Representing connections: how visualizations shape understandings of networks

visual methodologies
a postdisciplinary journal
http://journals.sfu.ca/vm/index.php/vm/index

Cover Photograph by Ray Gibson
Anna Wilson
University of Stirling, UK

Anna Wilson is a former nuclear structure physicist who found herself drawn more and more towards the social sciences. She is currently undertaking a second PhD in the field of education at the University of Stirling.

Abstract

This article raises questions about a type of image that is becoming increasingly ubiquitous: network visualizations. Such visualizations – particularly of social networks – are used to demonstrate an interconnectedness that seems to have taken on an almost ideological tone, in which connectedness has been naturalized as a social good. Images of networks that seem dense, well-connected and mixed are presented in a positive light, while images of networks that seem to show segregation, low levels of connectedness or isolation are presented as evidence that something needs to change. They are seductive in their visual appeal, their apparent readability, the fixity they confer on both the networks they represent and the sense that they are conveying facts. However, this paper uses a case study to argue that images of networks are far from neutral, and that they need to be approached with a high level of criticality.

Keywords: data visualization, network visualizations, social network analysis, visual literacy, Twitter
Introduction

This article raises questions about a type of image that is becoming increasingly ubiquitous, appearing not only in publications aimed at academics but also frequently in news media, blogs, commercial documents, business cases, political analyses and documents aiming to influence policy. Images representing networks – particularly social networks – are used to demonstrate an interconnectedness that seems to have taken on an almost ideological tone, in which connectedness has been naturalized as a social good. Images of networks that seem dense, well-connected and mixed are presented in a positive light, while images of networks that seem to show segregation, low levels of connectedness or isolation are presented as evidence that something needs to change.

The apparently scientific nature of such images confers on them an aura of objectivity that suggests they represent some kind of ‘truth’ – that the networks they map are real and well-defined. Yet the processes involved in creating, viewing and contextualizing such images are complex. This paper considers the difficulty in developing a generic and widespread ‘data visualization literacy’ (Börner et al., 2015; Maltese et al., 2015; Skiba, 2014; Speth et al., 2010) independent of expertise in the analytical approaches that generate network visualizations and of cultural biases. It also suggests that mere literacy falls short of the criticality that needs to be brought to such images, and that rather than focusing on the development of technical data visualization literacies, we must learn to see visualizations as non-transparent, and to ask ourselves critical questions about the manipulations that may be occurring in the four sites of meaning-making identified by Rose (2016) – production, image, circulation and audience.

This paper starts by considering the importance of the notion of networks in contemporary society. It then considers some examples of contexts in which network visualizations are used to demonstrate something about social relationships and to influence viewers. Finally, it uses a case study based on a network visualization of an on-going conversation among teachers held on Twitter. The case study demonstrates the many ways in which choices made during the production of images, together with the ways new images blend with memories of those previously seen, manipulate the meanings that they purport to carry.

A network age?

It seems that we are living in an age of networks.

The pace and level of comfort of modern life is facilitated by physical infrastructure networks such as transport, communications and power networks. Connectedness is seen as massively important to economic development (see, e.g., Henderson et al., 2002), and the impact of increasing connectedness is often discussed in revolutionary terms. Indeed, some predict that the ultimate interconnection of everything in an Internet of Things will simultaneously disrupt everything and enable the prediction and solution of all problems (Burrus, 2014).

Network approaches have also risen in importance in the sciences, with the spread of complexity theory and systems approaches from the physical sciences to biological, environmental and even social sciences. Neural networks (Haykin 2009) are the model for artificial intelligence, and the notion of networks in the brain now dominates popular conceptions of memory and learning (see, for example, Pearce Stevens, 2014).
Furthermore, society itself is increasingly described in network terms. For example, we talk of sociocultural networks such as family, professional and friendship networks, including “virtual” networks created through social media. Such networks are talked of in largely positive terms – the more connections we have, the more support we are likely to have, and the better anchored into society we are likely to be (see, e.g., Berkman and Glass 2000). The more extended and interconnected our business networks, the more successful we are likely to be at finding suppliers for our production lines or markets for our products or services (see, e.g., Ter Wal and Boschma 2009). Particularly by the social media platforms that facilitate virtual networks, we are encouraged to see ourselves as living in a highly connected world, and to see our degree of connectedness as a measure of our success as humans.

The combination of the notion of the brain-as-network and social networks has even spawned a new theory of learning, connectivism. Connectivism asserts that ‘personal knowledge is comprised of a network’ (Siemens 2005, np.) which may include non-brain-based elements such as databases and computers; ‘learning is a process of connecting specialized nodes or information sources’ (ibid.) and again, that may occur outside of individuals’ minds; ‘nurturing and maintaining connections is needed to facilitate continual learning’ (ibid.); and ‘social network analysis is an additional element in understanding learning models’ (ibid.).

But what do we mean when we talk of networks? Simply speaking, a network is comprised of (physical or virtual) objects, people, or places – network nodes or vertices – and the relationships that connect them – network lines or edges. Figure 2 shows an example of a real, physical network – the connections between floats and knots in a fishing net. In the “fishing net” network, floats and knots are nodes, and the lines are the lengths of rope and twine that run between them.

Just as in the fishing net, social networks may contain more than one type of connection, and more than one type of node. And, as in the fishing net, the relationships may be extensive, tangled and difficult to comprehend. In addition, they may be – indeed are very likely to be – constantly changing, with some connections enduring, others coming into being only fleetingly, some bringing nodes closer together and others pushing them farther apart.

The rise of network visualizations

The apparent importance of networks raises the question: how can we understand both how we generate and shape networks, and how they in turn may shape us? A first step in answering this question is to find a way to represent network information in a way that viewers can understand.

The photograph in Figure 2 seems to be an objective, neutral image that captures a particular moment in the life of the “fishing net network”. However, it does not present the viewer with an easy way to untangle or analyse it. Instead, we might prefer to produce a diagrammatic rendering in the form of a network visualization. Diagrammatic representations not only abstract away some of the messiness that might be a source of confusion, they also allow for quantification. For example, measurements can be made of the lengths of different paths between selected nodes; the importance of individual nodes in holding the network together; and overall properties of the network such as connectedness, density and so on. They thus carry an air of objectivity and scientific weight – they pose as a-theoretical representations that allow us to see underlying truths more clearly.

Network visualizations are becoming increasingly embedded in both research and public life. As social network analytical approaches (Scott 2012) are applied to investigations of everything from adolescent cigarette smoking (Mercken et al., 2012), through the spread of happiness (Fowler & Christakis, 2008), to data about online learning (Rabbany et al., 2014), we are increasingly presented with data visualizations that seem both quickly digestible and aesthetically appealing. But
the apparent ease with which we read such images may be misleading. A great many choices are made, both consciously and unconsciously, by humans and by software, in the production and selection of such images; these choices silently and subtly shape the understanding of those who view them.

Network visualizations are often used to highlight connectedness and community clustering. For example, a presentation on behalf of the American Society of Plant Biologists made use of a brightly-coloured network visualization to encourage its audience to see themselves as ‘hubs’ and so to contribute to community science projects (Williams 2015). Pakistan’s Community Motivation and Development Organization illustrates its own community network with another brightly-coloured, although unexplained, network diagram (CMDO nd.), presumably indicating the connectedness of different community groups.

Alternatively, visualizations may be used to show that some presumed members of communities are in fact isolated (see the case study below for an example), or that communities are more segregated than we are aware of (or would wish for). An influential example of this kind of visualization can be found in Moody’s (2001) study of racial segregation in high school friendship groups. Similar visualizations are used by members of the activist reporter community Global Voices to illustrate how members of different camps in the Israel/Palestine conflict access very different news sources and reports, thus apparently further embedding the separation of their worldviews (Lotan 2014).

Sometimes, network visualizations are used to show who the most influential figures in a particular community or section of society are. For example, the business of the company Relationship Science comprises compiling, measuring and monitoring social networks in order to determine individuals’ “relationship capital.” Network diagrams produced by this company and showing connections between delegates at the World Economic Forum appeared in the Economist 2014 (Economist Online 2014). In a similar network visualization and quantification project, researchers at GfK and the University of Vienna appeared to come to the rather surprising conclusion that Alan Rusbridger (then editor of The Guardian) and Alberto Nardelli (founder of Tweetminster and then about to become Data Editor at The Guardian) were the most important individuals on Twitter in relation to the ‘European political Twittersphere’ (GfK 2014, np.).

Visualizations are sometimes used to simultaneously illustrate community clustering or segregation and identify top influencers. For example, Stray (2013) uses a network visualization to simultaneously show influencers and segregation in a study of conversations about gun control on Twitter. In this case, the influencers are tweets rather than Twitter users, and his visualization suggests that there is little dialogue between pro- and anti-gun control lobbies on Twitter. Leetaru (2013) uses a network visualization to illustrate how Edward Snowden had supplanted Julian Assange in terms of media coverage, here discovering (among other things) that Barack Obama and Vladimir Putin are important influencers in relatively distinct networks.

These types of use of visualization are fast becoming part of our everyday lives, and are used to show us aspects of our own connectivity. In education, they are increasingly being included as features of the dashboard displays that display tracking data of activities in Learning Management Systems (Duval 2011; Ferguson and Shum 2012; Shum and Ferguson 2012; Siemens et al. 2011). The suggestion seems to be that by seeing how connected we are in a “learning network”, and perhaps by seeing the connectedness of our peers, we may learn to learn better.

Network visualizations are also being proposed as forms of assessment through the European Commission-funded PREATY project (PRoposing modern E-Assessment approaches and Tools to Young and experienced in-service teachers) (PREATY nd.). However, understanding what exactly such network visualizations tell us about our learning, or that of our children or students, is a non-trivial task.

Visualizations such as those used in the examples above are seductive in their apparent readability, the fixity they confer on both the networks they represent and the sense that they are conveying facts. However, this paper argues that they are far from neutral. They are often presented with little explanation of how the layout of nodes and lines has been decided upon, how colours and shapes have been assigned, and so on. An example of an exception can be found in Leetaru (2013). Here, the author gives some detail about how he produced the visualization, but still not enough for an untutored eye to critically consider the inferences he draws from them. Part of the danger appears to lie in the way that some aspects of the visualizations quickly confirm what we think we already know – of course Obama and Putin are important influencers – and in so doing appear to confirm the “correctness” or truth of the image. But at the same time choices made in their construction – for example the use of colour to visually separate clusters depending on a measure of interconnectedness – can make what is actually complex appear simpler than it really is. The following case study illustrates
the impact of choices made by the creators of a network visualization (both human and algorithmic) can have on its apparent message.

**A Twitter chat among teachers**

The network visualizations constructed in the case study presented below are part of a larger, ongoing study exploring the role of images in professional learning as it unfolds through Twitter conversations, and the potential pedagogical use of such images in the formal education of pre-service professionals (Wilson, 2015; Wilson, 2016a). The larger study seeks to explore the flows of knowledge, practice and affect in such conversations, and so to contribute to an improved understanding of the potential for, and barriers to, learning in these exchanges. The current case study uses data gathered in the first phase of the larger project, which involved observation of regular ‘conversations’ (exchanges based around common hashtags) held among teachers on Twitter. The conversation occurs each Friday, although sometimes teachers tweet with the conversation’s defining hashtag both before and after. It consists of teachers tweeting about the best thing that has happened in their teaching during the preceding week. The data gathered included both details of the tweeted images themselves and the visible user-image interactions that occurred during the observation period – that is, when and by whom tweeted images were retweeted, favoured or replied to.

Although, as its popularity has increased and stabilised, Twitter has become the subject of a great deal of research attention, the sharing of images on Twitter remained under-researched until recently. Vis et al.’s (2013) work on the role of images in the transmission of both eyewitness reporting and rumour during the 2011 “London Riots” is a notable exception, and represents a recent interest in the use of images in public and institutional responses to crisis events. However, images shared on Twitter within professional groups remain largely neglected by researchers. Yet they offer a powerful means to subvert Twitter’s 140-character limit – as one of the participants in the research below said, ‘pictures really do speak a thousand words’; as another observed, images seem to be processed more immediately and somehow more intuitively than text: ‘they’re in your brain very quickly, and they stay in your brain a long time’. They therefore offer a potentially rich and complementary alternative to analysis that focuses on the text of tweets, quickly, and they stay in your brain a long time’. They therefore offer a potentially rich and complementary alternative to analysis that focuses on the text of tweets, as another observed, images seem to be processed more immediately and somehow more intuitively than text: ‘they’re in your brain very quickly, and they stay in your brain a long time’. They therefore offer a potentially rich and complementary alternative to analysis that focuses on the text of tweets, opening up new possibilities for investigating the learning that is inevitably unfolding in these exchanges.

**Case study: a Twitter retweet network**

Because the aim of the larger study was to investigate how images might shape flows of knowledge and affect in the conversations, ways of visualising the interactions and relationships between images and participants were developed (Wilson, 2016a; Wilson, 2016b). The present case study focuses on one such visualization – that of relationships between participants who retweeted other participants’ tweets with new hashtags. The nodes in the network are participants in the conversations, and the lines that connect them indicate that one participant has retweeted a tweet originally posted by another.

The visualizations have been created using the software NodeXL (Hansen, Shneiderman & Smith, 2010). This is the first of many choices that shape the possible visualizations. This particular software package has its own menu structure and user interface that en- and discourage certain processes and approaches that might be quite different in other packages. For example, the ease with which one can edit the properties of individual nodes and lines may be greater in NodeXL than some others, because one can directly edit the entries in the spreadsheet from which the visualization is created without having to edit and re-load external data files. It is thus a package that allows the user to both correct and insert errors relatively easily. On the other hand, it is not a package that encourages the user to think in terms of bimodal (or multimodal) networks. Thus by choosing to use NodeXL, the researcher is perhaps rendered more likely to look for (and therefore create) unimodal networks in which participants appear to be genuinely and directly connected with each other, rather than linked via mediating objects (in this case, tweets).

The software also has a specific set of lay-out algorithms, its own (quasi) random number generator, a limited number of shapes representing nodes, and so on. These, too, differ between packages. The user interface makes some algorithms and lay-out options easier to discover and try out than others; for example, a user can quickly change between layout algorithms, but it is harder to make quick changes to clustering algorithm choice (see below). NodeXL also makes it very difficult to reproduce exactly the same visualization from the same data, so that if a user changes from one lay-out algorithm to another and then decides to revert to the original, they will not produce the same visualization as they had originally.

Despite the constraints introduced by choosing a single software package, a large range of visualizations can still be generated from a single set of unedited, unchanging data. Figure 3 is a representation of retweet relationships between participants in the Twitter conversation. Nodes represent individual Twitter users (or at least accounts); the size of the node scales according to the number of times that particular user has had tweets retweeted by others during the observation period. Arrows connecting nodes point from the retweeter to the original tweeter.
**Figure 3:** A retweet network for tweets containing images posted during teachers’ Twitter chats

**Figure 4:** As Figure 3.

Figure 4 is another representation of exactly the same network. And so is Figure 5.
These three images illustrate how exactly the same data can be rendered in ways that are visually radically different. These apparently very different representations of the same data have not been shaped by decisions to include or reject particular nodes or lines – i.e. conversation participants or retweet relationships between them. Rather, they are the result simply of different choices of algorithm governing the layout. All three are examples of what are called ‘force directed’ layouts – data layouts that do not have axes that regulate the placement of particular points, but that use the strength of the connection between two nodes (in this case, the number of times one participant has retweeted tweets posted by another participant) to determine how close they should lie. However, they differ in the degree to which an overall shape is imposed, and the choices made about the relative positioning of different nodes. Figure 3 has laid the data out using a random pattern. Figure 4 has been produced using an algorithm which aims for a roughly circular shape overall, placing the least connected nodes at the circumference and the most connected nodes towards the centre. It also aims to minimize the variability in edge lengths. Figure 5 uses a different algorithm again, which this time starts by making a high-dimensional (digital) representation of the data and then projects down onto a particular, but randomly chosen, two-dimensional plane.

Figure 5 can also be used to illustrate the contingency of meaning-making on personal and cultural backgrounds. For the author, this image calls to mind a whirling dancer, skirts flying. For one of her colleagues, the visualization in Figure 5 recalled an Inuit ulu – a “woman’s knife” used to scrape meat from hide and

Figure 6: “Mevlevi” by kT Lindsay (https://www.flickr.com/photos/ktlindsay/3409225057) - licensed under CC BY-SA 2.0; “Ulu” by Aileen Ireland, reproduced with permission; Japanese Warrior, courtesy of Pixabay, reproduced under CC0 licence.
occasionally to eat with. For another, it triggered recall of an image of a Samurai in full armour. That is, each of us had a response that was itself a visual memory, blending the image in Figure 5 with one of the images in Figure 6. We thus each created different image-responses, which in turn shaped the ways we were likely to interpret the meaning of the visualization in Figure 5. Blended with a memory of a whirling dervish, the image seems to carry connotations of spirituality, of fulfilment and of trance and entrancement. Blended with a memory of an ulu, the image seems to carry very different connotations of daily life, the earthy, solid fleshiness of meat and blood and food. Blended with a memory of an image of a Samurai, the image seems to carry different connotations again – this time of war, violence, honour and ceremony. Thus these network visualizations seem to have a function rather like Rorschach tests.

The choices made so far in constructing these images (both by the author and the algorithms, through their rule-based and random elements) thus have a significant impact on both their appearance and the possible meanings a viewer may make from it.

However, there are still more choices to be made. As discussed in the introduction, most of the network visualizations that appear in the media are brightly-coloured, so my next step is to add some colour. Concerned that Figure 5 seemed to produce such variable responses even in the small sample of three forty-something women, all full time PhD students in the same School of Education (and all in fact with the same primary supervisor), I return to the visualization of Figure 4, which appeared to us to be a little more neutral.

Perhaps I want to use the colour to highlight clusters of nodes that are more interconnected with each other than with other nodes. To do this, I first use the software to calculate some metrics about each node and then group them on the basis of those metrics. This will, I hope, help the ‘reader’ of the image see how the overall Twitter network consists of several subnetworks, some of which are more interconnected with other subnetworks and others relatively isolated. Doing this produces the image shown in Figure 7.

How do we interpret an image like this? Now, where in Figures 3-5 we had an apparently dense network of multiply-connected nodes, patterns start to appear. Some nodes seem to be central to their own subnetworks, such as the large dark blue node in the bottom right quadrant of the figure. We might interpret this to mean that this particular Twitter user is an important influencer, or a stabiliser of the network. Others who are less connected could be interpreted as being held in the network by this hub.

However, I could have made a different choice. Rather than colour nodes based on their clustering, I could instead make the software assign them into groups depending on their absolute connections. The resulting visualization, in which each group is not only assigned a different colour but also spatially separated, is shown in Figure 8.

Unlike Figure 7, this image now draws attention to the apparent disconnectedness of a small number of chat participants, indicated by the nodes and edges on the right hand edge of the graph. Not only that, but the apparently relatively separate subnetworks such as the one coloured in dark blue in Figure 7 can no longer be made out by the viewer – not even by an expert viewer.
This image gives an impression of one large and highly-interconnected group in which almost all of the participants are connected to each other, and a small number of participants who are not, really, participating. Perhaps I would choose to show this if I wanted the viewer to worry about those insufficiently-connected participants, or if I wanted to show how successful this conversation is overall in creating a highly-connected community. (I should note, I have added a further feature to this particular image, more out of aesthetic considerations than anything else. Now, I have used the software to vary the opacity of each node according to the type of action they have carried out during the conversations. The palest nodes represent those who only retweet; midtones represent those who only tweet and do not retweet anyone else’s tweets; the darkest nodes indicate those who both tweet and retweet. Such variation in colour does not actually increase the information potential of the image, as the edges connecting the nodes already show whether someone retweets or is retweeted by the direction of the arrow, but I think the overall effect is to make the image prettier.)

Of course, I might want to emphasise the formation of subnetworks even more than I had done in producing Figure 7. I can do this by not only colouring groups according to the clustering metric, but also by asking the software to lay the nodes out spatially so that the groups are well-separated. If I do this, I produce the visualization shown in Figure 9.

Here, rather than grouping according to the existence of at least one path through the edges and nodes connecting the nodes in each group as in Figure 8, the nodes have again been grouped according to their relative connectedness or clustering, exactly as in Figure 7. The difference is simply that now, both colour and spatial lay-out have been determined by group. Thus this visualization doubly emphasises segregation, by visually separating nodes into groups on both colour and spatial layout. A quick reading of such an image might suggest that instead of one big, interconnected community conversation, the Twitter chats are more like over twenty effectively separate, relatively closed conversations among subgroups – and that many of them seem to be dominated by a central Twitter user who is retweeted by a circle of followers, who (perhaps?) do not originate tweets themselves – or at least not tweets that are retweeted by others.
Which of these visualizations represents the truth? The answer, of course, is both all and none. Does it matter? Perhaps. The same kind of double emphasis of segregation – via colour and spatial layout – can make social fragmentation and isolation seem shocking and urgent. Maybe this does not matter for a Twitter chat among teachers (although it might lead a viewer to underestimate the extent to which practice knowledge may spread through such exchanges). However, when visualizations colour groups along racial (as in Moody (2001)), ethnic (as in Lotan (2014)) or political (as in Stray (2013)) lines, they may lead to an exaggerated sense of social breakdown. Indeed, an expert reader of the images in Moody’s (2001) study recognizes that they, too, have been constructed with a double emphasis 9colour and clustering) on racial difference, and that the segregation among high school friendships is not as severe as an untutored glance might be led to believe. Conversely, visualizations such as those in Figures 3 and 7 may generate an exaggerated sense of connectedness for the majority.

Conclusions

The inherent multiplicity in possible visualizations illustrated above has not even begun to address the additional variability in how one chooses the type of relationships to record as lines connecting nodes. Here, again, I (like any other network visualizer) have made a choice in defining what counts as a relationship. When participants actively engage with someone else’s tweet during a twitter conversation, they have three options – they can retweet it, favourite it, or reply to it with their own tweet. One might consider which of these represents a “real” connection between users. Retweeting may be done by anyone who happens across a tweet – there’s no guarantee that the original tweeter and the retweeter have ever had any interaction, or engaged in any form of mutual communication. Favouriting is perhaps even more tenuously related to genuine communicative connectivity: although some people intend favouriting to indicate approval or appreciation, others use it as a personal bookmarking or logging function (Wilson, 2016b). Commenting using the reply function does, at least, imply that some form of intentional communication between the connected nodes has taken place. Figure 10 shows an alternative visualization of the conversation network, where the lines now represent “replied-to” relationships. The much lower density of points and lines indicates that this kind of relationship is much less common, suggesting that the image of dense, highly-connected networks or subnetworks conjured up by the other figures is perhaps something of an exaggeration.
the ability to ask such questions, individuals may not understand how to or even whether they should try to adjust their position and connectedness.

When they are used to provide evidence intended to influence policy or practice, it is important to recognize that they may exaggerate subtle effects. Depending on choices made about how to lay out and colour nodes, they may give too strong an impression of segregation, or conversely of integration. This may have serious consequences if policy decisions are based on these impressions. It is thus important that, as such visualizations become increasingly common, those viewing them recognize that they are not neutral representations of fixed truths, and learn how to critically question their construction and apparent meaning.

References


