# Discovering the mechanics behind heterogeneous stochastic diffusion phenomena


Yoshiharu Maeno, Ph.D.

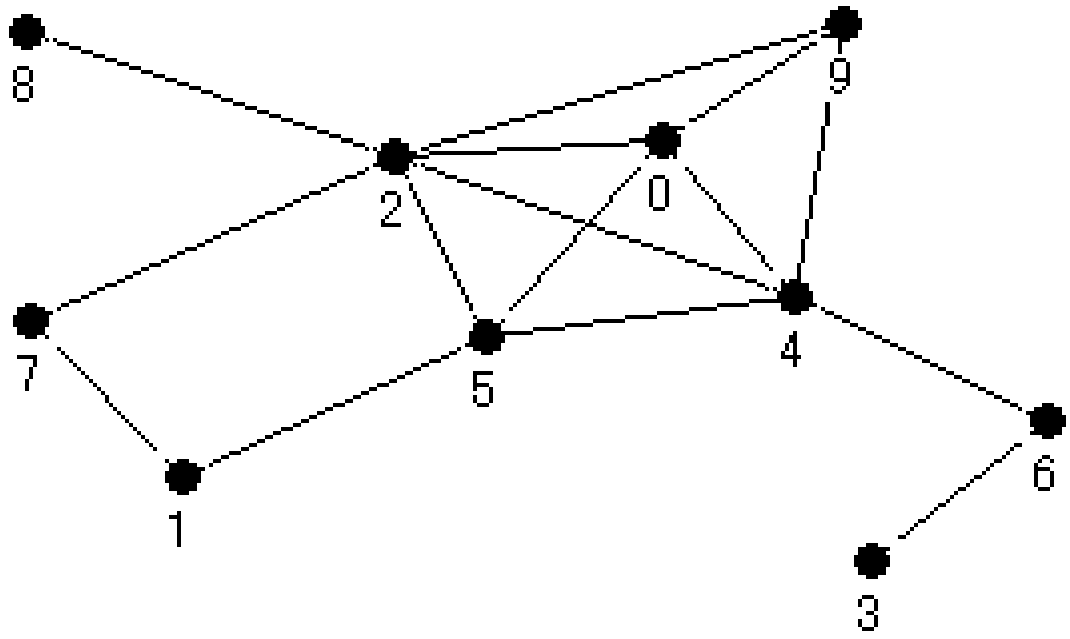
Founder management consultant and scientist at Social Design Group

Principal researcher at NEC Corporation

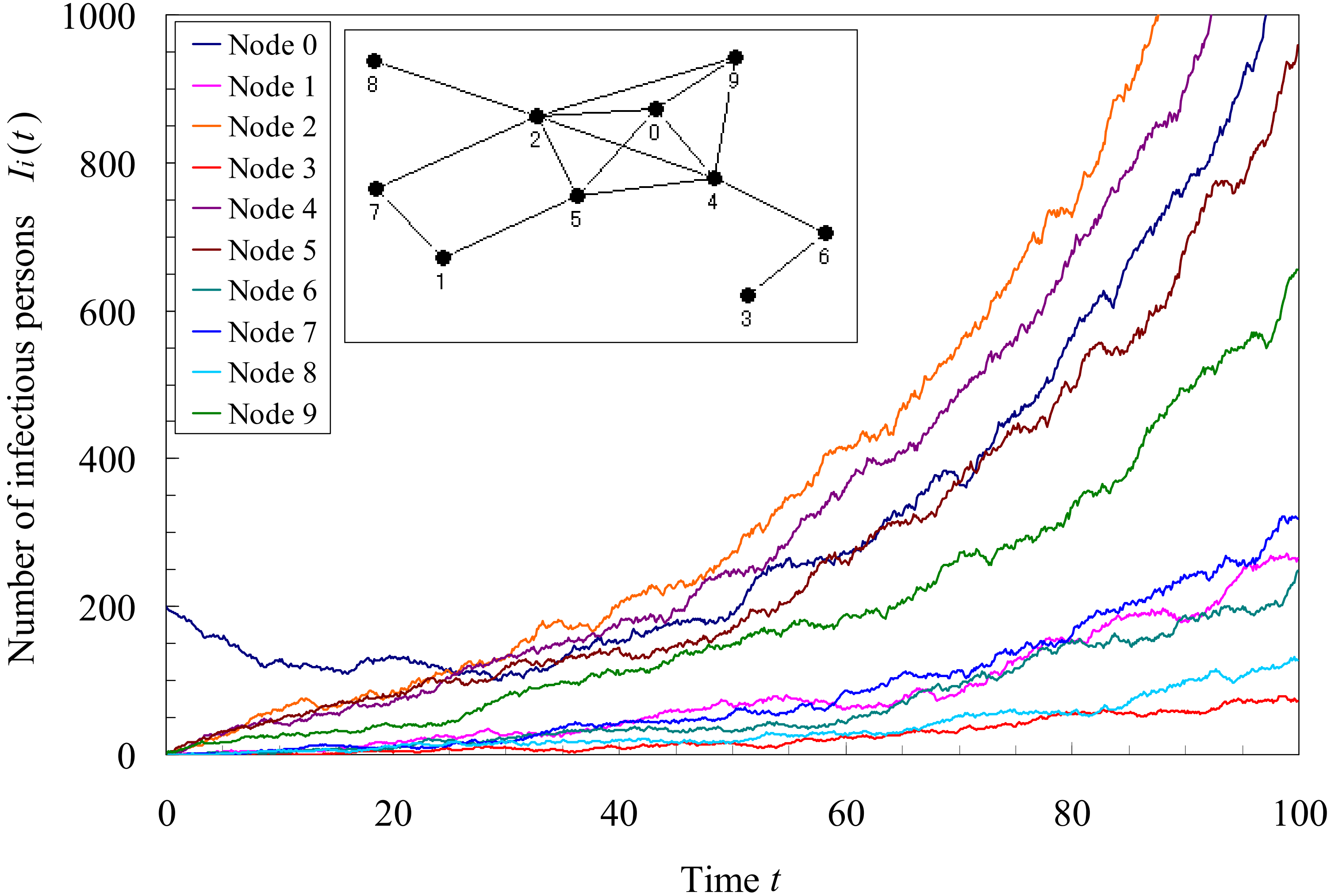
# Scope

- Diffusion phenomena arise from the stochastic spread over spatially heterogeneous media.
  - Spread of an infectious disease over a social network
  - Movement of persons over a transportation network
  - Diffusion of information over the Internet

# Example of a random netwok



# Spread of an infectious disease



# Inverse problem

**System characteristics**

- Stochasticity
- Spatial heterogeneousness

**Observation**

- Diffusion phenomena

Forward problem →

← Inverse problem

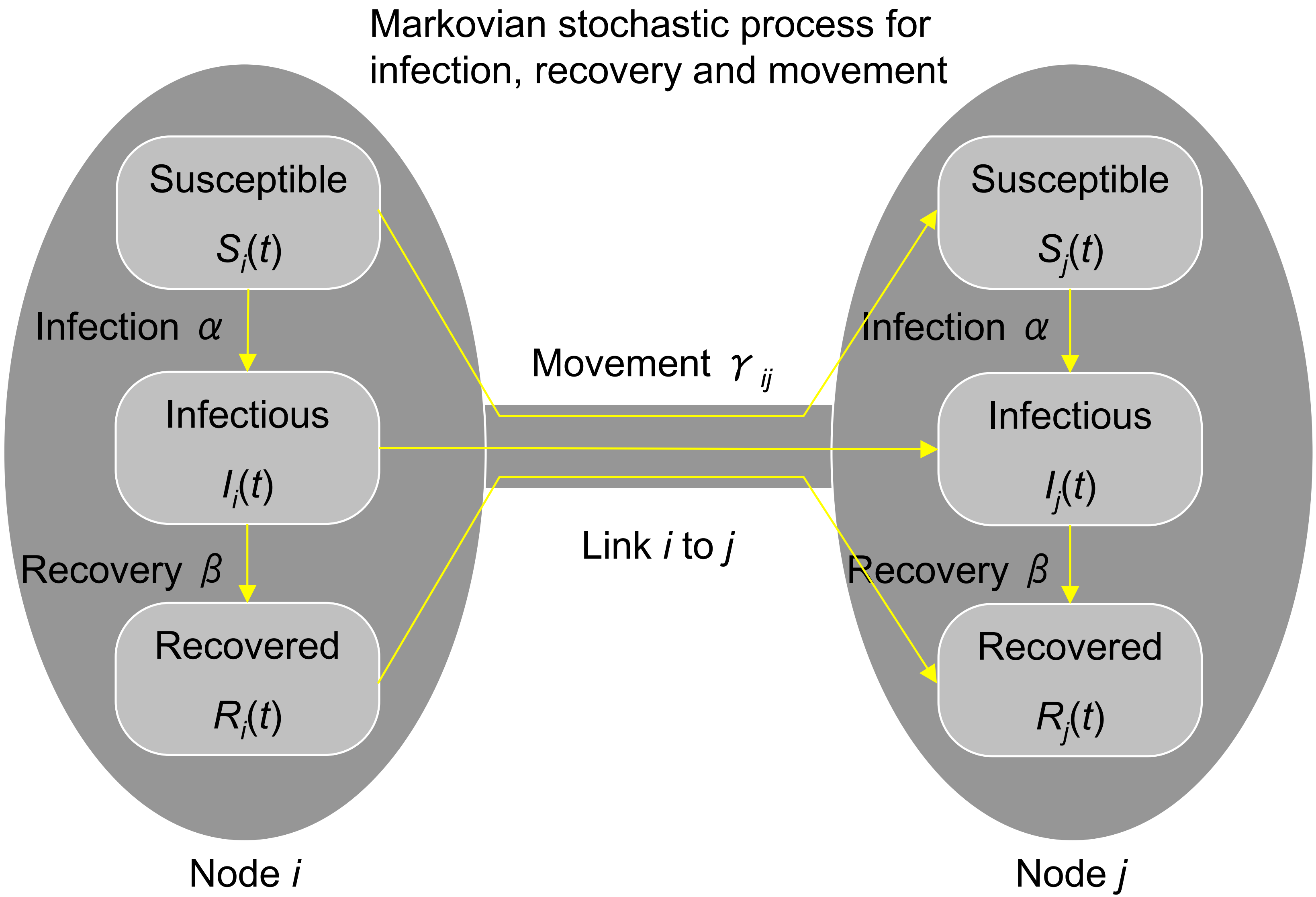
# Task 1: Profiling

- Profiling is solving an inverse problem in discovering the underlying network topology and revealing the transmission parameters from the observation on the diffusion governed by stochasticity and spatial heterogeneousness.

# Task 2: Node discovery

- Node discovery is identifying the nodes within a dataset of observation, diffusion over which is influenced by some unknown neighboring nodes.

# SIR model in a meta-population network



# Time evolution of $I_i(t)$

- SDE for the number of the infectious $I_i(t)$

$$\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} = \alpha \frac{S_i(t)}{S_i(t)+I_i(t)+R_i(t)} I_i(t) - \beta I_i(t) + \sum_{j \neq i} (\gamma_{ji} I_j(t) - \gamma_{ij} I_i(t)) + \Xi_i^{[\mathrm{I}]}(t).$$

Fluctuation for the combinatorial kinetics in chemical reactions

$$\Xi_i^{[\mathrm{I}]}(t) = \sqrt{\alpha \frac{S_i(t)}{S_i(t)+I_i(t)+R_i(t)} I_i(t)}\xi_i^{[\alpha]}(t) - \sqrt{\beta I_i(t)}\xi_i^{[\beta]}(t) + \sum_{j \neq i} \sqrt{\gamma_{ji} I_j(t)}\xi_{ji}^{[\gamma]}(t) - \sum_{j \neq i} \sqrt{\gamma_{ij} I_i(t)}\xi_{ij}^{[\gamma]}(t).$$

Statistical characteristics of the fluctuation

$$\langle \xi_{i'}(t) \rangle = 0.$$

$$\langle \xi_{i'}(t) \xi_{i''}(t') \rangle = \delta_{i'i''}\delta(t-t').$$

$$\langle \xi_{i'}(t) \xi_{i''}(t') \xi_{i'''}(t'') \cdots \rangle = 0.$$

# Time evolution of $J_i(t)$

- SDE for the cumulative number of new cases $J_i(t)$

$$\frac{\mathrm{d}J_i(t)}{\mathrm{d}t} = \alpha \frac{S_i(t)}{S_i(t)+I_i(t)+R_i(t)} I_i(t) + \Xi_i^{[\mathrm{J}]}(t).$$

Fluctuation for the combinatorial kinetics in chemical reactions

$$\Xi_i^{[\mathrm{J}]}(t) = \sqrt{\alpha \frac{S_i(t)}{S_i(t)+I_i(t)+R_i(t)} I_i(t)} \xi_i^{[\alpha]}(t).$$

# Maximal likelihood estimator (1)

1. Conversion from SDE to the time evolution of PDF

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = \mu_i(x_0(t), \cdots, x_{N-1}(t)) + \sum_{a=0}^{M-1} \sigma_{ia}(x_0(t), \cdots, x_{N-1}(t))\xi_a(t).$$

Fokker-Planck equation for PDF

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_{i=0}^{N-1} \frac{\partial}{\partial x_i} A_i(\boldsymbol{x}) p(\boldsymbol{x},t) + \frac{1}{2}\sum_{i,j=0}^{N-1} \frac{\partial^2}{\partial x_i \partial x_j} B_{ij}(\boldsymbol{x}) p(\boldsymbol{x},t).$$

where

$$A_i(\boldsymbol{x}) = \mu_i(\boldsymbol{x}). \quad B_{ij}(\boldsymbol{x}) = \sum_{a=0}^{M-1} \sigma_{ia}(\boldsymbol{x})\sigma_{ja}(\boldsymbol{x}).$$

Time evolution of the moments

$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = \langle A_i(\boldsymbol{x})\rangle_t.$$

$$\frac{\mathrm{d}v_{ij}(t)}{\mathrm{d}t} = \langle B_{ij}(\boldsymbol{x})\rangle_t + \langle x_i A_j(\boldsymbol{x})\rangle_t + \langle A_i(\boldsymbol{x}) x_j\rangle_t.$$

# Maximal likelihood estimator (2)

2. Approximation of PDF by the multivariate Gaussian with $m_i(t)$ and $v_{ij}(t)$

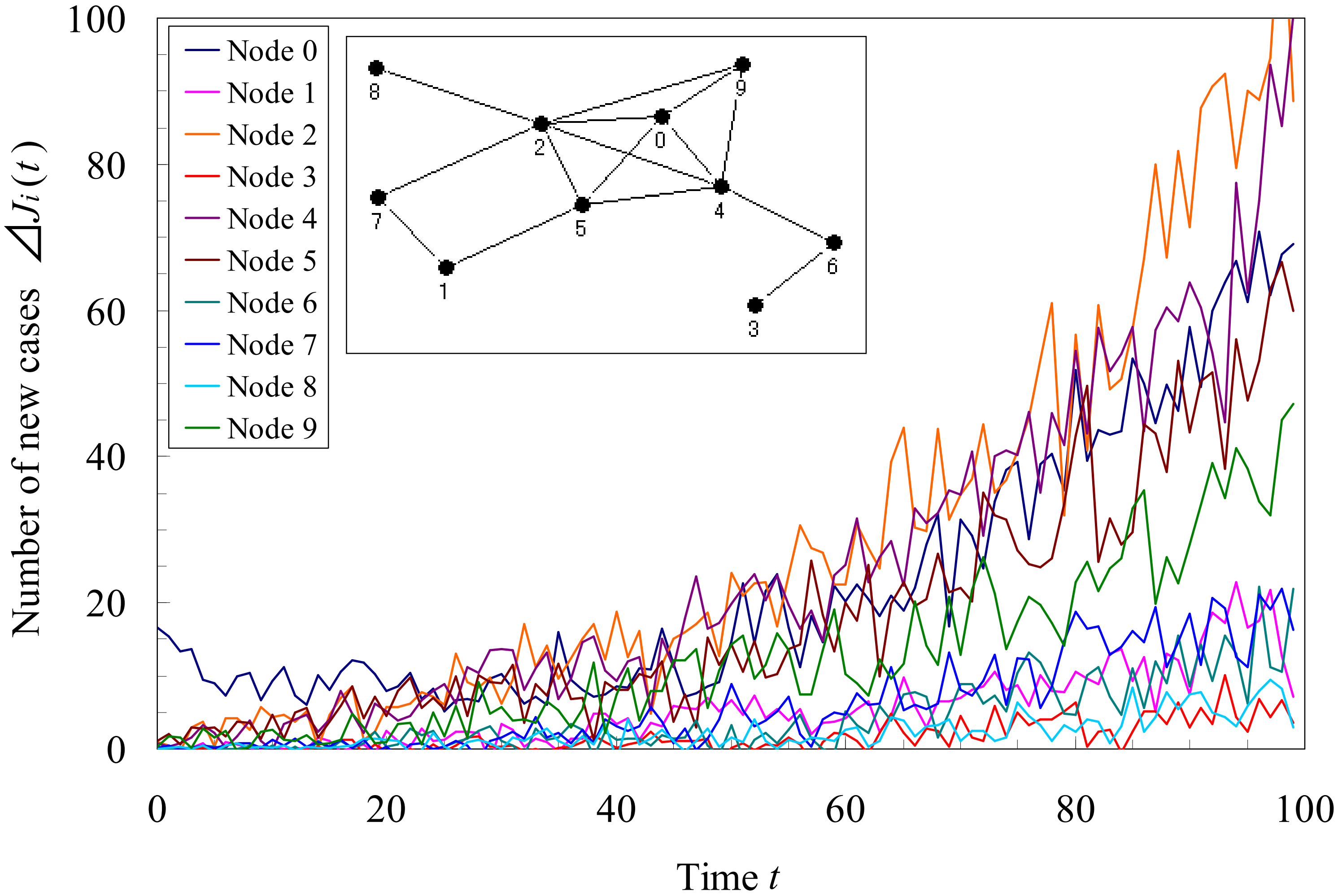
$$p(\boldsymbol{I}, t_{d+1}|\boldsymbol{\theta}) = \frac{\exp(-\frac{1}{2}(\boldsymbol{I} - \boldsymbol{m}(t_{d+1}|\boldsymbol{\theta}))\boldsymbol{v}(t_{d+1}|\boldsymbol{\theta})^{-1}(\boldsymbol{I} - \boldsymbol{m}(t_{d+1}|\boldsymbol{\theta}))^{\mathrm{T}})}{\sqrt{(2\pi)^N \det \boldsymbol{v}(t_{d+1}|\boldsymbol{\theta})}}.$$

3. Likelihood function for a given dataset

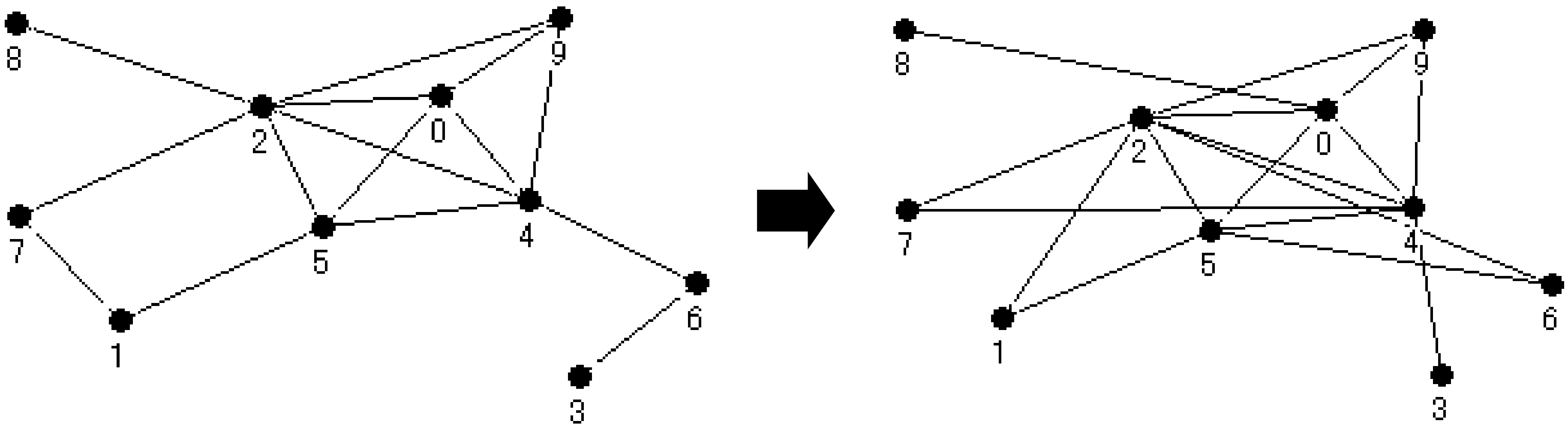
$$L(\boldsymbol{\theta}) = \sum_{d=0}^{D-2} \log p(\boldsymbol{I}(t_{d+1}), t_{d+1}|\boldsymbol{\theta}).$$

4. Maximal likelihood estimator of the transmission parameters and network topology

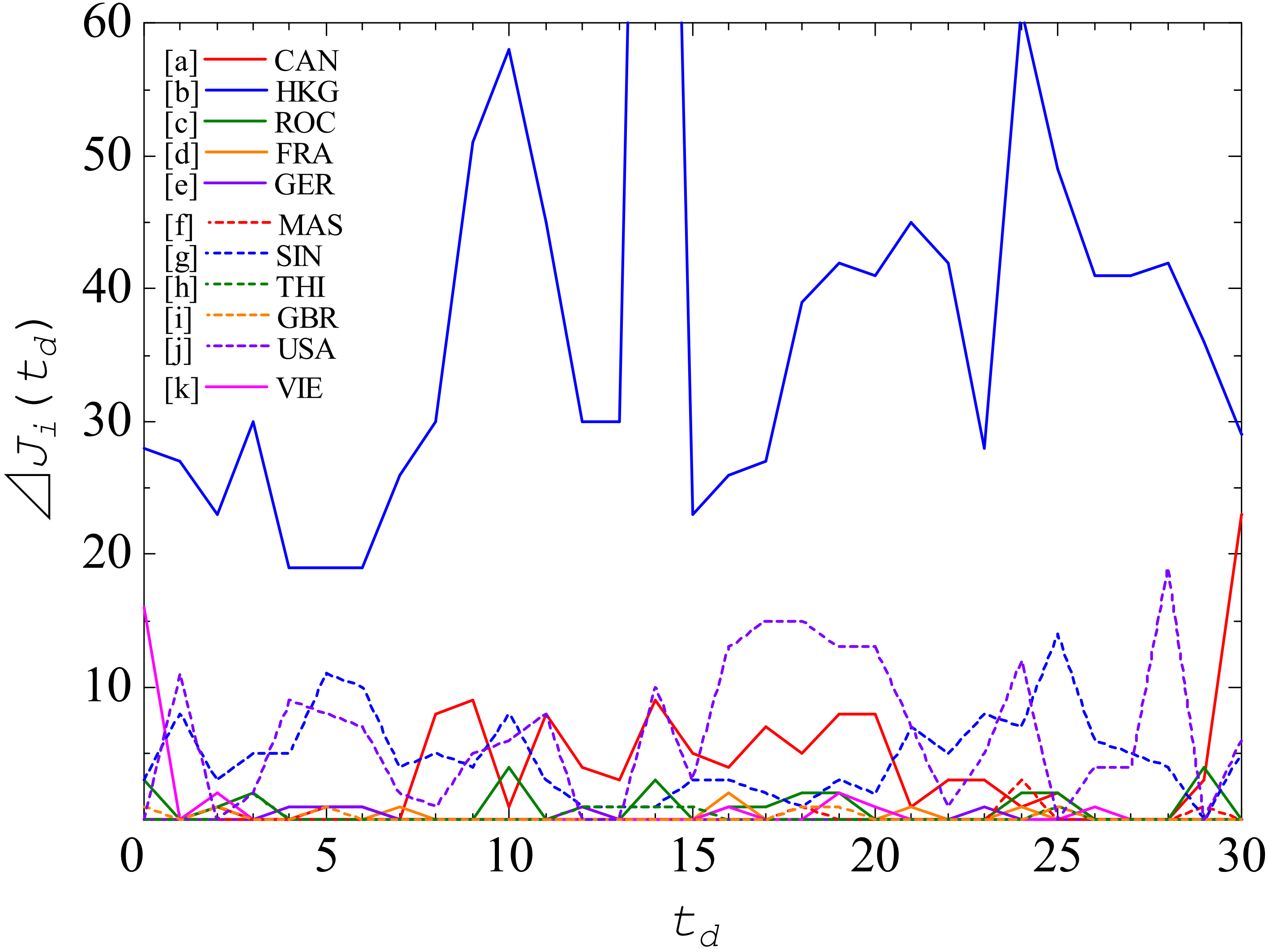
# Number of new cases



# Estimated topology

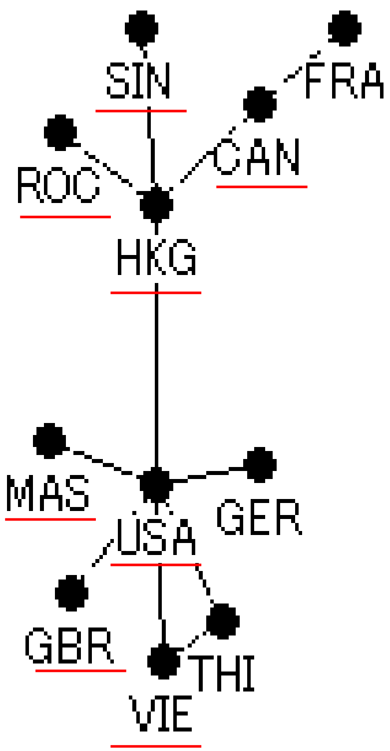


# WHO SARS dataset

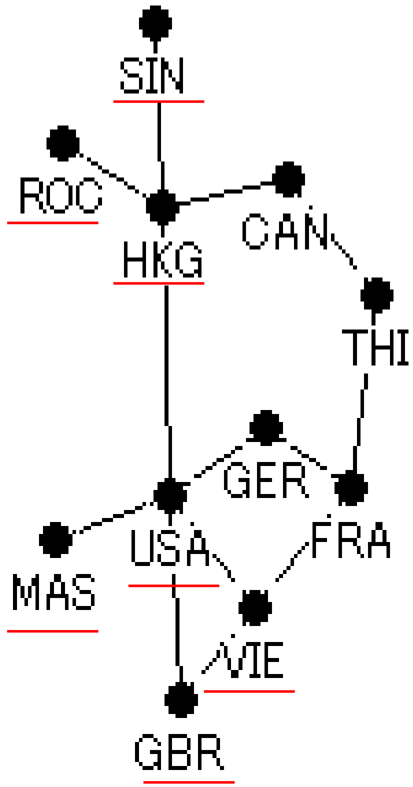


# Discovered topology

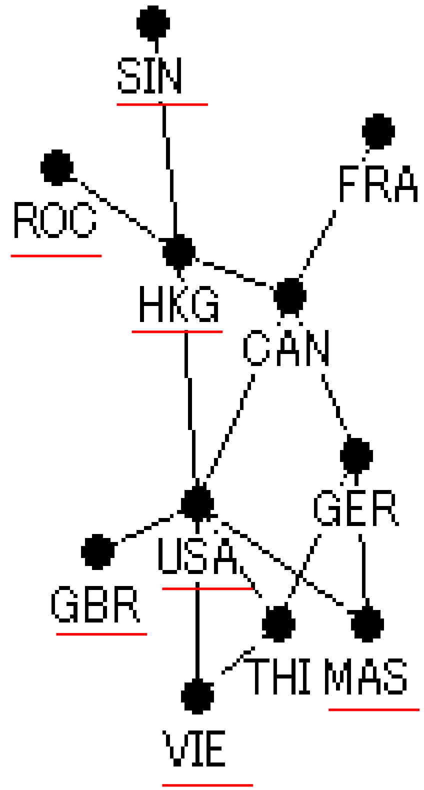
[A] most likely



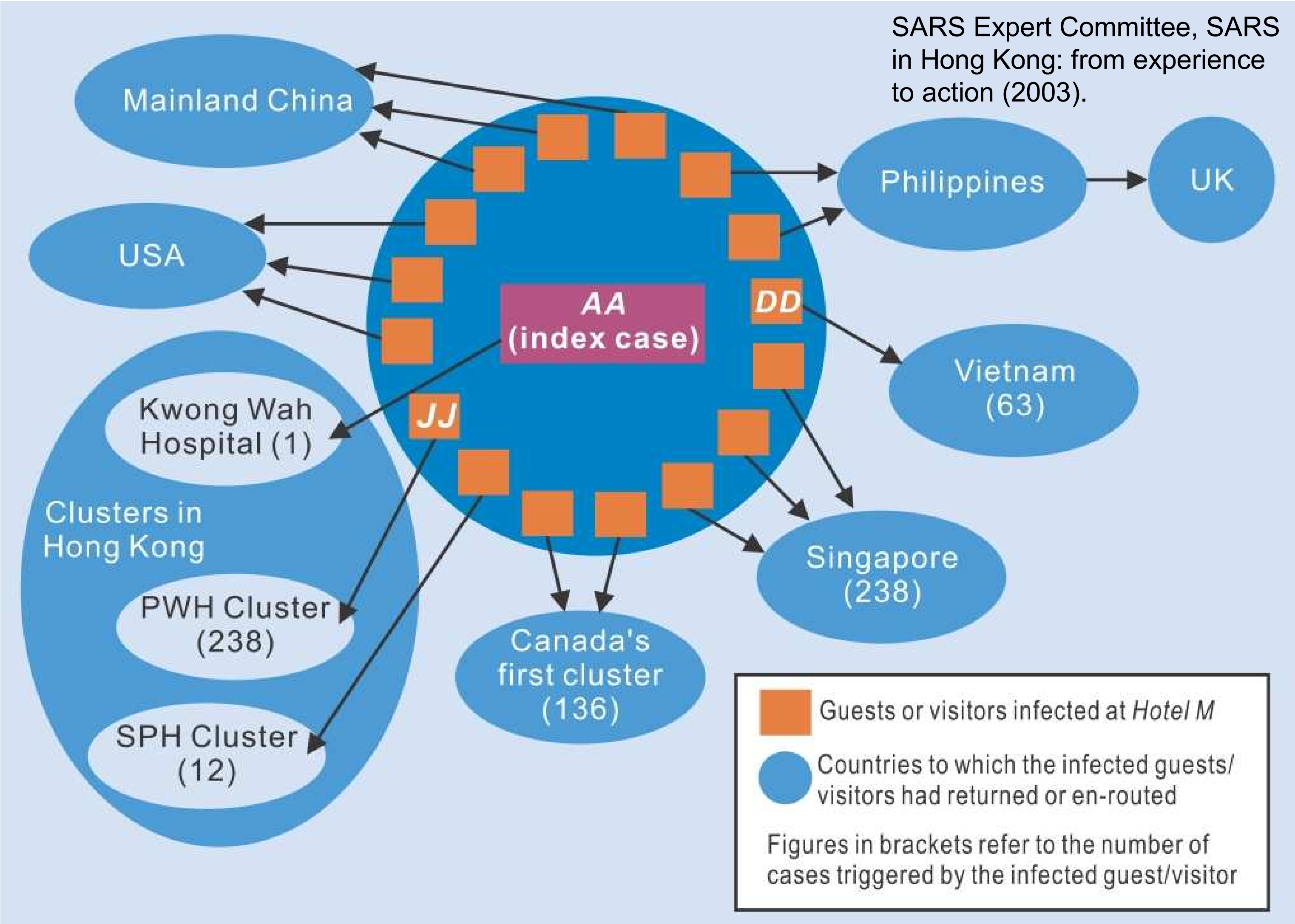

[B]




[C]



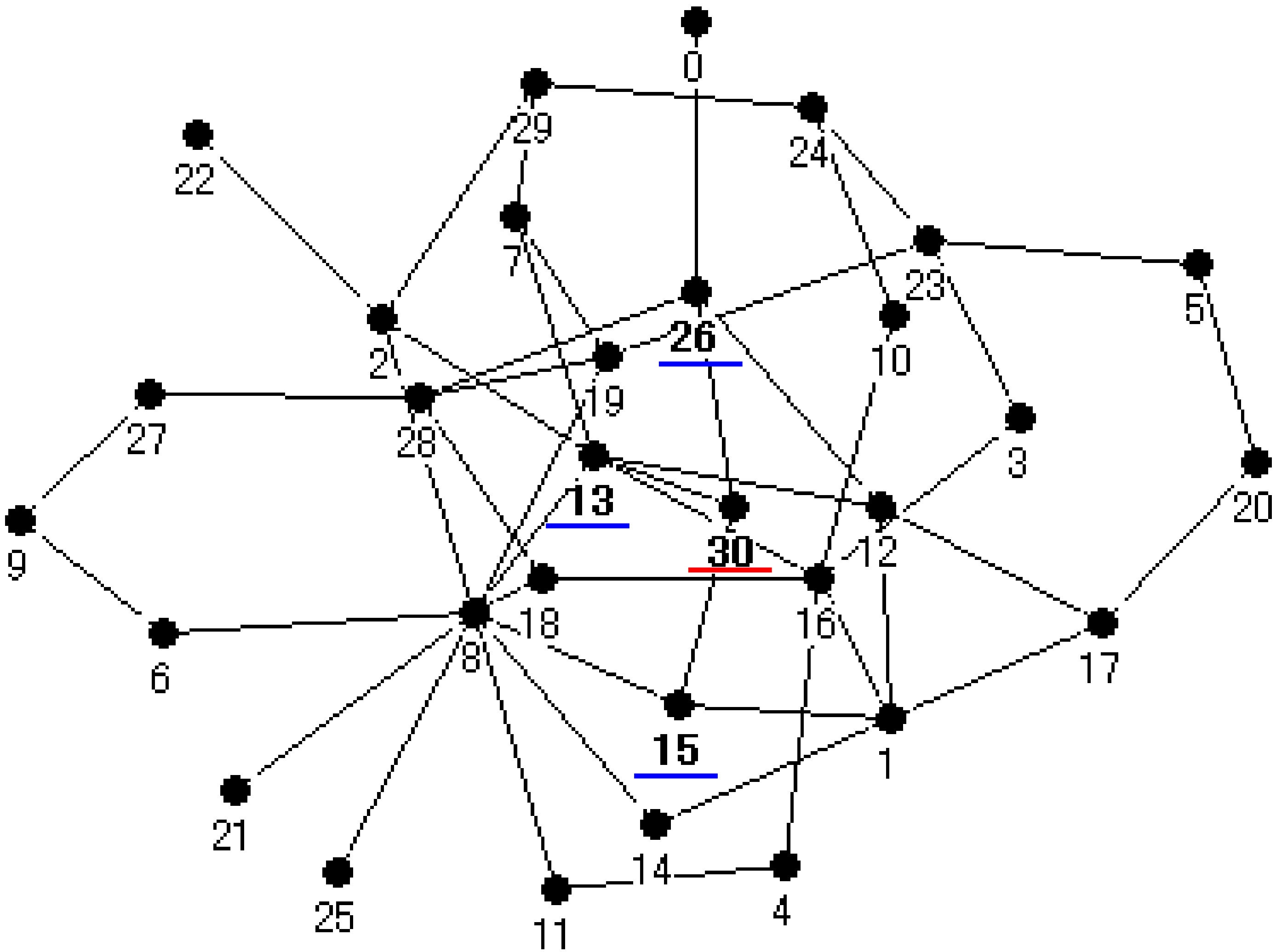
# Real series of events (1)



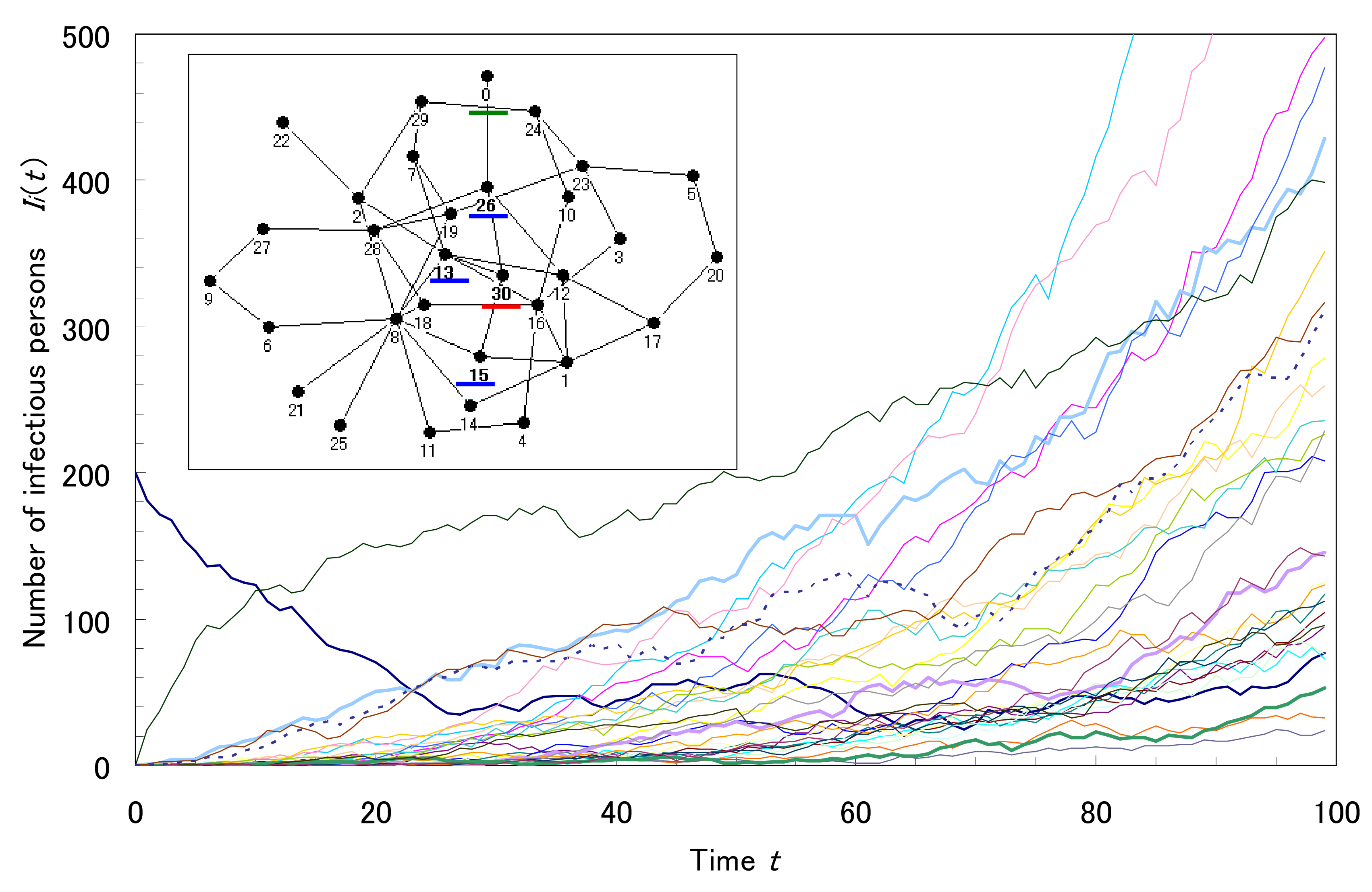
# Real series of events (2)

- A garment manufacturer from US became infected during the stay in HKG, showed symptoms in VIE, and was evacuated to a hospital in HKG.
- An Italian physician, who treated him at a hospital in VIE, showed symptoms in THI where he would attend a conference.
- Interactions are potentially present among HKG, USA, VIE, and THI which result in a chain of transmission.

# Unknown neighboring node



# Dataset



500
400
300
200
100
0
Number of infectious persons $I_i(t)$
0
20
40
60
80
100
Time $t$
0
29
24
22
7
23
5
2
26
10
27
19
28
3
13
30
12
20
9
18
16
17
8
6
15
1
21
14
4
25
11

# Extreme time sequence detection

1. Profiling of a network from a given dataset
2. Detection of the extreme time sequence $I_i(t)$ from a node $i$ which is not likely to realize when the sequences from the other nodes realize in the dataset

# Extremeness

- Conditional PDF to obtain $I_i(t_{d+1})$ under the maximal likelihood estimators in profiling

$$p(I_i, t_{d+1} | \boldsymbol{I}_{\bar{i}}, \hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{2\pi v_{ii}(t_{d+1}|\hat{\boldsymbol{\theta}})^*}} \exp\left(-\frac{(I_i - m_i(t_{d+1}|\hat{\boldsymbol{\theta}})^*)^2}{2v_{ii}(t_{d+1}|\hat{\boldsymbol{\theta}})^*}\right).$$

where

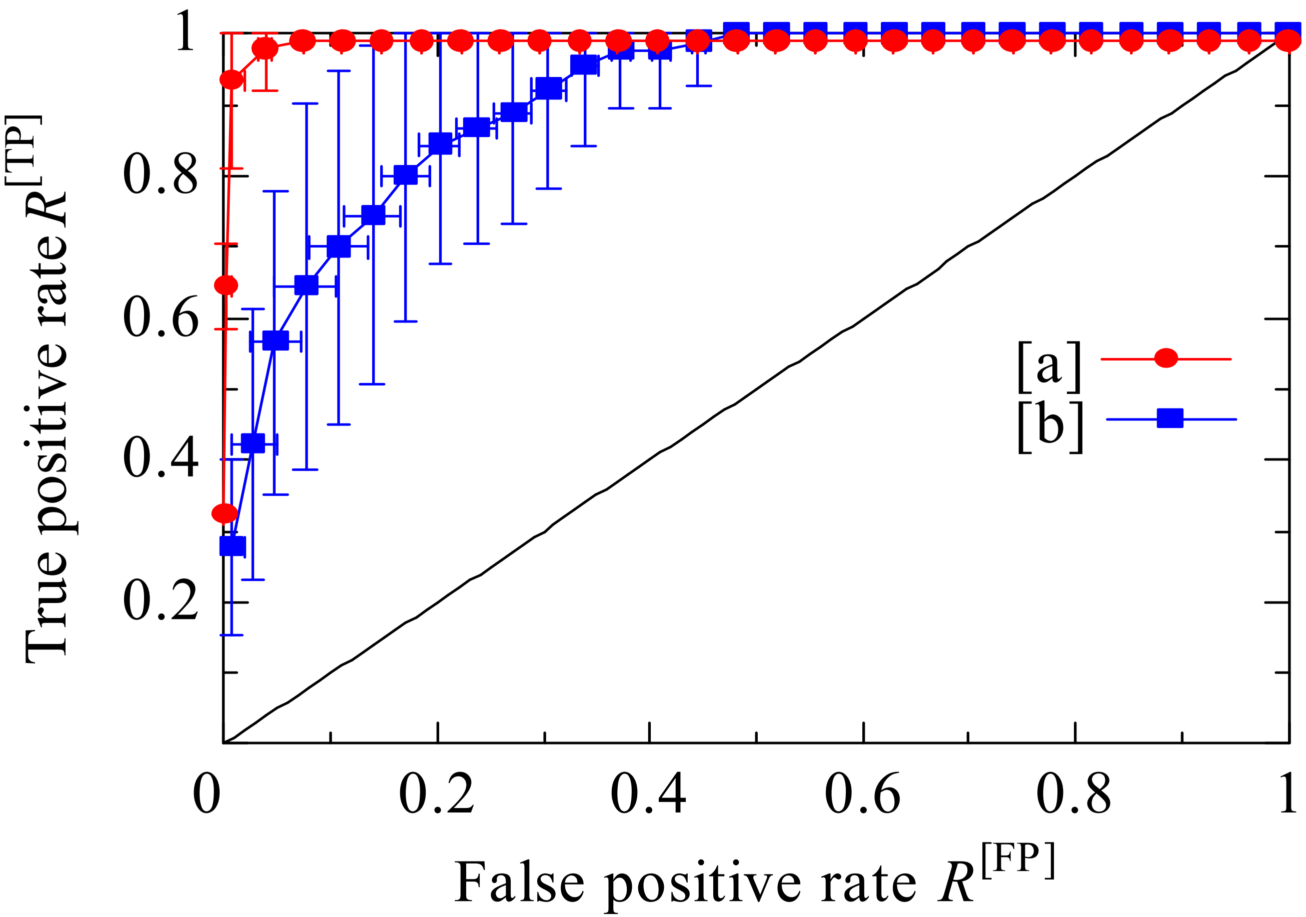
$$m_i(t_{d+1}|\hat{\boldsymbol{\theta}})^* = m_i(t_{d+1}|\hat{\boldsymbol{\theta}}) + \boldsymbol{v}_{i\bar{i}}(t_{d+1}|\hat{\boldsymbol{\theta}})\boldsymbol{v}_{\bar{i}\bar{i}}(t_{d+1}|\hat{\boldsymbol{\theta}})^{-1}(\boldsymbol{I}_{\bar{i}} - \boldsymbol{m}_{\bar{i}}(t_{d+1}|\hat{\boldsymbol{\theta}}))^{\mathrm{T}}.$$

$$v_{ii}(t_{d+1}|\hat{\boldsymbol{\theta}})^* = v_{ii}(t_{d+1}|\hat{\boldsymbol{\theta}}) - \boldsymbol{v}_{i\bar{i}}(t_{d+1}|\hat{\boldsymbol{\theta}})\boldsymbol{v}_{\bar{i}\bar{i}}(t_{d+1}|\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{v}_{\bar{i}i}(t_{d+1}|\hat{\boldsymbol{\theta}}).$$

- Identification of the most extreme time sequence

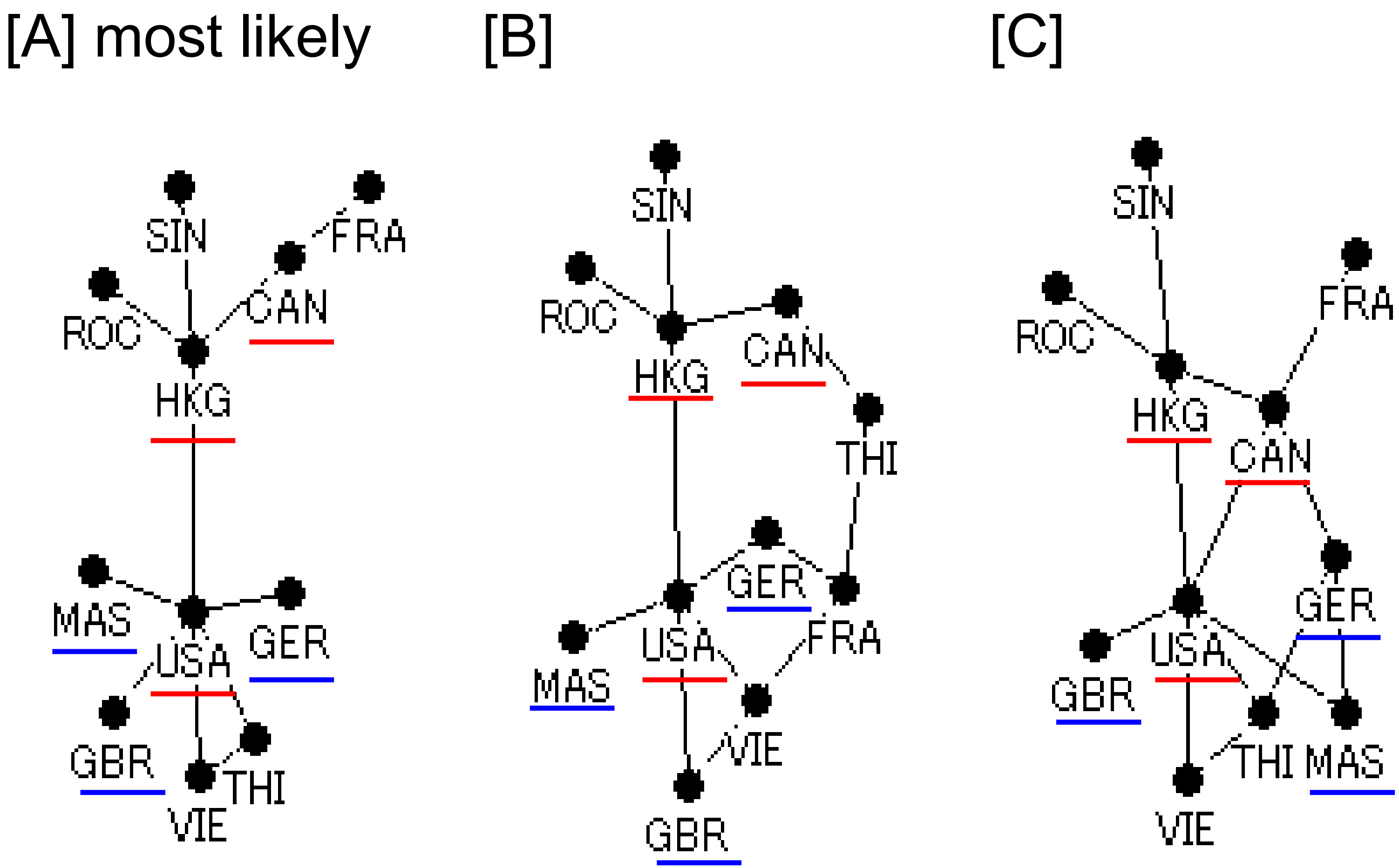
$$i_{\mathrm{e}} = \arg\min_i \prod_{d=0}^{D-2} p(I_i, t_{d+1} | \boldsymbol{I}_{\bar{i}}, \hat{\boldsymbol{\theta}}).$$

# Receiver operating characteristics



# Node discovery in SARS epidemic

1. HKG
2. USA
3. CAN
4. ROC
5. SIN
6. THI
7. VIE
8. FRA
9. MAS
10. GER
11. GBR



# Conclusion

- Profiling is solving an inverse problem in discovering the topology and revealing the parameters from the observation on stochastic heterogeneous diffusion.
- Node discovery is identifying the nodes, diffusion over which is influenced by unknown neighboring nodes.
- Some characteristics of the global transportation network over which SARS coronavirus spread in 2003 can be reproduced from a seemingly very noisy dataset.

# Reference


- Y. Maeno: Profiling of a network behind an infectious disease outbreak, *e-print* http://arxiv.org/abs/0905.3582.
- Y. Maeno: Node discovery in meta-population network behind infectious disease outbreak, *e-print* http://arxiv.org/abs/1006.2322.

# Related works

- How does the heterogeneousness govern the spread of a disease?
  - V. Colizza *et al*., Invasion threshold in heterogeneous meta-population networks, Phys. Rev. Lett. **99**, 148701 (2007).
- How does the stochasticity govern the spread of a disease?
  - C. E. Dangerfeld *et al*., Integrating stochasticity and network structure into an epidemic model, J. R. Soc. Interface, doi:10.1098/rsif.2008.0410 (2009).
- How accurately can the transmission of SARS be reproduced?
  - L. Hufnagel *et al*., Forecast and control of epidemics in a globalized world, Proc. Nat'l Acad. Sci. USA **101**, 15124 (2004).
- How accurately can a communication network topology be inferred?
  - M. G. Rabbat *et al*., Network Inference from co-occurrences, IEEE Trans. Info. Theory **54**, 4053 (2008).
  - Y. Maeno, Node discovery problem for a social network, Connections **29**, 62 (2009).