Towards Business Partnership Recommendation Using User Opinion on Facebook

Diego P Tsutsumi¹*, Amanda T Fenerich² and Thiago H Silva¹

¹Correspondence: diego.tsutsumi@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

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, 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 Collection\textsuperscript{[1]}, and a similar trend is observed in other scientific databases.

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

For instance, Cheng et al.\textsuperscript{[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.\textsuperscript{[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.\textsuperscript{[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.\textsuperscript{[26]}. Barbier et al.\textsuperscript{[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.\textsuperscript{[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\textsuperscript{[15]} formally defines communities and presents a comparison of some algorithms for detecting communities. Rossetti and Gazabet\textsuperscript{[16]} also give a background on community discovery; however, in contrast to\textsuperscript{[15]}, they state community structures in dynamic networks.

Wu et al.\textsuperscript{[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.\textsuperscript{[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.\textsuperscript{[31]}, and some challenges of community detection in social media were presented by Tang and Liu\textsuperscript{[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.

\textsuperscript{[1]}These values represent average growth using the term “Social Media”, and “Social Media” and “Business” in conjunction.
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.

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 | Like |
| Number of Check-ins | 39627 | Number of Fans | 15532 |
| Average Evaluation | 4.6 | |

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. 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 API[2].

3.2 Data Collection

Figure 1 illustrates the main steps performed in the collection process. The data were collected using the Facebook Graph API[3], all business data collected are

[2]https://developers.facebook.com.

[3]https://developers.facebook.com/docs/graph-api/overview.
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 Figure 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.

We collected a total of 1,986 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, Figure 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 1,926 pages, all representing businesses, and approximately 260 million reactions.

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 Figure 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 $B = \{b_1, b_2, ..., b_{n_b}\}$ 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):

$$\text{BusinessGraph} = (V, E, W),$$  \hspace{1cm} (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:

$$W(i, j) = \begin{cases} 
\frac{|U_i \cap U_j|}{|U_i \cup U_j|} & \text{if } (i, j) \in E \\
0 & \text{if } (i, j) \notin E 
\end{cases}$$  \hspace{1cm} (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 equation 3.

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

As depicted in Figure 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 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 Equations (4) and (5).

\[
\mu = E[X] = \frac{n_r}{n_c} \tag{4}
\]

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

Where, \(n_r = \sum_{i,j \neq j} |U_i \cap U_j|\) 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: \(|U_i \cap U_j| \leq \text{lowerBound}\), as defined in Equation (6), following the \(3\sigma\) statistics for the binomial distribution, which includes 99.73% of random edges.

\[
\text{lowerBound} = \mu + 3\sigma \tag{6}
\]

For the data collected in this study the calculated values were: \(\mu = 120.289\), \(\sigma = 10.96\). Thus, \(\text{lowerBound} = 153.195\). 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 |

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

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 Figure 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 Figure 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 \( \text{vector}(c) \) 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 \textit{allCommunities} as in Algorithm 1) and returns a set of clusters, where clusters are disjoint non-empty subsets of \textit{allCommunities}:

\[ \text{clusters} = \text{kmeans}(\text{allCommunities}), \]

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

### 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 \textit{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 (\( \in \mathbb{R} \)) among all the communities from that cluster, as in Equation 7:

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

Cluster centroids are used to build \textit{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 \textit{Food and Beverage}, \textit{Shopping and Retail} and \textit{Entertainment}, because they meet a 70% threshold defined in a particular application.

More formally, Algorithm 2 specifies how to capture a \textit{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 \textit{threshold} \( \in \ [0.5, 1] \), a number representing the minimum percentage considered as majority. 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 \( \text{threshold} \times 100\% \) of all dimensions. Its time complexity is \( O(|v|^2) \).

Finally, 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 Figure 7), and each cluster has a centroid (as in Equation 7), both are captured in Algorithm 3 as \textit{vector(community)} and \textit{centroid(cl)}, respectively. In order to capture the most representative categories of each cluster, the function \textit{getSignature} (Algorithm 2)
Algorithm 2: Function that returns the signature (greatest dimensions) of the vector.

Function `getSignature(v, threshold)`:

1. `s ← ∑v_i`
2. `accThrs ← 0`
3. `signature ← ∅`
4. for `i ∈ {1, 2, ..., |v|}` do
5.   `m ← max(v)` /*Max value in vector*/
6.   `j ← argmax(v)` /* Category (index) of the vector’s maximum value*/
7.   if `|m + accThrs - threshold| < |accThrs - threshold|` then
8.     `signature ← signature ∪ j`
9.     `accThrs ← accThrs + m`
10. else
11.    break
12. end
13. end
14. return `signature`

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.

Algorithm 3: Outlier Detection Algorithm.

Data: Clusters - set containing business community clusters
Result: Clusters with tagged businesses

1. `taggedClusters ← ∅`
2. for `cl ∈ clusters` do
3.   `newCluster ← ∅`
4.   `clSignature ← getSignature(centroid(cl), 0.7)`
5.   for `community ∈ cl` do
6.     `vc ← vector(community)`
7.     for `i ∈ {1, 2, ..., |vc|}` do
8.       if `vc_i > 0 and i ∉ clSignature` then
9.         /*tagBusiness(community, cat) tags businesses of category cat inside community*/
10.        `newCluster ← newCluster ∪ {tagBusinesses(community, i)}`
11.     end
12.   end
13. `taggedClusters ← taggedClusters ∪ {newCluster}`
14. end
15. return `taggedClusters`

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 $allCommunities$ 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 Figure 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 Appendix A by Tables 2 and 3). N1 represents the city hall of Curitiba, and it is a very popular Facebook page [4] not only in the city of Curitiba but also nationally. N2 is the largest TV channel in Curitiba. Both strongly impact the analysis.

[4] https://www.facebook.com/PrefsCuritiba.
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 9(a) illustrates a community containing entertainment businesses (e.g., “Blood Rock Bar”, and “SSCW - 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 Figures 9(b) and 9(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 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 Figure 10. Besides, Figure 11 shows the number of user reactions in each cluster.

In those word clouds, we did not perform the subcategory 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 Figure 10 and Figure 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 Figure 9(c), 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 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
(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

Figure 9 Communities detected by Algorithm 1

partnerships. Although outliers are not part of the dominant context, they still have strong connections to businesses from that context.

Figure 12(a) shows the *egonet* of Rubiane, a seafood restaurant, which was arbitrarily chosen for analysis, and Figure 12(b) 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 Figure 12(a), but it appears in a community where Rubiane is also included, shown in Figure 12(b). Also in Figure 12(b) 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.

**Competing interests**
The authors declare that they have no competing interests.

**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.

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

1. Bergquist, W.H., Betwee, J., Meuel, D.: Building Strategic Relationships: How to Extend Your Organization's Reach Through Partnerships, Alliances, and Joint Ventures. Jossey-Bass Publishers, San Francisco, USA (1995)

2. Elmuti, D., Kathawala, Y.: An overview of strategic alliances. Management decision 39(3), 205-218 (2001)

3. Hoffman, D.L., Fodor, M.: Can you measure the roi of your social media marketing? MIT Sloan Management Review 52(1), 41 (2010)

4. Tuten, T.L., Solomon, M.R.: Social Media Marketing. Sage, Thousand Oaks, CA, USA (2017)

5. Ferrari, V.C.: Content marketing and brand engagement on social media: a study of facebook’s posts in the ecommerce industry in brazil. PhD thesis, FGV - Fundação Getúlio Vargas (2016)

6. Chaffey, D.: Global social media research summary 2016. Smart Insights: Social Media Marketing (2016)

7. Culotta, A., Cutler, J.: Mining brand perceptions from twitter social networks. Marketing science 35(3), 343-362 (2016)

8. Trainor, K.J., Andzulis, J.M., Rapp, A., Agnihotri, R.: Social media technology usage and customer relationship performance: A capabilities-based examination of social crm. Journal of Business Research 67(6), 1201-1208 (2014)

9. Agnihotri, R., Dingus, R., Hu, M.Y., Krush, M.T.: Social media: Influencing customer satisfaction in b2b sales. Industrial Marketing Management 53, 172-180 (2016)

10. Hudson, J., Thal, K.: The impact of social media on the consumer decision process: Implications for tourism marketing. Journal of Travel & Tourism Marketing 30(1-2), 156-160 (2013)

11. Cranshaw, J., Schwartz, R., Hong, J.J., Sadeh, N.: The livehoods project: Utilizing social media to understand the dynamics of a city. In: Proc. of ICWSM'12, Dublin, Ireland (2012)

12. Silva, T.H., de Melo, P.O.V., Almeida, J.M., Musolesi, M., Loureiro, A.A.: You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In: Proc. of ICWSM'14, Ann Arbor, USA (2014)

13. Mueller, W., Silva, T.H., Almeida, J.M., Loureiro, A.A.: Gender matters! analyzing global cultural gender preferences for venues using social sensing. EPJ Data Science 6(1), 5 (2017). doi:10.1140/epjds/s13688-017-0101-0

14. Brito, S., Baldykowski, A., Miczewska, S., Silva, T.H.: Cheers to untappd! preferences for beer reflect cultural differences around the world. In: Proc. of AMCIS'18, New Orleans, USA (2018)

15. Fortunato, S.: Community detection in graphs. Physics reports 469(3-5), 75–174 (2010)

16. Rossetti, G., Cazabet, R.: Community discovery in dynamic networks: a survey. ACM Computing Surveys 51(2), 35 (2018)

17. Mukhopadhyay, S.: Opinion mining in management research: the state of the art and the way forward. OPSEARCH 55(2), 221–250 (2018)

18. 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 Computing Surveys 52(1), 17–11739 (2019)

19. Ferreira, A.P.G., Silva, T.H., Loureiro, A.A.F.: Beyond sights: Large scale study of tourists’ behavior using foursquare data. In: Proc. of IEEE ICDMW’15 Workshops, pp. 1117–1124 (2015)

20. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Lucas, R.E., Agrawal, M., Park, G.J., Lakshminarathan, S.K., Jha, S., Seligman, M.E., et al.: Characterizing geographic variation in well-being using tweets. In: Proc. of ICWSM’13, Boston, USA (2013)

21. De Choudhury, M., Sharma, S., Kiciman, E.: Characterizing dietary choices, nutrition, and language in food deserts via social media. In: Proc. of CSCW’16, pp. 1157–1170. ACM, San Francisco, USA (2016)

22. Aiello, L.M., Schifanella, R., Queercia, D., Aletta, F.: Chatty maps: constructing sound maps of urban areas derived from social media data. Open Science 3(1), 150690 (2016)

23. Santos, F.A., Silva, T.H., Loureiro, A.A.F., Villas, L.A.: Uncovering the Perception of Urban Outdoor Areas Experienced in Social Media. In: Proc. of IEEE ACM Web Intelligence (WI), Santiago, Chile (2018)

24. Magno, G., Weber, I.: International Gender Differences and Gaps in Online Social Networks, pp. 121–138. Springer, Barcelona, Spain (2014)

25. Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z.: Exploring millions of footprints in location sharing services. In: Proc. of ICWSM’11, Barcelona, Spain, pp. 81–88 (2011)

26. 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, 1–24 (2018)

27. Barbier, G., Tang, L., Liu, H.: Understanding online groups through social media. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(4), 330–338 (2011)

28. Yang, J., Zhang, M., Shen, K.N., Ju, X., Guo, X.: Structural correlation between communities and core-periphery structures in social networks: Evidence from twitter data. Expert Systems with Applications 3(3), 205–218 (2010)

29. Wu, Z., Gao, G., Bu, Z., Cao, J.: Simple: a simplifying-ensembling framework for parallel community detection from large networks. Cluster Computing 19(1), 211–221 (2016)

30. 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. Applied Intelligence, 1–16 (2017)

31. Alamsyah, A., et al.: Social network data analytics for market segmentation in indonesian telecommunications industry. In: Proc. of ICICT’17, pp. 1–5 (2017). IEEE

32. Tang, L., Liu, H.: Understanding group structures and properties in social media. In: Link Mining: Models, Algorithms, and Applications, pp. 163–185. Springer, New York, NY, USA (2010)

33. Grizane, T., Jurgelzane, I.: Social media impact on business evaluation. Procedia Computer Science 104, 190–196 (2017)

34. 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, Australia, p. 21 (2018). ACM
35. Kafeza, E., Makris, C., Rompolas, G.: Exploiting time series analysis in twitter to measure a campaign process performance. In: Proc. of SCC’17, pp. 68–75 (2017). IEEE
36. Giatsoglou, M., Chatzakou, D., Vakali, A.: User communities evolution in microblogs: A public awareness barometer for real world events. World Wide Web 18(5), 1269–1299 (2015)
37. 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. Computers & Industrial Engineering 112, 450–458 (2017)
38. Palsetia, D., Patwary, M.M.A., Agrawal, A., Choudhary, A.: Excavating social circles via user interests. Social Network Analysis and Mining 4(1), 170 (2014)
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, Canada, pp. 93–102 (2016). ACM
40. Karamshuk, D., Noulas, A., Scellato, S., Nicosa, V., Mascolo, C.: Geo-spotting: Mining online location-based services for optimal retail store placement. In: Proc. of ACM KDD’13, Chicago, Illinois, USA, pp. 793–801 (2013). doi:10.1145/2487575.2487616. http://doi.acm.org/10.1145/2487575.2487616
41. Tsutsumi, D., Fenerich, A., Silva, T.H.: Identificando a relação virtual entre empresas explorando reações de usuários no facebook. In: Proc. of CoUrb’18, Campos do Jordão, Brazil (2018)
42. Tasse, D., Liu, Z., Sciuolo, A., Hong, J.I.: State of the geotags: Motivations and recent changes. In: Proc. of ICWSM’17, Montreal, Canada, pp. 250–259 (2017)
43. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical review E 76(3), 036106 (2007)
44. Hartigan, J.A.: Clustering algorithms (1975)
(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.

**Figure 12** Framework output for Rubiane.
A Supplementary information on business reaction database and business relationship graph

Figure 13 Histogram of the top twenty companies in terms of user reaction number.

Figure 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 figure 6.
Figure 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 Muller                       | 1236                      |
| 7       | Dr. Frazão                            | 1200                      |
| 8       | CWB Brasil                            | 1200                      |
| 9       | Curitiba Comedy Club                  | 1199                      |
| 10      | Shopping Palladium - Curitiba         | 1196                      |
| 11      | ParkShopping Bangui                   | 1196                      |
| 12      | Disk Impressões                       | 1184                      |
| 13      | Ripa na Chulipa                       | 1148                      |
| 14      | Taco et Paroch                         | 1148                      |
| 15      | Integra - Cursos & Treinamentos Curitiba | 1138                  |
| 16      | Salsa Justa Vestidos                  | 1132                      |
| 17      | Cosméticos Curitiba Centro            | 1100                      |
| 18      | Leve Sabor                            | 1099                      |
| 19      | Zapaia Bar                            | 1099                      |
| 20      | Universidade Tuitui do Paraná - UTP   | 1093                      |
| 21      | Espaço Gourmet Escola de Gastronomia  | 1086                      |
| 22      | Pale Moneta Lingente                  | 1086                      |
| 23      | Invisible Braces Odontologia          | 1083                      |
| 24      | Max 55                                | 1077                      |
| 25      | Daju                                  | 1077                      |
| 26      | La Passion Palladium                  | 1074                      |
| 27      | Vida Leve Refeições Saudáveis         | 1071                      |
| 28      | Viva                                  | 1063                      |
| 29      | Blood Rock Bar                        | 1061                      |
| 30      | Rádio Rádio FM Curitiba               | 1052                      |
| 31      | Bar Guerrinasse                       | 1047                      |
| 32      | Seven Entretenimento                  | 1047                      |
| 33      | Power Airport                         | 1045                      |
| 34      | Bar a Restaurante Hora Extra          | 1035                      |
| 35      | Halleiro Jr                           | 1032                      |
| 36      | PoliShop                              | 1025                      |
| 37      | Neil Carneiro Hair Stylist            | 1025                      |
| 38      | BandNews FM Curitiba                 | 1020                      |
| 39      | Cervejaria Curitiba                  | 1010                      |
| 40      | Rádio Mundo Vivo FM                   | 1000                      |
| 41      | Império da Prata                     | 995                       |
| 42      | DOCK Premium                          | 991                       |
| 43      | Autoescola Sello                      | 991                       |
| 44      | A Barateira Feira do Calçados        | 989                       |
| 45      | Hospital Poente Príncipe              | 988                       |
| 46      | Barão Cogumelos                      | 985                       |
| 47      | Tatiana Imóveis                       | 984                       |
| 48      | Shopping Total                        | 981                       |
| 49      | Casa da Bruxa                         | 973                       |
| 50      | Pizzaria Sant’Eugenio Médicas Ltd.    | 975                       |
### Table 3 Business graph edges ranked by weight.

| Ranking | Edge Node 1          | Edge Node 2          | Edge Weight |
|---------|----------------------|----------------------|-------------|
| 1       | RPC                  | Prefeitura de Curitiba | 127958      |
| 2       | Portal Banda B       | Prefeitura de Curitiba | 82904       |
| 3       | Loja Us Store        | Vascothic             | 75629       |
| 4       | Portal Banda B       | RPC                   | 74094       |
| 5       | Clube Atlético Paranense | Prefeitura de Curitiba | 73948      |
| 6       | Associação Evangélico É Precisio | TV Evangélico | 72184       |
| 7       | Loja Us Store        | Pantufas.com.br      | 57844       |
| 8       | Clube Atlético Paranense | RPC                | 56555       |
| 9       | Shopping Mueller     | Prefeitura de Curitiba | 54423      |
| 10      | Clube Atlético Paranense | Portal Banda B       | 54437       |
| 11      | ParkShopping Barigui | Prefeitura de Curitiba | 51352      |
| 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 | 46587      |
| 16      | Vascothic             | Pantufas.com.br      | 45213       |
| 17      | Arquiteta - Arquitetura & Arte | Prefeitura de Curitiba | 45194     |
| 18      | Nood Rock Bar        | Prefeitura de Curitiba | 44105      |
| 19      | Rádio Mundo Livre FM | Prefeitura de Curitiba | 44033      |
| 20      | ParkShopping Barigui | Shopping Mueller      | 43942       |
| 21      | Saveiro Brasil       | ADRENALINA MX        | 42377       |
| 22      | Dr. Finocchioro      | Prefeitura de Curitiba | 41481      |
| 23      | Ripa na Chupipa      | Aliança Móveis       | 41188       |
| 24      | CWB Brasil           | Prefeitura de Curitiba | 40461      |
| 25      | Portal Banda B       | Aliança Móveis       | 40095       |
| 26      | Bella Morena Boutique | Rua Toffy          | 40042       |
| 27      | Shopping Mueller     | Aliança Móveis       | 39563       |
| 28      | Pela Morena Lingerie | Aliança Móveis       | 39093       |
| 29      | RPC                  | Aliança Móveis       | 38312       |
| 30      | Prefeitura de Curitiba | Curitiba Comedy Club | 38383     |
| 31      | Shopping Mueller     | RFC                  | 38406       |
| 32      | Prefeitura de Curitiba | Aliança Móveis       | 38298       |
| 33      | Shopping Palladium - Curitiba | RPC | 38193       |
| 34      | Partituba.com.br     | Shopping Mueller     | 38159       |
| 35      | Halloniron Jr        | RPC                  | 37496       |
| 36      | ParkShopping Barigui | RPC                  | 37155       |
| 37      | Hospital Pequeno Príncipe | Prefeitura de Curitiba | 36904    |
| 38      | Associação Evangélico É Precisio | Exper. de Deus com padr. Reg. Marotti | 36605  |
| 39      | Disk Ingressos       | Oficina de Estudos   | 36248       |
| 40      | variedades - Cursos & Treinamentos Curitiba | Aliança Móveis | 35482   |
| 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           | 34263       |
| 46      | Portal Banda B       | Shopping Palladium - Curitiba | 34461   |
| 47      | TV Evangélico        | Exper. de Deus com padr. Reg. Marotti | 34956  |
| 48      | Invisible Braces Ortodontia | Aliança Móveis | 33152   |
| 49      | Aliança Móveis       | Barone Construções  | 33150       |
| 50      | Portal Banda B       | Shopping Mueller     | 29699       |