

CRIAÇÃO DE UM FILTRO PARA MAPAS: LOCALIZAÇÃO DE ÁREAS DE INTERESSE BASEADA NA PROXIMIDADE DE SERVIÇOS

Yan Manenti Martins¹, Allan Farias Fávaro²

Resumo: O mercado imobiliário digital e os serviços de mapas carecem de filtros de busca que considerem a proximidade simultânea de múltiplos pontos de interesse, o que limita a capacidade do usuário de identificar locais que atendam a necessidades mais complexas. Este trabalho desenvolve um filtro de mapas voltado a localizar áreas de interesse com base na proximidade a um conjunto de serviços definidos pelo usuário. A metodologia adota a modelagem da consulta do usuário (por exemplo, “restaurante, escola, mercado”) como um grafo de consulta e da base de dados espacial como um grafo de dados, tratando o problema como uma correspondência de subgrafos. Um algoritmo é utilizado para encontrar, no grafo de dados, padrões que atendam aos requisitos espaciais e de categoria especificados. Os experimentos foram realizados com dados do Yelp e do OpenStreetMap, resultando na identificação de padrões e locais que atendem às combinações de proximidade previstas na consulta.

Palavras-chave: busca em grafos; correspondência de subgrafos; filtro de mapas; pontos de interesse; busca por proximidade.

ABSTRACT: The digital real estate market and map services lack search filters that consider the simultaneous proximity of multiple points of interest, which limits users’ ability to identify locations that meet more complex needs. This work develops a map-based filter designed to locate areas of interest based on proximity to a set of services defined by the user. The methodology models the user query (for example, “restaurant, school, market”) as a query graph and the spatial database as a data graph, treating the problem as a subgraph-matching task. An algorithm is used to identify patterns in the data graph that meet the specified spatial and category requirements. Experiments were conducted using Yelp and OpenStreetMap datasets, resulting in the identification of patterns and locations that correspond to the proximity combinations specified in the query.

¹Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), y.a.martins@unesc.net

²Orientador. Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense (Unesc), allan.favaro@unesc.net

Keywords: graph search; subgraph matching; map filter; points of interest; proximity search.

1 INTRODUÇÃO

Com a chegada da internet ao Brasil nos anos 1980 e o avanço das tecnologias de comunicação, organizações precisaram se adaptar para atingir seus objetivos de forma eficaz (Abreu, 2019). O acesso à internet no país cresceu de 67% para 70% entre 2017 e 2018, totalizando cerca de 126,9 milhões de pessoas conectadas (Lavado, 2019).

O mercado de imóveis residenciais também se expandiu, com 98.562 unidades lançadas em 2018, aumento de 3,1% em relação a 2017 (Conteúdo, 2019), e um crescimento de 18,6% nos lançamentos em comparação com 2024 (Construção, 2025). As vendas registraram elevação de 20,9% em relação a 2023, e a procura por imóveis via plataformas digitais aumentou significativamente: em 2016, 90% dos consumidores já utilizavam sites para essa finalidade (Abreu, 2019).

A internet tornou-se essencial no setor imobiliário, reduzindo os custos, otimizando a promoção de vendas, facilitando a captação de imóveis e diminuindo a necessidade de visitas presenciais, tornando o processo mais ágil e acessível (Correia; Reis, 2011). O uso das redes sociais também disparou; por exemplo, o Facebook registrava 2,9 bilhões de usuários ativos mensais em 2021, consolidando os dados como recurso estratégico para os negócios (Kraus et al., 2022).

Nesse contexto, grafos se destacam por sua maior expressividade em comparação com modelos tradicionais, sendo aplicáveis em redes sociais, análise de dados biológicos, sistemas de recomendação e detecção de plágio. Nesses casos, nós representam usuários ou atividades, e as arestas indicam suas relações. A popularidade dos grafos impulsionou a busca eficiente em grandes grafos, tema de interesse crescente na indústria e na academia (Ma et al., 2016).

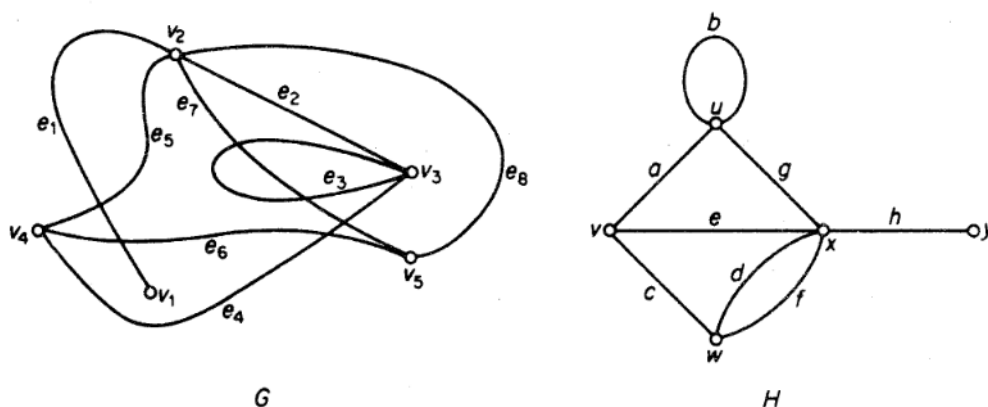
2 FUNDAMENTAÇÃO TEÓRICA

Com o objetivo de alinhar a pesquisa imobiliária às necessidades do usuário, este trabalho desenvolve um filtro capaz de correlacionar imóveis e pontos de interesse com precisão. A complexidade do cálculo de distâncias na malha viária exige, inicialmente, uma abordagem algorítmica baseada em Grafos. Uma vez estabelecida essa lógica de conectividade, ela é aplicada ao contexto do mundo real por meio de um SIG.

2.1 GRAFOS

Um grafo é formado por um conjunto de vértices (ou pontos) e um conjunto de arestas (ou ligações) entre esses vértices. Cada aresta conecta dois vértices, sem se importar com a ordem. A forma mais comum de representar um grafo é por meio de um desenho, onde os vértices são mostrados como pontos e as arestas como linhas ligando esses pontos. Muitas vezes, esse próprio desenho é chamado de grafo (Zhang; Chartrand, 2006). Exemplo de dois grafos na Figura 1.

Figura 1 - Diagrama de dois grafos G e H.



Fonte: Bondy, Murty et al. (1976).

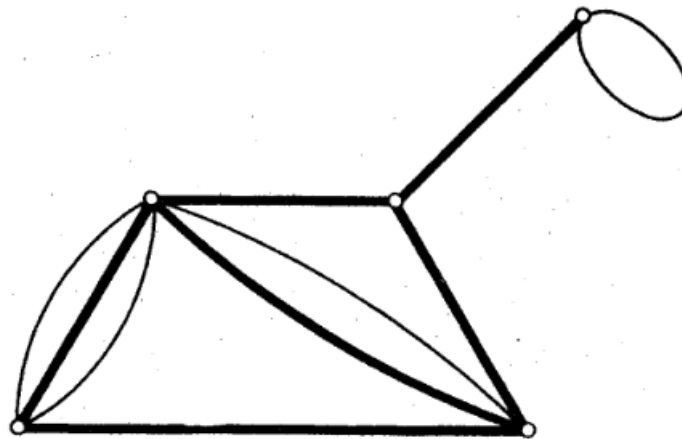
Em um grafo, dois vértices conectados por uma mesma aresta são considerados vizinhos, assim como duas arestas que compartilham um mesmo vértice. Uma aresta cujos extremos são o mesmo vértice é chamada de *loop*, enquanto uma aresta que une vértices diferentes é chamada de *ligação* (Bondy; Murty et al., 1976).

Um grafo é considerado subgrafo de outro quando todos os seus vértices e arestas fazem parte do grafo maior, e essas ligações conectam os mesmos pontos nos dois grafos. Um subgrafo pode ser interpretado como estando contido em (ou fazendo parte de) outro grafo (Deo, 2016). Exemplo de um subgrafo na Figura 2.

Em um grafo, cada conexão (ou ligação) associa-se um número real, chamado de peso. Um grafo com esse tipo de valor atribuído às conexões é conhecido como grafo ponderado. Grafos ponderados são comuns em diversas aplicações da teoria dos grafos (Bondy; Murty et al., 1976).

Por exemplo, em um grafo que representa amizades, os pesos podem indicar o nível de proximidade entre as pessoas. Já em um grafo de comunicações, esses valores podem representar os custos de construção ou manutenção das ligações entre os pontos (Bondy; Murty et al., 1976).

Figura 2 - Grafo e um de seus possíveis subgrafos.



Fonte: Bondy, Murty et al. (1976).

Um algoritmo é uma sequência de instruções precisas que, seguidas passo a passo, levam à solução de um problema em um número finito de etapas. Ele funciona como uma receita, com cada etapa claramente definida e sem ambiguidades (Deo, 2016).

Na teoria dos grafos, os algoritmos são essenciais para lidar com grafos de grande escala, que não podem ser analisados manualmente. O avanço dos computadores de alta velocidade permitiu a aplicação desses algoritmos de forma mais eficiente. Isso impulsionou o interesse pela teoria dos grafos, tornando possível resolver problemas antes considerados complexos demais (Deo, 2016).

O problema conhecido como isomorfismo de subgrafos consiste em identificar se um determinado grafo menor está contido dentro de um grafo maior, preservando a estrutura das conexões entre os seus elementos. Trata-se de um problema desafiador do ponto de vista computacional, sendo classificado como NP-completo, o que significa que não há, até o momento, um algoritmo eficiente que resolva todos os casos rapidamente. Além disso, a versão do problema que exige contar todas as possíveis correspondências é ainda mais complexa (Cibej; Mihelič, 2014).

A ideia central do algoritmo de Ullmann é simples: tentar, passo a passo, associar os elementos do grafo menor (o padrão) aos elementos do grafo maior (o alvo), verificando durante esse processo se as conexões entre os elementos estão sendo preservadas corretamente (Cibej; Mihelič, 2014).

2.2 SISTEMA DE INFORMAÇÃO GEOGRÁFICA (SIG)

Um sistema de informação geográfica (SIG) é uma tecnologia computacional voltada para coletar, armazenar, consultar, analisar e exibir dados geoespaciais. Esses dados representam tanto a localização quanto as características de elementos no espaço. Por exemplo, ao descrever uma estrada, é possível identificar onde ela está situada e também suas informações associadas, como o nome, extensão, limite de velocidade e sentido de tráfego (Chang, 2019).

O que diferencia um SIG de outros sistemas de informação é justamente sua capacidade de lidar com dados espaciais, permitindo integrar essas informações com outros tipos de dados para gerar análises mais completas e úteis (Chang, 2019).

Em um SIG, os dados representam objetos do mundo real, como estradas ou montanhas, incluindo sua localização e atributos. Cada elemento é associado a um objeto espacial. Devido às limitações dos sistemas, nem todos os detalhes podem ser armazenados, e cabe ao desenvolvedor decidir quais características e níveis de detalhe são mais relevantes.

Dados vetoriais representam objetos geográficos por meio de coordenadas precisas, como pontos, linhas e polígonos. Diferente dos dados raster, eles permitem detalhar formas e localizações com mais exatidão. Além das posições *x* e *y*, o valor *z* pode indicar propriedades como elevação. Esses dados podem ser armazenados de forma simples, sem conexões, ou com topologia, que define relações entre objetos e facilita análises. As informações vetoriais são divididas entre dados espaciais e atributos, geralmente gerenciados por bancos de dados relacionais (Bolstad, 2012).

Sistemas de Informação Geográfica geralmente usam bancos de dados espaciais para organizar e acessar informações geográficas. O modelo mais comum é o banco de dados relacional, que armazena apenas os dados necessários e facilita o acesso eficiente, embora possa ser complexo quando envolve várias tabelas interligadas. Outra abordagem comum são os bancos orientados a objetos, que organizam dados em hierarquias, como pessoas dentro de casas em áreas administrativas. Uma boa estruturação do banco é essencial para acelerar análises espaciais, especialmente com grandes volumes de dados. Normalmente, os atributos dos dados vetoriais ficam em bancos relacionais, enquanto os dados *raster* são tratados como objetos independentes ou associados a tabelas quando têm valores inteiros (Bolstad, 2012).

Como a Terra possui uma forma esférica, é necessário utilizar

sistemas de projeção cartográfica para representá-la em superfícies planas. Cada projeção implica distorções, algumas afetam áreas, outras distâncias ou formas. Projeções conformes preservam as formas, enquanto as de área igual mantêm as proporções das áreas, sendo impossível conservar ambos simultaneamente. A escolha adequada da projeção é essencial, especialmente em análises de grande escala, como cálculos de área de países (Bolstad, 2012).

Apesar da existência de centenas de projeções, apenas algumas são amplamente usadas em sistemas de informações geográficas (SIG), como a projeção cônica conforme de Lambert e a Mercator transversal, aplicadas em diversas regiões do mundo. Essas projeções servem de base para sistemas padronizados utilizados em diferentes contextos geoespaciais (Bolstad, 2012).

Para que dados espaciais sejam úteis, é necessário vinculá-los a um sistema de referência espacial. O georreferenciamento consiste em associar posições da superfície terrestre a coordenadas, como as obtidas por GPS, ajustando imagens para que se alinhem corretamente a pontos de controle em campo, processo conhecido como georretificação. Já a geocodificação converte endereços em coordenadas geográficas, permitindo relacionar nomes e endereços a locais específicos. Essa etapa é essencial para análises espaciais e pode incluir estimativas de localização com base em redes viárias quando dados precisos não estão disponíveis (Bolstad, 2012).

Governos nacionais frequentemente desenvolvem, organizam, arquivam e distribuem conjuntos de dados espaciais em escala nacional. A padronização de medidas e unidades é uma função essencial desses governos, e os dados espaciais podem ser considerados como medições de terras, mares ou outros territórios do país. Além disso, cabe ao governo planejar, construir e gerenciar infraestruturas públicas, como estradas, hidrovias e sistemas de distribuição de energia, tarefas que exigem o uso de dados espaciais abrangentes em todo o território nacional.

O *OpenStreetMap* (OSM) é uma iniciativa notável que visa criar dados geográficos globais por meio da colaboração voluntária internacional, funcionando de forma semelhante à *Wikipedia*. Usuários registrados podem editar conjuntos de dados, geralmente digitalizando vias e outras infraestruturas a partir de imagens ou com o uso de GNSS (*Global Navigation Satellite System* em inglês). Há protocolos de revisão e resolução de conflitos, e os dados podem ser baixados em diversos formatos. Essa pla-

taforma costuma oferecer os melhores dados em regiões com infraestrutura cartográfica limitada (Bolstad, 2012).

No entanto, como é uma colaboração aberta, podem existir limitações quanto à padronização e documentação da qualidade. A variedade de fontes e métodos utilizados pode comprometer a uniformidade e a completude dos dados, que também depende do nível de envolvimento voluntário em determinada área. Por isso, é essencial verificar a precisão e a adequação das informações antes de utilizá-las. Apesar dessas limitações, o detalhamento do OSM costuma justificar o esforço de validação (Bolstad, 2012).

3 TRABALHOS CORRELATOS

A construção de uma ferramenta de busca espacial baseada em critérios estruturais e geográficos, como a proposta neste trabalho, dialoga com diferentes vertentes da literatura, desde a definição de critérios de relevância para o usuário até a complexidade algorítmica do processamento de grafos. A seguir, articulam-se essas pesquisas para fundamentar as escolhas metodológicas adotadas.

No contexto de busca por localização e moradia, a simples proximidade física não é o único fator determinante. Annisa e Khairina (2021) demonstram, através do uso de consultas *skyline*, que a atratividade de um imóvel depende de múltiplos critérios combinados, como a diversidade de serviços no entorno. Essa perspectiva valida o objetivo da presente pesquisa, que busca superar as limitações apontadas por Liu et al. (2019) em sistemas tradicionais (como o Google Maps). Segundo os autores, tais sistemas dependem excessivamente de palavras-chave exatas e conhecimento prévio, falhando em capturar a semântica de "vizinhança" desejada pelo usuário.

Para solucionar essa lacuna, a modelagem do problema migra da simples busca textual para a teoria dos grafos. Neste cenário, a tarefa de encontrar um conjunto de serviços (ex: escola, farmácia e mercado) em uma configuração espacial específica configura-se como um problema de isomorfismo de subgrafos. Cordella et al. (2004) tratam da alta complexidade computacional inerente a esse problema com o algoritmo VF2, focado em correspondências exatas. No entanto, para aplicações de busca em mapas do mundo real, onde os dados podem ser ruidosos e as distâncias variam, a rigidez de algoritmos exatos pode ser um obstáculo.

Nesse sentido, a abordagem de Khan et al. (2013) com o fra-

mework NeMa (*Neighborhood Matching*) oferece um suporte teórico importante para a estratégia de validação funcional com dados do OSM utilizada neste trabalho. Ao focar na similaridade de vizinhança e em contextos heterogêneos, Khan et al. justificam a necessidade de algoritmos que tolerem variações e priorizem a estrutura topológica em vez da identidade exata, conceito central na adaptação do algoritmo de busca aqui desenvolvido.

Dessa forma, a base metodológica central desta pesquisa fundamenta-se na proposta de Liu et al. (2019). Diferentemente de abordagens puramente combinatórias, os autores modelam a busca como um problema de correspondência de subgrafos otimizado por atributos (localização, distância e avaliações) e resolvido através de sistemas dinâmicos (equação do replicador). Esta pesquisa adota e estende essa metodologia, replicando o algoritmo de Liu et al. (2019). para validação com dados do Yelp e, subsequentemente, adaptando-o para um novo domínio de aplicação utilizando dados do OSM, onde a busca por padrões espaciais é realizada sem a necessidade de normalização de atributos qualitativos.

4 MATERIAIS E MÉTODOS

A implementação do filtro de mapas fundamentou-se em uma pesquisa com base tecnológica, cujo objetivo foi identificar locais que ofereçam os serviços desejados dentro de distâncias previamente definidas.

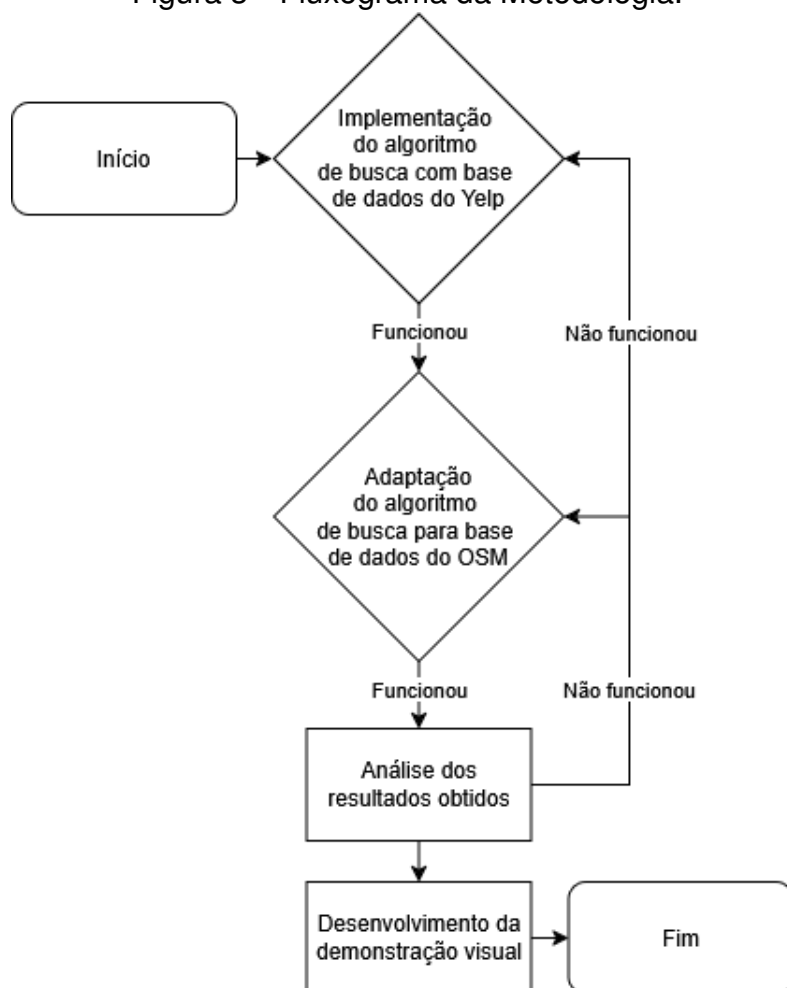
A aplicação é responsável por receber as entradas do usuário, que consistem na cidade principal, nos serviços desejados e na distância preferida até cada serviço. O filtro emprega dois algoritmos distintos para o processamento dos grafos derivados da entrada do usuário e das posições resultantes das consultas ao banco de dados: o primeiro identifica os grupos candidatos, enquanto o segundo detecta os padrões comuns entre eles. Ao final, a aplicação apresenta ao usuário os pontos encontrados por meio de uma visualização geográfica em um mapa.

A metodologia adotada neste trabalho baseia-se na proposta de Liu et al. (2019) e seguiu os passos mostrados na Figura 3. Para garantir a confiabilidade do sistema, a estratégia de validação foi dividida em duas etapas:

1. Validação por referência, utilizando o conjunto de dados Yelp³ para verificar a reprodutibilidade dos resultados em comparação à literatura base (Liu et al., 2019);
2. Validação funcional aplicada aos dados do OSM, na qual se avaliou a

precisão das distâncias e a coerência topológica dos resultados retornados em cenários de teste controlados.

Figura 3 - Fluxograma da Metodologia.



Fonte: Do autor.

4.1 IMPLEMENTAÇÃO DO ALGORITMO DE BUSCA COM OS DADOS DO YELP

O processo inicia-se com a leitura e o preparo dos dados, a partir do arquivo JSON do Yelp, que é carregado em partes (*chunks*). Durante essa etapa, são removidas as colunas desnecessárias e, ao final, os fragmentos são consolidados em um único *DataFrame* contendo apenas as informações relevantes sobre os estabelecimentos, como *business_id*, nome, cidade, localização geográfica, número de avaliações e nota média.

Em seguida, seleciona-se um conjunto de identificadores (*business_id*) correspondentes aos estabelecimentos da cidade de origem que

³<https://business.yelp.com/data/resources/open-dataset/>

o usuário deseja buscar. Esses identificadores são armazenados em uma lista para uso nas etapas posteriores de busca e comparação. Paralelamente, os dados do Yelp são filtrados de modo que apenas os registros selecionados anteriormente e da cidade de destino sejam mantidos, restringindo o escopo da análise às duas localidades envolvidas.

A etapa central do processo implementa o algoritmo de descoberta de padrões comuns. Esse algoritmo recebe a lista de identificadores da cidade de origem (por exemplo, Phoenix, EUA) e procura grupos de estabelecimentos semelhantes na cidade de destino (por exemplo, Las Vegas, EUA).

Ele começa realizando uma normalização dos atributos, de modo que medidas com diferentes escalas possam ser comparadas de forma coerente. Funções utilitárias aplicam a normalização Min-Max aos valores de estrelas e número de avaliações, convertendo-os para o intervalo $[0, 1]$ (Shalabi; Shaaban; Kasasbeh, 2006). Essa padronização é essencial, pois evita que atributos com magnitudes muito distintas (como notas de 1 a 5 e contagens de avaliações que podem chegar a milhares) distorçam o cálculo de similaridade.

Com os dados devidamente normalizados, procede-se ao cálculo da similaridade entre pares de estabelecimentos. Essa similaridade resulta da combinação de múltiplos fatores:

- A proximidade entre as notas médias (estrelas);
- A semelhança no número de avaliações;
- A correspondência entre as categorias (verificando se as categorias do estabelecimento de destino incluem aquelas do estabelecimento de origem);
- A distância geográfica entre eles.

Para isso, são utilizadas duas funções complementares: a primeira mede a similaridade local, avaliando diretamente os atributos de dois estabelecimentos; a segunda calcula a similaridade estrutural, incorporando também a distância espacial entre os pontos. Essa distância é obtida por meio da fórmula de Haversine, que determina a separação entre dois pontos na superfície terrestre a partir de suas latitudes e longitudes (Robusto, 1957). Dessa forma, o algoritmo não apenas compara as características

individuais dos negócios, mas também analisa se suas distribuições espaciais seguem padrões semelhantes entre as cidades consideradas.

Após essa etapa, são identificados os “candidatos”, estabelecimentos na cidade de destino que possuem categorias idênticas ou compatíveis com as dos estabelecimentos da cidade de origem. Essa filtragem garante que a comparação ocorra entre tipos equivalentes de negócios (por exemplo, restaurantes italianos sendo comparados apenas com outros restaurantes italianos).

Em seguida, o algoritmo constrói uma matriz de adjacência A , que representa as possíveis correspondências entre os estabelecimentos das duas cidades. Cada célula dessa matriz contém um valor de similaridade calculado com base nos atributos (como estrelas e número de avaliações) e na estrutura (como distância geográfica entre pares). Essa matriz constitui o núcleo matemático do processo, funcionando como um “grafo de similaridade” que descreve as conexões potenciais entre os grupos.

Para identificar o melhor padrão, o programa aplica uma técnica inspirada em sistemas dinâmicos, conhecida como equação do replicador (Sandholm, 2009). Essa equação atualiza iterativamente um vetor de pesos x , em que cada elemento representa o grau de confiança em uma correspondência entre negócios. A cada iteração, os pesos são ajustados conforme os valores da matriz A , até que o sistema alcance um estado de equilíbrio, isto é, quando as mudanças entre as iterações tornam-se insignificantes. Esse processo evidencia as correspondências mais fortes e consistentes entre os grupos analisados.

Por fim, com as correspondências estabelecidas, o programa avalia e classifica os padrões descobertos. Cada grupo de correspondência recebe uma pontuação calculada pela densidade do subgrafo correspondente, a qual expressa o grau de coesão e consistência do padrão entre as duas cidades. Os grupos são então ordenados do mais ao menos semelhante, e apenas aqueles que atingem um limiar mínimo de pontuação são exibidos. Para cada correspondência encontrada, o programa apresenta quais estabelecimentos da cidade de origem se alinham com os da cidade de destino.

Para a validação desta etapa, adotou-se como critério a consistência funcional do algoritmo. Devido à indisponibilidade do recorte exato do conjunto de dados utilizado originalmente por Liu et al. (2019), optou-se pela construção de um cenário de teste análogo utilizando uma nova consulta à base do Yelp. Embora os índices absolutos de similaridade e

a densidade dos grupos encontrados tenham sido inferiores aos do estudo de referência, tal divergência é atribuída às diferenças temporais e de amostragem entre as bases de dados. Independentemente dessa variação quantitativa, o algoritmo cumpriu sua função ao identificar correlações e formar clusters topologicamente coerentes. Dessa forma, considerou-se a implementação validada, uma vez que o sistema aplicou corretamente a equação do replicador e as métricas de similaridade para filtrar e estruturar os dados disponíveis.

4.2 IMPLEMENTAÇÃO DO ALGORITMO DE BUSCA COM OS DADOS DO OSM

Para a implementação com dados do OSM, o método empregado consistiu em uma adaptação da abordagem previamente aplicada aos dados do Yelp. Nessa adaptação, utilizou-se o PostgreSQL, em conjunto com sua extensão PostGIS, como sistema de gerenciamento de banco de dados espacial. Além disso, foram empregados os utilitários `osm2pgsql` e `shp2pgsql` para a importação dos dados do OSM e da Malha Municipal, respectivamente. A Malha Municipal é uma base cartográfica que contém os polígonos dos municípios brasileiros e foi utilizada para auxiliar na delimitação geográfica das análises realizadas no banco de dados.

4.2.1 Pré-processamento dos dados geográficos

Para otimizar as consultas executadas pelo script do programa, foram filtradas três tabelas a partir dos dados obtidos do OSM. A primeira tabela contém as colunas `osm_id`, `name`, `category`, `city`, `longitude` e `latitude`, referentes aos estabelecimentos extraídos da coluna `amenities`. As outras duas tabelas possuem a mesma estrutura, mas seus dados foram extraídos da coluna `building`.

Para chegar a essa estrutura, após o carregamento da Malha Municipal, foram executadas queries destinadas a associar os estabelecimentos às respectivas cidades (Apêndices C e D) e a criar as três tabelas mencionadas anteriormente (Imagem 4 e apêndices E e F).

4.2.2 Adaptação do algoritmo de busca para os dados do OSM

A adaptação do algoritmo de busca teve como objetivo possibilitar o uso dos dados provenientes do OSM na identificação de padrões espaciais urbanos, além de permitir a realização de pesquisas baseadas em um ponto central (como uma casa ou hotel) considerando os serviços localizados em determinadas distâncias.

Figura 4 - SQL para criação da tabela amenity_locations

```
DROP TABLE IF EXISTS amenity_locations;

CREATE TABLE amenity_locations AS
WITH unique_polygons AS (
    SELECT DISTINCT ON (osm_id)
        osm_id,
        name,
        amenity,
        cidade_nome,
        way
    FROM
        planet_osm_polygon
    WHERE
        amenity IS NOT NULL AND cidade_nome IS NOT NULL

    SELECT
        osm_id,
        name,
        amenity AS category,
        cidade_nome AS city,
        ST_X(ST_Transform(ST_PointOnSurface(way), 4326)) AS longitude,
        ST_Y(ST_Transform(ST_PointOnSurface(way), 4326)) AS latitude
    FROM
        unique_polygons

    UNION ALL

    SELECT
        p.osm_id,
        p.name,
        p.amenity AS category,
        p.cidade_nome AS city,
        ST_X(ST_Transform(p.way, 4326)) AS longitude,
        ST_Y(ST_Transform(p.way, 4326)) AS latitude
    FROM
        planet_osm_point AS p
    WHERE
        p.amenity IS NOT NULL
        AND p.cidade_nome IS NOT NULL
        AND NOT EXISTS (
            SELECT 1
            FROM unique_polygons up
            WHERE up.osm_id = p.osm_id
        );
```

Fonte: Do autor.

A abordagem proposta não realiza a normalização dos atributos dos estabelecimentos, concentrando-se exclusivamente na similaridade baseada em distância geográfica. Dessa forma, o cálculo de correspondência entre os pontos não considera características descritivas adicionais, como tamanho, popularidade ou avaliações, mas apenas a coerência espacial entre as posições dos elementos analisados e a compatibilidade por categoria.

Duas funções principais são utilizadas nesse processo. A primeira mede a similaridade local, avaliando diretamente a compatibilidade entre dois estabelecimentos de acordo com a categoria de serviço, por exemplo, se ambos pertencem à categoria “restaurante” ou “banco”. A se-

gunda função calcula a similaridade estrutural, que incorpora a diferença entre as distâncias esperadas (definidas pelo modelo de origem) e as distâncias reais observadas entre os pontos. Essa diferença é transformada em uma medida contínua de similaridade, de modo que valores menores de discrepância resultam em pontuações mais altas, indicando maior correspondência espacial.

A análise espacial é conduzida considerando cada acomodação (casa, hotel ou ambos) como um ponto central. Para cada uma delas, são identificados os serviços próximos pertencentes às categorias especificadas na consulta. Essa busca é restrita a uma distância máxima predefinida de 20 km, de modo a considerar apenas as amenidades realmente acessíveis a partir do ponto central. Os estabelecimentos encontrados são então relacionados à acomodação e avaliados com base na diferença entre a distância ideal (esperada) e a distância real observada no espaço.

A avaliação da similaridade entre a estrutura espacial esperada e a observada é realizada de forma contínua, utilizando uma função que atribui pontuações mais altas quando as distâncias medidas se aproximam das distâncias de referência. Dessa forma, padrões espaciais que mantêm proporções e distribuições semelhantes às desejadas recebem maior pontuação, refletindo uma maior compatibilidade com o modelo de busca definido.

Depois de criar o grafo de correspondência e de sua otimização, os padrões espaciais são avaliados de forma quantitativa. O algoritmo calcula uma pontuação final que combina a média das similaridades de distância com um fator de completude, que considera a proporção de serviços da consulta que foram efetivamente encontrados. Assim, são privilegiados os centros de acomodação que não apenas possuem serviços próximos, mas que também mantêm as distâncias relativas entre eles de forma coerente com o padrão de referência.

A verificação do funcionamento com dados do OSM foi realizada por meio de estudos de caso submetidos à verificação manual. Definiu-se um conjunto de consultas de teste com distâncias fixas (raios de busca) e categorias específicas. A validação consistiu na conferência da Acurácia Espacial, verificando se:

1. Todos os serviços retornados estavam, de fato, dentro do raio euclidiano ou de Haversine especificado em relação ao ponto central;
2. A filtragem de categorias eliminou corretamente falsos positivos (es-

tabelecimentos fora do escopo semântico);

3. O ranking de similaridade priorizou corretamente as configurações espaciais com menor desvio métrico em relação à consulta original.

Essa etapa assegurou que a migração da base de dados (Yelp para OSM) e a adaptação das métricas de similaridade mantiveram a integridade lógica da ferramenta.

Por fim, os resultados são ordenados de acordo com a pontuação obtida e os melhores padrões identificados são selecionados. Cada padrão é então composto por um ponto central (a acomodação) e pelos serviços associados a ele, formando um agrupamento espacial coerente. Essa abordagem é especialmente útil para compreender a distribuição e a organização dos serviços urbanos, bem como para apoiar a tomada de decisão em cenários de recomendação geográfica ou planejamento territorial.

5 RESULTADOS E DISCUSSÃO

Inicialmente, o software foi avaliado utilizando os dados do Yelp, cuja abordagem consiste em selecionar um conjunto de identificadores (*business_id*) correspondentes aos estabelecimentos da cidade de origem. Em seguida, foram implementadas adaptações para o uso com os dados do OSM, possibilitando a execução de buscas a partir de um ponto central (como uma residência ou hotel) e a consideração de serviços localizados dentro de faixas específicas de distância.

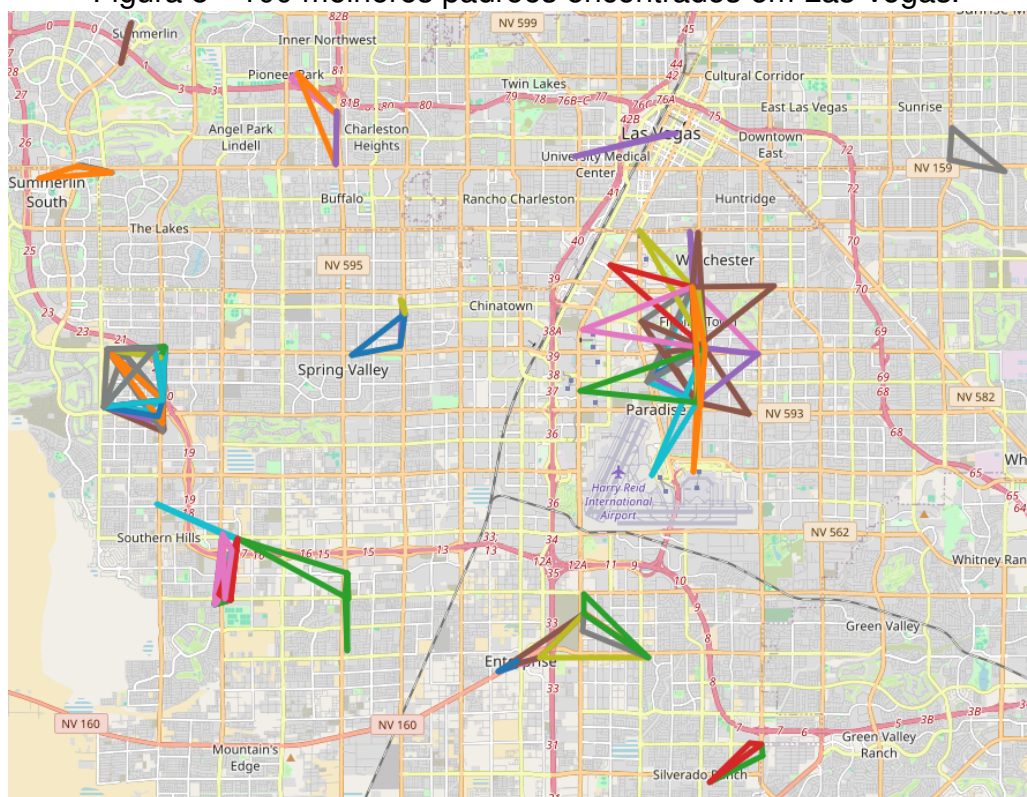
5.1 YELP

As cidades selecionadas para o estudo foram Phoenix, considerada a cidade de origem, e Las Vegas, definida como cidade-alvo. Os serviços incluídos na análise foram escolhidos com base em suas necessidades cotidianas, além de estarem situados próximos uns dos outros, o que é relevante para restringir o processo de busca.

O Apêndice B detalha os estabelecimentos específicos na cidade de origem, Phoenix, que foram utilizados como base para o padrão de busca. A imagem exibe um trecho de dados listando os três serviços selecionados: um restaurante (In-N-Out Burger), um centro de logística e impressão (The UPS Store) e um serviço financeiro (Bank of America).

A Figura 5 apresenta um mapa de Las Vegas, a cidade-alvo, exibindo a localização dos 100 melhores padrões encontrados que correspondem à busca. Os padrões são representados por formas geométricas

Figura 5 - 100 melhores padrões encontrados em Las Vegas.



Fonte: Do autor.

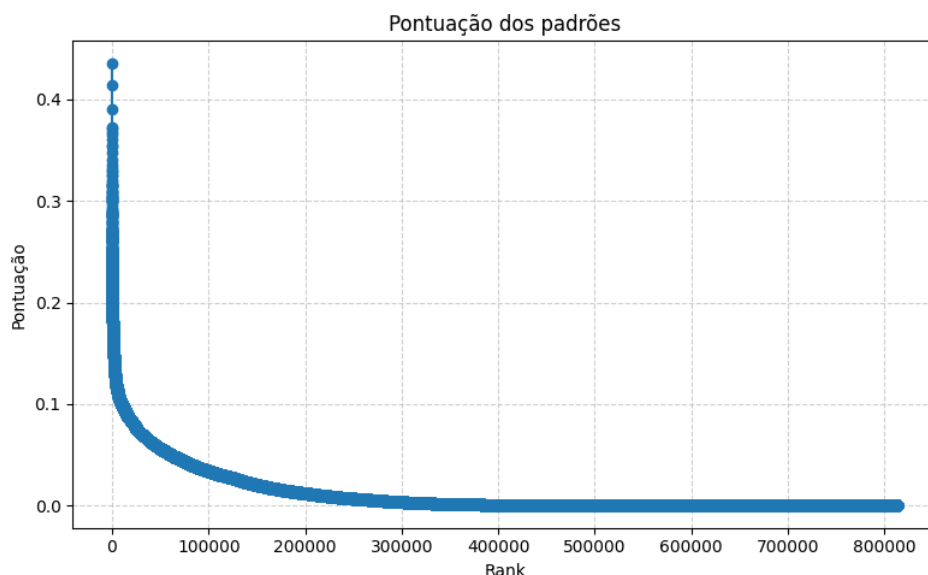
(polígonos) que conectam os serviços correspondentes, ilustrando sua distribuição espacial pela cidade.

A Figura 6 é um gráfico de dispersão que plota a pontuação (eixo Y) pelo "Rank"(eixo X) de todos os padrões encontrados no teste do Yelp. O gráfico demonstra que um número muito pequeno de padrões obteve pontuação alta (próxima de 0,4), havendo uma queda acentuada, com a grande maioria dos padrões classificados tendo pontuação próxima de zero.

Os trabalhos de Khan et al. (2013) e Liu et al. (2019) propõem abordagens para a busca de padrões em grafos. Khan et al. (2013) apresenta o NeMa, um algoritmo heurístico que identifica as k melhores correspondências aproximadas, otimizadas por uma métrica de custo. De forma similar, o framework de Liu et al. (2019), que serviu de inspiração metodológica para este TCC, também gera uma lista das top-k correspondências, classificadas de acordo com uma pontuação de similaridade multimodal (de 0 a 1).

No entanto, uma comparação quantitativa direta com Liu et al. (2019) é limitada, uma vez que os experimentos originais utilizaram dados e parâmetros diferentes. Além disso, a indisponibilidade do código-fonte e o desconhecimento das métricas exatas de normalização empregadas dificultam

Figura 6 - Pontuação de todos os padrões encontrados.



Fonte: Do autor.

tam qualquer comparação precisa, o que explica a discrepância observada nos resultados, já que Liu et al. (2019) relata pontuações máximas entre 0,8 e 0,9 e a implementação preliminar deste estudo atingiu pontuações pouco acima de 0,4.

Embora este trabalho utilize um framework de similaridade (com pontuações de 0 a 1) derivado de Liu et al. (2019), o foco dos resultados finais diverge: em vez de otimizar uma lista top-k de resultados multimodais (como Liu) ou aproximados (como Khan), a implementação final com OSM é utilizada para gerar uma lista exaustiva de todas as localizações que satisfazem exatamente um padrão espacial com restrições de distância precisas (ou seja, que atingem a pontuação de similaridade máxima).

Os resultados apresentados por Annisa e Khairina (2021) consistem em uma lista filtrada de locais "skyline"(ótimos), que representam as melhores combinações (trade-offs) baseadas na contagem de instalações (ex: o local com o melhor equilíbrio entre classificação e número de restaurantes). Diferentemente, a abordagem deste trabalho não visa classificar, mas sim gerar uma lista exaustiva de todos os locais que correspondem a um padrão espacial predefinido, permitindo restrições de distância individualizadas para cada serviço (ex: um hotel que possua um banco a 2 km e um restaurante a 1 km). Em termos de performance, enquanto o desempenho em Annisa e Khairina (2021) foi consideravelmente afetado pelo tempo de aquisição de dados via API do Google Maps, nossos resultados indicam que a abordagem baseada em grafos é mais sensível à densidade

de estabelecimentos (nós) extraídos do OSM.

O trabalho de Cordella et al. (2004) aborda o desafio algorítmico fundamental da correspondência de (sub)grafos, que forma a base metodológica deste estudo. Os resultados apresentados por Cordella et al. (2004) são métricas de desempenho que validam a eficiência do seu algoritmo (VF2) em termos de tempo e complexidade espacial, testado contra bancos de dados de grafos em larga escala, tanto sintéticos quanto de aplicações reais. Em contraste, os resultados deste trabalho, embora dependentes de algoritmos similares, são de natureza aplicada. Em vez de avaliar o desempenho algorítmico puro, nossos resultados medem a eficácia da aplicação dessa técnica na identificação de padrões espaciais em dados geográficos reais (OSM), demonstrando que o desempenho da consulta é sensível a características dos dados, como a densidade de estabelecimentos (nós) na área de busca.

5.2 OSM

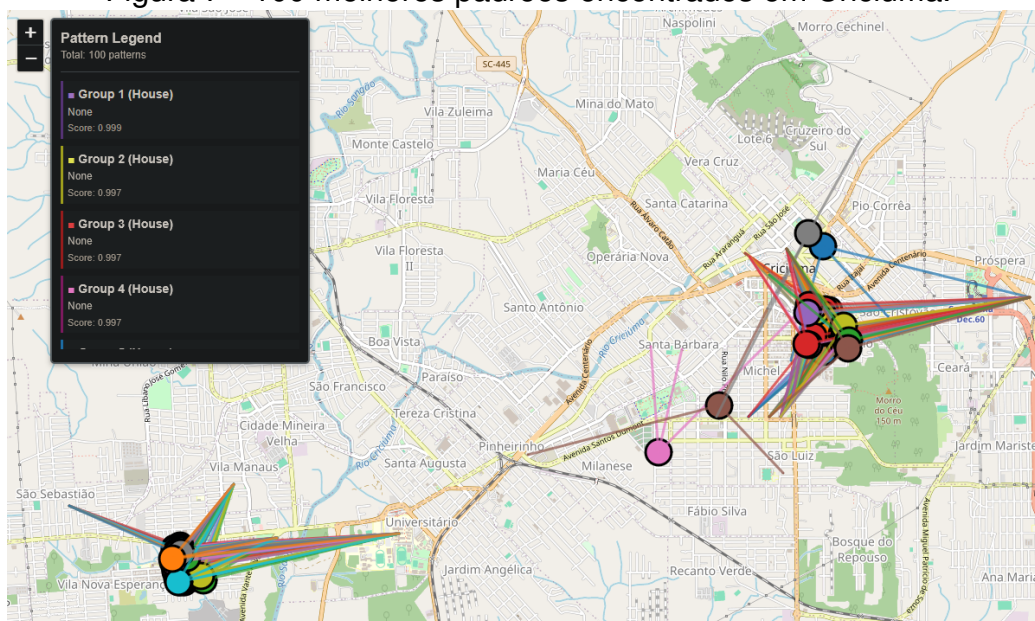
A cidade escolhida para a realização da busca foi Criciúma. Os serviços incluídos na análise, assim como suas respectivas distâncias, foram selecionados com base em necessidades cotidianas. Diferentemente de outros cenários, não foi necessário que esses serviços estivessem próximos entre si, já que o algoritmo não considera a proximidade entre eles.

O Apêndice A ilustra os parâmetros de entrada utilizados para a busca no teste com dados do OSM, especificamente para a cidade de Criciúma. A imagem exibe um trecho de código que define as três categorias de serviços selecionadas para o padrão: 'restaurant' (restaurante), 'bank' (banco) e 'place_of_worship' (local de culto), associadas às suas respectivas distâncias de busca (1, 2 e 1).

A Figura 7 exibe a visualização dos resultados em um mapa de Criciúma, destacando os 100 melhores padrões identificados. No mapa, os padrões são representados por agrupamentos de pontos e linhas, concentrados em regiões específicas da cidade. A legenda ("Pattern Legend") especifica os diferentes grupos de padrões detectados, vinculados a pontos centrais, como residências ou hotéis, sendo que cada linha indica a conexão com uma categoria dos parâmetros utilizados na análise.

A Figura 8 é um gráfico de dispersão que exibe a pontuação de todos os padrões encontrados no teste do OSM. O gráfico, intitulado "Pontuação das acomodações por rank", plota a "Pontuação" (eixo Y) contra o "Rank" (eixo X). Observa-se que a pontuação inicial é 1,0 e decai de forma

Figura 7 - 100 melhores padrões encontrados em Criciúma.



Fonte: Do autor.

lenta e gradual, mostrando uma queda em torno do rank 550.

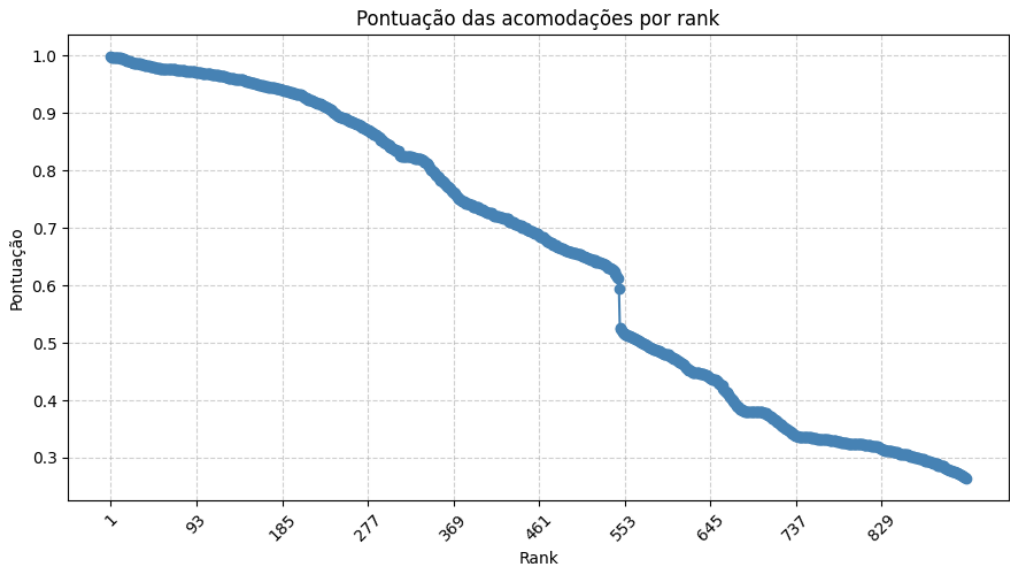
Ao final da avaliação, observaram-se algumas limitações relevantes relacionadas ao desempenho e à abrangência da busca. O algoritmo, na forma atual, considera apenas os serviços que se encontram exatamente na distância especificada em relação ao ponto central, desconsiderando estabelecimentos situados entre o ponto de origem e o limite de distância definido. Essa restrição reduz a flexibilidade da análise espacial e pode deixar de identificar padrões potencialmente relevantes. Além disso, verificou-se que o tempo de execução aumenta significativamente em cidades com um grande número de acomodações ou estabelecimentos, tornando o processo mais lento e exigindo maior poder computacional para a geração dos resultados.

6 CONCLUSÃO

Este trabalho teve como objetivo principal o desenvolvimento de um filtro para mapas capaz de localizar áreas de interesse com base na proximidade de serviços definidos pelo usuário. Para alcançar esse objetivo, a metodologia foi fundamentada na teoria dos grafos, especificamente em algoritmos de correspondência de subgrafos, e validada em duas etapas distintas.

Inicialmente, o algoritmo foi implementado e testado utilizando o conjunto de dados do Yelp, replicando a abordagem de Liu et al. (2019) para

Figura 8 - Pontuação de todos os padrões encontrados em Criciúma.



Fonte: Do autor.

encontrar padrões estruturais e atributos semelhantes entre duas cidades (Phoenix e Las Vegas). Os resultados deste primeiro teste indicaram que a correspondência exata de padrões complexos, incluindo atributos como avaliações e distância geográfica relativa entre estabelecimentos, é pouco frequente, produzindo um número reduzido de correspondências com pontuações elevadas e uma queda acentuada ao longo do ranqueamento.

Após essa etapa, o algoritmo foi adaptado para operar com dados geográficos do OSM. Essa adaptação modificou a natureza da busca. Em vez de comparar padrões entre duas cidades, o filtro passou a identificar locais em uma cidade-alvo que continham um conjunto de serviços essenciais, como restaurante, banco e local de culto, dentro de distâncias específicas definidas pelo usuário.

A aplicação no OSM gerou um conjunto amplo de correspondências baseadas exclusivamente na proximidade espacial a partir de um ponto central, permitindo observar um número significativo de padrões compatíveis com as combinações de serviços e distâncias estabelecidas. Diferentemente do teste com o Yelp, essa configuração resultou em distribuições de pontuação mais graduais e em um volume maior de ocorrências.

A pesquisa resultou no desenvolvimento de um filtro que transforma uma consulta baseada em múltiplos serviços e distâncias em um subgrafo de busca, aplicando-o a bases de dados geográficas para identificar localizações que atendem a esses critérios. A ferramenta apresenta potencial de uso em sistemas de busca de imóveis e em análises urbanas,

permitindo organizar regiões com base na proximidade a serviços essenciais.

Como trabalhos futuros, sugerem-se aprimoramentos que incluam um sistema de ponderação de serviços para refletir as prioridades do usuário, a incorporação de dados temporais, como horários de funcionamento, a substituição da distância linear por rotas baseadas na malha viária ou pelo tempo de deslocamento, além da integração com outras fontes de dados. Outras possibilidades incluem a expansão para atributos adicionais da área, mecanismos de busca inversa e estratégias de otimização voltadas à escalabilidade em bases extensas.

REFERÊNCIAS

ABREU, T. d. S. **Marketing digital: um estudo no mercado imobiliário de são luís do maranhão**. 2019, UFMA.

ANNISA, A.; KHAIRINA, S. **Location selection based on surrounding facilities in google maps using sort filter skyline algorithm**. Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, 2021, v. 7, n. 2, p. 65–72.

BOLSTAD, P. **GIS Fundamentals: A First Text on Geographic Information Systems**. [S.l.: s.n.], 2012. ISBN 978-0-9717647-3-6.

BONDY, J. A.; MURTY, U. S. R. et al. **Graph theory with applications**. [S.l.]: Macmillan London, 1976. v. 290.

CHANG, K.-T. **Introduction to geographic information systems**. [S.l.]: Mc Graw Hill, 2019.

CIBEJ, U.; MIHELİČ, J. Search strategies for subgraph isomorphism algorithms. In: . [S.l.: s.n.], 2014. v. 8321. ISBN 978-3-319-04125-4.

CONSTRUÇÃO, C. B. da Indústria da. **Vendas de imóveis crescem 20,9% e lançamentos aumentam 18,6% em 2024, diz CBIC**. 2025. Acesso em: 15 jul. 2025. Disponível em: <<https://cbic.org.br/vendas-de-imoveis-crescem-209-e-lancamentos-aumentam-186-em-2024-diz-cbic/>>.

CONTEÚDO, E. **Mercado imobiliário fecha 2018 com alta de 19,2% em vendas, diz CBIC**. 2019. Acesso em: 15 out. 2024. Disponível em: <<https://veja.abril.com.br/economia/mercado-imobiliario-fecha-2018-com-alta-de-192-em-vendas-diz-cbic/>>.

CORDELLA, L. P. et al. **A (sub) graph isomorphism algorithm for matching large graphs**. IEEE transactions on pattern analysis and machine intelligence, 2004, IEEE, v. 26, n. 10, p. 1367–1372.

CORREIA, P. G.; REIS, A. M. P. dos. **Pesquisa estruturada versus pesquisa em texto livre na seleção de imóveis em sites imobiliários.** Revista Portuguesa e Brasileira de Gestão, 2011, v. 10, n. 4, p. 2–10.

DEO, N. **Graph theory with applications to engineering and computer science.** [S.l.]: Courier Dover Publications, 2016.

KHAN, A. et al. **Nema: Fast graph search with label similarity.** Proceedings of the VLDB Endowment, 2013, VLDB Endowment, v. 6, n. 3, p. 181–192.

KRAUS, S. et al. **Facebook and the creation of the metaverse: radical business model innovation or incremental transformation?** International Journal of Entrepreneurial Behavior & Research, 2022, Emerald Publishing Limited, v. 28, n. 9, p. 52–77.

LAVADO, G. T. **Uso da internet no Brasil cresce, e 70% da população está conectada.** 2019. Acesso em: 15 jul. 2025. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2019/08/28/uso-da-internet-no-brasil-cresce-e-70percent-da-populacao-esta-conectada.ghtml>.

LIU, W. et al. **A common subgraph correspondence mining framework for map search services.** Multimedia Tools and Applications, 2019, Springer, v. 78, n. 1, p. 747–766.

MA, S. et al. **Big graph search: challenges and techniques.** Frontiers of Computer Science, 2016, Springer, v. 10, p. 387–398.

ROBUSTO, C. C. **The cosine-haversine formula.** The American Mathematical Monthly, 1957, JSTOR, v. 64, n. 1, p. 38–40.

SANDHOLM, W. H. Evolutionary game theory. In: **Encyclopedia of complexity and systems science.** [S.l.]: Springer, 2009. p. 3176–3205.

SHALABI, L. A.; SHAABAN, Z.; KASASBEH, B. **Data mining: A preprocessing engine.** Journal of Computer Science, 2006, v. 2, n. 9, p. 735–739.

ZHANG, P.; CHARTRAND, G. **Introduction to graph theory.** [S.l.]: Tata McGraw-Hill New York, 2006. v. 2.

A APÊNDICE - Categorias e distâncias escolhidas em Criciúma para busca.

```
1  [
2    {
3      "category": "restaurant",
4      "distance": 1
5    },
6    {
7      "category": "bank",
8      "distance": 2
9    },
10   {
11     "category": "place_of_worship",
12     "distance": 1
13   }
14 ]
```

Fonte: Do autor.

B APÊNDICE - Serviços escolhidos em Phoenix para busca.

```
1  [
2    {
3      "business_id": "1ku0t5u7dCAE0g-pZDQdAw",
4      "name": "In-N-Out Burger",
5      "city": "Phoenix",
6      "latitude": 33.5073002958,
7      "longitude": -112.0396049,
8      "stars": 4.0,
9      "review_count": 62,
10     "categories": "Burgers, Fast Food, Restaurants"
11   },
12   {
13     "business_id": "x9KsthqdtEFwydPVLvZpQw",
14     "name": "The UPS Store",
15     "city": "Phoenix",
16     "latitude": 33.506103467,
17     "longitude": -112.041238332,
18     "stars": 2.0,
19     "review_count": 41,
20     "categories": "Shipping Centers, Local Services, Notaries, Printing Services"
21   },
22   {
23     "business_id": "EPGpYU55Y9U1DtVwNgz2QA",
24     "name": "Bank of America",
25     "city": "Phoenix",
26     "latitude": 33.5095959,
27     "longitude": -112.0391184,
28     "stars": 2.5,
29     "review_count": 14,
30     "categories": "Financial Services, Home Services, Mortgage Brokers, Real Estate, Banks & Credit Unions"
31   }
32 ]
```

Fonte: Do autor.

C APÊNDICE - SQL para criação das colunas cidade_nome e cidade_codigo nas tabelas planet_osm_point e planet_osm_polygon

```
1  ALTER TABLE planet_osm_point
2  ADD COLUMN cidade_nome VARCHAR(255),
3  ADD COLUMN cidade_geocodigo VARCHAR(7);
4
5  ALTER TABLE planet_osm_polygon
6  ADD COLUMN cidade_nome VARCHAR(255),
7  ADD COLUMN cidade_geocodigo VARCHAR(7);
```

Fonte: Do autor.

D APÊNDICE - SQL para atualizar em qual cidade o serviço se encontra

```
1  UPDATE planet_osm_point AS points
2  SET
3      cidade_nome = municipios.nm_municipio,
4      cidade_geocodigo = municipios.cd_mun
5  FROM
6      public.municipios AS municipios
7  WHERE
8      ST_Within(
9          ST_Transform(points.way, 4674),
10         municipios.geom
11     );
12
13  UPDATE planet_osm_polygon AS polygons
14  SET
15      cidade_nome = municipios.nm_municipio,
16      cidade_geocodigo = municipios.cd_mun
17  FROM
18      public.municipios AS municipios
19  WHERE
20      ST_Within(
21          ST_Transform(ST_PointOnSurface(polygons.way), 4674),
22          municipios.geom
23     );
```

Fonte: Do autor.

E APÊNDICE - SQL para criação da tabela house_locations

```
1  CREATE TABLE house_locations AS
2  SELECT
3      osm_id,
4      name,
5      building as category,
6      cidade_nome as city,
7      ST_X(ST_Transform(way, 4326)) AS longitude,
8      ST_Y(ST_Transform(way, 4326)) AS latitude
9  FROM
10     planet_osm_point
11  WHERE
12     (building='house' OR
13     building='detached' OR
14     building='semidetached_house' OR
15     building='bungalow') AND
16     cidade_nome IS NOT NULL
17  UNION ALL
18  SELECT
19      osm_id,
20      name,
21      building as category,
22      cidade_nome as city,
23      ST_X(ST_Transform(ST_Centroid(way), 4326)) as longitude,
24      ST_Y(ST_Transform(ST_Centroid(way), 4326)) as latitude
25  FROM
26     planet_osm_polygon
27  WHERE
28     (building='house' OR
29     building='detached' OR
30     building='semidetached_house' OR
31     building='bungalow') AND
32     cidade_nome IS NOT NULL
```

Fonte: Do autor.

F APÊNDICE - SQL para criação da tabela hotel_locations

```
1  CREATE TABLE hotel_locations AS
2  SELECT
3      osm_id,
4      name,
5      tourism as category,
6      cidade_nome as city,
7      ST_X(ST_Transform(way, 4326)) AS longitude,
8      ST_Y(ST_Transform(way, 4326)) AS latitude
9  FROM
10     planet_osm_point
11  WHERE
12     tourism='hotel' AND cidade_nome IS NOT NULL
13  UNION ALL
14  SELECT
15     osm_id,
16     name,
17     tourism as category,
18     cidade_nome as city,
19     ST_X(ST_Transform(ST_Centroid(way), 4326)) as longitude,
20     ST_Y(ST_Transform(ST_Centroid(way), 4326)) as latitude
21  FROM
22     planet_osm_polygon
23  WHERE
24     tourism='hotel' AND cidade_nome IS NOT NULL;
```

Fonte: Do autor.