Similarity patterns, topological information and credit scoring models

Alessandro Spelta, Branka Hadji-Misheva and Paolo Giudici

Fintech laboratory
Department of economics and management
University of Pavia

July 2, 2019

A FINancial supervision and TECHnology compliance training programme

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825215 (Topic: ICT-35-2018 Type of action: CSA)

Motivation. Financial Technologies bring opportunities (competitive prices, improved user experience, wider inclusion) but also risks (credit risks, market risks, cyber risks), amplified by the interconnectedness of fintech platforms (systemic risks).

Aim. The Horizon2020 FIN-TECH project aims at building a fintech risk management platform, which measures risks to make fintech innovations sustainable, for both RegTech and SupTech purposes.

Method. The aims will be achieved creating a knowledge exchange hub, which will eventually lead to a research sandbox laboratory.

Participants. i) Project partners, who research and develop fintech risk management models; ii) National supervisors of 29 European countries, who give feedback in SupTech workshops; iii) European fintechs and banks, who give feedback in RegTech workshops; v) international regulators and advisors, who supervise and evaluate the developed models.

What we do?

How we do it?

RESEARCH FRAMEWORK:

6 RESEARCH WORKSHOPS

3 HORIZONTAL WORKSHOPS, TO DEVELOP USE-CASES BASED ON REGULATORS' PRIORITIES

3 VERTICAL WORKSHOPS, TO VALIDATE THE DEVELOPED USE CASES

SUPTECH FRAMEWORK:

3 X 29 SUPTECH WORKSHOPS

EACH WORKSHOP CONSISTS OF **16 HOURS OF TRAINING** FOR THE CORRESPONDING NATIONAL SUPERVISOR, BASED ON THE **USE-CASES** DEVELOPED BY THE PROJECT

REGTECH FRAMEWORK:

6 REGTECH WORKSHOPS

, EACH WORKSHOP CONSISTS **OF 6 HOURS OF PRACTICAL TRAINING ,** WHERE FINTECHS AND BANKS CAN REPLICATE THE PROJECT'S **USE CASES** THROUGH CODING SESSIONS

EVALUATION FRAMEWORK:

ALL PROJECT'S WORKSHOP FEEDBACKS ARE ELABORATED INTO AN INTERMEDIATE AND A FINAL EVALUATION REPORT.

ALL PROJECT'S DELIVERABLES ARE EVALUATED BY A PANEL OF ADVISORS WHO ELABORATE AN OVERALL EVALUATION REPORT

COMMUNICATION FRAMEWORK:

A DEDICATED WEBSITE INTEGRATES ALL PROJECT DELIVERABLES (RESEARCH PAPERS, USE CASES AND TRAINING SLIDES)

THE WEBSITE IS LINKED WITH **SOCIAL NETWORK CHANNELS**, TO ENGAGE ALL STAKEHOLDERS.

Who are we?

Project partner	Lead	Country
University of Pavia	Paolo Giudici	Italy
Humboldt University Berlin	Wolfgang K. Härdle	Germany
ZHAW Applied Sciences	Jörg Osterrieder	Switzerland
University College London	Tomaso Aste	UK
Bucharest University	Vasile Strat	Romania
WU Vienna	Ronald Hochreiter	Austria
Panteion University	Veni Arakelian	Greece
INESC-TEC	Paula Brito	Portugal
University of Paris 1	Prof. Christophe Henot	France
Politecnico of Milan	Emilio Barucci	Italy
University College Dublin	Andreas Hoepner	Ireland
University of Luxembourg	Radu State	Luxembourg
Jozef Stefan Institute	Marko Grobelnik	Slovenia
University of Warsaw	Piotr Wojcik	Poland
University of Rjeka	Saša Žiković	Croatia
Universidad Complutense de Madrid	Javier Arroyo	Spain
University of Economics in Bratislava	Jana Peliova	Slovakia
Kaunas University of Technology	Audrius Kabasinkas	Lithuania
Masaryk University Brno	Oleg Deev	Czech Republic
Varna University of Economics	Stefan Vachkov	Bulgaria
University of Tampere	Lasse Koskinen	Finland
B-Hive	Dave Remue	Belgium
Modefinance	Valentino Pediroda	Italy
Firamis	Jochen Papenbrock	Germany

Project network

The project network includes:

i) 24 partners: 21 universities, 3 fintechs

ii) 6 European fintech hubs

iii) The national supervisors of all 28 EU countries plus Switzerland

iv) 8 international regulators and supervisors (BIS, IMF, OECD, EC, EBA, ESMA, EIOPA, ECB)

v) A panel of International advisory board members

UNIVERSITIES and RESEARCH CENTRES	FINTECH HUBS and ASSOCIATIONS	REGULATORS and SUPERVISORS
UCL, zhaw, UCD Michael Smurfit Graduate Business School, WU, Tampere University, ktu, uni.lu, University of Warsaw, Universidad Complutense, Politecnico Milano 1863, Panthéon Sorbonne, MUNI, Jožef Stefan Institute, Panteion University, INESCTEC	ABI Lab, ASSOFINTECH, Swiss Finance + Technology Association, B-HIVE, Fintech District, modefinance, AFGC, FIRAMIS, Fintech e Insurtech, TechQuartier, ITALIAFINTECH, Swiss Fintech Innovations, Digital Magics, francefintech	AMF, Bank of Greece, Lietuvos Bankas, Central Bank of Ireland, CNMV, CNB Czech National Bank, CONSOB, Deutsche Bundesbank, MFSA, KNF, Banka Slovenije, FCA, Finantsinspektsioon, CMVM, Assemblée Nationale

P2P Platforms, introduction

- In the context of the consumer and commercial credit, **FinTech** solutions have introduced many opportunities for both lenders and borrowers redefining the role of traditional intermediaries
- Peer-to-peer lending platforms (**P2P**) allow private individuals to directly lend loans to private borrowers and/or small and medium enterprises (SME)
- The concept peer-to-peer captures the **interconnectdness** between units, which eliminates the need for a central intermediary

P2P Platforms, Pros&Cons

- **Advantages** of P2P Lending Platforms:
  - Improved financial inclusion
  - Customized user experience
  - Higher returns for lenders, lower fees for borrowers
- **Disadvantage** of P2P Lending Platforms:
  - P2P platforms may produce inadequate credit risk measurements
  - interconnectdness increases systemic risks x

Banks vs P2P Platforms: Risk concerns and data availability

- Both classic banks and P2P platforms rely on **credit scoring models** for estimating credit risk but the incentive for model accuracy is different:
  - Banks assume the risk so they are interested in having the most accurate possible model
  - In P2P lending platforms, the risk is fully borne by the lender
- P2P Platform often do not have access to historical data employed by banks
- P2P Platforms operate as **social network:** and, therefore, generate alternative transactional data.

Scope of the Paper

- Evaluate the **predictive performance** of the scoring models usually employed by P2P platforms.
- Improve scoring models using alternative transactional data
- Construct network based models using a proxy of trasactional data based on the similarity between balance sheet data,

Similarity Network

- In similarity networks **nodes** represent borrowing companies and the **edges** the distance between adjacent nodes.
- There exist different metrics to build up **distances** between objects: we apply the standardized Euclidean distance

$$d_{i,j} = (\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\Delta}^{-1}(\mathbf{x}_i - \mathbf{x}_j)'$$

- There exist different algorithm to extract a **sparse representation** of the fully connected distance matrix: we apply the Minimum Spanning Tree
- At each subsequent step, two clusters l_i and l_j are merged into a single cluster if:

$$d\left(l_i, l_j\right) = \min\left\{d\left(l_i, l_j\right)\right\}$$

with the distance between clusters being defined as:

$$d\left(l_i, l_j\right) = \min\left\{d_{rq}\right\}$$

with $r \in l_i$ and $q \in l_j$.

Topological Coefficients

- **Nodes Importance**
  - How many partners a company has (degree)?

$$k_i = \sum_{j=1}^{N} \hat{\mathbf{D}}_{ij} \tag{1}$$

  - How strong are the weights embedded in such connections (strength)?

$$k_i = \sum_{j=1}^{N} \mathbf{D}_{ij} \tag{2}$$

  - How crucial a node is in letting information to spread over the network (PageRank)

$$\pi_{in} = \varepsilon \left(\mathbf{D\Phi} + \mathbf{f}\mathbf{d}'_{out} \right) \pi_{in} + (1 - \varepsilon)\, \mathbf{f} \tag{3}$$

Similarity Network

- We compare different **credit scoring models**
  - Logistic regression
  - Discriminant analysis
  - Naive Bayes model
  - Support vector machine
  - Decision Trees
- We use **non parametric measures** to asses models' performance
  - Area under the ROC curve (AUC)
  - Area under the PR curve (AUPR)
  - Model accuracy (ACC)
  - F1-score (F1)
  - Net Reclassification Improvement (NRI)

Application

- We consider data supplied by the European External Credit Assessment Institution (ECAI) that specializes in credit scoring for P2P platforms focused on SME commercial lending
- The analysis relies on two separate data sets, that we refer to as DATA-SET A and DATA-SET B, composed of official financial information (balance-sheet and income statement variables) on 727 and 15045 SMEs respectively
- In DATA-SET A the proportion of defaulted companies is 6.01% whereas in DATA-SET B 10.85%

Results

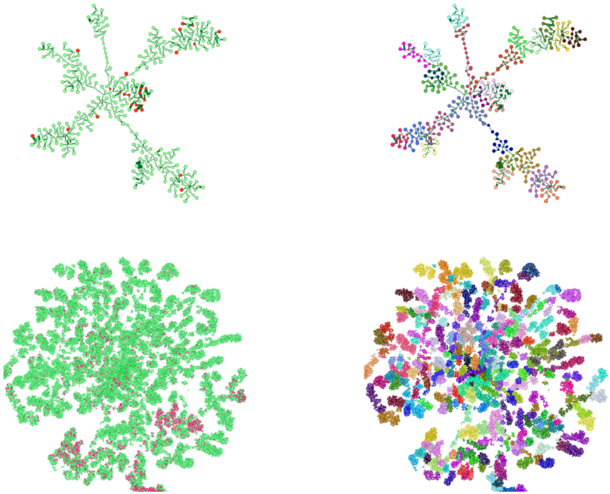

Figure 1: **Minimal spanning tree representation of the borrowing companies networks**

Results

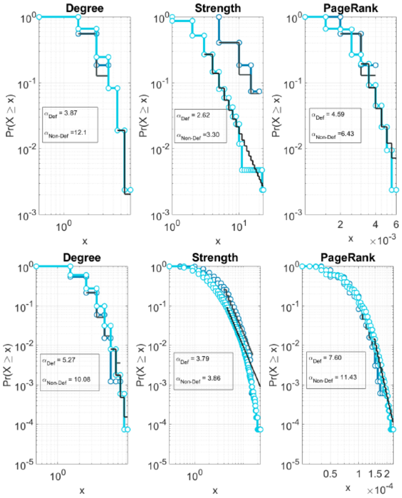

Figure 2: **Centrality measures distributions**

Results

DATA-SET A	AUC		AUPR		ACC		F1		NRI	
	Base	Network	Base	Network	Base	Network	Base	Network	Nri	P-val
Logistic Regression	0.7252	0.8021	0.1827	0.2653	0.9376	0.9510	0.9687	0.9688	0.5805	(0.09)
Discriminant Analysis	0.7197	0.7197	0.2590	0.2766	0.9443	0.9376	0.9684	0.9684	-0.0579	(0.54)
Naive Bayes	0.7404	0.7447	0.2358	0.2472	0.8730	0.8775	0.9655	0.9656	-0.039	(0.10)
Support VM	0.6014	0.7160	0.1361	0.1556	0.9398	0.9420	0.9689	0.9687	0.202	(0.55)
Decision Trees	0.7160	0.7178	0.2340	0.2416	0.9354	0.9376	0.9678	0.9679	-0.039	(0.20)

DATA-SET B	AUC		AUPR		ACC		F1		NRI	
	Base	Network	Base	Network	Base	Network	Base	Network	Nri	P-val
Logistic Regression	0.8155	0.8229	0.3434	0.3418	0.9034	0.9036	0.94208	0.94196	0.004	(0.79)
Discriminant Analysis	0.8011	0.8126	0.2942	0.3038	0.8888	0.8931	0.9400	0.9401	0.011	(0.08)
Naive Bayes	0.8064	0.8090	0.3097	0.3038	0.8029	0.8063	0.9310	0.9313	-0.015	(0.00)
Support VM	0.6997	0.7543	0.1615	0.2470	0.8932	0.8965	0.9424	0.9422	0.353	(0.00)
Decision Trees	0.7124	0.7097	0.1924	0.1899	0.8742	0.8705	0.9387	0.9384	-0.026	(0.00)

Figure 3: **Summary Statistics of non-parametric analysis**

Acknowledgment

This research has received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). In addition, the Authors thank ModeFinance, a European ECAI, for the data; the partners of the FIN-TECH European project, for useful comments and discussions.