

# Building Client's Risk Profile Based on Call Detail Records

ZALA HERGA, CASEY DOYLE, STEPHEN DIPPLE, CALEB NASMAN, GYORGY KORNISS, BOLESŁAW SZYMANSKI, JANEZ BRANK, JAN RUPNIK, AND DUNJA MLADENIĆ

---

Artificial Intelligence Laboratory, Jožef Stefan Institute

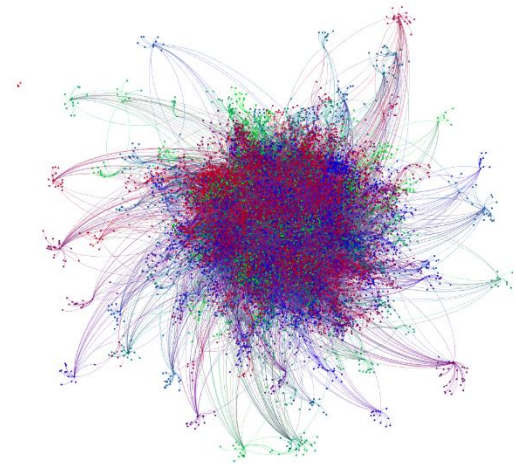
Rensselaer Polytechnic Institute, Troy, NY

SiKDD, October 9<sup>th</sup>, 2017

# Introduction

---

- Network built from mobile phone communication event logs
- Extraction of behavioral patterns
- Modeling user defaults



# Outline

---

## Data

### Mobile phone network

- Construction
- Network properties

## Modeling

- Feature extraction
- Evaluation

# Data

---

- Call detail records (CDR):
  - User A ID,
  - User B ID,
  - Communication event type (call, text, data)
  - Direction (incoming, outgoing)
  - Date and time,
  - Cell tower ID,
  - Duration
  - ...
- Three months
- Anonymized

# Mobile phone network

---

- Nodes = clients
- Edges = social link between clients
- Two approaches to network construction:
  - Weighted directed network (weight = frequency of communication)
  - Unweighted directed network (varying cutoffs)

# Network properties

---

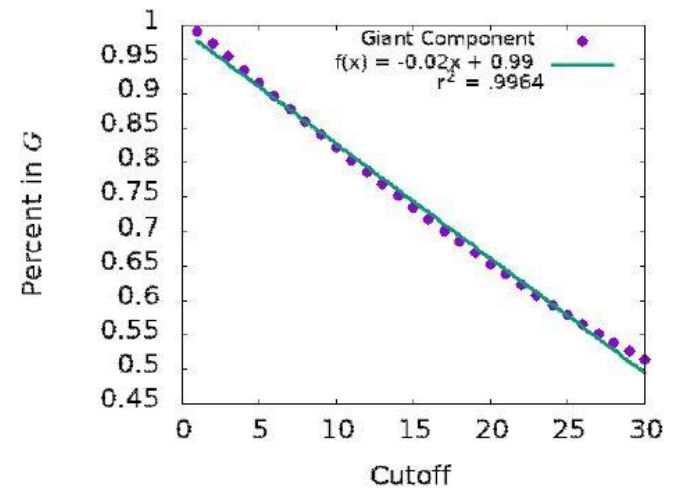
- Giant component: more than 99% of users

- Closeness centrality

  - shortest path between any two nodes is low

- Harmonic centrality

  - even the nodes that are not connected to the giant component tend to form small, tightly connected communities of their own



# Features

---

A few thousand features:

- Consumption
- Social network
- Mobility/Geographic
- Node level network measures

Many of those on granular level:

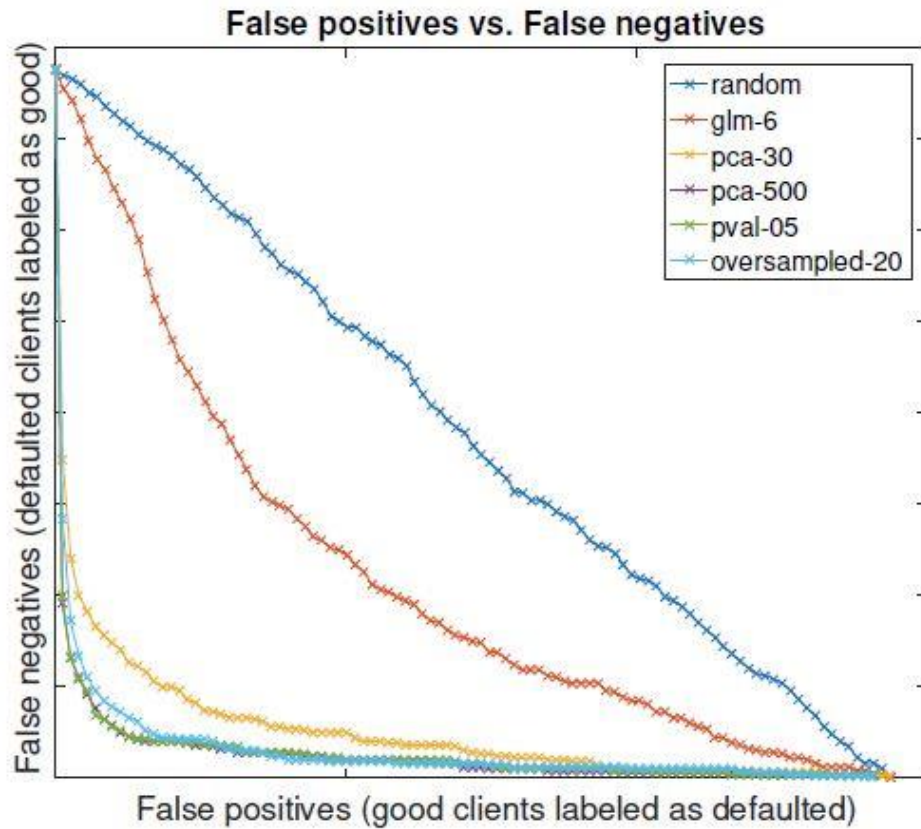
- incoming/outgoing,
- voice/data,
- per hour of week, night/morning/day/evening, day of week

# Modeling

---

- PCA analysis
- Linear model
  - used on PCA transformed features
- Unbalanced dataset: oversampling defaulted clients
- Train vs. test set: 70-30





Model	random	Glm-6	Pca-30	Pca-500	Pval-05	Oversampled-20
Recall	0.05	0.13	0.79	0.90	<b>0.91</b>	0.88
Precision	0.003	0.007	0.042	0.049	<b>0.049</b>	0.047

# Conclusion and Future work

---

- Behavioral patterns can be well described with network features
- Deepen our understanding of
  - the effect of communities
  - spread and propagation of individual's influence