Received 10 September 2022, accepted 20 October 2022, date of publication 26 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217467

RESEARCH ARTICLE

Hybrid Onion Layered System for the Analysis of Collective Subjectivity in Social Networks

LUIS GABRIEL MORENO-SANDOVAL AND ALEXANDRA POMARES-QUIMBAYA
Department of Engineering, Pontificia Universidad Javeriana, Bogota 110231594, Colombia

Corresponding author: Luis Gabriel Moreno-Sandoval (morenoluis@javeriana.edu.co)

This research was carried out by the Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA), led by the Pontificia Universidad Javeriana Colombia and funded by the Ministry of Information Technologies and Telecommunications of the Republic of Colombia (MinTIC) through the Ministry of Science, Technology and Innovation (MinCiencias) within contract No. FP44842- anex46-2015.

ABSTRACT This research aims to analyze the Digital Social Networks (DSN) behavior, constructed from the network's relationships, interactions, and expressions of users' private states through collective subjectivity. For this purpose, an onion-ring system called COSSOL has been built in a case study for Twitter, following a hybrid approach to integrate Machine Learning classifiers and structural metrics from Computational Linguistics and Computational Sociology disciplines, respectively. The paper designs two experimentation scenarios divided into cases of collective subjectivity analysis for Colombia under different levels of communities' granularity. The first case validates the system by performing a cointegration test on the metrics of each construct for the onion rings' communities. The results show that some communities better propagate their subjective expressions against the disclosed topic when they have a higher network density and a common polarity. Moreover, the most stable communities in polarity towards a topic are those whose members are highly connected. Conversely, communities with a higher centrality index in a subset of members do not exhibit stability in collective subjectivity towards a topic disclosed in that community. The second case validates the model with a series of Social Network Analysis (SNA) metrics with a polarity layer to describe the second onion ring subcommunities and their temporal variation through community recalculation. The results show no polar distributions similar to the bimodal ones representing consensus in the values of the common Thinking Acting and Feeling (TAF) forms. In addition, general negative sentiment is identified for the ten most representative nodes of the subcommunities analyzed.

INDEX TERMS Collective subjetivity analysis, digital social networks, network structure, sentiment analysis, social network analysis, subjectivity analysis.

I. INTRODUCTION

The Collective Subjectivity Analysis is one of the phenomena of study that has motivated sociologists, anthropologists, psychologists, and researchers interested in human interactions and the scope of communication. The notions of subjectivity are associated with the property of perceptions, arguments, and language that arise from the subject's point of view and are therefore influenced by the subject's particular interests and desires [1]. However, interactions should not be confined to processes developed by individual actors because collectives also interact [2], [3]. Therefore, Digital Social Networks (DSN) allow the identification of patterns that give rise to a community structure, i.e., cohesive groups or subgroups, which has its origins in the global structure and macrostructure of networks analysis [4], [5].

In recent years, DSNs have been consolidated as a privileged scenario for communication and interaction between people, becoming ideal spaces for socializing, debating, and generating new relationships around topics of interest. As a result, these scenarios present an exchange of high amounts of content, especially in textual format, reflecting the

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Bellan.

collective subjectivity through expressing opinions, emotions, ideas, and private states. In the linguistic field, the most cited authors are [6], [7], [8], and [9], who represent the most relevant notions of the semantics of linguistic expressions as an instrument of analysis of the elements of the message shared in the social network.

The elements that characterize social networks are the different contents (text, image, video) shared by users in their accounts. Among them are the texts called publications, which form a corpus from which it is possible to extract high-value information. [10] relate a group of linguistic features (lexical, syntactic, symbolic, participation, and complementary information) by analyzing the digital identity of the network actor from the collected corpus with the sociographic, demographic and psychographic characteristics of his real identity. In this way, the users' digital identities generate a research space that simultaneously studies the structure of the communities generated based on the interaction of the symbolic, participation, and complementary information features and the sentimental charges of the linguistic expressions in the lexical and syntactic features.

To perform the analysis of collective subjectivity in RSD, the *"Collective Subjectivity Communities in Onion Layers (COSSOL)"* system takes elements from two disciplinary constructs dedicated to the study of communication and human interaction phenomena in the digital scenario; in order to represent a hybrid system. The first is Social Network Analysis (SNA), which identifies the structural patterns in the relationships of actors in a social network, employing network analysis (graphs) framed within the contributions of the field of Computational Sociology. The second, created within the framework of Computational Linguistics, whose object of study is the linguistic expressions of private states present in communicative interactions, is called Subjectivity Analysis (SA), which uses Natural Language Processing (NLP) techniques (text classifiers).

A model is proposed to test the collective subjectivity comprised of the ways of Thinking, Acting, and Feeling (TAF) from the granularity of the communities and an axis from the linguistic construct for the identification, interaction analysis, and characterization of the communities. COSSOL integrates the two constructs based on the principle of "onion rings," which examines the interaction levels of communities in social networks from a granular analysis to analyze collective subjectivity. The COSSOL system was built following the scientific research approach within the design cycle proposed by [11]. The results of the experimentation scenarios will accept or reject the following hypotheses when analyzing the integration of the two constructs.

1) The communities with a higher index of centrality in a subset of members present greater stability in the collective subjectivity in the face of a topic disseminated in that community.
2) The most stable communities in polarity terms concerning a topic are those in which their members are highly connected.
3) The communities with a higher network density and a common polarity in a subset of members are more highly connected to a topic.

The article is in five sections. The second section introduces the fundamental constructs theoretically for the correct interpretation of the present research proposal, in addition to the evolution of each construct starting from recognizing the main milestones that have determined their development and the measurements' definitions used. The third section presents the methodology of the COSSOL system proposal, introducing the general overview of the hybrid system, the description of the onion rings principle, and the explanation of the implementation of the system for the different components developed. The fourth presents the results of the two experimentation scenarios performed, outlining the theoretical implications of the hypotheses raised and detailing the practical findings. Finally, the fifth section presents the conclusions and recommendations.

II. BACKGROUND

A. SOCIAL NETWORK ANALYSIS

A social network consists of relationships between a group of social actors and any additional information about those actors and their relationships [12]. The social network approach consists of patterns associated with social ties, in which actors are involved, and their interactions have significant consequences.

Social network analysts seek to discover various kinds of patterns, determine the conditions under which they emerge, and discover their consequences on the attitudes, perspectives, and behaviors of individuals, groups, or subgroups and the systems to which they belong [13]. According to [14], SNA comprehends four characteristics that have been consolidated since its emergence and whose integration gives rise to a research paradigm:

- It is motivated by a structural intuition based on the ties that bind social actors.
- It is an approach based on systematic and empirical data.
- It is based to a large extent on graphic representations and
- It is supported by using mathematical and computational models.

In addition to the above, there is a broad and growing range of applications, which, thanks to theoretical and applied research, have achieved an important level of generalization that goes beyond the limits of traditional disciplines and summons a wide range of scientific studies ranging from sociology to computer science.

Additionally, the availability of massive amounts of data in the web scenario has given a new statistical and computational impetus to the field of SNA. According to [15], hanks to this emerging phenomenon, social network analysis has taken a different direction related to new approaches to data analytics, which places it within computational social science paradigms along with text analysis (information extraction and classification), social complexity analysis (complexity sciences), and social simulations (agent-based models and cellular automata) [16].

Consequently, the computational challenges associated with the ability to perform mining and analytical processes to these information sources in the context of a social network constitute an unprecedented challenge and an opportunity to determine helpful information in a wide variety of fields, such as marketing, politics, social sciences, among others [4], [17], [18], [19].

In order to understand the processes that led to the SNA emergence and the characteristics that defined its development over time, this section reviews the main milestones in the scientific discipline evolution, which Table 1 summarizes.

The SNA evolution has brought with it the definition of several metrics. To address this area, the predominant approach in SNA has been graph theory, which originated in mathematical research [19], [20]. In this theoretical approach, individuals and groups are represented by points or nodes and their social relationships by lines, which may include a direction representing the influence flow. In addition, it is a theory that provides tools for analyzing the formal properties of the resulting sociograms through a matrix-based approach.

A second approach that has led to the measurements recorded below is Harvard structuralists Harrison White and Doug White [21] who focused on the characteristics of individuals' social positions, roles, and categories. This approach is called the "positional approach" and constitutes a rigorous method of matrices clustering that organizes networks into hierarchical positions to represent established relationships, emergent behaviors, and the consequences of these behaviors on the other actors in the network [4].

From the two approaches described above, measures emerge whose concepts and characteristics at the actor level, link, subgroups, and network are associated either with the structure or with the individual's position in the network [14], [20], [22].

Two concepts are distinguished at the **actor level analysis**: ego networks (ego-centered networks) and complete networks. Ego-centered networks, characterized by having the actor as a referent, consist of the identification and study of the personal networks of an individual actor; in addition, the comparison between the networks established by the actors in the network. Their center is the actor. Complete networks involve the identification and study of the network as a whole.

The **link-level analysis** purpose is to consider the characteristics of the ties that bind the actors in the network. The links are represented by lines connecting the nodes in the network and may or may not have directionality. When they have directionality, it refers to senders and receivers (sending agents and receiving agents), which can establish reciprocal relationships or mutual ties called arcs. If directionality is not, only the existence of a relationship between two actors or nodes is analyzed; in this case, the links are called edges.

TABLE 1. Summary of SNA contributions.

Year	Author	Contribution
1930	George Simmel	Emphasize the formal properties of social interaction to construct a "formal sociology."
1930	Alfred Vierkandt y Leopold von Wiese	Explicitly adopted points, lines, and connections terminology to describe social relationships.
1936	Lewin y Moreno	They introduce the concept of .
1936	Moreno y Jennings	"sociometry" and the idea of representing social structures as diagrams of networks of points and lines called Sociograms.
1939	Roethlisberger y Dickson	Structure approach of group relationships and the initiation of network diagrams to represent them.
1950	George Homans	Development of matrix methods for the explanation of exchange theories.
1953	Cartwright and Zander	Arising of group dynamics approach.
1953	Harary and Norman	Michigan University.
1958	Freeman	Conducts first structural study of community decision-making.
1963	White, Laumann	Explore using algebra to represent affinity structures and the multidimensional scaling methods employment in the social field.
1969	Mitchell,Bott y Barnes	First systematic summaries of a formal social network methodology.
1976	Boorman y White	Extend sociometric methods and matrix analysis methods to study social positions.
1978	Freeman	Founded the Social Network Journal.
1979	Wellman, Stokman, Scott y Griff	Emerge as the most notable exponents of SNA outside North America. Revive the tradition of sociometry with new mathematical and theoretical rigor.
1994	Wasserman y Faust	They redefine, codify and make publicly available the techniques developed by the generation of network analysts of the 1980s and 1990s.
1999	Watts,Strogatz, Barabasí y Albert	They mark a revolutionary change in the field by being the first physicists to publish on social networks, although they are criticized for ignoring all the previous literature built by sociologists.
2000	Scott	Identification of patterns in connections.
2002	Borgatti, Everett y Freeman	Convert the techniques of network analysts to computational algorithms.
2004	Freeman	Performs an analysis of the SNA history evidencing a division between the research done by physicists and sociologists, motivating an essential change in the direction of social network analysis.
2005	Carrington et al.; Burt	Contributions to the ideas of social capital, which connect the topological structure of social networks in the distribution of available resources.
2009	Borgatti et al.	They characterize dynamic structures, positions, and dimensions through graph theory properties to explore the overall structure and interconnections of the network.
2011	Scott	Simplify techniques for measurements of social network structure and dynamics to facilitate understanding mathematical models.
2012	Carley, Kathleen M.	Network modeling, segmentation, and analysis.
2013	Ferrara, Emilio	Use geospatial features for segmentation, community detection, and social movement analysis.
2014	Jung, Jason	Recommender systems, sentiment analysis based on fuzzy propagation in online social networks.

TABLE 1. ***(Continued.)*** **Summary of SNA contributions.**

2015	Farine, Damien R.	Development of a shared decision-making device in social networks.
	Xin, Yu	Algorithms for semantic community detection. Semantic social networks.
2016	Zhang, Bo	Information propagation in social networks. Reputation aggregation in mobile social networks.
2018	Domingues, Mauricio	Production of permanence and change in terms of public opinion, pointing out the mechanisms related to social memory and creativity and the rhythms of social development.
2019	Ureña, et al.	They examine progress in understanding the new possibilities and challenges of trust and reputation systems in social networks.
2019	Zucco, et al.	Methods and main tools from the definition of criteria for sentiment analysis.
2022	Peralta, et al	The exchange leads to archetypal public opinion states such as consensus and polarization.

Source: own elaboration based on [4], [14], [15]

There are at least two types of considerations when analyzing links in a social network. First, the directionality and reciprocity in the flow of information determine a measures group, such as degree centrality, closeness, eigenvector, and beta. The second has to do with identifying the type of established relationship, where the strength of ties, frequency of communication, and presence of trust are measured. The measures at the link level are balance, asymmetry, reciprocity, connectivity, frequency of contact, path length, and structural equivalence [22].

An analysis understands the **subgroup level** in two ways: a cohesive network subsection or a subset of actors sharing a ''position'' or ''role'' in a network. The first case refers to a subset of actors in a network, where a high proportion of these actors are connected through some positive communication link or friendship. It seeks to identify the subgroup's cohesion level, i.e., the degree to which the actors connect to each other.

The second case seeks to locate subsets of actors by studying the level of similarity between them in terms of the role or position they occupy in the network. This grouping is done in ''blocks'' or ''classes'' and is determined by structural characteristics and the individual attributes of the actors that comprise it. The determinant measures at the subgroup level are cohesion, roles, homophily, transitivity, and homogeneity [20].

Network level analysis: the concepts and measures that emerge when considering the network as a whole are primarily associated with sociology as they relate to an individual's membership in a community and the forces that hold people together in a group (network). These ties of belonging to a community imply that similar beliefs and values are shared [23]. Cohesion at the network level is the extent to which actors are connected directly or indirectly. The fundamental difference from the cohesion measure at the subgroup level is that it measures how the network members are associated.

Other associated measures at the network level analysis are density and centralization. Density, or the proportion of the total number of potential ties that are present in the network, does not always express a cohesive network; for example, in the case of vast networks where it is not feasible to achieve a high proportion of relationships, simply because of their size, or when there is a high proportion of ties to a single actor. Centralization evaluates the variability in centrality among network members and therefore needs to be combined with the previous one.

Studies of social relations and human communication address the attributes of each individual who is part of the group being analyzed as the content of the relationships and communication established. In this interrelation process, there are flows of information and resources that impact the form and content of the relationship between actors. SNA has been markedly structural, a perspective focused on the form analysis of relationships. However, as mentioned above, SNA has been addressing the dynamic nature of networks and has sought methods to advance the content analysis of the information shared among the actors involved, which has given rise to a series of advanced information processing techniques that seek to respond to the dynamism and complexity of social networks [15].

Mathematics, statistics, and computer science propose several methods for analyzing these interrelationships and resources, both in form and content. Thus, [20] present a detailed description of mathematical methods used to define structural properties, locations, positions, roles, and statistical models. Reference [24] organize the methods and models reported in the literature around the following concepts: data collection and measurements in networks, network modeling, centrality measures for groups, diffusion models, correspondence analysis for bimodal networks, statistical models, models for longitudinal network data, ways of drawing networks, and computer programs for SNA.

More recently, with the rapid emergence of digital scenarios, methods have focused on the analysis of large volumes of information through a variety of computational techniques, which constitutes a valuable resource for researchers in disciplines such as linguistics, sociology, and computer science, as well as emerging disciplines such as computational sociolinguistics. This discipline starts from the principle that language is an ever-changing social phenomenon and posits that the recent surge of interest among computational linguists in studying language in its social context is due, in large part, to the availability of information from social networks [25],

Of particular interest for SNA oriented to the analysis of collective subjectivity is the identification of patterns that give rise to a community structure [4], [20]. Reference [26] analyzed the influence problems by leveraging standard features to establish a minimum cost to find a set of users with a minimum cardinality that influenced a given fraction of users in multiple social networks.

The challenges of community analysis in social networks are associated with identifying structure, properties, and emerging behaviors of great interest and usefulness in different areas of society. However, they propose limitations referred, for example, to the topological properties that prevent the use of specific techniques, the requirements imposed by dynamic and directed networks, and the large number of nodes involved in the interactions [27], [28], [29].

The problem of community detection is associated with segmentation and identifying regions of the network, which are dense in terms of binding behavior. Because social networks are dynamic, integrating content behavior into the community detection process is vital. Since social networks are inherently dynamic entities in which groups or communities emerge and see members join and leave over time, content analysis can contribute to understanding the laws that govern changes in communities and their evolution within a network.

The most advanced methods in SNA concentrate on the representation of graphical models focused on understanding the structure and evolution of the network. However, according to [15], given the availability of information exposed in the free APIs (Application Programming Interface) of networks such as Facebook and Twitter, SNA has been activated from different angles and perspectives, especially through advances in Semantic Web, Visualization, Data Mining, and Machine Learning technologies.

The review's findings on table 2, which focused on articles that addressed collective phenomena through key terms such as community, propagation, and diffusion, gave rise to a series of emerging categories associated with the purpose of the analysis. In the case of SNA, the studies were classified as role classification, community detection, diffusion, social influence, interactions and reciprocity, dynamic network, topic detection, knowledge base, and private states. Similarly, the categories with the highest number of studies were identified as social influence, diffusion, and community detection. The social influence studies are mainly addressed through metrics associated with centralities, such as betweenness, closeness, indegree, alpha centrality, and eigenvector. The next category, diffusion, coincides with the importance of these metrics. For community detection, clustering algorithms and statistical methods are highlighted.

In summary, the results of this search allow us to conclude that research on the phenomena occurring in the context of digital social networks has been marked by the implementation of methods and techniques that allow taking advantage of the potential of the content available on the web, the increase in online interactions and technological evolution. In exponential growth, the collective behavior underlying social networks is undoubtedly a source of knowledge that requires further research. The application and integration of advanced computational techniques to exploit the potential of structural analysis of social networks may be a way to advance in these purposes.

TABLE 2. Applications and techniques /measures/algorithms used in SNA.

Application	Technique/measure/algorithm	Paper index
(RC) Role Clasification	(CM) Centrality Measures	[30]
(CD) Community Detection	(KM) K-means algorithm	[31]
	(SM) Statistical Methods	[32]; [33]; [34]
	(DLACD) Distributed learning automata based	[35]
	(CL) Clustering	[33]; [36]; [37]
	(WTS) Weak Tie Score	[38]
	(BCL) BIGClam Overlaping Community detection	[39]
	(CNN) Clauset-Newman-Moore	[40]
	(WIC) Within and Inter Community	[41]
	(CM) Centrality Measures	[42]; [43]; [44]; [45]; [46]
	(RW) Random Walk	[47]
(DI) Difussion	(MLR) Multivariate Linear Regression	[48]
	(CM) Centrality Measures	[49]; [50]; [51]; [52]; [53]; [54]; [55]
	(SM) Statistical Methods	[56]; [57]; [58]; [59]; [60]; [61]; [31]; [62]
	(LR) Logistic regression	[63]
	(LP) Label Propagation	[64]
(SI) Social Influence	(CM) Centrality Measures	[65]; [66]; [67]; [68]; [56]; [45]; [36]; [54]
	(PR) PageRank	[38]
	(UNS) User network score	[69]
	(CIR) Community Influence Ranking	[70]
	(HM) Homophily	[71]
	(PAC) Persuasiveness Aware Cascade	[72]
	(SM) Statistical Methods	[73]
	(PRE) Prestige	[43]
(IAR) Interactions and Reciprocity	(CM) Centrality Measures	[74]
	(MD) Modularity	[36]
	(CR) Concentration Reciprocity	[75]
(DN) Dynamic Network	(CM) Centrality Measures	[76]
	(CL) Clustering	[76]
(TD) Topic Detection	(CL) Clustering	[55]; [44]
	(AVPL) Average Path Length	[55]
	(TES) Topical Expertise Score	[69]
	(KNN) K Nearest Neighbors	[45]
	(TAP) Topical Affinity Propagation	[72]
	(RK) Ranking	[61]
	(LDA) Latent Dirichlet Allocation	[31]
	(SVM) Support Vector Machine	[47]
(KB) Knowledge Base	(LR) Lexical resources	[62]
(PS) Private states Analysis	(SC) Sentiment Score	[69]
	(BoW) Bag of Words	[34]
	(SVM) Support Vector Machine	[43]

B. SUBJECTIVITY ANALYSIS

Notions of subjectivity are relevant to many disciplines, including cultural studies, sociology, social theory, anthropology, psychology, linguistics, and computer science. Motivations range from identifying the processes by which subjectivities are produced, exploring subjectivity as a focus of social change, and examining how emergent subjectivities remake our social worlds.

The quality of the subjective links the subjectivity's notion, that is to say, to what belongs to the subject from the opinion or feeling of the one who expresses it, establishing an opposition to the objective, that is, to everything that refers to concrete and factual data. Subjectivity is a notion in linguistic literature in various ways, all of which refer to the systematic forms in which the speaking subject manifests itself in language.

In general terms, the concept refers to the ''presence of the subject'' in language and its use since it is through language that human beings constitute themselves as ''subjects'' and therefore, subjectivity is the capacity of the speaker to assume him/herself as such [77].

In the linguistic field, the most cited authors are [6], [7], [8], and [9], who represent the two most relevant notions with general application to the analysis of linguistic expressions. First, [6] and [7] recognized his notion of subjectivity as a semantic property of linguistic forms (morphemes or clusters of morphemes), that is, words or groups of words that have an inherently more objective meaning because they refer to "things" in the world around us: objects, events, and their properties, while others are inherently more subjective since they refer, for example, to the subject's evaluations of things in the world. Langacker's notion, on the other hand, is associated with the modality of linguistic expressions, i.e., with the "semantic dynamics according to the sender's attitude towards the enunciated thing and his interlocutor: certainty, probability, possibility, belief, obligation, assurance, permission, permission, desire, doubt, prediction, valuation, affectivity." [8], [9]

On the other hand, [78] and [79] defines subjectivity as the aspects of language used to express "private states," that is, mental or emotional states that cannot be directly observed or verified, which include opinions, emotions, appraisals, speculations, and feelings. The aspects of language have been studied mainly by the branch of linguistics known as sociolinguistics, which investigates the reciprocal influence between language and society and the aspects internal and external to the subject that influence communication [25].

The following is a summary of the evolution of the SA, as table 3 shows.

SA is a recently emerging branch of natural language processing; however, it is necessary to approach the phenomenon of subjectivity in a broader scope to understand how its analysis obeys the logic of the contributions of different disciplines whose research dates back to significantly more distant periods. Hence, describing the most critical milestones in the advances that have been made in everything that has to do with the "analysis of private states" is essential.

It should be made clear that interest in the subjective meaning and the affective, poetic-creative, social or interpersonal, and individual dimensions of language is not new to linguistics or computer science. On the contrary, language analysts, including computational linguists, have long recognized the importance of such concepts [80]. Therefore, although it is not in the interest of the present study to delve into the studies on subjectivity carried out in the field of linguistics, it is considered decisive to have as a reference the works of [78], [79], [82], and [81], on subjectivity in language.

Likewise, it is indispensable to consider the basis of the theories of grammar and syntax found in the work of Noam Chomsky [83] and [84], especially for the results recorded during the 50s and 60s around the concept of transformational grammar. These advances led to improvements in the automatic processing of grammar and syntax aspects through compilation and parsing techniques. During the 1970s, significant progress was made in refining these techniques, leading to more efficient algorithms.

TABLE 3. Summary of SA contributions.

Year	Author	Contribution
1976	William James	Theory of emotional categories: words denoting private states.
1977	Izard	Existence of a basic set of emotions.
1979	Carbonell	Interpretation of beliefs by artificial intelligence systems.
1980	James Russel	Emotional dimensions theory: pleasure and arousal.
1980	Lang	SAM standard in three dimensions: evaluation, activation and control.
1980	Plutchik	Set of basic or primitive emotions from which others are derived by conjugation.
1984	Allan Bell	Language style, language, and society.
1985	Watson y Tallegen	Dimensions of positive or negative affect with high or low intensity.
1987	Snow y Anderson	Verbal construction of personal identities.
1988	Ortony et al.	Standard OCC model in speech synthesis with 22 emotional categories.
1990	Wiebe	Identification of subjunctive characters in the narrative.
1992	Hearst	Interpretation of metaphors in texts.
1994	Wiebe Janyce	Impulses of the term Subjectivity analysis.
1994	Sack, Wiebe	Identification of points of interest.
2000	Huettner y Subasic	Affective analysis of the text.
2001	Parrot	Hierarchical structure of primary, secondary, and tertiary emotions.
2001	Dans y Chen; Tong	Prediction of judgments to analyze market behavior.
2002	Turney; Pang et al.	Approaches for the classification of texts according to their polarity.
2003	Dave et al.	Attribute-based processing of search results.
2003	Cowie y Cornelius	Interpret conceptual categories of emotions employing the words that denote them.
2005	Bucholtz y Hall	Identity and interaction: sociocultural, linguistic approach.
2008	Pang y Lee	Tasks: polarity classification, subjectivity detection, sentimental analysis by topics.
2008	Pang y Lee	Sentiment analysis and opinion mining as a subset of subjectivity analysis.
2009	Koller y Friedman	Computational modeling approaches: deep learning, neural networks.
2009	Wilson et al.	Sentiment analysis is not equated with opinion mining but is a subset of subjectivity.
2010	Liu	Opinion search and retrieval- Sentiment analysis of comparative sentences- Opinion spam and opinion utility.
2010	Liu	Document level sentiment classification. Subjective and sentimental classification at the sentence level. Feature-based sentiment analysis.
2011	Balahur	Methods and resources for sentiment analysis in multilingual documents.
2012	Dadvaret et al; Prabhakaran; Hovy	Linguistic variation related to the speaker's social identity.
2012	Montoyo et al.	Review subjectivity and sentiment analysis.
2013	Klaus Krippendorf	Content analysis.
2013	Volkova et al.	Gender differences in subjective language.
2013	Rajagopal et al.	Graph-based approach and semantic similarity.
2014	Xu et al.	Feature-based identification.
2015	Li, Zou y Li	Frequency-based methods with web-based similarity measures.
2016	Ribeiro et al. and Rojas-Barahona	Compare multiple methods of sentiment analysis based on lexicons, Deep Learning, and Machine Learning with different social network databases.
2017	Del Vicario et al. and Catal, Nangir	Development of sentiment classification systems with hybrid methods through topic detection tasks or decision process meta-algorithm to improve the performance of

TABLE 3. ***(Continued.)*** **Summary of SA contributions.**

		traditional classifiers.
2018	Islam et al.	Inclusion of linguistic and time series features to emotion classifiers for performance improvement of Machine Learning algorithms.
2019	Xu et al.	Capture context information with weighted word vectors by combining bidirectional recurrent neural network (BiRNN) models and LSTM units.
2020	Zhao et al.	Integrate semantic social network analysis with polarity layers in the hot search topics for co-occurrence network generation.
2021	Ali et al.	Enrich the ontology of a real-time detection system with sentiment analysis

Source: Own elaboration

TABLE 4. **Applications and techniques /measures/algorithms used in SA.**

Application	Technique/measure/algorithm	Paper index
(TD) Topic Detection	(TFF) Topic Fuzzy Fingerprinting	[90]
	(LDA) Latent Dirichlet Allocation	[91]; [92]; [93]; [94]; [95] [96]
	(BM) Naive Bayes, Bayesian Logistic	[97]; [98]; [99]
	(SVM) Support Vector Machine	[97]
	(RF) Random Forests	[97]
	(LR) Logistic Regression	[97]
	(MP) MultiLayer Perceptron	[97]
	(POS) Parts of Speech and N-gram	[100]; [101]
	(NER) Named Entity Recognition	[102]
	(NED) Named Entity Disambiguation	[102]
	(STT) Sentimental Term Tagger	[103]
	(TS) Time Series	[104]
	(BTM) Biterm Topic Model	[105]
(PS) Private states analysis	(LR) Logistic Regression	[90]
	(ME) Maxima Entropia	[90]; [106]
	(SS) SentiStrength	[107]; [91]; [108]; [109]; [110]
	(BoW) Bag of words	[97]; [111]; [102]; [112]; [113]; [114] [115]; [116]
	(BoE) Bag of embeddings	[97]
	(LBF) Lexical Based Features	[97]; [117]
	(MBP) Markov Based Probabilistic	[118]
	(LGR) Linguistic Rules	[111]
	(MLP) Multi Layer Perceptron	[119]
	(ST) Statistical techniques	[101]; [120]; [121]
	(SVM) Support Vector Machine	[93]; [122]; [123]; [104]
	(HSG) Hoeffding's Stochastic Gradient Descent Tree	[124]
	(TSSE) Tweet Sentiment Score Estimator	[103]
	(BM) Naive Bayes, Bayesian Logistic	[125]; [126]
	(LSA) Latent Semantic Analysis	[127]
	(LIWC) Linguistic Inquiry and Word Count	[128]; [129]; [130]
	(SANT) Sociological Approach to handling Noisy and short Texts	[131]
(SC) Sarcasm	(TPR) True Positive Ratio	[91]
	(SVM) Support Vector Machine	[91]
	(LRS) Linguistic Rules Sarcasm	[132]; [123]
(TC) Text Classification	(SVM) Support Vector Machine	[133]; [134]
	(ENS) Ensemble Classifiers	[134]; [135]
	(LECM) Latent Event Category Model	[136]
	(BM) Naive Bayes, Bayesian Logistic	[136]; [137]
	(RF) Random Forest	[138]
	(LR) Logistic Regression	[139]
(SE) Search	(FL) Fuzzy Logic	[140]
	(TF-IDF) Term Frecuency	[141]; [142]
(KB) Knowledge Base	(ON) Ontologias	[143]; [144]; [94]; [145]; [146]
(SI) Social Influence	(PN) Proximity Networks	[101]
	(PR) Pagerank	[112]
	(ST) Statistical techniques	[99]
	(BM) Naive Bayes, Bayesian Logistic	[110]
(DF) Diffusion	(RM) Rumor Model	[116]
	(BM) Naive Bayes, Bayesian Logistic	[126]
	(ST) Statistical techniques	[121]
	(VAM) Vector Autoregressive Model	[135]
	(MAC) Modified Adsorption with celebrity removal	[109]; [117]

The 1980s were notable for work on language as a cognitive process through research on the cognitive aspects of emotions and the creation of affective lexicons as frames of reference. During the 1990s [85], they delivered to the scientific community the vital WordNet tool for computational linguists and PLN. It is a decade characterized by the emergence of concepts such as semantic similarity, parts of speech, intelligent text-based systems, directionality (for, neutral, or against), subjective word extraction, and statistical methods in natural language parsing.

AS referred to sentiment analysis, opinion mining, and, in general, any attempt to identify and analyze human expressions associated with "private states" (feelings, emotions, opinions, assessments, beliefs, and speculations) [86], [87] through advanced computational techniques. As well as machine translation, information retrieval, response search systems, knowledge extraction, speech recognition, and generation, and summary generation, PLN applications respond to the interests of computational linguists who seek to improve the automatic processing of aspects of grammar and syntax in order to analyze the affective, social and individual dimensions of language more effectively. Reference [88] propose that sentiment analysis, opinion mining, and AS are interrelated research areas that use several techniques borrowed from machine learning, PLN, semantic analysis, Information Retrieval (IR), and structured and unstructured Data Mining.

PLN has been conceived as "a part of Artificial Intelligence that researches and formulates computationally effective mechanisms that facilitate the human/machine interrelationship and allow a much more fluid and less rigid communication than formal languages" [89]. Its ultimate goal is to build computational systems that can understand and generate natural language in the same way humans do, through an immediate objective of building systems that can process text and speech more efficiently. This processing, in turn, uses advanced statistical, machine learning, and text mining techniques.

The following contributions are the most significant advances in computational methods that have arisen from the interest in exploring private states such as opinions, emotions, feelings, valuations, beliefs, and speculations in natural language. The second search equation focused on identifying SA studies that involved the collective approach using keywords such as aggregation, social influence, and propagation. Extracting information from these studies gave rise to topic detection, knowledge base, private states analysis, sarcasm, text classification, and search like table 4 shows.

The results show that detection, extraction, and classification of emotions, feelings, and opinions are the main applications coded as private states analysis. This category records 33 techniques, algorithms, or methods for performing the analysis. Among the most exciting findings is that the most common technique is BoW (Bag of Words), followed by SVM and SENTIStrenght, a tool that generically performs sentiment calculation. Studies use them whose interest is not to deepen the sentiment analysis but that it becomes an input of a study at another level, such as detecting sarcasm and rumors.

The application occupies the second place regarding the number of studies registered in detecting topics. There are 19 techniques or methods where LDA and Bayesian Models stand out; this application was also evident in the SNA studies. Next is the text classification with six techniques where SVM algorithms and Bayesian models are present again. Finally, the literature reviewed shows a tendency of the scientific community to carry out studies at the linguistic level aimed at improving SA, such as sarcasm detection.

It is possible to conclude that the interest of the scientific community in studying the expression of private states in different textual sources has led to the emergence of different methods, techniques, and approaches that mainly are into Machine Learning methodologies, Lexicons, or a combination of the two previous ones called hybrid approaches. In turn, the literature reviews allow differentiating the contributions between ontological and non-ontological approaches to studies based on textual corpus processing.

C. SOCIAL NETWORK ANALYSIS AND SUBJECTIVITY ANALYSIS

Given that social networks are currently the leading platform for user communication on the Web, and given characteristics such as the ubiquity of these networks, the enormous amount of data available, and the diversity of topics discussed, researchers have come to develop advanced techniques to identify the expression of private states. Through text mining methods, new insights emerge to extract relevant information by analyzing and identifying large amounts of unstructured data [147]. However, it is necessary to consider that, although there are practical algorithms to detect this type of information, there is a significant value in the information that can be provided by analyzing the relationships established by users in these networks. This dimension of information can help to detect or infer opinions on specific topics from the orientation of the opinions of users interacting with each other [148].

One of the phenomena of study that has motivated sociologists, anthropologists, psychologists, and in general, researchers interested in human interactions and the scope of communication is the analysis of collective subjectivity. Reference [2] defines it as the set of common denominators of the TAF modes of the members of a social collective, including not only discourses and social representations but also their emotions, experiences, and actions.

For [2], interactions should not be confined to processes developed by individual actors since collectives also interact. However, this does not mean that the individual's actions should not be taken as the model from which the movement of collective subjectivities can be understood. The author suggests that the concepts underlying the notion of collective subjectivity are those of Marx's social classes and Parsons' collective actors, based on the model of individual actors who behave consciously and intentionally and influences collective discourses and representations.

TABLE 5. Studies integrates SNA and SA.

		SA studies associates SNA					
		(TD) Topic Detection	(PS) Private State Analysis	(SC) Sarcasm	(KB) Knowledge Base	(TC) Text Classification	(SE) Search
SNA studies associates SA	(RC) Role Classification						
	(CD) Community Detection	[44]; [45]; [47]; [154]	[34]; [43]				
	(DI) Difusion	[55]; [61]; [31]	[116]; [126]; [121]; [109]; [117]		[62]	[135]	
	(SI) Social Influence	[101]; [99]; [69]; [72]	[112]; [110]; [69]				
	(IAR) Interactions and Reciprocity						
	(DN) Dinamyc Network						

By analyzing this intersection, it is possible to determine that researchers from both fields have made progress in studying the collective phenomenon. For SNA analysts, the community is the actors that compose it in its ties (strong or weak), where the definition of roles plays an important role insofar as they speak of a social structure in the framework of which the community makes sense. On the other hand, SA researchers have concentrated on trying to group opinions, discourses, or texts according to a specific domain. In this sense, the actor's relationship with the topics detected in that domain is understood, creating communities of opinions by topic. The integration for the SNA field does not prioritize how the text analysis is done, but the link that the topic or polarity detected can generate between actors and the new relationships such links can create. SA are tools that automatically generate such information without considering the process, techniques, that they include. On the other hand, subjectivity analysts who contribute to the integration of techniques question the fact that in the aggregation of opinions, they all share the same weighting and use the SNA to add a characteristic to the collective polarity analysis, determined by the structural position of each actor in the network. They use the techniques used in SNA to identify influencers and start using propagation analysis techniques.

Table 5 shows the studies that record the integration of techniques. The intensity of the color in the table is associated with the research interest in each purpose, which means that the most significant number of studies that integrate techniques focus on the diffusion of private states. Then, there are researches focused on analyzing social influence on the topics detected on the network. Next in relevance are studies that determine communities around topics and, finally, those that create linguistic resources through analyzing how messages are disseminated within the network.

III. METHODOLOGY

The previous sections analyzed the nature, evolution, methods, techniques, and applications of two disciplinary

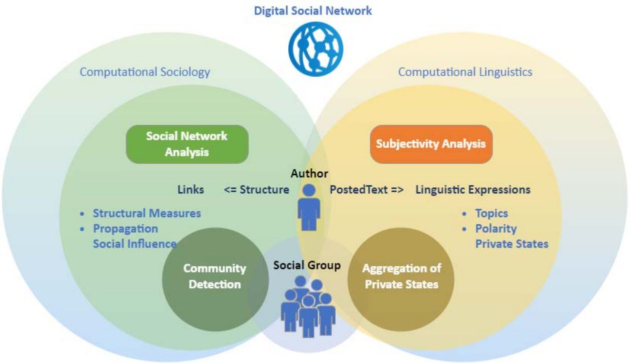

FIGURE 1. Overview of hybrid System COSSOL.

constructs dedicated to studying communication phenomena and human interaction in the digital scenario. The first one identifies the structural patterns in the relationships of actors in a social network (SNA) through network analysis (graphs) framed within the contributions of Computational Sociology. The second is within the framework of Computational Linguistics, whose object of study is the linguistic expressions of private states present in communicative interactions (CLI) using PLN techniques (text classifiers).

A. OVERVIEW OF THE PROPOSAL

To carry out the analysis of collective subjectivity, COSSOL takes elements from these two constructs to represent a hybrid system, as illustrated in Fig. 1, where one can see how computational sociology metrics are taken to evaluate the structure, density, and centrality of the network. In contrast, from computational linguistics, text classifiers are taken to characterize, identify and describe the private states of the communities shared in the social network.

COSSOL integrates the two constructs based on the principle of "*onion layers*," which examines the interaction levels of communities in social networks from a granular analysis to analyze collective subjectivity. The more internal the onion ring is (higher granularity), the more stable the collective subjectivity of the community or communities represented by the onion ring tends to be. Additionally, the interactions around specific issues and how these last over time determines a community; some communities remain for long periods, and others, being ephemeral, significantly impact their life cycle.

Since to identify communities, it is necessary to analyze the interactions of their members, COSSOL takes the concepts of actors, mentions, hashtags, comments (opinions), and sharing to describe the structure of a social network and the different scenarios that can occur in it. Consequently, a model is proposed to test the collective subjectivity from the communities' granularity and a thrust from the linguistic construct for the identification, interaction analysis, and characterization of the communities. All this information will be a fundamental input for the present proposal, and its detail is shown below.

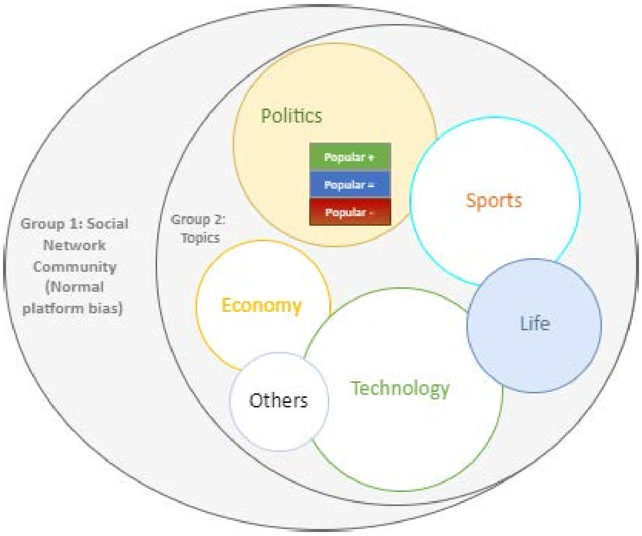

FIGURE 2. Community analysis like onion layer approach.

B. COLLECTIVE SUBJECTIVITY ANALYSIS OF COMMUNITIES BY ONION LAYERS

There is a large community made up of all the users who participate in a social network. Around it, different communities are created spontaneously on specific topics that last more or less in time. For example, the behavior analysis of publications in Colombia where a community related to political issues is identified, which in turn contains different sub-communities that support or oppose the current government; as a consequence, the political community is more stable than its subcommunities since they are dependent on the current characters or events in the region. In this way, and as there are other larger or more granular communities, the onion layers are intended to represent the existing relationship of contention between them.

In this sense, Fig. 2 represents the existence of different levels of interaction of the communities from the perspective of onion layers, starting with a general level of analysis toward a more specific one. A set of rings or circles represents each level of analysis that, in turn, contains others; that is, a smaller circle represents a more specific level of analysis with more defined communities or greater granularity generating a disaggregation of the network. The outer layer or circle corresponds to the conformation of a community with a broad spectrum representative of a social network; users interact in different scenarios generated on digital platforms such as Twitter, Facebook, and Instagram to propitiate different debates or comments without any particularity.

On the other hand, the inner circles represent those subcommunities where their actors are around a topic of common interest. For example, it can observe the different subcommunities generated during the interaction of national users in communities of other countries, created within the country or subcommunities of multiple countries. Thus, as the layer or circle is reduced (greater granularity), greater accuracy is

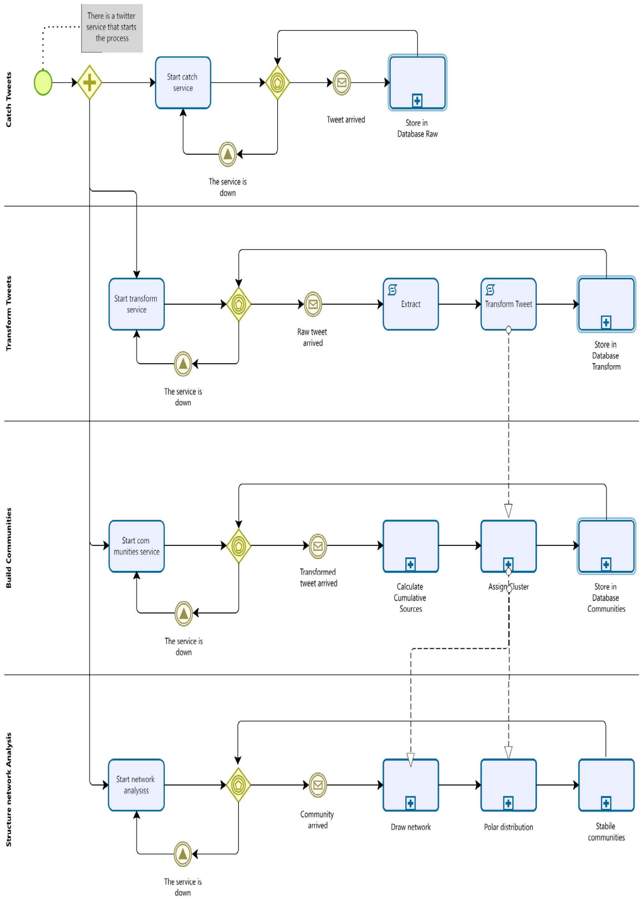

FIGURE 3. Computational model for collective subjectivity analysis on twitter.

obtained in identifying subjective communities, their respective interactions, and the characterization of each one of them.

C. COSSOL HYBRID APPROACH SYSTEM IMPLEMENTATION

This case study describes levels of designed processes in a system that allows the analysis of collective subjectivity in the social network Twitter. In order to perform such experimentation, the levels are divided into hierarchies based on the process size. The first level is the defined components represented in Fig. 3; it shows the workflow between the different computational components developed. The second level of processes comprises the activities that outline the main processes per component; the third level refers to the instructions to perform minor system processes to connect and relate the activities. Finally, the fourth level of processes contains the components that involve more complex computational development activities, called sub-processes.

The following sections describe the process flow of the COSSOL system for each component of Fig. 3 and the detail of the subprocesses within the computational model.

D. CATCH TWEETS

The process starts by connecting to developers' Twitter public API service, obtaining the specific tokens to authenticate and access the API. Subsequently, there is an activity that creates a service to collect tweets[1] in streaming; that is, it allows capturing in real time the tweets published. The process continues by checking the status of the catch service in order to verify that it is working normally and constantly, thus ensuring a recurring process to avoid erroneous analysis due to a lack of artifacts or metadata of tweets in lost time lapses. Once the tweet is collected, the component registers it by initiating the sub-process called "Storage in raw databases".

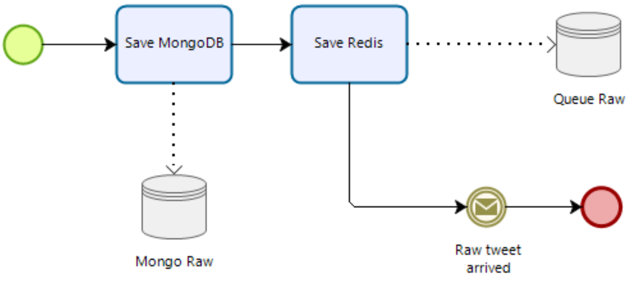

FIGURE 4. Subprocess for storing raw tweets.

1) STORAGE IN RAW DATABASES

The component's subprocess shows in Fig. 4, describing the generated database. This subprocess starts with the activity of saving the tweet that the component registered as collected into a NoSQL database (MongoDB). Once saved, a database engine (Redis) is integrated, which generates a queue structure of the managed tweets with the logic of performing the subprocess activities for each tweet; once it reaches the search engine, the subprocess activities can be started with another tweet, repeating this cycle until all tweets are managed. Finally, these tweets are arranged in the database engine, having them available to perform the search contemplated in other subsequent components.

E. TRANSFORM TWEETS

The purpose of this section is to describe the main component in charge of performing computational linguistic processes belonging to the AS construct. Like the previous component, the new service for transforming raw tweets ensures that all these tweets are available for the activities developed in this component. In this order of ideas, the service registers the arrival of the raw tweet to start the extraction activity, which comprises obtaining the characteristic artifacts of what is being done (I) such as the post raw text, the "likes", the number of times it was shared, the mentions present, as well as the hashtags, emojis, and user IDs.

Thus, the "tweet transformation" activity performs a set of profiling to the variables in the computational linguistics thrust; consequently, it becomes the main activity to elaborate the AS construct's techniques. The following sections provide the results and details of the sociographic, demographic and

[1] The Twitter public API is limited to downloading the last 3000 tweets from each account.

TABLE 6. **Distribution of articles by topic.**

Categories	Articles number
1- Sports	63,000
2- Culture	36,000
3- Economy	58,000
4- Politics	43,000
5- Technology	59,000
6- Life and leisure	41,000
Total	300,000

TABLE 7. **Results obtained with the Gold Standard comparison.**

	Precision	recall	f1-score	support
1- Sports	0.60	0.96	0.74	3,159
2- Culture	0.98	0.93	0.96	3,985
3- Economy	0.95	0.76	0.85	2,097
4- Politics	0.97	0.78	0.87	2,069
5- Technology	0.99	0.90	0.95	2,860
6- Life and leisure	0.95	0.78	0.85	3,365
Accuracy			0.87	17,535
Macro avg	0.91	0.85	0.87	17,535
Weighted avg	0.90	0.87	0.87	17,535

TABLE 8. **Description of the parameters used in the K-neighbor classifier.**

K-neighbors classifier	
Parameter	Value
Neighbors number (n_neighbors)	30
Weighted function used in forecasting (weights)	Uniform
Algorithm used to compute nearest neighbors (algorithm)	auto
Leaf size passed to KDtree (leaf_size)	30
Power parameter for Minkowski metric (p)	2
Distance metric to be used in the tree (metric) (metric)	Minkowski
Additional key arguments for the previous function (metric_params)	None
Number of parallel works to execute the search for neighbors (n_jobs)	None

*Note: The parameter label in parentheses is the code needed to define these in Python [153]

psychographic variables, saving these results in a database explained in the subprocess called ‘‘Storage in transformed databases’’.

1) PROFILING OF SOCIOGRAPHIC VARIABLES

The topic classifier contains a sociographic variable to give more granularity to the communities built in the next component. The following sections present the results of such a classifier to the COSSOL system under an F1-score performance metric applied to Twitter by going through the elaboration flow.

1) Dataset construction:

The detection of topics on the Twitter social network used a built dataset with articles published in the local press last year, which in Colombia was the newspaper **El Tiempo**. Within this consulted information source, the articles were classified into six different sections: *life and leisure*, *sports*, *culture*, *economics*, *politics*, and *technology*, resulting in a database composed of 300,000 newspaper articles. These articles were the input for the training of a topic classifier associating the classes of the classifier to the newspaper categories, where table 6 presents the distribution of articles collected by each category.

Now, from this resulting database, a sample of 70% was taken to train the classification algorithm on it. The first training sample started by developing a preprocessing of the raw tweets, which included lemmatizing and tokenizing the use of lexical and syntactic features within a vector. After this training stage with a 30 % sample, the text above processing and use of the Kneighbors classifier, as table 7 shows. The f1-score values indicate an overall performance of 87% for all topics, ranging between topics from 74% to 96%. The topics of *culture* y *technology* presented the best results with 96% and 95%, respectively. On the other hand, the topic of *sports* has the lowest value of 74%, and the remaining topics have similar performances of around 85%.

Table 8 presents the parameters used following the possible values that the scikit-learn library [152] offers for the k-neighbors classifier. The first parameter sets 30 neighbors for its queries; the second uses an equal (uniform) weighting on all points of each neighbor; the third adds a *auto* algorithm trying to decide the most appropriate algorithm based on the values supplied to the fitting method. The fourth sets an optimal value of 30 (leaf size), which affects the speed of construction and querying, as well as the memory needed to store the tree; the fifth and sixth set the power of the Euclidean distance parameter to realize the k neighbors; finally, the last parameters point out the non-existence of additional arguments on the previous metric and the number of parallel jobs to use.

2) PROFILING OF PSYCHOGRAPHIC VARIABLES

Psychographic variable profiling is an activity belonging to the sentiment analysis field that the AS construct offers. This activity extracts common TAF modes from each replica comment to group users’ private states aligned to a topic of interest. The following sections explain in detail the lexicons used to measure the TAF.

1) Polarity lexicon and performance

This section is devoted to analyzing the sentiment classification process under a methodological construction process explained in the article ‘‘*CSL: A combined Spanish lexicon - resource for polarity classification and sentiment analysis*’’ [150]; in it, the authors searched for each lemma found within the sentiment lexicon with a sentence or phrase. Subsequently, they obtained the weighted average of the rating of each of the words determining the polarity of the phrase or sentence. In this article, the polarity metric used three lexicons constructed by the authors for the Spanish language, choosing the one with the best performance, which consists of a combination of the CLS CAOBA and the Spanish lexicon CL. The CSL3 lexicon assembly had the best-tested performance of 62.05%. The table 9 presents the performance of the assemblies performed against different lexicon proposals to be classified in the polarity scenario.

3) PROFILING OF DEMOGRAPHIC VARIABLES

The profiling of the demographic variable arises from the description of the gender and age classifiers supplied to the COSSOL system. It is pertinent to mention that the

TABLE 9. Perfomances of polarity lexicons.

Lexicon	# Positive words	# Negative words	Accuracy
iSOL:	2,509	5,624	55.26
Elh Polar:	1,379	2,502	59.95
SEL:	631	931	54.33
SLS:	477	870	50.83
ML-SentiCon:	4,453	4,482	46.99
MS:	1,553	2,720	53.99
CLS_1:	1,901	1,910	60.66
CLS_2:	1,970	1,945	60.73
CLS_3:	11,634	3,305	**62.38**

TABLE 10. Identified age distributions by [151].

Age	Users
13-17	211
18-24	1067
25-34	566
35-49	267
50-64	109
65-+66	45
Total	2,265

TABLE 11. English and spanish gender classification by [149].

	Precision		Recall		F1-Score		Support	
Class	en	es	en	es	en	es	en	es
0	0.79	0.76	0.84	0.72	0.81	0.74	310	540
1	0.83	0.73	0.77	0.78	0.80	0.76	310	540
Micro avg	0.81	0.75	0.81	0.75	0.81	0.75	620	1080
Macro avg	0.81	0.75	0.81	0.75	0.81	**0.75**	620	1080

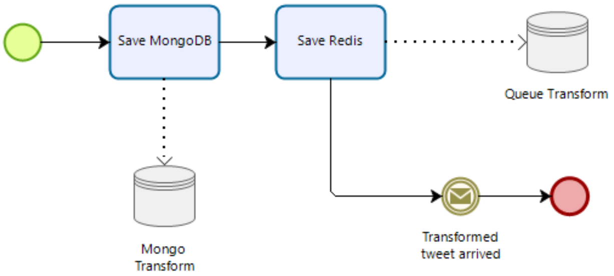

FIGURE 5. Subprocess for storing transformed tweets.

consumption of content on Twitter and, therefore, the discussion of this content changes as the levels of promotion of events or social happenings are included, giving an added value to studying the change generated in the different age groups or genders present.

1) Age classifier and performance
The age classifier uses the methodological construction process explained in the article "*Age Classification from Spanish Tweets - The Variable Age Analyzed by using Linear Classifiers*" [151]. The classifier uses a 45-word lexicon to profile Spanish expressions related to the concept "birthday" to automatically assign a label within six age ranges for Twitter users. It processed 50,819 accounts related to universities and 734,037 accounts related to celebrities, where 1,541 obtained an accurate automatic age label, and these increased to 2,265 users thanks to the use of a Gold Standard constructed. The classifier was the result of an effective validation of 120 models where seven models obtained an accuracy performance of the 66% to 69%. Next, an additional layer was applied to extract information from the users, bringing the accuracy of the best models to 72.96%.
Table 10 shows the distribution of correct and automatically assigned labels by the algorithm (SGDC hinge_none_optimal). The second and third age ranges obtained the highest correct label assignment; in contrast, the two higher age ranges obtained the lowest number of correctly assigned labels.
2) Gender classifier and performance

The age classifier uses the methodological construction process explained in the paper "*Bots and Gender Profiling on Twitter using Sociolinguistic Features Notebook for PAN at CLEF 2019*" [149]. It consists of extracting a set of characteristics from profiles and texts of posts made by Twitter users and developing different models of Machine Learning to have a model with the best performance for assigning a correct gender label. The text features aimed to explore the author's diversity by analyzing the features extracted in the user's profile features. On the other hand, the user profile features obtained the traditional values of word counts, hashtags, mentions, URLs, and emojis per tweet. The authors tested the models for the Spanish and English languages, where the Random Forest obtained the best performance with an 81% macro f1-score for the English language and 75% of macro F1-score coming from a Logistic Regression for the Spanish language. The table 11 presents the results obtained in more detail.

4) STORAGE IN TRANSFORMED DATABASES

The subprocess of the tweet transformation component is presented in Fig. 5, describing the database generated to receive the tweets transformed from the text classifiers mentioned in the variables of the previous sections. The subprocess starts with the activity of saving the transformed tweet into a NoSQL database (MongoDB); then, in the following activity, it integrates a database engine (Redis), where, similarly to the subprocess of the previous component, Redis performs the queue structure of the transformed tweets to have them available when the following community-building component searches.

F. BUILD COMMUNITIES

Building communities is the component aimed at identifying the target population to be analyzed: the communities. These are groupings of tweets transformed under the discussion that gave rise to the group. In the case of experimentation within the Twitter social network, the component designs three communities as representatives of onion layers following the logic of Fig. 2. The first community is the broad spectrum of discussion originating from the Twitter social

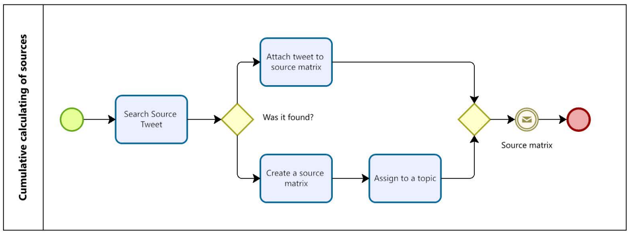

FIGURE 6. Cumulative source calculation.

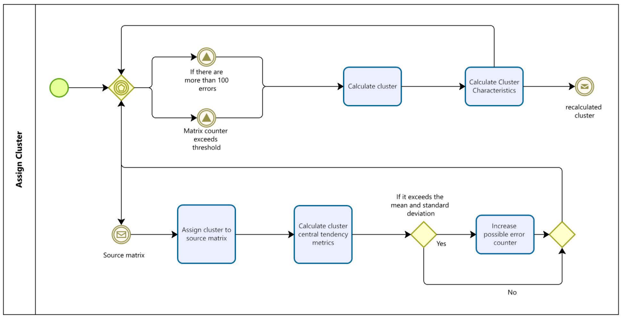
Assign Cluster
If there are more than 100 errors
Matrix counter exceeds threshold
Calculate cluster
Calculate Cluster Characteristics
recalculated cluster
Source matrix
Assign cluster to source matrix
Calculate cluster central tendency metrics
If it exceeds the mean and standard deviation
Yes
Increase possible error counter
No

FIGURE 7. Assigning clusters.

network; the second community has six groups resulting from a level of granularity granted by the sociographic variable. Finally, the third community associates popularity clusters with each group of the second community, exposing the variance compute of the collective subjectivity analysis as its temporal analysis feature.

The process starts with an instruction that constantly checks the status of the communities' service to verify that it receives all the transformed tweets from the previous component and is not inactive, avoiding incomplete processes resulting in communities' absence. Once the transformed tweet is received, the component starts executing the activities through the sub-process of cumulative calculation by origins, cluster assignment, and storage in the communities' database. The following sections describe the activities with subprocesses and the computational implications of each activity in the study of collective subjectivity for the COSSOL system.

1) CUMULATIVE CALCULATION OF TWEETS BY ORIGIN

This subprocess is mainly responsible for accumulating the tweets replicas of a source publication and generating the communities from this accumulation, as shown in Fig. 6. The subprocess starts by searching for those transformed tweets that are source posts in the social network; that is, they are posts made by a user about a particular event or situation. As a result of the search, the transformed tweets are handled in two different ways by incurring different instructions. First, the transformed tweets identified as source posts are compiled into distinct vectors to generate a matrix; the remaining transformed tweets, on the other hand, are understood to be replica comments by adding them to a source post since they are generators of discussions about the content of those in the first group.

This subprocess becomes an iterative procedure creating a matrix with a condition of the temporality of the replica comment on the original publication so that the latter accumulates the number of replica comments following the chronological order of publication, also known within the Twitter social network as a conversation thread. At this point, the subprocess infers the creation of the three communities mentioned in the component. The first community is all tweets representing the broad spectrum of discussions in the social network; the second community was created when the subprocess involves the application of the sociographic variable, where the origins matrix assigns to one of the six topics. Finally, the third community is the cluster assignment by popularity, which the next subprocess describes.

The cumulative calculation of replica comments by origin (referred to as threads hereafter) denotes collective TAF representing Twitter trends of certain events or situations, which implies the graphical visualization resulting from this subprocess. They are ascending lines with a growth factor dependent on the number of added replica comments, showing a higher growth factor presenting the largest number of replica comments accumulated by a source. Otherwise, there are rising lines where the life cycle of the source post is not long-lasting since its growth factor implied few accumulated replica comments. The conversation threads were configured according to time, where they are added immediately after the original publication is generated until 3900 minutes have elapsed.

2) ASSIGNING CLUSTERS BY POPULARITY

Popularity is a criterion of thread performance for a second community group that seeks to represent the activity levels over time to be grouped when the levels are similar or within an established range. The COSSOL system application divides into three clusters, each topic subcommunity implementing the popularity model. For example, the first cluster integrates those publications with high popularity, that is, comments that had a high level of participation of other users in the discussion of that publication.

The subprocess starts by receiving the source matrix resulting from the thread accumulation process, as shown in the workflow represented in Fig. 7. Then, the k-means technique computes the three clusters measuring the performance ranges for each cluster over the stored thread pool; then, a calculation of central tendency measures is performed to have some criteria to decide the need to perform a recalculation of the clusters. Each cluster recalculation is the evolution of the discussions due to the presence of new events or occurrences in the social network.

The subprocess starts with calculating each thread's mean and standard deviation for the cluster under study to establish a decision criterion. The criterion is designed as follows: if the calculation values for any thread do not exceed the range of the mean plus two times the standard deviation of the

TABLE 12. Number of cluster recalculations by topic.

Topic	Recalculation quantity
Culture	137
Politics	135
Sports	136
Life and leisure	134
Technology	134
Economy	132

cluster, then there is no reason to believe that there is a change in the TAF in the community; otherwise, if the calculation values of any thread exceed this range, then this thread will be marked as an error within the cluster and the subprocess flow. Now, there is an activity of counting these errors daily to register an instruction that notifies the exceeding of the defined limit. If there are more than ten accumulated errors within this time range, then the thread automatically performs the popularity recalculation using the k-means technique. The errors denote changes in the clustering characteristic; the discussions posted in the social network present new interaction phenomena generating new popularity levels. In addition, the recalculation of the clusters initiates another instruction in parallel, notifying the completion of the recalculation to keep track of the number of times the recalculation.

3) CLUSTER CALCULATION BY TOPIC

The COSSOL system presents a relationship between the components of "*tweet transformation*" and "*community creation*" so that the subprocess of "*cluster assignment*" is evaluated among the different groups belonging to the sociographic classifier. At the time of cluster calculation by topic, a minimum of one hundred threads be assigned to assign at a cluster that measures the performance of such threads. In this way, it is possible to show a granularity of the users' conversation to replicate the activities and subprocesses of the *Network Structural Analysis* component, recognizing with greater clarity the collective subjectivity. From the above, the need to establish key time windows or define recalculation number of the clusters that are tied to the dynamics of change, where the system presents the results of the application of the component "Network Structural Analysis" in the established time window or a defined number of recalculation of the clusters.

A metric count the number of times the popularity clusters' recalculations to mitigate the effect of TAF attitudinal changes. The table 12 presents such a count for the six clusters of the second community over a period, where this table becomes a template for integrating the results of the COSSOL system.

4) STORAGE IN COMMUNITIES DATABASE

The subprocess storing the communities created is in Fig. 8 describing the database generated for its subsequent management. The subprocess starts with the activity of saving the communities generated by the component into a NoSQL database (MongoDB), which is a resource connected to the next activity that integrates a text search engine (Elasticsearch). First, Elaticsearch stores the communities with an ID with all the artifacts of the "catch" activity, the metadata resulting from the classifiers of the "tweet transformation" activity, and the number of times the community was reassigned to the different clusters of the "assigning clusters" activity. Then, the communities are visualized to recreate the timeline that the "collection of tweets'" component managed. The process of storing the communities creates a time series with yearly, monthly, weekly, daily, hourly, minute, and secondly disaggregation on the created communities.

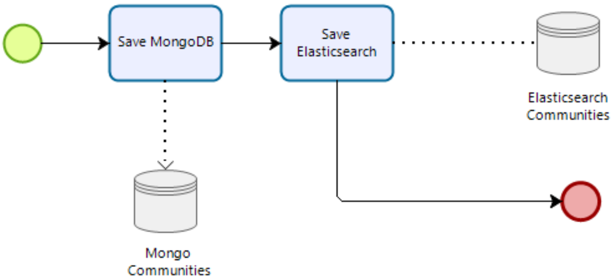

FIGURE 8. Subprocess for storing communities.

G. STRUCTURAL ANALYSIS OF THE NETWORK

This section describes the component that defines the activities of network creation, polar distribution analysis, and stability inference in communities with its subprocesses responsible for the results. The network's structural analysis comprises the COSSOL system's last step to integrate the classifiers developed in the "transformation of tweets" component associated with the Computational Linguistics axis and the communities identified from the "building communities" component associated with the Computational Sociology thrust. The following sections show the details of the subprocesses of the mentioned activities and the results generated from the relationships of this component with others addressed in previous sections.

1) WEAVING NETWORKS BY THEMES

Fig. 9 represents the subprocess showing the workflow to weave the networks to the different communities that the COSSOL system approach contemplates through the onion rings. Thus, the subprocess becomes an iterative task applied to the entire Twitter spectrum, the community with six different clusters, and finally to the different popularity clusters in each larger community group.

The subprocess starts with the notification of a recalculation performed for the popularity clusters; then, the notification connects to an event-driven gating instruction, where it starts with the arrival of the notification from the identified community. In the next step, a network graph is woven for the identified community by calculating the centrality by degree

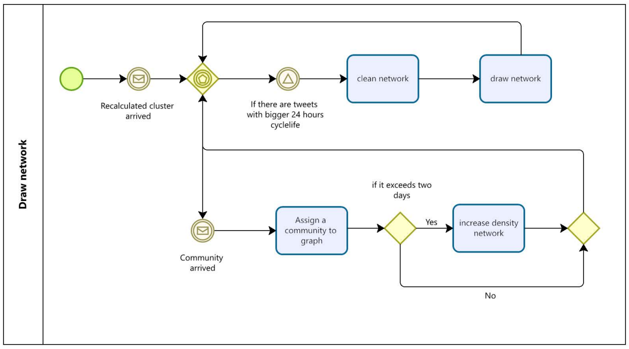
Draw network
Recalculated cluster arrived
If there are tweets with bigger 24 hours cyclelife
clean network
draw network
if it exceeds two days
Community arrived
Assign a community to graph
Yes
increase density network
No

FIGURE 9. Weaving networks using graphs.

TABLE 13. Descriptive statistics for degree centrality.

Topic	Varianaza degree centrality	Desviación estándar degree centrality
Culture	1.66%	227.99%
Politics	1.12%	150.83%
Sports	0.46%	61.90%
Vida	0.19%	24.90%
Technology	0.40%	53.75%
Economy	0.28%	37.40%

and the network density; then, the subprocess verifies if the network representing the community activity had exceeded two days reflected in denser networks due to higher activity when it exceeded the time. Subsequently, an event notification is designed to establish the existence of transformed tweets within the community with a life cycle longer than 24 hours, where a new network graph will wave.

1) Centrality metrics for the network
   The centrality metric by the network's degree aims to indicate the most and least essential terms in the conversation threads by the community with their frequency value; that is, to represent the number of links with other nodes. The table 13 presents the descriptive statistics for degree centrality by topics to establish a template for further analysis to study where the highest variance and the standard deviation are with the recalculation of the popularity clusters.
2) Density metrics for the network
   The network density metric establishes a limit on the number of nodes and connections present in the graph for the different communities, having a dynamic factor in exemplifying the TAF changes that contemplate the different communities to be analyzed. The COSSOL system maintains conversation threads with a life cycle of 24 hours, and after this time, the network is rewoven; therefore, the network avoids performance problems by generating very dense graphs. The table 14 presents some results of densities for the networks of the different topics establishing a template to analyze in the following chapter; the objective is to locate similarities and differences between the topics with the SNA construct metrics.

TABLE 14. Descriptive statistics for network density.

Topic	Varianza	Desviación estándar
Culture	0.000134%	0.02%
Politics	0.000206%	0.03%
Sports	0.000183%	0.02%
Vida	0.000222%	0.03%
Technology	0.000107%	0.01%
Economy	0.000109%	0.01%

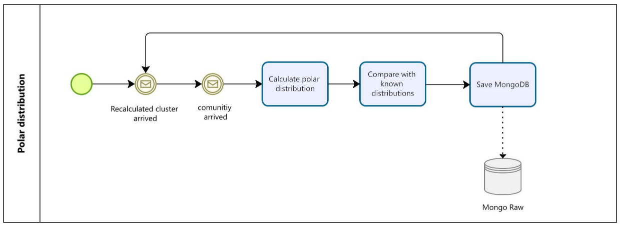
Polar distribution
Recalculated cluster arrived
comunitiy arrived
Calculate polar distribution
Compare with known distributions
Save MongoDB
Mongo Raw

FIGURE 10. Polar distribution calculation.

2) POLAR DISTRIBUTION ANALYSIS FOR COMMUNITIES

This subprocess analyzes the modes of feeling to identify the orientations of the discussions in each community. The polarity distribution makes it possible to analyze social phenomena of dissent or consensus manifested in the orientations of the ways of feeling that the debates generate within each of these communities; hence, the bimodal distribution shapes dissent in the communities, characterized by polarities at the extremes of the distributions. Otherwise, the distributions that shape consensus to some degree are the Weibull, Generalized Extreme Value, Uniform, Beta, T-student, Normal, or Natural Logarithm Range distributions, where their polarities are centered on the mean of the distribution. In conclusion, a distribution different from the bimodal distribution may denote settled discussions that tend to agreements within the communities.

The subprocess (see Fig. 10) starts with the arrival of the recalculated cluster notification causing the subprocess to identify a new subcommunity for its polar distribution study, in this case, when the third community is analyzed. Then the polar distribution is performed with the values present in the sentiment analysis metric (positive, neutral, and negative) to locate the place on the curve where most of the TAF is for the analyzed community. Further, the subprocess adds a known series of distributions to the observed distribution to model some degree of consensus or dissent; as a result, the comparison between these two distributions determines a clear pattern toward a dissent orientation if the discussions are not convergent. In contrast, a comparative distribution behavior toward an agreement orientation represents discussions toward the data mean or to an extreme of the TAF. Finally, the database stores these plots emphasizing the need to run a queueing process due to the requirement to save a new set of plots each time the clusters are recalculated.

3) STABILITY LEVELS IN THE COMMUNITIES

This activity describes the development of a subprocess to defining the stability levels by applying a collective subjectivity analysis for each onion layer. The stability in the communities presents a higher level as the community becomes more granular, allowing the COSSOL system to test the hypothesis of the present thesis. The following sections describe the cointegration subprocess in charge of performing the calculation, which concludes stability levels.

4) COINTEGRATION OF NETWORKS, CLUSTERS, AND SENTIMENTS

The subprocess notifies a network recalculation with structure and subjectivity metrics; therefore, recalculations are the object where these metrics are stored, and their accumulation makes up the community to be analyzed. Each time system reports a recalculation, it records the time at which it was created. The accumulation of records for the different recalculations defines per se the time horizon to which the application of the cointegration test uses the different structure and subjectivity metrics; that is, each community will have a different time horizon and, in turn, a unique time delta (unit of the time change). The typical behavior in any community can be plotted on an oscillating line along the time horizon showing its variance as the time delta, the delta (days) shows the interaction between the metrics. In the case of the COSSOL system, the **polarity** and **network density** are the metrics of subjectivity and network structure, respectively. They become the time series that are the instruments to apply the statistical test of cointegration.

Subsequently, the process reviews the time horizon defined by the network recalculations to avoid time gaps or dates with missing data in the network density and polarity series since if a gap is identified, it will represent an impediment to performing the test since the series must be continuous. Otherwise, the time gap problem is solved by resampling the original base by lowering the dimensional level of the time by interpolating the missing dates, obtaining the data, and the continuity in the series.

It continues with the series of steps of the Engle-Granger methodology to perform the cointegration test, which consists of two main steps: the first one checks the existence of unit roots in the series, and the second one performs a test to determine the existence of cointegration with the series that do not have a unit root, demonstrating the non-existence of spurious results. As the individual behaviors of the variables denote more pronounced trends over time, the variable would show signs of unit roots (non-stationary series). Thus, the first Engle-Granger step is to apply an Augmented Dickey-Fuller (ADF), Phillips- Perron (PP), and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test to each variable in levels. Each test will conclude the presence or absence of a unit root; however, the final determination of the unit root on each series will be made with the consensus of most of the conclusions coming from the above tests. For example, if two tests conclude unit root, it will be finally concluded that the series has a unit root; on the other hand, if only one of the three tests concludes unit root, it will be finally determined that there is no unit root.

The following paragraphs describe the hypothesis tests for each statistical test to provide insight into the results that will be shown later. First, the equation (1) represents a regression which model of a time series following the structure of an ARMA for an ADF test formulation:

$$\Delta Y_t = \pi Y_{t-1} + \beta' D_t + \sum_{j=1}^{\rho} \psi_j \Delta Y_{t-j} + \varepsilon_t \tag{1}$$

where:

Y_t is the analyzed series (polarity or network density).
Δ is the first difference operator of the series.
D_t is a vector of deterministic terms (constant, trend).
ρ is the differentiable lags number of the series.
ε_t is the error of the series.

The following are the hypotheses of this test:

$H_0 : \pi = 0$; There is a unit root.
$H_a : \pi < 0$; There is no unit root.

The test statistic values (ADF) are calculated as follows for the equition (2):

$$ADF = \frac{\hat{\pi}}{SE(\hat{\pi})} \tag{2}$$

The ADF statistic compares the critical value with the selected significance level. If the test statistic value is greater than the critical value in absolute value, the null hypothesis can be rejected, concluding that a unit root is not present.

Similarly, the regression presents the Phillips-Perron (PP) test represented in the equation (3):

$$\Delta Y_t = \pi Y_{t-1} + \beta' D_t + u_t \tag{3}$$

where:

u_t is the regression error for the series.

Consequently, the hypotheses of the PP test are as follows:

$H_0 : \pi = 0$; there is a unit root.
$H_a : \pi < 0$; There is no unit root.

The PP test statistic values are calculated as follows (4):

$$Z_\pi = T\hat{\pi} - \frac{1}{2}\frac{T^2 \cdot SE(\hat{\pi})}{\hat{\theta^2}}(\hat{\lambda^2} - \hat{\theta^2}) \tag{4}$$

where $\hat{\theta^2}$ and $\hat{\lambda^2}$ are constant estimators of the variance of the parameters. The PP test statistic value compares the critical value, keeping in mind the selected significance level, where this critical value turns out to be the same as the ADF test. If the value of the test statistic is greater than the critical value in absolute value, the null hypothesis can be rejected, concluding that a unit root is not present.

Finally, the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) hypothesis test is:

$H_0 : \theta - \varepsilon = 0$; It is stationary.
$H_a : \theta - \varepsilon > 0$; It is not stationary.

The test statistic values (KPSS) are calculated as follows (5):

$$KPSS = \left(T^{-2}\sum_{t=1}^{T}\hat{S_t^2}\right)/\hat{\lambda}^2 \tag{5}$$

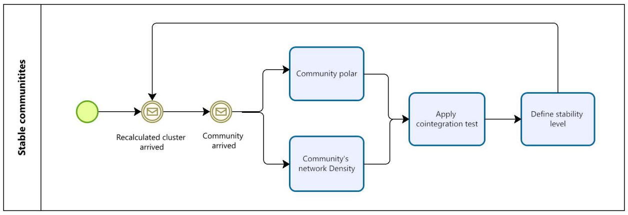

FIGURE 11. Cálculo de estabilidad para las comunidades.

where:

$\hat{S}_t^2 = \sum_{j=1}^{t} \hat{u}_j$ is the sampling variance of the model's sampling error (u_t).

$\hat{\lambda}^2$ is a consistent estimate of the long-run variance.

The test statistic (KPSS) value compares to the relevant critical value for the test with the selected significance level in mind. If the test statistic value is less than the critical value in absolute value, the null hypothesis can be rejected by concluding that a unit root is not present. Once most tests conclude the existence of a unit root problem, the next step is to transform the variable by differencing it so that the same ADF, PP, and KPSS tests are re-applied to conclude the non-existence of a unit root.

Then, cointegration is applied to measure the explanation of a long-run relationship, which infers a time-independent linear combination between the series; that is, a long-run relationship implying that the TAF in the analyzed community is constant. In other words, there are stable communities from the collective subjectivity since the TAF over time does not vary. The second step of the Engle-Granger methodology estimates the long-run equation or cointegration (6) between the pair of variables to be analyzed (polarity and network density) as follows:

$$Y_t = \delta_0 + \delta_1 X_t + u_t \tag{6}$$

where u_t is the error of the long-run relationship (disequilibrium) and can converts as $\hat{u}_t = Y_t - \hat{\delta}_0 + \hat{\delta}_1 X_t$. In other words, the cointegration test is a stationarity test of $\hat{u}_t$ by applying an ADF test. The test is applied to series that do not have unit roots, either in their levels or in their transformation (differenced series), that did not have this problem.

$H_0 : \hat{u}_t < 0$; Not stationary.

$H_a : \hat{u}_t = 0$; It is stationary.

The test statistic value (ADF) compares to the critical value for the Dickey-Fuller test with the selected significance level in mind. If the test statistic value is greater than the critical value in absolute value, the null hypothesis can be rejected, concluding that the error of the long-run equation is stationary and, hence, cointegration. Fig. 11 represents the activities flow for this component once the cointegration test is applied to the communities' different granularities following the COSSOL system onion-layer approach.

Once the general process has been explained for the cointegration test mentioned in the previous paragraphs, the COSSOL system applies the test for the different onion layers. The second onion layer consists of identifying the community to analyze under the different classes of the sociographic classifier (*economics*, *politics*, *culture*, *life and leisure*, and go on), which may or may not have a time gap problem that is solved as mentioned above. Thus the polarity and density series of the network are put under the different steps of the Engle-Granger methodology. Next, a tercile calculation applied to the network metric constructs the third onion layer, characterized by demonstrating the popularity within each second layer subcommunity; in this way, three new groups corresponding to **popular, average popularity, and unpopularity** are obtained. As more nodes are present in a network recalculation, the tweets posted in that recalculation were highly relevant as Twitter users joined the discussion that the original tweet contained. Meanwhile, a low level of nodes in the network present in the recalculations will mean users' lack of interest in discussing the tweets. Finally, the first onion layer is the agglomeration of the recalculations for each topic and, in turn, the sum across all topics.

IV. RESULTS

The experimentation scenarios in this work are two cases of analysis of collective subjectivity on Twitter under different levels of granularity. The first case studies the three onion layers by analyzing the communities from the SNA approach, where they belong in a general way to the total spectrum of the social network, and contemplates even those smaller ones generated by the application of the AS classifiers. The second case focuses on studying the communities belonging to the second onion layer and their recalculation to outline the temporal variation they present in a panorama of the hybrid COSSOL system proposal. In the latter case, designs a new digital ecosystem of ten new accounts associated with a different user, where these accounts were collected through a Gold-standard, ensuring the users' relevance to the topics within this scenario.

A. FIRST EXPERIMENTATION SCENARIO: COMMUNITY STABILITY

The first scenario refers to the three onion layers as a target of analysis against the proposed network structure integration proposal and the stability of the communities. In this case, the Engle-Granger methodology uses to apply a statistical cointegration test with which the long-term relationships are analyzed, examining the network density and polarity metrics. In addition, the scenario presents the quantitative and qualitative results in separate sections to describe the practical and theoretical implications of the results, respectively. The quantitative results present the descriptive statistical findings against the long-run relationships and conclusions to affirm the existence of these. In contrast, the qualitative results explain the implications of the long-run relationships on the hypothesized statements of the paper to reject or accept them, stating the reasons that led to the conclusion of such implications.

Table 15 shows the results of the first step of the [153] methodology on the three onion layers, demonstrating the presence or not of unit root for the polarity and density series of the network in levels and their first differences. The ADF and PP tests are applied to the series in levels, placing an asterisk when each test concludes the **not** existence of a unit root; however, the KPSS test is applied in levels when the conclusions of the two previous tests are contrary, indicating in a note below the network density series for each community analyzed.

Once the non-existence of unit root in levels has been defined for the series, it is not necessary to perform different tests on differentiated series since there is a stationarity property (no unit root). Now, the second and fifth columns of Table 15 contain the ADF test statistic in levels and first differences; the fourth and seventh columns present the critical value for the same test in levels, where the value of the statistic in absolute value must be greater than the critical value (5% significance). The third and sixth columns are the PP test in levels and first differences statistics, where its critical value is placed as a note at the bottom of the network density series for each community; as in the ADF test, the value of the statistic must be greater than the values placed in the notes. Finally, the KPSS test is applied for the first differences of the series with a value of the statistic observed in the eighth column, where a note contains its critical value looking for the statistic to be lower in absolute value to conclude the presence of unit root.

1st layer

- Neither of the two series has a unit root at a significance level of 5%, so performing their differences and corresponding tests was unnecessary.
- In both series, the same number of lags (3) were selected, which is statistically significant under a significance test applying a t-student statistic.

2nd layer

- The ADF and PP tests agree on the non-presence of a unit root problem in both series (polarity and network density) for almost all the communities except the *technology* topic. Therefore, the table shows larger statistics in absolute value concerning their critical value.
- In the topics of *life and leisure* together with *sports*, there were contrary conclusions under the ADF and PP tests in levels for the polarity series, so the KPSS test was performed, concluding the **NO** presence of unit root since its statistic was lower in absolute value than its critical value at 5% significance. Consequently, the PP and KPSS tests define a conclusion of not identifying the unit root, thus avoiding the need to calculate the first difference and its corresponding tests.
- In the case of the *technology* topic, there are multiple results. The first result evidence contrary conclusions between the ADF and PP tests on the network density series, showing that the latter mentioned test concludes the **NO** presence of unit root, and the ADF test, on the

TABLE 15. ADF, PP y KPSS tests of unit root.

Variable	ADF statistic (level)	PP statistic (level)	MacKinnon 5% critical value	ADF statistic first difference	PP statistic first difference	MacKinnon 5% critical values	KPSS statistics first difference
1st layer							
Polarity	-5.838(3)*	-5.584*	-2.860				
Network density	-8.822(3)*	-10.226*	-2.860				
5 % critical value PP for 1st onion layer is -2.860 at levels and first difference.							
2nd layer							
Culture							
Polarity	-3.139(0)*	-2.949*	-2.879				
Network density	-3.493(0)*	-3.604*	-2.879				
5 % critical value PP for 1st onion layer is -2.879 at levels and first difference.							
Politics							
Polarity	-3.512(1)*	-10.449*	-2.887				
Network density	-19.110(4)*	-22.976*	-2.887				
5 % critical value PP for 2nd onion layer is -2.887 at levels.							
Sports							
Polarity	-1.922(2)	-3.727*	-2.888				
Network density	-10.410(1)*	-110.085*	-2.888				
5 % critical value PP for 2nd onion layer is -2.888 at levels. Without a decision, the KPSS statistic was 0.45* with a 5 % critical value of 0.463, inferring NO unit root.							
Life an leisure							
Polarity	-2.843(2)	-3.4636*	-2.888				
Netwrok density	-5.576(4)*	-28.294*	-2.888				
5 % critical value PP for 2nd onion layer is -2.888 at levels and first difference. Without a decision, the KPSS statistic was 0.198* with a 5 % critical value of 0.463, inferring NO unit root.							
Technology							
Polarity	-1.812(1)	-1.924	-2.889	-0.392(0)	-0.391	-2,889	0.688
Network density	-1.855(1)	-3.702*	-2.889	-0.390(0)	-0.389	-2,889	0.688
5 % critical value PP for 2nd onion layer is -2.889 at levels and first difference. Without a decision, the KPSS statistic in polarity was 0.463* with a 5 % critical value, inferring unit root. Without a decision, the KPSS statistic in network density was 1.53 with a 5% critical value of 0.463, inferring the presence of unit root. In this case, ADF's statistics were -10.538* and -10.536* in second differences for the density and polarity variables, respectively, with a critical value of 5 % of -2.889. Likewise, the PP obtained a critical value of -10.536* with the same critical value of 5 % of the ADF test.							
Economy							
Polarity	-2.923(3)*	-3.011*	-2.888				
Network density	-5.448(3)*	-43.235*	-2.888				
5 % critical value PP for 2nd onion layer is -2.888 at levels							
3rd layer							
				Culture			
Popular							
Polarity	-0.806(4)	-1.657	-2.900	-5.630(3)*	-11.159*	-2.900	
Network density	-6.145(0)*	-11.145*	-2.900				
5 % critical value PP for 3rd onion layer is -2.900 at levels and first difference.							
Average popularity							
Polarity	-3.196(0)*	-3.306*	-2.900				
Network density	-10.055(2)*	-15.381*	-2.900				
5 % critical value PP for 2nd onion layer is -2.900 at levels.							
Unpopular							
Polarity	-1.821(0)	-1.720	-2.900	-9.808(0)*	-9.860*	-2,900	
Network density	-2.141(0)	-2.179	-2.900	-9.620(0)*	-9.627*	-2.900	
5 % critical value PP for 3rd onion layer is -2.900 at levels and first difference.							
				Politics			
Popular							
Polarity	-0.601(4)	0.421*	-2.933	-8.693(3)*	-17.250*	-2,933	0.0534(5)*
Network density	-0.004(0)	-0.005	-2.933	-7.323(0)*	-7.362*	-2.933	0.0546(5)*
5 % critical value PP for 3rd onion layer is -2.923 at levels and first difference. Without a decision, the KPSS statistic was 0.927 with a 5% critical value of 0.463, inferring unit root. The KPSS statistic with a 5% critical value of 0.463, inferring NO unit root.							
Average popularity							
Polarity	-0.235(4)	0.199	-2.933	-8.963(3)*	-16.697*	-2.933	0.071(6)*
Network density	-0.150(0)	0.161	-2.933	-7.088(0)*	-7.099(0)*	-2.933	0.0495(5)*
5 % critical value PP for 3rd onion layer is -2.933 at levels and -2.923 first difference. The KPSS statistic with a 5% critical value of 0.463 in first difference, inferring NO unit root.							
Unpopular							
Polarity	-2.054(0)	-5.665*	-2.936	-1.180(0)	-1.156	-2.936	0.815
Network density	-16.863(4)*	-11.806*	-2.944				
5 % critical value PP for 3rd onion layer is -2.936 at levels and first difference. Without a decision, the KPSS statistic was 0.845 with a 5% critical value of 0.463, inferring unit root in levels. In this case, ADF's statistics were -6.961* with a critical value of 5 % of -2.938. Likewise, the PP obtained a critical value of -6.969* with the same critical value of 5 % of the ADF test.							
3rd layer							
				Sports			
Popular							
Polarity	-1.999(0)	-2.002	-2.944	-6.735(1)*	-7.788*	-2.944	
Network density	-4.072(0)*	-3.925*	-2.944				
5 % critical value PP for 3rd onion layer is -2.936 at levels and first difference.							
Average popularity							
Polarity	-0.496(0)	-0.180	-2.944	-1.178(4)	-7.029*	-2.944	
Network density	-4.175(0)*	-4.176*	-2.944				
5 % critical value PP for 2nd onion layer is -2.944 at levels and first difference.							
3rd layer							
				Sports			
Unpopularity							
Polarity	-2.076(1)	-2.973*	-2.944	-13.456(0)*	-13.628*	-2,947	
Network density	-73.156(0)*	-72.619*	-2.944				
5 % critical value PP for 3rd onion layer is -2.944 at levels and -2.947 first difference.							
				Life and leisure			

TABLE 15. *(Continued.)* ADF, PP y KPSS tests of unit root.

Popular							
Polarity	-2.187(2)	-1.974	-2.947	-3.288(1)*	-6.839*	-2.947	
Network density	-2.641(4)	-2.367	-2.947	-5.226(0)*	-5.318*	-2.947	
5 % critical value PP for 3rd onion layer is -2.947 at levels and first difference.							
Average popularity							
Polarity	-1.304(0)	-1.187	-2.947	-7.296(0)*	-7.297*	-2.947	
Network density	-0.837(1)	-1.882	-2.947	-1.330(0)	-1.129	-2.947	
5 % critical value PP for 3rd onion layer is -2.947 at levels and first difference. In this case, ADF's statistics were -5.295* in second differences with the same critical value of 5 % of first differences. Likewise, the PP obtained a critical value of -5.078* with the same critical value of 5 % of first differences.							
Unpopularity							
Polarity	-3.244(2)*	-4.229*	-2.952				
Network density	-6.817(1)*	-22.029*	-2.950				
5 % critical value PP for 3rd onion layer is -2.947 at levels.							
3rd layer							
			Technology				
Popular							
Polarity	-0.993(0)	1.526	-2.964	-0.804(0)	-0.798	-2.964	0.711
Network density	-2.055(1)	-5.539*	-2.964	-0.802(0)	-0.796	-2.964	0.711
5 % critical value PP for 3rd onion layer is -2.964 at levels and first difference. Without a decision, the KPSS statistic was 0.746 with a 5% critical value of 0.463, inferring the presence of unit root at levels. The KPSS statistic with a 5% critical value of 0.463 in first differences, inferring unit root. In this case, ADF's statistics were -6.165* and -6.164* in second differences for the density and polarity variables, respectively, with a critical value of 5 % of -2.964. Likewise, the PP obtained a critical value of -6.167* and -6.166 with the same variables in a critical value of 5 % of the ADF test.							
Average popularity							
Polarity	-0.219(4)	-0.590	-2.964	-7.228(3)*	-21.653*	-2.964	0.055*
Network density	-0.317(2)	-0.133	-2.964	-14.476(1)*	-16.954*	-2.964	0.071*
5 % critical value PP for 3rd onion layer is -2.964 at levels and first difference. KPSS statistic was 0.463* with a 5% critical value of 0.463 in first differences, inferring NO unit root.							
Unpopularity							
Polarity	0.175(4)	-0.190	-2.978	-8.770(3)*	-21.151*	-2,978	0.069*
Network density	-0.988(2)	0.480	-2.972	-16.873(1)*	-15.663*	-2,972	0.0849*
5 % critical value PP for 3rd onion layer is -2.966 at levels and -2.969 first difference. KPSS statistic was 0.463* with a 5% critical value of 0.463 in first differences, inferring NO unit root.							
			Economy				
Popular							
Polarity	-3.151(3)*	-2.493	-2.947	-6.038(0)*	-6.084*	-2.947	
Network density	-1.557(0)	-0.639	-2.947	-4.951(0)*	-4.979*	-2.947	
5 % critical value PP for 3rd layer is -2.947 at levels and first difference.							
Average popularity							
Polarity	-1.658(0)	-1.540	-2.947	-7.328(0)*	-7.511*	-2.947	
Network density	-1.729(1)	-2.717	-2.947	-5.088(0)*	-5.087*	-2.947	
5 % critical value PP for 3rd onion layer is -2.947 at levels and first difference.							
Unpopularity							
Polarity	-2.115(0)	-2.162	-2.950	-6.464(0)*	-6.469*	-2.952	
Network density	-3.537(3)*	-35.681*	-2.958				
5 % critical value PP for 3rd onion layer is -2.950 at levels and -2.952 first difference.							

Notes: Critical values are from MacKinnon (1991).The terms in parentheses are the optimal number of lags chosen by the Akaike Information Criterion (AIC). Denote significance at the 5% correpondingly.*

contrary, concludes the presence of this problem. Consequently, the second result arises when the KPSS test is applied in levels, but contrary to the previous result, the test ratifies the presence of unit root, generating the need to create the first difference for this series.

- The third result manifests the need to make the ADF, PP, and KPSS tests for the first difference of the network density series, where all tests conclude the presence of unit root. The fourth result describes the need to make the first difference for the polarity series, which concludes that there is still a unit root problem since the ADF and PP statistics are not greater than their critical values and the KPSS statistic is not less than its critical value. Finally, the fifth result describes the tests performed for both series in second differences where a note describes the conclusion that there is no longer a unit root according to the statistics and critical values results.

3rd layer
Popularity clusters for the topics of culture and politics.

- The only cluster of popularity along the two analyzed topics, and where there is **NO** a unit root in both series at levels, is the one with average popularity for the topic of *culture*. Moreover, this group is the only one among the *culture* topic that meets the condition mentioned above.
- The popular and average popularity clusters for the topic of *politics* with the unpopular cluster for *culture* had unit roots in both series. However, the polarity series in the popular cluster topic of *politics* had to apply the KPSS test confirming the conclusion of the ADF test, which presented contradictions in the conclusions with the PP test. Therefore, the calculation of the first difference was necessary for both series conclude that these **NO** have unit root with the unanimous agreement of the available tests, making the caveat that the *political* clusters were able to apply the KPSS test since the dates of these clusters were continuous.
- The popularity cluster in the *culture* topic exhibits unit root in levels for the polarity series, requiring testing on the first difference in this series, where the conclusion affirms there is **NO** unit root. On the other hand, the network density **NO** had a unit root in levels by the consensus of the ADF and PP tests by presenting statistics greater than their critical values.
- In the case of the unpopularity cluster for the *politics* topic, it presents multiple results. The first result demonstrates the **NO** presence of a unit root in the network density series in levels. The second result evidences the presence of unit root where there were contrary conclusions between the ADF and PP tests in the polarity series in levels, arising the need to apply the KPSS test, which concludes the presence of such unit root. The third result reveals the need to perform the ADF, PP, and KPSS tests for the first difference of the polarity series, where all tests conclude the presence of unit root. Finally, the fourth result describes the tests performed for the polarity series in second differences, where it concludes that already **NO** unit root exists in a note according to the results of the statistics and critical values.

3rd layer
Popularity clusters for the topics of sports and life and leisure.

- The only cluster of popularity along the two analyzed topics where in both series at levels there is **NO** a unit root is unpopular for the *life and leisure* topic.
- The *sports* topic presents a particular case between the popular and unpopular clusters since the conclusions between the series are similar. The first result is the **NO** presence of unit root in the network density series, where this conclusion is more evident due to huge statistics in absolute value to the critical value of the ADF and PP tests. The second result is the presence of unit root in the polarity series but with a detail in the unpopular cluster where there were conflicts of conclusions between the ADF and PP tests, leaving the conclusion of the ADF test before the discontinuity of the recalculations belonging to that cluster. Finally, the third result is the need to differentiate the polarity series where both tests conclude the **NO** presence of unit root.

- The average popularity cluster for the *life and leisure* topic presents multiple results. The first result demonstrates the **NO** presence of unit root in the polarity series in the first difference, which translates to this series in levels reaching the same conclusion between the ADF and PP tests identifying unit root. The second result is the presence of unit root in the network density series in levels where the first difference of the series occurs, establishing tiny statistics in absolute value reflecting the fact of applying a second difference in the series so that the **NO** unit root is concluded with the ADF and PP tests.
- The average popularity for the *sports* topic in the polarity series presents a particular case since **NO** there was a consensus among the tests to conclude the **NO** presence of unit root in the first difference. First, the series at levels affirms the presence of unit root in the presence of statistics lower than its critical values, forcing differences in the series. Then the series in first differences contradicts the tests' conclusions with the impediment that a KPSS test cannot be performed since the recalculations belonging to this cluster have discontinuous dates. Therefore, it is not necessary to further differentiate the series before the PP test, which may be sufficient for not applying another difference.

3rd layer

Popularity clusters for the topics of technology and economy.

- With some exceptions, the *economy* topic presents similarities among its popularity clusters, specifically in the popular and average popularity clusters. First, the ADF and PP tests conclude the presence of unit root in levels on the network density series. Similarly, the polarity series in levels concludes unit root; however, the polarity for the popular cluster in this topic is at the limits to conclude the opposite. Finally, the series were differentiated by applying the ADF and PP tests that identified the **NO** presence of unit root when statistics of the respective tests for each series were greater in absolute value than its critical values. It is worth highlighting the selection of lags for the last mentioned cluster since no lags were added to perform the tests.
- The *technology* clusters for average popularity and unpopularity have unit roots in levels; that is, both network density and polarity in levels have statistics lower than their critical values at a significance level of 5%. Consequently, both series were differentiated to apply the ADF, PP, and KPSS tests to verify the existence of unit root, where the results concluded the **non**-existence of unit root. It is worth noting that the lags selected for each series are the same across the different popularity clusters; for example, the polarity series in the first difference selects three lags in both average and unpopularity.
- The unpopularity cluster in the *economy* topic presents a different behavior than its other clusters, represented in the network density series where it is concluded with the ADF and PP tests that **NO** exists a unit root in levels. However, the conclusions for the polarity series follow the line of its other clusters, which have unit root in levels, creating the first difference and concluding the **NO** existence of unit root.
- The popular cluster on the *technology* topic has a particular behavior that very few subcommunities have. Therefore, in both series, it is necessary to make the second difference to check for the presence of the unit root. In addition, the network density series in levels presented conflicts between tests, so it was necessary to apply the KPSS test, which rectified the conclusion presented by the ADF test

Table 16 presents the key results that support the conclusions of the hypotheses raised in this paper and the applications to the different fields of industry. The table describes the regressions created to apply the second step of the Englé-Granger methodology, which formally applies an ADF test to the regression errors to conclude the existence of a long-run relationship between the network density and polarity variables. The first column describes the layer to which the regression to be tested belongs, and in this case, all the communities of the different onion layers were tested since the table 15 showed that all the series **NO** have unit root problems in levels or their differences.

Now, concerning the regressions, these consist of selecting the polarity series as the dependent variable and the network density series as the independent variable. This series assignment to the order of each regression is on a logic of expected behavior in the relationship that these two series hold. For example, as network density rises, the polarity is expected to have a greater tendency toward one of its different categories (positive, neutral, or negative) since the inclusion of more nodes in the network means a more enriched discussion of different points of view with a more defined calculation over a polarity category.

The fourth, fifth, and sixth columns extract the regression data to analyze the constant impact, the impact that density has on polarity and the sense of the relationship between these, and finally, the explanation that the density of the network has on the behavior of the polarity series. Finally, the errors are calculated for each regression where the seventh, eighth, and ninth columns conclude the cointegration test result. The seventh column has the statistic of an ADF, the eighth column has the critical value for the mentioned test, and the ninth column expresses in words the results of the seventh and eighth columns.

Main results of the cointegration of the three onion layers of the COSSOL system

- The first onion layer regression shows the existence of cointegration meaning into a long-run relationship between the network structure and collective subjectivity series. The rate of the polarity series reflects its growth trend in the 3.822773 units of the constant coefficient. For its part, the coefficient of the explanatory vari-

TABLE 16. Engle–Granger cointegrating regressions.

Coeficientes de constante	Coeficientes de variable explicativa	Adjusted R2	ADF(*) para prueba en los residuos	MacKinnon 5% or 10% critical values	Cointegration conclusion
1st layer					
3.822773	126.0675	0.2269	-5.642(1)*	-2.9860	SÍ
2nd layer					
Culture					
3.81837	223.8647	0.7826	-3.372.(0)*	-2.879	SÍ
Politics					
3.027042	176.9898	0.4673	-4.403(1)*	-2.887	SÍ
Sports					
3.973571	66.27022	0.0242	-2.680(1)**	-2.888 ; -2.578	SÍ**
Life and leisure					
3.614344	-18.50468	0.0003	-2.831(2)	-2.888	NO
Technology					
2.593838	2595.697	0.7982	-1.273(1)	-2.889	NO
Economy					
3.644208	177.0627	0.0804	-3.945(3)*	-2.888	SÍ
3rd layer					
Culture-Popular					
5.903571	-23082.33	0.1321	-3.490(0)*	-2.879	SÍ
Culture-Average popularity					
4.113966	-2560.838	0.0031	-3.462(0)*	-2.879	SÍ
Culture-Unpopularity					
3.856621	219.5214	0.8701	-2.541(4)	-2.880	NO
Politics-Popular					
3.30469	-1095.055	1.000	-7.764(4)*	-2.887	SÍ
Politics-Average popularity					
3.304649	-1095.074	1.000	-7.764(4)*	-2.887	SÍ
Politics-Unpopularity					
3.061559	152.7961	0.4208	-4.172(1)*	-2.887	SÍ
Sports-Popular					
-9.457429	73340.53	0.1902	-10.538(1)*	-2.888	SÍ
Sports-Average popularity					
1.851913	9431.768	0.0294	-130.796(0)*	-2.888	SÍ
Sports-Unpopularity					
4.206861	40.13331	0.0155	-2.901(1)*	-2.988	SÍ
Life and leisure-Popular					
2.383036	7995.855	0.0034	-4.338(2)*	-2.888	SÍ
Life and leisure-Average popularity					
2.57702	4023.499	0.0148	-3.227(2)*	-2.988	SÍ
Life and leisure-Unpopularity					
3.500089	40.87842	0.0061	-2.868(2)**	-2.888 ; -2.578	SÍ**
Technology-Popular					
4.289993	-2910.367	0.2846	-1.847(1)	-2.889	NO
Technology-Average popularity					
2.248314	3426.491	1.000	0.000(1)	-2.889	NO
Technology-Unpopular					
2.248312	3426.496	1.000	-0.000(1)	-2.889	NO
Economy-Popular					
2.323263	5354.763	0.0201	-2.632(3)**	-2.888; -2.578	SÍ**
Economy- Average popularity					
0.659045	7522.296	0.1029	-2.932(3)*	-2.888	SÍ
Economy-Unpopular					
3.774501	161.5177	0.1306	-3.459(3)*	-2.888	SÍ

*Notes: The terms in parentheses are the optimal number of lags determined by the Akaike Information Criterion (AIC). Critical values for the Engle–Granger test are from MacKinnon (1991).∗∗ and * denote significance at the 10% level and 5%,correpondingly.*

able presents the sense and magnitude of the relationship that the network density has with the polarity series. For example, as can be seen in the table 16 for the first onion layer, an increase in polarity by 126.0675 units is generated in response to a one unit increase in network density, demonstrating a direct relationship between the increase in network density and the response variable (polarity).

- The first onion layer regression shows the existence of cointegration meaning into a long-run relationship between the network structure and collective subjectivity series. The rate of the polarity series reflects its growth trend in the 3.822773 units of the constant coefficient. For its part, the coefficient of the explanatory variable presents the sense and magnitude of the relationship that the network density has with the polarity series. For example, as can be seen in the table 16 for the first onion layer, an increase in polarity by 126.0675 units is generated in response to a one unit increase in network density, demonstrating a direct relationship between the increase in network density and the response variable (polarity).
- Among the 25 regressions calculated to conclude a long-run relationship in the network structure and collective subjectivity series, nine regressions identify the non-existence of a long-run relationship. Among the 36% of the communities, six whose statistics evaluated at a significance level of 5% were lower than the critical value. In comparison, the remaining three concluded the existence of a cointegration once the significance level at the critical value was relaxed by 10%; therefore, they belong within this group.
- The direction of the relationship between the network density series and polarity is crucial as it affects the interpretation of the coefficient reflected by the explanatory variable in the first result. Six of the 25 regressions (25% of the communities) showed an indirect relationship since polarity decreased as network density increased. However, four regressions (16% of the communities) had a long-term relationship, attracting the focus of attention for this result.
- The results demonstrate the presence of a long-term relationship between the polarity and net density series in all onion layers, with the nine exceptions mentioned below:
  - Layer 2: Sports
  - Layer 2: Life and leisure
  - Layer 2: Technology
  - Layer 3: Culture - unpopular
  - Layer 3: Life and leisure- unpopular
  - Layer 3: Technology - popular
  - Layer 3: Technology- average popularity
  - Layer 3: Technology - unpopular
  - Layer 3: Economy - popular

QUALITATIVE RESULTS

A long-term relationship refers to the non-dependence of the series on events over time; in other words, the metric of network structure (network density) and subjectivity (polarity) implies that the network nodes (Twitter users) form communities around a topic, and these preserve polarity in the long term, which is called stability in the communities. Therefore, stability is a phenomenon where a community's polarities and network density oscillate when an event or news is published on the social network. However, these oscillations are no longer significant since their behavior over the analyzed entire time horizon shows a constant pattern.

A long-term relationship for the polarity of the communities is understood as a short-term fluctuation in the new information discussed in the social network, adding new users and polarities of these nodes to the woven network of the community. However, the accumulation of their polarities coming from new events or news causes the networks to be recalculated, inferring a greater discussion within the community, accentuating the polarity towards a more evident trend with its SNA counterpart. The following paragraphs highlight the main elements for each hypothesis, the expected results of each element, as well as those obtained by combining the main elements, the exploration results incidence on the hypotheses raised in this paper, and the description of the reuslts implications to conclude the acceptance or rejection them.

Hypothesis 1: The communities with a higher index of centrality in a subset of members present greater stability in the collective subjectivity in the face of a topic disseminated in that community.

The first element is the presence of a higher centrality index in a subset of members, understanding that the centrality of the network is higher the more granular the community is, i.e., the centrality index increases as the analysis is carried out in smaller onion layers since the spectrum of the discussion on a topic is more specific.

The second element refers to the identification of an increase in the stability of collective subjectivity (polarity) in the same community analyzed, and it means stability when its behavior in the long term presents a constant pattern. Regarding the latter, the cointegration test concludes the presence or absence of stability concerning network metrics and collective subjectivity; therefore, the expected result of the cointegration test consists of identifying the absence of long-term relationships in the larger onion rings and, on the contrary, finding them in more granular layers.

By combining the expected results of the main elements, it is expected to find long-term relationships in more granular onion layers since there is a higher centrality index. Therefore the conformation of the network achieves relationships with the metric of collective subjectivity reaching constant patterns that opaque short-term fluctuations. The results show long-term relationships from the first onion layer, where the spectrum of discussion is broad since it includes all possible discussions published on the Twitter network. However, there are topics such as *technology* where the behavior is anomalous to the described logic for the set of expected results of the main elements. For example, the first onion layer is related in the long term to the metrics under consideration, but the second layer does not do so, and neither in all the popularity clusters.

The *life and leisure* topic behaves similarly, where the first layer presents a long-term relationship, but in the second one, it disappears, as does the popularity cluster (unpopular). In conclusion, the hypothesis is rejected because there is no stability in all the smaller onion layers, making the caveat for the *technology* topic.

Hypothesis 2: the most stable communities in polarity terms concerning a topic are those in which their members are highly connected.

The main elements present in this hypothesis are the stability of a community concerning a topic and the high connectivity of its members. The first element highlights the term with which the metric is used to define stability, expecting a more biased polarity in more granular communities since the members increase as new users identify with this bias TAF. It is called a tendency to associate with their peers (homophily); therefore, smaller onion layers are prone to offer these spaces when discussing topics in more detail. The second main element identifies those members who are highly connected, with the understanding that the network density metric allows showing whether they are cohesive; hence, more granular communities will have higher network densities since the distances between nodes are not as large compared to more extensive onion layers.

Thus, by combining the expected results of the principal elements, it is expected to have long-term relationships as the onion layer becomes more granular. The experiment confirms the expected results when combining the above main elements. Five of the six subjects: *sports*, *life and leisure*, *culture* y *economy*, present long-term relationships in most of the clusters belonging to the third onion layer (highly connected sets), corresponding to more stable communities compared to the second onion layer, where there is no long term relationship among them.

However, the *technology* topic is the exception to the expected results when combining the main elements, since methodologically, the communities are cumulative network constructions; so, the third onion layer conforms to the second one when the number does not make the distinction of clusters of nodes in the network. The first layer is the set of communities when the sociographic classifier is not applied, so if there is no long-term relationship in the third onion layer, the probability of this occurring in larger onion layers is very high. In conclusion, the hypothesis is accepted according to the results for the third and second onion layers.

Hypothesis 3:the communities with a higher network density and a common polarity in a subset of members are more highly connected to a topic.

This hypothesis highlights two main elements to be analyzed to conclude the rejection or acceptance. The first element is the presence of densities in networks for less granular communities (higher level onion layers), understanding that the expected result is the increase of this metric as the onion ring becomes more granular. The second element is the common polarity for the communities for each onion layer; the expected result focuses on finding communities that exhibit long-term relationships across the totality of the different onion rings.

By combining the expected results of the main elements, it is expected to find a topic with long-term relationships over the three onion layers. Keeping in mind that their densities increase as the hoop becomes more granular, a topic will evidence a better propagation of collective subjectivity when there is a uniformity of long-term relationships in its onion layers, contrary to those topics where some of its layers do not present such a relationship. The results of the experiment show long-term relationships in the three onion layers only in the political theme. In summary, the hypothesis is accepted as its analysis is applied to the topic of *politics*, affirming that this topic propagates collective subjectivity better. However, this conclusion applies to the topics of *life and leisure* and *culture* since their degree of less popularity (unpopular) and

TABLE 17. Summary of community centrality.

Topic	All degree Centrality	Betweenness Centrality	Closeness Centrality	Degree	Density
Culture	0.52315310	0.32620347	0.41220377	5.4672	0.00762514
Sports	0.40380101	0.28995701	0.42476387	4.9329	0.01743061
Economy	0.41653478	0.25921835	0.23470057	5.0761	0.00858907
Politics	0.56128530	0.29497507	0.49343354	5.6695	0.01615247
Technology	0.50959453	0.27053341	0.38651990	5.1589	0.00671726
Life and leisure	0.55862809	0.36926947	0.39298180	4.5667	0.01268519

therefore the one with less interaction among users, is the only cluster that does not present a long-term relationship, concluding the presence of uniformity in all its onion rings.

B. SECOND EXPERIMENTATION SCENARIO: RECALCULATION OF COMMUNITIES

Discussions on Twitter present temporal phenomena of change once a new event or occurrence has been posted. Therefore, variations of TAF's mode expressions are more evident in more granular communities since most community users stop discussing past events. The designed scenario consists of the topics in the second onion layer, which generates the communities to analyze. The distinction of the topics enriches the collective subjectivity analysis since the TAF are shared on the interest of a particular content present in the network, so scenarios provide a comparison of the temporal variation between the topics.

Table 17 presents a summary of the different centrality metrics for the topic networks corresponding to the second community, the first community being the whole of Twitter. The *politics* topic has the highest values for each type of centrality, which results in the largest amount of connections between users or the most connected network, a higher degree of information brokering by specific nodes, and finally, the smallest distance between nodes making the TAFs present have a greater reach to any node in the network. On the other hand, the *economy* topic has the lowest centrality values by intermediation and proximity, reflecting information widely shared by several users who become information nodes for others. In the same vein, the *economy* community has the information reach over other smaller nodes, which may reflect the low preference of users to read publications on this topic. In the end, the topic of *sports* presents the lowest connectivity or lowest number of connections between nodes.

Variation of all degree centrality metrics

The network connections indicate the connectivity degree between the topics and each node (user). Thus, a temporal analysis applied to examine the proportion of change or variance to the number of connections for each topic represents the life cycle that discussions may have: the longer the threads of conversation last, the higher the degree of connectivity since there are two conditions. The first one exists because of its inherent characteristic of having a degree of homophily for being in that community; the second one is fulfilled when the notifications of its closest relations increase the probability of making a comment replying to such a publication. The table 18 presents the variations for the degrees of connectivity for the topics in this scenario, seeking to look at lower proportions of variation since it would indicate longer life cycles.

TABLE 18. Descriptive statistics for all degree centrality.

Topic	Varianaza All degree centrality	Desviación estándar All degree centrality
Culture	1.66%	227.99%
Politics	1.12%	150.83%
Sports	0.46%	61.90%
Vida	0.19%	24.90%
Technology	0.40%	53.75%
Economy	0.28%	37.40%

The table shows that *culture* and *politics* topics have the highest variance of centrality by degree among the network recalculations; on the other hand, *life and leisure* and *economy* are topics whose variations are the smallest. Similarly, the standard deviation between network recalculations evidences the same phenomenon for the abovementioned topics. The results of the statistics shown are indications of the variability of the discussion for the different topics. Consequently, the topics of *culture* and *politics* present very different events or occurrences. In contrast, the topics of *life and leisure* and *economics* seem to have seasonal occurrences or events that are frequently discussed.

Network density metric variation

The number of relationships observed in the network times the total number of possible relationships in the community demonstrates the number of users sharing a particular topic, generating discussions to share their TAFs. Consequently, denser communities demonstrate a greater number of users sharing their TAFs driving the conversation towards a common polarity. The table 19 shows the variations of the network density metric by observing which are the topics with the highest level of variance indicating events or occurrences of greater interest for Colombians; that is, it can be observed which are the topics that have a bigger number of threads of conversation.

The table shows minimal variances for each topic among the recalculations of the networks. Thus, the topics of *politics* and *life and leisure* are communities that are renewing the users who generate the discussion stored in the conversation threads. Although the degree centrality result for the *life and leisure* topic concluded similar events each time there is a network recalculation, these results show that it is not the same users who are in charge of generating the discussion on these topics. Finally, the standard deviation shares in pairs the percentage of the distance of the number of users over their mean; *politics* and *life and leisure* topics are the first pair with the highest network density index; on the other hand, the topics of culture and sport are the second pair with a percentage of 0.02%; finally, there are the technology and economy topics.

The following sections present the dynamics of the metrics associated with the Computational Linguistics and Computational Sociology components for the scenarios constructed. In each, three types of analysis were established to explain

TABLE 19. Descriptive statistics for network density.

Topic	Varianza	Desviación estándar
Culture	0.000134%	0.02%
Politics	0.000206%	0.03%
Sports	0.000183%	0.02%
Vida	0.000222%	0.03%
Technology	0.000107%	0.01%
Economy	0.000109%	0.01%

TABLE 20. Structural metrics culture network.

All Degree		*Closeness*		*Betweenness*	
Value	**Label**	**Value**	**Label**	**Value**	**Label**
379	JacquesTD	0.6359	JacquesTD	0.3276	JacquesTD
311	DianAngel01	0.6167	SoyAndresParra	0.2431	SoyAndresParra
296	SoyAndresParra	0.5859	DianAngel01	0.2226	DianAngel01
262	jeabello	0.5226	oficialHASSAM	0.1457	jeabello
206	MFValdesV	0.5032	NoticiasCaracol	0.1131	oficialHASSAM
177	VelezMauricio	0.5032	ELTIEMPO	0.0994	MFValdesV
155	oficialHASSAM	0.5028	YouTube	0.0994	VelezMauricio
120	FelicianoValen	0.5028	BluRadioCo	0.033	FelicianoValen
63	SebastianYatra	0.5028	IvanDuque	0.0092	SebastianYatra
9	NoticiasCaracol	0.5025	elespectador	0.001	NoticiasCaracol

the practical contributions to the hybrid approach for the proposed analysis of collective subjectivity contemplated by the COSSOL system. The first analysis shows the structural metrics to demonstrate a ranking of the ten most important accounts to their structure, thus evidencing the role of these top accounts vis-à-vis the scenarios. The second type of analysis graphically presents the polarity (network graphs) associated with the ranking described for the previous type of analysis; consequently, each network graph shows the general sentiment present in the most abundant color of the network, the particular sentiment on each node and the level of relevance that particular sentiment has on the network, employing the size for each node. Finally, the third type of analysis makes a count of recalculations by polar distributions in a table to record the frequencies over the distributions that were theoretically added to observe the number of creations of new discussions represented in the variations of the TAF for each scenario. Additionally, the amount described in the table and the polar distribution graphical representation state of consensus and dissent in each scenario measures the closeness between these theoretical distributions and the observed polar distribution.

C. CULTURE

Regarding the *culture* community (onion layer), the degree centrality metric, as shown in the 20 table, presents the highest value in the @JacquesTD account, an account belonging to a well-known Colombian actor.

The value of the closeness metric for this same actor relates to the closeness of this user to any node in the network, demonstrating the incredible reach that the actor's publications can have on other users. For example, there is evidence of greater closeness to the other nodes of @SoyAndresParra and @DianAngel01, second and third place in this indicator, respectively. On the other hand, the betweenness metric refers to the intermediations necessary to flow messages between nodes. The accounts mentioned in the descriptions of the other metrics remain in the top 3, so their influence on information transfer is identified. On the other hand, it is worth mentioning the behavior of the news center @NoticiasCaracol, which, being in fifth place with a great reach of its publications to other users, presented the last place for the number of connections with other nodes. Additionally, it is essential to highlight that the value of the betweenness of this account is the lowest among all the accounts in the ranking found, even though it represents a recognized national media outlet.

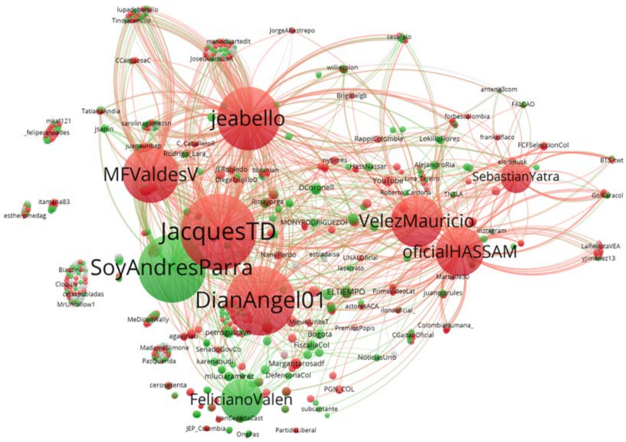

FIGURE 12. Polar layer network graph for the topic of culture.

Regarding the polar distribution, the Fig. 12 represents the node with the highest centrality *all degree* in red, indicating the negative comments that are evident in the mentions of the users regarding the actor @JacquesTD. This same node relates the most significant number of connections; hence, it is the largest node in the figure. However, the network graph evidences a slight change in the node sizes showing the differences in the centrality values *all degree* among the top 10 accounts. In addition, the top nodes' TAFs in the network show negative sentiment except for @SoyAndresParra and @FelicianoValen, located in the third and seventh place of *all degree* centrality.

The statistics of the table 21 for the topic of *culture* record the frequencies of the polar distributions throughout the analyzed time. There the frequency increased each time the cluster groups by popularity were recalculated. Recalculation is validating the existence of more than ten accumulated errors within a time range, which is performed automatically by repeating the popularity calculation using the k-means technique.

The table shows a total of 137 recalculations for the topic of *culture*, where the Beta distribution was assigned 125 times, representing 91.24% and obtaining a significant difference over the Uniform distribution, which can be observed in the Fig. 13.

D. POLITICS

In the *political* community (onion layer), the average distance metrics or closeness metrics, shown in the

TABLE 21. Frequencies of polar distributions by recalculation of culture clusters.

Distribución	Frecuencias
Weibull	2
Valor extremo generalizado	0
Uniforme	10
Beta	125
T-student	0
Normal	0
Gama logaritmo natural	0
Total	137

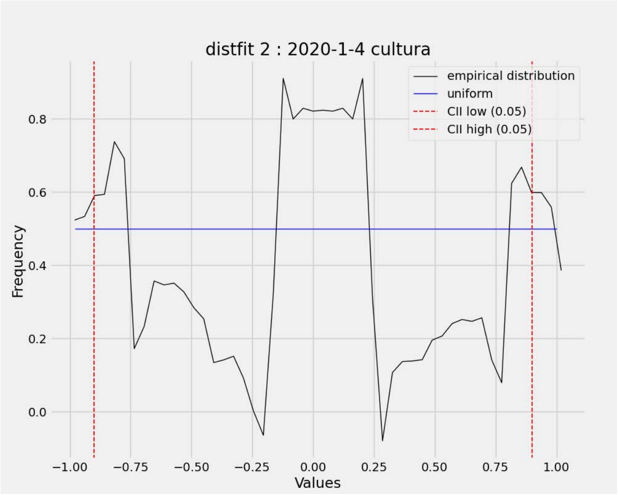

FIGURE 13. Polar distribution for the topic of culture.

TABLE 22. Structural metrics politics network.

Closeness		All Degree		Betweenness	
Value	Label	Value	Label	Value	Label
0.7014	JohnMiltonR_	201	JohnMiltonR_	0.2976	JohnMiltonR_
0.6641	sergio_fajardo	173	sergio_fajardo	0.227	sergio_fajardo
0.6506	FicoGutierrez	162	FicoGutierrez	0.2021	FicoGutierrez
0.625	Luis_Perez_G	140	Luis_Perez_G	0.164	Luis_Perez_G
0.6119	petrogustavo	128	petrogustavo	0.1341	petrogustavo
0.5993	IBetancourtCol	116	IBetancourtCol	0.1108	IBetancourtCol
0.5795	Enrique_GomezM	96	Enrique_GomezM	0.0788	Enrique_GomezM
0.5051	larepublica_co	7	larepublica_co	0	MinInterior
0.5051	BluRadioCo	7	BluRadioCo	0	angelamrobledo
0.5051	NoticiasCaracol	7	NoticiasCaracol	0	MinjusticiaCo

table 22, present in the first place the account @JohnMiltonR_, presidential candidate, followed by the candidates: @sergio_fajardo, @FicoGutierrez, @Luis_Perez_G, @petrogustavo, @IBetancourtCol, and @Enrique_GomezM, with average distances greater than 0.5795. This behavior is because it reflects two main phenomena; the first is the candidates' activity on Twitter as a response to their interest in obtaining a greater reach to other users a week before the first round of the presidential elections; while the second refers to the users' activity in terms of discussing the candidates' government proposals and thus influencing the opinion of other users. The order of these accounts along the centrality metrics does not change, thus reaffirming the candidates' leadership (@JohnMiltonR_ and @sergio_fajardo) over the analyzed community.

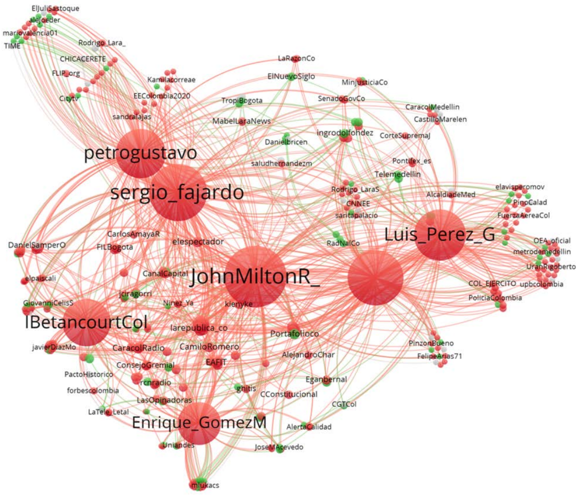

FIGURE 14. Polar layer network graph for the topic of politics.

TABLE 23. Frequencies of polar distributions by recalculation of politics clusters.

Distribución	Frecuencias
Weibull	8
Valor extremo generalizado	1
Uniforme	1
Beta	125
T-student	0
Normal	0
Gama logaritmo natural	0
Total	135

Fig. 14 shows the overall negative sentiment about all candidates, exposing the polar distributions of these users. Furthermore, the presence of the existing nodes in the *politics* community is primarily negative, with few exceptions such as Radio Nacional de Colombia (@RadNalCo), the economic newspaper Portafolio (@Potafolioco), and the channel Telemedellin (@Telemedellin). Finally, the node sizes do not differ due to the small distance between the centralities *all degree* for each node in the top 10.

The statistics of the polar distributions for the *politics* topic are presented in the table 23 over the entire period analyzed. There, the frequency increased each time the cluster groups by popularity were recalculated.

The table shows a total of 135 recalculations for the *politics* topic, where the Beta distribution was 125 times, representing 92.259%, followed by the Weibull, Generalized Extreme Value, and Uniform distributions. Similar to the results with the *culture* theme, the Beta distribution obtains a large difference over the remaining distributions. In Fig. 15 the Generalized Extreme Value distribution can be observed.

E. SPORTS

Regarding the sports community (onion layer), the account that has a higher value of metrics in closeness and *all degree* is that of the Olympic diver @Sebvilla92, which is surpassed

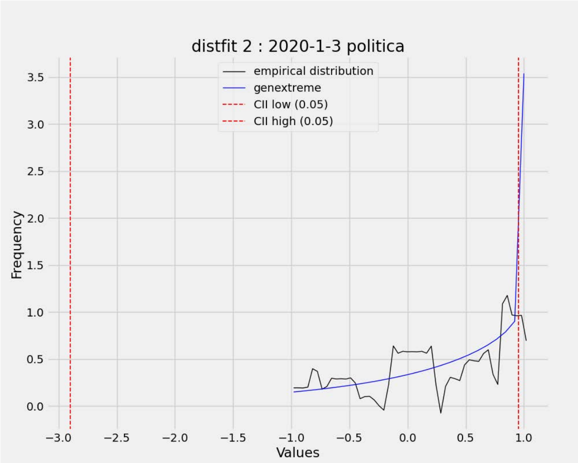

FIGURE 15. Polar distribution for the topic of politics.

TABLE 24. Structural metrics sport network.

Closeness		All degree		Betweenness	
Value	Label	**Value**	**Label**	**Value**	**Label**
0.6253	Sebvilla92	118	Sebvilla92	0.294	GHerreraCas
0.6225	GHerreraCas	111	GHerreraCas	0.281	Sebvilla92
0.59	jamesdrodriguez	90	jamesdrodriguez	0.2182	jamesdrodriguez
0.5709	Eganbernal	81	Eganbernal	0.1958	Eganbernal
0.5551	Cuadrado	76	FndoGaviria	0.1406	FndoGaviria
0.5465	FndoGaviria	74	carlossanchez6	0.1287	Cuadrado
0.5081	NairoQuinCo	69	Cuadrado	0.0999	carlossanchez6
0.5072	FALCAO	50	HiguitSergio	0.0428	HiguitSergio
0.5063	UranRigoberto	47	Ncadena98	0.0217	Ncadena98
0.5036	carlossanchez6	9	NairoQuinCo	0.0032	NairoQuinCo

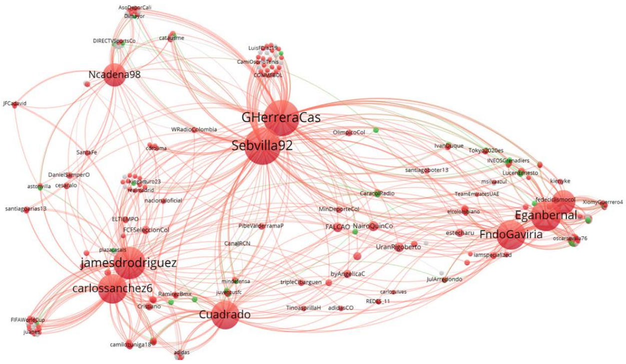

FIGURE 16. Polar layer network graph for the topic of sports.

TABLE 25. Frequencies of polar distributions by recalculation of sport clusters.

Distribución	Frecuencias
Weibull	2
Valor extremo generalizado	0
Uniforme	3
Beta	127
T-student	2
Normal	1
Gama logaritmo natural	1
Total	136

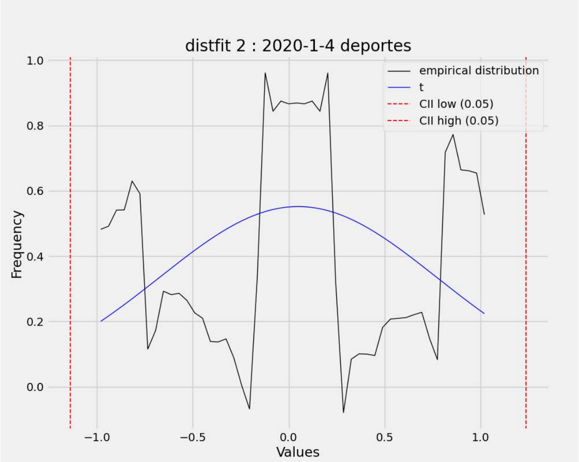

FIGURE 17. Polar distribution for the topic of spots.

in betweenness centrality by the account of the Minister of Sport @GHerreraCas, shown in the table 24.

Athletes @jamesrodriguez and @Eganbernal obtained third and fourth place in the three centrality metrics. Finally, the athlete @NairoQuinCo has an extensive reach to the remaining nodes of the network, but his betweenness and *all degree* value place him in tenth place; otherwise, for the athlete @carlossanchez6 located in tenth place in closeness centrality, he occupies higher positions in *all degree* and betweenness.

Fig. 16 shows a similar size in the nodes participating in the ranking above since the differences concerning the first *all degree* centrality value are lower values of 0.503. Additionally, the general feeling of Twitter users towards the top 10 athletes for the network is negative.

The statistics of the polar distributions for the topic of *sports* are presented in the table 25 over the entire time analyzed. There, the frequency increased each time the cluster groups by popularity were recalculated.

The table shows 136 recalculations for the *sports* topic where the Beta distribution was assigned 127 times, representing 93.38%. The Uniform, Weibull, and T-student distributions occupy the subsequent places presenting the same value for the last two mentioned.

In contrast to the previous issues, the distributions register a value different from zero, indicating the presence of a recalculation assigned from cluster to some distribution; however, the Beta distribution continues to obtain a great difference over the remaining distributions. The t-student distribution can be observed in Fig. 17.

F. LIFE AND LEISURE

Regarding the community of *life and leisure* (onion layer), the first two places are the accounts belonging to a writer and the director of the Bogota Philharmonic Orchestra, respectively, for the three metrics of closeness, degree, and betweenness (see table 26). In these accounts, a difference is observed between the values for the centrality of influence on information transfer, whose maximum value is 0.372. As for

TABLE 26. Structural metrics life and leisure network.

Closeness		All Degree		Betweenness	
Value	Label	Value	Label	Value	Label
0.6158	RSilvaRomero	204	RSilvaRomero	0.3722	RSilvaRomero
0.6074	davidgarciarod	196	davidgarciarod	0.3546	davidgarciarod
0.5	NoticiasCaracol	167	camilochara	0.255	camilochara
0.4972	Eganbernal	127	EseTonito	0.1651	EseTonito
0.4959	RevistaSemana	49	kikayis	0.0317	kikayis
0.4959	fdbedout	34	Orios8	0.0203	byfieldtravel
0.4959	fdbedout	34	Orios8	0.0203	byfieldtravel
0.4959	ELTIEMPO	29	byfieldtravel	0.0198	Orios8
0.4959	IvanDuque	10	Patoneando	0.0055	YouTube
0.4925	elespectador	8	AstrologiaCom	0.0045	Patoneando
0.4918	DanielSamperO	6	YouTube	0.0043	elespectador

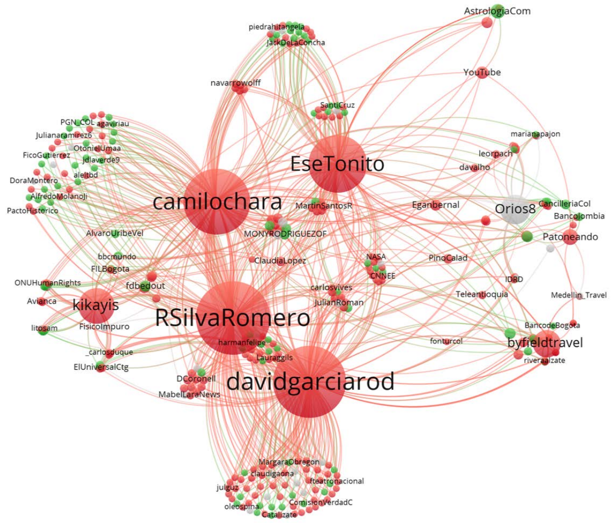

FIGURE 18. Polar layer network graph for the topic of life and leisure.

closeness centrality, which describes the distances between the other nodes, the whole column has a value close to 0.5.

Negative opinions predominate in the *life and leisure* topic. There are two clusters of nodes whose sizes are not very significant where the three types of polarity are found, located at the bottom of @davidgarciarod's node and on the left side of @camilochara's node. Compared to the topics described in previous sections, there is a node with a considerable *all degree* centrality with a neutral polarity, represented in a node size visible in the graph, corresponding to the account of the sports journalist @orios8.

The table 27 shows the statistics of the polar distributions for the topic of *life and leisure*; evidencing the increase in frequency each time the cluster groups by popularity were recalculated.

The table shows a total of 134 recalculations for the *life and leisure* topic, where the Beta distribution was 117 times, representing 87.31%, followed by the Weibull and Uniform distributions. Similar to the results with the topics of *culture*, *politics*, and *sports*, the Beta distribution obtains a significant difference over the remaining distributions; however, this topic presents the highest frequency so far in the Weibull distribution. The Weibull distribution can be seen in Fig. 19.

TABLE 27. Frequencies of polar distributions by recalculation of life and leisure clusters.

Distribución	Frecuencias
Weibull	14
Valor extremo generalizado	0
Uniforme	3
Beta	117
T-student	0
Normal	0
Gama logaritmo natural	0
Total	134

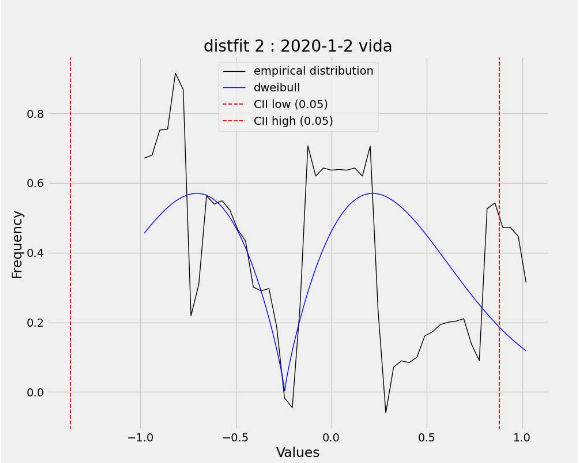

FIGURE 19. Polar distribution for the topic of life and leisure.

G. TECHNOLOGY

Regarding the *technology* community (onion layer), the director of appropriation of the TIC Ministry (@mafeardilalopez) presents the account with the highest value in all centrality metrics, as shown in the table 28. The above outlines the importance of the current government policies through the technology ministry to disseminate their policies on Twitter. As a result, they obtain an extensive reach to the users of this community (closeness), a high number of connections to reach the target populations (*all degree*), and a high closeness between the nodes as main actors that intermediate the information for the achievement of such policies (betweenness). In the same way, it is possible to infer the phenomenon of policy advocacy mentioned in the accounts of @MCarolina-HoyosT and @dgavalo, since these belong to state officers or ex-officers in the areas of communication or technology of their respective entities; however, the number of connections of @MCarolinaHoyosT places it in lower positions since it no longer holds its public position. Other accounts are essential in the ranking, such as @MauricioJaramil and @nmolano, which reach between 0.6073 and 0.5016.

On the other hand, the Fig. 20 presents an overall negative sentiment in the top 10 accounts for the topic of *technology*. Additionally, the size of the top six node locations differs from the rest, presenting a large difference in the metric of the *all degree* metric.

TABLE 28. Structural metrics technology network.

Closeness		All Degree		Betweenness	
Value	Label	Value	Label	Value	Label
0.6287	mafeardilalopez	395	mafeardilalopez	0.2719	mafeardilalopez
0.6176	MCarolinaHoyosT	367	dgavalo	0.2442	dgavalo
0.6151	dgavalo	351	nmolano	0.2183	MCarolinaHoyosT
0.6073	nmolano	303	CifuentesAura	0.2174	nmolano
0.5016	MauricioJaramil	295	MCarolinaHoyosT	0.1588	CifuentesAura
0.5016	RevistaSemana	205	ejramirezr	0.1243	ejramirezr
0.5016	BluRadioCo	60	mecheverry	0.0145	mecheverry
0.5016	ELTIEMPO	8	GPStrategyCO	0.0014	CaracolRadio
0.5016	Ministerio_TIC	7	MauricioJaramil	0.0011	GoogleColombia
0.5013	CaracolRadio	7	RevistaSemana	0.0009	MauricioJaramil

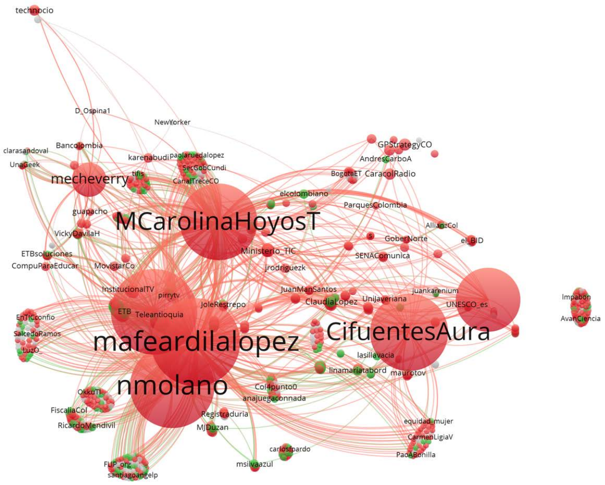

FIGURE 20. Polar layer network graph for the topic of technology.

TABLE 29. Frequencies of polar distributions by recalculation of technology clusters.

Distribución	Frecuencias
Weibull	6
Valor extremo generalizado	0
Uniforme	4
Beta	124
T-student	0
Normal	0
Gama logaritmo natural	0
Total	134

The statistics of the polar distributions for the topic of *technology* are presented in the table 29; evidencing the increase in frequency each time the cluster groups by popularity were recalculated.

The table shows 134 recalculations for the *technology* topic, where the Beta distribution was 124 times representing 92.53%, followed by the Weibull and Uniform distributions. Similar to the results with previous topics, the order of the top distributions is repeated for *technology* where the Beta distribution obtains a large difference over the remaining distributions. In Fig. 21 the Natural Logarithmic Range distribution can be observed.

H. ECONOMY

Regarding the *economy* community (onion layer), the density of this network is 0.008589; that is, it presents a small size.

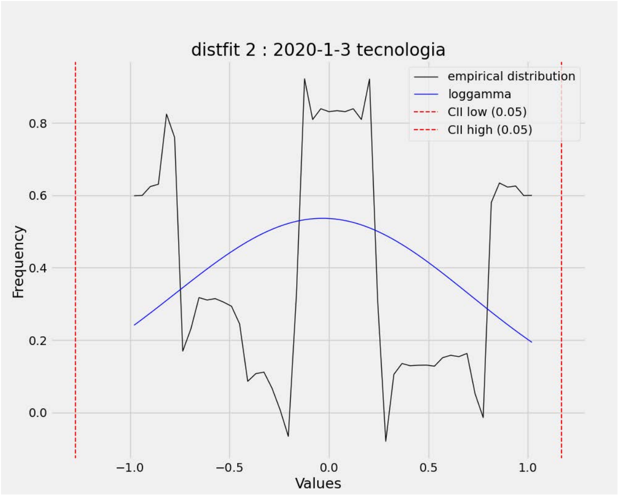

FIGURE 21. Polar distribution for the topic of technology.

TABLE 30. Structural metrics economy network.

Closeness		All degree		Betweenness	
Value	Label	Value	Label	Value	Label
0.5086	CamiloDeGuzman	250	CamiloDeGuzman	0.2615	CamiloDeGuzman
0.503	DNP_Colombia	228	DavidAlvaradoMu	0.2391	jagallegod
0.503	elespectador	216	jagallegod	0.1912	DavidAlvaradoMu
0.5026	ELTIEMPO	195	NicolasUribe	0.1829	NicolasUribe
0.5026	RevistaSemana	173	castellanosgd	0.1732	acocotero
0.5021	dgomezco	161	acocotero	0.1464	castellanosgd
0.5	agaviriau	148	adriana_guzman	0.0857	adriana_guzman
0.5	MauricioCard	126	amaldon19	0.0764	amaldon19
0.5	DeLaCalleHum	8	DNP_Colombia	0.005	DNP_Colombia
0.5	WRadioColombia	8	elespectador	0.005	elespectador

The table 30 shows the centrality metrics for the topic of *economy*, where the economist @CamiloDeGuzmán is leading in the three types of centrality metrics.

However, the values for these metrics are not very high since the centrality betweenness is below 30% for information intermediation, and the closeness to the nodes of the network only reaches 50%. Finally, most of the accounts in this ranking are individuals who may represent public or private entities, but only one state institution account (DNP) exists.

Fig. 22 does not present changes in the size of the top nodes of the *all degree* centrality metric. General negative sentiment is present in the community except for one node corresponding to the economist and researcher @acocotero.

The statistics of the polar distributions for the *economics* topic are presented in the table 31, evidencing the increase in frequency each time the cluster groups by popularity were recalculated.

The table shows a total of 132 recalculations for the *economics* topic, where the Beta distribution was assigned 107 times, representing 81.06%, followed by the Weibull and Uniform distributions. The Beta distribution obtains a great difference over the remaining distributions and the results with all the analyzed topics, while the Weibull distribution registers the highest frequency compared to the other topics. The Beta distribution can be seen in Fig. 23.

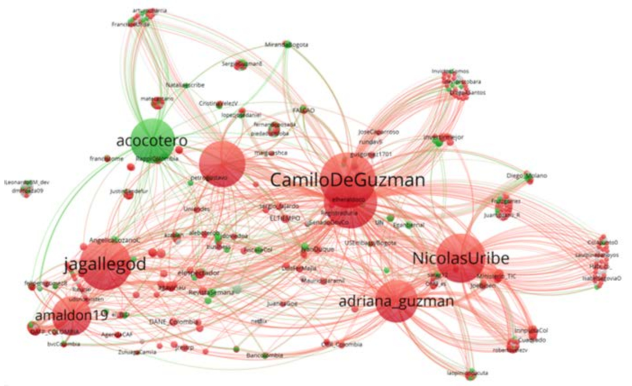

FIGURE 22. Polar layer network graph for the topic of economy.

TABLE 31. Frequencies of polar distributions by recalculation of economy clusters.

Distribución	Frecuencias
Weibull	18
Valor extremo generalizado	0
Uniforme	6
Beta	107
T-student	0
Normal	0
Gama logaritmo natural	1
Total	132

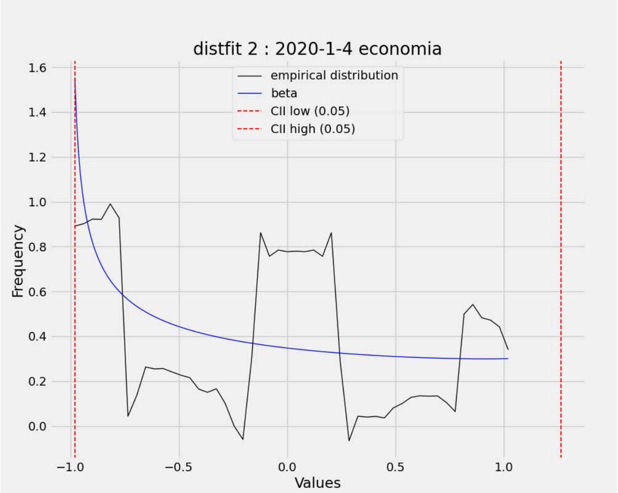

FIGURE 23. Polar distribution for the topic of economy.

V. CONCLUSION

One of the most significant features of the evolution of SNA has been the shift from structural analysis to content analysis. The mathematical and statistical methods that emerged between 1930 and the 1990s allowed sociologists, anthropologists, and researchers from other disciplines to enrich qualitative activities by developing increasingly sophisticated algorithms that seek to improve the precision of analysis in discursive contexts where pragmatic complexity increases.

The social network most widely used for this type of study is Twitter, a fact motivated by the availability of information and access to it through free APIs and for being the social network in which the number of characters imposes a series of conditions on linguistic expressions that allow greater control concerning other social networks, where the flow and amount of unstructured information are significantly higher.

The growing use of social networks, especially Twitter, by Internet users to express their opinions on a wide variety of topics has increased interest in the possibility of exploiting this information to understand their behavior based on public opinion. In this sense, the present research was based on the development of an alternative and novel method called *Collective Subjectivity Communities in Onion Layers (COSSOL)"* that would allow an analysis of collective subjectivity in the communities existing in the social network Twitter from the perspective of onion layers, providing greater granularity and detail in its analysis.

Design-based scientific research guides the methodological approach with three interlocking cycles: relevance, design, and rigor. The relevance cycle contributes to the analysis of private states in the framework of interactions in social networks, which are fundamental for the interaction of people and organizations since it is in this context where large volumes of information with diverse content are being generated and whose processing and analysis allows showing different patterns of behavior on the social dynamics analyzed.

The design cycle allowed the generation of the model for the analysis of collective subjectivity using Twitter data as the primary input. In contrast, the rigor cycle contributed to the consecutive and constant review of the theoretical, methodological, and structural contents of the SNA and the SA. It guarantees the quality in each of the stages of the project execution to know the literature relevant to the research advanced in the SNA and SA constructs, the identification of existing gaps to propose future areas of study, and the provision of a frame of reference that allows to appropriately position the research activities that correspond to the following phases of the process.

As for the generation of the COSSOL model, it was carried out within methodological cycle two, associated with the design. With this, it was possible to evaluate the users' behavior, recognizing that the structural links in common PAS modes are massively shared, called Collective Subjectivity. To perform such analysis, COSSOL took elements of the SNA and SA constructs and proposed a hybrid system of greater granularity in community analysis represented by onion layers to examine the levels of interaction of communities in social networks.

From the onion layers perspective, each level of analysis (from the most general to the most specific) is represented by a set of circles that, in turn, contain others; that is, a smaller circle represents a more specific level of analysis with more defined or higher granularity communities generating a disaggregation of the network.

In order to test the communities stability, the steps of the Engle-Granger methodology were executed to test the existence of long-term relationships. The first step tests all the communities since they are stationary in levels and in the first or second difference. The popularity clusters in the communities of *politics*, *technology* and *life and leisure* had a particular behavior that required the execution of such tests in order to prove stationarity, demonstrating the existence of cointegration in the errors of the relevant equations for polarity in the unpopular clusters of *politics* and popular clusters of *technology*. The same case was evidenced in the network density for the average popularity of *life and leisure* and in the popular ones of *technology*.

In this order of ideas, the existence of unit roots is evident as the onion ring becomes more granular; that is, ADF, PP, and KPSS tests in the first difference should be performed for the clusters of the communities in the third onion layer, except the *technology* topic. The cluster composed of the six topics with the highest popularity in levels presents unit root, which reflects a continuous change in the flow of sentiments posted on Twitter according to the events that occurred in the short term. On the other hand, there is a greater number of unit root problems in the polarity series in levels compared to its network density counterpart. Therefore, the SA metric fluctuates more over time once discussions are posted on Twitter.

Now, the results of the equations in the cointegration test for the second step of the Engle-Granger methodology demonstrated the highest coefficients of the explanatory variable at the popular and average popularity levels for the different topics, except *technology*; for example, the one unit increase in network density in the topic of *culture* caused the most significant increase in polarity units. In contrast, the levels of unpopularity in *sports* and *life and leisure* presented the lowest coefficients of the explanatory variables. Moreover, the latter phenomenon was present in the second onion layer for the topics of *sports* and *life and leisure*, where the increase caused by network density in *life and leisure* was the lowest.

As a result of the present research, we obtained the verification of 2 of the three hypotheses proposed; that is, both the second and third hypotheses were successfully demonstrated, which are associated with identifying more stable communities in terms of polarity that find highly connected members and communities with a higher density and a common polarity that better propagate their subjective expressions, respectively. However, regarding the first hypothesis, this was rejected since its logic only applied to the first onion ring, leaving the non-existence of long-term relationships in more granular layers for themes such as *technology* and *life and leisure*.

On the other hand, the second case of experimentation focused on the study of recalculations for the communities in their three onion layers, allowed concluding the inexistence of dissent represented in their polar distribution figures; that is, no bimodal polar distributions representing extremes of the common TAFs forms were evidenced. In addition, the descriptive statistics of the structural metrics for the constructed ecosystem point to the topics of *politics* and *life and leisure* as those of most significant interest for Colombians since they are the topics of greatest variation for each recalculation of their networks under the network density metric. Similarly, the all-degree centrality metric statistics show the topics of *economics* and *life and leisure* as those with the longest life cycles since their variations are the smallest.

REFERENCES

[1] L. Blackman, J. Cromby, D. Hook, D. Papadopoulos, and V. Walkerdine, "Creating subjectivities," *Subjectivity*, vol. 22, no. 1, pp. 1–27, May 2008.

[2] J. Domingues, *Sociological Theory and Collective Subjectivity*. London, U.K.: Palgrave Macmillan, 1995. [Online]. Available: https://books.google.com.co/books?id=X9aGDAAAQBAJ

[3] J. Domingues, *Social Creativity, Collective Subjectivity and Contemporary Modernity*. London, U.K.: Palgrave Macmillan, 2000. [Online]. Available: https://books.google.com.co/books?id=IfCFDAAAQBAJ

[4] P. Mika, "Social networks and the semantic web," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Sep. 2004, pp. 285–291, doi: 10.1109/WI.2004.10039.

[5] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994. [Online]. Available: http://books.google.se/books?id=CAm2DpIqRUIC

[6] E. C. Traugott and R. B. Dasher, *Regularity in Semantic Change*. Cambridge, U.K.: Cambridge Univ. Press, 2001. [Online]. Available: https://www.cambridge.org/core/books/regularity-in-semantic-change/F07CBB401A177975904C1E37BE0D9E07

[7] E. C. Traugott, "On the rise of epistemic meanings in English: An example of subjectification in semantic change," *Lang.*, vol. 65, no. 1, pp. 31–55, 1989.

[8] R. W. Langacker, *Grammar and Conceptualization*. Berlin, NY, USA: De Gruyter Mouton, 1999. [Online]. Available: https://books.google.com.co/books?id=a12FyDIEFgsC

[9] R. W. Langacker, *Cognitive Grammar: A Basic Introduction*. Oxford, U.K.: Oxford Univ. Press, 2008. [Online]. Available: https://books.google.com.co/books?id=UKVNKz0ZRqwC

[10] L. G. Moreno-Sandoval, A. Pomares-Quimbaya, and J. A. Alvarado-Valencia, "Celebrity profiling through linguistic analysis of digital social networks," *Comput. Social Netw.*, vol. 8, no. 1, pp. 75–105, Dec. 2021, doi: 10.1186/s40649-021-00097-w.

[11] R. H. Von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quart.*, vol. 28, no. 1, pp. 75–105, Mar. 2004.

[12] O. Bodin and C. Prell, *Social Networks and Natural Resource Management: Uncovering the Social Fabric of Environmental Governance*, 2nd ed. Cambridge, MA, USA: Cambridge Univ. Press, 2011, pp. 1–375. [Online]. Available: https://books.google.com.co/books?id=uClj6Heel5gC

[13] J. Scott, "Social network analysis: Developments, advances, and prospects," *Social Netw. Anal. Mining*, vol. 1, no. 1, pp. 21–26, 2011, doi: 10.1007/s13278-010-0012-6.

[14] L. C. Freeman, *Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, BC, Canada: ΣP Empirical Press, 2004. [Online]. Available: https://books.google.com.co/books/about/The_Development_of_Social_Network_Analys.html?id=VcxqQgAACAAJ&redir_esc=y

[15] C. Aggarwal, *Social Network Data Analytics*, 1st ed. New York, NY, USA: Springer, 2011, p. 502, doi: 10.1007/978-1-4419-8462-3.

[16] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social set analysis: A set theoretical approach to Big Data analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016, doi: 10.1109/ACCESS.2016.2559584.

[17] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas. - IMC*, 2007, pp. 29–42, doi: 10.1145/1298306.1298311.

[18] M. Zafar, P. Bhattacharya, N. Ganguly, K. Gummadi, and S. Ghosh, "Sampling content from online social networks: Comparing random vs. expert sampling of the Twitter stream," *ACM Trans. Web*, vol. 9, no. 3, pp. 1–33, Jun. 2015, doi: 10.1145/2743023.

[19] R. Zafarani, M. Abbasi, and H. Liu, *Social Media Mining: An Introduction*, Cambridge, MA, USA: Cambridge Univ. Press, 2014, pp. 1–382.

[20] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 1994. Cambridge, U.K.: Cambridge Univ. Press, 2004, p. 825. [Online]. Available: http://books.google.se/books?id=CAm2DpIqRUIC

[21] H. C. White, S. A. Boorman, and R. L. Breiger, "Social structure from multiple Networks. I. blockmodels of roles and positions," *Amer. J. Sociol.*, vol. 81, no. 4, pp. 730–780, Jan. 1976.

[22] W. De Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge, U.K.: Cambridge Univ. Press, 1976.

[23] Ö. Bodin and B. I. Crona, "The role of social networks in natural resource governance: What relational patterns make a difference?" *Global Environ. Change*, vol. 19, no. 3, pp. 366–374, Aug. 2009.

[24] P. J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005, doi: 10.2277/0521809592.

[25] D. Nguyen, A. S. Dogruöz, C. P. Rosé, and F. de Jong, "Computational sociolinguistics: A survey," *Comput. Linguistics*, vol. 42, pp. 537–593, Sep. 2016, doi: 10.1162/COLI_a_00258.

[26] H. Zhang, D. Nguyen, H. Zhang, and M. Thai, "Least cost influence maximization across multiple social networks," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1–11, Mar. 2015, doi: 10.1109/TNET.2015.2394793.

[27] A. Perer and B. Shneiderman, "Balancing systematic and flexible exploration of social networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 693–700, Nov. 2006, doi: 10.1109/TVCG.2006.122.

[28] B. Shneiderman and A. Aris, "Network visualization by semantic substrates," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 733–740, Sep. 2006, doi: 10.1109/TVCG.2006.166.

[29] B. Shneiderman and C. Dunne, "Interactive network exploration to derive insights: Filtering, clustering, grouping, and simplification," *Graph Drawing* (Lecture Notes in Computer Science), vol. 7704. Berlin, Germany: Springer, 2012, pp. 2–18, doi: 10.1007/978-3-642-36763-2_2.

[30] S. S. Bodrunova, A. A. Litvinenko, and I. S. Blekanov, "Please Follow Us," *J. Pract.*, vol. 12, no. 2, pp. 1–27, 2017, doi: 10.1080/17512786.2017.1394208.

[31] F. T. O'Donovan, C. Fournelle, S. Gaffigan, O. Brdiczka, J. Shen, J. Liu, and K. E. Moore, "Characterizing user behavior and information propagation on a social multimedia network," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2013, pp. 1–6, doi: 10.1109/ICMEW.2013.6618395.

[32] X. Zhou, B. Wu, and Q. Jin, "Analysis of user network and correlation for community discovery based on topic-aware similarity and behavioral influence," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 6, pp. 559–571, Dec. 2017, doi: 10.1109/THMS.2017.2725341.

[33] K. H. Lim and A. Datta, "An interaction-based approach to detecting highly interactive Twitter communities using tweeting links," *Book Web Intell.*, vol. 14, no. 1, pp. 1–15, 2016, doi: 10.3233/WEB-160328.

[34] Y. R. Lin and D. Margolin, "The ripple of fear, sympathy and solidarity during the Boston bombings," *EPJ Data Sci.*, vol. 3, no. 1, pp. 1–28, doi: 10.1140/epjds/s13688-014-0031-z.

[35] M. M. D. Khomami, A. Rezvanian, and M. R. Meybodi, "Distributed learning automata-based algorithm for community detection in complex networks," *Int. J. Modern Phys. B*, vol. 30, pp. 1–20, Mar. 2016, doi: 10.1142/S0217979216500429.

[36] S. J. Park, Y. S. Lim, and H. W. Park, "Comparing Twitter and YouTube networks in information diffusion: The case of the 'occupy wall street' movement," *Technol. Forecasting Social Change*, vol. 95, pp. 208–217, Jun. 2015, doi: 10.1016/j.techfore.2015.02.003.

[37] D. Murthy and J. P. Lewis, "Social media, collaboration, and scientific organizations," *Amer. Behav. Sci.*, vol. 59, no. 1, p. 149, 2015, doi: 10.1177/0002764214540504.

[38] C. Song, W. Hsu, and M. L. Lee, "Mining brokers in dynamic social networks," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2015, pp. 523–532, doi: 10.1145/2806416.2806468.

[39] A. Sowriraghavan and P. Burnap, "Prediction of malware propagation and links within communities in social media based events," in *Proc. ACM Web Sci. Conf. ZZZ (WebSci)*, 2015, pp. 1–2, doi: 10.1145/2786451.2786494.

[40] I. Himelboim and J. Y. Han, "Cancer talk on twitter: Community structure and information sources in breast and prostate cancer social networks," *J. Health Commun.*, vol. 19, no. 2, pp. 210–225, 2014, doi: 10.1080/10810730.2013.811321.

[41] J. Valverde-Rebaza and A. de Andrade Lopes, "Exploiting behaviors of communities of Twitter users for link prediction," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1063–1074, Dec. 2013, doi: 10.1007/s13278-013-0142-8.

[42] S. Y. Bhat and M. Abulaish, "Overlapping social network communities and viral marketing," in *Proc. Int. Symp. Comput. Bus. Intell. (ISCBI)*, Aug. 2013, pp. 243–246, doi: 10.1109/ISCBI.2013.56.

[43] Y. Tyshchuk, H. Li, H. Ji, and W. A. Wallace, "Evolution of communities on Twitter and the role of their leaders during emergencies," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2013, pp. 727–733, doi: 10.1145/2492517.2492657.

[44] S. Myneni, N. K. Cobb, and T. Cohen, "Finding meaning in social media: Content-based social network analysis of QuitNet to identify new opportunities for health promotion," *Stud. Health Technol. Informat.*, vol. 192, pp. 807–811, Jan. 2013, doi: 10.3233/978-1-61499-289-9-807.

[45] M. Thangaraj and V. T. Meenatchi, "Applying prefetching in Online Social Network to gain social intelligence," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2015, pp. 8–11, doi: 10.1109/ICCCI.2015.7218066.

[46] L. G. Moreno-Sandoval and L. M. Pantoja-Rojas, "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)," in *Proc. 21st Int. Conf. Enterprise Inf. Syst.*, 2019, pp. 488–495, doi: 10.5220/0007770804880495.

[47] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham, "Uncovering the wider structure of extreme right communities spanning popular online networks," in *Proc. 5th Annu. ACM Web Sci. Conf. (WebSci)*, 2013, pp. 276–285, doi: 10.1145/2464464.2464495.

[48] A. Utengen, D. Rouholiman, J. G. Gamble, F. J. Grajales, N. Pradhan, A. C. Staley, L. Bernstein, S. D. Young, K. A. Clauson, and L. F. Chu, "Patient participation at health care conferences: Engaged patients increase information flow, expand propagation, and deepen engagement in the conversation of tweets compared to physicians or researchers," *J. Med. Internet Res.*, vol. 19, no. 8, pp. 1–11, 2017, doi: 10.2196/jmir.8049.

[49] N. Alrajebah, M. Luczak-roesch, and T. Tiropanis, "Deconstructing diffusion on tumblr: Structural and temporal aspects," in *Proc. ACM Web Sci. Conf.*, 2017, pp. 319–328, doi: 10.1145/3091478.3091491.

[50] Z. Nasim and Q. Rajput, "Understanding role of Twitter in addressing social causes," in *Proc. Int. Conf. Innov. Electr. Eng. Comput. Technol. (ICIEECT)*, Apr. 2017, pp. 1–9.

[51] Z. Shoroye, W. Yaqub, A. A. Mohammed, Z. Aung, and D. Svetinovic, "Exploring social contagion in open-source communities by mining software repositories," in *Neural Information Processing* (Lecture Notes in Computer Science), vol. 9492. Cham, Germany: Springer, 2015, doi: 10.1007/978-3-319-26561-2_15.

[52] A. V. Mantzaris, "Uncovering nodes that spread information between communities in social networks," *EPJ Data Sci.*, vol. 3, no. 1, pp. 1–17, 2014, doi: 10.1140/epjds/s13688-014-0026-9.

[53] A. V. Kaiserx, J. Kröckel, and F. Bodendorf, "Simulating the spread of opinions in online social networks when targeting opinion leaders," *Inf. Syst. e-Business Manage.*, vol. 11, no. 4, pp. 597–621, 2013, doi: 10.1007/s10257-012-0210-z.

[54] L. A. Overbey, B. Greco, C. Paribello, and T. Jackson, "Structure and prominence in Twitter networks centered on contentious politics," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1351–1378, 2013, doi: 10.1007/s13278-013-0134-8.

[55] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield, "Analysing the connectivity and communication of suicidal users on Twitter," *Comput. Commun.*, vol. 73, pp. 291–300, Jan. 2016, doi: 10.1016/j.comcom.2015.07.018.

[56] A. Angadi and P. Suresh Varma, "Finding hubs and outliers in temporal networks," *Indian J. Sci. Technol.*, vol. 9, no. 20, pp. 1–5, 2016, doi: 10.17485/ijst/2016/v9i20/78483.

[57] K. H. Chu, J. B. Unger, J. P. Allem, M. Pattarroyo, D. Soto, T. B. Cruz, H. Yang, L. Jiang, and C. C. Yang, "Diffusion of messages from an electronic cigarette brand to potential users through Twitter," *PLoS ONE*, vol. 10, Dec. 2015, Art. no. e0145387, doi: 10.1371/journal.pone.0145387.

[58] E. Stattner, R. Eugenie, and M. Collard, "How do we spread on Twitter?" in *Proc. Int. Conf. Res. Challenges Inf. Sci.*, 2015, pp. 334–341, doi: 10.1109/RCIS.2015.7128894.

[59] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "Efficiency of human activity on information spreading on Twitter," *Social Netw.*, vol. 39, no. 1, pp. 1–11, 2014, doi: 10.1016/j.socnet.2014.03.007.

[60] D. Meng, L. Wan, and L. Zhang, ''A study of rumor spreading with epidemic model based on network topology,'' in *Trends and Applications in Knowledge Discovery and Data Mining*. Cham, Germany: Springer, 2014, doi: 10.1007/978-3-319-13186-3_35.

[61] M. Cataldi, L. Di Caro, and C. Schifanella, ''Personalized emerging topic detection based on a term aging model,'' *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, pp. 1–27, 2013, doi: 10.1145/2542182.2542189.

[62] I. B. Arpinar, U. Kursuncu, and D. Achilov, ''Social media analytics to identify and counter Islamist extremism: Systematic detection, evaluation, and challenging of extremist narratives online,'' in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, 2016, pp. 611–612, doi: 10.1109/CTS.2016.0113.

[63] I. Eleta and J. Golbeck, ''Multilingual use of Twitter: Social networks at the language frontier,'' *Comput. Hum. Behav.*, vol. 41, pp. 424–432, Dec. 2014, doi: 10.1016/j.chb.2014.05.005.

[64] G. Manju and T. V. Geetha, ''Concept similarity based academic tweet community detection using label propagation,'' in *Mining Intelligence and Knowledge Exploration* (Lecture Notes in Computer Science), vol. 8284. Cham, Germany: Springer, 2013, pp. 677–686, doi: 10.1007/978-3-319-03844-5_66.

[65] E. Lozano and C. Vaca, ''Crisis management on Twitter: Detecting emerging leaders,'' in *Proc. 4th Int. Conf. eDemocracy eGovernment (ICEDEG)*, Apr. 2017, pp. 140–147, doi: 10.1109/ICEDEG.2017.7962524.

[66] Y. G. Rykov, P. A. Meylakhs, and Y. E. Sinyavskaya, ''Network structure of an AIDS-denialist online community: Identifying core members and the risk group,'' *Amer. Behav. Scientist*, vol. 61, no. 7, pp. 688–706, 2017, doi: 10.1177/0002764217717565.

[67] A. M. Litterio, E. A. Nantes, J. M. Larrosa, and L. J. Gómez, ''Marketing and social networks: A criterion for detecting opinion leaders,'' *Eur. J. Manage. Bus. Econ.*, vol. 26, no. 3, pp. 347–366, 2017, doi: 10.1108/EJMBE-10-2017-020.

[68] G. Coronel-salas and C. M. Sanmartín, ''Twitter profile analysis of the institutions of science and technology in Ibero-America,'' in *Proc. 11th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2016, pp. 1–6, doi: 10.1109/CISTI.2016.7521588.

[69] T. Munger and J. Zhao, ''Identifying influential users in on-line support forums using topical expertise and social network analysis,'' in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 721–728, doi: 10.1145/2808797.2810059.

[70] Y. Li, X. Wu, and L. Li, ''Community influence analysis based on social network structures,'' in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Dec. 2015, pp. 247–254, doi: 10.1109/SmartCity.2015.79.

[71] C. Wukich and I. Mergel, ''Closing the citizen-government communication gap: Content, audience, and network analysis of government Tweets,'' *J. Homeland Secur. Emergency Manage.*, vol. 12, no. 3, pp. 707–735, 2015, doi: 10.1515/jhsem-2014-0074.

[72] T. Leung and F. L. Chung, ''Persuasion driven influence propagation in social networks,'' in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2014, pp. 548–554, doi: 10.1109/ASONAM.2014.6921640.

[73] J. Al-Sharawneh, S. Sinnappan, and M. A. Williams, ''Credibility-based Twitter social network analysis,'' in *Web Technologies and Applications* (Lecture Notes in Computer Science), vol. 7808. Berlin, Germany: Springer, 2013, doi: 10.1007/978-3-642-37401-2_33.

[74] R. R. M. Daga, ''Social network analysis of Tweets on Typhoon during Haiyan and Hagupit,'' in *Proc. 8th Int. Conf. Comput. Modeling Simul.*, 2017, pp. 151–154, doi: 10.1145/3036331.3036345.

[75] R. Cazabet, H. Takeda, and M. Hamasaki, ''Characterizing the nature of interactions for cooperative creation in online social networks,'' *Social Netw. Anal. Mining*, vol. 5, no. 1, pp. 1–17, 2015, doi: 10.1007/s13278-015-0284-y.

[76] P. Wadhwa and M. P. S. Bhatia, ''New metrics for dynamic analysis of online radicalization,'' *J. Appl. Secur. Res.*, vol. 11, no. 2, pp. 166–184, 2016, doi: 10.1080/19361610.2016.1137203.

[77] E. Benveniste, ''De la subjectivité dans le langage,'' *J. de Psychologie*, vol. 55, 1958.

[78] A. Banfield, *Describing the Unobserved: Events Grouped Around an Empty Centre*. Newcastle upon Tyne, U.K.: Cambridge Scholars Publishing, 1987, pp. 105–128, ch. 4.

[79] A. Banfield, *Unspeakable Sentences (Routledge Revivals): Narration and Representation in the Language of Fiction*. London, U.K.: Taylor & Francis, 2014. [Online]. Available: https://books.google.com.co/books?id=SGgKBAAAQBAJ

[80] C. O. Alm, ''Subjective natural language problems: Motivations, applications, characterizations, and implications,'' in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol. (ACL-HLT)*, vol. 2, 2011, pp. 107–112. [Online]. Available: https://books.google.com.co/books?id=SGgKBAAAQBAJ

[81] E. Benveniste, ''Subjectivity in language,'' *Problems in General Linguistics*, vol. 1. Oxford, OH, USA: Univ. of Miami Press, 1971, pp. 223–230.

[82] A. Banfield, ''Narrative style and the grammar of direct and indirect speech,'' *Found. Lang., JSTOR*, vol. 10, n. 1, pp. 1–39, 1973.

[83] M. B. H. Everaert, M. A. C. Huybregts, N. Chomsky, R. C. Berwick, and J. J. Bolhuis, ''Structures, not strings: Linguistics as part of the cognitive sciences,'' *Trends Cognit. Sci.*, vol. 19, no. 12, pp. 729–743, Dec. 2015, doi: 10.1016/j.tics.2015.09.008.

[84] M. D. Hauser, N. Chomsky, and W. T. Fitch, ''The faculty of language: What is it, who has it, and how did it evolve?'' *Science*, vol. 298, no. 5598, pp. 1569–1579, Nov. 2002, doi: 10.1126/science.298.5598.1569.

[85] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, ''Introduction to WordNet: An on-line lexical database,'' *Int. J. Lexicogr.*, vol. 3, no. 4, pp. 235–244, 1990, doi: 10.1093/ijl/3.4.235.

[86] J. Wiebe, ''Tracking point of view in narrative,'' *Comput. Linguistics*, vol. 20, no. 2, pp. 233–287, 1994.

[87] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, ''Multilingual subjectivity analysis using machine translation,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2008, pp. 127–135. [Online]. Available: https://aclanthology.org/D08-1014/

[88] K. Ravi and V. Ravi, ''A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,'' *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015, doi: 10.1016/j.knosys.2015.06.015.

[89] D. Jurafsky and J. H. Martin, ''Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition,'' *Comput. Linguistics*, vol. 26, no. 4, pp. 638–641, 2007, doi: 10.1162/089120100750105975.

[90] J. P. Carvalho, H. Rosa, G. Brogueira, and F. Batista, ''MISNIS: An intelligent platform for Twitter topic mining,'' *Expert Syst. Appl.*, vol. 89, pp. 374–388, 2017, doi: 10.1016/j.eswa.2017.08.001.

[91] D. Antonakaki, D. Spiliotopoulos, C. V. Samaras, P. Pratikakis, S. Ioannidis, and P. Fragopoulou, ''Social media analysis during political turbulence,'' *PLoS ONE*, vol. 12, no. 10, pp. 1–14, 2017, doi: 10.1371/journal.pone.0186836.

[92] M. Hajjem and C. Latiri, ''Combining IR and LDA topic modeling for filtering microblogs,'' *Proc. Comput. Sci.*, vol. 112, pp. 761–770, Jan. 2017, doi: 10.1016/j.procs.2017.08.166.

[93] R. Chatterjee and S. Agarwal, ''Twitter truths: Authenticating analysis of information credibility,'' in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 2352–2357.

[94] D. T. Vo, V. T. Hai, and C. Y. Ock, ''Exploiting language models to classify events from Twitter,'' *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–11, Sep. 2015, doi: 10.1155/2015/401024.

[95] L. Chen, C. Zhang, and C. Wilson, ''Tweeting under pressure: Analyzing trending topics and evolving word choice on Sina Weibo,'' in *Proc. 1st ACM Conf. Online social Netw. (COSN)*, 2013, pp. 89–100, doi: 10.1145/2512938.2512940.

[96] F. Gemci and K. A. Peker, ''Extracting Turkish tweet topics using LDA,'' in *Proc. 8th Int. Conf. Electr. Electron. Eng. (ELECO)*, Nov. 2013, pp. 531–534, doi: 10.1109/ELECO.2013.6713899.

[97] P. Saleiro, E. M. Rodrigues, C. Soares, and E. Oliveira, ''TexRep: A text mining framework for online reputation monitoring,'' *New Gener. Comput.*, vol. 35, no. 4, pp. 365–389, 2017, doi: 10.1007/s00354-017-0021-3.

[98] E. Fersini, E. Messina, and F. A. Pozzi, ''Earthquake management: A decision support system based on natural language processing,'' *J. Ambient Intell. Hum. Comput.*, vol. 8, no. 1, pp. 37–45, 2017, doi: 10.1007/s12652-016-0373-4.

[99] M. Shalaby and A. Rafea, ''Identifying the topic-specific influential users using SLM,'' in *Proc. 1st Int. Conf. Arabic Comput. Linguistics, Adv. Arabic Comput. Linguistics (ACLing)*, 2016, pp. 118–123, doi: 10.1109/ACLing.2015.24.

[100] R. Muppalla, M. Miller, T. Banerjee, and W. Romine, ''Discovering explanatory models to identify relevant tweets on Zika,'' in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1194–1197, doi: 10.1109/EMBC.2017.8037044.

[101] C. Lipizzi, D. G. Dessavre, L. Iandoli, and J. E. R. Marquez, ''Towards computational discourse analysis: A methodology for mining Twitter backchanneling conversations,'' *Comput. Hum. Behav.*, vol. 64, pp. 782–792, Nov. 2016, doi: 10.1016/j.chb.2016.07.030.

[102] D. Ulloa, P. Saleiro, R. J. F. Rossetti, and E. R. Silva, ''Mining social media for open innovation in transportation systems,'' in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 169–174, doi: 10.1109/ITSC.2016.7795549.

[103] M. A. M. Raja and S. Swamynathan, ''Tweet sentiment analyzer: Sentiment score estimation method for assessing the value of opinions in Tweets,'' in *Proc. Int. Conf. Adv. Inf. Commun. Technol. Comput.*, 2016, p. 83, doi: 10.1145/2979779.2979862.

[104] B. Peng, J. Li, J. Chen, X. Han, R. Xu, and K. F. Wong, ''Trending sentiment-topic detection on Twitter,'' in *Computational Linguistics and Intelligent Text Processing* (Lecture Notes in Computer Science), vol. 9042. Cham, Germany: Springer, 2015, pp. 66–77, doi: 10.1007/978-3-319-18117-2_5.

[105] W. Chen, J. Wang, Y. Zhang, H. Yan, and X. Li, ''User based aggregation for biterm topic model,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 2, 2015, pp. 489–494.

[106] H. Alnegheimish, J. Alshobaili, N. AlMansour, R. B. Shiha, N. Al Twairesh, and S. Alhumoud, ''AraSenTi-lexicon: A different approach,'' in *Social Computing and Social Media. Applications and Analytics* (Lecture Notes in Computer Science), vol. 10283. Cham, Germany: Springer, 2017, pp. 226–235, doi: 10.1007/978-3-319-58562-8_18.

[107] U. Yaqub, S. A. Chun, V. Atluri, and J. Vaidya, ''Analysis of political discourse on Twitter in the context of the 2016 U.S. presidential elections,'' *Government Inf. Quart.*, vol. 34, no. 4, pp. 613–626, 2017, doi: 10.1016/j.giq.2017.11.001.

[108] E. Shabunina, S. Marrara, and G. Pasi, ''An approach to analyse a hashtag-based topic thread in Twitter,'' in *Natural Language Processing and Information Systems* (Lecture Notes in Computer Science), vol. 9612. Cham, Germany: Springer, 2016, pp. 350–358, doi: 10.1007/978-3-319-41754-7_34.

[109] E. Ferrara and Z. Yang, ''Quantifying the effect of sentiment on information diffusion in social media,'' *PeerJ Comput. Sci.*, vol. 1, pp. 1–15, Sep. 2015, doi: 10.7717/peerj-cs.26.

[110] M. Jenders, G. Kasneci, and F. Naumann, ''Analyzing and predicting viral Tweets,'' in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 657–664, doi: 10.1145/2487788.2488017.

[111] Y. Pratama and P. Ratno, ''The addition symptoms parameter on sentiment analysis to measure public health concerns,'' *Telkomnika*, vol. 15, no. 3, pp. 1301–1309, 2017, doi: 10.12928/TELKOMNIKA.v15i3.4711.

[112] G. Apoorva, N. R. Vaishnav, E. D. Chowdary, and C. Uddagiri, ''An approach to sentiment analysis in Twitter using expert Tweets and retweeting hierarchy,'' in *Proc. Int. Conf. Microelectron., Comput. Commun. (MicroCom)*, Jan. 2016, pp. 1–8, doi: 10.1109/MicroCom.2016.7522482.

[113] N. Al-Twairesh, H. Al-Khalifa, and A. AlSalman, ''AraSenTi: Large-scale Twitter-specific Arabic sentiment lexicons,'' in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 697–705.

[114] L. Lin, J. Li, R. Zhang, W. Yu, and C. Sun, ''Opinion mining and sentiment analysis in social networks: A retweeting structure-aware approach,'' in *Proc. IEEE/ACM 7th Int. Conf. Utility Cloud Comput. (UCC)*, Dec. 2014, pp. 890–895, doi: 10.1109/UCC.2014.145.

[115] G. Sá, T. Silveira, R. Chaves, F. Teixeira, F. Mourão, and L. Rocha, ''LEGi: Context-aware lexicon consolidation by graph inspection,'' in *Proc. ACM Symp. Appl. Comput.*, 2014, pp. 302–307, doi: 10.1145/2554850.2554916.

[116] A. Dang, M. Smit, A. Moh'd, R. Minghim, and E. Milios, ''Toward understanding how users respond to rumours in social media,'' in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 777–784, doi: 10.1109/ASONAM.2016.7752326.

[117] A. Kanavos, I. Perikos, P. Vikatos, I. Hatzilygeroudis, C. Makris, and A. Tsakalidis, ''Modeling ReTweet diffusion using emotional content,'' *IFIP Adv. Inf. Commun. Technol.*, vol. 436, pp. 101–110, 2014.

[118] E. Fersini, F. A. Pozzi, and E. Messina, ''Approval network: A novel approach for sentiment analysis in social networks,'' *World Wide Web*, vol. 20, no. 4, pp. 831–854, 2017, doi: 10.1007/s11280-016-0419-8.

[119] J. Jotheeswaran and K. Seerangan, Mining medical opinions using hybrid genetic algorithm—Neural network,'' *J. Med. Imag. Health Informat.*, vol. 6, no. 8, pp. 1925–1928, 2016, doi: 10.1166/jmihi.2016.1950.

[120] J. Tang and A. Fond, ''Sentiment diffusion in large scale social networks,'' in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2013, pp. 244–245, doi: 10.1109/ICCE.2013.6486878.

[121] T. Cu, H. Schneider, and J. Van Scotter, ''New product diffusion: The role of sentiment content,'' in *Proc. ACM SIGMIS Conf. Comput. People Res. (SIGMIS-CPR)*, 2016, pp. 149–155, doi: 10.1145/2890602.2890627.

[122] G. Nunes, D. Lopes, and Z. Abdelouahab, ''Opinion analysis applied to politics: A case study based on Twitter,'' in *Proc. 3rd Annu. Int. Symp. Inf. Manage. Big Data*, 2016, pp. 35–42.

[123] M. Bouazizi and T. Ohtsuki, ''Opinion mining in Twitter how to make use of sarcasm to enhance sentiment analysis,'' in *Proc. IEEE ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 1594–1597, doi: 10.1145/2808797.2809350.

[124] S. S. Minab, M. Jalali, and M. H. Moattar, ''A new sentiment classification method based on hybrid classification in Twitter,'' in *Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK)*, Nov. 2015, pp. 295–298, doi: 10.1109/ICTCK.2015.7582685.

[125] P. Barnaghi, P. Ghaffari, and J. G. Breslin, ''Opinion mining and sentiment polarity on Twitter and correlation between events and sentiment,'' in *Proc. IEEE 2nd Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2016, pp. 52–57, doi: 10.1109/BigDataService.2016.36.

[126] M.-C. Yang, J.-T. Lee, and H. C. Rim, ''Using link analysis to discover interesting messages spread across Twitter,'' in *Proc. Workshop Graph-Based Methods Natural Lang. Process.*, 2012, pp. 15–19.

[127] M. Chong, ''Sentiment analysis and topic extraction of the Twitter network of #prayforparis,'' in *Proc. Assoc. Inf. Sci. Technol.*, vol. 53, no. 1, pp. 1–4, 2016, doi: 10.1002/pra2.2016.14505301133.

[128] J. Lee, B. A. Rehman, M. Agrawal, and H. R. Rao, ''Sentiment analysis of Twitter users over time: The case of the Boston bombing tragedy,'' in *E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life* (Lecture Notes in Business Information Processing), vol. 258. Cham, Germany: Springer, 2016, pp. 1–14, doi: 10.1007/978-3-319-45408-5_1.

[129] J. Kim and J. Yoo, ''Role of sentiment in message propagation: Reply vs. retweet behavior in political communication,'' in *Proc. ASE Int. Conf. Social Informat. (SocialInformatics)*, 2012, pp. 131–136, doi: 10.1109/SocialInformatics.2012.33.

[130] X. Liu, K. Tang, J. Hancock, J. Han, M. Song, R. Xu, and B. Pokorny, ''A text cube approach to human, social and cultural behavior in the Twitter stream,'' in *Social Computing, Behavioral-Cultural Modeling and Prediction* (Lecture Notes in Computer Science), vol. 7812. Berlin, Germany: Springer, 2013, pp. 321–330, doi: 10.1007/978-3-642-37210-0_35.

[131] X. Hu, L. Tang, J. Tang, and H. Liu, ''Exploiting social relations for sentiment analysis in microblogging,'' in *Proc. 6th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2013, pp. 537–546, doi: 10.1145/2433396.2433465.

[132] R. Archana and S. Chitrakala, ''Explicit sarcasm handling in emotion level computation of tweets—A big data approach,'' in *Proc. 2nd Int. Conf. Comput. Commun. Technol. (ICCCT)*, Feb. 2017, pp. 106–110, doi: 10.1109/ICCCT2.2017.7972260.

[133] E. V. Epure, R. Deneckere, and C. Salinesi, ''Analyzing perceived intentions of public health-related communication on Twitter,'' in *Artificial Intelligence in Medicine* (Lecture Notes in Computer Science), vol. 10259. Cham, Germany: Springer, 2017, pp. 182–192, doi: 10.1007978-3-319-59758-4_19.

[134] S. Agarwal and A. Sureka, ''Investigating the role of Twitter in E-governance by extracting information on citizen complaints and grievances reports,'' in *Big Data Analytics* (Lecture Notes in Computer Science), vol. 10721. Cham, Germany: Springer, 2017, pp. 300–310, doi: 10.1007/978-3-319-72413-3_21.

[135] A. Rahimi, T. Cohn, and T. Baldwin, ''Twitter user geolocation using a unified text and network prediction model,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 630–636.

[136] D. Zhou, L. Chen, X. Zhang, and Y. He, ''Unsupervised event exploration from social text streams,'' *Intell. Data Anal.*, vol. 21, no. 4, pp. 849–866, 2017, doi: 10.3233/IDA-160048.

[137] J. Cuzzola, D. Gasevic, and E. Bagheri, ''Product centric web page segmentation and localization,'' in *Proc. 4th Can. Semantic Web Symp.*, vol. 1054, 2013, pp. 29–32.

[138] L. G. M. Sandoval, E. Puertas, A. P. Quimbaya, and J. Alvarado, ''Assembly of polarity, emotion and user statistics for detection of fake profiles,'' in *Notebook for PAN at CLEF*, vol. 2696. Aachen, Germany: CEUR-WS, 2020.

[139] L. G. M. Sandoval, L. Gabriel, E. Puertas, F. M. Plaza-Del-Arco, A. P. Quimbaya, J. Alvarado, and L. A. Ureña-López, ''Celebrity profiling on Twitter using sociolinguistic features,'' in *Notebook for PAN at CLEF*, vol. 2380. Aachen, Germany: CEUR-WS, 2019.

[140] T. Praveen, K. Karthick, M. Thapasya, and S. S. Preethika, "FlierMeet: An extension to online social networking site (OSNs)," *IIOAB J.*, vol. 7, pp. 419–429, Aug. 2016.

[141] J. Li, Z. Wei, H. Wei, K. Zhao, J. Chen, and K. F. Wong, "Learning to rank microblog posts for real-time ad-hoc search," in *Natural Language Processing and Chinese Computing* (Lecture Notes in Computer Science). Cham, Germany: Springer, 2015, pp. 436–443.

[142] S. Itokawa, S. Shiramatsu, T. Ozono, and T. Shintani, "Estimating feature terms for supporting exploratory browsing of Twitter timelines," in *Proc. 2nd IIAI Int. Conf. Adv. Appl. Informat. (IIAI-AAI)*, 2013, pp. 62–67, doi: 10.1109/IIAI-AAI.2013.26.

[143] E. E. Küçük, K. Yapar, D. Küçük, and D. Küçük, "Ontology-based automatic identification of public health-related Turkish Tweets," *Comput. Biol. Med.*, vol. 83, pp. 1–9, Apr. 2016, doi: 10.1016/j.compbiomed.2017.02.001.

[144] C. De Boom, S. Van Canneyt, T. Demeester, and B. Dhoedt, "Representation learning for very short texts using weighted word embedding aggregation," *Pattern Recognit. Lett.*, vol. 80, pp. 150–156, Sep. 2016, doi: 10.1016/j.patrec.2016.06.012.

[145] Y. J. Tai and H. Y. Kao, "Automatic domain-specific sentiment lexicon generation with label propagation," in *Proc. Int. Conf. Inf. Integr. Web-Based Appl. (Services-IIWAS)*, 2013, pp. 53–62, doi: 10.1145/2539150.2539190.

[146] L. G. Moreno-Sandoval, C. Sánchez-Barriga, K. Espindola, A. Pomares-Quimbaya, and J. Garcia, "Spanish Twitter data used as a source of information about consumer food choice," in *Machine Learning and Knowledge Extraction*. Cham, Germany: Springer, 2018, doi: 10.1007/978-3-319-99740-7_9.

[147] F. Namugera, R. Wesonga, and P. Jehopio, "Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda," *Comput. Social Netw.*, vol. 6, pp. 1–21, Dec. 2019, doi: 10.1186/s40649-019-0063-4.

[148] J. Rabelo, R. B. C. Prudencio, and F. Barros, "Collective classification for sentiment analysis in social networks," in *Proc. Int. Conf. Tools Artif. Intell. (ICTAI)*, vol. 1, Nov. 2007, pp. 958–963, doi: 10.1109/ICTAI.2012.135.

[149] E. Puertas, L. G. Moreno-Sandoval, F. M. Plaza-Del-Arco, J. A. Alvarado-Valencia, A. Pomares-Quimbaya, and L. A. Ureña-López, "Bots and gender profiling on Twitter using sociolinguistic features notebook for pan at CLEF 2019," in *Proc. CLEF Labs Workshops*, L. Cappellato, N. Ferro, D. E. Losada, and H. Müller, Eds. Lugano, Switzerland, Sep. 2019, pp. 1–10.

[150] L. G. Moreno-Sandoval, P. Beltrán-Herrera, J. Vargas-Cruz, C. Sánchez-Barriga, A. Pomares-Quimbaya, J. Alvarado-Valencia, and J. Garcia-Díaz, "CSL: A combined Spanish Lexicon-resource for polarity classification and sentiment analysis," in *Proc. 19th Int. Conf. Enterprise Inf. Syst. (ICEIS)*, 2017, pp. 288–295, doi: 10.5220/0006336402880295.

[151] L. G. Moreno-Sandoval, J. Mendoza, E. Puertas, A. Duque-Marín, A. Pomares-Quimbaya, and J. Alvarado-Valencia, "Age classification from Spanish Tweets—The variable age analyzed by using linear classifiers," in *Proc. 20th Int. Conf. Enterprise Inf. Syst. (ICEIS)*, 2018, pp. 275–281, doi: 10.5220/0006811102750281.

[152] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[153] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.

[154] E. Puertas, L. G. M. Sandoval, J. Redondo, J. Alvarado, and A. P. Quimbaya, "Detection of sociolinguistic features in digital social networks for the detection of communities," *Cognitive Comput.*, vol. 13, pp. 518–537, Mar. 2021, doi: 10.1007/s12559-021-09818-9.

LUIS GABRIEL MORENO-SANDOVAL received the master's degree in business management and digital marketing from the ENAE Business School, Universitario de Espinardo, Murcia, Spain, the M.B.A. degree from the Externado University of Colombia, and the master's degree in information sciences and communications from the Universidad Distrital Francisco José de Caldas. He is currently pursuing the Ph.D. degree in engineering with the Pontificia Universidad Javeriana. He is also a Systems Engineer at the Los Libertadores University Foundation, Colombia, and a Researcher in computational linguistics and social networks at the Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA).

ALEXANDRA POMARES-QUIMBAYA received the master's degree in systems and computer engineering and the Ph.D. degree in engineering from the Universidad de los Andes, Colombia, and the Ph.D. degree in computer science from the University of Grenoble Alpes, France. Since 2001, she has been with the University of Grenoble Alpes, where she is a Full Professor, and has held various positions as the Director of the Systems Engineering Program. She has also been a Visiting Researcher at the Medical University of Graz, Austria; the University of Aalborg, Denmark; and the University of Jaén, Spain, and a Visiting Professor at the Pontificia Universidad Católica del Perú. She is currently a Systems Engineer with the Pontificia Universidad Javeriana. She is also a part of the Pontificia Universidad Javeriana (CAOBA) Research Team, the Center of Excellence and Appropriation in Big Data and Data Analytics in Colombia. She is the Research Director of the Research Vice-Rectory, Pontificia Universidad Javeriana.