



A Business Reputation Methodology Using Social Network Analysis

Antonio Galli, Michela Gravina, Vincenzo Moscato, Antonio Picariello,
Carlo Sansone, and Giancarlo Sperli^(✉)

University of Naples “Federico II”, Via Claudio 21, Naples, Italy
{anto.galli,mi.gravina}@studenti.unina.it,
{antonio.picariello,carlo.sansone,giancarlo.sperli}@unina.it

Abstract. Nowadays, every facet of people’s lifestyle is impacted by the continuous use of technologies and in particular the Internet. Social interactions have been radically changed since new technologies have removed communication barriers. The obstacles of time and space have been overcome, letting people from different places and cultures communicate. People tend to become part of a dense social network, where the distribution of information becomes an almost immediate process. This is the reason why companies, public institutions and business activities have opted for social networks as communication medium. As a consequence, people seem to incorporate information gained from the social networks into their decision-making processes.

In this paper we analyze how *what is said* in the social network could influence people’s decisions. As social network, we consider YELP community, that is a user generated content platform based on ‘word of mouth’, in which users can share their opinions about news, product, community, businesses. We propose a methodology for reviews analysis with the aim to compute business attractiveness, combining user’s sentiment and business reputation.

Our results show that reviews analysis should be performed because it may provide useful information to monitor how business public opinion changes over time.

Keywords: Natural language processing · Sentiment analysis · Social network analysis

1 Introduction

The proliferation of technology, especially the widespread use of the Internet, has completely changed people’s life letting to overcome the obstacles of time and space. Undoubtedly, the major impact has been on social life where connections between people from different places and cultures have been made possible. As a consequence, people become part of a dense social network, encouraging the transfer and sharing of information. The development of technology has, then, enhanced OSN features, enabling users to share their life by describing

multimedia content (text, audio, video, images) and interacting with such objects in order to provide feedback or comments or feelings compared to them. In the last years, several OSNs (such as Facebook, Twitter, and so on) provide several features to share and interact with multimedia contents (audio, video, images, texts). Social networks like Facebook are designed for social interaction, while others such as Flickr are designed for content sharing services, allowing people to improve their social connections. The analysis of these heterogeneous data assumes a key role for many companies. On one hand it represents the lifeblood for this company on other hand managing these information becomes more difficult. To gather knowledge from these raw data requires a new approach that involves the Big Data and related technologies. In fact using a simple ‘click’, it is possible to know people’s thoughts, opinions, to send and receive messages, keeping in touch with anyone you want. This is the incredible power that has made businesses exploit social networks for marketing, advertising and for being more easily connected to their customers. In fact, it implies that it is easy to be ‘influenced’ by the others if the majority of people shares the same idea or if somebody who is well known in the network (the so called influencer) expresses an opinion about a topic, such as a business object. In this way, people’s thought about a topic could be the result of what is extracted from the network.

In this paper we analyze the power of social networks, considering what people think about businesses. In particular we take into account people’s reviews in order to evaluate a business reputation, that is its *public opinion*. In our opinion, reviews analysis can be useful to a business owner for realizing activities aimed at improving the performed services and for ‘having a look’ of what people think, getting an immediate feedback. Furthermore, reviews analysis should be performed because customers tend to opt for businesses with a *good public opinion*.

As case of study, we considered YELP Dataset Challenge¹, containing about 5 million reviews, more than one million users, about 150 thousand businesses. Yelp is a user generated content platform in which users feed a virtual community based on ‘word of mouth’.

The rest of the paper is organized as follows: Sect. 2 outlines some related works; Sect. 3 introduces the proposed approach; Sect. 4 outlines system architecture; Sect. 5.1 introduces the dataset used and the experimental setup; Sect. 6 reports the obtained results, finally, Sect. 7 provides some conclusions.

2 Related Works

In last years with the advent of the Internet, Web 2.0 and the Online Social Network (OSN) the expression “social network” has become a common vocabulary. Their adoption has allowed people living in different parts of the world to establish different type of relationships and to share, comment and observe various types of multimedia content. OSN is often associated with the more general concept of “social media”: it refers to online technologies and procedures that

¹ <https://www.yelp.com/dataset/challenge>.

social network users use to share textual contents, images, videos, and audios. The analysis of these heterogeneous data allowed to support different applications, in particular related to marketing strategies. In fact, the interaction among firms and customers led to build brand loyalty through the promotion of products and services as well as the setting up of online communities of brand followers [4]. Furthermore, the interaction among customers allows firms to increase the trend reputation exploiting new means and channel [3]. Social media marketing campaign is also influenced by the type of industry and product. For instance, in the hotel industry Corstjens and Umblijs [2] show how the firm reputation impacts the effectiveness of social media efforts. Thus, Social media improves the relationships among customers and brand improving exposure time and spread of marketing campaign [8]. The aim of these marketing campaigns is to spread out the influence of a given product or technologies among users. The influence process is, then, a process where ideas or behaviors are spread with the initiator and the recipient unaware of any intentional attempt at influence by giving advice and recommendations, by serving as a role model that others can imitate, by persuading or convincing others, or by way of contagion [11]. Rogers [6] provides a definition based on the communication concept describing as a process in which participants create and share information with one another in order to reach a mutual understanding. In particular, Rogers defines the Diffusion as the process in which an innovation is communicated through certain channels over time among the members of a social system. For this reason the analysis of business attractiveness has grown in importance to identify main features of a given business object with respect to its competitors, as shown in [9], or to analyze how the users' ratings affect deal selection [10]. Social Media provide then useful information for computing business attractiveness because they provide different point of views about a given firms in terms of ratings, reviews, pricing and so on. In fact, in [12] the authors use Yelp to study human foraging patterns modern human food foraging patterns, with respect to both geography and cuisine. On-line reviews are also investigating by Bai et al. [1] for product marketing combining user's rating and useful score of its reviews. Zhao et al. [13] studies user's influence on local businesses using user sentimental deviations and the review's reliability. Nevertheless, the large amount and the heterogeneity of users' reviews led to define new challenges, as shown [7] that have to be address for properly supporting different applications. Furthermore, Li et al. [5] show how the Big Data methodologies can be used for supporting marketing analysis investigating also the connection between social media and marketing stocks.

The main novelties of the proposed approach are:

- The definition of new methodology for reputation computing in OSN based on the analysis of textual reviews;
- The adoption of a new strategy to evaluate business' attractiveness combining users' sentiment and business reputation;
- The implementation of the proposed approach on a scalable and parametrizable Big Data infrastructure.

3 Methodology

The aim of our approach is to combine user's sentiment and business object's reputation on the basis of users' reviews for computing business attractiveness. We analyze reviews written by different users who express their opinion using natural language assigning a star degree (from 1 to 5) to their experience in the business object. In particular our analysis is focused on the reviews expressed by users with the aim of extracting useful information to determine how much a business is appreciated. In the following we introduce some basic concepts to better explain the proposed methodologies as well as:

- the concept of review and the reason why it is important to analyze business reviews.
- the concept of *public opinion*.
- the concept of user and his main characteristics.

Review is a subjective opinion expressed by a user about a given business to judge its main features in both positive and negative terms. In particular a review is directed to potential customers, so its function should help them in the decision making process. For this reason, a review inevitably influences the user's thought, having an impact on what is considered *public opinion*.

On the other hand, reviews are useful for business owner who should be interested in improving business services and in fixing those that have not been successful.

Public Opinion is defined as the collective thinking of the business clients' majority. We argue that *user's opinion about a business, of which the user knows nothing, is strongly influenced by public opinion, that becomes the user's prejudice without direct experience of business services*. As a consequence, businesses with a 'good public opinion' are more likely to attract new customers, whereas a customer is unlikely to opt for a business with bad reviews.

User is someone who is able to make a review about a business or to provide feedback on reviews from other users, confirming or contradicting them. In Yelp community, each user has a personal page containing personal information and a set of attributes that allow to define how much the user is able to 'influence' the others.

In our opinion we can define an 'influencer' someone that respects the following characteristics:

- *Elite User* that is active in the Yelp community according to the number of well-written reviews, high quality tips, a detailed personal profile, an active voting and complimenting record. Yelp Elite is a privileged title granted to users by Yelp.
- *Large Number of fans*: this implies that the user is popular.
- *Large number of written reviews*: this means that he is active in the community.

- *Large number of friends*: this implies that the user is part of a wide social network.
- *Large number of compliments by other users*: this takes into account how ‘useful’, ‘funny’ and ‘cool’ the user’s reviews have been and how many compliments the user has received by other users of the community.

Our purpose is to assess a business public opinion in a quantitative manner, taking into account different influence factors (i.e. opinion value, number of compliments and so on). In fact, we argue that public opinion may depend on both ‘what’ has been written in the reviews and who has written that.

Definition 1. Let r and u_r be respectively a given review and user who has written the review r , we define the opinion value $Ov(r, u_r)$ according to the following relation:

$$Ov(r, u_r) = f(R(r), U(u_r), S(r)) \quad (1)$$

where $R(r)$ is the review value, $U(u_r)$ is the user value and $S(r)$ is the review success.

More specifically, we better describe the meaning of the $R(r), U(u_r), S(r)$ concepts.

$R(r)$ takes into account the number of starts of the review r , indicated by s_r , which is an user evaluation metric expressed as a number between 1 and 5, and the review text tx_r that is the user’s thought written in natural language. It is necessary to convert tx_r into a numeric value, processing it using Sentiment Analysis and Natural Language Processing *nlp* libraries. This step is referred by the notation $nlp(tx_r)$. Thus, $R(r)$ can be defined according to the following relation:

$$R(r) = g(s_r, nlp(tx_r)) \quad (2)$$

In turn, considering u a generic user, $U(u)$ is computed considering its attributes, including the number of friends, the number of written reviews, the number of received compliments, the number of fans and the title of Elite User. Then $U(u)$ can be computed according to the following equation:

$$U(u) = \sum_{i=1}^n w_i a_u^i \quad (3)$$

where w_i and $a_u = (a_u^1, a_u^2, \dots, a_u^n)$ are respectively the weight of the attribute a_u^i and the vector corresponding to the attributes of user u .

Finally, the review success $S(r)$ depends on how ‘useful’, ‘funny’ and ‘cool’ users consider the review r . It is possible to compute $S(r)$ according to the following equation:

$$S(r) = \beta us_r + \delta f_r + \gamma c_r \quad (4)$$

where β, δ, γ are the weights associated to us_r, f_r and c_r , us_r the number of times the review is found ‘useful’, with f_r the number of times the review is found ‘funny’ and with c_r the number of times the review is found ‘cool’.

According to the Eq. (1), an *opinion value* is computed for each review. However, the *public opinion of a specific business* can be computed taking into account the *opinion value* $Ov(r, u_r)$ of all the reviews r related to the business. In particular the *public opinion of a specific business* changes over time depending on the reviews that have been written up to a specific time t . Furthermore, each review is weighted in proportion to when it has been posted; in fact the most recent reviews seem to influence the opinion of users more. Thus, considering a specific business b and indicating with r_t the review posted at t , its *public opinion at specific time t* , $BPO(b, t)$ can be computed according to this recurrence relation:

$$BPO(b, t) = Ov(r_t, u_{r_t}) + \alpha BPO(b, t - 1) \quad (5)$$

where α is the weight associated with public opinion generated by the reviews preceding the one at time t .

4 Architecture

In this section system architecture is described. As shown in Fig. 1, three main layers have been selected:

- Data ingestion that has the aim to crawl data from different data sources as well as Yelp, Foursquare, TripAdvisor and so on;
- Data storage in which are stored the crawled reviews as document into the NoSQL document database;
- Data processing using Big Data methodologies to infer business object’s attractiveness.

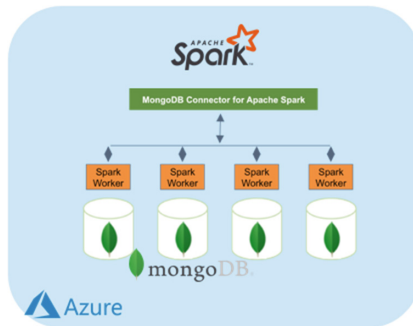


Fig. 1. System architecture

More in detail, we store users’ reviews into the NoSQL database MongoDB² because it allows to easily manage each review as a document in conjunction with its metadata (i.e. title, number of compliments, business object and so on).

² <https://www.mongodb.com/it>.

The data processing layer is based on Apache Spark³, an open-source distributed general-purpose cluster-computing framework, to process these large amount of heterogeneous data. Finally, we use Microsoft Azure platform as infrastructure as a service (IaaS) to support scalability of the proposed approach.

5 Experimental Evaluation

In this section we describe the evaluation results obtained according to the experimental protocol defined into the Sect. 5.1.

5.1 Experimental Protocol

Our evaluation concerns the following three types of analysis:

- Perform the Parameter estimation for identifying the best parameters for the computation of *Opinion value*;
- Evaluate the Natural Language Processing module with respect to the users evaluation (in terms of stars) about its reviews;
- Analyze how the business object reputation changes over the time.

We carried out our evaluation on YELP Dataset Challenge⁴. Yelp is a user generated content platform in which users feed a virtual community based on ‘word of mouth’. Yelp Dataset Challenge⁵ is composed of six json files:

- *business.json* contains business data including location data, attributes, and categories.
- *review.json* contains full review text data including the identifier of the user that wrote the review and the identifier of the business the review is written for. For each review the number of stars assigned by the user is given. The review date is necessary to track when it has been written. Furthermore, the attributes ‘useful’, ‘funny’ and ‘cool’ indicate the number useful, funny and cool votes respectively.
- *user.json* contains the user data including the user’s friend mapping and all the metadata associated with the user. Each user has several attributes, including the number of written review, the number of ‘useful’, ‘funny’ and ‘cool’ votes received, the years the user was Elite and the number of compliments received.
- *checkin.json* contains the checkins on a business.
- *tip.json* contains the tips written by a user on a business. Tips are shorter than reviews and tend to convey quick suggestions.
- *photo.json* contains photo data including the caption and classification (one of “food”, “drink”, “menu”, “inside” or “outside”).

Among all json files, only *business.json*, *review.json*, *user.json* have been taken into account during the analysis.

The evaluation has been performed on Microsoft Azure using a Linux virtual machine with the characteristics shown in Table 1.

³ <https://spark.apache.org/>.

⁴ <https://www.yelp.com/dataset/challenge>.

⁵ <https://www.yelp.com/dataset/challenge>.

Table 1. Virtual machine characteristics

Operating system	Linux Ubuntu 17.04
CPU	Intel Xeon E5
RAM	16 GB

The presented methodology has been implemented using PySpark, a collaboration of Apache Spark and Python, that exploits the simplicity of Python and the power of Apache Spark.

5.2 Experimental Results

As aforementioned, the *public opinion at a specific time t of a business b* (Eq. 5) can be computed by taking into account equations (1), (2), (3) and (4). Equation (1) represents the opinion value which depends on *what* has been written (review value), *who* wrote it (user value) and *how* successful it was (review success). Review value ($R(r)$) is computed by processing the text with the Natural Language Toolkit (NLTK)⁶, a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. In Yelp Community, when writing a review, the user is used to assign a number of stars (form 1 to 5) that can be interpreted as his brief opinion. For this reason, the number of stars should be taken into account when computing the review value $R(r)$ that is the average of the number of stars normalized between -1 and 1 as shown in Table 2, and the value NLTK provides by processing the review text. User value $U(u)$ is computed to estimate the ‘importance’ of the user in Yelp Community. In this context, a user should have a great ‘importance’ if his opinion has a strong impact. This means that his reviews are likely to be considered by other users. User value $U(u)$ is computed by taking into account user attributes, giving a high value to ‘influencer users’ and a low value to ‘fake users’ that are likely to write ‘fake reviews’. In this context a potential ‘fake user’ is someone who joins the community only to promote or disparage a particular business. He is identified by the characteristics of

Table 2. Normalized number of stars

Number of stars	Normalized value
1	-1
2	-0.5
3	0
4	0.5
5	1

⁶ <https://www.nltk.org/>.

having no friends or fans, not being an Elite User and having few reviews related to the same business.

Review success $S(r)$ estimates how much the review r is taken into account by the various users. It depends on how many times the review has been evaluated cool, useful and funny.

Combining the Eqs. (1), (2), (3) and (4), the *opinion value of a review r written by the user u_r* can be computer according to the following equation:

$$Ov(r, u_r) = R(r)(1 + U(u_r))(1 + \beta us_r + \delta f_r + \gamma c_r) \tag{6}$$

where $U(u_r)$ and review success $S(r) = \beta us_r + \delta f_r + \gamma c_r$ are normalized in range $[0,10]$ and $R(r)$ is limited in range $[-1,1]$. The Eq. (6) shows that $R(r)$ is amplified by user value and review success and the sign of $Ov(r, u_r)$ is established by $R(r)$. The additive term $+1$ is introduced to avoid a *null opinion value* in particular situations where the user has a zero value and the review has not been successful. The value of β, δ and γ are set so that $\beta + \delta + \gamma = 1$. The choice of the parameters is based on a tuning phase in with the brute-force or exhaustive search is applied to find the values that maximize review success.

6 Results

In this section the results obtained by applying the presented methodology to a business are reported. In particular the following values are chosen for the parameters in Eqs. 5 and 6: $\alpha = 0.5, \beta = 0.7, \delta = 0.2$ and $\gamma = 0.1$.

Fig. 2 shows the percentage of stars associated with the reviews of the considered business, while Fig. 3 shows the percentage of Positive and Negative reviews.

As seen, Review Value $R(r)$ is computed as average of the number of stars and the value NLTK provides by processing the review text. Figure 4 confirms the strict correlation between review text processed by NLTK and the number of stars associated by the user, showing that ‘what’ users say corresponds to the stars of the reviews. Figure 5 illustrates how public opinion about the business changes with time. As expected by the analysis of the reviews, business public opinion remains positive over time.

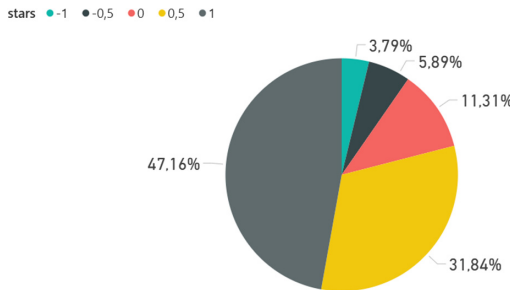


Fig. 2. Percentage of stars associated with reviews.

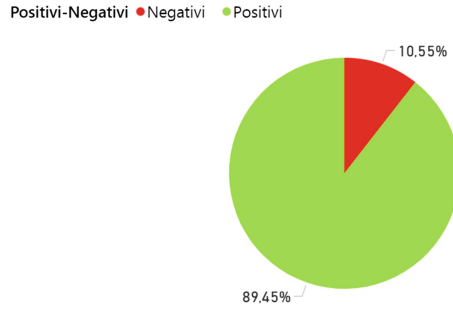


Fig. 3. Percentage of positive and negative reviews.

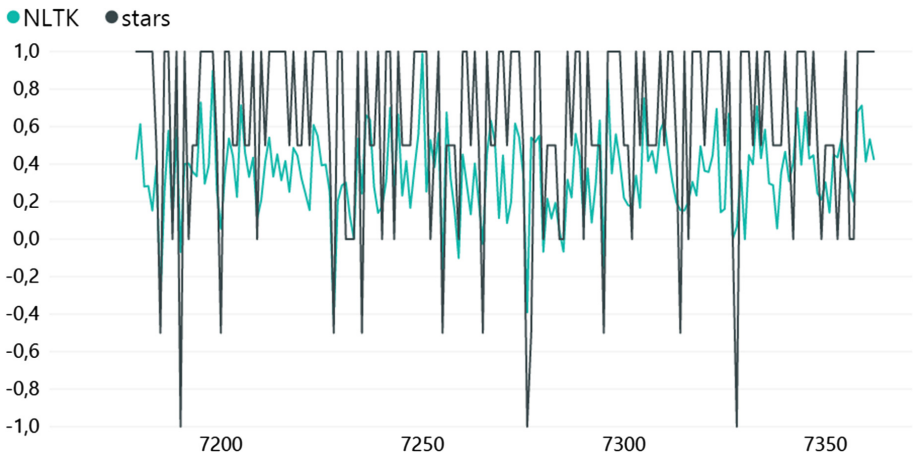


Fig. 4. Illustrates number of stars and NTLK value of each review. X-axis represents review identifier. The most recent reviews are considered.

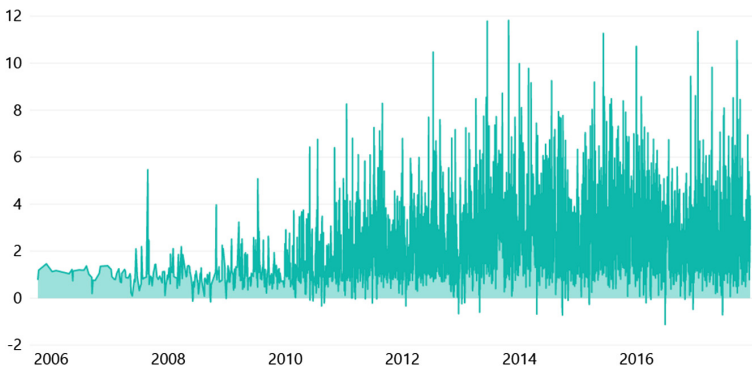


Fig. 5. Illustrates business public opinion over time.

7 Conclusion and Future Works

In this paper we describe a review analysis methodology with the aim to infer business attractiveness combining sentiment and business reputation. The obtained results on the Yelp Dataset Challenge show the effectiveness of the proposed approach.

Future works will be devoted to extend the proposed evaluation to other Online Social networks and to analyze how this approach can be useful for supporting several application as well as Viral Marketing, recommendation and so on.

Acknowledgment. This work is part of the Synergy-net: Research and Digital Solutions against Cancer project (funded in the framework of the POR Campania FESR 2014–2020).

References

1. Bai, T., Zhao, W.X., He, Y., Nie, J.Y., Wen, J.R.: Characterizing and predicting early reviewers for effective product marketing on e-commerce websites. *IEEE Trans. Knowl. Data Eng.* **30**(12), 2271–2284 (2018)
2. Corstjens, M., Umblijs, A.: The power of evil. *J. Advertising Res.* **52**(4), 433–449 (2012)
3. Gunelius, S.: *30-Minute Social Media Marketing: Step-by-step Techniques to Spread the Word About Your Business: Social Media Marketing in 30 Minutes a Day*. McGraw Hill Professional, New York (2010)
4. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. *Bus. Horiz.* **53**(1), 59–68 (2010)
5. Li, Q., Chen, Y., Wang, J., Chen, Y., Chen, H.: Web media and stock markets: a survey and future directions from a big data perspective. *IEEE Trans. Knowl. Data Eng.* **30**(2), 381–399 (2017)
6. Rogers, E.M.: *Diffusion of Innovations*. Simon and Schuster, New York (2010)
7. Shayaa, S., et al.: Sentiment analysis of big data: methods, applications, and open challenges. *IEEE Access* **6**, 37807–37827 (2018)
8. Tuten, T.L.: *Advertising 2.0: Social Media Marketing in a Web 2.0 World: Social Media Marketing in a Web 2.0 World*. ABC-CLIO (2008)
9. Valkanas, G., Lappas, T., Gunopulos, D.: Mining competitors from large unstructured datasets. *IEEE Trans. Knowl. Data Eng.* **29**(9), 1971–1984 (2017)
10. Wang, C.Y., Chen, Y., Liu, K.R.: Game-theoretic cross social media analytic: how yelp ratings affect deal selection on groupon? *IEEE Trans. Knowl. Data Eng.* **30**(5), 908–921 (2017)
11. Weimann, G.: *The Influentials: People Who Influence People*. SUNY Press, New York (1994)
12. Xuan, Q., et al.: Modern food foraging patterns: geography and cuisine choices of restaurant patrons on yelp. *IEEE Trans. Comput. Soc. Syst.* **5**(2), 508–517 (2018)
13. Zhao, G., Lei, X., Qian, X., Mei, T.: Exploring users' internal influence from reviews for social recommendation. *IEEE Trans. Multimedia* **21**(3), 771–781 (2018)