Business Relationship Network Model from Social Reactions Data

Diego Paolo Tsutsumi
Department of Informatics (DAINF)
Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba Paraná Brazil
diegotsu@utfpr.edu.br

Thiago Henrique Silva
Department of Informatics (DAINF)
Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba Paraná Brazil
thiagoh@utfpr.edu.br

ABSTRACT
One of the primary ways to expand a business or to keep it stable during a crisis is to create partnerships with other companies. With that, this study presents results regarding a new data model, which explores user reactions on social media to indicate strategic business partnerships. There are three main contributions of this study to the literature: (i) a business relationship network model; (ii) a business community detection algorithm; and (iii) a business outlier detection algorithm. The evaluation of the contributions was performed exploring real data of approximately 280 million user reactions on Facebook. Results suggest that business partnership recommendation is possible using the information available in social media.

KEYWORDS
Social Media, Community Detection, Business Partnership

1 Problem description and motivation
Businesses and companies are continuously exposed to uncertain and dynamic environments. For instance, markets go into crisis or change preferences, competitors strategies evolve, and newly elected politics may also affect business environments by changing regulations and making lobby. All those factors make it difficult for business owners to make strategic decisions. One of the ways to expand a business or to keep it stable during a crisis is to make strategic partnerships with other businesses.

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]. Thus, the task of choosing which partnerships are best for a particular business may be challenging and finding appropriate data for this issue is key.

Many known factors, such as customers' tastes and opinions, are reflections of their preferences for businesses [3, 4, 5]. Preferences implicitly manifested by users in actions in social media was also assumed to exist in previous works [6, 7, 8, 9]. From the social media perspective, since customers and companies are regularly feeding their social media profiles, users' opinion data are widely available in platforms such as Facebook and Twitter. For instance, user reactions on content shared by a particular business can represent users' preferences. A proper analysis of these data could be used along with traditional business analysis to improve the strategic decision quality of businesses.

Some studies analyzed social media data for business intelligence [10, 11, 12, 13], while other studies were carried out, without social media data, to understand business partnerships [1,2]; however, the exploration of social media data analysis in the context of business partnerships decision making, to the best of our knowledge, is a novelty in the literature.

In: II Concurso de Teses e Dissertações (CTD 2020), São Luís, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2020.

ISSN: 2596-1683

2 Objectives and Contributions
The general goal of this study is to propose and evaluate a model that helps business owners to choose strategic partnerships. There are three specific objectives to accomplish this goal:

1. To produce and evaluate a business relationship model by exploring network theory (Section 2.1);
2. To propose and explore a community detection algorithm specifically designed for business networks (Section 2.2);
3. To propose and explore a business outlier detection algorithm in the presented model (Section 2.3).

2.1 Business Relationship Model
The proposed business relationship model is based on two assumptions: (i) users who reacted to a particular business page are potential customers and (ii) the affinity between two businesses is proportional to the number of users who reacted to both businesses, the common users between them.

The model is a non-directed graph in which vertices represent businesses, and weighted edges represent relations between two businesses. This relation between two different businesses get stronger as their common users grow in proportion to the set of their own users. Thus, technically, edges are weighted using the Jaccard Index [14] of the set of users of each business, representing an index of affinity or similarity between the two sets, as following:

$$BusinessGraph = (V, E, W)$$

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

$$E = \{(i,j) : |U_i \cap U_j| > lowerBound\}$$

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

where B is the set of business pages and U_i is the set of users that reacted to the business i. Figure 1 illustrates the graph construction process, considering positive user reactions.

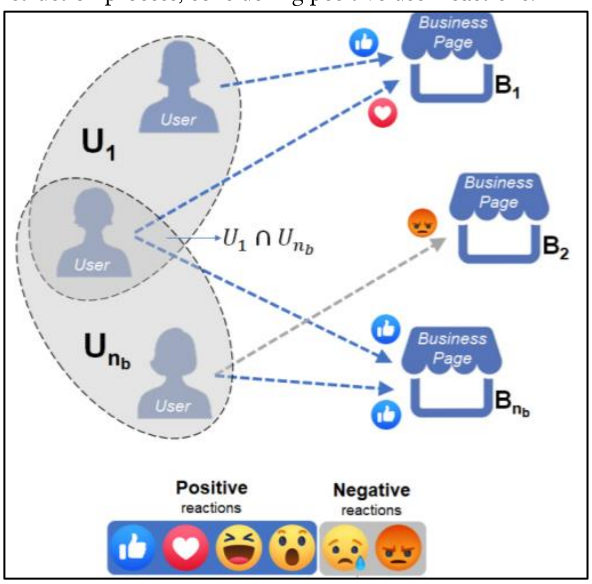


Figure 1: Users and Business pages reactions. Only positive reactions were considered in this study.

The higher the weight of an edge between two businesses the higher their degree of relationship. Weights have a proportionality in its formula, therefore balancing weights among businesses with different number of user reactions.

2.2 Business Community Detection Algorithm

Given a consistent network of business relationships, an essential step in achieving the study's goal is to detect business communities. A community detection algorithm based on label propagation (LP) was proposed [15] with advantages comparing to algorithms based on clique or dense subgraphs searches with optimal solution. One significant advantage is that this algorithm operates in almost linear time, and another advantage is that this LP-based algorithm does not need previous information.

Based on this, an iterative algorithm (Algorithm 1) is proposed for the detection of businesses' communities. 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.

Algorithm 1: Business Communities Detection Algorithm.

Data: BusinessGraph,minSize,maxSize
Result: Set of Communities of BusinessGraph

$G \leftarrow BusinessGraph$
$allCom \leftarrow \emptyset$
$minEdge \leftarrow \min_{i,j \in G} W(i,j)$
$counter \leftarrow 1$
/* $\sim O(|B|+|E|)$ */
while $|G| > minSize$ **do**
  $counter \leftarrow counter + 1$
  /* Method of (RAGHAVAN et al., 2007) */
  $detectedComm \leftarrow labelPropCommDetection(G)$ /* $\sim O(|B|+|E|)$ */
  $G \leftarrow empty\ graph$
  for $c \in detectedComm$ **do**
    if $|c| > minSize$ *and* $|c| < maxSize$ **then**
      $allCom \leftarrow allCom \cup \{c\}$
    else
      $G \leftarrow G \cup c$
    end
  end
  remove all edges of G with $W(i,j) < minEdge * counter$
  if *no edges removed from G* **then**
    break
  end
end
return $allCom$;

2.3 Business Outlier Detection Algorithm

Given a set of communities, a business dissimilar to a reference set of categories inside its community is considered an outlier. A clustering process is done to identify reference sets of categories. Considering that the category of a business *bi* can be represented by *cat(bi)*, the **feature vector** is defined by the following:

$$\forall com_i, com_j \in allCom; \quad \forall cat_k \in allCategories$$

$$B_{com_i,cat_k} = \{b \in com_i : cat(b) = cat_k\}$$

$$vector(com_i) = \left(\frac{|B_{com_i,cat_1}|}{\max\limits_{com_j}|B_{com_j,cat_1}|}, \frac{|B_{com_i,cat_2}|}{\max\limits_{com_j}|B_{com_j,cat_2}|}, ..., \frac{|B_{com_i,cat_{28}}|}{\max\limits_{com_j}|B_{com_j,cat_{28}}|}\right)$$

where $allCom$ is the set of all communities detected in Section 2.2; $allCategories$ is the set of all business categories in the dataset; $Bcom_i, cat_k$ is the number of business inside the community i that belongs to category k. The final vector represents proportions of business categories inside each community.

An algorithm to extract the most representative business categories in each community is described in Algorithm 2.

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

Function getSignature(v, *threshold*):
  $s \leftarrow \sum_i v_i$ /* $O(d)$ Used to normalize values to be comparable with the threshold */
  $accThrs \leftarrow 0$
  $signature \leftarrow \emptyset$
  /* For loop runs in $O(d^2)$ */
  for $i \in \{1, 2, ..., d\}$ **do**
    $m \leftarrow max(v)$ /* $O(d)$ Max value in vector */
    $j \leftarrow argmax(v)$ /* $O(d)$ Category (index) of the vector's maximum value */
    if $|\frac{m+accThrs}{s} - threshold| < |\frac{accThrs}{s} - threshold|$ **then**
      $signature \leftarrow signature \cup j$
      $accThrs \leftarrow accThrs + m$
      $v_j \leftarrow 0$ /* In next iteration, max(v) is the next greatest */
    else
      break
    end
  end
return *signature*

The final outlier detection algorithm (Algorithm 3) compares communities' signatures with its respective cluster signature to tell whether a business inside a community is an outlier.

Algorithm 3: Outlier Detection Algorithm.

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

$taggedClusters \leftarrow \emptyset$
for $cl \in clusters$ **do**
  $newCluster \leftarrow \emptyset$
  $clSignature \leftarrow getSignature(centroid(cl), 0.7)$ /* For all Clusters it runs in $O(|clusters| * n_{cat}^2)$ */
  for $community \in cl$ **do**
    $vc \leftarrow vector(community)$ /*Defined in Equation 4.10, for all communities it runs in $O(|B| + |allComm| * n_{cat})$ */
    for $i \in \{1, 2, .., |vc|\}$ **do**
      if $vc_i > 0$ *and* $i \notin clSignature$ **then**
        /*$tagBusiness(community, cat)$ tags businesses of category cat inside $community$, for all businesses it runs in $O(|B|)$ */
        $newCluster \leftarrow newCluster \cup \{tagBusinesses(community, i)\}$
      end
    end
    $taggedClusters \leftarrow taggedClusters \cup \{newCluster\}$
  end
end
return $taggedClusters$

3 Results and Discussion

The proposed model was tested using Facebook data in the city of Curitiba – Brazil, from November to December of 2017 [16, 17]. A total of 1,986 georeferenced pages and approximately 280 million user reactions related to those pages were collected. The pages with more connected edges were "Prefeitura de Curitba", with 1396 connections, following by "RPC" with 1357 connections. Also, the edge most weighted was found between these two nodes ("Prefeitura de Curitiba" and "RPC"), with a value of 127958.

After some data cleaning/filtering procedures [17,18] and running Algorithm 1 with the parameters considered, 144 communities were detected, each ranging from 4 to 30 businesses all located in the city of Curitiba. An example of these communities is illustrated in Figure 2, which containing entertainment businesses (e.g., "Blood Rock Bar", and "SSCWB - Shinobi Spirit") and food businesses (e.g., "Ca'dore Comida Descomplicada"), so we can notice they are businesses inside the "leisure" context.

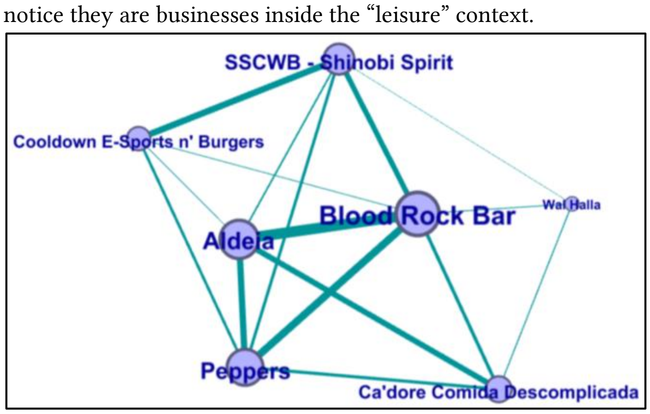
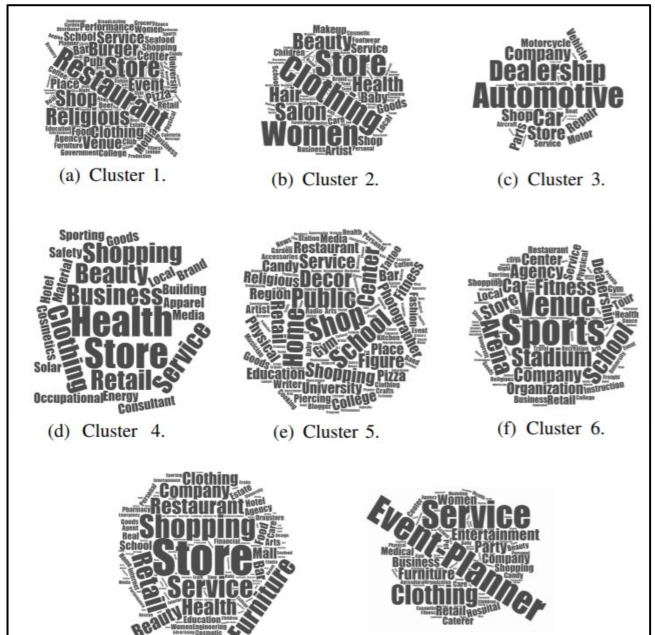


Figure 2: Community of businesses related to leisure

We can note that, even though both the business network construction and the Algorithm 1 did not use any information of the businesses themselves, all communities detected have similar strong semantics that bind 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 illustrated in Figure 3.



(a) Cluster 1. (b) Cluster 2. (c) Cluster 3.
(d) Cluster 4. (e) Cluster 5. (f) Cluster 6.
(g) Cluster 7. (h) Cluster 8.

Figure 3: Word clouds for categories of similar community clusters

Note the surprising similarity between the categories in each group. For example, Cluster 1 (Figure 3a) is related to leisure, containing predominantly food, drink and entertainment businesses, Cluster 2 (Figure 3b) contains most businesses related to beauty and style, while Cluster 3 (Figure 3c) is more related to establishments about automotive products and services. This analysis shows the existence of a predominant context in each community [18].

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.

The community illustrated in Figure 4, 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 Algorithms 2 and 3.

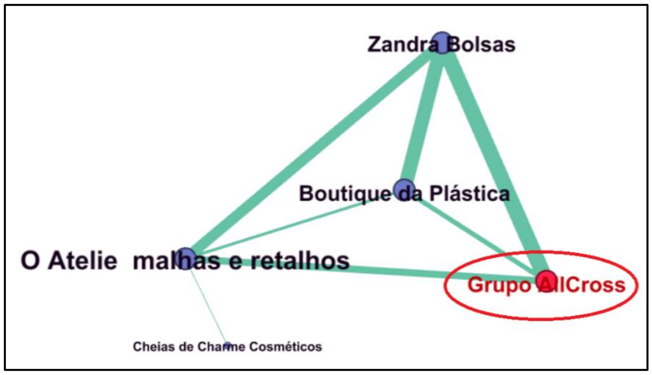
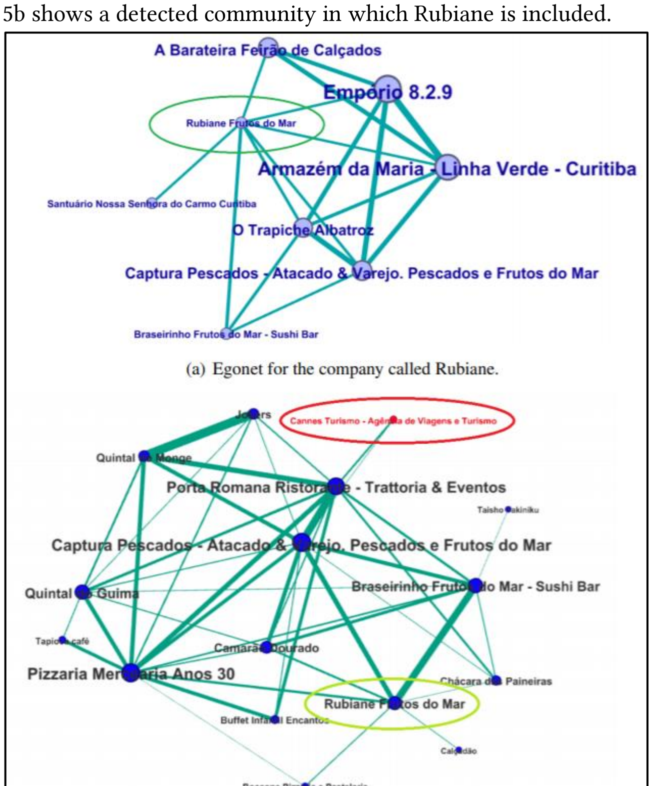


Figure 4: Community of businesses with an outlier

As an improvement of the results, outliers cannot be ignored in the results presented here, as they might represent non-trivial potential business partnerships [18]. Although outliers are not part of the dominant context, they still have strong connections to businesses from that context.

3.1 A real use case

In order to demonstrate a real business use case, an arbitrarily business was chosen for analysis, consisting in a seafood restaurant called Rubiane. Figure 5a shows the egonet (the subgraph of direct connections of a node) of Rubiane and Figure 5b shows a detected community in which Rubiane is included.



(a) Egonet for the company called Rubiane.

(b) Community of businesses related to Food (including Rubiane).

Figure 5: Framework output for Rubiane

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 named "Quintal do Monge" does not appear in the Rubiane's egonet shown in Figure 5a, but it appears in a community where Rubiane is also included, shown in Figure 5b. Also, in Figure 5b notice that the business called "Cannes Turismo" (highlighted in red) is a tourism related business and was tagged as an outlier by Algorithms 2 and 3.

ACKNOWLEDGMENTS

This work was partially supported by the project GoodWeb (process \#2018/23011-1 of FAPESP)

REFERENCES

[1] WH Bergquist, J. Betwee, D Meuel (1995). Building strategic relationships: How to extend your organization's reach through partnerships, alliances, and joint ventures. San Franscisco, USA: Jossey-Bass Publishers.
[2] D. Elmuti, Y. Kathawala (2001). An overview of strategic alliances. Management decision, MCB UP Ltd, v. 39, n. 3, p. 205–218.
[3] K. J. Trainor et al. (2014) Social media technology usage and customer relationship performance: A capabilities-based examination of social crm. Journal of Business Research, Elsevier, v. 67, n. 6, p. 1201–1208.
[4] R. Agnihotri, et al. (2016) Social media: Influencing customer satisfaction in b2b sales. Industrial Marketing Management, Elsevier, v. 53, p. 172–180.
[5] S. Hudson, K. Thal (2013). The impact of social media on the consumer decision process: Implications for tourism marketing. Journal of Travel & Tourism Marketing, Taylor & Francis, v. 30, n. 1-2, p. 156–160.
[6] J. Cranshaw, et al (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In: Proc. of ICWSM'12. Dublin, Ireland.
[7] T. H. Silva et al (2014). 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.
[8] W. Mueller, et al. (2017) Gender matters! analyzing global cultural gender preferences for venues using social sensing. EPJ Data Science, v. 6, n. 1, p. 5.
[9] S. Brito, et al. (2018) Cheers to untappd! preferences for beer reflect cultural differences around the world. In: Proc. of AMCIS'18. New Orleans, USA.
[10] J. Lin et al. (2016) Where is the goldmine?: Finding promising business locations through facebook data analytics. In: ACM. Proc. of Hypertext'16. Halifax, Canada, p. 93–102.
[11] D. Karamshuk, et al. (2013) Geo-spotting: Mining online location-based services for optimal retail store placement. In: Proc. of ACM KDD'13. Chicago, Illinois, USA. p. 793–801. ISBN 978-1-4503-2174-7.
[12] D. L. Hoffman, M. Fodor (2010). Can you measure the roi of your social media marketing? MIT Sloan Management Review, Massachusetts Institute of Technology, Cambridge, MA, v. 52, n. 1, p. 41.
[13] T. L. Tuten,; M. R. Solomon (2017). Social media marketing. Thousand Oaks, CA, USA: Sage.
[14] P.N. Tan, M. Steinbach, V. Kumar (2005). Introduction to data mining. 1st. ISBN 0-321-32136-7
[15] U. N. Raghavan, R. Albert, S. Kumara (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical review E, APS, v. 76, n. 3, p. 036106.
[16] D.P. Tsutsumi, A.T. Fenerich, T. H. Silva (2018). Identificando a relação virtual entre empresas explorando reações de usuários no facebook. In: Proc. of CoUrb'18. Campos do Jordão, Brazil.
[17] D.P. Tsutsumi, T. H. Silva (2018) Identifying virtual relations among businesses exploring user reaction on facebook. In: ACM. Proceedings of the International Conference on Web Intelligence.
[18] D.P. Tsutsumi, A.T. Fenerich, T.H. Silva (2019). Towards business partnership recommendation using user opinion on facebook. Journal of Internet Services and Applications, Springer, v. 10, n. 1, p. 11.