Database Support for Online Social Network Analysis

Paulo Pintor[1], Mariana D. A. Salgueiro[2], Sergio Lifschitz[2], Daniel Schwabe[2], José Moreira[1], Rogério Costa[3]

[1]Department of Electronics, Telecommunications and Informatics (DETI)
University of Aveiro – Aveiro, Portugal

[2]Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brasil

[3]CIIC – Polytechnic of Leiria
Leiria, Portugal

{paulopintor,jose.moreira}@ua.pt,

{msalgueiro, sergio}@inf.puc-rio.br,

dschwabe@gmail.com, rogerio.l.costa@ipleiria.pt

Abstract. *Online Social Networks (OSNs) have become part of our daily life, contributing to the vast amount of data generated online. The OSNs' data contribute to obtaining information about our society in various areas, such as misinformation, politics, marketing, and engagement metrics. OSN is a comparatively new concept, without a generic model, with problem-specific models used for the studies of it. This article demonstrates the use of logical models to analyze data from different OSNs, based on a generic conceptual model that reflects a general definition of social networks. These examples aim to show the feasibility and advantages of using the generic model as a basis for specific OSN models, to support the analyses performed from the data extracted.*

1. Introduction

OSNs have become a major means of communication and information dissemination among people. As a digital phenomenon, they generate a vast amount of data that is being leveraged to better understand and eventually influence social processes. Prior to OSNs, "traditional media" (e.g., newspapers, radio, tv news) were eminently the broadcast media, where a small minority of people (the so-callled "gatekeepers") [Thorson 2008, Mccombs 2008] were responsible for information dissemination regarding events taking place locally, nationally and internationally.

OSNs have changed that paradigm, allowing everyone to keep current and express a position about any topic. In [Yum 2020, Yousefinaghani et al. 2021, Bamiro and Assayad 2021, Vishwakarma and Chugh 2023], the authors used data from social networks to analyse people's opinions and apprehensions about political measures. Other good examples are elections and studies on the electoral polls [Ansari et al. 2020].

The data can also be used to study phenomena created through social networks, one of them being the growth and dissemination of misinformation [Belloir et al. 2022, Rath et al. 2020, Burbach et al. 2019]. Another phenomenon associated with OSNs (and their underlying algorithms) are filter bubbles[Matakos et al. 2020].

The literature shows previous analyses of social phenomena using OSN data, such as user engagement and content influence [Zhang et al. 2017, Fan et al. 2018a, Fan et al. 2018b, Fan et al. 2018c], but further examination shows that typically each study has created a specific model to answer only its objects of interest.

In this article, we demonstrate that typical OSNs analysis can be effectively supported by DBMSs whose logical schemas are derived from an extended common conceptual schema for OSNs.

The next section shows different conceptual models and their differences, justifying our choice. In section 3, we summarize the OSN definition and in 4, we explain the conceptual model. Section 5 shows the derivation of logical models from the conceptual model and examples of representative analyses that can be carried out through the resulting DB. The last section presents our conclusions.

2. Related Works

A literature review on OSNs shows the use of a variety of conceptual models, presenting four different models, with benefits and limitations, and justifying the model chosen and why.

Semantic-Interlinked Online Communities (SIOC) Ontology

The Semantic-Interlinked Online Communities, also known as SIOC [Breslin et al. 2005, Passant et al. 2010], is one of the first attempts to create an ontology (created in 2005) to unify and allow interoperability between social communities. SIOC aimed to allow browsing and querying over the data of the different social communities, almost as a search engine, by querying a representation of this data using the Resource Description Framework (RDF).

The SIOC ontology has six main entities to describe social communities, and everything is centered on the class *User*. The other important class is *Site*, representing a community or a group of communities, and the *Forum* is linked to it, representing the discussions composed by *Post*. All of these entities are also connected to the user. A *User* can participate in a *Event* or be part of a *Group*. There is an entity **Document** to support the posts having attachments.

Despite its update to help describe new social networks with new parameters, such as likes, mentions or shared_by, its main drawback is its applicability since in order to use it the communities need to migrate to its semantic representation to allow browsing.

SocIoS Core Ontology

The SocIoS Core Ontology [Tserpes et al. 2012] is an ontology built to support the SocIoS framework [Kardara et al. 2015]. The framework aimed to operate on top of multiple social networks, allowing uniform access to data, providing a single access point to aggregate network data and a set of analytical tools to explore them. The uniform data access aims to take advantage of the conceptual similarity between the different core social networks.

The ontology has two central entities: *Person* and *MediaItem*. A *Person* who belongs to a *Social Network* can connect with another *Person*, send *Messages*, and receive replies. The *MediaItem* is published and owned by a *Person*, and *Person* generates *Activities* (related with *Message*, *MediaItem* and *Event*). A *Person* can also rate *MediaItem*, and for that, there is a class called *Rating*. A *Group* has different *Persons* as members.

SocIos is a conceptual model created for a specific API. The API aims to collect data from different OSNs. The project's repository on GitHub has last changed ten years ago, while the OSN APIs have been constantly changing, and thus this API has become outdated.

sonSchema Conceptual Model

In [Bao et al. 2013] a conceptual model for social networks called sonSchema was proposed, following two guidelines: *Generality* and *Service-Orientation*. The first means that the model must be generic enough to support any network, and the latter states that the entities and relationships must match the social networks' activities and services.

The authors suggest four entities for the model: *user*, *group*, *post* and *social_product*. The last one is used to define events or some products that can be sold on social networks. Among entities there are four relationships: **friendship**, **membership**, **social_product_activity** and **social_product_relationship**. The **friendship** connects the *users*, **membership** connects the *user*s to a *group*, **social_product_activity** links a *user* to a product through an activity, and **social_product_relationship** correlates two different products.

The authors also mention one more activity and relationship separately, because interactions create them: *private_msg* and **response2post**. *Private_msg* is an entity that represents the private messages between users, and **response2post** is a relationship that can represent comments, for example. This model is not generic enough to represent relevant current OSN concepts, such as quotes, sharing, user mentions, reactions, and following hashtags. These models are relevant examples of approaches that tried to create a conceptual model for different social networks, falling short, however, in answering actual social network specificities or representing social networks as a whole.

3. Digital Social Networks Definition

This section will briefly present the work done in [Authors Hidden]. The authors presented a generic and minimalist conceptual schema. It includes definitions and entities common to most social networks.

The generic minimalist schema should, in use, be extended depending on the particular OSN and the types of analyses to be performed. Given an extended schema for a specific OSN, it can allow a DBMS to use that logical schema to support the specific analysis of some of its characteristics.

The primary goal of OSN is to allow people to communicate socially, mostly within groups of friends but also within (in)formal groups for discussing subjects of common interest. Therefore, the Online Digital Social Networks definition considers four essential dimensions: 1) Types of Targetless Communication; 2) Types of Connection; 3) Explicit Non-textual Reactions; 4) Tagging of Content.

Social networks have empowered users by enabling them to express and share their thoughts and opinions through their posts. The communication is spontaneous and not targeted, and a user can potentially reach everyone [Taprial and Kanwar 2012]. Communication can be accomplished through the timeline (e.g. Twitter Feed), non-private communication (e.g. "Twittersphere") or sharing (e.g. Twitter retweet).

Some social networks have the notion of a timeline. The user's timeline shows the user's posts chronologically [Wang et al. 2021], and the platform timeline where an algorithm groups the posts. Non-private communication includes the replies to content, allowing a dialogue in near real-time [Taprial and Kanwar 2012]. Sharing is the mechanism whereby the users can manifest their opinion about content by sharing it and expressing their position. A user's shares can reach different media and people.

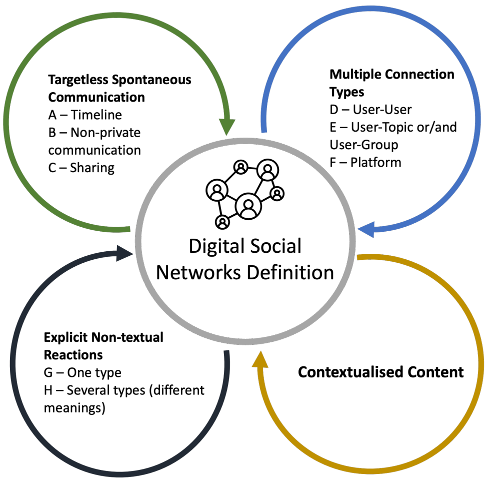

Figure 1. Summary of the Digital Social Networks Dimensions

Users connect through multiple connection types: user-user, usertopic and/or user-group and platform. The first is the typical social network connection, where a user can be a friend or following another user. The user-topic or user-group explicates the desire to connect through a common interest. The platform connection is a hidden type of connection, presenting how the platforms create indirect connections between the users, since their timeline will include content that is sponsored or posts from other users that the platform considers important. Even if a user does not establish any connections, they will still be indirectly connected to other users since their timeline will include sponsored content or posts from other users that the platform considers important.

Explicit non-textual reactions began with the "like" button, a hitherto unavailable functionality through which the user could react non-verbally to content, the semantics of which are vague and sometimes insufficient or inappropriate. Some platforms allow users to react to content with a variety of emojis and emoticons for better self-expression. Therefore the explicit non-textual reaction dimension indicates the variety of reactions enabled by OSNs.

The last dimension is contextualized content and concerns associating hashtags or topics to the content. The users can contextualize the content so that others know the topic of content, and some platforms also allow users to follow the topics or hashtags that represent things of interest to them. Figure 1 summarizes these dimensions.

4. Generic Conceptual Schema

Figure 2 illustrates the proposed conceptual schema, where it is possible to identify 3 main entities: *Account*, *Social Media Content (SMC)*, and *Group*. *Accounts* represent individuals or groups of people, that can generate some type of *SMC*. *SMCs* include combinations of texts or media (images, videos, animated gifs, etc...) and may **contain** *hashtags*. In some OSNs, Accounts are able to follow hashtags. It is also possible to create content within *Groups*, where users with common interests can gather.

There is the possibility of enriching data for both *Accounts* and *SMCs*, by includ-

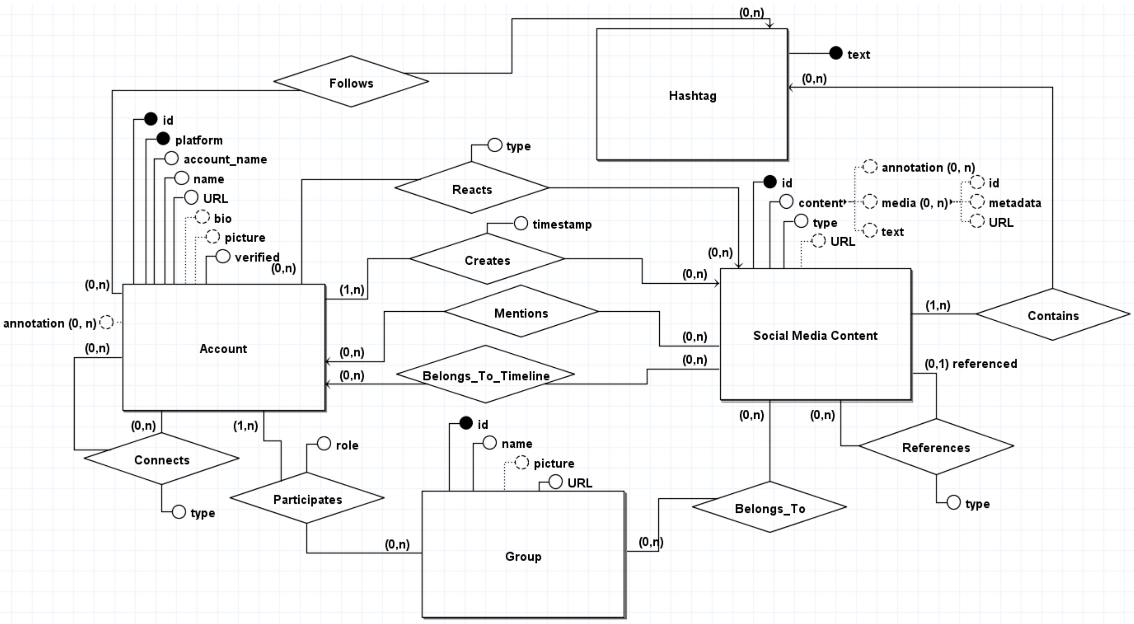

Figure 2. Generic Conceptual Schema

ing aspects retrieved from other sources or from a diversity of analyses, giving them an **annotation** attribute, representing information obtained in addition to the data initially collected. When modelling OSNs, observing the way entities relate to each other is vital. *Accounts* can **connect** to each other by friendship or following, or through their published *SMC*.

Accounts can create *SMCs* and react using "likes" or emojis. Alternatively, *SMCs* can mention accounts, indicating that the content is directed to that Account. Another way to direct content is by **referencing** it, by *replying, sharing,* and *quoting*, which can only be done to a single *SMC* at a time.

Apart from connections formed directly between *Accounts* and indirectly from content referencing, it is possible to **participate** in *Groups* and to **follow** *Hashtags*. In these cases, *Accounts* are connected by the theme or subject of the *Group* or *Hashtag*.

Another characteristic of OSNs is the concept of a timeline ("feed"). In the schema, the timeline is represented by the idea of an *SMC* **belonging to** a feed, published in a *Group*'s or an *Account*'s timeline, varying by its type and the particular OSN. A timeline can show all *SMCs* created by the user or just those published on its own timeline page. Contents can also **belong to** a single feed or to many. For example, *Accounts* are able to publish in someone's timeline and have the same *SMC* appear in theirs too.

5. Mapping to Logical Schemas

This section presents examples of specific conceptual schemas for two distinct OSNs and possible derivation of logical schemas, with examples of analyses for one of the OSN.

The relational logical schema for both OSNs was made using the following steps: 1) Conversion of entities and their attributes to tables and columns, respectively. Multivalued attributes are converted to tables, 2) Conversion of relationships and their attributes following cardinality's rules, 3) Conversion of generalizations/specializations using one table for the whole hierarchy.

5.1. Reddit

Figure 3 shows the conceptual schema of Reddit based on the generic schema, shows in blue new attributes, relationships and entities that needed to be created to represent the network fully. Two important new entities were introduced: the **Post** and **Comment** specializations.

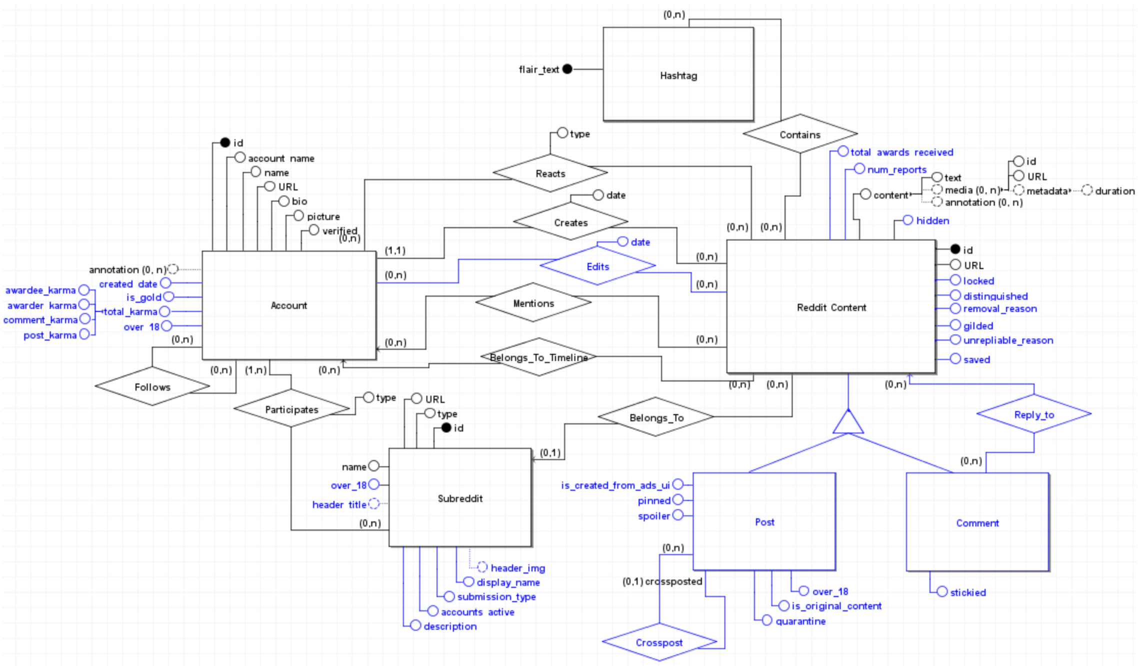

Figure 3. Reddit's Conceptual Schema. Elements in blue are specific extensions to the generic schema.

Posts are the topic of a discussion, while *comments* are replies to this topic or other comments. Only *posts* can be referenced by share (*crosspost*);flagged (as for example, *over 18*, *spoiler*, *pinned*). *Comments* can only be flagged as necessary (*stickied*) to a *post*. The entities *Account, Group* (Subreddit in this case) and *Reddit Content* have additional attributes and a relation **Edit** between *Account and Reddit Content* have been added. Another critical difference is that Reddit has timelines both for *Accounts* and *Subreddits*, which means that *Reddit Content* can belong to both or either. Table 1 shows the derived relational logical schema for Reddit's conceptual schema.

5.2. Twitter

Figure 4 presents the Twitter conceptual schema based on the generic schema shown in section 3. The new attributes and relationships which had to be created for a complete OSN representation are again illustrated in blue. Since it is a specific schema, we renamed the *Social Media Content* entity to *Tweet*. *Account* and *Tweet* have new attributes, and the entity *Group* does not exist because Twitter has no open groups.

The relationship **Reacts** is now called **Likes** because, in this OSN, it is only possible to react to content with a single type of reaction. Since, on Twitter, users cannot

Table 1. Reddit's Relational Logical Schema

Account (id, account_name, name, picture, verified, is_gold, awardee_karma, comment_karma, post_karma, awarder_karma, bio, created_date, over_18, URL)
Follows (*account1_id*, *account2_id*)
Accout_Annotation (*account_id*, *annotation_id*)
Annotation (id, annotation)
Reddit_Content (id, gilded, locked, URL, total_awards_received, hidden, date, distinguished, unrepliable_reason, saved, removal_reason, text, num_reports, stickied, is_original_content, spoiler, is_created_from_ads_ui, quarantine, over_18, pinned, type, *subreddit_id*, *account_id*)
Reddit_Content_Annotation (*reddit_content_id*, *annotation_id*)
Reacts (*reddit_content_id*, *account_id*, type)
Mentions (*reddit_content_id*, *account_id*)
Belongs_To_Timeline (*reddit_content_id*, *account_id*)
Belongs_To (*reddit_content_id*, *subreddit_id*)
Subreddit (id, header_title, header_img, submission_type, type, URL, over_18, name, description, display_name, accounts_active)
Participates (*account_id*, *subreddit_id*, type)
Edits (*account_id*, *reddit_content_id*, date)
References (*reddit_content_references*, *reddit_content_referenced*, type)
Hashtag (id, text)
Contains (*hashtag_id*, *reddit_content_id*)
Media (id, URL, duration, *reddit_content_id*)

follow *Hashtag*, the Hashtag entity type has been changed into a multiple-value attribute of a Tweet. The notion of hashtags can exist in different ways depending on the social media platform. Hence, this concept behaves differently on different platforms, justifying its existence as an attribute and entity.

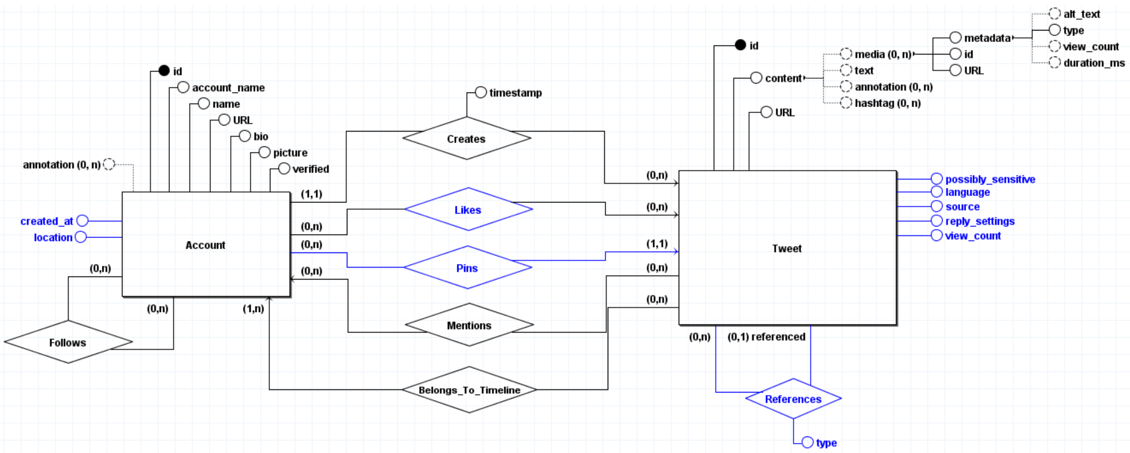

Figure 4. Twitter's Conceptual Schema

The transformation of Twitter's conceptual schema into a relational logical schema is presented in Table 2. The conceptual schema has also been transformed into a graph logical schema, shown in Figure 5.

Table 2. Twitter's Relational Logical Schema

Account (id, account_name, name, URL, bio, picture, verified, created_at, location, following, followers)
Follows (*account1_id*, *account2_id*)
Accout_Annotation (*account_id*, *annotation_id*)
Annotation (id, annotation)
Tweet (id, possibly_sensitive, URL, reply_settings, source, language, text, view_count, *account_id*, timestamp, like_count, quote_count, reply_count, retweet_count)
Tweet_Annotation (*tweet_id*, *annotation_id*)
Likes (*tweet_id*, *account_id*)
Mentions (*tweet_id*, *account_id*)
Belongs_To_Timeline (*tweet_id*, *account_id*)
Hashtag (id, hashtag)
Tweet_Hashtag (*tweet_id*, *hashtag_id*)
References (*tweet_id_references*, *tweet_id_referenced*, type)
Media (id, type, URL, alt_text, duration_ms, *tweet_id*)

5.2.1. Examples of analysis in OSNs supported by logical schemas

To illustrate how different logical schemas can be used in DBMSs implementations storing OSN data, two cases of the Twitter OSN are presented. There are two main types of analysis carried out on Twitter: network analysis and content analysis. Network analysis portrays the structural patterns existing within a network by studying the various types of relations within and content analysis is focused on characterizing the content of communications by analyzing the discourse in generated content.

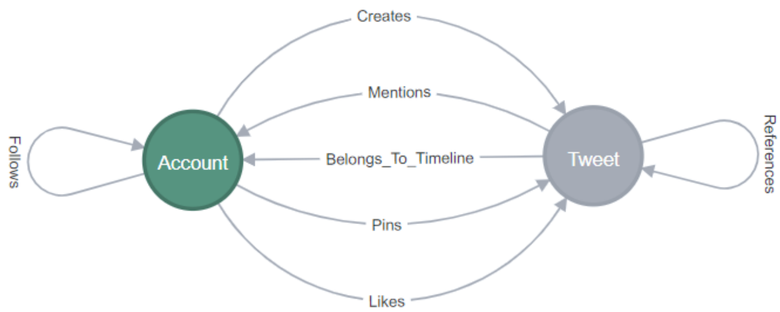

Figure 5. Twitter's Graph Logical Schema

Network Analysis: Retweets Path

Given a tweet and a chain of successive retweets, the current Twitter API does not include the connection between intermediate users in the chain, and the data reported disregards it. We now present functions that show how to obtain possible retweet paths based on a user's follower network.

The function shown in Listing 2 in SQL and the graph query in Cypher in Listing 1, implement the algorithm to find all the paths of followers between the au-

Listing 1. Function in Cypher that returns all the paths between the author of a tweet and the users who retweeted it

```
MATCH (a:Account)-[c:Creates]->()-[r:References {type: 'Retweet'}]->(t:Tweet {id: 'tweet_1'})<-[c1:Creates]-()
WITH DISTINCT (a)
MATCH c=(a)-[r:Follows*]->()-[c2:Creates]->(t:Tweet {id: 'tweet_1'})
return [n in NODES(c) where not n:Tweet | n.id] as possible_rt_path
```

thor of a tweet and all the users who retweet it. Both logical schemas can be used and choosing one among them depends on several possible criteria, as discussed in [Wycislik and Warchal 2014] and [Almabdy 2018], where the authors compare both schemas. It is not our goal to discuss which one is better suited, but rather to illustrate that our approach enables either choice equally.

Listing 2. Function in SQL that returns all the paths between the author of a tweet and the users who retweeted it

```
create or replace function possible_rt_path(retweeters text[], author text) returns table (accounts_path text[]) as $$
declare
  i text;
begin
  FOREACH i IN ARRAY retweeters
  loop
    return query
    with recursive graph_cte (fk_account_id, fk_account_id_, path) as (
      select fk_account_id, fk_account_id_, ARRAY[fk_account_id] as path from follows where fk_account_id = i
      union all
      select nxt.fk_account_id, nxt.fk_account_id_, array_append(prv.path, nxt.fk_account_id) from follows nxt, graph_cte prv where nxt.fk_account_id = prv.fk_account_id_ and nxt.fk_account_id != ALL(prv.path) )
    select array_append(path, fk_account_id_) from graph_cte where fk_account_id_ = author;
  end loop;
end;
$$ language plpgsql
select * from possible_rt_path((select * from retweeters('1')), (select * from author('1')))
```

Content Analysis: Frame Identification

In the context of political communication, *frames* refer to socially constructed interpretative frameworks that allow people to make sense of events and situations. In [Magalhães Firmino 2022], an automatic classification of tweets using a dictionary-based computational method is performed in order to automatically identify generic and specific *frames* present in collected tweets.

The author performed frame identification using the Python programming language over more than 31 million tweets collected between March 15 to June 15 2020. Each tweet is classified with a specific frame label, based on a set of keywords and expressions collected in a dictionary. A tweet can evoke none, one or several frames in its text.

By leveraging the relational logic schema in Table 2, it was possible to create a function in SQL defined in Listing 3 which efficiently performs the automatic identification previously done in Python. The function in Listing 3 receives the list of keywords and expressions representative of a specific frame as a parameter and returns all the tweets containing any of these words.

Listing 3. Frame analysis function 1

```
create or replace function frame_analysis(frame_keywords text[])
    returns table (tweet_id text, texto text, frame_keyword text) as
    $$
declare
  i text;
begin
  FOREACH i IN ARRAY frame_keywords
  loop
    return query
    SELECT id, text, i FROM tweet t WHERE lower(text) ~* ('(?<!-|¿)'
        || i || '(?!-|¿)');
  end loop;
end;
$$ language plpgsql
select * from frame_analysis(array['apos', 'a medida que', 'quanto
    maior', 'quanto mais', 'quanto menor', 'quanto menos', 'cria', '
    adiant', 'obrigad', 'devido a', 'devido ao', ...])
```

6. Conclusion

In this article, we have addressed the open issue of database support for the analysis of OSNs and employing DBMSs to support such analyses, the actual implementation of which have not been detailed in existing literature. Furthermore, as argued in [Authors Hidden], not even a consensus exists over what are exactly OSNs.

We have proposed an approach whereby a generic conceptual schema for OSNs is extended for a particular OSN, and this extended shcema is used to derive several alternative logical schemas. We have illustrated how typical kinds of analysis carried out over OSNs can be supported by queries over different DBMSs, each of which employing a different logical schema - relational and graph-based. For reference, a GitHub[1] repository containing the logical and physical schemas and the functions mentioned for each analysis.

Having a common conceptual schema facilitates the reuse of DBMS infrastructure for repeated analyses over the same OSN, as well as allows different integration approaches when analyzing data over multiple OSNs. It also enables the integration of

[1]https://github.com/marianadsalgueiro/database-suppor-for-OSNA

different analyses over the same OSN, each being carried out using a different logical schema. The study and evaluation of these alternatives are the focus of future research.

References

Almabdy, S. (2018). Comparative analysis of relational and graph databases for social networks. In *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, pages 1–4.

Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., and Singh, K. P. (2020). Analysis of political sentiment orientations on twitter. *Procedia Computer Science*, 167:1821–1828.

Bamiro, B. and Assayad, I. (2021). Data-based automatic covid-19 rumors detection in social networks. In *Proceedings of the 4th International Conference on Networking, Information Systems & Security*, NISS2021, New York, NY, USA. Association for Computing Machinery.

Bao, Z., Tay, Y., and Zhou, J. (2013). sonschema: A conceptual schema for social networks. In *Conceptual Modeling: 32th International Conference, ER 2013, Hong-Kong, China, November 11-13, 2013. Proceedings 32*, pages 197–211. Springer.

Belloir, N., Ouerdane, W., and Pastor, O. (2022). Characterizing fake news: A conceptual modeling-based approach. In Ralyté, J., Chakravarthy, S., Mohania, M., Jeusfeld, M. A., and Karlapalem, K., editors, *Conceptual Modeling*, pages 115–129, Cham. Springer International Publishing.

Breslin, J. G., Harth, A., Bojars, U., and Decker, S. (2005). Towards semantically-interlinked online communities. In Gómez-Pérez, A. and Euzenat, J., editors, *The Semantic Web: Research and Applications*, pages 500–514, Berlin, Heidelberg. Springer Berlin Heidelberg.

Burbach, L., Halbach, P., Ziefle, M., and Calero Valdez, A. (2019). Who shares fake news in online social networks? In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '19, page 234–242, New York, NY, USA. Association for Computing Machinery.

Fan, J., Qiu, J., Li, Y., Meng, Q., Zhang, D., Li, G., Tan, K.-L., and Du, X. (2018a). Octopus: An online topic-aware influence analysis system for social networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1569–1572.

Fan, J., Qiu, J., Li, Y., Meng, Q., Zhang, D., Li, G., Tan, K.-L., and Du, X. (2018b). Octopus: An online topic-aware influence analysis system for social networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1569–1572.

Fan, J., Qiu, J., Li, Y., Meng, Q., Zhang, D., Li, G., Tan, K.-L., and Du, X. (2018c). Octopus: An online topic-aware influence analysis system for social networks. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1569–1572.

Kardara, M., Kalogirou, V., Papaoikonomou, A., Varvarigou, T., and Tserpes, K. (2015). Socios api: A data aggregator for accessing user generated content from online social networks. In *WISE Workshops*, volume 9051.

Magalhães Firmino, L. (2022). *Comunicação Política, métodos computacionais e pandemia: os três primeiros meses da Covid-19 no Brasil e seu processo de enquadramento no Twitter*. PhD thesis, Pontifícia Universidade Católica do Rio De Janeiro PUC-Rio.

Matakos, A., Aslay, C., Galbrun, E., and Gionis, A. (2020). Maximizing the diversity of exposure in a social network. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1.

Mccombs, M. (2008). *Setting the agenda: The mass media and public opinion*.

Passant, A., Bojārs, U., Breslin, J. G., and Decker, S. (2010). The sioc project: Semantically-interlinked online communities, from humans to machines. In Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V. T., Matson, E., and Polleres, A., editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, pages 179–194, Berlin, Heidelberg. Springer Berlin Heidelberg.

Rath, B., Gao, W., and Srivastava, J. (2020). Evaluating vulnerability to fake news in social networks: A community health assessment model. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '19, page 432–435, New York, NY, USA. Association for Computing Machinery.

Taprial, V. and Kanwar, P. (2012). *Understanding Social Media*. Bookboon.

Thorson, E. (2008). Changing patterns of news consumption and participation. *Information, Communication & Society*, 11(4):473–489.

Tserpes, K., Papadakis, G., Kardara, M., Papaoikonomou, A., Aisopos, F., Sardis, E., and Varvarigou, T. (2012). Socios: A social media application ontology. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7567 LNCS:574–584.

Vishwakarma, A. and Chugh, M. (2023). Covid-19 vaccination perception and outcome: society sentiment analysis on twitter data in india. 13:84.

Wang, S., Yang, Z., and Chang, Y. (2021). Bringing order to episodes: Mining timeline in social media. *Neurocomputing*, 450:80–90.

Wycislik, L. and Warchal, L. (2014). *A Performance Comparison of Several Common Computation Tasks Used in Social Network Analysis Performed on Graph and Relational Databases*, volume 242. Springer Verlag.

Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., and Sharif, S. (2021). An analysis of covid-19 vaccine sentiments and opinions on twitter. *International Journal of Infectious Diseases*, 108:256–262.

Yum, S. (2020). Social network analysis for coronavirus (covid-19) in the united states. *Social Science Quarterly*, 101(4):1642–1647.

Zhang, F., Zhang, Y., Qin, L., Zhang, W., and Lin, X. (2017). When engagement meets similarity: Efficient (k,r)-core computation on social networks. *Proc. VLDB Endow.*, 10(10):998–1009.