Graph and Social Data Management
Spreading Phenomena

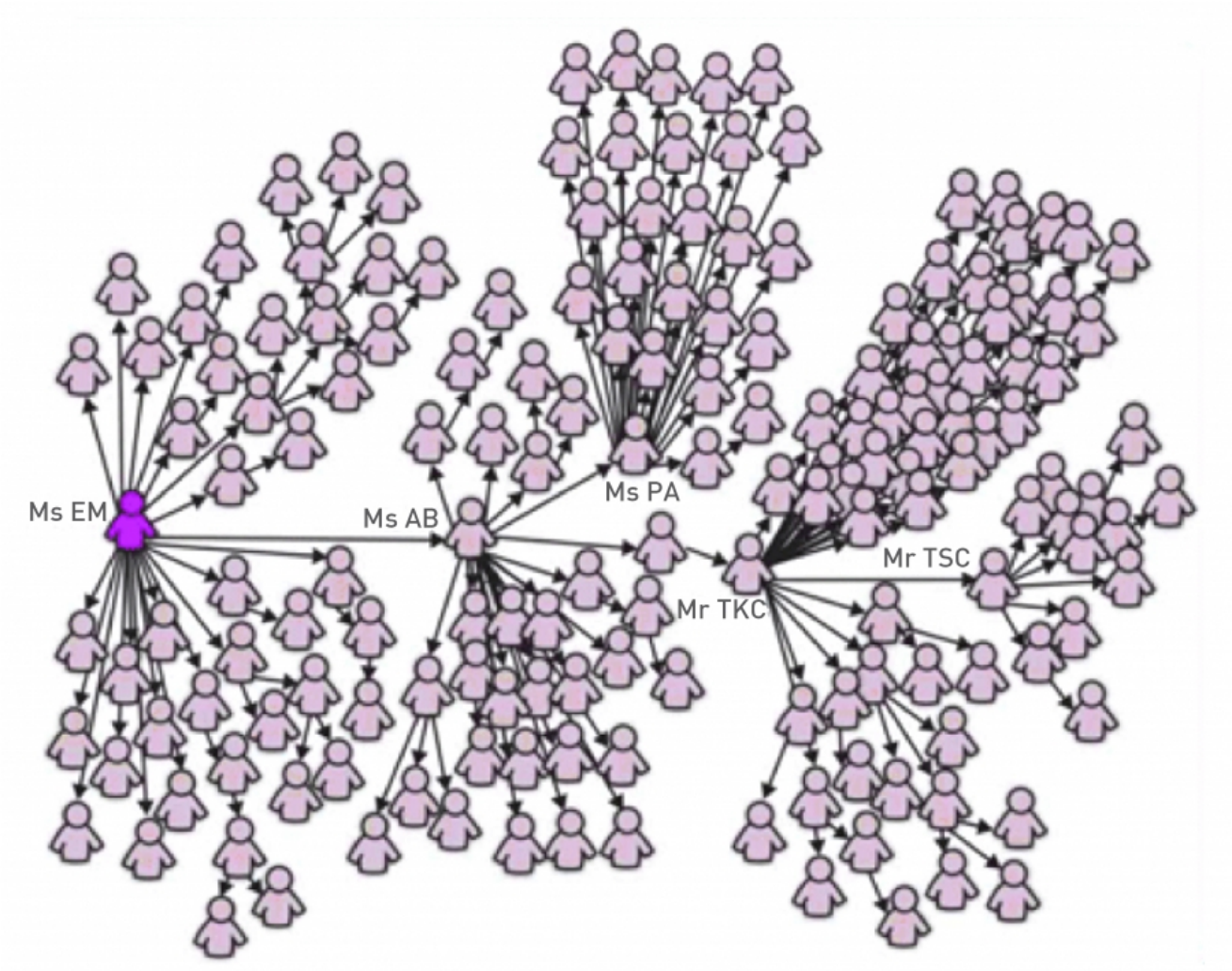
Benoît Groz (slides and content are mostly from Silviu Maniu)

December 1st, 2023

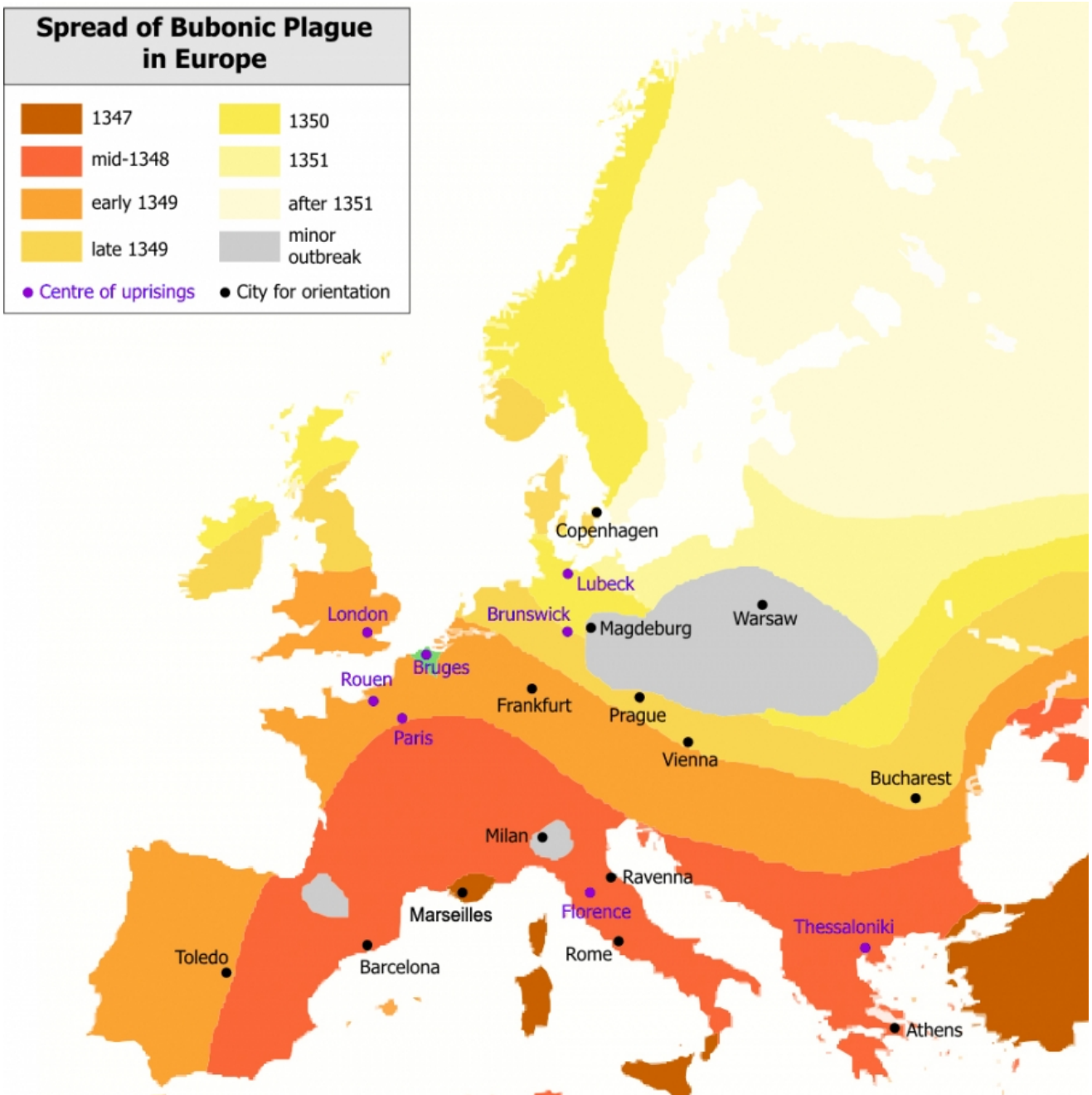
M2 Data Science

Table of contents

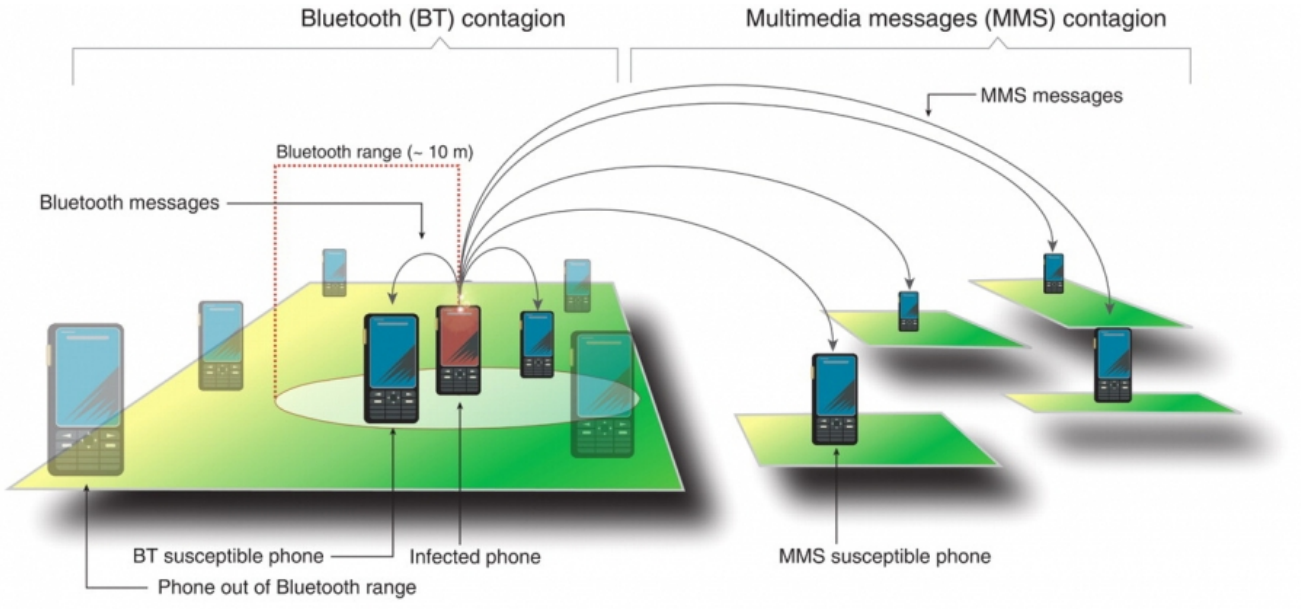
SARS and Super-spreaders



Bubonic Plague



Smartphone Viruses



Network Epidemics

Network epidemics – a framework allowing quantitative analysis and forecasting of infectious phenomena, in which *infections* spread through *networks* facilitated by *agents*

Occurs in diverse domains:

- biology: pathogens (influenza, SARS, Covid, tuberculosis, ...)
- digital: computer viruses and worms
- social: information cascades (innovation, products, memes)

Networks and Agents

phenomenon	**agent**	**network**
rumour spreading	information, memes	communication
innovation diffusion	ideas	communication
computer virus	malware	Internet
diseases	pathogen	human-human network
bedbugs	insects	hotel-guest network

Epidemic Modeling

The network epidemic modeling rests on two main hypotheses:

1. compartmentalization
2. homogeneous mixing

Compartmentalization

Classify each individual (node) depending of the state (or compartment) of infection:

- susceptible (S): healthy individuals
- infectious (I): contagious individuals having contracted the pathogen
- recovered (R): individuals having been infected before, but have recovered

Homogeneous Mixing

Anyone (infectious) can infect anyone, with the same probability.

Table of contents

Susceptible-Infected (SI) Model

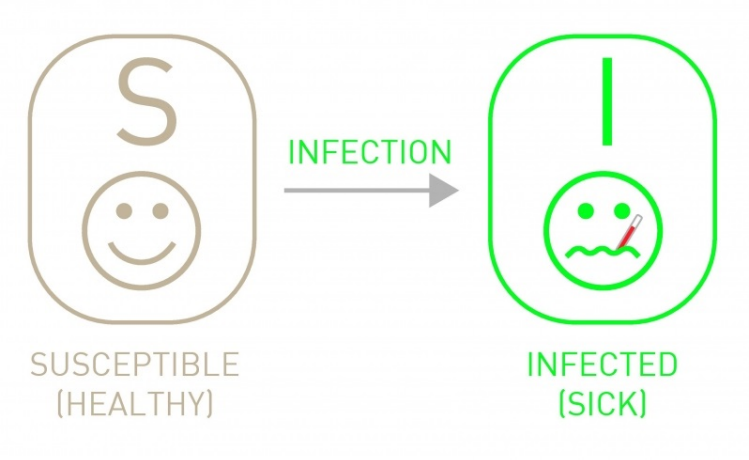
Assumptions and notation:

- consider N individuals, each having $\langle k \rangle$ contacts
- infections occur in time increments
- $S(t)$ number of susceptible individuals at time t
- $I(t)$ number of infected individuals at time t
- $S(0) = N$ (and by default $I(0) = 0$)
- likelihood of infection is a parameter β

Assuming $I(0) = 1$, how many will be infected at a later time t?

Susceptible-Infected (SI) Model

SI Model: an individual can be in two states, *healthy* (S) or *sick* (I), becoming infected at a rate β



Susceptible-Infected (SI) Model

$I(t)$ changes at the rate

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

Solving for $i(t) = I(t)/N$ (fraction of individuals infected):

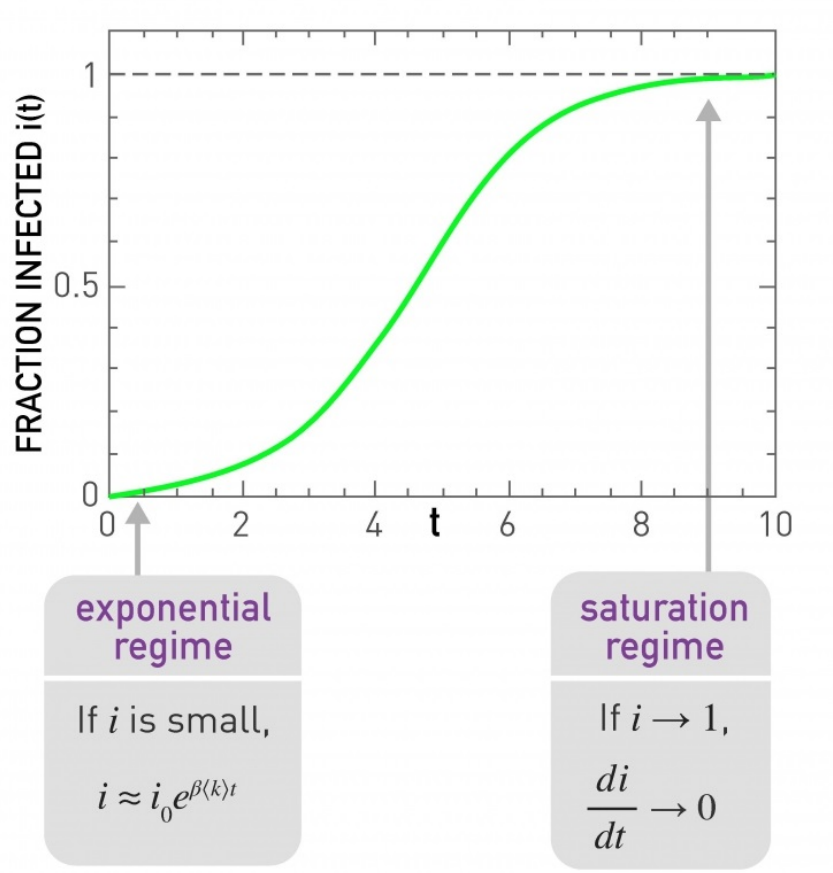
$$i(t) = \frac{i(0) e^{\beta \langle k \rangle t}}{1 - i(0) + i(0) e^{\beta \langle k \rangle t}}, \tag{1}$$

$\beta \langle k \rangle$ is called transmission rate

Susceptible-Infected (SI) Model

Characteristic time – time to reach a $1/e \approx 0.36$ fraction of infected individuals

$$\tau = \frac{1}{\beta \langle k \rangle}$$



Susceptible-Infected-Susceptible (SIS) Model

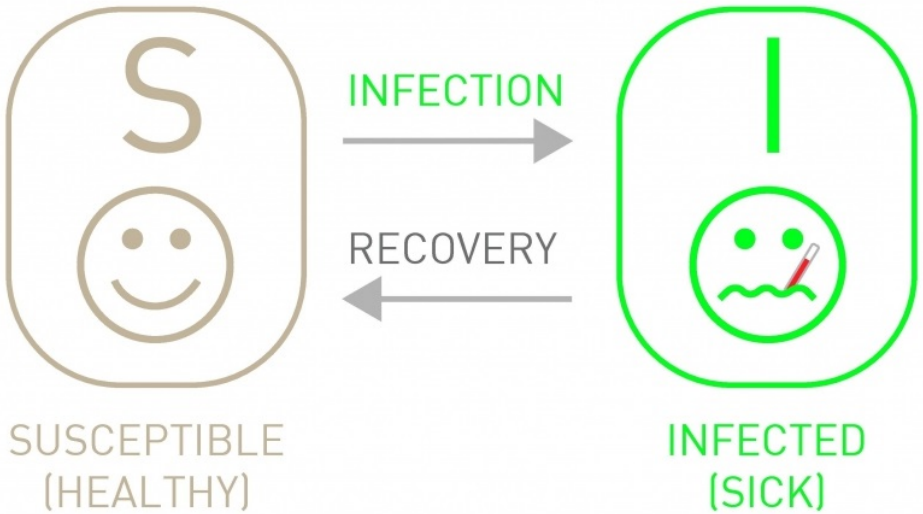
Same as SI, but nodes can recover, at a rate μ

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

and

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

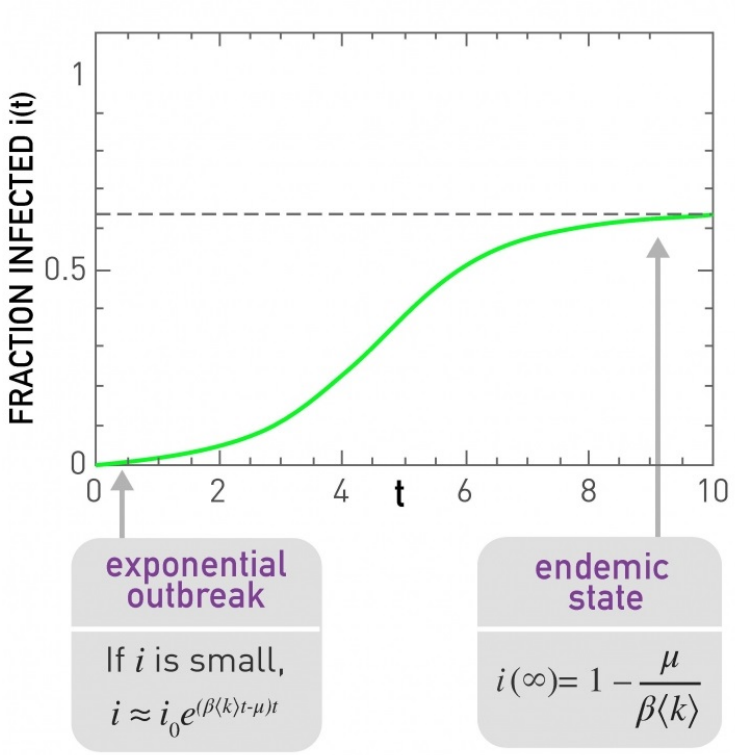
with $C = i_0/(1 - i_0 - \mu/(\beta \langle k \rangle)$



Susceptible-Infected-Susceptible (SIS) Model

Two possible outcomes:

- **endemic state**, $\mu < \beta\langle k \rangle$, not everyone is infected, but i reaches a plateau
- **disease-free state**, $\mu > \beta\langle k \rangle$, i decreases with time, so the disease dies out



SIS Model: basic reproductive number

Characteristic time

$$\tau = \frac{1}{\mu(R_0 - 1)}$$

depends on the basic reproductive number

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

Depending on R_0:

- $R_0 > 1$, $\tau > 0$, epidemic is in endemic state,
- otherwise disease free.

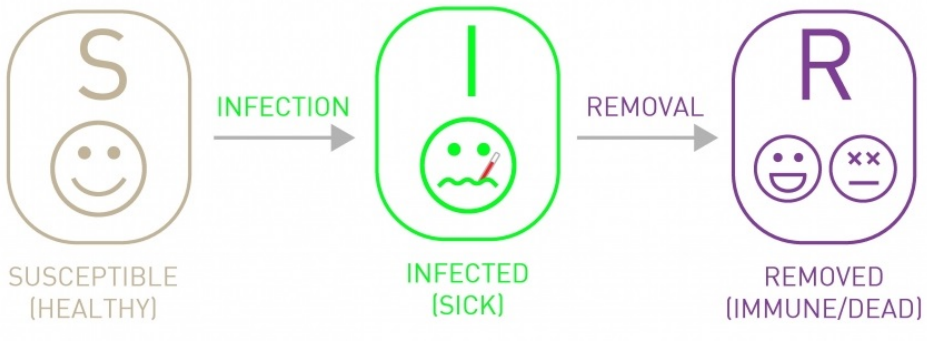
SIS Model: basic reproductive number

Disease	Transmission	R_0
Measles	Airborne	12-18
Smallpox	Social contact	5-7
HIV	Sexual contact	2-5
SARS	Airborne droplet	2-5
Influenza 1918	Airborne droplet	2-3

Susceptible-Infected-Recovered (SIR) Model

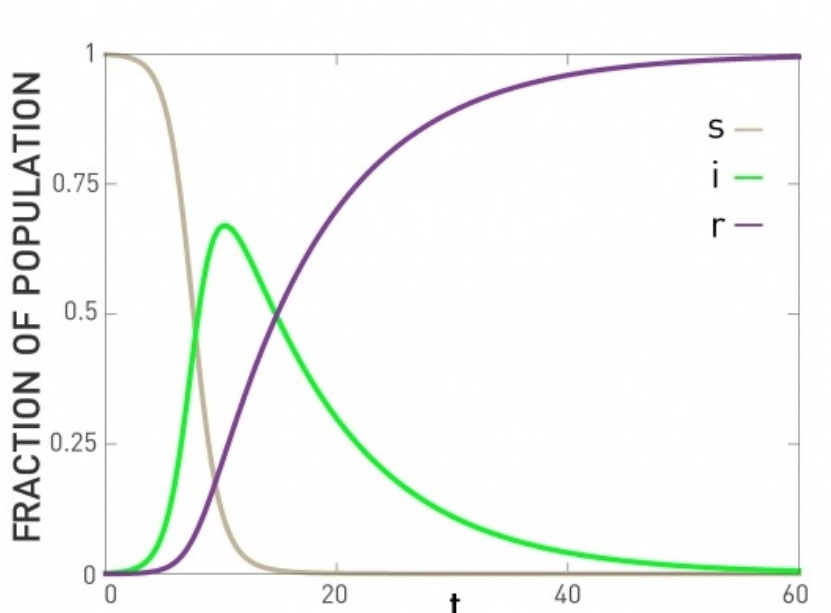
Nodes can be **recovered**, governed by a **recovery rate** r

No closed form solution for i – it depends on the rate of $s(t)$ and $r(t)$

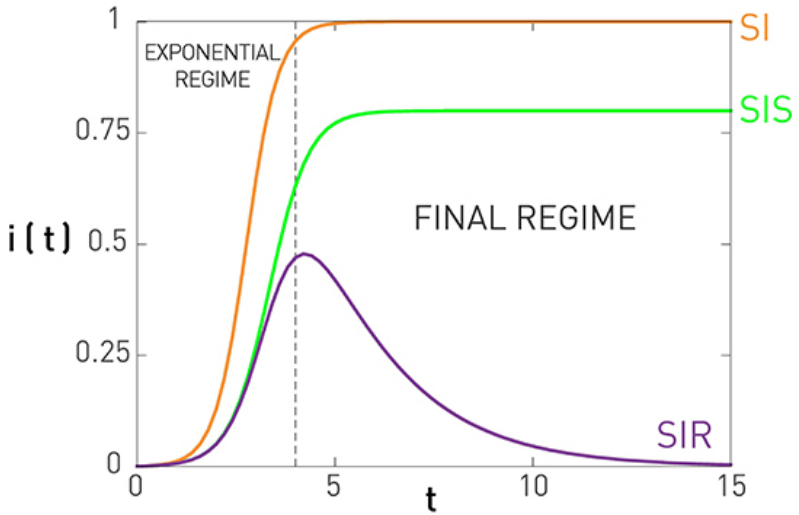


Susceptible-Infected-Recovered (SIR) Model

$$\frac{\partial s}{\partial t} = -\beta\langle k\rangle(1 - r - i)$$
$$\frac{\partial i}{\partial t} = -\mu i + \beta\langle k\rangle(1 - r - i)$$
$$\frac{\partial r}{\partial t} = \mu i$$



Model Overview



	SI	SIS	SIR
Exponential Regime: Number of infected individuals grows exponentially	$i = \frac{i_0 e^{\beta\langle k\rangle t}}{1 - i_0 + i_0 e^{\beta\langle k\rangle t}}$	$i = \left(1 - \frac{\mu}{\beta\langle k\rangle}\right) \frac{C e^{(\beta\langle k\rangle - \mu)t}}{1 + C e^{(\beta\langle k\rangle - \mu)t}}$	No closed solution
Final Regime: Saturation at t⟶=∞	$i(\infty) = 1$	$i(\infty) = 1 - \frac{\mu}{\beta\langle k\rangle}$	$i(\infty) = 0$
Epidemic Threshold: Disease does not always spread	No threshold	$R_0 = 1$	$R_0 = 1$

Table of contents

Taking Network into Account

Previous models do not use the actual network, and capture only behaviour on aggregate

They take into account $\langle k \rangle$ which is not always a good approximation for e.g., scale-free networks

We should study these models on some approximation of real networks – degree-block approximation

Degree-Block Approximation

SI Model on Networks

Assumes that nodes with the same degree are statistically equivalent

So i depends on k also:

$$\frac{\partial i_k}{\partial t} = \beta(1 - i_k)k\Theta_k,$$

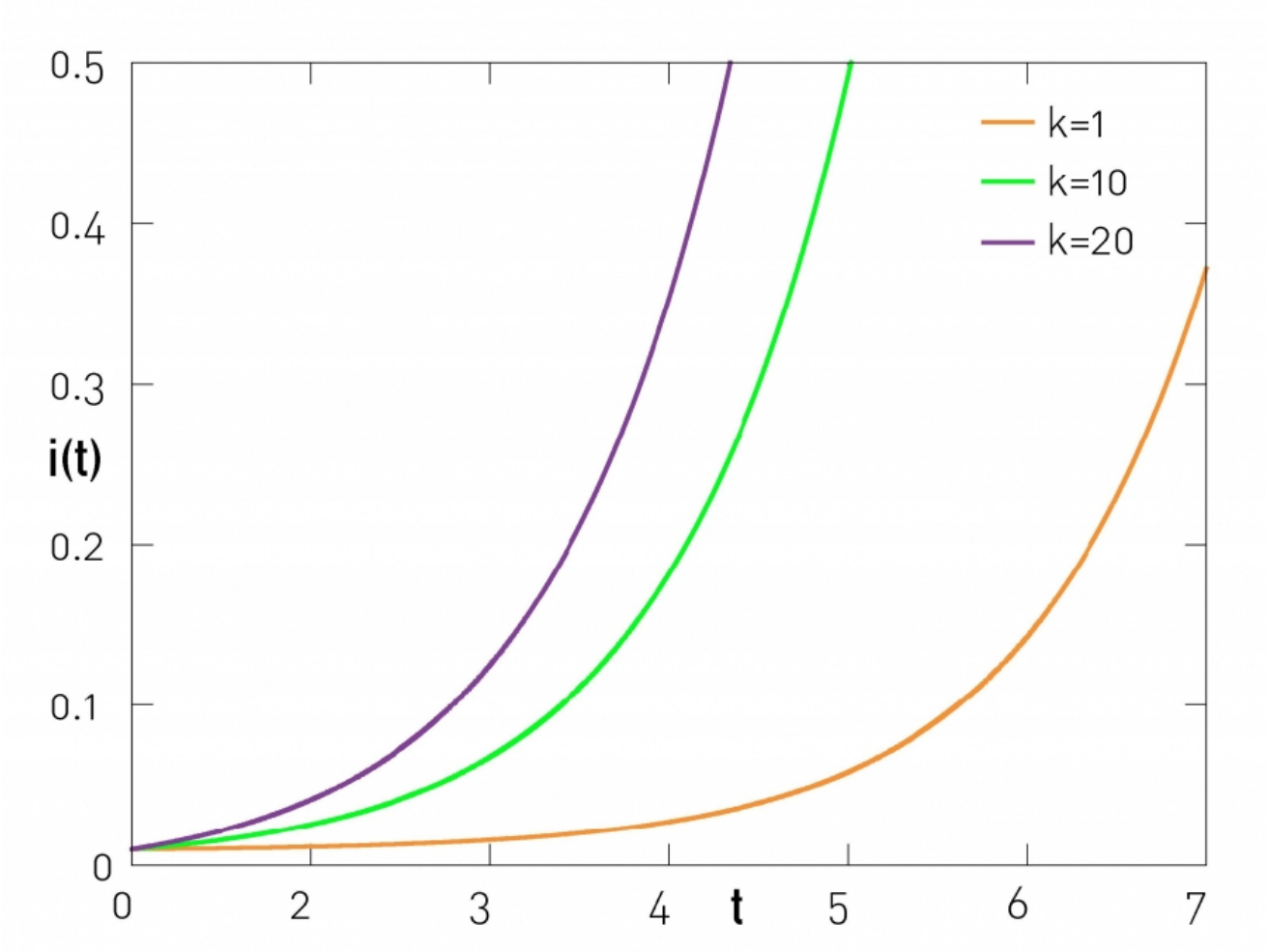
$$i_k = \frac{I_k}{N_k} = i_0 \left(1 + \frac{k\langle k\rangle - 1}{\langle k^2\rangle - \langle k\rangle}(e^{t/\tau^{SI}} - 1)\right),$$

where the characteristic time depends on the variance of the degree also at the beginning:

$$\frac{\partial i_k}{\partial t} \approx \beta k\Theta_k \approx \beta k i_0 \frac{\langle k\rangle - 1}{\langle k\rangle} e^{t/\tau^{SI}} \quad \text{with } \tau^{SI} = \frac{\langle k\rangle}{\beta(\langle k^2\rangle - \langle k\rangle)}$$

$$i = \int_0^{k_{max}} i_k p_k dk = i_0(1 + \frac{\langle k\rangle^2 - \langle k\rangle}{\langle k^2\rangle - \langle k\rangle}(e^{t/\tau^{SI}} - 1)$$

SI Model on Networks



SI Model on Networks

Depending on the type of network, we have different results:

- **Random networks**, where $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$ we are in the same case as homogeneous networks (so the classic SI model)
- **Scale-free networks**, $\gamma \geqslant 3$, $\langle k \rangle$ and $\langle k^2 \rangle$ are finite, so τ^{SI} is also finite, so similar to random networks
- **Scale-free networks**, $\gamma < 3$, $\langle k^2 \rangle$ diverges, which means that $\tau^{SI} \to 0$ – *spread in scale-free networks is instantaneous* (**vanishing characteristic time**) Why?

SIS Model on Networks

More realistic model: some nodes revert to initial state (S)

Straightforward extension from SI, by taking into account μ:

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

Characteristic time changes to:

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

SIS Model on Networks

In SIS, the spread depends both on β and μ and the difference between the two values

Spreading rate:

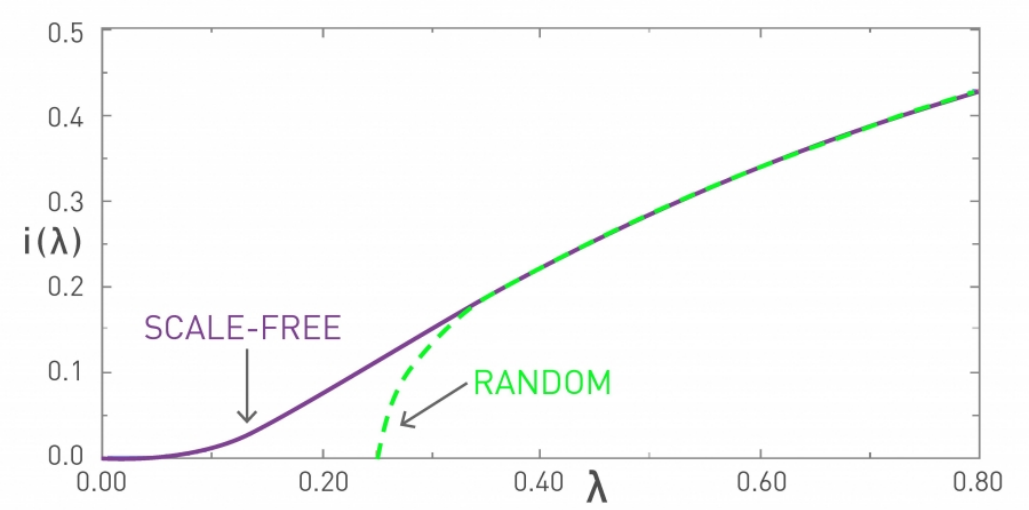
$$\lambda = \frac{\beta}{\mu}$$

We have to check how this relates to an epidemic threshold λ_c

Epidemic Threshold in SIS

Random Networks – infection persist after $\lambda_c = \frac{1}{\langle k \rangle + 1}$; if $\lambda < \lambda_c$ the network is disease free

Scale-Free Networks – epidemic threshold is $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \to 0$; the disease spreads even for very low λ values (**vanishing epidemic threshold**)



Epidemics in Scale-Free Networks

Main takeaways:

- characteristic time $\tau = \mathbf{o}$, viruses can reach most nodes instantaneously
- epidemic threshold $\lambda_c = \mathbf{o}$, viruses with small spreading rate can persist

Result from the fact that **hubs** can propagate to many neighbours

Effect of Degree Correlations

- degree correlations alter the threshold λ_c (assortativity decreases it)
- in scale free networks, the threshold still vanishes no matter its correlations
- since hubs are the first affected, assortativity accelerates the spread

Effect of Communities in Information Spread

Inside communities, ties between nodes are closer (**strong ties**), and between communities ties are **weak**

Direct influence over **information** spread:

- **information spreads fast inside communities** – due to the strong ties
- **information is trapped in a community** – due to the weak ties, it is less likely to “escape” a community

Types of Information Spread

Simple contagion: studied until now, i.e., simple contact suffices for infection

Complex contagion: information needs *reinforcement*, i.e., multiple sources of infection or information

Types of Information Spread

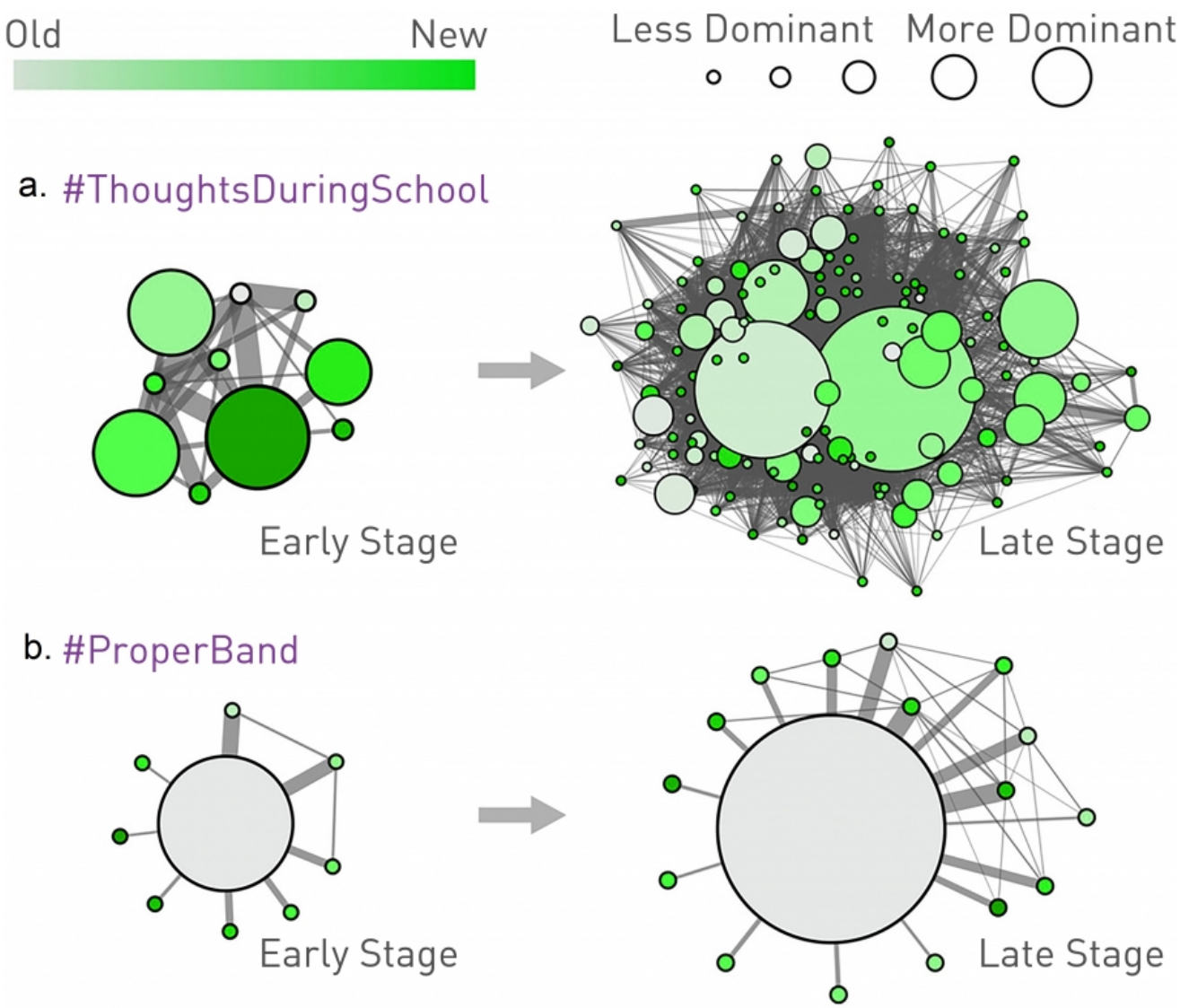


Table of contents

Immunization

How do we stop an infection in a network?

Objective: immunize a fraction g_c of nodes so that λ goes under λ_c; immunized nodes are "invisible"

Strategies:

- random
- selective

Random Immunization in Random Networks

Fraction g_c chosen randomly

We want that:

$$\frac{(1-g_c)\beta}{\mu} = \frac{1}{\langle k \rangle + 1},$$

so:

$$g_c = 1 - \frac{\mu}{\beta(\langle k \rangle + 1)}$$

The more nodes are immunized the better, but still less that the total number of nodes

Random Immunization in Heterogeneous Networks

Heterogeneous networks – high $\langle k^2 \rangle$

We want that:

$$(1 - g_c)\frac{\beta}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle},$$

so:

$$g_c = 1 - \frac{\mu \langle k \rangle}{\beta \langle k^2 \rangle}$$

We need to immunize a large fraction of nodes in the networks

For scale-free networks, $g_c \to 1$

Selective Immunization in Scale-Free Networks

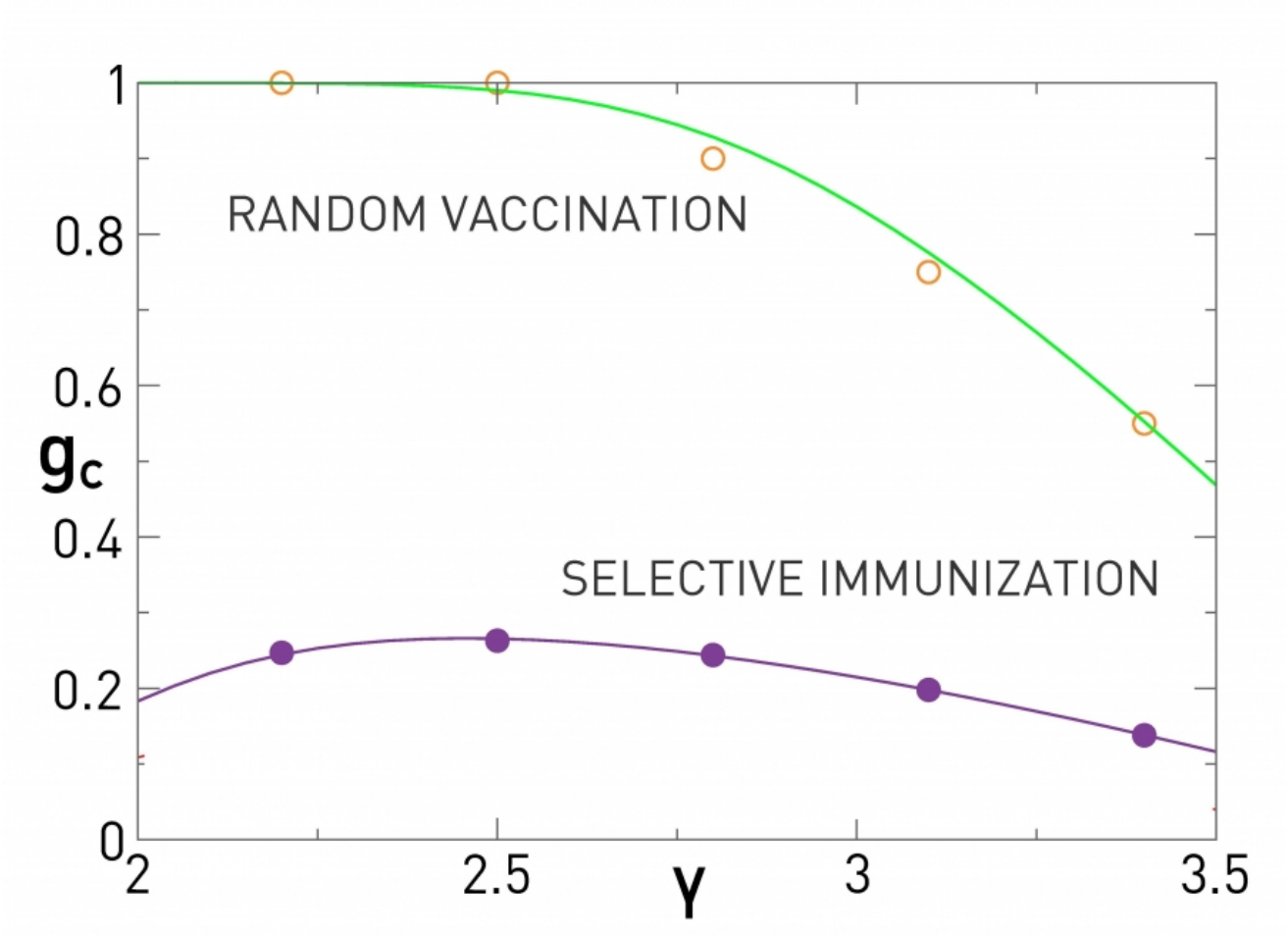
We should select **hubs** first, so that the network becomes *disconnected* – but the actual network is not always known

Strategy for selective immunization based on **friendship paradox**: *your friends are more popular than you*

1. choose **randomly** a fraction p of nodes – *Group 0*
2. for each node in *Group 0*, select a link randomly – put resulting nodes in *Group 1*
3. **immunize** *Group 1*

Why does it work? – nodes in *Group 1* have higher average degree than those in *Group 0*

Selective Immunization in Scale-Free Networks



Acknowledgments

Figures in slides 3, 4, 12, 14, 16, 18, 19, 20, 23, 25, 29, 33, 34, and 40 taken from the book “Network Science” by A.-L. Barabási. The contents is partly inspired by the flow of Chapter 10 of the same book.

`http://barabasi.com/networksciencebook/`

References i


Castellano, C. and Pastor-Satorras, R. (2010).
Thresholds for epidemic spreading in networks.
Phys. Rev. L., 105 21.


Pastor-Satorras, R. and Vespignani, A. (2001).
Epidemic spreading in scale-free networks.
Phys. Rev. L., 86.