

Luiz Eduardo Cartolano

Geração de contexto através da fusão de dados de sensoriamento participativo

Trabalho apresentado do curso de MC040 - Iniciação Científica
pelo aluno Luiz Eduardo Cartolano, email cartolanoluiz@gmail.com

Orientador: Prof. Dr. Leandro Villas

Campinas, SP, Julho de 2018

Conteúdo

1	INTRODUÇÃO	3
2	MOTIVAÇÃO E OBJETIVOS	4
3	FUNDAMENTAÇÃO TEÓRICA	5
3.1	Sistemas de Transporte Inteligentes	5
3.1.1	Definição	5
3.1.2	Aplicações	5
3.2	Mapas de Calor	6
3.2.1	Definição	6
3.2.2	Gmplot	6
3.3	Grafos	6
3.3.1	Definição	6
3.3.2	OpenStreetMap	7
3.3.3	Simulator of Urban MObility - SUMO	7
4	TRABALHOS RELACIONADOS	7
5	PROPOSTA	8
5.1	Revisão bibliográfica	8
5.2	Definição da estratégia de coleta e mineração	8
5.3	Definição da estratégia de geração de contexto	9
5.4	Avaliação e validação da estratégia	10
5.5	Escrita de relatório e documentação	10
6	CONCLUSÕES	10
	ANEXOS	11
	REFERÊNCIAS	13

RESUMO

O Twitter é, sem sombra de dúvidas, uma das plataformas que revolucionaram o mundo nos últimos anos, em especial, no âmbito de como as pessoas se comunicam. Nele, usuários falam sobre os mais diversos assuntos, desde seus problemas de saúde até problemas que encontram no seu dia a dia, como o trânsito, além de indicarem onde estão ou estiveram por meio de check-ins. Nesse paper, buscamos usar o Twitter como uma maneira de analisar o trânsito de grandes centros urbanos, para que, por meio de sistemas de transporte inteligente, possamos resolver o problema da mobilidade urbana no Brasil. O trabalho consistiu em dois momentos, em um primeiro período focou-se na coleta de dados do Twitter e do Google Maps. Enquanto que no segundo momento, buscou-se analisar a movimentação de pessoas usando a geo-localização dos tweets e, comparando com dados obtidos pelo Google Maps, prever o trânsito na cidade de Campinas.

1 INTRODUÇÃO

O Twitter está entre as redes sociais que modificaram a maneira como nos comunicamos. Somente no Brasil, existe, segundo [1], cerca de 18 milhões de usuários ativos no Brasil, e mais de 300 milhões no mundo. Estes são responsáveis por gerar cerca de 500 milhões de tweets (nome dado as mensagens postadas na rede) por dia, de acordo com [2]. Nesses tweets, usuários falam sobre os mais diversos assuntos, desde seus problemas de saúde até problemas que encontram no seu dia a dia, como o trânsito, além de indicarem onde estão ou estiveram por meio de check-ins.

Defini-se por mobilidade urbana a condição criada para as pessoas poderem se locomover entre as diferentes zonas de uma cidade. Atualmente, os automóveis particulares e os meios de transportes públicos são os meios de mobilidade urbana mais utilizados. Como nos mostra Vianna Young em [3], a mobilidade urbana é um dos grandes problemas enfrentados pela população brasileira atualmente, especialmente para os moradores de grandes centros urbanos. A baixa mobilidade dos grandes centros tem acarretado em uma série de problemas para a população. Alguns deles são o tempo perdido no trânsito, em São Paulo, por exemplo, a população em média 105,23 minutos por dia. Além disso, os problemas de saúde ocasionados que se tornam cada vez mais comuns, como: dores musculares, lombares e cervicais, problemas de circulação sanguínea, além do estresse e dos problemas respiratórios. Além disso, não podemos deixar de comentar sobre os problemas financeiros causados, o estado do Sudeste perde cerca de 3% do seu PIB por causa do trânsito, como nos mostra [3].

Neste projeto buscamos um método alternativo para resolver o problema destacado. A partir de uma fusão de dados de sensoriamento participativo deseja-se encontrar

uma relação entre o trânsito na cidade de Campinas-SP, e participação de usuários em redes sociais.

Por meio de *scripts* simples feitos em *Python* e usando *Interfaces de programação de aplicações - APIs*, disponibilizadas pelas próprias empresas (como a Google, Twitter, e.g), fizemos a parte inicial do projeto, que dizia respeito a coleta dos dados que serão utilizados no decorrer do trabalho. Depois de coletados os dados, foram feitas funções em *Python* que fossem capaz de, em primeiro lugar, tratar os arquivos obtidos, fazendo com que eles estivessem em condições de serem processados.

Em seguida, a partir dos dados processados do *Twitter* criou-se *heat maps* usando a *geolocalização* dos tweets, de modo que foi possível identificar a movimentação dos usuários da rede social. Em paralelo, também tratamos dados oriundos do *Google Maps*, a partir de um conjunto de rotas monitoradas, traçamos dois planos de uso para elas. Em primeiro lugar, montou-se uma página *HTML* simples, que exibe informações detalhadas de cada rota obtida (tempo de duração, distância, horário, e.g). Além disso, no momento, trabalha-se na criação de um *grafo*, no qual as ruas da cidade de Campinas serão suas arestas, e o tempo gasto para percorrer as rotas traçadas serão utilizados como "*peso*" para elas, a fim de criar uma visualização mais dinâmica e rápida.

Por fim, pretendemos criar um *serviço* no qual disponibilizaremos a solução criada para que todos os moradores da cidade possam utilizar. A implementação citada será explicada em mais detalhes na Seção 5.

2 MOTIVAÇÃO E OBJETIVOS

Nos dias atuais, os sistemas computacionais estão evoluindo rumo à consolidação do paradigma da computação ubíqua [4], permitindo um mundo totalmente monitorado, através de sensores que vão desde dispositivos integrados ao ambiente físico até as redes sociais. Neste contexto, estima-se que por volta do ano de 2020 a quantidade de dados gerados por estes sensores atinjam a ordem dos 25 ZB (zettabytes) [5].

Na mesma proporção em que cresce a quantidade de dados gerados por sensores, cresce também a quantidade de veículos e o trânsito nas cidades do Brasil, como pode ser visto em [6]. Por conseguinte, cada vez mais problemas de saúde surgem na vida dos moradores dessas cidades, caracterizando um claro e gravíssimo problema nacional.

Para que esta nova tendência - *da computação ubíqua* - torne-se realmente útil, e seja capaz mudar a vida das pessoas, faz-se necessário que novas aplicações sejam desenvolvidas tendo em mente a necessidade de se extrair um significado de todas estas diferentes fontes e tipos de dados, que seja relevante para a solução de problemas que afetem a vida das pessoas. Com base nesta necessidade, o objetivo deste trabalho é propor

o desenvolvimento de uma estratégia para a geração de contexto [7] através da fusão de diferentes fontes e tipos de dados sensorizados. Como prova de conceito para a viabilidade desta estratégia, será desenvolvida uma ferramenta para a realização de previsão do trânsito para o cenário de computação ubíqua.

3 FUNDAMENTAÇÃO TEÓRICA

Para realização do trabalho foi necessário adquirir alguns conhecimentos extras, que serão brevemente explicados nessa secção.

3.1 Sistemas de Transporte Inteligentes

3.1.1 Definição

Os Sistemas de Transporte Inteligentes (ITS) têm como objetivo melhorar a segurança e mobilidade dos transportes, como também o aumento da produtividade das pessoas e diminuição dos efeitos nocivos do trânsito. Essa melhoria é alcançada através da integração de tecnologias de comunicação nos veículos e na infraestrutura da cidade. As principais características, que serão apresentadas a seguir, se relacionam à arquitetura e as Redes Veiculares.

A arquitetura norte-americana (National ITS Architecture) [of Transport 2016], descreve como ocorre a comunicação entre seus elementos e subsistemas. Ela se divide em 4 classe: *Center* que define o centro de controle e gerenciamento de todo o sistema, no qual os serviços são executados; *Field* que engloba toda a parte de infraestrutura do ambiente (RSU, sensores de monitoramento, câmeras); *Vehicles* que são os veículos e sensores embarcados; e os *Travelers* que define-se pelos dispositivos usados pelas pessoas durante a viagem.

Já as *Redes Veiculares* são um tipo de rede emergente que tem atraído o interesse de muitos grupos de pesquisa. Estas redes são formadas por veículos com capacidade de processamento e comunicação sem fio, trafegando em ruas e rodovias, enviando e recebendo informações de outros veículos.

3.1.2 Aplicações

Os Sistemas de Transporte Inteligentes possuem uma série de aplicações práticas, de maneira simplificada, podemos caracterizá-las como:

- Aplicações de segurança

- Aplicações de eficiência de tráfego
- Aplicações de entretenimento e conforto
- Aplicações de sensoriamento urbano

As aplicações podem ser vistas com mais detalhes em [8].

3.2 Mapas de Calor

3.2.1 Definição

A melhor forma de entender um mapa de calor é pensar em uma tabela ou planilha a qual contém cores ao invés de números. O gradiente de cor padrão configura o menor valor no mapa de calor como azul escuro, o maior valor como um vermelho brilhante e valores medianos como cinza claro, com uma transição correspondente (ou gradiente) entre estes extremos. Os mapas de calor são bastante apropriados para visualizar grandes quantidades de dados multidimensionais e podem ser utilizados para identificar grupos de linhas com valores similares, conforme elas são mostradas nas áreas de cor similar.

3.2.2 Gmplot

O *Gmplot* é uma interface do tipo *matplotlib* para gerar o HTML e o javascript necessários para renderizar todos os dados que você deseja sobre o *Google Maps*. Vários métodos de plotagem tornam a criação de visualizações de mapas exploratórios sem esforço. Um dos modos permitidos é o de criação de *heatmaps*, que foi usado nesse trabalho. Seu uso será detalhado na seção 5.

3.3 Grafos

3.3.1 Definição

Um grafo é um animal formado por dois conjuntos: um conjunto de coisas chamadas vértices e um conjunto de coisas chamadas arestas; cada aresta está associado a dois vértices: o primeiro é a ponta inicial da aresta e o segundo é a ponta final. Você pode imaginar que um grafo é um mapa rodoviário idealizado: os vértices são cidades e as arestas são estradas (que podem ser de mão única ou dupla, dependendo do tipo de grafo utilizado). Para este trabalho faremos uma comparação um pouco diferente, enxergaremos uma cidade como um grafo, no qual a intersecção entre as ruas são os vértices, e as ruas serão arestas, como destacaremos na seção 5. Para gerarmos nosso grafo usamos duas ferramentas cruciais, o *OpenStreetMap* e o *Simulator of Urban MObility - SUMO*, que serão brevemente introduzidos nessa seção.

3.3.2 OpenStreetMap

O OpenStreetMap¹ é um serviço de mapeamento construído de maneira colaborativa por usuários, profissionais e entusiastas. O OpenStreetMap permite selecionar uma região específica a partir da qual o usuário deseja exportar informações de mapeamento. Esses arquivos de mapeamento podem ser processados por outras ferramentas. A partir deles geramos os *nós(vértices)* do nosso grafo.

3.3.3 Simulator of Urban MObility - SUMO

Pesquisadores e profissionais da comunidade de Sistemas de Transportes Inteligentes utilizam simuladores de tráfego como forma de estudar o impacto de algoritmos de roteamento de veículos, alterações no controle de semáforos e mudanças na infraestrutura viária antes das mesmas serem implementadas no mundo real. Um simulador de tráfego amplamente utilizado pela comunidade é o *SUMO*². O *SUMO* fornece um conjunto de ferramentas e bibliotecas que tem o objetivo de facilitar o desenvolvimento dos mais variados tipos de cenários, possibilitando o estudo de várias questões relacionadas ao tráfego de veículos e pedestres. No nosso trabalho o usamos em conjunto com o *OpenStreetMap* para conectar os vértices usando ruas como *arestas*.

4 TRABALHOS RELACIONADOS

O deslocamento constante das pessoas, em busca de bens e serviços de qualidade, empregos e outras situações, vem acarretando cada vez mais problemas inerentes a mobilidade urbana, tais como, longos engarrafamentos, aumento da poluição e prejuízos financeiros, principalmente nas regiões metropolitanas, devido a grande concentração de veículos nessas áreas [3]. A fim de minimizar tais problemas, os planejadores das cidades efetuam recorrentes melhorias na infraestrutura das vias, contudo, o alto custo e o longo tempo das obras são fatores que inviabilizam sanar esses problemas. Alternativamente, os Sistemas de Transportes Inteligentes (ITS na sigla em inglês) vêm sendo explorados para operar e prestar melhores serviços, e assim, garantir que a mobilidade urbana ocorra de forma mais eficiente, sem resultar em altos custos de implantação, tal prática já vem, inclusive, sendo usadas em alguns países, como mostra [9].

O uso de dados providos de redes sociais (como o Twitter, Instagram, Foursquare, etc) para prever situações do mundo real tem se tornado cada vez mais comum nos dias de hoje. Afim de compreender a dinâmica das cidades, em [10], os autores exploram dados compartilhados no Foursquare para identificar limites culturais e preferências alimentares.

¹<https://www.openstreetmap.org/>

²<http://sumo.dlr.de/index.html>

Em trabalhos como o de Gomide [11], no qual ele analisou como epidemias de dengue se refletem no uso do Twitter pelas pessoas das regiões afetadas pela doença. Ou ainda, Bollen [12], que também usou do *feed* do Twitter para analisar como o humor das pessoas influenciava no valor da *Dow Jones Industrial Average* (DJIA) ao longo do tempo. Já em [13] usuários do Twitter foram usados como sensores para identificar o acontecimento de terremotos.

No âmbito da mobilidade urbana, trabalhos como os citados também se popularizam. Trabalhos como [14], correlacionam mensagens compartilhadas por usuários do redes sociais com condições de tráfego para identificar padrões no comportamento das pessoas através de dados do Foursquare e do Instagram, as quais são correlacionadas com as condições de tráfego³ e foi possível identificar uma maior atividade das pessoas nos períodos de pico de movimentação, nos mostrando uma relação entre a movimentação das pessoas e a atividade em redes sociais. Similarmente, em [15] vemos uma série de exemplos de aplicações nas quais aparelhos celulares transmitem, por meio de *Near-Field Communication* (NFC) passageiros recebem, aproximando o aparelho celular do ônibus, informações cruciais à respeito do mesmo, como todos os horários e pontos nos quais aquela linha passou, entre outras. Auxiliando a passageiros que não possuem muito conhecimento sobre a rede da cidade, ou ainda aqueles que realizam muitas trocas de linhas.

5 PROPOSTA

Nossa proposta de trabalho pode ser dividida em etapas, que serão comentadas separadamente. O cronograma das etapas pode ser visto na Figura 1. Até o momento, já foram realizadas, como previsto no cronograma as seguintes atividades:

5.1 Revisão bibliográfica

A primeira parte do trabalho consistiu em um estudo bibliográfico de artigos científicos relacionados com os tópicos desse trabalho. A grande maioria dos artigos lidos, os mais significativos, ao meu ver, foram citados nesse trabalho na seção 4. Os demais, ainda que tenham dados bons *insights* para o desenvolvimento do projeto, eram, em sua grande maioria, de escopos muito diferentes para com o nosso.

5.2 Definição da estratégia de coleta e mineração

Os dados aqui obtidos são parte crucial para o projeto. Por isso, em um primeiro momento, dedicou-se os esforços a traçar uma estratégia adequada, definindo as fontes de

³fornecidas pelos mapas do Bing

informação e os espaços a serem monitorados. Para o primeiro, optou-se pelo *Twitter* e o *Google Maps*. Para a primeira fonte, decidiu-se a macro região de Campinas-SP como espaço a ser monitorado, enquanto que para o *Maps* as avenidas escolhidas podem ser vistas na Figura 2.

Primeiro, iremos introduzir a metodologia de trabalho para o *Twitter*, usou-se um código em Python que realiza requisições por meio da API (Interface de Programação de Aplicativos) disponibilizada pela própria empresa. Esta é responsável por conectar o código ao servidor do Twitter, e produzir uma stream com os dados, então filtramos essa stream para armazenarmos somente os dados que nos interessam (nesse caso usamos as coordenadas da cidade de Campinas - SP como filtro da stream). Os dados filtrados são salvos em um arquivo ".json", onde cada tweet é um objeto do JSON. Uma vez obtido os dados, partimos para uma segunda etapa do projeto, na qual foi feito um pré-processamento dos dados obtidos pelo Twitter. Nesta etapa, primeiro realizamos uma "leitura" dos dados salvos no arquivo JSON, separando os tweets em uma espécie de dicionário, no qual cada lista de coordenadas obtidas estava associada ao dia no qual ela foi coletada. Uma vez separados em dias de coleta, os dados passaram por uma análise que buscava encontrar possíveis bots (aplicação de software concebido para simular ações humanas repetidas vezes de maneira padrão). Excluídos os bots, fez se um plot do gráfico de calor produzido pelas coordenadas para cada dia no qual a coleta foi realizada. O resultado desse mapa pode ser visto na Figura 3.

Para os dados coletados junto ao Google Maps também usamos um código em Python que realiza requisições por meio de uma API disponibilizada pela própria empresa. Nesta informamos coordenadas de pontos iniciais e finais (isto é, origem e destino) que nos interessam e temos, como retorno, um arquivo ".json" que nos informa, para cada requisição feita, o tempo de viagem entre os pontos informados. O trabalho com os dados foi parecido com o feito com os do Twitter. Primeiro, realizamos uma "leitura" dos dados salvos no arquivo JSON, separando-os em uma espécie de dicionário, no qual cada rota foi separada junto a informações importantes sobre elas, como distância, tempo de viagem e data. A partir desses dados foi criada uma página HTML com as informações de cada rota coletada, um exemplo dessa página pode ser vista na Figura 4.

5.3 Definição da estratégia de geração de contexto

Fase atual de desenvolvimento do projeto, uma vez que já temos os scripts de coleta e mineração de dados funcionando. Para geração de contexto, concentraremos nossos esforços na criação de um grafo que representará a malha viária da cidade de Campinas-SP.

Para gerar o grafo, seguiremos os seguintes passos. Usando o *OpenStreetMaps*, im-

portaremos o mapa da cidade em um arquivo *.osm*, este será usado para gerar um arquivo *.net.xml*, a partir do simulador *SUMO*, ambos os arquivos serão cruciais na esquematização do grafo. A partir do arquivo *.osm* extrairemos os *nós*, enquanto que as arestas serão obtidas do arquivo *.net.xml*. Utilizaremos as informações das rotas sugeridas pelo *Google Maps* para ponderar as arestas do grafo com a velocidade média (distância/tempo). Dessa forma, saberemos qual é o padrão de mobilidade nas regiões das arestas que serão estudadas, e será possível saber o impacto que algum evento anômalo ocasiona na mobilidade dessas áreas.

5.4 Avaliação e validação da estratégia

A avaliação e validação da estratégia deve acontecer uma vez que a ferramenta esteja totalmente funcional, isto é, visto que já possuímos material necessário para coletar e minerar os dados, uma ferramenta poderá ser disponibilizada após a criação do grafo. Uma vez que o serviço esteja disponível para o público em geral será possível colher *feedbacks* a fim de validar a qualidade e funcionalidade do projeto.

5.5 Escrita de relatório e documentação

Ao fim do projeto pretendemos concentrar esforços na escrita de relatórios e na documentação do trabalho realizado até aqui, a fim de facilitar o uso do que foi desenvolvido por outras pessoas nos mais variados tipos de projetos.

Para o segundo semestre desse ano (2018), planeja-se o a criação e submissão de artigos em conferências de temas relacionados ao desenvolvido nesse projeto. Para que tal etapa se torne realidade, falta apenas a obtenção de resultados mais concretos, isto é, terminada a implementação do grafo representado a malha viária de Campinas-SP, teremos material suficiente para a submissão de tais artigos.

6 CONCLUSÕES

Com a implementação, execução e análise dos resultados obtidos até o momento, já é possível perceber uma relação entre os dados obtidos nos *scripts* de coleta, e o comportamento das pessoas no dia a dia.

Durante a análise dos *heatmaps* feitos a partir da geolocalização dos tweets, foi possível identificar um claro padrão no deslocamento das pessoas. Para os dados do *Google Maps* ainda não foi possível tirar muitas conclusões, visto que a principal fonte de análise, o grafo, ainda não foi implementada.

ANEXOS

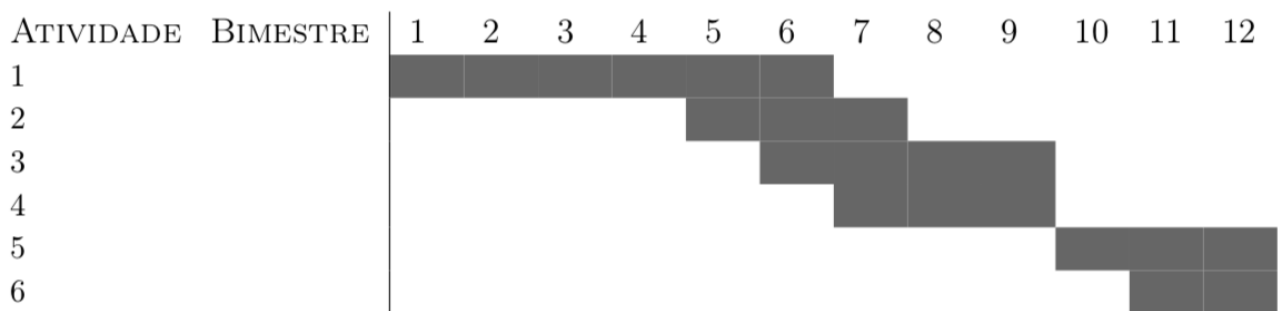


Figura 1: Cronograma de Atividades

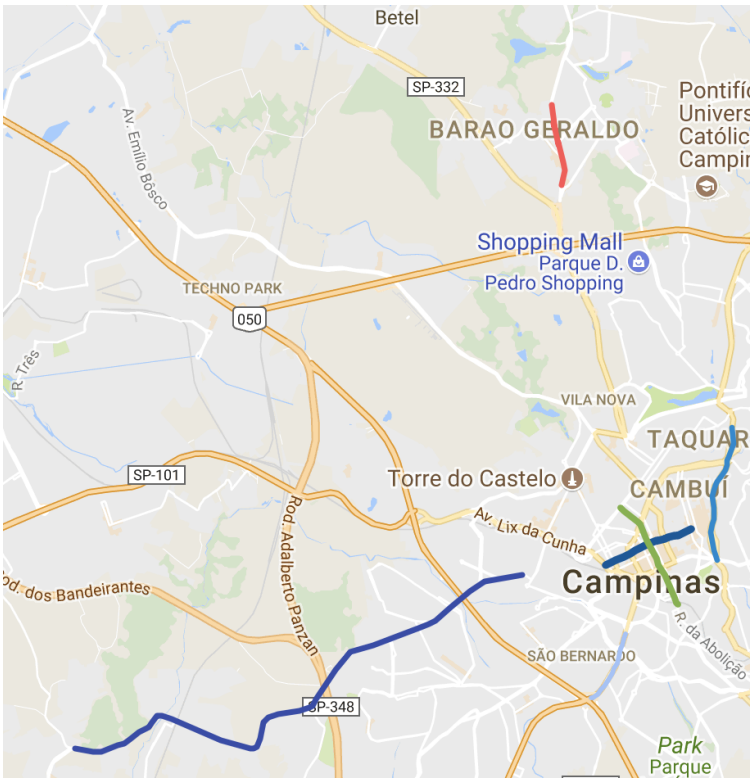


Figura 2: Avenidas monitoradas pelo *Google Maps*



Figura 3: *Heatmap* gerado a partir da geolocalização dos tweets

About trip:

Index:	Points:	Distance (m):	Time of trip (s):
1	-22.83324 , -47.0788791	916	91
2	-22.8251395 , -47.0797617		

GMT Time: 2018-01-21 17:48:13.957259-02:00

Duration in traffic (s): 91

Distance (m): 916

Figura 4: Página html feita com os dados obtidos do *Google Maps*

REFERÊNCIAS

- [1] Disponível em: <https://www.statista.com/outlook/digital-markets>, Acesso em: 17-04-2018.
- [2] Disponível em: <https://www.omnicoreagency.com/twitter-statistics/>, Acesso em: 17-04-2018.
- [3] Guilherme Szczerbacki Besserman Vianna and Carlos Eduardo Frickmann Young. Em busca do tempo perdido; uma estimativa do produto perdido em trânsito no brasil. *Revista de Economia Contemporânea*, 19(3):403–416, 2015.
- [4] Mark Weiser. The computer for the 21 st century. *Scientific american*, 265(3):94–105, 1991.
- [5] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. Context aware computing for the internet of things: A survey. *IEEE communications surveys & tutorials*, 16(1):414–454, 2014.
- [6] Andrade Rodrigo de Oliveira Machado-Filho Moysés Floriano Antenor, Samuel. Trânsito e aumento da frota de veículos tornam vulneráveis a saúde nas cidades. *Ciência e Cultura*, 62:8 – 10, 10 2010.
- [7] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International Symposium on Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.
- [8] Felipe Domingos da Cunha, Guilherme Maia, Clayson Celes, Daniel Guidoni, Fernanda de Souza, Heitor Ramos, and Leandro Villas. Sistemas de transporte inteligentes: Conceitos, aplicações e desafios. In *SBRC 2017 - Minicursos ()*, may 2017.
- [9] Disponível em: <http://www.its-portugal.com/its>, Acesso em: 28-06-2018.
- [10] Thiago H Silva, Pedro OS Vaz de Melo, Jussara M Almeida, Mirco Musolesi, and Antonio AF Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. 2014.
- [11] Janaína Gomide, Adriano Veloso, Wagner Meira Jr, Virgílio Almeida, Fabrício Bevenuto, Fernanda Ferraz, and Mauro Teixeira. Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proceedings of the 3rd international web science conference*, page 3. ACM, 2011.
- [12] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

- [13] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
- [14] Anna Izabel João Tostes Ribeiro, Thiago Henrique Silva, Fátima Duarte-Figueiredo, and Antonio AF Loureiro. Studying traffic conditions by analyzing foursquare and instagram data. In *Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks*, pages 17–24. ACM, 2014.
- [15] Victor Nassar and Milton Luiz Horn Vieira. teste. *urbe. Revista Brasileira de Gestao Urbana*, 9:327 – 340, 08 2017.