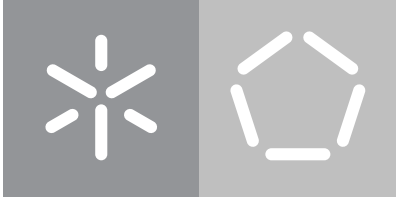
Universidade do Minho
Escola de Engenharia

João Lobarinhas Fernandes Miranda

Cálculo do poder de disseminação de um utilizador numa Rede Social Online como fator de determinação de risco para a sua privacidade digital

Junho, 2022



Universidade do Minho
Escola de Engenharia

João Lobarinhas Fernandes Miranda

Cálculo do poder de disseminação de um utilizador numa Rede Social Online como fator de determinação de risco para a sua privacidade digital

Dissertação de Mestrado

Mestrado em Engenharia Informática

Trabalho efetuado sob a orientação do(a)

Paulo Jorge Freitas de Oliveira Novais

Marco Filipe Vieira Gomes

Junho, 2022

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

_____, _____

(Localização) (Data)

_____

(João Lobarinhas Fernandes Miranda)

Agradecimentos

Para o desenvolvimento deste estudo foi essencial o trabalho do meu orientador Paulo Jorge Freitas de Oliveira Novais e do meu co-orientador Marco Filipe Vieira Gomes, sem eles seria impossível o trabalho apresentado, dada a sua supervisão e acompanhamento ao longo deste processo.

Resumo

A presença ubíqua das redes sociais faz com que estas sejam plataformas ideais para a disseminação de notícias, mas quando o conteúdo destas notícias é falso estas podem pôr em risco a privacidade digital de um utilizador. A disseminação de notícias nas redes sociais é principalmente feita com base em relações de confiança e confiabilidade entre os utilizadores, estas relações estão então diretamente ligadas com o poder de disseminação de um utilizador. A partilha e compartilha das notícias nas redes sociais é atualmente feita por utilizadores e contas *bot*, a disseminação das notícias falsas por *bots* põem a privacidade digital dos utilizadores das redes sociais em risco.

De forma a ter a recolher dados que pudessem ser livremente manipulados foi utilizado a ferramenta "FakeNewsNet", esta permitiu que fosse criado um *dataset* com os dados recolhidos da rede social Twitter. De forma a melhor compreender as redes de disseminação das notícias, os dados recolhidos foram aplicados ao "Community Health Assessment Model", este modelo é baseado nos modelos epidemiológicos, e que permite obter dados sobre a disseminação das notícias nas redes sociais entre utilizadores e dentro de comunidade ou "echo chambers". Os dados contidos no *dataset* foram também aplicados ao "Botometer", um modelo supervisionado de deteção de contas *bot* no Twitter, em que as contas dos disseminadores de notícias foram analisadas. Com os resultados obtidos dos modelos aplicados, é medido o impacto das contas *bot* na disseminação das notícias. O impacto das contas *bot* para a disseminação das notícias é entre 2,9% a 6,9% na disseminação de notícias dentro de comunidades e nas notícias falsas contribuem entre 2,8% a 0,7% na disseminação entre utilizadores, sendo que estas contas correspondem entre 0,272% a 0,296% da população que dissemina as notícias. Utilizando os dados dos *dataset* foi também feita uma análise de como as notícias são disseminadas. A análise foi feita em duas partes, uma parte foi dedicada às publicações e as relações que os utilizadores têm com contas *bot* e a outra parte foi dedicada ao poder de disseminação dos utilizadores. A partir da análise foi identificado que as contas *bot* desempenham um papel maior na disseminação das notícias falsas.

Utilizando esses dados são propostas medidas de modo a que os utilizadores consigam se proteger da influência das contas *bot*.

Palavras-chave: Poder Disseminação, Fake News, Bots, Redes Sociais, Modelos Epidemiológicos

Abstract

The ubiquitous presence of social networks makes them ideal platforms for spreading news, but when the content of this news is fake, it can put a user's digital privacy at risk. The dissemination of news on social networks is mainly done based on trust and trustworthiness relationships between users; these relationships are then directly linked with the dissemination power of a user. The *bot* users and accounts currently do the sharing and resharing of news on social networks, and the dissemination of fake news by *bots* puts the digital privacy of social network users at risk.

To collect data that could be freely manipulated, the tool "FakeNewsNet" was used, which allowed the creation of a *dataset* with the data collected from the social network Twitter. To better understand the news dissemination networks, the collected data was applied to the "Community Health Assessment Model", which is based on epidemiological models and allows to obtain data about the dissemination of news on social networks between users and within communities or "echo chambers". The data contained in the *dataset* was also applied to the "Botometer", a supervised Twitter *bot* account detection model, in which the accounts of news disseminators were analyzed. With the results obtained from the applied models, the impact of *bot* accounts on news dissemination is measured. The effect of *bot* accounts on news dissemination is between 2.9% and 6.9%on news dissemination within communities. On fake news, they contribute between 2.8% and 0.7% to dissemination among users, and these accounts correspond between 0.272% and 0.296% of the news disseminating population. An analysis of how the news is spread was also performed using the *dataset* data. The study was done in two parts. One was dedicated to the publications and the relationships that users have with *bot* accounts, and the other was to the users dissemination power. From the analysis, it was identified that *bot* accounts play a more significant role in disseminating fake news.

Using this data, measures are proposed so that users can protect themselves from the influence of *bot* accounts.

Keywords: Dissemination Power, Fake News, Bot Identification, Social Networks, Epidemiological Models

Índice

Lista de Figuras

Lista de Tabelas

Listagens

Glossário

Bot	É uma aplicação de software concebido para simular ações humanas.
Dataset	Um conjunto de dados ou "dataset" é uma colecção de dados normalmente tabulados.
Fact-Checking	Verificação de um algo apresentado como um facto.
Fake News	Fake news ou notícias falsas são uma forma de imprensa que consiste na distribuição deliberada de desinformação ou boatos via jornais, televisão, rádio, ou ainda online, como nas redes sociais. Este tipo de notícia é escrito e publicado com a intenção de enganar, a fim de se obter ganhos financeiros ou políticos.
Machine Learning	Machine Learning é um subcampo da engenharia e da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria de aprendizagem computacional em inteligência artificial. Utilizando algoritmos é possível criar modelos que vão aprendendo com os erros e assim fazendo previsões.

Siglas

ABM	Agent-Based Model ou Modelo Baseado em Agentes
API	Application Programming Interface
CNN	Convolutional Neural Networks
CSV	Comma-separated values
JSON	JavaScript Object Notation
LSTM	Long Short-Term Memory
ML	Machine Learning
TSM	Trust in Social Media
URL	Uniform Resource Locator

Capítulo 1

Introdução

1.1 Motivação

Num mundo em que cerca de 4,95 mil milhões de pessoas tem acesso a *internet*, o que corresponde a 59% da população mundial, tem existido proporcionalmente um aumento de utilizadores com acesso às redes sociais, sendo que de momento as redes sociais contêm cerca de 4,62 mil milhões de utilizadores ativos. [1]

A presença quase onipresente das redes sociais na vida dos seus utilizadores, faz com que estas sejam plataformas ideais para a disseminação de notícias. As notícias partilhadas variam muito no seu conteúdo e na sua generalidade, não impõem risco para os utilizadores das redes sociais, mas quando o conteúdo destas notícias é falso, as notícias podem pôr em risco a privacidade digital de um utilizador.

O risco à privacidade digital que as notícias falsas impõem para o utilizador é um risco diferente do normalmente associado com a privacidade digital. Normalmente os riscos que as redes sociais apresentam a privacidade digital de um utilizador estão associados com o risco dos dados privados do utilizador serem expostos, como por exemplo a sua morada ou número de cartão de crédito, mas com as notícias falsas este risco é menos literal. Neste caso as notícias falsas põem em risco outros fatores para além dos dados privados do utilizador, uma notícia falsas pode influenciar um utilizador nos seus pensamentos e formação de opiniões, o que pode levar até a afetar diretamente na tomada de decisões do utilizador na vida real, como por exemplo as notícias falsas em relação as vacinas, em que caso o utilizador acredite na notícia falsa pode pôr a vida deste em causa. Dado estes novos riscos é preciso medidas que possam proteger os utilizadores das redes sociais.

A disseminação de notícias nas redes sociais é principalmente feita com base em relações de confiança e confiabilidade entre os utilizadores, estas relações são importantes dado que ditam a eficácia da

[1] https://www.statista.com/statistics/617136/digital-population-worldwide/

informação ser disseminada numa rede social e então estão diretamente ligadas ao poder de disseminação de um utilizador [30][53]. As relações de confiança e confiabilidade também levam os utilizadores a criarem comunidades dentro das redes de disseminação, estas comunidades são muitas vezes criadas sem intencionalidade, mas são extremamente eficazes na disseminação de notícias, dado que os constituintes dessas mesmas têm normalmente relações altas de confiança e confiabilidade entre si. Este fenómeno tem o nome de "echo-chambers", este fenómeno normalmente está associado com a disseminação de notícias falsas, mas estão presentes em todas as cadeias de disseminação independentemente do conteúdo partilhado, dado que a criação destas comunidades apenas depende das relações entre os utilizadores.

Um outro factor que têm ganho maior relevância na disseminação de notícias é a utilização de contas *bot*. As contas *bot* no Twitter funcionam utilizando a API do Twitter, que permite fazer a maior parte das mesmas tarefas que uma conta "normal" consegue fazer. Depois da conta ter sido criada por um humano, este pode utilizar a API do Twitter para automatizar a sua atividade no Twitter (fazer publicações, retweets, etc). A utilização da API do Twitter tem os seus positivos e os seus negativos, por um lado permite a criação de contas *bot* legítimas, que apenas funcionam para a partilha de notícias de jornais ou blogs, mas também permite o surgimento de contas *bot* mal intencionadas [13].

As utilizações deste tipo das contas *bot* mal intencionadas são diversos, mas o seu objectivo final tende a ser influenciar um utilizador, neste caso de uso estas contas imitando o comportamento humano e criam ou compartilham uma publicação que contém uma notícia falsas, com o objectivo de aumentar o alcance da notícia ou dar a aparência para um utilizador de que a notícia partilhada é verdadeira, para este efeito são utilizadas um número considerável de contas de forma a um utilizador menos atento acredite que está em contacto com uma notícia verdadeira, dado que aparentemente existe um conjunto de utilizadores que validam naquela notícia e assim influenciando o utilizador a acreditar na notícia falsa. Estas interações dos utilizadores com as contas *bot* podem levar ao utilizador a criar relações de confiança e confiabilidade com um utilizador que dissemina notícias falsas e assim aumenta o poder de disseminação deste utilizador, ou até pode criar essas mesmas relações com a conta *bot*, dando então poder de disseminação a conta *bot*. As relações novas do utilizador para além de trazerem riscos para ele, podem também trazer riscos para os utilizadores que fazem parte da comunidade digital em que o utilizador se insere. É então importante quantificar o impacto que estas contas *bot* têm na disseminação de notícias e encontrar medidas que possam mitigar o seu impacto.

De forma a melhor compreender as diferenças entre as notícias falsas e as verdadeiras e o impacto das contas *bot* na disseminação das notícias é necessário efectuar a recolha de dados referentes a disseminação das notícias, esse dados serão então armazenados num "dataset" para posteriormente serem analisados. A recolha dos dados será feita utilizando ferramentas que permite extrair dados das publicações, como *FakeNewsNet* [67][68][69].

Para encontrar diferenças entre as redes de disseminação das notícias verdadeiras e das falsas, estas serão tratadas como se fossem doenças epidemiológicas, isto deve-se ao facto de a disseminação das notícias e a disseminação das doenças epidemiológicas terem semelhanças [28], este tipo de aplicação

posteriormente foi também utilizado, mas com a disseminação de rumores entre humanos [6] [10]. As semelhanças são principalmente associadas aos comportamentos que levam uma doença epidemiológica a espalhar-se pela população [6], uma doença epidemiológica é passada de um doente infectado para uma pessoa que tenha estado em contacto com ela, sendo que as pessoas mais próximas do infectado, como família ou amigos, têm um maior risco de serem infectadas e transmitirem a doença para outras pessoas próximas delas. No caso da disseminação das notícias nas redes sociais o mesmo comportamento é encontrado, um utilizador a partilhar uma notícia falsa põem em risco os seus seguidores, os seguidores agora "infectados" com a notícia falsas podem partilhar esta com os utilizadores próximos deles, criando assim cadeias de contacto. Enquanto que as cadeias de contacto das doenças doenças epidemiológicas são baseadas num contacto físico, as cadeias de contacto das noticias depende nas relações de confiança e confiabilidade entre os utilizadores, é nestas relações que a diferenciação entre as notícias falsas e as verdadeiras é encontrada, nas notícias falsas existe uma maior confiança e confiabilidade entre os utilizadores o que levam estas a serem mais "infecciosas", dado que existe uma maior probabilidade da notícia ser partilha entre os utilizadores. Ou seja, como nas redes de disseminação de notícias falsas os utilizadores têm uma maior confiança e confiabilidade entre eles, os utilizadores que disseminam essas notícias têm assim um maior poder de disseminação e impõem um maior risco para os utilizadores que estão em contacto com estes.

É então com estas semelhanças que os dados recolhidos com as notícias serão aplicados a modelos epidemiológicos, estas aplicações foram feitas e apresentaram bons resultados [54] [52] [28]. Com os dados obtidos com estas aplicações é pretendido encontrar factores que diferenciam as notícias e também tentar perceber o impacto das contas *bot* na disseminação das notícias.

De forma a identificar as contas *bot* serão utilizados algoritmos já existentes de identificação de contas *bot*, com uma preferência para algoritmos de aprendizagem supervisionada, dado que os comportamentos das contas *bot* estão em constante evolução e tentam ativamente não serem detectados pelos algoritmos existentes [84]. Os algoritmos de aprendizagem supervisionada apresentam a vantagem de terem uma maior facilidade em conseguir adaptar-se melhor a estas alterações, dado que recebem as opiniões dos utilizadores em relação à classificação que os algoritmos fornecem. Com os dados fornecidos por estes modelos é pretendido perceber melhor impacto e função que estas contas *bot* têm na disseminação de notícias.

Para além das análises feitas pelos modelos epidemiológicos, os dados recolhidos também serão processados de forma a encontrar diferenças entre as notícias falsas e as verdadeiras, compreender a função das contas *bot* na disseminação das notícias e perceber os riscos que o poder de disseminação pode ter para a privacidade digital de um utilizador.

1.2 Questões de Investigação

Dadas as implicações apresentadas na secção acima 1.1 surgem então as questões a que o trabalho pretende responder com o seu desenvolvimento:

- Como é que a utilização de modelos de identificação de *bots* nas redes sociais em conjunto com os modelos epidemiológicos resulta numa melhor identificação das notícias falsas?

  Esta questão pretende responder se a identificação das contas *bot* e a análise do impacto que estas contas têm na disseminação das notícias em conjunto com os resultados obtidos a partir dos modelos epidemiológicos, é possível obter uma melhor identificação em relação à veracidade da notícia.

- Qual é o impacto dos *bots* na disseminação de notícias falsas?

  Com os dados obtidos e a partir da análise dos mesmo e com auxílio dos modelos epidemiológicos, quantificar qual é o impacto das contas *bot* na disseminação das notícias e como é que estas contas impactam a disseminação das notícias.

- Com os resultados obtidos, como é que um utilizador de redes sociais consegue-se proteger melhor da influência dos disseminadores de notícias falsas?

  Utilizando os resultados obtidos da identificação dos *bot*, dos modelos epidemiológicos aplicados às notícias e a análise dos dados recolhidos, que medidas é que podem ser tiradas destes dados de forma ao utilizador se conseguir proteger da influência dos disseminadores de notícias falsas.

- Que medidas de protecção da integridade intelectual dos utilizadores poderão ser tomadas pelas redes sociais com os resultados obtidos?

  Como na questão acima, com os resultados obtidos que medidas as redes sociais podem utilizar para melhor protegerem a integridade intelectual dos seus utilizadores. A integridade intelectual refere-se algo que foi abordado na secção acima 1.1, mas que se refere ao facto de que as notícias falsas põem em risco os utilizadores em outros factores para além dos dados privados, uma notícia falsas pode influenciar um utilizador nos seus pensamentos e formação de opiniões, o que pode levar até a afetar diretamente na tomada de decisões do utilizador na vida real, como por exemplo as notícias falsas em relação as vacinas, em que caso o utilizador acredite na notícia falsa pode pôr a vida deste em causa.

1.3 Estrutura do Documento

O documento está estruturado em seis capítulos, incluindo este capítulo. Os dois primeiros capítulos, capítulo 1 "Introdução" e o capitulo 2 "Revisão da Literatura", dedicam-se a fazer uma introdução e enquadramento ao trabalho, nestes capítulos são debatida as razões e questões que motivaram o trabalho, é feito um levantamento de trabalhos relevante para o desenvolvimento do trabalho e as melhorias que o trabalho apresentam em relação às soluções existentes.

No capítulo 2 "Revisão da Literatura"este é dividido em seis seções: seção 2.1 "Modelos Epidemiológicos", em que são apresentados diferentes modelos epidemiológicos que podem ser aplicado a disseminação de informação; a seção 2.2 "Identificação de Contas *bot* nas Redes Sociais", onde são debatidos

os modelos e técnicas de deteção contas *bot*; seção 2.3 "Recolha de Dados nas Redes Sociais", em que são apresentados formas de como pode ser feita a recolha dos dados das redes sociais; a seção 2.4 "Identificação das Notícias Falsas nas Redes Sociais", são apresentados e debatidos os modelos existentes de identificação das notícias falsas; secção 2.5 "Trabalhos Relacionados", nesta secção são apresentados trabalhos que também eles tenta identificar o poder de disseminação das contas *bot* e as diferenças e melhorias que este trabalho apresenta em relação aos trabalhos relacionados, e a seção 2.6 "Análise do Estado de Arte", que funciona como uma conclusão para o capítulo.

Os tres seguintes capitulos, capitulo 3 "*Dataset*", capitulo 4 "Escolha de Modelos" e o capítulo 5 "Aplicação dos Modelos Selecionados e Cálculo do Poder de Disseminação", focam-se no desenvolvimento do trabalho e recolha de dados de forma responder às questões de investigação.

No capítulo *Dataset* são apresentadas as razões pela necessidade de ser feita uma recolha de dados, como esta recolha foi feita e os dados que foram recolhidos. O capítulo é composto por duas seções: seção 3.1 "*Dataset*" e seção 3.2 "Dados Recolhidos". A secção "*Dataset*" descreve como foi feita a recolha dos dados e a estrutura do *dataset*, enquanto que a secção "Dados Recolhidos" apresenta os dados que foram recolhidos e é feita uma apreciação crítica dos mesmos.

No "Escolha de modelos"são debatidos os diferentes modelos epidemiológicos aplicáveis a identificação das notícias falsas e os modelos aplicáveis a identificação das contas *bot* no Twitter. O capítulo é composto por três seções: seção 4.1 "Modelos Epidemiológicos aplicados à detecção de notícias falsas", seção 4.2 "Modelos de detecção de "bots"no Twitter" e seção 4.3 "Modelos selecionados". Nas secções "Modelos Epidemiológicos aplicados à detecção de notícias falsas"e "Modelos de detecção de "bots"no Twitter"são debatidos os diferentes modelos considerados e é feita a escolha de um desses modelos, na secção "Modelos selecionados"são apresentadas as razões pela qual os modelos escolhidos foram selecionados.

"Aplicação dos Modelos Selecionados e Cálculo do Poder de Disseminação"5 é um capítulo em que é abordado a aplicação dos modelos selecionados, o cálculo do impacto das contas *bot* na disseminação das notícias, e o cálculo do poder de disseminação. O capítulo é dividido em três seções: seção 5.1 "Aplicação dos Modelos Selecionados", secção 5.2 "Poder de Disseminação e as Contas *Bot*", e secção 5.3 "Medidas de Prevenção dos Utilizadores Contra a Disseminação de Notícias Falsas". A secção "Aplicação dos modelos e recolha de dados"é apresentada como foi feita a aplicação do modelo "Botometer"e "Community health assessment model", e como foi feita a aplicação dos resultados do "Botometer"no "Community health assessment model". Na seção "Poder de Disseminação e as Contas *Bot*", onde é apresentado os impacto das contas *bot* na disseminação das notícias, é feita uma análise a disseminação das notícias e o poder de disseminação dos utilizadores nas redes de disseminação das notícias.

O último capítulo é o capítulo 6 "Conclusão", que faz um resumo do trabalho feito e apresenta o trabalho futuro e os problemas encontrados durante o desenvolvimento.

1.4 Plano de Investigação

O plano de investigação está dividido em 5 diferentes eventos: levantamento de literatura Relevante, levantamento dos requisitos para o desenvolvimento do(s) programa(s), desenvolvimento e concretização do pipeline de processo e análise adequado ao problema, desenho e desenvolvimento da solução e escrita do documento da dissertação. Os eventos representam todos os diferentes passos necessários para o desenvolvimento das soluções para as questões de investigação.

O plano de investigação pode então ser visualizado pelas tabelas e diagrama de Gantt seguintes:

Tabela 1.1: Plano de investigação, com os eventos e data de início e fim de cada tarefa

Evento	Data Início	Data Fim
1 - Levantamento de literatura Relevante	16/10/2020	15/01/2021
2 - Levantamento dos requisitos para o desenvolvimento do(s) programa(s)	10/12/2020	15/02/2021
3 - Desenvolvimento e concretização do pipeline de processo e análise adequado ao problema	15/02/2021	25/03/2021
4 - Desenho e desenvolvimento da solução	25/03/2021	20/07/2021
5 - Escrita do documento da dissertação	05/03/2021	27/12/2021

A tabela 1.1 é composta por três colunas evento, data do início do evento e data do fim do evento. Cada evento é identificado por um número.

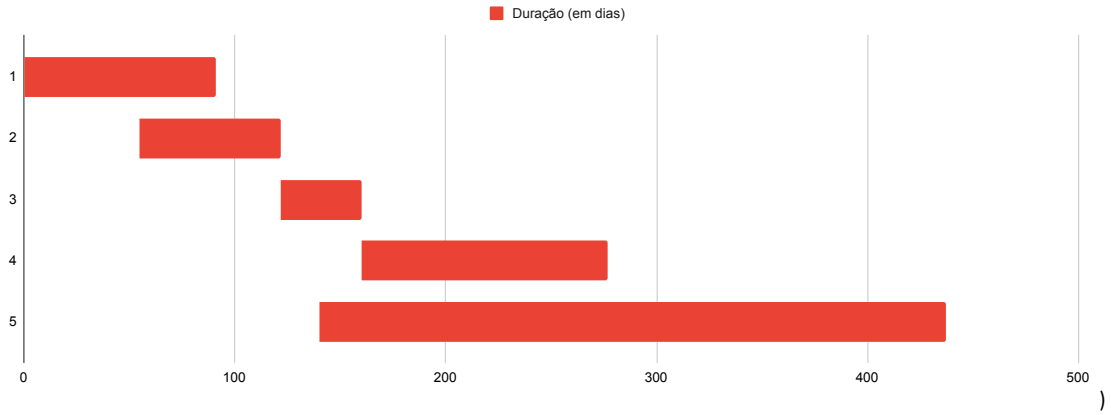


Figura 1.1: Diagrama de Gantt do plano de investigação

O diagrama 1.1 representa os eventos e a duração dos mesmos em dias. O eixo y é composto pelos números que identificam cada evento, como pode ser visualizado na tabela 1.1, o eixo x representa os eventos em espaço temporal e em duração de dias.

Capítulo 2

Revisão da Literatura

De forma a melhor abordar as questões de investigação referidas no 1º capítulo, foi necessário efetuar um estudo do estado da arte. O capítulo "Revisão da Literatura"foca-se no levantamento das áreas que foram consideradas relevantes para a realização do trabalho, essas áreas sendo:

- Modelos epidemiológicos - Esta área foi selecionada dada a aplicação destes modelos a redes de disseminação de rumores e notícias nas redes sociais, esta área também permite medir a vulnerabilidade que os utilizadores têm de se "enfeitarem"com um rumor ou notícia, o que assim podem permite compreender o impacto das contas *bot* na disseminação das notícias falsas. A utilização destes modelos também permite a identificação e diferenciação das notícias falsas das verdadeiras.

- Identificação de contas *bot* nas redes sociais - Dada a necessidade de compreender o impacto das contas *bot* na disseminação das notícias nas redes sociais e de como a identificação das contas *bot* pode levar a uma melhor identificação das notícias falsas, isto leva a esta área ter de ser analisada.

- Recolha de dados nas redes sociais - De forma a se obterem dados que possam ser aplicados aos modelos selecionados houve o interesse de se analisar esta área. Também a recolha de dados que possam ser manipulados livremente e que fossem recentes, é algo desejável para os modelos que foram selecionados.

- Identificação das Notícias Falsas nas Redes Sociais - Como uns dos objetivos é melhor compreender como a identificação das contas *bot* pode levar a uma melhor identificação das notícias falsas e o impacto das contas *bot* na disseminação das notícias falsas. Um melhor estudo dos diferentes modelos existentes na identificação das notícias falsas foi feito.

- Trabalhos Relacionados - Apesar da área do cálculo do poder de disseminação dos utilizadores nas redes sociais e o impacto que as contas *bot* podem ter para a sua privacidade ser um pouco

nicho, existem um conjunto de trabalhos nesta área e assim torna-se interessante o seu estudo e identificação de como este trabalho se diferencia dos outros já existentes.

Este estudo foi feito de forma a melhor compreender as diferentes áreas que irão ser abordadas ao longo do desenvolvimento, o estudo é importante de forma a levantar possíveis abordagens que possam vir a ajudar na resolução dos problemas que irão surgir.

2.1 Modelos Epidemiológicos

Os modelos epidemiológicos têm sido estudados desde do início do século 20 [60][61][62], e desde de então tem sofrido melhorias [34] [32] [25] e alterações [3] de forma a melhor modelarem a disseminação de doença epidemiológicas e também em outras aplicações. Estes modelos tenta fazer uma projeções de como uma doença epidemiológica se vai propagar de forma a ajudar os cientistas a medir o possível risco da doença para a população, os modelos utilizam dados estatísticos relativos a doença em conjunto com modelos matemáticos de forma a calcularem o risco de propagação.

Nesta tese os modelos epidemiológicos utilizados ou referidos serão principalmente os comportamentais em vez dos ABMs (*agent-based models* ou modelos baseado em agentes), isto deve-se a maior simplicidade e menor necessidade de poder computacional dos modelos comportamentais e também ao facto dos ABMs receberem críticas em relação a simplificar e fazerem assumpções irrealistas [71] [70]. Dentro dos modelos comportamentais existem diferentes tipos de modelos que categorizam os intervenientes em diferentes tipos de estados/comportamentos, o modelo mais simples e básico é o SI em que simplesmente existem dois comportamentos possíveis, ou S (suscetível) ou I (infeccionado), em que suscetível são os intervenientes susceptíveis a contrair a doença, e infeccionado são os intervenientes infectados. Dada a simplicidade do modelo SI, este modelo não é muito utilizado, mas é a base para os outros modelos comportamentais.

Dentro dos modelos epidemiológicos, há modelos mais aplicados do que outros, um dos modelos mais comuns no contexto da modelação das doenças epidemiológicos e ainda utilizado nesse mesmo contexto [35][66][85], e que já foi aplicado a disseminação das notícias é o modelo SIR [74], em que o SI tem o mesmo significado que no modelo SI referido acima, mas adiciona o comportamento R (Removido), este comportamento corresponde a um interveniente que tenha sido removido, porque este ou ganhou imunidade a doença ou morreu. Um outro modelo comum e que também já foi aplicado ao caso de uso é o SIS [74][28], que mais uma vez utiliza a base do SI, mas adiciona o S (suscetível), este comportamento corresponde ao facto de que nas doenças epidemiológicas os utilizadores depois de terem sido infectados, este podem voltar a contrair a doença.

Um dos modelos epidemiológicos, menos comuns é o modelo SEIZ[26], este modelo usa a base SI e adiciona dois outros comportamentos: E (exposto) e Z (cético). O comportamento "exposto" corresponde a alguém que foi exposto à doença, mas ainda está numa fase em que não sabe se está infetado com a doença ou não, o comportamento "cético" corresponde a alguém que esteve em contato com a doença

mas não a disseminou. Este modelo já foi aplicado à disseminação das notícias nas redes sociais com melhores resultados do que SIS ou SIR [28].

2.2 Identificação de Contas *Bot* nas Redes Sociais

Na identificação de contas *bot* nas redes sociais existem alguns modelos que podem ser aplicados a identificação deste tipo de contas no Twitter. Os modelos existentes baseiam na utilização de técnicas de *Machine Learning* (ML) em conjunto com *datasets* recolhidos das redes sociais, os *datasets* são depois utilizados com as diferentes técnicas de *Machine Learning*, de forma a criar um modelo que consiga classificar se uma conta é *bot* ou não.

Dentro das técnicas de ML utilizadas na detecção de contas *bot* existe uma maior preferência para a utilização de algoritmos de ML supervisionados [84], para este tipo de técnicas são necessários *data sets* extensos, estes *data sets* contêm contas que estão identificadas como *bot* e contas identificadas como normais. A identificação destas contas é normalmente feita por humanos [78], por métodos automatizados, como por exemplo métodos baseados em "potes de mel"[38], ou com uso a *botnets* que exibem comportamentos suspeitos [18]. Um problema com a identificação das contas *bot* é a falta de consenso no que define uma conta como *bot* numa rede social, isto torna-se ainda mais difícil nas contas que demonstram tanto comportamento humano como de *bot*, em que até investigadores experientes da área não conseguem discriminar facilmente. Independentemente das dificuldades, os *data sets* criados incluem um grande conjunto diferente de contas *bot*, usando as identificações das contas é possível criar modelos funcionais [65] [16].

De forma aos modelos criados identificarem de forma mais concreta as contas *bot* é necessária uma identificação das características principais que constituem uma conta *bot*. Em geral são consideradas 6 categorias de características que diferenciam uma conta *bot* de uma conta normal [76]: metadados da conta, metadados dos "amigos"da conta, as redes de partilha das publicações da conta, conteúdo e o tipo de linguagem utilizado pela conta nas suas publicações, os sentimentos expressos pelas publicações da conta (tentar compreender melhor os sentimentos expressos, como por identificar mensagens de ódio a partir das publicações, como por exemplo em "Hate speech classification in social media using emotional analysis"[43]), e os padrões temporais da conta. No caso dos modelos supervisionados, depois da extração e processamento das características que definem uma conta *bot*, estas são fornecidas aos modelos supervisionados para treino, os modelos depois de treinados são usados para avaliar contas desconhecidas ao modelo.

A maior parte dos modelos tentam detetar as contas *bot* no nível da conta, com base nas características da conta e as publicações da conta, mas existem modelos que utilizam uma abordagem baseada em *deep neural networks* no contexto da *long short-term memory (LSTM)* [36], esta arquitectura explora o conteúdo e os metadados para detectar se uma conta é *bot* no nível da publicação. Para isso, as características são extraídas dos metadados da conta e fornecidos como dados auxiliares ao LSTM de

aprendizagem profunda que processa a publicação. O modelo então consegue prever se uma publicação foi feita por uma conta *bot* ou não.

Apesar dos modelos supervisionados serem eficientes na identificação de contas *bot* na maior parte das situações, eles têm problemas na identificação de contas *bot* sociais coordenadas que publicam conteúdo gerado por humanos [15] [12], dado que essas contas *bot* quando analisadas individualmente não apresentam um comportamento suspeito. Para identificar as contas *bot* sociais coordenadas é necessário recolher dados em relação a sua coordenação, o que apenas se torna disponível depois da atividade de múltiplas contas *bot* é considerada. Para responder a esta dificuldade modelos não supervisionados são utilizados, nos primeiros modelos propostos combinavam várias características via distância euclidiana, sendo depois aplicados algoritmos de *cluster* ou algoritmos de deteção de comunidades [2] [44]. Noutra modelo o conteúdo das publicações é codificado num conjunto de símbolos, sendo depois consideradas as semelhanças entre os conjuntos de símbolos para identificar contas *bot* [14]. Há também modelos não supervisionados, que utilizando dados temporais, como data e hora das publicações, tentam identificar contas que publicam em sincronia, sugerindo então que estas possam ser contas *bot* [11] [12]. Apesar destes modelos não supervisionados terem a vantagem de considerarem o que as contas têm em semelhante, em vez do que as distingue, eles também têm a desvantagem de ter que considerar números quadráticos de pares de contas no pior caso [84].

A utilização dos modelos supervisionados então torna-se preferível, dada que estes são melhores na identificação das contas *bot* apenas utilizando os dados disponíveis da conta, para uma aplicação eficiente de modelos não supervisionados seria necessário a recolha de um grande conjunto de publicações feitas pela conta e também comparar essas publicações com outras de conteúdo semelhantes, de forma a corretamente identificar se é uma conta *bot* ou não. O que tornava a identificação das contas bastante mais complexa e difícil, sendo que a vantagem que os modelos não supervisionados têm sobre os supervisionados, é mais situacional do que facilmente generalizada.

2.3 Recolha de Dados nas Redes Sociais

A recolha de dados nas redes sociais é um fator importante para aplicação dos modelos de identificação de notícias falsas, identificação das contas *bot* e o cálculo do poder de disseminação. Apesar da existência de vários *datasets* públicos, em muitos casos estes já têm alguma idade e tornam difícil a identificação das contas *bot* e a identificação de notícias falsas, pois podem haver contas de utilizadores que podem ter sido removidas pela rede social ou eliminadas pelo utilizador, o que levaria a se obter resultados que não refletem a 100% a disseminação da notícia, outro factor a ter em conta é que os *datasets* podem não conter todos os dados necessários para as aplicações que estes iriam ser usados. Estes factores levam a que seja necessário efectuar a recolha de dados mais atuais.

Para recolher os dados necessários das redes sociais é necessário que seja possível ter acesso aos dados referentes à disseminação das notícias, das principais redes sociais foi escolhido o Twitter, pois existe um maior número de ferramentas que permitem a recolha de dados e análise dos mesmos. Então

a recolha foi feita apenas no Twitter, de forma a recolher os dados é utilizada a API do Twitter em conjunto com ferramentas de *data mining*. As ferramentas para a recolha de dados utilizam o Python em conjunto com módulos desenvolvidos para a recolha de dados utilizando a API do Twitter, isto permite que os dados referentes a publicações que partilharam uma determinada notícias possam ser recolhidos, com a recolha da publicação também é possível obter dados da conta que partilhou a notícia e saber quem compartilhou a publicação. Com os dados da conta é possível recolher quem são seus seguidores e as contas que esta conta segue, tornando assim possível identificar relações de "amizade" entre os utilizadores. Dentro destes requerimentos existe uma ferramenta que permite a extracção dos dados desta forma, *FakeNewsNet* [67], esta ferramenta permite recolher dados com base num *URL* de uma notícia que lhe for fornecido e recolhe dados referentes às publicações e aos utilizadores que fizeram essas publicações.

2.4 Identificação das Notícias Falsas nas Redes Sociais

Para identificação e distinção entre as notícias falsas e as verdadeiras, existem um conjunto de abordagens diferentes, umas que utilizam os modelos epidemiológicos e dados reais da disseminação de notícias, outras abordagens que preferem uma análises estatísticas de dados referentes a disseminação das notícias e ainda abordagens que utilizam técnicas de *machine learning*.

Nos modelos epidemiológicos aplicados à identificação de notícias falsas, existem diferentes tipos de abordagens. Em algumas abordagens são utilizados diretamente modelos epidemiológicos [74][28][41], nestes casos os modelos epidemiológicos já existentes como por exemplo SIS ou SIR, são aplicados a disseminação das notícias. Esta aplicação é feita tratando a disseminação da notícia como se fosse uma doença epidemiológica, para isso são recolhidos dados referentes a disseminação de notícias falsas e verdadeiras, esses dados são depois tratados e analisados de forma a poderem ser utilizados nos modelos epidemiológicos. Com os resultados obtidos a partir dos modelos epidemiológicos, é possível diferenciar e identificar as notícias falsas das verdadeiras. Outras abordagens usam modelos epidemiológicos como base, mas criam modelos mais adaptados para a disseminação das noticias nas redes sociais [54] [52] [73]. Nestes modelos ou são feitas algumas pequenas alterações aos modelos epidemiológicos de forma a estes estarem mais adaptados ao novo contexto a que estes são aplicados [73], ou então apenas usam parte da lógica existente nos modelos epidemiológicos, mas criam modelos relativamente diferentes ao esperado de um modelo epidemiológico convencional [54] [52].

Utilizando métodos estatísticos na disseminação das notícias é também possível distinguir entre as notícias falsas e verdadeiras [79]. Para este tipo de identificação é necessário um grande número de dados, para isso é necessário utilizar *datasets* existentes e também recorrer a recolha de dados da disseminação das notícias nas redes sociais. Os dados depois de recolhidos são analisados em diferentes campos, como o número de utilizadores que entraram em contacto com a notícia, a velocidade a que a notícia é disseminada, o alcance da disseminação da notícia, etc. As análises são feitas fazendo distribuições e utilizando algoritmos de estatística.

A aplicação de técnicas de *machine learning* é uma das abordagens mais populares para a identificação de notícias falsas nas redes sociais, mas dependendo da técnica de *machine learning* aplicada as abordagens podem ser muito diferentes entre si. Algumas abordagens preferem fazer uma análise de dados recolhidos sobre a disseminação das notícias nas redes sociais utilizando técnicas de *data mining* [68] [45], como o "processamento de língua natural"(PLN) aplicado ao texto das publicações ou a utilização de *convolutional neural networks* (CNN) ao conteúdo visual das publicações, como imagens ou vídeos, de forma a identificar os sentimentos expressos na publicação. Estas técnicas permitem uma análise profunda das características que diferenciam as publicações que contêm notícias falsas ou não, as características analisadas podem ser por exemplo o tipo de linguagem utilizado, os padrões de disseminação da publicação ou as contas dos utilizadores.

Outras abordagens preferem a aplicação de modelos de classificação de *machine learning*, estas técnicas permitem obter resultados precisos sem a necessidade de criar um modelo de raiz para a classificação das notícias falsas, para isso são identificadas características que o modelo precisa analisar e utilizar para desenvolver classificações [17]. Alguns exemplos da aplicação das técnicas de *machine learning* neste contexto são: a utilização de *Recurrent Neural Networks* [39] para a identificação das notícias falsas, aplicação de modelos baseados em redes neurais convolucionais [86] ou até a utilização de *Recursive Neural Network* [40] para este mesmo propósito.

A utilização de modelos epidemiológicos foi feita dada a vantagem de conseguir medir também a influência em diferentes níveis, que as contas *bot* podem ter na disseminação das notícias nas redes sociais.

2.5 Trabalhos Relacionados

Dentro da área em estudo existem um conjunto de trabalhos que relacionam a disseminação das notícias falsas nas redes sociais por contas *bot* e a sua relação e/ou impacto nos utilizadores destas redes, mas nenhuma das propostas existentes tenta utilizar os modelos de identificação de notícias falsas baseados em modelos epidemiológicos de forma a tentar compreender o impacto das contas *bot* na disseminação das notícias falsas e o risco que essa contas apresentam para os utilizadores.

Alguns trabalhos utilizam modelos de *machine learning* para identificar a disseminação de rumores e notícias falsas por grupos organizados [55] [78]. No "Detecting and Tracking Political Abuse in Social Media"[55] são utilizados modelos de *machine learning* de forma a identificar indivíduos e organizações com motivações políticas, que utilizam múltiplas contas controladas centralmente para criar a aparência de existir um apoio generalizado a um candidato ou opinião pública. O modelo sugerido utiliza recursos topológicos e características baseadas no conteúdo recolhido utilizando *crowdsourcing* de redes de difusão de informações no Twitter, para detectar a disseminação de rumores ou informações falsas numa fase inicial.

O "Detection of Promoted Social Media Campaigns"[78] é um modelo com a mesma motivação, mas em vez de analisar as redes de disseminação de um URL, este trabalho prefere analisar as publicações com

base na utilização de "hashtags" de forma a identificar assim as possíveis redes de disseminação, esta escolha é justificada dada que a partir "hashtags" é possível aumentar o alcance de uma publicação, pois o algoritmo do Twitter promove as "hashtags" mais populares durante um determinado período de tempo.

Ambos os trabalhos apresentaram um modelo que consegue distinguir entre publicações feitas por utilizadores e as publicações feitas por grupos ou pessoas organizadas com um bom grau de sucesso, mas "Detection of Promoted Social Media Campaigns"[78] identifica dificuldades em distinguir redes de disseminação política legítimas das mal intencionadas.

Um dos trabalhos relacionados tenta compreender a relação que os utilizadores das redes sociais tem com as contas mal intencionadas [19], neste trabalho é estudada a influência e interações que os utilizadores têm com conteúdo extremista, propaganda terrorista e campanhas de radicalização, com o objectivo de desenvolver uma ferramenta que analisa dados em tempo real e consiga detectar contas "extremistas", prever os utilizadores que vão interagir com contas "extremistas" e também prever os utilizadores com maior probabilidade de consumirem conteúdo "extremista". De forma a criarem a ferramenta foram recolhidos dados do Twitter, os dados continham 3395901 *tweets* recolhidos de 25538 contas identificados pelo Twitter como contas pertencentes a grupos terroristas e 29193267 *tweets* de utilizadores que interagiram com as contas identificadas como pertencentes a grupos terroristas. Os dados foram depois fornecidos a modelos de *machine learning* para classificarem 3 categorias de características diferentes, no total são 52 características distintas distribuídas pelas 3 categorias. Das 3 categorias definidas uma delas é "estáticas da rede", nesta categoria é abordada a difusão de informação entre utilizadores nas redes sociais a partir de *retweets* e menções em *tweets*, utilizando os *retweets* e as menções em *tweets* é possível fazer ligações diretas entre os utilizadores e assim compreender possíveis utilizadores que estão em contacto contas "extremistas"ou que consomem conteúdo "extremista". O trabalho conclui identificado que das 52 características, 11 delas tem um maior imapcto na detecção de contas "extremistas"e a possível detecção de interações com conteúdo "extremista", sendo que essas características estão mais associadas com a categoria das "estáticas da rede".

Apesar de não estar relacionado diretamente com as notícias falsas ou contas *bot* existem paralelos que podem ser feitos entre ambos, nomeadamente a utilização de dados do Twitter, análises a rede de disseminação de dados, a tentativa de compreensão de como informações inflamatórias se disseminam e influenciam os utilizadores das rede sociais e como contas mal intencionadas podem influenciar os utilizadores das redes sociais.

"Social bots distort the 2016 U.S. Presidential election online discussion"[5] estuda a influência que as contas *bot* tiveram na discussão política em torno das eleições presidenciais de 2016 nos Estados Unidos da América. O trabalho foca-se principalmente numa análise estática de dados recolhidos do Twitter em relação às eleições presidenciais de 2016, com a utilização do "Botometer"[65] [16] para a identificação das contas *bot*, com os resultados obtidos é concluído que as contas *bot* não interagem muito com utilizadores a partir de comentários, isto é justificado pela incapacidade das contas *bot* estudadas conseguirem imitar comportamento humano. Mas quando são analisados os "retweets" a publicações, os utilizadores

fazem "retweets" à publicações de contas *bot* com o mesmo rácio que fazem "retweets" à publicação de utilizadores reais, indicando assim que estas contas são tão eficazes a disseminar informação como um humano, o que pode levar a trazer riscos para os utilizadores das redes sociais.

Apesar dos trabalhos existentes tocarem em alguns dos tópicos que serão abordados, nenhum dos trabalhos existentes utiliza os modelos epidemiológicos, aplicados à detecção de notícias falsas nas redes sociais, de forma a melhor compreender o impacto das contas *bot* na disseminação de notícias falsas nas redes sociais. Com a utilização dos modelos epidemiológicos aplicados à detecção de notícias falsas como o sugerido em "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54] é a possibilidade de medir o impacto das contas *bot* na disseminação entre utilizadores dentro e fora das comunidades em que se encontram, e dessa forma melhor compreender o poder de disseminação das contas *bot* e as relações que estas têm com os utilizadores das redes sociais.

2.6 Análise do Estado de Arte

Com base nas secções anteriores e considerando os objectivos traçados para o trabalho na seção 1.2, atualmente existem ferramentas e trabalhos desenvolvidos que permitem o desenvolvimento de soluções para as questões levantadas.

O estudo feito as áreas dos modelos epidemiológicos e da identificação das notícias falsas nas redes sociais, permitiu melhor compreender como estas duas áreas se consegue conciliar, e de como a aplicação dos modelos epidemiológicos na identificação das notícias falsas nas redes sociais é uma vertente viável e aplicável que apresenta bons resultados como por exemplo "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54] e "Epidemiology inspired framework for fake news mitigation in social networks"[52].

A análise dos diferentes modelos de identificação de contas *bot* nas redes sociais permitiu que concluir que a utilização de um modelo supervisionado, com um maior foco na análise da conta em si do que no nas redes de partilha, seria a melhor aplicação para este caso, dado que para uma aplicação eficiente de modelos não supervisionados seria necessário a recolha de um grande conjunto de publicações feitas pela conta e também comparar essas publicações com outras de conteúdo semelhantes, de forma a corretamente identificar se é uma conta *bot* ou não.

Para a recolha de dados nas redes sociais ficou evidente que um maior foco seria dado à rede social Twitter, dado o suporte existente para a recolha de dados que esta plataforma fornece e as ferramentas desenvolvidas pela comunidade para esse mesmo efeito.

Comparando com o trabalho relacionado existente, este apresenta uma nova perspectiva na medição do impacto da disseminação de notícias por contas *bot* nas redes sociais, em que com a utilização de modelos epidemiológicos aplicados à disseminação de informação será possível medir o impacto destas contas a diferentes níveis, entre utilizadores e dentro de comunidades, o que é algo que os modelos anteriormente propostos não conseguem fazer.

Capítulo 3

Dataset

A criação de uma estrutura de dados existiu, pois era necessário obter dados que pudessem ser livremente manipulados, então isto levou a criação de um *dataset*. Neste capítulo serão abordadas as razões em mais detalhe e o porquê da necessidade da criação do *dataset*, como foi feita a recolha de dados, as ferramentas utilizadas e como foram organizados os dados.

3.1 Dataset

O *dataset* foi criado com base na necessidade de recolher um conjunto de dados recentes da disseminação de notícias nas redes sociais que pudessem ser livremente manipulados. Também existia o risco de que a utilização de dados que não fossem atuais pudesse comprometer a identificação de dados das contas *bot*, dado que estas contas podem ter sido banidas da rede social e assim seria impossível fazer a sua identificação.

De forma a se conseguir criar um *dataset* das redes de disseminação das notícias nas redes sociais, é necessário ter acesso aos perfis dos utilizadores, verificar se publicaram ou partilharam a notícia e também obter alguns dados pessoais como os seus seguidores e as contas que seguem, mas nem todas as APIs das redes sociais permitem que isto seja feito, deixando apenas o Twitter como a única em que se consegue recolher dados desta forma. Então o *dataset* foi construído apenas com dados obtidos a partir da API do Twitter.

A partir da API do Twitter é possível obter um conjunto de dados que foram úteis para o projecto, como recolher perfis dos utilizadores, pesquisar e obter dados de publicações com base em palavras ou frases chave, obter os números de identificação (ids) dos seguidores de um perfil e os ids das contas que este utilizador segue, recolher as publicações mais recentes de um utilizador, etc...

O levantamento de notícias verdadeiras e falsas utilizando *webjornais* que façam *fact-checking* como o

Politifact [1] ou Snopes [2], isto foi feito para que a decisão da veracidade ou não da notícia não dependesse de uma decisão feita pelo trabalho, mas sim de uma instituição independente ao mesmo.

O *dataset* recolhido contém 17 notícias, sendo 8 das notícias falsas e as restantes 9 são verdadeiras. Entre as 17 notícias foram recolhidos dados de 43,490 utilizadores que participaram ativamente na disseminação de notícias nas redes sociais e foram contabilizados 27,026,621 utilizadores que estiveram em contacto com as notícias recolhidas.

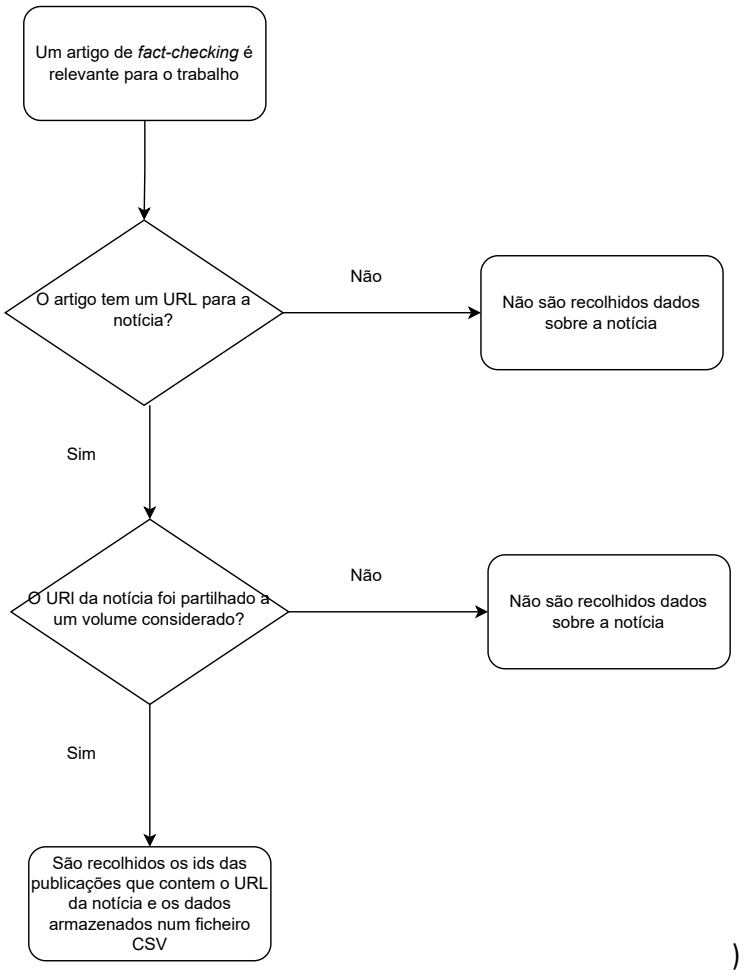
3.1.1 Recolha dos Dados

A recolha dos dados inicialmente focou-se na identificação de artigos de *fact-checking*, esta decisão foi feita de forma a que a determinação da veracidade da notícia não dependesse de uma decisão tomada pelo projeto, e também para que esta seja feita por um órgão mais reputáveis. A identificação é feita utilizando websites para o efeito como Politifact [3] ou Snopes [4]. A partir dos artigos de *fact-checking* é possível obter-se um *url* que redirecciona para a notícia que está a ser avaliada, com esse *url* é possível fazer uma recolha das publicações e partilhas feitas pelos utilizadores no Twitter.

A recolha foi feita utilizando o "FakeNewsNet". O "FakeNewsNet"[5][67][68][69] foi um projecto desenvolvido com o objectivo de criar um *dataset* que pudesse ser utilizado para o estudo dos comportamento das notícias falsas nas redes sociais, infelizmente com a idade do *dataset*, cerca de 3 anos, faz com que este não seja desejável para a aplicação que iria ter neste trabalho. O "FakeNewsNet" também contém uma ferramenta em Python que permite recolher dados referente a uma ou várias notícias contidas no *dataset* da ferramenta, dada a necessidade de se obter dados atuais o *dataset* existente não foi utilizado, então apenas foi utilizada a ferramenta de recolha de dados. Esta ferramenta foi escolhida dado o seu funcionamento simples e a maneira organizada e bem estruturada em que armazena os dados recolhidos. A ferramenta para obter dados lê um ficheiro CSV, o ficheiro contém dados com o *url* da notícia, titulo da notícia e os ids dos *tweets* em que a notícia foi partilhada.

A ferramenta do "FakeNewsNet" em si também foi alterada de modo a se adaptar às necessidades do projecto, nomeadamente a criação de um *script* em Python que com o *url* de uma notícia é possível obter os ids do *tweets* que partilharam aquela notícia. Depois dos dados serem recolhidos, estes são inseridos no ficheiro CSV para depois serem rastreados, também foi alterada a maneira como os dados são recolhidos, originalmente o "FakeNewsNet" recolhe as notícias todas que estão contidas no ficheiro CSV, como o objectivo é medir o impacto de um utilizador na partilha de uma notícia e não de um conjunto de notícias, a recolha é agora feita a apenas uma notícia. Para que o "FakeNewsNet" recolhesse a informação desta forma foi criando um CSV que contém o *dataset* todo e um outro CSV que apenas é utilizado para quando é necessário recolher a informação de uma notícia. Com estas alterações foi possível criar o *dataset* que foi utilizado para este projecto.

[1] https://www.politifact.com/
[2] https://www.snopes.com/
[3] https://www.politifact.com/
[4] https://www.snopes.com/
[5] https://github.com/KaiDMML/FakeNewsNet



)

Figura 3.1: Diagrama de decisão de como foi feita a escolha das notícias

Devido às limitações no número de pedidos feitos a API do Twitter, o processo de recolha de dados pode demorar entre 2 e 12 horas a 3 ou 4 dias, dependendo da dimensão da rede de disseminação da notícia. As notícias recolhidas e selecionadas podem ser visualizadas na tabela 3.1.

Em relação aos dados existe mais uma notícia verdadeira do que as falas, isto é devido ao facto que a notícia *Politifact_21* e a *Politifact_24* são sobre o mesmo tema, mas são partilhados por sites diferentes. O site que partilhou a notícia *Politifact_21* é normalmente um site que partilha notícias falsas enquanto que o que partilhou a *Politifact_24* é um website regular.

Tabela 3.1: Dataset recolhido.

ID Notícia	Titulo	Tipo	Disseminadores
Politifact_3	Per the CDC There Are Nearly Twice As Many Vaccine Related Deaths SO FAR in 2021	Falsa	2832
Politifact_4	REPORT: Pfizer Vaccine Confirmed To Cause Neurodegenerative Diseases - Study - National File	Falsa	4685
Politifact_5	Michael Jackson Absolved: Second Accusation of Molestation is Dismissed By Court in Final Ruling	Falsa	1583
Politifact_6	Italian hospital employee accused of skipping work for 15 years	Real	2753
Politifact_7	Israel, a world leader in vaccinations, lifts its outdoor mask mandate.	Real	601
Politifact_8	How will Biden's climate plan affect everyday Americans	Falsa	1222
Politifact_9	Out-of-control: Chinese rocket falling to Earth could partially survive re-entry	Real	1800
Politifact_12	Ellen DeGeneres to End Talk Show: Need Something New to Challenge Me (Exclusive)	Real	1886
Politifact_14	Spy Plane Identified Circling the Arizona Veterans Memorial Coliseum	Falsa	1439
Politifact_17	From the Desk of Donald J. Trump: The entire Database of Maricopa County in Arizona has been Hacked	Falsa	1003
Politifact_18	Ariana Grande Got Married to Dalton Gomez this Weekend	Real	4261
Politifact_19	Samuel E. Wright, Voice of Sebastian the Crab in The Little Mermaid Dies at 74	Real	500
Politifact_20	Black Lives Matter says it stands with Hamas terrorists in Israeli conflict	Falsa	4708
Politifact_21	Exclusive Photo: Gretchen Whitmer Violates Own Coronavirus Orders at Dive Bar	Real	4761
Politifact_22	China Says It Will Allow Couples to Have 3 Children, Up From 2	Real	659
Politifact_24	Whitmer apologizes after photo shows her at bar violating her own order	Real	731
Politifatc_27	Moderna Sent Coronavirus Vaccine To North Carolina University Weeks Before Pandemic	Falsa	2817

3.1.2 Estrutura do Dataset

Os dados depois de obtidos foram guardados no *dataset*, utilizando a mesma estrutura que o FakeNewsNet utiliza. Nesta estrutura existe um diretório raiz que armazena todos os dados referentes às notícia recolhidas, cada notícia tem o seu próprio diretório, o nome do diretório da notícia é criado utilizando um identificador e se a notícia é falsa ou não, como por exemplo Politifact_1_Fake.

Dentro do diretório da notícia existem outras quatro diretórios: um diretório que contém os dados dos perfis dos utilizadores em ficheiros de formato JSON, o nome dos ficheiros é atribuído utilizando o número de identificação da conta do Twitter dos utilizadores; dois diretórios que armazenam respectivamente os seguidores e as contas seguidas de um utilizador em ficheiros JSON, os ficheiros são também eles identificados pelo número de identificação da conta do Twitter do utilizadores; e um diretório que armazena os dados da notícia num ficheiro JSON, os *retweets* e os *tweets* referentes a notícia são armazenados em ficheiros JSON no seu respectivo diretório e são identificados pelo número de identificação do tweet.

3.2 Dados Recolhidos

A utilização da ferramenta de recolha de dados "FakeNewsNet"[67][68][69], foi essencial para a recolha dos dados da disseminação de notícias no Twitter. Apesar das limitações existentes na quantidade de pedidos que podem ser feitos a API do Twitter em conjunto com a grande quantidade de dados que são recolhidos, levou a que a recolha das notícias demorasse vários dias.

Os dados recolhidos 3.1 permitiram a criação de uma *dataset* de notícias de diversos campos, com populações diversas e tamanhos de rede diferentes, o que vai permitir fazer uma análise mais completa do impacto que as contas *bot* têm no poder de disseminação das notícias falsas e de que forma os utilizadores se podem proteger.

Capítulo 4

Escolha dos Modelos

Dada a existência de vários modelos nas áreas em estudo (modelos epidemiológicos aplicados à detecção de notícias falsas e modelos de detecção de contas *bot*) foi necessário realizar um estudo de quais seriam os modelos que melhor se aplicam. Inicialmente é feito um levantamento dos modelos considerados aplicáveis para o estudo, é feito um resumo do funcionamento dos modelos e como é que são aplicáveis ao caso em estudo, esses modelos são depois comparados uns contra os outros, indicando as vantagens de um sobre o outro e as desvantagens existentes nos modelos.

4.1 Modelos Epidemiológicos aplicados à detecção de notícias falsas

O interesse no estudo das redes de disseminação das notícias falsas ou rumores utilizando modelos epidemiológicos é uma área de estudo que existe quase a tanto tempo quanto os próprios modelos epidemiológicos, os primeiros modelos [51][49][50] de este género estudava como é que os rumores se disseminam em populações, estes modelos apresentavam as mesmas características que um modelo SIR (suscetível, infectado e recuperado), mas adaptado para a disseminação de rumores. Os modelos epidemiológicos também têm sido aplicados e adaptados para novos cenários, como para *markting* virtual [57][58][29], estudos na área da psicologia [42], estudos económicos ou financeiros [1], difusão de inovações [59], propagação de tendências [72], identificação de comportamentos fanáticos [27], utilização de drogas [22], etc.

Uma das novas aplicações tem sido para a identificação de noticias em relação a sua veracidade nas redes sociais, projectos como por exemplo "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], "Epidemiology inspired framework for fake news mitigation in social networks"[52], "Epidemiological modeling of news and rumors on Twitter"[28].

Dos projetos listados o maior foco foi dado "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], esse foco é explicado nas próximas sub-seções.

4.1.1 Epidemiological modeling of news and rumors on Twitter

Neste artigo são utilizados modelos epidemiológicos já existentes, como o SIS e SEIZ[26], e estes são aplicados a propagação de notícias no Twitter. Os resultados finais apresentados são positivos, mas a aplicação destes modelos na propagação das notícias tem as suas falhas, como não modelar as relações entre utilizadores de uma maneira mais realista em comparação com os outros modelos, como o "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54] que tenta computar a confiança e confiabilidade entre os utilizadores de forma a tentar determinar o poder de disseminação de um utilizador; e em relação ao "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], o modelo em "Epidemiological modeling of news and rumors on Twitter" apenas permite medir o quão "infecciosa" a notícia é, não permite medir o impacto das contas *bot* na disseminação das notícias num contexto entre utilizadores e dentro das comunidades como o "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54].

O modelo SIS (suscetível, infectado, suscetível), neste modelo os utilizadores susceptíveis (utilizadores que estiveram em contacto com a notícia) apenas conseguem passar para infectados (utilizadores que partilham a notícia), o que não realmente acontece, um utilizador pode acreditar no conteúdo de uma notícia falsa mas nunca partilhar a notícia, este tipo de comportamento não é abrangido pelo modelo SIS, esta crítica também é feita também pelo próprio trabalho [28].

O outro modelo sugerido foi o SEIZ(suscetível, exposto, infectado e séptico), neste caso de aplicação os diferentes estados que um utilizador pode estar são os seguintes:

- Suscetível - Um utilizador que ainda não ouviu falar da notícia.
- Exposto - Um utilizador que já ouviu falar da notícia, mas demorou algum tempo (*exposure delay*) antes de publicar ou partilhar notícia.
- Infectado - Um utilizador que publicou ou partilhou a notícia.
- Séptico - Um utilizador que já ouviu falar da notícia, mas decidiu não partilhar a mesma.

Em relação ao SIS este modelo permite que os utilizadores depois de entrarem em contacto com uma notícia não fiquem logo no estado infectado, apenas passando para este depois de partilharem a notícia, conseguindo assim replicar melhor o comportamento normal da partilha de uma notícia no Twitter.

De forma a conseguirem adaptar um modelo epidemiológico para este novo contexto, variáveis como a taxa de contato, o tempo de incubação, taxa de transição efectiva, etc. Tiveram de ser adaptadas e calculadas baseando-se nos comportamentos dos utilizadores na partilha das notícias, para o cálculo ser

possível são definidas algumas regras como por exemplo, um utilizador séptico só pode recrutar utilizadores do campo dos susceptíveis, desse contacto o utilizador contactado pode passar a ser um infectado ou séptico, a chance de sucesso é calculada baseada numa probabilidade deste evento acontecer.

Apesar das melhorias do modelo SIEZ em relação ao SIS, o "Community Health Assessment Model"[54] tem uma abordagem mais realista nas relações entre os utilizadores, dado que utiliza o TSM [64] para calcula a confiabilidade entre os utilizadores que partilham a notícia, utilizar o TSM é um ponto positivo porque os utilizadores não têm o mesmo nível de confiança entre eles, ao utilizar um probabilidade, como no SEIZ, é assumido que a probabilidade da transição entre os estados, por exemplo séptico para infectado, é igual para todas transições que vão acontecer.

4.1.2 Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model

"Community Health Assessment Model"[54] é um modelo criado para medir a vulnerabilidade de haver a partilha de notícias falsas numa rede de utilizadores numa rede social. O modelo é baseado no conceito de nas redes de partilha haver comunidades de utilizadores centrais, utilizadores vizinhos e utilizadores "fronteira", e com estes conceitos é proposto uma métrica que avalia a vulnerabilidade de um utilizador e de uma comunidade estarem expostos a uma notícia externa à comunidade.

Neste modelo é definido que as redes sociais são compostas por comunidades, que são estruturas que são grupos modulares, dentro dos grupos há utilizadores que têm uma conexão próxima entre eles, e entre os grupos há utilizadores vagamente conectados. Sendo que a modularidade é a proporção da densidade de utilizadores dentro da comunidade para os utilizadores fora da comunidade. Então uma comunidade está em risco de ser "infectada" com uma notícia falsa por utilizadores "vizinhos" a essa comunidade e assim que está exposta a notícia a probabilidade desta infectar toda a comunidade é mais alta. Então torna-se importante a identificação das comunidades vulneráveis a serem "infetadas" com a propagação das notícias falsas, para que estas possam ser protegidas e assim limitar a influência das notícias falsas na rede. Com esta necessidade em conta é introduzido primeiro o conceito de nodos/utilizadores centrais, nodos/utilizadores vizinhos e nodos/utilizadores "fronteira".

Os três diferentes tipos de nodos representam um tipo de comportamento que apresentam dentro de uma comunidade:

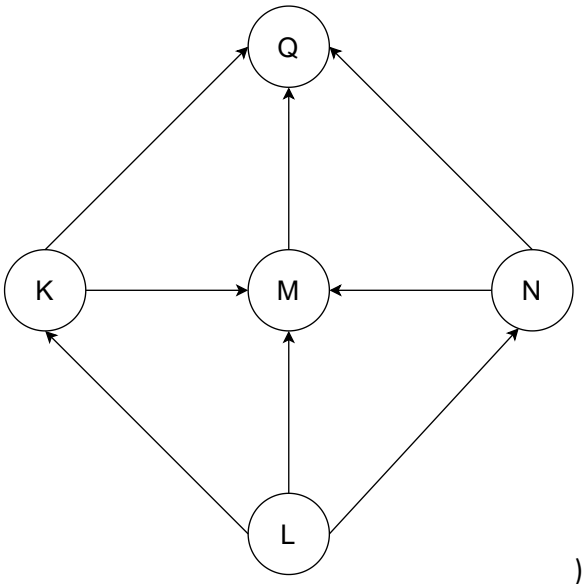
- **Nodos centrais** - Estes nodos apenas estão conectados a nodos dentro da comunidade.
- **Nodos vizinhos** - Estes nodos estão diretamente ligados a um membro da comunidade. Estes nodos não fazem parte da comunidade.
- **Nodos limite** - Os nodos da comunidade que estão conectados diretamente a pelo menos um nodo vizinho.

Adicionalmente é necessário compreender os conceitos de confiança, confiabilidade e credibilidade:

- **Confiança e Confiabilidade** - No contexto das redes sociais, de forma a compreender como as relações entre os utilizadores são criadas "Trustingness & Trustworthiness: A Pair of Complementary Trust Measures in a Social Network"[64] sugere utilização de uma pontuação de "confiança", para melhor compreender como estes se relacionam. Então é sugerido a utilização do modelo *Trust in Social Media* (TSM) [64], este modelo permite atribuir a um utilizador um par de pontuações em relação a sua "confiança" numa rede, a pontuação é composta pela confiança e confiabilidade. A confiança é relativa à confiança que um utilizador tem com os seus utilizadores "vizinhos", a confiabilidade é referente à confiança que os utilizadores vizinhos têm em relação ao utilizador. Modelos semelhantes que quantificam a confiança de utilizadores em redes também já utilizados em outros contextos e apresentaram bons resultados, como por exemplo "A Non-intrusive Approach to Measuring Trust in Opponents in a Negotiation Scenario"[24].

  Adicionalmente existe o conceito de "função de entrada" e "função de saída".A "função de entrada" de um nodo é definida por $inv(v)$, em que $v \in V$, sendo V o conjunto de nodos de origem para todas as arestas de entrada do nodo v. "Função de saída" de um nodo é definida por $out(v)$, em que $v \in V$, sendo v o conjunto de nodos de destino para todas as arestas de saída do nodo v.

  A pontuação da "confiança" (composta pela confiança e confiabilidade), tem como principal propriedade o *feedback* negativo da confiança. De forma a melhor compreender o conceito do *feedback* negativo da confiança a imagem 4.1 pode ser usada como referência. Na imagem o nodo L, tem a maior propriedade de confiança de todos nodos. O nodo L confia em todos só nodos exceto o nodo Q, então a confiança dada pelo nodo L tem menos "peso"do que a confiança dada por um nodo como o M, que é mais selectivo nos nodos que dá confiança.



)

Figura 4.1: Exemplo de uma rede, em que as arestas direcionadas indicam se um nodo confia no outro nodo

A partir da imagem 4.1 também é possível concluir que o nodo Q é um nodo muito confiável na rede, dado que o nodo Q tem a confiança dos nodos K, M e N, além disso os nodos que confiam

em $Q(K, M, N)$ são nodos que são seletivos nos nodos que confiam, o que torna o seu voto de confiança mais valioso do que o de L.

Utilizando a propriedade do *feedback* negativo da confiança, descrita acima, pode ser concluído que uma maior pontuação de confiança pouco afeta a confiabilidade dos seus nodos vizinhos. Mas uma maior pontuação de confiabilidade numa rede é o resultado dos seus nodos vizinhos terem pouca pontuação de confiança. Então a confiança de um nodo é traduzida pela seguinte equação:

$$confianca(v) = \sum_{\forall x \in out(v)} \frac{w(v,x)}{1 + confiabilidade(x)} \tag{4.1}$$

Na equação 4.1 é possível verificar que a confiança depende de 3 fatores: confiabilidade dos nodos de destino, número de ligações com outros nodos, e o peso de cada ligação com outros nodos.

A confiabilidade é calculada de forma semelhante:

$$confiabilidade(u) = \sum_{\forall x \in in(u)} \frac{w(x,u)}{1 + confianca(x)} \tag{4.2}$$

A equação 4.2 também depende de 3 fatores: confiança do nodo de origem, número de nodos que ligam ao nodo, e o peso de cada ligação ao nodo.

O peso das ligações entre os nodos, $w(x, v)$, pode ser uniforme ou variável, dependendo do contexto da rede.

Para o *Trust in Social Media* poder dar as pontuações este também usa o conceito de "envolvimento" numa rede social. O "envolvimento"[37] é definido como a uma construção orientada ao utilizador, o que significa que não reside num objecto, mas é sim uma construção mental do utilizador na percepção da relevância pessoal e a importância sobre o objeto, assim o envolvimento não pode ser adequadamente inferido a partir das propriedades da rede, porque é algo pessoal ao utilizador, então é sugerido a utilização de um questionário para se conseguir descobrir o envolvimento do utilizador na rede. O "envolvimento" numa rede [37] também é definido como o risco em fazer uma ligação errada entre dois intervenientes, então quanto maior for o risco numa rede, maior deve ser o impacto da confiança dos nodos vizinhos no cálculo da confiabilidade de um nodo, e vice versa. De forma ao "envolvimento" ser utilizado no TSM, uma função de decaimento exponencial é utilizada. Como foi verificado acima, o aumento de uma pontuação, como por exemplo a da confiança, inversamente afeta a segunda pontuação (confiabilidade) dos seus vizinhos. Uma função de decaimento exponencial ajuda a caracterizar esta propriedade como foi verificado nas fórmulas 4.1 e 4.2. Então como o envolvimento numa rede é definido como o risco de fazer uma ligação errada entre dois intervenientes, quanto maior for o envolvimento numa rede, maior é a influência da pontuação da confiança de um nodo vizinho, no cálculo da confiabilidade de um nodo, e vice versa. Usando uma distribuição de Zipf [47], pode ser concluído que a confiabilidade

é proporcionalmente inversa à soma dos expoentes do envolvimento das pontuações de confiança dos nodos vizinhos. Então as equações 4.1 e 4.2 passam a ser:

$$ti(v) = \sum_{\forall x \in out(v)} \frac{w(v,x)}{1 + (tw(x))^s} \tag{4.3}$$

$$tw(u) = \sum_{\forall x \in in(u)} \frac{w(x,u)}{1 + (ti(x))^s} \tag{4.4}$$

Em que $ti(v)$ corresponde a confiança de um nodo v, $tw(v)$ é a confiabilidade de um nodo v, e s corresponde a pontuação do "envolvimento" da rede. De forma a melhor compreender o impacto do "envolvimento" numa rede, se existirem duas redes uma rede "A" com $s = 0$ e outra rede "B" com $s = 1$. Na rede "A" não existem riscos em fazer ligações erradas entre os nodos de rede, então o cálculo da confiabilidade não é afetada pela confiança dos nodos vizinhos. Utilizando $s = 0$ na equação 4.4 ficamos com:

$$tw(u) = \sum_{\forall x \in in(u)} \frac{w(x,u)}{1 + (ti(x))^0} \tag{4.5}$$

Com o $s = 0$ converte o valor da confiança para 1, e então a confiabilidade fica uma relação entre a quantidade de conexões e a qualidade delas. Na rede "B" com o $s = 1$, existe um risco elevado em fazer uma ligação errada, então ao calcular a confiabilidade a confiança dos nodos vizinhos afeta extremamente o resultado da confiabilidade de um nodo. Utilizando $s = 1$ na equação 4.4 ficamos com:

$$tw(u) = \sum_{\forall x \in in(u)} \frac{w(x,u)}{1 + (ti(x))^1} \tag{4.6}$$

Neste caso, como o $(ti(x))^1$ é utilizado, a confiança dos vizinhos afeta completamente a confiabilidade dos nodos na rede "B". Apesar do "envolvimento" da rede desempenhar um papel importante no cálculo da confiança e confiabilidade, o "Community Health Assessment Model"[54] utiliza defini o "envolvimento" da rede como 1, em todas as redes utilizadas para testes. Esta decisão não é explicada, mas deve ter sido feita dado o tamanho das redes que foram utilizadas e o facto de que para redes grandes fazer um questionário de todas as relações que um utilizador tem na rede, não é prático, nem é completamente realista. Também nada garante que os utilizadores fossem responder a um questionário que seria extremamente extenso e prolongado, e mesmo que respondessem não existem garantias das respostas realmente terem sido completamente honestas.

- **Credibilidade** - A Credibilidade é uma pontuação derivada das pontuações da confiança e confiabilidade. Esta pontuação ajuda a quantificar a probabilidade de um utilizador acreditar numa

notícia falsa publicada por um determinado utilizador. A credibilidade de uma relação entre utilizadores é calculada em função da confiabilidade do remetente e da confiança do receptor. No contexto duma rede de partilha no Twitter, um relação de credibilidade de um utilizador A para um utilizador B, existe se um "tweet"(publicação) do utilizador A for "retweeted"(compartilhado) por B. A credibilidade quantifica então a "força" em que B confia em A, quando B faz um "retweet" a A. Portanto, é mais provável B acreditar em A se:

1. A tem uma pontuação elevada de confiabilidade na rede. A tem a confiança de muito intervenientes na rede, ou;
2. B tem uma pontuação elevada de confiança. B tem uma probabilidade alta de acreditar em outros utilizadores.

A credibilidade então é proporcional aos dois valores indicados acima, o que pode ser traduzida pela equação seguida:

$$Credibilidade(A \rightarrow B) = ti(A) * tw(B) \tag{4.7}$$

Com as noções acima definidas, é possível então definir as métricas que o modelo utiliza para medir a vulnerabilidade dos nodos. Estas métricas são criadas com base na ideia de que, para a disseminação de notícias falsas nas redes sociais é necessário existir um nível maior de confiança entre utilizadores intervenientes, dado que estas notícias como são falsas, se um utilizador tiver pouca confiança na rede, este simplesmente não confia no disseminador ou verifica a veracidade da notícia. Então são propostas métricas que medem a probabilidade de um utilizador com base nas ideia de definidas no *Trust in Social Media* [64], particularmente a pontuação da credibilidade para avaliar a saúde dos indivíduos e comunidades que encontram notícias falsas. As métricas definidas são as seguintes:

- **Vulnerabilidade dos nodos limite** $Vl(b)$: Esta métrica mede a probabilidade de nodo limite b se tornar num disseminador de uma notícia. A métrica é derivada de um nodo b acreditar num seu vizinho imediato n, esta métrica é calculada utilizando a confiabilidade do vizinho n (em que $n \in N_b$ e N_b é conjunto de todos os nodos que são vizinhos de b) e a confiança do nodo b, e é quantificada por $bel_{nb} = tw(n) * ti(b)$, ou seja $Credibilidade(n \rightarrow b)$. Então a probabilidade de b não ser vulnerável a n pode ser calculada por $(1 - bel_{nb})$. Generalizando, a probabilidade de b não ser vulnerável a nenhum dos seus nodos vizinhos é:

$$\prod_{\forall n \in N_b} (1 - bel_{nb}) \tag{4.8}$$

E a probabilidade de b acreditar em todos os seus vizinhos é (vulnerabilidade do nodo limite b):

$$Vl_b = 1 - \prod_{\forall n \in N_b} (1 - bel_{nb}) \tag{4.9}$$

- **Vulnerabilidade da comunidade** $Vc(C)$: Esta métrica mede a probabilidade dos nodos limite de uma comunidade $C(B_c)$ acreditarem em notícias recebidas por qualquer um de nodos vizinhos. A métrica é derivada da seguinte forma, utilizando a ideia da probabilidade de um nodo limite b não ser vulnerável aos seus vizinhos pode ser quantificada por $1 - Vl_b$, generalizando esta ideia a $b \in B_c$, a probabilidade de nenhum dos nodos da comunidade serem vulneráveis a influência dos seus vizinhos é:

$$\prod_{\forall b \in B_C} (1 - Vl(b)) \tag{4.10}$$

  Então probabilidade da comunidade C ser vulnerável a algum dos seus vizinhos (vulnerabilidade da comunidade):

$$Vc(C) = 1 - \prod_{\forall b \in B_C} (1 - Vl(b)) \tag{4.11}$$

Como já foi referido várias vezes este modelo mede o risco das comunidade existentes numa rede de partilha de notícias, com os dados obtidos é possível identificar as comunidades presentes numa rede de partilha de notícias, de forma a essa comunidades serem identificadas são utilizados algoritmos de detecção de comunidades. Estes algoritmos permitem criar comunidades numa rede por meio de representações grafos, estes métodos de detecção de comunidades encontram sub-redes com significativamente mais probabilidade de existirem conexões entre nodos no mesmo grupo do que nodos em grupos diferentes[21]. Entre os diversos algoritmos de detecção de comunidades foram utilizados pelos autores os seguintes: Louvain [7], Infomap [63], e Label Propagation [48], sendo que qualquer algoritmo de deteção de comunidades pode ser utilizado, como Leidenalg [75], Spinglass[56] ou Edge-Betweenness [21] .

Para medir o quão boas as métricas propostas são a quantificar a vulnerabilidade dos nodos e das comunidades, foram avaliadas as classificações dadas pelas métricas da vulnerabilidade dos nodos limite e das comunidades, em comparação com a veracidade das notícias recolhidas. Os autores então utilizam então medidas utilizadas em *Information Retrieval literature* [81].

Na vulnerabilidade dos nodos limite $Vl(b)$, um nodo vulnerável tem uma chance maior de ter uma credibilidade alta com os seus nodos vizinhos, com base nesta noção é considerado um nodo vulnerável um nodo que faz um "retweet", então a vulnerabilidade para um nodo limite é uma classificação binária, dado que apenas existem dados se o nodo fez um "retweet" ou não. Esta métrica é então avaliada utilizando a *Average Precision* e mAP (*mean average precision*) sobre todas as comunidades da rede.

Na vulnerabilidade das comunidades $Vc(C)$, uma comunidade com um número maior de nodos limite que são disseminadores é uma comunidade mais vulnerável. Como a maior parte das comunidades vai ter mais do que um nodo limite que é disseminador, não podem ser aplicadas as mesmas métricas que foram utilizadas na vulnerabilidade dos nodos limite. Então as comunidades são classificadas pela sua

pontuação de vulnerabilidade e as pontuações são comparadas com o número de nodos limite que são disseminadores. É também utilizado o coeficiente de correlação tau de *Kendall* (τ)[33], que é utilizado para medir a correlação de postos entre duas quantidades medidas.

Apesar de não ser diretamente um modelo epidemiológico, o "Community Health Assessment Model"apresenta um modelo comparável a um modelo SIR mas melhor adaptado para um contexto tecnológico das redes sociais. Neste modelo as notícias falsas são tratadas como se fossem a uma infecção viral e as redes sociais são a população, nos modelos epidemiológicos existem taxas de contato, que indicam a probabilidade de um utilizador ficar infectado no caso deste modelo a vulnerabilidade dos utilizadores no limite das comunidades e as comunidades ficam como indicadores dessa mesma taxa. A utilização do "Trust in Social Media" também permite uma melhor medição da taxa de vulnerabilidade para as redes sociais, dado que esta é calculada para todos os intervenientes na rede, e assim melhor simular as relações existentes entre os utilizadores. A aplicação de modelos que simulam os comportamentos de pessoas aplicados às redes sociais também têm apresentado resultados positivos, como por exemplo "Improving Conflict Support Environments with Information Regarding Social Relationships"[23], o que assim cria uma maior confiança na utilização deste modelo.

4.1.3 Epidemiology inspired framework for fake news mitigation in social networks

Este modelo pretende utilizar o "Community Health Assessment Model"[54] como base para construir um modelo que também consiga identificar os disseminadores de notícias falsas e um modelo de prevenção e controlo da disseminação das notícias falsas.

A identificação dos disseminadores de notícias falsas é feito utilizando o conceito de credibilidade existente no modelo "Community Health Assessment Model"[54], com as redes de disseminação das notícias representados na forma de grafo e utilizando as pontuações de credibilidade atribuídas ao nodos, (que representam um utilizadores,) são utilizados modelos de *network representation learning* para gerar *node embeddings*, o que permite representar os nodos como vetores, os resultados obtidos são depois aplicados a modelos *recurrent neural network* que permitem a identificação se um utilizador é um disseminador de notícias falsas ou não.

Na prevenção e controle da disseminação das notícias falsas, a motivação para isto pode ser explicada pela figura 4.2. O nodo D_1 é vizinho da comunidade 3 é um disseminador de notícias falsas, o nodo A_3 é um nodo limite da comunidade 3 e está exposto as notícias falsas disseminadas pelo nodo D_1, o que faz com que o nodo A_3 tenha uma percentagem alta de disseminar notícias falsas para a comunidade 3. De forma a prevenir este tipo de cenários é sugerida a identificação dos nodos limites de todas as comunidades de uma rede, que tem uma maior probabilidade de se tornarem disseminadores de notícias falsas. Considerando o cenário em que A_3 é um disseminador de notícias falsas, os nodos B_3, D_3 e E_3, que são seguidores do nodo A_3 estão agora expostos a notícias falsas, e a restante comunidade 3 estão em risco de serem expostos as notícias falsas. Dada a proximidade que estes nodos têm em relação a A_3,

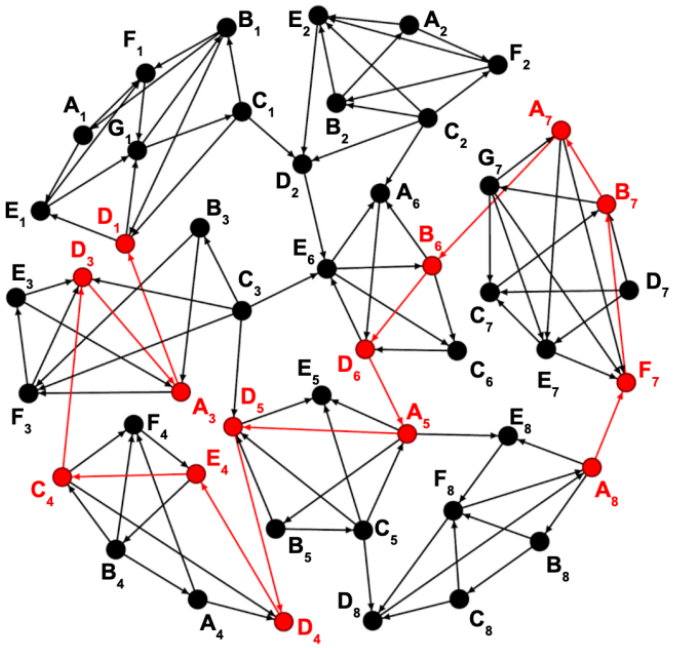


Figura 4.2: Exemplo de uma rede de disseminação de notícias. Os nodos vermelhos representam notícias falsas.

estes estão vulneráveis a acreditar em A_3 e assim disseminar notícias falsas pelo resto da comunidade. Então torna-se importante a identificação dos nodos centrais que quando em contacto com uma noticias falsa, tem uma maior probabilidade de se tornarem disseminadores quando uma noticias falsa chega aos nodos limite de uma comunidade.

Com as figuras 4.3 4.4 4.5 é possível compreender como é que o "Community Health Assessment Model"[54] pode ser aplicado na prevenção e controlo da disseminação das notícias falsas. Nodos dentro da oval pontilhado são nodos centrais, os nodos dentro da oval sólida são nodos limite e os restantes nodos fora das ovais são nodos vizinhos.

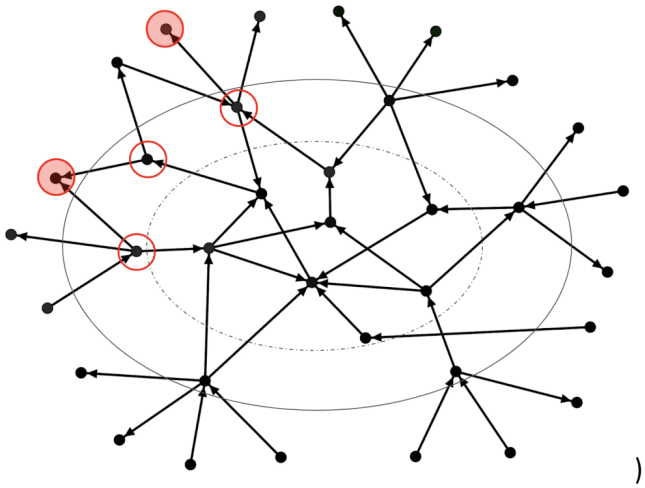
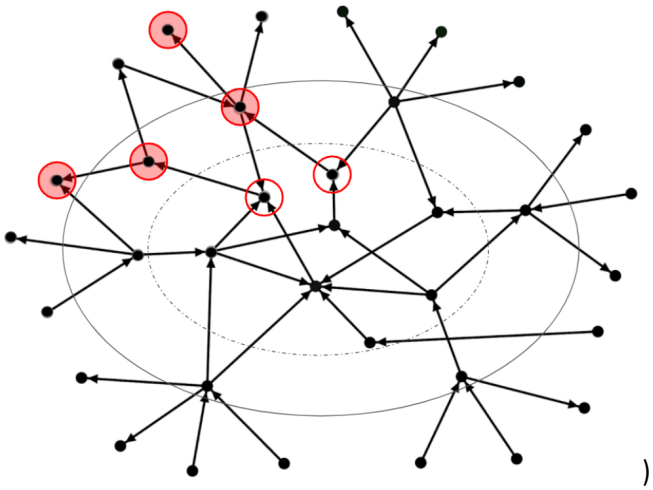


Figura 4.3: Notícias falsas chegam aos nodos vizinhos

Na figura 4.3 apresenta um exemplo em que as notícias falsas chegaram a 2 nodos vizinhos (os nodos

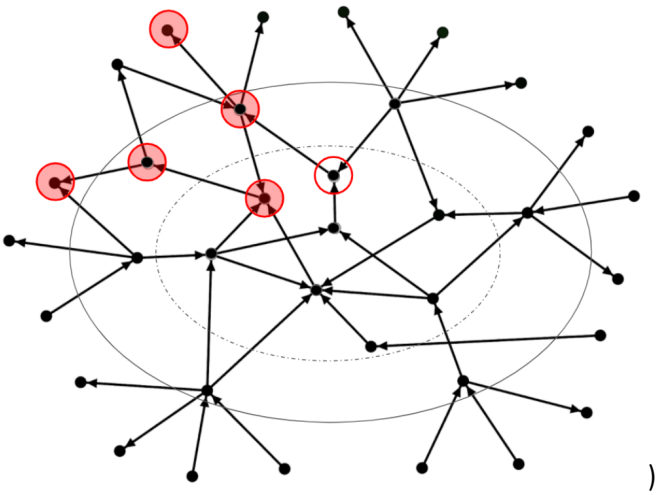
destacados a vermelho), o que faz com que 3 nodos limite (identificados pelo círculo vermelho) ficarem expostos a notícias falsas.



)

Figura 4.4: Notícias falsas chegam aos nodos limite.

Dos 3 nodos expostos à notícia falsa, 2 deles decidiram partilhar a notícia e assim tornarem-se disseminadores, o que marca o início da disseminação da notícia falsa dentro da comunidade. Como pode ser verificado na figura 4.4.



)

Figura 4.5: Notícias falsas chegam aos nodos centrais.

Com os nodos limite infectados e tornando-se disseminadores, põem em risco os nodos centrais de se tornarem também eles disseminadores da notícia falsa. Como pode ser verificado na figura 4.5 é o que aconteceu, assim pondo todos os nodos da comunidade em risco. Com a utilização do "Community Health Assessment Model"[54] é possível criar um modelo que consiga prever os nodos expostos e os nodos ainda não expostos, mas que tem maior probabilidade de se tornarem disseminadores. Com os resultados obtidos desses modelos podem depois ser criadas estratégias de mitigação da disseminação.

No artigo em que este modelo é sugerido, não existem exemplos de aplicação do modelo. Apenas existem sugestões de como o "Community Health Assessment Model"[54] pode ser usado para outros

contextos adicionais na disseminação de notícias falsas, idealmente os modelo sugeridos seriam utilizados dado que adicionam novas vertentes que poderiam ser úteis para o trabalho, como o modelo sugerido para a prevenção e controle da disseminação das notícias falsas seria útil para sugerir medidas de prevenção para os utilizadores.

4.2 Modelos de Detecção de Bots no Twitter

Na detecção de *bot* no Twitter existem vários modelos, mas dada a constante evolução e adaptação destas contas aos modelos de detecção existentes, existe a necessidade dos modelos receberem constantes atualizações de forma a este ainda serem relevantes e eficientes na identificação de contas *bot* no Twitter.

Dentro dos modelos de detecção de contas *bot*, o "Botometer"[65] [16] é generalizadamente mais utilizado e aplicado, tanto por utilizadores normais do Twitter como por trabalhos na área dos estudo do impacto das contas *bot* nas redes sociais [84] [5][65]. Este modelo supervisionado foi lançado em 2014, já recebeu 3 novas versões, estando atualmente na sua 4ª versão.

Um outro modelo que ira ser abordado é o "tweetbotornot2"[31], dado que numa recente trabalho que avaliava a deteção de contas *bot* [65], o "tweetbotornot2" apresentou bons resultados em comparação a 3ª versão do "Botometer". Este modelo tal como o "Botometer" também é supervisionado.

Entre os modelos supervisionados e os modelos não supervisionados, vai existir um maior foco nos supervisionados. Apesar dos modelos não supervisionados terem as suas vantagens como serem melhor a identificar publicações de contas *bot* que tem um comportamento semelhante ao humano, dado que estas publicações aparecerem para os modelos supervisionados como publicações isoladas e com comportamento humano, mas quando estas publicações são observadas em conjunto com outras publicações, facilmente são detectadas por utilizadores mais atentos ou por modelos de detecção de *bots* não supervisionados como "An Unsupervised Approach to Detect Spam Campaigns that Use Botnets on Twitter"[12] ou "DeBot: Twitter bot detection via warped correlation"[11]. Os modelos não supervisionados funcionam melhor para detectar publicações a imitar comportamento humano, por se focar primariamente no que os utilizadores publicam em comum, mas tem um problema de precisarem de um grande volume de dados para poder fazer classificações. Um outro problema é a constante evolução das contas *bot* e o conteúdo que publicam, o que cria dificuldades para os modelos se manterem constantemente atualizados. Os modelos supervisionados têm a vantagem de se conseguirem adaptar, dado que as suas classificações são supervisionadas. Com a recepção de *feedback* em massa por utilizadores os modelos supervisionados conseguem manter-se actualizados, mesmo que para isso demorem mais tempo a adaptar-se, o "Botometer", por exemplo, os utilizadores do "Botometer" podem avaliar as classificações feitas pelo modelo.

Um outro factor adicional, é que nas redes de disseminação de notícias nem todos os utilizadores que dão "retweet" a conteúdo publicado por contas *bot*, são também uma conta *bot*, muitos destes podem ser utilizadores normais, algo que num modelo não supervisionado que avalia o conteúdo publicado, as contas que fizeram "retweet" poderiam ser classificadas como uma contas *bot*. Nas redes de disseminação de

notícias também pode acontecer o inverso, ou seja as contas *bot* podem dar "retweet"a uma publicação feita por humanos, o que levaria estas contas que deram *bot* que deram "retweet" não serem identificadas. Sendo assim o foco será dado ao "Botometer"[65][16] e ao "tweetbotornot2"[31].

4.2.1 Tweetbotornot2

O "Tweetbotornot2"[31] é um modelo classificador supervisionado que considera cerca de 100 características, como por exemplo o conteúdo dos "tweets", conteúdo dos "retweets", data da publicação do "tweet", etc. As 100 características são organizadas em 3 categorias principais: atributos do utilizador, estatísticas das publicações e padrões do conteúdo das publicações.

O modelo apenas recolhe os últimos 200 "tweets" de cada utilizador, ou seja se um utilizador tiver mais de 200 "tweets" feitos, apenas os 200 mais recentes irão ser recolhidos para a classificação. Classificações feitas com menos do que 200 "tweets" não são consideradas confiáveis, excepto em casos em que a conta a ser classificada tenha feito menos do que 200 "tweets".

Infelizmente não existe muita informação em relação a como este modelo foi criado, que modelos classificadores de inteligência artificial ou que *datasets* que foram utilizados para treinar o modelo. Este também é o caso para outros modelos que foram tidos em consideração, como por exemplo o "Bot Sentinel"[8]. O "Tweetbotornot2" foi considerado dado o facto de que no "Detection of Novel Social Bots by Ensembles of Specialized Classifiers"[65], o "Tweetbotornot2" ter apresentado melhores resultados que o "Botometer"[16] [77] em alguns dos *datasets* que foram utilizados para medir o desempenho dos modelos na identificação de contas *bot*, mas dada a falta de informação em relação aos *datasets* que foram utilizados para treinar o modelo, não é possível concluir se o melhor desempenho se atribui a estes terem sido utilizados para o desenvolvimento do modelo ou se realmente o modelo para aqueles *datasets* tem melhor desempenho do que o "Botometer".

4.2.2 Botometer

Botometer é um modelo de identificação de *bots* no Twitter desenvolvido na Universidade de Indiana. O "Botometer" é baseado num modelo de aprendizagem de máquina supervisionado [16] [77], inicialmente o Botometer foi disponibilizado ao público em 2014, o acesso ao "Botometer" pode ser feito pelo web site [1] ou pela API. Nos últimos anos este modelo tem sido utilizado por vários estudos como "The spread of true and false news online"[80] e "Bots in the Twittersphere"[82], mas ainda mais importante o web site e a API são utilizados por utilizadores regulares todos os dias recebendo mais de 500 visitas por dia e cerca de um 250 000 pedidos por dia.

Quando uma conta é pesquisada utilizando o Botometer, este extrai mais de 1000 características relativas à conta do utilizador, os seus amigos, a estrutura da rede social envolta da conta do utilizador, padrões de atividade temporal, língua e sentimentos. As características são extraídas a partir dos dados disponíveis na API do Twitter, com os dados recolhidos estes são utilizados para atribuir uma pontuação,

[1] https://botometer.osome.iu.edu/

quanto mais alta for a pontuação, maior é a probabilidade de ser uma conta *bot* ou em parte *bot*. O Botometer é principalmente orientado para contas que utilizem inglês como língua principal, dado que algumas das características que o modelo procura são específicas a inglês. A conta depois de ser analisada oferece a pontuação global do utilizador ser *bot*, como também pontuações adicionais de diferentes comportamentos de contas *bot* como:

- **Echo-chamber**: Uma conta *bot* que participa activamente em grupos de *follow back* (grupos em que utilizadores seguem-se uns aos outros de forma planeada para pareceram mais credíveis) e partilham conteúdo político em grande volume.
- **Fake follower**: Conta *bot* que foi comprada para aumentar o número de seguidores de uma outra conta.
- **Financial**: *Bots* que fazem publicações utilizando *cashtags* (*cashtags* são como *hastags*, mas para o mundo financeiro).
- **Self declared**: Uma conta *bot* que está declarada na Botwiki [2].
- **Spammer**: Uma conta que publica conteúdo de *spam* identificada por datasets de *spambots*.
- **Other**: Uma conta que compartilha características com diversos outros bots obtidos de anotação manual, feedback dos utilizadores, etc.

O Botometer quando foi lançado em 2014 tinha o nome de "BotOrNot"e atualmente está na sua 4ª versão, este modelo é baseado em classificadores Random Forest [9]. A pontuação produzida pelo modelo deriva da fração de "árvores"que classificam a conta sob exame como um *bot*, como a pontuação é definida num intervalo de unidades existe o CAP (*Complete Automation Probability*). O CAP é essencialmente uma pontuação que deriva da probabilidade da conta ser *bot* ou não, ou seja uma conta com uma pontuação de 4,8 de 5 ou 0.96 de 1, mas um CAP de 90% existe uma probabilidade de 10% de não ser uma conta *bot* e uma probabilidade de 90% de o ser, isto permite que ao utilizar a API, dependo da taxa de erro pretendida concluir se uma conta é *bot* ou não. O CAP é também a pontuação que permite ao utilizadores da API identificar as contas *bots*, dado que quando uma conta é analisada é recebido um ficheiro JSON, como por exemplo IV, que fornece um conjunto de pontuações, apesar das outras pontuações fornecidas serem úteis para analisar o comportamento das contas, elas não fornecem informação suficiente para concluir que uma conta é *bot* ou não, então de forma de se conseguir concluir se uma conta é *bot* ou não, é fornecido o CAP, que simplesmente é a probabilidade que o modelo classifica se a conta é *bot* ou não.

Este modelo em comparação em modelos como "Detecting spammers and content promoters in online video social networks"[4], "Classification of twitter accounts into automated agents and human users"[20], "Deep Neural Networks for BOT detection"[36], "Bot Sentinel"[8] ou "Tweetbotornot2"[31],

[2] botwiki.org

tem a vantagem de receber atualizações frequentes, ser facilmente acessível por qualquer pessoa, e principalmente existirem artigos que detalham a maneira como o "Botometer"foi desenvolvido, testado e aplicado [16] [77] [83] [65]. Este último ponto é até o mais importante, dado que modelos como, "Bot Sentinel"[8] ou "Tweetbotornot2", são facilmente acessíveis, mas não apresentam qualquer explicação para como os modelos foram criados ou como foram treinados.

Uma outro ponto possitivo do "Botometer", é que este permitir que os utilizadores pesquisem abertamente a partir do web sites [3] contas de utilizadores do Twitter e assim recebam em tempo real uma pontuação da conta pesquisada, o que torna este modelo mais prático do que os modelos em que é necessário a instalação local, e com a pontuação obtida é também possível dar *feeedback* do resultado de forma a melhorar o modelo.

A possibilidade dar *feedback* na pontuação da conta é importante, como foi identificado em "Arming the public with artificial intelligence to counter social bots"[84], neste estudo é denotado que os modelos de identificação de *bots* supervisionados nas redes sociais, como o Botometer, tem problemas na detecção de publicações feita por contas *bot* que imitam o comportamento humano, muitas vezes estas publicações são feitas em massa e são coordenados por pessoas mal intencionadas de forma a influenciar utilizadores ou em ataques direcionados a utilizadores, figuras públicas, etc. Com a possibilidade de dar *feedback*, estas contas podem ser identificadas corretamente como *bot*.

4.3 Modelos Selecionados

Entre os modelos epidemiológico selecionados o "Community Health Assessment Model"[54] foi o escolhido, esta decisão baseia se em este medir a vulnerabilidade da disseminação de uma notícia entre utilizadores e entre comunidades de utilizadores (, nas aplicações mais diretas dos modelos epidemiológico a disseminação de notícias ou rumores apenas é avaliada a vulnerabilidade geral da rede), e a utilização do TSM [64] em conjunto com métrica de "credibilidade" do modelo fazem com que este consiga modelar de maneira mais realista as relações existentes entre utilizadores, e assim melhor modelar a vulnerabilidade dos intervenientes na rede. Outro ponto a ter em consideração é a forma como os dados foram recolhidos e aplicados no "Epidemiological Modeling of News and Rumors on Twitter", neste modelo os dados foram recolhidos com base em palavras chave e *hashtags*, a recolha utilizando palavras chave pode levar a recolha de "tweets" que não estejam relacionados com a notícia, o que vai influenciar os resultados finais, no "Community Health Assessment Model" os dados recolhidos foram feitos utilizando o URL de uma notícia, apesar de este tipo de recolha também ter problemas, como considerar que alguém que partilhe uma notícia falsa e no "tweet" desminta a mesma, vai ser incluído como disseminador apesar de estar a tentar ativamente desmentir a notícia e a tentar diminuir a disseminação da notícia falsa (este mesmo problema também está presente na recolha utilizando palavras chave e *hashtags*), utilizando o URL da notícia não há maneira de recolher um "tweet" que não esteja relacionado com a notícia em si. O número de contas *bot* nas redes de disseminação também são um fator, como foi verificado a presença

[3] https://botometer.osome.iu.edu/

das contas *bot* nos disseminadores de notícias é relativamente baixa, cerca de 0,284% em média (tabelas 5.25.5), então a maneira como "Community Health Assessment Model" tem em conta os relacionamentos entre os intervenientes nas redes de disseminação, foi considerado que dado o pequeno tamanho da população este seria mais adequado a medir o seu impacto, do que a aplicação de um modelo como SIS, SIR ou SEIZ. Um outro factor a ter em com é o suporte existente para o modelo, o "Community Health Assessment Model" é um modelo que é pode ser aplicado de maneira mais fácil do que os restantes, dado que o modelo está disponível online e é fornecido uma rede de disseminação, que funciona como uma referência para verificar se o modelo está a ser aplicado corretamente, enquanto que os modelos considerados não existe suporte online. Mesmo com a rede de disseminação como referência, a forma como os resultados dos modelos deteção das comunidades tem de estar formatados, o formato da tabela dos disseminadores e o significado de cada campo, ou até o formato que os resultados do "Trust in Social Media" devem estar de forma a serem aplicados ao modelo, nunca são apresentados e só por tentativa e erro é que foi possível concluir como estes podem ser aplicados ao "Community Health Assessment Model".

Dos modelos disponíveis para a detecção de contas *bot* no Twitter, o "Botometer"[65][16] foi o modelo selecionado. Esta decisão deve-se me em facto a este modelo está em constante desenvolvimento, ter um desempenho que é melhor ou igual a modelos comparáveis [65], existir informação disponível de como este modelo foi desenvolvido e os *datasets* que foram utilizados para teste e desenvolvimento, e o modelo ter ferramentas disponíveis que permitem que este seja facilmente aplicado. A escolha de um modelo que identifica as contas *bot* com base nos perfis em vez de do papel que aquela conta desempenhou numa rede, deve-se ao facto de que modelos baseados em comportamentos de rede necessitam de redes de grande dimensão [84] e mesmo com o tamanho das redes de disseminação de notícias recolhidas, estes modelos não seriam muito eficientes na detecção de contas *bot* com as redes recolhidas.

Capítulo 5

Aplicação dos Modelos Selecionados e Cálculo do Poder de Disseminação

Neste capítulo é abordado a aplicação dos modelos selecionados, o cálculo do poder de disseminação, a análise das redes de disseminação e as medidas de prevenção que os utilizadores contra a disseminação das notícias falsas.

Este capítulo é composto por 3 seções: "Aplicação dos Modelos Selecionados", nesta seção é abordada a aplicação do "Community Health Assessment Model"e do "Botometer", os resultados recolhidos desses mesmos modelos e como é que os resultados obtidos a partir do "Botometer"podem ser aplicados no "Community Health Assessment Model"; na seção "Poder de Disseminação e as Contas Bot", e é apresentado o impacto das contas *bot* nas disseminação notícias nas redes sociais, é feita uma análise a disseminação das notícias nas redes sociais, e como foi feito o calculado do poder de disseminação; na última seção do capítulo "Medidas de Prevenção dos Utilizadores Contra a Disseminação de Notícias Falsas", são apresentadas medidas para um utilizador proteger da disseminação das notícias falsas e que medidas as redes sociais podem aplicar de forma a proteger os seus utilizadores.

5.1 Aplicação dos Modelos Selecionados

Nesta secção é abordado como é que os modelos selecionados no capítulo 4 foram aplicados e são apresentados os resultados obtidos. A secção é composto em 2 subsecções, uma dedicada ao "Community Health Assessment Model"e outra ao "Botometer", nessas subseções são abordados com os dados foram recolhidos e tratados, a aplicação dos modelos com os dados recolhidos e os resultados obtidos com os modelos.

Durante a aplicação dos modelos selecionados foi utilizado apenas o Python, esta decisão deve-se ao facto de dos 2 modelos selecionados "Community Health Assessment Model"e "Botometer"o maior

suporte para a sua aplicação ser em Python e dado que o "Trust in Social Media"e a aplicação dos algoritmos de deteção de comunidades também poder ser feita utilizando o Python, este foi utilizado para todas as aplicações.

5.1.1 Community Health Assessment Model

Esta subseção é dedicada a explicar como o modelo "Community Health Assessment Model"foi aplicado no trabalho. O "Community Health Assessment Model"é um modelo utilizado em "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], as razões pela escolha deste modelo foram feitas no paragrafo 4.1.2.

Inicialmente é abordada a preparação dos dados, os algoritmos de deteção de comunidades que foram utilizados, e a utilização do "Trust In Social Media"(TSM). Depois é apresentado como o "Community Health Assessment Model"foi aplicado com os dados recolhidos e são apresentados os resultados obtidos.

5.1.1.1 Preparação dos dados

Os dados contidos no *dataset* necessitaram de ser manipulados de forma a serem aplicáveis aos algoritmos de deteção de comunidades, "Trust In Social Media"e "Community Health Assessment Model". De forma a estes manipulados foi criado um *script* em Python, que fornecendo um "id"de uma notícia gera uma *edge list* que contém os "ids"dos utilizadores e os "ids"dos seus seguidores e numa tabela que contém apenas dados dos utilizadores na rede que disseminaram a notícia.

O *script* recebe como argumento o "id"da notícia, este percorrer o *dataset* até encontrar a directório que tem um nome correspondente ao "id"da notícia, aí o *script* entra no directório que contém as publicações ("tweets"e "retweets") feitas pelos utilizadores, e recolhe os dados dos "tweets"e "retweets"e cria a tabela dos disseminadores, esta contém o "id", o nome de utilizador, o nome, o número de seguidores e o número de contas que o utilizador segue , a tabela é essencial para o cálculo do "Community Health Assessment Model".

Tabela 5.1: Exemplo de uma tabela dos disseminadores.

id	nome do utilizador	n° de seguidores	n° de contas seguidas
756542335098449920	desskysg	19	131
1305607525631504384	21Covik	485	387
1247250738297831425	digsbury	168	258

A *edge list* é gerada de maneira semelhante, o *script* acede a um directório que contém os ficheiros referentes aos seguidores dos utilizadores, os ficheiros contêm uma lista com os "ids"de todos os utilizadores que seguem um determinado utilizador. Assim o *script* cria pares ordenados associando o "id"de um utilizador aos "ids"dos seus seguidores, a *edge list* permite a representação das relações existentes na rede na forma de grafo orientado, em que os diferentes vértices representam os "ids"dos utilizadores

que estiveram em contacto com a notícia, e os arestas direcionadas representam a relação que um utilizador tem com o outro (se o segue ou se ambos se seguem). A *edge list* é utilizada pelos algoritmos de deteção de comunidades, "Trust In Social Media"e "Community Health Assessment Model".

Algoritmos de Deteção de Comunidades

De forma a aplicar o "Community Health Assessment Model"é necessário fornecer ficheiros que contêm comunidades, de forma a este calcular a vulnerabilidade nos nodos vizinho e nas comunidades de uma notícia. Dado que a partir dos dados contidos no *dataset* é difícil identificar essas comunidades foram utilizados algoritmos de deteção de comunidades. De forma aos resultados terem uma referência, foram utilizados os mesmo algoritmos de deteção de comunidades, que foram utilizados no "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], sendo esses os seguintes:

- **Louvain** [7]: baseia-se na melhoria da modularidade ao longo do processo do algoritmo. A modularidade é um valor numa escala, que vai de -0.5 até 1, que mede a densidade dos nodos limite dentre de uma comunidade em relação às comunidades externas dos nodos limite. Otimizando este valor teoricamente obtém-se o melhor agrupamento de nodos de uma rede, mas como percorrer todas as possíveis interações dos nodos em grupos não é algo prático, são utilizados algoritmos heurísticos.

- **Label Propagation** [48]: é um algoritmo de machine learning semi supervisionado que atribui etiquetas a nodos, as etiquetas correspondem às comunidades a que os nodos pertencem. No início da execução do algoritmo apenas um pequeno conjunto de nodos tem etiquetas, essas etiquetas são depois espalhadas pelos nodos ainda não identificados durante a execução do algoritmo. Consequentemente, grupos densamente conectados alcançam uma etiqueta comum rapidamente, os grupos depois continuam-se a expandir até atingirem um ponto em que se torna impossível.

- **Infomap** [63]: é um algoritmo de *cluster* para localizar comunidades numa rede, baseado na equação de *map*. A equação de *map* tenta encontrar um equilíbrio entre encontrar *clusters* em redes e minimizar o comprimento das "caminhadas aleatória"(random walk)[46]. O *random walker* anda pelos nodos da rede, atribuindo um peso a esses nodos, o quanto mais "pesada"é uma conexão de um nodo, o mais provável é o "walker" voltar a utilizar essa mesma conexão para chegar a um outro nodo, com o objectivo de criar um *cluster* em que o *random walker* fica o maior tempo possível.

De forma a aplicar os algoritmos de deteção de comunidades foi criado um *script* em Python que utiliza os pacote CDlib[1] e o Python-Louvain[2], a execução do *script* permite obter as comunidades e

[1] https://cdlib.readthedocs.io/en/latest/
[2] https://github.com/taynaud/python-louvain

armazená-las em lista em ficheiros. O *script* recebe a *edge list* da notícia a ser avaliada, a *edge list* é depois aplicada aos algoritmos de deteção de comunidades, dado que a *edge list* é um grafo orientado este que representa as relações entre os utilizadores numa rede de disseminação de notícias, isto permite assim gerar as comunidades com os algoritmos. A execução do *script* dura entre 1 a 3 horas, isto deve-se principalmente aos cálculos necessários para detectaram comunidades e ao tamanho das redes que são aplicados a estes algoritmos.

Trust In Social Media

Uma outra necessidade é calcular o TSM (Trust In Social Media), para o cálculo do TSM foi utilizado um *script* em Python com o algoritmo. O *script* utiliza a *edge list* de uma notícia, que é um grafo orientado que representa as relações entre os utilizadores numa rede de disseminação de notícias, para calcular a confiança e confiabilidade de cada utilizador, o *output* é armazenado num ficheiro. Uma descrição mais detalhada de como o modelo efetua os cálculos da confiança e confiabilidade foi feita na subseção 4.1.2.

5.1.1.2 Aplicação do Modelo

Utilizando os dados obtidos com a preparação dos dados, os algoritmos de deteção de comunidades e "Trust In Social Media"é possível utilizar o "Community Health Assessment Model", de forma a ser posto em prática foi utilizado um *script* em Python fornecido publicamente [3] por um dos autores do modelo. O modelo de forma a ser executado necessita de receber como *input* a *edge list* de uma notícia, a tabela dos disseminadores dessa mesma notícia, os resultados dos algoritmos de deteção de comunidades e os resultado do "Trust In Social Media"obtidos da *edge list* da notícia. Com os dados recebidos o modelo calcula a vulnerabilidade dos nodos limite e das comunidades.

A execução do *script* pode demorar entre 4 horas a 3 ou 4 dias, dependendo do número de intervenientes na rede disseminação da notícia. A duração do *script* deve-se na sua maior parte ao cálculo da credibilidade. Os resultados sobre a rede podem ser visualizados nas tabelas II.1, II.5, II.3 e II.7.

O L, I e LP correspondem aos algoritmos utilizados para gerar as comunidades, respectivamente Louvain, Info Map e Label Propagation, o AP é precisão média para a determinada comunidade e o MAP são os os resultados da função MAP aplicada aos quinze *melhores* resultados.

Os resultados obtidos com a execução do modelo utilizando os dado recolhidos a vulnerabilidade dos nodos limite vai em conta aos dados obtidos em "Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model"[54], mas a vulnerabilidade das comunidades são diferentes do esperado.

Na vulnerabilidade dos nodos limite nas notícias verdadeiras existem dois resultados que são atípicos, nomeadamente o *politifact_18* e *politifact_21*. No caso da *politifact_21* o resultado pode ser explicado pelo facto de que a notícia apesar de ser verdadeira, foi inicialmente relatada por um site que normalmente

[3] https://github.com/BhavtoshRath/Vulnerability-Metrics

partilha notícias falsas, o que pode ter levado a que esta notícia tenha tido o mesmo tipo de comportamento que as notícias falsas. Em relação a *politifact_18*, a notícia foi partilhada por um tablóide, o que pode justificar o seu comportamento.

A vulnerabilidade nas comunidades não permite a distinção entre as notícias falsas das verdadeiras em 2 dos 3 algoritmos de deteção de comunidades com os dados recolhidos, apenas no "Infomap"é possível fazer a distinção entre as notícias.

5.1.2 Botometer

Nesta subseção é abordada a aplicação do "Bototmeter". Inicialmente é apresentado como é que os dados foram recolhidos e tratados de forma a estes serem aplicados no modelo. Também são apresentados alguns dos dados mais relevantes, sendo que os restantes estão disponíveis em anexo.

5.1.2.1 Recolha de Dados

Para recolher os dados referentes a disseminadores que são contas *bot* foi seleccionado a utilização do Botometer, as razões pela foi feita esta escolha foram abordadas na seção 4.2.

De forma a recolha de dados ser feita foi utilizando a API disponibilizada pelo Botometer, entre os três planos disponíveis: "grátis", "pro"e "ultra", foi decidido utilizar o "pro"dado que não tem custos, mas tem um limite de dois mil pedidos por dia com um custo adicional de 0.001 dólares por pedidos extra. Então foi criado um *script* em Python para fazer os pedidos de forma automatizada, para ajudar a criar esse efeito foi utilizado o pacote fornecido pelo Botometer que tem dependência com o Tweepy, um pacote que permite obter dados a API do Twitter. Com este pacote é possível fazer pedidos em massa, mas dado o limite da API de dois mil pedidos, estes tiveram que ser limitados.

Com o limite em questão, houve a necessidade de apenas verificar as contas dos utilizadores que participaram ativamente na disseminação da notícia, ou seja aqueles que fizeram publicações que continham aquela notícia, esta decisão foi tomada porque a maior parte das redes tem perto ou mais de um milhão de intervenientes, o que tornava impossível a recolha de dados.

De forma ao *script* obter as avaliações das contas, é lhe fornecido a tabela dos disseminadores de uma notícia, o *scipt* utiliza o campo "id"disponível na tabela dos disseminadores da notícia, que corresponde ao número de identificação do utilizador no "Twitter", e faz um pedido síncrono a API do "Botometer"se a conta com aquele "id", a API como resposta envia um ficheiro JSON contendo os resultados da avaliação da conta que contêm a classificações CAP, echo-chamber, fake follower, financial, self declared e spammer, o ficheiro JSON é armazenado num diretório e tem uma estrutura semelhante à presente na listagem IV. O *scipt* vai fazendo os pedidos até terminar o número de "ids"existentes na tabela, os resultados obtidos são mais tarde processados por um outro *script*.

Dado o volume dos dados recolhidos para tratar os dados foi necessário criar um outro *scirpt* em Python de forma a automatizar o processo. O *script* em questão tinha a tarefa de com dados obtidos com o "Botometer"identificar as contas *bot*, calcular as médias das classificações recebidas, listar as contas

bot e categorizar as mesmas em listas referentes aos diferentes tipos de classificação (echo-chamber, fake follower, financial, self declared e spammer) e criar *edge lists* e tabelas de disseminadores (iguais as demonstradas na tabela 5.1) em que remove as contas *bot* da rede original.

A identificação das contas *bot* a partir dos dados é feita utilizando a classificação CAP dos utilizadores, dado que esta é correspondente a probabilidade de uma conta ser *bot*, a classificação é feita utilizando dois valores, nomeadamente uma classificação CAP acima ou igual a 96% e 90%. Foram escolhidas estas duas pontuações para demonstrar como uma pequena diferença de 6% pode ter impacto quando se fala de probabilidade de detecção de contas *bot*, sendo uma conta com 96% de CAP uma conta quase certa como *bot* e uma com 90% de CAP uma conta com alta probabilidade de ser *bot*.

O *script* também calcula as médias das classificações fornecidas pelo "Botometer", e também cria duas tabelas de disseminadores iguais as demonstradas na tabela 5.1 e duas *edge lists* da rede de disseminação da notícia, em que apenas contém as contas que o modelo não considerou como sendo contas *bot*, ou seja removendo as contas *bot* com classificações CAP acima ou igual a 96% e 90%.

Os dados obtidos das contas *bot* e do impacto das contas *bots* com a classificação CAP de 90% e 96% podem ser visualizados na tabelas seguintes:

Tabela 5.2: Resultados do Botometer com CAP 96 sobre as notícias Falsas.

	Contas Bot	Disseminadores	%
Politifact_3	2.00	2,832.00	0.071
Politifact_4	6.00	4,685.00	0.128
Politifact_5	0.00	1583.00	0.00
Politifact_8	6.00	1,222.00	0.491
Politifact_14	3.00	1,439.00	0.208
Politifact_17	5.00	1,003.00	0.499
Politifact_20	11.00	4,708.00	0.234
Politifact_27	0.00	2,817.00	0.000
Media			0.272

Tabela 5.3: Resultados do Botometer com CAP 90 sobre as notícias Falsas.

	Contas Bot	Disseminadores	%
Politifact_3	20.00	2,832	0.706
Politifact_4	51.00	4,685	1.089
Politifact_5	2.00	1,583	0.126
Politifact_8	30.00	1,222	2.455
Politifact_14	63.00	1,439	4.378
Politifact_17	31.00	1,003	3.091
Politifact_20	102.00	4,708	2.167
Politifact_27	4.00	2,817	0.142
Média			1.769

Tabela 5.4: Médias dos resultados Botometer sobre as notícias Falsas.

	Cap Score	Echo Chamber	Spammer	Fake Followers	Financial
Politifact_3	0.699	0.290	0.047	0.200	0.056
Politifact_4	0.691	0.273	0.057	0.198	0.051
Politifact_5	0.542	0.132	0.043	0.130	0.037
Politifact_8	0.644	0.373	0.037	0.192	0.057
Politifact_14	0.707	0.461	0.043	0.231	0.051
Politifact_17	0.645	0.375	0.052	0.200	0.040
Politifact_20	0.671	0.358	0.038	0.206	0.072
Politifact_27	0.698	0.275	0.048	0.195	0.064
Média Total	0.662	0.317	0.046	0.194	0.053

Tabela 5.5: Resultados do Botometer com CAP 96 sobre as notícias Verdadeiras.

	Contas Bot	Disseminadores	%
Politifact_6	5	2,753	0.182
Politifact_7	3	854	0.351
Politifact_9	3	1,800	0.167
Politifact_12	3	1,886	0.159
Politifact_18	3	4,261	0.070
Politifact_19	0	500	0.000
Politifact_21	12	4,761	0.252
Politifact_22	6	659	0.910
Politifact_24	2	731	0.274
Media			0.296

Tabela 5.6: Resultados do Botometer com CAP 90 sobre as notícias Verdadeiras.

	Contas Bot	Disseminadores	%
Politifact_6	24	2,753	0.872
Politifact_7	15	854	1.756
Politifact_9	19	1,800	1.056
Politifact_12	16	1,886	0.848
Politifact_18	13	4,261	0.305
Politifact_19	3	500	0.600
Politifact_21	103	4,761	2.163
Politifact_22	30	659	4.552
Politifact_24	13	731	1.778
Média			1.666

Tabela 5.7: Médias dos resultados Botometer sobre as notícias Verdadeiras.

	Cap Score	Echo Chamber	Spammer	Fake Followers	Financial
Politifact_6	0.526	0.174	0.038	0.137	0.039
Politifact_7	0.610	0.196	0.057	0.179	0.051
Politifact_9	0.618	0.228	0.053	0.183	0.048
Politifact_12	0.581	0.199	0.042	0.164	0.051
Politifact_18	0.636	0.154	0.067	0.215	0.061
Politifact_19	0.571	0.165	0.048	0.162	0.045
Politifact_21	0.644	0.355	0.027	0.189	0.069
Politifact_22	0.659	0.231	0.086	0.224	0.048
Politifact_24	0.655	0.341	0.035	0.186	0.057
	Cap Score	Echo Chamber	Spammer	Fake Followers	Financial
Média Total	0.611	0.227	0.050	0.182	0.052

Os resultados obtidos a partir do "Botometer"não apresentam uma grande distinção entre as notícias falsas e as verdadeiras. A presença percentual das contas *bot* disseminadoras é quase idêntica como pode ser verificado nas tabelas 5.2, 5.3, 5.5 e 5.6.

As pontuações médias entre as contas disseminadoras também não apresentam grandes diferenças, a pontuação de média de CAP das notícias recolhidas é quase idêntica, com as notícias falsas apresentando uma pontuação ligeiramente mais alta (tabelas 5.4 e 5.7), mas não suficiente para se conseguir fazer uma distinção entre uma notícia falsa de uma verdadeira a partir do CAP médio.

Por outro lado as contas disseminadoras das notícias falsas apresentam uma maior pontuação de "Echo Chamber"em média (tabelas 5.4 e 5.7), esta pontuação é 9% em média superior as notícias verdadeiras, o que indica que existe um maior comportamento de "Echo Chamber"na disseminação das notícias falsas.

Nas restantes pontuações não existe uma diferença suficiente para diferenciar as notícias falsas das verdadeiras.

5.1.3 Aplicação do resultados do Botometer no Community Health Assessment Model

Com os dados recolhidos com o "Botometer"foi possível identificar as contas *bot* nas redes de disseminação das notícias contidas no *dataset*, mas a identificação das contas não permite o cálculo do impacto das contas *bot* na disseminação das notícias nas redes sociais. Então de forma a medir o impacto que estas contas têm na vulnerabilidade dos utilizadores das redes sociais é utilizado o "Community Health Assessment".

Como foi referido na seção 5.1.2 o *scipt* criado para o processamento dos resultados obtidos do "Botometer"fornece uma *edge list* da disseminação de uma notícia. Estas *edge lists* representam um grafo orientado, composto por pares ordenados, cada par ordenado corresponde a um utilizador que é seguido por um outro utilizador. Como foi referido na subseção 5.1.1, estas *edge lists* que foram utilizadas ao longo do trabalho representam as relações existentes entre os utilizadores, e como tal são essenciais para todos o processo de aplicação do "Community Health Assessment", desde a deteção das comunidades, como no cálculo do "Trust in Social Media"e na própria aplicação do "Community Health Assessment". Ou seja removendo as contas *bot* das redes de disseminação e como o "Community Health Assessment"se baseia no conceito de que é necessário existir uma maior confiança entre os intervenientes nas redes de disseminação das notícias falsas do que nas verdadeiras, ao remover estas contas *bot* das *edge lists* da notícias, as relações de confiança e influência que estas têm com os utilizadores também deixariam de existir no "Trust in Social Media"e no "Community Health Assessment"e assim seria possível medir a vulnerabilidade de um utilizador disseminar uma notícia depois de entrar em contacto com ela, e assim medir a influencia e impacto que as contas *bot* tem nas redes de disseminação.

Então com as *edge lists* das notícias fornecidas pelo *scipt* de processamento dos resultados obtidos do "Botometer", são aplicadas aos algoritmos de deteção de comunidades e ao modelo "Trust in Social Media". Com os resultados obtidos de ambos modelos em conjunção com a *edge list* e a tabela de disseminadores gerada pelo *scipt* de processamento dos resultados obtidos do "Botometer"de uma determinada notícia, estes são aplicados ao "Community Health Assessment"apresentando assim os resultados da influência e impacto que as contas *bot* têm nas redes de disseminação das notícias das redes sociais. (Os resultados e a análise dos mesmo é feita na seção 5.2).

5.2 Poder de Disseminação e as Contas *Bot*

Com a aplicação dos modelos selecionados e com os dados obtidos, esta secção foca-se na análise dos dados obtidos e as conclusões que podem ser tiradas com esses mesmos dados. A secção é composta por 3 subseções, nas subseções "Impacto das contas *bot* nas vulnerabilidade dos utilizadores e das

comunidades" e "Poder de Disseminação e Distribuição dos Utilizadores", são apresentados os resultados obtidos e estes resultados são depois analisados. Na subseção "Medidas de Prevenção dos Utilizadores Contra a Disseminação de Notícias Falsas" são utilizados os dados recolhidos ao longo do trabalho de forma a fornecer medidas que possam proteger os utilizadores da disseminação de notícias falsas.

Nesta seção também foi utilizado extensivamente o Python, neste caso apenas deve-se a maior familiaridade com o Python.

5.2.1 Impacto das Contas *Bot* nas Vulnerabilidade dos Utilizadores e das Comunidades

Com os dados recolhidos com a aplicação dos resultados obtidos com o "Botometer" no "Community Health Assessment" feita na seção 5.1.3, é possível tentar deduzir o impacto que as contas *bot* possam têm vulnerabilidade dos utilizadores e nas comunidades. Para esse efeito, simplesmente foi calculada a diferença entre os resultados obtidos entre a execução do "Community Health Assessment" em redes de disseminação das notícias em que as contas *bots* foram removidas e as redes de disseminação das notícias. Apesar de simplesmente calcular a diferença entre os dois resultados parece ser simples, dada a maneira como o modelo "Community Health Assessment" é aplicado a remoção das contas *bot* nas redes de disseminação das notícias, tem um impacto significativo dado que as relações que os utilizadores de confiança que os utilizadores têm com as contas *bot* deixam de estar representadas. Os resultados das diferenças estão nas tabelas: III.1, III.5, III.24, III.3, III.2, III.7 e III.4.

A partir dos dados apresentados é possível concluir que as contas *bot* apresentam ser um factor que vá influenciar a vulnerabilidade nos nodos limite em particular utilizando o CAP a 96%, apesar de em alguma noticias a diferença ser mais significativa, como na *Politifact_8* na tabela III.1, em geral houve uma redução entre 2,8% a 0,7% (este 0,7% é relevante dado que foi obtido a partir do Label Propagation, que não utiliza *random walks*, logo a diferença é atribuída a remoção de *bots*). É de notar também que foi calculada uma diferença maior nas noticias falsas do que nas verdadeiras, até que nas verdadeiras houve um aumento da vulnerabilidade de 3,2% para o Louvain e de 0,1% para o Label Propagation, mas uma redução de 1,8% para o Infomap.

Por um outro lado os resultados do CAP a 90% (tabela III.24) não foram muito melhores do que com o CAP a 96%, havendo uma redução nas notícias falsas da vulnerabilidade nos nodos limite, mas sendo bastante reduzida em comparação com a do CAP a 96%, havendo apenas uma redução de 2,3% no Louvain, uma redução bastante insignificante de 0,2% no Infomap e uma redução de 0,5% no Label Propagation. Nas notícias verdadeiras houve uma melhoria em comparação com os resultados do CAP 96% apresentando o Louvain um ganho de 0,9%, o Infomap uma redução de 1,5% e no estranho caso do Louvain uma redução de 8,9%, sendo esta principalmente devida ao facto de que na notícia *Politifact_7* houve uma redução de 72,8%, se este resultado for removido existe na uma redução de 1,8%.

É possível concluir com os dados recolhidos, que a diferença da vulnerabilidade nos nodos limite possa ajudar na detecção de uma notícia falsa, mas não é um factor principal na distinção entre notícias

falsas e verdadeira, a diferença da vulnerabilidade nos nodos limite é mais um factor complementar, que em caso de dúvida pode ajudar na diferenciação.

Em relação a diferença na vulnerabilidade das comunidades a remoção das contas *bot* teve um efeito mais negativo do que positivo, sendo os resultados mais uma vez melhores no CAP a 96% do que no CAP a 90%. Os resultados apresentados é impossível inferir se uma notícia é falsa ou verdadeira utilizando apenas as diferenças, mas nos resultado com o CAP 90% houve em geral uma redução sendo a mais significativa de 6,9% nas comunidades geradas pelo Label Propagation nas notícias falsas, as tabelas III.29 e III.31 apresentam em geral uma redução significativa, o que pelo menos indica de que a remoção das contas *bot* reduz a vulnerabilidade das comunidades, estes dados podem indicar de que os *bots* são responsáveis entre 2,9% a 6,9% pela disseminação de notícias independentemente da veracidade das mesmas.

Com o Community Health Assessment Model é possível compreender melhor o impacto que as contas *bot* tem na disseminação de notícias dentro das comunidade e na propagação entre comunidades, sendo que o impacto destas contas parece ser maior na disseminação dentro das comunidades, do que na propagação entre as diferentes comunidades de uma rede. Este comportamento pode ser devido a utilização delas em comunidade de *echo chamber* ser mais significativo, tendo elas assim uma maior importância na disseminação dentro destas comunidades fechadas, do que entre comunidades vizinhas ou até desconhecidas. Conclui-se então que as contas *bot* podem ter um impacto entre 2,9% a 6,9% na disseminação de notícias nas comunidades em que se encontram, e um impacto entre 1,5% a 2,8% nas disseminação entre comunidades.

É de notar que as contas *bot* apenas representam 0,272% a 0,296% (tabelas 5.2 e 5.5) nas notícias falsas e verdadeiras com o CAP a 96% e para o CAP a 90% representam 1,769% e 1,666% (tabelas 5.3 e 5.6) para as notícias falsas e verdadeiras respectivamente. E que os resultados com a remoção completa das contas *bot* com CAP a 90% ou a 96%, não foi completamente concluído, dadas as limitações na API do Botometer só foi feito as contas que participaram de alguma forma na disseminação de notícias e não da rede completa. Uma análise na rede completa iria certamente afetar os resultados todos obtidos. Com uma análise completa da rede também iriam surgir outros problemas, nomeadamente o espaço que o *dataset* ocuparia.

5.2.2 Poder de Disseminação e Distribuição dos Utilizadores

De forma a identificar melhor o papel das contas *bot* nas redes sociais, foi feito um estudo sobre a distribuição dos utilizadores que participaram ativamente na disseminação de notícias e o seu impacto na disseminação de informação. Este estudo foi feito utilizando vários *scripts* em *Python* de forma a automatizar o processo de recolha e tratamento de dados, com resultados obtidos foram criadas tabelas de forma a estes resultados serem mais fáceis de compreender.

O poder de disseminação foca-se em analisar como é que um utilizador dentro de uma determinada

distribuição dissemina as notícias, o poder de disseminação é então constituído por 3 pontuações: pontuação do alcance, pontuação da influência e Pontuação do alcance nos seguidores. O objectivo destas pontuações é tentar medir se existem diferenças na maneira como os utilizadores disseminam as notícias e interagem uns com os outros dependendo se a notícia é falsa ou verdadeira. Com estas pontuações é então possível verificar, dependendo da distribuição, qual é o impacto ou poder de disseminação dos utilizadores dependendo do número de seguidores.

5.2.2.1 Distribuição dos Utilizadores

A distribuição dos utilizadores é referente ao número de seguidores que estes tem, este foram agrupados em diferentes grupos baseando-se no número de seguidores:

- 0 - 100 seguidores;
- 100 - 1k seguidores;
- 1k - 5k seguidores;
- 5k - 10k seguidores;
- 10k - 50k seguidores;
- 50k - 100k seguidores;
- 100k - 500k seguidores;
- 500k - 1M seguidores;
- 1M ou mais seguidores.

Eles foram agrupados desta maneira dado que os seus seguidores estão sempre em "contacto" directo quando este partilha uma notícia.

Para além da distribuição dos utilizadores por seguidores também foi observado: os tweets e retweets feitos pelos utilizadores, os utilizadores seguidos por *bots*, os utilizadores que seguem *bots* e a distribuição das contas *bot* pelos seus seguidores. Esta distribuição foi feita para tentar encontrar relações entre os utilizadores e as contas *bot* e as diferenças entre os comportamentos dos disseminadores que partilham notícias falsas e os que partilham notícias verdadeiras.

Para obter os dados foi criado um *scipt* em Python que utilizando os dados recolhidos do *dataset* e os dados obtidos do "Botometer", o *script* percorrer os dados e analisa os dados consoante os diferentes campos que foram selecionados. Depois de feita a analise o *script* exporta os dados em ficheiros "CSV".

Um primeiro factor que diferencia estes dois tipos de notícias é a distribuição do número de seguidores ao longo das redes de partilha, como se pode verificar nas tabelas III.9 e III.34 as noticias falsas tem uma maior percentagem de utilizadores com um menor número de seguidores em relação aos utilizadores das

notícias verdadeiras, que acrescentam uma melhor distribuição do número de seguidores ao longo dos diferentes grupos. Isto é relevante dado que demonstra que os disseminadores de notícias falsas tem uma menor "audiência" a que estes conseguem disseminar directamente, com este dado em conjunto com os dados do Community Health Assessment Model sobre as notícias falsas (tabela II.1), que indica que existe uma maior confiança entre os disseminadores e os seus seguidores, leva a concluir que a disseminação de notícias falsas depende das comunidades de *echo chamber*.

Um outro factor a ter consideração com estes dados é a diferença existente entre o número de disseminadores que são seguidos por contas *bot* e o número de disseminadores que seguem estas contas *bot* dependendo da veracidade da noticia. Nas tabelas III.10, III.12, III.14 e III.15 é possível verificar a discrepância existente entre o número médio de disseminadores seguidos por contas *bot* no diferente tipo de noticia, sendo que as noticias falsas apresentam ter em geral quase o dobro de disseminadores seguidos por contas *bot*. Sendo esta diferença ainda mais notável nas tabelas com o CAP a 96. Este comportamento também é verificável nos disseminadores que seguem contas *bot*, como pode ser observado nas tabelas III.16, III.17, III.14 e III.15. Isto leva a indicar que as contas *bot* tem um papel mais ativo nas redes de disseminação de noticias falsas do que nas verdadeiras e os disseminadores dependem mais de contas *bot* ou existe uma maior confiança que os disseminadores das noticias falsas nas contas *bot* e que existe uma associação directa entre as contas *bot* e os disseminadores de noticias falsas.

5.2.2.2 Poder de Disseminação dos Utilizadores para com os Seus Seguidores

Dados os dados apresentados nas tabelas III.10, III.12, III.16 e III.17, foi calculado o alcance que as publicações dos disseminadores tem nos seus seguidores e na rede. Esse cálculo foi utilizado o conceito de uma notícia falsa dentro de uma rede social tem um comportamento semelhante a uma doença epidemiológica, no sentido em que tal como uma doença epidemiológica as notícias falsas têm uma maior facilidade de propagação, e as notícias falsas nas redes sociais propagam-se de utilizador em utilizador, ou seja quando um utilizador faz uma publicação de uma notícia falsa, este expõem os seus seguidores a notícia falsa, se um seguidor ficar infectado este faz "retweet" da publicação, o seguidor ao fazer o "retweet" expõe também os seus seguidores a notícia falsa, podendo levar a mais infectados. Em comparação uma doença epidemiológica propaga-se de maneira semelhante, uma pessoa infectada ao entrar em contacto com um grupo de pessoas, põem estas em risco de ficarem infectadas, se uma dessas fica infectada ao entrar em contacto com um grupo diferente este expõe também elas a doença.

Como foi referido acima 5.2.2 o poder de disseminação foi medido em 3 pontuações, de forma a tentar compreender melhor como é que as notícias foram disseminadas com base nas distribuições definidas. De forma a ser mais fácil a abordar as 3 pontuações escolhidas, é necessário compreender como é calculada a dimensão da rede D, esta é a soma de todos os possíveis intervenientes na rede, ou seja os utilizadores que disseminaram e os seguidores desses mesmo, e também compreender o conceito de "utilizadores em contacto"(uc), o que se resume a soma dos utilizadores que fizeram "retweets", os favoritos e seguidores de um utilizador que tiveram em contacto uma notícia que foi publicada por um

utilizador. As 3 pontuações escolhidas podem ser definidas por:

- Pontuação do alcance nos seguidores (DC) - A média do resultado do somatório dos "utilizadores em contacto"uc dividindo pelos seguidores (s) de um utilizador (n) numa determinada distribuição (k).

$$DC = \frac{\sum_{k}^{n=1} \frac{uc_n}{s_n}}{k} \tag{5.1}$$

- Pontuação do alcance (PC) - Esta pontuação mede com base numa distribuição(k), a média do somatório dos "utilizadores em contacto"(uc) de um utilizador (n) dividindo esse pela dimensão da rede (D).

$$PC = \frac{\sum_{k}^{n=1} \frac{uc_n}{D}}{k} \tag{5.2}$$

- Pontuação da influência (PI) - A média do somatório dos utilizadores influenciados (ui) por um utilizador(n) numa determinada distribuição (k), ou seja os utilizadores que deram "retweet"e deram favorito, a dividir pelo seu número de seguidores (ns).

$$PI = \frac{\sum_{k}^{n=1} \frac{ui_n}{ns_n}}{k} \tag{5.3}$$

As pontuações foram criadas com base na vulnerabilidade dos nodos limite e das comunidades do "Community Health Assessment Model" em que indica que nas notícias falsas existe uma maior vulnerabilidade nos nodos nas notícias falsas, do que nos nodos nas notícias verdadeiras.

Com os dados recolhidos é possível concluir que as notícias falsas apresentam em média um melhor alcance, tanto nos para com o seus seguidores como em geral, como é possível verificar nas tabelas III.18 e III.19, dado que em média nas notícias falsas um disseminador tem um alcance 232,718 vezes maior do que o número dos seus seguidores, enquanto que um disseminador numa notícia verdadeira apenas alcança 150,357 vezes mais do que o número dos seus seguidores. A pontuação da influência é superior nas notícias verdadeiras as falsas, mas em deve-se principalmente a facto da notícia "Politifact_9" ter uma pontuação de influência extremamente alta em comparação com qualquer notícia recolhida, removendo os resultados da "Politifact_9" a média da pontuação da influência das notícias verdadeiras desce para 0.230, o que pode indicar que a influência nas notícias falsas também é superior ao das notícias verdadeiras.

Este maior alcance com os seguidores e maior influência removendo os dados do "Politifact_9" nas notícias falsas, pode dever-se em parte ao maior número de "tweets" e "retweets" que existem na redes de disseminação de notícias falsas, estes dados podem ser verificados nas tabelas III.35 e III.36 e nas tabelas III.37 e III.38. Apesar do maior número de "tweets" e "retweets" o maior alcance com os seguidores e maior influência, removendo os dados do "Politifact_9", estes dados também podem indicar que existe

uma maior confiança entre os utilizadores que estão presentes nas redes de disseminação de notícias falsas, dado o maior alcance dentro da comunidade e a maior pontuação de influência. Também indica que os disseminadores de notícias falsas tem um poder de disseminação maior, do que os disseminadores de notícias verdadeiras, dado que apesar destes terem um menor alcance os utilizador a que as notícias falsas alcançam têm uma maior probabilidade de serem influenciados pela notícia e também eles participarem na disseminação da notícia falsa.

Um outro dado interessante nas notícias falsas é a maior pontuação de influência que as contas com seguidores entre os 0-100 têm na rede (tabela III.20), estas contas têm sempre uma maior pontuação influência dada a forma como o cálculo é feito, mas as notícias falsas em geral apresentar uma maior pontuação de influência do que as verdadeiras para a essa mesma distribuição (tabela III.21).

Estes factores em conjunto com uma maior presença de seguidores *bot* nas notícias falsas (tabelas III.16 III.17 III.14 III.15), pode indicar que os *bots* tem um papel mais importante na disseminação de notícias falsas, sendo o papel deles ajudar na disseminação ou ajudar um utilizador mal intencionado parecer mais legítimo, dado que parece ter um maior número de seguidores e existe uma maior interação com as suas publicações.

5.3 Medidas de Prevenção dos Utilizadores Contra a Disseminação de Notícias Falsas

Com os dados recolhidos e analisados, nesta secção são abordadas as medidas de prevenção contra a disseminação de notícias falsas. As medidas de prevenção são sugeridas aos utilizadores das redes sociais e também as redes sociais, de forma a promover uma utilização mais segura das redes sociais.

Medidas de Prevenção Tomadas Pelos Utilizadores

Com os dados apresentados a um conjunto de medidas que os utilizadores podem aplicar de forma a se protegerem da disseminação de notícias falsas, nomeadamente:

- **Utilização de modelos de detecção de *bots* publicamente disponíveis como o Botometer[16] [77] ou o Bot Sentinel[8]**. Este modelos permitem que um utilizador do Twitter consiga verificar se a conta com que estão a interagir é de facto uma conta *bot* ou não, e dado que as contas *bot* são mais activas e tem um impacto maior nas notícias falsas, este pode ser um indicador de que a notícia em questão pode ser falsa.

- **Verificar a fonte da notícia**. As notícias *Politifact_21* e *Politifact_24* são sobre o mesmo tópico, mas partilhadas por web sites diferentes e isso afeta drasticamente os resultados como é possível verificar na tabela II.5 de uma forma que o utilizador pode ser induzido em erro, assim é preferível utilizar um motor de busca e pesquisar um titulo da noticia para verificar a veracidade da mesma.

A utilização de web site de *fact checking* também pode ser uma boa maneira de verificar uma notícia.

- **Verificar os perfis dos utilizadores que compartilharam uma publicação**. Caso existam dúvidas em relação a veracidade de uma publicação, o Twitter permite que se consiga verificar os utilizadores que compartilharam aquela publicação, e com esses dados verificar nos modelos de detecção de *bots* publicamente disponíveis um conjunto de utilizadores, se existirem um conjunto de contas que foram consideradas *bot* existe uma maior probabilidade da notícia ser falsa. Modelos como o "Botometer"[65] [16] ou "Bot Sentinel"[8] podem ser facilmente utilizados.

Medidas de Prevenção Tomadas Pelas Redes Sociais

Apesar do estudo apenas ser feito com dados do Twitter, as medidas sugeridas com base nos dados obtidos podem ser aplicadas nas restantes redes sociais.

- **Utilização de modelos de identificação de contas *bot* mais robustos**. As redes sociais em geral tem modelos de detecção de *bots* que são mais conservadores nas identificações das contas *bot*, o que é compreensível dado o facto de existirem consequências maiores numa identificação falsa do que para um modelo independente. Infelizmente isto pode trazer consequências para os seus utilizadores, em "Online Human-Bot Interactions: Detection, Estimation, and Characterization"[77] é indicado que cerca de 9% a 15% das contas ativas no Twitter apresentam comportamentos que podem indicar que são contas *bot*. Sendo que existem cerca de 330 milhões de utilizadores ativos por mês [4], se for considera o mínimo de 9%, isto corresponde a cerca de 27 milhões de contas por mês que apresentam comportamentos que podem indicar que são contas *bot*. Este número de contas pode *bot* pôr em risco os restantes utilizadores, a utilização de um modelo mais robusto poderia ajudar na redução do número de contas *bot*.

- **Limitar o alcance das comunidade que disseminam notícias falsas**. Nas notícias falsas as contas identificadas como *bot* tem uma maior presença nas redes, querem seja estas contas *bot* seguirem contas de disseminadores de notícias falsas ou as contas dos disseminadores seguirem contas *bot*, como pode ser verificado com os dados recolhidos III.39, III.10, III.11 e III.12, é também possível verificar que nas redes de disseminação de notícias falsas tem um maior alcance, como pode ser verificados nos dados seguintes III.18 e III.19 em conjunto com os dados recolhidos com o "Community Health Assessment Model"(tabelas II.1, II.5, II.3 e II.7), que indicam que existe uma maior vulnerabilidade nas para os utilizadores que estão dentro de comunidades em que são disseminadas notícias falsas e que existe uma maior vulnerabilidade de existir uma transmissão de notícias entre comunidade quando a notícias é falsa.

[4] https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/

Torna-se então importante limitar o alcance das notícias falsas, isto pode ser feito a partir de uma análise de comportamentos as redes sociais podiam identificar os utilizadores com maior impacto e "silenciar" ou ocultar as suas publicações de alcançarem mais utilizadores.

- **Indicar aos utilizadores que a notícia que estão a visualizar foi partilhada por um web site que não é fidedigno**. Indicar ao utilizador de uma forma que pareça óbvia para ele de que o website de onde a notícia é oriunda não é fidedigno, de forma a tentar mitigar a disseminação de notícias falsas. Outra opção é silenciar ou ocultar as notícias que são publicadas dependendo do web site.

Capítulo 6

Conclusão

Com os resultados obtidos ao longo do trabalho, foi possível responder às questões de investigação inicialmente propostas na seção 1.2. De forma a se obterem os resultados foi necessária a criação de um *dataset* que pudesse ser livremente manipulado, o que levou a que fossem recolhidos dados da disseminação de notícias nas redes sociais, devidas a algumas restrições os dados foram recolhidos apenas através do Twitter. Os dados foram recolhidos utilizando uma versão adaptada do "FakeNewsNet"[67][68][69], que é uma ferramenta que permite recolher dados referente a disseminação de uma notícia no Twitter. Os dados recolhidos são então aplicados a modelos e analisados de forma a responder às questões de investigação.

De forma a se obterem dados que pudessem responder às questões de investigação, os dados contidos no *dataset* foram aplicados a modelos de identificação de contas *bot*, modelos epidemiológicos de difusão de informação, modelos de deteção de comunidades e modelos de cálculo de confiança nas redes de disseminação nas redes sociais. Os dados também foram sujeitos a uma análise da disseminação das notícias e as relações dos utilizadores com as contas *bot*, e uma análise ao poder de disseminação.

Para a detecção das contas *bot* nas redes sociais foi selecionado o "Botometer" [65] [16], a fácil aplicação do modelo, os bons resultados em testes de grupo de modelos de detecção de contas *bot* e ao facto de apresentar papéis científicos que explica como o modelo foi criado e como foi treinado, algo que nem sempre está presente nestes modelos. O modelo também apenas foi aplicado apenas aos disseminadores das notícias, isto deve-se ao facto das redes terem um grande número de intervenientes e também devido às limitações do número de pedidos que pode ser feita a API do "Botometer" e da API do Twitter. Para a identificação das contas *bot* foi utilizada a métrica CAP, esta métrica é a indicada pela documentação do "Botometer" a ser utilizada para identificação das contas *bot*, dado que reflete a probabilidade da conta analisada ser uma conta *bot*, ou seja uma conta que tem um CAP de 0.96 ou 96%, é uma conta que o modelo considera ter a probabilidade de 96% de ser uma conta *bot*. Para a identificação das contas *bot* foram escolhidos dois "CAPs" diferentes, o de 90% e o de 96% de forma se

obter dados mais amplos da presença das contas *bot* nas redes de disseminação das notícias nas redes sociais.

A aplicação de modelos de difusão de informação a identificação de notícias falsas, foi feita utilizando o ”Community Health Assessment Model”[54], este modelo permite medir a vulnerabilidade de um utilizador partilhar uma notícia nas redes sociais. O modelo permite medir a vulnerabilidade de duas formas diferentes, a vulnerabilidade entre os utilizadores e a vulnerabilidade dentro de comunidades. De forma ao modelo poder ser aplicado é necessário a utilização de modelos de deteção das comunidades e modelos de cálculo de confiança nas redes de disseminação nas redes sociais. Os modelos de cálculo de confiança nas redes de disseminação nas redes sociais são utilizados dado que o ”Community Health Assessment Model” se baseia no facto de para disseminação de notícias falsas é necessário existir uma maior confiança entre os utilizadores, pois para a disseminação das notícias falsas é necessário que os utilizadores tenham uma maior confiança entre eles para que estas sejam disseminadas, dado que a veracidade das notícias falsas pode ser verificada, para o cálculo da confiança entre utilizadores é necessário a utilização do “Trust in Social Media”[64], que fornecendo uma rede de disseminação de uma notícia, calcula a confiança entre os utilizadores que estiveram em contacto com a notícia. Os modelos de deteção de comunidades são também eles utilizados para que o modelo consiga fornecer os dados da vulnerabilidade, os modelos selecionados foram o Louvain, Infomap, e Label Propagation, estes modelos utilizam os dados da rede disseminação de uma notícia e geram as comunidades existentes dentro dessa rede. Com os dados dos modelos de deteção de comunidades, do “Trust in Social Media” e os dados rede de disseminação de uma notícia, o ”Community Health Assessment Model” apresenta os resultados das vulnerabilidades.

As análises aos dados da disseminação das notícias e do poder de disseminação dos utilizadores foi feita utilizando os dados contidos no *dataset*. Para a análise aos dados da disseminação e do poder de disseminação foi feita uma distribuição das contas com base nos seus seguidores, a distribuição foi feita para tentar compreender se ter um maior número de seguidores realmente se obtém uma maior influência nos utilizadores dentro de uma rede e também para tentar compreender o papel que as contas com um determinado número de seguidores desempenham. Para a análise foram calculadas as médias dos utilizadores com base na sua distribuição de número de seguidores que disseminaram as notícias verdadeiras e as falsas, os utilizadores que fizeram “tweets” e os que fizeram “retweet”, os disseminadores que são seguidos por contas *bot*, os disseminadores que seguem contas *bot* e o número de seguidores das contas *bot* que disseminam notícias. Para a mediação do poder de disseminação dos utilizadores disseminadores de notícias foram criadas 3 métricas que tentam calcular diferentes campos que podem constituir o poder de disseminação, a própria constituição do que é o poder de disseminação não é muito concreta, o poder de disseminação poder ser considerado o número de utilizadores que um utilizador consegue alcançar, ou número de utilizadores que este consegue que se tornem disseminadores da notícia que partilharam ou então o número de utilizadores que este consegue influenciar com base nos seus seguidores. Com base nestas diferentes definições foram então criadas 3 diferentes métricas: pontuação do alcance nos seguidores, pontuação do alcance e pontuação da influência. Com base nas métricas foi

feita a análise aos disseminadores das notícias, para que se perceba melhor a influência que as contas com base nos seus seguidores têm na disseminação das notícias nas redes sociais.

De forma a se compreender como é que a identificação das contas *bot* nas redes sociais em conjunto com os modelos epidemiológicos resulta numa melhor identificação das notícias falsas, foram utilizados o "Botometer" e o "Community Health Assessment Model". Utilizando o dados do *dataset* e aplicando os dois modelos é possível medir que sem as contas *bot* na disseminação de notícias falsas entre os utilizadores existe uma redução da vulnerabilidade de 2,8% a 0,7%, enquanto notícias verdadeiras apenas existiu a redução de 1,8% num dos algoritmos de deteção de comunidades, enquanto que nos restantes os resultados mantiveram-se idênticos ou aumentaram, o que assim permite fazer identificar que a presença das contas *bot* desempenham um papel maior na disseminação das notícias falsas do que nas verdadeiras, o que permite assim distinguir as contas *bot* das verdadeiras apesar da distinção ser relativamente pequena, mas tendo em consideração que as contas *bot* representam em média 0,284% dos disseminadores das notícias nas redes sociais, os resultados são positivos.

De forma a medir o impacto das contas *bot* na disseminação das notícias falsas nas redes sociais foram utilizados os modelos "Botometer" e o "Community Health Assessment Model", em conjunto com uma análise da disseminação das notícias. Os dados contidos no *dataset* foram aplicados aos modelos, de forma a ser calculado o impacto das contas *bot* foram analisadas as redes de disseminação contendo as contas *bot* e removendo as contas *bot*. Assim permitiu medir que o impacto da contas *bot* é bastante maior do que o que seria inicialmente esperado dado que as contas *bot* no seu máximo representam apenas 4,5% dos disseminadores e em média representam 0,284% dos disseminadores, utilizando o "Community Health Assessment Model", o impacto na partilha das notícias dentro das comunidades é entre 2,9% a 6,9% e o impacto na disseminação de notícias falsas entre utilizadores é de 2,8% a 0,7%. Considerando a população pequena das contas *bot*, o impacto destas na disseminação é definitivamente significativo. Com a análise a disseminação, também foi possível compreender melhor o papel e o poder de disseminação das contas *bot*, nas redes de disseminação de notícias falsas os disseminadores são seguidos quase 2 vezes mais por contas *bot* do que nas notícias verdadeiras, nas redes de disseminação de notícias falsas as contas *bot* também são seguidas a 2 vezes mais por contas "normais" do que em comparação com as notícias verdadeiras, o que leva a concluir que nas redes de disseminação de notícias falsas as contas *bot* tem uma maior presença e até maior influência nas comunidades de disseminadores de notícias falsas. Os resultados da análise ao poder de disseminação as notícias falsas apresentam ter uma maior alcance do que em comparação com as notícias verdadeiras, as notícias falsas conseguem gerar um maior número de interações e "retweets" por publicação do que as notícias verdadeiras, o que pode justificar o maior alcance das notícias falsas, mas também pode indicar que estas têm um maior poder de disseminação.

Com os dados obtidos é então possível verificar que com a identificação das contas *bot* é possível distinguir as notícias falsas das verdadeiras e também compreender o impacto e o poder de disseminação que estas contas têm na disseminação de notícias falsas e verdadeiras. Considerando o facto de as contas *bot* apresentarem no seu máximo 4,5% dos disseminadores das notícias nas redes sociais, os resultados

apresentam que as contas *bot* desempenham uma função significativa considerando o seu tamanho da amostra.

Com os resultados recolhidos ao longo do trabalho foi possível fornecer medidas que permitem os utilizadores se protegerem da disseminação de notícias falsas por parte de contas *bot* e também fornecer medidas que permitam as redes sociais melhor proteger os seus utilizadores da disseminação das notícias falsas por parte destas contas. Medidas de prevenção para os utilizadores como a utilização de modelos de detecção de contas *bot* disponíveis online, uma melhor verificação das notícias e as fontes das notícias, e uma verificação cuidadosa das publicações e os utilizadores que interagem com mesma, podem proteger um utilizador da disseminação das notícias falsas e a possível influência de contas *bot*. Para a redes sociais proteger os seus utilizadores de contas *bot* e dos perigos das notícias falsas foi sugerido a utilização de modelos de identificação de contas *bot* mais robustos, limitar o alcance das comunidade que disseminam notícias falsas e indicar aos utilizadores que a notícia que estão a visualizar foi partilhada por um website que não é fidedigno. Estas medidas em conjunto poderão proteger os utilizadores e remover alguns dos riscos que estes atualmente correm.

Trabalho Futuro

No futuro caso existisse mais tempo e recursos seria interessante explorar melhor as redes dissemina-ção das notícias, de forma a melhor identificar o comportamento de uma conta disseminadora, de forma a que seja mais fácil compreender o comportamento e o impacto que uma conta tem na disseminação das notícias. Para que isto seja possível seria necessário fazer uma análise das notícias mais recentes partilhadas pelas contas intervenientes nas redes de disseminação, de forma a ser possível compreender o comportamento de uma conta e o seu poder de disseminação neste tipo de redes. Com esta melhor análise também seria melhor compreendido o papel das contas *bot* na disseminação das notícias falsas. Também seria interessante modificar e alterar os modelos existentes de forma ao cálculo da vulnerabili-dade dos utilizadores incluir o impacto que as contas *bot* tiveram na vulnerabilidade dos utilizadores. A aplicação de processamento de linguagem natural aos *tweets* numa rede de disseminação de forma a melhor a ter em conta o sentimento expresso pelos utilizadores ao disseminarem a notícia e com incluir este sentimentos expressos nos modelos existentes de forma a melhor representar a vulnerabilidade dos utilizadores.

Com os dados recolhidos também seria possível explorar em mais detalhe a disseminação das notí-cias, verificando de forma temporal a disseminação das mesmas e assim compreendendo que papel é que as contas *bot* têm ao longo da disseminação das notícias. A modelação da rede de disseminação em grafos e a possível verificação das relações entre os disseminadores também iria permitir uma me-lhor perceção das relações entre disseminadores e das relações dos disseminadores e as contas *bot*. Em geral, uma aplicação dos dados recolhidos em grafos, que permitisse uma melhor visualização e inter-pretação dos resultados recolhidos levaria uma melhor compreensão do papel que as contas *bot* têm na disseminação das notícias.

Problemas Encontrados no Desenvolvimento

Durante a realização deste estudo houve um conjunto de problemas encontrados, sendo eles a falta de poder de computação, a falta de estudos e informação sobre este tema e a dificuldade em obter dados. Os problemas de falta de poder de computação são devidos a necessidade de memória que os modelos de deteção de comunidades tem, para este modelos em alguns casos é necessário ter uma memória RAM superior a 64GB, um outro problema de computação é a necessidade de processamento que o "Community Health Assessment Model" tem. Estes problemas foram resolvidos com a utilização de computação em nuvem, mas mesmo assim os tempos de execução foram extensos. Dado que é uma área de estudo um pouco recente é de forma a encontrar informação, modelos relevantes e modelos que consigam ser aplicados de forma real é difícil, o que levou a necessidade de testar e descobrir os modelos que funcionam e os que não, o que é bastante frustrante e consome bastante tempo. As limitações das APIs em que os dados foram recolhidos limitaram e afetaram os resultados que foram apresentados, mas mesmo com essas limitações foi feita uma seleção do que é possível fazer com essas limitações e foi possível apresentar resultados satisfatórios.

Para concluir, mesmo com as limitações presentes foi possível desenvolver um estudo que apresenta resultados positivos da influência das contas *bot* na disseminação de informação e os riscos que apresentam para os utilizadores das redes sociais.

Bibliografia

[1] D. Aadland, D. Finnoff e K. Huang. “The Dynamics of Economic Epidemiology Equilibria”. Em: (jan. de 2011).

[2] F. Ahmed e M. Abulaish. “A generic statistical approach for spam detection in Online Social Networks”. Em: *Computer Communications* 36.10 (2013), pp. 1120–1129. issn: 0140-3664. doi: https://doi.org/10.1016/j.comcom.2013.04.004. url: https://www.sciencedirect.com/science/article/pii/S0140366413001047.

[3] R. Beckley, C. Weatherspoon, M. Alexander, M. Chandler, A. Johnson e G. S. Bhatt. “Modeling epidemics with differential equations”. Em: 2013.

[4] F. Benevenuto, T. Rodrigues, J. Almeida, M. Goncalves e V. Almeida. “Detecting spammers and content promoters in online video social networks”. Em: *IEEE INFOCOM Workshops 2009* (2009). doi: 10.1109/infcomw.2009.5072127.

[5] A. Bessi e E. Ferrara. “Social bots distort the 2016 U.S. Presidential election online discussion”. Em: *First Monday* 21.11 (2016). doi: 10.5210/fm.v21i11.7090. url: https://firstmonday.org/ojs/index.php/fm/article/view/7090.

[6] L. M. Bettencourt, A. Cintrón-Arias, D. I. Kaiser e C. Castillo-Chávez. “The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models”. Em: *Physica A: Statistical Mechanics and its Applications* 364 (2006), pp. 513–536. doi: 10.1016/j.physa.2005.08.083. url: https://doi.org/10.1016%2Fj.physa.2005.08.083.

[7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte e E. Lefebvre. “Fast unfolding of communities in large networks”. Em: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (2008). doi: 10.1088/1742-5468/2008/10/p10008.

[8] *botSentinel*. url: https://botsentinel.com/info/about.

[9] L. Breiman. “Random Forests”. English. Em: *Machine Learning* 45.1 (2001), pp. 5–32. issn: 0885-6125. url: http://dx.doi.org/10.1023/A%3A1010933404324.

[10] C. Budak, D. Agrawal e A. El Abbadi. “Limiting the Spread of Misinformation in Social Networks”. Em: *Proceedings of the 20th International Conference on World Wide Web*. WWW ’11. Association for Computing Machinery, 2011, pp. 665–674. isbn: 9781450306324. doi: 10.1145/1963405.1963499. url: https://doi.org/10.1145/1963405.1963499.

[11] N. Chavoshi, H. Hamooni e A. Mueen. “DeBot: Twitter bot detection via warped correlation”. Em: *2016 IEEE 16th International Conference on Data Mining (ICDM)* (2016). doi: 10.1109/icdm.2016.0096.

[12] Z. Chen e D. Subramanian. “An Unsupervised Approach to Detect Spam Campaigns that Use Botnets on Twitter”. Em: *CoRR* abs/1804.05232 (2018). arXiv: 1804.05232. url: http://arxiv.org/abs/1804.05232.

[13] Z. Chu, S. Gianvecchio, H. Wang e S. Jajodia. “Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?” Em: *IEEE Transactions on Dependable and Secure Computing* 9.6 (2012), pp. 811–824. doi: 10.1109/TDSC.2012.75.

[14] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi e M. Tesconi. “DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection”. Em: *IEEE Intelligent Systems* 31.5 (2016), pp. 58–64. issn: 1941-1294. doi: 10.1109/mis.2016.29. url: http://dx.doi.org/10.1109/MIS.2016.29.

[15] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi e M. Tesconi. “The Paradigm-Shift of Social Spambots”. Em: *Proceedings of the 26th International Conference on World Wide Web Companion - WWW ’17 Companion* (2017). doi: 10.1145/3041021.3055135. url: http://dx.doi.org/10.1145/3041021.3055135.

[16] C. A. Davis, O. Varol, E. Ferrara, A. Flammini e F. Menczer. “Botornot”. Em: *Proceedings of the 25th International Conference Companion on World Wide Web - WWW ’16 Companion* (2016). doi: 10.1145/2872518.2889302.

[17] A. Drif, Z. Ferhat Hamida e S. Giordano. “Fake News Detection Method Based on Text-Features”. Em: ago. de 2019.

[18] J. Echeverría e S. Zhou. *Discovery, Retrieval, and Analysis of ’Star Wars’ botnet in Twitter*. 2017. arXiv: 1701.02405 [cs.SI].

[19] E. Ferrara, W.-Q. Wang, O. Varol, A. Flammini e A. Galstyan. “Predicting Online Extremism, Content Adopters, and Interaction Reciprocity”. Em: *Social Informatics* (2016), pp. 22–39. issn: 1611-3349. doi: 10.1007/978-3-319-47874-6_3. url: http://dx.doi.org/10.1007/978-3-319-47874-6_3.

[20] Z. Gilani, E. Kochmar e J. Crowcroft. “Classification of twitter accounts into automated agents and human users”. Em: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (2017). doi: 10.1145/3110025.3110091.

[21] M. Girvan e M. E. J. Newman. “Community structure in social and biological networks”. Em: *Proceedings of the National Academy of Sciences* 99.12 (2002), pp. 7821–7826. issn: 0027-8424. doi: 10.1073/pnas.122653799. eprint: https://www.pnas.org/content/99/12/7821.full.pdf. url: https://www.pnas.org/content/99/12/7821.

[22] C. Godfrey, L. Wiessing e R. Hartnoll. *Modelling drug use: methods to quantify and understand hidden processes*. Jan. de 2001. isbn: 92-9168-056-7.

[23] M. Gomes, J. Alfonso-Cendón, P. Marqués-Sánchez, D. Carneiro e P. Novais. "Improving Conflict Support Environments with Information Regarding Social Relationships". Em: nov. de 2014, pp. 779–790. isbn: 978-3-319-12026-3. doi: 10.1007/978-3-319-12027-0_63.

[24] M. Gomes, J. Zeleznikow e P. Novais. "A Non-intrusive Approach to Measuring Trust in Opponents in a Negotiation Scenario". Em: *AICOL*. 2017.

[25] T. Harko, F. S. Lobo e M. Mak. "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates". Em: *Applied Mathematics and Computation* 236 (2014), pp. 184–194. doi: 10.1016/j.amc.2014.03.030. url: https://doi.org/10.1016%2Fj.amc.2014.03.030.

[26] R. Isea e K. Lonngren. "A New Variant of the SEIZ Model to Describe the Spreading of a Rumor". Em: *International Journal of Data Science and Analysis* 3 (2017), pp. 28–33. doi: 10.11648/j.ijdsa.20170304.12.

[27] M. Jiang. "Bioterrorism. Mathematical modeling applications in homelandsecurity.,Frontiers in Applied Mathematics, 28, Society for Industrial and Applied Mathematics (SIAM)". Em: *Biomedical Engineering Online - BIOMED ENG ONLINE* 4 (dez. de 2005), pp. 1–3. doi: 10.1186/1475-925X-4-69.

[28] F. Jin, E. Dougherty, P. Saraf, P. Mi, Y. Cao e N. Ramakrishnan. "Epidemiological modeling of news and rumors on Twitter". Em: cited By 151. 2013. doi: 10.1145/2501025.2501027. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84890706924&doi=10.1145%2f2501025.2501027&partnerID=40&md5=d687abf17c82e82382303ae438b2508a.

[29] K. Kandhway e J. Kuri. "How to run a campaign: Optimal control of SIS and SIR information epidemics". Em: *Applied Mathematics and Computation* 231 (2014), pp. 79–92. issn: 0096-3003. doi: 10.1016/j.amc.2013.12.164. url: http://dx.doi.org/10.1016/j.amc.2013.12.164.

[30] R. Karlsen e T. Aalberg. "Social Media and Trust in News: An Experimental Study of the Effect of Facebook on News Story Credibility". Em: *Digital Journalism* 0.0 (2021), pp. 1–17. doi: 10.1080/21670811.2021.1945938. eprint: https://doi.org/10.1080/21670811.2021.1945938. url: https://doi.org/10.1080/21670811.2021.1945938.

[31] M. Kearney. *Detect twitter bots*. url: https://tweetbotornot2.mikewk.com/index.html.

[32] D. Kendall. "Deterministic and Stochastic Epidemics in Closed Populations". Em: 1956.

[33] M. G. KENDALL. "A NEW MEASURE OF RANK CORRELATION". Em: *Biometrika* 30.1-2 (1938), pp. 81–93. issn: 0006-3444. doi: 10.1093/biomet/30.1-2.81. eprint: https://academic.oup.com/biomet/article-pdf/30/1-2/81/423380/30-1-2-81.pdf. url: https://doi.org/10.1093/biomet/30.1-2.81.

[34] W. O. Kermack, A. G. McKendrick e G. T. Walker. “A contribution to the mathematical theory of epidemics”. Em: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pp. 700–721. doi: 10.1098/rspa.1927.0118. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0118. url: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118.

[35] M Kröger e R Schlickeiser. “Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor”. Em: *Journal of Physics A: Mathematical and Theoretical* 53.50 (2020), p. 505601. doi: 10.1088/1751-8121/abc65d. url: https://doi.org/10.1088/1751-8121/abc65d.

[36] S. Kudugunta e E. Ferrara. “Deep Neural Networks for BOT detection”. Em: *Information Sciences* 467 (2018), pp. 312–322. doi: 10.1016/j.ins.2018.08.019.

[37] G. Laurent e J.-N. Kapferer. “Measuring consumer involvement profiles”. Em: *Journal of Marketing Research* 22.1 (1985), p. 41. doi: 10.2307/3151549.

[38] K. Lee, B. Eoff e J. Caverlee. “Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter”. Em: *ICWSM*. 2011.

[39] J. Ma, W. Gao, P. Mitra, S. Kwon, B. Jansen, K.-F. Wong e M. Cha. “Detecting rumors from microblogs with recurrent neural networks”. Em: vol. 2016-January. 2016, pp. 3818–3824. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85006173435&partnerID=40&md5=bc9932e9d906fb5859df3efa69a13f7e.

[40] J. Ma, W. Gao e K.-F. Wong. “Rumor detection on twitter with tree-structured recursive neural networks”. Em: vol. 1. 2018, pp. 1980–1989. doi: 10.18653/v1/p18-1184. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061740347&doi=10.18653%2fv1%2fp18-1184&partnerID=40&md5=b8f2c2ce92c738af964daf95faa71f25.

[41] M. Maleki, E. Mead, M. Arani e N. Agarwal. *Using an Epidemiological Model to Study the Spread of Misinformation during the Black Lives Matter Movement*. 2021. doi: 10.48550/ARXIV.2103.12191. url: https://arxiv.org/abs/2103.12191.

[42] R. Mann, J. Faria, D. Sumpter e J. Krause. “The dynamics of audience applause”. Em: *Journal of the Royal Society, Interface / the Royal Society* 10 (ago. de 2013), p. 20130466. doi: 10.1098/rsif.2013.0466.

[43] R. Martins, M. Gomes, J. Almeida, P. Novais e P. Henriques. “Hate Speech Classification in Social Media Using Emotional Analysis”. Em: out. de 2018, pp. 61–66. doi: 10.1109/BRACIS.2018.00019.

[44] Z. Miller, B. Dickinson, W. Deitrick, W. Hu e A. H. Wang. “Twitter spammer detection using data stream clustering”. Em: *Information Sciences* 260 (2014), pp. 64–73. issn: 0020-0255. doi: https://doi.org/10.1016/j.ins.2013.11.016. url: https://www.sciencedirect.com/science/article/pii/S0020025513008037.

[45] M. P. Moraes, J. de Oliveira Sampaio e A. C. Charles. “Data mining applied in fake news classification through textual patterns”. Em: *Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web*. SBC, 2020, pp. 49–52. doi: 10.5753/webmedia_estendido.2020.13061. url: https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/13061.

[46] P. Pons e M. Latapy. *Computing communities in large networks using random walks (long version)*. 2005. doi: 10.48550/ARXIV.PHYSICS/0512106. url: https://arxiv.org/abs/physics/0512106.

[47] D. M. W. Powers. “Applications and Explanations of Zipf’s Law”. Em: *New Methods in Language Processing and Computational Natural Language Learning*. 1998.

[48] U. N. Raghavan, R. Albert e S. Kumara. “Near linear time algorithm to detect community structures in large-scale networks”. Em: *Physical Review E* 76.3 (2007). doi: 10.1103/physreve.76.036106.

[49] A. Rapoport. “Spread of information through a population with socio-structural bias: II. Various models with partial transitivity”. Em: *Bulletin of Mathematical Biology* 15 (1953), pp. 535–546.

[50] A. Rapoport. “Spread of information through a population with socio-structural bias: III. Suggested experimental procedures”. Em: *Bulletin of Mathematical Biology* 16 (1954), pp. 75–81.

[51] A. Rapoport e L. I. Rebhun. “On the mathematical theory of rumor spread”. Em: *Bulletin of Mathematical Biology* 14 (1952), pp. 375–383.

[52] B. Rath e J. Srivastava. “Epidemiology inspired framework for fake news mitigation in social networks”. Em: vol. 2699. 2020. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097566839&partnerID=40&md5=42d303c0aa60566cf7692d5fded64194.

[53] B. Rath, W. Gao, J. Ma e J. Srivastava. “From Retweet to Believability: Utilizing Trust to Identify Rumor Spreaders on Twitter”. Em: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ASONAM ’17. Association for Computing Machinery, 2017, pp. 179–186. isbn: 9781450349932. doi: 10.1145/3110025.3110121. url: https://doi.org/10.1145/3110025.3110121.

[54] B. Rath, W. Gao e J. Srivastava. “Evaluating vulnerability to fake news in social networks”. Em: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2019). doi: 10.1145/3341161.3342920.

[55] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini e F. Menczer. “Detecting and Tracking Political Abuse in Social Media”. Em: jan. de 2011.

[56] J. Reichardt e S. Bornholdt. “Statistical mechanics of community detection”. Em: *Physical Review E* 74.1 (2006). doi: 10.1103/physreve.74.016110.

[57] H. S. Rodrigues e M. J. Fonseca. “Can information be spread as a virus? viral marketing as epidemiological model”. Em: *Mathematical Methods in the Applied Sciences* 39.16 (2015), pp. 4780–4786. issn: 0170-4214. doi: 10.1002/mma.3783. url: http://dx.doi.org/10.1002/mma.3783.

[58] H. S. Rodrigues e M. J. Fonseca. *Viral marketing as epidemiological model*. 2015. arXiv: 1507.06986 [physics.soc-ph].

[59] E. Rogers, A. Singhal e M. Quinlan. “Diffusion of Innovations”. Em: mar. de 2019, pp. 182–186. isbn: 9780203710753. doi: 10.4324/9780203710753-35.

[60] R. Ross. “An application of the theory of probabilities to the study of a priori pathometry.—Part I”. Em: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 92.638 (1916), pp. 204–230. doi: 10.1098/rspa.1916.0007. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1916.0007. url: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1916.0007.

[61] R. Ross e H. P. Hudson. “An application of the theory of probabilities to the study of a priori pathometry.—Part II”. Em: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 93.650 (1917), pp. 212–225. doi: 10.1098/rspa.1917.0014. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1917.0014. url: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1917.0014.

[62] R. Ross e H. P. Hudson. “An application of the theory of probabilities to the study of a priori pathometry.—Part III”. Em: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 93.650 (1917), pp. 225–240. doi: 10.1098/rspa.1917.0015. url: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1917.0015.

[63] M. Rosvall e C. T. Bergstrom. “Maps of random walks on complex networks reveal community structure”. Em: *Proceedings of the National Academy of Sciences* 105.4 (2008), pp. 1118–1123. doi: 10.1073/pnas.0706851105.

[64] A. Roy, C. Sarkar, J. Srivastava e J. Huh. “Trustingness & Trustworthiness: A pair of complementary trust measures in a social network”. Em: *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2016). doi: 10.1109/asonam.2016.7752289.

[65] M. Sayyadiharikandeh, O. Varol, K.-C. Yang, A. Flammini e F. Menczer. “Detection of Novel Social Bots by Ensembles of Specialized Classifiers”. Em: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. ACM, 2020. doi: 10.1145/3340531.3412698. url: https://doi.org/10.1145%2F3340531.3412698.

[66] R Schlickeiser e M Kröger. “Analytical solution of the SIR-model for the temporal evolution of epidemics: part B. Semi-time case”. Em: *Journal of Physics A: Mathematical and Theoretical* 54.17 (2021), p. 175601. doi: 10.1088/1751-8121/abed66. url: https://doi.org/10.1088/1751-8121/abed66.

[67] K. Shu, D. Mahudeswaran, S. Wang, D. Lee e H. Liu. “FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media”. Em: *arXiv preprint arXiv:1809.01286* (2018).

[68] K. Shu, A. Sliva, S. Wang, J. Tang e H. Liu. “Fake News Detection on Social Media: A Data Mining Perspective”. Em: *ACM SIGKDD Explorations Newsletter* 19.1 (2017), pp. 22–36.

[69] K. Shu, S. Wang e H. Liu. “Exploiting Tri-Relationship for Fake News Detection”. Em: *arXiv preprint arXiv:1712.07709* (2017).

[70] F. Squazzoni, J. G. Polhill, B. Edmonds, P. Ahrweiler, P. Antosz, G. Scholz, . Chappin, M. Borit, H. Verhagen, F. Giardini e et al. “Computational Models That Matter During a Global Pandemic Outbreak: A Call to Action”. Em: *Journal of Artificial Societies and Social Simulation* (2020). url: http://jasss.soc.surrey.ac.uk/23/2/10.html.

[71] D. Sridhar e M. S. Majumder. “Modelling the pandemic”. Em: *BMJ* 369 (2020). doi: 10.1136/bmj.m1567. eprint: https://www.bmj.com/content/369/bmj.m1567.full.pdf. url: https://www.bmj.com/content/369/bmj.m1567.

[72] D. Strang e M. W. Macy. “In Search of Excellence: Fads, Success Stories, and Adaptive Emulation”. Em: *American Journal of Sociology* 107.1 (2001), pp. 147–182. issn: 00029602, 15375390. url: http://www.jstor.org/stable/10.1086/323039 (acedido em 09/06/2022).

[73] M. Tambuscio e G. Ruffo. “Fact-checking strategies to limit urban legends spreading in a segregated society”. Em: *Applied Network Science* 4.1 (2019). doi: 10.1007/s41109-019-0233-1. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85075945041&doi=10.1007%2fs41109-019-0233-1&partnerID=40&md5=b5618d5aec9cfd49e266fd85df5c4fe2.

[74] M. Tambuscio, G. Ruffo, A. Flammini e F. Menczer. “Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks”. Em: 2015, pp. 977–982. doi: 10.1145/2740908.2742572. url: https://dl.acm.org/doi/10.1145/2740908.2742572.

[75] V. A. Traag, L. Waltman e N. J. van Eck. “From Louvain to Leiden: guaranteeing well-connected communities”. Em: *Scientific Reports* 9.1 (2019). doi: 10.1038/s41598-019-41695-z.

[76] O. Varol, C. A. Davis, F. Menczer e A. Flammini. “Feature Engineering for Social Bot Detection”. Em: *Feature Engineering for Machine Learning and Data Analytics*. Ed. por G. Dong e H. Liu. Data Mining and Knowledge Discovery Series. Chapman e Hall/CRC Press, 2018. Cap. 12, pp. 311–334. url: https://www.crcpress.com/Feature-Engineering-for-Machine-Learning-and-Data-Analytics/Dong-Liu/p/book/9781138744387.

[77] O. Varol, E. Ferrara, C. A. Davis, F. Menczer e A. Flammini. *Online Human-Bot Interactions: Detection, Estimation, and Characterization*. 2017. doi: 10.48550/ARXIV.1703.03107. url: https://arxiv.org/abs/1703.03107.

[78] O. Varol, E. Ferrara, F. Menczer e A. Flammini. *Early Detection of Promoted Campaigns on Social Media*. 2017. arXiv: 1703.07518 [cs.SI].

[79] S. Vosoughi, D. Roy e S. Aral. "The spread of true and false news online". Em: *Science* 359.6380 (2018), pp. 1146–1151. doi: 10.1126/science.aap9559. eprint: https://www.science.org/doi/pdf/10.1126/science.aap9559. url: https://www.science.org/doi/abs/10.1126/science.aap9559.

[80] S. Vosoughi, D. K. Roy e S. Aral. "The spread of true and false news online". Em: *Science* 359 (2018), pp. 1146 –1151.

[81] R. A. Wall. *Introduction to information retrieval*. Loughborough Univ. of Tech. Library, 1971.

[82] S. Wojcik, S. Messing, A. Smith, L. Rainie e P. Hitlin. "Bots in the Twittersphere". Em: 2018.

[83] K.-C. Yang, O. Varol, P.-M. Hui e F. Menczer. "Scalable and Generalizable Social Bot Detection through Data Selection". Em: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.01 (2020), pp. 1096–1103. doi: 10.1609/aaai.v34i01.5460.

[84] K. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini e F. Menczer. "Arming the public with artificial intelligence to Counter Social Bots". Em: *Human Behavior and Emerging Technologies* 1.1 (2019), pp. 48–61. doi: 10.1002/hbe2.115.

[85] W. Yang, D. Zhang, L. Peng, C. Zhuge e L. Hong. *Rational evaluation of various epidemic models based on the COVID-19 data of China*. 2020. doi: 10.48550/ARXIV.2003.05666. url: https://arxiv.org/abs/2003.05666.

[86] F. Yu, Q. Liu, S. Wu, L. Wang e T. Tan. "A Convolutional Approach for Misinformation Identification". Em: (2017), pp. 3901–3907. doi: 10.24963/ijcai.2017/545. url: https://doi.org/10.24963/ijcai.2017/545.

Anexo I

Tabelas adicionais

Dado o grande número de dados tratados, nem todas as tabelas podem ser apresentadas directamente nos capítulos, então este anexo é dedicado todas as tabelas que adicionais que não foram incluídas nos seus respectivos capítulos.

Anexo II

Aplicação dos Modelos e Recolha de Dados

Tabela II.1: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_3	0.946	0.999	0.750	0.927	0.945	0.958	0.897	0.900	0.999	0.826	0.745	0.940	0.746	0.667	0.999	0.849	0.820	0.959
Politifact_4	0.915	0.999	0.906	0.984	0.958	0.980	0.966	0.946	0.999	0.924	0.913	0.984	0.873	0.852	0.993	0.936	0.925	0.985
Politifact_5	0.913	0.882	0.712	0.847	0.738	0.999	0.835	0.779	0.999	0.839	0.828	0.999	0.783	0.844	0.999	0.838	0.827	0.929
Politifact_8	0.863	0.818	0.468	0.730	0.633	0.999	0.708	0.624	0.999	0.766	0.805	0.999	0.739	0.633	0.999	0.746	0.693	0.964
Politifact_14	0.999	0.999	0.714	0.940	0.999	0.999	0.949	0.899	0.999	0.892	0.649	0.999	0.788	0.577	0.999	0.894	0.761	0.969
Politifact_17	0.799	0.666	0.596	0.777	0.599	0.999	0.799	0.599	0.999	0.706	0.424	0.999	0.753	0.633	0.999	0.774	0.540	0.961
Politifact_20	0.925	0.866	0.615	0.881	0.794	0.999	0.868	0.854	0.999	0.911	0.759	0.949	0.845	0.683	0.799	0.893	0.770	0.904
Politifact_27	0.874	0.947	0.705	0.912	0.925	0.972	0.899	0.870	0.999	0.904	0.838	0.999	0.885	0.794	0.999	0.902	0.849	0.975

Tabela II.2: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.904	0.897	0.683	0.875	0.824	0.988	0.865	0.809	0.999	0.846	0.745	0.984	0.802	0.710	0.973	0.854	0.773	0.956

Tabela II.3: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas.

	L	I	LP
Politifact_4	0.115	0.428	0.287
Politifact_5	-0.143	-0.044	-0.186
Politifact_8	-0.070	0.153	-0.269
Politifact_11	-0.153	-0.252	-0.206
Politifact_14	-0.043	0.000	-0.341
Politifact_17	-0.005	0.466	-0.287
Politifact_20	0.111	0.191	-0.302
Politifact_27	-0.062	0.513	-0.310

Tabela II.4: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas.

	L	I	LP
Média	-0.047	0.244	-0.283

Tabela II.5: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias verdadeiras.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_6	0.691	0.666	0.432	0.612	0.587	0.999	0.564	0.638	0.999	0.549	0.743	0.999	0.551	0.624	0.999	0.578	0.655	0.928
Politifact_7	0.421	0.545	0.408	0.505	0.399	0.666	0.421	0.339	0.999	0.349	0.274	0.899	0.299	0.233	0.599	0.366	0.325	0.798
Politifact_9	0.677	0.593	0.528	0.683	0.580	0.749	0.662	0.715	0.849	0.637	0.766	0.799	0.555	0.699	0.783	0.638	0.695	0.785
Politifact_12	0.773	0.647	0.449	0.662	0.512	0.999	0.689	0.466	0.999	0.579	0.528	0.999	0.569	0.749	0.999	0.630	0.559	0.949
Politifact_18	0.715	0.874	0.495	0.711	0.803	0.999	0.705	0.838	0.999	0.682	0.757	0.999	0.706	0.698	0.999	0.704	0.786	0.944
Politifact_19	0.460	0.499	0.386	0.468	0.466	0.666	0.499	0.733	0.399	0.437	0.949	0.000	0.322	0.633	0.000	0.440	0.732	0.204
Politifact_21	0.960	0.636	0.693	0.925	0.944	0.999	0.933	0.959	0.999	0.874	0.739	0.999	0.802	0.616	0.999	0.890	0.772	0.946
Politifact_22	0.579	0.833	0.415	0.468	0.714	0.000	0.429	0.599	0.000	0.477	0.549	0.000	0.613	0.533	0.000	0.495	0.602	0.038
Politifact_24	0.812	0.666	0.590	0.773	0.733	0.999	0.776	0.849	0.999	0.754	0.805	0.999	0.741	0.783	0.999	0.747	0.787	0.956

Tabela II.6: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias verdadeiras.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.676	0.662	0.488	0.645	0.638	0.786	0.631	0.682	0.805	0.593	0.679	0.744	0.573	0.619	0.709	0.610	0.657	0.728

Tabela II.7: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras.

	L	I	LP
Politifact_6	-0.221	-0.137	-0.280
Politifact_7	-0.064	0.000	-0.147
Politifact_9	-0.082	-0.098	-0.245
Politifact_12	-0.240	0.052	-0.300
Politifact_18	-0.263	-0.040	-0.237
Politifact_19	-0.096	-0.071	-0.125
Politifact_21	-0.107	-0.200	-0.366
Politifact_22	-0.110	-0.216	-0.193
Politifact_24	-0.127	0.066	-0.342

Tabela II.8: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras.

	L	I	LP
Média	-0.136	-0.063	-0.244

Tabela II.9: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 90%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_3	0.946	1.000	0.760	0.900	0.944	0.967	0.853	0.900	0.925	0.852	0.728	1.000	0.826	0.662	1.000	0.865	0.813	0.958
Politifact_4	0.966	1.000	0.913	0.978	0.958	0.991	0.956	0.925	0.999	0.929	0.913	0.999	0.895	0.852	0.999	0.939	0.924	0.989
Politifact_5	0.872	0.889	0.734	0.848	0.762	1.000	0.822	0.780	1.000	0.881	0.829	1.000	0.862	0.833	0.833	0.853	0.826	0.905
Politifact_8	0.781	0.800	0.476	0.654	0.600	1.000	0.670	0.686	1.000	0.764	0.620	1.000	0.817	0.633	1.000	0.739	0.664	0.965
Politifact_14	0.937	1.000	0.743	0.862	1.000	1.000	0.905	0.900	0.900	0.835	0.625	1.000	0.820	0.556	1.000	0.858	0.749	0.954
Politifact_17	0.839	0.714	0.633	0.724	0.611	1.000	0.740	0.500	1.000	0.727	0.400	1.000	0.708	0.600	1.000	0.724	0.546	0.967
Politifact_20	0.943	0.765	0.629	0.891	0.714	0.733	0.903	0.836	1.000	0.928	0.720	0.950	0.867	0.642	0.800	0.914	0.724	0.891
Politifact_27	0.909	0.950	0.703	0.905	0.926	0.973	0.920	0.871	0.999	0.843	0.831	0.999	0.889	0.778	0.999	0.879	0.847	0.974

Tabela II.10: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 90%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.899	0.890	0.662	0.845	0.815	0.950	0.846	0.800	0.971	0.845	0.708	0.992	0.835	0.695	0.939	0.846	0.762	0.940

Tabela II.11: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 90%.

	L	I	LP
Politifact_3	-0.117	0.152	-0.309
Politifact_4	0.008	0.536	-0.275
Politifact_5	-0.040	0.046	-0.228
Politifact_8	0.025	0.061	-0.211
Politifact_14	0.121	0.000	-0.353
Politifact_17	-0.145	0.238	-0.327
Politifact_20	0.914	-0.033	-0.302
Politifact_27	-0.076	0.399	-0.370

Tabela II.12: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 90%.

	L	I	LP
Média	0.086	0.175	-0.297

Tabela II.13: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias reais, mas removendo as contas *bot* com CAP acima de 90%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_6	0.682	0.689	0.433	0.600	0.581	0.999	0.629	0.579	0.999	0.631	0.622	0.999	0.591	0.590	0.999	0.628	0.611	0.918
Politifact_7	0.424	0.500	0.409	0.469	0.500	0.167	0.400	0.327	0.100	0.200	0.267	0.000	0.233	0.244	0.000	0.233	0.321	0.070
Politifact_9	0.670	0.560	0.508	0.635	0.485	0.832	0.649	0.657	0.818	0.631	0.700	0.635	0.618	0.612	0.721	0.651	0.633	0.717
Politifact_12	0.775	0.562	0.432	0.684	0.417	0.999	0.672	0.440	0.999	0.594	0.450	0.999	0.625	0.667	0.999	0.643	0.505	0.937
Politifact_18	0.768	0.846	0.508	0.740	0.792	0.999	0.741	0.833	0.999	0.770	0.738	0.999	0.760	0.695	0.999	0.760	0.771	0.925
Politifact_19	0.473	0.500	0.419	0.438	0.467	0.667	0.432	0.733	0.400	0.414	0.950	0.000	0.367	0.667	0.000	0.440	0.750	0.238
Politifact_21	0.929	0.667	0.687	0.896	0.905	1.000	0.872	0.867	1.000	0.900	0.740	1.000	0.877	0.600	1.000	0.896	0.763	0.957
Politifact_22	0.651	0.737	0.418	0.564	0.643	0.000	0.517	0.554	0.000	0.644	0.667	0.000	0.556	0.587	0.000	0.570	0.604	0.045
Politifact_24	0.765	0.800	0.590	0.806	0.917	1.000	0.741	0.850	1.000	0.764	0.800	1.000	0.725	0.783	1.000	0.749	0.817	0.937

Tabela II.14: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias reais, mas removendo as contas *bot* com CAP acima de 90%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.682	0.651	0.489	0.648	0.634	0.740	0.628	0.649	0.702	0.616	0.659	0.626	0.595	0.605	0.635	0.619	0.642	0.638

Tabela II.15: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras, mas removendo as contas *bot* com CAP acima de 90%.

	L	I	LP
Politifact_6	-0.264	-0.158	-0.346
Politifact_7	-0.015	0.410	-0.187
Politifact_9	-0.070	-0.094	-0.234
Politifact_12	-0.234	0.111	-0.313
Politifact_18	-0.213	0.066	-0.329
Politifact_19	-0.102	0.000	-0.262
Politifact_21	-0.188	-0.152	-0.304
Politifact_22	-0.104	-0.232	-0.243
Politifact_24	-0.078	0.400	-0.373

Tabela II.16: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras, mas removendo as contas *bot* com CAP acima de 90%.

	L	I	LP
Média	-0.141	0.039	-0.288

Tabela II.17: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 96%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_3	0.947	0.999	0.749	0.894	0.936	0.909	0.896	0.889	0.942	0.831	0.733	0.883	0.796	0.652	0.999	0.853	0.812	0.919
Politifact_4	0.896	0.999	0.894	0.983	0.979	0.987	0.939	0.933	0.999	0.930	0.893	0.993	0.891	0.841	0.999	0.930	0.913	0.986
Politifact_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Politifact_8	0.777	0.727	0.460	0.636	0.666	0.999	0.566	0.714	0.999	0.645	0.774	0.999	0.718	0.616	0.999	0.667	0.704	0.964
Politifact_14	0.977	0.999	0.737	0.946	0.999	0.999	0.919	0.899	0.999	0.846	0.649	0.999	0.799	0.577	0.999	0.877	0.761	0.971
Politifact_17	0.796	0.666	0.612	0.763	0.599	0.999	0.735	0.599	0.999	0.684	0.424	0.999	0.758	0.599	0.999	0.741	0.551	0.965
Politifact_20	0.921	0.866	0.621	0.883	0.794	0.749	0.894	0.854	0.999	0.903	0.759	0.949	0.812	0.683	0.799	0.888	0.775	0.888
Politifact_27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabela II.18: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 96%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.886	0.876	0.679	0.851	0.829	0.940	0.825	0.815	0.990	0.807	0.705	0.970	0.796	0.661	0.966	0.826	0.753	0.949

Tabela II.19: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 96%.

	L	I	LP
Politifact_3	0.01	0.17	-0.33
Politifact_4	-0.03	0.63	-0.30
Politifact_5	0	0	0
Politifact_8	-0.09	0.36	-0.24
Politifact_14	-0.11	0.00	-0.40
Politifact_17	-0.06	0.07	-0.31
Politifact_20	-0.05	0.12	-0.32
Politifact_27	0	0	0

Tabela II.20: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias falsas, mas removendo as contas *bot* com CAP acima de 96%.

	L	I	LP
Média	-0.06	0.22	-0.32

Tabela II.21: Vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias reais, mas removendo as contas *bot* com CAP acima de 96%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_6	0.704	0.674	0.438	0.632	0.580	0.999	0.613	0.626	0.999	0.597	0.686	0.999	0.583	0.590	0.999	0.615	0.632	0.929
Politifact_7	0.445	0.428	0.408	0.477	0.393	0.555	0.416	0.327	0.399	0.349	0.266	0.199	0.399	0.244	0.000	0.382	0.305	0.284
Politifact_9	0.655	0.624	0.522	0.639	0.595	0.857	0.622	0.729	0.849	0.622	0.749	0.674	0.542	0.695	0.733	0.614	0.701	0.729
Politifact_12	0.771	0.624	0.460	0.671	0.410	0.999	0.679	0.399	0.999	0.572	0.385	0.999	0.498	0.644	0.999	0.621	0.478	0.939
Politifact_18	0.777	0.885	0.511	0.723	0.817	0.999	0.719	0.839	0.999	0.743	0.771	0.999	0.713	0.726	0.999	0.731	0.804	0.944
Politifact_19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Politifact_21	0.877	0.636	0.688	0.875	0.944	0.999	0.958	0.959	0.999	0.917	0.774	0.999	0.846	0.633	0.999	0.911	0.777	0.945
Politifact_22	0.642	0.789	0.413	0.462	0.599	0.333	0.449	0.538	0.199	0.533	0.585	0.000	0.533	0.586	0.000	0.505	0.602	0.119
Politifact_24	0.796	0.799	0.599	0.782	0.916	0.999	0.715	0.849	0.999	0.719	0.799	0.999	0.716	0.783	0.999	0.756	0.817	0.939

Tabela II.22: Médias da vulnerabilidade dos nodos limite do Community Health Assessment Model sobre as notícias reais, mas removendo as contas *bot* com CAP acima de 96%.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.708	0.682	0.505	0.658	0.657	0.843	0.646	0.658	0.805	0.632	0.627	0.734	0.604	0.613	0.716	0.642	0.640	0.729

Tabela II.23: Vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras, mas removendo as contas *bot* com CAP acima de 96%.

	L	I	LP
Politifact_6	-0.249	-0.117	-0.332
Politifact_7	-0.037	0.180	-0.218
Politifact_9	-0.118	-0.112	-0.282
Politifact_12	-0.329	0.098	-0.327
Politifact_18	-0.185	-0.147	-0.305
Politifact_19	0	0	0
Politifact_21	-0.210	-0.200	-0.334
Politifact_22	-0.100	-0.189	-0.280
Politifact_24	-0.217	0.199	-0.343

Tabela II.24: Médias da vulnerabilidade nas comunidades do Community Health Assessment Model sobre as notícias verdadeiras, mas removendo as contas *bot* com CAP acima de 96%.

	L	I	LP
Média	-0.159	-0.003	-0.304

Anexo III

Poder de Disseminação e as Contas *Bot*

Tabela III.1: Diferença entre a tabela II.17 e a tabela II.1

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_3	0.001	0.000	-0.001	-0.033	-0.009	-0.049	-0.001	-0.011	-0.057	0.005	-0.012	-0.057	0.050	-0.015	0.000	0.004	-0.008	-0.040
Politifact_4	-0.019	0.000	-0.012	-0.001	0.021	0.007	-0.027	-0.013	0.000	0.006	-0.020	0.009	0.018	-0.011	0.006	-0.006	-0.012	0.001
Politifact_5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Politifact_8	-0.086	-0.091	-0.008	-0.094	0.033	0.000	-0.142	0.090	0.000	-0.121	-0.031	0.000	-0.021	-0.017	0.000	-0.079	0.011	0.000
Politifact_14	-0.022	0.000	0.023	0.006	0.000	0.000	-0.030	0.000	0.000	-0.046	0.000	0.000	0.011	0.000	0.000	-0.017	0.000	0.002
Politifact_17	-0.003	0.000	0.016	-0.014	0.000	0.000	-0.064	0.000	0.000	-0.022	0.000	0.000	0.005	-0.034	0.000	-0.033	0.011	0.004
Politifact_20	-0.004	0.000	0.006	0.002	0.000	-0.250	0.026	0.000	0.000	-0.008	0.000	0.000	-0.033	0.000	0.000	-0.005	0.005	-0.016
Politifact_27	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Tabela III.2: Diferença entre a tabela II.19 e a tabela II.3

	L	I	LP
Politifact_3	0.005	0.166	-0.328
Politifact_4	0.084	0.202	-0.015
Politifact_5	n/a	n/a	n/a
Politifact_8	-0.017	0.205	0.032
Politifact_14	-0.066	0.000	-0.059
Politifact_17	-0.056	-0.400	-0.023
Politifact_20	-0.162	-0.074	-0.020
Politifact_27	n/a	n/a	n/a

Tabela III.3: Diferença entre a tabela II.13 e a tabela II.5

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_6	-0.009	0.023	0.001	-0.012	-0.006	0.000	0.065	-0.059	0.000	0.082	-0.121	0.000	0.040	-0.034	0.000	0.050	-0.044	-0.010
Politifact_7	0.003	-0.045	0.001	-0.036	0.101	-0.499	-0.021	-0.012	-0.899	-0.149	-0.007	-0.899	-0.066	0.011	-0.599	-0.133	-0.004	-0.728
Politifact_9	-0.007	-0.033	-0.020	-0.048	-0.095	0.083	-0.013	-0.058	-0.031	-0.006	-0.066	-0.164	0.063	-0.087	-0.062	0.013	-0.062	-0.068
Politifact_12	0.002	-0.085	-0.017	0.022	-0.095	0.000	-0.017	-0.026	0.000	0.015	-0.078	0.000	0.056	-0.082	0.000	0.013	-0.054	-0.012
Politifact_18	0.053	-0.028	0.013	0.029	-0.011	0.000	0.036	-0.005	0.000	0.088	-0.019	0.000	0.054	-0.003	0.000	0.056	-0.015	-0.019
Politifact_19	0.013	0.001	0.033	-0.030	0.001	0.001	-0.067	0.000	0.001	-0.023	0.001	0.000	0.045	0.034	0.000	0.000	0.018	0.034
Politifact_22	0.072	-0.096	0.003	0.096	-0.071	0.000	0.088	-0.045	0.000	0.167	0.118	0.000	-0.057	0.054	0.000	0.075	0.002	0.007
Politifact_24	-0.047	0.134	0.000	0.033	0.184	0.001	-0.035	0.001	0.001	0.010	-0.005	0.001	-0.016	0.000	0.001	0.002	0.030	-0.019

Tabela III.4: Diferença entre a tabela II.15 e a tabela II.3

	L	I	LP
Politifact_6	0.221	0.137	0.280
Politifact_7	-0.200	-0.158	-0.199
Politifact_9	0.067	0.508	0.058
Politifact_12	0.170	-0.146	0.066
Politifact_18	0.110	0.378	0.237
Politifact_19	-0.117	0.137	-0.204
Politifact_21	0.005	0.200	0.104
Politifact_22	-0.078	0.064	-0.111
Politifact_24	0.023	-0.298	0.099

Tabela III.5: Diferença entre a tabela II.21 e a tabela II.5

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_6	0.013	0.008	0.006	0.020	-0.007	0.000	0.049	-0.012	0.000	0.048	-0.057	0.000	0.032	-0.034	0.000	0.037	-0.023	0.001
Politifact_7	0.024	-0.117	0.000	-0.028	-0.006	-0.111	-0.005	-0.012	-0.600	0.000	-0.008	-0.700	0.100	0.011	-0.599	0.016	-0.020	-0.514
Politifact_9	-0.022	0.031	-0.006	-0.044	0.015	0.108	-0.040	0.014	0.000	-0.015	-0.017	-0.125	-0.013	-0.004	-0.050	-0.024	0.006	-0.056
Politifact_12	-0.002	-0.023	0.011	0.009	-0.102	0.000	-0.010	-0.067	0.000	-0.007	-0.143	0.000	-0.071	-0.105	0.000	-0.009	-0.081	-0.010
Politifact_18	0.062	0.011	0.016	0.012	0.014	0.000	0.014	0.001	0.000	0.061	0.014	0.000	0.007	0.028	0.000	0.027	0.018	0.000
Politifact_19	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Politifact_22	0.063	-0.044	-0.002	-0.006	-0.115	0.333	0.020	-0.061	0.199	0.056	0.036	0.000	-0.080	0.053	0.000	0.010	0.000	0.081
Politifact_24	-0.016	0.133	0.009	0.009	0.183	0.000	-0.061	0.000	0.000	-0.035	-0.006	0.000	-0.025	0.000	0.000	0.009	0.030	-0.017

Tabela III.6: Médias da diferença entre a tabela II.22 e a tabela II.6.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.032	0.020	0.016	0.012	0.019	0.056	0.015	-0.024	0.001	0.038	-0.052	-0.010	0.031	-0.006	0.007	0.032	-0.018	0.001

Tabela III.7: Diferença entre a tabela II.19 e a tabela II.3

	L	I	LP
Politifact_6	-0.028	0.020	-0.052
Politifact_7	0.027	0.180	-0.071
Politifact_9	-0.036	-0.014	-0.037
Politifact_12	-0.089	0.046	-0.027
Politifact_18	0.078	-0.107	-0.068
Politifact_19	n/a	n/a	n/a
Politifact_21	-0.103	0.000	0.032
Politifact_22	0.010	0.027	-0.087
Politifact_24	-0.090	0.133	-0.001

Tabela III.8: Distribuição dos disseminadores de notícias falsas pelo número de seguidores.

	0 - 100	%	100 - 1k	%	1k - 5k	%	5k - 10k	%	10k - 50k	%	50k - 100k	%	100k - 500k	%	500k - 1M	%	1M+	%
Politifact_3	817	0.288	1,270	0.448	595	0.210	88	0.031	52	0.018	5	0.002	5	0.002	0	0.000	0	0.000
Politifact_4	1,340	0.286	2,171	0.463	137	0.029	137	0.029	69	0.015	5	0.001	4	0.001	1	0.000	0	0.000
Politifact_5	406	0.256	873	0.551	245	0.155	20	0.013	34	0.021	3	0.002	2	0.001	0	0.000	0	0.000
Politifact_8	238	0.195	495	0.405	359	0.294	66	0.054	51	0.042	2	0.002	8	0.007	1	0.001	2	0.002
Politifact_14	209	0.145	528	0.367	496	0.345	96	0.067	99	0.069	7	0.005	4	0.003	0	0.000	0	0.000
Politifact_17	195	0.194	418	0.417	284	0.283	61	0.061	42	0.042	2	0.002	1	0.001	0	0.000	0	0.000
Politifact_20	1,325	0.281	1,836	0.390	1,146	0.243	206	0.044	161	0.034	14	0.003	13	0.003	4	0.001	3	0.001
Politifact_27	769	0.273	1,332	0.473	581	0.206	84	0.030	43	0.015	7	0.002	1	0.000	0	0.000	0	0.000

Tabela III.9: Distribuição total dos disseminadores de notícias falsas pelo número de seguidores em percentagens.

	0 - 100%	100 - 1k%	1k - 5k%	5k - 10k%	10k - 50k%	50k - 100k%	100k - 500k %	500k - 1M %	1M+ %
%	0.272	0.458	0.197	0.039	0.028	0.002	0.002	0.000	0.000

Tabela III.10: Media das distribuições dos disseminadores de notícias falsas que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	11.50	143.88	263.13	59.88	42.63	3.25	1.88	0.13	0.00

Tabela III.11: Distribuição dos disseminadores de notícias verdadeiras que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	0	3	28	17	15	3	1	0	0
Politifact_7	0	0	4	0	3	0	0	0	0
Politifact_9	0	9	22	11	15	1	1	0	0
Politifact_12	0	2	5	5	8	0	1	0	0
Politifact_18	0	1	0	0	0	0	0	0	0
Politifact_19	0	0	1	0	1	0	0	0	0
Politifact_21	44	912	813	196	150	10	6	0	0
Politifact_22	1	6	13	4	7	4	0	0	0
Politifact_24	0	11	48	17	5	1	1	0	0

Tabela III.12: Media das distribuições dos disseminadores de notícias verdadeiras que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	5.00	104.89	103.78	27.78	22.67	2.11	1.11	0.00	0.00

Tabela III.13: Distribuição dos disseminadores de notícias falsas que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	0	35	88	24	12	2	1	0	0
Politifact_4	3	72	165	39	20	0	1	0	0
Politifact_5	0	0	0	0	0	0	0	0	0
Politifact_8	0	21	86	33	28	0	1	0	0
Politifact_14	0	28	163	48	61	1	1	0	0
Politifact_17	0	41	82	30	26	0	0	0	0
Politifact_20	3	130	444	128	99	5	0	0	0
Politifact_27	0	0	0	0	0	0	0	0	0

Tabela III.14: Media das distribuições dos disseminadores de notícias falsas que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	0.75	40.88	128.50	37.75	30.75	1.00	0.50	0.00	0.00

Tabela III.15: Media das distribuições dos disseminadores de notícias verdadeiras que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	0.33	29.11	54.11	17.67	16.78	1.11	0.56	0.00	0.00

Tabela III.16: Media das distribuições dos disseminadores de notícias falsas que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	10.25	159.00	271.75	61.50	44.63	3.63	3.13	0.63	0.63

Tabela III.17: Media das distribuições dos disseminadores de notícias verdadeiras que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	3.89	106.44	108.00	28.67	24.56	2.44	3.00	0.67	0.89

Tabela III.18: Média das pontuações das noticia falsas.

	Pontuação Alcance	Pontuação Influência	Pontuação Alcance nos Seguidores
Politifact_3	0.100	0.335	85.929
Politifact_4	0.061	0.923	278.787
Politifact_5	0.216	1.358	312.665
Politifact_8	0.158	0.134	42.772
Politifact_14	0.170	0.234	79.731
Politifact_17	0.114	0.666	298.921
Politifact_20	0.134	0.647	168.889
Politifact_27	0.105	1.257	594.053
Média	0.132	0.694	232.718

Tabela III.19: Média das pontuações das noticia verdadeiras.

	Pontuação Alcance	Pontuação Influência	Pontuação Alcance nos Seguidores
Politifact_6	0.067	0.063	23.364
Politifact_7	0.310	0.281	78.917
Politifact_9	0.051	9.471	985.469
Politifact_12	0.199	0.051	9.084
Politifact_18	0.444	0.810	132.504
Politifact_19	0.616	0.068	25.272
Politifact_21	0.191	0.405	40.282
Politifact_22	0.131	0.078	23.844
Politifact_24	0.272	0.088	34.480
Média	0.254	1.257	150.357

Tabela III.20: Pontuação do Alcance distribuído pelo número de seguidores, para as notícias falsas.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	324.766	66.733	70.123	31.637	13.321	7.822	0.345	0.000	0.000
Politifact_4	1,383.762	139.014	48.821	34.855	24.927	12.604	0.334	0.000	0.000
Politifact_5	2,473.138	98.793	46.891	9.444	85.210	0.000	0.000	0.000	0.000
Politifact_8	149.978	55.372	39.464	10.973	8.332	0.000	5.312	0.000	0.461
Politifact_14	624.591	62.058	57.444	19.458	6.988	6.996	1.580	0.000	0.000
Politifact_17	2,078.174	56.601	25.151	11.422	9.065	0.838	2.518	0.000	0.000
Politifact_20	2,652.549	47.414	25.274	12.934	7.313	3.520	3.343	6.892	0.173
Politifact_27	5,305.777	118.420	55.447	39.077	8.300	17.460	0.000	0.000	0.000
Média	1,874.09	80.55	46.08	21.23	20.43	6.16	1.68	0.86	0.08

Tabela III.21: Alcance distribuído pelo número de seguidores, para as notícias verdadeiras.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	70.730	59.952	26.939	11.044	5.718	1.171	0.891	0.276	0.757
Politifact_7	0.000	482.754	37.560	30.250	6.007	5.218	3.808	0.000	0.000
Politifact_9	19,420.076	52.738	19.871	16.526	5.630	2.677	1.592	0.098	0.119
Politifact_12	17.500	18.957	16.578	8.970	4.663	1.651	0.750	0.000	3.466
Politifact_18	975.858	45.606	19.509	0.000	20.539	1.523	11.294	0.000	3.025
Politifact_19	0.000	11.962	41.005	0.000	4.739	4.211	0.000	0.000	0.703
Politifact_21	221.131	83.867	38.959	31.124	10.355	1.646	2.417	0.928	4.207
Politifact_22	229.985	70.484	21.961	10.596	5.764	3.188	1.234	0.000	0.177
Politifact_24	104.177	61.197	24.219	33.626	7.492	9.643	2.705	1.312	0.000
Média	2,337.717	98.613	27.400	15.793	7.879	3.436	2.743	0.291	1.384

Tabela III.22: Pontuação da influência distribuído pelo número de seguidores, para as notícias falsas.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	1.670	0.177	0.217	0.153	0.041	0.034	0.000	0.000	0.000
Politifact_4	4.844	0.377	0.126	0.142	0.085	0.088	0.000	0.000	0.000
Politifact_5	9.139	0.627	0.313	0.021	0.621	0.000	0.000	0.000	0.000
Politifact_8	0.764	0.108	0.095	0.015	0.015	0.000	0.030	0.000	0.003
Politifact_14	2.402	0.109	0.124	0.021	0.007	0.012	0.004	0.000	0.000
Politifact_17	4.808	0.092	0.024	0.010	0.012	0.000	0.030	0.000	0.000
Politifact_20	10.795	0.074	0.053	0.013	0.012	0.008	0.016	0.073	0.001
Politifact_27	10.968	0.302	0.130	0.138	0.010	0.093	0.000	0.000	0.000
Média	5.67	0.23	0.14	0.06	0.10	0.03	0.01	0.01	0.00

Tabela III.23: Pontuação da influência distribuído pelo número de seguidores, para as notícias verdadeiras.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	0.384	0.112	0.083	0.016	0.029	0.004	0.002	0.000	0.013
Politifact_7	0.000	1.779	0.098	0.104	0.015	0.041	0.016	0.000	0.000
Politifact_9	189.891	0.096	0.027	0.039	0.009	0.008	0.013	0.000	0.000
Politifact_12	0.667	0.161	0.043	0.015	0.006	0.002	0.011	0.000	0.023
Politifact_18	5.451	0.310	0.187	0.000	0.238	0.013	0.150	0.000	0.041
Politifact_19	0.000	0.023	0.099	0.000	0.019	0.055	0.000	0.000	0.005
Politifact_21	10.274	0.249	0.109	0.195	0.017	0.004	0.009	0.006	0.064
Politifact_22	1.448	0.133	0.024	0.009	0.012	0.008	0.015	0.000	0.000
Politifact_24	0.346	0.171	0.034	0.078	0.023	0.022	0.022	0.009	0.000
Média	23.162	0.337	0.078	0.051	0.041	0.018	0.027	0.002	0.016

Tabela III.24: Diferença entre a tabela II.9 e a tabela II.1

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Politifact_3	0.000	0.001	0.010	-0.027	-0.001	0.009	-0.044	0.000	-0.074	0.026	-0.017	0.060	0.080	-0.005	0.001	0.016	-0.007	-0.001
Politifact_4	0.051	0.001	0.007	-0.006	0.000	0.011	-0.010	-0.021	0.000	0.005	0.000	0.015	0.022	0.000	0.006	0.003	-0.001	0.004
Politifact_5	-0.041	0.007	0.022	0.001	0.024	0.001	-0.013	0.001	0.001	0.042	0.001	0.001	0.079	-0.011	-0.166	0.015	-0.001	-0.024
Politifact_8	-0.082	-0.018	0.008	-0.076	-0.033	0.001	-0.038	0.062	0.001	-0.002	-0.185	0.001	0.078	0.000	0.001	-0.007	-0.029	0.001
Politifact_14	-0.062	0.001	0.029	-0.078	0.001	0.001	-0.044	0.001	0.001	-0.057	-0.024	0.001	0.032	-0.021	0.001	-0.036	-0.012	-0.015
Politifact_17	0.040	0.048	0.037	-0.053	0.012	0.001	-0.059	-0.099	-0.099	0.021	-0.024	0.001	-0.045	-0.033	0.001	-0.050	0.006	0.006
Politifact_20	0.018	-0.101	0.014	0.010	-0.080	-0.266	0.035	-0.018	0.001	0.017	-0.039	0.001	0.022	-0.041	0.001	0.021	-0.046	-0.013
Politifact_27	0.035	0.003	-0.002	-0.007	0.001	0.001	0.021	0.001	0.000	-0.061	-0.007	0.000	0.004	-0.016	0.000	-0.023	-0.002	-0.001

Tabela III.25: Médias da diferença entre a tabela II.10 e a tabela II.2.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	-0.005	-0.007	0.015	-0.029	-0.009	-0.030	-0.019	-0.009	-0.021	-0.001	-0.037	0.010	0.034	-0.016	-0.019	-0.008	-0.012	-0.005

Tabela III.26: Diferença entre a tabela II.11 e a tabela II.3

	L	I	LP
Politifact_3	0.000	0.000	0.000
Politifact_4	-0.002	-0.276	-0.022
Politifact_5	0.151	0.580	-0.089
Politifact_8	0.030	-0.107	0.041
Politifact_14	-0.084	0.020	0.101
Politifact_17	0.126	-0.466	-0.066
Politifact_20	-0.256	0.047	-0.025
Politifact_27	0.062	-0.114	0.310

Tabela III.27: Médias da diferença entre a tabela II.18 e a tabela II.2.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	-0.019	-0.021	-0.004	-0.024	0.005	-0.048	-0.040	-0.006	-0.009	-0.040	-0.040	-0.013	-0.006	-0.049	-0.008	-0.028	-0.020	-0.007

Tabela III.28: Médias da diferença entre a tabela II.14 e a tabela II.6.

	AP@1			AP@3			AP@5			AP@10			AP@15			@MAP		
	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP	L	I	LP
Média	0.005	-0.011	0.001	0.003	-0.004	-0.046	-0.003	-0.024	-0.103	0.023	-0.020	-0.118	0.021	-0.014	-0.073	0.009	-0.015	-0.089

Tabela III.29: Media da diferença entre a tabela II.12 e a tabela II.4

	L	I	LP
Média	0.003	-0.040	0.031

Tabela III.30: Media da diferença entre a tabela II.20 e a tabela II.4

	L	I	LP
Média	-0.035	0.017	-0.069

Tabela III.31: Media da diferença entre a tabela II.16 e a tabela II.8

	L	I	LP
Média	0.040	0.086	0.067

Tabela III.32: Media da diferença entre a tabela II.24 e a tabela II.8

	L	I	LP
Média	-0.029	0.036	-0.039

Tabela III.33: Distribuição dos disseminadores de notícias verdadeiras pelo número de seguidores.

	0 - 100	%	100 - 1k	%	1k - 5k	%	5k - 10k	%	10k - 50k	%	50k - 100k	%	100k - 500k	%	500k - 1M	%	1M+	%
Politifact_6	559	0.203	1,331	0.483	629	0.228	109	0.040	100	0.036	13	0.005	8	0.003	3	0.001	1	0.000
Politifact_7	114	0.133	269	0.315	141	0.165	33	0.039	36	0.042	4	0.005	3	0.004	0	0.000	1	0.001
Politifact_9	439	0.244	751	0.417	428	0.238	70	0.039	95	0.053	7	0.004	6	0.003	1	0.001	3	0.002
Politifact_12	363	0.192	864	0.458	481	0.255	76	0.040	81	0.043	11	0.006	9	0.005	0	0.000	1	0.001
Politifact_13	130	0.097	664	0.495	403	0.300	74	0.055	62	0.046	5	0.004	4	0.003	0	0.000	0	0.000
Politifact_19	131	0.262	234	0.468	107	0.214	12	0.024	12	0.024	2	0.004	1	0.002	0	0.000	1	0.002
Politifact_21	776	0.163	2,529	0.531	1,037	0.218	216	0.045	169	0.035	15	0.003	13	0.003	4	0.001	2	0.000
Politifact_22	136	0.206	247	0.375	168	0.255	39	0.059	48	0.073	14	0.021	5	0.008	0	0.000	2	0.003
Politifact_24	170	0.233	288	0.394	191	0.261	40	0.055	32	0.044	3	0.004	6	0.008	1	0.001	0	0.000

Tabela III.34: Distribuição total dos disseminadores de notícias verdadeiras pelo número de seguidores em percentagens.

	0 - 100%	100 - 1k%	1k - 5k%	5k - 10k%	10k - 50k%	50k - 100k%	100k - 500k %	500k - 1M %	1M+ %
%	0.187	0.477	0.238	0.045	0.042	0.005	0.004	0.001	0.001

Tabela III.35: Distribuição dos tweets de notícias falsas pelo número dos seguidores dos disseminadores.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	18	49	53	11	16	4	1	0	0
Politifact_4	51	133	104	27	20	3	2	0	0
Politifact_5	7	39	23	2	11	0	0	0	0
Politifact_8	8	15	32	9	15	0	5	0	2
Politifact_14	9	25	53	16	27	3	2	0	0
Politifact_17	16	33	43	15	14	2	1	0	0
Politifact_20	7	21	28	17	35	6	6	3	3
Politifact_27	27	29	46	5	6	3	0	0	0

Tabela III.36: Distribuição dos tweets de notícias verdadeiras pelo número dos seguidores dos disseminadores.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	2	18	38	20	25	5	2	1	1
Politifact_7	0	4	6	4	12	2	2	0	0
Politifact_9	8	34	49	18	38	5	5	1	3
Politifact_12	1	4	10	2	12	3	5	0	8
Politifact_18	3	5	6	0	7	2	2	0	3
Politifact_19	0	1	9	0	3	2	0	0	1
Politifact_21	4	22	27	16	30	3	4	2	2
Politifact_22	2	5	13	8	12	11	3	0	1
Politifact_24	3	11	15	8	7	1	4	1	0

Tabela III.37: Distribuição dos retweets de notícias falsas pelo número dos seguidores dos disseminadores.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	701	1,139	515	70	38	1	3	0	0
Politifact_4	1,076	1,823	787	101	52	2	1	1	0
Politifact_5	451	1,158	375	24	38	4	2	0	0
Politifact_8	135	360	236	41	28	1	4	1	2
Politifact_14	116	363	317	50	45	4	1	0	0
Politifact_17	23	116	83	29	17	0	0	0	0
Politifact_20	1,214	1,740	1,050	183	133	10	8	2	0
Politifact_27	725	1,225	536	80	44	7	1	0	0

Tabela III.38: Distribuição dos retweets de notícias verdadeiras pelo número dos seguidores dos disseminadores.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	427	1,031	437	63	47	7	5	2	0
Politifact_7	85	236	109	23	20	3	1	0	1
Politifact_9	297	476	197	27	31	0	1	0	0
Politifact_12	324	810	430	56	61	6	4	0	0
Politifact_18	1,076	2,109	1,102	106	59	9	4	0	0
Politifact_19	123	214	89	9	8	0	1	0	1
Politifact_21	509	2,513	909	208	126	12	10	3	0
Politifact_22	62	124	73	14	20	4	2	0	1
Politifact_24	121	246	149	28	20	2	3	0	0

Tabela III.39: Distribuição dos disseminadores de notícias falsas que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	2	86	161	34	23	3	3	0	0
Politifact_4	17	186	329	72	31	3	1	1	0
Politifact_5	0	0	0	0	0	0	0	0	0
Politifact_8	6	69	158	46	38	1	3	0	0
Politifact_14	17	220	408	92	87	7	2	0	0
Politifact_17	2	74	182	49	34	1	0	0	0
Politifact_20	47	508	835	172	124	11	6	0	0
Politifact_27	1	8	32	14	4	0	0	0	0

Tabela III.40: Distribuição dos disseminadores de notícias verdadeiras que são seguidos por contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	0	2	7	1	5	1	0	0	0
Politifact_7	0	0	2	0	1	0	0	0	0
Politifact_9	0	0	12	5	8	0	0	0	0
Politifact_12	0	0	1	2	2	0	0	0	0
Politifact_18	0	0	0	0	0	0	0	0	0
Politifact_19	0	0	0	0	0	0	0	0	0
Politifact_21	3	260	461	151	131	7	5	0	0
Politifact_22	0	0	4	0	4	2	0	0	0
Politifact_24	0	0	0	0	0	0	0	0	0

Tabela III.41: Distribuição dos disseminadores de notícias falsas que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	1	101	163	33	23	3	3	0	0
Politifact_4	17	207	333	79	33	4	2	1	0
Politifact_5	0	4	3	0	3	0	0	0	0
Politifact_8	4	77	158	46	40	2	7	1	2
Politifact_14	9	219	419	93	88	7	3	0	0
Politifact_17	1	70	184	51	33	1	1	0	0
Politifact_20	50	582	880	175	131	12	9	3	3
Politifact_27	0	12	34	15	6	0	0	0	0

Tabela III.42: Distribuição dos disseminadores de notícias falsas que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	0	46	98	24	11	0	1	0	0
Politifact_4	6	89	173	45	21	2	1	0	0
Politifact_5	0	0	0	0	0	0	0	0	0
Politifact_8	0	25	88	33	29	0	1	0	0
Politifact_14	0	28	152	49	58	2	1	0	0
Politifact_17	0	34	77	30	27	0	0	0	0
Politifact_20	3	136	444	127	103	6	7	3	1
Politifact_27	0	0	0	0	0	0	0	0	0

Tabela III.43: Media das distribuições dos disseminadores de notícias falsas que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	1.13	44.75	129.00	38.50	31.13	1.25	1.38	0.38	0.13

Tabela III.44: Distribuição dos disseminadores de notícias verdadeiras que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	0	1	7	1	5	0	0	0	0
Politifact_7	0	0	1	0	3	0	0	0	0
Politifact_9	0	1	13	6	8	0	1	0	0
Politifact_12	0	0	0	2	3	0	0	0	0
Politifact_18	0	0	0	0	0	0	0	0	0
Politifact_19	0	0	0	0	0	0	0	0	0
Politifact_21	1	246	468	149	132	7	5	1	2
Politifact_22	0	0	5	0	4	2	1	0	1
Politifact_24	0	0	0	0	0	0	0	0	0

Tabela III.45: Media das distribuições dos disseminadores de notícias verdadeiras que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	0.11	27.56	54.89	17.56	17.22	1.00	0.78	0.11	0.33

Tabela III.46: Distribuição das contas *bot* que disseminaram notícias falsas pelo número dos seguidores dessas mesmas contas (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	2	0	5	7	6	0	0	0	0
Politifact_4	9	3	13	19	7	0	0	0	0
Politifact_5	2	0	0	0	0	0	0	0	0
Politifact_8	3	0	7	10	10	0	0	0	0
Politifact_14	2	0	18	19	23	1	0	0	0
Politifact_17	4	0	4	11	12	0	0	0	0
Politifact_20	9	1	31	31	29	1	0	0	0
Politifact_27	1	0	1	2	0	0	0	0	0

Tabela III.47: Media da distribuição das contas *bot* que disseminaram notícias falsas pelo número dos seguidores dessas mesmas contas (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	4.00	0.50	9.88	12.38	10.88	0.25	0.00	0.00	0.00

Tabela III.48: Distribuição das contas *bot* que disseminaram notícias verdadeiras pelo número dos seguidores dessas mesmas contas (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	4	1	6	7	6	0	0	0	0
Politifact_7	10	0	3	1	1	0	0	0	0
Politifact_9	5	1	6	2	5	0	0	0	0
Politifact_12	4	2	0	5	5	0	0	0	0
Politifact_18	6	3	2	1	1	0	0	0	0
Politifact_19	0	1	2	0	0	0	0	0	0
Politifact_21	0	0	20	35	46	1	1	0	0
Politifact_22	15	3	4	2	5	1	0	0	0
Politifact_24	3	0	4	4	2	0	0	0	0

Tabela III.49: Media da distribuição das contas *bot* que disseminaram notícias verdadeiras pelo número dos seguidores dessas mesmas contas (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	5.22	1.22	5.22	6.33	7.89	0.22	0.11	0.00	0.00

Tabela III.50: Distribuição das contas *bot* que disseminaram notícias falsas pelo número dos seguidores dessas mesmas contas (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	0	0	0	0	2	0	0	0	0
Politifact_4	1	0	0	4	1	0	0	0	0
Politifact_5	0	0	0	0	0	0	0	0	0
Politifact_8	0	0	0	2	4	0	0	0	0
Politifact_14	0	0	1	1	1	0	0	0	0
Politifact_17	0	0	1	1	3	0	0	0	0
Politifact_20	1	0	1	6	3	0	0	0	0
Politifact_27	0	0	0	0	0	0	0	0	0

Tabela III.51: Media da distribuição das contas *bot* que disseminaram notícias falsas pelo número dos seguidores dessas mesmas contas (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	0.25	0.00	0.38	1.75	1.75	0.00	0.00	0.00	0.00

Tabela III.52: Distribuição das contas *bot* que disseminaram notícias verdadeiras pelo número dos seguidores dessas mesmas contas (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	2	0	1	0	2	0	0	0	0
Politifact_7	2	0	0	0	1	0	0	0	0
Politifact_9	0	0	1	0	2	0	0	0	0
Politifact_12	1	1	0	0	1	0	0	0	0
Politifact_18	2	1	0	0	0	0	0	0	0
Politifact_19	0	0	0	0	0	0	0	0	0
Politifact_21	0	0	0	5	7	0	0	0	0
Politifact_22	4	0	0	2	0	0	0	0	0
Politifact_24	2	0	0	0	0	0	0	0	0

Tabela III.53: Media da distribuição das contas *bot* que disseminaram notícias verdadeiras pelo número dos seguidores dessas mesmas contas (CAP a 96%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Média	1.44	0.22	0.22	0.78	1.44	0.00	0.00	0.00	0.00

Tabela III.54: Distribuição dos disseminadores de notícias verdadeiras que seguem contas *bot* pelo número dos seguidores dos disseminadores (CAP a 90%).

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	0	4	28	17	17	3	3	1	0
Politifact_7	0	0	3	0	5	0	0	0	1
Politifact_9	0	8	25	12	19	2	3	0	2
Politifact_12	0	2	3	6	10	0	2	0	1
Politifact_18	0	0	0	0	1	0	1	0	1
Politifact_19	0	0	1	0	1	0	0	0	0
Politifact_21	34	932	856	201	156	14	12	4	2
Politifact_22	0	1	10	3	4	2	2	0	1
Politifact_24	1	11	46	19	8	1	4	1	0

Tabela III.55: Pontuação do alcance nos seguidores distribuído pelo número de seguidores, para as notícias falsas.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_3	324.766	66.733	70.123	31.637	13.321	7.822	0.345	0.000	0.000
Politifact_4	1,383.762	139.014	48.821	34.855	24.927	12.604	0.334	0.000	0.000
Politifact_5	2,473.138	98.793	46.891	9.444	85.210	0.000	0.000	0.000	0.000
Politifact_8	149.978	55.372	39.464	10.973	8.332	0.000	5.312	0.000	0.461
Politifact_14	624.591	62.058	57.444	19.458	6.988	6.996	1.580	0.000	0.000
Politifact_17	2,078.174	56.601	25.151	11.422	9.065	0.838	2.518	0.000	0.000
Politifact_20	2,652.549	47.414	25.274	12.934	7.313	3.520	3.343	6.892	0.173
Politifact_27	5,305.777	118.420	55.447	39.077	8.300	17.460	0.000	0.000	0.000
Média	1,874.09	80.55	46.08	21.23	20.43	6.16	1.68	0.86	0.08

Tabela III.56: Pontuação do alcance nos seguidores distribuído pelo número de seguidores, para as notícias verdadeiras.

	0 - 100	100 - 1k	1k - 5k	5k - 10k	10k - 50k	50k - 100k	100k - 500k	500k - 1M	1M+
Politifact_6	70.730	59.952	26.939	11.044	5.718	1.171	0.891	0.276	0.757
Politifact_7	0.000	482.754	37.560	30.250	6.007	5.218	3.808	0.000	0.000
Politifact_9	19,420.076	52.738	19.871	16.526	5.630	2.677	1.592	0.098	0.119
Politifact_12	17.500	18.957	16.578	8.970	4.663	1.651	0.750	0.000	3.466
Politifact_18	975.858	45.606	19.509	0.000	20.539	1.523	11.294	0.000	3.025
Politifact_19	0.000	11.962	41.005	0.000	4.739	4.211	0.000	0.000	0.703
Politifact_21	221.131	83.867	38.959	31.124	10.355	1.646	2.417	0.928	4.207
Politifact_22	229.985	70.484	21.961	10.596	5.764	3.188	1.234	0.000	0.177
Politifact_24	104.177	61.197	24.219	33.626	7.492	9.643	2.705	1.312	0.000
Média	2,337.717	98.613	27.400	15.793	7.879	3.436	2.743	0.291	1.384

Anexo IV

Listagens

Listagem IV.1: Botometer JSON

```
{
   "cap": {
      "english": 0.8018818614025648,
      "universal": 0.5557322218336633
   },
   "display_scores": {
      "english": {
         "astroturf": 0.0,
         "fake_follower": 4.1,
         "financial": 1.5,
         "other": 4.7,
         "overall": 4.7,
         "self_declared": 3.2,
         "spammer": 2.8
      },
      "universal": {
         "astroturf": 0.3,
         "fake_follower": 3.2,
         "financial": 1.6,
         "other": 3.8,
         "overall": 3.8,
         "self_declared": 3.7,
         "spammer": 2.3
```

```
        }
    },
    "raw_scores": {
        "english": {
            "astroturf": 0.0,
            "fake_follower": 0.81,
            "financial": 0.3,
            "other": 0.94,
            "overall": 0.94,
            "self_declared": 0.63,
            "spammer": 0.57
        },
        "universal": {
            "astroturf": 0.06,
            "fake_follower": 0.64,
            "financial": 0.3133333333333333,
            "other": 0.76,
            "overall": 0.76,
            "self_declared": 0.74,
            "spammer": 0.47
        }
    },
    "user": {
        "majority_lang": "en",
        "user_data": {
            "id_str": "11330",
            "screen_name": "test_screen_name"
        }
    }
}
```