UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

eXascale Infolab

UNIVERSITY OF FRIBOURG

BACHELOR THESIS

Thesis Title

Author:
Mattias Dürrmeier

Supervisor:
Prof. Dr. Philippe Cudré-Mauroux

Co-Supervisor:
Co-supervisor Name

February 27, 2023

eXascale Infolab
Department of Informatics

Boulevard de Pérolles 90 • 1700 Fribourg • Switzerland
phone +41 (26) 300 84 65 • diuf-secr@unifr.ch • www3.unifr.ch/inf

Abstract

Mattias Dürrmeier

Thesis Title

Write the thesis abstract here. Should be between half-a-page and one page of text, no newlines.

Keywords: keywords, list, here

Contents

List of Figures

Chapter 1

Experiments

1.1 Misinformation Dataset

FakeNewsNet is dataset made of political *tweets* collected from the Twitter social media. The entities present in this social network are heterogeneous. An entity can be a user, a news agency or a source.

The data contains interactions and reactions to news article posted by various news account. An edge between two entity represents an interaction. Interactions can be between entities of the same or different type: user to user, news to user, source to news, and so on.

Users reaction to the information presented can be positive, neutral or negative. Each news *tweet* can be either fake or real. The corresponding labels are available for each news.

1.2 Introduction to DeepWalk

DeepWalk allows to to learn latent representations of vertices in a social network. It performs random walks in a graph to learn latent features. The skipgram model is used to perform word embedding, in a similar fashion as word2vec.

1.3 From Heterogeneous to Homogeneous

DeepWalk only performs the embedding on homogeneous entities represented by integers. Because our data is heterogeneous, we first need to assign a unique value to each entity.

Using the available entity list, each entity is given a unique identifier. This is then mapped to the edgelists of all entities interactions.

We then build the graph using the homogeneous network. This simplify computing the adjacency list for each entity.

1.4 DeepWalk With Disinformation Data and Parameters Sensitivity

DeepWalk can now use the adjacency list to perform the graph embedding. The embedding is obtained with the entire network; scoring is however only performed with the news sources.

Once the embedding is obtained, the users and news entities are then removed from the embedding. The embedding is then analyzed and compared with each news label.

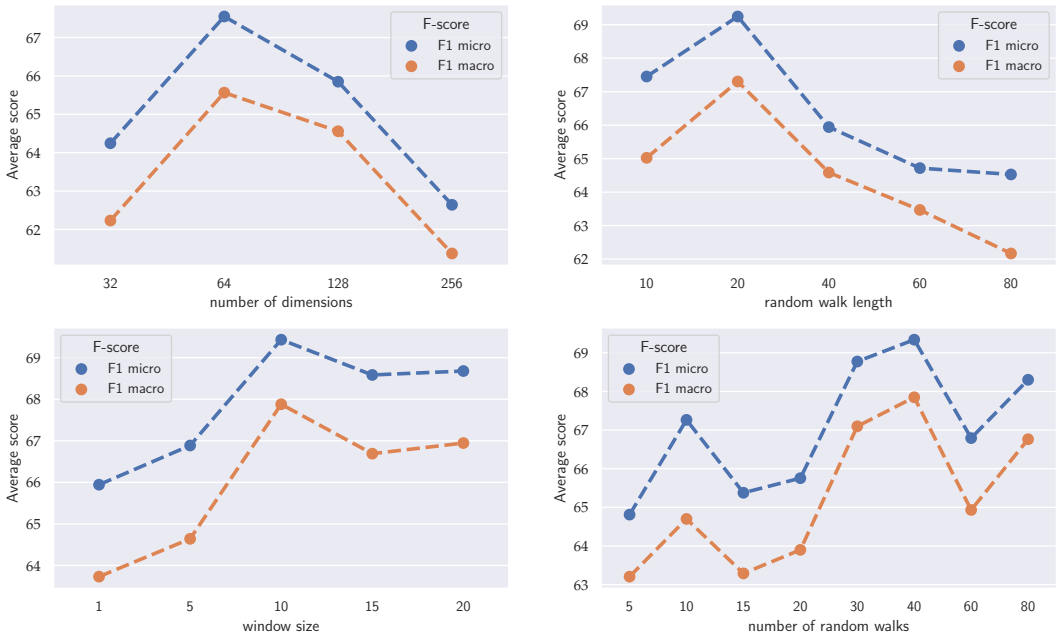
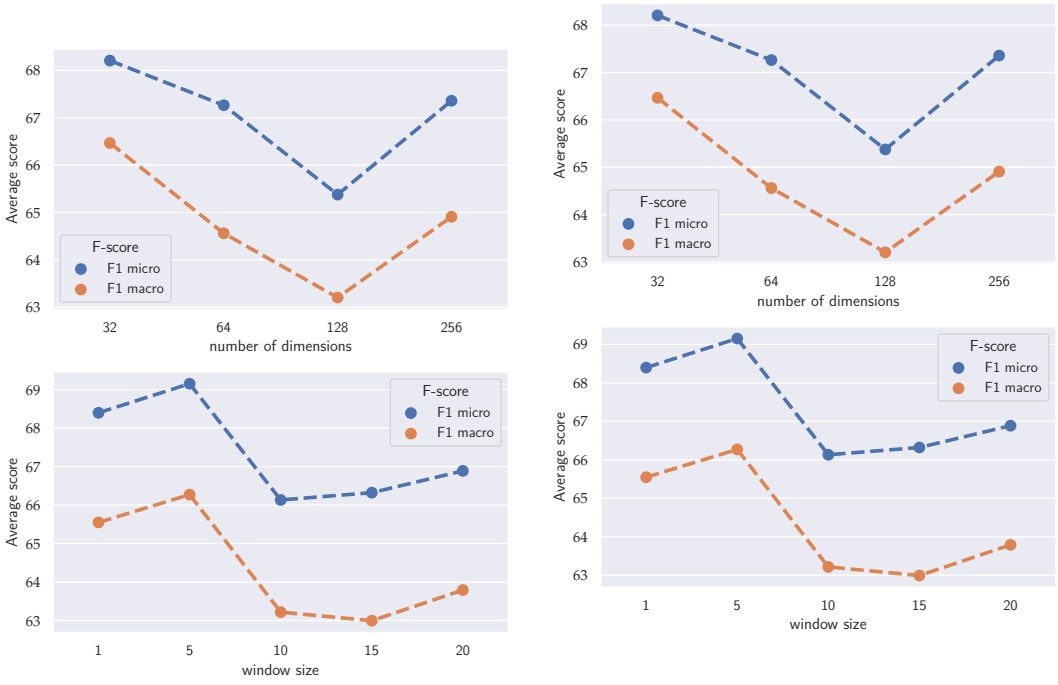


FIGURE 1.1: DeepWalk and FakeNewsNet results.

To obtain the best performance, the parameters of the program are tuned one after the other. First tuning one of them with all the others set to default, and then another one using the best value obtained for the previous run.

1.5 Node2vec

Similarly to DeepWalk, node2vec also allows to represent a network as a spatial representation.



(A) Placeholder since plot is missing.

FIGURE 1.2: Node2vec and FakeNewsNet results.