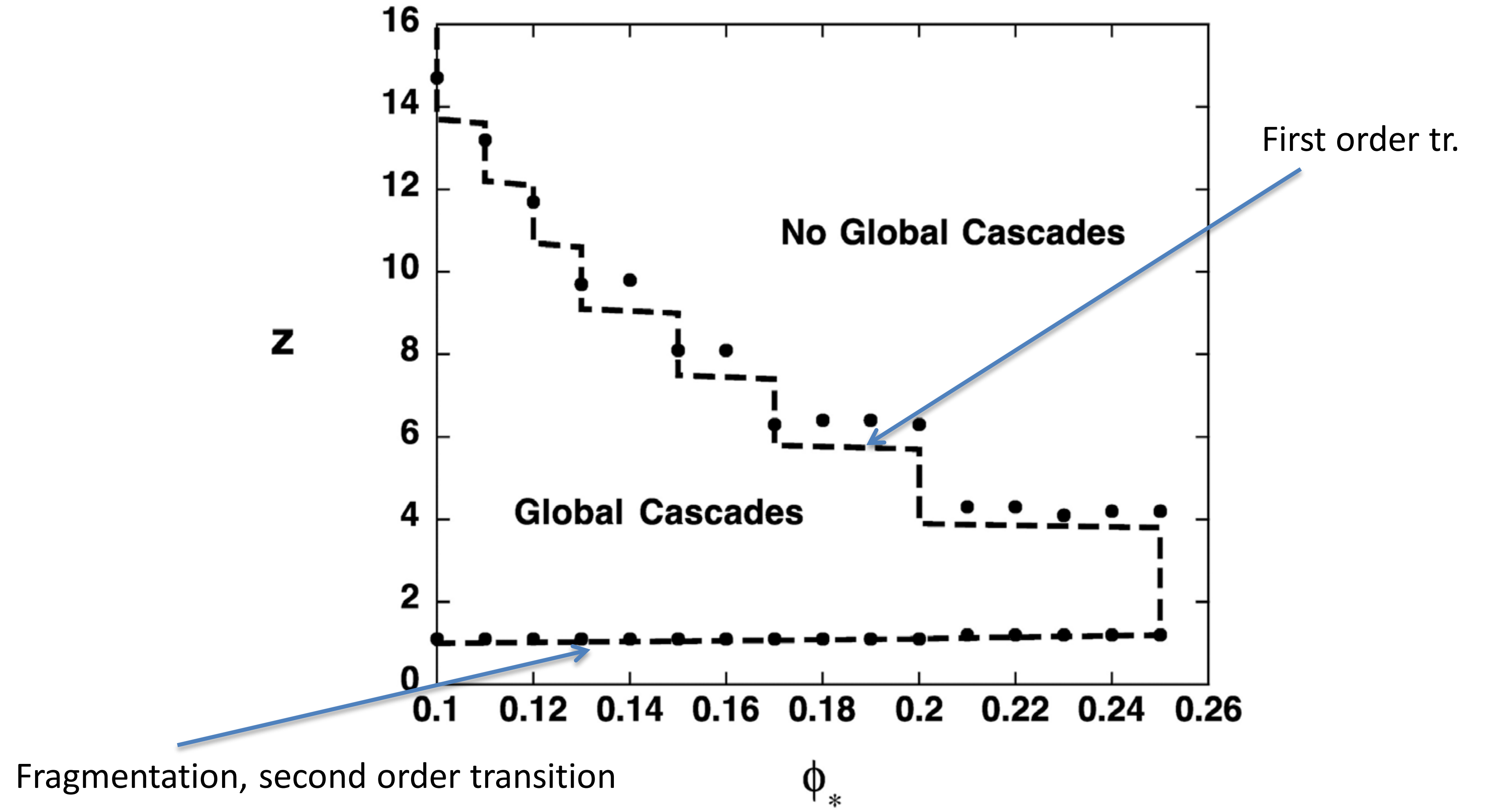
Random network with degree distribution P_k and average degree $\langle k \rangle = z$. Every node has a threshold ϕ indicating the **critical ratio** of adopting neighbors needed to make the node adopt. Initiate the process by infecting a node.

There are **vulnerable** nodes, which get infected if they have one adopting neighbor: $\phi \leq 1/k$.

The others are **stable**.

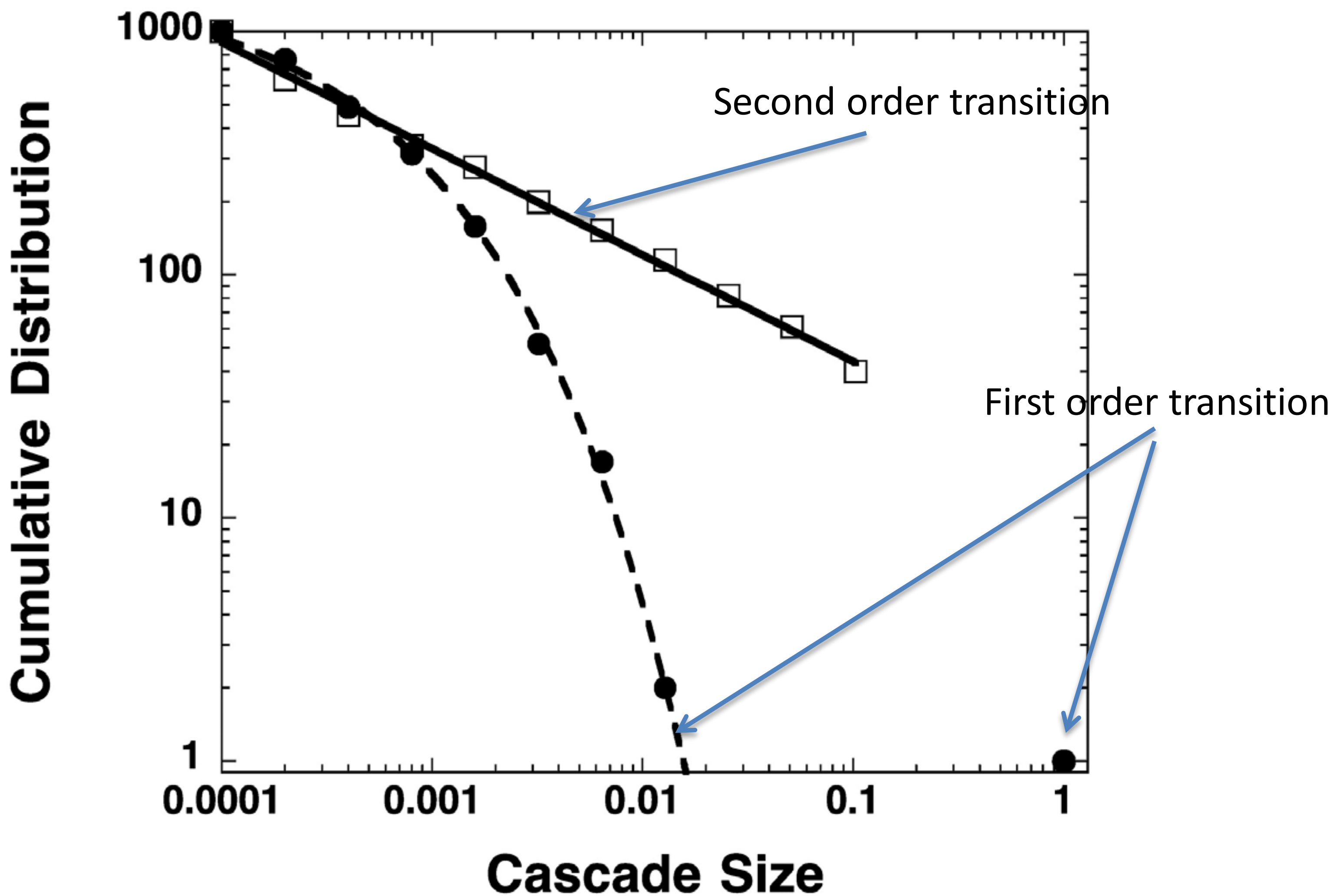
The phase diagram can be calculated.

Cascade windows for the threshold model. (ER graph)



Watts D J PNAS 2002;99:5766-5771

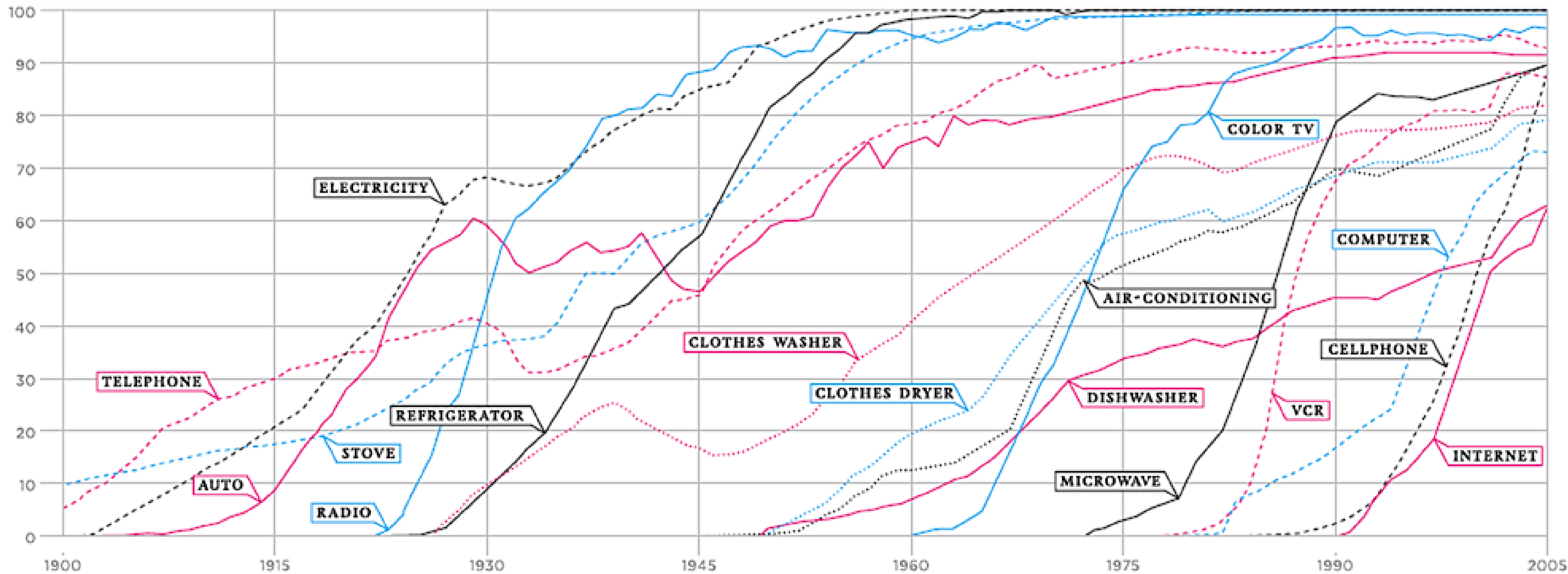
Cumulative distributions of cascade sizes at the lower and upper critical points, for n = 1,000 and z = 1.05 (open squares) and z = 6.14 (solid circles), respectively.



Watts D J PNAS 2002;99:5766-5771

PNAS

% US
Housholds



Adoption speed can be very different for different innovations

http://www.nytimes.com/imagepages/2008/02/10/opinion/10op.graphic.ready.html

Generalized Watts Model

In the Watts model the criterion for a dynamic process (global cascade) is traced back to a static problem, the existence of the **percolating vulnerable cluster.**

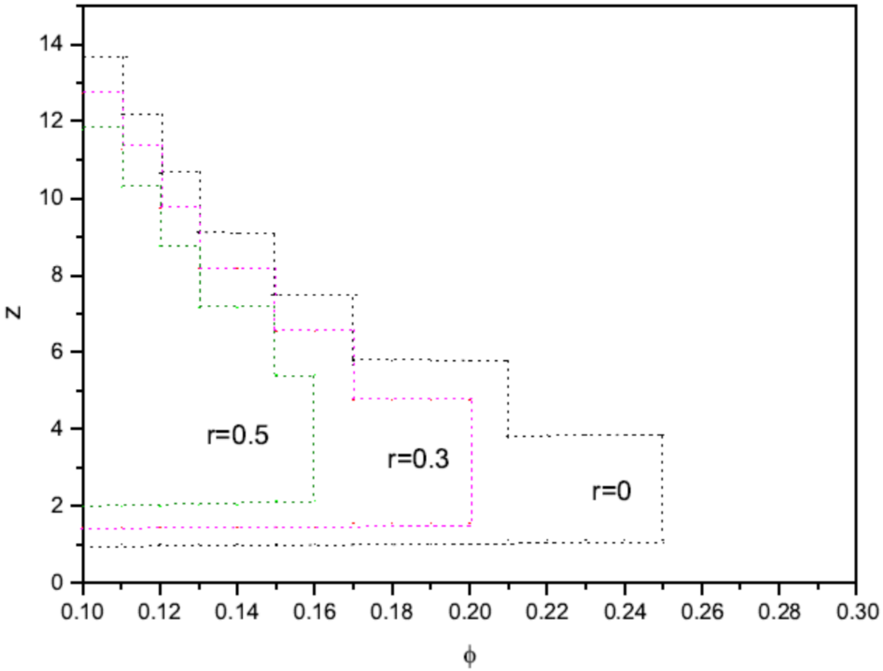
Incomplete picture

We extended the Watts model by two important elements

1. Some **nodes are blocked**. Some people are reluctant to adopt (have a satisfactory service, have some principal reasons etc.) (still static);
2. There are **spontaneous innovators appearing** as external information flows continuously (intrinsically dynamic).

Blocked Nodes

Nodes are blocked with probability *r* (quenched disorder).
Blocked nodes make it more difficult to fulfill the threshold criterion.
The problem can be solved similarly to the original Watts case.
The result is a **three-dimensional phase diagram**:



ER graph with average degree *z*, uniform threshold ϕ and blocking probability *r*.

Generating Function Method

p_k Prob that a node has degree *k*

ρ_k Prob that a node of degree *k* is vulnerable $(1/k > \phi)$

q_n Prob that a node belongs to vulnerable cluster of size *n*

w_n Prob that a node's neighbor –" – –" – –" – of size *n*

$G_0(x) = \sum_k p_k \rho_k (1-r)\, x^k$ gen. fn.: a node →vuln.

$G_1(x) = \sum_k \frac{k p_k \rho_k}{z} (1-r)\, x^{k-1}$ gen. fn.: a node's neighbor → vuln.

$G_1(x) = G_0'(x)/z$

$H_0 = \sum_n q_n\, x^n$ gen. fn.: node belongs to vuln. cluster

$H_1 = \sum_n w_n\, x^n$ gen. fn.: node's neighbor –" – –" – –" –

Sparse, random, uncorrelated networks are tree like

Generating Function Method

Using tree-like property:

$$H_1(x) = 1 - G_1(1) + xG_1(H_1(x))$$

$$H_0(x) = 1 - G_0(1) + xG_0(H_1(x))$$

$\langle n \rangle = H_0'(1) = G_0(1) + \frac{(G_0'(1))^2}{z - G_0''(1)}$ from which the criterion

$G_0''(1) = \sum_k k(k-1) p_k \rho_k (1-r) = z$ for the transition

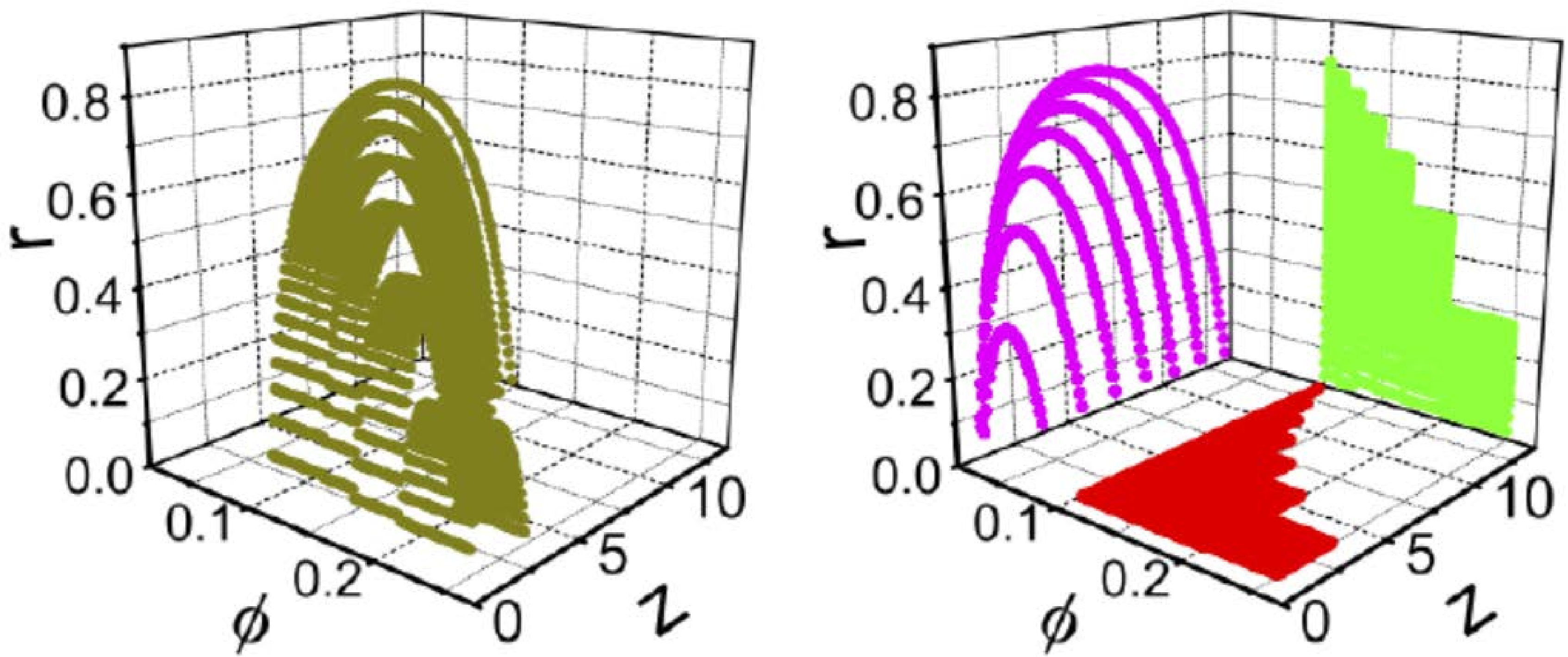
Up to $(1-r)$ the same as for the Watts model.

3D phase diagram for p_k and ϕ_i distributions.

3D Phase Diagram

For Erdős-Rényi graph p_k is Poisson, parametrized by *z*.
Assuming uniform ϕ with $k_c = \lfloor 1/\phi \rfloor$

$$(1-r)e^{-z}\sum_{k=2}^{k_c}\frac{z^k}{(k-2)!} - z = 0$$



Spontaneous Adopters

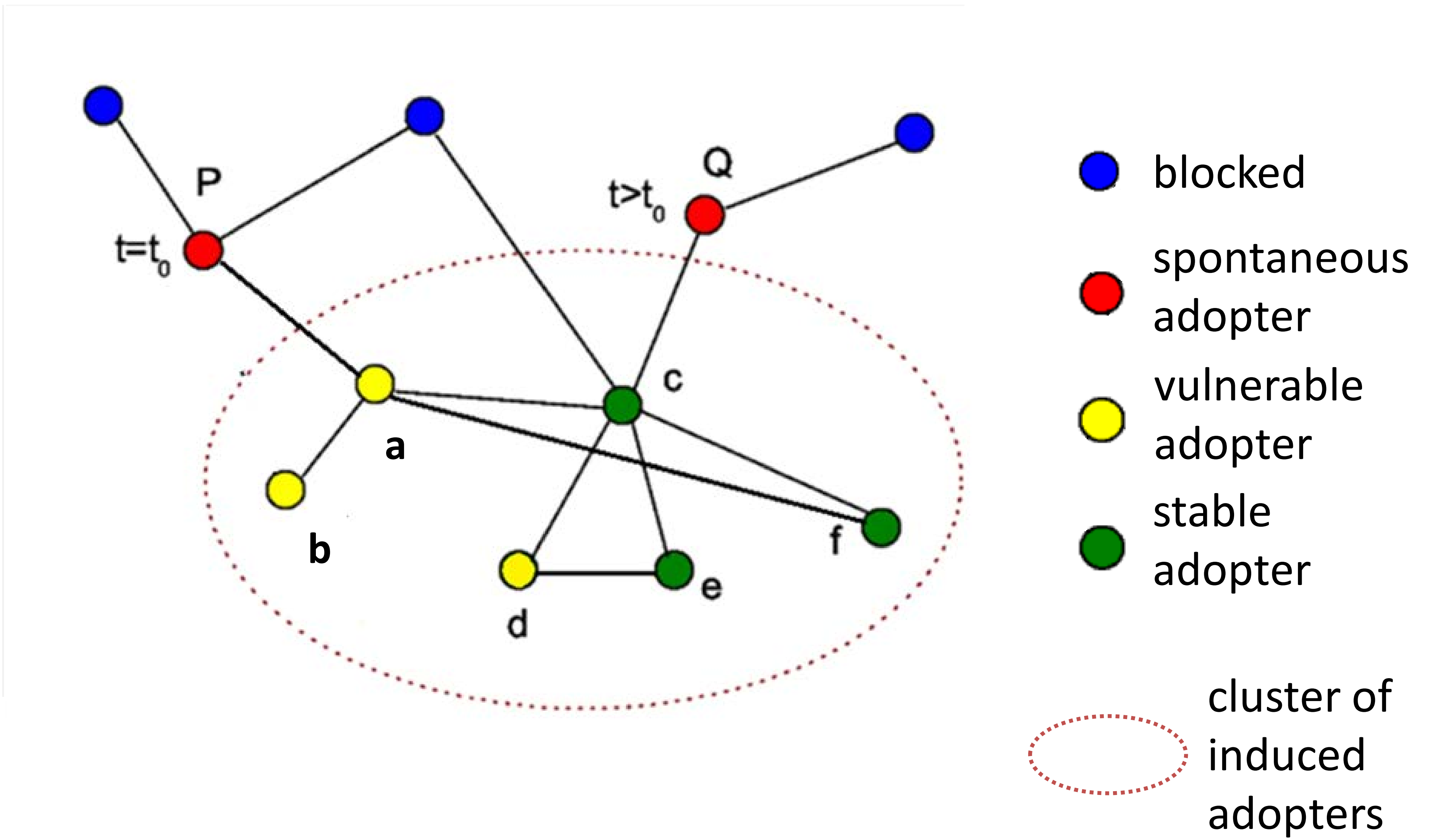
1. consider r=0

$p \to 0$ ⇒ original Watts model

$p \neq 0$ ⇒ unique final sate: everyone adopts

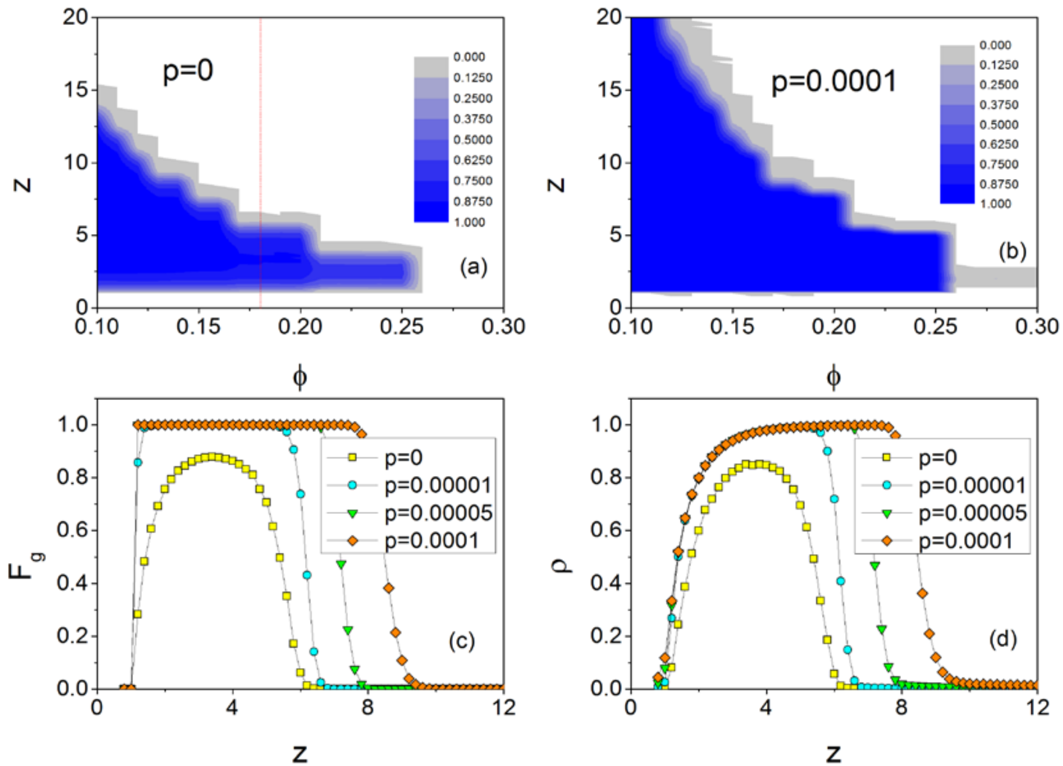
introduce a time window T

Node types


P
t=t$_0$
Q
t>t$_0$
a
b
c
d
e
f
blocked
spontaneous adopter
vulnerable adopter
stable adopter
cluster of induced adopters

Effect of Spontaneous Adopters

T=100



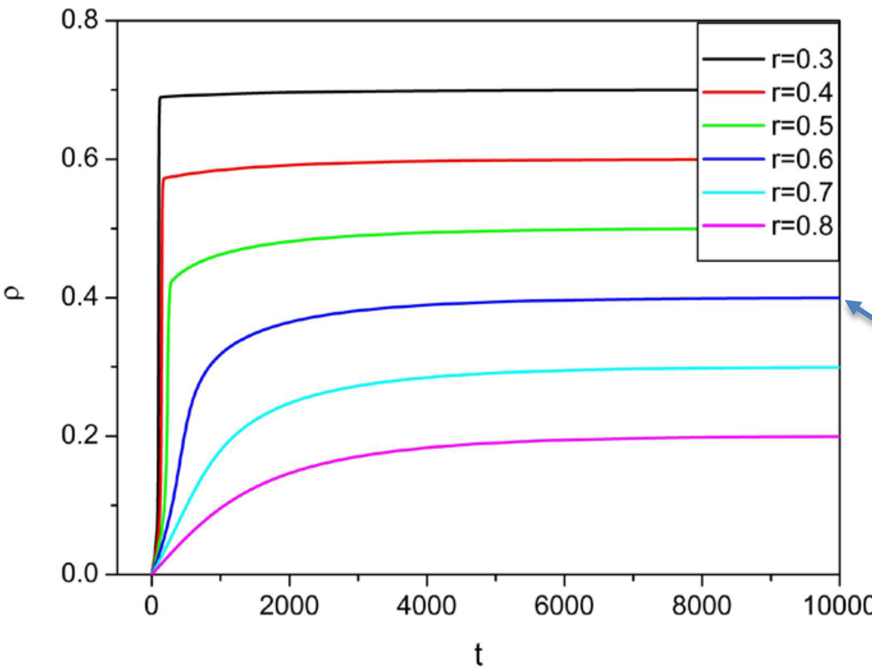
ER, $r = 0$, $\phi = 0.18$,

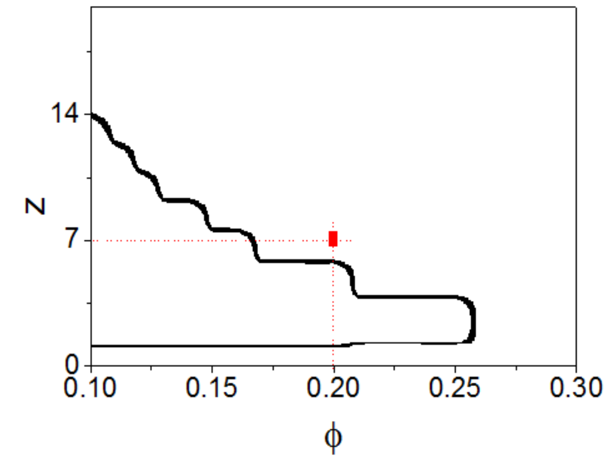
F_g: frequency of global cascades, (order parameter)

ρ: density of adopters

Spontaneous Adopters + Blocked Nodes

2. consider r>0





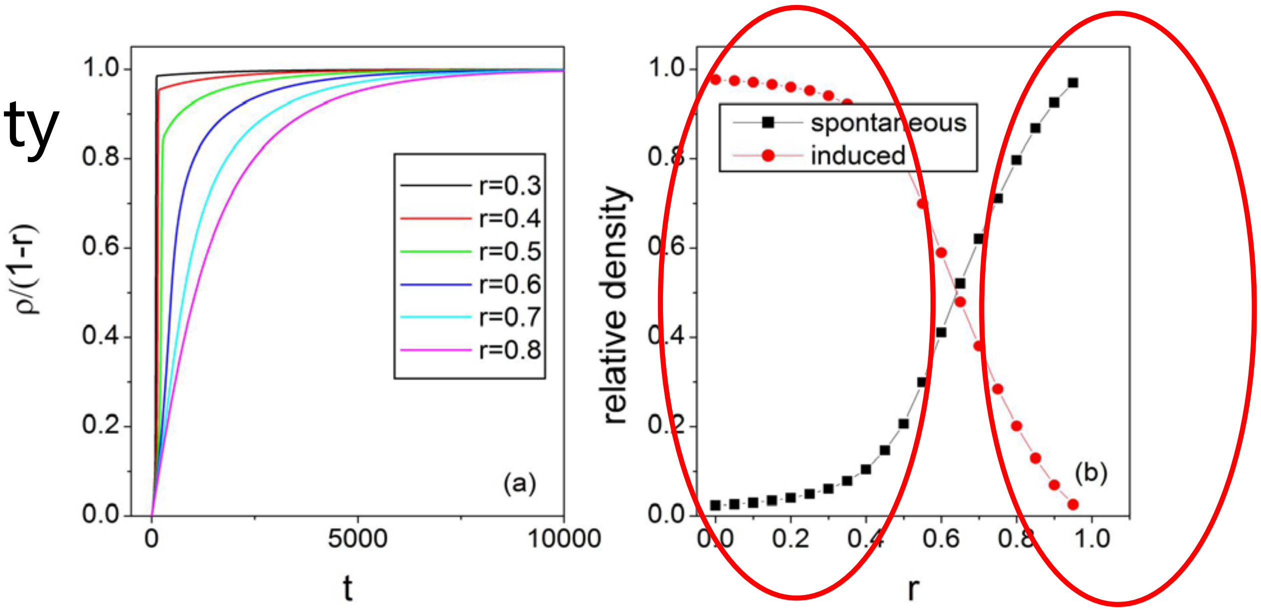
$1 - r$

ER, $z = 7,\ \phi = 0.2,\qquad p = 5 \times 10^{-4}$

Evolution of Adopter Density

Different mechnisms?

Normalized adopter density



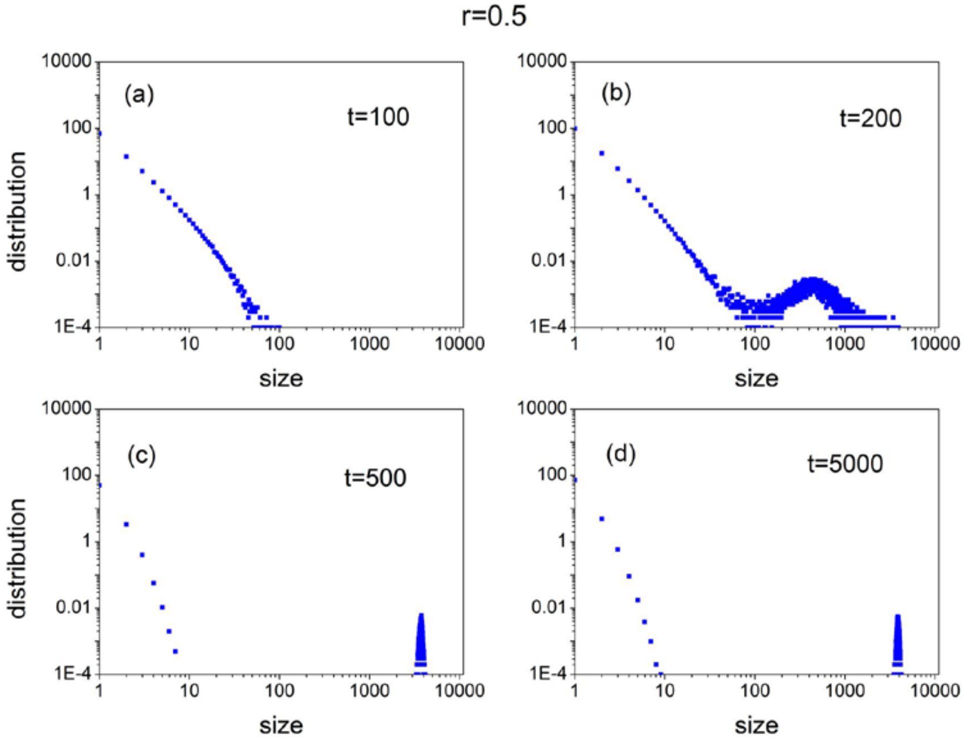
ER , $z = 7$, $\phi = 0.2$, $p = 5 \times 10^{-4}$

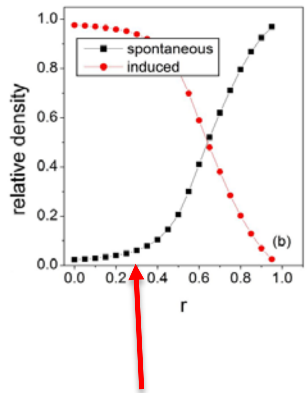
$r^* = 1 - 1/z = 0.86$ is the percolation threshold

Is there an $r_\times < r^*$ where the kinetics changes?

Distribution of Induced Clusters ($r < r_\times$)

scenario I

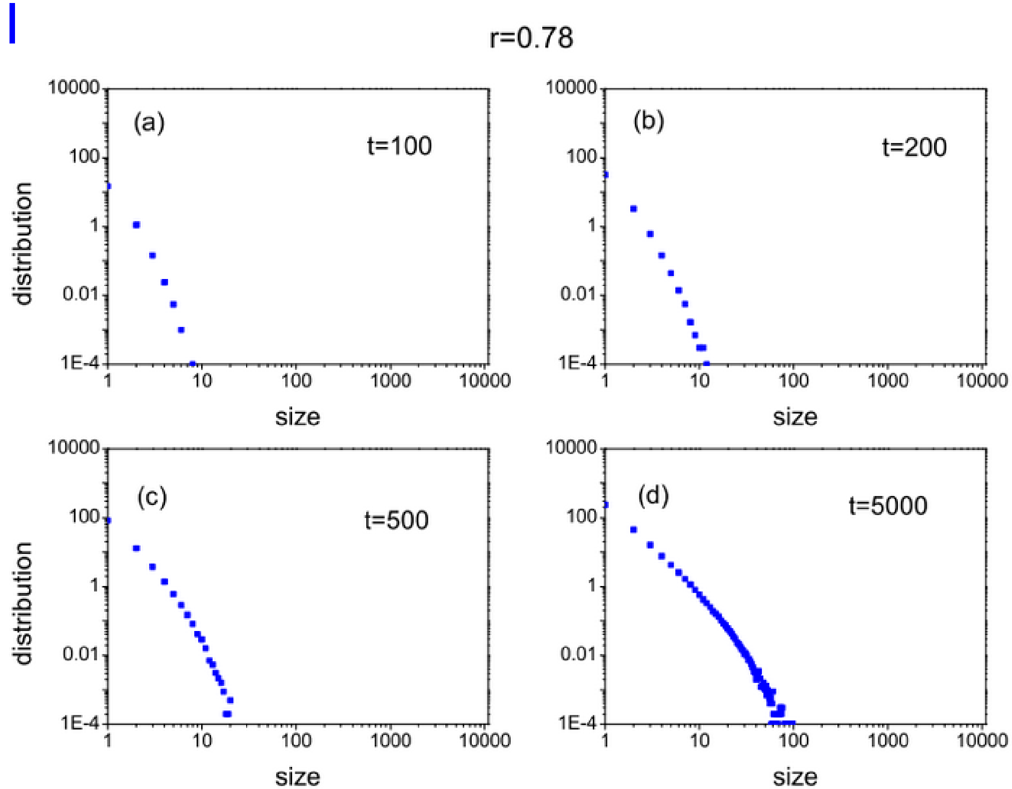


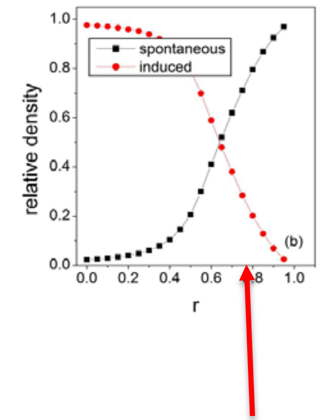


$$r = 0.5, \qquad z = 7, \qquad \phi = 0.2, \qquad p = 5 \times 10^{-4}$$

Distribution of Induced Clusters ($r_\times < r < r^*$)

scenario II



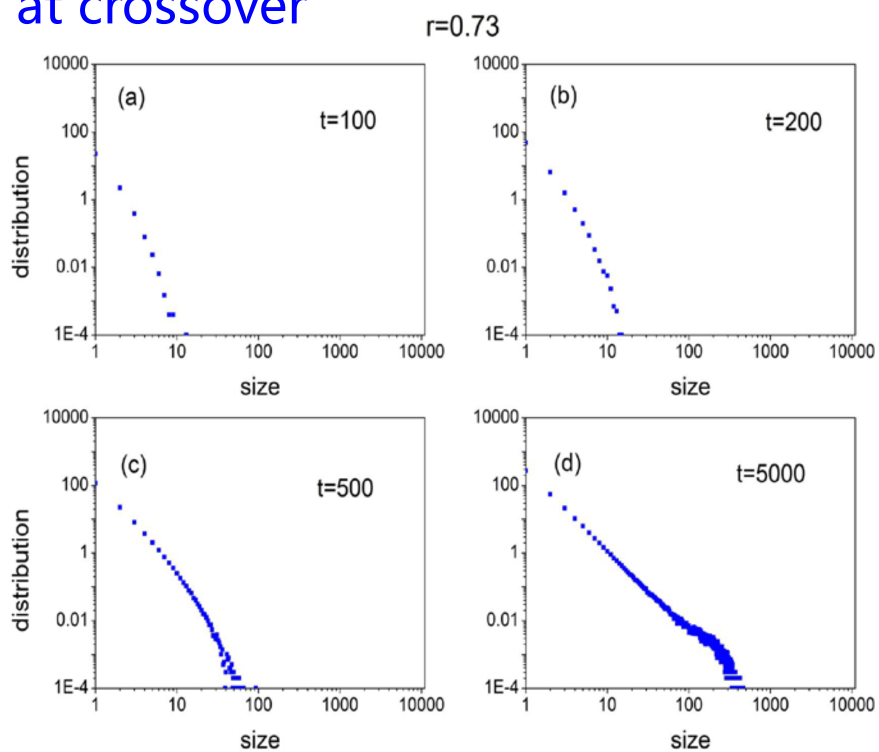


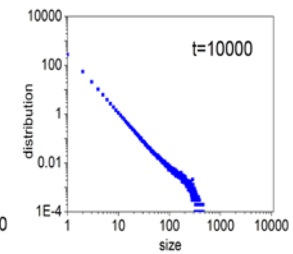
$$r^* = 0.86$$

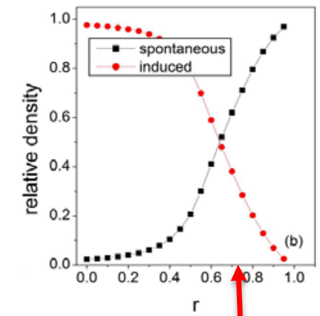
$$r = 0.78, \qquad z = 7, \qquad \phi = 0.2, \qquad p = 5 \times 10^{-4}$$

Distribution of Induced Clusters ($r \sim r_\times$)

scenario at crossover

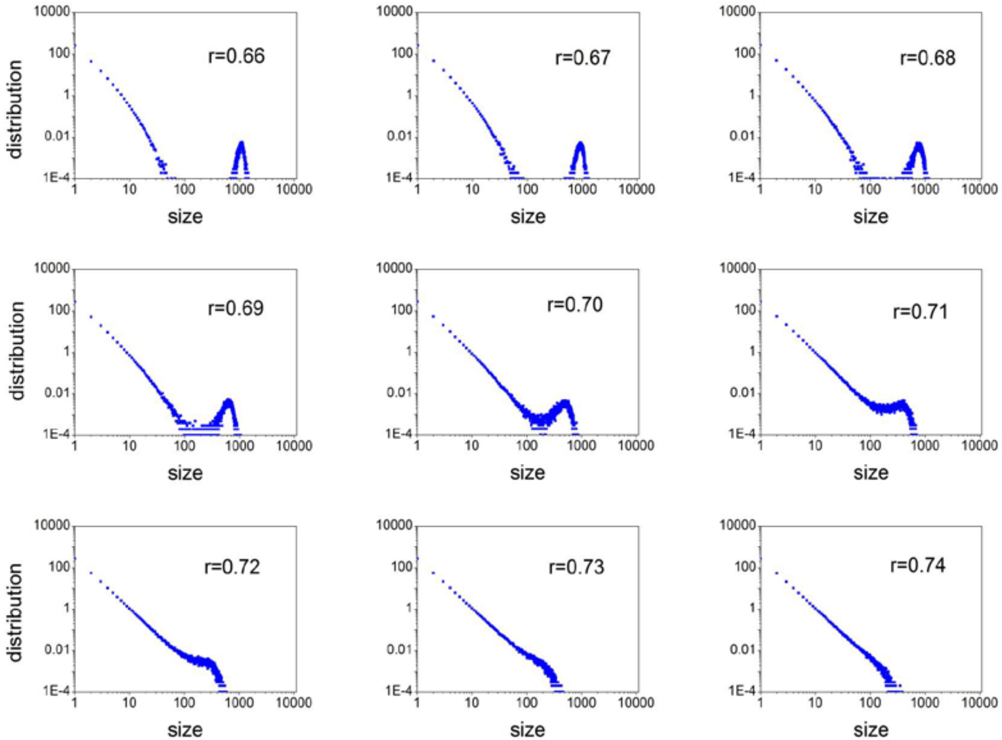






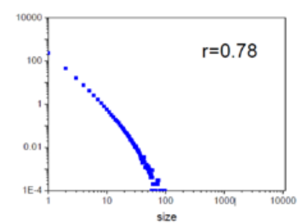
$$r = 0.73, \qquad z = 7, \qquad \phi = 0.2, \qquad p = 5 \times 10^{-4}$$

Asymptotic Distribution of Induced Clusters



$$r^* = 1 - 1/z = 0.86$$

percolation threshold



$$z = 7, \qquad \phi = 0.2, \qquad p = 5 \times 10^{-4}, \qquad t = 5000$$

Theoretical Treatment

$p > 0$. Dynamic treatment using extended Gleeson's* approach:

$$\frac{ds_{\mathbf{k},m}}{dt} = -F_{\mathbf{k},m}s_{\mathbf{k},m} - \beta_s(k-m)s_{\mathbf{k},m} + \beta_s(k-m+1)s_{\mathbf{k},m-1}$$

$\mathbf{k} = (k, c)$, with (k degree, c state variable);

$c = 0$, if node is blocked and $c = \phi$ otherwise

m : # adopter neighbors

$s_{\mathbf{k},m}$: # nodes with $(\mathbf{k}, m)$

$F_{\mathbf{k},m}$: prob. per unit time that a $(\mathbf{k}, m)$ node adopts;

$$F_{\mathbf{k},m} = \begin{cases} p & \text{if} \quad m < k\phi \\ 1 & \text{if} \quad m \geq k\phi \end{cases}, \qquad \forall m \text{ and } k, c \neq 0$$

β_s: rate that an s-s pair transforms to an s-a pair

$$\beta_s(t) = \frac{\sum_{\mathbf{k}} P_{\mathbf{k}} \sum_m (k-m) F_{\mathbf{k},m} s_{\mathbf{k},m}(t)}{\sum_{\mathbf{k}} P_{\mathbf{k}} \sum_m (k-m) s_{\mathbf{k},m}(t)}$$

*J. P. Gleeson, Phys. Rev. Lett. 107, 068701 (2011).

Theoretical Treatment

Initial condition:

$$s_{\mathbf{k},m}(0) = [1-\rho_{\mathbf{k}}(0)]B_{k,m}[\rho(0)]$$ with $$B_{k,m}(\rho) = \binom{k}{m}\rho^m(1-\rho)^{k-m}$$

Solution in terms of integral variables:

$$\rho(t) = 1 - \sum_{\mathbf{k}} P_{\mathbf{k}} \sum_m s_{\mathbf{k},m}(t)$$ density of adopters at time t

$$\nu(t) = \sum_{\mathbf{k}} P_{\mathbf{k}} \sum_m m s_{\mathbf{k},m}(t) / \sum_m k s_{\mathbf{k},m}(t)$$

Prob that random neighbor of an s node is s

Solution Ansatz: $s_{\mathbf{k},m}(t) = [1-\rho_{\mathbf{k}}(0)]B_{k,m}[\nu(t)]e^{-pt}$ for $m < k\phi$

leading to

$$\boxed{\begin{aligned}\dot{\rho} &= h(\nu,t) - \rho \\ \dot{\nu} &= g(\nu,t) - \nu\end{aligned}}$$

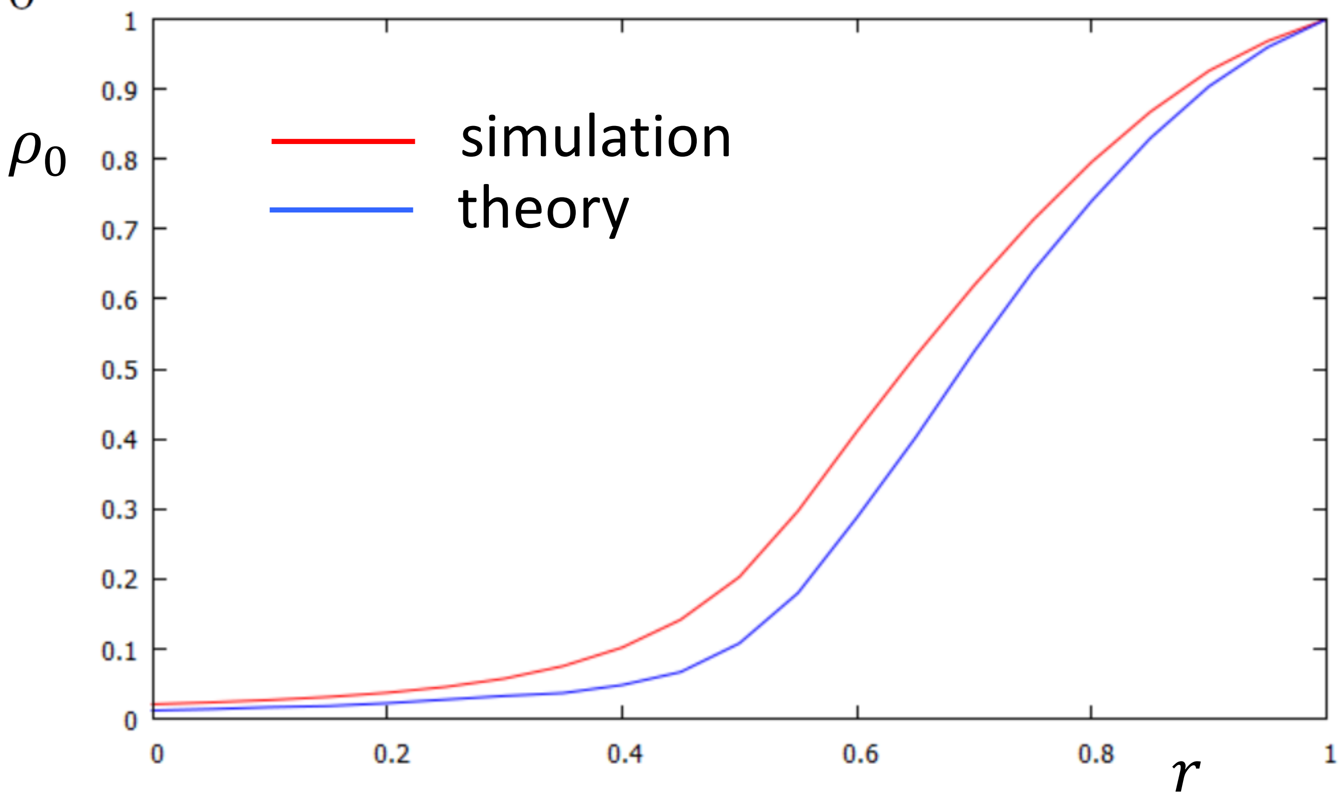
with $h(\nu), g(\nu)$ complicated, explicit functions of P_k, P_c and p.

Theoretical Treatment

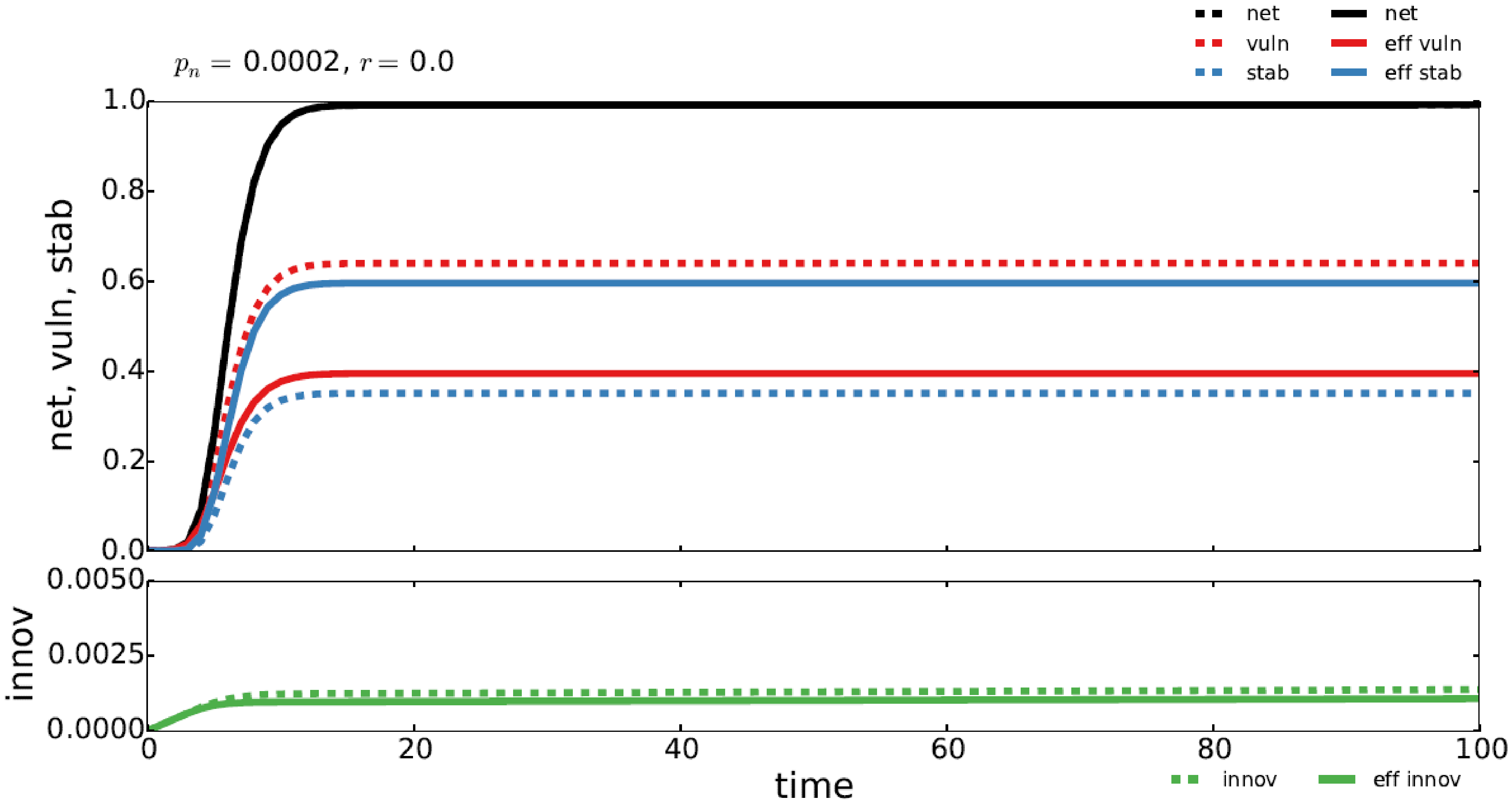
From $\rho(t) \rightarrow \rho_0(t)$: the fraction of spontaneous innovators

$\dot{\rho}_0 = p\rho_s$ with ρ_s fraction of susceptible. Using $1 - \rho = r + \rho_s$

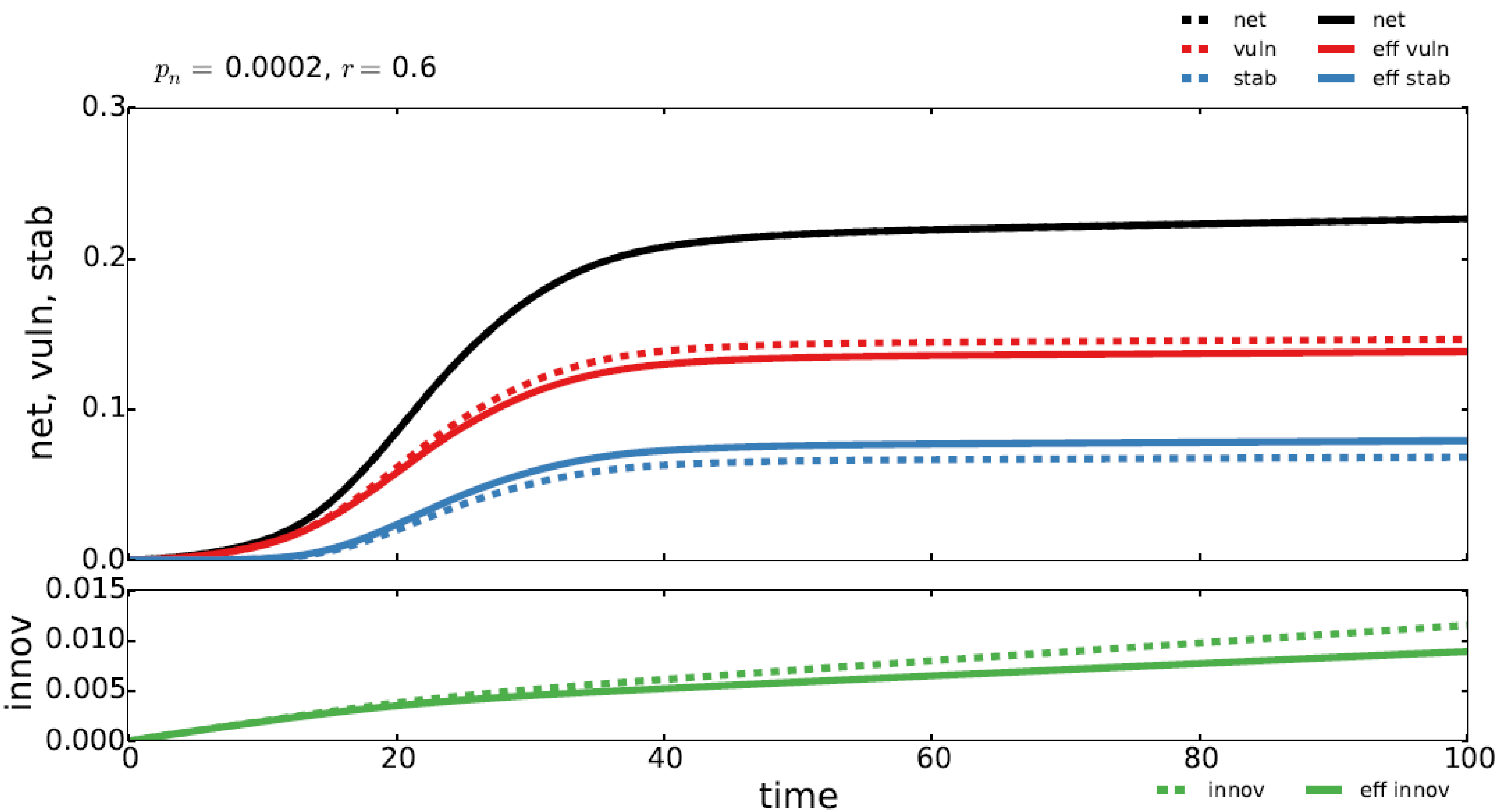
$$\rho_0(t) = p\int_0^t [1 - r - \rho(t)]dt$$



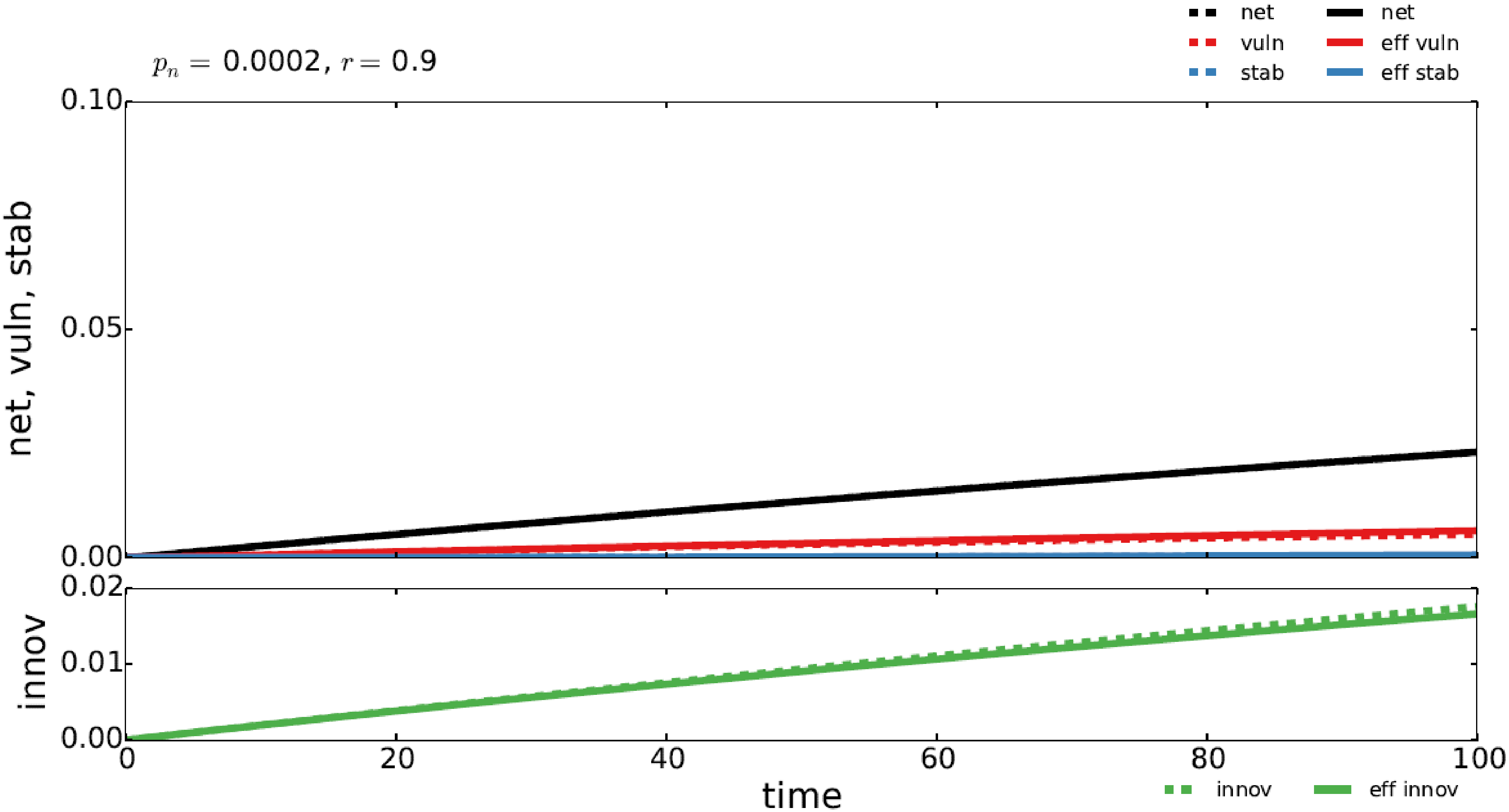
Model Calculations



Model Calculations



Model Calculations



Instead of anecdotes: Big Data

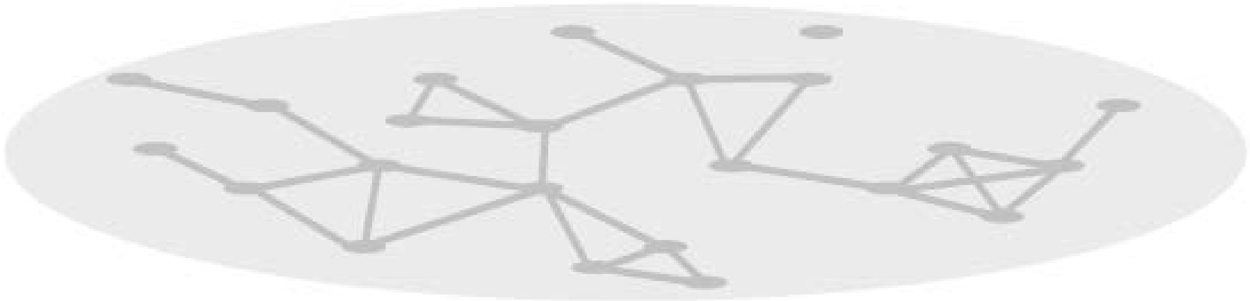
Information about:

- Basic service network
- Adoption of additional services
- Data about location (IP)

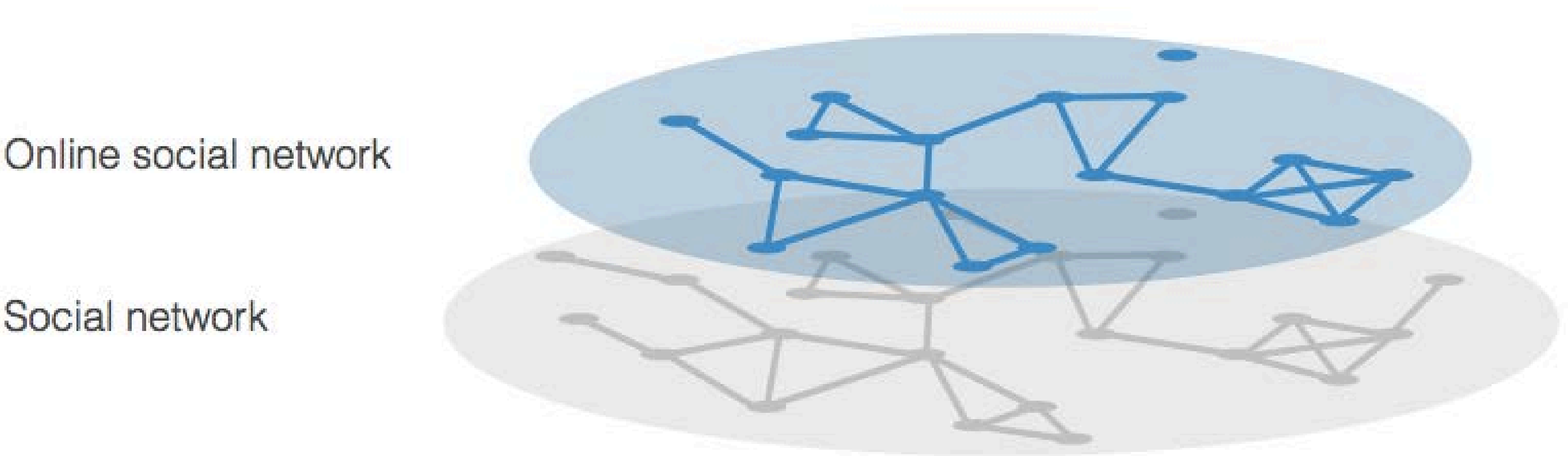
- 700+ million users world-wide
  - September 2003 - March 2011 (2738 days)
  - Registration dates
  - Location & self reported demographic data
  - Spamming accounts are removed
- Link creation dynamics
  - Time stamped link addition events
  - Only confirmed links
- Free and Payed services
  - 6 free and 9 payed services
  - Time of adoption
  - Usage activity sequences
- Country networks
  - For calculations we selected users in single countries
  - For selected users we considered all first neighbors
  - Look for the behaviour of country users only

Empirical Results

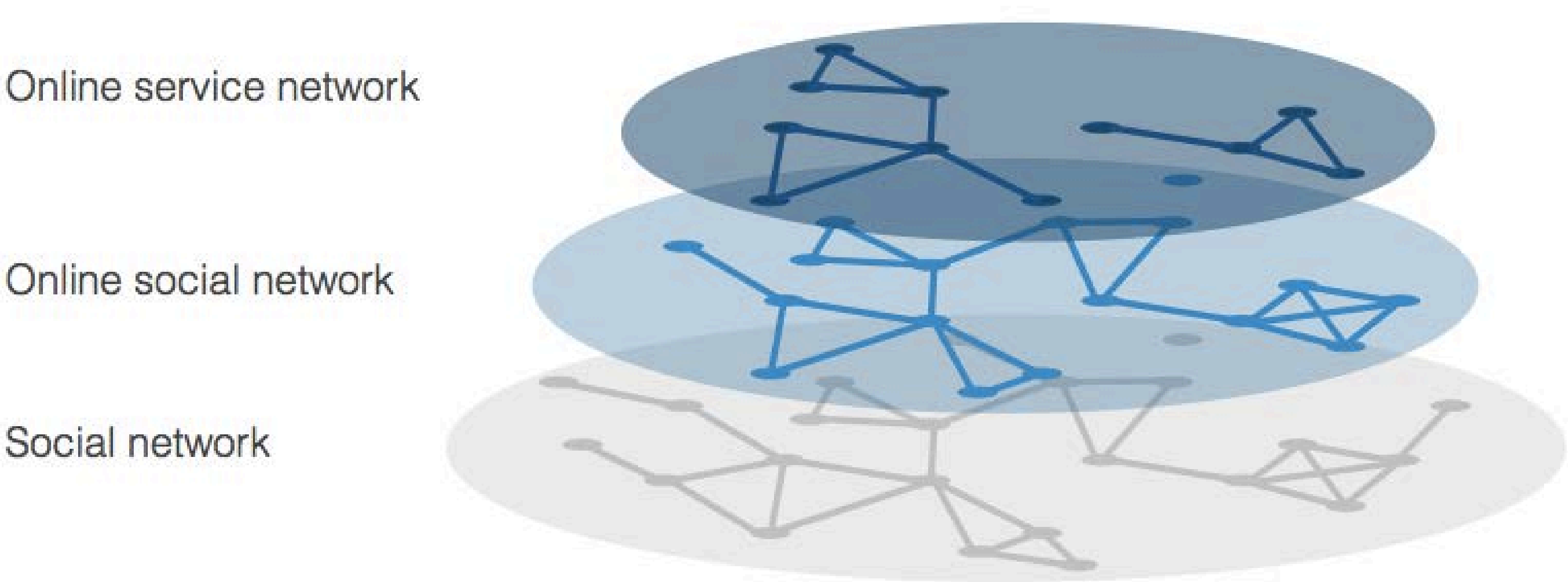
Social network



Empirical Results



Empirical Results



unknown

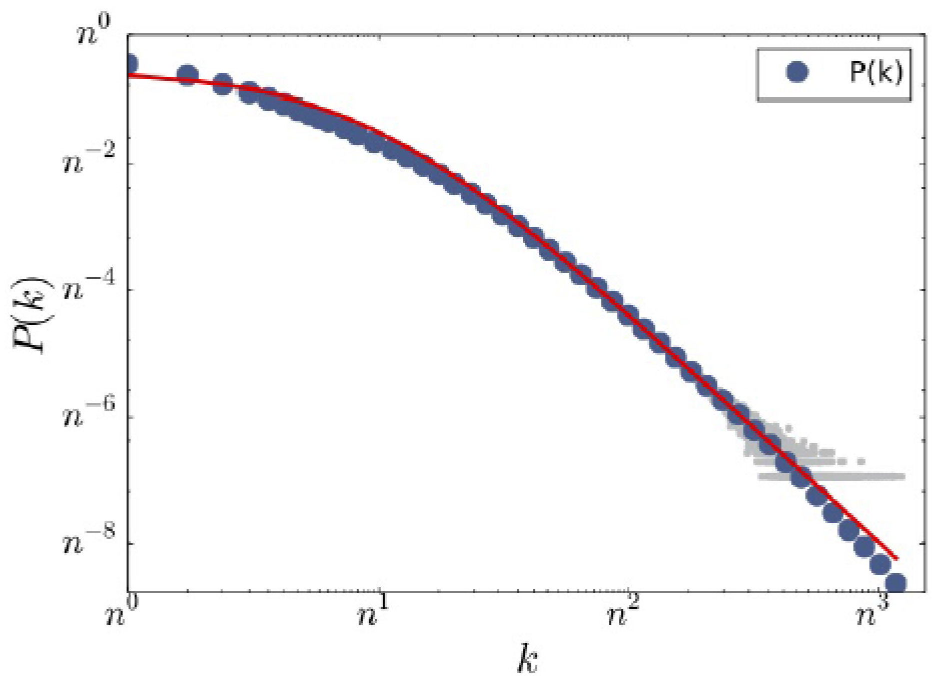
Empirical Results

Spreading of online service on the OSN

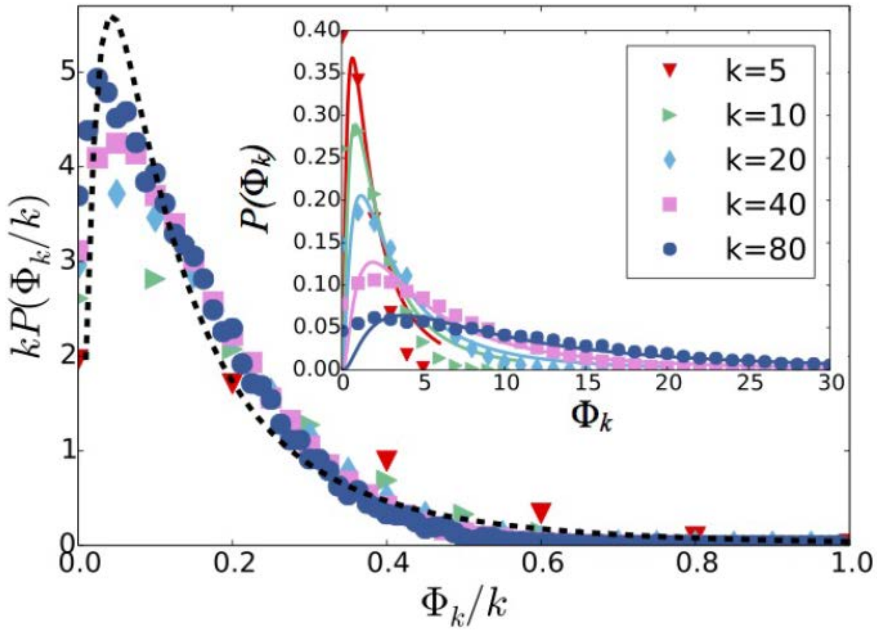
Here we know the underlying network: 520 M nodes of the Voice over Internet service.

r=0.95. The network is NOT ER, broad degree distribution.

Broad degree distribution

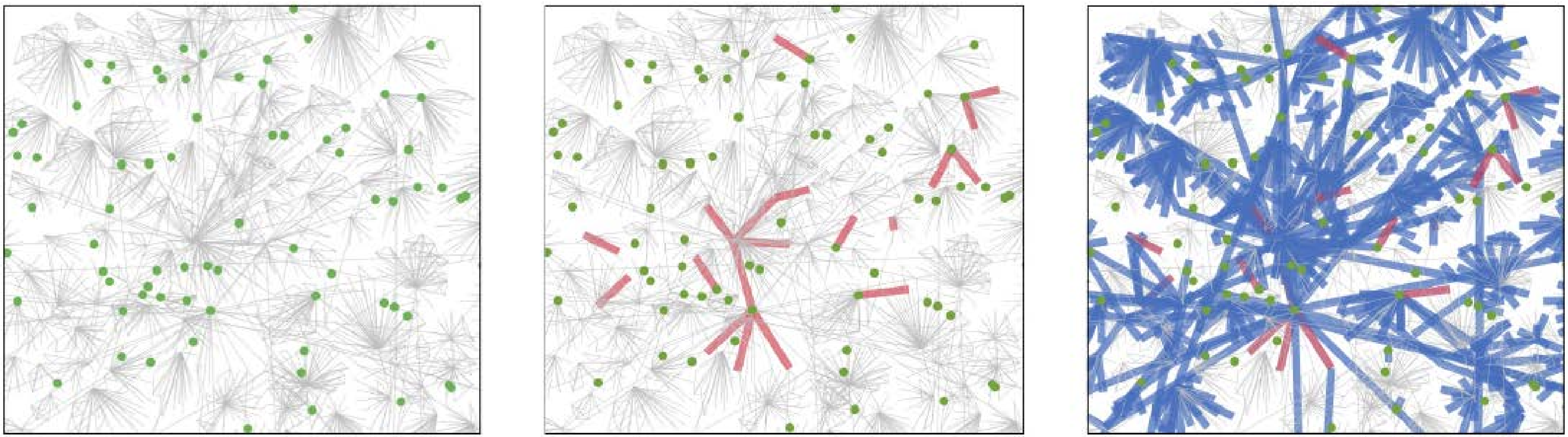


Empirical Results



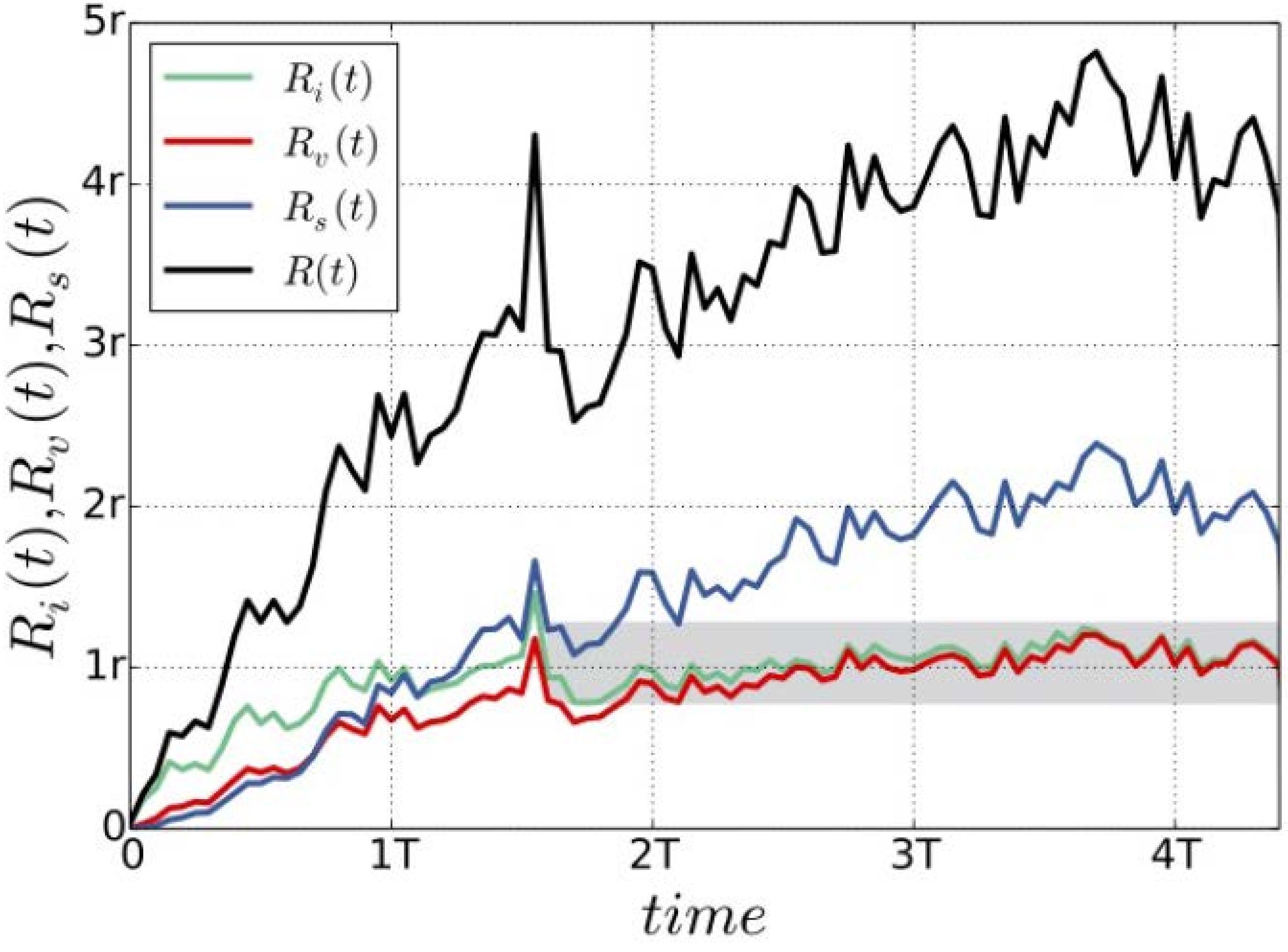
Empirical threshold distribution: log-normal with $\langle \phi \rangle = 0.19$
ϕ proper variable!

Empirical Results



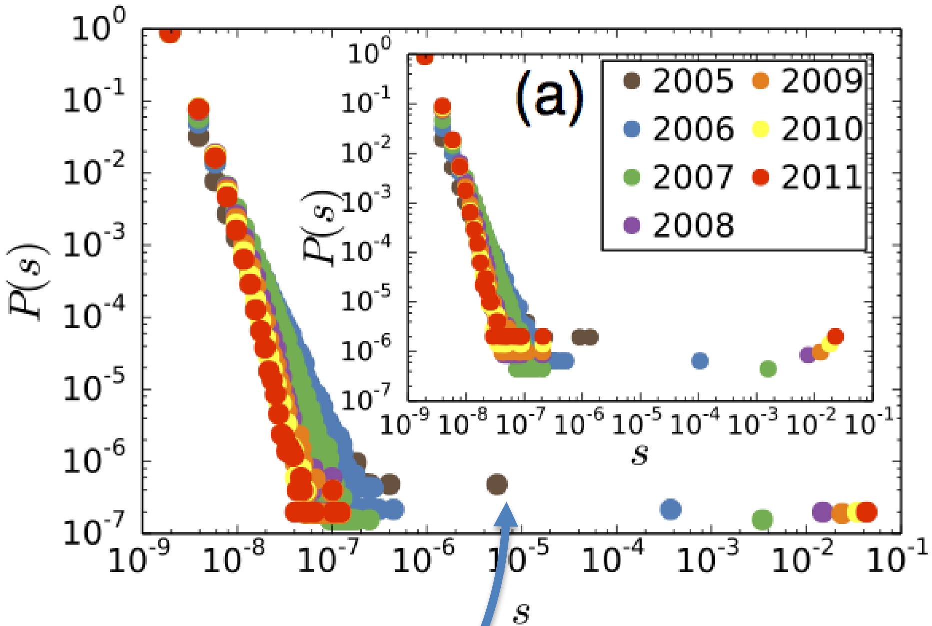
Initiators vulnerable clusters adopters

Empirical Results



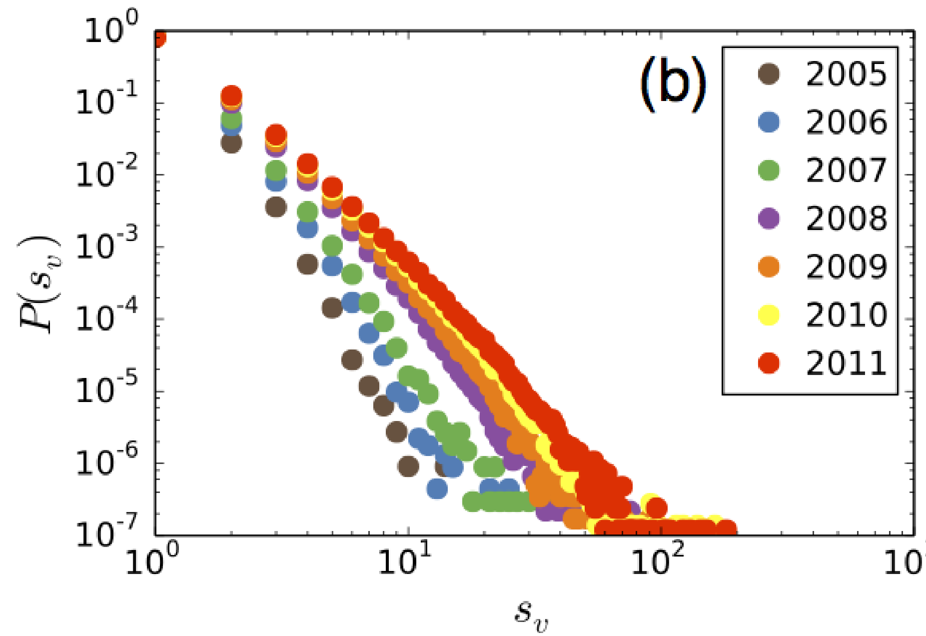
Rates

Empirical Results



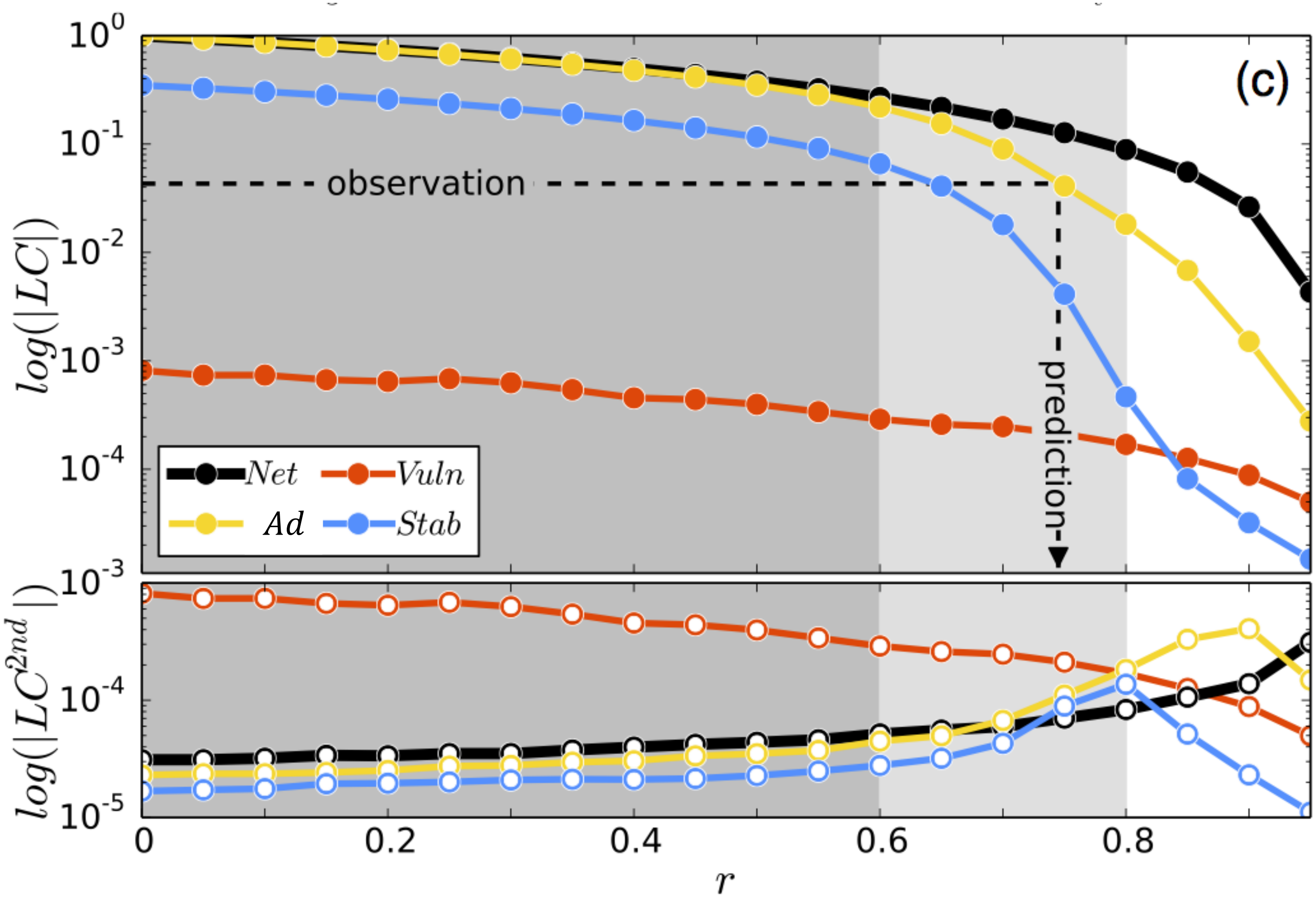
Size distribution of adopter clusters. Inset: stable adopters.

Giant component



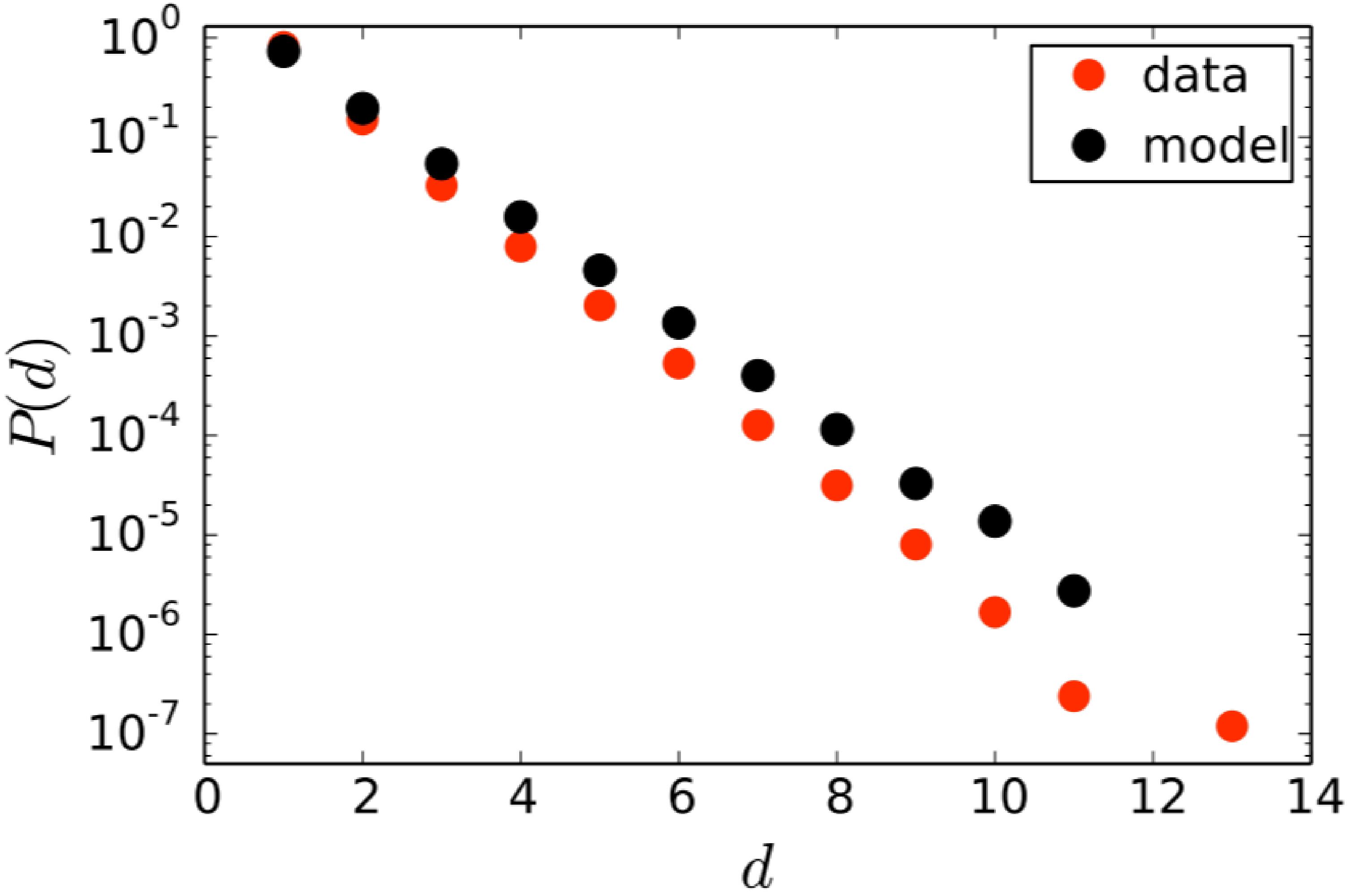
Size distribution of innovator induced vulnerable trees

Empirical Results – Comparison with Model



Model calculation with empirical threshold and degree distributions and evolution time. The density r_{emp} is determined from the plot: $r_{\mathrm{emp}} = 0.745$.

Empirical Results – Comparison with Model

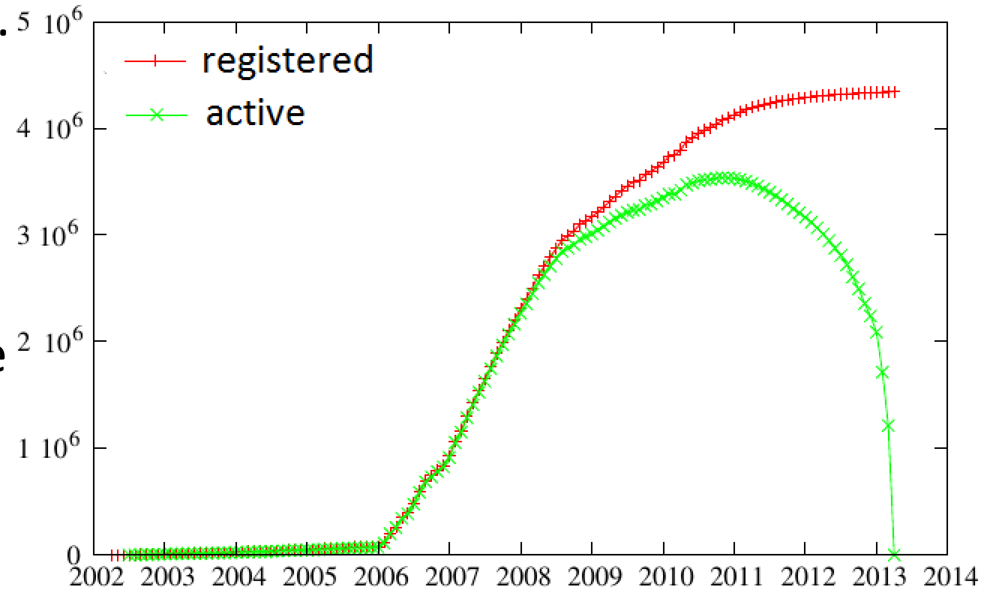


Distribution of depth of vulnerable trees (# generations)

CASCADIC COLLAPSE OF A NETWORK

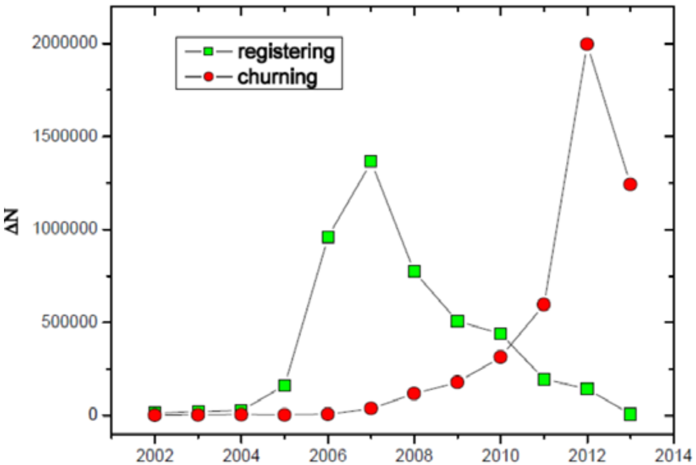
iWiW (originally: WiW, from Who is Who): Hungarian Online Social Network

Launched on April 14, **2002** (Facebook: 2004; worldwide: 2005)
2006: Acquired by a subsidiary of Hungarian Telekom → Deutsche Telekom (60%)
Early stage: Registration only by invitation. Find friends.
New members got invitation voucher(s).
After **2011**: unconditional.
Most visited site in Hungary 2005-2010
At peak 4.3 Mio registered users
(10 Mio Hungarians in Hungary +
~5 Mio in neighbor countries and in the
diaspora). ¾ of internet population.
Name, age, gender, location, school…
Closed June 30, **2014**



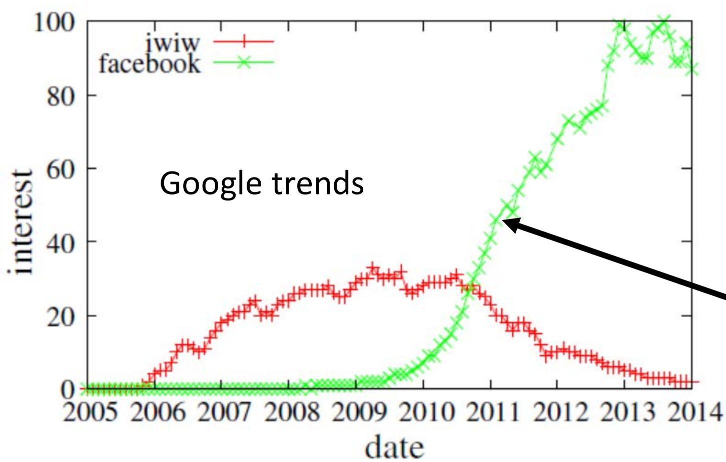
J. Török Z. Ruan, JK (unpublished)

DECLINE



Two origins of churning:

- External info (rise of Facebook)
- Pear pressure (not enough friends in the network)



The decline of iWiW was caused mainly by the rise of FB.

Hungarian version launched July 2008

~linear increase of interest

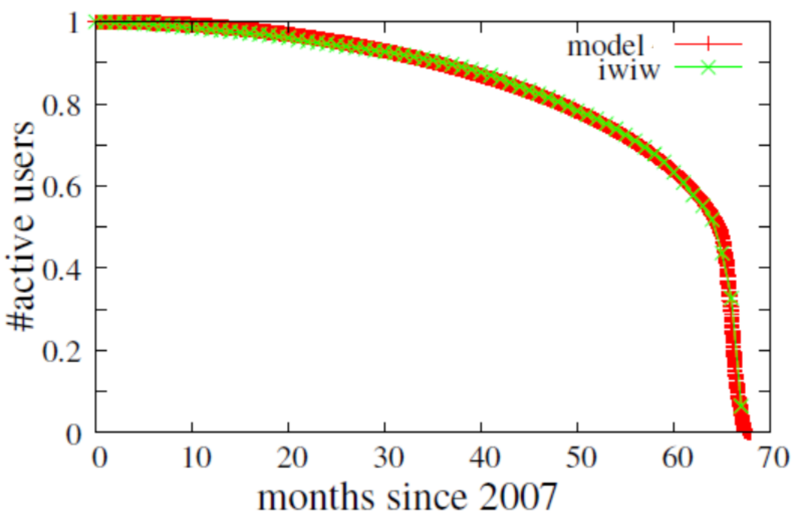
DECLINE

Cascade model of decline

- Assign a threshold $R_i = 45 \pm 10\%$ to each nodes
- Delete nodes with a linearly increasing rate: $\gamma = \mu t$
- If the ratio of a node i's alive neighbors < R_i delete it
- The created avalanches are treated instantaneously

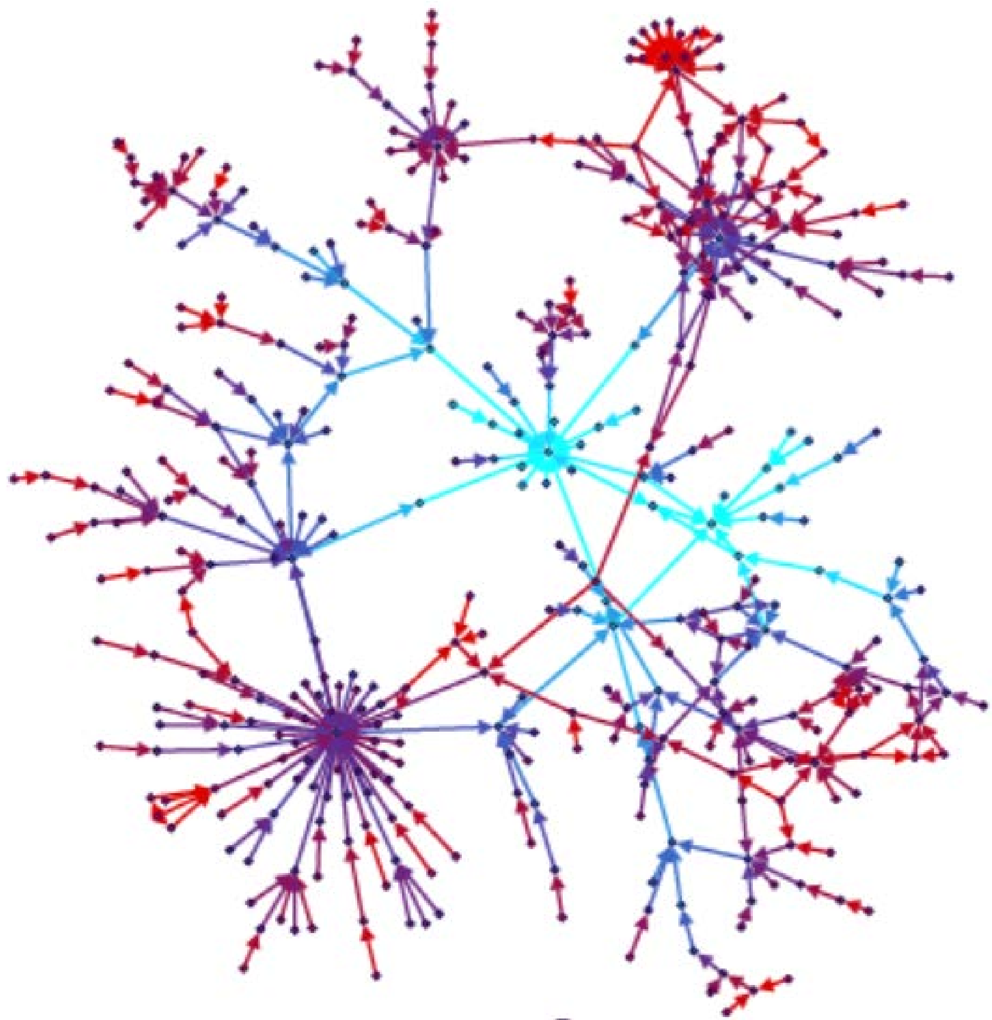
μ sets the time scale of the process, can be fitted.

Results on networks with
$\langle k_{\mathrm{ER}} \rangle = 6$ and $\langle k_{\mathrm{tc}} \rangle = 10$



CASCADES

In spite of the rapid process, only finite cascades always triggered by "spontaneous", i.e., externally driven events



Summary

- ICT based data help in understanding the laws of innovation spreading, an example of complex social contagion. Two levels of Skype data were used: Free and paid services
- Cascade model can be extended to describe the kinetics of spreading by inclusion of innovator rates and blocked nodes. Fast and slow regimes
- Generating function technique and general rate equation approaches were used to describe the model.
- Good agreement between empirical and model results was found. The spreading of paid service is relatively slow due to the large number of „blocked” individuals.