

Caracterização do TripAdvisor utilizando técnicas de Redes Complexas: Estudo de caso sobre os restaurantes cidade de Tiradentes-MG

Lucas G. S. Félix¹

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

lucasgsfelix@gmail.com

Resumo. *Este trabalho apresenta uma caracterização do site TripAdvisor possuindo como estudo de caso a cidade de Tiradentes onde foram avaliados seus restaurantes. Para tal avaliação foram utilizadas técnicas de Redes Complexas e Aprendizado de Máquina. Por meio de nossas avaliações foi possível perceber que a região em que um restaurante está localizado pouco influencia em sua reputação. Além disso foi percebido um viés nas avaliações de certos restaurantes, os quais possuem apenas avaliações entre si, formando assim uma componente desconexa a rede.*

1. Introdução

A indústria do turismo possui a cada ano um crescimento contínuo [Khatibi et al. 2018], gerando para cidades e comerciantes um fluxo de montantes que varia de acordo com temporadas, como estações do ano em que locais são mais visitados. E eventos, como shows, jogos esportivos, entre outros, os quais podem auxiliar na movimentação turística de um cidade.

Antes de ir um local para visita, usuários costumam se planejar, visando calcular gastos, e planejar áreas as quais o mesmo deseja conhecer. Para isso, a utilização de sites especializados em turismo e os chamados *travel social networks*, como *TripAdvisor*¹, *Gogobot*², *Virtual Tourist*³ e outros, tem sido cada vez maior [Leite et al. 2013].

Assim, proveniente destes e outros *websites*, a quantidade de dados disponível atualmente na rede sobre cidades, restaurantes, hotéis, transporte e atrações é ampla, se tornando difícil tanto para usuários, os quais pretendem visitar cidades, avaliar suas opções e adequá-las ao seu perfil. Também se tornando uma tarefa árdua, para empresas fornecedoras de serviços analisar todas suas avaliações disponíveis na rede. Podendo saber quais são suas fontes de reclamação e seus pontos positivos, avaliar quais são os períodos de alta e baixa demanda, além do perfil de usuários os quais visitam seu estabelecimento.

Pensando nisto, este trabalho tem por objetivo caracterizar, o comportamento de seus usuários por meio de dados disponíveis no site *TripAdvisor*. Destaca-se que a motivação para escolha do *TripAdvisor* como site para coleta dos dados se dá por este ser atualmente o maior site de viagens do mundo ⁴, com mais de 600 milhões de avaliações e 315 milhões de usuários apenas em dispositivos móveis.

¹<https://tripadvisor.com>

²<https://gogobot.com>

³<https://www.virtualltourist.com>

⁴<https://tripadvisor.mediaroom.com/BR-about-us>

Além disso, objetiva-se investigar quais são os motivos de um estabelecimento ter sua avaliação boa ou ruim, se são influenciados por fatores como localidade. Para realização de tal avaliação, será utilizado em nossa metodologia técnicas de mineração de dados (para coleta de dados e análise de sentimentos) e redes complexas (para caracterização e avaliação da rede). Para a modelagem da rede serão utilizadas três diferentes tipos de abordagens, considerando as seguintes ligações: (i) Usuário-Usuário, (ii) Estabelecimento-Estabelecimento e (iii) Usuário-Estabelecimento.

Para avaliação da metodologia proposta optou-se por limitar em apenas uma cidade para análise, sendo esta Tiradentes. A escolha desta cidade se motiva pelo fato de sua economia estar em grande maioria baseada no turismo [dat], além de ser uma importante cidade histórica de Minas Gerais. Além disso, o escopo foi limitado apenas aos restaurantes da cidade, pois os mesmos possuem a maior quantidade de atributos e dados que podem ser trabalhados de forma melhor.

2. Trabalhos Relacionados

Redes sociais podem ser utilizadas como fonte de informações sobre pessoas e lugares devido ao vasto volume de dados [Leite et al. 2013]. Nesse contexto diversos trabalhos se utilizam de redes sociais de variados fins visando explorar estes dados e trazer informações seja para usuários ou empresas. Como exemplo temos a avaliação de redes sociais de reclamação [Felix et al. 2018], caracterização do comportamento de usuários [Benevenuto et al. 2009, Hobbs and Burke 2017], avaliação de doenças [Souza et al. 2019], detecção de eventos [Corney et al. 2014] e artigos que se assemelham mais com a proposta deste trabalho: análise de dados para planejamento/recomendação/análise de viagens [Leite et al. 2013, Kim et al. 2009, Hao et al. 2010, Savage et al. 2012, Lu et al. 2012, Gionis et al. 2014, Kotiloglu et al. 2017, Guo et al. 2018, Khatibi et al. 2018].

O trabalho de [Leite et al. 2013], propõe uma ferramenta para ajuda na localização de pontos turísticos com base no gosto do usuário. Para isso os autores se utilizam de dados de diversas plataformas como: *Where Are You Now ? (WAYN)*, *GoGoBot*, *Virtual Tourist*, onde a partir dos dados são geradas *tags* as quais sumarizam os dados, sendo apresentadas uma nuvem de palavras com termos mais frequentes e análise sentimental das avaliações. Na proposta de [Kim et al. 2009], os autores visam a auxiliar na escolha do usuário de um próximo local a visitar com base em lugares anteriores de sua preferência. Para isso, os autores utilizam-se de dados gerados na própria ferramenta como: local de visita, características do local e motivos para realização da visita.

No trabalho de [Hao et al. 2010], é proposto um método para modelagem de tópicos com base em localização chamado *Location-Topic Model*. No modelo proposto é possível caracterização tópicos locais e globais (os quais são comuns para diversas localizações. Além disso, o trabalho propõe a representação e por meio da utilização dos tópicos revelar a similaridade entre locais. Em [Savage et al. 2012], os autores propõe um algoritmo para recomendação de lugares baseado em classificação textual o qual leva em consideração preferências previas, restrições de localidade e o sentimento do usuário naquele momento. O trabalho de [Lu et al. 2012], utiliza-se de dados de *check-in* para propor a recomendação de lugares, realizando podas por meio de múltiplas restrições. Neste trabalho são consideradas questões como restrições múltiplas, relações sociais, pro-

priedades temporais, além de ser proposto um *framework* que o qual funciona de maneira paralela.

Em [Gionis et al. 2014], os autores focam no problema de recomendação de áreas urbanas com base em dados de mídias sociais de localização, como o *FourSquare*. Na proposta são considerados fatores como tipos de locais os quais o usuário deseja visitar, limitações com relação a tempo gasto na visita e a ordem em que o usuário deseja visitar estes locais, além do mérito de se visitar um local. O trabalho de [Kotiloglu et al. 2017], possui como proposta o *rankeamento* de locais para posterior realização de uma visita. Para isso os autores se utilizam de dados de mídias sociais as quais são utilizados para realizar um *ranking* de locais os quais devem ser visitados. A proposta se auto-ajusta com base em uma função de otimização que define a satisfação de um usuário.

A proposta de [Guo et al. 2018], utiliza-se de dados medias sociais em uma abordagem multi-perspectiva para recomendação de locais. Além disso o trabalho aplica sobre dados de blogs de viagens algoritmos para identificação de rotas turísticas as quais trazem maior satisfação ao usuário. O trabalho de [Khatibi et al. 2018], utiliza-se de dados do *TripAdvisor* aliado com dados de clima para realização da predição da quantidade de visitas que uma atração pode ter.

Este artigo se diferencia dos trabalhos apresentados anteriormente por considerar uma metodologia a qual aplica redes complexas e aprendizado de máquina para avaliação e caracterização de dados de viagem do *TripAdvisor*, sendo este tipo de abordagem nunca utilizada antes para avaliação dados de viagem, dado nosso conhecimento.

3. Metodologia

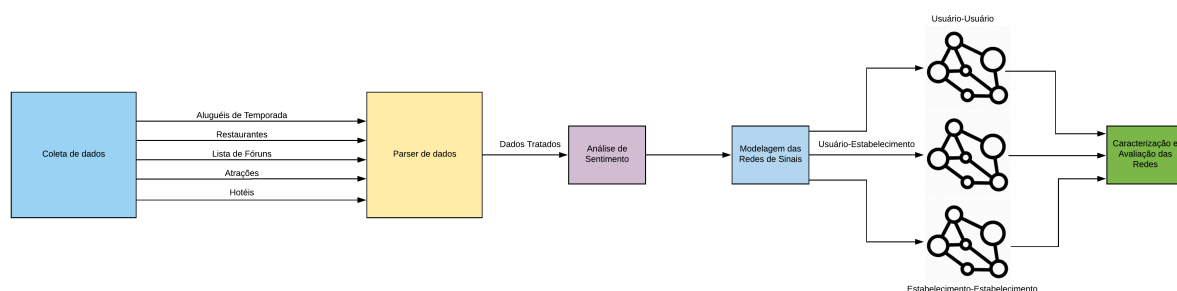


Figure 1. Descrição da metodologia proposta

O presente trabalho consiste em um estudo exploratório que aplica técnicas de mineração de dados, aprendizado de máquina e redes complexas, tendo como objetivo final caracterização do setor de turismo na cidade de Tiradentes por meio coletados na internet.

Por meio da Figura 1 é possível ver os passos da metodologia proposta. Para realização de tal, a metodologia proposta consiste de quatro etapas. (i) Coleta e parser de dados, responsável pela coleta dos dados do site *TripAdvisor* e por deixar os mesmos em um formato semiestruturado. (ii) Análise de sentimento, consiste em um método para definir a polaridade de um texto. A mesma é aplicada na metodologia proposta para definir a relação dos usuários com relação a estabelecimentos. (iii) Modelagem da rede, nessa

etapa são modeladas as redes formando diferentes relações a serem analisadas. Serão modeladas três diferentes tipos de interações **Usuário-Usuário**, **Usuário-Estabelecimento** e **Estabelecimento-Estabelecimento**. (iv) Caracterização e avaliação das redes geradas, onde é realizado o processo de extração de informação sobre os dados gerados.

3.1. Coleta e Parser de Dados

Para realização da coleta de dados, foram implementados 5 diferentes *crawlers* responsáveis pela coleta de diferentes categorias de dados do site *TripAdvisor*. Por meio do *crawler* desenvolvido, foram coletados as seguintes categorias de dados:

- **Aluguéis de Temporada:** Classe de dados que representa proprietários de casa que desejam alugar sua casa por um período de tempo que geralmente é superior a um dia.
- **Atrações:** Classe de dados que representa as atrações presentes em uma cidade, como museus, igrejas e parques.
- **Hotéis:** Classe de dados que representa estabelecimentos que alugam um quarto por períodos baseados em períodos diários.
- **Lista de Fóruns:** Classe de dados que representa dúvidas gerais sobre um tópico específico de algum futuro visitante. Essas dúvidas são em geral respondidas por outros turistas que já estiveram no local.
- **Restaurantes:** Classe de dados que representa os restaurantes.

Vale destacar que cada uma das categorias podem ter diferentes atributos de dados. Por exemplo: Restaurantes possuem um atributo cozinhas, os quais descrevem possíveis variedades de pratos que pode haver naquele restaurante, como: "Cozinha italiana, especializada em massas". Enquanto as outras categorias não apresentam tal atributo.

Para limitar o escopo do trabalho proposto e testar de maneira melhor a metodologia proposta, optou-se por selecionar apenas uma cidade para a avaliação. Deste modo, foi escolhida a cidade de Tiradentes para avaliar a metodologia. A escolha de Tiradentes se deu, pois a cidade possui um grande foco turístico em sua economia. Desta maneira, a mesma possui uma quantidade razoável de avaliações em estabelecimentos para que a análise fosse realizada.

A Tabela 1 mostra a quantidade de dados para cada uma das categorias avaliadas. Por meio dela é possível perceber que as categorias de aluguéis de temporadas e lista de fóruns possuem poucos dados para avaliação quando comparadas as outras categorias. Para este estudo selecionou-se apenas os dados de restaurantes da cidade para avaliação, dado que os mesmos possuem uma maior quantidade de dados sobre os quais será possível realizar uma avaliação mais detalhada.

	Aluguéis de Temp.	Atrações	Hotéis	Lista de Fóruns	Restaurantes
Quant. de instâncias	16	15237	13601	111	26552
Quant. de atributos	30	13	18	16	16

Table 1. Informações da base de dados

3.2. Análise de Sentimento

Antes da modelagem das redes, foi utilizado um método de análise de sentimento para definir qual o tipo de ligação (positiva ou negativa) entre indivíduos do grafo, gerando

assim a rede de sinais. A análise de sentimento, de maneira geral, compreende na técnica de detectar de maneira automática a polaridade de um texto, podendo esta polaridade ser positiva, negativa ou neutra [Araújo et al. 2014]. Na metodologia proposta foi utilizado como método de análise de sentimento o sistema [Araújo et al. 2014], o qual possui 18 implementações de diferentes métodos de algoritmos de análise sentimental. A Figura 2 ilustra a rede de sinais gerada após a análise de sentimento. Desta maneira, a ligação entre um estabelecimento e um usuário será caracterizada pelo sentimento agregado ao *review* feito pelo cliente sobre o serviço/produto daquele comércio.

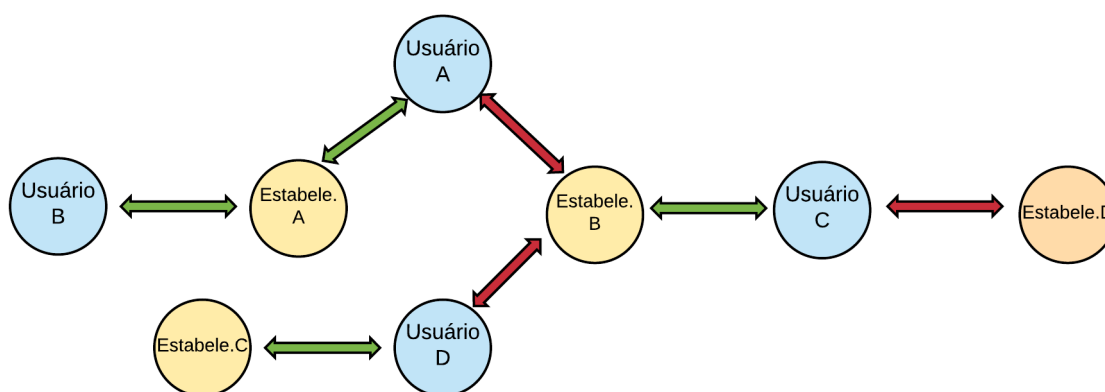


Figure 2. Exemplo da rede de sinais gerada após a análise de sentimento

3.3. Modelagem da Rede

Na etapa de modelagem das redes foram utilizadas três diferentes abordagens para avaliação: Usuário-Usuário, Usuário-Estabelecimento e Estabelecimento-Estabelecimento. Abaixo são definidos os processos que será feito em cada uma das redes:

- **Usuário-Usuário:** Por meio de ligações feitas em redes de Usuário-Usuário é possível verificar usuários influentes por compartilhar opiniões similares a outros usuários.
- **Usuário-Estabelecimento:** Por meio de grafos bipartidos e uma relação Usuário-Estabelecimento é possível modelar estabelecimentos mais conectados, além de verificar quais são os sinais em geral ligados a tal.
- **Estabelecimento-Estabelecimento:** Por meio de ligações de estabelecimentos é possível verificar locais em comum visitados por usuários, podendo estes serem usados para inferir o gosto de usuários.

3.4. Caracterização e Avaliação das Redes

Para realização da caracterização e avaliação das redes foram utilizadas diferentes abordagens de redes complexas. Para caracterização das foram aplicadas as seguintes técnicas: Densidade, diâmetro, coeficiente de clusterização, reciprocidade, assortatividade de grau, grau máximo e mínimo, e força máxima de entrada e saída.

Além dessas técnicas de caracterização, aplicou-se abordagens para uma análise mais profunda sobre as redes geradas. Para isso foi aplicada uma técnica de análise de comunidades com sobreposição proposta por [Lancichinetti et al. 2009].

Algoritmos de detecção de comunidades encontram subgrupos coesos dentro de grafos, permitindo assim uma análise mais minuciosa de diferentes *clusters*. A utilização de algoritmos que permitem a sobreposição auxilia nessas avaliações e atende ao dinamismo das redes sociais, possibilitando que usuários não fiquem engessados a apenas um subgrupo do grafo [Xie et al. 2013].

4. Resultados

Nesta seção são avaliados os resultados obtidos com nossa metodologia. Para isso primeiramente é descrito as características das redes geradas, sendo depois apresentada uma discussão sobre os resultados obtidos em nosso estudo de caso.

4.1. Características das Redes Geradas

Em nossa avaliação experimental foram geradas três redes as quais reúnem as principais características dos dados. Assim foram geradas redes conectadas por ligações Restaurante-Restaurante, Cliente-Restaurante e Cliente-Cliente. Por meio da tabela 2, é possível ver as características das redes. Por serem redes não direcionadas algumas métricas não nos trazem tantas informações. A reciprocidade por exemplo, é igual a 1 em todas as redes, já que esta métrica caracteriza a probabilidade de vértices serem mutuamente conectados. Assim como, força máxima de entrada e saída, que representa nós com maior número de graus de entrada e saída respectivamente.

Abaixo são apresentados os atributos que distinguem as redes, mostrando detalhes das mesas e como estas podem trazer informações úteis para clientes e restaurantes.

A rede Restaurante-Restaurante se caracteriza por possuir a menor quantidade de vértices e arestas, sendo formada pela ligação entre restaurantes que possuem clientes em comuns. A densidade da rede, que caracteriza a probabilidade de todo os nós serem conectados, apresenta um valor relativamente alto de 0.289. Isso nos mostra que, clientes tendem a visitar o mesmo conjunto de restaurantes. O coeficiente de *clustering* nos mostra que a rede possui características de mundo pequeno [Albert et al. 1999], dado o alto valor encontrado, mostrando assim que a proximidade entre os estabelecimentos é alta. A baixa assortatividade da rede nos mostra que os nós não tendem a se conectar com nós parecidos. Em outras palavras, isso pode mostrar que as ligações são formadas por clientes que vão a restaurantes com diferentes características: Por exemplo, um cliente que vai a um restaurante italiano não costuma realizar *reviews* em outro restaurante italiano.

A rede Cliente-Restaurante é um grafo bipartido, o qual é formado pela ligação entre um cliente o qual realiza um *review* em um estabelecimento. Esta rede possui a maior quantidade de vértices entre todas as redes, onde desses 13073 são usuários e 169 são restaurantes. Por este ser um grafo bipartido, a sua densidade é igual a 0, assim como seu coeficiente de *clustering*, também influenciando no fato da mesma ser disassortativa.

A rede Cliente-Cliente é caracterizada por possuir a maior quantidade de arestas entre as redes. A mesma é formada pela conexão entre clientes que fizeram *review* em um mesmo estabelecimento. O grafo possui um coeficiente de *clustering* alto, assim

como a rede Restaurante-Restaurante, mostrando assim também características de mundo pequeno. A densidade da rede mostra que a rede possui poucas conexões comparado ao número de vértices, mostrando assim que maiorias do usuários não vão aos mesmos restaurantes. Diferente das outras redes gerada, Cliente-Cliente possui uma tendência assortativa, mostrando que os usuários tendem a ir nos mesmos locais.

Métricas	Restaurante-Restaurante	Cliente Restaurante	Cliente-Cliente
# Vértices	167	13126	12954
# Arestas	4127	26552	10457947
Densidade	0.298	0.0	0.125
Diâmetro	4	10	5
Reciprocidade	1.0	1.0	1.0
Coef. de <i>Clustering</i>	0.693	0.0	0.683
Assortividade	0.003	-0.31	0.219
Grau Máx.	117	3298	8718
Força Saída	117.0	3298.0	8718.0
Força Entrada	117.0	3298.0	8718.0

Table 2. Informações das Redes Geradas

4.2. Análise de Comunidades com Sobreposição

Para realização das análises de comunidades com sobreposição foi utilizado o algoritmo proposto por [Lancichinetti et al. 2009]. Contudo, por termos redes muito grandes houve uma dificuldade na execução e avaliação dos resultados, sendo ao final obtidos resultados apenas de uma das três redes geradas. Das redes formadas foi possível obter resultados apenas da rede *Restaurante-Restaurante*, dado que a mesma é a menor dentre todas. A rede *Cliente-Restaurante*, por ser uma rede bipartida não nos traria tantas informações na análise de comunidades. Já a rede *Cliente-Cliente*, por ser uma rede muito densa e com muitos nós não foi possível computa-las em tempo viável.

Nos resultados da *Restaurante-Restaurante* foram encontradas duas comunidades, na qual a primeira possui 150 vértices, e a segunda possui apenas 17 vértices não possuindo a sobreposição como esperado dado que a rede possui duas componentes diferentes como é possível ver na Figura 3. Por meio da visualização da rede *Restaurante-Restaurante* é possível perceber que dada a construção da rede existem restaurantes que são visitados e avaliados por mesmo conjunto de pessoas formando uma clique. Enquanto outros restaurantes formam uma componente a qual nem todos visitantes passam por todos os locais.

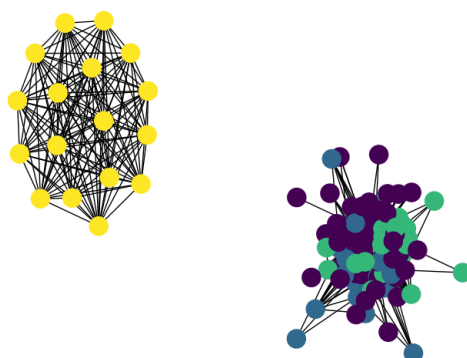


Figure 3. Visualização da rede Restaurante-Restaurante: Perceba como é possível ver as duas componentes isoladas uma da outra

4.3. Avaliação por localidade

Nesta seção nos propusemos a avaliar se a localidade de um restaurante infere em seus *reviews*. Assim, gostaríamos de analisar se restaurantes bem avaliados estão localizados em locais próximos, ou se região em que um estabelecimento está pouco afeta em suas avaliações.

Para que fosse realizada essa avaliação, foi utilizada a *API GoogleMaps* ⁵, a qual permite por meio do endereço de um local saber suas coordenadas. Após realizada a coleta, utilizando o agregador de métodos de análise sentimental *iFeel* [Araújo et al. 2014], foi feita a junção dos dados de localidade com a média do sentimento de cada local. Na Figura 4, é possível ver os resultados obtidos. Os pontos verdes representam locais que possuem boas avaliações, pontos vermelhos possuem avaliações ruins e pontos azuis avaliações neutras.

Avaliando a Figura 4, é possível ver que maioria dos estabelecimentos de Tiradentes se concentram no centro da cidade, enquanto há alguns em áreas periféricas da cidade. Por ser uma cidade pequena Tiradentes possui um centro comercial limitado, concentrado em apenas uma região. Isso fica evidente avaliando a localização dos restaurantes os quais se aglomeram em uma pequena região, enquanto há poucos em áreas diferentes. Assim, é possível observar que restaurantes bem qualificados e que possuem baixas qualificações dividem o mesmo espaço, mostrando assim que não há uma divisão e nem influência da localidade neste caso.

⁵<https://cloud.google.com/maps-platform/>

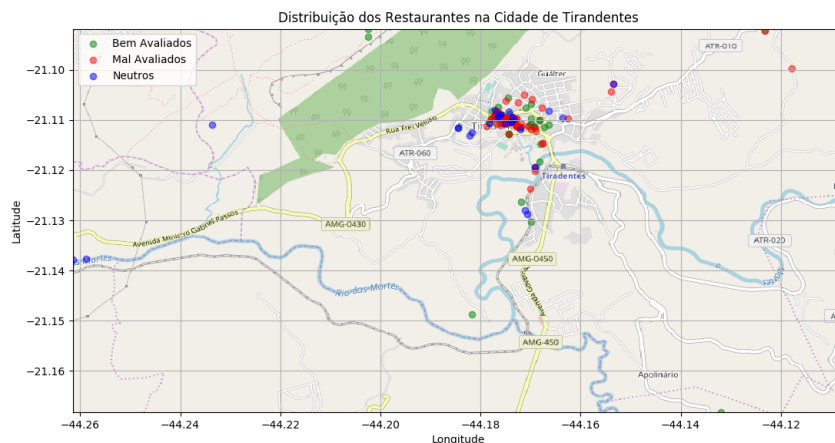


Figure 4. Mapa aproximado da cidade de Tiradentes localizando os restaurantes pelas avaliações

5. Conclusão

Este trabalho apresenta uma caracterização utilizando abordagens de Redes Complexas para os restaurantes da cidade de Tiradentes. Para que fosse feita essa análise foram coletados dados do site *TripAdvisor* que é atualmente o maior site de turismo mundial. Em adição a caracterização foram aplicados algoritmos de detecção de comunidades com sobreposição e feita uma análise por meio de geolocalização dos restaurantes, avaliando possíveis regiões as quais possuem uma maior quantidade de restaurantes bem avaliados.

Por meio deste estudo foi possível ter uma ideia do comportamento dos clientes da cidade de Tiradentes, a qual é uma cidade que possui sua economia basicamente associada ao turismo.

Através de nossas análises percebemos que há *clusters* de Restaurantes os quais possuem alto grau de *reviews* entre si, podendo revelar um viés para a realização destas avaliações. Além disso, podemos perceber que por ser uma cidade interiorana, Tiradentes não possui uma correlação entre localização e as avaliações que um restaurante recebe, dado que grande parte desses se localizam na região central da cidade.

Como trabalhos futuros deseja-se realizara avaliação da cidade utilizando todas as bases de dados coletadas, de atrações, hotéis e aluguéis de temporada. Além disso pretende-se realizar avaliações de maneira a caracterizar como estes dados podem auxiliar algoritmos clássicos de recomendação. Também avalia-se a utilização de métodos de *Aspect Level Sentiment Analysis* e a modelagem temporal dos grafos para avaliação de temporadas em que há um maior grau de visitas.

References

- Dataviva tiradentes. <http://dataviva.info/pt/location/4mg000214>. Accessed: 2019-04-22.
- Albert, R., Jeong, H., and Barabási, A.-L. (1999). Internet: Diameter of the world-wide web. *nature*, 401(6749):130.

- Araújo, M., Gonçalves, P., Cha, M., and Benevenuto, F. (2014). ifeel: a system that compares and combines sentiment analysis methods. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 75–78. ACM.
- Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. (2009). Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, pages 49–62. ACM.
- Corney, D., Martin, C., and Göker, A. (2014). Spot the ball: Detecting sports events on twitter. In *European Conference on Information Retrieval*, pages 449–454. Springer.
- Felix, L., Caetano, J. V., R., C., Luiz, W., Dias, D., and Rocha, L. D. (2018). Avaliação automática de conteúdo de aplicações de reclamação online. *The Symposium on Knowledge Discovery, Mining and Learning (KDMiLe)*, KDMiLe.
- Gionis, A., Lappas, T., Pelechrinis, K., and Terzi, E. (2014). Customized tour recommendations in urban areas. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 313–322. ACM.
- Guo, T., Guo, B., Ouyang, Y., Yu, Z., Lam, J. C. K., and Li, V. O. K. (2018). Crowdtravel: scenic spot profiling by using heterogeneous crowdsourced data. *Journal of Ambient Intelligence and Humanized Computing*, 9(6):2051–2060.
- Hao, Q., Cai, R., Wang, C., Xiao, R., Yang, J.-M., Pang, Y., and Zhang, L. (2010). Equip tourists with knowledge mined from travelogues. In *Proceedings of the 19th international conference on World wide web*, pages 401–410. ACM.
- Hobbs, W. R. and Burke, M. K. (2017). Connective recovery in social networks after the death of a friend. *Nature Human Behaviour*, 1(5):0092.
- Khatibi, A., Belem, F., Silva, A. P., Shasha, D., Goncalves, M. A., et al. (2018). Improving tourism prediction models using climate and social media data: A fine-grained approach. In *Twelfth International AAAI Conference on Web and Social Media*.
- Kim, J., Kim, H., and Ryu, J.-h. (2009). Triptip: a trip planning service with tag-based recommendation. In *CHI’09 Extended Abstracts on Human Factors in Computing Systems*, pages 3467–3472. ACM.
- Kotiloglu, S., Lappas, T., Pelechrinis, K., and Repoussis, P. (2017). Personalized multi-period tour recommendations. *Tourism Management*, 62:76–88.
- Lancichinetti, A., Fortunato, S., and Kertesz, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015.
- Leite, A. H., Benevenuto, F., and Moro, M. M. (2013). Triptag: Ferramenta de planejamento de viagens baseada em experiências de usuários de redes sociais. In *28 TH BRAZILIAN SYMPOSIUM ON DATABASES*, page 37.
- Lu, E. H.-C., Chen, C.-Y., and Tseng, V. S. (2012). Personalized trip recommendation with multiple constraints by mining user check-in behaviors. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pages 209–218. ACM.

- Savage, N. S., Baranski, M., Chavez, N. E., and Höllerer, T. (2012). I'm feeling loco: A location based context aware recommendation system. In *Advances in Location-Based Services*, pages 37–54. Springer.
- Souza, R., Neill, D. B., Assuncao, R. M., Meira Jr, W., et al. (2019). Identifying high-risk areas for dengue infection using mobility patterns on twitter. *Online Journal of Public Health Informatics*, 11(1).
- Xie, J., Kelley, S., and Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)*, 45(4):43.