Temer grampeado hoje!
Analisando narrativas múltiplas através da visualização de imagens[1]

Willian LOPES[2]
Fábio Gomes GOVEIA[3]
Universidade Federal do Espírito Santo, UFES

Resumo

Num momento como o atual, cheio de conflitos ideológicos, ânimos exaltados, memes e smartphones, cada evento carrega o potencial de produzir uma onda de interações digno de nota. Quando o evento é realmente grande esse potencial escala para um tsunami e o jornalista se vê ali, na praia, em frente a tal monstruosidade, com um bloco de notas nas mãos. Não há escapatória. O jornalista precisa de um lugar seguro para olhar para esse tsunami. Um lugar de onde ele possa ver o todo e entender os pequenos detalhes também. A Pulsão das Imagens é uma das tentativas de criar um jeito prático de olhar para os dados, em especial para as imagens. Não é um método definitivo, mas um método que se mostra eficiente e permite um olhar para esse tsunami de dados.

Palavras-chave: imagens, big data, visualização, política, redes sociais

1. Introdução

É cada vez mais evidente para olhos atentos que, nos dias de hoje, toda e qualquer disputa chegará mais cedo ou mais tarde, normalmente mais cedo, na internet, mais especificamente nas redes sociais. Muitas dessas disputas, inclusive, nascem lá. Seria inocência acreditar que qualquer evento de grande repercussão possa, hoje em dia, ser entendido sem levar as redes sociais em consideração. Alguns desses eventos muito populares, controversos ou até mesmo catárticos são especialmente desafiadores.

No dia 17 de maio de 2017, às 19:30, o jornal O Globo publicou uma matéria com o seguinte título: "Dono da JBS grava Temer dando aval para compra de silêncio

1 Trabalho apresentado na Divisão Temática Comunicação Multimídia, da Intercom Júnior – XIII Jornada de Iniciação Científica em Comunicação, evento componente do 40º Congresso Brasileiro de Ciências da Comunicação

2 Graduando do curso de Comunicação Social, habilitação em Jornalismo, na Universidade Federal dos Espírito Santo, UFES. E-mail: willianlopes404@gmail.com

3 Professor Doutor no departamento de Comunicação Social, UFES. E-mail: fabiogv@gmail.com

de Cunha"[4]. A internet "quebrou" logo em seguida. Não se falava de outra coisa. Muitos tweets. Muitos milhares de tweets. Esse é o nosso ponto de partida. Como entender esse momento? Como se deu a proliferação das principais imagens vinculadas à possível compra do silêncio do ex-deputado federal Eduardo Cunha pelo presidente da República.

Em eventos como esse, quando milhões de tweets e dezenas de milhares de imagens e algumas boas centenas de narrativas concorrentes surgem de uma hora para outra, a análise foge à capacidade humana de processamento. Assim, demanda formas alternativas de coleta, mineração e visualização de fenômenos que se desdobram especialmente nas redes sociais digitais, tais como Twitter, Facebook, Instagram, Youtube e outros.

Nesse cenário, o Laboratório de Estudos sobre Imagens e Cibercultura, LABIC[5], vem desenvolvendo e aprimorando técnicas a fim de lidar com esses grandes volumes de dados de forma eficiente. Para esse caso em específico, criamos uma visualização de dados que foi chamada de Pulsão das Imagens[6].

E é dessa visualização, a decisão de construí-la, os métodos aplicados e os resultados obtidos através dela, que discorreremos no presente trabalho.

2. Metodologia

A metodologia utilizada na produção dessa visualização pode ser dividida em quatro momentos fundamentais que podem, e devem, ser entendidos separadamente. São eles: Coleta, Desenho de solução, Tratamento dos dados, e Implementação.

Descreveremos cada um deles a seguir.

2.1 Coleta

Para coletar os dados usamos o Ford, um programa desenvolvido no laboratório, que usa a API pública do Twitter[7] para salvar os tweets com determinados termos. Para esse estudo foram coletados os seguintes termos: micheltemer OR temer OR diretas+ja OR diretasja + votei+dilma OR jbs; durante o período de 16 de Maio, às 19 horas, a 17

4 https://oglobo.globo.com/brasil/dono-da-jbs-grava-temer-dando-aval-para-compra-de-silencio-de-cunha-21353935 – visitado em 14/07/2017
5 http://labic.net
6 http://labic.net/micheltemer
7 https://dev.twitter.com/overview/api/tweets

de maio, às 22 horas. A saber, o operador "OR" é um operador lógico que significa que basta que um dos termos seja encontrado para satisfazer a condição. O "+", por sua vez, é um operador "AND" e exige que todos os termos seja encontrados. No total foram coletados 947418 tweets.

Para coletar as imagens referentes a esses tweets, utilizamos o Image Crawler, um programa em Java[8], também desenvolvido pelo laboratório, capaz de acessar as urls contidas nos tweets e salvá-las em disco rígido para posterior tratamento e análise. No total foram coletadas 12.637 imagens.

O conjunto que engloba essas imagens, junto com os tweets em si, serão o objeto da estratégia pelo qual nos será possível analisar o fenômeno.

2.2 Desenho da Solução

Quando falamos de visualizações de dados, e por visulizações me refiro a situações em que dados quantitativos que pela própria natureza não são visuais são transformados numa representação visual (MANOVICH, 2002), é de vital importância uma análise clara de quais são os elementos presentes nesses dados que podem possibilitar uma boa visualização e, baseado nesses elementos, escolher um modelo que melhor se encaixe com o dataset.

Um erro que pode ocorrer, e acontece com certa frequência, quando tratamos de jornalismo de dados é a preferência do pesquisador/jornalista por um determinado tipo de visualização e a tentativa de adaptar todo e qualquer tipo de dataset a esse modelo de visualização. Adaptação que perde o sentido algumas vezes. Reforço isso para reforçar a importância dessa parte do processo que é, muitas vezes, esquecido.

> [...] Jornalismo de Dados vai além do uso de dados no fazer jornalístico. Trata-se de saber trabalhar com os dados de modo que o profissional, ao fazer a mineração desse material bruto, consiga enxergar e disponibilizar, por meio de uma narrativa e da disponibilização dos próprios dados, a história que tais números tentam contar (LIMA, 2016, p. 281).

Olhando para o conjunto de dados (dataset) gerado, percebemos uma característica bastante comum nas redes sociais que é a existência de publicações

8 Uma linguagem de programação Orientada a Obejtos criada pela Sun Mycrosistems. Hoje é propriedade da Oracle Corporation. https://java.com/

focadas especialmente em imagens. Muitas das vezes uma publicação conta apenas com uma imagem, e essa imagem, que pode ser um meme textual, um print, um thumbnail de uma notícia, carrega todo o sentido da publicação. Dessa forma, uma visualização focada em imagens seria a escolha ideal (GOVEIA et. al., 2016).

Falando do dataset, olhando tanto para o dado em sua forma bruta, quanto para o evento real em si, observamos uma característica de temporariedade muito clara. Os temas coletados são temas que continuavam aparecendo na rede com alguma frequência, mas o pico ocorre quando a imprensa repercute a matéria de O Globo, o que evidencia ainda mais a necessidade de uma análise temporal dos dados.

Diante desses dois fatores, escolhemos utilizar uma visualização que já havíamos desenvolvido anteriormente. Uma das visualizações utilizadas para mapear as imagens da Copa do Mundo de Futebol de 2014[9]. Essa visualização possui ambas as características necessárias: Foco em imagens, e caráter temporal.

Essa aplicação chamada Cores da Copa, foi um conjunto de visualizações feitas pelo laboratório para entender a narrativa das cores das imagens durante a Copa. Uma visualização que trata das imagens em caráter temporal e se mostra muito interessante para esse dataset em particular.

2.3 Tratamento dos dados.

Durante nossas coletas da Copa do Mundo, tínhamos uma metodologia que calculava a quantidade de republicações das imagens em períodos determinados. Tratava-se de uma coleta específica pensada para a visualização (GOVEIA et. al. 2014).

Esse não era o caso para essa coleta e, por isso, foi preciso adaptar, tratar e preparar os dados para um visualização como essa.

O primeiro passo é tratar as imagens. Para isso utilizamos o script colors, um script em python[10] que calcula a cor predominante das imagens, e o script AISI[11], que verifica a similaridade das imagens. O colors é importante pois a visualização usa a cor das imagens para fazer a renderização. Não obstante, o AISI também tem suma importância por favorecer o entendimento das narrativas das imagens pois, muitas vezes, as imagens são "replicadas" de forma indireta. Existe uma discussão filosófica

9 http://labic.net/coresdacopa
10 Linguagem de programação aberta, gerida pela Python Software Foundation – http://python.org
11 http://www.labic.net/cartografia/aisi-script-de-identificacao-de-imagens-similares/

sobre a identidade dessas imagens, "Imagens 'iguais', mas de fontes diferentes são a mesma imagem?", mas não vou entrar nesse mérito aqui.

Analisadas as imagens, tanto no que se refere à suas cores, quanto no quesito similaridades, o próximo passo é adaptar o dataset do formato atual, resultado da coleta de dados, para um formato similar ao utilizado no cores da copa.

O formato resultante da coleta é um formato que tem como objeto os tweets. Cada tweet é uma linha de um arquivo csv (*comma separeted values*). Cada tweet possui, dentre outros valores, o tempo em que ele foi publicado, que chamamos de *timestamp*, e as imagens associadas a ele. Essa informação será importante.

O formato que precisamos é um formato json[12] que tem como objeto as imagens. Cada imagem contendo os seguintes dados: url da imagem, cor da imagem, e um mapa dos *timestamps* onde essa imagem aparecem, no caso os tweets aos quais ela está ligada, com sua quantidade de republicações. Esses tempos serão fundamentais para a visualização.

Para esse tratamento, foi desenvolvido um script capaz de ler os arquivos no formato de entrada e escrever um novo arquivo no formato desejado.

O script constrói um mapa, uma estrutura chave-valor, onde a chave é nome das imagens, que é único, e o valor são os dados necessários, principalmente o tempo, a quantidade de rts, a url original da imagem e sua cor. A cada linha do arquivo de entrada é feito uma busca da imagem no mapa, caso seja encontrada, os novos dados são inseridos junto aos dados já mapeados dessa imagem; caso não seja encontrada, a nova imagem é inserida como um novo registro. Mapeadas as imagens, o novo arquivo é escrito.

Com esse novo arquivo podemos, finalmente, implementar a visualização.

2.4 Implementação.

A visualização é implementada em d3js[13], uma biblioteca de visualiação e mapeamento de dados, construído sobre a linguagem de programação javascript.

Essa ferramenta permite o pareamento entre objetos conceituais (como um JSONObject, por exemplo) e objetos html, objetos de uma página web. Dessa forma é possível que cada imagem do arquivo, reprentada por um JSONObject, possa se tornar

12 Javascript Object Notation - http://www.json.org
13 Data-Driven Documents – http://d3js.org

os círculos que serão vistos na visualização. D3js permite também a criação de comportamentos para esses objetos, comportamentos que são, em parte, padronizados, mas que, por outro lado, dependem dos atributos dos objetos originais.

Na visualização (Figura 1) temos dois estados: 1)Um estado de visualização geral, onde o total dos compartilhamentos de cada imagem é mais importante; e 2) Um estado de animação onde podemos ver a narrativa das imagens através da relação dos compartilhamentos em função do tempo. Em ambos os estados, a posição das imagens no eixo x é baseada na matiz (hue) da cor predominante. Para os demais aspectos, cada estado tem regras separadas.

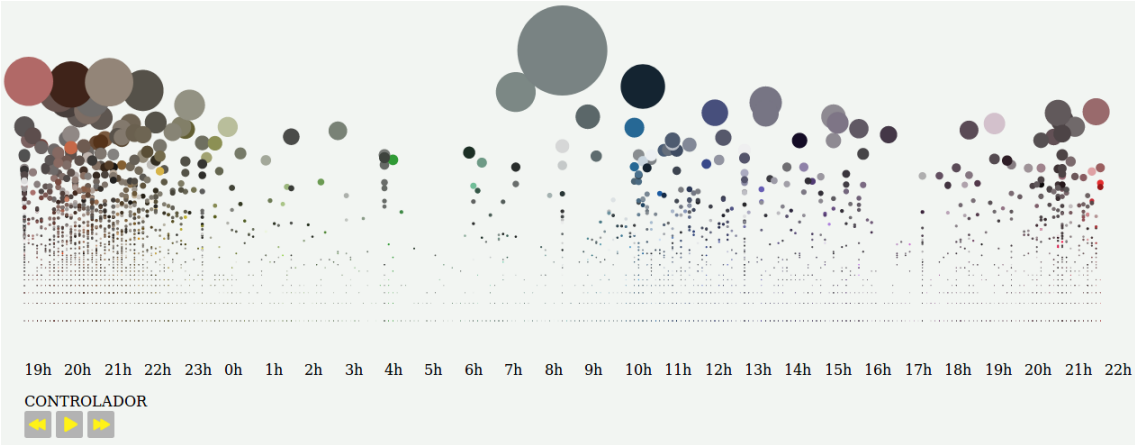

Figura 1. Print da visualização "Pulsão das Imagens". Visualização dinâmica disponível em http://labic.net/micheltemer/

No estado geral, tanto a posição no eixo y, quanto o raio do círculo, dependem do total das republicações.

No estado de animação, o eixo y e o raio do círculo variam com o tempo. Nessa visualização em particular, os recortes são de 1 hora. A cada momento, o eixo y representa a quantidade de retweets desse momento, e o raio do círculo remete ao total acumulado até aquele momento. Dessa forma podemos ver o momento em que as imagens surgem, e quanto tempo elas perduram. Falarei disso a seguir.

3. Resultados

A visualização de dados tem como principal objetivo permitir que a informação seja enxergada de diferentes formas. Aqui, vamos analisar algumas das coisas que

podem ser vistas e as conclusões que podem ser tiradas desse dataset, utilizando a visualização como ferramenta.

Tratando-se de uma visualização focada no caráter temporal, podemos fazer dois tipos de olhares distintos para entender a narrativa. Um olhar sobre a totalidade do conjunto em diferentes tempos, e um olhar sobre as narrativas de determinadas imagens durante o processo.

3.1 Geral

De forma geral, temos o fim do dia 16, relativamente tranquilo, mas com a aparição da maior imagem do dataset por volta das 2 horas da manhã. Trata-se de uma imagem que fala sobre a reforma da previdência, de um tweet do @RiKruschewky que tinha, no momento 39 mil retweets (Figura 2). Outras imagens aparecem, algumas com mais relevância, outras com menos.

Por volta do das 19 horas do dia 17, surge a imagem de Urgente, do plantão do @JornalOGlobo, imagem do thumbnail da notícia de que o "Dono da JBS gravou o então presidente Michel Temer dando aval para a compra do silêncio de Eduardo Cunha" (Figura 3). No momento das 19 horas já temos algumas reações, como a imagem de um homem segurando um cartaz escrito "Não é ódio. É nojo"; uma imagem da Sandra Annenberg, âncora do Jornal Hoje, com a legenda "Se fudeu"; e uma montagem com várias fotos do ex-presidente Lula gargalhando.

Às 20 horas, o tweet do @jornalOGlobo segue tendo muitos retweets enquanto uma profusão de outras imagens surgem. Imagens relembrando os protestos a favor do impeachment da ex-presidenta Dilma. Uma imagem do Jornal Nacional, que também repercutiu a notícia. Memes envolvendo o próprio Michel Temer. Imagem de outros jornais que também fizeram posts à respeito. Imagens do ex-presidente Lula ex-presidenta Dilma rindo em alguma ocasião. E cartazes convocando para manifestações com a hashtag #DiretasJa e #ForaTemer.

Às 21 horas começam a aparecer mais memes. Temer ouvindo sua fita, fazendo referência à série 13 Reasons Why[14]. Thumbnails de "Urgente" do Estadão e da Folha de São Paulo. Muitos cartazes contra o presidente.

14 Série da Netflix recente na época. Nela os personagens recebiam fitas que falavam sobre eles próprios

Figura 3. Print em twitter.com

No último período, o das 22 horas, temos mais memes menores surgindo, enquanto os grandes memes e as imagens maiores continuam a ser republicadas. Nesse momento temos quase uma estagnação nos retweets dos grandes jornais e uma profusão de novos memes e imagens com caráter humorístico. Alguns desses pequenos memes vão surgir e se manter pequenos, e outros crescerão no resto da noite e da madrugada, mas esse período sai do range dessa análise.

Nessa análise temporal, podemos perceber que a imagem relacionado ao post oficial do @jornalOGlobo tem um comportamento característico e, talvez, único. A figura do Urgente (Figura 3), que acontece por volta das 19 horas, apresenta um pico de republicações assim que é lançada, algo comum em grandes notícias, e permanece como figura mais relevante por algum tempo. Por sua vez, o que se sobressai no momento

seguinte são os memes, as paródias e as piadas envolvendo o caso. Podemos conjecturar, a partir desse movimento, os motivos para isso.

A impressão que temos, observando esse cenário, tendo por meio a visualização proposta, é que, em um primeiro momento, a surpresa da notícia leva a muitos retweets, mas esse comportamento é reduzido logo em seguida, devido ao fato de que o "fator novidade" se esvai tão rápido quanto o crescimento da notícia. Entenda, compartilhar uma notícia recém lançada, tem um fator de status social que é diluído com o tempo. Um pensamento do tipo "as pessoas precisam ver isso" é muito mais comum 5 minutos pós publicação, do que horas depois da publicação.

Outro fator importante é a multiplicidade e a completude das narrativas complementares, representadas aqui pelos memes, pelo escárnio e pela revolta. Explico.

As publicações de um perfil jornalístico tradicional, como o Jornal O Globo, tem uma narrativa limitada a um caráter mais técnico e factual. Essa narrativa, por ser limitada e por tentar ser uma narrativa imparcial, é cooptada por outros usuários que não estão presos a essas amarras. Essas novas narrativas carregam consigo a informação principal (ou pelo menos parte dela) acrescidas de algo mais. Esse algo mais pode variar desde um comentário cheio de ponderação e crítica, até o ódio e o desprezo em sua forma mais pura, passando ou não pelo humor. Para o usuário que chega mais tardiamente à discussão, é mais producente/interessante, me parece, compartilhar e unir sua voz em coro, a uma dessas narrativas mais complexas do que a original. “Por quê compartilhar a notícia, se eu posso compartilhar essa imagem que transmite a notícia e mais um sentimento com o qual eu concordo?”

É nesse cenário que, a meu ver, os memes, as releituras, as sátiras e as convocatórias ganham força. Vale ressaltar, porém, que isso vale tanto para as repercussões maiores, que podem começar com perfis de maior alcance, quanto com reações de usuários menores que vão ter menos retweets, mas não deixam de ser representativas, tanto para o panorama geral, quanto para seus grupos locais de discussão.

Usando a comparação que Richard Dawkins faz entre os memes e os genes no seu O Gene Egoísta (DAWKINS, 1976), e extrapolando para os tweets de forma geral, podemos dizer que o tweet da notícia oficial possui longevidade através da fidelidade, o que garante a ela uma vida longa, porém cada vez mais modesta, enquanto as novas

narrativas, que são descendentes da narrativa original, se espalham em novas árvores por possuir muita fecundidade, a capacidade de promover muitas réplicas. Assim como os genes, as narrativas também disputam pela sobrevivência.

3.2 Narrativas locais.

Outra forma de análise que esse tipo de visualização permite é uma análise da história de uma única imagem em separado. De forma geral, olhar para o todo é o melhor jeito de entender a história do todo, mas em alguns casos, algumas imagens, ou seus comportamentos, se destacam ou destoam do normal e torna-se essencial olhar para elas com maior atenção. Em resumo podemos dizer que, olhar para o todo nos permite entender as regras e chama a atenção para as exceções. Por outro lado, olhar as exceções a fundo também é extremamente importante.

A imagem que, obviamente, mais se destaca é a imagem mais retweetada (Figura 2) que é de um tweet do final de abril e que é retweetada durante o período de coleta. Temos alguns pontos de interesse aqui.

1) Essa imagem aparece apenas uma vez no período de coleta, mas, pelo fato de ela já ter um total acumulado de retweets, ela surge como a imagem mais relevante, logo maior, na visualização. Uma pergunta que decorre desse fato é "Essa imagem é mesmo relevante?" e a resposta não é tão simples quanto pode parecer. Seria razoável pensar que, se ela aparece apenas uma vez no período, ela deveria ser considerada apenas uma vez, por outro lado, é inegável a representatividade de uma imagem com 39 mil retweets. É claro, porém, que é mais fácil decidir, olhando para a visualização, se essa imagem, em particular, é relevante ou não. Mas é preciso entender que esse tipo de decisão deve ser tomada ainda na implementação e que essa discussão deve ser feita de forma minuciosa. Nesse caso, a exceção serve até mesmo para questionar o método.

E 2) Trata-se de uma imagem de um tweet de um usuário menos de 500 seguidores. Como essa imagem foi tão longe? Fica o questionamento para uma observação futura.

Outra imagem interessante é a imagem de um cartaz onde está escrito "Vai reformar seu cú Temer!" (Figura 4). Essa imagem aparece duas vezes na visualização, na mesma posição do eixo x, mas com quantidades de retweets diferentes. Procurando-a no dataset encontramos os usuários de onde os tweets originais surgem. Os dois tweets

são do dia 28 de abril e reaparecem no dataset por serem retweetados no período de coleta, assim como a imagem citada anteriormente. Os tweets originais são de @orgaabs e de @NatanPontia, às 6:43 e às 7:19, respectivamente. A proximidade dos horários pode significar que o segundo "copiou" o tweet do primeiro, ou que ambos tiveram acesso a essa imagem por via de uma mesma fonte, mas essa questão não é importante.

O importante é que, para olhos humanos, essas duas imagens são claramente as mesmas, mas isso não fica tão óbvio para a máquina. O fato de elas aparecerem na mesma posição no eixo x, significa que o script colors calculou suas cores e o resultado foi igual, ao menos no quesito matiz, mas o AISI não foi capaz de identificá-las como iguais por algum motivo, o que é razoável quando lembramos que o AISI usa um processo heurístico, ou seja, não exato.

Figura 4. Print em twitter.com

Esse comportamento aponta para o que queremos mostrar aqui, que uma visualização bem construída pode nos permitir observar e questionar coisas que não veríamos olhando apenas para o dataset, tanto nos dados, quanto nos próprios métodos.

Outras coisas podem ser descobertas olhando para as narrativas individuais das imagens e isso pode, e deve, ser feito com mais cuidado e tempo, afinal, tratam-se de quase 1 milhão de tweets com mais de 12 mil imagens.

4. Conclusão

Com esse exemplo de aplicação, fica suficientemente claro que uma boa visualização pode proporcionar uma perspectiva diferente de onde podemos observar os dados, e especial quando o volume desses dados é grande, e obter os mais diferentes resultados.

É claro, porém, que não existe uma visualização de dados definitiva que vá resolver todos os problemas do jornalismo de dados. Dessa forma, saber escolher uma visualização adequada e, se necessário, construir novas, se mostra essencial para o pesquisador ou grupo de pesquisa.

Fica evidente, também, que os métodos usados para o processamento dos dados pré-visualização precisam ser sempre, questionados e aprimorados para processos cada vez mais assertivos. Isso do ponto de vista técnico.

Olhando especificamente para este caso, temos como descoberta principal o enfraquecimento de uma narrativa oficial enquanto narrativas pessoais e mais localizadas surgem e ganham força. Esse fenômeno pode ser um fenômeno particular desse evento, ou um fato mais geral. Cabe observar eventos futuros e avaliar essa questão.

É possível que sejam necessárias ferramentas ainda mais sofisticadas sejam necessárias para essas análises. Pensar em novas formas de visualização e análise são os próximos passos da pesquisa.

5. REFERÊNCIAS

FLUSSER, Vilém. Comunicologia: reflexões sobre o futuro: as Conferências de Bochum. São Paulo: Martins Fontes, 2014.

GOVEIA. Fábio; CARREIRA, Lia Scarton; SOUZA, Tasso Gasparini de; HONORATO, Johanna Inácia; CYPRIANO, Lucas Oggioni; HAACKE, Verônica A. Ribeiro; LOPES, Willian. **Cores da Copa: ferramenta de visualização e análise das imagens compartilhadas no Twitter durante a Copa do Mundo de 2014.** VIII Simpósio Nacional ABCiber. ESPM . São Paulo, 2014. Disponível em http://www.abciber.org.br/simposio2014/anais/GTs/tasso_gasparini_de_souza_113.pdf

MANOVICH, Lev. A visualização de dados como uma nova abstração anti-sublime. Berlim, 2002. Disponível em http://www.ppgav.eba.ufrj.br/wp-content/uploads/2012/01/ae11_lev_manovich.pdf

LIMA, Soraia. Jornalismo de Dados no Brasil: tendências e desafios. In: SILVA, Tarcizio; STABILE, Max (orgs). Monitoramento e pesquisa em mídias sociais: metodologias, aplicações e inovações São Paulo: Uva Limão, 2016.

GOVEIA, Fábio; HONORATO, Johanna Inácia; SOUZA, Tasso Gasparini de; HAACKE, Verônica A. Ribeiro. Investigação de fotografia nas redes sociais: novos desafios, novos métodos. XXXIX Congresso Brasileiro de Ciências da Comunicação. USP, São Paulo, 2016. Disponível em http://portalintercom.org.br/anais/nacional2016/resumos/R11-1953-1.pdf

MACHADO, Arlindo. Arte e mídia. Rio de Janeiro: Zahar, 2007.

DAWKINS, R. The selfish gene. Oxford: OUP, 1976