

SCUOLA SUPERIORE DI CATANIA

Epcast: Controlled Dissemination in Human-based Wireless Networks by means of Epidemic Spreading Models

Salvatore Scellato

Joint work with Cecilia Mascolo (UCL), Mirco Musolesi
(Dartmouth College), Vito Latora (INFN-CT)

Scuola Superiore di Catania
Università degli Studi di Catania



Outline

- Motivation and definition of the research problem
- Design of a middleware interface for probabilistic anycast communication in mobile ad-hoc networks
- Models of epidemic spreading on complex networks and implementation of a new epidemic algorithm
- Simulation results with real traces
- Conclusions



Motivation

Mobile human networks can be frequently and temporarily disconnected and traditional protocols fail to offer any reliability when this happens.





Motivation

Epidemic-style protocols allow communication in dynamic and mobile networks, also in presence of temporary disconnections or network partitions.

However, existing epidemic algorithms do not permit to control the spreading of the information thus causing large system overloads.



Problem description

Design of an epidemic-style algorithm that allows the developers to tune the number of message sent for data dissemination in ad hoc networks

Existing solutions do not take into account the structure of the underlying network to drive and regulate the spreading!

- The original Epidemic Protocol by Vahdat and Becker
- Bimodal multicast by Birman et alii
- Frugal event dissemination by Baheni et alii
- Study of data dissemination in MANETs by Khelil et alii



What is an epidemic algorithm?

The analogy between information dissemination in networks and epidemic transmission in social systems is perfect

SPREADING OF AN EPIDEMIC	SPREADING OF INFORMATION
Individual is not infected	Host is not informed
Individual is infecting others	Host is spreading the information
Individual does not contract a disease any more	Host is no more interested



Our contribution

This work refines and enhances a previous **data dissemination algorithm** [Mascolo, Musolesi 2006] for mobile ad hoc networks that relies on epidemic models **taking into account the structure of the underlying networks**

We present the definition of an **interface** that allows the programmers to **tune the number of recipients** for anycast communication **reducing global system load**

Our probabilistic algorithm mimics the **spreading of an epidemic** among a set of individuals



Middleware primitive

We introduce a primitive for **probabilistic anycast communication** as follows:

`epcast (message , percentageOfHosts , time)`

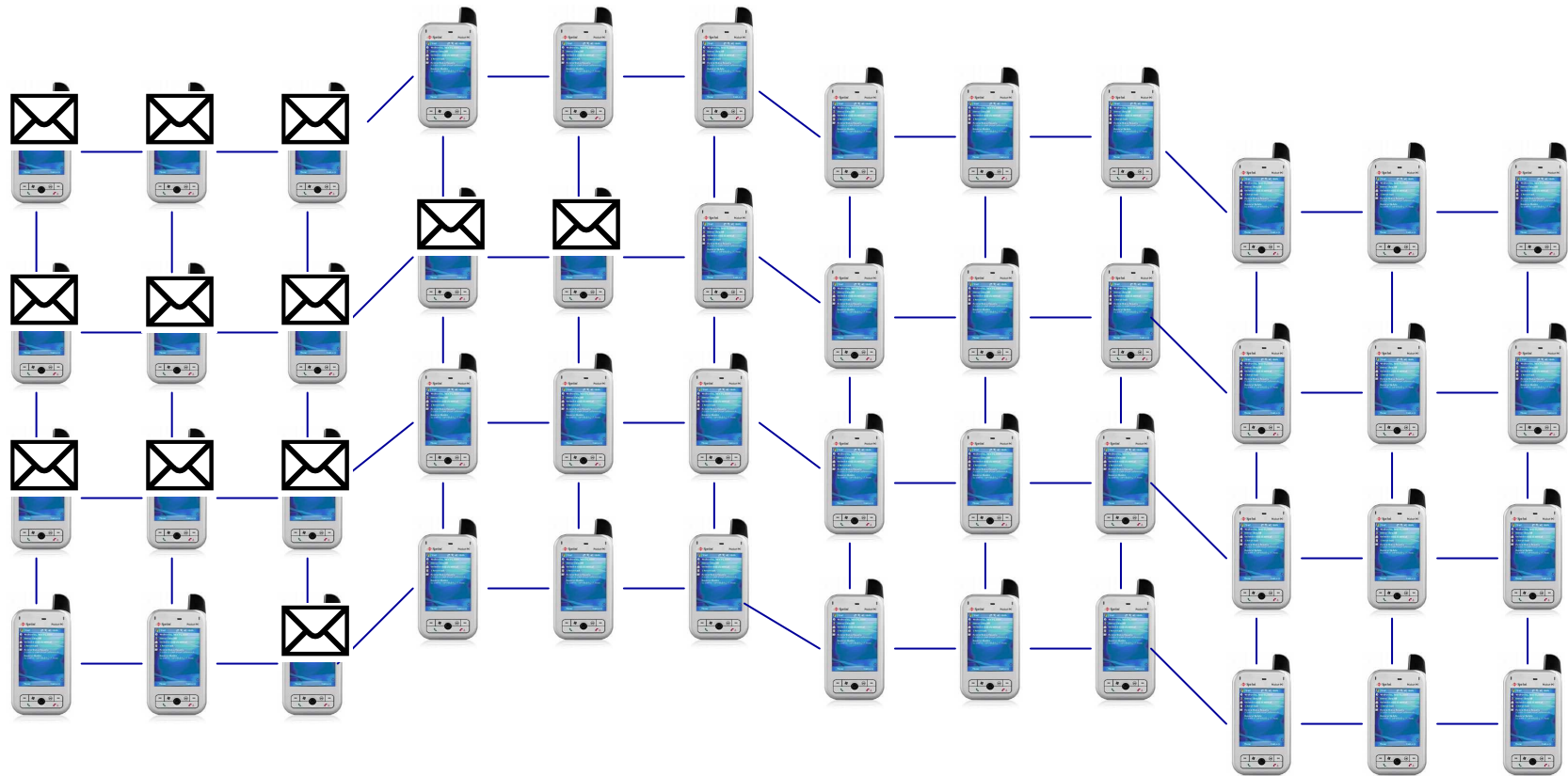
where a `message` has to be sent in a bounded `time` interval only to a `percentageOfHosts` in the system.

Our algorithm achieves its goal by sending only a **sufficient** number of messages, while existing epidemic-style protocols usually waste resources by sending a large number of messages on the network.



An example

$\text{epcast}(m1, 0.25, 30)$





Tuning the spreading



Problem: how to calculate the **probability of infection** of a message in order to have the desired number of copies in the system after a certain given time?

Recent results on epidemic spreading on complex networks shed new light on this issue



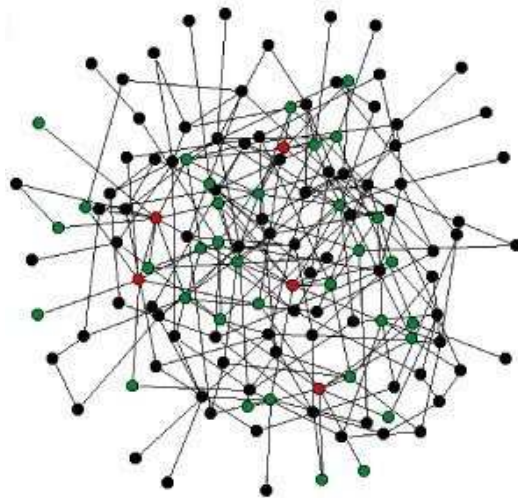
Some words on complex networks

Many recent results in the fields:

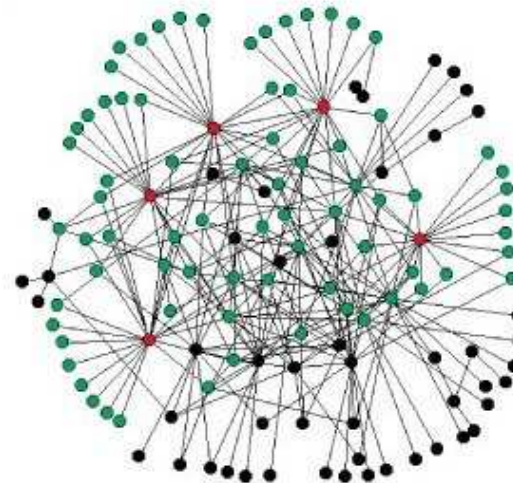
- Structural properties
- Dynamic processes

Two main classes of networks:

Homogeneous networks



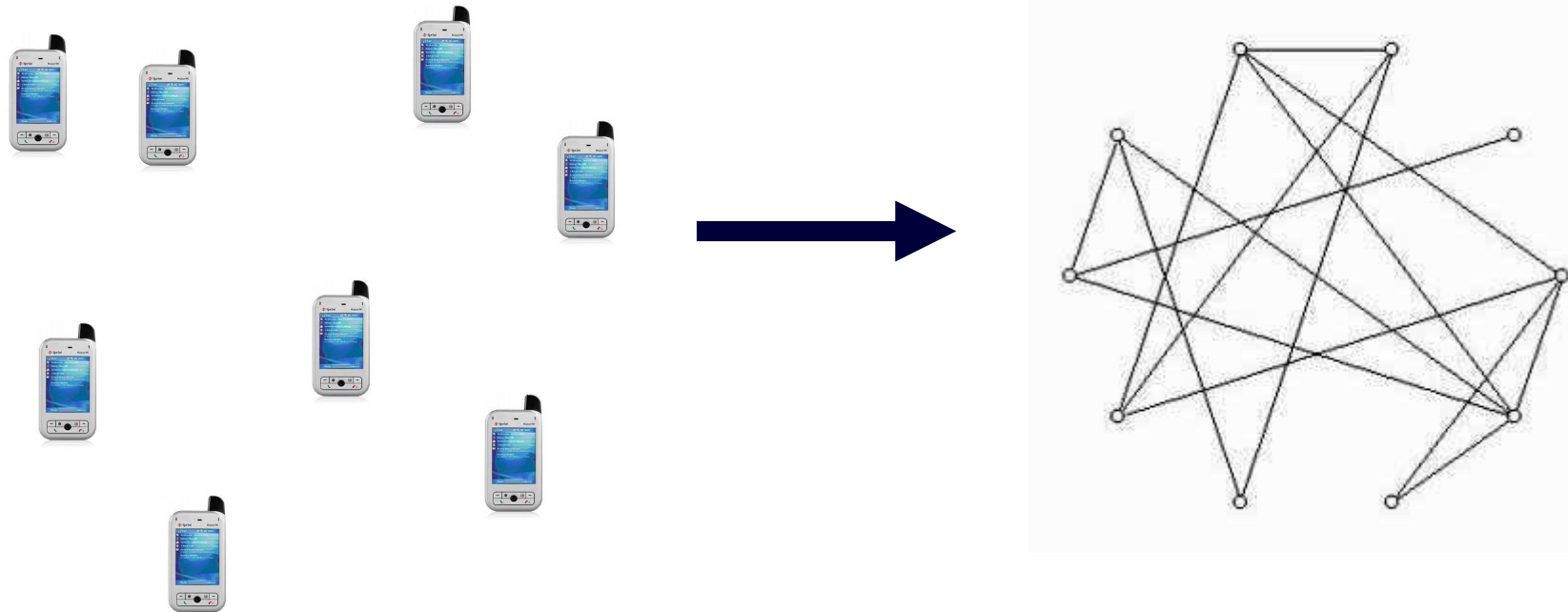
Heterogeneous networks





Mobile Ad Hoc Networks and Complex Networks

Complex networks can be used to model the pattern of connectivity among devices in a mobile ad hoc network: at a given time we can extract a graph where **two devices are connected only if they can communicate with each other.**





Mobile Ad Hoc Networks and Complex Networks

A mobile ad hoc network topology usually exhibits a **homogeneous network** structure: this is a realistic assumption in all cases characterized by a high density of independently moving hosts such as in large outdoor spaces (i.e., squares, stations, airports or around sport venues).

This model can be easily adapted to the case of constrained spaces and buildings with centers of aggregations, where the emergent pattern of connectivity is well described by a **heterogeneous network**



Assumptions

- All susceptibles in the population are equally at risk of infection from any infected host (*homogeneous mixing*)
- All infectives in the population have equal chances to recover
- The infectivity of a single host, per message, is constant
- The initial number of the nodes in the network is approximatively known a priori by each host
- Every host collaborates to the delivery process and no malicious nodes are present
- Each node has a buffer of the same size
- The number of hosts is considered constant during the spreading of the infection (no failing hosts)



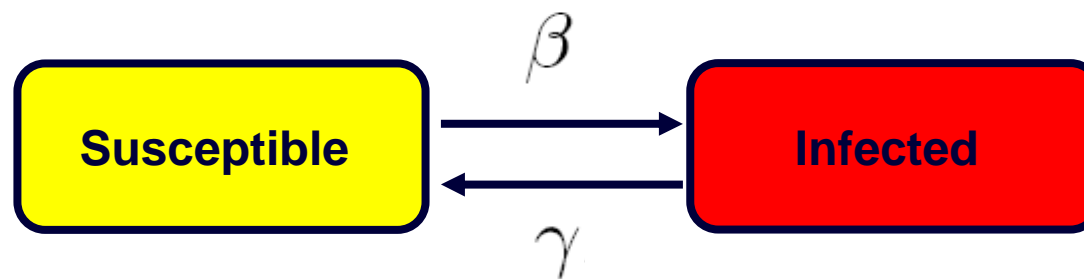
Epidemic compartmental models

Compartmental models are used to describe the infection dynamics in a population: the population is divided in classes and each individual progresses through these states as the infection evolves

SIR MODEL



SIS MODEL





The Kermack and McKendrick Model

The temporal evolution of the system is described by a set of nonlinear differential equations that relate the number of individuals in each class with the rates of infection and recovery

susceptible hosts at time t

infected hosts at time t

recovered hosts at time t

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\beta S(t)I(t) \\ \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \\ S(t) + I(t) + R(t) = N \end{array} \right.$$

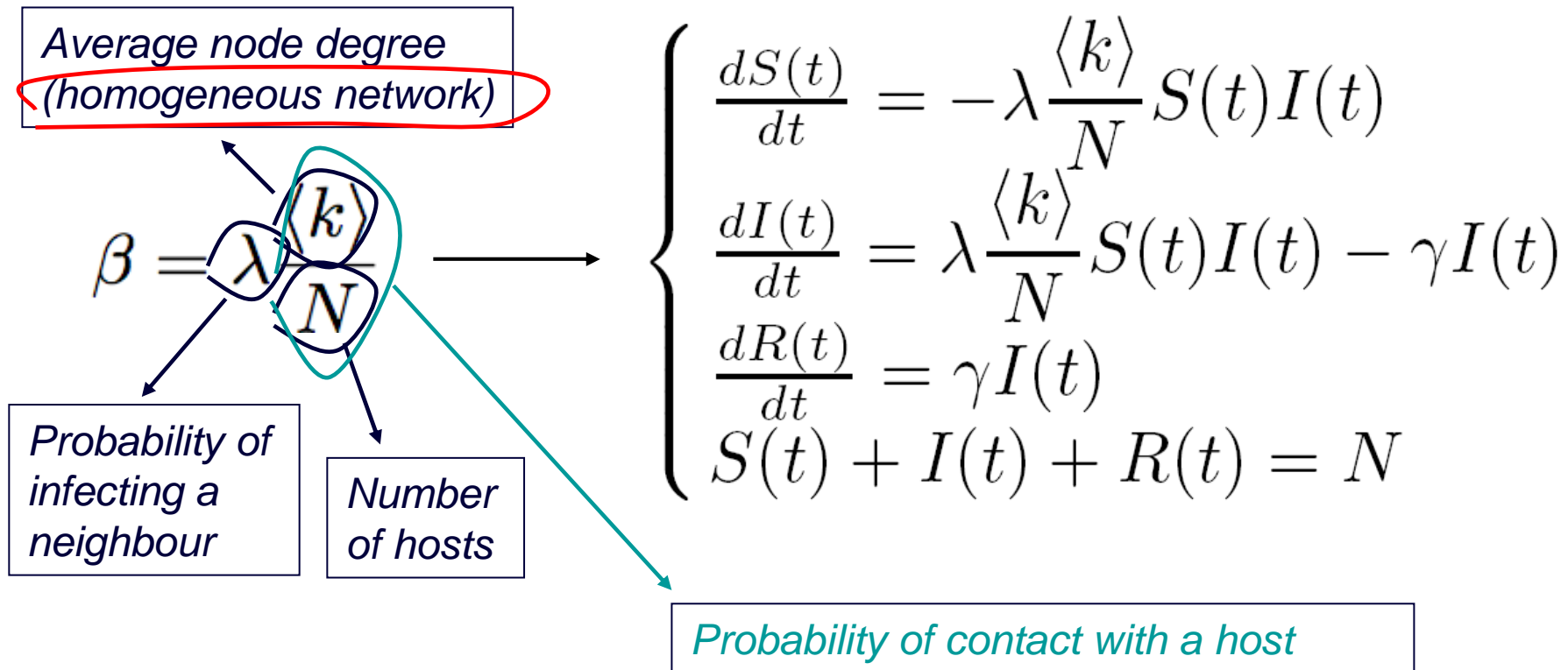
infection rate

recovery rate



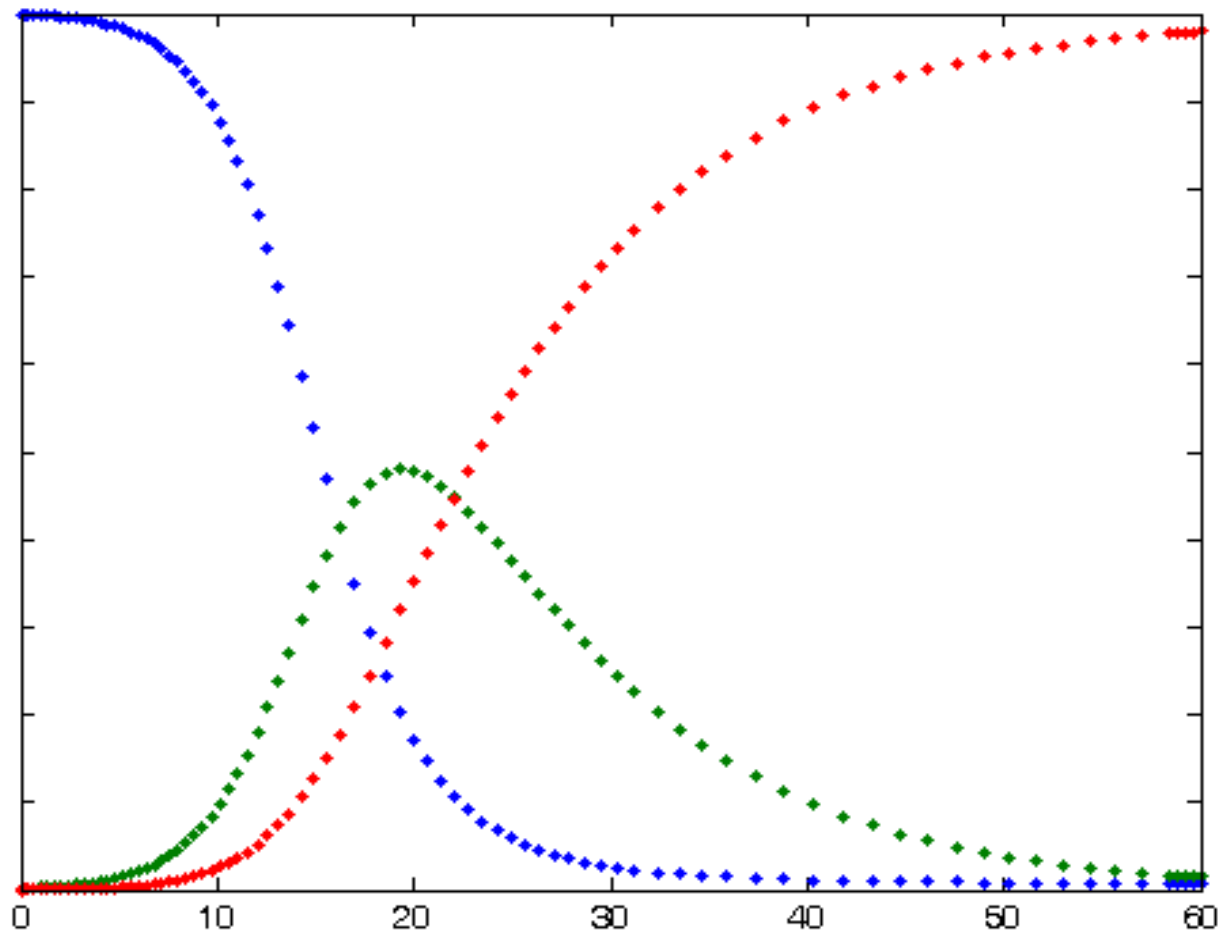
Epidemic Spreading Models on Complex Networks

In epidemic models on complex networks the infection may occur only between nodes connected by a link: we can thus define the **infectivity** λ





Solution of the system



- Susceptibles
- Infected
- Recovered

Initial conditions:

$$S = N-1$$

$$I = 1$$

$$R = 0$$



Tuning the epidemic

The system does not present an analytical solution:

however the numerical solution allows to compute the number of infected and recovered individuals at any instant t as a function of the **infection** and **recovery** rates.

The **recovery rate** γ is usually fixed by the local properties of the hosts: each host has a limited buffer, so messages can be deleted to make room for new ones.

Instead, the value of **infection rate** β can be calculated in order to have the desired number of reached hosts after a specific amount of time



Solution details

The calculation of the correct infectivity to reach a certain number of hosts in a given time depends also on the recovery rate: in an epidemic algorithm, this is represented by the **deletion probability** from the buffer, since a deleted message will not be accepted anymore.

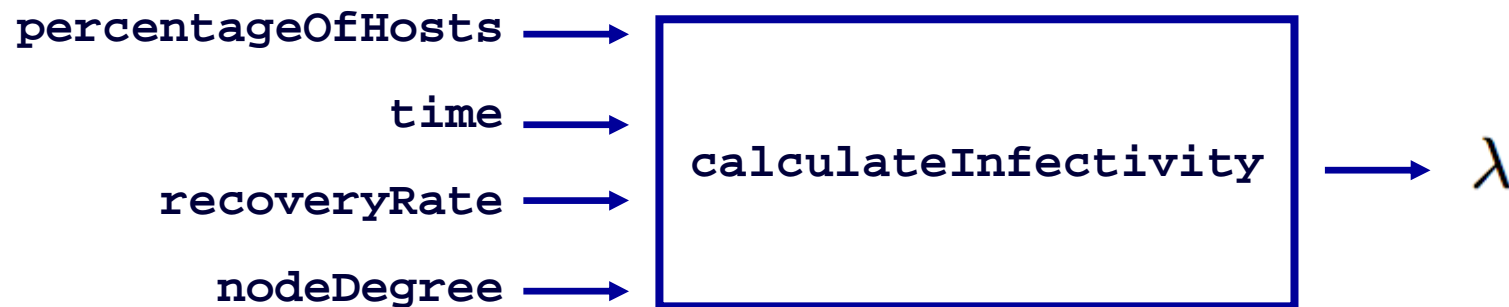
Each host observes how many deletions occur in its own buffer and uses this number as an approximate value of the global recovery rate in the system.



Infectivity calculation

When the `epcast` primitive is invoked to send a message, the correct **infectivity** λ must be assigned to it in order to obtain the desired results

The sender calculates the value of β to obtain the desired number of reached host: from this value it derives the correct infectivity λ and attach it to the message to be sent

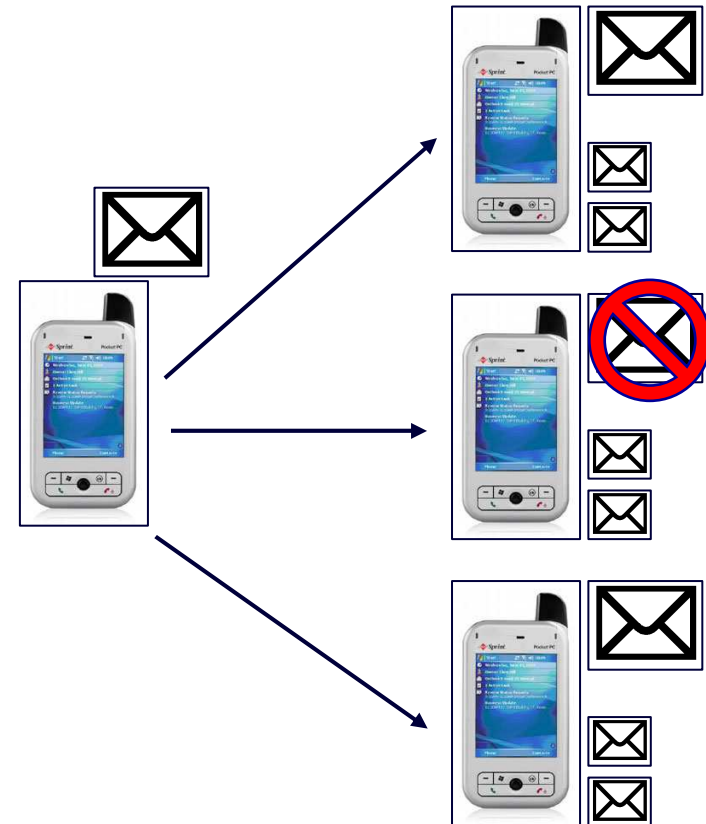




Implementation

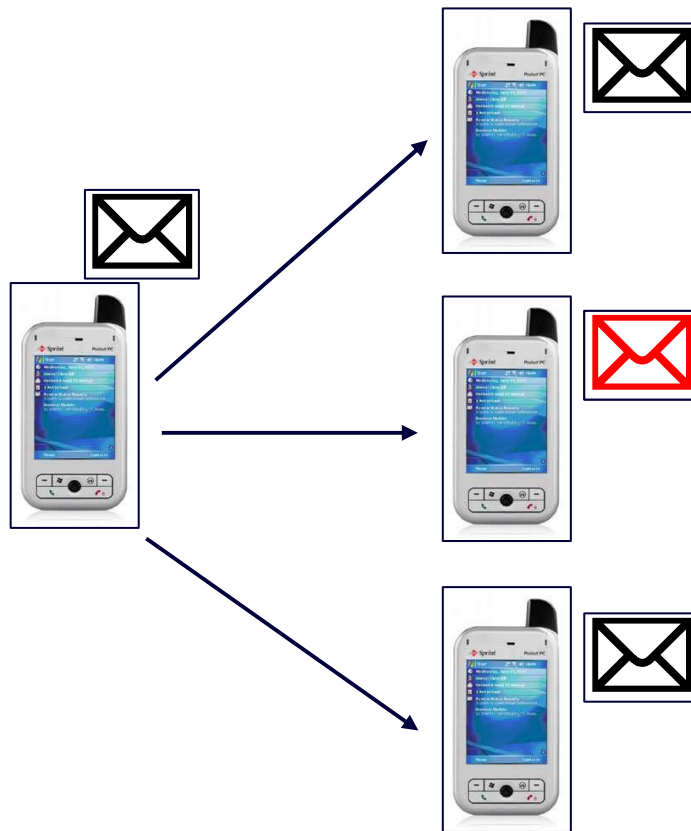
Periodically:

1. Each infected host **broadcasts** each message in its buffer **with the calculated infectivity**
2. If the message has never been received by a neighbour, it stores it **only with a probability equal to the infectivity** of the message
3. Older **messages are deleted** to make room for new ones, **never accepting them again** in the buffer





An example



```
SendMessage() {  
    for(int i=0;i<nMessages;i++) {  
        m=System.getMessage(i);  
        System.sendMessage(m);  
    }  
}
```

```
Receive(Message m) {  
    infectivity=m.getInfectivity();  
    rValue=random(0,1);  
    if(rValue <= infectivity &&  
        System.isNew(m))  
        System.storeMessage(m);  
}
```



Global System Overhead

An interesting quantitative parameter is the **total number of replicas** needed to disseminate the message to a certain percentage of hosts: a message is sent by each infected host in every round

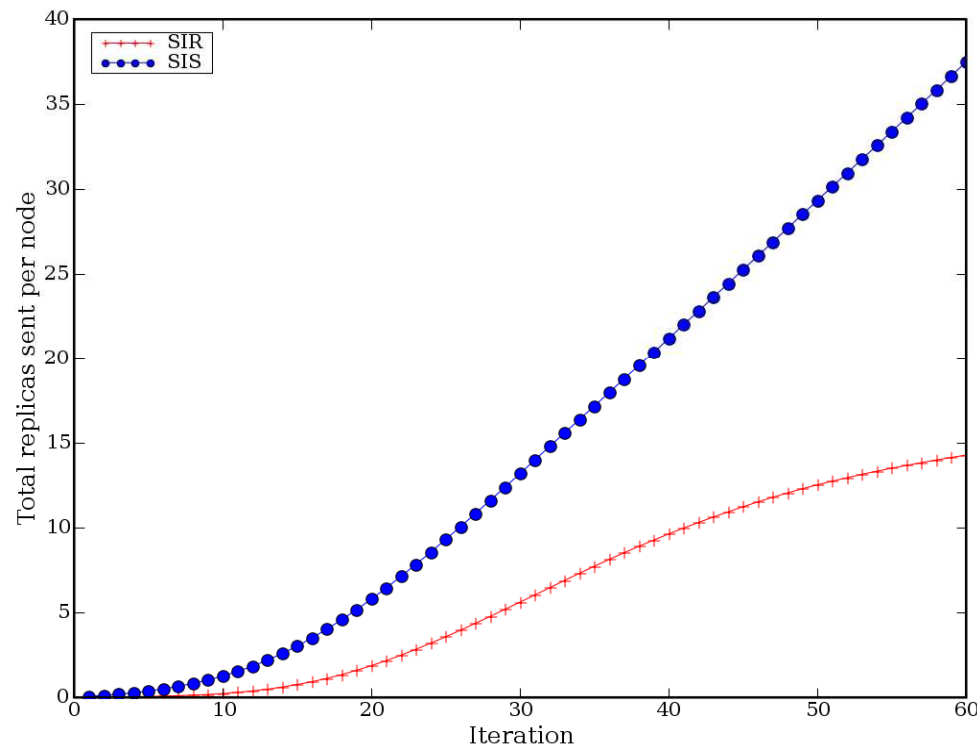
Thus the two models present different behaviours

- **SIR Model:** as soon as the host deletes a message it does not accept the same message again
- **SIS Model:** when the host deletes a message it can be accepted again



SIR vs. SIS

A comparison of SIR- and SIS-based protocols shows that in the SIR case the total number of replicas sent is lower, for hosts stop message spreading when the epidemic is already growing, thanks to the recovery process



● SIS protocol

● SIR protocol

Number of replicas per host per message for the SIS and the SIR Model VS. time



Adaptation to heterogeneous networks

Our approach can be extended to the general case of **heterogeneous networks**, where the approximation $k \approx \langle k \rangle$ is not valid: however the same probabilistic communication primitive can be used with a different semantic.

Given a node degree fluctuating in the range $[k_{\text{MIN}}, k_{\text{MAX}}]$, we observe that for a value of the infectivity computed with $k = k_{\text{MIN}}$, the obtained spreading of the infection will always be greater than the one obtained with another k .

By using k_{MIN} we obtain a guaranteed lower bound for final performance.



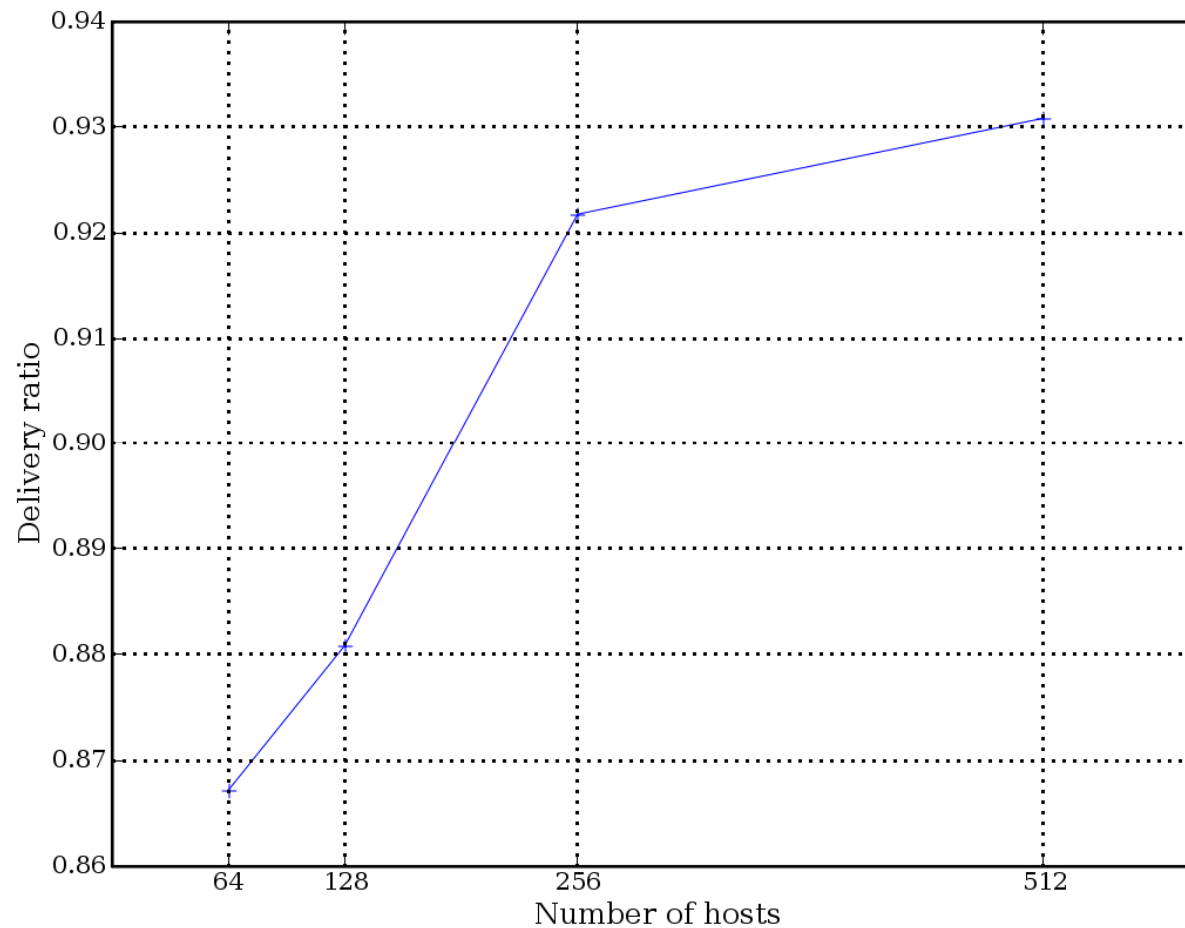
Simulation

We simulated a simple scenario of mobile nodes to test the performance of our approach

- Implemented using Python and NetworkX
- Square area $1km \times 1km$
- Transmission range $200m$
- 20 messages
- RWP Mobility Model \rightarrow Random Graph structure
- Speed in the range $[1-6] m/s$
- Scenarios with different number of hosts: 64, 128, 256, 512



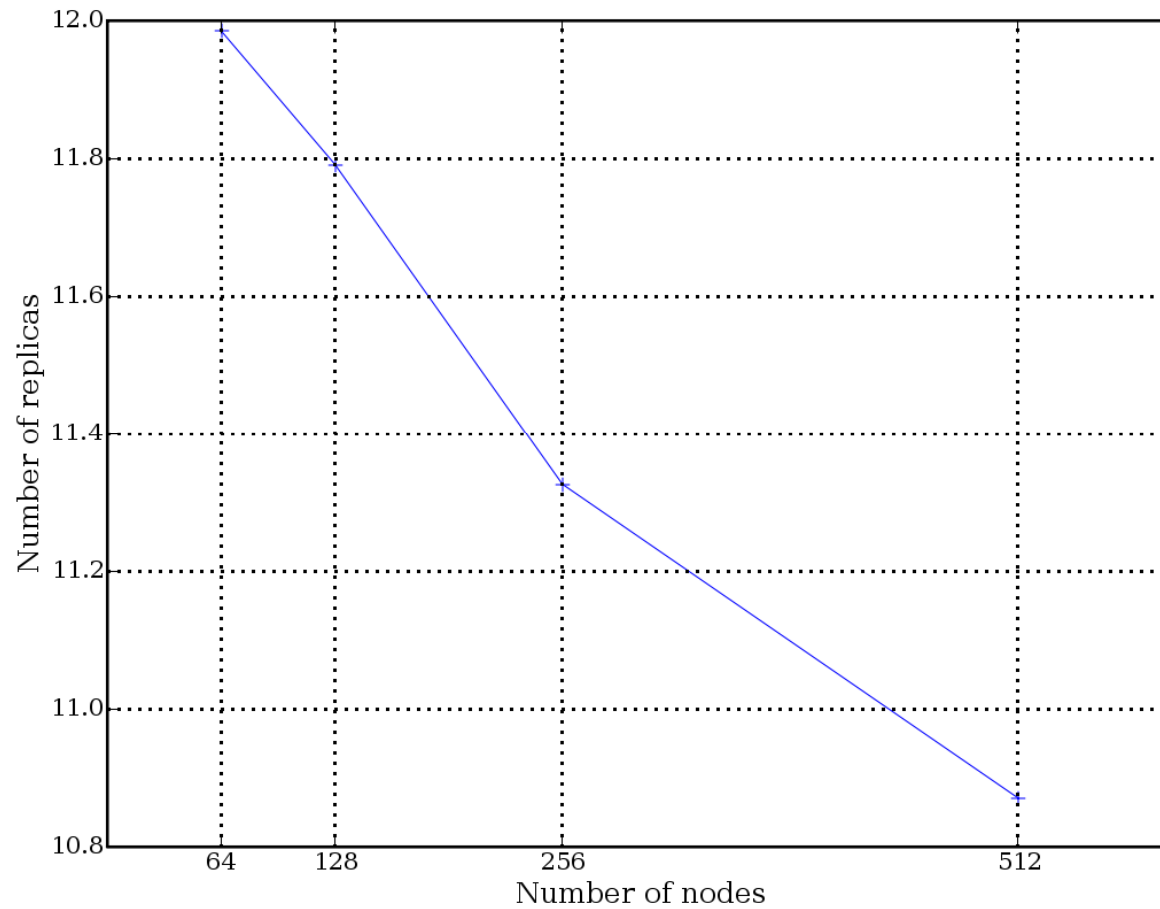
Delivery ratio vs. Population density



100% hosts to be reached



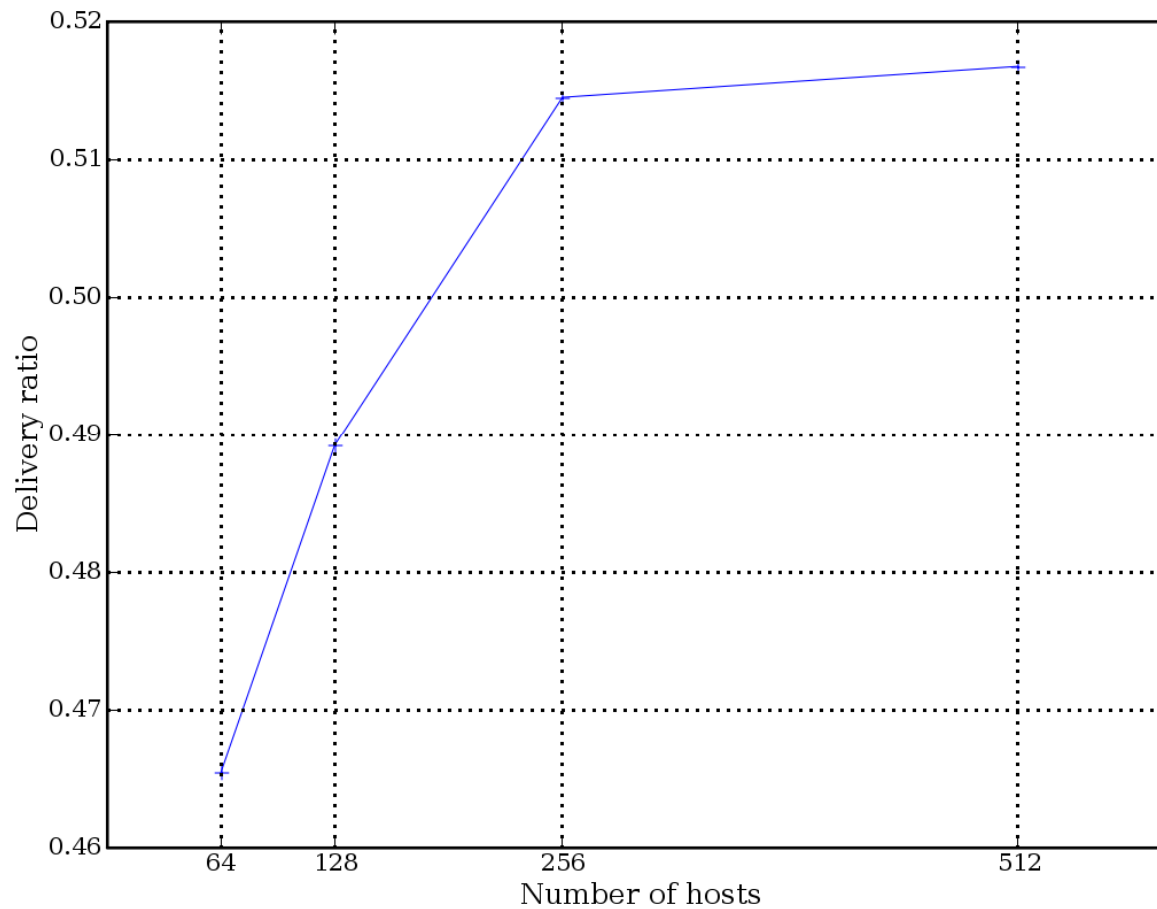
Number of replicas vs. Population density



100% hosts to be reached



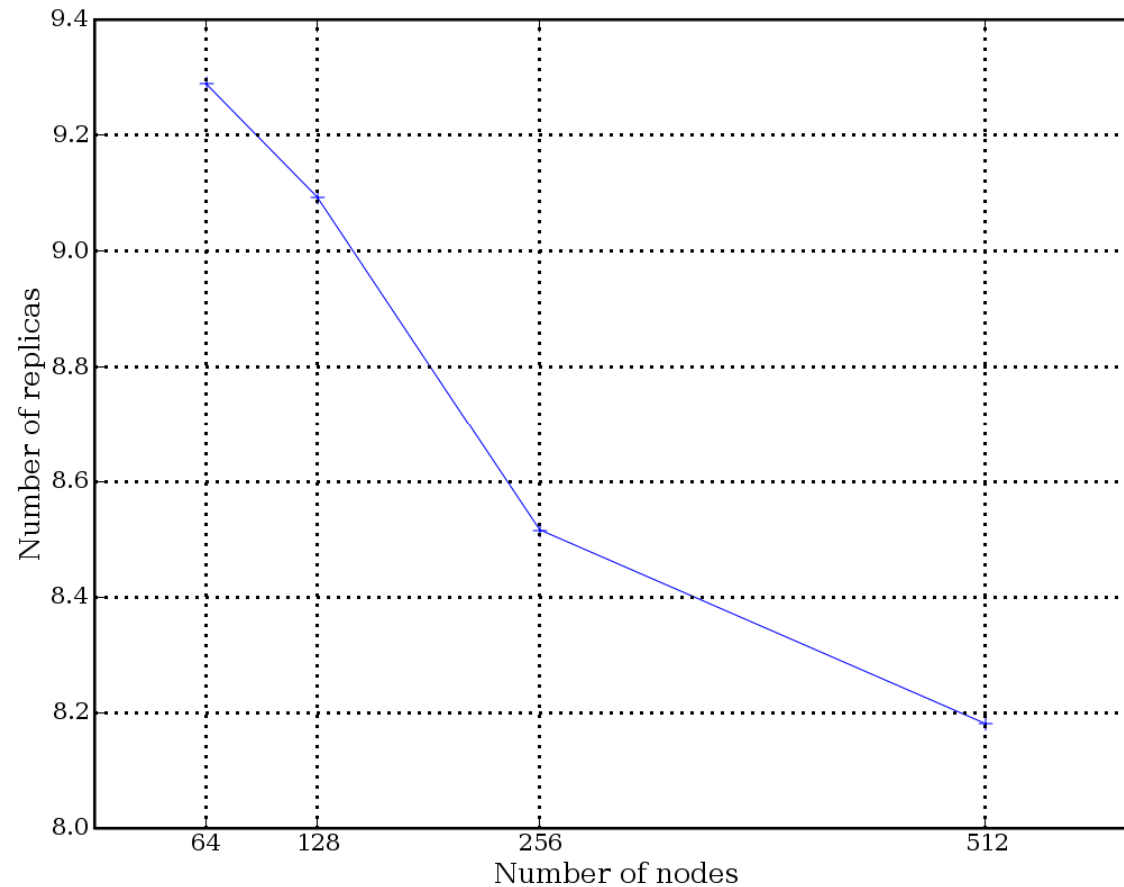
Delivery ratio vs. Population density



50% hosts to be reached



Number of replicas vs. Population density



50% hosts to be reached



Validation with real traces

A large amount of traces from the **802.11 campus network of Dartmouth College** is available through the **CRAWDAD Project**. We selected all the contacts between 9 am and 6 pm in one day during the term (*Wed. 21 April 2004*), discarding contacts with duration less than 60 seconds.

Two users are connected only while they are associated with the same AP. Our resulting data set had 2201 unique MACs and 11572 contacts with all access points.



Dartmouth College Traces

Protocol	Desired fraction	Delivered fraction	Messages sent
Epcast	0.50	0.43	17123
Epcast	0.75	0.68	24738
Epcast	1.00	0.90	32475
Epcast (heterogeneous)	1.00	0.90	57342
Epidemic (beta = 0.25)	1.00	0.64	95969
Epidemic (beta = 0.5)	1.00	0.87	121873
Epidemic (beta = 1.0)	1.00	0.92	155446



Applications and Future Work

- Mobile and fixed sensor networks
- Queries in P2P systems
- Vehicular networks



Questions?

Epcast: Controlled Dissemination in Human-based Wireless Networks by means of Epidemic Spreading Models

Salvatore Scellato, Cecilia Mascolo, Mirco Musolesi, Vito Latora.
Submitted for Publication. November 2007.

<http://www.ssc.unict.it/~sascellato/>