

BEWERTUNG VON COMMUNITY DETECTION ALGORITHMEN MIT DER CDLIB-BIBLIOTHEK

BACHELORARBEIT
ZUR ERLANGUNG DES AKADEMISCHEN GRADES
BACHELOR OF ENGINEERING (B.ENG.)

Markus Konietzka

Betreuerin:
Prof. Dr. rer. nat. Andrea Baumann

Tag der Abgabe: 30.06.2023

eingereicht bei
Universität der Bundeswehr München
Fakultät für Elektrotechnik und Technische Informatik



Universität der Bundeswehr München

Institut für
Software Engineering

Neubiberg, Juni 2023

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, insbesondere keine anderen als die angegebenen Informationen.

Der Speicherung meiner Bachelorarbeit zum Zweck der Plagiatsprüfung stimme ich zu. Ich versichere, dass die elektronische Version mit der gedruckten Version inhaltlich übereinstimmt.

Neubiberg, den 30.06.2023

Markus Konietzka

Zusammenfassung

Dieses Template soll als Formatvorlage für Bachelorarbeiten im Institut Etti6 dienen. Studierende können mithilfe dieser Vorlage, durch Ändern einiger weniger Variablen, eine konsistente Formatierung ihrer Abschlussarbeiten erreichen. Die zu verändernden Variablen sind im LaTeX-Quellcode entsprechend durch Kommentare markiert. Grundlegende LaTeX-Kenntnisse werden dennoch vorausgesetzt.

Inhaltsverzeichnis

Erklärung	III
1 Anleitung	1
1.1 Titelblatt und Erklärung	1
1.2 Abstract	1
1.3 Inhaltsverzeichnis	1
1.4 Hauptteil	1
1.5 Anhang	2
1.5.1 Abbildungen	2
1.5.2 Abkürzungen und Glossar	2
1.5.3 Quellcode	2
1.5.4 Literaturverzeichnis	2
1.5.5 Tabellenverzeichnis	3
2 Einleitung	5
2.1 Hintergrund und Motivation	5
2.2 Ziele	5
3 Grundlagen	7
3.1 Community Detection	7
3.1.1 Definition	7
3.1.2 Methriken	7
3.1.3 Algorithmen	7
3.2 NetworkX	7
3.3 CDlib - Biblitohek	7
4 Durchführung der Bewertung	9
4.1 Zusammenstellung der Testdaten	9
4.1.1 Synthetische Daten	9
4.1.2 Reale Daten	9
4.2 Bewertungsfunktionen	9
4.3 Experimente	9
5 Analyse und Bewertung	11
5.1 Ergebnisse der Algorithmen und Bewertungsfunktionen	11
5.2 Vergleich der Ergebnisse	11

6 Zusammenfassung	13
6.1 Diskussion der Ergebnisse	13
6.2 Grenzen der Studie	13
6.3 Ausblick	13
Abbildungsverzeichnis	V
Tabellenverzeichnis	VII
Quellcodeverzeichnis	IX

1 Anleitung

Diese Anleitung bietet einen Überblick über das Template für Bachelorarbeiten im Institut etti 6, und ist selbst ein Beispiel für die Anwendung Dieser.

1.1 Titelblatt und Erklärung

Das Titelblatt und die Erklärung sind bereits erstellt, und die Quelldateien müssen nicht mehr angepasst werden. Die Belegung der Variablen für Name, Datum, etc. findet im Einstiegspunkt, der Datei `thesis.tex` statt. Die zu ändernden Variablen sind durch Kommentare markiert.

1.2 Abstract

Standardmäßig wird in diesem Template ein Abstract erzeugt. Sollte dies nicht gewünscht sein, kann die Erzeugung des Abstract in der Datei `thesis.tex` durch auskommentieren deaktiviert werden. Der Text des Abstract kann in der Datei `source/002_abstract.tex` angepasst werden.

1.3 Inhaltsverzeichnis

Das Inhaltsverzeichnis wird automatisch erzeugt. Die Tiefe kann in der Datei `thesis.tex` angepasst werden.

1.4 Hauptteil

Es bietet sich an, für jedes Kapitel eine eigene `.tex`-Datei im `source`-Ordner zu erzeugen. Um diese einzubinden, müssen sie in der Datei `thesis.tex` referenziert werden.

The image shows the word "LATEX" in a large, black, serif font. The letters are bold and have a classic, slightly ornate design. The 'L' and 'T' are particularly prominent, with the 'T' having a long horizontal bar. The 'E' is also large and has a distinct shape. The 'X' is composed of two 'V' shapes joined at the top. The overall appearance is that of a high-quality typesetting result.

Abbildung 1.1: Beispielabbildung

1.5 Anhang

1.5.1 Abbildungen

Das Abbildungsverzeichnis wird automatisch erstellt, als Beispiel hierfür dient die oben gezeigte Beispiellabbildung. Dies kann verhindert werden, indem die entsprechende Anweisung in der Datei 801_appendix.tex auskommentiert wird.

1.5.2 Abkürzungen und Glossar

Abkürzungen wie etwa BA und Erläuterungen wie Latex-Template können in der Datei acronym.tex ergänzt werden. Die entsprechenden Verzeichnisse im Anhang werden automatisch erzeugt.

1.5.3 Quellcode

```
1 export const writeThesis = (literature, template) => {
2   var thesis;
3   while(!literature.isUnderstood){
4     literature.openAgain();
5   }
6   template.fill();
7   thesis = template.toPDF();
8   return thesis;
9 }
```

Quellcode 1.1: thesis.js

1.5.4 Literaturverzeichnis

Das Literaturverzeichnis wird automatisch erstellt. Dafür wird eine Literaturdatenbank (Dateiformat .bib) benötigt. Empfehlenswert hierfür sind die Programme Zitavi (Windows) oder Zotero (MacOs / Linux). Aus diesen können diese erzeugten Bibliotheken exportiert, und in die oberste Ebene der Ordnerstruktur des Templates unter dem Namen Literatur.bib eingefügt werden. Zitate, die im Text vorkommen, werden automatisch in das Literaturverzeichnis übernommen. Kennzeichnen von indirekten Zitaten [noauthor_software_nodate], und "direkten Zitaten"[noauthor_software_nodate] erzeugt somit dem IEEE-Zitierstil entsprechenden Zitate.

1.5.5 Tabellenverzeichnis

Das Tabellenverzeichnis wird automatisch erstellt, wie die Folgende Beispieltabelle zeigt:

a	b	c	d
1	6	11	16
2	7	12	17
3	8	13	18
4	9	14	19
5	10	15	20

Tabelle 1.1: Beispieltabelle

2 Einleitung

In der modernen Zeit der Digitalisierung spielt die effiziente und effektive Verarbeitung von Daten eine entscheidende Rolle. Dabei stellt sich auch die Frage, wie Datensätze am besten dargestellt werden können, um die richtigen Informationen leicht erkennbar zu machen. Neben den bekannten Formen wie Balken-, Kreis- oder Liniendiagrammen besteht in der Informatik auch die Möglichkeit, Datensätze in Form eines Graphen zu präsentieren.

2.1 Hintergrund und Motivation

2.2 Ziele

Das Hauptziel dieser Bachelorarbeit ist es, eine umfassende Bewertung von Community Detection Algorithmen mithilfe der CDlib-Bibliothek durchzuführen. Dabei liegt der Fokus auf der Analyse und Vergleich der Leistungsfähigkeit verschiedener Algorithmen bei der Erkennung von Gemeinschaftsstrukturen in komplexen Netzwerken. Die Arbeit strebt an, Erkenntnisse über die Stärken und Schwächen der Algorithmen zu gewinnen und ihre Anwendbarkeit in verschiedenen Domänen zu untersuchen. Darüber hinaus sollen konkrete Anwendungsbeispiele der CDlib-Bibliothek vorgestellt werden, um ihre Wirksamkeit in realen Szenarien zu demonstrieren. Das langfristige Ziel ist es, zur Weiterentwicklung und Verbesserung von Community Detection Techniken beizutragen und eine solide Grundlage für zukünftige Forschung und Anwendungen auf diesem Gebiet zu schaffen.

3 Grundlagen

3.1 Community Detection

3.1.1 Definition

3.1.2 Methriken

3.1.3 Algorithmen

3.2 NetworkX

NetworkX ist eine Bibliothek für die Scriptsprache Python und bietet umfassende Möglichkeiten Datensätze in Form von Graphen zu visualisieren. Es werden darüber hinaus auch Funktionen angeboten um künstliche bzw. zufällige Netzwerke zu generieren.

3.3 CDlib - Biblitohek

4 Durchführung der Bewertung

4.1 Zusammenstellung der Testdaten

4.1.1 Synthetische Daten

4.1.2 Reale Daten

4.2 Bewertungsfunktionen

4.3 Experimente

5 Analyse und Bewertung

5.1 Ergebnisse der Algorithmen und Bewertungsfunktionen

5.2 Vergleich der Ergebnisse

6 Zusammenfassung

6.1 Diskussion der Ergebnisse

6.2 Grenzen der Studie

6.3 Ausblick

Anhang

Abbildungsverzeichnis

1.1 Beispielabbildung	1
---------------------------------	---

Tabellenverzeichnis

1.1	Beispieltabelle	3
-----	---------------------------	---

Quellcodeverzeichnis

1.1	thesis.js	2
-----	---------------------	---

