



Universidade Federal
do Rio de Janeiro

Escola Politécnica

CLASSIFICAÇÃO DE PERFIS FALSOS EM REDES SOCIAIS ONLINE
UTILIZANDO A ESTRUTURA DA REDE

Rafael Gonçalves Damasceno

Projeto de Graduação apresentado ao Curso de Computação e Informação da Escola Politécnica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

Orientador: Daniel Ratton Figueiredo

Rio de Janeiro
Março de 2019

CLASSIFICAÇÃO DE PERFIS FALSOS EM REDES SOCIAIS ONLINE
UTILIZANDO A ESTRUTURA DA REDE

Rafael Gonçalves Damasceno

PROJETO SUBMETIDO AO CORPO DOCENTE DO CURSO DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE COMPUTAÇÃO E INFORMAÇÃO.

Examinadores:

Prof. Daniel Ratton Figueiredo, PhD

Prof. Fábio Happ Botler, DSc

Prof. Luidi Gelabert Simonetti, DSc

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2019

Gonçalves Damasceno, Rafael

Classificação de Perfis Falsos em Redes Sociais Online utilizando a Estrutura da Rede/Rafael Gonçalves Damasceno. – Rio de Janeiro: UFRJ/POLI – COPPE, 2019.

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

Orientador: Daniel Ratton Figueiredo

Projeto (graduação) – UFRJ/ Escola Politécnica/ Curso de Computação e Informação, 2019.

Referências Bibliográficas: p. 31 – 32.

1. Ciência das Redes. 2. Redes Complexas. 3. Redes Sociais. 4. Detecção de Bots. 5. Grafos. 6. Teoria dos Grafos. I. Ratton Figueiredo, Daniel. II. Universidade Federal do Rio de Janeiro, Escola Politécnica/ Curso de Computação e Informação. III. Título.

Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais, Gloria e Roberto, que sempre me apoiaram, incentivaram e acreditaram nos meus sonhos.

Agradeço a todos os amigos que estiveram ao meu lado durante esta trajetória, compartilhando momentos de alegrias e tristezas. O ambiente universitário não seria o mesmo sem a troca de experiências e as amizades.

A Universidade Federal do Rio de Janeiro, na qual eu tive o prazer de passar os meus últimos anos, o meu muito obrigado pelo ambiente de excelência acadêmica. Em especial, ao corpo docente que contribuiu com dedicação para esta jornada.

Ao professor Daniel Ratton, meu orientador neste trabalho, obrigado por compartilhar os seus conhecimentos e pelo auxílio durante esta realização.

Por fim, agradeço a todos que direta ou indiretamente fizeram parte da minha formação.

Resumo do Projeto de Graduação apresentado à Escola Politécnica/COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

CLASSIFICAÇÃO DE PERFIS FALSOS EM REDES SOCIAIS ONLINE UTILIZANDO A ESTRUTURA DA REDE

Rafael Gonçalves Damasceno

Março/2019

Orientador: Daniel Ratton Figueiredo

Curso: Engenharia de Computação e Informação

O estudo de Redes Complexas se mostra essencial para compreender diferentes interações e relacionamentos nas redes de múltiplos campos da ciência. Nesta perspectiva, o uso de uma abstração em rede permite codificar relacionamentos entre pares de objetos e, a partir disto, podem ser extraídas métricas estruturais da rede para entender o como, o porquê e as consequências destes relacionamentos. Dentro do âmbito de redes sociais online, as mesmas desempenham um papel central na sociedade moderna, principalmente pelo debate de opiniões. Entretanto, a presença de perfis falsos surge como uma preocupação diante de questões como disseminação de informações, fenômeno conhecido como epidemia em redes, e o uso das redes sociais para influenciar opiniões. Este trabalho busca, a partir da modelagem de uma rede induzida pelos comentários em redes sociais, calcular e avaliar métricas estruturais relevantes para o problema de identificação de perfis falsos. Ao final, estas métricas serão utilizadas por um modelo classificador com o objetivo de reforçar a importância do paradigma de rede e o campo de Ciência das Redes como ferramentas para resolução do problema.

Palavras-Chave: Ciência das Redes, Redes Complexas, Redes Sociais, Detecção de Bots, Grafos, Teoria dos Grafos.

Abstract of the Undergraduate Project presented to Poli/COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Computer and Information Engineer.

CLASSIFICATION OF FALSE PROFILES IN ONLINE SOCIAL NETWORKS USING THE NETWORK STRUCTURE

Rafael Gonçalves Damasceno

March/2019

Advisor: Daniel Ratton Figueiredo

Course: Computer and Information Engineering

The study of Complex Networks is fundamental to understand different interactions and relationships in the networks of multiple fields of science. In this perspective, the use of a network abstraction allows the encoding of relationships between pairs of objects and, from this, structural metrics can be extracted from the network to understand how, why and the consequences of these relationships. Within the scope of online social networks, they play a central role in modern society, mainly through the debate of opinions. However, the presence of false profiles emerges as a concern regarding issues such as dissemination of information, a phenomenon known as epidemic in networks, and the use of social networks to influence opinions. This project seeks, based on the modeling of a network induced by the comments in social networks, to calculate and evaluate structural metrics relevant to the false profiles identification problem. Finally, these metrics will be used by a classifier model with the goal of reinforcing the importance of the network paradigm and the Network Science field as tools to solve the problem.

Keywords: Network Science, Complex Network, Social Network, Bots Detection, Graphs, Graph Theory.

Sumário

Agradecimentos	iv
Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	1
1.2 Objetivo	2
1.3 Organização do Trabalho	3
2 Conceitos, Definições e Trabalhos Relacionados	4
2.1 Conceitos e Definições	4
2.1.1 Teoria dos Grafos e Redes Complexas	4
2.1.2 Aprendizado de Máquina	6
2.2 Trabalhos Relacionados	7
2.2.1 <i>The DARPA Twitter Bot Challenge</i>	7
2.2.2 <i>Online Human-Bot Interactions: Detection, Estimation, and Characterization</i>	8
3 Dataset e Modelagem da Rede	9
3.1 Reddit como Rede Social	9
3.2 <i>Dataset</i>	9
3.2.1 <i>Dataset de Comentários</i>	9
3.2.2 Lista de usuários <i>Bots</i>	10
3.3 Modelagem da Rede	11
4 Caracterização Estrutural de Usuários e Bots	15
4.1 Componentes da Rede Modelada	15
4.2 Métricas	17
5 Classificação de Perfis Falsos	24
5.1 Avaliação e Tratamento das Métricas para o Modelo	24

5.2	Aplicação e Avaliação de uma Rede Neural	26
6	Conclusão	29
6.1	Trabalhos Futuros	30
	Referências Bibliográficas	31
A	Métricas Complementares	33
A.1	Maior Componente Fracamente Conexa	33
B	Avaliação Complementar para o Modelo	38
B.1	Maior Componente Fracamente Conexa	38

Lista de Figuras

3.1	Exemplo de dados utilizados para modelar a rede	10
3.2	Arranjo da rede a partir do exemplo de dados de um único comentário	12
3.3	Relacionamento formado a partir do arranjo de um único comentário e as informações dos outros comentários	13
4.1	Função de distribuição cumulativa complementar (CCDF) do grau de entrada e peso de entrada dos vértices na Maior Componente Fortemente Conexa	18
4.2	Função de distribuição cumulativa complementar (CCDF) do grau de saída e peso de saída dos vértices na Maior Componente Fortemente Conexa	19
4.3	Função de distribuição cumulativa complementar (CCDF) da razão do grau e razão do peso dos vértices na Maior Componente Fortemente Conexa	20
4.4	Gráfico de dispersão entre o grau de saída e o grau de entrada dos vértices na Maior Componente Fortemente Conexa	21
4.5	Gráfico de dispersão entre o grau de saída e o coeficiente de clusteração dos vértices na Maior Componente Fortemente Conexa	22
5.1	Matriz de Correlação para as características dos vértices na Maior Componente Fortemente Conexa	25
5.2	Avaliação do modelo de Rede Neural para Recall, Precisão e F1 Score	27
5.3	Avaliação do modelo de Rede Neural no Espaço ROC	28
A.1	Função de distribuição cumulativa complementar (CCDF) do grau de entrada e peso de entrada dos vértices na Maior Componente Fracamente Conexa	34
A.2	Função de distribuição cumulativa complementar (CCDF) do grau de saída e peso de saída dos vértices na Maior Componente Fracamente Conexa	34

A.3	Função de distribuição cumulativa complementar (CCDF) da razão do grau e razão do peso dos vértices na Maior Componente Fracamente Conexa	35
A.4	Gráfico de dispersão entre o grau de saída e o grau de entrada dos vértices na Maior Componente Fracamente Conexa	36
A.5	Gráfico de dispersão entre o grau de saída e o coeficiente de clusteração dos vértices na Maior Componente Fracamente Conexa	36
B.1	Matriz de Correlação para as características dos vértices na Maior Componente Fracamente Conexa	38

Lista de Tabelas

3.1	Avaliação preliminar da rede induzida a partir dos comentários	13
4.1	Avaliação das Componentes Fortemente Conexas da Rede	16
4.2	Avaliação das Componentes Fracamente Conexas da Rede	17
4.3	Primeiras métricas para a Maior Componente Fortemente Conexa da Rede	21
4.4	Outras métricas para a Maior Componente Fortemente Conexa da Rede	23
A.1	Primeiras métricas para a Maior Componente Fracamente Conexa da Rede	35
A.2	Outras métricas para a Maior Componente Fracamente Conexa da Rede	37

Capítulo 1

Introdução

O estudo de interações e relacionamentos nos diferentes campos da ciência, como, natural, social, de engenharia e de informação, pode compreender para a computação um paradigma de rede. Este paradigma permite pesquisar aspectos e reconhecer padrões que definem o como, o porquê e as consequências destes relacionamentos. O campo acadêmico que se aprofunda neste estudo pode ser chamado de Ciência das Redes, do inglês *Network Science*, ou simplesmente Redes Complexas.

O uso de uma abstração em rede permite codificar relacionamentos entre pares de objetos. Esta formulação em grafo é amplamente explorada para abordar problemas reais e permite a resolução de problemas clássicos da computação. A Teoria dos Grafos apresenta formas de explorar a sua estrutura e aplicar algoritmos para, por exemplo, encontrar o menor caminho entre dois objetos da rede ou então colorir vértices de forma que os pares conectados não compartilhem a mesma coloração.

No entanto, muitas redes possuem propriedades topológicas não-triviais. Nessa perspectiva, os estudos acerca dos motivos e consequências desses relacionamentos, além das formações estruturais, mostram-se necessários dentro de uma nova área multidisciplinar.

As redes sociais, alvo recorrente para o estudo de redes, conseguem representar relacionamentos sociais humanos, gerando um grande volume de dados com as redes, por exemplo, de amizades e troca de informação. Cada rede social possui suas particularidades ou formas de interação, mas, independente disso, desempenha papel central na sociedade moderna, atuando como meio para divulgação de notícias e debate de opiniões.

1.1 Motivação

Atualmente, o papel das redes sociais vem sendo constantemente debatido, principalmente em momentos marcantes, como durante revoluções populares e eleições de mandatos. Esta discussão, em geral, busca entender o comportamento da rede e a

forma como as informações são difundidas por ela. Um ponto de destaque nestas discussões está acerca do fenômeno conhecido como epidemia em redes, o qual permite disseminar uma informação - verdadeira ou falsa - através dos relacionamentos existentes.

Neste contexto, o espaço das redes sociais acaba se comportando como um ambiente formador de opinião. Recentemente, escândalos envolvendo o uso de dados das redes para influenciar opiniões de eleitores em candidaturas presidenciais ocuparam as capas dos jornais - como o caso envolvendo a empresa de análise de dados Cambridge Analytica e a eleição do presidente norte americano Donald Trump¹. Esta nova preocupação em torno da manipulação de opiniões levanta a atenção para diferentes algoritmos ou comportamentos não humanos voltados unicamente para este propósito.

A presença de usuários não humanos, mais conhecidos como *bots*², ocupa um papel particular dentro da referida discussão. Estes usuários podem possuir objetivos bem definidos, como realizar uma determinada interação com outro usuário sempre que um dado padrão ocorrer. Em certos casos, todavia, este objetivo não é tão claro para os outros usuários, principalmente quando o usuário *bot* busca se comportar e interagir como um usuário humano comum.

O uso de usuários *bots* em redes sociais, em especial nos últimos anos, provoca uma série de discussões sobre o seu papel dentro da rede e as consequências geradas pelo seu comportamento. Estes usuários podem ser manipulados com múltiplos objetivos, dentre eles, influenciar a opinião dentro da rede ou propagar determinado conteúdo. Além disso, parte destes comportamentos podem ferir a política da rede social, demandando esforços para identificar, controlar e em alguns casos reduzir os efeitos causados.

A partir da definição deste problema, evidencia-se a importância de haver uma forma de identificar tais usuários dentro da rede, permitindo adotar as medidas cabíveis em cada caso.

1.2 Objetivo

O presente trabalho se concentra em estabelecer, fazendo uso do paradigma de rede, uma identificação de usuários falsos em redes sociais. Esta caracterização pretende explorar o campo de Ciência das Redes, buscando analisar a estrutura do grafo induzido a partir dos dados de comentários da rede social Reddit.

¹Segundo o The New York Times, os dados coletados pela Cambridge incluíam detalhes sobre as identidades dos usuários, redes de amigos e "curtidas". [1]

²Algoritmo que produz conteúdo automaticamente e interage com os humanos nas mídias sociais, tentando imitar e possivelmente alterar seu comportamento. FERRARA *et al.* [2]

Diferentemente de outras abordagens, este trabalho compreende a exploração dos relacionamentos entre pares de usuários como estrutura. Sendo assim, não existe nenhum tratamento do conteúdo escrito por usuários, como uma análise de sentimentos, ou mesmo uma comparação utilizando o nome do usuário frente a outros nomes empregados dentro da rede.

Tendo isto em vista, esta proposta tem o escopo de apresentar e comparar diferentes características da rede de comentários para usuários conhecidos como *bot* ou não. Ademais, busca demonstrar como um classificador de Aprendizado de Máquina, do inglês *Machine Learning*, poderia ser empregado a partir das métricas extraídas.

1.3 Organização do Trabalho

Este trabalho está dividido em 6 capítulos. Após esta introdução, o capítulo 2 explicitará conceitos e definições relacionados a grafos e redes complexas, possibilitando uma melhor compreensão do problema e das técnicas empregadas ao longo do trabalho. Ainda na abordagem deste capítulo, também apresentará uma revisão sobre propostas existentes dentro do contexto de identificação e comportamento de usuários *bots* em redes sociais.

Em seguida, o capítulo 3 discorrerá sobre quais dados e como os mesmos foram empregados para a construção da rede em análise. Este capítulo destacará, também, o comportamento de interações dentro da rede social e suas implicações para a modelagem.

O capítulo 4 examinará a rede modelada a partir do cálculo e comparação de diferentes métricas dos usuários contidos nela.

Após as métricas, o capítulo 5 buscará interpretar quais comportamentos da rede podem estar sendo refletidos nestas medidas e, por fim, como um modelo de classificação pode ser empregado para o problema.

O sexto e último capítulo apresentará as considerações finais e a avaliação sobre o uso do paradigma de rede para o problema de identificação de usuários *bots* em uma rede social qualquer. Por fim, pretenderá apontar possíveis trabalhos futuros a partir dos resultados e conclusões obtidas.

Capítulo 2

Conceitos, Definições e Trabalhos Relacionados

Este capítulo introduz conceitos e definições que servem de base para o desenvolvimento deste trabalho. Além disso, também apresenta trabalhos relacionados que estudam o comportamento de *bots* em redes sociais. Algumas destas propostas podem servir como referências e formas de comparar a técnica empregada neste trabalho.

2.1 Conceitos e Definições

2.1.1 Teoria dos Grafos e Redes Complexas

Grafos

Grafo é uma abstração que codifica relacionamentos entre pares de objetos. O conjunto de objetos é definido como vértices e os relacionamentos são as arestas que conectam um par de vértices KLEINBERG e TARDOS [3].

Usualmente, utiliza-se a notação $G = (V, E)$ para representar um grafo, onde V é conjunto de vértices e E é o conjunto de arestas que formam o grafo G . Cada aresta $e \in E$ é representada como uma tupla $e_{u,v} = (u, v)$, onde $u, v \in V$.

Uma aresta pode conter um valor associado, geralmente chamado de peso, podendo ser atribuído como $w_{u,v}$, onde (u, v) é uma aresta do grafo. Este peso codifica a intensidade do relacionamento entre o par de vértices, trazendo um valor como, por exemplo, o número de mensagens trocadas entre os vértices (quanto maior o número de mensagens trocadas, maior o valor do peso).

Grafos Direcionados

Arestas em um grafo podem indicar uma relação simétrica ou assimétrica [3]. Quando um grafo codifica relacionamentos assimétricos este é dito direcionado, enquanto um grafo com relacionamentos simétricos é dito não-direcionado.

Em um grafo direcionado, a existência de uma aresta entre dois vértices é denotada por $u \rightarrow v$, onde $u, v \in V$, e não implica necessariamente em $v \rightarrow u$.

Vizinhos

Em grafos não-direcionados, para cada vértice $v \in V$ existe um subconjunto de vértices que possuem aresta incidente em v . Este subconjunto é denominado conjunto de vizinhos de v . Para grafos direcionados, u é dito vizinho de v quando existe o relacionamento $v \rightarrow u$.

Grau

Em grafos não-direcionados, o grau de um vértice pode ser definido como o número de vizinhos, ou seja, o número de vértices que são adjacentes a ele.

Para grafos direcionados, existem os conceitos de grau de entrada e grau de saída. O grau de entrada de um vértice v está associado ao número de vértices que possuem aresta incidindo nele. Já o grau de saída, está associado ao número de vértices que possuem uma aresta de v incidindo.

Componente Conexa

Componente Conexa é um subgrafo $H \subseteq G$, onde para todo vértice $v \in V$ é possível chegar em qualquer outro vértice $u \in V$ [3], percorrendo algum caminho pelas arestas. Quando este conceito é utilizado para um grafo direcionado, a componente pode ser dita Fracamente Conexa, denotando que a condição definida aqui pode ser aplicada caso as arestas sejam consideradas não direcionadas.

Componente Fortemente Conexa

Em grafos direcionados, uma componente fortemente conexa é um subgrafo $H \subseteq G$, onde para quaisquer pares (u, v) existe ao menos um caminho de u para v e de v para u [3]. Isto quer dizer que, em um subgrafo, todos os pares de vértices são mutualmente acessíveis, podendo ser percorrido um caminho através das arestas direcionadas de u para v e de v para u .

Reciprocidade

GARLASCHELLI e LOFFREDO [4] definem reciprocidade como a tendência de pares de vértices estarem mutualmente conectados em uma rede. Em outras palavras, a reciprocidade está interessada na existência de arestas entre dois vértices com direções opostas. Considerando dois vértices, u e $v \in V$, a métrica pretende quantificar a existência de $u \leftrightarrow v$.

Definindo conexão como a existência de um aresta de entrada ou saída entre um par de vértices, a medida de reciprocidade r denota a proporção de conexões mútuas entre pares pelo somatório de todas as conexões (recíprocas ou não). Desta forma, pode-se usar a notação:

$$r = \frac{L^{\leftrightarrow}}{L} \quad (2.1)$$

Para calcular a reciprocidade local de um determinado vértice v , pode-se considerar o universo L_v como o conjunto de conexões com este vértice.

Clusterização Local

BARABÁSI [5] estabelece o coeficiente de clusterização como uma forma de capturar o grau em que os vizinhos de um determinado vértice se conectam entre si. Esta métrica pode ser interpretada também como a propensão de triângulos se formarem entre um vértice e os seus vizinhos.

Para um vértice v , com grau d_v e número de arestas entre vizinhos de v igual à E_v , em um grafo não direcionado, o coeficiente de clusterização local pode ser definido como:

$$C_v = \frac{2E_v}{d_v(d_v - 1)} \quad (2.2)$$

Essa métrica para grafos direcionados necessita de pequenas adaptações para considerar a propensão de triângulos direcionados se formarem entre um vértice e os seus vizinhos.

2.1.2 Aprendizado de Máquina

Matriz de Correlação

A Matriz de Correlação é uma visualização que utiliza o coeficiente de correlação de Pearson BENESTY *et al.* [6]. Esta visualização é utilizada para mostrar como diferentes variáveis são relacionadas entre si, positivamente ou negativamente.

O coeficiente de Pearson pode ser definido como:

$$\text{Corr}(X, Y) = \hat{\rho}_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2.3)$$

Redes Neurais

Em EVSUKOFF [7] Redes Neurais Artificiais são caracterizadas como modelos computacionais inspirados no cérebro humano. Uma rede neural possui um conjunto de neurônios artificiais definidos como unidades e organizados em uma série de camadas, cada uma das quais conectadas em uma estrutura hierárquica.

Este conjunto de algoritmos e modelos é construído com o objetivo de reconhecer padrões, interpretando os dados de entrada. Este trabalho explora o uso de uma rede neural *feedforward*, dado que a informação flui somente em direção à camada de saída, ou seja, não existe uma recorrência dentro da rede.

Os neurônios de uma rede neural possuem a capacidade de ativar os neurônios da camada seguinte, transmitindo os dados e suas respectivas interpretações do modelo até a saída. Cada ligação possui um peso associado que indica o quanto a informação da unidade anterior influenciará a próxima camada. E, por fim, a última camada da rede possui como número de neurônios o mesmo número de classes que podem ser escolhidas pelo classificador, refletindo o resultado do modelo.

2.2 Trabalhos Relacionados

2.2.1 *The DARPA Twitter Bot Challenge*

SUBRAHMANNIAN *et al.* [8] apresenta uma proposta específica em identificar "*bots de influência*", destacando o seu uso em campanhas explícitas para influenciar a opinião nas mídias sociais. O artigo descreve um desafio de identificação de *bots* na rede social do Twitter e detalha os métodos usados pelas três equipes mais bem classificadas.

O artigo também comenta como trabalhos anteriores, referentes a este problema de classificação, frequentemente possuem dificuldade em apresentar uma boa acurácia, uma vez que há uma limitação do número de usuários que são conhecidos como *bots*.

As técnicas utilizadas pelas equipes com melhor classificação se concentraram no uso de técnicas de aprendizado de máquina. As características destacadas como de interesse pelo artigo incluem: análise da sintaxe e semântica dos textos publicados; comportamento temporal do usuário; características do perfil; e métricas da rede.

2.2.2 *Online Human-Bot Interactions: Detection, Estimation, and Characterization*

VAROL *et al.* [9] apresenta uma *framework* para detectar entidades autônomas no Twitter. O artigo comenta que o sistema desenvolvido aproveita mais de mil características extraídas de dados públicos e meta-dados, como: amizades, conteúdo e sentimento do *tweet*, padrões de rede e séries temporais de atividade.

Com relação ao uso de características da estrutura da rede, o artigo informa que possui um grande impacto para a caracterização. Diante disso, para o sistema foram utilizadas três tipos de redes: *retweet*, menção e coocorrência de *hashtag*. As redes construídas são todas ponderadas de acordo com a frequência de interações ou coocorrências, e para cada uma, foi utilizado um conjunto de características como distribuições de entrada e saída dos pesos, densidade e clusterização.

O artigo destaca nas conclusões que foi possível obter acurácias entre 0,85% e 0,95% dependendo do *dataset* de teste. Entretanto, o modelo de treino desenvolvido contou com 15 mil contas de usuários verificadas como *bots* e 16 mil contas verificadas como usuários humanos, tornando o problema como uma classificação de classes balanceadas.

Capítulo 3

Dataset e Modelagem da Rede

Este capítulo apresenta como funciona o Reddit, rede social analisada por este trabalho, destacando os dados utilizados e a modelagem da rede para a análise.

3.1 Reddit como Rede Social

O Reddit¹ se define, em tradução livre, como lar de milhares de comunidades, conversas intermináveis e conexões humanas autênticas [10]. A rede social, com média de usuários ativos mensais maior que 330 milhões, possui como foco as interações através de postagem, votação (positiva ou negativa) e comentário.

As postagens são realizadas pela comunidade da rede, que pode compartilhar conteúdo postando histórias, links, imagens e vídeos. Em seguida, a comunidade pode comentar estas postagens - é importante destacar que cada comentário também pode receber comentários e assim sucessivamente. Por fim, tanto as postagens como os comentários podem ser votados, com o intuito de destacar o conteúdo mais interessante para a rede.

3.2 *Dataset*

3.2.1 *Dataset* de Comentários

O *dataset* utilizado neste trabalho é disponibilizado através do acervo pushshift BAUMGARTNER [11]. Neste acervo, é possível encontrar diferentes conjuntos de dados da plataforma Reddit coletados e separados por dia ou mês.

Para este trabalho, utilizou-se os dados de comentários da rede social, Reddit, no mês de janeiro de 2018. O arquivo compreende os comentários de todos os usuários neste período, um total de 91.558.594, onde cada linha se apresenta como

¹<https://www.reddit.com>

um *JSON* de um total de 20 atributos. Nestes atributos estão as informações daquele comentário, desde identificadores até seu conteúdo do mesmo.

Dentre as informações disponibilizadas, as relevantes para a construção da rede em interesse são apresentadas na Figura 3.1. Esta redução do número de atributos está relacionada ao fato deste trabalho se concentrar na estrutura da rede de comentários e, em especial, no relacionamento de um usuário responder o comentário de outro usuário.

```
{
  "author": "yeee_bot",
  "id": "dnqik1a",
  "parent_id": "t1_dnqijyp"
}
```

Figura 3.1: Exemplo de dados utilizados para modelar a rede

As informações apresentadas indicam apenas que o usuário *yeee_bot* (campo "*author*") realizou um comentário de identificador *dnqik1a* (campo "*id*") em cima do comentário de identificador *dnqijyp* (campo "*parent_id*", onde a marcação "*t1*" é apenas um prefixo, que de acordo com a documentação do Reddit, indica que o identificador é de um comentário).

Observa-se que, dentre as informações para cada comentário de forma individual contidas no *dataset*, o usuário que recebeu o comentário, ou seja, autor do comentário de identificador *dnqijyp* (campo "*parent_id*"), é desconhecido. As consequências desta característica para a modelagem da rede serão abordadas na seção 3.3.

3.2.2 Lista de usuários *Bots*

O Reddit conta com a presença de diferentes tipos de *bots*, assim como em outras redes sociais ou plataformas de comunicação. Além de *bots* que buscam se apresentar como um humano, existem outros com cunho, em geral, mais informativo e que deixam claro que não são humanos.

Para utilizar a estrutura da rede construída a partir do *dataset* apresentado anteriormente na seção 3.2.1, mostra-se necessário identificar, dentre os usuários presentes, ao menos uma parcela dos usuários *bots*. Esta caracterização permite comparar o comportamento de usuários *bots* ou não e, posteriormente, avaliar as métricas extraídas.

Com o objetivo de compor uma lista de usuários *bots*, foram agrupados os resultados de diferentes fontes descritas mais à frente. Os nomes dos usuários obtidos

como resultado foram adicionados na forma minúscula em um *set* para evitar duplicatas e permitir uma comparação sem ambiguidades com os usuários do *dataset* de comentários.

Inicialmente, foram realizadas requisições a partir da API da plataforma pushshift BAUMGARTNER [11] para as postagens do usuário *BotBust*², este usuário *bot* apresenta outros usuários *bots* identificados e banidos dentro de tópicos da rede social. Os critérios para o banimento de um *bot* incluem realizar comentários sem serem convidados, não fornecer valor para a comunidade ou simplesmente não funcionar corretamente. Desta maneira, caso um usuário seja banido de um tópico por ser considerado *bot*, é assumido que ele é um *bot* para qualquer outra interação que tenha realizado na rede. Este processo coletou um total de 578 usuários identificados como *bot*.

Em seguida, foram encontradas postagens³⁴ apresentando listagens dos usuários *bots* mais populares da plataforma. Agrupando cada uma das listas aos usuários já coletados, permitiu-se aumentar o número de usuários *bots* conhecidos para 989 e posteriormente 1055.

Dentre os 1055 usuários, é esperado que apenas uma parcela seja reconhecida dentro da rede construída para este trabalho. O motivo para isto está no tamanho limitado do *dataset* de comentários, restrito ao período de um mês, e o possível desaparecimento destes usuários *bots* coletados ao longo do tempo. Além disso, espera-se que uma parcela dos usuários presentes na rede construída sejam *bots* e não estejam aqui identificados.

3.3 Modelagem da Rede

A rede modelada para este trabalho, a partir do *dataset* apresentado na Seção 3.2.1, apresenta-se como uma rede de comentários. Os vértices da rede são os usuários e as arestas entre os pares de usuários indicam que o usuário de origem da aresta respondeu em algum comentário do usuário onde a aresta incide. Ou seja, utilizando a notação $G = (V, E)$ para representar um grafo, onde V é conjunto de vértices e E é o conjunto de arestas que formam o grafo G , existe uma aresta direcionada $u \rightarrow v$ somente se o usuário u comentou em um comentário do usuário v . Além disso, cada aresta possui um peso associado w que indica o número de comentários realizados de um usuário para o outro.

As informações disponibilizadas para cada comentário de forma individual, e utilizadas para a construção desta rede, indicam o usuário autor do comentário,

²<https://www.reddit.com/user/BotBust>

³<https://www.reddit.com/r/autowikibot/wiki/redditbots>

⁴https://www.reddit.com/r/dataisbeautiful/comments/9mh3pn/oc_the_50_most_active_bots_on_reddit_based_on

o identificador do comentário realizado por este autor e o identificador do outro comentário, que sofreu o comentário do autor. Logo, o usuário que recebeu o comentário é desconhecido ao analisar somente as informações para uma única entrada do *dataset*.

Para representar esta situação, pode-se analisar as informações da Figura 3.1. Partindo das informações deste comentário, é possível concluir um arranjo da formação da rede, como o apresentado na Figura 3.2.

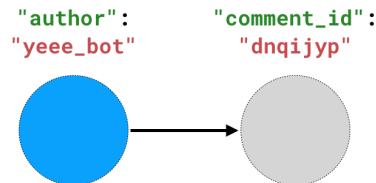


Figura 3.2: Arranjo da rede a partir do exemplo de dados de um único comentário

Como apontado na seção 3.2.1, o campo "`parent_id`" possui um prefixo "`t1_`" que indica que o identificador pertence a um comentário. Dessa forma, este prefixo é removido para permitir a associação com o identificador (campo "`id`" de outro comentário).

Como a rede deve ser concebida abstraindo um relacionamento entre usuários, constrói-se inicialmente uma série de pares de relacionamentos entre o usuário autor e o identificador do comentário que o autor está se dirigindo (que, por sua vez, pertence a algum usuário, ainda desconhecido). Ao mesmo tempo, também é construído um dicionário que indica para cada identificador de comentário quem é o usuário autor.

Desse modo, para construir as arestas da rede, percorre-se a série de pares de relacionamentos entre o usuário autor e o identificador do comentário que o autor está se dirigindo, substituindo o identificador do comentário pelo usuário correspondente - utilizando o dicionário construído. Caso o usuário correspondente não seja encontrado, esta aresta é descartada, visto que não poderia ser traçado o relacionamento entre dois usuários com os dados coletados.

Conclui-se que cada usuário que aparece na rede precisa necessariamente ter comentado ao menos uma vez, dado que somente desta forma o seu nome de usuário é conhecido. No entanto, isto não implica que o seu grau de saída na rede será obrigatoriamente maior que zero, na medida em que o usuário do comentário ao qual ele se dirige pode não aparecer na rede por não possuir o autor conhecido.

Seguindo esta construção, e ainda se baseando na Figura 3.2, pode-se concluir que, caso o comentário de identificador `dnqijyp` também esteja presente no *dataset*, será possível conhecer o seu autor e, assim, construir o relacionamento entre usuários apresentado na Figura 3.3.

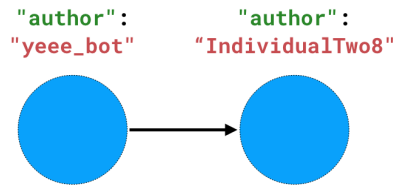


Figura 3.3: Relacionamento formado a partir do arranjo de um único comentário e as informações dos outros comentários

É importante ressaltar que a construção da rede, seguindo este processo, não será necessariamente uma linha, ou seja, cada usuário possuindo uma única aresta direcionada para outro usuário. Dentro do acervo de comentários, um mesmo usuário pode responder múltiplos comentários e receber respostas por outra gama de usuários em diferentes comentários de sua autoria. Também é importante notar que apesar de não estar representado na Figura 3.3, o fato de conhecer o autor *IndividualTwo8* indica que é conhecido uma resposta deste usuário para algum outro comentário.

Analisando os dados do *dataset*, foi possível perceber que parte dos autores dos comentários aparecem com o campo de nome de usuário igual à "[deleted]", o que indica que este usuário não existia mais na rede social no momento em que o *dataset* utilizado foi formado. Tratar estes usuários como parte da rede iria agrupar um conjunto de usuários diferentes como um único, deixando de representar a rede real e prejudicando as métricas de estudo. Logo, todos estes comentários foram ignorados para a construção da rede, restando um total de 85.006.711 comentários - aproximadamente 93% dos 91.558.594 comentários totais do *dataset* - realizados por 4.414.144 usuários diferentes.

Seguindo os procedimentos descritos acima para obter o usuário autor de cada comentário e construir os relacionamentos entre os pares, foi construída a rede induzida a partir dos comentários, apresentada na Tabela 3.1. Analisando as informações, podemos concluir o número de usuários presentes na rede, representado pelo número de vértices; o número de relacionamentos entre os usuários, representado pelo número de arestas; e o número total de comentários na rede, representado pela soma de peso das arestas.

Vértices	3.541.145
Arestas	38.452.098
Densidade	$3,07 \times 10^{-6}$
Soma do peso das arestas	48.754.283
Componentes Fracamente Conexas	21.872
Componentes Fortemente Conexas	969.314

Tabela 3.1: Avaliação preliminar da rede induzida a partir dos comentários

Analisando a Tabela 3.1 em comparação com dados anteriores, pode-se perceber uma redução no número de usuários e no número de comentários. Estas informações, em comparação com os dados do *dataset* após a remoção dos usuários que aparecem como deletados da rede, representam uma redução de aproximadamente 20% para o número de usuários e ainda mais significativa para o número de comentários, aproximadamente 43%. Esta redução está diretamente relacionada à limitação, imposta pelo *dataset*, de não conhecer o usuário autor do comentário que se foi dirigido.

Capítulo 4

Caracterização Estrutural de Usuários e Bots

Este capítulo apresenta e comenta a proposta deste trabalho, avaliando os dados extraídos a partir da Maior Componente Fortemente Conexa e da Maior Componente Fracamente Conexa construídas.

4.1 Componentes da Rede Modelada

Para a rede modelada neste problema são esperados diferentes grupos de usuários que se apresentam em diferentes componentes do grafo, de acordo com os seus relacionamentos. Seria possível imaginar, por exemplo, que determinados usuários só comentem em questões de um determinado assunto de interesse específico. Por outro lado, podem existir usuários que comentem em múltiplos assuntos, criando caminhos entre estes diferentes grupos.

Com o objetivo de saber mais sobre a organização da rede induzida, e principalmente a presença dos *bots* nela, pode-se observar a distribuição das componentes fortemente conexas e fracamente conexas nas Tabelas 4.1 e 4.2, respectivamente. Cabe destacar que foi encontrado um total de 207 usuários *bots* na rede como um todo comparando os nomes de usuários coletados.

Tamanho da Componente	Componentes Fortemente Conexas	Total de Bots
1	955.182	43
2	13.352	1
3	661	0
4	77	0
5	25	0
6	3	0
7	7	0
9	3	0
14	1	0
28	1	0
268	1	0
2.556.439	1	163
Total de Bots na Rede		207

Tabela 4.1: Avaliação das Componentes Fortemente Conexas da Rede

Analisando a Tabela 4.1, pode-se observar que existem 43 *bots* em componentes fortemente conexas de tamanho 1, ou seja, 43 componente formadas somente por um único *bot*. Verifica-se, também, um único *bot* em uma das componentes de tamanho 2. E, por fim, todos os *bots* restantes estão presentes na única maior componente fortemente conexa.

Tamanho da Componente	Componentes Fracamente Conexas	Total de Bots
1	3.605	1
2	16.659	0
3	1.290	0
4	230	0
5	52	0
6	14	0
7	9	0
8	1	0
9	3	0
10	1	0
11	2	0
13	1	0
14	1	0
17	1	0
18	1	0
22	1	0
3.498.874	1	206
Total de Bots na Rede		207

Tabela 4.2: Avaliação das Componentes Fracamente Conexas da Rede

Analisando a Tabela 4.2, pode-se observar que existe um único *bot* em componentes fracamente conexas de tamanho 1, ou seja, uma única componente formada somente por um único *bot*. Por fim, observa-se que todos os *bots* restantes estão presentes na única maior componente fracamente conexa.

Dessa forma, para este trabalho, serão avaliadas as métricas com base na maior componente fortemente conexa e na maior componente fracamente conexa. Estas possuem o maior número de usuários identificados como *bots* e concentram, devido ao seu tamanho, o maior número de comentários trocados.

4.2 Métricas

Esta seção avalia as métricas para a Maior Componente Fortemente Conexas da rede. Após a análise das métricas, verificou-se que a caracterização da Maior Componente Fracamente Conexas se apresenta muito semelhante a anterior e, por isto, está disponível no Apêndice A.1 para referência.

Nas Figuras 4.1, 4.2 e 4.3, a seguir expostas, são apresentadas algumas das principais características extraídas da Maior Componente Fortemente Conexas. As

marcações dentro dos gráficos como *bot* são referentes aos dados coletados de acordo com a seção 3.2.2.

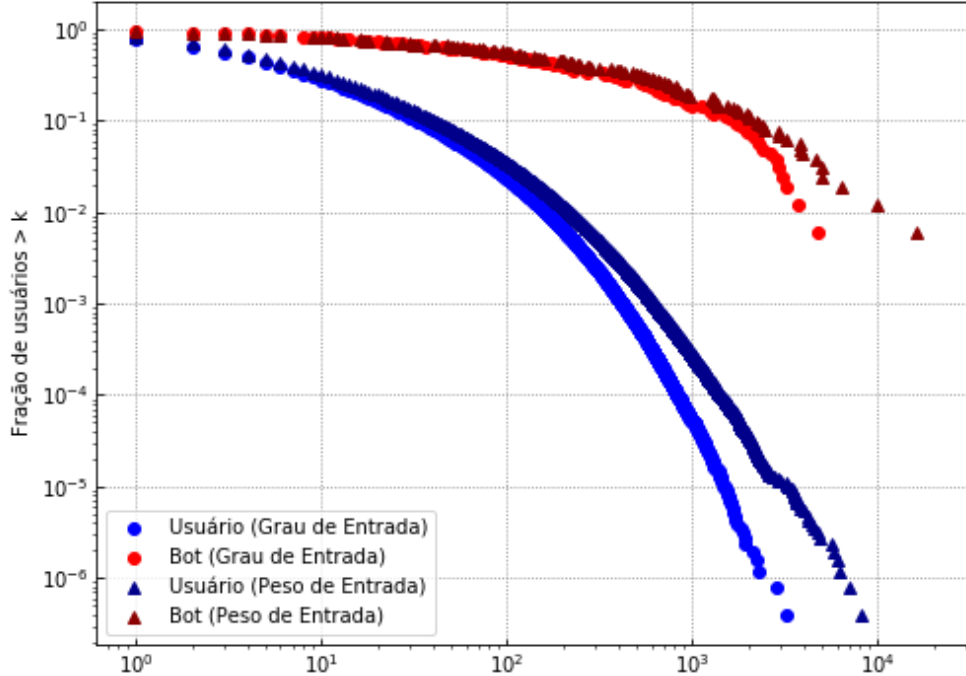


Figura 4.1: Função de distribuição cumulativa complementar (CCDF) do grau de entrada e peso de entrada dos vértices na Maior Componente Fortemente Conexa

Observando o gráfico da Figura 4.1, pode-se perceber que a diferença entre a curva de grau e de peso, para um mesmo tipo de usuário, começa a aumentar para valores em ordem menor para os usuários do que para os *bots*. Este comportamento pode indicar a presença mais provável de uma discussão entre usuários, dado que a curva de peso se afasta quando uma aresta direcionada entre um par de usuários está codificando a presença de mais comentários deste usuário para o outro.

Além disso, é possível perceber que os usuários *bots* recebem proporcionalmente mais comentários direcionados a eles do que os usuários comuns. Apesar de parecer, em um primeiro momento, um comportamento estranho, esta observação pode caracterizar situações onde o usuário emite um comentário relevante e/ou controverso, causando uma série de respostas apoiando ou contestando aquela informação. Desse modo, faz sentido esperar um grau maior para *bots* que desempenham este comportamento.

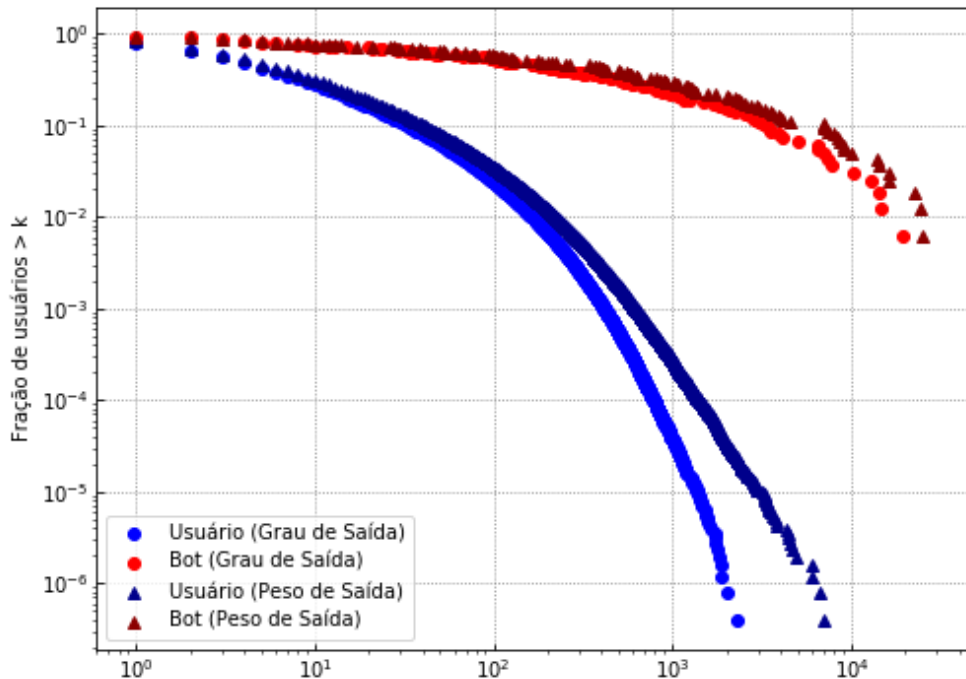


Figura 4.2: Função de distribuição cumulativa complementar (CCDF) do grau de saída e peso de saída dos vértices na Maior Componente Fortemente Conexa

Observando o gráfico da Figura 4.2, pode-se perceber que a curva do grau de saída e do peso de saída começam a se distanciar para os usuários comuns após um certo valor, enquanto permanecem praticamente juntas para toda a curva de usuários bots.

Como cada valor unitário de peso indica um comentário, o comportamento normal é que a curva do peso de saída apresentasse valores próximos ou maiores que o grau de saída. Entretanto, é notório que a curva de peso para os *bots* se distancia de forma limitada da curva de grau, o que indica a maioria de arestas com peso unitário. Em outras palavras, indica que há pouco envio de comentários como resposta para um mesmo usuário.

Dessa forma, os usuários *bots* parecem comentar de forma mais indiscriminada, com peso das arestas tendendo a ser mais próximo de 1, enquanto usuários acabam comentando no mesmo usuário mais de uma vez, podendo indicar, como na Figura 4.1, uma discussão entre os pares de usuários.

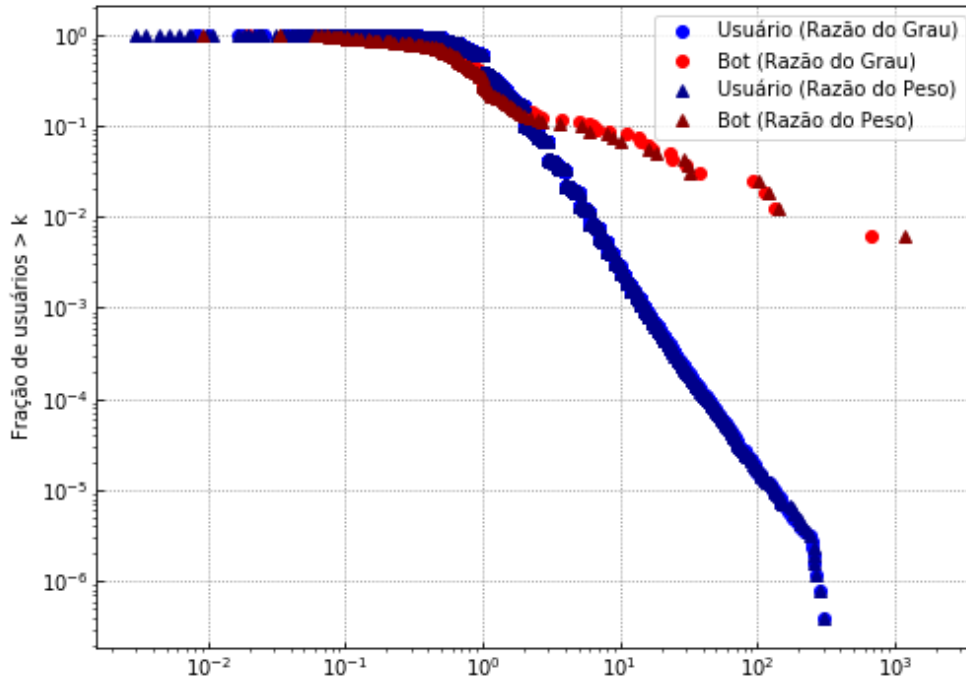


Figura 4.3: Função de distribuição cumulativa complementar (CCDF) da razão do grau e razão do peso dos vértices na Maior Componente Fortemente Conexa

Observando o gráfico da Figura 4.3, pode-se perceber que para valores abaixo de 1 ou pouco maiores (o grau ou peso de entrada é menor ou próximo ao de saída) existe pouca distinção entre usuários comuns e *bots*. Contudo, a partir de um certo valor é possível perceber uma proporção maior de usuários *bots* com o grau ou peso de entrada maiores que os de saída, quando comparado com usuários comuns.

Este comportamento indica que um mesmo usuário *bot* acaba, por vezes, recebendo mais comentários do que realizando. Como elucidado anteriormente, a propagação de certas informações pode acarretar no recebimento de múltiplas respostas concordando ou contestando aquela informação. Ademais, usuários *bots* parecem não se envolver em discussões diretas sobre seus comentários, como considerado a partir da Figura 4.2. Desta forma, o comportamento da curva se mostra justificável pelo comportamento dos usuários.

Diante das métricas apresentadas anteriormente nas Figuras 4.1, 4.2 e 4.3, pode-se conferir os valores mínimo, máximo, médio e a mediana para cada uma delas na Tabela 4.3.

	Mínimo	Máximo	Média	Mediana
Grau de Entrada	1	13.485	14,45	4
Soma dos Pesos de Entrada	1	20.660	18,45	5
Grau de Saída	1	19.740	14,45	4
Soma dos Pesos de Saída	1	29.930	18,45	5
Razão de Grau (entrada / saída)	$8,16 \times 10^{-3}$	1.124,00	1,28	1,00
Razão de Peso (entrada / saída)	$3,00 \times 10^{-3}$	2.005,00	1,29	1,00

Tabela 4.3: Primeiras métricas para a Maior Componente Fortemente Conexa da Rede

Nas Figuras 4.4 e 4.5, a seguir expostas são apresentados gráficos de dispersão entre dois pares de métricas, buscando visualizar a distribuição da rede e características mais comuns para usuários *bots*.

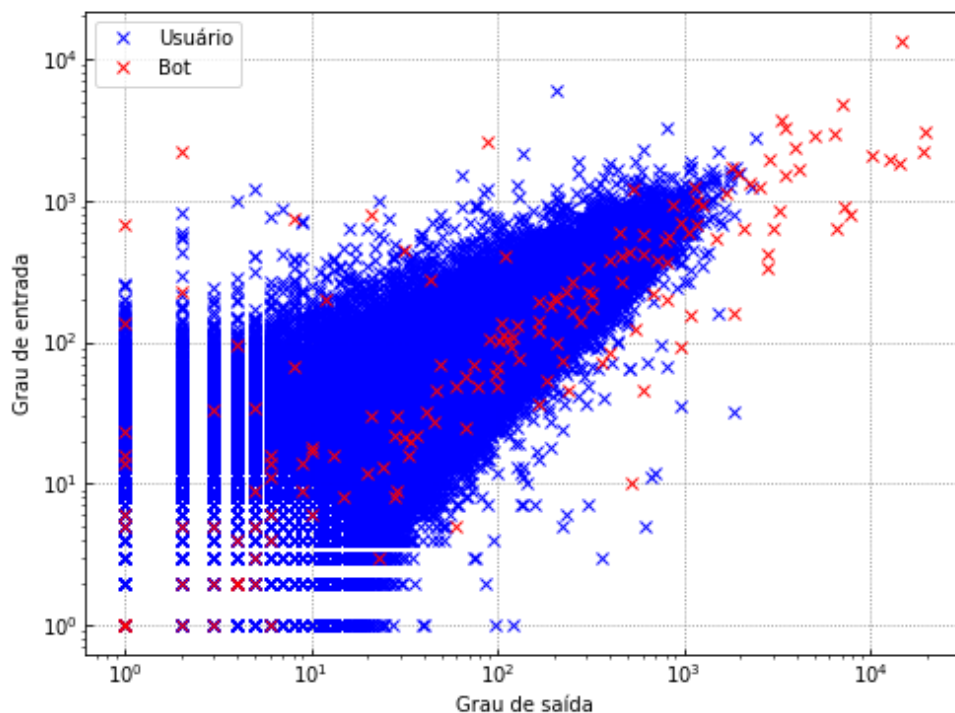


Figura 4.4: Gráfico de dispersão entre o grau de saída e o grau de entrada dos vértices na Maior Componente Fortemente Conexa

Observando o gráfico da Figura 4.4, pode-se notar que os usuários *bots* estão mais concentrados na diagonal principal entre o grau de entrada e o grau de saída, isto é, apresentando valores mais próximos um do outro, mesmo para valores menores de grau. Já para os usuários comuns, este comportamento aparece mais disperso em

valores menores, com um grupo de usuários apresentando valores mais altos de grau de entrada ou saída (se afastando da diagonal), passando a apresentar valores mais próximos um do outro com o aumento dos valores de ambos os graus.

Além disso, é possível verificar que para valores igualmente mais elevados de grau de entrada e grau de saída só existe a presença de usuários *bots*.

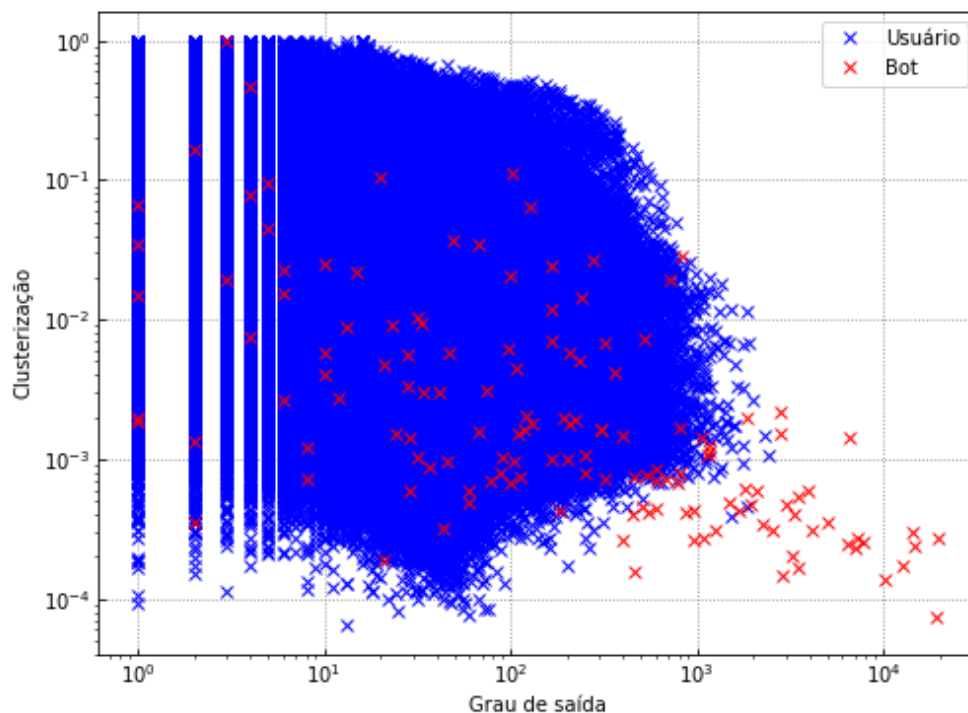


Figura 4.5: Gráfico de dispersão entre o grau de saída e o coeficiente de clusterização dos vértices na Maior Componente Fortemente Conexa

A Figura 4.5, para valores de grau de saída inferiores à 10^3 , possui um grande embaralhamento entre usuários comuns e usuários *bots*. Apesar disso, para valores de grau de saída acima deste ponto, permite observar que para coeficientes de clusterização muito baixos existe uma concentração quase exclusiva de usuários *bots*. Dentre os poucos usuários que aparecem como comuns nesta região, existe uma grande probabilidade de serem, na verdade, *bots* não classificados dentre os dados deste trabalho.

Esse padrão apresentado parece comum para um usuário *bot*, visto que este deve comentar de forma mais indiscriminada, como apontado na avaliação do gráfico na Figura 4.2, diminuindo a probabilidade de comentários ocorrerem entre os usuários respondidos pelo *bot*. Além disso, como o funcionamento do usuário *bot* acontece de forma autônoma, faz-se possível um alto grau de saída, ou seja, comentários para muitos usuários diferentes num dado período.

Outras métricas de interesse para a componente estão apresentadas na Tabela 4.4

	Mínimo	Máximo	Média	Mediana
Coefficiente de Clusterização	0	1	$4,04 \times 10^{-2}$	$9,07 \times 10^{-4}$
Reciprocidade Local	0	1	$3,48 \times 10^{-1}$	$2,92 \times 10^{-1}$
Reciprocidade Local (usuários)	0	1	$3,48 \times 10^{-1}$	$2,92 \times 10^{-1}$
Reciprocidade Local (bots)	0	1	$2,03 \times 10^{-1}$	$1,66 \times 10^{-1}$
Reciprocidade da Rede	-	-	$2,78 \times 10^{-1}$	-

Tabela 4.4: Outras métricas para a Maior Componente Fortemente Conexa da Rede

Analisando a Tabela 4.4, é interessante apontar os valores relativamente altos para a reciprocidade local. A média deste valor indica que 35% dos comentários realizados acabam sendo respondidos de volta, como uma discussão ou debate entre usuários. Obviamente, a métrica não garante que os usuários possuem um relacionamento recíproco diante dos mesmos comentários, mas devido ao tamanho da rede é mais esperado que isto ocorra em um mesmo tópico de discussão. Esta métrica é consideravelmente menor quando observada somente para usuários *bots*, reforçando a ideia de que estes usuários não se correspondem muito com outros usuários e mantém uma comunicação mais indiscriminada.

Os valores encontrados para o coeficiente de clusterização indicam que a rede não possui uma alta propensão para triângulos se formarem entre um vértice e os seus vizinhos. Em outras palavras, ocorrência de três comentários entre usuários que forme um ciclo (triângulo).

Por fim, vale ressaltar que, existem outras métricas estruturais da rede que poderiam ser exploradas nesta seção, como, por exemplo, o PageRank para determinar a importância de um vértice da rede.

Capítulo 5

Classificação de Perfis Falsos

Este capítulo aborda as técnicas de Aprendizado de Máquina utilizadas para demonstrar como as métricas estruturais da rede podem permitir uma classificação dos usuários da mesma.

A partir das métricas extraídas na Seção 4.2, é viável imaginar como estas características podem ser empregadas para classificação de um usuário qualquer, seguindo algum processo dentro da computação. Desta forma, é possível conceber algum modelo classificador de aprendizado de máquina (ou vários), que lide com as características de cada usuário e os classifique em uma das possíveis classes: "*bot*" ou "*não bot*".

Este trabalho explora um modelo e observa o seu resultado, verificando se consegue gerar uma predição razoável para o problema, indicando o potencial no emprego deste paradigma de rede e conceitos de Redes Complexas.

O problema deste trabalho trata de classes desbalanceadas, onde, como observado nas Tabelas 4.1 e 4.2, referentes às componentes utilizadas para avaliação, apenas uma pequena parcela de usuários são conhecidos como *bot*. Caso esta característica não seja tratada, o modelo tende a favorecer a classe que contém o maior número de amostras, neste caso, apontando todos os usuários como "*não bot*".

5.1 Avaliação e Tratamento das Métricas para o Modelo

Dentro desta ideia de conceber um modelo classificador que lide com as características de cada usuário, vale analisar a Figura 5.1, com a matriz de correlação para as características dos vértices na Maior Componente Fortemente Conexa da rede. Após a análise, verificou-se que o comportamento da Maior Componente Fortemente Conexa se apresenta muito semelhante a anterior e, por isto, a matriz está disponível no Apêndice B.1 para referência.

Como esperado, é possível reparar uma forte correlação entre métricas derivadas de outras, como, por exemplo, soma dos graus de entrada e saída com as métricas individuais. Mas também é possível reparar correlações entre métricas opostas como o grau de entrada e o grau de saída, como foi observado também em análise das métricas a partir dos gráficos da função de distribuição cumulativa complementar (CCDF) na seção 4.2.

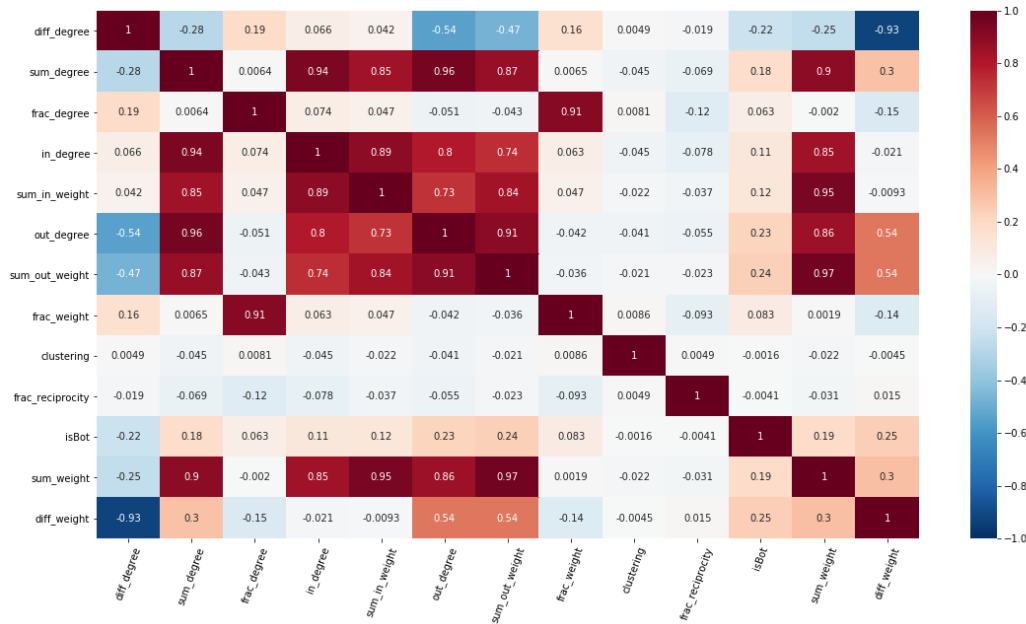


Figura 5.1: Matriz de Correlação para as características dos vértices na Maior Componente Fortemente Conexo

Analisando cuidadosamente a matriz de correlação é possível remover parte das características de alta correlação sem prejudicar um modelo classificador adotado. Este processo pode simplificar o modelo, por reduzir a dimensionalidade do problema, e melhorar a predição. Ressalta-se, todavia, que este trabalho não se preocupa em verificar o melhor conjunto de características para aplicação dentro de um modelo classificador, visto que este processo requer uma avaliação dentre diferentes combinações para evitar a perda de características importantes para a classe de interesse.

Por outro lado, como as métricas de avaliação possuem diferentes variações entre os valores mínimos e máximos, todas as métricas que não eram classificadas entre zero e um foram normalizadas. Este processo simples pode ser capaz de determinar se um modelo neural é capaz ou não de aprender com o passar de um número determinado de épocas. Quando as variações de valores são muito desiguais, isto pode fazer com que os gradientes oscilem muito até encontrar o caminho para o mínimo global ou local. Em contrapartida, as métricas normalizadas permitem que o gradiente possa convergir mais rapidamente.

5.2 Aplicação e Avaliação de uma Rede Neural

Para avaliar um classificador de rede neural, adotou-se um modelo relativamente simples. Como explicitado na Seção 2.1.2 o modelo adotado de rede neural foi o de *feedforward*, e contou com apenas duas camadas, sendo 62 neurônios presentes na camada de entrada e 2 neurônios de saída.

O modelo classificador foi utilizado com 12 variáveis de entrada (métricas extraídas a partir da estrutura da rede) e retorna a partir da sua camada de saída valores probabilísticos da classificação, onde os valores para as classes "bot" e "não bot" são complementares. Foram realizadas 200 épocas de treinamento, ou seja, o algoritmo estuda a amostra completa de dados um total de 200 vezes. Além disso, foram atribuídos pesos às classes de acordo com o desbalanceamento do problema.

Os dados de entrada foram divididos em 5 *folds* iguais, mantendo a proporção de cada classe "bot" e "não bot". Três destes *folds* foram agrupados e utilizados como conjunto de treinamento. Um *fold* foi utilizado para o conjunto de validação, utilizado ao final de cada época para testar o desempenho parcial do modelo. Após o final do treinamento da rede neural, o último *fold* foi utilizado como conjunto de teste e os valores da predição deste conjunto serão utilizados para avaliar a classificação.

Para avaliar o desempenho da classificação obtida, primeiramente, é importante definir algumas métricas que serão utilizadas:

- *Recall* é a fração de amostras classificadas corretamente de uma classe para todas as amostras da mesma classe. Desta forma, pode-se usar a notação:

$$Recall(C_1) = \frac{VP}{VP + FN} \quad (5.1)$$

- *Precisão* é a fração de amostras classificadas corretamente de uma classe para todas as amostras definidas pelo modelo como pertencentes da classe. Desta forma, pode-se usar a notação:

$$Precisão(C_1) = \frac{VP}{VP + FP} \quad (5.2)$$

- *F1 score* , ou simplesmente F1, se apresenta como uma combinação entre a precisão e o *recall* . A medida pode ser definida como:

$$F_1 = 2 \times \frac{Precisão \times Recall}{Precisão + Recall} \quad (5.3)$$

A Figura 5.2, apresentada a seguir, contém um gráfico com as métricas de *recall* para as duas classes, "bot" e "não bot", além de precisão e *F1 score* somente para a classe "bot". Neste gráfico, o eixo *x* representa o valor a partir do qual uma probabilidade do classificador para uma dada amostra é considerada como uma classificação em "bot".

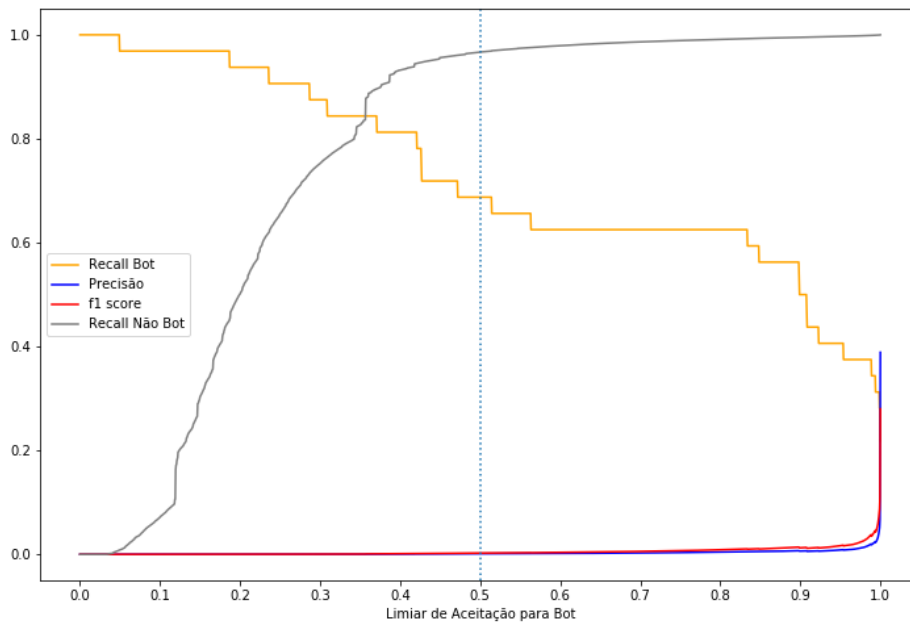


Figura 5.2: Avaliação do modelo de Rede Neural para Recall, Precisão e F1 Score

Avaliando as curvas das métricas, é possível perceber valores próximos de zero para as métricas de precisão e F1. Na verdade, estas métricas não são uma boa forma de avaliação para um problema desbalanceado como este. Isto pode ser verificado ao analisar a métrica de precisão, onde o valor máximo para o numerador, ou seja, verdadeiro positivo, será sempre muito pequeno devido a classe *bot* minoritária. Logo, qualquer valor de falso positivo (usuários comuns classificados como bots) irá prejudicar a métrica por estar no denominador. Fora isso, a classificação inicial do problema para *não bot* provavelmente não é totalmente correta, podendo existir usuários considerados comuns que são, na verdade, *bots*.

Para avaliar melhor as taxas de *recall* foi adotando um limiar de 0,5 para a aceitação de *bots*, ou seja, amostras avaliadas pelo classificador com probabilidades iguais ou acima deste valor classificam o usuário como *bot*. Para este valor de limiar, foi possível obter um *recall* de 0,69 para a classe *bot* e *recall* 0,97 para a classe *não bot*.

Apesar de um acerto maior para a classe *não bot*, este comportamento é importante num contexto de uma aplicação que pode, por exemplo, suspender tempora-

riamente usuários identificados como *bot*. Mesmo com interesse na classificação de usuários *bots* não seria interessante possuir alto erro para a classe oposta.

A Figura 5.3, a seguir exposta, apresenta a curva *ROC*. Em EVSUKOFF [7] é destacado que para problemas desbalanceados, a decisão pelo mínimo erro da classificação favorece a classe majoritária. Dessa forma, a curva *ROC* busca avaliar o desempenho relacionado com o quanto o classificador acerta da classe positiva e o quanto deixa de errar da classe negativa [7].

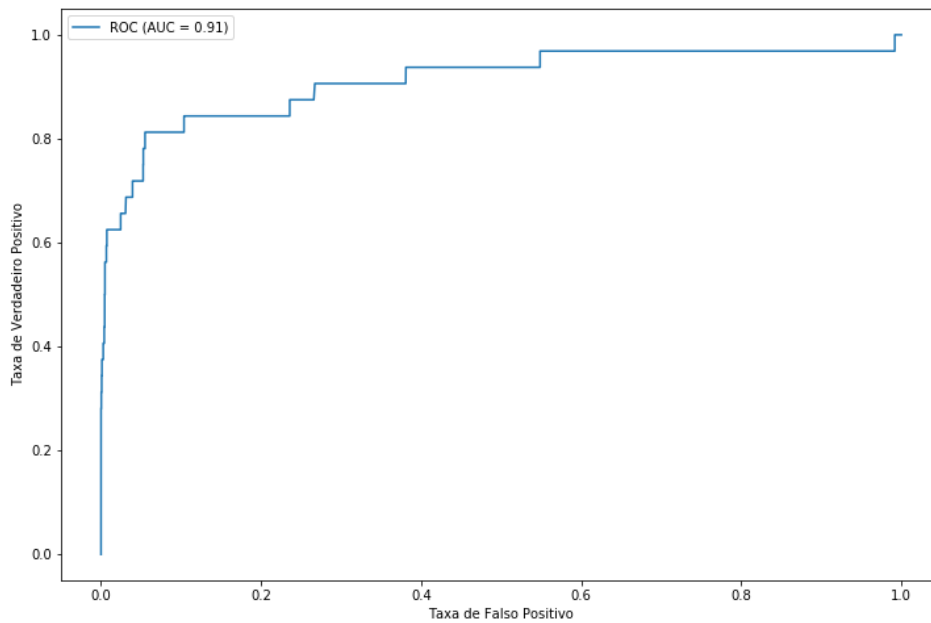


Figura 5.3: Avaliação do modelo de Rede Neural no Espaço ROC

Para avaliar a curva da Figura 5.3 é utilizada a métrica *AUC* que indica a área sob a curva *ROC*. Esta métrica vale para problemas de duas classes [7], como é o caso deste problema, possuindo valores entre zero e um.

Desta forma, podemos concluir que o valor de 0,91 encontrado caracteriza uma ótima métrica visto o problema desbalanceado entre duas classes, com interesse na classe minoritária *bot*.

Capítulo 6

Conclusão

Este capítulo discorre sobre as conclusões e considerações finais do trabalho.

O presente trabalho buscou apresentar uma abordagem utilizando o paradigma de rede para a identificação de usuários *bots* em uma rede social. Foram debatidos alguns conceitos de Redes Complexas, destacando a relevância deste campo para o problema em estudo. Além disso, demonstrou como a estrutura da rede expõe significados sobre os participantes da mesma e como os relacionamentos podem evidenciar estes significados.

O objetivo de classificar os usuários como *bots* ou *não bots* se mostrou viável, mesmo para um modelo classificador relativamente simples e sem muitas otimizações. Apesar das limitações dos dados utilizados para este trabalho e, conseqüentemente, para a construção da rede modelada - onde nem todas as arestas estão presentes - a classificação de usuários, mesmo se tratando de um problema desbalanceado, atingiu taxas significativas de acerto para a classe de interesse e também da classe onde o usuário não é *bot*.

Os resultados obtidas para o classificador foram importantes para a forma como o problema se apresenta. Além disso, a métrica estudada para avaliar o classificador de forma mais designada para este problema, *AUC*, apresentou resultado extremamente satisfatório.

Por fim, este projeto abrange a ideia do paradigma de rede e o campo de Ciência das Redes sendo utilizados como ferramentas para a classificação de usuários *bots* em redes sociais online. Desta forma, evita-se a subjetividade inerente ao conteúdo e utiliza-se aspectos mais robustos, como a troca de mensagens entre pessoas. Os resultados obtidos deixam claro que as métricas estruturais que podem ser extraídas de uma rede conseguem exemplificar fenômenos e comportamentos dos pares de objetos codificados.

6.1 Trabalhos Futuros

Esta seção apresenta pontos que poderiam ser tratados em trabalhos futuros para expandir ou melhorar a proposta apresentada.

- Avaliar a Proposta em outras Redes Sociais

Cada rede social se apresenta de uma forma diferente, possuindo foco em interações diferentes e comportamento dos usuários potencialmente diferentes. Dessa forma, a análise das mesmas métricas deste trabalho em outras redes sociais pode demonstrar novos fenômenos ou padrões alterados.

- Explorar Estrutura Local do Vértice

Existem outras métricas que poderiam ser extraídas da estrutura para os vértices. Algumas ideias, como calcular os vizinhos em comum, isto é, para um dado par de usuários u e v , o conjunto de usuários que recebem comentários a partir de u e também de v , poderiam trazer resultados interessantes que auxiliariam para a identificação de usuários *bot*.

- Utilizar Diferentes Redes que Codifiquem outros Relacionamentos

As métricas estruturais calculadas para este trabalho poderiam ser estabelecidas para um outro conjunto de redes com outros relacionamentos. Redes como de amizade, curtida, dentre outras, modeladas com o mesmo conjunto de usuários iriam permitir uma avaliação dentre diferentes codificações de relacionamentos. O uso combinado destas redes aparece abordado na literatura e pode aprimorar os resultados obtidos.

- Avaliar outros Modelos Classificadores ou Aperfeiçoar os Desenvolvidos

Este trabalho avaliou somente um modelo classificador usando redes neurais. As mesmas métricas calculadas poderiam ser utilizadas em outros modelos classificadores, utilizando diferentes técnicas. Desta forma, os novos modelos poderiam ampliar os resultados e deixar mais explícito o uso de Ciência das Redes para resolver o problema.

Referências Bibliográficas

- [1] “How Trump Consultants Exploited the Facebook Data of Millions”. . Disponível em: <https://nyti.ms/2GB9dK4>. [Online; accessed 12-march-2019].
- [2] FERRARA, E., VAROL, O., DAVIS, C., et al. “The Rise of Social Bots”, *Commun. ACM*, v. 59, n. 7, pp. 96–104, jun. 2016. ISSN: 0001-0782. doi: 10.1145/2818717. Disponível em: <http://doi.acm.org/10.1145/2818717>.
- [3] KLEINBERG, J., TARDOS, E. *Algorithm design*. Pearson Education India, 2006.
- [4] GARLASCHELLI, D., LOFFREDO, M. I. “Patterns of Link Reciprocity in Directed Networks”, *Phys. Rev. Lett.*, v. 93, pp. 268701, Dec 2004. doi: 10.1103/PhysRevLett.93.268701. Disponível em: <https://link.aps.org/doi/10.1103/PhysRevLett.93.268701>.
- [5] BARABÁSI, A.-L. “Network science”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 371, n. 1987, pp. 20120375, 2013.
- [6] BENESTY, J., CHEN, J., HUANG, Y., et al. “Pearson correlation coefficient”. In: *Noise reduction in speech processing*, Springer, pp. 1–4, 2009.
- [7] EVSUKOFF, A. “Ensinando Máquinas”. . Material disponibilizado na aula de Inteligência Computacional(COC361) do Centro de Tecnologia da UFRJ.
- [8] SUBRAHMANYAN, V., AZARIA, A., DURST, S., et al. “The DARPA Twitter Bot Challenge”, *Computer*, v. 49, n. 6, pp. 38–46, Jun 2016. ISSN: 0018-9162. doi: 10.1109/mc.2016.183. Disponível em: <http://dx.doi.org/10.1109/MC.2016.183>.
- [9] VAROL, O., FERRARA, E., DAVIS, C., et al. “Online Human-Bot Interactions: Detection, Estimation, and Characterization”. , 2017. Disponível em: <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817>.

- [10] “The Conversation Starts Here”. . <https://www.redditinc.com/>. Disponível em: <https://www.redditinc.com/>. [Online; accessed 15-march-2019].
- [11] BAUMGARTNER, J. “Reddit Comments dataset”. . <https://files.pushshift.io/reddit/comments/>, 2018. Disponível em: <https://files.pushshift.io/reddit/comments/>. [Online; accessed 10-march-2019].

Apêndice A

Métricas Complementares

Esta seção apresenta outras métricas das comentadas na seção 4.2, complementando a análise para a rede.

A.1 Maior Componente Fracamente Conexa

A seguir, serão apresentadas as métricas para a Maior Componente Fracamente Conexa da rede. As métricas equivalentes para a Fortemente Conexa podem ser encontradas na seção 4.2.

Nas Figuras A.1, A.2 e A.3, a seguir expostas, são apresentadas algumas das principais características extraídas da Maior Componente Fracamente Conexa. As marcações dentro dos gráficos como *bot* são referentes aos dados coletados de acordo com a seção 3.2.2.

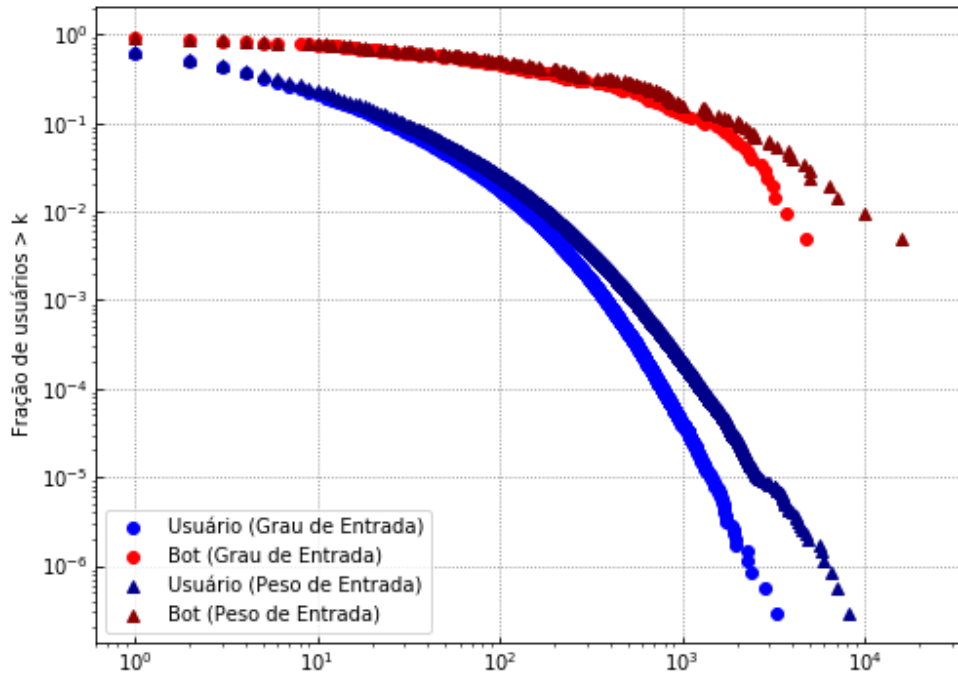


Figura A.1: Função de distribuição cumulativa complementar (CCDF) do grau de entrada e peso de entrada dos vértices na Maior Componente Fracamente Conexa

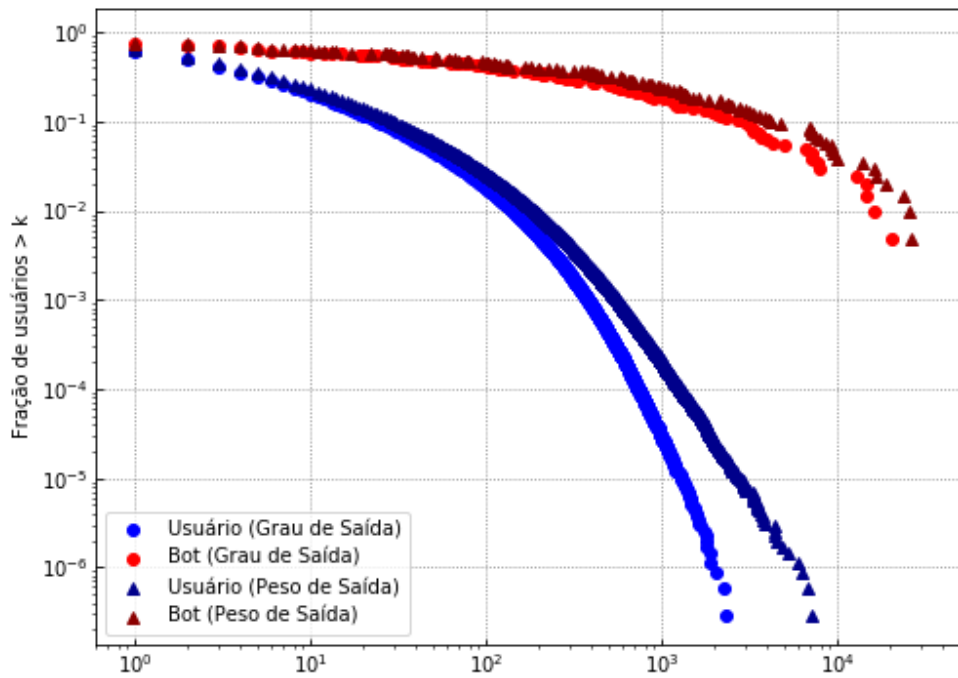


Figura A.2: Função de distribuição cumulativa complementar (CCDF) do grau de saída e peso de saída dos vértices na Maior Componente Fracamente Conexa

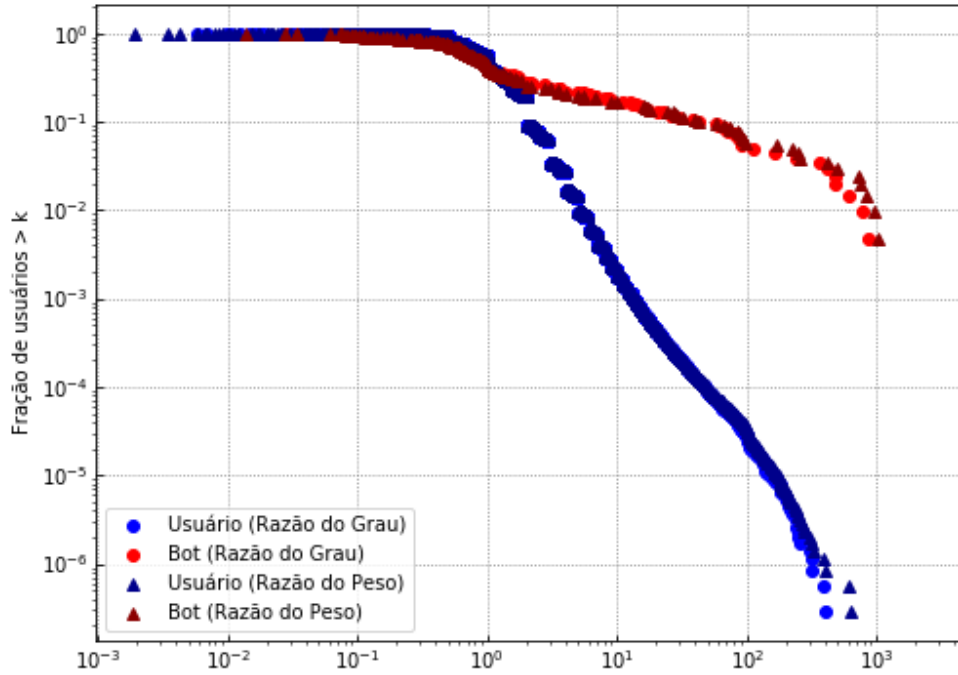


Figura A.3: Função de distribuição cumulativa complementar (CCDF) da razão do grau e razão do peso dos vértices na Maior Componente Fracamente Conexa

Diante das métricas apresentadas anteriormente nas Figuras A.1, A.2 e A.3, pode-se conferir os valores mínimo, máximo, médio e a mediana para cada uma delas na Tabela A.1.

	Mínimo	Máximo	Média	Mediana
Grau de Entrada	0	14.676	10,98	3
Soma dos Pesos de Entrada	0	21.937	13,92	3
Grau de Saída	0	21.001	10,98	2
Soma dos Pesos de Saída	0	32.355	13,92	3
Razão de Grau (entrada / saída)	$5,78 \times 10^{-3}$	1.818,00	1,26	1,00
Razão de Peso (entrada / saída)	$1,93 \times 10^{-3}$	2.446,00	1,27	1,00

Tabela A.1: Primeiras métricas para a Maior Componente Fracamente Conexa da Rede

Nas Figuras A.4 e A.5, a seguir demonstradas, são apresentados gráficos de dispersão entre dois pares de métricas, buscando visualizar a distribuição da rede e características mais comuns para usuários *bots*.

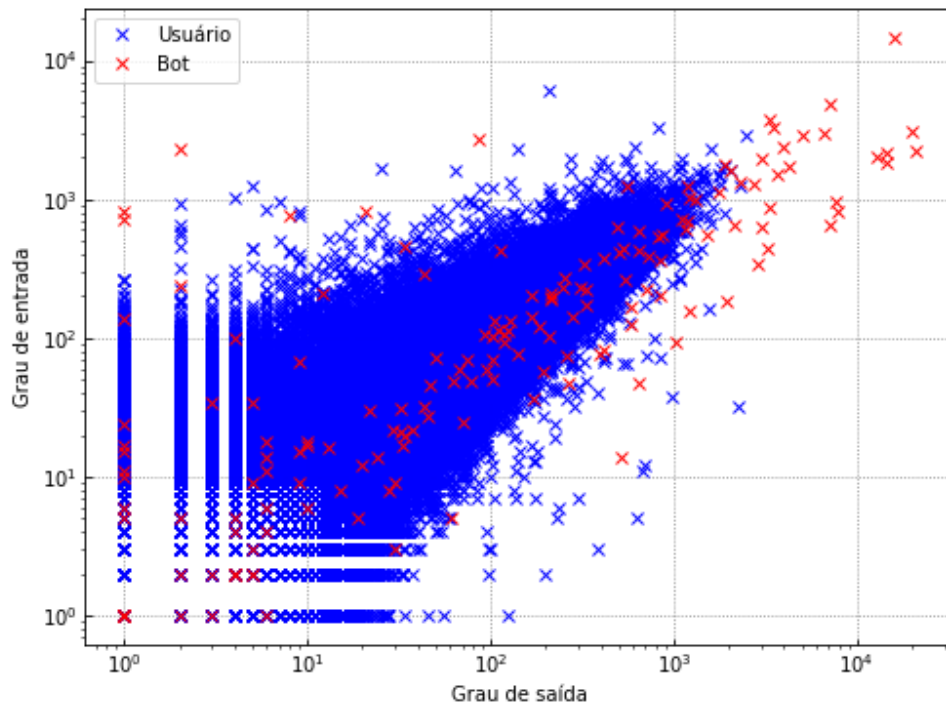


Figura A.4: Gráfico de dispersão entre o grau de saída e o grau de entrada dos vértices na Maior Componente Fracamente Conexa

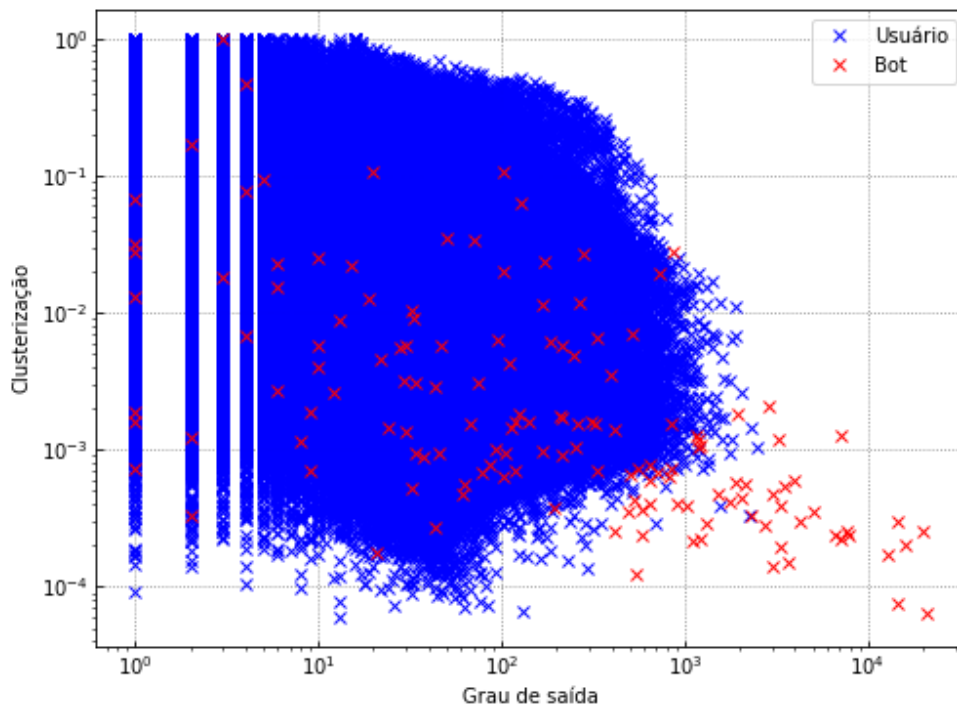


Figura A.5: Gráfico de dispersão entre o grau de saída e o coeficiente de clusterização dos vértices na Maior Componente Fracamente Conexa

Outras métricas de interesse para a componente estão apresentadas na Tabela A.2

	Mínimo	Máximo	Média	Mediana
Coefficiente de Clusterização	0	1	$3,26 \times 10^{-2}$	0,00
Reciprocidade Local	0	1	$2,52 \times 10^{-1}$	$2,00 \times 10^{-1}$
Reciprocidade Local (usuários)	0	1	$2,52 \times 10^{-1}$	$2,00 \times 10^{-1}$
Reciprocidade Local (bots)	0	1	$1,56 \times 10^{-1}$	$9,31 \times 10^{-2}$
Reciprocidade da Rede	-	-	$2,65 \times 10^{-1}$	-

Tabela A.2: Outras métricas para a Maior Componente Fracamente Conexa da Rede

Apêndice B

Avaliação Complementar para o Modelo

Esta seção apresenta outras avaliações das comentadas na seção 5.1, complementando a análise para a rede.

B.1 Maior Componente Fracamente Conexa

A seguir é apresentada a avaliação da matriz de correlação para a Maior Componente Fracamente Conexa da rede. As avaliações equivalentes de forma mais detalhada para a Fortemente Conexa podem ser encontradas na seção 5.1.

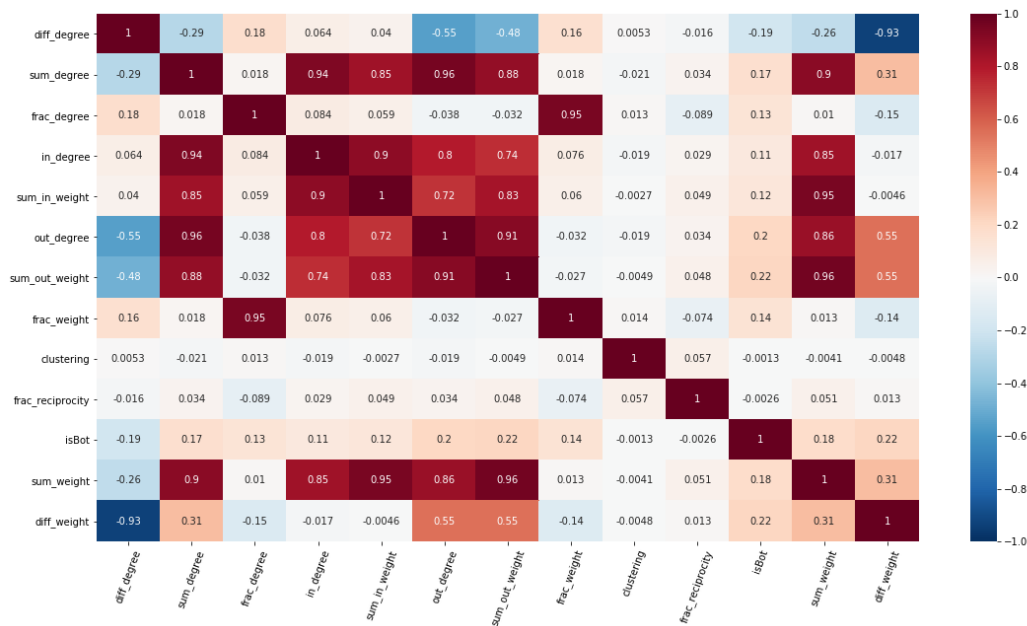


Figura B.1: Matriz de Correlação para as características dos vértices na Maior Componente Fracamente Conexa