



ANÁLISE DE COMPORTAMENTO DO USUÁRIO EM REDES SOCIAIS VEICULARES

Victor Ribeiro Neto

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Miguel Elias Mitre Campista

Rio de Janeiro
Junho de 2017

ANÁLISE DE COMPORTAMENTO DO USUÁRIO EM REDES SOCIAIS
VEICULARES

Victor Ribeiro Neto

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Miguel Elias Mitre Campista, D.Sc.

Prof. Igor Monteiro Moraes, D.Sc.

Prof. Luís Henrique Maciel Kosmowski Costa, Dr.

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2017

Ribeiro Neto, Victor

Análise de Comportamento do Usuário em Redes Sociais Veiculares/Victor Ribeiro Neto. – Rio de Janeiro: UFRJ/COPPE, 2017.

XII, 38 p.: il.; 29, 7cm.

Orientador: Miguel Elias Mitre Campista

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2017.

Referências Bibliográficas: p. 36 – 38.

1. Redes Sociais. 2. Redes Veiculares. 3. Comportamento do Usuário. I. Campista, Miguel Elias Mitre. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Dedico este trabalho à Rosenely
Ribeiro à qual devo tudo o que
sou hoje. À minha mãe, muito
obrigado pela dignidade e caráter
herdados, e por ter me dado
condições de chegar onde estou.*

Agradecimentos

Primeiro de tudo, gostaria de agradecer a Deus por me guiar e dar tranquilidade para seguir em frente com os meus objetivos não desanimando frente às dificuldades. Agradeço a Ele também por manter a minha família ao meu lado, com a saúde que estão até hoje.

Agradeço muito ao meu orientador Miguel Campista pela oportunidade concedida, pela orientação, empenho ímpar na ajuda da produção do artigo para o NoF 2016, motivação e conhecimento compartilhado, seja nas aulas ou na produção da dissertação, por compreender as frequentes interrupções em sua sala para solução de dúvidas e muita paciência no desenvolvimento do trabalho. Além disso, aqui deixo meu respeito e admiração pela sua disciplina e por conseguir extrair o melhor do potencial dos seus alunos.

À Dianne Scherly pela ajuda nessa primeira experiência científica também do artigo, resolvendo de problemas triviais aos mais questionadores, mostrando sempre boa vontade em solucioná-los. Agradeço muito pela sua amizade, momentos de descontração e respeito mútuo.

Ao Geraldo Ribeiro Junior, grande parceiro que mesmo distante, me influencia como ser humano e profissional no ambiente de trabalho.

Ao professor Luís Henrique pelo profissionalismo, competência e toda serenidade demonstrada.

Ao professor Otto Muniz pelos momentos de descontração no laboratório e toda a preocupação por manter o ambiente saudável.

Aos meus amigos João Batista (JB), Carlos Henrique e Martin pelas sugestões com formatação latex, opiniões técnicas e pelas palavras motivacionais nos momentos certos.

Agradeço a toda equipe do laboratório do Grupo de Teleinformática e Automação (GTA) onde trocamos experiências visando alcançar maior amadurecimento.

À Daniele da Silva pela presteza nas dúvidas nos procedimentos de inscrição em disciplinas e ao Maurício Machado pelos esclarecimentos das prestações de contas. Por último, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela bolsa de estudo concedida para o desenvolvimento da pesquisa.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE DE COMPORTAMENTO DO USUÁRIO EM REDES SOCIAIS VEICULARES

Victor Ribeiro Neto

Junho/2017

Orientador: Miguel Elias Mitre Campista

Programa: Engenharia Elétrica

A participação em redes sociais pode fornecer significativa quantidade de dados sobre usuários e o ambiente que os cerca. Quando adequadamente processados, esses dados podem ser usados como uma importante fonte de informação sobre o comportamento humano, se oferecerem informações confiáveis e de qualidade. Neste trabalho, usamos uma rede social veicular com o principal objetivo de avaliar o impacto de fatores externos sobre os usuários presentes nesses ambientes através de suas contribuições na rede. Verifica-se como a velocidade do usuário e o atraso da publicação influenciam na confiabilidade atribuída aos alertas. Estuda-se a tendência de melhora ou degradação da confiabilidade de cada usuário. Observa-se a associação entre pares de alerta que ocorrem em uma mesma rua. Verifica-se também o intervalo de contribuições consecutivas de cada usuário e a relação de intervalo de tempo entre sua primeira e a última contribuição. Os resultados foram obtidos através de um conjunto de dados público do aplicativo Waze, disponibilizado na Internet. Foi descoberto que os alertas mais postados são sobre congestionamentos, e que usuários o fazem principalmente nas horas de pico em dias úteis e durante a tarde nos fins de semana. Percebeu-se que os usuários que se movem em velocidades mais elevadas não contribuem para a rede e postagens que apresentam maiores atrasos para serem publicadas na rede são mal avaliadas. Além disso, percebeu-se também significativa associação entre alertas de risco climático e congestionamento. Como principal resultado, descobriu-se que usuários que recebem baixa confiabilidade em suas postagens tendem a manter baixa pontuação nas postagens seguintes. Por fim, foi possível notar que o intervalo entre as contribuições de cada usuário tem uma média de 10 minutos e não são feitas diariamente a rede social, mas quando o fazem, o intervalo de tempo entre os alertas possui um crescimento linear.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ANALYSIS OF USER BEHAVIOR IN VEHICULAR SOCIAL NETWORKS

Victor Ribeiro Neto

June/2017

Advisor: Miguel Elias Mitre Campista

Department: Electrical Engineering

Participation in social networks can provide a significant amount of data about users and their surroundings. When properly processed, such data can be used as an important source of information on human behavior if it provides reliable and quality information. In this work, we use a vehicular social network with the main objective of evaluating the impact of external factors on the users present in these environments through their contributions. We can see how speed of user and delay influence the reliability attributed to alerts. It also studies the tendency of improvement or degradation of the reliability attributed to the alerts of each user. It is possible to observe the association between pairs of alerts that occur on the same street in short intervals of time. We verify the interval of consecutive contributions of each user and the ratio of time interval between the first and last contribution and its total number of contributions. Results were obtained through a public dataset of the Waze application, available on the Internet. It was discovered that the most posted alerts are about congestion, and that users mostly do it during peak hours on weekdays and on weekends on the afternoon. It was found that users who move at higher speeds do not contribute to the network, and postings that present the longest delays to be published on the network are poorly evaluated. In addition, there was also a significant association between climate risk alerts and congestion. As the main result, it turned out that users who receive low reliability in their posts tend to keep score low on the following posts. Finally, it was also possible to notice that the interval between the contributions of each user has an average of 10 minutes and are not made daily to the social network, but when they do, the time interval between the alerts has a linear growth.

Sumário

Lista de Figuras	x
Lista de Tabelas	xii
1 Introdução	1
2 Revisão Bibliográfica	5
2.1 Estudo comportamental realizado com experimentos em campo	5
2.2 Estudo comportamental realizado de uma rede social	6
2.3 Estudo geral sobre gerência da informação no <i>Crowdsourcing</i>	7
3 O Waze	8
3.1 Visão geral	8
3.2 Uso da informação	9
3.3 Privacidade	10
3.4 Utilização do aplicativo	11
3.5 Comparação entre Waze e outros aplicativos	12
4 Descrição do Dataset e Metodologia	14
4.1 Descrição do dataset	14
4.2 Metodologia	15
5 Resultados e Discussões	18
5.1 Distribuição horária de alertas	18
5.2 Distribuição horária dos tipos de alertas	20
5.3 Correspondência de tipos e subtipos dos alertas	21
5.4 Distribuição geográfica dos tipos de alertas	23
5.5 Avaliação de confiabilidade atribuída aos alertas	24
5.6 Impacto da velocidade dos usuários	25
5.7 Impacto no atraso da publicação	26
5.8 Variação da confiabilidade de cada usuário	27
5.9 Associação entre pares de alerta na mesma rua	30

5.10	Frequência na postagem	31
6	Conclusões e Trabalhos Futuros	33
	Referências Bibliográficas	36

Lista de Figuras

1.1	Testes realizados para avaliar o comportamento do usuário de uma rede social veicular em função da influência de fatores externos ao sistema.	4
3.1	Captura de tela do programa Waze: Linhas vermelhas e laranjas para exibição de níveis de congestionamento e linha lilás para exibição de melhor trajeto.	9
3.2	Captura de tela do programa Waze: Os principais painéis de interação com o aplicativo.	11
3.3	Captura de tela do programa Waze: Tela para agradecimento e envio de mensagem ao alerta.	12
4.1	Captura de tela do programa Waze: Dados que apresentam velocidades irreais do Waze indicam para regiões cobertas por árvores. 16	
4.2	Captura de tela do programa Waze: Dados que apresentam velocidades irreais do Waze indicam para regiões cobertas por árvores. 17	
5.1	Número médio de alertas disparados pelos usuários por hora durante semana (de segunda a sexta). As linhas verticais pontilhadas demonstram os valores máximos e mínimos ao longo dos dias analisados. 19	
5.2	Número médio de alertas disparados pelos usuários por hora durante o final de semana (de sábado e domingo). As linhas verticais pontilhadas demonstram os valores máximos e mínimos ao longo dos dias analisados.	19
5.3	Distribuição acumulada de alertas dos usuários classificadas por tipo durante semana (de segunda a sexta).	21
5.4	Distribuição acumulada de alertas dos usuários classificadas por tipo durante o final de semana (de sábado a domingo).	21
5.5	Total de alertas categorizadas por subtipo.	23
5.6	Distribuição geográfica dos tipos de alertas durante dias úteis (de segunda a sexta).	24

5.7	Distribuição geográfica dos tipos de alertas durante o final de semana (de sábado a domingo).	24
5.8	Avaliação das notas de confiabilidade atribuída aos subtipos de alerta.	25
5.9	CDF da velocidade dos veículos quando o usuário dispara um alerta. .	25
5.10	Relação entre a velocidade dos usuários no disparo de cada alerta, número de alertas e a confiabilidade da alerta.	26
5.11	CDF do atraso dos alertas. As cores indicam as notas de confiabilidade.	27
5.12	CDF da confiabilidade dos alertas.	28
5.13	Variação absoluta de confiabilidade depois de receber a primeira classificação.	29
5.14	Variação relativa de confiabilidade depois de receber a primeira classificação.	29
5.15	CDF do intervalo de tempo entre contribuições consecutivas dos usuários.	31
5.16	Relação do intervalo de tempo entre a primeira e a última contribuição de cada usuário e seu número total de contribuições.	32

Lista de Tabelas

5.1	Número total de alertas por tipo.	20
5.2	Total de alertas categorizadas por tipo e subtipo.	22
5.3	Pares de alertas para intervalos de 5 minutos por rua	30

Capítulo 1

Introdução

O crescimento das redes sociais e o surgimento de novas tendências de mercado para personalização de produtos e serviços [1], faz com que o estudo sobre a compreensão humana se torne uma estratégia fundamental para melhoria dos lucros em diferentes setores da economia [2]. Categorizar os perfis dos usuários pelos seus ciclos sociais, preferências ou comportamentos pode trazer benefícios para tomada de decisões por parte de dois públicos. Primeiro público são as empresas que possam estar interessadas pelas informações para oferecerem produtos ou serviços aos seus clientes. Dessa forma, uma agência de turismo poderia por exemplo, oferecer pacotes de serviço para um usuário que apresente um determinado interesse ou uma loja virtual poderia oferecer produtos baseados em hábitos frequentes de um potencial cliente [1]. O segundo público interessado, é a própria rede social que pode aprender com a rotina do usuário e sugerir grupos ou informações de interesse. O desafio, porém, está na geração de informações úteis a partir da rede social, tendo em vista os milhões de usuários participantes [3].

O desafio de extrair informações úteis de grandes massas de dados pode se tornar ainda mais complexo dependendo do quão dinâmico for o ambiente onde as informações são geradas. Um exemplo típico de ambiente dinâmico são as redes veiculares [4], onde usuários conciliam as postagens nas redes sociais com a interação no trânsito. Nesses ambientes, as interações tendem a ser curtas e nem sempre precisas, já que o usuário está com a sua atenção voltada ao trânsito, o que pode prejudicar a qualidade das contribuições e dificultar ainda mais a geração de informação útil. Além disso, os participantes enfrentam problemas de conectividade intermitente e restrições de tempo em suas postagens [5]. Como consequência, postagens enviadas com certo atraso de publicação podem não ser completadas ou podem perder sua validade devido ao aspecto temporal.

As imprecisões das informações contidas nas redes sociais podem não ser apenas consequência das condições de postagem. Possíveis imprecisões podem também ser inseridas dependendo da interface com o usuário. Por exemplo, as interfaces

utilizadas poderiam permitir entradas que utilizassem a linguagem natural dos usuários, o que pode gerar imprecisões tanto acidentais quanto propositais. De maneira geral, esse mesmo problema acomete qualquer aplicativo que siga uma filosofia de cooperação entre usuários, como é o caso das aplicações que utilizam o *crowdsourcing*. O *crowdsourcing* surgiu como um método para a resolução distribuída de problemas complexos através da participação de vários indivíduos. Seu potencial tem sido expandido para áreas de estudo como transporte público inteligente [6], comunicação veicular [7], grandes massas de dados e segurança interna [8] e sistemas de recomendação para redes sociais [9]. Entretanto, o uso do *crowdsourcing* enfrenta desafios que podem comprometer a sua validade. Primeiro, a contribuição dos usuários na rede depende de estímulos ou incentivos; segundo, a contribuição dos usuários só tem valor, quando for feita com informações verídicas. Com isso, o estudo de mecanismos de incentivo a contribuições e a investigação de anomalias tornou-se um atrativo nesse segmento [8].

Na literatura, os trabalhos que analisam o desempenho das redes sociais estão muitas vezes focados na resolução de deficiências nos sistemas de recomendação, detecção de rotinas de comportamento do usuário ou semanticamente identificar o humor de usuários em uma postagem [10–12]. Em todos os casos, porém, os problemas na questão da qualidade e veracidade das informações obtidas continuam existindo. Em ambientes veiculares, os desafios mais específicos são abordados, como a proposta de detectar modelos de mobilidade ou recomendações de laços sociais [3, 13, 14]. Para conseguir isso, as informações podem ser obtidas a partir de aplicações móveis para redes sociais veiculares, na maioria das vezes com auxílio de coordenadas fornecidas pelo GPS ou roteadores distribuídos pelo acostamento das vias. Nesse sentido, o Waze é uma das aplicações mais populares para ambientes veiculares [15], necessitando do recurso de GPS e conectividade a Internet, com foco principal em troca de mensagens sobre as condições de tráfego. Posteriormente, essas informações de usuários podem ainda ser avaliadas por outros participantes, seja por comentários ou botão de aprovação. Com base na avaliação, a postagem tem uma pontuação de confiabilidade atribuída.

A análise conjunta do comportamento dos usuários e a influência de problemas de rede, como atraso de publicação normalmente não considerada, pode eventualmente revelar diferentes perspectivas relativas à confiabilidade das informações [16]. Em cenários veiculares, os desligamentos durante interações de redes sociais podem afetar a confiabilidade de alguns alertas por serem injetados tardiamente na rede social. Esses alertas podem estar suscetíveis a avaliações baixas de confiabilidade, principalmente quando o evento não existir mais no local e ainda estiver visível no aplicativo para o usuário. No entanto, dependendo do impacto sobre a rotina dos usuários e sobre a duração prevista do evento, os usuários podem se sentir mais

motivados a contribuir. Em qualquer caso, os dados obtidos podem ser compostos de informação não confiável, como consequência de condições da rede ou características do evento.

Esta dissertação tem como principal objetivo avaliar o impacto de fatores externos no comportamento dos usuários. Para conseguir isso, um conjunto de dados da rede social Waze disponível publicamente, é analisado. Esse conjunto de dados se adéqua bem aos principais propósitos deste trabalho, já que alia o comportamento dos usuários, através de suas postagens, às condições de mobilidade do veículo. Os testes realizados estão resumidos no esquema da Figura 1.1 e objetiva responder 2 principais questionamentos. Primeiro, quais fatores podem estimular uma contribuição e segundo, que elementos comprometem a qualidade da contribuição. A partir do conjunto de dados, é possível verificar as seguintes propriedades:

- a preferência dos horários de contribuições dos usuários e os principais alertas acionados;
- a avaliação da confiabilidade atribuída aos alertas;
- a influência da velocidade do veículo e como o atraso de transmissão dos dispositivos interfere na confiabilidade da informação postada;
- a tendência de melhora ou degradação da qualidade das informações de cada usuário após o disparo de seu primeiro alerta;
- a associação de ocorrência entre pares de alerta na mesma rua em curto intervalo de tempo;
- a frequência de contribuição do usuário e como se comportam os intervalos de participação na rede.

A partir da avaliação das propriedades citadas, os resultados relacionados a fatores que estimulam a contribuição mostram que o maior número de alertas é acionado no início e término das jornadas de trabalho durante os dias úteis. O número de alertas acionado também tem aumento no turno da tarde durante os finais de semana. Os resultados mostram ainda que os alertas de tráfego são os mais acionados, seguidos dos alertas de risco climático, rua fechada e, por fim, acidente. Os resultados também mostram que a velocidade dos veículos têm, de fato, influência no estímulo a contribuição de postagem e confiabilidade da informação respectivamente. É observado que o número de alertas reduz a partir de 10 km/h e torna-se desprezível acima de 30 km/h. Além disso, resultados demonstraram associação significativa entre alertas de risco climático e congestionamento em uma

mesma rua para intervalo de 5 minutos. Por último, o conjunto de dados mostra que os usuários não contribuem regularmente para a rede social mas, quando o fazem, o intervalo de tempo entre a primeira e a última contribuição cresce de forma linear em escala logarítmica.

Os resultados relacionados a elementos que comprometem a qualidade de uma contribuição mostram que os atraso na publicação da postagem possui influência sobre a confiabilidade adquirida. Os alertas com maior atraso de publicação têm menor confiabilidade e, uma vez mal avaliados, não tendem a melhorar posteriormente. Foi possível perceber também que alertas que indicam fechamento de rua possuem chances maiores de serem bem avaliados.



Figura 1.1: Testes realizados para avaliar o comportamento do usuário de uma rede social veicular em função da influência de fatores externos ao sistema.

O restante deste trabalho está estruturado da seguinte forma. No Capítulo 2 estão descritos os trabalhos relacionados. Capítulo 3 dá uma visão geral do aplicativo Waze e de suas principais diferenças em comparação com os demais aplicativos oferecidos no mercado. O Capítulo 4 descreve a metodologia e o conjunto de dados utilizados. Já o Capítulo 5 mostra os testes realizados, sendo este o objeto de estudo deste trabalho, os resultados obtidos e, finalmente, o Capítulo 6 conclui este trabalho e apresenta as direções futuras.

Capítulo 2

Revisão Bibliográfica

Este capítulo mostra esforços encontrados na literatura para descrever o estudo comportamental dos usuários em ambientes veiculares. O capítulo está dividido em três linhas de pesquisa: estudo comportamental de condutores realizados com experimentos em campo, estudo comportamental de condutores baseado em resultados fornecidos pela rede social e estudos sobre gerência da informação alimentada por *crowdsourcing*. A primeira linha de trabalho foca em modelos de comportamento dos usuários durante o processo de condução de veículos, tendo como principal objetivo modelar e aprender o comportamento do condutor ao receber diferentes tipos de informação do trânsito. A segunda linha visa compreender o comportamento de mobilidade humana em relação aos laços sociais de amizade adquiridos. Por fim, a terceira linha objetiva motivar usuários a participarem das redes sociais e quais técnicas estimulam a contribuição. Além disso, essa última linha de trabalhos busca desencorajar comportamentos maliciosos, observando o nível de degradação das informações quando tal comportamento ocorre. Sempre que possível a ideia é remover informações que são frutos de comportamentos maliciosos.

2.1 Estudo comportamental realizado com experimentos em campo

A maioria dos trabalhos encontrados na literatura, por exemplo os trabalhos de Bouhote et al. [17] e Choi et al. [18], tentam analisar o comportamento dos condutores para criar modelos que permitam classificar e identificar um comportamento de condução. Bouhote et. al *dataset* tem como objetivo principal modelar e aprender o comportamento do condutor na presença de diferentes tipos de informações de trânsito. Para isso, eles propõem uma nova abordagem formal para construir um modelo de comportamento de condução que é adaptado ao condutor individual. Para descrever o modelo, foi definido um formalismo de autômatos

que consistem em uma adaptação de conjunto de noções relacionadas ao conceito de autômatos híbridos. Um autômato híbrido é um modelo matemático para descrever precisamente sistemas onde processos computacionais digitais interagem com processos físicos analógicos. Em seguida, os autores propõem uma abordagem passiva baseada em aprendizagem para construir um modelo de acordo com o comportamento de condução observado. O modelo construído pode ser útil para prever o comportamento do condutor no futuro, evitar situações inseguras e proporcionar mais conforto ao condutor [17].

Choi et al. usa modelos de Markov ocultos (HMMs) para capturar a sequência de características de condução, adquirida das informações de um controlador do veículo. O comportamento do condutor é descrito e modelado utilizando dados do ângulo do volante, estado de freio, aceleração e velocidade do veículo. Dados obtidos e modelos foram avaliados em três tarefas distintas de classificação: 1) classificação da ação, 2) detecção da distração e 3) identificação do motorista. O objetivo da classificação de ação é categorizar os comportamentos de condução de longo prazo, como virar, mudar de faixa, parar e manter (condução neutra). O objetivo da tarefa de identificação do condutor é classificar os condutores das suas características de comportamento de condução. Ainda, a detecção de distrações identifica se o condutor está sob distração devido a tarefas secundárias ou não [18]. Outros exemplos de estudo comportamental de condutores realizados com experimentos em campo como [19] propõem uma abordagem de modelagem probabilística, onde os autores utilizam uma simples rede Bayesiana para prever o comportamento da condução em um futuro próximo.

Os trabalhos apresentados nessa seção se assemelham à proposta da dissertação pois objetivam aprender com as informações fornecidas no trânsito. O objetivo é também compreender o comportamento do usuário e fornecer maior segurança ao condutor. Entretanto, a análise comportamental desses trabalhos foca na captura da entrada dos dados observando comportamentos do condutor e do trânsito, enquanto o trabalho proposto na dissertação captura dados já processados e fornecidos pelo aplicativo Waze, ou seja, a saída.

2.2 Estudo comportamental realizado de uma rede social

Os trabalhos com dados fornecidos por redes sociais visam compreender o comportamento de mobilidade humana em relação aos laços sociais de amizade adquiridos. Embora os padrões de movimento e mobilidade humana tenham um grau de liberdade e variação, eles também exibem padrões estruturais devido a

restrições geográficas e sociais. Cho et al. [3] usam dados de localização de telefones celulares e redes sociais online para inferir o movimento humano. Foi descoberto que seres humanos experimentam uma combinação de movimento periódico que é geograficamente limitada e aparentemente seus saltos estão correlacionados com suas redes sociais. Os dados manipulados permitiram estudar os três principais aspectos da mobilidade: movimento geográfico (para onde se desloca?), dinâmica temporal (quantas vezes se desloca?) e a rede social (como relacionamentos sociais influenciam sobre o movimento?). Os autores afirmam que as viagens de curta distância e periódicas no espaço e no tempo não tem relação clara com as redes sociais, mas viagens de longa distância sim. Conclui-se que viagens de longa distância possuem maior influência dos laços de redes sociais. Backstrom et al. [14] investiga a relação entre a posição geográfica e a amizade usando endereços fornecidos pelo usuário e uma rede de associações entre os membros do Facebook.

Os trabalhos apresentados nessa seção diferem da proposta da dissertação pois o estudo comportamental foca na detecção de padrões de movimento em relação aos laços de amizade de rede social. Na dissertação, o estudo é concentrado na captura de outras informações para prever comportamento do usuário como horários, velocidades, intervalo entre contribuições, etc.

2.3 Estudo geral sobre gerência da informação no *Crowdsourcing*

No intuito de motivar os usuários a participarem das redes e desencorajar comportamentos maliciosos, Zhang et al. [20] criaram o MacroWiz, uma estrutura simples e eficaz para gerenciar as contribuições. Além disso, esse trabalho se preocupa com as respostas de baixa qualidade e como elas poderiam degradar acentuadamente a precisão dos resultados. Avançando mais além, criou-se um modelo de descarte de quaisquer informações, onde o artigo testa descarte de informações para classificação anônima de filmes.

O trabalho apresentado nessa seção, assim como proposta da dissertação, se preocupa com as respostas de baixa qualidade e como elas poderiam degradar acentuadamente a precisão dos resultados. A preocupação com o estímulo a contribuição também é algo em comum. Entretanto o trabalho relacionado fornece uma estrutura genérica para descarte de informações, sendo assim podendo não ser tão eficiente quando sua atuação se dá sobre cenários tão dinâmicos como os de uma rede social veicular.

Capítulo 3

O Waze

Entender o funcionamento do aplicativo Waze é essencial para a compreensão da análise realizada neste trabalho. Assim, este capítulo apresenta uma visão geral do aplicativo e suas principais características, como funciona, porque foi escolhido e a diferença entre ele e os aplicativos de navegação mais comuns.

3.1 Visão geral

O Waze é um aplicativo para dispositivos móveis baseado em navegação por satélite. Seu principal objetivo é o compartilhamento de dados das condições do trânsito, fornecendo informação em tempo real sobre os usuários e detalhes de trajetória, de acordo com a localização do dispositivo. O aplicativo oferece ainda, percursos planejados para definir uma hora de chegada aproximada, agendamento de eventos, velocímetro, avatar com indicações de humor, consulta de preços de gasolina e mapas interativos. Além disso, o Waze é capaz de aprender como os usuários se locomovem e utiliza as condições de trânsito reportadas pelos próprios usuários para fornecer trajetórias alternativas em tempo real [15]. Devido a essas características e às facilidades oferecidas pelo Waze, ele se tornou o aplicativo preferencial de muitos condutores, inclusive dos condutores do Uber [15].

Através do Waze, as pessoas podem relatar acidentes, congestionamentos, pontos de referência, numeração de casas etc. Alertas postados pelos usuários podem receber avaliações de agradecimento utilizando o botão de confirmação pelo usuário ou mesmo através de comentários. Com base nas avaliações recebidas, o Waze pode construir a confiabilidade de uma postagem. Além disso, o programa também atribui um valor de reputação aos usuários. À medida que o condutor contribui com o Waze e percorre limiares de quilometragem, são atribuídos pontos ao usuário para que ele possa usufruir de recursos oferecidos pela aplicativo. Um usuário iniciante do aplicativo, por exemplo, não tem permissões para modificar preços de gasolina ou atualizar características de um mapa.

O Waze também pode informar sobre congestionamentos nas vias através de ícones representativos de congestionamento ou linhas coloridas sobre as pistas. Na Figura 3.1 as linhas vermelhas indicam forte congestionamento, as laranjas indicam congestionamento moderado e fraco, enquanto que o lilás é responsável por traçar uma linha de orientação entre a origem e o destino com sugestão de melhor caminho.



Figura 3.1: Captura de tela do programa Waze: Linhas vermelhas e laranjas para exibição de níveis de congestionamento e linha lilás para exibição de melhor trajeto.

Atualmente o aplicativo investe em parcerias globais para traçar rotas de crescimento, investindo em acordos com órgãos públicos e mídias [15], entre elas emissoras de TV, estações de rádio e sites com seções dedicadas ao trânsito como The New Yorker e CBS NEWS para troca de informações locais e Financial Times para coleta de informações sobre cidades inteligentes [15]. O conjunto de dados utilizado neste trabalho é fruto da parceria da empresa Waze com o governo de Massachusetts, e está disponível publicamente na Internet [21]. O governo de Massachusetts tem se empenhado em disponibilizar gratuitamente para fins de pesquisa diversos conjuntos de dados sobre serviços gerados em seu estado [22]. São disponibilizadas informações sobre as obras de estrada em Massachusetts, além de incluir outros eventos realizados.

3.2 Uso da informação

As informações coletadas ou fornecidas pelo usuário podem ser utilizadas para vários propósitos como [15]:

- fornecer informações sobre fechamentos rodoviários planejados;
- anunciar sites, lojas e atrações em vizinhanças próximas;
- informar compromissos e reuniões importantes;

- melhorar e personalizar a experiência com os serviços, como por exemplo, usar informações de “casa” e “trabalho” ou histórico de localização para oferecer rotas com base em como foi escolhido viajar entre destinos favoritos no passado;
- melhorar a experiência de outros usuários e usar informações de rota para melhorar a precisão de dados de mapeamento e navegação;
- fornecer suporte e lidar com solicitações e reclamações;
- enviar atualizações, avisos, anúncios e informações adicionais relacionadas aos serviços pelo aplicativo;
- realizar pesquisas e questionários;
- prevenir fraudes, apropriações indevidas, infrações, roubo de identidade e outras atividades ilegais e uso indevido dos serviços.

3.3 Privacidade

O Waze oferece a opção de utilizar alguns serviços independentemente de escolher ou não configurar um nome de usuário para si, tornando-o anonimizado quando necessário. Se optar pela não configuração de nome, o programa associará todas as informações de conta do usuário a um identificador exclusivo gerado pelo aplicativo de acordo com uma política de privacidade. O usuário ainda assim conseguirá compartilhar informações de alertas, entretanto, para usufruir de determinadas funcionalidades e serviços como edição de mapa e humores, é obrigatório possuir um nome de utilizador.

Apesar de todas as possíveis informações, os usuários optam por compartilhar informações pessoais, como nome, sexo, informações de localização e relatórios de tráfego. Elas também podem ser compartilhadas através de serviços de terceiros, como outras redes sociais. O conteúdo que é compartilhado entre todos os usuários não é privado e nem confidencial. Isso quer dizer que os usuários participantes não devem ter quaisquer expectativas sobre privacidade em relação às informações que forem compartilhadas. As informações enviadas são postadas em conjunto com outras informações pessoais e eventos de alerta nas estradas relatados podem incluir o percurso de condução dentro da vizinhança. Assim, é recomendável exercer cautela e bom senso no compartilhamento de informações.

3.4 Utilização do aplicativo

A interação com o aplicativo é simples, direta e intuitiva. A ferramenta foi projetada para disponibilizar informações práticas de quem necessita dividir a atenção entre a condução de um veículo e interagir com um dispositivo móvel. O recurso por interação de voz permite com que rotas sejam solicitadas ou que informações sobre conversões e deslocamentos sejam fornecidas [15]. O painel inicial do aplicativo, pós-cadastro, apresenta-se com um mapa, velocímetro e botão para emissão de diferentes categorias de alertas. A Figura 3.2 exibe os três principais painéis de interação com o aplicativo.



Figura 3.2: Captura de tela do programa Waze: Os principais painéis de interação com o aplicativo.

Todo alerta acionado possui uma nota de confiabilidade atribuída. Devido à escassez de documentação do aplicativo, fraca documentação do *dataset* e ao fato do Waze ser de arquitetura fechada, não é possível garantir a origem da atribuição dessas notas. Assume-se que atualmente é possível avaliar a confiabilidade do alerta de duas formas. A Figura 3.3 extraída do aplicativo exibe os dois botões responsáveis por essa tarefa: botão de agradecimento de alerta e mensagem. Uma certa quantidade de aprovações acionadas seria um dado necessário para julgar a confiabilidade ou não de uma alerta. No caso das mensagens, é necessário mecanismos que interpretassem as palavras utilizadas nas orações para julgar o sentimento emitido [23].

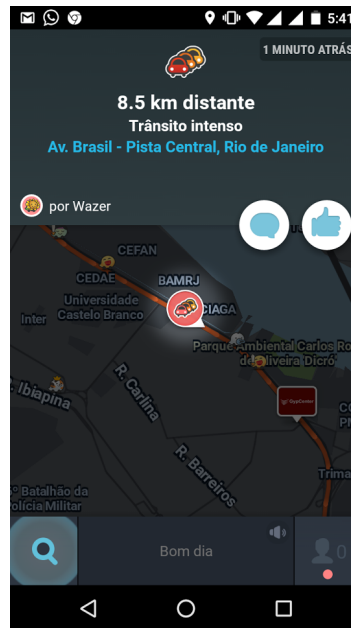


Figura 3.3: Captura de tela do programa Waze: Tela para agradecimento e envio de mensagem ao alerta.

3.5 Comparação entre Waze e outros aplicativos

O programa Waze difere de aplicativos navegadores de GPS tradicionais devido principalmente às suas funções relacionadas com a rede social, como o círculo de amigos e troca de mensagens. Atualmente, o aplicativo usufrui da API (*Application Programming Interface*) disponibilizada pela rede social Facebook para facilitar o cadastramento de novos usuários, obtendo dados públicos para evitar a burocratização de preenchimento de formulários. É possível também sincronizar seus contatos de amizade do Facebook com o programa.

Quando observa-se as semelhanças a outros aplicativos, assim como o Google Maps, o Waze também disponibiliza sua API publicamente para que outros programas utilizem seu mapa para geolocalização. Entretanto, alguns aplicativos apresentam recursos que o Waze ainda não apresenta. O Google Maps permite que o usuário possa traçar trajetos considerando outro meio de transporte que não seja por veículo, como os de deslocamento a pé [24]. O mesmo Google Maps também possui integração com Google Street View, serviço que fornece informações detalhadas sobre imagens das vias. Outro recurso ainda não tão eficiente apresentado pelo Waze, são as visualizações de mapas *offline* como o aplicativo Here Maps [25], que possibilita um flexibilidade no *download* considerando as divisas e mapas políticos. O usuário do Here Maps consegue definir um *download* de continente, país, estado ou cidade. No Waze para baixar um mapa, é necessário definir uma rota e no Google Maps é preciso posicionar a área desejada dentro do mapa, usando zoom para aumentar ou diminuir a abrangência [24]. Isso impede, que o usuário faça o download somente

de uma cidade ou estado. Porém, dá flexibilidade para que dispositivos com pouca memória baixem apenas um bairro ou parte dele. Atualmente, a maior semelhança do Waze com seus concorrentes é que todos fornecem também seus serviços por acesso Web.

Não foram encontrados *traces* disponíveis para estudo sobre os demais aplicativos. Esses aplicativos assim como o Waze também possuem arquiteturas fechadas, o que dificulta uma investigação mais precisa. A escolha pelo Waze como aplicativo para objeto de estudo se deu primeiramente pela disponibilidade do material encontrado e devido a relevância que o programa possui nos dias atuais para condutores de trânsito. Verificados os demais aplicativos, o único que apresenta elementos de rede social é o Waze, que permite um maior aprofundamento para estudos comportamentais.

Capítulo 4

Descrição do Dataset e Metodologia

Este capítulo descreve o conjunto de dados utilizado e a metodologia empregada para a obtenção dos resultados. Tais resultados visam compreender o comportamento dos usuários em redes sociais móveis.

4.1 Descrição do dataset

Apenas alguns conjuntos de dados de redes sociais veiculares, chamados aqui por simplicidade de *datasets*, estão disponíveis ao público. Devido às características de privacidade dessas redes, tornar públicas algumas informações viola os direitos dos usuários estabelecidos nas políticas de privacidade no ato do cadastro. As parcerias realizadas pelo Waze com o governo de Massachusetts propiciaram o fornecimento desse material para fins de pesquisa. Até onde se sabe, o único *dataset* público do Waze é o fornecido pelo governo de Massachusetts [26], nos Estados Unidos. O *dataset* oferece informações anonimizadas, não sendo possível identificar informações pessoais dos usuários envolvidos, mas somente diferenciá-los. Esse *dataset* é composto pela junção de três tabelas e suas principais informações estão divididas em dados de alerta, dados da pista, além de uma tabela de junção com as principais informações anteriores. As tabelas apresentam informações de eventos ocorridos em um intervalo de 8 dias, entre 22 de fevereiro e 1º de março de 2015, totalizando 100 MB os três arquivos juntos. Desses 8 dias, 1 foi descartado pois não foi disponibilizado informação correspondente a 24 horas, então foram aproveitados somente 7 dias para a realização do estudo.

A tabela Waze Alert Data é composta por atributos pertinentes a informações do alerta. Os atributos mais relevantes desse arquivo são: tipo e subtipo da alerta, data e hora de inserção da informação, nome e tipo da rua, cidade, identificação do usuário (ID), coordenadas geográficas do alerta e confiabilidade. É importante destacar que o Waze classifica os alertas em quatro principais tipos (Congestionamento, Risco climático, Rua fechada e Acidentes), e tais alertas são subdivididos em 24 subtipos

(que será informado no capítulo de resultado posteriormente em tabela 5.2). Além disso, o Waze atribui confiabilidade a uma postagem de alerta que varia de 5 como menor confiabilidade, a 10 como a mais elevada. A confiabilidade do usuário não pode ser observada pois não constavam informações no *dataset*. O usuário do aplicativo alcança reputações de acordo com suas participações na rede. A partir de um limiar de quilometragem percorrido pelo usuário, recursos do programa como alteração de preços de gasolina ou modificações no mapa tornam-se disponíveis para usuários de reputação mais alta. A tabela Waze Jam Data é composta por informações fornecidas pelas condições da via no momento que o alerta foi acionado. Os atributos mais relevantes são: velocidade do condutor, atraso de publicação e também informações já detectadas na tabela anterior como data e hora de inserção da informação, cidade, identificação do usuário - ID (anonimizado) e coordenada geográfica. A tabela Waze Point Data visa reunir informação de junção entre as duas tabelas anteriores como a identificação do usuário.

A documentação do *dataset* não descreve com precisão toda a informação existente e é necessário deduzir o significado de algumas colunas com base na semântica esperada. Assume-se que a coluna *startTime* é o momento em que um usuário interage com o aplicativo, ou seja, quando aciona o botão de alerta. A coluna *InjectDate* contém o momento em que a rede social publica o alerta das informações recebidas e finalmente, o *Reliability* é a avaliação de confiança recebida por cada postagem. A velocidade do usuário no momento do alerta não informa explicitamente a unidade, com isso, assumiu-se que foi representada em metros por segundo, devido a compatível velocidade permitida na áreas de Massachusetts. Nós também ajustamos o fuso horário do conjunto de dados para representar o fuso horário do Massachusetts para fevereiro de 2015. O conjunto de dados foi higienizado, removendo os alertas produzidos por veículos em velocidades consideradas irreais. O processo de higienização resultou na remoção de 0,016% dos dados existentes. Investigando estes *outliers*, descobriu-se que os IDs dos usuários que produzem registros com velocidades inviáveis são diferentes. Portanto, foi possível concluir que isso é uma consequência de problemas técnicos, provavelmente fornecido pelo GPS para a rede social. Em 4.2 serão mostradas possíveis razões para os problemas fornecidos pelo GPS.

4.2 Metodologia

O método proposto para atingir nosso resultado foi dividido em 2 partes:

- Higienização e conversão de valores

O processo de higienização do *dataset* capturado trouxe desafios para que se

chegasse nas condições ideais para manipular os dados. As informações referentes a data e tempo são exibidas no navegador de acordo com o fuso horário local quando carregadas. Entretanto, quando os arquivos do *dataset* são descarregados no disco, as informações de data e hora são disponibilizadas em fuso horário alternativo e um processo de atualização dessas informações teve de ser feito. A coluna *injectDate* apresenta uma estrutura de valor de data e hora semelhante a “2015-02-23 6:10PM” quando visualizada em navegador mas apresenta a estrutura “2015-02-23T16:10:00.000Z” quando descarregada no disco. A letra “Z” no final da estrutura significa que a informação de data e hora está configurada em um padrão chamado Zulu Time Zone, o fuso horário militar. O padrão Zulu Time Zone tem a mesma correspondência horária do $UTC+0$, um fuso horário utilizado por uma lista de países como Gana, Sibéria ou Islândia, mas o Estados Unidos não encontra-se nessa lista. Entre Janeiro e Fevereiro de 2015, Boston possuía o fuso horário EST e ao utilizar o conversor de horas entre $UTC+0$ e EST, descobriu-se um intervalo de 5 horas de diferença. Isso significa que os horários obtidos antes da conversão se mostravam 5 horas adiantado do horário real do cenário.

Foi necessário também descobrir a unidade de medida da velocidade utilizada pelo *dataset* para que pudesse adaptá-la à realidade. Baseado na investigação das velocidades das ruas de Massachusetts, descobriu-se que o valor apresentado estava em metros por segundo. Sendo assim, converteu-se para o padrão internacional quilômetros por hora. Nesta etapa do trabalho, percebeu-se que algumas

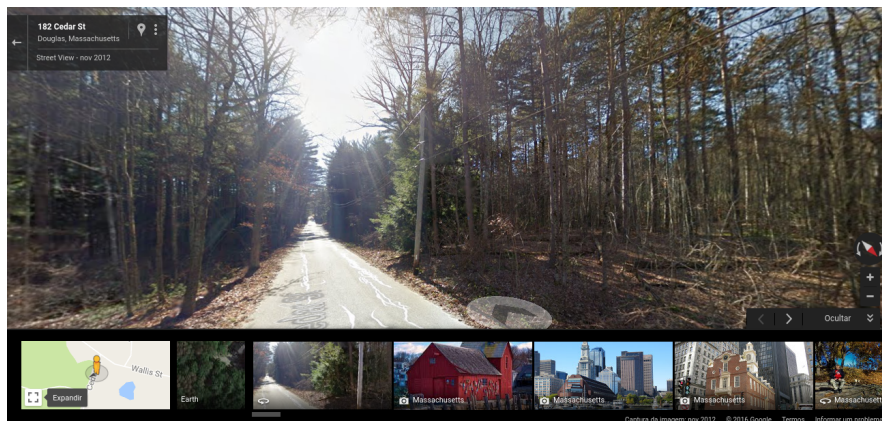


Figura 4.1: Captura de tela do programa Waze: Dados que apresentam velocidades irreais do Waze indicam para regiões cobertas por árvores.

postagens de alerta apresentavam velocidades irreais, superiores a 1.000 km/h. Essas postagens foram investigadas separadamente para que fosse possível identificar alguma anomalia que as explicassem. É possível encurtar o tempo de uma viagem e atingir a quantidade de pontos de reputação necessários mais rapidamente através da realização de viagens fictícias de longas distâncias, utilizando falsos localizadores [27]. Nessa direção, a posição geográfica dos veículos onde a postagem



Figura 4.2: Captura de tela do programa Waze: Dados que apresentam velocidades irreais do Waze indicam para regiões cobertas por árvores.

era originada foi observada. Descobriu-se que as informações de coordenada apontavam para regiões onde havia significativa cobertura de árvores e isso poderia influenciar no erro de transmissão das informações de GPS para o dispositivo, como pode ser visto nas Figuras 4.1 e 4.2. De acordo com Hedgecock et al. [28], receptores podem exibir até dezenas de metros de erro, o que pode ser uma explicação para essa discrepância encontrada nas velocidades. As postagens realizadas com velocidades muito além da máxima não foram consideradas nas análises deste trabalho.

- Elaboração de *scripts* para extração de resultados

Para obtermos os resultados propostos nesse estudo foi elaborado *scripts awk* e consultas na linguagem *SQL*.

Capítulo 5

Resultados e Discussões

A análise do conjunto de dados do Waze é dividida em dez partes. Em primeiro lugar, verifica-se os fatores externos que estimulam a contribuição de alerta pelos usuários, como os horários preferenciais de eventos representativos. Além disso, observa-se também se essas quantidades de alertas variam ao comparar dias úteis e finais de semana. Em segundo, verifica-se também os tipos de alerta acionados com mais frequência pelos utilizadores ao longo de um dia e, logo em terceiro, avalia-se a quantidade de alertas por subtipos, ordenando-os de maneira crescente em função do número total de alertas. Em quarto, verifica-se a distribuição geográfica dos tipos de alertas. Em quinto, a classificação de confiabilidade atribuída aos alertas. Em sexto, verifica-se ainda o interesse em contribuir dos usuários levando em consideração as suas velocidades e como as características de velocidade influenciam sobre a nota de confiabilidade de alerta adquirido. Em sétimo, observa-se a relação de atraso de transmissão dos alertas e a nota de confiabilidade do alerta. Em oitavo, investiga-se a tendência de melhoria ou deterioração das notas de confiabilidade atribuída aos alertas dos usuários. O objetivo é observar se a primeira nota de avaliação da postagem sofre modificações de valor posteriormente. Nono, verifica-se a associação entre pares de alertas que ocorrem na mesma rua em um intervalo de cinco minutos. Décimo, analisa-se o intervalo entre contribuições consecutivas de cada usuário e se o número de contribuições de alerta de um dado usuário possui relação com o intervalo de tempo entre cada um dos alertas. Cada um desses testes acrescenta conhecimento sobre o comportamento dos usuários do Waze.

5.1 Distribuição horária de alertas

Nessa primeira análise, investiga-se o número médio de alertas disparado por usuários por hora durante os dias da semana. Verifica-se primeiramente os dias úteis e em seguida não úteis. As Figuras 5.1 e 5.2 mostram, além da média, linhas verticais pontilhadas com valores máximos e mínimos da mesma hora considerando

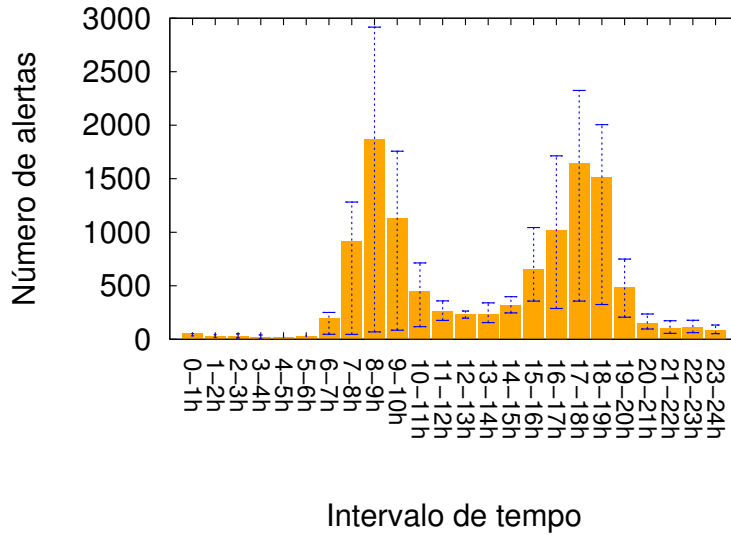


Figura 5.1: Número médio de alertas disparados pelos usuários por hora durante semana (de segunda a sexta). As linhas verticais pontilhadas demonstram os valores máximos e mínimos ao longo dos dias analisados.

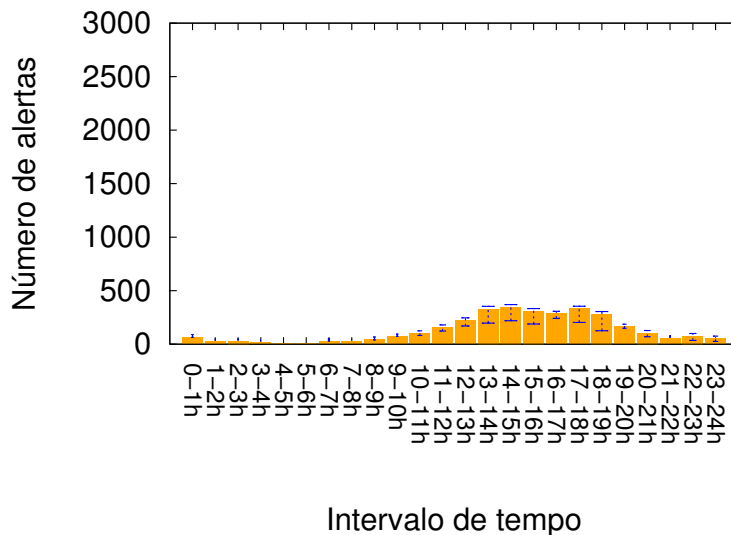


Figura 5.2: Número médio de alertas disparados pelos usuários por hora durante o final de semana (de sábado e domingo). As linhas verticais pontilhadas demonstram os valores máximos e mínimos ao longo dos dias analisados.

todos os dias. Na Figura 5.1, nota-se picos de alertas ao redor entre 8-9h e 17-19h, demonstrando que os usuários contribuem com a rede social veicular principalmente durante os horários mais comuns de início e término de expediente (horário do *rush*). O número de alertas é alto devido aos horários de almoço e fim do expediente pois um maior fluxo de veículos encontram-se em trânsito. Na Figura 5.2, por outro lado, a mesma tendência não existe e o número de alertas se concentra sobretudo na parte da tarde, sendo que o aumento ocorre às 9h e se mantém estável até às 20h. Isso mostra um comportamento completamente diferente dos usuários em comparação

com os dias de semana. Outro detalhe é que a tendência por contribuição nos finais de semana é bem inferior à contribuição durante a semana.

Justamente pelas Figuras 5.1 e 5.2 possuem o eixo Y na mesma escala, é possível notar que, em números absolutos, o número de alertas é bem menor na média no final de semana (Figura 5.2) que durante a semana. Por fim, é possível perceber que o intervalo entre a média, o máximo e o mínimo de ambos os gráficos demonstram tendências distintas. A média de alertas do final de semana tende a estar mais próxima do valor máximo de cada horário, enquanto que durante semana esse comportamento não acontece. Isso prova que a variação das amostras de alertas de cada horário durante a semana é maior. Tentativas de representação gráfica dos alertas com intervalo de confiança não foram bem sucedidas como consequência da dispersão e do número de dias do conjunto de dados.

5.2 Distribuição horária dos tipos de alertas

Tabela 5.1: Número total de alertas por tipo.

Tipo	Total
Acidente	1277
Congestionamento	50895
Rua fechada	3320
Risco climático	16013

O número de alertas é ainda investigado sob outro ponto de vista, a ideia é verificar a quantidade de alertas, separando-os conforme o tipo e o subtipo. A Tabela 5.1 mostra o número acumulado total de alertas separados por tipo, considerando todas as mensagens válidas no conjunto de dados do Waze. O tipo mais frequente de alerta é “Congestionamento”, que ocorre três vezes mais do que o segundo mais disparado, “Risco climático”. Isso indica que os usuários estão mais motivados a contribuir em caso de problemas de tráfego do que em caso de problemas climáticos, além de situações de congestionamento possibilitarem maior facilidade de interação com um dispositivo móvel.

As Figuras 5.3 e 5.4 mostram a distribuição acumulada horária dos alertas agrupados por tipo, durante a semana e durante o final de semana respectivamente. Na Figura 5.3, nota-se que os alertas de congestionamento são predominantes sobre os outros tipos, principalmente durante as horas de pico. Na Figura 5.4, nos finais de semana, apesar dos alertas de congestionamento ainda serem predominantes, a diferença em relação aos outros tipos não é mais tão alta. Os alertas de “Rua fechada” demonstraram distribuição regular durante os horários do dia, tanto durante semana quanto no final de semana. Já o tipo “Acidente”, apesar de ser proporcionalmente

menor, ocorre mais frequentemente durante o dia. Os números de alertas de acidentes de madrugada não foram significativos. Em resumo, a partir da análise desta seção, é possível reforçar que os usuários possuem horários de maior estímulo à contribuição.

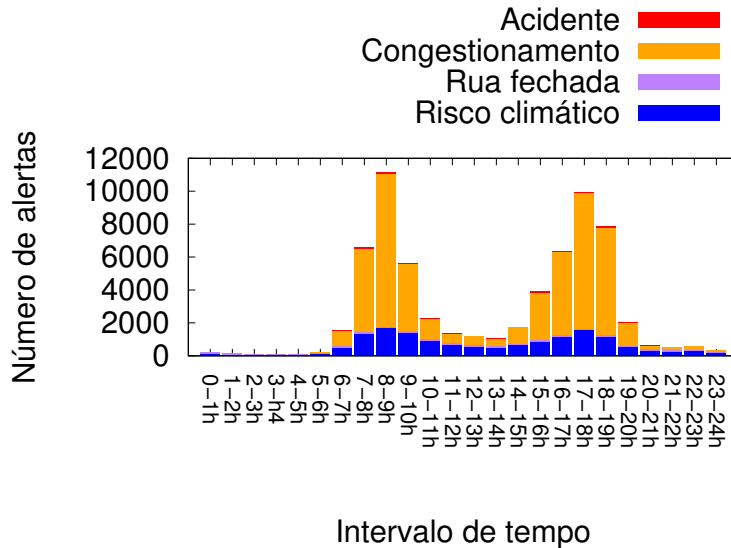


Figura 5.3: Distribuição acumulada de alertas dos usuários classificadas por tipo durante semana (de segunda a sexta).

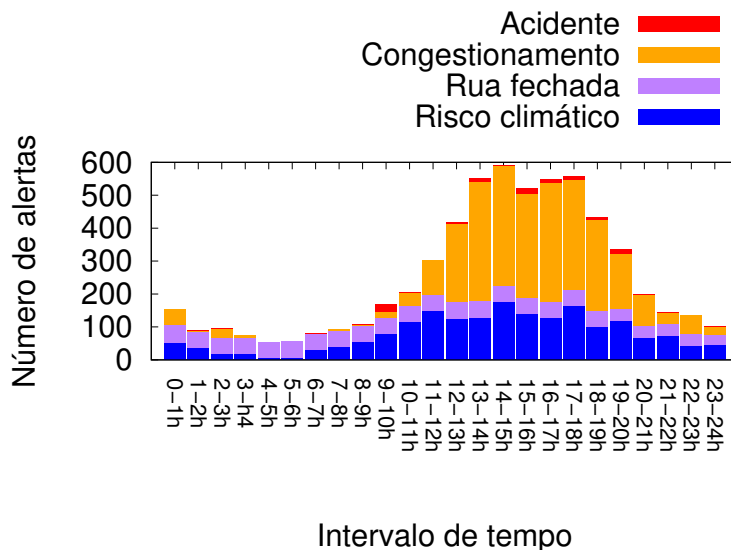


Figura 5.4: Distribuição acumulada de alertas dos usuários classificadas por tipo durante o final de semana (de sábado a domingo).

5.3 Correspondência de tipos e subtipos dos alertas

A Tabela 5.2 permite complementarmente verificar a distribuição total de alertas separados por subtipos e a partir dos resultados é possível observar que os usuários

postam mais determinados subtipos que outros. Quando o usuário posta que o trânsito encontra-se congestionado, o grau de detalhamento dessa informação é pequena. Em função disso, os usuários podem ainda escolher se vão apenas assinalar o congestionamento ou se vão oferecer mais informações sobre o quão congestionado está o trânsito. As três principais opções são os subtipos “Trânsito muito pesado”, “Trânsito parado” ou “Trânsito moderado”. O “Trânsito moderado” exige do condutor um grau de julgamento mais apurado ou não representa um nível de congestionamento suficiente para que os usuários se sintam interessados a contribuir. Para “Rua fechada”, o evento mais acionado foi “Evento com rua fechada”. Porém para “Risco climático”, em função das muitas opções distintas disponíveis, o condutor se concentra em quatro principais, sendo elas “Perigo de buraco na pista”, “Perigo de carro parado ao redor”, “Perigo de carro parado” e “Perigo de construção na rua”.

Tabela 5.2: Total de alertas categorizadas por tipo e subtipo.

Tipo	Subtipo	Total
Acidente	Nenhum subtipo	186
	Acidente grave	246
	Acidente leve	845
Congestionamento	Nenhum subtipo	2566
	Trânsito moderado	6160
	Trânsito parado	17659
	Trânsito muito pesado	24510
Rua fechada	Rua fechada com construção	9
	Perigo com rua fechada	13
	Nenhum subtipo	14
	Evento com rua fechada	3284
Risco climático	Risco climático á vista	1
	Risco climático com chuva de granizo	4
	Risco climático com nuvens	5
	Risco de animais ao redor	9
	Risco de morte na pista	11
	Perigo ao redor por perda de sinal	13
	Risco climático	13
	Perigo ao redor	29
	Perigo na pista	38
	Perigo de gelo na pista	131
	Nenhum subtipo	351
	Perigo com objeto na pista	407
	Perigo de construção na rua	1422
	Perigo de carro parado	2217
Perigo de carro parado ao redor	2871	
Perigo de buraco na pista	8491	

Olhando mais de perto os subtipos, é possível também confirmar que os alertas mais acionados são os de congestionamento, sendo os subtipos “Trânsito muito pesado” e “Trânsito parado” os mais usados. Somados, esses dois subtipos atingem mais de 40.000 alertas. Nota-se também através da Tabela 5.2 a preferência que alertas de “Congestionamento” e “Risco climático” possuem sobre “Acidentes” e “Ruas fechadas”.

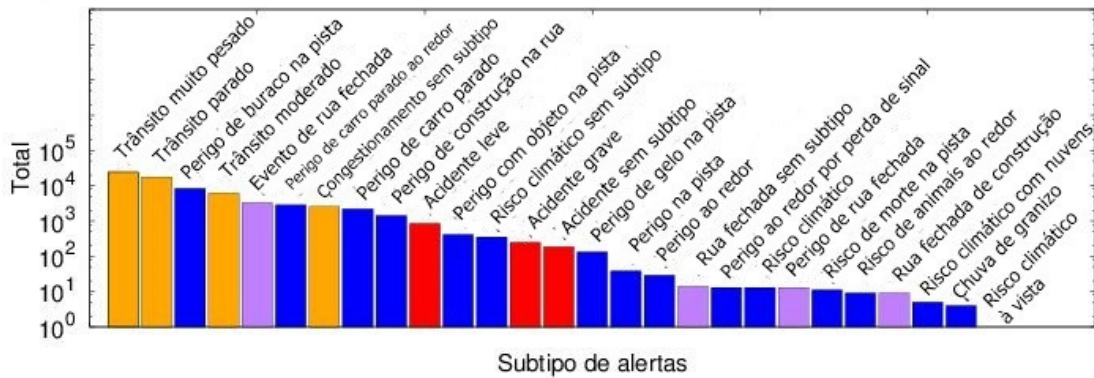


Figura 5.5: Total de alertas categorizadas por subtipo.

A Figura 5.5 por outra visão, permite analisar os subtipos de alerta e seu total de ocorrência ordenando de forma decrescente. É possível notar de forma mais clara o ranqueamento dos subtipos e perceber que eventos relacionados a congestionamento estão entre os mais usados.

5.4 Distribuição geográfica dos tipos de alertas

As Figuras 5.6 e 5.7 exibem a distribuição geográfica dos tipos de alertas de acordo com suas posições de latitude e longitude no mapa de Massachusetts onde o eixo X representa a latitude e o eixo Y a longitude. O *dataset* fornece as informações de coordenadas na unidade de grau decimal. As cores representam cada um dos tipos de alerta. É possível perceber na Figura 5.6 através da distribuição de alertas e posições geográficas que eventos como “Congestionamento” são predominantes sobre os demais como já observado em resultados anteriores. Nota-se também que alertas de “Acidente” são pouco numerosos e visíveis no gráfico, além de se distribuírem de forma aleatória comparados a “Congestionamento” e “Risco climático”. Por outro lado na Figura 5.7, fica visível uma menor distribuição de alertas e um maior equilíbrio entre os alertas de “Congestionamento” e “Risco climático”.

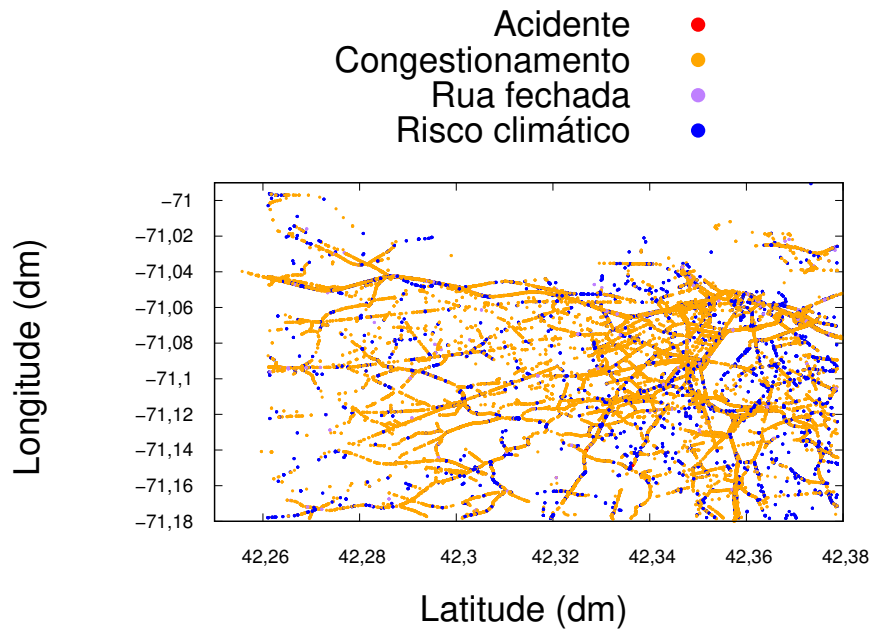


Figura 5.6: Distribuição geográfica dos tipos de alertas durante dias úteis (de segunda a sexta).

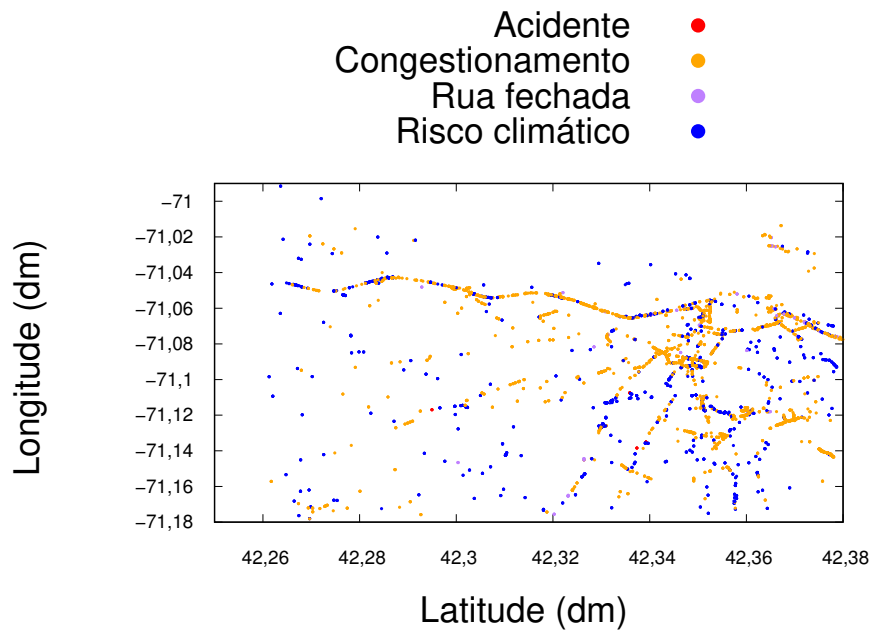


Figura 5.7: Distribuição geográfica dos tipos de alertas durante o final de semana (de sábado a domingo).

5.5 Avaliação de confiabilidade atribuída aos alertas

Nesta seção, a intenção é avaliar as notas de confiabilidade que foram atribuídas aos subtipos de alertas. O objetivo é verificar quais subtipos de alerta recebem as maiores notas. A Figura 5.8 apresenta a porcentagem das notas de confiabilidade atribuídas a cada subtipo. Ao observar os subtipos que se destacam, é possível perceber que o subtipo “Risco de animais ao redor” quando disparado, possui mais

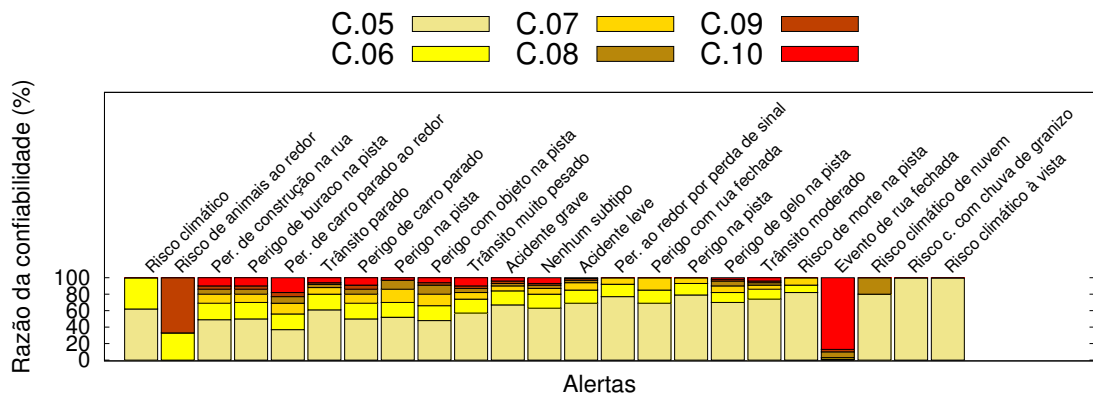


Figura 5.8: Avaliação das notas de confiabilidade atribuída aos subtipos de alerta.

de 60% de chance de receber nota 9 de confiabilidade. Além disso, o subtipo “Evento de rua fechada” apresentou mais de 80% de seus alertas com nota 10 de confiabilidade. Por outro lado, o subtipo “Perigo de gelo na pista” apresentou mais de 70% de seus alertas com nota 5, demonstrando ser um dos subtipos menos confiável. Nota-se que os subtipos de alerta “Risco de animais ao redor” e “Evento de rua fechada” demonstram significativa probabilidade de serem bem avaliados no quesito confiabilidade comparado aos demais alertas. É possível perceber também que o usuário avaliador tem uma considerável certeza ao confirmar esse tipo de evento quando o encontra.

5.6 Impacto da velocidade dos usuários

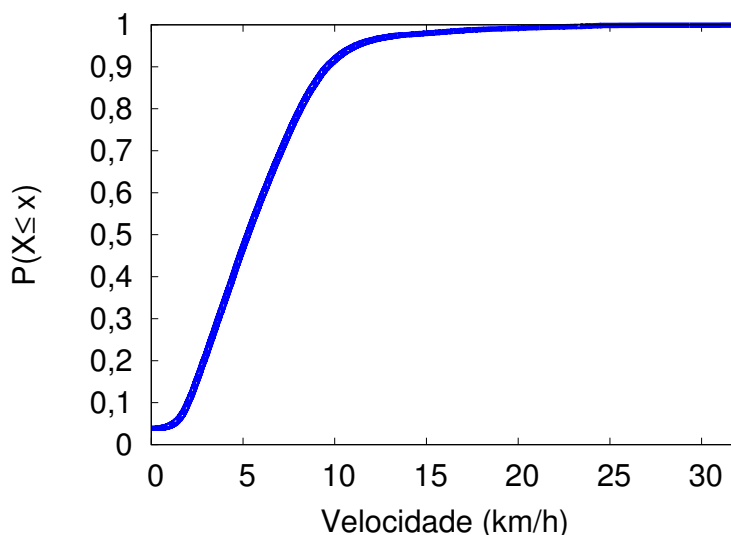


Figura 5.9: CDF da velocidade dos veículos quando o usuário dispara um alerta.

Nesta seção, a ideia é investigar a influência da velocidade do veículo sobre o número de contribuições de alerta e a confiabilidade correspondente das postagens.

A Figura 5.9 mostra a CDF(Função de distribuição acumulada) das velocidades mais comuns utilizadas pelos usuários ao contribuírem com a rede social veicular. Nota-se que 90% das postagem acontecem com velocidades abaixo de 10 km/h. Isso demonstra que os usuários estão mais propensos a contribuírem com a rede em velocidades mais baixas.

A Figura 5.10 apresenta em maiores detalhes a distribuição de velocidades às quais os usuários se deslocavam ao contribuir com a rede social. A Figura 5.10 mostra em números absolutos o número de postagens, separando a quantidade de postagens por valor de confiabilidade. Esse resultado confirma que os usuários interagem com a rede social a baixas velocidades. Ainda, é possível observar que o número de contribuições aumenta até 10 km/h e, depois disso, tende a diminuir rapidamente. Esse resultado corrobora a análise realizada anteriormente (Seção 5.2), onde observou-se que os usuários são mais propensos a contribuir quando o tráfego está mais congestionado. Além disso, pode-se esperar que os usuários tendem a interagir menos com a rede social em velocidades mais altas, pois precisam prestar mais atenção à pista. Nota-se também que a confiabilidade dos alertas postados é geralmente baixa e, acima de 30 km/h, o número de alertas se torna insignificante. Mesmo assim, é possível perceber que as poucas contribuições acima de 30 km/h não possuem alta confiabilidade.

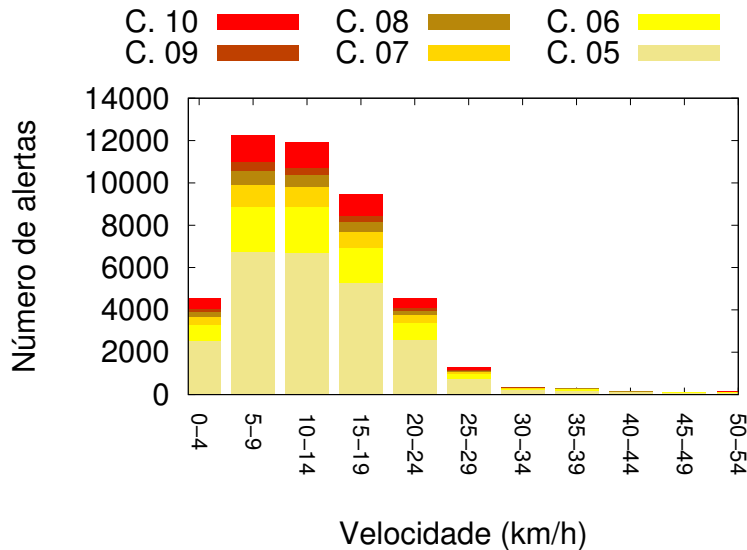


Figura 5.10: Relação entre a velocidade dos usuários no disparo de cada alerta, número de alertas e a confiabilidade da alerta.

5.7 Impacto no atraso da publicação

O atraso para postagem em uma rede social busca indicar o tempo necessário para que um bit de dados seja encaminhado do usuário até o servidor. Considerando

o dinamismo, densidade, problemas de conectividade intermitente e restrições de tempo em postagens de uma rede social veicular, avalia-se nesta seção a relação entre o atraso e a confiabilidade dos alertas correspondentes. A Figura 5.11 mostra a CDF da distribuição de atraso das contribuições. O resultado permite observar que as contribuições com maior atraso têm menor confiabilidade. A lógica por trás disso é que os alertas que passam mais tempo para serem publicados na rede social podem perder valor devido a aspectos temporais, como o exemplo de engarrafamento que não existe mais após um intervalo de tempo maior, ou uma rua que já não se encontra mais fechada. Em oposição a esse fato, podemos notar que as contribuições com maior confiabilidade são aquelas com atrasos inferiores e que são rapidamente publicadas.

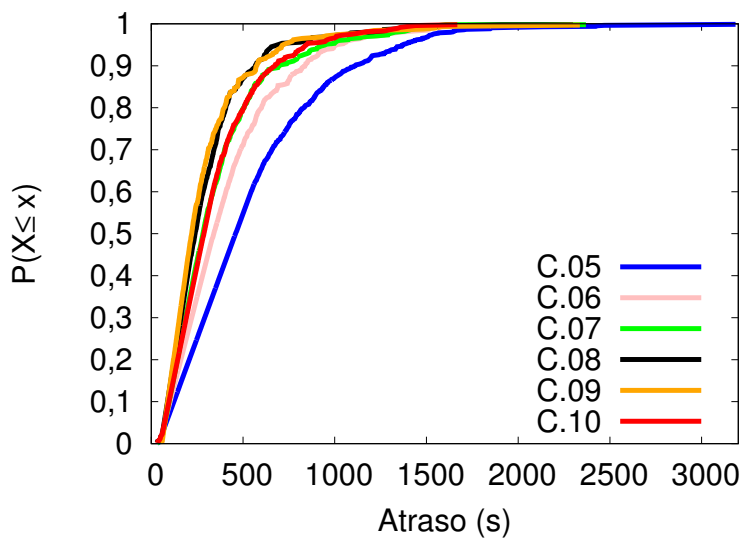


Figura 5.11: CDF do atraso dos alertas. As cores indicam as notas de confiabilidade.

5.8 Variação da confiabilidade de cada usuário

A possibilidade de interação entre usuários na rede social Waze ainda é limitada. Trocas de mensagens, buzinas, alertas ou a confirmação de agradecimento de alerta são as possibilidades de comunicação entre os participantes. Para cada alerta postado pelo usuário, há uma confiabilidade associada ao evento e, desse resultado, é possível analisar se há uma tendência de melhora ou degradação da nota de confiabilidade dos participantes, após a primeira postagem. O objetivo principal é averiguar se é possível perceber com facilidade usuários que não demonstram sinais de melhora após apresentar notas baixas, podem não estar com boas intenções ao inserir informações no Waze. Se as notas de confiabilidade dos usuários melhorarem para os alertas subsequentes, isso pode significar que o usuário simplesmente

disparou alertas sem comprometimento de sua veracidade, simplesmente por testar a funcionalidade do aplicativo.

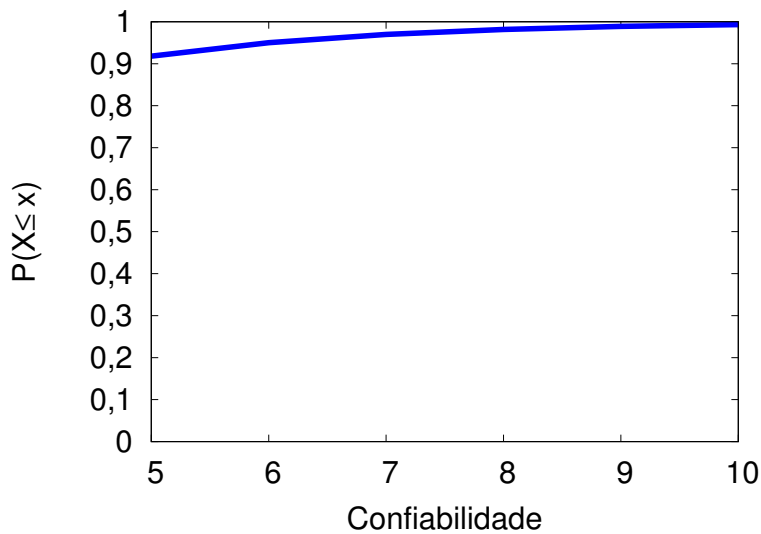


Figura 5.12: CDF da confiabilidade dos alertas.

Primeiramente, a CDF indicada na Figura 5.12 demonstra a variação das notas de confiabilidade no conjunto de dados analisado. É possível perceber que aproximadamente 90% das notas de confiabilidade encontram-se iguais a cinco, que é a nota mais baixa aplicada pelo Waze. As Figuras 5.13 e 5.14 mostram a tendência de melhoria ou degradação dos índices de confiabilidade que um usuário recebe após sua primeira contribuição. Nessas figuras, o eixo X representa a primeira avaliação dada a uma postagem (R_I), e o eixo Y denota a diferença entre a confiabilidade das avaliações seguintes em comparação com a primeira ($\Delta r_i = r - r_i$). Na Figura 5.13, a cor denota o número absoluto de avaliações com os mesmos Δr considerando todos os possíveis r_0 . Observa-se que a confiabilidade 5 é predominante e mantém-se inalterada nas avaliações seguintes, já que a cor avermelhada se sobressai em $(x, y) = (5, 0)$. Dessa forma, o Waze demonstra maior probabilidade de receber avaliações com pontuação mínima, e que na maioria das vezes não é alterada. Quando a avaliação muda, a variação de confiabilidade é na maioria das vezes pequena como a variação de 5 a 6, por exemplo.

A segunda pontuação de confiabilidade mais utilizada é a máxima, com valor 10. Nesse caso, observa-se que as mensagens bem avaliadas também permanecem sendo consideradas confiáveis. Já as outras avaliações de confiabilidade, as que geram pontuações entre 6 e 9, são pouco utilizadas e apresentam comportamento variado, pois enquanto as pontuações 6 e 7 tendem a se manter, as 8 e 9 demonstram uma tendência de aumento. Isso indica um comportamento bimodal do usuário à medida que os usuários avaliam as postagens de tal maneira que o Waze considere a contribuição com confiança 10 ou 5 na maioria das vezes. Olhando a variação

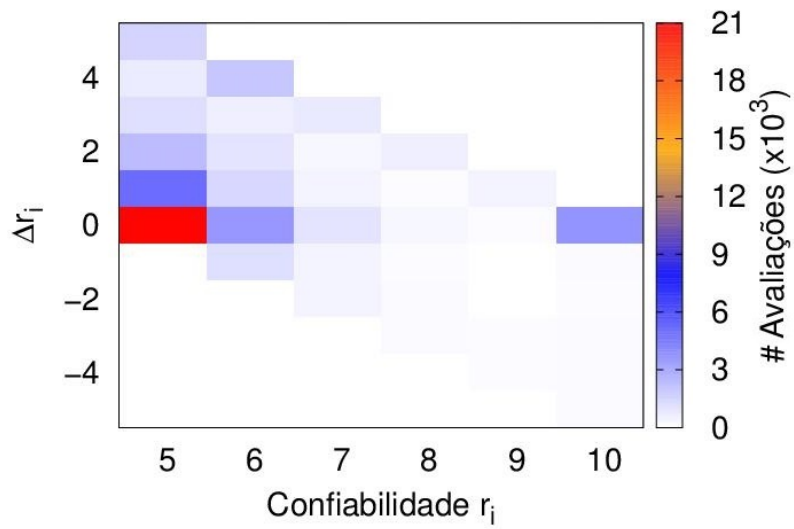


Figura 5.13: Variação absoluta de confiabilidade depois de receber a primeira classificação.

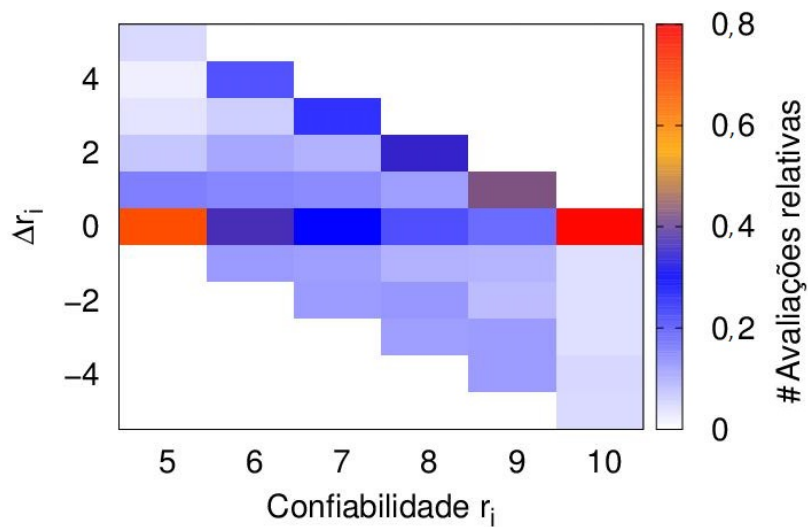


Figura 5.14: Variação relativa de confiabilidade depois de receber a primeira classificação.

da avaliação, nota-se que há uma maior tendência de melhora do que deterioração, como mostrado pela maior concentração de cores azuladas acima de $\Delta r = 0$.

A Figura 5.13 esconde a variação de confiabilidade quando os usuários começam

com um r_0 diferente de 5. Isso ocorre porque o número de ocorrências de avaliações iniciais com pontuação igual a 5 é muito maior proporcionalmente a todos os outros. Dessa forma, é interessante destacar a variação das avaliações para um dado r_0 . Para isso, calcula-se proporcionalmente a variação para o mesmo r_0 . A Figura 5.14, portanto, apresenta em cada ponto (x, y) , uma graduação de cor proporcional ao valor $\Delta r_0 / \sum_{-5}^5 \Delta r_0$. A partir desses resultados, observa-se que os usuários tendem a manter as suas pontuações, quando começado com 5 e 10. No entanto, quando eles começam com confiabilidade entre 6 e 9, a tendência de melhora é superior.

É necessário considerar que as notas de confiabilidade recebidas pelos usuários são dependentes da participação de demais contribuintes. Isso significa que se o usuário Waze se posicionar em regiões isoladas, ele estará menos suscetível a ser avaliado.

5.9 Associação entre pares de alerta na mesma rua

Esta seção investiga se há associação entre pares de alerta que ocorrem na mesma rua para curto intervalo de tempo. O objetivo é descobrir se algum evento exerce influência sobre outro ou se eles possuem alguma tendência a ocorrerem em instantes de tempo próximos. A associação é contabilizada a partir da ocorrência do primeiro alerta e então, para um intervalo de 5 minutos são verificados as associações. A Tabela 5.3 contabiliza a associação entre os pares de alertas que se repetem de diferentes tipos na mesma rua para intervalos de 5 minutos. Testes acima do intervalo de 5 minutos foram realizados, entretanto é possível que para intervalos maiores não haja dependência ou relação entre os pares, por isso foi descartado a possibilidade de incluí-los por não considerar tão precisos. Testes abaixo do intervalo de 5 minutos também não proporcionaram significativas associações e também foram descartados.

Tabela 5.3: Pares de alertas para intervalos de 5 minutos por rua

Pares de Alerta	Total
Congestionamento e Acidente	648
Congestionamento e Risco climático	3798
Congestionamento e Rua fechada	0
Acidente e Risco climático	421
Acidente e Rua fechada	0
Risco climático e Rua fechada	0

Nota-se que há significativa associação entre alertas do tipo “Risco Climático” e “Congestionamento”, “Acidente” e “Congestionamento” e “Risco Climático” e “Acidente”. A “Rua fechada” não representou nenhuma associação evidente no

intervalo de tempo testado. Isso mostra que esse tipo de evento, devido à sua natureza, não leva ao surgimento de novos eventos de alerta. A explicação para esse fato é que a rua fechada permite, se alertado previamente, que o usuário estabeleça novas rotas e altere seu plano de destino quando necessário.

5.10 Frequência na postagem

Em um primeiro momento, a motivação para contribuir com alertas pode ser circunstancial, pois à medida que os eventos surgem, a interação do usuário pode ser estimulada. Essa possibilidade pode ser examinada contabilizando a frequência com que o mesmo usuário tende a contribuir com a rede. Para isso, a frequência que cada usuário contribui para a rede social é calculada a partir do intervalo entre contribuições consecutivas do mesmo usuário. A Figura 5.15 mostra a CDF do intervalo de tempo, em minutos, entre as contribuições consecutivas disparadas pelo mesmo usuário. A maioria dos intervalos, aproximadamente 90%, estão perto de 10 minutos. Isso significa que a maioria dos usuários envia uma nova mensagem depois de 10 minutos da contribuição anterior.

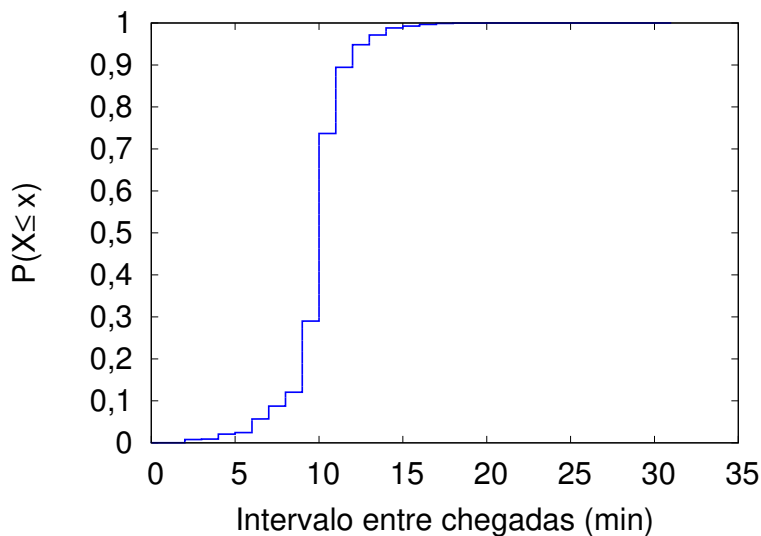


Figura 5.15: CDF do intervalo de tempo entre contribuições consecutivas dos usuários.

A Figura 5.16 mostra o intervalo de tempo, em horas, entre a primeira e a última contribuição do mesmo usuário e o número de alertas disparados por ele. Os resultados obtidos indicam que usuários com muitas contribuições tendem a ter intervalos mais longos entre as primeiras e últimas mensagens, enquanto os usuários com menos contribuições enviam mensagens consecutivas dentro de intervalos mais curtos. É possível observar, no entanto, que o intervalo de tempo entre a primeira e a última contribuição do mesmo usuário aumenta aproximadamente de forma linear.

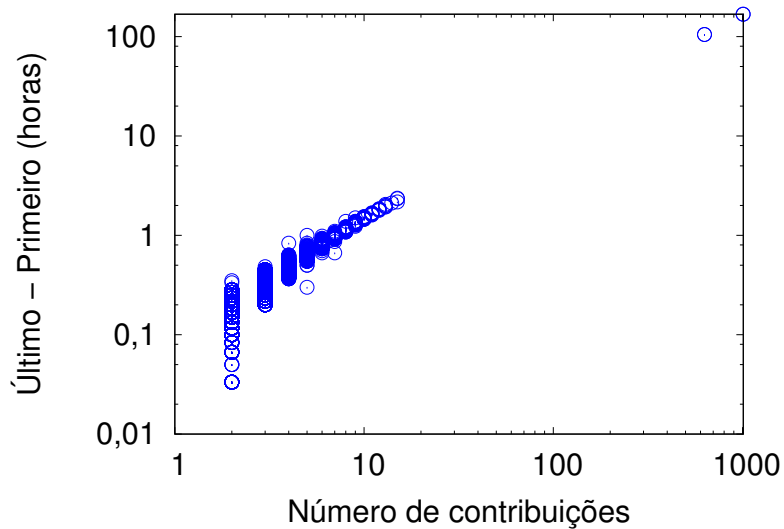


Figura 5.16: Relação do intervalo de tempo entre a primeira e a última contribuição de cada usuário e seu número total de contribuições.

Isso significa mais uma vez que os usuários tendem a manter uma frequência regular entre contribuições consecutivas. Além disso, podemos observar que a maioria dos usuários contribuem com menos de 20 postagens em uma semana. Podemos inferir também que os usuários não participam contribuindo diariamente, mas quando participam, tendem a fazer contribuições de forma consecutiva.

No presente estágio deste trabalho, identificamos dois usuários com muitas contribuições, cerca de 0,01% do total. Esses usuários são detentores de aproximadamente 1000 alertas cada um. Considerando ainda mais dois usuários com aproximadamente 500 alertas, percebe-se que os eventos disparados por eles encontram-se distribuídos em eventos de “Rua fechada” e “Risco climático”. Independente se forem mil ou quinhentas contribuições, esses valores podem ser considerados altos para uma base de dados que captura um intervalo de 7 dias. Mesmo que esse número de contribuições seja elevado para um único usuário, não há argumentos contrários para descartá-lo.

Capítulo 6

Conclusões e Trabalhos Futuros

Este trabalho foi o primeiro que se sabe a analisar um conjunto de dados da rede social Waze com informações de Massachusetts. O principal objetivo deste trabalho foi investigar o comportamento dos usuários em uma rede social veicular, levando em conta os possíveis fatores externos que influenciam a participação de cada usuário. Ainda, procura-se compreender como a qualidade das informações podem ser afetadas devido a esses fatores externos. O estudo sobre a motivação pela contribuição e qualidade das informações foram o foco deste trabalho, sendo esses os principais desafios enfrentados por aplicações de filosofia colaborativa.

Os resultados mostraram que a maioria das contribuições ao Waze foram acionadas durante as horas de pico, de 8h às 9h e das 17h às 18h em dias de semana. Nos finais de semana, os horários de maior contribuição mudaram, sendo que eles ocorreram com maior frequência na parte da tarde, apesar de apresentar uma quantidade significativamente inferior de contribuições em comparação aos dias de semana. As contribuições postadas no horário da madrugada se mostraram pouco representativas e outra explicação para o fato, é que vias movimentadas exigem maior necessidade de interação e busca por informação pelos usuários. Além disso, análises por hora sobre a quantidade máxima e mínima de alertas durante e finais de semana revelaram alguma variação que não impediu a obtenção de resultados. Tanto para os dias de semana quanto aos finais, a maioria das contribuições foram do tipo de “Congestionamento”. Ainda, pode-se perceber que os interesses por postagens de alerta relacionadas aos subtipos “Trânsito muito pesado”, “Trânsito parado” e “Trânsito moderado” são predominantes. Isso permite concluir que os usuários estão mais motivados a contribuir em casos que apresentam problemas dessa natureza, além de ser um momento de maior facilidade de interação do condutor com um dispositivo móvel.

A confiabilidade das contribuições dos usuários na rede social também foi verificada. Foi possível perceber quais alertas demonstraram maiores chances de serem bem avaliados. Os resultados indicaram que “Evento de rua fechada” e “Risco

de animais ao redor” apresentaram significativa probabilidade de obter altas notas de confiabilidade comparado aos demais alertas. Isso demonstra que a veracidade das avaliações recebidas por esse tipo de alerta são pouco contestadas e o usuário tem uma considerável certeza ao encontrar e confirmar esse tipo de evento. Os resultados mostraram também que os usuários que se movimentam a velocidades muito altas não contribuem para a rede social devido provavelmente a suas responsabilidades como condutor. Por outro lado, condutores em velocidades abaixo de 30 km/h são os que mais contribuem já que provavelmente encontram-se em congestionamentos.

Os resultados também mostraram influência da velocidade dos usuários no momento da postagem do alerta sobre a nota de confiabilidade adquirida. Percebe-se que condutores que postam alertas a velocidades mais altas possuem tendência a receberem notas mais baixas de confiabilidade, se comparado aos momentos em que as postagens ocorrem em velocidades mais baixas. A explicação para esse fato é que usuários que interagem com o dispositivo móvel em altas velocidades tendem a cometer mais erros na utilização do aplicativo do que em velocidades inferiores, pois sua dedicação entre as duas tarefas de forma simultânea é dificultada. Foi possível perceber também a relação entre o atraso na transmissão da postagem do usuário e a nota de confiabilidade atribuída ao seu alerta. Percebeu-se que contribuições com maior atraso de publicação, geralmente recebem notas inferiores de confiabilidade e que em oposição a esse fato, contribuições com menor atraso recebem notas de confiabilidade superior. A explicação para isso está no fato de que os alertas que passam mais tempo para serem publicados na rede social podem perder seu valor devido a aspectos temporais e talvez nem mais existam na realidade, quando disponível no Waze. Por exemplo, o caso dos congestionamentos que podem não existir mais após um intervalo de tempo. Outro exemplo poderia ser uma rua que já não encontra-se mais fechada. É possível que alertas postados muito rapidamente também sofram do mesmo efeito, pois os usuários muito próximos do evento podem não perceber ao passar por ele e não terem o tempo necessário para avaliá-los.

No interesse de avaliar as notas de confiabilidade atribuída aos alertas, observaram-se as tendências de degradação e melhoria de cada usuário após postarem sua primeira avaliação. Percebe-se que existe um expressivo número de alertas com baixa confiabilidade no conjunto de dados analisado. Isso motivou a realização de uma análise complementar que, em vez de forma absoluta, realiza a mesma avaliação de forma relativa. A partir dessa análise é possível notar que os usuários que recebem notas de confiabilidade baixa em sua primeira postagem, tendem a não alterar a avaliação das avaliações seguintes. Quando as avaliações mudam, elas sofrem apenas uma pequena variação de confiabilidade, como de 5 para 6, por exemplo. A segunda pontuação de confiabilidade mais usada é a 10

e observa-se que as postagens bem avaliadas também permanecem inalteradas. As outras avaliações de confiabilidade menos utilizadas, de 6 a 9, apresentam comportamento variado pois 6 e 7 tendem a se manter, enquanto 8 e 9 buscam melhorar. Isso indica um comportamento bimodal do usuário à medida que concentram as postagens com avaliações em 5 ou 10 com maior tendência de melhora sempre que possível.

Foi verificada também a associação entre alertas de uma mesma rua para curtos intervalos de tempo. Percebeu-se que houve significativa associação entre alertas do tipo ‘Risco Climático’ e ‘Congestionamento’ em uma mesma rua no intervalo de 5 minutos, além de ‘Acidente’ e ‘Congestionamento’ e ‘Risco Climático’ e ‘Acidente’. Alertas de ‘Rua fechada’ não representaram nenhuma associação evidente no intervalo de tempo testado. Isso mostra que esse tipo de evento, devido à sua natureza, não levou ao surgimento de novos eventos de alerta. A explicação para esse fato é que a rua fechada permite, se alertada previamente, que o usuário estabeleça novas rotas e altere seu plano de destino quando necessário.

A frequência de contribuição dos usuários também foi verificada. Os resultados mostraram que, não somente situações circunstanciais podem estimular os usuários a participarem do Waze pois, em média, as contribuições são realizadas de maneira consecutiva dentro de um intervalo de 10 minutos. Foi possível nesta etapa também reforçar a ideia de que usuários não contribuem todos os dias em uma semana. Nota-se, porém, que o intervalo de tempo entre a primeira e a última contribuição do mesmo usuário aumenta aproximadamente de forma linear. Isso significa que os usuários tendem a manter uma frequência regular entre contribuições consecutivas. Além disso, pôde-se observar que a maioria dos usuários contribuíram com menos de 20 postagens em uma semana e que tendem a realizar contribuições consecutivas.

Como trabalho futuro, planeja-se investigar ainda mais o conjunto de dados do Waze e tentar revelar novas informações a respeito dos eventos de alerta que se deve prestar atenção no trânsito de uma cidade e, posteriormente, prever quais que podem acontecer em seguida. Pretende-se também ampliar a verificação sobre comportamentos peculiares de usuários maliciosos e propor um modelo para descarte de informações falsas.

Referências Bibliográficas

- [1] CARVALHO, R. E., LOPES, C. T. “Social network analysis to understand behaviour dynamics in online health communities: A systematic review”. In: *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on*, pp. 1–7. IEEE, 2016.
- [2] BENEVENUTO, F., RODRIGUES, T., CHA, M., et al. “Characterizing user behavior in online social networks”. In: *ACM SIGCOMM*, pp. 49–62, 2009.
- [3] CHO, E., MYERS, S. A., LESKOVEC, J. “Friendship and mobility: User movement in location-based social networks”. In: *ACM SIGKDD*, pp. 1082–1090, 2011.
- [4] ALVES, R. D. S., CAMPBELL, I. D. V., COUTO, R. D. S., et al. “Redes veiculares: Princípios, aplicações e desafios”, *Minicursos do Simpósio Brasileiro de Redes de Computadores, SBRC*, 2009.
- [5] XIA, Y., YEO, C. K. “Mobile internet access over intermittent network connectivity”, *Journal of Network and Computer Applications*, v. 40, pp. 126–138, 2014.
- [6] NANDAN, N., PURSCHE, A., ZHE, X. “Challenges in crowdsourcing real-time information for public transportation”. In: *2014 IEEE 15th International Conference on Mobile Data Management*, v. 2, pp. 67–72. IEEE, 2014.
- [7] WU, D., ZHANG, Y., BAO, L., et al. “Location-based crowdsourcing for vehicular communication in hybrid networks”, *IEEE transactions on intelligent transportation systems*, v. 14, n. 2, pp. 837–846, 2013.
- [8] MARKOWSKY, G. “Crowdsourcing, big data and homeland security”. In: *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pp. 772–778. IEEE, 2013.
- [9] TINGHUAI, M., JINJUAN, Z., MEILI, T., et al. “Social network and tag sources based augmenting collaborative recommender system”, *IEICE transactions on Information and Systems*, v. 98, n. 4, pp. 902–910, 2015.

- [10] HE, J., CHU, W. W. “A social network-based recommender system (SNRS)”. In: *Data Mining for Social Network Data*, Springer, pp. 47–74, 2010.
- [11] PIANESE, F., AN, X., KAWSAR, F., et al. “Discovering and predicting user routines by differential analysis of social network traces”. In: *IEEE WoWMoM*, pp. 1–9, 2013.
- [12] MOGADALA, A., VARMA, V. “Twitter user behavior understanding with mood transition prediction”. In: *ACM DUBMOD*, pp. 31–34, 2012.
- [13] CRANSHAW, J., SCHWARTZ, R., HONG, J. I., et al. “The livelihoods project: Utilizing social media to understand the dynamics of a city”. In: *AAAI ICWSM*, pp. 58–65, 2012.
- [14] BACKSTROM, L., SUN, E., MARLOW, C. “Find me if you can: Improving geographical prediction with social and spatial proximity”. In: *ACM WWW*, pp. 61–70, 2010.
- [15] “Waze website - Outsmarting Traffic, Together”. Disponível em: <<https://www.waze.com/pt-BR/>>. Último acesso em 5 de Julho de 2016.
- [16] NETO, V. R., MEDEIROS, D. S., CAMPISTA, M. E. M. “Analysis of mobile user behavior in vehicular social networks”. In: *Network of the Future (NOF), 2016 7th International Conference on the*, pp. 1–5. IEEE, 2016.
- [17] BOUHOUTE, A., BERRADA, I., EL KAMILI, M. “A formal model of human driving behavior in vehicular networks”. In: *IEEE IWCMC*, pp. 231–236, 2014.
- [18] CHOI, S., KIM, J., KWAK, D., et al. “Analysis and classification of driver behavior using in-vehicle can-bus information”. In: *Biennial Workshop on DSP for In-Vehicle and Mobile Systems*, pp. 17–19, 2007.
- [19] KUMAGAI, T., SAKAGUCHI, Y., OKUWA, M., et al. “Prediction of driving behavior through probabilistic inference”. In: *Proc. 8th Intl. Conf. Engineering Applications of Neural Networks*, pp. 117–123, 2003.
- [20] ZHANG, X., SHANGGUAN, L., YUAN, Y. “A Crowd Wisdom Management Framework for Crowdsourcing Systems”, *IEEE Access*, v. 4, pp. 9764–9774, 2016.
- [21] “Website - Waze Dataset Mass”. . Disponível em: <<https://data.cityofboston.gov/Transportation/Waze-Jam-Data/yqgx-2ktq>>. Último acesso em 10 de Fevereiro de 2017.

- [22] “Website - The open data initiative for the Commonwealth of Massachusetts”. Disponível em: <<https://data.mass.gov/>>. Último acesso em 25 de Junho de 2017.
- [23] TORTELLA, P. L., COELLO, J. M. A. “Análise de Sentimentos em Mídias Sociais”, .
- [24] “Website - Google Maps”. Disponível em: <<http://www.google.com.br/maps/about/>>. Último acesso em 22 de Janeiro de 2017.
- [25] “Website - Here Maps”. Disponível em: <<https://here.com/en>>. Último acesso em 30 de Janeiro de 2017.
- [26] “Website - Waze Data From Massachusetts State”. mar 2015. Disponível em: <<https://data.cityofboston.gov/Transportation/Waze-Jam-Data/yqgx-2ktq>,<https://data.cityofboston.gov/dataset/Waze-Point-Data/b38s-xmkq>,<https://data.cityofboston.gov/Transportation/Waze-Alert-Data/h8aq-6mw7>>. Último acesso em 16 de Maio de 2016.
- [27] PREMCHAI SWADI, W., ROMSAIYUD, W., PREMCHAI SWADI, N. “Navigation without GPS: Fake location for mobile phone tracking”. In: *ITS Telecommunications (ITST), 2011 11th International Conference on*, pp. 195–200. IEEE, 2011.
- [28] HEDGE COCK, W., MAROTI, M., SALLAI, J., et al. “High-accuracy differential tracking of low-cost GPS receivers”. In: *Proceedings of the 11th annual international conference on Mobile systems, applications, and services*, pp. 221–234. ACM, 2013.