

RESEARCH

Open Access



Towards business partnership recommendation using user opinion on Facebook

Diego P. Tsutsumi^{1*} , Amanda T. Fenerich² and Thiago H. Silva¹

Abstract

The identification of strategic business partnerships can potentially provide competitive advantages for businesses; however, due to the dynamics and uncertainty present in business environments, this task could be challenging. To help businesses in this task, this study presents a similarity model between businesses that consider the opinions of users on content shared by businesses on social media. Thus, this model captures significant virtual relationships among businesses that are generated by users in the virtual world. Besides, we propose an algorithm for detecting business communities in the considered model. We also propose an algorithm to identify possible business outliers in the detected communities, which could represent an automatic way to identify non-obvious relations that might deserve particular attention of business owners. By exploring approximately 280 million user reactions on Facebook, we show that our results could favor the development of, for example, a new strategic business partnership recommendation service.

Keywords: Facebook, User reactions, Collective intelligence, Community detection, Business relation

1 Introduction

Strategic business partnerships are essential for various reasons. For instance, it could favor a competitive advantage for the business. A partnership with a true win-win intention could provide the edge a business needs to surpass its competitors. However, a poorly thought out partnership can hinder instead of help, making this procedure challenging [1, 2].

Recently, businesses from different segments have been exploring social media for various purposes, including for marketing. That involves producing and sharing content on social media platforms to promote a service or product, envisioning to achieve branding goals [3, 4]. Thus, social media platforms become also crucial for customer relationship management, among other things [5, 6].

Those interactions in social media generate a considerable amount of customer-business relationship data. Exploring these data could be an interesting alternative to complement traditional business analysis, such as market

analysis and market segmentation, which, typically, do not scale easily. Since the collection of customer-business relationship data on social media can be cheaper and faster, their proper analysis enables a market study in possibly shorter time and fewer effort [7].

We know that many factors, such as tastes and opinions, could affect the customers' preferences for business [8–10]. We believe that users' preferences for specific business could be implicitly manifested in the actions performed in the content shared by those companies in social media. In some social media platforms, such as Facebook, users can react to the companies content. Thus, these reactions could be a proxy to capture these implicit preferences. Preferences implicitly manifested by users in actions in social media was also assumed to exist in previous works [11–14].

To contribute to the task of identifying strategic business partnership, this study aims to find significant virtual relationships among businesses that are generated spontaneously by users in social media. The first step towards that is by proposing a similarity model between businesses that consider the opinions, i.e., reactions, of users on content shared by businesses on social media. For that,

*Correspondence: diegotsutsumi@alunos.utfpr.edu.br

¹Department of Informatics, Federal University of Technology - Parana, Sete de Setembro Avenue, 3165, Curitiba, Brazil

Full list of author information is available at the end of the article

it is used public data on social media. Particularly, we explored more than 280 million public user reactions collected from Facebook's users about businesses in Curitiba, PR, Brazil. Facebook, among many other social media platforms, has been chosen because it is a popular social media among customers and companies [5].

Besides, we also provide an iterative algorithm for detecting businesses communities in the presented model. A community can be viewed as a cohesive and densely connected subgraph in a larger graph structure; furthermore, algorithms for community detection in graph structures are well known in the literature [15, 16]. In this study, a specific domain algorithm for business community detection is proposed taking into account all the background in the literature.

We found that it enables the identification of business communities that have a surprising similarity regarding categories of businesses inside each community. Even not considering any information of the businesses themselves, all communities have strong semantic similarities in business that compose them, indicating that our approach has the potential to extract cohesive communities of business. We also proposed an automatic approach to extract possible relations outliers, business, on those communities. We observe that non-obvious relationships between businesses could be extracted by using that approach, which might deserve particular attention of business owners.

Our results favor the development of new services and applications. For example, a new strategic business partnership recommendation service that can be useful for entrepreneurs and business owners. This service can contribute to sales improvement, as well as to keep companies more sustainable in the market.

We organize the rest of the study as follows. Section 2 presents some of the main related work to this study. Section 3 describes the particularities of the data collected, which are used in the model proposed in Section 4. Section 5 presents the results obtained by this study. Finally, Section 6 concludes the paper and points out future work.

2 Related work

According to Mukhopadhyay [17], with the advent of online technologies and social media, individuals are increasingly sharing their views and opinions through the internet, which are influencing and affecting the business sociopolitical and personal contexts. A considerable amount of effort has been made by the academic community with the objective of extracting relevant information from social media, which is evidenced by the constant growth of literature. For instance, the average annual growth rate of the number of publications over the last five years is around 0.15 at Scopus and 0.37 at Web of Science

Core Collection¹, and a similar trend is observed in other scientific databases.

The information from social media can be used for different purposes [18], ranging from mobility understanding [10, 19] and well-being improvement [20, 21] to city semantics understanding [22, 23] and gender behavior study [13, 24]. Some of the studies in this direction have implications for existing and new business.

For instance, Cheng et al. [25] presented spatiotemporal analyses of users' displacement, exploring for that a dataset from a Foursquare-like system, i.e., a social media for location sharing. Their results can provide support in decisions about where and when to invest resources in a new business. By observing what people eat and drink in location-based social networks, Silva et al. [12] developed an approach to identify boundaries between different cultures, which could be useful, for example, to businesses from a particular country that desire to verify the compatibility of cultural preferences across different markets.

Cranshaw et al. [11] introduced the concept of Livehoods, which are regions of a city partitioned and grouped by similarity of users' behaviors. This behavioral targeting study may also be necessary for strategic decisions in companies. The geographic interaction characteristics of online public opinion propagation were also studied by Ai et al. [26]. Barbier et al. [27] proposed a method to understand the behavior and dynamics of online groups, showing that their results could have practical business implications, such as a better understanding of customers and influence propagation.

Yang et al. [28] presented the concept of core-periphery structures, that is a two-class partition of nodes (one is the core, and the other is the periphery), and provide empirical evidence that social communities always have this type of structure. This is an important study in the classic task of community detection, in social network analysis, which has implications on the different influences among users within the network. Thus, the core-periphery structures enable identification of such an influential actor, a business in our case, in a community. Regarding community detection, Fortunato [15] formally defines communities and presents a comparison of some algorithms for detecting communities. Rossetti and Gazabet [16] also give a background on community discovery; however, in contrast to [15], they state community structures in dynamic networks.

Wu et al. [29] were also concerned with studying a community detection task, presenting a new framework for detecting parallel communities, called SIMPLifying and Ensembling (SIMPLE). Similarly, Liu et al. [30] used the Markov-network for discovering latent links, i.e., links which are not directly observable but rather inferred, among people in social networks. Data analysis of social

media based on community detection was also used by Alamsyah et al. [31], and some challenges of community detection in social media were presented by Tang and Liu [32]. Social community detection studies are important because they can be applied to networks of different types. In fact, this study presents a business network demonstrating the utility of social community detection in the business context.

Grizane and Jurgelane [33] reinforce the importance of social media as a marketing tool for business, and present a model assessing the benefits of investing financial resources in social media. Mahony et al. [34] investigate the adoption of social media by small business, showing that the use of social media can bring benefits not only to large companies but also to small and medium businesses. Kafeza et al. [35] also focused on assessing business process performance using social media analysis and community detection methods, but with the differential of examining how communities change in time.

An interactive community mapping and detection scheme to reveal the dynamics of communities' evolution around an event is proposed by Giatsoglou et al. [36]. This approach to identifying static and evolutionary communities over time is of particular interest in investigating business dynamism. Dynamic evolution over time was also studied by Pepin et al. [37], who presented an approach based on a graph model easily adaptable to an interactive visualization. Visual analysis can be useful, for example, in the presentation of business communities and how they change over the years.

Considering that the contents of social media are dynamic, Palsetia et al. [38] presented an approach for detecting social communities based on posts, comments, and tweets of users. Their approach is, in some aspects, similar to Algorithm 1 presented in this study, they iteratively remove communities from the main graph making the detection of the current community not influenced by previously detected communities.

Closer to the proposal of this study, two studies were developed to help entrepreneurs and decision makers to find the best place in a city to open a new business, and the core decision process is based on social media data [39, 40]. Regarding the study of Lin et al. [39], business information, such as business type, location, and check-ins, is collected from public pages of Facebook in order to recommend the best places in the city of Singapore to open a new business. Similarly, Karamshuk et al. [40] collected information from Foursquare also aiming to recommend better locations to business, but in contrast to the static data explored in the study of Lin, Karamshuk also considered users' mobility.

The authors of this present study have previously performed a study in this direction [41]. To the best of our knowledge, [41] differs from all other studies available in

the literature, since it aims to identify virtual relationships among businesses that are generated spontaneously by users in social media. To achieve this goal, that study proposes a new way of modeling these relationships, as well as a strategy to extract relevant connections among businesses. Also, it presents a strategy for detecting businesses communities in the proposed model. The present study significantly builds upon our previous work [41]. First, it is proposed in this study a new approach to extract relations outliers on the communities detected. From the outliers, we observe that non-obvious relationships between businesses could be extracted by using that approach, improving the discussion and analysis of communities detected. Also, it is presented important properties of our dataset, and it is discussed more details about the proposed model, for instance, key statistics of the model.

It is important to note that our study could be used in conjunction with some of the previous efforts. For instance, the model proposed by Grizane and Jurgelane [33] could be used in conjunction with the model proposed in this research (explained in Section 4), in order to enrich the information provided to entrepreneurs about the impacts of the use of social media in business.

3 Data collection and processing

3.1 Data Choice

The decision of what data to collect is important to support further analysis, as well as to meet possible limitations imposed in the data collection process. For this study, data were collected from Facebook, because it is the most used social media platform in Brazil [5]. According to Ferrari et al. [5], in Brazil alone the number of Facebook users reached 74.8 million. Facebook is also highly relevant for businesses to create new and maintain ongoing relationships with their customers; therefore, data are widely available for analysis [5].

As the purpose of the model is to identify virtual relationships among businesses exploring user reaction data on Facebook, the data were chosen considering our objectives and what is publicly available on Facebook. Table 1 displays the structure of the considered data. We could choose similar information from other social media platforms; however, this assessment is outside the scope of this present study.

The first column of Table 1, called *Business Data*, represents data referring to the businesses themselves, such as their geographical location, their category (i.e., the market sector in which the business operates) and so forth. Therefore, each business in the dataset has all the information described in the first column. The second column contains *User Reaction Data* for businesses in our dataset. Each reaction expressed on Facebook comes from a user and refers to a particular business.

Table 1 Structure of the considered data

Business data	Example	User reaction data	Example
Business ID	166765230043005	User ID	154573625550
Business name	Rubiane Frutos do Mar	User Name	Fulano de Tal
Location	-25.516122, 49.231571	Business ID that user reacted to	166765230043005
Category	Seafood restaurant	Reaction type	<i>Like</i>
Number of check-ins	38627		
Number of fans	15532		
Average evaluation	4.6		

User reaction data makes it possible to create similarity connections among businesses, primarily by using the common reactions expressed by users regarding two or more businesses, as explained in Section 4. The business data is used mainly to extract names and categories of the businesses, which assists the reaction collection and evaluation of the results presented in Section 5.

All information collected is open and publicly available on the Facebook platform. More details can be found on the Facebook Graph API².

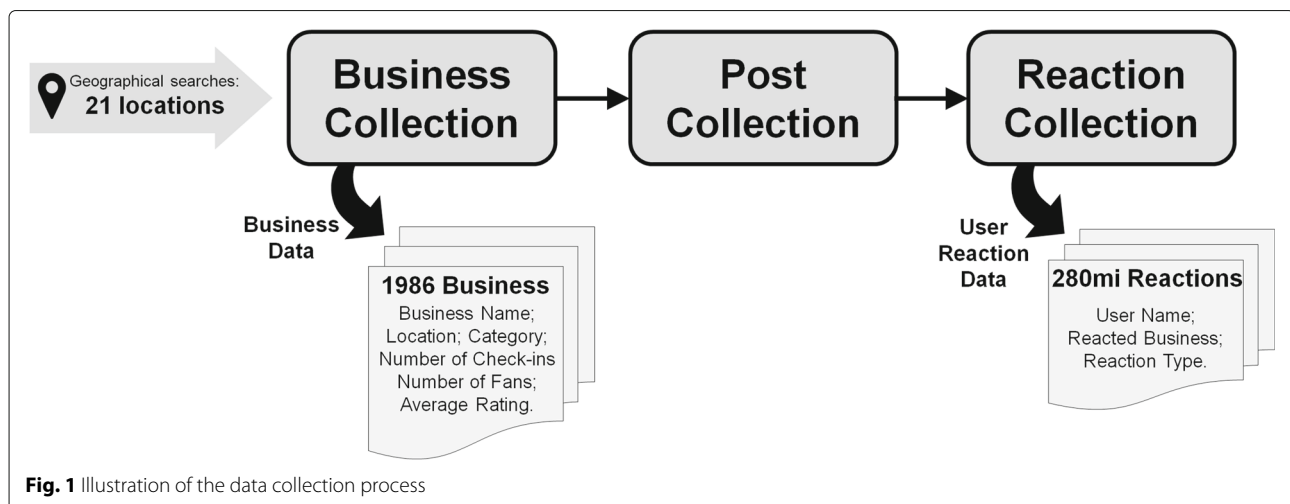
3.2 Data collection

Figure 1 illustrates the main steps performed in the collection process. The data were collected using the Facebook Graph API³, all business data collected are located in the city of Curitiba, Brazil, and all user reaction data are from November to December of 2017. First, we collected data referring to the first column of Table 1, i.e., *Business Data*.

Facebook Graph API requires a geographical coordinate and a radius in meters, then returns results considering the coordinate entered as the center of a circle with the radius informed, returning up to 800 results per search, that is, up to 800 businesses in this case. It is known that several regions of the studied city may have this number of businesses, then to increase the chance of getting most businesses of all regions of Curitiba, we considered twenty-one different geographic searches throughout the city. Each geographic search has a radius of 2000 meters and is centered in different regions of Curitiba, as shown in Fig. 2.

Figure 3 shows a heatmap representing the number of business found in different regions by the collection process just described. The redder the color, the more business were found in that particular location. As we can see, the central region of the city has a reddish coloration, indicating that more businesses were collected in that region, as expected. Despite that, it is also possible to notice that the resulting dataset includes businesses spread all around the city of Curitiba. Figure 15 in Appendix A shows how user reactions are distributed across regions of the city, as expected, it is easy to notice that the amount of reactions is more significant in the city center.

After obtaining the results of the geographical searches (*Business Data* in Table 1), containing basic data of the businesses in Curitiba, the reactions of the users (*User Reaction Data* in the Table 1) were collected from the business pages previously collected. For that, we obtained the reactions of the posts on the business pages. Because some businesses' pages have hundreds of posts and others have millions, we only collected reactions of the first one hundred posts. There are five types of reactions available in Facebook, namely *Like*, *Angry*, *Wow*, *Sad* and *Thankful*; we included all types in the database.

**Fig. 1** Illustration of the data collection process

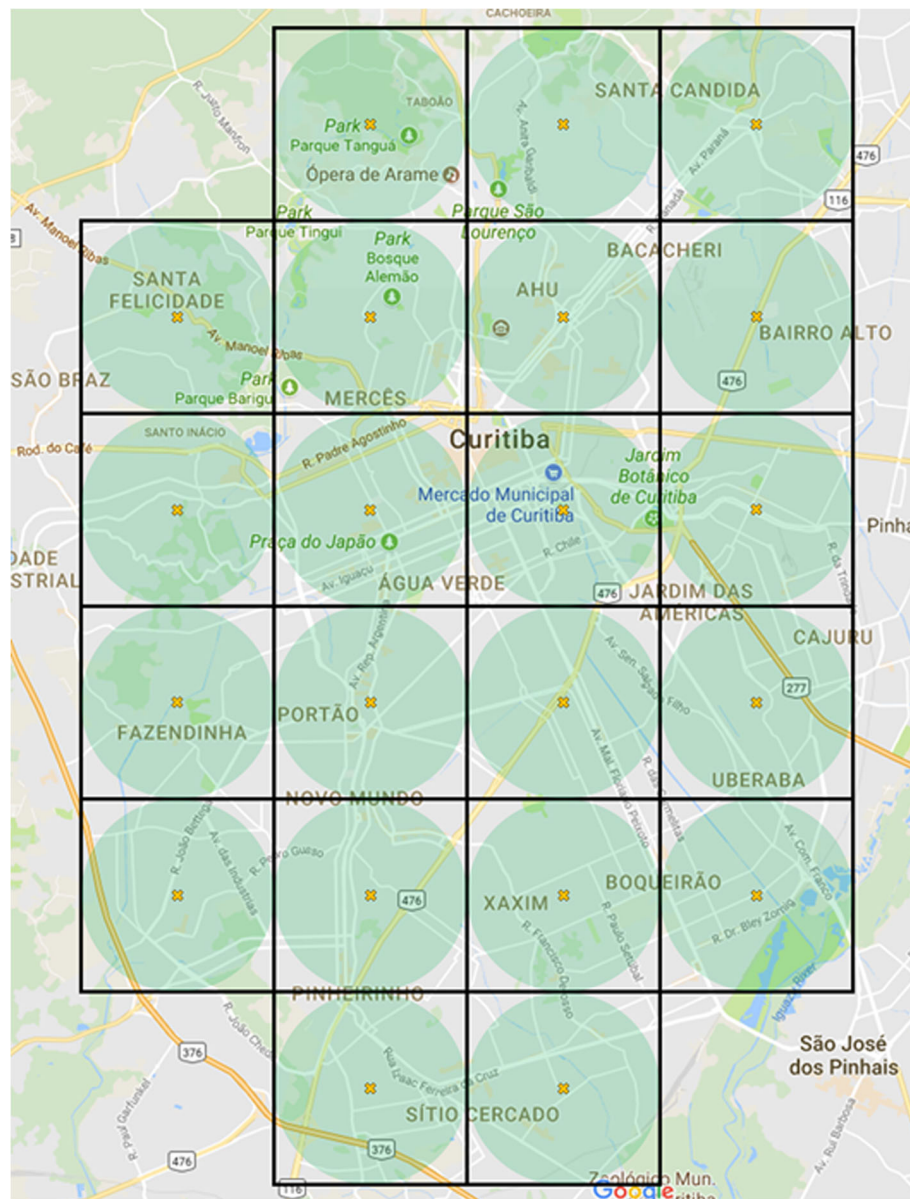


Fig. 2 Data collection points considered for Curitiba, PR, Brazil

We collected a total of 1986 georeferenced pages and approximately 280 million user reactions related to those pages. In Appendix A it is presented supplementary information about this dataset. Figure 13, shows the top twenty businesses regarding user reaction number, in addition, Fig. 14 shows the top twenty businesses categories concerning user reaction number.

3.3 Data cleaning

After obtaining all the data, an automatic and manual cleaning procedure was performed to increase the consistency of the dataset, consequently increasing the

consistency of the final results. We performed three main steps:

- Duplicate records removal (automatic procedure);
- Inconsistent records (e.g., unnamed pages, without location, and) removal (automatic procedure);
- Removal of pages that do not represent businesses (for example, a public square) and their reactions (automatic and manual procedure).

After the cleaning process, the dataset is left with 1926 pages, all representing businesses, and approximately 260 million reactions.

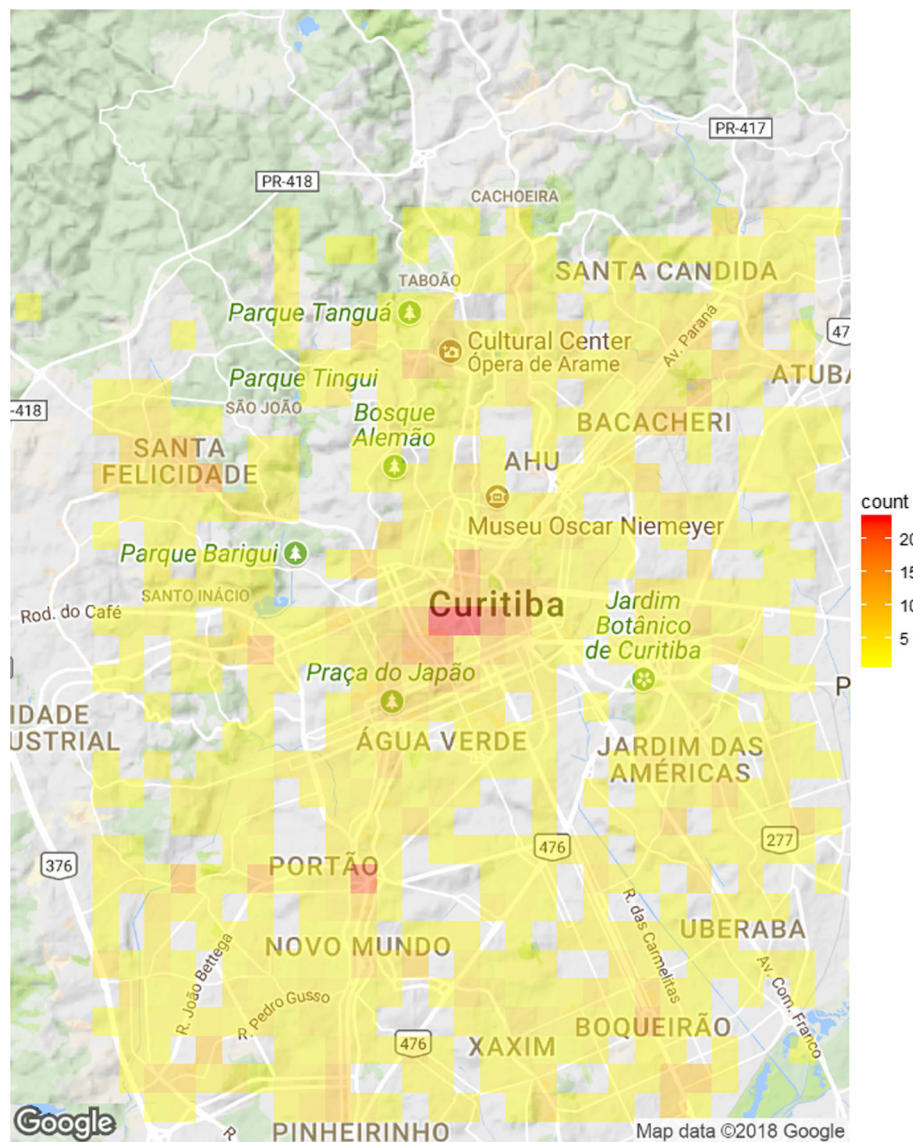


Fig. 3 Heatmap representing the number of businesses found in different regions by the business search process

4 Modeling and strategies for data analysis

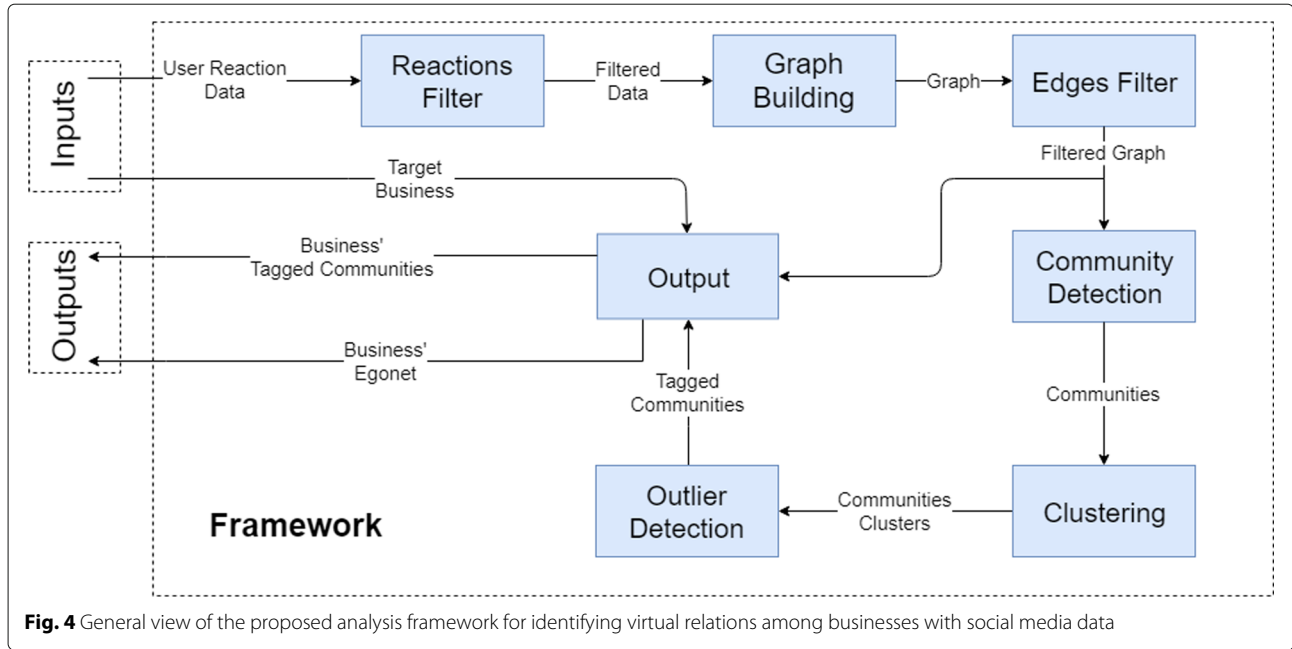
4.1 Overview

The main steps employed in this study to achieve the proposed objectives can be described in a framework, proposed by the authors of this study, illustrated in Fig. 4. The framework entries are the *Clean Data*, described in Section 3, and a *Target Business*, the business chosen to be analyzed. As outputs, we obtain the *Egonet of the Target Business*, a network of the most relevant direct connections of the target business, and the *Business Tagged Communities*, tagged communities of businesses in which the target business is part of. All non-standard businesses inside communities, considered outliers, are tagged. We

provide more details about those outputs next. A relevant feature of this framework is that it is designed to handle data from any data source. In this paper it is used Facebook data; however, this is not a restriction of the framework.

4.2 Business relationship graph

With the obtained data, we can then create a model to represent the virtual relations among businesses. The model is a non-directed graph in which vertices represent businesses, and weighted edges represent relations between two businesses. This relationship is built looking at the reactions of users in common between any two



businesses. The more common reactions two businesses have in proportion to their reactions, the stronger the relationship between them. Thus, we weight an edge by the Jaccard Index of the set of reactions of each business, representing an index of affinity or similarity between the two sets.

In a more formal way, consider $B = \{b_1, b_2, \dots, b_{n_b}\}$ being the set of all businesses, where n_b is the total number of businesses in the dataset. Now consider U_i being the set of all users who reacted to the i -th business. Thus, the graph is defined as in (1):

$$\text{BusinessGraph} = (V, E, W), \quad (1)$$

where vertices are businesses, $V = B$; edges exist if businesses have a minimum of users' reactions in common, $E = \{(i, j) : |U_i \cap U_j| > \text{lowerBound}\}$; and the weights of the edges are represented, as in Equation (2), by the Jaccard Index:

$$W(i, j) = \begin{cases} \frac{|U_i \cap U_j|}{|U_i \cup U_j|} & \text{if } (i, j) \in E \\ 0 & \text{if } (i, j) \notin E \end{cases} \quad (2)$$

4.3 Filters and graph consistency

In order to increase the consistency of the information about the graph structure, it is considered two essential filtering steps: a reactions filter, and a weak edges filter.

First, the reaction filter eliminates negative reactions (of the type *Angry* and *Sad*), because for possible partnerships between businesses what matters are positive reactions. Then, the reaction filter eliminates users who do not frequently express themselves about business in B .

So users with two or fewer reactions are eliminated from the model, leaving the filter lower bound as 3 reactions.

On the other hand, the filter also eliminates users with too many reactions for two reasons. (i) The proportion of the number of edges a that a user with m reactions generates in the graph is quadratic $a = \frac{m(m-1)}{2}$, therefore, for instance, users with 500 reactions generates 124,750 edges, unbalancing their influence over users with few reactions. (ii) Users with too many reactions could be robots (*bots*), a problem that appears in several Web systems [42]. In order to maintain a fair balance between representativeness over potential problems, 99.9% of the reactions (for users with 3 or more reactions) in the original database are kept. The filter upper bound is calculated counting the proportion of reactions R_{3-x} (reactions from 3 to x) over $R_{3-\infty}$ (all reactions above 3), as in Eq. 3.

$$\frac{R_{3-x}}{R_{3-\infty}} = 99.9\% \quad (3)$$

As depicted in Fig. 5, which shows the distribution of total users per number of reactions, the filter upper bound calculated for the studied dataset is $x = 174$; therefore the reaction filter considers only users who reacted from 3 to 174 times. After the reaction filtering process, the resulting dataset has 220 million reactions.

Once the reactions are filtered, the weak edge filter removes possible noises in the graph structure, thus removing edges classified as weak edges. We classify an edge as a weak edge by performing a random experiment, in which the typical reactions are randomly and uniformly

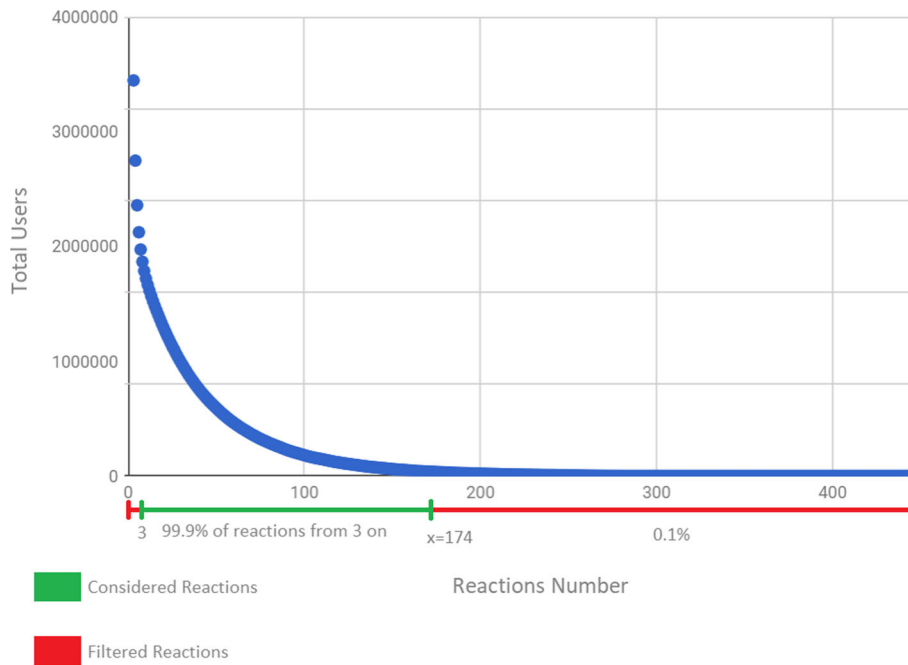


Fig. 5 Distribution of the number of users by number of reactions

distributed among all possible edges in the graph, forming a random graph. Then, after this simulation process, the edges of the original structure with similar weight to the experiment's graph can be considered weak edges.

When reactions are uniformly distributed among all possible edges, the weight of any edge in the random graph follows a binomial distribution. The expected value and the variance of the weight of any edge in the random graph (with binomial distribution) are, respectively, given by Eqs. (4) and (5).

$$\mu = E[X] = \frac{n_r}{n_c} \quad (4)$$

$$\sigma^2 = Var[X] = \frac{n_r}{n_c} \left(1 - \frac{1}{n_c}\right) \quad (5)$$

Where, $n_r = \frac{\sum_{i,j:i \neq j} |U_i \cap U_j|}{2}$ is the sum of weights in the original graph, $n_c = \frac{n_b(n_b-1)}{2}$ is the number of all possible edges and n_b is the number of businesses in the dataset.

In this way, an edge between businesses i and j is weak if: $|U_i \cap U_j| \leq lowerBound$, as defined in Eq. (6), following the 3σ statistics for the binomial distribution, which includes 99.73% of random edges.

$$lowerBound = \mu + 3\sigma \quad (6)$$

For the data collected in this study the calculated values were: $\mu = 120.289$, $\sigma = 10.96$. Thus, $lowerBound = 153.195$. Considering the dataset of this study, 978,410 edges were eliminated, resulting in a total of 223,939 edges in the graph with less probability of being random noise.

4.4 Detection of business communities

Given a consistent network of business relationships, an essential step in achieving the study's goal is to detect business communities. As the business graph diameter is 4, therefore not sparse considering the number of nodes and edges, community detection algorithms based on searching cliques or dense subgraphs with optimal solution, such as the Clique Percolation Method, have a very high computational time and space complexity; therefore they are not applicable in this study.

Raghavan et al. [43] proposed a community detection algorithm based on label propagation (LP), which iteratively uses the exchange of labels between adjacent vertices in such a way that promotes convergence of labels. One significant advantage is that this algorithm operates in almost linear time, which makes it tractable for dense graphs. Also, since previous information about the graph and communities are not available, another advantage is that this LP-based algorithm does not need previous information, such as heuristics of communities present in the input network, required in other methods cited in [43]. It is also true that LP-based algorithms sometimes give different solutions due to the random steps performed. To enhance the stability of the final solution this LP-based algorithm has aggregation steps to combine different solutions.

For the problem addressed here, it is interesting that communities have at least four businesses (*minSize*), and a maximum of thirty businesses (*maxSize*), since huge communities lose cohesion in possible recommendations.

Therefore, an iterative algorithm, Algorithm 1, is proposed for the detection of communities of businesses. The entries of this algorithm are the business graph (*BusinessGraph*), the minimum size (*minSize*) and maximum size (*maxSize*) of the communities, and the output is a set of business communities. The algorithm performs the following key steps:

- Detection of communities with the algorithm described in [43];
- Business communities with size of *minSize* and *maxSize* are saved for final return;
- The union of detected communities, and not saved for return, composes a new graph, named *G*;
- From this new graph *G*, weak edges are cut forming the graph for the next iteration.

Algorithm 1: Business Communities Detection Algorithm

Data: *BusinessGraph*, *minSize*, *maxSize*

Result: Set of Communities of *BusinessGraph*

```

1 G ← BusinessGraph
2 allCommunities ← ∅
3 minEdge ← mini,j ∈ G W(i, j)
4 counter ← 1
5 while |G| > minSize do
6   counter ← counter + 1
7   /* Method of [43] */
8   detectedComm ← labelPropCommDetection(G)
9   G ← empty graph
10  for c ∈ detectedComm do
11    if |c| > minSize and |c| < maxSize then
12      | allCommunities ← allCommunities ∪ {c}
13    else
14      | G ← G ∪ c
15    end
16  end
17  remove all edges of G with
  W(i, j) < minEdge * counter
18  if no edges removed from G then
19    | break
20  end
21 end
22 return allCommunities;

```

The communities detected according to Algorithm 1 are subgraphs that tend to be dense (many edges), so businesses within the same community have a greater cohesion than businesses randomly chosen in the graph.

This cohesion is formed by spontaneous user reactions, without additional information that could include a bias in communities detected. The time complexity of Algorithm 1 is $O(|V| + |E|)$, where $G = (V, E)$.

4.5 Clustering of business communities

In this section, it is described the steps to cluster business communities by their similarity. As we are dealing with Facebook data, the clustering process is done considering categories of businesses given by Facebook. However similar processes could be done with data from any other social media platform.

Facebook classifies all businesses into one of seven categories: *Interest*; *Community Organization*; *Media*; *Public Figure*; *Businesses*; *Non-Business Places*; and *Other*. All of them have subcategories, but the larger number of subcategories (22), as well as the greater diversity, are subordinated to the *Business* category. Therefore, we considered all subcategories within *Interest*, *Community Organization*, *Media*, *Public Figure*, *Non-Business Places*, and *Other* as their parent categories. For instance, all subcategories of *Interest* are transformed into *Interest*.

For the *Business* category, all subcategories have been considered. Under each of them, there are sub-subcategories. The sub-subcategories of *Business* were disregarded, as this level of specialization was not considered interesting for the analysis carried out. The *Advertising or Marketing* subcategory of *Business* has, for example, the sub-subcategories *Advertising Agency* and *Copywriting Service*. In this case, all sub-subcategories within *Advertising or Marketing* are considered *Advertising or Marketing*. We performed the same procedure for all other subcategories within *Business*.

After doing this process, a total of 28 business' category names (6 from the parent categories, and 22 subcategories from *Business*) were obtained, as illustrated in Fig. 6. We then built a feature vector with these 28 categories and counted for each community the number of occurrences of businesses in each of the 28 categories, as in Fig. 7. These values are then normalized based on the maximum number of locations in a given feature for each community. With this feature vector (represented formally as *vector*(*c*) for a given community *c*), it is possible to identify the similarity of different communities and perform a clustering process. The clustering algorithm used was the k-means, with the Euclidean distance [44]. For choosing the right *k* parameter of k-means algorithm, several different values were tested, and the value with the smallest sum of squared errors was chosen as the best fit for the data. For the dataset presented in this paper, the best fit was *k* = 8; thus this value is kept for this study.

Formally, k-means receives a set of all communities (called *allCommunities* as in Algorithm 1) and returns

Facebook Categories Hierarchy

1. Community Organization

- 1.1 Religions Organization
- 1.2 Social Club
- ...

2. Food & Beverage

- 2.1 Restaurant
- 2.2 Bakery
- ...

3. Automotive, Aircraft & Boat

- 3.1 Motor Vehicle Company
- 3.2 Automotive Dealership
- ...

... (28 Categories)

Fig. 6 Facebook subcategories renaming process

a set of clusters, where clusters are disjoint non-empty subsets of *allCommunities*:

$$clusters = kmeans(allCommunities),$$

$$\forall cl_1, cl_2 \in clusters; cl_1 \neq cl_2, \quad cl_1 \cap cl_2 = \emptyset, \quad cl_1 \neq \emptyset \text{ and } cl_2 \neq \emptyset.$$

4.6 Business outlier detection

To detect possible businesses outliers inside business communities, first a clustering of business communities, as described in Section 4.5, has to be performed. In possession of clusters of business communities, the detection is based on *cluster centroids*, which represent the common proportion of business categories for each cluster, and are a central piece in the business outlier detection. In a high level of abstraction, if communities differ significantly from its cluster centroids, they have some outlier businesses inside them. Formally, the cluster centroid is the average vector ($\in \mathbb{R}^{28}$) among all the communities from that cluster, as in Eq. 7:

$$\forall cl \in clusters, \quad centroid(cl) = \frac{\sum_{c \in cl} vector(c)}{|cl|} \quad (7)$$

Cluster centroids are used to build *cluster signatures*. A cluster signature is a set of the most representative business categories for a particular cluster. Each category (dimension) inside the cluster centroid vector represents a certain percentage of all categories present in the vector so that all percentages sum up to 100%. Therefore, picking the categories with the greatest percentages so that their percentages sum up to a threshold ($> 50\%$), makes them the set of the most representative categories for that cluster. For example, a cluster could have as its signature the categories *Food and Beverage*, *Shopping and Retail* and *Entertainment*, because they meet a 70% threshold defined in a particular application.

More formally, Algorithm 2 specifies how to capture a *cluster signature* given two inputs: A vector $v \in \mathbb{R}^n$, representing the centroid, where n is the total number of categories; and a *threshold* $\in]0.5, 1]$, a number representing the minimum percentage considered as majority.

	Community Organization	Food & Beverage	Automotive, Aircraft & Boat	... (28 categories)
Community 1	2	6	7	...
Community 2	7	1	1	...
...
Community n	1	14	4	...

Fig. 7 Examples of non-normalized feature vector for each community

As a result, Algorithm 2 returns a set containing the greatest dimensions, i.e. categories, of the vector v that corresponds to the closest possible number to $threshold * 100\%$ of all dimensions. Its time complexity is $O(|v|^2)$. Finally,

Algorithm 2: Function that returns the signature (greatest dimensions) of the vector

```

1 Function getSignature( $v, threshold$ ):
2    $s \leftarrow \sum_i v_i$ 
3    $accThrs \leftarrow 0$ 
4    $signature \leftarrow \emptyset$ 
5   for  $i \in \{1, 2, \dots, |v|\}$  do
6      $m \leftarrow \max(v)$  /*Max value in vector*/
7      $j \leftarrow \operatorname{argmax}(v)$  /* Category (index) of the
        vector's maximum value*/
8     if  $|\frac{m+accThrs}{s} - threshold| < |\frac{accThrs}{s} - threshold|$  then
9        $signature \leftarrow signature \cup j$ 
10       $accThrs \leftarrow accThrs + m$ 
11       $v_j \leftarrow 0$  /*In next iteration, max(v) is the
        next greatest*/
12    else
13      break
14    end
15  end
16 return signature

```

Algorithm 3 is responsible for tagging all businesses which categories are not included in its corresponding cluster signature, therefore considered as outliers. This algorithm takes as input a set of clusters and returns the same input structure, however with additional outlier tags. Note that firstly it calls Algorithm 2, to get each cluster signature, and next it tags businesses which categories are not in its cluster signature. Note that each community has a vector (as in Fig. 7), and each cluster has a centroid (as in Eq. 7), both are captured in Algorithm 3 as $vector(community)$ and $centroid(cl)$, respectively. In order to capture the most representative categories of each cluster, the function *getSignature* (Algorithm 2) is called. For this function, if the *threshold* value is high (e.g., 0.9), then too many categories are considered in the cluster signature, on the other hand, if the value is low (e.g., 0.5) then there is a chance that no category is considered in the cluster signature. According to our empirical analysis, the value 0.7 represents a good balance; therefore, we consider it as the threshold. The time complexity of Algorithm 3 is $O(|clusters| * |cat|^2 + |allCommunities| * |cat| + |V|)$, where $G = (V, E)$, cat is the set of all categories, and

Algorithm 3: Outlier Detection Algorithm.

Data: Clusters - set containing business community clusters

Result: Clusters with tagged businesses

```

1  $taggedClusters \leftarrow \emptyset$ 
2 for  $cl \in clusters$  do
3    $newCluster \leftarrow \emptyset$ 
4    $clSignature \leftarrow getSignature(centroid(cl), 0.7)$ 
5   for  $community \in cl$  do
6      $vc \leftarrow vector(community)$ 
7     for  $i \in \{1, 2, \dots, |vc|\}$  do
8       if  $vc_i > 0$  and  $i \notin clSignature$  then
9         /*tagBusiness(community, cat) tags
           businesses of category cat inside
           community*/
           $newCluster \leftarrow newCluster \cup$ 
             $\{tagBusinesses(community, i)\}$ 
10        end
11      end
12     $taggedClusters \leftarrow taggedClusters \cup \{newCluster\}$ 
13  end
14 end
15 return taggedClusters

```

allCommunities represent the set of all communities being studied.

5 Results and discussions

The two outputs of the framework are the tagged communities, i.e., tagged as an outlier or not, that includes the target business (informed by the user), and the *egonet* of the target business, consisting of a subgraph of all edges of the target business (see Fig. 4). As there are considerably large *egonets* for individual businesses, the *egonet* size in this study was limited to a maximum of seven adjacent vertices (with the strongest edges) plus the target business. Note that this parameter could be adjusted for each case under study.

Figure 8 illustrates the complete graph, constructed by the proposed framework, shown on a map of the city of Curitiba. To ease the visualization, we only show edges with more than 1,500 common reactions. The thicker the edge, the stronger the relationship between nodes. We show each node in the graph according to the location of the business it represents. Note that there are vertices far from the city center with a considerably high density of edges going towards the city center, indicating a strong activity also in outlying neighborhoods.

An important detail was captured during analysis. The nodes “Prefeitura de Curitiba” (N1) and “RPC” (N2) have a strong influence on business graph (as shown in

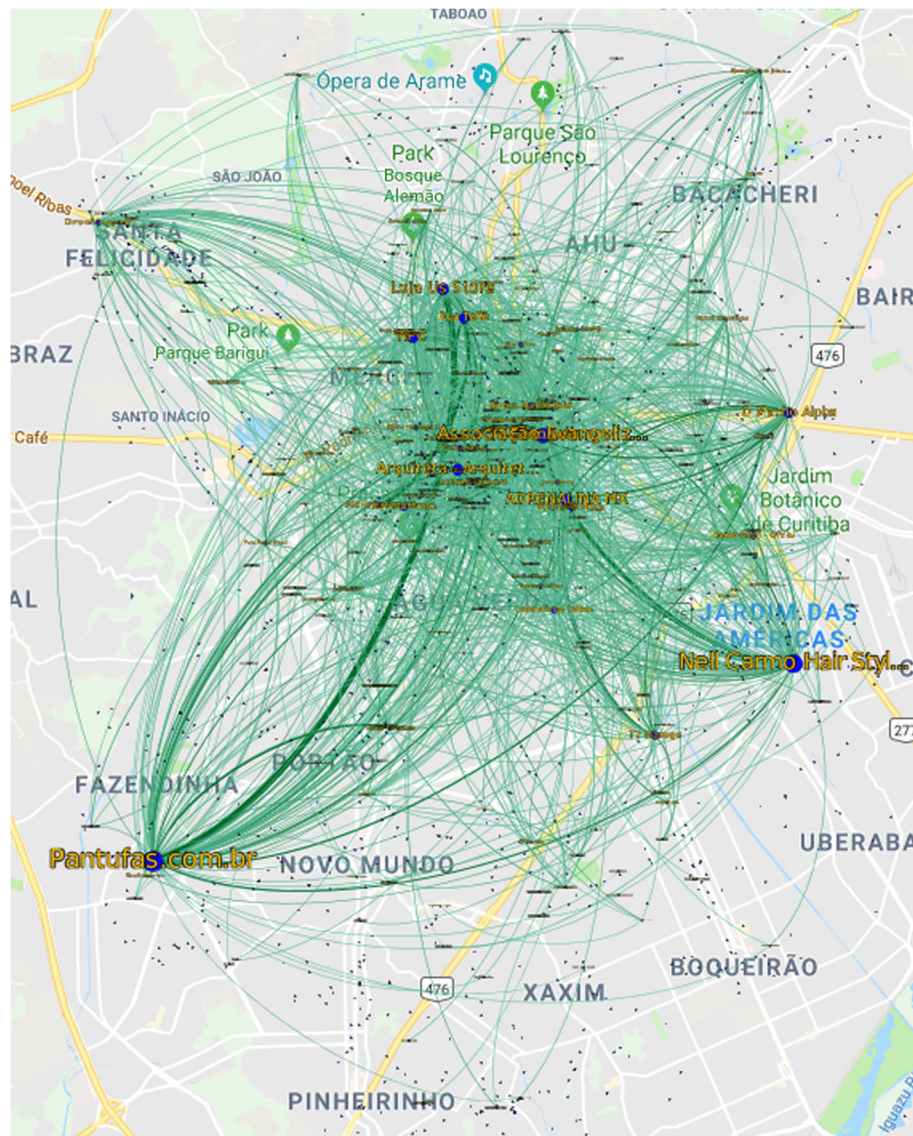


Fig. 8 Partial view of the business graph on the map of Curitiba

Appendix A by Tables 2 and 3). N1 represents the city hall of Curitiba, and it is a very popular Facebook page⁴ not only in the city of Curitiba but also nationally. N2 is the largest TV channel in Curitiba. Both strongly impact the analysis carried out by this study. N1 does not represent a business and N2 is an obvious TV media partnership. Their edges influence on the business relationship graph potentially hides interesting smaller business relationships. In order to favor more valuable relationships, the analysis was performed without both nodes.

As this graph is quite large, the extraction of useful information becomes complicated to human eyes, justifying the extraction of communities and *egonets*. After

running Algorithm 1 with the parameters considered, 144 communities were detected, each ranging from 4 to 30 businesses located in the city of Curitiba. Figure 9a illustrates a community containing entertainment businesses (e.g., “Blood Rock Bar”, and “SSCWB - Shinobi Spirit”) and food businesses (e.g., “Ca’dore Comida Descomplicada”), so they are businesses united by the “leisure” context. The two communities illustrated in Fig. 9b and c both have businesses bound together by the context that can be called “fashion”, because it contains businesses from the beauty salon sector (e.g., “Studio Andressa Mega Hair” and “Cheias de Charme Costméticos”), modeling agencies (e.g., “Nk Agencia de Modelos” and “South Models

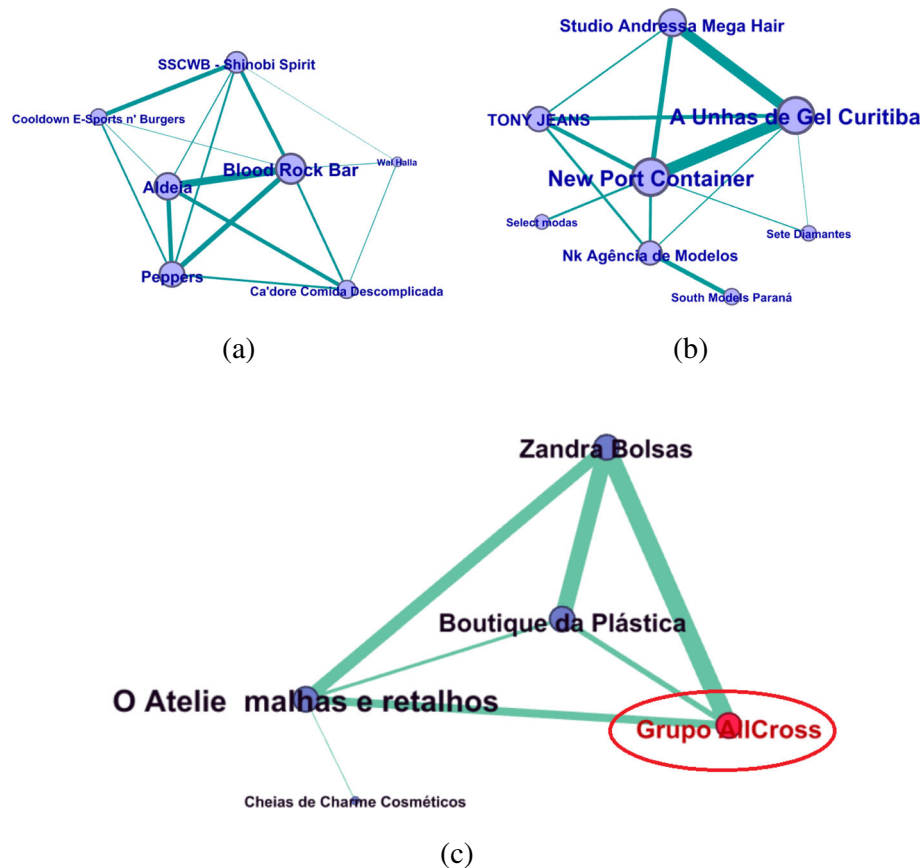


Fig. 9 Communities detected by Algorithm 1. **a** Community of businesses related to leisure. (No businesses tagged as outliers). **b** Community of businesses related to fashion. (No businesses tagged as outliers). **c** Another Community of businesses related to Fashion. The business tagged in red is a health plan related business and it was considered an outlier by Algorithm 3

Parana”) and fashion stores (e.g., “TONY JEANS” and “Zandra Bolsas”). We can note that, even though both the business network construction (see Section 4.2) and the Algorithm 1 did not use any information of the businesses themselves, all communities detected have similar strong semantics that binds businesses together inside each community.

The business category clustering analysis can illustrate those contexts in a more general view, considering all communities detected. The clustering step, then, unites all similar communities, by business categories, in eight different clusters (for $k = 8$ as discussed in Section 4.5). To illustrate the clusters, eight word clouds of businesses’ categories, inside each cluster, were generated and shown in Fig. 10. Besides, Fig. 11 shows the number of user reactions in each cluster.

In those word clouds, we did not perform the sub-category renaming process made to execute K-means, so we considered the original names. Note the surprising similarity between the categories in each group. For

example, Cluster 1 is related to leisure, containing predominantly food, drink and entertainment businesses, Cluster 2 contains most businesses related to beauty and style, while Cluster 3 is more related to establishments about automotive products and services. This analysis shows the existence of a predominant context in each community. Taking into account information in Figs. 10 and 11, it is possible to notice that the most popular contexts are leisure and food, represented by Cluster 1, followed by shopping malls, represented by Cluster 7.

Knowing that there is a tendency of having a predominant context of business in communities, outliers, i.e., business outside the predominant type of business, can be useful for decision makers. Next, we present results in this direction following the procedures described in Section 4.6.

The community illustrated in Fig. 9c, which is a fashion related community (its predominant context), has one outlier inside it, which is the business called “Grupo AllCross” (tagged in red). This business is a health

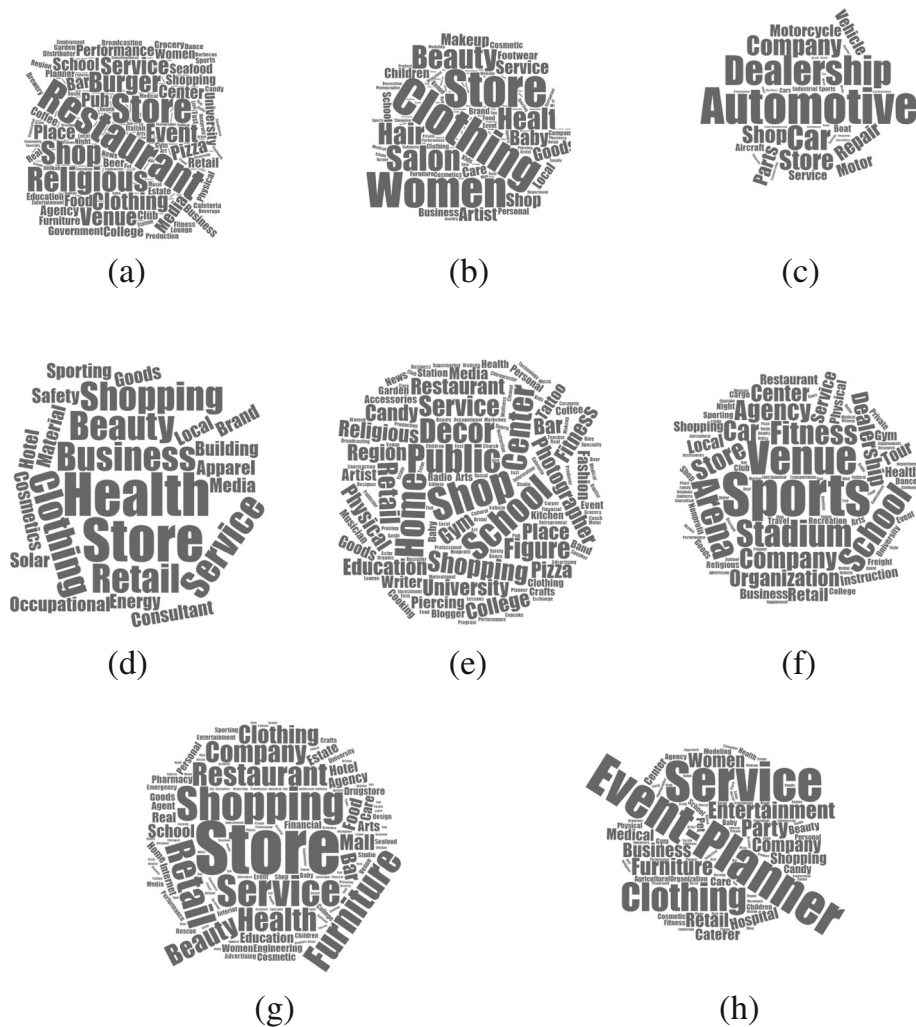
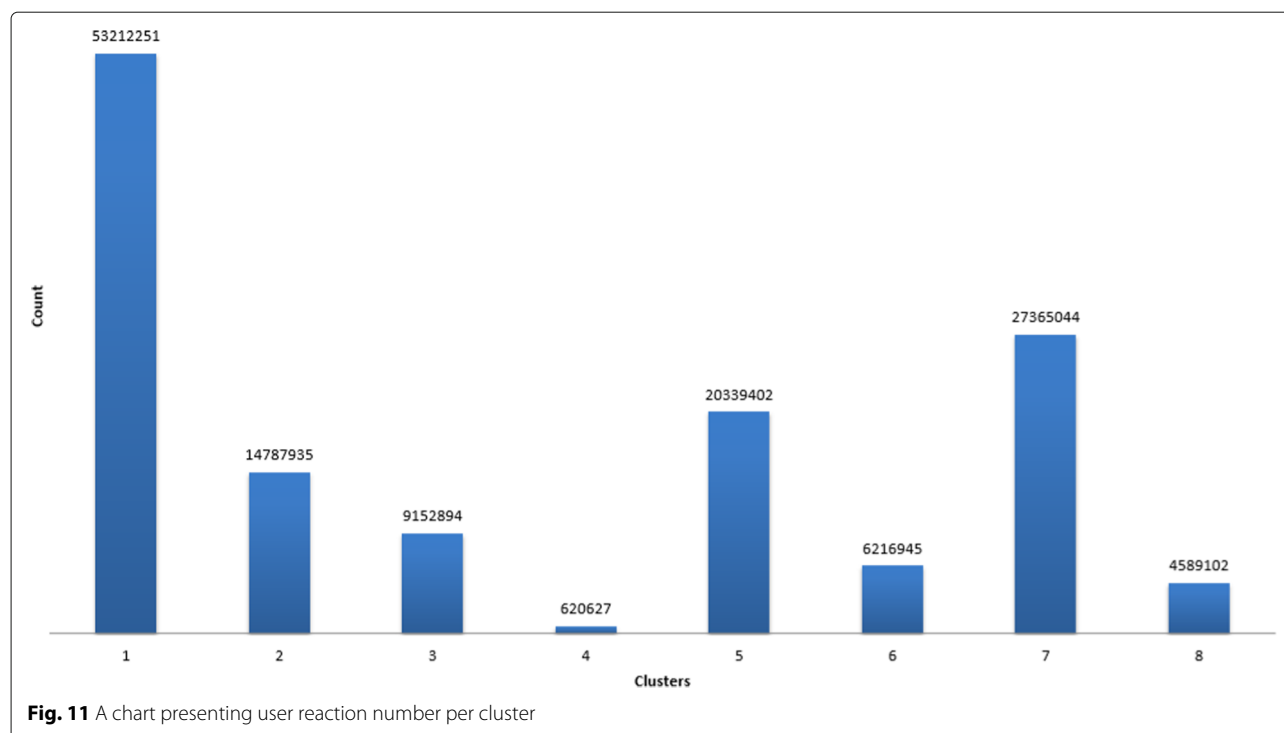


Fig. 10 Word clouds for categories of similar community clusters. Each cluster legend contains the three biggest super categories of its signature according to Algorithm 2. **a** Cluster 1. Arts & Entertainment; Local Service; Media News Company. **b** Cluster 2. Beauty Cosmetic & Personal Care; Shopping & Retail; Other. **c** Cluster 3. Automotive, Aircraft & Boat; Commercial & Industrial; Non-profit Organization. **d** Cluster 4. Media; Hotel & Lodging; Other. **e** Cluster 5. Education; Public Figure; Other. **f** Cluster 6. Sports & Recreation; Travel & Transportation; Advertising Marketing. **g** Cluster 7. Science, Technology & Engineering; Hotel & Lodging; Advertising Marketing. **h** Cluster 8. Local Service; Shopping & Retail; Medical & Health

plan consultant business, being not part of the “fashion” context and, thus, correctly identified as an outlier by Algorithm 3. As an improvement of the results in [41], outliers cannot be ignored in the results presented here, as they might represent non-trivial potential business partnerships. Although outliers are not part of the dominant context, they still have strong connections to businesses from that context.

Figure 12a shows the *egonet* of Rubiane, a seafood restaurant, which was arbitrarily chosen for analysis, and Fig. 12b shows a detected community in which Rubiane is included. On the one hand, having the business’ *egonet*, it is possible to visualize the direct connections that the

target business possesses with other businesses. On the other hand, having communities, it is possible to notice connections that may not be direct to the target business. Since these non-direct connections are within a community (detected by the Algorithm 1), they are cohesive (a dense subgraph) and may represent possible non-trivial partnerships for the business under evaluation. For example, the company “Quintal do Monge” does not appear in the Rubiane’s *egonet* shown in Fig. 12a, but it appears in a community where Rubiane is also included, shown in Fig. 12b. Also in Fig. 12b notice that the business called “Cannes Turismo” is a tourism related business and was tagged as an outlier by Algorithm 3.



Rubiane, for instance, could make use of this result to increase its sales by creating business partnerships, such as selling products and services along with the businesses found in the results, as well as marketing partnerships and joint marketing campaigns. For the case of the restaurant analyzed here, we observe that competitors appeared in the same community, for example, “Braseirinho Frutos do Mar”. For the case involving restaurants, this could be explained by the fact that users tend to attend several restaurants and some may be of the same type. However, this is not a problem with the proposed approach, since the entrepreneur is who decides the best strategy of how to explore the results. Note that a partnership could be made with competing establishments. However, these cases deserve special attention.

6 Conclusion and future work

The approach introduced in this study aims to provide a new way of identifying significant and non-trivial relations between business, which could ease the laborious task of strategic business partnerships recommendation. This study shows, using large-scale data from Facebook, that the proposed approach could be an important building block for the development of new applications and services, including a business partnership recommender. Furthermore, the presented results and discussions show that the data available in social media and other platforms

are indeed helpful for us to understand the dynamics of the world around us.

Considering this study as a basis for further research, there are many directions to follow. For instance, we observe the existence of competing businesses in the same community. As one of the possible implications of this study is to contribute to identifying new business partnerships, it may be interesting to determine a way to detect whether two businesses are competitors to improve the performance of possible recommendations. Besides, it is interesting to evaluate the results for a new dataset, especially in one representing a different culture. In addition, it is essential to perform a qualitative evaluation considering business owners or decision makers of the studied businesses. This is essential to understand how to explore the results in practice better. Another direction is to consider the temporality of the reactions, to evaluate, for example, the temporal correlation in the communities.

Endnotes

¹ These values represent average growth using the term “Social Media”, and “Social Media” and “Business” in conjunction.

² <https://developers.facebook.com>.

³ <https://developers.facebook.com/docs/graph-api/overview>.

⁴ <https://www.facebook.com/PrefsCuritiba>.

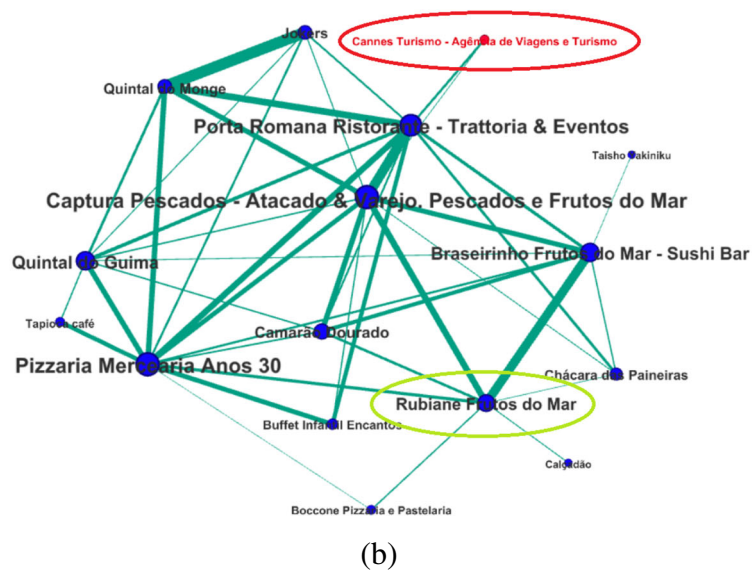
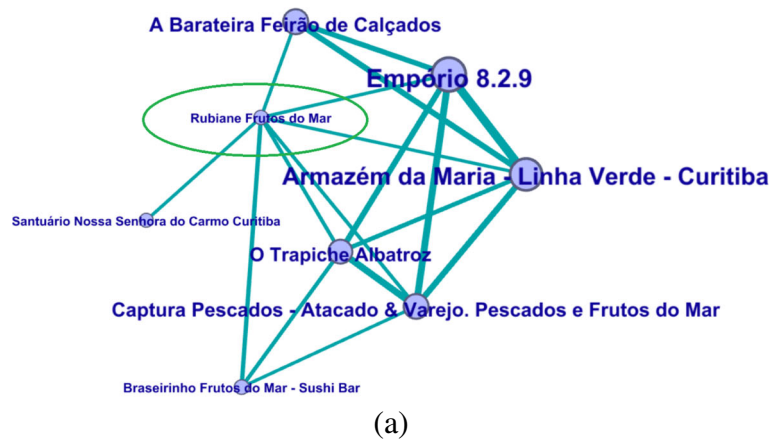
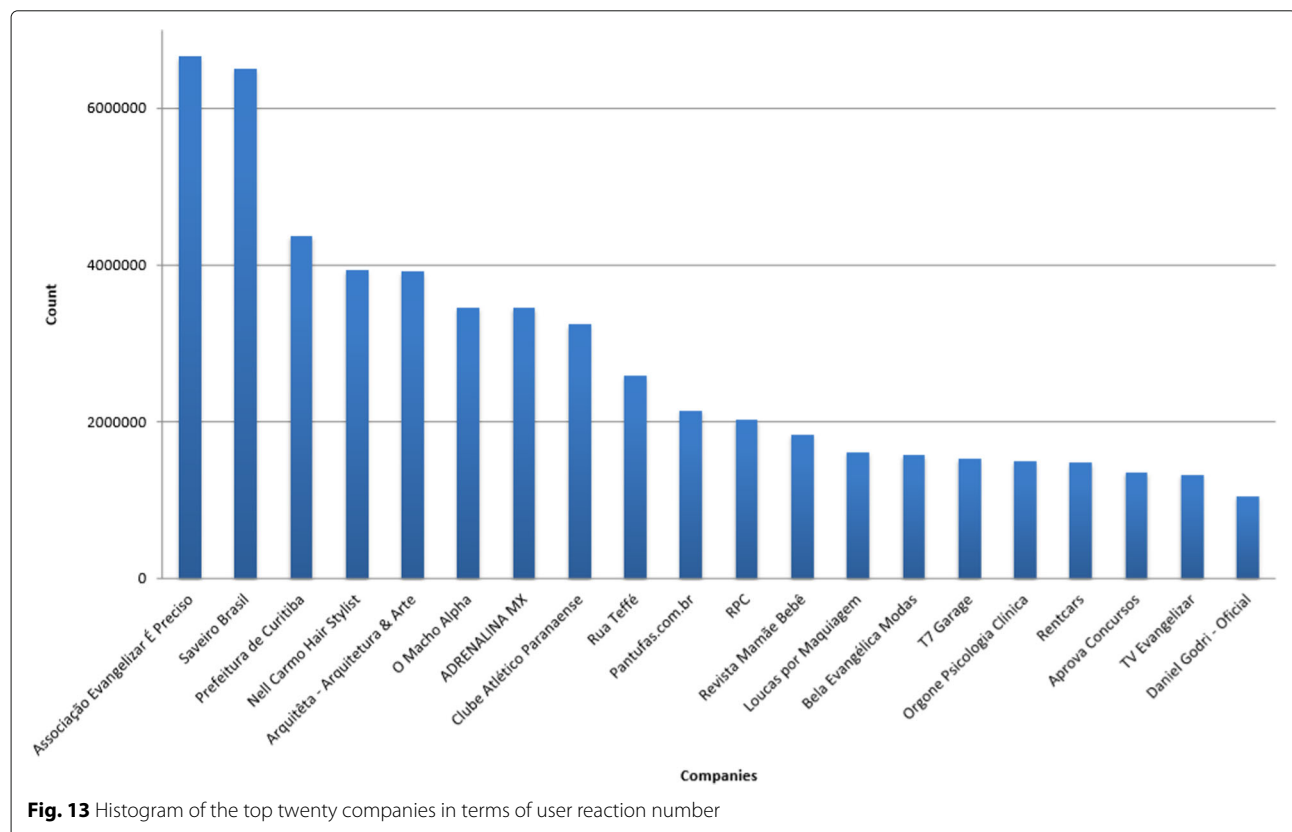


Fig. 12 Framework output for Rubiane. **a** Egonet for the company called Rubiane. **b** Community of businesses related to Food (including Rubiane). In red, a tourism related business was tagged as outlier by Algorithm 3

Appendix A: Supplementary information on business reaction database and business relationship graph



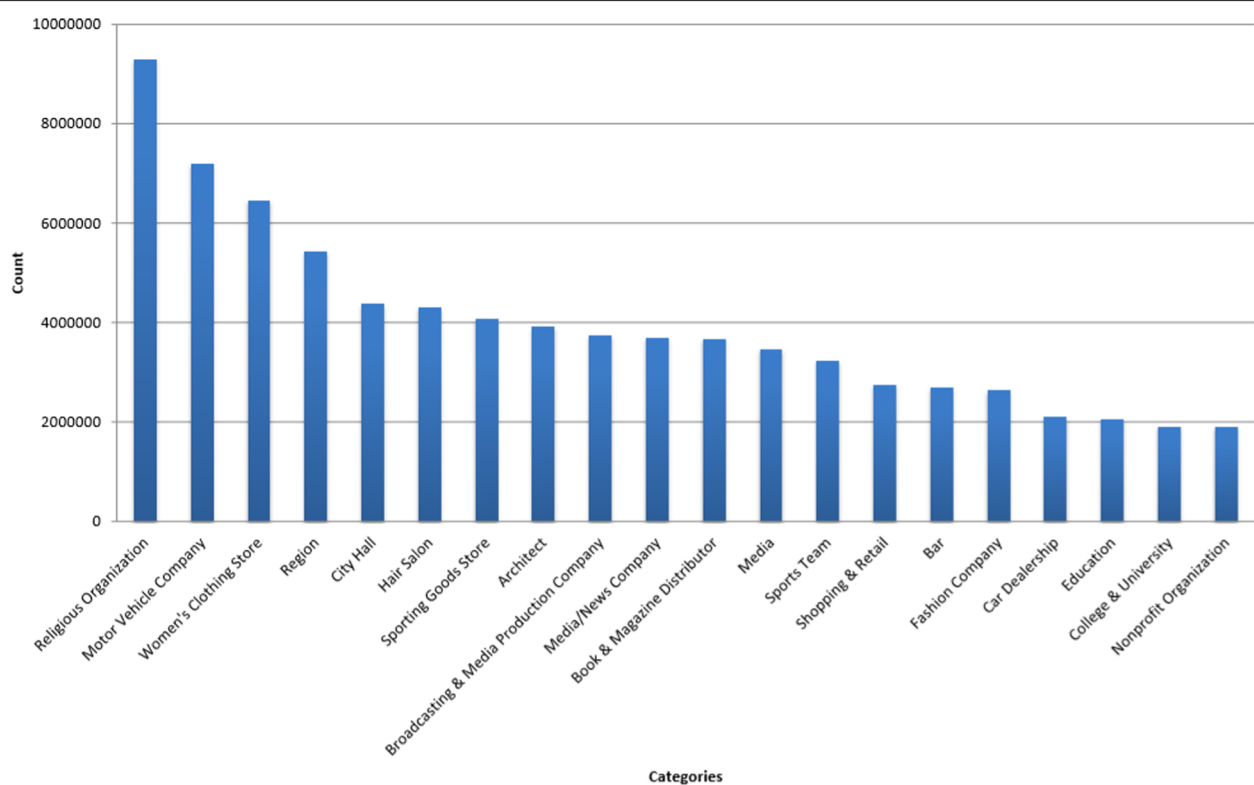


Fig. 14 Histogram of the top twenty categories in terms of user reaction number. The names are the original ones and do not reflect the renaming process shown in Fig. 6

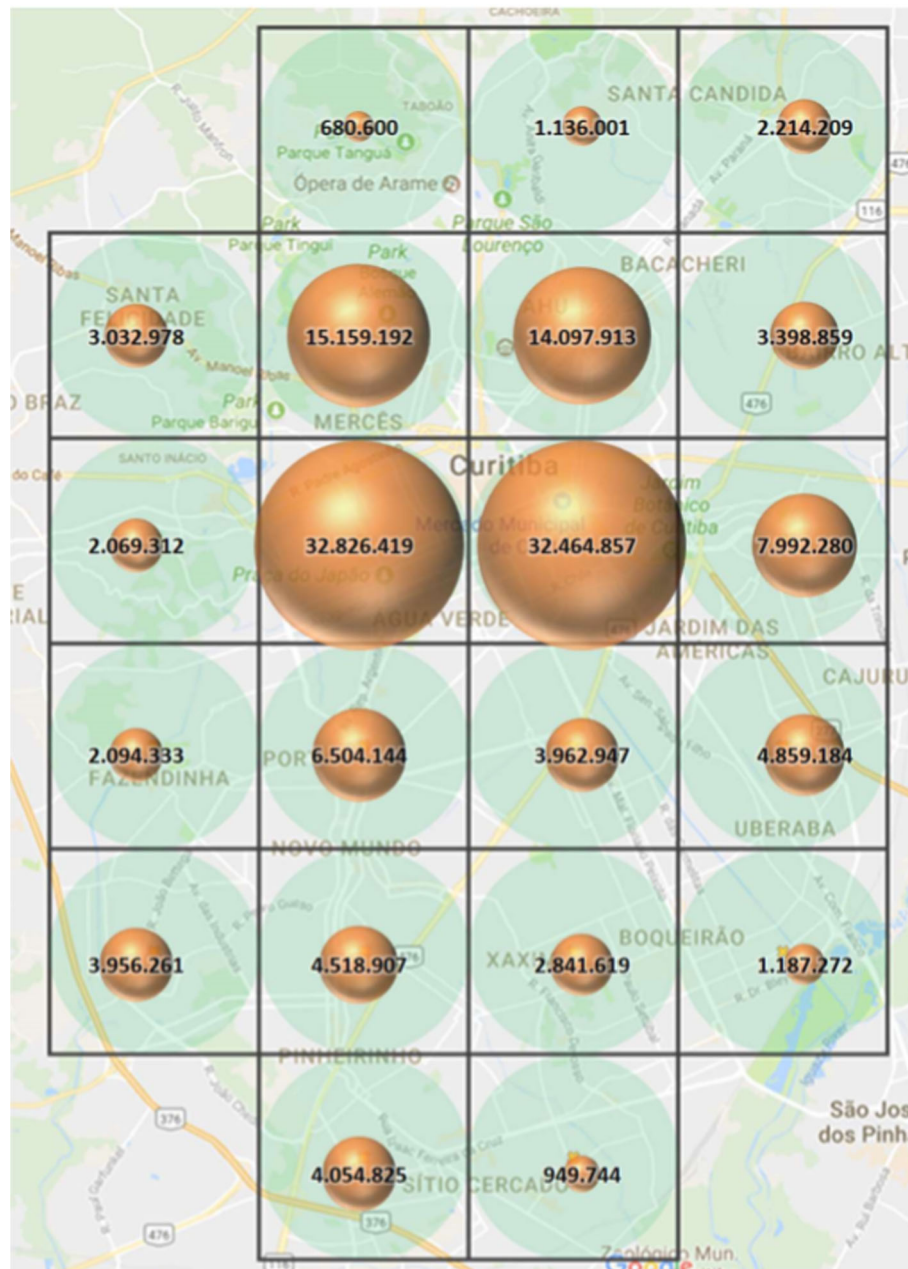


Fig. 15 A map showing the amount of user reactions per region considered in the collection process presented in Section 3

Table 2 Business graph nodes ranked by number of connections

Ranking	Business name	Number of connected edges
1	Prefeitura de Curitiba	1396
2	RPC	1357
3	Portal Banda B	1329
4	Clube Atlético Paranaense	1273
5	Aliança Móveis	1264
6	Shopping Mueller	1236
7	Dr. Freeze	1200
8	CWB Brasil	1200
9	Curitiba Comedy Club	1199
10	Shopping Palladium - Curitiba	1196
11	ParkShopping Barigüi	1196
12	Disk Ingressos	1164
13	Ripa na Chulipa	1148
14	Taco el Pancho	1148
15	Integra - Cursos & Treinamentos Curitiba	1138
16	Saia Justa Vestidos	1132
17	Cosmeticos Curitiba Centro	1100
18	Leve Sabor	1100
19	Zapata Bar	1099
20	Universidade Tuiuti do Paraná - UTP	1093
21	Espaço Gourmet Escola de Gastronomia	1086
22	Pele Morena Lingerie	1086
23	Invisible Braces Ortodontia	1083
24	Mais 55	1077
25	Daju	1077
26	La Passion Palladium	1074
27	Vida Leve Refeições Saudáveis	1071
28	Vivá	1063
29	Blood Rock Bar	1061
30	Rádio 98FM Curitiba	1052
31	Bar Quermesse	1047
32	Seven Entretenimento	1047
33	Power Airsoft	1045
34	Bar e Restaurante Hora Extra	1035
35	Hallorino Jr	1032
36	PolloShop	1025
37	Nell Carmo Hair Stylist	1025
38	BandNews FM Curitiba	1020
39	Cantinho da Bica	1010
40	Rádio Mundo Livre FM	1000
41	Império da Pizza	995
42	DOCK Premium	991
43	Autoescola Bello	991
44	A Barateira Feirão de Calçados	989
45	Hospital Pequeno Príncipe	988
46	Barone Consignações	985
47	Tatica Moveis	984
48	Shopping Total	981
49	Casa da Bruxa	973
50	Plus Santé Emergências Médicas Ltda	973

Table 3 Business graph edges ranked by weight

Ranking	Edge node	Edge node	Edge weight
1	RPC	Prefeitura de Curitiba	127958
2	Portal Banda B	Prefeitura de Curitiba	82904
3	Loja Us Store	Vissothi	75629
4	Portal Banda B	RPC	74084
5	Clube Atlético Paranaense	Prefeitura de Curitiba	73948
6	Associação Evangelizar É Preciso	TV Evangelizar	72184
7	Loja Us Store	Pantufas.com.br	57844
8	Clube Atlético Paranaense	RPC	56555
9	Shopping Mueller	Prefeitura de Curitiba	54823
10	Clube Atlético Paranaense	Portal Banda B	54437
11	ParkShopping Barigüi	Prefeitura de Curitiba	51522
12	Curitiba Cult	Prefeitura de Curitiba	50463
13	Shopping Palladium - Curitiba	Prefeitura de Curitiba	49198
14	BandNews FM Curitiba	Prefeitura de Curitiba	46635
15	Universidade Tuiuti do Paraná - UTP	Prefeitura de Curitiba	45870
16	Vissothi	Pantufas.com.br	45213
17	Arquitêta - Arquitetura & Arte	Prefeitura de Curitiba	45194
18	Blood Rock Bar	Prefeitura de Curitiba	44101
19	Rádio Mundo Livre FM	Prefeitura de Curitiba	44033
20	ParkShopping Barigüi	Shopping Mueller	43942
21	Saveiro Brasil	ADRENALINA MX	42377
22	Dr. Freeze	Prefeitura de Curitiba	41481
23	Ripa na Chulipa	Aliança Móveis	41188
24	CWB Brasil	Prefeitura de Curitiba	41041
25	Portal Banda B	Aliança Móveis	40995
26	Bella Morena Boutique	Rua Teffé	40642
27	Shopping Mueller	Aliança Móveis	39563
28	Pele Morena Lingerie	Aliança Móveis	38903
29	RPC	Aliança Móveis	38812
30	Prefeitura de Curitiba	Curitiba Comedy Club	38638
31	Shopping Mueller	RPC	38469
32	Prefeitura de Curitiba	Aliança Móveis	38298
33	Shopping Palladium - Curitiba	RPC	38193
34	Pantufas.com.br	Saveiro Brasil	38159
35	Hallorino Jr	RPC	37498
36	ParkShopping Barigüi	RPC	37350
37	Hospital Pequeno Príncipe	Prefeitura de Curitiba	36904
38	Associação Evangelizar É Preciso	Exper. de Deus com padre Reg. Manzotti	36505

Table 3 Business graph edges ranked by weight (*Continued*)

Ranking	Edge node	Edge node	Edge weight
39	Disk Ingressos	Prefeitura de Curitiba	36248
40	Integra - Cursos & Treinamentos Curitiba	Aliança Móveis	35842
41	Espaço Gourmet Escola de Gastronomia	Prefeitura de Curitiba	35641
42	CWB Brasil	RPC	35437
43	RPC	BandNews FM Curitiba	35045
44	Saia Justa Vestidos	Aliança Móveis	34837
45	Portal Banda B	CWB Brasil	34823
46	Portal Banda B	Shopping Palladium - Curitiba	34641
47	TV Evangelizar	Exper. de Deus com padre Reg. Manzotti	34596
48	Invisible Braces Ortodontia	Aliança Móveis	33152
49	Aliança Móveis	Barone Consignações	33150
50	Portal Banda B	Shopping Mueller	32969

Acknowledgements

The authors would like to thank Lucca Rawlyk, Fernanda Gubert, and Erik Almeida for their valuable help in this work. This study was partially supported by the project CNPq-URBCOMP (process 403260/2016-7), CAPES, CNPq, and Fundacao Araucaria.

Author's contributions

D.P.T. and T.H.S. designed the model and the computational framework and analyzed the data. D.P.T. carried out the implementation. D.P.T., A.T.F., and T.H.S. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Informatics, Federal University of Technology - Parana, Sete de Setembro Avenue, 3165, Curitiba, Brazil. ²Graduate Program in Industrial Engineering and Systems, PUC - Parana, Imaculada Conceição St., 1155, Curitiba, Brazil.

Received: 1 October 2018 Accepted: 9 April 2019

Published online: 18 June 2019

References

- Bergquist WH, Betwee J, Meuel D. Building Strategic Relationships: How to Extend Your Organization's Reach Through Partnerships, Alliances, and Joint Ventures. San Francisco: Jossey-Bass Publishers; 1995.
- Elmuti D, Kathawala Y. An overview of strategic alliances. *Management decision*. 2001;39(3):205–18.
- Hoffman DL, Fodor M. Can you measure the roi of your social media marketing?. *MIT Sloan Manag Rev*. 2010;52(1):41.
- Tuten TL, Solomon MR. *Social Media Marketing*. Thousand Oaks: Sage; 2017.
- Ferrari VC. Content marketing and brand engagement on social media: a study of facebook's posts in the ecommerce industry in brazil. Master dissertation (International Management), FGV - Fundação Getúlio Vargas, Escola de Administração de Empresas. São Paulo; 2016.
- Chaffey D. Global social media research summary 2016. *Smart Insights: Soc Media Mark*; 2016. <http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>. Accessed 20 July 2017.
- Culotta A, Cutler J. Mining brand perceptions from twitter social networks. *Mark Sci*. 2016;35(3):343–62.
- Trainor KJ, Andzulis JM, Rapp A, Agnihotri R. Social media technology usage and customer relationship performance: A capabilities-based examination of social crm. *J Bus Res*. 2014;67(6):1201–8.
- Agnihotri R, Dingus R, Hu MY, Krush MT. Social media: Influencing customer satisfaction in b2b sales. *Ind Mark Manag*. 2016;53:172–80.
- Hudson S, Thal K. The impact of social media on the consumer decision process: Implications for tourism marketing. *J Travel Tour Mark*. 2013;30(1-2):156–60.
- Cranshaw J, Schwartz R, Hong J, Sadeh N. The livehoods project: Utilizing social media to understand the dynamics of a city. In: *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12)*. Dublin: AAAI Press; 2012.
- Silva TH, de Melo POV, Almeida JM, Musolesi M, Loureiro AA. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In: *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*. Ann Arbor, MI. Menlo Park: AAAI Press; 2014.
- Mueller W, Silva TH, Almeida JM, Loureiro AA. Gender matters! analyzing global cultural gender preferences for venues using social sensing. *EPJ Data Sci*. 2017;6(1):5. <https://doi.org/10.1140/epjds/s13688-017-0101-0>.
- Brito S, Baldykowski A, Miczevski S, Silva TH. Cheers to untapped! preferences for beer reflect cultural differences around the world. In: *Proceedings of Americas Conference on Information Systems*, New Orleans (AMCIS18). New Orleans: Americas Conference on Information System; 2018.
- Fortunato S. Community detection in graphs. *Phys Rep*. 2010;486(3-5):75–174.
- Rossetti G, Cazabet R. Community discovery in dynamic networks: a survey. *ACM Comput Surv*. 2018;51(2):35.
- Mukhopadhyay S. Opinion mining in management research: the state of the art and the way forward. *OPSEARCH*. 2018;55(2):221–50.
- Silva T, Viana A, Benevenuto F, Villas L, Salles J, Loureiro A, Queercia D. Urban computing leveraging location-based social network data: a survey. *ACM Comput Surv*. 2019;52(1):17–11739.
- Ferreira APG, Silva TH, Loureiro AAF. Beyond sights: Large scale study of tourists' behavior using foursquare data. In: *Proc. of IEEE ICDMW'15 Workshops*. Atlantic City: IEEE; 2015. p. 1117–24. <https://doi.org/10.1109/ICDMW.2015.234>.
- Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Lucas RE, Agrawal M, Park GJ, Lakshmikanth SK, Jha S, Seligman ME, et al. Characterizing geographic variation in well-being using tweets. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*. Boston: AAAI Press; 2013. p. 583–91. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6138/6398>.
- De Choudhury M, Sharma S, Kiciman E. Characterizing dietary choices, nutrition, and language in food deserts via social media. In: *Proc. of CSCW'16*. San Francisco: ACM; 2016. p. 1157–70.
- Aiello LM, Schifanella R, Quercia D, Aletta F. Chatty maps: constructing sound maps of urban areas from social media data. *Open Sci*. 2016;3(3):150690.
- Santos FA, Silva TH, Loureiro AAF, Villas LA. Uncovering the Perception of Urban Outdoor Areas Expressed in Social Media. In: *Proc. of IEEE ACM Web Intelligence (WI)*. Santiago: IEEE; 2018. <https://doi.org/10.1109/WI.2018.00-99>.
- Magno G, Weber I. *International Gender Differences and Gaps in Online Social Networks*. Barcelona: Springer; 2014, pp. 121–38.
- Cheng Z, Caverlee J, Lee K, Sui DZ. Exploring millions of footprints in location sharing services. In: *Proc. of ICWSM'11*. Barcelona: AAAI Press; 2011. p. 81–88.
- Ai C, Chen B, He L, Lai K, Qiu X. The national geographic characteristics of online public opinion propagation in china based on wechat network. *Geoinformatica*. 2018;22:311–34. <https://doi.org/10.1007/s10707-017-0311-4>.
- Barbier G, Tang L, Liu H. Understanding online groups through social media. *Wiley Interdiscip Rev Data Min Knowl Disc*. 2011;1(4):330–8.
- Yang J, Zhang M, Shen KN, Ju X, Guo X. Structural correlation between communities and core-periphery structures in social networks: Evidence from twitter data. *Expert Syst Appl*. 2018;111:91–99. *Big Data Analytics for Business Intelligence*.
- Wu Z, Gao G, Bu Z, Cao J. Simple: a simplifying-ensembling framework for parallel community detection from large networks. *Clust Comput*. 2016;19(1):211–21.
- Liu W, Yue K, Wu H, Fu X, Zhang Z, Huang W. Markov-network based latent link analysis for community detection in social behavioral interactions. *Appl Intell*. 2017;48(8):2081–96. <https://doi.org/10.1007/s10489-017-1040-y>.
- Indrawati, Alamsyah A, et al. Social network data analytics for market segmentation in indonesian telecommunications industry. In: *Proc. of IColCT'17*. IEEE; 2017. p. 1–5. <https://doi.org/10.1109/IColCT.2017.8074677>.
- Tang L, Liu H. Understanding group structures and properties in social media. In: *Link Mining: Models, Algorithms, and Applications*. New York: Springer; 2010. p. 163–85.
- Grizane T, Jurgelane I. Social media impact on business evaluation. *Procedia Comput Sci*. 2017;104:190–6.
- Mahony T, Myers T, Low D, Eagle L. If we post it they will come: A small business perspective of social media marketing. In: *Proc. of ACSW'18*. Brisbane: ACM; 2018. p. 21.
- Kafeza E, Makris C, Rompolas G. Exploiting time series analysis in twitter to measure a campaign process performance. In: *Proc. of SCC'17*. IEEE; 2017. p. 68–75. <https://doi.org/10.1109/SCC.2017.17>.
- Giatsoglou M, Chatzakou D, Vakali A. User communities evolution in microblogs: A public awareness barometer for real world events. *World Wide Web*. 2015;18(5):1269–99.
- Pepin L, Kuntz P, Blanchard J, Guillet F, Suignard P. Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. *Comput Ind Eng*. 2017;112:450–8.
- Palsetia D, Patwary MMA, Agrawal A, Choudhary A. Excavating social circles via user interests. *Soc Netw Anal Mining*. 2014;4(1):170.

39. Lin J, Oentaryo R, Lim E-P, Vu C, Vu A, Kwee A. Where is the goldmine?: Finding promising business locations through facebook data analytics. In: Proc. of Hypertext'16. Halifax: ACM Digital Library; 2016. p. 93–102.
40. Karamshuk D, Noulas A, Scellato S, Nicosia V, Mascolo C. Geo-spotting: Mining online location-based services for optimal retail store placement. In: Proc. of ACM KDD'13. Chicago; 2013. p. 793–801. <https://doi.org/10.1145/2487575.2487616>. <http://doi.acm.org/10.1145/2487575.2487616>.
41. Tsutsumi DP, Fenerich AT, Silva TH. Identificando a relação virtual entre empresas explorando reações de usuários no facebook. In: WORKSHOP DE COMPUTAÇÃO URBANA (COURB_SBC), 2. , 2018, 1/2018. Anais do II Workshop de Computação Urbana (COURB 2018). Porto Alegre: Sociedade Brasileira de Computação (SBC); 2018. ISSN 2595-2706 <https://portaldeconteudo.sbc.org.br/index.php/courb/article/view/2336>.
42. Tasse D, Liu Z, Sciuto A, Hong JI. State of the geotags: Motivations and recent changes. In: Proc. of ICWSM'17. Montreal: AAAI Press; 2017. p. 250–9.
43. Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E. 2007;76(3): 036106.
44. Hartigan JA. Clustering algorithms. New York: Wiley; 1975.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)