

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Uso de heurística baseada em localidade para otimizar o número de influenciadores na divulgação de pessoas desaparecidas

Lívia Pereira Ozório

JUIZ DE FORA
JANEIRO, 2023

Uso de heurística baseada em localidade para otimizar o número de influenciadores na divulgação de pessoas desaparecidas

LÍVIA PEREIRA OZÓRIO

Universidade Federal de Juiz de Fora
Departamento de ciência da computação
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza
Coorientador: Lorenza Leão Oliveira Moreno
Coorientador: Victor Stroele de Andrade Menezes

JUIZ DE FORA
JANEIRO, 2023

USO DE HEURÍSTICA BASEADA EM LOCALIDADE PARA OTIMIZAR O NÚMERO DE INFLUENCIADORES NA DIVULGAÇÃO DE PESSOAS DESAPARECIDAS

Lívia Pereira Ozório

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA,
COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO
DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Jairo Francisco de Souza
Dr. em Informática (PUC-RJ)

Lorenza Leão Oliveira Moreno
Dra. em Informática (PUC-RIO)

Victor Stroele de Andrade Menezes
Dr. em Engenharia de Sistemas e Computação (UFRJ)

Stênio São Rosário Furtado Soares
Dr. em Computação (UFF)

Luciana Brugiolo Gonçalves
Dra. em Computação (UFF)

JUIZ DE FORA
09 DE JANEIRO, 2023

*Aos meus amigos pelo companheirismo.
À minha família pelo apoio e incentivo.*

Resumo

O problema de desaparecimento de pessoas é muito frequente e afeta várias famílias no Brasil e no mundo. Por mais que existam diversos esforços para se minimizar este problema, ele ainda não possui uma forma eficaz de solução. Uma alternativa para minimizar os danos e ajudar nas buscas é a divulgação dos casos de desaparecimento. Porém, essa divulgação ainda fica muito restrita, alcançando uma gama muito pequena de pessoas. Logo, este trabalho tem como propósito avaliar quem seria um bom influenciador utilizando-se de programação linear para fazer esta divulgação em redes sociais levando em consideração a localização do desaparecido. Objetivando então maximizar o valor de informações difundidas e minimizar a quantidade de influenciadores necessários para a divulgação.

Palavras-chave: Desaparecidos, Otimização, Difusão de informação

Abstract

The problem of people disappearing is very frequent and affects many families in Brazil and around the world. As much as there are several efforts to minimize this problem, it still does not have an effective solution. An alternative to minimize damage and help in searches is to publicize cases of disappearance. However, this disclosure is still very restricted, reaching a very small range of people. Therefore, this work aims to evaluate who would be a good influencer using linear programming to advertise on social networks, taking into account the location of the missing person. Aiming then to maximize the value of disseminated information and minimize the number of influencers needed for promotion.

Keywords: Missing Persons, Optimization, Information Dissemination

Agradecimentos

À minha família e amigos que sempre me deram suporte e incentivo.

Aos professores Jairo, Lorenza e Victor, que me apoiaram durante toda a trajetória desse trabalho dando todo o apoio para que o trabalho fosse realizado da melhor forma.

E a todos aqueles que, de alguma forma, fizeram parte do meu aprendizado, ressaltando os professores do DCC que sempre fizeram o melhor para que conseguisse captar da melhor forma todos os conhecimentos a mim transmitidos.

*Se tiver o hábito de fazer as coisas com
alegria, raramente encontrará situações
difíceis.* *Robert Baden-Powell*

Conteúdo

Lista de Figuras	7
Lista de Tabelas	8
Lista de Abreviações	9
Lista de Símbolos	10
1 Introdução	11
1.1 Justificativa	12
1.2 Hipótese	13
1.3 Objetivos	13
1.4 Organização	13
2 Fundamentação Teórica	15
2.1 Desaparecidos	15
2.2 Iniciativas para localização de desaparecidos	17
2.3 Problemas de difusão de informação	19
3 Metodologia	21
3.1 Bases de dados	21
3.2 Modelo	25
4 Análises	29
4.1 Cenário 1: sem uso de informação de localização	30
4.2 Cenário 2: com uso de informação de localização	31
4.3 Cenário 3: minimizando envio para fora do raio	32
5 Conclusão	36
5.1 Resultados obtidos	36
5.2 Limitações e Trabalhos futuros	38
Bibliografia	40

Lista de Figuras

3.1	Distribuição de desaparecidos e perfis, respectivamente.	23
3.2	Métricas estatísticas conforme a distância entre cada desaparecido e os perfil	25
4.1	Experimento - Cenário 1	30
4.2	Experimentos - Cenário 2	32
4.3	Experimentos para 5KM - Cenário 3	34
4.4	Experimentos para 10KM - Cenário - 3	34
4.5	Experimentos para 20KM - Cenário 3	35

Lista de Tabelas

3.1	Restrições do modelo	28
4.1	Experimentos cenário 1	30
4.2	Experimentos cenário 2	31
4.3	Experimentos cenário 3	33

Lista de Abreviações

DCC Departamento de Ciência da Computação

UFJF Universidade Federal de Juiz de Fora

PL Programação linear

PO Pesquisa operacional

Lista de Símbolos

Conjuntos:

D	conjunto dos nós de desaparecidos
P	conjunto dos nós de perfil
P_i	conjunto dos nós de perfis selecionados como influenciadores ($P_i \subset P$)
$\hat{P}^{(j)}$	conjunto dos perfis pertencentes ao raio do desaparecido j ($P^{(j)} \subset P$)

Elementos dos conjuntos:

k	nó perfil ($k \in P$)
i	nó influenciador ($k \in P_i$)
j	nó de desaparecido ($j \in D$)
i_k	nó perfil k que é um perfil seguidor do influenciador i ($i_k \in P$)

Variáveis:

r_{ij}	variável binária que indica se i divulga j ($r_{ij} \in \{0, 1\}$)
d_j	variável que indica quantidade total de informação recebida sobre j ($d_j \in \mathbb{R}$)
d_{kj}	variável binárias que indica se k recebe informação sobre j ($d_{kj} \in \{0, 1\}$)

Constantes:

f_{ik}	indica se o perfil k recebe informação do influenciador i ($f_{ik} \in \{0, 1\}$)
c_{ij}	indica a relevância de um influenciador i para um desaparecido j ($c_{ij} \in \mathbb{R}$)
R	indica qual a distancia máxima para ser considerado perto de um desaparecido ($R \in \mathbb{R}$)
W	indica o peso dado para a c_{ij} ($W \in \mathbb{R}$)
M	número máximo de influenciadores que podem divulgar um desaparecido ($M \in \mathbb{N}$)
N	número máximo de desaparecidos divulgados por um influenciador ($N \in \mathbb{N}$)

1 Introdução

O problema dos desaparecidos é um problema social que afeta diversas famílias. No Brasil, cerca de 80 mil pessoas desaparecem por ano (UOL, 2021). Este número se torna ainda maior quando se consideram as subnotificações existentes, como mostrado em (OLIVEIRA, 2014), as quais ocorrem quando os parentes não notificam o caso, ou os policiais se recusam a fazer o boletim de ocorrência, por exemplo, alegando que só o podem fazer 24 horas depois do desaparecimento e dificultando assim a localização do indivíduo antes que se realize a notificação.

Porém as primeiras 72 horas são fundamentais para que os procedimentos de busca sejam bem sucedidos (A&E BRASIL, 2021). Depois de 48 horas, as pistas começam a sumir e as testemunhas começam a se esquecer de fatos importantes sobre o desaparecimento. Além disso, se o desaparecido for uma pessoa vulnerável, como crianças, adolescentes ou dependentes químicos, o risco de morte depois deste tempo é alto.

Para a resolução dos casos, existem variados métodos e recursos utilizados. Há a utilização de *softwares* de processamento de imagens para simulação da aparência futura do indivíduo (SANTOS; ARTERO, 2011); identificação de DNA de pessoas encontradas já em estado de decomposição, que é a forma utilizada para a identificação em casos como o desastre de Brumadinho; além das investigações realizadas através de pistas coletadas e relatos de testemunhas. Outra ferramenta utilizada para ajudar na localização é a divulgação de desaparecidos em mídias sociais, o que faz com que o desaparecido seja encontrado mais rapidamente, já que outras pessoas as quais não estão diretamente ligadas ao caso podem dar mais pistas (A&E BRASIL, 2021).

Como abordado acima, a divulgação abrangente de dados de desaparecimento é um bom método para auxiliar na resolução de casos. Porém, essas divulgações, em geral, são realizadas por sites de governo ou ONGs, deixando tudo muito restrito e não sendo capaz de alcançar boa parte da população. Em (SILVA, 2019), a autora mostra como é possível utilizar as redes sociais para tornar essa divulgação mais acessível, identificando pessoas influenciadoras na rede para que a divulgação seja feita através delas. Porém,

não é trivial identificar as melhores métricas para auxiliar na seleção de influenciadores em mídias sociais para o contexto de pessoas desaparecidas.

1.1 Justificativa

O processo de busca de desaparecidos é bastante lento. Como dito anteriormente, um caso tem maiores probabilidades de ser resolvido durante as primeiras 72 horas. Assim, é necessário encontrar meios que auxiliem na rápida conclusão dos casos. Uma das formas de acelerar a localização de desaparecidos é a partir de divulgação em mídias. Um fato é que nos dias atuais a Internet vem desempenhando um papel fundamental para a divulgação de informações, como é apresentado em (KEMPE; KLEINBERG; TARDOS, 2003), principalmente as redes sociais as quais são muito utilizadas no mundo como um todo. Com elas, é possível conectar várias pessoas e passar uma informação para diversos usuários em apenas alguns segundos.

Dentro das redes é possível observar pessoas com maior influência, as quais são capazes de difundir de forma mais eficiente e rápida a informação para um determinado público alvo (LOU; YUAN, 2019). No contexto deste trabalho, são consideradas como público alvo pessoas que estão a uma determinada distância máxima do local do desaparecimento, já que estes são os que, provavelmente, poderiam ter alguma informação, além de se motivarem a auxiliar na divulgação e busca.

Então, neste ponto, é necessária a avaliação de quem seria um bom influenciador. A pessoa considerada influenciadora tem como objetivo fazer com que a informação a qual está divulgando chegue ao máximo de pessoas que pertencem ao público alvo da mensagem. Muitas das vezes um influenciador pode ser alguém que consegue alcançar uma quantidade muito alta de usuários, porém estes não pertencem ao público alvo do desaparecimento. Neste caso, influenciadores globais não seriam uma boa escolha, já que embora a informação enviada chegue a muitas pessoas, a maioria não possui interesse no assunto. Assim, haveria uma grande disseminação do desaparecimento, mas causando uma inundação da rede, fazendo com que um número muito alto de pessoas não interessadas recebam a mensagem.

1.2 Hipótese

O problema da busca de bons influenciadores em uma rede complexa é demorado. Porém, o uso de Pesquisa Operacional - PO tem sido usado em problemas similares (ALI; LI; CHEN, 2021) e, assim, espera-se ser adequado para auxiliar na escolha

Além disso, o uso da informação de localização também é importante para melhorar a filtragem dos usuários mais adequados para receberem a mensagem sobre o desaparecimento. Quando se utiliza dessas informações de localização no processo de busca de influenciadores pode acontecer de se obter uma menor dispersão da informação na rede, porém usuários mais importantes são atingidos.

Com isto, é esperado que o uso de técnicas de PO que considerem as informações sobre o local do desaparecimento e dos usuários de uma Rede Social gere soluções que melhorem a precisão da entrega da mensagem para o público alvo desejado sem, contudo, causar inundação da mensagem fora deste público.

1.3 Objetivos

Este trabalho tem como objetivo ajudar no problema de pessoas desaparecidas, utilizando-se da iniciativa de divulgação dos casos de desaparecimento. Com o auxílio de PO, este trabalho objetiva encontrar melhores métricas e heurísticas para selecionar influenciadores. Assim, espera-se obter um maior alcance na divulgação dos desaparecidos para o público alvo em questão (pessoas as quais se encontram próximas ao desaparecido) em redes sociais.

1.4 Organização

O restante do trabalho está assim estruturado: o Capítulo 2 discute o problema de desaparecimento civis através de trabalhos na área de segurança pública, além de apresentar o conceito de difusão de informação. No Capítulo 3 é mostrada a metodologia utilizada no trabalho, bem como as bases de dados utilizadas no estudo e o modelo matemático que descreve o problema abordado. Já no Capítulo 4 são mostrados os resultados obtidos

com os experimentos junto com as análises. Por último, no Capítulo 5 traz as conclusões e as considerações finais juntamente com discussões sobre trabalhos futuros.

2 Fundamentação Teórica

Este capítulo aborda referenciais teóricos, contendo uma contextualização sobre os desaparecimentos civis, explana também sobre algumas iniciativas para a localização de desaparecidos, e sobre difusão de informações através de redes sociais. Além de também trazer trabalhos relacionados ao tema.

2.1 Desaparecidos

Embora não tenha-se uma definição unânime sobre a expressão “pessoa desaparecida” em âmbito mundial, no Brasil, pela *lei nº 13.812 Art.2º*¹, de 16 de março de 2019, uma pessoa desaparecida é “todo ser humano cujo paradeiro é desconhecido, não importando a causa de seu desaparecimento, até que sua recuperação e identificação tenham sido confirmadas por vias físicas ou científicas”. Os desaparecimentos podem ser tanto voluntários, como involuntários. O primeiro trata-se de quando uma pessoa tem a intenção de sair sem comunicar a ninguém e não volta mais. Já o segundo são casos onde a pessoa desaparece sem ter esta intenção, sendo este desaparecimento provocado por diversas razões (HIRSCHEL; LAB, 1988).

O desaparecimento de pessoas é um problema social que impacta de diversas maneiras quem tem familiares e pessoas próximas que se encontram na condição de desaparecido. Como relata autora (OLIVEIRA, 2007), a nossa sociedade não é preparada para lidar de maneira devida com esta situação, os familiares de pessoas desaparecidas dependem de amparo jurídico e psicológico, já que enfrentam um período muito difícil e que pode se alongar por bastante tempo, porém é encontrado falta de preparo policial e de legislação específica para se lidar com isto. Os parentes e amigos sofrem com a agonia de não terem notícias, não saberem como a pessoa se encontra, se está morta ou passando por alguma situação perigosa.

Como mostra o artigo (CALMON, 2019), o número de casos no Brasil é bem

¹http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13812.htm

relevante quando comparado aos demais países da América Latina, chamando atenção de várias entidades internacionais. Somente durante o ano de 2021 foram registrados no país 65.225, ou seja uma média de 179 desaparecimentos diários segundo o *Anuário Brasileiro de Segurança Pública*². Mas é possível perceber que isto não acontece apenas no Brasil, este número é bem elevado em diversos outros países: no Reino Unido no ano de 2020 houve 353.000 reportados, sendo um caso a cada 90 segundos (MISSING PEOPLE, 2021).

Quando é falado sobre desaparecimento, vários fatores são levados em consideração para que seja possível encontrar a pessoa, sendo alguns destes fatores a idade, a localização, o ambiente social no qual a pessoa está inserida, o tempo de busca, entre diversos outros (MOREWITZ; COLLS, 2016). Isto leva a uma multidisciplinaridade quando falamos sobre o tema. Dentre estes fatores, tem um que é um ponto crucial para se resolver um caso de desaparecimento que é o tempo de busca: quanto mais se alonga este período, menos chance se tem de a busca ser bem sucedida. Existe o chamado tempo crítico que tem duração de 72 horas. E esse é o tempo em que normalmente quando uma pessoa desaparece ela é localizada com vida. Depois das 72 horas, a chance de se encontrar o desaparecido vai diminuindo exponencialmente. Durante as primeiras 48 horas é o tempo que os investigadores têm para coletar dados para a investigação, já que as pistas do desaparecimento vão sendo perdidas, as testemunhas vão se esquecendo dos fatos relevantes, e a cada hora as informações passam a ser mais confusas, como diz (A&E BRASIL, 2021).

Por conta disto, é necessário que a forma de divulgação desses casos chegue a um número grande de pessoas, um dos outros pontos em que é possível se perceber problemas dentro deste processo. As bases onde constam os casos são extremamente falhas, faltam diversas informações, além de não serem unificadas, o que impossibilita confrontar os dados, conforme é mostrado no artigo (BRASIL; SILVA, 2015). Com isto, o tempo de busca é prejudicado. Se as bases tivessem mais informações e elas chegassem a uma quantidade maior de pessoas essas velocidades poderiam ser bem mais rápidas do que a que temos no momento.

Então, há hoje ainda a necessidade de se encontrar uma forma eficiente de divulgar

²<https://forumseguranca.org.br/wp-content/uploads/2022/06/anuario-2022.pdf?v=5>

as informações destes casos para que seja possível localizar mais rapidamente o indivíduo desaparecido, já que quanto mais pessoas sabendo, mais provável é de alguma destas verem e saberem do paradeiro do desaparecido e darem informações que ajudem a solucionar o caso.

2.2 Iniciativas para localização de desaparecidos

Durante os anos surgiram variadas iniciativas para o auxílio no problema de pessoas desaparecidas. Dentre elas, umas com intuito mais voltado à identificação, outras para ajudar nas buscas policiais, e também para a divulgação de desaparecidos.

No artigo (EVISON et al., 2016) é mostrado como um algoritmo é capaz de, a partir de fósseis encontrados de pessoas, reconstruir a fisionomia do indivíduo, possibilitando assim, que seja mais fácil a identificação do indivíduo encontrado. Outra ferramenta utilizada para ajudar na identificação é a coleta de DNA de familiares do desaparecido para que, quando um fóssil for encontrado, seja possível comparar os materiais genéticos e possa ser feita a identificação, sendo utilizada em buscas de desaparecidos por tragédia, como a de Brumadinho, onde essa estratégia tenha sido utilizada (PARANAIBA, 2019).

Para ajudar nas buscas de desaparecidos foi desenvolvido no Zimbabué um algoritmo para ser colocado em câmeras na cidade. Este algoritmo é capaz de, a partir de imagens contidas do desaparecido em um banco de dados, identificar a pessoa desaparecida mandando uma mensagem contendo a localização onde a pessoa se encontra (MUYAMBO, 2018), dessa forma ajudando nas buscas. Uma segunda ferramenta que ajuda nas buscas é um *software* de processamento de imagem para simulação de fisionomia atual do indivíduo, com bastante utilidade na busca de crianças, já que estas mudam muito com o decorrer do tempo (SANTOS; ARTERO, 2011).

Além destas iniciativas, existem também as que são utilizadas para a divulgação dos casos. Dessa forma possibilitando pessoas que não estão diretamente ligadas ao caso deem mais pistas, como ocorreu no caso reportado em (AYRES, 2014), quando uma criança foi localizada por causa do aparecimento de sua foto na conta de luz. Em (LAMPINEN; ARNAL; HICKS, 2009), também é mostrado que divulgações feitas com posters em supermercados ajudam a localização de crianças desaparecidas.

Como as mídias têm uma grande importância no processo de divulgação de informação (A&E BRASIL, 2021), já que com elas é possível alcançar um número muito maior de pessoas. Baseando-se nisto, surgem iniciativas voltadas à divulgação nas mídias de maior porte, como em (BRASIL; SILVA, 2015), o qual cria na Internet um *site* contendo informações sobre desaparecidos, assim como muitas ONGs que utilizam da mesma estratégia. Porém como as informações ainda são muito descentralizadas, não existindo uma base única, existe uma dificuldade na visualização dessas informações. Pensando nisto, no artigo (JR; FERRANTI; SOUZA, 2019) é apresentado um *framework* para coleta de informações de diversos sites onde há informações de desaparecidos e monta um banco de dados com a compilação das mesmas. Este banco de dados pode ser acessado a partir de um *site*³, o qual foi criado para o projeto.

No artigo (CLASSE et al., 2019) apresentado o desenvolvimento do projeto "Encontre-me!" que visa a partir de aplicativos móveis, jogos, redes sociais e plataformas de interação divulgar desaparecidos. O projeto visa ser colaborativo, onde várias pessoas podem adicionar informações. Essas informações são divulgadas em redes sociais e em um jogo fazendo com que fique mais interativo e divertido o processo de divulgação.

Já no projeto (SILVA, 2019), utiliza-se do *Twitter*⁴ e de influenciadores para alcançar uma maior divulgação dos desaparecimentos. Nele é proposto um algoritmo que seleciona influenciadores para a difusão de informação sobre pessoas desaparecidas. No trabalho, a autora explana a dificuldade encontrada por causa da base adotada. Um diferencial deste trabalho é a utilização de uma base voltada a trabalhos que tratam de heurísticas voltadas à utilização de localização.

Este trabalho de conclusão de curso, assim como os outros, visa ajudar na localização de pessoas desaparecidas, utilizando-se da divulgação em redes sociais. Tem-se como objetivo encontrar melhores métricas e heurísticas para a realização da difusão de informação.

³<http://desaparecidos.nrc.ice.ufjf.br/index.php/sobre>

⁴<https://twitter.com/>

2.3 Problemas de difusão de informação

Difusão de informação é caracterizada, segundo (ROGERS, 2010), como o processo pelo qual uma inovação é comunicada através dos canais ao longo do tempo entre os membros de uma sociedade. Assim, o problema foca na forma que esta comunicação é realizada em uma rede de pessoas, definindo um grafo de difusão no qual os vértices representam as pessoas da rede e os arcos indicam quem influencia quem. Este tipo de solução é muito utilizado na área do *marketing* digital quando é tratado problemas de difusão.

Os algoritmos de difusão visam conseguir uma difusão mais eficiente das informações na rede em questão. Algumas abordagens utilizam como heurística a localização das pessoas pertencentes à rede. Outras apenas as ligações existentes entre os vértices da rede.

Em (WANG; STREET, 2018) é proposto um algoritmo para difusão de informação de um novo produto através das redes sociais. O trabalho tem como objetivo capturar a influência direta e indireta dos usuários, a atenuação da influência associada à profundidade na árvore, o declínio da influência temporal, que é o declínio da influência em relação ao tempo dela, e a dinâmica de difusão individual, através de um melhoramento no modelo linear. O trabalho mostra excelentes resultados com relação ao desempenho de tempo e propagação de influência nas quatro bases utilizadas.

Já em (BOUROS; SACHARIDIS; BIKAKIS, 2014) os autores utilizam uma abordagem de difusão com modelo de cascata independente utilizando um algoritmo de caminho em grafos e a localização dos usuários com o intuito de identificar usuários que influenciam um grande número de outros usuários em uma determinada região. A vantagem do modelo é que só funciona bem para áreas pequenas. À medida que aumenta o número de usuários na região, o processamento fica muito lento.

Em (ALI; LI; CHEN, 2021) é realizada uma busca por influenciadores para um produto, o qual se deseja escolher o local onde é relevante a divulgação e depois são selecionados os influenciadores. Durante o projeto foram utilizadas três heurísticas para se determinar qual era a melhor na escolha dos influenciadores. As três heurísticas testavam se é melhor que o influenciador deve pertencer à região à qual o produto deve ser divulgado ou não.

No contexto dos desaparecidos, o trabalho de Silva (SILVA, 2019) apresenta um algoritmo que considera a localização dos usuários e desaparecidos para fazer a difusão de informação em redes sociais. O objetivo foi maximizar o valor de informações de desaparecidos difundidas e minimizar a quantidade de influenciadores necessários para esta tarefa. Neste trabalho, é utilizado um modelo semelhante, onde objetiva-se maximizar a quantidade de informação difundida, porém levando em consideração também a minimização de mensagens enviadas para pessoas não interessadas na mensagem.

3 Metodologia

Este trabalho possui um caráter de pesquisa aplicada, já que é voltado a um problema específico, o qual pode ser replicado para o mundo real, sendo feitos testes simulados capazes de mostrar o comportamento da divulgação e da seleção de influenciadores utilizando-se diferentes parâmetros e critérios. Para que o estudo seja realizado, os dados são consumidos por um programa que se utiliza de Programação Linear - PL. Dessa forma, a resposta dada pelo modelo permitirá gerar estatísticas que comprovam a hipótese, sendo assim um trabalho caracterizado como quantitativo.

Para a realização desta pesquisa, serão utilizadas as seguintes etapas: (1) coleta dos dados de perfis da rede social e de desaparecidos a serem utilizados; (2) experimentos simulados de divulgação de desaparecidos; (3) análise dos dados obtidos pelos experimentos.

O modelo de programação linear formulado no trabalho (SILVA, 2019) com algumas modificações foi a base para que as simulações de divulgação aconteçam. O trabalho utilizou dois diferentes critérios para suas simulações: o primeiro selecionando influenciadores através da maximização da quantidade de informação recebida pelo público alvo da divulgação, sendo considerado público alvo pessoas próximos aos desaparecidos; o segundo leva em consideração também a minimização do número de pessoas fora do público alvo que recebem a mensagem. Além disto, em todos os critérios, o tamanho da área de abrangência a ser considerada como área do desaparecido foi variada, possibilitando, dessa forma, que fosse feita uma análise sobre os dados e a hipótese fosse comprovada.

3.1 Bases de dados

Como o estudo tem o intuito de achar uma forma de escolher melhor quem é um bom influenciador para divulgar desaparecidos utilizando-se de redes sociais, são utilizadas simulações com bases reais. Para isto, utiliza-se de duas bases: a primeira conta com dados de desaparecidos dos Estados Unidos, já a segunda, com dados de uma rede de

perfis extraída de uma rede social sendo um grafo direcionado $D(V, E)$, onde V é o conjunto de vértices dado pelos perfis da rede, os arcos definem a relação de seguir e ser seguido, de forma que se o arco $(i, j) \in E$, indica que o perfil i , segue o perfil j (ou o contrário). Essas bases são válidas para o estudo, já que a base contendo os perfis e seus *check-in* formam a rede pela qual a informação se propaga. Com a base contendo as localizações, consegue-se localizar quem na rede é o público alvo da publicação que é um subconjunto P_i do conjunto V . A base com os desaparecidos contém a informação a qual se deseja divulgar e a localização a qual cada desaparecido pertence, permite definir o raio de abrangência da divulgação. Além disso, a localização do desaparecido é o que delimita a área a qual o público alvo pertence.

A base de dados de desaparecidos foi extraída do site *NamUs*⁵ com auxílio do *framework* criado em (JR; FERRANTI; SOUZA, 2019). A base contém 13.888 dados de pessoas desaparecidas em todos os estados dos Estados Unidos, registrados até 2021 pelo projeto “*NamUs*”, conduzido pelo *National Institute of Justice*. Já a segunda⁶, contém a rede de perfis, ou seja, um grafo com 196.591 nós e 950.327 arestas, além de 6.442.890 *check-in* realizados também em sua maioria nos Estados Unidos. Essa base foi escolhida porque é destinada para experimentação de abordagens baseadas em localização para redes complexas, como utilizada no artigo (ALI; LI; CHEN, 2021). Esta base foi criada com base em dados oriundos da rede social “*Gowall*” em 2010.

Com os dados coletados, a Figura 3.1 mostra a dispersão dos desaparecidos (Figura 3.1.a) e perfis (Figura 3.1.b) nos estados americanos, nos locais mais escuros há uma concentração grande, enquanto em locais mais claros a concentração é baixa. Nesta figura percebe-se que os locais de maior concentração de desaparecidos também contêm uma grande concentração de perfis. Isto é importante para o trabalho já que, como é utilizada uma abordagem por localização, é necessário que existam uma quantidade relevante de perfis próximos aos desaparecidos.

⁵<https://www.namus.gov/MissingPersons/Search/>

⁶<http://snap.stanford.edu/data/loc-Gowalla.html>

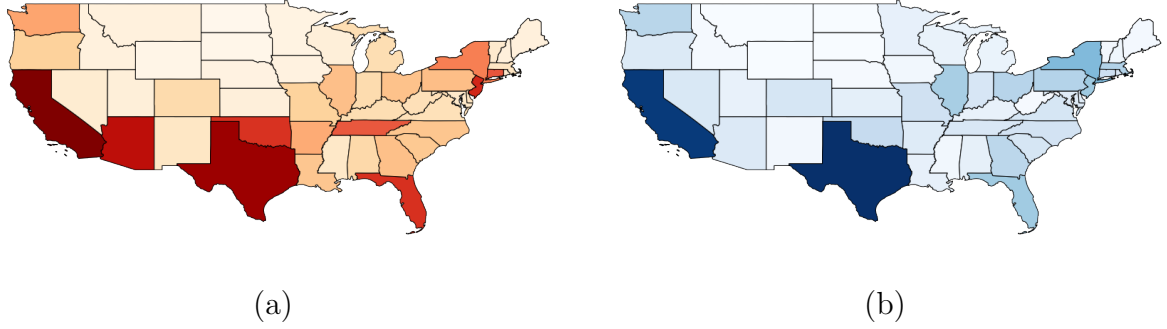


Figura 3.1: Distribuição de desaparecidos e perfis, respectivamente.

Neste trabalho é utilizado um grafo direcionado $D(V, E)$, onde V é um conjunto de perfis e desaparecidos, já o E é o conjunto de ligações que podem ser de um perfil para outro perfil, tendo assim uma relação de seguido e seguidor, ou de um perfil para um desaparecido que contém o valor de relevância do perfil divulgar o desaparecido. No problema é considerado como perfil um nó k pertencente ao conjunto P , sendo P um subconjunto de V . Os perfis alvos de um desaparecido j pertencente ao conjunto D , também um subconjunto de V , (denotados como $\hat{P}^{(j)}$) são todos os perfis que estão até uma distância R em relação à localização de j . Já os influenciadores denotados por i que pertencem ao conjunto P_i , subconjunto de P , são perfis selecionados para enviar uma mensagem sobre j para seus seguidores k . Sendo um seguidor de i denotado por i_k .

Para fazer o estudo foi determinado que um nó k candidato a influenciador deve possuir no mínimo 5 i_k e ter pelo menos 10% de i_k dentro do raio R de um nó j , isto foi realizado para que não ocorra de um influenciador influenciar um número muito pequeno de pessoas e ao mesmo tempo ter certeza de que um mínimo de pessoas as quais serão influenciadas pertencem ao raio do desaparecido.

A Figura 3.2 nas partes (a), (b) e (c) apresenta dados de distância entre os nós k e os nós j . No eixo X temos os nós j ordenados pelo valor crescente do eixo Y e no eixo Y, os valores referentes a cada um dos casos.

A Figura 3.2.a representa distância máxima entre um nó j e os nós k , variando os valores resultantes de 71 a 398km. A Figura 3.2.b representa a média de distância entre um nó j e os nós k variando os resultados de 32 a 79km. A Figura 3.2.c representa a mediana da distância entre um nó j e os nós k variando de 25 a 40km os valores obtidos.

A maior distância existente entre um desaparecido e uma pessoa a qual recebe sua informação é 509km, porém em grande parte dos desaparecidos esse valor fica em torno de 398km. Quando se olha a média de distância dos desaparecidos para os seu possíveis recebedores da mensagem, a grande maioria fica em torno de 49km; já a mediana fica em torno de 33km.

Com isto é possível determinar quais os valores de raio de abrangência devem ser utilizados para a realização dos experimentos, sendo escolhidos para este trabalho os valores 5, 10 e 20km. Foi escolhido um valor bem baixo para tentar selecionar influenciadores locais, mas que possuem alcances menores. Em contrapartida, 20km para testar um valor mais alto que possui mais seguidores na região de i , mas que seja favorável ao teste com localizações ainda restringindo-os para cada desaparecido.

Já a 3.2.d apresenta o percentual dos i_k que estão dentro do raio de um desaparecido caso um determinado perfil seja selecionado como influenciador. O eixo X representa os candidatos a i se fossem selecionados para divulgar j , esses dados são ordenados em relação ao valor de Y. Enquanto, o eixo Y representa a porcentagem de seguidores de k que estão dentro do raio do desaparecido, neste caso considerando R igual a 20Km.

Ao analisar a porcentagem de perfis que tem interesse em receber a mensagem, ou seja, estão dentro do raio do desaparecido, obtivemos perfis candidatos a influenciadores que possuem apenas 10% dos seus seguidores dentro do raio, mas também perfis que obtiveram 68%. Como a grande maioria está por volta de 30% e 50% faz sentido realmente cortar candidatos que têm menos de 10% de seus seguidores fora do raio. Isto porque ainda temos uma grande quantidade de candidatos, não deixando muito limitado.

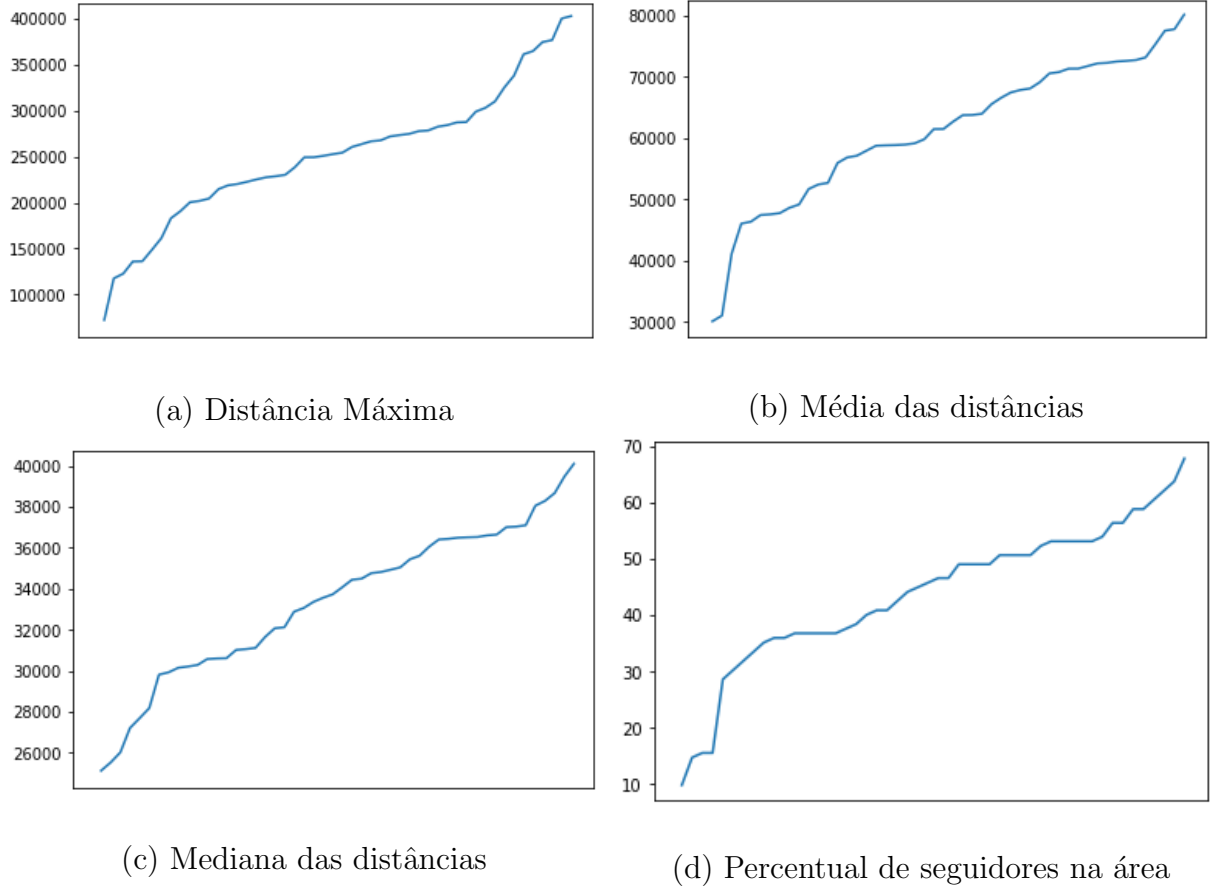


Figura 3.2: Métricas estatísticas conforme a distância entre cada desaparecido e os perfil

Sendo assim, esta pesquisa tem seu universo amostral formado pelos perfis k e pelos desaparecidos j contidos nas bases de dados acima citadas. Sendo candidatos a influenciador todos os perfis que respeitem as restrições de possuir pelo menos 5 seguidores e tenha pelo menos 10% de seus seguidores dentro do raio R .

3.2 Modelo

Para se realizar o estudo, o trabalho utiliza-se de um modelo baseado no modelo de programação linear proposto no trabalho (SILVA, 2019) para selecionar os influenciadores i que façam com que a informação ganhe maior alcance, sendo assim possível maximizar a divulgação.

Como já explanado na Seção 3.1, o modelo se utilizará das relações entre os perfis k contidos no conjunto P . Cada relação possui um peso, baseado na atividade dos perfis k na rede social, que representa a probabilidade de uma informação enviada por

um influenciador i , pertencente ao conjunto P_i , chegar ao seu seguidor i_k .

Para o modelo selecionar os influenciadores i , são construídas árvores de influência, uma estrutura de dados que computa a influência que um nó k tem sobre os demais nós k da rede. Nestas árvores, o peso da aresta é igual ao peso da ligação do nó para o outro multiplicado pelo valor das arestas pertencentes ao caminho da raiz até a ligação, ou seja, o valor de uma ligação decai a cada nível decidido. Neste trabalho é considerado que um nó k recebe a mensagem se a probabilidade de isto acontecer for maior que 0.8, realizando um corte na árvore a partir da aresta que for menor. Com isto é utilizada uma constante f_{ik} pertencendo ao conjunto $\{0, 1\}$, que indica no modelo se o perfil k recebe ou não a informação do influenciador i .

Um ponto importante é que em casos de divulgação, não se é interessante inundar a rede, caso uma pessoa receba muitas mensagens de diferentes pessoas desaparecidas possivelmente começará a ignorar todas, sendo ideal que receba apenas mensagens de pessoas em sua região. Com isto, é necessário se fazer um controle sobre a quantidade de informação enviada. Para isto o modelo limita a M a quantidade de influenciadores i que podem divulgar um determinado desaparecido j e a N , o número máximo de desaparecidos j que um influenciador i pode divulgar. Deste modo, evita-se que um influenciador i vire um spam na rede.

Outro ponto que merece atenção é a quantidade de mensagem enviada para um perfil k sem interesse na informação. Isto pode ser suavizado se pensar também em minimizar a quantidade enviada para perfis não interessados, que estão fora do público alvo, gerando-se um equilíbrio entre a maximização de mensagens recebidas dentro dos perfis alvos em $\hat{P}^{(j)}$ e a minimização de mensagens recebidas por perfis k fora de $\hat{P}^{(j)}$. Dessa forma, a constante c_{ij} é um peso dado a cada ligação de um influenciador i à um desaparecido j , sendo a constante uma classificação levando em consideração a porcentagem de seguidores do influenciador que estão dentro da área do desaparecido sendo o 100% todos os seguidores, fazendo assim com que o modelo priorize influenciadores locais da região. Já o W é uma constante a qual indica o quanto os influenciadores que atingem um percentual maior de perfis dentro da área do desaparecido serão privilegiados pelo modelo. Quando o W é colocado em zero, essa característica é desconsiderada passando

todas as ligações a terem o mesmo peso.

O modelo possui três variáveis sendo elas r_{ij} , d_j e d_{kj} . A primeira variável pertence ao conjunto $\{0, 1\}$, que indica se o influenciador deve ser selecionado ou não para divulgar informações sobre o desaparecido j . Já d_{kj} pertence ao conjunto $\{0, 1\}$, indicando se o perfil k recebe informação sobre o desaparecido j . Esta variável só é criada para os perfis em $\hat{P}_{(j)}$, ou seja, que teriam interesse na informação. Por último o d_j é uma variável do conjunto dos números reais positivos que possui o valor referente a quantidade de pessoas que receberam informação sobre o desaparecido j .

A função objetivo, como já mencionado anteriormente, tem como intuito maximizar a quantidade de informação recebida sobre os desaparecidos na rede de perfis, secundariamente, minimizar a quantidade de informação recebida por perfis sem interesse na informação, logo é a função objetivo é dada pelo somatório do número de pessoas que recebem mensagem mais a soma dos pesos dos influenciadores selecionados para cada desaparecido. As restrições do modelo se encontram na Tabela 3.1.

$$Max \sum_{j \in D} d_j + \sum_{i \in P_i} \sum_{j \in D} W c_{ij} r_{ij} \quad (3.1)$$

$\sum_{i \in P_i} r_{ij} \leq M \quad \forall j \in D$	Para não causar uma inundação na rede um desaparecido j só pode ser divulgado por no máximo M influenciadores. Sendo uma restrição para cada desaparecido em D
$\sum_{j \in D} r_{ij} \leq N \quad \forall i \in P_i$	Para que não acontecer de um influenciador virar um <i>spam</i> um influenciador pode divulgar no máximo N desaparecidos, sendo uma restrição para cada influenciador em P_i
$\sum_{i \in P_i} f_{ik} r_{ij} \geq d_{kj} \quad \forall j \in D, \forall k \in \hat{P}_{(j)}$	Esta restrição tem como intuito determinar se um perfil k de $\hat{P}_{(j)}$ recebe ou não informação sobre o desaparecido j, sendo uma restrição para cada desaparecido em D e para cada perfil em $\hat{P}_{(j)}$
$\sum_{k \in \hat{P}_{(j)}} d_{kj} = d_j \quad \forall j \in D$	A restrição conta quantos perfis do conjunto $\hat{P}_{(j)}$ receberam informação sobre o desaparecido j, sendo uma restrição para cada desaparecido em D
$r_{ij} \in \{0, 1\} \quad \forall j \in D, \forall i \in P_i$	Variável que indica se um desaparecido j será divulgado pelo influenciador i
$d_{kj} \in \{0, 1\} \quad \forall j \in D, \forall k \in \hat{P}_{(j)}$	Variável que indica se um perfil k recebe informação sobre um desaparecido j
$d_j \in \mathbb{R} \quad \forall j \in D$	Variável que indica quantas pessoas receberam informação sobre o desaparecido j

Tabela 3.1: Restrições do modelo

4 Análises

Neste capítulo é tratado sobre as simulações realizadas, como foram feitas, os parâmetros utilizados, e as análises dos resultados nelas obtidas. Além do capítulo também trazer uma discussão sobre como o modelo funcionou.

Com estas discussões, é possível responder a alguns questionamentos relacionados ao problema como: o que acontece com a disseminação de informação se não for utilizado um raio R entre o usuário e o desaparecido? Qual será a importância desse raio em abordagens de disseminação de informação quando o interesse de um nó k em receber ou não a mensagem tem relação com a distância?

Como discutido nas seções anteriores, o esperado é que, quando um influenciador i for escolhido sem a utilização de um raio de influência, uma grande quantidade das mensagens recebidas são enviadas para pessoas as quais não têm interesse, já que o modelo não prioriza as pessoas próximas à região do desaparecido e, dessa forma, vai escolher influenciadores globais e não influenciadores locais.

Utilizando o modelo previamente descrito na Seção 3.2, foram geradas simulações pensando-se em 3 cenários, o primeiro desconsidera as localizações, o segundo utiliza as localizações e o terceiro tenta minimizar os envios para fora do raio.

No total foram geradas 16 simulações sendo estas as permutações dos valores das constantes R e W . Nos três cenários as constantes M e N , que representam a quantidade máxima de influenciadores que podem divulgar um desaparecido j e a quantidade máxima de desaparecidos que pode ser divulgado por um influenciador i , respectivamente, não foram variadas durante as simulações, tendo seus valores fixados em 2 e 5 para a realização do estudo.

A constante W indica o peso dado para a classificação c_{ij} , ou seja, é o peso dado para o segundo critério de escolha, que seria a minimização das mensagens recebidas por perfis fora da área do desaparecido. Nota-se então que quando o valor de W é 0 desconsidera-se a adequação do influenciador i para divulgar o desaparecido j . Já com W nos demais valores utilizados, varia-se a importância dessa adequação na escolha do

influenciador. Sendo com isto possível avaliar se vale apenas perder um pouco do número de mensagem recebida dentro do raio, mas diminuir juntamente o número de mensagens enviadas para público não alvo.

4.1 Cenário 1: sem uso de informação de localização

O primeiro cenário desconsidera as localizações sendo escolhido como influenciadores os usuários que possuem maior número de seguidores, independente se os seguidores estão perto do desaparecido ou não. Com este cenário é possível avaliar a importância da utilização de localização observando se o número de pessoas atingidas que fariam parte do público alvo é grande. Este cenário utiliza apenas o W igual a 0 já que não existem pessoas fora do raio para minimizar.

Nome do Experimento	Valores das constantes
W00R ∞	$W = 0.00$; $R = \infty$

Tabela 4.1: Experimentos cenário 1

Quando é realizado o estudo do cenário 1 obtém-se uma dispersão de 698.236 mensagens. Porém, como mostrado nos gráficos da Figura 4.1, o número de mensagens enviadas para k próximo a j é bem baixo, sendo de aproximadamente 5%, 8% e 14% para usuários considerados perto até 5, 10 e 20km, respectivamente

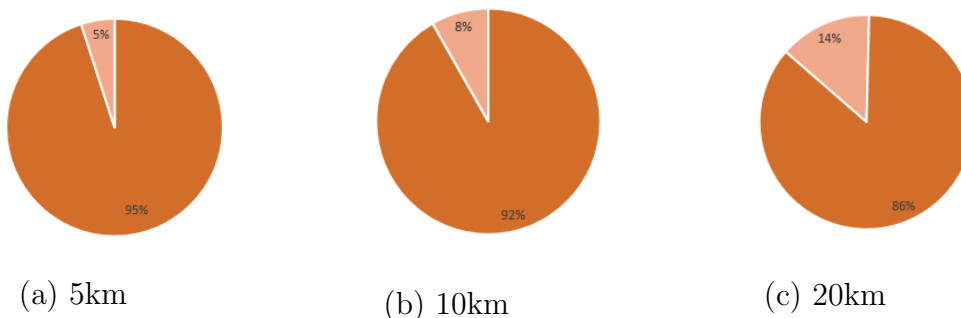


Figura 4.1: Experimento - Cenário 1

4.2 Cenário 2: com uso de informação de localização

O segundo cenário insere-se as localizações, mas deixa os pesos de um influenciador para um desaparecido desligado, ou seja, o valor de W ainda continua como 0. Com isto é possível comparar com o primeiro cenário e gerar análises sobre o quanto a localização influencia.

Nome do Experimento	Valores das constantes
W00R5	$W = 0.00; R = 5$
W00R10	$W = 0.00; R = 10$
W00R20	$W = 0.00; R = 20$

Tabela 4.2: Experimentos cenário 2

Para o raio de cobertura da região do desaparecido foram utilizados os valores de 5km, 10km, 20km. Quando uni-se os resultados do cenário 1 com o cenário 2 é possível comparar os resultados gerados. Com isto é pode-se analisar o quanto o raio influencia neste modelo, já que o intuito é maximizar a informação recebida por usuários perto da área dos desaparecidos.

O gráfico 4.2 mostra um comparativo do número de perfis que receberam as mensagens em cada uma das simulações. A parte azul do gráfico mostra os perfis que estão a uma distância de até 5Km do desaparecido, a parte em laranja perfis que se localizam entre 5 e 10km, a em cinza perfis que se localizam de 10km a 20km e por fim o amarelo representa perfis considerados distantes do desaparecido pelo modelo, ou que estão a mais de 20Km de distância.

É possível perceber que no experimento do cenário 1, o qual o raio de abrangência é desconsiderado, a grande maioria dos perfis que recebem estão longe do desaparecido, logo o interesse no recebimento é pequeno. Quando é se considera um limite de raio, levando-se em conta no modelo a proximidade do perfil que recebe a mensagem do desaparecido o qual está sendo divulgado, o número de mensagens recebidas cai drasticamente, porém as mensagens que chegam são mais direcionadas ao público alvo, ou seja, pessoas da região do desaparecido.

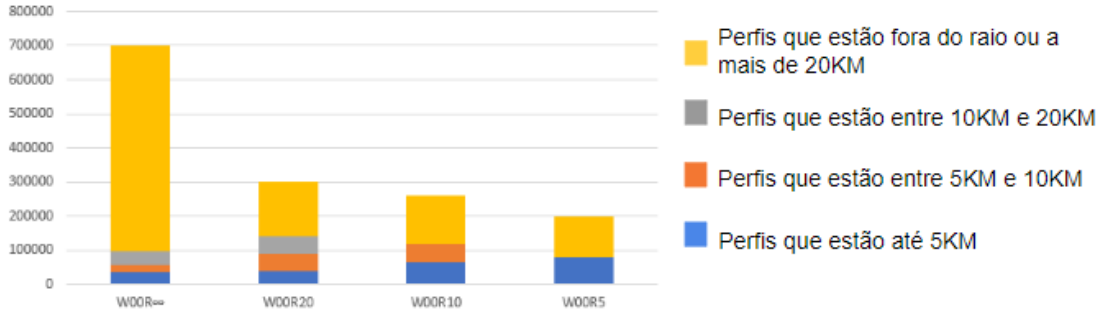


Figura 4.2: Experimentos - Cenário 2

4.3 Cenário 3: minimizando envio para fora do raio

Já o terceiro e último cenário liga o peso, ou seja, liga a constante W . Fazendo ela variar no conjunto $\{0,01; 0,05; 0,1; 0,15\}$. Com isto é possível visualizar se com o peso conseguimos além de alcançar um número alto de pessoas pertencentes ao público alvo conseguimos também minimizar o número de pessoas alcançadas fora do público alvo.

O valor da constante R continua variando nos valores de 5km, 10km e 20km assim como no cenário 2 sendo assim possível comparar os dois cenários para avaliar se vale a pena ligar este peso. Neste cenário o modelo permite que se tenha uma diminuição no valor disperso dentro do raio do desaparecido em relação ao cenário 2. Porém espera-se que seja uma redução pequena em comparação a diminuição da dispersão para usuários fora do raio de abrangência. Neste cenário será considerado cada um dos valores de R um caso diferente a ser analisado.

Nome do Experimento	Valores das constantes
W01R5	$W = 0.01; R = 5$
W05R5	$W = 0.05; R = 5$
W10R5	$W = 0.10; R = 5$
W15R5	$W = 0.15; R = 5$
W01R10	$W = 0.01; R = 10$
W05R10	$W = 0.05; R = 10$
W10R10	$W = 0.10; R = 10$
W15R10	$W = 0.15; R = 10$
W01R20	$W = 0.01; R = 20$
W05R20	$W = 0.05; R = 20$
W10R20	$W = 0.10; R = 20$
W15R20	$W = 0.15; R = 20$

Tabela 4.3: Experimentos cenário 3

Nos gráficos das Figuras 4.3, 4.4 e 4.5 a cada coluna representa o total (100%) de mensagens enviadas, a parte em azul é a porcentagem de dentro da área enquanto a em laranja é a de perfis fora da área. Já as linhas em cinza representam o número de perfis que receberam a mensagem e estavam fora da área enquanto a de amarelo é quem está dentro. Estes experimentos foram considerando o cenário 3.

Neste primeiro caso (5km) é possível observar que quando se utiliza um W acima de 0,1 a quantidade de perfis dentro da área e os fora da área ficam praticamente constantes dando para perceber que a partir daquele ponto a otimização já alcançou praticamente a melhor resposta possível sendo difícil uma otimização muito superior. Além disso dá para observar que quando W igual a zero ainda existe muitos perfis recebendo informações inúteis já que eles não fazem parte do público alvo da mensagem, sendo muito bem otimizado quando se aumenta até 0,1. E o valor de perfis dentro da área que deixam de receber é insignificante.

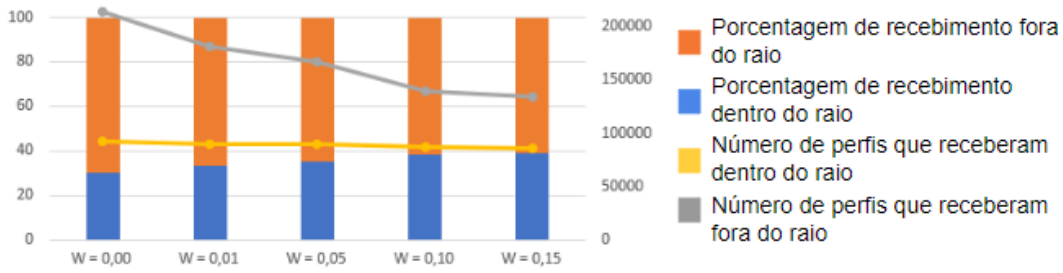


Figura 4.3: Experimentos para 5KM - Cenário 3

Já no segundo caso (10km) é possível perceber que nela ainda existe um caimento bom quando o W chega em 0,15. Como a quantidade de perfis dentro do raio é maior, já que o raio se expandiu, o número de combinações de influenciadores também aumenta, gerando dessa forma mais possibilidades de otimização, além disto ainda temos uma boa quantidade de perfis fora da área.

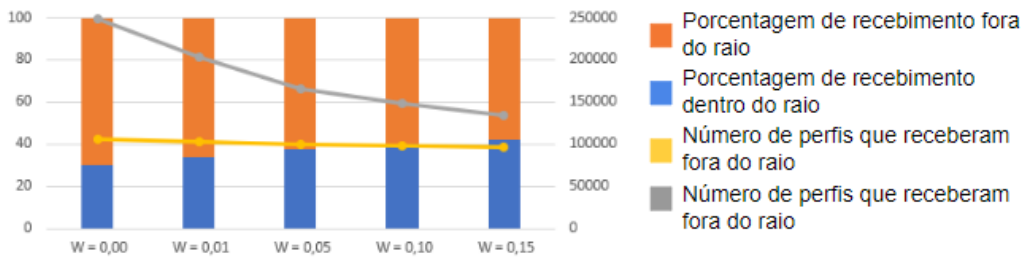


Figura 4.4: Experimentos para 10KM - Cenário - 3

Enquanto isso, no último caso (20km) é possível perceber que a queda de perfis fora da área que recebem a mensagem não é tão grande como nos outros casos, isto provavelmente se deve ao fato de a mediana das distâncias entre os perfis e os desaparecidos ser 30km, ficando assim uma quantidade bem menor de perfis fora do raio que nas outras simulações. Fazendo desta forma que a otimização focada na minimização da quantidade de mensagem recebida fora do raio não seja tão eficiente como nos outros casos.

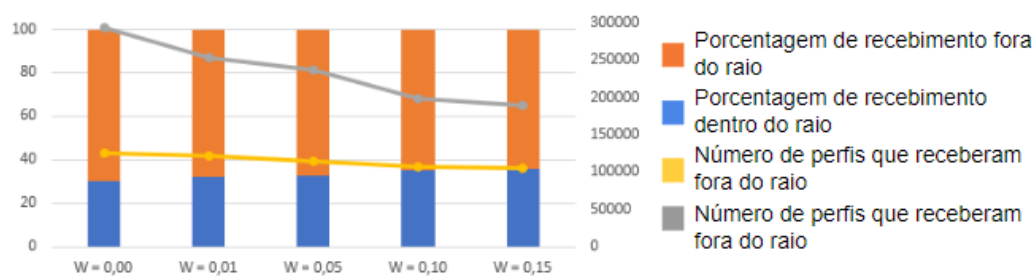


Figura 4.5: Experimentos para 20KM - Cenário 3

5 Conclusão

Como explanado durante todo o trabalho, o problema de desaparecimento de civis afeta diversas famílias ao redor do mundo. Durante as investigações, as primeiras 72 horas são fundamentais para os casos, este é o período em que se tem maior probabilidade de sucesso nas buscas. Com isto, surgem iniciativas para ajudar na aceleração das buscas, sendo uma delas a divulgação de pessoas desaparecidas. A partir das divulgações é possível que os casos cheguem a pessoas que possuam alguma informação e queiram ajudar.

O presente trabalho teve então como objetivo ajudar no problema de desaparecimentos civis a partir da divulgação de pessoas desaparecidas em redes sociais. Encontrando melhores métricas para selecionar influenciadores, com o intuito de obter um maior alcance na divulgação dos desaparecidos para o público alvo utilizando-se de redes sociais.

5.1 Resultados obtidos

Após a realização das simulações foi possível observar a difusão para cada uma das simulações propostas neste trabalho e a relação proporcional de usuários alcançados que pertenciam ou não ao público alvo.

É possível perceber pelos gráficos da Sessão 4.2 que ao utilizar-se de dados de localização as mensagens são enviadas de forma mais precisa. No cenário 1, que é sem raio, o valor de dispersão é bem maior que nos demais casos, porém ao se olhar o público alvo das mensagens ele é mais baixo para os três valores de R considerados. Ao não se utilizar das localizações o público alvo do envio das mensagens é desconsiderado fazendo com que se tenha um valor de dispersão elevado, mas não necessariamente atinge quem deveria ser atingido.

Além disto à medida que se aumenta o W os valores de perfis que recebem a informação e estão dentro do raio cai assim como os que estão fora, porém, a porcentagem de perfis que estão dentro do raio cresce o que é interessante para o problema. Isto ocorre

pois quando o valor de W é zero, a escolha dos influenciadores privilegia aqueles com maior número de seguidores dentro do raio. À medida que W aumenta, influenciadores com maioria de seguidores fora do raio perde um pouco sua importância no modelo. Assim o modelo permite que haja um valor um pouco menor de pessoas atingidas dentro do raio, para permitir uma diminuição significativa de pessoas que recebem a mensagem fora do raio.

Pensando que o envio de uma mensagem a pessoas não interessadas causa uma inundação da rede não desejada, ao se escolher um influenciador que tem uma boa influência na região, mas um percentual muito grande dos seus seguidores é de fora e tendo uma segunda escolha que possui somente alguns seguidores dentro do raio a menos, mas que possui muitos seguidores de fora a menos o custo benefício é melhor. Conseguindo assim diminuir o número de mensagens para pessoas não interessadas continuando com praticamente a mesma quantidade de perfis recebendo a mensagem dentro do raio

Ao observar os resultados é possível deduzir que quando o importante para quem quer fazer a divulgação é alcançar o maior número de pessoas sem se preocupar com o envio de mensagens para pessoas fora do público alvo, então utilizar-se de um raio maior é mais vantajoso. Porém no problema aqui discutido em questão o objetivo é alcançar o maior número de pessoas enviando o mínimo possível de mensagens. Com isto tem-se que é interessante um equilíbrio com o valor de raio 10 quilômetros, já que entrega para bastante perfis dentro do raio, mas ainda sim conseguindo minimizar os envios para fora do público alvo.

O estudo gerou resultados satisfatórios, sendo possível concluir que existe um equilíbrio entre a maximização de recebimento de informação por perfis que estão no máximo a uma distância R do desaparecido e a minimização de recebimento de informação por perfis não alvos, possibilitando, dessa forma que uma grande quantidade de perfis interessados receba, mas sem que se tenha um número muito alto de mensagens enviadas para perfis não interessados.

É possível também concluir que a utilização de uma abordagem utilizando localizações é interessante para este problema. Isto porque quando não se utiliza a localização mais mensagens são enviadas, mas menos usuários de interesse a recebem, o que gera um

maior esforço desnecessário de envio da mensagem.

Os resultados mostram que valores muito baixos de raio causam uma estagnação mais rápida, na otimização enquanto quando se aumenta o valor do raio para perto da mediana de distância, o segundo critério, não possui tanto impacto como nos demais casos.

5.2 Limitações e Trabalhos futuros

Apesar do trabalho ter sucesso no que se propôs a responder, ainda existem pontos que podem ser melhorados. A base de perfis utilizada para as simulações possui diversas localizações para cada usuário. Neste trabalho é escolhida uma localização para cada um. A escolha por outra localização pode afetar os resultados, deixando a base melhor ou pior para as análises. Já que o modelo utiliza a distância do perfil do desaparecido para escolher outra localização, muda-se a configuração da rede.

Além disso, o trabalho foi todo produzido utilizando-se de simulações com uma rede social voltada à coleta de *check-in*, ou seja, foi utilizada uma base que facilitou o estudo com as localizações dos perfis, uma ideia seria utilizar a localização com a hora mais próxima do desaparecido. Outras redes sociais podem influenciar nos resultados, sendo dessa forma necessário outros experimentos nestes cenários.

Outro ponto que pode ser explorado é a variação das constantes M e N , as quais não foram trabalhadas durante o estudo. A modificação destas constantes implica na quantidade de informação divulgada, já que M limita a quantidade de influenciadores que pode divulgar um desaparecido e N limita a quantidade de desaparecidos que um influenciador pode divulgar.

Para trabalhos futuros é interessante fazer uma variação nas constantes M e N . Pode-se realizar experimentos com dados reais, já que os testes deste trabalho foram todos simulados fazendo, desta forma, uma comparação com a literatura e prática no mundo real. Por último, seria possível utilizar uma nova métrica para selecionar qual a posição do perfil. Neste trabalho foi utilizada a primeira localização, mas não necessariamente essa localização representa a localização mais adequada para o perfil.

Sendo assim, os resultados aqui apresentados demonstram um caminho de experimentação para divulgar informação sobre desaparecidos, mas experimentos com dados

reais e mais volumosos são necessários.

Bibliografia

A&E BRASIL. *As primeiras 72 horas são as mais críticas para pessoas desaparecidas; entenda o porquê*. 2021. Disponível em: <https://br.aeplay.tv/noticia/primeiras-72-horas-sao-mais-criticas-para-pessoas-desaparecidas-entenda-o-porque>.

ALI, K.; LI, C.-T.; CHEN, Y.-S. Joint selection of influential users and locations under target region in location-based social networks. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 21, n. 3, p. 709, 2021.

AYRES, D. *Foto em conta de luz ajuda polícia a achar criança desaparecida em MG*. 2014. Acessado em: 2021-12-10. Disponível em: <http://g1.globo.com/mg/sul-de-minas/noticia/2017/01/crianca-desaparecida-em-mg-e-reconhecida-por-colega-de-escola.html>.

BOUROS, P.; SACHARIDIS, D.; BIKAKIS, N. Regionally influential users in location-aware social networks. In: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. [S.l.: s.n.], 2014. p. 501–504.

BRASIL, L. M.; SILVA, R. S. “de volta pra casa”: Portal de prevenção e divulgação de perfis de desaparecidos. 2015.

CALMON, M. Forensic anthropology and missing persons: A brazilian perspective. *Forensic science international*, Elsevier, v. 298, p. 425–e1, 2019.

CLASSE, T. et al. Projeto encontre-me! tics para a divulgação e busca de pessoas desaparecidas de maneira colaborativa. In: SBC. *Anais do XV Simpósio Brasileiro de Sistemas Colaborativos*. [S.l.], 2019. p. 18–23.

EVISON, M. P. et al. Forensic facial reconstruction and its contribution to identification in missing person cases. In: *Handbook of Missing Persons*. [S.l.]: Springer, 2016. p. 427–441.

HIRSCHEL, J. D.; LAB, S. P. Who is missing? the realities of the missing persons problem. *Journal of Criminal Justice*, Elsevier, v. 16, n. 1, p. 35–45, 1988.

JR, J. G.; FERRANTI, N.; SOUZA, J. F. de. Semantic enrichment of web data for the provision of an unified data repository of brazilian missing persons. In: *Proceedings of the XV Brazilian Symposium on Information Systems*. [S.l.: s.n.], 2019. p. 1–8.

KEMPE, D.; KLEINBERG, J.; TARDOS, É. Maximizing the spread of influence through a social network. In: ACM. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2003. p. 137–146.

LAMPINEN, J. M.; ARNAL, J.; HICKS, J. L. The effectiveness of supermarket posters in helping to find missing children. *Journal of interpersonal violence*, Sage Publications Sage CA: Los Angeles, CA, v. 24, n. 3, p. 406–423, 2009.

LOU, C.; YUAN, S. Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, Taylor & Francis, v. 19, n. 1, p. 58–73, 2019.

- MISSING PEOPLE. *Stats*. 2021. Disponível em: <https://www.missingpeople.org.uk/for-professionals/information-and-policy/information-and-research/key-information>.
- MOREWITZ, S. J.; COLLS, C. S. *Handbook of missing persons*. [S.l.]: Springer, 2016.
- MUYAMBO, P. An investigation on the use of lbph algorithm for face recognition to find missing people in zimbabwe. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2018.
- OLIVEIRA, D. D. d. Desaparecidos civis: conflitos familiares, institucionais e segurança pública. 2007.
- OLIVEIRA, D. D. de. *O desaparecimento de pessoas no Brasil*. [S.l.]: Cênone Editoração Ltda, 2014.
- PARANAIBA, G. *Polícia busca famílias de 24 desaparecidos de Brumadinho para coletar DNA*. 2019. Acessado em: 2021-12-10. Disponível em: https://www.em.com.br/app/noticia/gerais/2019/03/30/interna_gerais,1042399/policia-busca-familias-de-24-desaparecidos-de-brumadinho-para-dna.shtml.
- ROGERS, E. M. *Diffusion of innovations*. [S.l.]: Simon and Schuster, 2010.
- SANTOS, B. Malacrida dos; ARTERO, A. O. Efeitos especiais em computação gráfica–morphing. In: *Colloquium Exactarum*. [S.l.: s.n.], 2011. v. 3, n. 2.
- SILVA, T. T. de O. Descoberta de influenciadores no twitter para ampliação do alcance de informações sobre desaparecidos no brasil. 2019.
- UOL. *Cerca de 80 mil pessoas desaparecem a cada ano no Brasil, aponta levantamento*. 2021. Acessado em: 2021-12-02. Disponível em: https://cultura.uol.com.br/noticias/29118_cerca-de-80-mil-pessoas-desaparecem-a-cada-ano-no-brasil-aponta-levantamento.html.
- WANG, W.; STREET, W. N. Modeling and maximizing influence diffusion in social networks for viral marketing. *Applied network science*, SpringerOpen, v. 3, n. 1, p. 1–26, 2018.