

# Sensoriamento Participativo de Regiões de Interesse com Descrição Adaptativa das Taxas de Amostragem

Carlos Henrique de O. M. André<sup>1</sup>, Dianne S. V. de Medeiros<sup>2</sup> e Miguel Elias M. Campista<sup>1</sup> \*

<sup>1</sup>Universidade Federal do Rio de Janeiro - PEE-COPPE/Del-Poli/GTA

<sup>2</sup>Universidade Federal Fluminense - PPGEET/MídiaCom

{choma,miguel}@gta.ufrj.br, diannescherly@id.uff.br

**Abstract.** *Participatory Sensing (PS) is a new paradigm of collaborative networks which provides incentives for users to participate of sensing tasks on a Region of Interest (RoI). A challenge in wireless networking, however, is to balance the amount of data collected by each user so as not to impose an excessive load to the network. In this direction, this work proposes a centralized system to adapt the sample rate assigned to each participating sensor. The sample rate is computed as a function of the sample variation collected in a given RoI in the last time interval. The results obtained via simulations show a tradeoff between the sample rate and the number of participating users. The more participating users, the lower must be the individual sample rate and the lower will be the amount of data transferred. This strategy, even though requires a larger number of sensors, can increase the data delivery rate taking into account the available short contact times.*

**Resumo.** *O Sensoriamento Participativo (Participatory Sensing – PS) é um novo paradigma de redes colaborativas que incentiva a participação dos usuários no sensoriamento de uma Região de Interesse (Region of Interest – RoI). Um desafio em redes sem-fio, porém, é calibrar a quantidade de dados sensorizados por cada usuário para que a carga imposta à rede não se torne excessiva. Nessa direção, este trabalho propõe um sistema centralizado capaz de adaptar a taxa de amostragem a ser atribuída a cada sensor participante. A taxa de amostragem é calculada em função da variabilidade das amostras coletadas em uma dada RoI a partir das amostras recebidas no último intervalo de tempo. Os resultados obtidos através de simulação mostram o compromisso entre a taxa de amostragem e o número de sensores participantes. Quanto mais nós participantes, menor pode ser a taxa de amostragem individual e menor pode ser a quantidade de dados a serem transferidos. Essa estratégia, apesar do maior número de sensores, aumenta a taxa de entrega dos dados tendo em vista os curtos intervalos de tempo de contato disponíveis.*

## 1. Introdução

O paradigma de sensoriamento participativo (*Participatory Sensing – PS*) [Xiao et al. 2013] vem surgindo como uma forma de escalar sistemas de sen-

---

\*Os autores agradecem o apoio da CAPES, do CNPq, da FAPERJ e da FAPESP. Este trabalho foi parcialmente financiado pelos processos nº 15/24494-8 e nº 15/24490-2, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

soriamento usando recursos de usuários, como *smartphones*. Para isso, é necessário que os usuários cedam os seus dispositivos para a realização de tarefas de sensoriamento, mediante incentivos. Essas tarefas são escalonadas por alguma entidade capaz de identificar uma Região de Interesse (*Region of Interest – RoI*) para a coleta de dados. O sensoriamento participativo tem como principal benefício o aumento da taxa de dados amostrados de uma dada RoI a um baixo custo. As aplicações inseridas neste paradigma podem, portanto, aprimorar os seus serviços a partir de um número grande de fontes de dados, que se torna mais viável através da participação dos usuários. Um exemplo é o Waze, que oferece informações sobre as condições do trânsito usando dados dos próprios usuários [Ribeiro Neto et al. 2016].

O número elevado de fontes de dados pode impor desafios em termos de largura de banda, devido ao grande volume de dados gerados; acessibilidade, já que os dispositivos devem ser capazes de acessar a Internet mesmo estando espalhados em regiões geo-distribuídas; e mobilidade, já que as fontes podem pertencer a usuários móveis. Todos esses desafios podem ser encontrados simultaneamente nas redes ad hoc veiculares (VANETs) [Liu et al. 2013] que, na era da Internet das Coisas, evoluíram para o novo paradigma da Internet dos Veículos (*Internet of Vehicles – IoV*) [Miche and Bohnert 2009]. Na IoV, além dos usuários, sensores inteligentes e atuadores ainda podem ser instalados na infraestrutura veicular [Yu et al. 2013], aumentando ainda mais o ecossistema de sensoriamento. Assim, os serviços associados dependem de uma grande coleção de nós geo-distribuídos, sejam eles móveis ou não, que são responsáveis por sensoriar e enviar os dados coletados para uma entidade centralizadora, possivelmente na Internet. A quantidade de dados gerados, as restrições dos dispositivos e as dificuldades para transferir esses dados para a Internet representam um compromisso que pode se tornar um importante obstáculo ao sensoriamento participativo. É fundamental, portanto, adaptar a taxa de amostragem de cada nó participante para que o sensoriamento se torne mais inteligente.

Muitos sistemas de sensoriamento já foram propostos na literatura com intuítos diversos. O *SensingBus*, por exemplo, refina o monitoramento das condições climáticas de uma cidade a partir de sensores embarcados em ônibus urbanos [Cruz et al. 2017]. A ideia é um contraponto às soluções mais triviais que seriam aumentar o número de estações meteorológicas ou o poder de coleta de cada uma delas. Como as duas opções implicam investimentos, propôs-se adicionar aos dados já existentes, aqueles coletados pelos ônibus. A área monitorada é gratuitamente expandida através da mobilidade dos ônibus. Mohan et al. [Mohan et al. 2008] propuseram o Nericell, que usa sensores embarcados em *smartphones*, tais como acelerômetros, microfones e GPS, para determinar as condições da via (ruas e estradas) e do tráfego. Zhou et al. [Zhou et al. 2012] propuseram um sistema baseado em colaboração para prever a posição dos ônibus ao longo do tempo durante o percurso, utilizando a potência entre as torres de celular e os dispositivos móveis. Uma característica em comum dos trabalhos anteriores é que nenhum deles se preocupa com a taxa de amostragem utilizada. A aplicação As-air [Zeng and Xiang 2017], diferentemente, adapta a taxa de amostragem de acordo com as características do ambiente, que varia em dias de semana, finais de semana e feriados. O As-air é usado para medir a qualidade do ar que pode variar, necessitando de taxas de amostragem mais elevadas em dias de atividades mais intensas. Já Weinschrott et al. aumentam a taxa de amostragem por questões de robustez [Weinschrott et al. 2011]. Taxas de amostragem maiores são usadas em regiões onde se deseja aumentar a probabilidade de detecção de um evento de

interesse. Entretanto, até onde se sabe, nenhum dos trabalhos anteriores adapta a taxa de amostragem considerando tanto aspectos espaciais quanto temporais.

Este trabalho propõe um sistema de sensoriamento participativo, que diferente dos trabalhos anteriores [Zeng and Xiang 2017, Weinschrott et al. 2011], adapta a taxa de amostragem considerando tanto aspectos espaciais quanto temporais. Para isso, assume-se a existência de um nó centralizador, na nuvem, com capacidade elástica, que recebe os dados coletados pelos participantes do sistema de sensoriamento. Esses dados são utilizados para calcular a variação das medidas coletadas em uma dada região de interesse (RoI) durante um período de tempo. Neste trabalho, assume-se que o sistema é usado para avaliação das condições de trânsito de uma cidade e, portanto, são coletadas as velocidades dos veículos participantes. Caso haja uma grande variação nas medidas coletadas em um dado trecho (ruas ou estradas) da RoI, considera-se que é mais provável existirem medidas errôneas nesse trecho. Portanto, o sistema deve aumentar a taxa de amostragem para que sejam coletadas mais amostras apenas desses trechos. De forma semelhante, caso a variação no trecho seja baixa, considera-se a possibilidade de redução da taxa de amostragem. Dessa forma, evita-se esbarrar no compromisso entre o número de amostras coletadas, a carga de dados gerada e as limitações dos dispositivos dos usuários, por exemplo, a quantidade de energia disponível. As análises da proposta utilizam o conjunto de dados de movimentação da frota de ônibus da cidade de Seattle, nos Estados Unidos [Jetcheva et al. 2003]. Desse conjunto de dados, analisa-se um dia inteiro, assumindo que o valor sensoriado é a velocidade dos ônibus. Os resultados mostram que maiores cargas de dados só são possíveis de serem entregues com poucos nós competindo pelo meio e que, caso mais nós participem do sistema, é recomendável reduzir a carga de dados individuais para se conseguir maiores taxas de entrega. Os resultados ainda indicam que a amostragem adaptativa é capaz de reduzir os erros sem impactar na carga de dados.

Este trabalho está organizado da seguinte forma. A Seção 2 introduz noções de sensoriamento participativo. A Seção 3 descreve a arquitetura de rede assumida neste trabalho e a propõe um sistema de sensoriamento com taxa de amostragem adaptativa. A Seção 4 apresenta o conjunto de dados utilizado e o caracteriza. A Seção 5 apresenta a análise da proposta. Por fim, a Seção 6 conclui este trabalho e lista direções futuras.

## **2. Visão Geral do Sensoriamento Participativo**

Atualmente, o poder de sensoriamento embarcado em veículos ou mesmo em dispositivos móveis pessoais, como *smartphones*, vem aumentando a passos largos. Tal aumento é consequência da evolução da tecnologia móvel e da miniaturização e barateamento dos componentes eletrônicos. A partir dos dados coletados por esses sensores é possível vislumbrar uma série de aplicações que oferecem noções de contexto ou do ambiente em que o usuário se encontra [Melo et al. 2014]. Quando os dispositivos são usados em conjunto, combinando-se os dados coletados, é possível construir um sistema de sensoriamento mais poderoso, escalável e de baixo custo. Como consequência as aplicações podem ser mais sofisticadas, resultando inclusive em aplicações úteis para as chamadas Cidades Inteligentes [Melo et al. 2014]. O sensoriamento através da participação coletiva dos usuários e seus dispositivos são o foco de pesquisa do paradigma conhecido como Sensoriamento Participativo (*Participatory Sensing* – PS).

O sensoriamento participativo é composto de tarefas implementadas nos dispositi-

vos móveis que possibilitem a coleta, a análise e o compartilhamento do conhecimento gerado em uma região de interesse (RoI) [Burke et al. 2006]. Uma das principais vantagens desse tipo de sensoriamento é o aumento do conhecimento sobre a região monitorada, como consequência do compartilhamento dos dados sensoriados. Além disso, é possível ampliar o raio de cobertura da RoI, devido à mobilidade dos sensores. Os dados obtidos pelo coletivo precisam ser processados, possivelmente em uma entidade externa. Assim, existe um crescente uso da capacidade da nuvem para realização tanto do processamento quanto do armazenamento do conhecimento obtido para a RoI [Estrin et al. 2010]. Várias aplicações podem ser desenvolvidas segundo este paradigma e em diferentes áreas como: gestão de recursos naturais [McCall and Minang 2005], planejamento e monitoramento urbano [D'Hondt et al. 2013, Mohan et al. 2008] e saúde pública [Boulos et al. 2011].

A participação do usuário na coleta e transferência de dados deve ser estimulada através de mecanismos de incentivo. Nos sistemas de sensoriamento participativo, dependendo da recompensa, um usuário pode se sentir motivado a coletar um grande volume de dados em um curto intervalo de tempo. Por exemplo, um usuário que informe a velocidade do seu veículo pode receber em troca descontos em serviços ou até mesmo anúncios específicos de promoções conforme seu perfil. O benefício pode ocorrer também na forma de retorno do conhecimento coletivo, como a informação de trânsito no Google Maps, ou outros benefícios, como no Waze. Para não exaurir os recursos do usuário, que poderia levar à exclusão ou à não obtenção de um dado importante para o sistema, deve existir um compromisso entre a quantidade e a qualidade dos dados coletados.

Apesar de o sensoriamento participativo poder ser utilizado em diversos cenários, neste trabalho aborda-se a sua aplicação em redes veiculares, explorando as áreas de planejamento e monitoramento urbano. Nesse cenário, o sensoriamento participativo pode ser utilizado, por exemplo, para planejar uma nova rota de deslocamento, instalação de novos pontos de parada e planejamento de expansões viárias. Os dados são inseridos no sistema através dos próprios usuários, cuja participação é fundamental. No entanto, essa participação pode provocar problemas para a rede e para o próprio sistema. Para a rede, se o volume de dados coletados durante o trajeto em uma região for muito grande, transferi-los durante contatos curtos entre as Unidades de Bordo (*On Board Units* – OBUs) e as Unidades de Acostamento (*Road Side Units* – RSUs) pode não ser possível. Caso os dados não sejam recebidos, o próprio sistema estará comprometido.

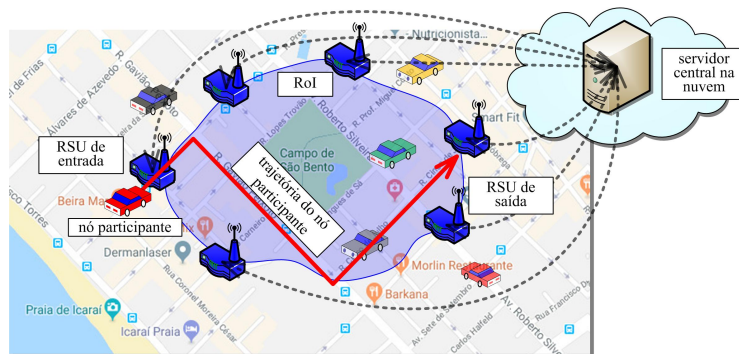
O estudo realizado neste trabalho utiliza o conjunto de dados da cidade de Seattle (Seção 4), no qual os participantes do sistema de sensoriamento são os ônibus da cidade. Outros conjuntos de dados podem ser utilizados sem perda de generalidade. Os participantes do sistema de sensoriamento são os próprios ônibus que circulam pela cidade, e todos eles medem a mesma variável física: a velocidade.

### **3. Proposta de Sensoriamento Participativo com Amostragem Adaptativa**

Esta seção apresenta a arquitetura da rede usada neste trabalho e descreve a operação do sistema de sensoriamento proposto.

#### **3.1. Arquitetura da rede veicular**

A arquitetura da rede veicular usada neste trabalho é composta por um servidor central, por pontos de acesso (*RoadSide Units* - RSUs) e por usuários do sistema de sensoriamento participativo equipados com unidades de bordo (*OnBoard Units* - OBUs). O



**Figura 1. Arquitetura da rede veicular usada neste trabalho. Os usuários participantes do sistema de sensoriamento iniciam a coleta de dados ao entrar na RoI e descarregam os dados sensorizados na RSU de saída.**

servidor central está alocado na nuvem e, portanto, possui uma capacidade elástica, não consistindo em um gargalo para o sistema. O servidor central se serve dos dados coletados pelos participantes para oferecer serviços a usuários externos. Por exemplo, neste trabalho considera-se que o servidor central pode oferecer informação das condições do trânsito a usuários externos. As RSUs estão conectadas à Internet através de uma rede de acesso normalmente cabeada. As OBU's não são capazes de se comunicarem utilizando infraestrutura celular e, portanto, alcançam o servidor central apenas através da comunicação com as RSUs, feita através de redes sem-fio IEEE 802.11p. A Figura 1 ilustra a arquitetura da rede veicular.

A Figura 1 mostra um nó participante que, ao entrar na Região de Interesse (*Region of Interest* - RoI), tem a oportunidade de receber solicitações para contribuição com o sistema de sensoriamento participativo. Tal solicitação é enviada pelo servidor central e entregue ao usuário através de uma RSU na entrada da RoI. Note que por questões de simplicidade, uma RoI é formada por trechos de ruas e estradas delimitadas por RSUs. Além disso, visando generalidade, não existem RSUs dentro da RoI, por isso a atualização da taxa não ocorre o tempo todo. Ao sair da RoI, o usuário participante do sistema de sensoriamento entrega os dados coletados ao servidor central através de uma RSU na saída. Munido de diferentes contribuições da mesma RoI, o servidor central pode consolidar os dados recebidos para possível oferta de serviços a usuários externos. Durante o período que o usuário permanece na RoI, ele realiza as medidas periodicamente conforme definido pelo servidor. Neste trabalho, a questão do incentivo não é abordada, mas assume-se que a participação no sistema de sensoriamento seja de alguma forma recompensada.

Um dos problemas do sensoriamento participativo é cobrir uma RoI de forma eficiente sem sobrecarregar os participantes do sistema, nem o próprio sistema. Uma possível solução para esse problema é aumentar a largura de banda para transmissão dos dados coletados ou o número de RSUs, o que aumentaria a granularidade dos locais de descarregamento dos dados. Consequentemente, os participantes do sistema precisariam descarregar menos dados em cada RSU. No entanto, isso onera o custo de implantação e manutenção do sistema e esbarra em outro problema, que seria a localização ideal dessas RSUs. Em vez disso, este trabalho propõe utilizar uma taxa de amostragem adaptável às condições dos diferentes trechos da RoI. A escolha da taxa depende da variabilidade dos dados amostrados, de forma que pode ser necessário aumentar ou diminuir essa taxa

nos diferentes trechos da mesma RoI ao longo do tempo. Isso, obviamente, é importante caso se deseje minimizar em toda a RoI o erro da amostragem obtido a partir dos dados consolidados. A dinamicidade da amostragem deve ser levada em conta para que: (i) a RoI seja totalmente sensoriada apresentando resultados com baixo erro em comparação à medida real e que (ii) não haja nem sobrecarga dos usuários participantes do sistema de monitoramento e nem da rede sem-fio ou da rede de acesso ao servidor central.

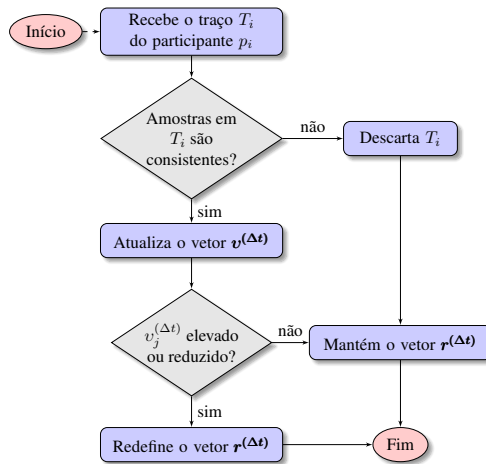
### 3.2. Operação do sistema de sensoriamento proposto

Este trabalho propõe um sistema de sensoriamento participativo cuja taxa de amostragem é adaptada ao erro das medidas nos diferentes trechos da mesma RoI, considerando tanto características espaciais quanto temporais. Para isso, o servidor central deve, periodicamente, inferir o erro da amostragem para cada trecho dentro da mesma RoI para que os próximos usuários do sistema de monitoramento possam adaptar suas taxas de amostragem conforme a necessidade dos trechos. Assim, denota-se o conjunto de participantes do sistema de sensoriamento como  $\mathcal{P} = \{p_1, \dots, p_n\}$ , onde  $p_i$  é um usuário participante e  $n$  é o número total de usuários no sistema. Cada usuário  $p_i$  coleta amostras e as registra em um traço de dados  $T_i$ . Cada amostra é uma tupla contendo posição e instante e valor da medida amostrada. No caso estudado, a medida amostrada é a velocidade do veículo. Ressalta-se que o sistema de sensoriamento proposto não perde generalidade caso outro sensor ou outros cenários sejam usados.

O servidor central, ao receber o trace  $T_i$ , verifica se as amostras possuem algum dado inconsistente que podem ser gerados a partir de falhas do sistema de sensoriamento ou por algum tipo de ação maliciosa<sup>1</sup>. Assumindo que todas as amostras são válidas, o servidor central passa para uma etapa de atualização da variação dos valores contidos nas amostras recebidas por trecho da RoI. A variação é usada como forma de inferir o erro nas medidas em cada trecho, já que este último não pode ser obtido diretamente. Neste trabalho, uma RoI  $\mathcal{R}$  é representada por um conjunto de trechos, ou seja,  $\mathcal{R} = \{\rho_1, \dots, \rho_m\}$ , onde  $\rho_j$  é um trecho e  $m$  é o número de trechos em  $\mathcal{R}$ . Ainda, denota-se como  $\mathbf{v}^{(\Delta t)} = \langle v_1^{(\Delta t)}, \dots, v_m^{(\Delta t)} \rangle$  o vetor atualizado da variação das amostras calculadas por trecho  $\rho_j$  referente ao intervalo de tempo  $\Delta t$ . Observe que  $v_j^{(\Delta t)}$ , correspondente à variação no trecho  $\rho_j$ , é calculada considerando amostras recebidas em traços anteriores no mesmo intervalo  $\Delta t$  e mais as amostras recebidas no último traço. Portanto, caso exista pelo menos um trecho com variação elevada ou reduzida, a taxa de amostragem ao menos nesse trecho deve ser mais bem adaptada. Caso a variação esteja elevada, deve-se aumentar a taxa de amostragem; ou reduzi-la, caso a variação esteja baixa. Define-se então o vetor  $\mathbf{r}^{(\Delta t)} = \langle r_0^{(\Delta t)}, \dots, r_m^{(\Delta t)} \rangle$  que contém as taxas de amostragem recomendadas para todos os trechos em  $\mathcal{R}$  no intervalo de tempo  $\Delta t$ . Novamente, destaca-se a correspondência entre o trecho  $\rho_j$  e a sua respectiva taxa de amostragem  $r_j^{(\Delta t)}$ . A Figura 2 ilustra todas as etapas descritas, executadas no servidor central para atualização do vetor de variação nos trechos  $\mathbf{v}^{(\Delta t)}$  e do vetor de taxas de amostragem por trecho  $\mathbf{r}^{(\Delta t)}$ .

Um usuário  $p_i$  do sensoriamento participativo, ao entrar na RoI através de uma RSU de entrada, tem a oportunidade de receber o vetor  $\pi$  contendo a descrição do sensoriamento a ser realizado em toda RoI no intervalo  $\Delta t$ , como pode ser visto na Figura 1. O vetor é composto por uma sequência de tuplas que descrevem os trechos da RoI e as

<sup>1</sup>O problema da consistência dos dados já foi atacado em trabalho anterior [André et al. 2017].



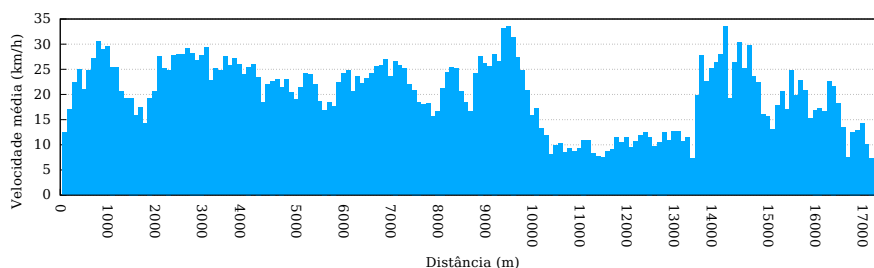
**Figura 2. Fluxograma do processo executado no servidor central ao receber um novo traço  $T_i$  de um participante  $p_i$ .**

respectivas taxas de amostragem recomendadas. O participante, ciente de sua trajetória, utiliza apenas as taxas de amostragem referentes aos trechos que vai atravessar. Caso o sistema seja ciente da trajetória do participante  $p_i$ , o vetor  $\pi$  pode ser resumido a apenas os trechos a serem seguidos por  $p_i$ . Mesmo nesse caso, não necessariamente todos os trechos da trajetória de  $p_i$  precisam estar contidos no vetor  $\pi$ . Caso a taxa de amostragem de dois trechos consecutivos seja a mesma, o sistema pode agregar dois trechos em uma única tupla. O vetor  $\pi$  é enviado ao usuário  $p_i$  através de uma estrutura de dados. Como comentado anteriormente, cada trecho em uma tupla de  $\pi$  pode ser indicado através da sua posição geográfica de início e término. Note que um trecho agregado também pode ser representado da mesma forma. As amostras sensoriadas são descarregadas através da RSU de saída da RoI. O servidor, ao receber as novas amostras, atualiza ou não o vetor  $r^{(\Delta t)}$ , usado no intervalo de tempo  $\Delta t$ . A seguir, o conjunto de dados utilizado na avaliação é descrito e, em seguida, os resultados são apresentados.

#### 4. Conjunto de Dados

A avaliação da proposta é feita através da análise do impacto da variação das taxas de amostragem no conjunto de dados Ad Hoc City [Jetcheva et al. 2003], que contém vários traços diários de um mês de movimento da frota de ônibus de passageiros da cidade de Seattle, no estado de Washington, nos EUA. A coleta foi realizada entre 31/10/2001 e 02/12/2001, totalizando 125 MB de dados. O conjunto de dados representa o movimento real dos ônibus, sendo os eventos de mobilidade de um ônibus  $id_{bus}$  que percorre uma rota  $id_{route}$  registrados no formato  $\langle d, t, id_{bus}, id_{route}, x, y \rangle$ , onde  $x, y$  são as coordenadas cartesianas (em pés) da posição do ônibus na data  $d$  no horário  $t$ .

Neste trabalho, analisa-se uma quarta-feira típica, 31/10/2001, que inclui a mobilidade de 236 rotas de ônibus. As velocidades dos ônibus não são fornecidas no conjunto de dados original e, portanto, são previamente calculadas e anexadas a cada tupla de mobilidade. Essas tuplas podem ser interpretadas como contribuições dos usuários para o sistema de sensoriamento. Assim, no cenário estudado existem 376.491 amostras, com uma média de 1.595 amostras por rota. Este trabalho considera apenas a rota com maior quantidade de amostras, que é a rota  $id_{route} = 007$ , contendo 15.601 amostras.



**Figura 3. Velocidade média dos ônibus da rota  $id_{route} = 007$  obtida a partir da subdivisão da rota em trechos de 100 m.**

A velocidade máxima dos ônibus na região metropolitana de Seattle é de aproximadamente 56 km/h [Seattle.gov 2017]. O cálculo das velocidades dos ônibus com base nos deslocamentos registrados nos dados originais revelam alguns valores irrealistas, que não são contabilizados (Figura 2). Para tanto, são executadas duas avaliações. Primeiramente verifica-se se a velocidade de uma amostra,  $v_i$ , está no intervalo  $[0, v_{max}]$  km/h que define uma faixa de velocidades plausíveis para o cenário. Neste trabalho, considera-se  $v_{max} = 110$  km/h, que é a velocidade máxima permitida na cidade de Seattle [Seattle.gov 2017]. Além disso, verifica-se se o posicionamento do ônibus registrado pelo GPS está dentro das coordenadas da cidade de Seattle. A rota  $id_{route} = 007$  escolhida também passa pelo procedimento de validação, que resulta na exclusão de 55 amostras. As 55 amostras equivalem a 0,35% do total de amostras da rota, o que é considerado desprezível.

#### 4.1. Caracterização da rota $id_{route} = 007$

Considerando-se a rota  $id_{route} = 007$  inteira, a velocidade média no decorrer do dia é de 19,91 km/h. Para melhorar o refinamento dessa medida, permitindo uma avaliação mais detalhada da rota, primeiramente divide-se a rota em trechos de tamanho 100 m. Em seguida, para cada trecho, calcula-se a velocidade média dos ônibus durante todo o dia. A Figura 3 mostra a velocidade média obtida para cada trecho. Nessa figura, observa-se que a velocidade média em alguns pontos da rota  $id_{route} = 007$  no decorrer do dia pode alcançar até 34 km/h. Isso indica que há congestionamentos na região, uma vez que os ônibus não se movem com a velocidade máxima permitida para esse tipo de veículo na cidade de Seattle [Seattle.gov 2017] em nenhum trecho por completo por todo o tempo. Porém, nota-se na figura que há trechos onde a velocidade média é mais elevada e trechos onde a velocidade média é menos elevada.

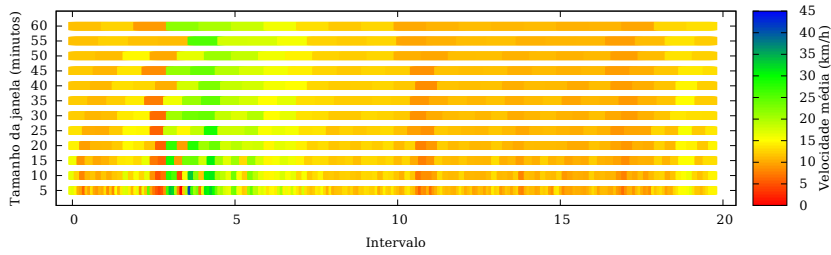
### 5. Análise do Sistema de Sensoriamento Proposto

O sistema proposto engloba tanto aspectos temporais quanto espaciais. O aspecto espacial é incluído através da divisão da RoI em trechos, enquanto o aspecto temporal é incluído a partir da divisão do tempo em intervalos  $\Delta t$ . Esta seção inicia a avaliação do sistema proposto através da definição de um valor para  $\Delta t$ . Em seguida, usando o intervalo de tempo selecionado, prossegue-se com a análise da proposta.

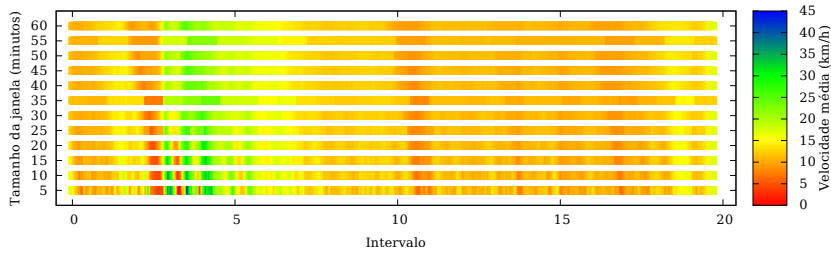
#### 5.1. Análise do tamanho da janela de tempo $\Delta t$

A janela de tempo  $\Delta t$  é usada para obter resultados como a velocidade média dos veículos ou o número de amostras sensoriadas por período. O tamanho ideal dessa janela





(a) Janela de tamanho fixo no tempo.



(b) Janela deslizante no tempo.

**Figura 4. Análise do tamanho da janela de tempo  $\Delta t$  comparando o uso de janelas de tamanho fixo e janelas deslizantes.**

é investigado utilizando-se duas abordagens: janela de tempo deslizante e janela de tempo fixa. Em ambos os casos, considera-se que a primeira janela se inicia no instante de tempo 00:00 e a última janela é finalizada no instante de tempo 20:00, que é a duração total do traço do dia 31/10/2001 da rota analisada Seção 4).

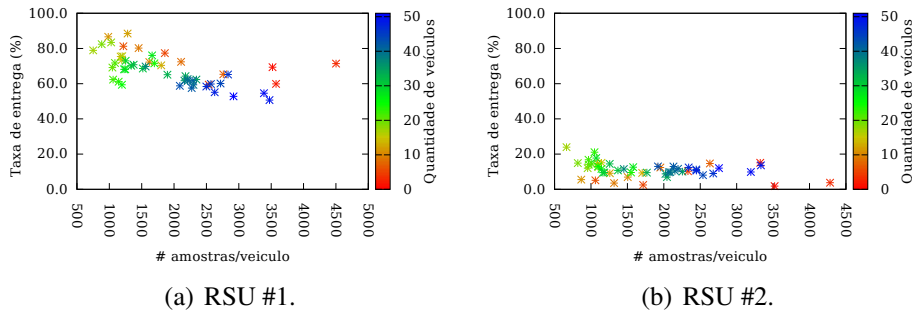
São analisados diversos valores para o tamanho da janela,  $\Delta t$ , tal que  $5 \leq \Delta t \leq 60$  min com passos iguais a 5 min. A Figura 4(a) mostra o resultado para janelas de tamanho fixo e a Figura 4(b), para janela deslizante. A cor de cada par  $(x,y)$  nessas figuras representa a velocidade média calculada na hora  $x$  usando um intervalo de  $y$  minutos. Em ambas as abordagens, observa-se que a utilização de janelas menores captura maiores variações nas medidas de velocidade média. As diferenças entre os resultados para cada tipo de janela são sutis, sendo mais perceptíveis ao se comparar o momento de transição entre janelas fixas consecutivas e o intervalo correspondente utilizando a janela deslizante. No entanto, nem a redução na granularidade da informação ao aumentar o tamanho da janela até 60 min, nem a perda de informação no momento de transição entre duas janelas consecutivas ao se utilizar janelas fixas, alteram significativamente a interpretação do estado momentâneo da rota. Por essa razão, no restante deste trabalho utiliza-se uma janela fixa de 60 min para obtenção e análise dos resultados.

## 5.2. Desempenho do sistema de sensoriamento proposto

A metodologia proposta é avaliada através da análise da taxa de entrega e do erro da amostragem, considerando que a RoI é a rota  $id_{route} = 007$  e a medida a ser sensoriada é a velocidade média dos ônibus nos trechos. Considera-se ainda que a rota  $id_{route} = 007$  está dividida em trechos de 100 m. A mobilidade e a comunicação no cenário são simulados utilizando o simulador de redes NS-3.26.

**Avaliação da taxa de entrega:** Esta análise verifica o compromisso entre o número de nós

e a quantidade de dados transferidos por nó, de forma que a RoI seja coberta e as RSUs sejam capazes de receber por completo todos os dados descarregados. Dessa forma, é possível fazer uma análise de capacidade da rede. Para isso, foram realizadas simulações variando o número de veículos e o número de amostras por veículo a serem entregues às RSUs #1 e #2. Ao final, calcula-se a taxa de entrega.



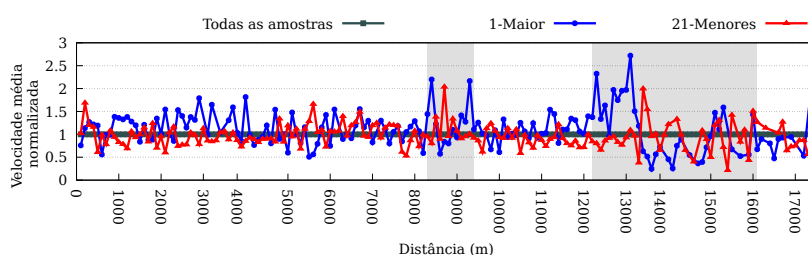
**Figura 5. Taxa de entrega nas duas RSUs nas extremidades da rota  $id_{route} = 007$  variando a quantidade de veículos que participam do sistema de sensoriamento e a carga de dados por veículo.**

A Figura 5 apresenta a relação entre a taxa de entrega, a carga média transmitida dos ônibus para as RSUs e o número de veículos que competem pelo meio durante uma transferência de dados. As RSUs estão localizadas nas extremidades da rota  $id_{route} = 007$ , o que significa que há uma RSU (RSU #1) na extremidade direita e outra RSU (RSU #2) na extremidade esquerda da Figura 3. A taxa de entrega na RSU #1 é alta para uma carga média por veículo mais baixa, por exemplo, 1.000 amostras por veículo, e diminui à medida que a carga média aumenta, chegando a  $\approx 60\%$  para uma carga de  $\approx 2.500$  amostras por veículo. Isso pode ocorrer devido à curta duração do contato com a RSU, sendo insuficiente para descarregar os dados. A quantidade de veículos competindo pelo meio também reduz a taxa de entrega. Por exemplo, para uma mesma carga, 1.500 amostras por veículo, a taxa de entrega é de  $\approx 65\%$  com  $\approx 30$  veículos, e  $\approx 80\%$  para  $\approx 10$  veículos. Essa redução pode ocorrer devido ao aumento do número de colisões quando mais de um veículo tenta descarregar os dados ao mesmo tempo na mesma RSU. Note que para cargas elevadas  $> 3.500$  amostras por veículo, apenas as configurações com um número menor de veículos competindo pelo meio conseguem entregar dados.

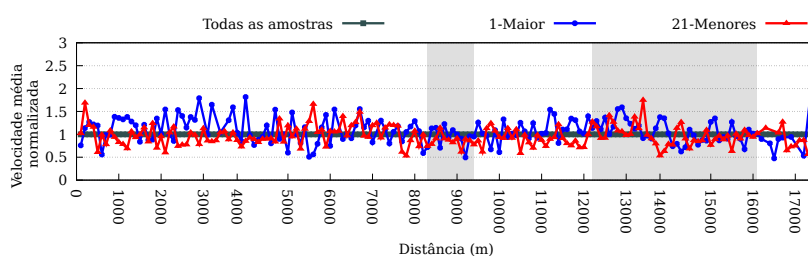
Na RSU #2 ambas as variáveis, quantidade de veículo e carga por veículo, influenciam a taxa de entrega de forma semelhante. Porém, mesmo com cargas mais baixas e menor quantidade de veículos, a taxa de entrega é pequena. A diferença entre os comportamentos das duas RSUs se deve ao menor tempo de contato entre os veículos e a RSU #2, que está localizada em uma região de maior velocidade da rota (Figura 3). Observa-se novamente que há alguma entrega de dados para cargas mais elevadas somente quando o número de nós competindo pelo meio é baixo. Isso significa que cargas mais elevadas são apenas possíveis de serem entregues em cenários com poucos carros competindo pelo meio. Para cargas baixas, porém, a taxa de entrega pode ser mais alta mesmo quando há mais carros competindo pelo meio. Essa última observação é interessante pois permite a entrega de maior carga de dados agregados, caso sejam utilizados vários veículos ao mesmo tempo com cargas baixas individuais. Sobrecarregar um único veículo pode não ser atraente sob o ponto de vista do usuário, já que pode levar mais rapidamente à

exaustão de recursos próprios.

**Avaliação do erro da amostragem:** Em um cenário ideal, o sistema conhece exatamente a quantidade de dados necessários para reproduzir com baixo erro a informação sensoriada. Isso evitaria o problema de subamostragem, que pode levar a erros na medida; e de superamostragem, que pode representar desperdício de recursos. O sistema proposto neste trabalho não consegue determinar com certeza esses valores, mas é capaz de estimar quais regiões necessitam de mais ou menos coletas, de forma adaptativa, com base na análise da variabilidade dos dados nos diferentes trechos.



(a) Sem o uso do sistema proposto, todos os trechos realizam amostragem média constante igual a  $R$ .



(b) Com o uso do sistema proposto, trechos com menor variação com taxa de amostragem igual a  $R$  e trechos com maior variação com taxa de amostragem igual a  $2R$ .

**Figura 6. Velocidades médias normalizadas usando um único veículo ('1-Maior') e 21 veículos com a mesma carga ('21-Menores') nos trechos de 100 m. A região cinza representa os trechos com maiores diferenças de velocidades médias.**

A Figura 6 mostra a variabilidade das medidas de velocidade média sem e com o uso do sistema de sensoriamento proposto. São apresentados os resultados para o sistema usando apenas o ônibus que coleta o maior número de amostras do conjunto de dados, chamado '1-Maior', e um conjunto de 21 ônibus que somados coletam a mesma quantidade de amostras do '1-Maior', chamado '21-Menores'. As médias estão normalizadas, sendo todas elas divididas pela velocidade média obtida com todas as amostras do conjunto de dados. A adaptação da taxa de amostragem ocorre quando o sistema determina que mais coletas devem ser realizadas em uma região onde existe elevada variabilidade. Essas regiões estão preenchidas em cinza na Figura 6, onde a variação em torno da linha horizontal de velocidade média normalizada igual a 1 é maior. A Figura 6(a) mostra os resultados considerando uma taxa de amostragem média ( $R$ ) em todos os trechos de 7,5, ou seja, foram coletadas 1.304 amostras ao longo dos 173 trechos da rota  $id_{route} = 007$ . Note que há uma maior variação das médias nos trechos preenchidos com cinza. Dobrando a taxa de amostragem para  $2R = 15$ , apenas nos trechos de maior variação (preenchidos

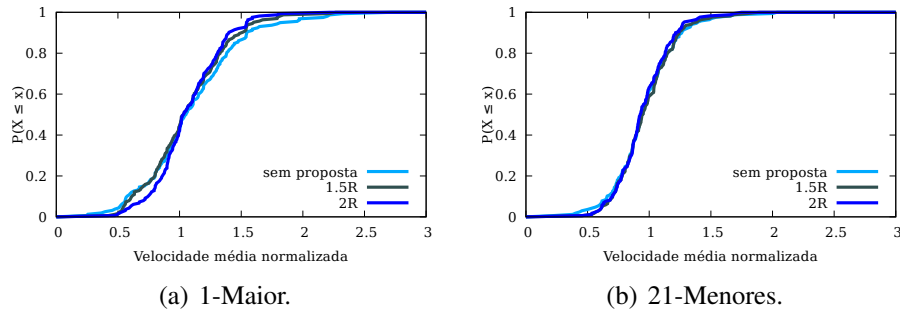


Figura 7. CDF das diferenças de toda a rota  $id_{route} = 007$ .

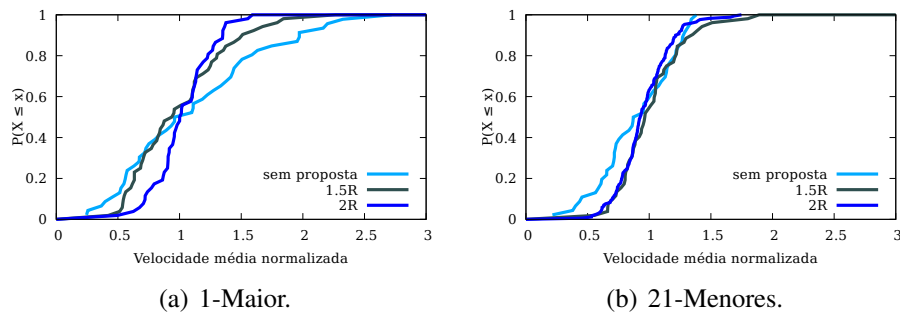


Figura 8. CDF das diferenças das velocidades médias nos trechos de maior variação: curva sem proposta, com taxas de  $1,5R$  e  $2R$ .

com cinza), já há uma melhora, como visto na Figura 6(b).

A Figura 7 mostra a CDF das velocidade médias normalizadas usando também uma taxa de amostragem intermediária,  $1,5R$ , na região preenchida com cinza para enriquecer os resultados. Note que há uma redução das diferenças tanto para o caso de um único ônibus, '1-Maior', quanto para o caso dos 21 ônibus, '21-Menores'. Tal redução ainda é tímida porque o aumento da taxa foi realizado apenas em alguns trechos. Caso se isole a CDF apenas dos trechos de maior variação, como visto na Figura 8, é possível perceber com mais clareza a redução das maiores diferenças de velocidade. As velocidades medidas se tornam mais próximas da existente no conjunto de dados conforme a curva da CDF se aproxima de um degrau em  $x = 1$ . Valores medidos menores que 1 significam velocidades médias menores que a do conjunto de dados, enquanto valores medidos maiores que 1 representam valores de velocidades médias maiores que a do conjunto de dados. Ao utilizar 21 ônibus, o impacto é menor porque as amostras utilizadas são aleatórias dentro da região de maior variação e, no momento da amostragem, pode haver trechos que não foram contemplados. A Tabela 1 reforça os resultados mostrando a redução na RMSE (*Root Mean Square Error*) das velocidades médias normalizadas. Percebe-se que há uma tendência de redução do erro para toda a rota com o aumento das taxas de amostragem nos trechos de maior variação.

É importante ressaltar que a taxa de amostragem em toda a rota poderia ser aumentada, já que o sistema ainda está distante do número máximo de amostras contidas no conjunto de dados e da capacidade máxima da rede, como visto na Figura 5. A rota  $id_{route} = 007$  possui 15.546 amostras no total, sendo que nos experimentos com mais

**Tabela 1. RMSE das velocidades médias considerando toda a rota  $id_{route} = 007$ .**

Configuração	Sem Proposta	Proposta com $1.5R$	Proposta com $2R$
1-Maior	0,422	0,342	0,29
21-Menores	0,441	0,387	0,366

amostras (trechos com taxa  $2R$ ) foram coletas apenas 1.696 amostras. Esse valor não representa problema pois está abaixo do máximo contido no conjunto de dados e ainda possui probabilidade de ser entregue com sucesso à RSU durante o contato.

## 6. Conclusões e Trabalhos Futuros

Este trabalho propôs um sistema de sensoriamento participativo que considera tanto características espaciais quanto temporais de regiões de interesse (RoI). Para isso, divide-se a RoI em trechos e assume-se que as medidas coletadas podem variar dentro de um intervalo de tempo  $\Delta t$ . Os participantes coletam dados obedecendo a taxas de amostragem definidas por um servidor central e a boa acurácia do sistema depende dessa taxa. Quanto maior for a taxa, melhores serão os resultados. No entanto, isso esbarra no compromisso entre o número de amostras coletadas, limitações dos dispositivos e carga de dados gerada. Considerando o cenário abordado, monitoramento das velocidades dos veículos levando em conta amostras coletadas por veículos, esse compromisso é ainda mais severo devido ao meio sem-fio. O sistema proposto ataca esse compromisso ajustando a taxa de amostragem de cada trecho da RoI em função da variação das medidas de velocidade dos veículos. Os trechos com maior variação devem ter maior taxa de amostragem, enquanto os de baixa variação podem ter taxas menores. Os resultados obtidos com conjunto de dado real mostram que maiores cargas de dados só são possíveis de serem entregues com poucos nós competindo pelo meio. Os resultados ainda mostram que a amostragem adaptativa reduz os erros sem impactar na carga de dados. Como trabalhos futuros, pretende-se expandir os resultados para mais rotas e tornar a identificação da melhor taxa de amostragem por trecho autônoma.

## Referências

- André, C. H. O. M., Medeiros, D. S. V., and Campista, M. E. M. (2017). A methodology to assess data consistency in vehicular networks using participatory sensing. Technical Report GTA-17-31, GTA/PEE/UFRJ.
- Boulos, M. N. K., Wheeler, S., Tavares, C., and Jones, R. (2011). How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. *Biomedical engineering online*, 10(1):24:1–24:14.
- Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M. B. (2006). Participatory sensing. In *World Sensor Web Workshop, ACM Sensys*, pages 1–5.
- Cruz, P., da Silva, F. F., Pacheco, R. G., Couto, R. S., Velloso, P. B., Campista, M. E. M., and Costa, L. H. M. (2017). SensingBus: um sistema de sensoriamento baseado em Ônibus urbanos. In *Salão de Ferramentas do XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1–14.

- D'Hondt, E., Stevens, M., and Jacobs, A. (2013). Participatory noise mapping works! an evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring. *Pervasive and Mobile Computing*, 9(5):681–694.
- Estrin, D., Chandy, K. M., Young, R. M., Smarr, L., Odlyzko, A., Clark, D., Reding, V., Ishida, T., Sharma, S., Cerf, V. G., et al. (2010). Participatory sensing: applications and architecture. *IEEE Internet Computing*, 14(1):12–42.
- Jetcheva, J. G., Hu, Y.-C., PalChaudhuri, S., Saha, A. K., and Johnson, D. B. (2003). CRAWDAD dataset rice/ad\_hoc\_city (v. 2003-09-11). [http://crawdad.org/rice/ad\\_hoc\\_city/20030911/bus\\_mobility](http://crawdad.org/rice/ad_hoc_city/20030911/bus_mobility). traceset: bus\_mobility.
- Liu, Y., Niu, J., Ma, J., Shu, L., Hara, T., and Wang, W. (2013). The insights of message delivery delay in vanets with a bidirectional traffic model. *Journal of Network and Computer Applications*, 36(5):1287–1294.
- McCall, M. K. and Minang, P. A. (2005). Assessing participatory GIS for community-based natural resource management: claiming community forests in Cameroon. *Geographical Journal*, 171(4):340–356.
- Melo, P. C. F. et al. (2014). CSVM: Uma plataforma para crowdsensing móvel dirigida por modelos em tempo de execução. Master's thesis, Universidade Federal de Goiás.
- Miche, M. and Bohnert, T. M. (2009). The internet of vehicles or the second generation of telematic services. *ERCIM News*, 77:43–45.
- Mohan, P., Padmanabhan, V. N., and Ramjee, R. (2008). Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *ACM conference on Embedded network sensor systems*, pages 323–336.
- Ribeiro Neto, V., Medeiros, D. S. V., and Campista, M. E. M. (2016). Analysis of mobile user behavior in vehicular social networks. In *International Conference Network of the Future (NOF 2016)*, pages 1–5.
- Seattle.gov (2017). Seattle department of transportation. <http://www.seattle.gov/transportation/sdotfaqs.htm>. Acessado: 20/07/2017.
- Weinschrott, H., Weisser, J., Durr, F., and Rothermel, K. (2011). Participatory sensing algorithms for mobile object discovery in urban areas. In *IEEE International Conference on Pervasive Computing and Communications (PERCOM)*, pages 128–135.
- Xiao, Y., Simoens, P., Pillai, P., Ha, K., and Satyanarayanan, M. (2013). Lowering the barriers to large-scale mobile crowdsensing. In *14th Workshop on Mobile Computing Systems and Applications*, pages 9:1–9:6.
- Yu, R., Zhang, Y., Wu, H., Chatzimisios, P., and Xie, S. (2013). Virtual machine live migration for pervasive services in cloud-assisted vehicular networks. In *8th International ICST Conference on Communications and Networking in China (CHINACOM)*, pages 540–545.
- Zeng, Y. and Xiang, K. (2017). Adaptive sampling for urban air quality through participatory sensing. *Sensors (Basel)*, 17(11):1–16.
- Zhou, P., Zheng, Y., and Li, M. (2012). How long to wait?: predicting bus arrival time with mobile phone based participatory sensing. In *10th international conference on Mobile systems, applications, and services*, pages 379–392.