

# **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)**

Luis Gabriel Moreno Sandoval<sup>1</sup> and Liliana María Pantoja Rojas<sup>2</sup>

<sup>1</sup>*Pontificia Universidad Javeriana, Facultad de Ingeniería, Colombia*

<sup>2</sup>*Universidad Nacional de Colombia, Facultad de Ingeniería, Colombia*

**Keywords:** Reputational Risk, Twitter, Polarity, Social Network, Computational Linguistic.

**Abstract:** Within a society increasingly technological and immersed in the different possibilities offered by the Internet as a channel of interaction and communication, there is a need for private and state entities to explore the information contained in social networks, an opportunity that implies new challenges associated with emerging risks such as reputational. For this reason, through the study of social networks, reputational risk and computational linguistics, an analysis of the strategic accounts of the El Dorado Airport of the city of Bogotá (Colombia) on Twitter (@bog\_eldorado) is carried out, taking into account the mentions and most relevant hashtags, to identify the polarity through of BoW (Bag of Words) of your comments and infer the user experience, customer satisfaction and how this influences reputational risk.

## **1 INTRODUCTION**

The availability of massive amounts of data in the scenario of the web has given new statistical and computational impetus to the field of social network analysis. According to Aggarwal (2011), thanks to this emerging phenomenon, the study of social networks has taken a new direction concerning the results of data analytics, changes in the paradigms of the computational social sciences along with the analysis the text of social extraction, the complexity sciences and social simulations.

Consequently, the computational challenges associated with the ability to perform mining and analytical processes to these sources of information in the context of a social network, constitute an unprecedented challenge and an opportunity to determine useful information in a great variety of fields (Mika , 2004; Mislove, Marcon, Gummadi, Druschel, and Bhattacharjee, 2007; Zafar, Bhattacharya, Ganguly, Gummadi, and Ghosh, 2015; Zafarani, Abbasi, and Liu, 2014), including reputational risk.

Participation in contemporary social life through networks has become increasingly complicated; people accumulate hundreds of friends and acquaintances through social media, but more and

more report a smaller number of friends they can trust (Burt, Kilduff, and Tasselli, 2013).

For the foregoing and in order to analyze as a case study the reputational risk at El Dorado Airport in Bogotá (Colombia) taking into account the most relevant mentions and hashtags, to identify the polarity of their comments and infer the user experience and customer satisfaction, six sections are presented. The second refers to the conceptual framework, the third is the case study. In the fourth and fifth sections is the methodology and development and in the sixth section the conclusions and future work.

## **2 CONCEPTUAL REVIEW**

The present study has three fundamental pillars: reputation, social networks and computational linguistics that will be described below:

### **2.1 Reputational Risk**

The traditional types of risk that can affect reputation are mainly strategic, operational, financial and compliance risks, which can often be quantified, but their collective impact on reputation is difficult to

foresee (Lamont, 2015). However, reputation can also be affected by additional risks that fall outside these areas, especially those that come from social networks.

It should be understood that reputation is one of the main intangible assets that helps organizations not only to obtain competitive advantages, but also to survive in times of economic turbulence (Fernández-Gómez, Gil-Corral, and Galán-Valdivieso, 2016), which is why it is necessary for organizations to use a multifaceted approach.

The organization managers, estimate that an intangible increasingly important in the organizational field is the corporate reputation (CR), and although it is the most important asset, it is also the most difficult to protect (E., De la Fuente Sabaté, and Delgado García, 2005). In the eighties, many experts in organizational valuation assured that 70% of the value of a business depended on its tangible assets and 30% of what its intangible assets contributed. At present, it is thought that 75% of the value of a business depends on the quality and management of intangible assets (Tomás Garicano Rojas, 2011).

Now, for Hernández Velasco (2012), from the corporate point of view, reputation is the recognition that interest groups or "stakeholders" make of the behavior of an organization in the satisfaction of its expectations and reputational risk in relation to the response that these "stakeholders" can have when their expectations are not met.

## 2.2 Social Network Analysis (SNA)

A social network consists of a set of relationships established between a group of social actors, as well as any additional information about those actors and their relationships (Ö Bodin and Prell, 2011). The social network's approach is based on the notion of having patterns associated with social ties, in which some actors are involved, and their interactions have significant consequences for them. Social network analysts seek to discover different kinds of patterns, determine the conditions under which they arise and explore their consequences on the attitudes, perspectives, and behaviors of individuals, groups or subgroups and the systems to which they belong (Scott J, 2011 ).

### 2.1.1 Basic elements of a Network

According to Bodin and Crona, (2009), we have the following definitions of the essential elements of a network:

**Nodes or actors:** These are the people or groups of people who are around a common goal. Usually, the nodes or actors are represented by circles. The sums of all the nodes represent the size of the network.

**Link:** These are the links that exist between two or more nodes. In a network, a direct connection with another actor is shown. Lines represent the links or relationships.

**Flow:** Indicates the link address. An arrow indicating the direction represents the streams. It is also possible that there are also mutual or bidirectional flows. When an actor does not have any current, which in turn implies no link, it is said that this node is loose within the network.

### 2.1.2 Analysis of Social Networks-Properties

According to Freeman (2004), the analysis of social networks is based on four characteristics that have been consolidated since their appearance and whose integration gives rise to a research paradigm:

- The study of social networks is motivated by a basic intuition based on the bonds that unite social actors.
- It is an approach that is based on systematic and empirical data.
- It is found in no small extent on graphic representations and,
- It relies on the use of mathematical and computational models.

According to the thesis: "Evolution of the policy network for innovation in Colombia: The emerging phenomenon of confirmation of governance networks" (Moreno, 2015), the structural characteristics of the network that the theory on the analysis of social networks distinguish the following elements:

**Level of cohesion of the network:** an essential feature of social networks is their level of coherence, that is, to what extent the network "works together" when presenting relations between nodes instead of being divided into subgroups cohesive among themselves. This means that a network with high structural cohesion lacks a set of differentiable subsets, understanding that the existence of subgroups represents challenges to unite the action towards common initiatives.

**Position and influence of the network:** this is a structural characteristic that is analyzed to understand the location of the actors in the structure to influence the system. Being better placed can access valuable information (Burt et al., 2013). There are several ways to define and measure centrality in social

networks. The first is by a degree of centrality and the second by a degree of intermediation.

The degree of centrality refers to the number of links an actor has or can be understood as the degree to which an individual actor connects with other actors and is associated with the level of activity in the network (Burt et al., 2013). In the analysis of social networks, the degree to which an actor indirectly connects to other actors is often quantified using the degree of intermediation (Freeman, 2004) with which it is determined that an actor that is among many other actors in The network has a high degree of mediation, which implies that it could act as a bridge between others that would be disconnected without their existence.

### 2.3 Computational Linguistics

According to Montes-y-Gómez, (2002) computational linguistics is *"the science that deals with the application of computational methods in the study of natural language. This science is a combination of two larger sciences; linguistics, which studies the laws of human language, and artificial intelligence, which investigates computational methods for the management of complex systems."*

Now, the sentimental analysis aims to determine the attitude of a speaker or a writer concerning some topic, or the global polarity of a document. According to Alborno (2011) the polarity classification aims to obtain a score that indicates whether the text expresses a positive or negative opinion, within a range where 0 would mean a neutral subjective load, 1 a positive personal load and -1 a load personal negative.

It is possible to use many techniques in the classification of polarity, these can be supervised algorithms such as SVM (Support Vector Machine), LR (Logistic Regression), NB (Naive Bayes) among others, there are other types of models are based on lexicons; in this case study was used a polarity lexicon called "CSL: A Combined Spanish Lexicon - Resource for Polarity" through a process of BoW (Bag of Words) (Moreno et al., 2017). It was possible to determine the polarity of comments, hashtags, and mentions in the tweets, finally, these results are consolidated in the kind cluster file that is read for the SNA process.

## 3 CASE STUDY: EL DORADO AIRPORT

The El Dorado Airport is part of the urban and architectural history of Colombia and Bogotá. OPAIN S.A. is the company set up with the sole objective of managing, modernizing, commercially developing, expanding, operating and maintaining the El Dorado International Airport.

It is the first airport in Latin America in cargo volume and the third largest airport in Latin America in passenger volume, after the International Airport of Mexico City and the International Airport of São Paulo-Guarulhos.

He won the Skytrax award for the best personnel in South America in 2016, as well as the first place on the list of the best airports in Latin America for the second time in a row. On the other hand, in the list of the best airports in the world, El Dorado achieved the 42nd position in the list of 2017.

A total of 42 lines arrive at the airport, discriminated against as follows: 21 international airlines, six domestic airlines, and 15 cargo airlines.

## 4 METHODOLOGY

This is empirical research, that is, based on experimentation, to test the hypothesis that under the evolution of the structural properties of the social network Twitter it is possible to establish relationships that show reputational risk.

The analysis of the structural properties of the El Dorado Airport network (as a case study) is done by systematizing and examining the relationships established through Twitter, for which the methodological design is based on the use of the Pajek and Voswiever software for the construction of graphs corresponding to each of the structural properties defined for the study. All this, from the data obtained by "crawling" (crawling) the public access information of Twitter through the use of application programming interfaces (APIs) whose complexities describe Mislove et al. (2007) extensively.

This is how to achieve the objectives of this empirical research is carried out in three phases in the methodological design:

- a) Conceptual exploration for the identification of reputational risk categories and structural analysis.

- b) Application of an exploratory methodology to describe the structural properties of the El Dorado Airport's digital ecosystem.
- c) Empirical evaluation of the behavior of the network in its general interactions of the El Dorado Airport network.

The steps to carry out the following investigation are summarized below:

1. A network is woven to analyze it empirically, and the social structure is explained to establish its degree of significance and relevance for the case study.
2. The types of account to be analyzed are found because they are not homogeneous since they are mixed reports of organizations with personal accounts.
3. All friendship links are "downloaded" on Twitter and friends are established for each of the accounts.
4. The "mentions network" and the "hashtag network" are established.
5. The user is chosen, and the tweets are taken, and then each of the posts is decomposed, and the polarity is calculated (mentions and hashtag).
6. The list of remarks and hashtag is stored; the value of the vector is normalized. The amount of mentions and hashtag of the moment is accumulated in a vector per user mentioned, and the number suggested to determine the importance.  
The normalized value is equal to:  
$$\text{normalized value} = (\ln \text{value to normalize} + 1) / (\ln \text{calculated max} - \ln \text{calculated min} + 1) * 50$$
7. Each post is read and polarity is extracted through the iSOL and CSL Caoba lexicon, accumulating the polarity of the user. The information is accumulated as positive, negative and neutral.
8. Pajek files are written.
9. VOSviewer is used; According to its official website <http://www.vosviewer.com/>, as a software tool for the construction and visualization of networks.

## 5 DEVELOPING

The present study was conducted during April and May of 2018. The ecosystem of social network accounts: 72 (Appendix)

The data collected are the following:

Table 1: Dates.

Description	Count
<b>Follower</b>	41826
<b>Hashtag</b>	65013
<b>Mention</b>	146922
<b>Profile</b>	41841
<b>Tweets</b>	176251

Phase 1: The study is closed for Twitter, all tweets from the ecosystem timeline of social media accounts were taken, and followers, hashtags and mentions at the structural level were analyzed, showing the density of the interconnections between the followers of the reports of the digital ecosystem of El Dorado Airport.

Phase 2 Polarity is defined as positive, negative and neutral.

According to the information analyzed, the following graphs are obtained.

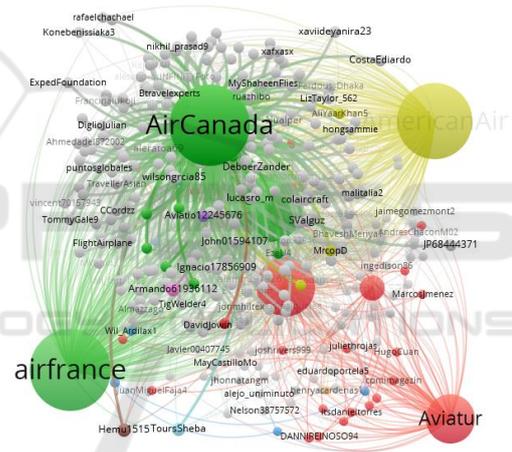


Figure 1: Followers.

The graph shows how the density of the interconnections between the followers of the digital ecosystem accounts of the El Dorado Airport has manifested and also suggests that the links of the topics discussed are stable due to a large amount of message exchange.

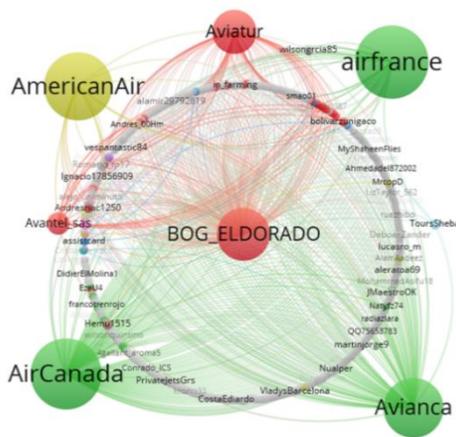


Figure 2: Degree- El Dorado Airport.

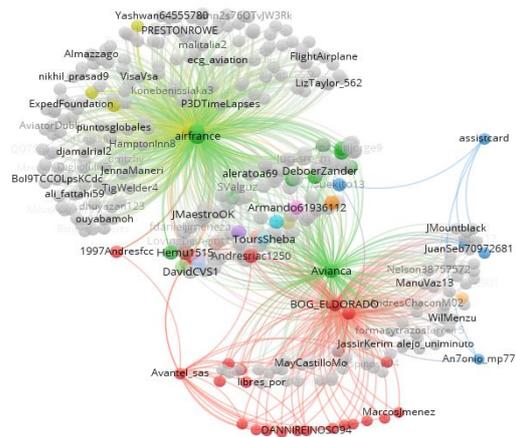


Figure 4: Clonnes Centrality – El Dorado Airport.

On Twitter, the nodes with the most significant number of followers when issuing a trill, have a higher probability of generating an impact on the network through the simple fact of reaching a more substantial amount of users. This impact does not depend solely on the number of followers but the number of responses or mentions generated as a result of the trills of the user.

Therefore, it is evident that Air Canada, American Air, AirFrance, Aviatu and Avianca are the ones with the highest degree of centrality concerning the airport in question.

The centrality by proximity determines how close a node is to all the other nodes of the network, establishing the importance of calculating the distance that exists between the node with the rest of the system and a node with greater proximity. The graph shows great importance with Avianca and AirFrance.

Hashtags are essential for any digital marketing campaign that is carried out, as they can frame the success or failure of an idea, slogan or concept of any brand, product or organization. Outstanding #AttendeesTravellers, #Happy Monday, #Cofee

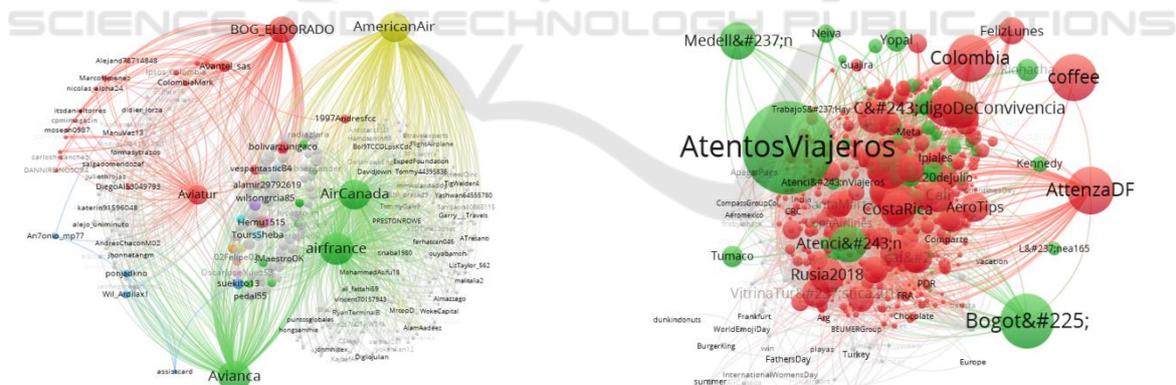


Figure 3: Between Centrality - El Dorado Airport.

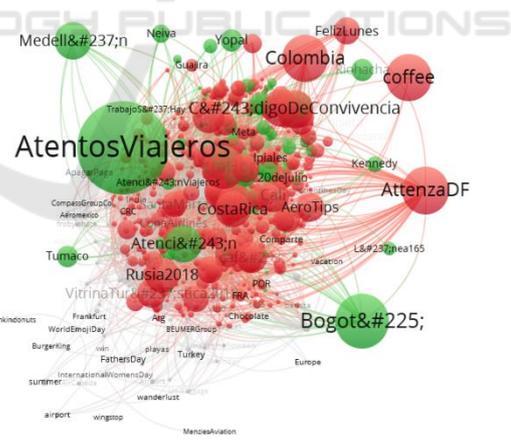


Figure 5: Polarity Hashtag - El Dorado Airport.

Nodes with high intermediation play an essential role in the communication of the entire network. Therefore, it is observed that the nodes that connect more effectively within the system are again Air Canada, American Air, AirFrance, Aviatu and Avianca.

When observing the polarity of "AtentosViajeros", it is found that it is positive because it allows the users to know the situations of the last moment of the airport and to be attentive to the events that are presented. When analyzing



#AttendeesTraveling Our Warriors of the @FCFSeleccionCol are coming home and we are ready to receive them. If you travel from @BOG\_ELDORADO tomorrow, keep in mind that congestions and vials closings are planned on 26th street from 10:00 a.m.



But for #HappyMonday it is observed that the algorithms used are able to abstract the sarcasm since the message is opposed to the hashtag #HappyMonday #Holiday less for airport users the golden one who are locked in a plane waiting for a runway #Planretorno



The coffee for example that indicates. "I just copied an aromatic and a cake in the airport Oma and then I was left without the rent"

The foregoing demonstrates aspects that the Airport did not contemplate but that its partners were affecting their reputation.

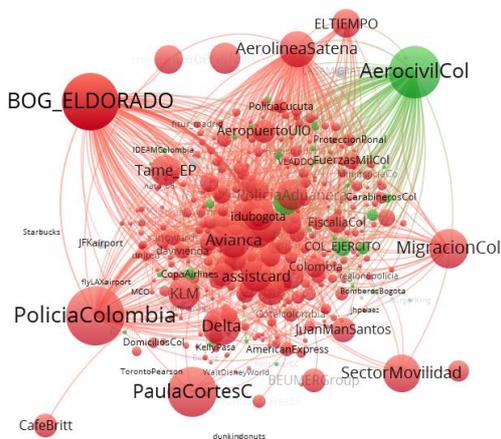


Figure 6: Polarity of Mentions.

When carrying out the polarity analysis, for this case a reputational risk contagion issue is observed between the relationship of Aerocivil de Colombia and El Dorado Airport; this was due to the fact that situations such as the lack of light at the airport were attributed to the airport, although it is an obligation of Aerocivil; however, users directly related the responsibility to El Dorado because there is no distinction between the public of what activities each of these entities performs.

## 6 CONCLUSIONS

Social networks have great potential to identify groups of interest, know their needs, expectations, tastes, preferences, and opinions on specific initiatives and / or products; therefore, this type of studies allows identifying events that can generate reputational risk in an organization.

Currently, the volume of comments on the network is very high, so an organization does not have the capacity to process them manually; therefore, it is necessary to have technological tools such as those presented in this study, which help to automatically transform all the opinions of interest and identify their type of polarity (positive, negative or neutral) for decision making.

It is possible the integration of computational linguistic analysis with the structural analysis of networks linking the results of the polarities associated with the mentions and hashtag, which allows understanding the dynamics of the reference risk information. To the extent that actions to control reputational risk are identified, monitored and executed, the gap closes, and the probability that the crisis in an organization is managed will be much more effective.

The information contained in social networks is an opportunity that involves new risks associated with risks such as reputational risk. Through the SNA on Twitter and computational linguistics, it is possible to identify the polarity of comments and hashtag to infer the user experience and customer satisfaction.

According to the structural analysis, foreign airlines such as AirCanada and AirFrance are fundamental for strategic alliances in the handling of information since the reputation comes from third parties and what is sought with the protection of the status is the creation of value of an organization.

The polarity of the comments and hashtag as well as the identification of topics at El Dorado Airport allowed us to find aspects that the Airport was not contemplating but that were affecting its reputation as

the coffee issue related mainly to prices and service, which was changing his status.

In conclusion, this work contributes in the emerging study of reputational risk through the Twitter network is a very nascent topic in Latin America; therefore, it should be considered as a new field of research to be used by different organizations, since the reputation is of the main assets intangible of a company.

For future work, the reputation risk can be supported by supervised techniques of polarity classification, this would help to improve the calculation of risk in digital social networks. It has structural information which can be integrated into the training data set as the number of followers or the average number of words and sentences for tweets, adding more characteristics in the models of polarity classification.

## ACKNOWLEDGEMENTS

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

## REFERENCES

- Aggarwal C.C. (2011) An Introduction to Social Network Data Analytics. In: Aggarwal C. (eds) Social Network Data Analytics. Springer, Boston, MA.
- Becerra Bonza, D. C., and Guerrero García, T. M. (2012). Social representations workers of a public organization have about psychosocial risks and harassment in the workplace towards the Promotion of Health, 17 (1), 118-132.
- Bodin, Ö., and Crona, B. I. (2009). The role of social networks in natural resource governance: What relational patterns make a difference ?. *Global environmental change*, 19 (3), 366-374.
- Bodin, Ö., and Prell, C. (Eds.). (2011). *Social networks and natural resource management: uncovering the social fabric of environmental governance*. Cambridge University Press.
- Bourcier, D. (2003). *Artificial intelligence and law*. Editorial UOC
- Burt, R. S., Kilduff, M., and Tasselli, S. (2013). *Social Network Analysis: Foundations and Frontiers on Advantage*. <http://doi.org/10.1146/annurev-psych-113011-143828>.
- Contreras Barrera, M. (2014). Text mining: a current vision University Library, vol. 17, no. 2, July-December, 2014, pp. 129-138 National Autonomous University of Mexico Federal District, Mexico. University Library, 17 (2), 129-138
- From Albornoz, J.C., Plaza, L., Gervás, P., and Díaz, A. (2011, April). A joint model of feature mining and sentiment analysis for product review rating. In *European conference on retrieval information* (pp. 55-66). Springer, Berlin, Heidelberg
- E., from the Sabaté Source, J. M., and Garcia, J. B. D. (2005). Corporate reputation and value creation. A theoretical framework of a circular relationship. *European Investigations of Management and Economy of the Company*, 11 (2), 81-97.
- Fernández-Gámez, M.A., Gil-Corral, A.M., and Galán-Valdivieso, F. (2016). Corporate reputation and market value: Evidence with generalized regression neural networks. *Expert Systems with Applications*. <http://doi.org/10.1016/j.eswa.2015.10.028>
- Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, 1.
- Garicano, T. (2011). Corporate governance and corporate reputation. *Corporate Governance Center* <http://doi.org/10.1016/j.dss.2013.09.004>
- Hernández Velasco, J. (2012). Reputational risk: Assess to prevent; prevent to act. *Give in*, 32, 1-10.
- Hernández M, and Gómez, J. (2013). Natural Language Processing Applications. *Polytechnic Journal*, Vol. 32, N, 87-96.
- Kanayama, H., and Nasukawa, T. (2006, July). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 355-363). Association for Computational Linguistics.
- Kim, S. M., and Hovy, E. (2004, August). Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics* (page 1367). Association for Computational Linguistics.
- Lamont, J. (2015). Risk Management: Reputation is everything. *KMWorld*, 60 (1), 8. Retrieved from <http://ra.ocls.ca/ra/login.aspx?inst=conestoga&url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=100327144&site=bsi-live&scope=site>.
- Martín De Castro, G. (2008). *Business reputation and competitive advantage*. Esic Editorial. Madrid. pp. 32-35.
- Mika, P. (2004). Social networks and the semantic web. *Proceedings of the 2004 IEEE / WIC / ACM International Conference on Web Intelligence*. <http://doi.org/10.1109/WI.2004.10039>
- Mislove, A., Campus, E., Marcon, M., Gummadi, K. P., and Druschel, P. (2007). *Measurement and Analysis of Online Social Networks*.
- Montes-y-Gómez, M., Gelbukh, A., and López-López, A. (2002, July). Text mining at detail level using conceptual graphs. In *International Conference on Conceptual Structures* (pp. 122-136). Springer, Berlin, Heidelberg.

Moreno-Sandoval L. (2015). Master's Thesis: Evolution of the policy network for innovation in Colombia: The phenomenon emergent form of governance networks. Faculty of Administration. university externship of Colombia.

Moreno-Sandoval L., Beltrán-Herrera P., Vargas-Cruz J., Sánchez-Barriga C., Pomares-Quimbaya A., Alvarado-Valencia J. and García-Díaz J. (2017). CSL: A Combined Spanish Lexicon - Resource for Polarity Classification and Sentiment Analysis. In Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-247-9, pages 288-295. DOI: 10.5220/0006336402880295

Pang, B., Lee, L., and Vaithyanathan, S. (2002, July). Thumbs up ?: feeling classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical Methods in Natural Language Processing, Volume 10 (pages 79-86). Association for computational linguistics.

Scott, J. (2011). Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining*, 1 (1), 21-26. <http://doi.org/10.1007/s13278-010-0012-6>

Tomás Garicano Rojas. (2011). Introduction to the management of reputational risks. IE Bussines School and Corporate Reputation Forum, 1-97.

Zafar, M. B., Bhattacharya, P., Ganguly, N., Gummadi, K. P., and Ghosh, S. (2015). Sampling content from online social networks: Comparing random vs. expert sampling of the twitter stream. *ACM Transactions on the Web (TWEB)*, 9 (3), 12.

Zafarani, R., Abbasi, M.A., and Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press

beumergroup	elcorral_	aerolineasatena	trafficaicolom
cafebritt	elmarketco	spiritairlines	anatonacional
budget	frisbylohace	starbucks	reportcolombia
burgerking	ussfontibon	tame_ep	idubogota
cafeygourmet	iberia	turkishairlines	policia colombia

## APPENDIX

### Analyzed Digital Ecosystem Airport El Dorado

bog_eldorado	xuecafe	iemark	vivaaircolombia
aerocambiar	cambiosdorado	interjet	wingstop
aerolineas_ar	carolinaherrera	jetblue	ziropevzla
aeromexico	colsubsidio_ofi	klm	cancilleriacol
aviatur	compassgroupcol	kokoriko_col	migracioncol
aircanada	prioritypasscom	latam_co	diancolombia
aireuropaco	copairlines	lasamigration	icacolombia
airfrance	crepeswafflesco	lufthansa	policia bogota
americanair	davivienda	unibaggage	policia aduanera
assistcard	directvã	mh_colombia	policia antinar
attenzadf	worlddutyfreees	mcdonalds	aerocivilcol
avantel_sas	dunkindonuts	menziesaviation	redirecciona el dorado
avianca	easyflyvuelos	omacolombia	invimacolombia.