



6º SIMPÓSIO  
INTERNACIONAL DE  
CIBERJORNALISMO

Performance em Ciberjornalismo:  
tecnologia, inovação e eficiência

Performance in cyberjournalism: technology, innovation and efficiency

1 a 3 de junho/2015 na UFMS  
em Campo Grande-MS - Brasil

## A aplicação de visualizações gráficas para a análise jornalística de dados do Twitter<sup>1</sup>

Me. Caroline Q. Santos, Dra. Silvia M.W. Moraes, Dra. Milene S. Silveira, Dr.

Marcelo Träsel, Dr. Roberto Tietzmann, Dra. Isabel H. Manssour<sup>2</sup>

**Resumo:** A penetração das redes sociais entre os usuários da Internet vem crescendo continuamente desde a metade dos anos 2000, gerando um grande volume de dados que, para além das expressões de sociabilidade, refletem o cotidiano social e acontecimentos de interesse jornalístico. Os dados gerados como subproduto do uso de redes sociais podem ser analisados através de técnicas da informática, como visualizações de dados, com o objetivo de iluminar padrões e tendências que de outra forma permaneceriam ocultos. Técnicas de análise de redes sociais vêm sendo usadas pela mídia para ilustrar o comportamento social durante eventos específicos, como competições esportivas, eleições, festivais, entre outras manifestações sociais. Ao longo da Copa do Mundo de futebol de 2014 no Brasil, muitos veículos jornalísticos publicaram visualizações gráficas com base na resposta da população ao evento nas redes sociais. Entretanto, em muitas situações os gráficos simples não são suficientes para uma análise aprofundada destas manifestações individuais. O objetivo da pesquisa é compreender como as diferentes técnicas de visualização podem ajudar jornalistas e pesquisadores a entender o comportamento e o sentimento dos usuários das redes sociais. Tweets publicados durante a Copa do Mundo de 2014 foram coletados, processados e usados para criar visualizações referentes às palavras-chave mais comuns e ao sentimento da população que mantém perfis no serviço Twitter. As visualizações passíveis de interação por parte dos participantes do grupo focal mostraram potencial para permitir a identificação de padrões e tendências de interesse jornalístico.

**Palavras-chave:** Cibercultura. Redes sociais. Jornalismo Guiado por Dados. Twitter. Visualização gráfica.

<sup>1</sup> Artigo enviado na modalidade comunicação oral para o GT 1 – Mídias Sociais.

<sup>2</sup> O artigo é resultado parcial de projeto de pesquisa contemplado pelo edital PRAIAS da Pontifícia Universidade Católica do Rio Grande do Sul, reunindo docentes e discentes das faculdades de Comunicação e Informática. Contato com os coordenadores: [rtietz@pucrs.br](mailto:rtietz@pucrs.br) e [isabel.manssour@pucrs.br](mailto:isabel.manssour@pucrs.br).

## 1 INTRODUÇÃO

O uso de redes sociais na Internet por uma quantidade crescente de indivíduos vem se ampliando continuamente desde a metade da década de 2000, gerando grandes volumes de dados que refletem as opiniões e o humor de parte da população mundial a respeito de assuntos do cotidiano. Tecnologias que permitam a compreensão destes dados se tornaram estratégicas em diversos cenários, tais como o marketing, campanhas políticas, saúde pública, ou especulação financeira, por exemplo. As organizações jornalísticas também vêm buscando formas de usar as tecnologias de coleta, tratamento e análise de dados gerados por usuários de redes sociais nas rotinas produtivas das redações. Este artigo descreve métodos de coleta, tratamento e visualização de dados oriundos da rede social Twitter e discute algumas de suas possíveis aplicações no jornalismo.

Conforme Recuero (2009, p.164), as redes sociais na Internet são uma fonte privilegiada de informação a respeito de dinâmicas sociais na sociedade em rede:

Quando se focam as expressões sociais na Rede como expressões de grupos sociais, temos um ecossistema único: na Internet, é possível examinar em minúcias boa parte das trocas sociais que são mediadas pelo computador. Isso porque essas trocas, suas conversações e interações, ao contrário, por exemplo, da linguagem oral, tendem a permanecer gravadas nesse espaço. Ali, é possível observar um histórico de interações, as apropriações realizadas pelos atores sociais, a constituição dos grupos e sua possível ruptura. Compreender essas redes é essencial, portanto, para compreender também a apropriação da Internet como ferramenta da organização social e informação contemporânea. É essencial para compreender os novos valores construídos, os fluxos de informação divididos e as mobilizações que emergem no ciberespaço.

Sua adoção cada vez mais ampla pela parcela da humanidade que goza de acesso à Internet torna serviços como Twitter, Facebook, Pinterest, YouTube, entre outros, ambientes acessíveis e convenientes para acompanhar a evolução da opinião pública. Este cenário cria oportunidades para combinar os subprodutos da sociabilidade no ciberespaço com técnicas computacionais de coleta, análise e visualização de dados, de forma a apoiar profissionais especializados na extração de informação.

Exemplos de aplicação destas técnicas podem ser encontrados em estudos sobre campanhas eleitorais e eleições (HANNA et al. 2013), inferência de orientação política de usuários do Twitter (CONOVER et al., 2011), ou eventos esportivos (SAHAMI et al., 2011). Técnicas de aprimoramento de visualização e interação com dados vem sendo desenvolvidas

na área da Ciência da Computação, com o objetivo de contribuir com a pesquisa nessas e outras linhas de estudo. Este trabalho apresenta resultados parciais do projeto de pesquisa interdisciplinar “Aplicação de mineração visual de dados em redes sociais”, desenvolvido em conjunto pelas faculdades de Comunicação e de Informática da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). O foco do artigo são as contribuições da visualização de dados estática ou interativa para a rotina produtiva dos jornalistas.

Conforme Heer, Bostock e Ogievetsky (2010), a meta das visualizações é auxiliar na compreensão de dados, ao apoiar a habilidade do sistema visual humano para o reconhecimento de padrões, identificação de tendências e percepção de registros desviantes. Se as representações visuais são bem planejadas e executadas, elas podem substituir os cálculos cognitivos intuitivos com inferências perceptivas simples e incentivar a compreensão, a construção de memórias e o processo de tomada de decisões. Em termos jornalísticos, os gráficos, sejam estáticos ou manipuláveis, permitem identificar informação dotada de valor-notícia em bases de dados que, devido a seu volume, seriam incompreensíveis a curto prazo, se examinadas “a olho nu”, isto é, se o repórter fosse obrigado a ler todos os registros e realizar cálculos mentais ou anotações em papel. Neste sentido, as visualizações gráficas funcionam como interfaces entre os dados estocados numa planilha ou banco e seu observador humano. No contexto do jornalismo, se costuma usar o termo infografia, que conforme Teixeira (2009, p.4) seria

a modalidade discursiva do jornalismo informativo, no qual a presença indissociável de imagem e texto em uma construção narrativa permite a compreensão de um fenômeno específico, como um acontecimento jornalístico, ou o funcionamento de algo complexo ou difícil de ser descrito em uma narrativa textual convencional.

Técnicas para a visualização de bases de dados geradas por usuários de redes sociais vêm sendo usadas por diversos veículos jornalísticos para ilustrar o comportamento das pessoas em relação a um tema ou a um evento específico. Na Copa do Mundo de 2014 no Brasil, visualizações gráficas relacionadas à competição foram destacadas e amplamente compartilhadas em redes sociais na Internet.<sup>3</sup> A produção destes gráficos frequentemente exige o uso de técnicas computacionais, como processamento de linguagem natural, mineração de

---

<sup>3</sup> Ver, por exemplo: <http://t.co/jOoei6z69j>, <https://interactive.twitter.com/wcmentions/>, <http://t.co/lvIF36lqzz> e <http://goo.gl/rPzfme>.

dados e de textos, entre outras, bem como processos analíticos complexos para a combinação de fontes variadas de dados. Além disso, é preciso atingir um equilíbrio entre os aspectos computacionais e os aspectos estéticos da visualização de dados ao se produzir tabelas, gráficos e outros tipos de imagens, de modo que elas favoreçam uma compreensão rápida e correta da informação.

Pesquisadores como Elias e Bezerianos (2011), Elias, Aufare e Bezerianos (2013), Heer, Viegas e Wattenberg (2007), Heer, Bostock e Ogievetsky (2010) vêm empreendendo estudos sobre a evolução destas técnicas. Em muitas situações, gráficos simples não permitem traduzir a complexidade de bases de dados para uma audiência leiga e, portanto, a primeira fase deste projeto de pesquisa procura entender como técnicas de visualização podem ajudar a imprensa a analisar o comportamento e o sentimento dos usuários de redes sociais. O foco recai sobre o serviço Twitter<sup>4</sup>, criado em 2006. Nesta rede social, os clientes publicam textos, imagens, hyperlinks e palavras-chave expressando suas opiniões e sentimentos sobre um determinado tema, que são automaticamente enviados aos atores com os quais mantêm conexões e podem ser propagados por estes para outros nós e núcleos. A identificação do sentimento de cada mensagem é sujeita a erros, ambiguidade e apresenta muitos desafios do ponto de vista da tecnologia e da linguagem (BARBOSA et al., 2012). Assim, a questão norteadora da pesquisa aqui relatada se dirige ao problema de como apresentar os dados, de forma a apoiar a compreensão dos mesmos e a extração de informação pelo leitor<sup>5</sup>. Noutras palavras, *como a visualização de dados pode favorecer a obtenção de informação que seria muito difícil de perceber e adquirir de outra forma?*

Para tentar responder a esta pergunta, foram coletados e processados textos publicados no Twitter durante a Copa do Mundo de 2014. A partir deste *corpus*, foram geradas visualizações dos sentimentos dos usuários brasileiros da rede social durante a partida semifinal entre Brasil e Alemanha, bem como das palavras-chave (*hashtags*) mais usadas: #vatercopa, #vaitercopasim e #nãovaitercopa. A partir deste material, foi realizado um grupo focal com alunos do Programa de Pós-Graduação em Comunicação Social da PUCRS, cujos resultados serão discutidos ao final deste artigo. É preciso ressaltar, porém, que este primeiro grupo focal

---

<sup>4</sup> <http://twitter.com>.

<sup>5</sup> Usa-se aqui o termo leitor em relação a visualizações gráficas seguindo o raciocínio de SANTAELLA (2004) sobre a terminologia usada para se referir aos “usuários” de produtos hipertextuais que combinam linguagem escrita e audiovisual.

teve um caráter exploratório, voltado para a preparação das próximas fases da pesquisa, nas quais jornalistas profissionais serão os sujeitos e o método será a entrevista semi-estruturada. Esta comunicação, portanto, não apresenta resultados finais, nem muito menos generalizáveis, dos estudos em andamento, mas se propõe principalmente a levantar questões e receber sugestões de colegas pesquisadores, além de compartilhar técnicas de investigação computacional de redes sociais.

Na próxima seção, a visualização gráfica será discutida no contexto do jornalismo contemporâneo. Em seguida, serão apresentadas as técnicas usadas na coleta, análise e representação de dados gerados por usuários do Twitter durante a partida de futebol entre Brasil e Alemanha na Copa do Mundo de 2014. Posteriormente, serão debatidas as visualizações gráficas resultantes e as impressões que elas suscitaram no grupo focal. Ao final, serão apresentadas questões e os caminhos previstos para o desenvolvimento do restante do projeto de pesquisa.

## 2 O JORNALISMO GUIADO POR DADOS

O jornalismo guiado por dados (JGD) ou *data-driven journalism* compreende diversas práticas profissionais, cujo ponto em comum é o uso de bases de dados como principal fonte de informação para a produção de notícias. As práticas de JGD envolvem técnicas de reportagem assistida por computador (RAC), visualização de dados, infografia, criação e manutenção de bases de dados e a política de acesso à informação e transparência pública de governos. Contemporaneamente, outras expressões usadas para se referir a essa especialidade profissional são “jornalismo de dados” ou *data journalism* (GRAY et al., 2012; PARASIE et al., 2013) e “jornalismo computacional” ou *computational journalism* (ANDERSON, 2012; DIAKOPOULOS, 2012; COHEN, 2011; HAMILTON et al., 2009; LEWIS et al., 2013; LIMA JR., 2011). Abaixo, se apresenta o conceito de jornalismo guiado por dados.

Para Matzat (2011), o termo JGD seria o mais adequado, de fato, para se referir a essas práticas, porque ressalta o caráter dinâmico da apuração e apresentação de informações através da navegação e interação com bancos de dados online.

De fato, o jornalismo de dados vai um passo decisivo além: ele se baseia em registros numéricos não só como fontes de pesquisa, mas também torna os dados o tema central da história e sua apresentação. Ele reside na intersecção de três áreas:

primeiro, jornalismo visual, ou infografia, segundo, multimídia e narrativas interativas, terceiro, jornalismo investigativo. (MATZAT, 2011, s.p.)<sup>6</sup>

O jornalista guiado por dados assume uma postura exploratória, característica do leitor imersivo experto descrito por Santaella (2004, p.60), que busca a informação na Internet a partir do raciocínio dedutivo, deixando-se guiar tanto pela experiência quanto por pistas disponíveis na rede para encontrar seu objetivo. Não há na bibliografia sobre o tema nenhuma discussão aprofundada a respeito de uma possível diferença conceitual entre *data journalism* e *data-driven journalism*. A maioria dos jornalistas e alguns pesquisadores – Dagiral e Parasie (2013), por exemplo – usam os termos como sinônimos.

Os repórteres passam a, cada vez mais, assumir também o papel de técnicos, ao deixarem de delegar a operação de hardware e software para especialistas e passarem eles mesmos a assumir a operação dessas tecnologias.

Por volta de 1991, Meyer afirma na segunda edição de seu livro que jornalistas interessados em RAC eram uma raça particular de repórteres, que frequentemente compravam seus próprios computadores antes mesmo de as redações os adotarem. [...] Entretanto, o reporter assistido por computador era ainda primordialmente um jornalista, não um técnico; a meta subjacente era produzir uma notícia melhor. Hoje, porém, o programador-jornalista se diferencia deste reporter de RAC ao enxergar o produto final não como uma notícia, mas como um “artefato produtivo” de “filtragem de informação”.<sup>7</sup> (LEWIS e USHER, 2013, p.605)

A partir desta percepção, os autores consideram o JGD uma fusão das ciências da computação e jornalismo, a partir da qual os programadores começam a assumir a centralidade nas redações e participar ativamente da elaboração de formatos jornalísticos como visualizações de dados, aplicativos, algoritmos de geração e recomendação de notícias, entre outros tipos de peças noticiosas derivadas da informática.

Parasie e Dagiral (2013) seguem numa vertente semelhante, argumentando que os projetos de JGD dentro das redações repousam sobre o tratamento, análise e visualização de dados relacionados a uma grande variedade de domínios, como criminalidade, demografia, transportes, infraestrutura, orçamento público, meio ambiente, entre outros, e aos agentes sociais responsáveis pelo gerenciamento de ações governamentais nestas áreas, como políticos, a respeito dos quais se busca levantar informações sobre financiamento de campanhas, uso de

---

<sup>6</sup> Tradução livre.

<sup>7</sup> Tradução livre.

verbas, votos, ou projetos de lei apresentados, por exemplo) em todas as esferas administrativas. Estes dados são reunidos então em “programas informáticos apresentados como constituintes intrínsecos de produtos jornalísticos”<sup>8</sup> (PARASIE et al., 2013, p.53).

Lima Jr. (2011), Anderson (2012), Diakopoulos (2012) e Lewis e Usher (2013) preferem usar a expressão “jornalismo computacional”, proposta por Cohen et al. (2011), em lugar de JGD. O jornalismo computacional seria, inicialmente, “a mudança na forma como as notícias são descobertas, apresentadas, agregadas, monetizadas e arquivadas”<sup>9</sup> (COHEN et al., 2011, p.1). A computação, nesse caso, é vista como um fator de evolução do jornalismo, permitindo inovações como detecção automática de tópicos na rede mundial de computadores, análise de vídeos, personalização, agregação, visualização e produção de sentido. Num estudo anterior, porém, Hamilton e Turner definem o jornalismo computacional como o uso de “algoritmos, dados e ciências sociais para reforçar as funções de fiscalização do jornalismo”<sup>10</sup> (HAMILTON et al., 2009, p.2). A diferença entre o conceito de jornalismo computacional e JGD, entretanto, não fica clara nos estudos destes pesquisadores. As práticas atribuídas a ambos os conceitos se sobrepõem frequentemente.

A partir das noções expostas acima, o conceito de Jornalismo Guiado por Dados com o qual a comunidade profissional jornalística vem operando pode ser definido da seguinte forma: o jornalismo guiado por dados é a aplicação da computação e dos saberes das ciências sociais na interpretação de dados, com o objetivo de ampliar a função da imprensa como defensora do interesse público.

A principal promessa do JGD para a profissão jornalística especificamente e para a sociedade, de maneira geral, está inscrita já na primeira obra bibliográfica a respeito da prática: Philip Meyer (1973) garantia que a aplicação do método científico poderia conferir maior objetividade, no sentido de precisão factual, ao noticiário. Os proponentes da RAC e, mais recentemente, do JGD citados na discussão conceitual acima seguem na mesma linha, apresentando como principal vantagem da aplicação da informática à rotina produtiva das notícias a redução de erros e o embasamento de reportagens em fatos mais sólidos do que declarações de fontes humanas como políticos, empresários e mesmo o transeunte comum,

---

<sup>8</sup> Tradução livre.

<sup>9</sup> Tradução livre.

<sup>10</sup> Tradução livre.

sempre passíveis de contrabandear seus próprios interesses na avaliação dos acontecimentos. Em resumo, o sentido do JGD, para esta comunidade profissional, parece ser a eliminação do fator humano nas rotinas produtivas do jornalismo através da aplicação de técnicas da informática na apuração de notícias.

Não apenas o JGD permitiria eliminar os erros humanos na produção de notícias, como também poderia vir a eliminar os erros humanos na própria recepção do noticiário, em especial através do uso de visualizações gráficas, que simplificam relações estatísticas complexas (CAIRO, 2013), e facilitam a compreensão dos acontecimentos e fatos sociais. No editorial de estreia de um canal do New York Times dedicado ao JGD, denominado *The Upshot*, por exemplo, Leonhardt (2014, s.p.) cita como a primeira razão para o lançamento a necessidade de explicar o noticiário para os leitores do jornal ao qual o website está atrelado:

Acreditamos que muitas pessoas não entendem o noticiário tão bem como elas gostariam. Elas querem entender matérias importantes, complicadas, [...] tão bem que possam explicar os porquês e os comos dessas notícias para os seus amigos, parentes e colegas. Acreditamos que podemos ajudar os leitores a chegar a esse nível de compreensão escrevendo de forma direta, de forma simples, com a mesma voz que poderíamos usar ao escrever um e-mail a um amigo. Seremos conversacionais sem sermos estúpidos. Nós vamos construir sobre o excelente jornalismo que o The New York Times já está produzindo, ajudando os leitores a fazer conexões entre diferentes matérias e entender como essas matérias se encaixam.<sup>11</sup>

A estratégia adotada por *The Upshot* para a consecução deste objetivo é a apresentação de gráficos e visualizações de análises de dados relacionados ao noticiário do jornal. A natureza interativa de grande parte das visualizações gráficas e aplicativos noticiosos, ademais, contribuiria para este objetivo, ao permitir que o leitor escolha os caminhos que vai trilhar dentro do hipertexto e atualize uma narrativa individual a partir dos múltiplos caminhos virtuais possíveis. A imersão proporcionada pela experiência de interação com hipertextos mobiliza a cognição do leitor e o incentiva a adotar uma atitude dedutiva (SANTAELLA, 2004). Os proponentes do JGD consideram a possibilidade de interação com um aplicativo ou gráfico como uma forma de transferir o controle da narrativa para a audiência.

A visualização de dados é bastante mencionada na bibliografia a respeito do JGD, embora seja pouco discutida em seus elementos conceituais próprios. Como lembra Teixeira (2007, p.112), a infografia não é uma técnica exclusiva do jornalismo, mas é usada em

---

<sup>11</sup> Tradução livre.



publicações como livros didáticos, periódicos científicos ou manuais. Talvez por isso a infografia e, mais atualmente, a visualização de dados sejam pouco examinadas pelos pesquisadores do campo do jornalismo, com exceções como a academia espanhola (DE PABLOS, 1999; VALERO SANCHO, 2001), Cairo (2013) e a própria Teixeira (2007; 2009). A bibliografia a respeito da visualização de dados usada nos estudos de jornalismo costuma ser oriunda principalmente dos campos do design e da informática.

De qualquer modo, a infografia ou a visualização de dados são em geral analisadas como produto a ser desenvolvido pela redação, mas não como ferramenta de apoio ao repórter em sua rotina produtiva. Exceções são os estudos de Diakopoulos e Shamma (2010), que desenvolveram uma metodologia analítica e representações visuais para auxiliar jornalistas e relações públicas a melhor compreender as dinâmicas temporais do sentimento expresso em redes sociais, e de Satyanarayan e Heer (2014), que criaram uma ferramenta de visualização para uso em narrativas jornalísticas. As ferramentas desenvolvidas no âmbito desta pesquisa se pretendem menos um gênero jornalístico e mais um instrumento de apuração jornalística, na medida em que possam auxiliar repórteres a perceber nas bases de dados padrões e tendências de outra forma invisíveis a “olho nu”. Noutras palavras, se acredita que visualizações de dados como as descritas abaixo possam servir para ampliar as habilidades profissionais dos repórteres, em especial no que tange à identificação de informação dotada de valor-notícia e no entendimento de fenômenos que possam auxiliar na compreensão de acontecimentos jornalísticos.

### **3 COLETA, PREPARAÇÃO E VISUALIZAÇÃO DOS DADOS**

A Copa do Mundo, realizada entre junho e julho, foi um dos principais temas de discussão no Brasil em 2015, alcançando grande popularidade no Twitter e outras redes sociais. A previsão de uma grande quantidade de dados sociais relacionados a esse evento nos motivou para selecioná-los como um tema para a análise. Para fazer esta análise, decidimos começar com um estudo de caso sobre o jogo "Alemanha 7 x 1 Brasil" e dividimos o trabalho em várias etapas que estão descritas a seguir e ilustradas na figura 1.



Figura 1 – Etapas para elaboração deste trabalho

Os *tweets* utilizados na pesquisa foram coletados entre 30 de maio e 13 de Julho de 2014 e armazenados em um banco de dados. O processo de captura dos *tweets* foi baseado em palavras chave como “copa”, “vencedor”, “turistas”, “hexa”, entre outras. Para cada *tweet* foram armazenadas informações como o seu identificador, número de *retweets*, identificador do usuário que enviou o *tweet* e *hashtags*. O conjunto de dados coletados foi estruturado num *corpus* chamado de WorldCupBrazil2014. Devido à repercussão do resultado do jogo “Alemanha x Brasil”, ocorrido no dia 8 de julho de 2015, focamos nosso trabalho em mensagens publicadas durante este jogo. Para isso, extraímos um subconjunto do *corpus* WorldCupBrazil2014 para formar o *corpus* 7x1PT, com 2.728 de *tweets* em português. Este *corpus* possui duas informações adicionais para cada *tweet*: texto da mensagem pré-processado e polaridade. A mensagem pré-processada corresponde à mensagem original sem *hashtags*, *links* e abreviações – por exemplo, a abreviação “oq” foi substituído por “o que”).

A anotação de polaridade (negativo, neutro ou positivo) foi baseada no sentimento que as pessoas expressavam em relação à seleção brasileira e foi realizada por dois anotadores humanos. Nós anotamos manualmente cada *tweet* como negativo, neutro ou positivo. Nós consideramos como positivos os *tweets* que elogiavam ou encorajavam a seleção brasileira, e

como negativos aqueles que criticavam ou expressavam sentimentos pessimistas quanto ao desempenho do time brasileiro. As demais mensagens foram classificadas como neutras. A Tabela 1 mostra a distribuição atual de polaridade dos *tweets*.

Tabela 1 . Distribuição de Polaridade do *corpus* 7x1

<b>Polaridade</b>	<b># Tweets (%)</b>
Negativo	800 (29 %)
Neutro	1,771 (65%)
Positivo	157 (6%)

Técnicas de visualização são capazes de trazer mais informações, bem como auxiliar a tomada de decisão em diferentes níveis e contextos. No contexto do jornalismo, estas técnicas são usadas para traduzir enormes quantidades de dados em uma visualização que auxilia a compreensão de questões complexas. Como as visualizações apresentadas a seguir sugerem, e considerando a anotação da polaridade, os gráficos são capazes de refletir mudanças no humor e opinião.

Algumas das *hashtags* mais utilizadas pelos brasileiros sobre a Copa do Mundo antes do evento acontecer foram: #vaitercopa, #vaitercopasim e #naovaitercopa. Elas eram frequentemente utilizadas com um viés político, enfatizando a opinião do usuário sobre os preparativos para o evento esportivo (ZAGO, 2014). Durante o processo de coleta de dados foram criados alguns gráficos simples relacionados com a visualização dessas *hashtags* ao longo do dia da partida (figura 2).

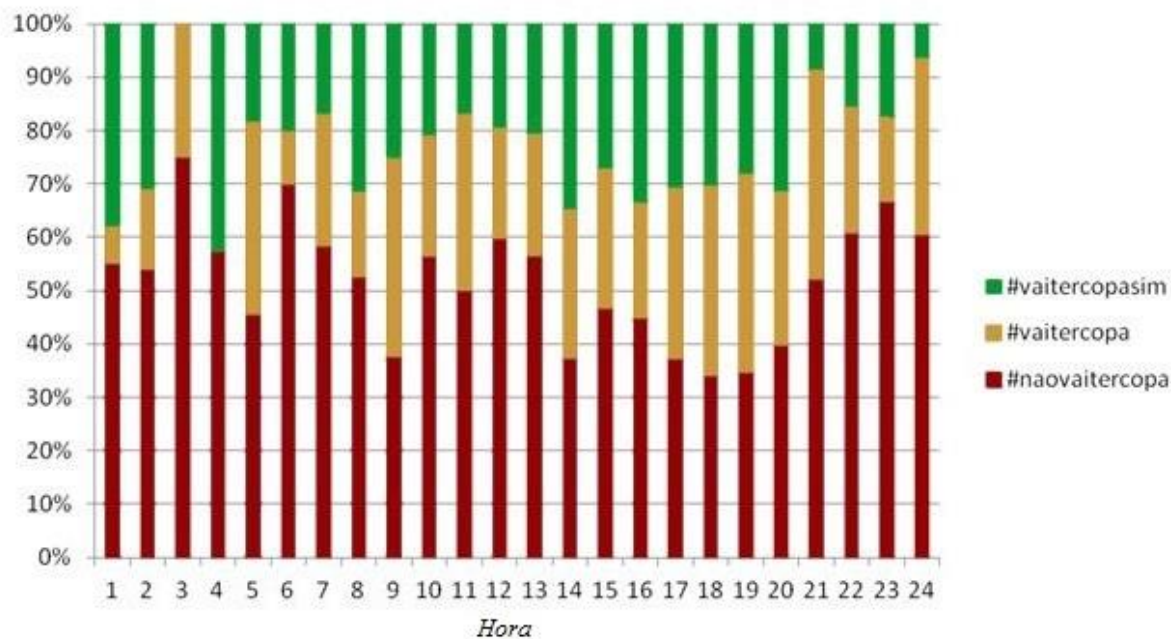


Figura 2 – Proporção de *tweets* por hora do dia 8 de julho com *hashtags* populares.

Além disso, havia interesse em saber quais eram as dez *hashtags* mais utilizadas pelos brasileiros no *Twitter* durante o dia 8 de julho, o que é mostrado na figura 3. Na figura 4, se pode visualizar o sentimento geral no *Twitter* a respeito do desempenho dos jogadores atacantes da seleção brasileira, a partir de sua classificação manual como positivo, neutro ou negativo. O eixo X mostra o placar do jogo ao longo do tempo, incluindo o intervalo entre o primeiro e o segundo tempo, e o eixo Y mostra o total de *tweets*.

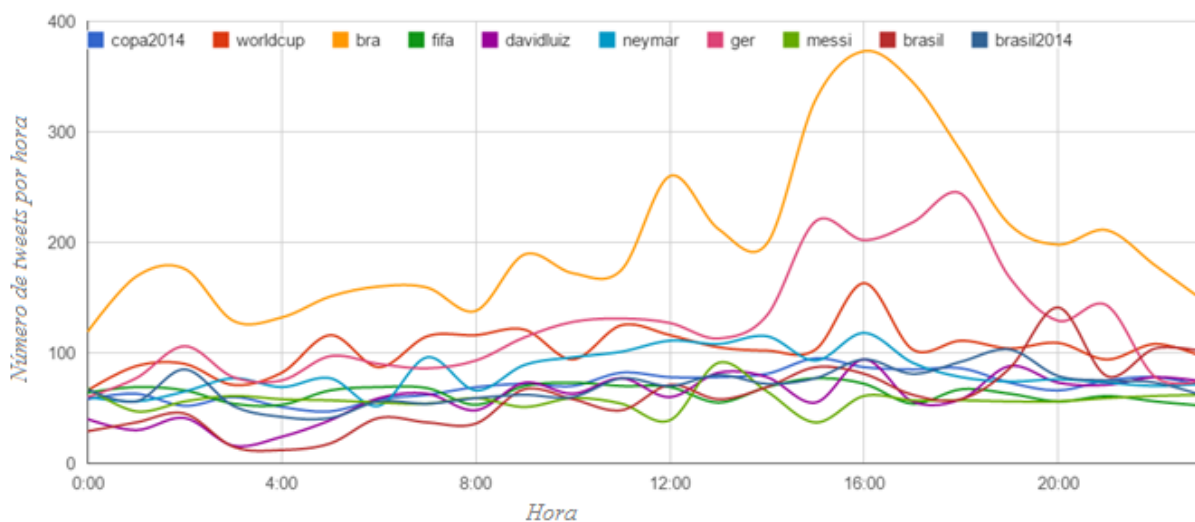


Figura 3 – *Hashtags* mais frequentes no dia 8 de julho, por hora.

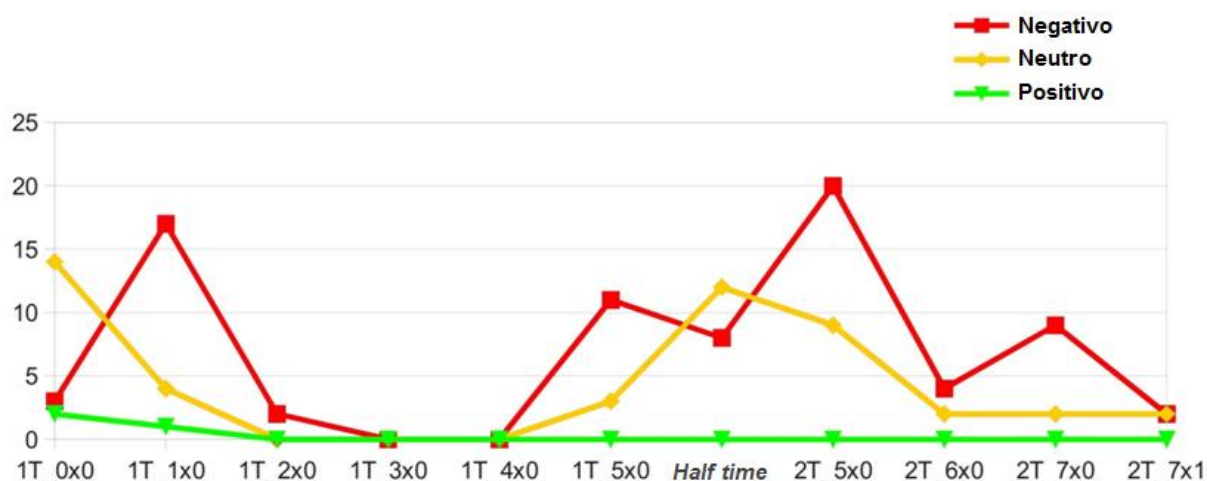


Figura 4 – Total de *tweets* sobre os jogadores atacantes da seleção brasileira classificados como positivo, neutro e negativo, considerando o placar no primeiro e no segundo tempo.

Além do uso dos gráficos estáticos em barras e linhas apresentados acima, mais comuns no jornalismo, também desenvolvemos um protótipo para visualização interativa da análise de sentimento obtida com a anotação de polaridade. As figuras 5 e 6 mostram duas visualizações manipuláveis de análise de sentimento sobre o jogo da semifinal, em que os *tweets* foram classificados como positivo, neutro ou negativo. Na figura 5, a visualização corresponde à média dos *tweets* classificadas pelos três tipos de sentimentos. O usuário pode interagir com a linha do tempo que aparece abaixo do gráfico, ampliando ou diminuindo o período de tempo a ser mostrado no gráfico através da seleção de um intervalo. Na figura 6, a visualização mostra a quantidade de *tweets* por tipo de sentimento durante o jogo, e a interação ocorre da mesma forma.

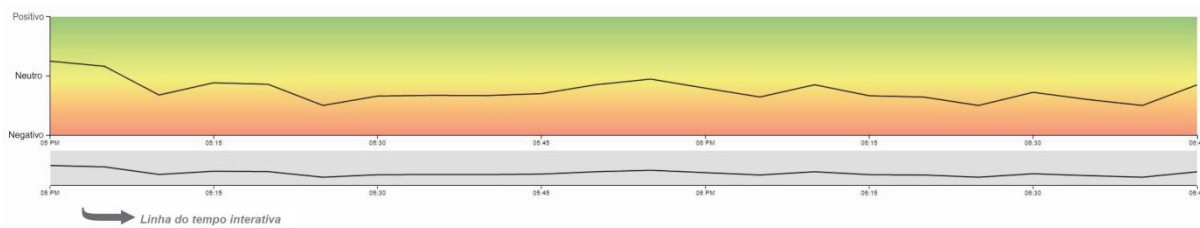


Figura 5 – Visualização interativa do jogo Alemanha 7 x 1 Brasil: média dos *tweets* classificados como positivo, neutro e negativo.

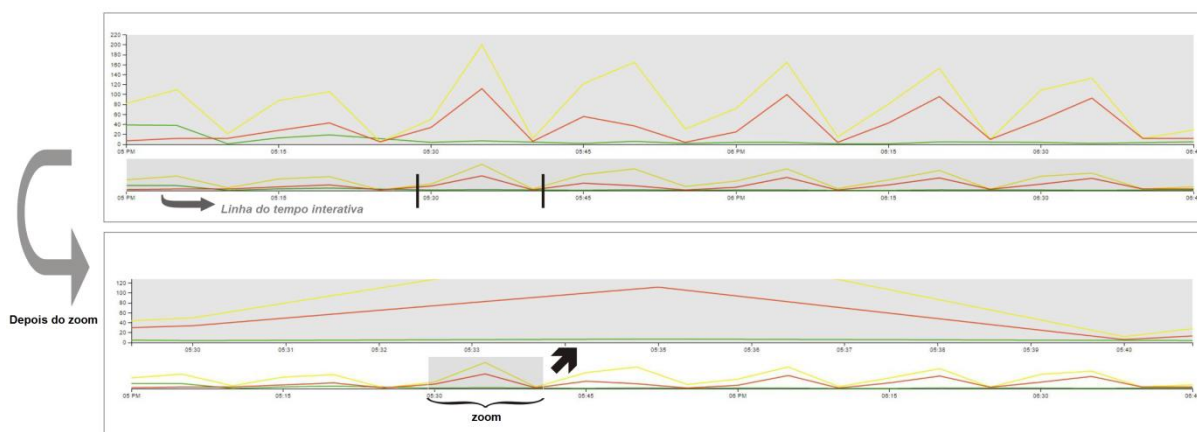


Figura 6 – Visualização interativa do jogo Alemanha 7 x 1 Brasil: exemplo da seleção de um pequeno período de tempo.

#### 4 CONVIDANDO OLHARES EXTERNOS

Após a coleta, análise e criação de visualizações gráficas a partir dos *corpora*, foi conduzido um grupo focal para entender como os protótipos poderiam auxiliar – ou não – na compreensão de dinâmicas sociais complexas em redes sociais na Internet. Buscamos a opinião dos participantes sobre as visualizações criadas com os dados coletados ao longo da partida Brasil x Alemanha na Copa do Mundo de 2014. A principal questão a nortear o grupo focal foi: quais informações podem ser obtidas a partir das visualizações apresentadas?

Na escolha desta metodologia tomamos como referência a definição de Morgan (1988), para quem o principal objetivo de um grupo focal é obter intuições sobre uma área definida de interesse num ambiente próximo do natural. Os dados resultantes são qualitativos e consistem em opiniões, ideias, experiências e motivações para determinados comportamentos. Tremblay et al. (2010), por outro lado, destacam a capacidade de grupos focais de permitir o surgimento de ideias ou opiniões geralmente difíceis de ser descobertos em entrevistas individuais e oferecer aos pesquisadores a oportunidade de coletar dados ricos em várias etapas de um projeto.

Com este objetivo, o grupo focal foi realizado em uma sala privada na universidade, com duração de cerca de 2 horas em que 4 visualizações de dados foram apresentadas aos participantes por um mediador. Quatro indivíduos do sexo feminino participaram do estudo, todos jornalistas e estudantes de pós-graduação, com uma faixa de idade entre 26 e 30 anos.

Todas as participantes do grupo focal administram contas no Twitter e muitas vezes usam este meio como fonte de informação, mas apenas metade usa o Twitter para divulgar informações. A maioria dos participantes não costuma interagir com outros perfis de amigos ou desconhecidos no Twitter. Todas elas afirmaram não usar qualquer aplicativo para análise de dados em redes sociais. Quando perguntadas sobre os benefícios do uso de redes sociais em atividades jornalísticas, as participantes apontaram vantagens como: encontrar informações complementares; servir de fonte de pautas para notícias (agendamento); buscar por fatos e casos; facilmente disseminar informações e obter suprimentos; ou monitorar o impacto de acontecimentos.

Durante o grupo focal, após uma breve explicação de cada visualização, o mediador encorajou o grupo a falar sobre suas impressões, sem limite de tempo, e a discutir abertamente as suas percepções e experiências, bem como as qualidades pertinentes das visualizações uns com os outros. Para as duas visualizações interativas, o mediador convidou os participantes a ir ao computador e interagir com a interface/visualização. O grupo focal foi gravado utilizando duas câmaras de vídeo e um dispositivo de gravação de áudio. As transcrições dessas gravações são a principal fonte da análise que se segue.

Do ponto de vista de comunicação, os resultados do grupo focal sugerem que a visualização de dados, em geral, oferece a possibilidade de identificar tendências, padrões e anomalias nas bases de dados que não seriam visíveis por outros meios. No caso da figura 3, por exemplo, os membros do grupo identificaram na curva ascendente e no pico de menções da palavra-chave "BRA" a expectativa dos fãs brasileiros para o jogo. Embora um usuário acompanhando o fluxo de dados no Twitter através de sua linha do tempo talvez fosse capaz de ver um grande número de referências a "BRA" na ocasião da partida, dependendo das preferências e interesses dos perfis que seguisse, ele dificilmente seria capaz de determinar as diferenças de proporção entre menções deste e de outros termos. A apresentação destas proporções como linhas em um gráfico torna as proximidades e afastamentos mais claras.

A apresentação gráfica dos dados também permite identificar padrões não-intuitivos. Analisando a mesma figura 3, a Participante 1 perguntou, por exemplo, por que o nome do jogador Messi, da seleção argentina, estava entre as palavras mais citadas num dia em que não havia jogo de sua equipe. Não está no escopo desta pesquisa investigar as razões para um pico

de menções a Messi naquele dia, mas a surpresa da Participante 1 sugere a identificação de um fenômeno inesperado, o que poderia, no contexto do jornalismo, gerar uma pauta para uma matéria. O grupo especulou que as referências a Messi poderiam ser derivadas de uma publicação de uma notícia importante sobre o jogador, considerando as funções do Twitter como um espaço de reverberação para informações publicadas na imprensa durante a Copa do Mundo de 2014, como afirma Zago (2014). Da mesma forma, observando a figura 5, a respeito da análise de sentimento durante o jogo, a Participante 2 comentou: “O neutro e positivo estão bem dentro do esperado, esses dois grandes picos negativos é que eu não sei, só olhando para o gráfico”. Isto é, a visualização apresentava anormalidades que poderiam ser investigadas por diversas estratégias: uma avaliação do vídeo da partida e a análise do conteúdo dos *tweets* considerados negativos, por exemplo.

Nesse sentido, as visualizações de dados são uma ferramenta para identificar padrões desviantes. Para os jornalistas, esses padrões são importantes, pois geralmente indicam eventos com valor noticioso. Parafraseando a fórmula tradicional: os gráficos podem muitas vezes mostrar os momentos em que “o homem mordeu o cachorro”, isto é, quando as expectativas sobre o andamento da vida diária foram invertidos. A partir da identificação de tais fenômenos, os repórteres podem procurar fontes humanas e documentos que expliquem seu contexto, a fim de informar o público sobre esse evento. As visualizações são, portanto, ferramentas auxiliares à produção de notícias.

A figura 6 ofereceu às participantes uma visualização manipulável, na qual se pode alternar entre a média dos sentimentos medidos nos *tweets* sobre o jogo e linhas separadas mostrando a proporção de sentimentos positivos, negativos e neutros. Esta segunda alternativa foi considerada a mais útil pelo grupo, pois permite acompanhar melhor a variação no humor dos usuários do Twitter ao longo da partida com maior precisão. Na visualização agregada, há pouca variação ao longo do tempo, devido ao efeito de cancelamento entre os *tweets* negativos e positivos em meio a uma grande quantidade de *tweets* classificados como neutros. Para o leitor, portanto, a média pode dar a impressão de que houve pouca variação na atitude dos espectadores ao longo do jogo. A visualização de cada sentimento em linhas separadas desfez esta impressão entre as participantes do grupo focal, pois permitiu identificar picos de negatividade ou positividade.



Uma das participantes, com efeito, declarou que, se houvesse visto as linhas de sentimento em separado antes da visualização da média, não poderia imaginar que a resultante de sua agregação seria neutra, mas concluiria que a atitude geral durante toda a partida fora negativa. A discrepância entre as percepções geradas pelas duas opções de visualização dos mesmos dados confirma a importância de se levar em conta, na criação deste tipo de produto, as habilidades e limitações do sistema visual humano para a identificação de padrões, como apontam Heer, Bostock e Ogievetsky (2010). Por outro lado, a possibilidade de interação emerge como uma característica que facilita a compreensão de grandes volumes de dados, porque permite ao leitor observar e comparar diferentes tipos de padrões e, desta forma, observar o mesmo banco de dados sob diferentes perspectivas.

Entre as sugestões de aprimoramentos possíveis para as visualizações apresentadas, o grupo recomendou ampliar as possibilidades de manipulação, como, por exemplo, mostrar o número absoluto de *tweets* quando se passa o ponteiro do mouse sobre as linhas e sobrepor as linhas a um vídeo da partida, para facilitar a análise dos picos de positividade e negatividade em relação aos movimentos do jogo, como gols e faltas. Outra sugestão fornecida pelas participantes foi criar uma conexão entre os eixos cronológicos das visualizações que os oferecem e notícias publicadas na imprensa durante a Copa do Mundo de 2014, para que se possa estabelecer relações entre o cotidiano e a dinâmica social no Twitter.

## 5 CONSIDERAÇÕES FINAIS

A questão norteadora da pesquisa, relatada aqui em seus resultados preliminares, é: *Como a visualização de dados pode favorecer a obtenção de informação que seria muito difícil de perceber e adquirir de outra forma?*. Para encaminhar uma resposta a esta pergunta, foram coletados e processados textos publicados no Twitter durante a Copa do Mundo de 2014. A partir deste *corpus*, foram geradas visualizações dos sentimentos dos usuários brasileiros da rede social durante a partida semifinal entre Brasil e Alemanha, bem como das palavras-chave (*hashtags*) mais usadas: #vatercopa, #vaitercopasim e #nãovaitercopa. Para realizar uma avaliação inicial dos protótipos desenvolvidos em parceria por pesquisadores das áreas de Informática e Comunicação, foi realizado um grupo focal com alunos do Programa de Pós-Graduação em Comunicação Social da PUCRS.

Os resultados não oferecem respostas definitivas a respeito da adequação dos protótipos à extração de informação a partir de fluxos de dados sobre o comportamento dos usuários de redes sociais na Internet às rotinas produtivas das redações jornalísticas. No entanto, o grupo focal, em que pese as limitações impostas pela pouca quantidade de participantes e sua ausência de relação profissional com o jornalismo no momento da realização do estudo, foi útil no sentido de levantar problemas e questões a serem levados em conta no desenvolvimento de novos protótipos e nas próximas fases da pesquisa, nas quais jornalistas profissionais serão os sujeitos e o método será a entrevista semi-estruturada. De qualquer forma, esta linha de investigação parece promissora tanto para o desenvolvimento de tecnologias aplicadas ao jornalismo, como para a compreensão de como as visualizações de dados podem contribuir na etapa de apuração de reportagens.

## 6 REFERÊNCIAS

ANDERSON, C. W.. Notes Towards an Analysis of Computational Journalism. **HIIG Discussion Paper Series**, n. 1, 2012.

BARBOSA, G. A. et al.. Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment. In: **CHI'12 Extended Abstracts on Human Factors in Computing Systems**, 2012, ACM, p. 2621–2626.

CAIRO, Alberto. **The functional art**: an introduction do information graphics and visualization. Berkeley: New Riders, 2013.

COHEN, S.; HAMILTON, J. T.; TURNER, F.. Computational journalism. **Communications of the ACM**, v. 54, n.. 10, 2011, p. 66-71.

CONOVER, M. D. et al.. Predicting the political alignment of twitter users. In Privacy, security, risk and trust (passat). In: **IEEE Third International Conference on Social Computing – Socialcom**, 2011, IEEE, p.192-199.

DE PABLOS, J. M.. **Infoperiodismo**: el periodista como creador de la infografía. Madrid: Síntesis, 1999.

DIAKOPOULOS, N.. **Cultivating the landscape of innovation in computational journalism**. City University of New York, abr. 2012. Disponível em: [http://cdn.journalism.cuny.edu/blogs.dir/418/files/2012/04/diakopoulos\\_whitepaper\\_systematic\\_innovation.pdf](http://cdn.journalism.cuny.edu/blogs.dir/418/files/2012/04/diakopoulos_whitepaper_systematic_innovation.pdf). Acesso em: 10 mai. 2013.

DIAKOPOULOS, N.; SHAMMA, D. A.. Characterizing debate performance via aggregated twitter sentiment. In: **Proceedings of the SIGCHI Conference on Human Factors in computing Systems**, Nova York, 2010, ACM, p. 1195–1198.

ELIAS, M.; BEZERIANOS, A.. Exploration views: understanding dashboard creation and customization for visualization novices. In: **Human-Computer Interaction–INTERACT 2011**, 2011, Springer, p.274-291.

ELIAS, M.; AUFAURE, M.A.; BEZERIANOS, A.. Storytelling in visual analytics tools for business intelligence. In: **Human-Computer Interaction–INTERACT 2013**, 2013, Springer, p.280-297.

HAMILTON, James; TURNER, Fred. **Accountability Through Algorithm: Developing the Field of Computational Journalism**. Stanford, 2009. Disponível em: <http://www.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>. Acesso em: 9 dez. 2013.

HANNA, A. et al.. Partisan alignments and political polarization online: A computational approach to understanding the french and us presidential elections. In: **Proceedings of the 2nd workshop on Politics, elections and data**, 2013. ACM.

HEER, J.; BOSTOCK, M.; OGIEVETSKY, V.. A tour through the visualization zoo. **Communications of the ACM**, v.53, n.6, 2010, p.59–67.

HEER, J.; VIEGAS, F. B.; WATTENBERG, M.. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In: **Proceedings of the SIGCHI conference on Human factors in computing systems**, 2007, ACM, p. 1029–1038.

GRAY, J.; CHAMBERS, L.; BOUNEGRU, L. (orgs.). **Manual de jornalismo de dados: como os jornalistas podem usar dados para melhorar suas reportagens**. São Paulo: Abraji/EJC, 2013.

LEONHARDT, D.. Navigate News With The Upshot. **The Upshot**, 22 abr. 2014. Disponível em: <http://www.nytimes.com/2014/04/23/upshot/navigate-news-with-the-upshot.html>. Acesso: 22 mai. 2014.

LEWIS, Seth; USHER, Nikki. Open source and journalism: toward new frameworks for imagining news innovation. **Media, culture & society**, v. 35, n. 5, p. 602-619, 2013.

LIMA JR., Walter T. Jornalismo computacional em função da “Era do Big Data”. **Revista Líbero**, v. 14, n. 28, p. 45-52, 2011.

MATZAT, Lorenz. Datenjournalismus. In: **Dossier Open Data**. Bonn: Bundeszentrale für politische Bildung, 2011.

MCCHESENEY, Robert. **Digital disconnect: how capitalism is turning the Internet against democracy**. Nova York: The New Press, 2013. [Ebook]

MEYER, P.. **Precision Journalism: a reporter’s introduction to social science methods**. Bloomington: Indiana University Press, 1973.

MORGAN, D. L.. **Focus groups as qualitative research**. Nova York: Sage, 1988.

PARASIE, S.; DAGIRAL, É.. Des journalistes enfin libérés de leurs sources? Promesse et réalité du “journalisme de données”. **Sur le journalisme**, v. 2, n. 1, 2013.

RECUERO, R.. **Redes sociais na Internet**. Porto Alegre: Sulina, 2009.

SAHAMI S. A. et al.. Real-time nonverbal opinion sharing through mobile phones during sports events. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, 2011, ACM, p.307-310.

SANTAELLA, Lucia. **Navegar no ciberespaço: o perfil cognitivo do leitor imersivo**. São Paulo: Paulus, 2004.

SATYANARAYAN, A.; HEER, J.. Authoring narrative visualizations with ellipsis. **Computer Graphics Forum**, v. 33, 2014, p. 361–370.

TEIXEIRA, T.. A presença da infografia no jornalismo brasileiro: proposta de tipologia e classificação como gênero jornalístico a partir de um estudo de caso. **Revista Fronteira – Estudos Midiáticos**, v.9, n.2, 2007.

\_\_\_\_\_. O futuro do presente: os desafios da Infografia jornalística. **Ícone**, v.1, n.2, dez/2009.

TREMBLAY, M.; HEVNER, A.;BERNDT, D.. Focus groups for artifact refinement and evaluation in design research. **Communications of the ASI**, v.26, n.1, 2010.

VALERO SANCHO, J.L.. **La infografía: técnicas, análisis y usos periodísticos**. Bellaterra: UAB, 2001.

ZAGO, Gabriela. Enquanto você lia este título, a Alemanha fez mais um gol: a Copa da zoeira nos sites de rede social. **Estudos em Jornalismo e Mídia**, v.11, n.2, 2014, p.415–423.