

# Um *Middleware* para Sensores Virtuais Baseados em Localização no Contexto de Cidades Inteligentes

Alice A. F. Menezes<sup>1</sup>, Carlos Maurício S. Figueiredo<sup>2,3</sup>

<sup>1</sup>Instituto de Computação (iComp) – Universidade Federal do Amazonas (UFAM)  
Manaus – AM – Brasil

<sup>2</sup>Escola Superior de Tecnologia (EST) – Universidade do Estado do Amazonas (UEA)  
Manaus – AM – Brasil

<sup>3</sup>Samsung Ocean  
Manaus – AM – Brasil

alice.menezes@icomp.ufam.edu.br, cfigueiredo@uea.edu.br

**Abstract.** *Urban areas are growing over the years and this, together with the modernization of cities, allows us to have access to heterogeneous data that can be used to create new services and use existing resources and infrastructures efficiently. With this, several platform architectures for Smart Cities were proposed with the purpose of managing a large volume of data and providing solutions for urban issues. In this way, this work proposes a middleware to facilitate the knowledge extraction from the data generated in the cities, monitoring abstract information generated from low level data. As proof of concept, we made a case study with a large volume of open data from an urban area.*

**Resumo.** *As áreas urbanas estão crescendo ao longo dos anos, e este crescimento, em conjunto com a modernização das cidades, nos permite ter acesso aos dados heterogêneos que podem ser utilizados para a criação de novos serviços e utilização mais eficiente de recursos e infraestruturas já existentes. Com isto, diversas arquiteturas de plataformas para Cidades Inteligentes foram propostas com o objetivo de gerenciar um grande volume de dados e prover soluções para problemas urbanos. Neste sentido, este trabalho propõe um middleware para facilitar a extração de conhecimento dos dados gerados nas cidades, monitorando informações abstratas geradas a partir de dados de baixo nível. Como prova de conceito, realizamos um estudo de caso com um grande volume de dados abertos de uma área urbana.*

## 1. Introdução

Com a modernização das cidades, o progresso da tecnologia e o aumento da população, questões como a poluição do ar, o consumo de energia e os congestionamentos de trânsito estão se tornando recorrentes. Tais problemas podem ser identificados, analisados e tratados do ponto de vista tecnológico utilizando o conceito de Cidades Inteligentes, cujo objetivo é monitorar áreas urbanas para melhorar continuamente a vida dos habitantes e os serviços oferecidos pelo governo e outras instituições, além de ajudar na utilização de recursos e infraestruturas da cidade [Santana et al. 2017; Zheng et al. 2014].

Para estudar e entender os eventos urbanos, fenômenos que ocorrem nas cidades, há a necessidade de coletar e processar dados de múltiplas fontes, tais como sensores

físicos, dados abertos do governo e de outras instituições, e cidadãos que atuam como sensores através de redes sociais e dispositivos pessoais [Berntzen et al. 2018]. Gerenciar, analisar e fundir os dados destas fontes torna possível não só a resolução dos problemas citados, como também proporciona planejamentos eficientes e rápidos, e a implementação e integração de novos serviços.

No entanto, extrair e fundir conhecimentos dos dados de forma eficiente para agregar valor e identificar áreas de atuação tornou-se um desafio, visto que os dados gerados nas cidades são dispersos e estão em diversos formatos, além de que alguns eventos urbanos podem não ser identificados por todas as fontes de dados referentes a uma mesma região.

Trabalhos existentes, tais como os de Wu et al. [2014], Costa e Santos [2016] e Souza et al. [2018], propõem arquiteturas para resolver alguns dos problemas de aquisição, gerenciamento e agregação dos dados, mas são limitados quanto à extrair conhecimento em diferentes níveis analisando múltiplas fontes, principal contribuição deste trabalho.

Desta forma, neste trabalho propomos um *middleware* para Sensores Virtuais baseados em localização, cujo objetivo é : (i) proporcionar a integração e a criação de novos serviços, além da identificação de melhorias em áreas urbanas, fazendo uso de fontes heterogêneas de dados para o auxílio na tomada de decisões voltadas para o planejamento urbano, e (ii) facilitar o processo de extração e fusão de conhecimento, a partir de dados geolocalizados, por meio de métodos implementados.

Os Sensores Virtuais são uma emulação de um grupo de sensores físicos presentes em uma área urbana monitorada [Madria et al. 2014]. Estes sensores podem realizar medições e monitoramentos indiretos de condições abstratas, após receberem dados de baixo nível como entrada [Kabadayi et al. 2006].

A utilização dos Sensores Virtuais permite a extração de conhecimento a partir de diversas fontes presentes na região urbana monitorada, como *smartphones* e câmeras de segurança. Com a implementação destes sensores neste trabalho, passamos a monitorar a mobilidade urbana a partir de dados de baixo nível, como latitudes e longitudes de origem e destino de um determinado meio de transporte, que são obtidos por meio de sensores físicos.

A validação do *middleware* proposto neste trabalho é realizada por meio de um estudo de caso com dados abertos do governo e de uma instituição privada da cidade de Nova Iorque. Como resultado, padrões de mobilidade foram identificados e utilizados para detectar regiões afetadas por um evento urbano.

## **2. Trabalhos Relacionados**

Com o surgimento do conceito de Cidades Inteligentes e a busca por solucionar problemas decorrentes do aumento da população nas cidades, diversas arquiteturas e plataformas tem sido apresentadas com o objetivo de coletar, gerenciar e agregar um grande volume de dados de múltiplas fontes.

Wu et al. [2014] propõem uma plataforma chamada Concinnity, que permite o desenvolvimento de aplicações e fluxos de processamento de dados baseados em sensores físicos geolocalizados. Para áreas onde não existem sensores físicos disponíveis, a

plataforma estima medidas e cria Sensores Virtuais por meio de uma quantidade mínima de sensores físicos de áreas adjacentes. Como prova de conceito, os autores apresentaram uma aplicação para a exploração visual da poluição do ar na cidade de Londres.

Costa e Santos [2016] apresentam uma arquitetura de *Big Data* para cidades inteligentes com o objetivo de criar novos serviços. Para a validação da arquitetura, os autores apresentaram um estudo de caso em que o agrupamento de dados de atrasos de voos permite a criação de perfis de atraso para áreas urbanas, aeroportos e companhias aéreas.

Souza et al. [2018] propõem um *middleware* chamado SGeoL, que visa integrar dados de diferentes fontes em um ambiente de cidade inteligente. Os autores apresentam um modelo de camadas unificado, no qual os dados armazenados na plataforma possuem coordenadas geográficas para a representação de entidades no mapa da cidade. O *middleware* possui funcionalidades que permitem que os dados armazenados sejam agregados, visualizados e analisados. Como prova de conceito, o SGeoL foi utilizado como base para a criação de uma aplicação de planejamento urbano para entidades governamentais.

Assim como nos trabalhos apresentados, nós buscamos processamento eficiente para milhões de dados. Porém, nosso diferencial e contribuição está na etapa de agregação dos dados com o objetivo de gerar conhecimento para a solução de problemas e propostas de melhorias em regiões urbanas. Para isto, criamos métodos para monitorar informações abstratas geradas a partir de dados de baixo nível, transformando-os para melhor análise de um determinado aspecto, como a mobilidade entre regiões.

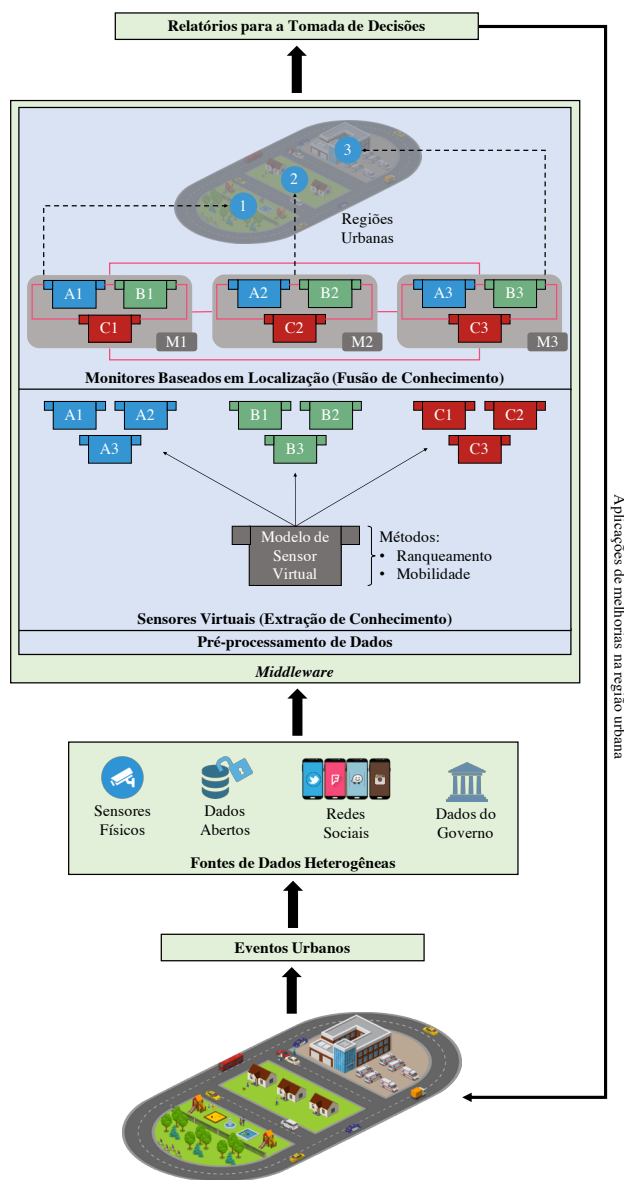
### 3. Arquitetura do *Middleware*

Para prover a melhoria contínua em áreas urbanas, nós propomos nesta seção a arquitetura de um *middleware* que permite a geração e fusão de conhecimento por meio de Sensores Virtuais (Figura 1). Estes sensores contêm informações de dados heterogêneos, que são gerados pelos habitantes, empresas e instituições da própria área urbana.

O *middleware* proposto pode funcionar como uma ferramenta auxiliar na tomada de decisões baseadas em dados, separando um determinado perímetro urbano em regiões que podem ser analisadas sozinhas ou em conjunto. Os dados geolocalizados que são gerenciados pelo *middleware* podem ser não apenas aqueles que incluem localização precisa, como latitude e longitude, mas também aqueles nas quais a localização pode ser extraída a partir de fontes diversas e não precisas, como textos contidos em mensagens de *microblogs*.

Embora o *middleware* tenha como foco dados geolocalizados, normalmente estes dados estão associados a dados temporais, como data e hora da ocorrência de um evento urbano. Isto nos permite realizar uma análise em diferentes granularidades, tanto espaciais quanto temporais, o que é relevante devido à rotina dos habitantes de uma determinada região, que pode mudar de acordo com diversos fatores, como mudanças climáticas, crises econômicas e implementações de novas infraestruturas relacionadas ao planejamento urbano.

A arquitetura proposta foi baseada no conceito de Internet das Coisas (IoT), que estabelece que objetos estão conectados à Internet. Estes objetos devem ter um identificador único, além de terem localização e *status* conhecidos [Coetzee e Eksteen 2011].



**Figura 1. Visão geral da arquitetura do *middleware*.**

No contexto de IoT, um *middleware* é um componente responsável por receber dados de sensores físicos, demandando armazenamento e processamento. Além disso, os usuários podem acessar, visualizar e analisar os dados extraídos dos sensores por meio de um componente de apresentação [Gubbi et al. 2013].

Desta forma, nossos Sensores Virtuais, espalhados nas regiões urbanas como uma rede de sensores sem fio, possuem um identificador e monitoram informações abstratas a partir de dados de baixo nível. Estes sensores podem se integrar com sensores em uma mesma região ou em regiões diferentes para a extração de conhecimento.

Como exemplo do funcionamento destes Sensores Virtuais, podemos citar os dados de redes sociais, que geralmente são coletados no formato de texto. Após o pré-processamento e inserção nos sensores, é possível monitorar diferentes eventos, como acidentes de trânsito [Menezes e Figueiredo 2017] ou as características e anomalias de

uma região ao longo do tempo [Menezes et al. 2017; Xavier et al. 2018].

### **3.1. Componentes do *Middleware***

Nesta Seção, apresentamos os componentes do *middleware* informando quais são dos dados de entrada, como é feito o processamento dos dados e qual é a saída esperada, além de identificar possíveis usuários e como as informações finais geradas podem ser utilizadas para a tomada de decisões baseadas em dados.

#### **3.1.1. Pré-processamento de Dados**

O componente de *pré-processamento de dados* é responsável por mapear os dados de entrada para as zonas urbanas em análise, o que é importante para a integração de dados heterogêneos. Para isto, é necessário obtermos as zonas de estudo em conjunto com seus identificadores.

A forma destas zonas pode estar de acordo com fatores diversos e podem representar bairros, distritos escolares ou industriais, área de atendimento de serviços específicos, entre outros. No geral, estas zonas e seus identificadores são disponibilizadas em um arquivo *shapefile*, contendo uma ou mais camadas de zoneamento. Caso o usuário não tenha um arquivo de zonas, o *middleware* pode criar zonas genéricas em forma de quadrados na área estudada.

A saída deste componente são os dados de entrada correspondentes à área urbana estudada, associados com os identificadores de suas respectivas zonas. Desta forma, os dados que não são pertencentes a esta área urbana, não estarão presentes nos dados de saída.

#### **3.1.2. Sensores Virtuais**

Este componente é responsável por gerar conhecimento por meio dos dados de entrada para auxiliar em decisões. Cada tipo de dado permite a utilização de diferentes métodos de um sensor modelo. Os dados do tipo *categórico*, por exemplo, podem utilizar métodos de *ranqueamento* para a criação de regiões funcionais [Menezes et al. 2017; Xavier et al. 2018].

O método de *mobilidade*, uma importante funcionalidade do *middleware*, pode ser utilizado por dados do tipo *transição*. Este método recebe os dados do componente de pré-processamento como entrada. A saída são informações que apresentam a densidade de fluxos de uma região para outra da cidade, tornando possível a identificação de padrões. Resultados referentes a este método são apresentados na Seção 4.3.

#### **3.1.3. Monitores Urbanos Baseados em Localização**

Este componente é responsável por encontrar padrões na cidade e identificar as mudanças decorrentes de eventos urbanos, por meio da interoperabilidade das informações dos Sensores Virtuais do componente anterior. Os monitores são mutáveis e podem possuir mais ou menos Sensores Virtuais dependendo da granularidade da análise. Ressaltamos que a

quantidade de dados heterogêneos e de Sensores Virtuais pode influenciar na qualidade da análise realizada.

A saída deste componente são informações no formato de dados tabulares, gráficos e mapas, que podem ser utilizadas de várias maneiras, desde a criação de *dashboards* em aplicações até documentos de relatório. Dentre os possíveis usuários do *middleware* proposto, podemos citar empreendedores e empresas que visam oportunidade de negócio, pessoas que atuam em atividades estratégicas do governo e que podem identificar e remediar problemas urbanos, e habitantes que buscam informações gerais ou que querem conhecer a cidade.

Ressaltamos que os resultados produzidos podem resultar em melhorias na cidade, que eventualmente ocasionam mudanças no comportamento dos habitantes e, eventualmente, geram novos dados de entrada, que permitem a identificação de novos padrões.

## 4. Estudo de Caso

Nesta Seção, apresentamos um estudo de caso da utilização do *middleware* na cidade de Nova Iorque. Todos os parâmetros de entrada são disponibilizados como dados abertos da cidade.

### 4.1. Área Urbana Estudada

A cidade de Nova Iorque possui cinco grandes regiões ou distritos: Manhattan, Queens, Brooklyn, Bronx e Staten Island. Estas regiões comumente podem ser divididas de diversas maneiras, como bairros do censo ou áreas escolares. Neste estudo de caso, dividimos estes distritos em zonas de táxi disponibilizadas pela agência governamental *Taxi and Limousine Commission* (TLC)<sup>1</sup>. Desta forma, podemos agregar os dados em dois níveis: distritos e zonas de táxi.

Na Figura 2(a), apresentamos as cinco regiões e suas divisões. As regiões com o maior número de zonas de táxi são Manhattan e Queens, contendo 69 zonas (Figura 2(b)). Em seguida estão as regiões do Brooklyn (61 zonas), Bronx (43 zonas) e Staten Island (20 zonas). Além destas zonas, também foi disponibilizada uma zona referente ao aeroporto internacional de Newark (EWR), que apesar de ser localizado em Nova Jersey fornece serviços para a região metropolitana de Nova Iorque.

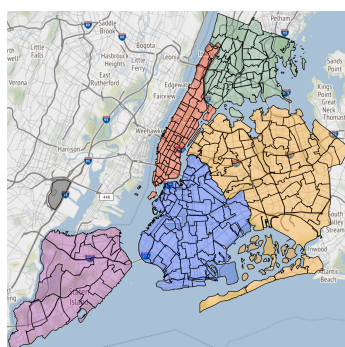
### 4.2. Bases de Dados: *Taxi and Limousine Commission* e *Citi Bike*

Como bases de dados, utilizamos 123.070.119 registros de percursos de táxi, fornecidos pela TLC, e 16.364.255 registros de percursos bicicletas alugadas, fornecidos pela empresa Citi Bike<sup>2</sup>. As bases de dados são do tipo *transição* e estão no formato de origem e destino, ou seja, são registrados o horário e o local do ponto de partida e de chegada, além de outras informações.

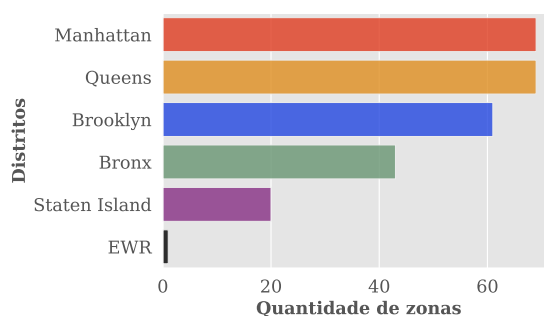
Em relação aos registros de táxi, são 111.377.780 percursos de *táxis amarelos* e 11.692.339 percursos de *táxis verdes*. Os táxis amarelos são os únicos veículos licenciados para pegar passageiros em qualquer lugar de Nova Iorque. Os táxis verdes atuam ao norte de Manhattan e nas regiões mais distantes. No que diz respeito às bicicletas alugadas, as estações do serviço encontram-se nas regiões de Manhattan, Queens e Brooklyn.

<sup>1</sup><https://www1.nyc.gov/site/tlc/index.page>

<sup>2</sup><https://www.citibikenyc.com/>



(a) Mapa de zonas de táxi estudadas por distrito.



(b) Quantidade de zonas de táxi por distrito.

**Figura 2. Zonas de táxi e distritos na área urbana estudada.**

### 4.3. Sensores Virtuais

Quando os dados dos percursos de táxis e bicicletas são inseridos nos Sensores Virtuais, é possível extrair informações de mobilidade das regiões monitoradas por estes sensores. Desta forma, assim como os sensores físicos registram dados de percursos de cada táxi e bicicleta individualmente, os sensores virtuais registram as transições em cada zona.

O monitoramento efetuado pelos Sensores Virtuais pode ser realizado em diversas granularidades espaciais ou temporais, tornando possível ter uma visão geral das transições por serviço durante todo o ano ou apenas por um dia, assim como podemos verificar as transições que ocorrem entre zonas ou entre os distritos. Isto torna possível uma análise mais detalhada para a melhoria dos serviços da cidade, além de possibilitar um estudo a respeito dos impactos causados por algum evento específico, como um desastre natural.

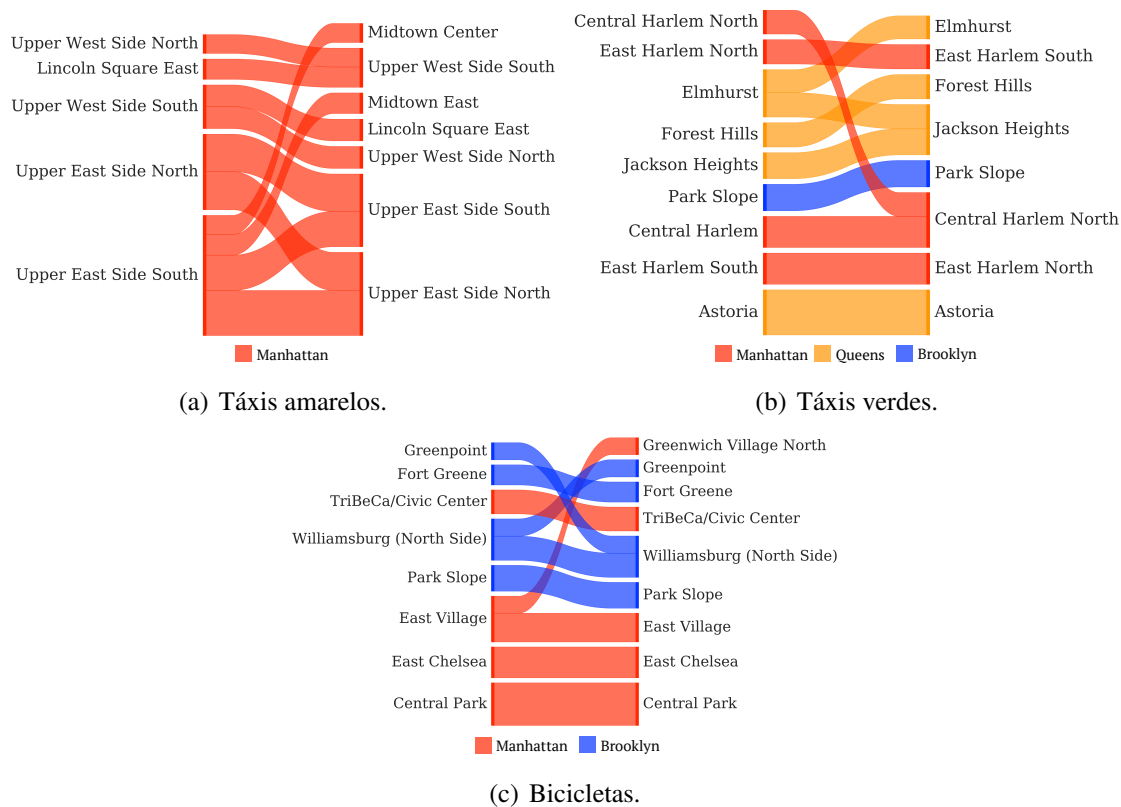
Na Figura 3, apresentamos os dez percursos que mais ocorrem entre as zonas de táxi na cidade de Nova Iorque para cada serviço, sendo a origem do percurso à esquerda e o destino à direita. É possível observar, por meio do monitoramento dos Sensores Virtuais, que grande parte dos percursos são realizados dentro da mesma zona ou entre zonas próximas que pertencem à mesma região.

Para os táxis amarelos (Figura 3(a)) os percursos mais frequentes são na ilha de Manhattan, que é a principal região de atuação deste serviço, sendo que a maioria dos percursos tem início no *Upper East Side South* e o destino mais requisitado é o *Upper East Side North*. Estas zonas encontram-se ao lado do *Central Park* e são conhecidas por serem umas das mais ricas e com os maiores custos habitacionais da cidade.

Os táxis verdes (Figura 3(b)) possuem maior atuação no Queens, na zona de *Astoria*. Esta região da cidade é conhecida pela grande quantidade de imigrantes. Já as bicicletas alugadas (Figura 3(c)) têm percursos mais presentes no *Central Park*, um dos principais pontos turísticos da cidade.

### 4.4. Monitores Urbanos

Os monitores urbanos são resultado da fusão dos Sensores Virtuais de uma mesma área em análise. Com eles é possível identificar padrões e a ocorrência de eventos. Desta forma, é possível se preparar para eventos semelhantes que possam ocorrer no futuro,



**Figura 3. Percursos que mais ocorrem na cidade de Nova Iorque.**

além de identificar possíveis regiões para a implementação de melhorias ou alocação de recursos.

No dia 05/11/17 (Domingo), por exemplo, verificamos que houve uma queda de 5,2% na quantidade de percursos de táxis amarelos (representando 498.711 percursos), 5,5% na quantidade de percursos de táxis verdes (representando 51.122 percursos) e 29,9% na quantidade de percursos de bicicletas alugadas (representando 566.889 percursos) em relação ao mesmo período no mês anterior (dia 01/10/17).

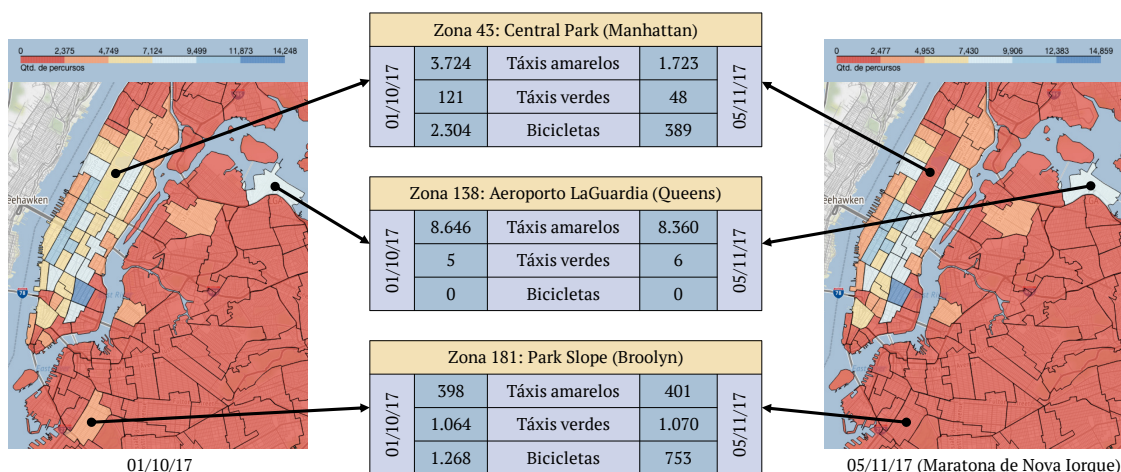
O evento ocorrido neste dia foi a maratona de Nova Iorque, uma das seis maiores maratonas do mundo, fazendo parte do *World Marathon Majors*. Com os monitores urbanos, foi possível identificar as regiões mais afetadas pelo evento, comparando as medidas entre os Sensores Virtuais.

Caso nenhum sensor tivesse uma diferença brusca, era estabelecido que não houve impacto. Com um ou dois dos sensores identificando uma diferença entre os padrões dos dados, o monitor alertava um impacto baixo ou moderado. Com os três sensores identificando diferenças, o impacto é alto.

A zona referente ao *Central Park*, por exemplo, foi identificada como de impacto alto (Figura 4). A queda na medição de percursos iniciados nesta região ocorreu porque este era o ponto de chegada da maratona. Outras zonas, como a de *Park Slope*, foram identificadas como de baixo impacto, visto que o percurso ocorria apenas na fronteira da zona. Em zonas como a do aeroporto *LaGuardia* não houve impacto, pois o percurso da corrida não passava por lá e nem em zonas próximas.



Por meio destes exemplos, ressaltamos a importância da utilização dos dados de múltiplas fontes e dos monitores urbanos para identificar regiões para análise na cidade. No caso da zona de *Park Slope*, dependendo da fonte de dados utilizada não seria identificado nenhum impacto relacionado ao evento ocorrido.



**Figura 4. Monitores urbanos apresentando as diferenças nas regiões durante a maratona de Nova Iorque.**

Desta forma, com os monitores urbanos, é possível que as partes interessadas identifiquem melhorias e soluções de acordo com o seus casos. Os moradores da região podem se planejar quanto à locomoção em um evento futuro semelhante a este. As empresas e agências prestadoras dos serviços monitorados pelos Sensores Virtuais, podem analisar o impacto financeiro em relação à queda de percursos realizados nas regiões afetadas. O governo pode alocar de forma eficiente a utilização de recursos não só para as vias das zonas em que ocorreu alto impacto, mas também para suas zonas adjacentes.

## 5. Conclusão e Trabalhos Futuros

Neste trabalho, apresentamos um *middleware* para Sensores Virtuais baseados em localização. A camada de Sensores Virtuais tem como objetivo extrair conhecimento a partir de dados de baixo nível. Além disso, há uma camada de monitores urbanos baseados em localização, que visa fundir conhecimento para a identificação de regiões da área estudada em que ocorreu algum tipo de mudança. Ressaltamos que os métodos implementados nestas camadas são genéricos e podem ser instanciados para a análise de diferentes eventos urbanos.

Para a validação do *middleware*, realizamos um estudo de caso utilizando dados abertos da cidade de Nova Iorque, em que identificamos padrões de mobilidade a partir de registros geolocalizados de táxis e bicicletas alugadas. Além disso, identificamos e analisamos algumas das regiões da cidade durante a maratona de Nova Iorque.

Dentre os desafios durante a implementação deste estudo de caso, ressaltamos a necessidade de paralelização do processamento do grande volume de dados, que foi necessário para extrair conhecimento de forma rápida e eficiente. Com isto, tornou-se possível processar cerca de 110 milhões de dados em aproximadamente 4 minutos.

Como trabalhos futuros, propomos a implementação de novos métodos de extração de conhecimento, utilizando técnicas de mineração de dados e aprendizagem de máquina, para identificar padrões e eventos. Além disso, propomos novos estudos de caso com a agregação de novos dados para verificar circunstâncias diversas.

## Referências

- Berntzen, L., Johannesen, M. R., Böhm, S., Weber, C., e Morales, R. (2018). Citizens as sensors: human sensors as a smart city data source. In *Proceedings of the 7th International Conference on Smart Cities, Systems, Devices and Technologies Citizens*, pages 11–18. IARIA.
- Coetzee, L. e Eksteen, J. (2011). The internet of things - promise for the future? an introduction. In *IST-Africa Conference Proceedings*, pages 1–9. IEEE.
- Costa, C. e Santos, M. Y. (2016). Basis: A big data architecture for smart cities. In *2016 SAI Computing Conference (SAI)*, pages 1247–1256. IEEE.
- Gubbi, J., Buyya, R., Marusic, S., e Palaniswami, M. (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660.
- Kabadayi, S., Pridgen, A., e Julien, C. (2006). Virtual sensors: Abstracting data from physical sensors. In *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, pages 587–592. IEEE Computer Society.
- Madria, S., Kumar, V., e Dalvi, R. (2014). Sensor cloud: A cloud of virtual sensors. *IEEE software*, 31(2):70–77.
- Menezes, A. A. F., dos Santos, J. W. V., Souza, B. Á., Almeida, T. G., Nakamura, F. G., Nakamura, E. F., e Figueiredo, C. M. S. (2017). Um método de partição de regiões funcionais utilizando dados de redes sociais. In *XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 860–873. SBC.
- Menezes, A. A. F. e Figueiredo, C. M. S. (2017). Um método de agrupamento incremental para a detecção de eventos em redes sociais. In *XXXVII Congresso da Sociedade Brasileira de Computação*, pages 1556–1565. SBC.
- Santana, E. F. Z., Chaves, A. P., Gerosa, M. A., Kon, F., e Milojevic, D. S. (2017). Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture. *ACM Computing Surveys (CSUR)*, 50(6):78.
- Souza, A., Pereira, J., Batista, T., Cavalcante, E., Cacho, N., Lopes, F., e Almeida, A. (2018). A geographic-layered data middleware for smart cities. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web, WebMedia '18*, pages 411–414, New York, NY, USA. ACM.
- Wu, C., Birch, D., Silva, D., Lee, C.-H., Tsinalis, O., e Guo, Y. (2014). Concinnity: A generic platform for big sensor data applications. *IEEE Cloud Computing*, 1(2):42–50.
- Xavier, A. S., Menezes, A. A. F., Nakamura, F. G., e Figueiredo, C. M. S. (2018). Analysis of urban regions popularity using foursquare. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web, WebMedia '18*, pages 379–386, New York, NY, USA. ACM.
- Zheng, Y., Capra, L., Wolfson, O., e Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38.