



Seminários do
Programa Interinstitucional de Pós-Graduação em Estatística

Epidemic Processes in Complex Networks

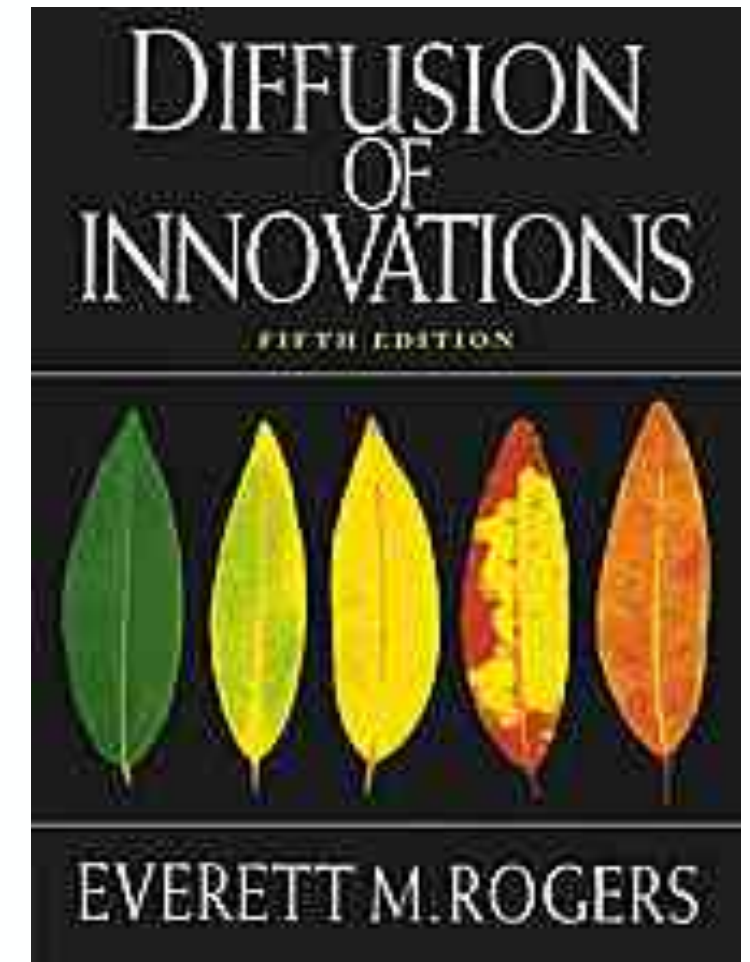
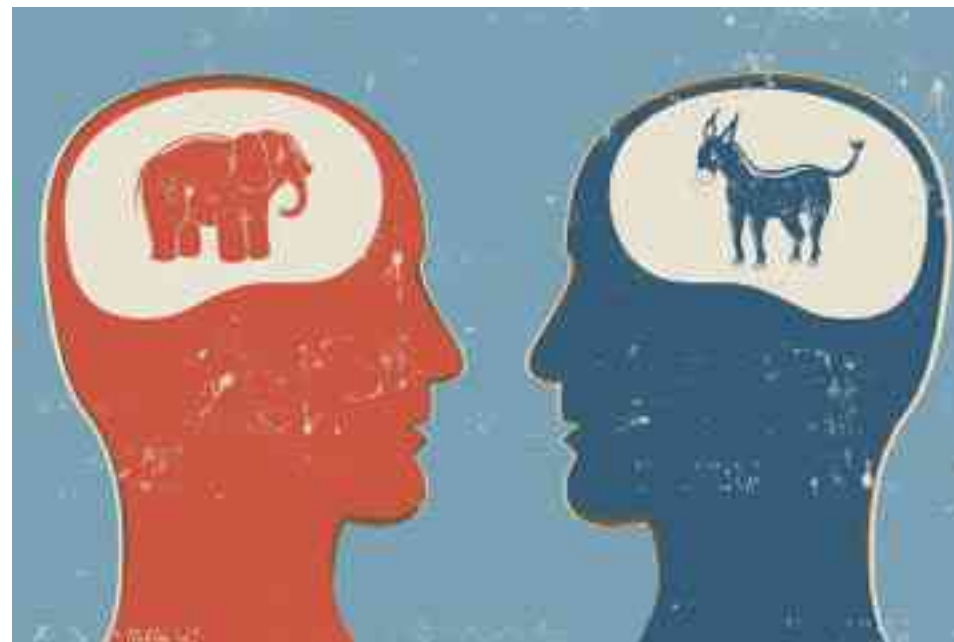
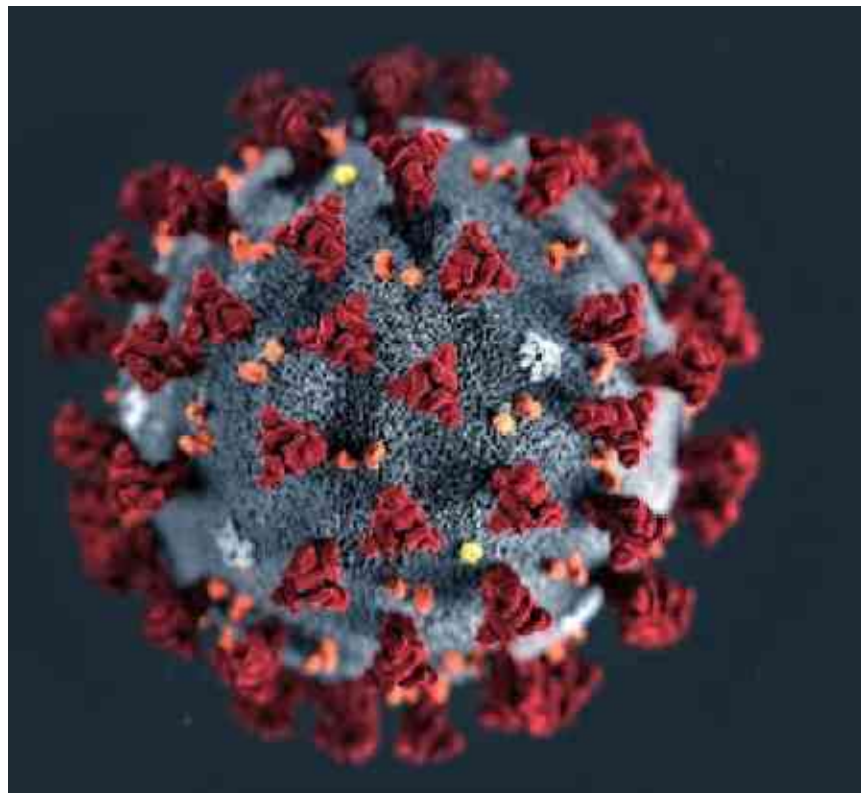
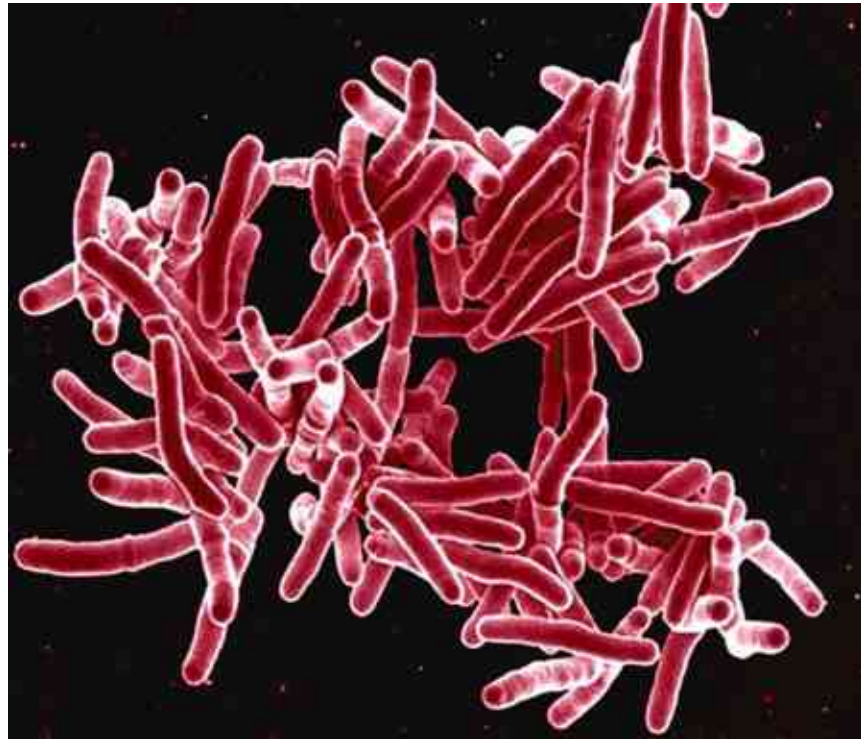
Francisco A. Rodrigues

Instituto de Ciências Matemáticas de Computação

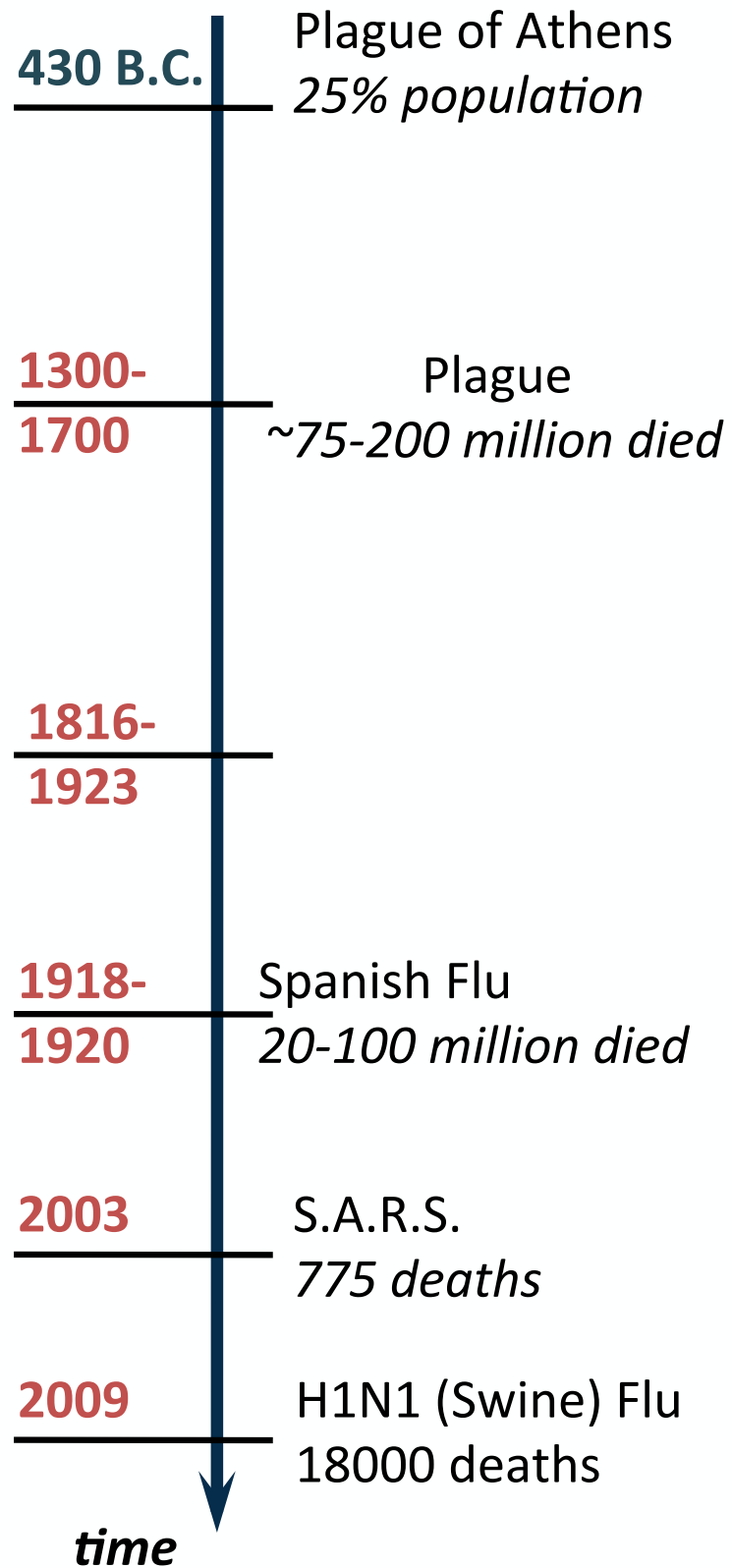
Universidade de São Paulo

francisco@icmc.usp.br

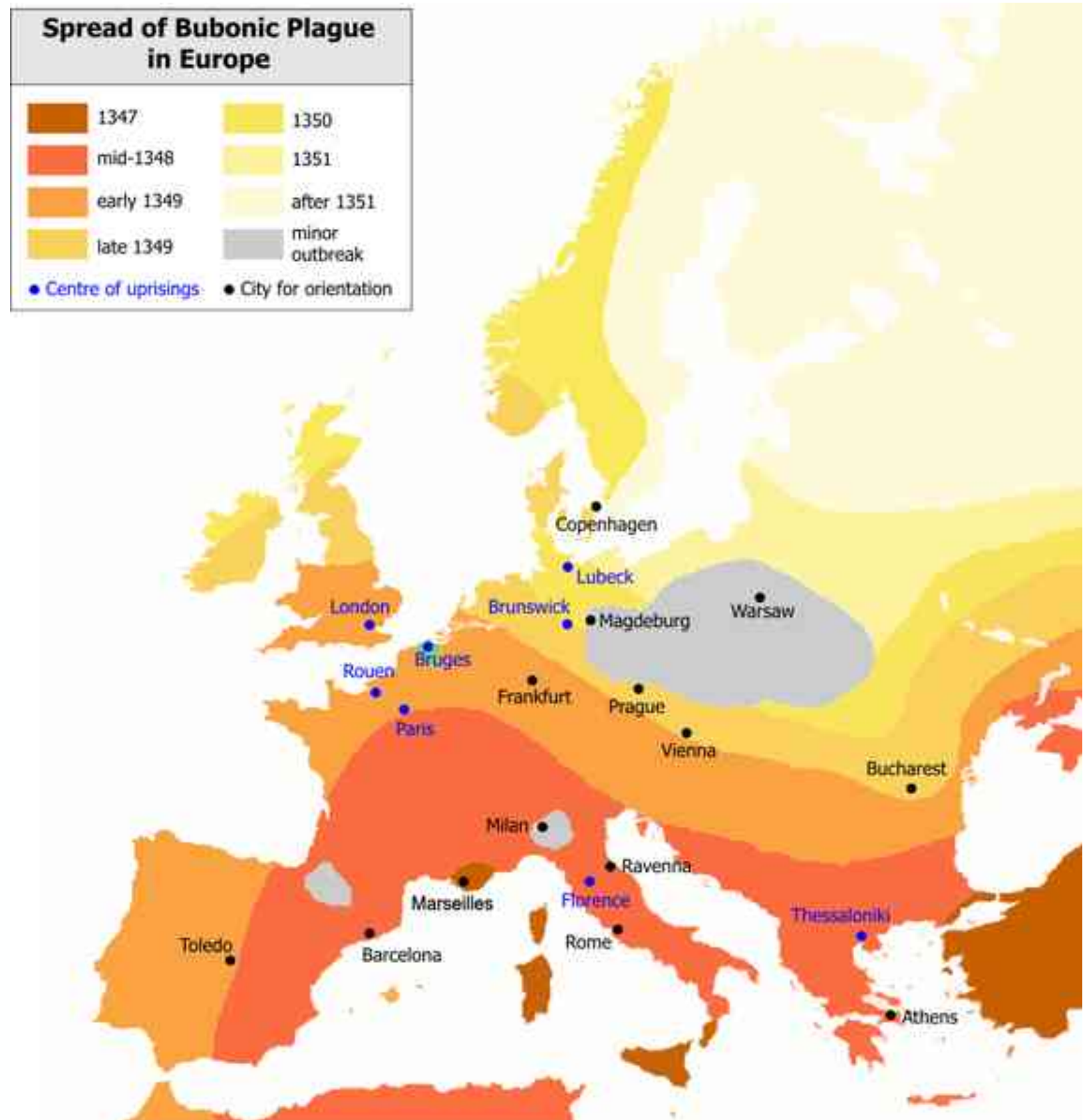
Spreading



Diseases

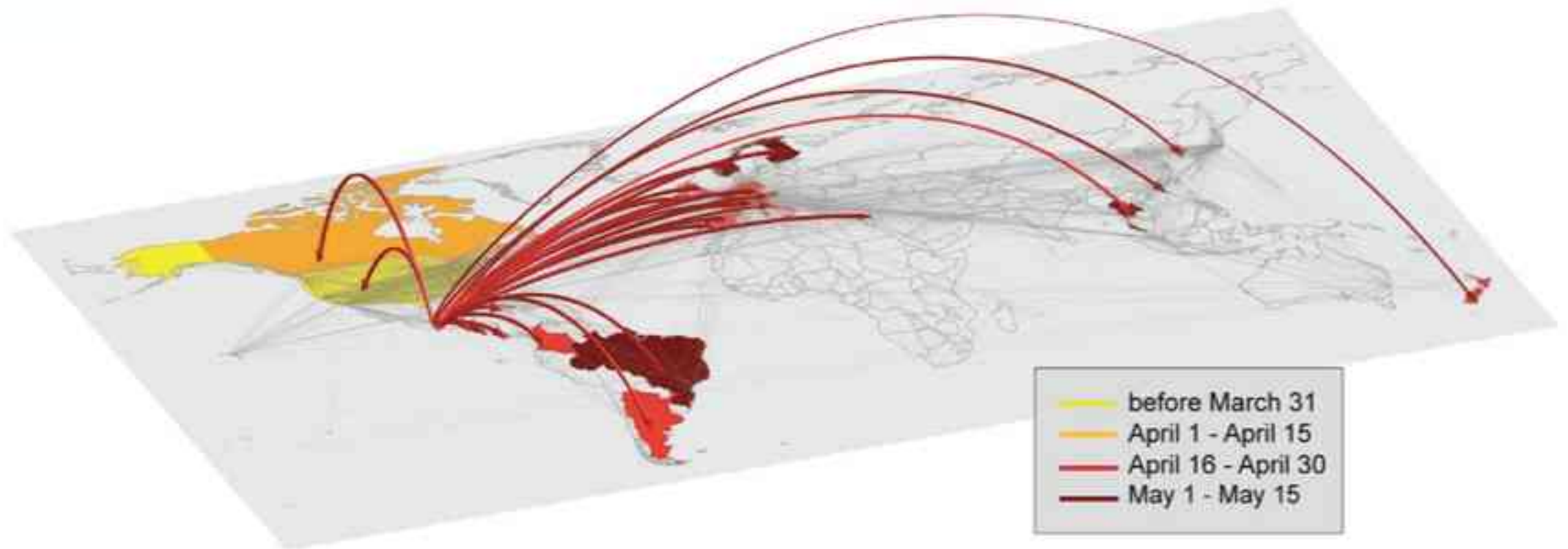


Bubonic Plague



H1N1

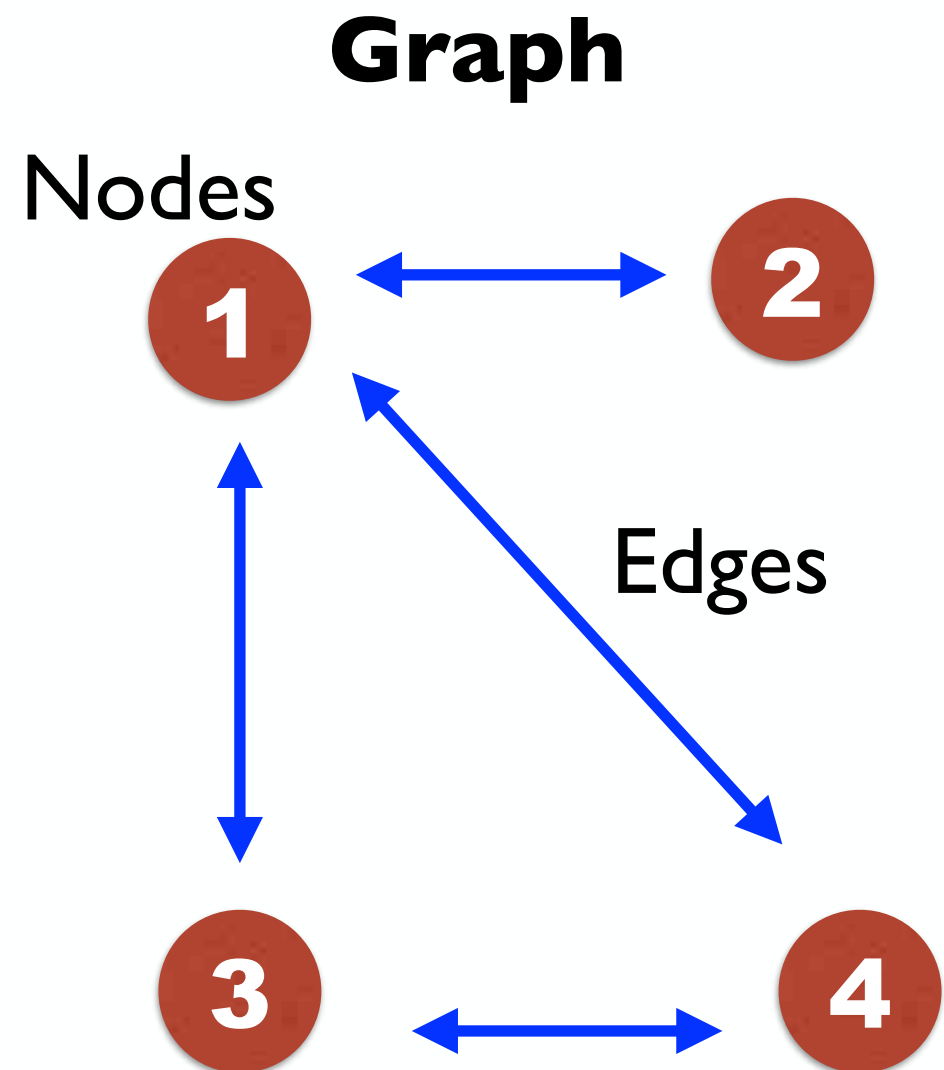
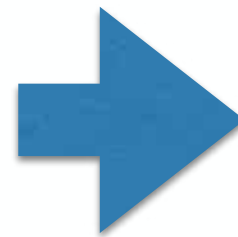
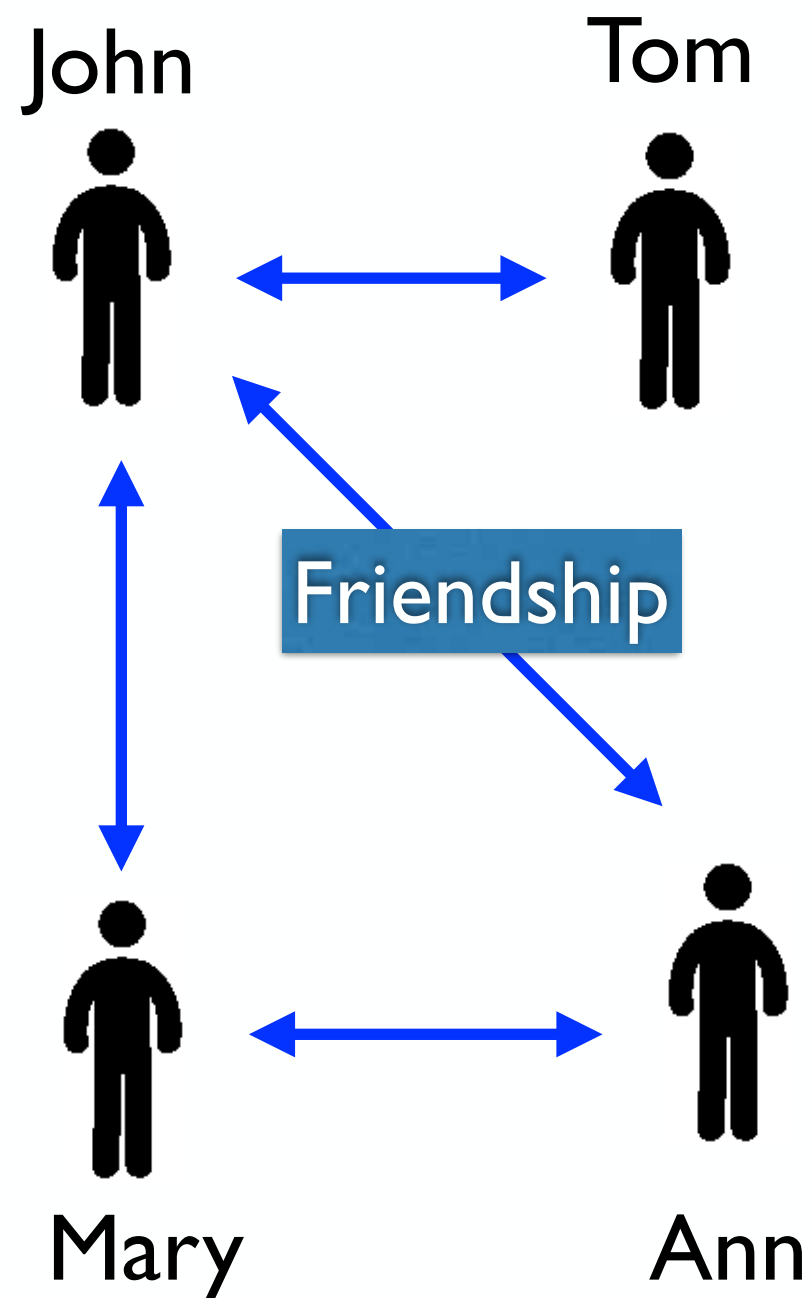
2009 flu pandemic

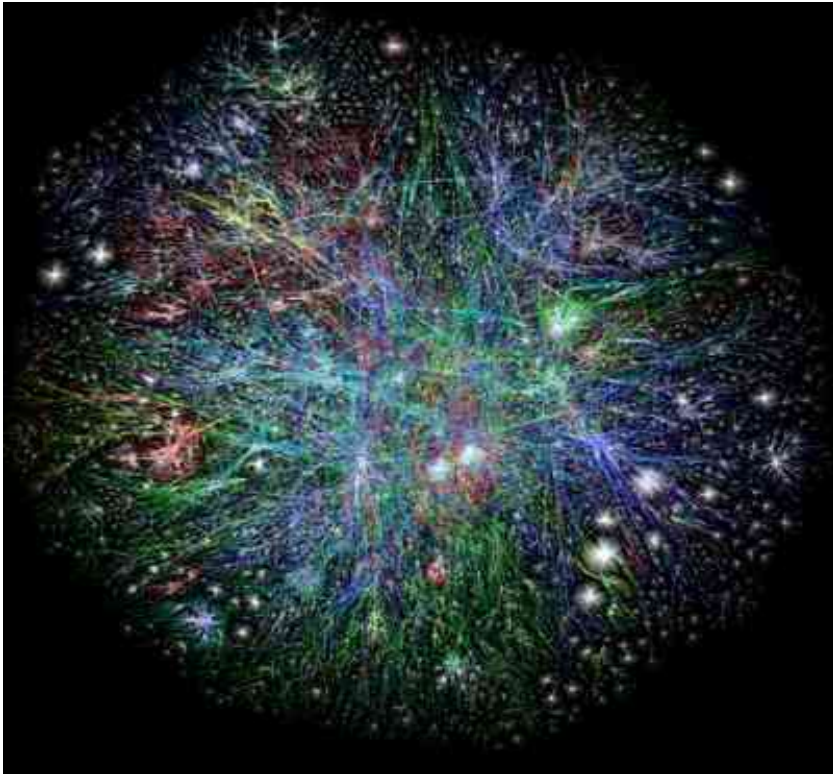


Spreading depends on the network structure!

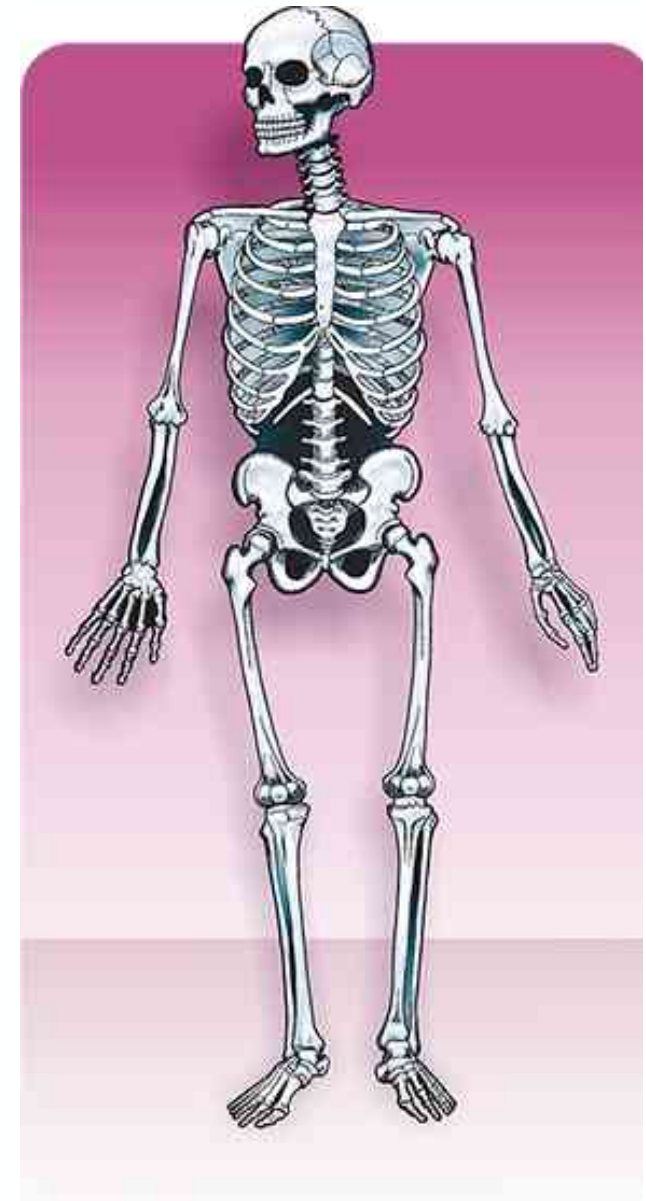


What is a network?



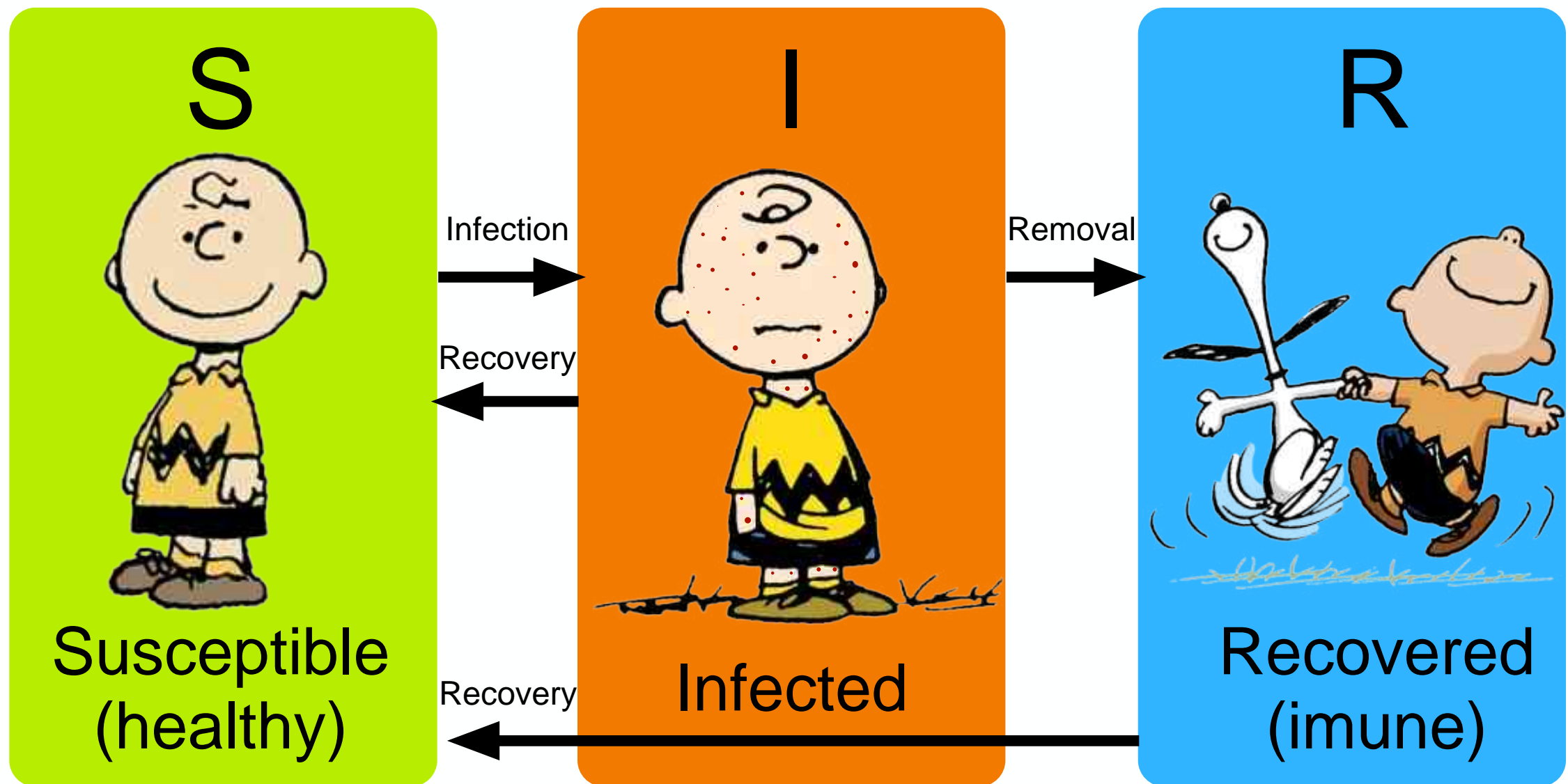


=

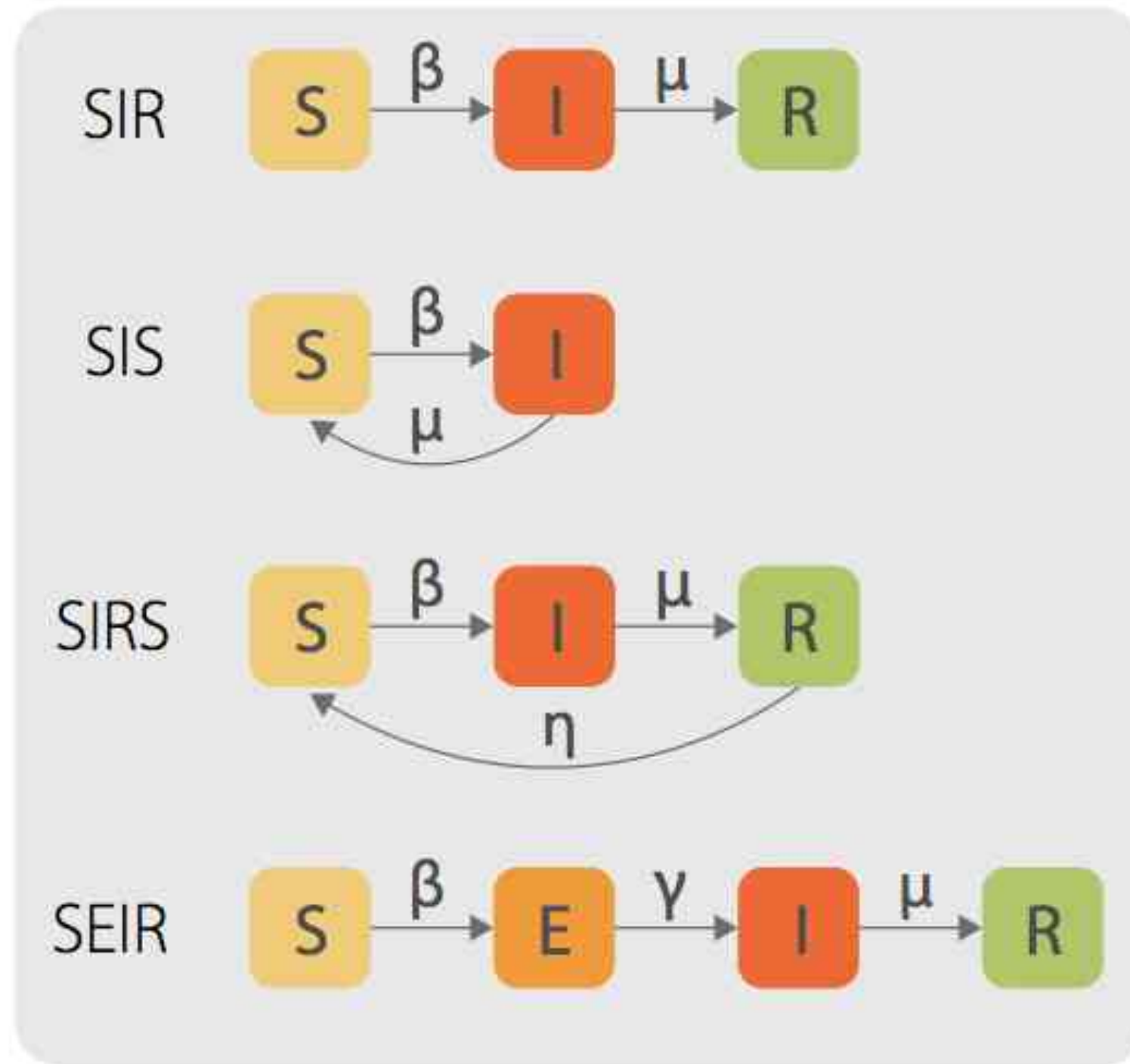


Networks represent the structure of complex systems.

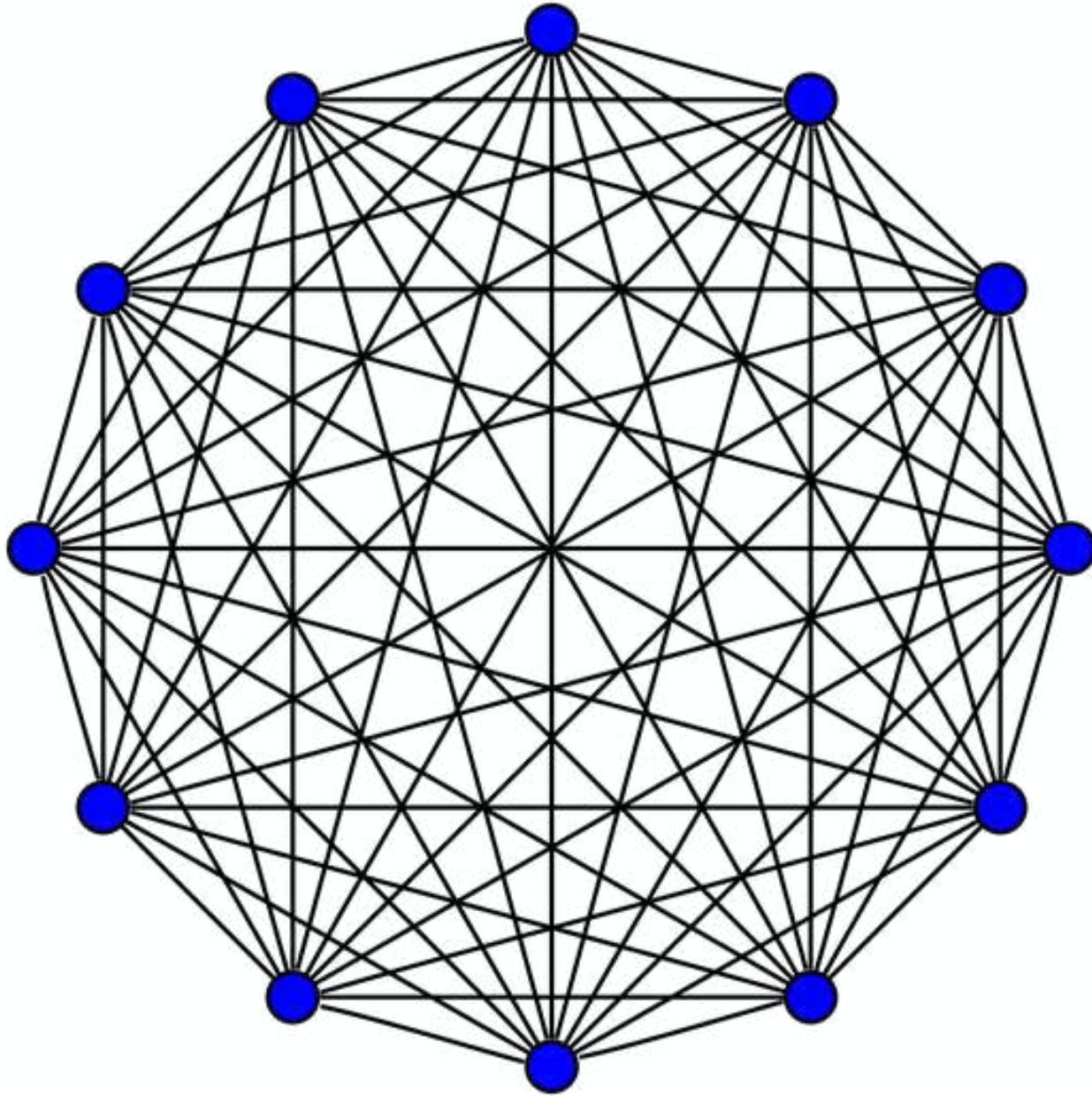
Epidemic models



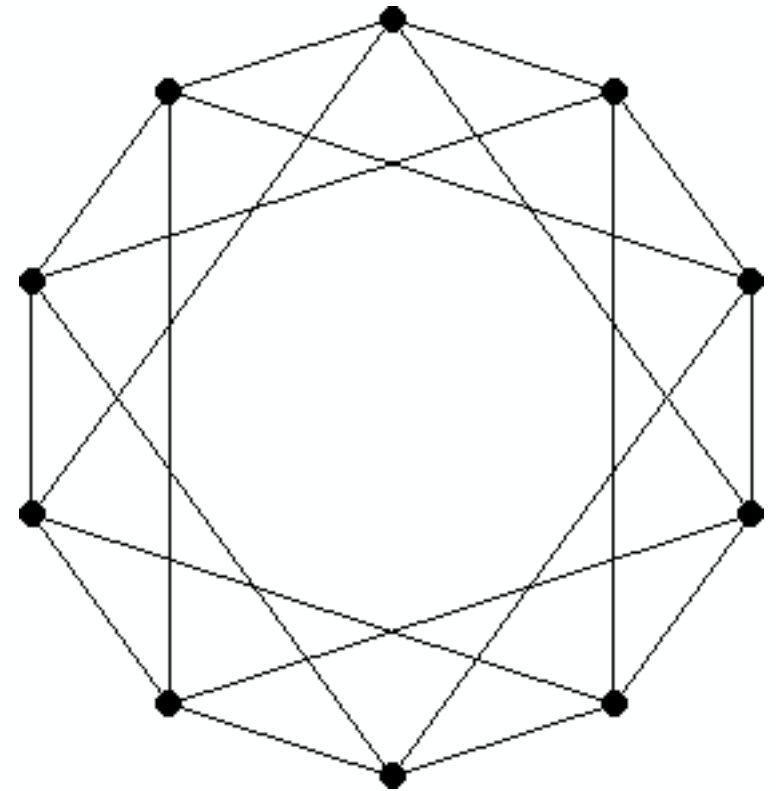
Epidemic models



Epidemic models



Fully connected graph



Each vertex with
k neighbors

Epidemic models

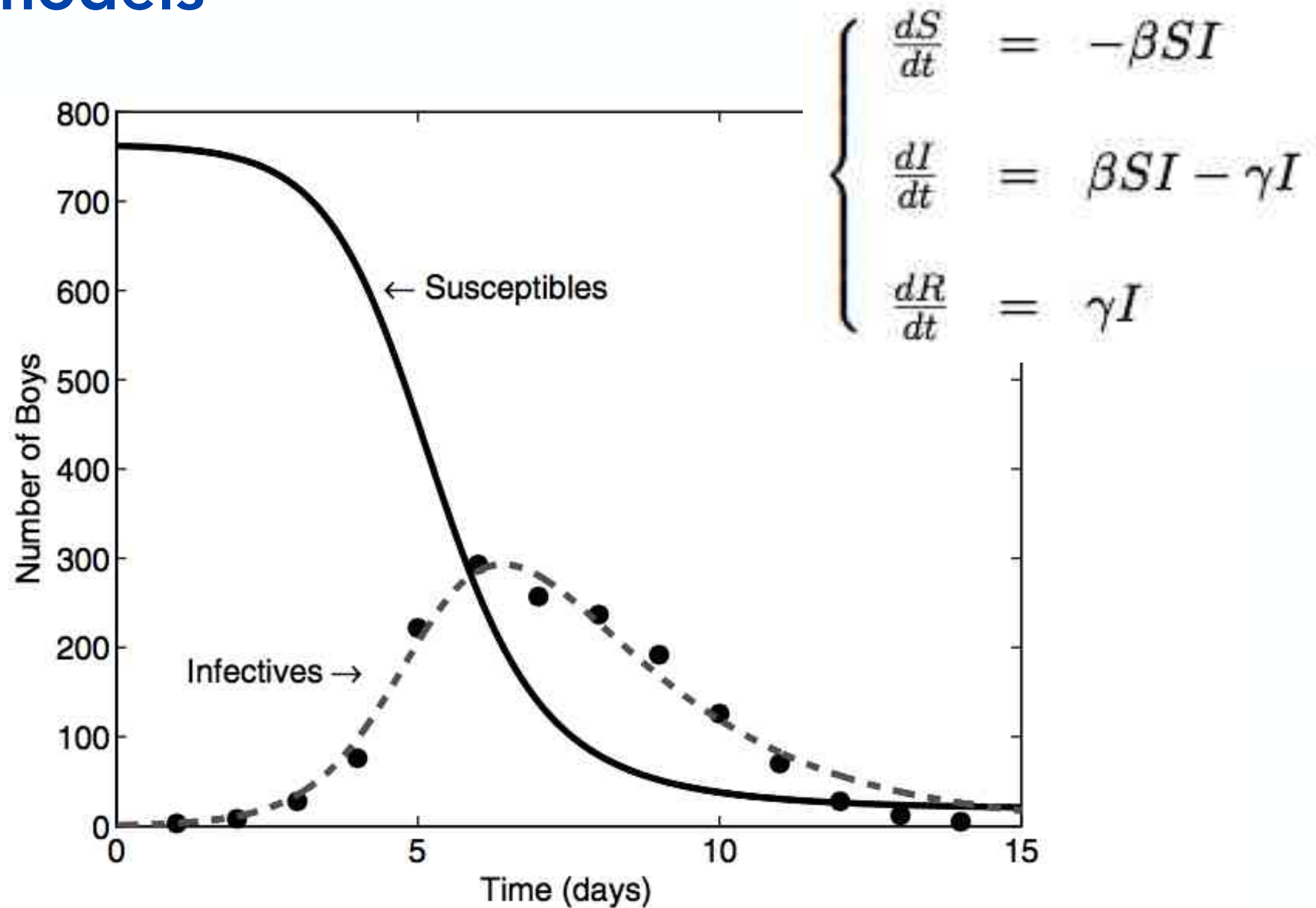
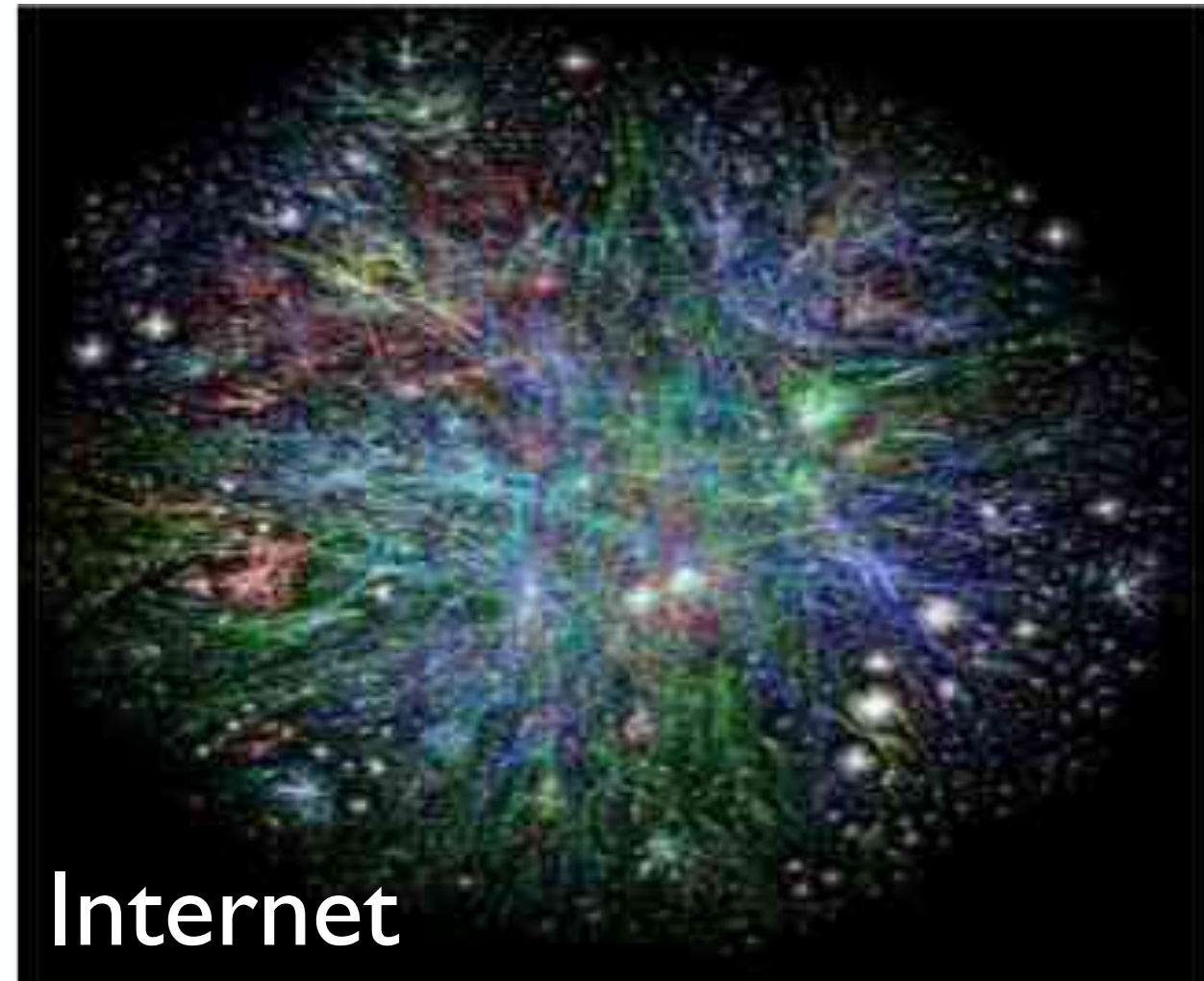
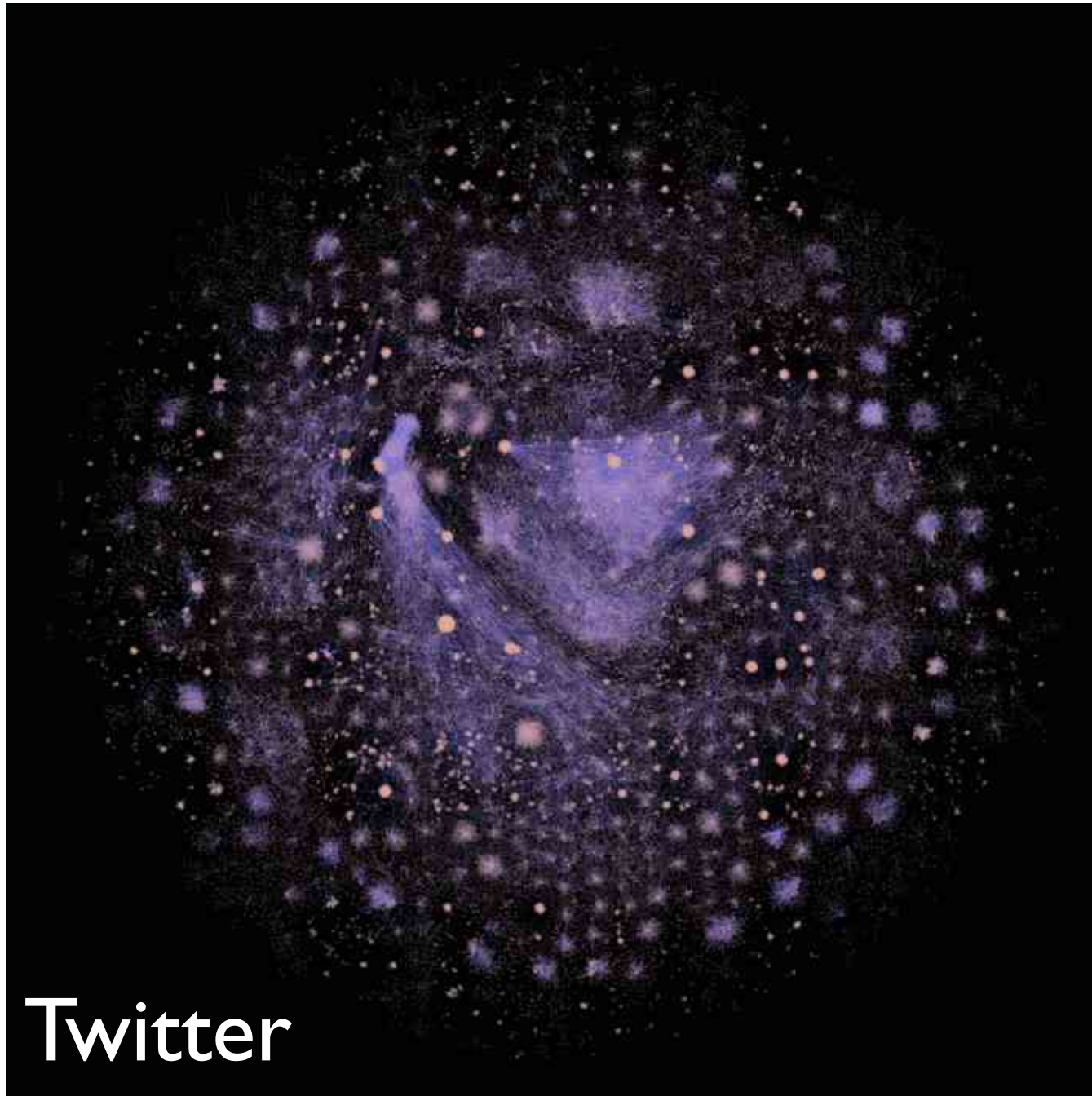


Figure 2.4. The *SIR* dynamics. The filled circles represent the number of boys with influenza in an English boarding school in 1978 (data from the March 4th edition of the *British Medical Journal*). The curves represent solutions from the *SIR* model fitted to the data using least squares. Estimated parameters are $\beta = 1.66$ per day and $1/\gamma = 2.2$ days, giving an R_0 of 3.65.

Networks

$$P(k) \approx k^{-\gamma}$$



Epidemic models in networks

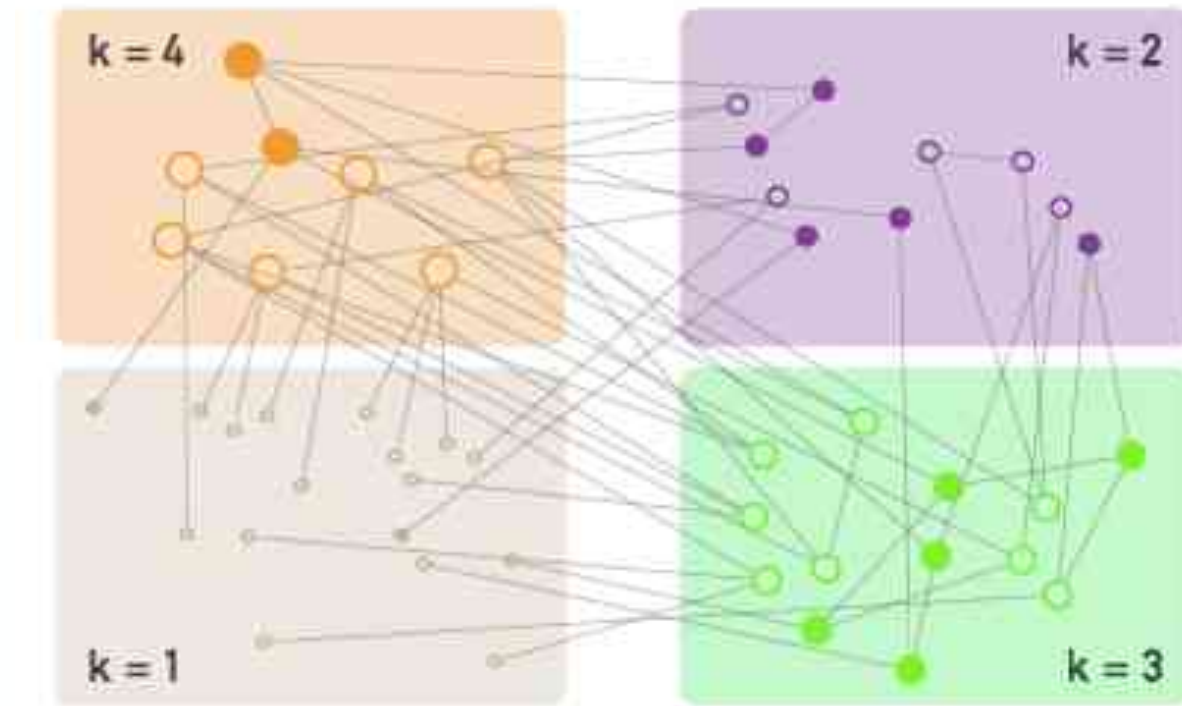
Degree-based mean field: SIS model

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

v

$$\Theta_k = \frac{\sum_{k'} k' p_{k'} i_{k'}}{\langle k \rangle} = \Theta$$

the fraction of
infected neighbors of
a susceptible node k



Keeping only the first order terms:

$$\frac{di_k}{dt} = \beta k \Theta - \mu i_k$$

Multiplying the equation with $(k-1)p_k / \langle k \rangle$ and summing over k

$$\frac{d\Theta}{dt} = \left(\beta \frac{\langle k^2 \rangle}{\langle k \rangle} - \mu \right) \Theta \longrightarrow \boxed{\Theta(t) = C e^{t/\tau}}$$

$$\boxed{\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle \mu}}$$

characteristic
time

Epidemic models in networks

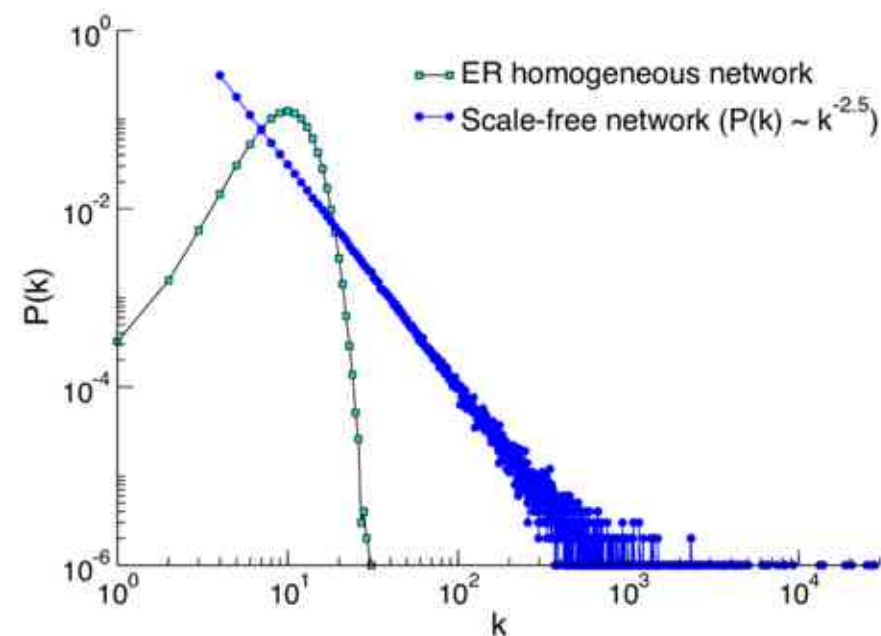
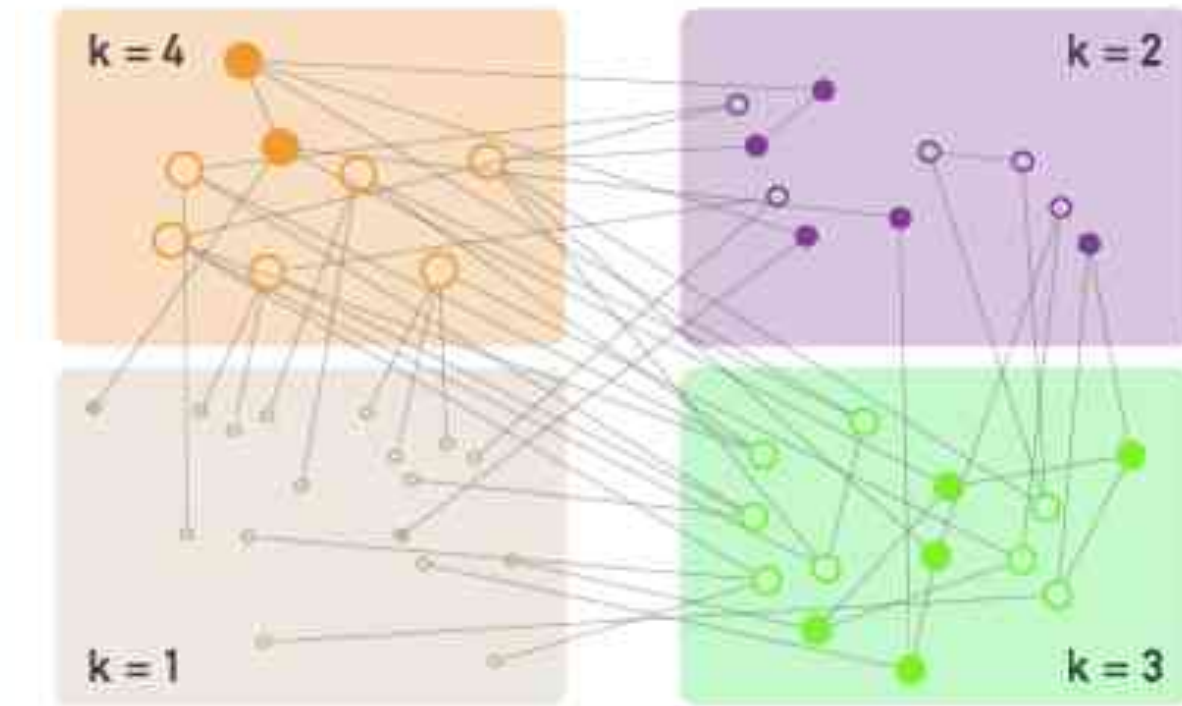
Degree-based mean field: SIS model

$$\Theta(t) = Ce^{t/\tau}, \quad \tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle \mu}$$

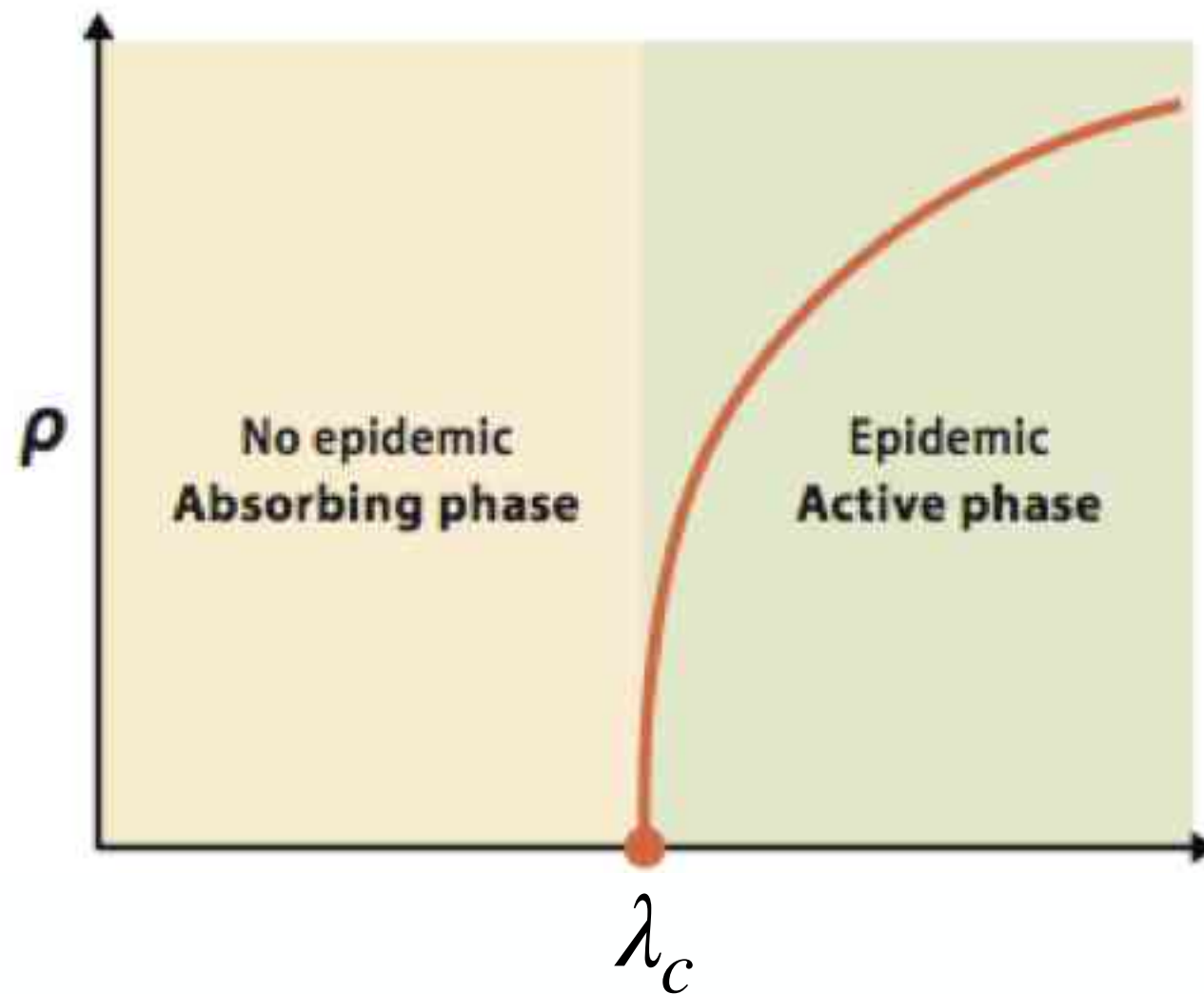
A global outbreak is possible if $\tau > 0$, which yields the condition for a global outbreak as

$$\lambda = \frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle}$$

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

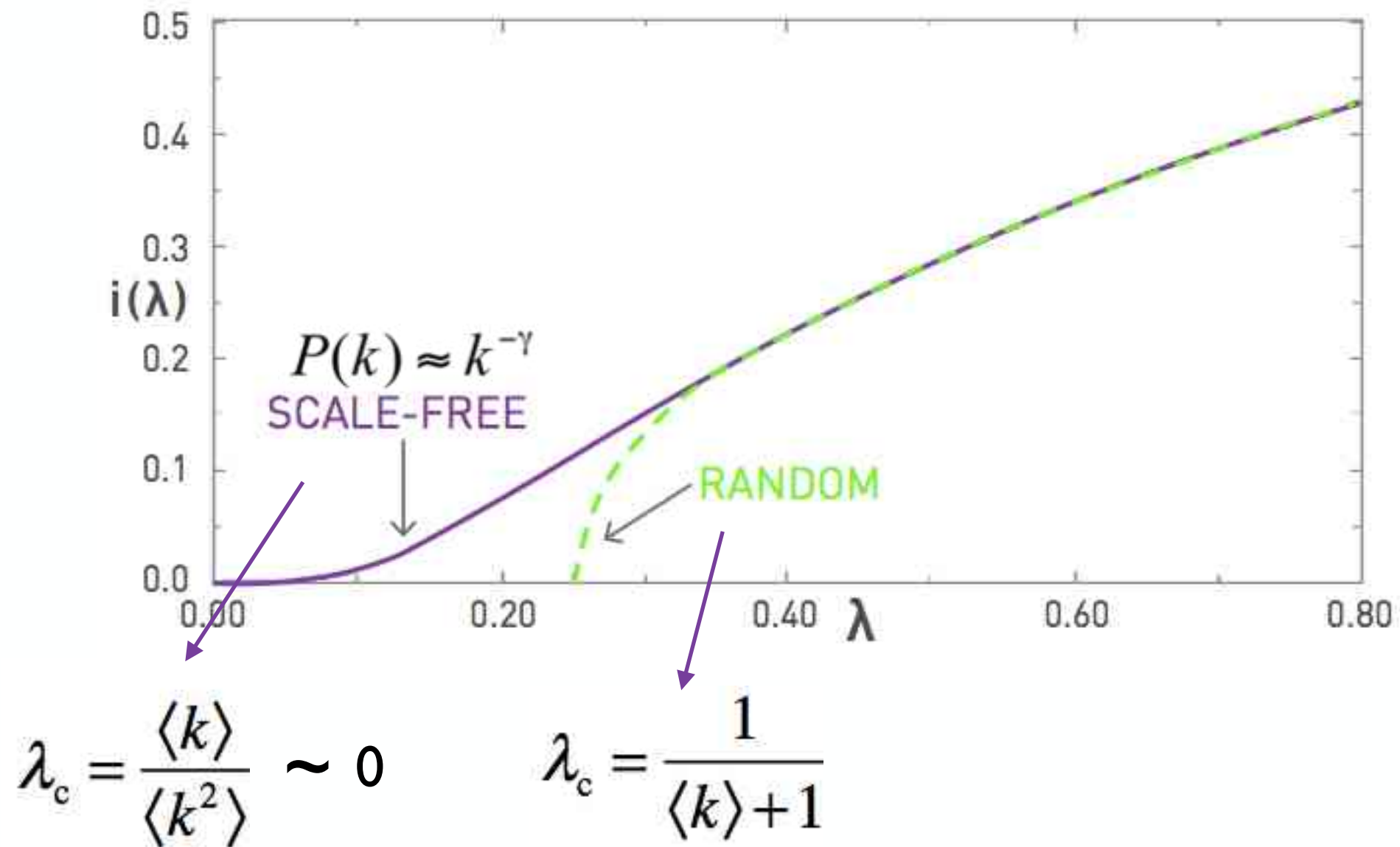


Epidemic models in networks

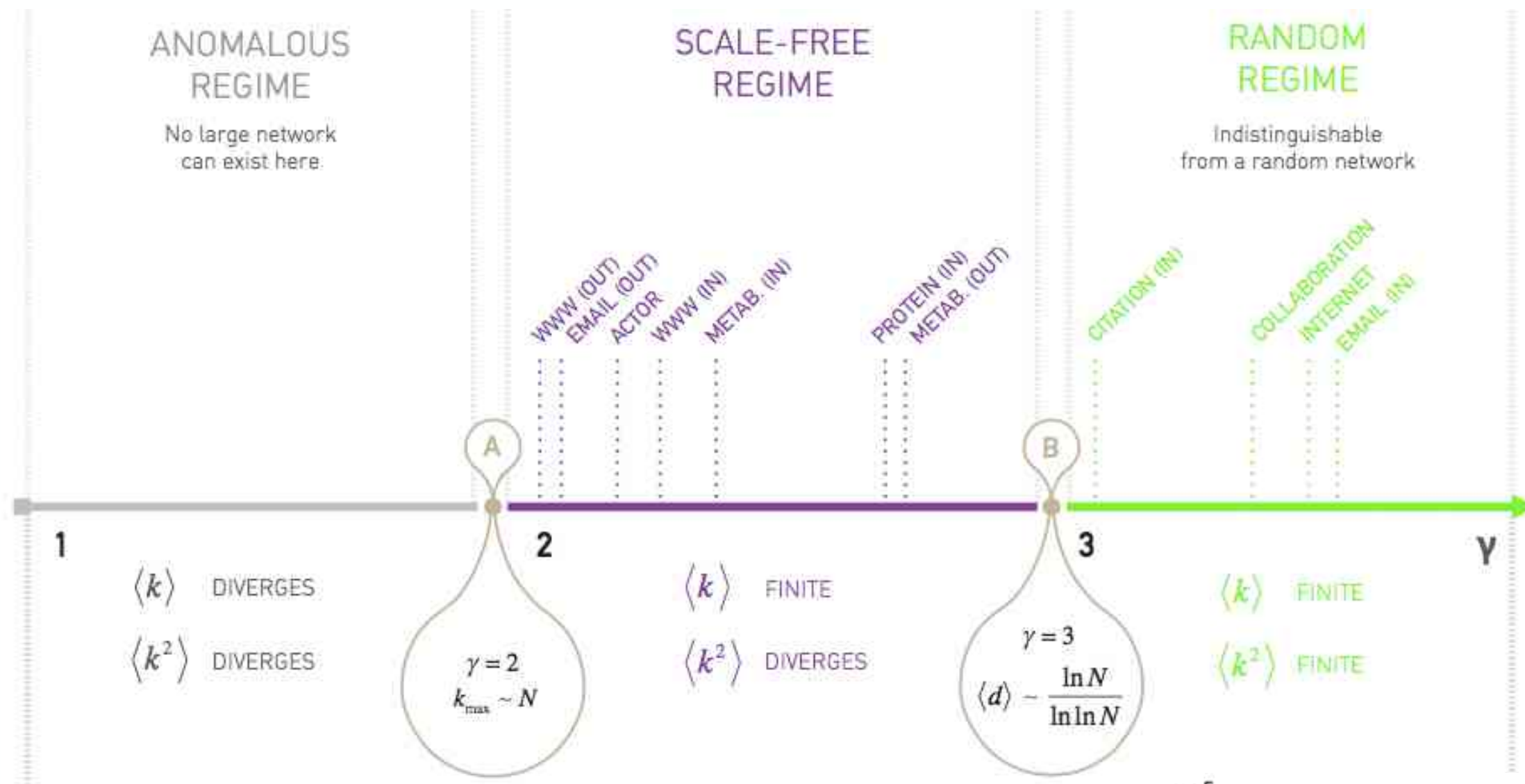


$$\frac{\beta}{\mu} = \lambda_c = \frac{1}{\Lambda_{max}} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

Epidemic models in networks

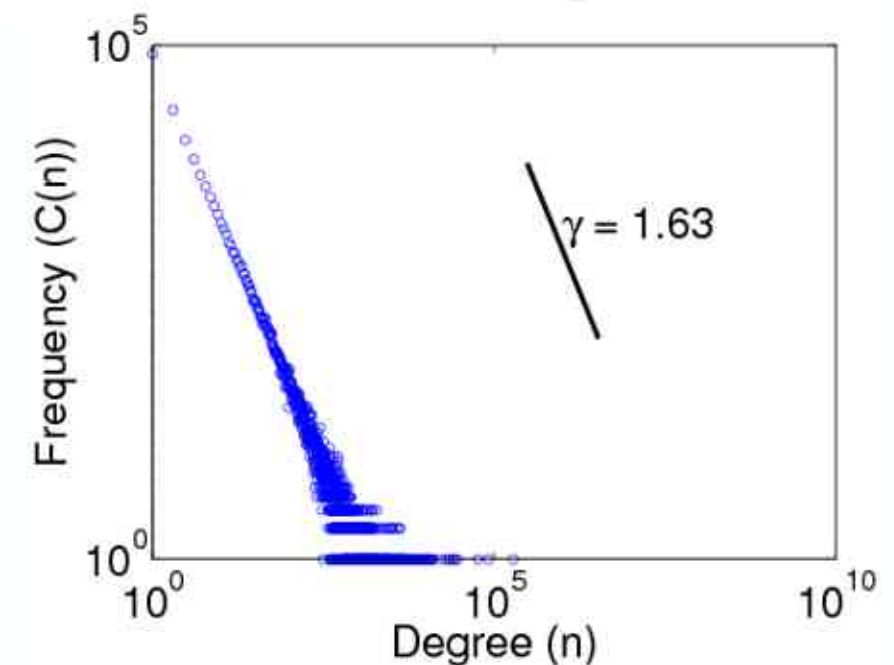


Scale-free networks



$$P(k) \approx k^{-\gamma}$$

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$



Discrete-time Markov chain approach

The evolution of the probability of infection of any node i :

$$p_i(t+1) = (1 - q_i(t))(1 - p_i(t)) + (1 - \mu)p_i(t) + \mu(1 - q_i(t))p_i(t)$$

probability that node i is susceptible ($1 - p_i(t)$) and is infected ($1 - q_i(t)$) by at least a neighbor.

the probability that node i is infected at time t and does not recover

the probability that an infected node recovers ($\mu p_i(t)$) but is re-infected by at least a neighbor ($1 - q_i(t)$).

where the probability of node i not being infected by any neighbor is:

$$q_i(t) = \prod_{j=1}^N (1 - \beta r_{ji} p_j(t))$$

Discrete-time Markov chain approach

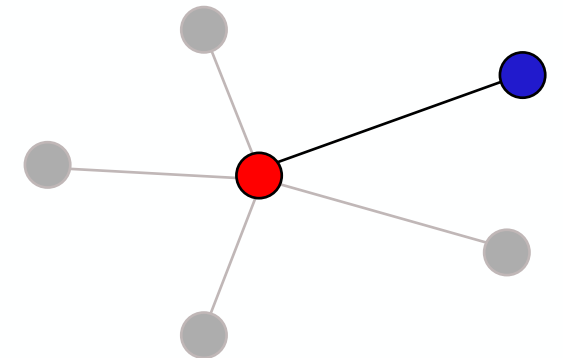
Contact probabilities

$$r_{ij} = 1 - \left(1 - \frac{w_{ij}}{w_i}\right)^{\lambda_i}$$

number of random
walkers leaving node
i at each time step

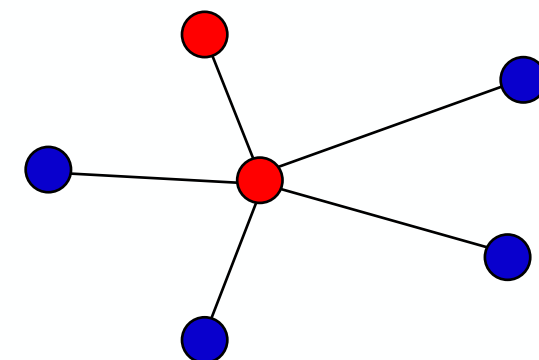
Contact process: R is the transition probability matrix

$$r_{ij} = w_{ij} / w_i$$



Reactive process: R is the adjacency matrix

$$r_{ij} = a_{ij}$$



Discrete-time Markov chain approach

Considering that when $\beta \rightarrow \beta_c$, the probabilities $p_i \approx \epsilon_i$, where $0 < \epsilon_i \ll 1$,

$$q_i(t) = \prod_{j=1}^N (1 - \beta r_{ji} p_j(t)) \longrightarrow q_i \approx 1 - \beta \sum_{j=1}^N r_{ji} \epsilon_j$$

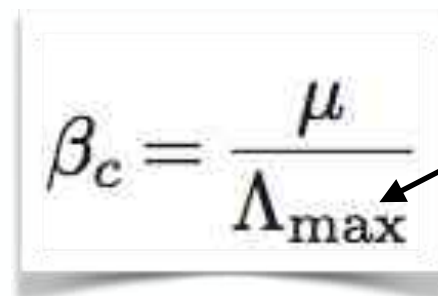
Including in:

$$p_i(t+1) = (1 - q_i(t))(1 - p_i(t)) + (1 - \mu)p_i(t) + \mu(1 - q_i(t))p_i(t)$$

Neglecting second-order terms in ϵ we get

$$\sum_{j=1}^N \left(r_{ji} - \frac{\mu}{\beta} \delta_{ji} \right) \epsilon_j = 0, \quad \forall i = 1, \dots, N,$$

This system has non trivial solutions if and only if μ/β is an eigenvalue of the matrix R . Since we are looking for the onset of the epidemic, the lowest value of β satisfying

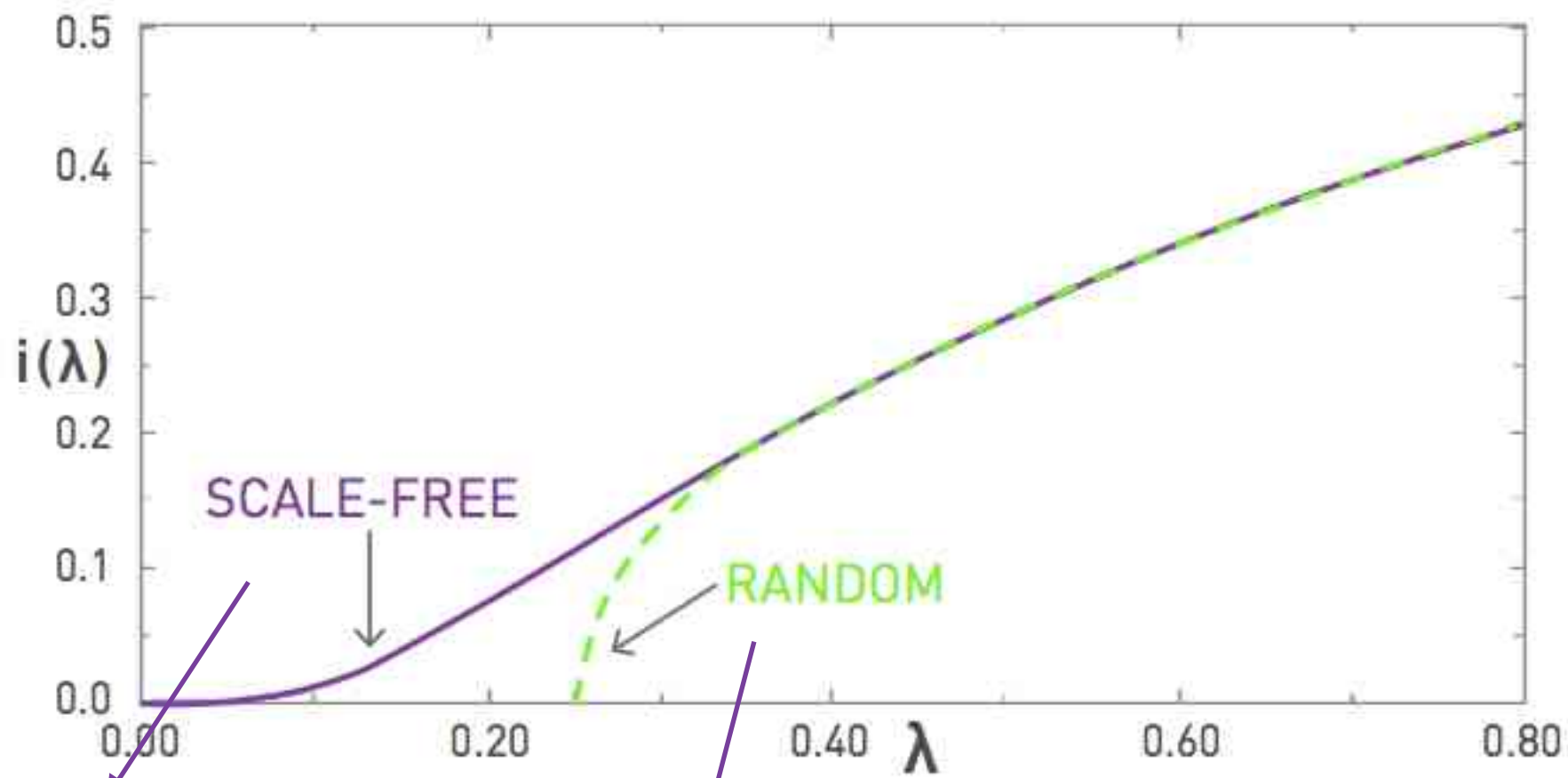
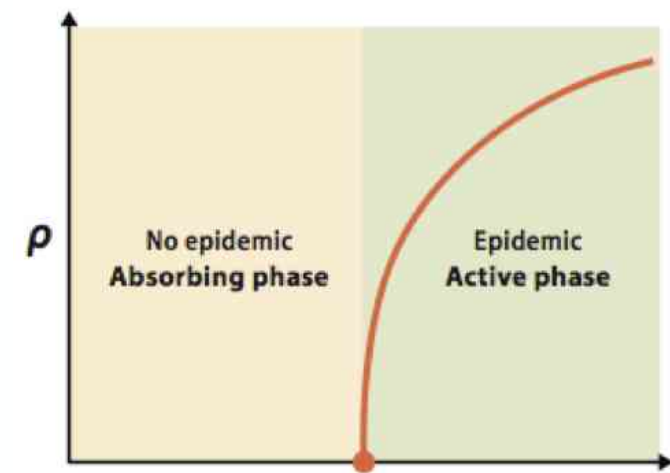

$$\beta_c = \frac{\mu}{\Lambda_{\max}}$$

the largest
eigenvalue of the
matrix R

$$\Lambda_{\max} = \langle k^2 \rangle / \langle k \rangle$$

Gomez et al. EPL 2010.

Epidemic models in networks



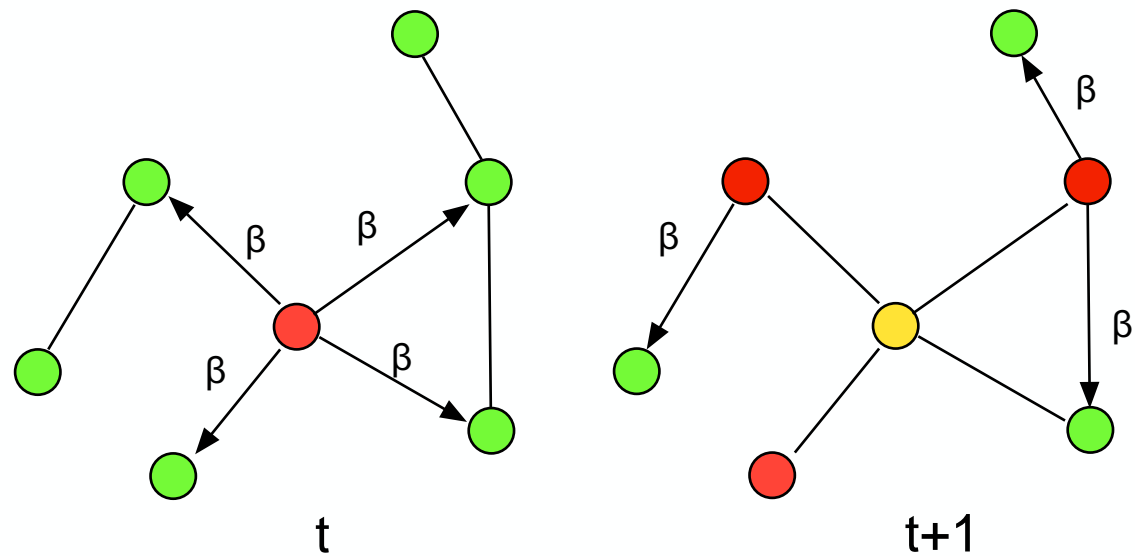
$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \sim 0$$

$$\lambda_c = \frac{1}{\langle k \rangle + 1}$$

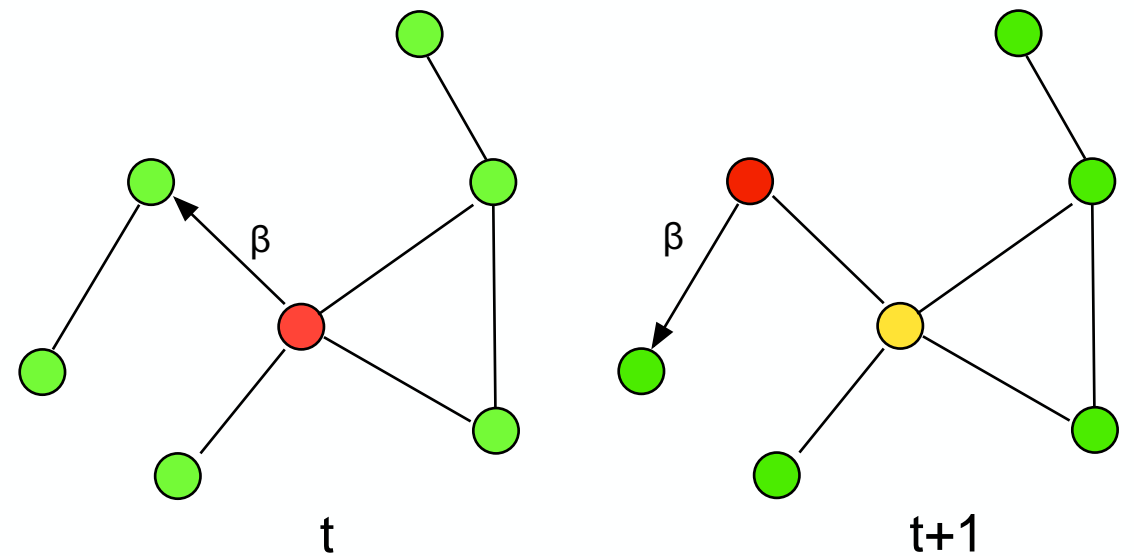
Simulation

Discrete time:

Reactive process

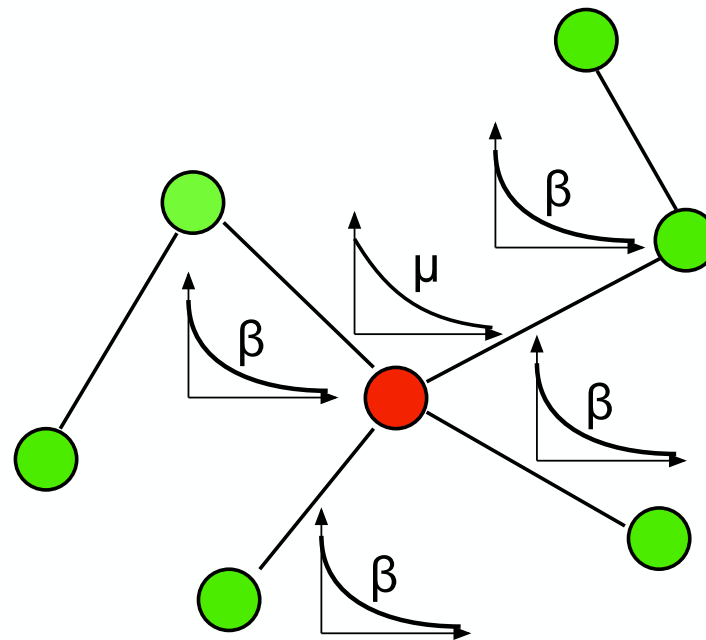


Contact process

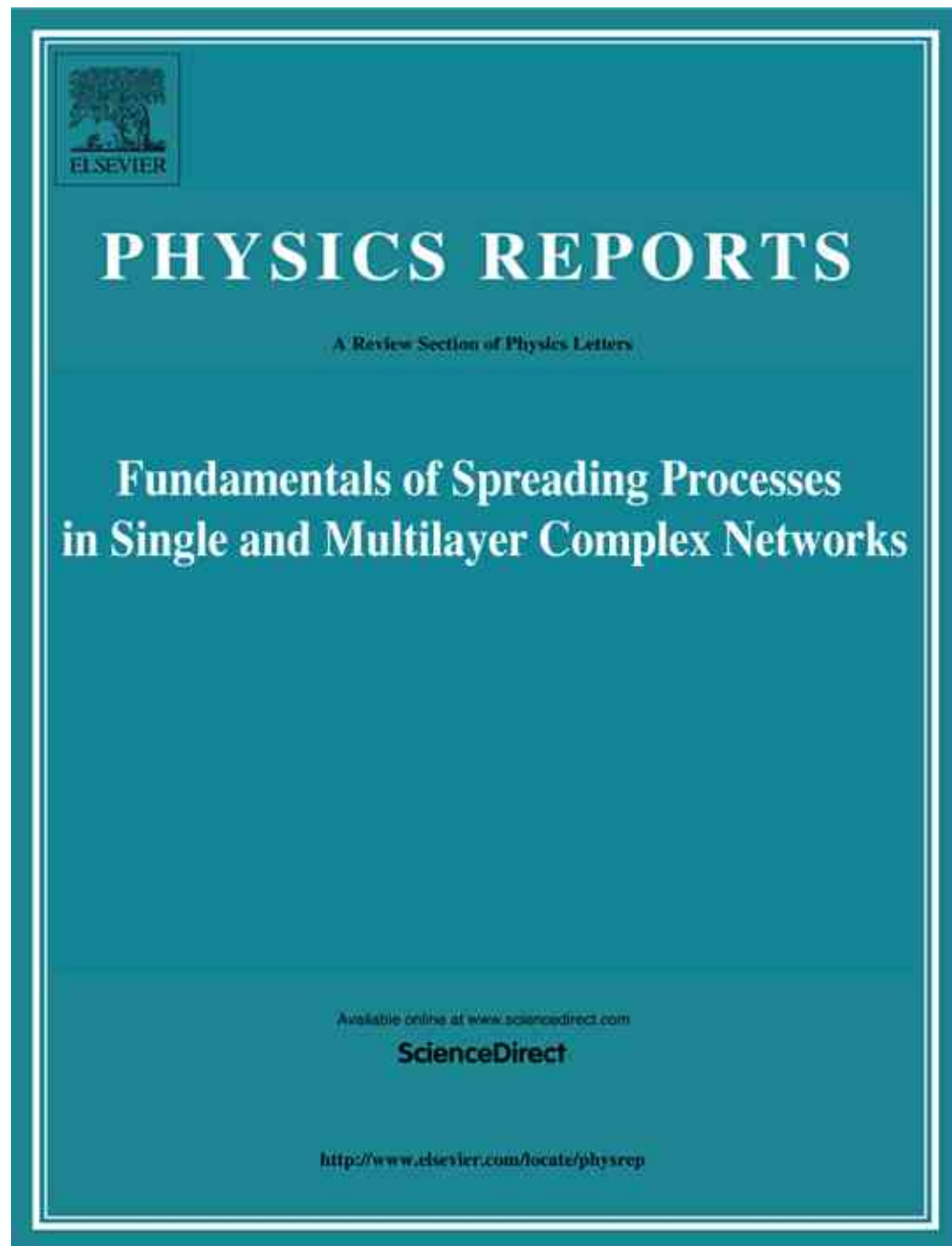


Continuous time:

Poisson Process
Gillespie Algorithm



Epidemic spreading depends on the network structure



1. Mean-field
2. Markov chain
3. Quenched-MF (QMF)
4. Pair approximation
5. Individual based MF
6. Message passing
7. ...

Arruda, Rodrigues and Moreno, 2018

Epidemic spreading with awareness

Disease control

Different strategies:

- Vaccination
- Quarantine
- Disease awareness
- ...



Disease awareness



Disease awareness

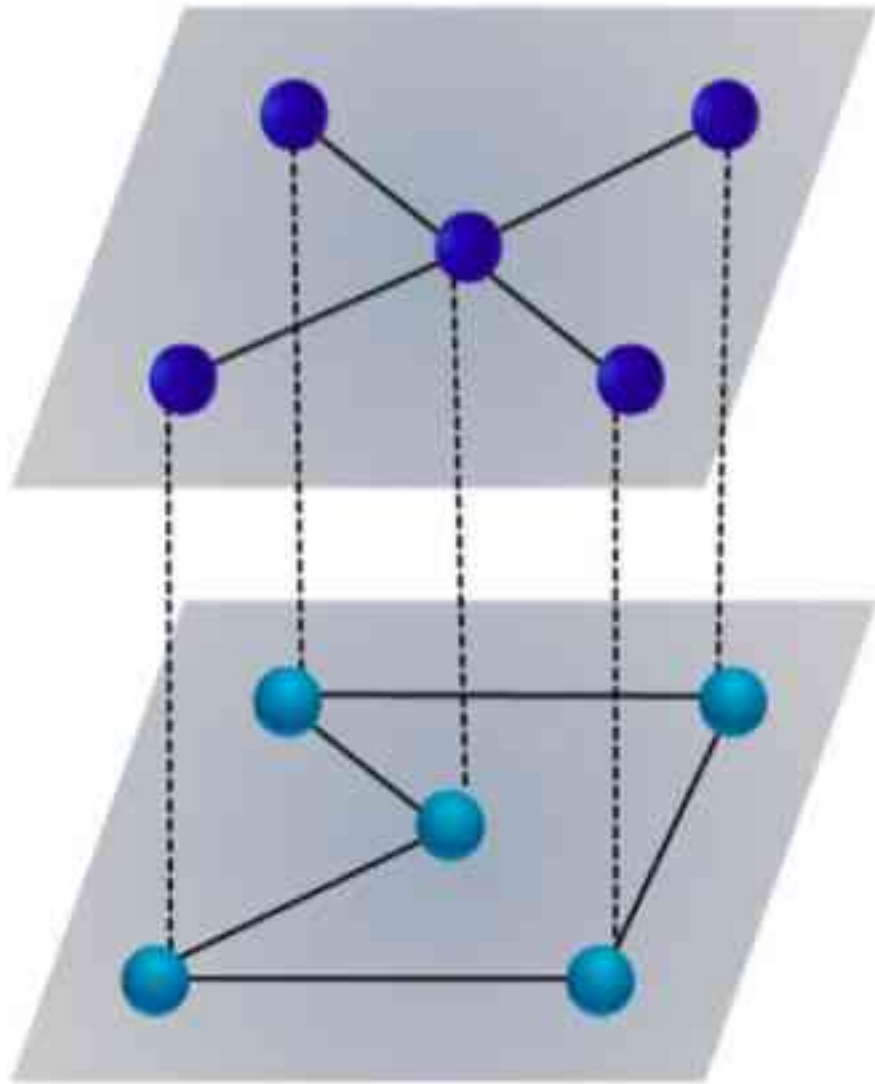
- If you know about the disease, then you try to avoid the infection.
- The probability of infection is reduced.

Cover
your
Cough



Clean
your
Hands

Epidemic spreading with awareness



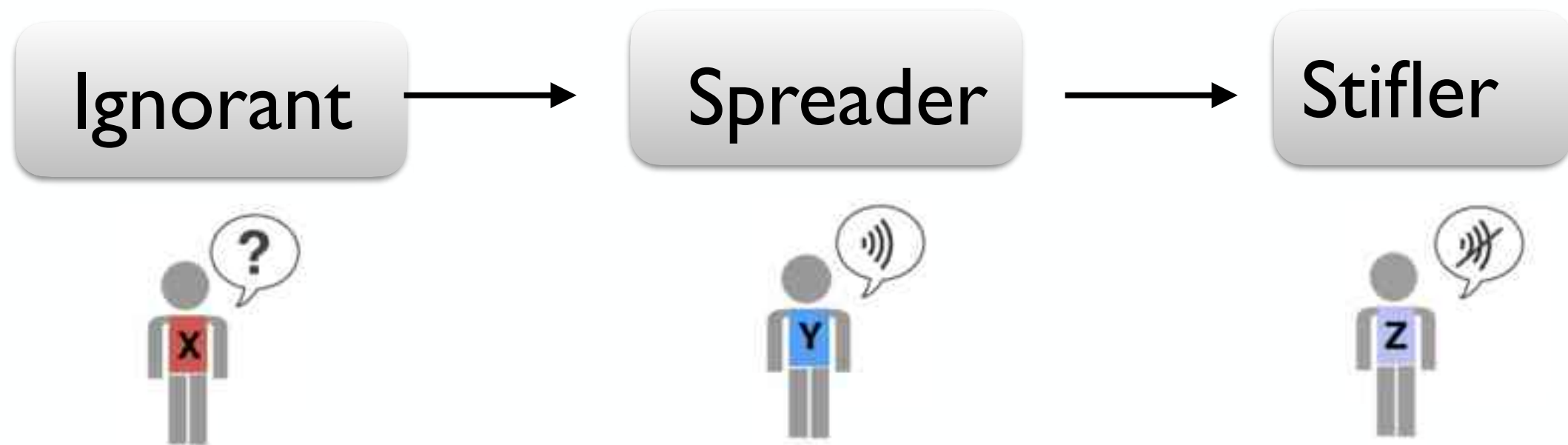
Rumour spreading



Disease transmission

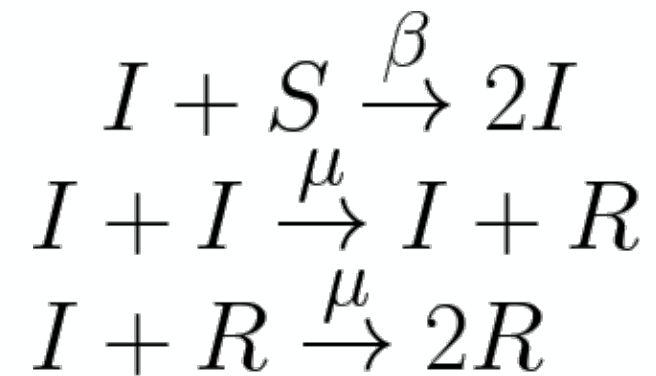


Awareness = Rumour propagation



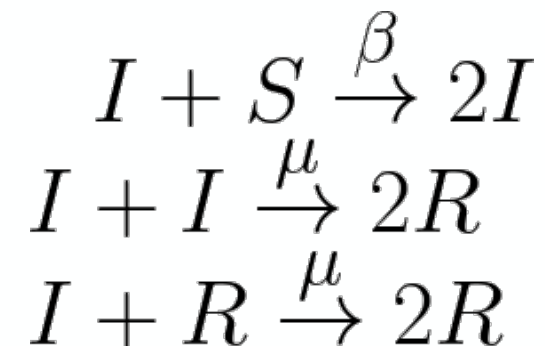
Deterministic compartmental models

Maki-Thompson

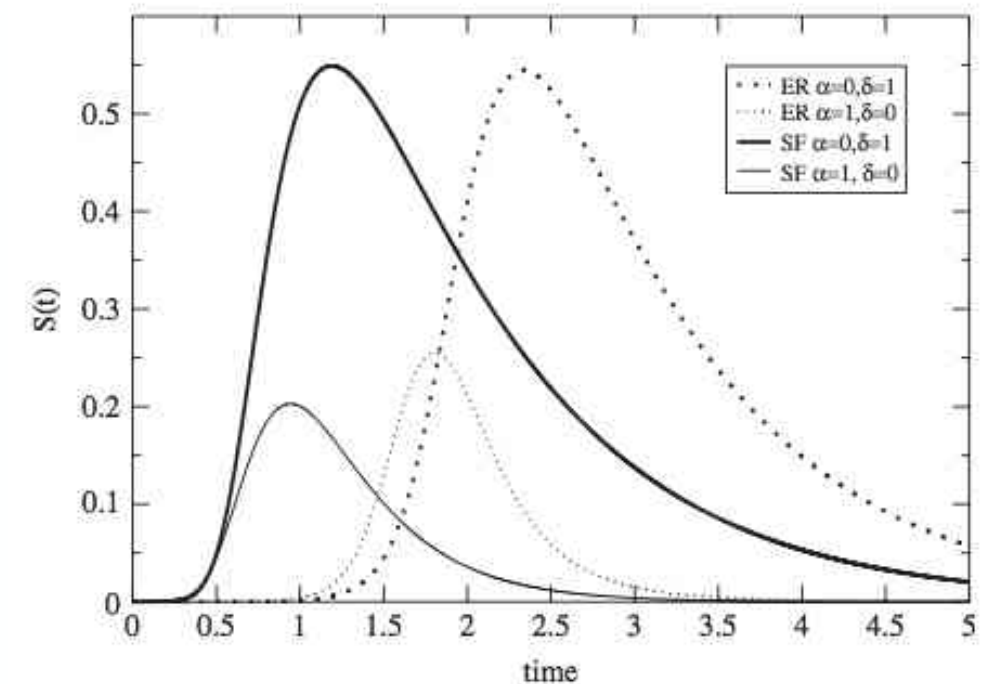
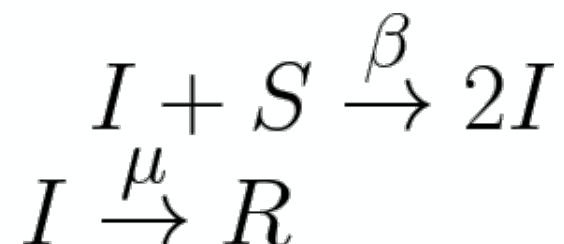


S: Spreader
I: Ignorant
R: Stifler

Daley-Kendal

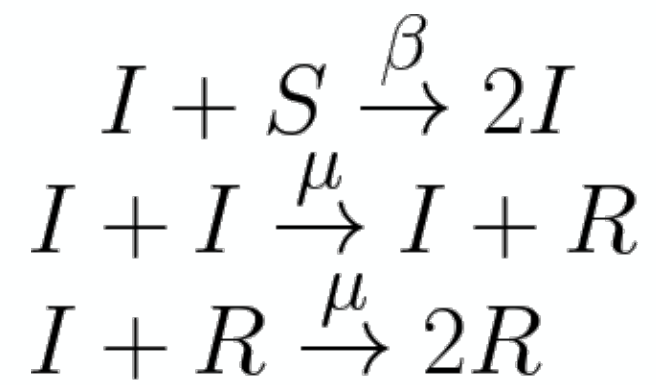
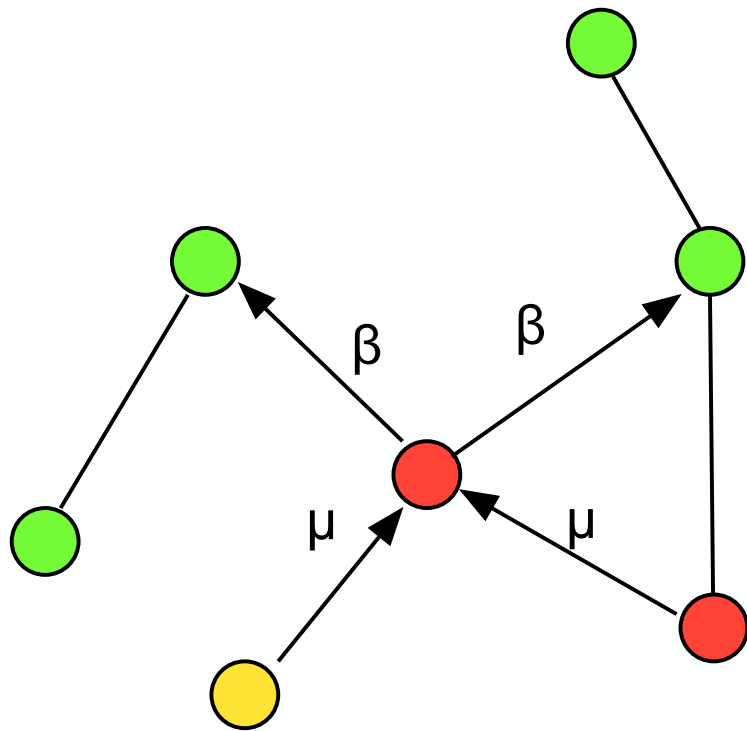


SIR

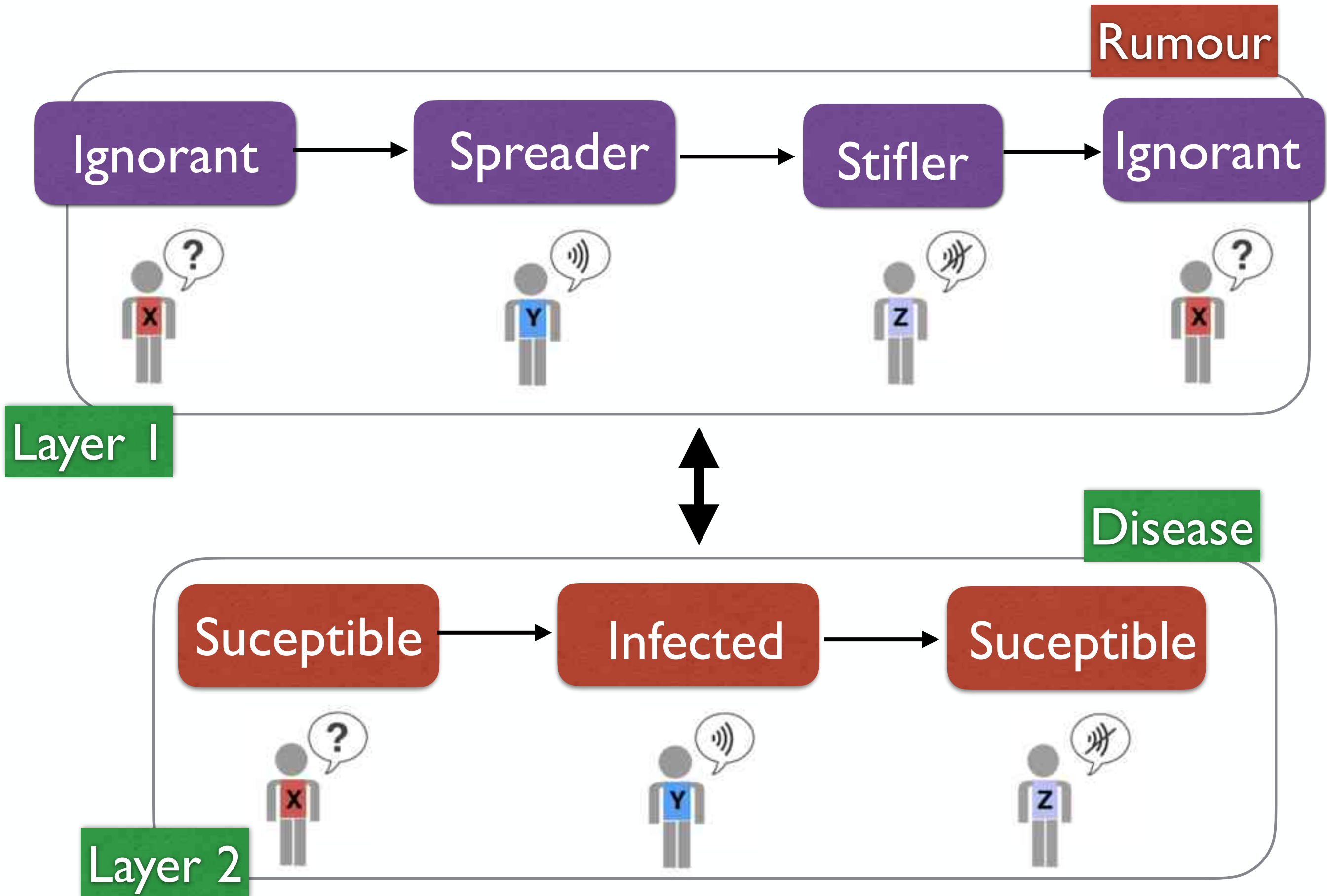


Simulation

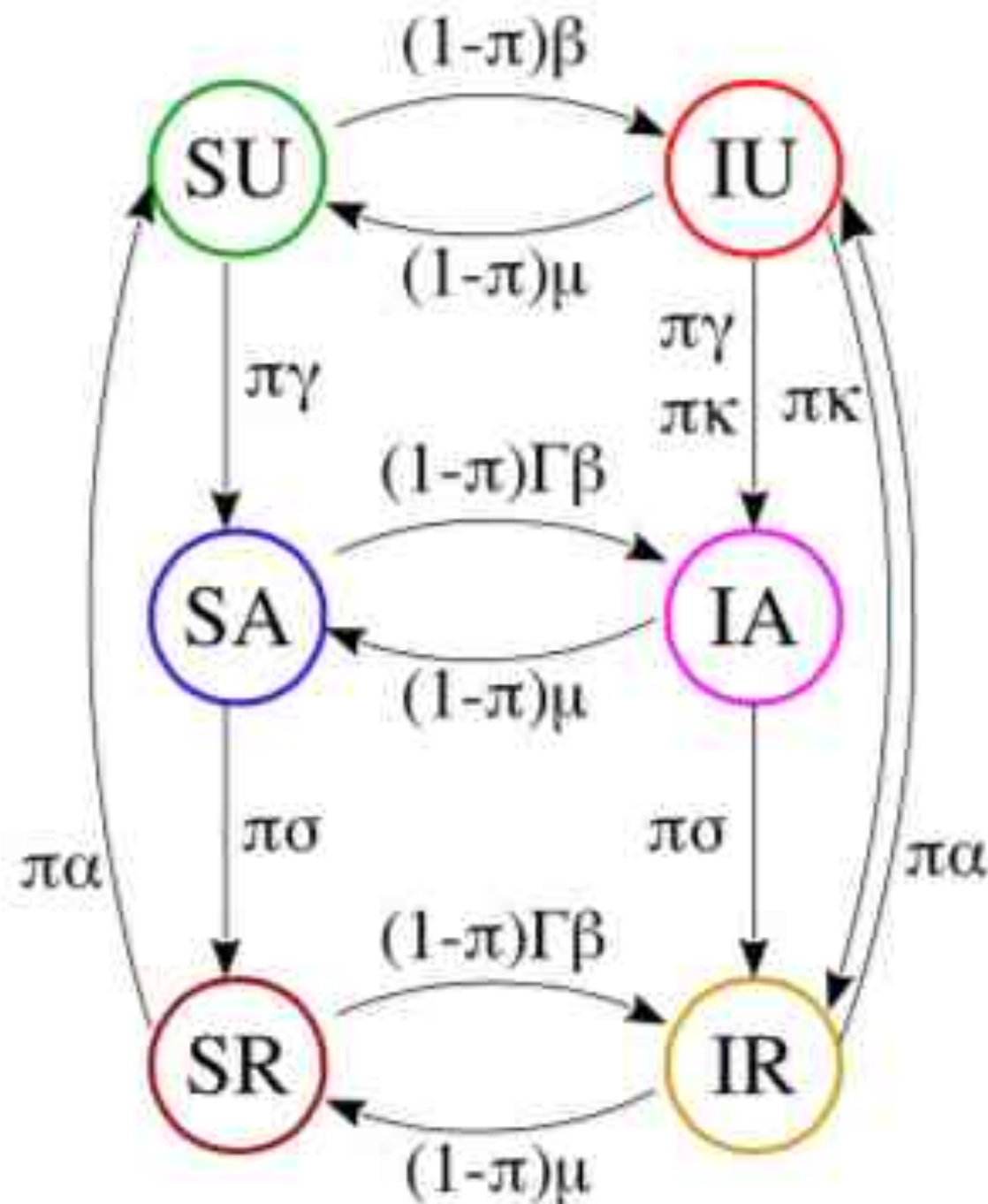
Truncated process



Epidemic spreading with awareness



Epidemic spreading with awareness



- The rumour and disease propagate with different velocities.
- At each time step:
 π : information
 $1 - \pi$: disease

Epidemic spreading with awareness

Markov chain formulation:

π : information
 $1 - \pi$: disease

$$p_{SU}^i(t+1) = \overset{U \rightarrow U}{p_{SU}^i(t)[\pi(1 - r_U^i(t))]} + \overset{S \rightarrow S}{(1 - \pi)(1 - q_U^i(t))}]$$
$$+ \overset{R \rightarrow U}{p_{SR}^i(t)[\pi\alpha]} + \overset{I \rightarrow S}{p_{IU}^i[(1 - \pi)\mu]}$$

q_U^i : infection of an unaware node

r_U^i : awareness by contacting an aware neighbor

r_A^i : stifling - lost of interest

$$\begin{aligned}
p_{SU}^i &= p_{SU}^i[\pi(1 - r_U^i) + (1 - \pi)(1 - q_U^i)] + \\
&\quad + p_{SR}^i[\pi\alpha] + \\
&\quad + p_{IU}^i[(1 - \pi)\mu] \quad (A5)
\end{aligned}$$

$$\begin{aligned}
p_{SA}^i &= p_{SU}^i[\pi r_U^i] + \\
&\quad + p_{SA}^i[\pi(1 - r_A^i) + (1 - \pi)(1 - q_A^i)] + \\
&\quad + p_{IA}^i[(1 - \pi)\mu] \quad (A6)
\end{aligned}$$

$$\begin{aligned}
p_{SR}^i &= p_{SA}^i[\pi r_A^i] + \\
&\quad + p_{SR}^i[\pi(1 - \alpha) + (1 - \pi)(1 - q_A^i)] + \\
&\quad + p_{IR}^i[(1 - \pi)\mu] \quad (A7)
\end{aligned}$$

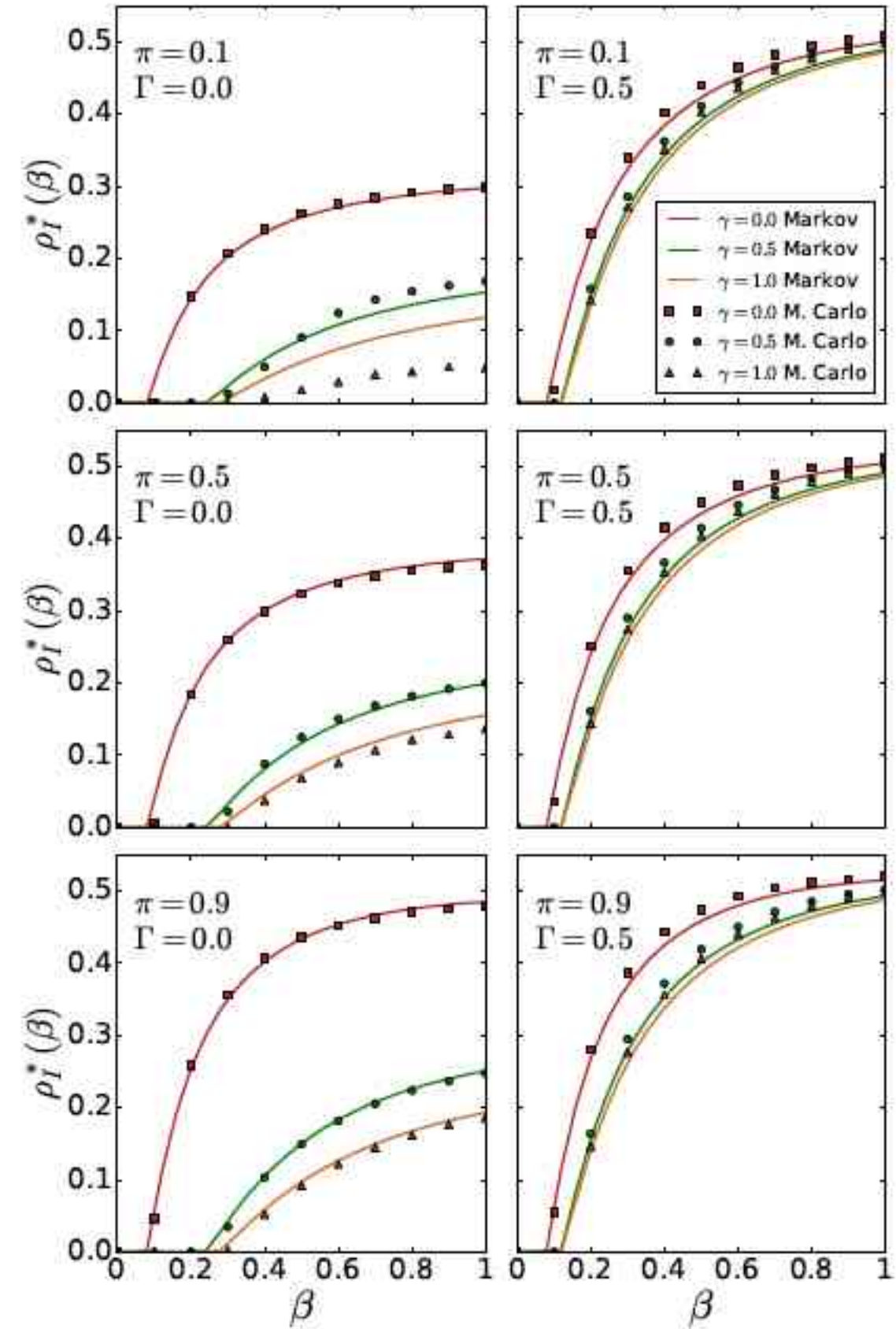
$$\begin{aligned}
p_{IU}^i &= p_{SU}^i[(1 - \pi)q_U^i] + \\
&\quad + p_{IU}^i[\pi(1 - r_U^i)(1 - \kappa) + (1 - \pi)(1 - \mu)] + \\
&\quad + p_{IR}^i[\pi\alpha(1 - \kappa)] \quad (A8)
\end{aligned}$$

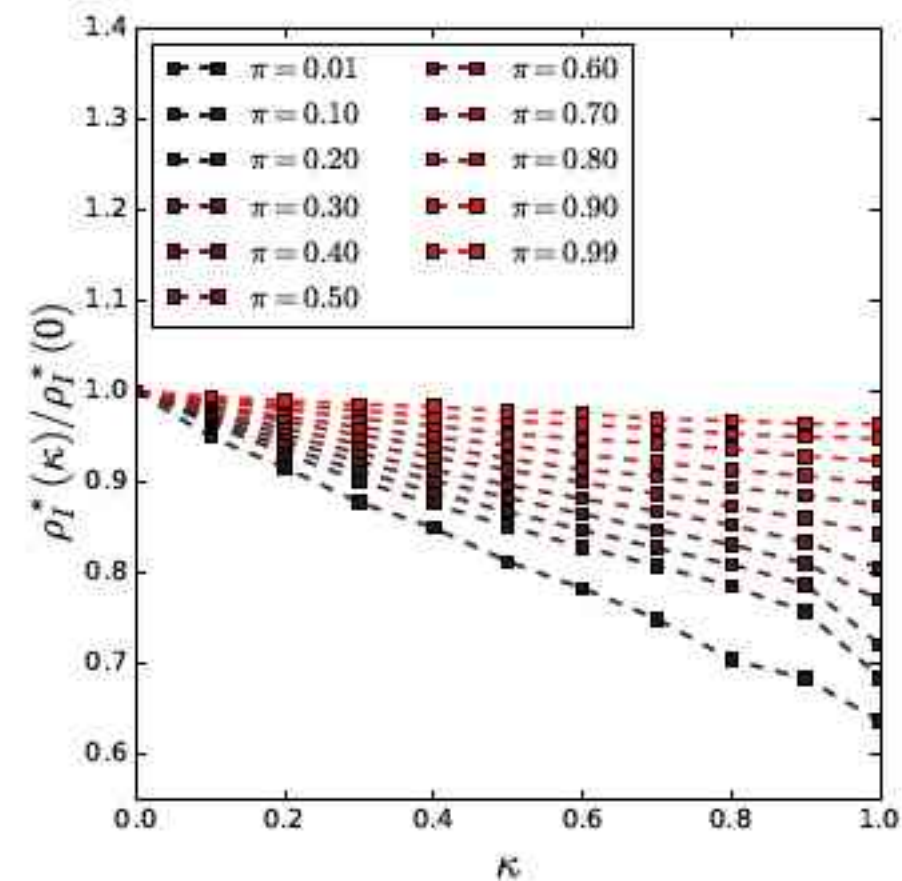
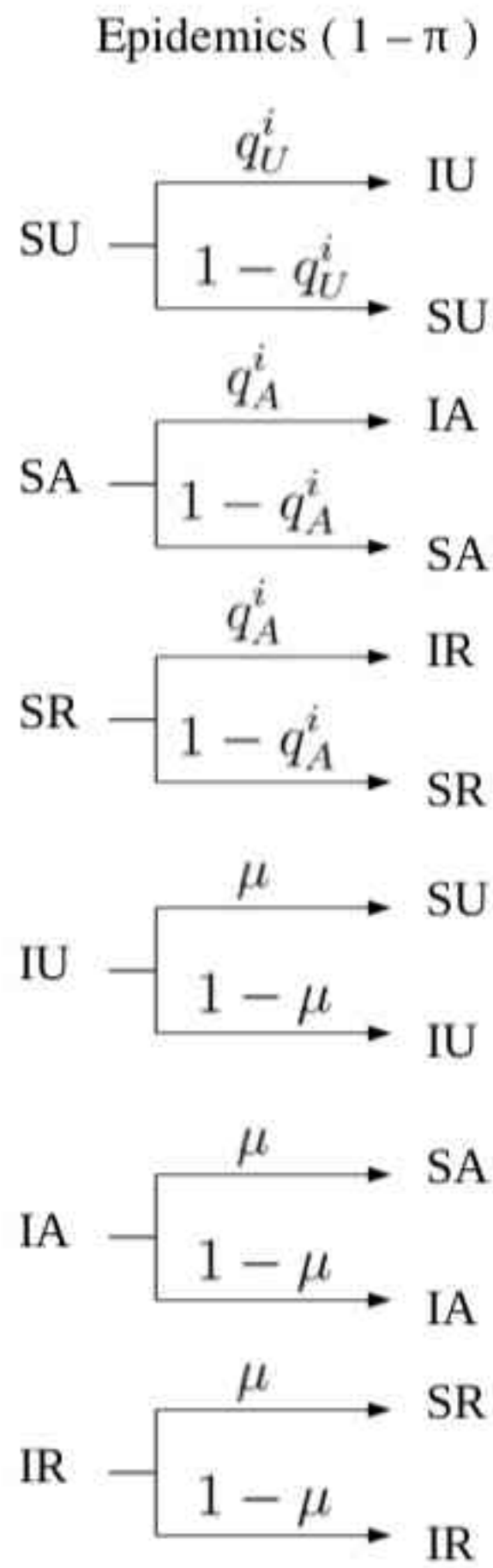
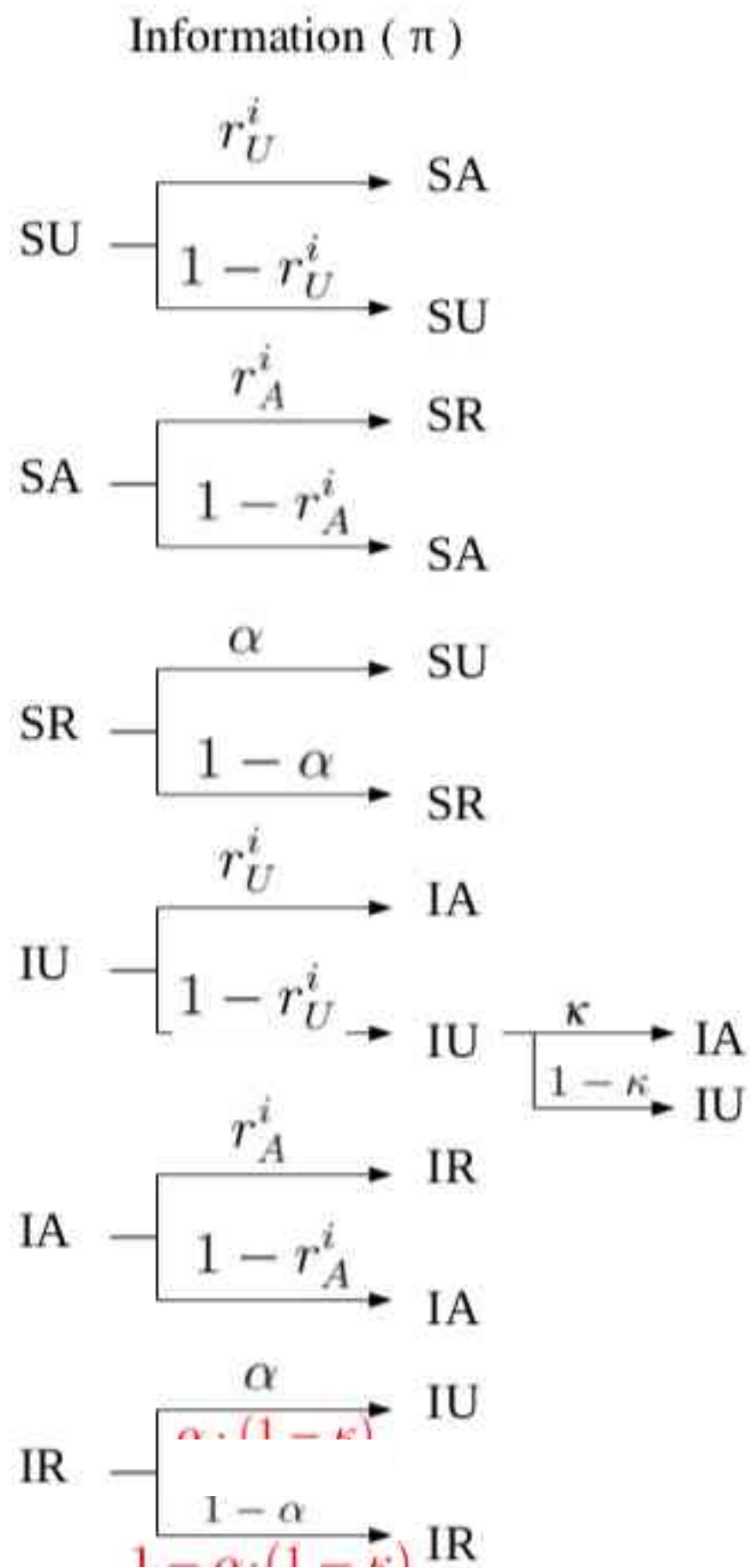
$$\begin{aligned}
p_{IA}^i &= p_{SA}^i[(1 - \pi)q_A^i] + \\
&\quad + p_{IU}^i[\pi(r_U^i + (1 - r_U^i)\kappa)] + \\
&\quad + p_{IA}^i[\pi(1 - r_A^i) + (1 - \pi)(1 - \mu)] \quad (A9)
\end{aligned}$$

$$\begin{aligned}
p_{IR}^i &= p_{SR}^i[(1 - \pi)q_A^i] + \\
&\quad + p_{IA}^i[\pi r_A^i] + \\
&\quad + p_{IR}^i[\pi(\alpha\kappa + 1 - \alpha) + (1 - \pi)(1 - \mu)] \quad (A10)
\end{aligned}$$

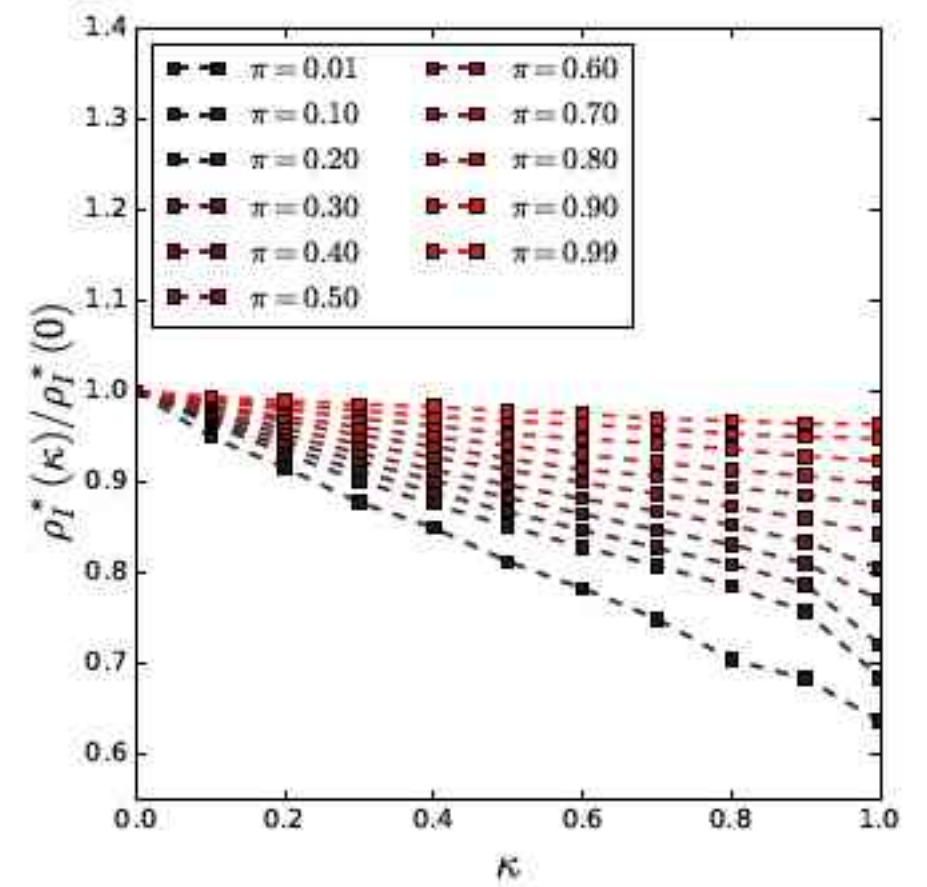
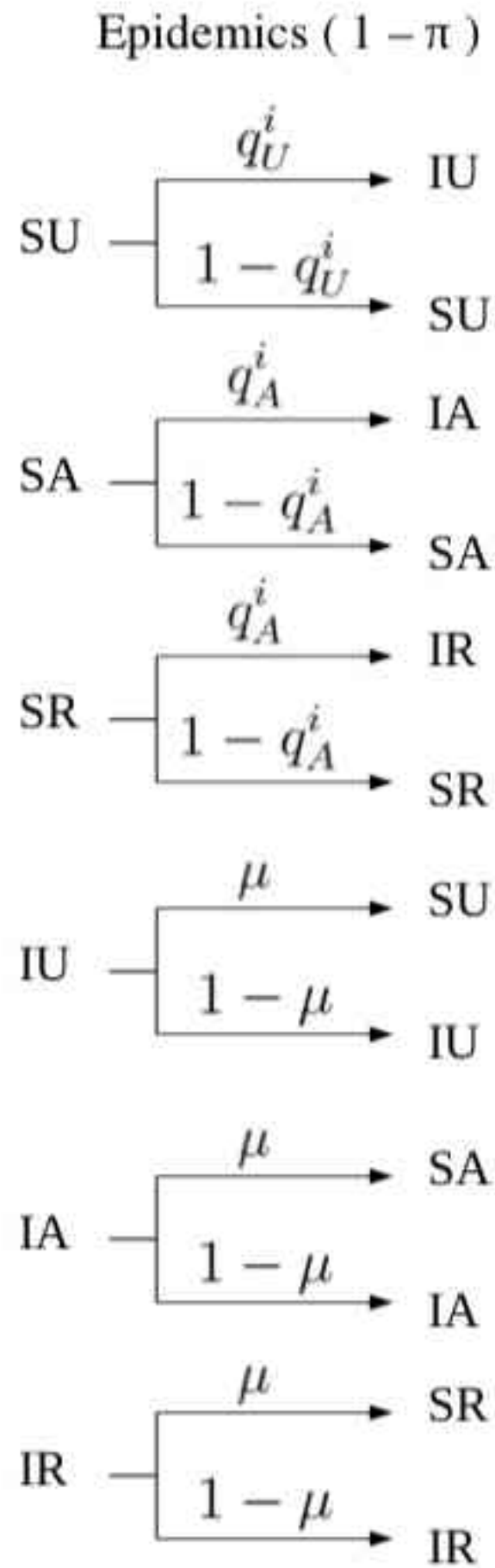
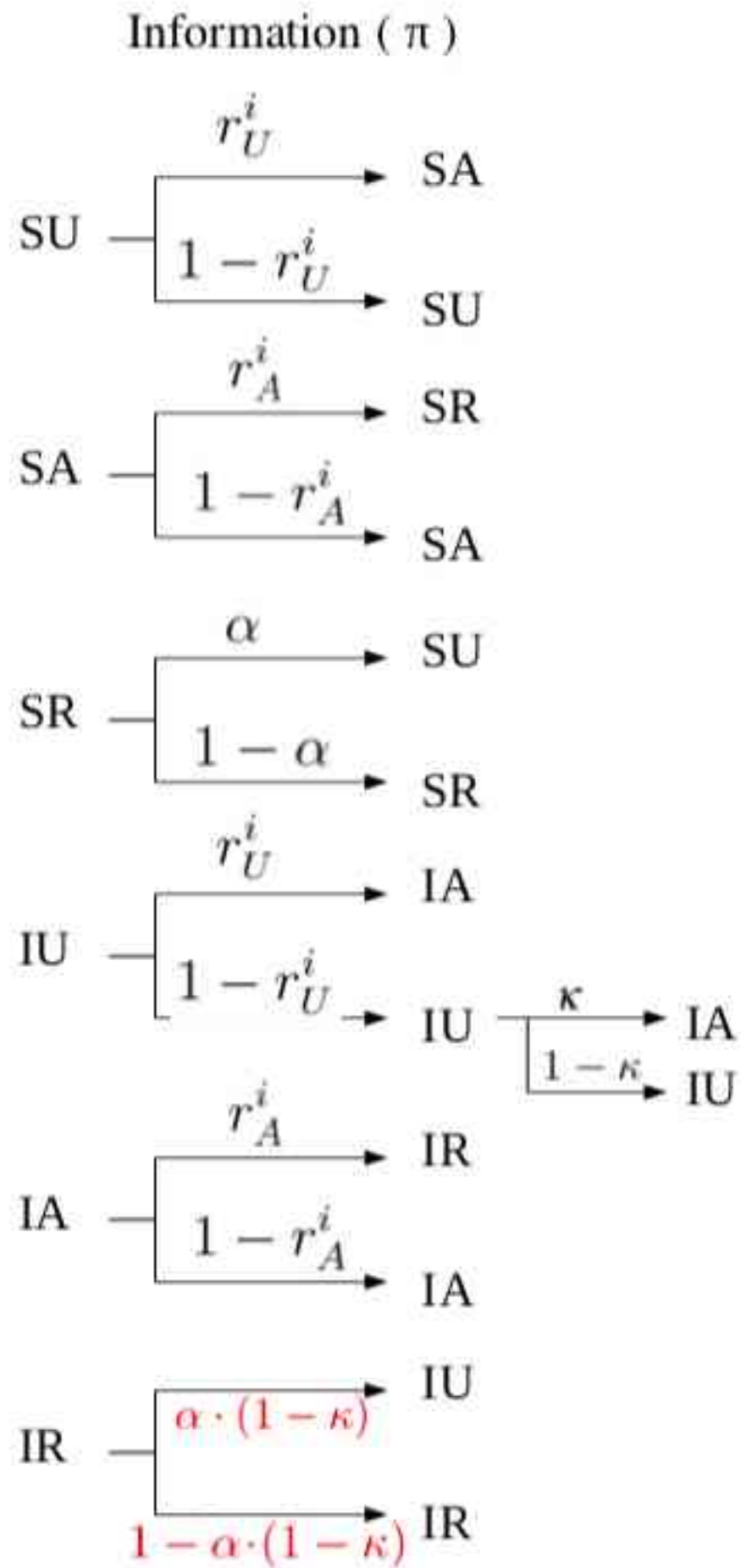
$$\begin{aligned}
p_{IU}^i &= p_{SU}^i[(1 - \pi)q_U^i] + \\
&\quad + p_{IU}^i[\pi(1 - r_U^i)(1 - \kappa) + (1 - \pi)(1 - \mu)] + \\
&\quad + p_{IR}^i[\pi\alpha] \quad (A11)
\end{aligned}$$

$$\begin{aligned}
p_{IR}^i &= p_{SR}^i[(1 - \pi)q_A^i] + \\
&\quad + p_{IA}^i[\pi r_A^i] + \\
&\quad + p_{IR}^i[\pi(1 - \alpha) + (1 - \pi)(1 - \mu)] \quad (A12)
\end{aligned}$$

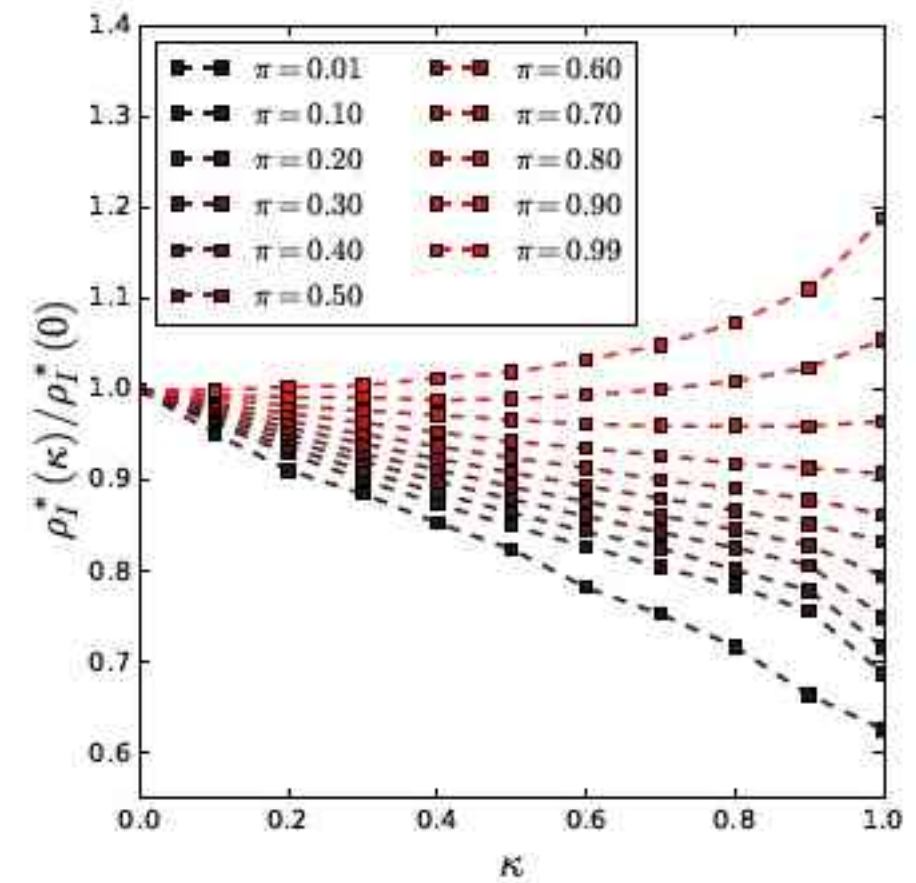




(a) Baseline model

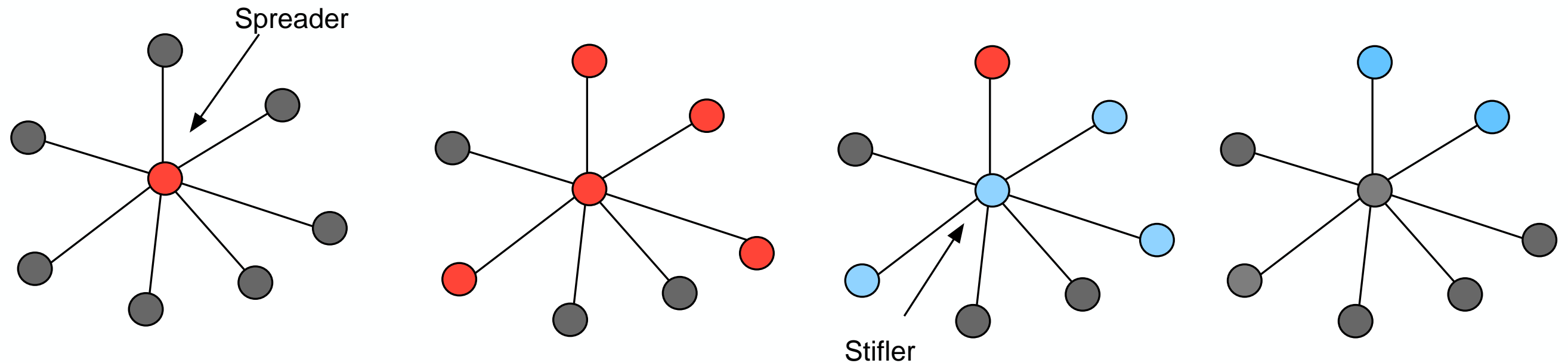


(a) Baseline model

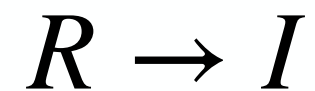
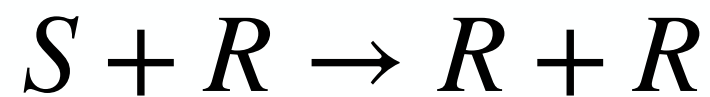
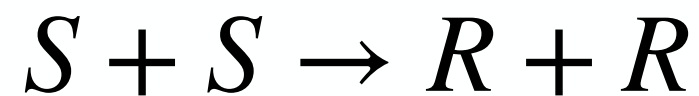
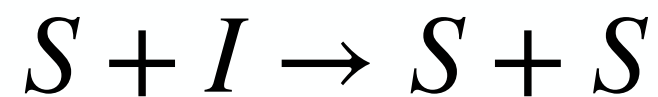


(b) Modified model

Epidemic spreading with awareness



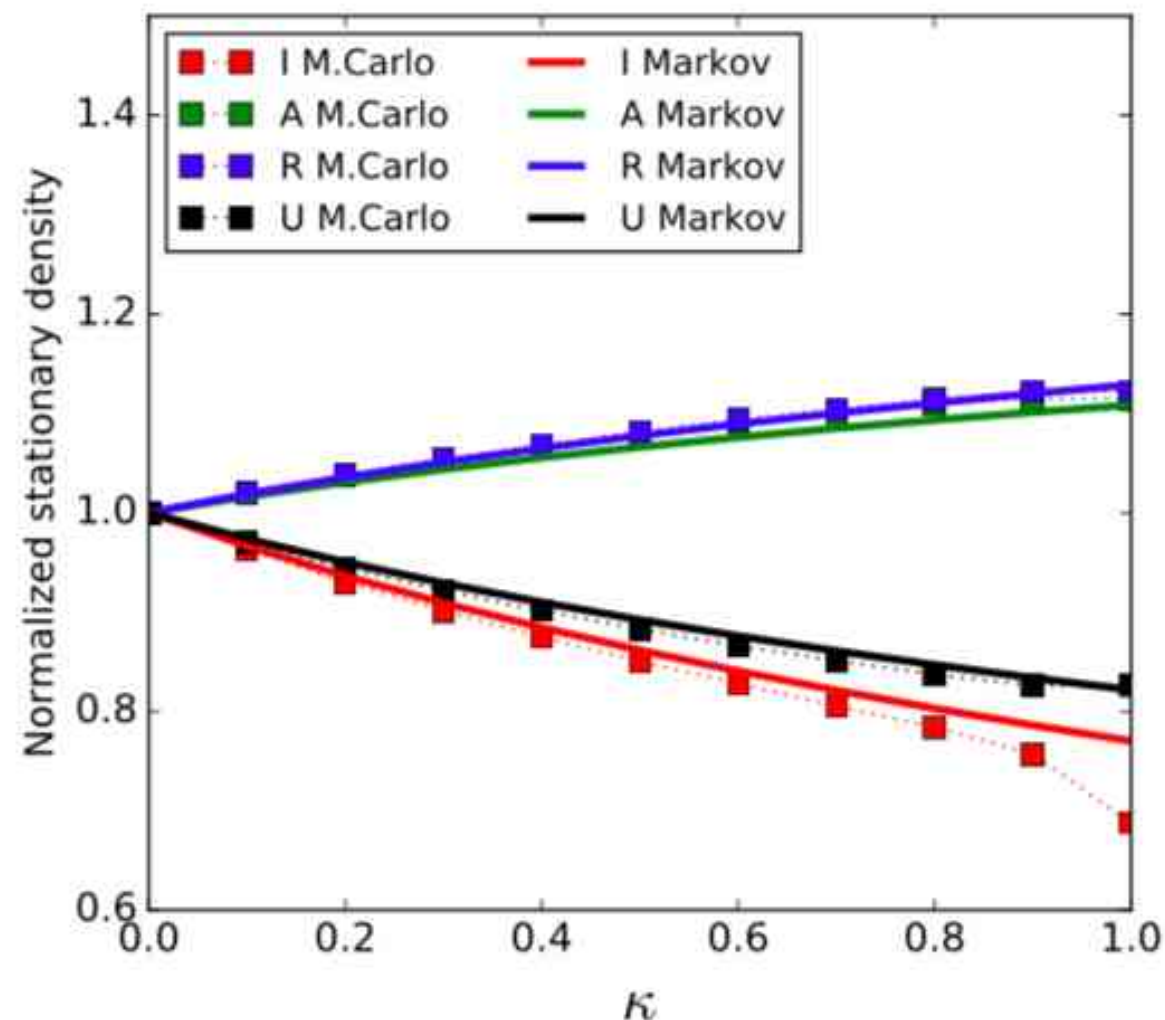
→ Time



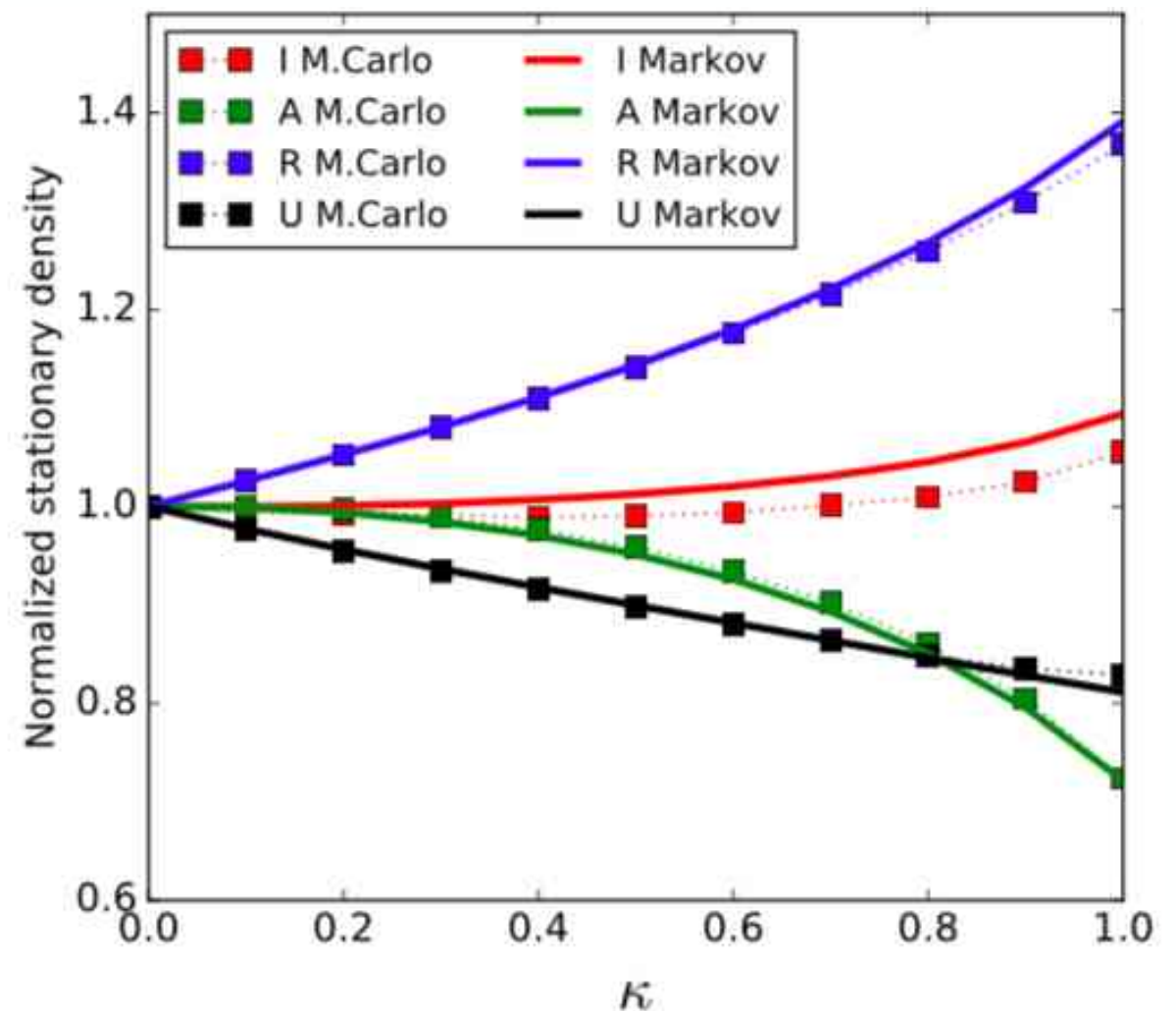
S: Spreader
I: Ignorant
R: Stifler

Epidemic spreading with awareness

$$\pi = 0.1$$



$$\pi = 0.9$$



- If the rumor propagation is too fast, then the outbreak increases!

Ventura et al. Physical Review E, 100, 032313 (2019).

Epidemic spreading with awareness

Applications

- Infectious diseases with no symptoms (Sexually transmitted diseases (STD)).



**Time scale plays a
fundamental role on the
epidemic spreading with
awareness.**

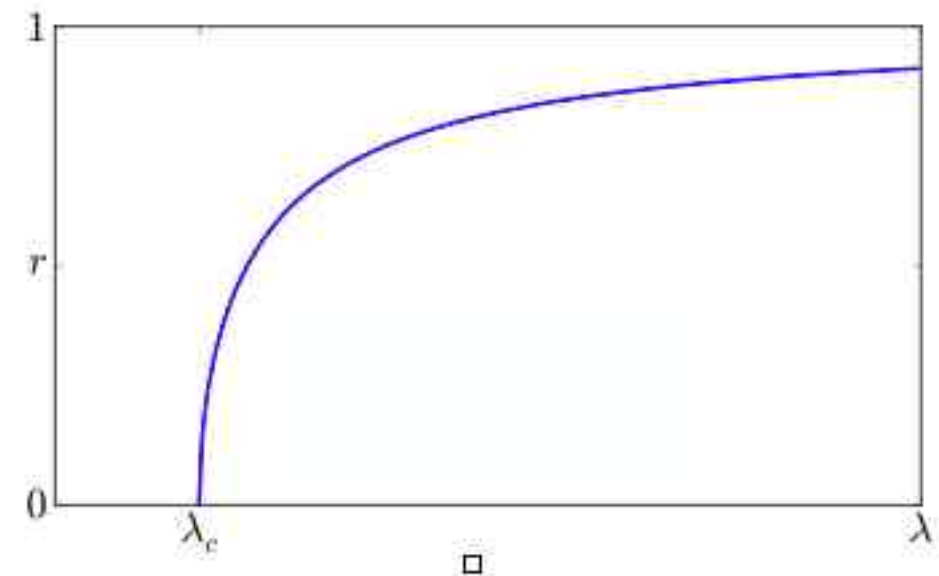
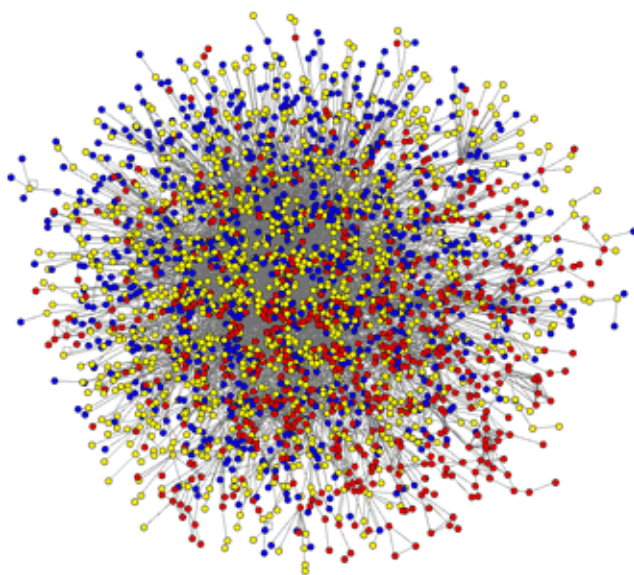


Ventura et al. Physical Review E, 100, 032313 (2019).

Predicting epidemic spreading

Epidemic spreading depends on the network structure.

Can we predict this dynamics process from the network structure?



Inference X Prediction

Inference: the goal is to understand the role of the input variables in explaining the outcome.

- To explain the data we have.

Prediction: the goal is to predict the outcome from the income data.

- Based on the data we have, predict new data.

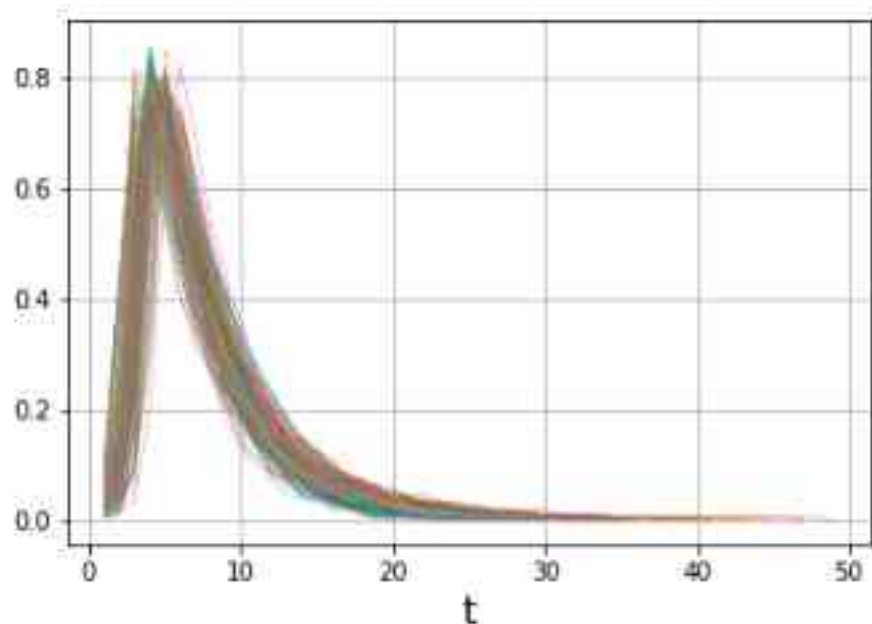
Inference and Prediction

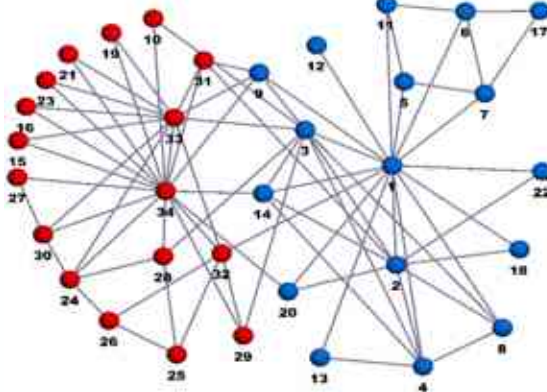
Global level: Consider the large-scale network properties to predict the outbreak size.

Local level: Consider the node properties to predict the average outbreak size when the disease starts at each node.

Structure X Dynamics: global level

Hypothesis:



$$= f(\text{Network}) + \varepsilon$$


$$Y = f(X) + \varepsilon$$

Regression model

$$f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$$

d : number of features

Structure X Dynamics: global level

Dataset:

Each observation: a network

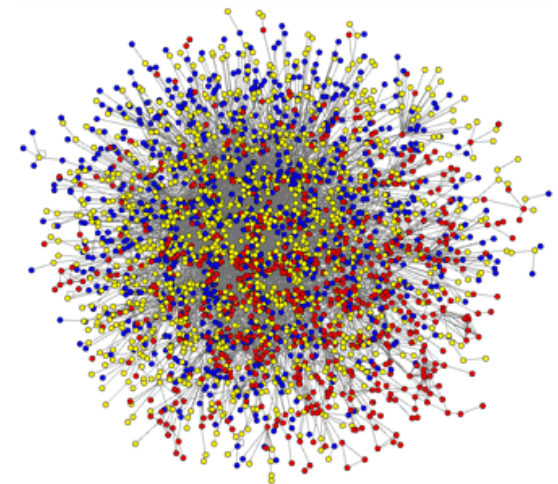
Attributes: Large scale network measures

Models

- ER
- BA
- Non-linear BA
- WS
- Waxman
- Spatial scale-free
- Assortative scale-free
- Dissortative scale-free



50 networks
for each model

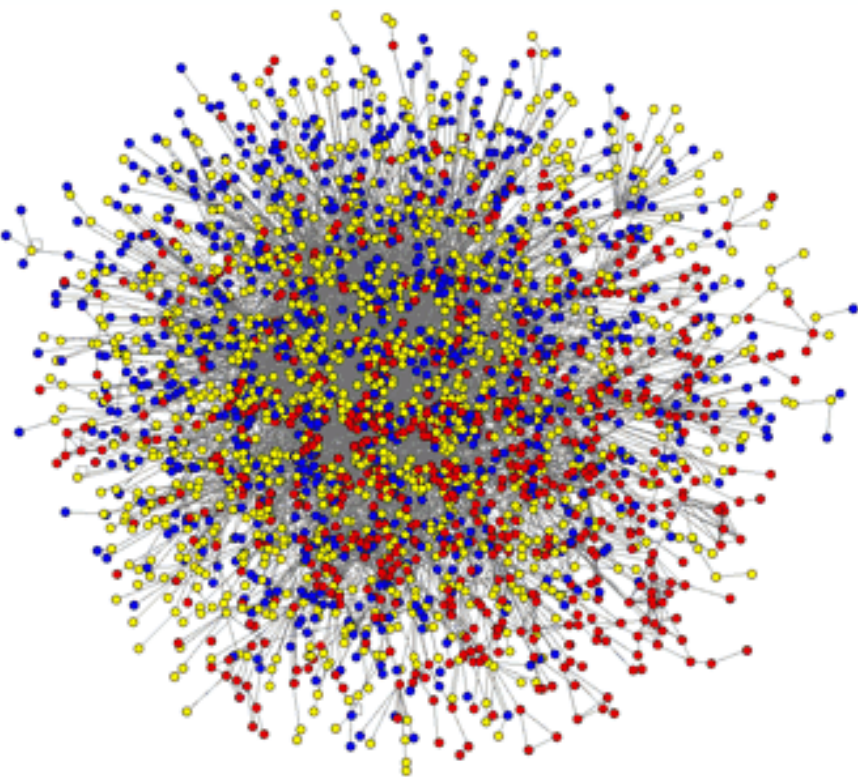


Structure X Dynamics: global level

Dataset:

Each observation: a network

Attributes: Large scale network measures



Measures:

- Average search information
- Number of articulation points
- Second moment of degree distribution
- Shannon entropy of the degree distribution
- Average efficiency
- Assortativity coefficient
- Average clustering coefficient
- Average coreness (hkci),
- Transitivity
- Variance of the betweenness centrality

Structure X Dynamics: global level

Beta regression

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}$$

$$E(y) = \mu$$

$$\text{var}(y) = \frac{\mu(1-\mu)}{1+\phi}.$$

$$g_1(\mu) = \beta_0 + \beta_1 S + \beta_2 A + \beta_3 \langle k^2 \rangle + \beta_4 E + \beta_5 H \\ + \beta_6 r + \beta_7 \langle cc \rangle + \beta_8 \langle kc \rangle + \beta_9 C,$$

$$g_2(\phi) = \gamma_0 + \gamma_1 V(B) + \gamma_2 \langle cc \rangle + \gamma_3 H + \gamma_4 V(B) \langle cc \rangle \\ + \gamma_5 V(B) H + \gamma_6 \langle cc \rangle H + \gamma_7 V(B) \langle cc \rangle H.$$

Structure X Dynamics: global level

Beta regression

Epidemic spreading

	Mean	Std. error	p-value	Measure
β_0	1.913	0.001	≈ 0	—
β_1	-0.200	0.008	≈ 0	S
β_2	-0.123	0.004	≈ 0	A
β_3	-0.097	0.011	≈ 0	$\langle k^2 \rangle$
β_4	0.108	0.013	≈ 0	E
β_5	0.063	0.008	≈ 0	H
β_6	-0.134	0.002	≈ 0	r
β_7	0.002	0.012	0.843	$\langle cc \rangle$
β_8	0.066	0.008	≈ 0	$\langle kc \rangle$
β_9	0.119	0.012	≈ 0	C
γ_0	9.611	0.151	≈ 0	—
γ_1	0.065	0.158	0.681	$V(B)$
γ_2	-0.562	0.304	0.064	$\langle cc \rangle$
γ_3	-0.388	0.201	0.053	H
γ_4	-0.483	0.346	0.163	$V(B) : \langle cc \rangle$
γ_5	-0.191	0.333	0.566	$V(B) : H$
γ_6	-0.01	0.263	0.968	$\langle cc \rangle : H$
γ_7	-0.223	0.183	0.224	$V(B) : \langle cc \rangle : H$

Rumour spreading

	Mean	Std. error	p-value	Measure
β_0	1.43	0.001	≈ 0	—
β_1	-0.290	0.010	≈ 0	S
β_2	-0.142	0.005	≈ 0	A
β_3	0.196	0.017	≈ 0	$\langle k^2 \rangle$
β_4	-0.231	0.018	≈ 0	E
β_5	0.043	0.008	≈ 0	H
β_6	0.000	0.002	0.813	r
β_7	-0.188	0.020	≈ 0	$\langle cc \rangle$
β_8	0.127	0.008	≈ 0	$\langle kc \rangle$
β_9	0.072	0.017	≈ 0	C
γ_0	8.828	0.151	≈ 0	—
γ_1	-1.863	0.158	≈ 0	$V(B)$
γ_2	-1.007	0.303	0.001	$\langle cc \rangle$
γ_3	-1.194	0.201	≈ 0	H
γ_4	0.134	0.346	0.698	$V(B) : \langle cc \rangle$
γ_5	-0.875	0.333	0.009	$V(B) : H$
γ_6	0.338	0.263	0.198	$\langle cc \rangle : H$
γ_7	-0.021	0.183	0.911	$V(B) : \langle cc \rangle : H$

The most influential topological measure in both dynamics is the average search information (S).

Structure X Dynamics: global level

Beta regression

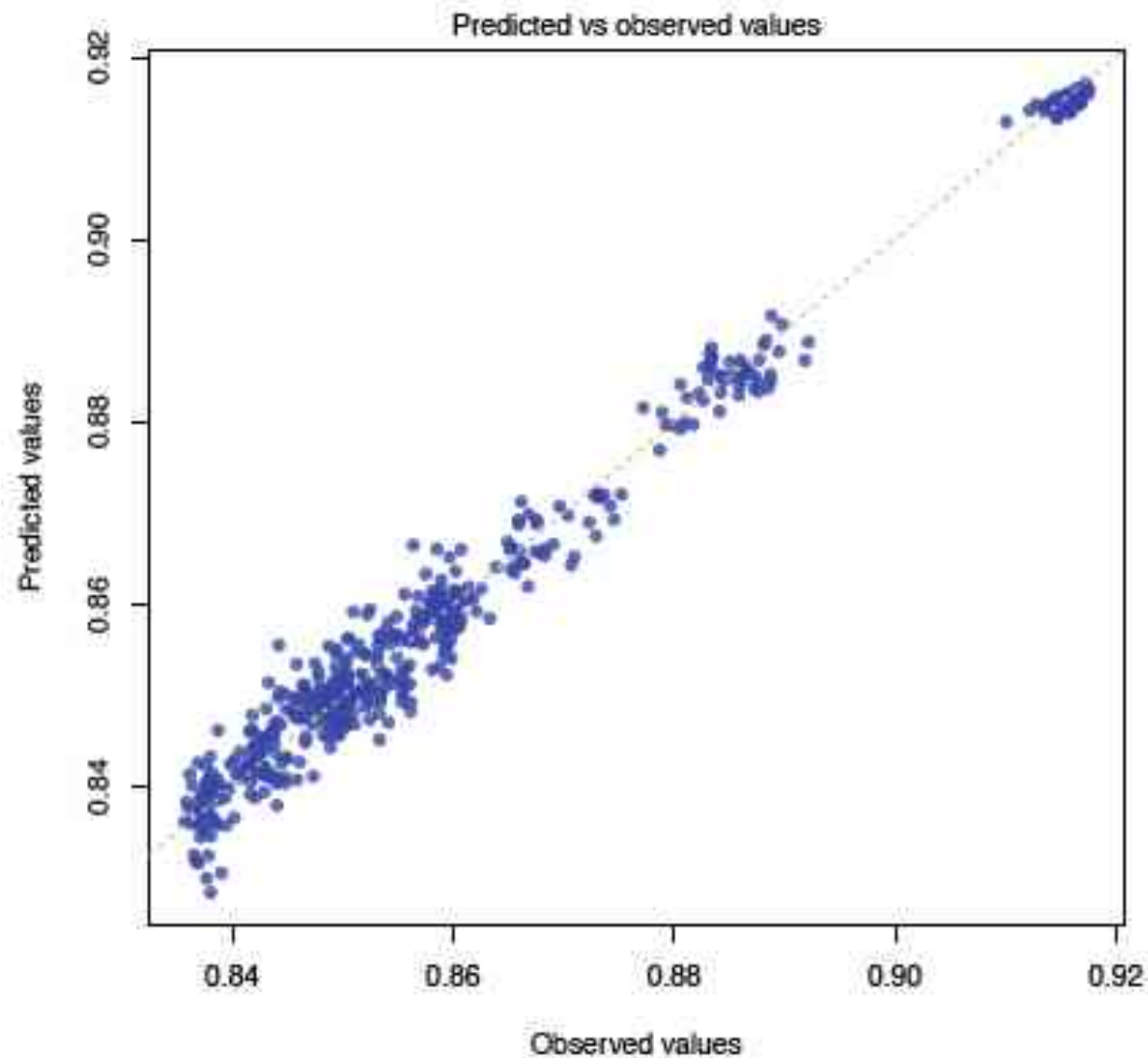


FIG. 3: Predicted vs Observed values for the SIR model.

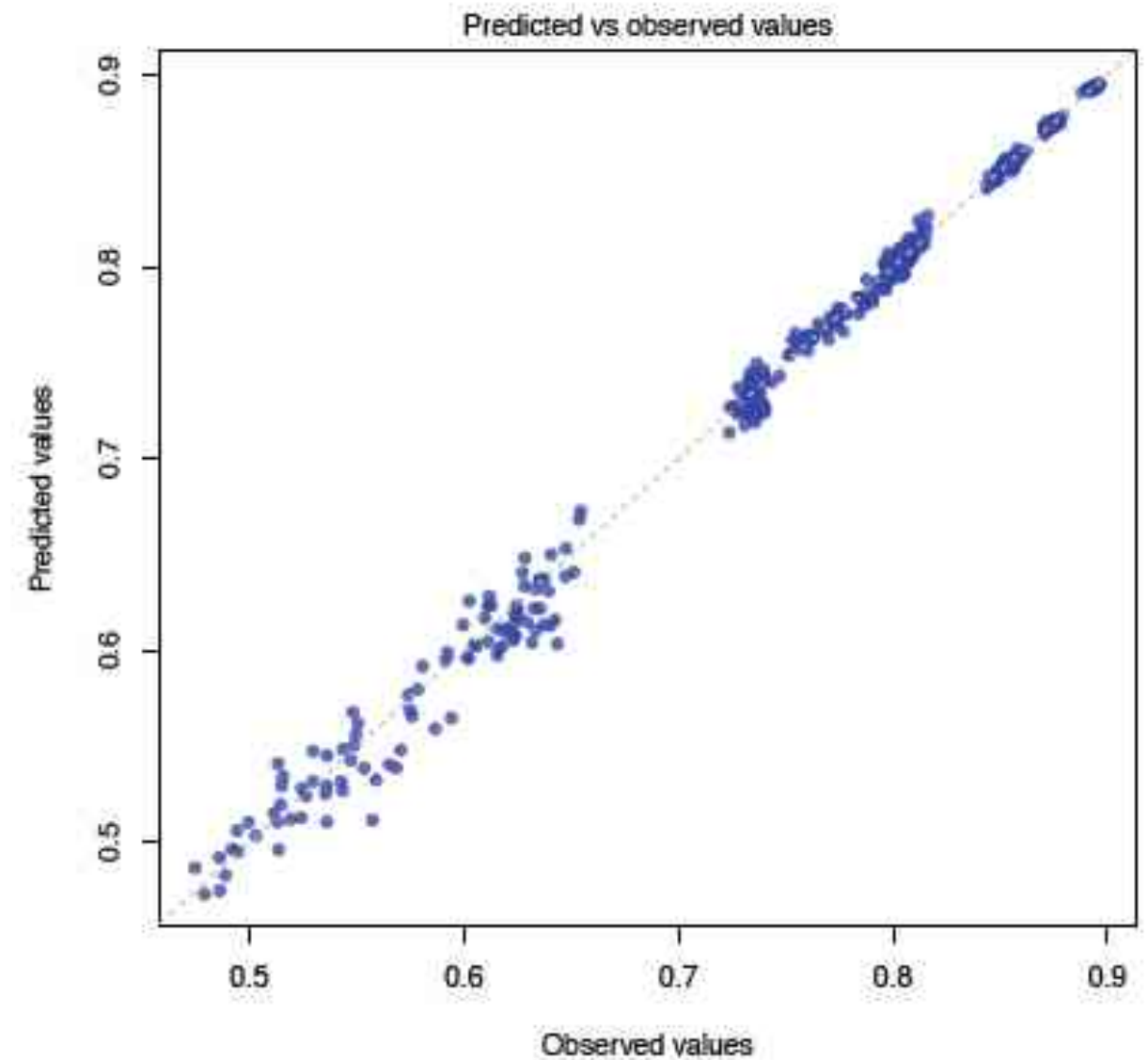
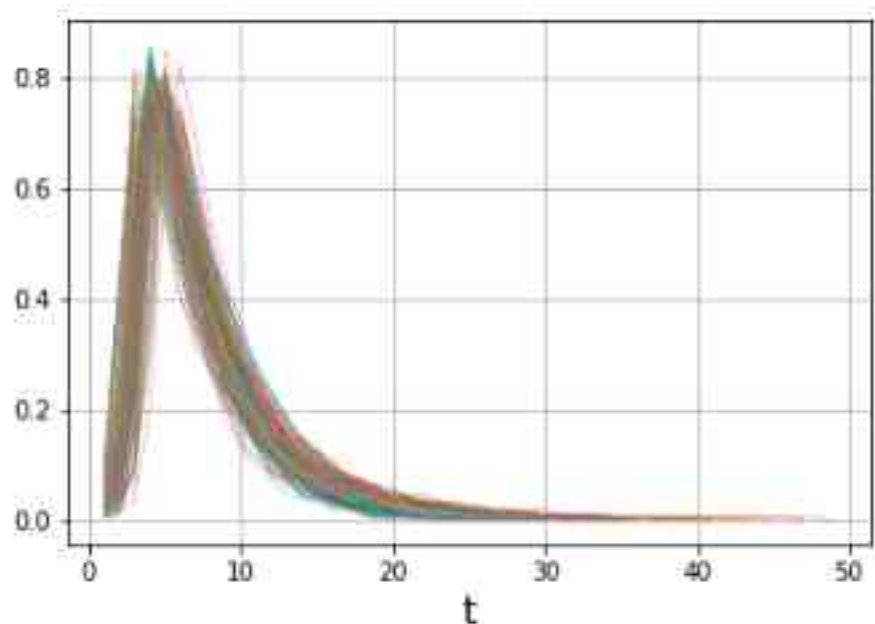


FIG. 4: Predicted vs Observed values for the ISR model.

Structure X Dynamics: Local level

Hypothesis:



$$= f(\text{graph icon}) + \varepsilon$$

$$Y_i = f(X_i) + \varepsilon$$

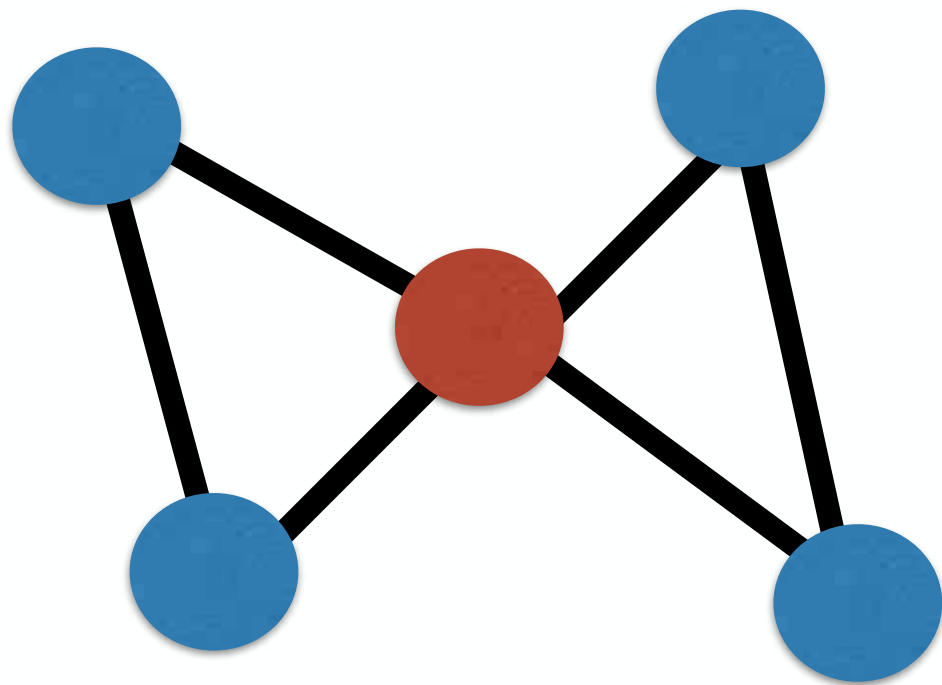
Regression model

$$f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$$

d : number of features

Structure X Dynamics: Prediction

Local level: predict the average outbreak size of a disease starting in a single node from the node properties.



- Nodes (observations) are not independent, then traditional regression models cannot be considered.

Structure X Dynamics: global level

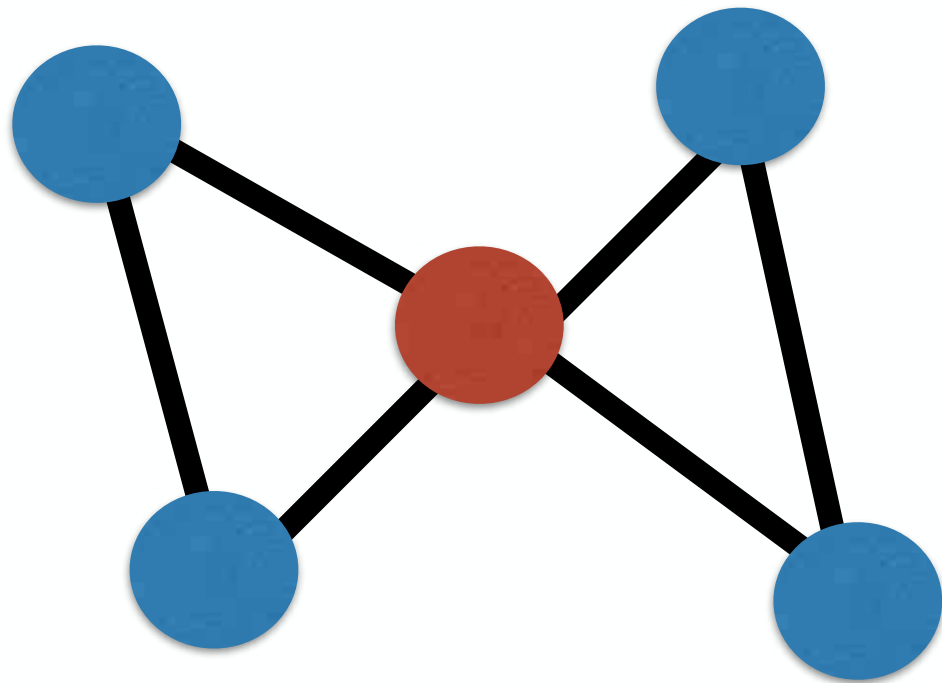
$$Y = f(X) + \varepsilon$$

$$f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$$

- The function f is very complicated due to the presence of non-trivial patterns of connections, nonlinear effects and correlations between variables...

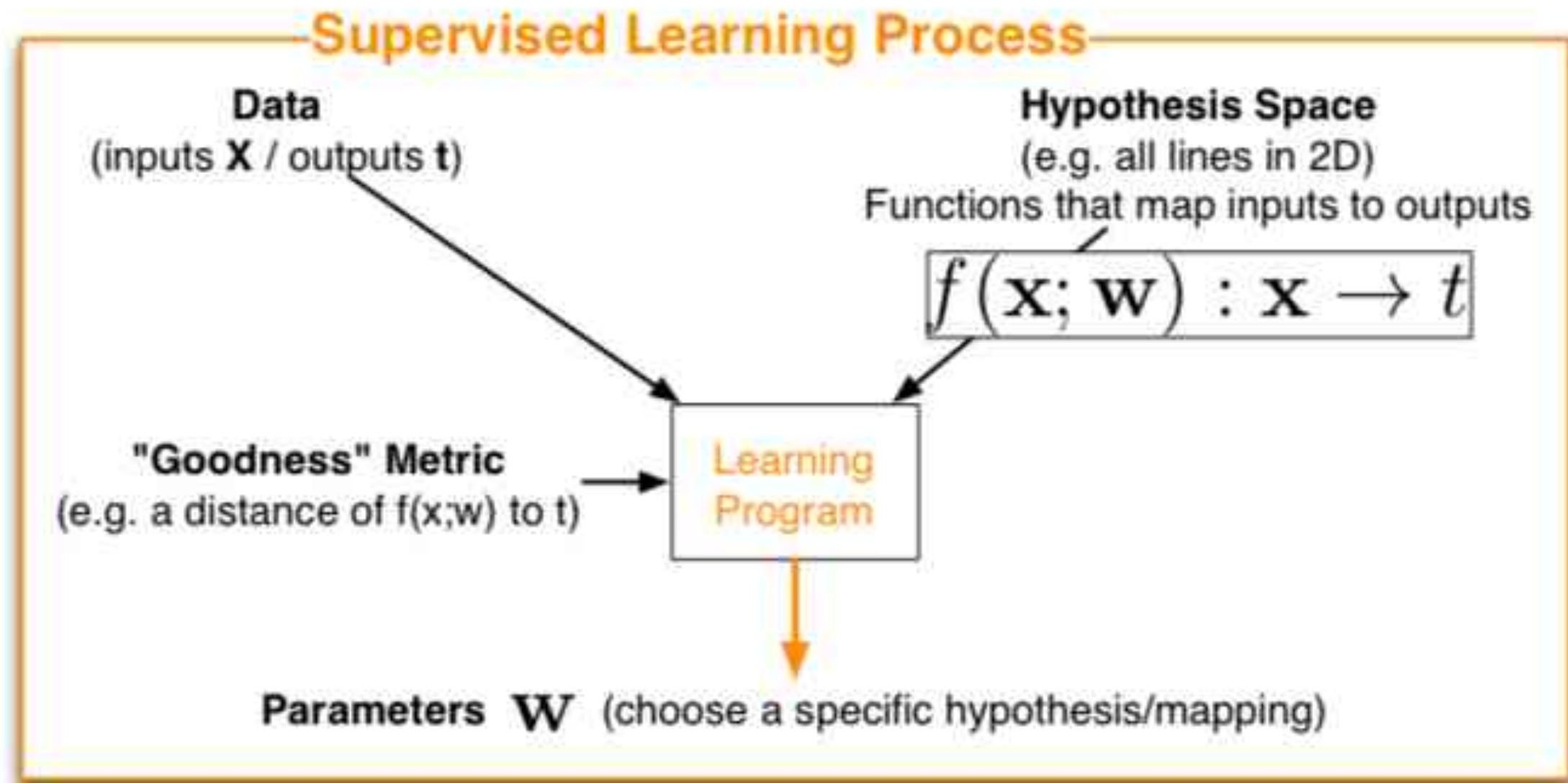
Structure X Dynamics: Prediction

$$Y_i = f(X_i) + \epsilon_i$$

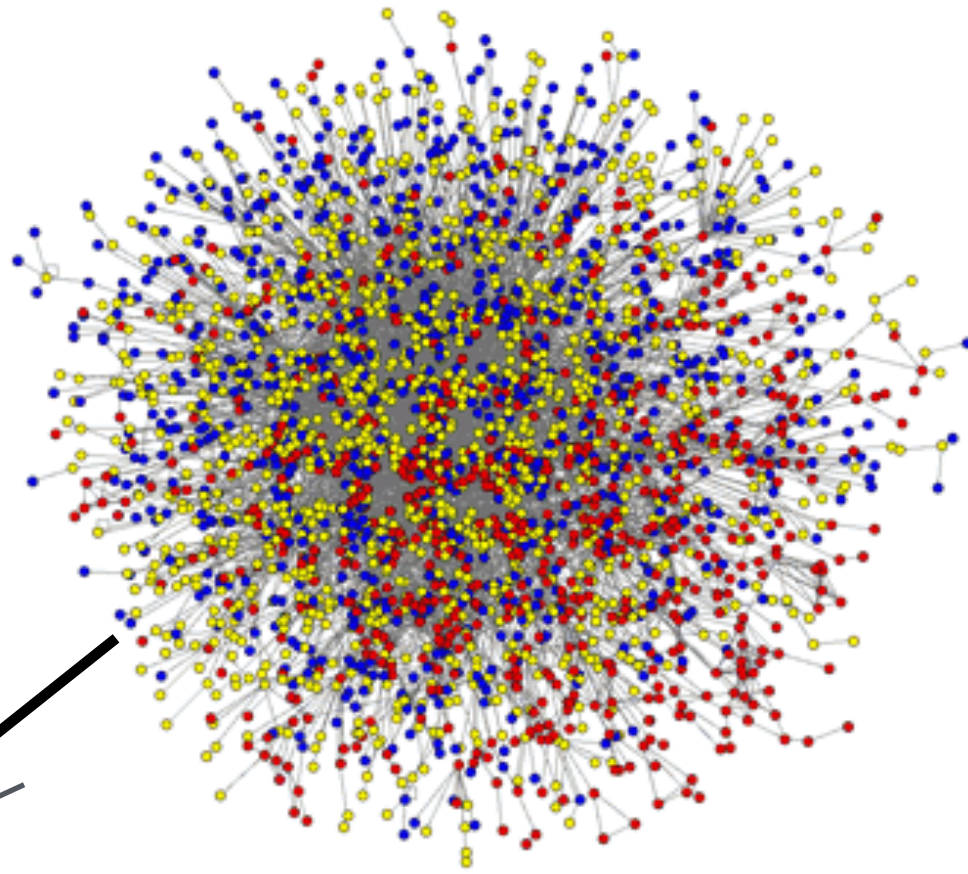


Solution:
Machine Learning

Supervised (inductive) Learning



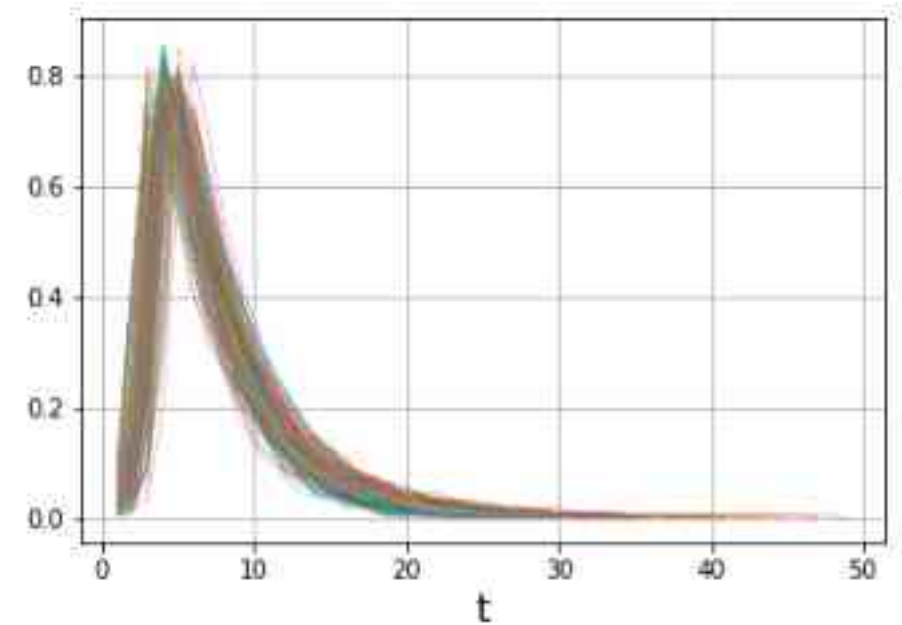
Data



node i

X

$$\begin{bmatrix} k(i), cc(i), B(i), PR(i), kc(i), ec(i) \\ \dots \\ k(j), cc(i), B(i), PR(i), kc(i), ec(i) \\ \dots \end{bmatrix}$$



Y

$$\begin{bmatrix} y_i \\ \dots \\ y_j \\ \dots \end{bmatrix}$$

Epidemic spreading

- We define Y_i as the expected fraction of infected nodes when the disease starts in i

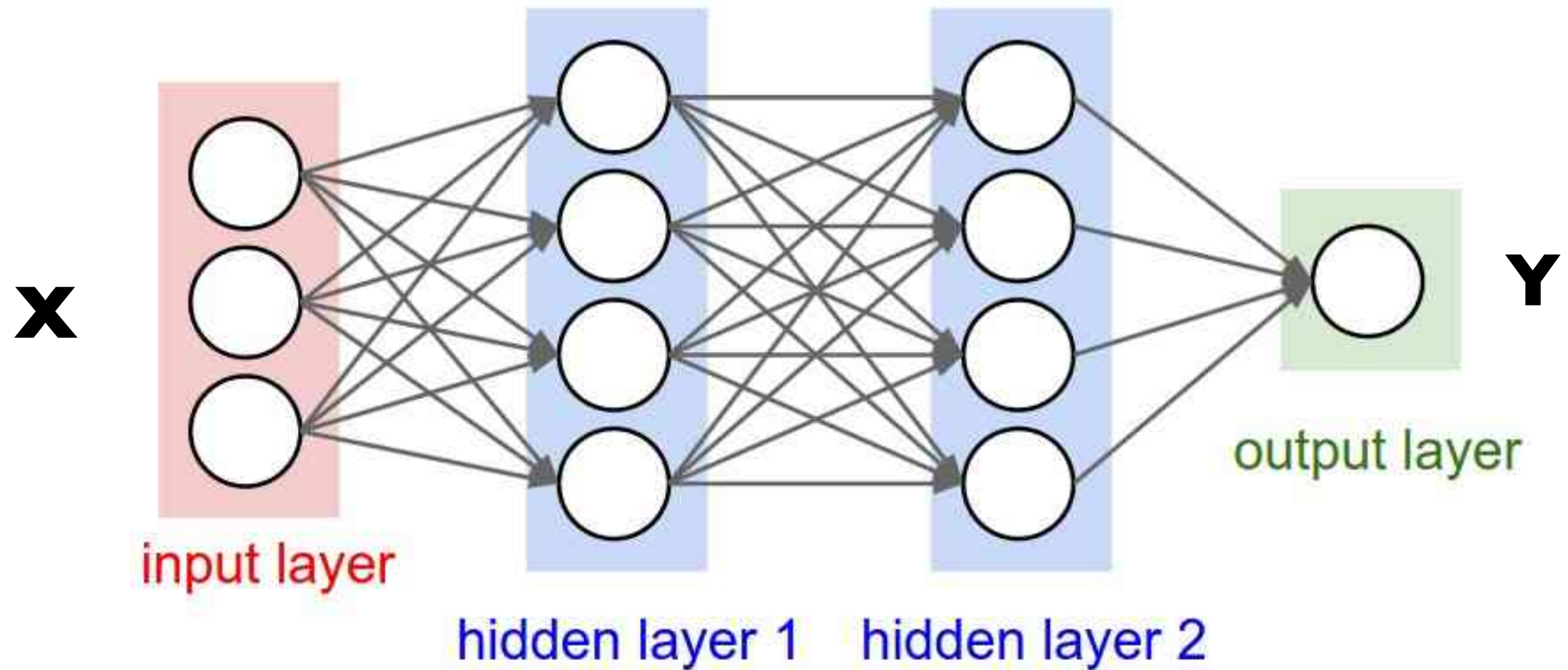
$$Y_i = \lim_{t \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N Z_i(t),$$

$$Y_i = f(X_i) + \varepsilon$$

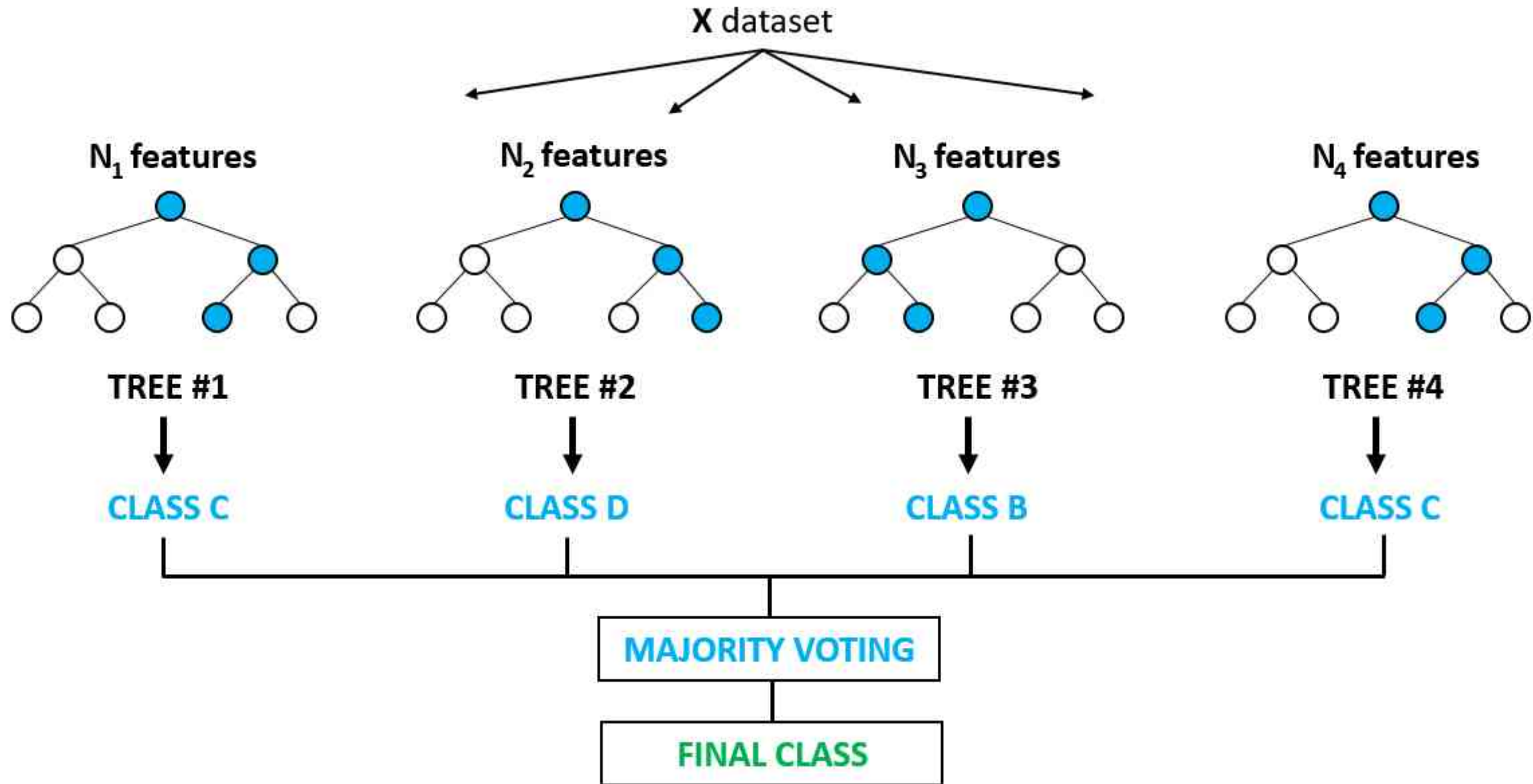
Machine learning:

- To obtain the function $f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$
 - Random forests
 - Neural Networks

Artificial neural networks



Random forests

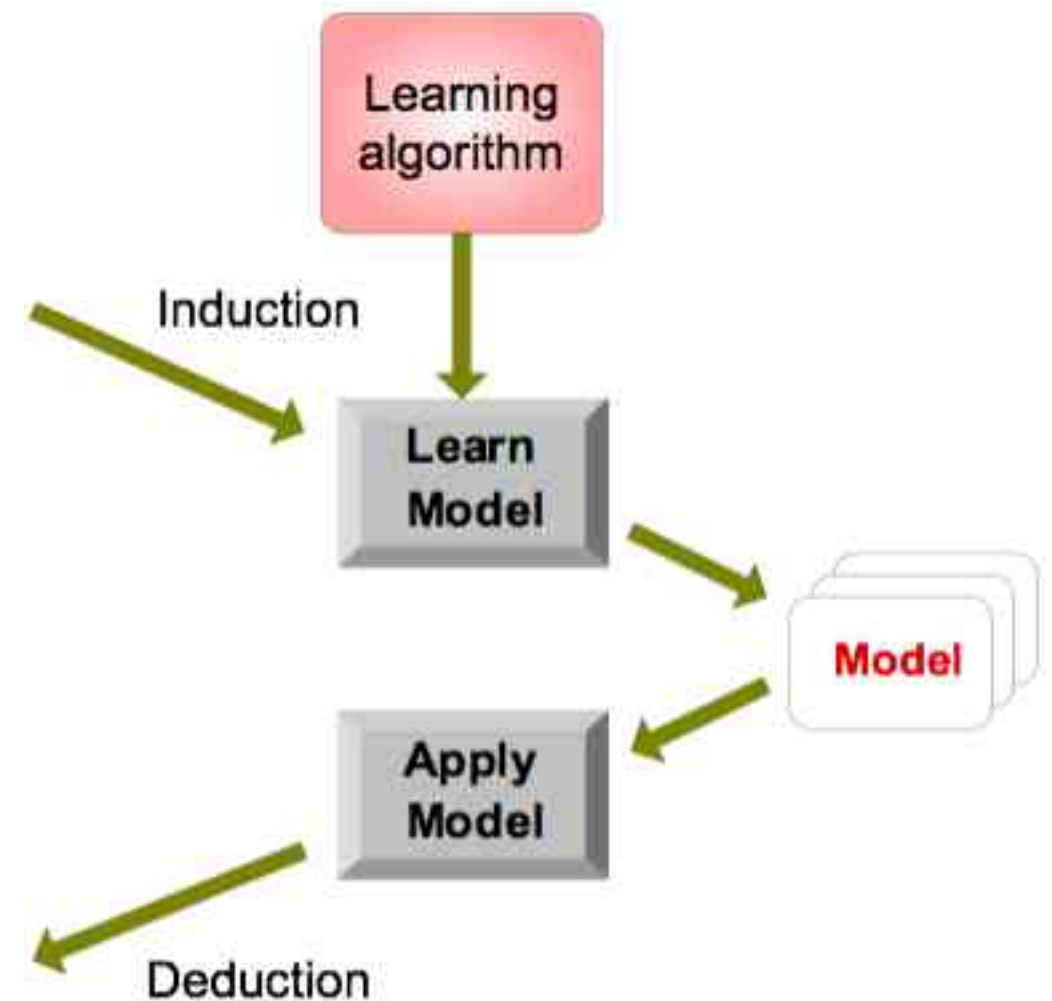


Predictive learning

Training set

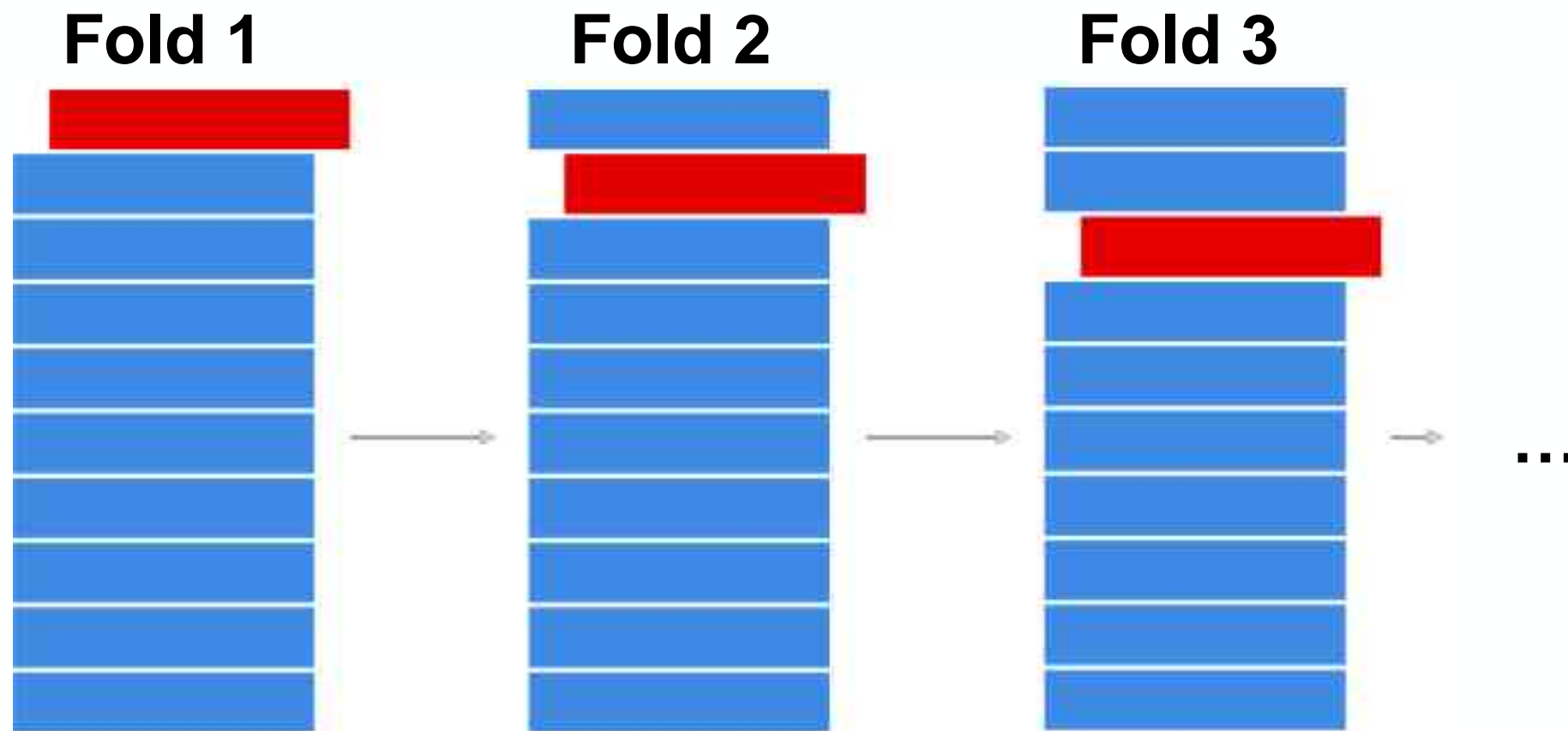
$$\begin{bmatrix} k(1), cc(1), \dots, kc(1), ec(1) \\ \dots \\ k(l), cc(l), \dots, kc(l), ec(l) \end{bmatrix}$$

Testing set

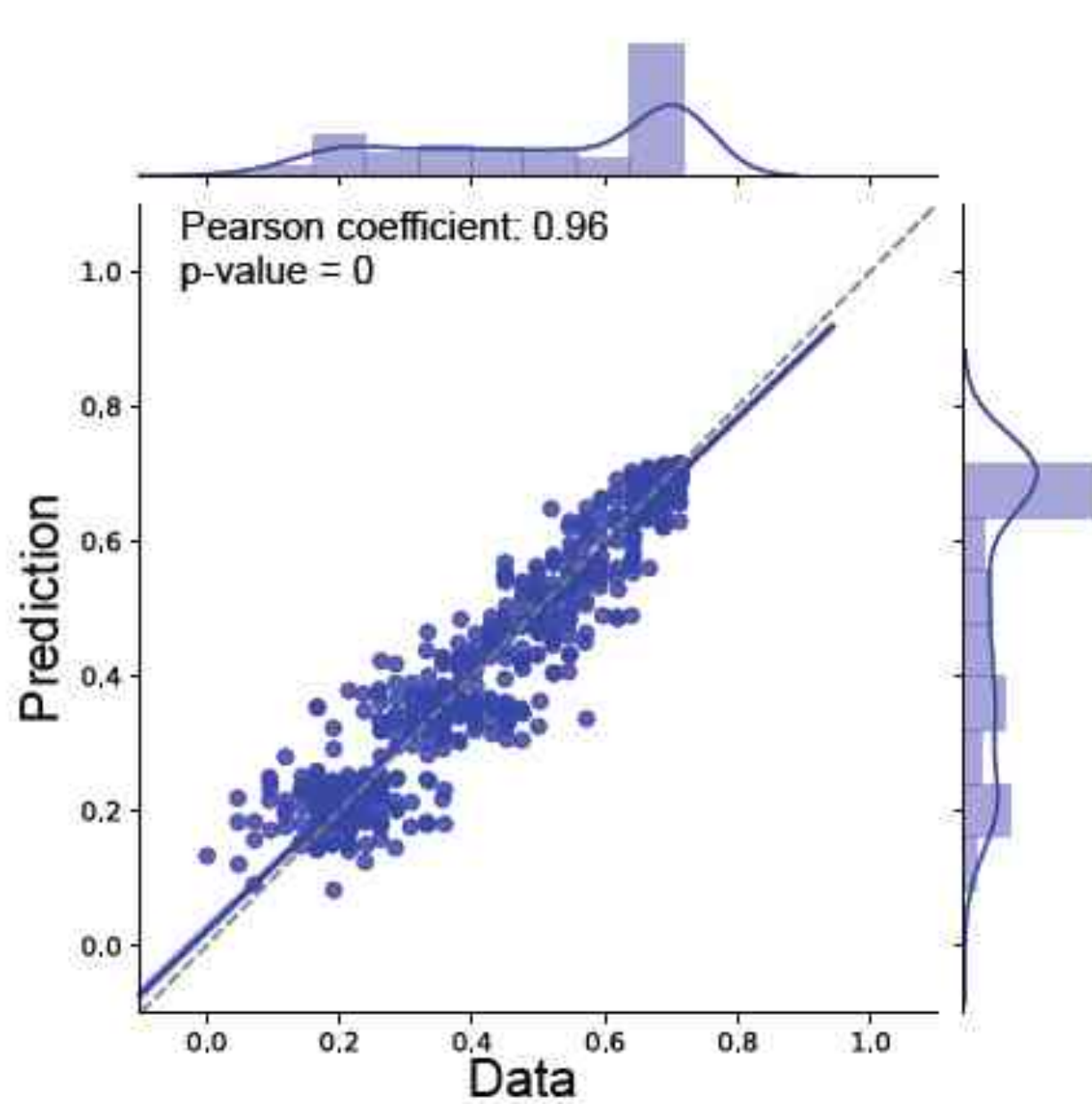
$$\begin{bmatrix} k(l+1), cc(l+1), \dots, kc(l+1), ec(l+1) \\ \dots \\ k(N), cc(N), \dots, kc(N), ec(N) \end{bmatrix}$$


K-fold cross validation

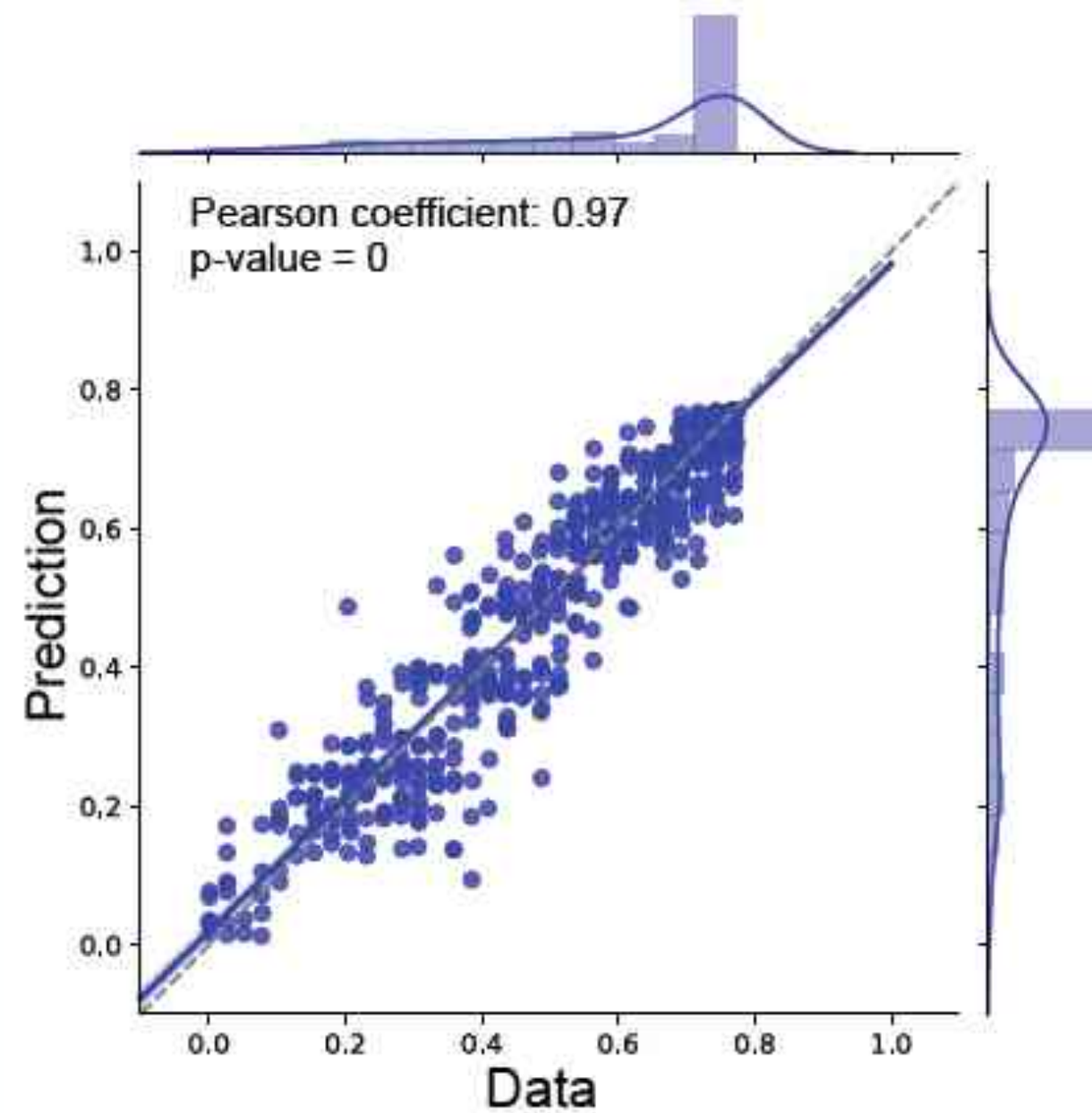
Training data 
Testing data 



Prediction

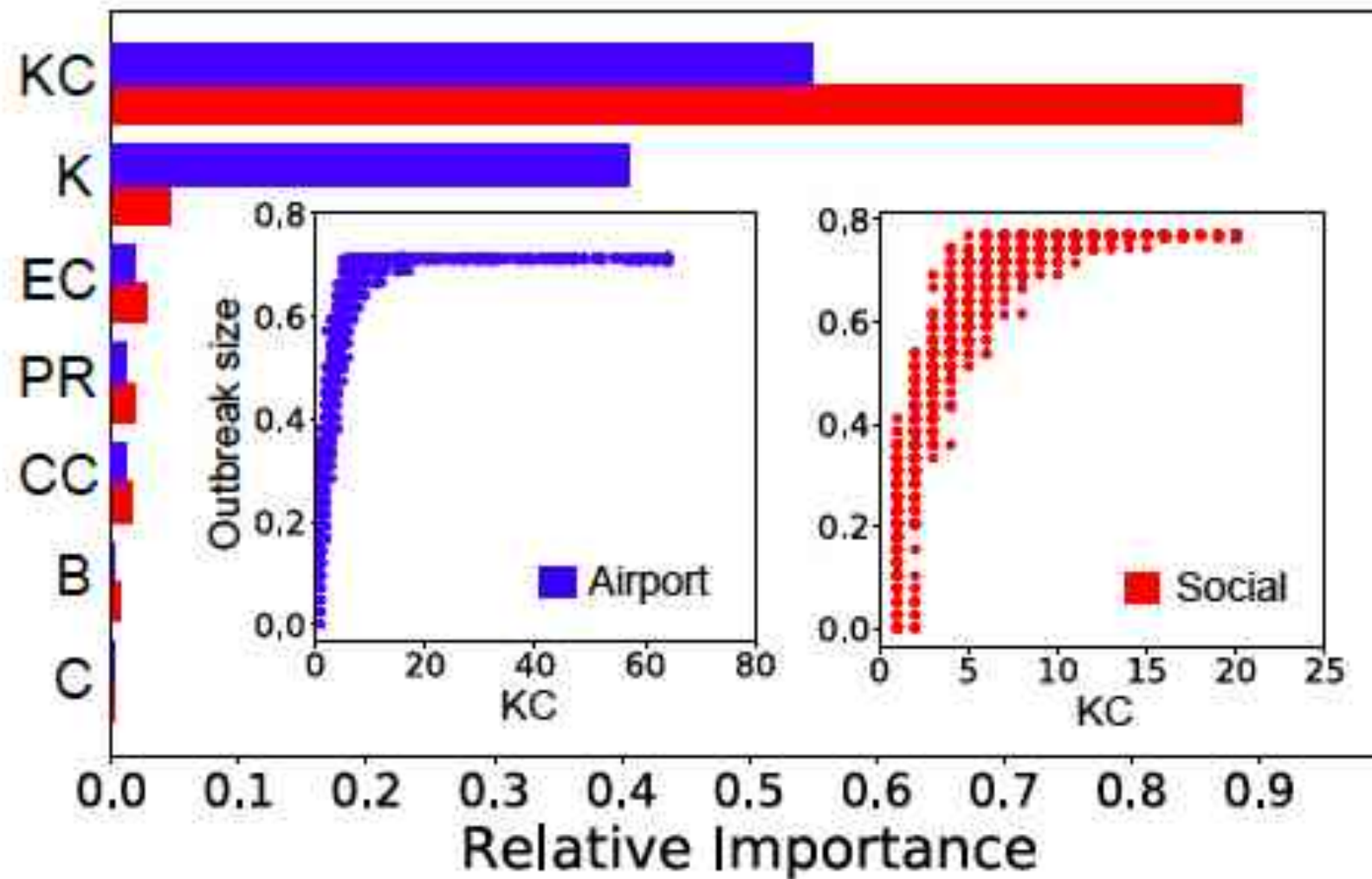


US air transportation
network



Hamsterster social
network

Features' importance



Our approach solve the problem of the identification of most influential spreaders!



nature
physics

Letter | Published: 29 August 2010

Identification of influential spreaders in complex networks

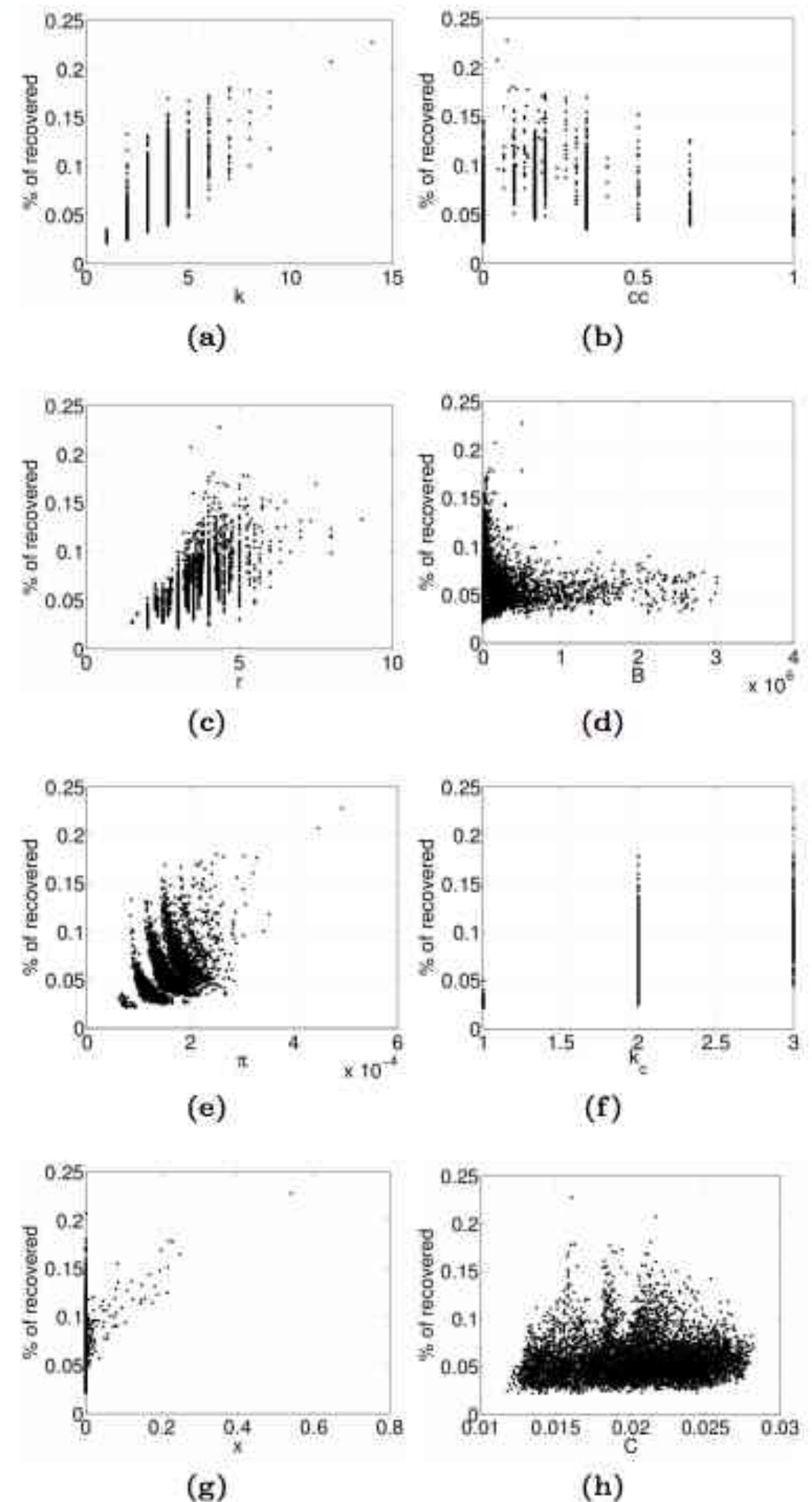
Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley & Hernán A. Makse

nature
International journal of science

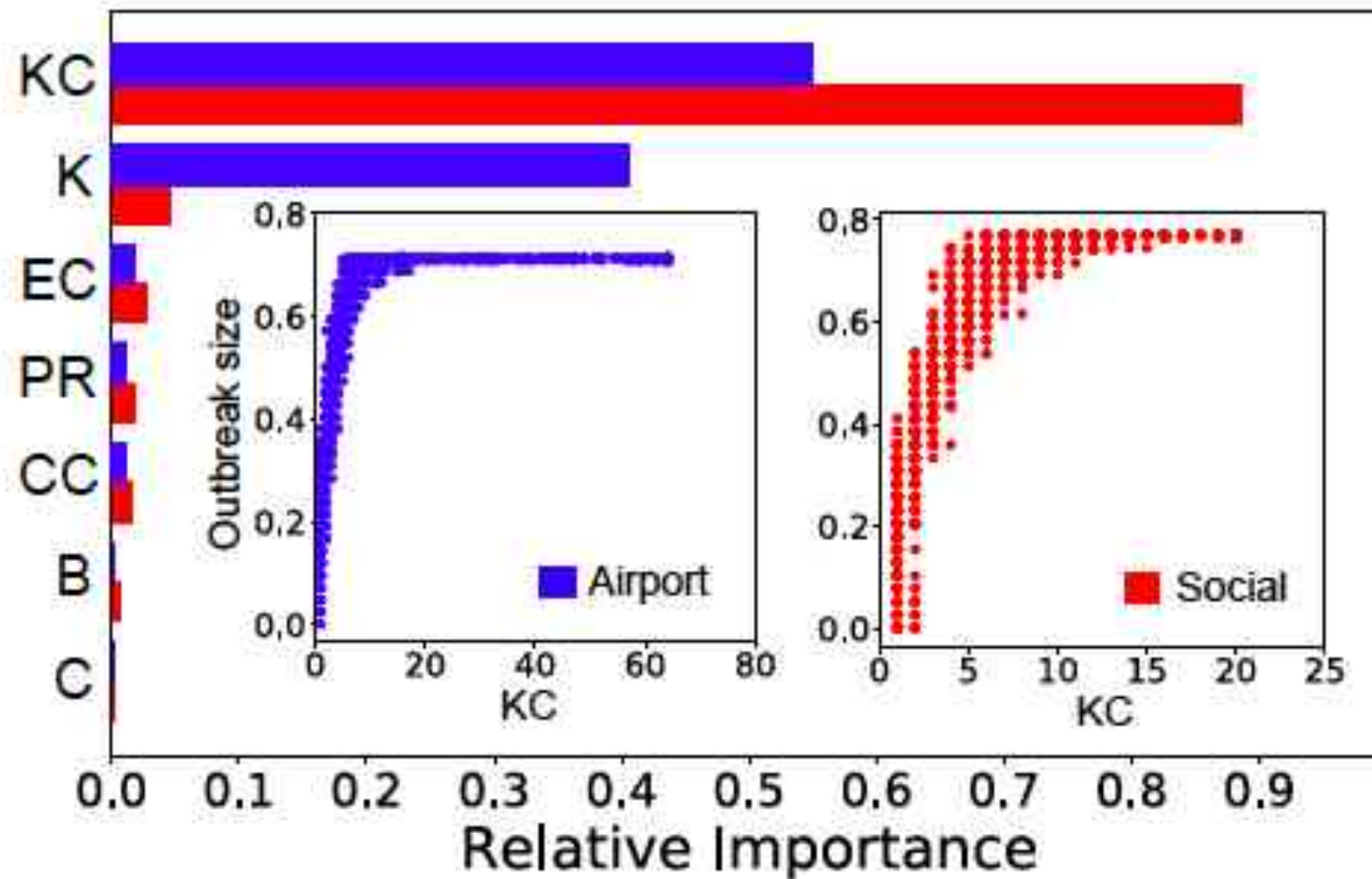
Letter | Published: 01 July 2015

Influence maximization in complex networks through optimal percolation

Flaviano Morone & Hernán A. Makse



Features' importance

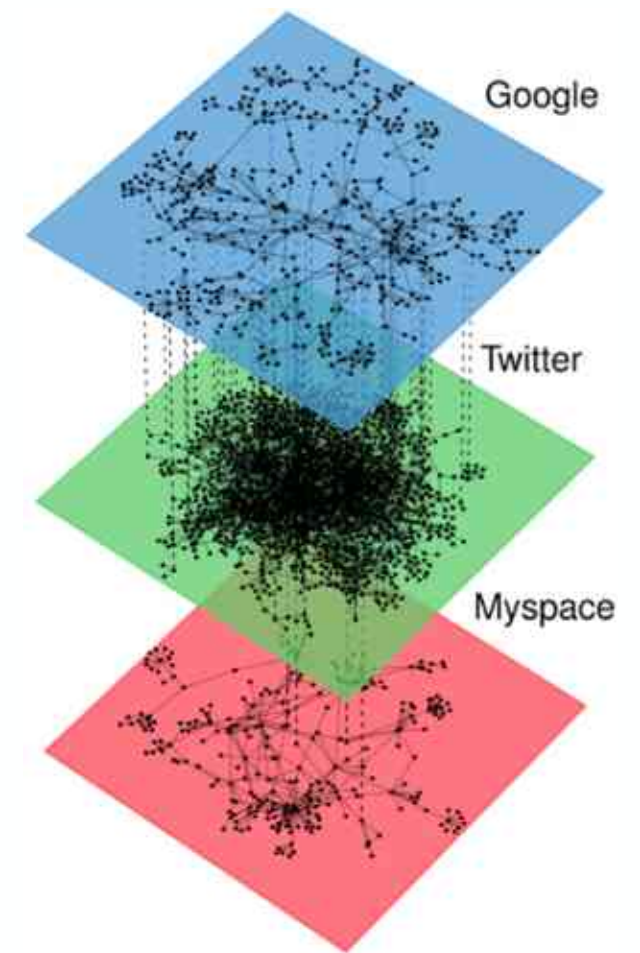


Our approach solve the problem of the identification of most influential spreaders!

Statistics and machine learning are very useful tools to predict dynamical processes from the network structure.

Challenges

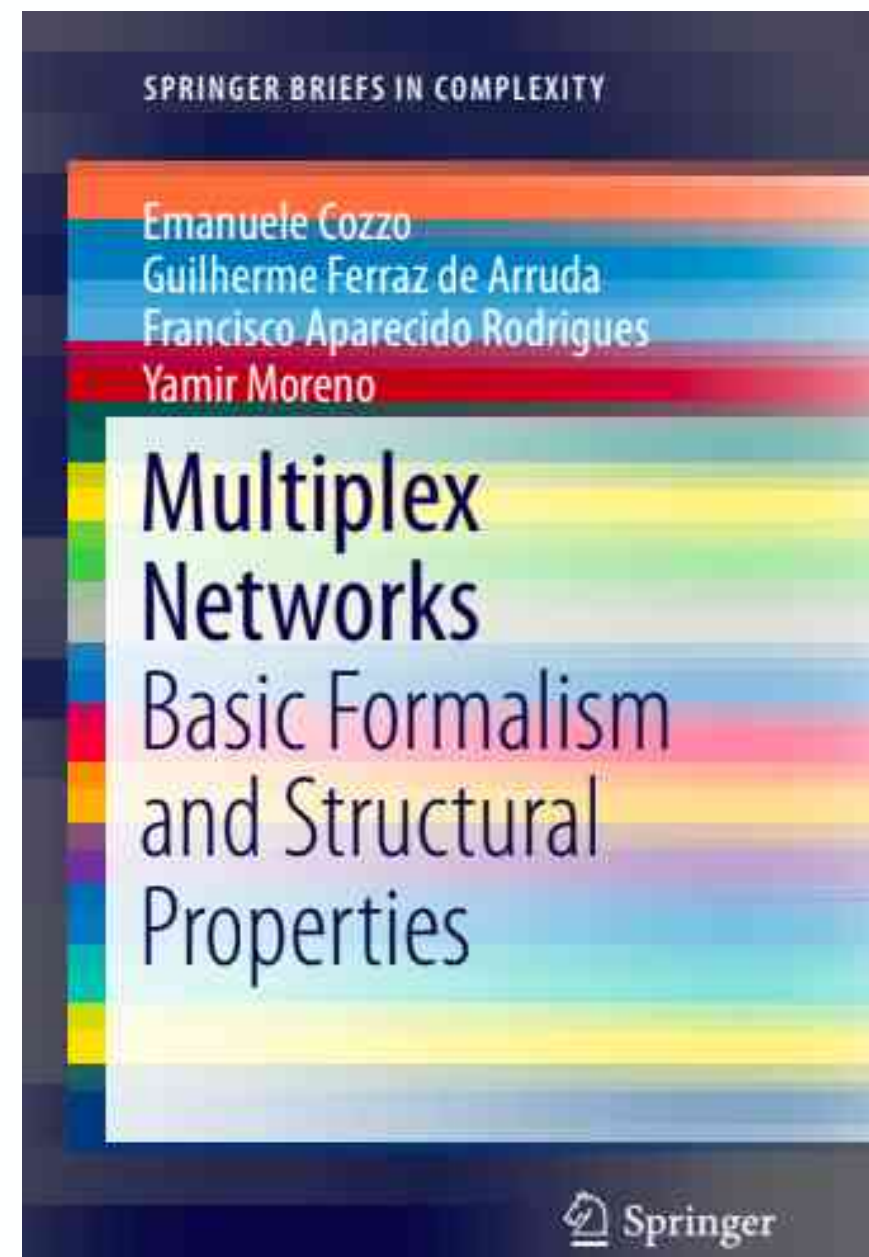
- Temporal networks
- Multilayer networks
- Networks with noise
- Heterogeneous dynamics
- ...
- A new statistics?



Thank you!

<https://sites.icmc.usp.br/francisco>

francisco@icmc.usp.br



Spreading in Multilayer Networks

Multilayer networks

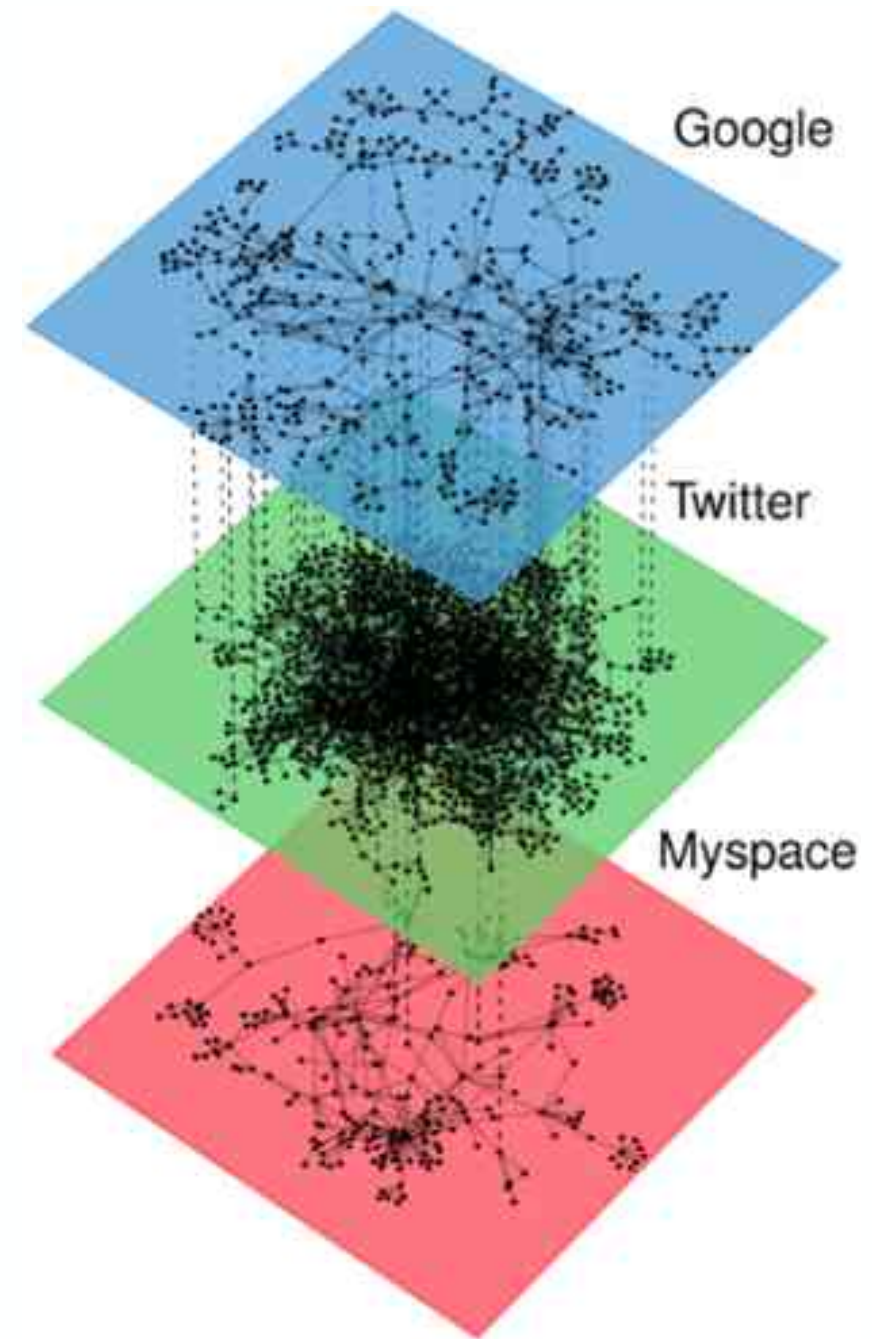
TRAIN



METRO

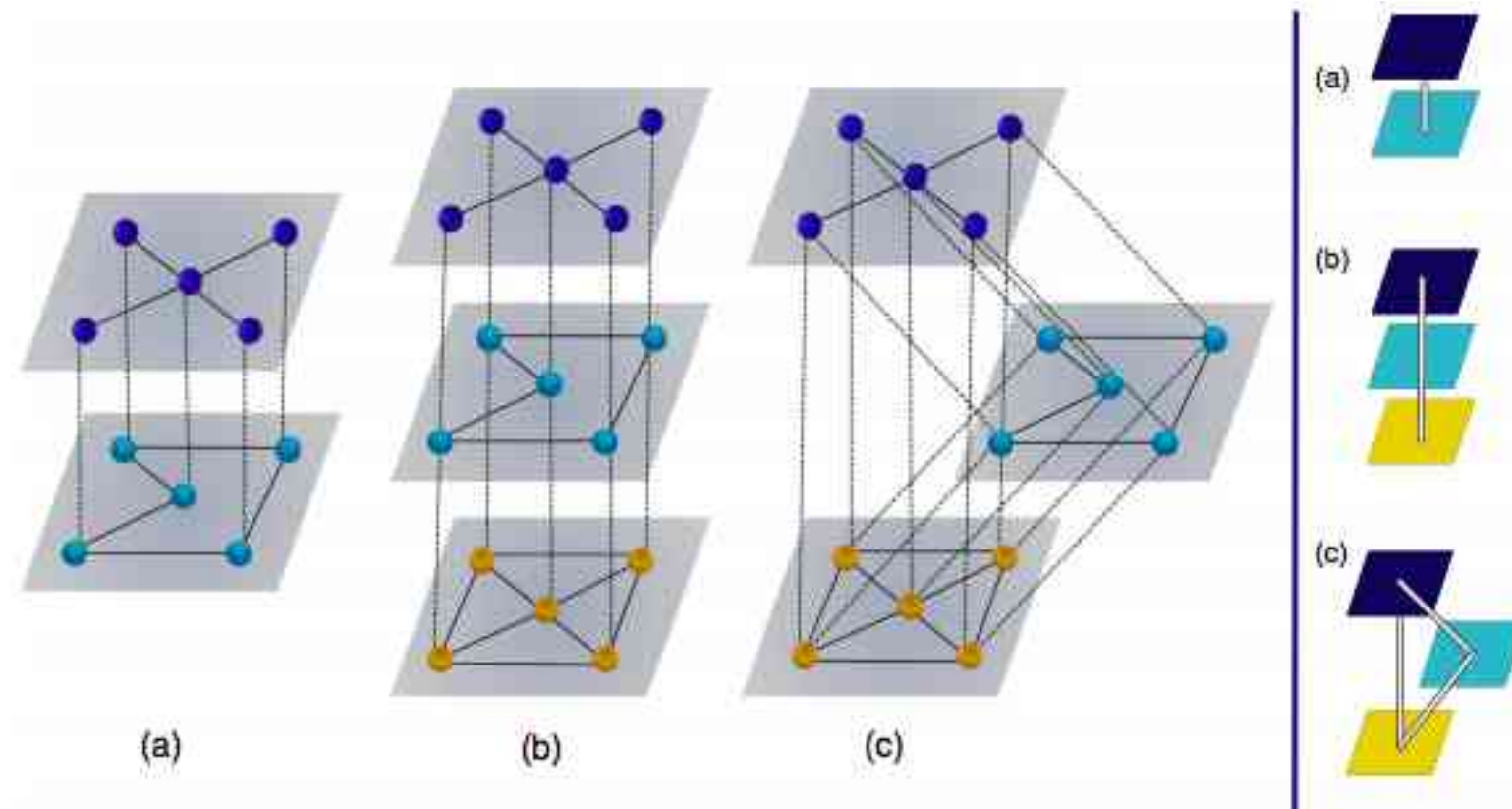


ROAD



Multilayer networks

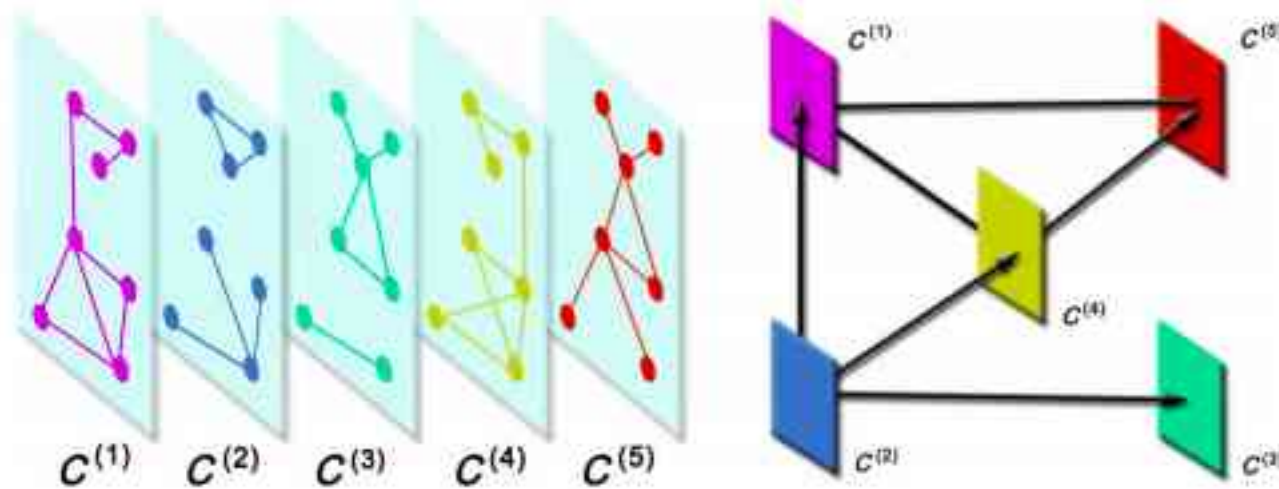
- A multilayer network is a network made up of several layers:



Theoretical formulation

- A multilayer network is represented as the fourth-order adjacency tensor $M \in \mathbb{R}^{n \times n \times m \times m}$

$$M_{\beta\tilde{\gamma}}^{\alpha\tilde{\delta}} = \sum_{\tilde{h}, \tilde{k}=1}^m C_{\beta}^{\alpha}(\tilde{h} \tilde{k}) E_{\tilde{\gamma}}^{\tilde{\delta}}(\tilde{h} \tilde{k})$$



- The adjacency matrix representing the connections inside layers:

$$M_{\beta\tilde{\gamma}}^{\alpha\tilde{\delta}} E_{\tilde{\delta}}^{\tilde{\gamma}}(\tilde{r} \tilde{r}) = C_{\beta}^{\alpha}(\tilde{r} \tilde{r}) = A_{\beta}^{\alpha}(\tilde{r})$$

Multilayer networks

- The equations describing the system dynamics:

$$\frac{dX_{\beta\bar{\delta}}}{dt} = -\mu X_{\beta\bar{\delta}} + (1 - X_{\beta\bar{\delta}})\lambda\mathcal{R}_{\beta\bar{\delta}}^{\alpha\bar{\gamma}}(\lambda, \eta)X_{\alpha\bar{\gamma}}$$

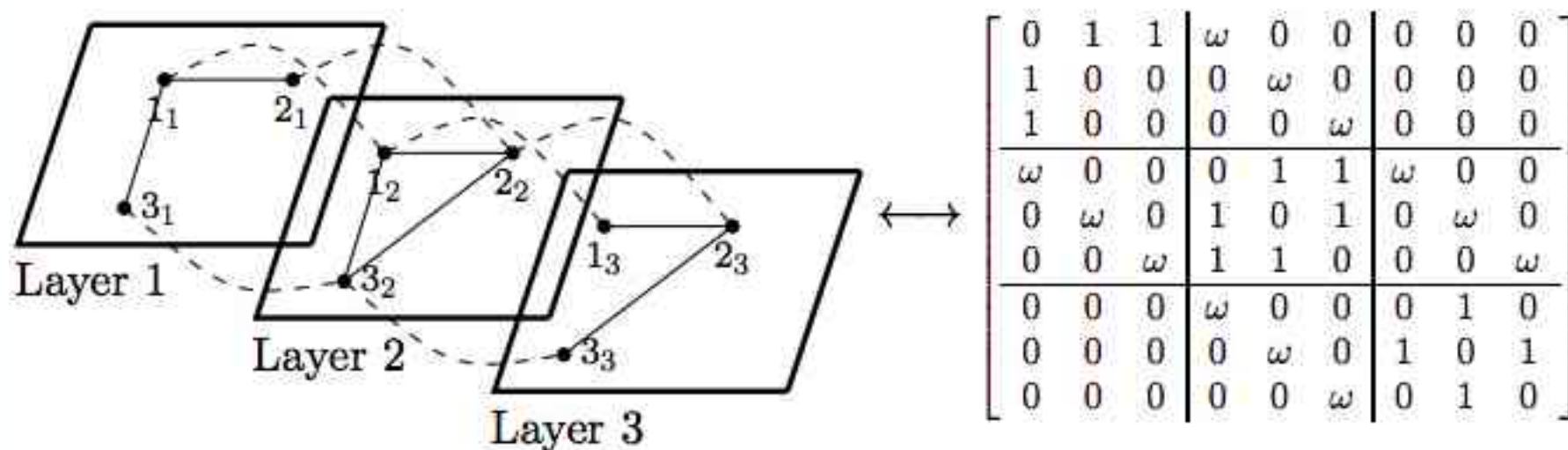
Multilayer networks

- Following the tensorial formulation, we can show:

$$\left(\frac{\mu}{\lambda}\right)_c = \Lambda_1$$

- where Λ_1 is the largest eigenvalue of the supra-adjacency matrix:

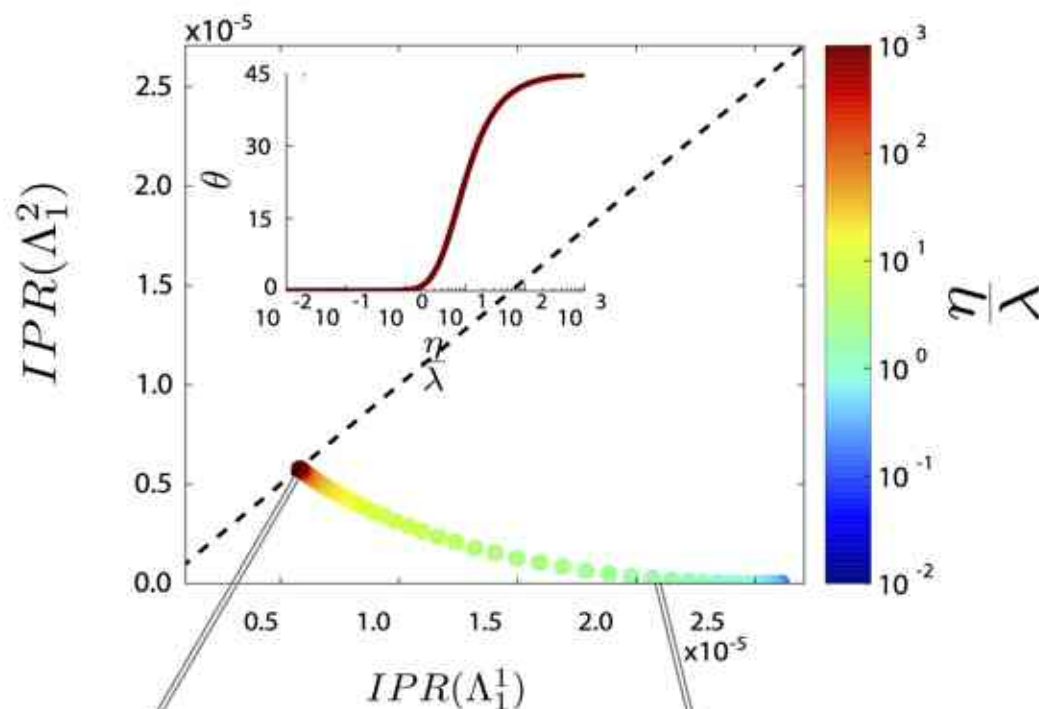
$$\mathcal{R}_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}}(\lambda, \eta) = M_{\beta\tilde{\sigma}}^{\alpha\tilde{\eta}} E_{\tilde{\eta}}^{\tilde{\sigma}}(\tilde{\gamma} \tilde{\delta}) \delta_{\tilde{\delta}}^{\tilde{\gamma}} + \frac{\eta}{\lambda} M_{\beta\tilde{\sigma}}^{\alpha\tilde{\eta}} E_{\tilde{\eta}}^{\tilde{\sigma}}(\tilde{\gamma} \tilde{\delta}) (U_{\tilde{\delta}}^{\tilde{\gamma}} - \delta_{\tilde{\delta}}^{\tilde{\gamma}})$$



Disease localization

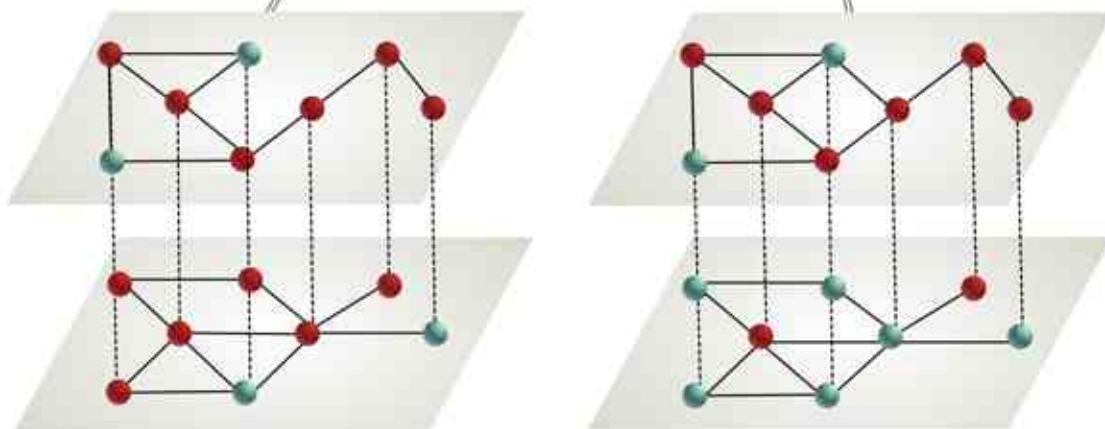
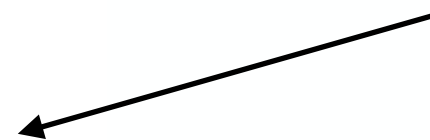
- Defining the inverse participation ratio

$$\text{IPR}(\Lambda) \equiv (f_{\beta\tilde{\delta}}(\Lambda))^4 U^{\beta\tilde{\delta}}$$



- The localisation of the disease in a certain layer means that most of the spreading is expected to take place on the nodes of that layer.

Disease localisation



Outline

1. Epidemic spreading
2. Networks
3. Awareness
4. Prediction