

Fontes de dados gerados por usuários: quais plataformas considerar?

Gustavo R. T. Almeida¹, Isabelle Guimarães¹
Antonio F. L. Jacob Jr.², Fábio M. F. Lobato^{1,2}

¹Instituto de Engenharia e Geociências –
Universidade Federal do Oeste do Pará (UFOPA) – Santarém – BR – Brasil

²Departamento de Engenharia da Computação –
Universidade Estadual do Maranhão (UEMA) – São Luís – MA – Brasil

gr.almeida00@gmail.com, fabio.lobato@ufopa.edu.br

Resumo. *É inegável a pervasividade das mídias sociais em nosso cotidiano, tanto que o fácil acesso e a sensação de anonimato influenciou usuários a compartilharem ativamente suas opiniões. Como consequência, o crescimento exponencial do volume de dados gerou novos desafios em relação a seleção de estudos de caso e formas de coleta dos dados. Neste contexto, uma classificação das mídias sociais para facilitar o reconhecimento de suas estruturas e fluxos de dados pode otimizar processos de seleção e coleta de dados. No presente estudo é conduzido um levantamento das mídias sociais online mais relevantes e que disponibilizam conteúdo gerado pelos usuários. Também, apresentam-se os mecanismos de acesso de dados destas mídias, suas características e funcionalidades identificadas. Como impacto direto do estudo destaca-se a perspectiva mais ampla na busca de fontes de dados, de modo a reduzir o viés de pesquisa e permitir uma maior diversidade nos estudos de fenômenos socioeconômicos.*

Palavras-chave. *Mídias Sociais; Dados gerados por usuários; Levantamento exploratório.*

Abstract. *The pervasiveness of social media in our daily lives is unquestionable, so much so that the easier access and their inherent anonymity influence its users to share their opinions actively. As a consequence, the exponential growth of the data volume generated challenges for researchers as for the selection of case studies and data collection methods. In this context, a classification of social media to easily identify their data structure and flow can enhance data selection and collection processes. This study presents a exploratory survey of the most relevant social media platforms that provide user-generated content. The data access mechanisms, characteristics and features of these media are also discussed. As a direct impact of the study, a broader perspective in the search for data sources stands out, reducing the research bias and allowing a greater diversity in the study of socioeconomic phenomena.*

Keywords. *Social media; User-generated content; Exploratory survey.*

1. Introdução

A explosão além das expectativas das mídias sociais é inegável [Stieglitz et al. 2018]. Mesmo com a grande quantidade de novas interações e a contínua geração de dados,

o crescimento das mídias sociais facilita a condução de análises das mais diversas, de assuntos relacionados a economia, política à movimentos sociais [Lobato et al. 2017, Vaganov et al. 2018]. Ao mesmo passo, afeta a forma como informações são transmitidas e consumidas [Perrin 2015].

De forma errônea, é comum a utilização dos termos Mídias Sociais e Redes Sociais para determinar apenas um conceito. [Ellison and Boyd 2013] descreve as Redes Sociais como um conjunto de indivíduos que comungam de um mesmo espaço e possuem interesses semelhantes, promovendo a troca de informações sobre os mais diversos assuntos. Já as Mídias Sociais são os veículos disponibilizados de forma eletrônica que dão suporte ou geram redes sociais utilizando conhecimento informacional e métodos de comunicação. Logo, plataformas como *Facebook* e *Twitter* são **mídias sociais** que incorporam seus usuários em diversas **redes sociais**.

Na área de negócios, muitas empresas transformam as mídias sociais em ferramentas de comunicação [Lobato et al. 2017, Lee 2018, Sousa et al. 2019, Barata et al. 2018]. Um exemplo dessa transformação são as plataformas de reclamações *online* (e.g. ReclameAqui). Elas são facilitadoras do movimento chamado de boca-a-boca virtual (*Electronic Word-of-Mouth*), que é a adaptação do boca-a-boca tradicional ao cenário da comunicação *online* [Almeida et al. 2017, Sousa et al. 2020].

Em um contexto geral, a forma mais importante de dados para pesquisas encontradas em mídias sociais são conhecidas por *User-generated Content* (UGC), que se refere as informações que um usuário produz no momento que expõe suas opiniões ou questiona as opiniões de terceiros na web [Smith et al. 2012]. Compreender os problemas e desejos dos clientes e incorporá-los nos processos de compra e venda de uma empresa é o objetivo fundamental do Social CRM [Cheng and Shiu 2019], uma metodologia que facilita e otimiza as interações cliente-empresa por meio das mídias sociais. Ademais, [Barreto 2014] afirma que o aspecto eletrônico destes dados permite: i) Acesso posterior a sua criação para análise ou leitura; ii) Assumir diversos formatos digitais (Vídeo, imagem, som e texto); iii) Usuários podem manter-se anônimos de suas opiniões. Estes detalhes se integram e auxiliam o processo do Social CRM na identificação, utilização e proteção dos dados dos usuários [Lobato et al. 2017].

Mesmo com a diversidade de mídias sociais, a preferência dos usuários se concentrava nas mídias sociais mais famosas, com destaque ao *Facebook* que mantinha 79% do público adulto em sua plataforma no ano de 2016 [Greenwood et al. 2016] e no ano de 2019 mantêm 69% do público mesmo com a chegada de novas mídias [Perrin and Anderson 2020]. Isto se reflete na quantidade de trabalhos na literatura que estudavam apenas esta plataforma. Porém, novas pesquisas apontam uma mudança significativa na dinâmica das mídias sociais, e atualmente, jovens possuem um histórico mais variado no uso destas plataformas [Anderson and Jiang 2018].

Esta alteração na dinâmica social online pode diversificar as pesquisas sobre UGC, tendo em vista que cada plataforma sempre tem um enfoque diferente em como permitem o compartilhamento de informações por parte dos usuários. Neste ponto encontramos uma lacuna na literatura, pois o grande enfoque das pesquisas em mídias mais populares fez com que estudos sobre mídias menos conhecidas não possuíssem referências suficientes para permitir boas descrições de suas metodologias. Além disso, a falta de pesquisas

reduz a replicabilidade de trabalhos nesse contexto.

Com intuito de verificar estas lacunas realizou-se um levantamento exploratório de ambientes virtuais. Para este fim, os objetivos traçados para este trabalho foram: i) Identificar quais são os métodos de acesso aos UGC das mídias sociais levantadas; ii) classificá-las de acordo com a sua finalidade; e iii) identificar suas funcionalidades. O objetivo principal deste levantamento é elencar as plataformas em volume de acesso e que hospedam UGC que potencialmente possam ser usados em pesquisas envolvendo mídias sociais ou de comportamento do usuário, além de classificá-las de acordo com seus tipos de interação. Estes levantamentos auxiliam pesquisadores de diversas áreas na tomada de decisões sobre onde e como buscar os seus dados a serem estudados.

O restante do artigo está organizado como segue. Na Seção 2 alguns trabalhos relacionados em análises de dados sociais são discutidos. Na Seção 3 apresenta-se referencial teórico do trabalho. Na Seção 4 são apresentadas as metodologias de levantamento de mídias sociais. Nas Seções 5 e 6 os resultados dos levantamentos são descritos e discutidos, respectivamente. Por fim, na Seção 7 são feitas as considerações finais da pesquisa.

2. Trabalhos Relacionados

A análise de dados de mídias sociais é uma investigação recorrente na literatura. A densa quantidade de dados gerados por usuários, estes carregados de sentimentalismo e opiniões das mais diversas, são material essencial para análises precisas de comportamento humano, tanto social como econômico [Chen et al. 2014].

Análises sociais são as mais comuns devido ao maior número de mídias sociais que pertencem a este nicho (e.g. plataformas de seguidores, troca de mensagens, postagens, etc.). No trabalho de [Rodrigues et al. 2019] comentários de uma página de notícias são analisados com intuito de identificar a presença do posicionamento machista frente a um evento que foi consequência direta desta forma de discriminação. Em um contexto mais abrangente, os autores de [Pelle and Moreira 2017] investigam o discurso de ódio ao anotar um *dataset* próprio de comentários e gerar uma análise classificatória sobre este conteúdo. Nestes trabalhos percebe-se como os dados sociais possibilitam o estudo das interações humanas e problemas sociais importantes. Porém, mesmo análises sem vertentes sociais podem ser alcançadas através da análise destes dados, como identificado a seguir.

Os autores de [Almeida et al. 2017] investigam reclamações de usuários à empresas com foco em identificar os problemas comuns de clientes frente a sua experiência de compra de um produto ou serviço, assim como, o atendimento prestado pelas empresas. Ainda no contexto de cliente-empresa, [Vaganov et al. 2018] investiga padrões de consumo para gerar um modelo de predição de categorias de compra de cada consumidor. Como observado, mesmo análises sem contexto social oferecem resultados de mesmo impacto. Ainda fora do contexto social e econômico, os dados de mídias sociais podem ser utilizados para entender o funcionamento das plataformas e como funciona o fluxo de usuários entre elas.

Em um contexto mais específico, [Marcolin et al. 2019] apresentaram uma análise baseada em coleta de dados disponíveis na plataforma *TripAdvisor*. O *dataset* utilizado foi constituído de 23.229 comentários no período de 2011 a 2016 em perfis de hotéis localizados em Porto Alegre (RS). Os autores utilizaram do LSA (*Latent Semantic Analysis*)

para extrair os tópicos mais comuns nos dados coletados. Como resultado, obteve-se um conjunto de tópicos mais recorrentes nos relatos dos visitantes, o que auxilia os hoteleiros da região na tomada de decisões sobre seus negócios.

Mesmo compreendendo o potencial que estas análises possuem os autores de [Sivarajah et al. 2017] afirmam que o verdadeiro desafio nestas análises está na definição de suas metodologias de coleta e análise, em destaque aos métodos de coleta de dados que são pouco explorados nos trabalhos, sendo ofuscados pelas descrições mais aprofundadas de métodos de análise. Portanto, ao invés de focar nas etapas de análise, neste artigo utilizou-se uma perspectiva focada em otimizar as etapas de seleção e coleta de dados em pesquisas. Este foco permite uma análise sistemática das mídias sociais, a fonte mais comum de dados para análises. Essa análise é importante para promover uma base para novas pesquisas, tanto em como selecionam os dados aos quais desejam analisar, assim como, o método de coleta a ser utilizado.

3. Referencial Teórico

Nesta seção são apresentados conceitos sobre mídias sociais relacionados a suas características de funcionamento mais predominantes. Tais conceitos servirão para mapear as mídias em categorias distintas e possibilitar a discussão de quais dados podem ser mais relevantes para pesquisas científicas.

Os autores de [Tabassum et al. 2018] descrevem as redes de preferência similar como aquelas nas quais os usuários constroem relações com base na interseção de assuntos de natureza diversa. Neste mesmo contexto, descrevem as redes usuário-usuário como as que promovem o contato direto entre os indivíduos, sem a necessidade de interseções. Já as redes de co-autoria são caracterizadas pela possibilidade de criação colaborativa de conteúdo entre os usuários, como autores de artigos científicos [Barabã et al. 2002], repórteres de notícias em jornais [Park and Leydesdorff 2013] e autores de histórias originais, assim como membros comuns de uma mídia social *online*, a qual permite menções entre os usuários. Outro tipo de rede são as de interação, as quais são responsáveis por promover a comunicação entre usuários, seja por mensagens diretas ou por comentários e respostas em publicações ou fóruns específicos [Recuero 2007].

No trabalho de [Recuero and Zago 2009] é definido que redes de seguidores são identificadas quando existe a possibilidade de acompanhar o conteúdo compartilhado por uma determinada pessoa ou empresa sem necessariamente possibilitar a interação direta entre os usuários, como por exemplo a plataforma *LinkedIn*. Redes de estudantes são destinadas a alunos tanto de nível escolar quanto acadêmico para compartilhamento de estudos, videoaulas ou reforço [Basso et al. 2013]. Redes de amizade, onde é possível conversar diretamente entre os usuários com a finalidade de construção de laços afetivos ou *networking*. Redes de co-atores, que podem ser destinadas a atores de cinema, desenvolvedores e criadores de conteúdo e redes de viralização¹, onde conteúdos são criados com a finalidade de atingir o maior público possível.

Em mídias sociais *online* existem funcionalidades internas que podem prover dados específicos. O Perfil de Usuário, por exemplo, é onde encontram-se as

¹O termo Viralização remete a quando um conteúdo compartilhado atinge proporções muito grandes, de maneira a atingir a maior parte de uma população.

informações pessoais sobre os usuários cadastrados na plataforma, além de seus comportamentos sociais e preferência de conteúdo [Pereira et al. 2014]. Os autores de [Benevenuto et al. 2011] apresentam a função de Postagens, também conhecidas como Atualizações, que são compartilhamentos de conteúdo imediatamente visíveis na página do usuário. Estes mesmo autores caracterizam as Interações como a avaliação de um conteúdo por parte dos usuários que o visualizam, sendo estas importantes na identificação de conteúdo relevante e auxiliam na construção do perfil do usuário. Com relação aos comentários/respostas tratam-se de interações diretas com o conteúdo compartilhado por um usuário [Benevenuto et al. 2011].

Outras funcionalidades que tem destaque são as mensagens diretas, que estão relacionadas as troca de mensagens privadas entre um ou mais usuários, a depender das limitações da mídia social. Co-autoria é a construção colaborativa de conteúdo, como possibilidade de citações e marcações em postagens e/ou comentários [Maia and Caregnato 2008]. Multimídia se refere ao compartilhamento de conteúdo por meio de fotos, vídeos e/ou áudios por meio das plataformas *online* [Luz and Woitowicz 2017]. Login para acesso é quando a mídia social requer acesso em conta/cadastro para visualização dos conteúdos postados [Alexopoulos et al. 2014]. Na mesma linha, Login para interação requer o acesso em conta cadastrada para interação e/ou comentário em conteúdo.

4. Metodologia

Inicialmente, conduziu-se um levantamento de mídias sociais mais populares e seus mecanismos de acesso aos dados implementados, com intuito de identificar possíveis alvos para coleta e análise de dados nas plataformas atuais. A coleta de mídias sociais foi procedida de forma manual por meio de uma lista disponível na plataforma *Wikipedia*² e por uma pesquisa exploratória na web. O critério de escolha das mídias a serem apresentadas foi a sua relevância no contexto geral e no seu nicho de funcionamento.

As mídias selecionadas foram mapeadas de acordo com as formas de disponibilização dos dados. Esta verificação e mapeamento foram conduzidas baseadas nas informações contidas no repositório online *Programmable Web*³, onde verificava-se a disponibilidade de *Application Programming Interface* (API) de coleta de dados oficiais de cada mídia social em questão. No caso negativo, verificava-se a existência de uma forma alternativa de acesso aos dados, como *datasets* disponíveis publicamente em repositórios *online* difundidos entre a comunidade científica, como o *Kaggle*⁴. Todos estes repositórios online foram escolhidos especificamente por serem atualizados constantemente e permitirem e incentivarem a submissão de novos conteúdos, continuamente enriquecendo as plataformas.

Também, foi realizado um levantamento de funcionalidades disponíveis nas redes sociais utilizadas neste estudo. As funções levadas em consideração foram: existência de um perfil de usuário com informações pessoais (e.g. nome, idade, cidade, etc.) e/ou conteúdo compartilhado na plataforma; possibilidade de fazer postagens textuais com ou sem delimitação de caracteres; possibilidade de interagir com as postagens por meio de

²https://en.wikipedia.org/wiki/List_of_social_networking_websites

³<https://www.programmableweb.com>

⁴<https://www.kaggle.com/>

reações (e.g. *like*, *amei*, *uau*, etc.); permissão de comentários e/ou respostas nas postagens de terceiros; permitir a troca de mensagens direta entre os usuários por meio de *chats* privados; construção colaborativa de conteúdo por meio de citações e/ou marcações em postagens; permissão de postagens multimídia como áudios e/ou vídeos; necessidade de *login* para ter acesso ao conteúdo integral de uma postagem; e necessidade de *login* para interação com o conteúdo por meio de comentários e/ou reações. Este levantamento foi feito manualmente em cada plataforma.

Além disso, foi conduzido um mapeamento das mídias sociais estudadas a partir das atividades internas presentes em cada mídia e da forma de interação dos usuários. As atividades internas foram baseadas nos conceitos apresentados na Seção 3, além de uma observação empírica feita nas próprias plataformas. As principais características foram: formas de interação (e.g. mensagens diretas, comentários, respostas, etc.), disponibilização de conteúdo (e.g. aberto, por solicitação, etc.) e acompanhamento (e.g. seguidores, curtidores, inscritos, filiados, etc.).

5. Resultados

A partir da condução manual do levantamento foram selecionadas 58 mídias sociais. Dessas, apenas 26 possuem API's oficiais. As outras ou possuem acesso direto aos dados ou não disponibilizam as informações. Todas as plataformas dispostas na Figura 1 armazenam e permitem a criação e visualização de UGC em suas páginas. Como exemplo, mídias sociais *online*, sistemas de reputação, *reviews* ou reclamações, assim como, plataformas de notícias, pertencem a este nicho.

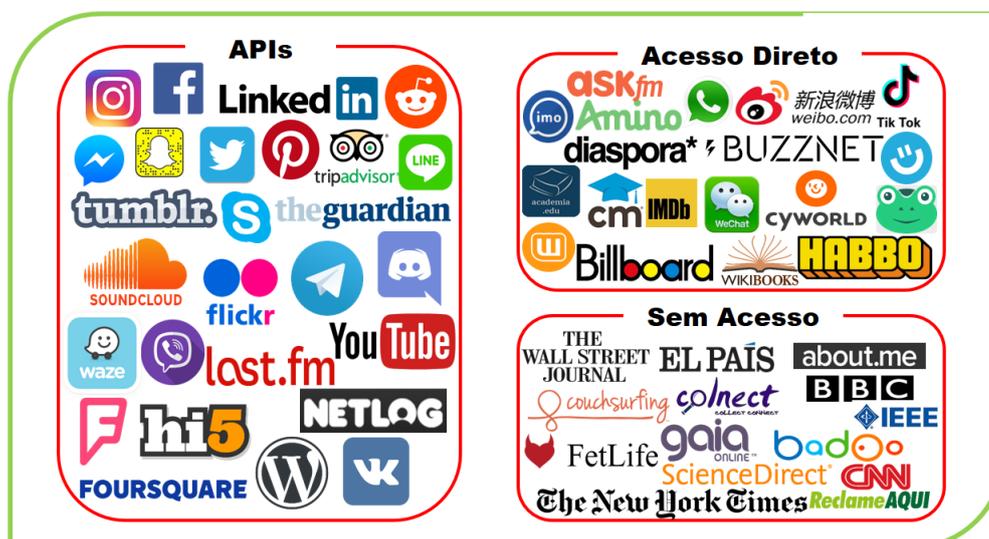


Figura 1. Mídias sociais e seus respectivos acesso aos dados.

A Tabela 1 apresenta o levantamento de funcionalidades das dez mídias sociais mais acessadas no mundo, segundo o ranqueamento *Alexa*⁵. O levantamento completo pode ser visualizado na Tabela A no apêndice deste artigo. É importante ressaltar que plataformas como *El Pais*, *NYT* e *IEEE* (Plataformas de notícias e acadêmicas, respectivamente) são consideradas mídias sociais, pois permitem a interação entre seus usuários.

⁵<http://www.alexa.com>

Tabela 1. Plataformas e suas funcionalidades.

	Perfil de usuário	Postagens	Interação	Comentário / resposta	Mensagem direta	Co-autoria	Multimídia	Login para acesso	Login para interação
Youtube	✓	✓	✓	✓	×	✓	✓	×	✓
Facebook	✓	✓	✓	✓	✓	×	✓	×	✓
Weibo	✓	✓	✓	✓	✓	×	✓	×	✓
Reddit	✓	✓	✓	✓	✓	×	✓	×	✓
VK	✓	✓	✓	✓	✓	×	✓	×	×
Instagram	✓	✓	✓	✓	✓	×	✓	×	✓
Twitter	✓	✓	✓	✓	✓	×	✓	×	✓
WhatsApp	✓	✓	×	✓	✓	✓	✓	✓	✓
IMDb	×	✓	✓	×	×	×	✓	×	✓
WordPress	×	✓	✓	✓	×	✓	✓	✓	✓

Em grande parte destas mídias os dados não são da mesma natureza, visto que dados textuais e gráficos, assim como os metadados, ficam dispostos nos mesmos ambientes [Stieglitz and Dang-Xuan 2013]. Entretanto, observa-se que a maioria dos trabalhos focam no conteúdo textual das mídias sociais [He et al. 2013, Lobato et al. 2017, Gavilanes et al. 2018], isto facilita a replicabilidade de pesquisas no quesito de método de coleta utilizado. Em relação a classificação das mídias de acordo com suas funcionalidades apresentada na Tabela 2, as principais categorias identificadas foram: redes de preferência similar, redes de co-autoria e redes de interação. As plataformas podem se encaixar em mais de uma categoria, portanto apenas a mais expressiva é apresentada.

Tabela 2. Caracterização das Mídias Sociais

Característica Predominante	Mídia Social
Redes de Preferência Similar	Reddit; Soundcloud; Tumblr; Pinterest; Amino; Badoo. Flickr; Waze; Last.fm; Ask.fm; Couchsurfing; Fetlife;
Redes de Co-autoria	LinkedIn; BBC; The New York Times; The Guardian; Science Direct; Wattpad; Buzznet; The Wall Street Journal; É! País; CNN; IEEE.
Redes de Interação	Weibo; Discord; Skype; Hi5; Cyworld; Haboo; TripAdvisor; ReclameAqui.
Redes Usuário-usuário	WhatsApp; Telegram; WordPress; Messenger; Line; Viber; WeChat.
Redes de Amizade	Facebook; VK; Netlog; Diaspora; Gab; Imo; About.me.
Redes de Viralização	Youtube; Twitter; TikTok; Foursquare.
Redes de Seguidores	Instagram; SnapChat; Colnect.
Redes de Co-atores	IMDb; Gaia; Billboard.
Redes de Estudantes	Academia.edu; Wikibooks; Classmates.

A Tabela 2 auxilia pesquisadores em prontamente identificarem qual tipo de característica desejam analisar dentre as mídias sociais. O discernimento e seleção destas mídias pode facilitar a etapa de coleta de dados, visto que cada mídia é formada por diferentes estruturas e fluxo de dados. Nesse contexto, um mesmo método de coleta pode tornar-se inviável em diferentes plataformas.

6. Discussões

Os resultados encontrados através do primeiro objetivo desta pesquisa, principalmente demonstrados pela Figura 1, mostram uma grande variedade de fontes de dados possíveis para pesquisas. Porém, como pesquisas em grande parte focam seus estudos de casos em

mídias sociais mais conhecidas, a lacuna identificada nesta pesquisa, muitos dados são pouco utilizados. Neste contexto, tais resultados facilitam a escolha de estudos de caso, ao mapear as mídias por método de coleta necessário a ser implementado, incentivando novas pesquisas a diversificarem a análise de dados *online*.

Em relação ao mapeamento por métodos de coleta, ao longo da pesquisa, não foi encontrado, a nível acadêmico, uma ilustração das formas de acesso aos dados de mídias *online*, e quais destas necessitam da aplicação de mecanismos de coleta de dados. Nesta etapa a escolha da fonte ou do tipo de dado a ser coletado deve ser criteriosa, tendo em vista como o tempo de execução desta tarefa pode estender-se em meses ou até anos para a produção de um conjunto de dados de qualidade.

Em decorrência da classificação das mídias sociais *online* de acordo com sua finalidade, prevista no segundo objetivo do trabalho, e através da Tabela 2 observou-se que as redes de preferência similar e redes de co-autoria são as de características mais predominantes nas mídias sociais estudadas. Isso corrobora com o fato de que as mídias sociais são principalmente formadas por indivíduos que compartilham de interesses semelhantes. Tal caracterização também facilita identificar qual mídia poderá fornecer um melhor estudo de caso para sua pesquisa, além de facilitar a definição do método de coleta a ser utilizado, tendo em vista que cada tipo de mídia possui um fluxo de dados distinto.

O terceiro objetivo da pesquisa tratou da identificação das funcionalidades das mídias sociais. Por meio da Tabela 1 constatou-se que das 58 mídias, apenas 5 não apresentam uma página de perfil de usuário e 29 plataformas não permitem a visualização dos conteúdos postados pelos usuários que não possuem um cadastro na mídia social em questão. Com isso, infere-se que é necessário que o usuário compartilhe suas informações e delimite suas preferências a fim de construir uma experiência mais acurada durante a utilização das plataformas *online*. Tal resultado reitera a importância das redes de preferência similar no cenário online atual.

7. Conclusão

Atualmente, não somente as mídias sociais online permitem que os usuários compartilhem comentários e opiniões e interajam com o conteúdo gerado por outros usuários, outras plataformas também incorporaram aspectos de mídias sociais. Esta profusão de plataformas abre um leque de possibilidades para pesquisas envolvendo mineração de opinião. Apesar disso, a maior parte dos estudos na área focam em redes bem consolidadas como *Twitter* e *Facebook*, o que pode representar um viés de pesquisa em potencial.

Considerando o valor presente nos dados gerados pelos usuários, aliado à lacuna existente na literatura quanto a um levantamento das mídias sociais, o presente trabalho apresentou um levantamento sistemático de plataformas que abrigam dados gerados pelos usuários considerando a sua relevância. Estas plataformas também foram categorizadas em relação aos mecanismos de acesso. Além de apresentar uma classificação de acordo com as finalidades de cada mídia social, o trabalho mostrou que existe uma tendência em relação à utilização das mídias sociais para a facilitar a interação entre os usuários. Este fato corrobora com [Ashley and Tuten 2015] que diz que o conteúdo gerado pelo usuário é uma das maiores fontes de informação disponíveis na atualidade.

Também foi apresentada uma tabela de classificação das mídias sociais estudadas de acordo com suas finalidades. Foram detectados nove grupos de maior relevância: re-

des de preferência similar, redes de co-autoria, redes de interação, redes usuário-usuário, redes de amizade, redes de viralização, redes de seguidores, redes de co-atores e redes de estudantes.

Como contribuição técnico-científica do trabalho, destaca-se que os resultados obtidos por meio do levantamento conduzido servem de guia para novas pesquisas na seleção de potenciais fonte de dados, diminuindo assim vieses na pesquisa. Como trabalhos futuros pretende-se realizar um levantamento complementar das bases de dados publicamente disponíveis em repositórios como o *Kaggle* e *GitHub* e expandir a análise quanto aos mecanismos para coleta de dados. Pretende-se também analisar um conjunto maior de Mídias Sociais e classificá-las de forma mais acurada.

8. Agradecimentos

Os autores agradecem a Universidade Federal do Oeste do Pará (UFOPA) pelo suporte e concessão de bolsa de extensão universitária. E igualmente agradecem ao Serviço Alemão de Intercâmbio Acadêmico (DAAD - *Alemanha*) e FAPEMA pelo suporte nesta pesquisa.

Referências

- Alexopoulos, C., Loukis, E., and Charalabidis, Y. (2014). A Platform for Closing the Open Data Feedback Loop Based on Web2.0 Functionality. *JeDEM - eJournal of eDemocracy and Open Government*, 6(1):62–68.
- Almeida, G. R. d., Cirqueira, D. R., and Lobato, F. M. (2017). Improving social crm through eletronic word-of-mouth: a case study of reclameaqui. *XIV Workshop de Trabalhos de Iniciação Científica*.
- Anderson, M. and Jiang, J. (2018). Teens, social media & technology 2018. *Pew Research Center*, 31:2018.
- Ashley, C. and Tuten, T. (2015). Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement. *Psychology & Marketing*, 32(May 2014):15–27.
- Barabã, A. L., Barabãsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3):590–614.
- Barata, G. M., Viana, J. A., Reinhold, O., Lobato, F., and Alt, R. (2018). Social CRM in digital marketing agencies: an extensive classification of services. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 750–753. IEEE, IEEE.
- Barreto, A. M. (2014). The word-of-mouth phenomenon in the social media era. *International Journal of Market Research*, 56(5):631–654.
- Basso, M. V. d. A., Bona, A. S., Pescador, C. M., Koehler, C., and Fagundes, L. d. C. (2013). Redes sociais: espaço de aprendizagem digital cooperativo//social networks: collaborative digital learning space. *Conjectura: filosofia e educação*, 18(1):135–149.
- Benevenuto, F., Almeida, J. M., and Silva, A. S. (2011). Explorando Redes Sociais Online: Da Coleta e Análise de Grandes Bases de Dados às Aplicações. *Mini-cursos do Simpósio Brasileiro de Redes de Computadores (SBRC)*, pages 63–101.

- Chen, M., Mao, S., and Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209.
- Cheng, C. C. and Shiu, E. C. (2019). How to enhance smes customer involvement using social media: The role of social crm. *International Small Business Journal*, 37(1):22–42.
- Ellison, N. B. and Boyd, D. M. (2013). *Sociality Through Social Network Sites*, volume 1. Oxford University Press.
- Gavilanes, J. M., Flatten, T. C., and Brettel, M. (2018). Content Strategies for Digital Consumer Engagement in Social Networks: Why Advertising Is an Antecedent of Engagement. *Journal of Advertising*, 47(1):4–23.
- Greenwood, S., Perrin, A., and Duggan, M. (2016). Social media update 2016. *Pew Research Center*, 11(2).
- He, W., Zha, S., and Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3):464–472.
- Lee, I. (2018). Social media analytics for enterprises: Typology, methods, and processes. *Business Horizons*, 61(2):199 – 210.
- Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2017). Social crm: Biggest challenges to make it work in the real world. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 221–232. Springer International Publishing, Cham.
- Luz, A. L. L. and Woitowicz, K. J. (2017). Jornalismo nas Redes Sociais: Instagram e Facebook como Plataformas para Produção de Conteúdo Multimídia pelo Sítio Cultura Plural. *Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação*, pages 1–14.
- Maia, M. d. F. S. and Caregnato, S. E. (2008). Co-autoria como indicador de redes de colaboração científica. *Perspectiva em Ciência da Informação*, 13(2):18–31.
- Marcolin, C., Becker, J. L., Wild, F., Schiavi, G., and Behr, A. (2019). Business analytics in tourism: Uncovering knowledge from crowds. *BAR - Brazilian Administration Review*, 16(2).
- Park, H. W. and Leydesdorff, L. (2013). Decomposing social and semantic networks in emerging "big data" research. *Journal of Informetrics*, 7(3):756–765.
- Pelle, R. P. d. and Moreira, V. P. (2017). Offensive comments in the brazilian web: a dataset and baseline results. In *6º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2017)*, volume 6, Porto Alegre, RS, Brasil. SBC.
- Pereira, C. K., Campos, F., Ströele, V., David, J. M. N., and Braga, R. (2014). Extração de Características de Perfil e de Contexto em Redes Sociais para Recomendação de Recursos Educacionais. In *Anais do XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014)*, volume 1, page 506.
- Perrin, A. (2015). Social media usage: 2005-2015. *Pew Internet and American Life Project*.

- Perrin, A. and Anderson, M. (2019 (Acessado dia 23 de Junho de 2020)). *Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018*. <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>.
- Recuero, R. (2007). Considerações sobre a difusão de informações em redes sociais na internet. *CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO DA REGIÃO SUL*, 8., 2007, *Passo Fundo*, pages 1–16.
- Recuero, R. and Zago, G. (2009). Em busca das "redes que importam": redes sociais e capital social no Twitter. *Líbero*, 12(24):81–94.
- Rodrigues, L. D., da Silva Júnior, J. L., and Lobato, F. M. (2019). A culpa é dela! e isso o que dizem nos comentários das notícias sobre a tentativa de feminicídio de elaine caparroz. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 47–58. SBC.
- Sivarajah, U., Kamal, M. M., Irani, Z., and Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70:263 – 286.
- Smith, A. N., Fischer, E., and Yongjian, C. (2012). How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2):102–113.
- Sousa, G. N., Almeida, G. R., and Lobato, F. (2019). Social network advertising classification based on content categories. In Abramowicz, W. and Corchuelo, R., editors, *Business Information Systems Workshops*, pages 396–404, Cham. Springer International Publishing.
- Sousa, G. N. D., DA SILVA GUIMARÃES, I., Augusto, J., Viana, N., Reinhold, O., Fernando, A., Jacob, L., and MANOEL FRANÇA LOBATO, F. (2020). Análise do setor de telecomunicação brasileiro : Uma visão sobre Reclamações. *RISTI -Revista Ibérica de Sistemas e Tecnologias de Informação*, pages 31–48.
- Stieglitz, S. and Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4):1277–1291.
- Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. (2018). Social media analytics – challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39:156 – 168.
- Tabassum, S., Pereira, F. S., Fernandes, S., and Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):1–21.
- Vaganov, D., Sheina, E., and Bochenina, K. (2018). A comparative study of social data similarity measures related to financial behavior. *Procedia Computer Science*, 136:274 – 283. 7th International Young Scientists Conference on Computational Science, YSC2018, 02-06 July 2018, Heraklion, Greece.

A. Apêndice - Plataformas e suas funcionalidades (Versão completa)

	Perfil de usuário	Postagem	Interação	Comentário / resposta	Mensagem direta	Co-autoria	Multimídia	Login para acesso	Login para interação
Facebook	✓	✓	✓	✓	✓	×	✓	×	✓
Instagram	✓	✓	✓	✓	✓	×	✓	×	✓
Twitter	✓	✓	✓	✓	✓	×	✓	×	✓
WhatsApp	✓	✓	×	✓	✓	✓	✓	✓	✓
SnapChat	✓	✓	✓	×	×	×	✓	✓	✓
TikTok	✓	✓	✓	✓	×	✓	✓	×	✓
Messenger	×	×	×	×	✓	✓	✓	✓	✓
Pinterest	✓	✓	✓	×	×	✓	✓	×	✓
Line	✓	×	×	×	✓	×	✓	✓	✓
The Guardian	✓	×	×	✓	×	✓	✓	×	✓
Flickr	✓	✓	✓	✓	×	×	✓	×	✓
Waze	✓	×	×	×	✓	✓	×	✓	✓
Youtube	✓	✓	✓	✓	×	✓	✓	×	✓
Netlog	✓	×	×	×	✓	×	×	✓	✓
Reddit	✓	✓	✓	✓	✓	×	✓	×	✓
Tumblr	✓	✓	✓	✓	✓	×	✓	×	×
Telegram	✓	✓	×	✓	✓	✓	✓	✓	✓
Viber	✓	✓	×	✓	✓	✓	✓	✓	✓
Foursquare	✓	×	×	✓	×	✓	×	✓	✓
Linkedin	✓	✓	✓	✓	✓	✓	✓	✓	✓
TripAdvisor	✓	✓	✓	✓	×	×	✓	×	✓
Skype	✓	×	×	✓	✓	✓	✓	✓	✓
Soundcloud	✓	✓	✓	✓	×	✓	✓	×	✓
Discord	✓	✓	×	✓	✓	✓	✓	✓	✓
Last.fm	✓	✓	✓	✓	✓	✓	✓	×	✓
Hi5	✓	✓	✓	✓	✓	×	×	✓	✓
WordPress	×	✓	✓	✓	×	✓	✓	✓	✓
VK	✓	✓	✓	✓	✓	×	✓	×	×
Ask.fm	✓	✓	✓	×	✓	×	×	✓	✓
Cyworld	✓	✓	×	×	✓	✓	×	✓	✓
IMDb	×	✓	✓	×	×	×	✓	×	✓
Haboo	✓	✓	×	×	✓	✓	×	✓	✓
Wattpad	✓	✓	✓	✓	✓	✓	×	✓	✓
Amino	✓	✓	✓	✓	✓	✓	✓	×	✓
Diaspora	✓	✓	✓	×	✓	✓	×	✓	✓
Academia.edu	✓	✓	×	×	✓	✓	×	✓	✓
WeChat	✓	✓	×	✓	✓	✓	✓	✓	✓
Wikibooks	×	✓	×	×	×	✓	✓	×	✓
Gab	✓	✓	✓	✓	✓	×	✓	✓	✓
Imo	✓	×	×	×	✓	×	✓	✓	✓
Weibo	✓	✓	✓	✓	✓	×	✓	×	✓
Buzznet	×	✓	✓	×	×	×	×	×	×
Class Mates	✓	✓	×	×	✓	×	×	✓	✓
Billboard	✓	×	✓	×	×	×	✓	×	✓
The WallStreet Journal	✓	✓	×	✓	×	×	✓	✓	✓
El Pais	✓	✓	×	✓	×	✓	✓	✓	✓
About.me	✓	✓	×	×	×	×	✓	×	×
Couchsurfing	✓	✓	✓	✓	✓	✓	✓	×	✓
Colnect	✓	✓	×	×	✓	×	×	×	✓
BBC	✓	×	✓	✓	×	×	✓	×	✓
Fetlife	✓	✓	✓	✓	✓	✓	×	✓	✓
Gaia	✓	✓	✓	✓	×	×	✓	×	✓
Badoo	✓	✓	✓	✓	✓	×	×	✓	✓
The New York Times	✓	×	✓	✓	×	×	✓	×	✓
ReclameAqui	✓	✓	✓	✓	×	×	×	×	✓
CNN	✓	×	✓	✓	×	×	✓	×	✓
Science Direct	✓	×	×	×	×	✓	×	×	×
IEEE	✓	×	×	×	×	✓	×	✓	×