

UNIVERSITY OF LJUBLJANA
FACULTY OF COMPUTER AND INFORMATION SCIENCE

Damir Varešanović

**Programming library for network
community detection based on label
propagation**

DIPLOMA THESIS

UNIVERSITY STUDY PROGRAMME
FIRST CYCLE
COMPUTER AND INFORMATION SCIENCE

MENTOR: doc. dr. Lovro Šubelj

Ljubljana, 2018

COPYRIGHT. The results of the diploma work are the intellectual property of the author and the Faculty of Computer and Information Science of the University of Ljubljana. For the publication and use of the results of the diploma the written consent of the author, the Faculty of Computer and Information Science and the mentor is required.

The text was created with text editor \LaTeX .

Faculty of Computer and Information Science issues the following thesis:
Programming library for network community detection based on label propagation

The subject of the thesis: In progress.

I would like to thank my mentor doc. dr. Lovro Šubelj for his advice and support throughout the thesis work and my friends and family for their encouragement throughout the years. Lastly, I would like to thank my fiancée, you knew it would be a long and sometimes bumpy road, but encouraged and supported me along the way. Thank you.

Contents

Abstract

| | | |
|----------|---------------------------------------|-----------|
| 1 | Introduction | 1 |
| 2 | Label propagation | 3 |
| 2.1 | Label Ties Resolution | 3 |
| 2.2 | Label Propagation Order | 4 |
| 2.3 | Convergence Criterium | 4 |
| 2.4 | Complexity of the algorithm | 4 |
| 3 | Advances of Label Propagation | 7 |
| 3.1 | Consensus clustering | 7 |
| 4 | Conclusion | 9 |
| | Literatura | 11 |

List of used abbreviations

| Abbreviation | English | Slovenian |
|--------------|-------------------------|-------------------------|
| CA | classification accuracy | klasifikacijska točnost |
| CC | consensus clustering | konsenzno gručenje |
| ... | ... | ... |

Abstract

Title: Programming library for network community detection based on label propagation

Author: Damir Varešanović

In this diploma thesis, we present the basic implementation of label propagation, review different advances of the original method, and highlight its equivalences with other approaches. We will show how label propagation can be used effectively for large-scale community detection. We will achieve this by implementing the algorithm in the programming language Python.

Keywords: computer, computer, computer.

Chapter 1

Introduction

Among several methods in network analysis, label propagation is not the most accurate and robust one. It is, however, one of the simplest and fastest clustering methods, which is the main motive for implementing the algorithm. Label propagation can be implemented with several lines of code and applied on networks with millions of nodes.

The main goal of Label propagation is community detection in networks, while the method can also be used for other types of network clustering. In this assignment, we present the basic implementation of label propagation, examine different advances of the method, and highlight its similarities with other approaches.

For this assignment, we will use programming language Python and corresponding libraries. One of those libraries is called NetworkX, which is useful for building and working with graph structures. As the algorithm looked for communities within the graph through several iterations, we will create different tests with synthetic graphs which will be used to check if the correct number of communities has been assigned at the end.

Similar topics have been mentioned in the following articles and publications [1] [2] [4] [3].

In the 2. chapter we will show different approaches of the algorithm and how it is actually implemented.

Chapter 2

Label propagation

Action plan:

1. Read and investigate the problem in different publications.
2. Basic implementation.
3. Write introduction of the chapter.

2.1 Label Ties Resolution

Action plan:

1. Read and investigate the problem in different publications.
2. Implementation.
3. Write section and corresponding subsections.

2.1.1 Random Resolution

2.1.2 Inclusion Resolution

2.1.3 Retention Resolution

2.2 Label Propagation Order

Action plan:

1. Read and investigate the problem in different publications.
2. Implementation.
3. Write section and corresponding subsections.

2.2.1 Synchronous

2.2.2 Asynchronous

2.3 Convergence Criterium

Action plan:

1. Read and investigate the problem in different publications.
2. Implementation.
3. Write section and corresponding subsections.

2.3.1 Label equilibrium

2.3.2 Strong communities

2.4 Complexity of the algorithm

Action plan:

1. Read and investigate the problem in different publications.
2. Write section.

Chapter 3

Advances of Label Propagation

Action plan:

1. Read and investigate the problem in different publications.
2. Write introduction of the chapter.

3.1 Consensus clustering

Action plan:

1. Read and investigate the problem in different publications.
2. Implementation.
3. Write section and corresponding subsections.

Chapter 4

Conclusion

Action plan:

1. Analyse the problem as a whole and form a conclusion.
2. Write the chapter.

Bibliography

- [1] Lovro Šubelj and Marko Bajec. Group detection in complex networks: An algorithm and comparison of the state of the art. 397:144–156, 2014.
- [2] Lovro Šubelj, Marko Bajec, Slavko Žitnik, and Neli Blagus. Node mixing and group structure of complex software networks. 17, 2014.
- [3] Lovro Šubelj, Marko Bajec, Slavko Žitnik, Dejan Lavbič, and Olegas Vasilecas. General context-aware data matching and merging framework. 24:119–152, 2013.
- [4] Lovro Šubelj, Ludo Waltman, and NessJan vanEck. Clustering scientific publications based on citation relations: A systematic comparison of different methods. 11, 2016.