Social Evolving Networks: Models and Information Spreading

Joint work of

IIT-CNR, Univ. Roma “Sapienza”, Univ. Roma 2, Univ. Salerno

Outline

1. Motivation and models overview

(P. Santi)

2. Information spreading in social evolving networks

(F. Pasquale)

3. Conductance and information spreading

(L. Becchetti)

Part I:
Motivation and models overview

Introduction

Goal:

Investigating the dynamics of information spreading in mobile social networks (**MSNs**)

What are MSNs?

MSNs are a specific type of opportunistic (or delay-tolerant) network in which mobile nodes are individuals (hence, ***social entities***) carrying smart phone/PDA or similar devices. Nodes in an MSN can establish direct wireless communication links and exchange msgs when close to each other

Features of MSN:

The network is ***very sparse*** and ***always disconnected***; small "connectivity islands" – communication opportunities – arise thanks to node mobility; mobility is essentially the only communication means within the network

State-of-the-art

Where we are:

- Some recent results on information spreading in Markovian Evolving Graphs (MEG) – ***discrete time model***: Given any two nodes *u*,*v* in the network, existence of edge (*u*,*v*) is modeled as a two-state Markov chain, with state 0 = **“No link”**, state 1 = **“Link”**, and transition probabilities *p* (link birth rate) and *q* (link death rate)

[CMMPS08] A. Clementi, C. Macci, A. Monti, F. Pasquale, R. Silvestri, *“Flooding Time in Edge-Markovian Dynamic Graphs”*, Proc. ACM PODC, 2008.

…

- Some recent results on bounding unicast delivery time in opportunistic networks – ***continuous time model***: Given any two nodes *u*,*v* in the network, the inter-meeting time between nodes *u*,*v* is modeled as an exponential r.v. with a certain, fixed parameter λ

[GNK05] R. Groenvelt, P. Nain, G. Koole, *“The Message Delay in Mobile Ad Hoc Networks”*, Performance Evaluation, 2005.

…

What about "social structure"?

Shortcoming of existing approaches:

"Social structure" of the collection of individuals forming an MSN is completely ignored: the *"connectivity properties"* (probability of having a communication opportunity) between two network nodes *u*,*v* are statistically equivalent to those between any other pair of nodes *w*, *z*. ***This is very distant from reality!!***

How can we take social structure into account in the analysis?

First attempt in a recent manuscript: analysis of unicast performance in MSNs in the continuous-time model, where meeting rate λ_{uv} depends on the degree of **"interest similarity"** between *u* and *v*

[DMMSS11] J. Diaz, A. Marchetti-Spaccamela, D. Mitsche, P. Santi, J. Stefa, *"Social-Aware Forwarding Improves Routing Performance in Pocket Switched Networks"*, submitted for publication, 2011.

Our goal

Our goal in this work is **gaining an understanding of the dynamics of information propagation in MSNs**

The following questions are of interest to us:

1. **What is the effect of "social structure" on information propagation speed?** Given the same "density of contacts", does a "social structure" ***increase*** or ***decrease*** information propagation speed? Intuition says: ***increase***, but formally proving this fact is not at all trivial
2. **What is the effect of "social structure" on the total number of messages (message complexity) to be sent to reach all nodes in the network?**

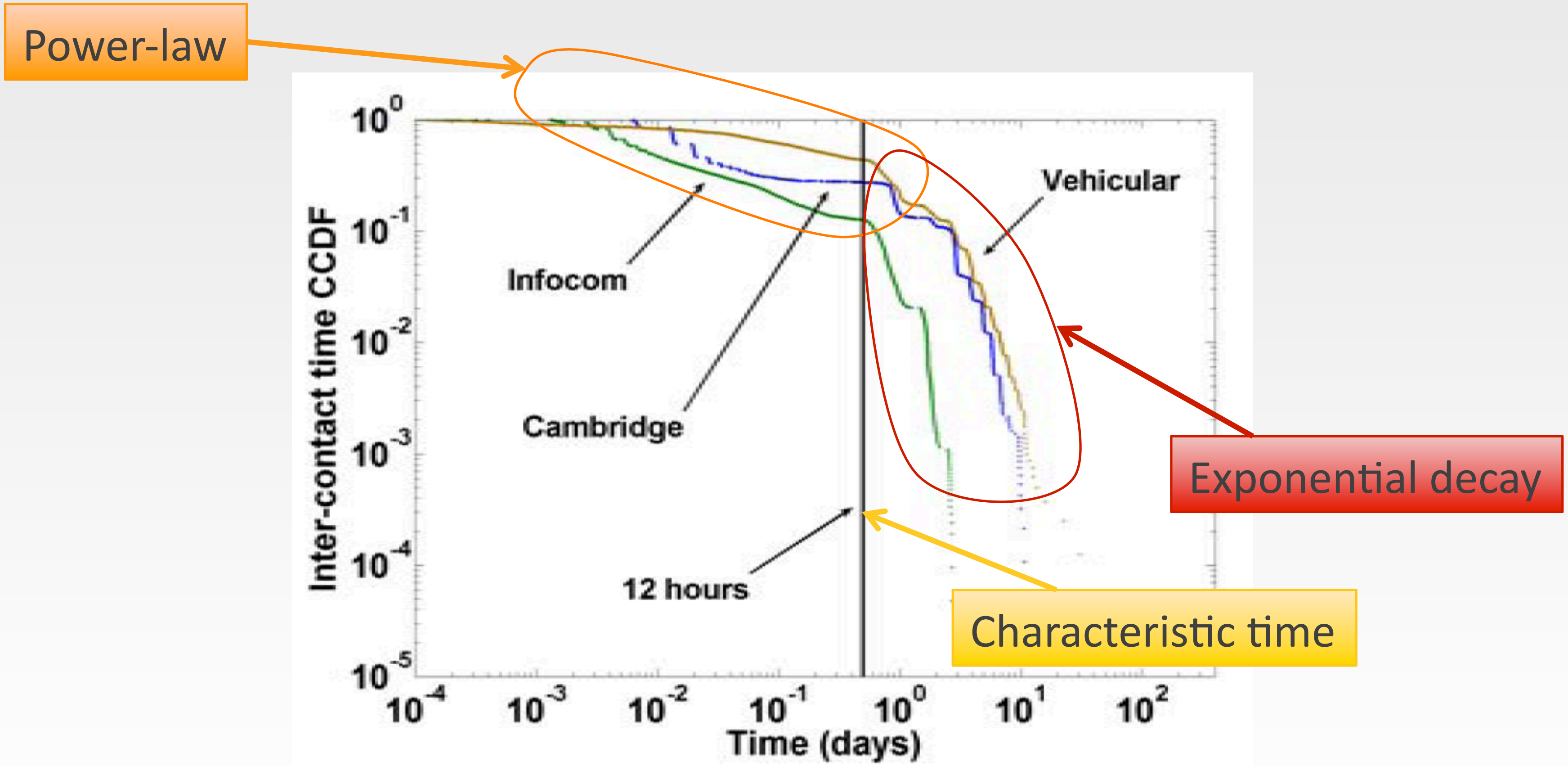
Modeling MSNs

A first challenge to address to tackle questions 1. and 2. is to define an ***analytically tractable model*** of MSN ***accounting for "social structure"***

Should we go for a continuous or discrete model?

Our choice is **discrete**, so that we can re-use the machinery of the recently proposed **MEG** approach

The dichotomy of inter-contact time distribution



Aggregated inter-contact time ccdf for three data sets
(taken from **[KLV07]**)

Inter-contact time distribution dichotomy (2)

Main finding of [KLV07]:

Inter-meeting time distribution displays a dichotomy:

There exists a *characteristic time* *T* (about 12 hours) such that inter-meeting time distribution behaves as a **power-law** before time *T*, and behaves as an **exponential distribution** after time *T*

Can the exponential tail of the distribution be ignored in analyzing opportunistic network performance?

No, because the **mean inter-meeting time** is often **larger than the characteristic time**, so the exponential tail cannot be ignored

[KLV07] T. Karagiannis, J.-Y. Le Boudec, M. Vojnovic, *"Power Law and Exponential Decay of Inter Contact Times between Mobile Devices"*, Proc. ACM Mobicom, 2007.

Modeling the ICT distribution dichotomy

Can we define a ***simple***, ***analytically tractable***, ***discrete-time*** model which is able to **reproduce the inter-contact time distribution dichotomy** observed in real world traces?

OPEN PROBLEM IN THE LITERATURE

To address the above question, let's go back to [KLV07]. The authors give a possible explanation of the observed inter-contact time distribution dichotomy

Dichotomy: possible explanation

Which could be a possible explanation of the observed inter-contact time distribution dichotomy?

In [KLV07], the authors attempt to answer this question by analyzing the relationship between the **return time** and the **inter-contact time** distribution

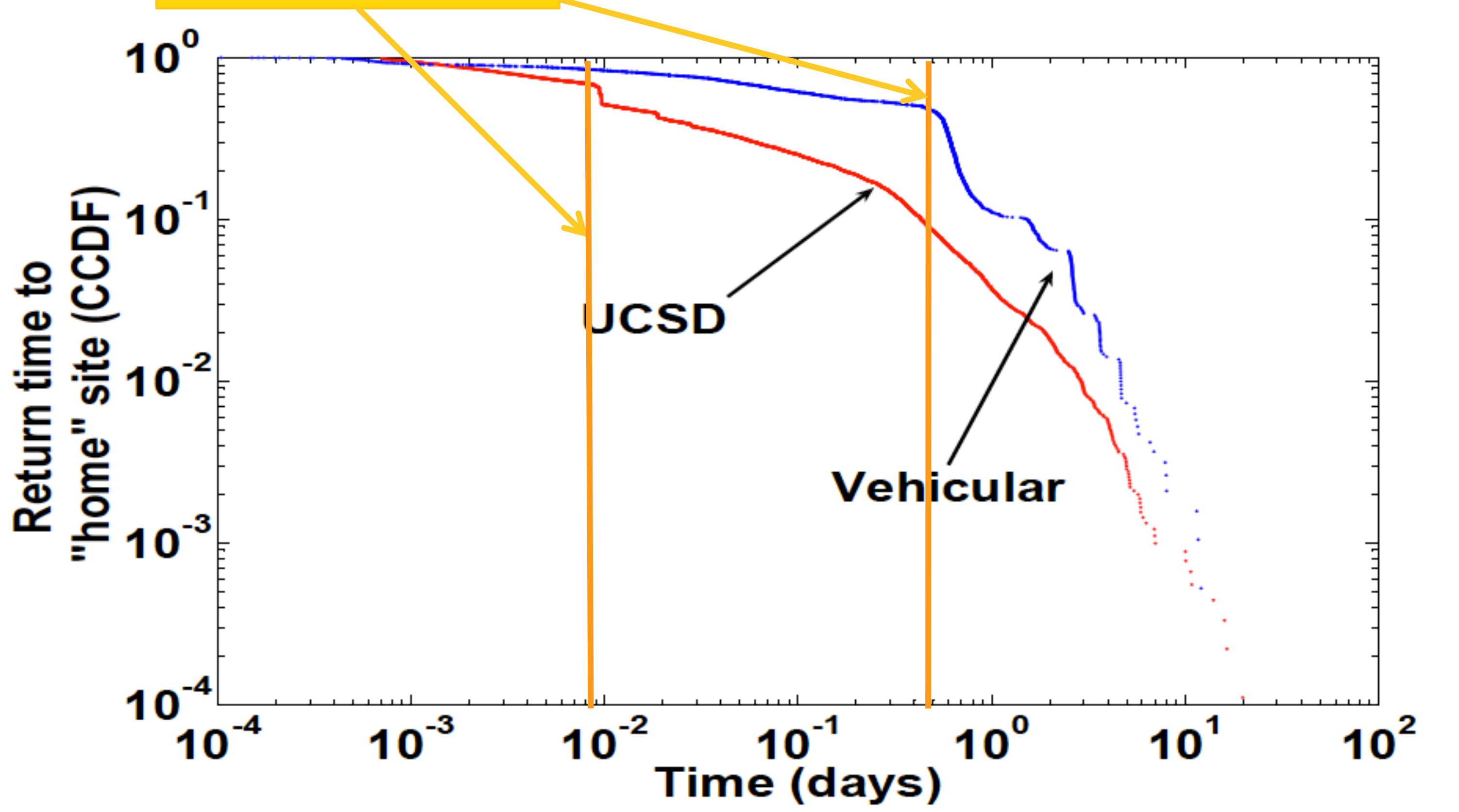
Return time: time for a node to return to its **"home site"**

"Home site": location where the node spends most of the time

In real-world traces, **"home site"** is defined as the most visited AP/cell, or geographical region (for vehicular traces)

Return time distribution

Characteristic time



Return time distribution for two real-world traces
(taken from **[KLV07]**)

Return vs. inter-contact time

Why are return and inter-meeting time related?

Hypothesis:

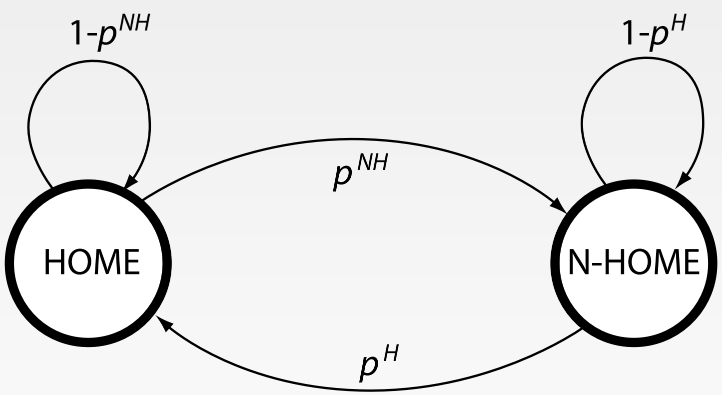
Two mobile nodes always meet at a particular site – **"meeting site"**

Under the above hypothesis, **inter-contact time** is *stochastically larger* than **return time** of any of the two nodes to the **"meeting site"**

If the two devices are *time-synchronized*, then **return time to "meeting site"** would closely characterize **inter-contact time** between the two nodes

The Home-MEG model

The **Home-MEG** model builds upon the intuition that ***nodes tend to meet in a single place*** (Home location). Thus, the probability of having a contact opportunity between nodes *u*,*v* is p_{high} if the two nodes are at home, and p_{low} if one of the two nodes (or both) are in the outside world



The **Home-MEG** model for a node pair *u*,*v* is thus a simple two-state Markov chain, where state is HOME when both *u*,*v* are at home location, and NotHOME otherwise

Home-MEG model for a network of *n* nodes: *n*(*n*-1)/2 replicas of statistically identical Home-MEGs

The Home-MEG model (2)

The **Home-MEG** model thus has four parameters:

1. p_{NH} = probability of transition to state NH
2. p_H = probability of transition to state H
3. p_{high} = probability of having a (***instantaneous***) contact opportunity when in state H
4. p_{low} = probability of having a (***instantaneous***) contact opportunity when in state NH

Can we set the values of (p_{NH} ,p_H, p_{high}, p_{low}) so to resemble inter-contact time distribution of real-world traces?

The ICT distribution in the Home-MEG model

Prob($ICT=k$) = Prob ($H|Contact$) P_{kH} + Prob ($NH|Contact$) P_{kN}

where

$P_{iH} = p^{NH} (1-p_{low})P_{(i-1)N}+(1- p^{NH})(1-p_{high})\ P_{(i-1)H}$

$P_{iN} = (1- p^{H})\ (1-p_{low})P_{(i-1)N}+ p^{H} (1-p_{high})\ P_{(i-1)H}$

for $i = 2,\ldots,k$ ***and***

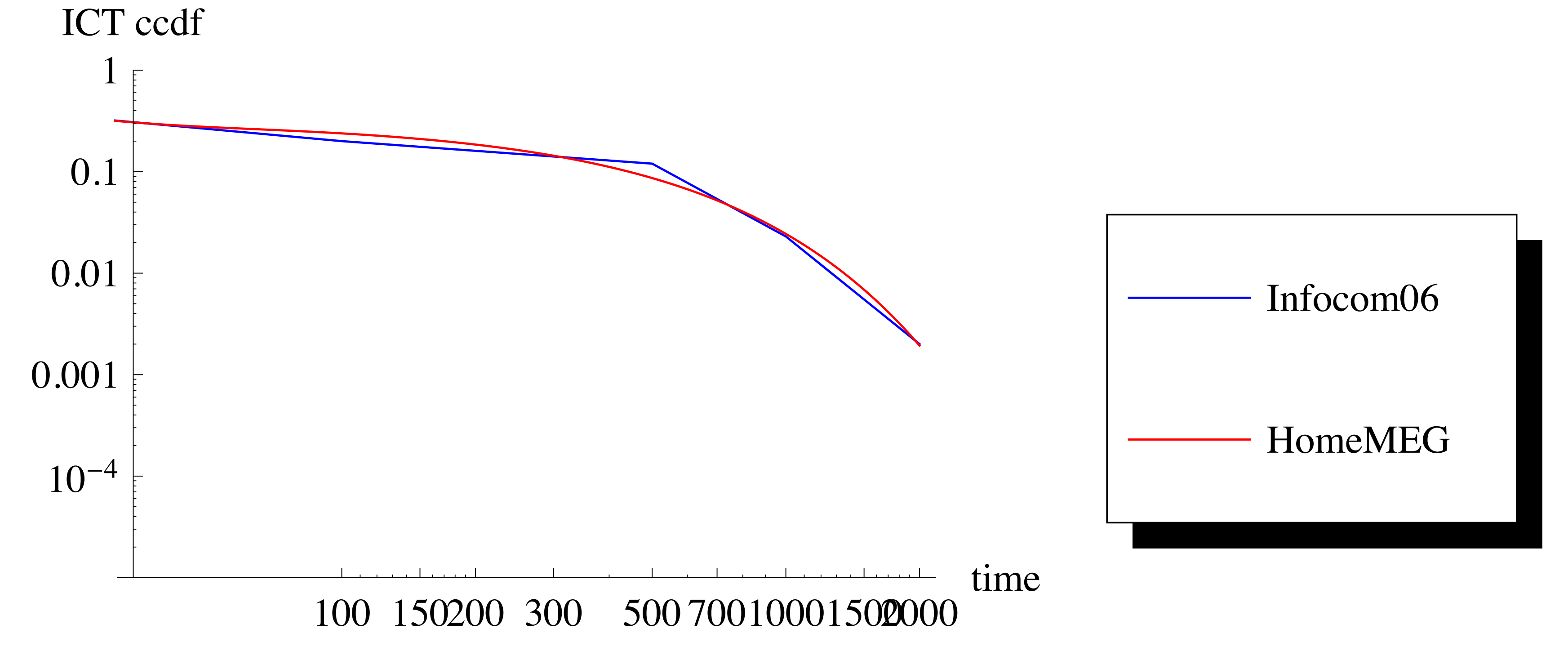
$P_{1H} = p^{NH}\ p_{low} + (1- p^{NH})p_{high}$

$P_{1N} = (1- p^{H})\ p_{low}+ p^{H}\ p_{high}$

Prob ($H|Contact$) = $(p^{H}\ p_{high})/(p^{H}\ p_{high} + p^{NH}\ p_{low})$

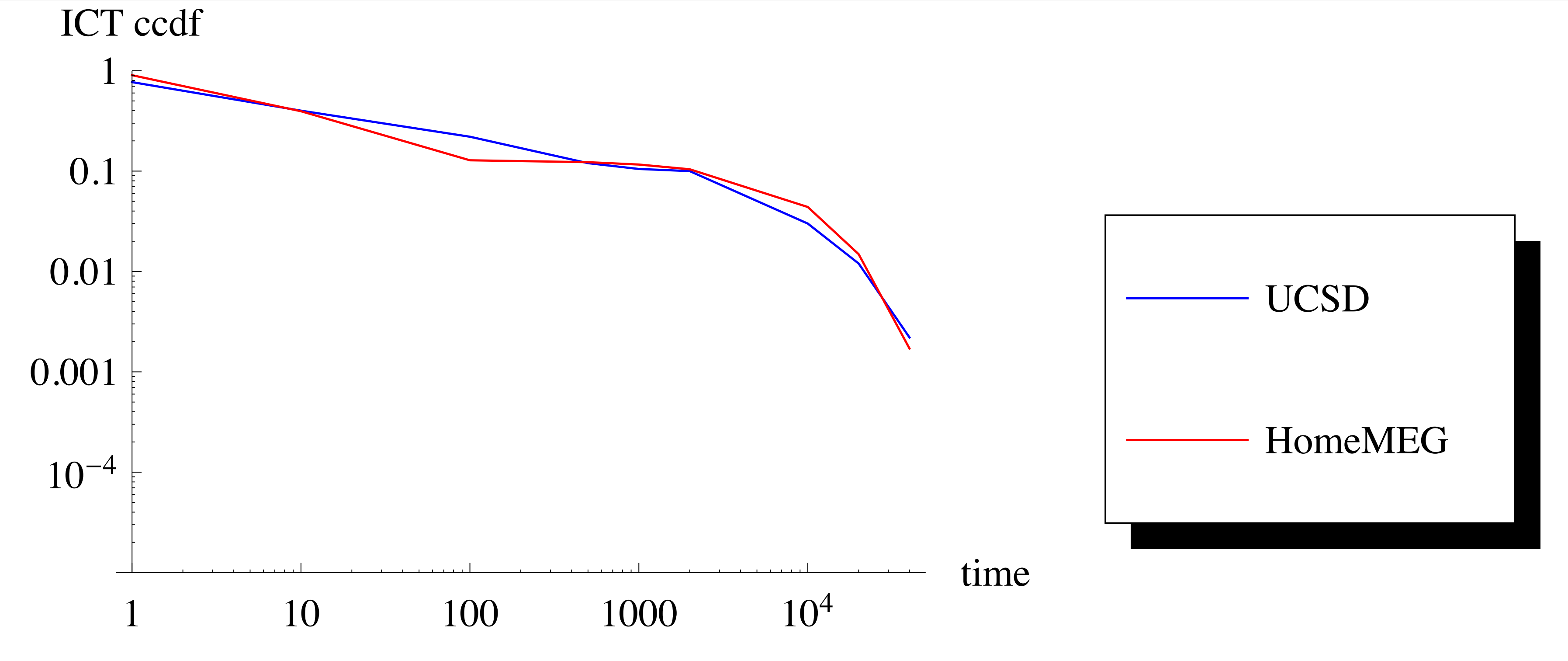
Prob ($NH|Contact$) = $(p^{NH}\ p_{low})/(p^{H}\ p_{high} + p^{NH}\ p_{low})$

Home-MEG model: validation



HomeMEG model (p_{NH} = 0.025, p_H = 0.003, p_{high} = 0.07, p_{low} =0.0003) vs. Infocom06 trace

Home-MEG model: validation (2)



HomeMEG model (*p_{NH}* = 0.0133,*p_H* = 0.00011, *p_{high}* = 0.1, *p_{low}* =0.00001) vs. UCSD trace

Looking at parameters

Let us give a look to the values of parameters of best fit Home-MEG model for Infocom06 and UCSD trace

Parameter	Infocom 06	UCSD
p^{H}	3×10^{-3}	1.1×10^{-4}
p^{NH}	25×10^{-3}	13.3×10^{-3}
p_{high}	7×10^{-2}	10×10^{-2}
p_{low}	3×10^{-4}	1×10^{-5}
p_{HOME}	0.107	0.008
P_{high}/p_{low}	233.33	10000

Useful assumptions in the analysis: $p_{HOME} << p_{NHOME} = 1 - p_{HOME}$, and $p_{low} << p_{high}$

To do list and open problems

1. Can we ***formally prove*** the power law/exponential tail dichotomy in the Home-MEG model?

   A formal proof of the above mentioned dichotomy seems complex: the generic term Prob(*ICT*=*k*) is a high order polynomial with a number of terms exponential in *k*

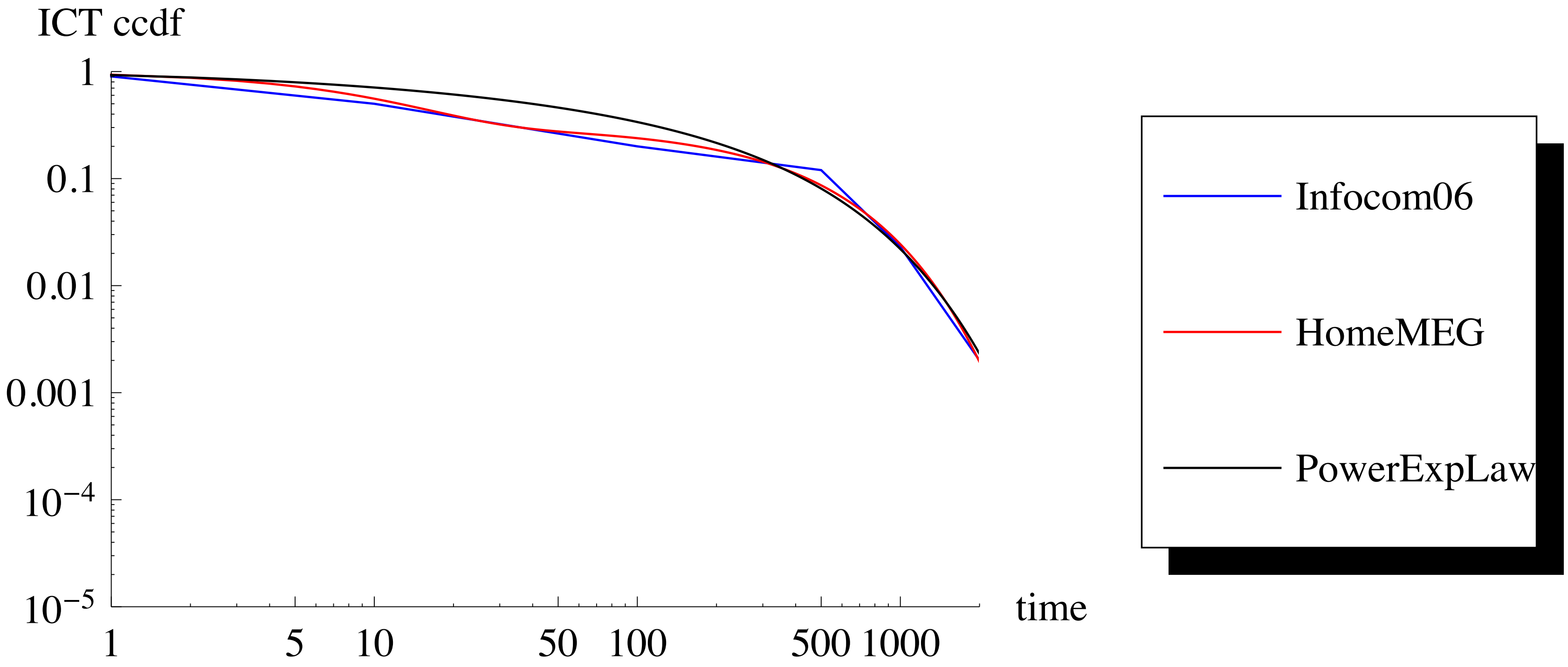
   We have empirically proven that Prob(*ICT*=*k*) can be approximated by an **power law distribution** with **exponential cutoff**:

   $\mathrm{Prob}(ICT=k) \approx (k-1)^{-\alpha} e^{-\beta(k-1)} / E_\alpha(\beta)$ and $\mathrm{Prob}(ICT > k) \approx (k-1)^{1-\alpha} E_\alpha(\beta(k-1)) / E_\alpha(\beta)$

   where $E_\alpha(\beta)$ is the exponential integral function defined as:

   $$E_\alpha(\beta) = \int_1^\infty \frac{e^{-\beta t}}{t^\alpha} dt$$

Home-MEG model: validating dichotomy



HomeMEG model vs. Infocom06 trace vs. Power law with Exponential cutoff (α = 0.829, β = 0.0018)

To do list and open problems

2. Study the dynamics of **information propagation** in Home-MEG networks

3. The **"social structure"** is only ***implicitly*** accounted for in the Home-MEG model. Can we generalize the Home-MEG model ***explicitly*** taking into account **"social structure"**?

To do list and open problems (2)

Possible Social-HMEG model:

- ✓ A network of n nodes is modeled through m_1 Home-MEGs of **type 1**, and $n(n\text{-}1)/2 - m_1$ Home-MEGs of **type 2**
- ✓ **type 1** Home-MEG: models contacts between nodes in the same **"community"** → $p_{high} >> p_{low}$
- ✓ **type 2** Home-MEG: models contacts between nodes in different **"communities"** → $p_{high} \approx p_{low} \approx 0$

Open question:

Does the (aggregate) ICT distribution in the Social-HMEG model display the power-law+exponential tail dichotomy?