DIGITALES ARCHIV

Sandoval, Luis G. Moreno; Rojas, Liliana M. Pantoja; Pomares-Quimbaya, Alexandra et al.

Article

Computational linguistic and SNA to classify and prevent systemic risk in the Colombian banking industry

International journal of e-business research

This Version is available at:
http://hdl.handle.net/11159/654583

Computational Linguistic and SNA to Classify and Prevent Systemic Risk in the Colombian Banking Industry

Luis G. Moreno Sandoval, Pontificia Universidad Javeriana, Colombia*

https://orcid.org/0000-0001-8853-2455

Liliana M. Pantoja Rojas, Universidad Distrital Francisco José de Caldas, Colombia

Alexandra Pomares-Quimbaya, Pontifica Universidad Javeriana, Colombia

Luis Antonio Orozco, Universidad Externado de Colombia, Colombia

ABSTRACT

The banking sector has been one of the first to identify the importance of social media analysis to understand customers' needs to offer new services, segment the market, build customer loyalty, or understand their requests. Users of Social Networking Sites (SNS) have interactions that can be analyzed to understand the relationships between people and organizations in terms of structural positions and sentiment analysis according to their expectations, opinions, evaluations, or judgments, what can be called collective subjectivity. To understand this dynamic, this study performs a social network analysis combined with computational linguistics through opinion mining to detect communities, understand structural relationships, and manage a Colombian case study's reputation and systemic risk in the banking industry. Finagro and BancoAgrario are the network leaders in both centralities, most of the main actors have a negative polarity, and MinHacienda and cutcolombia with totally different orientations appear in all methods.

KEYWORDS

Banking Sector, Community Detection, Content Analysis, Natural Language Processing, Opinion Mining, Social Network Analysis, Topic Modeling, Twitter

INTRODUCTION

Social networks are widely recognized as a new communication technology that has transformed social contexts (Bohlin et al., 2018). For example, banks have exploited this technology using mainly Facebook, Twitter, and YouTube for marketing, financial advice, information support, customer service, sales, surveys, polls, and other services (Mabic et al., 2017). Thus, services offered through

*Corresponding Author

mobile devices and social networks are disruptive because they can replace other banking channels and make branch-oriented banking redundant (Boyd & Ellison, 2007; Miranda et al., 2013).

Users of social networks and e-commerce websites express their opinions or comments on banks' products and services, which is used to analyze service quality perceptions, which improves responses to maintain loyalty and trust (Nicoletti, 2017). Advances in the development of algorithms allow measuring interactivity and content posted about a given bank, taking advantage of the potential of social networks to identify stakeholders and learn about their needs, expectations, interests, preferences, and opinions; however, the integration of services with social networks is not an easy task.

Technological advances have increased its subsidiaries' management and control capacity (Berger & DeYoung, 2005). Since 1992, when some USA banks started using electronic banking services, the Internet and social networks became the industry's backbone (Mucan & Özelturkay, 2014). These developments have enabled the industry to promote an organizational identity, develop public relations with stakeholders, solicit feedback, share content, or analyze conversations to handle complaints and grievances to create new solutions (Keskar & Pandey, 2018; Ozdora & Atakan, 2016; Schulte, 2018).

According to Bohlin et al. (2018), banks that are pioneers in innovation use social networks, thus revealing their best practices; therefore, social networks become one of the main channels for managing reputational risk (Porras & Orozco, 2019). For example, if there is a concentration of liquidity among a small group of banks and this spreads in social networks, the collapse may be greater, and the whole network may collapse (Leitner, 2005) so that the defaults of one bank could influence the failure of another in a domino effect (Ozdora & Atakan, 2016). Uhde & Michalak (2010) securitization of credit risk has a positive impact on increasing the systematic risk of European and Swiss banks as they seek financial leverage, achieving standards such as the publication of their securitizations to control expectations of external investors and bank managers.

Information about the banking sector structure using the interbank network topology is a valuable tool for directing policies, managing contagion risk, and preventing financial system failures (Boss et al., 2004; Houston et al., 2018; Leitner, 2005). For instance, Elsinger et al. (2006) showed that contagion scenarios among Austrian banks would be a rare event; however, such scenarios can occur when there are many fundamental defaults due to the exposure of correlated portfolios affecting most of the banking sector. In such a case, 12% of the banking system's total securities would have to be added to recover from a fundamental default and 1% for a systematic risk of contagion.

The analysis of social networks such as Twitter uses Bayesian risk models that can predict a bank's failures and default conditional on the information disseminated by its network (Cerchiello et al., 2017); models to identify and classify stakeholders and define the structure of the banking sector using social network analysis with opinions and sentiments in networks are scarce in the literature. Keskar et al. (2020) determined the main features of a network model for creating an index to assist banks in achieving customer satisfaction on the Internet.

The emergence of digital social networks has transformed society, social groups, and institutions in terms of communication and their opinions. Determining how linguistic variations enable the detection of communities and the relevance of specific vocabulary in social networks could lead to a better understanding of their dynamics and social foundations, resulting in better service (Puertas et al., 2021).

Therefore, this research develops a methodology to integrate Computational Linguistic (CL) analysis and structural techniques of Social Network Analysis (SNA) (Moreno S & Pantoja R, 2019) in the Colombian banking industry using Twitter accounts. This model shows the structural metrics of relevant actors, their polarity, and community detection and classifies them to understand customer relationships according to their way of thinking, feeling, and acting (collective subjectivity).

The article is structured as follows: the first section presents the literature review on the main topics for detecting communities in social networks, the structural analysis of the networks that perform the banking sector, and the identification of the sentiment in tweets regarding the banking industry's services. The second section shows the methodology and analysis model describing the

corpus captured and the detail of integrating the SNA and NLP techniques. The third part presents the development section, which analyzes the most outstanding results before the Colombian banking industry, the conclusions describing the theoretical and practical implications, and finally, a section with the limitations of the research and future work under an SNA and NLP integration approach.

BACKGROUND

Research on social networks and the banking sector is an emerging field studied in the last decade; however, there is very little literature concerning social networks and their support in better service with community sense. The following subsections outline the study's theoretical background, highlighting the literature on the applications of fundamental concepts of Natural Language Processing (NLP). First, the essential background of community detection is outlined; second, related work on SNA in banks is reviewed; finally, work addressing opinion mining in the banking industry to measure reputational risk is reviewed.

Community Detection

A complex system's collective behavior is better understood by detecting communities and classifying them according to similarities and differences (Loe & Jensen, 2015). As Lancichinetti and Fortunato (2009) mentioned, the community structure is one of the network's essential characteristics because it shows the nodes' internal organization.

The complex networks are composed of many nodes with difficulties in identifying communities that can summarize and explain the entire system's behavior. The partition of a network does not have a single solution since the new nodes are according to different criteria; therefore, detecting communities means discovering a small number of interactions involving many unitary elements (Capocci et al., 2004) and some similarities.

It has recently been observed that different networks in various fields often share important concepts or common themes, showing a view of their structures, properties, and emerging behaviors (Barabási & Bonabeau, 2003). However, there is no universally accepted formal definition because the community's concept depends on the problem domain and the observer who discovers it.

Generally, studies in SNA privilege the actors' structural position over the content of the relationships (Barabási & Bonabeau, 2003). Furthermore, the comparison of the graph structure, as a generalized task in machine learning analysis, can highlight the various applications in neuroscience, cybersecurity, social network analysis, and bioinformatics (Wills & Meyer, 2019), analyzing the representative network of the communities. In summary, network analysis can discover significant patterns and potentially illuminate the essential properties of networks, showing a robust modular nature or structure of the (Parthasarathy, 2011) community.

Social Networks and Communities in the Banking Sector

The banking industry's social network structure has been studied in recent years due to systemic risk and financial crisis implications, especially in contagion phenomena (Leventides et al., 2019). The literature often analyzed lender-borrower relationships showing a core-periphery structure in which few nodes are strongly connected, while a subset of several nodes presents links to the core with few reciprocal connections among them (Bargigli et al., 2015; Sui et al., 2020). The definition of a core-periphery network is as follows:

Core banks are large in balance sheet size and are all bilaterally linked with each other; periphery banks are relatively small and link only with the core banks; core banks intermediate liquidity among themselves and for their directly linked periphery banks. (Sui et al., 2020, p. 1)

The core-periphery structure is found in several cases. Houston et al. (2018) examined 99 large global banks in the Boardex database between 2000–2010. They found a highly centralized structure with two hubs of banks (the U.S. and European banking corporations), and small banks formed small groups in the periphery with few connections among them. Boss et al. (2004) found that interbank networks in Austria's financial system show a tiered community structure with a low degree of separation between banks, a low clustering in which two small banks linked to a significant financial corporation are not connected between themselves.

The case of the interbank market in Italy, analyzed in a multilayer network approach, shows a core-periphery and disassortative structure in which few banks hold the power to control the financial system (Bargigli et al., 2015). In Latin America, Mexico's case shows that inter-banking networks based on payment relationships also fit a core-periphery model (Martinez-Jaramillo et al., 2014). Meanwhile, the case of Brazil (Cont et al., 2012) shows substantial similarity to the inter-bank network structure presented by Boss et al. (2004) for the Austrian financial system.

Social and economic networks understand how a shock or default in one bank might spread an impact to other banks, i.e., a banking sector could present a cascade of failures in the financial system, and those failures will depend on network structure. Thus, studies of contagion phenomena in banking industry areas state that "beyond a certain point, dense interconnections serve as a mechanism for the propagation of shocks, leading to a more fragile financial system" (Acemoglu et al., 2015, p. 56).

Gong et al. (2019) indicate some empirical research works to find that Chinese financial entities' causal network possesses small world and scale-free properties. The number of connections increases dramatically in crisis periods, indicating more substantial interconnectedness in the financial system. On the other hand, when a bank defaults, it does not repay its obligations in full, which may precipitate other collapses; Rogers & Veraart (2012) notify a scenario where many banks close, letting banks fail in succession until only solvent banks remain, suggesting a state intervention. Thus, they establish the conditions for failing banks that consortia of other banks might rescue.

Opinion Mining in the Banking Sector

The SNS study as a channel to measure social networks around the banking industry has been neglected in the literature. Therefore, Opinion Mining (O.M.) in sentiment analysis is a new field of application and research that arose to give tools (Montoyo et al., 2012). The sentiment analysis aims to determine the attitude of a speaker or a writer concerning some topic or the global polarity of a document (Moreno S & Pantoja R, 2019). According to Albornoz et al. (2011), the polarity classification aims to obtain a score that indicates whether the text expresses a positive or negative opinion within a range where 0 would mean a neutral subjective load, 1 is a positive personal load, and -1 is a load personal negative.

A proposal like the Facebook Assessment Index developed for measuring the content, interactivity, and popularity of banks profiles (Miranda et al., 2013) goes in the direction of knowing better the behavior of banks in SNS. Afolabi et al. (2017) studied the activities of Nigerian banks from Facebook and Twitter sites. They categorized the banks' activities into three levels: very active, averagely active, and least active; they defined the score of sentiment analyses along the five largest and leading financial chosen banks and discovered the patterns on the Twitter and Facebook accounts for each bank are similar.

Houston et al. (2018) observed the influence of shared social connections among the largest 99 global banks in the Boardex database. They demonstrated that connected banks are more likely to partner in the syndicated loan market, and banks that play a more central role in the social network are more likely to play a leading role in syndicated loan origination.

Nopp & Hanbury (2015) developed a survey that obtained quantitative data where they analyzed through a lexicon the capability to measure a bank's attitude and opinions showing each entity's potential risk level. Then, they showed that sentiment scores from letter CEO from annual bank reports reflect major economic events between 2002 and 2014 very well. There is a strong correlation between uncertainty, negativity, and the Tier 1 Capital Ratio evolution over time.

The banking industry is not the only sector that has harnessed the potential of sentiment analysis to its advantage, as several political interest groups produced policy imperatives in the face of the recurrent use of this type of analysis. For example, Karavitis and Kazakis (2022) studied the borrowing costs of U.S. firms with foreign subsidiaries whose political exposure was evident and concluded that an increase in the positive political sentiment of the firm produces a drop in borrowing costs. In addition, the negative effect of a high lending cost in countries with foreign subsidiaries where political polarization is higher is mitigated. Similarly, Khatua et al. (2020) used a multinomial logistic regression model in the case of the 2014 Indian elections, showing that tweet patterns where various political parties are mentioned successfully reveal users' political opinions.

Election campaigns are a precise scenario where sentiment analysis is necessary to predict outcomes. Bermingham and Smeaton (2011) used a model to predict the Irish general election case by comparing the election polls and their final election results; in the end, they demonstrated the success of the predictive factor that cross-party sentiment analysis has due to the closed nature of the system. Liu and Lei (2018) used machine learning-based techniques to capture the sentiment and central themes of the 2016 U.S. presidential candidates' speeches, which explained the victory of President Trump by spreading negative sentiment posts on the networks and repeating those tweets to increase his electoral base and mobilize these network users to the boxes.

Private interest group interventions are another side of the economy that benefits from sentiment analysis, where the stock market becomes a perfect setting for very high returns if such groups know about the impacts or relationships of certain actors in the stock market. Tiwari et al. (2022) demonstrate the possible benefits in the Australian market by understanding, for example, the causalities of the sentiments of people in the 45-year age group and stock returns in all nine industry sectors. Similarly, Yi et al. (2022) warn about Chinese companies' overvaluation of stock prices due to investors' high positive sentiment when they find a connection between the CEO of private companies with politics, leaving it as a systematic risk factor.

In addition, there are the alliances of these private groups to obtain profitable future projects through sentiment analysis. De Silva et al. (2021) examine university-industry collaborations (UICs) with a set of 415 final reports finding "*there is a negative relationship between collaborators' perceived challenges and the benefits of UICs, mediated by negative affective evaluation. In contrast, a positive affective evaluation of UIC is positively correlated with its perceived benefits*" (p. 1).

On the other hand, sentiment analysis includes government interventions as participants in their autonomous central banks' monetary policy-making decisions. Picault et al. (2022) constructed a new index to capture media sentiment on the decisions made by the European central bank to analyze the transmission of policy to the remaining economic actors, where press conferences such as the bank president's meetings significantly affect the indicator. Debata et al. (2021) examine the predictive impact of monetary policy on liquidity during different scenarios of investor sentiment; therefore, they concluded that a greater effect of monetary policy on liquidity during a low value in the sentiment index compared to a high sentiment or optimism.

METHODOLOGY

For the analysis of the different communities from the analysis of structural networks on the tweets captured for the digital ecosystem designed, the detection of the topics, and the visualization of the sentiment identified, Figure 1 shows each stage to achieving the process as mentioned earlier. It is worth mentioning that the accounts belong to the different actors of the banking sector in Colombia for the selected social network.

In the first stage, the accounts ecosystem to be analyzed is defined; then, the ecosystem is entered through the Twitter API to obtain the profile data of each account, its followers, and comments within the timeline. For example, the accounts contain the financial institutions of the banking industry, as well as the official accounts publishing tweets for the promotion of financial services about the entities they manage or the dissemination of new news about their corresponding institutions.

Figure 1. Process of analysis of topics and communities in digital social network
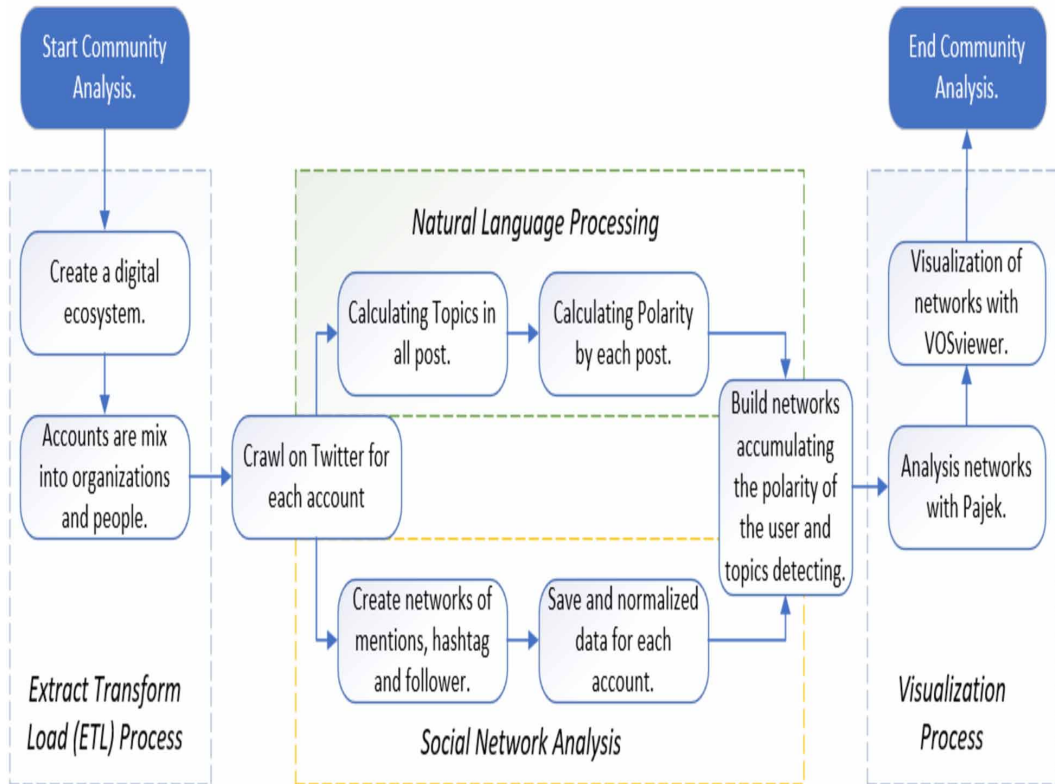
Subsequently, the processes are carried out to "weave" or "build" the networks associated with mentions, hashtags, and followers. After that, two processes of NLP are carried out: the topics of all the collected comments are modeled from the analyzed accounts, and with comments, the polarity is calculated; as a result, a file is created that defines positive, negative and neutral clusters for the mentions and hashtags networks.

Next, the list of mentions and hashtags is stored; the value of the vector is normalized. Then, the amount of mentions and hashtags per user mentioned are accumulated to determine the importance; in this case, a polarity lexicon called "CSL: A Combined Lexicon in Spanish" was used through a BoW process (Word Bag) (Moreno-Sandoval et al., 2017).

Hence, one of the methods for detecting community is the structure of networks by hierarchical grouping, where the weight W_{ij} is calculated for each pair i, j of vertices in the network, which represents how connected the vertices are; as more nodes are added, the resulting graph shows a nested set (Girvan & Newman, 2002). With this in mind, the main structural measures to analyze in the present study are:

Degree Centrality: The node's importance falls to the number of adjacent nodes (degree of connections). As a higher density, the node gets more influence on the network (Parthasarathy, 2011). Most digital social networks suffer the power distribution phenomena, where the minority of nodes have a much higher density of connections than the rest of the network.

The degree centrality is defined by the equation (1):

$$C_D\left(v_i\right) = d = \sum_j A_{ij}$$ (1)

Where C_D is the density of connections over a node v_i. At the same time, it is the sum of the adjacency matrix A_{ij} of node connections.

Formally defined by the equation (2):

$$D_{avg}\left(v_i\right) = \frac{1}{n-1}\sum_{j\neq 1}^{n} g(v_i, v_j) \tag{2}$$

Where n is the number of nodes, and $g(v_i, v_j)$ is the distance between the nodes v_i and v_j. The average distance is the time spent by the node v_i to reach the entire network. It means that a node is more valuable if its proximity is greater.

The model uses Latent Dirichlet Allocation (LDA), a probabilistic generative model that includes three levels of structure: words, topics, and documents (Cao et al., 2009). Given a corpus D containing V unique words and M documents, each document contains a sequence of words $d\left\{W_1, W_2 \cdots W_{Nd}\right\}$.

Given an appropriate number of topical K, the generative process for each document d is as follows:

Given a K-vector θ_d of the distribution $\rho(\theta / \alpha)$, where θ_d is the proportion of mixed topics in the document d. For $i = 1 \cdots N_d$ given the w_i in document d with a specific multinomial distribution $\rho(w_n / \theta_d, \beta)$ where α is a vector-k of parametric Dirichlet and $\rho(\theta / \alpha)$ in the equation (3) is given by:

$$\rho\left(\theta / \alpha\right) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\theta_1^{\alpha_1 - 1}\cdots\theta_{1k}^{\alpha_k - 1} \tag{3}$$

Where the equation (4) contains to β like a $K * V$ word probability matrix.

$$\beta_{ij} = \rho(w_j = 1 / z_i = 1), i = 0,1\cdots K; j = 0,1\cdots V \tag{4}$$

The detection of topics was done through the LDA technique calculating 10 clusters of words that describe ten different topics. Then, these were analyzed with the base banking domain classification provided by Moro et al. (2015); this classification uses dictionaries with a list of composed terms of one or more words (n-grams) using the stemming technique to reduce similar words to its root (for example, "banking" and "banks" are transformed in "bank"). Next, an extended list of related terms that includes other concepts in the same domain was created, for example, "fleteo" [2] and "embezzlement," to analyze user satisfaction and customer service.

Finally, the Louvain algorithm executes the task of detecting communities; the objective is to identify user groups with different roles in the banking system, which seeks to profile the interests of the clients. According to De Meo et al. (2011), the technique Louvain (L.M.) is based on local information and is suitable for analyzing large weighted networks. It consists of two steps: i) each node is assigned to the chosen community with maximum modularity Q; the increase derived from moving a node i to a community C can be calculated through the equation (5) defined by Pasquale.

$$\Delta Q = \frac{\sum_{c} + k_i^c}{2m} - \left(\frac{\sum_{c} * + k_i}{2m}\right) - \left[\left(\frac{\sum_{c}}{2m}\right) - \left(\frac{\sum_{c} *}{2m}\right)^2 + \left(\frac{k_j}{2m}\right)\right] \tag{5}$$

Where $\sum_{c}$ is the sum of the weights of the inner edges inside C, $\sum_{c} *$ is the sum of the weights of the edges incident to a node in C, k_i is the sum of the weights of the edges of i to the nodes in C, m is the sum of the weights of all the edges in the network; ii) The second step consists in making a new network of nodes where are all communities previously found. Then, the process is iterative until a significant improvement of the network modularity is found (De Meo et al., 2011, p. 3).

In conclusion, the bank communities' data process a structural analysis of networks, Louvain, and LDA to test the approach's applicability. In addition, a visualization process is developed on the different techniques using software for network scientometry analysis (Pajek) and visualization of the resulting networks (VOSViewer).

DEVELOPMENT

This study consists of two phases to integrate a hybrid approach to describe the banking industry in Colombia from a corpus collected on Twitter; the first phase consists of applying SNA techniques to describe the structure formed from the interactions of financial institutions, their officers, and network actors. In the second phase, a series of PLN algorithms are integrated with a visualization process of networks to measure and analyze the sentiment perceived in the mentions of the network actors, stakeholders, their topics, and their relationships are identified and classified. It was conducted from the first of January to March 2019, with 53 accounts corresponding to Colombian banks, state entities, and related banking sector entities. Consequently, Table 1 summarizes the collected data:

Structure Network

This study takes the social network Twitter, for which reason, the study took all tweets in the ecosystem of social network accounts and analyzed the hashtags at the structural level; Figure 2 produces the structural measures showing the hashtags interconnections density of the banking sector accounts:

In Twitter, the nodes with the largest number of followers (degree) when posting a tweet are more likely to impact the network only by reaching a greater number of users (Moreno & Pantoja, 2019). Also, the level of degree centralization is 0.1902; in other words, 19% of the edges are connected to each node in the graph. Also, applying the degree metric to the 1081 nodes; thus, Table 2 shows the first ten entities in the digital ecosystem, and these are:

Nevertheless, the impact of a tweet on the network does not necessarily depend on the number of followers but on the number of responses or mentions generated as a result of the user tweets; in this case, Table 2 can be observed the high importance of second-tier banking to provide loans to farmers

Table 1. Dataset

Description	Count
Follower	25286
Hashtag	72581
Mention	116731
Profile	25341
Tweets	141471

Figure 2. Structural measures of hashtag network: Degree centrality
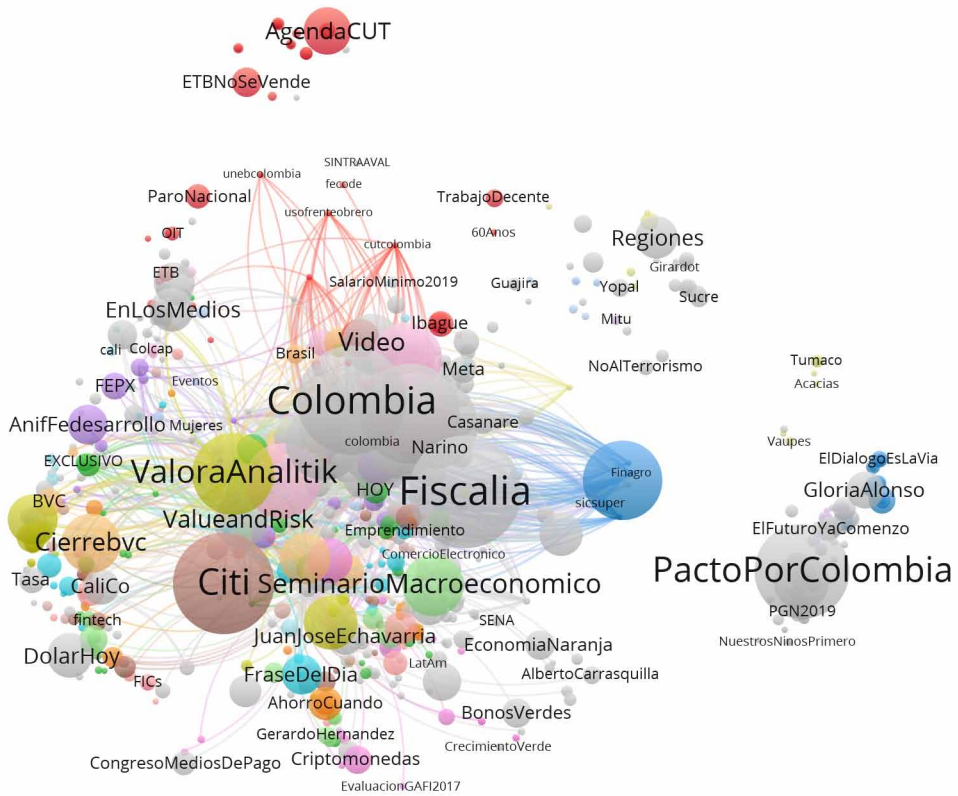
Table 2. Degree-banking sector

Rank	Vertex	Value	Id
1	1396	351	Finagro
2	1232	322	BancoAgrario
3	1628	289	MinHacienda
4	1418	242	BancoWOficial
5	86	234	Asobancaria
6	497	228	cutcolombia
7	1410	226	Fogafin
8	1801	225	Findeter
9	666	222	ValueandRisk
10	1081	221	Bancoldex

through financial intermediaries (Finagro), the sate financial entity to provide banking services to the rural sector (BancoAgrario), ministry of finance and public credit (MinHacienda) and banking institution to grant loans to the micro and small business sector (BancoWOficial).

On the other hand, the structural analysis of intermediation (betweenness centrality) identifies the shortest paths to a user. Intermediation analysis considers an influential node when the number of connections that arrive at the node and interactions like answers, mentions, and retweets are considerable due to its control over information within the network; standing out that Finagro, BancoAgrario, MinHacienda, and Colombian Association of Cooperatives (AscoopColombia) generates the network's highest intermediation as show Figure 3.

Table 3 contains the ten entities that stand in the metric betweenness centrality.

According to Table 3, Finagro, and the Association of Banks and Financial Institutions (Asobancaria) are the nodes of banking entities with the most significant influence on network information. In all three cases, their values exceed 10% as a proportion of the number of times each node is in one of them to learn about banking industry information. In contrast, the degree of intermediation of the remaining nodes does not exceed 9%, and the difference between their values is minimal; surprisingly, MinHacienda is not in the top 3, and the cutcolombia intermediaries are the same as the securities rating agency (ValueandRisk).

Now, a more thorough analysis integrating the results of Table 2 and Table 3 let us identify that almost all nodes (actors) with greater centrality or hashtag number have a high degree of intermediation or influence on the network; in other words, their positions do not vary between these two measures.

Finagro and Bancoagrario are the network leaders since their measures are the highest in both centralities. In particular, the central union of workers (cutcolombia) has many followers, but it lacks influence, contrary to the development bank that promotes business growth and foreign trade (Bancoldex), which degree of intermediation or influence is more considerable with a smaller number of followers than cutcolombia.

Figure 3. Structural measures of networks betweenness centrality - hashtag
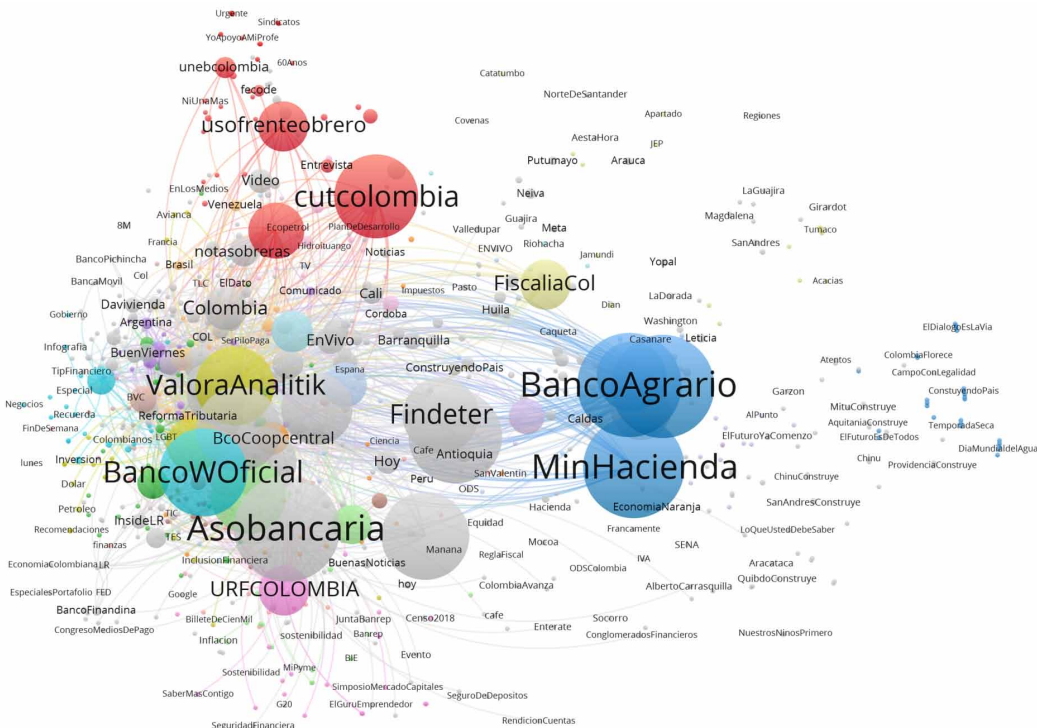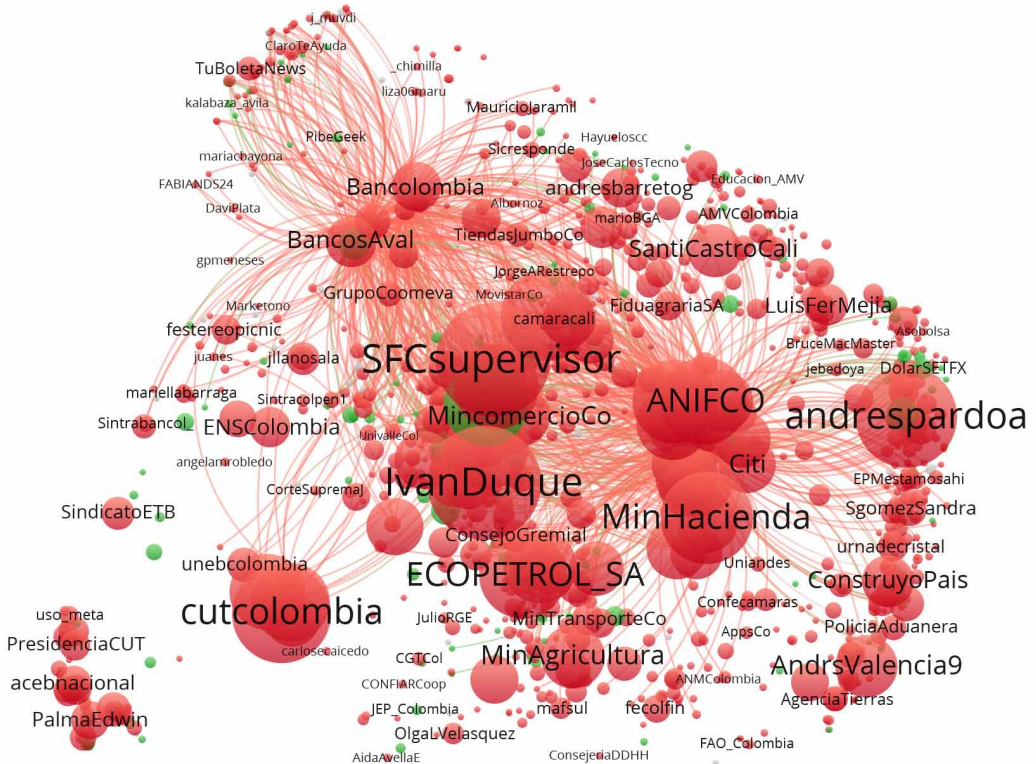
Table 3. Betweenness centrality: Banking sector

Rank	Vertex	Value	Id
1	1396	0.11	Finagro
2	1232	0.106	BancoAgrario
3	86	0.101	Asobancaria
4	1628	0.088	MinHacienda
5	792	0.085	AscoopColombia
6	1418	0.084	BancoWOficial
7	1410	0.079	Fogafin
8	1081	0.078	Bancoldex
9	666	0.074	ValueandRisk
10	497	0.074	cutcolombia

Figure 4. Polarity of mentions

Polarity Analysis

According to the mentions information analyzed, Figure 4 is obtained.

As Figure 4 shows, negative polarity is evident in Colombia's banking sector, pointing to the Ministry of Finance and Public Credit and its ex-Vice Minister Andres Pardo, who served as chief

adviser to President Duque; in contrast, the bank Davivienda stands out for its positive polarity. Likewise, the leading magazine in news and information of economy and businesses with the highest circulation and readership in Colombia (Portafolioco) presents a positive polarity that indicates credibility on economic issues in the sector.

Table 4 shows the most significant nodes or actors concerning the mentions made in the social network. In summary, most of the main actors have a negative polarity, as mentioned in Figure 4. Likewise, Table 4 shows the mention's polarity values representing the magnitude of the sentiment identified as shown by the sizes of the nodes in the network of mentions observed in Figure 4; for example, the first two actors have values above 1900, emphasizing the big negative sentiment identified, which show a more considerable difference of their values to the third node than that observed in the remaining seven nodes.

Additionally, the method for detecting communities Louvain as modularity analysis (Berrocal et al., 2013) is in Figure 5, where six different clusters or communities were generated by followers and identified by color. Therefore, three of the clusters represented by green, red, and blue colors visually have a higher number of nodes representing the closeness of the trait of those nodes concerning the calculated cluster; in contrast, the blue and purple clusters have the least number of nodes.

Table 5 presents some entities representing the clusters identified by the Louvain technique. The results show an essential fact about some of the communities; an example is the case of communities 1 and 2, where the nodes are representatives of public institutions and private sector banks, respectively.

On the other hand, the model Latent Dirichlet Allocation (LDA) was applied. Table 6 shows the ten (10) topics identified and the actors who relate their tweets to the respective topic. Results show that banking's main topics refer to tax reform, payment capacity, industrial property, currency exchange, customer service, debit card, national development planning, line of credit, investment fund, and economic indicators.

Table 7 analyzes each of the four clusters from the dendrogram shown in Figure 6 with the mentions to describe how they are composed. According to Figure 6, the most similar nodes are dataIFX, fiduciary association (Asofiduciaria) with Davivienda, and central bank (BancoRepublica) since the link's height is the smallest. Then, there are the bank BancodeBogota, state development planning entity (DNP_Colombia), pension and severance fund manager (PorvenirOficial), teachers' union (fecode), worker union of Aval group bank (SINTRAAVAL), and AscoopColombia with the leading investor in the structuring, management, and administration of companies and projects in Colombia (corficolombiana), regulatory projection and financial regulation studies office (URFCOLOMBIA) and Colpatria bank. Then, the most similar nodes can be discovered by reiterating the height observation in the dendrogram.

Table 4. Polarity of mentions

Rank	Vertex	Value	Id	Polarity
1	9	1942	Fcsupervisor	-
2	105	1930	IvanDuque	-
3	221	1538	andrespardoa	-
4	469	1475	cutcolombia	-
5	130	1295	MinHacienda	-
6	151	1260	BancoRepublica	-
7	213	1144	ECOPETROL_SA	-
8	113	1128	ANIFCO	-
9	66	1091	larepublica_co	-
10	7	1089	Portafolioco	+

Figure 5. Louvain followers
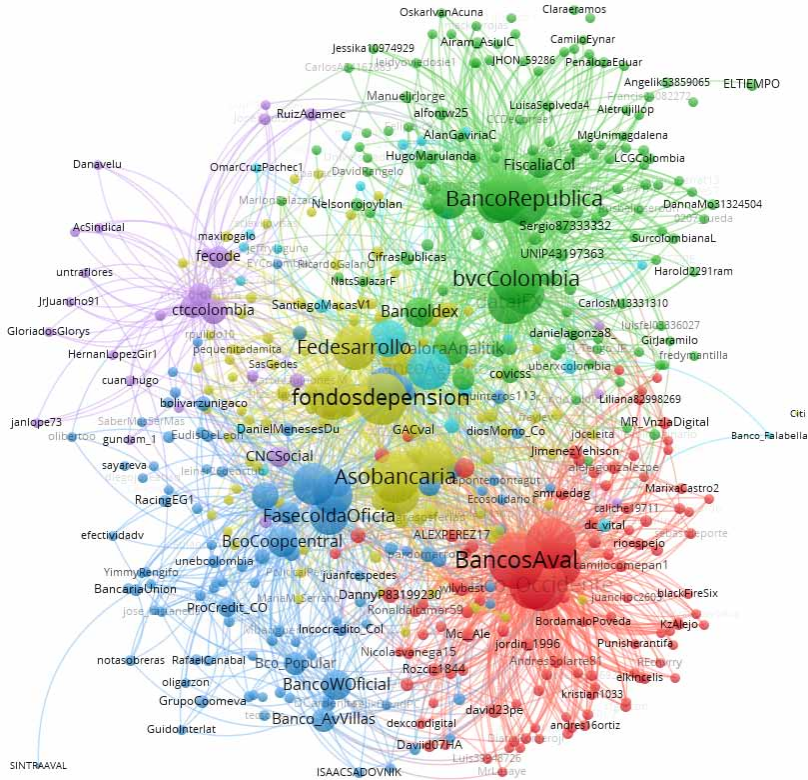
Table 5. Louvain followers

Cluster	Entities
1	@BancoRepublica @FiscaliaCol, @bvcColombia,@Bancoldex, @dataiFX
2	@Banco_AvVillas, @BancoWOficial, @Banco_Popular, @BancodeBogota, @BancoOccidente
3	@fecode, @ctccolombia, @cutcolombia, @asSindical, @CNCSocial
4	@Fedesarrollo, @fondospension, @Asobancaria, @corficolombiana, @ANIFCO
5	@Fogafin, @ValuandRisk, @Asofiduciarias, @FasecoldaOficia, @BcoCoopCentral
6	@Finagro, @BancoAgrario, @Banco_Falabella, @dc_vital, @ExpressLogstic2

When reviewing the different methodologies of the network, that is, using degree and the structural analysis of intermediation (betweenness centrality), and polarity analysis, two organizations of totally different characters and orientations are entities that appear in all methods; they are the MinHacienda and cutcolombia.

However, when reviewing these entities' orientations, there is at least one labor union in clusters 1, 3, and 4. Reviewing the clusters according to the "Ministry of Finance" in every cluster appears an entity associated with the Colombian government, either a bank such as Agriculture Bank and BancoRepublica or a control and regulation entity like prosecutor office (Fiscalia), superintendence of industry and commerce (Sicsuper), tax and customs national authority (DIAN). When analyzing the clusters for the "Asobancaria" entity, it is observed that everyone has at least one association of companies or entities related to finances' management.

Table 6. Related entities: Topic banking

No	Topics	Banks or Related Entities
1	Tax Reform	Banco AVVillas, Porvenir, SINTRAVAL, CTCColombia, Banco Caja Social, Grupo Coomeva, ANIF, Findeter.
2	Payment capacity	Value and Risk, AscoopColombia, Banco Davivienda, Superfinanciera
3	Industrial propriety	Notas obreras, CNC Social, SICSuper
4	Currency exchange	Banco de la República, USO Colombia, Fecode, CUT Colombia, uneb, DataIFX, Bolsa de Valores de Colombia, Banco Pichincha, URFCOLOMBIA
5	Customer service	Bancolombia, Colpatria, MinHacienda, DIAN
6	Debit card	Bancos Aval, Banco de Occidente, Banco Popular, Banco Agrario, Bank ProCredit, Banco Finandina
7	National development planning	Corficolombiana, ValoraAnalitik, Fedesarrollo, BBVA en Colombia, Asobancaria, Fogafín, DNP, Fasecolda, Asofondos
8	Debit Line	Banco de Bogotá, Bancamía, Banco Mundo mujer, Bancoldex
9	Investment fund	Citibank, Asofiduciarias
10	Economic indicators	Fiscalía, El Tiempo, Banco W, Banco Falabella

Table 7. Cluster characterization

| Cluster	Description	Members
1	It is a group made up of banking entities, public and private control, and rating entities.	ProCredit, BancoWoficial, Asobancaria, usofrenteobrero, ValueandRisk, Fiscalia, and BancoPopular
2	This group is made up of banks, insurance companies, and as a not-so-common element, the newspaper El Tiempo.	Asofiduciaria, Davivienda, Banco de la República, Banco de Occidente, El Tiempo, Banco Agrario, Fasecolda y BBVA Colombia
3	This group has a more substantial presence of the labor unions, especially Grupo Aval's labor union and banks and control entities.	SINTRAVAL, AscoopColombia, corficolombiana, URFCOLOMBIA, Colpatria, Sicsuper, Fedesarrollo, and cutcolombia
4	This cluster has an outstanding feature a second-level bank, marked with labor unions, a pension fund, and a regulator and tax collector.	Finagro, notasobreras, ANIFCO, ctccolombia, Fondos de pensiones and DIANColombia

CONCLUSION

The proposed methodology allows analyzing a mixed approach that integrates NLP, CL, and SNA techniques to study social phenomena in any country and domain, in this case, the analysis of banking entities in digital social networks. The integration of these techniques creates layers of information to study the actors by analyzing the content of the node texts, detecting the most relevant nodes of the network for the intermediation and connection of information, identifying the feelings about the most mentioned nodes, detecting the topics in which these nodes are pigeonholed and identifying communities from their digital interaction.

Theoretical Implications

The study used information from Twitter users since this social network has great potential to identify interest groups and know their needs, expectations, tastes, preferences, and opinions of the banking sector. From the linguistic analysis, it was possible to determine the polarity of the users' comments understanding whether these were positive, negative, or neutral. With the structural analysis, it was possible to determine the social role that each of the actors has within the network, thus valuing the

Figure 6. Dendrogram of mentions
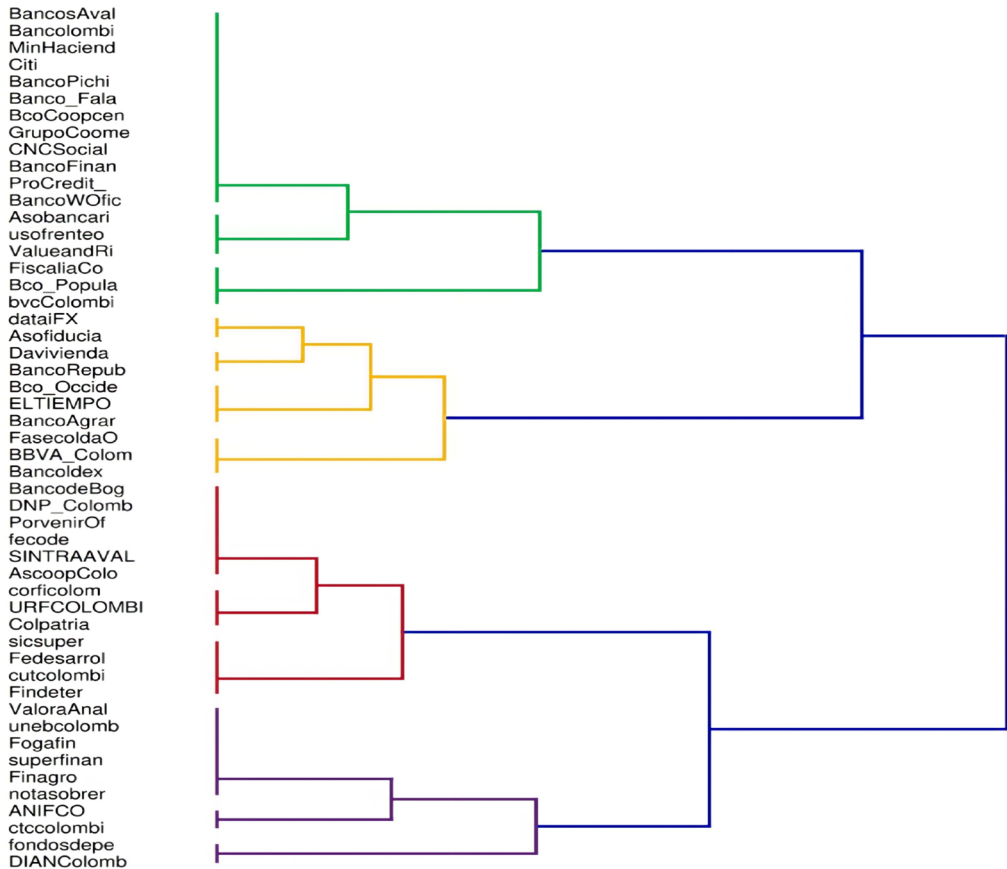
polarity not only from the agglomeration of comments but from the social position of the user from its centrality or intermediation of information, as suggested to us by authors such as Domingues J. (2017) from the perspective of collective subjectivity.

The integration of SNA, as mentioned by Parthasarathy (2011), discovered significant patterns to analyze the structural form in which they are organized and visualize the results of NLP algorithms utilizing network graphs. Like Miranda et al. (2013) wanted better understand the banks' behavior in SNS, the integration of the techniques highlighted the most relevant nodes in terms of connections and information brokering to be the pillars of change of the negative sentiment calculated for the prevention of systematic risk.

Government interventions also played an essential role in the analysis of the Colombian banking industry; therefore, mentions as MinHacienda in all the techniques applied, the great centrality of the institutions BancoRepublica, Bancoldex, and Bancoagrario, and the great sentiment of their officials such as President Ivan Duque, andrespardoa, follow the line of determining the primary factors by (Debata et al., 2021; Picault et al., 2022) on the part of the state.

Managerial Implications

According to the polarity study, the structural model shows that the banking system's confidence in Colombia has a significant negative polarity, which can give clues to the banks to improve their policies with topic detection in which they were framed and improve their management with the disclosure in

networks about the connections that the centrality measures highlighted. However, Davivienda bank stands out for its positive polarity in the digital ecosystem. Furthermore, the leading economic and business news magazine "Portafolio" presents a positive polarity that indicates credibility in handling the sector's financial information.

The topic detection model identified that in social networks (Twitter), the banks associated with customer service are, for example, Bancolombia and Colpatria, and the credit line users associate with Banco de Bogota. These types of connections generate essential information to identify if their clients perceive the associations of the banks' approaches correctly.

A shared language is a key distinguishing characteristic of a human being, which is still present in digital social networks. The present work has obtained exciting results through a community detection approach. A study in the banking sector through language on Twitter allowed the detection of specific sociolinguistic characteristics that reveal a specific vocabulary for each community and therefore show users' interests. Thus, recognizing the vocabulary to standardize the terms used in the tweets, taking as an example the banks or financial institutions with a positive sentiment that belong to the same topic, can reduce the systematic risk with the idea of generating a contagion effect in their favor.

LIMITATIONS AND FUTURE WORKS

One of the predominant entities in the different clusters are the trade union entities; It is essential to check whether the different users use the trade union associations to support their complaints or whether the relations between the user and the union reflect a negative score in the bank employees regarding their working conditions. This hypothesis must be thoroughly reviewed, making it open to future work.

While the different CL and NSA techniques yielded positive results given the novel approach to the systematic risk problem, it is pertinent to highlight the importance of exploring more granular levels with more complex community detection algorithms to examine sentiment variations in these new communities against the backdrop of negative sentiment in the banking industry.

Additionally, NLP is a whole world in itself and has challenges that are still being solved; that does not discount the contribution obtained in the development of this article; however, it is recognized that there are still algorithms, such as neural networks, that can give greater accuracy to the calculated sentiment.

At this point, it is also worth highlighting the importance of deeper network models that can draw from other sources, such as specialized lexicons for the Spanish language spoken in Colombia, to trace new relationships between the actors of the digital ecosystem.

Finally, the importance of the data shared in social networks is a limitation because although these represent the primary input for the experiments described access to data from these official accounts is sometimes restricted. Therefore, the inability to process such data is a limitation that directly affects the results and the degree of analysis of the experiments performed.

ACKNOWLEDGMENT

REFERENCES

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *The American Economic Review*, *105*(2), 564–608. doi:10.1257/aer.20130456

Afolabi, I. T., Ezenwoke, A. A., & Ayo, C. K. (2017). Competitive analysis of social media data in the banking industry. *International Journal of Internet Marketing and Advertising*, *11*(3), 183–201. doi:10.1504/IJIMA.2017.085644

Amaya-cristancho, H. A., & Cortés-vargas, Y. L. (2014). El fleteo: "la abstracción de un riesgo criminal". Una experiencia de inteligencia criminal. *Revista Criminalidad*, *56*(1), 97–131.

Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 60–69. doi:10.1038/scientificamerican0503-60 PMID:12701331

Bargigli, L., di Iasio, G., Infante, L., Lillo, F., & Pierobon, F. (2015). The multiplex structure of interbank networks. *Quantitative Finance*, *15*(4), 673–691. doi:10.1080/14697688.2014.968356

Berger, A. N., & DeYoung, R. (2005). Technological Progress and the Geographic Expansion of the Banking Industry. SSRN *Electronic Journal, 38*(6), 1483–1513. 10.2139/ssrn.318703

Bermingham, A., & Smeaton, A. F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. *Proceedings of the Workshop on Sentiment Analysis Where A.I. Meets Psychology (SAAIP 2011)*, 2–10. https://www.researchgate.net/publication/267250109

Berrocal, J. L., Figuerola, C., & Medarno, Jo. (2013). Recolección, detección de comunidades y visualización de información web. *II Seminario Hispano-Brasileño de Investigación En Información, Documentación y Sociedad (2013)*, 255–267.

Bohlin, E., Shaikh, A. A., & Hanafizadeh, P. (2018). Social network banking: A case study of 100 leading global banks. *International Journal of E-Business Research*, *14*(2), 1–13. doi:10.4018/IJEBR.2018040101

Boss, M., Elsinger, H., Summer, M., & Thurner, S. (2004). Network topology of the interbank market. *Quantitative Finance*, *4*(6), 677–684. doi:10.1080/14697680400020325

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7–9), 1775–1781. doi:10.1016/j.neucom.2008.06.011

Capocci, A., Servedio, V. D. P., Caldarelli, G., & Colaiori, F. (2004). Communities Detection in Large Networks. Lecture Notes in Computer Science, 3243, 181–187. doi:10.1007/978-3-540-30216-2_15

Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion. *Neurocomputing*, *264*, 50–56. doi:10.1016/j.neucom.2016.10.101

Cont, R., Moussa, A., & Santos, E. B. e. (2012). Network Structure and Systemic Risk in Banking Systems. SSRN *Electronic Journal*. 10.2139/ssrn.1733528

de Albornoz, J. C., Plaza, L., Gervás, P., & Díaz, A. (2011). A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. *Proceedings of the European Conference on Information Retrieval (ECIR), 6611*, 784–787. doi:10.1007/978-3-642-20161-5_8

de Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. *International Conference on Intelligent Systems Design and Applications, ISDA,* 88–93. doi:10.1109/ISDA.2011.6121636

de Silva, M., Rossi, F., Yip, N. K. T., & Rosli, A. (2021). Does affective evaluation matter for the success of university-industry collaborations? A sentiment analysis of university-industry collaborative project reports. *Technological Forecasting and Social Change*, *163*, 120473. Advance online publication. doi:10.1016/j.techfore.2020.120473

Debata, B., Dash, S. R., & Mahakud, J. (2021). Monetary policy and liquidity: Does investor sentiment matter? *IIMB Management Review*, *33*(3), 257–277. doi:10.1016/j.iimb.2021.07.001

Domingues, J. M. (2017). Subjetividad Colectiva: Un concepto, una teoría. *Cuadernos de Teoría Social*, *6*, 38–48.

Elsinger, H., Lehar, A., & Summer, M. (2006). Risk assessment for banking systems. *Management Science*, *52*(9), 1301–1314. doi:10.1287/mnsc.1060.0531

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. doi:10.1073/pnas.122653799 PMID:12060727

Gong, X. L., Liu, X. H., Xiong, X., & Zhang, W. (2019). Financial systemic risk measurement based on causal network connectedness analysis. *International Review of Economics and Finance, 64*, 290–307. 10.1016/j.iref.2019.07.004

Houston, J. F., Lee, J., & Suntheim, F. (2018). Social networks in the global banking sector. *Journal of Accounting and Economics*, *65*(2–3), 237–269. doi:10.1016/j.jacceco.2017.11.006

Karavitis, P., & Kazakis, P. (2022). Political sentiment and syndicated loan borrowing costs of multinational enterprises. *Journal of International Financial Markets, Institutions and Money*, *78*, 1–23. doi:10.1016/j.intfin.2022.101537

Keskar, M. Y., & Pandey, N. (2018). Internet banking: A review (2002–2016). *Journal of Internet Commerce*, *17*(3), 310–323. doi:10.1080/15332861.2018.1451969

Keskar, M. Y., Pandey, N., & Patwardhan, A. A. (2020). Development of conceptual framework for internet banking customer satisfaction index. *International Journal of Electronic Banking*, *2*(1), 55–76. doi:10.1504/IJEBANK.2020.105417

Khatua, A., Khatua, A., & Cambria, E. (2020). Predicting political sentiments of voters from Twitter in multi-party contexts. *Applied Soft Computing Journal, 97*(Part A). 10.1016/j.asoc.2020.106743

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *80*(5), 1–11. doi:10.1103/PhysRevE.80.056117 PMID:20365053

Leitner, Y. (2005). Financial Networks: Contagion, Commitment, and Private-Sector Bailouts. SSRN *Electronic Journal, 02*. 10.1111/j.1540-6261.2005.00821.x

Leventides, J., Loukaki, K., & Papavassiliou, V. G. (2019). Simulating financial contagion dynamics in random interbank networks. *Journal of Economic Behavior & Organization*, *158*, 500–525. doi:10.1016/j.jebo.2018.12.017

Liu, D., & Lei, L. (2018). The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 U.S. presidential election. *Discourse, Context & Media*, *25*, 143–152. doi:10.1016/j.dcm.2018.05.001

Loe, C. W., & Jensen, H. J. (2015). Comparison of communities detection algorithms for multiplex. *Physica A*, *431*, 29–45. doi:10.1016/j.physa.2015.02.089

Mabic, M., Gašpar, D., & Lucovic, D. (2017). Presence of banks on social networks in Bosnia and Herzegovina. *Business Systems Research*, *8*(2), 59–70. doi:10.1515/bsrj-2017-0016

Martinez-Jaramillo, S., Alexandrova-Kabadjova, B., Bravo-Benitez, B., & Solórzano-Margain, J. P. (2014). An empirical study of the Mexican banking system's network and its implications for systemic risk. *Journal of Economic Dynamics & Control*, *40*, 242–265. doi:10.1016/j.jedc.2014.01.009

Miranda, F. J., Chamorro, A., Rubio, S., & Morgado, V. (2013). Evaluation of Social Networks Sites in the Banking Sector: An Analysis of Top 200 International Banks. *Journal of Internet Banking and Commerce*, *18*(2), 17.

Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, *53*(4), 675–679. doi:10.1016/j.dss.2012.05.022

Moreno, S. L. G., & Pantoja, R. L. M. (2019). Analytics applied to the study of reputational risk through the analysis of social networks (Twitter) for the El Dorado airport in the city of Bogotá (Colombia). *ICEIS 2019 - Proceedings of the 21st International Conference on Enterprise Information Systems, 1*, 476–483. doi:10.5220/0007770804880495

Moreno-Sandoval, L. G., Beltrán-Herrera, P., Vargas-Cruz, J. A., Sánchez-Barriga, C., Pomares-Quimbaya, A., Alvarado-Valencia, J. A., & García-Díaz, J. C. (2017). CSL: A Combined Spanish lexicon: Resource for polarity classification and sentiment analysis. *ICEIS 2017 - Proceedings of the 19th International Conference on Enterprise Information Systems, 1*, 288–295. doi:10.5220/0006336402880295

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, *42*(3), 1314–1324. doi:10.1016/j.eswa.2014.09.024

Mucan, B., & Özelturkay, E. (2014). Social Media Creates Competitive Advantages : How Turkish Banks Use This Power? A Content Analysis of Turkish Banks through Their Webpages. *Procedia: Social and Behavioral Sciences*, *148*, 137–145. doi:10.1016/j.sbspro.2014.07.027

Nicoletti, B. (2017). The Future of A.I. Artificial Intelligence for. NET: Speech, Language, and Search. doi:10.1007/978-1-4842-2949-1_10

Nopp, C., & Hanbury, A. (2015). Detecting risks in the banking system by sentiment analysis. *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, September*, 591–600. doi:10.18653/v1/D15-1071

Ozdora, E., & Atakan, S. (2016). Gaining legitimacy through CSR: An analysis of Turkey's 30 largest corporations. *Business Ethics (Oxford, England)*, *25*(3), 238–257. doi:10.1111/beer.12114

Parthasarathy, S. (2011). *Social Network Data Analytics*. doi:10.1007/978-1-4419-8462-3

Picault, M., Pinter, J., & Renault, T. (2022). Media sentiment on monetary policy: Determinants and relevance for inflation expectations. *Journal of International Money and Finance*, *124*(102626). Advance online publication. doi:10.1016/j.jimonfin.2022.102626

Porras, H., & Orozco, L. (2019). Riesgo Reputacional. *Experto*, *9*(2), 117–152. doi:10.2307/j.ctvkrkk8z.10

Puertas, E., Moreno-Sandoval, L. G., Redondo, J., Alvarado-Valencia, J. A., & Pomares-Quimbaya, A. (2021). Detection of Sociolinguistic Features in Digital Social Networks for the Detection of Communities. *Cognitive Computation*, *13*(2), 518–537. doi:10.1007/s12559-021-09818-9

Rogers, L. C. G., & Veraart, L. A. M. (2012). Failure and Rescue in an Interbank Network. SSRN *Electronic Journal, 2007*, 1–33. 10.2139/ssrn.1932911

Schulte, P. (2018). Mobile Technology: The New Banking Model Connecting Lending to the Social Network. In Handbook of Blockchain, Digital Finance, and Inclusion (Vol. 2). Elsevier Inc. doi:10.1016/B978-0-12-812282-2.00013-9

Sui, P., Tanna, S., & Zhou, D. (2020). Financial contagion in a core-periphery interbank network. *European Journal of Finance*, *26*(7–8), 691–710. doi:10.1080/1351847X.2019.1630460

Tiwari, A. K., Abakah, E. J. A., Bonsu, C. O., Karikari, N. K., & Hammoudeh, S. (2022). The effects of public sentiments and feelings on stock market behavior: Evidence from Australia. *Journal of Economic Behavior & Organization*, *193*, 443–472. doi:10.1016/j.jebo.2021.11.026

Uhde, A., & Michalak, T. C. (2010). Securitization and systematic risk in European banking: Empirical evidence. *Journal of Banking & Finance*, *34*(12), 3061–3077. doi:10.1016/j.jbankfin.2010.07.012

Wills, P., & Meyer, F. G. (2019). Change Point Detection in a Dynamic Stochastic Blockmodel. *Complex Networks 2019: Complex Networks and Their Applications VIII, 881*, 211–222. https://www.cambridgescholars.com/complex-networks-and-their-applications-9

Yi, S., Wang, J., Wang, X., & Feng, H. (2022). CEO political connection and stock sentiment beta: Evidence from China. *Pacific-Basin Finance Journal*, *74*, 101813. Advance online publication. doi:10.1016/j.pacfin.2022.101813

ENDNOTES

[1] The public "Application Programming Interface" APIs on Twitter are limited to downloading the last 3000 tweets from each account. The information associated with these APIs can be consulted in https://developer.twitter.com/en/docs/api-reference-index.

[2] Fleteo is to rob bank clients of their money just after they withdraw funds from their accounts (Amaya-cristancho & Cortés-Vargas, 2014).

Luis G. Moreno Sandoval is a Ph.D. in Engineering, Pontificia Universidad Javeriana. Master in Business Management and Digital Marketing Business School ENAE Campus Universitario de Espinardo, Murcia (Spain). MBA Master in Business Administration. The Externado University of Colombia. Master in Information Sciences and Communications from Universidad Distrital Francisco José de Caldas. Systems Engineer, Los Libertadores University Foundation. Researcher and profesor in Computational Linguistics and Social Networks Analysis at Pontificia Universidad Javeriana.

Liliana M. Pantoja Rojas is a Ph.D. student in Engineering at Universidad Distrital Francisco José de Caldas. Master in Industrial Engineering from Universidad Distrital Francisco José de Caldas, Master in Integral Risk Management from Universidad del Externado and Master in Integral Risk Management from ENAE (Murcia, Spain). Industrial Engineer from Universidad Nacional de Colombia, Bachelor in Mathematics from Universidad Distrital Francisco José de Caldas and Specialist as Educational Manager from Universidad de la Sabana. PMP since 2015, Professor of quantitative analysis at Pontificia Universidad Javeriana, researcher of Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA), researcher of Ministry of Environment and Sustainable Development, researcher of Special Jurisdiction for Peace and Analytics Project Leader at LUMON.

Alexandra Pomares-Quimbaya is Professor at the Pontificia Universidad Javeriana in courses on information systems and data mining and research foudamentals. Throughout her research work, she has been involved in the area of data analytics and data integration. Currently, she is working in a Colombian initiative that aims to improve the quality of the medical information available to promote value based medicine. She is the director of research vice-rectory at Pontificia Universidad Javeriana.

Luis Antonio Orozco is a Ph.D. in administration at the Los Andes University. Adjoined researcher at Science, Technology and Policy Studies (STePS) Faculty of Management and Governance at University of Twente, Netherlands. Besides, Adjoined researcher at the National Center for Digital Government at the University of Massachusetts. Senior researcher in the field of management in science, technology, and innovation with specialties in Organizational Theory, Networks, and Institutions, Sociology of Science, Scientometrics.