Representation Learning in Graphs for Credit Card Fraud Prediction

Rafaël Van Belle, Sandra Mitrovic, Jochen De Weerdt
Research Centre for Information Systems Engineering, KU Leuven

MIDAS Workshop@ECMLPKDD, 16 September 2019

Credit: Sofia Varano

Outline

- Introduction
- Methodology: inductive representation learning for fraud detection
- Experiments & Results
- Conclusion

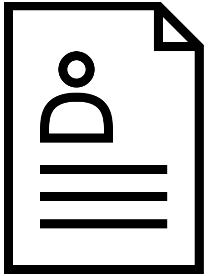
Introduction

$4.57 billion in 2016, up 34 percent wrt 2015

(Federal Reserve Payments Study)

Value of fraudulent transactions with SEPA cards:

€1.8 billion in 2016



Application fraud

Behavioral Fraud

Every transaction leaves a **Data-Trail** in credit card transaction logs.

⇒ **Data mining and machine learning** can be applied to help detect fraud.

Supervised Fraud Detection Techniques:

- Logistic Regression
- Support Vector Machines
- Random Forests
- Neural Networks

⇒ Often relying on **customer profiling** based on historical spending behaviour

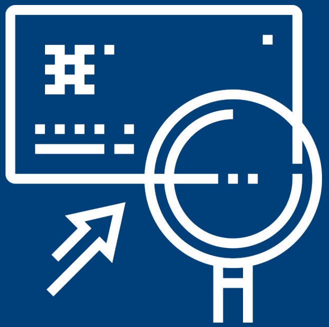
- Tedious, intricate feature engineering
- Relational features are ignored
- Lack of methodology
- Case-dependent

RQ1: Can relational/structural information from the transaction network be captured **holistically** with graph representation learning avoiding hand-crafted featurization?

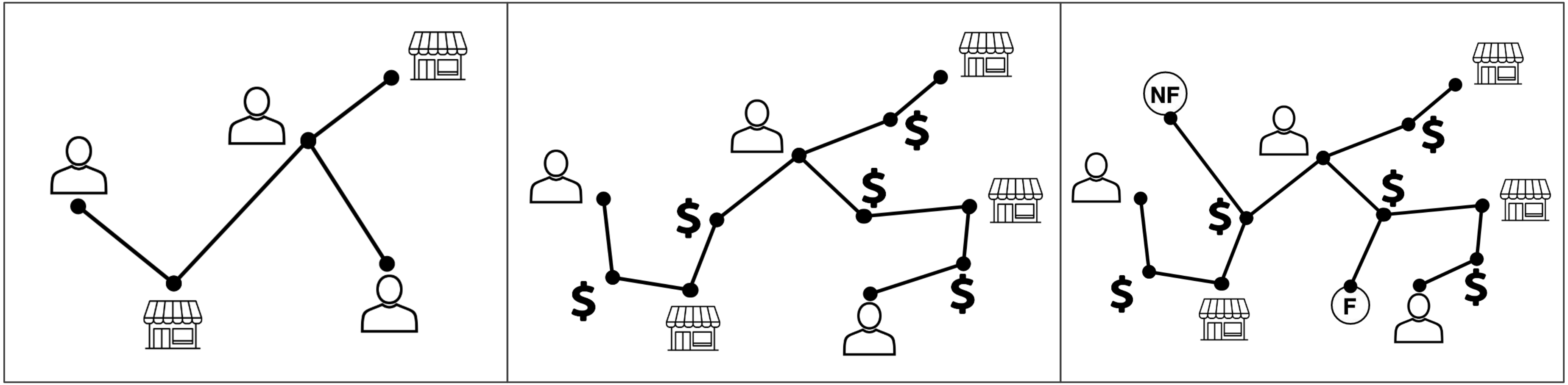
RQ2: What is the impact of alterations to the transaction network architecture in the form of **artificial nodes** on predictive performance?

RQ3: Can existing transductive representational learners be adapted to generalize to **unseen graph elements** such that they work incrementally without full retraining.

Methodology:
Inductive Representation Learning for Fraud Detection



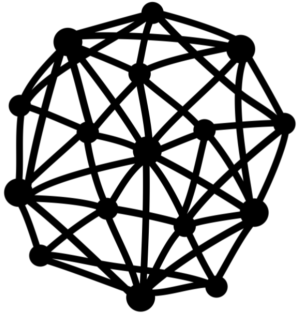
Graph Structure



Bipartite Graph

Tripartite Graph

Tripartite Graph
with Artificial Nodes



$$V \mapsto \mathbb{R}_d$$

Machine learning
Statistical Analysis
...

Network

Node Embedding

Downstream task

- Matrix-factorization (e.g. GraRep, HOPE)
- Random walk-based (e.g. Deepwalk, Node2Vec)
- Deep learning-based (e.g. SDNE, DNGR)

Random Walk-Based Node Embedding

- Deepwalk (Perrozzi et al., 2014)
- Inspired by Word2Vec model (Natural Language Processing)
- Shallow neural network with ‘fake’ learning task

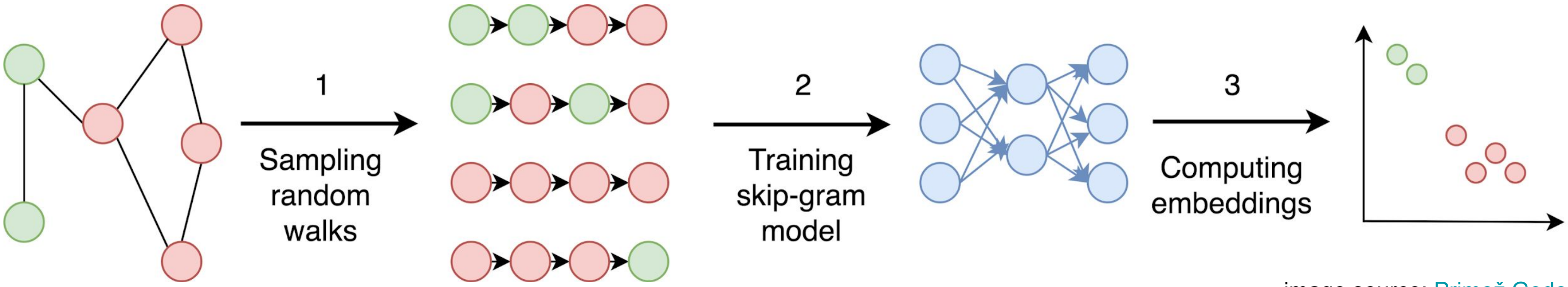


image source: Primož Godec

Inductive representation learning

- DeepWalk = transductive ⇒ cannot generalize to unseen nodes
- Continuous stream of new transactions = new unseen nodes

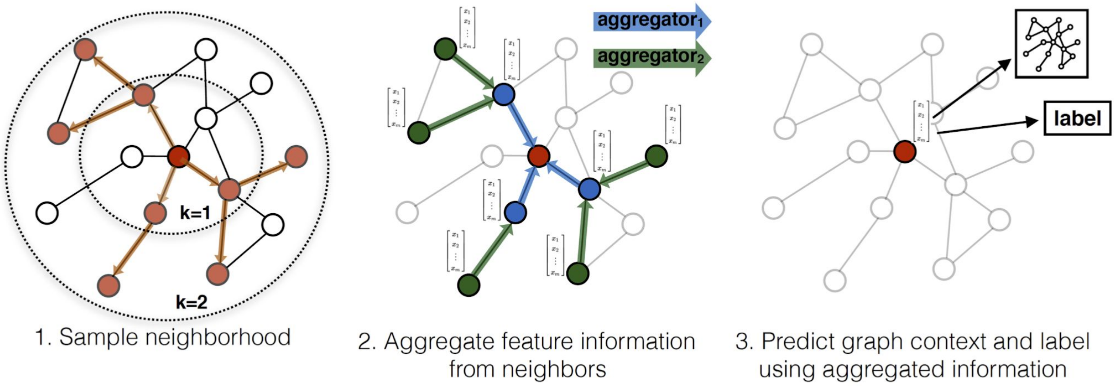
Retraining DeepWalk?

- Time demanding
- Computationally expensive
- Vector space is not preserved

⇒ Fast and efficient inductive solution?

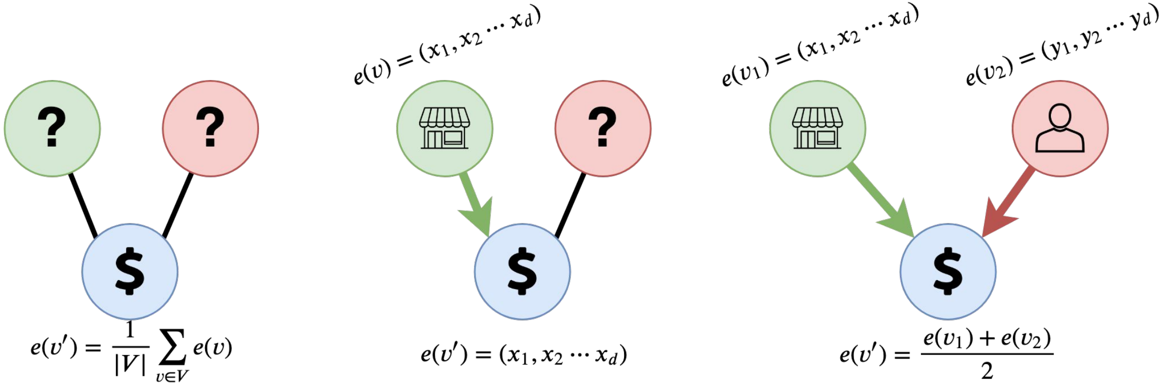
GraphSAGE

- Hamilton et al. (2017)
- Iterative aggregation of neighbourhood attribute information
- Deep learning
- Extends GCN



Pooling

- Inductive extension for shallow embedding methods.
- Pooling of neighbour embedding information
- Fast and memory efficient



Experiments & Results

Dataset

- # of transactions: **3.2 Mio**
- # of cardholders: **1.2 Mio**
- # of merchants: **130K**
- # of fraudulent transactions: **13K**
- Fraud rate: **0.32%**

TX	Cardholder	Merchant	Cat.	Country	Amount	Timestamp	Fraud
t0	AC83FD	m000174	4816	USA	7.37	2013-10-01 01:00:06	False
t1	1CD10E	m207001	5735	LUX	6.25	2013-10-01 01:00:08	False
t2	4ECA55	m003020	7523	CAN	7.18	2013-10-01 01:00:08	False
t3	74186F	m800002	4812	USA	154.93	2013-10-01 01:00:09	True
t4	8777F3	m000102	7399	BEL	15.00	2013-10-01 01:00:10	False

Experimental design

4 +1 days
5 replications

4 options

4 options

3 metrics

Data split → **Sampling** → **Embedding** → **Prediction** → **Metrics**

1. Pagerank — Traditional featurization

---

2. Graphsage Meanpool
3. Graphsage Maxpool

— Inductive method

---

4. Pooling -
5. Pooling +

— Proposed inductive extension

KU LEUVEN

Technique	AUC avg	AUC max	F1 avg	F1 max	Lift avg	Lift max
Pagerank	0.62 ± 0.10	0.71	0.16 ± 0.08	0.39	1.25 ± 0.56	2.66
Graphsage mean-pool	0.63 ± 0.05	0.64	0.04 ± 0.05	0.10	2.01 ± 0.84	2.59
Graphsage max-pool	0.71 ± 0.05	0.73	0.05 ± 0.05	0.13	2.76 ± 0.97	3.82
Pooling-	0.76 ± 0.03	0.80	0.05 ± 0.05	0.23	2.91 ± 0.93	$\underline{4.89}$
Pooling+	$\underline{0.77 \pm 0.04}$	$\underline{0.83}$	$\underline{0.18 \pm 0.12}$	$\underline{0.40}$	$\underline{3.83 \pm 0.57}$	4.53

Results Nemenyi Test

Critical value: 62.50

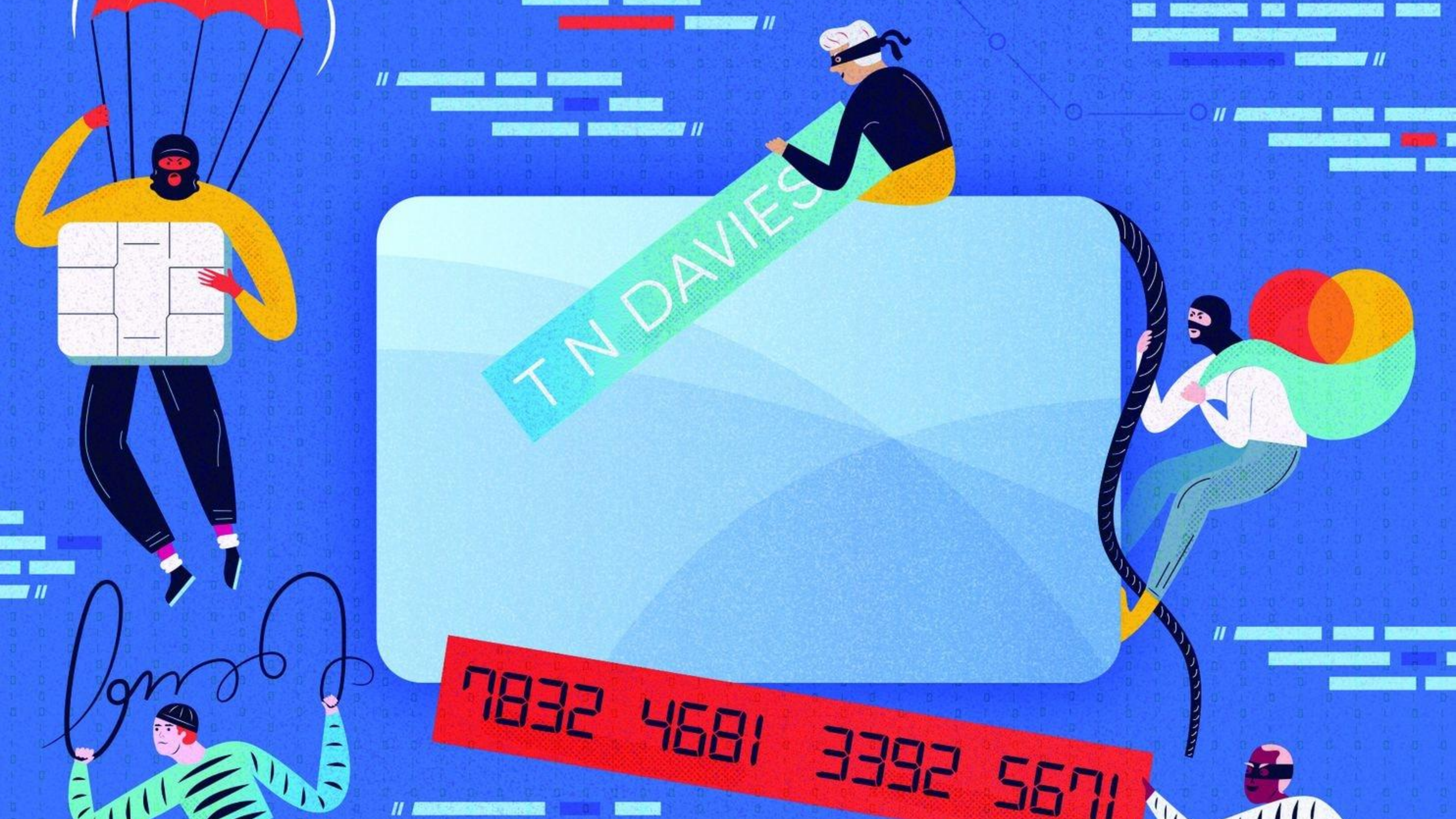
Technique + Sampling + Classifier	pooling-RO XGB	pooling+RO XGB	pooling+SMT XGB	pooling+ADS XGB
Pagerank US SVM	56.6	59.4	61.6	**63.0**
Pagerank US LR	62.2	**65.0**	**67.2**	**68.6**
Pagerank ADS LR	59.8	**62.6**	**64.8**	**66.2**
Pagerank RO RF	**62.7**	**65.5**	**67.7**	**69.1**
Pagerank SMOTE RF	**63.3**	**66.1**	**68.3**	**69.7**
Graphsage meanpool RO RF	58.0	60.8	**63.0**	**64.4**
Graphsage meanpool US LR	56.6	59.4	61.6	**63.0**
Graphsage meanpool SMT XGB	56.2	59.0	61.2	**62.6**

Conclusion

- RQ1: can representation learning be applied for fraud detection → YES
- RQ2: what is the impact of artificial nodes → artificial nodes improved the predictive performance.
- RQ3: how to generalize to unseen nodes → inductive pooling operator outperformed state-of-the-art inductive framework GraphSAGE.

Future research

- Take into consideration node type heterogeneity.
- Theoretical underpinning of the effect of artificial nodes.
- Take financial performance measures into account.


T N DAVIES
7832 4681 3392 5671