

Exploiting Social Capital for Recommendation in Social Networks*

Paulo Roberto de Souza
Federal University of Bahia
Institute of Computing

Frederico Araujo Durao
Federal University of Bahia
Institute of Computing

João Paulo Dias de Almeida
Federal Institute of Sergipe

Abstract—Many computational techniques have been proposed by social networks to analyze the users' behaviors to recommend relevant content for them. Social networks generate a huge volume of information, which users cannot consume, generating a problem known as information overload. This way, filtering relevant information to help users with this problem becomes necessary. Social networks have many available features, such as relationships and interactions, which can be used to investigate the users' behaviors regarding news on their feed. The value of news can be defined as Social Capital, which is used by this work to model the user's preferences. This paper aims to investigate, model, and quantify interactions on social networks by exploiting social capital to develop a recommender system. Hence, in order to evaluate recommendations, an experiment was conducted with real users. Results show that our proposal was able to generate relevant recommendations on at least 62% of the scenarios.

I. INTRODUCTION

THE EXPONENTIAL growth of Web 2.0 has been driven by key innovations such as Online Social Networks (OSN), in which users have become the information drivers on the Web. Online Social Networks provide a virtual environment in which people can share information, experiences, opinions, interests and specially make connections [1]. For example, on the microblogging social network Twitter, more than 300 million users post 140 million of tweets every day, generating a huge amount of information which is shared and consumed by millions of others users.

Although online social networks are primarily used to communicate and relate with others, over the last few years OSNs have become important tools of mass communication, particularly as a way to disseminate news and influence others. The energy emanating from social interactions and available resources has been investigated in literature as Social Capital (SC). [2] formalizes Social Capital as an aggregated value of resources that are available on relations' networks. [3] claim that these social networks features can be analyzed to model user preferences. The underlying assumption of SC is that individuals benefit from various norms and values that a social network fosters and produces, such as trust, reciprocity, information, and cooperation [4], [5]. All these observations aim to solve one of the most challenging problems of online social networks: the Information Overload, which is the natural human incapacity to process the huge amount of information

produced in social networks [6]. Too much information can quickly cross users' cognitive limits in processing news and can make them feel overwhelmed and overloaded. [7] warn that 66% of users on Twitter felt overloaded when they received a lot of posts, and more than one-half reported needing a tool to filter irrelevant posts. After realizing this, Twitter developed a tool for marking comments as irrelevant or offensive, but not to analyze social capital as well. In this scenario, even users are able to mark posts as irrelevant or offensive, but there are no guarantees that similar posts will not come up in the future, or even that other posts that consider the social capital features will be recommended.

Recommender Systems (RS) therefore ascend as key tools to cope with the Information Overload problem by filtering relevant information according to the user's interest [8]. In literature, the problems about Information Overload (IO) have been discussed by many authors through analyses such as User's Influence [9], Interactions [10], and News Timeline [11], but none of them focus on the analysis of exploiting SC. [12] claims that RSs are more common among the generations of Web applications due to collaboration, interaction, and sharing of information. These systems use various piece of information to model each user's preferences such as clicks, website history, purchases, and items' evaluation. The term item, in general, describes what is used by RS to recommend to the user, like news or tweets on Twitter, or books on Amazon [13].

The power of SC in social networks is naturally affected by irrelevant information, which can overload users. This way, once the SC inherited from OSNs is neglected, RS perform an important role in content filtering on social networks. Given the complexity and resources offered by social networks, the value of the recommended item can consider other aspects, such as collaboration [14], interaction [15], and influence [16]. For instance, news can supposedly be more informative if there is good discussion on its comments. Different points of view can help the reader see that information from various perspectives. For example, a user can encounter solutions for a particular problem from others' opinions in social networks. In other words, the value of a recommendation item is associated with the social engagement behind it, beyond the traditional similarity metrics of interest.

This paper aims to investigate, model, and quantify interactions and available resources on social networks by exploiting SC, besides developing a RS through this exploitation. This

This work was partially funded by the Coordenacao de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Grant number 001

approach analyzes and values news that have more interactions or comments across a set of news, which are interesting for a user in order to suggest what is most relevant. We can point out that this paper does not consider misinformation or news published by bots yet.

The main contributions of this work are: i) *The Social Capital Recommendation Model* that calculates the value of news recommendations grounded from social interactions in a social network, ii) *The User Model* that comprises a user profile from user interactions and feedbacks from recommended news, and iii) *A Social Network Interaction Dataset* containing data from user activities in the social network.

The remaining aspects of this paper are structured as follows: Section II discusses related works. Section III introduces the proposed recommendation algorithm. Section IV presents the RecSocial application as well as the experiment setup and results. Section V discusses the results. Finally, Section VI concludes this paper and presents future work.

II. RELATED WORKS

Recommender systems on social networks have been attracting attention for a long time. These systems aim to aid users in decision-making, for example: which news to read [17], which account to follow [18], or which Wiki Pages to read [19]. In this context, many techniques are used to generate recommendations, such as machine learning, content-based and user-based filtering, besides collaborative filtering. Moreover, RS are widely used in literature, but both social capital and RS are rarely addressed working together. Nevertheless, we use similar studies to compare their techniques with our work such as: Text Analysis (TA), Topic Modeling (TM), Semantic Enrichment (SE), User Popularity (UP), Natural Language Processing (NLP), Sentiment Analysis (SA), Conceptual Relations (CR), Collaborative-Filtering (CF), Content-Based-Filtering (CB), Social Capital (SC), Empiric Analysis (EA), and Centrality Measure (CM).

[20] explore Twitter to analyze and recommend news through a CB algorithm, which has the goal to use different strategies (topic modeling, semantic enrichment, and temporal restrictions) to create a user's profile. Furthermore, they have developed a framework that models the user's preferences by adopting a semantic enrichment approach. So, this work depicts different techniques from our proposal, focusing more on semantic and topic modeling. Our proposal, on the other hand, focuses on the news' content, besides interactions and other features. Furthermore, [19] have developed a tag-based RS, which aims to recommend wiki pages used in corporate environments. In other words, pages are created for knowledge sharing about a particular subject within the organization. This way, they proposed a method called Wiki Page Collaborative Value (WPCV), which calculates the collaborative value of wiki pages. This approach analyzes the collaborative activities on a wiki page, such as edition, evaluation, tagging, or comments. Besides that, other features are adopted as a user experience about the subject and interaction among social ties. As a result, the score obtained through the WPC method is

used to rank and recommend wiki pages. [19] employ features similar to those used in this paper. [21] have developed an analysis on how the user's personality is manifested through different features on Facebook. They use many features such as the user's actions, including the number of published photos, events, and groups he has uploaded or created and the amount of news that he has liked. In addition, they use aspects of the profile that depends on the actions of a user and their friends, including the number of times a user has been tagged in photos, and the size and density of their friendship network. Thereby, [21] present a feature analysis similar to the one in our work. [22] added a new strategy assigning weights to topics/concepts of user's interest. This approach helps specify to what extent the user is interested in a topic. In this manner, they observed that, by considering the change of interest for a long time, the recommendations' quality was improved. Regardless of how they approached different techniques when compared to this work, we have adopted a weighting sentiment analysis to improve the recommendations.

In turn, [23] have developed a CB RS called Lumi Social News, using a mobile approach, which aims to recommend extracted news from the active user's timeline in their social networks based on their geographic location. The news are ranked according to their popularity and local trending. In other words, they are measured according to the frequency of sharing and interactions (likes and sharing, for example) among individuals of the same geographic location as the active user on the system. [17] approach a RS called TGS-post in order to recommend tweets based on conceptual relations among interest topics of a target user. The main goal of TGS-post is to present a new timeline for a user, which is ranked according to their interest. [24] proposed a framework, which has the goal to analyze tweets and identify which are the criteria that make them popular. The basic idea is to analyze the tweets in order to find the basis of the popularity of a person and extract the reasons supporting the popularity. As a result, they claim that this analysis can be used to recommend a list of the most popular users according to the personal interests of each individual. [10] have developed a RS, which uses the interactions maintained among users on social networks, therefore recommending users with similar preferences. These three studies, approach similar features in our work, using them to measure popularity, influence, and others to identify the user's preferences and behavior.

[18] have developed a RS called EIRank, which aims to recommend Twitter users to be followed. They depict the importance of interactions among users from which they can be measured and integrated into a proximity metric that factors the relationships among users and the number of interactions. Furthermore, they consider influent users those that have a large number of followers and with a high frequency of interaction to recommend them. Hence, we can observe how the news shown in a user's timeline can contribute to overloading him/her or even reduce his/her engagement. In such manner, this study approaches features which are analyzed so as to construct our proposed social capital model.

III. THE PROPOSAL

A. Twitter Features

This paper uses Twitter for evaluation purposes. However, the recommendation model may be implemented by other platforms (i.e. Instagram, Facebook), as it implements the basic concepts of social networks. Twitter is a micro-blogging service that allows users to share messages, called tweets, which contain various resources and offer a glossary¹ with a wide list: **Timeline**: A list of tweets published by accounts that a user follows; **News (Tweet)**: It is a message which contains a max of 280 chars. The term news will be adopted to refer the content created by the user on Twitter; **Retweet**: It is the action to forward a tweet; **Favorite**: It is the user action of liking a tweet; **Lists**: It is a functionality that allows creating personalized lists of users or topics. For example, a user can create lists with users or topics that are most interesting for him/her such as “soccer players”, “top 10 TV shows”; **Reply**: It is the users’ action to comment on a tweet; and **(Un)Follow**: The act of un/following another account.

1) *Twitter’s API*: Twitter provides an API to access information on its platform. This way, endpoints are available to extract the content of a timeline, post news, interact, or even follow any account. Moreover, the platform provides access plans that have limitations on extracting data in a range of time. As an example, we adopted a standard plan, which allows us to make 450 requests within a 15-minute window. So, if a high number of users makes requests in a short period of time, we will not be able to access that information.

B. Notation

In this section we present the mathematical notations that will be used along this proposal. **U**: The set of all users; **N**: The set of all extracted news; **TC**: The number of received likes; **TE**: The number of retweeted news; **TCP**: The number of comments that a news has; **TS**: The number of followers; **TLS**: The number of lists that a user belongs; **TNP**: The number of published news; **TXT**: The text of published piece of news; ps_n : The sentiment weight of a comment or news; **STM**: The reputation score of a mentioned user; and **STC**: The social capital score of a comment or a piece of news.

1) *Users (U)*: Social networks are constituted of inter-linked users that post, comment, mention or forward some news. Users are represented as $U = \{u_1, u_2, \dots, u_n | 1 \leq n \leq \mathbb{N}\}$, where each $u \in U$ is defined as a tuple $u = \langle TC, TS, TLS, TNP, RN \rangle$ with four metadata, which are used to calculate the influence of users in the social network (see Section III-E). RN is the assigned rating for news when a user participates in an online experiment. This was created to demonstrate that the user’s profile is updated constantly.

2) *News (N)*: News are the recommended items for a given user, and they are formally represented as $N = \{n_1, n_2, \dots, n_z | 1 \leq z \leq \mathbb{N}\}$, where each $n \in N$ is a tuple $n = \langle TC, TE, STM, STC, TCP, TXT \rangle$ with six metadata,

¹<https://help.twitter.com/pt/glossary>

which are also used to calculate the reputation and influence of users in the social network (see Section III-E).

C. News Pre-processing and Modeling

All extracted news are pre-processed and modeled according to the Fig. 1. As a result of the pre-processing, the news are represented as $NM_n = \{feat, (term, frequency) | term \in BW, feat \in n\}$, where each $feat_i$ represents a news metric (i.e. amount of likes, retweets, comments, followers, followees), $term_i$ is a news’ term (i.e. keywords from comments or posted news) and $frequency$ is a relevancy derived from TF-IDF calculation. Fig. 1 shows the news modeling scheme.

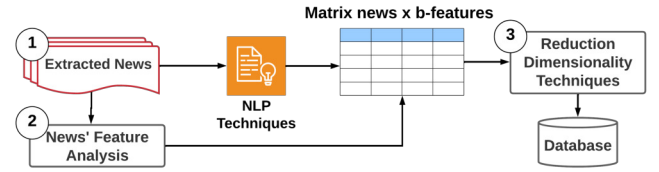


Fig. 1. Steps to model the news.

Extracted News: The news are extracted from the public timelines using the Twitter API. Besides the text itself, all interaction data (features) associated to the news is also withdrawn, including retweets, likes, comments, hyperlinks and hashtags; **News Feature Analysis**: Once the feature vector is created, the relevance of each term of the news is calculated using TF-IDF metric. Moreover, each extracted feature is added to the news model. and **Sparseness**: A multidimensional news matrix $news \times b\text{-features}$ is created containing all features of all extracted news. Principal Component Analysis technique is used to reduce the matrix dimension [25].

D. User’s Influence

This work approaches a new influence metric represented by the Equation 3, which is a part of the Equation 4 to measure how influential a user is, considering their popularity and reputation, besides his activity according to post frequency and feedback received by his followers. The Equation 3 calculates a user’s influence based on his popularity (Equation 1) [26],

$$PScore_u = 1 - e^{-\lambda \cdot TS} \quad (1)$$

where $PScore_u$ represents the popularity score of a user u , λ is a constant that by default is 1, which helps provide a fine adjustment, and TS is the number of followers.

In addition, the Equation 3 employs the Equation 2 to measure the user’s reputation considering the number of lists in which he/she is inserted [27]. The idea behind this metric is to identify the credibility, reputation, and reliability of a user compared to other users that shared similar interests. As a result, the Equation 2 shows that the lower the score, the better is a user’s reputation:

$$RScore_u = \begin{cases} \frac{TS_u}{TLS_u}, & \text{if } TLS_u \neq 0 \\ TS_u & \text{if } TLS_u = 0 \end{cases} \quad (2)$$

where u is the active user, TS is the number of followers, and TLS is the number of lists that a user u is inserted in. In other words, if a user is popular and has high posting frequency and feedback on their network, their score will be high. However, it is necessary to observe other users that do not have a good reputation, as this metric factors the number of followers and lists. Hence, new users or even those with few followers can produce interesting content and should be taken into account.

$$IScore_u = \frac{PScore_u + TC_u + TNP_u}{RScore_u} \quad (3)$$

where TC is the number of likes that u received for the news that they posted, and TNP is the number of news published by the user u .

E. The Social Capital Metric

Every news impact on people, the position where it appears in the timeline represents relevance to the user. In other words, each social network adopts different methods to rank news with the goal of satisfying its users or even helping them in the decision-making process (e.g. finding a healthy restaurant nearby). [28] and [29] claim that an interesting approach is to exhibit news with a high number of comments, interactions, connections and positive sentiment at the top of the news' list. The rationale is that much attention is being focused on that particular set of news.

We try to calculate the Social Capital as the power of a news ($n \in N$) based on the amount of interaction and its impact on the network over a period of time. This way, Equation 4 sums up these interactions pondered by the sentiment classification expressed by the text and influence of a user, calculated by the $IScore_u$. The sentiment analysis is used to provide new insights to understand the user's preference. Moreover, this classification can be used to model news' comments to predict their relevance [30]. The Social Capital metric is shown as:

$$SCScore(n, u) = \begin{cases} (TC_n + TE_n + STM_n + STC_n + TCP_n) \\ \cdot ps_n \cdot IScore_u, \text{ if } ps_n \neq 0 \\ (TC_n + TE_n + STM_n + STC_n + TCP_n) \\ \cdot IScore_u, \text{ if } ps_n = 0 \end{cases} \quad (4)$$

where n is a news, u is a user who posted it, TC , TE , STM , STC , TCP are described in Section III-B, and ps_n is the sentiment score, which is calculated from the Algorithm 2. Our intuition behind this metric is that news with more repercussions within a context of the user's interest can be useful to provide information, reduction of the information overload problem, and improve the user's engagement

F. Recommendation Model

Besides the social capital, the recommendation model contemplates the user's reputation and the sentiment on the recommended news. Indeed, our intuition behind this approach is that news with more comments, interactions, etc. can be

relevant according to the user's preferences. In this case, the following algorithms provide the recommendation model according to the aforementioned equations.

1) *Reputation's Algorithm*: The Algorithm 1 calculates the user's influence and has as an input a user ($u \in U$), who posts, comments, or was mentioned in news. It has as an output the user's influence score. Line 1 initializes variables according to each user's features. Line 2 verifies if a user has followers, if true, the max number of followers is assigned to TS . Line 5 verifies if the user has been inserted in any lists, if true, the constant β is assigned to TLS . The value used in this work was 1, once it is not less or equal to 0. Line 8 calculates the user's reputation. Line 9 verifies if a user is authentic, which is used to establish if the account is authentic, active, and remarkable. So, in case an account has those characteristics the score is increased, which is carried out by the constant θ in line 10, and the value used was 1. Finally, line 12 calculates the score.

Algorithm 1 User's influence pseudo-code.

Require: User $u \in U$
Ensure: $IScore_u$
 $TS, TC, TLS, RNP \in u$
2: **if** $TS = 0$ **then**
 $TS \leftarrow maxFollowers$
4: **end if**
 if $TLS = 0$ **then**
6: $TLS \leftarrow \beta$
 end if
8: $reputationScore \leftarrow RScore_u$
 if user is verified _{u} **then**
10: $reputationScore \leftarrow reputationScore + \theta$
 end if
12: **return** $\frac{PScore_u + TC + TNP}{reputationScore}$

2) *Sentiment Analysis Algorithm*: The Algorithm 2 shows how the text of the news is analyzed and processed in order to classify the user's sentiment while typing it.

Algorithm 2 Sentiment analysis of a text/comment.

Require: Text of a news/comment $text_n \mid n \in N$
Require: Classification of the news sentiment $label$
text \leftarrow preprocess($text_n$)
2: **if** text \neq null and text.size $>$ 0 **then**
 return nlp.classify(preProcessedText)
4: **end if**
return null

Thus, the Algorithm 2 has the news text as input and the classification of this text as an output. Line 1 processes the text removing irrelevant data. Line 2 verifies if a text is null and if its length is greater than zero. If true, line 3 uses the Amazon Comprehend service to classify the sentiment expressed by the text, which can be negative, positive, neutral, or mixed. This service has a set of available tools to analyze texts in natural language. We can point out that this service uses a

pre-trained model to gather insights about a document or a set of documents. Besides that, according to Amazon, this model is continuously trained on a large body of text so that there is no need to provide training data. Finally, in the same line mentioned above, the label is returned.

Algorithm 3 Pseudo-code of a recommendation model.

Require: News $n \in N$, User $u \in U$
Require: $SCScore(n, u)$

```

URS ← influenceScore(u)
2: TC, TE, TCP ∈ n
STM, STC, score ← 0
4: for user ∈ Mn do
    STM ← STM + influenceScore(user)
6: end for
for com ∈ Cn do
8:   STC ← STC + scScore(com, com.whoPosted)
end for
10: score ← score + ((TC+TE+STM+TCP+STC) · URS)
    sentimentWeight ← 0
12: if shouldAnalyzeSentiment then
    sentimentLabel ← sentimentAnalysis(textn)
14:   if sentimentLabel is positive then
    sentimentWeight ← α
16:   else
    if sentimentLabel is negative then
18:     sentimentWeight ← θ
    else
20:     sentimentWeight ← β
    end if
22:   end if
    score ← score · sentimentWeight
24: end if
return score + Similarity(u, n)
  
```

3) *Recommendation Model With Social Capital:* The Algorithm 3 calculates the social capital of news, which receives as an input a piece of news n and a user u , and generates the social capital score as an output. The Algorithm 3 employs techniques from both Algorithms 1 and 2. Lines 2 initializes variables according to each news features. Line 4 iterates each mentioned user of the news, so the algorithm calculates the user's influence. Line 7 iterates each comment, and line 8 calculates its social capital. Line 10 calculates the partial score, which is weighted by the user's influence. Line 12 verifies if the sentiment analysis should be considered, if so, the analysis is undertaken (line 13) as shown in Algorithm 2. Lines 13-22 verify each condition to assign the weight to the final score, in which α , θ , and β are the weights used on the social capital score. In this work, the values 1.5, 1, and 0.5 were used as alpha, omega, and beta, respectively. These values were adjusted empirically. Line 25 calculates the cosine similarity [31] between the user's profile and the current news. We consider a high similarity, scores greater than 0.7. In other words, we use the user's ratings from each recommended news in order to update his profile and then, recommend a set of

news that is closest to his previous preferences. In this case, we increase the score of news with the user's profile similarity score. Finally, a set of news that will be presented to an active user is ranked by the social capital score in order to show news with the highest score at the top of the list.

IV. EXPERIMENTAL EVALUATION

The proposed approach was evaluated in a user trial with the goal of assessing the quality of recommendations. To conduct the experiment, we developed the RecSocial web-based system, an application used for collecting the users' feedbacks and evaluating the generated recommendations using the approach proposed by this work.

A. The RecSocial Application

The RecSocial was created for evaluation purposes exclusively. All data collected from the user trial resulted in dataset², which is available freely online. For this, some steps are realized: 1) **News Extraction:** RecSocial is populated with news extracted from Twitter; 2) **Pre-processing:** Data is pre-processed in order to remove redundancies, inconsistencies, noises and irrelevant data. After that, we calculate the statistics about the data on the RecSocial including the number of followers, likes and comments in a news, etc; 3) **Social Capital Model:** The Social Capital score for each piece of news imported from Twitter is calculated; and 4) **Recommendation Model:** Finally, the recommendations are generated so that users are able to evaluate them.

1) *Using RecSocial:* Initially, the user creates their account and logs in. After authentication, a set of topics and subtopics of interest (Sports, Trips, News, etc.) are shown so that user selects those which will form their timeline. Bear in mind that we are trying to simulate Twitter's timeline. After authentication, the system shows a set of topics, which have a set of subtopics associated with a set of extracted news.

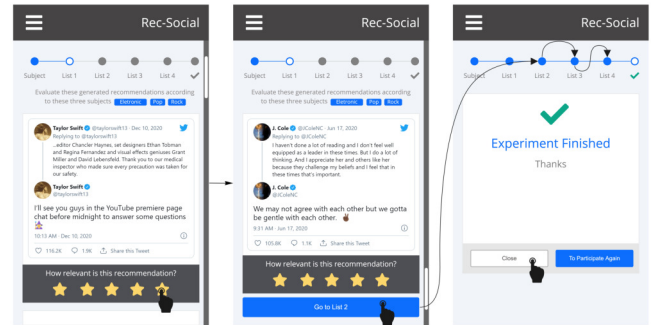


Fig. 2. Recommendations and user's feedback.

After selecting the topics of interest, recommendations are generated as shown in Fig. 2. In this step, the user evaluates each generated recommendation providing his/her feedback by a 5-point Likert scale. A "Go To List x" button appears at the bottom guiding the user to the next round of recommendations.

²shorturl.at/qBJQ3

For these experiments in particular, four rounds were undertaken. Finally, the logged user can close the app or participate again for a new round of recommendations. This app code is available on <https://github.com/pauloprdesouza/recsocial-api>.

B. Methodology

For the online experiment, 80 volunteers took part in the experiment from 3/15/21 to 6/6/21. They were 20 to 40 years old and all very familiar with social networks, using them for various purposes including work, family and friendship. Before the experiment, the participants were introduced about the RecSocial application and received the two major instructions: 1) Select three subcategories from their interests; and 2) Evaluate four lists of recommendations, each with 10 items (tweets about news). In summary, 40 recommendations were evaluated by each participant using the 5-point Likert scale [32] (see Fig. 2). In total, there were 80 volunteers, 3.030 recommendations evaluated, 12.185 Twitter accounts used, 9.533 (78%) users mentioned and 32.252 news extracted.

C. Simulation of User's Timeline

In general, the participants do not feel comfortable in giving access to their Twitter personal accounts, once their behaviors on the microblogging can show a lot about them. In other words, they do not want to expose themselves in order to allow that an experiment can analyze their interactions and produced content over time. For this reason, we decided to simulate the user's timeline (on RecSocial) by allowing them to make up a timeline with Tweets from a wide range of categories such as music and news, for example. These categories represent the user interests and were used to retrieve Twitter accounts associated with them. For instance, if one chooses "Sports/Football", then @FIFACOM is an eligible account from which its tweets will compose that user's timeline. In this case, our intuition behind this approach is to construct a user's timeline more as diversified as possible. In other words, the number of news was defined considering the diversity criteria for the RS.

Our approach was compared against other methods: i) Social Capital (SC); ii) Social Capital with Sentiment Analysis (SC+SA); iii) Cosine Similarity (CS-Plus); and iv) Baseline (B1) [33]. Inspired by [34], who uses an approach of incremental evaluation, called Prequential Evaluation, this experiment methodology updates the user's profile according to his preferences after every recommendation round. The following metrics were used: Mean Reciprocal Rank (MRR) [22], Precision@N [22], Mean Average Precision (MAP) [17], and Normalized Discounted Cumulative Gain (NDCG) [35].

D. Results

The experiment results are organized by the metrics analyzed. The MRR results are SC (0.75), SC+SA (0.68), CS+PLUS (0.67) and B1 (0.68). The SC algorithm achieves the highest MRR scores. The SC MRR score is 7% higher than both B1 and SC+SA MRR scores. The lowest MRR results are observed by the CS-PLUS algorithm. From these results, we can observe that SC algorithm was effective to generate

relevant recommendations at the top of the list. In other words, users evaluated recommendations between ratings 4-5.

Fig. 3 shows the precision results. We can observe that the SC algorithm achieves higher scores than the other approaches. In other words, the proposed approach was able to generate relevant recommendations up to the fifth evaluated time. The compared methods present slight variations among them. We can point out that SC algorithm generates relevant recommendations in at least 60% of cases, demonstrating that SC approach can be useful to generate recommendations.

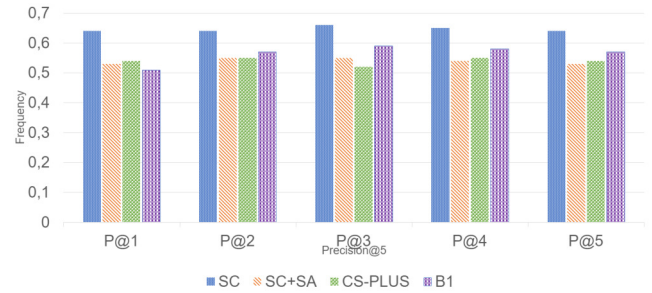


Fig. 3. Results of Precision@N metric.

The MAP results are SC (0.62), SC+SA (0.53), CS+PLUS (0.52) and B1 (0.55). The SC algorithm achieves the highest MAP scores. The SC MAP score is 7% and 9% higher than B1 and SC+SA MAP scores respectively. The lowest MAP results are observed by the CS-PLUS algorithm. In this case, we can observe that SC algorithm was able to generate the most interesting recommendations at the top of the list, in other words, MAP metric rewards first-loaded relevant recommendations.

The NDCG results are SC (0.85), SC+SA (0.81), CS+PLUS (0.79) and B1 (0.79). The SC algorithm achieves the highest NDCG scores. The SC NDCG score is 4% and 6% higher than SC+SA and B1 scores respectively. The CS-PLUS presents a score equal to B1. In this case, once the NDCG metric considers the order of recommended items versus the ideal order of recommended items, we observe that the SC algorithm was able to generate relevant recommendations in 85% of cases at least. We notice that the SC+SA algorithm has the second-best results when compared to the others, generating relevant recommendations in 81% of cases.

Fig. 4 shows the rate frequency by each comparison method. We can observe that the algorithm SC is the most frequently top rated (5 stars), and the least frequent for lower ratings (1-3). The top rating at SC is 3% higher than the B1 method, the second best observed. The other methods do not present relevant results when compared with SC.

Fig. 5 shows the correlation between evaluated recommendations and the news features. We calculate the average of all features of each news recommended and correlate it with ratings. We can observe that the SC algorithm achieves the highest scores. The SC score is 24% higher than B1, the second best observed. The other methods do not present relevant results when compared to SC. Therefore, we could

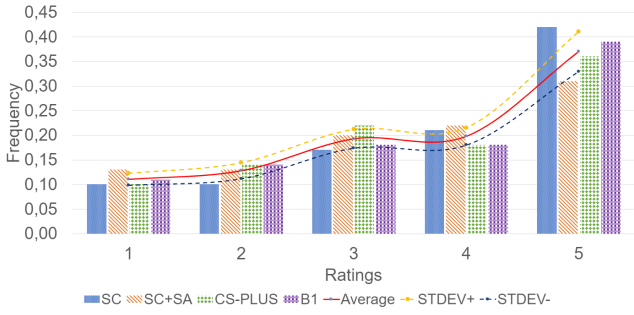


Fig. 4. Evaluation's rating frequency.

state that we have the intuition that users may prefer reading news with more interactions and with more features.

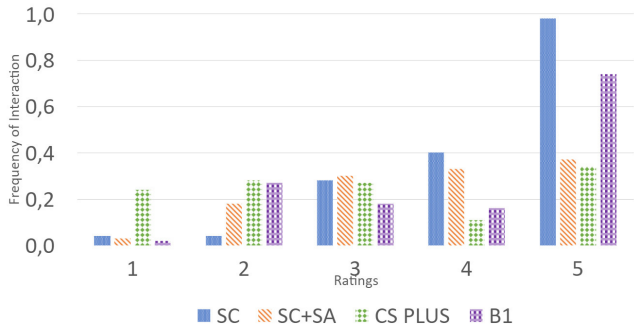


Fig. 5. Evaluated recommendations X news features.

Fig. 6 presents the correlation between recommendations evaluated with the IScore (see Section 3) metric. We can observe that SC is the most frequently top rated (4-5), and the least frequent for lower rating 3. Besides, the SC+AS achieves the SC rates (4 stars). In other words, recommendations published by influential users received better evaluations according to the proposed algorithms. These results can be expressed from the fact that the user's influence in the social capital approach can generate relevant recommendations according to the user's interests.

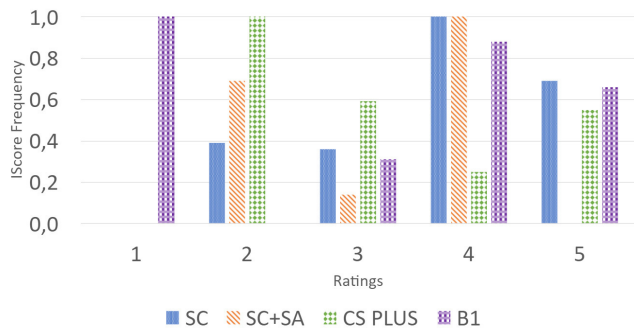


Fig. 6. Evaluated recommendations X news features.

V. DISCUSSION

Personalization methods aim to provide the construction of the user's profile according to their interests. In this case, [11] discuss how the timeline order influences what a user sees, and they take many interactions and features of news into account to analyze it. Our work has found empiric results that the proposed social capital algorithm, besides the user influence metric can better order the user's timeline by providing relevant news. An example of this can be seen through Fig. 6, which demonstrates that users liked (ratings 4-5) news recommended through popular accounts. Besides that, both proposed metrics IScore and SCSScore when used together have the capacity to measure the value of news, which provides a score that can be used to rank a user's timeline, as we can see from the previously aforementioned ranking metrics. In other words, taking interactions, relationships and other news features into account can be useful to generate relevant recommendations.

We can see that results from the other algorithms were more similar, indicating that the incremental approach adopted by this work updates the user's profile according to the last evaluation. Moreover, we observed that our second SC+SA algorithm does not present relevant results when compared with other implementations. This occurs because weights employed in sentiment analysis need other adjustments through empiric tests to perform online and offline experiments. In addition, texts with few words may be biased, which influences the weighting step.

Finally, this work does not deal with misinformation problems or even bots. Moreover, we know that these problems mentioned before may have a high social capital score if they present a high number of interactions, comments, etc. We must also point out yet that users can be overwhelmed or even overloaded with news that talks about the same context, but we need to consider the scope of this work, which is exploiting the social capital on social networks as a method to generate news recommendations.

VI. CONCLUSION

This work proposes a social capital metric to measure the news value used in generating personalized recommendations, according to the individual's profile. Thus, the results show that exploiting social capital on social networks can be useful to cope with the information overload problem. The hypothesis raised by this proposal is to have a timeline that contains the most interesting and relevant news at the top of the list, presenting a significant gain on the knowledge and on information overload reduction.

As future work, we intend to test our proposal with other approaches in order to measure how misinformation impacts the news with high social capital. Besides that, we plan to implement other techniques to cope with the synonymy and polysemy problem using WordNet, as well as to construct a bigger dataset in order to improve the user's profile. After that, we plan to conduct a long-term experiment using linear regression techniques to learn about user's behavior and further

improve their profile. Finally, we plan to resort to certain strategies in order to deal with misinformation and bots as previously mentioned by using available datasets in literature, and applying other available metrics in literature so as to compare the results with our proposed method.

REFERENCES

- [1] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in social media research: Past, present and future," *Information Systems Frontiers*, vol. 20, no. 3, pp. 531–558, Nov. 2017. doi: 10.1007/s10796-017-9810-y
- [2] P. Bourdieu, "The forms of capital," in *Readings in Economic Sociology*. Blackwell Publishers Ltd, 2002, pp. 280–291.
- [3] J. Nahapiet and S. Ghoshal, "Social capital, intellectual capital, and the organizational advantage," *Academy of Management Review*, vol. 23, no. 2, pp. 242–266, Apr. 1998. doi: 10.5465/amr.1998.533225
- [4] C. Spratt, J. Hong, K. McAreavey, and W. Liu, "Community-based measures for social capital," in *Studies in Computational Intelligence*. Springer International Publishing, Dec. 2018, pp. 327–338.
- [5] J. Coleman, *Foundations of social theory*. Cambridge, Mass: Belknap Press of Harvard University Press, 1990.
- [6] G. Pijpers, *Information Overload: A System for Better Managing Everyday Data*, ser. Microsoft executive leadership series. John Wiley and Sons Ltd, 2010.
- [7] K. Bontcheva, G. Gorrell, and B. Wessels, "Social media and information overload: Survey results," 06 2013.
- [8] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems: Introduction and Challenges*. Springer US, 2015, ch. 1, pp. 1–34.
- [9] S. Forouzandeh, A. Sheikhamadi, A. R. Aghdam, and S. Xu, "New centrality measure for nodes based on user social status and behavior on facebook," *International Journal of Web Information Systems*, vol. 14, no. 2, pp. 158–176, Jun. 2018. doi: 10.1108/ijwis-07-2017-0053
- [10] Y. Li, J. Liu, and J. Ren, "Social recommendation model based on user interaction in complex social networks," *PLOS ONE*, vol. 14, no. 7, p. e0218957, Jul. 2019. doi: 10.1371/journal.pone.0218957
- [11] N. Bartley, A. Abeliuk, E. Ferrara, and K. Lerman, "Auditing algorithmic bias on twitter," in *13th ACM Web Science Conference 2021*. ACM, Jun. 2021. [Online]. Available: <https://doi.org/10.1145/3447535.3462491>
- [12] T. O'Reilly, "What is web 2.0," Sep 2005. [Online]. Available: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- [13] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer, 2011.
- [14] S. Berkovsky, J. Freyne, and G. Smith, "Personalized network updates: Increasing social interactions and contributions in social networks," in *User Modeling, Adaptation, and Personalization*. Springer Berlin Heidelberg, 2012, pp. 1–13.
- [15] W. Feng and J. Wang, "Retweet or not?" in *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*. ACM Press, 2013. doi: 10.1145/2433396.2433470
- [16] C. Li and F. Xiong, "Social recommendation with multiple influence from direct user interactions," *IEEE Access*, vol. 5, pp. 16 288–16 296, 2017. doi: 10.1109/ACCESS.2017.2739752
- [17] D. P. Karidi, Y. Stavarakas, and Y. Vassiliou, "Tweet and followee personalized recommendations based on knowledge graphs," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 6, pp. 2035–2049, Apr. 2017.
- [18] H. Bo, R. McConville, J. Hong, and W. Liu, "Social network influence ranking via embedding network interactions for user recommendation," in *Companion Proceedings of the Web Conference 2020*. ACM, Apr. 2020. doi: 10.1145/3366424.3383299
- [19] F. Durao and P. Dolog, "Improving tag-based recommendation with the collaborative value of wiki pages for knowledge sharing," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 1, pp. 21–38, Apr. 2012. doi: 10.1007/s12652-012-0119-x
- [20] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing user modeling on twitter for personalized news recommendations," in *User Modeling, Adaption and Personalization*. Springer Berlin Heidelberg, 2011, pp. 1–12.
- [21] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of facebook usage," in *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12*. ACM Press, 2012. doi: 10.1145/2380718.2380722
- [22] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Twitter-based user modeling for news recommendations," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, ser. IJCAI '13. AAAI Press, 2013. doi: 10.5555/2540128.2540558 p. 2962–2966.
- [23] G. Kazai, I. Yusof, and D. Clarke, "Personalised news and blog recommendations based on user location, facebook and twitter user profiling," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, Jul. 2016. doi: 10.1145/2911451.2911464
- [24] R. K. Mudgal, R. Niyogi, A. Milani, and V. Franzoni, "Analysis of tweets to find the basis of popularity based on events semantic similarity," *International Journal of Web Information Systems*, vol. 14, no. 4, pp. 438–452, Nov. 2018. doi: 10.1108/ijwis-11-2017-0080
- [25] M. E. Timmerman, "Principal component analysis," *Journal of the American Statistical Association*, vol. 98, no. 464, pp. 1082–1083, dec 2003. doi: 10.1198/jasa.2003.s308
- [26] A. Aleahmad, P. Karisani, M. Rahgozar, and F. Oroumchian, "OLFinder: Finding opinion leaders in online social networks," *Journal of Information Science*, vol. 42, no. 5, pp. 659–674, Jul. 2016. doi: 10.1177/0165551515605217
- [27] J. Bullas, "What is your twitter reputation?" 2011. [Online]. Available: <https://www.jeffbullas.com/what-is-your-twitter-reputation/>
- [28] T. H. P. Silva, A. H. F. Laender, and P. O. S. V. de Melo, "On knowledge-transfer characterization in dynamic attributed networks," *Social Network Analysis and Mining*, vol. 10, no. 1, Jun. 2020.
- [29] L. Camacho, J. Faria, S. Alves-Souza, and L. Filgueiras, "Social tracks: Recommender system for multiple individuals using social influence," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. SCITEPRESS - Science and Technology Publications, 2019. doi: 10.5220/0008166503630371
- [30] Y. Bae and H. Lee, "Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 12, pp. 2521–2535, Nov. 2012. doi: 10.1002/asi.22768
- [31] F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261–273, nov 2015. doi: <https://doi.org/10.1016/j.eij.2015.06.005>
- [32] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*. Cambridge. Cambridge University Press, 2010.
- [33] E. Lee, Y. J. Kim, and J. Ahn, "How do people use facebook features to manage social capital?" *Computers in Human Behavior*, vol. 36, pp. 440–445, Jul. 2014. doi: 10.1016/j.chb.2014.04.007
- [34] J. Vinagre, A. M. Jorge, and J. Gama, "Evaluation of recommender systems in streaming environments," 2014. doi: 10.13140/2.1.4381.5367
- [35] R. Zhang, H. Bao, H. Sun, Y. Wang, and X. Liu, "Recommender systems based on ranking performance optimization," *Frontiers of Computer Science*, vol. 10, no. 2, pp. 270–280, Jul. 2015. doi: 10.1007/s11704-015-4584-1