Epidemics on Networks

Piet Van Mieghem

in collaboration with Eric Cator, Ruud van de Bovenkamp, Cong Li, Stojan Trajanovski, Dongchao Guo, Zhidong He, Qiang Liu, Annalisa Socievole and Huijuan Wang

Leiden Complex Networks Network (LCN2) Seminar
Universiteit Leiden, 28 Oktober 2016

TU Delft

Motivation for virus spread in networks

- **Digital world:**
  - Information spread in on-line social networks
  - security threats to Internet (Code Red worm, Conficker, …: several billion $ in damage)
- **Real world: Biological epidemics** (e.g. Mexican flue, Ebola, Zika)

- **Understanding the virus spread**: first step in prevention
  - *Time*: How fast to disinfect nodes so that the virus dies quickly? Which nodes?
  - *Amount*: How many nodes will eventually be infected?

History

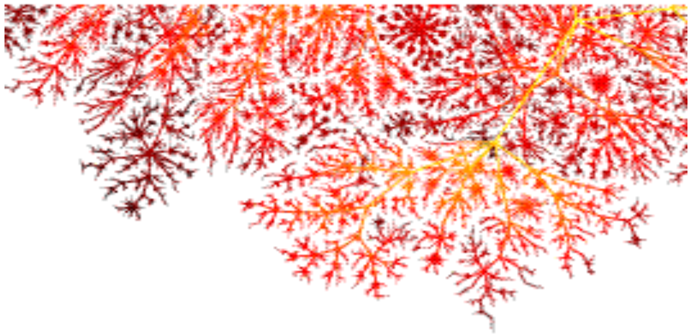
First epidemic model (Daniel Bernoulli,1760)

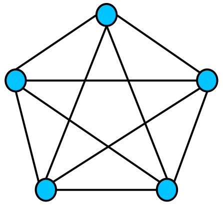
Homogeneous mixing model (biology)

Contact model (Harris, 1978): lattice

Epidemic on K_N (Kephart and White, 1991): computer viruses

Epidemics on graphs (Pastor-Satorras and Vespignani, 2001): birth of network epidemics





D. J. Daley and J. Gani, "Epidemic modelling – an introduction"

Local rule – Global emergent property models on networks

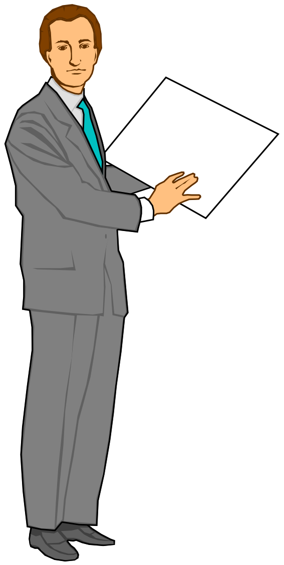
- Opinion models
- Synchronization
- Automata
- Ising-Spin model
- Sandpile models
- …

- **Epidemics on networks**

Many feature a phase transition

All crucially depend on the graph

Outline



Exact SIS model

NIMFA: N-intertwined MF approximation

Tanh-formula

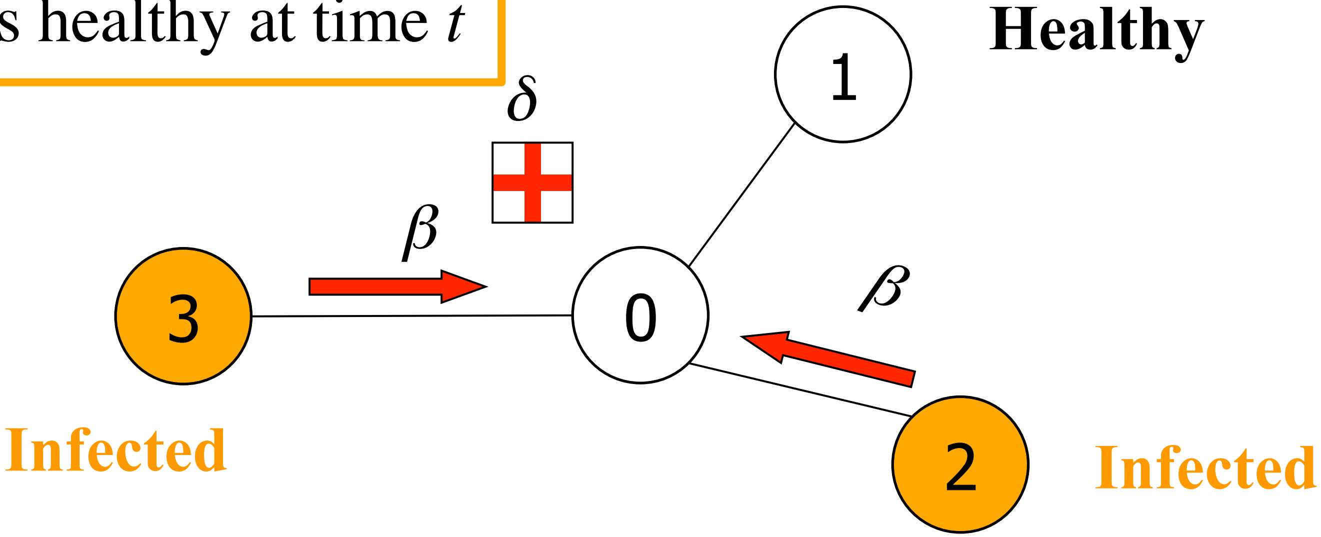
Continuous-time Markovian SIS epidemics on networks

- Constant infection rate β on all links
- Constant curing rate δ for all nodes

$\tau = \beta / \delta$: effective spreading rate

$X_j(t) = 1$ node j is infected at time t

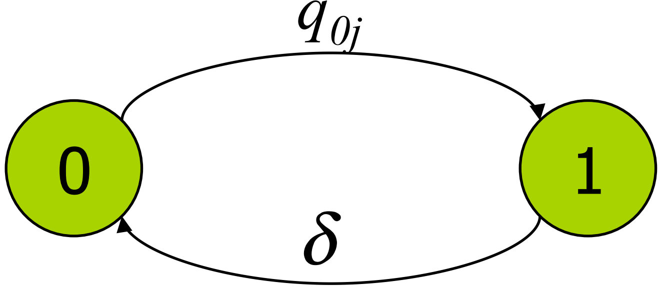
$X_j(t) = 0$ node j is healthy at time t



Infection and curing are independent Poisson processes

Markov theory: SIS on networks (1)

Local (each node): Two-states



Infinitesimal generator:

$$Q_j(t) = \begin{bmatrix} -q_{0j} & q_{0j} \\ \delta & -\delta \end{bmatrix}$$

The infection rate is due all infected neighbors of node *j:*

$$\boxed{q_{0j}(t) = \beta \sum_{k=1}^{N} a_{jk} X_k(t)}$$

where the adjacency matrix of the graph is

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}$$

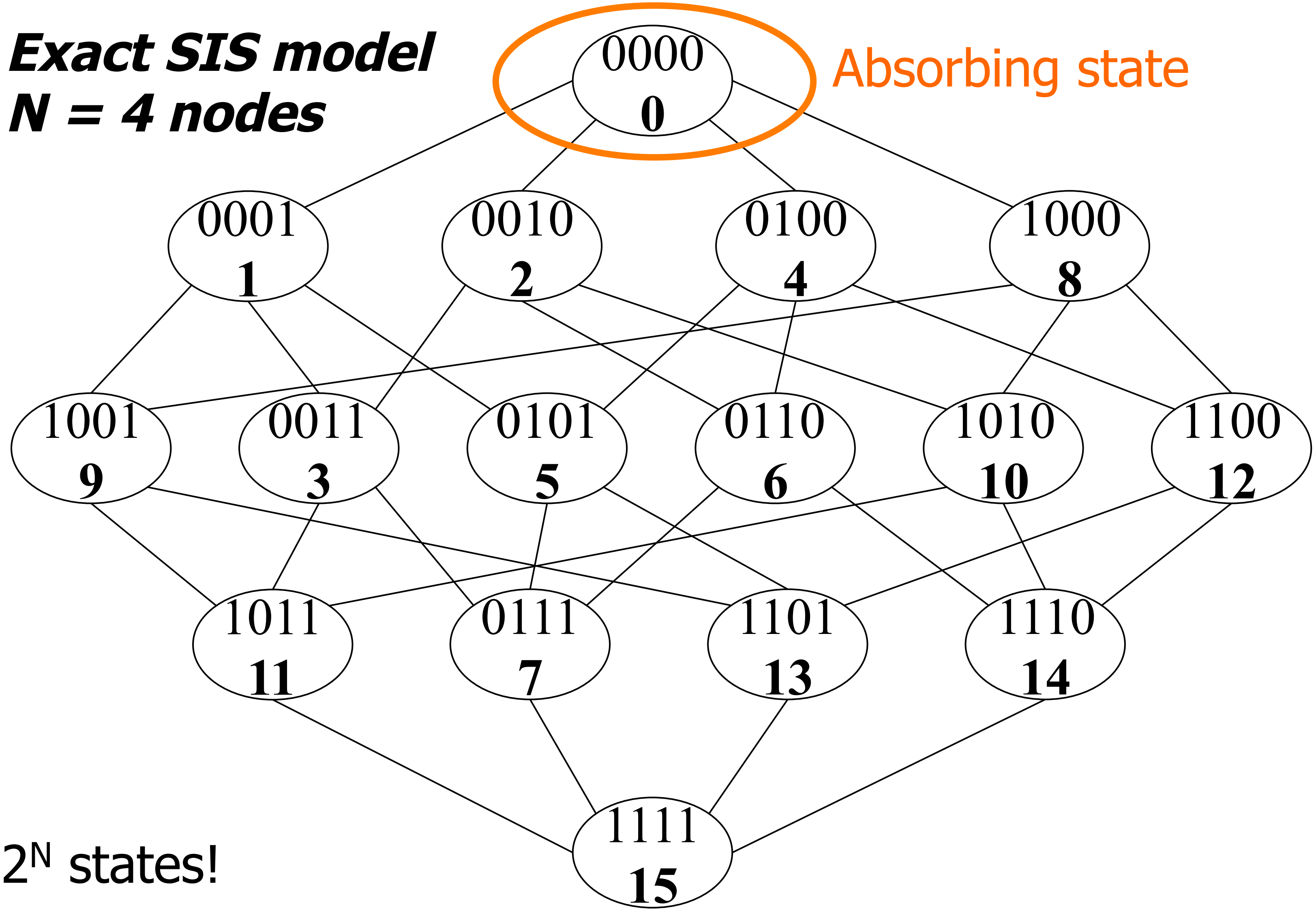
Markov theory: SIS on networks (2)

- Markov theory requires that the infinitesimal generator is a matrix whose elements are **not** random variables
- However, this is not the case in our simple model:

$$q_{0j}(t)=\beta\sum_{k=1}^{N}a_{jk}X_k(t)\xrightarrow{\text{NIMFA}}q_{0j}(t)=\beta\sum_{k=1}^{N}a_{jk}E\left[X_k(t)\right]$$

- By conditioning to each possible combination of infected states, we finally arrive to the exact Markov continuous SIS model
- *Drawback*: this exact model has 2^N states, where N is the number of nodes in the network.

Exact SIS model
N = 4 nodes

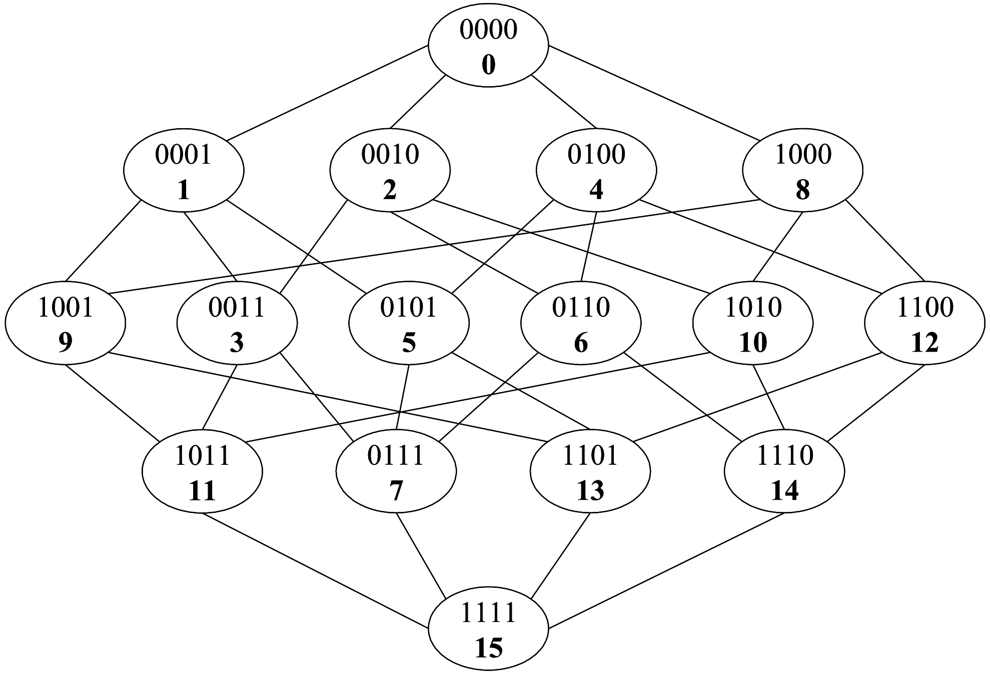


Absorbing state

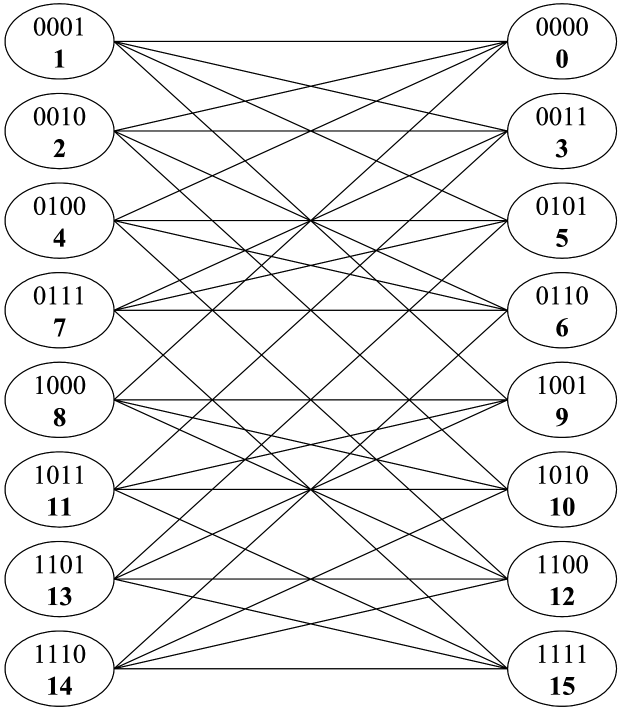
2^N states!

P. Van Mieghem, J. Omic, R. E. Kooij, “Virus Spread in Networks”,
IEEE/ACM Transaction on Networking, Vol. 17, No. 1, pp. 1-14, (2009).

Markov theory



Regular bipartite Markov graph



Recursive structure of infinitesimal general Q_N

Simon, P., M. Taylor and I. Z. Kiss, Exact epidemic models on graphs using graph-automorphism driven lumping, Mathematical Biology, Vol. 62, pp. 479-508, 2011

Van Mieghem, P. and E. Cator, ε-SIS epidemics and the epidemic threshold, Physical Review E, vol. 86, No. 1, July, p. 016116, 2012

Governing SIS equation for node *j*

$$\frac{dE[X_j]}{dt} = E\left[-\delta X_j + (1-X_j)\beta\sum_{k=1}^{N} a_{kj}X_k\right]$$

time-change of $E[X_j] = \Pr[X_j = 1]$, probability that node *j* is infected

if *infected*: probability of curing per unit time

if *not infected (healthy)*: probability of infection per unit time

$$\frac{dE[X_j]}{dt} = -\delta E\left[X_j\right] + \beta\sum_{k=1}^{N} a_{kj}E\left[X_k\right] - \beta\sum_{k=1}^{N} a_{kj}E\left[X_jX_k\right]$$

R. Pastor-Satorras, C. Castellano, P. Van Mieghem and A. Vespignani, "Epidemic processes in complex networks", Review of Modern Physics, Vol. 87, No. 3, pp. 925-979, 2015

Joint probabilities

$$\frac{dE\left[X_iX_j\right]}{dt} = E\left[\left\{-\delta X_i + \beta(1-X_i)\sum_{k=1}^{N} a_{ik}X_k\right\}X_j + X_i\left\{-\delta X_j + \beta(1-X_j)\sum_{k=1}^{N} a_{jk}X_k\right\}\right]$$

$$= -2\delta E\left[X_iX_j\right] + \beta\sum_{k=1}^{N} a_{ik}E\left[X_jX_k\right] + \beta\sum_{k=1}^{N} a_{jk}E\left[X_iX_k\right] - \beta\sum_{k=1}^{N}\left(a_{jk}+a_{ik}\right)E\left[X_iX_jX_k\right]$$

Next, we need the $\binom{N}{3}$ differential equations for E[$X_iX_jX_k$]...

In total, the SIS process is defined by $2^N = \sum_{k=1}^{N}\binom{N}{k} + 1$ linear equations

E. Cator and P. Van Mieghem, 2012, "Second-order mean-field SIS epidemic threshold", Physical Review E, vol. 85, No. 5, May, p. 056111.

Outline

Exact SIS model

NIMFA: N-intertwined MF approximation

Tanh-formula

NIMFA: N-intertwined mean-field approxim.

$$\frac{dE[X_j]}{dt} = -\delta E\left[X_j\right] + \beta\sum_{k=1}^{N} a_{kj}E\left[X_k\right] - \beta\sum_{k=1}^{N} a_{kj}E\left[X_jX_k\right]$$

$$\Longrightarrow \quad Cov\left[X_jX_k\right] = E\left[X_jX_k\right] - E\left[X_j\right]E\left[X_k\right] \geq 0$$

E. Cator and P. Van Mieghem, 2014, "Nodal infection in Markovian SIS and SIR epidemics on networks are non-negatively correlated," Physical Review E, Vol. 89, No. 5, p. 052802.

$$\boxed{\frac{dE[X_j]}{dt} = \underbrace{-\delta E\left[X_j\right] + \beta\left(1 - E\left[X_j\right]\right)\sum_{k=1}^{N} a_{kj}E\left[X_k\right]}_{\text{NIMFA: upper bounds SIS}}}\underbrace{- \beta\sum_{k=1}^{N} a_{kj}Cov\left[X_jX_k\right]}_{R_j > 0}$$

P. Van Mieghem and R. van de Bovenkamp, 2015, "Accuracy criterion for the mean-field approximation in SIS epidemics on networks," Physical Review E, Vol. 91, No. 3, p. 032812.

NIMFA: replace rv by its mean

$$\frac{dE[X_j]}{dt} = E\left[-\delta X_j + (1 - X_j)\beta \sum_{k=1}^{N} a_{kj} X_k\right]$$

NIMFA

$$X_j \Rightarrow E\left[\breve{X}_j\right]$$

$$\frac{dE\left[E\left[\breve{X}_j\right]\right]}{dt} = E\left[-\delta E\left[\breve{X}_j\right] + (1 - E\left[\breve{X}_j\right])\beta \sum_{k=1}^{N} a_{kj} E\left[\breve{X}_k\right]\right]$$

Bernoulli rv

$$E\left[\breve{X}_j\right] = \Pr\left[\breve{X}_j = 1\right] = v_j$$

$$\frac{dv_j}{dt} = -\delta v_j + (1 - v_j)\beta \sum_{k=1}^{N} a_{kj} v_k$$

Lower bound for the epidemic threshold

$$\frac{dv_j(t)}{dt} = -\delta v_j + \beta \sum_{k=1}^{N} a_{kj} v_k - \beta \sum_{k=1}^{N} a_{kj} E\left[X_i X_k\right] \qquad v_k(t) = E[X_k(t)]$$

Ignoring the correlation terms

$$\frac{dV(t)}{dt} \le \left(-\delta I + \beta A\right) V(t) \Longrightarrow V(t) \le \mathrm{e}^{(-\delta I + \beta A)t}\, V(0)$$

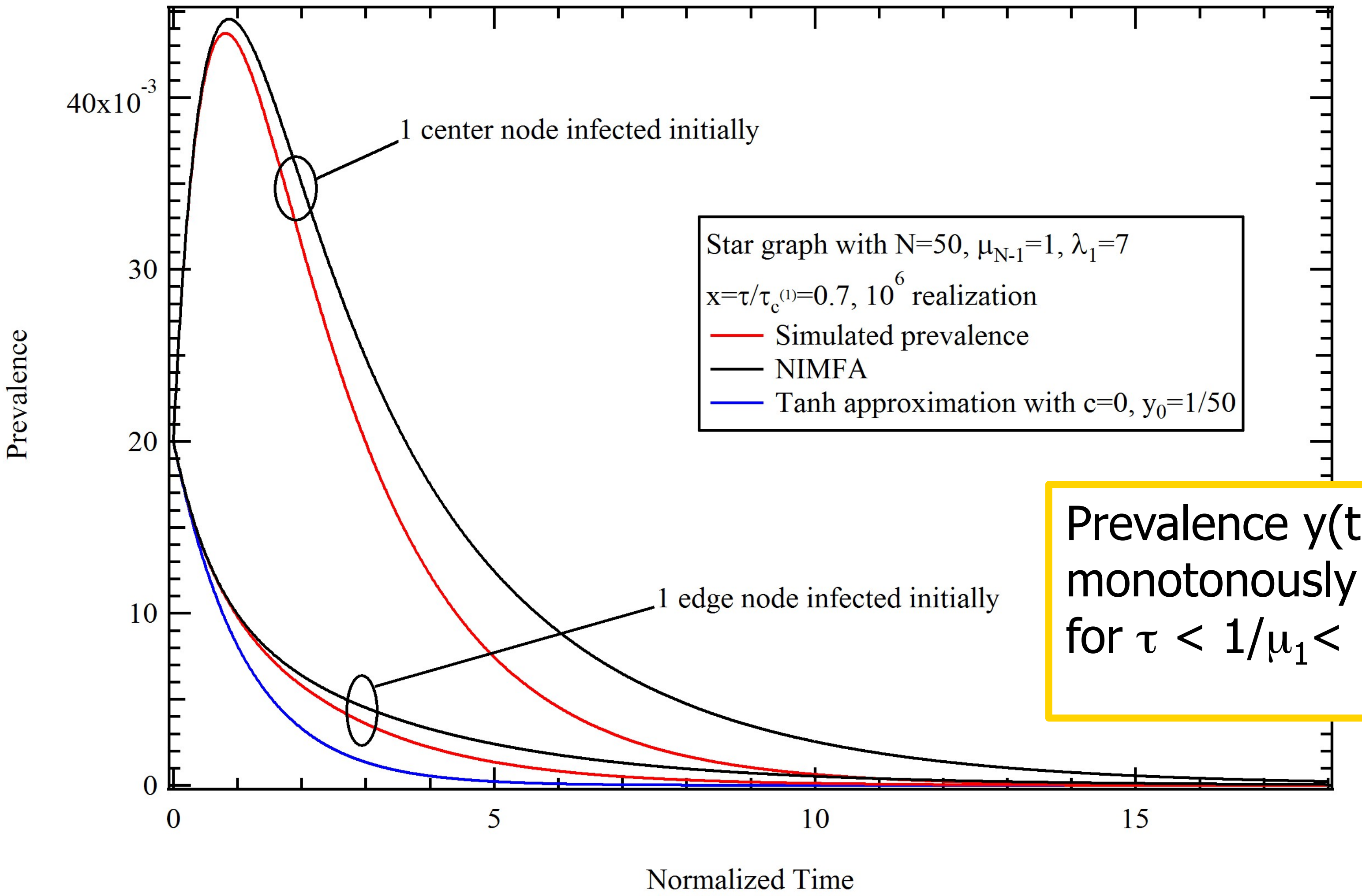
If all eigenvalues of $\beta A - \delta I$ are negative, v_j tends exponentially fast to zero for sufficiently large time *t*. Hence, if

$$\beta \lambda_1(A) - \delta < 0 \Longrightarrow \tau = \frac{\beta}{\delta} < \frac{1}{\lambda_1(A)} < \tau_c$$

The NIMFA epidemic threshold is precisely

$$\boxed{\tau_c^{(1)} = \frac{1}{\lambda_1(A)} < \tau_c} \qquad \boxed{\tau_c^{(1)} = \frac{1}{\lambda_1(A)} < \tau_c^{(2)} = \frac{1}{\lambda_1(H)} < \tau_c}$$

Below the epidemic threshold: $x = \lambda_1 \tau < 1$



Prevalence y(t) decreases monotonously
for $\tau < 1/\mu_1 < 1/\lambda_1$

P. Van Mieghem, 2016, "Approximate formula and bounds for the time-varying SIS prevalence in networks", Physical Review E, Vol. 93 No. 5, p. 052312.

What is so interesting about epidemics?

network protection

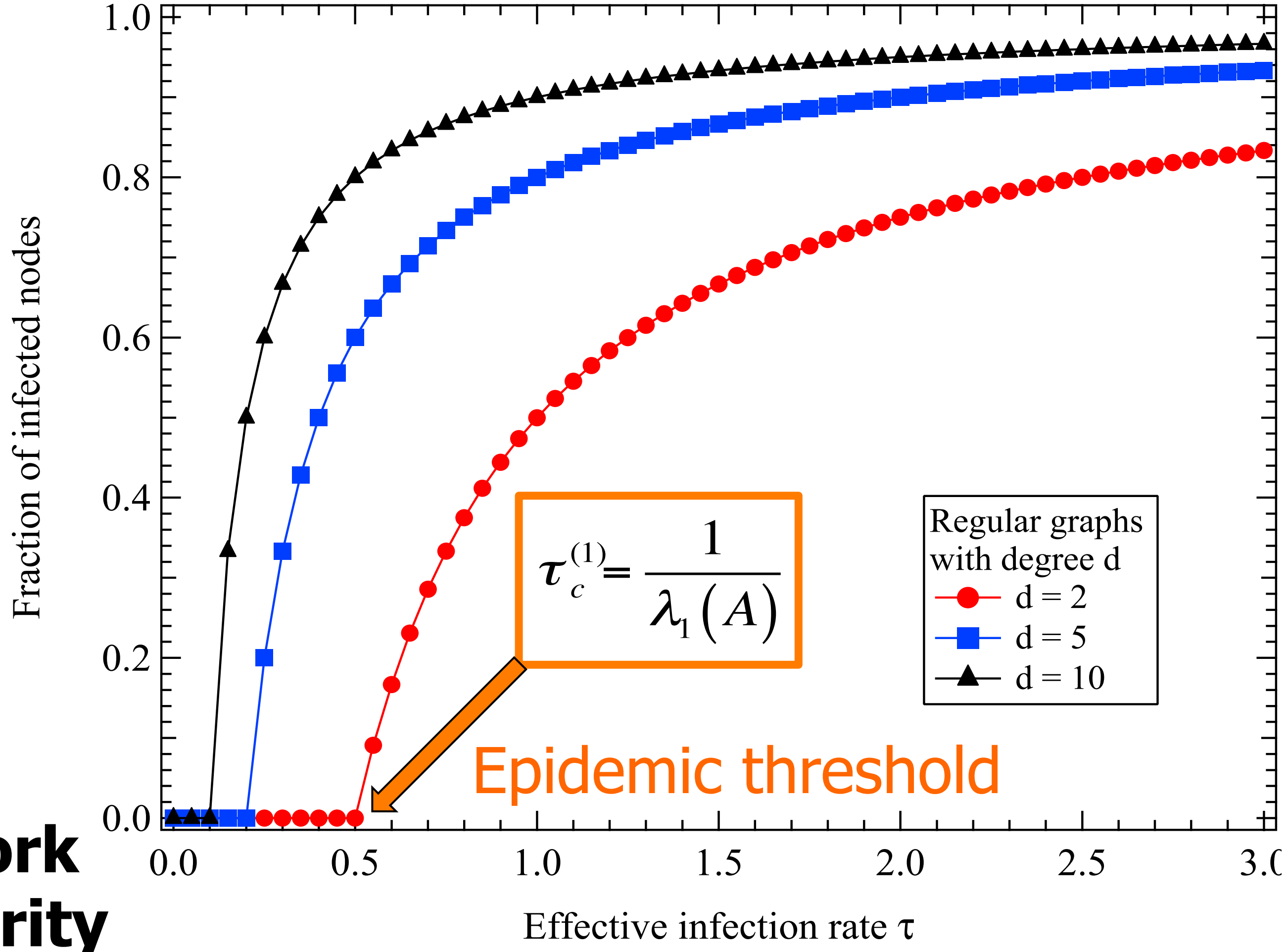
self-replicating objects (worms)

propagation errors

rumors (social nets)

epidemic algorithms (gossiping)

cybercrime : network robustness & security

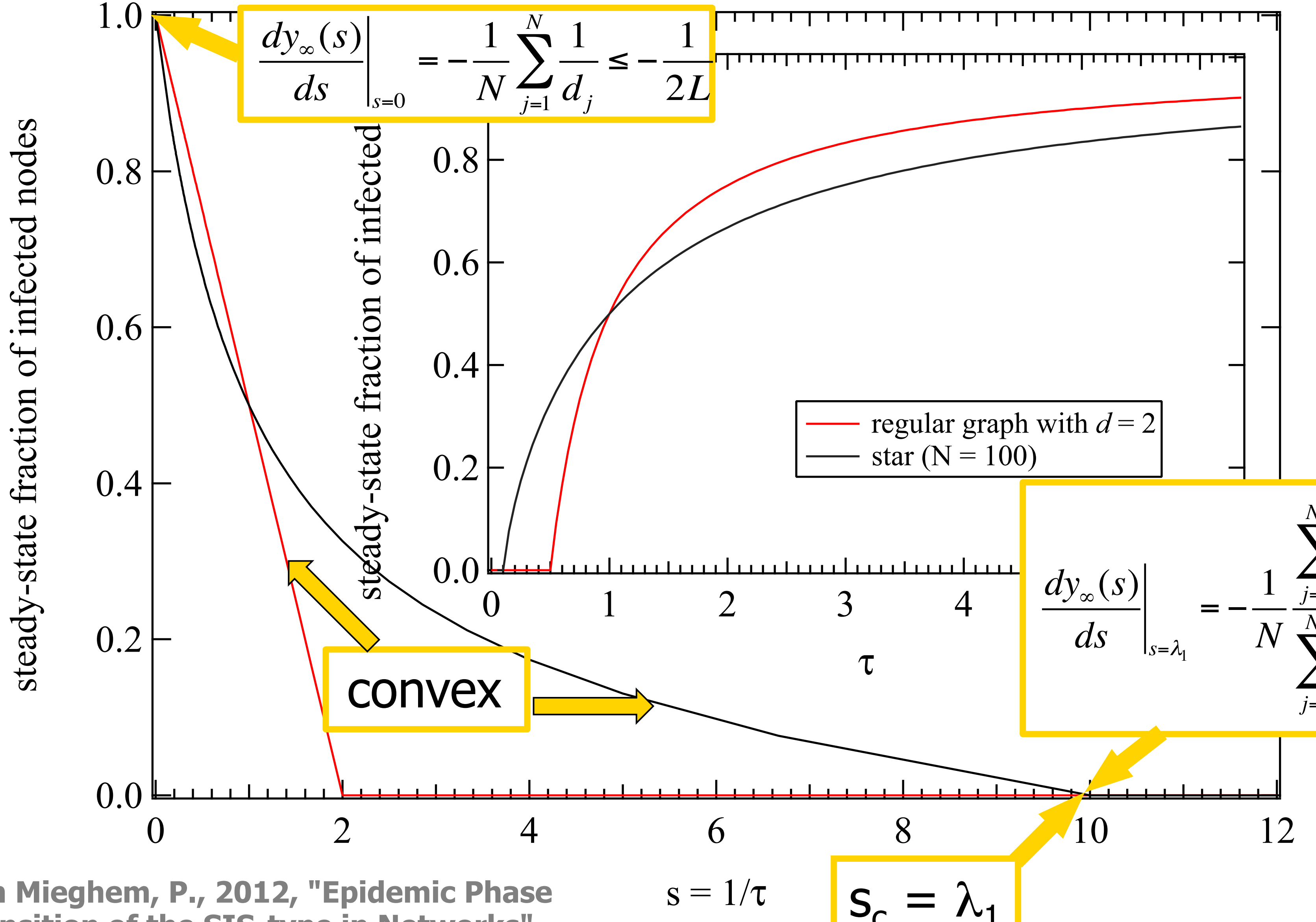


$$\tau_c^{(1)}=\frac{1}{\lambda_1(A)}$$

Epidemic threshold

$$\max\left(E[D]\sqrt{1+\frac{Var[D]}{(E[D])^2}},\sqrt{d_{\max}}\right)\le \lambda_1(A)\le d_{\max}$$

Transformation $s = \frac{1}{\tau}$ & principal eigenvector



$$\left.\frac{dy_\infty(s)}{ds}\right|_{s=0} = -\frac{1}{N}\sum_{j=1}^{N}\frac{1}{d_j} \le -\frac{1}{2L}$$

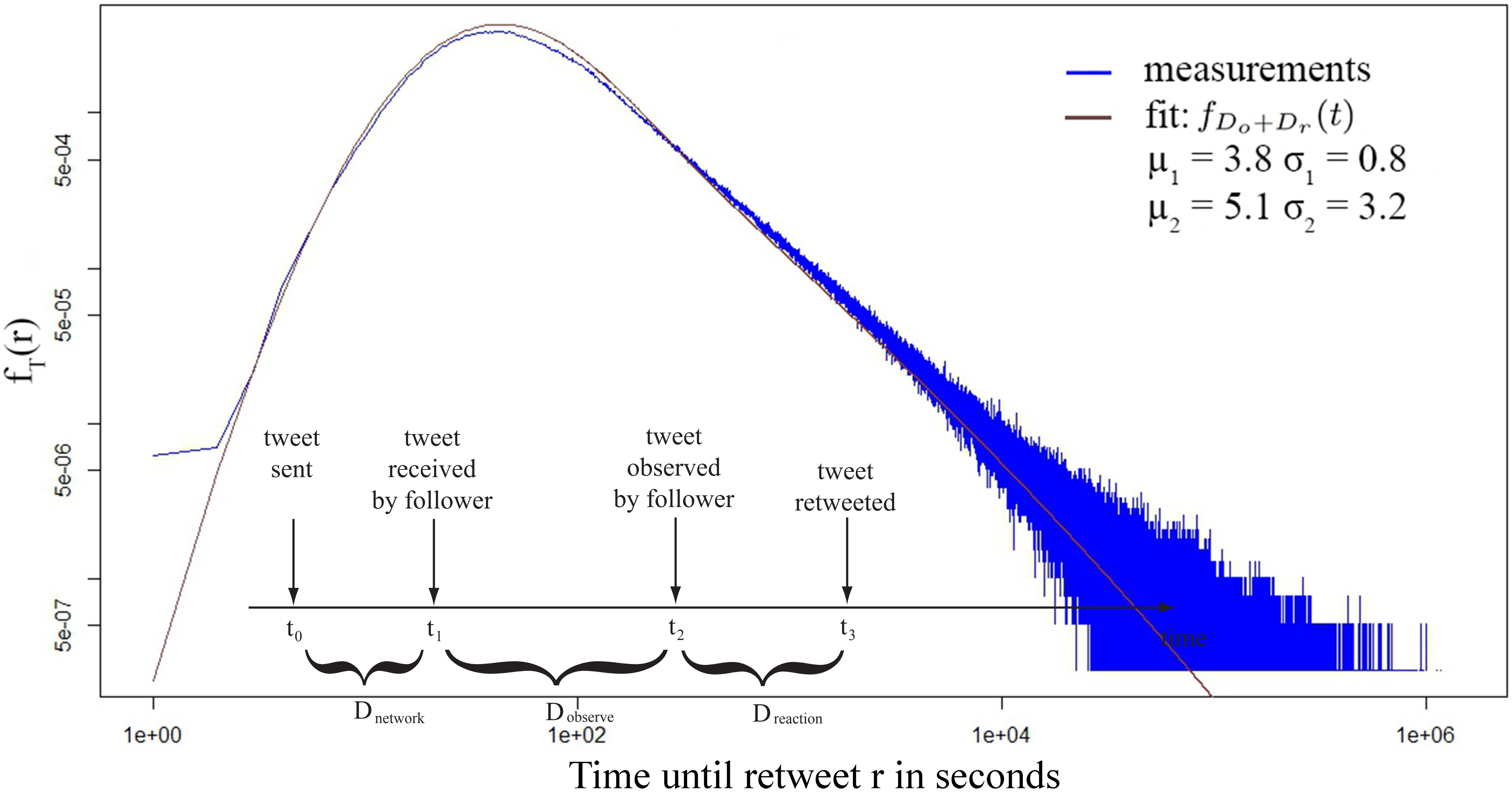
convex

$$\left.\frac{dy_\infty(s)}{ds}\right|_{s=\lambda_1} = -\frac{1}{N}\frac{\sum_{j=1}^{N}(x_1)_j}{\sum_{j=1}^{N}(x_1)_j^3}$$

$s_c = \lambda_1$

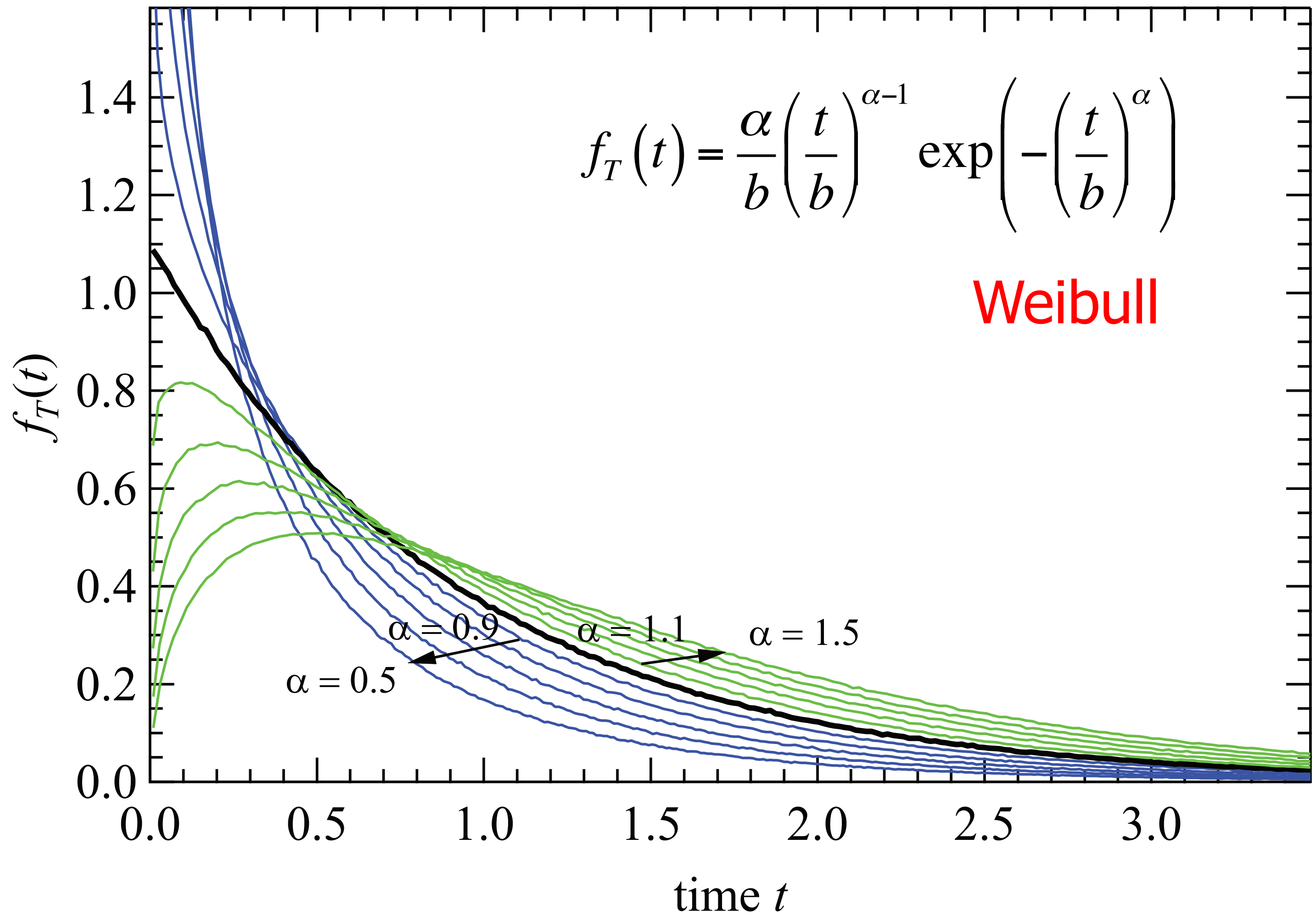
Van Mieghem, P., 2012, "Epidemic Phase Transition of the SIS-type in Networks", Europhysics Letters (EPL), Vol. 97, Februari, p. 48004.

Epidemic times are not exponential



C. Doerr, N. Blenn and P. Van Mieghem, "Lognormal infection times of Online information spread", PLOS ONE, Vol. 8, No. 5, p. e64349, 2013

Non-Markovian infection times



$$f_T(t) = \frac{\alpha}{b}\left(\frac{t}{b}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{b}\right)^{\alpha}\right)$$

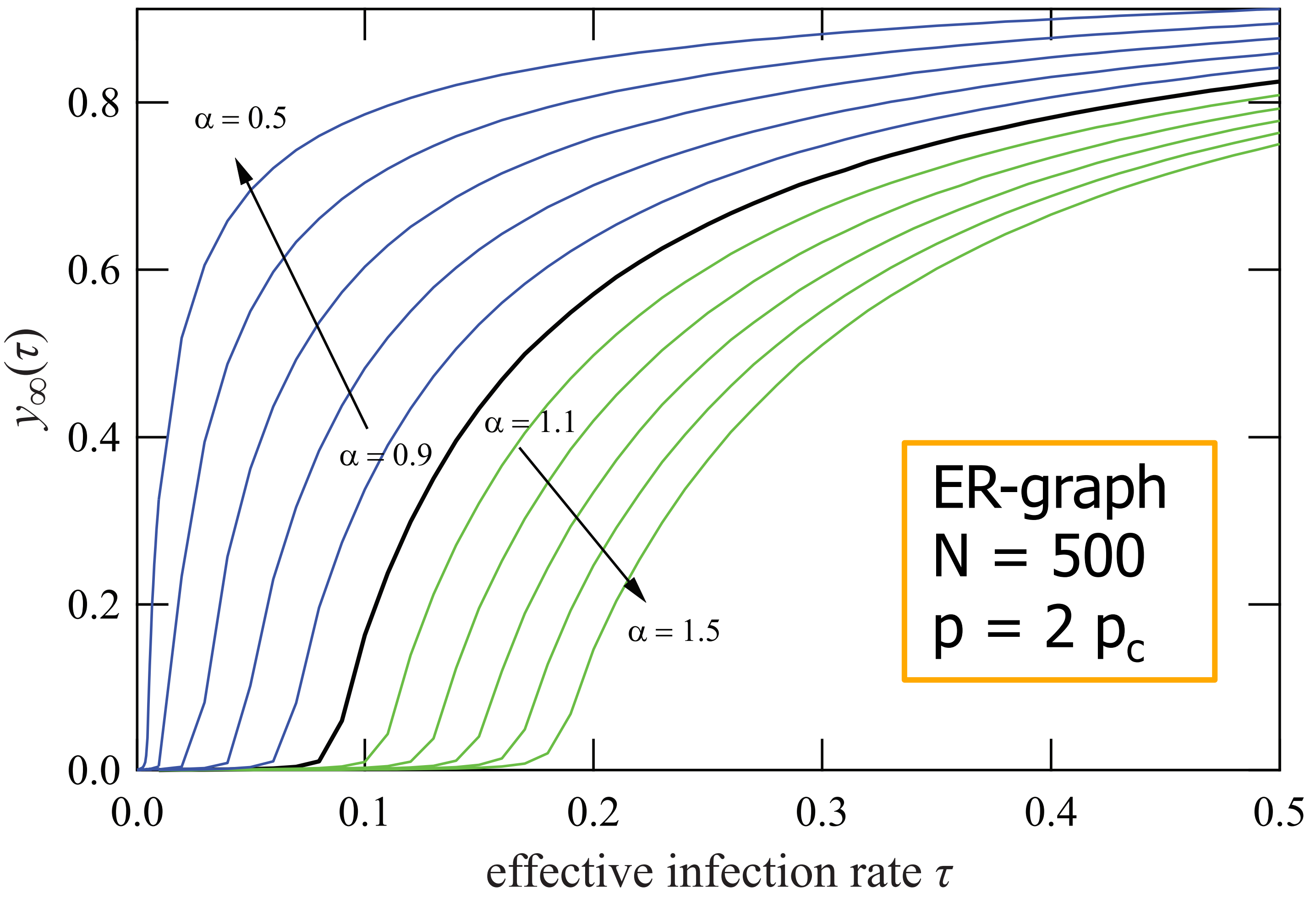
Weibull

Same mean E[*T*]:

$$b = \frac{1}{\beta\Gamma\left(1+\frac{1}{\alpha}\right)}$$

T is the time to infect a neighboring node

Non-Markovian epidemic threshold



Non-exponential infection time has a dramatic influence!

P. Van Mieghem and R. van de Bovenkamp, "Non-Markovian infection spread dramatically alters the SIS epidemic threshold", Physical Review Letters, vol. 110, No. 10, March, p. 108701.

GSIS: SIS with general infection times

NIMFA is valid provided the effective infection rate $\tau=\beta/\delta$ is replaced by the average number $E[M]$ of infection events during a healthy period (related to the basic reproduction number R_0):

$$\boxed{E[M]=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\phi_T(z)\phi_R(-z)}{1-\phi_T(z)}\frac{dz}{z}}$$

with pgf $\phi_X(z)=E[e^{-zX}]$

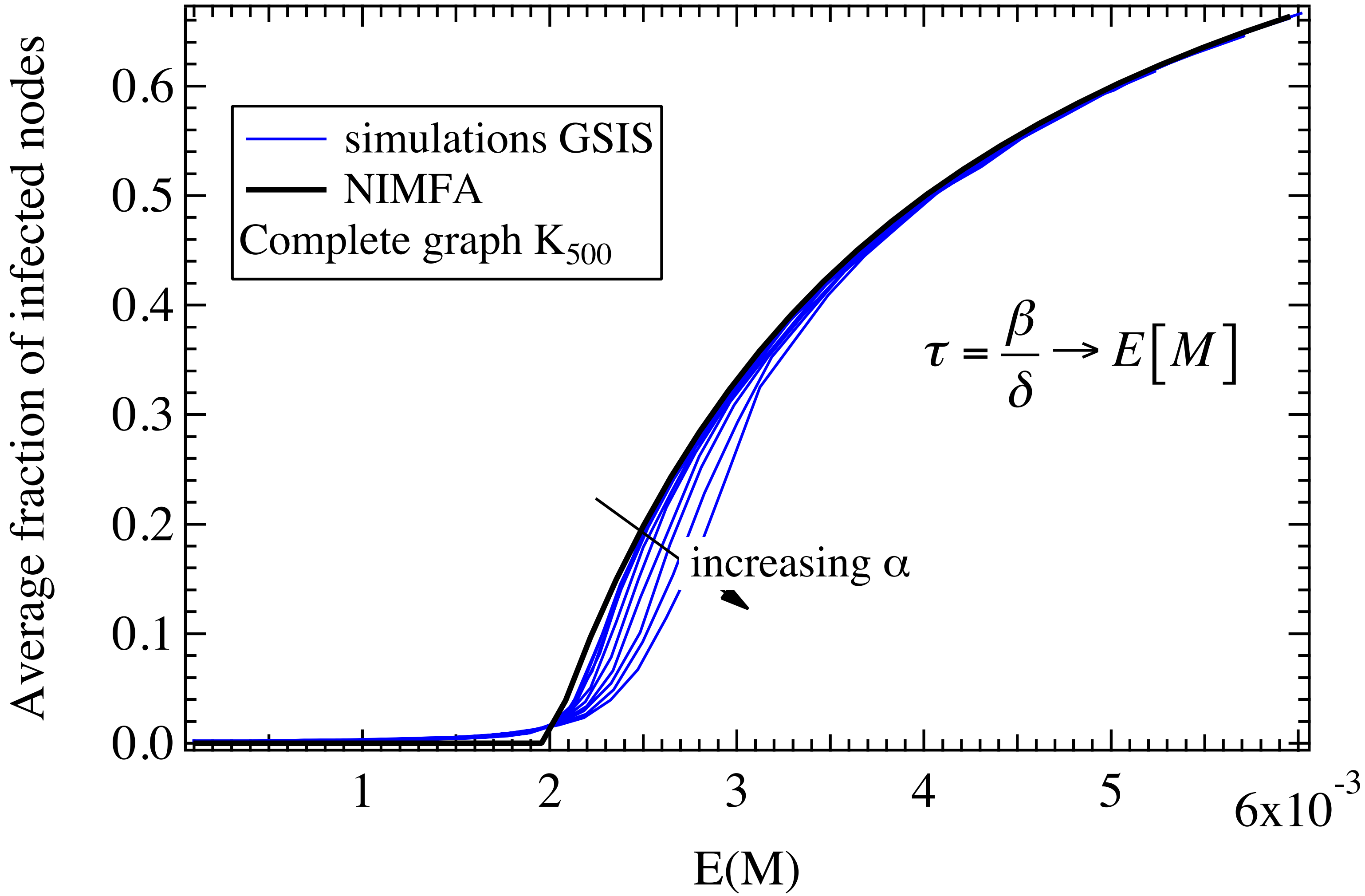
Generalized criterion for the epidemic threshold: $\boxed{E[M_c]=\frac{1}{\lambda_1}}$

Scaling law for large N
when infection time T is Weibullian
and recovery time R is exponential:

$$\boxed{\tau_c=\frac{q(\alpha)}{\lambda_1^{1/\alpha}}}\qquad q(\alpha)=O(1)$$

E. Cator, R. van de Bovenkamp and P. Van Mieghem, "SIS epidemics on networks with general infection and curing times", Physical Review E, Vol. 87, No. 6, p. 062816, 2013.

GSIS: E[M] gives the right scaling



$$f_T(t) = \frac{\alpha}{b}\left(\frac{t}{b}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{b}\right)^{\alpha}\right)$$

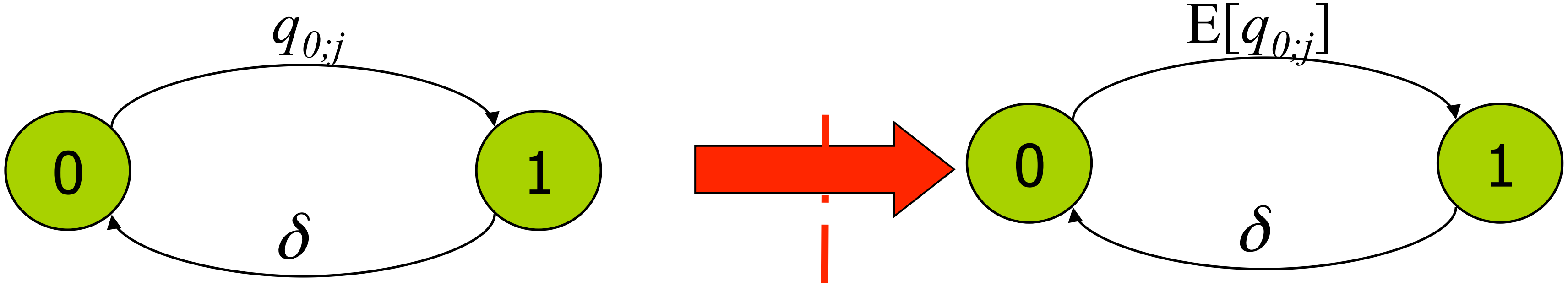
Extensions of the NIMFA

- **In-homogeneous**: each node i has own β_i and δ_i:
  P. Van Mieghem and J. Omic, 2008, "In-homogeneous Virus Spread in Networks", (arxiv.org/ 1306.2588)
- **SAIS** (Infected, Susceptible, Alert) and **SIR** instead of SIS:
  F. Darabi Sahneh and C. Scoglio, 2011,"Epidemic Spread in Human Networks", 50th IEEE Conf. Decision and Contol, Orlando, Florida.
  "M. Youssef and C. Scoglio, 2011, An individual-based approach to SIR epidemics in contact networks" Journal of Theoretical Biology 283, pp. 136-144.
- Generalized Epidemic mean-field model (**GEMF**): general extension of NIMFA to *m* compartments (includes both SIS, SAIS, SIR,...):
  F. Darabi Sahneh, C. Scoglio, P. Van Mieghem, 2013, "Generalized Epidemic Mean-Field Model for Spreading Processes over Multi-Layer Complex Networks",, IEEE/ACM Transactions on Networking, Vol. 21, No. 5, pp. 1609-1620.
- NIMFA on **Interdependent networks**
  Wang, H., Q. Li, G. D'Agostino, S. Havlin, H. E. Stanley and P. Van Mieghem, 2013, "Effect of the Interconnected Network Structure on the Epidemic Threshold", Physical Review E, Vol. 88, No. 2, August, p. 022801.

Extensions of the NIMFA

- **Time-dependent rates β(t) and δ(t) in regular graphs**:
  P. Van Mieghem, 2014, "SIS epidemics with time-dependent rates describing ageing of information spread and mutation of pathogens", Delft University of Technology, report20140615.
  Maertens, M., H. Asghari, M. van Eeten and P. Van Mieghem, 2016, "A Time-dependent SIS-model for Long-term Computer Worm Evolution", IEEE Conference on Communications and Network Security, 17-19 October, Philadelphia, PA USA.

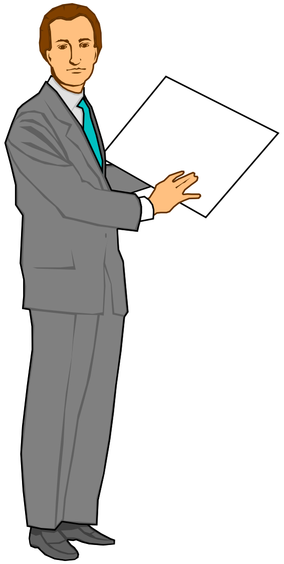
Mean-field approximation (NIMFA)



- *2^N* linear equations
- Steady-state
    - absorbing (healthy) state
    - reached after unrealistically long time
- difficult to analyze
- **only** for exponential infection and curing times

- *N* non-linear equations
- Meta-stable state:
    - phase-transition
    - epidemic threshold
    - realistic
- analytically tractable
- lower bound epidemic threshold
- valid for **any** infection and/or curing time distribution

Outline



Exact SIS model

NIMFA: N-intertwined MF approximation

Tanh-formula

SIS Prevalence

- Fraction of infected nodes in the graph *G*

$$S(t) = \frac{1}{N}\sum_{j=1}^{N} X_j(t)$$ (random variable!)

- Prevalence: Expected fraction of infected nodes in *G*

$$y(t) = E[S(t)] = \frac{1}{N}\sum_{j=1}^{N} \Pr[X_j(t) = 1]$$

P. Van Mieghem, 2016, "Approximate formula and bounds for the time-varying SIS prevalence in networks", Physical Review E, Vol. 93 No. 5, p. 052312.

Differential equation prevalence (1)

$$\frac{dE[X_j]}{dt} = E\left[-\delta X_j + (1 - X_j)\beta \sum_{k=1}^{N} a_{kj} X_k\right]$$

After summing over all nodes:

$$\frac{d}{\delta dt}\left(\frac{1}{N}\sum_{j=1}^{N} E[X_j]\right) = E\left[-\frac{1}{N}\sum_{j=1}^{N} X_j + \frac{\tau}{N}\sum_{j=1}^{N}\sum_{k=1}^{N}(1 - X_j)a_{kj}X_k\right]$$

Using the definition of prevalence: $y(t) = \frac{1}{N}\sum_{j=1}^{N} E\left[X_j\right]$

$$\frac{dy(t^*)}{dt^*} = -y(t^*) + \frac{\tau}{N} E\left[\sum_{j=1}^{N}\sum_{k=1}^{N}(1 - X_j)a_{kj}X_k\right]$$

Normalized time t* = δ t

Differential equation prevalence (2)

$$\frac{dy\left(t^*\right)}{dt^*} = -y\left(t^*\right) + \frac{\tau}{N} E\left[\sum_{j=1}^{N}\sum_{k=1}^{N}(1-X_j)a_{kj}X_k\right]$$

Executing the double sum:

$$\sum_{j=1}^{N}\sum_{k=1}^{N}(1-X_j)a_{kj}X_k = 2\sum_{l\in L}(1-X_{l^+})X_{l^-}$$

For Bernoulli rv's:

$$\left(X_{l^-}-X_{l^+}\right)^2 = X_{l^-}^2 - 2X_{l^-}X_{l^+} + X_{l^+}^2 = X_{l^-} - 2X_{l^-}X_{l^+} + X_{l^+}$$
$$= X_{l^-}\left(1-X_{l^+}\right) + X_{l^+}\left(1-X_{l^-}\right)$$

Thus:

$$\sum_{j=1}^{N}\sum_{k=1}^{N}(1-X_j)a_{kj}X_k = \sum_{l\in L}(X_{l^+}-X_{l^-})^2 = w^TQw$$

where *w* is the nodal random Bernoulli vector

$$w = \left(X_1, X_2, \ldots, X_N\right)$$

Differential equation prevalence (3)

Finally, in terms of the Laplacian $Q = \Delta - A$ and the normalized time t* = δ t

$$\boxed{\frac{dy\left(t^*\right)}{dt^*} = -y\left(t^*\right) + \frac{\tau}{N} E\left[w^T Q w\right]}$$

P. Van Mieghem, F. Darabi Sahneh and C. Scoglio, 2014, "Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold", Proceedings of the 53rd IEEE Conference on Decision and Control (CDC'14), December 15-17, Los Angeles, CA, USA (also on http://arxiv.org/abs/1402.1731).

“Local rule - global emergent properties” class

$$\frac{dE[X_j(t)]}{dt} = E\left[-\delta X_j(t) + (1 - X_j(t))\beta \sum_{k=1}^{N} a_{kj} X_k(t)\right]$$

Local SIS rule

Global emergent SIS spread

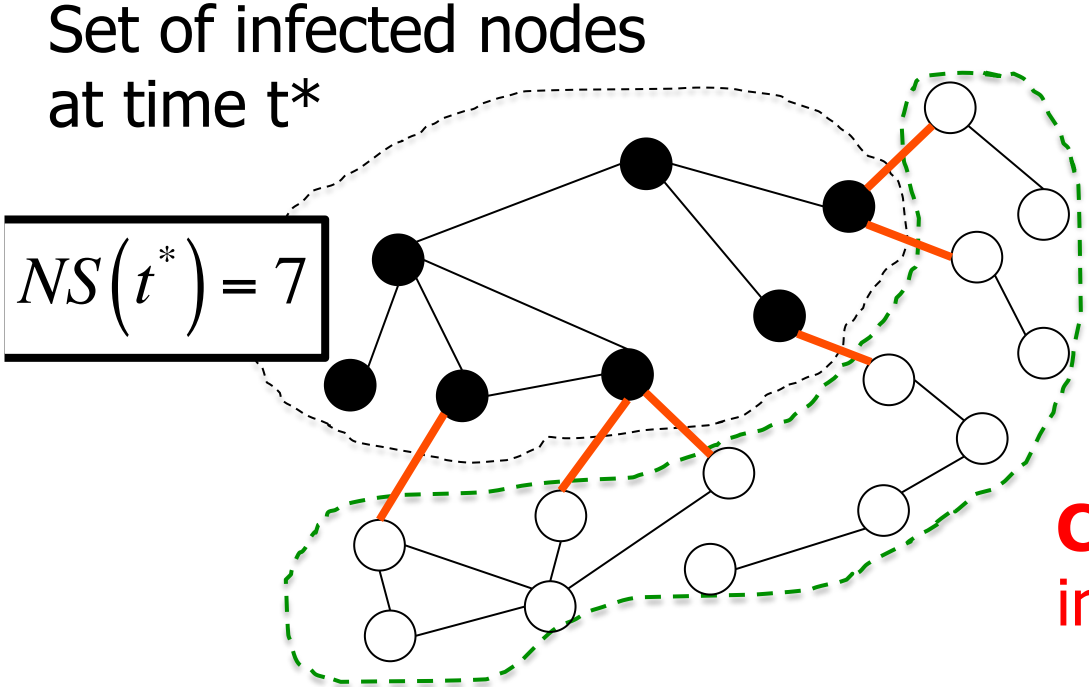
$$\frac{dy(t^*)}{dt^*} = -y(t^*) + \frac{\tau}{N} E\left[w^T(t^*)\, Q\, w(t^*)\right]$$

The Laplacian $Q = \Delta - A$
The normalized time $t^* = \delta t$
Bernoulli state vector
$w(t^*) = \left(X_1(t^*), X_2(t^*), \ldots, X_N(t^*)\right)$

P. Van Mieghem, F. Darabi Sahneh and C. Scoglio, 2014, "Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold", Proceedings of the 53rd IEEE Conference on Decision and Control (CDC’14), December 15-17, Los Angeles, CA, USA (also arxiv.1402.1731).

SIS prevalence dynamics

$$\frac{dy\left(t^*\right)}{dt^*} = -y\left(t^*\right) + \frac{\tau}{N} E\left[w^T\left(t^*\right) Q w\left(t^*\right)\right]$$



Set of infected nodes at time t*

$NS\left(t^*\right) = 7$

$w^T\left(t^*\right) Q w\left(t^*\right) = 6$

Cut-Set: set of links with 1 infected node at time t*

Set of susceptible nodes at time t*

Spectral decomposition

Expanding *w* in terms of all eigenvectors of *Q*; x_k belongs to μ_k

$$w = \sum_{k=1}^{N} \zeta_k x_k \quad \Longrightarrow \quad w^T Q w = \sum_{k=1}^{N} \mu_k \zeta_k^2 \quad \text{with} \quad \zeta_k = w^T x_k$$

Using *Qu=0* and Bernoulli rv properties lead to

$$w^T Q w = \mu_{N-1} N \left(S - S^2 \right) + R$$

and

$$\boxed{\frac{dy\left(t^*\right)}{dt^*} = \left(\tau\mu_{N-1} - 1\right) y\left(t^*\right) - \tau\mu_{N-1} y^2\left(t^*\right) - \Psi}$$

with

$$\boxed{\Psi = \tau\mu_{N-1}\left(Var\left[S\right] - \frac{E\left[R\right]}{N\mu_{N-1}} \right)}$$

$$R = \sum_{k=1}^{N-2} \left(\mu_k - \mu_{N-1}\right) \zeta_k^2(t) \geq 0 \qquad R_{K_N} = 0$$

Spectral decomposition: tanh-formula

Bounding $c_L \le \Psi \le c_U$ yields a Riccati differential equation

$$\frac{d\breve{y}\left(t^*\right)}{dt^*} = \left(\tau\mu_{N-1} - 1\right)\breve{y}\left(t^*\right) - \tau\mu_{N-1}\breve{y}^2\left(t^*\right) - c$$

with solution

$$\boxed{\breve{y}\left(t^* \middle| c\right) = \frac{1}{2}\left(1 - \frac{1}{\tau\mu_{N-1}}\right) + \frac{\Xi}{2}\tanh\left(\frac{\tau\mu_{N-1}\Xi}{2}t^* + \Omega\right)}$$

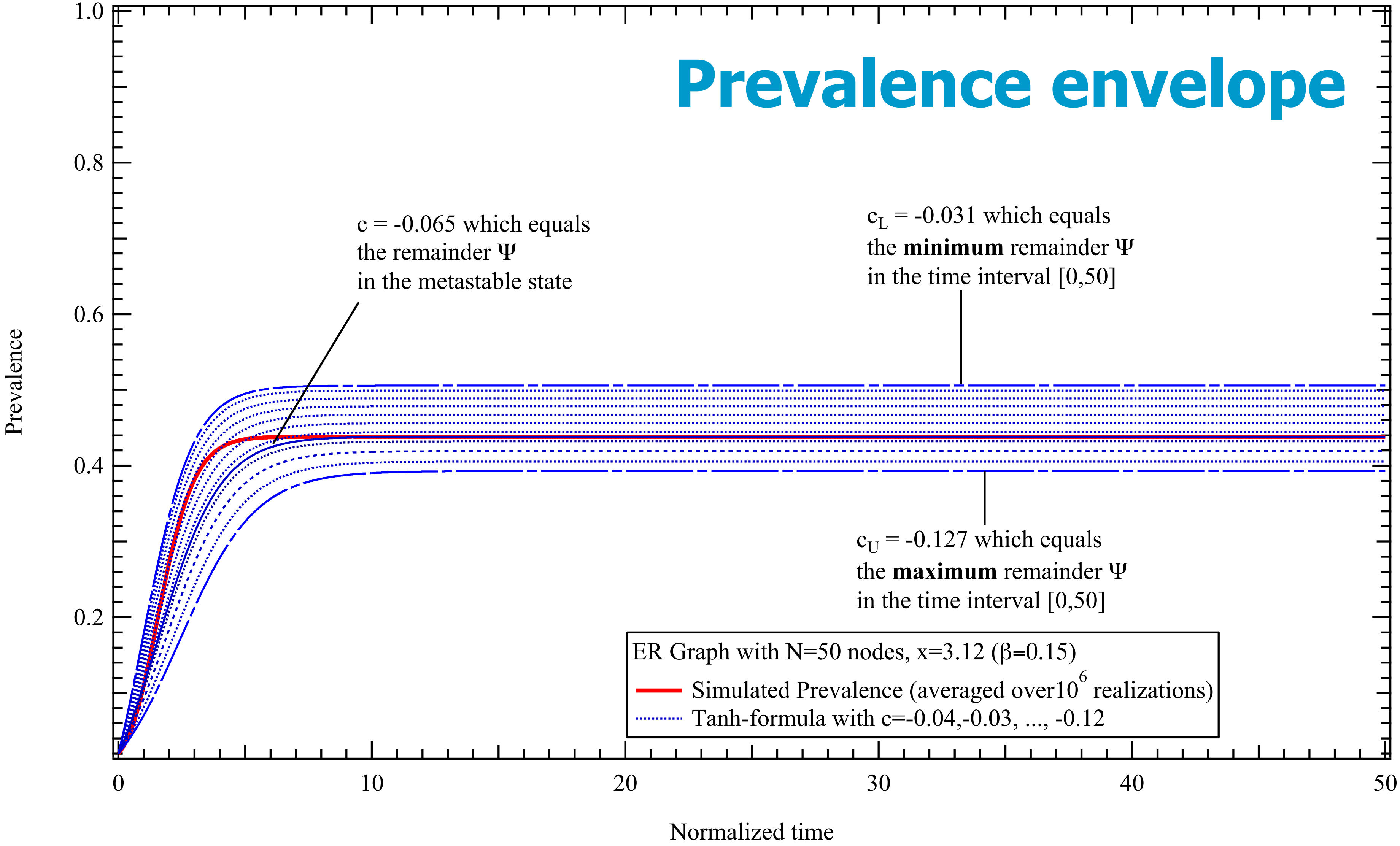
where

$$\boxed{\Xi = \sqrt{\left(1 - \frac{1}{\tau\mu_{N-1}}\right)^2 - \frac{4c}{\tau\mu_{N-1}}}} \qquad \boxed{\Omega = \operatorname{arctanh}\left(\frac{1}{\Xi}\left(2y_0 - \left(1 - \frac{1}{\tau\mu_{N-1}}\right)\right)\right)}$$

- The prevalence is upper and lower bounded by the tanh-formula:

$$\boxed{\breve{y}\left(t^* \middle| c_U\right) < y\left(t^*\right) < \breve{y}\left(t^* \middle| c_L\right)}$$

Prevalence envelope



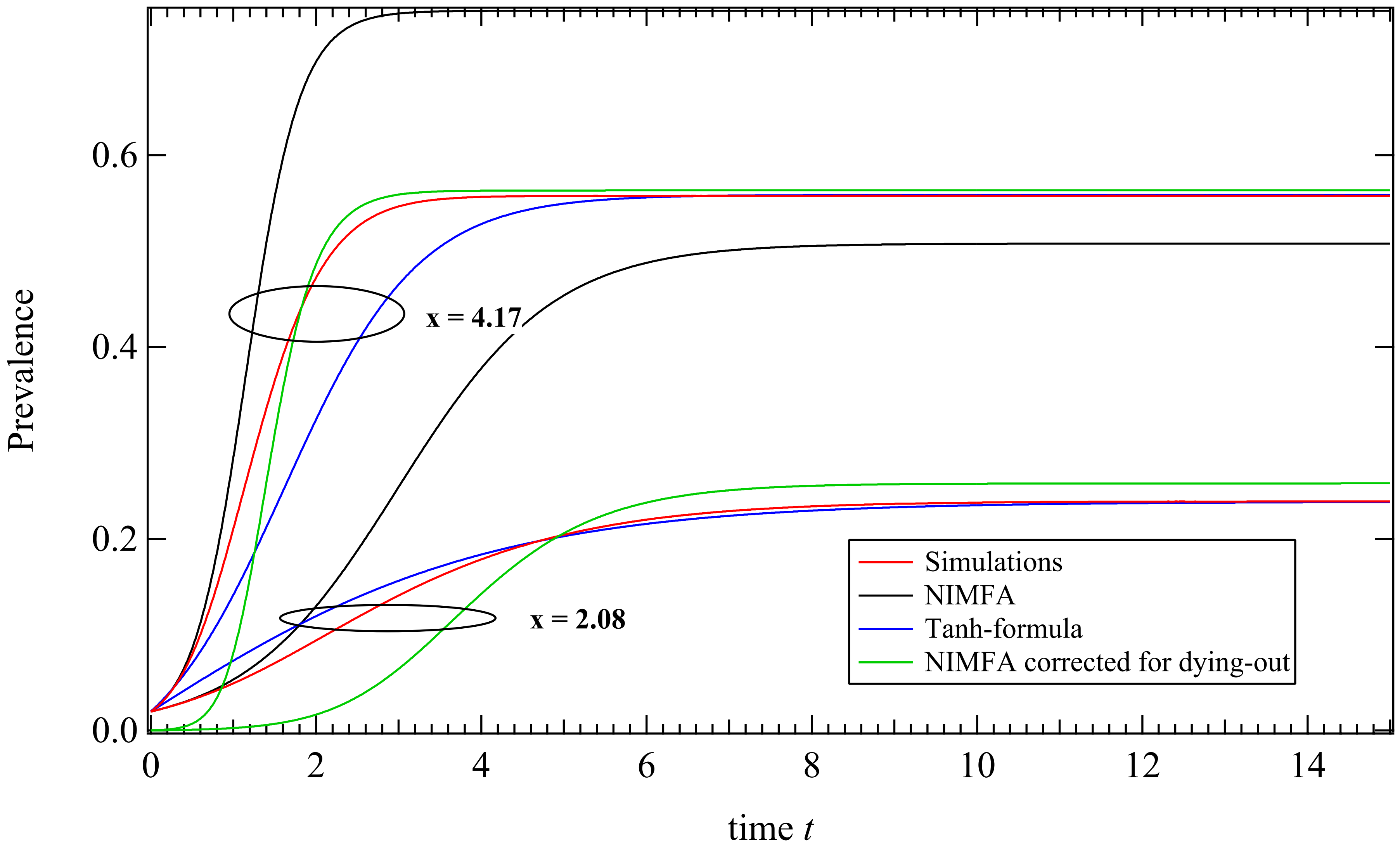
c = -0.065 which equals
the remainder Ψ
in the metastable state

c_L = -0.031 which equals
the **minimum** remainder Ψ
in the time interval [0,50]

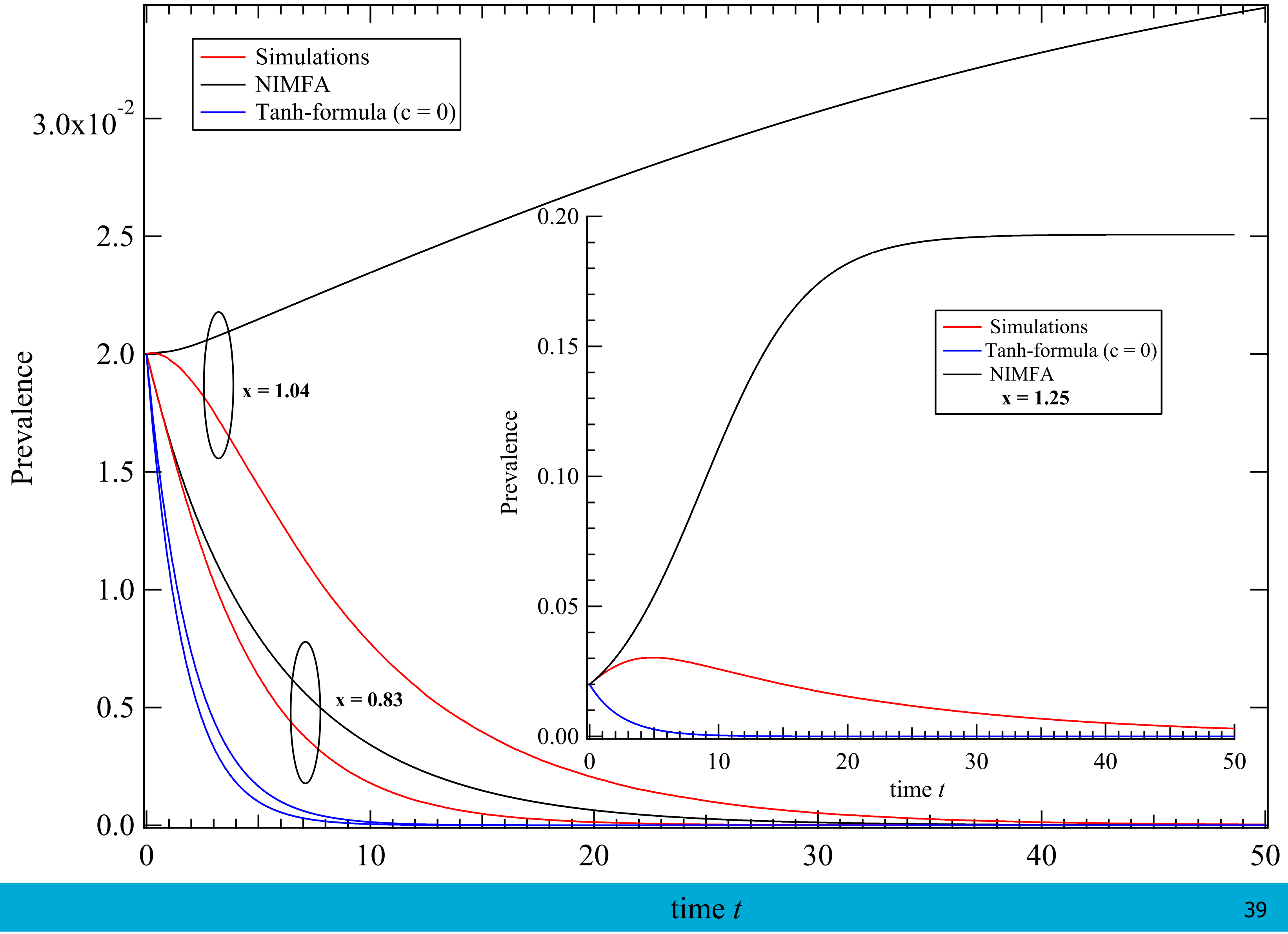
c_U = -0.127 which equals
the **maximum** remainder Ψ
in the time interval [0,50]

$$\boxed{\breve{y}\left(t^*\middle|c_U\right) < y\left(t^*\right) < \breve{y}\left(t^*\middle|c_L\right)}$$

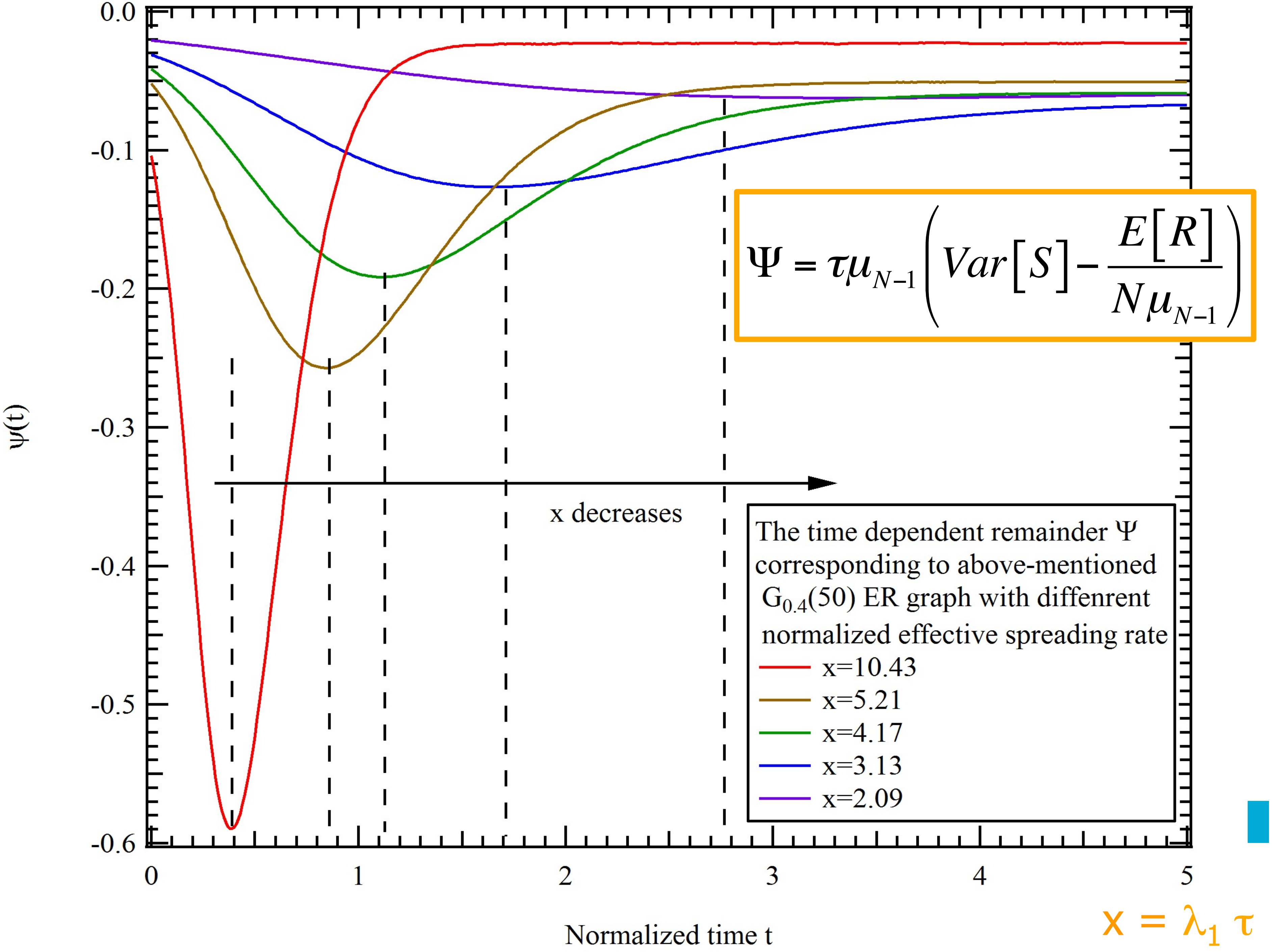
Comparison in ER $G_{0.4}(50)$



λ_1 = 20.85 and μ_{N-1} = 10.11 and $x = \lambda_1 \tau$


Simulations
NIMFA
Tanh-formula (c = 0)
3.0×10^{-2}
2.5
2.0
1.5
1.0
0.5
0.0
Prevalence
x = 1.04
x = 0.83
0
10
20
30
40
50
time t
0.20
0.15
0.10
0.05
0.00
Prevalence
Simulations
Tanh-formula (c = 0)
NIMFA
x = 1.25
0
10
20
30
40
50
time t

ER $G_{0.4}(50)$



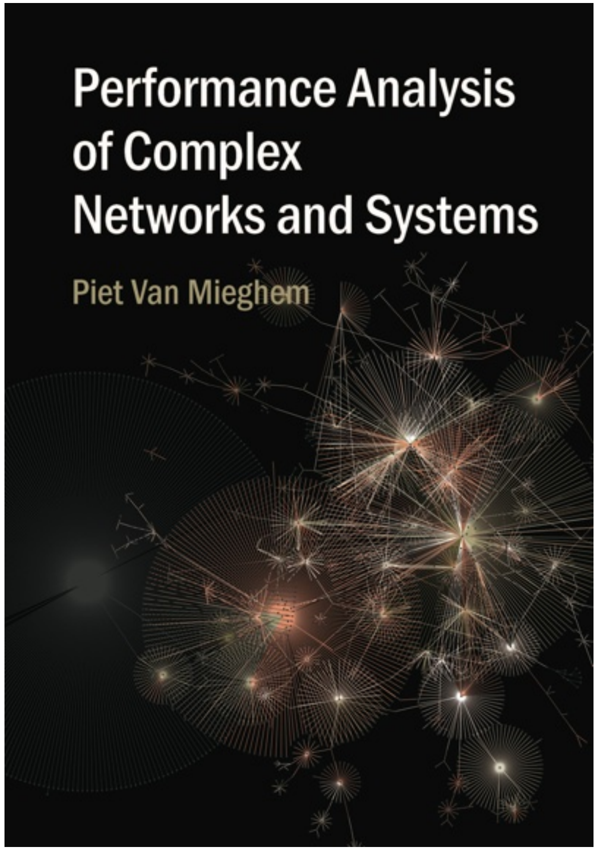
$$\Psi = \tau\mu_{N-1}\left(Var\left[S\right] - \frac{E\left[R\right]}{N\mu_{N-1}}\right)$$

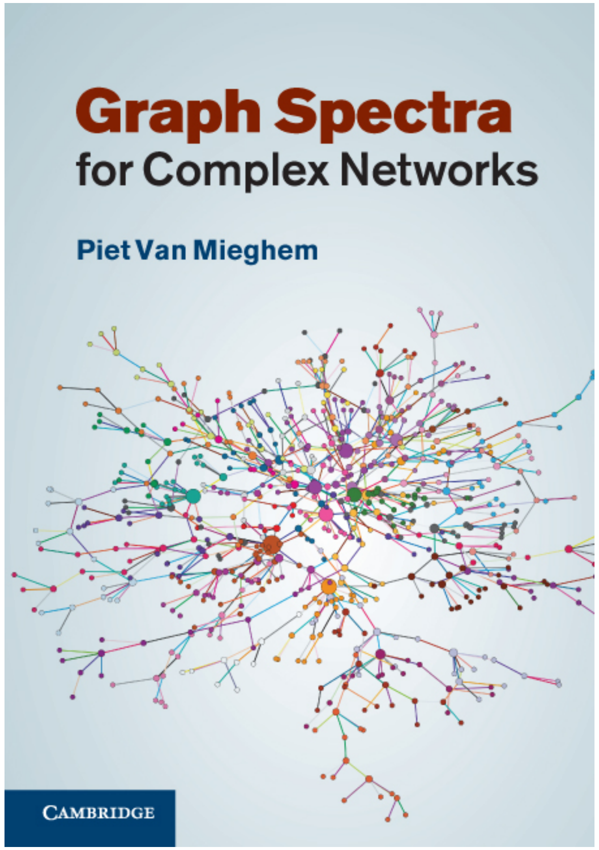
$x = \lambda_1 \tau$

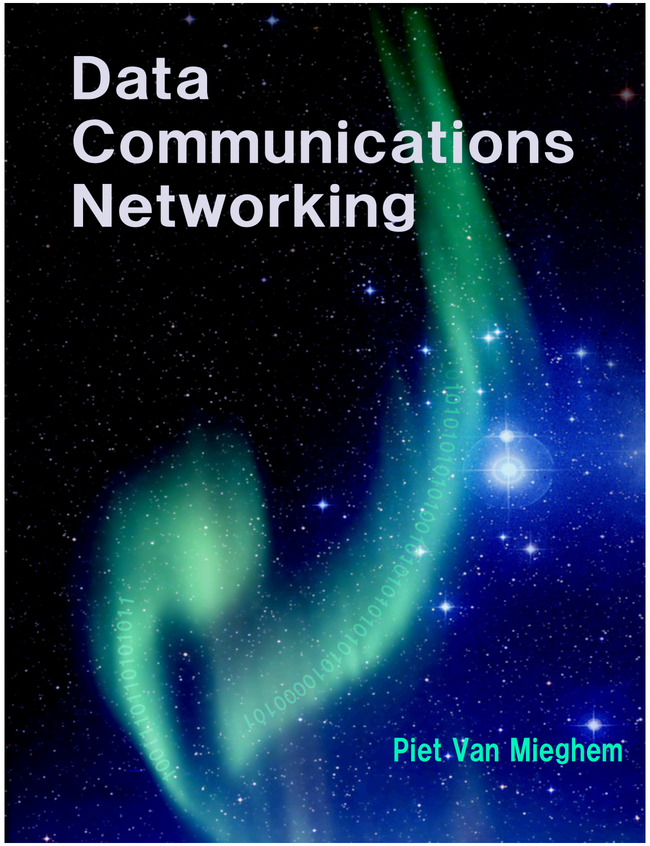
Challenges for epidemics on networks

- **Theory**:
  - A general mean-field criterion: for which graphs is NIMFA accurate? Is the conjecture true?
  - Non-unimodality of the prevalence
  - Tight upper bound of the epidemic threshold (for any graph), or near to exact determination of the epidemic threshold τ_c
  - Time-dependent analysis of SIS epidemics: beyond the tanh-formula and non-Markovian epidemics
- Epidemics on *evolving, adaptive* and *temporal* networks
- Competing and mutating viruses on networks
- Modeling of social contagion
- Control of epidemics on networks
- **Measured data** of epidemics (e.g. fraction of infected nodes & the underlying topology of the 'contact' network) in real-world networks!

Books







Articles: **http://www.nas.ewi.tudelft.nl**

Thank You

Piet Van Mieghem
NAS, TUDelft
P.F.A.VanMieghem@tudelft.nl