Network representation learning for credit scoring

By Ricardo Luis Muñoz Cancino

Department of Industrial Engineering

PhD Thesis Defense to obtain the degree of Doctor in Engineering Systems

PhD Advisor:

Sebastián A. Ríos Pérez

Co-Advisors:

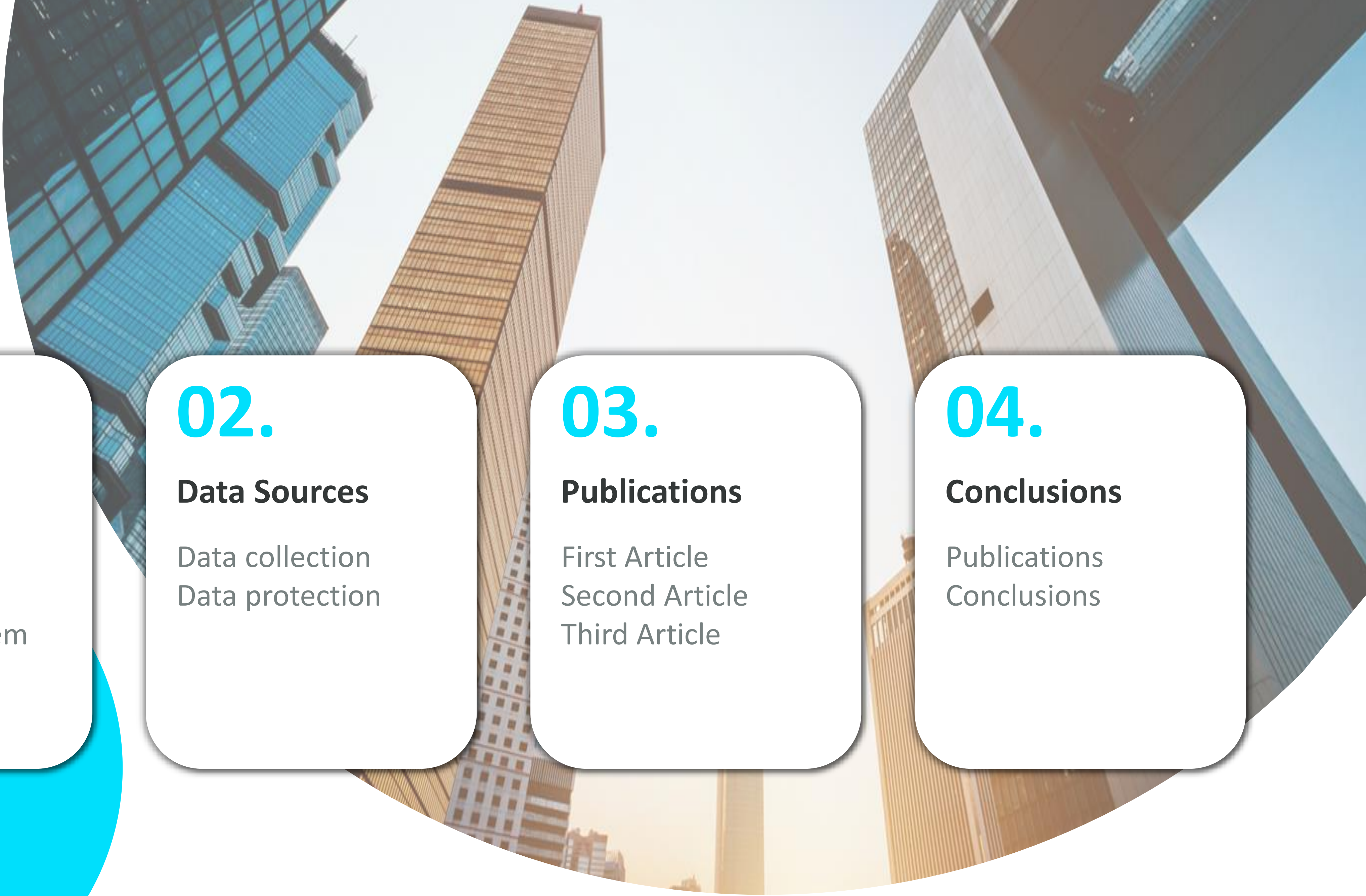
Cristián Bravo Román

Manuel Graña Romay

Committee Members:

Marcos Orchard Concha

Contents



Motivation

Motivation

Motivation


Your Credit Score
805
Excellent
Monitor Your Credit Activity
See Recent Credit Alerts
See Ways To Protect Your Identity
See Your Free Report
Ways To Boost Your Credit Score
CREDIT APPLICATION
CUSTOMER INFORMATION
NAME
ADDRESS
CITY
PHONE NO.
EMAIL
ACCOUNT NUMBER


USD

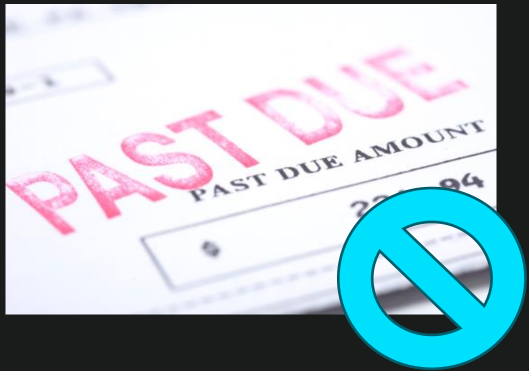

PAST DUE
PAST DUE AMOUNT
224.94


LOAN

Motivation


Your Credit Score
805
Excellent
Monitor Your Credit Activity
See Recent Credit Alerts
See Ways To Protect Your Identity
See Your Free Report
Ways To Boost Your Credit Score
CREDIT APPLICATION
CUSTOMER INFORMATION
NAME
ADDRESS
CITY
PHONE NO.
EMAIL
ACCOUNT NUMBER


USD


PAST DUE
PAST DUE AMOUNT


LOAN

Motivation

Family

Employer-Employee

Professional

Transactional

Motivation

Alternative Data and social interactions.

Telephone call data (Óskarsdóttir et al., 2017; Óskarsdóttir, Bravo, Vanathien, & Baesens, 2018a; Óskarsdóttir, Bravo, Sarraute, Vanthienen, & Baesens, 2019)	**Written risk assessments** (Stevenson, Mues, & Bravo, 2021)	**Data generated by an app-based marketplace** (Roa, Correa-Bahnsen, et al., 2021; Roa, Rodríguez-Rey, Correa-Bahnsen, & Valencia, 2021)	**Social media data** (Tan & Phan, 2018; Cnudde et al., 2019; Putra, Joshi, Redi, & Bozzon, 2020)
Network information (Ruiz, Gomes, Rodrigues, & Gama, 2017)	**Behavioral and psychological surveys** (Goel & Rastogi, 2021)	**Fund transfers datasets** (Shumovskaia, Fedyanin, Sukharev, Berestnev, & Panov, 2020; Sukharev, Shumovskaia, Fedyanin, Panov, & Berestnev, 2020)	**Psychometric data** (Rabecca, Atmaja, & Safitri, 2018; Djeundje, Crook, Calabrese, & Hamid, 2021; Rathi, Verma, Jain, Nayyar, & Thakur, 2022)

These sources share the common use of social-interaction information gathered from the graph formed by the interactions among individuals recorded in alternative data sources.

Gaps

In this thesis work, we will address three main gaps identified.

- Partial network data sources
- Scarce behavioral data
- Hand-made feature engineering
- Application Scoring VS Behavioral Scoring

Research Problem

General Objective

The main idea is to investigate the value of incorporating network data in credit risk management.

We defined three specific aims to achieve this goal. Each specific aim in this thesis is driven by its own research questions, which have been thoroughly investigated and documented in their corresponding articles. These articles have been published in peer reviewed journals and conferences, providing insights and findings for further research.

Specific Aims

Aim 1: On the combination of graph data

Research Questions:

- When combining different GRL techniques over complex graph structures, is there a performance improvement compared to merely applying hand-crafted feature engineering or graph neural networks?
- What insights are obtained into the combined network features, and what value do these insights add to credit risk assessment?
- Where does social information help the most? Is the most significant performance enhancement obtained in personal credit scoring or business credit scoring? What can we gather from this information? Does it influence which network and which features are the most relevant?

Specific Aims

Aim 2: On the dynamics of graph data features and their impact on performance

Research Questions:

- Knowing that borrowers' repayment history increases creditworthiness assessment performance, at which point in time since the loan is granted, does this information become meaningful? For how long do we need to observe borrowers' repayment history to assess their creditworthiness accurately?
- Knowing that social-interaction data contributes more value to application scoring, that is when behavioral information is scarce. For how long is it beneficial to rely on these sources of information?
- What insights and benefits to credit risk management are obtained from studying the dynamics of both the creditworthiness assessment performance and the value of alternative data sources?

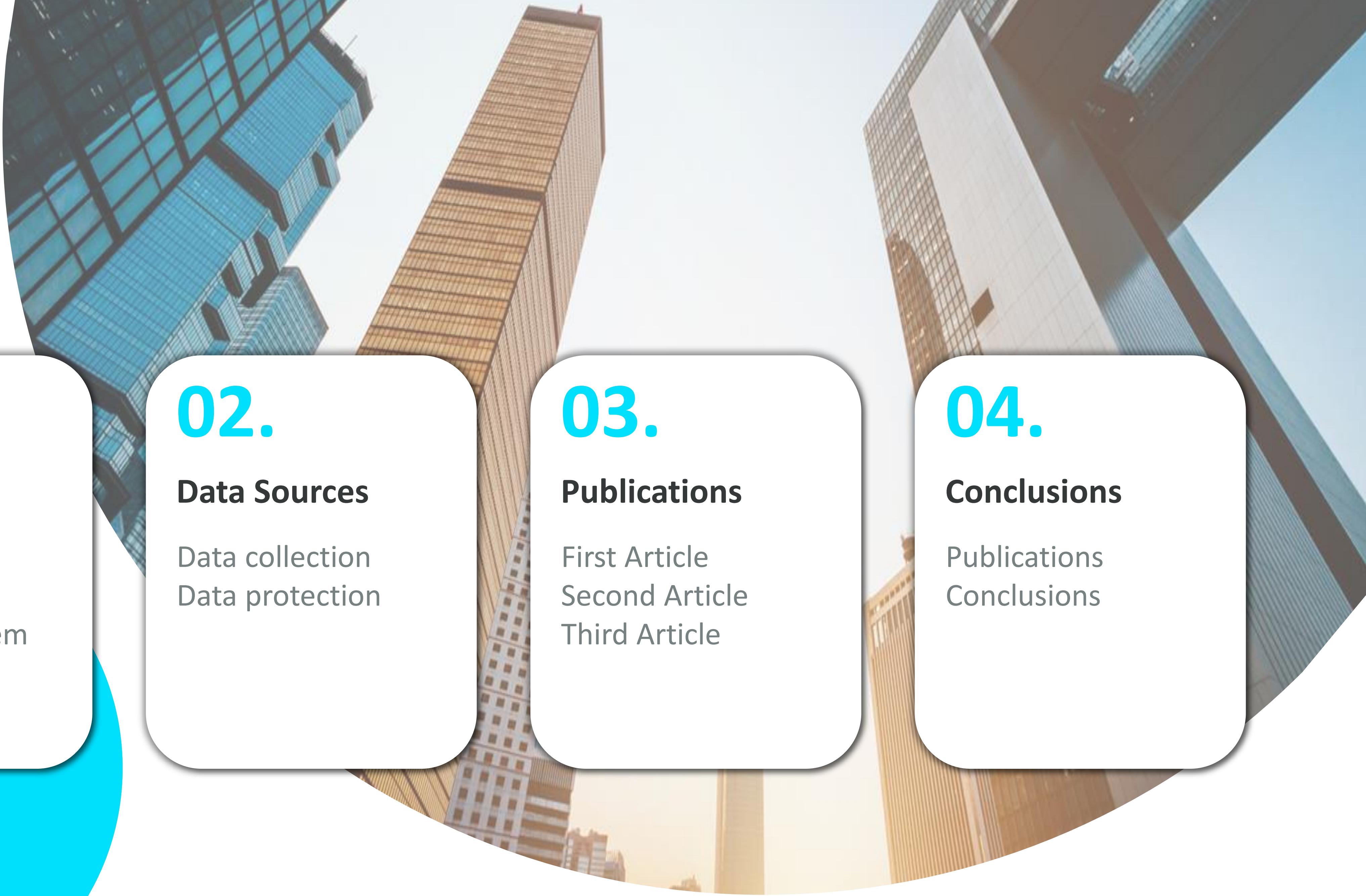
Specific Aims

Aim 3: On the training of credit scoring models using synthetic data

Research Questions:

- Can a model trained on synthetic data perform well in real-world scenarios?
- How does increasing the features impact synthetic data quality?
- Is there a performance cost for working in a privacy-preserving environment?

Contents



Data Sources

Data sources used in this research

Financial Information

7.65 million people
0.245 million firms
Jan-18 to Mar-20
+Benchmark Score

Network of marriages

5.5 million nodes
2.8 million edges
0% Fraction of business

Enterprise's ownership

2.6 million nodes
2.5 million edges
4.6% Fraction of business

Parents & Children

19.8 million nodes
27.7 million edges
0% Fraction of business

Transactional services

7.3 million nodes
23.4 million edges
4.4% Fraction of business

Employment Network

0.8 million nodes
0.8 million edges
2.9% Fraction of business

Data Protection

Privacy protection and ethical guidelines

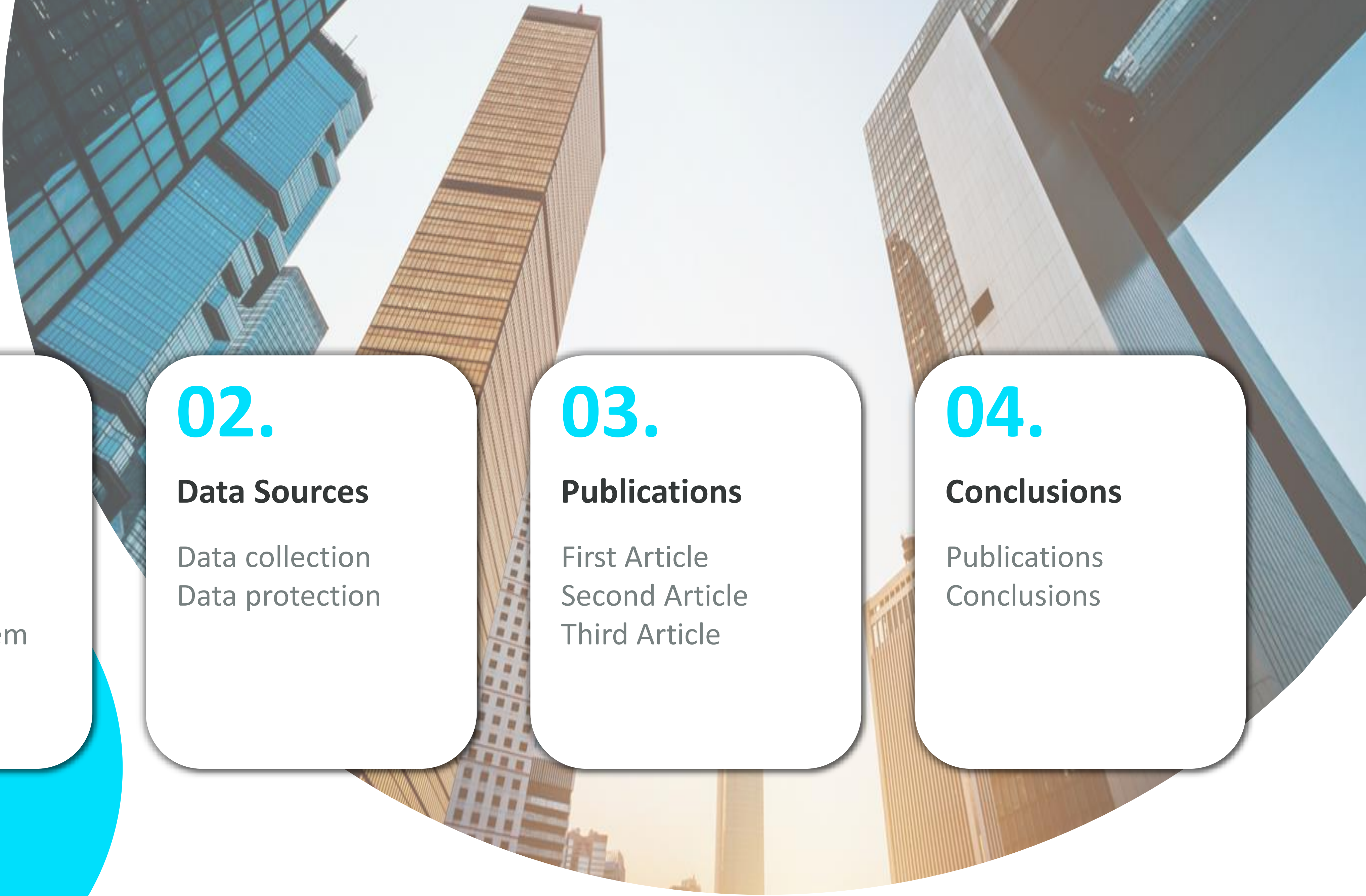
Data sources do not reveal any individual's identity or personal information.

Customer identifiers and any personal data were removed before starting the analysis

This project has the approval of the ethics committee of the University of Chile.

The final data produced by this research does not compromise customers' privacy and cannot leak any personal private information.

Contents



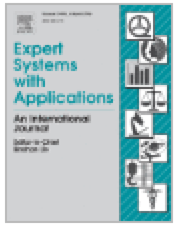
First Article

On the combination of graph data for assessing thin-file borrowers’ creditworthiness

ELSEVIER

Expert Systems with Applications

Volume 213, Part A, 1 March 2023, 118809



On the combination of graph data for assessing thin-file borrowers’ creditworthiness

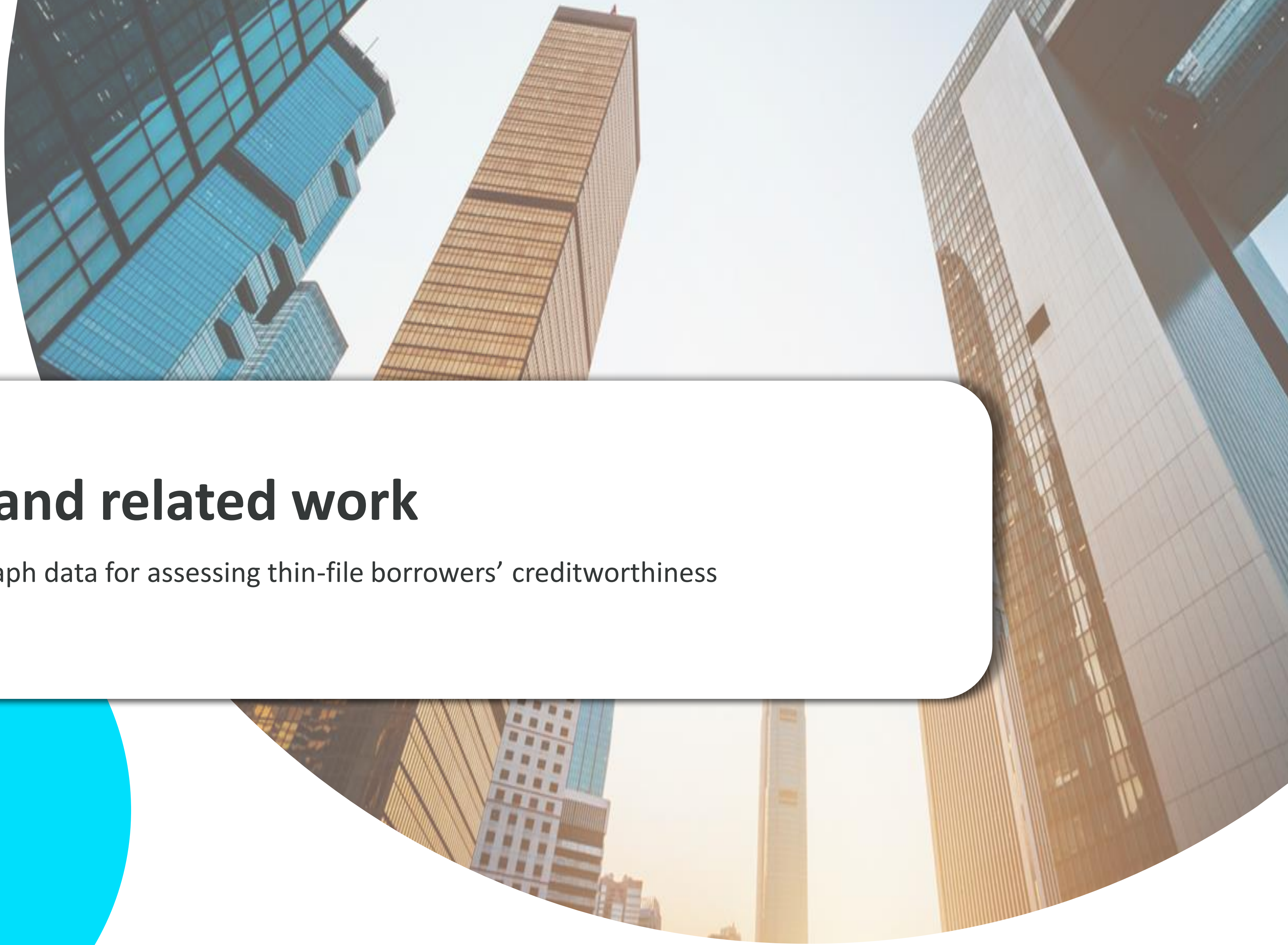
Ricardo Muñoz-Cancino [a], Cristián Bravo [b], Sebastián A. Ríos [a], Manuel Graña [c]

Research Questions

On the combination of graph data for assessing thin-file borrowers' creditworthiness

- When combining different GRL techniques over complex graph structures, is there a performance improvement compared to merely applying hand-crafted feature engineering or graph neural networks?
- What insights are obtained into the combined network features, and what value do these insights add to credit risk assessment?
- Where does social information help the most? Is the most significant performance enhancement obtained in personal credit scoring or business credit scoring? What can we gather from this information? Does it influence which network and which features are the most relevant?

First Article



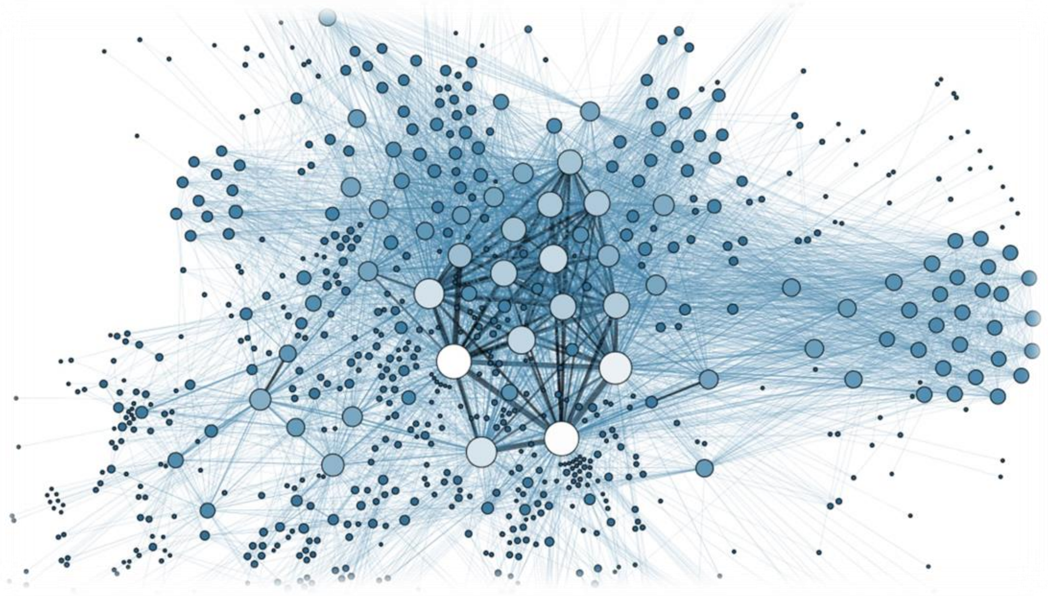
Background and related work

On the combination of graph data for assessing thin-file borrowers' creditworthiness

Social Network Analysis

Background and related work

Social network analysis foundation is that individuals are connected through complex relationships that we do not yet fully understand forming networks. Links among people enable us to comprehend how they are related, how they gather to form part of groups (Easley & Kleinberg 2010)



Social Network Analysis

Background and related work

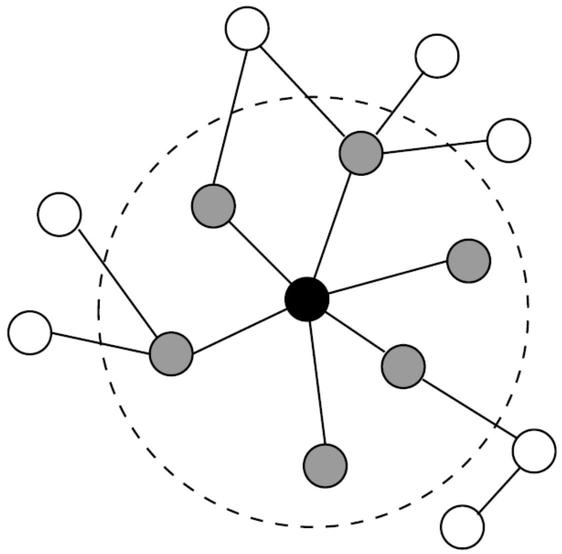
Mathematically, we describe a network by a graph $g = (V, E, A)$. V is the set of nodes. Let $V = v_1, \dots, v_N$ where N is the number of nodes, E is the set of edges, and adjacency matrix A is a $N \times N$ matrix with $A_{ij} = 1$ if there is an edge e_{ij} from v_i to v_j, $A_{ij} = 0$ otherwise. A graph can be associated with a node attributes matrix $X \in \mathbb{R}^{N \times F}$ $, where\ X_i \in \mathbb{R}^F$ represents the feature vector of node v_i. When the feature matrix X evolves over time T the graph is called a spatial-temporal graph and it is defined as $g = (V, E, A, X)$, where $X \in \mathbb{R}^{N \times F \times T}$

Representation Learning on Networks

Background and related work

- The machine learning subfield that attempts to apply techniques to graph-structured data is known as Graph Representation Learning.

- Each node has a variable number of neighbors, meaning that operations that are easy to calculate on other types of data, such as convolutions on images, cannot be applied directly to graphs.



Representation Learning on Networks

Background and related work

Researchers have proposed many methodologies to extract knowledge for networks, here we present a nomenclature and characteristics for the primary methods.

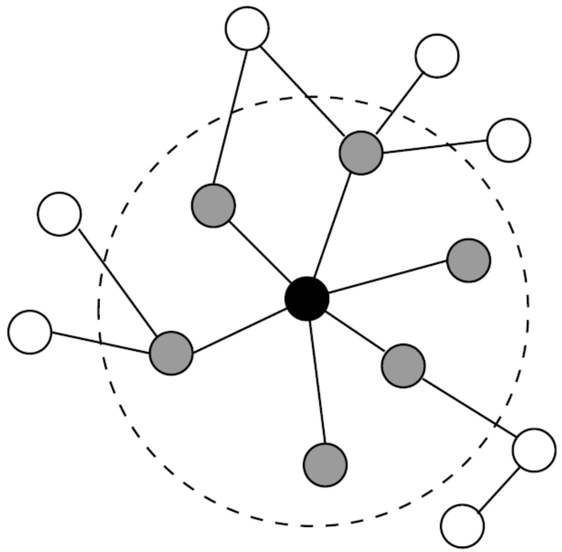
- Feature Engineering
- Network Embedding
    - Matrix Factorization
    - Random Walks
- Graph Neural Networks
    - Recurrent graph neural networks
    - Convolutional neural networks
    - Graph autoencoders
    - Spatial-temporal graph neural networks

Feature Engineering

Background and related work

Consists of generating attributes from the raw data to later be used in training predictive models

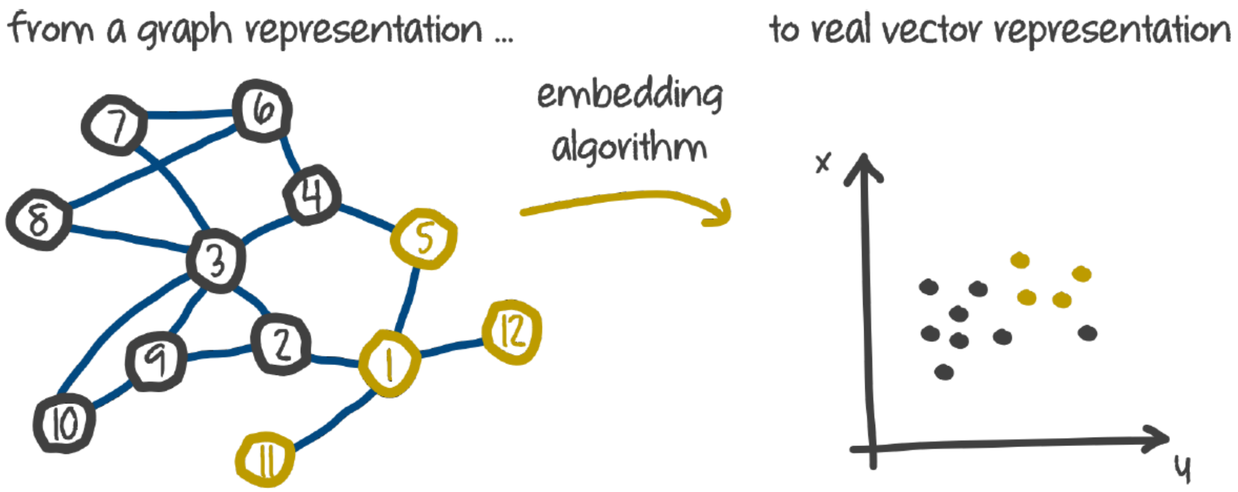
- This process is manual, iterative, time-consuming, and involves a series of steps from conceptualization, implementation, and evaluation.
- Requires domain-expertise and creativity.
- Art and Science



Network Embedding

Background and related work

Network embedding methods are mostly unsupervised learning techniques aiming to learn a low-dimensional representation of each node in Euclidean space by optimizing an objective function that measures the similarity among these representations.
Formally, a node embedding is a function $f\colon G(V, E, A) \rightarrow \mathbb{R}^d$ that maps each node v to a feature vector (embedding) $\{Z_v\}_{v \in V} \in \mathbb{R}$^$d$



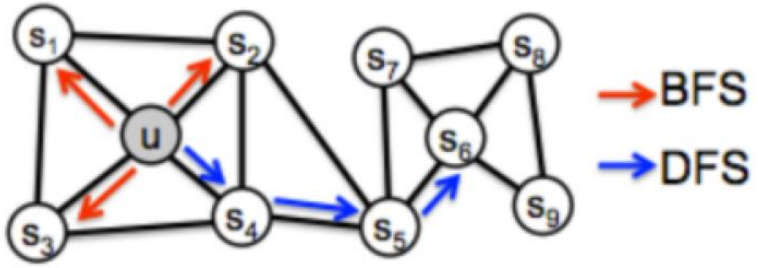
Network embedding

Network Embedding – Node2Vec

Background and related work

This approach addresses a neighborhood vector's sparseness nature by characterizing a node's neighborhood through the co-occurrence rate in a random walk. It is inspired by the fact that the distribution of nodes that appear in random walks is similar to the distribution of words in a natural language problem. In this way, the Skip-Gram model is adapted to represent nodes instead of words.

- **Intuition:** Optimize embeddings to maximize likelihood of random walk co-occurrences
- $\mathcal{L} = \sum_{u \in V} \sum_{v \in N_R(u)} -\log\left(\frac{\exp(z_u^T z_v)}{\sum_{z \in \mathrm{V}} \exp(z_u^T z_n)}\right)$
- $\log(\phi) \rightarrow \phi$: predicted probability of u and v co-occuring on random walk



Node2Vec

(Grover & Leskovec, 2016)

GNN - Convolutional neural networks (ConvGNNs)

Background and related work

ConvGNNs generalizes the convolution operation to be applied in networks. In this case, the purpose is to produce a node's representation Z_v by adding its attributes X_v and neighbors $\{X_u\}_{u\in N(v)}$

- **Spectral-based ConvGNNs**: Mathematical foundation inherited from graph signal processing. This theory is based on the properties of the Laplacian matrix ($L = \mathrm{I}_n - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$). The graph convolution of feature vector X_i with filter $g \in \mathbb{R}^N$ is represented as $X_i * g = \mathcal{F}^{-1}(\mathcal{F}(X_i) \odot \mathcal{F}(g))$. The ChebNet employs Chebyshev polynomial of eigenvalues' diagonal matrix as a filter.

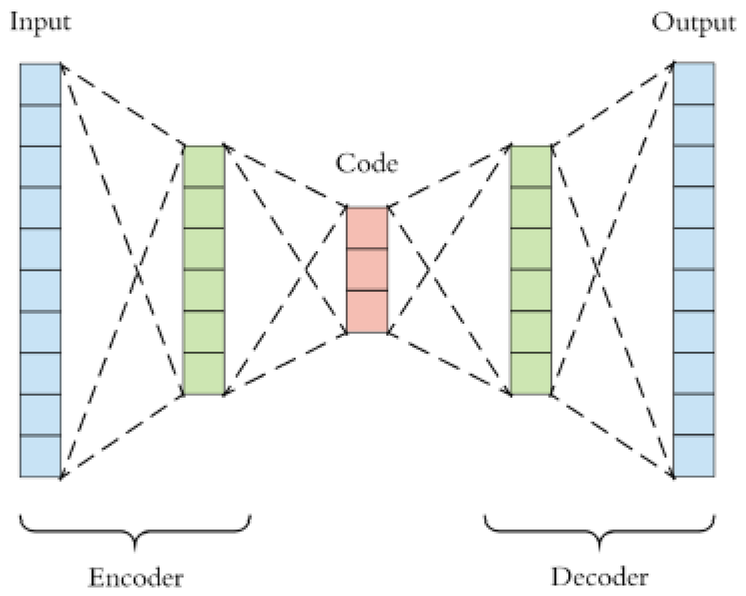
- **Spatial-based ConvGNNs**: Spatial-based methods represent the graph convolution based on the spatial relationships of a node. GraphSage displays a fixed number of neighbors for each node. In this way, the graph convolution is defined as $h_v^k = \sigma(A^k \cdot f_k(h_v^{k-1}, \{h_u^{k-1}, \forall u \in S_{N(v)}\}))$, where $h_v^0 = X_v, f_k(\cdot)$ is an aggregation function. $S_{N(v)}$ is a random sample of the node v neighborhood.

GNN - Graph Autoencoders (GAEs)

Background and related work

GAEs is an unsupervised learning framework that encodes the nodes in a low-dimensional representation and then reconstructs the original network data from the encoded information.

GAE calculates the network embedding matrix Z and the reconstruction of the adjacency matrix A' as follows:
$A' = \sigma(ZZ^T)$ with $Z = GCN(X, A)$ where X is the node attributes matrix.



AutoEncoder
(towardsdatascience.com)

First Article

Data Description & Preprocessing

On the combination of graph data for assessing thin-file borrowers' creditworthiness

Financial & Social-Interaction Data

Data Description and Preprocessing

The information used in this research originates from a Latin American bank. The information provided by the financial institution to create networks originates from varied sources and can be cataloged as follows

- **WeddNet** - Network of marriages
- **TrxSNet** - Transactional services Network
- **EnOwNet** - Enterprise's ownership Network
- **PChNet** - Parents & Children Network
- **EmpNet** - Employment Network

Financial Data

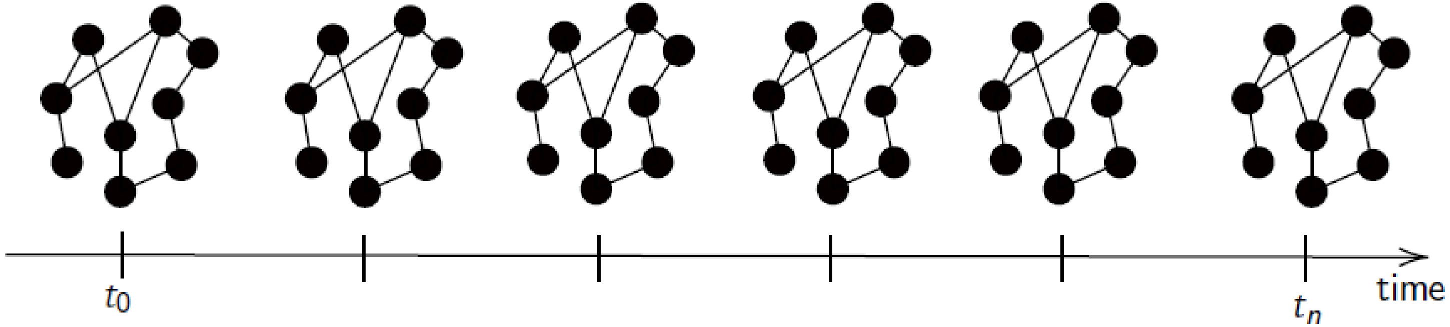
The node dataset contains information on the consolidated indebtedness of each debtor in the financial system from January 2018 until March 2020, reporting monthly the debt decomposition from 7.65 million people and 245,000 firms.

Network Construction

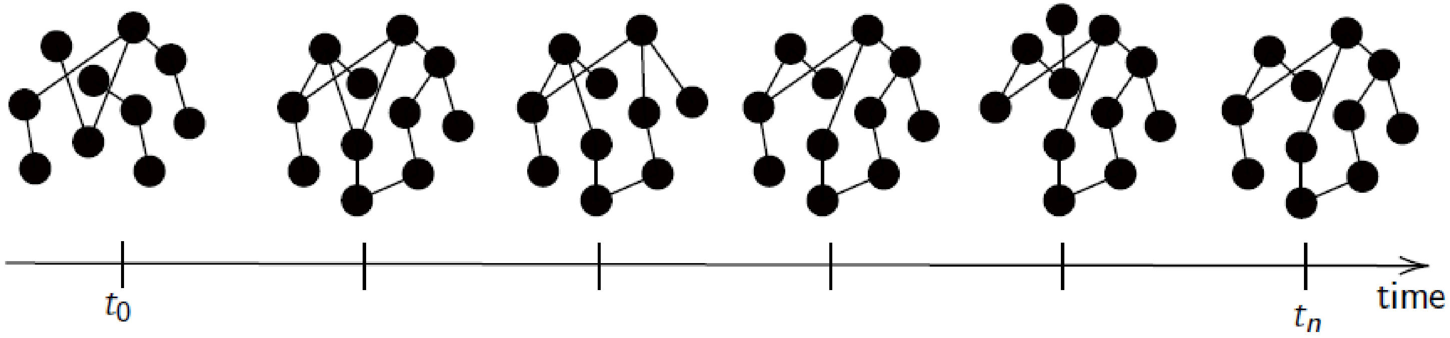
Data Description and Preprocessing

All the data sources can be represented by a network. However, not all of them have the characteristics of social networks. To manage this, we combine them into two primary data sources. From them, we develop networks that characterize people and businesses.

[FamilyNet] Family Network: WeddNet and PChNet.



[EOWNet] Enterprise's ownership Network and Workers: TrxSNet, EnOwNet and EmpNet.



First Article

Experimental Design and Methodology

On the combination of graph data for assessing thin-file borrowers' creditworthiness

Datasets

Experimental Design and Methodology

- Credit scoring models are built with information about the financial system for 24 months.
- However, the models are trained over **23 months** to avoid target leakage
- For the **unbanked application scoring model**, individuals and companies are considered only in the month that they enter the financial system.
- For the **behavioral scoring** model individuals and companies are considered six or more months after entering the financial system
- **Target:** A person or company is considered defaulter when it is 90 or more days past due within twelve months from observing him.

Scoring application	Model	Observations	# Features
Unbanked Application Scoring	Business Credit Score	29,044	687
	Personal Credit Score	192,942	1,283
Behavioral Scoring	Business Credit Score	931,910	687
	Personal Credit Score	1,978,664	1,283

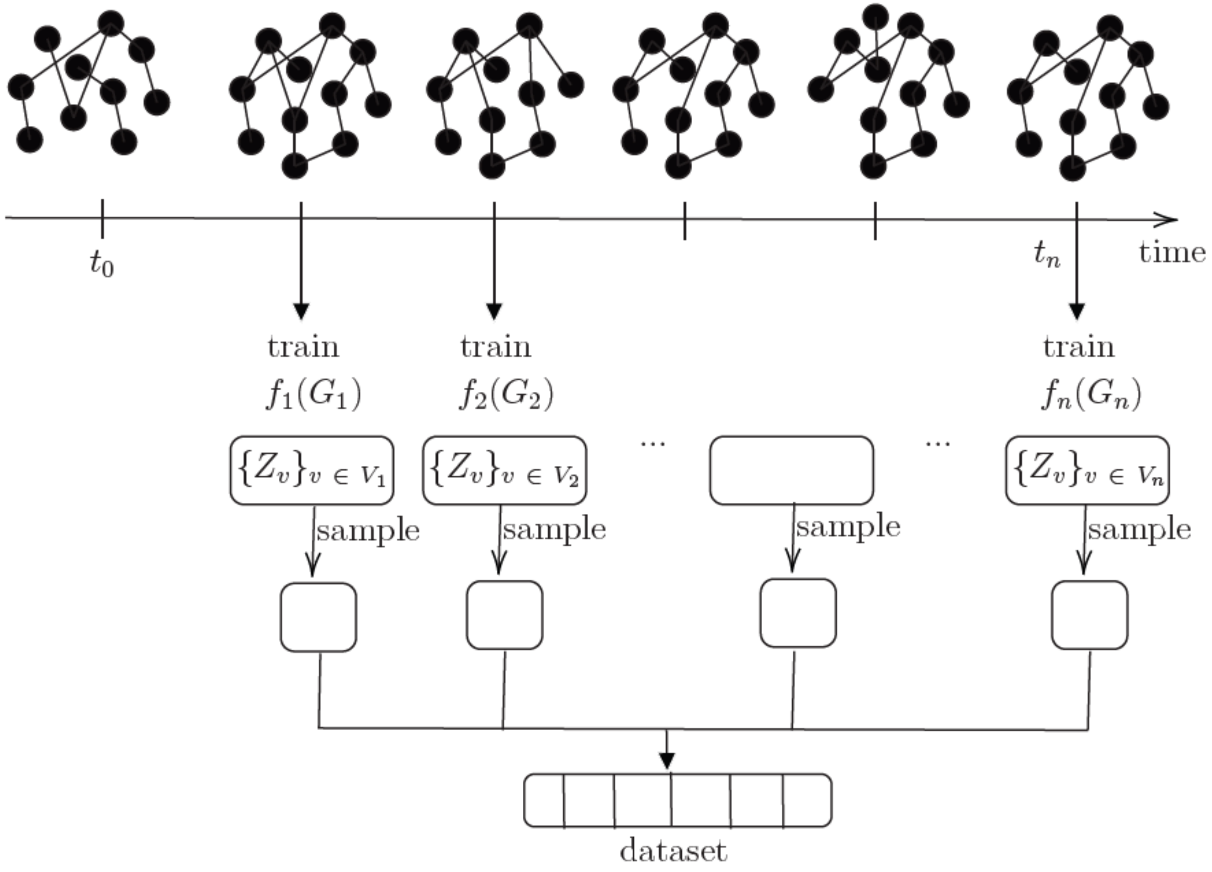
Traditional and Graph Representation Learning Features

Experimental Design and Methodology

- **[Node] Financial Information**: This dataset contains information from January 2018 until March 2020, and it monthly reports debt decomposition from 7,65 million people and 245,000 firms.
- **[NodeBench] Benchmark Score**: The BenchScore corresponds to the probability of default for the coming 12 months. This probability was assessed and provided by the financial institution.
- **[NodeStats] Node Statistics:** Correspond to nodes' statistics based on their position and characteristics within the network (Degree Centrality, number of Triads, PageRank Score, authority and hub score, and whether the node is an articulation point flag).
- **[EgoNet] EgoNetwork Agreggation and EgoNet Weighted Aggregation Features**
- [**N2V] Node2Vec Features:** Each node is characterized by a feature vector obtained applying Node2Vec.
- **[GNN] Graph Neural Network Features:** Each node is characterized by several feature vectors obtained applying a group of Graph Neural Networks Techniques (Graph Convolutional Networks and Graph Autoencoders)

Node2Vec Features

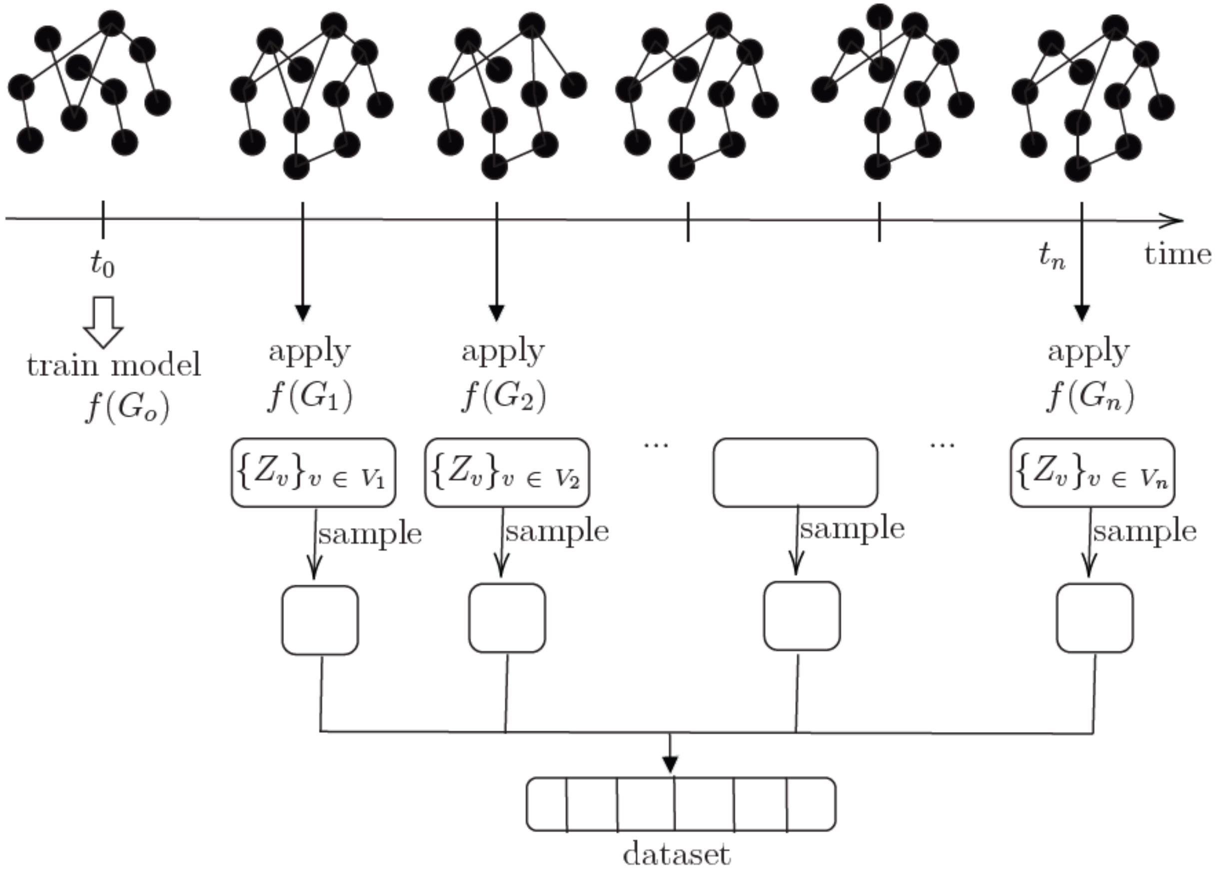
Experimental Design and Methodology



Node2Vec to Features

GCN and GAE Features

Experimental Design and Methodology



Graph Convolutional Networks and Graph Autoencoders to Features

Features Subsets

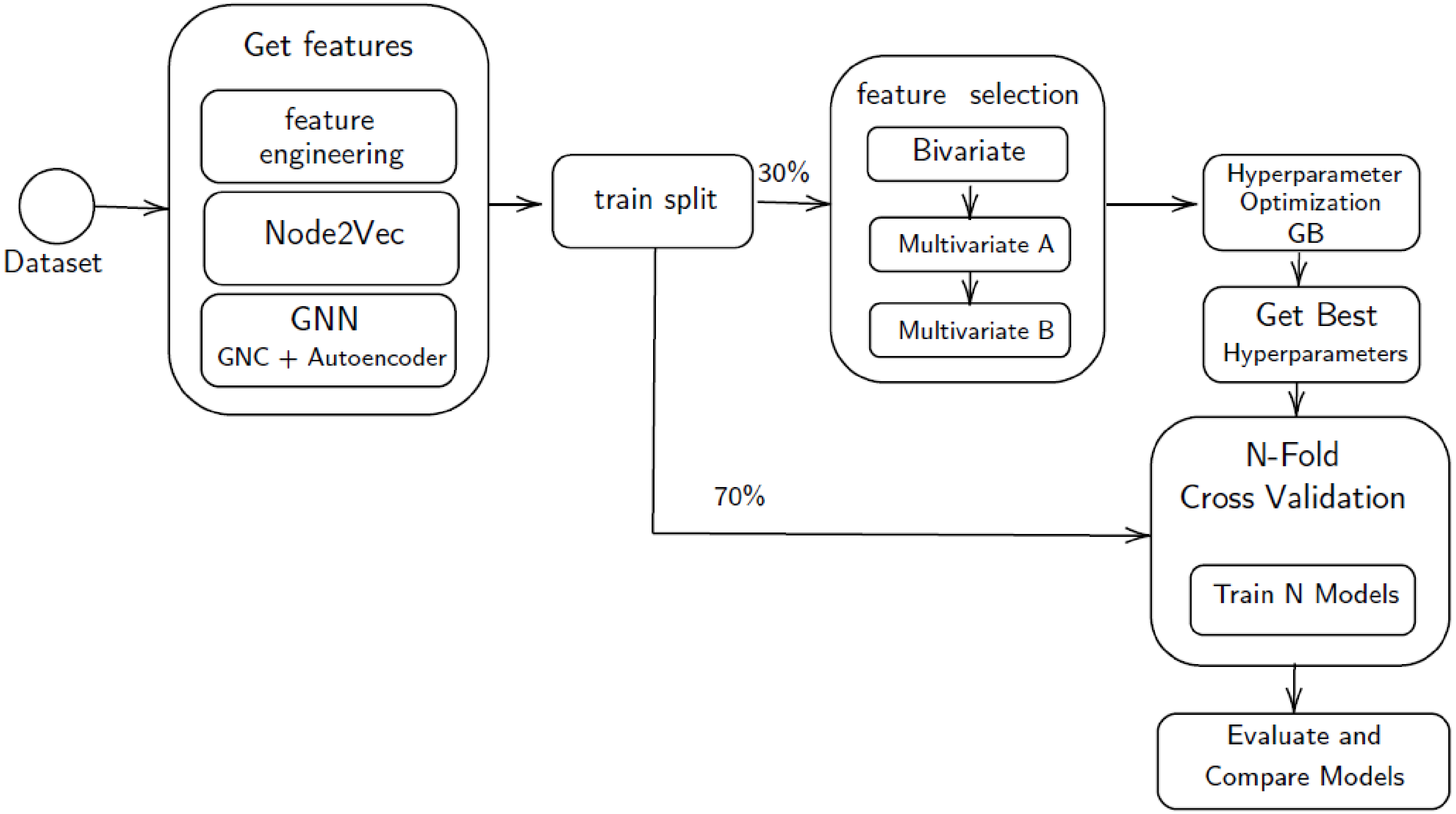
Experimental Design and Methodology

Experiment Id	Feature Group
A	$X = \{X_{Node}\}$
A+B	$X = \{X_{Node} + X_{BenchScore}\}$
A+B+C	$X = \{X_{Node} + X_{BenchScore} + X_{NodeStats}\}$
A+B+D	$X = \{X_{Node} + X_{BenchScore} + X_{EgoNet}\}$
A+B+E	$X = \{X_{Node} + X_{BenchScore} + X_{GNN,N2V}\}$
A+B+C+D	$X = \{X_{Node} + X_{BenchScore} + X_{NodeStats} + X_{EgoNet}\}$
A+B+C+E	$X = \{X_{Node} + X_{BenchScore} + X_{NodeStats} + X_{GNN,N2V}\}$
A+B+C+D+E	$X = \{X_{Node} + X_{BenchScore} + X_{NodeStats} + X_{EgoNet} + X_{GNN,N2V}\}$

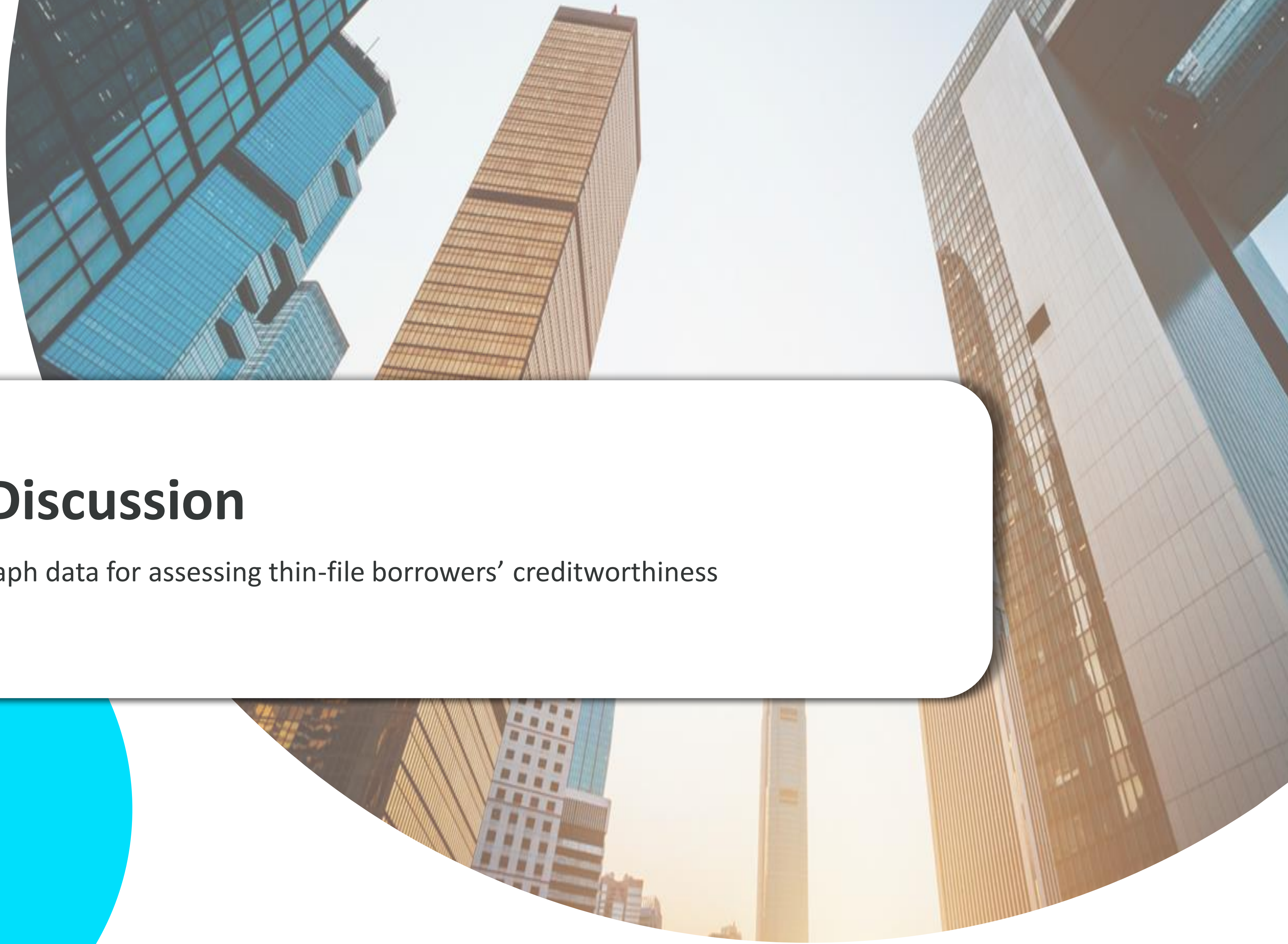
Experiments Setup

Methodology

Experimental Design and Methodology



First Article



Results and Discussion

On the combination of graph data for assessing thin-file borrowers’ creditworthiness

Implementations Details

Software:

Networkx v2.6.3, Snap-stanford v5.0.0, PyTorch v1.6.0 and PyTorch Geometric v2.0.1

Hardware:

- Laptop with 8 CPU cores Intel i7 and 32 GB of RAM for network construction and hand-crafted feature engineering.
- Server with a driver node with 140 GB of RAM and 20 CPU cores and between 2 and 8 auto-scaling worker nodes with 112Gb of RAM and 16 CPU cores for the Node2Vec, GCN, GAE, and model training phase

Execution Time

Results and Discussion

[NodeStats] Node Statistics: **625** minutes $(25m \times 25)$.

[EgoNet] EgoNetwork Agreggation: **300** minutes.

[N2V] Node2Vec Features: **7,500** minutes $(300m \times 25)$.

[GNN] Graph Neural Network Features: **6,920** minutes (train + apply: 3,320 + 3,600).

Gradient Boosting Training: **160** minutes $(40m \times 4)$.

The total execution time of our methodology is **15,500** minutes.

Model Performance Using Traditional Features

Results and Discussion

Improvement in **AUC** relative to the benchmark model (mean and std). We only report results when the equal performance hypothesis is rejected, with a confidence level of 95%; otherwise, we display *. The best performance in each column is shown in bold; more than one bold value indicates that the hypothesis of equal performance between those models cannot be rejected.

Feature Set	Business Credit Score		Personal Credit Score	
	Application	Behavioral	Application	Behavioral
A	-3.52% ± 2.87%	-0.90% ± 0.21%	-0.74% ± 0.63%	-0.63% ± 0.09%
A+B	*	0.58% ± 0.06%	1.45% ± 0.39%	0.95% ± 0.06%
A+B+C	*	1.13% ± 0.12%	2.02% ± 0.49%	1.07% ± 0.06%
A+B+D	**8.96%** ± 3.37%	2.33% ± 0.15%	2.31% ± 0.64%	1.25% ± 0.08%
A+B+E	3.92% ± 2.03%	1.77% ± 0.13%	3.17% ± 0.55%	1.96% ± 0.04%
A+B+C+D	**9.00%** ± 3.47%	2.37% ± 0.16%	2.39% ± 0.60%	1.32% ± 0.08%
A+B+C+E	4.25% ± 1.84%	1.94% ± 0.16%	3.26% ± 0.48%	2.03% ± 0.05%
A+B+C+D+E	**8.43%** ± 2.83%	**2.80%** ± 0.16%	**3.58%** ± 0.61%	**2.18%** ± 0.04%

Model Performance Using Traditional Features

Results and Discussion

Improvement in **KS** relative to the benchmark model (mean and std). We only report results when the equal performance hypothesis is rejected, with a confidence level of 95%; otherwise, we display *. The best performance in each column is shown in bold; more than one bold value indicates that the hypothesis of equal performance between those models cannot be rejected.

Feature Set	Business Credit Score		Personal Credit Score	
	Application	Behavioral	Application	Behavioral
A	*	-4.15% ± 0.94%	-5.25% ± 2.40%	-2.39% ± 0.46%
A+B	*	1.56% ± 0.40%	4.38% ± 1.19%	1.95% ± 0.35%
A+B+C	*	3.21% ± 0.71%	6.27% ± 1.02%	2.23% ± 0.39%
A+B+D	**20.69%** ± 16.73%	7.69% ± 0.92%	6.79% ± 1.36%	2.69% ± 0.47%
A+B+E	**12.22%** ± 10.89%	5.83% ± 0.74%	8.64% ± 2.13%	4.68% ± 0.28%
A+B+C+D	**21.28%** ± 17.10%	8.09% ± 0.95%	7.12% ± 1.52%	2.83% ± 0.52%
A+B+C+E	**12.88%** ± 10.11%	6.33% ± 0.70%	8.93% ± 1.98%	4.93% ± 0.26%
A+B+C+D+E	**19.32%** ± 14.77%	**9.45%** ± 0.85%	**10.83%** ± 1.98%	**5.15%** ± 0.42%

The Advantages of GRL Blending

Results and Discussion

Blended Graph Representation Learning performance. The performance enhancement of training a model using all GRL methods (A+B+C+D+E) is measured as the relative increase in AUC.

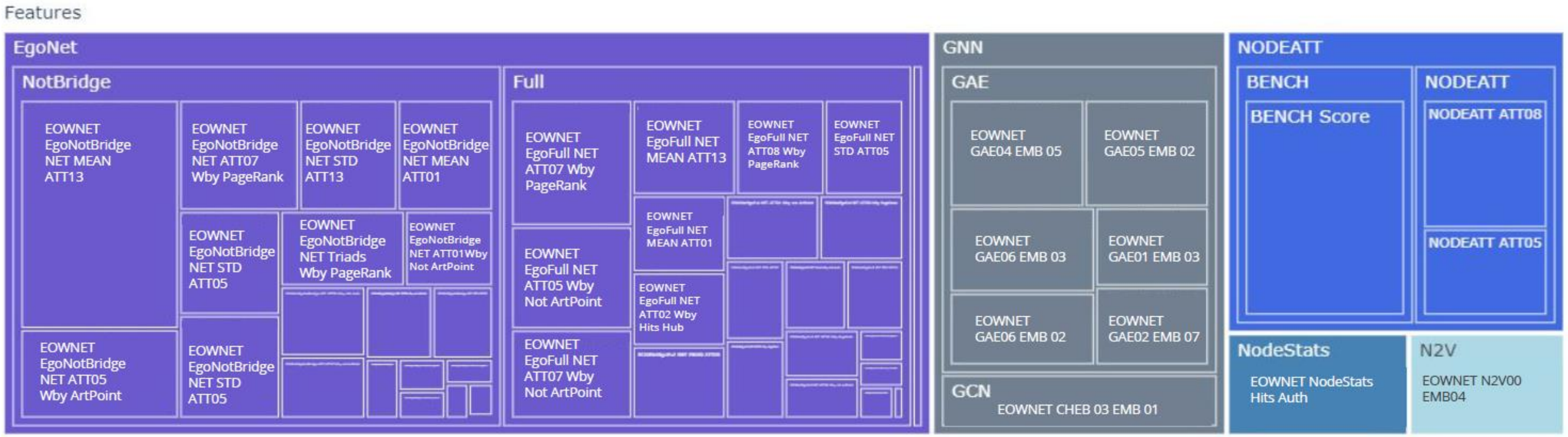
Scoring	Model	Feature Set			
		A+B+D	A+B+C+D	A+B+E	A+B+C+E
Application Scoring	Business Credit Score	*	*	4.33%	4.00%
	Personal Credit Score	1.23%	1.16%	0.39%	0.31%
Behavioral Scoring	Business Credit Score	0.47%	0.43%	1.02%	0.85%
	Personal Credit Score	0.92%	0.84%	0.22%	0.15%

Blended Graph Representation Learning performance. The performance enhancement of training a model using all GRL methods (A+B+C+D+E) is measured as the relative increase in KS.

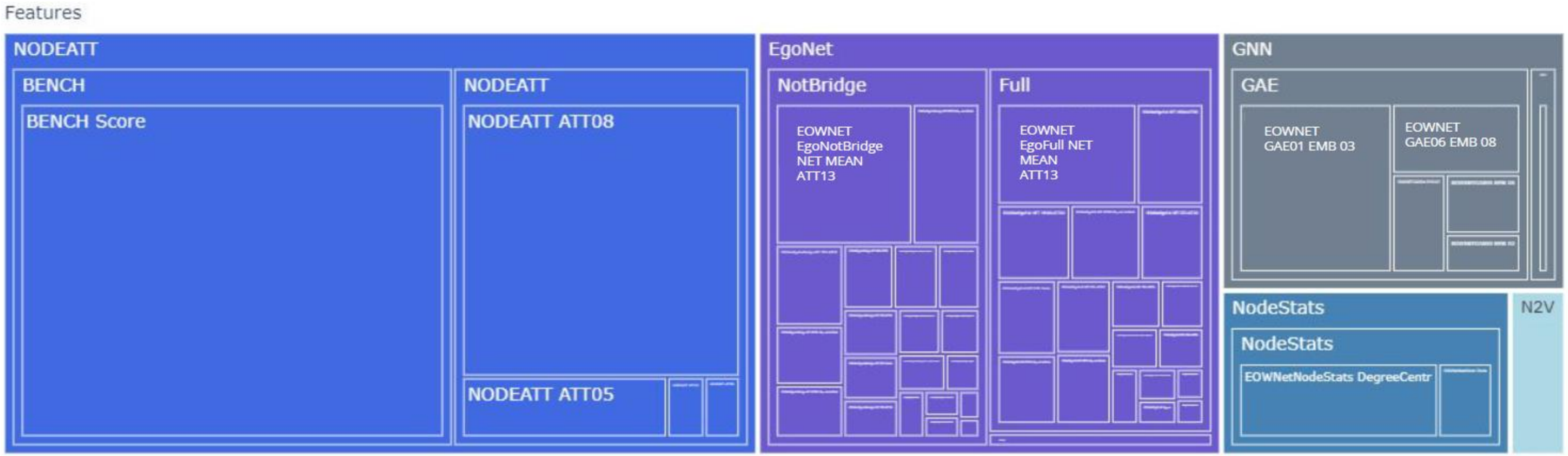
Scoring	Model	Feature Set			
		A+B+D	A+B+C+D	A+B+E	A+B+C+E
Application Scoring	Business Credit Score	*	*	*	*
	Personal Credit Score	3.79%	3.47%	2.02%	1.75%
Behavioral Scoring	Business Credit Score	1.68%	1.31%	3.47%	2.99%
	Personal Credit Score	2.40%	2.26%	0.45%	0.21%

Business Credit Scoring: Treemap of Feature Importance

Results and Discussion



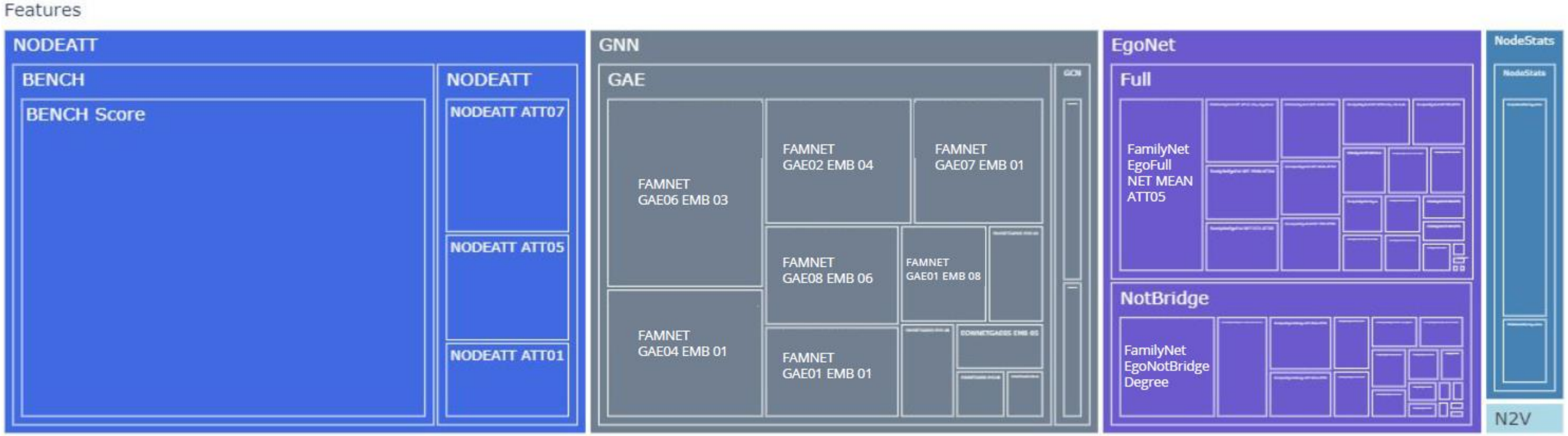
Application Scoring



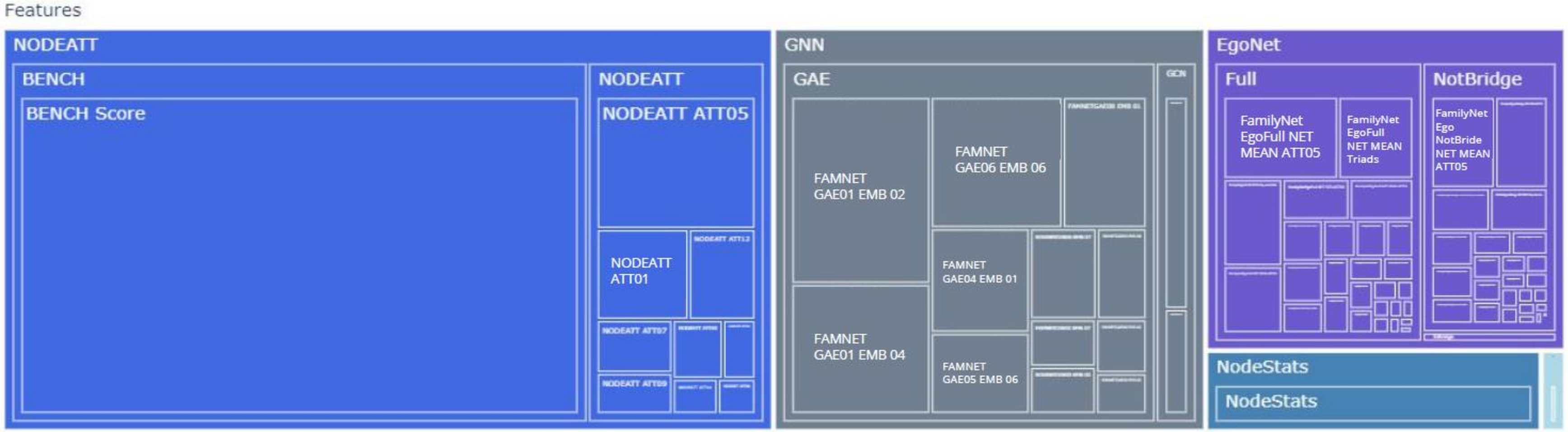
Behavioral Scoring

Personal Credit Scoring: Treemap of Feature Importance

Results and Discussion



Application Scoring



Behavioral Scoring

Contributions

On the combination of graph data for assessing thin-file borrowers' creditworthiness

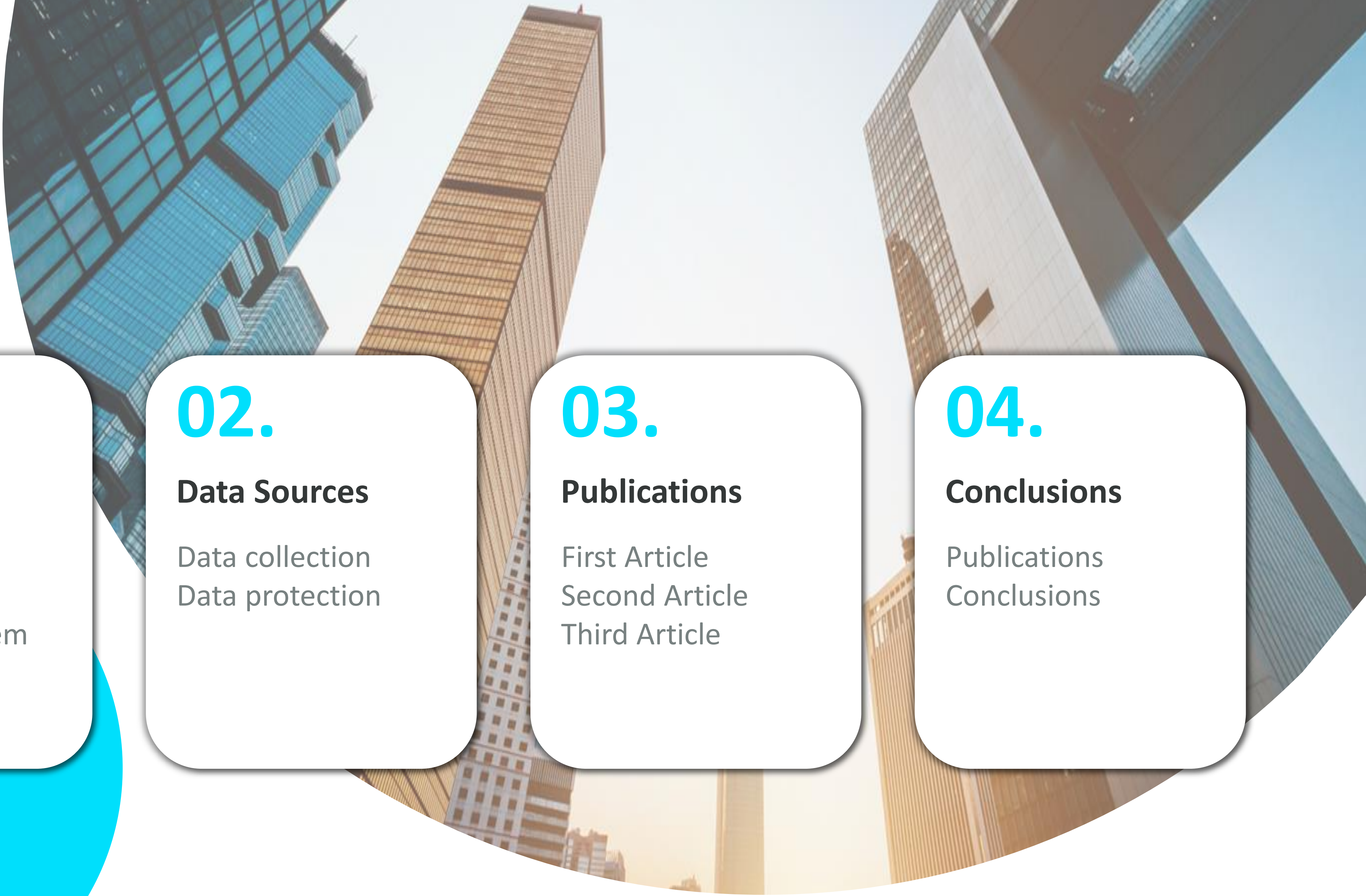
- Our framework combines hand-engineered features, graph embeddings, and GNN attributes into a single score. This helps loan granting decisions, simplifying whether to approve or reject a credit.

- Our findings validate graph data's role in corporate and consumer lending. They demonstrate varying impacts on borrowers—individuals or companies—reflected in predictive power and relevant features. This guides us on when to employ social-interaction data and anticipate its effects on creditworthiness prediction.

- This study examines a whole country's credit behavior alongside extensive social networks. These networks encompass various relationships—parental, marital, business, employer-employee, and transactions.

- This paper also contributes to the growing literature in credit scoring and network data, proposing a mechanism to achieve better results than the popular hand-crafted feature engineering and the novel GNN approach.

Contents



Second Article

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

ELSEVIER

Expert Systems with Applications

Volume 218, 15 May 2023, 119599

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

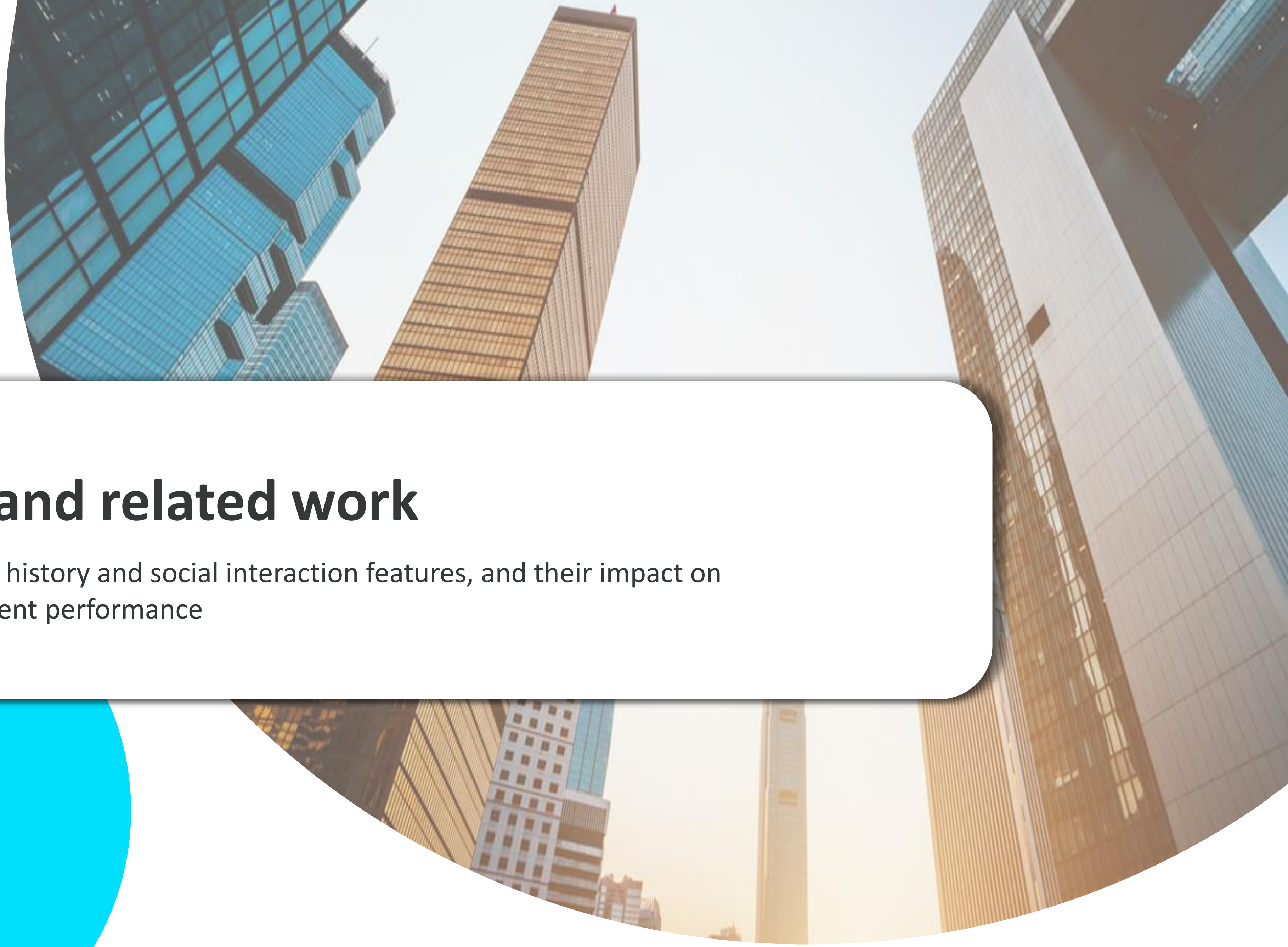
Ricardo Muñoz-Cancino [a], Cristián Bravo [b], Sebastián A. Ríos [a], Manuel Graña [c]

Research Questions

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

- Knowing that borrowers' repayment history increases creditworthiness assessment performance, at which point in time since the loan is granted, does this information become meaningful? For how long do we need to observe borrowers' repayment history to assess their creditworthiness accurately?
- Knowing that social-interaction data contributes more value to application scoring, that is when behavioral information is scarce. For how long is it beneficial to rely on these sources of information?
- What insights and benefits to credit risk management are obtained from studying the dynamics of both the creditworthiness assessment performance and the value of alternative data sources?

Second Article



Background and related work

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

Literature review

Background and related work

Authors	Entry	Performance Period (Months)	Conclusions on the Performance Period	Datasets	Algorithms	Methodology	Alternative Data
(Kennedy et al., 2013)	Journal Article	6, 12, 18	The 12-month window achieved the best performance but was limited to some scenarios	2,500 customers from the Irish Credit Bureau	LR	Train Test Split & Bootstrapping	
(Neto, Jorge Adeodato, & Carolina Salgado, 2017)	Journal Article	6, 12, 24		Two datasets: 682 records from the PKDD1999 Challenge and 30,000 customers from Brazilian retail	MLP, ANN, RF, KNN	K-Fold Cross Validation	
(Nikolaidis et al., 2017)	Incollection	1, 3, 6, 12	The 12-month window achieved the best performance but only slightly better	20,000 borrowers and their 86,082 credit lines	LR, SVM	K-Fold Cross Validation	
(Ruiz et al., 2017)	Inproceedings	12		Two datasets: first loans and all loans granted	LR, SVM	K-Fold Cross Validation	Mobile phone network usage
(Óskarsdóttir et al., 2019)	Journal Article	1, 3		22,000 observations. Customers from a telecommunications operator and a commercial bank	LR, DT, RF	Train Test Split	Mobile phone and graph data
(Djeundje et al., 2021)	Journal Article	3,6		Two datasets: 1,826 records from Mexico and 16,358 from Nigeria; both were supplied by Lenddo	LR, ANN, GB	Train Test Split	Psychometric data and customer's email activity
(Kyeong, Kim, & Shin, 2022)	Journal Article	6		200,000 records from KakaoBank in Korea	LR	Train Test Split & Bootstrapping	Log data recorded by the mobile application
(L. C. Thomas, 2000)	Journal Article	12	The 12-month window is used as an example				
(Liu, 2001)	Tech Report	12	The 12-month window is used as an example				
(Siddiqi, 2012)	Book	6, 12	(Siddiqi, 2012) stated, "For behavior scorecard development, accounts are chosen at one point in time, and their behavior analyzed over, typically, a 6- or 12-month period."				
(Bhalla, 2016)	Blog Entry	1	(Bhalla, 2016) stated that "No fixed window for all the models. Depends on the type of model."				
(Mashanovich, 2017)	Blog Entry	12	(Mashanovich, 2017) stated that "The length of the observation and performance windows will depend on the industry sector for which the model is being designed."				

Literature review of performance window analysis for behavioral credit scoring

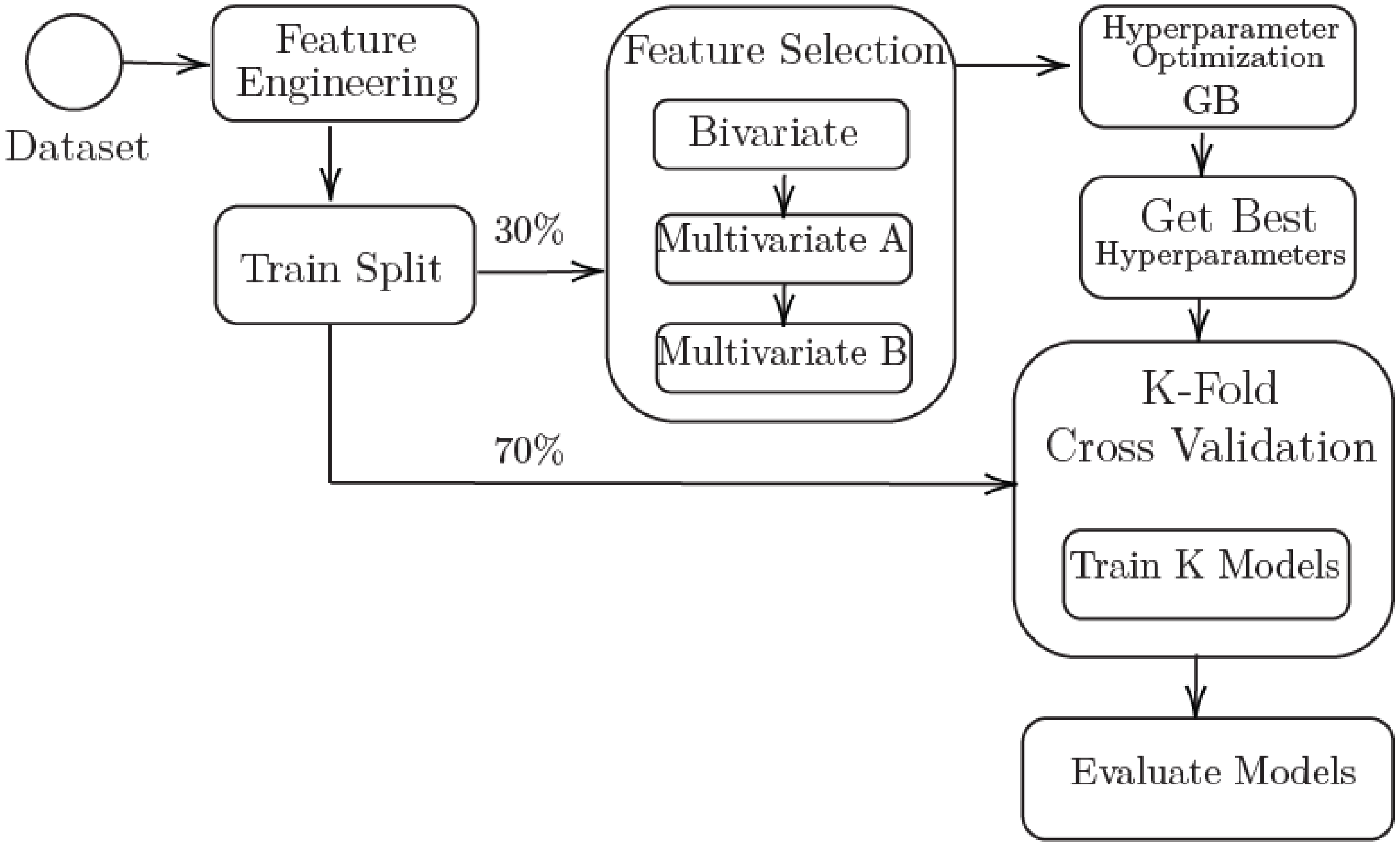
Second Article

Experimental Design and Methodology

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

Methodology

Experimental Design and Methodology

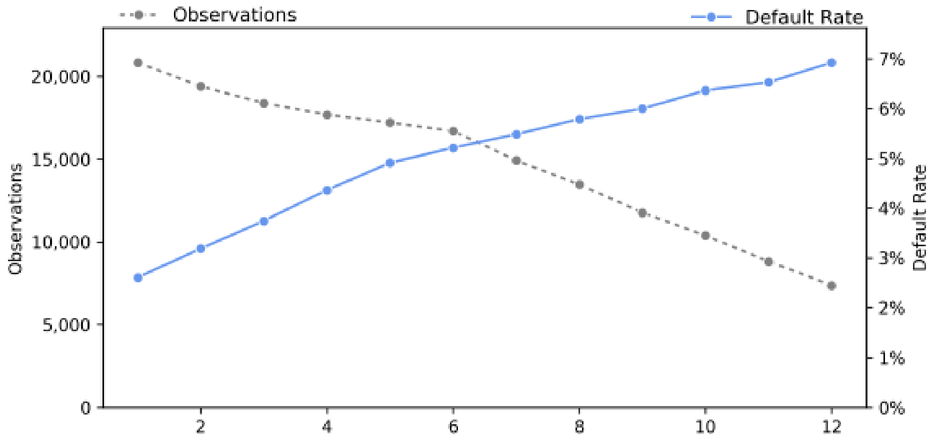


Dataset

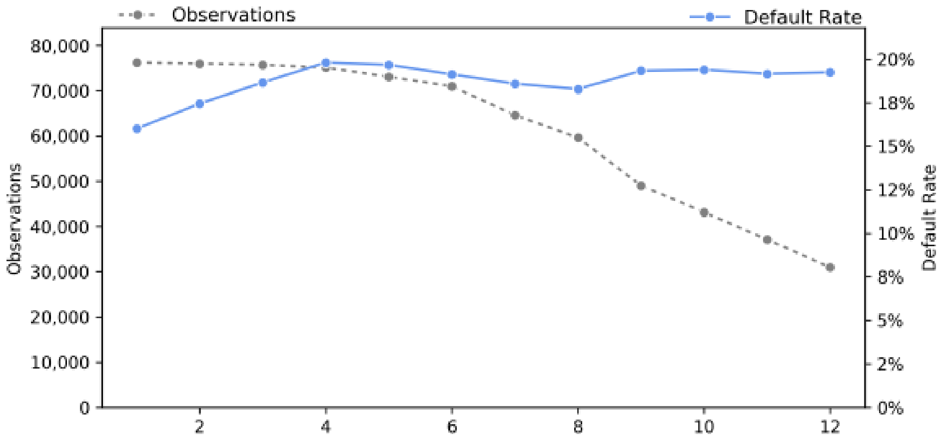
Experimental Design and Methodology

Dataset description. Borrowers correspond to the total number of individuals and companies that are part of our analysis, which will be observed from the moment they obtain a loan until 12 months later.

Model	Borrowers	Features
Business Credit Score	20,835	585
Personal Credit Score	76,209	936



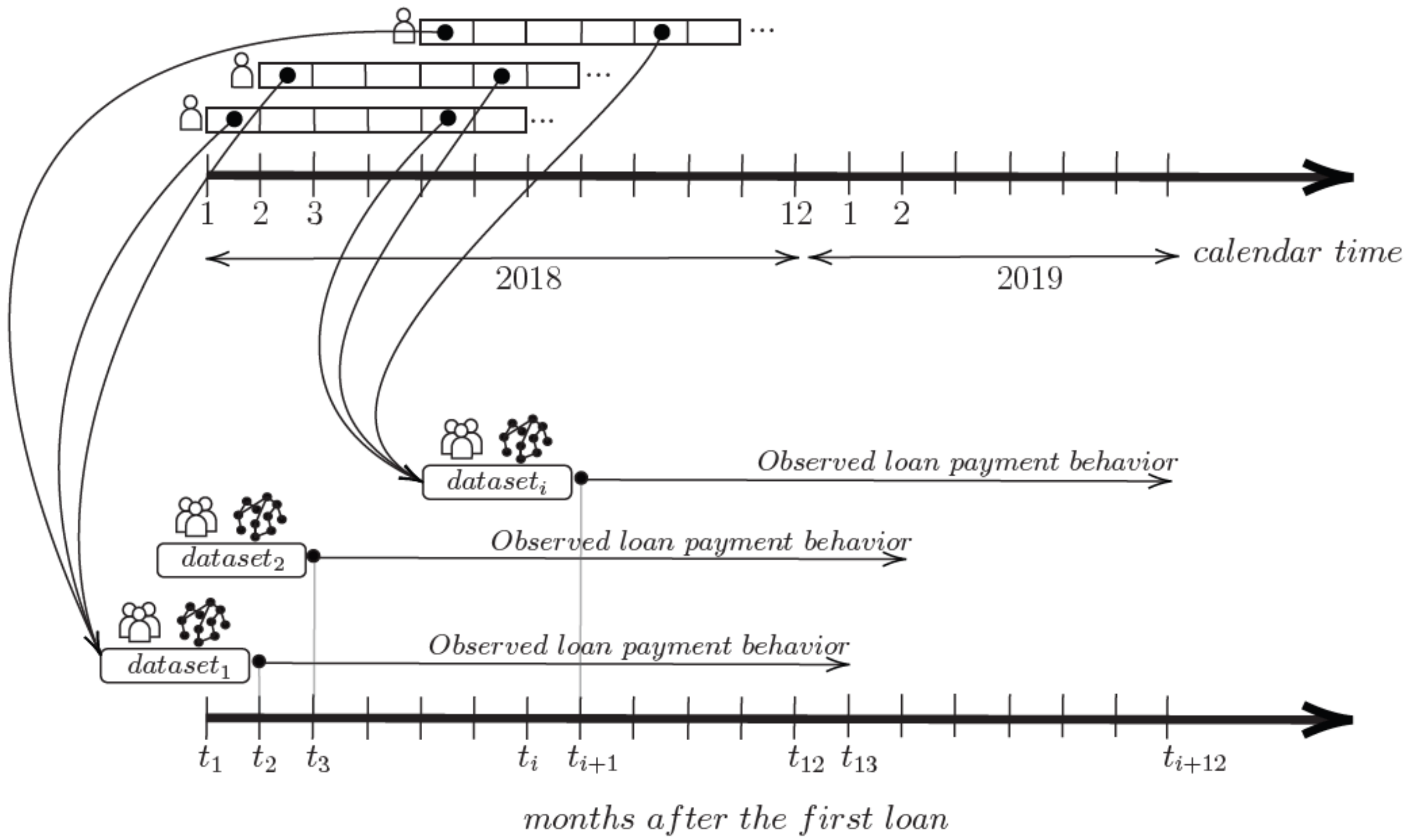
Business Scoring



Personal Scoring

Dataset

Experimental Design and Methodology



Dataset construction. Upper timeline corresponds to calendar dates. Lower timeline corresponds to the relative time from first loan granting.

Experiments

Experimental Design and Methodology

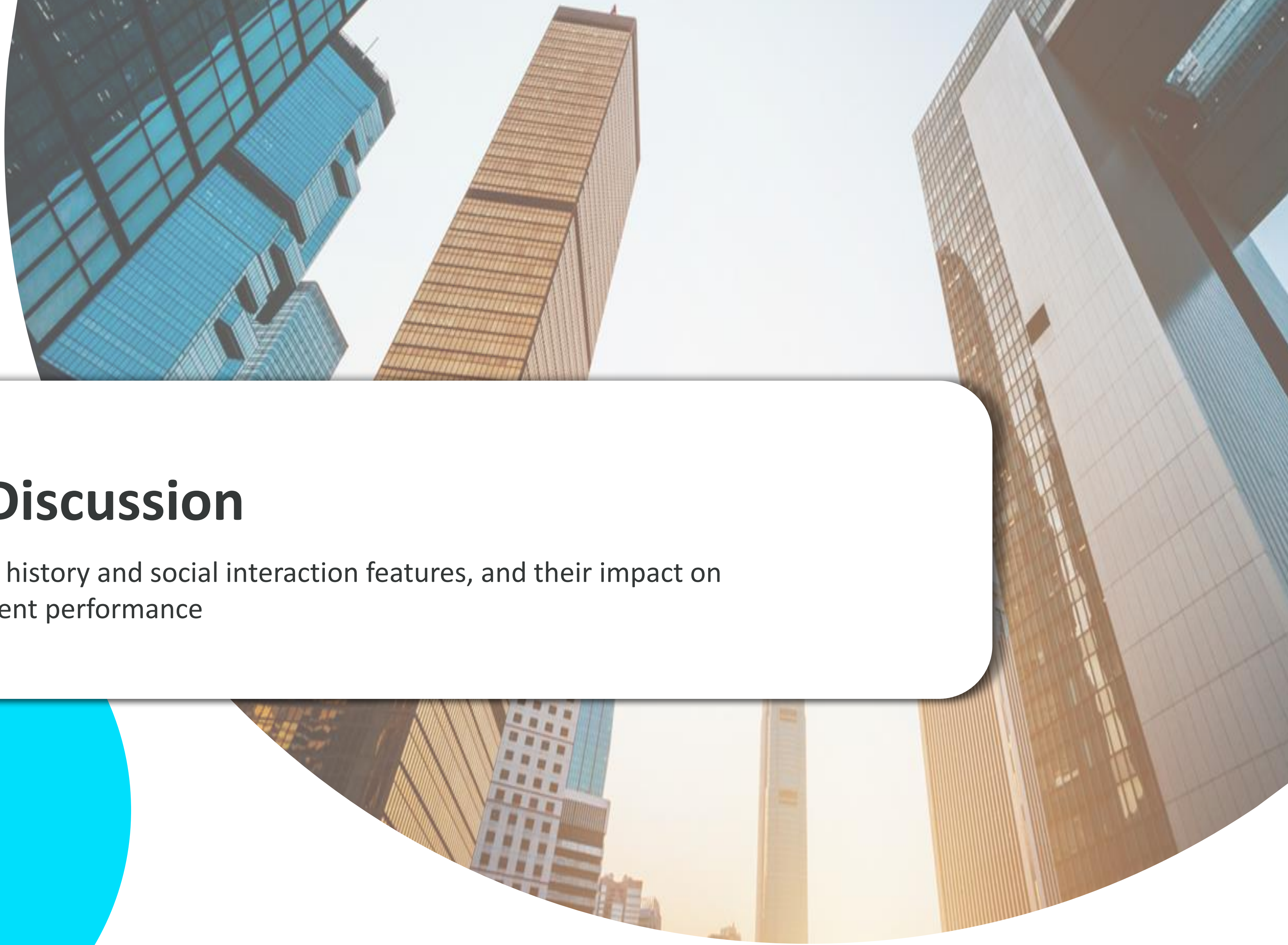
- Borrower’s Financial Features
- Node Statistics
- Social Interaction Features

We devised a series of experiments to analyze the effects on performance dynamics of credit history, repayment history, and social network features.
With each feature sets trained twelve independent 12 models.

Experiment Id	Feature Group
E1	$X = \{X_{Fin}\}$
E2	$X = \{X_{Fin} + X_{FinHist}\}$
E3	$X = \{X_{Fin} + X_{FinHist} + X_{NodeStats} + X_{SocInt} + X_{SocIntHist}\}$

Experiments setup

Second Article

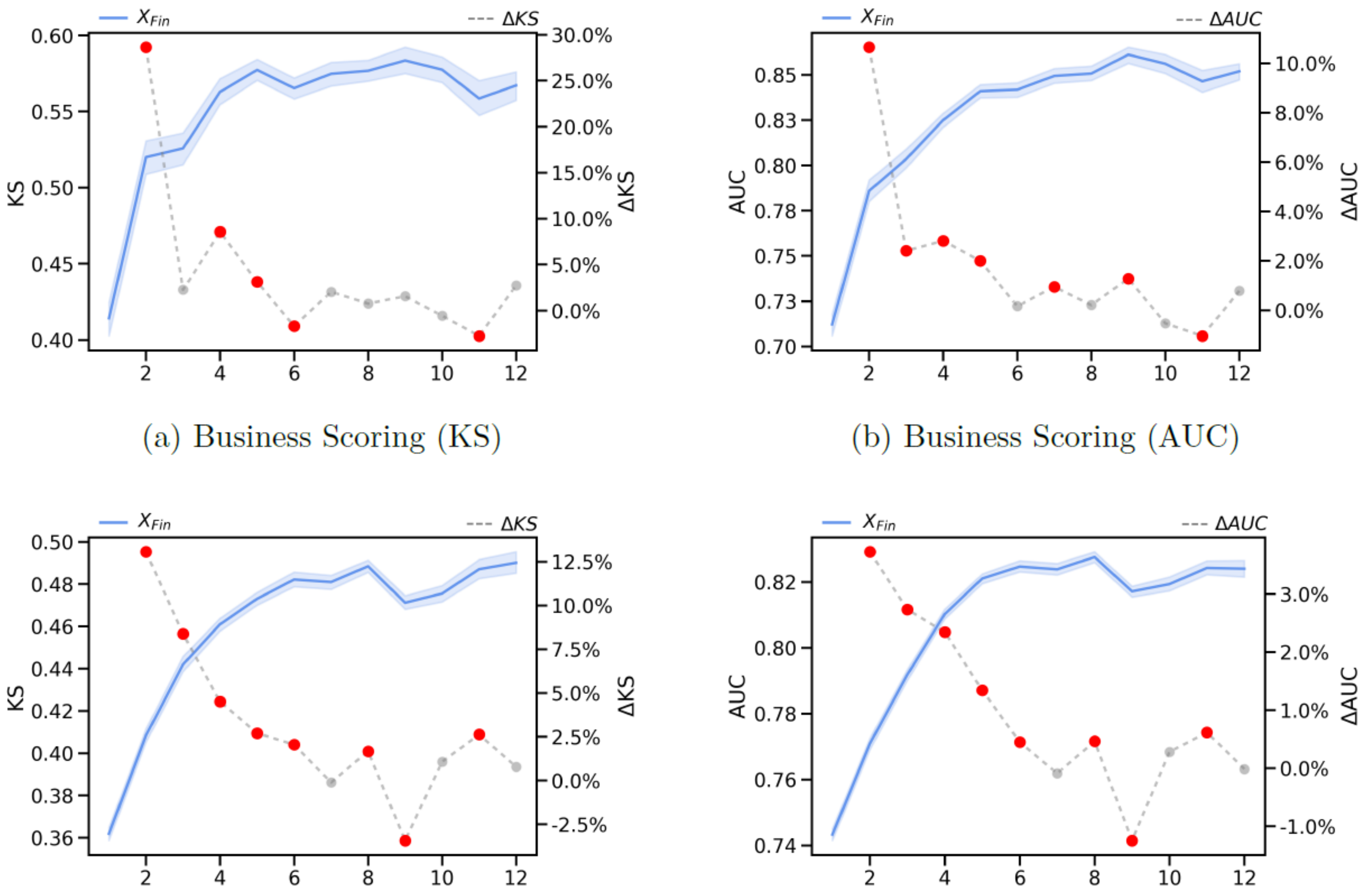


Results and Discussion

On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

Experiment E1: Borrower Credit History

Results and Discussion



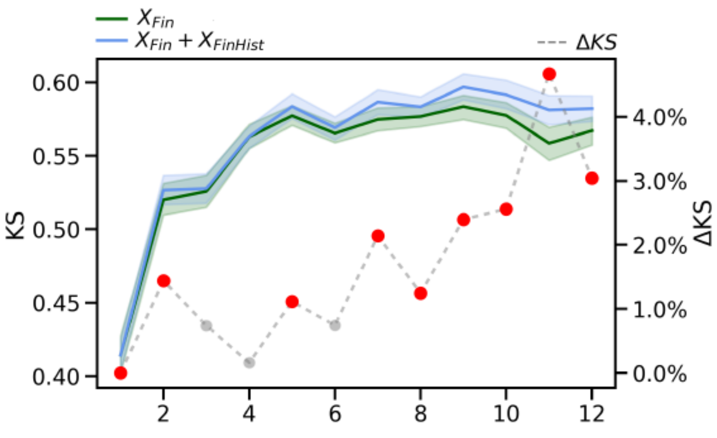
(a) Business Scoring (KS) (b) Business Scoring (AUC)

(c) Personal Scoring (KS) (d) Personal Scoring (AUC)

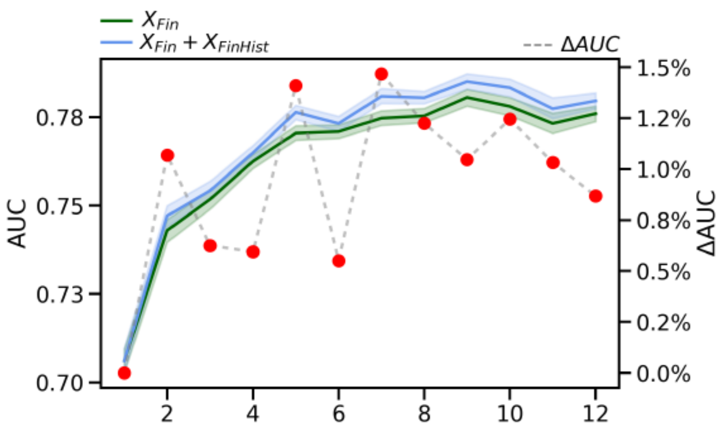
KS and AUC scores for the business scoring and personal scoring problems. The X-axis displays the number of months elapsed since the loan granting. The blue line shows the creditworthiness assessment performance (left Y-axis) for experiment E1, using only X_{Fin}: borrower features. The dotted gray line (right Y-axis) shows the percentage increment between consecutive periods; when this increment is statistically significant, the dots are colored red. Otherwise, they are colored gray.

Experiment E2: Borrower Credit History and Repayment Features

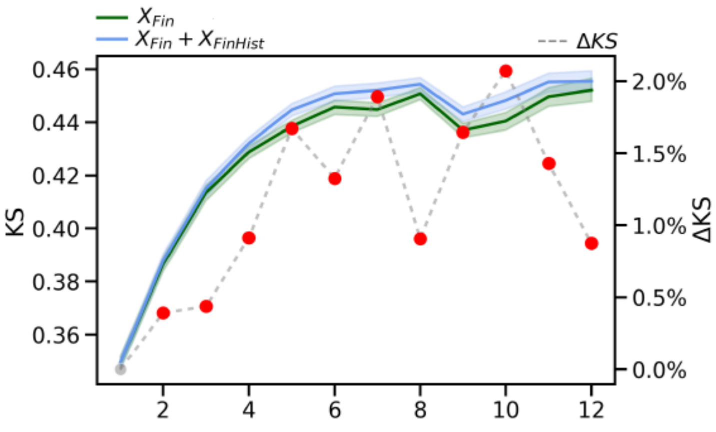
Results and Discussion



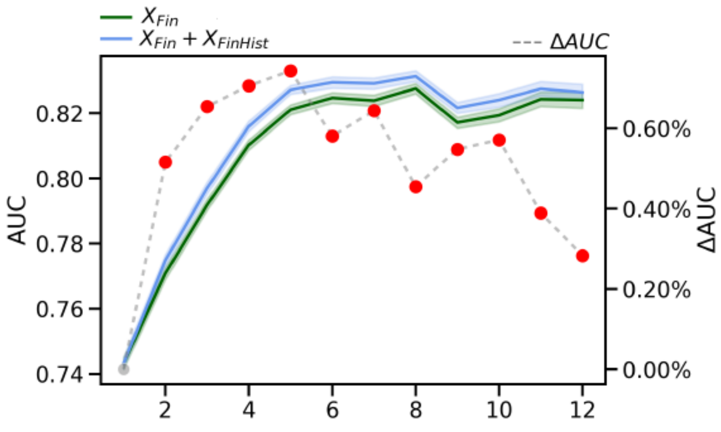
(a) Business Scoring (KS)



(b) Business Scoring (AUC)



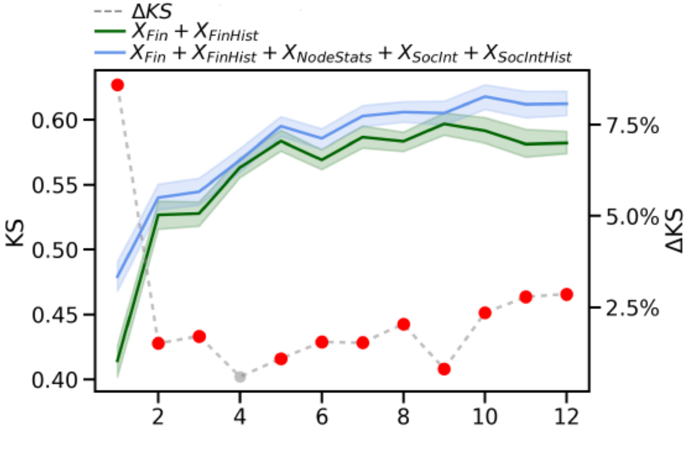
(c) Personal Scoring (KS)



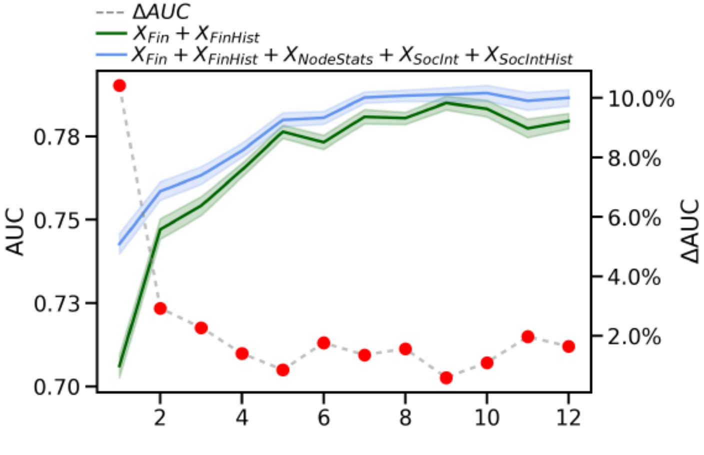
(d) Personal Scoring (AUC)

KS and AUC scores for the business scoring and personal scoring problems. The X-axis displays the number of months elapsed since the granting of the loan. The blue line and the green line show the creditworthiness assessment performance (left Y-axis) for experiment **E2** and **E1**, respectively. The dotted gray line (right Y-axis) shows the percentage increment between **E2** and **E1**; when this increment is statistically significant, the dots are colored red. Otherwise, they are colored gray.

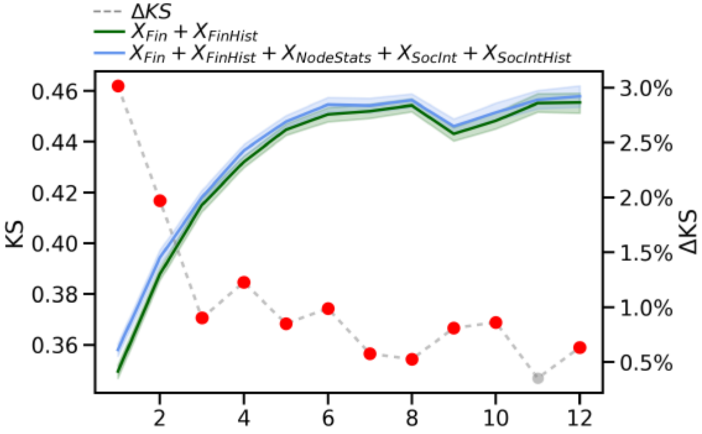
Experiment E3: Borrower credit history, repayment features and social interaction features



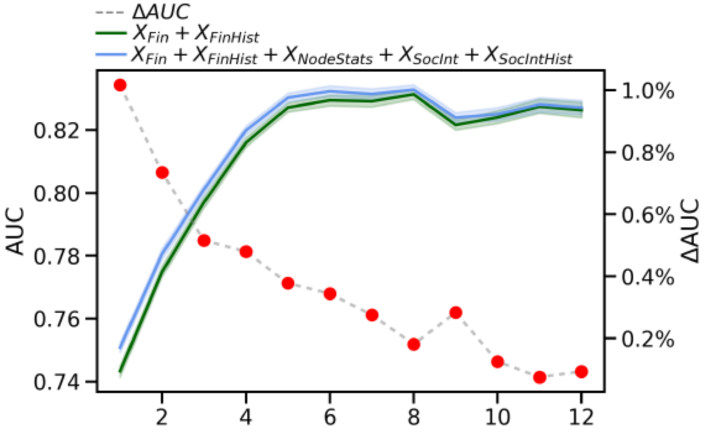
(a) Business Scoring (KS)



(b) Business Scoring (AUC)



(c) Personal Scoring (KS)

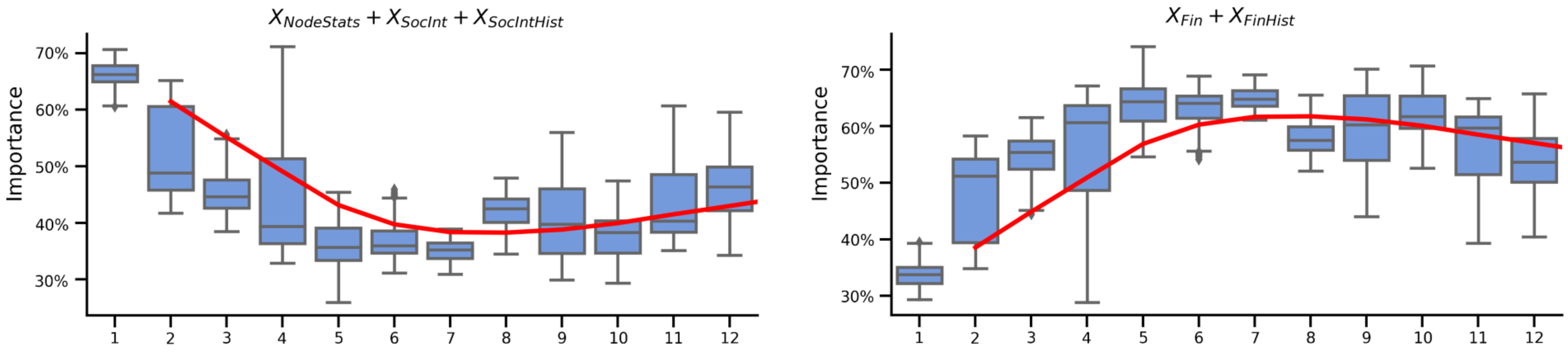


(d) Personal Scoring (AUC)

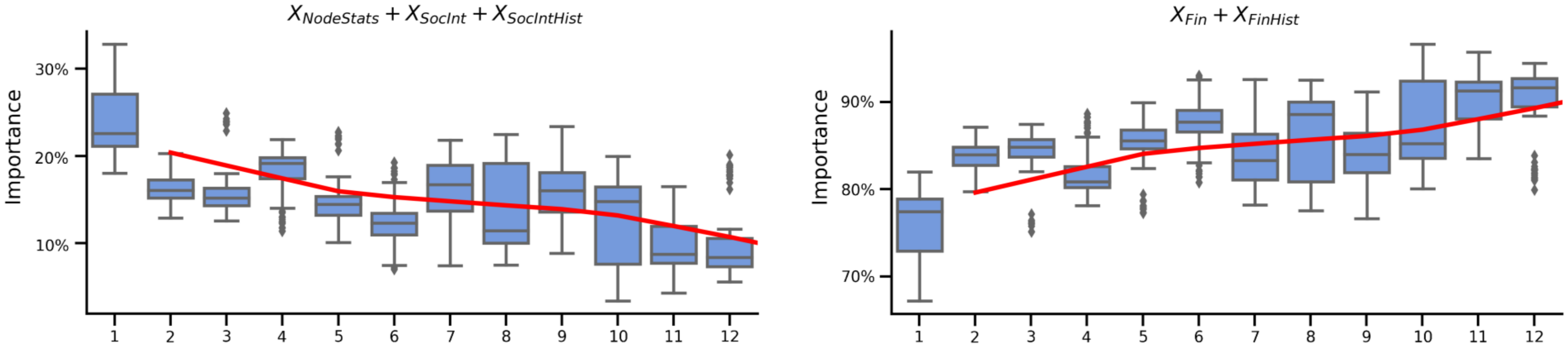
KS and AUC scores for the business scoring and personal scoring problems. The X-axis displays the number of months elapsed since the granting of the loan. The blue line and the green line show the creditworthiness assessment performance (left Y-axis) for experiment **E3** and **E2**, respectively. The dotted gray line (right Y-axis) shows the percentage increment between **E3** and **E2**; when this increment is statistically significant, the dots are colored red. Otherwise, they are colored gray.

Importance of Social Interaction Features Over Time

Results and Discussion



Business Credit Scoring



Personal Credit Scoring

Feature Importance Analysis using Shapley Values.

Figures presents the Business Scoring problem and Personal Scoring problem, in both using the Experiment E3 feature set.

The features are grouped into two categories, the borrower's features, and the social interaction features.

The X-axis displays the number of months elapsed since the loan granting. The Y-axis shows the relative feature importance. The boxplots show the feature importance in the 10-fold cross-validation, and the red line is a LOWESS regression fitted using these results.

Contributions

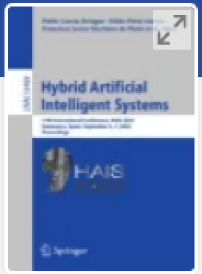
On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance

- Enhanced understanding of credit scoring and the use of social network data, challenging the conventional distinction between application and behavioral scoring.
- Focus on borrower-oriented analysis instead of process-centric evaluation, revealing the evolving performance of credit scoring models relative to the borrower's credit history growth.
- Analysis of social-interaction features contribution and their diminishing significance compared to behavioral attributes.
- Introduction of an innovative dataset that profiles individuals and companies from their initial loan acquisition through subsequent credit history, repayment actions, and social network interactions.
- Addressing the scarcity of data for behavioral model research, overcoming limitations highlighted by earlier studies.
- Novel investigation into the dynamics of credit assessment performance.

Contents

Third Article

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data



International Conference on Hybrid Artificial Intelligence Systems

↳ HAIS 2022: **Hybrid Artificial Intelligent Systems** pp 375–384 | Cite as

Home › Hybrid Artificial Intelligent Systems › Conference paper

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

Ricardo Muñoz-Cancino, Cristián Bravo, Sebastián A. Ríos & Manuel Graña

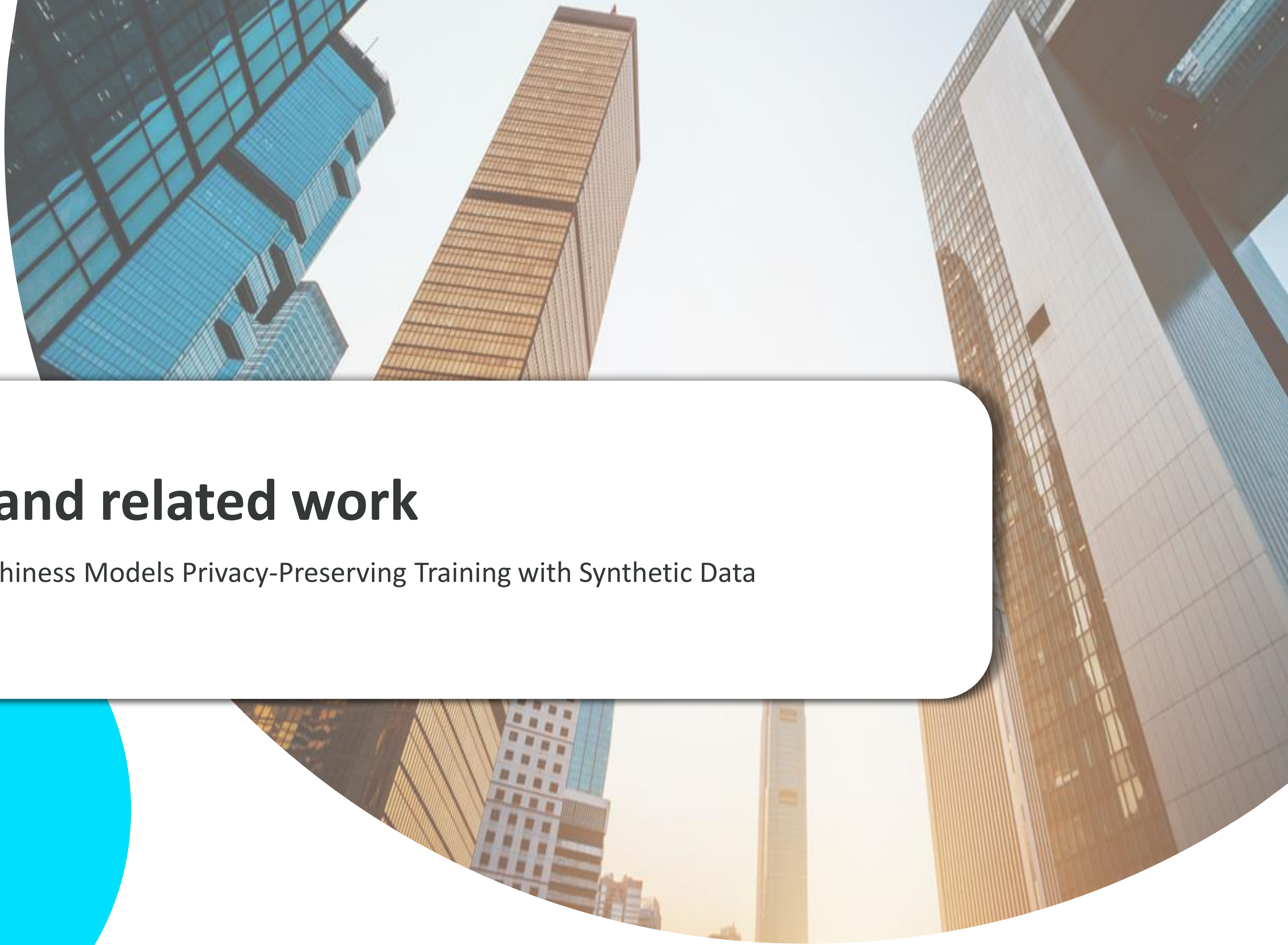
Conference paper | First Online: 12 September 2022

Research Questions

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

- Can a model trained on synthetic data perform well in real-world scenarios?
- How does increasing the features impact synthetic data quality?
- Is there a performance cost for working in a privacy-preserving environment?

Third Article



Background and related work

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

Generative models for synthetic data generation

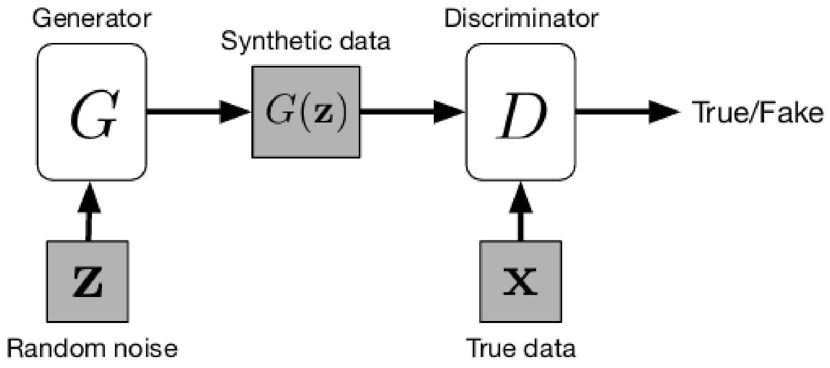
Background and related work

- Generative models in machine learning aim to learn real-data distribution and generate consistent samples from it.
- Synthetic data is valuable for problems with expensive or sensitive real-data, or when a large dataset is needed for training.
- Traditional statistical methods like Gaussian Mixture Models and Bayesian Networks were common for estimating joint distributions but struggled with mixed variable types and complex data.
- Deep learning models, like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have gained popularity for generating synthetic data due to their performance and versatility.

Generative Adversarial Networks

Background and related work

- Generative adversarial networks (GANs) are a deep learning framework based on a competitive scenario between a generator network (G) and a discriminator network (D).
- GANs aim to produce synthetic data that resembles real data, while the discriminator network tries to distinguish between real and synthetic samples.
- In its basic form, vanilla GAN, **G** maps a vector from a Gaussian distribution to data samples, and **D** outputs a probability indicating whether a sample is real or fake.
- Vanilla GANs face challenges with unbalanced categorical features and numerical features with multiple modes.
- Conditional generator (CTGAN) addresses these issues by sampling records based on log-frequency for categorical features and using kernel density estimation for handling numerical features with multiple modes (Xu et al., 2019).



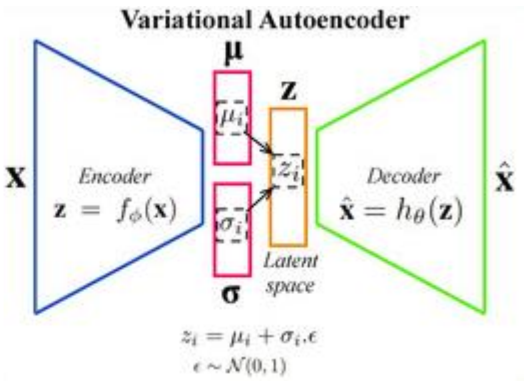
GAN Structure

Source: akiraaptx.blog

Variational autoencoders

Background and related work

- Autoencoders (AE) are unsupervised machine learning models with two primary objectives: low-dimensional representation and synthetic data generation.
- Variational Autoencoders (VAE) interpret the latent space as a probability distribution modeling training samples as independent random variables.
- VAE assumes a posterior distribution defined by the encoder and a generative distribution defined by the decoder.
- The encoder outputs vectors of means and standard deviations, which are optimized model parameters.
- TVAE, presented by Xu et al. (2019), is a VAE adaptation for tabular data, utilizing similar pre-processing as CTGAN and the evidence lower bound (ELBO) loss.



VAE Structure

Source: Zemouri, Ryad. 2020

Third Article

Experimental Design and Methodology

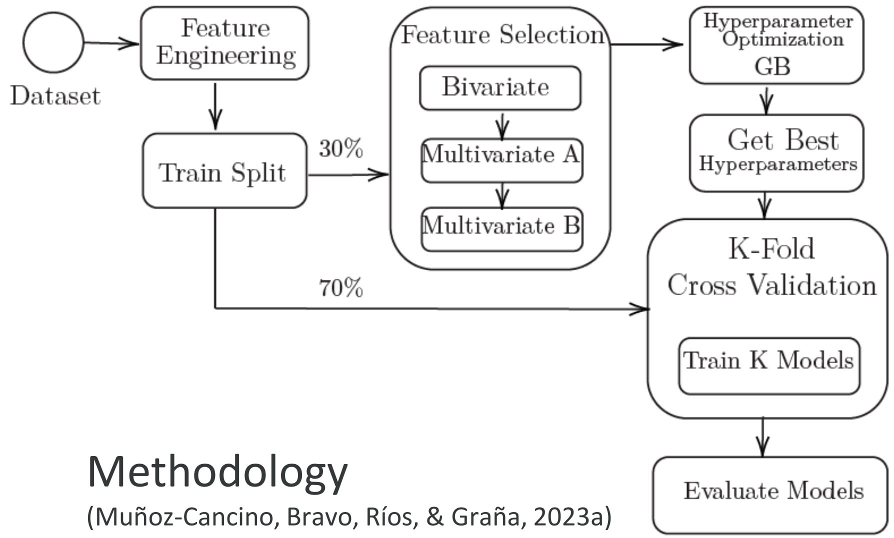
Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

Methodology

Experimental Design and Methodology

Methodology

- Framework based on previous research (Muñoz-Cancino, Bravo, Ríos, & Graña, 2023a)
- Ensuring generalization with K-fold cross-validation.
- Two holdout datasets: Holdout 2018 and Holdout 2019.
- Results stored for t-test comparison of different models.



Methodology
(Muñoz-Cancino, Bravo, Ríos, & Graña, 2023a)

Experimental Setup

- **S01** compares CTGAN and TVAE methods using borrowers' features to generate synthetic data.
- **S02** where a new synthesizer is trained using the best architecture identified in S01. It uses borrowers' features and one feature from the network data, the node degree.
- **S03** is the third experiment, where borrowers' features, node degree and social interaction features are used to train a synthesizer to generate synthetic data.

Third Article

Results and Discussion

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

Synthetic Data Generation Performance

Results and Discussion

Experiment	Synthesizer training features	Synthesizer	Arch	Exec Time (m)	CSTest	KSTest	Logistic Detection
S01	X_{Fin}	CTGAN	A	410	0.836	0.864	0.697
			B	260	0.861	0.846	0.749
		TVAE	A	230	0.962	0.868	0.803
			B	130	0.952	0.861	0.756
S02	$X_{Fin} + X_{Degree}$	TVAE	B	140	0.935	0.836	0.644
S03	$X_{Fin} + X_{Degree} + X_{SocInt}$	TVAE	A	400	0.924	0.809	0.539
S03	$X_{Fin} + X_{Degree} + X_{SocInt}$	TVAE	B	320	0.907	0.825	0.542
S03	$X_{Fin} + X_{Degree} + X_{SocInt}$	TVAE	B	465	0.930	0.819	0.513

Synthetic data generators performance

Creditworthiness assessment performance on **real data**

Results and Discussion

Classifier training features	Classifier	Holdout 2018		Holdout 2019	
		AUC	KS	AUC	KS
X_{Fin}	GB	0.88 ± 0.001	0.59 ± 0.002	0.82 ± 0.001	0.50 ± 0.002
X_{Fin}	LR	0.87 ± 0.001	0.58 ± 0.001	0.82 ± 0.001	0.50 ± 0.002
$X_{Fin} + X_{Degree} + X_{SocInt}$	GB	0.88 ± 0.001	0.59 ± 0.002	0.82 ± 0.001	0.50 ± 0.002
$X_{Fin} + X_{Degree} + X_{SocInt}$	LR	0.87 ± 0.001	0.58 ± 0.002	0.83 ± 0.001	0.50 ± 0.002
$X_{Degree} + X_{SocInt}$	GB	0.61 ± 0.002	0.17 ± 0.002	0.62 ± 0.001	0.18 ± 0.002
$X_{Degree} + X_{SocInt}$	LR	0.60 ± 0.001	0.17 ± 0.002	0.61 ± 0.001	0.18 ± 0.002

Creditworthiness assessment performance for models trained on real data

Creditworthiness assessment performance on **real data**

Results and Discussion

Classifier training features	AUC diff (%)	KS diff (%)	AUC diff p-value	KS diff p-value
X_{Fin}	0.70%	1.65%	0.000	0.000
$X_{Fin} + X_{Degree} + X_{SocInt}$	0.84%	1.91%	0.000	0.000
$X_{Degree} + X_{SocInt}$	1.65%	2.36%	0.000	0.000

Gradient boosting and logistic regression comparison on real data (holdout 2018)

Creditworthiness assessment performance on **synthetic data**

Results and Discussion

Synthesizer Experiment	Classifier training features	holdout 2018		holdout 2019	
		AUC	KS	AUC	KS
S01	X_{Fin}	0.85 ± 0.003	0.53 ± 0.002	0.82 ± 0.002	0.48 ± 0.002
S02	X_{Fin}	0.82 ± 0.001	0.51 ± 0.001	0.80 ± 0.001	0.46 ± 0.002
S03	X_{Fin}	0.85 ± 0.002	0.55 ± 0.002	0.80 ± 0.002	0.46 ± 0.002
S03	$X_{Fin} + X_{Degree} + X_{SocInt}$	0.85 ± 0.002	0.56 ± 0.003	0.80 ± 0.002	0.47 ± 0.003
S03	$X_{Degree} + X_{SocInt}$	0.60 ± 0.002	0.16 ± 0.002	0.61 ± 0.003	0.18 ± 0.002

Creditworthiness assessment performance on real data for model trained on synthetic data

Creditworthiness assessment performance on **synthetic data**

Results and Discussion

Synthesizer Experiment	Classifier training features	holdout 2018		holdout 2019	
		AUC diff	KS diff	AUC diff	KS diff
S01	X_{Fin}	-3.59%**	-10.09%**	-0.86%**	-3.92%**
S02	X_{Fin}	-6.24%**	-13.24%**	-3.32%**	-6.48%**
S03	X_{Fin}	-2.81%**	-6.01%**	-3.21%**	-6.70%**
S03	$X_{Fin} + X_{Degree} + X_{SocInt}$	-3.12%**	-5.68%**	-2.54%**	-4.73%**
S03	$X_{Degree} + X_{SocInt}$	-1.85%**	-4.31%**	-0.69%**	1.10%*

Comparison between models trained using synthetic data and models trained on real data. $**$ denotes when the difference is statistically significant using 0.05 as the p-value threshold, while $*$ uses 0.1.

Contributions

Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data

- We used synthetic data to train creditworthiness assessment models.
- We utilized a massive dataset of 1 million individuals and state-of-the-art synthesizer methods to generate synthetic data.
- We developed a training framework for analyzing model performance with synthetic data and assessing susceptibility to data drift.
- We found that the quality of synthetic data decreases as the number of attributes in the synthesizer increases. Despite lower-quality synthetic data, models trained with these data achieved good results in real-world scenarios.
- We observed only a minor reduction in predictive power (approximately 3% in AUC and 6% in KS) when using synthetic data.
- Our findings enable the training of accurate creditworthiness models while maintaining borrower privacy, fostering collaboration between financial institutions and academia, and promoting research in credit scoring without privacy and security restrictions.

Contents

Publications

Network representation learning for credit scoring

These articles are a direct outcome of the research conducted in this thesis and were produced and published as part of the Ph.D. program:

- Ricardo Muñoz-Cancino, Cristián Bravo, Sebastián A. Ríos, Manuel Graña, On the combination of graph data for assessing thin-file borrowers' creditworthiness, Expert Systems with Applications, Volume 213, Part A, 2023, 118809, ISSN 0957-4174

- Ricardo Muñoz-Cancino, Cristián Bravo, Sebastián A. Ríos, Manuel Graña, On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance, Expert Systems with Applications, Volume 218, 2023, 119599, ISSN 0957-4174

- Ricardo Muñoz-Cancino, Cristián Bravo, Sebastián A. Ríos, Manuel Graña (2022). Assessment of Creditworthiness Models Privacy-Preserving Training with Synthetic Data. In: , et al. Hybrid Artificial Intelligent Systems. HAIS 2022. Lecture Notes in Computer Science(), vol 13469. Springer, Cham.

Publications

Network representation learning for credit scoring

Other articles were developed during the doctoral studies; however, they do not pertain to the subject matter of this research:

- Muñoz-Cancino, R., Rios, S. A., Goic, M., & Graña, M. (2021). Non-Intrusive Assessment of COVID-19 Lockdown Follow-Up and Impact Using Credit Card Information: Case Study in Chile. International Journal of Environmental Research and Public Health, 18(11), 5507

- Muñoz-Cancino, R., Rios, S. A., & Graña, M. (2023). Clustering Cities over Features Extracted from Multiple Virtual Sensors Measuring Micro-Level Activity Patterns Allows One to Discriminate Large-Scale City Characteristics. Sensors. 2023; 23(11):5165

- Muñoz-Cancino, R., Ríos, S.A., Graña, M. (2023). Predicting Innovative Cities Using Spatio-Temporal Activity Patterns. In: García Bringas, P., et al. Hybrid Artificial Intelligent Systems. HAIS 2023. Lecture Notes in Computer Science(), vol 14001. Springer, Cham.

- Ricardo Muñoz-Cancino, Sebastián A. Ríos, Manuel Graña. Spatiotemporal city activity patterns from digital traces using dynamic topic models (Work in progress)

Conclusions

Network representation learning for credit scoring

- Credit is a crucial driver for economic growth. It enables individuals to achieve personal goals like housing, healthcare, and education.
- Banks and financial institutions face **credit risk**, the risk of borrowers defaulting on loans.
- Banks use mathematical and statistical **models to assess creditworthiness**.
- Incorporating **alternative data from graphs** into traditional credit scoring methods has shown its value in enhancing credit assessment.
- This **doctoral thesis explored** the potential value of utilizing network information for credit risk assessment.

Conclusions

Network representation learning for credit scoring

- **First Study:**
  - Combines various graph representation learning methods for improved creditworthiness assessment.
  - Emphasizes the importance of social-interaction data in credit scoring, especially for thin-file borrowers.

- **Second Study:**
  - Explores creditworthiness assessment performance dynamics based on credit history, loan behavior, and social data.
  - Incorporating social interaction features adds significant value, especially early in lending.
  - Creditworthiness assessment performance improves with more credit history, stabilizing after six months.

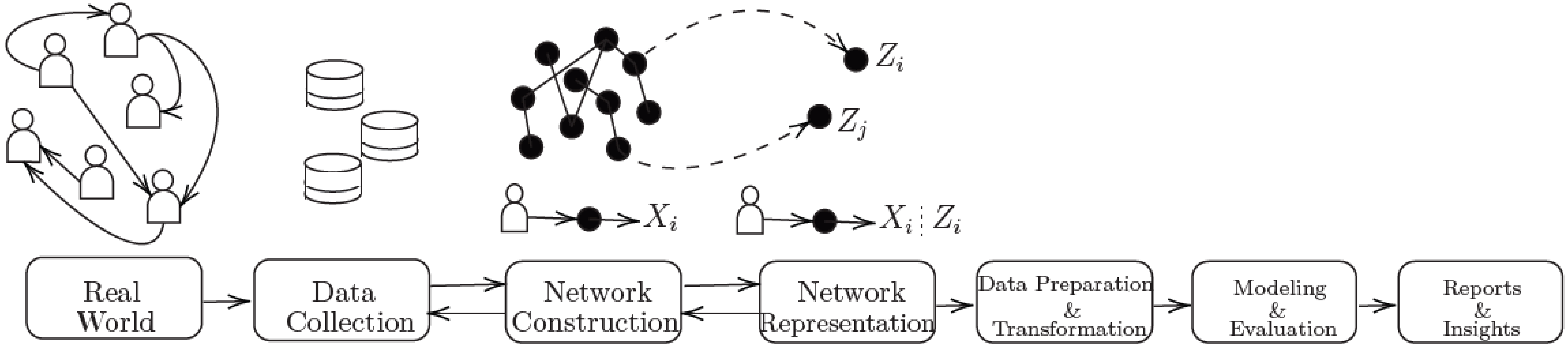
- **Third Study:**
  - Introduces the use of synthetic data for training creditworthiness assessment models.
  - Shows models perform well while preserving borrower privacy.
  - Encourages collaboration between financial institutions and academia for innovation without compromising privacy.

Thank you!

Any Question?

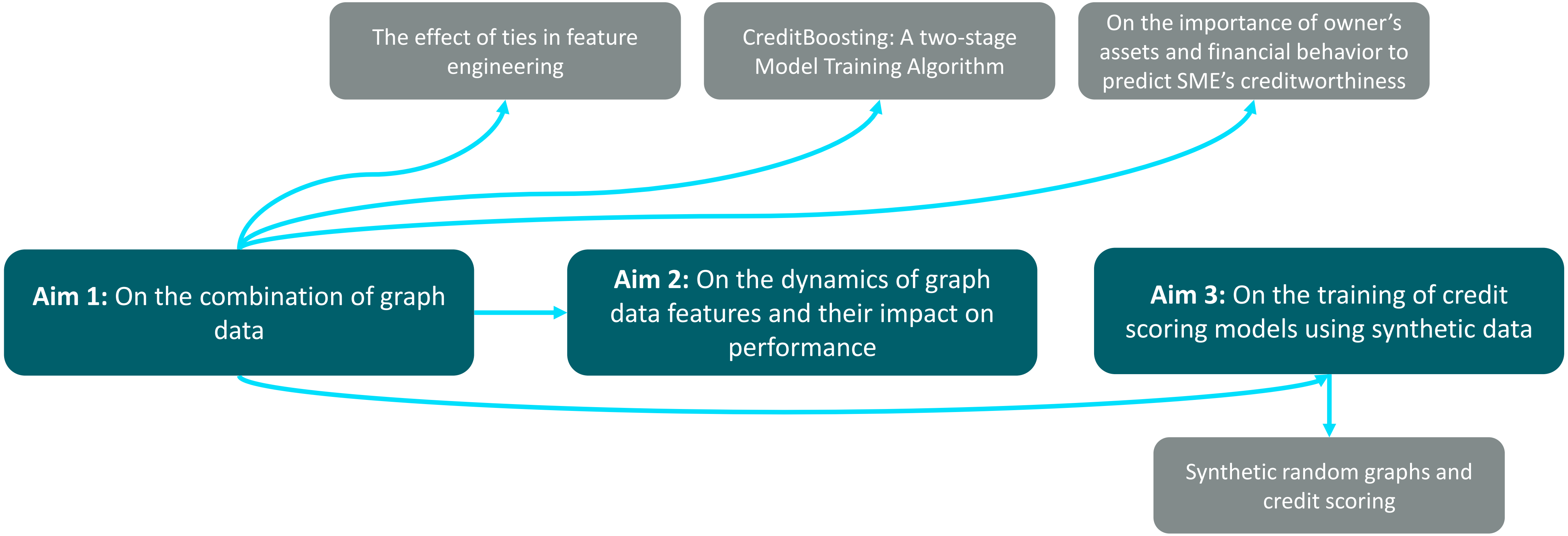
Methodology

A general approach



The Journey

that guided the progress of this investigation.



GCN Propagation Rule

Network representation learning for credit scoring

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$