

## Rotas Veiculares Cientes de Contexto: Arcabouço e Análise Usando Dados Oficiais e Sensoriados por Usuários sobre Crimes

Frances A. Santos<sup>1</sup>, Diego O. Rodrigues<sup>1</sup>, Thiago H. Silva<sup>2</sup>,  
Antonio A. F. Loureiro<sup>3</sup>, Leandro A. Villas<sup>1</sup>

<sup>1</sup>Instituto de Computação, Universidade Estadual de Campinas. Campinas, Brasil

<sup>2</sup>Dep. de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, Brasil

<sup>3</sup>Dep. de Ciência da Computação, Univ. Federal de Minas Gerais. Belo Horizonte, Brasil

**Abstract.** *An increasing number of users have been adopting route recommendation systems, many of them motivated by the convenience that these systems bring to their traffic experiences. These systems observe the current traffic conditions in order to evaluate the fastest path. However, contextual information, such as pollution levels, weather conditions and security of the route, might not be taken into account in the recommendation process. With this in mind, we propose a framework to support context-aware route recommendation systems. Furthermore, we present possible approaches for development of two key components of this framework: (1) identification of contextual areas based on different urban data sources; and (2) identification of typical routes. The proposed framework might improve existing recommendation algorithms, or also enable the proposal of new ones. To validate this hypothesis, we used a set of routes suggested by Google Maps in the city of Curitiba to identify frequent patterns on these routes. Afterwards, some insecure zones were identified in Curitiba using data provided by the government of the city and also data generated by the citizens via their mobile devices. The obtained results showed the existence of an opportunity for route planners to provide differential services to users who desire it, which is an important step towards the development of context-aware vehicular networks.*

**Resumo.** *Cada vez mais usuários utilizam sistemas de recomendação de rotas, pois esse serviço torna a vida no trânsito mais prática. Tipicamente esses serviços levam em consideração a situação atual do trânsito para fornecer a rota mais rápida. No entanto, requisitos contextuais, como nível de poluição, condições climáticas e segurança da rota, podem não ser levados em consideração na recomendação. Nessa direção, propomos um arcabouço para apoiar o oferecimento de recomendação de rotas veiculares cientes de contexto. Além disso, apresentamos possíveis abordagens para a implementação de dois componentes chave desse arcabouço: (1) identificação de áreas contextuais em fontes de dados urbanos distintas; e (2) identificação de rotas frequentes. O arcabouço proposto possibilita o aprimoramento de algoritmos de recomendação existentes, bem como a proposição de novos. Com o intuito de estudar essa hipótese, utilizamos um conjunto de rotas sugeridas pelo Google Maps na cidade de Curitiba para identificar padrões frequentes nessas rotas. Em seguida, identificamos áreas de insegurança nessa mesma cidade utilizando dados abertos fornecidos pela prefeitura e dados sensoriados pelos usuários sobre ocorrências de crimes. Nossos resultados indicam que existe uma oportunidade para recomendadores de rotas oferecem serviços diferenciados para usuários que assim desejarem, sendo um passo importante para a construção de uma rede veicular sensível ao contexto de rotas.*

## 1. Introdução

Sistemas de recomendação de rotas estão sendo cada vez mais utilizados pelos motoristas. Alguns dos motivos que explicam essa tendência são a praticidade proporcionada e a possibilidade de auxiliar em problemas de congestionamento no trânsito, que têm se agravado em grandes cidades [Wang et al. 2014, Vo 2015]. De modo geral, esses serviços visam permitir aos usuários encontrarem a melhor rota para atingir um determinado destino a partir de um determinado ponto da cidade. Com isso, uma pergunta importante que surge é: como definir a melhor rota de um ponto a outro na cidade?

Comumente, os sistemas de recomendação de rotas veiculares, tais como, Waze<sup>1</sup> e Google Maps<sup>2</sup>, consideram as condições históricas e atuais de trânsito para recomendação de rotas mais rápidas. No entanto, outros requisitos contextuais, como o nível de poluição, condições climáticas, eventos e segurança da rota, podem não ser levados em consideração na recomendação. Por exemplo, em 2015 um casal seguiu uma recomendação de rota fornecida pelo Waze e acabaram sendo baleados ao entrar na área da comunidade do Caramujo em Niterói, RJ [Rio 2015]. Episódios como esse mostram a importância da consideração de contexto inerente às rotas além de aspectos tradicionais, como velocidade atual da via, para fornecer *rotas veiculares cientes de contexto*. Desta forma, os usuários podem escolher uma rota considerando diferentes interesses pessoais.

Nessa direção, propomos um arcabouço que visa auxiliar no processo de recomendação de rotas veiculares cientes de contexto. O arcabouço é composto por dois componentes chave: (i) identificação de áreas contextuais em fontes de dados urbanos distintas; e (ii) identificação de rotas frequentes. Além de discutir esses componentes, também são apresentadas possíveis abordagens para as suas implementações. Por meio do arcabouço proposto, torna-se possível o aprimoramento de algoritmos de recomendação existentes, bem como a proposição de novos. Com o intuito de estudar essa hipótese, assim como ilustrar uma instanciação do arcabouço, foram coletados dados oficiais e sensoriados por usuários sobre incidências de crime em Curitiba, Brasil, e também rotas sugeridas pelo Google Maps nessa cidade. Observamos indícios de que quesitos de segurança não são considerados na recomendação de rotas, pois rotas que são frequentemente recomendadas passam por áreas reconhecidas por sua insegurança. Diante disso, verificamos que rotas alternativas, que não passam por essas áreas, não adicionam atrasos excessivos. Nossos resultados indicam que existe uma oportunidade para recomendadores de rotas oferecerem serviços diferenciados para usuários que assim desejarem, sendo um passo importante para a construção de uma rede veicular sensível ao contexto de rotas.

O restante deste trabalho está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve o arcabouço proposto. A Seção 4 apresenta as bases de dados analisadas. A Seção 5 investiga o emprego do arcabouço em um cenário com dados reais. Por fim, a Seção 6 apresenta as conclusões e trabalhos futuros do estudo.

## 2. Trabalhos Relacionados

Existem diversas iniciativas que extraem informações contextuais da cidade a partir de fontes de dados heterogêneos para fornecer serviços mais especializados aos cidadãos, visando melhorar a qualidade de vida na cidade. Exemplos desses serviços incluem segurança [Lee et al. 2006, Calavia et al. 2012], sistemas de energia [Vaghefi et al. 2014, Jain et al. 2016] e planejamento de tráfego [Demiryurek et al. 2009, Melnikov et al. 2015].

---

<sup>1</sup>www.waze.com.

<sup>2</sup>maps.google.com.br.

Particularmente em relação ao planejamento de tráfego, um tópico que vem recebendo grande atenção dos pesquisadores é o de recomendação de rotas para veículos ou mesmo pedestres [Wang et al. 2014, Dai et al. 2015]. Para esse fim, são propostas várias abordagens para obter o “melhor” caminho a ser utilizado pelo usuário, tais como [ter Mors et al. 2010, Bader et al. 2011], onde uma das principais premissas adotadas é que o usuário deseja economizar tempo em sua viagem. Contudo, encontrar o *melhor caminho* é um problema que vai além de encontrar o *caminho mais rápido*. Isto porque o caminho mais rápido pode passar por vias inseguras ou com muita poluição do ar, o que pode resultar em experiências desagradáveis aos usuários.

Desta forma, além de reduzir o tempo gasto no trânsito, aplicações de recomendação de rotas também podem ter outros objetivos que dependem dos interesses dos usuários. O CrowdSafe [Shah et al. 2011], por exemplo, é um sistema para dispositivos móveis que permite aos usuários reportarem informações sobre crimes que eles sofreram ou presenciaram. Tais informações, são utilizadas para o sistema identificar áreas de insegurança na cidade, fornecer um serviço de recomendação de rotas ciente dos locais de insegurança e, também, para geração de estatísticas de crimes da cidade. O mapeamento de insegurança não difere entre as possíveis categorias de crime (e.g., furto, arrombamento, assalto, homicídio, etc) e consequentemente, pode não refletir o grau de segurança requerido. Além disso, quando a rota mais segura não for distinta da menor rota, ou ainda, quando os usuários não requerem rotas mais seguras, então a menor rota é sempre recomendada, o que pode gerar congestionamentos e outros problemas inerentes.

A abordagem proposta por [Quercia et al. 2014], para recomendação baseada em informações contextuais, considera a percepção das pessoas sobre as rotas na cidade, a fim de recomendar rotas mais agradáveis. A participação ativa das pessoas, durante todas as etapas do sistema de recomendação, é um ponto crítico dessa abordagem e, por isso, manter os usuários motivados é essencial para o sucesso desse sistema. Em [De Domenico et al. 2015], a mobilidade dos agentes (e.g., veículos) é modelada como um sistema de partículas, onde informações de diferentes contextos da cidade tais como, tráfego, poluição, crimes e eventos, representam forças exercidas sobre as partículas. Nesse modelo, os agentes são partículas que devem se mover entre pares origem-destino em uma matriz, que representa uma área geográfica de interesse. A resultante das forças de atração/repulsão deve guiar a partícula pela matriz. Como as partículas que representam os indivíduos se movem por um espaço 2D livre, i.e., sem barreiras, esse modelo não considera as limitações da malha de ruas e avenidas existente nas cidades.

Este trabalho diferencia dos anteriores por propor um arcabouço capaz de identificar padrões em rotas frequentemente utilizadas por pessoas no mundo real, utilizando contextos, ao invés de inferir melhores rotas baseando-se na rota mais curta. Como [De Domenico et al. 2015], o arcabouço proposto possibilita a identificação de áreas contextuais em fontes de dados urbanos. Contudo, realizamos a análise agregada dos dados urbanos para minimizar erros e enriquecer a qualidade da informação obtida.

### **3. Arcabouço para Apoiar a Recomendação de Rotas Cientes de Contexto**

#### **3.1. Visão geral**

A ampla disponibilidade de fontes de dados sobre diversos aspectos da cidade permite entender e tratar problemas enfrentados pelos centros urbanos e, com isso, oferecer serviços mais sofisticados que visam melhorar a vida dos usuários na cidade. Dessa forma, estudamos uma abordagem para o oferecimento de recomendação de rotas veiculares cientes de contexto.

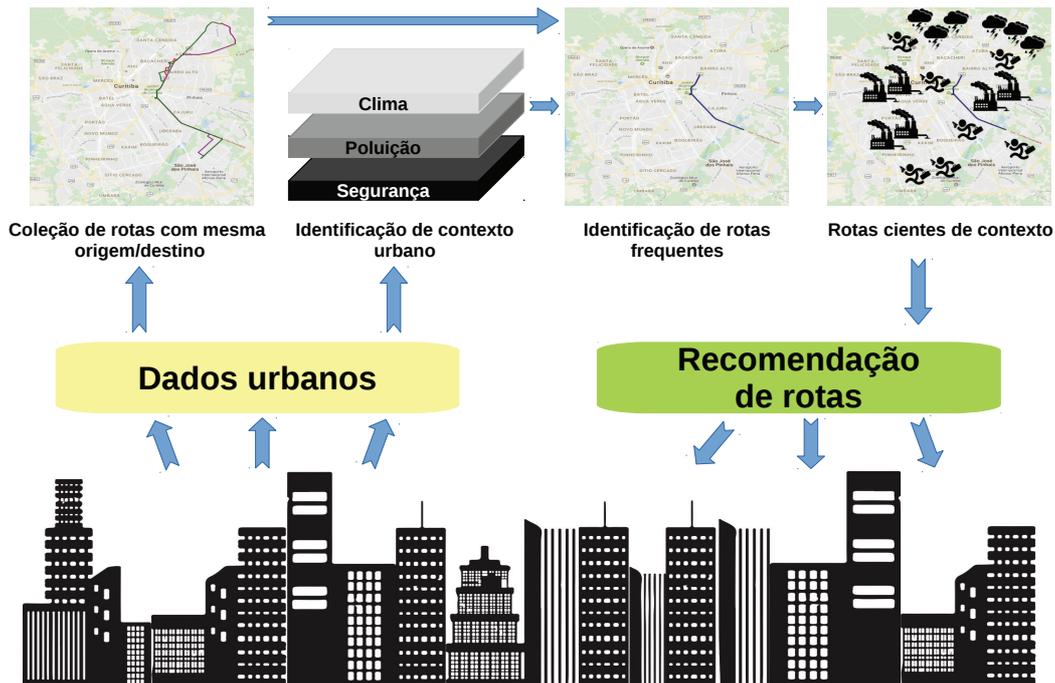


Figure 1. Visão geral do arcabouço proposto.

A Figura 1 ilustra o arcabouço proposto para apoiar serviços de recomendação de rotas cientes de contexto. Considerando o serviço que temos interesse, precisamos de um conjunto de rotas, bem como dados sobre um ou mais aspectos da cidade. Após a obtenção desses dados, que podem ocorrer de diversas formas [Silva and Loureiro 2015], eles são processados para gerar informações de interesse. Em particular, neste trabalho o arcabouço é avaliado na Seção 5 utilizando dados contribuídos por usuários e de órgãos oficiais para detectar áreas de insegurança em áreas urbanas.

A primeira informação de interesse obtida é através de um conjunto de rotas. Seja  $U = \{u_1, u_2, \dots, u_n\}$ ,  $n \geq 1$ , o conjunto de usuários,  $R = \{r_1, r_2, \dots, r_m\}$ ,  $m \geq 1$ , o conjunto de rotas e  $f_i$  é a frequência de uma rota  $r_i \in R$ , para um dado tipo de contexto (e.g., poluição do ar) definida como  $0 \leq f_i \leq 1$ . Cada rota  $r_i \in R$  é percorrida por um usuário  $u_i \in U$ , que pode percorrer uma ou mais rotas. Dado  $R$ , precisamos identificar as rotas que desejamos conhecer o contexto. Essa identificação depende da recomendação que estamos interessados em oferecer. Para isso, devemos responder a algumas perguntas, tais como:

- A recomendação será feita para algum usuário específico  $u_i$ ? Se sim, precisamos filtrar todas as rotas em  $R$  que podem ser utilizadas por  $u_i$ . A mesma ideia vale para um subconjunto de  $U$ . Caso contrário, consideramos todas as rotas sem aplicar um filtro específico;
- Qual a frequência  $f_i$  adequada de uma dada rota  $r_i$  que desejamos estudar o contexto? O valor da frequência depende do contexto. Considerando um contexto de poluição do ar, talvez o valor de frequência seja maior do que para o contexto de segurança, já que percorrer caminhos em áreas com elevado índice de poluição do ar pode ser mais tolerável do que caminhos que coloquem em risco a vida do usuário. Mais detalhes de como realizar esse passo são fornecidos na Seção 3.2.

Em seguida, precisamos analisar os dados sobre algum determinado aspecto da cidade para identificar informações contextuais que serão consideradas nas rotas estudadas. Note que podemos ter várias fontes de dados sobre diferentes aspectos da cidade e, assim, temos a opor-

tunidade de incorporar diversos contextos. Dessa forma, é possível criar camadas de contexto sobre vários aspectos das cidades, tais como, segurança, poluição e clima. Com isso, uma ou mais camadas podem ser consideradas na recomendação de rotas de acordo com os requisitos de interesse do usuário. Para trabalhar com diferentes contextos simultaneamente existem várias estratégias. Neste trabalho, demonstramos esse caso utilizando uma estratégia que considera a área geográfica de todos os contextos identificados para uma região da cidade. Em outras palavras, sobreponemos as áreas dos diferentes contextos no mesmo espaço geográfico. O arcabouço proposto aqui é genérico o suficiente para permitir a utilização de outras estratégias mais sofisticadas, como a utilizada em [Silva et al. 2014], que propõe um modelo para agrupar distintas fontes de dados urbanas. No entanto, uma avaliação do custo/benefício de cada abordagem está fora do escopo deste trabalho.

Nosso foco é a identificação de áreas contextuais em fontes de dados urbanos distintas,  $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$ , onde cada  $F_i \in \mathcal{F}$  é um conjunto de dados sensoriados por pessoas. Esses dados possuem uma localização geográfica  $d$  e o tempo de registro do dado  $t$ . Por exemplo, uma ocorrência de crime possui o endereço de onde o crime ocorreu ( $d$ ) e o momento de sua ocorrência ( $t$ ). De posse do conjunto  $\mathcal{F}$ , o objetivo é identificar áreas contextuais que representam de forma mais adequada a área de cobertura  $a$  de dados no elemento  $F_i$ , em vez de considerar apenas o ponto específico representado por  $d$ .

Note que a cobertura mencionada é dependente do contexto a ser identificado. Por exemplo, considerando  $F_i$  como um conjunto de dados sobre criminalidade e  $F_j$  como um conjunto de dados sobre a poluição do ar, pode ser desejável considerar um valor de  $a$  menor para  $F_i$  do que para  $F_j$ . A ocorrência de um crime em  $d$  não indica apenas que o ponto  $d$  pode ser inseguro, mas sim pelo menos em alguns quarteirões. Já a medição de um nível elevado de poluição em  $d$  pode indicar que uma área maior esteja afetada, por exemplo, um bairro inteiro. A Seção 3.3 é dedicada a discutir uma estratégia para identificar áreas contextuais.

Como também ilustra a Figura 1, de posse das rotas de interesse, bem como de uma ou mais área contextual, podemos gerar rotas cientes de contexto através da combinação dessas informações. Rotas que são cientes de um determinado contexto podem ser exploradas em algoritmos de recomendação de rotas. Por exemplo, usuários podem ativamente requerer rotas que evitem áreas com alto grau de criminalidade, ou rotas que evitem vias que podem comprometer a segurança deles em condições climáticas adversas, etc. Outra possibilidade é oferecer uma recomendação automática, ou seja, sem a requisição ativa dos usuários por determinados tipos de rotas contextuais. Para isso é necessário analisar as rotas que são frequentemente utilizadas por um determinado usuário para fornecer informações que o ajude a tomar melhores decisões, como uma rota alternativa para evitar problemas respiratórios futuros.

### 3.2. Identificação de Rotas Frequentes

Motoristas podem percorrer rotas existentes da cidade por conta própria ou podem utilizar sistemas de recomendação de rotas para ajudá-los em suas viagens diárias. Em ambos os casos, podem existir rotas que são frequentemente utilizadas por eles, já que há uma rotina no trânsito, como observado em [Karnadi et al. 2007]. A identificação desses padrões é especialmente interessante para o arcabouço proposto, pois possibilita o estudo e a recomendação de rotas veiculares cientes de contexto

Considere um usuário específico  $u_i$ , que se desloca diariamente de sua casa para o seu trabalho e vice-versa. Seja  $R$  o conjunto de rotas, filtramos  $R$  para obter  $R_i \subseteq R$  o conjunto de rotas utilizadas por  $u_i$  para percorrer esse trajeto. Seja  $r_i$  a rota mais frequente em  $R_i$  e seja  $f : 0 \leq f \leq 1$ , a frequência que  $r_i$  ocorre em  $R_i$ . Dependendo do contexto de interesse

do usuário e do valor de  $f$ , é possível que uma rota alternativa seja recomendada para  $u_i$ . Por exemplo, suponha que  $u_i$  tem interesse em evitar áreas com elevados índices de poluição do ar, que  $r_i$  tem frequência  $f \geq 0,90$  (limite hipoteticamente alto) e que  $r_i$  passa por várias áreas poluídas (identificadas, por exemplo, com a abordagem discutida na Seção 3.3).

Com isso, o sistema de recomendação tem a oportunidade de fornecer a  $u_i$  uma nova rota  $r_j \in R_i$  (ou não) para evitar áreas poluídas, mesmo que  $r_j$  seja mais longa e aumente o tempo do percurso. Caso  $u_i$  não tenha interesse em evitar áreas poluídas, ainda assim o sistema de recomendação poderia alertá-lo sobre o risco dele contrair doenças graves. No entanto, se  $f \leq 0,30$  (considerado hipoteticamente um limite baixo) e  $r_i$  for um trajeto significativamente menor que  $r_j$ , então  $u_i$  provavelmente não irá optar por  $r_j$ . Similarmente, podemos considerar as rotas de todos ou um subconjunto dos usuários  $U$ . Se considerarmos o conjunto de rotas  $R$  utilizadas por todos usuários, é possível identificar um subconjunto de rotas  $R' : R' \subseteq R$ , onde  $\forall r' \in R'$ ,  $r'$  é uma rota frequente a todos usuários. Dessa forma, se considerarmos as áreas contextuais que  $r'$  permeia, é possível fornecer recomendações como as que discutimos anteriormente.

Para identificar as rotas frequentes, representamos cada rota por um grafo direcionado conexo  $G(V, E)$ , em que um nó  $v_i \in V$  é um ponto específico de uma rota (por exemplo, uma esquina) e uma aresta direcionada  $e_{i,j} \in E$  representa uma via específica entre os pontos  $v_i$  e  $v_j$ . O grafo  $G$  possui um vértice que representa a origem e um vértice que representa o destino da rota, que são os únicos vértices de grau 1 de  $G$ . Se  $R$  contém  $n$  rotas, então denotamos por  $\mathcal{G} = \{G_1, G_2, \dots, G_n\}$ , a coleção de  $n$  grafos  $G$  que representam as  $n$  rotas  $r \in R$ . A coleção  $\mathcal{G}$  é utilizada para identificar a coleção de subgrafos maximais  $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_m\}$ , onde cada  $Q_i \in \mathcal{Q}$ ,  $1 \leq i \leq m$ , é um subgrafo maximal que ocorre com frequência  $f$  em  $\mathcal{G}$ , e  $Q_i$  representa um trecho (i.e., uma parte da rota ou a rota em si) que é comum a  $f$  rotas contidas em  $R$ .

Note que encontrar cada subgrafo maximal  $Q_i$  por meio de testes de isomorfismo é conhecido por ser um problema  $\mathcal{NP}$ -completo [Cook 1971]. Então, para contornar esse problema, podemos utilizar, por exemplo, o algoritmo chamado *graph-based Substructure pattern mining* (gSpan), que minimiza esse problema na sua estratégia adotada [Yan and Han 2002]. O gSpan espera como parâmetro uma lista de grafos  $\mathcal{G}$  e um valor de frequência mínimo  $f$  que um determinado subgrafo deve apresentar. De posse desses parâmetros, o algoritmo retorna uma lista de subgrafos  $\mathcal{Q}$ .

### 3.3. Identificação de Contexto Urbano

Nesta seção, discutimos mais detalhes de uma etapa chave do arcabouço discutido neste estudo: a identificação de contexto. Dados urbanos, quando analisados individualmente, podem não refletir precisamente o estado atual de um contexto específico. Isto porque o dado pode possuir diversos problemas, como estar desatualizado, incorreto ou não possuir a granularidade necessária para representar corretamente um determinado contexto.

Para ilustrar um tipo de problema relacionado à localização do evento, discutimos o que é observado para dados de criminalidade. Dados oficiais sobre criminalidade podem não possuir a localização precisa do crime, por meio de geolocalização ou logradouro completo. Ao invés disso, a localização é correspondente a uma área na cidade, por exemplo, nome da rua e bairro, como é o caso dos dados fornecidos pela prefeitura de Curitiba<sup>3</sup>. Fontes alternativas para esse tipo de dado, como as do site `ondefuirobado.com.br`, que permitem a usuários compartilharem a ocorrência de um crime, normalmente possuem uma localização baseada em

<sup>3</sup><http://www.curitiba.pr.gov.br/dadosabertos>

geolocalização. No entanto, os usuários podem informar erroneamente a localização em que eles presenciaram algum crime, já que eles podem, por exemplo, não recordar o local exato da ocorrência do crime no momento de compartilhar o dado. Diante disso, é importante realizar a análise agregada dos dados urbanos com o intuito de minimizar erros e enriquecer a qualidade da informação obtida.

Para este fim, uma possível estratégia para a identificação de contexto urbano é a clusterização. Clusterização é a organização não-supervisionada de uma coleção de dados em grupos de acordo com alguma medida de similaridade [Jain et al. 1999]. Assim, quando é aplicado o procedimento de clusterização a um conjunto de dados  $F_i$ , são identificados subconjuntos disjuntos de dados  $C_i^1, C_i^2, \dots, C_i^n$  chamados *clusters*, onde dados em um mesmo *cluster* são mais semelhantes entre si do que com dados pertencentes a *clusters* diferentes [Jain et al. 1999]. A medida de similaridade adotada é a distância geográfica provida pela fórmula de haversine [Sinnott 1984].

Dentre as técnicas existentes de clusterização, o algoritmo *Density Based Spatial Clustering of Application with Noise* (DBSCAN) [Ester et al. 1996] é particularmente interessante, porque é capaz de gerar *clusters* com diferentes formatos e tamanhos e por ter bom resultado de clusterização mesmo na presença de ruídos. Por essa razão, esta é a estratégia adotada neste estudo. DBSCAN identifica clusters baseando-se na densidade da vizinhança dos pontos, onde pontos que estão em regiões de baixa densidade (de acordo com o que é especificado para o algoritmo) são chamados *outliers*. Cada *cluster*  $C$  é formado por um conjunto de pontos, onde o número mínimo de pontos é  $\eta$ , os quais estão em uma vizinhança de raio  $\varepsilon$ . Os parâmetros  $\eta$  e  $\varepsilon$  devem ser fornecidos como entrada do algoritmo DBSCAN. Seja  $p_i \subseteq C$  um ponto, que pode ser um ponto *central* se  $p_i$  tiver pelo menos  $\eta - 1$  pontos que estejam à distância máxima  $\varepsilon$  de  $p_i$ . Caso contrário,  $p_i$  é um ponto *periférico* de um ponto central  $p_j \subseteq C$ . *Outliers* são os pontos que não são centrais nem periféricos e são descartados pelo algoritmo. Essa característica é interessante, pois pode auxiliar no descarte de ruídos e dados imprecisos.

Note que a necessidade de ajustar os parâmetros  $\eta$  e  $\varepsilon$  do DBSCAN é fundamental para lidar com múltiplos contextos. Isto porque dois contextos distintos  $F_i$  e  $F_j$ , provavelmente possuem propriedades diferentes. Por exemplo, suponha que  $F_i$  seja o contexto de crime e  $F_j$  seja o contexto de clima. Assim, um *cluster*  $C_i$  obtido a partir de  $F_i$  representa uma área de insegurança na cidade, enquanto que  $C_j$  obtido a partir de  $F_j$  representa as condições climáticas de uma área da cidade. Provavelmente, o raio  $\varepsilon_j$  do *cluster*  $C_j$  será definido como sendo maior do que o raio  $\varepsilon_i$  do *cluster*  $C_i$ , já que uma área de insegurança é tipicamente definida na granularidade de quarteirões ou bairro, enquanto que uma condição climática pode ser referente a uma área que abrange uma cidade inteira.

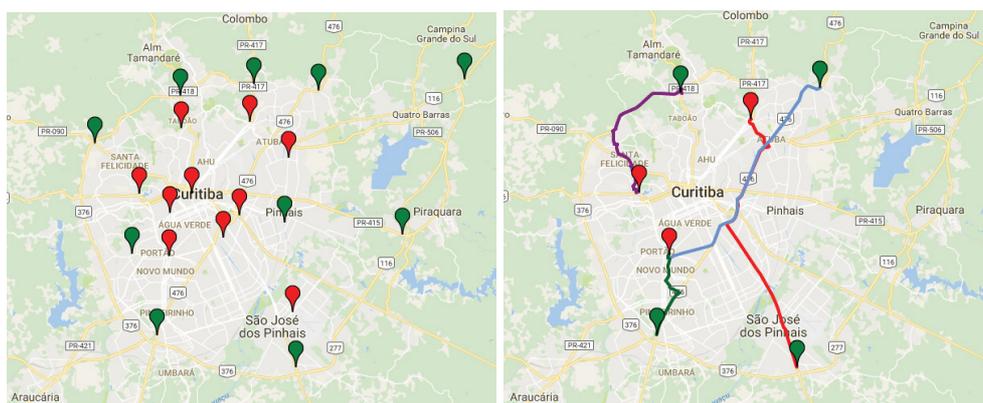
Além disso, a “validade” do *cluster* também pode diferir de acordo com o contexto que ele representa. De fato, dados sobre clima são significativos por um curto período de tempo (e.g., horas ou dia) e, por isso, devem ser atualizados com maior frequência. Enquanto dados sobre crime são relevantes por períodos maiores de tempo (e.g., semanas ou meses), já que uma área de insegurança não tende a se tornar segura em poucos dias. Assim, cada *cluster*  $C$  deve ter definido uma janela temporal  $\mathcal{T} = [t_{min}, t_{max}]$ , onde qualquer ponto  $p \in C$  representa uma localização  $d$  e um instante de tempo  $t$ , tal que  $t_{min} \leq t \leq t_{max}$ .

#### 4. Bases de Dados

Esta seção apresenta as bases de dados analisadas neste trabalho. A Seção 4.1 apresenta as rotas coletadas que foram sugeridas pelo Google Maps. A Seção 4.2 apresenta dados extra-oficiais (Subseção 4.2.1) e oficiais sobre crimes ocorridos em Curitiba (Subseção 4.2.2).

#### 4.1. Rotas Recomendadas pelo Google Maps

Um sistema de recomendação de rotas é uma poderosa ferramenta utilizada pelas pessoas para ajudá-las durante a navegação pela cidade. As pessoas devem informar o ponto de origem (geralmente obtido de forma automática), o ponto de destino e o modo de navegação (e.g., automóvel, transporte público ou a pé) e, em seguida, o sistema recomenda uma rota baseada nas informações fornecidas. Google Maps e Waze são dois sistemas de recomendação de rotas bem conhecidos, que são comumente utilizados pelas pessoas devido à capacidade deles em recomendar “melhores” rotas. Tais sistemas, consideram informações sobre as condições históricas de trânsito e as informações em tempo real para estimar a melhor rota em um determinado momento.



(a) Pontos de origem (verde) e destino (vermelho).

(b) Exemplos de rotas recomendadas.

**Figure 2. Rotas recomendadas pelo Google Maps em Curitiba.**

Neste trabalho, consideramos a cidade de Curitiba como cenário para identificação de padrões na recomendação de rotas do Google Maps. Para isso, foram determinados dois conjuntos distintos de endereços para representar os pontos de origem e destino, onde cada conjunto possui dez endereços distintos. A Figura 2(a) mostra o mapa da cidade de Curitiba com os pontos de origem e destino representados em verde e vermelho, respectivamente. Com ambos conjuntos definidos, é construída a matriz de origem/destino, denotada por  $M_{10 \times 10}$ , que é utilizada para solicitar ao Google Maps Directions API<sup>4</sup> recomendações de rotas a partir de cada origem a todos os destinos definidos. O modo de navegação adotado é o automóvel. A cada iteração, são obtidas 100 rotas recomendadas, com informações do percurso e as condições de trânsito ao longo dele.

Ao todo, foram realizadas 420 iterações, durante 60 dias, com intervalo de 7 horas entre as iterações. Assim, o conjunto de rotas recomendadas em Curitiba tem 42.000 rotas, em diferentes períodos do dia (i.e., manhã, tarde e noite), com instruções detalhadas em cada trecho do percurso (e.g., distância à percorrer e duração prevista), tempo estimado de chegada ao destino e distância total. A Figura 2(b) mostra quatro exemplos de rotas sugeridas pelo Google Maps, rotas representadas pelas cores azul, vermelho, verde e roxo.

#### 4.2. Dados Sobre Ocorrências de Crimes

Considerando o contexto de segurança pública, foram coletados dois conjuntos de dados de ocorrências de crime em Curitiba: um reportado por usuários do serviço “Onde Fui Roubado”

<sup>4</sup><https://goo.gl/kiXllw>

(dados alternativos provenientes da colaboração coletiva), e o outro disponibilizado pela Guarda Municipal (dados oficiais). A seguir, apresentamos uma breve descrição dessas bases de dados.

#### 4.2.1. Dados Contribuídos por Usuários

O sítio “Onde Fui Roubado”<sup>5</sup> é um sistema online onde os usuários contribuem voluntariamente com dados sobre denúncias de crimes sofridos ou presenciados por eles. Com isso, o sistema é capaz de mapear as regiões de insegurança das cidades, o que ajuda os usuários a se prevenirem. Encontra-se disponível no sítio registros de crimes que ocorreram nos últimos quatro meses, com relação a uma data atual, em cidades brasileiras. Os principais atributos acerca de cada ocorrência são: endereço da ocorrência (logradouro, bairro, cidade, estado, país e CEP), geolocalização (latitude e longitude), descrição do crime, data e horário de registro no sistema, natureza do crime, prejuízo estimado e se foi realizado o boletim de ocorrência junto ao órgão competente. Ao longo de quatro meses, de julho à outubro de 2016, foram registradas 331 ocorrências de crime em Curitiba: roubo (36,8%), furto (26,5%), arrombamento domiciliar (7,8%), arrombamento veicular (7,5%), assalto a grupo (7,25%), roubo de veículo (6,35%), tentativa de assalto (6,0%), arrombamento de estabelecimento comercial (0,9%), sequestro relâmpago (0,3%), “saidinha” bancária (0,3%) e arrastão (0,3%). Dessas ocorrências, cerca de 39% não foram registradas oficialmente, o que pode indicar que uma parte significativa dos usuários preferem manifestar sua insatisfação com a insegurança por meio de um sistema extra-oficial. Desta forma, é clara a importância dos dados provenientes da colaboração coletiva para enriquecer o conhecimento acerca da insegurança nas cidades.

#### 4.2.2. Dados Oficiais

A base de dados oficial<sup>6</sup> contém registros sobre ocorrências de crime atendidas pela Guarda Municipal de Curitiba desde 2009 e é atualizada mensalmente, onde a última atualização da base de dados utilizada neste trabalho ocorreu em 1º de novembro de 2016. Ao todo, são 182.267 registros de ocorrências criminais na região metropolitana de Curitiba, sendo 18.235 em 2016. Os principais atributos acerca de cada ocorrência são: o ano de atendimento, bairro, logradouro, se houve flagrante, a natureza da ocorrência (e.g., roubo), a sub-categoria da natureza (e.g., veículo), data e horário de registro da ocorrência no sistema e descrição da situação no momento de atendimento. Como a base de dados oficial possui registros de crimes de diversas naturezas e durante um longo período de tempo, utilizamos apenas o subconjunto de dados que ocorreram no mesmo período que os dados extra-oficiais e cuja natureza do crime seja uma categoria existente nos dados extra-oficiais, totalizando 1.313 ocorrências de crime.

### 5. Análise Experimental

Como prova de conceito do arcabouço proposto, nesta seção consideramos rotas recomendadas pelo Google Maps e dados de criminalidade para Curitiba, com o intuito de analisar as rotas recomendadas com relação à segurança.

#### 5.1. Identificação de Áreas de Insegurança

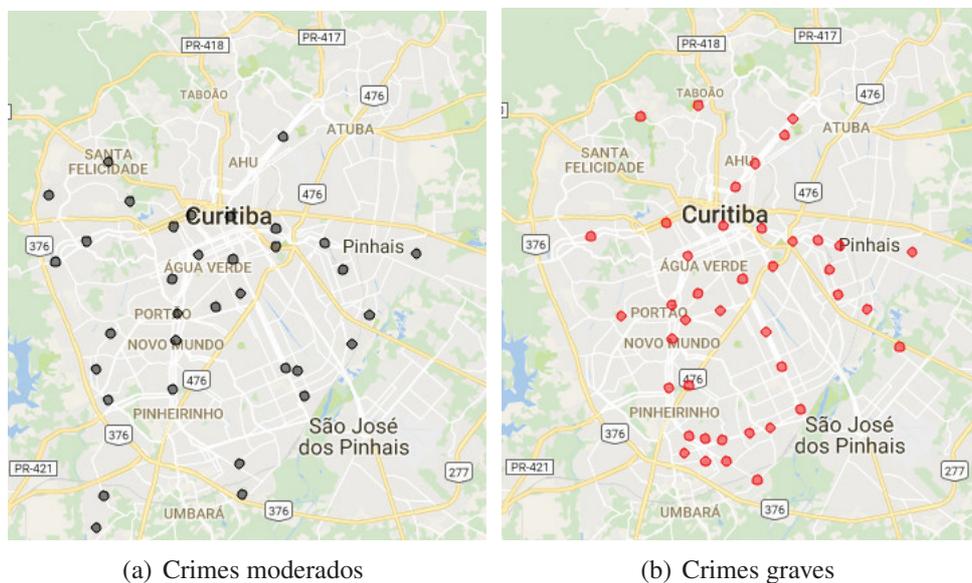
Para identificar as áreas de insegurança de Curitiba, aplicamos a metodologia proposta em nosso arcabouço (Seção 3.3). Consideramos que cada ocorrência de crime atinge diretamente uma área

<sup>5</sup>Disponível em: <https://goo.gl/HPOAHB>

<sup>6</sup>Disponível em: <https://goo.gl/Hf0aP2>

equivalente a um quarteirão (i.e.,  $100 \text{ m}^2$ ). Representamos cada ocorrência de crime que ocorreu em um local  $d$  no instante  $t$  pela variável  $o$ . Isto significa que para a mesma localidade  $d$  podemos ter diversas ocorrências  $o$ . Com isso, um *cluster*  $C = \{o_1, o_2, \dots, o_n\}$ , que contém  $n$  ocorrências de crimes, possui uma área de cobertura  $a$  representando uma área de insegurança em Curitiba. O tamanho de  $a$  pode variar de acordo com a densidade da vizinhança dos pontos, podendo ser equivalente a alguns quarteirões ou até um ou mais bairros. Por simplicidade, consideramos que  $a$  é uma área circular, cujo centro de  $a$  é o centroide de  $C$  e o raio de  $a$  é o raio  $\varepsilon$  utilizado no algoritmo para a identificação de  $C$ .

Dadas essas informações, definimos neste experimento que  $\varepsilon = 250 \text{ m}$  e  $\eta = 5$ , pois acreditamos que esses parâmetros resultem em *clusters* que representem razoavelmente bem áreas de insegurança na cidade. Além disso, consideramos a janela temporal de quatro meses ( $\mathcal{T} = [\text{Junho}/2016, \text{Outubro}/2016]$ ), que é o mesmo período em que os dados sobre crimes são válidos na plataforma `ondefuirobado.com.br`. Em seguida, filtramos os dados de crimes de acordo com o seu grau de severidade. A divisão foi feita em duas classes: (1) crimes moderados, por exemplo furto, e (2) crimes graves, por exemplo, sequestro.

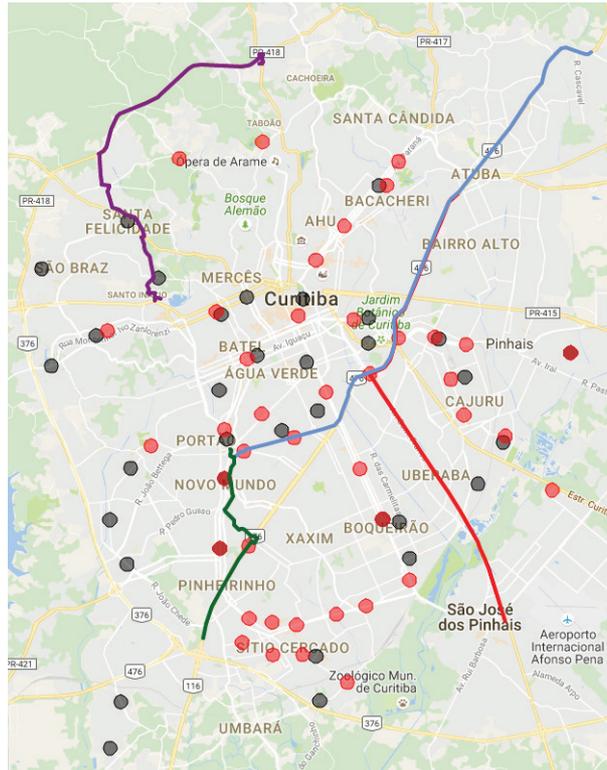


**Figure 3. Áreas ameaçadas por crimes moderados e crimes graves em Curitiba.**

A Figura 3 mostra as áreas de insegurança em Curitiba, onde os *clusters* de crimes moderados (furto, crime tentado, arrombamento, arrastão e “saidinha”) são as áreas em preto mostradas na Figura 3a e os de crimes graves (roubo, sequestro e assalto) são as áreas em vermelho na Figura 3b. Ao todo, foram identificadas 75 áreas de insegurança, sendo 34 de crimes moderados e 41 de crimes graves, havendo sobreposição de áreas em alguns casos.

## 5.2. Segurança de Rotas Frequentes

Nesta seção, identificamos trechos (i.e., subgrafos maximais) das rotas que são regularmente recomendadas pelo Google Maps para verificar se elas tendem a passar por áreas inseguras. Para isso, consideramos as rotas coletadas  $R^i$  com mesmo par origem/destino  $i : 1 \leq i \leq 10$ , como descrito na Seção 4.1. Em seguida, utilizamos a abordagem de identificação de rotas frequentes descrita na Seção 3.2. Definimos a frequência  $f = 0,85$ , o que significa que  $Q^i$  está contida em pelo menos 85% das rotas  $R^i$ . Assim, se  $Q^i$  permeia áreas de insegurança em Curitiba, então os usuários que solicitarem uma rota com origem/destino  $i$ , têm grandes chances de passarem por áreas inseguras.



**Figure 4. Trechos frequentes contidos em rotas recomendadas pelo Google Maps, que permeiam áreas de insegurança em Curitiba.**

A Figura 4 traça no mapa de Curitiba alguns exemplos de trechos frequentes ( $f = 0,85$ ), que estão contidos nas rotas exibidas na Figura 2(b) e disponíveis em  $R^i$ , juntamente com as áreas inseguras encontradas de acordo com a nossa abordagem. Todos esses exemplos ilustram rotas que passam por pelo menos uma área insegura e, em alguns casos, passam por quatro. Esse resultado é um indício de que aspectos de segurança não são considerados na recomendação de uma determinada rota. Ressaltamos que as rotas frequentemente sugeridas nem sempre passam por áreas inseguras e, em alguns casos, o padrão identificado é um trecho pequeno, o que é insuficiente para determinar se as rotas que contêm o padrão passam por áreas de insegurança.

Nessa direção, uma possibilidade é oferecer ao usuário a oportunidade de definir o grau de segurança da rota que será recomendada a ele. Como descrevemos anteriormente, áreas de insegurança são divididas em áreas com crimes moderados ou graves. Dessa forma, o usuário pode estar disposto a assumir o risco de seguir por uma rota que permeia áreas de insegurança com crimes moderados, como o trecho de cor roxa da Figura 4. Alternativamente, o usuário pode optar por uma rota livre de áreas inseguras, mesmo que isso aumente a distância do percurso e o tempo de viagem. Nesse caso, o sistema de recomendação ciente de contexto deve sugerir rotas que evitem áreas inseguras. A Figura 5 mostra possíveis opções de rotas mais seguras, cuja origem/destino são as mesmas exemplificadas na Figura 2(b).

Considerando as métricas (i) distância total e (ii) tempo de viagem, realizamos uma comparação entre as rotas que contêm os trechos frequentes ilustrados na Figura 4, denominadas  $r_1$  (rota roxa),  $r_2$  (rota verde),  $r_3$  (rota azul),  $r_4$  (rota vermelha), com as rotas alternativas ilustradas na Figura 5, denominadas  $r'_1, r'_2, r'_3, r'_4$ . A Tabela 1 mostra os resultados das métricas (i) e (ii) para as rotas estudadas. Os valores de ambas as métricas são calculados por meio da API do Google Maps, onde o tempo de viagem é obtido considerando o mesmo horário de partida

para todas as rotas. Como podemos observar, apesar das rotas mais seguras serem maiores que as rotas frequentes, elas não adicionam um atraso muito grande nas rotas, em média 9,75 min, e nenhum atraso entre as rotas  $r_1$  e  $r'_1$ . Isso ilustra que os mecanismos atuais de recomendação de rotas poderiam ser adaptados para atender a essa nova funcionalidade.

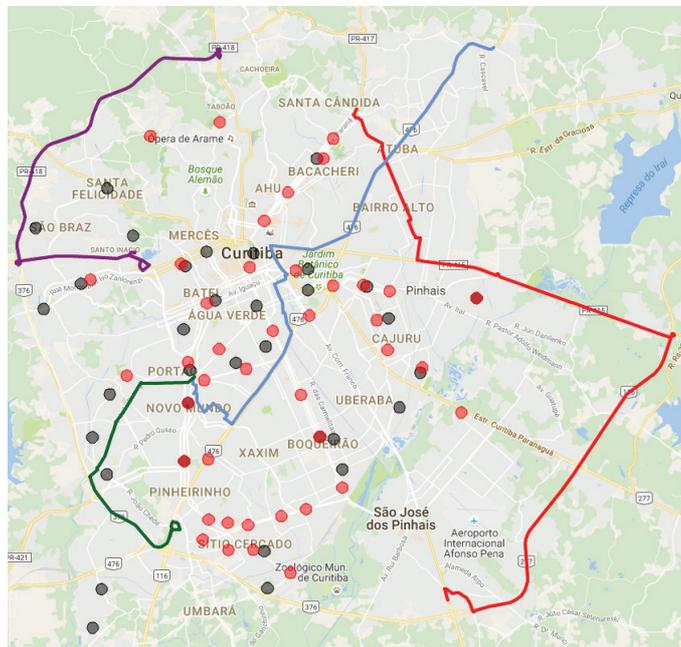


Figure 5. Opções de rotas mais seguras em Curitiba.

Table 1. Rotas frequentes *versus* rotas mais seguras

	$r_1$	$r'_1$	$r_2$	$r'_2$	$r_3$	$r'_3$	$r_4$	$r'_4$
Distância total (km)	15,2	23,1	8,5	14,1	21,8	27,6	25,1	37,5
Tempo de Viagem (min)	22	22	15	22	36	53	35	50

## 6. Considerações Finais e Trabalhos Futuros

O presente trabalho explora a utilização de fontes de dados distintas sobre diferentes aspectos das cidades, juntamente com rotas que são frequentemente utilizadas pelas pessoas em suas viagens diárias, para fornecer mecanismos importantes para a construção de uma rede veicular sensível ao contexto. Para atender esse objetivo, definimos um arcabouço para auxiliar no processo de recomendação de rotas veiculares. Considerando o contexto de segurança, o arcabouço proposto demonstra que pode possibilitar o aprimoramento de sistemas de recomendação de rotas, identificando rotas que são frequentemente recomendadas e que passam por áreas de insegurança. No entanto, é importante observar que o mesmo princípio poderia ser aplicado a qualquer outro tipo de contexto individualmente ou tratado de forma conjunta. Essa é uma das vantagens desta proposta: a flexibilidade em tratar diferentes contextos de forma homogênea.

Como trabalho futuro, pretende-se evoluir este arcabouço para incorporar outras fontes de dados heterogêneas, tais como, redes sociais, redes oportunistas e redes veiculares, para possibilitar a identificação de múltiplos contextos, sem depender de fontes de dados específicos.

## Agradecimentos

Os autores gostariam de agradecer o apoio financeiro concedido da CAPES para o bolsista de doutorado Frances Albert Santos. Ainda, os autores também agradecem a FAPEMIG, Fundação

Araucária e CNPq por financiarem partes dos seus projetos de pesquisas. Por fim, Leandro Villas agradece o apoio financeiro da FAPESP por meio do processo no 2015/07538-1, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

## References

- Bader, R., Neufeld, E., Woerndl, W., and Prinz, V. (2011). Context-aware poi recommendations in an automotive scenario using multi-criteria decision making methods. In *Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation, CaRR '11*, pages 23–30, New York, NY, USA. ACM.
- Calavia, L., Baladrón, C., Aguiar, J. M., Carro, B., and Sánchez-Esguevillas, A. (2012). A semantic autonomous video surveillance system for dense camera networks in smart cities. *Sensors*, 12(8):10407.
- Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the third annual ACM symposium on Theory of computing*, pages 151–158. ACM.
- Dai, J., Yang, B., Guo, C., and Ding, Z. (2015). Personalized route recommendation using big trajectory data. *2015 IEEE 31st International Conference on Data Engineering (ICDE)*, 00(undefined):543–554.
- De Domenico, M., Lima, A., González, M. C., and Arenas, A. (2015). Personalized routing for multitudes in smart cities. *EPJ Data Science*, 4(1):1.
- Demiryurek, U., Banaei-Kashani, F., and Shahabi, C. (2009). Transdec: A data-driven framework for decision-making in transportation systems. 50th Annual Transportation Research Forum, Portland, Oregon, March 16-18, 2009 207726, Transportation Research Forum.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.
- Jain, A., Mangharam, R., and Behl, M. (2016). Data predictive control for peak power reduction. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, BuildSys '16*, pages 109–118, New York, NY, USA. ACM.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323.
- Karnadi, F. K., Mo, Z. H., and c. Lan, K. (2007). Rapid generation of realistic mobility models for vanet. In *2007 IEEE Wireless Communications and Networking Conference*, pages 2506–2511.
- Lee, U., Zhou, B., Gerla, M., Magistretti, E., Bellavista, P., and Corradi, A. (2006). Mobeyes: Smart mobs for urban monitoring with a vehicular sensor network. *Wireless Commun.*, 13(5):52–57.
- Melnikov, V., Krzhizhanovskaya, V., Boukhanovsky, A., and Sloot, P. (2015). Data-driven modeling of transportation systems and traffic data analysis during a major power outage in the netherlands. *Procedia Computer Science*, 66:336–345.
- Quercia, D., Schifanella, R., and Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. *CoRR*, abs/1407.1031.
- Rio, G. (2015). *Mulher morre após casal entrar por engano em comunidade em Niterói*. G1. <http://g1.globo.com/rio-de-janeiro/noticia/2015/10/mulher-morre-apos-entrar-por-engano-em-comunidade-em-niteroi-rj.html>.

- Shah, S., Bao, F., Lu, C.-T., and Chen, I.-R. (2011). Crowdsafe: Crowd sourcing of crime incidents and safe routing on mobile devices. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11*, pages 521–524, New York, NY, USA. ACM.
- Silva, T. H., de Melo, P. O. V., Almeida, J. M., Vianna, A. C., Salles, J., and Loureiro, A. A. (2014). Definição, modelagem e aplicações de camadas de sensoriamento participativo. In *Brazilian Symposium on Computer Networks and Distributed Systems (SBRC'14), Florianópolis, Brazil*.
- Silva, T. H. and Loureiro, A. A. (2015). Computação urbana: Técnicas para o estudo de sociedades com redes de sensoriamento participativo. In *Anais da XXXIV Jornada de Atualização em Informática*, volume 8329, pages 68–122. Sociedade Brasileira de Computação – SBC.
- Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, 68(2):159+.
- ter Mors, A. W., Witteveen, C., Zutt, J., and Kuipers, F. A. (2010). *Context-Aware Route Planning*, pages 138–149. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Vaghefi, A., Jafari, M., Bisse, E., Lu, Y., and Brouwer, J. (2014). Modeling and forecasting of cooling and electricity load demand. *Applied Energy*, 136(C):186–196.
- Vo, H. T. (2015). Enabling smart transit with real-time trip planning. In *Proceedings of the ACM First International Workshop on Understanding the City with Urban Informatics, UCUI '15*, pages 13–18, New York, NY, USA. ACM.
- Wang, H., Li, G., Hu, H., Chen, S., Shen, B., Wu, H., Li, W.-S., and Tan, K.-L. (2014). R3: A real-time route recommendation system. *Proc. VLDB Endow.*, 7(13):1549–1552.
- Yan, X. and Han, J. (2002). gspan: Graph-based substructure pattern mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 721–724. IEEE.