



Meta Social Media

Carlos de Souza Lima
Leonardo Leite Meira dos Santos

Work carried out under the guidance of
Prof. Paulo Matos

Master in Information Systems
2022

Meta Social Media

Semantic Web Project
Master in Information Systems
Escola Superior de Tecnologia e de Gestão

Carlos de Souza Lima,Leonardo Leite Meira dos Santos
novembro 2022

A Escola Superior de Tecnologia e de Gestão não se responsabiliza pelas opiniões expressas neste relatório.

Certifico que li este relatório e que, na minha opinião, é adequado no seu conteúdo e forma como demonstrador do trabalho desenvolvido no âmbito da UC de Projecto.

Prof. Paulo Matos (Orientador)

Aceite para avaliação da UC de Projecto.

Abstract

This research project aims to explore the concept of social meta-network, as a way of integrating social networks, such as Facebook, LinkedIn and others, providing an integrated perspective and transversal functionalities of added value for its users. The present approach uses graph databases, implemented using Neo4j. The nodes/vertices of the graph are used to represent people; properties on nodes are used to represent personal information of each individual; the relationships/edges represent directional relationships between individuals according to each social network; and the properties about relationships serve to maintain information about the relationship between individuals for each social network. The model created intends to be a mixed solution focused on relationships by type of social network, referring details to the information existing in the native network itself. Native networks are seen as distinct dimensions in which the present solution makes it possible to relate information between these dimensions, even when there are no relationships between the networks/user in question. It also allows for a more complete perspective of each individual, integrating the professional, playful and other more specialized aspects, which leads to results that are much higher than what can be achieved by looking individually at each network - whether for problems/challenges of centrality/influence, identification, and characterization of communities, identification of similarities, identification of potential relationships, among many other approaches. In practical terms, a considerable asset for marketing purposes.

Contents

Abstract	vii
1 Introduction	1
1.1 Goals	1
1.2 Art State	2
2 Tools	3
2.1 Neo4J	3
2.2 Python	3
2.3 React	4
3 Development	5
3.1 Social Medias	5
3.2 Person	5
3.3 Relationships	6
3.4 Queries	7
Bibliography	9

List of Figures

3.1	"Alice" entity with the vertices for your social networks	6
3.2	Entity "Alice" with its social network connecting to other networks	7

Chapter 1

Introduction

Social media is software that is constantly present in a large part of people's lives. Nowadays contents more than 4.7 billion active users, according to [1], which is over 93% of the internet's users, and it tends to grow more until 2027, when this statistic will reach on 5.85 billion active users [2]. So, it presents itself as a market in full growth that contains large companies, such as Meta owner of Facebook, Whatsapp and Instagram and also Microsoft, owner of LinkedIn. More recent emerged TikTok too, company that grown more in 2021 when grown 142% in market value [3].

Therefore, the level of influence and social strength that social media has conquered to this day is perceptible. A common question is how these big companies make their money, as most of them are free. The answer to that question is not so easy and a detailed description of this subject is beyond the scope of this work. By [4] the highest percentage of revenue from these companies comes from AdSense. For this a fundamental resource is the large number of users and also a large dataset of their frequently use, with the objective of the AdSenses to indicate a large number of people.

Otherwise, the user's data that are maintained are also important to another functions, for example to make a recommendation of new connections or and content exhibitions. As a result of this, sometimes the data can be skewed when we consider the actives users objectives at which social media, as an example, in the most part of LinkedIn users the media are used in a professional way, meanwhile the TikTok users have an entertainment objective.

Thus, the project purpose is the construction of a system that adds information about different social medias, considering the objectives interests of every person in each networks, to do the most correct and relevant continent recommendations of connections and exhibition. This work will be followed by a mor detailed presentation of the project, and subsequently by your supplementation and presentation of the user interface idealization. Finally, having the results, they are going to be presented and after having the conclusion.

1.1 Goals

The goal of the present work is to propose a model of database to trace the profile of the users in a multisocial aspect, that is, we are researching and crossing information from different social media, taking from this a more complete profile, which can provide more accurate information. This information would be used to propose connection recommendations between different social media, based on the current connection and the interest of each person and each mapped profile.

Another goal of this work is to use the stored dataset to identify the level of influence of a person on a social media, creating a general and tag influence score, serving as parameters

for choosing influential people for marketing ads on different networks.

1.2 Art State

At some related papers we may see different approaches to do the recommendations of connections in social medias. Accordingly, in [5] were is purpose a new approach to recommendations were are consider some information about the users, for example the localization, and such information are used as entries to one SVM (Support Vector Machine).

We also see an similar approach in [6], a work in which the users information are utilized in a Convolutional Neural Network and offer as an exits better recommendations based in more complex information of the users.

While in [7] the work developed an algorithm of recommendation based in semantic analysis of social medias into a context of learning environments, for this the authors use some semantic web tools which assists at analysis realizations and recommendations.

To conclude in [8] is purposed a recommendation algorithm based on graphs, where is also done the comparison between the purposed algorithm and the traditional recommendation algorithms.

A great difference between those works and the present one becomes from a fact that the most part of them utilize the user's data that comes from a single social media. However much, the implementations or the utilized information vary from work to work, there wasn't find until then one solution that are based un multiples social networks to make those recommendations.

Chapter 2

Tools

2.1 Neo4J

In order to achieve the objectives of the work, the tool chosen to provide the database is Neo4J [9], which provides graph database management and is recommended for use with large volumes of data that maintain a navigational relationship between them. This database model is classified as a non-relational database, but based on graphs, that is, the database is designed on the concept of nodes and edges.

This tool allows mapping different users of different social media and their connection, in order to make extrapolations on the recommendation format. A database in the Neo4J system is composed of four main types of components, when through them it is possible to map and extract information from our databases, these components are:

- Nodes: Equivalent to the vertex of the graph, being the main data within the database. They can contain information such as labels and properties.
- Relationship: Relationships are equivalent to the edges of a graph and describe the relationship between nodes. They can also contain properties.
- Labels: Labels are classifications or types of nodes, being used to categorize and classify the node into a certain type.
- Properties: Attributes of nodes or relationships, they are composed of key and value.

For graph database manipulation, Cypher [10] is used, which is a graph manipulation language in Neo4J. Through this language we can manipulate our database in Neo4J, where we can search, insert, update and perform other functions.

2.2 Python

Python [11] is a general purpose programming language that in the scope of work was used to automate processes and manipulate Neo4J, abstracting the necessary steps and also applying the necessary validations and transformations before performing the insertions. This language is also used to provide the information stored in the database through an API (Application Programming Interface), which makes it possible to create a graphical interface for user interactions.

2.3 React

React is a library for the JavaScript language, used to build graphical interfaces for web platforms. In the scope of this work we will use it to create the graphical interface, which accesses the API to display the information in the web system. The web interface prototype can be accessed through this link.

Chapter 3

Development

The first necessarily step to the development of the work is make the abstractions, for that we classify and separate the entities, relationships and attributes of work. The data base is composed by a couple of entities called by "Social Media" and "Person", a couple types of relationship, the "HAS" to connect one person to a social network, and another relationship to represent connections type between the social networks, as the possible following labels: "PERSONAL", "INFLUENCE", "PROFESSIONAL_RELATIONSHIP", e "CONTENTS".

Such abstractions that were citated anteriorly are codified in a program that automatize the insertion process and population of Neo4J, applying some validations and a defined logic to several connections as the verification of a relationship type, and can be find at the following repositório. To use is necessary substitute connections data of Neo4J to the local machine.

3.1 Social Medias

The Social Media is responsible for represent one social network of one single person. A social network can be connected to many other social media and in just one person. By this way, the social networks that were chosen to be part of the project initially are currently the main utilized, being them the Instagram, Facebook, Ticktok, Youtube, Pinterest e Twitter. The attributes of each social network are

- Identifier: Single identifier of the entity
- Name: Social network name that the entity refers to
- User name: Responsible for identify the user name chosen by the person that created that social network
- Identifier of the user: Responsible for being a reference to a person identifier ("Person" entity")

3.2 Person

The "Person" entity represent the person that utilize determinated social networks. One person can be related to many others different social networks, and yours relations with another people happen through those social networks.

One instance of the "Person" entity have the following attributes:

- Identifier: A single identifier to that user

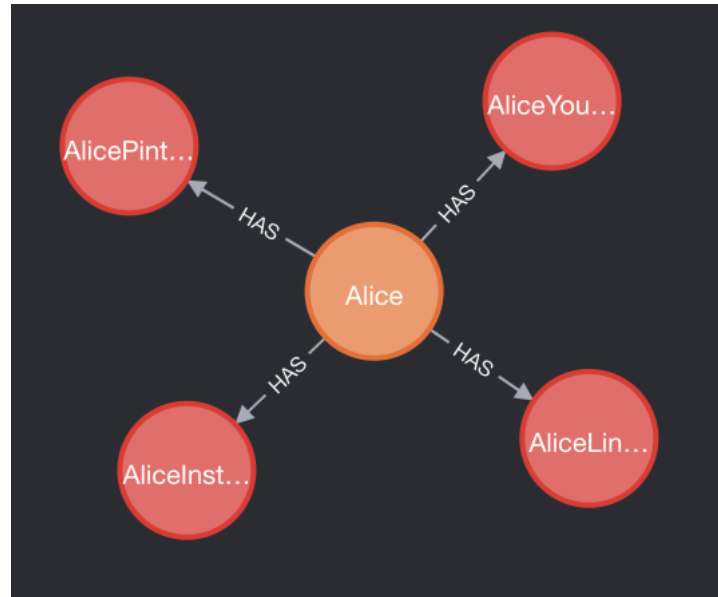


Figure 3.1: "Alice" entity with the vertices for your social networks

- Name: User's first name
- Surname: User's last name

3.3 Relationships

The relationships between the entities describe the behavior that are being charted by the graphs, in this case exist a couple type of present relationship at the system, one to represent the connection between one person and you social networks and another one to represent the connections between the social networks.

The first relation, between one person and four social networks are described for the vertex "Has", each person must be in a social network set. Were the connection are defined based at this relation. The image 3.1 shows the relation of the "Alice" person with your social networks through the vertice "HAS".

The other type of relationship would be one that describes the connection between two social networks of different people. This relationship takes into account the type of connection and the social networks used by two users, and can be of the following types:

- Personal: Follow each other on networks for personal use
- Professional: Follow each other in professional networks
- Content consumption - Follow each other on networks focused on video content
- Influence: Represents the relationship of influence, when a person does not follow another but is followed by him. The vertical arrow points to who is followed, that is, who influences the relationship.

In figure 3.2 it is possible to visualize the entity "Alice" linked to the social network "AliceInstagram", and its relationship with other social networks of other people.

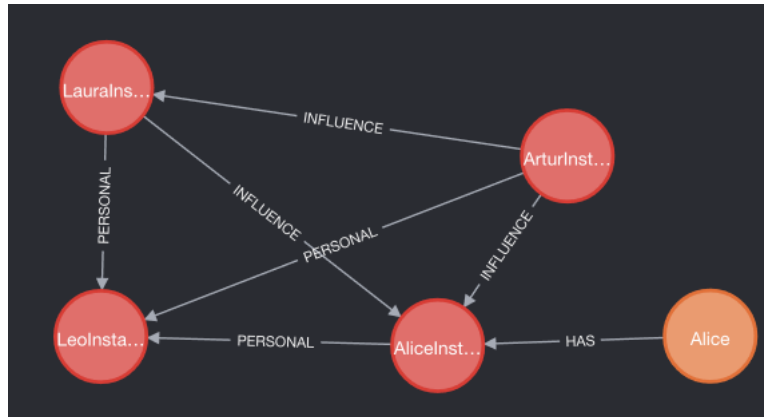


Figure 3.2: Entity “Alice” with its social network connecting to other networks

3.4 Queries

Given the database structure presented and the objectives of this work, some queries were created in Cypher to return the desired information. The query below returns the social media of a given person. In the example, you are looking for a person’s social networks with the “Name” property as “Laura”.

```
match(person:Person {Name: "Laura"})-[HAS]->(socialMedia:SocialMedia)
return socialMedia
```

The next query is intended to return all people connected to someone else’s social networks. In the example, the query is looking for people connected to a person’s social networks with the “Name” property as “Laura”.

```
match(p:Person{Name:"Laura"})-[]-(s:SocialMedia)
match(s)-[]-(s1:SocialMedia)
match(s1)-[HAS]-(f:Person)
return f
```

The following query looks for the first-level connections for a relationship type. In the example, it searches for people who have a social network (in this case, instagram), then it searches for the social networks connected in a type of relationship (in the example, the relationship is “PERSONAL”), and finally, it searches for the social networks connected to a person (in the Leonardo case). With this query it is possible to list all connections for a type of relationship that has been identified.

```
match (p:Person)-[]-(s:SocialMedia{Name:"Instagram"})
match (s)-[f:PERSONAL]-(s1)
match (s1)-[:HAS]-(se:Person{Name:"Leonardo"})
return p
```

The following query performs an in-depth search using the plugion APOC [?] available in the Neo4J system. In the first line, the social network with “Name” defined as “Instagram” of the person “Alice” is selected. In the second line, the search for node S (Alice’s Instagram) begins, then the search for vertices is performed, with a maximum depth of 5. The result of this query are the suggestions of possible new connections of a personal nature that can be made .

```

match (p:Person{Name:"Alice"})-[]-(s:SocialMedia{Name:"Instagram"})
call apoc.path.spanningTree(s,{relationshipFilter: "PERSONAL",minLevel:0,maxLevel: 5})
YIELD path
RETURN path, p;

```

For other types of recommendations just replace the "relationshipFilter" property to "INFLUENCE", to have influence recommendations, to "PROFESSIONAL_RELATIONSHIP" to have professional connection recommendations, or "CONTENTS" to have content recommendations.

The following query is intended to return the influence score that a given person has. In the example below, all of a person's social networks are selected, in this case "Alice", and then the number of influence connections is returned.

```

match(person:Person {Name: "Alice"})-[HAS]->(socialMedia:SocialMedia)
match(socialMedia)<-[f:INFLUENCE]-()
return count(f)

```

This last query aims to identify someone's level of influence within a network. In the first line, the social networks of a person are selected to be evaluated for influence. In the second line, all the social networks of a network that must be evaluated, which are influenced are selected, and finally it is defined which network the level of influence is being evaluated, in this case a social network with the property "Username" with the value "LauraInstagram".

```

match(person:Person {Name: "Leonardo"})-[HAS]->(socialMedia:SocialMedia)
match(socialMedia)<-[f:INFLUENCE]-(socialMediaInOtherSubGraph:SocialMedia)
match(socialMediaInOtherSubGraph:SocialMedia)-
    []-(subGraph:SocialMedia{Username:"LauraInstagram"})

return count(f)

```

Bibliography

- [1] Statista, “Number of internet and social media users worldwide as of july 2022,” 2022. <https://www.statista.com/statistics/617136/digital-population-worldwide/>, (accessed: 13.11.2022).
- [2] Statista, “Number of social media users worldwide from 2018 to 2027.” <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>, (accessed: 13.11.2022).
- [3] B. of Apps, “Tiktok revenue and usage statistics (2022).” <https://www.businessofapps.com/data/tik-tok-statistics/>, (accessed: 13.11.2022).
- [4] G. MCFARLANE, “How facebook (meta), twitter, social media make money from you.” <https://www.investopedia.com/stock-analysis/032114/how-facebook-twitter-social-media-make-money-you-twtr-lnkd-fb-goog.aspx>, (accessed: 13.11.2022).
- [5] M. Xin and L. Wu, “Using multi-features to partition users for friends recommendation in location based social network,” *Information Processing & Management*, vol. 57, no. 1, p. 102125, 2020.
- [6] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang, “Self-supervised multi-channel hypergraph convolutional network for social recommendation,” in *Proceedings of the Web Conference 2021, WWW ’21*, (New York, NY, USA), p. 413–424, Association for Computing Machinery, 2021.
- [7] A. Khaled, S. Ouchani, and C. Chohra, “Recommendations-based on semantic analysis of social networks in learning environments,” *Computers in Human Behavior*, vol. 101, pp. 435–449, 2019.
- [8] L. Dong, Y. Li, H. Yin, H. Le, and M. Rui, “The algorithm of link prediction on social network,” *Mathematical Problems in Engineering*, vol. 2013, p. 125123, Sep 2013.
- [9] Neo4J, “Neo4j graph database.” <https://neo4j.com/cloud/platform/aura-graph-database/?ref=nav-get-started-cta>, (accessed: 13.11.2022).
- [10] Neo4J, “Cypher query language.” <https://neo4j.com/developer/cypher/>, (accessed: 13.11.2022).
- [11] Python, “Beginner’s guide to python.” <https://wiki.python.org/moin/BeginnersGuide>, (accessed: 13.11.2022).