



Design da Informação e Análise de Redes Sociais: explorando formas de visualização de redes

Information Design and Social Network Analysis: exploring forms of network visualization

Gabriela da Silva Zago

design da informação, análise de redes sociais, visualização, sites de rede social, acontecimento

O artigo tem por objetivo discutir formas de visualização de dados obtidos a partir de sites de rede social. Para tanto, parte-se de um referencial teórico que aborda design da informação e análise de redes sociais, para a seguir propor um exercício de diferentes formas de visualização de um mesmo conjunto de dados. O conjunto de dados utilizado como recorte no presente trabalho envolve postagens feitas no Twitter sobre um acontecimento relacionado à Copa do Mundo de 2014. Os resultados apontam para diferentes percepções de uma mesma rede a partir da escolha de diferentes layouts e algoritmos de visualização.

information design, social network analysis, visualization, social network sites, news event

In this article, we aim to discuss forms of network visualization of data obtained from social network sites. In order to do so, we first present a theoretical background on information design and social network analysis, and, later, propose an exercise of different forms of visualization of the same dataset. The data set we used in this study encompasses posts made on Twitter regarding one event related to the 2014 FIFA World Cup. Results point to different perceptions of a single network based on the selection of different layouts and visualization algorithms.

1 Introdução

O artigo tem por objetivo discutir formas de visualização de dados obtidos a partir de sites de rede social. Para tanto, utiliza um referencial teórico que aborda design da informação e análise de redes sociais.

Para operacionalizar o estudo, faz-se um conjunto de explorações de formas de visualização de um mesmo conjunto de dados. Essas diferentes visualizações são utilizadas para discutir o papel dos algoritmos na percepção dos usuários sobre as informações mostradas.

O conjunto de dados selecionado para este trabalho faz parte de uma pesquisa que procurou abordar os aspectos comunicacionais das narrativas de usuários no Twitter em torno da Copa do Mundo de 2014 (Zago, 2014). Neste artigo, entretanto, o foco recai sobre as formas de visualização de redes.

Na primeira parte do trabalho, abordam-se aspectos gerais relacionados ao design da informação e à análise de redes sociais. Em um segundo momento, os procedimentos metodológicos são explicitados e as formas de visualização são apresentadas.

2 Design da informação

O Design da Informação é o ramo do Design dedicado a estudar as formas de distribuição da informação em um determinado suporte ou interface. A partir de um conteúdo já definido, o Design da Informação busca pensar em soluções visuais e informativas para dispô-lo da melhor forma para o usuário. Para tanto, alguns princípios são seguidos, em termos cognitivos, comunicacionais e estético-formais (O'Grady & O'Grady, 2008). Enquanto nos princípios cognitivos o foco é projetar um produto que possa ser compreendido pelo público-alvo pretendido, os princípios comunicacionais envolvem pensar a melhor forma de passar a mensagem, ao passo que os princípios estético-formais auxiliam na definição das melhores soluções visuais para uma determinada peça visual.

Dentre as soluções visuais que podem ser pensadas a partir do Design da Informação, estão os infográficos e as visualizações. Em termos gerais, um infográfico apresenta informação: o designer organiza os dados para contar uma história. Já uma visualização permite aos usuários explorar os dados e chegar a suas próprias histórias. Para Cairo (2012), a diferença entre infográfico e visualização não é muito clara. Muitos artefatos incorporam elementos de ambos. Para o autor, todo infográfico e toda visualização possuem dois elementos: apresentação e exploração, ou seja, eles apresentam informações, mas também facilitam a análise e exploração do que mostram. Nesse sentido, pensar a apresentação do artefato contribui para seu potencial de exploração pelos usuários.

3 Análise de Redes Sociais

Redes sociais se refere a um conjunto de atores e suas conexões (Recuero, 2009). Um grupo de pessoas relacionados entre si por algum motivo qualquer – mesmo local de trabalho, colegas de faculdade, etc. – pode ser considerado uma rede social. Com a internet, entretanto, surgem espaços propícios para a formação e manutenção de redes sociais, os sites de rede social (Ellison & Boyd, 2013). Embora as redes sociais não tenham surgido com a internet, elas se tornam mais visíveis nesses espaços, na medida em que é possível visualizar as relações entre os diferentes atores (representados por perfis) a partir dos rastros visíveis deixados pelos mesmos.

Para Watts (2010), o papel da internet foi tornar mensurável o imensurável – se antes era impossível colocar uma rede social inteira dentro de um laboratório para estudá-la, atualmente é possível analisar uma rede social inteira com um único computador.

A Análise de Redes Sociais fornece subsídios para que esse estudo aconteça, mediante métricas voltadas para o estudo de atores e suas conexões, permitindo determinar tanto os papéis dos nós quanto características das redes como um todo.

A Análise de Redes Sociais (ARS) tem sua origem na interseção entre a matemática, a informática e a sociologia. Ela envolve um conjunto de métricas e técnicas de pesquisa utilizado para descrever a relação entre nós (atores) e suas conexões (arestas) (Scott, 2013; Wasserman & Faust, 1994). O foco da ARS está no estudo da estrutura dos grupos sociais, buscando identificar as relações entre os atores dos grupos (Wasserman & Faust, 1994). A ARS se utiliza de dados relacionais para analisar os nós em relação à rede. Diversas métricas são consideradas para o estudo da rede como um todo (como centralização, densidade, coeficiente de clusterização) ou de cada um de seus nós (como grau de conexão, centralidade, grau de intermediação). O resultado dessa análise normalmente é apresentado na forma de um grafo, mostrando as relações entre os nós e suas propriedades.

Duas métricas em especial são consideradas neste trabalho: modularidade e grau de conexão. A modularidade é uma medida da rede como um todo. Ela divide a rede em grupos, de acordo com a força das conexões entre os diversos nós. Nós mais conectados entre si do que aos demais são incluídos em um mesmo grupo. Já o grau de conexão é uma medida do nó, e permite identificar a força da conexão de cada nó. O grau de conexão representa o somatório de todas as conexões partindo ou direcionadas a um determinado nó. Assim, se A está conectado a B e C, ele possui um grau de conexão 2. Em redes direcionadas (ou seja, em que é possível determinar o sentido da conexão), o grau de conexão pode ser desdobrado em dois tipos: grau de conexões recebidas (*indegree*) e grau de conexões enviadas (*outdegree*). O

indegree é particularmente relevante pois permite identificar os nós que mais receberam conexões em redes direcionadas, como no caso de identificar usuários que foram mais mencionados ou retuitados por outros no Twitter.

4 Procedimentos metodológicos

Para operacionalizar o estudo, partiu-se de um conjunto de dados relacionais obtidos no Twitter sobre um acontecimento jornalístico específico relacionado à Copa do Mundo de 2014. O acontecimento escolhido foram as notícias sobre eventuais vencedores de bolões da Copa após o resultado da partida em que o Brasil foi derrotado por 7 a 1. A partida foi realizada no dia 08 de julho de 2014. Os dados foram coletados com a ferramenta NodeXL no dia 09 de julho de 2014.

O NodeXL considera cada usuário autor de tweet ou mencionado em um tweet como um nó da rede. As conexões são estabelecidas a partir de relações de menção ou retweet. Ao retuitar um tweet, um usuário estabelece uma conexão com o usuário que foi retuitado.

A coleta compreende um período de cerca de 12 horas, entre às 22h30min do dia 08 e às 10h10min do dia 09 de julho. Ao todo foram obtidas 1.349 conexões (relações de menção ou retweet) e 1.344 nós (autores dos tweets e usuários mencionados ou retuitados) a partir de tweets contendo os termos “bolao” e “copa”.

Esses dados foram analisados com procedimentos e métricas de Análise de Redes Sociais. Após o tratamento dos dados, diferentes formas de visualização da rede foram experimentadas. Para gerar as visualizações, utilizou-se o software Gephi. Esse software de Análise de Redes Sociais permite aplicar diferentes métricas e algoritmos de visualização aos dados, possibilitando a realização do exercício visual proposto. Assim, diferentes grafos da rede foram gerados, buscando discutir diferentes formas de visualização a partir de um mesmo conjunto de dados.

O Quadro 1 procura listar algumas das características dos algoritmos dos layouts considerados no trabalho.

Quadro 1: Características dos algoritmos dos layouts considerados no trabalho (Fontes consultadas: <http://www.slideshare.net/gephi/gephi-tutorial-layouts>, <https://www1.ethz.ch/sg/research/graphlayout>)

Layout da rede	Ênfase	Autor
ForceAtlas2	complementariedades	Jacomy (2011)
ARF	complementariedades	Geipel (2009)
Circular	ranking	Groeninger (2010)
OpenOrd	divisão	Martin et al (2011)

Após incluir os dados das redes no Gephi, diferentes algoritmos de visualização podem ser aplicados para mapear a rede. Cada algoritmo possui características específicas, que vão resultar em visualizações diferentes.

O layout ForceAtlas2 tem como principal característica fazer com que os nós desconectados se afastem e as conexões sejam reforçadas (Jacomy et al, 2012).

O layout ARF é construído em resposta a layouts baseados em molas. Ao balancear as forças de atração e repulsão, o algoritmo deste layout permite a constituição de redes compactas que mostram claramente a separação entre grupos. Por conta dessas características, esse layout é indicado para redes pequenas, com cerca de 1.000 nós (Geipel, 2009).

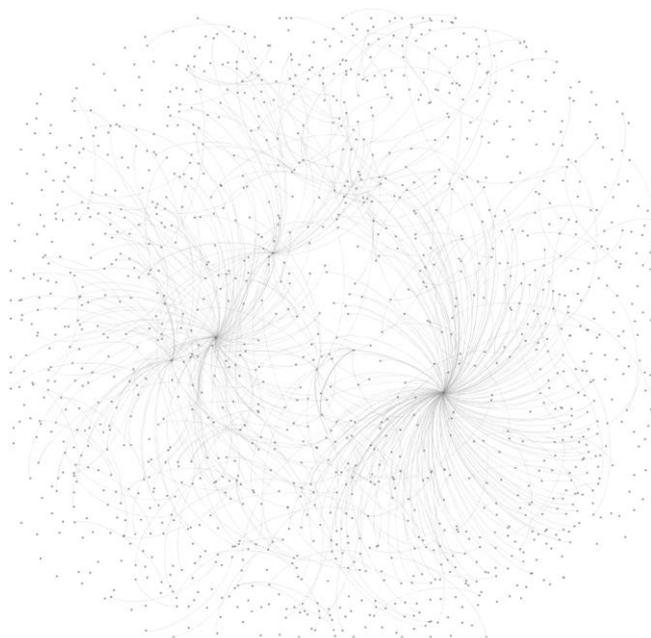
O layout Circular, por sua vez, possui uma ênfase em ranqueamento. Sua principal característica é permitir visualizar rapidamente quais nós possuem maiores índices para uma determinada métrica. Para tanto, ele dispõe os dados ao redor de um círculo, com os nós nas bordas e as conexões no interior da esfera.

Por fim, o layout OpenOrd enfatiza as divisões da rede, reforçando separações entre nós que não estão conectados entre si (Martin et al, 2011).

5 Resultados

O primeiro grafo foi gerado no Gephi sem usar nenhum filtro ou configuração prévia. O resultado pode ser visto na Figura 1.

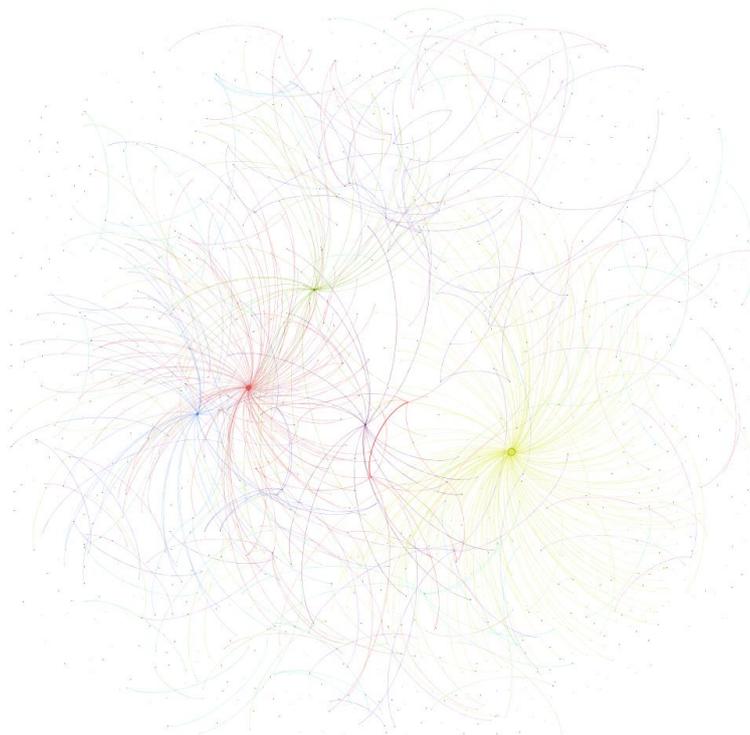
Figura 1: Layout padrão da rede gerado pelo software Gephi.



O grafo mostra os autores dos tweets (pontos cinza) e as linhas que conectam uns usuários aos outros, através de relações de menção e retweet. Nessa primeira configuração, todos os nós e conexões estão do mesmo tamanho. É possível perceber dois grandes aglomerados de nós conectados entre si, sugerindo o papel proeminente de alguns usuários da rede (aqueles que receberam ou enviaram mais conexões).

O segundo grafo foi criado partindo do mesmo layout, porém aplicando configurações de tamanho e cor (Figura 2). Aplicou-se a métrica de modularidade para colorir os diferentes grupos do grafo. O tamanho do nó está definido de forma proporcional ao grau de conexão *indegree*. Nós maiores representam usuários que receberam mais conexões (foram mais mencionados ou retuitados). Nesta segunda visualização, é possível perceber mais claramente os grupos de nós, e também os nós mais centrais da rede

Figura 2: Rede com a aplicação de cor nas conexões e tamanho nos nós.

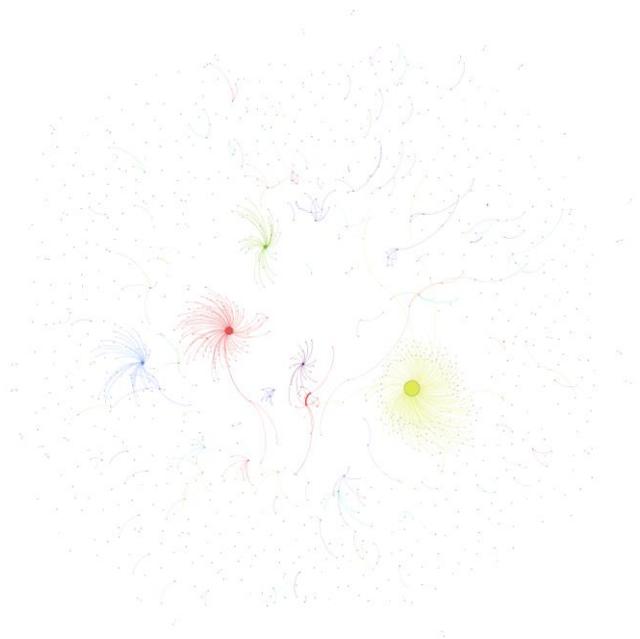


Em relação à Figura 1, um avanço que pode ser percebido é o fato de o grafo destacar outros usuários com papel importante (vê-se, por exemplo, que o emaranhado de nós mais à esquerda do grafo corresponde a pelo menos três grandes módulos –azul, amarelo e verde).

A terceira forma de visualização experimentada envolve a aplicação do layout ForceAtlas 2.

Esse layout atua sobre a rede no sentido de atrair os nós mais conectados entre si, ao mesmo tempo em que afasta os nós sem conexão. Com isso, os nós interconectados aparecem no centro, ao passo que os nós desconectados aparecem na periferia do grafo. O resultado pode ser visto na Figura 3.

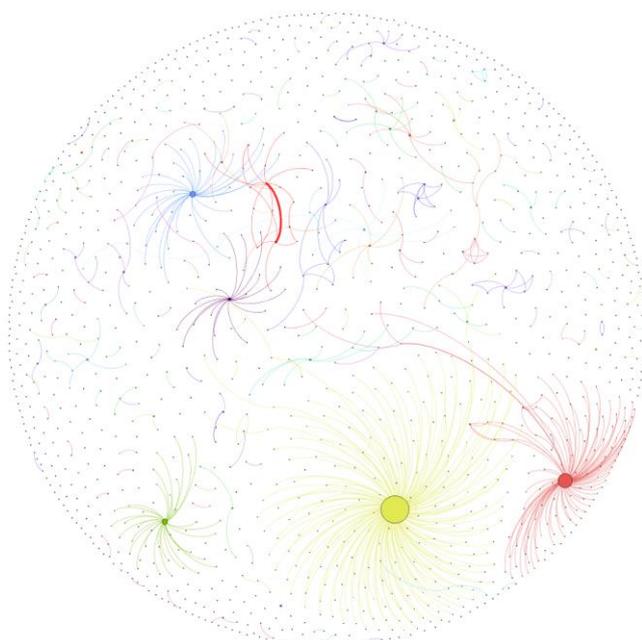
Figura 3: Rede com a aplicação do layout ForceAtlas2.



Essa forma de visualização permite identificar mais claramente os usuários que mais estabeleceram conexões, ou seja, mais mencionaram ou foram mencionados por outros usuários. Também é possível identificar alguns grupos de nós em torno de usuários que foram bastante retuitados. Além disso, distingue-se mais claramente os nós que estão conectados a outros dos nós que não estão conectados. Os módulos verde, azul e vermelho passam a ser exibidos de forma separada, sem sobreposições.

O layout a seguir (Figura 4) utiliza a configuração ARF. O layout é o resultado do balanceamento das forças de atração e repulsão.

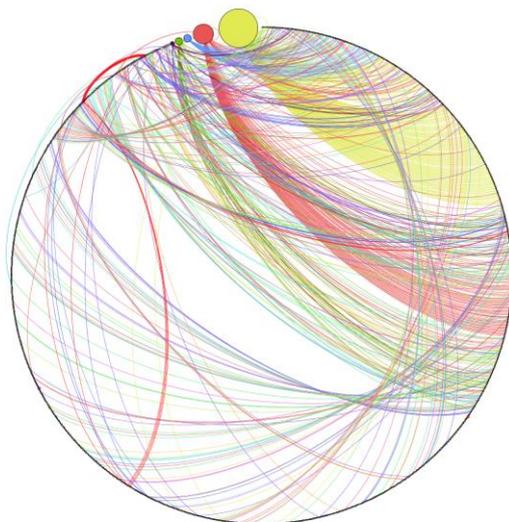
Figura 4: Visualização da rede com o layout ARF.



Nesse layout, pode-se perceber o peso de cada nó na medida em que cada usuário recebe um peso conforme o número de conexões recebidas (menções ou retweets feitos ao usuário). Também é possível perceber mais claramente o peso relativo de cada módulo, uma vez que o layout procura evitar sobreposições entre os módulos.

Uma outra possibilidade de visualização foi gerada com a aplicação do layout Circular, que dispõe todos os nós em um círculo e traça as conexões entre eles na parte interna da esfera (Figura 5).

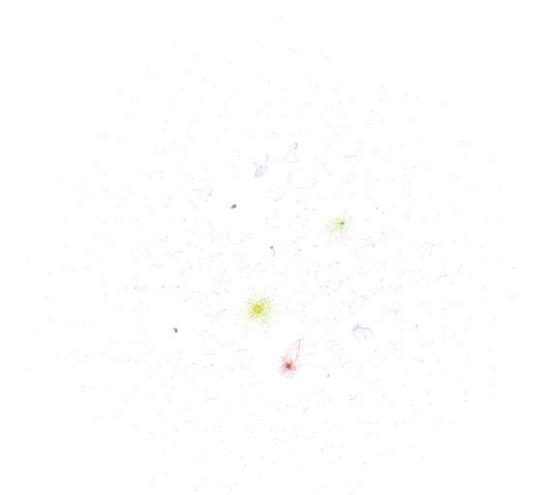
Figura 5: Visualização da rede com o layout circular.



Esse layout ajuda a perceber melhor relações de hierarquia entre os nós. No caso da Figura 5, os nós estão ordenados por *indegree* – quanto maior o grau de conexões recebidas, maior o nó, e mais no topo do círculo se encontra o usuário. Apesar de favorecer a exibição dos nós, a visualização das conexões fica prejudicada, com várias linhas (conexões) sobrepostas. Mesmo assim, é possível ter uma noção geral de quais usuários receberam mais conexões.

Na Figura 6, por sua vez, a rede é representada com o uso do layout OpenOrd. Esse layout reforça a divisão da rede – cada módulo é representado de forma isolada.

Figura 6: Visualização da rede com o layout OpenOrd.



Uma característica importante dessa rede é reforçada nesta forma de visualização: o fato de a rede ser pouco interconectada entre si. Assim, a rede possui 714 módulos, mas apenas 152 possuem dois ou mais nós. Ao tentar representar cada módulo isoladamente, percebe-se um predomínio de nós isolados (pontos ao redor do grafo) em relação a grupos de nós.

6 Discussão: design de redes

Ao gerar visualizações de redes sociais, inúmeros aspectos podem ser levados em consideração. Métricas diferentes podem ser aplicadas, e diferentes layouts podem ser usados para ‘desenhar’ a rede.

Geipel (2009), ao discutir as propriedades do layout ARF, sugere que um bom layout de rede é aquele que evita ao máximo a sobreposição das conexões, mantendo a harmonia entre os elementos representados. O autor aponta como critérios para analisar o layout de uma rede:

- número de conexões sobrepostas;
- estética do layout;
- separação de grupos;
- uso do espaço disponível.

Assim, para o autor, um layout com uma estrutura limpa (sem sobreposições, com uma clara separação de grupos e com um bom aproveitamento do espaço disponível), torna mais fácil de discernir diferentes grupos de nós.

Em termos de design da informação, pode-se dizer que o layout Circular prioriza aspectos cognitivos e comunicacionais – ele permite distinguir claramente uma informação (grau de importância do nó), mas faz isso em prejuízo de outras informações (como conexões, que aparecem sobrepostas no interior círculo).

Já o layout OpenOrd prioriza aspectos estético-formais em detrimento de aspectos cognitivos. Ao levar em consideração a configuração dos módulos, faz com que a rede seja representada de forma muito esparsa e difícil de acompanhar.

Em termos visuais, os layouts ARF e ForceAtlas 2 parecem balancear melhor a distribuição da informação. Enquanto o ForceAtlas 2 parece priorizar aspectos estético-formais, no ARF a clareza de informações é reforçada, ao evitar sobreposições, o que leva muitas vezes módulos que estariam próximos a serem afastados, como no caso do módulo vermelho, que nas outras representações aparece mais próximo ao módulo azul (o que significa que possuem conexões em comum).

7 Considerações finais

O trabalho procurou aproximar design da informação e análise de redes sociais a partir da realização de um exercício de experimentação na visualização de dados de uma determinada rede. Para tanto, utilizou-se de um conjunto de dados do Twitter relacionado à Copa do Mundo de 2014.

Ainda que se tenha partido de um mesmo conjunto de dados, diferentes formas de visualização puderam ser aplicadas. Cada uma dessas formas de visualização enfatiza diferentes aspectos da rede, ainda que se tratem dos mesmos *tweets*, dos mesmos atores, das mesmas conexões.

Escolher um layout envolve fazer escolhas em termos de design da informação. Escolher o que mostrar e o que ocultar traz impactos para a percepção daquele que visualiza a imagem, nos diferentes contextos em que um grafo social pode ser utilizado. Seja no contexto de uma pesquisa acadêmica ou de um infográfico na mídia, a compreensão de uma rede social a partir de um grafo envolve escolhas quanto ao que mostrar e ao que priorizar, e essas escolhas estão associadas aos dados selecionados e ao layout escolhido para visualização. Nesse sentido, entender como essas visualizações são formadas é um passo essencial para decidir como aplicá-las.

Referências

- CAIRO, A. 2012. *The Functional Art*. Berkeley: New Riders.
- ELLISON, N. B. & BOYD, D. 2013. Sociality through Social Network Sites. In: DUTTON, W. H. (Ed.). *The Oxford Handbook of Internet Studies*: 151-172. Oxford: Oxford University Press.
- GEIPEL, M. 2009. *Self Organization applied to Dynamic Network Layout*.
- JACOMY, M.; HEYMANN, S.; VENTURINI, T. & BASTIAN, M. 2012. *ForceAtlas2, A Continuous Graph Layout Algorithm for Handy Network Visualization*. In: *Controversy Mapping Reader*: <http://www.medialab.sciences-po.fr/publications/Jacomy_Heymann_Venturini-Force_Atlas2.pdf>. 13/08/2014.
- MARTIN, S.; BROWN, W.M.; KLAVANS, R. & BOYACK, K. 2011. OpenOrd: an open-source toolbox for large graph layout. *Visualization and Data Analysis*.
- O'GRADY, J. & O'GRADY, K. 2008. *The information design handbook*. Switzerland: RotoVision.
- RECUERO, R. 2009. *Redes Sociais na Internet*. Porto Alegre: Sulina.
- SCOTT, J. 2013. *Social Network Analysis*. 3.ed. Londres: Sage Publications.
- WASSERMAN, S. & FAUST, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- WATTS, D. J. 2011. *Everything is Obvious*. New York: Crown Business.
- ZAGO, G. 2014. *Circulação e Recirculação de Narrativas do Acontecimento no Jornalismo em Rede: A Copa de 2014 no Twitter*. Tese de doutorado não publicada. Departamento de Comunicação e Informação Universidade Federal do Rio Grande do Sul, Porto Alegre.

Sobre a autora

Gabriela da Silva Zago, PhD, UFPel, Brazil <gabriela.zago@ufpel.edu.br>