Estratégias de pré-processamento para posicionamento de pontos de coleta em redes de sensores móveis

Pedro Cruz
Universidade Federal do Rio de Janeiro – COPPE/PEE/GTA
Cidade Universitária - Rio de Janeiro/RJ
cruz@gta.ufrj.br

Rodrigo S. Couto
Universidade do Estado do Rio de Janeiro – PEL/DETEL/FEN
Campus Maracanã - Rio de Janeiro/RJ
rodrigo.couto@uerj.br

Abílio Lucena
Universidade Federal do Rio de Janeiro – COPPE-PESC
Cidade Universitária - Rio de Janeiro/RJ
abiliolucena@cos.ufrj.br

Luís Henrique M. K. Costa
Universidade Federal do Rio de Janeiro – COPPE/PEE/GTA
Cidade Universitária - Rio de Janeiro/RJ
luish@gta.ufrj.br

RESUMO

Cidades inteligentes podem utilizar sensores sem fio embarcados em ônibus para obter dados e emitir alertas aos cidadãos. Dados são entregues oportunisticamente a pontos de coleta acoplados a pontos de ônibus, que utilizam uma infraestrutura em nuvem pra fornecer informações aos usuários finais. Com a comunicação oportunística, surge o problema de atrasos da entrega de dados. Esses atrasos dependem do intervalo entre contatos dos ônibus com pontos de ônibus escolhidos como pontos de coleta. Este trabalho propõe um método para a escolha ótima de pontos de ônibus a receber equipamentos de coleta quando há restrição no número desses equipamentos. Modela-se o problema como P-Centros e sua entrada é pré-processada com base em uma heurística. Aplica-se o problema aos ônibus e pontos de ônibus da Zona Sul do Rio de Janeiro. Os resultados para a instância estudada mostram que o pré-processamento possibilita a obtenção da solução ótima, impraticável sem o mesmo.

PALAVRAS CHAVE. Cidades Inteligentes, Redes de Sensores Móveis, P-Centros

TEL&SI - PO em Telecomunicações e Sistemas de Informações; OC - Otimização Combinatória

ABSTRACT

Smart Cities can attach wireless sensors onto urban buses in order to collect data and trigger alerts to its citizens. Data is delivered in opportunistic contacts between buses and gateways coupled onto bus stops, that use a cloud infrastructure to deliver information to final users. Opportunist communication brings the problem of delays on data delivery. Delays are related to the interval between contacts of buses and bus stops that are chosen as gateways. This work proposes

https://proceedings.science/p/85320?lang=pt-br

a method to choose optimally bus stops to receive gateways, when the number of such equipment is constrained. The problem is modelled as a P-Center and the input data is pre-processed using a heuristic. The problem is applied to the buses and bus stops of the South Zone of Rio de Janeiro. Results for the chosen instance show that the pre-processing allows an optimal solution to be found, not possible otherwise.

KEYWORDS. Smart Cities, Mobile Wireless Sensors Networks, P-Center

TEL&SI - OR in Telecommunications and Information Systems; CO - Combinatorial Optimization

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

https://proceedings.science/p/85320?lang=pt-br

1. Introdução

As cidades inteligentes possuem diversas aplicações que necessitam de sensoriamento. A cobertura completa da cidade por sensores estáticos é geralmente custosa, pois pode necessitar de um número proibitivo de sensores. Além disso, esses sensores devem ser integrados a uma rede que cubra toda a cidade. Para contornar esses problemas, podem ser adotadas soluções que utilizem sensores móveis. Esse tipo de sensor toma proveito da mobilidade dos nós para reduzir a quantidade de sensores utilizados para cobrir a área da cidade. Assim, com esses sensores é possível aumentar a cobertura espacial, em detrimento da cobertura temporal, ou seja, as informações levam mais tempo pra chegar aos usuários Liu et al. [2005]. Uma outra vantagem é a possibilidade de utilizar comunicação oportunística entre sensores e pontos de coleta, dispensando a necessidade de uma rede implantada por toda a cidade, como demonstrado por Ekici et al. [2006] e Zhao et al. [2016]. Uma opção para dotar sensores com mobilidade é embarcar os sensores em ônibus urbanos, e instalar pontos de coleta em pontos de ônibus. Os dados são coletados pelos ônibus e entregues por redes sem fio aos pontos de ônibus. Esse pontos de ônibus são conectados à Internet, por rede cabeada ou rede sem fio, enviando as informações recebidas a um servidor de aplicações ou diretamente às aplicações.

O sensoriamento em cidades inteligentes pode ser utilizado por diversas aplicações, incluindo aplicações sensíveis ao atraso Zanella et al. [2014]. A utilização de comunicação oportunística significa que, uma vez que um fenômeno é sensoreado, ainda há um atraso na comunicação desse fenômeno ao usuário. Esse atraso é dado pelo intervalo de tempo entre o instante em que um ônibus detecta o fenômeno até o envio da informação para a aplicação. Esse tempo é dominado pelo tempo que o ônibus leva desde a detecção até realizar contato com um ponto de coleta. Se não for possível equipar todos os pontos de ônibus com pontos de coleta, é esperado um aumento no atraso na entrega da informação em alguns trechos. O maior de todos os atrasos determinará o maior atraso que um alerta qualquer poderá ser emitido pela rede. Assim, se o número de pontos de coleta for limitado por algum orçamento, é necessário escolher quais pontos de ônibus receberão pontos de coleta.

Este trabalho propõe um método para a seleção de pontos de ônibus a serem equipados com pontos de coleta com o objetivo de minimizar o maior atraso possível na rede. O método modela o problema como um problema de p-centros e utiliza a heurística apresentada por Cruz et al. [2017a] para pré-processar os dados, possibilitando a obtenção de uma solução ótima para determinadas instâncias do problema. O método é aplicado a dados reais de trajetos de ônibus e posições de pontos de ônibus na Zona Sul da cidade do Rio de Janeiro.

Este artigo está organizado da seguinte forma. A Seção 2 aborda os trabalhos relacionados. A Seção 3 discute o comportamento dos atrasos da rede quando há variação no número de pontos de coleta pelos pontos de ônibus. A Seção 4 apresenta uma formulação do problema como um problema de P-Centros, enquanto a Seção 5 apresenta um algoritmo de pré-processamento para redução do espaço de busca do problema. A Seção 6 apresenta a aplicação do método em um cenário real. Por fim, a Seção 7 conclui o trabalho e apresenta trabalhos futuros.

2. Trabalhos Relacionados

Este trabalho se relaciona com trabalhos que utilizam a modelagem de P-Centros, com trabalhos de posicionamento de pontos de coleta para redes de sensores e com trabalhos de sensoriamento urbano com veículos. Tradicionalmente, o problema de P-Centros é utilizado para localizar instalações de atendimento de emergência, como em Pereira et al. [2004]. Nesse trabalho, Pereira *et al.* utiliza o modelo de P-Centros para distribuir cabines de segurança por um bairro na cidade do Rio de Janeiro. A ordem de grandeza das instâncias analisadas pelo trabalho permite a enumeração para obtenção de solução ótima.

Davidović et al. [2011] utilizam a estratégia da colônia de abelhas para encontrar soluções aproximadas para o problema de P-Centros, cujo resultado é dependente de parâmetros escolhidos pelo usuário. A abordagem do presente trabalho é independente de parâmetros e produz soluções

https://proceedings.science/p/85320?lang=pt-br

ótimas, com a condição de que a distribuição dos dados de entrada possibilitem o pré-processamento eficiente dos mesmos.

Em termos de posicionamento de pontos de coleta para redes de sensores, os trabalhos de Wong et al. [2004] e Youssef e Younis [2007] consideram o atraso de rede como o número de saltos na rede. Assim, Wong et al. [2004] buscam a melhor configuração de pontos de coleta para minimizar o atraso da rede elaborando um problema de otimização combinatória. Já Youssef e Younis [2007], propõe uma solução baseada no algoritmo genético para obter uma solução de posicionamento de pontos de coleta para redes maiores. Ambos os trabalhos, porém, têm como escopo redes de sensores estáticos.

Em termos de sensoriamento urbano utilizando veículos, se destacam os projetos Opensense, de Marjovi et al. [2015], e Mosaic, de Dong et al. [2015]. Nesses projetos, veículos urbanos recebem sensores e coletam dados ambientais. Entretanto, ao contrário do presente trabalho, a comunicação não é oportunística, sendo desnecessária a preocupação com o atraso de comunicação. O trabalho de *et al.* Bonola et al. [2016] também utiliza comunicação oportunística por meio da mobilidade da frota de táxis de uma cidade, mas não realiza uma análise do atraso na comunicação.

No trabalho de Cruz et al. [2017c], ônibus são equipados com nós de sensoriamento e entregam os dados coletados a pontos de coleta nos pontos de ônibus. Nesse tipo de cenário, os trabalhos de Cruz et al. [2017a] e Cruz et al. [2017b] elaboram um algoritmo baseado em uma heurística para escolher pontos de ônibus a serem equipados com pontos de coleta. Essa heurística, entretanto, não possui garantia de otimalidade. O presente trabalho modela o problema de escolha de pontos de coleta como um problema do tipo P-Centros e utiliza a abordagem de Cruz et al. [2017a] para pré-processar os dados de entrada do problema e reduzir o espaço de candidatos a soluções. Assim, é possível obter resultados ótimos para instâncias do problema da ordem de bairros ou zonas de uma cidade como o Rio de Janeiro.

3. Modelo de Rede de Sensores com Pontos de Coleta Limitados

No esquema de sensoriamento estudado, nós de sensoriamento são embarcados em ônibus urbanos e pontos de coleta são acoplados aos pontos de ônibus. Ao trafegar pela cidade junto aos ônibus, os nós de sensoriamento coletam e acumulam dados ambientais a respeito das regiões que fazem parte dos trajetos dos ônibus. Tais dados podem ser sobre temperatura, poluição, quantidade de chuvas ou outros dados de interesse. Ao realizar contato com um ponto de coleta, o nó de sensoriamento entrega a esse ponto os dados acumulados durante o trajeto. O ponto de coleta pode, então, encaminhar os dados pela Internet para o gerenciador de operação. Esse último pode, por exemplo, ser um servidor hospedado em um serviço de nuvem. O esquema é ilustrado na Figura 1, adaptada do trabalho de Cruz et al. [2017a]. As notações utilizadas por este trabalho estão listadas na Tabela 1.

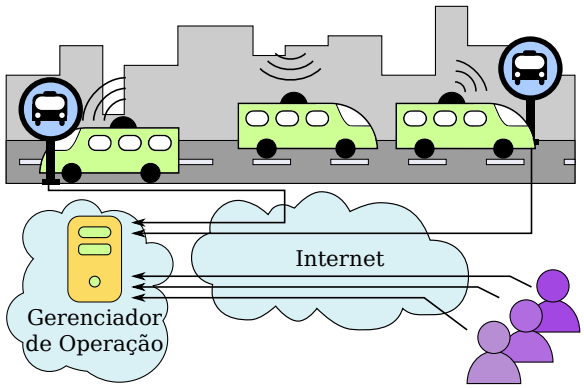


Figura 1: Esquema de coleta de dados com sensores embarcados em ônibus urbanos.

A partir do instante em que os dados são coletados até que estejam finalmente disponíveis

https://proceedings.science/p/85320?lang=pt-br

para os usuários e aplicações, os mesmos sofrem um atraso, denotado por D. Esse atraso é dado pelo intervalo entre a aquisição dos dados e o contato com um ponto de coleta, mais o tempo necessário para envio dos dados entre ônibus e ponto de coleta e, depois, entre o ponto de coleta e o gerenciador de operação. Como o atraso de envio dos dados é tipicamente várias ordens de magnitude menor do que o intervalo entre contatos de um ônibus com dois ponto de ônibus, considera-se que o atraso de envio de dados seja desprezível e que D seja apenas o tempo desde que um dado é sensoreado por um ônibus até o momento em que é entregue a um ponto de coleta. Uma vez que os ônibus podem coletar dados constantemente, o maior atraso D_{max} sofrido por um dado coletado por essa rede é o maior tempo que um ônibus leva entre dois pontos de coleta quaisquer em seu trajeto.

Este trabalho assume que, uma vez que um nó de sensoriamento embarcado em um ônibus realiza contato com um ponto de coleta, o nó de sensoriamento consegue descarregar todos os dados acumulados anteriormente. Tal afirmativa se mostra razoável a partir dos resultados experimentos obtidos por Rubinstein et al. [2009], que foi capaz de transferir, 1,5 MB de dados entre dois carros que se cruzam com velocidade relativa de 120 km/h de uma única vez. Utilizando como referência a massa de dados geradas de 53 B/s no protótipo de Cruz et al. [2017c], pode-se afirmar que um ônibus deve ficar mais do que 7 h sem sequer passar por nenhum ponto de coleta para que o nó de sensoriamento embarcado não consiga descarregar todos os dados armazenados. Os resultados deste trabalho mostram que os atrasos máximos da rede são pelo menos uma ordem de grandeza inferiores a esse valor.

3.1. Relação entre atrasos e as configurações de pontos de coleta

Denota-se por $\mathcal{B}$ o conjunto de todos os ônibus e $\mathcal{P}$ o conjunto de todos os pontos de ônibus elegíveis a serem equipados com pontos de coleta. Assim, pode-se representar o trajeto de $b \in \mathcal{B}$ como uma sequência $(p_{b1}, p_{b2}, ..., p_{bn})$ dos pontos de ônibus com os quais o ônibus pode realizar contato, com $p_{b1}, p_{b2}, ..., p_{bn} \in \mathcal{P}$. O ponto p_{bi} é o i-ésimo ponto com o qual o ônibus b faz contato. Consequentemente, é possível determinar uma sequência $(d_{b1}, d_{b2}, ..., d_{b(n-1)})$, na qual cada dbi é o intervalo de tempo entre o ônibus perder contato com o i-ésimo ponto de seu trajeto até estabelecer contato com o (i+1)-ésimo ponto do trajeto.

Caso um ponto de ônibus p_{bi} que faça parte do trajeto de b não seja equipado com um ponto de coleta, há um aumento no atraso na entrega dos dados, pois os dados coletados a partir de p_{bi-1} somente podem ser entregues ao ponto de coleta em p_{bi+1}. Tal efeito é ilustrado na Figura 2, na qual p, q, r e s são pontos de ônibus candidatos a receberem pontos de coleta que estão no trajeto do ônibus b. No caso (i), todos os pontos recebem pontos de coleta e os dados sofrem atrasos de d_{bpq}, d_{bqr} e d_{brs}. No caso (ii), o candidato r não é escolhido como ponto de coleta e, assim, não pode receber dados. Dessa forma, os dados coletados no instante em que o contato com q é perdido sofrem atraso de d_{bqs}, que é a soma dos atrasos d_{bqr} e d_{brs}.

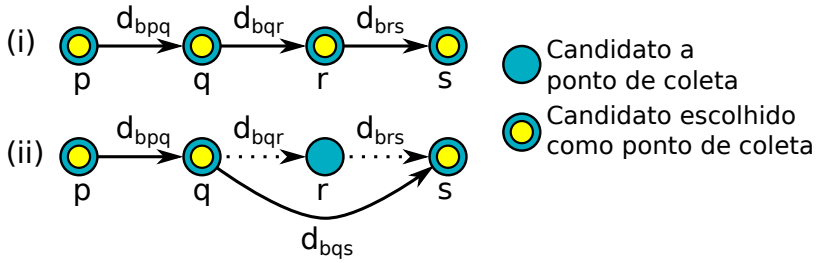


Figura 2: Efeito nos atrasos de entrega quando um ponto de ônibus não é um ponto de coleta.

Em um cenário no qual todos os pontos de ônibus sejam equipados com pontos de coleta, obtém-se o menor valor possível para D_{max}. Entretanto, caso haja uma limitação no número de equipamentos a serem implantados, pode haver um aumento em D_{max}. Existe uma gama de aplicações que, para utilizar a rede, precisam de determinadas garantias com relação ao atraso

https://proceedings.science/p/85320?lang=pt-br

máximo que os dados utilizados podem sofrer. Um exemplo é o caso de aplicações de alerta de chuvas, que devem alertar condições de chuvas forte em tempo hábil para as autoridades tomarem decisões. Por isso, é importante que a escolha dos pontos de ônibus a serem equipados com pontos de coleta seja capaz de minimizar D_{max}.

Uma observação a ser feita a respeito da escolha dos pontos de ônibus a receberem equipamentos de pontos de coleta é que o primeiro e o último ponto do trajeto de algum ônibus devem sempre receber tais equipamentos. Caso o ponto inicial não receba, não é possível garantir que o ônibus iniciou seu trajeto sem dados já atrasados e, caso o ponto final não seja ponto de coleta, o ônibus terminará seu trajeto sem entregar alguns dados. Neste trabalho, o conjunto de pontos inicial ou final no trajeto de algum ônibus é denotado por $\mathcal{I}$. Este trabalho assume que os trajetos dos ônibus não são cíclicos.

Tabela 1: Notações utilizadas no trabalho.

Notação	Descrição	Tipo
$\mathcal{B}$	Ônibus que circulam pela cidade	Conjunto
$\mathcal{P}$	Pontos de ônibus candidatos a serem pontos de coleta	Conjunto
$\mathcal{C}$	Pares (b, p) para cada ônibus b que faz contato com o ponto p	Conjunto
$\mathcal{I}$	Pontos de ônibus que são os primeiros ou os últimos pontos de qualquer um dos ônibus	Conjunto
$N_{desejado}$	Quantidade desejada de pontos de coleta	Parâmetro
d_{bpq}	Tempo total entre o ônibus b fazer contato com o ponto p e o ponto q, com $p, q \in \mathcal{P}$	Parâmetro
n_b	Quantidade de pontos de ônibus no trajeto do ônibus b	Parâmetro
D_{max}	Maior atraso possível na rede entre dois pontos de coleta	Variável
x_p	Valor binário que indica se o ponto p é escolhido como ponto de coleta	Variável
y_{bpq}	Valor binário indicando, para um ônibus b, se q é o próximo ponto de coleta a partir do ponto p	Variável

4. Formulação P-Centros do Problema

O problema dos P-Centros é um problema de localização NP-Difícil, como demonstrado por Kariv e Hakimi [1979]. Nele, um conjunto $\mathcal{P}$ de candidatos a instalações dispostos em um mapa devem atender a um conjunto $\mathcal{C}$ de demandas, também dispostas nesse mapa. Entre cada demanda e candidato a instalação existe uma distância. Como exemplo, as instalações atendem a demandas de natureza emergencial, como incêndios ou resgates. Um candidato a instalação pode ser escolhido ou não para nele seja construída uma instalação. Considera-se que uma demanda é atendida pelo candidato mais próximo que for escolhido para receber uma instalação. Adicionalmente, existe um limite $N_{desejado}$ de candidatos que podem receber instalações. Assim, é importante que os candidatos a instalação sejam escolhidos de forma a garantir que toda demanda seja atendida e que seja minimizada a maior distância entre uma demanda qualquer e a instalação que a atende. Uma instância do problema P-Centros é ilustrada na Figura 3.

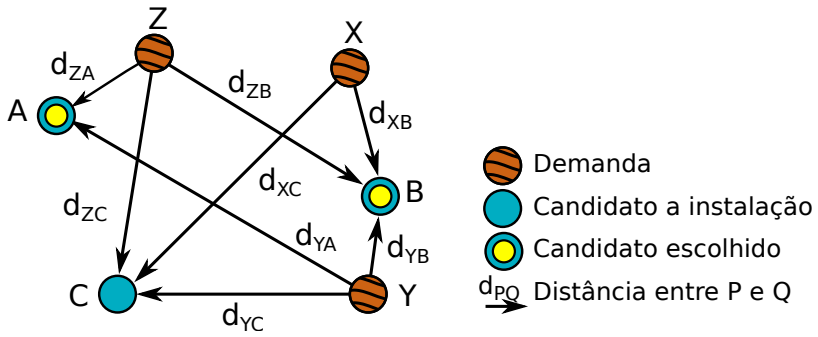


Figura 3: Uma instância do problema de p-centros com dois candidatos escolhidos.

É possível modelar o problema de escolha dos pontos de coleta como um problema de P-Centros. Para tal, cada encontro de um ônibus b com um candidato p a ponto de coleta gera uma

demanda $(b, p) \in \mathcal{C}$ e cada candidato a ponto de coleta é um candidato a instalação. As distâncias entre uma demanda e uma instalação é o tempo d_{bpq} que o ônibus b demora para ir do ponto p até o ponto q. A Figura 4 ilustra a transformação de uma instância do problema estudado em uma instância de problema P-Centros. O modelo em Programação Mista Inteira Linear (*Mixed Integer Linear Programming* - MILP) para o problema é mostrado a seguir.

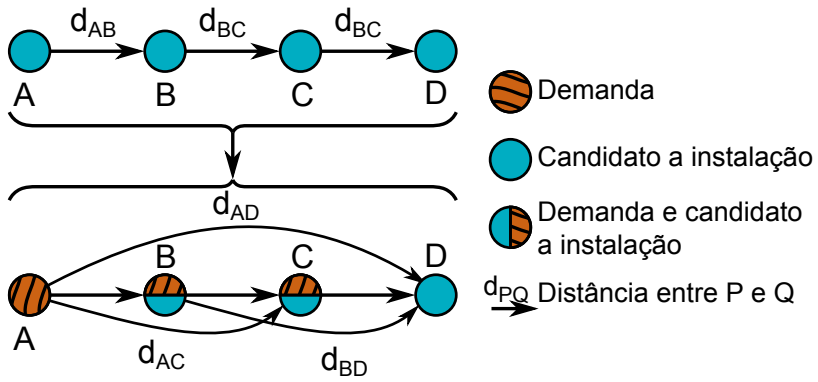


Figura 4: Transformação dos atrasos dos trajetos em uma instância do problema p-centros.

$$\text{minimizar:} \quad D_{max}; \tag{1}$$

$$\text{sujeito a:} \quad \sum_{q \in \mathcal{P}} y_{bpq} = 1\,,\ \forall (b,p) \in \mathcal{C}; \tag{2}$$

$$y_{bpq} - x_q \leq 0, \forall (b,p) \in \mathcal{C}, q \in \mathcal{P} \tag{3}$$

$$\sum_{p \in \mathcal{P}} x_p = N_{desejado}; \tag{4}$$

$$\sum_{q \in \mathcal{P}} d_{bpq} y_{bpq} - D_{max} \leq 0\,,\ \forall (b,p) \in \mathcal{C},\ \forall p \in \mathcal{P}; \tag{5}$$

$$x_p = 1,\ \forall p \in \mathcal{I}; \tag{6}$$

$$D_{max} \in \mathbb{Z};\ y_{bpq} \in \{0,1\},\ \forall b \in \mathcal{B},\ \forall p,q \in \mathcal{P};\ x_p \in \{0,1\}\ \forall p \in \mathcal{P}. \tag{7}$$

No problema exposto, a Equação 1 é a função objetivo do problema e minimiza D_{max}, que é o maior atraso da rede. As variáveis binárias x_p indicam se o ponto de ônibus p é escolhido para ser ponto de coleta, enquanto as variáveis y_{bpq} são variáveis binárias que indicam se, uma vez que o ônibus b saia do alcance do ponto de coleta p, o ponto q é o próximo ponto a ser encontrado. O parâmetro d_{bpq} é o tempo que o ônibus b demora para ir do ponto p ao ponto q. A Equação 2 é um conjunto de restrições para garantir que, uma vez que b saia do ponto p, haja apenas um ponto q no qual os dados sejam descarregados. A Equação 3 diz que se um ponto q não é escolhido como ponto de coleta, então um ônibus b não pode descarregar dados nele. A Equação 4 limita o número de pontos escolhidos ao orçamento definido pelo problema. A Equação 5 define D_{max} como o maior entre os atrasos sofridos pelos dados. A Equação 6 força que os pontos em I sejam sempre escolhidos. As Equações 7 definem os limites e domínios para as variáveis do problema.

5. Pré-processamento Baseado em Limitante Superior por Heurística Gulosa

Na formulação proposta, o número de variáveis x_p é igual ao número de pontos de ônibus em $\mathcal{P}$. Entretanto, o número de variáveis y_{bpq} e parâmetros d_{bpq} obedece a seguinte expressão:

https://proceedings.science/p/85320?lang=pt-br

$$\sum_{b \in \mathcal{B}} \frac{(1+n_b)n_b}{2}, \tag{8}$$

sendo n_b o número de pontos de ônibus no trajeto do ônibus b. Este trabalho procura reduzir o número de variáveis y_{bpq} e parâmetros d_{bpq} para que o problema se torne computacionalmente viável para instâncias grandes. São utilizadas três estratégias de pré-processamento para eliminar esses elementos. Na primeira, uma solução sub-ótima é utilizada para eliminar os $ybpq$ e d_{bpq} para os quais d_{bpq} é maior que o valor sub-ótimo. Na segunda, são eliminados $ybpq$ e d_{bpq} para os quais d_{bpq} é zero. Na terceira estratégia, são eliminados $ybpq$ e d_{bpq} para os quais existe, no trajeto de b, algum ponto que obrigatoriamente é escolhido (isto é, que está no conjunto $\mathcal{I}$) entre p e q. Essas estratégias são detalhadas a seguir e exemplificadas na Figura 5:

- **Eliminação por solução heurística**: No problema dos P-Centros, o objetivo D_{max} é sempre o maior tempo em d_{bpq} para o qual y_{bpq} é igual a um, como definido na Equação 6. Assim, é possível reduzir o espaço de soluções do problema a partir de uma solução viável qualquer. Seja D_{max} a solução ótima do problema proposto e D^*_{max} uma solução viável qualquer. Se D^*_{max} é viável, sabe-se que a solução D_{max} é ou menor ou igual a D^*_{max}. Assim, nenhum $d_{bpq}y_{bpq}$ pode ser maior que D^*_{max}. Portanto, pode-se eliminar todos os atrasos em d_{bpq} que forem maiores do que D^*_{max}, reduzindo o espaço de busca do problema e garantindo a otimalidade. Na Figura 5, essa estratégia é utilizada para eliminar a variável d_{BD}.

- **Eliminação por nulidade**: No problema estudado, um ônibus pode estar ao mesmo tempo em contato com diversos pontos de ônibus. Dessa forma, pode surgir uma sequência com eles, com atraso igual a zero. Essa situação, quando transformada em um problema de P-Centros, gera elementos $d_{bpq}y_{bpq}$, com d_{bpq} igual a zero. Como não há adição de atraso entre esses pontos, pode-se eliminar o elemento $d_{bpq}y_{bpq}$ sem haver perda da garantia de otimalidade. Na Figura 5, essa estratégia é utilizada para eliminar a variável d_{BC}.

- **Eliminação por garantia de escolha**: O problema possui alguns pontos $p \in \mathcal{P}$ que são obrigatoriamente escolhidos. Se o trajeto de um ônibus $b \in \mathcal{B}$ contém p, é possível eliminar $d_{boq}y_{boq}$ para todos os pontos $o, q \in \mathcal{P}$ tais que o ocorra antes de p e q ocorra após p, no trajeto de b. Isso ocorre pois, se p é obrigatoriamente escolhido, todos os dados coletados antes de p são, no máximo, entregues a p. Portanto, não é possível que um dado coletado antes de p seja submetido a um atraso que ocorre após a passagem de b por p. Na Figura 5, essa estratégia é utilizada para eliminar a variável d_{AC}.

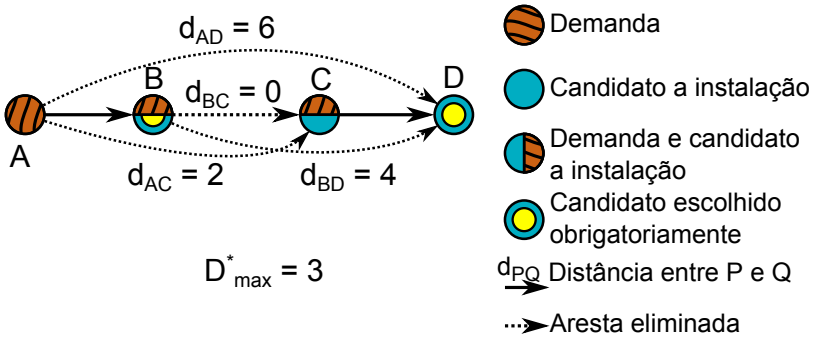


Figura 5: Exemplos de elementos eliminados pelo pré-processamento.

As estratégias de pré-processamento necessitam que seja determinada uma solução viável e um conjunto de pontos que são obrigatoriamente escolhidos. Para tal, utiliza-se um algoritmo

proposto por Cruz et al. [2017a], baseado em uma heurística gulosa que inicialmente assume que todos os candidatos estão escolhidos e, iterativamente, remove pontos de coletas até que o número de pontos de coleta seja igual a $N_{desejado}$. A escolha do ponto de coleta a ser removido é baseada em um critério denominado atraso de remoção. O atraso de remoção de um ponto p é o maior atraso gerado na rede caso p seja removido. Ou seja, o atraso de remoção de p é o maior $d_{bop} + d_{bpq}$ para cada ônibus b que passa por p saindo de o e indo para q. A cada iteração do algoritmo, remove-se o ponto com menor atraso de remoção até que número de pontos restantes seja igual a $N_{desejado}$. O maior atraso na solução heurística é denotado por D^*_{max}.

Na primeira iteração do algoritmo, calcula-se o atraso de remoção para todos os pontos com a rede completa. Se, com a rede completa, a remoção de algum ponto $p \in \mathcal{P}$ pode gerar um atraso maior do que D^*_{max}, então p não pode ser removido, e é obrigatoriamente escolhido. O conjunto de pontos que não podem ser removidos é a união de $\mathcal{I}$ com o conjunto de pontos com atraso de remoção inicial maior do que D^*_{max}.

6. Aplicação das Estratégias em um Cenário Real

O método proposto neste trabalho é utilizado em um cenário real. São coletados dados sobre os ônibus da Zona Sul da cidade do Rio de Janeiro, utilizando a ferramenta oferecida pelo IPLANRIO [2016a]. Essa ferramenta disponibiliza os dados de posição medidos com GPS dos ônibus da cidade do Rio de Janeiro a cada minuto. São armazenados dados desde 0h do dia 5 de novembro de 2017 até 0h do dia de 6 de novembro de 2017 dos ônibus circulando pela Zona Sul da cidade. Também são obtidos os dados das posições dos pontos de ônibus da cidade do Rio de Janeiro, disponibilizados por IPLANRIO [2016b]. Para este trabalho, um ônibus é considerado na Zona Sul caso esteja dentro do quadrado definido pelos pares ordenados de latitude e longitude (-23.013874,-43.296139) e (-22.918251,-43.149197). São considerados apenas os ônibus que, durante o período observado, não saem da Zona Sul. A partir das posições dos ônibus e dos pontos de ônibus, é construído um conjunto de dados com os contatos entre nós de sensoriamento embarcados nos ônibus e pontos de coleta instalados nos pontos de ônibus. Considera-se que se um ônibus encontra-se a 300 m de um ponto de ônibus, então existe contato entre ambos. Tal distância é utilizada pois é uma distância típica para comunicação em redes veiculares, como mostrado no trabalho de Gozálvez et al. [2012].

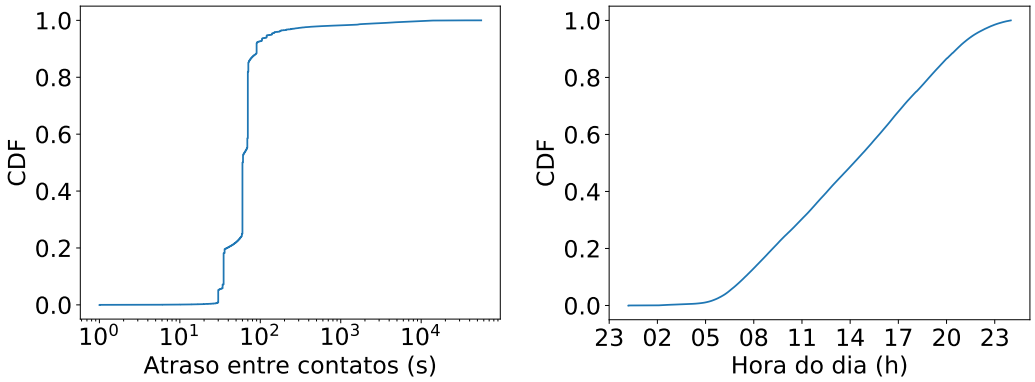
Segundo Dias e Costa [2016], os dados coletados nos ônibus do Rio de Janeiro podem conter inconsistências. A fim de detectar tais problemas, a Seção 6.1 realiza uma análise dos dados e mostra-se os filtros aplicados antes de utilizar os dados no método proposto.

6.1. Análise e preparação do conjunto de dados

Pela observação do conjunto de dados construído, é possível notar que alguns ônibus permanecem por diversas horas sem encontrar nenhum ponto de ônibus. O gráfico na Figura 6(a) mostra a distribuição cumulativa do tamanho dos atrasos. É possível notar que mais de 90% dos atrasos são menores do que 1000 s. A Figura 6(b) apresenta a distribuição cumulativa de contatos em função da hora do dia. Pode-se notar que durante a madrugada existe uma redução no número de contatos. Uma hipótese é a de que haja mais ônibus fora de circulação. Assim, são considerados apenas os contatos acontecidos entre 8 h e 22 h.

Utilizando os valores de atraso definidos por Zanella et al. [2014], considera-se que atrasos maiores do que 30 minutos são atrasos grandes. O valor serve de referência para a detecção dos pontos inicial e final de cada trajeto, para cálculo do conjunto $\mathcal{I}$. Todos os trajetos são analisados, e define-se como o primeiro ponto de um trajeto o ponto imediatamente depois do primeiro atraso de 30 minutos ou mais daquele trajeto e, como último ponto de um trajeto, o ponto imediatamente antes do último atraso de 30 minutos ou mais do trajeto. São eliminados do conjunto de dados os ônibus cujo trajeto tenha menos de dois encontros com pontos e também os ônibus que possuem algum atraso com mais de 30 minutos. Considera-se que estes ônibus não são adequados para garantir a entrega apropriada das mensagens. A Tabela 2 apresenta algumas características do conjunto de dados obtidos.

https://proceedings.science/p/85320?lang=pt-br



(a) Distribuição cumulativa dos atrasos sem nenhum filtro.

(b) Distribuição de contatos ao longo de um dia.

Figura 6: Atributos do conjunto de dados coletado antes de filtragem inicial.

Tabela 2: Características do conjunto de dados obtido.

Parâmetro	**Valor (#)**
Ônibus	116
Pontos de ônibus	6.310
Pontos de ônibus com algum contato	744
Total de contatos	485.208
Pontos de ônibus iniciais e finais	116

A Figura 7(a) ilustra a distribuição cumulativa dos atrasos após a eliminação dos encontros ocorridos antes de 8h e após 22h, e após a eliminação dos ônibus cujo trajeto possua algum intervalo com mais de 30 minutos sem nenhum contato. A Figura 7(b) ilustra a distribuição cumulativa da quantidade de contatos de cada ônibus no conjunto de dados. O conjunto de dados obtido é utilizado para escolher quais pontos de ônibus devem receber pontos de coleta.

6.2. Resultados obtidos pelas estratégias de pré-processamento

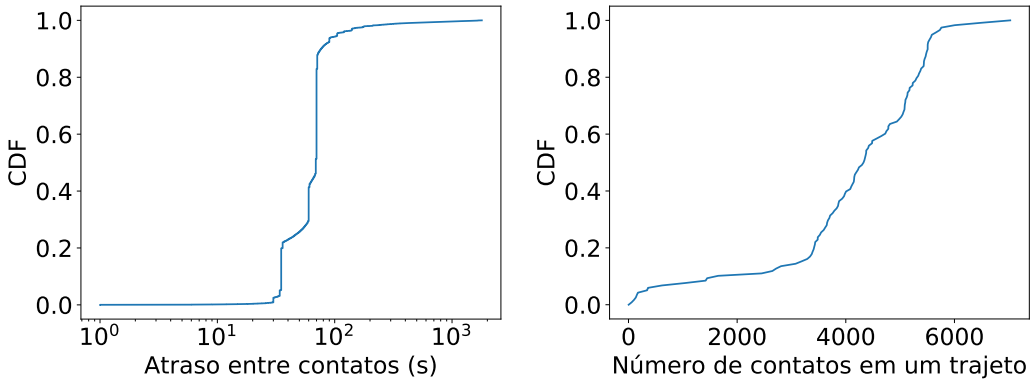
O pré-processamento definido na Seção 5 é aplicado ao conjunto de dados obtido na Seção 6.1. A Tabela 3 exibe os valores de atraso máximo D^*_{max} obtido pelo algoritmo, para cada orçamento $N_{desejado}$. Pode-se observar que, a partir de 120 pontos escolhidos, não existe melhora no atraso máximo da rede, segundo a heurística. Os resultados são obtidos em um processador *Intel(R) Xeon(R) X5570* de 2.93GHz, com 100 GB de memória RAM.

Tabela 3: Resultados obtidos pela heurística.

Número de pontos escolhidos	**Maior atraso da rede**	**Tempo de execução (s)**
116	2067	2,35
117	1917	2,37
118	1848	2,39
119	1801	2,40
120 ou mais	1794	2,42 (para 120)

Os valores obtidos pela heurística são usados para criar o conjunto de dados de entrada para o problema formulado na Seção 4. A Tabela 4 exibe a quantidade de restrições resultantes da Equação 6 antes e após o filtro. Os valores de $N_{desejado}$ empregados na solução ótima de Seção 6.3 são todos maiores do que 120. Assim, para construir as variáveis do problema, o pré-processamento

https://proceedings.science/p/85320?lang=pt-br



(a) Distribuição cumulativa dos atrasos após filtragem. (b) Distribuição do número de contatos após filtragem.

Figura 7: Atributos do conjunto de dados coletado após filtragem por horário e por atrasos maiores que 30 minutos.

é realizado apenas uma vez, pois os resultados da heurística são todos iguais para valores maiores do que 120 pontos escolhidos.

Tabela 4: Quantidade de restrições do tipo $d_{bpq}y_{bpq}$ no problema.

Tipo de pré-processamento	**Número de elementos do tipo** $d_{bpq}y_{bpq}$
Sem pré-processamento	1.122.629.838
Pré-processamento para 120 pontos escolhidos	393.157

6.3. Resultado obtidos pela solução ótima

Os dados obtidos servem de entrada para o problema elaborado na Seção 3. São realizados experimentos com $N_{desejado}$ igual a 200, 300, 400, 500, 600 e 700, uma vez que há 744 pontos com os quais há algum contato e 116 devem ser obrigatoriamente escolhidos. O problema é executado utilizando a ferramenta IBM ILOG CPLEX 12.5.1, em um processador *Intel(R) Xeon(R) X5570* de 2.93GHz, com 100 GB de memória RAM. O trabalho considera como principais resultados os valores numéricos de D_{max}, como também o tempo de execução para cada um dos valores de $N_{desejado}$ executado. Esses resultados se encontram na Tabela 5.

Tabela 5: Resultados das execuções do problema de escolha de pontos de ônibus.

$N_{desejado}$	D_{max} **(s)**	**Tempo de execução (s)**
200	1786	1601
300	1786	1643
400	1786	1354
500	1786	1376
600	1786	1378
700	1786	1381

As instâncias do problema que sofreram o pré-processamento convergem para a solução ótima em até 1601 s, em todos os casos executados, enquanto as instâncias que não sofreram o pré-processamento demoram mais de 50 h, sem convergir. Após as 50 h, a execução das instâncias sem pré-processamento não começou a resolver o problema, encontrando-se ainda na fase de adição de restrições lineares.

7. Conclusões e Trabalhos Futuros

O sensoriamento urbano é uma importante etapa na criação de cidades inteligentes. A utilização de redes de sensores móveis pode reduzir custos, a partir da redução de número de nós

https://proceedings.science/p/85320?lang=pt-br

necessários para sensoriamento e também do número de pontos de coleta para o recebimento das informações coletadas. Uma estratégia possível é a de utilizar ônibus para obter dados e pontos de ônibus para servir de pontos de coleta. Como os ônibus realizam comunicação oportunística com os pontos de ônibus, existe a preocupação no atraso das informações coletadas. Adicionalmente, se houver alguma restrição orçamentária que impeça que todos os pontos de ônibus sejam equipados com pontos de coleta, é importante que a distribuição dos equipamentos pelos pontos de ônibus leve em consideração o atraso na entrega de mensagens. Do ponto de vista da coleta de dados que visem algum tipo de alerta, é importante saber qual o maior atraso que uma mensagem pode sofrer por toda a rede.

Este trabalho propôs um método para a escolha ótima dos pontos de coleta de maneira a minimizar o maior atraso sofrido pelas mensagens na rede. Para tal, modelou-se o problema da escolha de pontos de ônibus como um problema do tipo P-Centros e reduziu-se o espaço de busca do problema utilizando uma solução sub-ótima obtida a partir de uma heurística. O método foi aplicado em dados dos ônibus da Zona Sul da cidade do Rio de Janeiro. Os resultados mostraram que o método utilizado permitiu que fosse possível obter uma solução para o problema em menos de 25 minutos. Em contraste, não foi possível encontrar solução ótima para o problema utilizando o *hardware* disponível sem a utilização do pré-processamento proposto.

Como trabalhos futuros, pretende-se aplicar o algoritmo de escolha de pontos de coleta para toda a cidade do Rio de Janeiro, de forma a criar bases para uma rede que possa servir toda a cidade. Duas estratégias possíveis são reduzir o conjunto de dados contendo todos os ônibus da cidade ou resolver o problema da escolha de pontos de coleta para diferentes partes da cidade, obtendo uma solução sub-ótima, e analisar a qualidade dessa solução.

Referências

Bonola, M., Bracciale, L., Loreti, P., Amici, R., Rabuffi, A., e Bianchi, G. (2016). Opportunistic communication in smart city: Experimental insight with small-scale taxi fleets as data carriers. *Ad Hoc Networks*, 43:43–55.

Cruz, P., Couto, R. S., e Costa, L. H. M. K. (2017a). An algorithm for sink positioning in bus-assisted smart city sensing. *Future Generation Computer Systems*.

Cruz, P., Couto, R. S., e Costa, L. H. M. K. (2017b). Um algoritmo de posicionamento de pontos de coleta para uma rede de sensores baseada em Ônibus urbanos. *XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuıdos*, 14(1).

Cruz, P., da Silva, F. F., Pacheco, R. G., Couto, R. S., Velloso, P. B., Campista, M. E. M., e Costa, L. H. M. K. (2017c). Sensingbus: um sistema de sensoriamento baseado emˆonibus urbanos. *XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuıdos*, 14(1).

Davidović, T., Ramljak, D., Šelmić, M., e Teodorović, D. (2011). Bee colony optimization for the p-center problem. *Computers & Operations Research*, 38(10):1367–1376.

Dias, D. e Costa, L. H. M. K. (2016). Análise da capacidade de dados de uma rede de Ônibus urbanos. In *XXXIV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. SBrT.

Dong, W., Guan, G., Chen, Y., Guo, K., e Gao, Y. (2015). Mosaic: Towards city scale sensing with mobile sensor networks. In *IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, p. 29–36. IEEE.

Ekici, E., Gu, Y., e Bozdag, D. (2006). Mobility-based communication in wireless sensor networks. *IEEE Communications Magazine*, 44(7):56 – 62.

Gozálvez, J., Sepulcre, M., e Bauza, R. (2012). IEEE 802.11p vehicle to infrastructure communications in urban environments. *IEEE Communications Magazine*, 50(5).

https://proceedings.science/p/85320?lang=pt-br

IPLANRIO (2016a). Descrição do dataset conjunto gps ônibus. Disponível em `http://data.rio/dataset/gps-de-onibus`.

IPLANRIO (2016b). Documentação de paradas das linhas de ônibus. Disponível em `http://data.rio/dataset/pontos-de-parada-de-onibus`.

Kariv, O. e Hakimi, S. L. (1979). An algorithmic approach to network location problems. i: The p-centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538.

Liu, B., Brass, P., Dousse, O., Nain, P., e Towsley, D. (2005). Mobility improves coverage of sensor networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*. MobiHoc.

Marjovi, A., Arfire, A., e Martinoli, A. (2015). High resolution air pollution maps in urban environments using mobile sensor networks. In *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, p. 11 – 20. IEEE.

Pereira, R., Garcia, L., Melo, V., Texeira, P., e Boaventura-Netto, P. (2004). Distribuição de cabines de segurança em parte do bairro do Leblon na cidade do Rio de Janeiro. In *Simpósio Brasileiro de Pesquisa Operacional–SBPO*.

Rubinstein, M. G., Abdesslem, F. B., De Amorim, M. D., Cavalcanti, S. R., Alves, R. D. S., Costa, L. H. M. K., Duarte, O. C. M. B., e Campista, M. E. M. (2009). Measuring the capacity of in-car to in-car vehicular networks. *IEEE Communications Magazine*, 47(11).

Wong, J. L., Jafari, R., e Potkonjak, M. (2004). Gateway placement for latency and energy efficient data aggregation [wireless sensor networks]. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, p. 490–497. IEEE.

Youssef, W. e Younis, M. (2007). Intelligent gateways placement for reduced data latency in wireless sensor networks. In *IEEE International Conference on Communications, 2007. ICC'07.*, p. 3805–3810. IEEE.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., e Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32.

Zhao, D., Ma, H., Li, Q., e Tang, S. (2016). A unified delay analysis framework for opportunistic data collection. *Wireless Networks*, p. 1–13.

https://proceedings.science/p/85320?lang=pt-br