Social and Technological Network Data Analytics

Lecture 10: Epidemics Spreading

Prof Cecilia Mascolo

UNIVERSITY OF CAMBRIDGE

In This Lecture

- In this lecture we introduce the process of spreading epidemics in networks.
  - This has been studied widely in various disciplines from different perspectives from theoretical models to simulations of real events.
  - But it also has important parallels and applications in information/idea diffusion in social and technological networks.

Epidemics vs Cascade Spreading

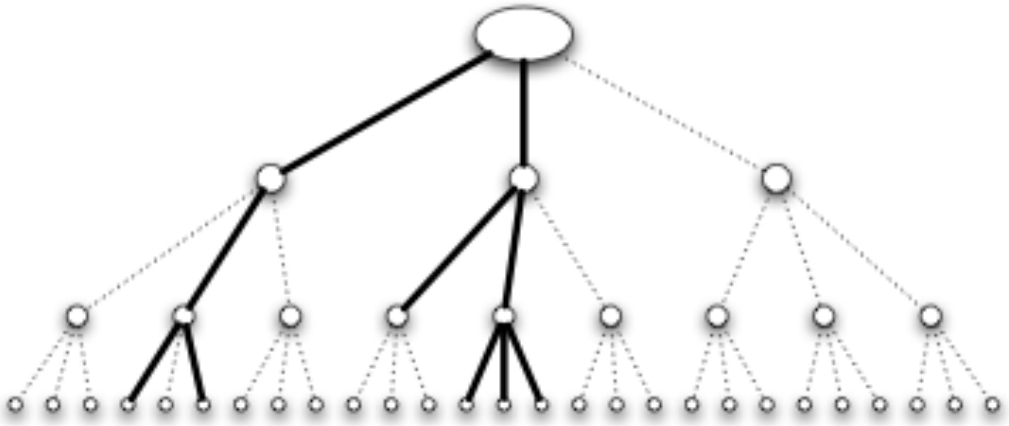
- In cascade spreading nodes make decisions based on pay-off benefits of adopting one strategy or the other.
- In epidemic spreading
  - Lack of decision making.
  - Process of contagion is complex and unobservable
    - In some cases it involves (or can be modeled as) randomness.

Branching Process

- Simple model.
- **First wave:** A person carrying a disease enters the population and transmits to all he meets with probability p. He meets k people: a portion of which will be infected.
- **Second wave:** each of the k people goes and meet k different people. So we have a second wave of $k*k=k^2$ people.
- **Subsequent waves:** same process.

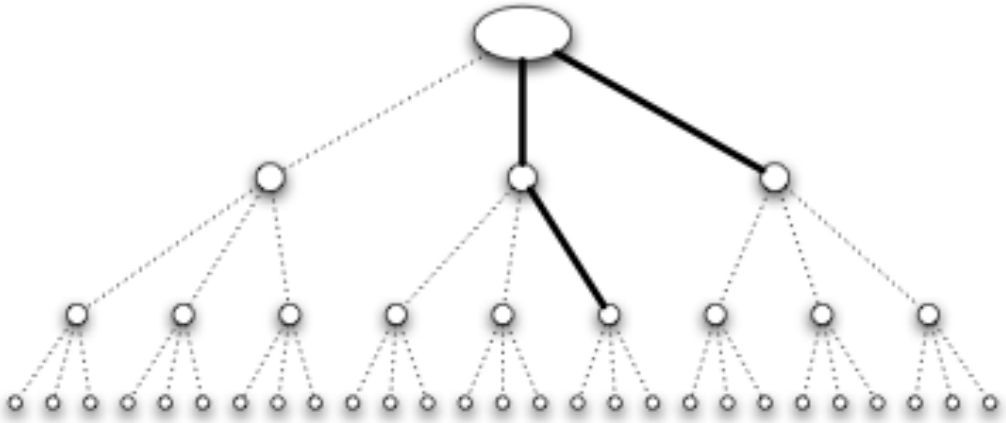
UNIVERSITY OF CAMBRIDGE

Example with k=3



High contagion probability:
The disease spreads

Low contagion probability:
The disease dies out



UNIVERSITY OF CAMBRIDGE

Basic Reproductive Number

- Basic Reproductive Number $R_0=p*k$
  - It determines if the disease will spread or die out.
- In the branching process model, if $R_0<1$ the disease will die out after a finite number of waves. If $R_0>1$, with probability >0, the disease will persist by infecting at least one person in each wave.

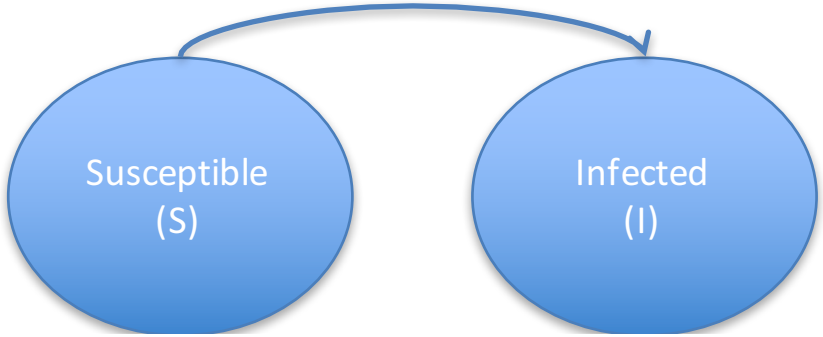
Measures to Limit the Spreading

- When R_0 is close 1, slightly changing p or k can result in epidemics dying out or happening.
  - Quarantining people/nodes reduces k.
  - Encouraging better sanitary practices reduces germs spreading [reducing p].
- Limitations of this model:
  - No realistic contact networks: no triangles!
  - Nodes can infect only once.
  - No nodes recover.

UNIVERSITY OF CAMBRIDGE

Formal Epidemics Models
The SI Model

- S: susceptible individuals.
- I: infected individuals, when infected they can infect others continuously.
- n: total population.
- <k> average contacts per individual
- $\beta=\lambda\langle k\rangle$ is the infection rate per individual ($0\leq\lambda\leq1$)
- Susceptible contacts per unit of time $\beta S/n$.
- Overall rate of infection $I\beta S/n$.



UNIVERSITY OF CAMBRIDGE

SI Model

$$\frac{dI}{dt} = \beta \frac{SI}{n}$$

$$\frac{dS}{dt} = -\beta \frac{SI}{n}$$

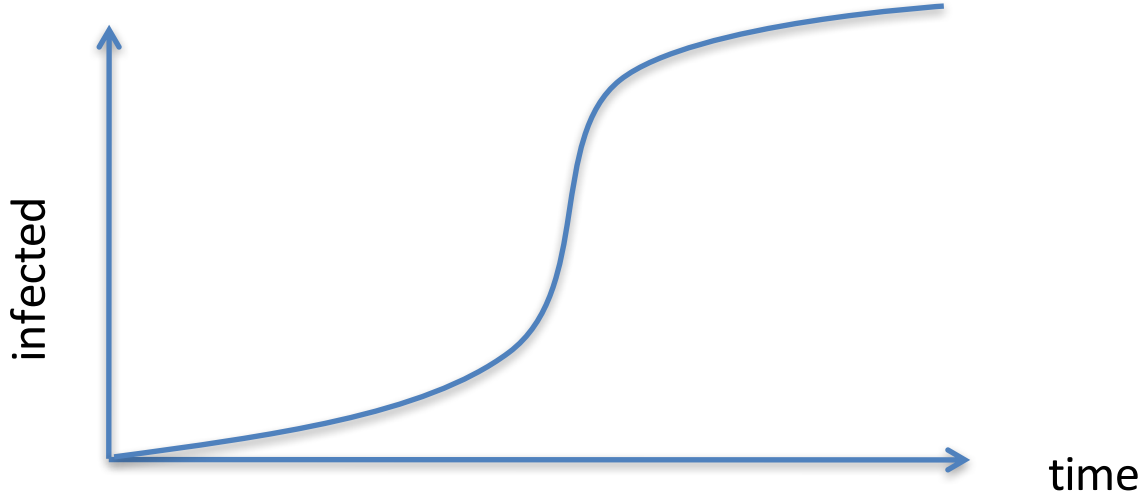
$$s = \frac{S}{n} \qquad i = \frac{I}{n}$$

$$s = 1 - i$$

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

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

Logistic Growth Equation



UNIVERSITY OF CAMBRIDGE

SIR Model

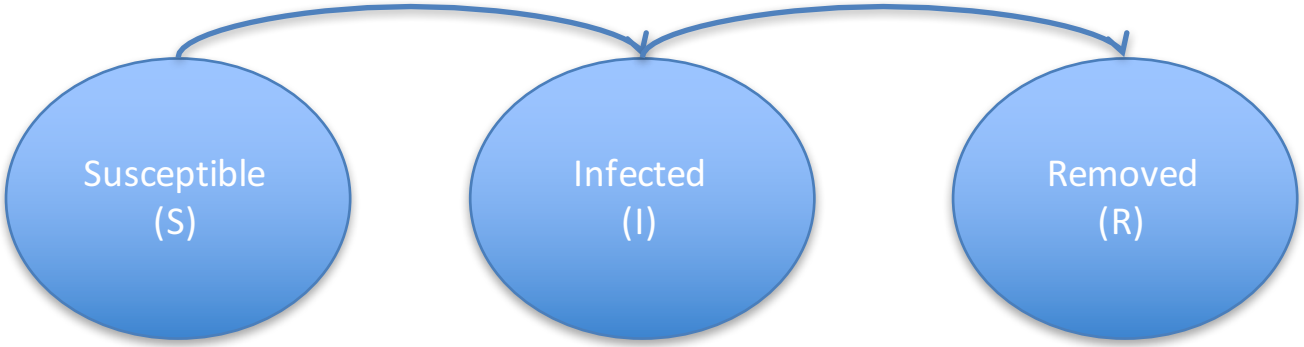
- Infected nodes recover at a rate γ.
- A node stays infected for τ time.
- Branching process is SIR with τ=1.

$$\frac{ds}{dt} = -\beta si$$

$$\frac{di}{dt} = \beta si - \gamma i$$

$$\frac{dr}{dt} = \gamma i$$

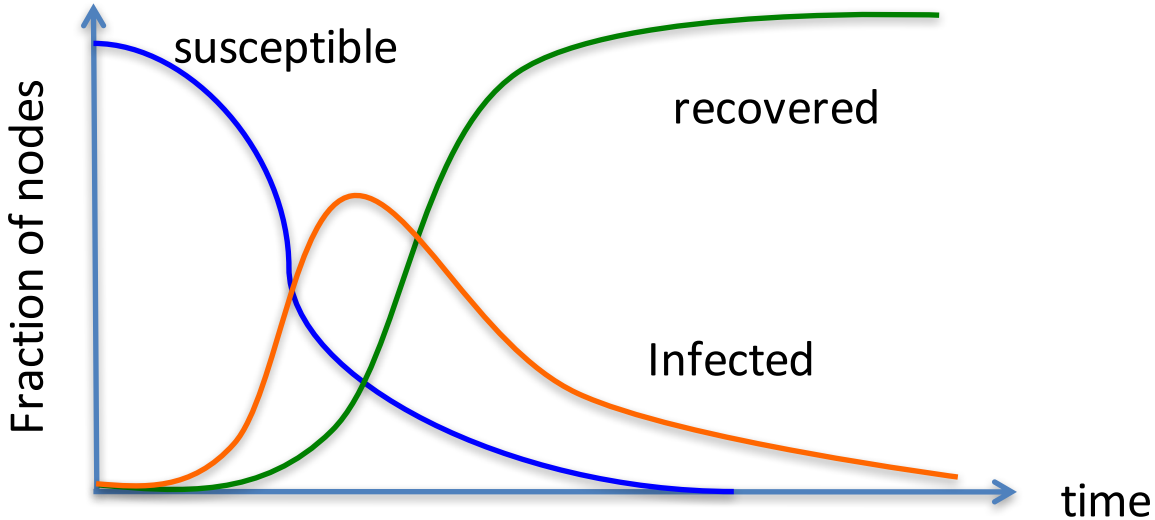
$$s + i + r = 1$$



UNIVERSITY OF CAMBRIDGE

Example

- Numerical examples of solution:
- β=1, γ=0.4, s(at start)=0.99, i(at start)=0.01, r(at start)=0



UNIVERSITY OF CAMBRIDGE

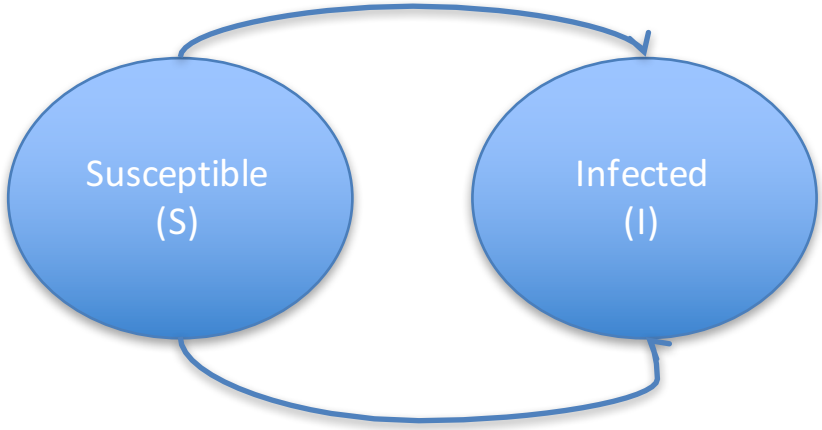
SIS Model

$$\frac{ds}{dt} = \gamma i - \beta si$$

$$\frac{di}{dt} = \beta si - \gamma i$$

$$s + i = 1$$

- If $\beta > \gamma$ growth curve like in SI but never reaching all population infected. The fraction of infected->0 as β approaches γ.
- If $\beta < \gamma$ the infection will die out exponentially.
- SIS has the same R_0 as SIR.



UNIVERSITY OF CAMBRIDGE

Epidemic Threshold

- When would the epidemic develop and when would it die out?

- It depends on the relationship of β and γ:
  - Basic Reproductive Number $R_0=\beta/\gamma$
  - If the infection rate [per unit of time] is higher than the removal rate the infection will survive otherwise it will die out.
  - In SI, γ=0 so the epidemics always happen.

Limitations of SIR

- Contagion probability is uniform and “on-off”
- Extensions
  - Probability q of recovering in each step.
  - Infected state divided into intermediate states (early, middle and final infection times) with varying probability during each.
  - **We have assumed homogenous mixing :** assumes all nodes encounter each others with same probability: we could assume different probability per encounter.

UNIVERSITY OF CAMBRIDGE

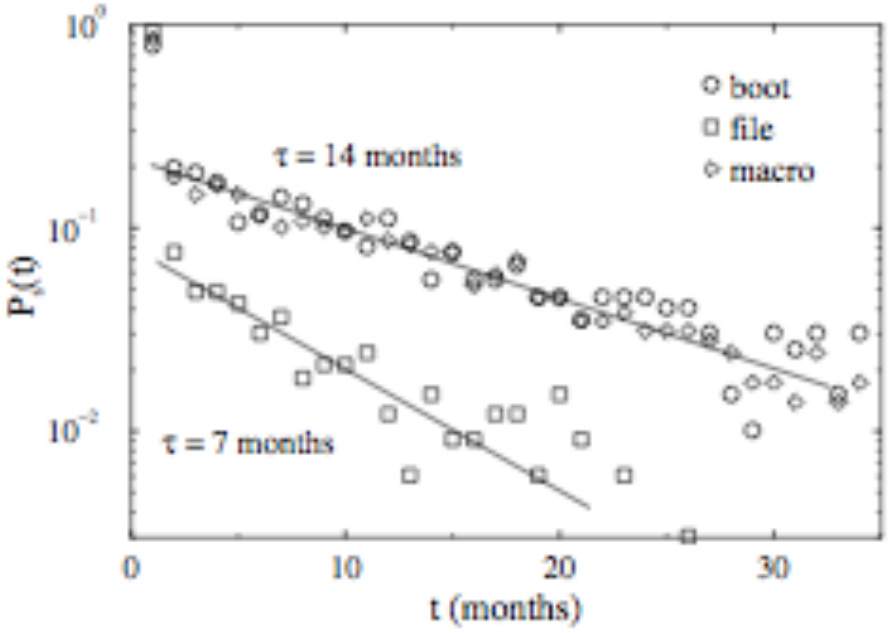
Relaxing Assumptions

- Homogeneous Mixing: a node connects to the same average number of other nodes as any other.
- Most real networks are not Erdos-Renyi random networks (for which the homogeneous mixing assumption holds).
- Most networks have heterogeneous degree distributions.
  - Scale free networks!

Would the Model Apply to SF?

- Pastor-Satorras and Vespignani [2001] have considered the life of computer viruses over time on the Internet:

Surviving probability of virus



Virus survived on average 6-9/14 months depending on type

UNIVERSITY OF CAMBRIDGE

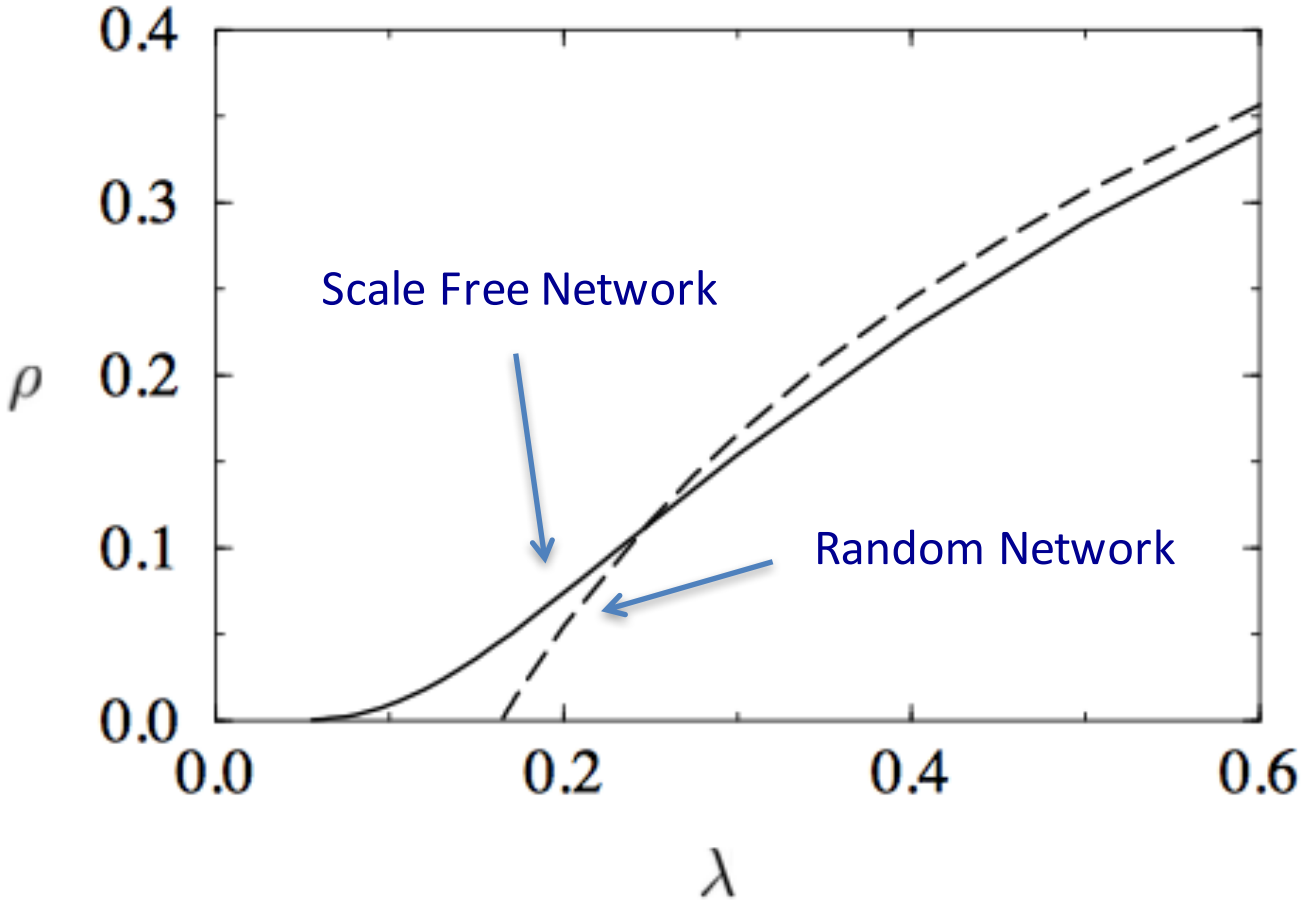
How to Justify This Survival Time?

- The virus survival time is considerably high with respect to the results of epidemic models of spreading/recovering:
  - Something wrong with the epidemic threshold!

- Experiment: SIS over a generated Scale Free network (exponent -3).

UNIVERSITY OF CAMBRIDGE

No Epidemic Threshold for SF!



UNIVERSITY OF CAMBRIDGE

Infections proliferate in SF networks independently of their spreading rates!

Network Immunisation

- Random network can be immunized with some sort of uniform immunization process [oblivious of the characteristics of nodes].
- **Random immunisation does not work in SF networks** no matter how many nodes are immunized [unless it is all of them].
- Targeted immunization in SF must be used instead.
  - Keeping into account degree!

UNIVERSITY OF CAMBRIDGE

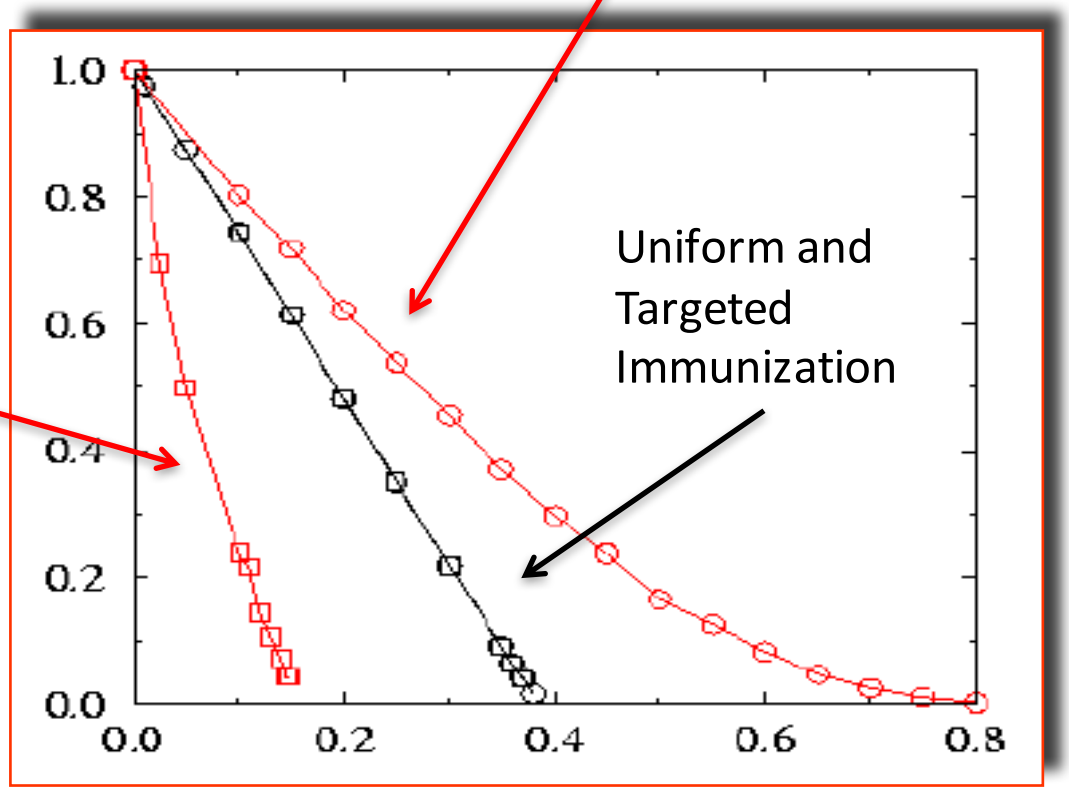
Immunization on SF Networks

- Red=SF
- Black= Random

Uniform Immunization

Targeted Immunization

Density of Infected



Fraction of immunized nodes

UNIVERSITY OF CAMBRIDGE

Local Immunization

- Global knowledge on the network structure is rarely available (more on this later)
- Local immunization strategy:
  - Select *g* nodes at random
  - Ask to each of them to pass over the vaccine to one of their neighbors
  - As a result, a node with degree k is immunized with a probability kP(k) (**hubs are immunized with higher probability**!)

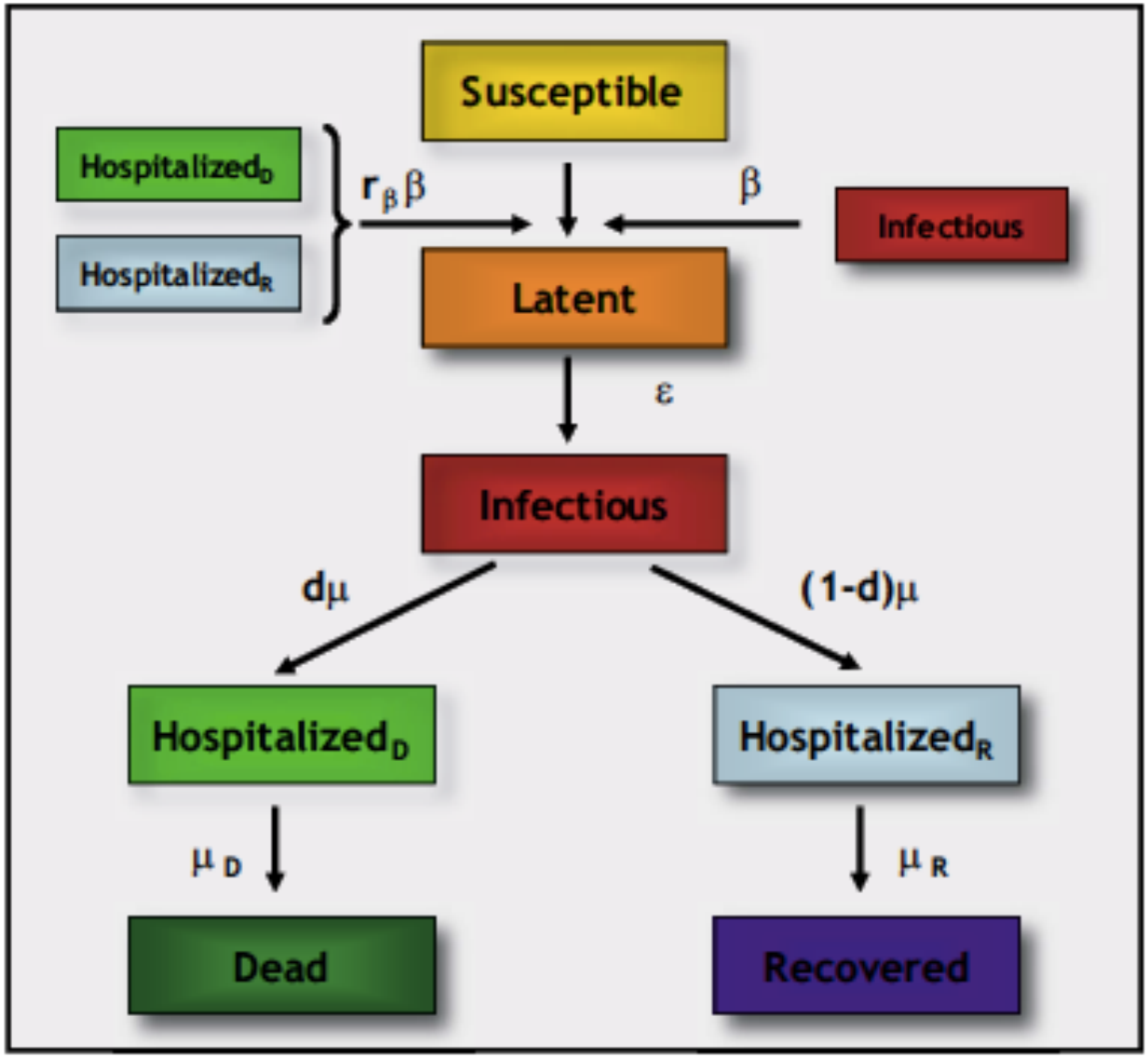
UNIVERSITY OF CAMBRIDGE

Modelling SARS Spreading

- SARS: severe acute respiratory syndrome
- SIR like model with more parameters and homogenous mixing
- Travel data and census data
- WHO data about SARS spreading to evaluate the model
- Outbreak starts in Hong Kong

UNIVERSITY OF CAMBRIDGE

Epidemics Flow



UNIVERSITY OF CAMBRIDGE

The Parameters of the Model

- Parameters used:

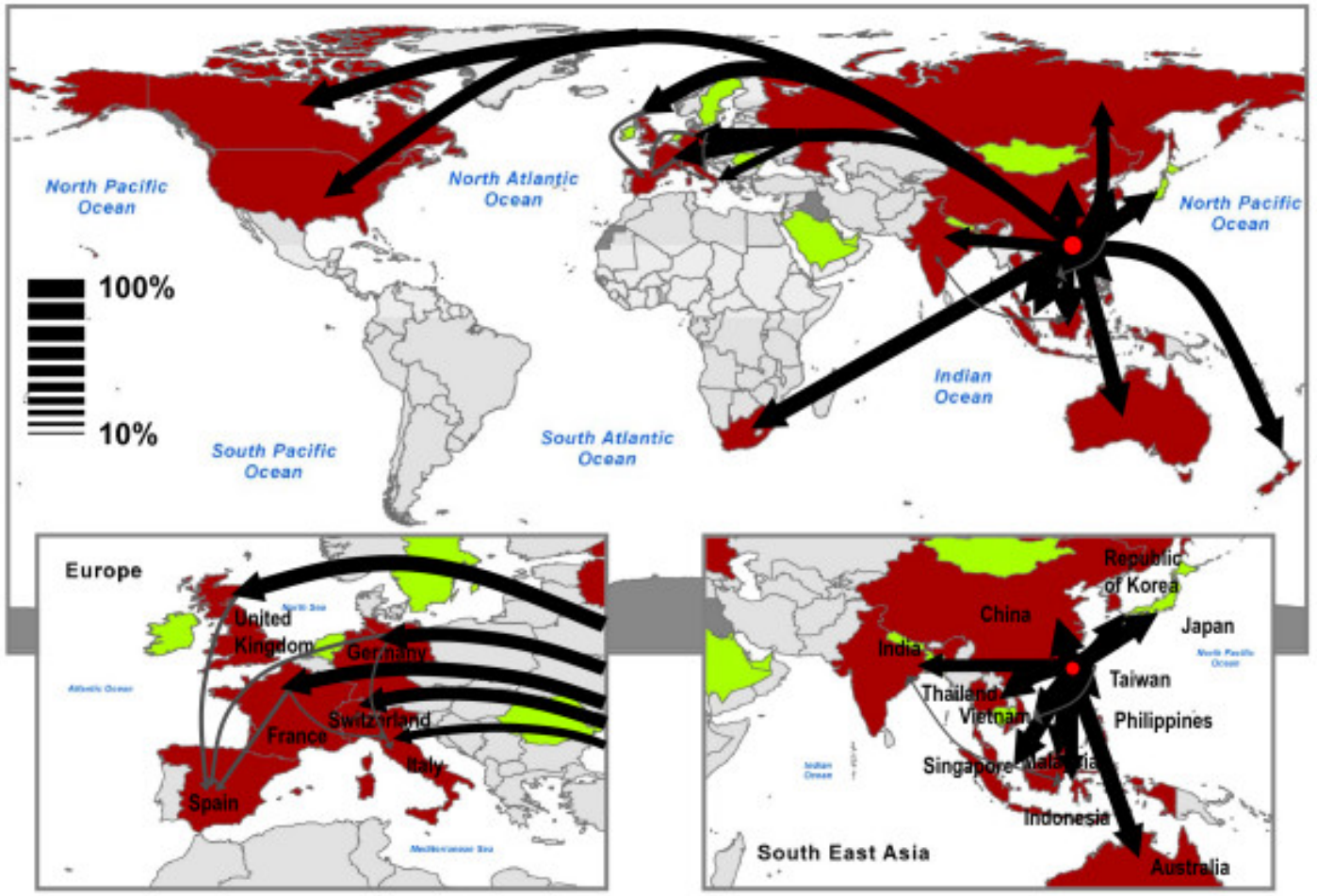
Parameter	Description		Baseline value
T_0	Initial offset from 21 February (days)		3*
β	Rate of transmission		0.57*
$L(t = 0)$	Number of initial latent individuals		10*
$s_f(t)$	Scaling factor for the rate of transmission	21 February + T_0-20 March	1.00
		21 March – 9 April	0.37
		10 April – 11 July	0.06
r_β	Relative infectiousness of patients at the hospital		0.2
ε^{-1}	Average latency period (days)		4.6
$\mu^{-1}(t)$	Average period from onset of symptoms to admission (days)	21 February + T_0-25 March	4.84
		25 March – 1 April	3.83
		2 April – 11 July	3.67
μ_R^{-1}	Average period from admission to recovery (days)		23.5
μ_D^{-1}	Average period from admission to death (days)		35.9
d	Case fatality rate		0.2

UNIVERSITY OF CAMBRIDGE

Adding Travel and Geography

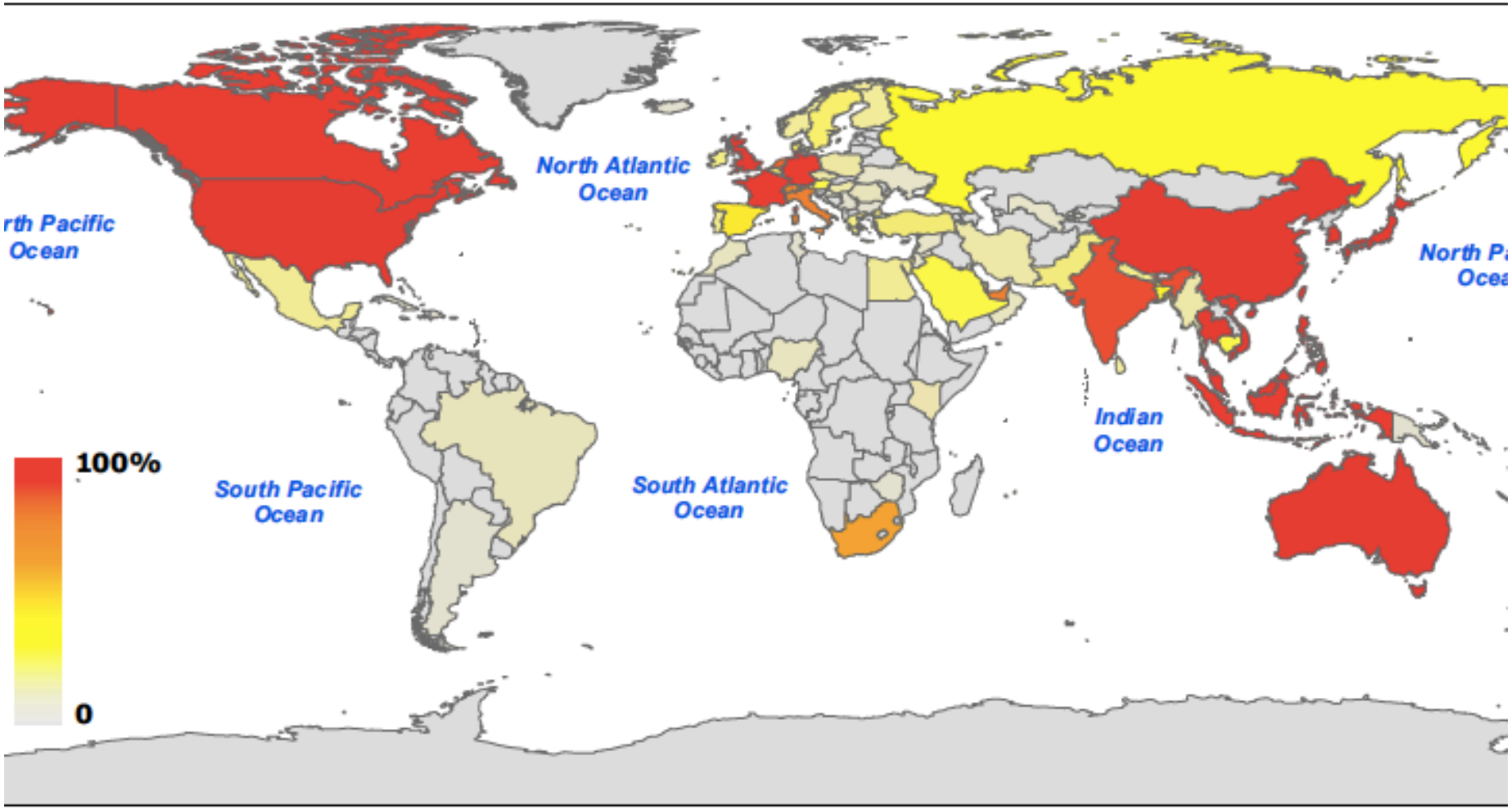
- A model per city ("meta-population model")
  - Each compartmental model describes the epidemics in a given city.
  - Models per city are coupled using information from air travel
    - Possible simplification of the reality but effective abstraction.
- Stochastic model
  - Probabilities of an individual of moving from one city to the next follows proportions of traffic observed in the air travel data.

Epidemic Pathways



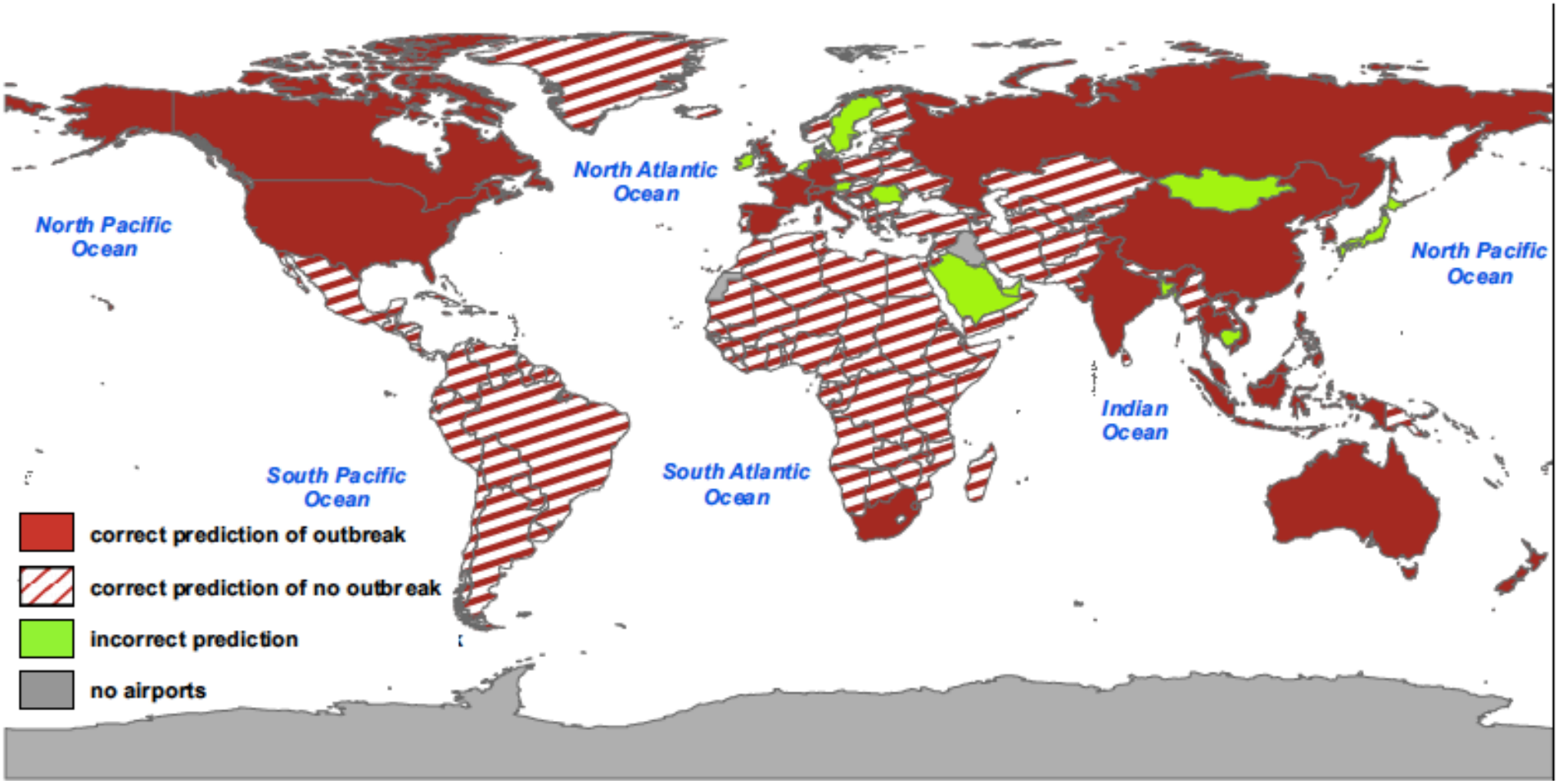
UNIVERSITY OF CAMBRIDGE

Predicted Outbreak Likelihood



UNIVERSITY OF CAMBRIDGE

Comparison with Data



UNIVERSITY OF CAMBRIDGE

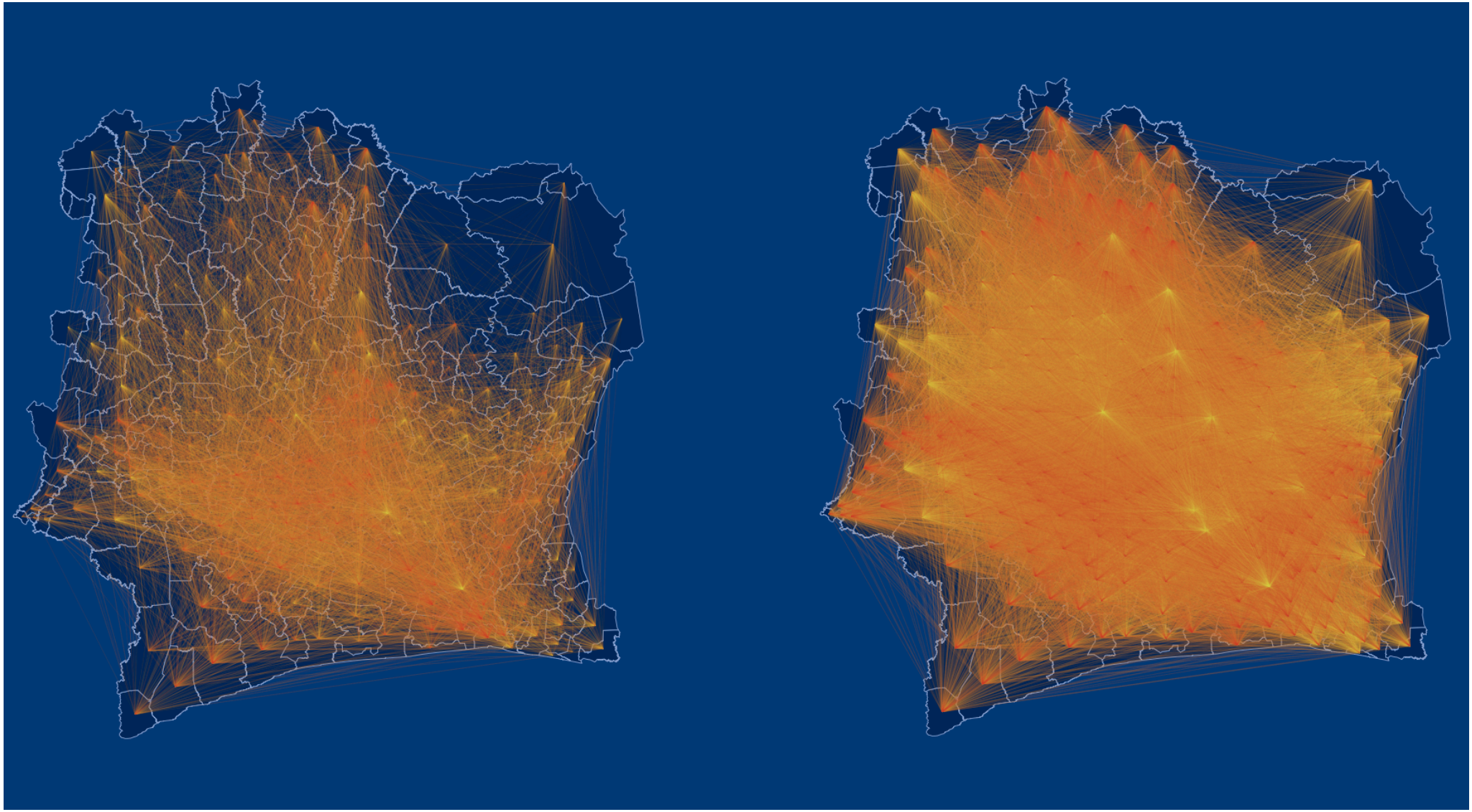
Epidemic Spreading Models and Real Data

- A key problem is how to extract information for modelling the spreading the disease from real data.
- One possibility is to use information coming from the cellular network:
  - Transitions between base-stations for modelling the mobility;
  - Phone calls graph for modelling the underlying social networks between callers and callees.

UNIVERSITY OF CAMBRIDGE

Epidemic Spreading Models and Real Data

- In order to study possible strategies of containment of epidemics, an important aspect is how to model information campaigns for example for mass vaccination.
- Vaccination can happen through the population by exploiting the “strong” ties between them (family ties or friendships, etc.)
- Presence of two concurrent processes:
  - Epidemic spreading
  - Information diffusion (spreading of “immunising information”)



Mobility

Calls

UNIVERSITY OF CAMBRIDGE

Data for Development Challenge

- Data mining competition using a data set containing movement and call data of 5 million individuals in Ivory Coast in order to help to address society development questions in novel way
- Information extracted from the Call Data Records (CDRs) of the Orange network in Ivory Coast

Massachusetts Institute of Technology

orange™

Call Data Records

- A Call Data Record (CDR) entry contains information about a specific phone call (usually called “metadata”), including:
  - The phone number of the subscriber originating the call;
  - The phone number of the called party;
  - The identification of the equipment writing the record (base station).
    - From this information it is possible to extract the geographic location of the caller.

UNIVERSITY OF CAMBRIDGE

Mobility Matrix

- Movement data extracted from the registration patterns to the cellular infrastructure (i.e., CDRs) are used to evaluate the influence of human mobility on the spreading of the disease in a given geographic area.
- From the data it is possible to extract the probability of transitions between different areas, in this case “sub-prefectures” (counties) of Ivory Coast.
- Using this information, we build a mobility matrix representing movement in the country as a Markov process.

UNIVERSITY OF CAMBRIDGE

Mobility Matrix

Number of times a user *u* moves from the sub-prefecture *i* to the sub-prefecture *j*

$$m_{ij} = \frac{\sum_u \mathcal{M}_{ij}^u}{\sum_k \sum_u \mathcal{M}_{ik}^u}$$

Probability that an individual moves from the sub-prefecture *i* to the sub-prefecture *j*

Number of times a user *u* moves from the sub-prefecture *i* to the sub-prefecture *k*

UNIVERSITY OF CAMBRIDGE

Call Graph

Number of phone calls initiated from the sub-prefecture *i* and directed to the sub-prefecture *j*

$$c_{ij} = \frac{C_{ij}}{\sum_k C_{ik}}$$

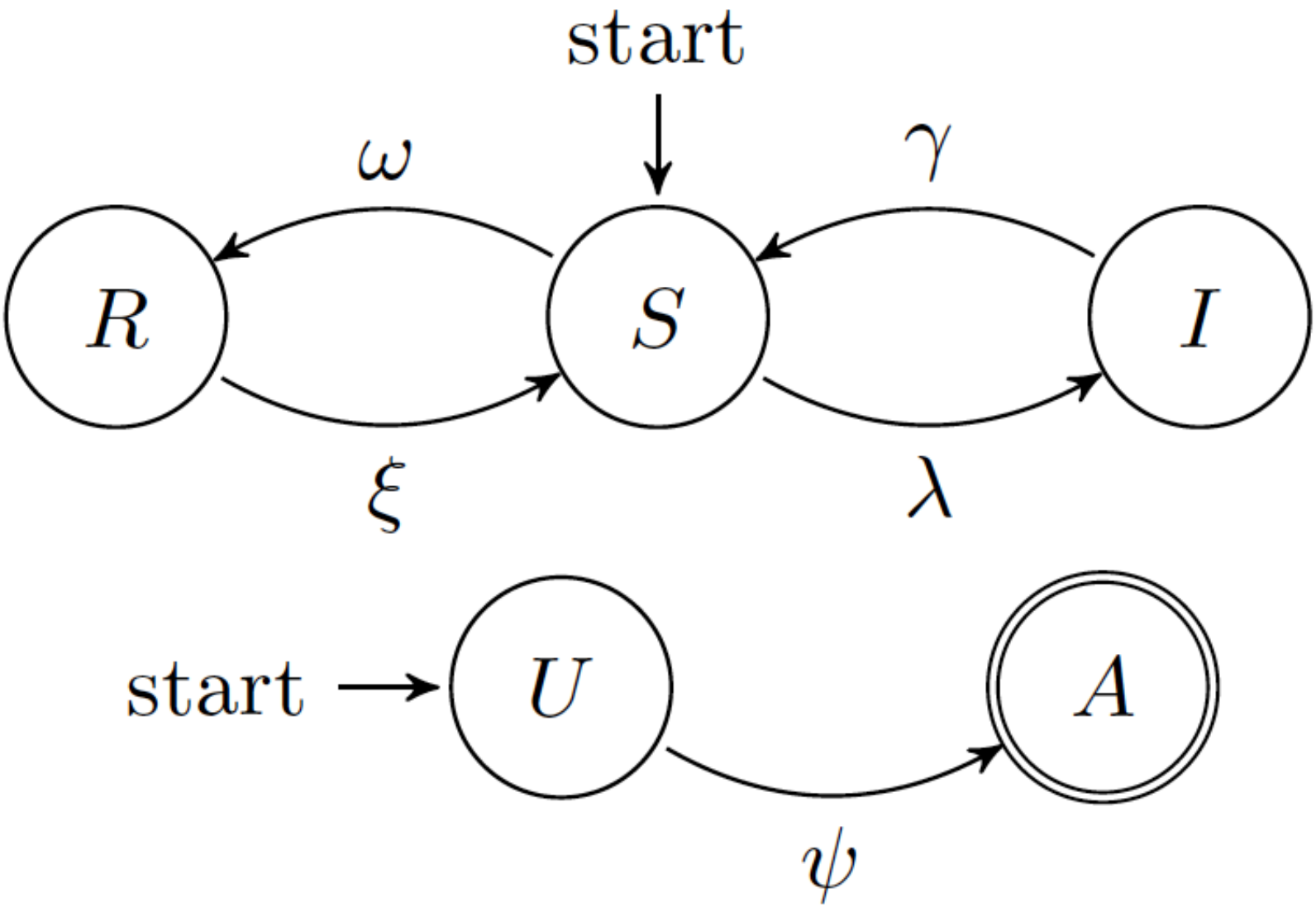
Probability of a call being established between sub-prefectures *i* and *j*

Number of phone calls initiated from the sub-prefecture *i* and directed to the sub-prefecture k

UNIVERSITY OF CAMBRIDGE

States

- Disease spreading process:
  - S: Susceptible
  - I: Infected
  - R: Resistant/recovered
- Information spreading process:
  - U: unaware
  - A: aware
- $N(t)=S(t)+I(t)+R(t)=U(t)+A(t)$


start
ω
γ
R
S
I
ξ
λ
start
U
A
ψ

$$
\begin{aligned}
I_i[t+1] &= \sum_{j=1}^{n} m_{ji} \left[I_j[t] \ + \lambda \frac{S_j[t]}{N_j[t]} I_j[t] - \gamma I_j[t] \right] \\
S_i[t+1] &= \sum_{j=1}^{n} m_{ji} \left[S_j[t] \ - \lambda \frac{S_j[t]}{N_j[t]} I_j[t] + \gamma I_j[t] + \xi R_j[t] + \right. \\
&\qquad \left. - \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \\
R_i[t+1] &= \sum_{j=1}^{n} m_{ji} \left[R_j[t] - \xi R_j[t] + \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \\
A_i[t+1] &= \sum_{j=1}^{n} m_{ji} \left[A_j[t] + \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \\
U_i[t+1] &= \sum_{j=1}^{n} m_{ji} \left[U_j[t] - \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \qquad (3)
\end{aligned}
$$

Mobility Matrix

$$I_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[I_j[t] + \lambda \frac{S_j[t]}{N_j[t]} I_j[t] - \gamma I_j[t] \right]$$

$$S_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[S_j[t] - \lambda \frac{S_j[t]}{N_j[t]} I_j[t] + \gamma I_j[t] + \xi R_j[t] + - \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

$$R_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[R_j[t] - \xi R_j[t] + \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

$$A_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[A_j[t] + \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

$$U_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[U_j[t] - \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \qquad (3)$$

Mobility Matrix

$$I_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[I_j[t] + \lambda \frac{S_j[t]}{N_j[t]} I_j[t] - \gamma I_j[t] \right]$$

$$S_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[S_j[t] - \lambda \frac{S_j[t]}{N_j[t]} I_j[t] + \gamma I_j[t] + \xi R_j[t] + \right.$$

Call Matrix

$$\left. - \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

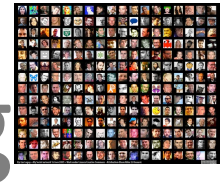
$$R_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[R_j[t] - \xi R_j[t] + \omega S_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

$$A_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[A_j[t] + \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right]$$

$$U_i[t+1] = \sum_{j=1}^{n} m_{ji} \left[U_j[t] - \psi U_j[t] \frac{\sum_{k=1}^{n} c_{kj} A_k[t]}{\sum_{k=1}^{n} c_{kj} N_k[t]} \right] \qquad (3)$$

UNIVERSITY OF CAMBRIDGE

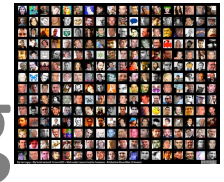
Real-time Predictive Modelling



- By using a model like this one, real-time predictive modelling might be possible
- Policy-makers might extract the parameters of the model, such as the mobility matrix, in real-time.
- Transmission models ha been widely used for assessing potential strategies for containing diseases such as influenza.
- “Mobile big data” might help in developing more accurate models.

UNIVERSITY OF CAMBRIDGE

Real-time Predictive Modelling



- One of the key problems is understanding the uncertainty associated to the model.
- It is difficult to understand the contribution of many factors such as:
  - Climatic factors
  - Transmission seasonality
  - Long-term immunity of a population
  - …
- Scenario-based modelling is routinely used in order to predict future evolution of epidemics.

UNIVERSITY OF CAMBRIDGE

Summary

- Epidemics are very complex processes.
- Existing models have been increasingly capable of capturing their essence.
- However there are still a number of open issues related to the modelling of real disease spreading or information dissemination.
- Mixing geographic factors with epidemics model is very relevant and important.
- The availability of “big data” might help in building more realistic and possibly real-time epidemic models.

UNIVERSITY OF CAMBRIDGE

References

- Chapter 21
- Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic Spreading in Scale-Free Networks. Phys. Rev. Lett.(86), n.14. Pages = 3200--3203. 2001.
- Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of Complex Networks. Physical Review E 65. 2002.
- Marcelo Kuperman and Guillermo Abramson. Small world effect in an epidemiological model. Physical Review Letters, 86(13):2909–2912, March 2001.
- Vittoria Colizza, Alain Barrat, Marc Barthélemy and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. BMC Medicine 2007, 5:34

References

- Antonio Lima, Manlio De Domenico, Veljko Pejovic and Mirco Musolesi. Exploiting Cellular Data for Disease Containment and Information Dissemination Campaigns Strategies in Country-wide Epidemics. In *Proceedings of NetMob'13*. Boston, MA, USA. May 2013.
- Marc Lipsitch, Lyn Finelli, Richard T. Heffernan, Gabriel M. Leung, and Stephen C. Redd. Improving the Evidence Base for Decision Making during a Pandemic: The Example of 2009 Influenza A/H1N1. In Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science. Volume 9, Number 2, 2011.
- Matt J. Keeling and Pejman Rohani. Modeling Infectious Diseases in Humans and Animals. Princeton University Press. 2007

UNIVERSITY OF CAMBRIDGE