Network models in FINancial TECHnology

Paolo Giudici, Paola Cerchiello, Alessandro Spelta, Giancarlo Nicola, Branka Hadji-Misheva, Iman Hashish, Paritosh Jha, Paolo Pagnottoni, Gloria Polinesi, Kamonchai Rujirarangsan, Anca Toma

Data Science Laboratory
Department of Economics and Management, University of Pavia
giudici@unipv.it; Twitter: @stateconomist; Facebook: Data Science Network

Background

- FinTech denotes companies that combine financial services with (on-line) innovative technologies.
- Advances in big data and big data modelling (data science) have enabled Fintechs to provide **competitive** (e.g. faster and cheaper) banking and finance services: creditech, paytech, wealthtech, insurtech, regtech,...
- More competitive services may bring, however, higher risks to the consumers: e.g. **cyber risk** (the risk of financial losses due to operational failures in the fintech IT systems) and **scoring risk**: the risk that consumers' choices may be misguided by wrong information, particularly in terms of credit rating. Both are amplified by **systemic risk**, due to the high interconnectdness of Fintech companies.

Contribution

- The **Data Science laboratory** since 2001, currently with 2 Faculty, 1 post-doc and 8 Phd students, carries out research and training in collaboration with Banks, Fintechs, and Regulators.
- Our current aim is to improve data science methodologies and, specifically, **the modelling of information coming from correlation, social and transactional networks**, to evaluate and improve scoring and rating of both banks and fintech platforms.
- Examples: credit scoring models in P2P lending; sentiment analysis to predict financial distress; matching risk profiles in roboadvisory; price discovery in cryptocurrency markets; assessment and prioritisation of cyber risks.

P2P Lending: business model

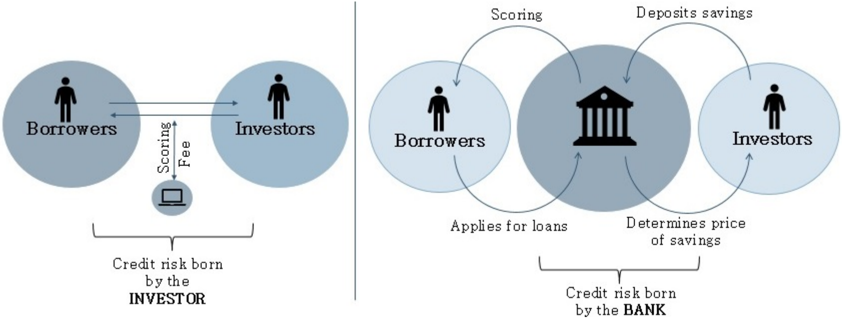

Figure: The business models of a P2P lender (left) and of a bank (right)

P2P lending: contribution

- Build a framework to test the predictive performance of scoring models employed by P2P platforms.
- To investigate whether network information can improve predictive performance, to further protect lenders, in a financial stability context.

Data

- So far we have collected data from: Lending Club, Mode Finance and N26 (in progress).
- Different statistical and machine learning methods have been applied to the data, with the aim of finding the best predictive model (in terms of default predictive accuracy).
- Here we refer to Mode Finance data, made up of SMEs which have applied for a loan via a P2P lending platform. The proportion of defaults in the sample is equal to 23%.

Scoring Models

- The most widely used model for building and estimating the probability of default is the logistic regression

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij},$$

from which:

$$p_i = \frac{1}{1+e^{\alpha+\sum_j \beta_j x_{ij}}}$$

Correlation networks

- Companies are related by their past financial behaviour. These relationships can be embedded in a correlation network.
- If each company is a node in the network and we associate different time series with different nodes of the network, each pair of nodes can be connected by an edge with a weight equal to the correlation coefficient:

$$w_{xy} = \frac{T(\sum_t x_t y_t) - (\sum_t x_t)(\sum_t y_t)}{\sqrt{[T\sum_t x_t^2 - (\sum_t x_t)^2][T\sum_t y_t - (\sum_t x_t)^2)]}}$$

Centrality measures

- **Degree centrality.** Degree centrality refers to the number of ties (edges) a node has to other nodes:

$$d_x = \sum_{y \neq x} e_{xy}.$$

- **Closeness centrality.** Closeness is defined as an inverse function of the distance of one node to all others:

$$c_x = \frac{1}{\sum_{y \neq x} d_{xy}},$$

Default-conditioned centrality measures

- **Default-conditioned degree centrality.**

$$d_x = \sum_{y \neq x} (e_{xy} | s(y) = 1)$$

- **Default-conditioned closeness centrality.**

$$c_x = \frac{1}{\sum_{y \neq x} (d_{xy} | s(y) = 1)}$$

Network-based scoring models

- We propose to extend scoring models including network centrality components:

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij} + \sum_i \gamma g_i$$

- from which:

$$p_i = \frac{1}{1+e^{\alpha+\sum_j \beta_j x_{ij}+\gamma g_i}}$$

Results: the network based on the activity indicator

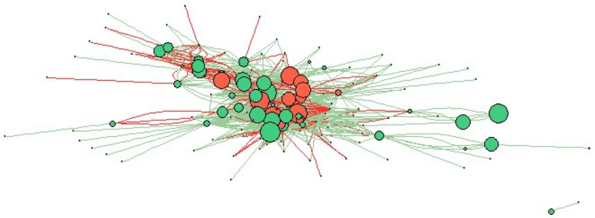

Figure: Correlation network based on the activity indicator

The network based on the solvency indicator

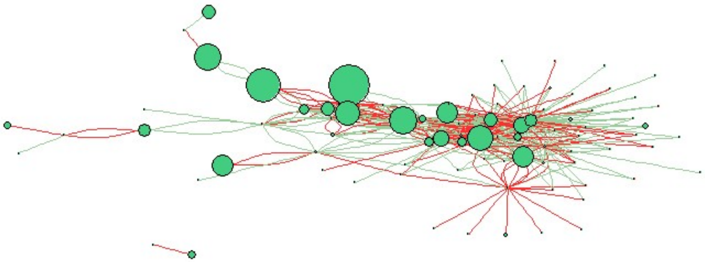

Figure: Correlation network based on the solvency indicator

Predictions: Standard Model

Variable	Estimate	P-value	Significance
Intercept	-3.39	0.011	*
Solvency ratio	0.01	0.539	
Debt to equity ratio	-0.07	0.517	
Current ratio	0.21	0.032	*
Cash over total assets	-2.51	0.579	
Return on equity	-0.08	0.008	**
Return on assets	0.01	0.963	
Return on Capital Employed	0.09	0.044	*
Coverage	-0.01	0.875	
Activity ratio	-1.92	0.001	***
Predictive accuracy (AUROC)			**0.71**

Table: The estimated baseline regression model

Predictions: Network based model

Variable	Estimate	P-value	Significance
Intercept	-1.53	0.033	*
Solvency ratio	-0.02	0.012	*
Debt to equity ratio	-0.00	0.576	
Current ratio	0.24	0.072	*
Cash over total assets	1.08	0.443	
Return on equity	-0.11	0.000	***
Return on assets	0.02	0.876	
Return on capital employed	0.01	0.212	
Coverage	0.02	0.248	
Activity ratio			
Degree Centrality	0.01	0.026	*
Closeness	1.05	0.002	**
Predictive accuracy (AUROC)			0.82

Main findings

- Network models can improve default predictions and, therefore, better protect P2P lending users. In addition, they provide useful information that can be used to monitor companies that may trigger and spread contagion.
- We expect that further network information (e.g. s transactional networks) can further improve model performance. This is research in progress.
- Similar results are emerging in sentiment based financial stress prediction, roboadvisory risk profile matching, crypto price discovery and in the prioritisation of cyber risks.

Policy suggestions

- Data Science methods can be very helpful to protect investors and to preserve financial stability, encouraging the diffusion of Fintech innovations.
- This requires investing in financial data science education, that intersects Finance, Statistics and Computer Programming, and especially in "collaborative" education, through cooperation between Universities, the Financial industry, and Regulators.
- Such cooperation can be boosted by "light budget" initiatives, such as: a) "outsourcing" the activity of data science laboratories in Fintech districts, innovation hubs and sandboxes, through joint industrial Phd programmes; b) issuing "dedicated" research projects, that include funding of Phd and Post-doc positions; c) create a "pooled database" of network data that can augment existing databases, for the benefit of improved model predictions.

Some Recent papers

- (2017) Pejman Abedifar, Paolo Giudici, Shatha Hashem. Heterogeneous market structure and systemic risk: evidence from dual banking systems. **Journal of Financial Stability.**
- (2017) Paolo Giudici, Peter Sarlin, Alessandro Spelta. The interconnected nature of financial systems: direct and common exposures. **Journal of Banking and Finance.**
- (2017) Paolo Giudici, Laura Parisi. Sovereign risk in the Euro area: a multivariate stochastic process approach. **Quantitative Finance.**
- (2017) Raffaella Calabrese, Johan Elkink, Paolo Giudici. Measuring bank contagion using binary spatial regression models. **Journal of the Operational Research Society.**
- (2017) Paola Cerchiello, Paolo Giudici, Giancarlo Nicola. Twitter data models for bank risk contagion. To appear in **Neurocomputing.**
- (2016) Paolo Giudici, Alessandro Spelta. Graphical network models for international financial flows. **Journal of Business and Economic Statistics**, 34 (1), pp. 126-138.
- (2017) Paolo Giudici, Laura Parisi. Correlation networks to measure the systemic implications of bank resolution. Submitted to Review of Finance
- (2017) Paolo Giudici, Laura Parisi. Corisk: measuring systemic risk through default probability contagion. Submitted to Quantitative Finance
- (2017) Stefan Avdjiev, Paolo Giudici, Alessandro Spelta. Measuring contagion risk in international banking. Bank for International Settlements research paper. Submitted to Journal of Financial Stability
- (2017) Paolo Giudici, Branka Hadji Misheva. Scoring models for P2P lending platforms: a network approach. Submitted to Journal of Banking and Finance