

Identificando a Relação Virtual Entre Empresas Explorando Reações de Usuários no Facebook

Diego P. Tsutsumi¹, Amanda T. Fenerich², Thiago H. Silva¹

¹Programa de Pós-Graduação em Computação Aplicada
Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba – PR – Brasil

²Programa de Pós-Graduação em Engenharia de Produção e Sistemas
Pontifícia Universidade Católica do Paraná (PUCPR)
Curitiba – PR – Brasil

diegotsu@gmail.com, amanda.fenerich@pucpr.edu.br, thiagoh@utfpr.edu.br

Resumo. *Vivemos hoje em um mundo altamente competitivo em várias áreas, inclusive no mundo dos negócios. Diante disso, parcerias empresariais estratégicas são importantes para o sucesso dos negócios, principalmente quando expostos a um mercado dinâmico e incerto. Com isso, este estudo tem como objetivo identificar relações virtuais expressivas entre empresas que são geradas espontaneamente por usuários em mídias sociais. Para isso, são propostos neste estudo um modelo de similaridade entre empresas baseado em reações comuns espontâneas de usuários em mídias sociais, particularmente o Facebook, bem como um novo algoritmo de detecção de comunidades de empresas neste modelo proposto. Para demonstrar a metodologia proposta, foram coletadas mais de 280 milhões de reações de usuários no Facebook sobre empresas de Curitiba, Brasil. Explorando esses dados e a metodologia proposta, o estudo mostra o potencial para identificar relações não óbvias entre empresas, o que pode habilitar o desenvolvimento de novos mecanismos de recomendação de parcerias empresariais estratégicas para empreendedores e donos de negócios.*

Abstract. *We live today in a highly competitive world in many areas, including the business world. Thus, strategic business partnerships are important for a business to succeed, especially for those exposed to a dynamic and uncertain market. Therefore, this study aims to identify expressive virtual relationships between companies that are generated spontaneously by users in social media. To this end, a business similarity model based on common spontaneous reactions of users in social media, particularly Facebook, is proposed, as well as a new algorithm for detecting business communities within the proposed model. To demonstrate the proposed methodology, the study uses more than 280 million user reactions on Facebook about companies in Curitiba, Brazil. Exploring these data and the proposed methodology, this study shows the possibility to identify non-trivial relations between companies, which may enable the development of new mechanisms for recommending strategic business partnerships for entrepreneurs and business owners.*

1. Introdução

Dada a alta competitividade do mundo dos negócios, atualmente muitas empresas são pressionadas a inovar e explorar novas estratégias para permanecerem competitivas no mercado. Uma ação fundamental a ser realizada nesse sentido é a identificação de parcerias empresariais estratégicas.

Atualmente as empresas exploram, cada vez mais, mídias sociais para diversos propósitos. Isso é interessante porque as mídias sociais são amplamente utilizadas pela nossa sociedade moderna, o que pode ser utilizado como um canal de comunicação, bem como valiosas fontes de informações sociais [Chen *et al.* 2016]. Uma das mídias sociais mais utilizadas pelas empresas para expor seus produtos, interagir com seus clientes, e também para manter uma relação com suas clientelas é o *Facebook*¹ [Ferrari 2016, Chaffey 2016]. Isso gera uma vasta quantidade de dados de relacionamento cliente-empresa que podem proporcionar *insights* valiosos se forem analisados apropriadamente.

Devido à alta complexidade, incerteza e dinamismo de grande parte do mundo dos negócios, torna-se desafiador para empresários e empreendedores possuírem um elevado grau de certeza em suas decisões estratégicas, principalmente quando o mercado está em crise. Deste modo, técnicas tradicionais como, análise de mercado, planejamento estratégico, segmentação de mercado e outras, podem não suprir completamente as necessidades de mercados dinâmicos e incertos. Assim, análises de dados de mídias sociais se mostram uma alternativa interessante, tanto para complementar análises tradicionais quanto para ter uma visão do mercado em menor tempo e esforço de análise [Culotta and Cutler 2016].

As preferências das pessoas por empresas ao efetuar uma compra podem ser afetadas por várias variáveis como cultura, rotinas, gostos, opiniões, eventos e situações do cotidiano [Trainor *et al.* 2014, Agnihotri *et al.* 2016, Hudson and Thal 2013]. Assumindo que as mídias sociais são representações do que acontece no mundo real, as preferências por empresas podem ser implicitamente manifestadas por usuários nestes canais.

Com isso, este estudo tem como objetivo identificar relações virtuais, geradas espontaneamente por usuários em mídias sociais, que se demonstram expressivas entre duas empresas. A modelagem proposta de similaridade entre empresas considera, os usuários que duas empresas possuem em comum, estabelecendo relações empresariais através da possível clientela em comum entre empresas, bem como a estrutura dessas relações empresariais em larga escala, particularmente dados públicos do *Facebook*, que é objeto de estudo deste trabalho. Além disso, é proposto também neste trabalho um algoritmo iterativo de detecção de comunidades de empresas.

Para demonstrar a utilidade da abordagem proposta, foram coletadas mais de 280 milhões de reações públicas de usuários no *Facebook* sobre empresas de Curitiba, PR, Brasil. O *Facebook*, entre tantas outras plataformas de mídias sociais, foi escolhido porque é uma mídia social amplamente utilizada tanto pelos clientes quanto pelas empresas [Ferrari 2016].

Utilizando os dados coletados e a metodologia proposta foram encontradas comunidades que possuem uma similaridade surpreendente entre as categorias de locais que

¹<https://www.facebook.com>

compõem essas comunidades. Isso ilustra o potencial para identificar relações não óbvias entre empresas, o que pode habilitar o desenvolvimento de novos serviços e aplicações. Por exemplo, um novo serviço de recomendação de parcerias empresariais estratégicas que pode ser útil para empresários e empreendedores. Esse serviço pode contribuir para aumentar o desempenho de vendas das empresas, bem como para que elas se mantenham mais sustentáveis no mercado.

O restante do trabalho está organizado da seguinte forma. A Seção 2 apresenta alguns dos principais trabalhos com objetivos similares a este estudo. A Seção 3 descreve as particularidades dos dados coletados, que são utilizados na modelagem proposta na Seção 4. A Seção 5 apresenta os resultados obtidos por este estudo. Por fim, a seção 6 conclui o trabalho e aponta possíveis trabalhos futuros.

2. Trabalhos Correlatos

Uma quantidade considerável de esforço foi realizada pela comunidade acadêmica, com o objetivo de extrair informações relevantes de mídias sociais. [Silva *et al.* 2014] usaram algumas técnicas para criar um modelo a partir de dados coletados do *Twitter* e *Foursquare*, a fim de identificar diferentes culturas observando o que as pessoas comem e bebem. [Barbier *et al.* 2011] criaram um método para entender o comportamento e a dinâmica do que chamam de *grupos online (online groups)*, categorizando e classificando entidades como, indivíduos, grupos de pessoas e empresas, em diferentes categorias e classes.

[Cheng *et al.* 2011] apresentaram análises espaço-temporais e de mobilidade, baseadas em uma base de dados coletada de uma mídia social para compartilhamento de localização (*Location Sharing Social Media*). Do ponto de vista de negócios, os resultados dessa análise podem oferecer sustentação em decisões sobre onde e quando investir recursos em um novo negócio. [Cranshaw *et al.* 2012] introduziram o conceito de *Livehoods*, que são regiões de uma cidade particionadas e agrupadas por similaridade de comportamentos dos moradores, este estudo de segmentação de comportamento pode também ser importante para decisões estratégias em empresas.

Mais próximas à proposta deste estudo, duas aplicações foram desenvolvidas por [Lin *et al.* 2016, Karamshuk *et al.* 2013] para ajudar empresários e analistas a encontrar o melhor local em uma cidade para abrir um novo negócio, e todo o processo de decisão é baseado em informações de mídias sociais. No caso da aplicação de [Lin *et al.* 2016], são coletadas informações das empresas como, tipo da empresa, localização, *check-ins*, de várias páginas públicas do *Facebook* visando recomendar os melhores locais na cidade de Cingapura para abrir um novo negócio. Já [Karamshuk *et al.* 2013] coletam informações do *Foursquare* para, também, recomendar melhores locais, porém em contraste com *check-ins* estáticos de [Lin *et al.* 2016], [Karamshuk *et al.* 2013] formularam modelos de mobilidade dos usuários para encontrar os melhores locais.

No melhor do nosso conhecimento, este presente trabalho se diferencia de todos os outros trabalhos disponíveis na literatura, pois objetiva identificar relações virtuais entre empresas que são geradas espontaneamente por usuários em mídias sociais. Além disso, para alcançar esse objetivo, este trabalho propõe uma nova forma de modelar essas relações, bem como uma estratégia para extrair ligações relevantes entre empresas.

3. Obtenção e Tratamento dos Dados

3.1. Escolha dos Dados

A decisão de quais dados coletar é importante para apoiar as análises programadas, bem como para atender às limitações de requisições impostas na coleta de dados. Para este estudo, foram coletadas informações do *Facebook*, pois é a plataforma de mídia social mais utilizada por empresas brasileiras para criar e manter um relacionamento com seus clientes [Ferrari 2016], portanto as informações estão amplamente disponíveis. Como o objetivo do modelo é identificar relações virtuais entre empresas usando dados de reações dos usuários no Facebook, os dados foram escolhidos considerando os objetivos do projeto, e o que existe disponível publicamente no *Facebook*. A Tabela 1 apresenta a estrutura dos dados coletados. Informações similares poderiam ser escolhidas de outras plataformas de mídia sociais, no entanto essa avaliação está fora do escopo do presente trabalho.

Tabela 1. Estrutura dos Dados

Dados de Empresas	Exemplo	Dados de Reações de Usuários	Exemplo
Id da Empresa	166765230043005	Id do Usuário	154573625550
Nome da Empresa	Rubiane Frutos do Mar	Nome do Usuário	Fulano de Tal
Localização	-25.516122, 49.231571	Id da Empresa que o Usuário Reagiu	166765230043005
Categoria	Seafood Restaurant	Tipo da Reação	<i>Like</i>
Número de Check-ins	38627		
Número de Fãs	15532		
Avaliação Média	4.6		

A primeira coluna da Tabela 1, chamada de *Dados de Empresas*, representa dados referentes às empresas propriamente ditas, como a localização geográfica de cada empresa, a categoria de cada empresa, se referindo ao setor de mercado em que a empresa está inserida, entre outras. Assim sendo, cada empresa na base de dados possui todas as informações descritas na primeira coluna. Já a segunda coluna contém dados de reações dos usuários referentes às empresas existentes na primeira coluna. Cada reação do *Facebook* parte de um usuário que se refere a uma empresa e é de um determinado tipo, podendo ser *Like;Angry;Wow;Sad;Thankful*.

Os dados da coluna *Dados de Reações de Usuários*, na Tabela 1, possibilitam criar conexões de similaridade entre empresas, principalmente se valendo das reações em comum que duas empresas possam ter. Já a coluna *Dados de Empresas* é utilizada principalmente para extrair os nomes e categorias das empresas em uma mesma comunidade, o que auxilia na avaliação dos resultados encontrados.

Todas as informações coletadas, tanto as representadas pela primeira coluna, quanto pela segunda, são abertas e publicamente disponibilizadas pela plataforma do *Facebook*. Mais detalhes podem ser obtidos no website da *API do Facebook*².

²<https://developers.facebook.com>.

3.2. Coleta de Dados

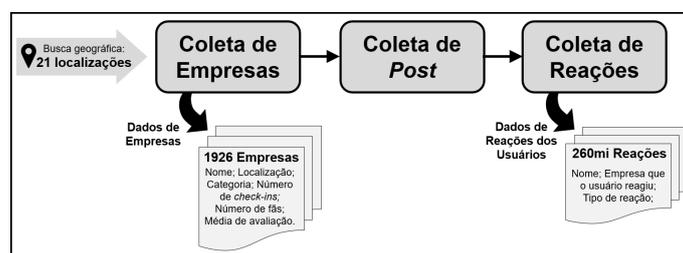


Figura 1. Processo de coleta de dados

A Figura 1 ilustra os principais passos da coleta feita neste estudo. Os dados foram coletados utilizando o *Facebook Graph API*³ e a coleta de dados foi realizada em empresas da cidade de Curitiba-PR entre novembro e dezembro de 2017. Primeiramente foram coletados dados descritos pela coluna *Dados de Empresas* da Tabela 1. O *Facebook Graph API* solicita uma coordenada geográfica e um raio em metros, em seguida retorna resultados considerando o ponto informado como o centro de uma circunferência com o raio informado, retornando até 800 resultados por busca, ou seja, até 800 empresas no nosso caso. É sabido que várias regiões podem possuir esse número de empresas. Com isso, para uma garantia maior de englobar empresas de todas as regiões de Curitiba, foram executadas vinte e uma buscas geográficas diferentes espalhadas pela cidade. Cada busca geográfica tem um raio de 2000 metros e são centradas em diferentes regiões de Curitiba, como mostrado na Figura 2.

A Figura 3 mostra o resultado dos dados coletados para essa primeira etapa em forma de um mapa de calor. Percebe-se que a região central possui uma coloração mais avermelhada, indicando que foram coletadas mais empresas naquela região, como já era esperado. Apesar disso, é possível notar também que a base de dados resultante engloba empresas espalhadas por toda a cidade de Curitiba.

Depois de obter os resultados das buscas geográficas (*Dados de Empresas* na Tabela 1), contendo dados básicos das empresas de Curitiba, foram coletadas as reações dos usuários para as páginas de cada empresa (*Dados de Reações de Usuários* na Tabela 1). Para isso, foram obtidas as reações dos *posts* das páginas das empresas. Como o histórico das páginas de algumas empresas possuem centenas de *posts* e de outras milhões, foram coletadas as reações dos 100 *posts* mais recentes de cada empresa. Existem cinco tipos de reações disponíveis no *Facebook*: *Like*; *Angry*; *Wow*; *Sad*; *Thankful*; todos os tipos foram incluídos na base de dados.

Foi coletado um total de 1986 páginas georreferenciadas, e aproximadamente 280 milhões de reações de usuários, referentes às páginas.

3.3. Limpeza de Dados

Após obtidos todos os dados, um procedimento automático de limpeza foi utilizado para aumentar a consistência das análises, o que, conseqüentemente aumenta a consistência dos resultados finais. Três principais etapas foram executadas:

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

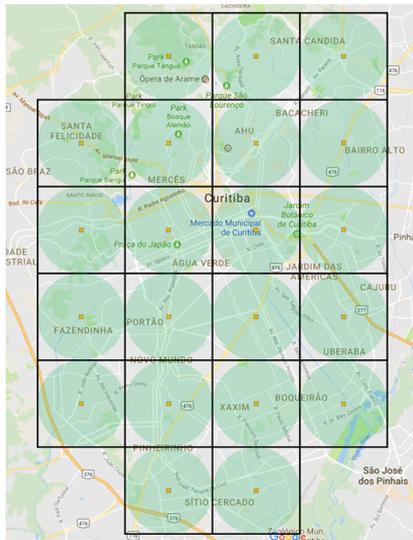


Figura 2. Pontos de coleta de dados de Curitiba/PR

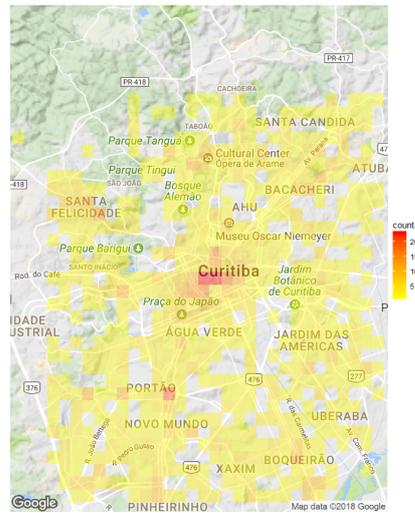


Figura 3. Mapa de calor da busca de empresas

- Remoção de registros duplicados;
- Remoção de registros inconsistentes (por exemplo, páginas sem nome, sem localização, etc.);
- Remoção de páginas que não representam empresas (por exemplo, uma praça pública) e suas reações;

Após a limpeza sobraram 1926 páginas, agora todas representando empresas, e aproximadamente 260 milhões de reações.

4. Modelagem e Estatísticas Utilizadas

4.1. Visão Geral

Os principais passos empregados neste trabalho para alcançar os objetivos propostos compõem um *framework* que é ilustrado pela Figura 4. As entradas do *framework* são os *Dados Limpos*, já tratados, como descrito na Seção 3, e a *Empresa Avaliada*, empresa escolhida para extrair suas relações virtuais. Como saídas o modelo produz a *Egonet da Empresa*, uma rede de conexões da *Empresa Avaliada*, e as *Comunidades da Empresa*, comunidades em que a *Empresa Avaliada* está inserida.

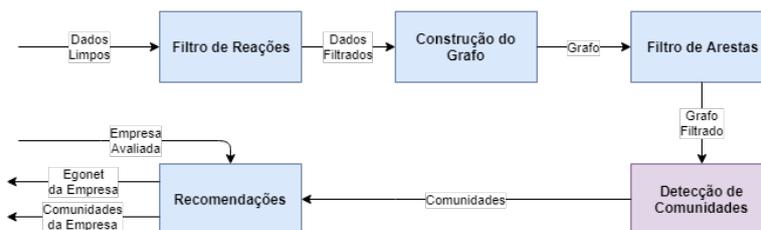


Figura 4. Visão geral do *framework* proposto para a identificação de relações virtuais entre empresas explorando reações de usuários em mídias sociais

4.2. Definição do Grafo de Relações Virtuais Entre Empresas

Com os dados obtidos, pode-se então criar um modelo para representar as relações virtuais entre empresas. O modelo escolhido foi um grafo de relações, possibilitando a extração de métricas, a criação de estruturas e a realização de análises. O grafo de empresas é um grafo não-direcionado no qual, vértices representam empresas e arestas (com peso) representam relações entre duas empresas.

O modelo de relacionamento entre empresas foi construído olhando para as reações de usuários em comum entre duas empresas quaisquer. Quanto mais reações em comum duas empresas possuírem, proporcionalmente às suas próprias reações, mais forte é a relação entre elas. Desse modo, o peso da aresta é dado pelo *Jaccard Index* [Real and Vargas 1996] do conjunto de reações de cada empresa, que representa um índice de afinidade ou similaridade entre dois conjuntos.

De uma forma mais formal, considere $B = \{b_1, b_2, \dots, b_{n_b}\}$ o conjunto de todas as empresas, onde n_b é o número total de empresas da base de dados. Considere agora U_i o conjunto de todos os usuários que reagiram à i -ésima empresa. Dessa forma, o grafo é definido como em (1).

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

Onde os vértices são empresas, $V = B$; arestas existem se empresas têm um mínimo de reações de usuários em comum, $E = \{(i, j) : |U_i \cap U_j| > \text{lowerBound}\}$; e os pesos das arestas são representadas, como em (2), pelo *Jaccard Index*:

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

4.3. Filtros e Consistência do Grafo

Com o objetivo de aumentar a consistência da informação vinculada à estrutura do grafo, a análise se vale de duas etapas importantes, um filtro de reações na base de dados e um filtro de *arestas fracas*.

Primeiramente o filtro de reações elimina reações negativas (do tipo *Angry;Sad*), pois para possíveis parcerias entre empresas o que importam são reações positivas. Em seguida, o filtro de reações elimina do modelo usuários que não se expressam frequentemente sobre o conjunto de empresas B em questão, portanto usuários com duas reações ou menos são eliminados do modelo. Por outro lado, usuários com muitas reações também são eliminados pois, como a proporção do número de arestas a que um usuário com m reações gera é quadrática $a = \frac{m(m-1)}{2}$, usuários com muitas reações contribuem com muitas arestas no grafo, da mesma forma usuários com muitas reações podem possivelmente ser robôs (*bot*), problema que aparece em vários sistemas Web [Tasse *et al.* 2017]. São, então, consideradas as reações de usuários que realizaram de 3 a $x = 174$ reações que, como mostrado na distribuição de quantidade de usuários por número de reações da Figura 5, representam 99,9% do total de reações a partir de 3.

Para os dados deste estudo de 260 milhões de reações foram filtradas aproximadamente 40 milhões de reações sobrando 220 milhões.

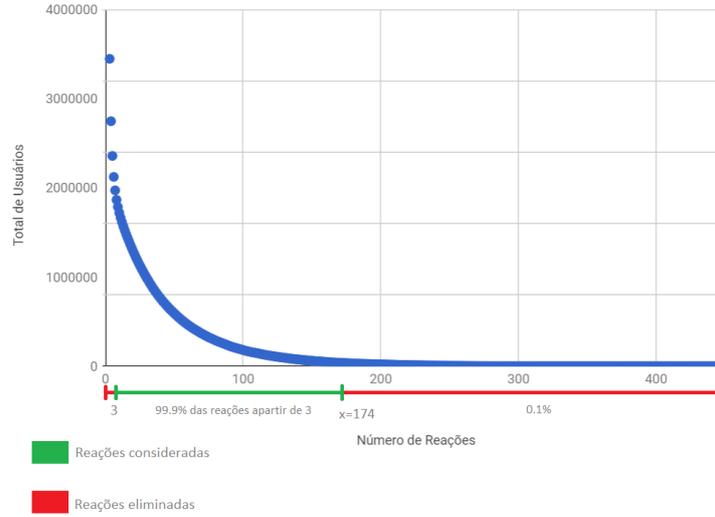


Figura 5. Distribuição da quantidade de usuários por número de reações

O filtro de *arestas fracas* elimina possíveis ruídos na estrutura do grafo, desse modo são removidas arestas classificadas como *arestas fracas*. A classificação de uma aresta, em *aresta fraca* ou não, é feita realizando um experimento aleatório, em que as reações são aleatoriamente distribuídas entre todas as possíveis arestas do grafo, e então arestas da estrutura original com peso similar ao experimento podem ser consideradas *arestas fracas*. O valor esperado e a variância do peso de uma aresta qualquer, no experimento aleatório (com distribuição binomial), são dados respectivamente por (3) e (4):

$$\mu = E[X] = \frac{n_r}{n_c} \quad (3)$$

$$\sigma^2 = Var[X] = \frac{n_r}{n_c} \left(1 - \frac{1}{n_c}\right) \quad (4)$$

Onde, $n_r = \frac{\sum_{i,j:i \neq j} W(i,j)}{2}$, $n_c = \frac{n_b(n_b-1)}{2}$ e n_b é o número de empresas da base de dados.

$$lowerBound = \mu + 3\sigma \quad (5)$$

Assim sendo, uma aresta entre as empresas i e j é fraca se: $W(i, j) \leq lowerBound$, como definido em (5).

Para os dados coletados neste estudo os valores calculados foram: $\mu = 120,289$, $\sigma = 10,96$ e $lowerBound = 153,195$. Para os dados deste estudo foram eliminadas 978410 arestas, com $W(i, j) \leq 153,195$, sobrando então um total de 223939 arestas no grafo com menor probabilidade de serem aleatórias.

4.4. Detecção de Comunidades

Dada uma rede consistente de relacionamento entre empresas, um passo importante para alcançar o objetivo do estudo é detectar comunidades de empresas. Como o grafo de empresas deste estudo não é esparso, algoritmos de detecção de comunidades que se

baseiam na busca de cliques ou subgrafos densos com solução ótima, como o *Clique Percolation Method* [Palla *et al.* 2005], apresentam complexidade muito alta, portanto não são aplicáveis neste estudo.

[Raghavan *et al.* 2007] propuseram um algoritmo de detecção de comunidades em redes baseado em *Label Propagation* que usa iterativamente a troca de *labels* entre vértices até uma concordância, portanto promovendo uma convergência. Este algoritmo opera em tempo quase linear e não utiliza informações prévias, como heurísticas, sobre as comunidades presentes na rede de entrada. Para o nosso problema são interessantes comunidades que tenham, no mínimo quatro empresas (*minSize*), e no máximo trinta empresas (*maxSize*), pois comunidades muito grandes perdem a coesão nas recomendações.

Diante disso, é proposto um algoritmo iterativo, o Algoritmo 1, para detecção de comunidades de empresas. As entradas desse algoritmo são, o grafo de empresas (*BusinessGraph*), o tamanho mínimo (*minSize*) e tamanho máximo (*maxSize*) das comunidades, e a saída é um conjunto de comunidades de empresas. Em cada iteração do algoritmo são realizados os seguintes passos:

- Detecção de comunidades com o algoritmo descrito em [Raghavan *et al.* 2007];
- Comunidades de empresas com tamanho de *minSize* a *maxSize* são guardadas para retorno final;
- União das comunidades detectadas, e não guardadas para retorno, compõe um novo grafo, nomeado de *G*;
- Do grafo *G*, cortam-se arestas fracas formando o grafo para a próxima iteração.

Algoritmo 1: Algoritmo de Detecção de Comunidades de Empresas

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

```

1  $G \leftarrow BusinessGraph$ ;
2  $communities \leftarrow \emptyset$ ;
3  $minEdge \leftarrow \min_{i,j \in G} W(i, j)$ ;
4  $counter \leftarrow 1$ ;
5 while  $|G| > minSize$  do
6    $counter \leftarrow counter + 1$ ;
7   /* Método de [Raghavan et al. 2007] */
8    $detectedComm \leftarrow labelPropCommDetection(G)$ ;
9    $G \leftarrow empty\ graph$ ;
10  for  $c \in detectedComm$  do
11    if  $|c| > minSize$  and  $|c| < maxSize$  then
12       $communities \leftarrow communities \cup \{c\}$ ;
13    else
14       $G \leftarrow G \cup c$ ;
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 return  $communities$ ;

```

As comunidades detectadas pelo Algoritmo 1 são subgrafos que tendem a ser densos (mais arestas), portanto as empresas dentro de uma mesma comunidade possuem uma coesão maior do que empresas aleatoriamente escolhidas no grafo. Esta coesão é formada por arestas criadas espontaneamente por usuários, sem informações adicionais que poderiam incluir um viés na detecção de comunidades.

5. Resultados

Como saída do *framework* discutido são apresentadas para o usuário todas as comunidades, detectadas pelo Algoritmo 1, que contém a empresa a ser avaliada pelo usuário, e a *Egonet* desta empresa, que consiste de um subgrafo de todas as arestas da empresa avaliada. Como existem empresas com *Egonets* consideravelmente grandes, neste trabalho foi limitado o tamanho da *Egonet* para no máximo oito empresas incluindo a empresa avaliada. Observe que esse é um parâmetro que pode ser ajustado para cada caso, assim como a limitação do tamanho das comunidades.

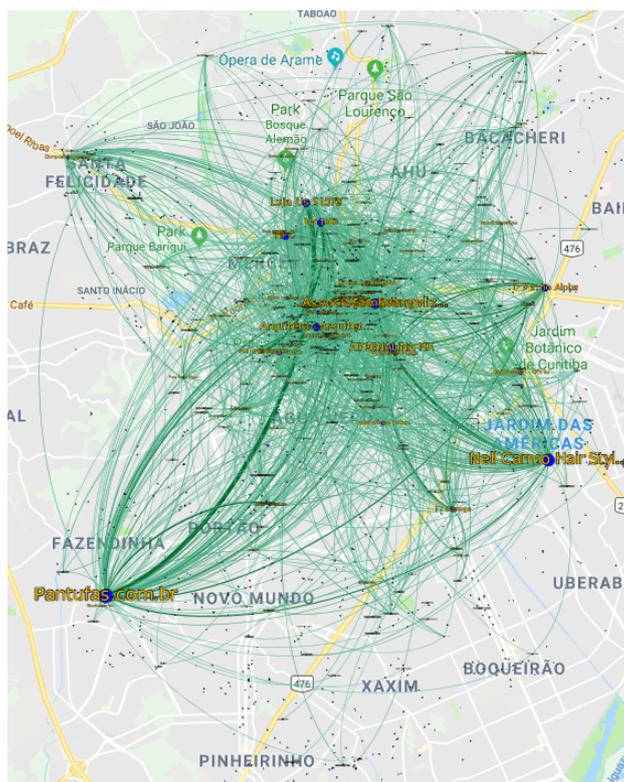


Figura 6. Grafo de empresas no mapa de Curitiba

A Figura 6 ilustra o grafo completo, construído pelo *framework* proposto, impresso no mapa de Curitiba/PR. Para facilitar a visualização, foram impressas somente as arestas com peso maior que 1500. Quanto mais espessa a aresta mais forte é a relação entre os nós. Cada nó no grafo é mostrado de acordo com a localização da empresa que ele representa. Pode-se de notar que existem pontos longe do centro da cidade com uma densidade considerável de arestas que vão em direção ao centro da cidade, indicando uma atividade forte também em bairros. Como esse grafo é muito grande, a extração de informações úteis fica muito complicada a olho nu, justificando a extração de comunidades e de *egonets*.

Após a execução do algoritmo de identificação de comunidades, conforme discutido na seção anterior, foram detectadas 141 comunidades de empresas, com tamanho entre 4 e 30 empresas, na cidade de Curitiba/PR. A Figura 7 ilustra uma comunidade que contém empresas de entretenimento ("Blood Rock Bar", "Peppers" e "SSCWB - Shinobi Spirit") e comida ("Ca'dore Comida Descomplicada"), portanto são empresas unidas pelo

contexto "lazer". Já a comunidade ilustrada na Figura 8, estão coesas pelo contexto que pode ser chamado de "moda", pois contém empresas do setor de salão de beleza ("Studio Andressa Mega Hair"), lojas de roupas ("New Port Container") e agências de modelos.

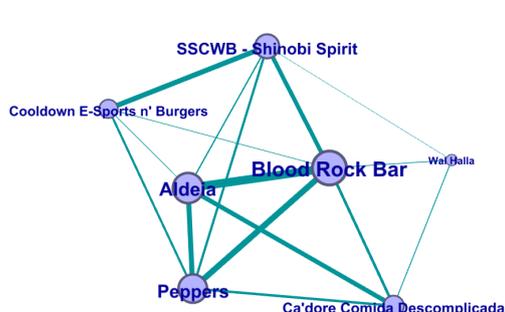


Figura 7. Comunidade de empresas relacionadas à lazer

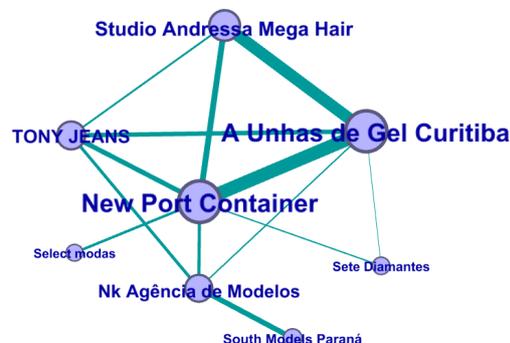


Figura 8. Comunidade de empresas relacionadas à moda

Com uma inspeção visual de todas as comunidades encontradas, nota-se que os estabelecimentos pertencentes às comunidades possuem forte relação semântica. Discutir e apresentar o resultado de todas as comunidades é impraticável. Com isso, foi realizado um processo de agrupamento visando sumarizar a intuição observada.

Todas os locais coletados do *Facebook* são classificados em sete categorias mãe: *Interest*; *Community Organization*; *Media*; *Public Figure*; *Businesses*; *Non-Business Places* e *Other*. Todas possuem subcategorias, mas o número maior de subcategorias (22), bem como a maior diversidade, estão subordinadas à categoria *Business*. Com isso, todas as subcategorias dentro de *Interest*, *Community Organization*, *Media*, *Public Figure*, *Non-Business Places* e *Other* foram consideradas como a categoria mãe. Assim, todas as subcategorias de *Interest*, por exemplo, são transformadas em *Interest*.

Para a categoria *Business* foram consideradas todas as subcategorias. No total existem 22, e cada uma possui outras subcategorias (sendo então "subsubcategorias"). As subsubcategorias de *Business* foram desconsideradas, pois o nível de especialização não foi considerado interessante para as análises realizadas. A subcategoria *Advertising or Marketing* de *Business* possui, por exemplo, as subcategorias *Advertising Agency* e *Copywriting Service*. Nesse exemplo, todas as subcategorias dentro de *Advertising or Marketing* são consideradas *Advertising or Marketing*. Esse mesmo procedimento foi feito para todas as outras subcategorias dentro de *Business*.

Após feito esse procedimento, foi obtido um total de 28 nomes de categorias de locais (6 provenientes das categorias mãe, sem utilizar as subcategorias, e 22 subcategorias de *Business*). Foi, então, construído um vetor de *features* com essas 28 categorias, e contabilizado, para cada comunidade, o número de ocorrência de estabelecimentos em cada uma das 28 subcategorias. Em seguida, esses valores são normalizados com base no número máximo de locais em uma determinada *feature* para cada comunidade. De posse desse vetor de *features*, foi calculada a similaridade de diferentes comunidades e realizado um agrupamento. O agrupamento utilizado foi o k-means, com a distância euclidiana. Foram testados vários valores para *k*, e o valor com a menor soma dos erros quadráticos (*Sum of Squared Errors*, SSE) foi o $k = 8$. Com isso, foi mantido esse



Figura 9. Nuvem de palavras para as categorias de grupos de comunidades similares

valor neste trabalho. Em seguida, foi criada uma nuvem de palavras para cada grupo encontrado. Essas nuvens de palavras são mostradas na Figura 9.

Nessas nuvens de palavras foram utilizados os nomes das categorias de locais consideradas. Observe a similaridade surpreendente entre as categorias de locais em cada grupo. Por exemplo, o grupo 1 é relacionado com lazer, contendo predominantemente estabelecimentos relacionados com comida, bebida e entretenimento, o grupo 2 contém a maioria das localidades relacionadas a beleza e estilo, já o grupo 3 é mais relacionado com estabelecimentos sobre produtos e serviços automotivos.

A Figura 10 mostra a Egonet da empresa Rubiane, um restaurante de frutos do mar, que foi escolhida arbitrariamente para a análise, e a Figura 11 mostra uma comunidade detectada em que a empresa Rubiane faz parte. Pela Egonet da empresa, é possível visualizar as conexões diretas que a empresa já possui virtualmente com outras empresas, contudo pelas comunidades é possível notar conexões que não são diretas com a empresa avaliada. Estas conexões, por estarem inseridas dentro de uma comunidade detectada pelo Algoritmo 1, são coesas (um subgrafo denso) e podem representar possíveis parcerias não-triviais para a empresa avaliada. Por exemplo, a empresa Quintal do Monge

não aparece na *Egonet* da empresa Rubiane na Figura 10, porém aparece na comunidade mostrada na Figura 11.

A empresa Rubiane poderia aproveitar esse resultado para aumentar suas vendas criando parcerias comerciais, como vender produtos e serviços em conjunto com as empresas que aparecem nos resultados, bem como efetuar parcerias de marketing e divulgações em conjunto. Para o caso do restaurante analisado, é observado que concorrentes apareceram na mesma comunidade: "Braseirinho Frutos do Mar". Para o caso envolvendo restaurantes, isso pode ser explicado pelo fato de que usuários tendem a frequentar diversos restaurantes e alguns podem ser do mesmo tipo. No entanto, isso não é um problema da abordagem proposta, uma vez que o empresário é que decide a melhor estratégia de como aproveitar os resultados, inclusive uma parceria poderia ser feita com estabelecimentos concorrentes. De qualquer forma, esses casos merecem uma atenção especial.



Figura 10. Egonet Rubiane

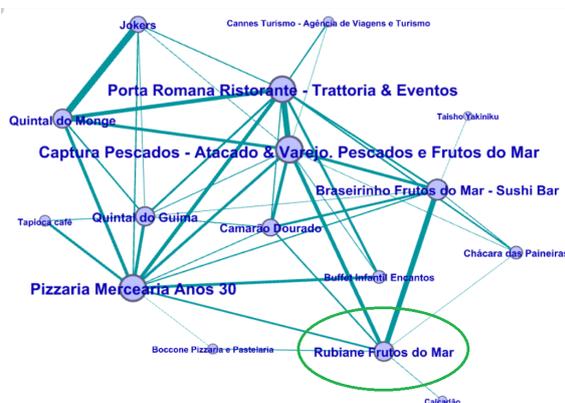


Figura 11. Comunidade de empresas relacionadas à Comida (incluindo a empresa Rubiane)

6. Conclusão e Trabalhos Futuros

Existem muitos desafios acadêmicos em conseguir sistemas de recomendação que produzam resultados satisfatórios. Um dos maiores desafios é o fato de uma recomendação ser algo subjetivo, passível de interpretações diferentes sob diferentes pontos de vista. O *framework* apresentado neste estudo visa contribuir para a comunidade científica com uma nova forma de abordar de forma objetiva, com um modelo concreto, a tarefa subjetiva de recomendar parcerias empresariais. Este trabalho mostra, utilizando dados em larga escala do Facebook, que a nossa abordagem é promissora e tem o potencial de alavancar pesquisas que podem gerar novos mecanismos e serviços.

Uma questão interessante que surge da análise deste estudo é o fato de aparecerem empresas concorrentes na mesma comunidade. Como o objetivo do trabalho é apontar novas parcerias empresariais, pode ser interessante determinar uma forma de detectar se duas empresas são concorrentes, para melhorar o desempenho das recomendações. Além disso, existe um grande espaço para trabalhos futuros a serem feitos a partir deste. Uma das possibilidades é analisar a correlação espacial nas comunidades, e outra seria analisar a dinâmica temporal das comunidades, tarefas que estão sendo estudadas pelos pesquisadores deste trabalho.

Agradecimentos

Este trabalho foi parcialmente apoiado pelo projeto CNPq-URBCOMP (processo 403260/2016-7).

Referências

- Agnihotri, R., Dingus, R., Hu, M. Y., and Krush, M. T. (2016). Social media: Influencing customer satisfaction in b2b sales. *Industrial Marketing Management*, 53:172–180.
- Barbier, G., Tang, L., and Liu, H. (2011). Understanding online groups through social media. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4):330–338.
- Chaffey, D. (2016). Global social media research summary 2016. *Smart Insights: Social Media Marketing*.
- Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., Zhang, X. L., and Zhang, J. (2016). Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics*, 22(1):270–279.
- Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. In *Proc. of ICWSM'11*, pages 81–88, Barcelona, Spain.
- Cranshaw, J., Schwartz, R., Hong, J. I., and Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In *Proc. of ICWSM'12*, Dublin, Ireland.
- Culotta, A. and Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing science*, 35(3):343–362.
- Ferrari, V. C. (2016). *Content marketing and brand engagement on social media: a study of Facebook's posts in the ecommerce industry in Brazil*. PhD thesis.
- Hudson, S. and Thal, K. (2013). The impact of social media on the consumer decision process: Implications for tourism marketing. *Journal of Travel & Tourism Marketing*, 30(1-2):156–160.
- Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., and Mascolo, C. (2013). Geo-spotting: Mining online location-based services for optimal retail store placement. In *Proc. of ACM KDD'13*, pages 793–801, Chicago, Illinois, USA.
- Lin, J., Oentaryo, R., Lim, E.-P., Vu, C., Vu, A., and Kwee, A. (2016). Where is the goldmine?: Finding promising business locations through facebook data analytics. In *Proc. of Hypertext'16*, pages 93–102, Halifax, Canada. ACM.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.
- Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106.
- Real, R. and Vargas, J. M. (1996). The probabilistic basis of jaccard's index of similarity. *Systematic biology*, 45(3):380–385.
- Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (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.
- Tasse, D., Liu, Z., Sciuto, A., and Hong, J. I. (2017). State of the geotags: Motivations and recent changes. In *Proc. of ICWSM'17*, pages 250–259, Montreal, Canadá.
- Trainor, K. J., Andzulis, J. M., Rapp, A., and Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social crm. *Journal of Business Research*, 67(6):1201–1208.