

# Classificação GeoSocial de Contatos para Disseminação de Dados em Redes Veiculares Oportunistas

Carlos R. Guerber<sup>1,2</sup>, Eduardo L. Gomes<sup>1,2</sup>, Mauro Fonseca<sup>1</sup>, Anelise Munaretto<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEl)  
Universidade Tecnológica Federal do Paraná (UTFPR)  
Centro – 80.230-901 – Curitiba – PR – Brasil

<sup>2</sup>Departamento de Informática  
Instituto Federal de Santa Catarina (IFSC) – Canoinhas – SC – Brasil

carlos.guerber@ifsc.edu.br, maurofonseca@utfpr.edu.br

**Abstract.** *Research in opportunistic vehicular networks has attracted attention in contact selection due to its application in data communication networks. We present a GeoSocial model for contact selection and message routing in urban environments based on the amplitude and frequency of movement and on the vehicle's social structure. We consider the temporality in the formation of the network links to extract these pieces of information. The model was evaluated in a real taxi drive movement database. The results show that with a small number of vehicles it is possible to achieve superior outcome in the delivery rate with low relative overhead compared to other routing protocols.*

**Resumo.** *As pesquisas em redes veiculares oportunistas têm atraído atenção no que se refere à seleção de contatos para aplicação em redes de dados. Apresentamos um modelo GeoSocial para seleção de contatos e roteamento de mensagens em ambientes urbanos baseado na amplitude e frequência de movimentação e na estrutura social dos veículos. Consideramos a temporabilidade na formação dos enlaces na rede para extrair estas informações. O modelo foi avaliado em uma base de dados real de movimentação de táxis. Em decorrência mostramos que com um pequeno número de veículos é possível atingir resultados superiores na taxa de entrega com baixo overhead relativo em relação a outros protocolos de roteamento.*

## 1. Introdução

Estudos recentes têm utilizado grandes volumes de dados da movimentação de veículos em regiões urbanas [Vaz De Melo et al. 2015, Celes et al. 2017]. Uma atenção especial tem sido dada à dinâmica da movimentação de veículos que não sejam de uso particular (e.g. táxis, ônibus, veículos do corpo de bombeiros). Estas redes são conhecidas como VANETs (*Vehicular Adhoc NETWORKS*).

Compreender a mobilidade destes veículos tem grande importância para o desenvolvimento de protocolos de comunicação que venham a explorar seus encontros oportunistas. O principal problema está em escolher, da melhor forma, os contatos que serão utilizados no futuro para comunicação V2V (Vehicle to Vehicle). As atividades veiculares diárias geram dados de movimentação de espaço e tempo, e podem ser utilizados para

estimar o potencial de um veículo que atue como transmissor de mensagens em uma rede de dados.

Mesmo que os veículos apresentem regularidade em sua movimentação, situações aleatórias acontecem. Os táxis são um exemplo e estas situações são de difícil previsão, pois podem não se repetir. Neste caso uma rede de contatos formada por veículos é um exemplo de uma rede veicular oportunista, pois sua mobilidade cria conexões aleatórias e intermitentes entre os componentes da rede no decorrer do tempo.

Estas redes geram relacionamentos sociais e para os identificar é preciso descobrir como a regularidade da estrutura da rede influencia no comportamento de um determinado indivíduo e assim, como este será classificado de acordo com o tipo de relacionamento que mantém com outros indivíduos [Granovetter 1977]. Neste trabalho propomos um modelo para selecionar veículos a partir das informações temporais existentes em uma base de dados de movimentação real a fim de utilizar como contatos. Em seguida foi proposto um protocolo de roteamento GeoSocial para realizar o roteamento de mensagens em redes de dados veiculares oportunistas.

O modelo extrai seis características sociais e uma característica geoespacial da movimentação de táxis na cidade de Roma, Itália. Utilizamos medidas sociais de centralidade para verificar a importância dos veículos e medidas sociais de coesão para verificar a intensidade do vínculo existente entre dois veículos. Ainda, utilizando os dados das coordenadas geográficas, o modelo calcula a amplitude e a frequência de movimentação geoespacial por meio do cálculo do raio de giro dos veículos. Os valores das características são ponderados em uma função de custo de probabilidade de contato futuro e utilizadas pelo protocolo de roteamento. De forma resumida este trabalho contribui com:

- Uma nova forma de seleção que utilize um número reduzido de contatos prioritários para transferências de mensagens por meio de um protocolo de roteamento baseado em características sociais e geoespaciais.

Este trabalho está organizado da seguinte forma: na seção 2 apresentamos os trabalhos correlatos; na seção 3 a descrição da base de movimentação de táxis; na seção 4 apresentamos as características extraídas para os veículos e as correlacionamos; na seção 5 explicamos a aplicação da função custo e o funcionamento do protocolo GeoSocial; na seção 6 avaliamos os resultados e na seção 7 apresentamos as conclusões.

## **2. Trabalhos Relacionados**

Estudos recentes têm realizado a análise de grandes volumes de dados da movimentação de veículos em regiões urbanas. As pesquisas sobre encaminhamento e disseminação de mensagens em VANETs abrangem assuntos como grafos temporais, padrões frequentes em grafos, análise de dados geoespaciais e análise de redes sociais.

Os autores em [Costa et al. 2015], investigaram a centralidade em grafos temporais e a importância de um nó ao longo da evolução temporal da rede, mostraram a relevância da centralidade na disseminação de dados. [Cunha et al. 2014] analisaram bases de dados reais que descrevem a mobilidade dos veículos. Obtidas as métricas sociais compararam os resultados com grafos randômicos e concluíram que redes veiculares apresentam comportamento social ou seja, possuem interações e rotinas em comum.

Os autores de [Vaz De Melo et al. 2015] ao utilizar dados reais de movimentação de veículos e pessoas, propuseram um modelo eficiente de seleção de contatos e roteamento de mensagens baseado em relacionamentos sociais para classificar a interação entre os nós. Os autores [Cunha et al. 2015] analisaram detalhadamente a mobilidade de veículos a partir de dados reais de movimentação de táxis em grandes cidades. Por meio de técnicas estatísticas, de teoria dos grafos e análise de redes sociais concluíram que os táxis compartilham encontros em comum, porém, não são frequentes.

Em sua pesquisa [Villas et al. 2013] apresentaram um algoritmo geográfico eficiente de partições na rede para disseminação de dados em redes veiculares. O algoritmo elimina *broadcast* e maximiza a capacidade de disseminação na rede. Em seu trabalho [Thakur et al. 2010] utilizaram dados reais de movimentação de pessoas para verificar o impacto das informações geoespaciais na rede. No trabalho [Hui et al. 2011] apresentaram um protocolo de roteamento oportunista que utiliza a análise de redes sociais e separação em agrupamentos/comunidades.

As pesquisas relacionadas à seleção de contatos e encaminhamento de mensagens em redes veiculares são abrangentes e diversificadas, porém, abre espaço para investigar técnicas de movimentação geoespacial combinadas com técnicas de análise de redes sociais.

### **3. Dados de Movimentação Veicular**

Neste trabalho utilizamos dados da movimentação de táxis na cidade de Roma, Itália [Bracciale et al. 2014]. As capturas ocorreram entre os dias 01/02 a 02/03/2014, totalizando trinta dias de 24 horas. As leituras foram realizadas a cada sete segundos por GPS instalados no interior de 320 táxis. O arquivo contém aproximadamente 1,5GB com 21,8 milhões de linhas, cada linha corresponde a uma leitura. Os dados coletados possuem o identificador único dos veículos, data e hora da leitura e posição geográfica do veículo (LATITUDE, LONGITUDE) no formato decimal.

Grafos são formados por nós e arestas. Os nós são os veículos. Para encontrar as arestas calculamos a distância em metros entre todos os pares de veículos. Conforme proposto em [Hui et al. 2011], a distância entre 1 e 200 metros<sup>1</sup> entre um par de nós é classificada como uma aresta. Calculamos a distância com a fórmula de *Harversine* [Robusto 1957], que fornece a distância entre dois pontos em uma esfera a partir da latitude e longitude. O raio da Terra utilizado é igual a 6372km. O arquivo de saída contém: o par de identificadores dos táxis os quais formam as arestas do grafo não direcionado, o tempo em segundos (horário) em que foi realizada a leitura, a latitude e a longitude.

### **4. Caracterização da Mobilidade Veicular**

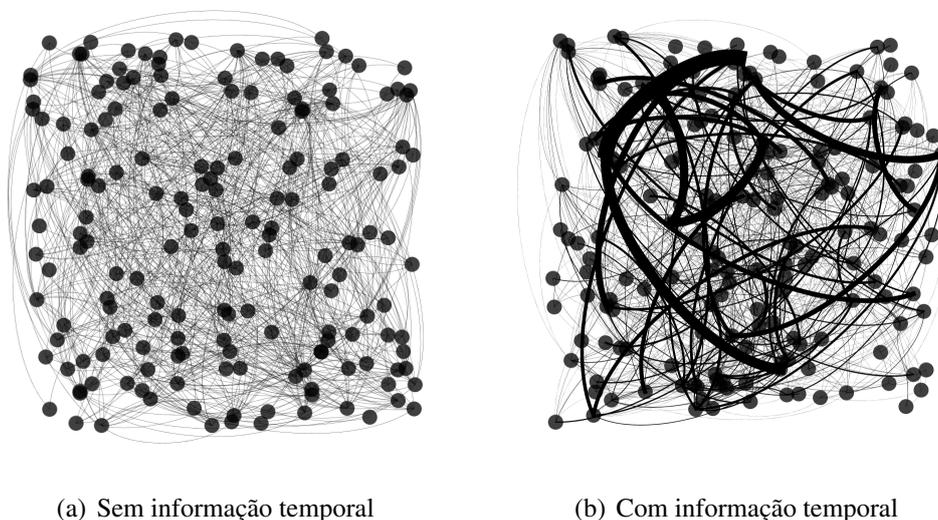
Redes veiculares não podem ser representadas de maneira estática devido a característica dinâmica da formação das arestas no grafo no decorrer do tempo [Costa et al. 2015], pois muitas informações se perdem ao descartar a informação temporal. Isto inclui as possíveis correlações entre a ativação e desativação de arestas, conforme estas arestas aparecem e desaparecem no decorrer do ciclo de atividades. Em cada aresta há um registro do tempo em que ocorreu a ligação, fundamental para a classificação.

---

<sup>1</sup>Estes valores foram assumidos para determinar a área de alcance para comunicação via rádio em redes de comunicação de dados sem fio.

A formação dinâmica das arestas é observada ao especificar um intervalo de tempo entre um instante inicial  $t_i$  e um instante final  $t'_i$  de uma janela de tempo  $(0, T_{max})$  (e.g. 0 a 900 segundos de uma hora, 3 às 6 horas de um dia, de um dia para o próximo dia), onde  $t'_i > t_i$  e  $(t_i, t'_i) \subset (0, T_{max})$ .

Na Figura 1 foi escolhido o dia com a menor densidade<sup>2</sup> para representar o grafo (173 veículos e 735 arestas), apenas para facilitar a visualização e ilustração do exemplo. O mesmo procedimento foi realizado para todos os dias e independente da densidade ou da janela de tempo, o comportamento da visualização é análogo. Assim observamos pela espessura das arestas no grafo, Figura 1(b), que para cada aresta é atribuído um peso ou valor, ao contrário do que mostra a Figura 1(a), em que nenhum peso ou valor é considerado às arestas do grafo além de uma simples ligação.



**Figura 1. Distribuição dos veículos no grafo para a janela de tempo de 24 horas (0-86400 segundos) do dia 02-03-2014, um domingo: (a) informação temporal de ativação das arestas desconsiderada; (b) informação temporal de ativação das arestas considerada.**

Assim, dada qualquer janela de tempo  $(t_i, t'_i)$  o objetivo é extrair as informações de  $(t_i, t'_i)$  para classificar os veículos e os utilizar em um intervalo de tempo futuro  $(t_f, t'_f)$ , onde  $0 \leq t_i < t'_i < t_f < t'_f$ . A rigor  $(t_i, t'_i)$  é uma janela de tempo de treinamento e  $(t_f, t'_f)$  é uma janela de tempo de teste, ambos devem corresponder a um mesmo intervalo de tempo em janelas de tempo diferentes, respeitando as condições apresentadas acima.

#### 4.1. Métricas Sociais

Nas janelas de tempo a formação do grafo permite extrair características sociais e assim realizar a análise de redes sociais [Scott 2017, Masuda and Lambiotte 2016]. O interesse está em analisar as ligações formadas pelos táxis e identificar os nós com maior importância de acordo com as seguintes medidas de centralidade: Grau (Gra), Intermediação (Int), Proximidade (Pro) e medidas de coesão: Laços Sociais (Ausentes (Wa), Fortes (Ws)

<sup>2</sup>Para considerar a densidade utilizamos a quantidade de veículos e a quantidade de arestas formadas durante uma janela de tempo qualquer.

e Fracos(Wu)). Os cálculos das medidas de centralidade foram feitos por meio da biblioteca *Networkx* em Python [Hagberg et al. 2008, Borgatti and Halgin 2011].

A análise das interações entre os veículos pode ser feita por meio de duas propriedades: centralidade e coesão [Scott 2017]. Quando a análise é feita por meio da centralidade o nível de análise é sobre o veículo, quando feito por meio de coesão o nível é sobre pares de veículos.

É importante conhecer as medidas de centralidade do ponto de vista de aplicação. O *grau* é a quantidade de contatos diretos que um determinado nó possui. A importância de um veículo está relacionada à quantidade de veículos ao qual está conectado isto é, quanto maior, mais importante se torna na rede. Cada veículo  $v_i$  terá um grau  $C_G(v_i)$  (Equação 1), onde  $j$  é o veículo ao qual  $i$  forma a aresta  $a_{ij}$ :

$$C_G(v_i) = \sum_j a_{ij} \quad (1)$$

A *intermediação* é a capacidade do veículo em participar como ponto de articulação e fazer parte do caminho entre veículos de regiões diferentes. A importância de um veículo está relacionada com a capacidade de fazer parte de vários caminhos isto é, quanto mais caminhos participar, mais importante se torna na rede. Cada veículo  $v_i$  terá uma intermediação  $C_I(v_i)$  (Equação 2), onde  $v, a, b \in V$ , sejam  $g_{ab}$  o número de geodésicas<sup>3</sup> entre o veículo  $a$  e o veículo  $b$  e  $g_{avb}$  o número destas que passam pelo veículo  $v$ .

$$C_I(v_i) = \sum_{a,b \neq v} \frac{g_{avb}}{g_{ab}} \quad (2)$$

A *proximidade* considera ser mais importante estar próximo de outros veículos do que ter uma grande quantidade de conexões. A importância do veículo está relacionada a sua proximidade aos demais veículos isto é, quanto mais próximo mais importante se torna da rede. Cada veículo  $v_i$  terá uma proximidade  $C_P(v_i)$  (Equação 3), que é dada pelo recíproco da soma de suas distâncias  $d(v, a)$  no grafo  $V$  com os demais veículos  $a$ .

$$C_P(v_i) = \frac{1}{\sum_{a \in V} d(v, a)} \quad (3)$$

Para explorar a coesão por meio dos laços sociais dos veículos no grafo, definimos um Peso Agregado  $w_{ij}$  (Equação 4) [Masuda and Lambiotte 2016], onde cada peso entre o veículo  $i$  e o veículo  $j$  corresponde ao número de contatos observados e recebe uma marcação a qual identifica o contato em três níveis de relacionamento:  $S = Forte(Ws); U = Fraco(Wu); A = Ausente(Wa)$ . O conjunto de dados obtidos em  $W_i$  é ordenado de forma crescente, isso permite calcular o quartil e marcar os laços como  $25^{o}percentil = A$ ,  $50^{o}percentil = U$  e o  $75^{o}percentil = S$ . O Peso Agregado  $w_{ij}$ , onde o valor dado para cada ocorrência de contato é  $c_{ij} = 1$  é dado por:

$$w_{ij} = \sum c_{ij} \quad (4)$$

Conforme propôs [Granovetter 1977] com a existência maior de laços fracos ( $Wu$ ) os nós têm menor vínculo no ambiente social mais próximo e as conexões são

<sup>3</sup>Geodésica é o caminho mais curto (ou mais longo) entre dois pontos em um espaço tridimensional.

mais variadas e menos intensas. São mais eficazes em atingir outros indivíduos e criar mais ligações. Os laços fortes ( $Ws$ ) permitem alto grau de confiança entre um par de nós, pois sua relação é muito próxima, são importantes localmente. A Tabela 1 resume as propriedades de coesão e centralidade conforme suas aplicações.

**Tabela 1. Aplicação das propriedades de centralidade e coesão**

Medidas	Característica
Grau	Conexões diretas, importância local na troca de informações.
Intermediação	Ponte entre grupos ou veículos. Intermedia e direciona a informação.
Proximidade	Transmite a informação com facilidade. Mais independente.
Laços Fortes	Troca de informação com alto grau de confiança. Relação muito próxima.
Laços Fracos	Troca de informação entre mais indivíduos. Relação mais fraca no ambiente local.
Laços Ausentes	Relações de troca de informação casual com baixo grau de confiança.

## 4.2. Característica Geoespacial

Outra característica existente no grafo são as mudanças da posição geográfica, veículos têm uma grande atividade de movimentação em várias direções. A partir da latitude e longitude é possível calcular a amplitude da movimentação espacial média, conhecida como Raio de Giro  $r_g$  [Cheng et al. 2011], que retorna as distâncias percorridas entre as diferentes localizações e o centro de massa de um veículo, como mostra a Equação 5. Onde  $r_i$  é uma leitura de movimentação dada pelo par (latitude, longitude),  $r_{cm}$  é o centro de massa das movimentações dos veículos (média de todas as movimentações). O valor  $(r_i - r_{cm})$  é a distância entre uma leitura e o centro de massa.  $r_g$  é dado em metros.

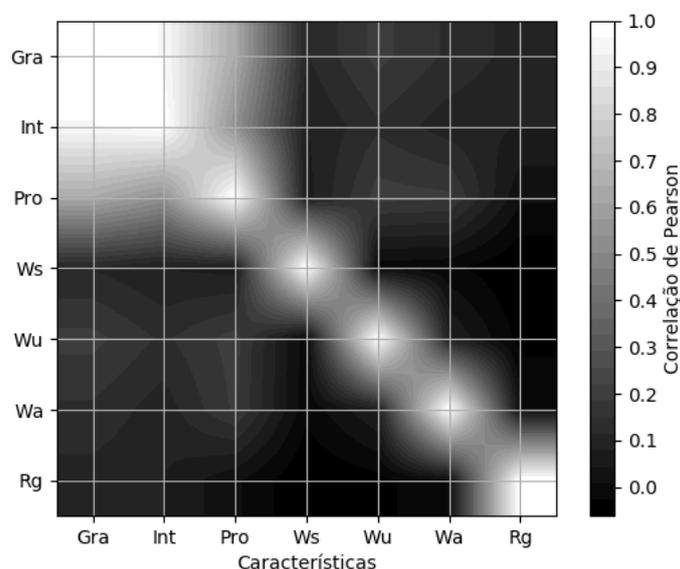
$$r_g = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - r_{cm})^2}, r_{cm} = \left\{ \left( \frac{1}{n} \sum Lat \right), \left( \frac{1}{n} \sum Lon \right) \right\} \quad (5)$$

Veículos com maior movimentação se afastam com mais frequência e a maiores distâncias do seu centro de massa. Um raio de giro baixo indica movimentação local e com pouca variação de distância, estes veículos tendem a ser mais importantes localmente. Um raio de giro alto indica que a movimentação é mais variável a maiores distâncias, estes veículos tendem a ser importantes não só localmente, mas também na interligação de localidades.

### 4.2.1. Correlação Entre as Características

Por meio da aplicação da matriz de correlação (Figura 2) realizamos a análise simultânea da associação entre as características dos veículos extraídas da base de movimentação. O Coeficiente de Correlação de Pearson [Benesty et al. 2009] expressa o grau de dependência linear entre as características. Os valores variam entre  $-1$  e  $+1$ , quanto mais próximo de *zero* menor a correlação, se igual a *zero* não existe correlação entre as características. A correlação apresentada corresponde a média dos resultados para cada uma

das características extraídas nos trinta dias. O intervalo de confiança de 95% com precisão  $p$  para: Grau ( $p = 0,032$ ), Intermediação ( $p = 0,003$ ), Proximidade ( $p = 0,014$ ), Laços Fortes ( $p = 1,258$ ), Laços Fracos ( $p = 0,015$ ), Laços Ausentes ( $p = 0,004$ ) e Raio de Giro ( $p = 0,028$ ).



**Figura 2. Matriz de Correlação Linear de Pearson entre as características sociais (Gra=Grau, Int=Intermediação, Pro=Proximidade, Ws=Laços Fortes, Wu=Laços Fracos, Wa=Laços Ausentes) e geográfica (Rg=Raio de Giro) obtidas para os táxis.**

Observamos que as características sociais, sejam elas de centralidade ou coesão, não apresentam correlação com a característica geoespacial. Da mesma forma as características sociais de centralidade e coesão não apresentam correlação entre si. Estas duas afirmações são importantes, pois estas características são utilizadas de formas distintas na composição da função custo de seleção de veículos e se fossem correlacionadas sua aplicação poderia ser redundante. Apenas as características de Grau e Intermediação se mostram correlatas.

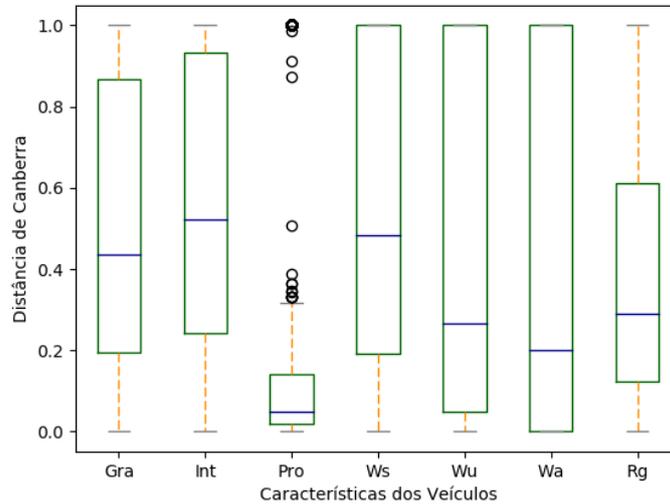
## 5. Classificação GeoSocial

Calculamos a distância dos resultados de cada característica de um intervalo de tempo para o outro a fim de verificar se os valores encontrados na janela de tempo de treinamento tendem a se repetir na janela de tempo de teste. Utilizamos a Distância de Canberra<sup>4</sup>[Lance and Williams 1966], que quanto mais próximo de *zero* for a distância mais próximas as variáveis estão.

A Figura 3 apresenta a dispersão da distância para as sete características avaliadas nos 30 dias. Se observarmos pela média as distâncias tendem a ficar abaixo de 0,5, isto indica que os veículos tendem a repetir os valores das características. Também observamos a existência de muitas distâncias próximas de 1, pode ser causado por veículos que

<sup>4</sup>A distância de Canberra é uma medida de distância numérica utilizada em pares de pontos em um vetor. Se aplica como métrica para comparação de listas.

não mantém as características, estão desativados, ou ausentes naquela janela de tempo de teste. O destaque fica para a medida de centralidade de *proximidade*, em que as distâncias estão muito próximas, indicando um comportamento de repetição dos valores desta característica para os veículos de uma janela de tempo para outra.



**Figura 3. Distribuição das distâncias para as sete características obtidas para os táxis. O intervalo de confiança utilizado foi de 95% com precisão  $p$  para: Grau ( $p = 0,008$ ), Intermediação ( $p = 0,008$ ), Proximidade ( $p = 0,009$ ), Laços Fortes ( $p = 0,008$ ), Laços Fracos ( $p = 0,009$ ), Laços Ausentes ( $p = 0,009$ ) e Raio de Giro ( $p = 0,009$ ).**

### 5.1. Função Custo

A função custo de probabilidade, que propomos na Equação 6, realiza a ponderação dos valores sociais com o valor geográfico como forma de verificar o efeito destas características na seleção dos contatos para encaminhamento de mensagens. Os pesos da função são dados por:  $\alpha$  (relacionado às características sociais  $Social_i$ ) e  $\beta$  (relacionado à característica geoespacial  $Geo_i$ ), sendo que  $0 \geq (\alpha, \beta) \leq 1$ . Assim as ponderações adotadas para os pesos são:  $(\alpha = 1, \beta = 0)$ ,  $(\alpha = 0, \beta = 1)$ ,  $(\alpha = 0,75, \beta = 0,25)$ ,  $(\alpha = 0,25, \beta = 0,75)$  e  $(\alpha = 0,5, \beta = 0,5)$ . Todo  $\alpha + \beta = 1$ .

$$FC_i = \alpha(Social_i) + \beta(Geo_i) \quad (6)$$

A entrada de dados para a função custo  $FC$  ocorre a partir dos valores obtidos das características sociais e da característica geográfica. Estes cálculos são realizados de acordo com uma janela de tempo qualquer. Neste trabalho a janela de tempo de treinamento tem 3 horas (180 minutos). Então a função custo retorna uma lista com o identificador dos veículos em ordem decrescente. Quanto maior o valor calculado como resultado da função custo para um veículo, melhor candidato este veículo será. Esta lista de classificação de veículos aplica-se como métrica para disseminação de mensagens.

### 5.2. Protocolo GeoSocial

Para testar a classificação dos veículos propomos um protocolo de roteamento, desenvolvido no simulador *The One*, chamado *GeoSocial*. O protocolo possui dois parâmetros

de entrada conforme pode ser visualizado no Algoritmo 1: 1) a lista  $FC$  resultado da função custo e 2) o número de hosts  $numH$  que é utilizado para definir quantos veículos da lista  $FC$  são utilizados para realizar o encaminhamento das mensagens. O número de hosts pode variar de 1 até a quantidade total de veículos contida na lista.

---

**Algoritmo 1: PROTOCOLO GEOSOCIAL**

---

**Entrada:**  $FC, numH$   
**Saída:** Mensagem Transferida

```
1 início
2   se (Nó está transferindo) ou (Nó não pode transferir) então
3     | retorna nulo
4   fim
5   se (Nó conectado for destinatário) então
6     | retorna entregar mensagem
7   fim
8   se (Nó conectado estiver em  $FC$ ) e ( $p_i \leq numH$ ) então
9     | retorna transferir mensagem
10  fim
11 fim
```

---

O protocolo atua de maneira simples e atende as seguintes condições: *Linha 2:* verifica se o veículo atual está transferindo alguma mensagem, caso esteja, não pode iniciar outra transmissão ao mesmo tempo; ou verifica se o veículo não pode iniciar uma transmissão, neste caso precisa existir ao menos uma mensagem e uma conexão ativa com outro veículo; *Linha 5:* verifica se o veículo que faz a transmissão está conectado com o destinatário da mensagem, se estiver realiza a entrega diretamente e *Linha 8:* verifica na lista de veículos  $FC$  se os veículos estão listados, desta forma fará o encaminhamento da mensagem somente aos nós classificados pela função custo. Ainda deve respeitar número de hosts que podem ser usados, só utiliza os  $numH$  primeiros elementos da lista, os veículos com a melhor classificação GeoSocial.

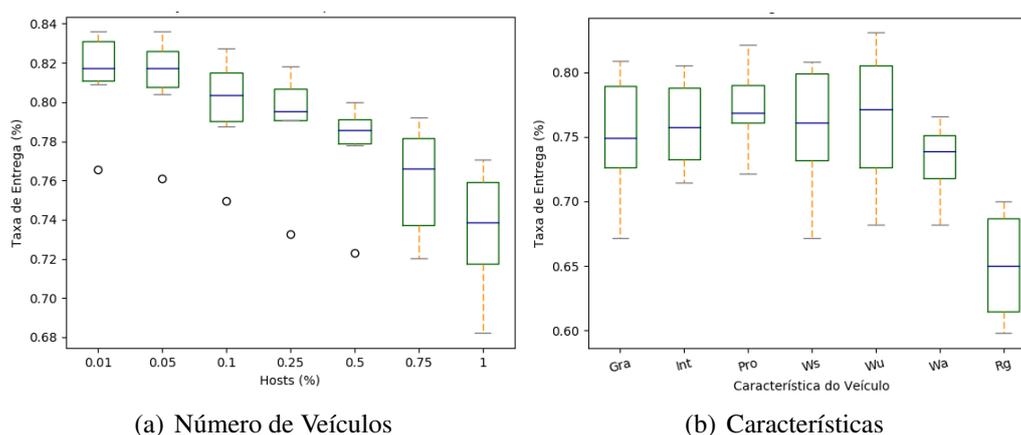
## 6. Avaliação e Resultados

Para realizar a avaliação do modelo de classificação dos veículos e do protocolo GeoSocial são apresentados os seguintes resultados: 1) o impacto do número de veículos (parâmetro  $numH$ ) em relação às características sociais; 2) o impacto dos pesos na função custo em relação às características sociais e 3) da taxa de entrega de mensagens e do *overhead* em relação a outros protocolos de roteamento. Os experimentos foram rodados 18 vezes com janelas de tempo diferentes (janelas consecutivas) e todos os resultados apresentados possuem 95% de intervalo de confiança.

Para realizar as simulações algumas configurações são necessárias. No arquivo de configuração do simulador *The One* o modelo de movimentação é externo, com os dados de movimentação dos táxis de Roma. A janela de tempo utilizada contém períodos de três horas (180 minutos), mesmo tempo da simulação. Outras configurações: Velocidade de transmissão = 9Mbps, alcance do sinal = 200 m, TTL=180min, tamanho do *buffer* = 1 GB e tamanho da mensagem = entre 1 e 100 MB.

## 6.1. Impacto do Número de Veículos no Envio de Mensagens

Neste experimento variamos a quantidade de veículos, entre 1% e 100%, utilizados para realizar o encaminhamento das mensagens no protocolo GeoSocial. Esta variação foi realizada para cada uma das características sociais e da característica geográfica avaliadas. Esta variação ocorre com o ajuste do parâmetro  $numH$ .



**Figura 4. Impacto do número de veículos (parâmetro  $numH$ ) em relação às características na taxa de entrega de mensagens.**

A Figura 4(a) mostra que quanto menor a quantidade de veículos utilizados maior a taxa de entrega de mensagens. Observamos que quando o parâmetro estiver ajustado entre 1 e 10% ocorre uma estabilização na taxa de entrega, e neste caso, uma equiparação na média e desvio padrão. O número máximo (100%) de veículos em cada janela de tempo varia (e.g. número máximo = 61, 152, 184, 284) então quando nos referimos a 1% pode significar mais de um veículo. A taxa de entrega aumenta com o menor grupo de veículos porque somente os veículos que detêm os maiores valores das características sociais ficam à disposição na lista de classificação. Isto implica em dizer que o repasse da mensagem ocorre somente para os veículos que têm maior importância na rede, ao contrário de enviar a mensagem para muitos veículos porém com pouca importância.

Na Figura 4b definimos o parâmetro  $numH = 4$  para mostrar seu impacto em cada uma das características analisadas. As medidas sociais se sobressaem na taxa de entrega em relação à medida geoespacial. A variação entre as características sociais, em torno de 10%, não destaca nenhuma medida como mais apropriada. Todas contribuem para que o repasse aconteça com poucos veículos como nós centrais em um ambiente de rede dinâmico, em que os nós roteadores são móveis.

A medida de centralidade de *Proximidade* obteve melhor resultado em relação às medidas de centralidade de *Grau e Intermediação*, o que nos leva a afirmar que neste cenário os veículos centrais estão mais próximos uns dos outros. Por isso, os veículos são mais independentes na rede o que facilita a troca de informações e enfatiza que a importância dos veículos está na proximidade uns dos outros e não na quantidade de conexões que possuem ou na capacidade de intermediar a troca de informações. Como a medida de centralidade de *Intermediação* tem por característica interligar e ser ponte entre caminhos ou grupos, indica que neste cenário há a formação de agrupamentos ou comu-

nidades de veículos, o que leva a outras abordagens de encaminhamento de mensagens que não fazem parte deste trabalho.

Os resultados obtidos com as medidas de coesão  $Ws$  e  $Wu$  indicam que neste cenário os veículos estabelecem uma forma de vínculo social entre pares, mas cabe uma investigação mais apropriada para definir esta situação. Porém, ao utilizar a informação temporal existente nas conexões entre os veículos, foi mostrado que os laços sociais fortes ( $Ws$ ) são presentes e que a troca de informação ocorre entre poucos veículos que possuem vínculos próximos e assim formam conexões entre si com frequência e se postulam como excelentes nós roteadores. Já com os laços sociais fracos ( $Wu$ ), há uma grande quantidade de veículos que realizam uma maior e mais variada quantidade de conexões com outros veículos atingindo maiores taxas de entrega em relação as demais características e se aplicam como nós roteadores. A medida de coesão ( $Wa$ ) teve o pior desempenho em relação às outras características sociais e seus laços sociais são casuais, as conexões ocorrem com baixa frequência e o vínculo entre os veículos é considerado inexistente.

A característica geoespacial do raio de giro ( $Rg$ ) teve desempenho inferior a todas as características sociais. Mesmo assim, a utilização desta característica mostra que a amplitude de movimentação permite que os veículos sejam escolhidos como nós roteadores. O raio de giro médio de todos os veículos para os 30 dias é de  $4,6 km$  e um intervalo de confiança de 95% com precisão  $p=0,51 km$ . A amplitude do raio de giro varia do *máximo*= $24 km$  ao *mínimo*= $0,1 km$ . Os veículos se posicionam em maior quantidade nas regiões centrais, o que indica que a troca de mensagens ocorre, em sua maioria, a curtas distâncias.

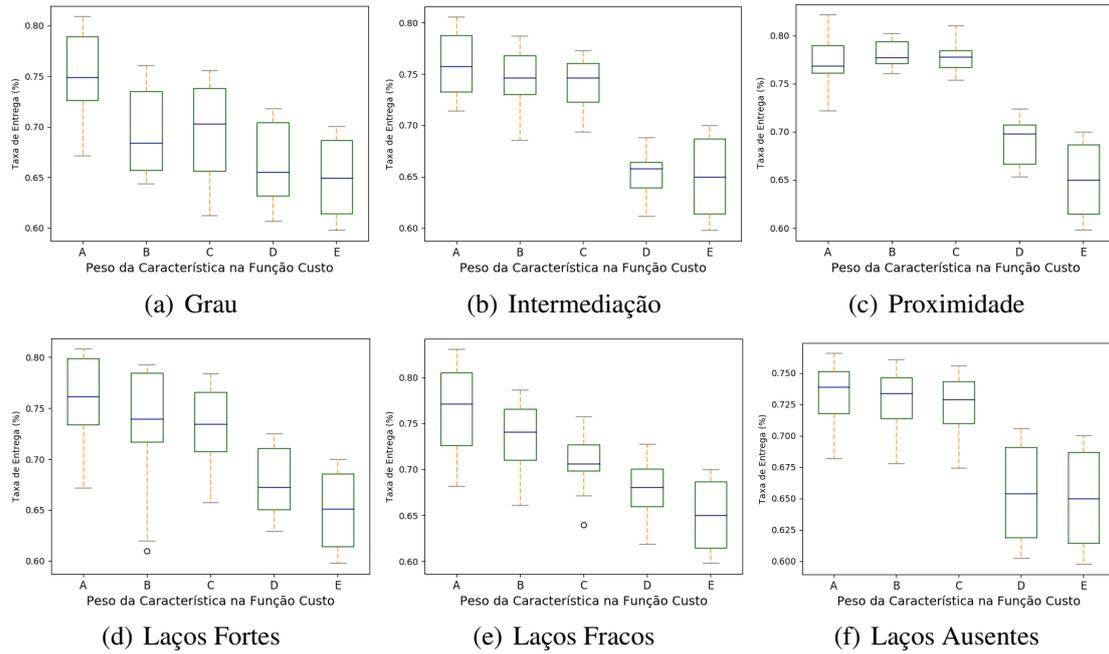
## 6.2. Impacto do Peso das Características dos Veículos no Envio de Mensagens

Neste experimento variamos o peso de cada uma das características sociais, em relação ao raio de giro dos veículos para verificar o impacto na taxa de entrega de mensagens. De acordo com a *Função Custo* proposta, os pesos foram divididos da seguinte forma:  $A = (\alpha = 1; \beta = 0)$ ,  $B = (\alpha = 0,75; \beta = 0,25)$ ,  $C = (\alpha = 0,5; \beta = 0,5)$ ,  $D = (\alpha = 0,25; \beta = 0,75)$  e  $E = (\alpha = 0; \beta = 1)$ . O parâmetro  $num.H = 4$ .

A Figura 5 mostra que conforme o peso das características sociais diminui, a taxa de entrega também diminui. Quando  $\alpha = 1$  atinge a maior taxa de entrega e quando  $\beta = 1$  ocorre a menor taxa. A aplicação das características sociais é mais eficiente que a característica geoespacial para o repasse de mensagens. Neste caso, somente aplicando alguma característica social sem levar em consideração o raio de giro, há obtenção de melhores resultados. Isto não é conclusivo, sendo que outras características geoespaciais como *Centro de Massa* e *Centro de Gravidade* podem ser exploradas.

Quando os valores de  $\alpha = 0,5$  e  $\beta = 0,5$ , nos casos das medidas de *Proximidade* e *Laços Fracos*, por exemplo, notamos que a variação das taxas de entrega de mensagens é muito pequena em relação às melhores taxas. Assim, é possível considerar que o impacto na aplicação das características sociais em relação à característica geoespacial pode melhorar a taxa de entrega de mensagens neste cenário.

Também é importante observar que a troca de mensagens se concentra na região central e em curtas distâncias. Desta forma testes devem ser executados em outras bases de movimentação com características diferentes e com amplitude de movimentação maior para a troca de mensagens.



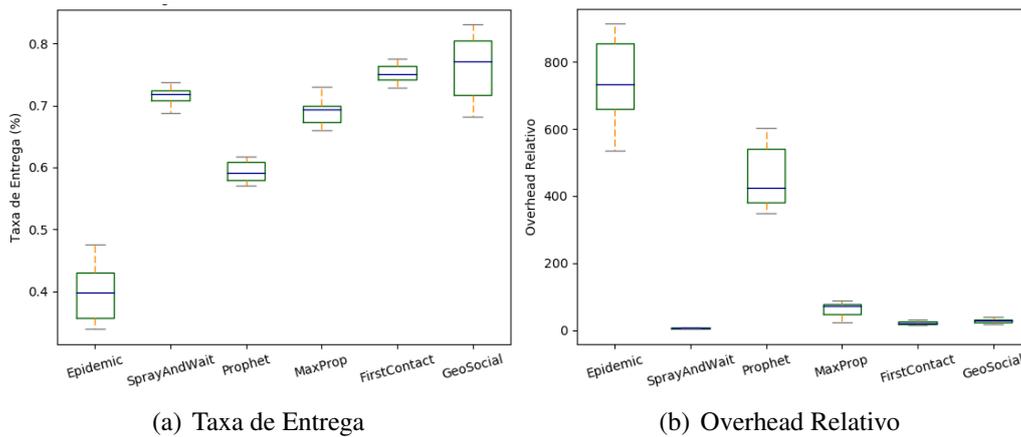
**Figura 5. Impacto do peso das características dos veículos na taxa de entrega de mensagens. (Eixo x pesos das características:  $A = (\alpha = 1; \beta = 0)$ ,  $B = (\alpha = 0, 75; \beta = 0, 25)$ ,  $C = (\alpha = 0, 5; \beta = 0, 5)$ ,  $D = (\alpha = 0, 25; \beta = 0, 75)$  e  $E = (\alpha = 0; \beta = 1)$ ), (eixo y taxa de entrega (%)).**

### 6.3. Comparação com Outros Protocolos de Roteamento

Neste teste comparamos o protocolo GeoSocial com outros cinco protocolos de roteamento: *Epidemic*, *Spray and Wait*, *Prophet*, *MaxProp* e *First Contact*. Para os protocolos *Prophet* e *Spray and Wait* foram utilizados os parâmetros de acordo com o proposto em suas publicações. Os demais protocolos não exigem nenhum parâmetro adicional. O protocolo GeoSocial tem o parâmetro  $numH = 4$  e  $\alpha = 1; \beta = 0$  para a medida de coesão laços fracos ( $Wu$ ).

A Figura 6(a) mostra o desempenho dos algoritmos em relação a taxa de entrega de mensagens. O GeoSocial manteve melhor desempenho que os demais protocolos de roteamento neste cenário. O GeoSocial não realiza inundação de mensagens na rede, realiza a entrega para nós prioritários reduzindo assim o uso de recursos. Os resultados mostram que o GeoSocial manteve em média os melhores resultados em relação aos demais protocolos como o *Epidemic* (92%), *Prophet* (28%), *Spray and Wait* (6%), *MaxProp* (10%) e *FirstContact* (1%).

A Figura 6(b) mostra o desempenho dos algoritmos em relação ao *overhead* relativo, que é a quantidade de mensagens transmitidas na rede para cada mensagem entregue ao destinatário. O GeoSocial manteve resultado superior ao *Epidemic* (2600%), *Prophet* (1600%) e *MaxProp* (200%). Isto se deve ao fato da entrega de mensagens ser feita diretamente aos nós prioritários selecionados por meio da função custo isto é, há um filtro e as entregas são mais pontuais, há menor geração de mensagens e, conseqüentemente, menor consumo de recursos. Já em relação aos demais protocolos o GeoSocial teve desempenho inferior, *Spray and Wait* (530%) e *FirstContact* (32%).



**Figura 6. Comparação da taxa de entrega e *overhead* relativo entre os protocolos de roteamento**

## 7. Conclusão

Neste trabalho propusemos uma nova e simples forma de seleção de contatos que pondera características sociais e características geoespaciais por meio de uma função custo para seleção de contatos em redes veiculares oportunistas. Também desenvolvemos um protocolo de roteamento ao qual chamamos de GeoSocial, que tem por objetivo aumentar a taxa de entrega de pacotes e diminuir o overhead de comunicação por meio da utilização de poucos nós roteadores. Identificamos que com 1% dos nós ativos em uma janela de tempo é possível atingir altas taxas de entrega e em consequência, obter baixo overhead.

Analisamos nossa proposta utilizando uma base de dados de movimentação real dos táxis da cidade de Roma, Itália. Extraímos desta base sete características, sendo seis sociais e uma geoespacial. Verificamos que os veículos podem criar vínculos sociais entre si que permitem atingir altas taxas de entrega de pacotes e overhead relativo menor que protocolos citados na literatura como o Epidemic, Prophet e MaxProp.

Também verificamos que, neste cenário, as características sociais têm desempenho superior à característica geoespacial. Acreditamos que isso se deve à concentração de veículos nas regiões centrais e à mobilidade reduzida dos veículos.

Em trabalhos futuros pretendemos testar a proposta em outras bases de dados e investigar qual caminho os veículos percorrem para nos certificarmos de que tanto as características sociais quanto as características geográficas são relevantes no roteamento das mensagens. Com isso poderemos analisar o impacto do gerenciamento de descarte de mensagens e seguir um modelo diferente do modelo de filas (FIFO). Também utilizaremos novas características geoespaciais como Centro de Massa e Centro de Gravidade.

## Referências

- Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). *Pearson correlation coefficient*. Springer.
- Borgatti, S. P. and Halgin, D. S. (2011). Analyzing affiliation networks. *The Sage handbook of social network analysis*, 1:417–433.

- Bracciale, L., Bonola, M., Loreti, P., Bianchi, G., Amici, R., and Rabuffi, A. (2014). CRAWDAD dataset roma/taxi (v. 2014-07-17). *Computer Communications*.
- Celes, C., Silva, F., Boukerche, A., Andrade, R., and Loureiro, A. (2017). Improving vanet simulation with calibrated vehicular mobility traces. *IEEE Transactions on Mobile Computing*.
- Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011). Exploring Millions of Footprints in Location Sharing Services. *Icwsn*, 2010(Cholera):81–88.
- Costa, E. C., Vieira, A. B., Wehmuth, K., Ziviani, A., and Paula, A. (2015). Time Centrality in Dynamic Complex Networks. *ArXiv*, pages 1–12.
- Cunha, F. D., Alvarenga, D. A., Viana, A. C., Mini, R. A., and Loureiro, A. A. (2015). Understanding interactions in vehicular networks through taxi mobility. In *Proceedings of the 12th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 17–24. ACM.
- Cunha, F. D., Vianna, A. C., Mini, R. A., and Loureiro, A. A. (2014). Is it possible to find social properties in vehicular networks? In *Computers and Communication (ISCC), 2014 IEEE Symposium on*, pages 1–6. IEEE.
- Granovetter, M. S. (1977). The strength of weak ties. pages 347–367.
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA.
- Hui, P., Crowcroft, J., and Yoneki, E. (2011). BUBBLE Rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11):1576–1589.
- Lance, G. N. and Williams, W. T. (1966). Computer programs for hierarchical polythetic classification (“similarity analyses”). *The Computer Journal*, 9(1):60–64.
- Masuda, N. and Lambiotte, R. (2016). *A Guidance to Temporal Networks*. World Scientific.
- Robusto, C. C. (1957). The Cosine-Haversine Formula. *Source: The American Mathematical Monthly*, 64(1):38–40.
- Scott, J. (2017). *Social network analysis*. Sage.
- Thakur, G. S., Kumar, U., Helmy, A., and Hsu, W.-J. (2010). Analysis of spatio-temporal preferences and encounter statistics for dtn performance. *arXiv preprint arXiv:1007.0960*.
- Vaz De Melo, P. O. S., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Le Mouël, F., Loureiro, A. A. F., Addepalli, L., and Guangshuo, C. (2015). RECAST: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19–36.
- Villas, L. A., Boukerche, A., Loureiro, A. A., and Ueyama, J. (2013). Um novo algoritmo geográfico ciente de partições na rede para disseminação de dados em redes veiculares. *XXXI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos-SBRC*.