Performance Evaluation

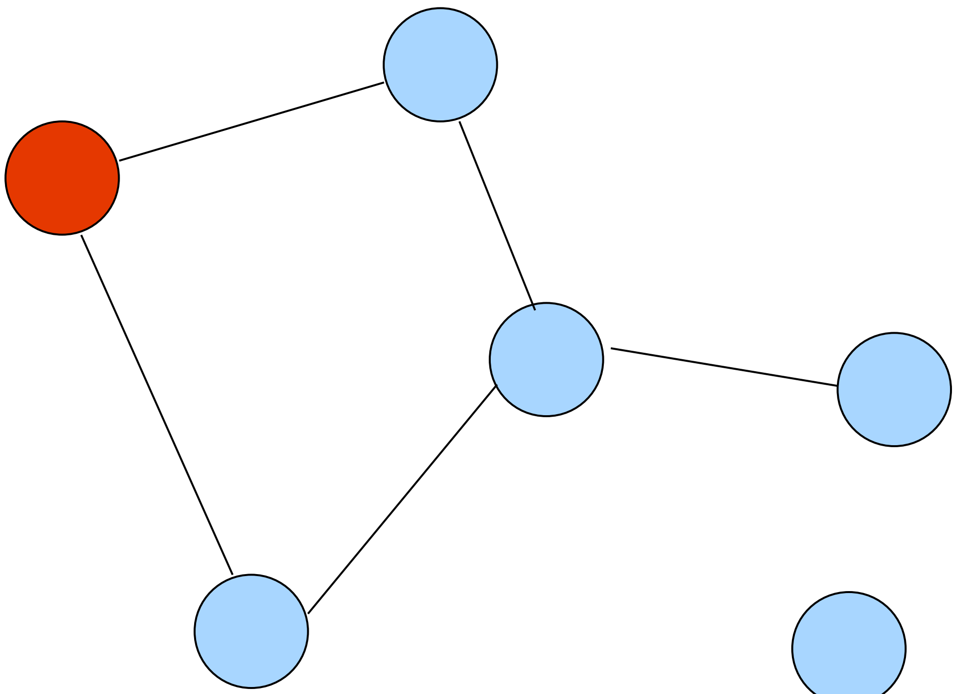
Lecture 4: Epidemics

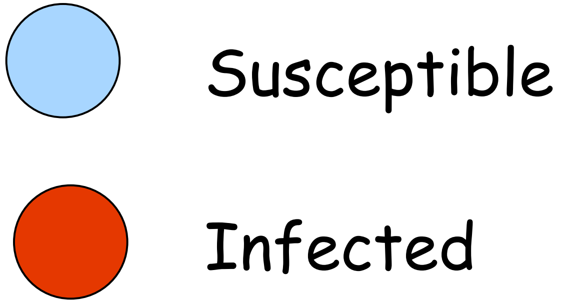
Giovanni Neglia
INRIA – EPI Maestro
12 January 2013

Outline

- Limit of Markovian models
- Mean Field (or Fluid) models
  - exact results
  - Extensions
    - Epidemics on graphs
      - Reference: ch. 9 of Barrat, Barthélemy, Vespignani "Dynamical Processes on Complex Networks", Cambridge press
  - Applications to networks

SI on a graph

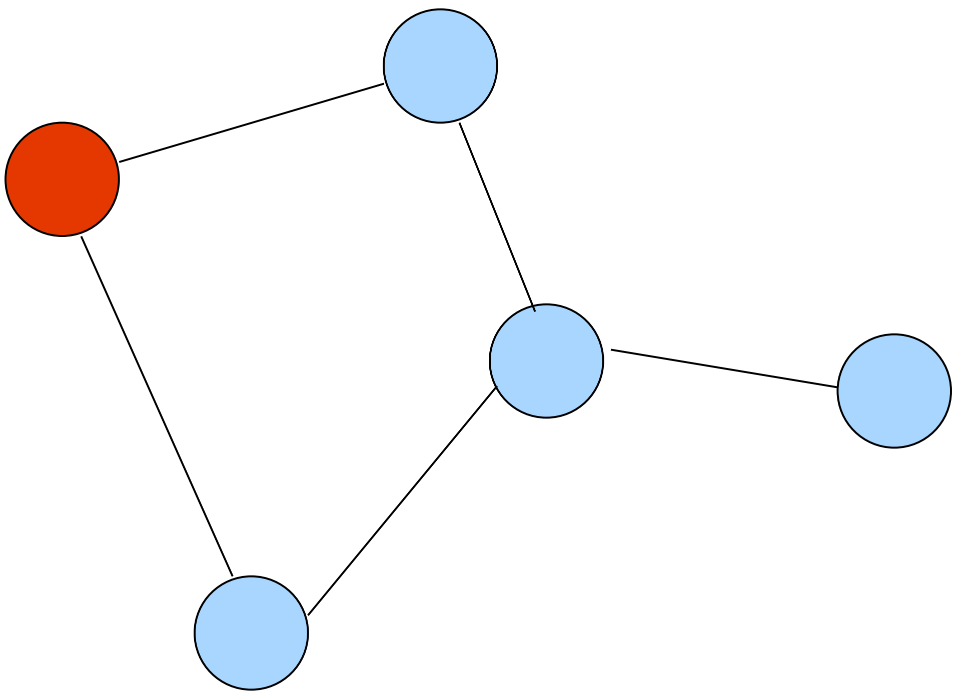




At each time slot, each link outgoing from an infected node spreads the disease with probability p_g

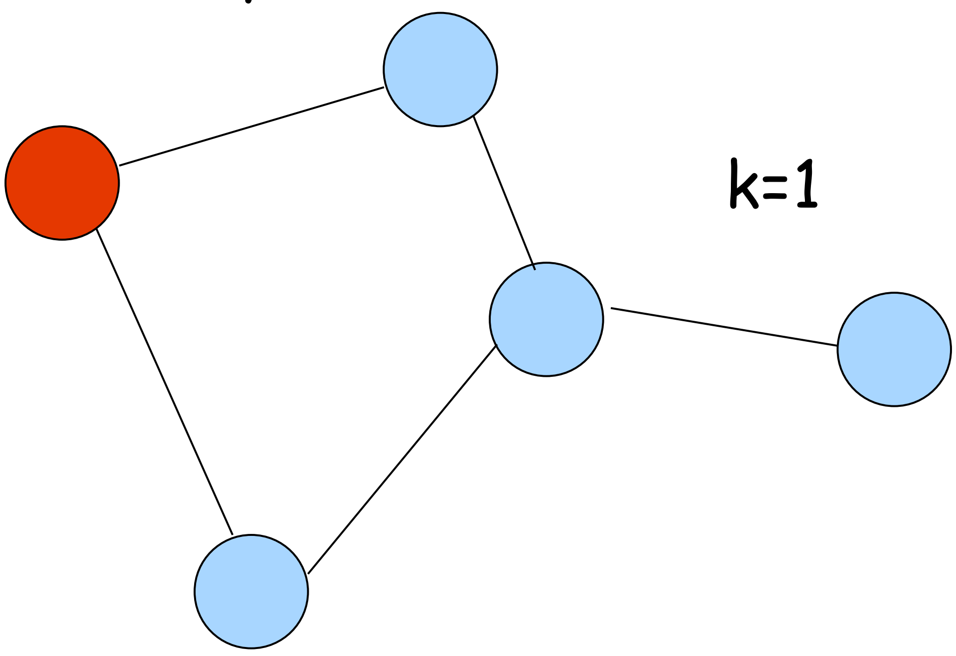
Can we apply Mean Field theory?

- Formally not, because in a graph the different nodes are not equivalent…
- …but we are stubborn



Derive a Mean Field model

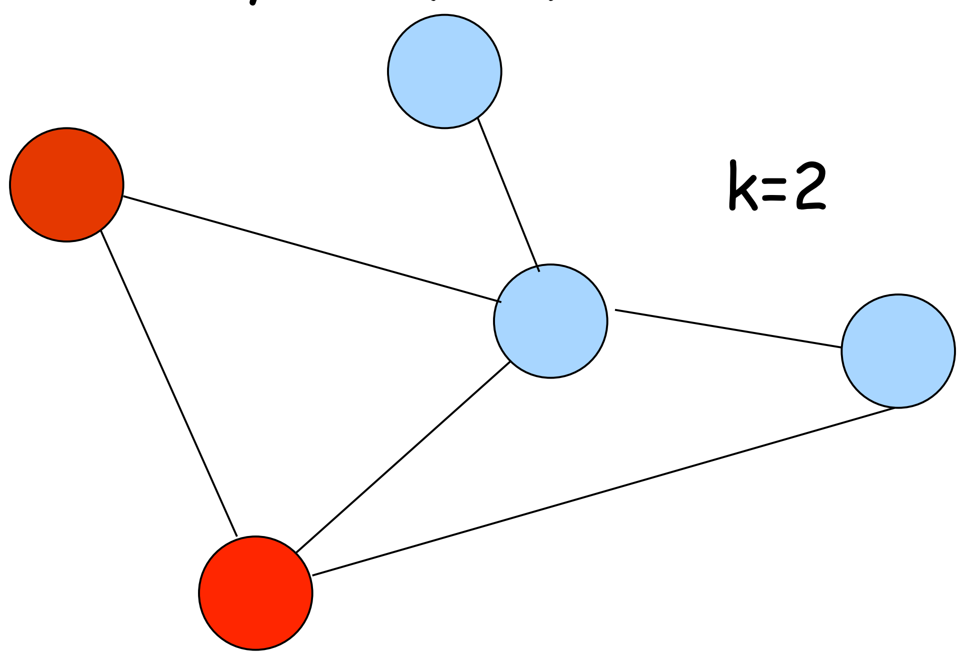
- Consider all the nodes equivalent
- e.g. assume that at each slot the graph changes, while keeping the average degree <d>
    - Starting from an empty network we add a link with probability <d>/(N-1)



k=1

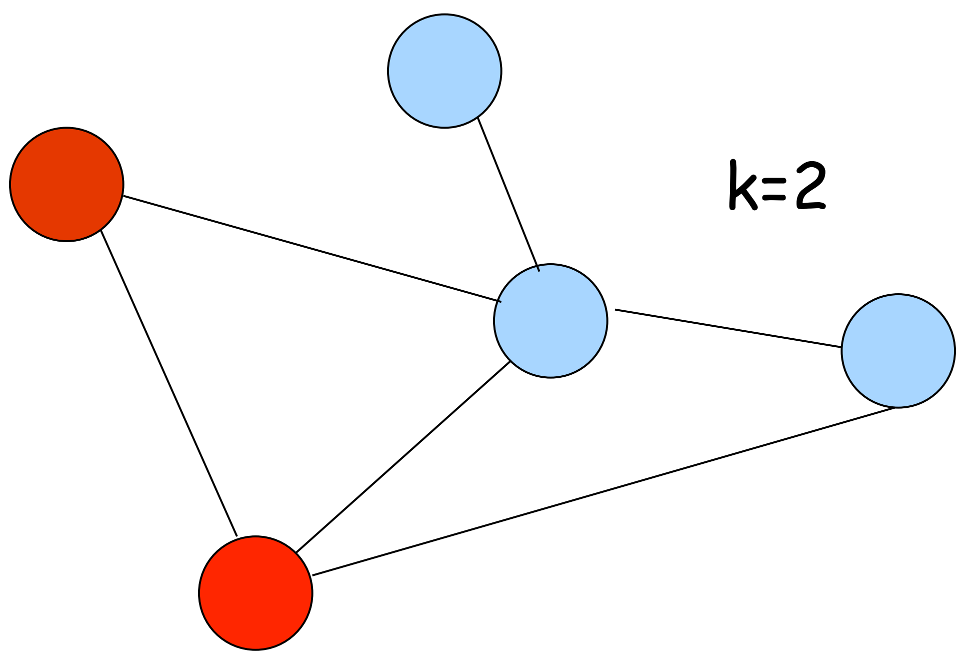
Derive a Mean Field model

- Consider all the nodes equivalent
- e.g. assume that at each slot the graph changes, while keeping the average degree <d>
  - Starting from an empty network we add a link with probability <d>/(N-1)



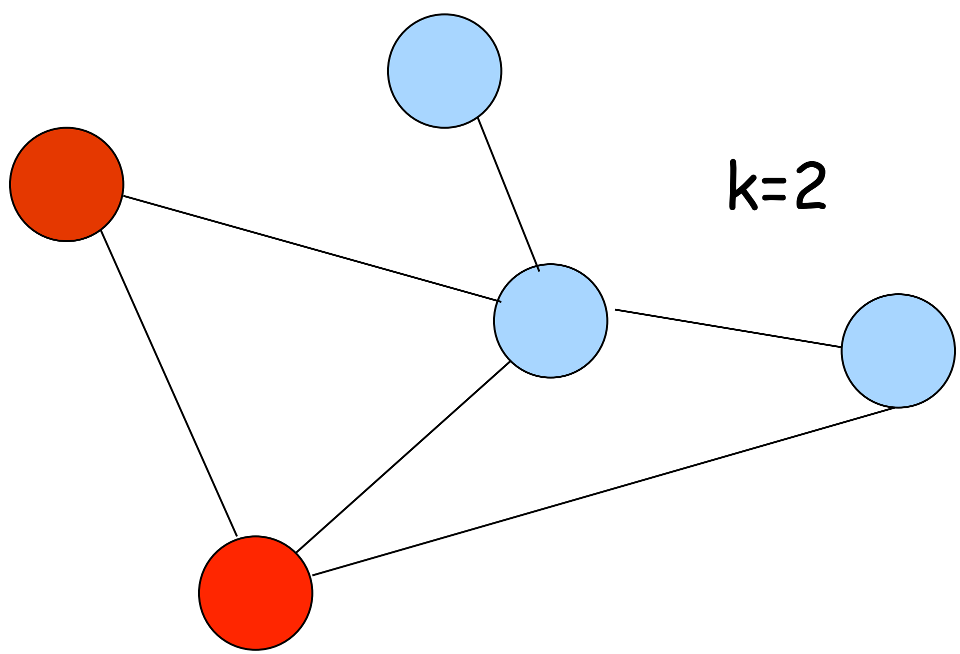
Derive a Mean Field model

- i.e. at every slot we consider a sample of an ER graph with N nodes and probability <d>/(N-1)
  - Starting from an empty network we add a link with probability <d>/(N-1)



Derive a Mean Field model

- If $I(k)=I$, the prob. that a given susceptible node is infected is $q_I=1-(1-\langle d\rangle/(N-1)\ p_g)^I$
- and $(I(k+1)-I(k)|I(k)=I) =_d Bin(N-I,\ q_I)$



Derive a Mean Field model

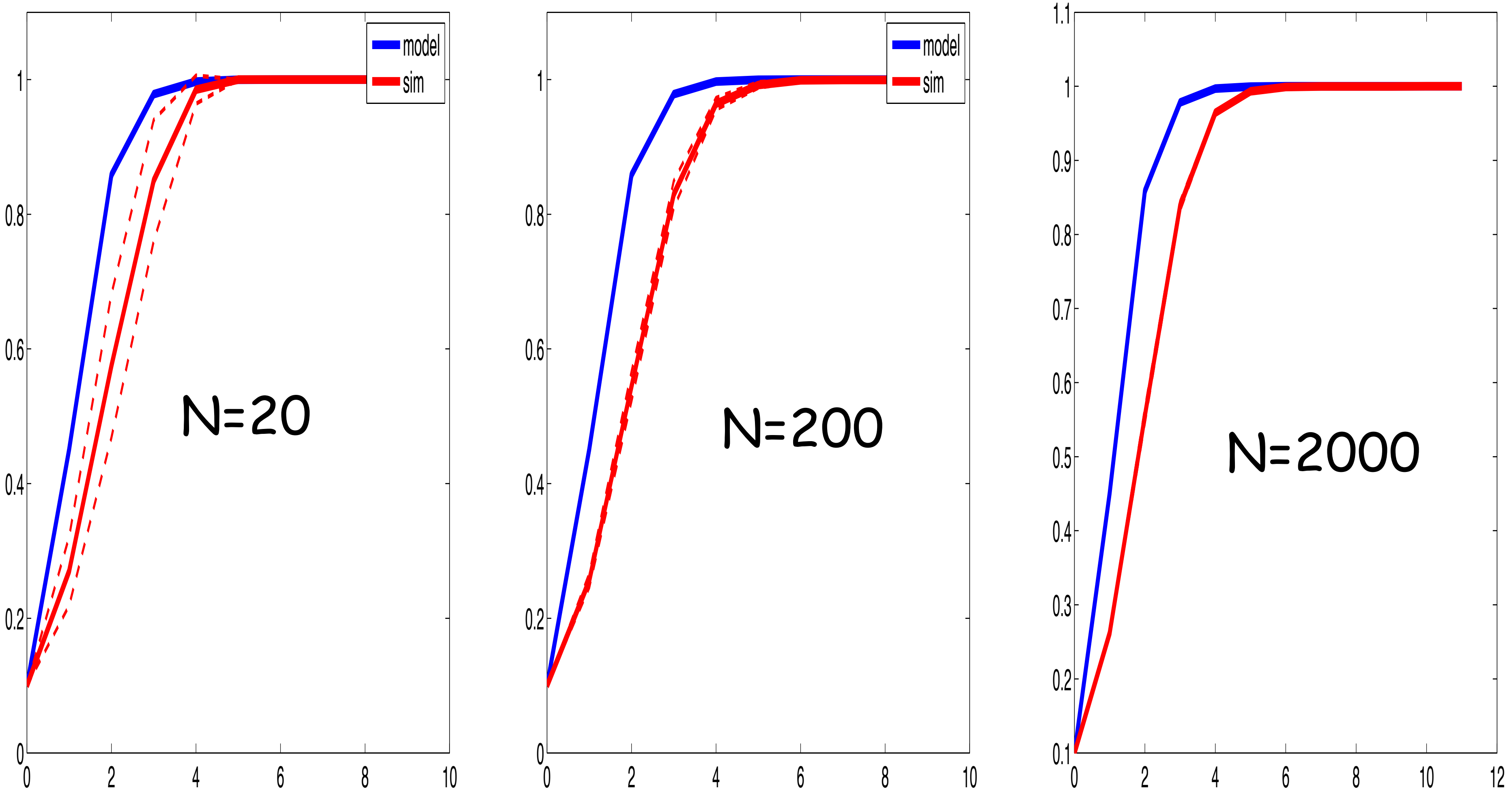
- If $I(k)=I$, the prob. that a given susceptible node is infected is $q_I=1-(1-\langle d\rangle/(N-1)\ p_g)^I$
- and $(I(k+1)-I(k)|I(k)=I) =_d \mathrm{Bin}(N-I,\ q_I)$
  - Equivalent to first SI model where $p=\langle d\rangle/(N-1)\ p_g$
  - We know that we need $p^{(N)}=p_0/N^2$
- $i^{(N)}(k) \approx \mu_2\,(k\varepsilon(N))=1/((1/i_0-1)\exp(-k\ p_0/N)+1)=$ $= 1/((1/i_0-1)\exp(-k\ \langle d\rangle\ p_g)+1)$
  - The percentage of infected nodes becomes significant after the **outbreak time** $1/(\langle d\rangle p_g)$
- How good is the approximation practically?
  - It depends on the graph!

Let's try on Erdös-Rényi graph

- Remark: in the calculations above we had a different sample of an ER graph at each slot, in what follows we consider a single sample

ER $\langle d \rangle$=20, p_g=0.1, 10 runs

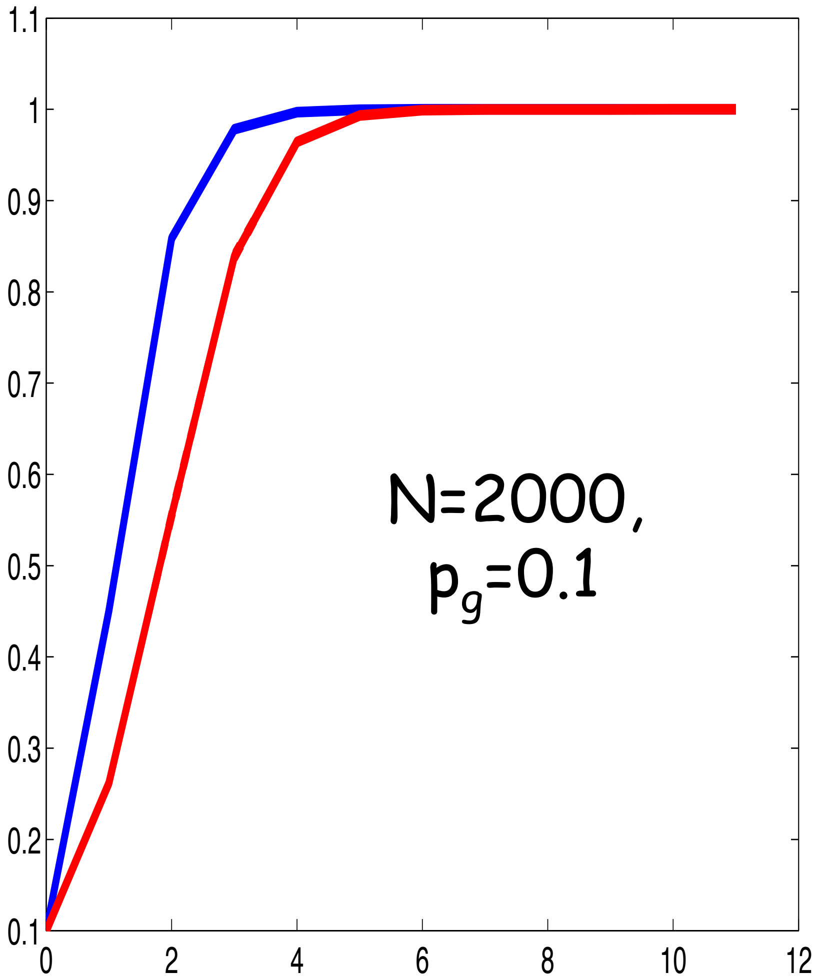
$$i^{(N)}(k) \approx 1/((1/i_o - 1)\exp(-k \langle d \rangle p_g) + 1)$$

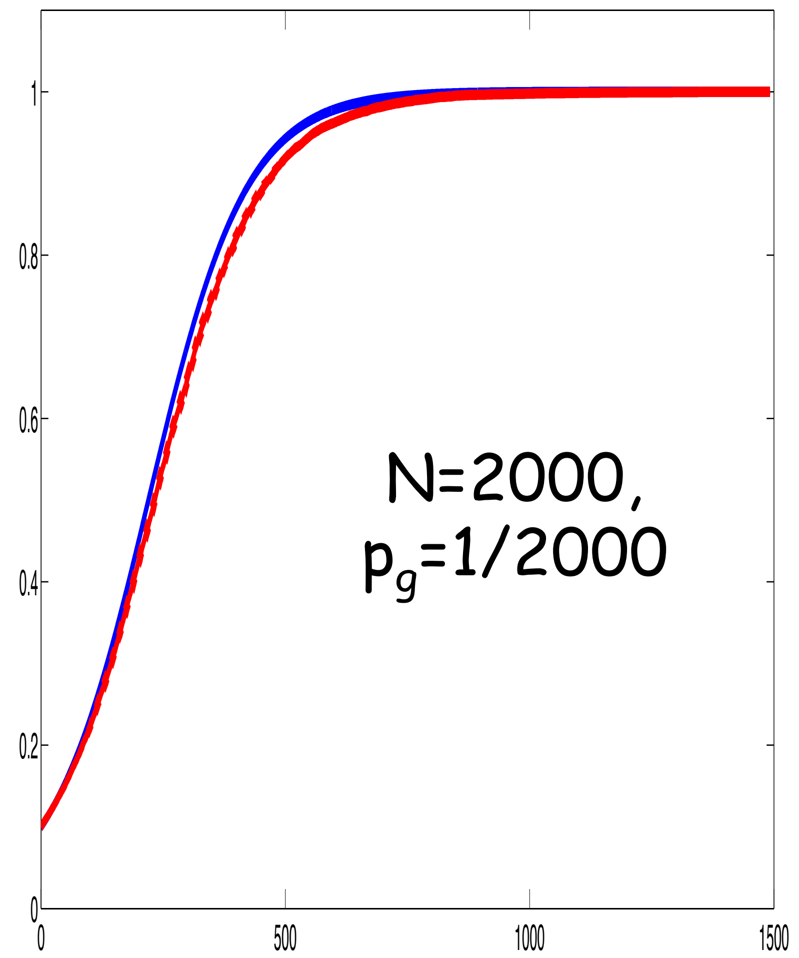


Lesson 1

- System dynamics is more *deterministic* the larger the network is
- For given <d> and p_g, the MF solution shows the same relative error

ER <d>=20, 10 runs

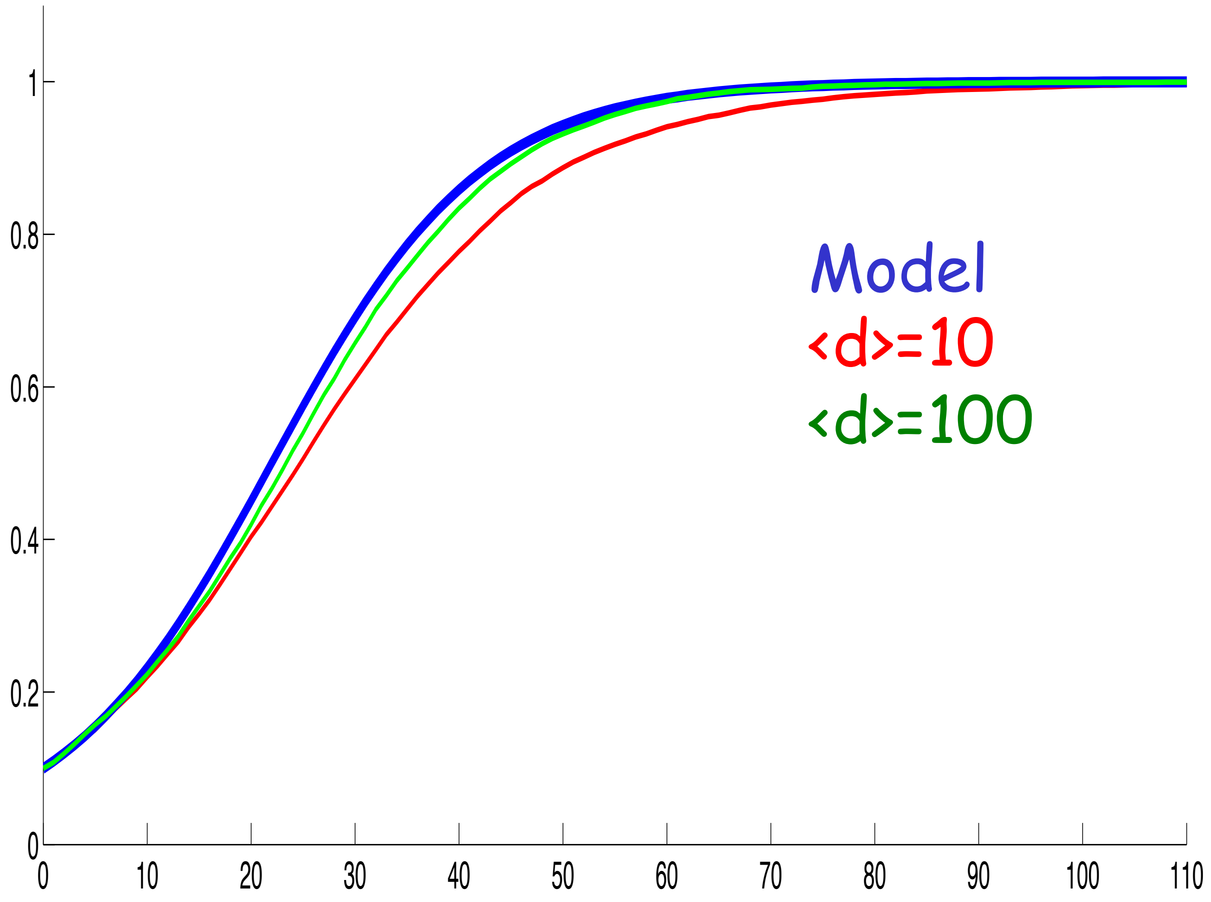




Lesson 2

- For given <d>, the smaller the infection probability p_g the better the MF approximation
  - Why?

Changing the degree
ER N=1000, <d>p_g=0.1, 10 runs

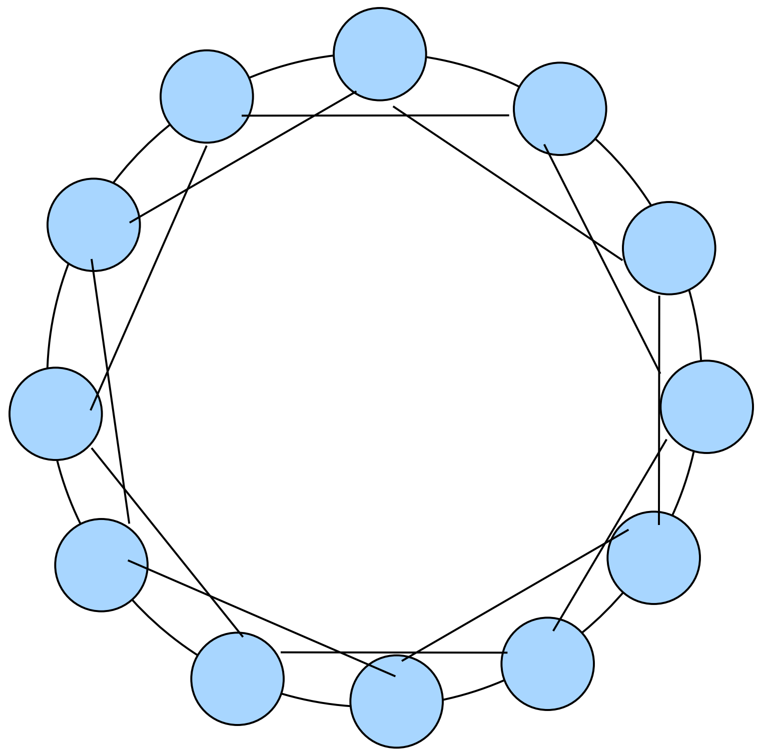


$$i^{(N)}(k) \approx 1/((1/i_0-1)\exp(-k \langle d \rangle p_g)+1)$$

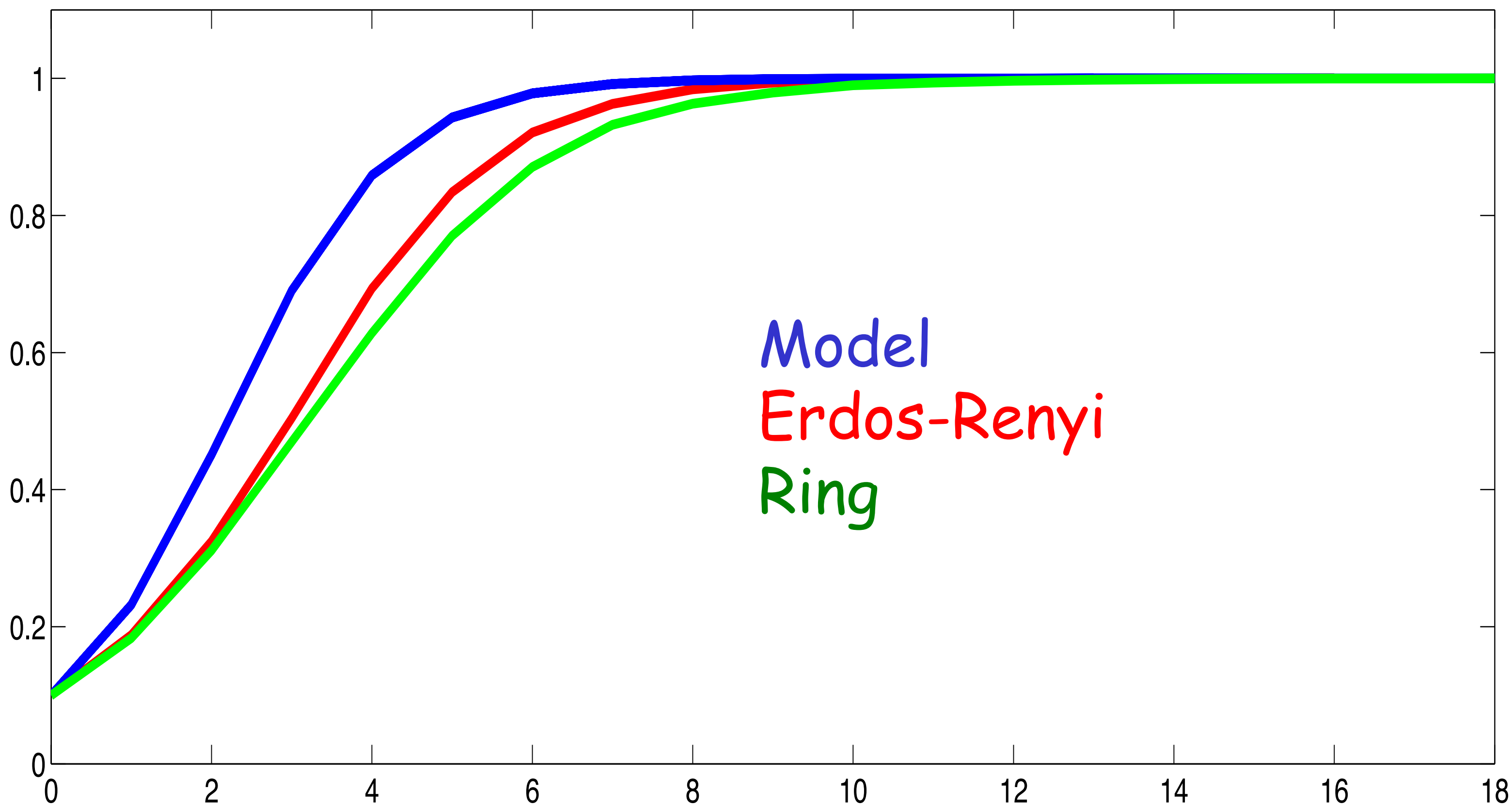
Lesson 3

- Given $\langle d \rangle p_g$, the more the graph is connected, the better the MF approximation
  - Why?

A different graph Ring(N,k)



Ring vs ER, N=2000, <k>=10



Lesson 4

- The smaller the clustering coefficient, the better the MF approximation
  - Why?

Heterogeneous Networks

- Denote P(d) the probability that a node has degree d
- If the degree does not change much, we can replace d with <d>
  - what we have done for ER graphs (N,p)
    - Binomial with parameters (N-1,p)
- How should we proceed (more) correctly?
  - Split the nodes in degree classes
  - Write an equation for each class
- Remark: following derivation will not be as rigorous as previous ones

Heterogeneous Networks

- N_d number of nodes with degree d (=N*P(d))
- I_d: number of infected nodes with degree d
- Given node i with degree d and a link e_{ij}, what is the prob. that j has degree d'?
  - P(d')? NO
- and if degrees are uncorrelated? i.e. Prob(neighbour has degree d'|node has a degree d) independent from d,
  - P(d')? NO
  - Is equal to $d'/\langle d\rangle\, P(d')$

Heterogeneous Networks

- Given node i with degree d and a link e_{ij}
- Prob. that j has degree d′ is
  - $d'/\langle d\rangle\, P(d')$
- Prob. that j has degree d′ and is infected
  - $d'/\langle d\rangle\, P(d')\, I_{d'}/N_{d'}$
  - more correct $(d'-1)/\langle d\rangle\, P(d')\, I_{d'}/N_{d'}$
- Prob. that i is infected through link e_{ij} is
  - $p = p_g \sum_{d'} (d'-1)/\langle d\rangle\, P(d')\, I_{d'}/N_{d'}$
- Prob. that i is infected through one link
  - $1-(1-p)^d$

Heterogeneous Networks

- $E[(I_d(k+1)-I_d(k)|\mathbf{I}(k)=\mathbf{I})] = (N_d-I_d)(1-(1-p)^d)$
  - $p = p_g \Sigma_{d'} (d'-1)/\langle d\rangle P(d')\, I_{d'}/N_{d'}$
- $f_d^{(N)}(i)=(1-i_d)(1-(1-p)^d)$
  - $i_d = I_d/N_d$
  - if we choose $p_g = p_{g0}/N$
  - $f_d(\mathbf{i})= p_{g0}(1-i_d)\, d\, \underbrace{\Sigma_{d'}(d'-1)/\langle d\rangle P(d')\, i_{d'}}_{\Theta}$
- $di_d(t)/dt=f_d(\mathbf{i}(t))=p_{g0}(1-i_d(t))\, d\, \Theta(t)$

Heterogeneous Networks

- $di_d(t)/dt = f_d(i(t)) = p_{g0}\,(1-i_d(t))\,d\,\Theta(t)$,
  - for $d=1,2\ldots$
  - $\Theta(t)=\Sigma_{d'}(d'-1)/\langle d\rangle\,P(d')\,i_{d'}(t)$
  - $i_d(0)=i_{d0}$, for $d=1,2\ldots$
- If $i_d(0) \ll 1$, for *small* t
  - $di_d(t)/dt \approx p_{g0}\,d\,\Theta(t)$
  - $d\Theta(t)/dt = \Sigma_{d'}(d'-1)/\langle d\rangle\,P(d')\,di_{d'}(t)/dt$

$$\approx p_{g0}\,\Sigma_{d'}(d'-1)/\langle d\rangle\,P(d')\,d'\,\Theta(t) =$$

$$= p_{g0}\,(\langle d^2\rangle - \langle d\rangle)/\langle d\rangle\,\Theta(t)$$

Heterogeneous Networks

- $d\Theta(t)/dt \approx p_{g0}(\langle d^2\rangle - \langle d\rangle)/\langle d\rangle\ \Theta(t)$
  - Outbreak time: $\langle d\rangle/((\langle d^2\rangle - \langle d\rangle)\ p_{g0})$
    - For ER $\langle d^2\rangle = \langle d\rangle(\langle d\rangle + 1)$, we find the previous result, $1/(\langle d\rangle p_{g0})$
    - What about for Power-law graphs, $P(d) \sim d^{-\gamma}$?
- For the SIS model:
  - $d\Theta(t)/d \approx p_{g0}(\langle d^2\rangle - \langle d\rangle)/\langle d\rangle\ \Theta(t) - r_0\ \Theta(t)$
  - Epidemic threshold: $p_{g0}\ (\langle d^2\rangle - \langle d\rangle)/(\langle d\rangle r_0)$

Outline

- Limit of Markovian models
- Mean Field (or Fluid) models
  - exact results
  - extensions
  - Applications
    - Bianchi's model
    - Epidemic routing

Decoupling assumption in Bianchi's model

- Assuming that retransmission processes at different nodes are independent
  - Not true: if node i has a large backoff window, it is likely that also other nodes have large backoff windows
- We will provide hints about why it is possible to derive a Mean Field model…
- then the decoupling assumption is guaranteed asymptotically

References

- Benaïm, Le Boudec, "A Class of Mean Field Interaction Models for Computer and Communication Systems", LCA-Report-2008-010
- Sharma, Ganesh, Key, "Performance Analysis of Contention Based Medium Access Control Protocols", IEEE Trans. Info. Theory, 2009
- Bordenave, McDonarl, Proutière, "Performance of random medium access control, an asymptotic approach", Proc. ACM Sigmetrics 2008, 1-12, 2008

Bianchi's model

- N nodes,
- K possible stages for each node, in stage i (i=1,…V) the node transmit with probability $q^{(N)}_i$ (e.g. $q^{(N)}_i = 1/W^{(N)}_i$)
- If a node in stage i experiences a collision, it moves to stage i+1
- If a node transmits successfully, it moves to stage 0

Mean Field model

- We need to scale the transmission probability: $q^{(N)}_i = q_i/N$
- $\mathbf{f}^{(N)}(\mathbf{m})=E[\mathbf{M}^{(N)}(k+1)-\mathbf{M}^{(N)}(k)|\mathbf{M}^{(N)}(k)=\mathbf{m}]$
- $f_1^{(N)}(\mathbf{m})=E[M_1^{(N)}(k+1)-M_1^{(N)}(k)|M_1^{(N)}(k)=\mathbf{m}]$
- $P_{idle}=\Pi_{i=1,\ldots V}(1-q_i^{(N)})^{m_i N}$
- The number of nodes in stage 1
  - increases by one if there is one successful transmission by a node in stage i<>1
  - Decreases if a node in stage 1 experiences a collision

Mean field model

- $P_{idle} = \Pi_{i=1,\dots V}(1-q_i^{(N)})^{m_i^N} \to \exp(-\Sigma_i q_i\, m_i)$
  - Define $\tau(m) = \Sigma_i q_i\, m_i$
- The number of nodes in stage 1
  - increases by one if there is one successful transmission by a node in stage i<>1
    - with prob. $\Sigma_{i>1}\, m_i\, N\, q_i^{(N)}\, P_{idle}/(1-q_i^{(N)})$
  - Decreases if a node in stage 1 experiences a collision
    - with prob. $m_1\, N\, q_1^{(N)}\, (1-P_{idle}/(1-q_1^{(N)})$
- $f_1^{(N)}(\mathbf{m}) = E[M_1^{(N)}(k+1) - M_1^{(N)}(k) \mid M_1^{(N)}(k) = \mathbf{m}] =$

  $= \Sigma_{i>1} m_i q_i^{(N)} P_{idle}/(1-q_i^{(N)})$

  $- m_1 q_1^{(N)}(1-P_{idle}/(1-q_1^{(N)}))$

Mean field model

- $P_{idle}=\Pi_{i=1,\dots V}(1-q_i^{(N)})^{m_i N}$ -> $\exp(-\Sigma_i q_i\, m_i)$
  - Define $\tau(\mathbf{m})= \Sigma_i q_i\, m_i$
- $f_1^{(N)}(\mathbf{m})=\Sigma_{i>1} m_i q_i^{(N)} P_{idle}/(1-q_i^{(N)})$
  $- m_1 q_1^{(N)}(1-P_{idle}/(1-q_1^{(N)}))$
- $f_1^{(N)}(\mathbf{m}) \sim 1/N \left(\Sigma_{i>1} m_i q_i\, e^{-\tau(\mathbf{m})} - m_1 q_1 (1-e^{-\tau(\mathbf{m})})\right)$
- $f_1^{(N)}(\mathbf{m})$ vanishes and $\varepsilon(N)=1/N$, continuously differentiable in $\mathbf{m}$ and in $1/N$
- This holds also for the other components
- Number of transitions bounded

=> We can apply the Theorem

Outline

- Limit of Markovian models
- Mean Field (or Fluid) models
  - exact results
  - extensions
  - Applications
    - Bianchi's model
    - Epidemic routing

Mean fluid for Epidemic routing (and similar)

1. Approximation: pairwise intermeeting times modeled as independent exponential random variables
2. Markov models for epidemic routing
3. Mean Fluid Models

Inter-meeting times under random mobility (from Lucile Sassatelli's course)

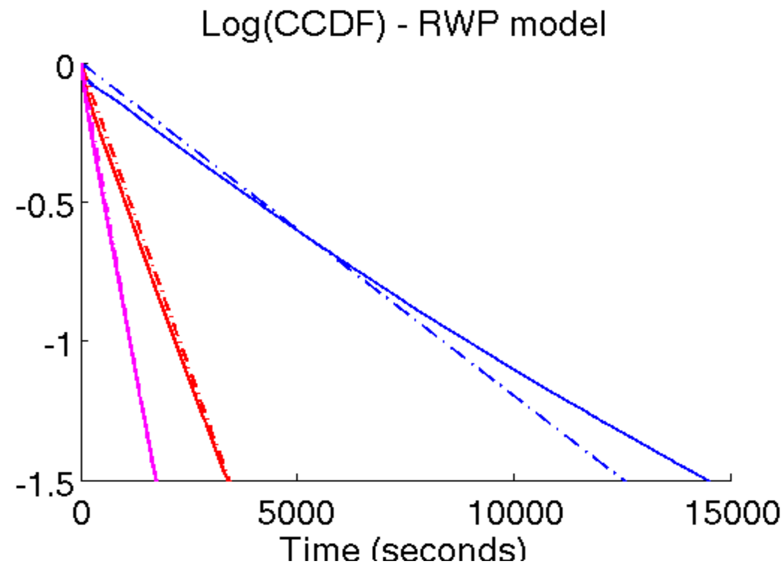
Inter-meeting times mobile/mobile have shown to follow an exponential distribution

[Groenevelt et al.: The message delay in mobile ad hoc networks. Performance Evalation, 2005]

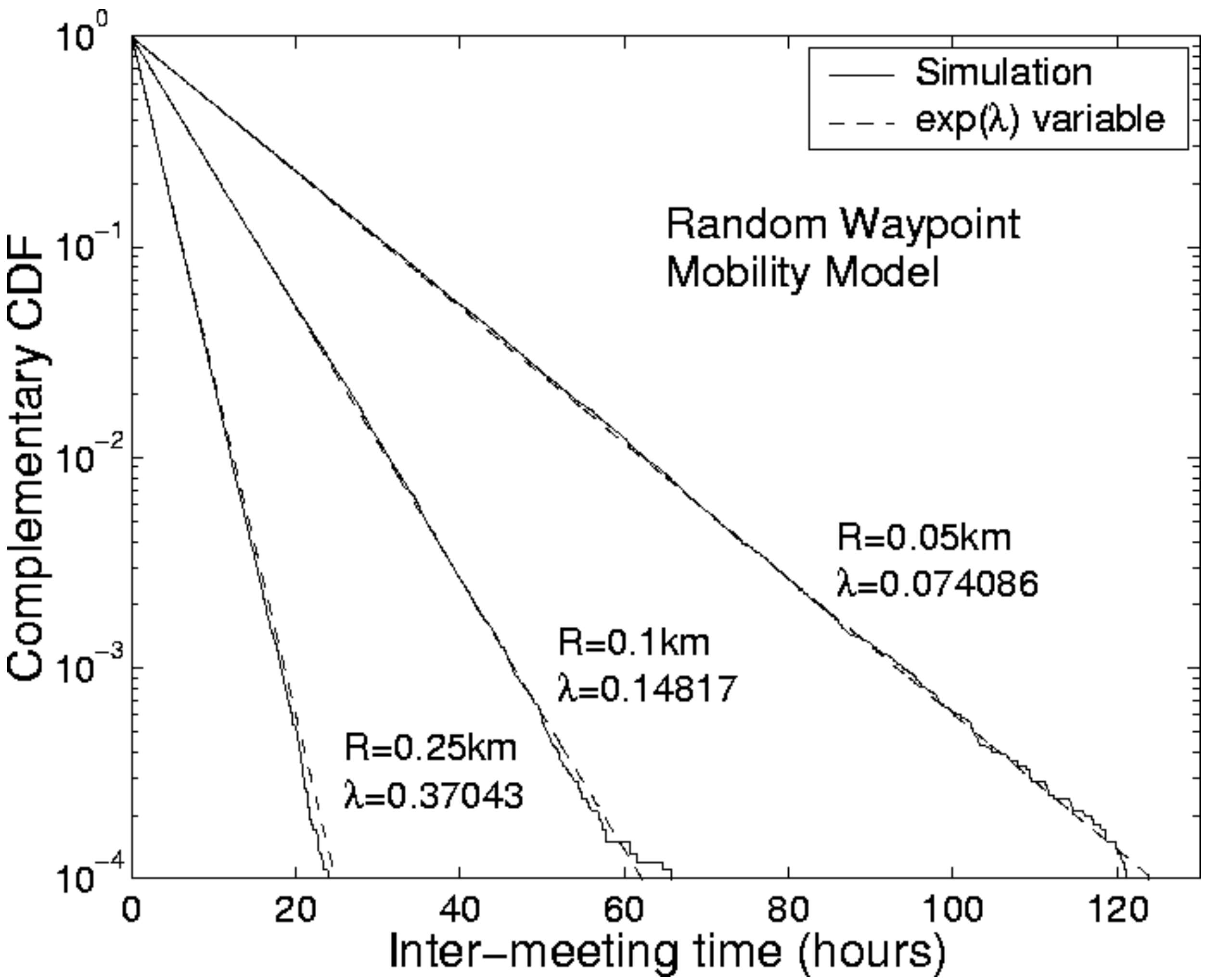
MAESTRO

$\Pr\{ X = x \} = \mu \exp(- \mu x)$

CDF: $\Pr\{ X \leq x \} = 1 - \exp(- \mu x)$, CCDF: $\Pr\{ X > x \} = \exp(- \mu x)$



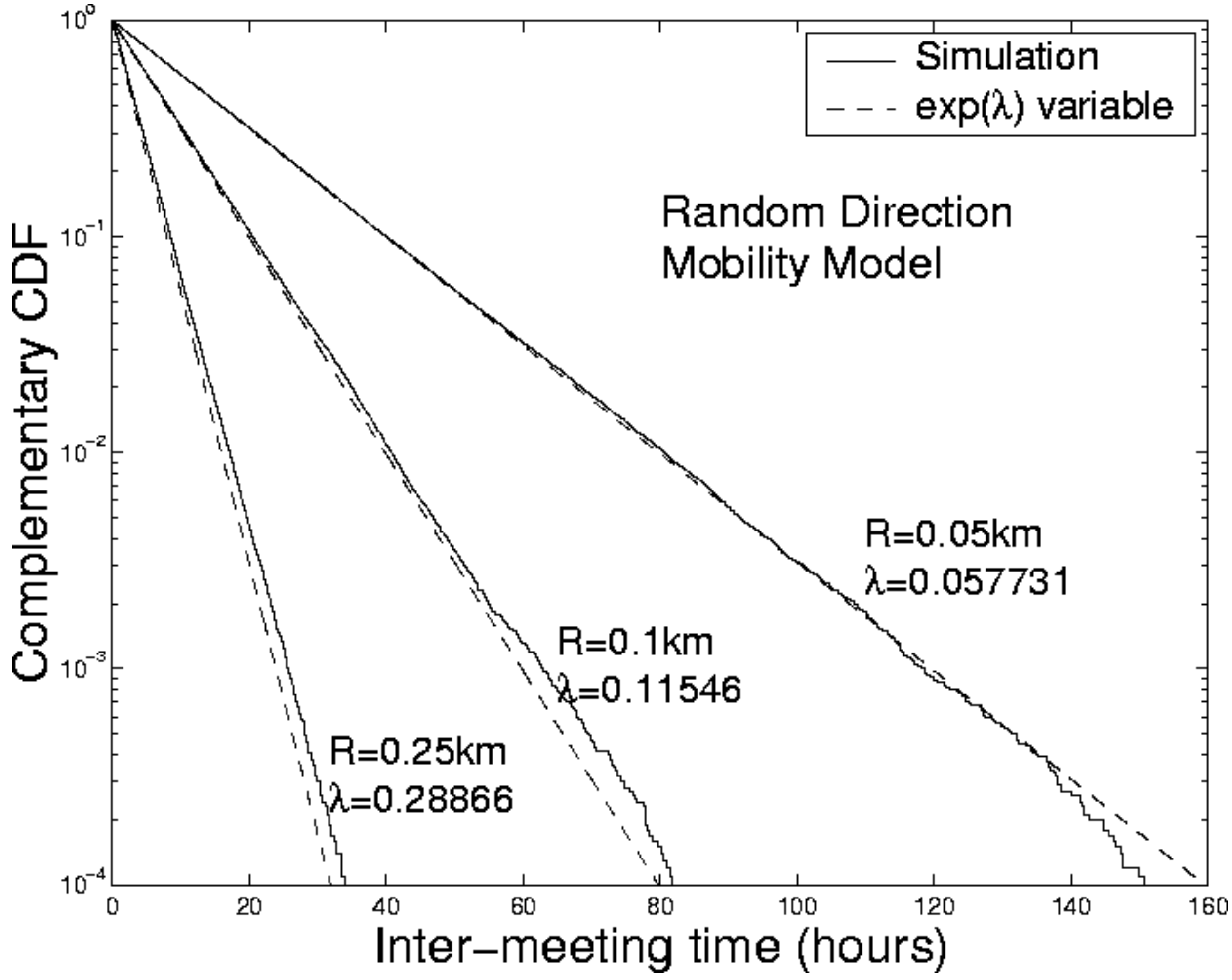
Pairwise Inter-meeting time



$$P(T > t) = e^{-\lambda t}$$

$$\log P(T > t) = -\lambda t$$

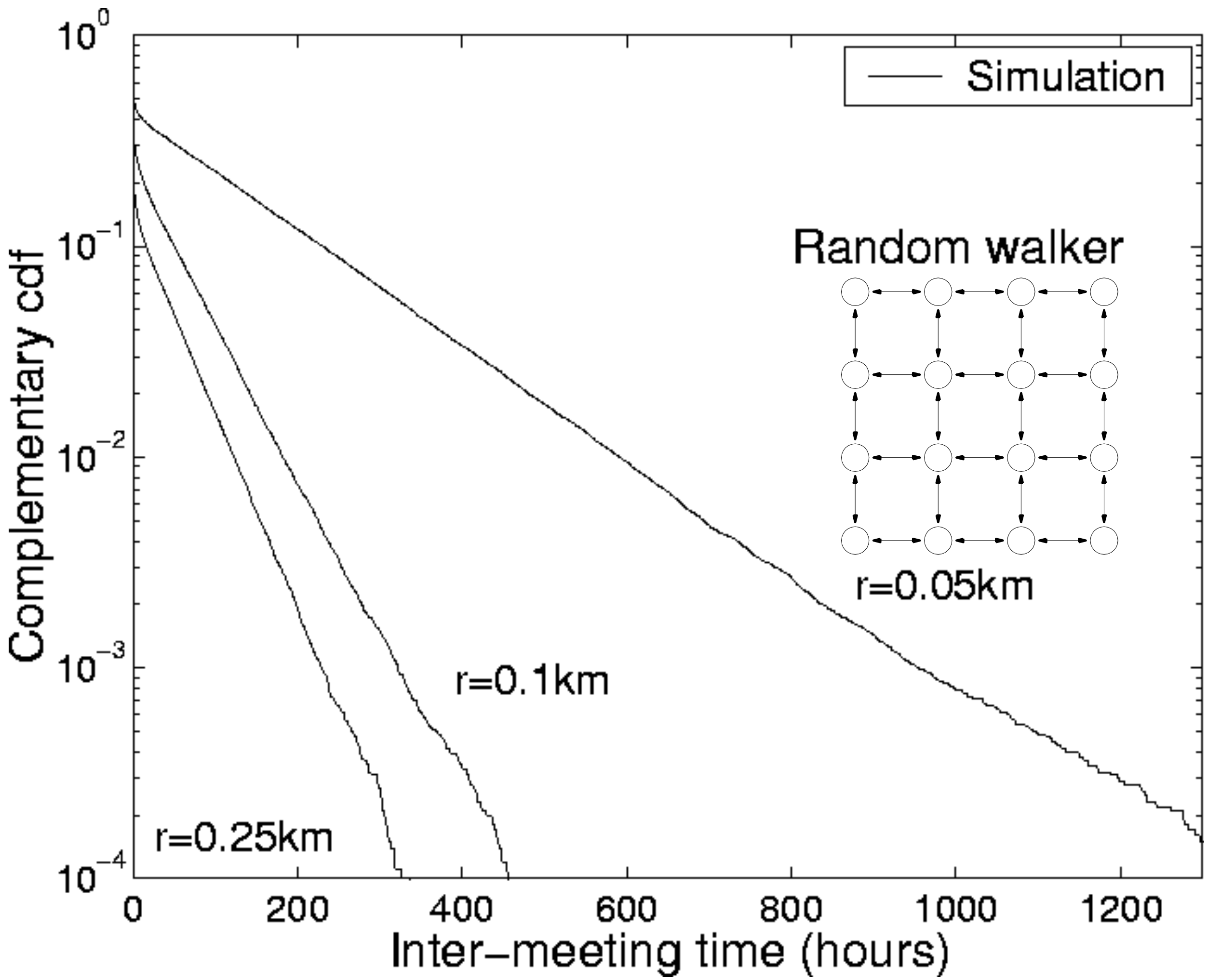
Pairwise Inter-meeting time



$$P(T > t) = e^{-\lambda t}$$

$$\log P(T > t) = -\lambda t$$

Pairwise Inter-meeting time

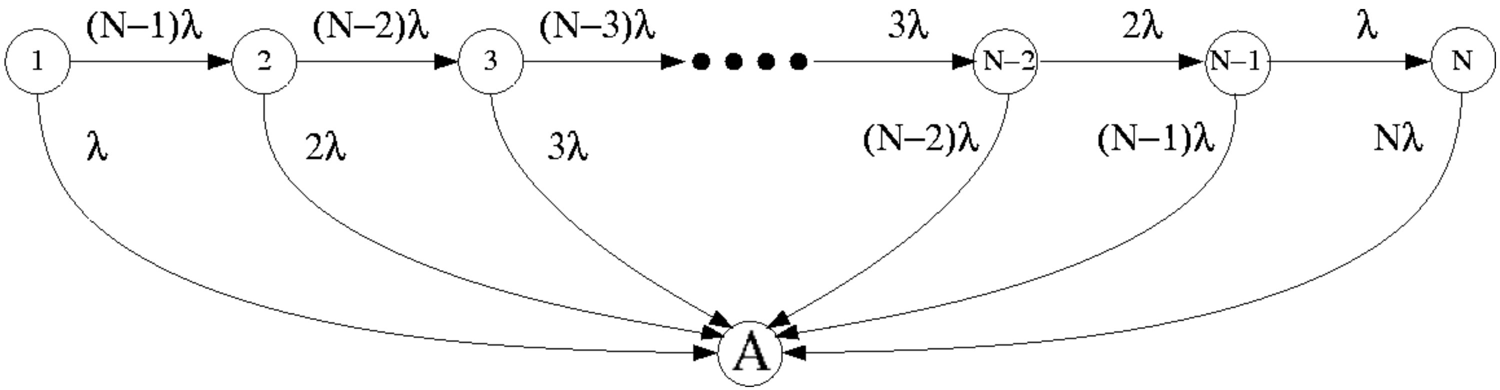


$$P(T > t) = e^{-\lambda t}$$

$$\log P(T > t) = -\lambda t$$

2-hop routing

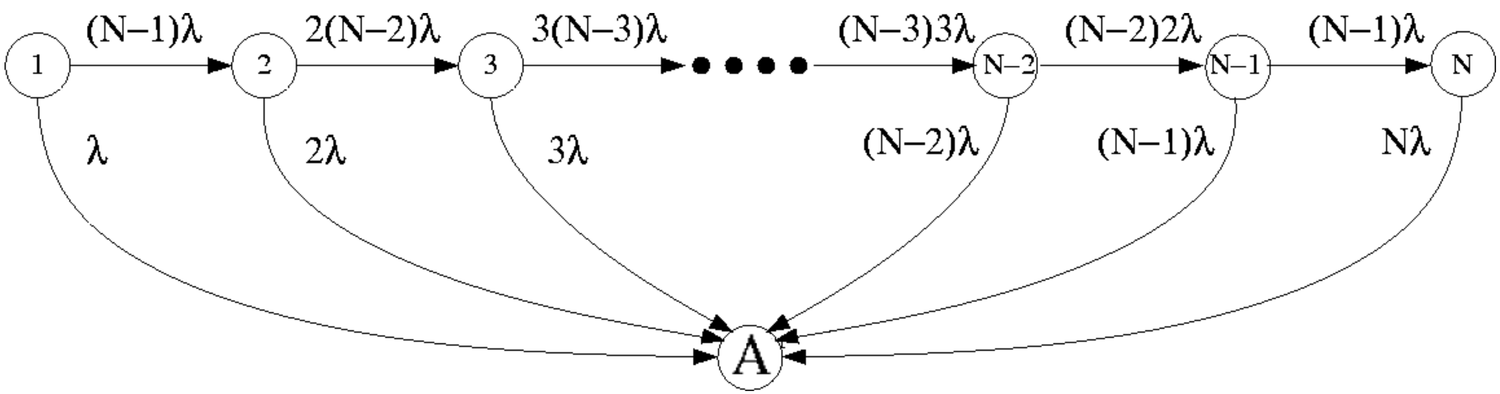
Model the number of occurrences of the message as an absorbing Continuous Time Markov Chain (C-MC):



- State i∈{1,…,N} represents the number of occurrences of the message in the network.
- State A represents the destination node receiving (a copy of) the message.

Epidemic routing

Model the number of occurrences of the message as an absorbing C-MC:



- State $i \in \{1,\ldots,N\}$ represents the number of occurrences of the message in the network.
- State *A* represents the destination node receiving (a copy of) the message.

A mean-field interaction model for modeling dissemination

(from Lucile Sassatelli's course)

- Time $t \in \mathbb{N}$ is discrete. There are N objects.
- Object n has state $Z_n^{(N)}(t)$ in $S = \{0, 1\}$.
- We assume that $\mathbf{Y}^{(N)}(t) = (Z_1^{(N)}(t), \ldots, Z_N^{(N)}(t))$ is a homogeneous Markov chain on S^N.
- We assume that we can observe the state of an object but not its label, i.e.,

$$\mathcal{K}^N(i_1, \ldots, i_N; i'_1, \ldots, i'_N) =$$

$$Pr\{Z_1^{(N)}(t+1) = i_1, \ldots, Z_N^{(N)}(t+1) = i_N | Z_1^{(N)}(t) = i'_1, \ldots, Z_N^{(N)}(t) = i'_N\}$$

is stable under any permutation.

$\rightarrow$ The process $\mathbf{Y}^{(N)}(t)$ is called a mean-field interaction model with N objects.

T. G. Kurtz, *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, Journal of Applied Probability, vol. 7, no. 1, pp. 49–58, 1970.
M. Benaïm and J.-Y. Le Boudec, *A class of mean field interaction models for computer and communication systems*, Performance Evaluation, vol. 65, no. 11-12, pp. 823–838, 2008.

A mean-field interaction model for modeling dissemination

(from Lucile Sassatelli's course)

- Define the occupancy measure $\mathbf{M}^{(N)}(t)$ as the vector of frequencies of states $i \in S$ at t:
$M_i^{(N)}(t) = \frac{1}{N}\sum_{n=1}^{N} 1_{\{Z_i^{(N)}(t)=i\}}$. $\mathbf{M}^{(N)}(t)$ that takes vales in Δ.
$\mathbf{M}^{(N)}(t)$ is a homogeneous Markov chain.
- Let us define the drift $\mathbf{f}(\mathbf{m})$ for $\mathbf{m} \in \Delta$ as the expected change to $\mathbf{M}^{(N)}(t)$ in one time-slot:

$$\begin{aligned}
\mathbf{f}^{(N)}(\mathbf{m}) &= \mathbb{E}[\mathbf{M}^{(N)}(t+1) - \mathbf{M}^{(N)}(t) | \mathbf{M}^{(N)}(t) = \mathbf{m}] \\
&= \sum_{\{i,i'\}\in S, i\neq i'} m_i P_{i,i'}^{(N)}(\mathbf{m})(\mathbf{e}_{i'} - \mathbf{e}_i)
\end{aligned}$$

where $P_{i,i'}^{(N)}$ is the marginal transition probability:
$P_{i,i'}^{(N)}(\mathbf{m}) = Pr\{Z_n^{(N)}(t+1) = i' | Z_n^{(N)}(t) = i, \mathbf{M}^{(N)}(t) = \mathbf{m}\}$.

T. G. Kurtz, *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, Journal of Applied Probability, vol. 7, no. 1, pp. 49-58, 1970.
M. Benaïm and J.-Y. Le Boudec, *A class of mean field interaction models for computer and communication systems*, Performance Evaluation, vol. 65, no. 11-12, pp. 823-838, 2008.

Convergence to the mean-field limit

(from Lucile Sassatelli's course)

If $\lim_{N\infty} \mathbf{f}^{(N)}(\mathbf{m}) = \mathbf{f}(\mathbf{m})$ exists for all $\mathbf{m} \in \Delta$,
Then $\mathbf{M}^{(N)}(t)$ converges to a deterministic process $\mu(t)$ that satisfies:

$$\begin{cases} \frac{d\mu(t)}{dt} = \mathbf{f}(\mu(t)) \\ \mu(0) = \mu_0 \text{ constant in } N \end{cases}$$

More exactly (Kurtz Th 3.1), $\forall \delta$:

$$\lim_{N\infty} Pr\{\sup_{s \leq t} ||\mathbf{M}^{(N)}(t) - \mu(t)|| > \delta\} = 0$$

T. G. Kurtz, *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, Journal of Applied Probability, vol. 7, no. 1, pp. 49–58, 1970.
M. Benaïm and J.-Y. Le Boudec, *A class of mean field interaction models for computer and communication systems*, Performance Evaluation, vol. 65, no. 11-12, pp. 823–838, 2008.

Performance modeling of dissemination under two-hop routing or epidemic routing

(from Lucile Sassatelli's course)

$$\mathbf{M}^{(N)}(t) = \begin{bmatrix} M_0^{(N)}(t) \\ M_1^{(N)}(t) \end{bmatrix} = \begin{bmatrix} 1 - M_1^{(N)}(t) \\ M_1^{(N)}(t) \end{bmatrix}$$

- Two-hop routing: $f_1(m_1) = \lambda s(1 - m_1)$, where s is the fraction of sources (constant in N)
- Epidemic routing: $f_1(m_1) = \lambda m_1(1 - m_1)$
- Let us rename $\mu_1(t)$ as $x(t)$, standing for the fraction of infected nodes.
- Let $X^{(N)}(t)$ be the number of infected nodes: $X^{(N)}(t)$ can be approximated by $Nx(t)$.

T. G. Kurtz, *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes*, Journal of Applied Probability, vol. 7, no. 1, pp. 49–58, 1970.
M. Benaïm and J.-Y. Le Boudec, *A class of mean field interaction models for computer and communication systems*, Performance Evaluation, vol. 65, no. 11-12, pp. 823–838, 2008.

Performance modeling of dissemination under two-hop routing or epidemic routing

(from Lucile Sassatelli's course)

From that we approximate $X^{(N)}(t)$ by the solution of:

Epidemic

$$\frac{dX^{(N)}(t)}{dt} = \beta X^{(N)}(t)(N - X^{(N)}(t)), \quad X^{(N)}(0) = 1$$

Two-hop

$$\frac{dX^{(N)}(t)}{dt} = \beta 1(N - X^{(N)}(t)), \quad X^{(N)}(0) = 0$$

- Defining T_d as the packet delivery delay, we can derive $P(t) = Pr\{Td < t\}$:

$$\frac{dP(t)}{dt} = \lambda x(t)(1 - P(t))$$

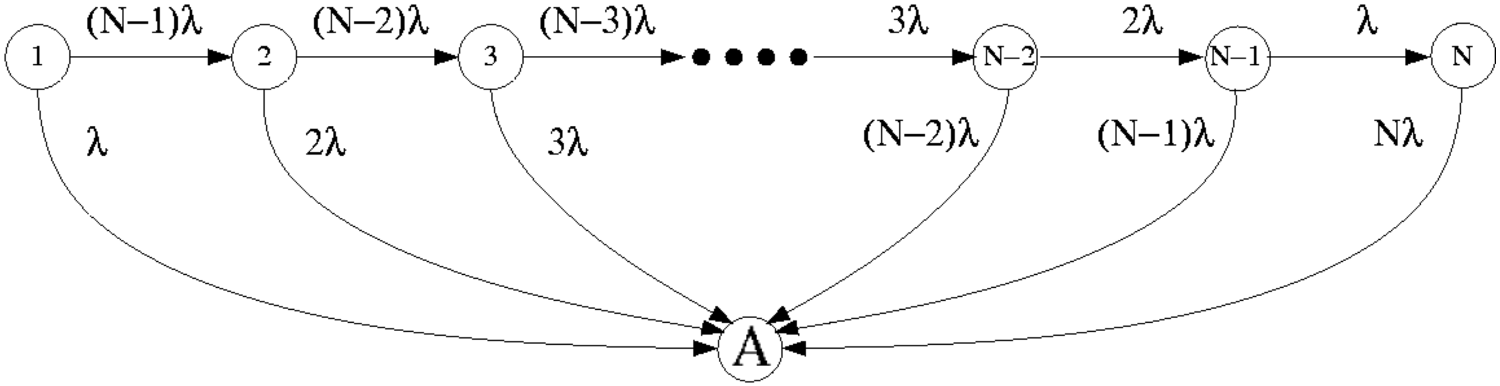
Proof: Exercise class

MAESTRO

Zhang, X., Neglia, G., Kurose, J., Towsley, D.: Performance Modeling of Epidemic Routing. Computer Networks 51, 2867–2891 (2007)

A further issue

Model the number of occurrences of the message as an absorbing Continuous Time Markov Chain (C-MC):



- We need a different convergence result

[Kurtz70] Solution of ordinary differential equations as limits of pure jump markov processes, T. G. Kurtz, Journal of Applied Probabilities, pages 49-58, 1970

[Kurtz1970]

$\{X_N(t)$, N natural$\}$

a family of Markov process in Z^m

with rates $r_N(k,k+h)$,   k,h in Z^m

It is called density dependent if it exists a continuous function f() in R^m such that

$r_N(k,k+h) = N\, f(1/N\, k,\, h)$, $h<>0$

Define $F(x)=\Sigma_h\, h\, f(x,h)$

Kurtz's theorem determines when $\{X_N(t)\}$ are ***close*** to the solution of the differential equation:

$$\frac{\partial x(s)}{\partial s} = F(x(s)),$$

The formal result [Kurtz1970]

Theorem. Suppose there is an open set E in R^m and a constant M such that

$|F(x)-F(y)|<M|x-y|$, x,y in E

$\sup_{x \text{ in } E}\Sigma_h |h|\ f(x,h) <\infty$,

$\lim_{d\to\infty}\sup_{x \text{ in } E}\Sigma_{|h|>d}|h|\ f(x,h) =0$

Consider the set of processes in $\{X_N(t)\}$ such that

$\lim_{N\to\infty} 1/N\ X_N(0) = x_0$ in E

and a solution of the differential equation

$$\frac{\partial x(s)}{\partial s} = F(x(s)), \quad x(0) = x_0$$

such that x(s) is in E for $0<=s<=t$, then for each $\delta>0$

$$\lim_{N\to\infty} \Pr\left\{ \sup_{0\le s\le t} \left| \frac{1}{N} X_N(s) - X(s) \right| > \delta \right\} = 0$$

Application to epidemic routing

$r_N(n_I)=\lambda\, n_I\, (N-n_I) = N\, (\lambda N)\, (n_I/N)\, (1-n_I/N)$

assuming $\beta = \lambda N$ keeps constant (e.g. node density is constant)

$f(x,h)=f(x)=x(1-x)$, $F(x)=f(x)$

as $N\rightarrow\infty$, $n_I/N \rightarrow i(t)$, s.t.

$$i'(t) = \beta i(t)(1 - i(t))$$

with initial condition

$$i(0) = \lim_{N\to\infty} n_I(0)/N$$

multiplying by N

$$I'(t) = \lambda I(t)(N - I(t))$$