

PolRoute-DS: um *Dataset* de Dados Criminais para Geração de Rotas de Patrulhamento Policial*

Bruno Cunha Sá¹, Gustavo Muller¹, Maicon Banni¹, Wagner Santos^{1,2}
Marcos Lage¹, Isabel Rosseti¹, Yuri Frota¹, Daniel de Oliveira¹

¹Instituto de Computação - Universidade Federal Fluminense (IC/UFF)
Av. Gal. Milton Tavares de Souza, s/nº – Niterói – RJ – Brasil

² Polícia Militar do Estado do Rio de Janeiro (PMERJ)

{bcunha, gustavomuller, maiconbanni, wagnergs}@id.uff.br

{mlage, rosseti, yuri, danielcmo}@ic.uff.br

Resumo. *Esse artigo apresenta o dataset PolRoute-DS, criado para viabilizar o desenvolvimento e testes de abordagens de geração de rotas policiais em centros urbanos. O PolRoute-DS combina a estrutura espacial da cidade de interesse, representada como um grafo conexo e direcionado de segmentos de vias, com dados criminais obtidos de fontes públicas (no contexto deste artigo os dados são providos pela Secretaria da Segurança Pública de São Paulo). O PolRoute-DS se encontra disponível para uso da comunidade sob a licença Creative Commons By Attribution 4.0 International (versões CSV e PostgreSQL), e pode ser obtido em <https://osf.io/mxrqu/>.*

Abstract. *This paper introduces the PolRoute-DS dataset, created to foster the development and evaluation of police routing approaches in urban centers. The PolRoute-DS combines the spatial structure of the city of interest (in the context of this paper the city of São Paulo) represented as a connected and directed graph of street segments with criminal data obtained from public sources. PolRoute-DS is available for usage under the Creative Commons By Attribution 4.0 International license (CSV and PostgreSQL versions), and can be downloaded at <https://osf.io/mxrqu/>.*

1. Introdução

O conceito de *Cidades Inteligentes* (i.e., *Smart Cities*) vem ganhando bastante relevância na última década [Shapiro 2006]. Diversas iniciativas têm sido propostas nas mais variadas áreas e.g., saúde [Caban and Gotz 2015], transportes [Ota et al. 2017], planejamento urbano [Miranda et al. 2020, Miranda et al. 2017], etc. De acordo com o *Cities in Motion Index*, da *IESE Business School* [Berrone and Ricart 2020], uma cidade inteligente pode ser avaliada pelas políticas e soluções adotadas nas áreas de (i) Governança, (ii) Administração Pública, (iii) Planejamento Urbano, (iv) Tecnologia, (v) Meio Ambiente, (vi) Conexões Internacionais, (vii) Apoio Social, (viii) Capital Humano e (ix) Economia. Nesse artigo, focaremos na dimensão *Planejamento Urbano*, e em especial na área de Segurança Pública.

O combate à criminalidade nas cidades brasileiras é um desafio histórico, especialmente se considerarmos grandes centros urbanos como São Paulo, Rio de Janeiro, Fortaleza e

*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores ainda gostariam de agradecer ao CNPq e a FAPERJ por financiarem parcialmente esse trabalho.

Recife [Lourenço et al. 2018]. Nos últimos anos, impactado pela crise econômica enfrentada desde 2015, o índice de criminalidade nessas cidades tem apresentado uma trajetória inconstante, mas ascendente. Por exemplo, o total de delitos registrados na cidade de São Paulo¹ saiu de 152.481 no primeiro trimestre de 2018 para 155.307 no mesmo período de 2019. Em 2020, apesar da política de distanciamento social imposta pela pandemia da COVID-19, o número de delitos chegou a 158.063 no primeiro trimestre. Por fim, o índice caiu para 145.598 no três primeiros meses de 2021 (período ainda impactado pela pandemia). De fato, por conta da complexidade desses grandes centros urbanos (*e.g.*, disparidade populacional e socio-econômica das regiões da cidade, existência de áreas com ocupação informal de difícil acesso, disponibilidade de malhas viárias com topologias intrincadas, *etc.*), definir políticas públicas capazes de reduzir os índices criminais não é uma tarefa trivial.

Uma das políticas possíveis, e que tem sido adotada com bons resultados, é o policiamento ostensivo ou proativo. A ideia do policiamento ostensivo é tornar evidente para a população a atuação policial, seja pela presença de policiais em pontos estratégicos da cidade ou por meio da ronda de patrulhas policiais. Utilizando esta estratégia de policiamento, a autoridade policial tem como objetivo manter a ordem pública, evitando a ocorrência de crimes, contravenções ou infrações. No contexto de policiamento ostensivo, uma das tarefas mais complexas de serem realizadas é a definição de rotas das patrulhas policiais. Em geral, os recursos existentes são escassos, *i.e.*, a quantidade de viaturas e policiais disponíveis é bem menor do que a necessária. Assim, a autoridade policial deve definir *a priori* quais vias específicas serão cobertas pelas patrulhas. Esse tipo de definição de rotas acaba priorizando os chamados *Hot Spots* [Reis et al. 2006], *i.e.*, áreas da cidade em que o índice de criminalidade é mais alto. Dessa forma, considerando a escassez de recursos, a alocação eficiente de rotas para as viaturas se torna prioridade.

Apesar da definição das rotas (também chamado de roteamento) das patrulhas poder ser realizada de forma manual, esse processo pode se tornar laborioso e propenso a erros se realizado dessa forma. O cenário ideal é que a definição de rotas seja realizada de forma automática e baseada em dados históricos de índices criminais (também chamado de policiamento orientado a dados ou *Data-Driven Policing* [Dewinter et al. 2020]). De fato, há um grande volume de dados sobre crimes em regiões urbanas disponíveis para análise e uso, seja por institutos governamentais tais como a Secretaria da Segurança Pública de São Paulo (SSP-SP)² ou o Instituto de Segurança Pública (ISP)³ no Rio de Janeiro, ou ainda por *sites* que mapeiam crimes usando *crowdsourcing* como o *Onde fui Roubado*⁴ e WikiCrimes⁵.

A automatização do roteamento de patrulhas pode utilizar diferentes estratégias computacionais, como por exemplo técnicas de Otimização [Saint-Guillain et al. 2021]. De fato, os autores deste artigo vêm desenvolvendo uma solução para este problema utilizando Metaheurísticas⁶ como o GRASP [Resende and Ribeiro 1997]. Para que este tipo de solução possa ser construída, uma base de dados com as quantidades de ocorrências de crimes por tipo deve estar disponível e as ocorrências devem estar associadas as vias das cidades. Dessa forma, representamos as vias da cidade por meio de um grafo $G^c = (V, E, Q)$. Nesta representação, o conjunto $E = E^1 \cup E^2$ é composto de segmentos de vias, porém repre-

¹<http://www.ssp.sp.gov.br/Estatistica/Trimestrais.aspx>

²<http://www.ssp.sp.gov.br/>

³<http://www.isp.rj.gov.br/>

⁴<https://www.ondefuiroubado.com.br/>

⁵<http://www.wikicrimes.org/>

⁶Os detalhes desta proposta fogem ao escopo deste artigo.

sentam segmentos de mão única (*i.e.*, $E^1 = E^2$) e mão dupla ($E^1 \neq E^2$). Reforçamos que o grafo G^c é composto por segmentos de vias, *i.e.*, dividimos todas as vias da cidade em segmentos de comprimento entre 150 e 200 metros. Esta estratégia evita que vias longas sejam associadas a índices criminais muito altos (destoantes do resto do grafo) e/ou que rotas muito compridas sejam artificialmente construídas em resposta a um grupo de ocorrências associado a um único ponto de uma via grande. Além disso, o grafo G^c é formado pelo conjunto de vértices V de ordem $n = |V|$ que representa esquinas (conexões entre duas vias) ou junções de segmentos de uma mesma via. A Figura 1 apresenta um exemplo de grafo gerado para a região em torno do Museu de Arte de São Paulo Assis Chateaubriand (MASP) na Avenida Paulista em São Paulo.

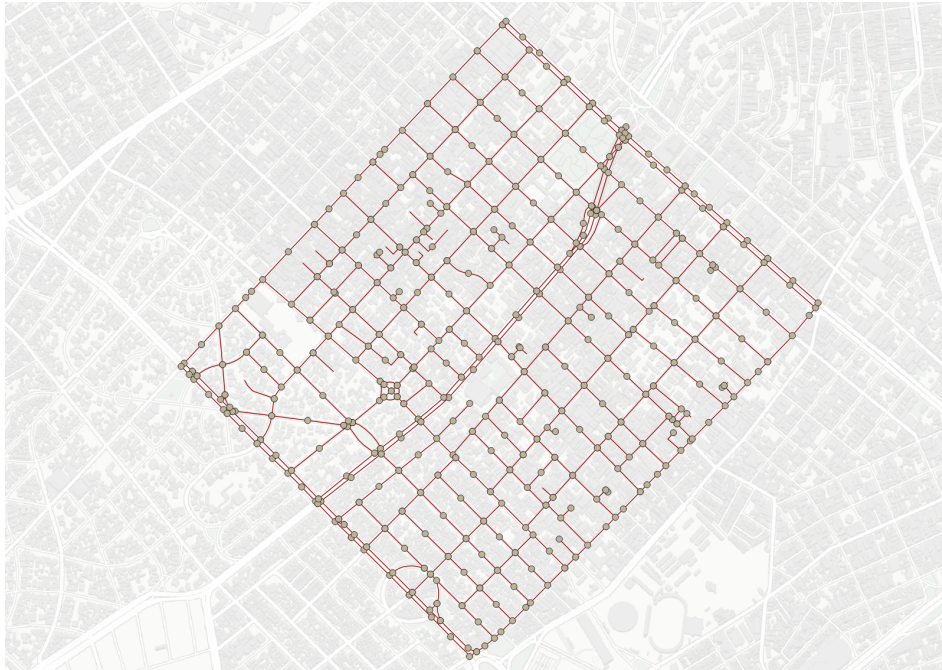


Figura 1. Grafo representando a região em torno do Museu de Arte de São Paulo Assis Chateaubriand (MASP) na Avenida Paulista.

Neste artigo apresentamos o *dataset* intitulado `PolRoute-DS`, criado para viabilizar o desenvolvimento e testes de abordagens de geração de rotas policiais em centros urbanos. O `PolRoute-DS` combina a estrutura espacial da cidade de interesse (*i.e.*, um grafo conexo de segmentos de vias) com índices criminais obtidos de fontes públicas. O grafo é representado em um banco de dados relacional que segue o esquema floco-de-neve tradicionalmente utilizado no desenvolvimento de *Data Warehouses* [Inmon 1996]. O objetivo desta abordagem é permitir a representação dos valores totais de cada tipo de crime por segmento e o cálculo de agregações na dimensão temporal. Estas operações são fundamentais não só para algoritmos de roteamento de viaturas, mas também para aplicações de análise dados tradicionais. Na versão atual, o `PolRoute-DS` contém os dados da cidade de São Paulo.

O restante deste artigo está organizado da seguinte forma: Na Seção 2, discutimos os trabalhos relacionados. Na Seção 3, apresentamos a estrutura do `PolRoute-DS` e seu processo de geração. Na Seção 3.3, analisamos algumas estatísticas do *dataset*. Na Seção 3.4 é apresentado um exemplo de uso do *dataset*. Na Seção 3.5 apresentamos instruções de *download* e citação do *dataset*. Por fim, a Seção 4 apresenta algumas considerações finais.

2. Trabalhos Relacionados

Alguns trabalhos relacionados a este artigo descrevem a estruturação e disponibilizam publicamente *datasets* semelhantes ao PolRoute-DS.

Os autores de [Yoo 2019] discutem como materializar, em um único banco de dados, informações espaciais (*e.g.*, mapa de uma cidade) e de ocorrências criminais que possam ser úteis para construir soluções de combate a violência. Assim como o PolRoute-DS, a solução proposta no trabalho utiliza como base os conceitos de *Data Warehouse*. Já o *Crime Data-Warehouse*, da *Royal Canadian Mounted Police*⁷, reúne em um *Data Warehouse* múltiplos *datasets* sobre eventos de crime relatados oficialmente para as autoridades policiais da cidade de Burnaby, no Canadá. Esse *dataset* possui aproximadamente 4,4 milhões de eventos descritos por atributos como a data, hora, local e informações do criminoso.

Usando uma abordagem baseada em redes complexas, [Spadon et al. 2017] propõe métricas para detectar regiões com alto índice de criminalidade em uma cidade. Para avaliar sua proposta, os autores utilizam um *dataset* chamado G-FranC [Scabora et al. 2019] que possui a estrutura da rede complexa combinada com dados de crimes ocorridos na cidade de São Francisco, nos Estados Unidos. Assim como no PolRoute-DS, o G-FranC propõe a representação das vias da cidade como um grafo. Entretanto o G-FranC não leva em consideração a mão das vias (*i.e.*, se são de mão única ou mão dupla), já que seu objetivo é identificar regiões com alto índice de criminalidade e não definir rotas para patrulhamento.

3. Apresentação do PolRoute-DS

Conforme mencionado anteriormente, o PolRoute-DS é criado a partir da junção espacial da geometria da malha viária da cidade com dados de ocorrências de crimes, disponibilizados por fontes externas. Em sua versão atual, o PolRoute-DS contém dados da cidade de São Paulo. A seguir, apresentamos o processo de aquisição e transformação dos dados, o esquema do PolRoute-DS, estatísticas do *dataset* e exemplos de uso.

3.1. Processo de Aquisição e Transformação dos Dados

Para a criação do PolRoute-DS, utilizamos o arcabouço apresentado na Figura 2. O processo se inicia (Passo ❶ na Figura 2) com a importação dos dados a partir de fontes públicas de informação. O componente ETL (*Extract, Transform, Load*) tem a responsabilidade de realizar o *download* de arquivos ou acessar *APIs* públicas para obter dos dados. Na versão atual do PolRoute-DS utilizamos os dados de crimes disponibilizados pela Secretaria da Segurança Pública de São Paulo (SSP-SP)⁸. O *download* dos dados utilizados para a construção do PolRoute-DS foi realizado em Agosto de 2019. Os dados obtidos no *site* da SSP-SP são disponibilizados em arquivos XLS organizados por ano e tipo de crime. No escopo deste trabalho, foram considerados os seguintes tipos de crime: (i) Femicídio, (ii) Furto de Celular, (iii) Roubo de Celular, (iv) Homicídio Doloso, (v) Latrocínio, (vi) Furto de Veículo e (vii) Roubo de Veículo. Apesar de estarem disponíveis, não foram utilizados os dados de (i) Morte em decorrência de intervenção policial, (ii) Morte suspeita e (iii) Registro de óbitos. A quantidade total de arquivos obtidos após *download*, o volume de dados e a janela de tempo considerada para cada tipo de crime são apresentados na Tabela 1.

Todos os arquivos obtidos possuem o mesmo conjunto de atributos. Neste trabalho consideramos somente os atributos Data da Ocorrência, Hora da Ocorrência, Período da

⁷<https://www.rcmp-grc.gc.ca/en/evidence-and-reports-data-warehouse>

⁸<http://www.ssp.sp.gov.br/transparenciassp/Consulta.aspx>

Tabela 1. Informações dos arquivos obtidos.

Tipo	Nº de Arquivos	Tamanho	Janela de Tempo
Feminicídio	52	1,2 MB	04/2015 - 07/2019
Furto de Celular	115	1,3 GB	01/2010 - 07/2019
Roubo de Celular	115	2,2 GB	01/2010 - 07/2019
Homicídio Doloso	198	126,6 MB	01/2003 - 07/2019
Latrocínio	196	7,6 MB	04/2003 - 07/2019
Furto de Veículo	199	1,6 GB	01/2003 - 07/2019
Roubo de Veículo	199	1,9 GB	01/2003 - 07/2019

Ocorrência (*e.g.*, Manhã, Tarde, Noite, Madrugada), Logradouro, Número, Latitude, Longitude, Cidade e Estado. Em uma etapa de pre-processamento, todos os arquivos baixados foram condensados em um único arquivo contendo todas as ocorrências de todos os tipos de crimes. Para isso, um novo atributo relativo ao Tipo de Crime foi adicionado (essa informação podia ser extraída do título de cada arquivo). Além disso, foram substituídos campos vazios (;;) por uma *string* para identificação de valor nulo (;NULL;). Uma vez formatado, o arquivo foi carregado para um banco de dados no PostgreSQL, cujo esquema se assemelha a um *Data Mart* de crimes (que será apresentado na Subseção 3.2).

Uma vez que os dados de ocorrências de crimes já se encontram disponíveis no *Data Mart*, se inicia a criação do grafo de vias da cidade (Passo 2 na Figura 2). Essa etapa é executada com o apoio do *software* QGIS. O QGIS é uma ferramenta de Sistema de Informação Geográfica (GIS) que permite a visualização, edição e análise de dados georreferenciados em múltiplas camadas (*layers*). Utilizamos o QGIS para particionar as vias em segmentos de tamanho que variam de 150 a 200 metros (já que dependendo do comprimento da via, a divisão em segmentos de mesmo tamanho não é possível). Em seguida, os dados de ocorrência de crimes foram carregados no *software* para serem combinados com as arestas do grafo. Neste processo, foram descartadas tuplas com latitude/longitude fora da cidade de São Paulo (*e.g.*, existiam ocorrências localizadas na Europa ou no oceano Atlântico).

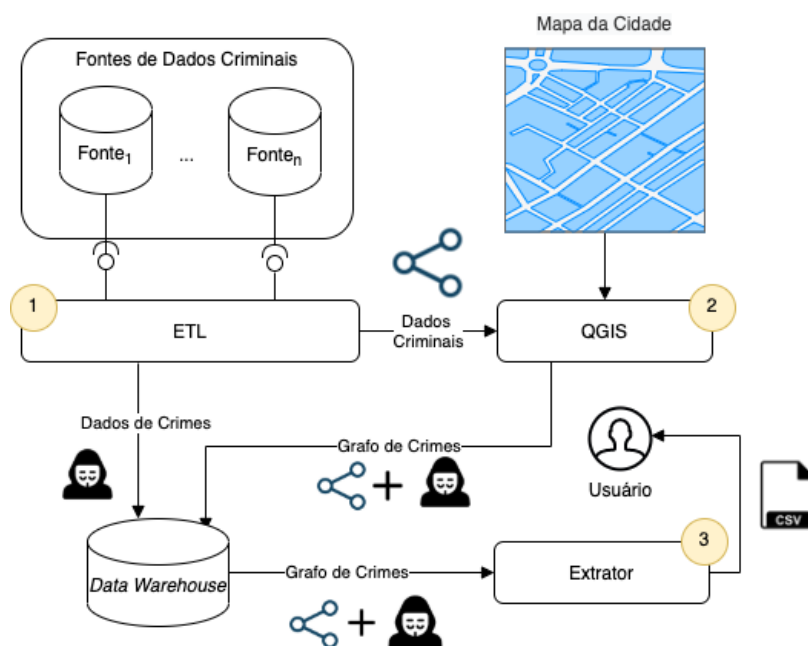


Figura 2. Processo de Aquisição dos Dados do PolRoute-DS.

Para cada ocorrência de crime encontrada no *Data Mart*, foi executado um *Spatial Join* no QGIS com o grafo das vias, usando a operação *Join Attributes by Nearest*. Essa operação requer duas camadas como entrada, uma com os dados de ocorrências e outra com os dados de vias. Para cada *feature* da camada de ocorrências, as *features* mais próximas da camada de vias são identificadas. A saída do algoritmo é composta pelas geometrias das *features* da camada de vias associadas aos atributos das *features* da camada de ocorrências. Dada uma via *E*, apenas crimes que ocorreram a uma distância de até 0.001 graus decimais (*i.e.*, aproximadamente 100 metros) de *E* são associados a via. Feita a associação, a quantidade de ocorrências por segmento é calculada. O grafo resultante (chamado deste ponto em diante de *grafo de crimes*) é então armazenado no banco de dados. Finalmente, no Passo ③, o componente extrator submete consultas ao banco de dados e gera arquivos CSV contendo os dados armazenados, de forma que possam ser facilmente importados em ferramentas existentes. Esses arquivos podem ser usados em abordagens de roteamento baseado em *hot spots* como a que nosso grupo vem desenvolvendo.

3.2. Esquema do Banco de Dados

Após as etapas de ETL, que incluem a importação, formatação e mapeamento da quantidade de crimes por segmento do grafo, os dados são inseridos em um banco de dados modelado como um *Data Mart* que segue o esquema floco de neve [Inmon 1996]. A Figura 3 apresenta o esquema do banco de dados em questão. O esquema é composto por sete tabelas: (i) Crime, (ii) Segmento, (iii) Vértice, (iv) Tempo, (v) Distrito, (vi) Zona, e (vii) Bairro.

A tabela Crime armazena as quantidades de crimes associadas a um determinado segmento do mapa em um determinado instante ou período de tempo. A tabela possui nove atributos: *segmento_id* (chave primária e chave estrangeira) que caracteriza a qual segmento as quantidades de crimes se referem, *tempo_id* (chave primária e chave estrangeira) que define a qual data as quantidades se referem, *total_feminicidio* que define a quantidade de ocorrências de feminicídio, *total_homicidio* que representa o total de homicídios, *total_latrocinio* que representa o total de ocorrências de latrocínio, *total_furto_celular* que representa o total de furtos de celular, *total_roubo_celular* que representa o total de roubos de celular, *total_furto_automovel* que representa o total de furtos de veículos e *total_roubo_automovel* que representa o total de roubos de veículos. Ressaltamos que a tabela crime é capaz de armazenar dados pré-agregados no tempo, de forma a acelerar as consultas a serem realizadas.

A tabela Tempo armazena os períodos de tempos com ocorrências dos crimes em várias resoluções temporais. Ela é composta pelos atributos *tempo_id* (chave primária), *dia*, *mes*, *ano*, *dia_de_semana* (booleano que indica se uma determinada data é dia de semana ou final de semana) e o *período* (manhã, tarde, noite ou madrugada). Sendo assim, a tabela de Tempo pode armazenar uma data/hora específica em que um crime aconteceu e também representar um mês do ano para o qual se pretende saber o total de ocorrências de crime.

A tabela Segmento armazena os segmentos que fazem parte do grafo de vias. Os segmentos são associados a índices criminais, através da tabela Crime. A tabela segmento é composta pelos atributos *id_segmento* (chave primária), *vertice_inicio_id* e *vertice_fim_id* (chaves estrangeiras) que representam os vértices inicial e final do segmento, *geometria* (representação interna do QGIS para o segmento), *mao_unica* que informa se a via é de mão única ou dupla e *comprimento* que representa o tamanho de um segmento.

A tabela Zona representa áreas que o usuário pode demarcar no mapa para uso em abordagens de roteamento. A zona não é uma organização oficial da cidade, e sim definida pelo usuário de acordo com a necessidade. Ela é composta por *zona_id* (chave primária),

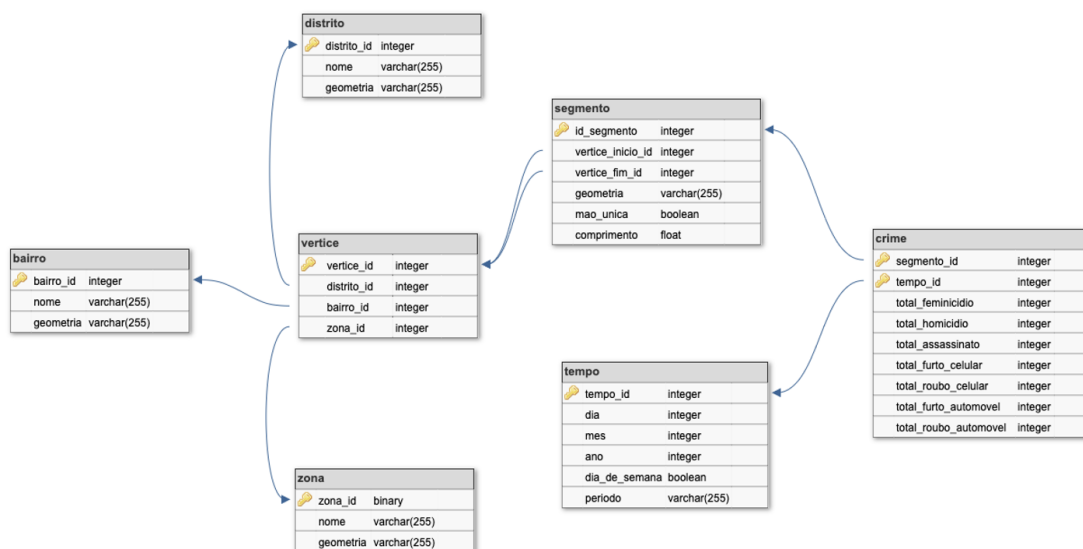


Figura 3. Esquema do PolRoute-DS.

nome e *geometria*. A tabela Distrito define os distritos de uma cidade. Ela é composta pelos atributos *distrito_id* (chave primária), *nome* e *geometria*. A tabela Bairro representa os bairros ou vizinhanças oficiais de uma cidade. Ela é composta pelos atributos *bairro_id* (chave primária), *nome* e *geometria*. Finalmente, a tabela Vertice representa os vértices do grafo de vias. Ela é composta pelos atributos *vertice_id* (chave primária), *nome* e *geometria*. Além disso, ela possui chaves estrangeiras *distrito_id*, *bairro_id* e *zona_id* que referenciam as tabelas distrito, bairro e zona, respectivamente.

3.3. Estatísticas do PolRoute-DS

Nesta subseção apresentamos algumas estatísticas da versão atual do PolRoute-DS. A Figura 4 apresenta a quantidade de ocorrências por tipo de crime ($\times 1.000$). É importante ressaltar que nem todas as ocorrências que foram obtidas no *site* da SSP-SP foram utilizadas na construção do PolRoute-DS, pois muitas ocorrências não possuíam informações de localização que permitissem associar a mesma a um segmento do grafo de vias. Assim, a Figura 4 apresenta o *Total* de ocorrências obtidas e as a quantidade de ocorrências que foram *Úteis* para a construção do PolRoute-DS.

Além da quantidade de ocorrências utilizadas na construção do PolRoute-DS, apresentamos na Tabela 2 a quantidade de tuplas em cada uma das tabelas do modelo de dados. É importante reforçar que por seguir o esquema floco de neve de modelagem dimensional, o PolRoute-DS já possui diversas tuplas agregadas no tempo, *e.g.*, para um determinado segmento já possuímos o total de crimes por dia, mês, ano pré-agregados, facilitando tanto a consulta, quanto a geração da entrada para a heurística de roteamento que vem sendo desenvolvida.

3.4. Exemplo de Uso - Geração de Rotas

Nesta seção apresentamos um exemplo de uso para ilustrar o potencial do PolRoute-DS. O exemplo se encontra no contexto de geração de rotas de patrulhamento policial por meio de metaheurísticas, e que consomem os dados do PolRoute-DS. Tomamos como exemplo

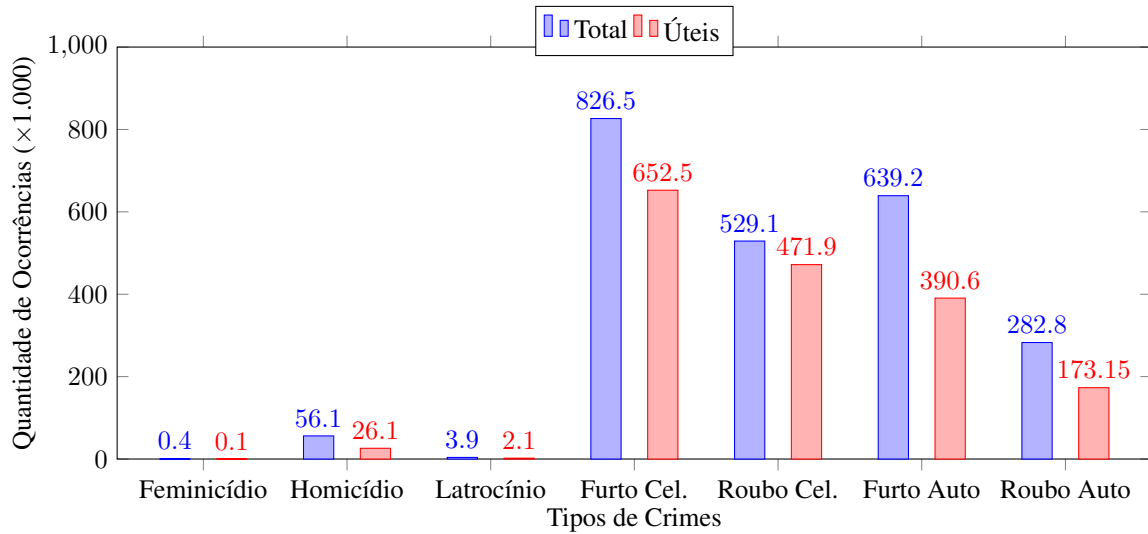


Figura 4. Quantidade de Ocorrências por Tipo de Crime (x1.000).

o distrito de Sacomã, na cidade de São Paulo. Ele é dividido entre os bairros de Anchieta, Moinho Velho e São João Clímaco. Na Figura 5(a), definimos 2 zonas no distrito. Essas zonas definem regiões em que as rotas serão construídas. Existem dois tipos de rotas: intra-zona e inter-zona. A partir desse momento, o mapa com o índice criminal de cada segmento pode ser visualizado (Figura 5(b)). Os segmentos em tons mais fortes de vermelho são aqueles com maiores valores para o índice criminal. Nesse exemplo, calculamos o índice criminal de acordo com a seguinte fórmula:

$$I_{Criminal} = (Total_{Feminicidio} \times 10 + Total_{Homicidio} \times 10 + Total_{Latrocínio} \times 10 + Total_{RouboCel.} \times 5 + Total_{RouboAuto} \times 5 + Total_{FurtoCel.} \times 1 + Total_{FurtoAuto} \times 1) / 42$$

É importante ressaltar que os pesos foram atribuídos empiricamente com o objetivo de atribuir maior importância para crimes em que houvesse ameaça à vida. É importante ressaltar que o usuário pode definir diferentes pesos de acordo com sua necessidade. Ainda, o usuário pode considerar diferentes períodos de tempo, pois determinadas zonas podem ser mais perigosas na data presente do que eram em períodos passados. O usuário pode ainda visualizar quais segmentos serão considerados por zona (Figura 5(c)). Finalmente, de posse das informações de zona e dos segmentos em questão (com seus índices criminais associados), a heurística pode definir as melhores rotas de acordo com os recursos informados. No caso do exemplo foram definidos sete policiais fixos (*pins* azuis na Figura 5(d)) e oito viaturas para realizarem patrulhas de no máximo 2,4 km de distância percorrida. Na Figura 5(d) podemos visualizar as rotas produzidas pela heurística. É importante ressaltar que o PolRoute-DS

Tabela 2. Quantidade de Tuplas nas Tabelas do Esquema.

Tabela	Quantidade de Tuplas
Crime	7.267.353
Tempo	28.593
Zona	115
Segmento	226.975
Vértice	180.598
Distrito	97
Vizinhança	323

pode ser usado por qualquer abordagem de roteamento para segurança pública, e não somente pela abordagem que vem sendo desenvolvida pelo nosso grupo.

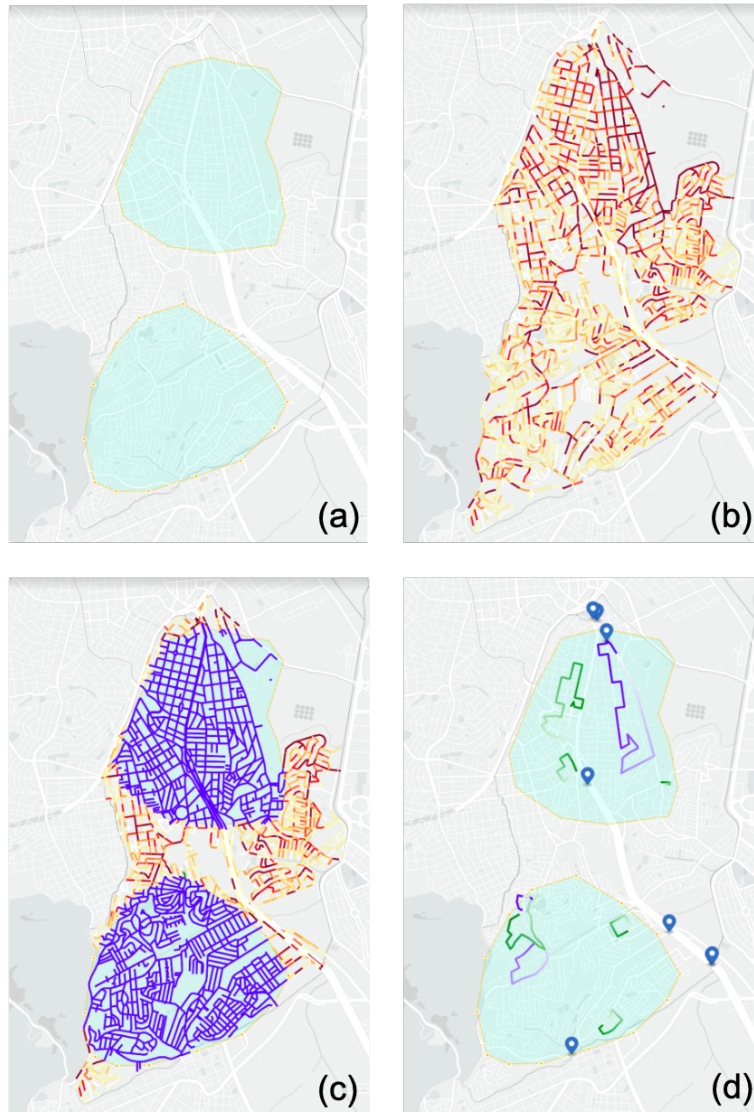


Figura 5. (a) Definição das Zonas; (b) Visualização dos índices de criminalidade por segmento; (c) Visualização dos segmentos cobertos por cada zona; (d) Rotas geradas pela abordagem de roteamento

3.5. Download e Citação

O `PolRoute-DS` (versão CSV e `dump` do PostgreSQL) e os `scripts` de geração do QGIS podem ser obtidos em <https://osf.io/mxrgu/>. O conjunto de dados está disponível para uso da comunidade sob a licença *Creative Commons By Attribution 4.0 International*. O uso do `PolRoute-DS` requer citação para o DOI do *dataset* (10.17605/OSF.IO/MXRGU) e para o presente artigo.

4. Considerações Finais

Neste artigo, apresentamos o *dataset* `PolRoute-DS`, que combina a geometria da malha viária da cidade de São Paulo (representada como um grafo) com dados de ocorrências de crimes, disponibilizados por fontes externas (*e.g.*, SSP-SP). O grafo de crimes gerado pode ser

usado como entrada para abordagens de roteamento baseadas em dados criminais, em especial aquelas baseadas em *hot spots*. A contribuição desse artigo se encontra no detalhamento do processo de ETL e na disponibilização do *dataset* para uso público. O PolRoute-DS pode ser obtido em <https://osf.io/mxrqu/>.

Apesar de apresentar um período amplo de abrangência (de 2003 até 2019), podemos considerar como limitação do *dataset* o fato dos dados finalizarem no ano de 2019. Entretanto, cabe salientar que as ocorrências relativas aos anos de 2020 e 2021 podem ser carregadas *a posteriori*. Como trabalhos futuros, pretendemos expandir o Polroute-DS para considerar dados de regiões dominadas por facções criminosas, que embora apresentem poucas ocorrências oficiais de crimes, são regiões bastante perigosas e que necessitam de patrulhamento ostensivo. Ainda pretendemos incluir dados pluviométricos e de ruas bloqueadas, de forma que essa informação possa ser usada na geração de rotas (*e.g.*, não faz sentido propor uma rota por uma via que se encontra bloqueada no momento).

Referências

- Berrone, P. and Ricart, J. (2020). Iese cities in motion index 2020. *IESE Business School University of Navarra: Barcelona, Spain*.
- Caban, J. J. and Gotz, D. (2015). Visual analytics in healthcare—opportunities and research challenges.
- Dewinter, M., Vandeviver, C., Vander Beken, T., and Witlox, F. (2020). Analysing the police patrol routing problem: A review. *ISPRS International Journal of Geo-Information*, 9(3).
- Inmon, W. H. (1996). The data warehouse and data mining. *Commun. ACM*, 39(11):49–50.
- Lourenço, V., Mann, P., Guimaraes, A., Paes, A., and de Oliveira, D. (2018). Towards safer (smart) cities: Discovering urban crime patterns using logic-based relational machine learning. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8. IEEE.
- Miranda, F., Doraiswamy, H., Lage, M., Zhao, K., Gonçalves, B., Wilson, L., Hsieh, M., and Silva, C. T. (2017). Urban pulse: Capturing the rhythm of cities. *IEEE Trans. Vis. Comput. Graph.*, 23(1):791–800.
- Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., and Silva, C. T. (2020). Urban mosaic: Visual exploration of streetscapes using large-scale image data. In Bernhaupt, R., Mueller, F. F., Verweij, D., Andres, J., McGrenere, J., Cockburn, A., Avellino, I., Goguy, A., Bjørn, P., Zhao, S., Samson, B. P., and Kocielnik, R., editors, *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, pages 1–15. ACM.
- Ota, M., Vo, H. T., Silva, C. T., and Freire, J. (2017). Stars: Simulating taxi ride sharing at scale. *IEEE Trans. Big Data*, 3(3):349–361.
- Reis, D., Melo, A., Coelho, A. L. V., and Furtado, V. (2006). Towards optimal police patrol routes with genetic algorithms. In Mehrotra, S., Zeng, D. D., Chen, H., Thuraisingham, B. M., and Wang, F., editors, *Intelligence and Security Informatics, IEEE International Conference on Intelligence and Security Informatics, ISI 2006, San Diego, CA, USA, May 23-24, 2006, Proceedings*, volume 3975 of *Lecture Notes in Computer Science*, pages 485–491. Springer.
- Resende, M. G. C. and Ribeiro, C. C. (1997). A GRASP for graph planarization. *Networks*, 29(3):173–189.

- Saint-Guillain, M., Paquay, C., and Limbourg, S. (2021). Time-dependent stochastic vehicle routing problem with random requests: Application to online police patrol management in brussels. *Eur. J. Oper. Res.*, 292(3):869–885.
- Scabora, L. d. C., Spadon, G., Rodrigues, L. S., Cazzolato, M. T., Araújo, M. V. d. S., Sousa, E. P. M. d., Traina, A. J. M., Rodrigues Junior, J. F., and Traina Junior, C. (2019). G-franc: a dataset of criminal activities mapped as a complex network in a relational dbms. In *Brazilian Symposium on Databases - SBBD*, Fortaleza. SBC.
- Shapiro, J. M. (2006). Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital. *The Review of Economics and Statistics*, 88(2):324–335.
- Spadon, G., Scabora, L. C., Oliveira, P. H., Araujo, M. V. S., Machado, B. B., de Sousa, E. P. M., Jr., C. T., and Jr., J. F. R. (2017). Behavioral characterization of criminality spread in cities. In Koumoutsakos, P., Lees, M., Krzhizhanovskaya, V. V., Dongarra, J. J., and Sloot, P. M. A., editors, *International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland*, volume 108 of *Procedia Computer Science*, pages 2537–2541. Elsevier.
- Yoo, J. S. (2019). Crime data warehousing and crime pattern discovery. In *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, DATA '19*, New York, NY, USA. Association for Computing Machinery.