5 Network Visualization Literacy

Novel Approaches to Measurement and Instruction

Angela Zoss

Duke University

Adam Maltese

Indiana University

Stephen Uzzo

New York Hall of Science

Katy Börner

Indiana University

Abstract

Network visualizations, a particular kind of data visualization, can be a useful way to visually represent the relationships in real or theoretical social, physical or biological systems. Network data can be generated and analyzed without being visualized, but the visualizations are often more compelling and may be more easily understood than numbers that summarize network properties. With the growth of network science research across a variety of domains, there is an increased call for basic literacies in networks and the ability to use network visualization as a powerful tool to understand interactions in complex systems.

In this chapter, we discuss the current status of the research on network visualization literacy (NVL), how it is measured, what the current research says about NVL across a variety of contexts, ways experts are teaching to develop NVL, and recommendations based on our current understanding of best ways to improve NVL.

Introduction

Broadly speaking, we visualize data to investigate complex relationships among variables and to communicate these relationships to others. Network visualizations translate network data into a visual representation of some combination of the actors, relationships, clusters, and data attributes. The value of visualizing the structure of relationships and connections is being recognized as an increasingly important 21st Century skill due to the need for all people to have a better understanding of complex phenomena across disciplines.

A call for increased literacy about networks, writ large, resulted in the development of essential concepts that can be taken as a set of goals for what a network literate person should know by the time they graduate high school (Sayama, Cramer, Porter, Sheetz, & Uzzo, 2016). For the purposes of this chapter, we define network visualization literacy (NVL) as the ability to read, interpret, and create visualizations of various types of networks. Research on NVL is still in its early phases, and recent studies suggest that NVL, and more generally data visualization[1] literacy, of youth and adults is not very high or broad (Börner, Balliet, Maltese, Uzzo, & Heimlich, 2015; Maltese, Harsh, & Svetina, 2015).

Given this, we focus this chapter on a set of topics that together constitute an attempt to build a more comprehensive vision for NVL, including how to measure NVL, the role of NVL in teaching and learning, what the current research says about NVL across a variety of learning contexts, and recommendations based on our current understanding of best ways to improve NVL. Before we move into discussing these topics, we first define how we conceive of NVL.

Network Visualizations

The simplest network visualization is an adjacency list, where each node is itemized and followed by a list of all of the other nodes with which that node shares a link (its neighbors). In the example in Figure 1A, entity A has connections with B and D, and entity B has connections with A and C. Entities C (with B) and D (with A) only have singular connections with other nodes.

Large networks are more likely to be visualized as matrices or node-link diagrams and can be displayed using one or more of several organizing principles. A matrix visualization (Figure 1B, representing the same network data as Figure 1A) is a tabular visualization where a node is represented by either a row or a column (or both) and a link is represented by a numerical value placed in the cell where a node row and a node column intersect. For example, in a matrix visualization of a network of individuals who send text messages to each other, a two-dimensional

[1] In this chapter, "data visualization" is being used broadly to refer to information visualizations, scientific visualizations, and other conceptual or diagrammatic visualizations.

table is created where the same names appear in the row and column headers. Numerical values representing the number of texts sent between the two people will appear in the cell where the row of one individual and the column of the other intersect. Columns and rows can be ordered to highlight patterns in the data values, such as social cliques where all members text each other a lot (Eliassi-Rad & Henderson, 2010).

A

Node	Neighbors
A	B,D
B	A,C
C	B
D	A

B

	A	**B**	**C**	**D**
A		1	0	1
B	1		1	0
C	0	1		0
D	1	0	0	

Fig. 1. Sample network visualizations: adjacency list (A), matrix (B).

In contrast to a matrix, a node-link diagram represents each actor as a single point using some graphical icon or symbol (often a circle). The presence of a link between two actors is visualized by the addition of a line or arc between the nodes (Figure 2). These components are often laid out such that smaller distances between nodes represent higher similarity (Figure 3), but nodes can also be arranged in a circular layout, perhaps in order of a certain property (e.g., a node's number of links), or against a separate reference system like a geospatial map or a science map, where scientific disciplines are arranged in space using citation- or topic-based similarity algorithms. (Figure 4).

Researching Network Visualization Literacy

In general, we define data visualization literacy as the ability to make meaning from and interpret patterns, trends, and correlations in visual representations of data (Börner, Maltese, Balliet, & Heimlich, 2016). In order to interpret visualizations, users need the ability to complete a combination of the following tasks: read text, interpret data arrangements (e.g., to see correlations, trends), and compare object properties (e.g., compare the sizes of nodes in a network given a legend). Users of any information visualization form may engage in a variety of tasks, including both low-level tasks like data foraging and high-level tasks like problem-solving and composing (i.e., making decisions based on data trends) (Card, Mackinlay, & Shneiderman, 1999).
As a subset of data visualization, network visualization is subject to many of the same kinds of interpretation issues present in other approaches to data visualization. Given the range of abilities needed to interpret visualizations and the myriad

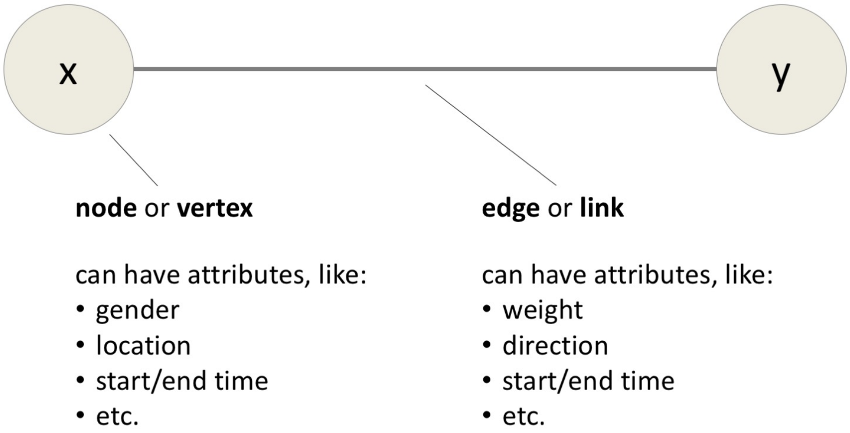

Fig. 2. Node-link diagrams typically represent nodes as circles and links as lines or arcs.

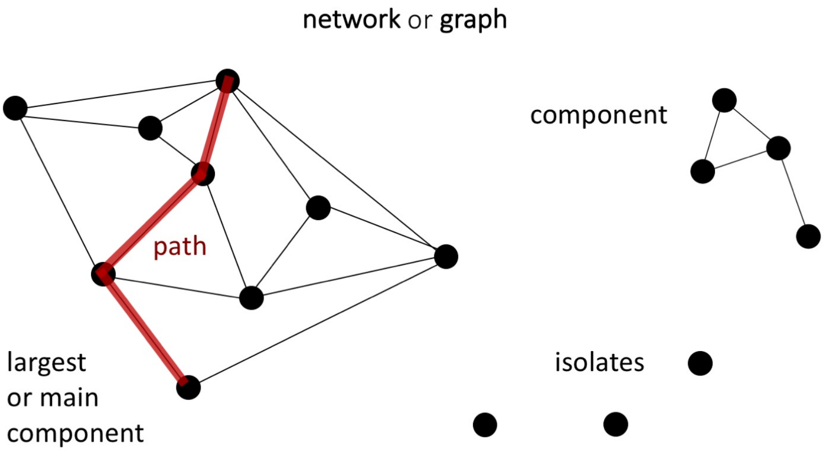

Fig. 3. A simple node-link diagram, labeled with common network-related terminology.

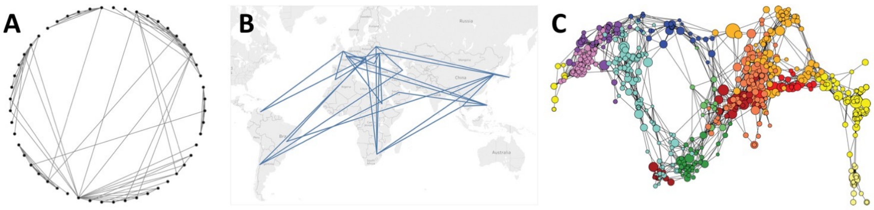

Fig. 4. Sample network visualizations, using a circular layout algorithm (A), a geographic layout (B), and a science map (C).

tasks possible, it is important to acknowledge the opportunities for network visualizations to be easily misinterpreted. These challenges arise through lack of clarity about the limits of network visualizations in interpreting very complex systems, and the many ways that characteristics and behaviors of network components can be represented.

Representational literacy: Do individuals understand how network data are converted into visuals?

There has been an ongoing concern among network science practitioners about the trajectory of network science as a way into deepening data understanding, particularly of large data sets. A call for increased literacy about networks, writ large, has resulted in the articulation of a set of seven essential concepts and core ideas (Sayama et al., 2016). These essential concepts can be taken as a set of goals for what a network literate person should know by the time they graduate high school. The fourth essential concept is: "Visualizations can help provide an understanding of networks," and the core ideas subsumed by it include:

- Networks can be visualized in many different ways
- Diagrams of a network can be drawn by connecting nodes to each other using edges
- There are a variety of tools available for visualizing networks
- Visualization of a network often helps to understand it and communicate ideas about connectivity in an intuitive, non-technical way
- Creative information design plays a very important role in making an effective visualization
- It is important to be careful when interpreting and evaluating visualizations because they typically do not tell the whole story about networks

These essential concepts are relatively new, and scaling them into wide adoption will require robust validation and transformative professional development that integrates new curriculum and learning materials on network visualization into rigorous content knowledge and pedagogical approaches.

Research on the skills required to interpret network visualizations and the prevalence and quality of those skills is still in early phases. Small-scale studies investigated the comprehension of the basic metaphors used by the diagrams (Fabrikant, Montello, Ruocco, & Middleton, 2004), the specific structural properties of network data (Ghoniem, Fekete, & Castagliola, 2005), and the graph design aesthetics that are most likely to improve performance on quantitative interpretation tasks (Bennett, Ryall, Spalteholz, & Gooch, 2007). In the sections that follow we outline some general questions related to NVL that have been investigated through research along with initial findings.

Metaphoric literacy: How intuitive is the arrangement of nodes and links in a network visualization?

Network visualization literacy studies might address whether users understand the metaphoric properties of the visualization – that is, the implicit structures the visualization is using to represent network data. Most node-link diagrams have conventions that guide interpretation of the diagram, such as:

- The positions of nodes are an approximation of the similarity between the nodes, based on an analysis of the links between nodes (and possibly also the weights of those links)
- Nodes that are close together are more similar than nodes that are far apart (the distance-similarity metaphor)
- The positions of nodes may be influenced by aesthetic choices that are encoded into the layout algorithm (e.g., to minimize edge crossing) or that are used to make manual adjustments (e.g., to eliminate overlap of two nodes by manual shifting)
- Network visualizations can be rotated or reflected in space arbitrarily
- Some network visualizations omit a portion of the links to better focus attention on the node positions and the most important link structures
- Shorter links are usually stronger than longer links, even though longer links may draw the eye and shorter links may be so short that they almost disappear

Studies of the metaphoric properties of network visualizations are rare and have focused primarily on the distance-similar metaphor. Fabrikant and colleagues (Fabrikant & Montello, 2008; Fabrikant et al., 2004; Fabrikant, Ruocco, Middleton, Montello, & Jörgensen, 2002) explored the judgments of novice users of network visualizations regarding the presumed similarity of two pairs of target nodes, manipulating a variety of topological and aesthetic variables: the Euclidean distance between the nodes, the cumulative measured length of links between the nodes, the number of intervening nodes on the path between the target nodes, and the width, darkness or hue of links. In all studies, participants overwhelmingly associated similarity with the length of the path between two nodes (in terms of geometric length or direct-line distance, not the number of links in the path). Nodes close to each other "as the crow flies" were considered less similar to each other than nodes that had a shorter network connection. The only design features of a network that contradicted this powerful intuition were the width of a link and, to a lesser extent, the darkness of a link; wider links especially made nodes seem more similar to each other, even if those nodes had a longer measured path.

In a final study, Fabrikant and Montello (2008) compared judgments of node similarity to judgments of node distance by making a slight change to the task instructions from their previous studies, such that participants answered questions about similarity and distance separately. When asked about distance, participants focused on Euclidean distance. When asked about similarity, participants focused on the geometric length of links. These results are encouraging, in that network layout algorithms may make compromises about where a node is positioned, thereby rendering "as the crow flies" distances less meaningful than the presence of links. On the other hand, the length of links can be determined both by the layout of the nodes and by whether the layout algorithm has a constraint on link length. Novices without a sophisticated understanding of layout algorithms will be likely to make judgments based on the length of the lines.

One way of interpreting these findings is through the lens of basic perceptual skills. Even without special training, users of visualizations have natural skills for interpreting spatial information. These skills were described over a century ago by German psychologists as "Gestalt laws" (Ware, 2013), and they can help explain how components of data visualizations are understood on a very fundamental, perceptual level. These laws are especially relevant for network visualizations, where training and even exposure are uncommon among a general population. The arrangements of nodes in space and the connect of those nodes by lines have very strong connotations for users, and visualization designers must anticipate how that will affect interpretation.

Topological literacy: Do individuals understand basic network properties when reading network visualizations?

Beyond a user's intuition about a network visualization, researchers may also want to investigate whether users can glean topological information (i.e., the mathematical or statistical properties of the network data underlying the diagram) from the visualization. Depending on the field of study, different parts of a network dataset may be considered especially important. In some fields, the clusters of nodes in the network are most important, whereas in other fields it is important to identify specific nodes that are highly influential. For example, a study could measure a user's ability to identify nodes with a high betweenness centrality score (i.e., those nodes that lie on heavily travelled paths between node clusters) from the network visualization. Testing whether a user can read or estimate topological information about network data from a visualization can be an important way of assessing either the user's literacy or the visualization's success.

A user's topological literacy is dependent on many factors: the user's prior training with both network data and network visualizations, the choice of network layout algorithm (and by extension, the topological properties that are emphasized by the network layout algorithm), any additional design choices made by the producer of the network visualization (e.g., adding color coding to emphasize a particular topological property), the specific properties of that particular network dataset (e.g., the size of the network, whether some nodes have notably more links than others), and the choice of topological property (i.e., "task") to measure.

Effect of layout or base map choice

Node-link diagrams have a wide variety of layout algorithms (Figure 5) that determine the position of nodes and edges. The most common layout algorithms, especially for small or medium-sized networks, are algorithms drawn from physical analogies likes springs and forces, pushing and pulling the nodes into place based on the presence and/or weight of edges. The complexity of network data means that there is no one "correct" layout of the nodes and edges – two nodes may have a strong link to each other, but they may also be strongly connected to other nodes that are very far apart. Because of this complexity problem, different layout algorithms have been developed to prioritize either different features of the data or to make certain types of visual judgments easier.

One way of evaluating layout algorithms is to explore the extent to which the layout follows guidelines for graph design aesthetics (Börner, Chen, & Boyack, 2003; Brandes, 2001). Many such graph aesthetic principles have been identified (Bennett et al., 2007), including:

- Global and local symmetry
- Non-overlapping nodes
- Minimized edge crossings
- Edges of equal length
- Evenly spaced nodes
- Visual representation or emphasis of clusters (e.g., intra-cluster edges are shortened, inter-cluster edges are lengthened)
- Space-filling algorithms
- Node-area awareness

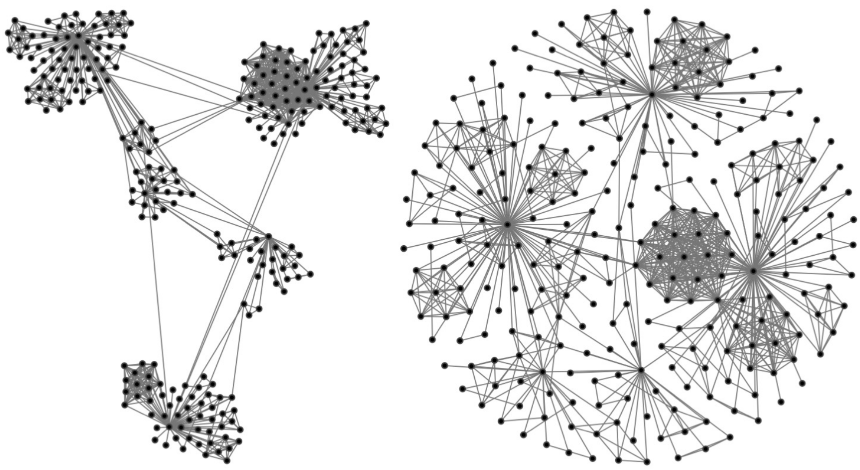

Fig. 5. Two visualizations of the same network data. The layout algorithm on the left prioritizes clusters, while the layout algorithm on the right prioritizes even node distribution.

The most widely studied aesthetic properties for network visualizations have been edge crossings and the angles created by those crossings, as these features have been found to have a large impact on topological literacy, and different layout algorithms vary greatly in their performance on these aesthetic properties.

Seminal work by Purchase and colleagues (Purchase, 1997, 2000; Purchase, Carrington, & Allder, 2002; Purchase, Cohen, & James, 1997; Ware, Purchase, Colpoys, & McGill, 2002) manipulated and tested a series of aesthetic properties of network visualizations to determine user performance on three tasks for finding: a) the length of the shortest path between two nodes, b) the number of nodes that need to be removed to destroy a path between two nodes, and c) the number of edges that need to be removed to destroy a path between two nodes. Through a sequence of related studies, Purchase and colleagues (Purchase, 1997, 2000; Purchase et al., 2002; Purchase et al., 1997) systematically investigated the effects of edge bends, edge crossings, layout symmetry, angles between links as they leave a node, and use of an orthogonal grid for nodes and links. Results consistently emphasized that higher numbers of edge crossings and high numbers of edge bends generally reduce performance, measured via task accuracy and response time. Related work by Huang and colleagues (Huang, 2013, 2014; Huang, Eades, Hong, & Lin, 2013; Huang & Huang, 2011; Huang, Huang, & Lin, 2016) supports these results and suggests that edge crossings with small angles, especially, inhibit performance (measured by accuracy, response time, and self-reported mental effort) on tasks that require users to follow paths.

As a follow-up to the original studies by Purchase and colleagues, Ware et al. (2002) introduced the concept of path continuity, or the lack of abrupt changes in direction of the path. This study focused on a single task – length of the shortest path between two nodes – and found that response time on this task increased as a result of the following changes (in order of influence): increase in number of edges in shortest path, decrease in continuity of shortest path, increase in number of crossings on shortest path, and increase in number of branches off nodes in the shortest path. This suggests that for tasks requiring users to follow a path, anything increasing the number of additional candidate paths or that makes it harder to focus on the shortest path will increase the time needed to complete the task. Rather than focusing on properties of the entire network visualization, it may make more sense to optimize visualizations for specific tasks and the aesthetic properties that will make those tasks easier.

Effect of data overlay design choices

Regardless of the layout algorithm used, other basic graphic design properties that apply to all visualizations should be considered when designing network visualizations. A series of core perceptual studies have addressed basic human perceptual abilities as they relate to interpreting information visualizations. Both early studies and more recent replications (Cleveland & McGill, 1985; Heer & Bostock, 2010) suggest that humans have aptitude for comparisons related to position in space and length of an object. Accuracy suffers when tasks require comparisons of area (like comparisons between two circles) or color value (like the comparisons between two shades of red). Node-link diagrams employ only relative positioning, and even those relative positions are the result of algorithms that may not have an optimal solution for a 2D visualization. In data visualization, there is often a tension between accuracy and aesthetics. The differences in positions in a node-link diagram are not meant to be interpreted with great accuracy, despite human acuity for position comparisons. Conversely, node-link diagrams often employ size- and color-coding to emphasize topological data in the visualization despite our relatively low acuity with those visual encodings. These mismatches between node-link diagrams and our basic human perceptual systems suggest challenges for the use of network visualizations without supplemental numerical information.

Effect of network data properties

The basic properties of a network dataset can also have a large impact on the effectiveness of the visualization. Ghoniem, Fekete, and Castagliola (2005) compared task performance of users viewing matrices and node-link diagrams, varying the size and densities of sample data sets. They found performance on all experimental tasks deteriorated for node-link diagrams as the size increased from 20 nodes to 50 nodes, and again between 50 nodes and 100 nodes. Increases in density between 0.2 and 0.6 had mixed effects on task performance. They concluded that certain tasks are much harder with high-density networks, while others show no significant drop in accuracy as density increases. Similarly, Purchase, Cohen, and James (1997) found that an increase in density of node-link diagrams relates to a decrease in accuracy on tasks dealing with the connectivity of a network.

Teaching Network Visualization Literacy

As reviewed above, research within the visualization community focuses primarily on experimental studies of network visualization comprehension, limited to specific pre-determined tasks. A more robust understanding of network visualization literacy must also take into account both how users understand network visualizations when they encounter them in their daily life and how individuals can gain the expertise necessary to produce their own network visualizations. Thus, a combination of formal and informal education is desirable for empowering many to read and make network data visualizations. Here we present and discuss three existing approaches: Connections: the Nature of Networks (a public science museum exhibition at the New York Hall of Science); NetSci High (a research program for high school students sponsored by Boston University, Binghamton University, USMA West Point, and the New York Hall of Science); and the Information Visualization MOOC course at Indiana University.

Network Visualization in Informal Learning Environments

Informal learning environments (which includes unstructured learning opportunities such as museums, and personal learning) provide opportunities for acquainting the public with network visualization to increase NVL. Because of the unstructured nature of these environments and the relative novelty of the use of network visualization as a tool for understanding complex systems, significant scaffolding is required for effective learning and knowledge transfer.

Connections: The Nature of Networks

The first public museum exhibition on network science was developed by the New York Hall of Science in 2004 (Uzzo & Siegel, 2010). The pedagogical goal of this exhibition was to acquaint museum visitors with the fundamentals of network science, including the basic ways networks are represented as a series of links and nodes, but also, the generalizability and value of how most kinds of complex connected systems can be represented as networks, the benefits of these kinds of representations, and a basic characterization of complex network concepts (small worlds, scale-free properties and emergence). To address a very diverse audience (including all ages), it was theorized that the experience overall should engage

Fig. 6. *Ropes and Pulleys* (left) conveys the complexity and dynamic of networks. Visitors turn the wheels to change the topology of the pulleys and ropes creating clusters and isolated nodes. NEAR (bottom right) simulates the dynamics of social networks using nearest neighbor algorithms.

visitors in network concepts in a variety of ways, including visual representations, sound, and embodied or physical interaction with networks and network concepts.

A significant challenge to developing this experience was that the notion of networks as a general principle was a new idea to visitors. In a preliminary visitor study (Cohen, 2002) visitors could readily identify computer and communications networks with little or no prompting, but they could not readily identify networks in either social or natural contexts. A summative evaluation of the exhibition (Rothenberg & Hart, 2006) indicated an increase in the number of visitors who identified networks in a broad spectrum of applications, particularly environmental and social and indicate that networks are a way to understand the world through the *Connections* exhibition. To achieve these outcomes, however, required intervention by floor staff to explain relevant network ideas represented in the exhibition. By far the most popular and effective aspects of the experience were where visitors physically interacted with networks.

On balance, beyond positive affect and recognition of the ubiquity of networks, there was little transfer of knowledge. It was a much more difficult task for visitors to deepen their understanding of the properties of networks. The implications of this work for NVL is that effective engagement of museum visitors in complex network ideas: a) is deeper when visualization is combined with hands-on activities where the visitor is engaged in network concepts, b) requires intervention of floor staff for understanding specific network properties, and c) exists within a reality of an overall lack of understanding of networks among a diverse visitorship, indicating the need for more learning opportunities for the public about the importance and utility of networks.

Network Visualization in Formal Learning Environments

For the sake of this chapter, formal learning environments circumscribe teaching and learning in primary, secondary, and post-secondary school-based environments. The following section details NVL activities in both university and high school classrooms.

NetSci High

Network visualizations have been taught at the university level for some time, but secondary educational environments only recently started exploring this topic. As with the introduction of any new idea, finding a way to fit networks into existing curriculum is difficult for teachers, who are accountable primarily for their student's performance on standardized tests. Because networks align well with mathematics and science content standards through the Common Core (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and Next Generation Science Standards (NGSS Lead States, 2013), it is reasonable to infuse these ideas into curriculum, create professional development opportunities, and also student and teacher research through guidance by researchers and university faculty.

NetSci High started in 2010 and is a first of its kind program to train high school teachers and students in network analysis techniques and have them apply it to research mentored by university researchers and graduate students (Cramer et al., 2015). An important aspect of this program is training of students and teachers, which between 2012 and 2015 took the form of a 2-week "boot camp" at Boston University, where teachers and students were immersed in network concepts and trained in tools to equip them for mentored research during the school year. Research projects culminate in display and defense of research at the International School and Conference for Network Science. Findings from the evaluation of the *Connections* exhibition at the New York Hall of Science were useful to inform the development of the training and the role of network visualization in the program. Specifically, this involved exposing participants to a wide variety of applications of networks and how they are visualized, as well as providing hands-on and embodied ways to demonstrate and engage with network concepts. Central to the training was skills development in the use of network analysis tools. NetworkX and Gephi were the primary tools taught, and various other analysis environments and techniques were included in the actual research phase. Program evaluation by Davis Square Research Associates (Faux, 2015) indicated significant gains in the understanding of a) the value of network visualization and its role in analysis of complex networks, b) the intimate relationship between analysis and visualization, c) the process of representing a variety of network attributes, which can be accomplished through a variety of tools, and d) the importance of an intensive approach to teaching novices network visualization as a tool to analyze and communicate findings in network science.

Information Visualization MOOC

In a long-running teaching and research program at Indiana University, teaching university students to understand and create network visualizations began by developing a systematic process for designing effective visualizations. This framework for creating visualizations has then been embedded into course structure, books, activities, software, and digital teaching aids, all of which allowed the graduate-level Information Visualization course to expand into a Massive Open Online Course (MOOC).

Frameworks for Network Visualization Education

Börner (2015) proposes a general process for converting data into a visualization (Figure 7), each step of which is based on an analysis of what the users of the visualization need and want from the visualization. First, data need to be parsed and

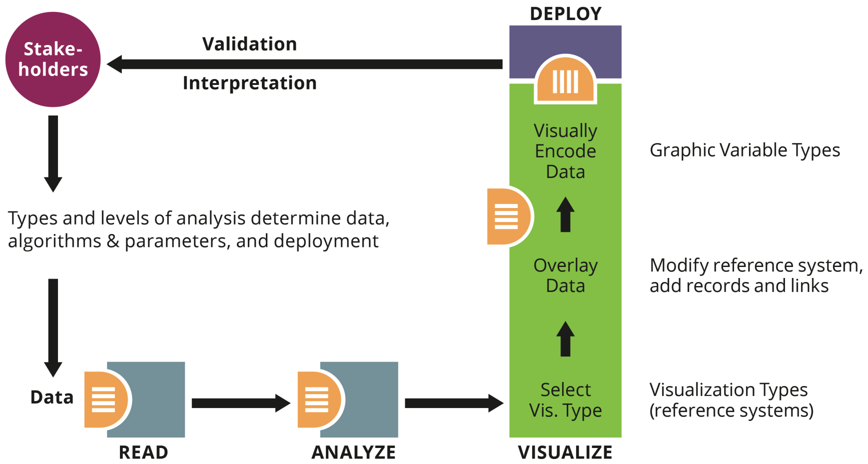

Fig. 7. Needs-driven workflow design with science map network example on right.

read (READ). Extensive cleaning and preprocessing might be needed. Temporal, geospatial, topical, and network analyses might be performed to identify trends and patterns (ANALYZE). The visualization phase (VISUALIZE) comprises three major steps. First, the appropriate reference system must be identified. This reference system becomes the stable base map onto which data are layered. Second, the reference system might be modified (e.g., an axis may undergo a logarithmic transformation). Third, additional data variables are visually encoded using diverse graphic variable types. Ultimately, the visualization must be deployed (DEPLOY) (i.e., printed, published online, etc.) Last but not least, the visualization is presented to stakeholders for validation and interpretation. Frequently, new visual insights lead to new questions, requiring additional data analysis and visualization—the cycle repeats.

This detailed formulation of different steps involved in visualization design open each step up for the critical discussions that are necessarily for gaining data visualization literacy. For example, different needs from stakeholders, combined with different properties of the data, will lead to different visual design recommendations. In the case of network visualizations, the framework is especially helpful in guiding students through a series of design choices that are easy to dismiss as arbitrary because of the lack of standardized guidelines and training within the broader information visualization community.

The Information Visualization MOOC (IVMOOC) (CNS Center at Indiana University, 2017) is a graduate-level course that has been continually developed and taught at Indiana University (IU) since 2013. Most students, hailing from over 100 countries, take the course for free to earn a personalized letter of accomplishment and digital Mozilla badge. Additionally, in Spring 2016, more than 120 students registered for three IU credits as part of the Information and Library Science

M.S. program and the online Data Science M.S. Program offered by the School of Informatics and Computing.

Course Structure

The IVMOOC course aims to improve data visualization literacy—the expertise and skills needed to read and make data visualizations. It teaches theoretical foundations and advanced tools that help turn data into insights.

The course uses a combination of hands-on case studies showing how to read, analyze and visualize; theory lectures; client projects; homework assignments; and exams to empower students to design effective visualizations that take the needs of users into account. In the first week of the course, students are introduced to the visualization framework, which is used to structure the course's schedule and exams, textbook (Börner & Polley, 2014), tools, and an IVMOOC flashcard app (discussed more below). In weeks two to six students use the framework to learn about a variety of types of visualizations, including network visualizations. In the last seven weeks of the course, students collaborate on real-world projects for a variety of clients. Results from previous student projects are published in Börner and Polley (2014).

Each unit includes theory and hands-on sections. Each theory section comprises:

- Examples of exemplar visualizations
- Visualization goals
- Key terminology
- General visualization types and their names
- Workflow design, and
- Discussion of specific algorithms

Each hands-on section guides students through user- and task-analysis; data preparation, analysis, and visualization; deployment; and the interpretation of visualizations. The sections feature in-depth instruction on how to navigate and operate several software programs used to visualize information. Furthermore, students learn the skills needed to visualize their own data, allowing them to create unique visualizations.

The theory component and the hands-on component are standalone, meaning that participants can read/watch whichever section they are more interested in first, and then review the other section. After the theory videos there are self-assessments, and after the hands-on videos are short homework assignments.

Textbook

The *Visual Insights* textbook (Börner & Polley, 2014) was designed as a companion resource for students taking the IVMOOC. It contains all theory and workflows covered in the course. While the *Atlas of Knowledge* (Börner, 2015) aims to feature timeless knowledge, or principles that are indifferent to culture, gender, nationality, or history, the IVMOOC and associated textbook cover "timely knowledge," or the most current data formats, tools, and workflows used to convert data into insights.

Analogous to the IVMOOC course, Chapter 1 introduces the visualization framework intended to help non-experts assemble advanced analysis workflows and design different visualization layers. It also showcases how the framework can be applied to "dissect visualizations" for optimization or interpretation. Chapters 2–7 in the textbook introduce the different types of analysis: temporal (*when*), geospatial (*where*), topical (*what*), and trees and networks (*with whom*). Chapter 8 presents exemplary case studies that resulted from IVMOOC real-world client projects.

Software

Every student who registers for the IVMOOC gets experience using the Sci2 Tool (Sci2 Team, 2009), a software application for data analysis and visualization developed by Börner at IU. The NSF-funded tool has been in development since 2008 and benefits from more than 10 years of tool development and feedback from many of the more than 150,000 tool users in academia, industry, and government. The tool supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels. It implements the visualization framework to help users assemble more than 180 algorithms into proper workflows. Specifically, it organizes the main menu structure by workflow steps (from reading and preprocessing data to analyzing and visualizing data and saving out results) and by visual analysis type (Temporal, Geospatial, Topical, Networks) using the visualization framework discussed above.

Flashcard App

Visualization designers and users must have a basic understanding of different visualizations – their types and the visual encodings used. They must be able to recognize and name visualizations in order to refer to and talk about visualizations. The IVMOOC Flashcard app lets users browse more than 60 information visualizations. Users swipe to navigate through visualizations, pinch in/out to zoom, and tap to turn the card to access information about: name of the visualization, visualization type (e.g., graph, map, network layout), visual encoding used (graphic symbol types, and graphic variable types), and reference to additional information provided in the *Atlas of Knowledge* (Börner, 2015). The Flashcard app, created in Unity3D, supports both Android and iOS.

What are our best ways forward?

Based on our review of relevant research and experiences with teaching and learning with network visualization in formal and informal settings, we make the following recommendations for improving network visualization literacy:

- Use data that are meaningful to the learner. Data that are personally relevant to, directly collected from/by or selected by the learner will increase their engagement and familiarity with the data. Similarly, when asking questions of the data, instructors are urged to pick tasks that make sense for network visualizations but also those that make sense for the selected data and for the research questions the user is investigating.
- When introducing novices to networks and visualization techniques, best practices would suggest the use of small networks with low density to increase understanding.
- When designing visualizations, leverage core perception mechanisms by following Gestalt grouping, continuity, and proximity principles. Additionally, endorsing certain types of aesthetic principles like minimal edge crossing and path continuity should improve the likelihood of understanding by the public.
- Following best instructional practices, educators should engage novices in low complexity tasks with greater support and move toward higher complexity tasks and networks, withdrawing support as the learners gain competence.
- Network scientists need to provide explicit instruction on how readers/users should read the visualizations they create. Clearly outlining what conclusions are and are not valid will help users in interpretation. This should include use of standardized terminologies to leverage prior knowledge.

Discussion

In conclusion, there is ample evidence that network visualization is an important tool for understanding complex connected systems. But it is important that it be thoughtfully combined with other pathways into understanding what networks are, their characteristics and behaviors.

To enact these recommendations and advance the nascent research on NVL, we invite close collaboration with others on developing both widely adoptable visualization frameworks that can be used to teach information visualization theory and methods, and also custom development of a more refined and meaningful definition and framework for NVL. Efforts must be made to develop guidelines that recommend skills and learning outcomes and competencies for both learners that have taken information visualization courses in formal settings and a wide audience of citizens and policymakers. Additionally, the visualization community should work together openly to standardize terminology, theoretical frameworks and visualization techniques. This work should involve the development, testing, and implementation of course designs, tools, materials and activities to increase student competency with interpreting and implementing visualizations, preparing them to evangelize these methods and practices in research, practice, and training. Finally, open data, open code, and open education are true enablers that can empower anyone to convert data into visual insights.

Acknowledgements

This work was partially supported by the National Institutes of Health under awards P01 AG039347 and U01CA198934 and the National Science Foundation under awards NCSE 1538763, EAGER 1566393, NRT 1735095, AISL 1713567, and NCN CP Supplement 1553044. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Bennett, C., Ryall, J., Spalteholz, L., & Gooch, A. (2007). The Aesthetics of Graph Visualization. In D. W. Cunningham, G. Meyer, & L. Neumann (Eds.), *Computational Aesthetics in Graphics, Visualization, and Imaging* (pp. 57-64): The Eurographics Association.

Börner, K. (2015). *Atlas of knowledge: Anyone can map*. Cambridge, MA: MIT Press.

Börner, K., Balliet, R., Maltese, A. V., Uzzo, S. M., & Heimlich, J. E. (2015). *Meaning Making Through Data Representation Construction and Deconstruction*. Paper presented at the AERA 2015 Annual Meeting, Chicago, IL.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*(1), 179-255.

Börner, K., Maltese, A. V., Balliet, R., & Heimlich, J. E. (2016). Investigating Aspects of Data Visualization Literacy Using 20 Information Visualizations and 273 Science Museum Visitors. *Information Visualization, 15*(3), 198-213.

Börner, K., & Polley, D. E. (2014). *Visual Insights: A Practical Guide to Making Sense of Data*. Cambridge, MA: The MIT Press.

Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology, 25*, 163-177.

Card, S., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization : using vision to think*. San Francisco: Morgan Kaufmann Publishers.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 299*(4716), 828-833.

CNS Center at Indiana University. (2017). IVMOOC: Information Visualization MOOC 2017. Retrieved from http://ivmooc.cns.iu.edu

Cohen, S. (2002). *Connections The Nature of Networks: Front End Evaluation*. Retrieved from Program Evaluation and Research Group:

Cramer, C., Sheetz, L., Sayama, H., Trunfio, P., Stanley, H. E., & Uzzo, S. (2015). *NetSci High: Bringing Network Science Research to High Schools*. Paper presented at the Complex Networks VI: Proceedings of the 6th Workshop on Complex Networks CompleNet 2015, New York.

Eliassi-Rad, T., & Henderson, K. (2010). Literature search through mixed-membership community discovery. In S.-K. Chai, J. Salerno, & P. L. Mabry (Eds.), *Advances in Social Computing: Third International Conference on Social Computing, Behavioral Modeling and Prediction, SBP10* (pp. 70-78). Bethesda, MD: Springer.

Fabrikant, S. I., & Montello, D. R. (2008). The effect of instructions on distance and similarity judgements in information spatializations. *International Journal of Geographical Information Science, 22*(4), 463-478. doi:10.1080/13658810701517096

Fabrikant, S. I., Montello, D. R., Ruocco, M., & Middleton, R. S. (2004). The Distance-Similarity Metaphor in Network-Display Spatializations. *Cartography and Geographic Information Science, 31*(4), 237-252.

Fabrikant, S. I., Ruocco, M., Middleton, R., Montello, D. R., & Jörgensen, C. (2002). *The first law of cognitive geography: Distance and similarity in semantic space*. Paper presented at the GIScience 2002, Boulder, CO.

Faux, R. (2015). *Evaluation of the NetSci High ITEST Project: Summative Report*. Retrieved from Boston:

Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2005). On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Information Visualization, 4*(2), 114-135. doi:10.1057/palgrave.ivs.9500092

Heer, J., & Bostock, M. (2010). *Crowdsourcing graphical perception: using mechanical turk to assess visualization design*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA. http://dl.acm.org/citation.cfm?doid=1753326.1753357

Huang, W. (2013). An aggregation-based overall quality measurement for visualization. Retrieved from https://arxiv.org/abs/1306.2404

Huang, W. (2014). Evaluating overall quality of graph visualizations indirectly and directly. In W. Huang (Ed.), *Handbook of Human Centric Visualization* (pp. 373-390). New York: Springer-Verlag.

Huang, W., Eades, P., Hong, S.-H., & Lin, C.-C. (2013). Improving multiple aesthetics produces better graph drawings. *Journal of Visual Languages & Computing, 24*(4), 262-272. doi:10.1016/j.jvlc.2011.12.002

Huang, W., & Huang, M. L. (2011). Exploring the relative importance of number of edge crossings and size of crossing angles: A quantitative perspective. *International Journal of Advanced Intelligence, 3*(1), 25-42.

Huang, W., Huang, M. L., & Lin, C.-C. (2016). Evaluating overall quality of graph visualizations based on aesthetics aggregation. *Information Sciences, 330*, 444-454.

Maltese, A. V., Harsh, J. A., & Svetina, D. (2015). Data Visualization Literacy: Investigating Data Interpretation Along the Novice—Expert Continuum. *Journal of College Science Teaching, 45*(1), 84-90.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards. Retrieved from http://www.corestandards.org/

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Purchase, H. C. (1997). *Which aesthetic has the greatest effect on human understanding?* Paper presented at the Graph Drawing. GD 1997. Lecture Notes in Computer Science, vol 1353.

Purchase, H. C. (2000). Effective information visualisation: a study of graph drawing aesthetics and algorithms. *Interacting with Computers, 13*(2), 147-162.

Purchase, H. C., Carrington, D., & Allder, J. (2002). Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering, 7*(3), 233-255.

Purchase, H. C., Cohen, R. F., & James, M. I. (1997). An experimental study of the basis for graph drawing algorithms. *Journal of Experimental Algorithmics, 2*, No. 4. doi:10.1145/264216.264222

Rothenberg, M., & Hart, J. (2006). *Analysis of Visitor Experience in the Exhibition Connections: the Nature of Networks at the New York Hall of Science*. Retrieved from Northampton, MA:

Sayama, H., Cramer, C., Porter, M. A., Sheetz, L., & Uzzo, S. (2016). What are essential concepts about networks? *Journal of Complex Networks, 4*(3), 457-474. doi:10.1093/comnet/cnv028

Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.

Uzzo, S., & Siegel, E. (2010). Connections: The Nature of Networks, Communicating Complex and Emerging Science. In A. Filippoupoliti (Ed.), *Science Exhibitions, Communication and Evaluation*. Edinburgh: MuseumsEtc.

Ware, C. (2013). *Information visualization: perception for design* (3rd ed.). Waltham, MA: Morgan Kaufmann Publishers.

Ware, C., Purchase, H. C., Colpoys, L., & McGill, M. (2002). Cognitive measurements of graph aesthetics. *Information Visualization, 1*(2), 103-110. doi:10.1057/palgrave.ivs.9500013