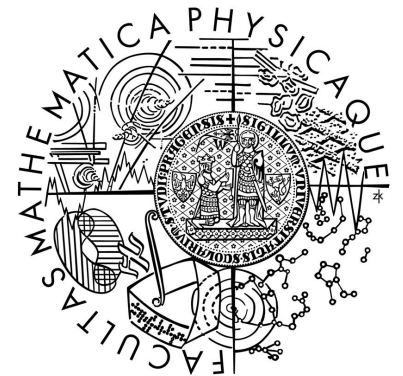


Czech Insolvency Proceedings Data: Social Network Analysis

Iveta Mrázová, Peter Zvirinský

Department of Theoretical Computer Science and Mathematical Logic
Faculty of Mathematics and Physics
Charles University in Prague



Outline

▷ Introduction

- Insolvency proceedings (IPs) in the Czech Republic (and EU and US)
- Insolvency Registers

▷ Dataset preparation

- Extraction of missing data

▷ Supporting experiments

- Dynamic social network analysis
- Link prediction

▷ Conclusions

Insolvencies in the Czech Republic

➔ **180 000**
insolvencies since 2008

➔ **\$460 millions**
of company debts in 2012 only

➔ **7%**
average return rate to creditors

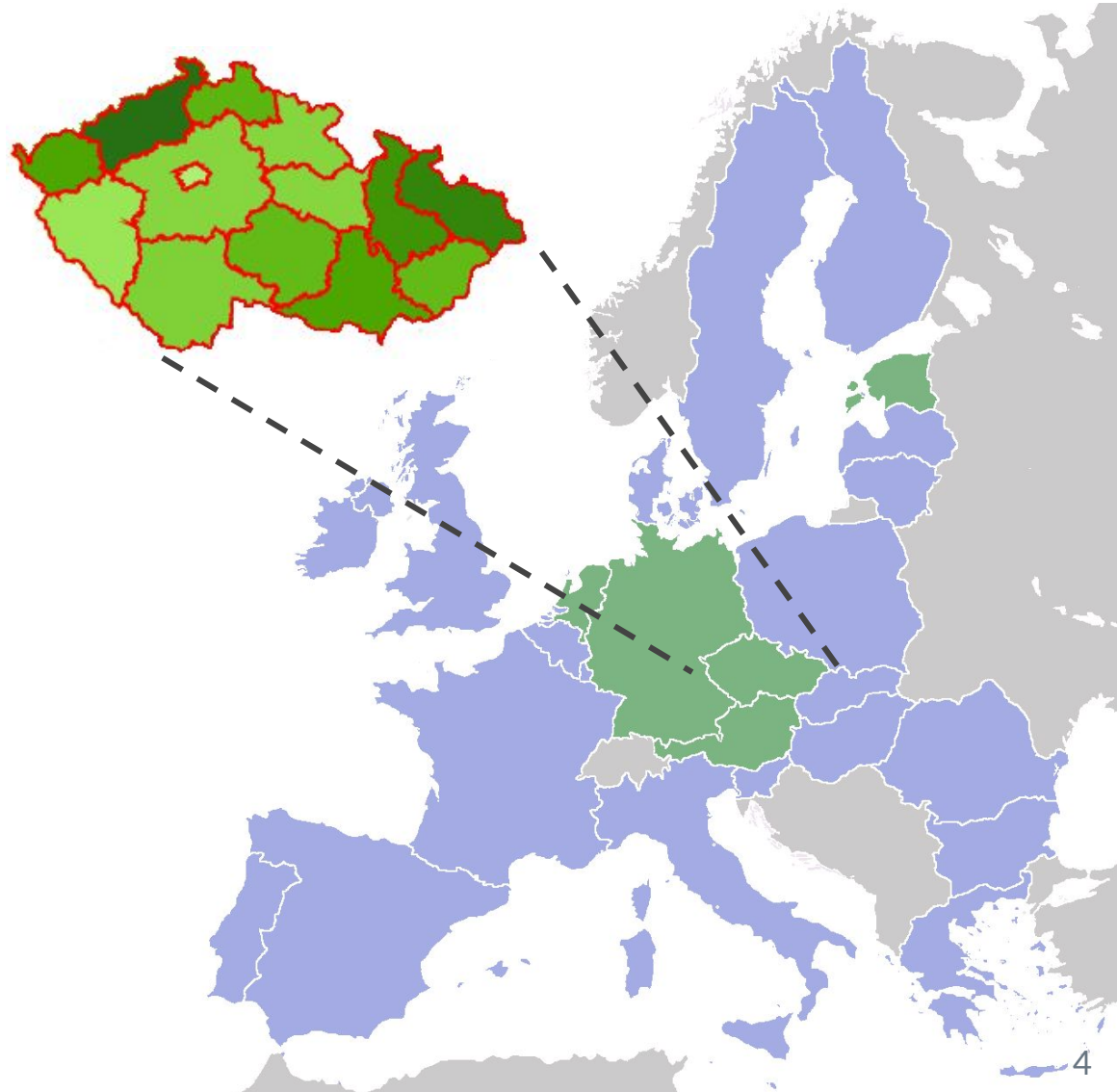
Introduction

Insolvency registers in EU

European e-Justice Portal supports unified search across insolvencies from:

- ▷ Austria
- ▷ **Czech Republic**
- ▷ Estonia
- ▷ Germany
- ▷ Netherlands

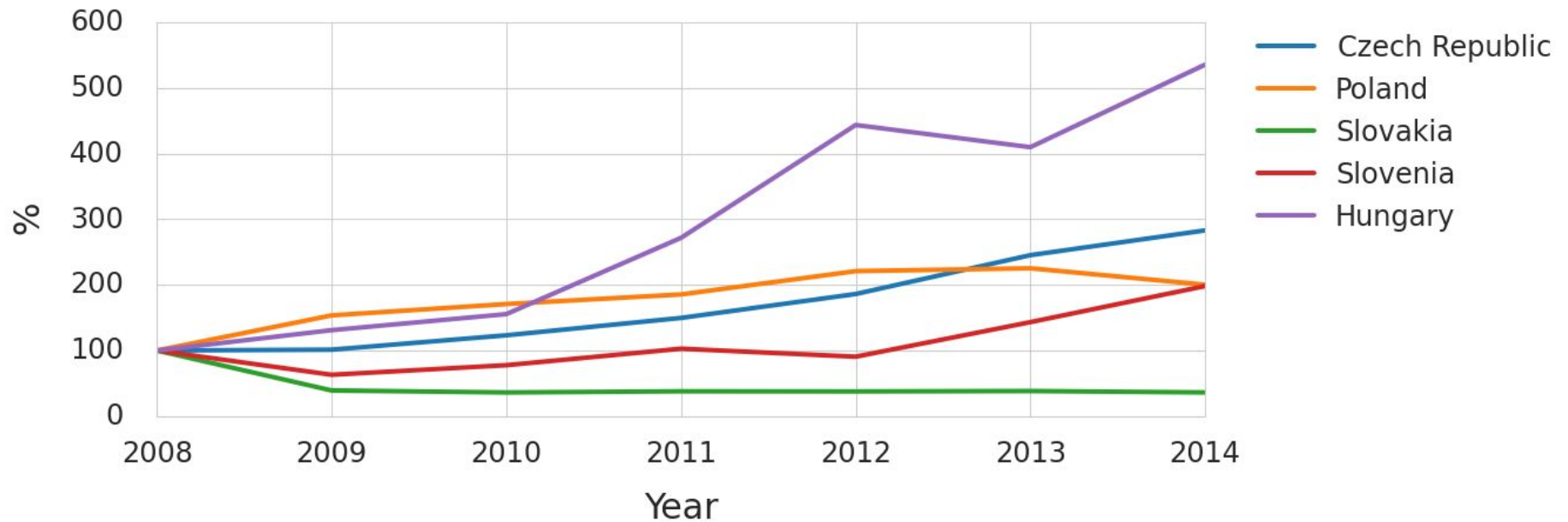
That is **only 5 out of current 28** member states.



Introduction

Insolvencies - Eastern Europe

Development of insolvencies in selected countries

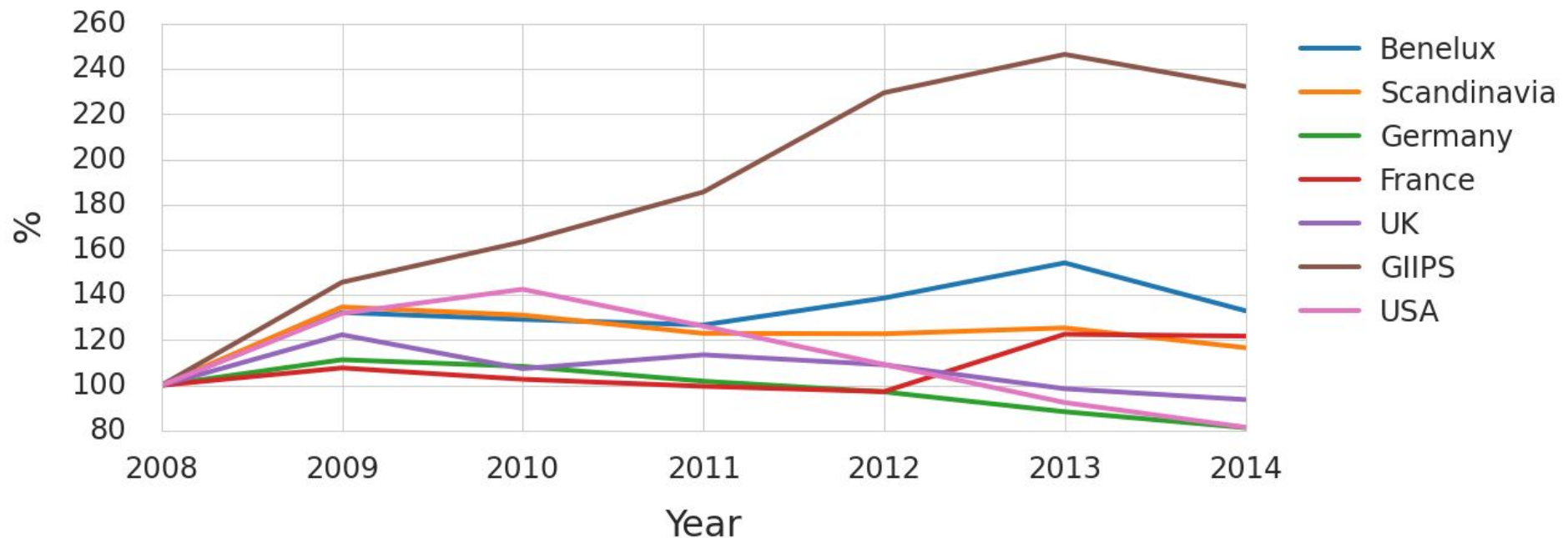


2008 = 100%

Introduction

Insolvencies - Western Europe (and US)

Development of insolvencies in selected countries



Benelux = Belgium, Netherlands, Luxemburg

GIIPS = Greece, Ireland, Italy, Portugal, Spain

Introduction

Insolvency Register of the Czech Republic

Insolvency proceeding detail **Petra Dudová**

Actual state:

Checked Off

Reference Number:

KSOS 14 INS 1311 / 2010 at Region court in Ostravě

Name:

Identification

Petra Dudová

Company ID:

Date of birth:

Address:

Haviřov - Město, Mozartova 1060/1, PSČ 736 01

Insolvency administrator:

Insolvency administrators

Ing. Václav Dlouhý

Last update:

Historie insolvenčního řízení

31.05.2010

senate

debtor

administrator

creditor

Section A - proceeding before bankruptcy | Section B - proceeding after bankruptcy | Section P - Applications of receivables

	Publication date		Description	Document	Creditors	Senate number
P1 - 1.	01.03.2010	14:08	Application of receivables	Full text (581 kB)	GE Money Bank, a.s.	

Dataset preparation

Missing data

- Missing creditors
 - available since late 2011
 - missing in 300 000 cases
 - available only in scanned documents
- Optical Character Recognition
 - Tesseract OCR

PŘIHLÁŠKA POHLEDÁVKY

Soud: **Krajský soud v Ústí nad Labem** Spis. značka: **KSUL 45** INS: **26550** / 2013

Dlužník

☒ 01 Fyzická osoba ☐ 02 Právnícká osoba Státní příslušnost¹

Osobní údaje: Příjmení: **Šindlerová** Jméno: **Věra**
 Titul za jm.: Titul před jm.:
 Datum narození²: **13.8.1957** Rodné číslo: **575813/2039**

Údaje o podnik.³ IČ: **7 2 5 8 2 5 6 1** Jiné registr.č.:

Bydliště/sídlo: Ulice: **Strojetice** Č.p./č.e.: **34** Co.:
 Obec: **43986** PSČ: **75363** Část obce: **Kryry**
 Stát: **Česká republika**

Věřitel

☒ 03 Fyzická osoba ☐ 04 Právnícká osoba Státní příslušnost¹

Osobní údaje: Příjmení: **Vaňková** Jméno: **Jarmila**
 Titul za jm.: Titul před jm.:
 Datum narození²: Rodné číslo:

Údaje o podnik.³ IČ: **4 2 2 0 2 8 7 6** Jiné registr.č.:

Bydliště/sídlo: Ulice: **Psinice** Č.p./č.e.: **8** Co.:
 Obec: **Libáň** PSČ: **50723** Část obce: **Libáň**
 Stát: **Česká republika** Číslo účtu: **2088792143/0800**

☒ 05 Korespondenční adresa⁴ Ulice: **Petra Bezruč** Č.p./č.e.: **362** Co.:
 Obec: **Nymburk** PSČ: **28842** Část obce: **Nymburk**
 Stát: **Česká republika**

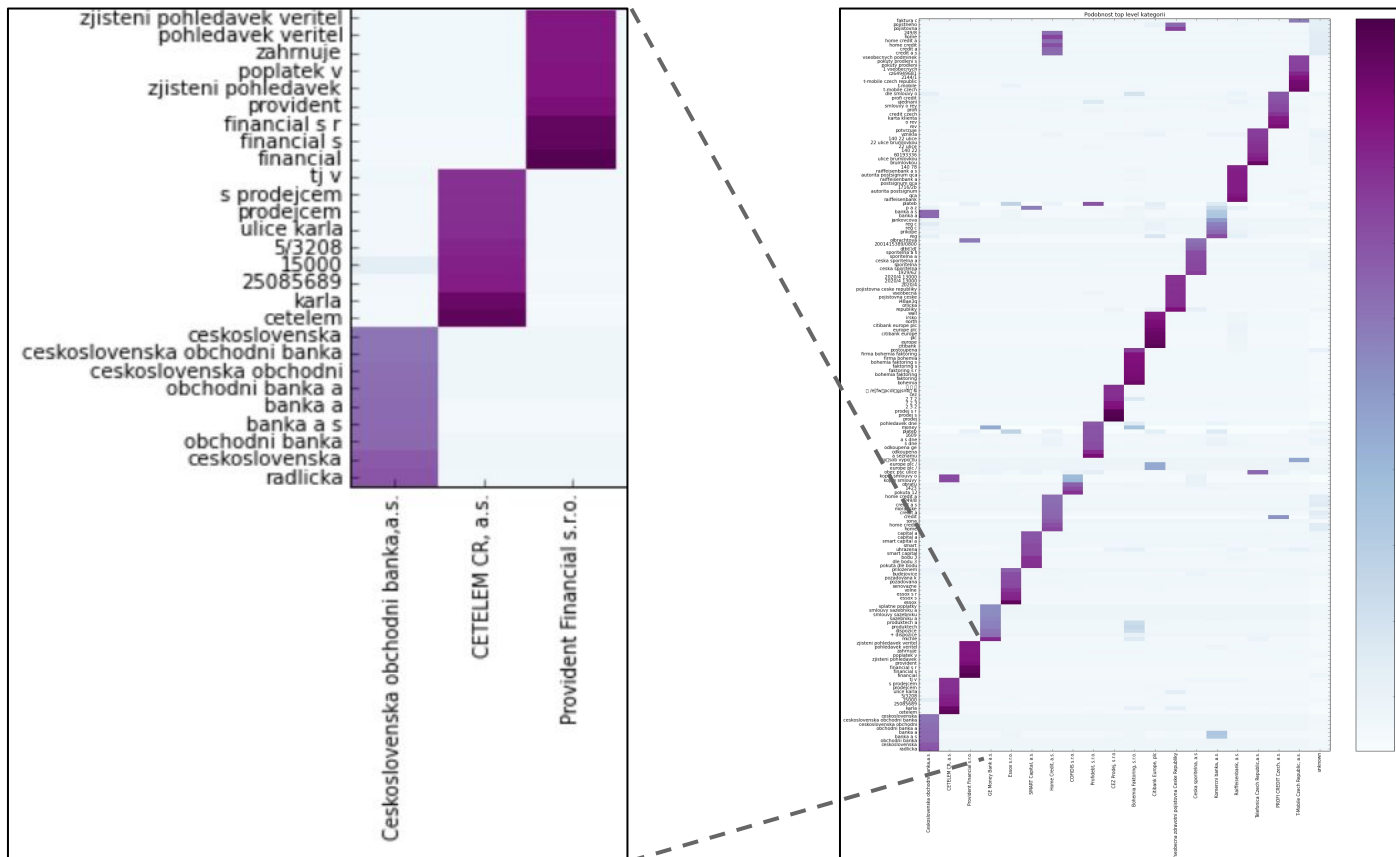
Elektronická adresa: Akreditovaný poskytovatel certifikačních služeb:

¹ vyplní se pouze u zahraničních osob
² datum narození se vyplní, pokud nebylo přiděleno rodné číslo
³ vyplní se pouze tehdy jedná-li se o fyzickou osobou podnikatele
⁴ vyplní se pokud se liší od sídla či trvalého bydliště

Dataset preparation

Creditors' extraction

- ▷ Text classification (Logistic Regression)
 - Recall: 0.96, Precision: 0.96



Experiments

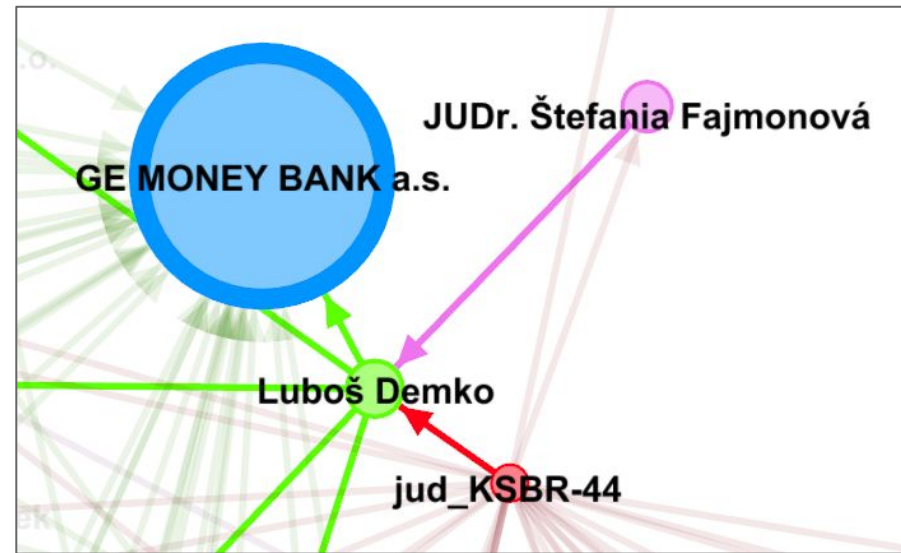
Insolvency Register as a Social Network

▷ nodes:

- debtors
- creditors
- senates
- insolvency administrators

▷ relations

- debtor owes a creditor
- administrator manages the insolvency of a debtor
- senate handles the insolvency proceeding of debtor



Experiments

How to analyze the insolvency network?

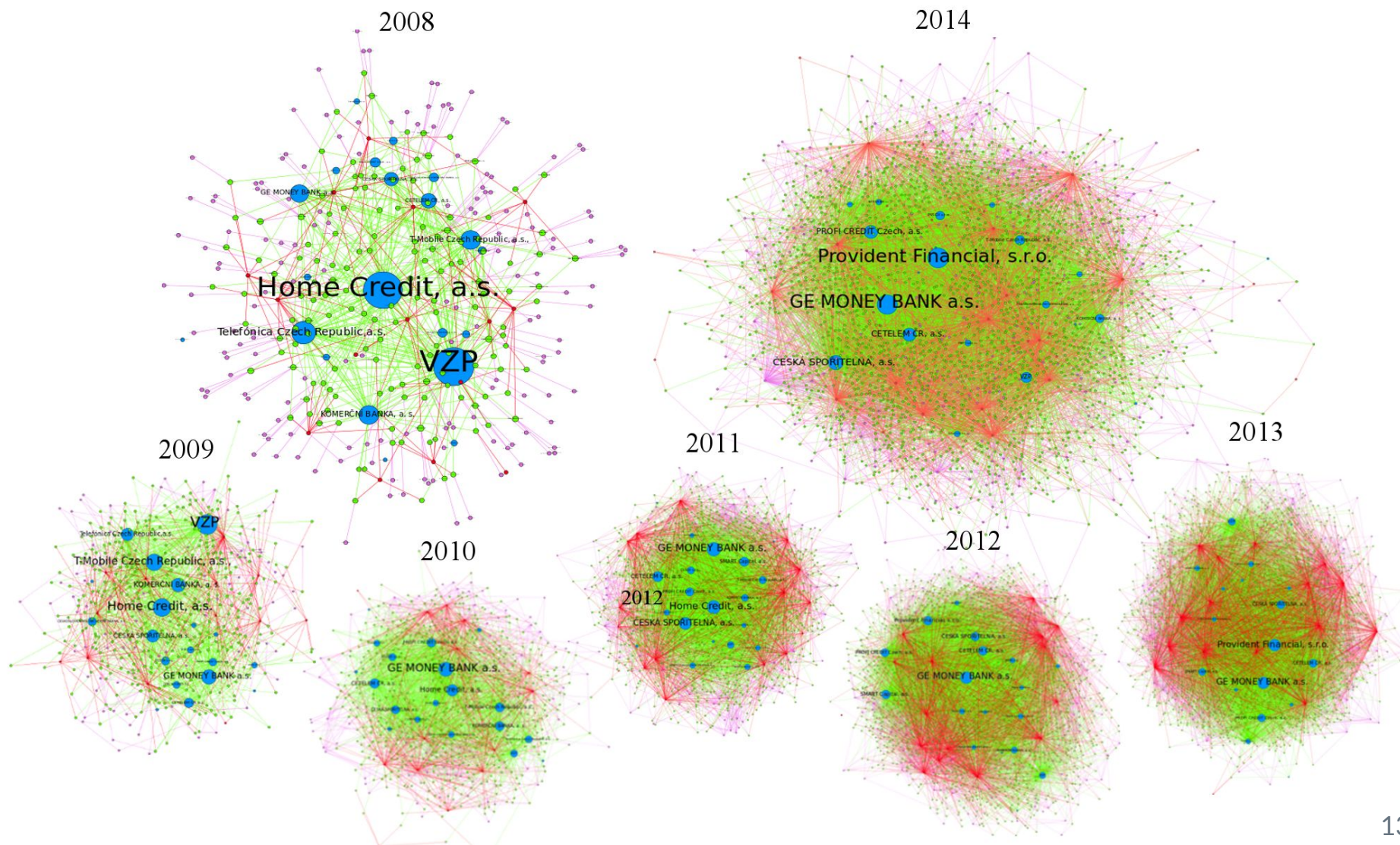
- ▷ study social entities and their relationships in the insolvency network
- ▷ main areas of interest:
 - the importance of the nodes
 - Centralities, PageRank, HITS
 - the structure of the network
 - community discovery, clustering
 - network dynamics and evolution

Social network of the Insolvency Register

-

Experiments

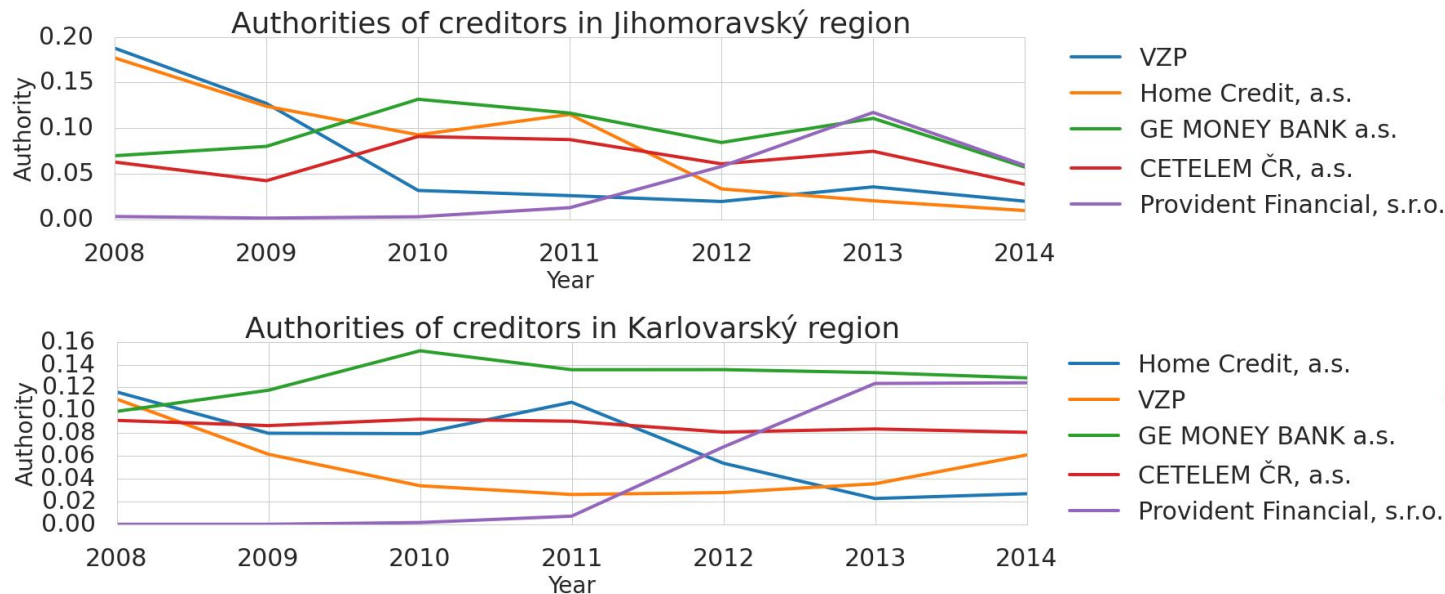
Does the network change over time?



Experiments

Capturing trends

- Can we effectively track the importance of the node in the network over time?

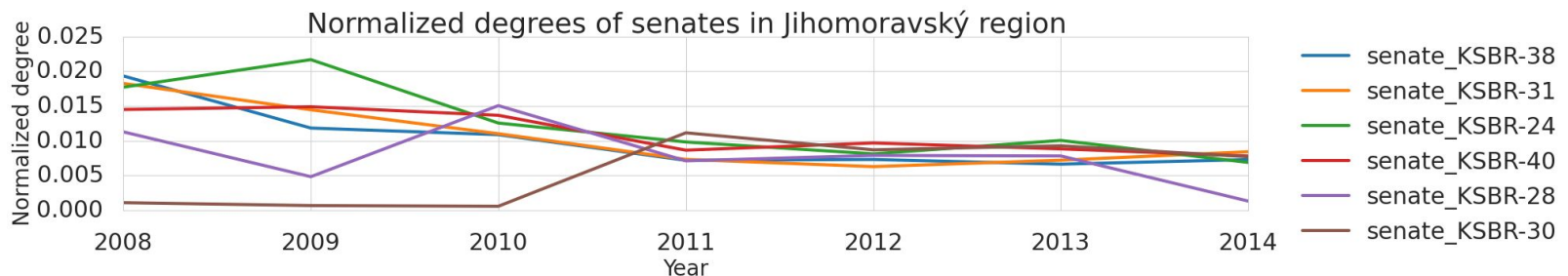
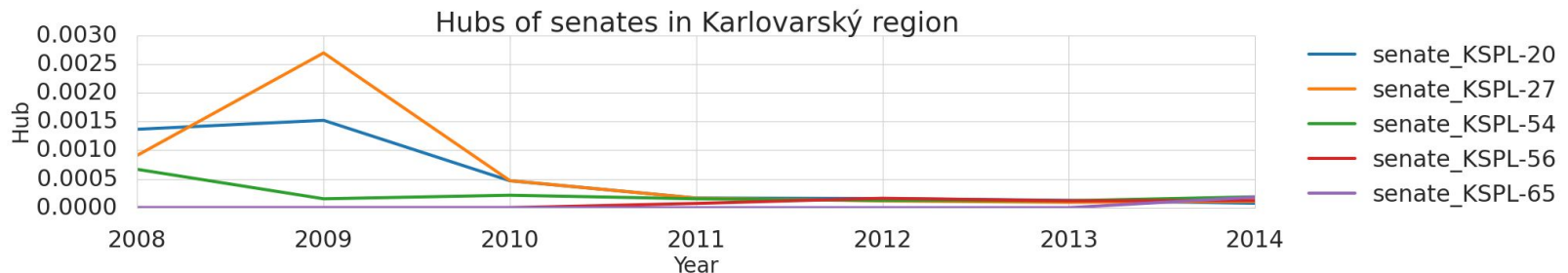
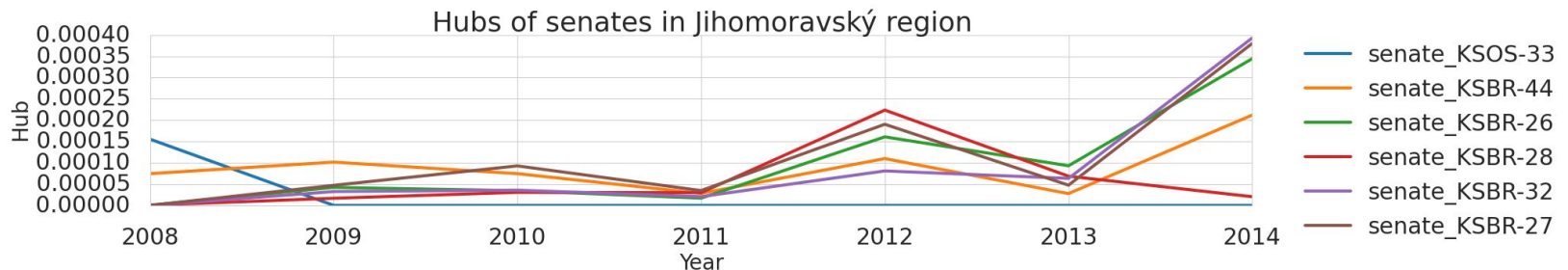


- Best results obtained by the HITS algorithm
 - authorities, hubs

Experiments

Capturing trends

▷ Why are simpler methods not good enough?



▷ They fail to capture any changes in the network

Experiments

Link prediction

- ▶ Can we predict what subjects are likely to participate in an IP together given a senate of city Brno?
- ▶ Association rule mining
 - Apriori, FP-growth

LHS	RHS	Support	Confidence	Lift
PersonType: Natural 2009	Administrator: Mgr. Vladimíra Zúkalová	9	0.26	165.32
Creditor: VZP PersonType: Legal 2011	Creditor: ČSSZ	10	0.71	40.07
Creditor: VZP PersonType: Legal 2012	Creditor: ČSSZ	9	0.69	38.83

Conclusions

- ▷ machine learning methods can extract additional information from unstructured public data
 - OCR frameworks together with advanced text mining algorithms
- ▷ link analysis algorithms can identify the main players in dynamic networks
 - advanced methods such as HITS outperform simpler methods
- ▷ association rules can be used to predict future changes in the network

Thanks!

Any questions?

You can find us at:

iveta.mrazova@mff.cuni.cz

peter.zvirinsky@gmail.com