7th CIDI
7th Information Design International Conference
7th CONGIC
7th Information Design Student Conference
Blucher Design Proceedings
Setembro, 2015 – num. 2, vol.2
proceedings.blucher.com.br

Visualização sobre dados de redes sociais
Visualization on social networks data

Carolina Riff, Amanda Pereira, Barbara Castro, Doris Kosminsky

visualização, dados, redes sociaisl

Tendo em vista a recente expansão das redes sociais, este artigo aborda o desenvolvimento de um processo semi-automático para o design de visualizações, empregando dados extraídos do Twitter, e seguindo o método desenvolvido por Ben Fry (2007). A partir da repercussão negativa do pronunciamento da Presidente Dilma Roussef no dia 8 de Março de 2015 no Twitter, obtivemos as quantidades das principais hashtags utilizadas pela população e sua disseminação ao longo do tempo. Com esses dados criamos uma visualização que nos permitiu explorar os acontecimentos daquele dia e seu impacto na rede social.

visualization, data, social network

In view of the recent expansion of social networks, this article discusses the development of a semi-automatic process for creating visualizations over extracted data from Twitter, relied on Ben Fry's method (2007). We observed the negative impact of President Dilma Rousseff's pronouncement on March 8, 2015 and obtained the amounts of the hashtags posted on Twitter and their spread over time. With these data we created a visualization that allowed to explore the events of that day and its impact on the social network.

1 Introdução

O crescimento das redes sociais nos últimos anos vem tornando estas estruturas alvo de interesse crescente. Produtoras de milhões de dados diários, redes como Twitter, Facebook, Tumblr, dentre outras, disseminam conteúdos que podem rodar o mundo em questão de minutos, compartilhados muitas vezes por milhares de usuários.

O aumento da popularidade destas plataformas como difusoras de informação as torna foco de estudos acerca de sua capacidade como ferramenta social formadora de opiniões e transmissora de ideias.

Levando em consideração a utilização de dados obtidos nas redes para a construção de visualizações de dados e informações (que é cada vez mais crescente) e o papel do designer na criação de representações de dados e informações, neste artigo abordaremos a produção de uma visualização utilizando dados extraídos do Twitter sobre a rejeição da Presidente da República, Sra. Dilma Roussef. Os dados foram obtidos no dia 8 de março de 2015, quando houve um pronunciamento pelas redes de rádio e televisão, onde a Presidente parabenizou as mulheres pelo seu dia e abordou a atual crise econômica, medidas políticas, e a criação de novas leis. Nos apoiaremos no método desenvolvido por Ben Fry (2007) para a criação da

C. Riff, A. Pereira, B. Castro, D. Kosminsky. 2015. Visualização sobre dados de redes sociais. In: C. G. Spinillo; L. M. Fadel; V. T. Souto; T. B. P. Silva & R. J. Camara (Eds). **Anais [Pôster] do 7º Congresso Internacional de Design da Informação/Proceedings [Poster] of the 7th Information Design International Conference | CIDI 2015** [Blucher Design Proceedings, num.2, vol.2]. São Paulo: Blucher, 2015. ISSN 2318-6968, DOI 10.5151/designpro-CIDI2015-congic_21

visualização de dados sobre a rejeição da Presidente da República a partir de dados extraídos do Twitter.

2 Problema

A visualização de dados é uma área interdisciplinar que tem como objetivo revelar, e explorar a estrutura por trás de grandes conjuntos de dados, representando-os de forma visual.

Existem diversas metodologias para a criação de visualizações. Neste projeto, optamos pela utilização da metodologia desenvolvida por Ben Fry, um conjunto de sete passos que consistem em: extrair os dados, analisa-los, filtrá-los de acordo com o que se deseja, minerá-los, representa-los de alguma forma visual, refinar essa representação e adicionar elementos de interação desses dados.

Tal metodologia representa bem o trabalho do designer atualmente uma vez que há uma necessidade de envolvimento no processo como um todo, delineando um perfil profissional que abrange mais funções. O que apresenta o oposto dos processos que ocorriam em meados do século XX como, por exemplo, o *Isotype,* sistema de pictogramas projetado por Otto Neurath para comunicar informações de forma objetiva através do emprego das linguagens verbal e visual. Os artistas gráficos de então eram responsáveis apenas por desenhar os elementos que eram passados já extraídos (normalmente por profissionais de outras áreas) e configurados (pelos transformadores), cabendo aos artistas apenas finalizar o processo.

Figura 1: Funcionamento do sistema ISOTYPE

3 Método

Seleção da Rede Social

Na contagem geral, os questionários aplicados com a sequência 1 e 3 apresentaram diferença entre a recepção da imagem sem e com cor. A imagem PB teve menos acertos que a imagem colorida nos dois casos (figura 7 e 8). A fala das crianças durante o exercício indica que na ausência de cor as crianças se guiaram mais pela forma do alimento, enquanto que na colorida pela cor: "*É um limão claro, porque tem cor de limão ...* (pausa para observação da imagem) *... e porque tem a forma de um limão*" (criança de 7 anos).

Extração

Tomando como base a metodologia de Ben Fry (2007) e escolhendo como objeto da visualização as manifestações de opiniões referentes às questões políticas brasileiras (dentro do Twitter) iniciou-se o processo de busca de ferramentas que fossem capazes de extrair esses dados baseadas nas *hashtags*, palavras-chaves utilizadas por esta rede social e que indicam a popularidade de um assunto.

Após o período de pesquisas, que incluiu testes com diferentes ferramentas, optou-se pela utilização de um programa escrito em *Python (*uma linguagem de programação orientada a objetos) e que, fazendo uso de uma aplicação do Twitter (recurso disponibilizado pela própria rede social) extrairia tuítes contendo os termos selecionados e os converteria para um formato .csv (*comma-separated values* ou um formato de tabelas que tem suas colunas delimitadas por vírgulas), de modo que pudesse ser lido por outros programas.

As seguintes *hashtags* foram observadas como as mais populares no Twitter no dia do pronunciamento da presidente Dilma: #ForaPT, #ForaDilma, #panelaço, #VaiaaDilma, #VemPraRua e #ImpeachmentDilma. Todos os tuítes relativos a essas *hashtags* foram extraídos, num intervalo de 12 horas começando cerca de 7 horas antes do pronunciamento televisionado, que se iniciou às 20:40.

Quantificação

Seguindo a metodologia, após a extração dos dados, foi necessária a contabilização destes para que estes pudessem ser devidamente analisados e filtrados. Neste caso, optamos pela contabilização dos tuítes a cada hora, para observar as oscilações em sua quantidade com o passar do tempo, dentro do recorte que fizemos (filtragem), até o seu momento de pico que se deu após as 20h. Para contabilizar, utilizamos os valores das planilhas de cada hashtag, somando os valores de cada hora. Os números dessa contabilização foram guardados em um arquivo csv.

Figura 2: Tabela com os valores de cada Hashtag

Horas	#ForaDilma	#Panelaço	#VemPraRua	#ForaPT	#Impeachment	#VaiaaDilma
14h-15h	24	20	30	14	18	1
15h-16h	77	21	26	22	8	1
16h-17h	221	4	95	54	13	1
17h-18h	97	22	65	47	20	7
18h-19h	89	64	47	23	5	19
19h-20h	128	14	112	51	30	76
20h-20:40h	171	58	336	53	41	118
20:40h-21h	2828	143	177	600	352	1197
21h-22h	4474	552	462	750	338	7270
22h-23h	1802	263	174	376	210	14148
23h-00h	1980	211	617	398	256	12337
00h-1h	706	56	160	195	164	7179
1h-2h	554	18	76	244	382	3403
2h-3h	179	10	15	80	86	991
3h-4h	70	4	14	20	22	546

Desenvolvimento da Visualização

O objetivo nesse projeto foi criar uma visualização que permitisse acompanhar as manifestações da população sobre o pronunciamento da Presidente a partir da dinâmica das relações da quantidade de tuítes para cada *hashtag* com os acontecimentos e com outras *hashtags* (uma vez que o Twitter permite que mais de uma *hashtag* seja adicionada a um tuíte) e a forma como o total de tuítes e retuítes se relacionavam de acordo com o passar das horas.

A partir da análise dos dados, que revelou quantidades muito maiores de tuítes nos horários de pico em relação aos outros horários, concluímos que a característica mais forte dos dados obtidos era o fenômeno de explosão de informações que observamos e então decidimos que este padrão dos dados deveria estar em destaque.

Para tal ponderamos acerca de opções de layout que fossem de fácil entendimento e que, ao mesmo tempo, favorecessem a possibilidade de explorar esta dinâmica de expressão da população. Foi decidido então que a visualização seria dividida em duas partes, uma para representar as diversas hashtags e outra para indicar o número total de tuítes ao longo do tempo. Optamos pelo uso de um gráfico de linhas, por este permitir a observação das *hashtags* em conjunto e individualmente além de ser um formato já assimilado, objetivo e possível de ser produzido de forma mais automática. E gráficos de círculos cuja escolha se deu devido ao fato dos círculos favorecerem a leitura e ressaltarem o aspecto curioso do número de tuítes ser maior que o número de retuítes em quase todo o período analisado. Neles, a área dos círculos representa os valores totais de tuítes e retuítes.

Considerações técnicas

Visando uma automação de parte do processo, a primeira parte, composta pelas linhas de cores variadas, foi realizada com a utilização de *Processing*, linguagem de programação de código aberto, voltada para artistas e designers que também permite a leitura de arquivos .csv.

Utilizando os dados numéricos guardados em uma tabela, foi desenvolvido um algoritmo em *Processing* que lesse os valores da tabela e criasse um gráfico de linhas a partir do número de tuítes de cada *hashtag* em relação ao tempo a cada hora. Nele, cada linha representava uma *hashtag* diferente, e sua alternância na vertical a forma como seus valores se modificaram, subindo e descendo. O tempo foi representado no eixo horizontal, dividido por horas. A escala no eixo vertical tinha seus valores dados em escala logarítmica, uma vez que os números máximos e mínimos de tuítes eram muito extremos, tornando inviável a utilização de uma escala linear.

A execução dos círculos que representam os valores totais dos tuítes e retuítes, no entanto, se deu de forma manual. Utilizando os valores totais obtidos, foi definida uma equação matemática para calcular a área referente a cada um dos círculos responsáveis por representar os valores totais de cada hora para que se pudesse, a partir da mesma, extrair o valor do diâmetro, necessário para a realização do desenho. Os círculos posicionados alinhados com a hora correspondente.

Após o processo de estruturação, o aspecto gráfico de ambas as visualizações foi refinado, através da apresentação das visualizações para a equipe do laboratório de modo a garantir a utilização de cores e corpo tipográfico adequados.

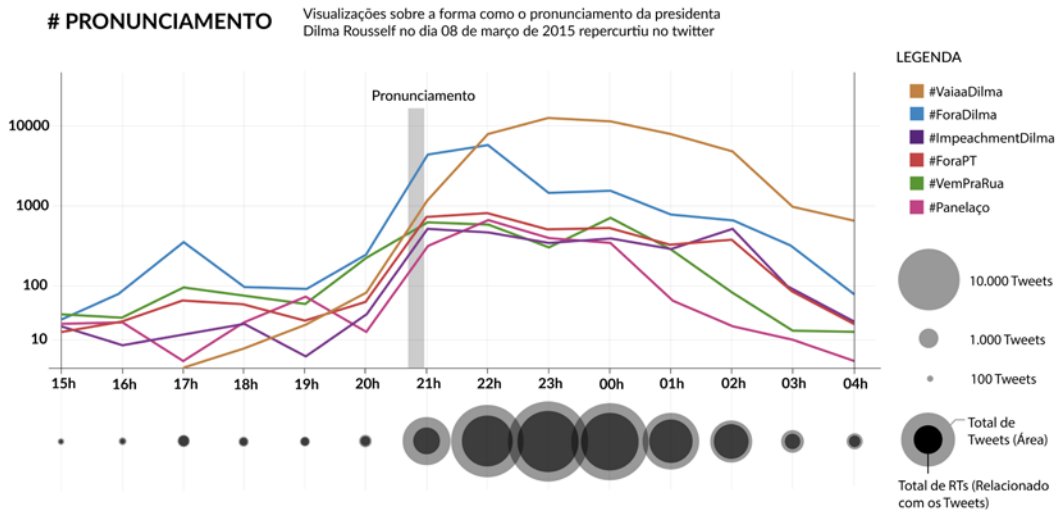
4 Resultados

Ao final do processo, concebemos uma visualização de dados que permitiu a observação e análise do comportamento das *hashtags* negativas em relação à presidente. Embora a maior parte das *hashtags* tenham tido um comportamento similar, aumentando a partir do pronunciamento e gradativamente diminuindo com o passar do tempo, algumas *hashtags* tiveram ocorrência muito maior que outras. Pudemos observar o surgimento da hashtag #VaiaDilma às 17h que atinge seu auge entre as 22h e 23h em que teve 14148 ocorrências enquanto #ForaDilma contabilizou apenas 1802 ocorrências. O gráfico tornou visível o fenômeno de viralização baseado na manifestação espontânea em massa da população, mostrando como as redes sociais estão conectadas em tempo real com os eventos, o que as faz poderosas ferramentas na disseminação de informações.

Neste projeto, desenvolvemos um método passo a passo para a criação de visualizações com dados extraídos de redes sociais que pode ser utilizado para outras pesquisas

sobre informações veiculadas no Twitter. Esse método compreende as diversas etapas, deste a extração até a representação visual, associando a programação ao design ao complementar a visualização produzida por meio de códigos com ajustes posteriores em softwares de manipulação gráfica.

Figura 3: Visualização sobre o Pronunciamento



5 Conclusão

Considerando a expansão das redes sociais e o consequente aumento na produção de dados e informações, a visualização de dados coloca-se como uma ferramenta poderosa para a análise e compreensão da forma como as informações se comportam nesse ambiente e suas consequências para o ambiente social.

Neste projeto, aplicamos o método de produção de visualizações de dados criado por Ben Fry no desenvolvimento de um processo que nos conduziu à criação de uma visualização a partir de tuítes em oposição à Presidente Dilma no dia de seu pronunciamento televisionado. Acreditamos que as diversas soluções que desenvolvemos para cada uma dessas etapas poderão ser aplicadas em outros projetos de visualização de dados que se proponham a empregar dados do Twitter.

Agradecimentos

Agradecemos o apoio recebido: ao pesquisador Marlus Araújo, que nos ajudou no processo de extração dos dados; ao professor Cláudio Esperança (PESC-COPPE-UFRJ) pelo auxílio com as etapas que envolveram programação.

Referências

FRY, B. Visualizing Data. 1ª Edição. Estados Unidos: O'Reilly Media; 2008. Cap 1:The Seven Stages of Visualizing Data; 1-18

HARRYS, J. & KAMVAR S. 2006. We feel fine and searching the emotional. Diponivel em: http://wefeelfine.org/wefeelfine.pdf - acesso 09/3/15.

LUPTON, E. 1986, Reading Isotype. Design Issues. V.3, n.2: 47-58.

MANOVICH, L. 2010, What is Visualization? Em: Articles: What is visualization? Diposnivel em: http://manovich.net/index.php/projects/what-is-visualization - acesso 04/3/15.

YAU, N. 2013, Data Points: Visualization that means something. Estados Unidos: John Willey & Sons, Inc

Sobre os autores

Carolina Riff, bolsista de Iniciação Científica (PIBIC), Universidade Federal do Rio de Janeiro, Brasil, <carolinariff@gmail.com>

Amanda Pereira, bolsista de Iniciação Científica, Artística e Cultural (PIBIAC), Universidade Federal do Rio de Janeiro, Brasil <aspd.contato@gmail.com>

Barbara Castro, doutoranda, Universidade Federal do Rio de Janeiro, Brasil <barbarap.castro@yahoo.com.br>

Doris Kosminsky, PhD, Universidade Federal do Rio de Janeiro, Brasil <doriskos@gmail.com>