# A study of neighborhood influence over SME's creditworthiness with Graph Neural Networks

Lextrait, B. *

June, 2022

**Abstract**

We propose a first attempt at quantifying risk transmission between geographically close firms, as the economic environment is assumed to have an impact on the growth and survival rates of Small and Medium-sized Enterprises (SMEs). We assume that individual Probability of Default is driven by two risk factors: the first is linked to firm fundamentals, the second external and transmitted by neighbor-to-neighbor contagion. We test this hypothesis on the whole French SME population. We first apply modern machine learning techniques to estimate the internal risk of each firm. We then propose to use graphs and Message Passing Neural Networks (MPNN) to simulate risk contagion between neighboring firms. We apply explanatory methods to extract information from the unique structure of MPNN. We reveal statistical evidence for several candidate phenomena discussed in the literature: agglomeration economies, detrimental local competition, impact of multi-company leadership. We confirm that risk transmission is facilitated between companies sharing common characteristics such as size, sector and representatives.

*Keywords*: SMEs, Credit scoring, Risk transmission, Company clusters, Graph Neural Network, GNN Explainer.

*lextrait.bastien@parisnanterre.fr, Economix, University Paris Nanterre

# 1 Introduction

The prediction of corporate failure events has been the focus of the literature for decades [Cochran, 1981 ; Dimitras, Zanakis & Zopoudinis, 1996]. Although little of the research effort was focused on SMEs, a consensus slowly emerged on the internal and global factors that could lead to a default event. Age, size and ownership type are amongst the most commonly cited internal factors related to growth and failure rate [Dunne, Roberts & Samuelson, 1989 ; Cefis & Marsili, 2005]. Traditional financial ratios of leverage, profitability and liquidity also have strong predictive power over incoming bankruptcies [Pompe & Bilderbeek, 2005 ; Altman & Sabato, 2007]. On more global scales, unemployment rate and retail sales are important indicators of the regional firms' financial health [Everett & Watson, 1998]. However, the existence and the importance of local factors affecting SMEs' survival rate is a quite recent and still ongoing debate.

The matter emerged as a consequence of the broad consensus reached on macroeconomics and global policies to support economic progress, which was progressively deemed insufficient at picturing the complexity of the real economy [Ketels , 2003]. Research efforts then reoriented towards microeconomics and the emerging cluster theory, iterating on the works of early $XX^{th}$ century [Krugman, 1991]. Porter [1998] laid the foundation of this new field, defining clusters as "geographical groups of interconnected companies and associated institutions in a particular field, linked by similarities and complementarities". Defenders of the agglomeration economies principle stated that clustering essentially boosts innovation and productivity by reducing transaction costs, increasing information flows and adapting the business environment to the needs of the cluster members, which in turn stimulates the creation of new businesses [Baptista & Swann, 1998 ; Porter, 2003].

Pollard [2003] was one of the first to conjecture that clusters play a non-negligible role in the financial prosperity of their constituents. He stressed that thorough understanding of the underlying process would require the collaboration of several research fields, namely geographics, econometrics, and statistics. He was joined on his proposal by McCann & Folta [2008], who proposed a research agenda on the topic. They adviced to specifically focus on clusters' creation and temporal dynamics, their attractiveness on isolated firms, their impact on survival rates and the relationships between their constituents, the last two point being adressed in this paper. First studies were conducted on the special case of high-tech company clusters, essentially relying on the concept of knowledge spillover as regional-scale vector of development [Gilbert & al., 2008 ; Depret & Hamdouch, 2009]. Although findings aknowledged the existence of such spillovers, concerns were expressed regarding their ability to significantly influence the financial growth of companies. It was recommended to extend the scope of analysis beyond the sole Research and

Development sector.

The main goals of this paper are threefold. We first propose a new and simple mathematical model able to illustrate fundamental and local risk factors. This model is designed to adapt to the considerations of internal risk literature based on access to finance, growth and bankruptcy, and external risk literature based on cluster theory, competition and local attractiveness. We then develop a practical methodology to test the model at large scales. We show the relevance of Message Passing Neural Networks (MPNN) as a modern graph-based machine learning algorithm able to grasp the phenomena at stake. We finally apply explanative tools to the fitted MPNN and provide statistical evidence of several phenomenon debated in the literature, which are detailed in the following paragraphs. We also introduce and discuss the notion of "influence" as the propention of certain firms to convey risk to their neighbors.

The current state of research is divided in two unconciliable branches, none of which providing enough evidence to build consensus [Frenken, Cefis & Stam, 2015]. The first one supports the agglomeration economy hypothesis, according to which spatial clustering of companies facilitates the emergence of asymetrical benefits, which positively impacts their financial health and henceforth their survival rates [Delgado et al., 2010 ; Wennberg & Lindqvist, 2010 ; McCann & Folta, 2011]. The second one - competitive economies - claims that spatial clustering has mixed effects on firm productivity. These effects might depend on the stage of the cluster in its life cycle [Kukalis, 2010 ; Branco & Lopes, 2018 ; Pavelkova et al., 2021], be aggravated by excessive interconnectedness [Molina-Morales, 2012], and even turn fully detrimental [Boschla & Wenting, 2007]. The latest works attempt to blend in the notions of localisation, human capital and entreprepreneurship strategies to explain the micro-spatial effects upon firm survival. All those research areas are considered of equal importance so far [Huggins, Prokop & Thomson, 2017].

Company representatives also constitute an other speculated factor of influence over their survival rate. Very few studies ever attempted to produce evidence on the subject as related data is very scarce and hard to collect. Those who did however highlighted a potentially impactful leadership effect, the consequences of which could be beneficial as well as detrimental, depending on the representative's profile and relationships with peers [Ceci & Lubatti, 2012 ; Khelil, 2016].

In practice, we assume that risk contagion between companies is an asymetrical process. We model firm interactions within a graph embedding. A baseline predictive model then estimates the Probability of Default (PD) of each firm from internal factors, including - but not limited to - its descriptive information and latest balance sheets. To that end we use a Light Gradient

Boosting Machine (LGBM) from the Gradient Boosting Decision Trees (GBDT) family. The methodology used and the characteristics of this model are reported in a previous study [Lextrait, 2022]. The estimated PD and the graph embedding are both submitted to a Graph Neural Network (GNN). GNNs are part of the Message Passing Neural Network (MPNN) algorithm family, mainly used in chemical molecular analysis [Gilmer et al., 2017 ; Mater & Coote, 2019 ; C. St. John et al., 2019], but its applications seem versatile enough to be applied in our context. Its usage here is twofold. First, it can be used as a prediction corrector, taking into account cluster information to refine and hopefully outperform the base estimations. Second, once trained, it can be submitted to explanative methods to reveal the specific features which play an important role in risk contagion. The final step of our methodology therefore consists of reviewing several methods to achieve GNN interpretability. Among those, we apply the new model-agnostic GNN Explainer recently proposed by Ying & al. [2019]. All experiments are led on a network of approximately 600,000 French firms, over a period covering the financial years 2016 to 2018.

The results confirm GNN's abilities to simulate local risk contagion effects. Similarities between neighboring firms turn out to be decisive in transmitting the risk, in line with general insights of cluster theory and spatial econometrics. More specifically, sharing an activity sector with neighboring firms increases the risk, while sharing the same size decreases it. Geography aside, sharing common representatives with other companies also raises risk. The intensity of risk contagion through those three similarities is doubled in average when neighboring firms recently experienced a default. Results also indicate that sharing the same age - in decades - with neighboring firms conveys some risk, however this result does not hold when the neighbors experience default. No conclusion can be drawn over similarities of revenue magnitude. Finally, a first practical application of GNN is proposed to determine which neighbors of a specific company are considered the most influential providers of risk - or lack thereof.

This paper proceeds as follows. Section 2 provides a literature review on the subject of externalities and their influence over firm growth and survival rates. Section 3 describes the chosen embedding framework for the data we use in this study, the GNN algorithm to which it is submitted, as well as the assumptions supporting the use of these tools. Section 4 describes the data used in this study. In section 5 the performances of all candidate GNN architectures are compared to each other and the LGBM baseline. Interpretability methods are applied to the best performing ones to better understand the risk contagion phenomenon. Section 6 provides our conclusions.

# 2 Review of literature : the role of neighboring in credit risk modeling

The following definitions apply to the rest of the paper. The notion of 'proximity' applies to companies that are geographically close, or that share a common representative. Companies sharing a proximity relationship are designated as 'neighbors'. The notion of 'similarity' further describes the likeness of neighbor firms in terms of age, size, activity sector or revenue.

In the literature, the very first external factor assumed to impact the PD of a company is its lender, through which the event of failure occurs. In the specific population of SMEs that role generally falls to the banks, although the recent development of the crowdlending market in the last decade challenges that traditional organization [Ziegler et al., 2018]. A systemic theory of the traditional lending landscape established the existence of an anti-selection process, emerging from information asymetry between borrowers and lenders [Berger & Udell, 2002 ; Pollard, 2003 ; Beck & Demirguc-Kunt, 2006]. In addition to this now well-understood mecanism, recent findings also highlighted a spatial discrimination phenomenon, according to which distance to lender impacts loan rates as well as loan defaults. Degryse and Ongena [2005] were the first to reach that conclusion on a sample of 15,000 belgian bank loans established between 1975 and 1997. By applying ordinary least squares estimation on loan rates, with distance to lender and closest competitor banks as variables of interest, they found that loan rates decrease with the lender-firm distance and increase with lender-competing banks distance. The explanation lies in the ease of soft information gathering at close proximity of the lender, which gives him informational advantage over its distant competitors [Agarval & Hauswald, 2010]. More precisely, it has been suggested that small local banks would oppose less credit constraints to SMEs than local branches of bigger centralized banks with distant headquarters [Zhao & Jones-Evans, 2017]. DeYoung, Glennon & Nigro [2008] complemented those findings by applying a discrete-time hazard model on 30,000 US loans in the same time period, and found that loan probability of default also rise with distance to lender.

Networks of mutually interdependant counterparties is the generalization of the previous lender-borrower problem to multiple agents. Such a system allows to consider macroeconomical phenomenons, such as partner-choice dynamics or resiliency to financial shocks. Using Monte-Carlo simulations, Neu & Kühn [2004] showed that systems of mutually supportive firms are inherently generative of high counterparty risks, which could lead to avalanches of defaults. Such systemic risks are dramatically undermined by standard credit risk models, for which the borrower is the only source of data and not considered as member of an interdependant network. To limit that risk in the context of loan portfolio management, the authors suggested to include

mutually supportive as well as mutually competitive firms in the diversification process. Later findings suggests that network topology plays an important role in its endogenous risk diffusion. While a moderate amount of connectivity provides a stabilizing effect though risk diversification, it may act as a shock amplifier at too important levels [Battiston & al., 2012]. This effect is all the more serious as financial networks are not fixed structures and may naturally evolve to a state of heavy-interconnectedness through preferred-partner choice rule [Gatti & al., 2009]. The dynamics between the evolving topology of networks, their resiliency to initial shocks and so-called default cascading behavior currently constitues a whole research area.

The previous findings are not directly applicable when considering the problem of risk transmission within the SME population, as most of its elements are not intricated in financial networks of mutually-dependent agents. The literature on the subject rather focused on geographical clustering, assuming distance as the best vector of risk transmission between companies that would otherwise be unrelated to each other. Boschma & Wenting [2007] were among the first to propose a systemic approach bringing together evolutionary economics and economic geography. They studied the case of the British automotive industry between 1895 and 1968, using Cox regression over a sample of several hundreds manufacturers, 89% of which were located in the Coventry-Birmingham area, constituting *de facto* a cluster. Their findings highlighted a significant negative impact of spatial company distribution on the survival rate of the new entrants : traditional actors present from the very beginning of the sector development benefitted higher survival odds, while no spillover effects seemed to favorise young contenders. However, another study conducted over a sample of 4,400 Swedish firms from 1993 to 2002 led to the complete opposite conclusion [Wennberg & Lindqvist, 2010]. Using pooled time-series regression, the authors found evidence that a high concentration of employment in same or related industry sectors is correlated with better chances of survival, as firms benefit from agglomeration economies. This conclusion had also previously been reached in other similar studies, mostly in high technology, manufacturing and communication sectors [Beaudry & Swann, 2001 ; Dumais, Ellison & Glaeser, 2002 ; Rosenthal & Strange, 2005 ; Pe'er & Vertinsky, 2006].

More recently, Weterings & Marsili [2015] provided a more nuanced conclusion after analysing exit rates over a 24,000 firms sample in the Netherlands. Using multivariate logit regression, they found that whether positive agglomeration economies or negative competition effects prevails depends on the industry being studied. De Silva & McComb [2012] refined the Cox proportional hazard model application to a mile-radius precision, over a software and high-tech firms sample located in Texas. Reconciling the observations from previous literature, their findings suggested that there might actually be a shift of the influence sign depending on the considered inter-firms distance : very close firm concentration at a 1-mile radius or less tends to increase competition

and lower survival rates, while spillover effects favorising survival rates tend to appear in a 1 to 25 miles radius.

Cluster theory aside, some methods also provided interesting results in the alternative field of spatial filter approaches. Barreto & Artes [2013] applied kriging to approximate spatial credit risk and used the results as a logistic regression's explanatory variable, significantly increasing its performance. Later, Mate-Sanchez-Val, Lopez-Hernandez & Fuentes [2018] searched for spatial co-localized patterns of surviving and defaulted madrilenian companies through Join-Count statistics. Interestingly, they used k-nearest-neighbors to determine neighborhood relationship between close firms, and found significative evidence of co-localized failed companies for $k$ varying from 4 to 12. In the same paper, using spatial probit models, they also studied the existence of a potential economic "buffer zone" around different urban facilities, and found that logistic centers and industrial zones reduce the exposure to failure of nearby firms. Calabrese, Andreeva & Ansell [2019] introduced a refined methodology on the spatial probit model, arguing that the traditional euclidean distance supporting the adjacency matrix may be too weak of an indicator. They proposed instead to use similarity measures between companies, such as Jaccard's or Gower's, allowing to take their formal characteristic descriptors into consideration, such as legal form, sector code or staffsize categorization. The method we introduce in this article further explores the similarity-based approach. Finally Agosto, Giudici and Leach [2019] introduced their own spatial regression model, and proposed to substitute a firm trade flow matrix to the traditional firm adjacency matrix, based on the observed trade flows between activity sectors. Applied on a sample of 1180 Italian SMEs, this methodology revealed evidence of a spatial autocorrelation, interpreted as a contagion vector of default.

# 3 Methodology

## 3.1 Baseline default risk modeling without neighboring influence

### 3.1.1 Model specification

Let $\mathbf{X} = x_1, ..., x_N \in \mathbb{R}^{N \times P}$ be the observation sample of our study. For the rest of the paper we will use $v, w \in [1, N]$ as observation indexes and $p \in [1, P]$ as feature index. The 12-month business failure risk we are trying to estimate can be represented by a latent continuous variable $\hat{Y}^{\iota} \in \mathbb{R}^N$ in our machine learning models, where the true response vector $Y = Y_1, ..., Y_N$ is such that $Y_v$ equals 1 if the $v^{th}$ balance sheet from our sample is associated with a bankruptcy event and 0 elseway. A decision threshold $t$ links the two variables such that:

$$Y_v = \begin{cases} 1 \text{ if } \hat{Y}_v^{\iota} > t \\ 0 \text{ else} \end{cases} \quad \forall v \in [|1; N|] \tag{1}$$

Our objective is to train the best function $\hat{H}^{\iota} \in \mathbb{R}^P \to [0,1]$ that minimises the expected loss $L(.,.)$ between its predictions and the true binary answer, as follows :

$$\hat{H}^{\iota} \approx \underset{H \in \mathbb{R}^P \to [0,1]}{\operatorname{argmin}} \mathbb{E}_{Y,\mathbf{X}} L(Y, H(\mathbf{X})) \tag{2}$$

### 3.1.2 LightGBM

The literature already provides a vast array of statistical and algorithmic methods to generate satisfying approximate solutions to (2). We choose in this paper to apply Light Gradient Boosting Machine (LightGBM), which is a specific algorithmic implementation of the Gradient Boosting Decision Tree (GBDT) family. We motivate our choice by the robust and highly performing results that LightGBM provides when applied to our problematic [Lextrait, 2021].

Gradient Boosting Decision Trees (GBDT) is an ensemble model, whose objective is to train a great number of simple base learners, then aggregating their predictions into a single output. Here, base learners are binary decision trees $(h_k)_{k \in 1,\dots,K} \in \mathbb{R}^P \to \mathbb{R}$ trained by boosting : the residual errors $Y - H^{t-1}(\mathbf{X})$ from learning step $t-1$, $(t > 0)$ are considered during learning step $t$, for $t \in [|0; t_{max}|]$. Depending on the implementation, the global aggregating function $\Phi \in \mathbb{R}^K \to [0,1]$ can usually be a summation, an averaging, or a majority voting operation :

$$H^t(x) = \underset{k \in 1,\dots,K}{\Phi} h_k^t(x) \tag{3}$$

The objective of each binary decision tree $h_k$ is to generate a partition of the feature space $\mathbb{R}^P$ through a series of binary tests. Each test splits the feature space in two depending on whether a feature $m$ is superior to a splitting attibute $\theta_m$. Each final partition element $R^j$ is attributed a response value $c_j$. A single tree can be expressed as :

$$h_k(x) = \sum_{j=1}^{J} c_j . \mathbb{1}_{x \in R_j} \tag{4}$$

LightGBM was introduced by Ke et al. [2017] as a fast and computationally light implementation of GBDT relying on two optimisation techniques called Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS favors the use of the most contributive fraction of the learning sample when recursively building the trees. EFB bundles together features that never simultaneously take nonzero values, thus specifically adressing numerical issues emerging in high-dimensional space, such as data sparcity.

## 3.2 Default risk with neighboring influence

This section's objective is to formalize the key notion of neighboring influence, which we consider to be the missing link between intrinsic risk and cluster effects over firm survival. Through it, risk or absence of risk can be spread from all companies to their neighbors, and the intensity of that transmission depends on the shared similarities between the neighbors. By visual analogy, if we think of risk as an electrical impulse within a brain cell, then neighbor influence would be the axon transferring part of that impulse to related neurons.

### 3.2.1 Model specification

From the recent findings in literature described in section 2), we emit a series of working hypothesis :

- The general latent Probability of Default $\hat{Y}$ can be seen as the addition of the intrinsic risk $\hat{Y}^{\iota}$ inherent to the company, an external risk $\hat{Y}^{\nu}$ imported from microeconomical surroundings, and a global risk $\hat{Y}^{\eta}$ resulting from macroeconomical circumstances, such that $\hat{Y} = \hat{Y}^{\iota} + \hat{Y}^{\nu} + \hat{Y}^{\eta}$.
- We assume that $\hat{Y}^{\eta}$ can be marginalized as a constant bias $\eta$ in our study. This decision is purely technical as our sample only covers three years of data, which is too short to reflect the evolving circumstances within a global economic cycle.
- We assume that the external microeconomical risk $\hat{Y}^{\nu}$ emerges from a combination of the intrinsic risk of all surrounding firms. This hypothesis relies on the existence of a risk transfer phenomenon, which we denote $f$, responsible for bringing risk or absence of risk from any company to its neighbors.

Under these assumptions, the latent Probability of Default of any specific company $v$ can be expressed as follows:

$$\hat{Y}_v = \hat{Y}_v^{\iota} + f\left(\{(\hat{Y}_w^{\iota}, e_{v,w}) | w \in \Gamma_v\}\right) + \eta \tag{5}$$

where $\Gamma_v$ is the company neighborhood of $v$ and $e_{v,w}$ describes the similarities between $v$ and any of its neighbor $w$. Thus, $f$ extracts the external risk $\hat{Y}_v^{\nu}$ of $v$ from its interactions with its neighborhood. The function $f$ and $e_{v,w}$ vectors will be more precisely described in the following sections.

Our methodology consists in two iterative steps. We first train an intrinsic risk predictor $\hat{H}^{\iota}$ as a prime candidate for (2), using $\mathbf{X}$, as described in section 3.1. Its predictions can be considered as an estimation of the intrinsic risk $\hat{Y}^{\iota}$. We then train a second predictor $\hat{H}$ built upon the knowledge of $\hat{H}^{\iota}$, which takes into account neighborhood structure and similarities. Its

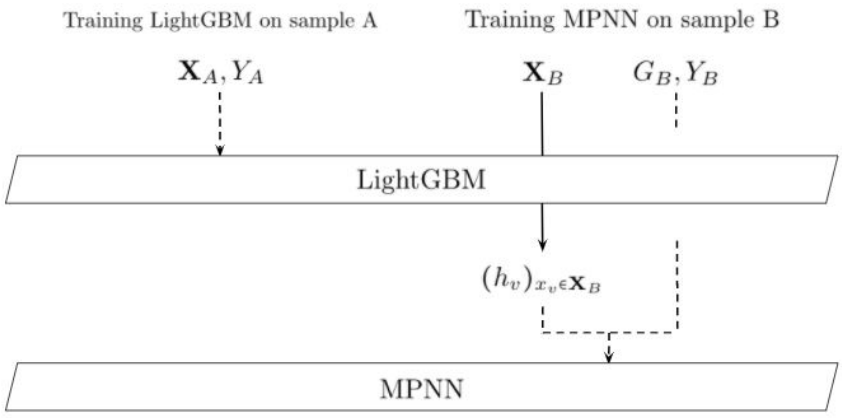
role is to use this new information to refine $\hat{Y}^{\iota}$ into a better estimation $\hat{Y}$ of the global risk, as defined in (5). The chosen learning algorithm is a Message Passing Neural Network (MPNN) specifically designed to learn over graph-structured data. If we denote $G$ the new information - which is described in section 3.2.2 - it follows :

$$\begin{aligned} \hat{H} &\approx \underset{H\in[0,1]\to[0,1]}{\operatorname{argmin}} \mathbb{E}_{Y,G} L\left(Y, H(\hat{Y}^{\iota})\right) \\ &\approx \underset{H\in[0,1]\to[0,1]}{\operatorname{argmin}} \mathbb{E}_{Y,G} L(Y, (H \circ \hat{H}^{\iota})(\mathbf{X})) \end{aligned} \tag{6}$$

with $\circ$ the function composition operator.

The ability to approximate $Y$ in (6) heavily depends on the algorithm chosen to solve (2), and a poor estimation of the intrinsic risk will only lead to a poor understanding of the risk transfer phenomenon. Figure 1 summarises the whole training process. Another practical issue we have to consider is the possibility that the MPNN fails to extract significant information value from the graph embedding of our data. In this case it will simply return the identity function $\hat{H} = I : \mathbb{R} \to \mathbb{R}$. We will have to carry out a statistical analysis to check that we do not meet this scenario.

Figure 1: Iterative training of LightGBM and MPNN




Dotted arrows indicate learning phases of the process, solid arrows indicate scoring phases. The first step consists in training the LightGBM on a sample $A$. The learner gains knowledge about the intrinsic risk of default $\hat{Y}^{\iota}$ through association of internal firm data $\mathbf{X}_A$ and the observed event $Y_A$. Then, the MPNN is trained on a second sample $B$ similar to $A$, enriched through embedding the data in a relational graph structure $G_B$ detailing neighboring information between all the observations. The output $(h_v)_{x_v \in \mathbf{X}_B}$ of the LightGBM is part of MPNN's input data. This way, the MPNN learns to refine the intrinsic risk prediction using neighbor influence. MPNN's output is an approximation of the final risk $\hat{Y}$ detailed in (5).

### 3.2.2 Neighborhood embedding structure

The second term of (5) requires that we embed our collection of observations - here, SMEs - within a mathematical graph $G(\Gamma, V)$ able to represent individual as well as relational features. $\Gamma = \{\hat{y}_1^\iota, ..., \hat{y}_N^\iota\} \in \mathbb{R}^{N \times \Pi}$ is the node information matrix, storing the intrinsic predictions of $\hat{H}^\iota$ as well as some additional information. $V = \{e_{v,w} \in \mathbb{R}^Q\}$ is the edge collection bearing the relational information for all pairs of connected nodes[1]. More specifically, each row $e_{v,w}$ of $V$ designates a pair of geographically neighboring nodes $(v, w)$ and contains a series of boolean values answering the question : *"What characteristics do nodes v and w have in common ?"*. The considered characteristics first include *ActivitySector* and *StaffSize* as they were used by Calabrese, Andreew & Ansell [2019] to build their similarity measure. We also include *Revenue* the order of magnitude of the company turnover, to qualify its size as a regional economical actor. *Age* also seems to play an important role within company clusters, as successive batches of new entrants may benefit from identical growth conditions during their lifetime [Boschma & Wenting, 2007]. Finally, we make the assumption that entrepreneur profiles may also be vectors of risk contagion, and add a characteristic *Representative* indicating whether companies are lead by common individuals. Indeed management styles, although being mostly absent from economic research topics, are thought to impact company growth towards success or failure [Khelil, 2016]. Table **1** summarizes the content of matrixes $\Gamma$ and $V$. Their construction process from the available data is described in section 4.2.

This structure allows us to extract for any specific node $v$ its neighborhood $\Gamma_v \subset \Gamma$, which is the ensemble of other nodes $\{w\}$ for which $e_{v,w}$ exists in $V$. Similarly, we can also consider $V_v \subset V$ the restriction of the edge collection to the sole information related to $v$.

### 3.2.3 Estimation of the external risk

According to our third working hypothesis defined in section 3.2.1, the external risk $\hat{Y}^\nu$ of any company $v$ is derived from the internal risk of all the elements of its neighborhood $\Gamma_v$ as well as the similarities it shares with its neighbors. The extraction and combination of this information is performed by the risk transfer function $f$, so that in (5) we have $\hat{Y}_v^\nu = f\left(\{(\hat{Y}_w^\iota, e_{v,w}) | w \in \Gamma_v\}\right)$. Our main goal is to determine the best candidate function for $f$.

We consider single-layer Message Passing Neural Networks (MPNN) as a candidate for that role. MPNNs are a subcategory of Neural Networks applied to graph-structured data. Their general form is the following :

[1] In this study, $\Pi = 3$ and $Q = 5$

Table 1: Feature embedding of $G(\Gamma, V)$

| Vector | Variable name | Variable support | Description |
|---|---|---|---|
| $\Gamma$ | BaseValue | $[0,1]$ | Intrinsic risk predictions from $\hat{H}^{\iota}$ |
| | BaseValueRange | $[0,1]$ | Extent of *BaseValue*'s uncertainty margin |
| | FailPrevious | $\{0;1\}$ | 1 if the company experienced a failure the year before the current studied financial closure, 0 if not |
| $V$ | SameAgeDecade | $\{0;1\}$ | 1 if both companies share the same *Decade*, 0 if not |
| | SameSector | $\{0;1\}$ | 1 if both companies share the same *SectorCode*, 0 if not |
| | SameStaffSize | $\{0;1\}$ | 1 if both companies share the same *StaffSize*, 0 if not |
| | SameRevenue | $\{0;1\}$ | 1 if both companies share the same *Revenue* order of magnitude, 0 if not |
| | SameRepresentative | $\{0;1\}$ | 1 if both companies share at least 1 common representative, 0 if not |

$$\begin{aligned} \hat{y}_v &= f(\gamma, \Psi, \Phi, \hat{y}_v^{\iota}, \Gamma_v, V_v) \\ &= \gamma(\hat{y}_v^{\iota}, \Psi_{\Phi}(\Gamma_v, V_v)) \\ &= \gamma\left(\hat{y}_v^{\iota}, \underset{w \in \Gamma_v}{\Psi} \Phi(\hat{y}_w^{\iota}, e_{v,w})\right) \end{aligned} \tag{7}$$

First, the Message function $\Phi$ extracts relevant information from each neighbor $w$ of $v$ according to shared similarities between both nodes, and returns a message in the form of a single scalar. $\Psi$ agregates the messages transmitted by the entire neighborhood. It needs to be a permutation-invariant operator of $\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ : the mean, sum or max operators are the most common choices. Finally, the Update function $\gamma$ refines the value of the input $\hat{y}_v^{\iota}$ with the agregated message transmitted from the neighborhood. During learning, we use the binary cross-entropy $L = -\sum_v y_v log(\hat{y}_v) + (1 - y_v) log(1 - \hat{y}_v)$ as loss function.

Literature provides us with a vast array of $(\Phi, \Psi, \gamma)$ functions to choose from. To guide our choice, we asked ourselves the following questions : *What are the most relevant Message functions* $\Phi$ *given our current problematic ? Which* $\Psi$ *operator must be chosen among the summation, averaging or maximum ?* and *Does the update function* $\gamma$ *have to learn an additional bias ?* We decided to try four specific implementations of message functions $\Phi$, namely GraphConv, EdgeConv, NNConv and RGCNConv, whose corresponding architectures are described in Appendix **A**. We also decided to try all three possibilities for $\Phi$ and both possibilities for $\gamma$, resulting in 24 possible $(\Phi, \Psi, \gamma)$ settings to experiment with. The training methodology is

described in Appendix **B**.

After completion of the learning phase, we dispose of 25 models for each setting (600 models in total), ranked from best to worst on the validation set. For the rest of the paper, each trained model will be denoted as $H^i_{\Phi,\Psi,\gamma}$ with $i$ its rank. Filtering those models to keep only the most relevant ones is a necessity. The procedure followed to achieve this objective is detailed in Section 5.1.2.

# 4 Data

## 4.1 Balance sheets and labels

Our raw data sample consists of the 1,685,493 French companies' balance sheets nonconfidential recordings, corresponding to the closing of the financial years 2016 to 2018 and available at the INPI opendata service[2]. Those balance sheets are separated in two categories : the Complete report, which is the most detailed form of report as it includes 428 financial items, and the Simple report which is a more condensed form only including 190 financial items.

The classification labels are deduced from the public records of the SME's commercial court rulings given between 2016/01/01 and 2019/12/31, also available at the INPI opendata service. For each balance sheet, we associate a positive event if we find any court-ordered liquidation or recovery procedure within the 12 first months following the financial year-end date, and a negative event otherwise. Table **2** indicates the repartition of our sample within the three studied financial years and report categories, along with the population's fail rate in each subcategory. The 2016 data serves as training set for the LightGBM, the 2017 data serves as training set for the MPNN, the 2018 data serves as test set.

Table 2: Sample repartition over financial years and report categories

| Type of report | Year | Dataset size | Fail rate |
|---|---|---|---|
| Simple | 2016 | 208 348 | 0.31% |
| | 2017 | 169 824 | 0.39% |
| | 2018 | 117 760 | 0.37% |
| Complete | 2016 | 408 668 | 0.36% |
| | 2017 | 415 520 | 0.34% |
| | 2018 | 365 373 | 0.28% |

[2]See https://data.inpi.fr/content/editorial/Serveur_ftp_entreprises.

## 4.2 Relational data

Features described in Table **3** summarize essential information about each company in order to establish the graph embedding described in section 3.2.2. Figure 2 describes the distribution of numerical features. The company and leadership descriptive information is available at the INPI opendata service, while the precise geographical coordinates of each company are retrieved from the National Address Database (BAN)[3] opendata service.

Table 3: Description of features used in graph-embedding process

| Feature Name | Type | Feature description |
|---|---|---|
| **Siren** | Identifier | Company ID |
| Revenue | Categorical | Order of magnitude of the annual revenue |
| Decade | Categorical | Age of the company, rounded to the lowest decade |
| RepID | Identifier | List of company representatives, each assigned an ID |
| ReprAppDate | Date | List of company representatives' first date of appointment, respectively associated with each RepID |
| **Siret** | Identifier | Physical business unit ID. Several Siret can be attached to the same Siren |
| Lon | Coordinate | Business unit longitude, WGS 84 |
| Lat | Coordinate | Business unit latitude, WGS 84 |
| SectorCode | Categorical | French Activity Nomenclature (NAF) code categorizing the different business activity sectors |
| StaffSize | Categorical | Workforce size categorization among 15 size layers from 0 to more than 10 000 employees |

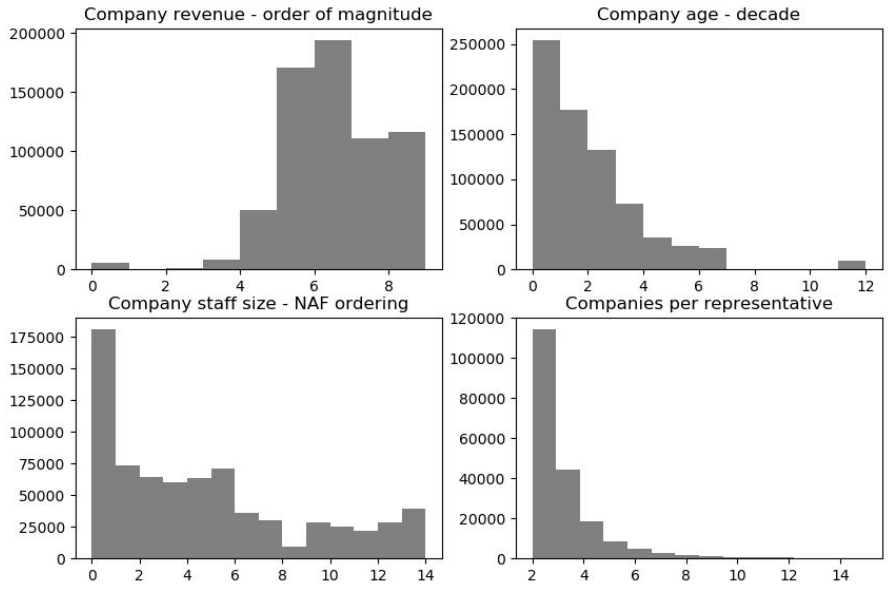
We consider two rules to generate the graph's edge collection $V$ from available data. The first rule is based on geographical k Nearest Neighbors (k-NN) computed from the *Lat* and *Lon* coordinates. With $k$ as integer parameter, all euclidean k-NN of a given company $v$ are considered to be neighbors of $v$. As companies may possess several physical business units, the aforementioned rule is applied to each of those physical business units. If we denote $Bu_v$ the ensemble of physical business units of a company $v$, this rule can be expressed as follows :

$$\exists b_v, b_w \in Bu_v \times Bu_w : b_w \in k_{NN}(b_v) \Rightarrow e_{v,w} \in V \tag{8}$$

The second rule consists in considering as neighbors all companies that share a representative. If we denote $Rep_v$ the ensemble of representative IDs of a company $v$, this rule can be expressed as follows :

[3]See https://adresse.data.gouv.fr/donnees-nationales.

Figure 2: Distribution of categorical features among the study sampled observations. Representatives of a single company are not taken into account as they do not generate neighbor-related information




$$Rep_v \cap Rep_w \neq \emptyset \Rightarrow e_{v,w} \in V \tag{9}$$

Note that because of the non-symmetrical nature of (8), $e_{v,w} \in V \Leftrightarrow e_{w,v} \in V$ is not necessarily true.

# 5 Results and discussion

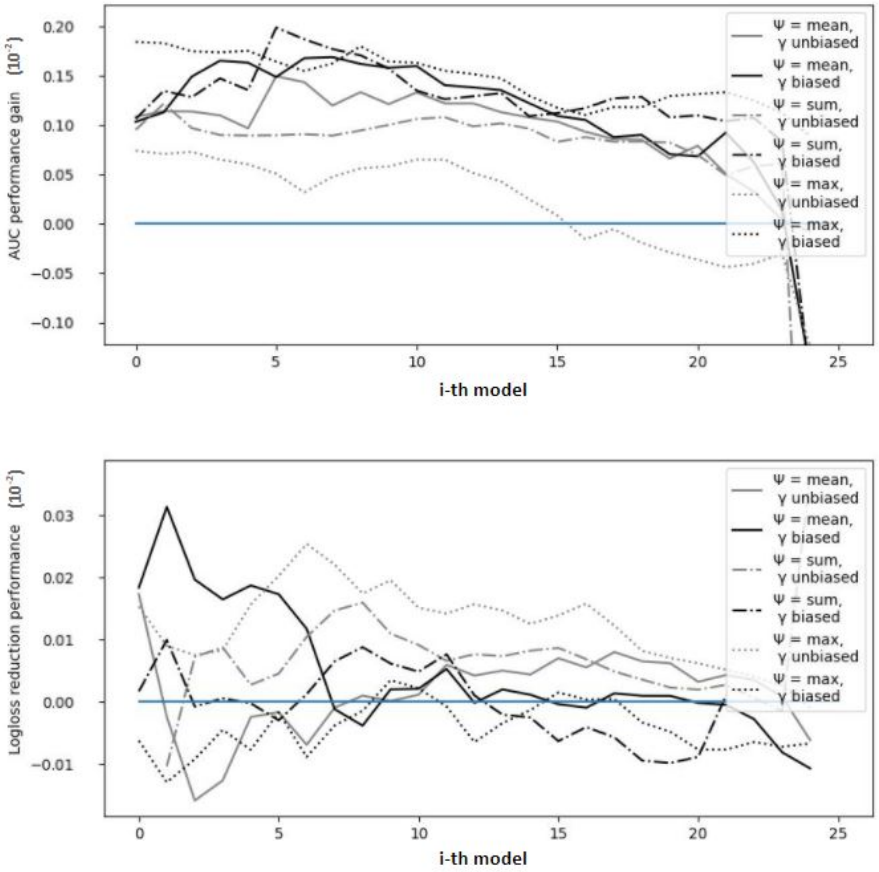
## 5.1 Model relevance

### 5.1.1 Performances on test set

RGCNConv proved to be too expensive from a computational point of view during the experimental process. This architecture is therefore abandoned, although we recognise its theoretical interest.

GraphConv was the fastest algorithm to train given its simplicity. However, the learning process regularly fails to converge. We assume that its simplicity is responsible for the instability of the learning process : the model might just not be detailed enough to correctly represent the studied phenomena. The most promising architectures seem to be NNConv and unbiased EdgeConv, which both provided satisfying convergence properties during training.

Figure 3: Performance gain of NNConv settings over the test dataset




Models' performances are displayed comparatively to LightGBM's, which serves as baseline. Each one of the six NNConv architecture settings features 25 models resulting from (15) in Appendix B. Top graph displays the gain for the Area Under Curve (AUC) metric. Bottom graph displays Logloss variation. The models which seem the most promising combine a positive gain in AUC and negative variation in Logloss.

Figure 3 illustrates the performances of the NNConv models over the test set, both in terms of AUC gain and Logloss variation compared to the LGBM baseline $\hat{H}^{\iota}$. The most promising models combine a positive AUC gain and a negative Logloss variation, although it is possible that some other models that don't reach those conditions still manage to learn information about risk contagion. Amongst the 25 models displayed for each of the six NNConv settings, a few manage to reach $+1.10^{-3}$ AUC gain and $-1.10^{-4}$ Logloss reduction, namely the first five max-biased and mean-unbiased settings, and the 18th to 20th ones of the sum-biased setting.

To further our analysis, we need to filter this batch of models on a statistical significance criterion. The objective is triple : we want to reduce this population of models to a handlable subset, maximize our chances to select those which learnt information about failure contagion,

and confirm that any model of the selected subset overperforms the baseline in a statistically significant way. In other terms, we need to apply a statistical filter that is strict enough to keep only a handful of models, and verify that the baseline has been excluded from that resulting subset. The method is described in the following section.

### 5.1.2 Model Confidence Set (MCS) filtering methodology

All $\hat{H}^i_{\Phi,\Psi,\gamma}$ predictors - along with the baseline LightGBM predictor - are submitted to a Model Confidence Set (MCS) experiment. MCS was introduced by Hansen, Lunde & Nason [2009] as a statistical method to identify the set of best elements from a collection of models. Formally, it is defined as the subset of objects whose expected loss over the test sample is smaller than the rest of the collection's. If we denote $\mathcal{H} = (\hat{H}^i_{\Phi,\Psi,\gamma}) \cup \hat{H}^\iota$ the set of all candidate predictors, then :

$$MCS_{\mathcal{H}} = \left\{ H^i \in \mathcal{H} \mid \mathbb{E}_Y \left[ L(Y, H^i) - L(Y, H^j) \right] \leq 0, \quad \forall H^j \in \mathcal{H} \right\} \tag{10}$$

The MCS set is recursively built following an *elimination rule* $e$ and an *equivalence test* $\delta$ on confidence level $\alpha$ determining the stop condition of the procedure. $\delta$ tests the null hypothesis that all expected losses of the current model collection are statistically equal. On rejection of that hypothesis, a new collection is built from the current one through elimination of the weakest model according to $e$. Once the assumption $\delta$ is met, the procedure ends and the current model collection is considered to be the MCS. The procedure is illustrated on Algorithm **1**.

**Algorithm 1** Model Confidence Set determination

```
1: procedure MCS
2:     M ← H
3: test:
4:     if δ_M at level α is rejected then
5:         M ← M\e_M
6:         goto test
7:         close;
8:     MCS_H ← M
```

We use Binary cross-entropy, which we already defined in section 3.2.3, as support loss function for this experiment. We apply a separate MCS procedure on each model family of the same message and agregation functions $(\Phi, \Psi)$. For each family, the test is failed if the MCS still includes the baseline LGBM predictor : no significant evidence is found that an element of that family provides better predictions than the intrinsic predictions of $\hat{H}^\iota$. In this case, the family is discarded from further consideration. In the opposite case, we keep one element from the MCS for further analysis. Should we prove that this element is different from a homothecy of $\hat{H}^\iota$, we will have achieved our goal of generating a model learning from firm neighboring interactions.

In total, 150 predictors $\hat{H}^{i}_{\Phi,\Psi,\gamma}$ from the NNConv model family, 75 from the unbiased EdgeConv family and 75 from the unbiased GraphConv family are submitted to the MCS procedure at 95% confidence level, alongside with the baseline predictor $\hat{H}^{\iota}$. The resulting set of selected models is the following :

$$\{\hat{H}^{2}_{\text{NN,sum,unbiased}}, \hat{H}^{2}_{\text{NN,max,biased}}, \hat{H}^{3}_{\text{Edge,max,unbiased}}, \hat{H}^{3}_{\text{Graph,max,unbiased}}\}$$

## 5.2 Empirical analysis of the risk transfer phenomenon

One of the main objectives of this paper is to determine if the MPNN architecture is actually able to learn information concerning the risk transfer phenomenon from neighboring companies. The difficulty lies in the fact that there is currently no robust methodology to bring explainability and interpretability to MPNN models. Indeed, these algorithms are recent and at crossroads between neural networks and graph theory, inheriting the black box issue which specifically plagues such complex models.

The following sections introduce our answers to the subsequent questions concerning the four previously selected models : *Which of those models differ enough from the identity function ? What relationships have those models established between neighboring risk and individual risk ? Can we assert from the results that similarities between neighbors facilitate risk transfer ?*

### 5.2.1 Linear regression setup

To test the working hypothesis that led to (5), we apply a Linear Regression of the form : $\hat{y} \sim \hat{y}^{\iota} + \hat{y}^{\nu} + \eta$ to each model's inputs $(\hat{y}^{\iota}, \hat{y}^{\nu})$ and output $\hat{y}$.

The first objective is to determine if the coefficient of $\hat{y}^{\nu}$ is attributed a significantly non-null value. A model not fulfilling this condition would just end up being a homothetic function between pre-correction and post-correction risk prediction, not attributing any value to the neighboring information. Such a model would not necessarily be wrong, since it passed the MCS test and therefore performs better than the baseline, but it would be irrelevant to our investigation of neighboring risk.

The second objective is to learn if the external risk is imported from all neighbors at random, or if a specific group of neighbors is responsible for transmitting the risk. To do so, let's first split the firm population in two groups $\Gamma^{1}$ and $\Gamma^{0}$ which respecticely experienced a failure or not during the years preceding the experiment (variable *FailPrevious* from Table **1**). Let's then consider $\rho$ a specific similarity of the vector $V$ detailed in table **1**. We can similarly split the neighbors of any company $v$ in two groups $\Gamma^{\rho}$ and $\Gamma^{\bar{\rho}}$ respectively sharing or not that similarity with $v$ : $\Gamma^{\rho} = \{\hat{y}^{\iota}_{w} \mid e^{\rho}_{v,w} = 1\}$ and $\Gamma^{\bar{\rho}} = \{\hat{y}^{\iota}_{w} \mid e^{\rho}_{v,w} = 0\}$.

From there, we can compute for each node $v$ the sum $\sigma^{\rho,1} = \sum_{w \in V^\rho \bigcap \Gamma^1} h_w$ of the surrounding risk from all defaulting neighbors sharing the characteristic $\rho$, and $\sigma^{\rho,0} = \sum_{w \in V^\rho \bigcap \Gamma^0} h_w$ its counterpart for all non-defaulting neighbors sharing the characteristic $\rho$. Similarly, we also generate $\sigma^{\bar{\rho},1} = \sum_{w \in V^{\bar{\rho}} \bigcap \Gamma^1} h_w$ and its counterpart $\sigma^{\bar{\rho},0} = \sum_{w \in V^{\bar{\rho}} \bigcap \Gamma^0} h_w$ the sum of the surrounding risk for all defaulting and non-defaulting neighbors which doesn't share any characteristic in common with $v$. We finally dispose of a detailed decomposition of the neighboring risk inputed in the MPNN :

$$\hat{y}^\nu = f\left(\sum_\rho \left(\sigma^{\rho,0} + \sigma^{\rho,1}\right) + \sigma^{\bar{\rho},0} + \sigma^{\bar{\rho},1}\right) \tag{11}$$

We aknowledge that neighbor firms occasionnally share more than one similarity and as a result their risk may be counted several times in (11). However, review of the data reveals that this effect is negligible in the global computation. The linear regression can then be expressed as :

$$\hat{y} \sim \alpha \hat{y}^\iota + \sum_\rho \alpha^{\rho,0}\sigma^{\rho,0} + \sum_\rho \alpha^{\rho,1}\sigma^{\rho,1} + \alpha^{\bar{\rho},0}\sigma^{\bar{\rho},0} + \alpha^{\bar{\rho},1}\sigma^{\bar{\rho},1} + \eta \tag{12}$$

All covariates are submitted to logit transformation before fitting the model. Given the nature of the $\sigma$ covariates, the $\alpha$ coefficients can be interpreted in two different ways once the regression fitted. Indeed for a specific similarity $\rho$, there are two possibilities to obtain a higher $\sigma^\rho$ : either the surrounding risk in $\Gamma^\rho$ is globally high, or $\Gamma^\rho$ simply contains a lot of neighbors. As a result, it is impossible to determine if a higher $\alpha$ would reveal a high surrounding risk, or numerous similar companies. To solve that ambivalence, we also lead a second regression by replacing all the risk-sum covariates by risk-average covariates $\mu^{\rho,0/1} = \sigma^{\rho,0/1} \ / \ |V^\rho \bigcap \Gamma^{0/1}|$.

### 5.2.2 Results of the linear regression

When applied to the two EdgeConv and GraphConv models selected in section 5.1.2, the linear regression only returns expressions of the form $\hat{y} = \alpha\hat{y}^\iota + \eta + o(\alpha)$, which reveals that those two architectures do not grant any importance to neighboring information and only generate homotheties of $\hat{y}^\iota$. Those two architectures are therefore discarded. The results of the explanative linear regression analysis are summarized on Table **4** for the two remaning NNConv settings, with risk-sum and risk-average covariates.

Over those settings, we can first notice that the influence from neighbors with a similar order-of-magnitude revenue is not attributed any statistically significant effect. The same absence of conclusion holds for defaulting neighbors of the same age and defaulting neighbors which do not

Table 4: Linear regression explaining NNConv architecture's predictions.

| Variable | $\Psi$=sum, $\gamma$ unbiased | $\Psi$=max, $\gamma$ biased | $\Psi$=sum, $\gamma$ unbiased | $\Psi$=max, $\gamma$ biased |
|---|---|---|---|---|
| | Sum-aggregated covariates | | Mean-aggregated covariates | |
| InternalRisk | **0.3234** | **0.3255** | **0.3233** | **0.3257** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| NeighborRisk_SameStaffSize_NoFail | -0.0013 | 0.0003 | **-0.0219** | **-0.0153** |
| | (0.381) | (0.874) | (0.000) | (0.000) |
| NeighborRisk_SameStaffSize_Fail | **0.0248** | **0.0258** | **0.0250** | **0.0234** |
| | (0.000) | (0.000) | (0.000) | (0.001) |
| NeighborRisk_SameAgeDecade_NoFail | **0.0067** | **0.0093** | **0.0272** | **0.0400** |
| | (0.001) | (0.000) | (0.000) | (0.000) |
| NeighborRisk_SameAgeDecade_Fail | -0.0067 | -0.0086 | -0.0052 | -0.0084 |
| | (0.189) | (0.108) | (0.309) | (0.115) |
| NeighborRisk_SameSector_NoFail | **0.0121** | **0.0173** | **0.0095** | **0.0067** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| NeighborRisk_SameSector_Fail | **0.0332** | **0.0530** | **0.0350** | **0.0533** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| NeighborRisk_SameRevenue_NoFail | **-0.0040** | **0.0139** | 0.0014 | 0.0010 |
| | (0.000) | (0.000) | (0.309) | (0.500) |
| NeighborRisk_SameRevenue_Fail | -0.0083 | 0.0010 | **-0.0100** | 0.0022 |
| | (0.100) | (0.844) | (0.046) | (0.681) |
| NeighborRisk_SameRepresentative_NoFail | **0.0101** | **0.0281** | **0.0119** | **0.0257** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| NeighborRisk_SameRepresentative_Fail | 0.0220 | **0.0757** | **0.0300** | **0.0821** |
| | (0.136) | (0.000) | (0.042) | (0.000) |
| NeighborRisk_NothingInCommon_NoFail | **-0.0252** | -0.0041 | **-0.0179** | **0.0432** |
| | (0.000) | (0.267) | (0.000) | (0.000) |
| NeighborRisk_NothingInCommon_Fail | 0.0029 | -0.0042 | 0.0020 | -0.0058 |
| | (0.416) | (0.263) | (0.577) | (0.118) |

The regression coefficients indicate the estimated strength of the relationship between the internal and external origins of the risk and its aggregated value provided by the best performing GNN architectures. Coefficients whose p-value is below 0.05 are considered statistically significant. The two left columns present the results of the regression where the transmitted risk - the $\sigma$ covariates of equation (12) - has been summed over the corresponding neighbor populations. The two right columns present the case where the transmitted risk has been averaged.

share any common characteristic with the targeted company. For the other characteristics, all coefficient are statistically significant and mostly positive.

Our analysis for the resulting regression coefficients is the following :

***Activity sector***

Risk rises with the presence of defaulting neighbors within the same activity sector, which can be interpreted as the consequence of a local sectorial crisis. We could argue that companies may *a contrario* benefit from the fall of local competitors, however this effect doesn't seem to be predominant : the regression coefficient *NeighborRisk_SameSector_Fail* is among the strongest positives. This is in line with Enright's analysis of direct competition in clusters [1998], according to which same-sector firm concentration raises the whole cluster's interregional competitiveness. As such, the fall of competitors would weaken the sector's resources - in terms of sectorial expertise as well as accessibility to specialised workforce - and impact the surviving members.

Although being weaker, the previous effect holds with the presence of non-defaulting neighbors within the same activity sector. Coefficients from the sum-aggregated neighbor risk also outweights the mean-aggregated ones. This suggests that the mere presence of close competitors is sufficient to generate risk, which is consistent with the literature's indication that compact competitive company networks have a detrimental effect on survival rates [Shaver & Flyer, 2000 ; Cabral, Wang & Ju, 2018]. It is indeed suggested that, as most developed firms have the most to lose due to spillover effects to the benefit of the weakest ones, geographical clustering results in anti-selection processes within the same activity sector.

***Representatives***

Risk rises with the presence of defaulting companies sharing the same representatives, which can be interpreted as the consequence of a poor management strategy from those companies' leaders. Said otherwise, this result provides statistical evidence that the risk of failure increases when the company manager experienced previous failures.

Risk also rises with the presence of non-defaulting companies sharing the same representatives, although being weaker than its counterpart with defaulting companies. This suggests an interesting fact about leadership : simultaneously managing more than one company is enough to generate risk, or at least to transmit it from one to another.

The current literature does not directly address the subject of multi-company management, although a few studies focused on relationships between managerial profiles - mostly entrepreneurs

- and firm performance. The data needed to describe and classify entrepreneur profiles is scarce and difficult to normalize, as most of it derive from sociology and psychology studies. Gimeno and Folta [1997] argued that entrepreneurs who inherited businesses are more likely to sustain it at higher costs and tolerate underperformance than would 'serial founders'. Being managed by dedicated profiles would then not necessarily be the sign of growth and performance, but rather resilience to failure. However, some other studies temper the impact of managerial profiles on firm performance [Westhead & Wright, 2001].

***Age and size***

Risk rises with the average risk of non-defaulting neighbors of the same age, but decreases with the average risk of non-defaulting neighbors of the same size, as suggested by the two covariates *NeighborRisk_SameStaffSize_NoFail* and *NeighborRisk_Same_AgeDecade_NoFail*. Those seemingly counterintuitive results are more difficult to interpret, however answers may be found in the literature. We can argue that they are a combination of compact clustering effects. The detrimental effect of same-age relationship may be a reflection of the current state of the cluster within its life cycle. Cluster productivity is thought to erode over time [Menzel & Fornahl, 2010]. The beneficial effect of same-size relationship may *a contrario* reflect the strength of homogeneous agglomeration economies and knowledge spillovers. This balancing combination of opposed effects has already been highlighted by Weterings & Marsili [2015].

Risk rises with the presence of defaulting neighbors of the same size. This may be the most difficult result to analyse, as it cannot be interpreted as the consequence of a local sectorial crisis and there is no trace of it in the literature. We assume that it results from the sudden irruption of a defaulting link in well-established company clusters benefiting from agglomeration economies and knowledge spillover effects. Such a default would act as a shock propagating to closely related businesses, ultimately weakening the local economic equilibrium.

***Revenue and absence of commonalities***

The amplitude of the other statistically significant effects is generally lower and there is some uncertainty about their sign, so we cannot conclude that they play a determinant role in risk transmission.

### 5.2.3 GNN Explainer

In this section, we propose an alternative analysis of the trained MPNN models. The considered tool is a GNN explainer, which was recently introduced as the first interpretability model-agnostic method, designed to tackle any variant of the Graph Neural Network family.

From a trained MPNN algorithm based on a specific graph $G(\Gamma, V)$ and a specific node $v$, its objective is to extract a subsample graph $G_v(\Gamma_v, V)$ composed only of the most relevant nodes and features used for $v$'s risk scoring. The notion of "most relevant" subgraph is defined as the result of a constrained optimisation problem looking for the smallest subgraph in size, containing $v$, and maximizing mutual information MI over $P_\Phi$ the distribution of $Y$ learnt by the GNN. This can be formalized as :

$$\begin{aligned} G_v(\Gamma_v, V) &= \underset{G_S, \Gamma_S}{\operatorname{argmin}} \operatorname{MI}(Y_v, G_S(\Gamma_S, V)) \\ &= \underset{G_S, \Gamma_S}{\operatorname{argmin}} E(Y_v) - E(Y_v | G = G_S, \Gamma = \Gamma_S) \\ &\text{subject to : } |G_S| \leq c \end{aligned} \tag{13}$$

with $E(Y|G = G_S, \Gamma = \Gamma_S) = -\mathbb{E}_{Y|G_S,\Gamma_S}[\log P_\Phi(Y|G = G_S, \Gamma = \Gamma_S)]$ the conditional entropy of $Y$ on the subgraph $G_S$, $E(Y) = -\mathbb{E}[\log P_\Phi(Y)]$ the entropy of $Y$ on the initial graph $G$ and $c$ a factor constraining the size of $G_S$.

Applied to our problematic, GNN Explainer can filter the neighborhood $\Gamma_v$ of any firm $v$ to keep only the fraction of elements assumed to be the most relevant in the computation of $\hat{y}_v$ by the MPNN. Said otherwise, it aims to reveal from which specific neighbors of $v$ the MPNN imports the risk, given a confidence probability $\alpha$. From there, we can test whether firms sharing common characteristics with their neighbors are or not surrepresented within the population of relevant risk transmitters.

To formalize it, let $\Xi$ be the extraction function of the GNN Explainer. For each pair of firms, it returns 1 if one of the node is part of the most relevant subgraph of the other, as follows :

$$\begin{aligned} \Xi : \quad & V \to \{0; 1\} \\ & e_{v,w} \longmapsto \mathbb{1}_{w \in G_v \text{ or } v \in G_w} \end{aligned} \tag{14}$$

We then conduct a t-test experiment comparing $\mathbb{E}[\Xi(V^\rho)]$ and $\mathbb{E}[\Xi(V^{\bar{\rho}})]$ for each of the five similarity indicators $\rho$. Table **5** summarizes the results of the test, conducted over a sample of 5,000 observations from the test set with an average of 47.8 neighbors each.

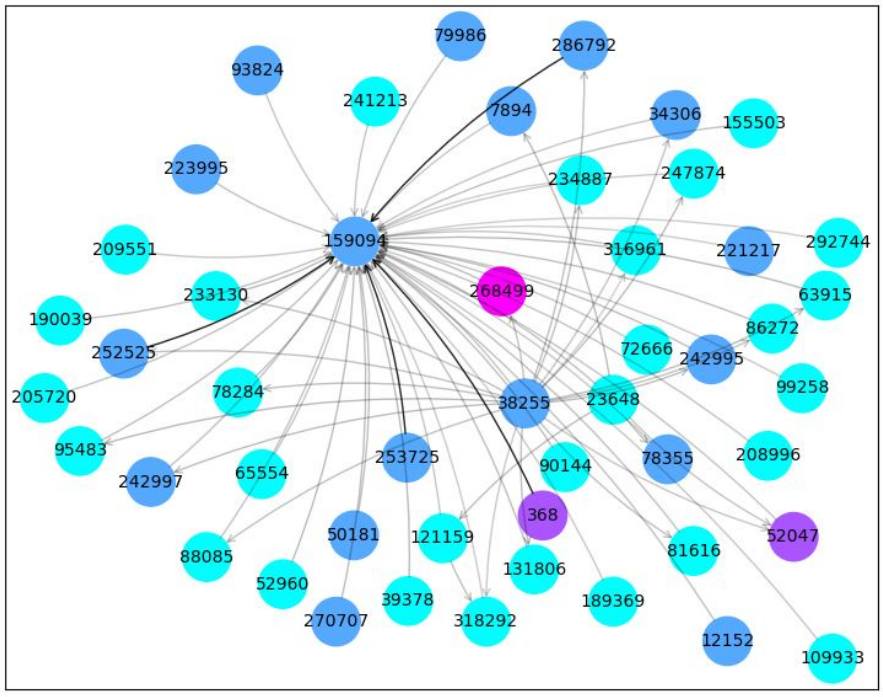
We can first notice that simply sharing a similarity with a neighbor makes it more likely to be considered among the relevant risk transmitters. This effect is statistically significant for

Table 5: T-test analysis for important neighbors, as categorized by the GNN Explainer

| Variable $\rho$ | $\mathbb{E}[\Xi(V^{\bar{\rho}})]$ $(10^{-2})$ | $\mathbb{E}[\Xi(V^{\rho})]$ $(10^{-2})$ | t-test statistic | p-value |
|---|---|---|---|---|
| SameStaffSize | 5.26 | 5.83 | -4.34*** | $1.45e^{-5}$ |
| SameAgeDecade | 5.29 | 5.48 | -1.84* | $6.64e^{-2}$ |
| SameSector | 5.27 | 6.00 | -4.62*** | $3.81e^{-6}$ |
| SameRevenue | 5.30 | 5.49 | -1.61 | $1.07e^{-1}$ |
| SameRepresentative | 5.32 | 6.40 | -3.63*** | $2.78e^{-4}$ |

companies sharing the same staff size, activity sector, or representatives. It is also in accordance with the conclusions of the previous section.

Figure 4: An example of GNN Explainer output for a single observation




The neighborhood of the targeted firm 159094 is displayed as a mathematical graph. Color indicates the number of characteristics from the edge vector $V$ that each neighbor shares with target firm (cyan : 0 - blue : 1 - violet : 2 - pink : 3). Neighbors identified by the GNN Explainer as the most important in terms of risk transmission have thick arrows.

As a side note, it is also possible to apply GNN Explainer on an individual scale to determine which neighbors of a specific company are considered to convey the main external risk - or absence of risk. An illustrative example of this idea is provided on Figure 4

# 6 Conclusion

To the best of our knowledge, this study is the first to model risk as a transmissible potential between neighboring SMEs, as well as using MPNNs to simulate this contagion on a large-scale dataset. We assumed risk to be the combination of an internal factor and a local one. We first estimated PD for each firm with modern predictive algorithms such as LGBM, and used PD as a proxy for internal risk. We then assumed that risk to be transmissible among close geographical neighbors. We embedded the economic landscape of French companies in a mathematical graph structure, used node features to store internal risk, and edge features to act as a similarity measure for each pair of connected companies. We then applied GNN to simulate interactions between companies of the same geographical clusters.

We found that risk - or absence of risk - is more likely to be transmitted between companies sharing the same size, activity sector, and managers or representatives. The interpretation of GNN's outputs is in accordance with several aspects of cluster theory. Same-sector firm concentration contributes to the strength of the cluster in interregional competitiveness, at the expense of its individual constituents. Indeed, we found that firms suffer from both the presence and fall of local competitors, suggesting that joining a sector cluster is only viable when there is much to gain and little to lose. Sectoral clusters may then become antiselectors in the long run, leading to their decline. We also found that intra-cluster firms suffer from age uniformity and benefit from size homogeneity. Although we did not study the effects of massively dissimilar size relationships, those findings suggest that healthy clusters would be those regularly accepting new entrants - there to benefit from spillover effects - while ensuring that local labor accessibility remains balanced across all firms. On a final note, we found that multi-company leaders and representatives are likely to generate and pass on risk between their firms, especially if one of it recently failed.

GNNs prove to be robust and adaptive tools fit to learn information from complex systems. More complete systems could integrate the presence of local growth facilitators such as the presence of universities, warehouses and major roadways. Similarity metrics can be improved and extended to not only consider commonalities between firms but also between their internal risk - notably with the emergence of SHAP value methodologies. Predictive practical applications are also possible, as we briefly demonstrated with the use of GNN Explainer to determine from which neighbors the external risk of a target company is more likely to originate.

# Appendix A MPNN architectures

Table **6** and the following paragraphs introduce the different candidate architectures of MPPN used in the study. Each architecture implements a specific neighbor message function $\Phi$.

Table 6: Architectures of MPNN

| Architecture name | Mathematical expression | Parameters |
|---|---|---|
| GraphConv | $\dot{h}_v = \Theta_1 h_v + \Psi_{h_w \in \Gamma_v}(\Theta_2 h_w)$ | $(\Theta_1, \Theta_2) \in \mathbb{R}^{1\times\Pi}$ |
| EdgeConv | $\dot{h}_v = \Psi_{h_w \in \Gamma_v}(A(h_v \| h_v - h_w))$ | $A \in \mathbb{R}^{2\Pi} \to \mathbb{R}$ |
| NNConv | $\dot{h}_v = \Theta h_v + \Psi_{h_w \in \Gamma_v}(A(e_{v,w})h_w)$ | $\Theta \in \mathbb{R}^{1\times\Pi}$<br>$A \in \mathbb{R}^{Q} \to \mathbb{R}^{1\times\Pi}$ |
| RGCNConv | $\dot{h}_v = \Theta h_v + \sum_{\rho=1}^{Q}\left[\Psi_{h_w \in \Gamma_v}\left(A^{\rho}(e_{v,w})h_w \cdot \mathbb{1}_{e^{\rho}_{v,w}=1}\right)\right]$ | $\Theta \in \mathbb{R}^{1\times\Pi}$<br>$(A^{\rho}) \in \mathbb{R}^{Q} \to \mathbb{R}^{1\times\Pi}$ |

GraphConv is the simplest implementation of MPNN. It is based on the hypothesis that similarity or dissimilarity between neighbor nodes does not play any role in the message transmission process : the information held in $V$ is not considered. Data from neighboring nodes is simply submitted to a linear combination of parameter $\Theta_2$ then aggregated by $\Psi$. Applied to our subject, this MPNN implementation treats neighbor influence as a constant effect, meaning that risk transmission occurs with the same intensity regardless if neighbor firms are from completely different activity sectors, size or age categories. In view of the literature we detailed in section 2, this hypothesis clearly lacks refinement.

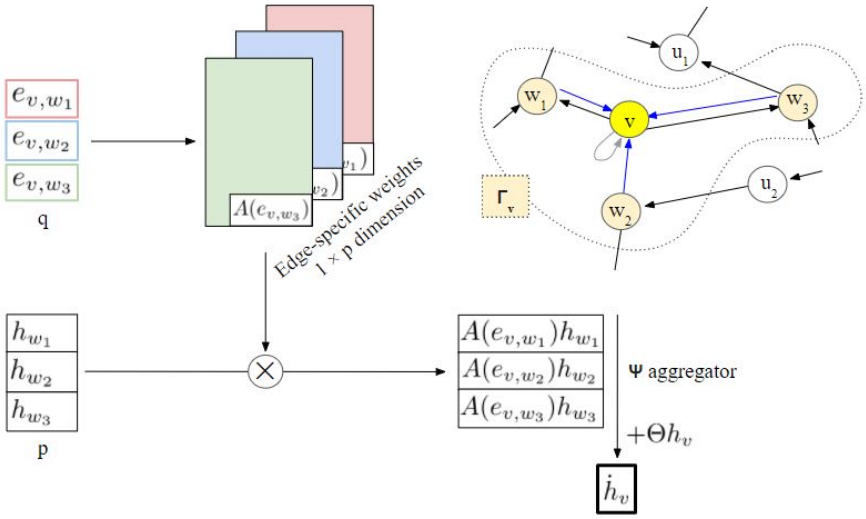
EdgeConv is almost similar to GraphConv in that it does not consider the notion of similarity / dissimilarity between pair of nodes. Its architecture brings the knowledge of target score $h_v$ within the message computation. This message now results from direct interpretation of target score and its difference with neighbor scores $h_v - h_w, \forall w \in V_v$ by a learned subneural network $A \in \mathbb{R}^{2\Pi} \to \mathbb{R}$.

NNConv is based on the hypothesis that the message transmission process depends on the nature of the relationship between neighboring nodes. It differentiates from GraphConv in the way the neighboring information $h_w$ is weighted before transmission. The constant weight vector $\Theta_2$ is replaced by a transformation of the edge features $e_{v,w}$ through a learned subneural network $A \in \mathbb{R}^{Q} \to \mathbb{R}^{\Pi}$. Applied to our subject, this architecture allows finetuning risk transfer intensity depending on the nature of similarities between neighor firms. This architecture is visually detailed on Figure 5.

Relational Graph Convolutional Network (RGCN)'s architecture is a more refined implementation of the idea motivating NNConv's architecture. Here, each feature $\rho$ of $e_{v,w}$ is attributed

a distinct subneural network $A^\rho \in \mathbb{R}^Q \to \mathbb{R}^\Pi$ which only activates if the two neighbor nodes $v$ and $w$ share the similarity defined by $\rho$. In practice, the authors recommend to use simple linear combinations for each $A^\rho$ as the learning process of this architecture is computationally expensive [Schlichtkrull et al., 2018].

Figure 5: Illustration of NNConv




In this illustrative example, NNConv is applied to the neighborhood $\Gamma_v$ of node $v$, which is composed of the three neighbors $(w_1, w_2, w_3)$. First, the Q-dimensional edge features $(e_{v,w})$ are submitted to a subneural network $A \in \mathbb{R}^Q \to \mathbb{R}^\Pi$. The resulting $\Pi$-dimensional weights are applied to their respective node features $(h_w)$ to generate one message per neighbor. The message aggregator $\Psi$ reduces all the messages to a single scalar, which is then added to $\Theta h_v$ the refined information from target node $v$.

For each of those architectures we have to specify the behavior of the two functions $\Psi$ and $\gamma$. More precisely, the neighbor message aggregator $\Psi$ must be a permutation-invariant operator of $\mathbb{R}^\mathbb{N} \to \mathbb{R}$ and three usual choices are the average, summation and maximum aggregators. We can also specify the update function $\gamma$ to learn or not a bias, and will subsequently refer to those options as *biased* and *unbiased* $\gamma$ for the rest of the paper. We will not discuss the choice of the subneural networks $A$ and $(A^\rho)$ for NNConv and RGCNConv as it is of little interest in this paper. The choice of $\Psi$'s and $\gamma$'s behavior leaves us with six distinct settings for each architecture.

# Appendix B MPNN Training methodology

From the French national company directed graph $G^{2017}(\Gamma, V)$ built with all data from the financial year 2017, we sample $\kappa$ independant training sets $G^{2017}_\kappa(\Gamma_\kappa, V)$ of fixed size $\lfloor \frac{n}{\kappa+1} \rfloor$

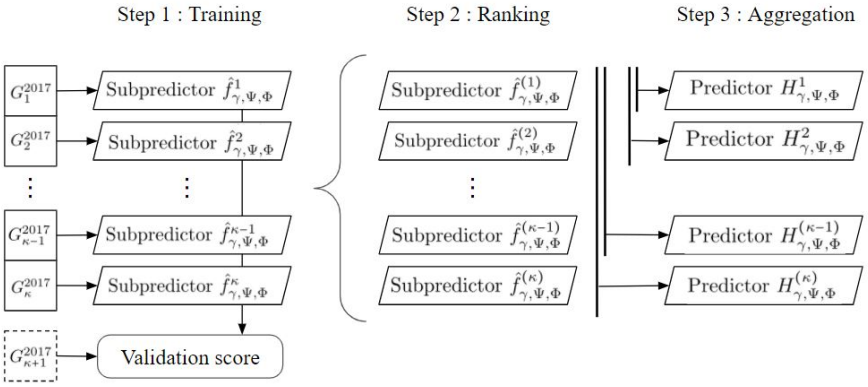
with one additional holdout sample kept as validation set. Each of those training sets is then used to feed a MPNN learner $\hat{f}^{\kappa}_{\gamma,\Psi,\Phi} : \Gamma_\kappa \times \mathcal{P}(V) \to \mathbb{R}$ for 1000 epochs. We finally generate $(\hat{f}^{(1)}, \hat{f}^{(2)}, ..., \hat{f}^{(\kappa)})_{\gamma,\Psi,\Phi}$ the best-to-worst ordered set of those $\kappa$ learners according to their performance evaluated on the validation set.

The final $\kappa$ predictors considered to evaluate a MPNN algorithm are the successive mean-aggregates of the i-th best learners generated by this algorithm :

$$H^i_{\gamma,\Psi,\Phi} : \begin{array}{c} \Gamma \times \mathcal{P}(V) \to \mathbb{R} \\ (h_v, V_v) \mapsto \frac{1}{i} \sum_{z=1}^{i} \hat{f}^{(z)}_{\gamma,\Psi,\Phi}(h_v, V_v) \end{array} \tag{15}$$

This methodology is illustrated on Figure 6

Figure 6: Generation of the predictors



# References


[1] Agarwal, S., & Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies, 23*(7), 2757-2788.

[2] Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. *Abacus, 43*(3), 332-357.

[3] Baptista, R., & Swann, P. (1998). Do firms in clusters innovate more?. *Research policy, 27*(5), 525-540.

[4] Barreto, G., & Artes, F. (2013). *Spatial correlation in credit risk and its improvement in credit scoring.* Working Paper WPE: 321/2013.

[5] Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of economic dynamics and control, 36*(8), 1121-1141.

[6] Beaudry, C., & Swann, P. (2001). Growth in industrial clusters: A bird's eye view of the United Kingdom. *Stanford Institute for Economic Policy Research Discussion Paper 00-38.*

[7] Beck, T., & Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & finance, 30*(11), 2931-2943.

[8] Berger, A. N., & Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The economic journal, 112*(477), F32-F53.

[9] Boschma, R. A., & Wenting, R. (2007). The spatial evolution of the British automobile industry: Does location matter?. *Industrial and corporate change, 16*(2), 213-238.

[10] Branco, A., & Lopes, J. C. (2018). Cluster and business performance: Historical evidence from the Portuguese cork industry. *Investigaciones de Historia Económica, 14*(1), 43-53.

[11] Cabral, L., Wang, Z., & Xu, D. Y. (2018). Competitors, complementors, parents and places: Explaining regional agglomeration in the US auto industry. *Review of Economic Dynamics, 30*, 1-29.

[12] Calabrese, R., Andreeva, G., & Ansell, J. (2019). "Birds of a feather" fail together: Exploring the nature of dependency in SME defaults. *Risk Analysis, 39*(1), 71-84.

[13] Ceci, F., & Iubatti, D. (2012). Personal relationships and innovation diffusion in SME networks: A content analysis approach. *Research policy, 41*(3), 565-579.

[14] Cefis, E., & Marsili, O. (2005). A matter of life and death: innovation and firm survival. *Industrial and Corporate change, 14*(6), 1167-1192.

[15] Cochran, A. B. (1981). Small business mortality rates: A review of the literature. *Journal of Small Business Management (pre-1986), 19*(000004), 50.

[16] Degryse, H., & Ongena, S. (2005). Distance, lending relationships, and competition. *The Journal of Finance, 60*(1), 231-266.

[17] Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. Journal of economic geography, 10(4), 495-518.

[18] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7, 1-30.

[19] Depret, M. H., & Hamdouch, A. (2009). Clusters, réseaux d'innovation et dynamiques de proximité dans les secteurs high-tech. Une revue critique de la littérature récente. *Revue d'économie industrielle*, (128), 21-52.

[20] De Silva, D. G., & McComb, R. P. (2012). Geographic concentration and high tech firm survival. *Regional Science and Urban Economics, 42*(4), 691-701.

[21] DeYoung, R., Glennon, D., & Nigro, P. (2008). Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation, 17*(1), 113-143.

[22] Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European journal of operational research, 90*(3), 487-513.

[23] Dumais, G., Ellison, G., & Glaeser, E. L. (2002). Geographic concentration as a dynamic process. *Review of economics and Statistics, 84*(2), 193-204.

[24] Dunne, T., Roberts, M. J., & Samuelson, L. (1989). The growth and failure of US manufacturing plants. *The Quarterly Journal of Economics, 104*(4), 671-698.

[25] Enright, M. J. (1998). Regional clusters and firm strategy. *The dynamic firm: The role of technology, strategy, organization and regions,* 315-42.

[26] Everett, J., & Watson, J. (1998). Small business failure and external risk factors. *Small business economics, 11*(4), 371-390.

[27] Frenken, K., Cefis, E., & Stam, E. (2015). Industrial dynamics and clusters: a survey. *Regional studies, 49*(1), 10-27.

[28] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association, 32*(200), 675-701.

[29] Gatti, D. D., Gallegati, M., Greenwald, B. C., Russo, A., & Stiglitz, J. E. (2009). Business fluctuations and bankruptcy avalanches in an evolving network economy. *Journal of Economic Interaction and Coordination, 4*(2), 195.

[30] Gilbert, B. A., McDougall, P. P., & Audretsch, D. B. (2008). Clusters, knowledge spillovers and new venture performance: An empirical examination. *Journal of business venturing, 23*(4), 405-422.

[31] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

[32] Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative science quarterly*, 750-783.

[33] Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453-497.

[34] Huggins, R., Prokop, D., & Thompson, P. (2017). Entrepreneurship and the determinants of firm survival within regions: human capital, growth motivation and locational conditions. *Entrepreneurship & Regional Development, 29*(3-4), 357-389.

[35] Ketels, C. (2003, December). The Development of the cluster concept–present experiences and further developments. In *NRW Conference on Clusters, Duisberg, Germany* (Vol. 5).

[36] Khelil, N. (2016). The many faces of entrepreneurial failure: Insights from an empirical taxonomy. *Journal of business venturing, 31*(1), 72-94.

[37] Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy, 99*(3), 483-499.

[38] Kukalis, S. (2010). Agglomeration economies and firm performance: the case of industry clusters. *Journal of Management, 36*(2), 453-481.

[39] Mater, A. C., & Coote, M. L. (2019). Deep learning in chemistry. *Journal of chemical information and modeling, 59*(6), 2545-2559.

[40] Maté-Sánchez-Val, M., López-Hernandez, F., & Fuentes, C. C. R. (2018). Geographical factors and business failure: An empirical study from the Madrid metropolitan area. *Economic Modelling, 74*, 275-283.

[41] McCann, B. T., & Folta, T. B. (2008). Location matters: where we have been and where we might go in agglomeration research. *Journal of management, 34*(3), 532-565.

[42] McCann, B. T., & Folta, T. B. (2011). Performance differentials within geographic clusters. *Journal of Business Venturing, 26*(1), 104-123.

[43] Menzel, M. P., & Fornahl, D. (2010). Cluster life cycles—dimensions and rationales of cluster evolution. *Industrial and corporate change, 19*(1), 205-238.

[44] Molina-Morales, F. X., & Expósito-Langa, M. (2012). The impact of cluster connectedness on firm innovation: R&D effort and outcomes in the textile industry. *Entrepreneurship & Regional Development, 24*(7-8), 685-704.

[45] Neu, P., & Kühn, R. (2004). Credit risk enhancement in a network of interdependent firms. *Physica A: Statistical Mechanics and its Applications, 342*(3-4), 639-655.

[46] Pavelkova, D., Zizka, M., Homolka, L., Knapkova, A., & Pelloneova, N. (2021). Do clustered firms outperform the non-clustered? Evidence of financial performance in traditional industries. *Economic Research-Ekonomska Istraživanja*, 1-23.

[47] Pe'er, A., & Vertinsky, I. (2006). Determinants of survival of de novo entrants in clusters and dispersal. *Available at SSRN 940477.*

[48] Pollard, J. S. (2003). Small firm finance and economic geography. *Journal of Economic Geography, 3*(4), 429-452.

[49] Pompe, P. P., & Bilderbeek, J. (2005). The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business venturing, 20*(6), 847-868.

[50] Porter, M. E., & Porter, M. P. (1998). Location, clusters, and the" new" microeconomics of competition. *Business Economics*, 7-13.

[51] Porter, M. E. (2003). Building the microeconomic foundations of prosperity: Findings from the business competitiveness index. *The global competitiveness report, 2004*, 29-56.

[52] Rosenthal, S. S., & Strange, W. C. (2005). The geography of entrepreneurship in the New York metropolitan area. *Federal Reserve Bank of New York Economic Policy Review, 11(2), 29-54.*

[53] Myles Shaver, J., & Flyer, F. (2000). Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic management journal, 21*(12), 1175-1193.

[54] St. John, P. C., Phillips, C., Kemper, T. W., Wilson, A. N., Guan, Y., Crowley, M. F., ... & Larsen, R. E. (2019). Message-passing neural networks for high-throughput polymer screening. *The Journal of chemical physics, 150*(23), 234111.

[55] Wennberg, K., & Lindqvist, G. (2010). The effect of clusters on the survival and performance of new firms. *Small Business Economics, 34*(3), 221-241.

[56] Weterings, A., & Marsili, O. (2015). Spatial concentration of industries and new firm exits: Does this relationship differ between exits by closure and by M&A?. *Regional Studies, 49*(1), 44-58.

[57] Westhead, P., & Wright, M. (1998). Novice, portfolio, and serial founders: are they different?. *Journal of business venturing, 13*(3), 173-204.

[58] Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). Gnn explainer: A tool for post-hoc explanation of graph neural networks. *arXiv preprint arXiv:1903.03894.*

[59] Zhao, T., & Jones-Evans, D. (2017). SMEs, banks and the spatial differentiation of access to finance. *Journal of Economic Geography, 17*(4), 791-824.

[60] Ziegler, T., Shneor, R., Garvey, K., Wenzlaff, K., Yerolemou, N., Rui, H., & Zhang, B. (2018). Expanding horizons: The 3rd European alternative finance industry report. *Available at SSRN 3106911.*