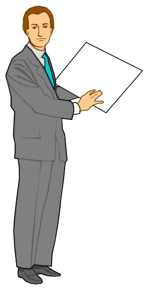
Virus Spread in Networks

Piet Van Mieghem

work in collaboration with Jasmina Omic, Huijuan Wang, Rob Kooij, Mina Youssef and Caterina Scoglio

Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

Viral Conductance

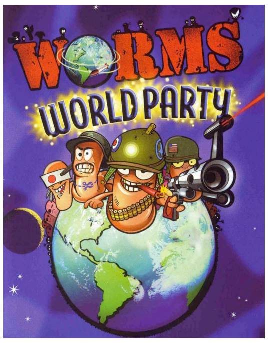
Modifying the epidemic threshold

Summary

Motivation for virus spread in networks

- Computer viruses
  - security threat to Internet
  - annoyance
  - very costly
    - Code Red worm: several billion $$ in damage
- Why do we care?
  - Understanding the spread of a virus is the first step in preventing it
  - How fast do we need to disinfect nodes so that the virus dies quickly? Which nodes?

Applications of virus spread models



- **Computer virus and worms modelling**
- **Epidemic algorithms**
- **Error propagation in networks**
- **Any self-replicating object on a dynamic network**
- **Emotions as infectious diseases in social networks**

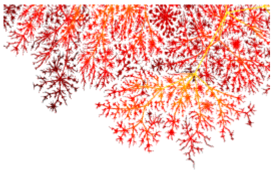
History*

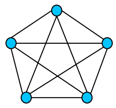
First epidemic model by Bernoulli (1760)

Homogeneous mixing model (biology)

Epidemic on a graph (Kephart and White)

Power-law graphs (Pastor-Satorras and Vespignani)





* “Epidemic modelling – an introduction” – D. J. Daley and J. Gani

State of the art of epidemic modeling

- Timing parameters modelling.
  - “Modelling the Effects of Timing Parameters on Virus Propagation” (Y. Wang and C. Wang)
- Topology influence on epidemic.
  - “The Effect on Network Topology on the Spread of Epidemic” (Ganesh, Massoulie, Towsley)
  - “Threshold for Virus Spread on Networks” (Draief, Ganesh, Massoulie)
- Percolation and generating function approach.
  - “The spread of epidemic disease on networks” (Newman)
- Contact Processes (mathematical theory)
  - Branch of continuous-time Markov theory (book by Liggett)

Spreading process

- *States in which a node can be*:
  - susceptible (S)
  - infected (I)
  - incubating
  - alert
  - removed (either cured or death)
- *Time description of the model*:
  - continuous-time
  - discrete-time

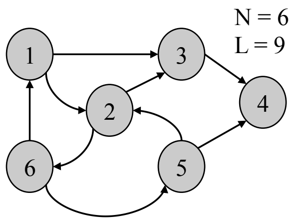
- *Focus here is a simple model*:
  - a SIS model in continuous time applicable to **any network**

A few graph metrics…

- Number of **nodes** N and **links** L in a graph G
- **Degree** d_i of a node i: number of directly linked neighbors
  - degree distribution of all your first hop neighbors, all second hop neighbors, etc…
- **Hopcount** (number of links) of a shortest path between two nodes
  - **Diameter:** largest possible hopcount (distance)
  - ***k*-th level set**: number of nodes at hop k in the shortest path tree rooted at a node (also called the **expansion** of G)
- **Connectivity** and number of **components (clusters)**
  - strongly connected components
  - number of **cliques** (= complete subgraphs K_m with $m < N$)
  - **edge/vertex connectivity**: minimum number of links/node whose removal disconnects G
- **Betweenness** of a link/node: number of shortest paths between all pair of nodes in G that traverses the link/node
- **… and many more**…

Algebraic graph theory

Any graph G can be represented by an adjacency matrix A and an incidence matrix B, and a Laplacian Q



$$A_{N\times N} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} = A^T$$

$$B_{N\times L} = \begin{bmatrix} 1 & 1 & -1 & \dots & 0 \\ -1 & 0 & 0 & & 0 \\ 0 & -1 & 0 & & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & & -1 \\ 0 & 0 & 1 & & 1 \end{bmatrix}$$

$$Q = BB^T = \Delta - A$$

$$\Delta = diag(d_1 \quad d_2 \quad \dots \quad d_N)$$

TU Delft

A wealth of relations…

Degree law: $\sum_{j=1}^{N} d_j = 2L$

Number of k-hop walks between node i and j: $(A^k)_{ij}$

Any **real symmetric matrix S** can be written as $S = X \Lambda X^T$, where X is the orthog. matrix with real eigenvectors in the columns and $\Lambda = diag(\lambda_1,\ldots,\lambda_N)$, where λ_j is the j-th real eigenvalue. Eigenvalues can be ordered as $\boxed{\lambda_N \le \lambda_{N-1} \le \cdots \le \lambda_2 \le \lambda_1}$

Spectrum of A: 1) all eigenvalues lie in the interval $(-d_{max}, d_{max}]$

2) $\sum_{j=1}^{N} \lambda_j = 0 \qquad \sum_{j=1}^{N} \lambda_j^2 = 2L \qquad \sum_{j=1}^{N} \lambda_j^k = Trace(A^k) = \sum_{j=1}^{N} (A^k)_{jj}$

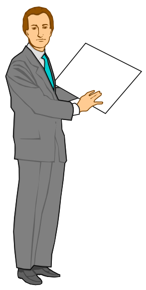
Spectrum of Q: 1) any eigenvalue μ_k is non-negative and smallest $\mu_N = 0$.

2) complexity (number of spanning trees) is $\xi(G) = \frac{1}{N}\prod_{k=1}^{N-1} \mu_k$

3) the second smallest eigenvalue (**algebraic connectivity μ_{N-1}**) is related to how strongly a graph is connected

TU Delft

Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

Viral Conductance

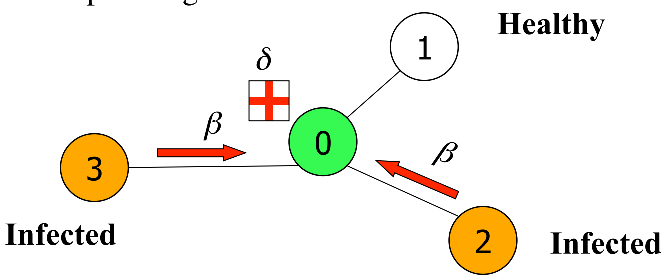
Modifying the epidemic threshold

Summary

Simple SIS model (1)

- Homogeneous birth (infection) rate β on all edges between infected and susceptible nodes
- Homogeneous death (curing) rate δ for infected nodes

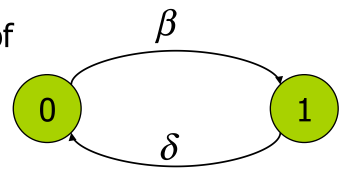
$\tau = \beta / \delta$: effective spreading rate



12

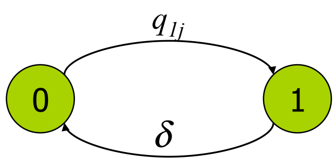
Simple SIS model (2)

- Each node *j* can be in either of the two states:
  - "0": healthy
  - "1": infected
- Markov continuous time:
  - infection rate β
  - curing rate δ
- Mathematically:
  - X_j is the state of node j
  - infinitesimal generator $Q_j(t) = \begin{bmatrix} -q_{1j} & q_{1j} \\ q_{2j} & -q_{2j} \end{bmatrix} = \begin{bmatrix} -q_{1j} & q_{1j} \\ \delta & -\delta \end{bmatrix}$



Simple SIS model (3)

- Nodes are interconnected in graph:
  $$Q_j(t) = \begin{bmatrix} -q_{1j} & q_{1j} \\ \delta & -\delta \end{bmatrix}$$



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

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

and where the adjacency matrix of the graph is

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

Simple SIS model (4)

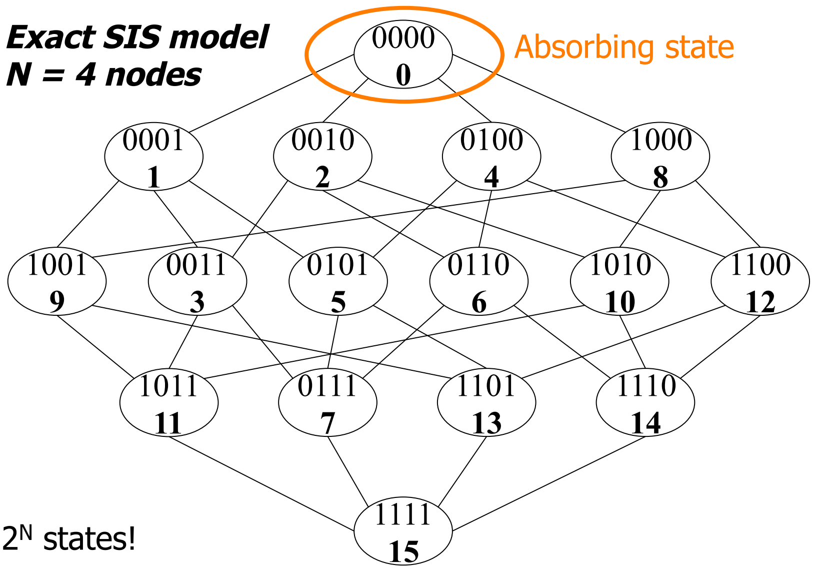
- Markov theory requires that the infinitesimal generator is a matrix whose elements are NOT random variables
- However, this is not the case in our simple model:

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

- 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.

TUDelft

Exact SIS model
N = 4 nodes



2^N states!

TUDelft

Simple SIS model (5): mean field

- The infinitesimal generator

$$Q_j(t) = \begin{bmatrix} -q_{1j} & q_{1j} \\ \delta & -\delta \end{bmatrix} \qquad \boxed{q_{1j}(t) = \beta \sum_{k=1}^{N} a_{jk} 1_{\{X_k(t)=1\}}}$$

is replaced by its mean (***the only approximation***!)

$$Q_j = \begin{bmatrix} -E[q_{1j}] & E[q_{1j}] \\ \delta & -\delta \end{bmatrix} \qquad \boxed{E[q_{1j}(t)] = \beta \sum_{k=1}^{N} a_{jk} \Pr[\{X_k(t)=1\}]}$$

- Being able now to apply ordinary Markov theory, we arrive at our **_N_-intertwined model for virus spread**

$$\begin{cases} \frac{dv_1}{dt} = (1-v_1)\beta \sum_{k=1}^{N} a_{1k} v_k - \delta v_1 \\ \frac{dv_2}{dt} = (1-v_2)\beta \sum_{k=1}^{N} a_{2k} v_k - \delta v_2 \\ \vdots \\ \frac{dv_N}{dt} = (1-v_N)\beta \sum_{k=1}^{N} a_{Nk} v_k - \delta v_N \end{cases}$$

where $v_k(t) = \Pr[X_k(t) = 1]$

TU Delft

N-intertwined virus spread model

- Non-linear matrix equation:

$$\boxed{\frac{dV(t)}{dt} = \beta A.V(t) - diag(v_i(t))(\beta A.V(t) + \delta u)}$$

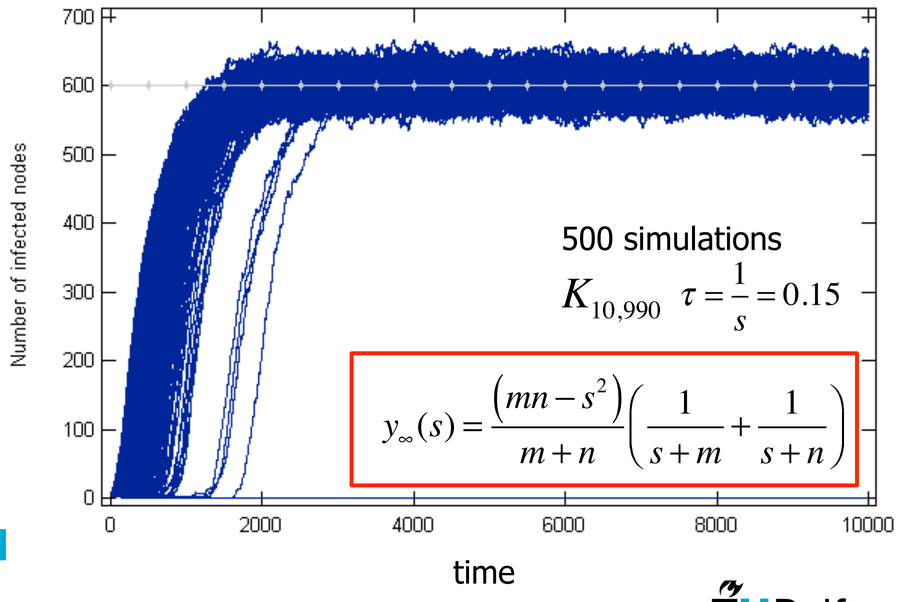
where the vector $u^T = [1\ 1\ \dots\ 1]$ and $V^T = [v_1\ v_2\ \dots\ v_N]$

- *Results:*
  - Probability of infection v_k for each node k separately
  - Number of infected nodes in the steady state
  - Phase transition phenomena for any network (largest eigenvalue of the adjacency matrix A)
  - Analytic computations feasible:
    - expansions of v_k as a function of the effective infection rate around the epidemic threshold and around infinity

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).

TU Delft

Simulations



TU Delft

Kephart-White model

Assume perfect homogeneity & symmetry: a graph of degree *r*

$$\begin{cases} \frac{dv_1}{dt} = (1-v_1)\beta \sum_{k=1}^{N} a_{1k} v_k - \delta v_1 \\ \frac{dv_2}{dt} = (1-v_2)\beta \sum_{k=1}^{N} a_{2k} v_k - \delta v_2 \\ \vdots \\ \frac{dv_N}{dt} = (1-v_N)\beta \sum_{k=1}^{N} a_{Nk} v_k - \delta v_N \end{cases}$$

$$\frac{dv}{dt} = (1-v)\beta r v - \delta v$$

steady-state

$$v = 1 - \frac{1}{r\tau} \quad \text{where} \quad \tau = \frac{\beta}{\delta}$$

threshold

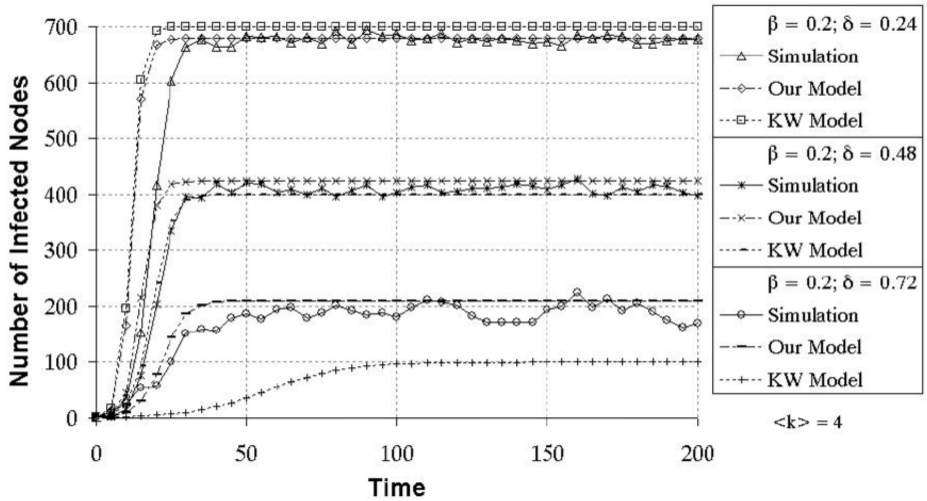
$$\tau \geq \tau_c = \frac{1}{r} \quad \text{then} \quad v \geq 0$$

20

J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," *Proc. IEEE Comput. Soc. Symp. Research in Security and Privacy*, May 1991, pp. 343–359.

TU Delft

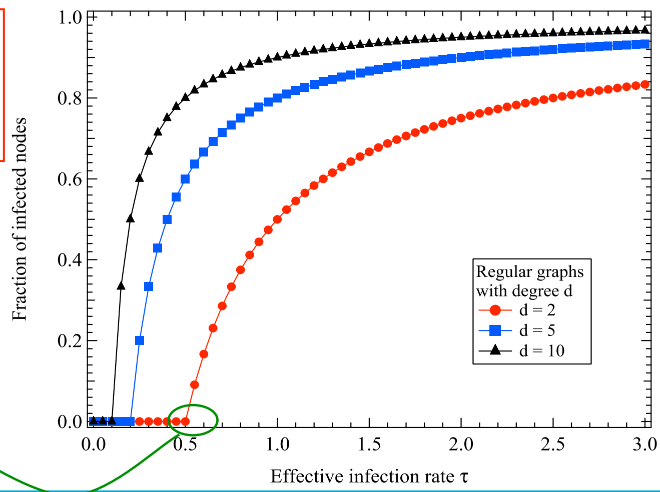
Comparison with simulation



TUDelft

What is so interesting about epidemics?

β : infection rate per link
δ : curing rate per node
τ= β/ δ : effective spreading rate



- **Final epidemic state**
- **Rate of propagation**
- **Epidemic threshold**

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

$$E[D] = \frac{2L}{N} \le \lambda_1(A) \le d_{\max}$$

TUDelft

Application example: distributed storage

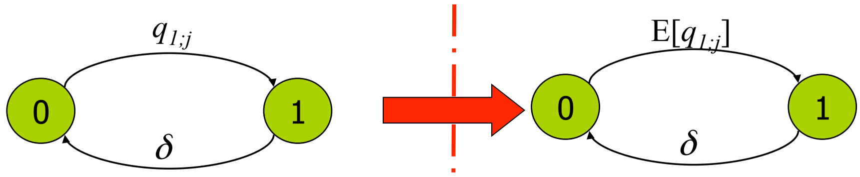
- each network node reserves memory space for other nodes
- each node spreads own files with rate β
- each node deletes files from neighbors with rate δ

$$\tau = \frac{\beta}{\delta} > \tau_c = \frac{1}{\lambda_1(A)}$$

Then your file is stored at node *j* in the network with probability $v_j > 0$

TU Delft

Mean-field approximation



- 2^N linear equations
- Steady-state
  - absorbing (healthy) state
  - reached after unrealistically long time
- difficult to analyze

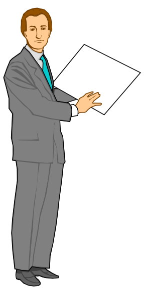
- *N* non-linear equations
- Meta-stable state:
  - phase-transition
  - epidemic threshold
  - realistic
- analytically tractable

TU Delft

Open question

- What is the accuracy of the *N*-intertwined model?
  - due to single approximation of a mean-field type
- Why is this a difficult to determine:
  - exact Markov chain cannot be compute for $N > 20$
  - simulations have own "accuracy limits"
  - mathematical derivation (contact networks)
    - assume asymptotic analyses ($N \rightarrow$ infinity)
    - absorbing state is reached with almost zero probability
    - only very few results exist (for specific graphs)

Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

Viral Conductance

Modifying the epidemic threshold

Summary

Heterogeneous virus spread

- The N-intertwined model is extended to a hetergeneous setting:

$$\boxed{\frac{dV(t)}{dt} = A\,diag(\beta_i)V(t) - diag(v_i(t))\left(A\,diag(\beta_i)V(t) + C\right)}$$

where the curing rate vector $C^T = [\delta_1\ \delta_2 \ldots \delta_N]$

- *Results:*
  - Extended multi-dim. threshold for virus spread
  - Generalized Laplacian that extends the classical Laplacian of a graph: $\boxed{Q(q_k) = diag(q_k) - A}$
  - Strong convexity v_k with respect to δ_k, concave with respect to others δ_j (j different from k).
  - Choose C vector in network protection via game theory

27

- J. Omic, P. Van Mieghem, and A. Orda, "Game Theory and Computer Viruses", *IEEE Infocom09.*
- P. Van Mieghem and J. Omic, "In-homogeneous Virus Spread in Networks", TUDelft report (see my website)

TUDelft

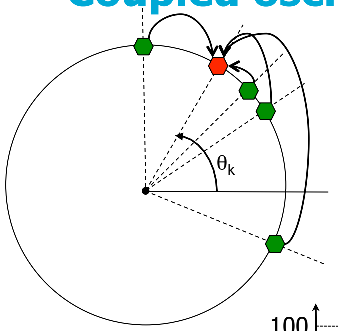
Extension of the N-intertwined model

- SAIS instead of SIS:
  - From 2 states (Infected and Susceptible) to a 3-states (Infected, Susceptible, Alert)
  - "Epidemic Spread in Human Networks", F. Darabi Sahneh and C. Scoglio, 50th IEEE Conf. Decision and Contol, Orlando, Florida (2011)
- SIR instead of SIS:
  - "R" state for "removed"
  - "An individual-based approach to SIR epidemics in contact networks", M. Youssef and C. Scoglio, Journal of Theoretical Biology 283, pp. 136-144, (2011).

28

TUDelft

Coupled oscillators (Kuramoto model)

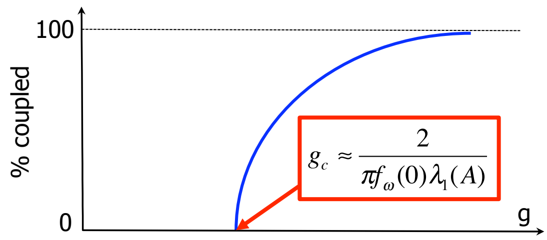


Interaction equals sums of sinus of phase difference of each neighbor:

$$\dot{\theta}_k = \omega_k + g\sum_{j=1}^{N} a_{kj}\sin\left(\theta_j - \theta_k\right)$$

coupling strength

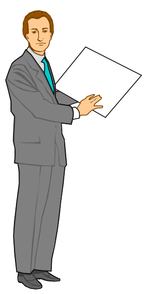
natural frequency



J. G. Restrepo, E. Ott, and B. R. Hunt. Onset of synchronization in large networks of coupled oscillators, Phys. Rev. E, vol. 71, 036151, 2005

TUDelft

Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

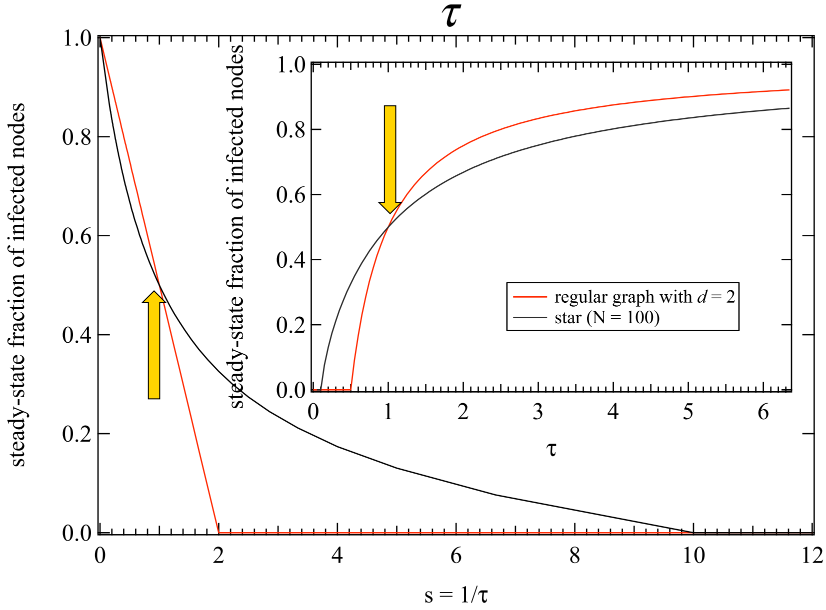
Viral Conductance

Modifying the epidemic threshold

Summary

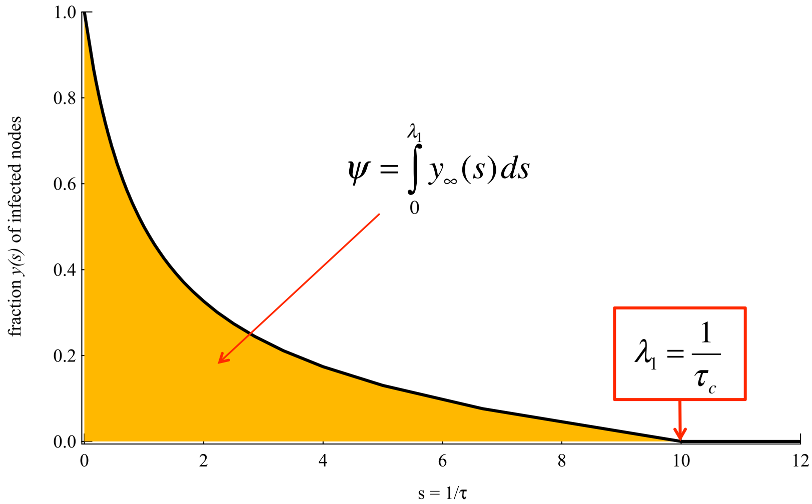
TUDelft

Transformation $s = \frac{1}{\tau}$



TU Delft

Viral Conductance: definition



$$\psi = \int_0^{\lambda_1} y_\infty(s)\,ds$$

$$\boxed{\lambda_1 = \frac{1}{\tau_c}}$$

32

R. E. Kooij, P. Schumm, C. Scoglio and M. Youssef, "A new metric for robustness of networks with respect to virus spread", IFIP Networking 2009

TU Delft

Viral Conductance: properties

$$\psi = \int_0^{\lambda_1} y_\infty(s)\,ds$$

Bounds: $\frac{1}{2} d_{\min} \le \psi \le \frac{1}{2}\lambda_1$

Viral Robustness metrics: (τ_c, ψ)

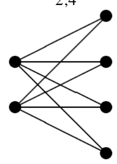
What are the graphs with maximum and minimum ψ ?

Some analytic evaluations:

$$\psi_{regular} = \frac{d}{2}$$

$$\psi_{K_{m,n}} = \sqrt{mn} - \frac{mn}{m+n} + \frac{(n-m)}{m+n}\log\frac{\left(1+\sqrt{n/m}\right)^m}{\left(1+\sqrt{m/n}\right)^n}$$

$K_{2,4}$

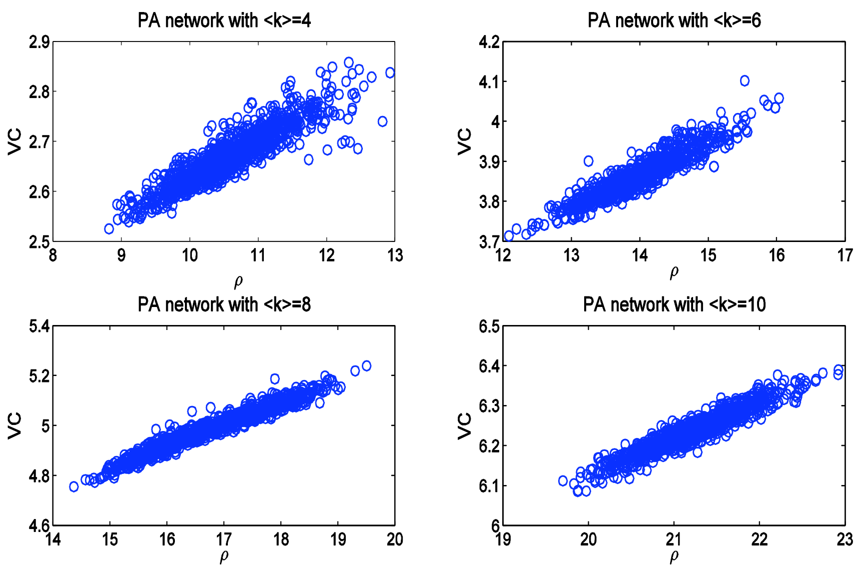


33

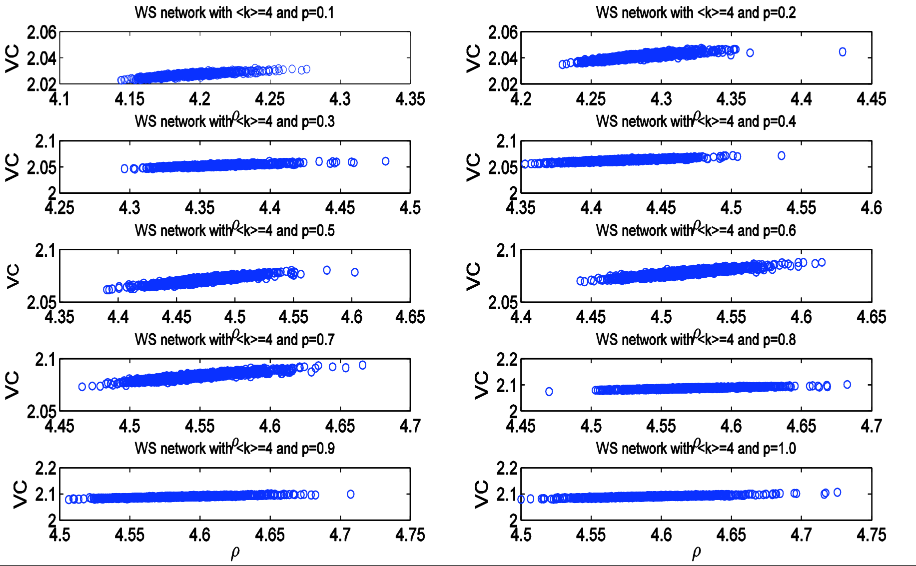
M. Youssef, R. E. Kooij and C. Scoglio, "Viral conductance: Quantifying the robustness of networks with respect to spread of epidemics", Journal of computational science, to appear 2011

TU Delft

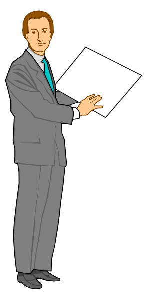
Preferential Attachment nets with E[D]=<k>= 4, 6, 8, and 10



Watts-Strogatz nets with <k>=4 and $0 \le p \le 1$



Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

Viral Conductance

Modifying the epidemic threshold

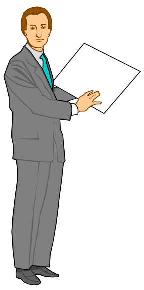
Summary

TU Delft

Affecting the epidemic threshold

- Degree-preserving rewiring
  - Changing the assortativity of the graph
  - Van Mieghem, P., H. Wang, X. Ge, S. Tang and F. A. Kuipers, 2010, "Influence of Assortativity and Degree-preserving Rewiring on the Spectra of Networks", The European Physical Journal B, vol. 76, No. 4, pp. 643-652.
- Removing links/nodes (optimal way is NP-complete)
  - Van Mieghem, P., D. Stevanovic, F. A. Kuipers, C. Li, R. van de Bovenkamp, D. Liu and H. Wang, 2011, "Decreasing the spectral radius of a graph by link removals", Physical Review E, Vol. 84, No. 1, July, p. 016101.
- Quarantining: Removing inter-module links
  - Omic, J., J. Martin Hernandez and P. Van Mieghem, 2010, "Network protection against worms and cascading failures using modularity partitioning", 22nd International Teletraffic Congress (ITC 22), September 7-9, Amsterdam, Netherlands.

Outline



Introduction & Definitions

The N-intertwined model for virus spread

Extensions

Viral Conductance

Modifying the epidemic threshold

Summary

Summary (1)

- Epidemic theory can model many processes: information spread, real viruses, storage, emotions in social nets, ...
- Real epidemics: phase transition at $\tau_c = 1/\lambda_1$
- The topology plays an important role in spreading
- Networks can be designed to protect individual nodes via its curing strength δ (virus software, fire-wall, etc...). Protection, however, does cost money...
- Epidemic threshold engineering:
  - Degree-preserving *assortative* rewiring increases λ_1, while degree-preserving *disassortative* rewiring decreases λ_1.
  - Removing links/nodes to maximally decrease λ_1 is NP-hard.

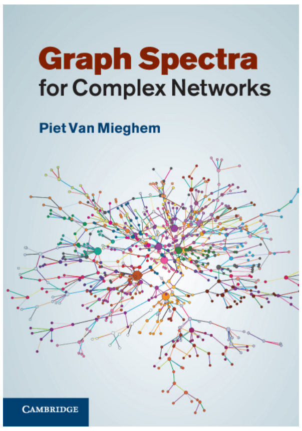
TU Delft

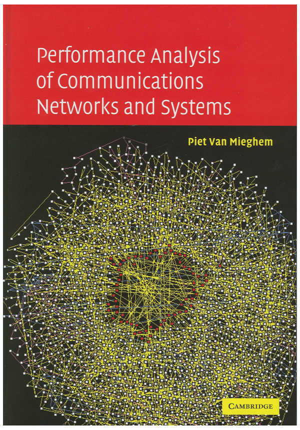
Summary (2)

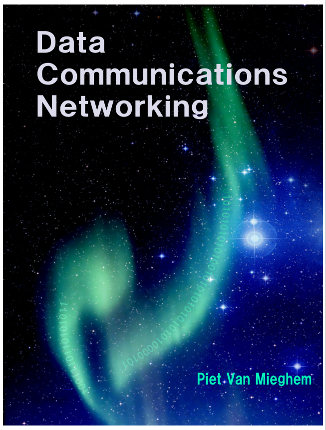
- *Agenda for future research:*
  - accuracy of *N*-intertwined model, i.e. of the mean-field approximation
  - coupling of the virus spread process and the underlying topology
  - most general extensions of the *N*-intertwined model (to $m > 2$ states per node)
  - Multiple, simultaneous viruses on a network

TU Delft

Books







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

TUDelft

IWSOS 2012 in Delft: March 15-16

supported by

euronf NETWORK OF EXCELLENCE ifip

IWSOS 2012

Sixth International Workshop on Self-Organizing Systems
Delft, The Netherlands, March 15-16, 2012

TUDelft

Thank You

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

TU Delft