

Tourism Recommendation System using complex network approaches

Antônio P. S. Alves², Lucas G. S. Félix³, Carlos Magno G. Barbosa¹, Vinícius da Fonseca Vieira¹, Carolina Ribeiro Xavier¹

¹ Programa de Pós Graduação em Ciência da Computação
Universidade Federal de São João del Rei (UFSJ)
São João del-Rei - MG - Brasil
² Pontifícia universidade Católica
Rio de Janeiro - MG - Brasil
³ Programa de Pós Graduação em Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte - MG - Brasil
carolinaxavier@ufs.j.edu.br

Abstract. The amount of available data on the web has grown exponentially, mostly due to the emergence of the Collaborative Internet, in mid-2006, which turns the process of obtaining information into a hard task. This way, several computational techniques have been used in order to automate the exploitation and analysis of data, such as Text Mining techniques, Topic Modeling (TM), which establishes relationships between text documents and discussion topics through the present words, and Sentiment Analysis (SA), whose objective is to identify sentences' polarity; Complex Networks modeling, which seek to capture the dynamics of complex systems, present in social networks; and Recommendation Systems, which assist with decision-making and whose operation resides in the suggestion of items that have not yet been evaluated by a user, such as traveling to a new place or trying another meal from a menu. The Tourism scenario is also included in the context of massive data generation and advances in techniques to deal with them. In this case, specialized travel platforms, like Tripadvisor, have a major role since they concentrate a large amount of data about users and their experience in Points-of-Interest (POI). Therefore, this work proposes a new approach to a predictive model for POI recommendation systems based on the construction of a Complex Network and the use of specific techniques for its structural analysis. The city chosen to validate these objectives was the city of Tiradentes, Minas Gerais, whose geographic proximity and tourism-oriented economy make it a good choice. The results obtained show (*i*) that a predictive model based on Complex Networks does not overcome the error obtained by *baselines* algorithms, however, it brings a good ranking correlation between what was predicted and the real result, which makes it a good option for recommendation systems.

Categories and Subject Descriptors: H.2.8 [Database Management]: Database Applications; I.2.6 [Artificial Intelligence]: Learning

Keywords: Complex Networks, Graph mining, Recommendation Systems, Tourism.

1. INTRODUÇÃO

Por muitos anos, planejar uma viagem era sinônimo de comprar mapas e revistas especializadas sobre o destino, além de passar horas decidindo qual trajeto realizar, quais atrações visitar e onde se hospedar. Esta situação mudou abruptamente com o surgimento da Internet Colaborativa, em meados de 2006, cuja geração de conteúdo por usuários e o compartilhamento de informações se tornou regra em ambientes virtuais. Assim, a migração de clientes e empresas para a web, além do aumento no volume de dados, tornaram a busca por informações uma tarefa desafiadora para todos os interessados em dados web. Não poderia diferir para os turistas, que possuem a sua disposição milhares de sites e blogs especializados em turismo, bem como redes sociais voltadas para este setor com milhões de opiniões a respeito de estabelecimentos, feitas por viajantes de todo o mundo. Alguns exemplos são

*Foursquare*¹, *Yelp*² e *Tripadvisor*³. O *Tripadvisor* sozinho conta com aproximadamente 867 milhões de opiniões e comentários (ou *reviews*) sobre 8.7 milhões de lugares espalhados pelo globo e cadastrados na plataforma, tornando-a a maior plataforma sobre viagem do mundo [Tripadvisor 2020].

A partir da vasta disponibilidade de dados gerados por usuários em ambientes virtuais, a indústria sugere o uso de *Sistemas de Recomendação (SR)* para analisar e criar associações pertinentes e únicas, que auxiliam na tomada das melhores decisões [Pantano et al. 2019]. Considerando o contexto geral, os SRs atacam o problema da sobrecarga de informação por meio de sugestões personalizadas de itens mais adequados para usuários específicos. Em suma, estes algoritmos têm por objetivo prever o interesse do usuário consumidor e sugerir (recomendar) itens que melhor se adequam a este interesse [Lu et al. 2015]. Entre as diversas abordagens presentes na literatura de SRs, se destacam as abordagens baseadas em Filtragem Colaborativa (FC), as quais utilizam interações passadas entre usuários e itens para realizar previsões sobre itens similares que podem ser consumidos [Schafer et al. 2007].

Algoritmos que pertencem à classe de FC se dividem em duas partes: algoritmos baseados em memória (vizinhança) e algoritmos baseados em modelo (fatores latentes) [Koren 2010]. As técnicas baseadas em vizinhança (e.g. k-Nearest Neighbors - kNN), focam na recomendação baseada em itens populares entre os vizinhos de um usuário [Koren 2010]. Já as técnicas baseadas em fatores latentes (e.g. Singular Value Decomposition - SVD, Non-Negative Matrix Factorization - NMF), transforma itens e usuários em um mesmo espaço latente, fazendo com que itens e usuários sejam diretamente comparáveis [Ricci et al. 2011]. Enquanto os algoritmos baseados em modelos conseguem capturar ligações intrínsecas entre os itens recomendados, mostrando, em geral, melhores resultados quando comparados com técnicas baseadas em vizinhança, as técnicas baseadas em vizinhança mantêm como principal vantagem a interpretabilidade dos resultados.

Uma vez que o comportamento de usuários pode ser influenciado por conteúdos direcionados [Miguéns et al. 2008], tem-se que a recomendação voltada para o turismo busca, dentre várias possibilidades, sugerir destinos de viagens [Zheng et al. 2018], auxiliar na construção de roteiros turísticos [Cenamora et al. 2017] e sugerir visitas em locais específicos a partir de preferências definidas em redes sociais diversas, como *Facebook*, *Twitter*, *Gowalla*, *Foursquare*, *Yelp* e *Tripadvisor* [Missouli et al. 2018]. Esta última possibilidade de recomendação, é muitas vezes referida como *recomendação de pontos de interesse (Points-of-Interest - POI)*, que consiste na identificação de lojas, atrações turísticas, restaurantes ou parques que potencialmente maximizam a preferência de um usuário [Zhao et al. 2020].

Sistemas que recomendam POIs se diferem de outros tipos de SRs, dado que são extremamente voláteis a informações espaço-temporais, como clima e restrições de horários desses locais [Qian et al. 2019]. A volatilidade e variedade do conteúdo dos dados para recomendadores de POI permitem que diferentes abordagens sejam estabelecidas, de modo a explorar nuances e características únicas dos dados e que beneficiam a recomendação em si. Nesse sentido, é importante criar um recomendador eficiente, que consiga capturar a essência dos usuários para sugerir opções atrativas aos mesmos. Além disso, atualmente, o interesse dos usuários em entender suas recomendações tem aumentado, algo que vem sendo estudado dentro de campos da interpretabilidade de modelos de aprendizado máquina [Carvalho et al. 2019]. Assim, ainda que abordagens mais complexas, especialmente aquelas baseadas em modelo, possuam melhores resultados com relação a métricas que avaliam SRs, abordagens mais simples, como modelos baseados FC, possuem a explicabilidade como grande vantagem.

Pensando nisso, o objetivo deste trabalho é definir uma metodologia de recomendação de POIs que consiga minimizar as métricas de erro, e maximizar a interpretabilidade dos modelos. Para isso, propomos uma abordagem que se utiliza de técnicas de redes complexas para tal.

¹<https://pt.foursquare.com/>

²<https://www.yelp.com/>

³<https://www.tripadvisor.com.br/>

Uma rede complexa é definida em termos da estrutura de um grafo $G = (N, E)$, cujo conjunto N é o conjunto de vértices, ou nós (redes complexas), e o conjunto E é o conjunto de arestas, ou links que conectam dois nós [Zou et al. 2019]. Tal estrutura, pode ser diretamente comparada as abordagens feitas por técnicas de FC baseadas em vizinhança, dado que as mesmas utilizam-se das ligações entre itens e usuários para realizar inferências sobre os gostos dos usuários. Entretanto, a utilização de uma abordagem de redes complexas, distancia a abordagem proposta de técnicas como o kNN ao acrescentar métodos de detecção de comunidades e *Link Prediction*, permitindo assim melhorar os resultados de técnicas baseadas em FC sem perder o fator da interpretabilidade.

Assim, no presente trabalho propomos uma nova abordagem para predição de avaliações, necessária para o desenvolvimento de um sistema de recomendação, baseada na modelagem de uma Rede Complexa e uso de técnicas de *Link Prediction* e Detecção de Comunidades. Pode-se, assim, definir nossa principal pergunta de pesquisa: Quão eficiente é um modelo preditivo de avaliações turísticas baseado em comunidades e *Link Prediction* para um sistema de recomendação de POI? Por se tratar de uma nova abordagem, a comparação com *baselines* é imprescindível para a validação deste novo método. Assim, neste trabalho comparamos nossa proposta com técnicas de FC baseada em vizinhança e em modelo. Para avaliar nossa proposta, coletamos da plataforma Tripadvisor avaliações de restaurantes da cidade de Tiradentes - Minas Gerais [Alves et al. 2022]. A cidade foi escolhida por ter sua economia voltada ao turismo, e ser considerada por um estudo da Universidade de Oregon e o instituto *The Creative Urban* uma das 10 cidades "mais legais" a serem visitadas no mundo [LatestLY 2020].

2. MATERIAIS E MÉTODOS

2.1 Dados

No Tripadvisor, Tiradentes possui 4 categorias de lazer e acomodação, são elas as **atrações**, **hotéis**, **aluguel de temporada** e **restaurantes**. Nesse sentido, usuários do Tripadvisor só podem avaliar estabelecimentos ou locais que estejam associados a uma dessas categorias. Todos os dados associados a Tiradentes foram coletados, usando um coletor de implementação própria, e os dados dos restaurantes serão utilizados neste trabalho.

2.2 Métodos

2.2.1 Modelagem e Construção da Rede. Os dados dos *reviews* disponíveis sobre Tiradentes no Tripadvisor consistem de alguns atributos, entre eles o **perfil do usuário**, **nota** e **lugar visitado**. A relação entre esses três dados permite modelar a estrutura fundamental de qualquer SRs de FC, a **matriz de utilidade**.

A matriz de utilidade será responsável por estabelecer arestas ponderadas entre os usuários e, para tanto, duas possibilidades de matrizes são possíveis, tanto a de *usuário - usuário* quanto a de *usuário - item*. Pela interpretabilidade e ampla utilização na literatura [Ahuja et al. 2019; Dascalu et al. 2016; Chen and Lee 2018], a matriz de *usuário - item* foi escolhida, implicando na seguinte composição: os usuários que avaliaram restaurantes em Tiradentes pelo Tripadvisor são os nós da rede e as arestas são definidas segundo a similaridade entre dois usuários a partir das notas dadas para os diversos restaurantes que ambos visitaram, isto é, a partir das linhas que representam cada usuário na matriz de utilidade, modela-se um vetor de notas sobre restaurantes que ambos avaliaram e, então, extrai-se a similaridade deste vetor. Essas medidas estão disponíveis na biblioteca *Surprise*⁴, a escolha, entretanto, de qual medida utilizar demanda algumas considerações.

No contexto de POI, usuários visitarem um mesmo lugar pode soar intuitivo, mas não significa que eles são similares, posto que eles podem ter tido experiências totalmente diferentes, apesar de terem

⁴<https://surprise.readthedocs.io/en/stable/index.html>

visitado um mesmo lugar. Apesar de amplamente utilizada, a Similaridade de Cosseno, no contexto de avaliação sobre POI, não captura a distinção qualitativa existente entre notas no contexto do Tripadvisor, que possui uma escala de 1 a 5 para notas dadas em quaisquer das categorias disponíveis. Brevemente, se o usuário u avaliou 3 lugares com nota máxima, isto é, 5, e o usuário v avaliou os mesmos 3 lugares com nota mínima (1), temos que a similaridade entre eles seria máxima, pois os vetores possuem a mesma orientação. Essa limitação não é da medida em si, mas do contexto da base de dados, em que se busca que as notas dadas por usuários seja utilizada como medida de similaridade, isto é, quanto mais próximas às notas dadas, mais similares são os usuários. Dessa forma, foi escolhido o **Mean Squared Difference (MSD)**, que apesar de não balancear o total de avaliações, considera a ideia mais importante nesta nova abordagem, a diferença absoluta das avaliações em comum feitas pelos usuários.

A Figura 1 mostra as etapas para modelagem da rede complexa, que será utilizada na predição de notas considerando um caso de três usuários que avaliaram pontos turísticos em cinco grandes cidades do mundo cadastradas no Tripadvisor, as notas variam de 1 a 5.

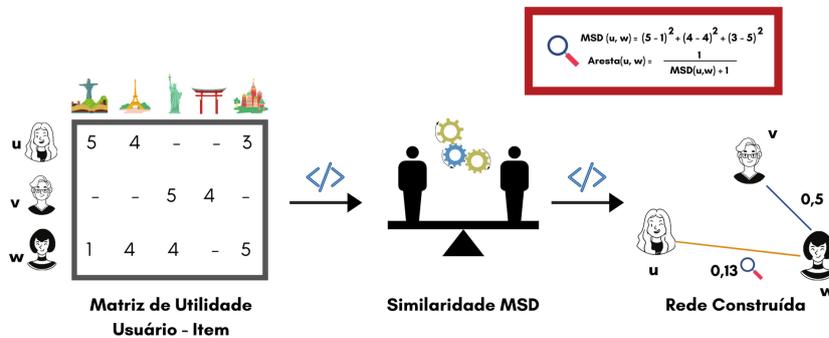


Fig. 1: Etapas para a construção da rede.

Um ponto importante sobre a rede construída é o seu tamanho, na categoria de restaurantes existem **29541 reviews** de **16478** usuários distintos, o que resultaria em uma rede com **16478** nós. Para reduzir o tamanho da rede escolheu-se por focar em usuários mais ativos, e somente os usuários que fizeram *reviews* de, no mínimo, 7 restaurantes foram mantidos. A escolha por este valor representa a média do número de restaurantes visitados por cada usuário nesta base, o que reduziu a rede a **261** nós.

Para criação das ligações entre os usuários foi estabelecido um *threshold* mínimo de 3 lugares, isto é, se ambos visitaram até 2 lugares somente, a similaridade entre eles é considerada **zero**, 3 lugares ou mais, o MSD é calculado normalmente e uma rede densa é gerada.

Para reduzir a densidade de arestas, foram testados diferentes *thresholds* nos pesos das arestas: (i) rede completa; (ii) somente as arestas com valor superior a 0.2 são mantidas (13897 arestas); ou somente as arestas com valor superior a 0.4 são mantidas (9843 arestas), os valores são empíricos sendo estabelecidos a partir da observação da redução do número de arestas. Observou-se que uso de valores superiores a 0.4 inviabiliza as análises pela drástica redução do tamanho da rede.

2.2.2 Sistemas de recomendação baseado em redes complexas. Com os dados dos usuários modelados como uma rede, é possível explorar a estrutura utilizando as técnicas de *link prediction* e de detecção de comunidades para preencher as arestas faltantes na rede, ou seja, as posições faltantes na matriz de utilidade, que define a recomendação.

Pretende-se aplicar técnicas de *link prediction* e informações de comunidades para realizar a recomendação. Para cada posição faltante da matriz, ou melhor, para cada restaurante ainda não avaliado por um usuário, estima-se uma avaliação. A grande questão para problemas de predição e modelos de aprendizado de máquina no geral, é: **como saber se o que foi predito estava correto?** Através do histórico de avaliações. No contexto de POI, usuários possuem um histórico de quais lugares já foram visitados e como foi sua experiência nestes locais.

As técnicas de *link prediction* os algoritmos *Coeficiente de Jaccard*, *Adamic-Adar Index*, *Preferential Attachment* e *Resource Allocation Index* para prever arestas faltantes. Toda técnica de *link prediction* retorna um valor, um número que reflete a probabilidade de uma nova relação entre dois vértices existir. Sem nenhum tipo de poda, tem-se, após o *link prediction*, uma rede completa com diferentes valor de probabilidade. De modo a evitar uma rede tão densa, as arestas preditas só são reinseridas se estiverem acima de um *threshold* mínimo. Na metodologia, três valores empíricos são utilizados como *threshold*, são eles **0.3**, **0.4** e **0.5** - valores inferiores a *0.3* prejudicaram as análises, enquanto superiores a *0.5* diminuem muito a contribuição e efetividade do *link prediction* para a abordagem proposta. Vale ressaltar que, o intervalo de valores de saída para cada algoritmo varia, logo, para unificar a adoção de um *threshold* mínimo para reinserção, as saídas dos algoritmos foram normalizadas para o intervalo de 0 a 1 através da aplicação *min-max*.

As comunidades são estruturas importantes nas redes, uma vez que se revelam grupos de nós densamente conectados entre si e com poucas conexões com nós de outros grupos [Cherifi et al. 2019]. Em redes sociais é perceptível a formação de grupos, sobretudo pelo modo como os membros conversam e se relacionam, mas da perspectiva matemática, essa afinidade não é mensurável explicitamente e sim por propriedades de estruturais [Javed et al. 2018], que podem ser definidas por ferramentas e técnicas distintas da Física, Biologia, Matemática, Sociologia e Ciência da Computação [Lancichinetti and Fortunato 2009].

Nesse sentido, nesta nova abordagem de predição utilizam-se os principais algoritmos de detecção de comunidades para identificar quais usuários possuem um comportamento de visita similar e tiveram experiências parecidas nos restaurantes explorados. Para a detecção de comunidades foram experimentados os algoritmos: *Fastgreedy*[Clauset et al. 2004], *Blondel*[Blondel et al. 2008] e *Leading Eigenvector*[Newman 2006], baseados em maximização de modularidade, para identificar os usuários similares do ponto de vista das avaliações publicadas. Para utilizar as comunidades obtidas para guiar o preenchimento da matriz de utilidade, isto é, um usuário sem nota para um determinado restaurante, o avaliará a partir da média das notas que os membros da comunidade da qual faz parte. Ressalta-se que como o objetivo final é preencher totalmente a matriz de utilidade, tem-se que, caso a comunidade não consiga avaliar um restaurante pela média das notas dadas a este - porque não tem nenhum membro que visitou aquele restaurante - utiliza-se a média de notas gerais dos membros (ou nota média de todas as avaliações da comunidade), para este restaurante não avaliado.

Para escolher os algoritmos, esses foram submetidos a comparações para se obter uma configuração que tenha boa desempenho na minimização do erro e apresente bons resultados em termos de modularidade. A utilização do **Algoritmo de Blondel**, em conjunto com o **Coeficiente de Jaccard para o link prediction** com **threshold** de reinserção de **0.5**, sob uma rede com **peso mínimo da aresta igual a 0.4** é a melhor combinação para comparar com os *baselines*. Esta combinação, especificamente, possui uma **acurácia de 85.10%** e uma precisão de 40.71%, o que também a qualifica como escolha.

Os principais algoritmos voltados para FC foram executados e comparados utilizando o erro médio quadrado (RMSE) e o erro médio absoluto (MAE) e o coeficiente de Spearman. Os algoritmos foram: **KNN Básico**, **KNN com Média**, **KNN com Z-Score**, **SVD**, **NMF** e **SVD++**.

3. RESULTADOS E DISCUSSÕES

A Figura 2(a) mostra como a abordagem proposta neste capítulo se porta frente aos *baselines* considerando o RMSE, enquanto a Figura 2(b) compara segundo o MAE. Percebe-se que, para ambas métricas, a metodologia proposta apresenta um erro superior aos *baselines*, no entanto ficou abaixo de um (1 estrela). O algoritmo que se destacou foi o KNN Básico, que obteve os menores RMSE e MAE médios.

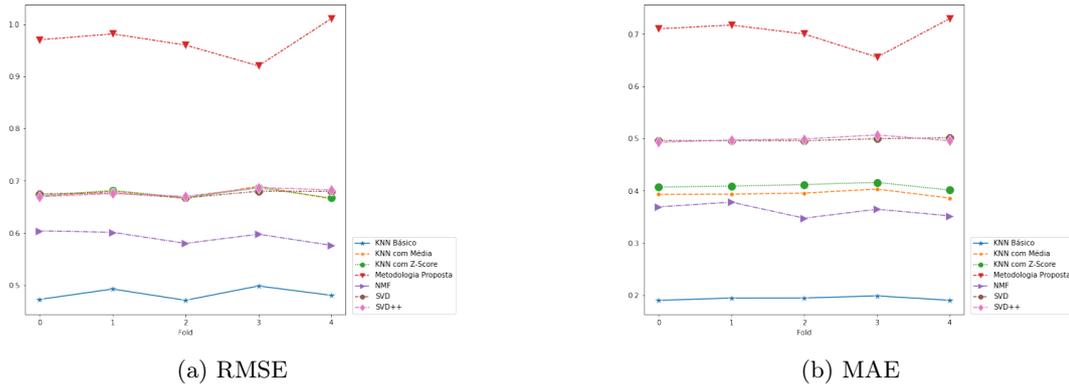
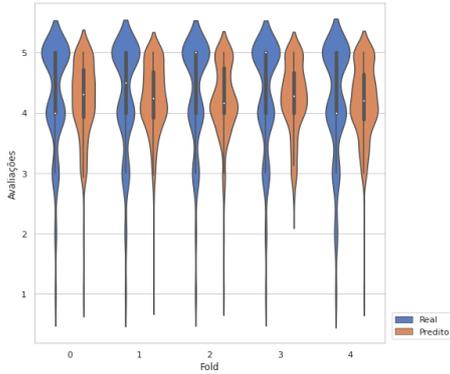


Fig. 2: Comparativo dos erros obtidos para a metodologia proposta e para os *baselines*.

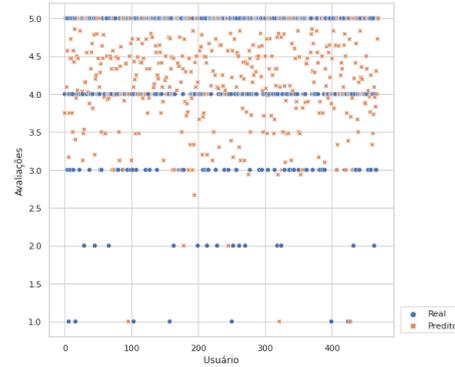
O resultado inferior aos algoritmos da literatura pode ter sido influenciado pelo o cenário proposto. Existem 163 restaurantes em Tiradentes no Tripadvisor, mas há uma concentração de avaliações com notas 4 e 5 para estes estabelecimentos, sabe-se que um modelo preditivo *naive* que assuma sempre notas médias, neste caso entre 4 ou 5, poderia obter resultados tão bons quanto as técnicas comparadas, levando a crer que a utilização de uma base com maior distribuições de notas em suas avaliações, apresente maiores diferenças, exigindo mais do desempenho dos *baselines* e da metodologia proposta. A Figura 3a mostra a diferença da distribuição das notas ao longo dos *fold*s entre o que foi predito e o que era a nota real para a rede utilizada na configuração escolhida.

Percebe-se que as avaliações originais possuem a concentração nas notas 4 e 5, a metodologia proposta captura uma concentração similar nas mesmas notas ao longo dos *fold*s, sob a diferença que as avaliações preditas tratam valores no domínio dos reais e assumem muitas avaliações neste intervalo, enquanto as avaliações reais tratam de valores inteiros. A estratégia proposta se distanciou mais da predição de valores 4 e 5. Essa distinção pode ser observada em detalhes no *Fold 0* na Figura 3b, em que cada avaliação predita e real são apresentadas como pontos no espaço. Observa-se que a distribuição da maioria das predições está no meio do intervalo de 4 a 5, assegurando o resultado positivo de que a metodologia proposta não supera um erro maior que 1, que é o erro mínimo possível no Tripadvisor (1 estrela), mas, em contrapartida, piora o desempenho frente aos *baselines*.

Outra medida de avaliação utilizada foi o Coeficiente de Correlação de Spearman, que define a correlação entre duas variáveis (notas dadas x notas preditas, por exemplo)[Zheng et al. 2017]. Ao trocar as variáveis pelo conjunto de avaliações preditas e avaliações originais, o coeficiente de Spearman ordenará esses vetores e permitirá avaliar se o que está sendo predito está relacionado com o real. Considerando que a recomendação é a próxima etapa do modelo preditivo proposto e dos *baselines* comparados, obter valores altos para a mesma é um bom indicativo sobre a qualidade da recomendação a ser proposta. A Figura 4 apresenta o desempenho das estratégias sob a ótica desta correlação e mostra que a abordagem proposta é a melhor nesse quesito. Com destaque, novamente, para o KNN Básico, tem-se que a estratégia proposta possui resultados mais consistentes que os demais, tornando seu uso interessante nesta perspectiva.



(a) Comparativo entre as notas estimadas e as notas reais ao longo dos *fold*s.



(b) Comparativo entre as notas estimadas e as notas reais no *fold* 0.

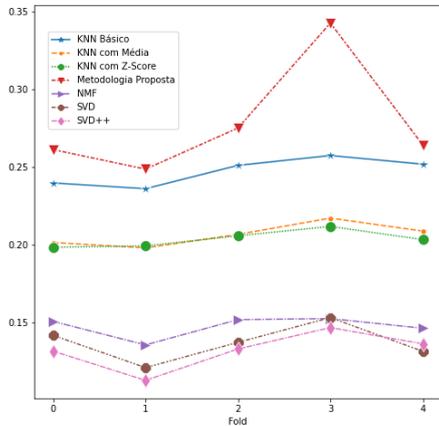


Fig. 4: Comparativo do coeficiente de correlação do Spearman obtido entre a metodologia proposta e os *baselines*.

Algoritmo	Spearman	RMSE	MAE
Abordagem Proposta	0.2781	0.9686	0.7024
KNN Básico	0.2469	0.4832	0.1937
KNN com Média	0.2061	0.6743	0.3942
KNN com Z-Score	0.2034	0.6751	0.4089
NMF	0.1471	0.5919	0.3622
SVD	0.1366	0.6757	0.4977
SVD++	0.1318	0.6771	0.4982

Table I: Comparativo das Métricas de Erro e coeficiente de Spearman entre todas as estratégias.

A Tabela I resume o exposto de todas as estratégias, frente às métricas de erro e ao coeficiente de Spearman, consolidando os dados dos *fold*s por uma simples média.

4. CONCLUSÕES

O presente trabalho apresenta uma abordagem baseada em Redes Complexas, utilizando técnicas de *Link Prediction* e Detecção de Comunidades. Para tanto, a partir da modelagem de uma rede cuja os nós são usuários do Tripadvisor e as arestas são relações de similaridade sobre o que foi visitado em comum entre dois nós, estabelece-se o uso do Coeficiente de Jaccard para predição de links, com o valor da "força" do link predito devendo ser superior à 0.5, e posteriormente o uso do Algoritmo de Blondel como estratégia de detecção de comunidades. Com a formação destas, para cada avaliação faltante de um usuário sobre um restaurante, utiliza-se a nota média dos membros da comunidade deste usuário para este restaurante. Para avaliar o resultado da predição obtida, comparou-se a abordagem proposta com o desempenho de algoritmos linha de base de Filtragem Colaborativa utilizando para validação o processo de *K-Fold Cross-Validation*, com $k = 5$. Os resultados obtidos em termos de métricas de erro não permitem afirmar que a estratégia proposta é melhor ou pior que os *baselines*, haja vista que todas estiveram abaixo do erro médio de 1 estrela, que é o erro mínimo no Tripadvisor. Contudo, a avaliação do coeficiente de correlação de Spearman para as predições realizadas indicam

que a abordagem apresentada se mostra eficiente para o contexto de um SRs, respondendo à pergunta de pesquisa.

REFERENCES

- AHUJA, R., SOLANKI, A., AND NAYYAR, A. Movie recommender system using k-means clustering and k-nearest neighbor. *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2019.
- ALVES, A., FÉLIX, L., BARBOSA, C., VIEIRA, V., AND XAVIER, C. Tiradentes no tripadvisor - o que se fala sobre essa simpática cidade histórica? In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*. SBC, Porto Alegre, RS, Brasil, pp. 145–156, 2022.
- BLONDEL, V., GUILLAUME, J.-L., LAMBIOTTE, R., AND LEFEBVRE, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* vol. 2008, pp. 10008, 2008.
- CARVALHO, D. V., PEREIRA, E. M., AND CARDOSO, J. S. Machine learning interpretability: A survey on methods and metrics. *Electronics* 8 (8): 832, 2019.
- CENAMOR, I., ROSA, T., NÚÑEZ, S., AND BORRAJO, D. Planning for tourism routes using social networks. *Expert Syst. Appl.* vol. 69, pp. 1–9, 2017.
- CHEN, T. T. AND LEE, M. Research paper recommender systems on big scholarly data. In *PKAW*, 2018.
- CHERIFI, H., PALLA, G., SZYMANSKI, B., AND LU, X. On community structure in complex networks: challenges and opportunities. *Applied Network Science* vol. 4, pp. 1–35, 2019.
- CLAUSET, A., NEWMAN, M., AND MOORE, C. Finding community structure in very large networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* vol. 70 6 Pt 2, pp. 066111, 2004.
- DASCALU, M., BODEA, C., MIHAILESCU, M. N., TANASE, E. A., AND PABLOS, P. O. D. Educational recommender systems and their application in lifelong learning. *Behaviour e Information Technology* vol. 35, pp. 290 – 297, 2016.
- JAVED, M. A., YOUNIS, M. S., LATIF, S., QADIR, J., AND BAIG, A. Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications* vol. 108, pp. 87–111, 2018.
- KOREN, Y. Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4 (1): 1–24, 2010.
- LANCICHINETTI, A. AND FORTUNATO, S. Community detection algorithms: a comparative analysis. *Physical review. E, Statistical, nonlinear, and soft matter physics* vol. 80 5 Pt 2, pp. 056117, 2009.
- LATESTLY. The 10 coolest small towns in the world — and why you should consider moving to one. <https://www.latestly.com/auto/the-10-coolest-small-towns-in-the-world-and-why-you-should-consider-moving-to-one-2003725.html>, 2020. Acessado em 04/05/2021.
- LU, J., WU, D., MAO, M., WANG, W., AND ZHANG, G. Recommender system application developments: a survey. *Decision Support Systems* vol. 74, pp. 12–32, 2015.
- MIGUÉNS, J., BAGGIO, R., AND COSTA, C. Social media and tourism destination: Tripadvisor case study, 2008.
- MISSAOUI, S., KASSEM, F., VIVIANI, M., AGOSTINI, A., FAIZ, R., AND PASI, G. Looker: a mobile, personalized recommender system in the tourism domain based on social media user-generated content. *Personal and Ubiquitous Computing* vol. 23, pp. 181–197, 2018.
- NEWMAN, M. Finding community structure in networks using the eigenvectors of matrices. *Physical review. E, Statistical, nonlinear, and soft matter physics* vol. 74 3 Pt 2, pp. 036104, 2006.
- PANTANO, E., PRIPORAS, C.-V., STYLOS, N., AND DENNIS, C. Facilitating tourists’ decision making through open data analyses: A novel recommender system. *Tourism Management Perspectives* vol. 31, pp. 323–331, 2019.
- QIAN, T., LIU, B., NGUYEN, Q. V. H., AND YIN, H. Spatiotemporal representation learning for translation-based poi recommendation. *ACM Trans. Inf. Syst.* 37 (2), Jan., 2019.
- RICCI, F., ROKACH, L., AND SHAPIRA, B. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, pp. 1–35, 2011.
- SCHAFFER, J. B., FRANKOWSKI, D., HERLOCKER, J., AND SEN, S. Collaborative filtering recommender systems. In *The adaptive web*. Springer, pp. 291–324, 2007.
- TRIPADVISOR. About tripadvisor. <https://tripadvisor.mediaroom.com/US-about-us>, 2020. Accessed: 2020-06-01.
- ZHAO, P., LUO, A., LIU, Y., ZHUANG, F., XU, J., LI, Z., SHENG, V. S., AND ZHOU, X. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2020.
- ZHENG, X., LUO, Y., SUN, L., ZHANG, J., AND CHEN, F. A tourism destination recommender system using users’ sentiment and temporal dynamics. *Journal of Intelligent Information Systems* vol. 51, pp. 557–578, 2018.
- ZHENG, X., XU, L., AND CHAI, S. Qos recommendation in cloud services. *IEEE Access* vol. 5, pp. 5171–5177, 2017.
- ZOU, Y., DONNER, R. V., MARWAN, N., DONGES, J. F., AND KURTHS, J. Complex network approaches to nonlinear time series analysis. *Physics Reports* vol. 787, pp. 1–97, 2019. Complex network approaches to nonlinear time series analysis.