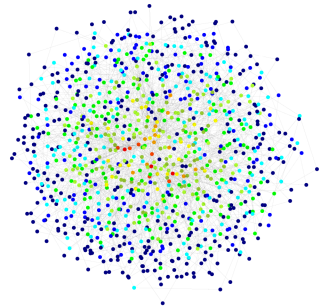
Information Spreading in Last.fm Online Social Network

Róbert Pálovics, András Benczúr

Informatics Laboratory, Department of Computer and Automation Research Institute, Hungarian Academy of Sciences



supported by EC FET Open project NADINE

July 26, 2012

Introduction

Last.fm

Densification in artist subgraphs

Temporal influences

Summary

Problem

- Characterize *information diffusion*, or *information spreading* by investigating online social networks

LAST.FM

- About Last.fm
  - Leading online service in music based social networking
  - "Scrobbling": collecting listening activity of users
  - Recommendation system for users
  - Social network
  - Public statistics
- Influences
  - People often share their musical taste
  - They recommend each other new artists, albums, tracks
  - Directed influences

Main goal

- User-user social network, with scrobble time series
- Justify the existence of influences, i.e. correlation between individuals and the listening behavior of their contacts

Large dataset

- Available for us under NDA for Last.fm
- Selection criteria
  - Location is stated as UK
  - Date of birth between 1/12/1961 and 1/12/1997
  - Profile displays scrobbles publically
  - $5 \cdot 10^{-5} 1/s \leq Act < 10^{-2} 1/s$, where

    $$Act = \frac{\text{\# scrobbles}}{\text{elapsed time from registration}}$$
  - Scrobbles between January 2010 and December 2011
- Size
  - 10^5 users, $3 \cdot 10^5$ edges, 10^9 scrobbles

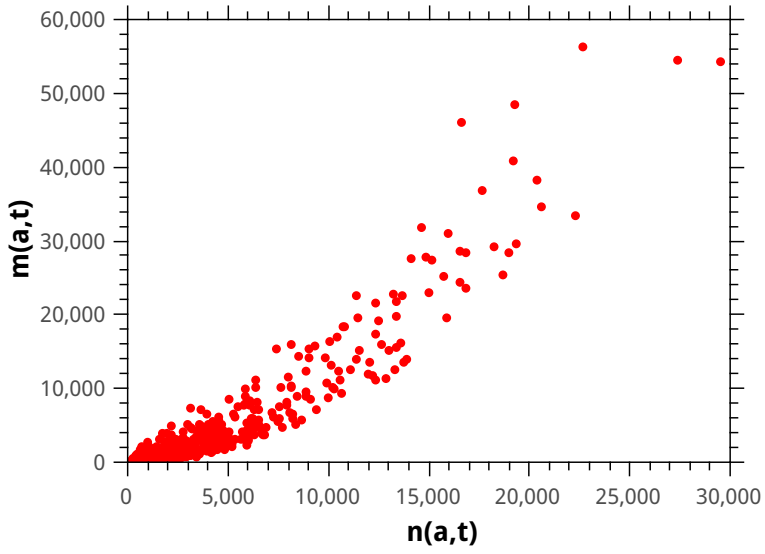
ARTIST SUBGRAPHS

- For artist a in time t

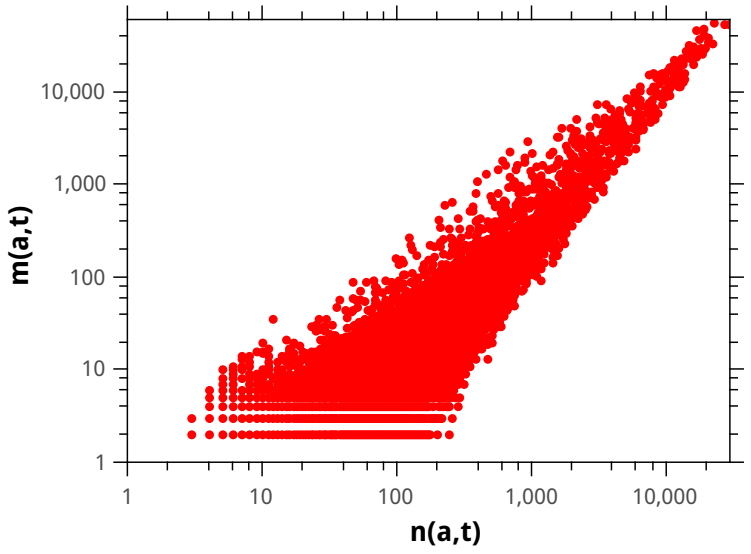
  $$G(a,t) = \{\text{subgraph of users who listened to } a \text{ before } t\}$$

- *Main result:*
  - Increased edge density in $G(a,t)$
  - The number of edges $m(a,t)$ is power-law function of the number of users $n(a,t)$ in the subgraph with exponent ≈ 1.535

MEASUREMENTS



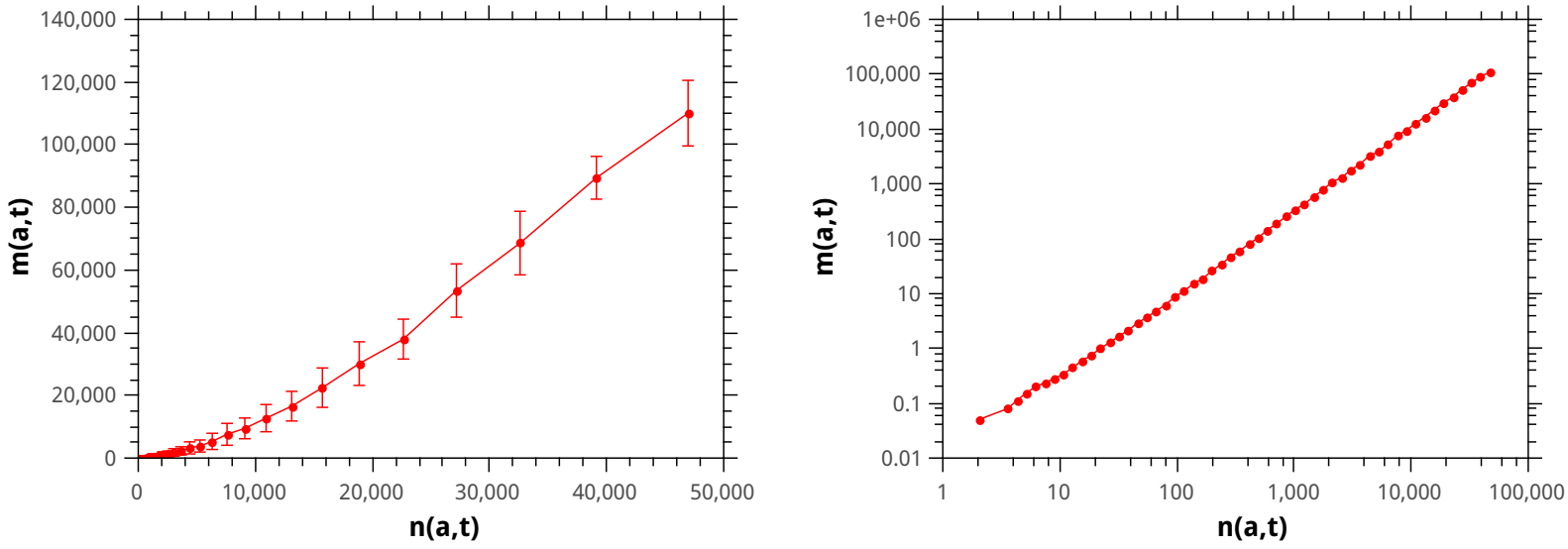
(a) linear scale



(b) log scale

Figure: # edges as the function of vertex number

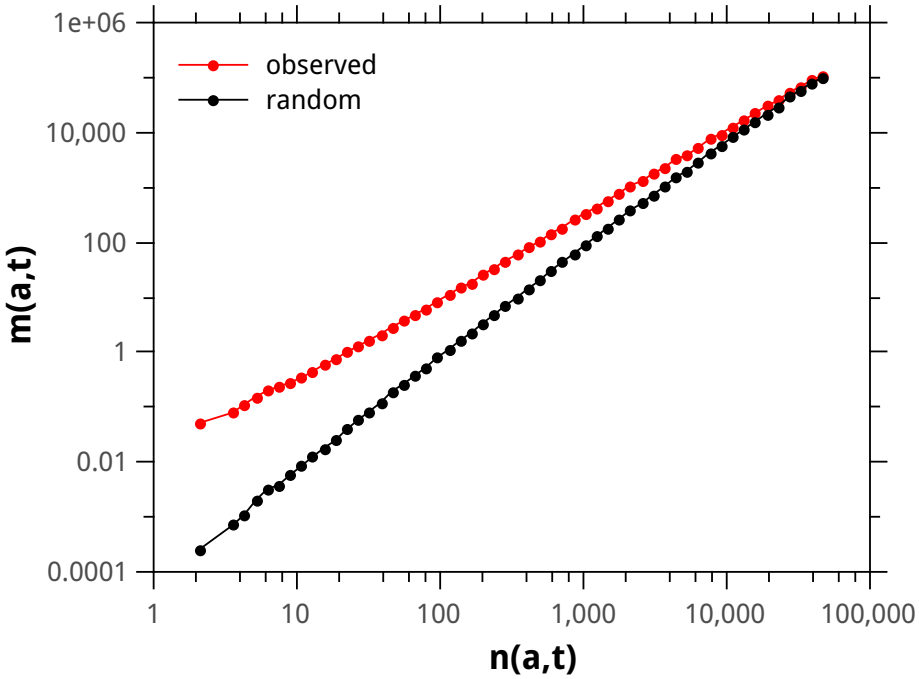
MEASUREMENTS



(a) linear scale (b) log scale

Figure: # edges as the function of vertex number

Difference from the uncorrelated case



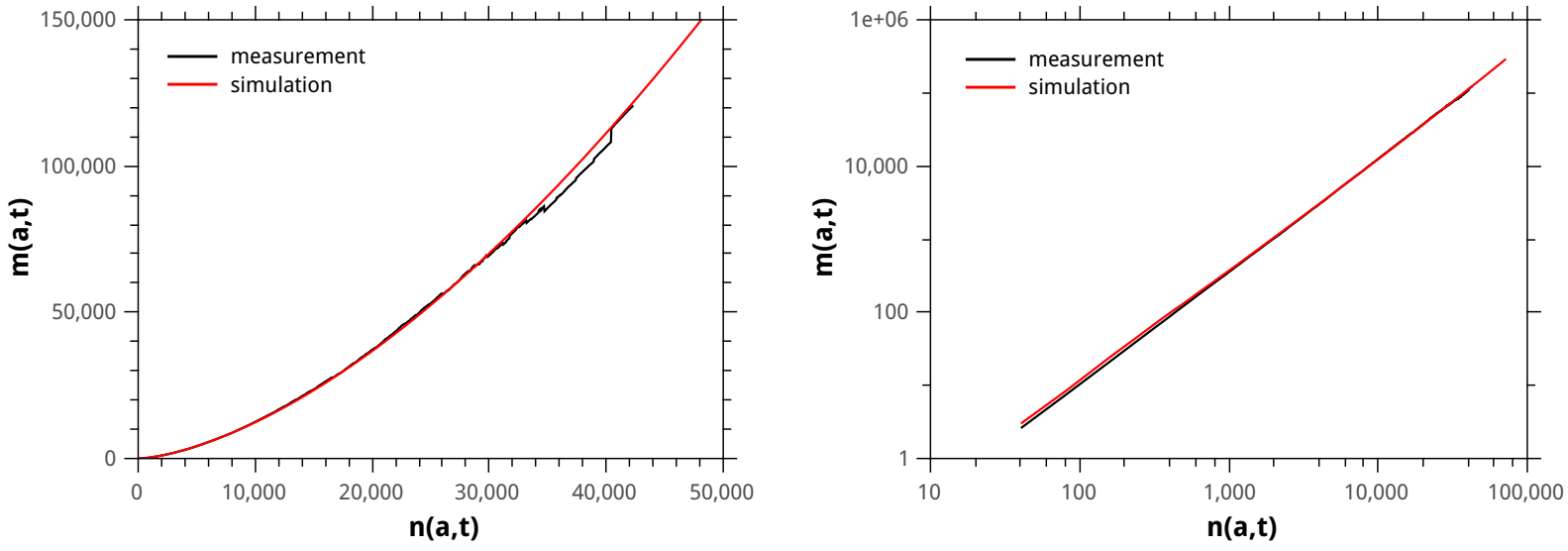
- Larger graphs are denser
- But small artist subgraphs are much denser than random subgraphs

Proposed model

- Fixed friendship network
- At each time, select a new fan of a
- Select randomly proportional to its weight $w(i)$
- Initial weight $w(i)$ is uniform
- User i gets weight at time t when first neighbor j listens to a

$$w(i) = 1 + \frac{C}{d(j)} \left(\frac{n(a,t)}{N} \right)^{-\alpha}$$

Simulation result



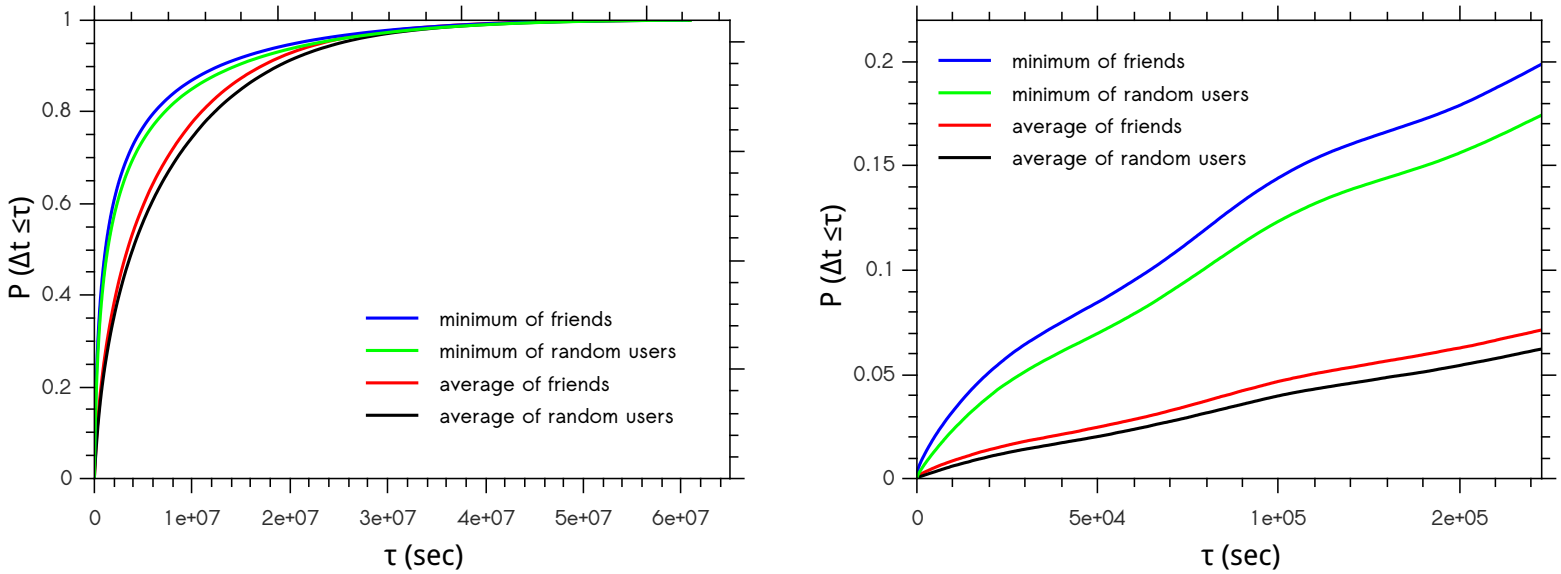
(a) (b)

Figure: Simulation result

Temporal influence

- User j is influenced by user i
- User j listens to a at the first time at t
- If i listens to a at time $t - \Delta t$
- We compute $\overline{\Delta t}$ and Δt_{min} in case of friends and random users.

CDF CURVES



(a) whole distribution (b) first 48 hour

Figure: CDF curves

STRENGTH OF INFLUENCES

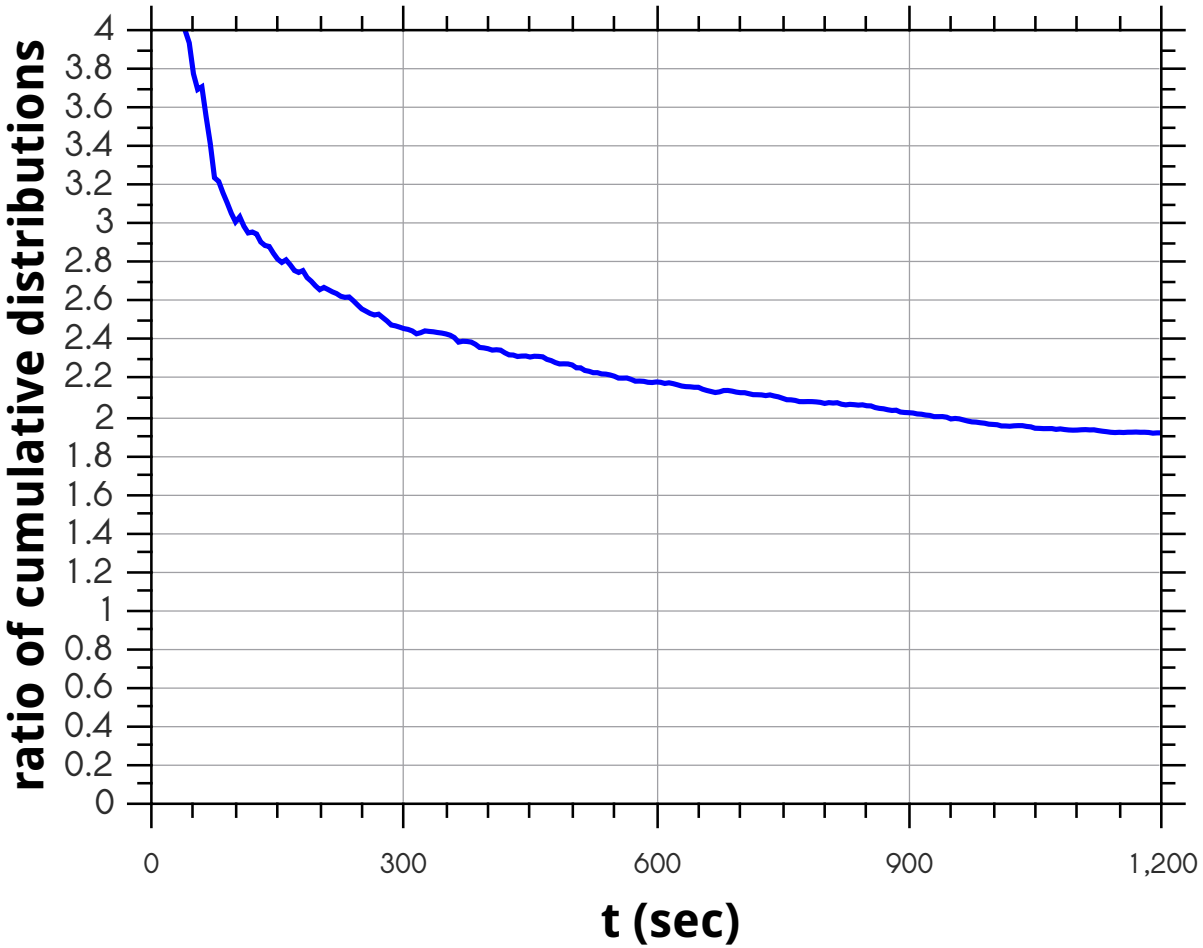
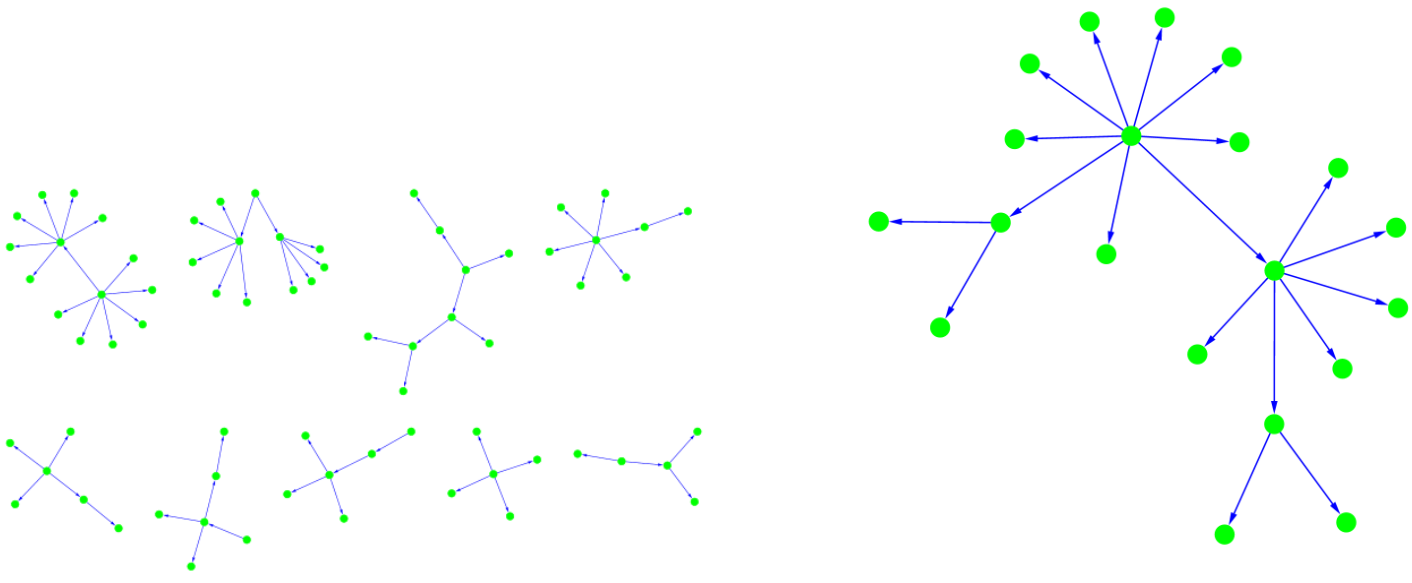


Figure: Ratio of cumulative distributions in the first 20 minutes

INFLUENCES



(a) (b)

Figure: Influences

Summary

- Work in progress over the Last.fm community
- Large dataset
- Increased edge density in artist subgraphs
- Proposed model for artist subgraphs
- Temporal influences
- Strength of influences