Controvérsias na Wikipédia lusófona: pode o olhar CTS ser apoiado por ferramentas qualitativas?

Henrique Rabelo de Andrade
PESC/COPPE/UFRJ
Rio de Janeiro, Brazil
handrade@cos.ufrj.br

Resumo. *O presente estudo revisa esforços prévios de mapeamentos de controvérsias na Wikipédia e analisa o funcionamento da ferramenta Contropedia, explicitando suas funcionalidades e a utilizando para observar algumas controvérsias, a fim de entender como ferramentas quantitativas podem apoiar estudos de controvérsias que utilizem abordagens CTS.*

Abstract. *This The present study reviews previous efforts of mapping controversies on Wikipedia and analyzes the Contropedia tool, detailing its features and using it to observe some controversies in order to understand how quantitative tools can support controversies studies that use STS approaches.*

Introdução

O presente trabalho se propõe a estudar como são feitos os mapeamentos de controvérsias na Wikipédia, buscando entender o que significa uma controvérsia para distintos pesquisadores e como (ou se) esses indicadores podem ser utilizados na Wikipédia em português. Além disso, buscaremos compreender se esses esforços podem ser compatíveis com pesquisas que utilizem uma abordagem CTS (Ciência-Tecnologia-Sociedade).

Seguindo a proposição de Bruno Latour, "*nossa entrada no mundo da ciência e da tecnologia será pela porta de trás, a da ciência em construção, e não pela entrada mais grandiosa da ciência acabada*" (LATOUR, 1987 p. 6). Assim, nosso ponto de partida serão os estudos sobre controvérsias na Wikipédia, onde veremos como elas são definidas

e abordadas por diferentes pesquisadores, assim como a relação entre seus trabalhos.

O primeiro ponto de parada deste estudo é a tese de doutorado defendida em 2014 por Bernardo Esteves, "*As controvérsias da ciência na Wikipédia em português: o caso do aquecimento global*". Em um grande esforço para estudar as controvérsias sobre mudanças climáticas, Esteves (2014, p. 295) observa que "*na Wikipédia, cada ação dos editores deixa rastros disponíveis para consulta de pesquisadores e demais interessados, abrindo uma janela para um repositório riquíssimo de informações sobre como os usuários negociam suas versões de verdade*", e, para dar conta desses rastros, ele acompanha editores, regras editoriais, robôs, bloqueios, discussões e proteções de páginas, narrando processos de resolução de controvérsias e de estabilização de textos enciclopédicos.

Em sua pesquisa, Esteves fez tanto análises quantitativas como qualitativas, e em seus esforços de construir um olhar ao mesmo tempo amplo como o da águia e míope como o da formiga, chegou a cogitar a criação de um indicador de controvérsias que pudesse ser utilizado na Wikipédia em português. Compondo esse movimento, o autor apresentou esforços já feitos por outros pesquisadores na construção de indicadores, e seguiremos esse rastro na próxima seção para visitar esforços de mapeamento de controvérsias.

Indicadores de controvérsia na Wikipédia

Nesta seção vamos observar indicadores que fomentaram o trabalho de Esteves, e buscar entender o que cada um deles chama de controvérsia. Estarão buscando a mesma coisa ou será controverso o uso termo "controvérsia"?

Em um trabalho bastante utilizado por Esteves, "*Edit wars in Wikipedia*", Sumi et al (2011) propõe o índice M de controvérsia, observando "*como o número de edições e reversões desviam em algumas páginas dos números seguidos pela maioria dos artigos*" [1]. Mapeando duplas de editores que se revertem em um artigo e seu volume de edições no mesmo texto, o autor testa seu modelo em 6 diferentes idiomas e observa que menos de 1% dos artigos apresentam um índice alto.

1 Todas as citações desta seção foram traduzidas livremente pelo autor.

A avaliação anterior foi a base do estudo "*The most controversial topics in Wikipedia: A multilingual and geographical analysis*" de Yaseri, que utilizou o índice M como ponto de partida para observar versões em diferentes idiomas da enciclopédia buscando aproximações e similaridades entre grupos de idiomas próximos, testando "*um indicador multilíngue e independente da cultura*" (YASERI, 2014). Em sua pesquisa ele identifica que a maior parte das controvérsias são localizadas, e não se repetem tão comumente em outros idiomas (menos ainda se forem de grupos linguísticos diferentes), o que reforça a dificuldade já apontada por Esteves sobre critérios globais sendo utilizados para comunidades distintas.

Já Vuong et al. (2008), em "*On ranking controversies in wikipedia: models and evaluation*", resolvem levar em consideração não somente a atividade editorial em um artigo como também o perfil dos usuários envolvidos. Utilizando também a idade do artigo (em número de revisões salvas, e não em tempo corrido desde sua criação), o autor propõe diferentes índices partindo da seguinte premissa: "*um usuário é mais controverso se participa de disputas em artigos menos controversos e um artigo é mais controverso se nele participam de disputas usuários menos controversos*."

Brander e Lerner (2008), em "*Visual analysis of controversy in user-generated encyclopedias*", estão interessados em saber quem edita depois de quem, e qual a diferença de tempo entre essas edições. Com essas informações os autores criam uma visualização de usuários agrupados, simbolizando "facções" em disputas de edições. Após gerar seus mapas, os pesquisadores foram ler os artigos e estudar o perfil dos usuários envolvidos nas disputas, tanto na versão inglesa como na alemã da Wikipédia. Após observarem de perto a questão, concluíram que o modelo tende a aproximar vândalos e combatentes de vandalismo, pois "*ambos podem apresentar o comportamento 'um contra todos'*" (BRANDER, 2008). Pensando em melhorias de seu trabalho, propõem que sejam adicionados futuramente na equação registros de bloqueios dos usuários e conversas realizadas nas páginas dos usuários.

Em "*There is no deadline: time evolution of Wikipedia discussions*", Kaltenbrunner e Laniado (2012) se voltam para as páginas de discussão dos artigos, onde inicialmente não encontraram padrões de pico correlacionados com a página do domínio principal.

Aprofundando a inferência, criaram o indicador *h*, que aponta o maior número de discussões que seja igual ao número mínimo de respostas nelas. Observando o tempo que um artigo leva para incrementar seu *h* o estudo procura "*identificar escaladas ou estabilizações de controvérsias*".

Indicadores globais e abordagem CTS

Todos os estudos apresentados na seção anterior apresentam instrumentos de inferência que confiam nos metadados como fiéis porta-vozes dos artigos estudados, e pouco se ocupam de ler o conteúdo de fato escrito pelos usuários. Como conciliar tentativas de olhares generalizantes, como o voo de uma águia, com o olhar míope da formiga da abordagem CTS?

Para dar sequência a nossa jornada, lançaremos mão da Teoria Ator-Rede, utilizada para tratar dos enredamentos e estabilizações de fatos científicos. Sabemos que fatos científicos têm sua dinâmica de gênese própria, mas a construção de artigos na Wikipédia se aproxima bastante dela e com isso experiências acadêmicas de estudos da construção de textos científicos podem servir de base para estudos de escritas enciclopédicas. Como diz Esteves (2014 p. 102),

> "*a Teoria Ator-Rede propôs definir a factualidade de uma alegação não em termos de uma suposta veracidade intrínseca, mas nos termos de sua resistência aos ataques que ela venha a sofrer. A analogia com a Wikipédia é clara. Latour parecia descrever a verificabilidade nessa passagem de Ciência em ação: "E a que resiste [a realidade]? Aos testes de força. Se, numa dada situação, nenhum dissidente for capaz de modificar a forma de um novo objeto, então é isso, é realidade, ao menos enquanto os testes de força não forem modificados" (LATOUR, 1987, p. 93).""*.

Nesse caso, os especialistas e os laboratórios não estão imersos na controvérsia, mas os editores da enciclopédia assumem o papel de seus porta-vozes e a disputa, apesar de deslocada, segue dinâmicas parecidas.

Assim como o fazer da ciência, a escrita da Wikipédia também segue um "*caminho muito estranho porque é invisível quando tudo vai bem*" (LATOUR, 2010 p. 44). Porém, quando existem divergências entre editores a cozinha fica visível, e publicações científicas,

relatórios da ONU, notícias de jornais e demais fontes são arregimentadas por wikipedistas para defender uma versão do texto. A busca por aliados mais poderosos para reforçar uma posição é tão similar em ambos processos que a seguinte passagem de Latour (2010, p. 48) poderia muito bem ter sido escrita sobre a Wikipédia: "*A presença ou ausência de referências, citações e notas de rodapé é um sinal tão importante de que o documento é ou não sério que um fato pode ser transformado em ficção ou uma ficção em fato apenas com o acréscimo ou a subtração de referências*".

Outro ponto que não podemos deixar de mencionar aqui é a diferença entre os idiomas da Wikipédia e suas comunidades. Se a busca por indicadores generalistas já é complexa e questionável *per se*, a utilização desse ferramental pra tratar de enciclopédias com regras editoriais distintas e editores de locais, culturas e formações também tão diversas se torna um desafio ainda maior.

Um dos trabalhos mais populares que buscou tal empreitada é "*Global disease monitoring and forecasting with Wikipedia*", onde Generous et al (2014) buscaram identificar surtos de doenças epidêmicas a partir do comportamento dos usuários das Wikipédias. Mesmo buscando realizar calibragens locais, em seus resultados fica claro que enquanto muito bem-sucedido em alguns casos, o modelo fracassou em outros.

Um novo actante entra em cena: Contropedia

O impasse apresentado ao final da seção anterior entre dinâmicas locais e generalização pode parecer, a um primeiro olhar, incompatível com os estudos CTS. Porém, nem tudo está perdido. Mesmo com alguma ressalva, Latour (2010 p. 167) relata em Cogitamus que

> "*sim, reconheço, as ferramentas digitais são um veneno. Mas, talvez, também ofereçam um remédio. Ao menos, isso é o que exploro há dez anos com os alunos dos cursos chamados 'mapeamento de controvérsias'. Talvez fosse possível aprender a se orientar nas disputas, com a condição de contar com uma plataforma suficientemente calibrada e padronizada, para dar a um público virtual – ainda a ser inventado – hábitos comuns*".

Tomando esse ponto de vista, e tendo o devido cuidado de não buscar resolver as

disputas mas orientá-las, a aproximação que parecia improvável de ferramentas de análise de rastros digitais com estudos CTS se torna possível.

Seguindo a linha de "*Mapping Controversies*" citada acima, o médialab do Instituto de Estudos Políticos de Paris (Science Po) de Latour foi enredado junto da Fundação Barcelona Media, ferramentas de visualização de dados, DesityDesign Lab do Politécnico de Milano, metadados da Wikipédia, Digital Methods Initiave da Universidade de Amsterdam e orçamentos de pesquisa da Comissão Europeia no Projeto Contropedia, em um esforço para "*construir uma plataforma de visualização e análises em tempo real de controvérsias na Wikipédia*" (Contropedia, 2013)[2].

Criado em novembro de 2013, o projeto é brevemente citado por Esteves em sua tese, mas no momento de seu estudo praticamente não havia informações publicadas. Já agora, em 2017, o projeto tem resultados publicados e uma versão beta de sua plataforma está disponível para uso. Desta feita, desbravaremos esse esforço que busca "*prover um melhor entendimento de fenômenos sociotécnicos que acontecem na internet e equipar cidadãos com ferramentas para desenrolarem plenamente a complexidade de controvérsias*" (Contropedia, 2013).

Diferente de outros estudos citados aqui, a Contropedia não se propõe a medir quais artigos são controversos dentro de uma Wikipédia, e sim quais tópicos são controversos dentro de um determinado artigo. O índice de controvérsia da Contropedia é associado a *wikilinks*[3],e é calculado somando as edições (com alguma remoção de conteúdo) de frases onde o *wikilink* apareça no artigo, sendo que, caso mais de um *wikilink* apareça na mesma frase, o peso de controvérsia atribuído é proporcional ao número de *wikilinks* na frase editada. (BORRA, 2015)

A partir deste indicador, a plataforma Contropedia oferece algumas visualizações de dados. A primeira e mais simples de todas é uma linha do tempo que exibe a atividade editoral do artigo. Abaixo dela é apresentada a ferramenta *Layer*, que oferece visualização da versão mais recente do artigo com os *links* internos sendo coloridos conforme foram disputados no histórico de edições. Outra ferramenta oferecida é a *Dashboard*, com dados

2 Todas as citações de textos da Contropedia foram traduzidas livremente pelo autor.
3 Link de um artigo da Wikipédia pra outro.

agrupados sobre os *wikilinks* disputados, onde podem ser vistos também links que não estão mais na versão atual do artigo. Nesta visão pode ser visto tanto o grau histórico de controvérsia de um *link* como sua distribuição ao longo do tempo, assim como o número de usuários envolvidos na disputa. Outra visualização possível são os usuários envolvidos na disputa em torno de um *wikilink*, agrupados pela lógica de que "o inimigo de meu inimigo é meu amigo". Por fim, a plataforma fornece uma visualização micro da disputa de cada *wikilink*, exibindo o histórico de edições onde um determinado link foi disputado, destacando na edição apenas a frase relativa ao assunto, e não toda a modificação feita por aquela edição.

Essa abordagem desloca os esforços de indicadores de controvérsia de uma posição onde buscam resolver quantitativamente o grau de controvérsia de um verbete para uma posição alternativa onde passam a atuar como ferramenta de apoio para pesquisadores que busquem um olhar detalhado sobre a construção dos artigos. Segundo o próprio projeto, "*Contropedia destaca conhecimento instável em oposição a fatos estáveis*" (Contropedia, 2013). Parafraseando Latour em Cogitamus para falar dos indicadores de controvérsia, "*é como se houvéssemos passado da pesquisa dos matters of fact à exploração dos matters of concern*." (LATOUR, 2010 p. 160)

Cabe destacar que o projeto não alcançou todos os objetivos previstos, e a até hoje em seu site existe apenas a versão beta da plataforma. Esta versão não faz análises ao vivo dos dados da Wikipédia, e precisa que um *script* seja executado *offline* para gerar os dados que a alimentam. Inicialmente o grupo do projeto disponibilizou a análise de artigos em inglês sobre aquecimento global, e disponibilizou um e-mail onde receberiam pedidos de exibição de outros artigos.

Atualmente, a plataforma apresenta a análise de 6 artigos em português: Mesquita de Al-Aqsa, Budas de Bamiyan, Tauromaquia, Mármores de Elgin, Homossexualidade e Zwarte Piet, que foram extraídos entre janeiro de 2015 e julho de 2016. Todos esses artigos apresentam versões em vários idiomas, o que nos leva a supor que a versão em português tenha sido parte do esforço de algum pesquisador de comparar diferentes idiomas, e provavelmente nenhuma pesquisa específica sobre artigos em língua portuguesa tenha sido feita até então utilizando a plataforma.

Atualmente, após o fim do financiamento da Comissão Europeia (entre 2013 e 2015 o projeto teve recursos do NEIS - *Network of Excellence in Internet Science* e do EMAPS - *Electronic Maps to Assist Public Science*, ambos financiados pelo programa FP7 "*Science in Society*"[4]), o projeto encontra-se adormecido, esperando novos enredamentos que o façam realidade novamente (palavras de um dos membros do projeto ouvido informalmente por este pesquisador)[5].

Contropedia para Wikipédia lusófona?

Apesar de planejada originalmente para funcionar na Wikipédia em inglês, a opção da Contropedia de se guiar pelas edições em torno de *wikilinks* (que existem em abundância em todos os idiomas da enciclopédia) a torna potencialmente útil para versões em outros idiomas. Por outro lado, algumas funções de filtros da visualização de edições, que escondem edições de vandalismo recorrente para focar a pesquisa em disputas de narrativas, foram desenvolvidas a partir de especificidades da edição anglófona e, com certeza, serão problemáticas em outras versões. Sabendo dessa limitação observaremos se a plataforma resiste a testes com artigos lusófonos.

A fim de analisar o funcionamento da ferramenta para a Wikipédia em português resolvemos escolher dois artigos já disponíveis em nosso idioma, ver quais controvérsias a ferramenta apontaria e buscar detalhes de seu desenrolar. Ao abrir o artigo Homossexualidade (com *dataset* extraído em janeiro de 2015) a Contropedia consegue facilmente destacar uma disputa que ocorreu em torno de um termo que não mais está presente no artigo: “ex-gay”. Um passeio pelo histórico de edições nos mostra uma intensa disputa, com várias nuances, sobre a atuação de psicólogos que buscam “curar a homossexualidade”. Causou estranheza ver que, segundo a plataforma, essa controvérsia havia se estabilizado por volta do ano de 2010, pois o assunto ganhou as manchetes dos jornais em 2013 por ter sido pauta da Comissão de Direitos Humanos da Câmara de

4 O 7th *Framework Programme* de financiamento de pesquisa da União Europeia nasceu em 2002 como “*Science and Society*” e hoje se chama "*Science With And For Society*",com orçamento de 462,2 milhões de euros a serem executados entre 2014 e 2020.

5 Era previsto para esse estudo uma entrevista de um pesquisador que trabalhou no projeto Contropedia, mas por problemas logísticos o encontro se resumiu a uma conversa de *coffee-break*. Espera-se que em trabalhos futuros olhares de atores que trabalharam nos laboratórios sejam incluídos na pesquisa.

Deputados, e essa exposição midiática teria grande potencial de reabrir a discussão. Um mergulho mais profundo no histórico de edições do artigo apresentou a explicação: por questões estilísticas, o *wikilink* para "ex-gay" passou a ser apresentando dentro de um *template*, levando a Contropedia a acreditar que ele havia sido removido do artigo. Na verdade, a controvérsia continuou se acalorando ao longo dos anos apesar de seus metadados não mais se apresentarem conforme esperados pelo sistema.

Já no artigo Tauromaquia podemos observar uma pequena controvérsia em torno do termo "gado bovino", mas o detalhamento da Contropedia das edições relativas a essa disputa apresentavam sequências de vandalismo mas não suas reversões, narrando uma estranha história de edições onde acontecem várias remoções seguidas sem que conteúdo fosse readicionado. Ao desligarmos os filtros que já imaginávamos serem problemáticos fora do inglês, a visualização (apesar de conter algumas redundâncias) ficou mais clara e podemos identificar que a disputa era mais específica sobre "corrida de touros", que, por não ter um artigo próprio na Wikipédia, acaba sendo *linkada* para "gado bovino". Vale destacar também que o caloroso debate ocorrido na página de discussão desse artigo, girando principalmente em torno da "parcialidade" do mesmo, se desdobra no artigo curiosamente por elementos que acabaram não mapeados pela Contropedia, como a escolha da infocaixa, da foto em destaque e das categorias.

Um olhar apressado poderia concluir que as situações relatadas acima são uma derrota da Contropedia, mas como a ferramenta se propõe a ser um mapa que aponte indícios para pesquisadores que pretendam fazer descrições densas, ela cumpriu (mesmo que parcialmente) seu papel, chamando a atenção do pesquisador para tópicos dentro do artigo que realmente ensejavam um olhar mais aprofundado. Assim, atentos à observações de Esteves (2014, p. 295) sobre como

> "*até o fim do século XX, os cientistas sociais tinham que escolher entre análises quantitativas robustas que os distanciavam de seu tema de pesquisa ou análises qualitativas detalhadas que corriam o risco de perder de vista o contexto mais amplo em que seu objeto de estudo se inseria. Em outras palavras, precisavam escolher entre falar muito sobre pouco ou pouco sobre muito. A difusão das ferramentas computacionais de análise de dados e a grande disponibilidade de registros deixados pelos usuários dos sistemas digitais na internet abriram novas possibilidades para as ciências sociais e tornaram viável superar ao menos em parte esse dilema*",

podemos concordar com Borra, ao afirmar que "*reconhecer o potencial do histórico de edições da Wikipédia como provedor de insights sobre controvérsias sociais, e reconhecer que cada link em um artigo pode ser visto como um ponto focal de debate, nos permite utilizar a Wikipédia como um interessante local para mapear controvérsias*" (BORRA, 2015 p. 196).

Um futuro promissor

O presente estudo, apesar de breve, conseguiu identificar que três anos após a pesquisa de Esteves os estudos sobre controvérsias na Wikipédia receberam significativas contribuições, especialmente para pesquisadores que utilizem abordagens advindas do campo de estudos CTS. Nessa trajetória encontramos a Contropedia, um enredamento complexo que não se manteve estável por um longo período de tempo, mas que nos proporcionou contribuições que merecem atenção.

A utilização de rastros digitais (ou metadados) para aproximações dos estudos CTS, mesmo com algumas dificuldades encontradas, se mostrou um tanto quanto promissora para o estudo de controvérsias na Wikipédia, com seu índice de controvérsias intra-artigos resistindo a provas de força (bem limitadas) em artigos em português. Sendo assim, é pertinente propor que futuros trabalhos busquem expandir os ferramentais apresentados pela Contropedia para o estudo de controvérsias na Wikipédia.

Algumas melhorias necessárias, já mapeadas por Borra (2015), se mostraram relevantes neste estudo, como o mapeamento de outros elementos dos artigos além dos *wikilinks;* a saber: *templates*, referências e figuras. Além disso também cabe destacar a simplicidade da ferramenta *timeline*, que exibe a atividade editorial em um artigo ao longo do tempo. Se apropriando de outros estudos já feitos sobre controvérsias no nível do artigo, essa visualização poderia apresentar indicadores como o índice M de Sumi e o indicador H de Kaltenbrunner e seu cruzamento temporal com eventos externos à enciclopédia relacionados ao tema. Tais adições tornariam a plataforma potencialmente mais útil para pesquisadores interessados em utilizar a Wikipédia para estudos de controvérsias, quiçá a tornando um ponto de passagem obrigatório para tais empreitadas.

Referências Bibliográficas

Borra, Erik, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. "**Societal Controversies in Wikipedia Articles**." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 193–196. ACM, 2015.

Brandes, Ulrik, and Jürgen Lerner. "**Visual Analysis of Controversy in User-Generated Encyclopedias**." *Information Visualization* 7, no. 1, 2008.

Contropedia. *Website*, 2013. Disponível em <http://www.medialab.sciences-po.fr/projets/contropedia/>, acessado em 02 de agosto de 2017.

Esteves, Bernardo. "**As Controvérsias Da Ciência Na Wikipédia Em Português: O Caso Do Aquecimento Global**," 2014.

Generous, Nicholas; Geoffrey Fairchild; Alina Deshpande; Sara Y. Del Valle and Reid Priedhorsky. "**Global Disease Monitoring and Forecasting with Wikipedia.**" *PLoS Computational Biology* 10, no. 11 (2014): e1003892.

Kaltenbrunner, Andreas, and David Laniado. "**There Is No Deadline: Time Evolution of Wikipedia Discussions**." In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, 6. ACM, 2012.

Latour, Bruno. ***Ciência em ação: como seguir cientistas e engenheiros sociedade afora.*** *Cambridge: Harvard University Press,* 1987.

Latour, Bruno. ***Cogitamus: Seis cartas sobre as humanidades científicas***. *Paris: Editora 34,* 2010.

Sumi, Róbert, Taha Yasseri, and others. "**Edit Wars in Wikipedia**." In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference On*, 724–727. IEEE, 2011.

Vuong, Ba-Quy, Ee-Peng Lim, Aixin Sun, Minh-Tam Le, Hady Wirawan Lauw, and Kuiyu Chang. "**On Ranking Controversies in Wikipedia: Models and Evaluation.**" In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 171–182. ACM, 2008.

Yasseri, Taha; Anselm Spoerri; Mark Graham and János Kertész. "**The Most Controversial Topics in Wikipedia: A Multilingual and Geographical Analysis.**" In *Global Wikipedia: International and Cross-Cultural Issues in Online Collaboration*. Scarecrow Press, 2014.