Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

[Analysis of Social Networks]

Bachelor Thesis

2024
Matej Zelenák

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

[Analysis of Social Networks]

Bachelor Thesis

Study Programme:	Applied Informatics
Field of Study:	Applied Informatics
Department:	Department of Applied Informatics
Supervisor:	doc. RNDr. Damas Gruska, PhD.

Bratislava, 2024
Matej Zelenák

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:
Študijný program:

Študijný odbor:
Typ záverečnej práce:
Jazyk záverečnej práce:
Sekundárny jazyk:

Názov:

Anotácia:

Vedúci:
Katedra:
Vedúci katedry:

Dátum zadania:

Dátum schválenia:

garant študijného programu

.....

študent

.....

vedúci práce

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:
Study programme:

Field of Study:
Type of Thesis:
Language of Thesis:
Secondary language:

Title:

Annotation:

Supervisor:
Department:
Head of department:

Assigned:

Approved:

Guarantor of Study Programme

.....
Student

.....
Supervisor

Acknowledgments: Tu môžete poďakovať školiteľovi, prípadne ďalším osobám, ktoré vám s prácou nejako pomohli, poradili, poskytli dáta a podobne.

Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: jedno, druhé, tretie (prípadne štvrté, piate)

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords:

Contents

List of Figures

List of Tables

Chapter 1

Literature Review

1.1 Social Networks

Social networks [1, 2] are the intricate patterns of social connections that form among individuals, groups, organizations, or other entities within a social system. These connections are established through various interactions such as communication, friendship, collaboration, or resource exchange.

In essence, social networks serve as the underlying framework through which people interact, share information, and collaborate with one another. They encompass both strong ties, such as close relationships with family and friends, and weak ties, such as casual acquaintances or connections through professional networks.

Social networks play a fundamental role in shaping individual behavior, facilitating the spread of information and ideas, and influencing social dynamics within communities and societies. They provide opportunities for social support, resource mobilization, and collective action, while also influencing individuals' access to resources, opportunities, and social capital.

Understanding the structure and dynamics of social networks is crucial for comprehending how social systems function and evolve over time. By studying the patterns of connections between individuals and groups within a network, researchers can gain insights into the mechanisms driving social interactions, the formation of social norms, and the emergence of collective behavior.

Overall, social networks represent the complex web of relationships that underlies human social life, and studying them provides valuable insights into the dynamics of social behavior, communication, and interaction.

1.1.1 Facebook

Established in 2004 by Mark Zuckerberg, Facebook [3, 4] is a social media platform connecting users worldwide. Users can actively participate by sharing content, includ-

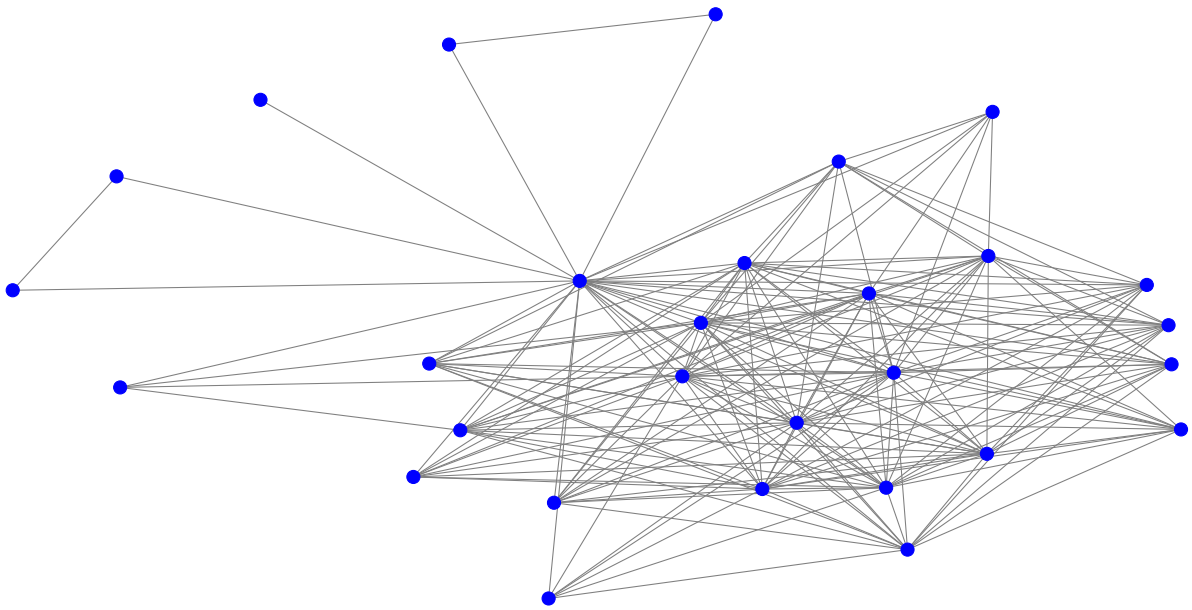

Figure 1.1: Example of connected actors in social network

ing posts, photos, and videos. Key features like groups facilitate connections among users with shared interests; events coordinate gatherings and activities, significantly enhancing interaction among individuals with similar interests and opinions.

One of the biggest feature of Facebook is its news feed, which delivers an infinite stream of content from friends, pages, and groups based on relevance and user preferences. Public figures, including politicians, use Facebook to connect with followers, share news, and engage in political discourse.

Despite its widespread popularity, Facebook has frequently faced challenges such as privacy concerns, data breaches, and distibution misinformation. However, despite these issues, it continues to maintain its position as the most widely utilized platform by billions of users worldwide. These challenges highlight the importance of ongoing efforts to address and mitigate the spread of misinformation and enhance user privacy and security on the platform.

In summary, Facebook represents more than just a social networking site; it's a dynamic platform that shapes how people connect, communicate, and collaborate online, while also serving as a fertile ground for academic inquiry and research.

1.1.2 Instagram

With a user base exceeding one billion globally, Instagram [5, 6] has established itself as a widely utilized social media platform. It enables users to share photos, videos, and messages. Through various features like Stories, Feed, Live, IGTV (for longer videos), and Direct messaging, individuals, including teenagers, use Instagram for diverse purposes such as documenting significant moments, connecting with friends and family,

Figure 1.2: Example of post on Instagram with likes, comments `https://www.instagram.com/zuzana_caputova/`

building communities, and exploring shared interests. It is compatible with Apple iOS devices and Android smartphones and tablets. See example of post shown in Figure 1.2

Users can follow others and be followed, although reciprocity isn't required like on Facebook. Users can follow accounts privately, controlling who can view their posts. However, by default, content is visible to anyone unless the account is set to private. Users maintain control over their privacy settings, with the option to make their account private and approve followers.

Posting on Instagram involves uploading photos or videos. Users can add captions, locations, tag people, and choose whether to share content exclusively with their followers or extend it beyond the app.

1.2 Scraping

As shown in Figure 1.3 Web scraping [7, 8, 9] is a data extraction technique employed to retrieve information from websites across the internet. It encompasses a series of systematic steps designed to access, parse, extract, transform, and store data from web pages. At its core, web scraping involves accessing the underlying HTML code of web pages, extracting relevant data elements, and transforming them into a structured format for analysis or further processing.

The process begins with accessing web pages using software tools or programming languages capable of fetching the HTML content. These tools facilitate the retrieval of web page data, providing a foundation for subsequent parsing and extraction tasks.

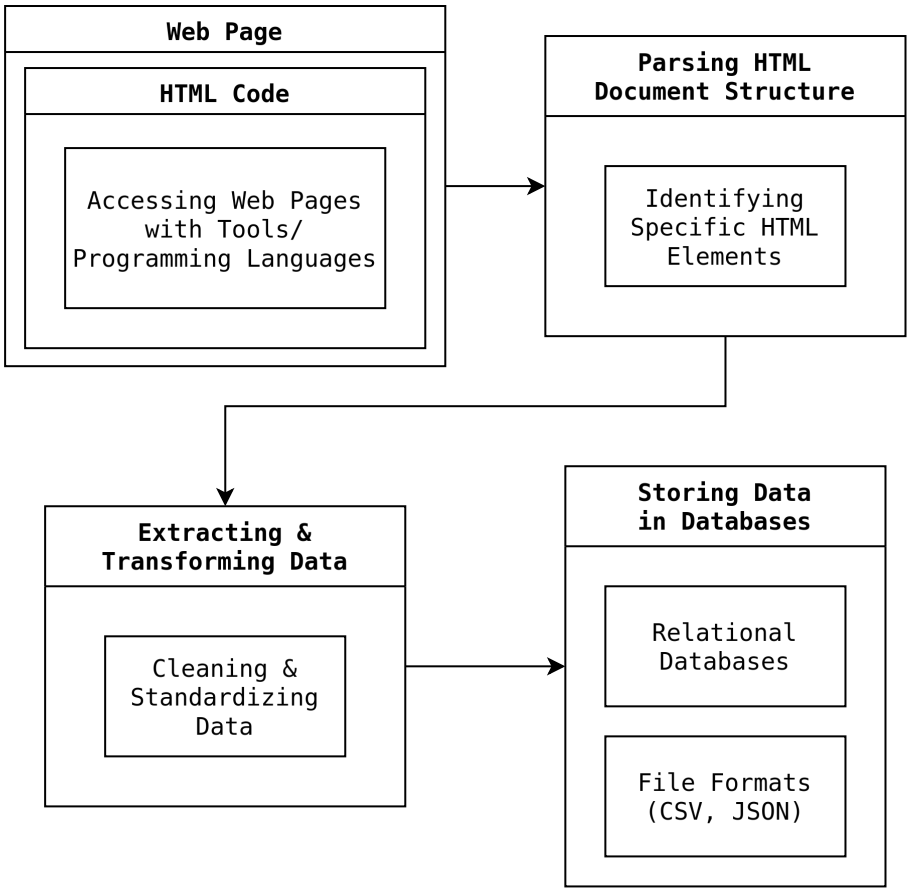

Figure 1.3: Process of Scraping

Once the HTML content is obtained, the parsing stage comes into play, where the structure of the HTML document is analyzed to identify specific elements containing the desired information.

Following parsing, the extracted data undergoes a transformation phase to convert it into a structured format suitable for analysis. This may involve cleaning and standardizing the data, removing HTML tags, formatting dates or numbers, and handling any inconsistencies or anomalies encountered during extraction. By transforming the data into a structured format, it becomes easier to manipulate and analyze, enabling insights to be gleaned effectively.

Finally, the transformed data is stored in a database, spreadsheet, or other storage systems for future use. Storing the scraped data facilitates easy access, retrieval, and analysis, allowing researchers, analysts, and developers to leverage the extracted information for various applications. Common storage options include relational databases like MySQL or PostgreSQL, or simple file formats like CSV or JSON.

1.2.1 Facebook Scraper

Facebook Scraper [10] is a Python library developed by Kevin Zúñiga, available on GitHub, designed for scraping public data from Facebook. With this tool, users can extract various types of information from Facebook pages, including posts, comments, reactions, and other metadata.

The library leverages web scraping techniques to access and retrieve data from Facebook's web pages, mimicking the behavior of a web browser. It does not rely on Facebook's official API, making it suitable for extracting data that may not be accessible through official channels.

Facebook Scraper provides a convenient interface for specifying the target Facebook page and the type of data to extract. It also includes features for pagination, filtering, and sorting the retrieved data, allowing users to customize their scraping operations according to their specific requirements.

However, it's essential to note that scraping data from Facebook may raise ethical and legal considerations, particularly concerning user privacy and data usage policies. Users of Facebook Scraper should exercise caution and ensure compliance with Facebook's terms of service and relevant laws and regulations when using the library for data extraction purposes.

1.2.2 Instaloader

Developed by Alexander Graf, Instaloader [11] is a powerful Python tool designed for interacting with Instagram data programmatically. It provides a comprehensive set of functionalities for accessing and downloading Instagram content without the need for an official API.

One of the primary features of Instaloader is its ability to fetch various types of data from Instagram profiles, including posts, stories, IGTV videos, likes, comments, and more. Users can specify the type of data they want to retrieve and define filters based on parameters such as hashtags, geotags, or user IDs.

Moreover, Instaloader facilitates the downloading of media content from Instagram, allowing users to save images, videos, and profile pictures locally on their machines. This capability is particularly useful for researchers, data analysts, and enthusiasts who wish to archive or analyze Instagram content offline.

Additionally, Instaloader supports advanced functionalities such as profile crawling, enabling users to traverse through followers, followings, and interactions of Instagram profiles. This feature is valuable for studying social network dynamics, user engagement patterns, and influencer marketing strategies on the platform.

Overall, Instaloader serves as a valuable tool for accessing, downloading, and analyzing Instagram data in a flexible and efficient manner. Its wide range of functionalities

makes it a popular choice among researchers, data scientists, journalists, and social media enthusiasts seeking to gain insights into Instagram's vast ecosystem of content and interactions.

1.3 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a process in natural language processing (NLP) that involves determining the sentiment expressed in a piece of text. The goal is to understand whether the text expresses positive, negative, or neutral sentiment towards a particular subject or entity. [12]

Sentiment analysis algorithms typically work by analyzing the words, phrases, and context within the text to infer the sentiment. This can involve various techniques, such as lexical analysis to identify positive or negative words, machine learning algorithms trained on labeled datasets, or more advanced methods like deep learning models such as BERT 1.3.1.

1.3.1 BERT

Crafted by Google, this NLP model is tailored to process and grasp human language in a manner more aligned with human cognition. BERT [13, 14] and other language models are trained on vast amounts of text from the internet, absorbing the patterns and meanings of words and sentences. What distinguishes it is its capability to simultaneously analyze words in a sentence from both directions, thereby considering their contextual significance comprehensively. This bidirectional approach helps to understand nuances and relationships between words better than previous models. Essentially, BERT acts as a sophisticated tool for computers to comprehend and generate human-like language

1.3.2 SlovakBERT

A language model tailored specifically for Slovak, SlovakBERT stands out in the field of transformers-based models. Despite the availability of multilingual models that support Slovak language, the development of a dedicated Slovak model holds promise for achieving superior results and optimizing language processing efficiency.

It adopts the RoBERTa [15] architecture and undergoes training using a web-crawled corpus. This selection of architecture and training data is customized to enhance performance specifically for the Slovak language.

The authors of SlovakBERT [16] tackle the absence of established evaluation standards for the Slovak language by devising their own set of evaluation tests. These tests

encompass various tasks including part-of-speech tagging, semantic textual similarity, sentiment analysis, and document classification. By establishing these benchmarks, they provide a comprehensive framework for evaluating Slovak language models, including SlovakBERT, which could become a standard for future assessments.

Sentiment Analysis model based on SlovakBERT

It utilizes the specialized features and contextual understanding of the Slovak language provided by SlovakBERT to accurately determine the sentiment expressed in Slovak text. Leveraging the bidirectional contextual understanding of words and phrases, the model comprehensively analyzes the text to discern whether the sentiment expressed is positive, negative, or neutral. By fine-tuning SlovakBERT on sentiment-related tasks and datasets, the model learns to identify nuanced expressions of sentiment, capturing subtle nuances and contextual cues specific to Slovak language usage. This allows the sentiment analysis model to provide precise and contextually appropriate assessments of sentiment in Slovak text.

1.4 Related Work

[TODO]

Bibliography

[1] Stanley Wasserman and Katherine Faust. Social network analysis: Methods and applications. 1994.

[2] Linton Freeman et al. The development of social network analysis. *A Study in the Sociology of Science*, 1(687):159–167, 2004.

[3] Facebook. Facebook, 2004. Prominent social media platform for connecting users globally.

[4] Robert E Wilson, Samuel D Gosling, and Lindsay T Graham. A review of facebook research in the social sciences. *Perspectives on psychological science*, 7(3):203–220, 2012.

[5] Instagram. Instagram, 2010. Social media platform for sharing photos and videos.

[6] POSTING ON Instagram. Instagram. *Online unter: https://www. wissenschaft-skommunikation. de/format/instagram*, 2016.

[7] Bo Zhao. Web scraping. *Encyclopedia of big data*, 1, 2017.

[8] Moaiad Ahmad Khder. Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3), 2021.

[9] Ismael Camargo-Henríquez and Yarisel Núñez-Bernal. A web scraping based approach for data research through social media: An instagram case. In *2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)*, pages 1–4. IEEE, 2022.

[10] kevinzg. Facebook scraper. `https://github.com/kevinzg/facebook-scraper`, 2024.

[11] aandergr. Instaloader. `https://github.com/instaloader/instaloader`, 2024.

[12] Bing Liu et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666, 2010.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[14] Ian Tenney, Dipanjan Das, and Ellie Pavlick. Bert rediscovers the classical nlp pipeline. *arXiv preprint arXiv:1905.05950*, 2019.

[15] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[16] Matúš Pikuliak, Štefan Grivalský, Martin Konôpka, Miroslav Blšták, Martin Tamajka, Viktor Bachratý, Marian Simko, Pavol Balážik, Michal Trnka, and Filip Uhlárik. SlovakBERT: Slovak masked language model. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 7156–7168, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.