

Master Thesis

# Gaining Customer Insights using Machine Learning on Graphs

University of Basel

Author:

Michael von Siebenthal

Supervisor:

Prof. Dr. Dietmar Maringer

April 5, 2021

# Abstract

Abstract goes here

# Dedication

To mum and dad

# Declaration

I declare that..

# Acknowledgements

I want to thank...

# Contents

<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	General Overview . . . . .	8
1.2	Specific Setting . . . . .	8
1.3	Challenges . . . . .	9

# List of Figures

# List of Tables



# Chapter 1

## Introduction

### 1.1 General Overview

The aim of this Master Thesis is to analyze to what extent machine learning methods on graphs such as embedding strategies like node2vec (Grover & Leskovec 2016) or Graph Neural Networks (GNN) (Kipf & Welling 2016, Hamilton et al. 2017, Veličković et al. 2017, Vaswani et al. 2017) can be applied for bank client classification. Graph machine learning methods have proven to be very successful for classification or prediction tasks among others. The advantage of these methods is that they can consider the network structure of a dataset (e.g. social network). This is a valuable feature if one for instance compares it to the capabilities of traditional machine learning methods such as Convolutional Neural Networks (CNN) which can "only" work with grid structures (e.g. image classification). Graph machine learning methods are currently widely used in recommendation systems for social media (e.g. Pinterest, Ying et al. (2018)) and could be similarly beneficial for banks when classifying clients. Credit Scoring for instance appears to be an application in which GNN perform very well as shown by Sukharev et al. (2020).

### 1.2 Specific Setting

Another interesting application and the focus for this thesis is to classify bank clients according to their investment preference (e.g. which type of products should be advertised to which client?). This would be especially useful in the retail banking segment where advisers typically cannot know their clients personally due to the large number of clients being serviced. Investor classification is the intended main focus of this Master Thesis.

## 1.3 Challenges

This topic faces many different hurdles due to the low availability and mostly absence of available bank client data. To the extent possible, appropriate data sets will be retrieved (thus far 1 dataset found). The main difficulty however is to find a dataset which both includes attribute/feature data and the network structure of the customer data. For this reason, mostly methods to create synthetic data will be used to create the dataset for the subsequent simulation/testing.

# Bibliography

- Grover, A. & Leskovec, J. (2016), node2vec: Scalable feature learning for networks, *in* ‘Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining’, pp. 855–864.  
**URL:** <https://doi.org/10.1145/2939672.2939754>
- Hamilton, W. L., Ying, R. & Leskovec, J. (2017), ‘Inductive representation learning on large graphs’, *arXiv preprint arXiv:1706.02216* .  
**URL:** <https://arxiv.org/abs/1706.02216>
- Kipf, T. N. & Welling, M. (2016), ‘Semi-supervised classification with graph convolutional networks’, *arXiv preprint arXiv:1609.02907* .  
**URL:** <https://arxiv.org/abs/1609.02907>
- Sukharev, I., Shumovskaia, V., Fedyanin, K., Panov, M. & Berestnev, D. (2020), ‘Ews-gcn: Edge weight-shared graph convolutional network for transactional banking data’, *arXiv preprint arXiv:2009.14588* .  
**URL:** <https://arxiv.org/abs/2009.14588>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017), ‘Attention is all you need’, *arXiv preprint arXiv:1706.03762* .  
**URL:** <https://arxiv.org/abs/1706.03762>
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. & Bengio, Y. (2017), ‘Graph attention networks’, *arXiv preprint arXiv:1710.10903* .  
**URL:** <https://arxiv.org/abs/1710.10903>
- Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L. & Leskovec, J. (2018), Graph convolutional neural networks for web-scale recommender systems, *in* ‘Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining’, pp. 974–983.  
**URL:** <https://doi.org/10.1145/3219819.3219890>