UNIVERSIDADE FEDERAL DA FRONTEIRA SUL
CAMPUS CHAPECÓ
CURSO DE CIÊNCIA DA COMPUTAÇÃO

MARIANA PEREIRA

ANÁLISE EXPLORATÓRIA DE TWEETS UTILIZANDO MODELAGEM DE TÓPICOS PARA TEXTOS CURTOS: CASO OLIMPÍADAS RIO 2016

CHAPECÓ
2019

MARIANA PEREIRA

ANÁLISE EXPLORATÓRIA DE TWEETS UTILIZANDO MODELAGEM DE TÓPICOS PARA TEXTOS CURTOS: CASO OLIMPÍADAS RIO 2016

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do grau de Bacharel em Ciência da Computação da Universidade Federal da Fronteira Sul.

Orientador: Prof. Dr. Denio Duarte

CHAPECÓ
2019

Pereira, Mariana

Análise Exploratória de Tweets Utilizando Modelagem de Tópicos para Textos Curtos: Caso Olimpíadas Rio 2016 / por Mariana Pereira. – 2019.

59 f.: il.; 30cm.

Orientador: Denio Duarte

Monografia (Graduação) - Universidade Federal da Fronteira Sul, Ciência da Computação, Curso de Ciência da Computação, SC, 2019.

1. Textos Curtos. 2. Modelagem de Tópicos. 3. BTM. 4. Twitter. 5. Olimpíadas. I. Duarte, Denio. II. Título.

E-mail: mariana.pereira_@hotmail.com

MARIANA PEREIRA

ANÁLISE EXPLORATÓRIA DE TWEETS UTILIZANDO MODELAGEM DE TÓPICOS PARA TEXTOS CURTOS: CASO OLIMPÍADAS RIO 2016

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do grau de Bacharel em Ciência da Computação da Universidade Federal da Fronteira Sul.

Orientador: Prof. Dr. Denio Duarte

Este trabalho de conclusão de curso foi defendido e aprovado pela banca em: 09 / 12 / 2019

BANCA EXAMINADORA:

Dr. Denio Duarte - UFFS

Dr. Guilherme Dal Bianco - UFFS

Ma. Andressa Sebben - UFFS

RESUMO

A modelagem de tópicos é uma forma de mineração de texto que visa extrair, dada uma coleção de documentos, os principais tópicos que representem os assuntos abordados nos documentos da coleção. Um documento, que pode ser curto ou longo, pode ser definido como uma mistura de tópicos, sendo um conjunto de palavras ordenadas por suas probabilidades de ocorrência. Abordagens convencionais de modelagem de tópicos como LDA e PSLA foram desenvolvidas para serem aplicadas em documentos longos e, quando aplicados em textos curtos, não são tão eficientes pelo fato de não conseguir lidar bem com a dispersão dos dados. Sendo assim, para a extração de tópicos em textos curtos, se faz necessário a utilização de modelos de tópicos específicos para tal. Posto isso, neste trabalho será realizada uma análise exploratória na base de dados do Twitter, fazendo a utilização do modelo BTM (*Biterm Topic Model*) para descobrir os principais tópicos discutidos durante o período das Olimpíadas Rio 2016. Como resultado, os tópicos obtidos representaram a coleção e refletiram os acontecimentos ocorridos nos Jogos Olímpicos, principalmente os que fizeram referência ao Brasil.

Palavras-chave: Textos Curtos. Modelagem de Tópicos. BTM. Twitter. Olimpíadas.

ABSTRACT

Topic modeling is a data mining problem that aims to extract, given a document collection, the main topics that represent the subjects covered in the collection documents. A document, which can be short or long, can be defined as a mixture of topics, being a set of words ordered by their probability of occurrence. Conventional approaches for topic modeling such as LDA and PSLA have been used in long documents and when used it in short texts may not work well since conventional topic models suffer from the severe data sparsity. So, in this paper, an exploratory analysis is performed in the Twitter database applying the Biterm Topic Model (BTM) to discover the main topics discussed during the Rio 2016 Olympic Games. As a result, the topics obtained represents the collection of documents and reflects the events that occurred at the Olympic Games, especially those that made reference to Brazil.

Keywords: Short Text, Topic Model, BTM, Twitter, Olympic Games.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE APÊNDICES

LISTA DE ABREVIATURAS E SIGLAS

BTM	*Biterm Topic Model*
LDA	*Latent Dirichlet Allocation*
PLSA	*Probabilistic Latent Semantic Analysis*

SUMÁRIO

1 INTRODUÇÃO

Modelagem de Tópicos são algoritmos que visam resolver problemas de mineração de dados e descobrir os principais assuntos abordados em uma coleção de documentos. Utilizada para organizar e/ou resumir grandes quantidades de dados, a modelagem de tópicos fornece uma solução para o problema de gerenciar grandes arquivos de documentos [Blei, 2012].

Segundo [Steyvers and Griffiths, 2007], tange-se na ideia de que nos modelos de tópicos, os documentos são uma combinação de tópicos, os quais são arranjos de palavras sobre tópicos. O modelo torna-se generativo, especificando procedimentos probabilísticos simples onde os documentos podem ser originados. Ao gerar um novo documento, será escolhida uma composição sobre tópicos e um tópico, aleatoriamente, de forma que possa-se gerar palavras a partir desse tópico. Quando for necessário inverter esse processo, na geração de documentos, podem ser usadas técnicas estatísticas.

Modelos convencionais de modelagem de tópicos, como LDA, proposto por [Blei et al., 2003], e PLSA, proposto por [Hofmann, 2017], são modelos que foram desenvolvidos para serem aplicados em documentos grandes, ou com pelo menos grande quantidade de palavras. Quando aplicados em documentos com textos curtos, podem não funcionar da forma esperada pelo fato de que não conseguem lidar bem com a dispersão dos dados.

Com o crescimento das redes sociais e a disseminação dos textos curtos na web, torna-se cada vez mais necessário o desenvolvimento de modelos de tópicos que consigam analisar textos curtos e que resolvam o problema dos dados esparsos. Tendo isso em vista, [Yan et al., 2013] propuseram um novo modelo, mais básico, para modelar textos curtos, conhecido como *Biterm Model Topic (BTM)*.

Conforme [Yan et al., 2013], o BTM faz a modelagem evidenciando a co-ocorrência de palavras padrões, assim, melhorando o aprendizado do tópico e resolvendo o problema de palavras esparsas no documento. Deste modo, torna-se mais eficaz o processo de análise de documentos que incluem textos curtos.

Desta forma, a descoberta de tópicos em textos curtos é determinante para a execução de várias tarefas de análise de conteúdo. Posto isto, este trabalho propõe a extração e análise de um conjunto de dados textuais a fim de explorar, agrupar e classificar os principais tópicos abordados em uma rede social para um determinado evento. A análise será conduzida na base de dados do microblog *Twitter*, selecionando apenas *tweets* ocorridos durante o evento das

Olimpíadas, sediada na cidade do Rio de Janeiro no ano de 2016, tendo como foco a utilização do modelo BTM (*Biterm Topic Model*) para descobrir os principais tópicos discutidos durante o período dos Jogos.

1.1 Objetivos

1.1.1 Objetivo Geral

Analisar e identificar os principais tópicos discutidos no *Twitter*, no período dos Jogos Olímpicos de 2016, utilizando como técnica de extração abordagens para textos curtos.

1.1.2 Objetivos Específicos

- Definir o método a ser utilizado no trabalho;
- Definir o período de datas para obtenção dos *tweets*;
- Definir o número de tópicos que serão extraídos;
- Identificar a melhor forma para obter os dados (*tweets*) de acordo com as métricas pré estabelecidas;
- Realizar o pré-processamento e eliminar *stop words* da coleção de textos obtidos;
- Extrair os tópicos utilizando o modelo BTM;
- Verificar e analisar os principais assuntos abordados com referência aos Jogos Olímpicos Rio 2016;
- Comparar e classificar os resultados obtidos.

1.2 Justificativa

As mídias sociais têm se tornado uma importante plataforma de interação entre as pessoas do mundo todo. Usadas também como um canal de comunicação, as mídias sociais transmitem informação até o usuário de forma muito fácil e rápida, possibilitando que seja compartilhado conteúdo de assuntos diversos e em tempo real [Lee et al., 2011]. Como exemplo, pode-se citar o microblog *Twitter*.

Diariamente, milhões de *tweets*, conhecidos como textos curtos de até 140 caracteres, são postados pelos usuários do *Twitter*. Englobando os mais diferentes assuntos e gerando uma grande quantidade de dados, os *tweets* variam entre expressão de opinião, propagação de informação, conteúdo relacionados à política, desastres, celebridades e até mesmo grandes eventos internacionais como os Jogos Olímpicos.

Conforme estimado pelo *Twitter*, no período do evento das Olimpíadas Rio 2016 foram registrados cerca de 187 milhões de *tweets*[1]. Esses registros, ao serem analisados e compreendidos, podem fornecer importantes dados acerca de quais assuntos dentro do microblog foram os mais discutidos. Segundo [Fukuyama and Wakabayashi, 2018], a extração de tópicos em microblogs é uma abordagem promissora para descobrir tópicos populares no mundo. Além disso torna-se perceptível, através de uma análise temporal, o comportamento dos elementos supracitados.

Para realizar a análise, visa-se preparar dados da base de dados do *Twitter*, removendo palavras e fragmentos de texto que não venham a ser produtivos no problema em questão, para posteriormente fazer a extração dos tópicos. Como resultado, serão obtidos os tópicos que representam a base de dados de acordo com a estrutura proposta ao modelo.

1.3 Estrutura do Trabalho

O presente trabalho está estruturado da seguinte forma: nos Capítulos 2 e 3 é apresentado o referencial teórico. No capítulo 4 são apresentados os trabalhos relacionados. O Capítulo 5 apresenta o projeto de experimento que define as etapas e procedimentos utilizados para atingir os objetivos definidos. No Capítulo 6 é apresentado o experimento realizado e os resultados obtidos. Por fim, no Capítulo 7 são apresentadas as conclusões. A posteriori, estão organizadas as referências bibliográficas.

[1] blog.twitter.com/pt_br/a/pt/2016/rio2016-a-emo-o-dos-jogos-ol-mpicos-no-twitter.html

2 TWITTER/OLIMPÍADAS

2.1 Olimpíadas

Em 2009, o Comitê Olímpico Internacional escolheu o Rio de Janeiro para sediar os Jogos Olímpicos de 2016. O anúncio marcou a primeira vez em que as Olimpíadas foram concedidas a uma cidade da América do Sul, um momento extremamente significativo na história olímpica e também na história do Brasil como comunidade esportiva [Millington and Darnell, 2014].

Ocorrido entre os dias 05 e 21 de agosto de 2016, os jogos Olímpicos tiveram aproximadamente 10.500 atletas de 206 países, inscritos em 42 modalidades diferentes nas quais foram disputadas 306 medalhas de ouro, 306 medalhas de prata e 359 medalhas de bronze [2].

Nos 19 dias de competição, dentre as 306 provas ocorridas, 136 foram femininas, 161 masculinas e 9 mistas. Destacaram-se como os esportes mais populares entre os brasileiros: futebol, vôlei de quadra e de areia. Já no ranking mundial, destacaram-se o futebol, vôlei de quadra e ginástica olímpica [3].

Com mais de 62 milhões de pessoas assistindo às competições na TV aberta, o índice de popularidade do assunto tornou-se mundial e atingiu todos os tipos de mídia, incluindo as redes sociais [4]. De acordo com dados fornecidos pelo *Twitter*, foram mais de 180 milhões de *tweets* enviados sobre o assunto dos Jogos Olímpicos. Destacou-se como o momento mais tweetado do período, o gol decisivo marcado pelo jogador Neymar durante a partida do Brasil contra Alemanha no futebol, o qual garantiu a primeira medalha de ouro da história para o país [5].

Ainda segundo o *Twitter*, nomes como Michael Phelps, Usain Bolt e Neymar Jr. lideraram o ranking dos atletas mais comentados da plataforma. Já nas modalidades olímpicas, natação, futebol e atletismo foram as mais mencionadas.

Por fim, acerca dos assuntos que estiveram na mídia, pode-se salientar que houveram três momentos distintos relacionados aos Jogos Olímpicos. Em um primeiro momento, houve muita expectativa de como seria o evento e a mídia estrangeira destacava possíveis problemas de segurança, infraestrutura e o risco da *zika*. Após a cerimônia de abertura ter superado as expectativas, assuntos relacionados às competições marcaram o segundo momento. No terceiro

[2] www.ebc.com.br/esportes/rio2016/2016/07/conheca-modalidades-das-olimpiadas-rio-2016
[3] br.blastingnews.com/brasil/2016/05/serao-42-modalidades-esportivas-nas-olimpiadas-rio-2016-00909597.html
[4] exame.abril.com.br/marketing/os-10-esportes-mais-assistidos-pelos-brasileiros-na-rio-2016
[5] blog.twitter.com/pt_br/a/pt/2016/rio2016-a-emo-o-dos-jogos-ol-mpicos-no-twitter.html

e último momento, após o encerramento do evento, tópicos positivos de superação e agradecimento sobressaíram nas mídias [6].

Este trabalho pretende explorar os assuntos que foram discutidos três dias antecedendo o início dos Jogos Olímpicos, durante todo o evento e três dias após o término. Para identificar tais assuntos, será utilizada a abordagem de extração de tópicos para textos curtos BTM.

2.2 Twitter

O *Twitter* é um microblog, que têm sido usado com o objetivo de disseminação de informação em tempo real. Os usuários têm acesso aos mais variados assuntos, que vão desde celebridades, política, eventos esportivos e notícias. É uma rede social muito utilizada por várias campanhas tais como publicitárias, eleitorais e como uma mídia de notícias [Lee et al., 2011].

Com o limite de 140 caracteres, as publicações dos usuários são chamadas de *tweets*. Geralmente são textos curtos que são caracterizados por expressarem opiniões, críticas, comentários a respeito de algo ou alguém. Os *tweets* são públicos e podem ser visualizados pelo próprio autor, seguidores do autor ou até mesmo alguém interessado nos *tweets*.

Normalmente, para identificar o tema ou assunto do *tweet* publicado, faz-se o uso das *hashtags*. Caracterizadas como palavras-chaves antecedidas pelo símbolo da cerquilha (#), as *hashtags* facilitam a exibição dos *tweets* na busca do *Twitter* e ao se referenciarem a assuntos populares, são incluídas na lista de assuntos do momento.

A Figura 2.1 apresenta um exemplo de um *tweet* que foi publicado no microblog *Twitter* em 20 de agosto de 2016. O mesmo contém inicialmente o nome do usuário em destaque (ex.: cleytu), logo ao lado, a imagem do perfil do mesmo, seguido do id do usuário e a data em que o texto foi publicado. O texto segue logo abaixo, podendo ou não ser seguido de uma imagem ou vídeo de escolha do usuário. Ao final do texto, são utilizadas hashtags que referenciam o assunto do *tweet* (ex.: #Rio2016, #Futebol).

Desde que foi lançado, em 2006, o Twitter teve um crescimento muito rápido e sua popularidade aumentou consideravelmente nos últimos anos. Estima-se que mais de 500 milhões de *tweets* são enviados por dia e que muitos deles acabam sendo sobre um tópico que se tornou popular [7]. Quando isso ocorre, o assunto em questão estará listado em uma lista de popularidade

[6] g1.globo.com/rio-de-janeiro/olimpiadas/rio2016/blog/brasil-visto-de-fora-na-olimpiada/
[7] www.omnicoreagency.com/twitter-statistics/

Figura 2.1: Exemplo de um *tweet* publicado durante o período das Olimpíadas Rio 2016.

fornecida pela plataforma, chamada de *Trending Topics*, contendo os assuntos do momento [Lee et al., 2011].

Figura 2.2: *Trending Topics* do Brasil retirados em 02/06/2019 às 15:54h.

Na Figura 2.2 pode-se observar uma lista dos assuntos mais comentados no Brasil, ou seja, *Trending Topics*, os quais são formados por nomes próprios (ex.: Anderson Martins) ou *hashtags* (ex.: #TamanhoFamília). No entanto, esta lista está em constante atualização, mos-

trando apenas os assuntos daquele exato momento e que, por ser variável, não é possível saber quais foram os *Trending Topics* de datas anteriores.

A Figura 2.3 exibe exemplos de *tweets* em português postados no período dos Jogos Olímpicos. Os textos contém a palavra-chave #Rio2016 que os classifica como sendo relacionados ao assunto das Olimpíadas.

Figura 2.3: Exemplos de *tweets* publicados durante o período das Olimpíadas Rio 2016.

O *Twitter* será utilizado como objeto de estudo neste trabalho e servirá como fonte da base de dados. Sendo assim, será feita a extração de *tweets* públicos para a análise dos assuntos mais frequentemente discutidos durante os Jogos Olímpicos Rio 2016.

3 MODELAGEM PROBABILÍSTICA DE TÓPICOS

Algoritmos de modelagem de tópicos são métodos utilizados para analisar o conteúdo de um documento e o significado das palavras contidas no mesmo. Apesar de existirem pequenas diferenças estatísticas entre os modelos, no geral, todos seguem a mesma ideia de que os documentos são uma mistura de tópicos [Steyvers and Griffiths, 2007]. Posto isso, uma vez que os tópicos são gerados a partir dos documentos e textos originais, não existindo a necessidade de entradas rotuladas, considera-se que são algoritmos de aprendizagem não supervisionada.

Segundo [Blei, 2012], os modelos de tópicos visam descobrir padrões latentes (ocultos) em documentos, que além de sua aplicação em textos, podem ser utilizados em outros tipos de dados. Considerando um texto sobre determinado tópico, aparecerão palavras relacionadas ao tema com certa frequência em que serão rotulados como sendo os tópicos descobertos pelo algoritmo de modelagem de tópicos.

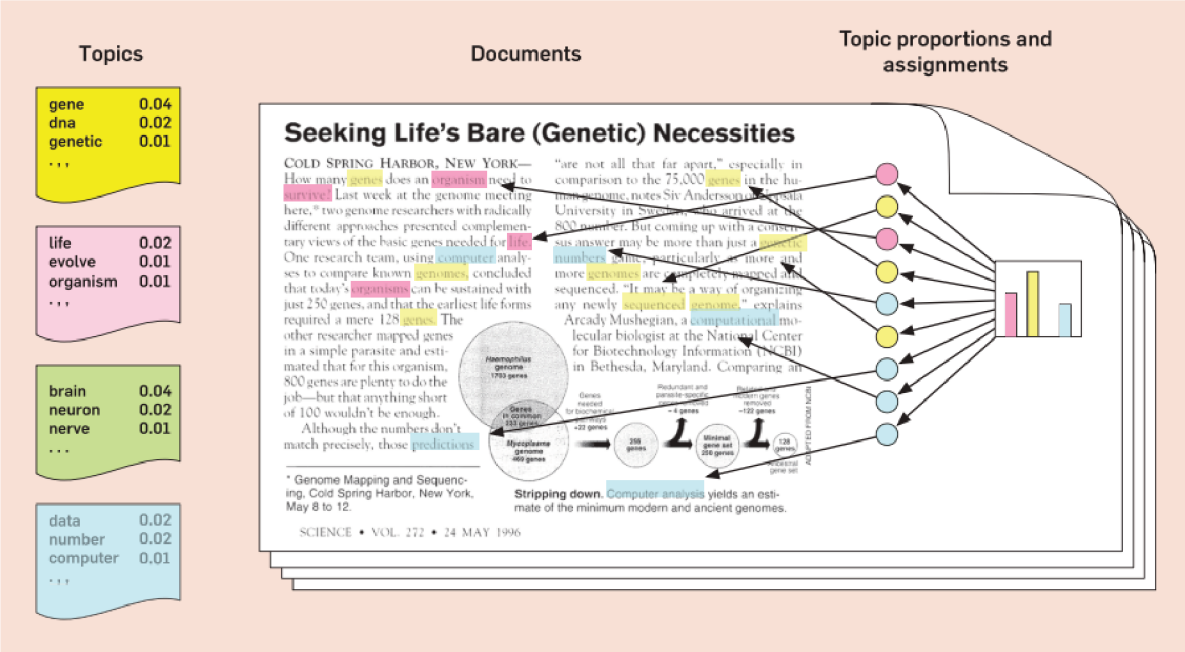

Figura 3.1: Exemplo de Modelagem de Tópicos [Blei, 2012].

A Figura 3.1 mostra um exemplo de artigo intitulado *Seeking Life's Bare (Genetic) Necessities*, o qual disserta sobre o uso da análise de dados para determinar o número de genes que um organismo precisa, em um sentido evolucionário, para sobreviver. Manualmente, foram destacadas diferentes palavras usadas no artigo. Em azul, destacam-se as palavras sobre análise de dados. Palavras sobre biologia evolucionária foram destacadas em rosa e em amarelo, palavras referentes à genética. Segundo [Blei, 2012], se fossem destacadas cada palavra do artigo, seria

possível visualizar que há combinações entre os assuntos genética, análise de dados e biologia evolutiva em diferentes proporções. Além disso, saber que este artigo combina esses tópicos, ajudaria a situá-lo em uma coleção de artigos científicos.

Deste modo, a modelagem de tópicos é um conjunto de algoritmos que foram desenvolvidos com o objetivo de obter informações em grandes arquivos de texto. Segundo [Blei, 2012], estes algoritmos são métodos estatísticos que analisam as palavras dos textos originais, visando descobrir os temas abordados, como eles se conectam e como mudam ao longo do tempo.

3.1 Tópicos

Pode-se considerar que os tópicos são uma distribuição de palavras a respeito de um documento, sendo que cada tópico é um padrão recorrente de co-ocorrência de palavras. Segundo [Blei, 2012], pode-se definir um tópico como como sendo uma distribuição sobre um vocabulário fixado. Por exemplo, nos tópicos sobre *genética*, terão uma maior probabilidade de aparecerem palavras relacionadas à genética, assim como em tópicos sobre *biologia evolucionária*, terão maior probabilidade de aparecem palavras relacionadas com biologia evolucionária.

"Genetics"	"Evolution"	"Disease"	"Computers"
human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

Figura 3.2: Exemplo de tópicos gerados [Blei, 2012].

Na Figura 3.2 são apresentados quatro tópicos distintos encontrados em um mesmo documento, cada tópico com as *top-15 words* (quinze palavras mais frequentes), sendo que as palavras estão ordenadas conforme a sua maior probabilidade de ocorrência. Tendo como exemplo o tópico sobre *Evolução*, as palavras *evolução*, *evolucionário* e *espécies* são as palavras com maior probabilidade de ocorrência no tópico citado.

3.2 Documentos

Documento é toda informação registrada em um suporte material, sendo também uma fonte de informação. Podem ser considerado longos ou curtos, dependendo de sua extensão. Por exemplo, artigos científicos, notícias, podem ser considerados documentos longos, já comentários, postagens, *tweets*, são considerados documentos curtos.

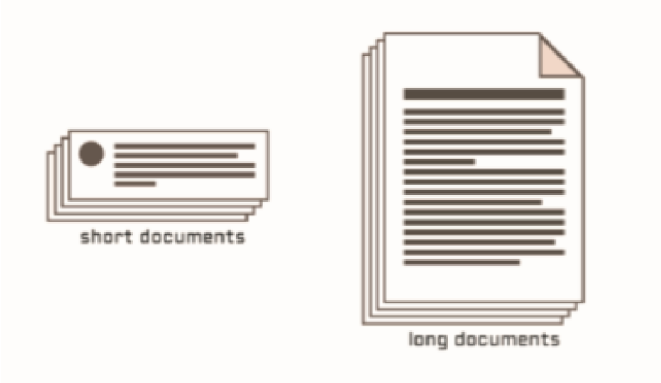

Figura 3.3: Exemplo de documento curto e documento longo.

Abordagens tradicionais de extração de tópicos consideram que um documento é composto por uma mistura de tópicos, ou seja, vários tópicos podem descrever um mesmo documento considerando probabilidades diferentes. Conforme [Blei et al., 2003], a ideia básica é que os documentos são representados como misturas aleatórias sobre tópicos latentes (ocultos), em que cada tópico é caracterizado por uma distribuição de probabilidade sobre palavras.

Modelos como, por exemplo, *Latent Dirichlet Allocation* (LDA) e *Probabilistic Latent Semantic Analysis* (PLSA) assumem que um documento pode ser visto como uma *bag-of-words* (sacola de palavras) cuja ordem das palavras no documento não importa. No entanto, considerando que modelos convencionais foram feitos para serem aplicados em documentos longos, quando utilizados em documentos curtos, os dados tornam-se esparsos, tornando a abordagem menos eficaz.

Nos documentos curtos, algumas características podem ser citadas, como por exemplo, a maioria das palavras aparecem apenas uma vez no texto, não dando contexto suficiente para identificar ambiguidades. O contexto também é muito limitado e devido à extensão do documento, palavras relevantes aparecem com menos frequência. Daí, surge a necessidade de utilização de abordagens específicas para utilização em textos curtos, como por exemplo, *Biterm Topic Model* (BTM).

3.3 Biterm Topic Model (BTM)

Biterm Topic Model (BTM) é um modelo generativo de extração de tópicos para textos curtos proposto por [Yan et al., 2013] que visa resolver o problema da dispersão dos dados. O mesmo extrai tópicos modelando diretamente a geração de *biterms* (termos-pares) em toda a coleção de documentos. Termo-par é um par de palavras não ordenadas em um contexto curto, ou seja, uma forma de explicitar a co-ocorrência de palavras relacionadas em documentos.

Em textos curtos com contexto limitado, como por exemplo, *tweets* e mensagens de texto, considera-se cada documento como uma unidade de contexto individual. Nesse caso, quaisquer duas palavras distintas em um documento formam um *biterm*. Por exemplo, um documento com três palavras distintas gera três *biterms*:

$$(p1, p2, p3) \Rightarrow \{(p1, p2), (p2, p3), (p1, p3)\}$$

Após a extração dos *biterms* em cada documento, a coleção de documentos passa a ser um conjunto de *biterms* [Yan et al., 2013].

No modelo BTM, considera-se que toda a coleção de documentos é uma mistura de tópicos, onde cada *biterm* é extraído de um tópico específico independentemente. Sendo assim, se duas palavras são encontradas frequentemente juntas, então elas provavelmente pertencem ao mesmo tópico. Suponha que α e β sejam hiperparâmetros que definem a distribuição a priore de *Dirichlet* relacionados à distribuição documento-termo e tópico-palavra, o processo generativo do BTM pode ser representado graficamente conforme a Figura 3.4.

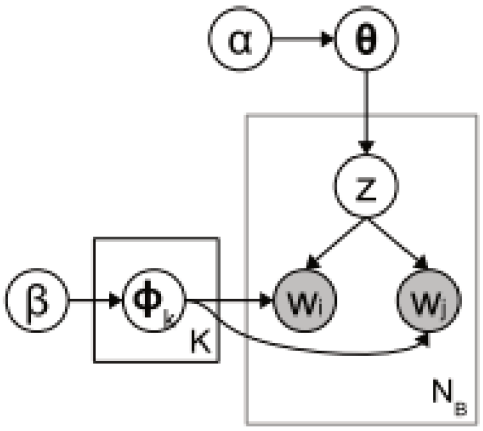

Figura 3.4: Representação gráfica do modelo BTM [Yan et al., 2013].

Conforme [Yan et al., 2013], o modelo BTM pode ser descrito formalmente da seguinte forma:

1. Para cada tópico z:

   (a) extrai uma distribuição de palavras específicas por tópico $\phi_k \sim$ Dirichlet(β)

2. Extrai uma distribuição de tópico $\theta \sim$ Dirichlet(α) para toda a coleção

3. Para cada *biterm* b na coleção de *biterms* B:

   (a) extrai uma distribuição de tópico $z \sim$ Multinomial (θ)

   (b) extrai duas palavras: w_i, $w_j \sim$ Multinomial (ϕ_k)

Na Figura 3.4, cada retângulo representa um processo de repetição, no qual o número de vezes é rotulado pela variável em seu interior. Sendo assim, dada uma coleção de documentos N_D, supõe-se que a coleção contém N_B *biterms* $B = \{b_i\}N_b$, i = 1, com $b_i = (w_i, 1, w_i, 2)$, e K tópicos expressos sobre W palavras únicas no vocabulário.

O modelo apresentado possui uma implementação na linguagem *C++*[8] e será utilizado neste trabalho como o método de extração dos tópicos para textos curtos, a fim de identificar os assuntos discutidos durante os Jogos Olímpicos Rio 2016.

[8] https://github.com/xiaohuiyan/BTM

4 TRABALHOS RELACIONADOS

Este capítulo apresenta alguns dos trabalhos relacionados a esta proposta. Foram selecionados três trabalhos [Fukuyama and Wakabayashi, 2018, Ligutom et al., 2017, Oliveira et al., 2018], que utilizam técnicas de extração de tópicos para textos curtos e que também serviram de base para esta proposta.

O trabalho de [Fukuyama and Wakabayashi, 2018] propõe um método para calcular a frequência de ocorrência de tópicos extraídos de um microblog, para examinar se os tópicos obtidos estão relacionados a eventos reais que ocorreram no mesmo período de tempo. Para o experimento, os dados utilizados foram *tweets* obtidos do microblog *Twitter*, publicados em um período de doze dias em agosto de 2012. Além do método citado, os autores também tiveram como objetivo fornecer uma abordagem online para extração consistente de tópicos.

Para a execução do método proposto, primeiramente os tópicos foram extraídos utilizando-se do modelo BTM. Os *tweets* foram agrupados em lotes por hora e treinados online de forma cronológica. Além disto, obteve-se também a frequência de ocorrência de cada tópico, utilizando uma estimativa do número de *tweets* para cada tópico em cada período de tempo.

Como uma análise de popularidade, [Fukuyama and Wakabayashi, 2018] calcularam o *burstiness* de cada tópico, a partir da frequência de ocorrência do tópico, onde o valor de 0 a 1 indica a intensidade de sincronismo entre o *tweet* e o evento real.

Por fim, foram examinados manualmente os tópicos que tiveram maior frequência de ocorrência e foi possível obter 5 tópicos que corresponderam a eventos reais, conforme Figura 4.1 e Figura 4.2.

Em [Ligutom et al., 2017] é feita uma análise exploratória de *tweets*, também utilizando-

topic description	burstiness	
	date/time (JST)	score
About Japanese gymnasts	2012-8-2/01:00 - 02:00	0.979
About Kohei Uchimura, Japanese gymnasts	2012-8-5/22:00 - 23:00	0.977
About Japanese football players	2012-8-4/21:00 - 22:00	0.975
About Japanese swimmers	2012-8-2/03:00 - 04:00	0.967
The Hiroshima peace memorial ceremony	2012-8-6/08:00 - 09:00	0.964

Figura 4.1: Tabela com os 5 tópicos que tiveram maior ocorrência no experimento realizado por [Fukuyama and Wakabayashi, 2018].

event description	starting time of the event date/time (JST)
The all-around gymnastics in the London Olympics Games	2012-8-2/00:30
The event Final gymnastics in the London Olympics Games	2012-8-5/22:00
The football match Japan vs Egypt in the London Olympics Games	2012-8-4/20:00
The men's 200M breaststroke final in the London Olympics Games	2012-8-2/03:30
The Hiroshima peace memorial ceremony	2012-8-6/08:00

Figura 4.2: Tabela contendo os eventos correspondentes aos tópicos apresentados na Figura 4.1 [Fukuyama and Wakabayashi, 2018].

se da base de dados do *Twitter*. O objetivo do trabalho foi verificar como os tópicos descobertos refletem o comportamento dos filipinos em momentos de grandes desastres. No experimento, foram utilizados tópicos discutidos em relação aos tufões ocorridos entre os anos de 2013 e 2014.

Para realizar o estudo, foram usados 9.714 *tweets*, coletados no período de Fevereiro de 2013 a Novembro de 2014, utilizando como parâmetros de busca a geolocalização da região metropolitana da cidade de Manila, a capital nacional das Filipinas e palavras-chave como *typhoon* e *bagyo* (termo local para tufão).

Para identificar os tópicos dentre os *tweets* coletados, da mesma forma que o trabalho anterior, foi utilizado o modelo de extração para textos curtos, BTM. No experimento, após a limpeza dos dados e a remoção de *stop words*, definiu-se que 15 era o número aceitável de tópicos para verificar quais seriam os comportamentos notados.

Como resultado, comportamentos como união e resiliência, determinação e antagonismo, foram encontrados a partir dos tópicos. [Ligutom et al., 2017] determinaram que os filipinos expressaram diferentes pensamentos e opiniões ao longo dos *tweets* coletados e que os mesmos são essenciais para revelar seus comportamentos perante a passagem do tufão.

A Tabela 4.1 apresenta um extrato dos resultados obtidos no experimento, no qual dentre os 15 tópicos obtidos no trabalho, foram selecionados aqueles que condizem com os comportamentos observados. As palavras mais frequentes nos tópicos são apresentadas em inglês e filipino e refletem os comportamentos de preocupação com a tempestade, conforme tópico 01, com palavras "chuva", "forte", "Deus", "tempo". O tópico 3 expressa sentimentos com os esforços de ajuda no qual possui palavras como "vítimas", "obrigado", "ajuda", orar". O tópico 5 é sobre vida de estudante contendo palavras como "faculdade", "escola", "estudantes", "trabalho". Nos tópicos 10, 11 e 14, são apresentadas palavras que condizem com os sentimentos de

esforços de reconstrução, antagonismo e recomeço.

Label	**Topic models**
Topic 01: Worrying about the storm	safe, rain, manila, stay, weather, philippines, classes, heavy, typhoonmaring, suspended, flood, home, today, typhoons, morning, rainy, super, god, strong, tomorrow
Topic 03: Relief efforts	haiyan, victims, philippines, world, affected, thank, super, help, tacloban, pray, puso, hit, typhoonhaiyan, walk, helping, typhoonyolanda, leyte, yolandaph, relief, phthankyou
Topic 05: Student life	college, elementary, isda, trabaho, preschool, highschool, immortal, school, imortal, high, pre, tamad, hinahangad, estudyanteng, wahahahahahaha, students, waterproof, nagtatrabaho, lol, tatrabaho
Topic 10: Rebuilding efforts after the typhoon	presidential, assistant, phl, rehabilitation, survivors, sector, rebuilds, filipino, live, caves, private, agaton, senator, hit, low, pressure, genii, geniimeqqamundiial, team, recorded
Topic 11: Antagonism	napoles, nasalanta, cuddle, weather, pinas, corrupt, 24oras, anyare, napabayaan, tatama, report, super, flash, bitterness, eastern, ipangalan, wreaked, havoc, barrel, malakas
Topic 14: Starting again	start, lives, vice, binay, president, jejomar, survivors, phl, tuesday, led, team, genii, geniimeqqamundiial, repeat, convinced, napakasaya, andy, x9, pumasok, habang

Tabela 4.1: Extrato dos tópicos obtidos em [Ligutom et al., 2017].

O trabalho de [Oliveira et al., 2018] propõe um método para identificar *tweets* políticos de congressistas brasileiros e classificá-los como: políticos e não políticos. Para o método, foram coletados *tweets* de todos os deputados que tiveram uma conta ativa no *Twitter* e que trabalharam no parlamento brasileiro entre Outubro de 2013 à Outubro de 2017. O objetivo deste método é distinguir comunicações com foco em opiniões políticas significativas daquelas com foco em trivialidades, como mensagens pessoais e individualistas.

Para a base de dados, foram extraídos 1.1 milhões de *tweets* públicos de 692 contas de deputados brasileiros. Após a coleta, foi feita uma validação manual de todas as contas para verificar a veracidade dos *tweets*, bem como foi realizada a limpeza dos dados, removendo pontuações, *tweets* duplicados, *stop words*, *hashtags*, URLs e palavras com menos de dois caracteres.

Primeiramente, para identificação dos *tweets* políticos, foram realizados vários processos desde seleção e rotulação manual dos posts e técnicas de redes neurais como método de classificação. Para uma melhor visualização das palavras mais frequentes nos textos, foram utilizados gráficos na forma de nuvem de palavras, os quais mostram as principais palavras relacionadas à política no período eleitoral, conforme Figura 4.3[9]. As palavras mais proeminentes são "campanha", "federal", "apoio", "governo"e "dilma", que indicam claramente os *tweets* da

[9] Todos os resultados foram traduzidos do português para o inglês.

campanha. Em contraste, a nuvem de palavras à direta, mostra as palavras mais predominantes dos *tweets* não políticos para o mesmo período.

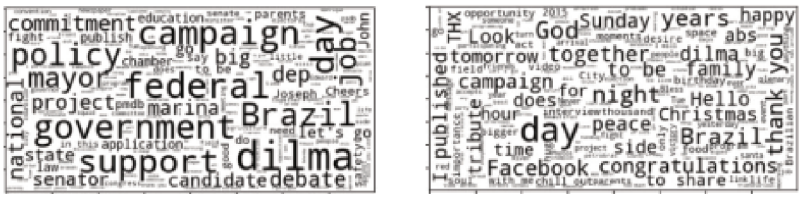

Figura 4.3: Nuvens de palavras de *tweets* políticos (esquerda) e não políticos (direita) durante as eleições de 2014 [Oliveira et al., 2018].

Após a classificação dos *tweets*, foi utilizada a abordagem de extração de tópicos para textos curtos, BTM, para a identificação dos principais tópicos dos *tweets* dos parlamentares. Utilizando parâmetros como número de tópicos $K = 10$ e número de iterações = 1.000, o método destacou o uso do personalismo pelos parlamentares brasileiros e evidenciou as cinco palavras mais frequentes em cada tópico, tanto para *tweets* políticos quanto para *tweets* não políticos.

#	Non-Political Topics	% tweets	Political Topics	% tweets
1	brazil can be country go	12	government dilma against politics brazil	5
2	today city big day friends	5	mayor deputy today president meeting	3
3	day all god friends week	3	dilma lula temer against president	4
4	thousand brazil millions go project	6	meeting about today national chamber	8
5	for friend big brazil today	5	health federal government resources state	4
6	day today brazil years congratulations	13	chamber commission deputies pec project	3
7	program live day radio interview	2	about law against education project	5
8	together hugs see moments relaxation	1	day today every year god	3
9	facebook posted photos album posted	5	government dilma brazil year millions	6
10	food soul grace lauds los	1	deputy about federal today chamber	6

Figura 4.4: Tabela contendo as 5 palavras mais frequentes em cada tópico [Oliveira et al., 2018].

Conforme resultados mostrados na Figura 4.4, em tópicos não políticos, existem muitos tópicos de assuntos pessoais representados por palavras como “parabéns”, “amigo”, “abraços”, “momentos” e "juntos”. No que diz respeito aos tópicos políticos, observa-se indícios de patrocínio ou de um comportamento mais pragmático dos políticos, como os tópicos 2 e 5, outros relacionados ao cronograma dos deputados como os tópicos 6, 7 e 10. Como a classificação não é perfeita, foi encontrado um tópico de assunto pessoal, 8, no meio dos tópicos dos *tweets* de assunto político. No entanto, os outros temas refletem fatos e questões importantes relacionados ao cenário político brasileiro, como a crise política, tópicos 1 e 3.

5 PROJETO DE EXPERIMENTO

Os experimentos foram realizados conforme a seguinte ordem de execução: criação da base de dados com os *tweets* coletados seguindo parâmetros pré-definidos; pré-processamento com a limpeza dos dados; aplicação do algoritmo de modelagem de tópicos BTM; e por fim, pós-processamento dos dados.

5.1 Configuração de Ambiente

Para o desenvolvimento do experimento, utilizou-se de uma máquina virtual com o sistema operacional *CentOS Linux release 7.6.1810 (Core)*, configurada com 8 vCPUs, 16GB de memória RAM e 240GB de armazenamento SSD, e um computador com dois processadores *Intel(R) Xeon(R) CPU E5-2670 @ 2.60GHz (32 cores)*, 64GB de memória RAM, placa gráfica NVIDIA GeForce GT 730, 480GB de armazenamento SSD e sistema operacional Windows 10 Pro 64-bit.

5.2 Base de Dados

Para a base de dados desse projeto, foram utilizados *tweets* extraídos do microblog *Twitter* utilizando utilizado a ferramenta desenvolvida em *Python*, Get Old Tweets[10]. Para a extração, foram utilizadas como parâmetros as seguintes palavras-chave: "rio2016", "olimpiadas", "olimpiada", "cerimoniadeabertura", "cerimoniadeencerramento", "jogosolimpicos", "olympics", "olympicgames", "openingceremony", "closingceremony", e o período de dias de 02 de agosto de 2016 a 24 de agosto de 2016.

Inicialmente, foram criados 230 arquivos (um arquivo para cada palavra-chave x 23 dias), com um tamanho total de 576MB e contendo 2.806.785 *tweets*. Posteriormente, os arquivos foram agrupados por dia e, por fim, divididos em 9 arquivos. A Tabela 5.1 detalha as características da base de dados, com o período de dias que compreende cada arquivo, tamanho dos arquivos e quantidade de *tweets*.

Na primeira coluna da Tabela 5.1, nomeada como *Arquivo*, é apresentado um sequencial de números com o objetivo de separar os arquivos utilizados na base de dados. A segunda coluna, nomeada *Dias*, apresenta a fatia de tempo utilizada para cada arquivo, ou seja, a data

[10] https://github.com/Jefferson-Henrique/GetOldTweets-python

em que os *tweets* foram coletados. É importante observar que nos arquivos 2 e 8, devido à grande quantidade de *tweets* para esses dias em que ocorreram importantes eventos, tais como cerimônia de abertura e cerimônia de encerramento dos Jogos Olímpicos, foram utilizados apenas um dia e não uma fatia de três dias como nos outros arquivos. O tamanho em *Megabytes* de cada arquivo é descrito na terceira coluna e a última coluna, *Quantidade de Tweets*, apresenta a quantidade de *tweets*/documentos contidos em cada arquivo.

Tomando como exemplo a quinta linha da Tabela 5.1, pode-se observar que o arquivo 5 corresponde a *tweets* coletados entre os dias 12 de agosto de 2016 a 14 de agosto de 2016, com um tamanho de 52,1 MB e possuindo 245.149 *tweets* coletados.

Arquivo	Dias	Tamanho	Quantidade de Tweets
1	02.08.2016 a 04.08.2016	27,2MB	128.527
2	05.08.2016	65,9MB	334.125
3	06.08.2016 a 08.08.2016	127MB	624.444
4	09.08.2016 a 11.08.2016	51MB	239.912
5	12.08.2016 a 14.08.2016	52,1MB	245.149
6	15.08.2016 a 17.08.2016	64,7MB	311.581
7	18.08.2016 a 20.08.2016	59,2MB	279.432
8	21.08.2016	53,9MB	260.063
9	22.08.2016 a 24.08.2016	74MB	350.763

Tabela 5.1: Características da base de dados

Para decidir qual seria a fatia de tempo a ser utilizada para separar os *tweets*, foi conduzida uma pesquisa empírica. Inicialmente foi definida uma fatia igual cinco dias. Porém, analisando os resultados parciais no experimento, verificou-se que, com exceção do arquivo correspondente à cerimônia de abertura dos Jogos Olímpicos, os resultados não estavam abrangendo todos os eventos ocorridos durante o período de dias selecionado. Posto isto, este resultado empírico utilizado foi corroborado com os eventos que ocorreram durante as olimpíadas, no qual, devido à grande quantidade de modalidades, vários jogos ocorreram no mesmo dia. Então, definiu-se que o período a ser utilizado seria de três dias, excluindo os dias das cerimônias de abertura e encerramento dos Jogos Olímpicos.

5.3 Pré-processamento dos dados

Para a etapa de preparação dos dados, foi realizada uma limpeza com o objetivo de evitar que os resultados saíssem incorretos. Os *tweets* coletados continham informações irrelevantes e símbolos indesejados, além de caracteres maiúsculos e minúsculos. A fim de resolver isso,

todos os *tweets* foram convertidos em letras minúsculas, bem como foi realizada a remoção das seguintes informações:

- URLs;
- Menções, ou seja, palavras seguidas do caractere especial '@';
- *Hashtags*, ou seja, palavras seguidas do caractere especial '#';
- Pontuação e acentuação;
- Caracteres especiais;
- *Stop words*;
- Dígitos, exceto 2016 quando seguido da palavra rio;
- Palavras com dois ou menos caracteres, exceto quando estivesse representando país, por exemplo, br, fr, us;
- *Tweets* contendo apenas duas ou menos palavras;
- *Tweets* duplicados.

Para a limpeza dos *tweets* foi tomado como exemplo o processo utilizado em [Ligutom et al., 2017], em que processos de *lemmatization*, utilizado para reduzir palavras ao seu lema, e *stemming*, utilizado para reduzir palavras flexionadas ao seu radical, não foram utilizados.

A Tabela 5.2 exibe um exemplo de um *tweet* antes e depois da limpeza na qual foram removidos todas as informações e caracteres indesejados. Por exemplo, o texto original continha dez palavras, sendo uma *hashtag* (#Rio2016) fazendo referência aos Jogos, uma menção (@Arena Corinthians) citando o estádio de futebol do time Corinthians, localizado na zona leste do município de São Paulo, popularmente conhecida como Itaquerão e uma URL correspondendo a uma foto publicada na rede social Instagram. Após o processamento, foram retiradas todas as informações indesejadas, restando apenas cinco palavras posteriormente utilizadas no experimento.

De um total de 2.806.785 *tweets* coletados, após a limpeza, obteve-se um conjunto de 1.548.759 *tweets*, com um total de 8.611.136 palavras e com uma média de 5,56 palavras por *tweet*.

Original	Após limpeza
Dia de futebol feminino nessas Olimpíadas. #Rio2016 @Arena Corinthians - Itaquerão https://www.instagram.com/p/BIyVNoOBHvD/	futebol feminino olimpiadas corinthians itaquerao

Tabela 5.2: Exemplo de *tweet* antes e depois do processo de limpeza.

5.4 Aplicação do BTM

Após realizado o pré-processamento dos dados, para identificar os tópicos na coleção de documentos e realizar o experimento, foi utilizado um *script* disponibilizado pelos próprios autores do modelo[11]. O arquivo `runExample.sh` foi testado e executado no Linux, no qual após informado o arquivo de entrada contendo os *tweets* coletados e limpos, o *script* executa o BTM e mostra como saída os tópicos gerados. O processo de execução do *script* consiste em três passos:

1. Indexação das palavras nos documentos:
   Inicialmente, utilizando a linguagem *Python*, é feito um mapeamento a partir do arquivo de entrada e atribuído um identificador único para cada palavra encontrada nos documentos. Por exemplo, nos arquivos de entrada a serem indexados, cada linha é um documento com o formato "palavra palavra palavra...". Após a indexação, no arquivo de saída, cada linha é um documento com o formato "palavraId palavraId palavraId...". Por fim, é criado o arquivo de vocabulário no qual cada linha está no formato "palavraId palavra".

2. Aprendizado do tópico:
   Após o processo de indexação, é necessário definir os hiperparâmetros a serem utilizados na execução do código do BTM e treinar o modelo usando os documentos representados pelos ids das palavras. Inicialmente, definiu-se o número de tópicos, denotado por K, em 5, 10, 15, 20 e 30. Além disso, foram utilizados os hiperparâmetros $\alpha = 50/K$ e β = 0.01 tal qual sugere os autores do modelo. Para todos os experimentos, o vocabulário W é definido automaticamente no script de acordo com os arquivos criados anteriormente no formato "palavraId palavra". Por fim, foi definido um número de 1.000 iterações para todas as execuções deste trabalho.

[11] https://github.com/xiaohuiyan/BTM/tree/master/script

3. Exibição dos resultados:

   Por fim, é utilizado um *script* na linguagem *Python* para exibir as principais palavras dos tópicos e suas probabilidades. A Tabela 5.3 representa um exemplo genérico de saída gerada pelo *script*. Na primeira coluna é informada a probabilidade de ocorrência do tópico na coleção de documentos $P(z)$ e na coluna *Top words*, o conjunto de palavras do tópico seguido pela sua probabilidade de ocorrência $P(w|z)$. A exibição dos tópicos também foi utilizada para auxiliar na definição da quantidade de tópicos a serem usados no experimento.

P(z)	**Top words**
$p(z)$	palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$ palavra:$p(w\|z)$

Tabela 5.3: Exemplo de arquivo de saída com os tópicos e probabilidades.

5.5 Pós-processamento dos dados

Após a obtenção dos tópicos, utilizando K igual a 5, 10, 15, 20 e 30, fez-se necessário a definição da quantidade de tópicos que seriam utilizados no experimento. Para tal definição, foram levadas em consideração notícias relacionadas ao período de tempo de cada arquivo. Após experimentos preliminares e análise dos tópicos obtidos nos resultados, observou-se que K = 15 tinha o grupo de palavras-chave mais representativo em comparação com os outros experimentos.

6 EXPERIMENTOS

Com base no projeto de experimento e na obtenção da base de dados necessária para a realização do trabalho, foi executado novamente o *script* para a obtenção e visualização dos tópicos, tendo em consideração os hiperparâmetros definidos anteriormente. Após a execução, houve a necessidade de continuar o processo de limpeza devido a existência de palavras indesejadas nos tópicos.

Para a execução do *script* de limpeza foram criados dois arquivos: `dicionario.replace` e `dicionario.delete`. No primeiro arquivo, foram acrescentadas palavras que deveriam ser substituídas na base de dados, como por exemplo, gírias, abreviações, palavras com letras duplicadas e palavras escritas de forma errada. Já no segundo, foram adicionadas palavras que deveriam ser apagadas dos dados, tais como palavras ofensivas e onomatopeias.

O processo de limpeza dos dados e execução do *script* para a obtenção dos tópicos foi realizado diversas vezes até que os tópicos gerados estivessem satisfatórios, levando em consideração para a análise todo o conhecimento empírico obtido sobre os Jogos Olímpicos Rio 2016.

Após a obtenção da versão final dos tópicos, observou-se que as palavras Rio e Brasil eram predominantes e constavam respectivamente em 60% e 90% dos tópicos. Posto isso, foi realizado novamente o processo de limpeza visando a retirada dessas duas palavras e gerando os tópicos uma última vez.

A Tabela 6.1 é um extrato dos tópicos e probabilidades de ocorrência obtidos a partir do *Arquivo 2* (Tabela 5.1) da base de dados. Os cinco tópicos apresentados são referentes aos *tweets* coletados do dia 05 de agosto de 2016, data em que ocorreu o evento da cerimônia de abertura dos Jogos Olímpicos. Os tópicos estão em ordem crescente, ou seja, o tópico mais provável aparece na primeira posição, seguidos da sua probabilidade de ocorrência e do grupo de palavras que o representam.

Tomando como exemplo o tópico 3 da Tabela 6.1, pode-se observar que a probabilidade de ocorrência do tópico na coleção de documentos $P(z)$ é igual a 0.094771, ou seja, 9,48%. Na terceira coluna, o grupo de palavras representando o tópico é composto por 'lindo', 'deus', 'gisele', 'demais', 'cerimonia', 'orgulho', 'maravilhosa', 'mundo', 'amo', 'parabens', que faz alusão a aparição da modelo Gisele Bündchen desfilando no Maracanã ao som da música "Ga-

Tópico	P(z)	Palavras
1	0.159879	pokemon, vendo, hoje, cerimonia, mundo, casa, assistir, assistindo, estar, tv.
2	0.124873	pais, lindo, copa, mundo, brasileiro, povo, dinheiro, bonito, mal, festa.
3	0.094771	lindo, deus, gisele, demais, cerimonia, orgulho, maravilhosa, mundo, amo, parabens.
4	0.089204	mundo, orgulho, pais, brasileiro, lindo, melhor, hoje, momento, amo, atleta.
5	0.068853	anitta, caetano, gil, cantando, cantar, mc, fernanda, karol, montenegro, gilberto.

Tabela 6.1: Extrato dos tópicos obtidos a partir do *Arquivo 2* (Tabela 6.1).

rota de Ipanema" cantada por Daniel Jobim, neto do compositor Tom Jobim.

Para obter uma melhor visualização dos tópicos gerados, foi realizado um processo de rotulação manual dos dados. Primeiramente foram selecionadas notícias de dois sites [12] [13] que correspondiam ao mesmo período de dias de cada arquivo da base de dados, conforme descrito na Tabela 5.1. Após, foi verificado se as palavras contidas nos tópicos existiam nas notícias e a quais assuntos ou modalidades esportivas estavam ligadas. Para a definição do rótulo, selecionou-se a modalidade e/ou assunto que correspondia à maioria das palavras. Os tópicos que não se encaixaram em nenhum assunto descrito nos sites, foram rotulados de forma empírica.

Na Tabela 6.2 é possível visualizar o tópico 6 e o respectivo grupo de palavras, que foram obtidos através do *Arquivo 3* (Tabela 5.1) da base de dados, correspondendo aos *tweets* coletados entre os dias 06 de agosto de 2016 a 08 de agosto de 2016. As palavras 'ouro', 'medalha', 'rafaela', 'silva', 'primeira', 'parabens', 'mulher', 'judo', 'primeiro' e 'tiro' foram utilizadas como parâmetro de busca nas notícias que correspondiam aos dias 06, 07 e 08 de Agosto de 2016.

Tópico	Palavras
6	ouro, medalha, rafaela, silva, primeira, parabens, mulher, judo, primeiro, tiro.

Tabela 6.2: Tópico obtido a partir do *Arquivo 3* (Tabela 5.1).

A Figura 6.1 é um extrato de uma notícia do terceiro dia dos jogos olímpicos que ocorreu em 08 de agosto de 2016. Pode-de observar que das palavras contidas no tópico, 'ouro', 'medalha', 'rafaela', 'silva', 'primeira' e 'judo' também estão contidas no site de notícias e estão relacionadas à modalidade judô. Tendo em vista que 60% das palavras do tópico estão na notícia e relacionadas à modalidade esportiva judô, o rótulo selecionado para este tópico foi

[12] https://recordtv.r7.com/rio-2016/
[13] https://www.olympic.org/rio-2016

judô. Caso a porcentagem fosse menor que 50%, não seria considerado este rótulo para este tópico.

Figura 6.1: Extrato de notícia ocorrida em 08 de Agosto de 2016.

A Tabela 6.3 representa 5 dos 15 tópicos já rotulados que foram obtidos a partir do *Arquivo 4* (Tabela 5.1). Citando como exempo o tópico 05, pode-se observar que a partir do grupo de palavras 'gol', 'selecao', 'neymar', 'jogo', 'jesus', 'galvao', 'gabriel', 'hoje', 'dinamarca' e 'time', o rótulo atribuído para o mesmo foi futebol.

O processo de rotulação manual utilizado foi o mesmo para todos os tópicos gerados a partir dos nove arquivos de entrada que constituem a base de dados, conforme descrito na Tabela 5.1.

Após a finalização do experimento, tendo em vista a realização da limpeza dos dados e a rotulação manual dos tópicos, obteve-se como resultado nove tabelas, contendo quinze tópicos cada, que referenciam o evento das Olimpíadas ocorrido entre os dias 05 a 21 de agosto de 2016 e também os três dias que antecederam e sucederam o evento sediado na cidade do Rio de Janeiro. Todas as nove tabelas estão dispostas no Apêndice A e todos os links das notícias utilizadas no processo de rotulação estão dispostos no Apêndice B.

Rótulo	Palavras
Tópico 01: Vôlei	jogo, esporte, volei, melhor, futebol, mulher, assistir, galvao, globo, vendo.
Tópico 02: Transmissão das Olimpíadas	jogo, hoje, assistir, vendo, casa, assistindo, ficar, vamos, tv, tempo.
Tópico 03: Ginástica	atleta, brasileiro, esporte, medalha, pais, mundo, futebol, jogo, torcida, copa.
Tópico 04: Futebol	gol, selecao, neymar, jogo, jesus, galvao, gabriel, hoje, dinamarca, time.
Tópico 05: Futebol	masculino, feminino, jogo, selecao, futebol, volei, hoje, argentina, basquete, brasileira.

Tabela 6.3: Tópicos obtidos a partir do *Arquivo 4* (Tabela 5.1).

As nove tabelas apresentadas a seguir, iniciando pela Tabela 6.4 e finalizando pela Tabela 6.12, mostram as palavras contidas nos tópicos obtidos, agrupadas por seus respectivos rótulos e excluindo palavras que apareceram repetidas.

A Tabela 6.4 apresenta as palavras contidas nos quinze tópicos gerados, a partir dos *tweets* coletados dos dias 02 a 04 de agosto de 2016, ou seja, o período que antecedeu o início dos Jogos Olímpicos. Pode-se observar que o assunto referente à cerimônia de abertura dos Jogos foi predominante quanto aos outros assuntos que aparecem na tabela. Já os tópicos sobre futebol fazem referência aos jogos de futebol masculino e feminino, da primeira fase, que ocorreram antes do início oficial das Olimpíadas, entre os dias 03 e 04 de agosto de 2016.

As palavras rotuladas como *Pokémon Go*, apesar de não fazerem parte do assunto das Olimpíadas, representam um assunto muito comentado no microblog *Twitter* e fazem referência ao lançamento oficial do jogo no Brasil, em 03 de agosto de 2016, bem como as palavras referentes ao filme Esquadrão Suicida, lançado oficialmente em 04 de agosto de 2016 no País.

Rótulo	Palavras
Cerimônia de Abertura	jogo, copa, mundo, futebol, pais, selecao, atleta, melhor, vamos, brasileiro, seguranca, coi, comite, crise, legado, saude, dinheiro, tocha, olimpica, temer, dilma, cidade, janeiro, gisele, cerimonia, golpe, passagem, olimpico, vaia, organizacao, estadio, engenhao, especial, hoje, phelps, vila, eua, tenis, russos, primeira, medalha, ouro, portugal, ganhar, modalidade, argentina, esporte, video, musica, anitta, fifth, harmony, clipe, cantar, gostei, katy, hino, pele, pira, acender, solo, zika, nadal, hope, torcida, ole.
Futebol	jogo, hoje, amanha, causa, casa, feriado, sexta, boa, bom, ficar, futebol, feminino, selecao, sul, africa, estreia, gol, masculino, china, meninas, surf, brasileira, ouro, esporte, gol, neymar, marta, bola, jesus, cristiane, africa, gabriel.
Ingresso para os Jogos	ingresso, jogo, bomba, assistir, comprar, ganhei, reais, estadio, barra, mae.
Mídia	jogo, globo, canais, sportv, tv, record, cobertura, esporte, band, vivo.
Pokémon Go	pokemon, jogo, povo, brasileiro, falar, mundo, causa, falando, esquadrao, suicida.

Tabela 6.4: Palavras contidas nos tópicos obtidos a partir do *Arquivo 1* (Tabela 5.1) agrupadas por seus respectivos rótulos.

A Tabela 6.5 apresenta as palavras dos tópicos relacionados à cerimônia de abertura dos Jogos Olímpicos, que aconteceu no dia 05 de agosto de 2016. Pode-se observar que muitas das palavras fazem referência a momentos ocorridos no evento, por exemplo, palavras como 'anitta', 'caetano', 'gil', 'cantando', 'cantar', 'mc', 'fernanda', 'karol', 'montenegro', 'gilberto', mencionam as seguintes apresentações: Anitta, Caetano Veloso e Gilberto Gil cantando músicas de Ary Barroso e João Gilberto, Karol Conka e Mc Soffia cantando rap e Fernanda Montenegro declamando trechos do poema "A Flor e a Náusea", de Carlos Drummond de Andrade. Já as palavras 'pira', 'olimpica', 'acender', 'vanderlei', 'tocha', 'guga', 'cordeiro', 'lima' e 'chama', citam o momento em que a chama olímpica aparece no Maracanã nas mãos do extenista Guga Kuerten, e em seguida, o ex-maratonista Vanderlei Cordeiro de Lima a recebe, acendendo a pira e oficializando o início dos Jogos Olímpicos.

Rótulo	**Palavras**
Cerimônia de Abertura	pais, lindo, copa, mundo, brasileiro, povo, dinheiro, bonito, mal, festa, deus, gisele, demais, cerimonia, orgulho, maravilhosa, amo, parabens, pais, melhor, hoje, momento, atleta, anitta, caetano, gil, cantando, cantar, mc, fernanda, karol, montenegro, gilberto, jogo, medalha, ouro, futebol, esporte, selecao, olimpico, delegacao, bandeira, alemanha, mulher, vem, porta, grecia, aquecimento, indios, portugueses, global, parte, mostrar, historia, hino, musica, paulinho, viola, nacional, zeca, flamengo, pira, olimpica, acender, vanderlei, tocha, guga, cordeiro, pele, lima, chama, maracana, janeiro, cidade, estado, paises, samba, aula, carnaval, geografia, escola, falar, regina, ingles, case, falando, portugues, tremendo, santos, fala, homem.
Mídia	galvao, boca, cala, gloria, maria, globo, ouvir, bueno, falando, falar.
Político	temer, vaia, dilma, presidente, vaiado, golpista, michel, lula, medo, povo.
Transmissão das Olimpíadas	pokemon, vendo, hoje, cerimonia, mundo, casa, assistir, assistindo, estar, tv.

Tabela 6.5: Palavras contidas nos tópicos obtidos a partir do *Arquivo 2* (Tabela 5.1) agrupadas por seus respectivos rótulos.

A Tabela 6.6 representa os tópicos obtidos dos *tweets* extraídos entre os dias 06 a 08 de agosto de 2016, ou seja, os três primeiros dias dos Jogos Olímpicos. Apesar do assunto cerimônia de abertura ainda estar em voga, é possível observar que os dados contidos na Tabela 6.6 já abrangem eventos ligados a modalidades esportivas, jogos ocorridos e conquista de medalhas.

Tomando como exemplo as palavras rotuladas como *Futebol*, observa-se que as mesmas se referem a dois eventos: o jogo feminino entre Brasil e Suécia ocorrido no primeiro dia de competições, no qual a jogadora Marta destacou-se por ter marcado os três gols da partida, garantindo a vitória contra a seleção sueca e também o jogo ocorrido no segundo dia de competições, entre Brasil e Iraque pela seleção masculina de futebol.

Já as palavras rotuladas como *Judô* fazem referência ao terceiro dia de competições, com a vitória da judoca Rafaela Silva contra o mongol Sumiya Dorjsuren, conquistando a primeira

medalha de ouro para o Brasil.

Rótulo	**Palavras**
Cerimônia de Abertura	pais, atleta, brasileiro, esporte, copa, mundo, futebol, pessoa, povo, dinheiro, lindo, orgulho, cerimonia, parabens, melhor, festa, hoje, pira, anitta, olimpica, gisele, vanderlei, tocha, musica, acender.
Futebol	selecao, gol, neymar, jogo, futebol, iraque, marta, galvao, masculino, time.
Ginástica	ginastica, lindo, artistica, flavia, solo, deus, atleta, hypolito, diego, rebeca, perna, video, frances, ginasta, gostei, quebrou, ciclista, prova, fratura.
Judô	ouro, medalha, rafaela, silva, primeira, parabens, mulher, judo, primeiro, tiro, final, sarah, jogo, masculino, vence, tenis, feminino, vamos, mesa.
Mídia	globo, jogo, tv, canais, esporte, assistir, sportv, galvao, melhor, canal.
Transmissão das Olimpíadas	jogo, vendo, assistir, hoje, assistindo, esporte, casa, pokemon, ficar, tv.
Político	temer, dilma, vaia, lei, lula, jogo, presidente, protestos, manifestacoes, vaiado.
Vôlei	feminino, selecao, futebol, masculino, jogo, volei, meninas, brasileira, mulher, handebol, set, djokovic, dupla, tenis, praia, vitoria, bruno, estreia.
Parque Olímpico	jogo, olimpico, parque, janeiro, arena, ingresso, hoje, veja, atleta, comida.

Tabela 6.6: Palavras contidas nos tópicos obtidos a partir do *Arquivo 3* (Tabela 5.1) agrupadas por seus respectivos rótulos.

Entre os primeiros três dias dos Jogos Olímpicos Rio 2016, ocorreram mais de 160 disputas entre mais de 20 modalidades diversas. Até então o Brasil conquistava apenas duas medalhas, sendo que a primeira foi de prata e garantida pelo atirador esportivo Felipe Wu no primeiro dia de competições. Já no terceiro dia, a judoca Rafaela Silva conquistou a primeira medalha dourada para o Brasil.

A Tabela 6.7 apresenta os dados obtidos nos três dias seguintes, ou seja, 09 de agosto de 2016 a 11 de agosto de 2016. Nesse período, dentre os assuntos que se destacaram, pode-se citar a conquista de Phelps, que entrou para a história com a sua 26ª medalha em Jogos Olímpicos (22 medalhas de ouro) e a quarta medalha dourada no Rio de Janeiro. Segundo dados informados pelo *Twitter*, Michael Phelps foi o atleta mais comentado dentre os 187 milhões de *tweets* postados no período das Olimpíadas fazendo com que também a modalidade natação fosse a mais citada[14]. Pode-se observar que as palavras rotuladas como *Natação* representam a conquista do nadador norte-americano.

Dentre as palavras rotuladas como *Judô* na Tabela 6.7, está o nome da atleta Mayra Aguiar, fazendo referência à disputa que fez com que a judoca conquistasse a terceira medalha para o Brasil nos Jogos Olímpicos. O nome da judoca Rafaela Silva ainda é citado devido à conquista da medalha dourada no dia 07 de agosto de 2016.

Na Tabela 6.8, constam as palavras referentes aos tópicos obtidos tomando como fonte os *tweets* extraídos do período de 12 de agosto de 2016 a 14 de agosto de 2016. No período de

[14] https://www.techtudo.com.br/noticias/noticia/2016/08/rio-2016-phelps-bolt-neymar-biles-e-lochte-fecham-top-5-do-twitter.html

Rótulo	Palavras
Cerimônia de Abertura	musica, casamento, brasileira, hino, jogo, atleta, historia, pedido, gay, mundo.
Futebol	gol, selecao, neymar, jogo, jesus, galvao, gabriel, hoje, dinamarca, time, masculino, feminino, futebol, volei, argentina, basquete, brasileira.
Ginástica	ginastica, lindo, atleta, brasileiro, esporte, medalha, pais, mundo, futebol, jogo, torcida, copa, ginastica, jade, rebeca, biles, simone, artistica, lindo, solo, final, nota.
Judô	silva, medalha, mayra, mariana, judo, bronze, aguiar, ouro, rafaela, luta.
Natação	phelps, medalha, ouro, michael, thiago, pereira, historia, ganhar, natacao, final, onibus, piscina, jogo, verde, agua, ingresso, seguranca, comite, jornalistas, nacional, atleta, lugar, final, prova, confira, hoje, veja, oliveira, brasileiro.
Transmissão das Olimpíadas	jogo, hoje, assistir, vendo, casa, assistindo, ficar, vamos, tv, tempo.
Parque Olímpico	olimpico, video, jogo, janeiro, olimpica, gostei, parque, clima, hoje, arena.
Político	manifestacoes, protestos, juiz, decisao, temer, politicas, repressao, federal, libera, proibe.
Tênis	final, quartas, bellucci, tenis, masculino, nadal, brasileiro, marcelo, vence, bruno.
Vôlei	jogo, esporte, volei, melhor, futebol, mulher, assistir, galvao, globo, vendo, set, vamos, pedro, evandro, praia, basquete, vitoria, primeiro.

Tabela 6.7: Palavras contidas nos tópicos obtidos a partir do *Arquivo 4* (Tabela 5.1) agrupadas por seus respectivos rótulos.

dias citados houveram conquistas de três medalhas para o Brasil nas modalidades ginástica e judô, sendo duas de bronze e uma de prata.

Tomando as palavras rotuladas como *Ginástica*, é possível identificar os nomes dos ginastas Diego Hypólito e Arthur Nory e também as palavras 'medalha', 'prata' e 'bronze' que referenciam a conquista dos brasileiros no dia 14 de agosto de 2016. Já sobre a conquista da medalha de bronze do judoca Rafael "Baby" Silva, não há um tópico específico que referencie a conquista. Ainda é possível observar que há palavras referenciando os eventos ocorridos no período em várias outras modalidades tais como Atletismo, Boxe, Futebol, Tênis, Natação e Vôlei.

Rótulo	Palavras
Atletismo	solo, hope, zika, torcida, eua, mundo, brasileiro, mae, bolt, casa, usain, final, prova, phelps, olimpico, recorde, salto, metros, atletismo.
Boxe	olimpico, ouro, janeiro, estadio, marcha, parque, tiro, atletica, engenhao, hoje.
Cerimônia de Abertura	video, atleta, jogo, veja, gostei, brasileiro, confira, olimpica, pokemon, doping, piscina, agua, comite, temer, verde, seguranca, nacional, forca, justica.
Futebol	atleta, esporte, brasileiro, medalha, pais, futebol, mundo, jogo, melhor, copa, meninas, feminino, selecao, vamos, coracao, barbara, volei, penalti, neymar, gol, colombia, bola, time, falta, luan.
Ginástica	diego, nory, parabens, hypolito, medalha, cai, arthur, ginastica, prata, bronze.
Mídia	melhor, jogo, globo, guga, galvao, hino, volei, espn, comentarista, narrador.
Natação	medalha, ouro, phelps, michael, historia, simone, ganhar, biles, ganhou, natacao.
Transmissão das Olimpíadas	jogo, hoje, assistir, vendo, assistindo, futebol, volei, esporte, tv, casa.
Tênis	set, jogo, nadal, bellucci, vamos, argentina, primeiro, tempo, bola, segundo, rafael, medalha, silva, ouro, bronze, final, robson, murray, delpotro.
Vôlei	feminino, masculino, volei, futebol, jogo, final, selecao, praia, quartas, alemanha.

Tabela 6.8: Palavras contidas nos tópicos obtidos a partir do *Arquivo 5* (Tabela 5.1) agrupadas por seus respectivos rótulos.

Após passado quase dez dias do início dos Jogos Olímpicos no Brasil, mais de 470 competições já haviam ocorrido entre as mais de trinta modalidades esportivas. No período subsequente, havia a expectativa da conquista de mais medalhas conforme os atletas brasileiros classificavam-se para as disputas finais.

Na Tabela 6.9 são apresentadas as palavras que representam os tópicos gerados dos *tweets* coletados dos dias 15, 16 e 17 de agosto de 2016. Nesse período houve importantes conquistas de medalhas para o Brasil, destacando-se o boxeador brasileiro Robson Conceição e o atleta Thiago Braz, especializado no salto com vara.

As palavras rotuladas como *Atletismo* na Tabela referenciam o ouro conquistado para o Brasil pelo atleta Thiago Braz, após atingir a marca de 6.03m no salto com vara. Thiago, além de ser recordista olímpico, é um dos nove atletas no mundo que saltaram acima dos seis metros de altura. As palavras rotuladas como *Boxe* citam Robson Conceição que foi destaque no décimo primeiro dia de competição ao conquistar sua primeira medalha olímpica. O boxeador, que ao conquistar a primeira medalha de ouro do boxe brasileiro nos Jogos Olímpicos, colocou seu nome na história do País. Já as palavras rotuladas como *Ginástica* e *Vôlei*, fazem referência às conquistas de medalhas de prata pelos atletas Arthur Zanetti e a dupla Ághata e Bárbara, respectivamente.

Rótulo	**Palavras**
Atletismo	thiago, ouro, braz, vara, salto, olimpico, medalha, recorde, silva, parabens final, bolt, usain, metros, prova, fernando, semifinal, cleubercunha, pimenta.
Boxe	ouro, medalha, robson, conceicao, isaquias, boxe, prata, parabens, queiroz, brasileiro.
Cerimônia de Abertura	video, jogo, olimpico, olimpica, janeiro, gostei, hoje, pokemon, veja, arena.
Ciclismo	olimpico, camera, parque, pessoa, cai, acidente, caiu, hoje, estadio, duas.
Futebol	frances, brasileiro, torcida, mundo, copa, vaia, pais, atleta, ouro, futebol, gol, neymar, selecao, jogo, honduras, jesus, suecia, bola, final, gabriel.
Ginástica	atleta, medalha, esporte, brasileiro, pais, ouro, jogo, militares, outros, pessoa, zanetti, flavinha, bronze, prata, arthur, poliana, biles, parabens.
Mídia	melhor, globo, jogo, galvao, guga, musica, narrador, tv, espn, volei.
Político	temer, justica, comite, ingresso, jogo, coi, lula, policia, presidente, dilma.
Vôlei	jogo, hoje, assistir, vendo, volei, ficar, acabar, assistindo, esporte, tv, feminino, meninas, futebol, selecao, ouro, medalha, parabens, mulher, final, masculino, praia, franca, set, vamos, coracao, deus, meninas, china, penalti.

Tabela 6.9: Palavras contidas nos tópicos obtidos a partir do *Arquivo 6* (Tabela 5.1) agrupadas por seus respectivos rótulos.

A Tabela 6.10 apresenta as palavras contidas nos quinze tópicos gerados a partir dos *tweets* coletados do período de 18 a 20 de Agosto de 2016. Nesse período foram conquistadas 7 medalhas para o Brasil nos Jogos Olímpicos, totalizando um número de 18 medalhas para o país. Pode-se observar na Tabela 6.10 que a modalidade *Futebol* tem destaque entre as outras

modalidades.

No primeiro dia do período descrito e décimo terceiro dia de disputa da Rio 2016, ou seja, 18 de agosto de 2016, o atleta brasileiro Isaquias Queiroz garantiu a medalha de bronze para o Brasil na canoagem velocidade. É possível observar que as palavras rotuladas como *Canoagem* na Tabela fazem referência à conquista do canoísta. Ainda neste dia, as atletas brasileiras Martine Grael e Kahena Kunze garantiram medalha de ouro na vela no qual as palavras rotuladas como *Vela* referem-se a conquista. Já na modalidade vôlei de praia, a dupla brasileira Alison e Bruno Schmidt garantiram o ouro para o País. Palavras como 'jogo', 'final', 'ouro', 'masculino', 'bruno' e 'praia', rotuladas como *Vôlei* fazem menção a vitória dos brasileiros contra o time italiano por 2 sets a 0.

No terceiro dia do período descrito e penúltimo dia de disputa da Rio 2016, ou seja, 20 de agosto de 2016, o Brasil conquistou mais quatro medalhas, sendo uma de ouro no Futebol, duas de prata na Canoagem e uma de bronze no Taekwondo, subindo algumas posições no quadro geral de medalhas das Olimpíadas.

Citando as palavras rotuladas como *Futebol* da Tabela 6.10, observa-se que fazem referência a vitória da seleção brasileira contra a seleção da Alemanha. O Brasil venceu a partida nos pênaltis por 5 a 4 e segundo dados do *Twitter*, o jogador brasileiro Neymar foi o segundo atleta mais mencionado nos *tweets* postados no período das Olimpíadas. Ainda segundo o microblog, o *Top Momento* mais comentado foi o gol decisivo marcado pelo Neymar na partida[15].

Rótulo	**Palavras**
Atletismo	bolt, usain, prova, ouro, final, medalha, olimpico, lugar, phelps, atleta.
Cerimônia de Abertura	olimpico, jogo, video, olimpica, final, janeiro, hoje, atleta, gostei, parque.
Canoagem	medalha, isaquias, ouro, queiroz, brasileiro, bronze, prata, historia, parabens, canoagem.
Futebol	medalha, atleta, futebol, esporte, brasileiro, melhor, ouro, parabens, jogo, pais, hino, alemanha, galvao, torcida, lindo, neymar, mundo, copa, selecao, campeao, ganhar, final, gol, vamos, penalti, bola, tempo, deus, coracao, jesus, gabriel, luan, weverton.
Mídia	globo, waack, cris, melhor, william, galvao, marcha, atletica, jornal, guga.
Nado Sincronizado	ginastica, lindo, musica, ritmica, melhor, amo, jogo, mascote, nado, sincronizado.
Natação	nadadores, americanos, lochte, caso, assalto, ryan, desculpas, eua, atleta, brasileiro.
Político	temer, dilma, lula, jogo, ingresso, governo, coi, mil, atleta, dinheiro.
Vela	ouro, grael, martine, kahena, vela, kunze, medalha, parabens, familia, olimpico.
Vôlei	jogo, hoje, final, acabar, assistir, amanha, vamos, volei, fim, acabando, ouro, masculino, feminino, bruno, praia, parabens, selecao.

Tabela 6.10: Palavras contidas nos tópicos obtidos a partir do *Arquivo 7* (Tabela 5.1) agrupadas por seus respectivos rótulos.

[15] https://www.techtudo.com.br/noticias/noticia/2016/08/rio-2016-phelps-bolt-neymar-biles-e-lochte-fecham-top-5-do-twitter.html

A Tabela 6.11 apresenta os dados referentes ao último dia de competições da Rio 2016 que ocorreu em 21 de agosto de 2016. Neste dia, apesar de ter ocorrido várias disputas finais, o assunto que mais se destacou dentre os tópicos obtidos foi o evento da cerimônia de encerramento dos Jogos Olímpicos. Pode-se observar que as palavras rotuladas como *Cerimônia de Encerramento* fazem menção aos eventos ocorridos naquela noite, como por exemplo, 'mario', 'ministro', 'primeiro', 'super', 'vestido', 'apresentacao', 'cano', 'tokyo', que fazem referência a aparição do primeiro-ministro japonês Shinzo Abe vestido de Super de Mario para apresentar o país que sediará as Olimpíadas de 2020. Outro assunto que também se destacou, foi a possibilidade de o presidente do Comitê Olímpico Brasileiro, Carlos Arthur Nuzman, estar sofrendo do Mal de Parkinson, após estar tremendo muito suas mãos no discurso de encerramento da Rio 2016. As palavras rotuladas como *Nuzman (Parkinson)* fazem menção ao fato.

Rótulo	**Palavras**
Cerimônia de Abertura	santos, dinheiro, maracana, atleta, dumont, cerimonia, ingresso, estadio, novo, chuva.
Cerimônia de Encerramento	saudades, lindo, sentir, jogo, crush, falta, chorando, saudade, acabou, coracao, mundo, melhor, copa, pais, brasileiro, festa, povo, japao, orgulho, atleta, carnaval, gringos, samba, tenis, gringo, maracana, fim, amanha, triste, hoje, volta, jogo, acaba, vamos, ouro, volei, medalha, masculino, italia, selecao, futebol, wallace, parabens, obrigado, esporte, tokyo, proxima, novo, daqui, estar, vem, hino, criancas, musica, cantando, cerimonia, nacional, bandeira, nordeste, frevo, olimpico, maratona, bolt, esporte, isaquias, musica, cantar, tocar, katy, esperando, ivete, funk, rise, pira, olimpica, chama, apagar, embora, chuva, deixa, tocha, fogo, deus, mario, ministro, primeiro, super, pokemon, vestido, apresentacao, cano.
Mídia	galvao, temer, globo, dilma, lula, gloria, maria, boca, paes, eduardo.
Nuzman (Parkinson)	tremendo, homem, ingles, nuzman, parkinson, falando, mal, bandeira, red, desse.

Tabela 6.11: Palavras contidas nos tópicos obtidos a partir do *Arquivo 8* (Tabela 5.1) agrupadas por seus respectivos rótulos.

Por fim, a Tabela 6.12 apresenta as palavras referentes ao período que sucederam o evento das Olimpíadas, ou seja, 22 de agosto de 2016 a 24 de agosto de 2016. Apesar do assunto sobre a cerimônia de encerramento ainda se destacar, assuntos referente a medalha conquistada na modalidade futebol pela seleção brasileira também ainda estão sendo comentados.

De acordo com as tabelas apresentadas, observa-se que os tópicos obtidos abrangeram diversos eventos ocorridos nos Jogos Olímpicos Rio 2016, principalmente nos momentos em que os atletas brasileiros conquistaram medalhas para o País.

A Figura 6.2 representa um mapa de calor com a quantidade de tópicos por período, conforme tabelas apresentadas no Apêndice A e tomando como base para agrupamento seus rótulos. Para a montagem do mapa, inicialmente foi realizado a identificação da quantidade de tópicos do mesmo rótulo por tabela e verificado a representação em porcentagem da quantidade identificada, tomando como 100% a quantidade total de tópicos, ou seja, 15. Após a identifi-

Rótulo	Palavras
Atletismo	video, gostei, bolt, neymar, playlist, adicionei, ouro, usain, bruna, final.
Cerimônia de Encerramento	tokyo, proxima, japao, acabou, copa, mundo, dinheiro, pais, vamos, falando, fim, volta, acabaram, hoje, voltar, bom, vida, semana tv, saudades, jogo, triste, saudade, assistir, lindo, parabens, melhor, brasileiro, festa, obrigado, atleta, paraolimpiadas, esporte, paralimpicos, veja, ingresso, milhoes, mil, turistas, legado, olimpico, dinheiro, venda, janeiro, olimpica, brasileira, historia, fotos, robotica, nacional, facebook, mario, super, ministro, japones, primeiro, pokemon, cerimonia, medalha, prata, medalhista, ajudar, cancer, etiope, jogo, contas, leiloa.
Futebol	ouro, futebol, medalha, volei, selecao, final, jogo, masculino, feminino, neymar, copa, mundo, melhor, pais, atleta, historia, quadro, brasileiro, luan, lochte, tite, jogou, gabigol, ryan, jesus, taison, convocacao.
Político	lula, dilma, sucesso, temer, brasileiro, pais, mundo, povo, atleta, copa.
Vôlei	fim, musica, agatha, barbara, neymar, ultima, midia, volta, diario, cores.

Tabela 6.12: Palavras contidas nos tópicos obtidos a partir do *Arquivo 9* (Tabela 5.1) agrupadas por seus respectivos rótulos.

cação dos parâmetros necessários e obtida a porcentagem correspondente de cada tópico por tabela, foi utilizada essa porcentagem para definir a transparência da cor usada no mapa. Os dados apresentados na vertical do mapa de calor, referem-se as fatias de dias conforme Tabela 5.1, e na horizontal, os rótulos dos tópicos.

Tomando como exemplo a primeira linha do mapa de calor, que se refere aos tópicos obtidos através dos *tweets* coletados dos dias 02 a 04 de agosto de 2016, pode-se observar que tópicos sobre o assunto *Cerimônia de Abertura* foram mais proeminentes quanto aos demais, ou seja, a porcentagem de representação do assunto foi maior tornando a cor mais intensa. Apesar do evento de abertura ter ocorrido no dia 05 de agosto de 2016, é possível observar que o assunto continuou em voga por praticamente todo o período dos Jogos, exceto na última fatia de dias referente ao período pós Olimpíadas.

Tópicos rotulados como mídia e político também estão presente em quase todas as fatias de dias. Os tópicos rotulados como mídia fazem menção tanto aos narradores das partidas que ocorreram quanto aos canais de televisão que fizeram as transmissões dos Jogos Olímpicos. Já os tópicos sobre política, apesar de não estarem diretamente ligados ao assunto dos Jogos, fazem menção ao momento político que o país estava enfrentando no ano de 2016.

Na Figura 6.3 é apresentado o gráfico com a quantidade de tópicos, agrupados por seus rótulos, dispostos nos 135 tópicos obtidos (15 tópicos x 9 tabelas). Os assuntos que mais se destacaram e foram mencionados: *Cerimônia de Abertura* com quase 30 tópicos, *Cerimônia de Encerramento* com mais de 20 tópicos e *Futebol* com 20 tópicos.

As Figuras 6.4, 6.5 e 6.6 são nuvens de palavras representando os tópicos, agrupados por seus rótulos, que tiveram maior quantidade conforme apresentado na Figura 6.3. Ou seja,

Figura 6.2: Mapa de calor com a quantidade de tópicos por fatia de dias.

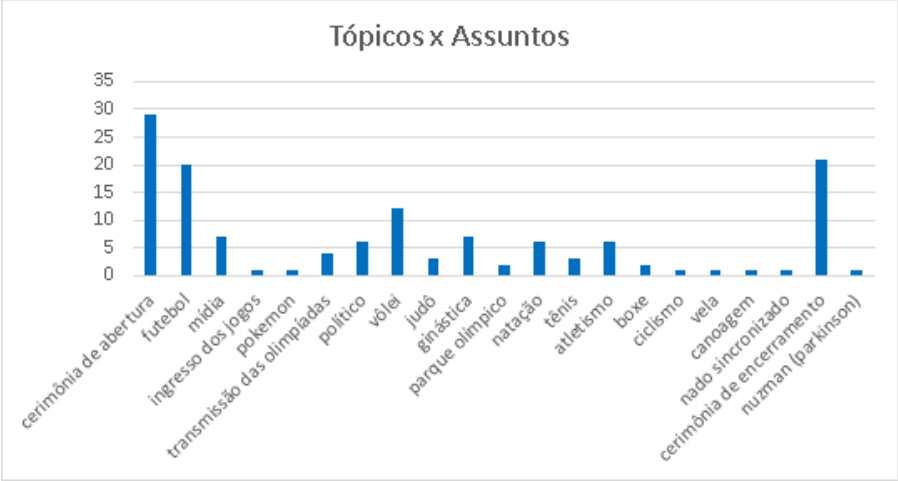

Figura 6.3: Gráfico da quantidade de tópicos por assunto

tópicos sobre a *Cerimônia de Abertura*, *Futebol* e a *Cerimônia de Encerramento*.

Para a montagem da nuvem de palavras, foi realizada a multiplicação da probabilidade de ocorrência da palavra no tópico por mil e utilizado esse valor como parâmetro para repetição da palavra em um arquivo texto. Para a criação da imagem foi utilizado uma ferramenta web, chamada *Wordclouds.com*[16].

Na Figura 6.4 são apresentadas as palavras que representam todos os tópicos rotulados como *Cerimônia de Abertura.* Pode-se observar que palavras como 'lindo', 'jogo', 'olimpica', 'atleta', 'pais', se destacam entre as demais e ocorreram mais frequentemente nos tópicos obti-

[16] www.wordclouds.com

dos.

Figura 6.4: Nuvem de palavras das palavras rotuladas como *Cerimônia de Abertura*.

Na Figura 6.5 são apresentadas as palavras que fazem parte dos tópicos sobre o assunto *Futebol*. Palavras como 'selecao', 'ouro', 'jogo', 'futebol', 'feminino', e 'gol' foram as palavras mais proeminentes nos tópicos obtidos.

Figura 6.5: Nuvem de palavras das palavras rotuladas como *Futebol*.

Por fim, na Figura 6.6 são mostradas as palavras que representam todos os tópicos rotulados como *Cerimônia de Encerramento*. Palavras como 'jogo', 'japao', 'tokyo', 'mario', 'acabou', 'atleta', se destacaram entre as demais e ocorreram com mais frequência nos tópicos obtidos.

Baseando-se nos resultados apresentados e suas análises, constata-se que os tópicos obtidos refletiram diversos acontecimentos ocorridos nos Jogos Olímpicos. Embora os tópicos não compreendam todas as disputas que aconteceram, considerando que foram mais de 600, é possível afirmar que a partir dos *tweets* extraídos e da escolha do número de tópicos $K = 15$, o resultado obtido foi satisfatório, apresentando um extrato dos assuntos comentados no microblog *Twitter*, acerca do evento multiesportivo realizado no segundo semestre de 2016, na cidade do Rio de Janeiro.

Figura 6.6: Nuvem de palavras das palavras rotuladas como *Cerimônia de Encerramento*.

7 CONCLUSÃO

Neste trabalho de conclusão de curso foi realizada uma análise exploratória na base de dados do microblog *Twitter*, selecionando apenas *tweets* ocorridos durante o evento das Olimpíadas Rio 2016. Para viabilizar a execução do trabalho, foram realizadas etapas de pré-processamento, com a limpeza dos dados, aplicação do algoritmo de modelagem de tópicos BTM para a extração dos tópicos, e pós-processamento, com a definição da quantidade de tópicos por período que seriam utilizados no experimento.

Para avaliação dos tópicos obtidos, após a realização do experimento, foram analisadas as palavras contidas em cada tópico, reconhecendo as relações entre elas a fim de rotular o conjunto, identificando o tópico e relacionando-o a um determinado assunto. Assim, foram observados os tópicos mais recorrentes no período especificado, ligados principalmente a eventos ocorridos durante as Olimpíadas como as cerimônias de abertura e encerramento e também jogos e disputas ocorridos dentre as diversas modalidades.

Além da análise sobre as palavras e a rotulação dos tópicos, foram geradas nove tabelas, dispostas no Apêndice A, contendo 15 tópicos cada, conforme definido anteriormente, que representam um extrato dos assuntos comentados no microblog *Twitter* no período das Olimpíadas.

Através do estudo das tabelas, gráficos e nuvens de palavras geradas, constatou-se que os tópicos obtidos refletiram os acontecimentos decorridos nos Jogos Olímpicos, principalmente aos que fizeram referência ao Brasil.

7.1 Trabalhos Futuros

A fim de abranger as análises realizadas neste trabalho, algumas alternativas podem ser utilizadas durante todo o processo. Uma delas é a execução do algoritmo BTM e a extração dos tópicos utilizando apenas um arquivo de entrada contendo toda a base de dados para identifica os possíveis tópicos gerados e comparar com a abordagem atual. Outra alternativa, é a utilização de diferentes palavras como parâmetro para a extração dos *tweets*, bem como a ampliação do processo de limpeza dos dados. Para aprofundar as análises, pode ser feita a utilização de algoritmos para a rotulação dos tópicos tomando como base de dados palavras que compreendam assuntos esportivos. Ainda, é possível a utilização de um conjunto de tópicos maior e até mesmo tópicos que contenham mais palavras.

REFERÊNCIAS

[1] D. M. Blei. Probabilistic Topic Models. *Communications of the ACM*, 55(4):77–84, 2012.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] S. Fukuyama and K. Wakabayashi. Extracting time series variation of topic popularity in microblogs. In *Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services - iiWAS2018*, pages 365–369, New York, New York, USA, 2018. ACM Press.

[4] T. Hofmann. Probabilistic Latent Semantic Indexing. *ACM SIGIR Forum*, 51(2):211–218, 2017.

[5] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary. Twitter trending topic classification. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 251–258, 2011.

[6] C. Ligutom, J. V. Orio, D. A. M. Ramacho, C. Montenegro, R. E. Roxas, and N. Oco. Using Topic Modelling to make sense of typhoon-related tweets. *Proceedings of the 2016 International Conference on Asian Language Processing, IALP 2016*, pages 362–365, 2017.

[7] R. Millington and S. C. Darnell. Constructing and contesting the Olympics online: The internet, Rio 2016 and the politics of Brazilian development. *International Review for the Sociology of Sport*, 49(2):190–210, apr 2014.

[8] L. S. Oliveira, P. O. S. V. de Melo, M. S. Amaral, and J. A. G. Pinho. When Politicians Talk About Politics: Identifying Political Tweets of Brazilian Congressmen. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, pages 664–667, 2018.

[9] M. Steyvers and T. Griffiths. Probalistic Topic Models. *Handbook of latent semantic analysis*, 427(7):424–440, 2007.

[10] X. Yan, J. Guo, Y. Lan, and X. Cheng. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, pages 1445–1456. ACM Press, 2013.

APÊNDICES

APÊNDICE A – Tópicos por Fatia de Dias

Tópico	P(z)	Palavras	Rótulo
01	0,165237	jogo:0,016942, copa:0,012868, mundo:0,01157, futebol:0,010352, pais:0,007221, selecao:0,006392, atleta:0,005826, melhor:0,005542, vamos:0,005318, brasileiro:0,00514.	cerimônia de abertura
02	0,102747	jogo:0,022707, hoje:0,016948, amanha:0,014056, causa:0,010454, casa:0,009257, feriado:0,008085, sexta:0,007535, boa:0,007005, bom:0,006881, ficar:0,006033.	futebol
03	0,099167	futebol:0,053966, jogo:0,045826, feminino:0,044769, selecao:0,028135, sul:0,027293, africa:0,024895, estreia:0,022436, gol:0,022329, masculino:0,021225, china:0,016999.	futebol
04	0,090619	pokemon:0,069382, jogo:0,007753, povo:0,007522, brasileiro:0,006518, falar:0,005494, mundo:0,005294, causa:0,005134, falando:0,004855, esquadrao:0,004683, suicida:0,00462.	pokémon go
05	0,079602	futebol:0,045032, selecao:0,043349, feminino:0,040604, masculino:0,019434, meninas:0,017042, jogo:0,016144, surf:0,013288, brasileira:0,013056, ouro:0,012243, esporte:0,011976.	futebol
06	0,075438	gol:0,035330, neymar:0,020062, selecao:0,017128, marta:0,01596, bola:0,012894, jesus:0,009984, jogo:0,008793, cristiane:0,008736, africa:0,008618, gabriel:0,008533.	futebol
07	0,059095	seguranca:0,012925, jogo:0,012522, pais:0,009794, coi:0,008671, comite:0,006724, atleta:0,006664, crise:0,006219, legado:0,006201, saude:0,006177, dinheiro:0,005895.	cerimônia de abertura
08	0,050649	tocha:0,040072, olimpica:0,025930, temer:0,012355, dilma:0,007786, cidade:0,006875, janeiro:0,006132, gisele:0,005817, cerimonia:0,005691, golpe:0,005123, passagem:0,004779.	cerimônia de abertura
09	0,049550	jogo:0,030571, temer:0,013051, olimpico:0,012929, vaia:0,008209, janeiro:0,007356, organizacao:0,007163, estadio:0,006303, engenhao:0,00626, especial:0,005981, hoje:0,005437.	cerimônia de abertura
10	0,046941	jogo:0,029816, globo:0,027163, canais:0,01642, sportv:0,016027, tv:0,012292, record:0,011673, cobertura:0,01004, esporte:0,009828, band:0,009246, vivo:0,007711.	mídia
11	0,040511	atleta:0,023719, jogo:0,010297, olimpica:0,008519, phelps:0,007565, vila:0,006698, olimpico:0,006278, eua:0,006033, tenis:0,005542, russos:0,005271, primeira:0,005131.	cerimônia de abertura
12	0,039284	medalha:0,030190, ouro:0,027969, portugal:0,012549, atleta:0,012242, jogo:0,008667, ganhar:0,007719, modalidade:0,006916, argentina:0,006347, olimpico:0,006293, esporte:0,005823.	cerimônia de abertura
13	0,037877	ingresso:0,038947, jogo:0,016188, bomba:0,007808, assistir:0,006301, comprar:0,005945, ganhei:0,005383, reais:0,004166, estadio:0,004054, barra:0,003745, mae:0,003717.	ingresso para jogos olimpicos
14	0,037510	video:0,023152, musica:0,016837, anitta:0,013339, fifth:0,011987, harmony:0,010711, clipe:0,009766, cantar:0,008839, gostei:0,008489, katy:0,007119, hino:0,006826.	cerimônia de abertura
15	0,025773	pele:0,022609, pira:0,022032, acender:0,018941, olimpica:0,016634, solo:0,01559, zika:0,015494, nadal:0,015329, hope:0,014711, torcida:0,010961, ole:0,010755.	cerimônia de abertura

Tabela A.1: Tópicos obtidos a partir dos *tweets* extraídos dos dias 02 a 04 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,159879	pokemon:0,013100, vendo:0,010790, hoje:0,009274, cerimonia:0,009052, mundo:0,00866, casa:0,007701, assistir:0,007672, assistindo:0,007258, estar:0,006347, tv:0,006145.	transmissão das olimpíadas
02	0,124873	pais:0,026923, lindo:0,017974, copa:0,017289, mundo:0,01249, brasileiro:0,011829, povo:0,008836, dinheiro:0,008776, bonito:0,007514, mal:0,006931, festa:0,006574.	cerimônia de abertura
03	0,094771	lindo:0,070607, deus:0,018086, gisele:0,017362, demais:0,015383, cerimonia:0,013843, orgulho:0,011245, maravilhosa:0,011106, mundo:0,010144, amo:0,00963, parabens:0,009159.	cerimônia de abertura
04	0,089204	mundo:0,029393, orgulho:0,024327, pais:0,023629, brasileiro:0,021462, lindo:0,017826, melhor:0,014759, hoje:0,012185, momento:0,011305, amo:0,009581, atleta:0,009225.	cerimônia de abertura
05	0,068853	anitta:0,054571, caetano:0,015449, gil:0,014939, cantando:0,013987, cantar:0,011288, mc:0,009323, fernanda:0,0082, karol:0,007988, montenegro:0,00794, gilberto:0,007827.	cerimônia de abertura
06	0,059024	jogo:0,029520, hoje:0,014365, cerimonia:0,009212, atleta:0,009189, medalha:0,008154, ouro:0,007865, futebol:0,007727, esporte:0,005858, selecao:0,005041, olimpico:0,004946.	cerimônia de abertura
07	0,053799	delegacao:0,018858, bandeira:0,018753, atleta:0,015766, alemanha:0,010757, lindo:0,010062, mulher:0,009678, vem:0,00653, pais:0,006394, porta:0,006359, grecia:0,005972.	cerimônia de abertura
08	0,049012	aquecimento:0,010647, indios:0,010599, lindo:0,010509, portugueses:0,010286, global:0,010235, parte:0,00944, atleta:0,008425, mostrar:0,007199, mundo:0,006423, historia:0,005928.	cerimônia de abertura
09	0,047924	galvao:0,075064, boca:0,035239, cala:0,02139, gloria:0,02137, maria:0,016447, globo:0,015879, ouvir:0,01475, bueno:0,014446, falando:0,012739, falar:0,011635.	mídia
10	0,045591	temer:0,092321, vaia:0,038747, dilma:0,018993, presidente:0,016381, vaiado:0,015627, golpista:0,014614, michel:0,010785, lula:0,00887, medo:0,008633, povo:0,008359.	político
11	0,043754	hino:0,035729, cantando:0,020968, musica:0,015859, paulinho:0,014039, viola:0,013857, nacional:0,012877, zeca:0,009728, lindo:0,009688, cantar:0,009402, flamengo:0,009209.	cerimônia de abertura
12	0,043544	pira:0,066917, olimpica:0,054836, acender:0,043894, vanderlei:0,034324, tocha:0,02746, guga:0,025983, cordeiro:0,024629, pele:0,017724, lima:0,017358, chama:0,011089.	cerimônia de abertura
13	0,043366	maracana:0,014572, jogo:0,012152, janeiro:0,007073, cerimonia:0,006709, hoje:0,006561, olimpica:0,006334, atleta:0,005721, cidade:0,00536, estado:0,004931, olimpico:0,004556.	cerimônia de abertura
14	0,040193	paises:0,029980, samba:0,026077, historia:0,016996, aula:0,015673, carnaval:0,014809, pais:0,014681, geografia:0,013132, escola:0,01272, escolas:0,010159, falar:0,008887.	cerimônia de abertura
15	0,036214	regina:0,041894, ingles:0,035047, case:0,031541, falando:0,021163, portugues:0,012697, tremendo:0,011143, falar:0,010702, santos:0,009964, fala:0,009839, homem:0,009748.	cerimônia de abertura

Tabela A.2: Tópicos obtidos a partir dos *tweets* extraídos do dia 05 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,156192	jogo:0,015767, vendo:0,013368, assistir:0,013159, hoje:0,009805, assistindo:0,009203, esporte:0,008822, casa:0,008541, pokemon:0,007889, ficar:0,007721, tv:0,006667.	transmissão das olimpíadas
02	0,120929	pais:0,014216, atleta:0,012542, brasileiro:0,012007, esporte:0,011377, copa:0,010833, mundo:0,010117, futebol:0,008453, pessoa:0,006383, povo:0,006117, dinheiro:0,005907.	cerimônia de abertura
03	0,105978	selecao:0,029382, gol:0,016747, neymar:0,013934, jogo:0,01361, futebol:0,013367, iraque:0,012863, marta:0,011934, galvao:0,009707, masculino:0,009167, time:0,008554.	futebol
04	0,087924	feminino:0,048502, selecao:0,041073, futebol:0,040102, masculino:0,036665, jogo:0,024599, volei:0,02394, meninas:0,019658, brasileira:0,01376, mulher:0,010324, handebol:0,009893.	vôlei
05	0,082672	lindo:0,029588, orgulho:0,017254, mundo:0,015379, cerimonia:0,012705, brasileiro:0,012383, parabens:0,01227, pais:0,011986, melhor:0,01032, festa:0,007477, hoje:0,006352.	cerimônia de abertura
06	0,068391	ouro:0,072684, medalha:0,060327, rafaela:0,052388, silva:0,039973, primeira:0,028417, parabens:0,020792, mulher:0,015491, judo:0,015109, primeiro:0,014611, tiro:0,014342.	judô
07	0,061676	globo:0,019140, jogo:0,016103, tv:0,013035, canais:0,010363, esporte:0,009789, assistir:0,009208, sportv:0,008639, galvao:0,007741, melhor:0,007599, canal:0,007309.	mídia
08	0,055601	ginastica:0,022887, lindo:0,014214, artistica:0,010657, flavia:0,010541, solo:0,009135, deus:0,008267, atleta:0,008051, hypolito:0,00734, diego:0,007078, rebeca:0,006936.	ginástica
09	0,050208	jogo:0,019489, olimpico:0,018522, parque:0,007407, janeiro:0,006865, arena:0,006862, ingresso:0,00642, hoje:0,005905, veja:0,004502, atleta:0,004257, comida:0,004122.	parque olímpico
10	0,044027	final:0,015859, sarah:0,012515, judo:0,012293, jogo:0,011989, masculino:0,011781, vence:0,011737, tenis:0,011491, feminino:0,011252, vamos:0,010257, mesa:0,009778.	judô
11	0,041192	pira:0,018527, anitta:0,016355, olimpica:0,015593, gisele:0,012645, vanderlei:0,012115, cerimonia:0,008877, tocha:0,008399, musica:0,007776, lindo:0,006896, acender:0,006337.	cerimônia de abertura
12	0,038491	temer:0,042443, dilma:0,019465, vaia:0,012498, lei:0,011978, lula:0,009663, jogo:0,007493, presidente:0,007025, protestos:0,006846, manifestacoes:0,006808, vaiado:0,006173.	cerimônia de abertura
13	0,038359	medalha:0,017236, recorde:0,015330, ouro:0,014173, final:0,013135, natacao:0,011362, mundial:0,0107, prova:0,010389, phelps:0,00987, feminino:0,008275, masculino:0,007961.	natação
14	0,032815	set:0,038192, jogo:0,019625, volei:0,017387, djokovic:0,013993, dupla:0,013717, tenis:0,011824, praia:0,011282, vitoria:0,010719, bruno:0,010603, estreia:0,009189.	vôlei
15	0,015544	perna:0,025643, video:0,023724, frances:0,021338, ginasta:0,015143, gostei:0,009621, quebrou:0,008719, atleta:0,00751, ciclista:0,007166, prova:0,007067, fratura:0,006784.	ginástica

Tabela A.3: Tópicos obtidos a partir dos *tweets* extraídos dos dias 06 a 08 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,139616	jogo:0,015011, esporte:0,010608, volei:0,010073, melhor:0,008372, futebol:0,006813, mulher:0,006352, assistir:0,006302, galvao:0,006097, globo:0,005991, vendo:0,005785.	vôlei
02	0,137306	jogo:0,020216, hoje:0,013598, assistir:0,011349, vendo:0,010046, casa:0,009048, assistindo:0,007567, ficar:0,007488, vamos:0,006459, tv:0,006193, tempo:0,005949.	transmissão das olimpíadas
03	0,130762	atleta:0,017530, brasileiro:0,012894, esporte:0,012161, medalha:0,011025, pais:0,009225, mundo:0,006934, futebol:0,006486, jogo:0,005503, torcida:0,004975, copa:0,004821.	ginástica
04	0,085209	gol:0,033675, selecao:0,026198, neymar:0,01603, jogo:0,012634, jesus:0,01157, galvao:0,010609, gabriel:0,009997, hoje:0,008962, dinamarca:0,008805, time:0,008154.	futebol
05	0,083462	masculino:0,039511, feminino:0,037515, jogo:0,036802, selecao:0,032704, futebol:0,02931, volei:0,022133, hoje:0,015274, argentina:0,012533, basquete:0,011861, brasileira:0,011339.	futebol
06	0,079347	phelps:0,051170, medalha:0,050969, ouro:0,043319, michael:0,019336, thiago:0,018402, pereira:0,011028, historia:0,010239, ganhar:0,008932, natacao:0,008276, final:0,007185.	natação
07	0,050561	silva:0,038189, medalha:0,031965, mayra:0,030627, mariana:0,02729, judo:0,023994, bronze:0,023958, aguiar:0,021103, ouro:0,020824, rafaela:0,015963, luta:0,015016.	judo
08	0,047840	ginastica:0,023219, jade:0,019586, rebeca:0,019381, biles:0,013265, simone:0,013035, artistica:0,011283, lindo:0,010052, solo:0,009462, final:0,008799, nota:0,008218.	ginástica
09	0,046332	set:0,042990, jogo:0,029365, volei:0,015995, vamos:0,01412, pedro:0,009602, evandro:0,009129, praia:0,008812, basquete:0,008454, vitoria:0,008031, primeiro:0,007854.	vôlei
10	0,040964	olimpico:0,020999, video:0,015317, jogo:0,010921, janeiro:0,009305, olimpica:0,007628, gostei:0,007085, parque:0,006251, clima:0,005059, hoje:0,004395, arena:0,004236.	parque olímpico
11	0,039275	final:0,028799, quartas:0,015219, bellucci:0,0141, tenis:0,01396, masculino:0,013294, nadal:0,011707, brasileiro:0,011447, marcelo:0,010703, vence:0,010479, bruno:0,009985.	tênis
12	0,036866	onibus:0,013755, piscina:0,013700, jogo:0,013683, verde:0,011843, agua:0,010707, ingresso:0,009089, seguranca:0,007532, comite:0,006939, jornalistas:0,006667, nacional:0,006551.	natação
13	0,035025	musica:0,009594, casamento:0,009449, brasileira:0,00816, hino:0,007302, jogo:0,006748, atleta:0,00592, historia:0,005325, pedido:0,005016, gay:0,004812, mundo:0,004584.	cerimônia de abertura
14	0,030861	atleta:0,011473, lugar:0,009918, final:0,009766, prova:0,00874, jogo:0,007086, confira:0,00704, hoje:0,006239, veja:0,006173, oliveira:0,006127, brasileiro:0,005684.	natação
15	0,016575	manifestacoes:0,023579, protestos:0,023370, juiz:0,020324, decisao:0,018912, temer:0,017758, politicas:0,01653, repressao:0,01621, federal:0,014602, libera:0,014442, proibe:0,013558.	político

Tabela A.4: Tópicos obtidos a partir dos *tweets* extraídos dos dias 09 a 11 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,172663	jogo:0,029173, hoje:0,011550, assistir:0,010152, vendo:0,008869, assistindo:0,006818, futebol:0,006369, volei:0,006357, esporte:0,006322, tv:0,005967, casa:0,00579.	transmissão das olimpíadas
02	0,126433	atleta:0,017913, esporte:0,016092, brasileiro:0,013837, medalha:0,01037, pais:0,007995, futebol:0,007584, mundo:0,006469, jogo:0,006207, melhor:0,005737, copa:0,004869.	futebol
03	0,088783	jogo:0,036449, meninas:0,024488, feminino:0,021553, selecao:0,018341, futebol:0,017076, vamos:0,016903, coracao:0,014846, barbara:0,014665, volei:0,01262, penalti:0,012608.	futebol
04	0,070612	solo:0,010077, hope:0,009391, zika:0,007781, torcida:0,007265, eua:0,00725, mundo:0,005872, brasileiro:0,005812, mae:0,005628, bolt:0,00541, casa:0,005317.	atletismo
05	0,067847	feminino:0,047137, masculino:0,032405, volei:0,032085, futebol:0,027563, jogo:0,026616, final:0,024776, selecao:0,019373, praia:0,015537, quartas:0,015114, alemanha:0,010645.	vôlei
06	0,062749	bolt:0,051732, usain:0,018179, final:0,015189, prova:0,012804, phelps:0,011838, olimpico:0,010938, recorde:0,010747, salto:0,009717, metros:0,009485, atletismo:0,007737.	atletismo
07	0,062015	neymar:0,027441, gol:0,024837, colombia:0,020102, jogo:0,017857, selecao:0,013488, futebol:0,009608, bola:0,009319, time:0,007886, falta:0,007788, luan:0,007296.	futebol
08	0,054187	medalha:0,068826, ouro:0,053708, phelps:0,033184, michael:0,011638, historia:0,011614, simone:0,010946, ganhar:0,010561, biles:0,01029, ganhou:0,009043, natacao:0,008021.	natação
09	0,053072	melhor:0,018604, jogo:0,012746, globo:0,012322, guga:0,009336, galvao:0,007297, hino:0,007043, volei:0,00692, espn:0,006686, comentarista:0,006127, narrador:0,005528.	mídia
10	0,050989	set:0,042525, jogo:0,030336, nadal:0,023303, bellucci:0,019985, vamos:0,015152, argentina:0,010954, primeiro:0,008969, tempo:0,007516, bola:0,006926, segundo:0,006232.	tênis
11	0,047420	diego:0,068006, nory:0,030719, parabens:0,030551, hypolito:0,029224, medalha:0,028221, cai:0,027272, arthur:0,026996, ginastica:0,025049, prata:0,022856, bronze:0,021756.	ginástica
12	0,046523	rafael:0,027651, medalha:0,022224, silva:0,021781, nadal:0,020171, ouro:0,019841, bronze:0,017937, final:0,017819, robson:0,015146, murray:0,014201, delpotro:0,013487.	tênis
13	0,036200	video:0,016113, atleta:0,015613, jogo:0,012639, veja:0,008173, gostei:0,007859, brasileiro:0,00722, confira:0,006906, olimpica:0,005588, pokemon:0,004676, doping:0,004565.	cerimônia de abertura
14	0,034933	piscina:0,015877, agua:0,012976, comite:0,01176, temer:0,011375, verde:0,008943, seguranca:0,008685, nacional:0,008504, forca:0,007613, justica:0,006897, jogo:0,006807.	cerimônia de abertura
15	0,025575	olimpico:0,027539, ouro:0,009615, janeiro:0,009311, estadio:0,008055, marcha:0,007611, parque:0,007226, tiro:0,007209, atletica:0,006495, engenhao:0,006355, hoje:0,005805.	boxe

Tabela A.5: Tópicos obtidos a partir dos *tweets* extraídos dos dias 12 a 14 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,162476	jogo:0,022658, hoje:0,012497, assistir:0,010723, vendo:0,008364, volei:0,007909, ficar:0,00645, acabar:0,006139, assistindo:0,006068, esporte:0,005952, tv:0,00565.	vôlei
02	0,112806	frances:0,013691, brasileiro:0,013508, torcida:0,012738, mundo:0,011982, copa:0,010833, vaia:0,007522, pais:0,007044, atleta:0,006758, ouro:0,00583, futebol:0,005595.	futebol
03	0,107202	feminino:0,027587, meninas:0,025128, futebol:0,020816, volei:0,02065, hoje:0,017324, selecao:0,016109, ouro:0,011796, medalha:0,00882, parabens:0,008729, mulher:0,008533.	vôlei
04	0,087242	atleta:0,026583, medalha:0,022740, esporte:0,018948, brasileiro:0,013726, pais:0,007681, ouro:0,006493, jogo:0,006058, militares:0,004984, outros:0,004218, pessoa:0,004203.	ginástica
05	0,084218	volei:0,044456, feminino:0,035079, final:0,03331, masculino:0,027022, futebol:0,022114, jogo:0,022099, praia:0,019289, hoje:0,017032, selecao:0,013668, franca:0,011736.	vôlei
06	0,074065	set:0,027582, vamos:0,027261, jogo:0,025975, coracao:0,013138, deus:0,01235, meninas:0,010288, volei:0,009693, hoje:0,00938, china:0,008544, penalti:0,008186.	vôlei
07	0,060302	thiago:0,073002, ouro:0,056044, braz:0,050671, vara:0,047805, salto:0,041893, olimpico:0,023869, medalha:0,02275, recorde:0,016796, silva:0,01513, parabens:0,015062.	atletismo
08	0,057674	melhor:0,016316, globo:0,009912, jogo:0,00926, galvao:0,009232, guga:0,00685, musica:0,005029, narrador:0,004847, tv:0,004523, espn:0,00452, volei:0,004327.	mídia
09	0,049801	gol:0,038727, neymar:0,022461, selecao:0,018177, jogo:0,015677, honduras:0,014304, jesus:0,012329, suecia:0,010358, bola:0,010234, final:0,009902, gabriel:0,009501.	futebol
10	0,049155	medalha:0,041884, zanetti:0,021523, flavinha:0,020036, bronze:0,016725, prata:0,016199, ouro:0,015376, arthur:0,014856, poliana:0,012894, biles:0,011704, parabens:0,011639.	ginástica
11	0,039650	video:0,012711, jogo:0,010772, olimpico:0,008061, olimpica:0,006442, janeiro:0,006286, gostei:0,00597, hoje:0,005942, pokemon:0,005038, veja:0,004846, arena:0,004754.	cerimônia de abertura
12	0,036560	ouro:0,077165, medalha:0,053280, robson:0,045684, conceicao:0,031295, isaquias:0,020204, boxe:0,02007, prata:0,018609, parabens:0,015248, queiroz:0,013258, brasileiro:0,011949.	boxe
13	0,029319	final:0,023443, bolt:0,020606, usain:0,00892, metros:0,008623, prova:0,007515, fernando:0,006797, semifinal:0,006602, cleubercunha:0,006375, salto:0,006332, pimenta:0,005938.	atletismo
14	0,025646	olimpico:0,028788, camera:0,019324, parque:0,015311, pessoa:0,008211, cai:0,008038, acidente:0,007847, caiu:0,007094, hoje:0,005847, estadio:0,005705, duas:0,005365.	ciclismo
15	0,023884	temer:0,022341, justica:0,012108, comite:0,011525, ingresso:0,009848, jogo:0,008423, coi:0,008264, lula:0,007012, policia:0,006269, presidente:0,006024, dilma:0,005786.	político

Tabela A.6: Tópicos obtidos a partir dos *tweets* extraídos dos dias 15 a 17 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,164754	jogo:0,023519, hoje:0,009647, final:0,007286, acabar:0,007197, assistir:0,007081, amanha:0,007053, vamos:0,006451, volei:0,006403, fim:0,006096, acabando:0,006058.	vôlei
02	0,132994	medalha:0,018859, atleta:0,014614, futebol:0,013628, esporte:0,013393, brasileiro:0,011596, melhor:0,010461, ouro:0,009996, parabens:0,007522, jogo:0,007176, pais:0,006704.	futebol
03	0,097410	ouro:0,037744, volei:0,035308, final:0,031569, masculino:0,024259, feminino:0,022275, jogo:0,019013, bruno:0,018151, praia:0,014325, parabens:0,013679, selecao:0,01322.	vôlei
04	0,094932	ouro:0,011081, brasileiro:0,010518, hino:0,010343, alemanha:0,010193, galvao:0,008389, torcida:0,008018, lindo:0,006276, medalha:0,006192, neymar:0,005963, mundo:0,005732.	futebol
05	0,073709	ouro:0,052049, copa:0,038877, alemanha:0,035464, futebol:0,030263, mundo:0,02124, selecao:0,020729, campeao:0,02056, ganhar:0,017309, medalha:0,014406, final:0,01138.	futebol
06	0,067436	gol:0,044956, alemanha:0,044138, jogo:0,027441, vamos:0,020175, penalti:0,016963, bola:0,01439, final:0,013508, tempo:0,012406, deus:0,010414, coracao:0,010156.	futebol
07	0,066280	neymar:0,035828, selecao:0,014838, jesus:0,011966, gol:0,011208, melhor:0,011168, ouro:0,008713, jogo:0,008234, gabriel:0,008183, luan:0,008055, weverton:0,007949.	futebol
08	0,061818	medalha:0,080151, isaquias:0,048744, ouro:0,030236, queiroz:0,022695, brasileiro:0,021476, bronze:0,018003, prata:0,016414, historia:0,014336, parabens:0,012541, canoagem:0,010389.	canoagem
09	0,048273	bolt:0,059007, usain:0,020063, prova:0,012567, ouro:0,012121, final:0,009711, medalha:0,009229, olimpico:0,00921, lugar:0,008486, phelps:0,007254, atleta:0,007222.	atletismo
10	0,045898	olimpico:0,022209, jogo:0,012684, video:0,012314, olimpica:0,007941, final:0,007797, janeiro:0,007612, hoje:0,007505, atleta:0,007484, gostei:0,006999, parque:0,006929.	cerimônia de abertura
11	0,038091	ginastica:0,015530, lindo:0,014103, musica:0,012324, ritmica:0,009549, melhor:0,008622, amo:0,005386, jogo:0,004916, mascote:0,004658, nado:0,004529, sincronizado:0,004524.	nado sincronizado
12	0,034775	temer:0,013750, dilma:0,013061, lula:0,012052, jogo:0,00872, ingresso:0,007792, governo:0,007499, coi:0,006609, mil:0,006105, atleta:0,006007, dinheiro:0,006001.	político
13	0,034634	nadadores:0,030298, americanos:0,023106, lochte:0,016198, caso:0,011943, assalto:0,01143, ryan:0,011114, desculpas:0,011049, eua:0,010619, atleta:0,0093, brasileiro:0,007682.	natação
14	0,019908	ouro:0,054224, grael:0,032256, martine:0,032162, kahena:0,029987, vela:0,024625, kunze:0,02018, medalha:0,017211, parabens:0,014809, familia:0,008785, olimpico:0,007509.	vela
15	0,019088	globo:0,024375, waack:0,016191, cris:0,013854, melhor:0,013696, william:0,013588, galvao:0,012986, marcha:0,011852, atletica:0,010669, jornal:0,00851, guga:0,008135.	mídia

Tabela A.7: Tópicos obtidos a partir dos *tweets* extraídos dos dias 18 a 20 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,100277	saudades:0,016620, lindo:0,014392, sentir:0,013978, jogo:0,013949, crush:0,011696, falta:0,010611, chorando:0,010345, saudade:0,009381, acabou:0,008744, coracao:0,008593.	cerimônia de encerramento
02	0,097294	mundo:0,031407, melhor:0,024072, copa:0,02126, pais:0,021155, brasileiro:0,019432, lindo:0,011099, festa:0,010954, povo:0,009544, japao:0,008153, orgulho:0,007919.	cerimônia de encerramento
03	0,095406	atleta:0,023904, carnaval:0,022967, gringos:0,017506, samba:0,013274, povo:0,01158, festa:0,009558, tenis:0,008857, mundo:0,00842, gringo:0,007984, maracana:0,007391.	cerimônia de encerramento
04	0,095272	acabou:0,030424, fim:0,025533, amanha:0,025322, triste:0,013866, hoje:0,013155, volta:0,012439, jogo:0,012152, voltar:0,010677, acaba:0,009466, vamos:0,008903.	cerimônia de encerramento
05	0,091574	ouro:0,052332, volei:0,030658, medalha:0,020311, masculino:0,016792, italia:0,0157, melhor:0,013369, selecao:0,013028, jogo:0,012167, futebol:0,011494, wallace:0,01045.	cerimônia de encerramento
06	0,083236	lindo:0,024028, parabens:0,022227, obrigado:0,016, orgulho:0,01544, pais:0,014652, brasileiro:0,014407, atleta:0,013837, esporte:0,008708, jogo:0,008665, festa:0,008124.	cerimônia de encerramento
07	0,076066	tokyo:0,026174, vamos:0,023653, proxima:0,019879, jogo:0,01409, japao:0,009996, copa:0,008748, novo:0,008094, daqui:0,008082, estar:0,008034, vem:0,007015.	cerimônia de encerramento
08	0,053142	lindo:0,041399, hino:0,029395, criancas:0,011246, musica:0,010933, cantando:0,010673, cerimonia:0,008734, nacional:0,008693, bandeira:0,008548, nordeste:0,008515, frevo:0,007686.	cerimônia de encerramento
09	0,052697	galvao:0,021184, temer:0,015097, globo:0,011829, dilma:0,010397, lula:0,009933, gloria:0,009197, maria:0,00865, boca:0,0077, paes:0,007069, eduardo:0,005748.	mídia
10	0,050782	medalha:0,035030, ouro:0,016434, atleta:0,013256, olimpico:0,009269, maratona:0,00837, bolt:0,006968, jogo:0,006918, esporte:0,006335, melhor:0,005995, isaquias:0,005686.	cerimônia de encerramento
11	0,047096	santos:0,012829, dinheiro:0,011691, maracana:0,011051, atleta:0,00942, dumont:0,007585, cerimonia:0,006406, ingresso:0,006247, estadio:0,005278, novo:0,005094, chuva:0,004941.	cerimônia de abertura
12	0,041973	musica:0,016091, cantando:0,015576, cantar:0,009405, tocar:0,0092, katy:0,008875, esperando:0,008651, ivete:0,008591, cerimonia:0,007866, funk:0,006234, rise:0,006148.	cerimônia de encerramento
13	0,041695	pira:0,031960, olimpica:0,029155, chama:0,022083, apagar:0,021868, embora:0,018601, chuva:0,016522, deixa:0,015638, tocha:0,015194, fogo:0,013739, deus:0,012604.	cerimônia de encerramento
14	0,039979	japao:0,060870, mario:0,057930, tokyo:0,037662, ministro:0,023875, primeiro:0,021903, super:0,020899, pokemon:0,015261, vestido:0,008807, apresentacao:0,008803, cano:0,008697.	cerimônia de encerramento
15	0,033511	tremendo:0,020320, homem:0,017431, ingles:0,016257, nuzman:0,014111, parkinson:0,012451, falando:0,012171, mal:0,011787, bandeira:0,010698, red:0,01067, desse:0,009931.	Nuzman (parkinson)

Tabela A.8: Tópicos obtidos a partir dos *tweets* extraídos do dia 21 de Agosto de 2016.

Tópico	P(z)	Palavras	Rótulo
01	0,134760	tokyo:0,013867, proxima:0,007137, japao:0,007122, acabou:0,006183, copa:0,005539, mundo:0,005109, dinheiro:0,005105, pais:0,004845, vamos:0,004626, falando:0,004611.	cerimônia de encerramento
02	0,134476	acabou:0,044962, fim:0,017214, volta:0,013889, acabaram:0,012892, vamos:0,009691, hoje:0,009113, voltar:0,008985, bom:0,008497, vida:0,008237, semana:0,007413.	cerimônia de encerramento
03	0,127592	acabou:0,020387, tv:0,019006, hoje:0,017039, saudades:0,014996, jogo:0,01494, fim:0,01418, triste:0,012408, volta:0,010457, saudade:0,009748, assistir:0,009298.	cerimônia de encerramento
04	0,098535	saudades:0,013349, lindo:0,013205, parabens:0,009855, melhor:0,009772, saudade:0,008195, jogo:0,008015, mundo:0,007984, brasileiro:0,007809, festa:0,007242, obrigado:0,007195.	cerimônia de encerramento
05	0,078644	lula:0,021321, dilma:0,018855, sucesso:0,012587, temer:0,012361, brasileiro:0,009121, pais:0,007843, mundo:0,007204, povo:0,006939, atleta:0,006642, copa:0,006255.	político
06	0,069337	jogo:0,031846, atleta:0,025215, paraolimpiadas:0,014727, paralimpicos:0,011155, esporte:0,010306, acabou:0,006822, veja:0,006514, brasileiro:0,006339, vamos:0,00628, ingresso:0,005768.	cerimônia de encerramento
07	0,061911	ouro:0,032993, futebol:0,024478, medalha:0,021176, volei:0,017636, selecao:0,015923, final:0,013923, jogo:0,010227, masculino:0,010113, feminino:0,009752, neymar:0,009219.	futebol
08	0,056287	medalha:0,044895, ouro:0,027629, copa:0,015572, mundo:0,012733, melhor:0,012418, pais:0,008678, atleta:0,008669, historia:0,006236, quadro:0,005416, brasileiro:0,005272.	futebol
09	0,044183	ingresso:0,014281, jogo:0,012343, milhoes:0,012262, mil:0,011838, turistas:0,0092, legado:0,009039, olimpico:0,008488, dinheiro:0,006802, venda:0,006791, janeiro:0,006458.	cerimônia de encerramento
10	0,043081	luan:0,036643, lochte:0,019186, tite:0,01688, melhor:0,015622, jogou:0,014587, gabigol:0,013657, ryan:0,013446, jesus:0,012422, taison:0,011081, convocacao:0,010011.	futebol
11	0,037295	video:0,052025, gostei:0,026888, bolt:0,02003, neymar:0,012928, playlist:0,010837, adicionei:0,009655, ouro:0,007917, usain:0,007279, bruna:0,006817, final:0,005826.	atletismo
12	0,033967	jogo:0,014889, olimpica:0,014695, brasileira:0,011715, historia:0,01161, fotos:0,00936, melhor:0,006349, robotica:0,006334, nacional:0,006051, hoje:0,005992, facebook:0,005902.	cerimônia de encerramento
13	0,029506	mario:0,046874, japao:0,033158, tokyo:0,030312, super:0,017608, ministro:0,015156, japones:0,012293, primeiro:0,011522, pokemon:0,01111, cerimonia:0,01087, jogo:0,010733.	cerimônia de encerramento
14	0,026409	atleta:0,043933, medalha:0,034649, prata:0,013841, medalhista:0,010337, ajudar:0,008461, cancer:0,007906, etiope:0,007217, jogo:0,006719, contas:0,006662, leiloa:0,00626.	cerimônia de encerramento
15	0,024017	fim:0,020826, musica:0,012412, agatha:0,012201, barbara:0,011612, neymar:0,010561, ultima:0,010245, midia:0,010203, volta:0,009972, diario:0,009193, cores:0,009193.	vôlei

Tabela A.9: Tópicos obtidos a partir dos *tweets* extraídos dos dias 22 a 24 de Agosto de 2016.

APÊNDICE B – Links

B.1 02 a 04 de Agosto de 2016:

http://www.adorocinema.com/filmes/filme-144185/
http://agenciabrasil.ebc.com.br/rio-2016/noticia/2016-08/brasil-e-africa-do-sul-termina-em-0-0
http://g1.globo.com/tecnologia/games/noticia/2016/08/pokemon-go-comeca-funcionar-no-brasil.html
http://globoesporte.globo.com/rj/olimpiadas/jogo/03-08-2016/feminino-brasil-china/
http://globoesporte.globo.com/df/olimpiadas/jogo/04-08-2016/masculino-brasil-africa-do-sul/

B.2 05 de Agosto de 2016:

https://oglobo.globo.com/esportes/rio-celebra-diversidade-passa-mensagem-de-esperanca-na-abertura-dos-jogos-no-maracana-19866473
https://www.olympic.org/news/a-brazil-style-opening-ceremony
https://recordtv.r7.com/rio-2016/siga-tudo-o-que-rolou-na-cerimonia-de-abertura-da-rio-2016-30012019

B.3 06 a 08 de Agosto de 2016:

https://oglobo.globo.com/esportes/destaques-do-primeiro-dia-oficial-de-jogos-da-olimpiada-do-rio-19871158
https://oglobo.globo.com/esportes/veja-que-foi-destaque-no-segundo-dia-de-olimpiadas-19874647
https://oglobo.globo.com/esportes/rio-2016/veja-que-foi-destaque-no-terceiro-dia-de-olimpiadas-19882091
https://www.olympic.org/news/when-virginia-thrasher-won-the-first-medal-in-rio
https://www.olympic.org/news/michael-phelps-strikes-again
https://www.olympic.org/news/australia-writes-a-chapter-of-olympic-history
https://recordtv.r7.com/rio-2016/acompanhe-tudo-o-que-rolou-no-primeiro-dia-de-competicoes-dos-jogos-olimpicos-do-rio-30012019

https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-segundo-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-terceiro-dia-de-disputa-da-rio-2016-30012019

B.4 09 a 11 de Agosto de 2016:

https://www.olympic.org/news/michael-phelps-and-simone-biles-take-centre-stage
https://www.olympic.org/news/king-kohei-uchimura-retains-his-title-right-at-the-death
https://www.olympic.org/news/michael-phelps-unsurpassable-simone-biles-brilliant
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-quarto-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-quinto-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-sexto-dia-de-disputa-da-rio-2016-30012019

B.5 12 a 14 de Agosto de 2016:

http://g1.globo.com/hora1/noticia/2016/08/diego-hypolito-e-arthur-nory-conquistam-medalhas-na-ginastica.html
https://www.olympic.org/news/almaz-ayana-shatters-the-10-000m-world-record
https://www.olympic.org/news/phelps-bows-out-with-a-23rd-title-elaine-thompson-becomes-the-new-queen-of-the-sprint
https://www.olympic.org/news/van-niekerk-bolt-murray-whitlock-biles-kenny-and-rose-shine
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-setimo-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-oitavo-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-nono-dia-de-disputa-da-rio-2016-30012019

B.6 15 a 17 de Agosto de 2016:

http://globoesporte.globo.com/olimpiadas/boxe/noticia/2016/08/ouro-inedito-robson-bate-frances-e-conquista-titulo-historico-no-boxe.html
https://www.olympic.org/news/drama-in-the-men-s-pole-vault-and-the-women-s-beam
https://www.olympic.org/news/simone-biles-goes-down-in-history-a-british-couple-at-their-peak
https://www.olympic.org/news/kaori-icho-s-four-out-of-four-elaine-thompson-is-queen-of-the-sprint
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-10-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-11-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-12-dia-de-disputa-da-rio-2016-30012019
https://pt.wikipedia.org/wiki/Thiago_Braz

B.7 18 a 20 de Agosto de 2016:

https://www.olympic.org/news/usain-bolt-on-top-of-the-world
https://www.olympic.org/news/new-triple-for-bolt-germany-at-the-summit-of-women-s-football
https://www.olympic.org/news/park-in-bee-makes-history-neymar-farah-and-felix-at-the-top-of-their-game
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-13-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-14-dia-de-disputa-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-penultimo-dia-de-disputa-da-rio-2016-30012019

B.8 21 de Agosto de 2016:

https://brasil.elpais.com/brasil/2016/08/22/deportes/1471846331_102160.html
https://oglobo.globo.com/rio/imprensa-internacional-destaca-superacao-da-organizacao-para-realizar-os-jogos-do-rio-19969365
https://www.olympic.org/news/the-final-fireworks
https://recordtv.r7.com/rio-2016/acompanhe-tudo-sobre-o-ultimo-dia-de-disputas-da-rio-2016-30012019
https://recordtv.r7.com/rio-2016/acompanhe-a-cerimonia-de-encerramento-da-rio-2016-30012019

B.9 22 a 24 de Agosto de 2016:

https://oglobo.globo.com/esportes/rio-2016-uma-olimpiada-com-recordes-historias-marcantes-19972316