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
2023

Meta Social Media

Carlos de Souza Lima,Leonardo Leite Meira dos Santos
janeiro 2023

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. 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

List of Figures

x

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 Adenses 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 Organization

This document is separated into objectives to be achieved with the development of the project, then the tools used in the development are explained, and finally the technical specification of the database and API.

1.2   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. Based on this objective, it will have a parameter to select people with a higher level of influence among social networks (or some networks in particular) for marketing campaigns and partnerships between content producers and companies, thus improving content and marketing segmentation.

Beside with the database structure, an API is also available to access both the connection recommendations and the influence score. To use this API, it is necessary to register a user to receive the access token, and then access the endpoints that provide the information.

1.3   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

This section is used to describe the tools used in the development of the project, going through the structuring of the database and the construction of the API.

2.1 Protégé

Protégé [9] is a free, open source ontology editor and a knowledge management system, was developed by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. The software is the most popular and widely used ontology editor in the world, works in Windows, Linux, and MacOS.

This tool allows mapping different users of different social media and their connection, and and build queries to search the database. The details of entities and relationships are described in chapter X.

2.2 Python

Python [10] is a general purpose programming language that in the scope of work was used to automate processes and manipulate databse, 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. The framework used to build the API was FastAPI [11].

Chapter 3

Conceptual solution

3.1 Entities

System data is based on social medias entities (SocialMedia), users (User), and the connection entity (Connection), each with its own properties.

3.1.1 Social Medias

The Social Media is responsible for represent one social network. A social network can be connected with a user and connection. 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, LinkedIn, Ticktok, Youtube, Pinterest e Twitter.

The attributes of each social network are:

- Id: Single identifier

- Name: Social network name that the entity refers to

- Description: Text that represent a description of social media

3.1.2 User

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

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

- Identifier: A single identifier to that user

- Email

- Initial date: Date the user signup into the system

- Updated date: Update date of the information in the system

3.1.3 Connection

The connection entity represents the connection between two users of a social network. A connection is linked to two different users, representing the relationship between them, and also to a social media, storing where the relationship between those users takes place. The connection entity has the following properties:

- Identifier: A single identifier to that connection

- Initial date: Date the connection was registered in the system

- Final date: Date the connection was broken

- Relationship type: Represents the type of relationship between those two users

The relationship type 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. This type is used for influence score.

3.2 Relationships

The relationships between the entities describe the behavior that are being charted by the graphs. Within the proposed system there are three types of relationship that are described as follows:

- Relationship between a user and the social media: Vertex with the properties username, date joined, status, date deleted (optional).

- Relationship between user and connection: Used to store connection information between two users in a social network.

- Relationship between connection and social media: Used to represent in which social media the relationship between the two users exists

3.3 Queries

Given the database structure presented and the objectives of this work, some queries were created to return the desired information. The querie below returns the social media of a given person.

```
SELECT ?socialName
WHERE {
 ?subject rdf:type :User;
         :email ?name ;
         :id ?id ;
         :IsOwnerOf ?social .
  ?social :name ?socialName
  FILTER (?id = 2)
}
```

The next query is intended to return all people connected to someone else's social networks. In this example is filtered by type "Personal".

```
SELECT DISTINCT ?email ?id ?idConnection ?connectionEmail ?type
WHERE {
?subject rdf:type :User;
         :email ?email ;
         :id ?id ;
         :HasRelationship ?relations .
?relations :With ?connection .
?relations rdf:type ?type .
?connection :id ?idConnection ;
            :email ?connectionEmail
FILTER (?id = 1 && ?type = :Personal)
}
```

The following query looks for the first-level connections for a relationship type and return has recommendations.

```
SELECT DISTINCT ?recommends
WHERE {
 ?subject rdf:type my:Person;
          my:PersonID ?id ;
          my:HAS ?social;
         OPTIONAL {
            ?type rdfs:subPropertyOf* my:RELATIONSHIP .
            ?social ?type ?conections .
            ?social ?type ?recommends .
            FILTER NOT EXISTS {?recommends ?type ?social}
          }
   FILTER ( ?id = 1 && !isBlank(?recommends) )s
}
```

The following query is intended to return the influence score that a given person has.

```
SELECT (count ( distinct ?relations ) AS ?level)
WHERE {
    ?relations rdf:type :Influence .
    ?relations :With ?user .
    ?user :id ?id .
    FILTER (?id = 4)
}GROUP BY ?id
```

3.4   API

To use the system [12], it is necessary to create an access, and log in to the platform. In this way, it is possible to have access to endpoints that return recommendations and level of influence. Existing endpoints are:

- List Social Medias of a user: This endpoint receives in the path parameter the id of the user that must be listed the social medias. Once the user exists in the database, a list of social media names referring to the informed user will be returned.

- A user's connections: This endpoint returns a list of people who are connected with a given user. The id of a user to be searched for is a mandatory parameter (path parameter), and as optional parameters (query parameter) we have the type of connection,

and a certain social media. If any mandatory parameter is not informed, the response will contain all possibilities.

- Connection suggestions: This endpoint returns a list of possible connections for a given user whose id is informed as a parameter (path parameter). As optional parameters (query parameter) there are the type of connection, and a certain social media, to suggest connections only from a certain context. The return from this endpoint is basically a search in the graph of connections that the user's connections make (small in-depth search).

- Level of influence: This endpoint returns the level of influence of a given user whose id is passed as a parameter (path parameter). As optional parameters (query parameter) there are a specific social media, and a sub-graph that would be someone else's social network (user id). This calculation is based on the number of influence vertices that arrive in a user's social media, which can be specifically from a social media, or from another user's social network, making it possible to analyze someone's level of influence within another context.

- Create access: Endpoint to create access to API. It receives in the body of the request in JSON (Java Script Object Notation) format, email, name and password.

- Login: Receive formdata with username (email) and password, and them return a JWT token with 24 hours validity

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] N. I. of General Medical Sciences., "A free, open-source ontology editor and framework for building intelligent systems." `https://protege.stanford.edu/`, (accessed: 07.01.2023).

[10] Python, "Beginner's guide to python." `https://wiki.python.org/moin/BeginnersGuide`, (accessed: 13.11.2022).

[11] S. Ramírez, "Fast api." `https://fastapi.tiangolo.com/`, (accessed: 07.01.2023).

[12] L. L. . C. de Souza Lima, "Meta social media api."