

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Redes sociais no contexto da conservação de mamíferos terrestres: conectando a
dimensão humana ao universo digital

Yuri Geraldo Gomes Ribeiro

Dissertação apresentada para obtenção do título de Mestre em
Ciências, Programa: Recursos Florestais. Opção em: Conserva-
ção de Ecossistemas Florestais

Piracicaba
2020

Yuri Geraldo Gomes Ribeiro
Bacharel em Engenharia Florestal

Redes sociais no contexto da conservação de mamíferos terrestres: conectando a dimensão humana ao universo digital

Orientadora:
Profa. Dra. **KATIA MARIA PASCHOALLETTO MICCHI DE BARROS FERRAZ**

Dissertação apresentada para obtenção do título de Mestre em Ciências, Programa: Recursos Florestais. Opção em: Conservação de Ecossistemas Florestais

Piracicaba
2020

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP**

Ribeiro, Yuri Geraldo Gomes

Redes sociais no contexto da conservação de espécies de mamífero: conectando a dimensão humana com o universo digital/ Yuri Geraldo Gomes Ribeiro. - - Piracicaba, 2019.

92p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Redes sociais 2. Conservação de mamíferos terrestres 3. Dimensão de dados I. Título

AGRADECIMENTOS

Gostaria de começar agradecendo á vida que me proporcionou este período de experiência e aprendizado.

À minha orientadora, mestra e amiga Prof^a Dra. Katia M.P.M.B. Ferraz. Você acreditou neste trabalho e fez parte de toda a trajetória que me trouxe até aqui, me amparou quando precisei e me deu todas as condições e chances para que eu pudesse ser o profissional que sou hoje.

Ao meu co-orientador Prof^o Dr. Pedro Pizzigatti Corrêa. A primeira vez que tive contato com a ideia deste trabalho a achei completamente absurda e impossível, você me mostrou que também sou um cientista dos dados em desenvolvimento.

Ao meu amigo e co-orientador nas horas vagas Dr. Silvio Marchini. Você abriu a persiana para janela do mundo e me fez ver quão grande podemos sonhar.

Ao meu time de ajudantes: Beatriz Silva, Beatriz Trivellato, Rafael e Rodrigo. Obrigado queridos, obrigado por me permitirem exercer uma das funções mais bonitas da academia, ensinar e aprender o dobro. Aos meus ajudantes de campo, embora o esforço de entrevistas não esteja incluído aqui, ele também é parte dessa dissertação, obrigado Mayara, Carla e especialmente Leandro e Gabriela que seguraram muitas broncas comigo vocês todos se tornaram bons amigos.

Ao Prof^o Dr. Ricardo, por participar dessa pesquisa e colaborar com a orientação deste trabalho. Aos pós-doutorandos: Dra. Laila Sandroni e Dr. Juliano Bogoni, mesmo não formalmente, vocês cederam seu tempo e ajudaram muito na construção e elaboração dessa dissertação. Ao Alex Bovo, pelos pitacos e correções. Á Luiza, por ajudar na revisão ortográfica do trabalho.

Às agências de fomento. CAPES por proporcionar o período de bolsa para elaboração deste trabalho, que nosso país possa gozar do apoio á ciência para construção de conhecimento de qualidade que sirva a sociedade e nos permita o desenvolvimento. Á Fundação Boticário pelo financiamento ao projeto 201710015. Á FAPESP pelo financiamento ao processo 2018/16662-6. Ao Programa de Excelência Acadêmica (CAPES) que permitiu a impressão desta dissertação. Ao Banco Bradesco que financiou as buscas através da Twitter API.

Aos amigos de laboratório: Beatriz, Maristela, Ana, Daiane, Roberta, Mari e Vinicius. Especialmente aos amigos Alex e Letícia. E aos que cruzaram meu caminho, meus gringos favoritos Jake e Anna. Vocês deixaram meus dias mais leves e felizes. Á amiga Dra. Fernanda Abra, que me estendeu a mão em um período sem financiamento, serei eternamente grato.

Aos amigos da vida, esses que não entendem muito bem do que nosso trabalho trata. Queridos de São Paulo: Bruno, Thiago, Aline, Victor e Ketilly. Aos amigos de Piracicaba e ESALQ, que seguem no meu coração: Ingrid, Isabella, Yohanna, Érica, Laura, Luã e Maria Elisa. Á Laís, que me divertiu e foi minha companheira nessa reta final. A surpresa da amizade da Luiza. Ao companherismo do amigo Igor que esteve comigo em boa parte da jornada. Aos muitos personagens que essa jornada fantástica me revelou.

Á minha mãe que me nutriu de amor e sempre acreditou nas desventuras do seu filho. Minha irmã Penelope, que ajudou muito em tudo da maneira que pôde. Meu irmão Adolfo e sua esposa Michelle, além de toda sua família. Aos meus sobrinhos queridos que trazem tanta alegria pro meu coração: César, Lorena, Wendel, Carlinhos e Larissa. Ao meu pai Carlos César, que seja onde estiver segue comigo e sempre estará na minha memória.

***Dedico este trabalho á meu pai Carlos César Ribeiro.
Espero que você esteja bem e feliz, seja onde estiver.***

"And what is the use of a book," thought Alice, "without pictures or conversation?"

Lewis Carroll, Alice's Adventures in Wonderland

"We should not feel guilty to do, what it is, at a minimum, the most decent."

Nigel Collar piece of thought, given in front a bonfire, about conservation practice

SUMÁRIO

RESUMO.....	8
ABSTRACT.....	10
1. INTRODUÇÃO.....	15
2. REVISÃO BIBLIOGRÁFICA.....	21
3. MATERIAIS E MÉTODOS.....	27
3.1 Plataformas de busca	27
3.2 Área de estudo	27
3.3 Esquema de procedimentos metodológicos	29
3.4 Coleta e organização de dados	31
3.4.1 Facebook	31
3.3.2 Twitter	37
3.5 Análise dos dados	319
3.6 TESTES DE CATEGORIZAÇÃO NA BASE DO TWITTER UTILIZANDO APRENDIZADO DE MÁQUINA.....	40
4. RESULTADOS.....	43
Facebook - <i>O que?</i>	43
Facebook - <i>Quando?</i>	47
Facebook - <i>Onde?</i>	49
Facebook - <i>Quem?</i>	51
Facebook - <i>Engajamento</i>	52
Twitter - <i>O que?</i>	60
Twitter – <i>Quando?</i>	63
Twitter – <i>Quem?</i>	64
Twitter – <i>Engajamento</i>	65
Classificação da base do Twitter utilizando aprendizado de máquina.....	68
5. DISCUSSÃO	69
6. CONCLUSÕES.....	81
REFERÊNCIAS.....	82

RESUMO

Redes sociais no contexto da conservação de mamíferos terrestres: conectando a dimensão humana ao universo digital

O volume de dados de redes sociais e seus usos têm crescido em número e complexidade e vem sendo utilizado nas mais variadas áreas, de marketing a pesquisas de opinião. Estratégias para sua obtenção, tratamento e análise são necessárias já que variam enormemente em relação a características (dimensões) dos dados. Este conjunto de dados apresenta forte potencial para contribuir, por exemplo, com o monitoramento da opinião pública em relação a eventos relacionados à conservação. Apesar da crescente popularidade das redes sociais, estudos que abordam seu potencial como fonte de informação para conservação e manejo de fauna silvestre continuam novos e merecem uma atenção especial. Assim, objetivamos neste estudo explorar dados de redes sociais sob a ótica de cinco dimensões: do que se trata, quando foi gerado, onde foi gerado, quem gerou e o engajamento popular (i.e., interação pública com conteúdo *online* através de recursos disponibilizados por redes sociais) que recebeu, afim de explorarmos o potencial de aprendizado com conjuntos de dados menores e aplicações em processos automatizados. Obtivemos dados a partir de duas plataformas de busca: Facebook e Twitter, relacionados a quatro espécies icônicas da fauna brasileira: onça-pintada (*Panthera onca*), onça-parda (*Puma concolor*), bugio (*Alouatta* spp.) e bicho-preguiça (*Bradypus* spp.; *Choloepus* spp.). Utilizamos a combinação de busca “espécie+local” no Facebook e apenas “espécie” no Twitter, através de API. Obtivemos 317 postagens para as espécies no Facebook e 38.516 no Twitter (com uma estimativa de 10% de postagens não correspondendo às espécies). Classificamos os dados em 12 categorias, sendo que por espécie, observamos que as categorias mais proeminentes foram “avistamento de fauna fora de área protegida”, “cativo”, “resgate/tratamento e soltura de fauna” e “saúde de fauna silvestre”. A análise específica do texto por espécie revelou palavras principalmente relacionadas a estas categorias no Facebook, e no Twitter a inclusão de termos relacionados ao bioma Pantanal e outras áreas. Observamos padrões temporais que destacam a importância das categorias para entender picos de informação no Facebook. No Twitter, os padrões temporais parecem estar relacionados às mesmas categorias, mas utilizamos ranking de palavras para entender picos de informação. Palavras como “Juma” (onça-pintada abatida logo antes dos jogos olímpicos de 2016) foram importantes para determinar eventos que geraram picos de postagens. De modo geral, a distribuição espacial das postagens se concentrou no município de São Paulo, com poucos dados apresentando localização precisa de eventos. Encontramos 9 categorias de tipos de usuários para o Facebook, onde páginas pessoais, páginas de notícias, páginas de negócios e zoológicos e aquários foram as categorias com mais postagens encontradas. No

Twitter, observamos grande variação em relação aos valores obtidos para seguidores e amigos de usuários que postaram sobre as espécies de interesse deste estudo. Quando analisamos as variáveis de engajamento *online*, observamos a predominância de reações positivas no Facebook e altos valores de compartilhamento e comentários em postagens, com exceção do bugio, que apresentou reações negativas. Identificamos a importância de temas como a febre amarela na determinação de engajamento. Além disso, encontramos variáveis com associação positiva com engajamento através de regressão linear múltipla, como “avistamento de fauna em área protegida” e o tipo de usuário “instituição governamental”. A partir de algoritmos de aprendizagem de máquina conseguimos resultados variando entre 0,58 e 0,71 de acurácia para a classificação automatizada na base do Twitter. Este trabalho teve como principal conclusão a importância da utilização de agrupamento de dados menores para lidar com grandes conjuntos de dados de redes sociais no contexto da conservação de espécies, utilizando o Brasil e estado de São Paulo como áreas de estudo, além da necessidade de expandir pesquisas neste segmento no Brasil que contribuam com conservação aplicada.

Palavras-chave: Redes sociais; Conservação de mamíferos terrestres; Dimensão de dados

ABSTRACT

Social media in the context of land mammals conservation: connecting human dimension with the digital universe

The volume of data from social media and their uses have grown in number and complexity and have been used in the most diverse areas, from marketing to opinion monitoring. Strategies for obtaining, treating and analyzing them are necessary, since they vary enormously in relation to its characteristics (dimensions) of data. These data have a strong potential to contribute, for example, to the monitoring of public opinion in relation to conservation-related events. Despite the growing popularity of social media, studies that address its potential as a source of information for conservation and management of wildlife remain new and deserve special attention. Thus, in this study we aim to explore data from social media from the perspective of five different dimensions: what is it about, when was it generated, where was it generated, who generated it and engagement (i.e., public interaction with *online* content through metrics available by social media) it received, in order to explore and learn from smaller data sets and apply it in automated processes. We obtained data from two search platforms: Facebook and Twitter, related to four central species from Brazilian wildlife: jaguar (*Panthera onca*), puma (*Puma concolor*), howler monkey (*Alouatta spp.*) and sloth (*Bradypus spp.*; *Choloepus spp.*). We use the combination of “species + local” search on Facebook and only “species” on Twitter through API (*Application Programming Interface*) use. We obtained 317 posts for the species on Facebook and 38,516 on Twitter (with an estimated 10% of posts not corresponding to the species) on Twitter. We classified data in 12 categories, we observed that the most prominent categories were “sighting of fauna outside protected areas”, “captivity”, “rescue / treatment and release of wildlife” and “wildlife health”. The specific analysis of the text by species revealed words mainly related to these categories on Facebook with the inclusion of terms related to the Pantanal biome and other areas on Twitter. We observed temporal patterns that highlight the importance of categories to understand information peaks on Facebook. On Twitter, temporal patterns seem to be related to the same categories found in Facebook, but we use word ranking to understand peaks of information. Words like “Juma” (jaguar shot dead before 2016 Brazil Olympic Games) were important in determining events that generated peaks in posting. In general, the spatial distribution of posts was concentrated in the municipality of São Paulo with little data showing precise location of events. We found 9 categories of users types on Facebook, where personal pages, news pages, business and zoos and aquariums were the categories with most posts found. On Twitter, we observed a great variation in relation to the values obtained for followers and friends of users who posted about the species of interest in this study. When we analyze engagement variables *online*, we see a predominance of positive reactions on Facebook and high values for sharing and commenting on

posts, with the exception of the howler monkey, who showed negative reactions. We identified the importance of topics such as yellow fever in determining engagement. In addition, we found variables with a positive association with engagement through multiple linear regression, such as “wildlife sighting in a protected area” and the type of user “government institution”. Using machine learning algorithms, we obtained results varying between 0.58 and 0.71 of accuracy for the automated classification performed on Twitter database. The main conclusion of this work was the importance of using smaller data clusters to deal with large social network data sets in the context of species conservation using Brazil and the state of São Paulo as areas of study. In addition, we found the need to expand research in this segment in Brazil that contribute to applied conservation.

Keywords: Social media; Mammals; Dimensions of data

LISTA DE FIGURAS

- Figura 1. Limites adotados para buscas em redes sociais neste estudo. A área em cinza escuro representa o estado de São Paulo, utilizado nas buscas do Facebook, e a área em cinza claro, o Brasil, utilizados nas buscas do Twitter. 29
- Figura 2. Detalhamento dos procedimentos metodológicos adotados para este estudo. A figura está dividida entre as redes sociais Facebook e Twitter..... 31
- Figura 3. Agrupamento de categorias para postagens coletadas no Facebook. 34
- Figura 4. Proporção de postagens por categorias, onde: (1) Ação de conservação ou evento público, (2) Atropelamento, (3) Avistamento de fauna em área protegida, (4) Avistamento de fauna fora de área protegida, (5) Cativeiro, (6) Dano infligido à fauna ou caça, (7) Educação, (8) Informação falsa, (9) Invasão de propriedade por fauna, (10) Registro científico, (11) Resgate, tratamento ou soltura de fauna e (12) Saúde de fauna silvestre. O último gráfico corresponde ao total com todas as espécies juntas 43
- Figura 5. Proporção de postagens por tipos de mídia para cada uma das espécies na amostra do Facebook. Os ícones de mídia descrevem, respectivamente, da esquerda pra direita, texto, foto e vídeo. 45
- Figura 6. Histogramas com representação do número de palavras por espécie, incluindo mediana, desvio padrão e número amostral..... 46
- Figura 7. Linha temporal de postagens por categorias de postagens. 48
- Figura 8. Linha temporal de postagens por categorias de postagens para espécies do estudo..... 49
- Figura 9. Distribuição de postagens por municípios e espécie. A legenda apresenta os percentuais de postagem por município por espécie..... 50
- Figura 10. Categorias de usuário por espécie. 51
- Figura 11. Variáveis de engajamento para onça-pintada. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários. 52
- Figura 12. Variáveis de engajamento para a onça-parda. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas, em ordem da esquerda para direita os

símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.....	53
Figura 13. Variáveis de engajamento para o bugio. A representação das variáveis está em escala logaritmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas, em ordem da esquerda para direita os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.....	54
Figura 14. Variáveis de engajamento para o bicho-preguiça. A representação das variáveis está em escala logaritmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas, em ordem da esquerda para direita os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.....	55
Figura 15. Proporção de postagens por tipos de mídia para cada uma das espécies na amostra do Twitter. Os ícones de mídia descrevem, respectivamente, foto e texto.	61
Figura 16. Histogramas com representação do número de palavras por espécies, incluindo mediana, desvio padrão e número amostral.	62
Figura 17. Linha temporal de postagens do Twitter para onça-pintada e onça-parda.	64
Figura 18. Variáveis descritivas de usuários. A representação das variáveis está em escala logaritmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, onde 1 – seguidores e 2 – amigos.....	65
Figura 19. Variáveis de engajamento para o Twitter. A representação das variáveis está em escala logaritmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas, em ordem da esquerda para direita, os símbolos representam: compartilhamento e favoritado/like.	66

LISTA DE TABELAS

Tabela 1. Relação de variáveis coletadas manualmente no Facebook por dimensão dos dados.	33
Tabela 2. Categorias obtidas a partir de dados do Facebook.....	35
Tabela 3. Relação de variáveis obtidas no Twitter.	38
Tabela 4. Ranking de palavras por espécie, incluindo proporção relativa ao número total de palavras encontrada por espécie	45
Tabela 5. Resultados de regressão linear múltipla utilizando engajamento como variável resposta. Onde slope=coeficientes da regressão, IC=intervalo de confiança, p=significância, GL=graus de liberdade, R ² =coeficiente de determinação e (*) quando uma categoria é utilizada como parâmetro para regressão.....	56
Tabela 6. Ranking de palavras por espécie, incluindo proporção relativa ao número total de palavras encontrada por espécie	63
Tabela 7. Resultados de regressão linear múltipla utilizando engajamento como variável resposta para o Twitter. Onde slope=coeficientes da regressão, IC=intervalo de confiança, p=significância, GL=graus de liberdade, R ² =coeficiente de determinação e (*) quando uma categoria é utilizada como parâmetro para regressão.	68
Tabela 8. Resultados de desempenho para categorização da base do Twitter com a utilização de algoritmos.	68

1. INTRODUÇÃO

A explosão no uso da internet tem orquestrado uma verdadeira revolução na forma como pessoas se conectam. O desenvolvimento da web foi responsável por severas transformações na maneira como pessoas se relacionam com tópicos de interesse e, por consequência, umas com as outras (NIE; ERBRING, 2002). Estima-se que, mundialmente, cerca de 4,33 bilhões de pessoas utilizam os serviços da web diariamente (CLEMENT, 2019). O advento da internet também culminou na disponibilização de serviços *online* que possibilitam troca e criação de comunidades, seja por meio de microblogs ou organização em redes, popularmente conhecido como redes sociais.

A ideia de rede social não é, todavia, exclusiva do que hoje entendemos por *websites* destinados a troca de mensagens ou experiências. A internet herdou um conceito local de redes de trocas estabelecidas entre pessoas em diferentes localidades do mundo (KAPLAN; HAENLEIN, 2010). O conceito original versa sobre a troca de informações ou experiências entre diferentes pessoas, o que localmente pode corresponder a redes entre pessoas que dividem um espaço geográfico como um bairro, ou mesmo o ambiente de trabalho. Além disso, o conceito herdado das ciências sociais foi expandido a níveis nunca antes vistos. A ideia de redes sociais, não necessariamente *online*, perpassa uma série de fenômenos sociais e uma tendência humana de organização em teias sociais. O desenvolvimento da web deste modo cria um espaço propício para propagação de conteúdo. Não por acaso, outros setores têm se aproveitado desta oportunidade para lidar com dinâmicas de engajamento e manifestação *online*.

A ideia de rede social que temos hoje é global e traz implicações sociais fortes para a cultura e norma social (WEBSTER, 2010). Redes sociais fazem parte do cotidiano e têm, cada vez mais, desempenhado papel relevante no desenvolvimento da opinião pública. Pessoas passam horas conectadas e o número de usuários em redes sociais margeia 3,5 bilhões de usuários. Mais especificamente no Brasil, o número de usuários de internet é de 149 milhões, dos quais 140 milhões são usuários ativos de redes sociais (HOOTSUITE, 2019), sendo Facebook (2,1 milhões de usuários ativos), Instagram (895 milhões de usuários ativos) e Twitter (251 milhões de usuários ativos) as principais redes sociais no contexto global (WE ARE SOCIAL; HOOTSUITE, 2019). Mais especificamente em relação ao Facebook,

o índice de alcance de usuários gira em torno de 130 milhões no Brasil, segundo relatório do Hootsuite de 2019.

Sites como Facebook e Twitter possuem grande apelo e permitem que pessoas se conectem com conteúdos que não seriam normalmente explorados. Com efeito, o Facebook, maior rede social no Brasil em número de usuários, é responsável por permitir o acesso da população à notícias, humor, cultura e serviço de informação em tempo real. Existe aí uma via importante que caracteriza também a fragilidade da informação compartilhada, já que nem tudo o que está disponível em redes sociais corresponde necessariamente à verdade (WINGFIELD et al., 2016). Setores mais tradicionalmente voltados para públicos de mercado tem explorado profundamente as oportunidades disponíveis em sites como Facebook e Twitter. Por exemplo, Sajid (2016) afirma que, a partir de redes sociais, perfis podem ser traçados e estratégias de alcance podem ser desenvolvidas. Outros, como Harlow (2011), se preocupam com as transformações sociais e potencial de influência que redes sociais podem desempenhar.

Neste contexto, a conservação da biodiversidade carece de atenção e muitas vezes tem pouca força frente à população que nem sempre possui interesse direto no assunto. Iniciativas vêm sendo desenvolvidas para analisar efeitos de redes sociais sobre a problemática da conservação. Duas escolas proeminentes neste campo correspondem ao que é conhecido como conservação digital (DI MININ et al., 2015; TOIVONEN et al., 2019) e, mais recentemente, ao termo culturômica da conservação, o qual vem ganhando espaço. Estas duas frentes têm explorado dados digitais, sociais ou não, buscando entender e traçar perfis que possam, em última análise, servir à conservação de espécies. No primeiro caso, a ideia de conservação digital perpassa diferentes campos, quase sempre lidando com a questão tecnológica frente ao campo da conservação. Mais especificamente, quando tratam de redes sociais, Toivonen et al. (2016) foram capazes de sintetizar avanços relacionados a este campo do conhecimento e caracterizar como não é homogêneo o conjunto de metodologias e iniciativas. Por outro lado, a culturômica da conservação (LADLE et al., 2016) trata da análise metódica de um grande aporte de dados de texto e termos de busca para compreender a relação entre pessoa e conteúdo representacional relacionado à uma espécie, além da significância de determinados temas e sua saliência na cultura. Além disso, outras iniciativas não vinculadas necessariamente a nenhuma das duas escolas também ganham espaço

e possibilitam o desenvolvimento de diferentes metodologias possíveis para lidar com essas informações.

A diversidade de abordagens se dá por duas possibilidades primordiais no tratamento e aquisição de dados de redes sociais. Primordialmente, é possível obter dados em buscas não-automatizadas (manual) e automatizadas, cada qual com suas vantagens e desvantagens. A busca manual apresenta limitada capacidade de obtenção de uma gama grande de informação, é laboriosa e limitada em volume. Entretanto, trabalhos elaborados com obtenção manual de informações trazem grande possibilidade no agrupamento de dados e o desenvolvimento de modelos teóricos replicáveis que, por sua vez, podem alimentar sistemas automatizados (CAMPBELL; ENGELBRECHT, 2018; DI CAMILLO et al., 2018). A busca automatizada se aproxima mais do conceito difundido *Big Data*, onde grandes bancos de informação podem ser processados e são caracterizados pela velocidade com a qual são gerados, seu volume, variedade e veracidade. A busca automatizada é capaz de criar grandes bancos de dados com séries temporais e seleção completa do acervo disponível em sites específicos (google, portais de notícias, redes sociais, etc). Entretanto, lidar com grandes volumes de dados pode ser problemático na elaboração de modelos pertinentes, para entender do que necessariamente os dados tratam (FOSTER et al., 2016). Modelos menores, acurados e precisos, são essenciais para análises com maior quantidade de informação, para seu tratamento e limpeza.

De todo modo, redes sociais permitem acessar a relação pública com diversos conteúdos através do engajamento que estes conteúdos recebem. Em alguns trabalhos é possível compreender como pessoas percebem ou lidam com espécies (FUKANO; SOGA, 2019), em outros, como conteúdos relacionados a caça e depredação da fauna impactam pessoas. Além disso, usuários podem, a partir de posts, revelar comportamentos ilegais relacionados à fauna. Através de métricas próprias, redes sociais permitem acesso ao engajamento público em conteúdo relacionado à conservação. O termo engajamento ganhou força a partir de sua adoção por escolas, principalmente associadas ao marketing digital e relações públicas (GAMBETTI; GRAFFIGNA, 2010). O engajamento *per se* é também uma interessante ferramenta para acessar a importância dada por segmentos sociais a tópicos (KHAN, 2017). No caso específico da biodiversidade, é possível saber como pessoas reagem ou elegem temas relacionados à conservação. É necessário,

porém, manter em mente as diferentes dimensões associadas a conteúdo gerado em redes sociais. Uma forma de abordagem é a agregação das informações em cinco camadas, conhecidas como 5 w's: onde uma informação foi gerada, por quem, quando, como e porque (DI MININ et al., 2015). Esta categorização tem classicamente respondido a compreensão de conteúdo, sendo uma técnica comum a campos como jornalismo e ética. Observar a dimensão dos dados sob esta ótica pode ajudar a entender padrões de engajamento e auxiliar na sua utilização.

Mais especificamente, quando falamos da fauna brasileira, existe uma lacuna grande em análises que explorem características dos dados e engajamento *online* para espécies. Entender a dinâmica característica do conteúdo/engajamento associada às espécies pode ser chave na elaboração de estratégias de conservação digital, importantes para organizações dedicadas à conservação de espécies. Este trabalho tem como objetivo, desta forma, compreender como a dimensão de dados no Facebook e Twitter se dá, suas características e sua relação com engajamento para quatro espécies icônicas da fauna de mamíferos terrestres de médio-grande porte: onça-parda (*Puma concolor*), onça-pintada (*Panthera onca*), bugio (*Alouatta* spp.) e bicho-preguiça (*Bradypus* spp.; *Choloepus* spp.).

As onças são grandes símbolos da fauna brasileira. A onça-pintada possui distribuição geográfica que vai desde o norte do México até parte da Argentina (IUCN, 2016) e é considerada predador de topo em boa parte de sua distribuição. A onça-pintada vêm sendo representada ao longa da história como um símbolo de força e poder (CRISTANCHO; VINING, 2004; SHEPARD, 2014). A onça-parda é igualmente importante na cultura brasileira, embora sua representação seja distinta da que encontramos para onça-pintada, com percepções menos intensas de medo em relação a onça-parda (CONFORTI; AZEVEDO, 2003). A espécie apresenta maior distribuição e uso do espaço, sendo mais generalista quando comparada à onça-pintada (IUCN, 2019). Em regiões onde apresentam alta densidade de indivíduos, tais espécies são associadas a danos às atividades pecuárias e perda de rebanhos (PALMEIRA et al., 2008; MIOTTO et al., 2012). Apesar disso, ambas as espécies demonstram ser atrativas em relação a população que não reside em regiões onde os animais ocorrem em maiores densidades.

O termo bugio compreende espécies do gênero *Alouatta* de macacos neotropicais com ampla distribuição, do México ao norte da Argentina, com 10 espécies (ICMBIO, 2019; IUCN, 2019). O gênero, embora não possua o mesmo

apelo que espécies de grandes felinos, se enquandra na afeição gerada por características filogeneticamente próximas à humanos (GUNNTHORSDOTIRR, 2001; BATT, 2009). Além disso, primatas possuem grande adaptabilidade de hábitos à ambientes com presença humana (MCLENNAN et al., 2017). Desta forma, o bugio possui uma representação cultural baseada em seus traços próximos de seres humanos, sendo uma espécie carismática, com grande atração por parte do público. O termo bicho-preguiça também é utilizado para descrever espécies do gênero *Bradypus* e *Choloepus* distribuídos da América Central até a América do Sul, do qual fazem parte 5 espécies (ICMBIO, 2019; IUCN, 2019). O grupo das preguiças se enquadra no efeito comum para grandes mamíferos em relação à seu papel enquanto espécie carismática (DUCARME et al., 2012). De todo modo, estudos que avaliaram a relação entre pessoas e preguiças encontraram alto nível de tolerância e apreço pela espécie (PEREIRA et al., 2018). Além disso, preguiças são amplamente conhecidas e embutidas no imaginário coletivo, principalmente, por serem relacionadas à traços da personalidade humana por seu comportamento vagaroso (GREANEY et al., 2016).

Os resultados aqui obtidos poderão ser úteis à conservação aplicada, apresentando potencial para uso de dados de redes sociais na conservação de espécies da fauna. Deste modo a questão central deste trabalho é: quais são os tipos de informação disponíveis em duas redes sociais (Facebook e Twitter) relacionados à estas espécies? E como essas informações podem ser analisadas para servirem de subsídio à conservação?

O objetivo geral deste trabalho é examinar padrões em bases de dados em relação a quatro das cinco diferentes dimensões das informações obtidas em redes sociais (o que, quando, onde e quem). Os objetivos específicos são: (1) explorar maneiras de obtenção de dados em redes sociais; (2) determinar padrões nos dados obtidos com auxílio de análises quantitativas descritivas; (3) avaliar engajamento como uma quinta dimensão dos dados e (4) avaliar a possibilidade de categorizar a base de dados do Twitter de maneira automatizada (não manual).

2. REVISÃO BIBLIOGRÁFICA

Convivemos com a internet há, pelo menos, 29 anos. Desde então, sua existência tem trazido severas mudanças em nossa sociedade. É notório que a criação da internet só se fez possível graças ao desenvolvimento de tecnologias como o telefone, rádio e, por último, o computador. De toda forma, a internet representa um dos exemplos de maior sucesso do benefício de investimentos e compromisso com a pesquisa em comunicação. Desenvolvida inicialmente como uma maneira de trocar arquivos e informações, a internet tem como base uma criação com profunda conexão com estratégias políticas de proteção da informação, remontando aos tempos da ARPA net (rede de comunicação desenvolvida para gestão de dados de inteligência e pesquisa durante a guerra fria) (KEEFER; BAIGET, 2001; LEINER et al., 2009). Em 6 de agosto de 1991, a internet se tornou disponível para o público geral a partir de sua privatização, tornando-se uma ferramenta que interfere em todos os setores da sociedade, provendo para internautas a possibilidade de gerar e modificar conteúdo. Em 1996, a primeira pesquisa com usuários de internet contabilizava cerca de 40 milhões de usuários e, em 2013, este número era de cerca de 2,5 bilhões (CASTELLS, 2014). De fato, Castells (2014) afirma que vivemos em uma nova estrutura social, global, caracterizada pelo advento de uma cultura mutável e baseada na autonomia. Assim, estes são destacados como os pilares para o desenvolvimento da *web 2.0*, termo utilizado para denotar a segunda geração de comunidades e serviços *online*, fundamentado no conceito da rede como plataforma (FORREST, 2006). Ainda que uma definição controversa (BRODKIN, 2007), a idéia de um novo paradigma da internet pode ser utilizada como uma separação temporal em serviços web onde a participação de usuários é proeminente na geração e modificação de conteúdo.

Mesmo assim, o conceito de redes sociais parece perpassar a própria história da internet como a conhecemos, sendo uma forma de voltar ao desejo inicial da *world wide web* como uma plataforma que facilita a troca de informações entre usuários. Deste modo, redes sociais pertenceriam a um grupo de aplicações baseadas em internet, firmadas na base ideológica e tecnológica da *web 2.0* e que permitem a criação e troca de conteúdo gerado por usuários (KAPLAN; HAELEIN, 2010). Os mesmos autores classificam os tipos de redes sociais em blogs, sites de redes sociais (Facebook, Twitter, etc.), mundos sociais virtuais, projetos colaborativos (Wikipedia), comunidades de conteúdo (YouTube) e mundos de jogos virtuais. Sites de re-

des sociais são definidos como ambientes de média interação social, ricos em compartilhamento de mídia e com altos níveis de apresentação pessoal e compartilhamento de informações pessoais. É certo que interações, anteriormente limitadas ao ambiente real, ganham espaço para uma amplificação nunca antes vista. Whiting e Willians (2013) determinam que pessoas utilizam redes sociais por gratificações, como interações sociais, busca de informações, passatempo, entretenimento, lazer, comunicação, conveniência, expressão de opinião, compartilhamento de informações e monitoramento/conhecimento sobre outras pessoas.

Neste contexto, os últimos dez anos foram responsáveis pelo desenvolvimento das redes sociais como as conhecemos hoje. Desde a criação da rede social *Six Degrees* em 1997, até Facebook e Twitter se tornarem disponíveis para o público em 2006, outros vários sites de redes sociais e aplicativos se popularizaram e ganharam espaço com o público. Sites de redes sociais têm estado, então, presentes em setores como marketing (SAJID, 2016; ALALWAN et al., 2017), ciência política (STIEGLITZ; DANG-XUAN, 2013), negócios (JUSSILA; KÄRKKÄINEN, 2014) e saúde pública (TRACKERAY et al., 2012), como uma ferramenta majoritariamente estratégica de engajamento e análise de percepção pública.

Todos estes campos possuem em comum a necessidade de uma forma de operação interdisciplinar, já que o uso de dados de redes sociais normalmente envolve a conexão de diferentes campos do conhecimento. Normalmente, temos a necessidade da combinação do conhecimento em tecnologia e computação à ciências sociais e áreas de interesse, como, por exemplo, prevenção de desastres naturais (STIEGLITZ et al., 2014; FELT, 2016; ZHANG, et al., 2019). Esse campo novo e potencialmente importante é conhecido como *social media analytics* (STIEGLITZ et al., 2014) e trata principalmente do conjunto de operações e ferramentas para lidar com dados de redes sociais. O aporte de dados advindos destes sites é enorme e são necessárias estratégias para lidar com um volume grande de dados. Popularmente conhecido como *Big Data*, este termo se refere muito mais a um modelo de análises para lidar com a complexidade destes dados, caracterizados por terem grande volume, serem gerados com alta velocidade e alta variedade de informações (MCAFFE et al., 2012). O uso de *Big Data* também ganha cada vez mais espaço em áreas como marketing, epidemiologia e psicologia (BATRINCA; TRELEAVEN, 2018). A complexidade e quantidade dessas informações trazem desafios que justificam o desenvolvimento de um arcabouço teórico para obtenção, tratamento e definição de análi-

ses estatísticas possíveis para estas informações (DEKTHYAR, 2016). O autor resume o processo no ciclo da ciência dos dados:

(1) Definição da questão de pesquisa: a etapa de definição da questão de pesquisa é essencial para definição das estratégias subsequentes, desde a coleta à análise dos dados.

(2) Obtenção de dados: a obtenção dos dados obedecerá ao objetivo final da pesquisa, sendo necessária a adequação da metodologia e estratégia para obtenção dos dados. No que tange redes sociais especificamente, esta estratégia pode ser automatizada ou não automatizada, o que traz implicações para as outras etapas.

(3) Limpeza de dados: a etapa de limpeza dos dados é completamente diferente no método não automatizado e automatizado. No primeiro caso, a coleta sistemática dos dados já é responsável por agrupá-los nas categorias principais, sendo a limpeza apenas a detecção de incongruências, adaptação do conteúdo coletado e eliminação de ruídos. Em relação à busca automatizada, os dados carecem de receber maior atenção nesta etapa, visto que não existe um filtro prévio para a base de dados coletada.

(4) Modelagem de dados: a modelagem dos dados nesta etapa é mais teórica, com agrupamento das informações e elaboração de modelos com intuito de responder às questões básicas do experimento. Neste caso, a modelagem foi previamente estabelecida com base nos dados obtidos no Facebook, agrupamento de informações com base nas cinco dimensões e separação de métricas da rede social para engajamento. O modelo foi então aplicado à base do Twitter, para reconhecimento de métricas semelhantes e definição de barreiras que necessitam de aprofundamento para comparação entre as bases.

(5) Análise de dados: a etapa de análise dos dados corresponde a análises que expliquem o que é compreendido com as questões base do estudo.

O termo ciência dos dados se refere à prática de obter informações importantes através de um conjunto de dados, utilizando ferramentas que possibilitem lidar com a complexidade de grandes conjuntos de dados. Para Batista, Corrêa e Palanisamy (2016), a partir dos processos de análises advindos da ciência dos dados, poderemos lidar com dispositivos que possibilitem análises em tempo real. A ciência dos dados é uma área de pesquisa multidisciplinar que possibilita a indivíduos de diferentes formações utilizarem este tipo de análise para avanços consideráveis em suas áreas de interesse.

De fato, Cleveland (2001) ressalta a necessidade de um conjunto de habilidades versáteis no trabalho com ciência dos dados. De todo modo, o próprio conceito para ciência dos dados ainda está se estabelecendo, sendo que a utilização de ciência dos dados, dentro da perspectiva científica, já está incorporada como uma das frentes deste campo, segundo o *Committeon Data Science and Technology* (CODATA).

Na conservação, ainda é tímido o avanço do uso de dados de internet para inferir padrões sociais relacionados à conservação de espécies ameaçadas. Os esforços para lidar com este tipo de análise ainda estão em escala ascendente e se estabelecendo enquanto um campo da ciência da conservação. É necessário destacar o papel do trabalho de Di Minin et al. (2015) ao levantar as oportunidades em se utilizar conteúdo de redes sociais para trabalhos de conservação. Um crescente de trabalhos vem trazendo contribuições relevantes para o desenvolvimento deste campo. Toivonen et al. (2019) examinou diferentes estudos utilizando redes sociais e encontrou uma gama potencial de aplicações, além de constatar a heterogeneidade entre abordagens e métodos para extração, tratamento e obtenção de resultados. Em diferentes trabalhos, o termo conservação digital ganha força como uma possível alcunha a este campo em ascensão (DI MININ et al., 2015, TOIVONEN et al., 2019; FINK, et al., 2019).

De todo modo, por corresponder à uma ciência que emerge de diferentes pontos o trato de dados de redes sociais ganha abordagens e considerações complementares de diferentes autores. Do mesmo modo, é notória a contribuição de Lacle et al. (2016) no desenvolvimento do conceito de cultorômica da conservação, que, de modo simples, trata da manifestação cultural relacionada à biodiversidade associada a grandes corpos textuais. O mesmo autor estabelece a idéia de universo representacional para aspectos e espécies da biodiversidade, que geralmente estão relacionados com uma percepção macro da população a respeito daquele aspecto/espécie. Semelhante a outras vertentes conhecidas do que se estabeleceu como cultorômica, ou a observação de padrões culturais e sociais através de grandes conjuntos de texto em diferentes áreas, como as ciências sociais (SHERREN et al., 2017).

Entre outras utilidades aos dados de redes sociais, podemos citar o monitoramento de mudanças no interesse público em relação a espécies (SORIANO-REDONDO et al., 2018), monitoramento do interesse em relação às espécies em

áreas protegidas (HAUSSMANN et al., 2018), análise automatizada de sentimentos em relação à fauna ou ecossistemas (FINK, 2019; BECKEN et al., 2017). Neste ponto, são destacadas duas diferenças substanciais na maneira como lidar com dados de redes sociais; Alguns estudos optam por lidar com conjunto de dados coletados de maneira não automatizada (manual) (CAMPBELL; ENGELBRECHT, 2018; DI CAMILLO et al., 2018), enquanto outros trabalhos utilizam metodologia automatizada (DI MININ et al. 2019; BECKEN et al. 2017). Segundo Poorthuis et al. (2016), ganha-se muito em se utilizar analistas para compreender padrões em conjuntos de dados menores, embora o volume de dados atualmente obtido requeira esforços que só podem ser alcançados utilizando modelos automatizados para análise de informações. A junção dessas duas formas no processamento dos dados poderia contribuir para que processos automatizados ganhem em acurácia e adaptabilidade a diferentes realidades de conservação.

A dimensão humana da conservação é um campo que vem ganhando cada vez mais atenção e desenvolvimento, com o objetivo de inserir a relação humano-fauna na biologia da conservação. Reverter o *status* de espécies ameaçadas pode parecer, simplesmente, uma questão de manejar adequadamente a paisagem ou suas populações para atingir uma determinada meta de conservação. Mas, quando inserimos o complexo prisma dos efeitos da interação humano-fauna, podemos perceber a necessidade de se trabalhar a esfera humana na conservação (DECKER et al., 2002). A fauna, de modo geral, é capaz de despertar efeitos nas populações humanas, que em alguns casos é resultado da interferência direta da fauna no modo de vida humano, ou apenas a constatação da sua existência (RILEY et al., 2002).

Além disso, o trabalho com redes sociais se aproxima de campos já conhecidos da conservação, como as dimensões humanas na conservação de fauna. Este campo está originalmente muito atrelado ao manejo (as primeiras publicações tinham grande viés do manejo de populações para caça) (MANFREDO et al., 1989), até se expandir para outros âmbitos, como turismo e recreação, interface doença-homem-fauna e conflitos humano-fauna (MANFREDO, 2008). Assim, estas análises puderam ser aperfeiçoadas, tornando-se mais sofisticadas, aprofundando o entendimento de variáveis, como percepção e tolerância, e adotando conceitos e metodologia da psicologia e ciências sociais (MANFREDO et al., 1989; MANFREDO et al., 2008).

As redes sociais podem prover variáveis que complementam variáveis clássicas de dimensões humanas, como engajamento. Engajamento, segundo Gluck (2012), faz referência a uma ação iniciada por um usuário de rede social, que pode levar à “co-criação” de valor (BRODIE et al., 2013). Outros conceitos para engajamento o determinam como um conceito multidimensional, que faz referência à comportamento, funções cognitivas e emocionais, através de ações, pensamentos e sentimentos. Para Khan (2017), o conceito pode ser dividido entre engajamento “através de um clique”, chamado de engajamento de participação, referindo-se à *likes*, e compartilhamentos ou engajamento através de visualização e leitura, chamado de engajamento de consumo. É interessante traçar um paralelo com as características das diferentes métricas de engajamento com variáveis comuns a dimensões humanas, como atitude, percepção e tolerância, mas geradas e acessadas em uma esfera diferente, o universo digital. A junção destas, e outras diferentes ferramentas, é indicada como etapa fundamental, ao que Marchini et al. (2019) tratam como planejamento para coexistência. Planejar para coexistência é integrar informações e ferramentas disponíveis pela dimensão humana às ferramentas utilizadas pelo planejamento da conservação e, neste caso, poderia se incluir também dados de redes sociais.

3. MATERIAIS E MÉTODOS

3.1 Plataformas de busca

As redes sociais selecionadas para este estudo foram Facebook e Twitter. Mundialmente conhecidas, as duas redes possuem desafios e potenciais distintos de uso (RUAS et al., 2019). A rede social Facebook foi inicialmente planejada como espaço de troca por alunos de um campus universitário. Em 2006, a rede social foi aberta ao público e, em 2018, a companhia alegou ter pelo menos 2,3 bilhões de usuários ativos ao menos uma vez por mês (GEBEL, 2019). O trabalho com dados do Facebook é laborioso, já que a rede social não permite acesso a seus dados, desde uma decisão por restrição dos dados em 2018 (FACEBOOK, 2018), restringindo também procedimentos de extração automatizada. Algumas companhias são conhecidas por manter uma política de compartilhamento de dados por meio de interfaces de programação (API, do inglês *Application Programming Interface*), como é o caso do Twitter. Deste modo, muitos dos estudos feitos com base no Facebook necessitam da utilização de artifícios como a coleta manual de dados (PAPWORTH et al., 2015).

O Twitter foi lançado com proposta diferente do Facebook, constituindo uma rede social originalmente criada para *microblogging*, consagrada pelas características dos *tweets*, que correspondem à textos curtos e diretos. O Twitter possui uso mais restrito no Brasil, quando comparado ao Facebook. A rede social tornou-se aberta ao público em 2006 e, desde sua criação, ganhou extensa notabilidade e popularidade mundial. Diferente do Facebook, estudos utilizando Twitter tendem a dispor de uma gama maior de recursos para obtenção e armazenamento de dados, já que a rede social disponibiliza um serviço de API exclusivo (TWITTER, 2019).

3.2 Área de estudo

Definimos áreas distintas para cada plataforma de busca (Figura 1). Para as buscas realizadas na plataforma Facebook, a área definida foi o estado de São Paulo. Optamos por limitar o espaço territorial da busca no Facebook para evitar a pulverização dos dados, principalmente pelas limitações na obtenção de resultados pelo algoritmo da rede social.

O estado de São Paulo é considerado o estado brasileiro mais desenvolvido e abriga em sua extensão um dos maiores maciços de Mata Atlântica no Brasil. Sua extensão é de 248,209 km² e população estimada em 42,919 milhões de habitantes

em 2019 (IBGE, 2019). No estado, também é encontrado o segundo maior valor para população com acesso a internet entre os estados brasileiros (59,5%) (IBGE, 2011). O estado de São Paulo sofreu rápida conversão de seu uso da terra, principalmente devido ao crescimento populacional exacerbado e avanços agrícolas (DEAN, 1995). Atualmente, o estado possui intenso uso da terra baseado no cultivo sucroalcooleiro. Permanecem no estado cerca de 4,7 milhões ha de vegetação nativa, distribuídas em fragmentos menores (SPAROVEK et al., 2010). Em relação ao uso de redes sociais, os dados para o estado, especificamente, estão dispersos, mas, apenas na capital do estado, é estimado que 88,8% dos usuários de internet utilizam também redes sociais. Na capital paulista, o Facebook é a rede social dominante, com 43,3% dos usuários ativos afirmando priorizar seu uso (PREFEITURA DE SÃO PAULO, 2014).

Em relação ao Twitter, a seleção de dados teve como base o limite territorial do Brasil, já que a busca pela API não possibilitava a inclusão de uma palavra adjacente para localização, como “São Paulo”. O Brasil possui extensão territorial de ~8,514,876 km², com população total de 209,3 milhões (IBGE, 2019). Além disso, o país figura entre os 17 países mega diversos (MITTERMEIER, 1999), hospeda dois *hotspots* (Mata Atlântica e Cerrado) (MYERS et al., 2000), e seis biomas terrestres, sendo um deles considerado uma das grandes áreas selvagens do mundo (Amazônia) (MITTERMEIER, 2003). Segundo relatório do IBGE (2013), há sete anos o país contava com 35,5 % de seus domicílios com microcomputadores conectados à internet, e 86,4% com telefones móveis. No mesmo ano, 46,5% dos brasileiros usavam internet, sendo o maior acesso encontrado na população com mais escolaridade (90,2%). O Twitter, embora mundialmente conhecido, atinge apenas 5% da população que utiliza redes sociais (WE ARE SOCIAL; HOOTSUITE, 2019). A rede social conta com cerca de 8,5 milhões de usuários no Brasil.



Figura 1. Limites adotados para buscas em redes sociais neste estudo. A área em cinza escuro representa o estado de São Paulo, utilizado nas buscas do Facebook, e a área em cinza claro, o Brasil, utilizados nas buscas do Twitter.

3.3 Esquema de procedimentos metodológicos

Para esquematização do trabalho, nos propusemos a organizar os procedimentos metodológicos em um fluxograma (Figura 2). Inicialmente, identificamos a necessidade de determinar estratégias para obtenção e organização dos dados em ambas as redes sociais. As estratégias foram baseadas em conceitos de ciência dos dados, principalmente nas etapas descritas anteriormente por Dhakar (2010), onde, após a definição do tema, a forma como estas informações serão obtidas é definida. Para as duas redes sociais, nos baseamos em Toivanen (2019) para conduzir a

busca automatizada e não automatizada, principalmente na definição de variáveis de interesse e exemplos para organização de dados coletados manualmente.

A partir da obtenção das duas bases de dados, o passo seguinte foi aplicar conceitos de ciência dos dados para organização dos dados. Organizamos a base do Facebook com base no que obtivemos de exemplo na literatura, e adotamos o modelo de organização e separação das variáveis como exemplo para a base maior, do Twitter. As variáveis foram agrupadas em texto e outras variáveis extraídas, como data, nome de usuário e localização (quando disponível). No caso da base do Twitter, os dados foram obtidos a partir de linguagem de programação e hospedados na própria API.

A sistematização inicial dos dados foi baseada na proposta de dimensão de dados de redes sociais feita por Di Minin, Tenkanen e Toivonen (2015), sendo possível para este trabalho apenas trabalhar com 4 dimensões (“o que”, “quando”, “onde” e “quem”). Para as dimensões “o que”, “onde” e “quem” foi necessária a triagem e categorização das informações na base do Facebook. No caso específico da dimensão “onde”, não foi possível explorar a base de dados do Twitter, visto que a identificação e localização foram feitas manualmente. O aprendizado obtido a partir do Facebook foi utilizado para organizar e extrair informações da base do Twitter *a posteriori*.

Na sequência, análises foram elaboradas para cada dimensão dos dados, incluindo uma dimensão distinta (engajamento), em ambas as bases. Um quarto passo foi incluir um teste experimental com dados classificados da dimensão “o que”, aplicado na base crua do Twitter com utilização de algoritmos de aprendizado de máquina.

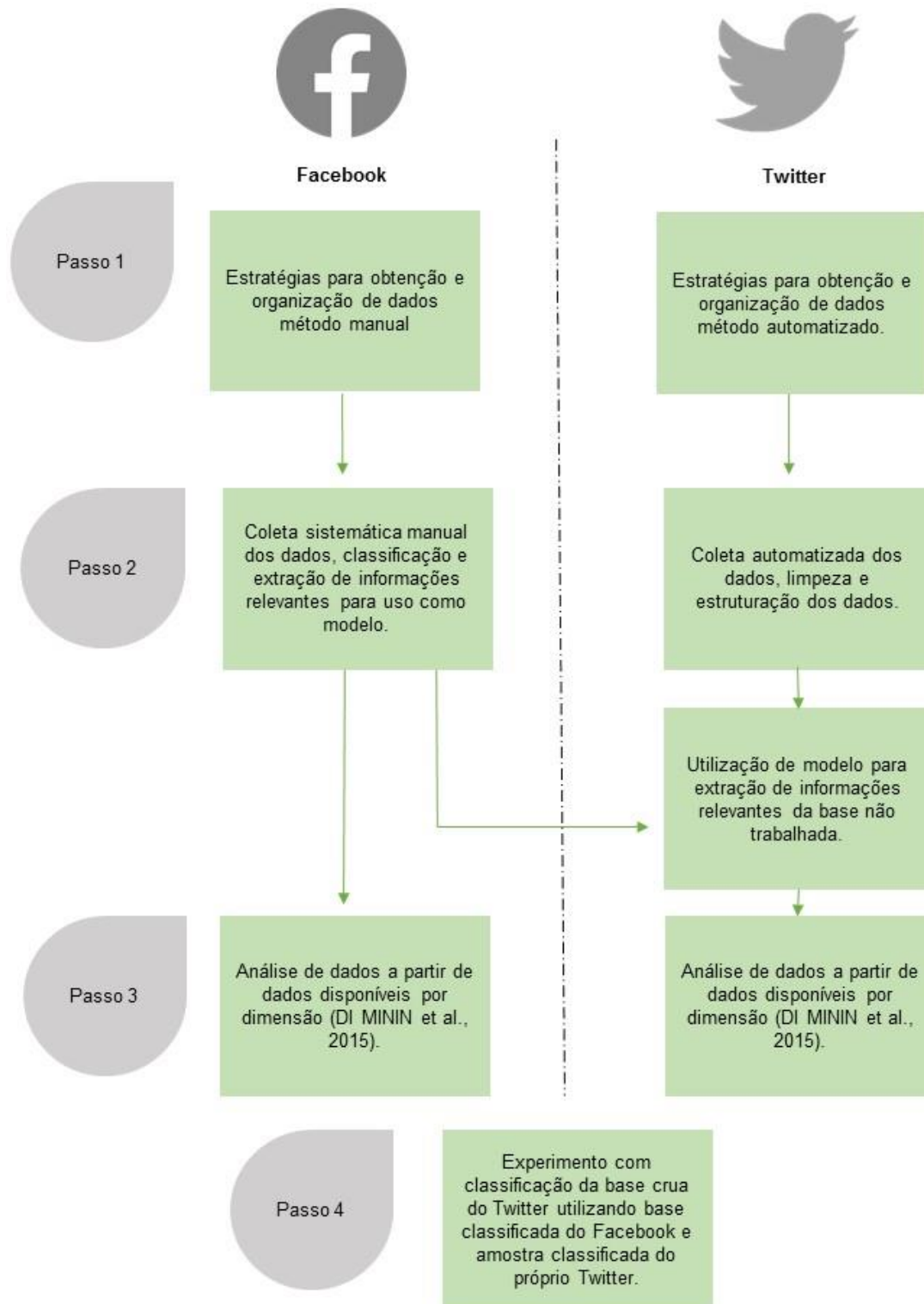


Figura 2. Detalhamento dos procedimentos metodológicos adotados para este estudo.

3.4 Coleta e organização de dados

3.4.1 Facebook

Através de contas pessoais, fizemos buscas utilizando os termos “onça-parda são paulo”, “onça-pintada são paulo”, “macaco bugio são paulo”, “bicho-preguiça são paulo”. Utilizamos contas pessoais de dois usuários diferentes, a fim de minimizar o efeito do algoritmo da própria rede social. O algoritmo do Facebook pode acabar causando uma repetição de tópicos ao favorecer a exibição de postagens que correspondam potencialmente aos interesses do usuário (SCHÄFER, 2020).

A partir dos termos de interesse, analisamos postagens públicas nas seções “Publicações”, “Fotos” e “Vídeos” da ferramenta de busca do Facebook. Consideramos somente as postagens relacionadas às espécies e eventos ocorridos no estado de São Paulo na coleta de dados, de forma que conteúdos com localização diferente e/ou com assuntos sem relação direta com as espécies foram descartados. Também excluimos da coleta postagens com inconsistência na localização ou com informações cuja veracidade não pôde ser comprovada.

Ao todo, foram coletadas 317 postagens, entre textos, fotos e vídeos, sendo 79 relacionados à onça-parda, 80 à onça-pintada, 78 ao bugio e 80 ao bicho-preguiça. O número de publicações coletadas foi determinado pelo efeito do algoritmo do Facebook e ausência de novas postagens. No caso específico da onça-pintada, o algoritmo da rede social gerou uma repetição de tópicos, de modo que obtemos uma grande quantidade de fotos similares, de indivíduos em cativeiro no Parque Zoológico de São Paulo. No caso do Facebook, além de variáveis comuns a muitas redes sociais como compartilhamento e número de comentários, a própria rede social disponibiliza reações. As reações correspondem a sentimentos e tem geralmente significados distintos, oferecendo camadas a mais de informação sobre o engajamento público em relação a determinado tópico. A relação de variáveis está descrita na Tabela 1.

Tabela 1. Relação de variáveis coletadas manualmente no Facebook por dimensão dos dados.

Variáveis	Descrição	Tipo de variável
<i>O que?</i>		
Categorias de postagem	Categorização baseada no tópico principal apresentado em uma postagem.	Categórica nominal
Tipo de mídia	Tipo de mídia compartilhada: texto, vídeo ou foto.	Categórica nominal
Texto completo	Texto completo encontrado em um post.	Texto
Contagem de palavras	Contagem de palavras de um post.	Numérico
<i>Quando?</i>		
Data	Marca temporal por data.	Numérica
<i>Onde?</i>		
Localização do evento da postagem	Localização encontrada na postagem relacionada ao evento que motivou a postagem. Não necessariamente fornecida e nem sempre relacionada com o local onde a postagem foi criada.	Localização (Municípios)
Localização precisa do evento da postagem	Localização precisa do evento que motivou a postagem, podendo ser acessada no nível de coordenadas geográficas.	Localização (Coordenadas geográficas)
<i>Quem?</i>		
Tipo de usuário	Categorização baseada no agrupamento de diferentes tipos de perfil de usuário.	Categórica nominal
Nome de usuário	Nome disponível na rede social. Não disponível no documento por razões éticas.	Texto
<i>Engajamento</i>		
<i>Like</i>	Representa apreciação pública a um determinado conteúdo. Likes são a variável mais comum, encontrados em diferentes plataformas de redes sociais.	Numérica
<i>Love</i>	Representa afeição do público em relação a um post. Love está associada com fortes sentimentos positivos em relação a um conteúdo.	Numérica
<i>Funny</i>	Representa reação a um post mais comumente ligado a humor.	Numérica
<i>Wow</i>	Representa surpresa em relação a um post. Esta métrica é ambígua e pode ter significado positivo ou negativo.	Numérica
<i>Cry</i>	Representa tristeza em relação a um post. É uma reação comumente negativa, mas pode variar em relação a contexto.	Numérica
<i>Angry</i>	Representa raiva do público em relação a um post. É considerada uma métrica que varia em relação à seu significado dependendo do contexto.	Numérica
Compartilhamentos	Compartilhamento é uma métrica de importância dada ao conteúdo, o público tende a utilizá-la para conteúdos considerados relevantes.	Numérica
Comentários	Comentários é uma métrica de interesse público, mas varia em relação à importância do conteúdo do post.	Numérica

Para gerar a variável “categorias de postagens”, optamos por uma categorização a partir das informações disponíveis (Figura 3). Com base em Bryman (2001),

os assuntos foram sumariamente agrupados utilizando critérios para sua organização. O texto principal foi dividido em tópicos relacionados à natureza do evento que motivou a postagem. Por exemplo, uma postagem relacionada à atropelamento foi colocada na categoria atropelamento. Ao todo, 19 categorias foram definidas com a primeira categorização, resultando em 12 categorias finais. Categorias com baixo número de postagens foram agrupadas por possuírem relação, restando apenas uma categoria com uma única postagem (Informação falsa), por não possuir relação com as outras categorias.

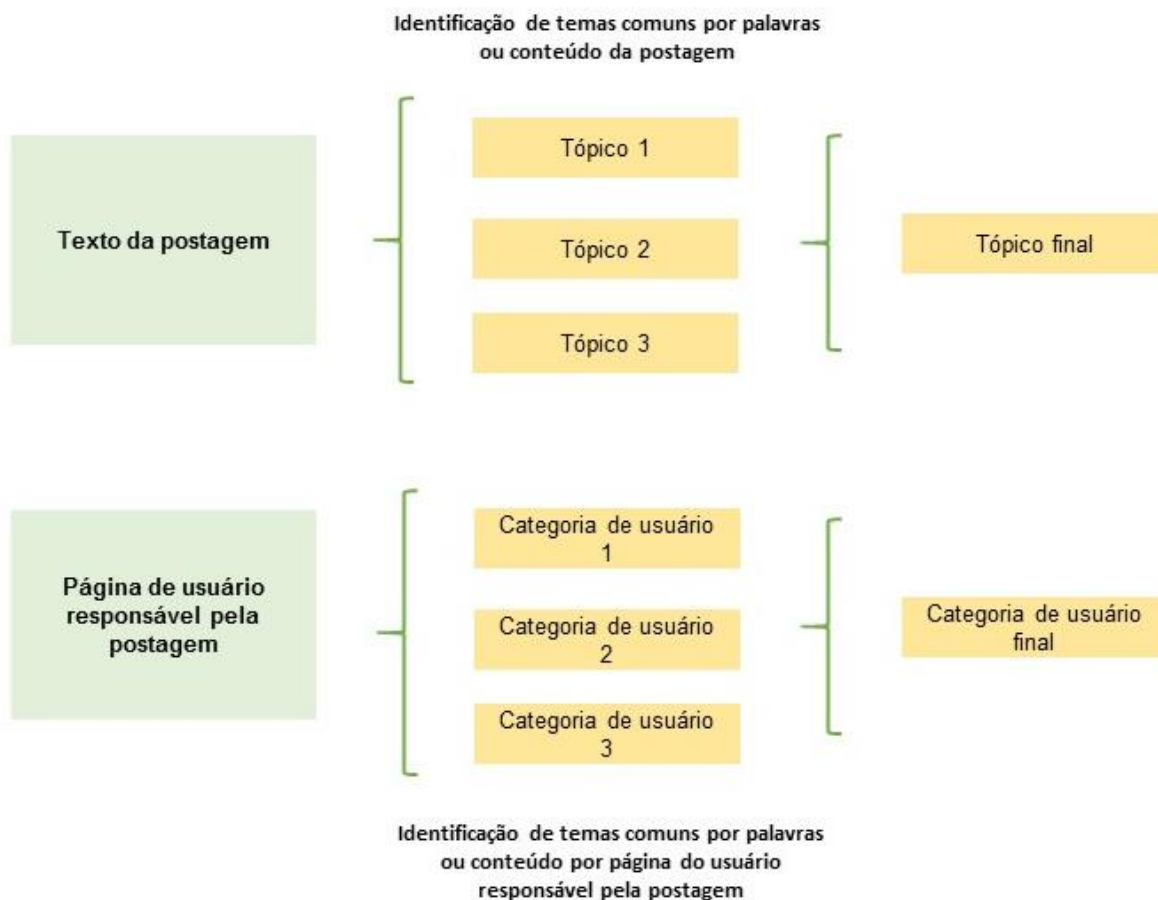


Figura 3. Agrupamento de categorias para postagens coletadas no Facebook.

Para a variável “tipo de usuário” seguimos o mesmo modelo de categorização para os tipos de usuários encontrados na rede social, com 19 categorias iniciais, restando 10 categorias finais. As categorias estão descritas na Tabela 2.

Tabela 2. Categorias obtidas a partir de dados do Facebook.

Identificação	Categoria	Descrição
		<i>Postagens</i>
1	Ação de conservação ou evento público	Postagens contendo: anúncio de ações feitas por institutos, ONGs, governo ou sociedade civil. Eventos relacionados à conservação (seminários, workshops).
2	Atropelamento	Postagens contendo: conteúdo relacionado á atropelamento (notícias, mídia gerada por usuário) e ações relacionadas à mitigação de atropelamentos.
3	Avistamento de fauna em área protegida	Postagens contendo: avistamento de fauna com identificação geográfica dentro de áreas protegidas.
4	Avistamento de fauna fora de área protegida	Postagens contendo: avistamento de fauna fora de áreas protegidas, assim sendo, áreas rurais, urbanas ou não urbanas.
5	Cativeiro	Postagens contendo: espécies expostas em cativeiro, cuidados em cativeiro como tratamento ou dieta; postagens de zoológicos ou instituições relacionadas ao cativeiro com foco na temática.
6	Outros danos infligidos à fauna ou caça	Postagens contendo: postagens relacionadas à caça, dano infligido a fauna direto ou indiretamente (como queimadas ou perda de indivíduos por ações de manejo agrícola).
7	Educação	Postagens contendo: iniciativas de educação ambiental, pedagógicas ou postagens com curiosidades ou informações relacionadas à espécie. Postagens podem ser parte de um projeto ou ação, ou apenas feitas espontaneamente por usuários.
8	Informação falsa (<i>fakenews</i>)	Postagens contendo: conteúdo falso relacionado à espécie ou reportando um acontecimento falso relacionados à espécie.
9	Invasão de propriedade por fauna	Postagens contendo: situações de fauna adentrando em propriedades, incluindo acidentes quando acontecerem.
10	Registro científico	Postagens contendo: achados científicos, registros de localização de espécie (ex: armadilha fotográfica) ou rastros. Resultados de projetos científicos ou conteúdo científico diverso compartilhado.
11	Resgate, tratamento ou soltura de fauna	Postagens contendo: ações de resgate, reabilitação de fauna em situações vulneráveis, soltura de fauna.

12	Saúde de fauna silvestre	Postagens contendo: notificação de animais mortos devido a doenças, explicação do risco e consequências de doenças relacionadas a fauna ou com fauna silvestre como parte de seu ciclo.
<i>Usuários</i>		
1	Associação/Fundação/Instituto/ONG	Páginas de representantes do terceiro setor. Esta categoria inclui apenas páginas que se caracterizam desta forma, incluindo páginas que não produzem diretamente conteúdo relacionado com conservação ou meio ambiente.
2	Instituição governamental	Páginas de prefeituras, governo ou outras instituições públicas.
3	Página conteúdo ambiental	Páginas que produzem ativamente conteúdo relacionado a meio ambiente ou conservação.
4	Página de entretenimento	Páginas de humor ou outros temas ligados a entretenimento.
5	Página de negócios	Páginas de empresas (pequenas a grandes) ou pequenos negócios e iniciativas.
6	Página de notícias	Páginas ligadas a vinculação de notícias ou imprensa. Podem ser locais ou de maior porte.
7	Página pessoal	Página de pessoa física.
8	Zoológico e Aquários	Página de aquários e zoológicos.
10	Outros	Páginas com perfil díspare, mas com número de postagens muito baixo para criação de uma nova categoria.

Para a variável “tipo de mídia”, não foi necessária nenhuma alteração, apenas sua contabilização. A variável “contagem de palavras” foi obtida a partir da análise do texto completo das postagens. Além disso, o texto completo foi incluído como variável, para que a frequência de palavras fosse analisada.

Em relação à dimensão “quando”, data e hora foram coletadas, mas, devido ao número amostral baixo e para comparação com a base do Twitter, optamos por trabalhar apenas com a data das postagens, que foram transformadas de dia-mês-ano para apenas mês-ano. Em relação à dimensão “quando”, conseguimos resultados apenas para base do Facebook. Agrupamos a variável em duas camadas: municípios onde os eventos que motivaram a postagem ocorreram e localização acurada quando possível e disponível no texto da postagem (p. ex., quilômetro de uma rodovia onde um atropelamento ocorreu).

3.3.2 Twitter

Para obtenção de dados do Twitter utilizamos uma API. Uma API, de modo simples, permite que desenvolvedores interajam com a plataforma da rede social por meio de programação (CLARK, 2004). A Twitter API é uma interface que permite que usuários cadastrados acessem serviços da rede social. A API permite a busca do histórico de *Tweets* (postagens na plataforma), *streaming* de *Tweets* (ou sua obtenção à medida que são gerados), a integração de serviços à plataforma *online* e publicação automatizada, dentre outros.

Neste trabalho, utilizamos a função de busca histórica de *Tweets* (*Search API*). A função permite que um conjunto de palavras chaves seja especificado para busca em toda a base de dados do Twitter, além de outras variáveis. O resultado da busca retornou um JSON (*Java Script Object Notation*) contendo as mensagens relacionadas à busca. A função oferece uma versão padrão e uma versão premium, sendo que a versão padrão permite apenas a busca de mensagens postadas no máximo de uma semana e a segunda com dois módulos de busca: últimos 30 dias e mensagens postadas a qualquer momento.

Especificamente em relação ao Twitter, fizemos sucessivas tentativas com termos para retorno de tweets. Inicialmente utilizamos a junção “nome da espécie + localização”, do mesmo modo que fizemos no Facebook. Esta estratégia não resultou em *tweets* relacionados à espécie. Fizemos todas as buscas em português, utilizando nomes relacionados à maneira como a espécie é conhecida no Brasil. De todo modo, fizemos uma busca inicial com 7 espécies de mamíferos (termos: onça-pintada, onça-parda, bugio, bicho-preguiça, veado, macaco-prego e gambá), após triagem de uma amostra de 100 tweets para cada busca, concluímos que o aproveitamento para três das espécies era excessivamente baixo e optamos por manter apenas duas: onça-pintada e onça-parda. Para onça-pintada e onça-parda encontramos 90% de aproveitamento, para as outras espécies respectivamente: bugio (15%), bicho-preguiça (7%), veado (10%), macaco-prego (6%) e gambá (4%).

O modelo de organização e limpeza dos dados trouxe algumas dificuldades, dado a quantidade de postagens extraídas das redes sociais. Foram no total 23.853 postagens extraídas para onça-pintada e 14.663 para onça-parda. Para que o conjunto pudesse ser analisado, algumas etapas de mineração de dados foram necessárias. Postagens incompletas ou com lacunas de informação foram descartadas na

etapa de limpeza. Dos resultados obtidos, optamos por fazer a seleção das variáveis que seriam consideradas, tanto para características das postagens, quanto para engajamento (Tabela 3). A partir destas variáveis, integramos as bases da onça-pintada com as de onça-parda, porém, devido às dificuldades para classificar as postagens, não avançamos na etapa de transformação dos dados.

Tabela 3. Relação de variáveis obtidas no Twitter.

Variável	Descrição	Tipo de variável
<i>O que?</i>		
Tipo de mídia	Tipo de mídia compartilhada: texto, vídeo ou foto.	Categórica nominal
Texto completo	Texto completo encontrado em um post.	Texto
Contagem de palavras	Contagem de palavras de um post.	Numérica
<i>Quando?</i>		
Data	Marca temporal por data.	Numérica
<i>Quem?</i>		
Seguidores	Número total de seguidores do usuário responsável pela postagem.	Numérica
Amigos	Número total de amigos do usuário responsável pela postagem.	Numérica
<i>Engajamento</i>		
Retweets	Semelhante a compartilhamento, retweets reflete o engajamento do usuário em compartilhar uma postagem ou informação.	Numérica
Favoritado/ Like	Representa apreciação pública á um determinado conteúdo. Likes são a variável mais comum, encontrados em diferentes plataformas de redes sociais. A variável favorite foi adaptada para like após 2015 no Twitter, com significado semelhante a variável favorite permitia o usuário também salvar postagens.	Numérica

3.5 Análise dos dados

Optamos por analisar, especificamente, cada dimensão dos dados por rede social. Os tratamentos foram comuns por dimensão, sendo abordados neste tópico de forma conjunta. Utilizamos estatística descritiva para a maior parte das análises, sendo que, para a organização dos dados e seu processamento, utilizamos o *software* R (R CORE TEAM, 2017) e Atlas ti.8 (FRIESE, 2019). Abaixo, segue a descrição de análise para cada dimensão:

O que: optamos por analisar a frequência por categoria e frequência geral para a variável “categorias de postagem”. Utilizamos histogramas baseados na frequência do número de palavras para representar à variável “contagem de palavras”, e contagem de postagens por mídia para a variável “tipos de mídia”. Utilizamos o

software Atlas ti.8 para analisar a frequência de palavras para todas as postagens por espécie, relacionada a variável “texto”. As análises com a base do Twitter seguiram o mesmo padrão.

Quando: para esta dimensão, elaboramos gráficos de séries temporais por espécie. No caso dos dados do facebook, utilizamos a categorização feita para as postagens para visualização de categorias com mais postagens.

Onde: para esta dimensão, na base do Facebook, filtramos dos resultados aqueles que possuíam localização exata do evento que motivou a postagem. Utilizamos o *software* ArcGis 10.3 (ESRI, 2015) e a base de municípios do IBGE (IBGE, 2019) para visualização de localização das postagens.

Quem: para esta dimensão, seguimos a mesma abordagem utilizada para a variável “categorias de postagem” com os dados do Facebook. Em relação aos dados do Twitter, optamos pela representação gráfica da distribuição dos dados para as variáveis “amigos” e “seguidores”.

Engajamento: para esta dimensão, seguimos a mesma abordagem de descrição estatísticas das frequências por variável. Além disso, testamos modelos de regressão linear múltipla com o intuito de investigar a influência da dimensão “o que” e “quem” em relação ao engajamento. Os modelos de regressão linear múltipla foram uma boa alternativa, por serem modelos simples que permitem a inclusão de variáveis categóricas nominais e numéricas. Os dados obedeciam à normalidade necessária para aplicação do método (UYANIK; GÜLER, 2013). Para o desenvolvimento dos modelos, utilizamos o somatório de engajamento por espécie, por exemplo, a soma total de reações, compartilhamento e comentários. Utilizamos o valor de p (probabilidade de significância) e o R^2 / R^2 ajustados (coeficientes de determinação) para verificar a significância dos modelos. As associações positivas ou negativas entre preditores e variável principal foram utilizadas a partir do *Slope* ou coeficientes da regressão.

3.4 Testes de categorização na base do Twitter utilizando aprendizado de máquina

Um último teste efetuado foi a verificação de sucesso de categorização por *machine learning*. A base de dados do Twitter, devido a seu volume, não pôde ser categorizada manualmente. Desta forma, resolvemos utilizar a base de dados classificada do Facebook como entrada para classificação da base do Twitter. Além disso,

fizemos uma classificação manual da base do Twitter para onça-parda e onça-pintada, com 303 e 359 postagens classificadas, respectivamente. Na base classificada do Twitter, o número de categorias foi reduzido a 9, sendo que as seguintes categorias não foram encontradas: educação, cativo, e ação ou evento de conservação.

Esquematizamos comandos para selecionar e avaliar modelos de predição de forma automatizada. Além dos algoritmos de *machine learning*, também foi necessário adotar técnicas de representação de texto para chegarmos em modelos finais. Os procedimentos foram feitos utilizando duas bibliotecas (pacotes) no python (*Gensim* e *Scikit-learn*). A partir da base de dados fornecida, o texto é transformado de formato texto para vetor, com o auxílio de *embeddings* (representação matricial do texto). A geração de *embeddings* é um processo não supervisionado, então não foi necessária a exclusão de postagens. Os algoritmos de transformação (*embeddings*) apresentam diferenças quanto à transformação de texto em matrizes que influenciam no desempenho final dos algoritmos de classificação.

Com isso, pudemos fazer o treino e avaliação dos algoritmos de *machine learning*, adotando a divisão da base de dados entre classificados e não classificados. Deste modo, treinamos classificadores para prever a variável resposta (categorias), assim pudemos obter a acurácia da previsão por algoritmo. A acurácia foi obtida através de validação cruzada com amostras obtidas aleatoriamente no conjunto de dados ou seguindo a ordem dos dados de maneira contínua (não aleatória) (HASTIE; TIBSHIRANI; FRIEDMAN, 2001). A validação cruzada foi feita utilizando o pacote *Scikit-learn*. Os algoritmos para transformação e classificação utilizados para a análise foram:

Algoritmos de transformação (Embeddings)

TF-IDF (*Term frequency – inverse document frequency*): Representação da matriz termo-documento com pesos calculados, onde não é necessária a utilização de parâmetros (JURAFSKY; MARTIN, 2000).

LSA (Análise de Semântica Latente): Execução com uso da matriz termo-documento (com pesos calculados por Tfidf). Neste teste, utilizamos um espaço dimensional de 100 e 300. Foi utilizada a reconstrução da matriz termo-documento como *embedding* (DEERWESTER et al., 1990).

TopicLSA: Alteração do algoritmo anterior, no qual, ao invés de utilizarmos reconstrução da matriz como *embedding*, utilizamos a matriz $V'S'$. O nome é utilizado devido ao embedding gerado não ser representado em relação ao vocabulário, e sim, às dimensões resultantes do método (chamadas pelo *Gensim* de tópicos). Para seleção do número de dimensões foram testados os mesmos valores do LSA (DERWESTER et al., 1990).

Classificadores

Naivebayes: algoritmo onde foi necessário testar se a distribuição condicional de valores é gaussiana, sem necessidade de outros parâmetros (ZHANG, 2004).

Regressão logística: no caso da regressão logística, o principal parâmetro utilizado é o fator de regularização, responsável por penalizar valores dos parâmetros do modelo. Utilizamos o valor 10 com uma baixa penalização aplicada (BÖHNING, 1992).

SVM (*Support vector machine*): no caso do SVM o que é mais importante é a penalização sobre classificações erradas e a função Kernel utilizada. No primeiro caso, adotamos os valores da regressão logística, com a diferença de que a lógica de penalização neste caso é inversa, assim, valores menores são mais penalizados. Adotamos a função linear para a Kernel (LORENA; DE CARVALHO, 2007).

Random Forest: para o algoritmo, é necessário definir o número de árvores de decisão. Escolhemos 500 árvores de decisão. Além disto, foi selecionado como parâmetro o critério de avaliação de quebra das árvores em número mínimo de amostras por folha, o valor adotado foi 5 (BREIMAN, 2001).

4. RESULTADOS

Facebook - *O que?*

Das 317 postagens obtidas no Facebook, encontramos três categorias principais: cativoiro (24,6%), avistamento de fauna fora de área protegida (21,1%) e resgate, tratamento ou soltura de fauna (16,7%) (Figura 4). Analisando por espécie, as duas principais categorias encontradas para a onça-pintada foram avistamento de fauna fora de área protegida (25,6%) e cativoiro (35,9%). Para a onça-parda, as duas principais categorias foram invasão de propriedade por fauna (21,5%) e registro científico (17,7%). Enquanto que para o bugio, as duas categorias principais foram avistamento de fauna fora de área protegida (25,6%) e saúde de fauna silvestre (35,9%). Para o bicho-preguiça, as duas categorias predominantes foram avistamento de fauna fora de área protegida (30,0%) e resgate, tratamento ou soltura de fauna (45,0%).

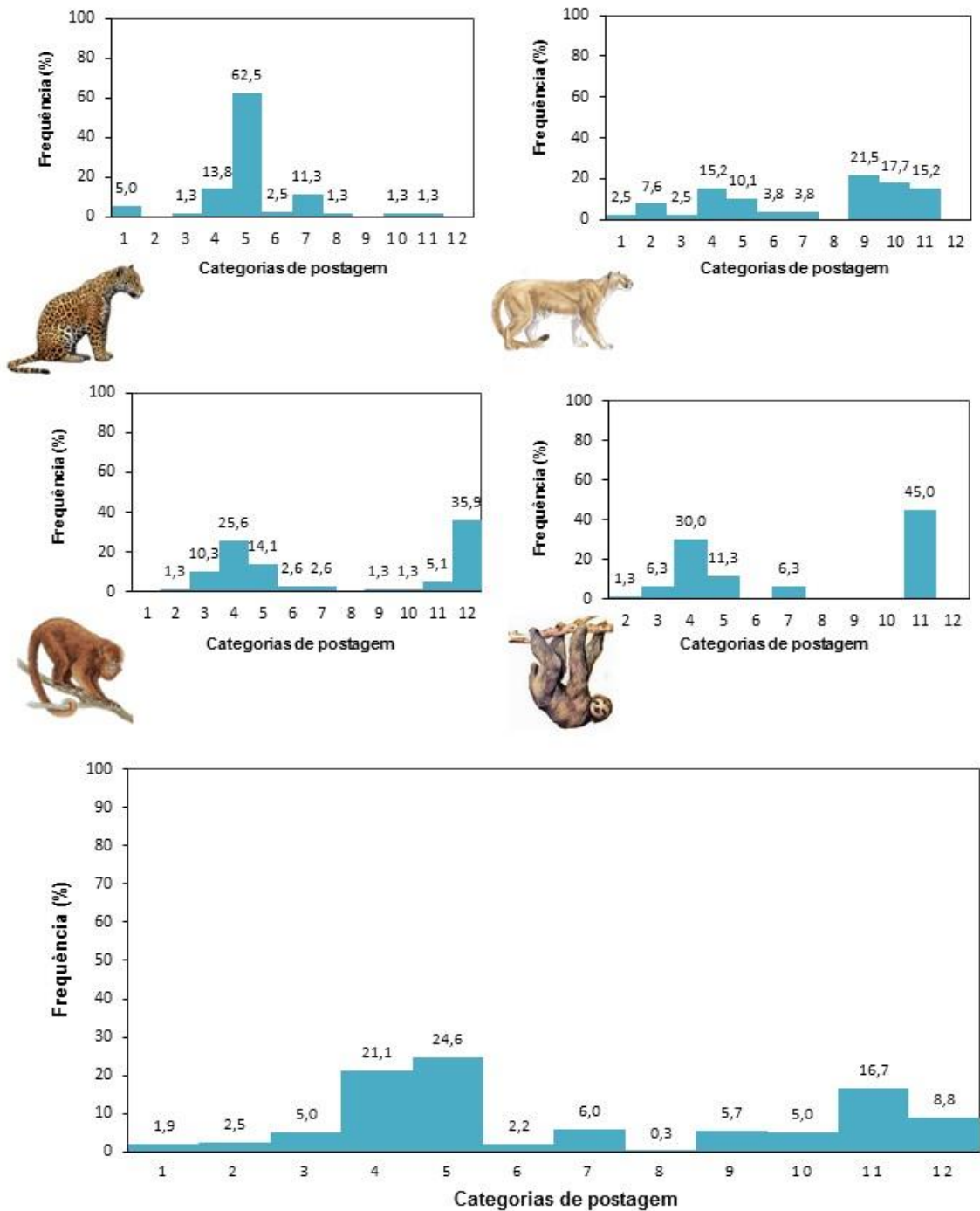


Figura 4. Proporção de postagens por categorias, onde: (1) Ação de conservação ou evento público, (2) Atropelamento, (3) Avistamento de fauna em área protegida, (4) Avistamento de fauna fora de área protegida, (5) Cativo, (6) Dano infligido à fauna ou caça, (7) Educação, (8) Informação falsa, (9) Invasão de propriedade por fauna, (10) Registro científico, (11) Resgate, tratamento ou soltura de

fauna e (12) Saúde de fauna silvestre. O último gráfico corresponde ao total com todas as espécies juntas.

Em relação aos tipos de mídia, 11,98% eram postagens apenas com conteúdo de texto, 68,40% incluindo foto e 18,61% incluindo vídeos. Tratando especificamente das espécies, a tendência de postagens incluindo majoritariamente fotos se manteve para onça-pintada, bugio e bicho preguiça. Para onça-parda, embora postagens com fotos tenham sido dominante, postagens contendo vídeos tiveram um alto valor de frequência (Figura 5).

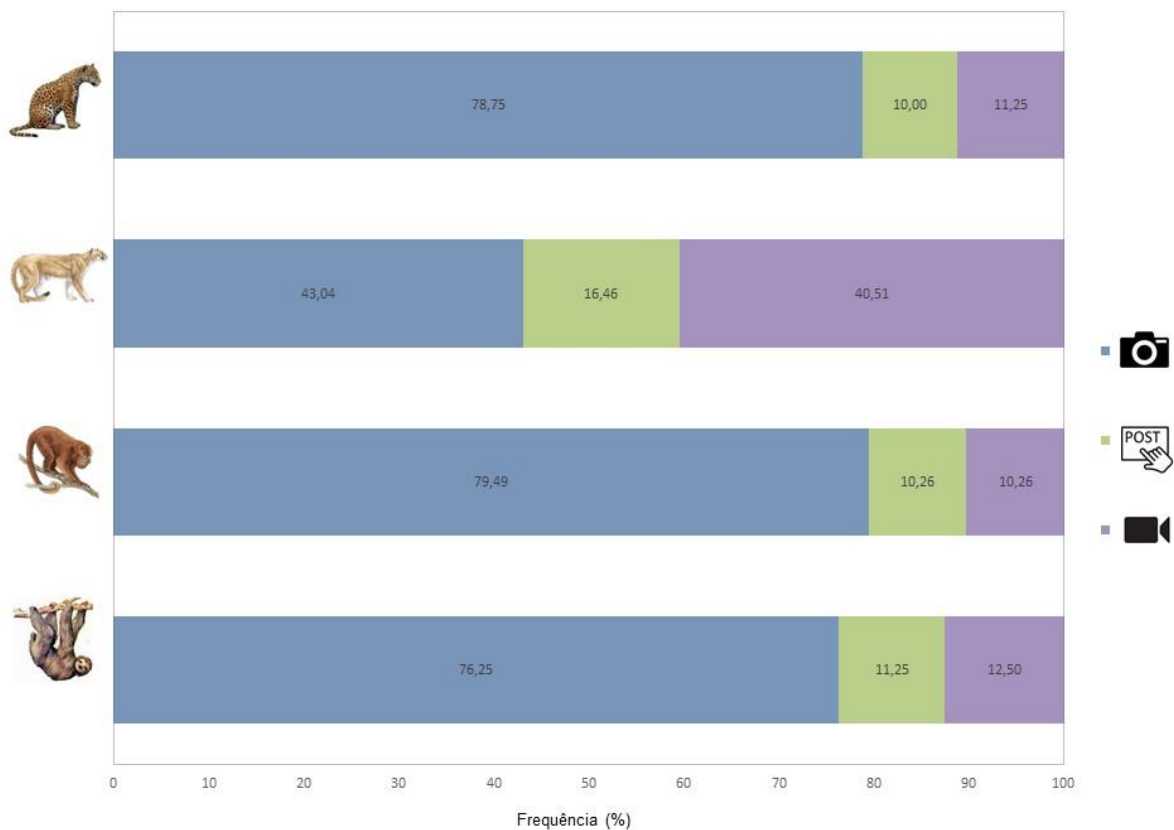


Figura 5. Proporção de postagens por tipos de mídia para cada uma das espécies na amostra do Facebook. Os ícones de mídia descrevem, respectivamente, da esquerda para a direita, texto, foto e vídeo.

A tendência geral para a variável contagem de palavras foi uma mediana de 26 palavras. Encontramos que os valores de mediana por espécie variaram entre 12 para a onça-pintada, 22 para a onça-parda, 37,5 para o bugio e 42,5 para o bicho-preguiça (Figura 6). Todas as espécies apresentaram valores *outliers*, com a distribuição geral dos dados se concentrando em torno de 150 palavras por postagem para onça-pintada e onça-parda, 200 para bugio e bicho-preguiça. Os valores totais

de palavras foram respectivamente 4.298 para onça-pintada, 3.470 para onça-parda, 7.179 para bugio e 6.002 para bicho-preguiça.

No que concerne aos *rankings* de texto, alguns termos apresentaram repetição entre espécies, sendo eles: zoo ou zoológico, parque, conservação, mata, ambiental, foto, filhote, interior, cidade e menções a sistema viário (estrada e rodovia) (Tabela 4). Em alguns casos, cidades, como Apiaí para onça-pintada e Piracicaba para onça-parda, apresentaram destaque no ranking. Outras palavras em destaque por espécie foram, para a onça-pintada: rodovia, filhote e grande; para a onça-parda: cidade, parque e secretaria; para o bugio: febre-amarela, saúde e morto e para bicho-preguiça: litoral, resgatado e estrada.

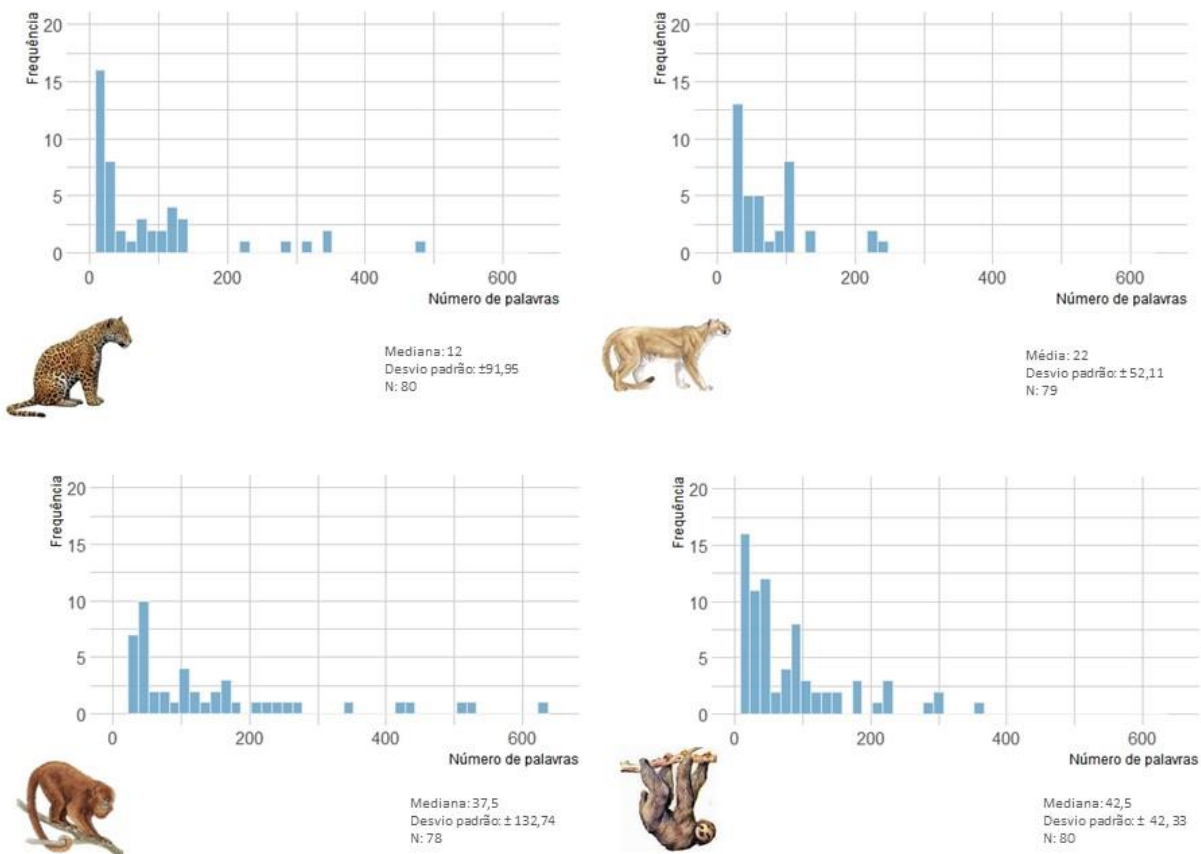


Figura 6. Histogramas com representação do número de palavras por espécie, incluindo mediana, desvio padrão e número amostral.

Tabela 4. Ranking de palavras por espécie, incluindo proporção relativa ao número total de palavras encontrada por espécie.

<i>Onça-pintada</i>			<i>Onça-parda</i>			<i>Bugio</i>		
Palavras	Contagem	Proporção (%)	Palavras	Contagem	Proporção (%)	Palavras	Contagem	Proporção (%)
Zoo	17	0,43	Mata	13	0,40	Febre	75	1,00
Parque	14	0,35	Interior	14	0,34	Amarela	74	1,00
Conservação	13	0,33	Zoológico	11	0,34	Parque	38	0,50
Mata	13	0,33	Ambiental	11	0,34	Saúde	32	0,42
Rodovia	12	0,30	Ambiente	9	0,28	Cidade	27	0,36
Apiáí	11	0,28	Fauna	9	0,28	Secretaria	22	0,29
Ambiental	10	0,35	Cidade	6	0,25	Vacina	22	0,29
Foto	10	0,25	Conservação	8	0,25	Cantareira	21	0,28
Filhote	9	0,23	Parque	8	0,25	Morto	20	0,27
Grande	9	0,23	Piracicaba	8	0,25	Zoológico	20	0,27
Registro	9	0,23	Secretaria	8	0,25	Doença	19	0,25
Felinos	8	0,20	Vídeo	8	0,25	Horto	19	0,25
Interior	8	0,2	Bombeiros	8	0,20	Jardim	19	0,25

Facebook - Quando?

Identificamos na linha temporal que os picos de postagem ocorrem entre setembro e novembro de 2017, sendo que a categoria saúde de fauna silvestre contou com 7 postagens, com um pico menor entre janeiro em fevereiro de 2016. É importante ressaltar que todas as postagens de saúde de fauna silvestre se referem à bugio (Figura 7). É notável também um segundo pico de postagens, ocorrendo entre junho e julho de 2019, com a categoria avistamento de fauna fora de área protegida, sendo importante para onça-parda neste conjunto de dados (Figura 8). Focando especificamente as espécies, podemos notar picos para a categoria educação entre dezembro de 2018 e abril de 2019, e avistamento de fauna em área fora de área protegida entre junho e agosto de 2019. Para onça-parda, o único pico que se destaca está relacionado à categoria avistamento de fauna fora de área protegida, entre novembro e janeiro de 2019 (Figura 8). Para o bugio, encontramos picos em outubro de 2017 e fevereiro de 2019, com postagens relacionadas à categoria saúde de fauna silvestre (Figura 8). Enquanto que, para o bicho-preguiça, a categoria resgate, tratamento ou soltura de fauna se manteve contínua, com um pico entre julho de 2018 e março de 2019 (Figura 8).

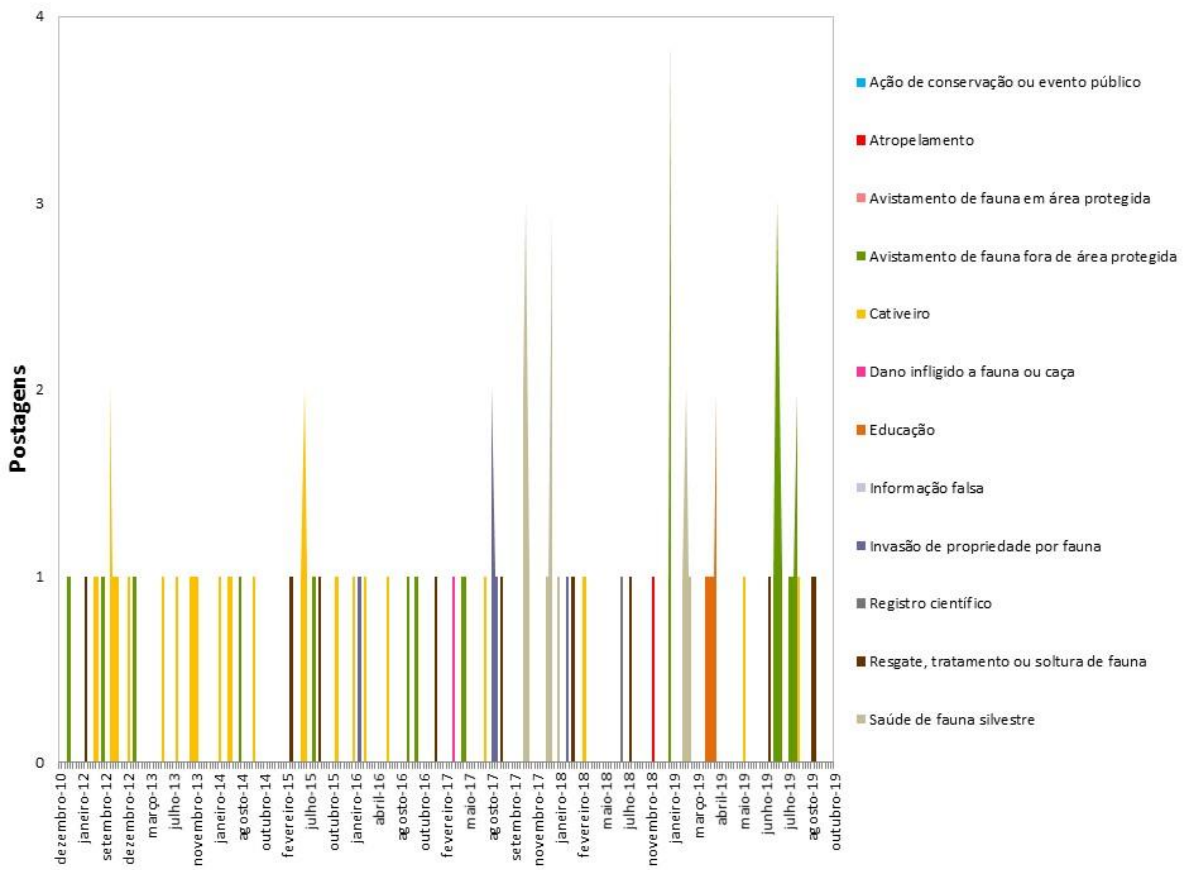
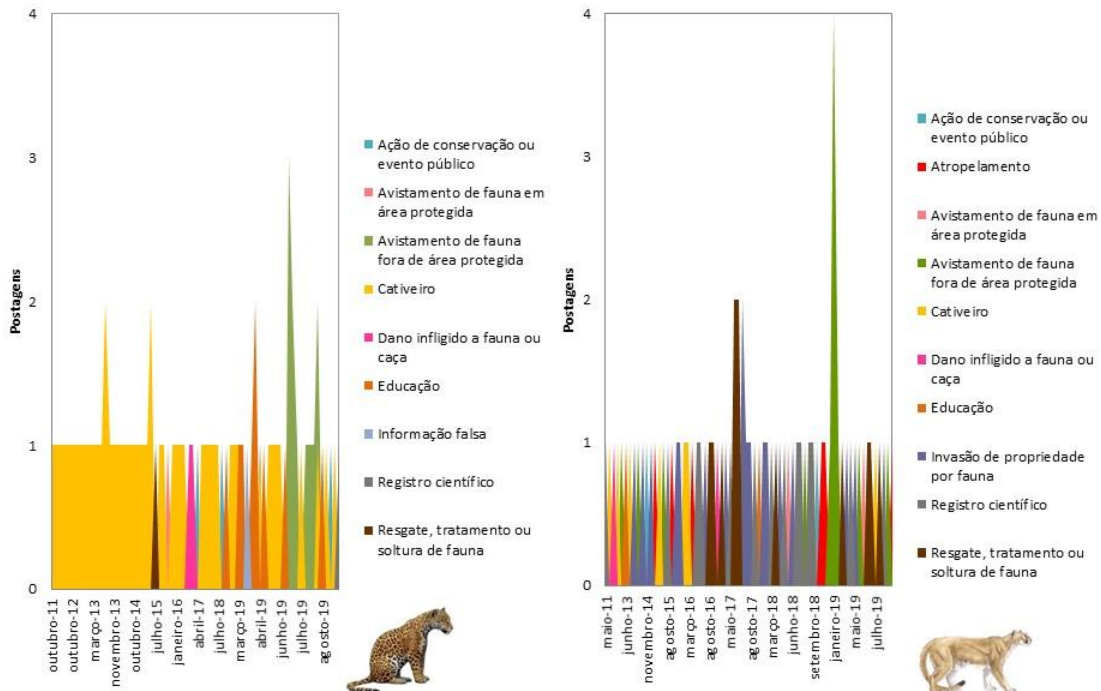


Figura 7. Linha temporal de postagens por categorias de postagens.



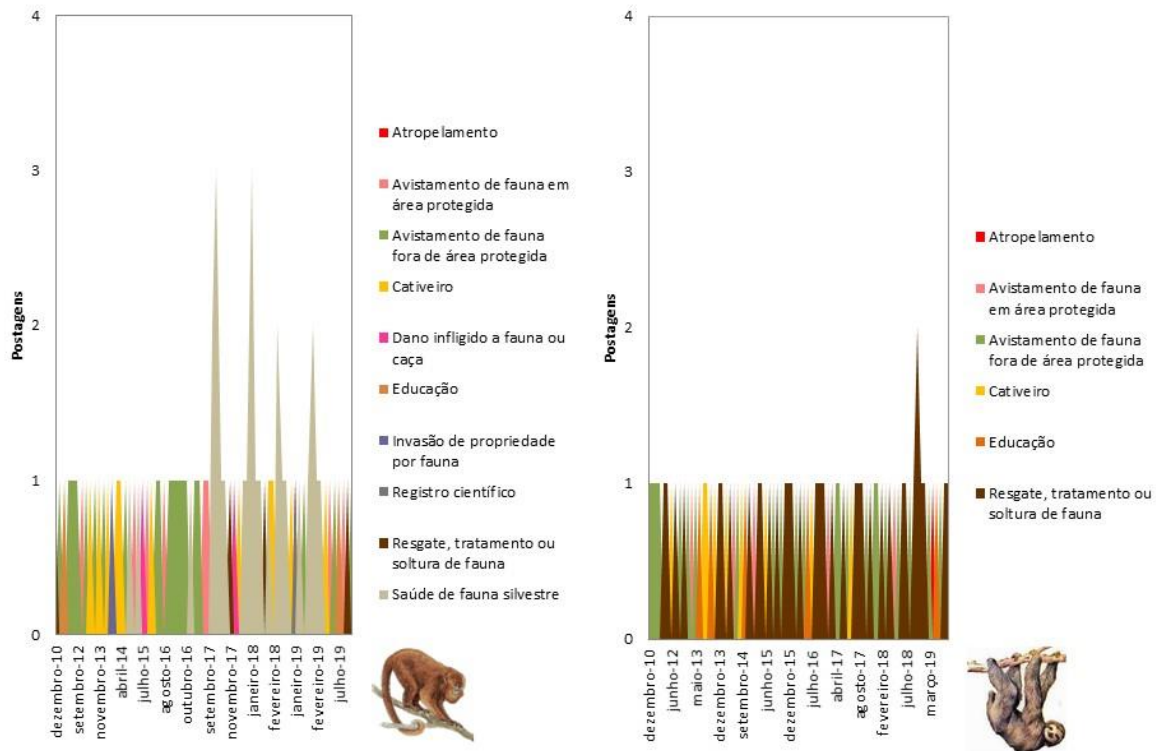


Figura 8. Linha temporal de postagens por categorias de postagens para espécies do estudo.

Facebook - Onde?

Os dados da variável localização por município revelaram diferentes concentrações de postagem para o estado de São Paulo (Figura 9), com o município de São Paulo como destaque para todas as espécies. Para onça-pintada encontramos postagens de cinco municípios, sendo São Paulo (80%) e Apiaí (14,66%) as duas com maior quantidade. Para onça-parda, encontramos postagens de 40 municípios, sendo São Paulo (14,29%) e São Carlos (5,71%) as duas cidades com maior número de postagens, respectivamente. Para o bugio, encontramos ao todo 18 municípios, com um maior número de postagens para São Paulo (72%). Para o bicho-preguiça, encontramos 22 municípios, com um maior número de postagens para São Paulo (41,03%).

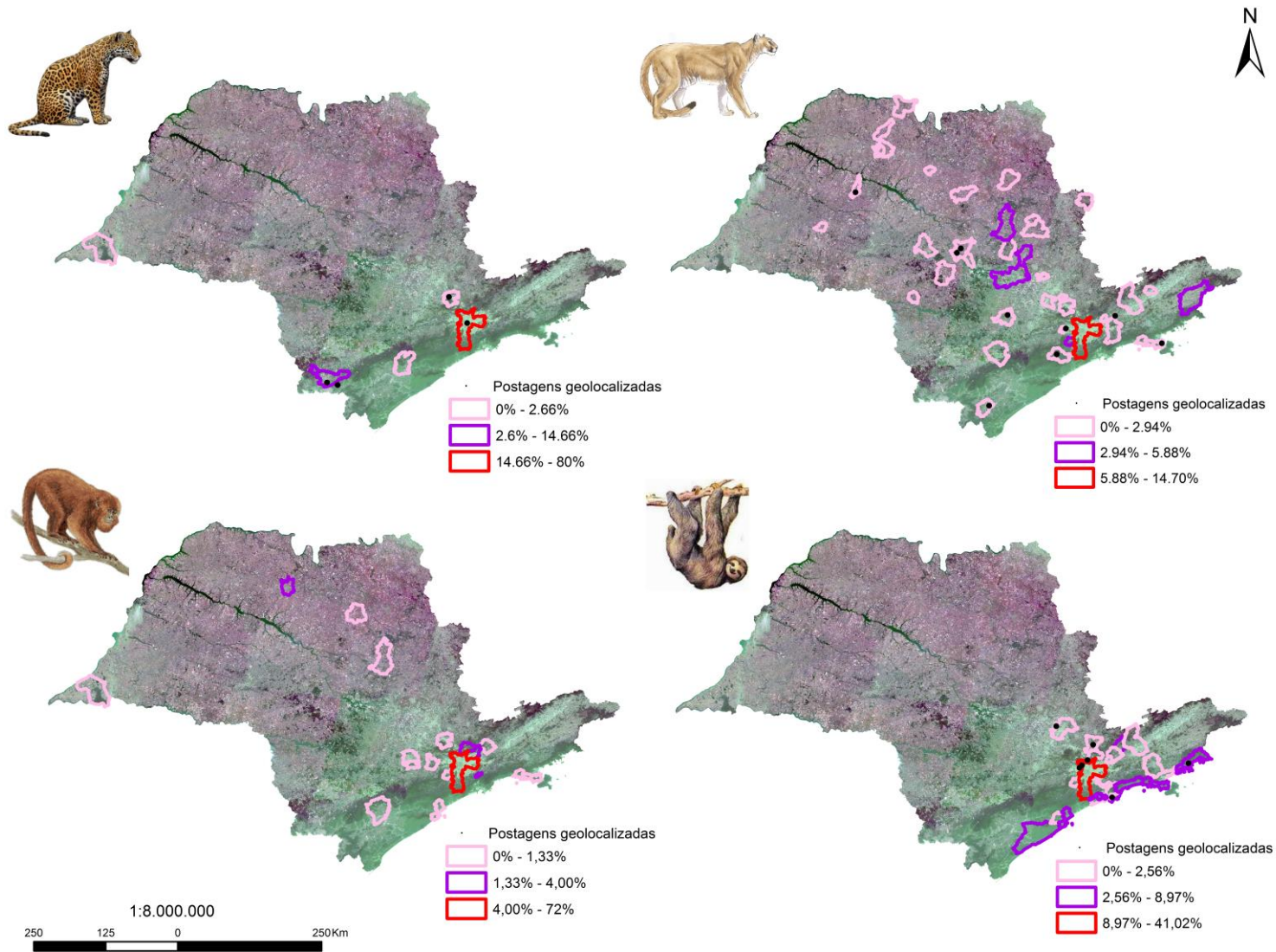


Figura 9. Distribuição de postagens por municípios e espécie. A legenda apresenta os percentuais de postagem por município por espécie.

Facebook - Quem?

Das postagens obtidas no Facebook, encontramos que as três categorias de usuários com maior representação no conjunto de dados foram: página pessoal (42,59%), página de notícias (14,51%) e página de negócios (8,52%). As categorias apresentaram valores similares por espécie (Figura 10), sendo que página pessoal foi a categoria dominante para onça-pintada (75%). Para a onça-parda, página pessoal também foi a categoria dominante (32,91%), seguida por página de notícias (22,78%). Para o bugio, página pessoal apresentou frequência semelhante à da onça-parda (34,62%) seguido por zoológicos e aquários (20,51%). E, para o bicho-preguiça, página pessoal também foi a categoria dominante (34,62%) seguido de página de notícias (18,75%).

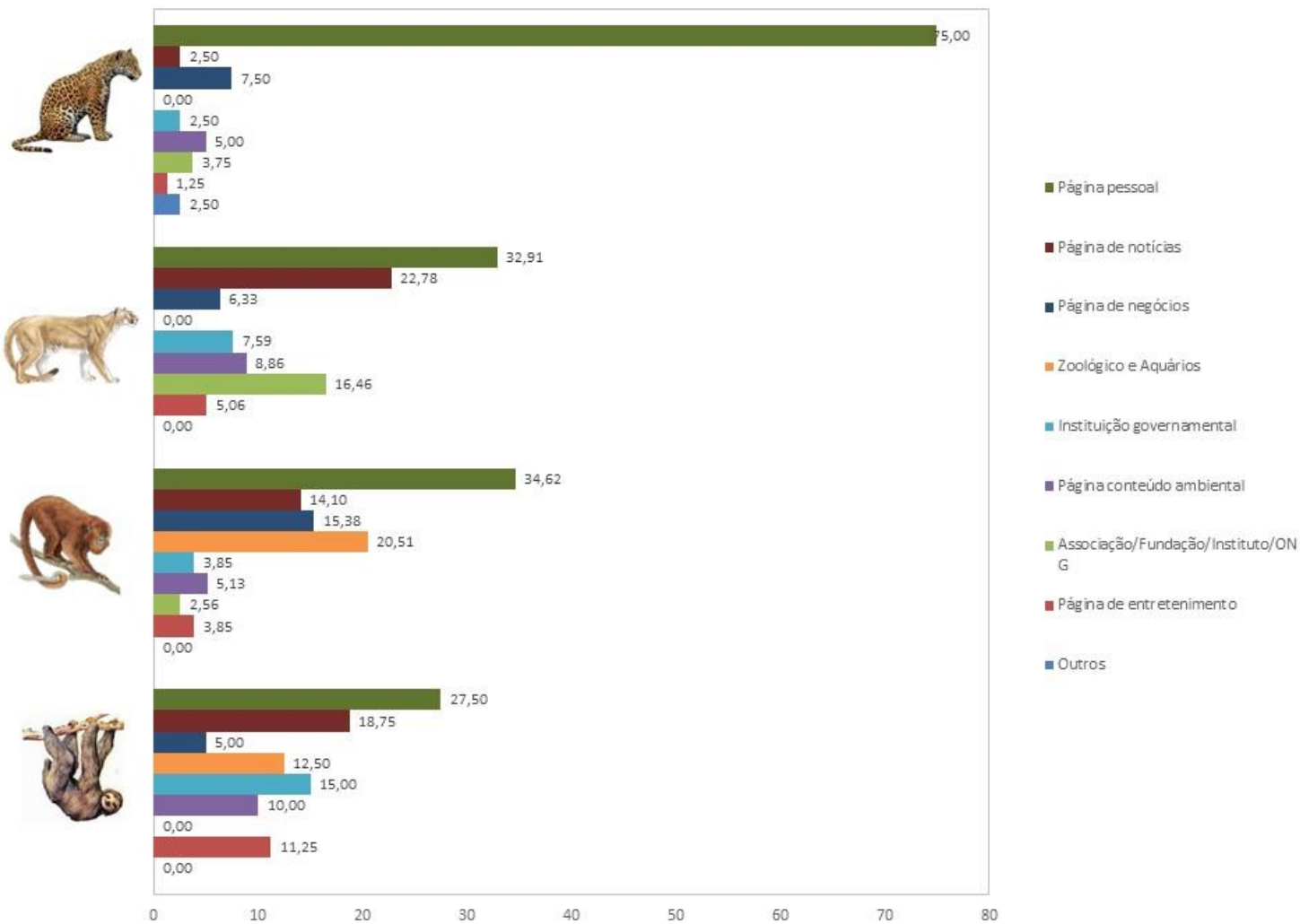


Figura 10. Categorias de usuário por espécie.

Facebook - Engajamento

Das variáveis de engajamento coletadas na base do Facebook, encontramos grande variação, com as variáveis *like*, compartilhamento e *love* apresentando os valores mais expressivos. Para onça-pintada, a variável *like* apresentou mediana de 8,5, sendo que o maior valor observado foi 698. O segundo maior valor de mediana encontrado para onça-pintada foi de compartilhamento (1), sendo 1622 o maior valor encontrado, para comentários, encontramos uma mediana de 0,5, sendo 80 o maior valor encontrado. Para *love*, mesmo com valor 0 de mediana, encontramos o segundo maior valor para variáveis de engajamento utilizando reações do Facebook (288) (Figura 11).

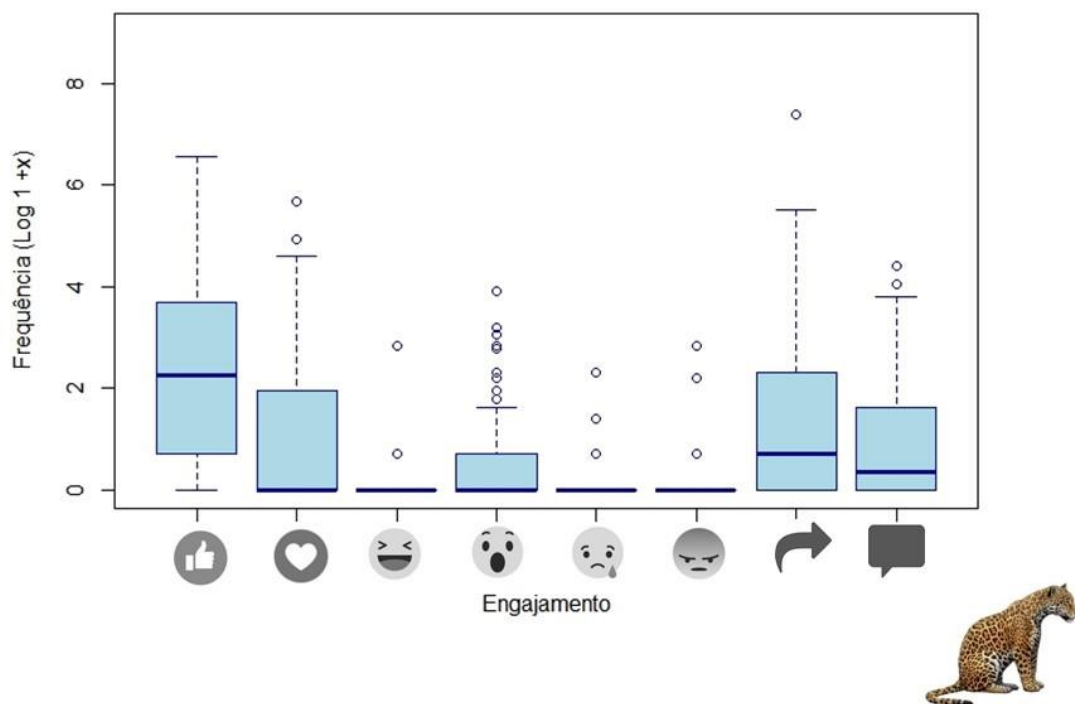


Figura 11. Variáveis de engajamento para onça-pintada. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.

Para onça-parda, as variáveis de engajamento de participação *like*, compartilhamento, comentários e *love* também apresentaram os valores mais expressivos. Para a variável *like*, a mediana encontrada foi de 22, sendo 5.700 o maior valor en-

contrado. Para a variável compartilhamento, a mediana encontrada foi de 4, sendo 1195 o maior valor encontrado. Para a variável comentários, encontramos uma mediana de 2, sendo 259 o maior valor encontrado, e para *love*, mesmo com valor 0 de mediana, encontramos o terceiro maior valor de engajamento utilizando reações do Facebook (308). Para onça-parda, um aumento para a variável *wow* é visto em comparação à onça-pintada (Figura 12).

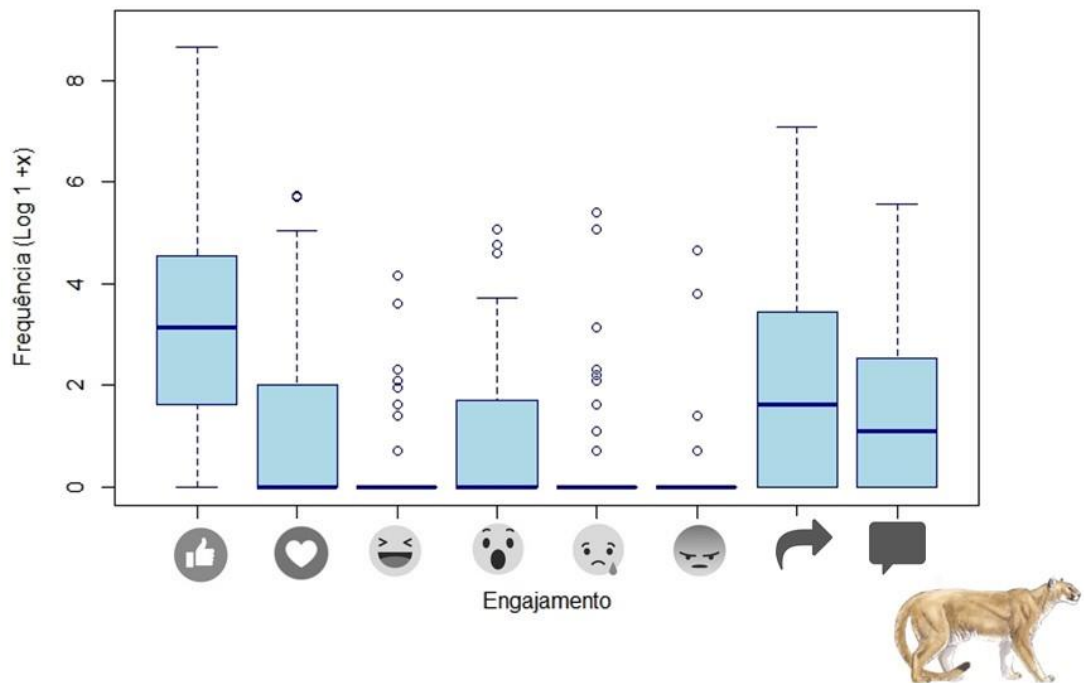


Figura 12. Variáveis de engajamento para a onça-parda. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.

Para o bugio, as variáveis de engajamento de participação *like*, compartilhamento, comentários e *sad* apresentaram os valores mais expressivos. Para a variável *like*, a mediana encontrada foi de 11,5, sendo 611 o maior valor encontrado. Para a variável compartilhamento, a mediana encontrada foi de 2, sendo 1.400 o maior valor encontrado. Para a variável comentários, encontramos uma mediana de 1,5, sendo 743 o maior valor encontrado, para a variável *sad*, mesmo com valor 0 de

mediana, encontramos o terceiro maior valor de engajamento utilizando reações do Facebook (505) (Figura 13).

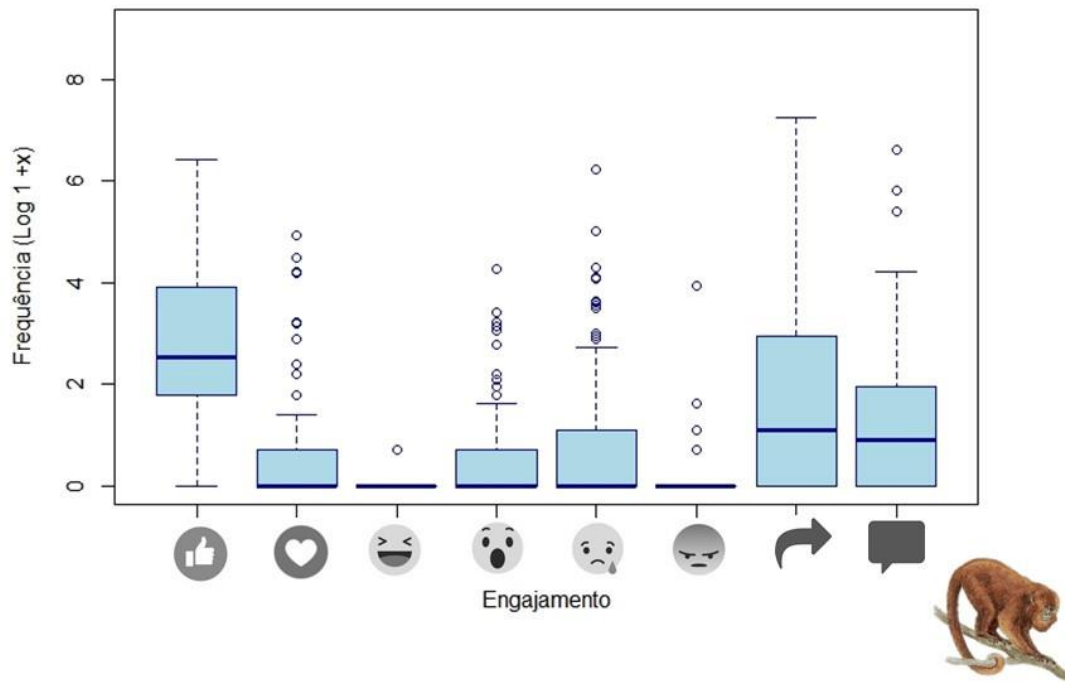


Figura 13. Variáveis de engajamento para o bugio. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.

Para o bicho-preguiça, as variáveis de engajamento de participação *like*, compartilhamento, comentários e *love* apresentaram os valores mais expressivos. Para a variável *like*, a mediana encontrada foi de 13, sendo 5000 o maior valor encontrado. Para a variável compartilhamento, a mediana encontrada foi de 2,5, sendo 473 o maior valor encontrado. Para a variável comentários, encontramos uma mediana de 1 e maior valor de 328. Para a variável *love*, mesmo com valor 0 de mediana, encontramos o terceiro maior valor de engajamento utilizando reações do Facebook (760) (Figura 14).

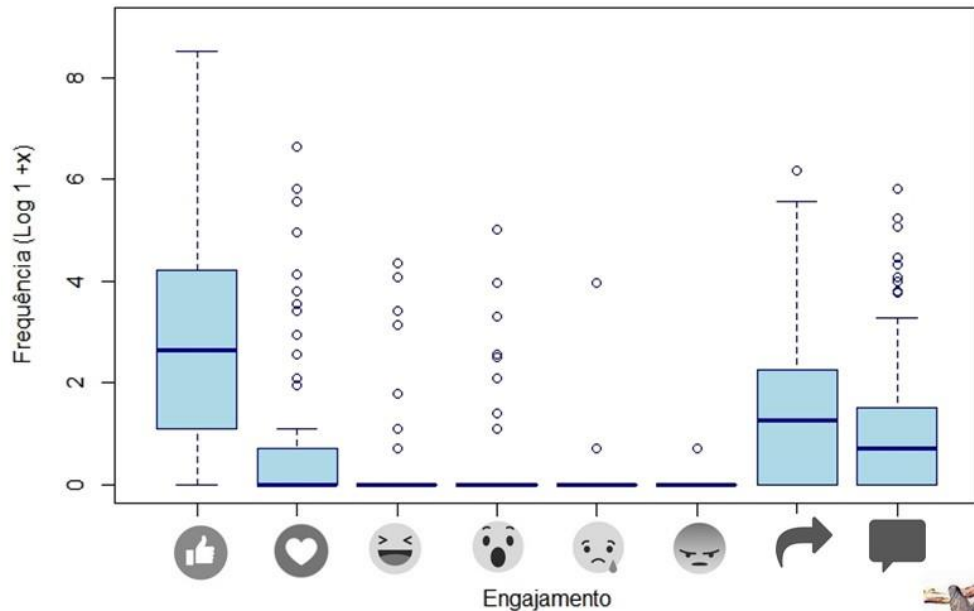


Figura 14. Variáveis de engajamento para o bicho-preguiça. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, os símbolos significam: like, love, funny, wow, sad, angry, compartilhamento e comentários.

Os modelos de regressão apresentaram resultados com grande diferença por espécie (Tabela 5). Para a onça-pintada, no modelo 1, mídias de texto e vídeo apresentaram associação positiva com engajamento e valores de p significativos ($p < 0,001$ e $p < 0,01$). Além disso, para o mesmo modelo, postagens de usuários classificadas páginas de entretenimento também apresentaram associação positiva com engajamento ($p < 0,001$), a categoria outros apresentou associação negativa com engajamento ($p < 0,03$). Para o modelo 2, apenas página de entretenimento apresentou associação positiva com engajamento ($p < 0,001$). As categorias foto e outros apresentaram associação negativa com engajamento ($p < 0,001$, $p < 0,02$). Ambos os modelos apresentaram um valor de R^2 e R^2 ajustados significativo (0,739, 0,622), indicando a validade do modelo.

Para a onça-parda, o modelo 1, a categoria “avistamento de fauna em área protegida” apresentou associação positiva com engajamento ($p < 0,001$). As categori-

as de usuário “instituição governamental” e “página pessoal” apresentaram associação negativa com engajamento (p 0,027, p 0,045). No modelo 2, “avistamento de fauna em área protegida” se manteve com associação positiva com engajamento ($p < 0,001$). A categoria “associação/fundação/instituto/ONG” se revelou como outra categoria com associação positiva com engajamento (p 0,03). Ambos os modelos apresentaram R^2 e R^2 ajustado significativos (0,538, 0,402).

Para o bugio, os modelos não revelaram associações positivas ou negativas com engajamento, o valor de R^2 e R^2 ajustado para ambos os modelos foi de 0,193, não significativo. Para bicho-preguiça, o modelo 1 apresentou todas as categorias de postagem, com exceção de atropelamento, como tendo associação negativa com engajamento (Tabela 5). Todas as categorias de tipo de usuário, com exceção de instituição governamental, também apresentaram associação negativa com engajamento. No modelo 2, a categoria de postagem “atropelamento” apresenta associação positiva com engajamento ($p < 0,001$). Em tipo de usuário, a categoria “instituição governamental” também apresenta associação positiva com engajamento ($p < 0,00$). Ambos os modelos apresentaram R^2 e R^2 ajustado significativos (0,455, - 0,338).

Tabela 5. Resultados de regressão linear múltipla utilizando engajamento como variável resposta. Onde slope=coeficientes da regressão, IC=intervalo de confiança, p=significância, GL=graus de liberdade, R^2 =coeficiente de determinação e (*) quando uma categoria é utilizada como parâmetro para regressão.

Onça-pintada			
Preditor	Slope	IC	p
Modelo 1			
Valor da constante	14,53	-236,65 – 265,71	0,91
Categorias de postagem			
Ação de conservação ou evento público	(*)		
Avistamento de fauna em área protegida	-101,96	-602,02 – 398,11	0,69
Avistamento de fauna fora de área protegida	4,41	-274,64 – 283,45	0,98
Cativeiro	157,06	-89,71 – 403,84	0,21
Dano infligido a fauna ou caça	6,49	-373,38 – 386,36	0,97
Educação	169,1	-134,43 – 472,62	0,27
Informação falsa	214,96	-380,23 – 810,16	0,47
Registro científico	215,02	-296,31 – 726,35	0,40
Resgate, tratamento ou soltura de fauna	107,62	-324,22 – 539,47	0,62
Tipo de mídia			
Foto	(*)		
Texto	283,42	144,33 – 422,52	<0,001
Vídeo	159,84	45,50 – 274,18	0,01
Contagem de palavras	0,33	-0,41 – 1,06	0,38
Tipo de usuário			

Associação/Fundação/Instituto/ONG	(*)		
Instituição governamental	1,02	-316,78 – 318,82	1,00
Outros	-481,67	-910,94 – -52,41	0,03
Página conteúdo ambiental	-337,45	-770,51 – 95,61	0,12
Página de entretenimento	1670,23	1158,51 – 2181,95	<0,001
Página de negócios	-119,48	-470,96 – 232,01	0,50
Página de notícias	-185,48	-617,96 – 247,01	0,394
Página pessoal	-166,44	-494,34 – 161,46	0,314
R^2/R^2 ajustado = 0,739/0,622	GL= 61		
IC = 95%			
Modelo 2			
Valor da constante	197,01	-388,15 – 782,18	0,50
Categorias de postagem			
Ação de conservação ou evento público	101,96	-398,11 – 602,02	0,69
Avistamento de fauna em área protegida	(*)		
Avistamento de fauna fora de área protegida	106,36	-277,04 – 489,77	0,58
Cativeiro	259,02	-136,49 – 654,53	0,20
Dano infligido a fauna ou caça	108,45	-349,77 – 566,67	0,64
Educação	271,05	-141,37 – 683,48	0,19
Informação falsa	316,92	-265,57 – 899,41	0,28
Registro científico	316,98	-247,26 – 881,21	0,27
Resgate, tratamento ou soltura de fauna	209,58	-332,12 – 751,28	0,44
Tipo de mídia			
Foto	-283,42	-422,52 – -144,33	<0,001
Texto	(*)		
Vídeo	-123,58	-264,08 – 16,91	0,08
Contagem de palavras	0,33	-0,41 – 1,06	0,38
Tipo de usuário			
Associação/Fundação/Instituto/ONG	-1,02	-318,82 – 316,78	1,00
Instituição governamental	(*)		
Outros	-482,69	-897,67 – -67,71	0,02
Página conteúdo ambiental	-338,47	-755,40 – 78,47	0,11
Página de entretenimento	1669,21	1177,09 – 2161,34	<0,001
Página de negócios	-120,5	-450,18 – 209,19	0,47
Página de notícias	-186,49	-590,26 – 217,27	0,359
Página pessoal	-167,46	-470,83 – 135,91	0,274
R^2/R^2 ajustado = 0,739/0,622	GL= 61		
IC = 95%			
Onça-parda			
Preditor	<i>Slope</i>	<i>IC</i>	<i>p</i>
Modelo 1			
Valor da constante	192,21	-1264,62 – 1649,05	0,793
Categorias de postagem			
Ação de conservação ou evento público	(*)		
Atropelamento	305,55	-1112,38 – 1723,48	0,668
Avistamento de fauna em área protegida	4482,64	2575,32 – 6389,96	<0,001
Avistamento de fauna fora de área protegida	396,77	-1147,74 – 1941,29	0,609
Cativeiro	519,34	-1025,01 – 2063,69	0,504
Dano infligido a fauna ou caça	395,83	-1296,29 – 2087,96	0,642
Educação	455,1	-1207,14 – 2117,33	0,586
Invasão de propriedade por fauna	306,88	-1133,16 – 1746,92	0,672
Registro científico	621,76	-874,98 – 2118,49	0,409
Resgate, tratamento ou soltura de fauna	355,2	-1131,24 – 1841,63	0,634
Tipo de mídia			

Foto	(*)		
Texto	196,99	-487,23 – 881,21	0,567
Vídeo	406,54	-166,46 – 979,54	0,161
Contagem de palavras	0,99	-3,79 – 5,76	0,681
Tipo de usuário			
Associação/Fundação/Instituto/ONG	(*)		
Instituição governamental	-1003,24	-1890,49 – -115,98	0,027
Página conteúdo ambiental	-816,52	-1698,55 – 65,52	0,069
Página de entretenimento	-893,94	-1947,60 – 159,72	0,095
Página de negócios	-783,12	-1729,26 – 163,03	0,103
Página de notícias	-605,76	-1276,44 – 64,91	0,076
Página pessoal	-721,43	-1425,75 – -17,11	0,045
R^2/R^2 ajustado = 0,538/0,402	GL= 61		
IC = 95%			
Modelo 2			
Valor da constante	-308,48	-1477,82 – 860,86	0,60
Categorias de postagem			
Ação de conservação ou evento público	-305,55	-1723,48 – 1112,38	0,67
Atropelamento	(*)		
Avistamento de fauna em área protegida	4177,09	2564,84 – 5789,34	<0,001
Avistamento de fauna fora de área protegida	91,22	-1001,99 – 1184,44	0,87
Cativeiro	213,79	-898,90 – 1326,48	0,70
Dano infligido a fauna ou caça	90,28	-1186,18 – 1366,74	0,89
Educação	149,54	-1114,10 – 1413,19	0,81
Invasão de propriedade por fauna	1,33	-960,88 – 963,54	1,00
Registro científico	316,21	-727,66 – 1360,07	0,55
Resgate, tratamento ou soltura de fauna	49,65	-993,36 – 1092,65	0,92
Tipo de mídia			
Foto	-196,99	-881,21 – 487,23	0,57
Texto	(*)		
Vídeo	209,55	-639,47 – 1058,56	0,62
Contagem de palavras	0,99	-3,79 – 5,76	0,68
Tipo de usuário			
Associação/Fundação/Instituto/ONG	1003,24	115,98 – 1890,49	0,03
Instituição governamental	(*)		
Página conteúdo ambiental	186,72	-860,36 – 1233,81	0,72
Página de entretenimento	109,3	-1039,54 – 1258,14	0,85
Página de negócios	220,12	-872,57 – 1312,82	0,688
Página de notícias	397,47	-485,07 – 1280,02	0,371
Página pessoal	281,81	-610,53 – 1174,15	0,53
R^2/R^2 ajustado = 0,538/0,402	GL= 61		
IC = 95%			
Bugio			
Modelo 1			
Valor da constante	365,7	-853,95 – 1585,36	0,55
Categorias de postagem			
Atropelamento	(*)		
Avistamento de fauna em área protegida	-538,2	-1612,83 – 536,43	0,32
Avistamento de fauna fora de área protegida	-596,66	-1644,16 – 450,83	0,26
Cativeiro	-477,08	-1453,65 – 499,50	0,33
Dano infligido a fauna ou caça	-430,87	-1666,64 – 804,91	0,49
Educação	-485,9	-1720,71 – 748,91	0,43
Invasão de propriedade por fauna	-469,29	-2003,06 – 1064,47	0,54
Registro científico	-557,7	-1872,52 – 757,13	0,40

Resgate, tratamento ou soltura de fauna	-256,61	-1399,28 – 886,06	0,66
Saúde de fauna silvestre	-316,17	-1322,74 – 690,40	0,53
<hr/>			
Tipo de mídia			
Foto	(*)		
Texto	-103,22	-533,97 – 327,53	0,63
Vídeo	-32,76	-438,27 – 372,74	0,87
<hr/>			
Contagem de palavras			
	0,69	-0,45 – 1,83	0,23
<hr/>			
Tipo de usuário			
Associação/Fundação/Instituto/ONG	(*)		
Instituição governamental	4,04	-929,88 – 937,96	0,99
Página conteúdo ambiental	80,62	-819,71 – 980,95	0,86
Página de entretenimento	196,87	-739,90 – 1133,64	0,68
Página de negócios	400,62	-394,31 – 1195,56	0,32
Página de notícias	125,34	-711,52 – 962,21	0,765
Página pessoal	151,28	-620,79 – 923,35	0,696
Zoológico e aquários	191,99	-614,86 – 998,85	0,636
<hr/>			
R^2/R^2 ajustado = 0,193/-0,071			
GL= 59			
IC = 95%			
<hr/>			
Modelo 2			
Valor da constante	-271,67	-1244,98 – 701,64	0,58
<hr/>			
Categorias de postagem			
Atropelamento	538,2	-536,43 – 1612,83	0,32
Avistamento de fauna em área protegida	(*)		
Avistamento de fauna fora de área protegida	-58,47	-468,34 – 351,41	0,78
Cativeiro	61,12	-508,21 – 630,44	0,83
Dano infligido a fauna ou caça	107,33	-710,97 – 925,63	0,79
Educação	52,3	-708,35 – 812,94	0,89
Invasão de propriedade por fauna	68,9	-1176,37 – 1314,18	0,91
Registro científico	-19,5	-1096,78 – 1057,79	0,97
Resgate, tratamento ou soltura de fauna	281,59	-434,74 – 997,91	0,44
Saúde de fauna silvestre	222,03	-227,26 – 671,32	0,33
<hr/>			
Tipo de mídia			
Foto	103,22	-327,53 – 533,97	0,63
Texto	(*)		
Vídeo	70,46	-480,66 – 621,57	0,80
<hr/>			
Contagem de palavras			
	0,69	-0,45 – 1,83	0,23
<hr/>			
Tipo de usuário			
Associação/Fundação/Instituto/ONG	-4,04	-937,96 – 929,88	0,99
Instituição governamental	(*)		
Página conteúdo ambiental	76,58	-820,08 – 973,24	0,87
Página de entretenimento	192,83	-729,66 – 1115,31	0,68
Página de negócios	396,58	-397,11 – 1190,27	0,32
Página de notícias	121,3	-684,22 – 926,82	0,764
Página pessoal	147,24	-618,79 – 913,27	0,702
Zoológico e aquários	187,95	-605,53 – 981,43	0,637
<hr/>			
R^2/R^2 ajustado = 0,193/-0,071			
GL= 59			
IC = 95%			
<hr/>			
Bicho-preguiça			
Modelo 1			
Valor da constante	4846,28	3219,53 – 6473,02	<0,001
<hr/>			
Categorias de postagem			
Atropelamento	(*)		
Avistamento de fauna em área protegida	-3773,43	-5648,88 – -1897,99	<0,001
Avistamento de fauna fora de área protegida	-3791,22	-5533,27 – -2049,17	<0,001

Cativeiro	-4021,97	-5822,71 – -2221,22	<0,001
Educação	-4206,59	-6117,21 – -2295,97	<0,001
Resgate, tratamento ou soltura de fauna	-3406,06	-5109,09 – -1703,02	<0,001
Tipo de mídia			
Foto	(*)		
Texto	161,63	-549,37 – 872,63	0,65
Vídeo	-124,29	-714,16 – 465,58	0,68
Contagem de palavras	-1,37	-4,59 – 1,86	0,40
Tipo de usuário			
Instituição governamental	(*)		
Página conteúdo ambiental	-1248,01	-2042,70 – -453,33	0,00
Página de entretenimento	-1220,84	-1975,82 – -465,85	0,00
Página de negócios	-1063,42	-2061,77 – -65,06	0,04
Página de notícias	-893,92	-1619,63 – -168,21	0,02
Página pessoal	-940,65	-1602,63 – -278,68	0,01
Zoológico e aquários	-691,17	-1527,20 – 144,85	0,10
R^2/R^2 ajustado = 0,455/-0,338	GL= 65		
IC = 95%			
Modelo 2			
Valor da constante	-13,54	-1090,14 – 1063,06	0,98
Categorias de usuário			
Atropelamento	3773,43	1897,99 – 5648,88	<0,001
Avistamento de fauna em área protegida	(*)		
Avistamento de fauna fora de área protegida	-17,79	-835,11 – 799,53	0,97
Cativeiro	-248,53	-1198,07 – 701,00	0,60
Educação	-433,15	-1583,35 – 717,04	0,46
Resgate, tratamento ou soltura de fauna	367,38	-500,16 – 1234,91	0,40
Tipo de mídia			
Foto	-161,63	-872,63 – 549,37	0,65
Texto	(*)		
Vídeo	-285,92	-1117,57 – 545,73	0,50
Contagem de palavras	-1,37	-4,59 – 1,86	0,40
Tipo de usuário			
Instituição governamental	1248,01	453,33 – 2042,70	0,00
Página conteúdo ambiental	(*)		
Página de entretenimento	27,18	-774,79 – 829,14	0,95
Página de negócios	184,60	-899,32 – 1268,51	0,74
Página de notícias	354,10	-394,77 – 1102,97	0,35
Página pessoal	307,36	-477,00 – 1091,72	0,44
Zoológico e aquários	556,84	-402,05 – 1515,73	0,25
R^2/R^2 ajustado = 0,455/-0,338	GL= 65		
IC = 95%			

Twitter - O que?

Na base de dados do twitter, postagens incluindo apenas texto corresponderam a quase totalidade dos dados. Analisando a base extraída, encontramos apenas 5,17% dos dados com informação sobre presença de foto para onça-pintada e 2,75% dos dados para onça-parda (Figura 15).

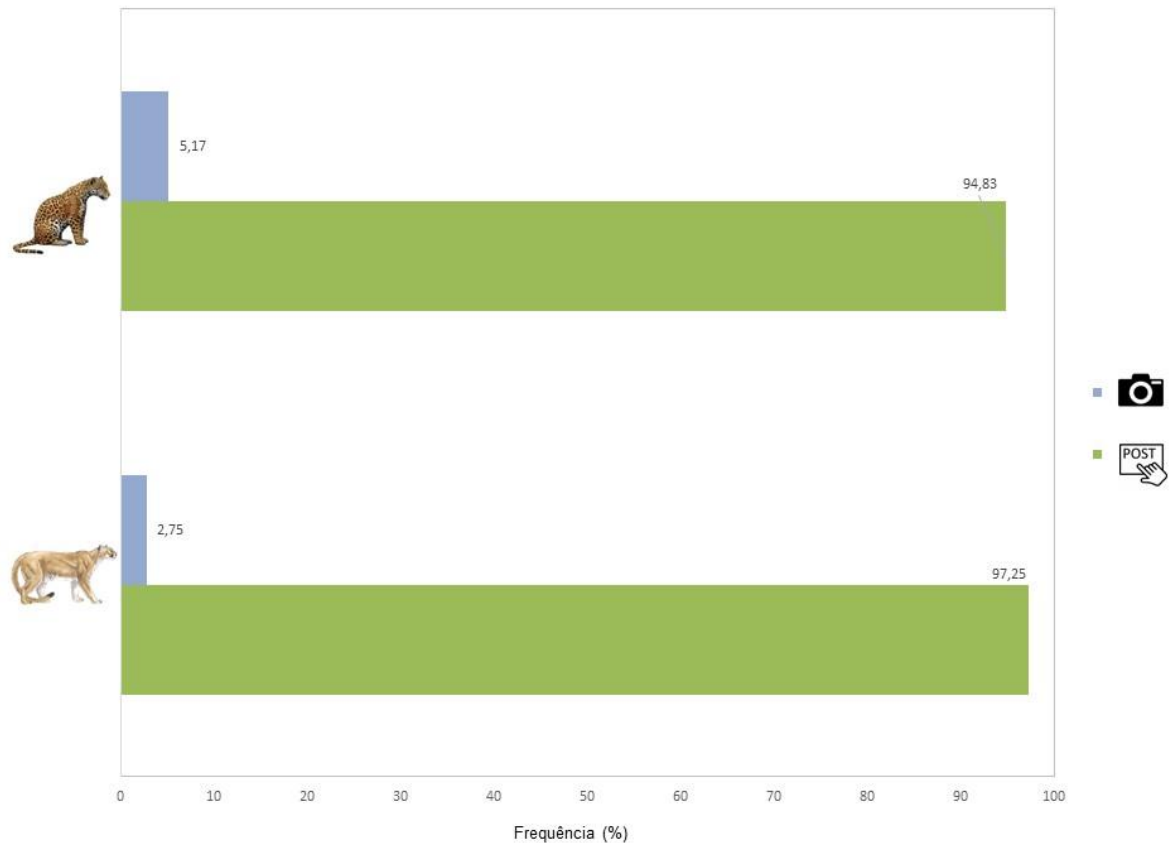


Figura 15. Proporção de postagens por tipos de mídia para cada uma das espécies na amostra do Twitter. Os ícones de mídia descrevem, respectivamente, foto e texto.

Em relação à variável contagem de palavras no Twitter, a distribuição foi mais próxima de uma distribuição normal (Figura 16). A mediana geral encontrada foi de 15 para as espécies, sendo uma mediana de 15 para onça-pintada e 16 para onça-parda. Não encontramos altos valores de *outliers*, com a distribuição geral dos dados se concentrando em torno de 5 e 25 palavras para onça-pintada e 10 e 25 palavras para onça-parda. Os valores totais de palavras foram 263.473 para onça-pintada e 260.748 para onça-parda.

Os *rankings* de texto (Tabela 6) encontrados na base de dados do Twitter destacaram palavras em comum para as espécies como filhote, parque e casa. Para espécies, encontramos palavras como pantanal, extinção e tiro para onça-pintada e rodovia, capturada e morre para onça-parda. Também nos *rankings* foi possível encontrar cidades e também biomas com destaque entre as palavras.

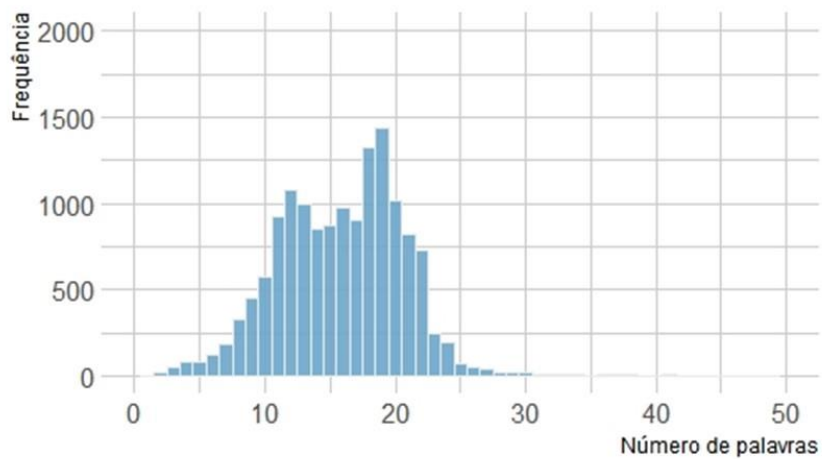
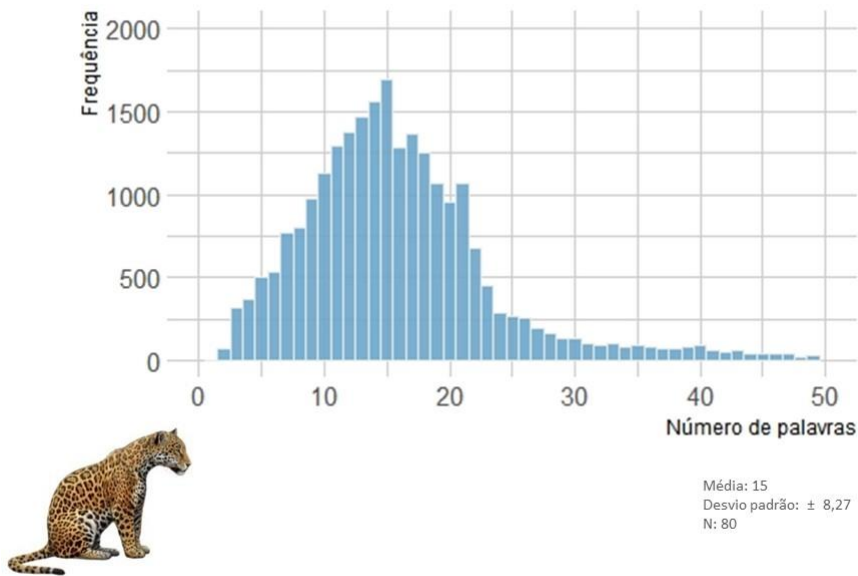


Figura 16. Histogramas com representação do número de palavras por espécies, incluindo mediana, desvio padrão e número amostral.

Tabela 6. Ranking de palavras por espécie, incluindo proporção relativa ao número total de palavras encontrada por espécie.

<i>Onça-pintada</i>			<i>Onça-parda</i>		
Palavras	Contagem	Proporção (%)	Palavras	Contagem	Proporção (%)
Youtube	616	0,25	Atropelada	2442	0,97
Vídeo	608	0,24	Rodovia	2159	0,86
Pantanal	524	0,21	Filhote	1858	0,74
Extinção	502	0,20	Capturada	1219	0,49
Tiro	475	0,19	Interior	1110	0,44
Casa	414	0,17	Morre	960	0,38
Parque	369	0,15	Casa	786	0,31
Gente	367	0,15	Morta	778	0,31
Mundo	365	0,15	Resgatada	708	0,28
Filhotes	338	0,14	Fazenda	673	0,27
Rodovia	312	0,12	Parque	633	0,25
Mata	308	0,12	Árvore	592	0,24
Iguaçu	298	0,12	Paraná	578	0,23

Twitter – Quando?

Na linha temporal obtida para o Twitter (Figura 17), encontramos um pico de postagens começando no final de maio de 2016, com ápice entre junho e julho do mesmo ano. O termo “Juma” consta na base de palavras das postagens do Twitter, embora não tenha composto o *ranking* com as quinze palavras mais citadas. De todo modo, o número de postagens relacionado à espécie apresenta pouca variação entre outubro de 2016 e junho de 2019. Para onça-parda, a linha temporal possui dados dos últimos dez anos com três picos proeminentes de postagens entre setembro e dezembro de 2011, agosto e setembro de 2015, e julho e agosto de 2016. A ausência de categorização na base permite apenas inferir as razões para picos pela presença de palavras como cidade e atropelamento, o que seriam indicativos de postagens relacionadas à avistamento da onça-parda próxima a cidades.

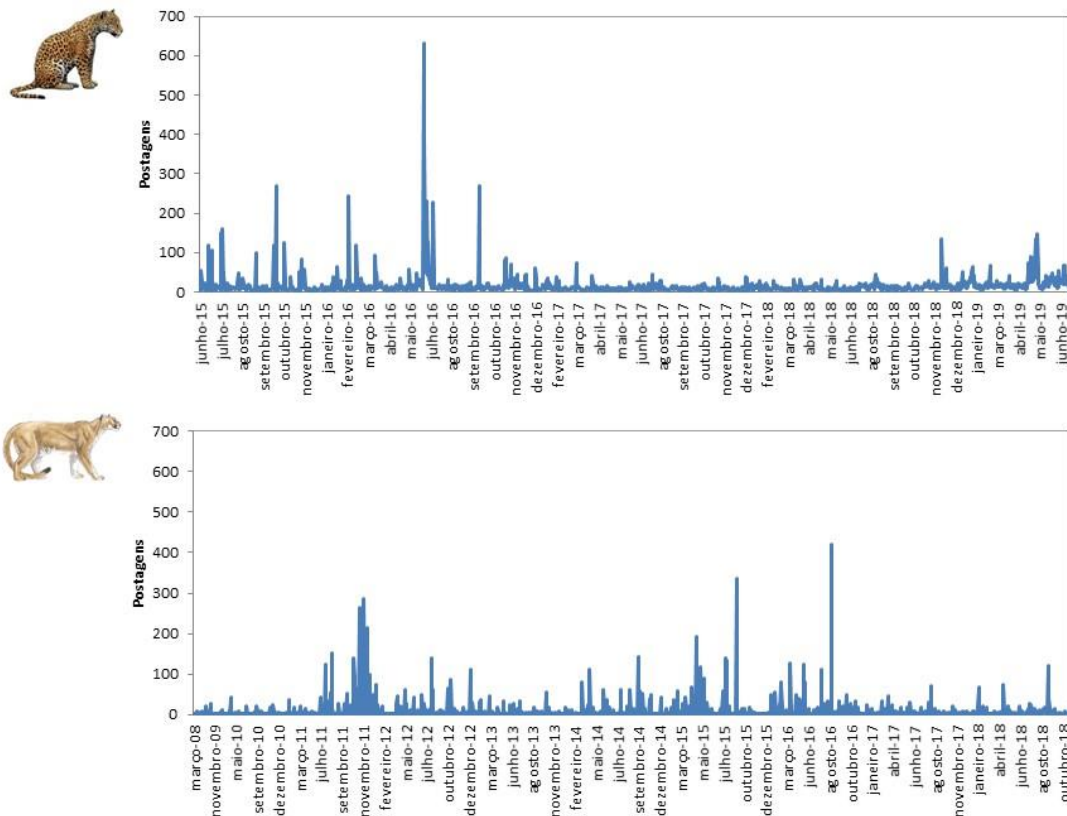


Figura 17. Linha temporal de postagens do Twitter para onça-pintada e onça-parda.

Twitter – Quem?

A distribuição dos dados obtidos para as variáveis que descrevem o perfil de usuários para o twitter revelou uma grande variação de dados para usuários (Figura 18). Para onça-pintada o valor de mediana encontrado para seguidores foi de 451 (com valor máximo de 1.044.712) e para amigos uma mediana de 340 (valor máximo de 999.676). Para onça-parda, encontramos o valor de mediana para seguidores de 627 (valor máximo 10.447.139) e para amigos, o valor de 329 (valor máximo 999.934).

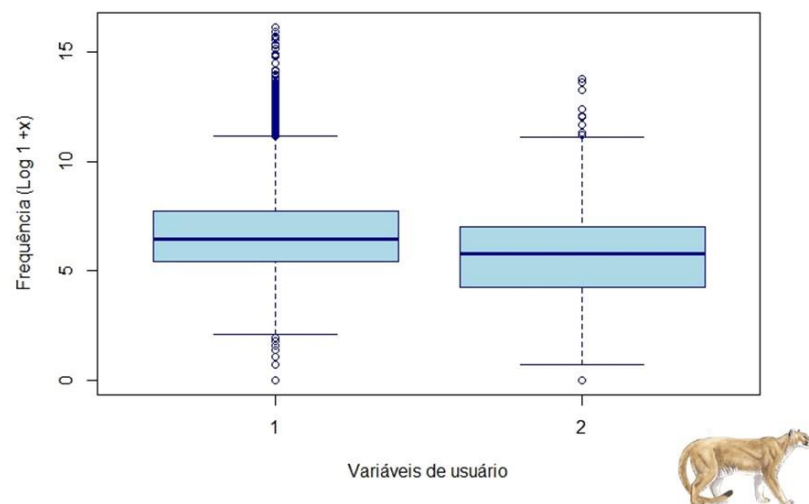
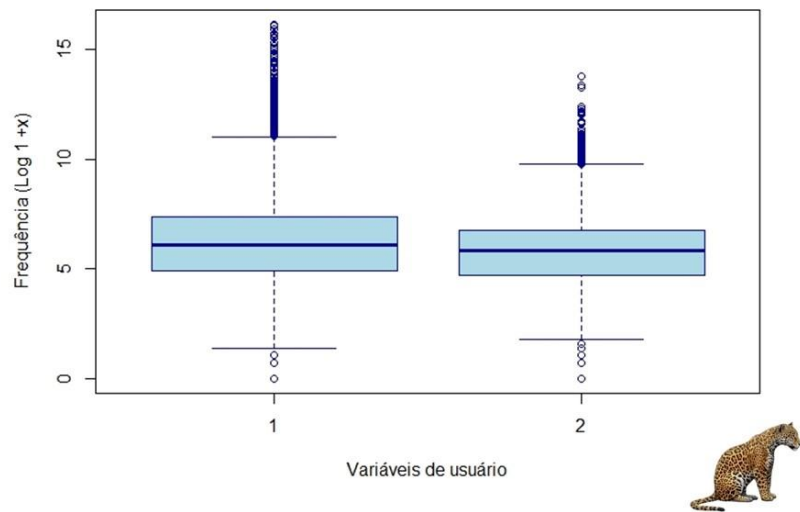


Figura 18. Variáveis descritivas de usuários. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, onde 1 – seguidores e 2 – amigos.

Twitter – *Engajamento*

As variáveis de engajamento de consumo para o Twitter apresentaram todas as medianas com valor 0, com grande variação nos dados (Figura 19). Os valores máximos de compartilhamentos para onça-pintada foi de 3.359 e de favoritado/ *like* foi de 8.097. Para onça-parda, o valor máximo de compartilhamento foi de 500 e favoritado/ *like* de 2.865.

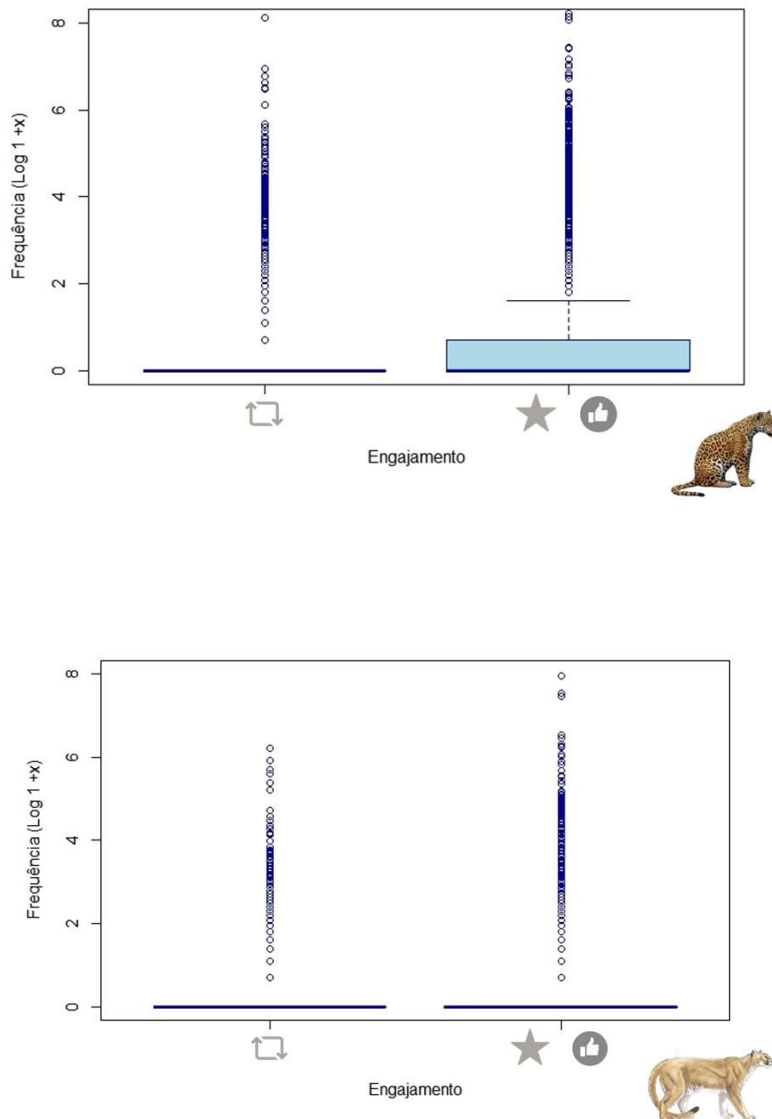


Figura 19. Variáveis de engajamento para o Twitter. A representação das variáveis está em escala logarítmica para diminuir o efeito de outliers e permitir a visualização de quartis e medianas. Em ordem da esquerda para direita, os símbolos representam: compartilhamento e favoritado/like.

Para os modelos obtidos na análise de regressão utilizando dados do Twitter, observamos que a maior parte das categorias e variáveis possui associação positiva com engajamento (Tabela 7). Para onça-pintada, o modelo 1 destaca contagem de palavras e seguidores como tendo associação positiva com engajamento ($p < 0,001$, $p < 0,001$) e texto com associação negativa com engajamento ($p < 0,001$). O modelo 2 destaca foto ($p < 0,001$), seguidores ($p < 0,001$) e amigos ($p < 0,044$) com associação positiva com engajamento. Ambos os modelos apresentaram R^2 e R^2 ajustado signifi-

ficativos (0,135, 0,135). Para onça-parda, o modelo 1 destaca foto ($p<0,001$) e seguidores ($p<0,001$) como tendo associação positiva com engajamento e amigos ($p=0,044$) como tendo associação negativa com engajamento. O modelo 2 destaca seguidores ($p<0,001$) e amigos ($p<0,001$) como tendo associação positiva com engajamento e texto ($p<0,001$) como tendo associação negativa com engajamento. Ambos os modelos apresentaram R^2 e R^2 ajustado significativos (0,135, 0,135).

Tabela 7. Resultados de regressão linear múltipla utilizando engajamento como variável resposta para o Twitter. Onde slope=coeficientes da regressão, IC=intervalo de confiança, p=significância, GL=graus de liberdade, R^2 =coeficiente de determinação e (*) quando uma categoria é utilizada como parâmetro para regressão.

Onça-pintada			
Preditor	Slope	IC	p
Modelo 1			
Valor da constante	-8,63	-12,31 – -4,95	<0,001
Tipo de mídia			
Texto	(*)		
Foto	10,9	4,53 – 17,26	0,001
Contagem de palavras	0,78	0,58 – 0,98	<0,001
Seguidores	0	0,00 – 0,00	<0,001
Amigos	0	-0,00 – 0,00	0,498
R^2/R^2 ajustado =0,015/0,015		GL= 23848	
IC = 95%			
Modelo 2			
Valor da constante	2,27	-4,33 – 8,87	0,501
Tipo de mídia			
Texto	-10,9	-17,26 – -4,53	0,001
Foto	(*)		
Contagem de palavras	0,78	0,58 – 0,98	<0,001
Seguidores	0	0,00 – 0,00	<0,001
Amigos	0	-0,00 – 0,00	0,50
R^2/R^2 ajustado =0,015/0,015		GL= 23848	
IC = 95%			
Onça-parda			
Preditor	Slope	IC	p
Modelo 1			
Valor da constante	12,76	8,92 – 16,61	<0,001
Tipo de mídia			
Texto	(*)		
Foto	11,35	7,91 – 14,79	<0,001
Contagem de palavras	-0,09	-0,23 – 0,05	0,224
Seguidores	0	0,00 – 0,00	<0,001
Amigos	0	-0,00 – -0,00	0,044
R^2/R^2 ajustado =0,135/0,135		GL= 14664	
IC = 95%			
Modelo 2			
Valor da constante	1,41	-1,01 – 3,84	0,253
Tipo de mídia			

Texto	-11,35	-14,79 – -7,91	<0,001
Foto	(*)		
Contagem de palavras	-0,09	-0,23 – 0,05	0,224
Seguidores	0	0,00 – 0,00	<0,001
Amigos	0	-0,00 – -0,00	0,044
<hr/>			
R^2/R^2 ajustado =0,135/0,135	GL= 14664		
IC = 95%			

Classificação da base do Twitter utilizando aprendizado de máquina

Na classificação feita para a base do twitter obtivemos diferentes resultados de classificação por algoritmo para categorias (Tabela 8). Para a base de dados da onça-pintada, a melhor acurácia na classificação com validação a partir de amostragem aleatória ocorreu com o algoritmo SVM *embedding* LSA Topic (0,71) e a melhor acurácia utilizando validação sem amostragem aleatória ocorreu com o algoritmo SVM com *embedding* LSA (0,62). Para a base de dados da onça-parda, a melhor acurácia na classificação com validação a partir de amostragem aleatória ocorreu com o algoritmo Naive Bayes *embedding* LSA (0,64) e a melhor acurácia utilizando validação sem amostragem aleatória ocorreu com os algoritmos Naive Bayes *embedding* TI-IDF e Random Forest *embedding* LSA (0,58, 0,58).

Tabela 8. Resultados de desempenho para categorização da base do Twitter com a utilização de algoritmos.

Algoritmo	<i>Embedding</i>	Acurácia
Onça-pintada		
Validação com amostra aleatória		
Naive Bayes	LSA	0,66
Regressão logística	TI-IDF	0,68
SVM	LSATopic	0,71
Random Forest	LSA	0,65
Validação sem amostra aleatória		
Naive Bayes	LSA	0,58
Regressão logística	LSA	0,61
SVM	LSA	0,62
Random Forest	LSA	0,56
Onça-parda		
Validação com amostra aleatória		
Naive Bayes	LSA	0,64
Regressão logística	LSA	0,62
SVM	TI-IDF	0,58
Random Forest	LSATopic	0,61
Validação sem amostra aleatória		
Naive Bayes	TI-IDF	0,58
Regressão logística	LSA	0,57
SVM	LSA	0,53
Random Forest	LSA	0,58

5. DISCUSSÃO

A inclusão da dimensão digital traz consigo grandes desafios para estudos e experimentos adequados no trabalho com redes sociais. Com este trabalho, a partir da compreensão de diferentes características dos dados para onça-pintada, onça-parda, bugio e bicho-preguiça, obtivemos resultados que indicam que utilizar conjuntos de dados menores de redes sociais para identificar padrões e características chave pode contribuir para a compreensão dos dados em outras plataformas e em esferas maiores. Espécies apresentam particularidades quando representadas em redes sociais, que, por sua vez, constituem um poderoso meio de comunicação social que pode amplificar a relação com setores e o potencial de ações.

Nosso principal resultado confirma o benefício em utilizar um analista para classificação de dados de redes sociais (POORTHUIS et al., 2016). Embora, para grandes conjuntos de dados, como obtivemos no Twitter, esta função não seja necessariamente aplicável. A classificação automática é, sem dúvida, uma via de mão mais rápida para tratar grandes conjuntos de dados (DI MININ, 2015), mas, como observamos, há mais benefícios quando partimos para uma análise automatizada utilizando o aprendizado obtido com os dados e tratamento em uma escala menor.

Levando em consideração o que outros autores afirmam sobre a construção do universo representacional que a *web 2.0* proporciona, existe uma distinção na maneira como cada espécie é representada. Podemos nos apoiar no conceito da cultorômica da conservação para acessar um panorama dessas diferenças. Por exemplo, se levamos em consideração uma das espécies abordadas neste estudo, a onça-pintada, culturalmente temos uma representação do animal como símbolo de força e ícone da biodiversidade brasileira (CRISTANCHO; VINI, 2004). Já a onça-parda possui destaque diferente do observado para países da América do Norte, mas ainda figura entre um dos ícones da fauna brasileira. Quando observamos os dados obtidos em ambas as redes sociais para as duas espécies, a principal diferença está no tipo de informação disponibilizado e no impacto que ela causa nas pessoas. Enquanto fotos da onça-pintada em cativeiro podem contribuir para a presença carismática da espécie, notícias sobre onças-pardas em áreas urbanas podem aumentar o medo e a desconfiança em relação à espécie.

Em relação às outras duas espécies, a representação para o bugio no ambiente digital foi influenciada por um fator que apresenta variação no tempo: o surto de febre amarela que ocorreu no sudeste brasileiro entre 2016 e 2018 (REZENDE et

al., 2018). Para a preguiça, encontramos sua representação dominada por uma visão positiva, muito provavelmente ligada à boa recepção que a espécie possui em interações reais (PEREIRA et al., 2018) ou a identificação da espécie com traços de personalidade exclusivamente humanos (GREANEY et al., 2016). Todavia, temos ciência de que grandes mamíferos ou animais mais próximos filogeneticamente a seres humanos atraem maior atenção pública por carisma e por despertar sentimentos positivos na população (GUNNTHORSOTTIR, 2001; BATT, 2009; HOME et al., 2009; DUCARME et al., 2012), o que deve ser sempre levado em consideração ao analisar a representação *online* dessas espécies.

Dimensões dos dados: “O que?” e “Quando?”

Quando tratamos cada dimensão da base de dados do Facebook, notamos que principalmente a organização categórica do “o que?” foi fundamental para entender padrões nas demais dimensões. Além disso, encontramos uma tendência quando olhamos para as categorias, a representação de encontros reais com essas espécies (avistamentos de fauna fora e em áreas protegidas, invasão de propriedade por fauna, resgate/ tratamento e soltura de fauna, etc) e a representação apenas virtual da espécie (educação, ação de conservação, etc).

A presença dominante da categoria cativo para onça-pintada, por exemplo, pode ser um indicador da distância entre pessoas e grandes predadores. A representação da espécie a partir dos dados é, majoritariamente, baseada em animais de cativeiro, o que pode corresponder ao fato de que muitas dessas pessoas nunca poderão realmente observar a espécie em seu habitat natural. Entretanto, devemos considerar também o efeito do algoritmo do Facebook ao mostrar um número grande de postagens relacionadas ao Zoológico de São Paulo quando fazemos essa consideração. É interessante que a categoria avistamento de fauna fora de área protegida também tenha aparecido como uma categoria relevante para a onça-pintada. No caso da base de dados do Facebook, isso é relacionado principalmente ao vídeo viral de uma onça-pintada avistada na região de Apiaí, o que é corroborado também pela presença da cidade de Apiaí entre os termos destaque para a espécie. Pessoas se interessam pela possibilidade de ver a espécie em seu habitat. Outro termo em destaque, “filhote”, se relaciona com aspecto carismático da espécie.

Em relação à onça-parda, há uma importância clara para a categoria invasão de propriedade por fauna à registro científico, e em menor escala, resgate, tratamen-

to e soltura de fauna junto. Esta tendência encontrada nas categorias corresponde fortemente ao que estudos de ecologia vêm identificando para a espécie, uma proximidade cada vez maior com agrupamentos humanos (MURPHY; MACDONALD, 2010; MIOTTO et al., 2014). De todo modo, o crescimento de conteúdo relacionado às onças-pardas em áreas urbanas não vinculado às ações de comunicação ou educação podem gerar medo e desconfiança por parte da população nestes lugares.

As categorias no caso do bugio estão, majoritariamente, relacionadas com o tema saúde de fauna silvestre e cativo. Os resultados contribuem para o que muitos autores destacam como sendo uma das funções do tratamento de espécies em cativeiro, auxiliar como ponte entre sociedade e conservação da biodiversidade (CONSORTE-MCCREA, 2019). Além de, como podemos observar com nossos resultados, contribuir para a construção da representação digital de espécies de fauna. É clara a influência do surto de febre amarela no computo geral das categorias, já que essa foi a categoria predominante para o bugio, o que coincide com o que outros autores afirmam sobre o tema para a espécie (BICCA-MARQUES; DE FREITAS, 2010). Da mesma forma, as análises corroboram o impacto da perda de indivíduos na representação digital da espécie. As palavras destacadas, majoritariamente, citaram “morte” e outros elementos relacionados à mortandade de indivíduos, bem como termos relacionados ao bem-estar humano, como “vacina” e “saúde”.

O bicho-preguiça apresentou categorias de postagem que se relacionam com características da espécie, vagarosidade e facilidade na observação de indivíduos quando próximos à agrupamentos humanos (POOLA et al., 2012; PEREIRA et al., 2018). Existe uma constante presença de resgate, tratamento ou soltura de fauna, o que também foi observado na escala temporal para bicho-preguiça. Embora comporte as características de grande mamífero carismático, a percepção positiva da espécie para a população não é largamente reportada.

A utilização da contagem de palavras não se mostrou sozinha como uma variável efetiva para compreender a representação da espécie na rede social. Embora a quantidade de palavras utilizadas tenha grande influência em métricas de engajamento no conteúdo, pessoas gostam de textos concisos e diretos (postagens curtas recebem 66% mais engajamento, sendo 80 caracteres ou menos o ideal para o Facebook e entre 71 e 100 caracteres no Twitter; WE ARE SOCIAL; HOOTSUITE, 2019), os resultados observados para a base do Facebook não revelaram um pa-

drão na construção do texto. Os resultados para o Twitter, todavia, indicam uma distribuição mais compatível com o que se espera da comunicação na rede social.

Além de entender o quanto atores estavam escrevendo sobre as espécies foi fundamental a complementação com o que escreviam. Por exemplo, no caso do bugio, onde postagens estavam mais concentradas em assuntos como febre amarela e morte de indivíduos, textos mais longos foram necessários para comunicar ao público toda a informação sobre o risco da doença. Do mesmo modo, para as outras espécies conseguimos perceber a importância da categorização do tipo de informação disponível nas redes sociais.

A partir da perspectiva do que trata a informação coletada, fomos capazes de obter outras percepções sobre as demais variáveis. Neste ponto, o tipo de mídia demonstrou ser uma variável com um comportamento semelhante por espécie, sendo que a maior parte das postagens relacionadas às espécies continha pouco texto. Isso demonstra a importância de imagens na comunicação, o que é ainda mais claro quando olhamos o crescimento de redes sociais baseadas no compartilhamento de fotos e vídeos, como o Instagram (TING et al., 2015). Complementar à isto, vemos também que a sobreposição de categorias com os padrões temporais nos ajudam a compreender porquê determinados eventos tem maior significância para a espécie. Por exemplo, os dois picos encontrados para a onça-pintada e onça-parda estão relacionados a eventos específicos onde indivíduos foram avistados em áreas externas a parques e ambientes urbanos. No caso da onça-parda em 2019, as publicações se referiam ao viral “onça-parda na rua do porto” que ganhou grande repercussão na mídia, embora, especificamente para a onça-parda, a categoria de resgate, tratamento ou soltura de fauna tenha sido a categoria mais relevante em número absoluto de postagens. Os picos temporais para bugio também seguiram a tendência encontrada para saúde de fauna silvestre, e para bicho-preguiça foram majoritariamente associados à resgate/soltura e tratamento de fauna e avistamento de fauna fora de área protegida.

Dimensões dos dados: “Onde?”

Quando analisamos os padrões espaciais na base de dados do Facebook podemos destacar a influência do município de São Paulo em concentrar eventos que posteriormente estariam disponíveis na rede social. Destacamos a importância em se considerar o efeito do algoritmo do Facebook, que pode acabar causando repeti-

ção de tópicos e favorecer a exibição de determinado conteúdo em função do usuário (SCHÄFER, 2020).

O efeito do município de São Paulo, embora possa corresponder a um viés pelo próprio algoritmo se mostrou menos proeminente quando comparamos o conteúdo por espécie. No que diz respeito à onça-parda e bicho-preguiça, temos padrões de eventos em municípios que seguem características da paisagem condizentes com a realidade biológica da onça-pintada, já que os municípios que não correspondem diretamente à região metropolitana de São Paulo se distribuem por áreas do contínuo Mata Atlântica (CONDE et al., 2010; CULLEN Jr et al., 2013). Embora, bichos-preguiças apresentem certo nível de tolerância a paisagens fragmentadas (NEAM; LATCHER, 2015) o resultado parece demonstrar um maior índice de eventos relacionados à espécie também na área do contínuo. Sabemos que o número de onças-pardas vem crescendo no interior do estado de São Paulo (MIOTTO et al., 2014), neste caso observamos um grande número de eventos para onça-parda na região central do estado. Embora o número de postagens para bugio esteja quase totalmente relacionado à eventos que ocorrem no município de São Paulo, ainda conseguimos observar padrões espaciais associados a municípios dentro do corredor da Serra do Mar (ANZURES-DADDA; MANSON, 2007).

Os dados por espécie além de permitirem visualizar padrões espaciais para a concentração de eventos que motivam postagens em redes sociais levantam dois pontos importantes. O primeiro está relacionado à grande concentração de parques e espaços de conservação nos municípios em destaque (MINISTÉRIO DO MEIO AMBIENTE, 2019), principalmente quando olhamos para a localização precisa do evento que motivou a postagem. Isto também aparece na frequência com que palavras como “parque” surgem nas postagens. Estes dados, mais uma vez, confirmam a importância que estas áreas possuem para biodiversidade de modo geral (TABARELLI, et al. 2005; RYLANDS, BRANDON, 2005) e, por consequência, para representação de espécies no ambiente virtual.

O segundo está relacionado à disponibilidade de internet no estado de São Paulo e na região sudeste, que estão localizados na área com a melhor relação qualidade/cobertura de internet no Brasil atualmente (SIMET, 2020), além de concentrar a maior população de pessoas que utilizaram internet entre 2016 e 2017 (IBGE, 2017). Mesmo num estado desenvolvido como São Paulo, existem desafios em se acessar a percepção pública em municípios com baixa cobertura de rede e baixo

acesso à internet. Uma grande parte destes municípios ocorre em trechos importantes de Mata Atlântica conservada, onde a percepção pública seria especialmente interessante para traçar estratégias de conservação. De fato, em entrevistas realizadas em áreas remotas ao longo do Parque Estadual Turístico do Alto Ribeira e do Parque Estadual da Serra do Mar, moradores demonstram baixo interesse em reportar eventos naturais não associados com a beleza cênica em redes sociais (RIBEIRO dados não publicados¹; ARTS et al., 2018).

Existe um potencial no aproveitamento de dados espaciais advindos de redes sociais, o que já foi anteriormente reportado por outros autores (WOOD et al., 2013; SPALDING, et al., 2017; ELQADI, et al., 2017). É importante estar atento à esta dimensão em especial por sua importância no planejamento estratégico da conservação (MARGULES et al., 2000). Entretanto, existe uma diferença grande nos usos de dados espaciais a partir de análises de dados de redes sociais. A abordagem adotada aqui, a filtragem por eventos, necessitaria de maior atenção para ser executada de forma automatizada. Normalmente, lida-se com mais facilidade com a localização do dispositivo que gerou a postagem. Mesmo com o uso de localização, a cobertura de rede disponível em países sulamericanos ainda não alcança por completo o território, como, por exemplo, no Brasil (SIMET, 2020), e possivelmente não obteríamos os mesmos resultados encontrados para análises semelhantes feitas em países com larga cobertura e acesso, como a Finlândia (JÄRV, 2020).

Dimensões dos dados: “*Quem?*”

Em relação aos responsáveis por gerar conteúdo em redes sociais, encontramos um padrão que divide este grupo entre páginas pessoais e páginas de segmentos. No primeiro caso é interessante observar o peso que páginas pessoais apresentaram no compartilhamento de conteúdo relacionado às espécies. Esse é um resultado positivo por levantar a importância das pessoas na disseminação de conteúdo sobre espécies. As redes sociais têm revolucionado a maneira como nos moldamos enquanto sociedade e como comunicamos uns com os outros. Deste modo, quando observamos que pessoas são as grandes responsáveis por postagens relacionadas à fauna, embora seja um resultado que necessite de mais dados para

¹ Dados coletados para o Projeto nº 1927 20171 “Conservação de mamíferos ameaçados de extinção no Corredor de Biodiversidade da Serra do Mar: integrando ecologia, dimensões humanas e planejamento” financiado pela Fundação Boticário.

uma validação de tendência, podemos traçar um paralelo com o poder do compartilhamento pessoal de informação, já observado em movimentos sociais (KIDD; MCINTOSH, 2016).

Outro ponto importante é a presença de páginas relacionadas a órgãos governamentais e organizações para representação destas espécies no universo digital. Esta *cyber* realidade acaba por criar um universo onde o contato de pessoas e fauna, em muito, ocorre apenas através da mídia disponibilizada em redes sociais. Vemos isso em outros casos, como espécies de fauna silvestre traficadas (DI MININ et al., 2018), exploradas como *pets* e vídeos virais relacionados a elas (NEKARIS et al., 2013), onde advertências são geradas por instituições de conservação e, em última análise, por usuários que demonstram comportamento empático em relação às espécies. Esse campo possibilita identificar potenciais *stakeholders* para conservação destas espécies (DI MININ et al., 2015; KNIGHT et al., 2006). De todo modo, uma análise mais profunda sobre motivações e perfil de usuários que postam sobre a fauna seria necessária para entender por completo esta esfera e torná-la aplicável na consideração de ações de conservação.

Outro fato importante destacado pelos resultados é o papel de atores como páginas de entretenimento e negócios. Muitas vezes, essas páginas são responsáveis por compartilhar vídeos impressionantes ou divertidos sobre a fauna que acabam também por cumprir um papel importante ao disseminar informações e alcançar um público que nem sempre está interessado em conservação. Por fim, páginas de notícias foram importantes para muitas espécies e são um dos principais veículos na comunicação de eventos relacionados à fauna. Como Killion et al. (2018) observam, notícias são capazes de influenciar a percepção pública em relação às espécies ao longo do tempo, ou mesmo em relação a percepção de achados científicos, como no caso da categoria registro científico para onça-parda.

Facebook e Twitter

Quando comparamos o que conseguimos apreender com os dados da base do Facebook com a base do Twitter percebemos uma dificuldade em se aplicar a maioria das análises obtidas com um conjunto menor de dados em uma base maior como a do Twitter. De modo geral não foi possível identificar categorias de informação, o que dificultou a compreensão sobre do que se tratavam os dados obtidos no Twitter. O *ranking* de palavras utilizado, porém, pode facilitar a identificação de se-

melhanças entre as bases. Termos relacionados a atropelamento, avistamento de fauna fora de área protegida ou invasão de propriedade por fauna demonstram que essas categorias possivelmente possuem uma importância dentro deste conjunto de dados. É importante notar que, mesmo com outro recorte espacial, termos ainda aparecem em destaque. Entretanto, foi possível perceber a importância de outras características dos dados quando olhamos para o conjunto do Twitter. Vemos Pantanal aparecer como uma palavra relativamente importante para onça-pintada, o que é condizente com a realidade da espécie, já que no bioma sua presença e contato com seres humanos é muito mais comum (CAVALCANTI et al., 2012; MARCHINI; MACDONALD, 2012). De todo modo, não notamos menção à outros biomas, como a Amazônia, por exemplo. Também não foi possível inferir sobre a distribuição espacial dos eventos que motivaram postagens no Twitter. Acreditamos que isso levante o ponto da dificuldade que eventos relacionados à fauna ocorrendo em áreas remotas do país possam chegar às redes sociais. Outro ponto importante é o fato de que a representatividade do Twitter no Brasil é baixa quando comparado à outras redes sociais como o Facebook (HOOTSUITE, 2009), o que o coloca em uma posição de rede social complementar.

O texto apresentado no Twitter é mais curto do que o observado no Facebook, o que condiz com o padrão de mensagens curtas e mais diretas encontrados para a rede social. Outro ponto importante nos dados obtidos para o Twitter é o da análise da linha temporal de postagens. Quando olhamos para a da onça-pintada, vemos um efeito grande da morte da onça-pintada Juma, pouco antes do início dos jogos olímpicos de 2016. Esse mesmo pico não aparece em linhas do tempo para o termo onça-pintada no *google trends*, mas apresenta efeito semelhante a repercussão da morte do leão Cecil (MACDONALD et al., 2016), embora em uma escala menor. Os picos para a onça-parda são menos acentuados, mas mais constantes, em uma checagem das postagens, em muitos casos relacionados a eventos de proximidade da espécie com seres humanos em eventos de invasão de propriedade ou avistamento em áreas não protegidas. Esses resultados são condizentes com o que encontramos para o Facebook.

Outro ponto importante na base obtida com o Twitter é a influência que usuários têm na disseminação da informação. Não obtivemos dados dos usuários em relação ao número de amigos ou número de seguidores em páginas no Facebook, mas conseguimos essas informações para o Twitter. Em uma rede social com um

alcance menor em relação à população brasileira, o poder que perfis de usuários possuem em disseminar dados importantes para a conservação da fauna é de extrema relevância (ANGER; KITTL, 2011). A dispersão destes dados mostra que não existe um padrão em relação aos perfis de usuário que disseminam informações de fauna no Twitter, tanto para a quantidade de amigos quanto para quantidade de seguidores, embora esse seja um dado relevante para entendermos a quantidade de pessoas que essas postagens podem alcançar.

A classificação dos dados da base do Twitter demonstrou o potencial que aliar análises em menor escala com grandes agrupamentos de dados possui. Embora o método ainda necessite de ajustes, é possível perceber que muito do que aprendemos na base do Facebook pode ser aplicado e validado em uma base maior. Essa abordagem é inovadora e segue uma tendência em se utilizar algoritmos para grandes classificações de dados (ANGUS et al. 2010; DI MININ et al., 2018). Os resultados ainda podem ser melhorados e outros algoritmos poderiam ser utilizados para análise de desempenho e capacidade de classificação.

Engajamento como quinta dimensão dos dados

A quinta dimensão dos dados obtida neste estudo, engajamento, contribuiu para elaborar conclusões sobre como pessoas reagem ou sobre o impacto que informações relacionadas à fauna têm sobre a população. Além disso, engajamento se mostrou também como uma variável importante que pode aproximar o que entendemos como dimensão humana da conservação com os avanços em conservação digital. De modo geral, nossos resultados em relação à engajamento corroboram a presença icônica destas espécies na composição do imaginário popular. Além disso, vimos que a significância que o engajamento possui pode revelar quais são os tipos de informação com maior impacto sobre a população. Por exemplo, para onça-pintada observamos padrões altos de engajamento positivo (reações positivas no Facebook), e mesmo com medianas baixas no Twitter, encontramos outliers com grandes valores de engajamento positivo e compartilhamento.

Os valores indicam que as pessoas se engajam positivamente em relação aos tópicos relacionados à onça pintada. O mesmo padrão é observado para onça-parda, onde reações de surpresa (*wow*) também despontam, assim como para bicho-preguiça, quase totalmente dominado por reações positivas. Além disso, onça-parda e bicho-preguiça são as duas espécies que agregam o maior número de rea-

ções no Facebook, embora no Twitter onça-parda e onça-pintada tenham atraído valores muito semelhantes de engajamento. Ainda que não exista uma literatura específica relacionada à engajamento e fauna silvestre em redes sociais, podemos aplicar conceitos de marketing e como tipos de informação tem influência sobre como o público se engaja com postagens (STEPHEN et al., 2015). O bugio é uma exceção, sendo o único a apresentar reações negativas em relação à engajamento, muito provavelmente pela perda de indivíduos e sua relação com a febre amarela. Para alguns autores, isso se refere ao efeito que postagens com conteúdo negativo têm ao receber engajamento, significando a preocupação popular com os temas (BITTER; GRABNER-KRÄUTER, 2016).

O resultado da regressão para encontrar potenciais preditores de engajamento revelou primeiro uma variação grande do que pode contribuir para altos níveis de engajamento entre espécies. Não há também literatura específica que debata preditores de sucesso para engajamento em conteúdo de fauna, embora esse tipo de análise já tenha sido realizado por diferentes autores (HEISS et al., 2018, CARD et al., 2018). Para onça-pintada, o tipo de mídia foi muito importante para determinar engajamento, talvez pela presença icônica da espécie e pelo fato de sua representação ser suficiente para atrair atenção e interesse público. Assim como no Twitter, para onça-pintada, o tipo de mídia foi também importante. Como esperado, a influência de usuários foi igualmente importante, embora o tipo de mídia “texto” tenha apresentado associação negativa com engajamento.

Para a onça-parda, é destacado o papel de páginas de associações, fundações, institutos e ONGs, já que as demais com resultado significativo apresentaram associação negativa com engajamento. Para onça-parda no Facebook, uma categoria com baixa representatividade em número acabou se destacando em engajamento: avistamento de fauna em área protegida, o que tem efeito positivo, reforçando a importância destas áreas para conservação. Quando olhamos para a base do Twitter, percebemos que novamente postagens com fotos atraem maiores valores de engajamento. Em contrapartida, número de amigos parece apresentar associação negativa com altos valores de engajamento. Não encontramos fontes que corroborem ou expliquem o porquê específico destas associações no que tange o escopo deste trabalho.

Em relação ao bicho-preguiça, apenas duas variáveis apresentaram associação positiva com engajamento. Neste caso, destacamos mais uma vez o impacto

que algumas categorias possuem em relação à opinião pública, já que uma delas é a categoria atropelamento. Outro importante ponto é a categoria instituição governamental demonstrar o potencial da espécie para aproximar ações com a população, já que foi a única categoria de usuário com associação positiva com engajamento. Não encontramos bons preditores para bugio, o que, infelizmente, pode indicar que mesmo em casos onde é necessária a atenção para perda de indivíduos ainda não conseguimos encontrar determinantes para que essas informações sejam amplamente compartilhadas pelo público. De todo modo, estes resultados necessitam de validação uma vez que a literatura científica disponível não nos permitiu um debate aprofundado sobre a assertividade dos resultados obtidos.

Este estudo traz reflexões importantes sobre como as análises de dados de redes sociais podem variar. A realidade brasileira ainda não permite que muitas das informações estejam disponíveis *online*, assim como a realidade em outros países da América Latina. Ainda não conseguimos bancos de dados completos como autores têm obtido em estudos conduzidos em outras áreas, e isso varia também em relação ao tipo de estudo. Assim como estamos limitados a atores com acesso às redes sociais e tecnologia, e muitas vezes não temos informações sobre atores que convivem com determinadas espécies. Todavia, o desenvolvimento tecnológico permitirá que possamos ter cada vez mais acesso a estes atores através de redes sociais, dado que é esperado um crescimento na utilização destes serviços (HOO-TSUITE, 2019). Além disso, como observado por Correia et al. (2019), a utilização de dados de redes sociais para inferir interesse público requer cuidado ao discutir e gerar generalizações.

O método ainda tem limitações. O acesso à informação pertencente a empresas como o Facebook não está disponível em muitos casos, e companhias que permitem a utilização de dados pela ciência, como o Twitter, nem sempre tem a mesma força e alcance público. Além disso, o fato de ainda estarmos caminhando para solidificação de uma metodologia no assunto torna difícil a comparação entre achados ou mesmo a estruturação de uma metodologia de pesquisa adequada e replicável.

Implicações para conservação

As implicações para conservação deste estudo residem, justamente, nos primeiros passos para integração e consideração da dimensão virtual na conservação aplicada. Ainda, consideramos o potencial de disseminação da informação digital em

planos de ação de forma tímida. Nos planos de ação para Grandes Felinos (ICMBIO, 2018) e para Primatas da Mata Atlântica e da Preguiça-de-Coleira (ICMBIO, 2018) ações que tratam da comunicação e amplificação de informações não citam diretamente redes sociais, embora tenham um grande potencial de envolvê-las.

Certamente, o desenvolvimento de estratégias pode contribuir para que o público entenda e atue em conjunto para conservação e aplicação de ações de conservação. Além disso, este trabalho evidencia que a dimensão humana se estende para o meio digital, o que em uma conservação 2.0 levará futuros profissionais a lidarem com a complexidade de incorporar este tipo de análise na tomada de decisão. Ladle (2018) pontua bem que profissionais precisam se expor aos desafios em lidar com estes dados e gerar informação para que consigamos incorporar este tipo de análise em conservação e manejo. Por exemplo, se queremos tratar de assuntos como a febre amarela, ou qualquer outro, precisaremos traçar estratégias pensando em engajamento e características de espécies com bons resultados em comunicação digital. Igualmente, se quisermos que ações tenham sua dimensão amplificada, podemos representar espécies com grande importância local a partir de características relacionadas ao sucesso de postagens com outras espécies mais conhecidas.

Além disso, o desenvolvimento de sistemas de monitoramento da opinião pública em tempo real pode ser uma realidade e colaborar com a efetividade de ações, elevando tanto o alcance de ações de conservação como o de instituições ligadas à conservação. A utilização de forma organizada dessas ferramentas já é uma realidade para diversos setores e pode ser também para a conservação de fauna.

6. CONCLUSÕES

1) Dados de redes sociais são importantes para entender como espécies icônicas da fauna estão sendo representadas no ambiente virtual.

2) Uma análise deste tipo necessita considerar de maneira integrada as diferentes dimensões que compõe os dados de redes sociais relacionados à fauna.

3) Em relação às espécies, no Facebook e Twitter encontramos que: para a onça-pintada há presença majoritária de informações associadas com cativeiro e avistamento de fauna fora de área protegida, a espécie recebe grande atenção pública positiva, sua representação na esfera digital é composta por encontros distantes de seu habitat e que estratégias de engajamento para o felino podem ser amparadas no poder da imagem da espécie em relação ao público.

4) Para a onça-parda, há presença majoritária de informações associadas com invasão de propriedade por fauna e registro científico, a espécie apresenta marcante número de encontros reais em áreas urbanas divulgados nas redes sociais e estratégias de engajamento podem contribuir na redução do medo e desconfiança em relação à presença ascendente da espécie próxima a agrupamentos humanos.

5) Para o bugio, há presença majoritária de informações relacionadas à saúde de fauna silvestre (febre amarela) e avistamento de fauna fora de área protegida. A análise do engajamento recebido pela espécie demonstra preocupação pública com a mortandade de indivíduos, mas ações de comunicação precisam ser aplicadas a fim de garantir maior disseminação da informação e participação de *stakeholders* através do engajamento.

6) Para o bicho-preguiça, encontramos que os principais temas relacionados à espécie estão associados à resgate, soltura e tratamento de fauna silvestre e avistamento de fauna fora de área protegida, de modo geral a espécie apresenta auto nível de engajamento positivo e potencial para sua utilização em ações de conservação, o que poderia beneficiar espécies ameaçadas do mesmo gênero.

7) Engajamento é uma variável que complementa nossa compreensão sobre dados e informa como a percepção da população é atraída por determinado conteúdo ou por tipo de páginas.

8) Estudos com metodologia automatizada apresentam vantagens quando planejados sobre pequenos agrupamentos de dados, que podem conter detalhes que melhoram análises e complementam achados.

REFERÊNCIAS

- ALALWAN, A. A. et al. Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, v. 34, n. 7, p. 1177-1190, 2017.
- ANGER, I.; KITTL, C. Measuring influence on Twitter. Em: Proceedings of the 11th international conference on knowledge management and knowledge technologies, p. 1-4, 2011.
- ANGUS, E.; STUART, D.; THELWALL, M. Flickr's potential as an academic image resource: an exploratory study. *Journal of Librarianship and Information Science*, v. 42, n. 4, p. 268-278, 2010.
- ANZURES-DADDA, A.; MANSON, R. H. Patch-and landscape-scale effects on howler monkey distribution and abundance in rainforest fragments. *Animal Conservation*, v. 10, n. 1, p. 69-76, 2007.
- ARTS, K. et al. *Online* and offline representations of biocultural diversity: A political ecology perspective on nature-based tourism and indigenous communities in the Brazilian Pantanal. *Sustainability*, v. 10, n. 10, p. 3643, 2018.
- BATISTA, A. F. M.; CORREA, P. L. P.; PALANISAMY, G. Visual Analytics Improving Data Understandability in IoT Projects: An Overview of the US DOE ARM Program Data Science Tools. In: 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). IEEE, 2016. p. 349-354.
- BATRINCA, B.; TRELEAVEN, P. C. Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, v. 30, n. 1, p. 89-116, 2015.
- BATT, S. Human attitudes towards animals in relation to species similarity to humans: a multivariate approach. *Bioscience horizons*, v. 2, n. 2, p. 180-190, 2009.
- BECKEN, S. et al. Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. *Journal of environmental management*, v. 203, p. 87-97, 2017.
- BICCA-MARQUES, J. C.; DE FREITAS, D. S. The role of monkeys, mosquitoes, and humans in the occurrence of a yellow fever outbreak in a fragmented landscape in south Brazil: protecting howler monkeys is a matter of public health. *Tropical Conservation Science*, v. 3, n. 1, p. 78-89, 2010.

- BITTER, S.; GRABNER-KRÄUTER, S. Consequences of customer engagement behavior: when negative Facebook posts have positive effects. *Electronic Markets*, v. 26, n. 3, p. 219-231, 2016.
- BÖHNING, D. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, v. 44, n. 1, p. 197-200, 1992.
- BREIMAN, Leo. Random forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.
- BRODIE, R. J. et al. Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, v. 66, n. 1, p. 105-114, 2013.
- BRODKIN, J. *Web 2.0: Buzzword, or Internet revolution?* O'Reilly Radar, 24 jan. 2007. Disponível em: <<https://web.archive.org/web/20070603171459/http://www.networkworld.com/news/2007/012407-web-20.html>>. Acessado em: 10 de Outubro de 2019.
- BRYMAN, A. The Debate about Qualitative and Quantitative Research. A Question of Methods or Epistemology. *The British Journal of Sociology*. V. 35(1), p. 75-92, 1984.
- CAMPBELL, H.; ENGELBRECHT, I. The Baboon Spider Atlas—using citizen science and the 'fear factor' to map baboon spider (Araneae: Theraphosidae) diversity and distributions in Southern Africa. *Insect Conservation and Diversity*, v. 11, n. 2, p. 143-151, 2018.
- CASTELLS, M. The impact of the internet on society: a global perspective. *Change*, v. 19, p. 127-148, 2014.
- CAVALCANTI, SANDRA MC et al. The status of the jaguar in the Pantanal. *Cat News Special Issue*, v. 7, p. 29-34, 2012.
- CLARK, S. Measuring API Usability. Dr. Dobbs's The World of Software Development, mail. 2004. Disponível em: <<https://www.drdoobbs.com/windows/measuring-api-usability/184405654>>. Acessado em 10 de Dezembro de 2019.
- CLEMENT, J. Number of Internet Users Worldwide 2005-2018. Statista Research De, 2019.
- CLEVELAND, W. S. Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, v. 69, n. 1, p. 21-26, 2001.
- CONDE, D. A. et al. Sex matters: Modeling male and female habitat differences for jaguar conservation. *Biological Conservation*, v. 143, n. 9, p. 1980-1988, 2010.

- CONFORTI, V. A.; DE AZEVEDO, F. C. C. Local perceptions of jaguars (*Panthera onca*) and pumas (*Puma concolor*) in the Iguazu National Park area, south Brazil. *Biological conservation*, v. 111, n. 2, p. 215-221, 2003.
- CONSORTE-MCCREA, Adriana et al. Large carnivores and zoos as catalysts for engaging the public in the protection of biodiversity. *Nature Conservation*, v. 37, p. 133, 2019.
- CORREIA, R. A. et al. Inferring public interest from search engine data requires caution. *Frontiers in Ecology and the Environment*, v. 17, n. 5, 2019.
- CRISTANCHO, S.; VINING, J. Culturally defined keystone species. *Human Ecology Review*, p. 153-164, 2004.
- CULLEN JUNIOR, L. et al. Selection of habitat by the jaguar, *Panthera onca* (Carnivora: Felidae), in the upper Paraná River, Brazil. *Zoologia (Curitiba)*, v. 30, n. 4, p. 379-387, 2013.
- DEAN, W. *With broadax and firebrand: the destruction of the Brazilian Atlantic Forest*. University of California Press, 1997.
- DECKER, D. J.; LAUBER, T. B.; SIEMER, W. F. *Human-wildlife conflict management*. 2002.
- DEERWESTER, S. et al. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, v. 41, n. 6, p. 391–407, 1990.
- DEKHTYAR, A. *Lecture Notes on Data Science - DATA 301*. California Polytechnic State University. 2016. Disponível em: <http://users.csc.calpoly.edu/~dekhtyar/Data301-Spring2016/lectures/lec03.301.pdf>>. Acessado em 10 de Setembro de 2018.
- DI CAMILLO, C. G. et al. Building a baseline for habitat-forming corals by a multi-source approach, including Web Ecological Knowledge. *Biodiversity and Conservation*, v. 27, n. 5, p. 1257-1276, 2018.
- DI MININ, E. et al. A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation biology*, v. 33, n. 1, p. 210-213, 2019.
- DI MININ, E. et al. Machine learning for tracking illegal wildlife trade on social media. *Nature ecology & evolution*, v. 2, n. 3, p. 406-407, 2018.

- DI MININ, E.; MOILANEN, A. Improving the surrogacy effectiveness of charismatic megafauna with well-surveyed taxonomic groups and habitat types. *Journal of Applied Ecology*, v. 51, n. 2, p. 281-288, 2014.
- DI MININ, E.; TENKANEN, H.; TOIVONEN, T. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, v. 3, p. 63, 2015.
- DUCARME, F.; LUQUE, G. M.; COURCHAMP, Franck. What are “charismatic species” for conservation biologists. *BioSciences Master Reviews*, v. 10, n. 2013, p. 1-8, 2013.
- ELQADI, M. M. et al. Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. *Ecological informatics*, v. 39, p. 23-31, 2017.
- ESRI (ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE). ArcGIS 10.3. 1 for desktop. 2015.
- FACEBOOK. Informações sobre a rede social. Disponível em:<<https://www.businessinsider.com/facebook-has-2-billion-plus-users-after-15-years-2019-2>>. Acessado em: 10 de Outubro de 2019.
- FACEBOOK. Restrições de uso de API. Disponível em:<<https://about.fb.com/news/2018/04/restricting-data-access/> Facebook API>. Acessado em: 10 de Outubro de 2019.
- FELT, M. Social media and the social sciences: How researchers employ *Big Data* analytics. *Big Data & Society*, v. 3, n. 1, p. 2053951716645828, 2016.
- FINK, C.; HAUSMANN, A.; DI MININ, E. *Online* sentiment towards iconic species. *Biological Conservation*, p. 108289, 2019.
- FORREST, B. Controversy about our “*Web 2.0*” service mark. *O’reilly Radar*, 25 mai. 2006. Disponível em:<https://web.archive.org/web/20080129105950/http://radar.oreilly.com/archives/2006/05/controversy_about_our_web_20_s.html>. Acessado em: 10 de Outubro de 2019.
- FOSTER, I. et al. *Big Data* and social science: A practical guide to methods and tools. Chapman and Hall/CRC, 2016.
- FRIESE, S. *Qualitative data analysis with ATLAS. ti*. SAGE Publications Limited, 2019.

- FUKANO, Y.; SOGA, M. Spatio-temporal dynamics and drivers of public interest in invasive alien species. *Biological Invasions*, v. 21, n. 12, p. 3521-3532, 2019.
- GAMBETTI, R. C.; GRAFFIGNA, G. The concept of engagement: A systematic analysis of the ongoing marketing debate. *International Journal of Market Research*, v. 52, n. 6, p. 801-826, 2010.
- GEBELS, M. Número de usuários facebook. *Business Insider*, fev. de 2019. Disponível em: < <https://www.businessinsider.com/facebook-has-2-billion-plus-users-after-15-years-2019-2>>. Acessado em: 10 de Dezembro de 2019.
- GLUCK, M. Digital Ad Engagement: An industry overview and reconceptualization. Interactive Advertising Bureau, 2013.
- GREANEY, M. Laziness: a literary-historical perspective. In: *The Restless Compendium*. Palgrave Macmillan, Cham, 2016. p. 183-190.
- GUNNTHORSOTTIR, A. Physical attractiveness of an animal species as a decision factor for its preservation. *Anthrozoös*, v. 14, n. 4, p. 204-215, 2001.
- HARLOW, S. Social media and social movements: Facebook and an *online* Guatemalan justice movement that moved offline. *New Media & Society*, v. 14, n. 2, p. 225-243, 2012.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc., 2001. (Springer Series in Statistics).
- HAUSMANN, A. et al. Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, v. 11, n. 1, p. e12343, 2018.
- HEISS, R.; SCHMUCK, D.; MATTHES, J. What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, v. 22, n. 10, p. 1497-1513, 2019.
- HOME, R. et al. Selection criteria for flagship species by conservation organizations. *Environmental Conservation*, v. 36, n. 2, p. 139-148, 2009.
- HOOTSUITE. Comprimento de teto para diferentes redes sociais. Disponível em: <<https://blog.hootsuite.com/ideal-social-media-post-length>>. Acessado em: 11 de Dezembro de 2019.
- IBGE. Acesso á internet e posse de telefone móvel celular para uso pessoal. Rio de Janeiro: IBGE, 2013. 157 p. Pesquisa Nacional por Amostra de Domicílios.

- IBGE. Estimativas da população residente com data de referência de 1º de julho de 2019. Brasília, 2019.
- IBGE. Pesquisa Nacional de Acesso Digital – Apresentação com resultados. Disponível em: <https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/9e88a636785c573625be2c5632bd3087.pdf>. Acessado em: 10 de Outubro de 2019.
- ICMBIO. Informações sobre espécies. Disponível em: <<http://www4.icmbio.gov.br/cpb/index.php/tatus-tamanduas-e-preguicas-brasileiros>>. Acessado em: 10 de Outubro de 2019.
- ICMBIO. Matriz de planejamento: Plano de Ação Nacional para Conservação dos Grandes Felinos. 2018 Disponível em: <<http://www.icmbio.gov.br/portal/faunabrasileira/planos-de-acao/9326-plano-de-acao-nacional-para-a-conservacao-dos-grandes-felinos>>. Acessado em: 02 de Janeiro de 2019.
- ICMBIO. Matriz de planejamento: Plano de Ação Nacional para Conservação dos Primatas da Mata Atlântica e da Preguiça-de-coleira. 2018. Disponível em: <<http://www.icmbio.gov.br/portal/faunabrasileira/planos-de-acao/8330-plano-de-acao-nacional-para-a-conservacao-dos-primatas-e-preguica-mata-atlantica>>. Acessado em: 02 de Janeiro de 2019.
- IUCN. Consulta de informações gerais sobre espécies. Disponível em: <<https://www.iucnredlist.org/species>>. Acessado em: 10 de Outubro de 2019.
- JÄRV, O. Can we use Twitter data to estimate population distribution in Finland? Digigeolab Blog, 12 jan. 2020. Disponível em: <<https://blogs.helsinki.fi/digital-geography/2020/01/12/estimating-finnish-population-from-twitter-data/>>. Acessado em: 10 de Outubro de 2019.
- JURAFSKY, D.; MARTIN, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 1st. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.
- JUSSILA, J. J.; KÄRKKÄINEN, H.; ARAMO-IMMONEN, H. Social media utilization in business-to-business relationships of technology industry firms. Computers in Human Behavior, v. 30, p. 606-613, 2014.
- KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, v. 53, n. 1, p. 59-68, 2010.

- KEEFER, A. C.; BAIGET, T. How it all began: a brief history of the Internet. *Vine*, v. 31, n. 3, p. 90-95, 2001.
- KHAN, M. L. Social media engagement: What motivates user participation and consumption on YouTube?. *Computers in Human Behavior*, v. 66, p. 236-247, 2017.
- KIDD, D.; MCINTOSH, K. Social media and social movements. *Sociology Compass*, v. 10, n. 9, p. 785-794, 2016.
- KILLION, A. K. et al. Tracking a half century of media reporting on gray wolves. *Conservation Biology*, v. 33, n. 3, p. 645-654, 2019.
- KNIGHT, A. T.; COWLING, R. M.; CAMPBELL, B. M. An operational model for implementing conservation action. *Conservation biology*, v. 20, n. 2, p. 408-419, 2006.
- LADLE, R. J. et al. Conservation culturomics. *Frontiers in Ecology and the Environment*, v. 14, n. 5, p. 269-275, 2016.
- LADLE, R. J. Mining social media The world of abundant “messy” data and what marine conservation and management can learn from it. [Entrevista]. *Meam – Marine Ecology and Management*, 08 mai. 2018.
- LEINER, B. M. et al. A brief history of the Internet. *ACM SIGCOMM Computer Communication Review*, v. 39, n. 5, p. 22-31, 2009.
- LORENA, A. C.; DE CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43-67, 2007.
- MACDONALD, C. et al. Conservation potential of apex predator tourism. *Biological Conservation*, v. 215, p. 132-141, 2017.
- MACDONALD, D. W. et al. Cecil: a moment or a movement? Analysis of media coverage of the death of a lion, *Panthera leo*. *Animals*, v. 6, n. 5, p. 26, 2016.
- MANFREDO, M. J. Human dimensions of wildlife management. *Wildlife Society Bulletin (1973-2006)*, v. 17, n. 4, p. 447-449, 1989.
- MANFREDO, M. J. Who cares about wildlife?. In: *Who Cares About Wildlife?*. Springer, New York, NY, 2008. p. 1-27.
- MARCHINI, S. et al. Planning for coexistence in a complex human-dominated world. *Human-Wildlife Interactions: Turning Conflict into Coexistence*; Frank, B., Glikman, JA, Marchini, S., Eds, p. 414-438, 2019.

- MARCHINI, S.; MACDONALD, D. W. Predicting ranchers' intention to kill jaguars: case studies in Amazonia and Pantanal. *Biological Conservation*, v. 147, n. 1, p. 213-221, 2012.
- MARGULES, C. R.; PRESSEY, R. L. Systematic conservation planning. *Nature*, v. 405, n. 6783, p. 243, 2000.
- MCAFEE, A. et al. *Big Data*: the management revolution. *Harvard business review*, v. 90, n. 10, p. 60-68, 2012.
- MCLENNAN, M. R.; SPAGNOLETTI, N.; HOCKINGS, K. J. The implications of primate behavioral flexibility for sustainable human–primate coexistence in anthropogenic habitats. *International Journal of Primatology*, v. 38, n. 2, p. 105-121, 2017.
- MCLENNAN, M. R.; SPAGNOLETTI, N.; HOCKINGS, K. J. The implications of primate behavioral flexibility for sustainable human–primate coexistence in anthropogenic habitats. *International Journal of Primatology*, v. 38, n. 2, p. 105-121, 2017.
- MINISTÉRIO DO MEIO AMBIENTE. Mapa de áreas protegidas. Disponível em:< <https://www.mma.gov.br/areas-protegidas.html>>. Acessado em: 10 de Dezembro de 2019.
- MIOTTO, R. A. et al. Estimating puma *Puma concolor* population size in a human-disturbed landscape in Brazil, using DNA mark-recapture data. *Oryx*, v. 48, n. 2, p. 250–257, 2014
- MIOTTO, R. A. et al. Monitoring a puma (*Puma concolor*) population in a fragmented landscape in southeast Brazil. *Biotropica*, v. 44, n. 1, p. 98-104, 2012.
- MITTERMEIER, R. A. et al. Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences*, v. 100, n. 18, p. 10309-10313, 2003.
- MITTERMEIER, R. A.; MITTERMEIER, C. G.; GIL, P. R. Megadiversity: Earth's biologically wealthiest nations. *Chelonian Conservation and Biology*, v. 3, n. 3, p. 537-537, 1999.
- MURPHY, T.; MACDONALD, D. W. Pumas and people: lessons in the landscape of tolerance from a widely distributed felid. *Macdonald DW*, p. 431-451, 2010.
- MYERS, N. et al. Biodiversity hotspots for conservation priorities. *Nature*, v. 403, n. 6772, p. 853, 2000.

- NEAM, K. D.; LATCHER, T. Spatial distribution, resource use, and behavior of brown-throated sloths (*Bradypus variegatus*) in a multi-use landscape. *Edentata*, v. 16, p. 46-56, 2015.
- NEKARIS, K. A. et al. Tickled to death: analysing public perceptions of 'cute' videos of threatened species (slow lorises—*Nycticebus* spp.) on *Web 2.0* Sites. *PloS one*, v. 8, n. 7, 2013.
- NIE, N. H.; ERBRING, L. Internet and society: A preliminary report. *IT & society*, v. 1, n. 1, p. 275-283, 2002.
- PALMEIRA, F. B. L. et al. Cattle depredation by puma (*Puma concolor*) and jaguar (*Panthera onca*) in central-western Brazil. *Biological conservation*, v. 141, n. 1, p. 118-125, 2008.
- PAPWORTH, S. K. et al. Quantifying the role of *online* news in linking conservation research to Facebook and Twitter. *Conservation Biology*, v. 29, n. 3, p. 825-833, 2015.
- PEREIRA, K. F. et al. Urban Sloths: Public Knowledge, Opinions, and Interactions. *Animals (2076-2615)*, v. 8, n. 6, 2018.
- POOLA, M. et al. Sloths in the city: unexpectedly high density of pale-throated three-toed sloths (*Bradypus tridactylus*) found in an urban forest patch in Paramaribo, Suriname.
- POORTHUIS, A. et al. 16 Using Geotagged Digital Social Data in Geographic Research. *Key methods in geography*, p. 248, 2016.
- R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Austria, Vienna, 2017.
- RILEY, S. J. et al. The essence of wildlife management. *Wildlife Society Bulletin*, p. 585-593, 2002.
- REZENDE, I. M. et al. Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. *PLoS neglected tropical diseases*, v. 12, n. 6, p. e0006538-e0006538, 2018.
- RUAS, P. H. B. et al. Identification and characterisation of facebook user profiles considering interaction aspects. *Behaviour & Information Technology*, v. 38, n. 8, p. 858-872, 2019.
- RYLANDS, A. B.; BRANDON, K. Brazilian protected areas. *Conservation biology*, v. 19, n. 3, p. 612-618, 2005.

- SAJID, S. I. Social media and its role in marketing. *Business and Economics Journal*, v.7, 203-208, 2016.
- SCHÄFER, S. Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions. *Computers in Human Behavior*, v. 103, p. 1-12, 2020.
- SHEPARD, G. Old and in the way: jaguar transformation in Matisgenka. In: IX Sesquiannual Conference of the Society for the Anthropology of Lowland South America (SALSA) Gothenburg, Sweden, June. 2014. p. 26-29.
- SHERREN, K. et al. Digital archives, *Big Data* and image-based culturomics for social impact assessment: Opportunities and challenges. *Environmental Impact Assessment Review*, v. 67, p. 23-30, 2017.
- SIMET. Mapa de velocidade e acesso de internet. 2019. Disponível em:<<https://simet-publico.ceptro.br/mapas7/> SIMET MAPA>. Acessado em: 10 de Outubro de 2019.
- SORIANO-REDONDO, A. et al. Internet-based monitoring of public perception of conservation. *Biological conservation*, v. 206, p. 304-309, 2017.
- SPALDING, M. et al. Mapping the global value and distribution of coral reef tourism. *Marine Policy*, v. 82, p. 104-113, 2017.
- SPAROVEK, G. et al. Brazilian agriculture and environmental legislation: status and future challenges. 2010.
- STEPHEN, A. T.; SCIANDRA, M.; INMAN, J. Is it what you say or how you say it? How content characteristics affect consumer engagement with brands on Facebook. *How Content Characteristics Affect Consumer Engagement with Brands on Facebook* (October 1, 2015). Saïd Business School WP, v. 19, 2015.
- STIEGLITZ, S. et al. Social media analytics. *Business & Information Systems Engineering*, v. 6, n. 2, p. 89-96, 2014.
- STIEGLITZ, S.; DANG-XUAN, L. Social media and political communication: a social media analytics framework. *Social network analysis and mining*, v. 3, n. 4, p. 1277-1291, 2013.
- TABARELLI, M. et al. Challenges and opportunities for biodiversity conservation in the Brazilian Atlantic Forest. *Conservation Biology*, v. 19, n. 3, p. 695-700, 2005.

- THACKERAY, R. et al. Adoption and use of social media among public health departments. *BMC public health*, v. 12, n. 1, p. 242, 2012.
- TING, H. et al. Beliefs about the use of Instagram: An exploratory study. *International Journal of business and innovation*, v. 2, n. 2, p. 15-31, 2015.
- TOIVONEN, Tuuli et al. Social media data for conservation science: A methodological overview. *Biological Conservation*, v. 233, p. 298-315, 2019.
- TWITTER. API Information. Disponível em: < <https://developer.twitter.com/en/docs> > . Acessado em: 10 de Dezembro de 2019.
- UYANIK, G. K.; GÜLER, N. A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, v. 106, n. 1, p. 234-240, 2013.
- WE ARE SOCIAL; HOOTSUITE. Digital 2019 essential insights into how people around the world use the internet, mobile devices, social media, and e-commerce. Disponível em: <<https://wearesocial.com/global-digital-report-2019>>. Acesso em: 20 dez. 2018
- WE ARE SOCIAL; HOOTSUITE. Digital Report 2019. Disponível em: <<https://hootsuite.com/pages/digital-in-2019>>. Acessado em: 10 de Outubro de 2019.
- WEBSTER, J. G. User information regimes: How social media shape patterns of consumption. *Nw. UL Rev.*, v. 104, p. 593, 2010.
- WHITING, A.; WILLIAMS, D. Why people use social media: a uses and gratifications approach. *Qualitative Market Research: An International Journal*, 2013.
- WINGFIELD, N.; ISAAC, M.; BENNER, K. Google and Facebook take aim at fake news sites. *The New York Times*, v. 11, p. 12, 2016.
- WOOD, S. A. et al. Using social media to quantify nature-based tourism and recreation. *Scientific reports*, v. 3, p. 2976, 2013.
- ZHANG, C. et al. Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*, v. 49, p. 190-207, 2019.
- ZHANG, Harry. The optimality of naive Bayes. *AA*, v. 1, n. 2, p. 3, 2004.