



Descoberta de influenciadores no *Twitter*
para ampliação do alcance de informações

sobre desaparecidos no Brasil
Versão Completa através de:
coord.computacao@ice.ufjf.br

Thalita Thamires de Oliveira Silva

Descoberta de influenciadores no *Twitter*
para ampliação do alcance de informações
sobre desaparecidos no Brasil
Versão Completa através de:
coord.computacao@ice.ufjf.br

THALITA THAMIREs DE OLIVEIRA SILVA

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza
Coorientador: Lorenza Leao Oliveira Moreno

JUIZ DE FORA
NOVEMBRO, 2019

DESCOBERTA DE INFLUENCIADORES NO *Twitter* PARA
AMPLIAÇÃO DO ALCANCE DE INFORMAÇÕES SOBRE
DESAPARECIDOS NO BRASIL
VERSÃO COMPLETA ATRAVÉS DE:
COORD.COMPUTACAO@ICE.UFJF.BR

Thalita Thamires de Oliveira Silva

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

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

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

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

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

JUIZ DE FORA
28 DE NOVEMBRO, 2019

Aos meus pais, pelo apoio e sustento.

Resumo

Casos de desaparecimento de pessoas no Brasil possuem um número elevado, 82.684 boletins de ocorrência registrados, segundo Anuário Brasileiro de Segurança Pública 2018. Porém, o cenário pode ser ainda mais grave, uma vez que nem todos os desaparecimentos são notificados. Devido a esse alto número e falta de locais centralizados sobre informações de desaparecimentos no Brasil, surge um problema no que tange a divulgação de pessoas desaparecidas. A Internet a cada dia ganha uma posição cada vez mais alta no processo de difusão de informações. Dentre as várias opções que Internet oferece, as redes sociais têm atraído muita atenção. O presente trabalho tem como objetivo construir um sistema, com auxílio de modelos de Programação Linear, com o propósito de otimizar a difusão de informação sobre desaparecimentos no Brasil no *Twitter*. O modelo construído tem como objetivos encontrar influenciadores para cada pessoa desaparecida, maximizar a difusão da informação e minimizar o tamanho do grupo de influenciadores. **O texto completo está em sigilo e pode ser requisitado através de:** coord.computacao@ice.ufjf.br

Palavras-chave: Desaparecimentos, Pessoas desaparecidas, Difusão de informação, Programação Linear, Otimização.

Abstract

Cases of missing persons in Brazil have a high number, 82,684 reported bulletins, according to the Brazilian Public Security Yearbook 2018. However, the scenario can be even more serious, since not all disappearances are reported. Due to the high number of cases and the lack of a centralized information system on missing persons in Brazil, the problem regarding the disclosure of missing persons arises. Everyday The Internet gains a higher and higher position in the process of information dissemination. Among the many options that the Internet offers, social networking has attracted a lot of attention. The present work aims to build a system, with the aid of Linear Programming models, with the purpose of optimizing the dissemination of information about disappearances in Brazil on Twitter. The built model aims to find influencers for each missing person, maximize the diffusion of information and minimize the size of the influencer groups. **The full text is confidential and may be requested through:** coord.computacao@ice.ufjf.br

Keywords: Disappearances, Missing Persons, Information Dissemination, Linear Programming, Optimization.

Agradecimentos

A todos os meus familiares e amigos, pelo encorajamento e apoio.

Aos professores Jairo e Lorenza, pela orientação, incentivo e dedicação para a realização deste trabalho.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

A vitória de um homem às vezes se esconde num gesto forte que só ele pode ver.

O Rappa

Sumário

Lista de Figuras	7
Lista de Tabelas	8
Lista de Quadros	9
Lista de Abreviações	10
1 Resumo Estendido	11
1.1 Justificativa	12
1.2 Metodologia	13
1.2.1 Caracterização da pesquisa	13
1.2.2 Delimitação do universo	13
1.3 Otimização da divulgação de informações de desaparecidos	14
1.3.1 Base de Dados	14
1.3.2 Definição do Problema	16
1.3.3 Modelo Matemático	17
2 Conclusão	20
Referências Bibliográficas	22

Lista de Figuras

Lista de Tabelas

Lista de Quadros

1.1	Variáveis do modelo	18
-----	-------------------------------	----

Lista de Abreviações

CICV Comitê Internacional da Cruz Vermelha

ONGs Organizações não Governamentais

OFT Teoria Ideal de Forrageamento

1 Resumo Estendido

Segundo o Anuário Brasileiro de Segurança Pública 2018 (PÚBLICA, 2018), foram registrados 82.684 boletins de ocorrência informando o desaparecimento de pessoas em todo país. Porém, o cenário pode ser ainda mais grave, uma vez que nem todos os desaparecimentos são notificados, seja por falta de conhecimento ou por despreparo policial e/ou legislativo. Para o Comitê Internacional da Cruz Vermelha (CICV), “pessoas desaparecidas” são todos os indivíduos cujo paradeiro é desconhecido pelas suas famílias ou que foram dados como desaparecidos, segundo fontes fidedignas, devido a um conflito armado, violência interna, desastre natural ou outras crises humanitária. Nesse trabalho é utilizado essa definição. Os motivos dos desaparecimentos são variados, como violência doméstica, envolvimento com drogas, venda de órgãos, trabalho escravo, prostituição, dentre outros.

A pessoa desaparecida é a primeira vítima, mas essa tragédia também afeta muitas outras. A família ou os conhecidos do desaparecido enfrentam uma situação de angústia e impotência. Estes tornam-se reféns, uma vez que ficam dependentes da boa vontade das delegacias, da repercussão do caso na mídia ou da influência a partir de relacionamentos pessoais. Assim, um problema surge no que tange a divulgação de desaparecimentos. Atualmente, o Brasil não possui política pública específica para tratar do tema das pessoas desaparecidas e seus familiares, porém existem iniciativas governamentais com o objetivo de localização de pessoas desaparecidas, como o registro online, disponibilizado no site (<https://desaparecidos.mj.gov.br>), lançado em 2010, que tinha como objetivo criar um registro central de desaparecidos no Brasil. Porém, o site encontrou-se indisponível por longos períodos e possui uma baixa popularidade.

Em geral, a divulgação de desaparecimentos é realizada em outros sites do governo ou ONGs brasileiras. Porém a propagação das informações fica restrita ao acesso direto a esses sites.

1.1 Justificativa

A Internet está desempenhando um papel cada vez mais central no processo de propagação da informação, devido à sua ampla adoção em todo o mundo. Entre as várias opções que a internet disponibiliza, as redes sociais têm atraído muita atenção. Estas possuem um crescente número de usuários, que almejam conquistar uma grande influência nesses espaços. As redes sociais fornecem uma plataforma para os usuários construírem ou manterem relacionamentos com outros usuários.

De acordo com (KEMPE; KLEINBERG; TARDOS, 2003), redes sociais desempenham um papel fundamental durante o processo de disseminação de informação, ideias e influências dentro da rede de relacionamento criada através da interação entre os usuários nesse ambiente. Em um local no qual usuários encontram-se em constante interação, compartilhando assuntos de interesse comum, é esperado que as informações ali discutidas facilmente se propaguem entre os usuários de uma rede social. Exemplos do poder que as redes sociais possuem sobre a disseminação de informação atualmente, seria a utilização destas como canais de comunicação de presidentes, como Trump (presidente dos EUA, eleito nas eleições de 2016) e Jair Bolsonaro (presidente do Brasil, eleito na eleições de 2018); O uso do *Facebook* durante as campanhas eleitorais dos EUA em 2016; Outro exemplo ainda, seria a mobilização de pessoas para manifestações no Brasil a favor ou contra o governo regente durante os anos de 2018 e 2019.

Muitos casos de desaparecimento permanecem sem solução por um longo período de tempo ou, às vezes, jamais são resolvidos. Por isso, em virtude da sua gravidade, o desaparecimento de pessoas é um fenômeno importante que precisa ser estudado.

Um dos problemas que dificulta encontrar desaparecidos no Brasil é a divulgação dos dados desse fato de forma não efetiva. Para alcançar maior visualização das informações, uma opção seria utilizar as redes sociais, já que estas possuem um importante papel na atualidade na difusão de informações.

1.2 Metodologia

A metodologia utilizada no trabalho envolve as etapas: (1) construção do modelo para encontrar um grupo de perfis (pessoas) que, a partir destes, a informação é difundida, chamaremos esse grupo de influenciadores; (2) difusão da informação; e (3) análise dos resultados obtidos.

O modelo construído com auxílio de técnicas de programação linear e tem como objetivos encontrar influenciadores para cada pessoa desaparecida, maximizar a difusão da informação e minimizar o tamanho do grupo de influenciadores. Na abordagem do presente trabalho, a métrica adotada no cálculo do peso, para caracterizar um perfil como influenciador, considera o número de seguidores, que este possui, e a que região estes pertencem, aspirando assim encontrar influenciadores regionais.

Para realizar a análise da difusão, citada anteriormente, foi utilizada uma abordagem de modificação de parâmetros e bases. Para assim visualizar os diferentes resultados.

1.2.1 Caracterização da pesquisa

O estudo se configura numa abordagem quantitativa, do ponto de vista da forma de abordagem do problema, já que haverá análises estatísticas dos resultados obtidos. Quanto à natureza, esta é uma pesquisa aplicada, visto a necessidade da utilização de requisitos específicos apresentados nas áreas que envolvem difusão de informação e programação linear.

1.2.2 Delimitação do universo

Existem três grupos de perfis de pessoas: desaparecidos, candidatos a influenciadores e influenciadores. Os perfis de desaparecidos foram coletados de vários sites que disponibilizam informações de pessoas desaparecidas no Brasil, sejam órgãos públicos ou não. Depois dessa coleta houve uma normalização e pré-processamento dos dados. Os perfis de candidatos a influenciadores são perfis do *Twitter* que foram extraídos no ano de 2018. Por fim, a resposta obtida pelo modelo fornece o terceiro grupo, os influenciadores.

1.3 Otimização da divulgação de informações de desaparecidos

A proposta do presente trabalho é dar grande visibilidade a um conjunto de informações sobre desaparecidos, e para atingi-la optou-se por utilizar como rede social o *Twitter*. Porém, enviar esse conjunto para todos os perfis possíveis não seria uma abordagem viável. Assim, esse trabalho irá tratar do problema de maximização da influência, que consiste em encontrar os perfis que mais influenciam a rede. Os fatores que são considerados importantes, neste trabalho, para determinar se um perfil é influente são: local que o perfil se encontra, quantidade de seguidores deste e o local que esses seguidores se encontram.

Analisando a divulgação de informação, acredita-se que é mais vantajoso divulgar um desaparecimento às pessoas que estão próximas ao local do evento, pois as pessoas mais próximas teriam uma maior probabilidade de divulgar a informação sobre alguém da mesma região, por um senso de comunidade, além de existir uma questão emotiva envolvida, visto que o perfil influenciado saber do desaparecimento de uma pessoa de sua região tende a fazê-lo se interessar mais pela informação divulgada. Como resultado, a probabilidade deste difundir a informação é maior, o que conseqüentemente aumenta as chances de encontrar a pessoa desaparecida. Sendo assim, no presente trabalho foi considerado um fator externo à rede, a distância real. Como já foi exposto por (CONGRAM; KENYHERCZ; GREEN, 2017), fatores espaciais tem relevância para busca de pessoas.

1.3.1 Base de Dados

A base de desaparecidos que utilizada nesse trabalho foi adquirida a partir do trabalho de (GOMES JR.; FERRANTI; SOUZA, 2019). Nesse trabalho, foi construída uma base de dados centralizada de pessoas desaparecidas no Brasil. Essa base foi construída a partir da coleta de vários sites disponíveis na web, além de ter sido enriquecida através de um *framework* também desenvolvido no trabalho. Este *framework* funciona da seguinte forma: é feita uma coleta de informações sobre desaparecidos em sites e, logo depois dessa coleta, as informações passam por várias funções de normalização e/ou atribuição de metadados. Funções para atribuição de metadados é um tipo de função que recebe um

certo dado bruto e faz uso de bases de dados externas ou de ferramentas para encontrar informações adicionais a serem incluídas na base. Um exemplo de função de atribuição de metadados é a função para complementação de informação de sexo. A função utiliza uma ferramenta externa, uma API do IBGE (Instituto Brasileiro de Geografia e Estatística) que recebe o nome de uma pessoa e retorna o sexo provável dessa pessoa. Todas essas informações são armazenadas em um banco de dados não relacional. Essa base foi coletada durante o ano de 2018 e contém 10499 desaparecidos. Entretanto, apenas 7352 deles possuíam metadados correspondentes aos estados em que os desaparecimentos ocorreram. A distribuição de desaparecidos por estado pode ser visualizada na Tabela ??.

Para o presente modelo, as informações que serão utilizadas desse conjunto são: (1) o *id* de cada desaparecido, que será sua identificação no modelo, (2) a latitude e (3) a longitude do desaparecido, que foi calculada a partir da informação de cidade destes. As demais informações de cada um poderão ser utilizadas em uma etapa posterior de envio prático de informação para cada um dos perfis influenciadores, para estes efetivamente fazerem a divulgação em seus perfis no *Twitter*.

A base de perfis do *Twitter* também foi coletada durante o ano de 2018. Para essa coleta foi utilizado um critério de seleção. Foram pesquisadas na rede social *hashtags*, *keywords*, *posts* e perfis que realizassem alguma difusão de informação sobre desaparecimentos. Esse critério aumenta a possibilidade de re-compartilhamento da informação caso o perfil seja selecionado. Por conta da base de desaparecidos só contar com desaparecimentos ocorridos no Brasil, foram realizadas filtragem aos perfis que continham apenas informação relacionadas ao Brasil. Dessa pré-análise, foram selecionadas 16 entidades centrais que faziam o papel de divulgação de informação sobre desaparecimento. Definido o nível inicial da rede, foram coletados todos os dados de seguidores dois níveis abaixo. Ou seja, as 16 entidades centrais compõem o nível 0, seus seguidores o nível 1 e o nível 2 é formado pelos seguidores dos perfis do nível 1. Todos os perfis coletados que não tivessem postagem em 2018, não possuíssem dados de localização, não fossem brasileiros ou fossem perfis padrões foram retirados da base. As informações coletadas de cada perfil foram: (1) quantidade de seguidores, (2) data do último *tweet*, (3) cidade, (4) país, (5) perfil padrão e os (6) relacionamentos com outros perfis.

A base total de perfis do *Twitter* possui 157.823 perfis sendo que as informações que foram utilizadas no modelo são: (1) nome do perfil, (2) latitude e (3) longitude calculada a partir da cidade, estado ou país contido nas informações deste, além (4) do peso de influência sobre cada seguidor. Esse peso entre os perfis foi calculado a partir da média de atividade do perfil e de seu seguidor. A atividade de um perfil é a diferença de tempo entre o primeiro e o vigésimo *tweet* na *timeline* de perfil.

1.3.2 Definição do Problema

O problema de otimização abordado neste trabalho consiste em selecionar um grupo restrito de influenciadores, para quem as informações de um ou mais desaparecidos serão passadas.

O objetivo é maximizar a divulgação dessas informações na região do desaparecido. Para concentrar a divulgação dos dados para usuários próximos ao local de desaparecimento, foi levada em consideração a localidade dos perfis e dos desaparecidos. Seja D o conjunto de desaparecidos e $j \in D$ um índice que representa um desaparecido

Para concentrar a divulgação dos dados para usuários próximos ao local de desaparecimento, foi levada em consideração a localidade dos perfis e dos desaparecidos. Para representar esse fato, para cada desaparecido foi definido um raio. Dentro da área deste raio estão localizados os perfis alvos para receberem as informações do desaparecido. Esse conjunto de perfis aqui é denominado P_j e $k \in P_j$ como um destes perfis.

Além disso para quantificar a importância de cada perfil para cada desaparecido foi considerado a distância real entre eles. Distância real é calculada a partir das coordenadas geográficas de cada um. Essa quantificação foi utilizada no conjunto de constantes g_{kj} , onde k representa o índice de um perfil e j o índice de um desaparecido.

Essas constantes, g_{kj} , são o ponto do modelo que carregam a influência de fatores externos. Para favorecer a importância dos perfis próximos aos desaparecidos foi necessário inverter o valor da distância, ou seja, o perfil que estiver mais próximo terá uma importância maior. Uma solução para inverter esse valor seria dividir 1 pelo valor da distância, é uma solução correta, que retorna os valores com as importâncias proporcionalmente desejáveis. Porém existem distâncias com valores muito altos, o que acarretaria

em constantes com valores muito pequenos. Assim, para o modelo trabalhar com valores maiores, a inversão da distância foi feita a partir da divisão de 5.000.000 pela distância real entre desaparecido e perfil. Esse valor foi escolhido, pois a maior distância entre dois pontos dentro do Brasil é de 4.394.000 metros, assim foi selecionado um valor inteiro que garante que todos os valores de g_{kj} são maiores que 0. Esse conjunto de perfis influenciadores, perfis que divulgam informação de algum desaparecido, está contido no conjunto I_j , onde $i \in I_j$ como um índice representando um influenciador. Todos os perfis podem ou não influenciar outro perfil. Essa influência é calculada analisando a rede de perfis que estes seguem, diretamente ou não. Para representar o valor dessa influência foi construído o conjunto de constantes f_{ik} .

f_{ik} é calculada para os pares de perfis, através de uma árvore de influência gerada para cada perfil i . Essa árvore funciona da seguinte maneira: i é o nó raiz dessa árvore, no próximo nível estão todos os perfis que seguem o perfil i , e no próximo todos que seguem os perfis do nível anterior. O peso máximo de uma arco é 1, sendo que o peso dos arcos vão decrescendo quanto mais baixo na árvore o nó está situado. Para delimitar os seguidores mais importantes para cada perfil, houve um corte nas árvores. Quando o valor do arco é menor que 0.8 a construção da árvore é terminada.

É importante ressaltar que pode haver interseção entre P_j e I_j , isto é, um mesmo perfil pode ser influenciado e influenciador, ou seja, $f_{ik} = 1$ quando $i = k$;

Para resolver o problema, com essas considerações, serão utilizadas técnicas de programação linear, sendo necessária a construção de um modelo matemático que represente o problema. Em (HILLIER, 2012) o modelo matemático de um problema é definido como o sistema de equações e de expressões matemáticas relativas que descrevem sua essência.

1.3.3 Modelo Matemático

Conforme a definição do problema na seção anterior, um modelo matemático será utilizado (1) para definir que influenciadores serão selecionados para a divulgação da informação a respeito de cada desaparecimento j e (2) para quantificar (e maximizar) a divulgação destas informações nas respectivas regiões de cada desaparecimento. A Tabela 1.1 apresenta

as variáveis de decisão do modelo.

Variável	Significado	Domínio
r_{ij}	Representa se o perfil influenciador i divulga informação sobre o desaparecido j	$r_{ij} \in \{0, 1\}$
d_j	Representa a quantidade total de informação dispersa sobre o desaparecido j	$d_j \geq 0$
d_{jk}	Representa a quantidade de informação que o perfil k recebe sobre o desaparecido j	$d_{jk} \in [0, 1]$

Quadro 1.1: Variáveis do modelo

Para criar as restrições desse modelo foi necessário o uso de duas constantes: f_{ik} , que representa a importância da força de influência que o perfil i tem sobre k , e g_{kj} , que representa importância do perfil k em relação ao desaparecido j .

A função objetivo é definida pela Equação 1.1. Essa equação tem como propósito maximizar a soma da quantidade de informação dispersa sobre cada desaparecido na rede. Ou seja, a equação tem como propósito maximizar o total de informação dispersa sobre o conjunto de desaparecidos.

$$\text{Max} \sum_{j \in D} d_j \quad (1.1)$$

As restrições são:

$$\sum_{i \in P} r_{ij} \leq 1 \quad \forall j \in D \quad (1.2)$$

As restrições 1.2 garantem que todos os desaparecidos são divulgados por no máximo um perfil. O intuito dessa restrição é limitar por quantos perfis um desaparecido será divulgado, assim o modelo diminui a possibilidade de sobreposição de influências. Sobreposição de influências seria o caso onde os conjuntos de perfis que são influenciados, por cada indivíduo da rede, possuam uma interseção não vazia, assim alguns dos perfis contidos na interseção poderiam receber a informação do mesmo desaparecido, através de perfis diferentes e estas informações serem contabilizadas como informações diferentes.

$$\sum_{j \in D} r_{ij} \leq N \quad \forall i \in P \quad (1.3)$$

As restrições 1.3 garante que cada perfil divulguem no máximo N desaparecidos. Essa restrição tem como propósito limitar o número de desaparecidos que cada perfil pode divulgar, assim não há a possibilidade de sobrecarregamento de perfis da rede, algo que seria possível sem essa restrição, pois um dos objetivos do modelo é encontrar o menor grupo de influenciadores.

$$\sum_{i \in P} f_{ik} r_{ij} \geq d_{kj} \quad \forall j \in D \quad \forall k \in P \quad (1.4)$$

As restrições 1.4 tem como objetivo quantificar a informação recebida sobre cada desaparecido, e essa quantificação é feita pela soma das divulgações realizadas pelos perfis influenciadores, multiplicadas pelas influências entre os perfis envolvidos. Assim essa restrição garante que a quantidade de informação recebida é proporcional a quantidade de informação difundida. Logo é essa restrição que faz o modelo tender a selecionar os perfis mais influentes.

$$\sum_{k \in P} g_{kj} d_{kj} = d_j \quad \forall j \in D \quad (1.5)$$

As restrições 1.5 tem como intuito quantificar a informação difundida sobre cada desaparecido. A quantificação é feita pela soma da quantidade de informação recebida multiplicada pela posição invertida de cada perfil. Ou seja, é nessa restrição que o modelo pretende garantir que os perfis que receberão as informações estarão próximos ao desaparecido.

Como foi apresentado no presente capítulo, a programação linear será utilizada para resolver o problema de maximização da influência, através da solução do modelo linear definido por 1.1 a 1.5. O modelo proposto utiliza dois fatores de grande importância no presente trabalho, a influência e a localidade dos perfis. O objetivo final é (1) a seleção de perfis influentes, o que traz como consequência uma maior dispersão de informação na rede, e (2) que os perfis influenciados estejam próximos a cidade do desaparecido, resultando em uma maior probabilidade do desaparecido ser encontrado, pois mais pessoas da região estarão cientes do desaparecimento.

2 Conclusão

O presente trabalho teve como objetivo otimizar a difusão de informação sobre desaparecimentos no Brasil utilizando o *Twitter* levando em consideração a localidade dos perfis e dos desaparecidos como característica decisória no processo de seleção de influenciadores. Através da realização dos experimentos foi possível concluir que o trabalho obteve resultados satisfatórios na seleção de perfis influenciadores, levando em consideração as regiões dos desaparecidos e dos perfis. Além disso, o trabalho apresentou uma solução funcional para o problema da divulgação de desaparecimentos no Brasil.

Apesar do trabalho obter sucesso nos objetivos traçados, ainda existem pontos de melhorias. A quantidade de informação dispersa não está intrinsecamente relacionada com o número de perfis influenciados é um ponto. No modelo utilizado, o objetivo é maximizar o quantidade de informação dispersa. Porém, a quantidade de informação dispersa é calculada levando em consideração a distância entre perfil e desaparecido, além do valor de influência entre os perfis. Um ponto de melhora seria considerar também o número de perfis atingidos e o tempo para geração de resultados de instâncias maiores.

Encontrar pessoas desaparecidas é considerado um problema no Brasil e no mundo. Falta de informações, agilidade e ampla divulgação são alguns dos elementos que dificultam descobrir o paradeiro dessas pessoas. Existem milhares de casos que nunca foram resolvidos. O presente trabalho apresenta uma contribuição para o problema mostrando que uso de redes sociais pode ser uma fonte de divulgação de dados que pode ajudar na solução deste. Contudo, há outros pontos que devem ser atingidos para um melhor resultado, como: (1) a conscientização dos cidadãos brasileiros (com mais pessoas conscientes desse problema um maior número de pessoas irão tender a divulgar informações de desaparecimentos, e aumentar o conjunto de candidatos a influenciadores); (2) Extração de bases de dados mais fiel, pois apesar de alguns perfis no *Twitter* conterem localização, essa localização pode não ser verídica (durante o trabalho foram descartados perfis por possuírem localização fictícia); (3) Medidas sociais feitas pelo estado, utilizando outros canais de comunicação, como TV, SMS, entre outros.

Houve limitações tecnológicas para não realizar os experimentos com a base de dados do Brasil. Como o número de perfis da base é muito alto, 157.823, é necessário a criação de um conjunto proporcionalmente grande de variáveis. Isso acarreta em grande uso de memória e tempos elevados para geração dos resultados. Foram feitas diversas otimizações no código com o propósito de reduzir principalmente o uso de memória, pois extrapolando a capacidade de memória não é possível obter nenhum resultado. Após estas otimizações, aparentemente é possível obter resultados para uma instância maior que possui 14 mil perfis. Porém, o tempo de processamento é elevado, não viabilizando a realização dos experimentos para o presente trabalho.

Trabalhos futuros podem realizar os experimentos em bases maiores, como a do Brasil e além disso, de fato construir um sistema para o envio dessa informações aos perfis influenciadores no *Twitter*. Assim, seria possível uma atuação prática para resolução do problema de divulgação. Além disso, seria possível analisar se com a divulgação houve um maior número de pessoas desaparecidas encontradas, e também encontrar possíveis ajustes que devem ser feitos no trabalho para aplicação no mundo real.

Referências Bibliográficas

CONGRAM, D.; KENYHERCZ, M.; GREEN, A. G. Grave mapping in support of the search for missing persons in conflict contexts. *Forensic science international*, Elsevier, v. 278, p. 260–268, 2017.

GOMES JR., J. a.; 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*. New York, NY, USA: ACM, 2019. (SBSI'19), p. 57:1–57:8. ISBN 978-1-4503-7237-4.

HILLIER, F. S. *Introduction to operations research*. [S.l.]: Tata McGraw-Hill Education, 2012.

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.

PÚBLICA, F. B. de S. 12º anuário brasileiro de segurança pública. *Fórum Brasileiro de Segurança Pública*. Disponível em <http://www.forumseguranca.org.br/wp-content/uploads/2019/03/Anuario-Brasileiro-de-Seguranca-Publica-2018.pdf>. Acesso em Nov. 2019, 2018.