# Economic Geography and Social Networks Diffusion in networks

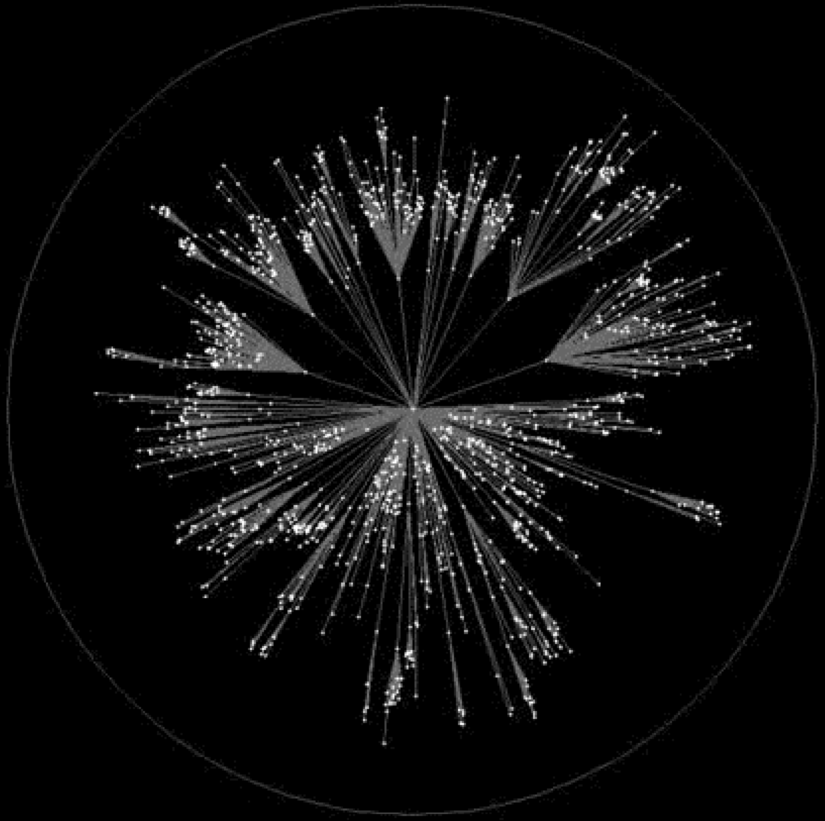
Lengyel Balázs

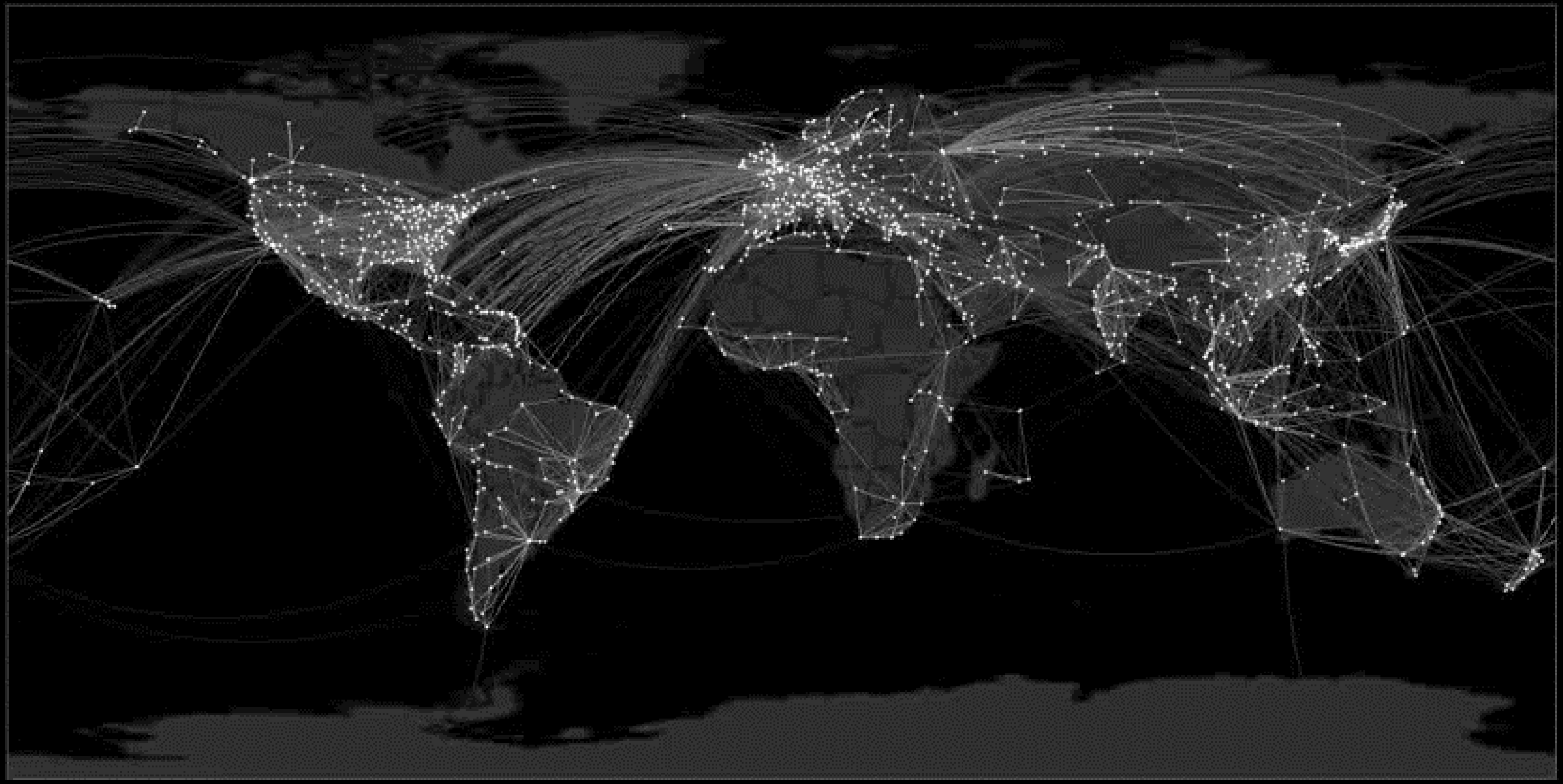
ANETI Lab, HUN-REN KRTK and CIAS

Institute of Data Analytics and Information Systems, Corvinus University of Budapest

2024. May 14.

# Virus diffusion on the global flight network

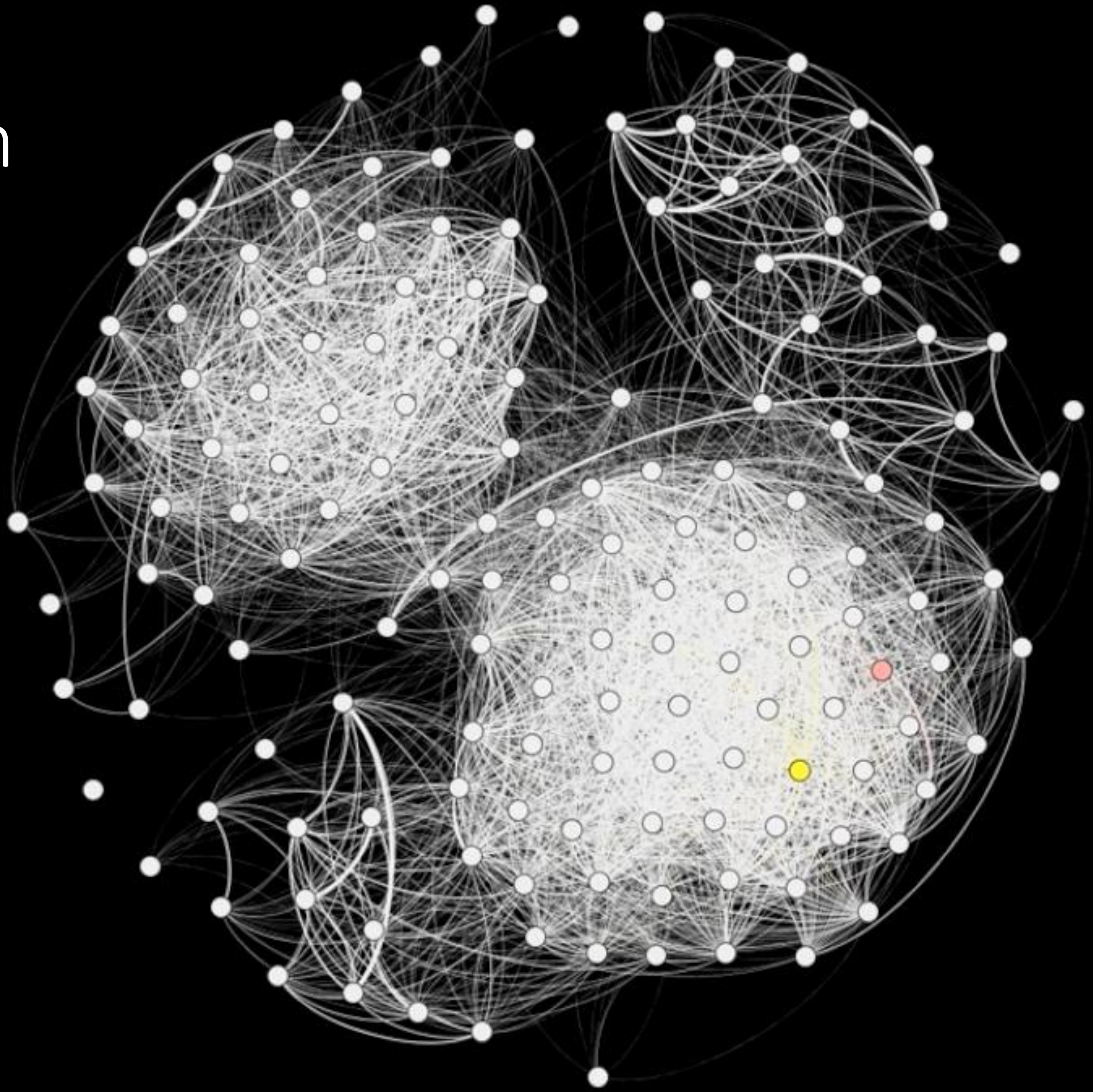




Brockman D., Helbing D (2013) The hidden geometry in complex, network-driven contagion phenomena. Science 342 (6164) pp. 1337-1342.

*Aplin et al. 2015. Nature 518.*
Doi:10.1038/nature13998

# Diffusion in networks



# Today

I. Diffusion in networks

1. Diffusion models: Bass modell {innovation} and SI models {virus} (SIS, SIR)

2. The impact of network structure on diffusion process

3. Simple versus Complex diffusion: spreading of viruses versus innovation

II. The spatial diffusion of innovation in networks

# 1. Diffusion models

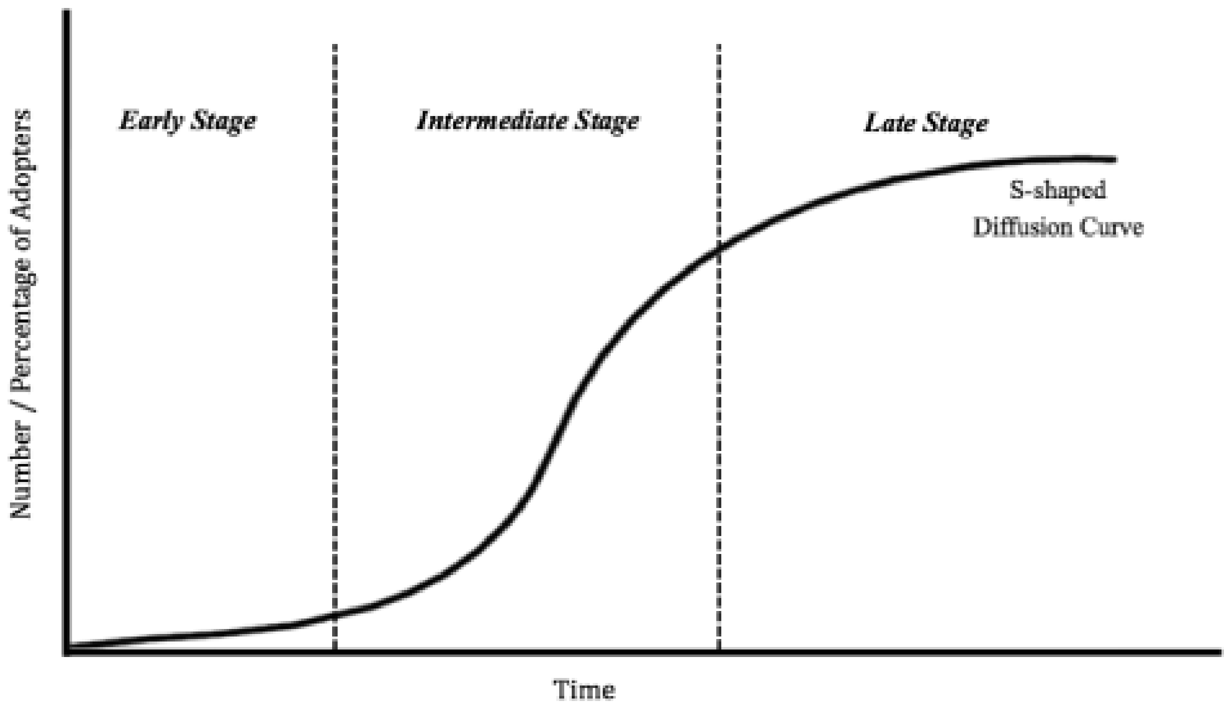
# The S curve of diffusion

Coleman, Katz, Menzel (1966): how fast do doctors prescribe doctors new medicins?

- diffusion: prescriptions
- network: survey information about advice links between doctors
- High degree doctors started to prescribe earlier.

Griliches hybrid corn example: how big fraction of farmers use it?

- Initially slow diffusion accelerates and slows down again.



# Bass model

Two states / activity: 0 or 1 – yes or no

- *F(t)* – fraction of the population who started the activity until time t.
- *p* is the rate of spontaneous adoption
- *q* is the rate of imitation

$$\frac{dF(t)}{dt} = (p + qF(t))\,(1 - F(t))$$

- Second term: fraction of population who have not adopted until time t.
- They can adopt spontaneously or follow the adopters.
- Solution to differential equation is:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

# Bass model

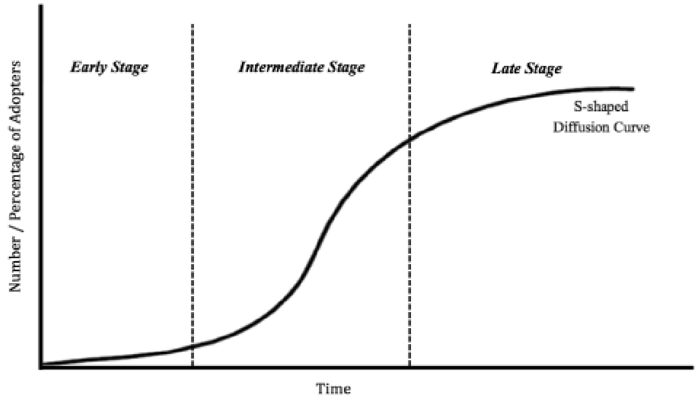
$$\frac{dF(t)}{dt} = (p + qF(t))\,(1 - F(t))$$

If *F(t)* is close to 1, diffusion slows down

- Second term goes to 0, growth goes to 0

In the beginning, only *p* matters, but later *q* becomes more important

- In the beginning *F(t)* is small, thus *qF(t)* goes to 0
- There is noone to adopt but anyone can adapt spontaneously
- The slope of spread is p in the beginning, but then q speeds this up.



# Bass model

The diffusion curve has S shape if $q>p$, that saturates at 1

If $F(t)$ is close to 1, then $d(F)/d(t) = 0$

If $F(t)=0$, then $d(F)/d(t)=p$

If $F(t)= \varepsilon$, then $d(F)/d(t) = (p + q\varepsilon)(1- \varepsilon)$

Convexity will be realized in the case:

$$(p + q\varepsilon)(1- \varepsilon) > p$$

$$(q)(1- \varepsilon) > p$$

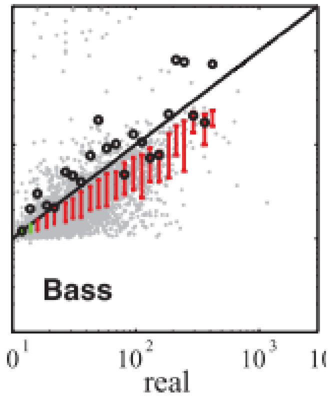
$$q > p$$

Usage:

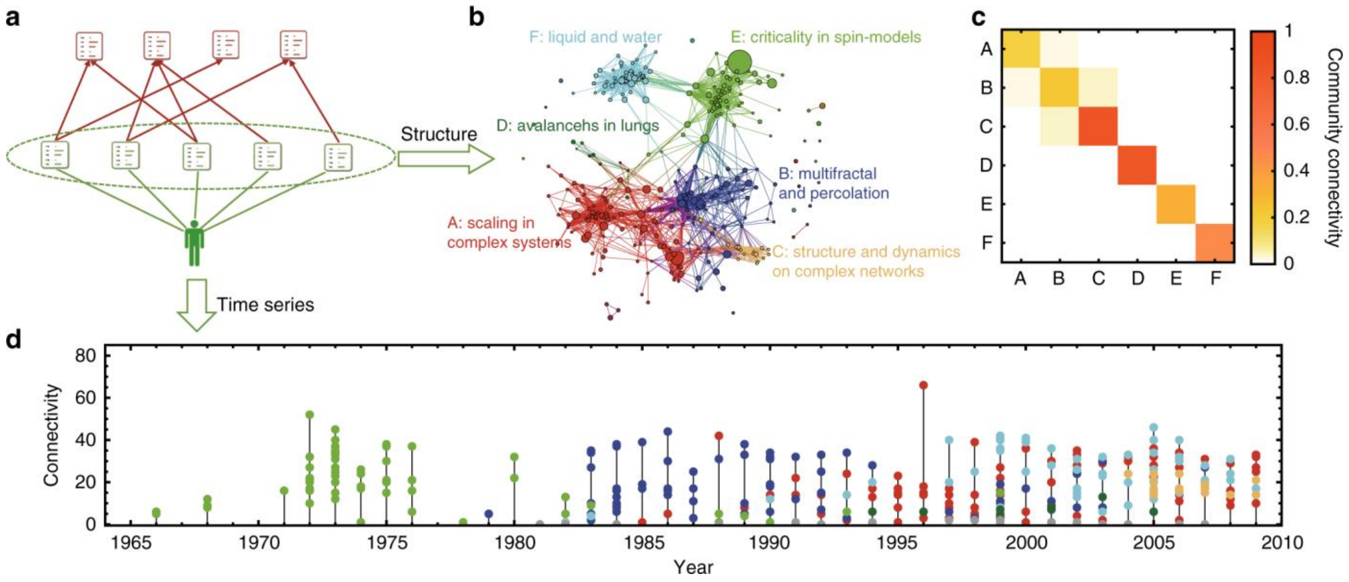
p and q can be estimated at small ε and then predictions can be made

# Bass model research to predict impact (the LI70 hypothesis)

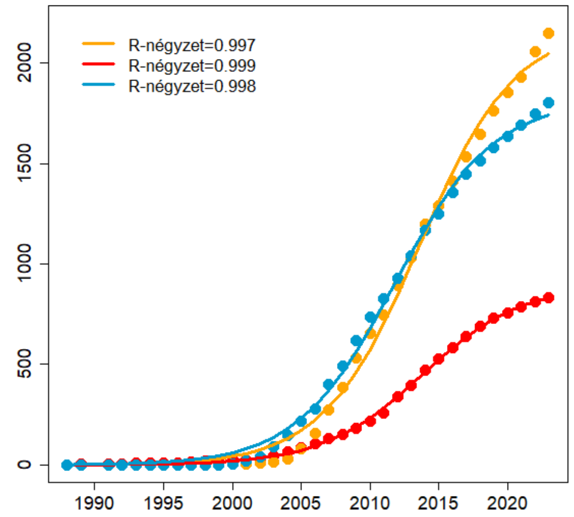
p and q can be estimated early and then predictions can be made




Wang et al. 2013 *Science*




Zeng et al. 2019 *Nature Communications*

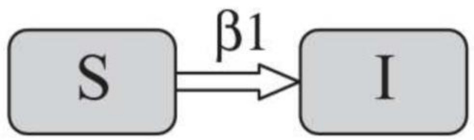



***LI70 hypotesis:** „Changes in research topics along the academic career make the prediction of success more difficult.”*

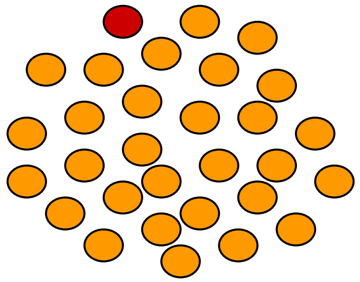
# SI model

- N: size of population
- S: susceptible population. Those who can be potentially infected.
- I: infected population. Those who can infect others.
- S(t): susceptibles at time t.
- I(t): infected at time t.
- β: probability of transmission

$$N = S(t) + I(t)$$



# SI model



- Each individual has <k> contacts with randomly chosen others individuals per unit time.

- The likelihood that the disease will be transmitted from an infected to a healthy individual in a unit time: β

If there are $I$ infected individuals and $S$ susceptible individuals, the average rate of new infection is $bSI / N$

$$\beta\langle k\rangle \frac{S(t)I(t)}{N} dt.$$

$$\frac{dI(t)}{dt} = \beta\langle k\rangle \frac{S(t)I(t)}{N}.$$

# SI model

- Homogenous mixing
- In every period t Infected individual will infect βS individual.
- In t+1, βIS individual will be infected.
- $i_0$ is the number of Infected at t=0.

$$\frac{dI(t)}{dt} = \beta\langle k\rangle\frac{S(t)I(t)}{N}.$$

$$s = S/N, \qquad i = I/N$$

$$\frac{di}{dt} = bsi = bi(1 - i)$$

$\beta\langle k\rangle$ is called the *transmission rate* (or transmissibility).

$$\frac{di}{i} + \frac{di}{(1-i)} = \beta\langle k\rangle dt. \qquad \ln i - \ln(1-i) + c = \beta\langle k\rangle t.$$

$$i(t) = \frac{i_0 \exp(bt)}{1 - i_0 + i_0 \exp(bt)}$$

# SI model

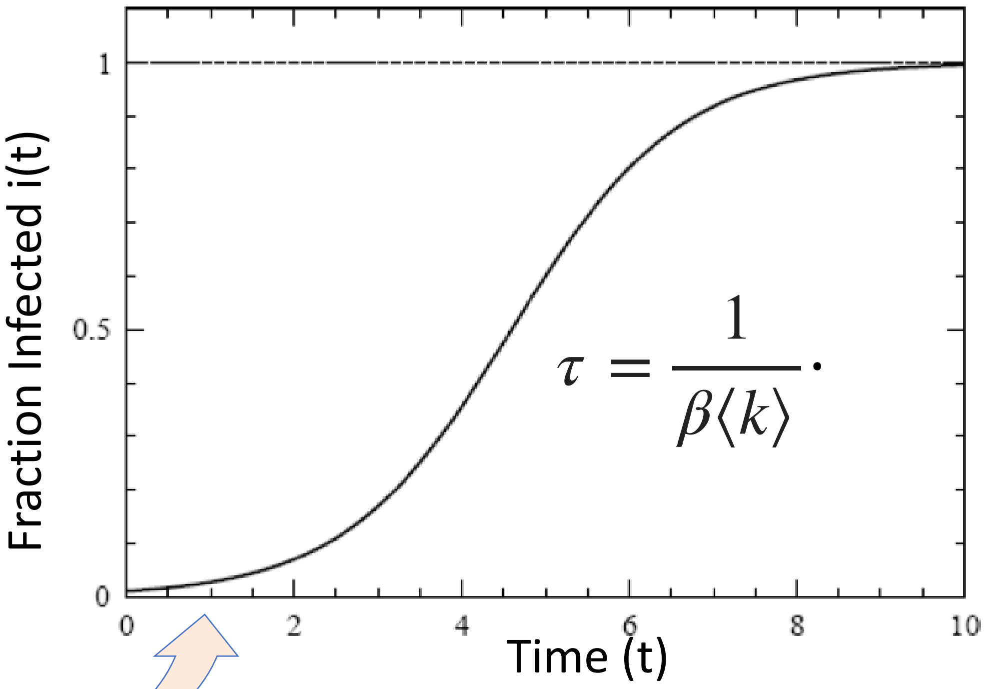
Recovered individuals become Susceptibles again.

If *i(t)* is small,

$$\frac{di}{dt} \approx \beta i$$

$$i \approx i_0 \exp(\beta t)$$

**exponential outbreak**




As i(t)→ 1.

$$\frac{di}{dt} \to 0$$

**saturation**

**SI model:** the fraction infected increases until everyone is infected.

# SIS model

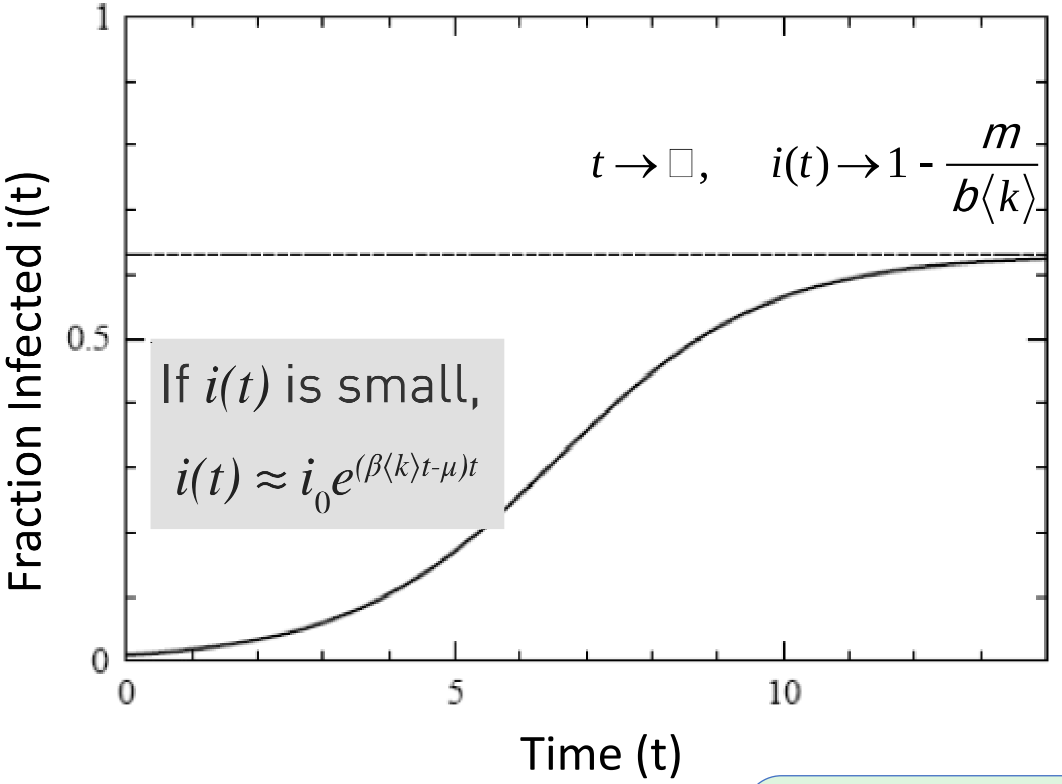
$$\frac{di}{dt} = \beta\langle k\rangle i(1-i) - \mu i$$

I S I→S

$$i = \left(1 - \frac{\mu}{\beta\langle k\rangle}\right)\frac{Ce^{(\beta\langle k\rangle-\mu)t}}{1 + Ce^{(\beta\langle k\rangle-\mu)t}}.$$

**Endemic state** ($\mu<\beta\langle k\rangle$):

**Disease-free state** ($\mu>\beta\langle k\rangle$):




Stationary state:

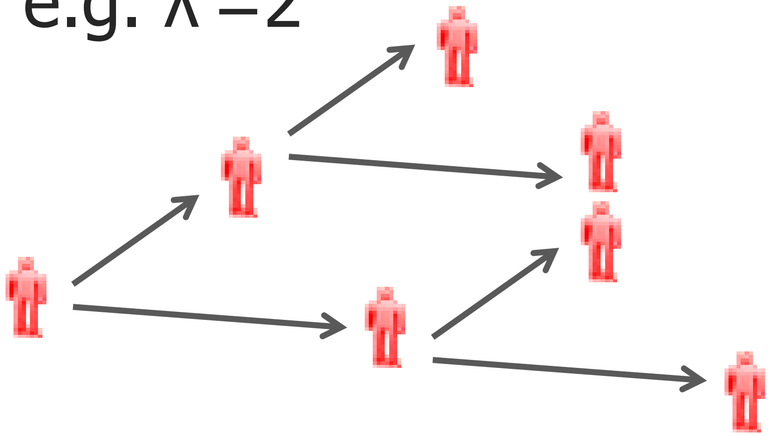
$$\frac{di}{dt} = \beta i(1-i) - \mu i = 0$$

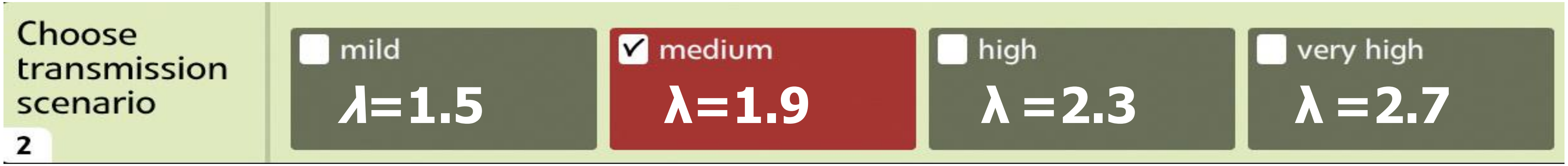
# SIS model

**reproductive number $R_0$ (λ):** average # of infectious individuals generated by one infected in a fully susceptible population.

$$R_0 = \frac{\beta\langle k\rangle}{\mu}.$$

e.g. λ =2



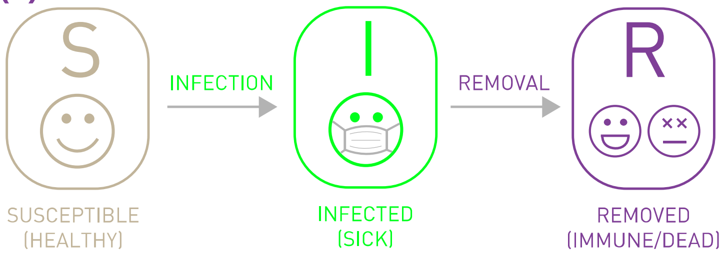


# SIR model

- Besides I and S, individuals have R (recovery/removal) state as well.
- Individuals cannot transform from R to I or S.

- r is the recovery probability in t.

(a)



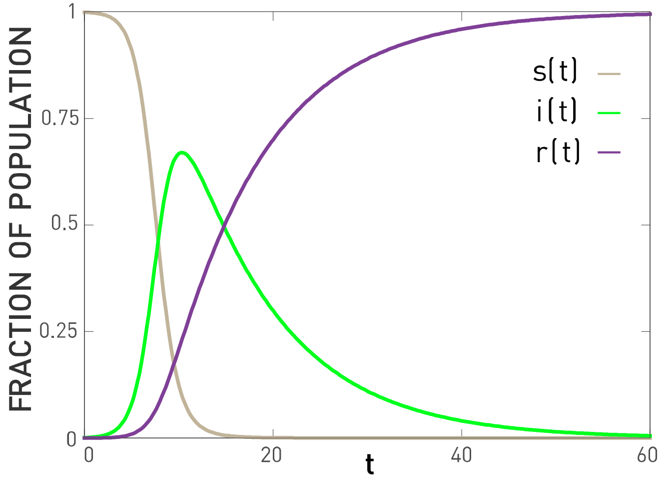

(b)

$$\frac{ds(t)}{dt} = -\beta\langle k\rangle i(t)[1 - r(t) - i(t)]$$

$$\frac{di(t)}{dt} = -\mu i(t) + \beta\langle k\rangle i(t)[1 - r(t) - i(t)]$$

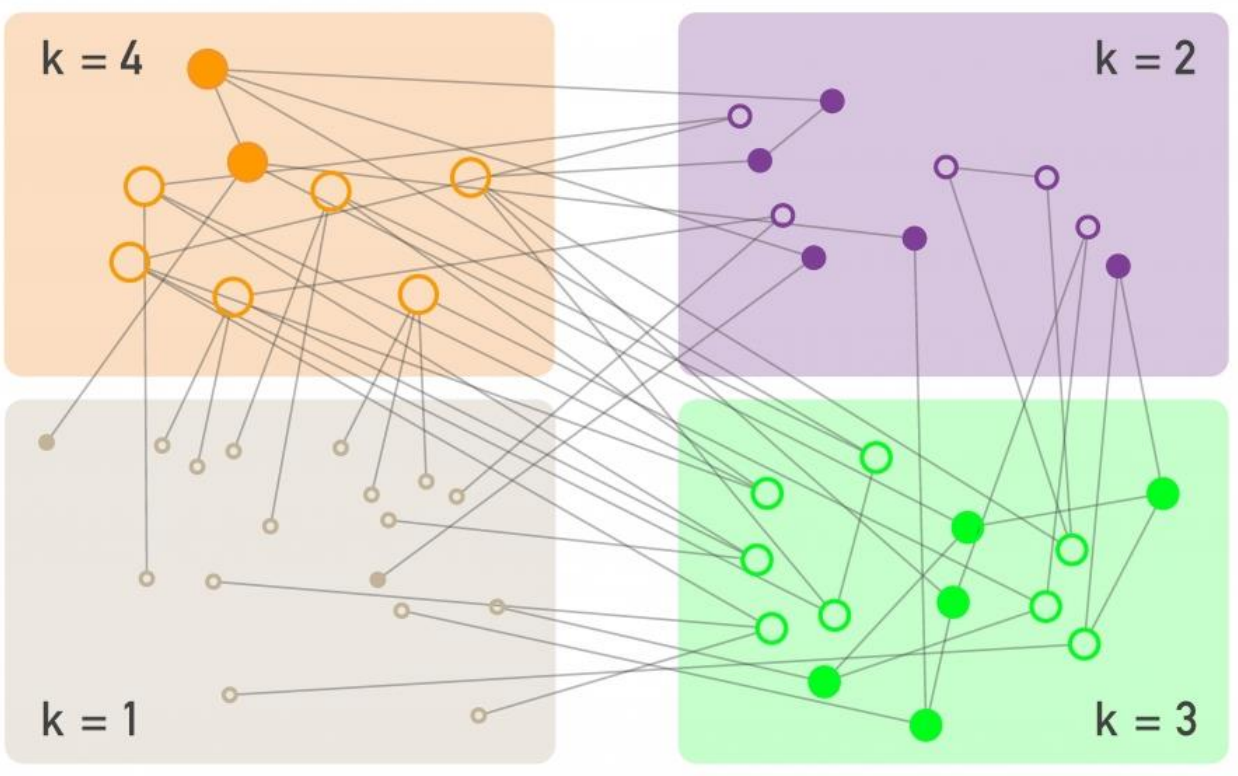
$$\frac{dr(t)}{\mathrm{dt}} = \mu i(t).$$

(c)



# 2. The impact of network structure on diffusion process

# SIS model on a network: degree block approximation




Split nodes by their degrees

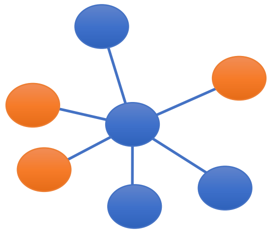
$$i_k = \frac{I_k}{N_k}, \quad i = \sum_k P(k) i_k$$

SIS model:

$$\frac{di_k(t)}{dt} = \beta(1 - i_k(t))k\Theta_k(t) - \mu i_k(t)$$

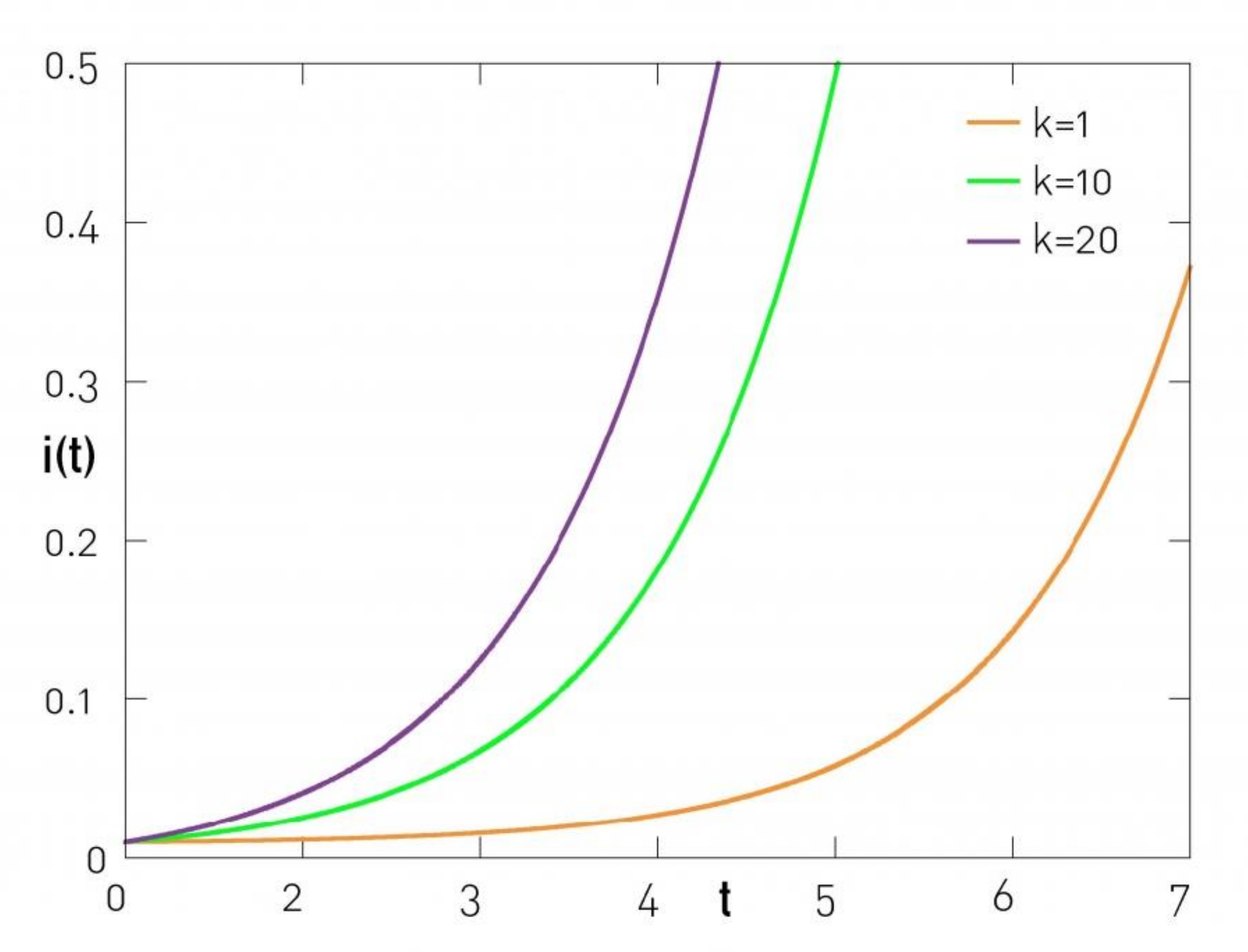
Proportional to *k*

Density of infected neighbors of nodes with degree *k*

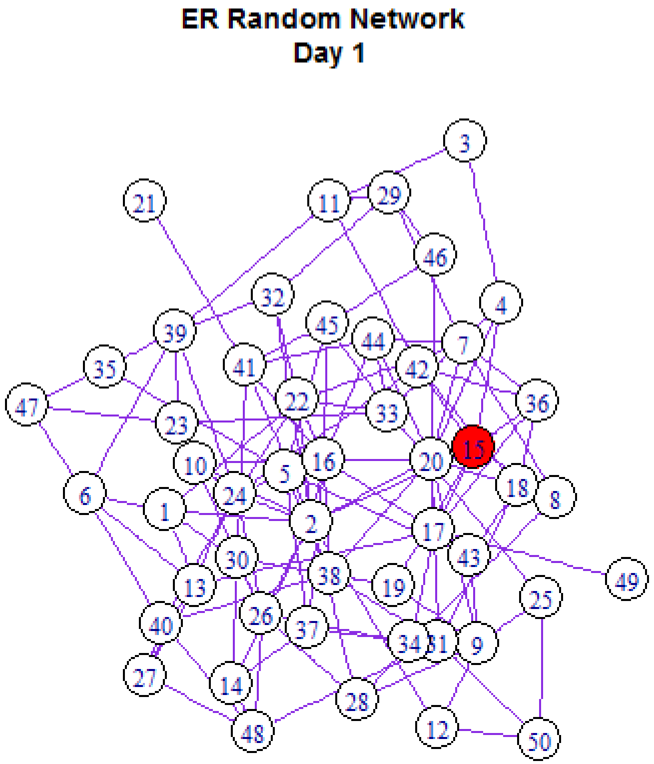


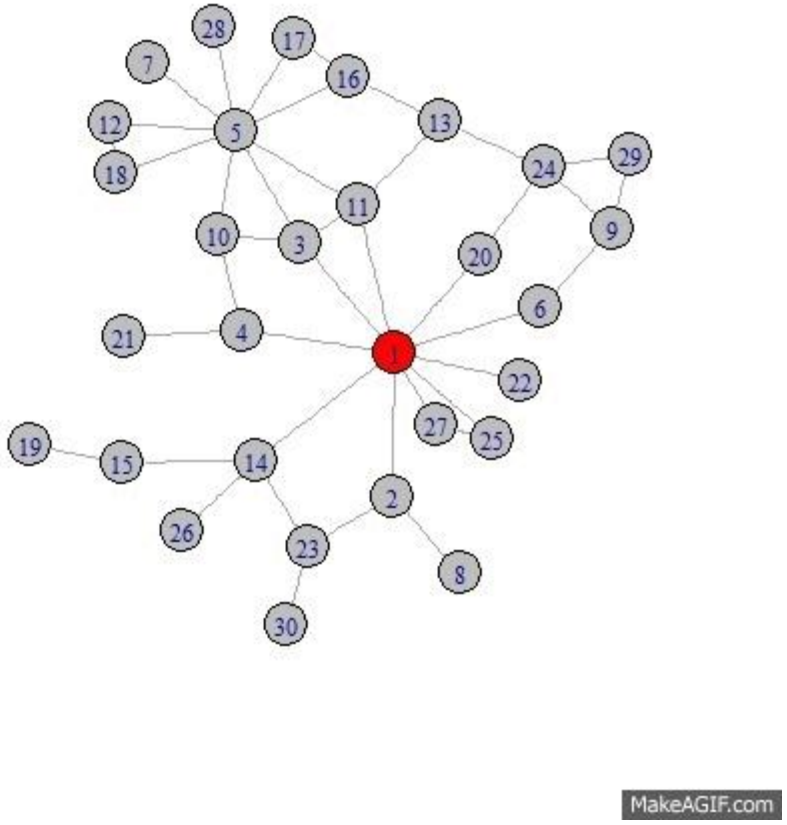
I am susceptible with *k* neighbors, and $\Theta_k(t)$ of my neighbors are infected.

# Degree is key in diffusion

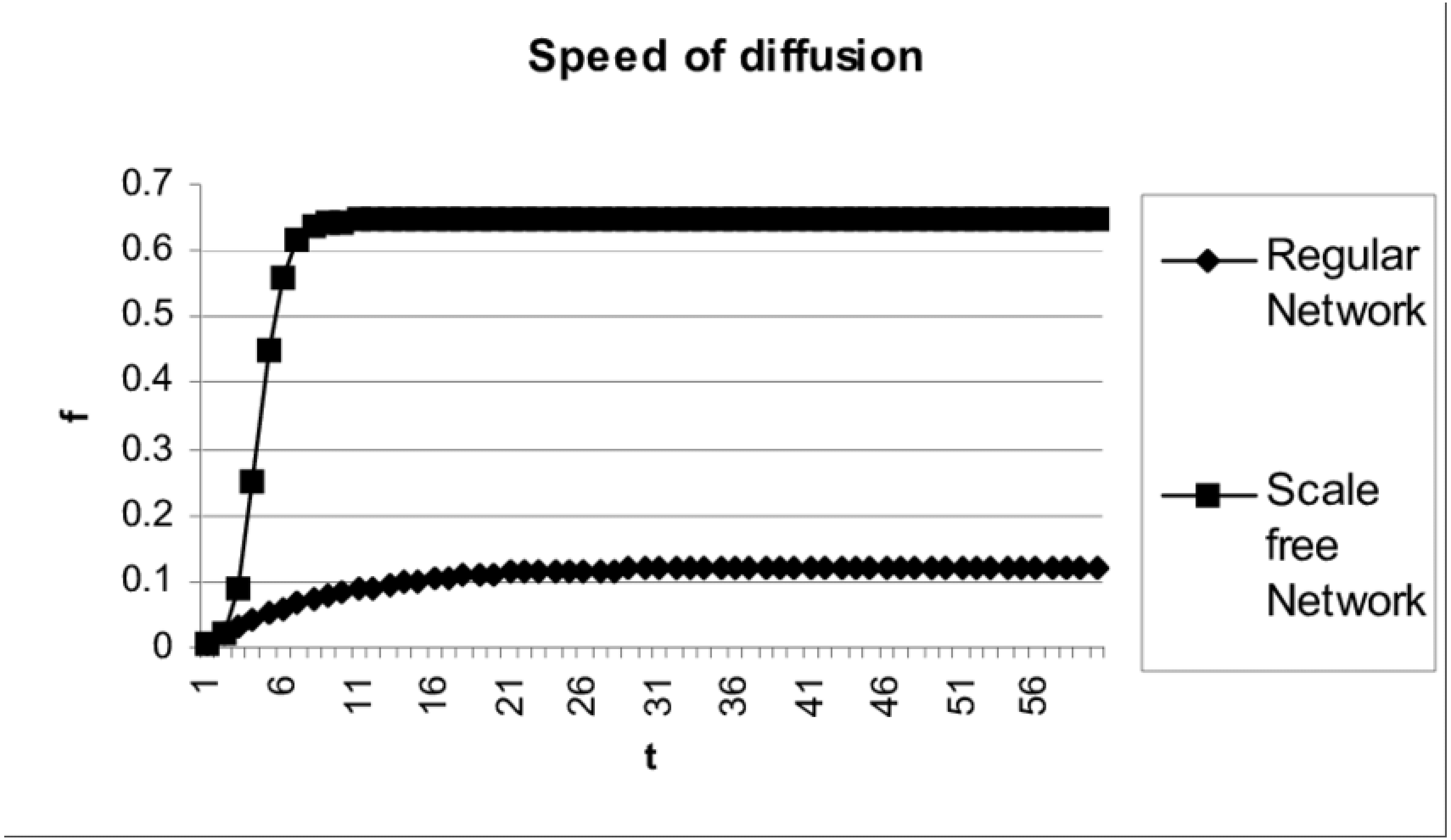


# The role of hubs



# Speed of diffusion



## 3. Simple versus Complex diffusion: spreading of viruses versus innovation

# Threshold models of adoption

## Shelling: segregation in cities

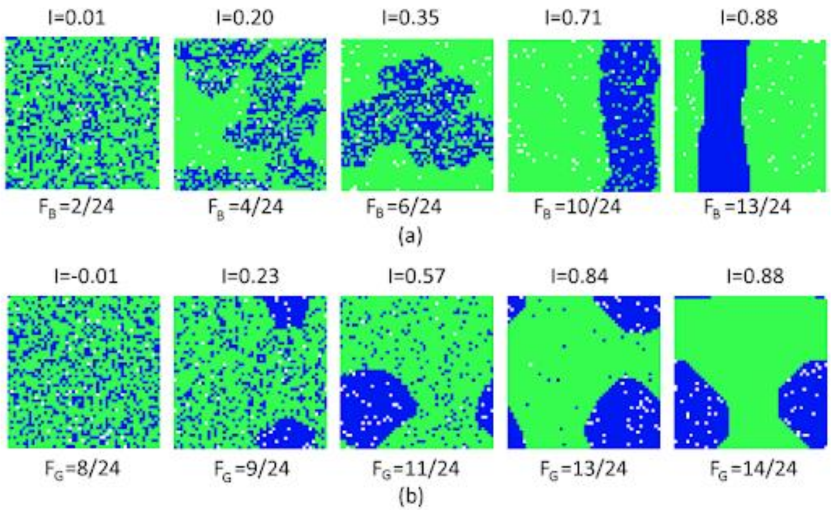
- Individual threshold in terms of ratio of tolerated different neighbors is heterogenous
- Low threshold individuals move away first that increases the probability of movement of high threshold individuals

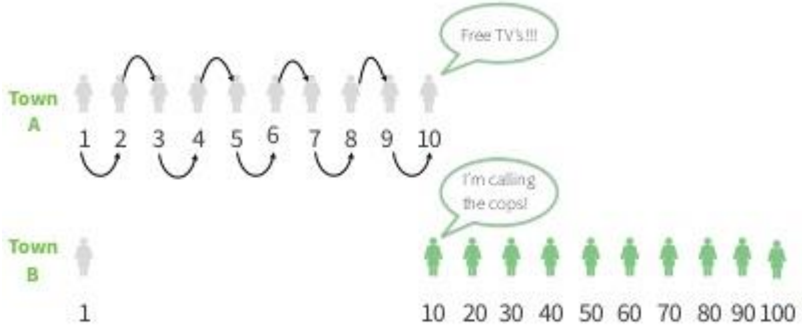
## Granovetter: behavior in a protest

- Individual threshold: how many individuals is needed to start the activity before the individual will decide to start

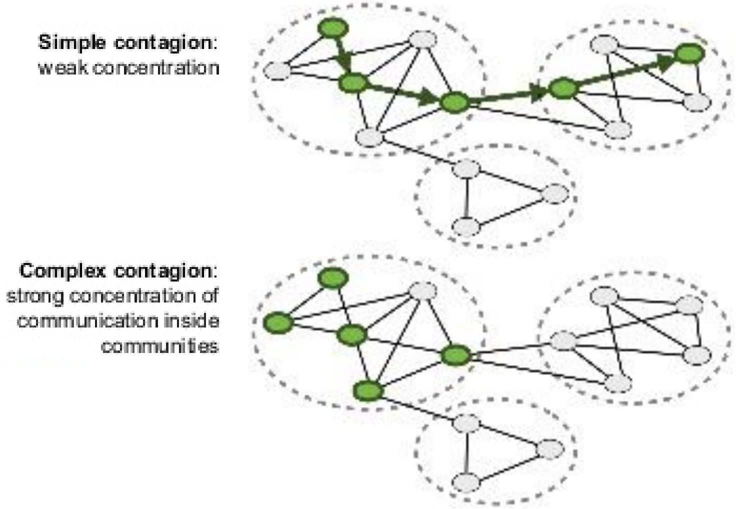
## Watts: cascading behavior in networks

- A behavior will diffuse only if the network structure follows the threshold sequence
- Innovators are connected to Early Adopters

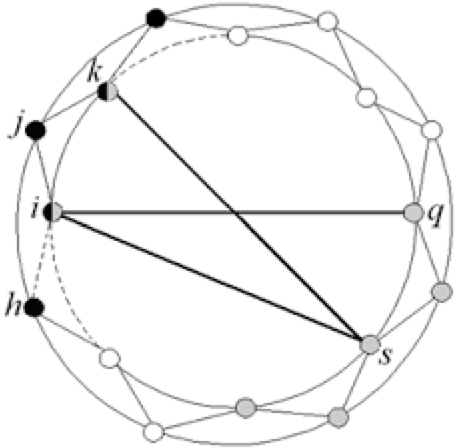


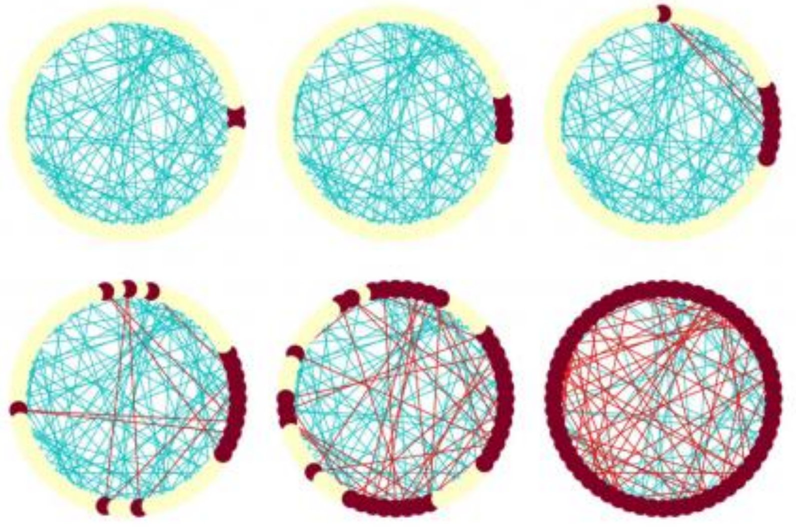



# Simple vs. complex contagion

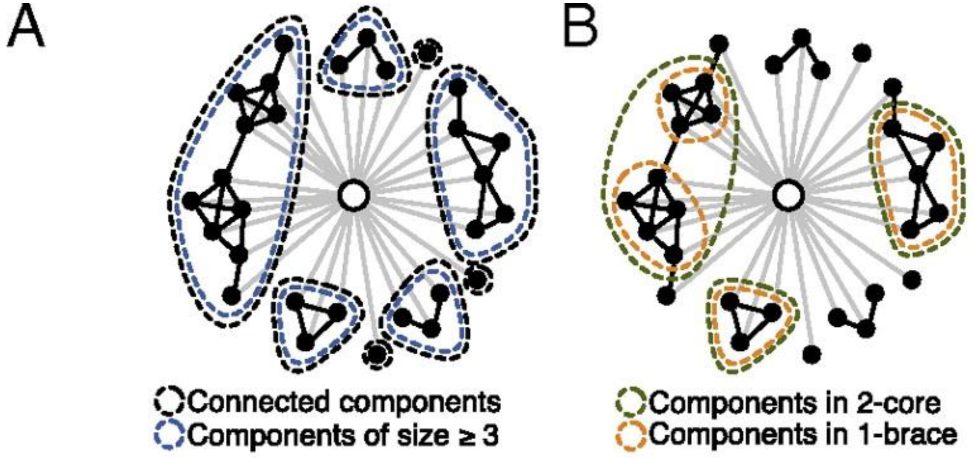


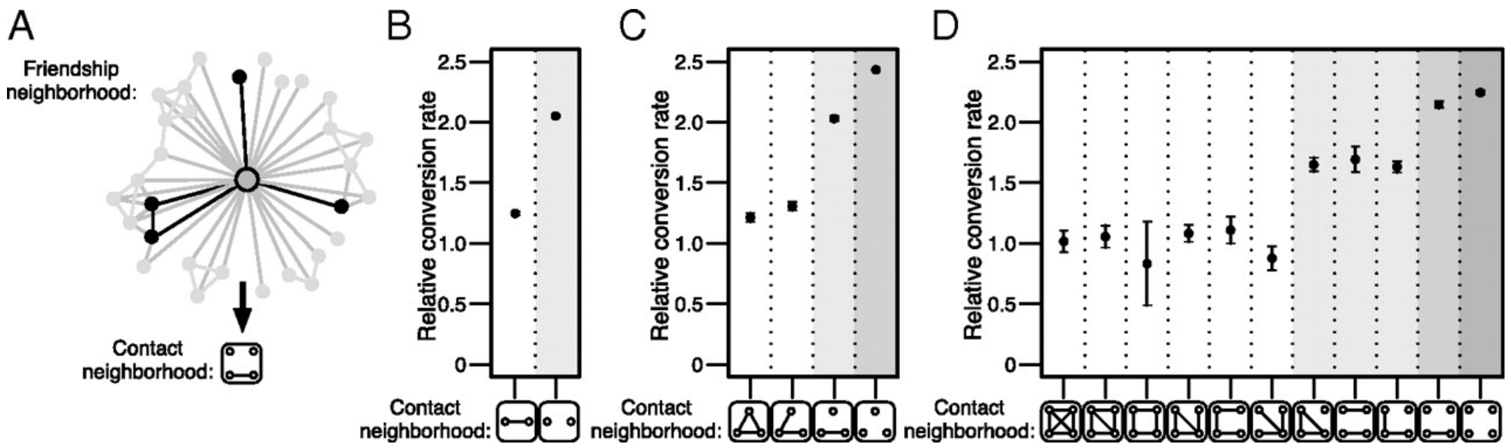

- Simple: infection depends on dyadic probability
- Complex: adoption is a process of convincing-decision and depends on fraction/number of adopting friends
- Innovation diffuses through complex contagion.
- Small-world networks: high clustering speeds diffusion up, bridges slow it down.



# Role of communities: adoption probability increases after adoption in diverse communities



# The role of geography in the complex diffusion of innovations


Lengyel Balázs



Bokányi Eszter
KRTK



Riccardo di Clemente UCL

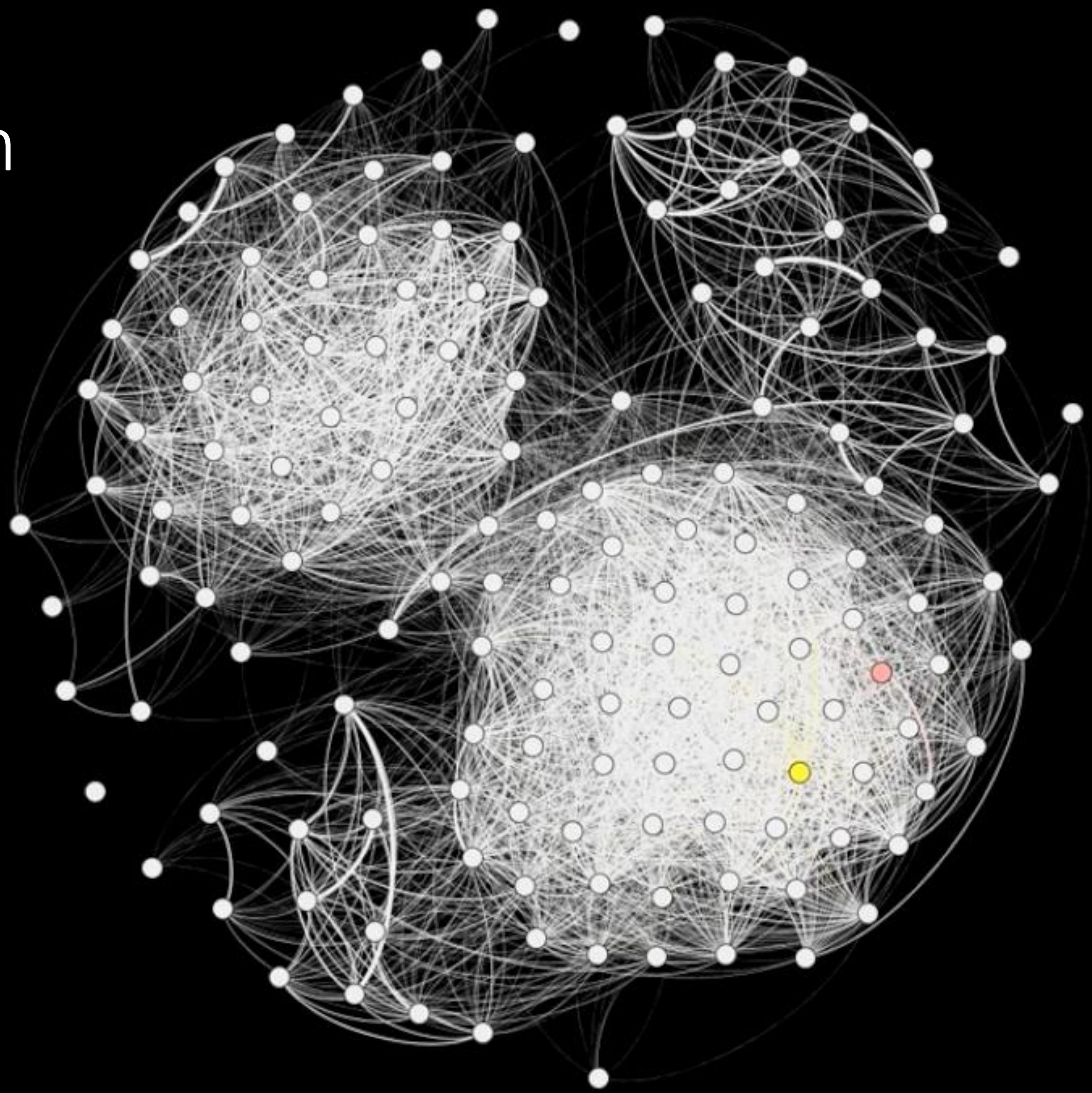


János Kertész CEU



Marta González
UC Berkeley

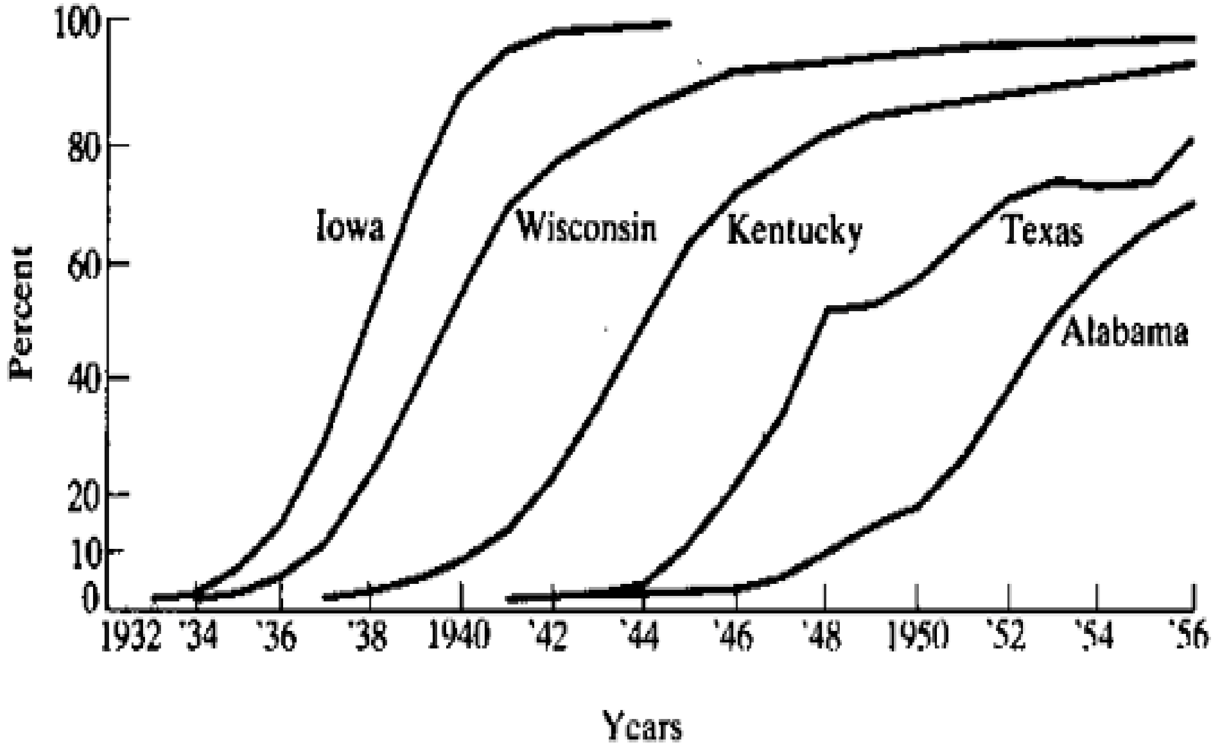
# Diffusion in networks



*Aplin et al. 2015. Nature 518.*
Doi:10.1038/nature13998

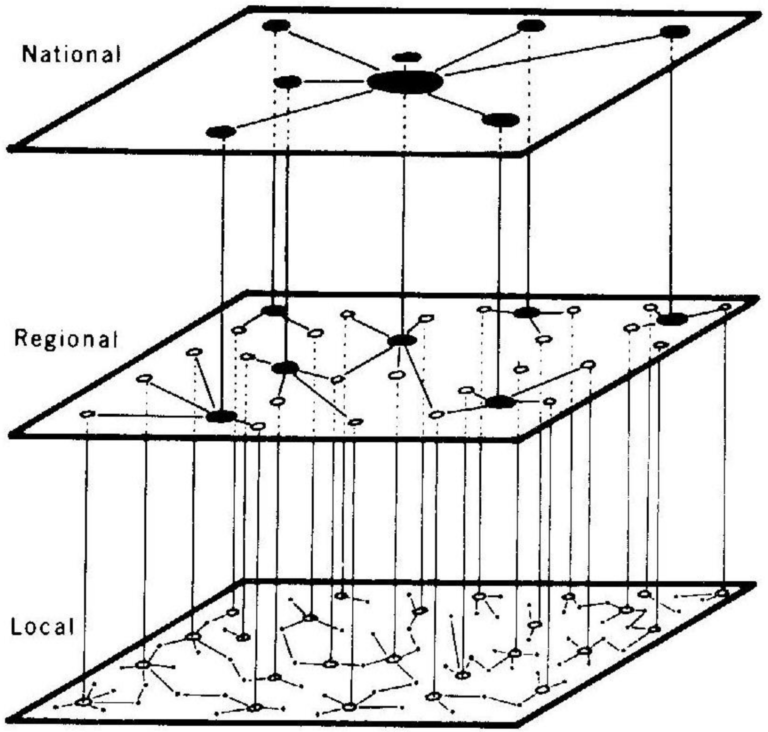
# Spatial diffusion

Innovation gets quicker to proximate places.

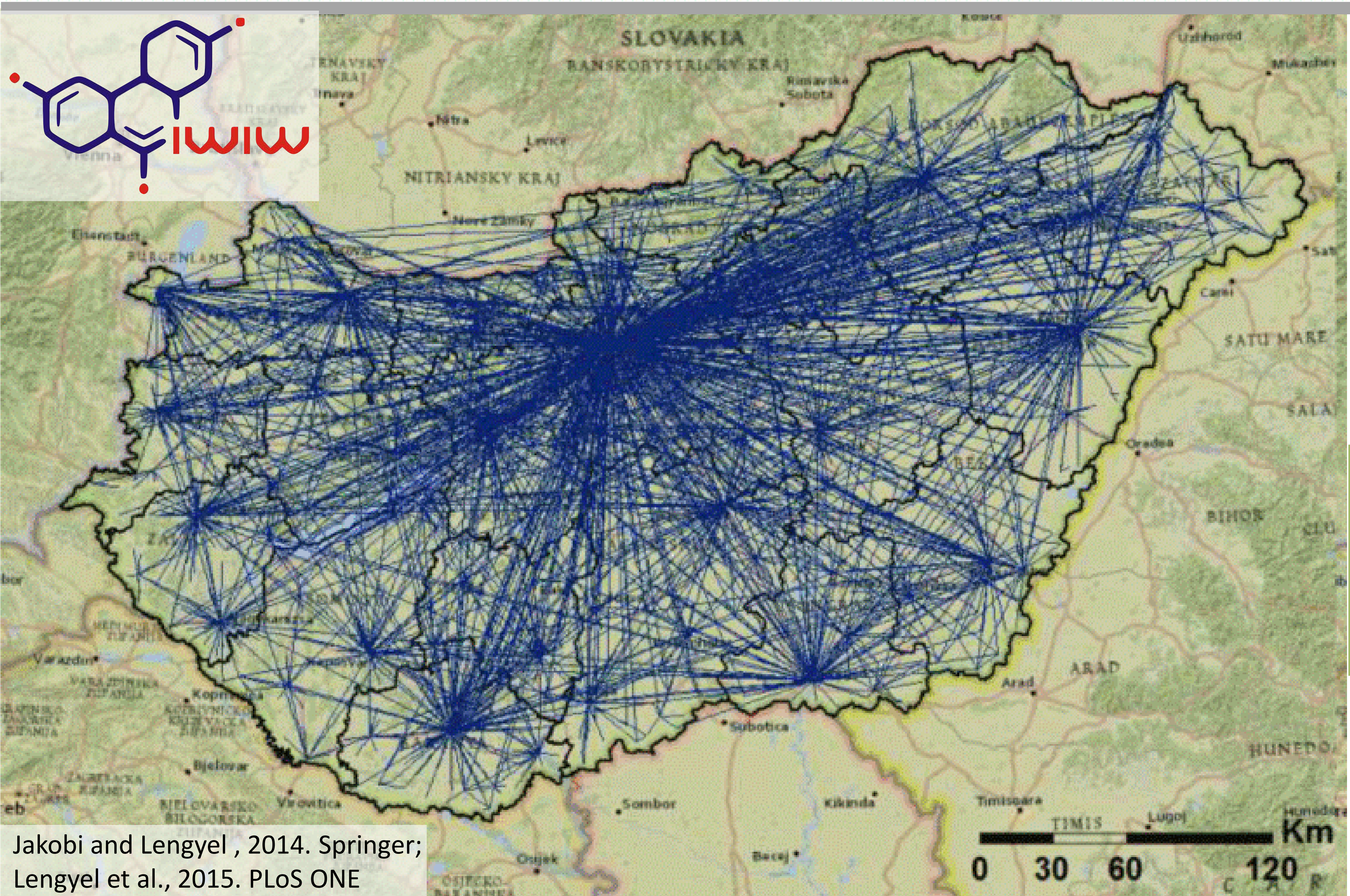



Griliches (1957, Econometrica)

Innovation spreads from big to medium cities and then to small towns.




Haegerstrand (1953, Chicago UP)



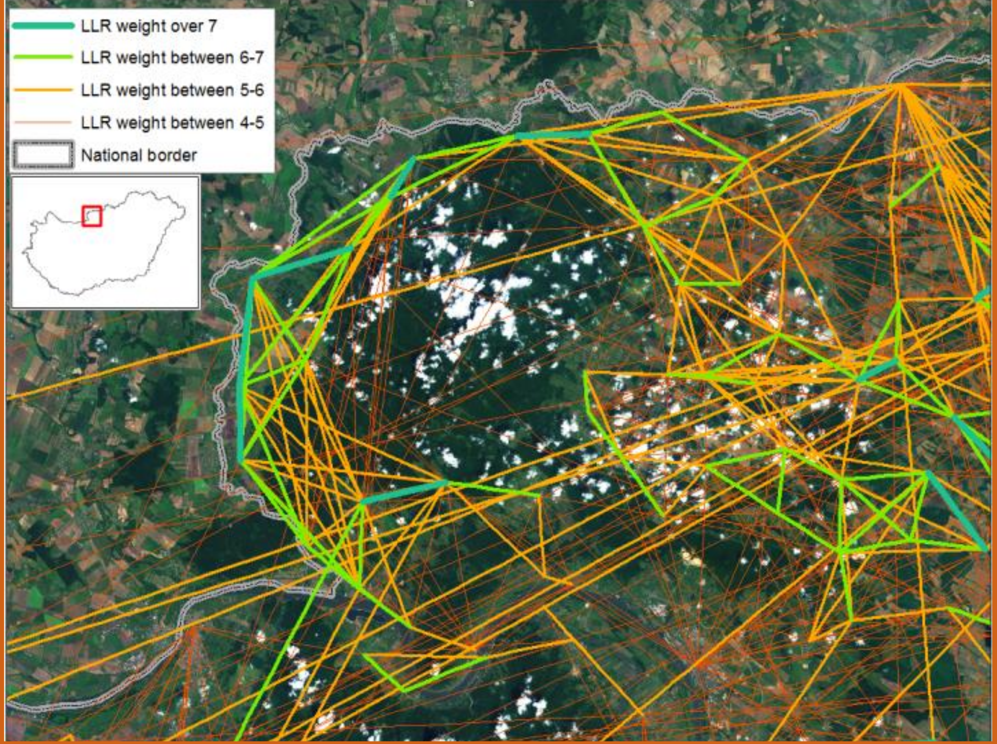

Jakobi and Lengyel , 2014. Springer;
Lengyel et al., 2015. PLoS ONE

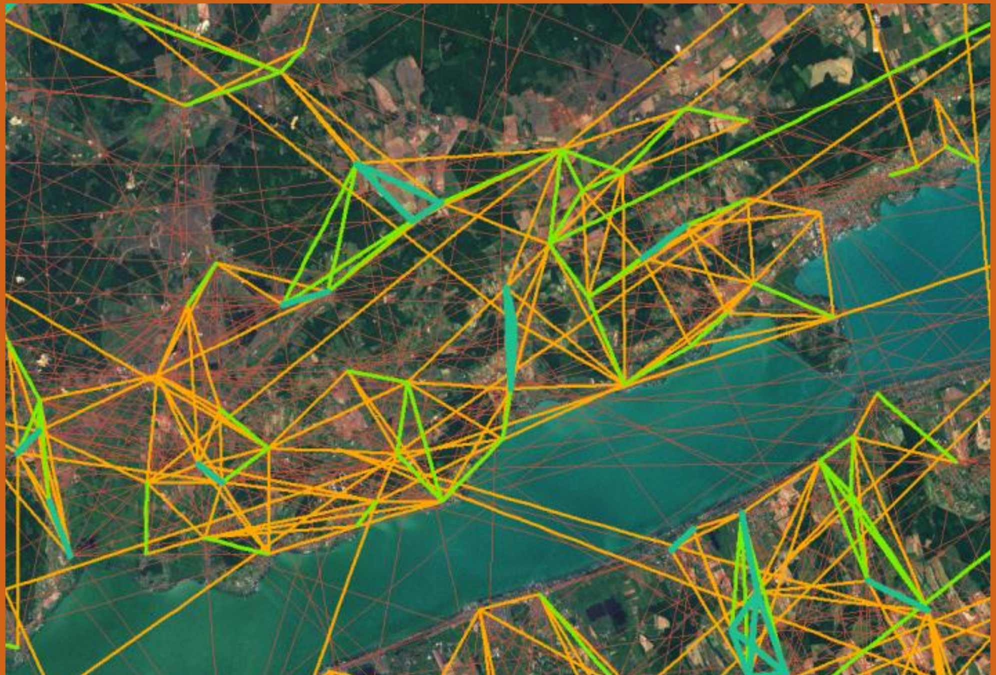
# Full life cycle of a social media platform

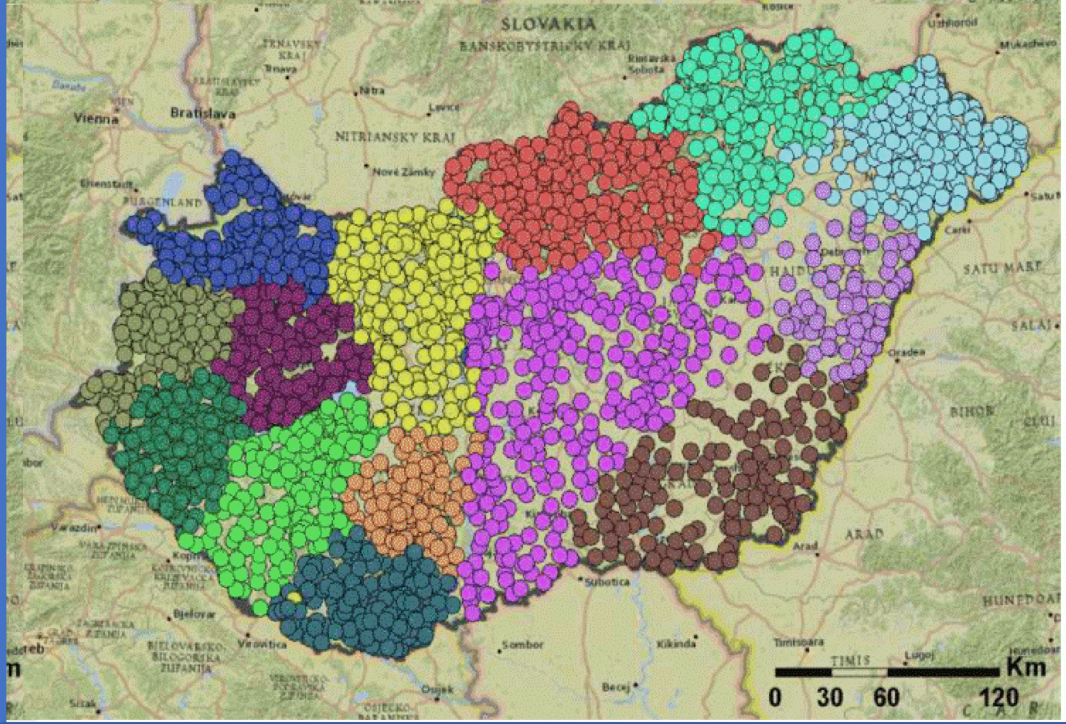
## Users

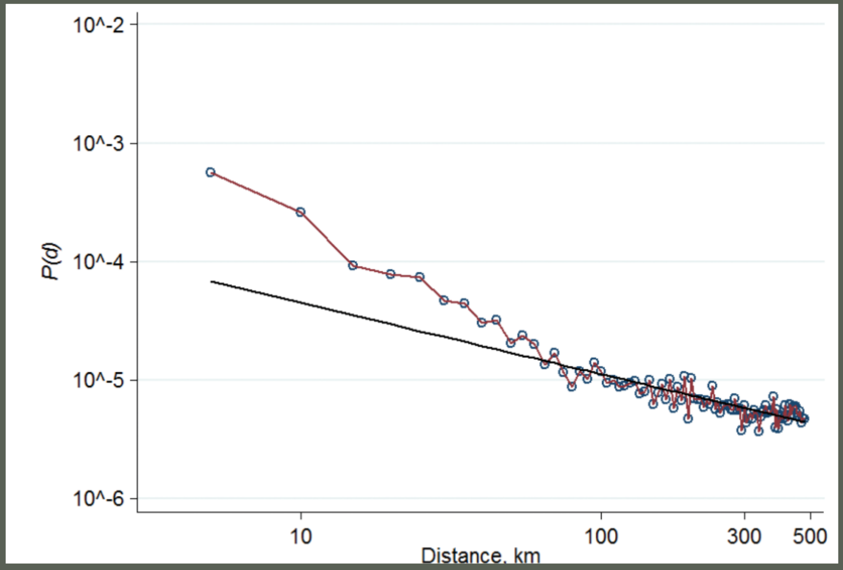
- **Location** (self-reported),
- **Date of registration,**
- **Date of last login,**
- **ID of friends,**
- **ID of invitor**

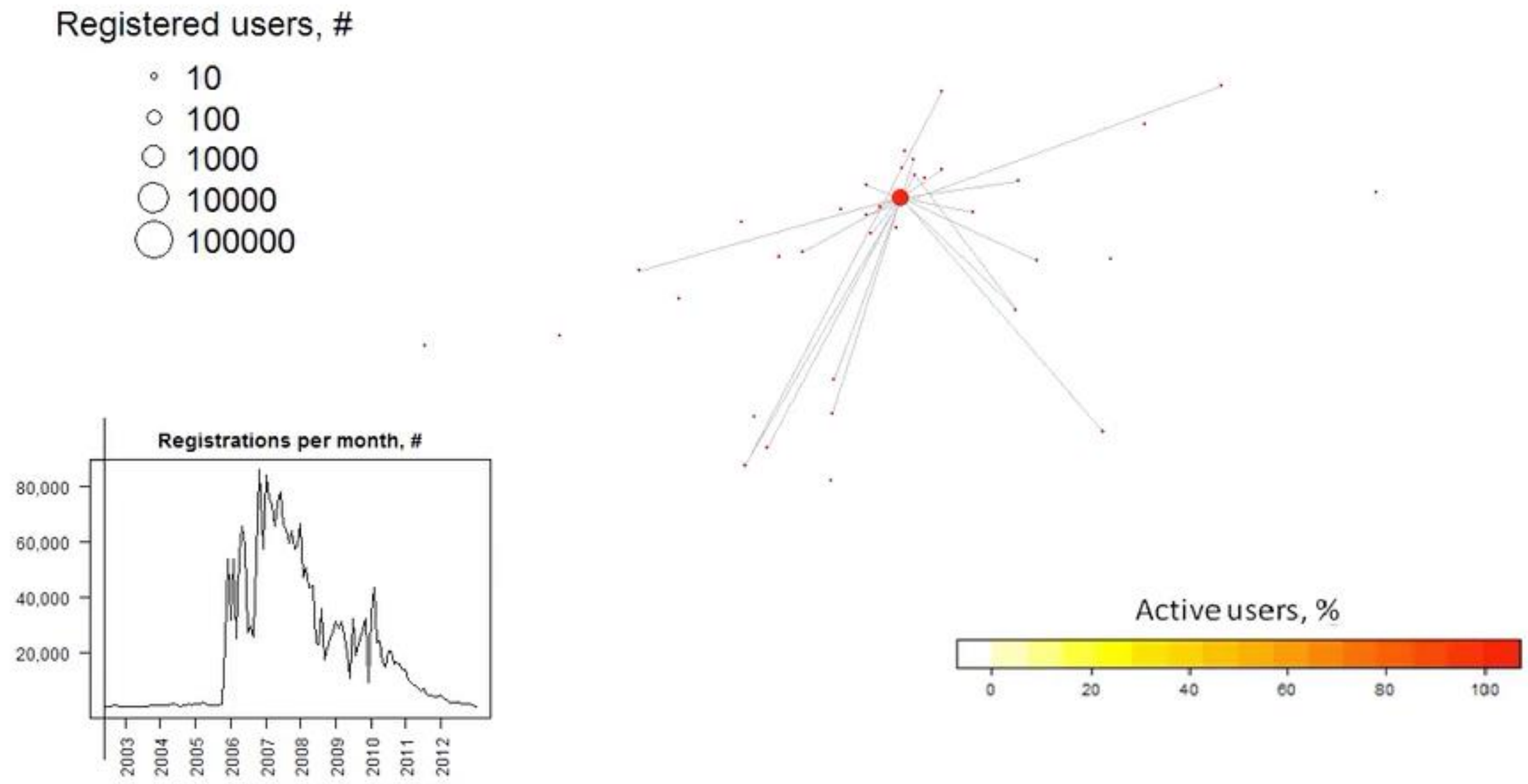
# Geographies of the network



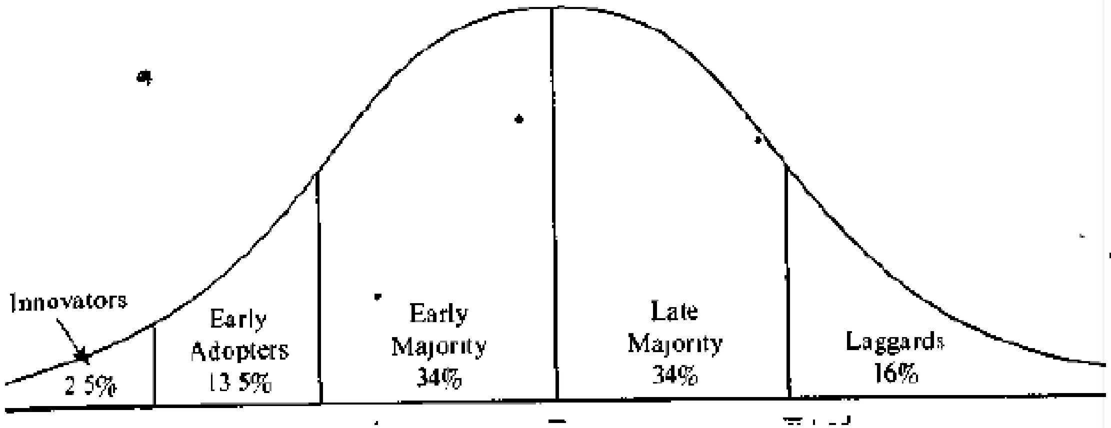


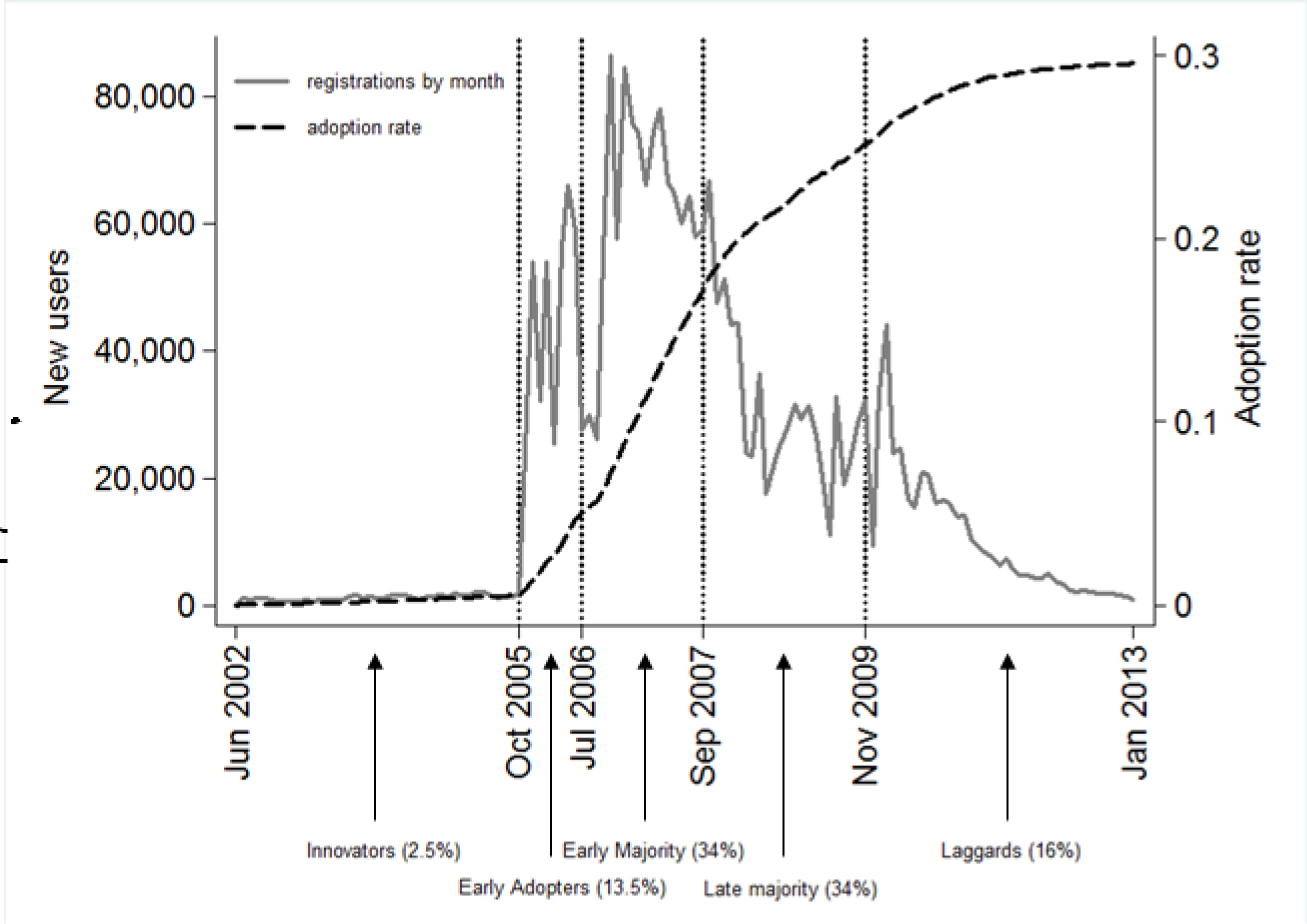





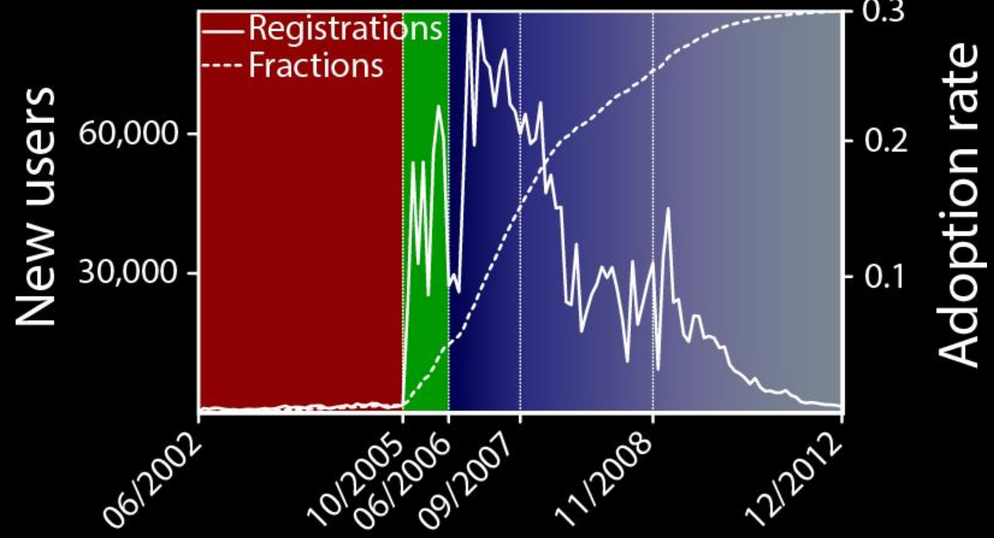
Registered users, #
10
100
1000
10000
100000
Registrations per month, #
80,000
60,000
40,000
20,000
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Active users, %
0
20
40
60
80
100

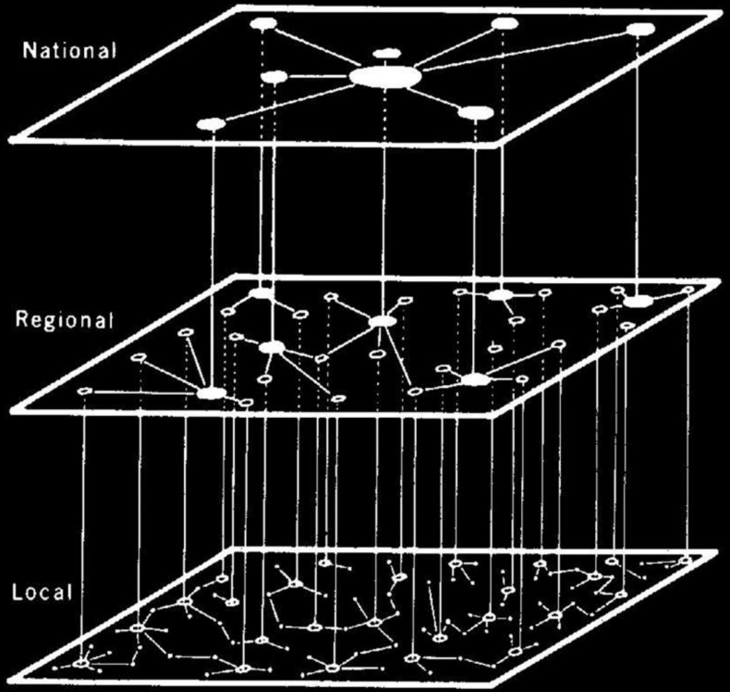
# The iWiW life cycle by Rogers categories



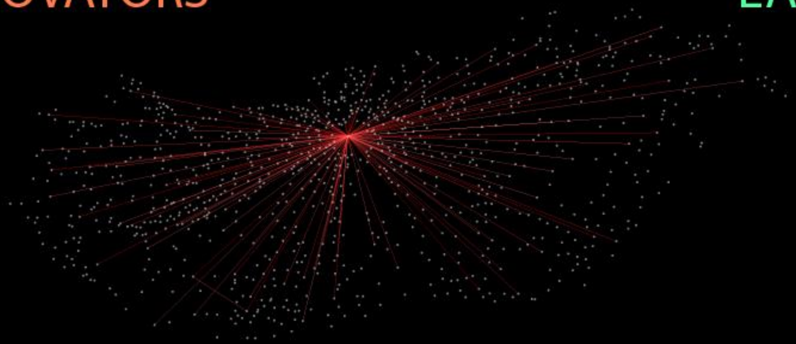




Forrás: Rogers (The Free Press, 1962),



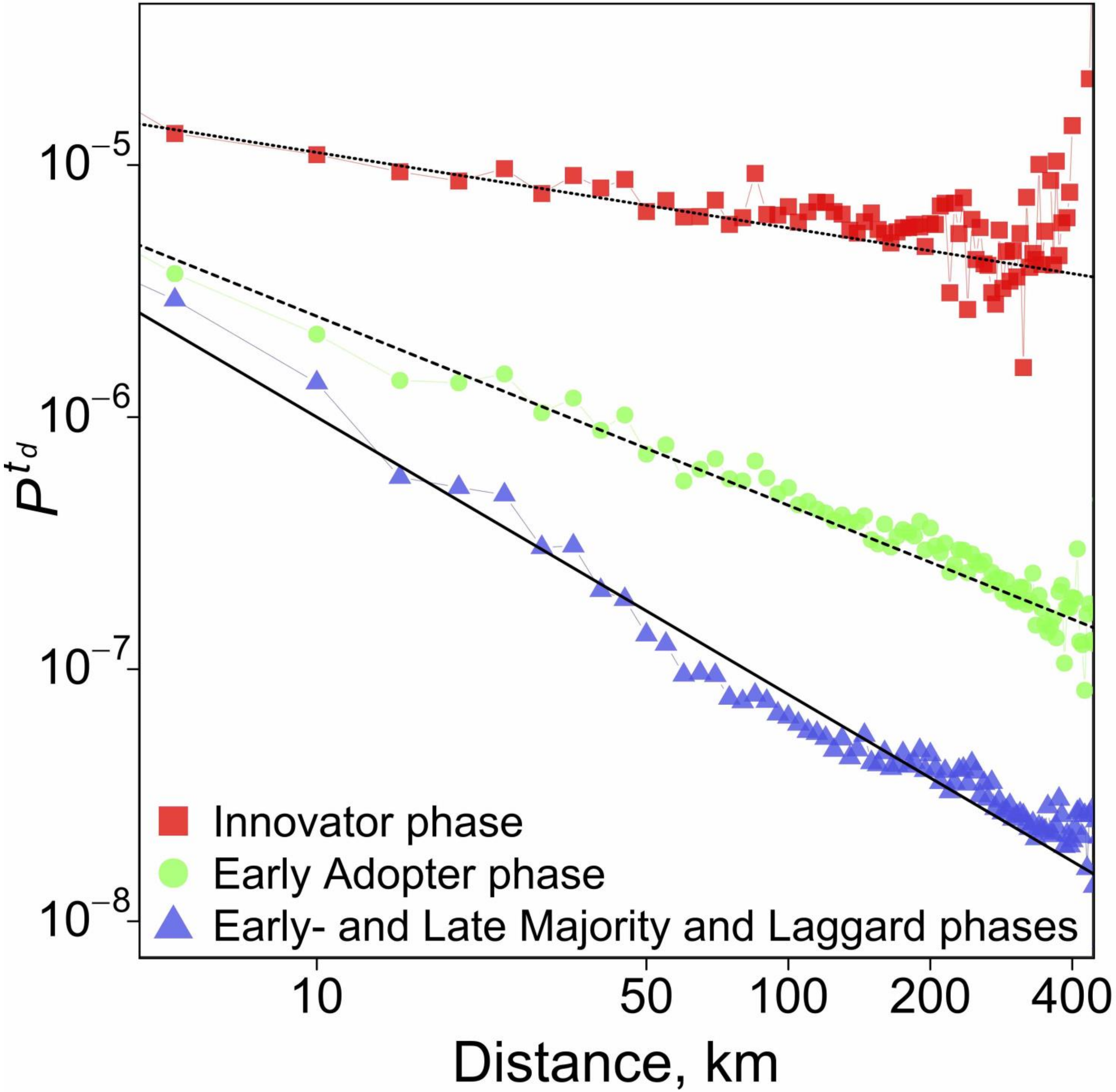

INNOVATORS



EARLY ADOPTERS

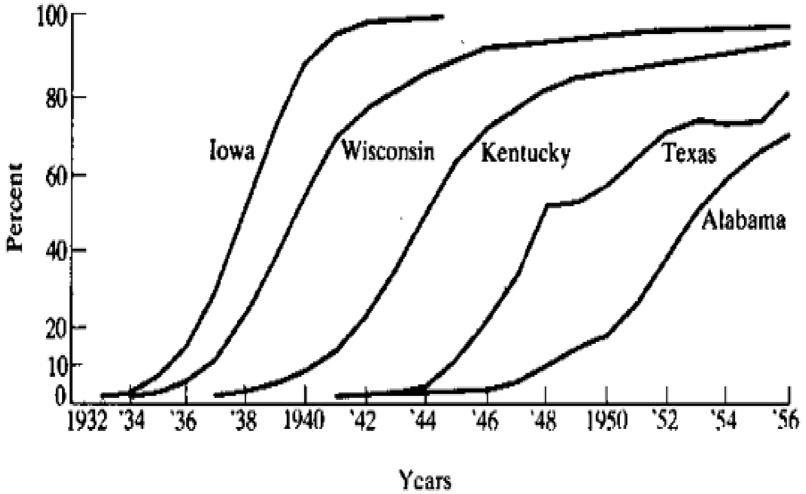
MAJORITY & LAGGARDS

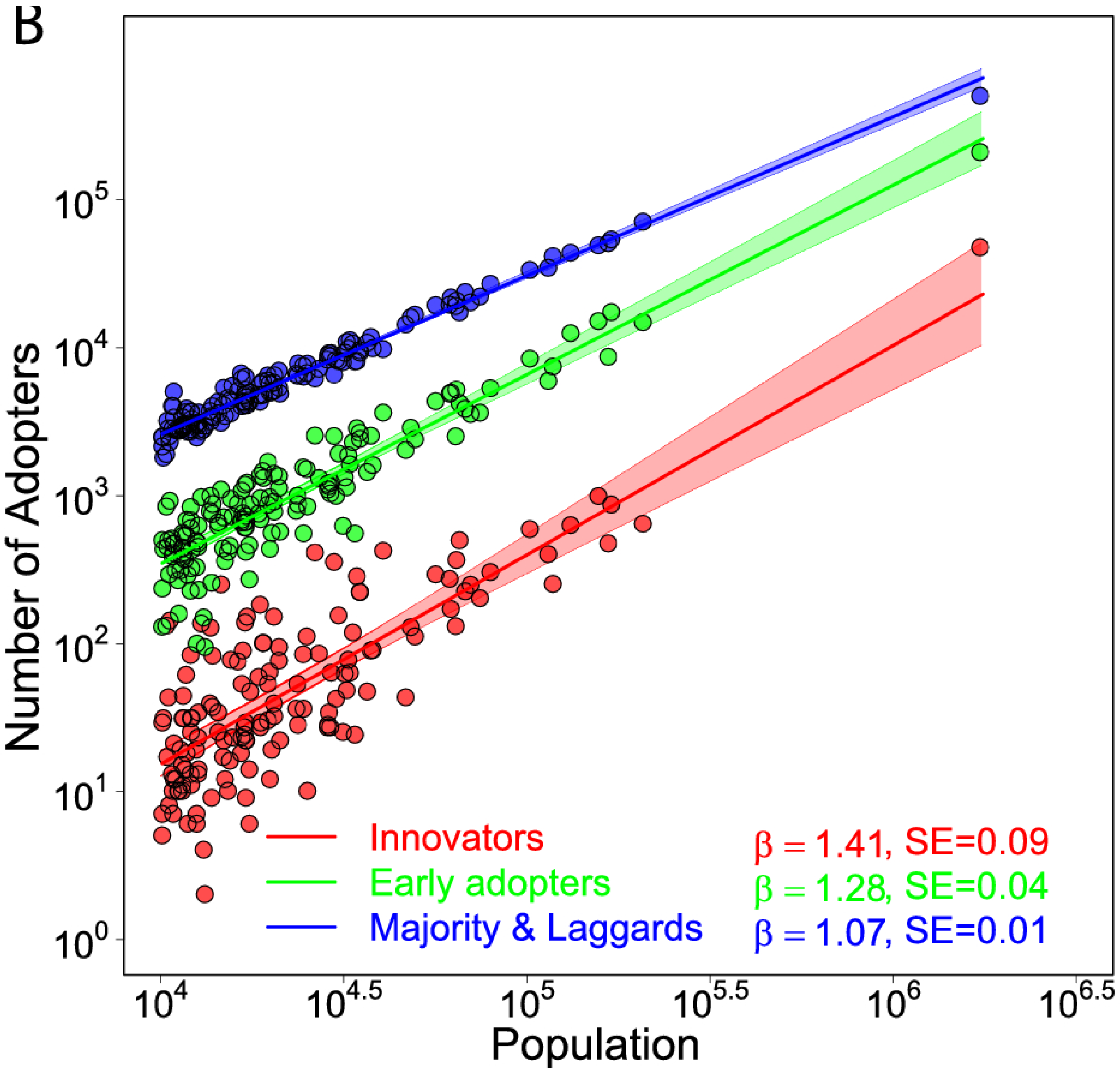
# The role of distance




Distance only slightly decreases the probability of invitaiton in early phases

Diffusion becomes local in later phases.

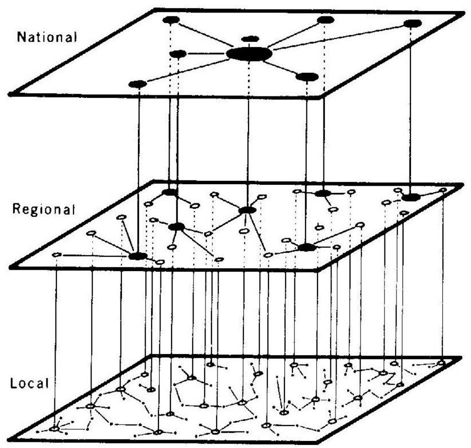





# The role of size

Innovators and Early Adopters concentrate in cities.

Majority and Laggards are proportional to town size.



# Adoption Rate

(Bass Curve)

$$\frac{dy_a(t)}{dt} = (p_a + q_a y_a(t))(1 - y_a(t))$$

$$y_a(t) = m\frac{1 - e^{-(p_a+q_a)t}}{1 + \frac{q_a}{p_a}e^{-(p_a+q_a)t}},$$

# of adopters

$y_a(t)$

$t$

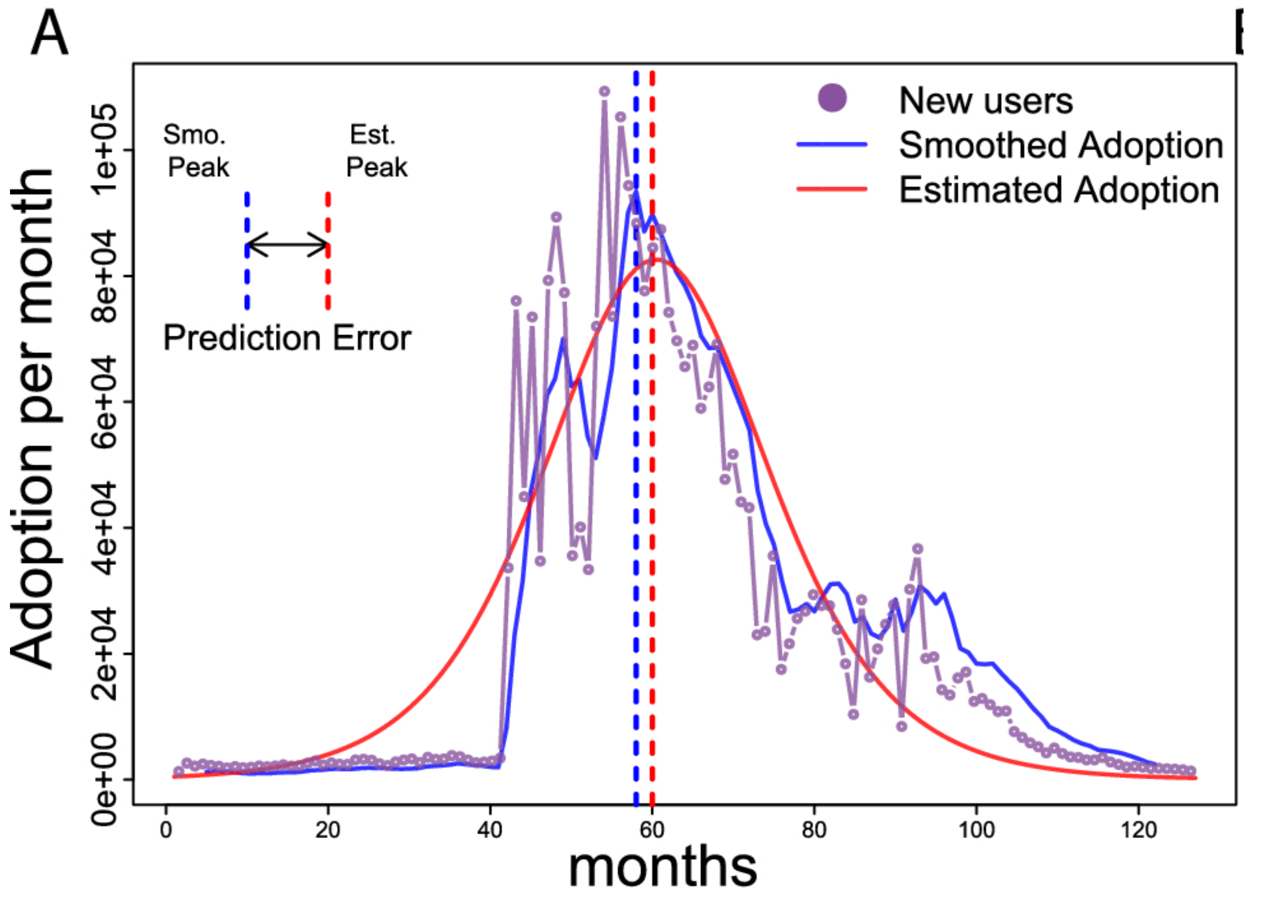
Time (months)

Innovation Par.

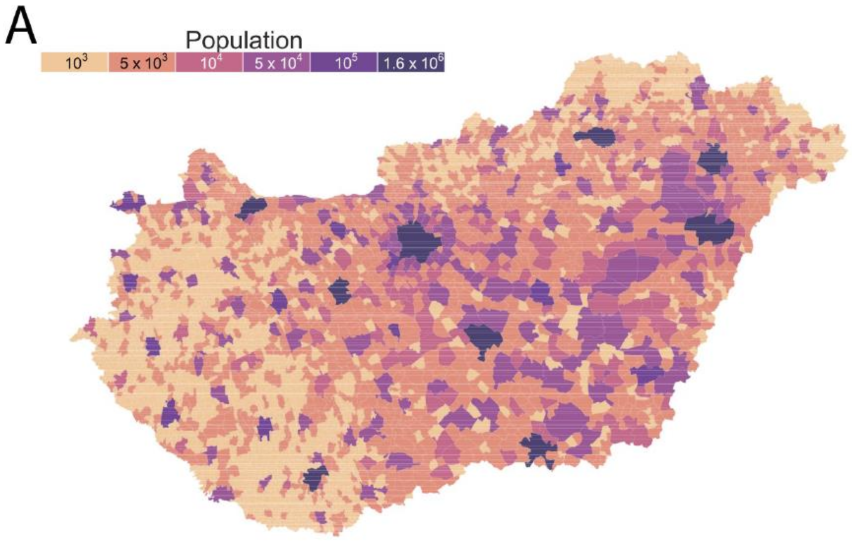
$p_a$

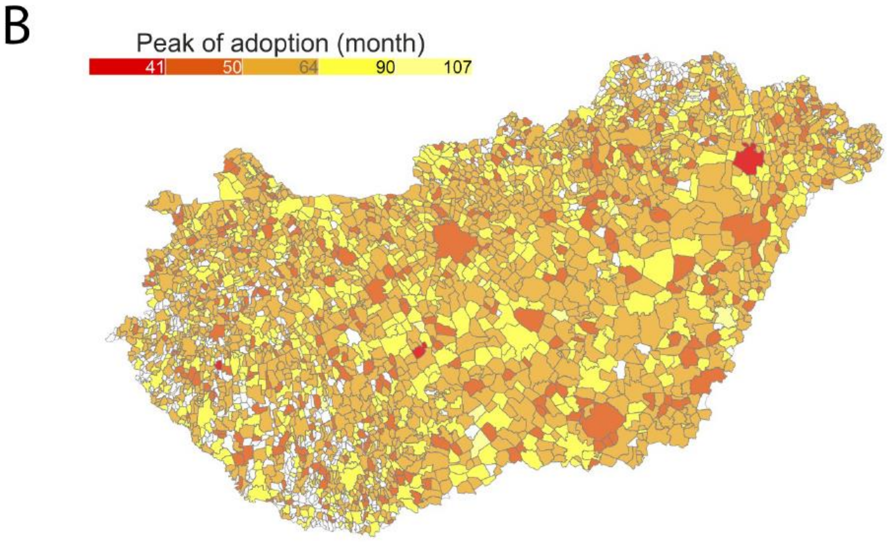
$q_a$

Imitation Par.

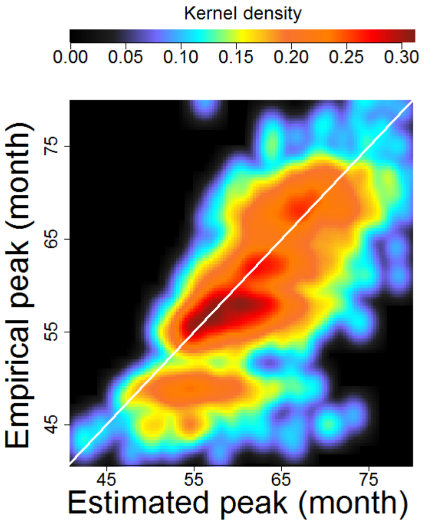


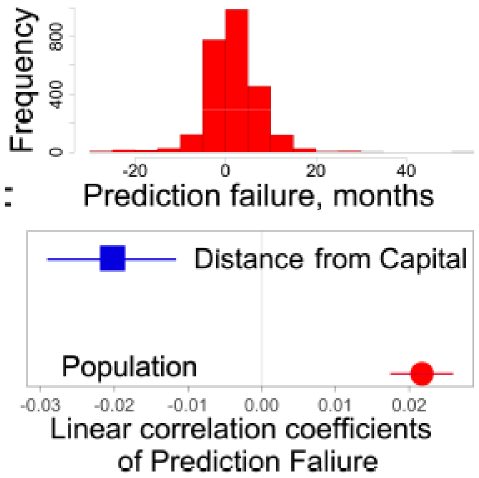
# Local adoption peaks and prediction errors




$$T^i_{peak-t} = \frac{\ln p^i_a + \ln q^i_a}{p^i_a + q^i_a}$$

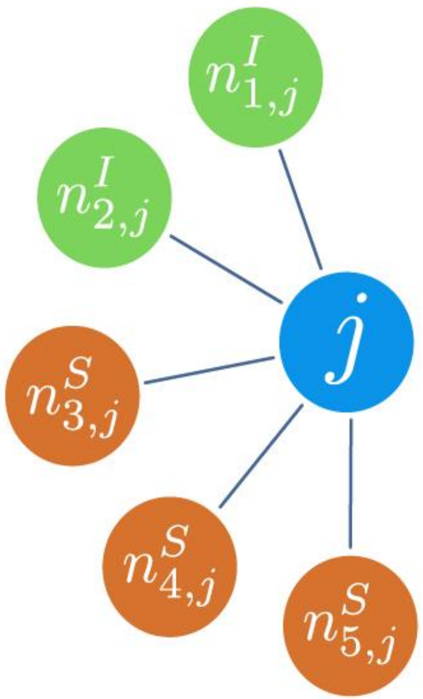


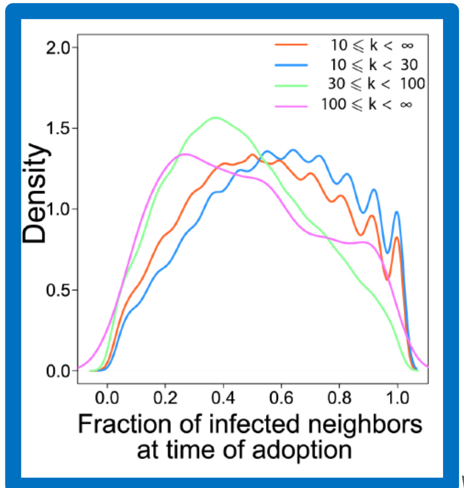
# Agent-based model of diffusion in the network

- We fix the network.
- Inactives are not removed.

- 10% sample (300 K users)
- Sampling stratified by towns and network moduls




Innovation or Marketing parameter



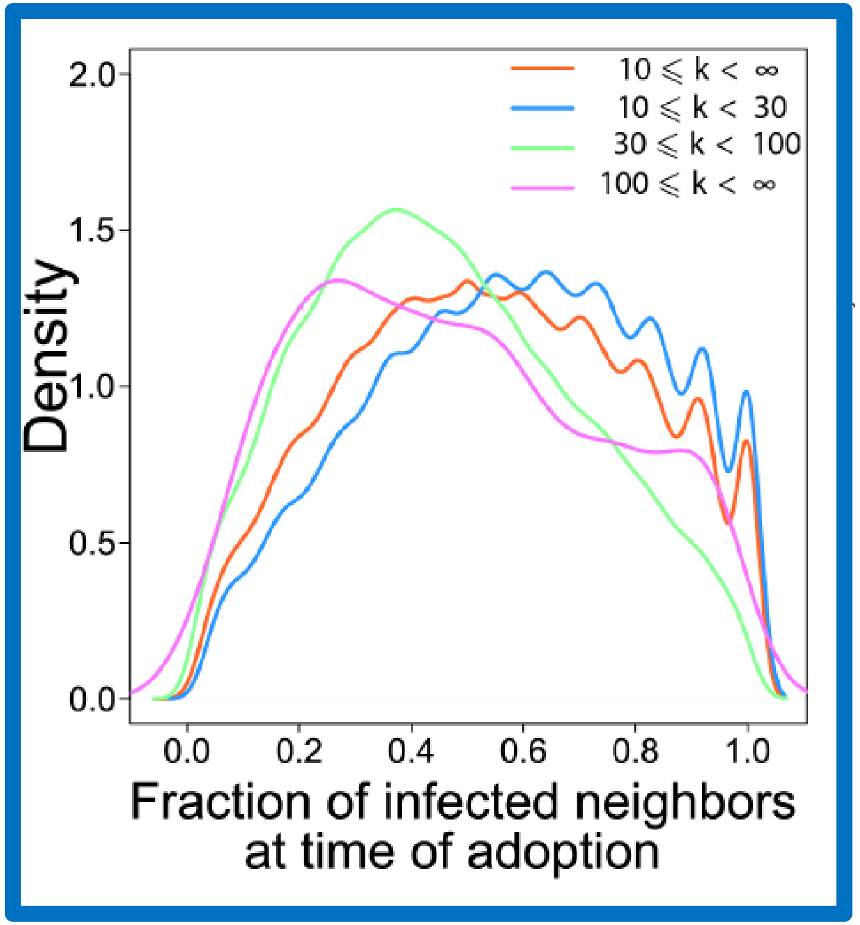

Network effect

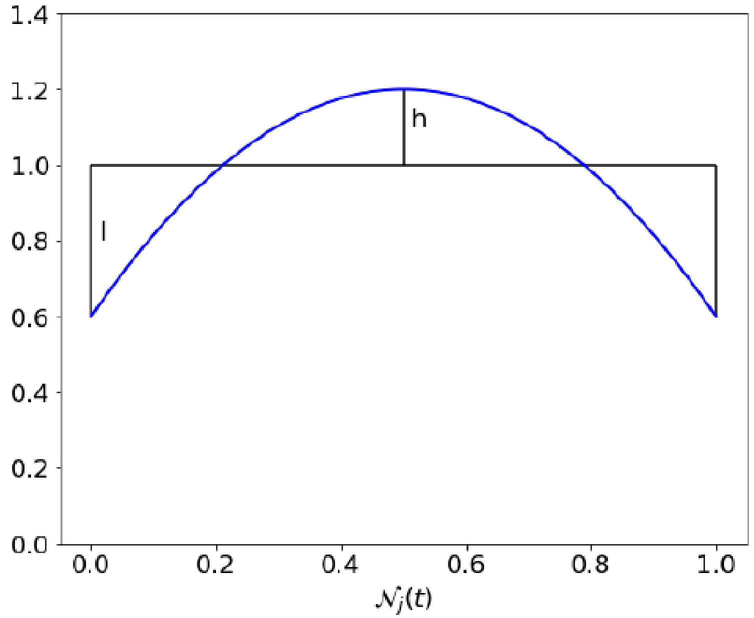
$$\mathcal{N}_j(t) = \frac{\#n_j^I(t)}{\#n_j^I(t) + \#n_j^S(t)}$$

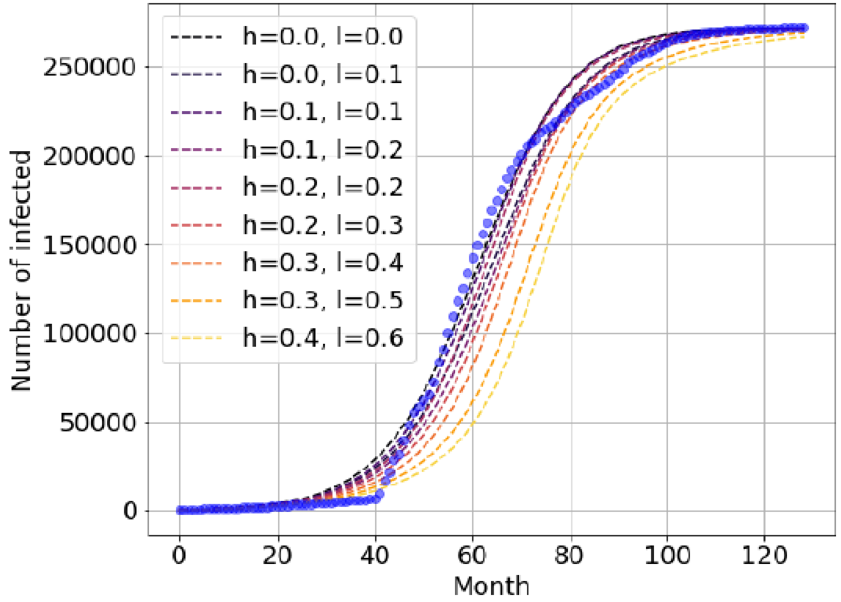
$$F_j(t+1) = \begin{cases} I & \text{if } U(0,1)_{jt} < \hat{p}^{\mathrm{ABM}} + T(\mathcal{N}_j(t), h, l) \times \mathcal{N}_j(t) \times \hat{q}^{\mathrm{ABM}} \\ S & \text{otherwise} \end{cases}$$

Imitation parameter

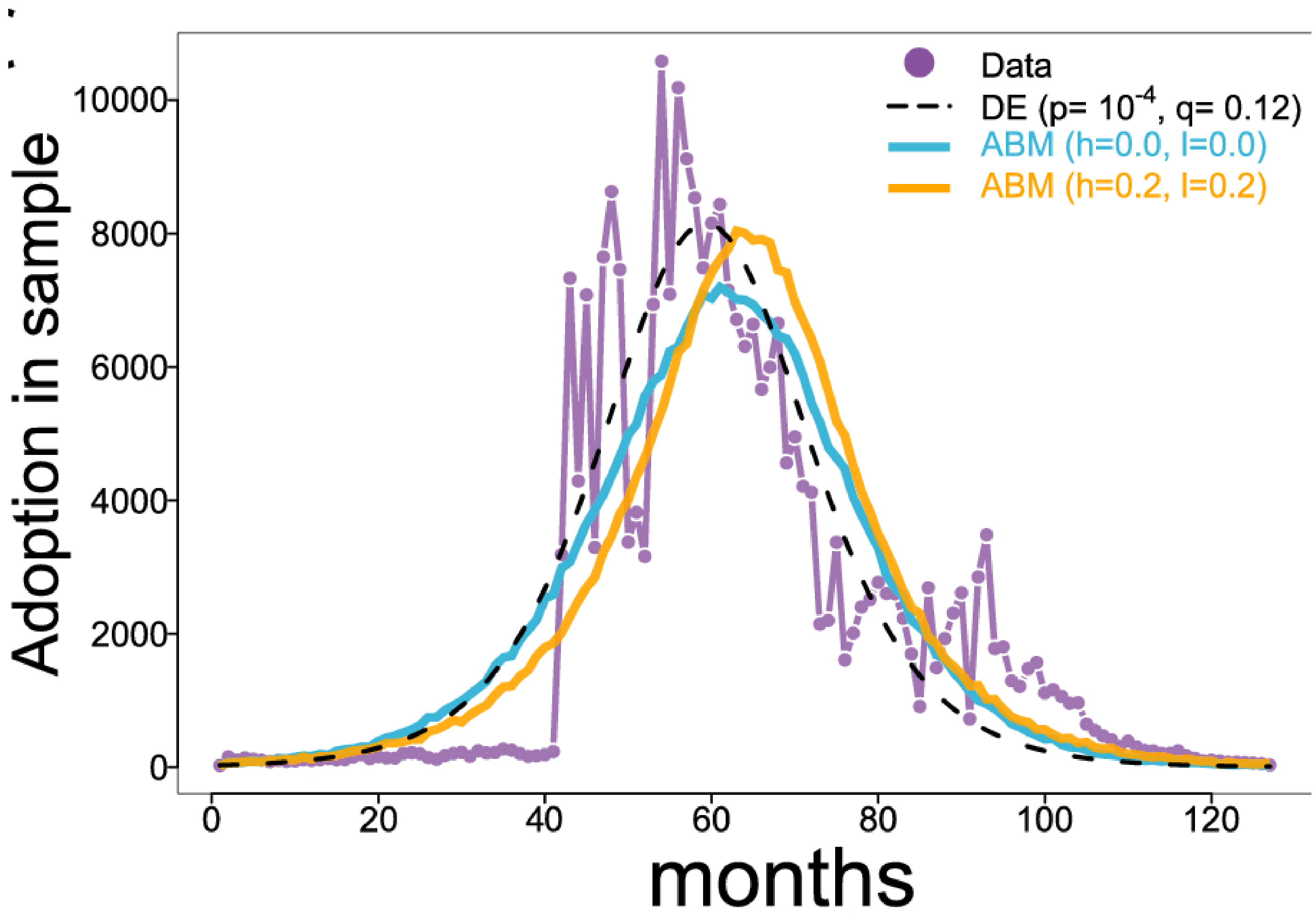
# Threshold parameters



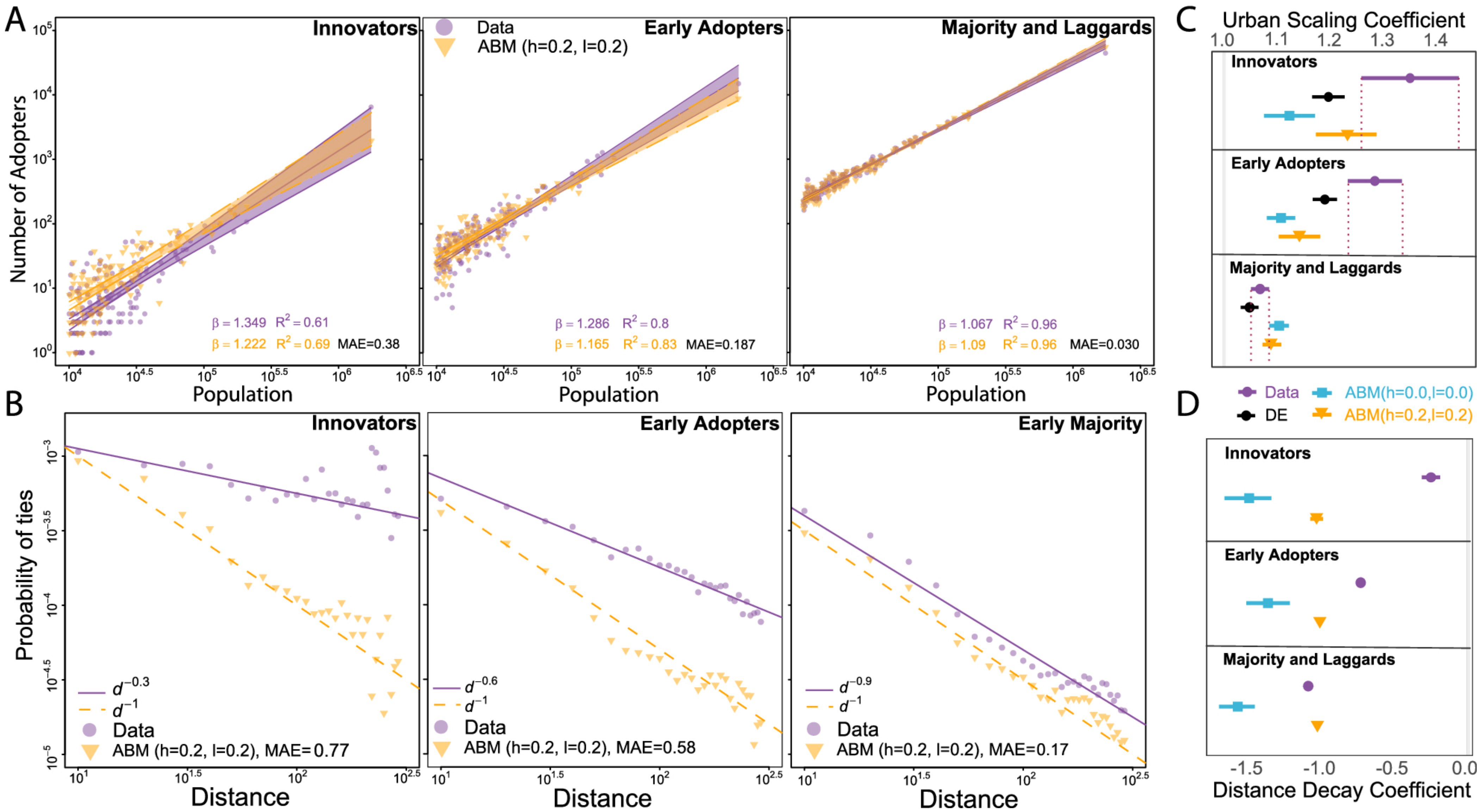



# Results



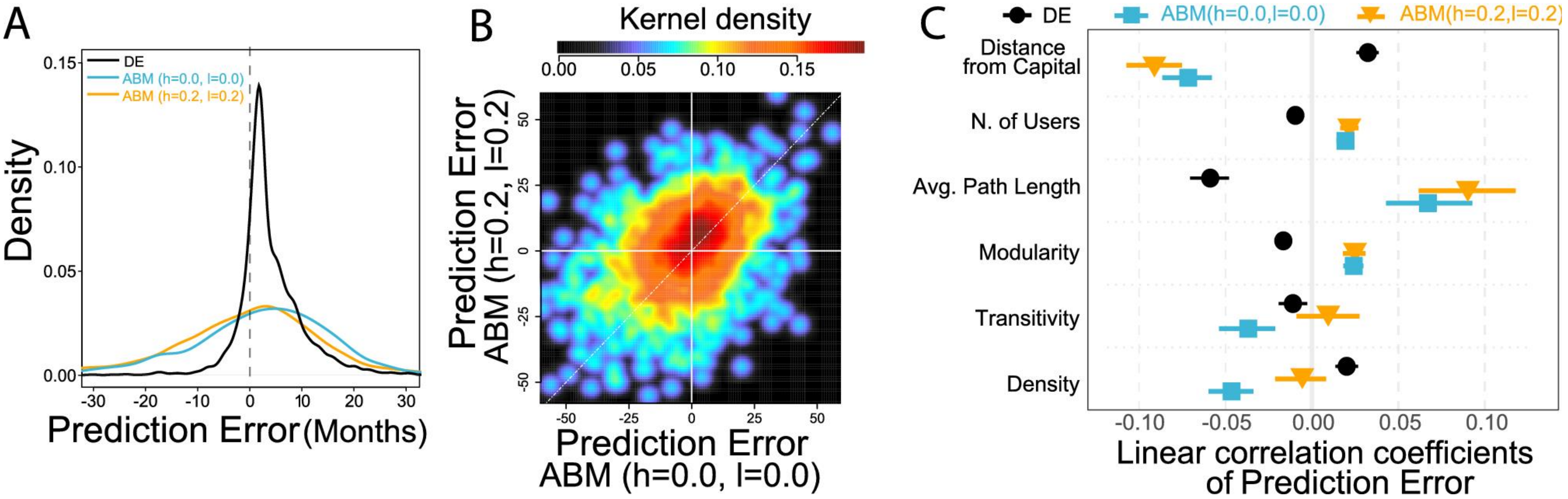
Data
DE (p= 10-4, q= 0.12)
ABM (h=0.0, l=0.0)
ABM (h=0.2, l=0.2)
Adoption in sample
0
2000
4000
6000
8000
10000
0
20
40
60
80
100
120
months

# Results



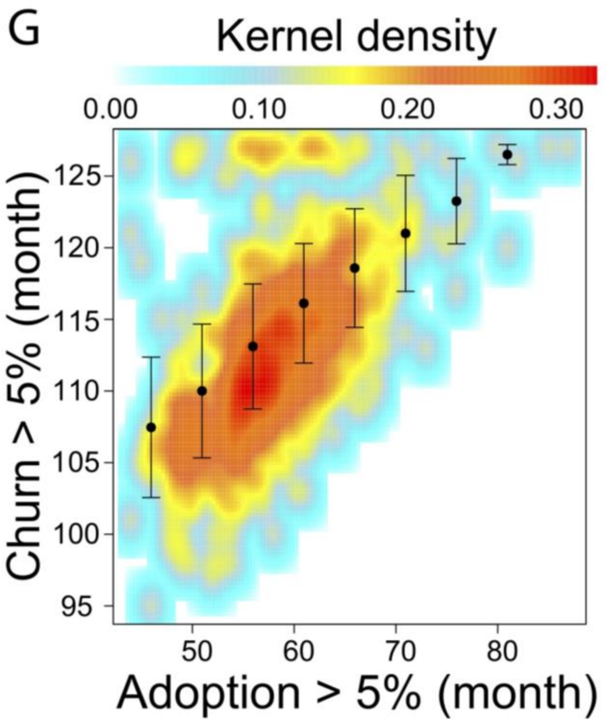
A
Innovators
Early Adopters
Majority and Laggards
Data
ABM (h=0.2, l=0.2)
Number of Adopters
Population
β = 1.349 R² = 0.61
β = 1.222 R² = 0.69 MAE=0.38
β = 1.286 R² = 0.8
β = 1.165 R² = 0.83 MAE=0.187
β = 1.067 R² = 0.96
β = 1.09 R² = 0.96 MAE=0.030
B
Innovators
Early Adopters
Early Majority
Probability of ties
Distance
d^-0.3
d^-1
Data
ABM (h=0.2, l=0.2), MAE=0.77
d^-0.6
d^-1
Data
ABM (h=0.2, l=0.2), MAE=0.58
d^-0.9
d^-1
Data
ABM (h=0.2, l=0.2), MAE=0.17
C
Urban Scaling Coefficient
1.0
1.1
1.2
1.3
1.4
Innovators
Early Adopters
Majority and Laggards
D
Data
ABM(h=0.0,l=0.0)
DE
ABM(h=0.2,l=0.2)
Innovators
Early Adopters
Majority and Laggards
-1.5
-1.0
-0.5
0.0
Distance Decay Coefficient

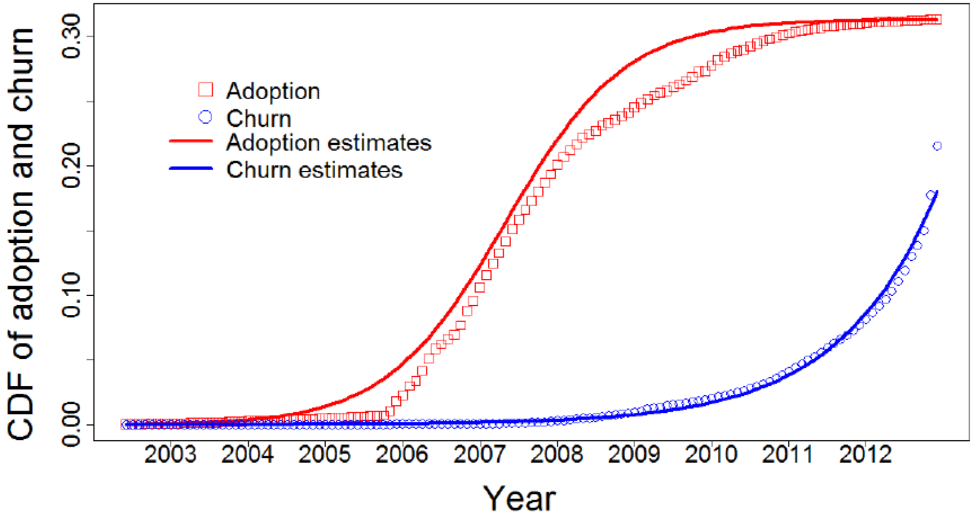
# Results

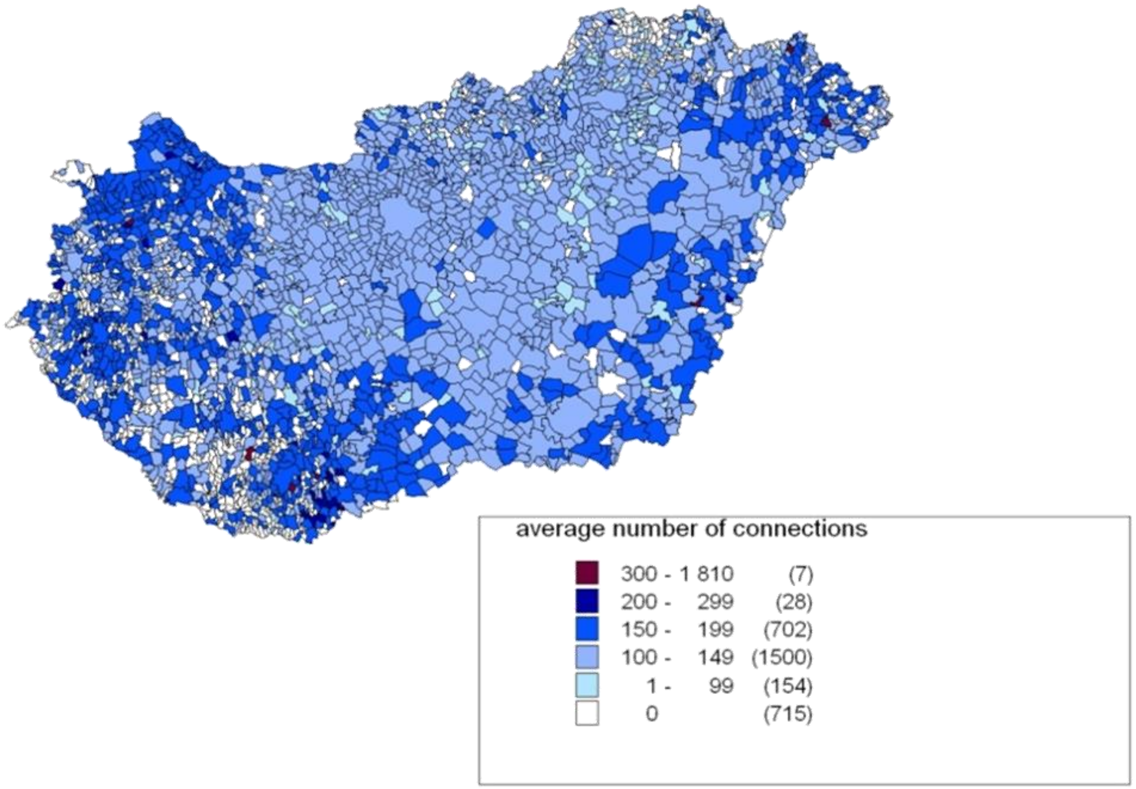


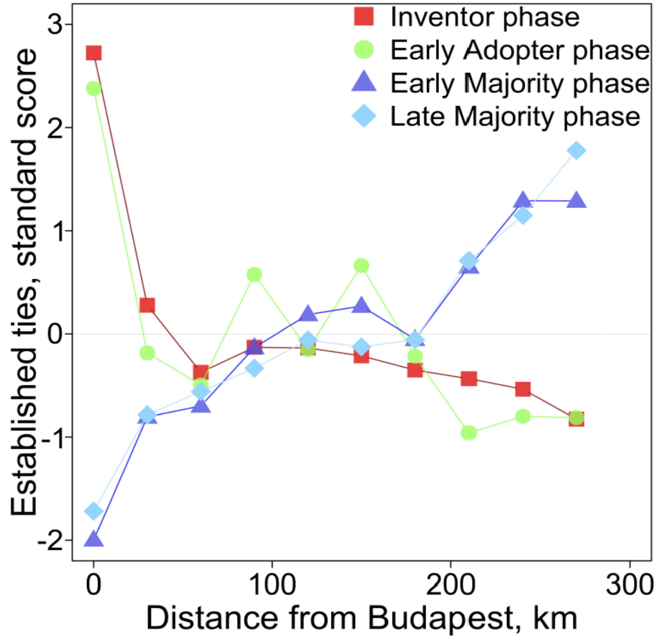
A
Density
Prediction Error (Months)
DE
ABM (h=0.0, l=0.0)
ABM (h=0.2, l=0.2)
B
Kernel density
Prediction Error
ABM (h=0.2, l=0.2)
Prediction Error
ABM (h=0.0, l=0.0)
C
DE
ABM(h=0.0,l=0.0)
ABM(h=0.2,l=0.2)
Distance from Capital
N. of Users
Avg. Path Length
Modularity
Transitivity
Density
Linear correlation coefficients of Prediction Error

# CHURN


G
Kernel density
0.00
0.10
0.20
0.30
125
120
115
110
105
100
95
50
60
70
80
Churn > 5% (month)
Adoption > 5% (month)


Adoption
Churn
Adoption estimates
Churn estimates
CDF of adoption and churn
0.00
0.10
0.20
0.30
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year


average number of connections
300 - 1 810 (7)
200 - 299 (28)
150 - 199 (702)
100 - 149 (1500)
1 - 99 (154)
0 (715)


Inventor phase
Early Adopter phase
Early Majority phase
Late Majority phase
Established ties, standard score
3
2
1
0
-1
-2
0
100
200
300
Distance from Budapest, km