Using Network Visualizations to Engage Elementary Students in Locally Relevant Data Literacy

SCHOLARONE™
Manuscripts

Using Network Visualizations to Engage Elementary Students in Locally Relevant Data Literacy

Abstract

Purpose – We explore how network visualization provides opportunities for learners to explore data literacy concepts using locally and personally relevant data.

Design/methodology - We designed six locally relevant network visualization activities to support students' data reasoning practices toward understanding aggregate patterns in data. Cultural Historical Activity Theory (Engeström, 1999) guides our analysis to identify how network visualization activities mediate students' emerging understanding of aggregate datasets.

Findings - Pre/post-test findings indicate that this implementation positively impacted students' understanding of network visualization concepts, as they were able to identify and interpret key relationships from novel networks. Interaction Analysis (Jordan &Henderson, 1995) of video data revealed nuances of how activities mediated students' improved ability to interpret network data. Some challenges noted in other studies, such as students' tendency to focus on familiar concepts, are also noted as teachers supported conversations to help students move beyond them.

Value – This is the first study we are aware of that supported elementary students in exploring data literacy through network visualization. We discuss how network visualizations and locally/personally meaningful data provide opportunities for learning data literacy concepts across the curriculum.

Keywords: *Network visualization, Data literacy, Mediation, Interaction Analysis*, Locally/personally relevant data

Introduction

The massive amount of information generated everyday has driven an increasing need for people to comprehend and utilize data. Traditional data literacy teaching relied on data devoid of context, emphasizing quantitative reasoning skills using pre-existing data sources (Acker & Bowler, 2018), while more recent initiatives prioritize data relevant to students' daily lives (e.g., Acker & Bowler, 2018; Bowler et al., 2017; Deahl, 2014; Rubin, 2020). These efforts have also expanded to include elementary-aged students (Bowler et al., 2017; Jiang et al., 2022). However, it is challenging for young learners to understand how data may be relevant to their lives. Previous research employing student-generated data from social networks (e.g., Twitter) revealed that, despite achieving a general understanding of data lifecycles (e.g., data creation, collection, etc.), students struggled to establish personal connections with their data (Bowler et al., 2017). Additionally, young students find it difficult to shift attention from isolated/individual data points toward meaningful inferences from patterns across the entire dataset (Ben-Zvi & Arcavi, 2001).

This study leverages the value of data representations, particularly network visualizations, to make obscure data literacy concepts salient to students. We regard networks as a promising entry point for young students to learn about both dimensions of data literacy supported by regular data graphs (e.g., bar charts) and other dimensions that network visualizations specialize in, such as interrelationships between different data points. This study explores network visualization as a unique approach to support young learners' data literacy development without attending to challenging mathematical concepts and their exploration of relationships in locally relevant data. Our goal is not to replace other data representations but to provide a potentially powerful alternative for learners who may not find those approaches interesting/approachable. To explore the potential of network visualization in supporting data literacy, we involved fifth and sixth graders in a three-week curriculum where they explored multiple locally relevant networks using an open-source network visualization tool (Net.Create; Authors, 2018). To examine if and how the network visualization tool and accompanying activities helped students develop an understanding of network visualization and data literacy, we asked:

1. Did students show an improved understanding of network visualization of data in post-tests compared to pre-tests?
2. How did network-visualization-centered curriculum mediate students' ability to understand data literacy concepts and practices?

Theoretical framework

The present analysis uses the concept of mediation from Cultural Historical Activity Theory (Engeström, 1999) to examine how different network visualization activities shaped students' emergent understanding of data literacy. Mediation refers to the idea that activities are shaped by social and cultural elements in the environment (Vygotsky, 1978; Wertsch, 1981). These mediators- including tools, rules, and the community, can impact how learners

understand the object/shared goals they are pursuing and how they pursue them (Engeström et al., 1990). Examining mediators can help researchers better understand how the design helped learners succeed and how it might be refined in the future (Authors, 2022). Mediation is evident in the video-based analysis of learning activity as learners shift their actions in response to mediators and comment on mediators' impacts (Authors, 2014). For example, this study asked students to indicate their interests and explore peers' interests using Net.Create. Despite months of acquaintance, students still discussed previously unknown interests of their peers, indicating the network visualization tool influenced how they saw each other. This study explores how six network activities mediated students' exploration of data and data literacy concepts, helping them to see information in new ways as they worked with visualizations.

Data literacy and network visualization

Data literacy was defined by Curcio (1987) as the ability to read data, read between data points, and read beyond the data (as cited in Friel et al., 2001). This framework was expanded by Friel et al. (2001) as "extracting information from the data, finding a relationship in the data, and moving beyond the data" (p.131). Calzada Prado and Marzal (2013) further note the value of data collection and structuring skills before data is "read." Our work builds on this view of working with a full data cycle, aiming to understand how learners benefit from creating datasets and interpreting relationships within both self-created and second-hand datasets. One way to examine students' ideas about data is to characterize their descriptions of relationships they see in data, including individual data points, specific cases, and how they can be viewed together to convey broader ideas being represented (Konold et al., 2015). In many implementations, these ideas are rooted in basic statistics (e.g., mean) to describe aggregate patterns in data mathematically. While we recognize the value of such descriptions, our goal in this study was to begin with qualitative descriptions of aggregate patterns revealed in network visualizations, helping learners to go beyond individual data points to understand the datasets. Activity theory also highlights the need to attend to learner's goals of working with data, as their goals are presumed to inform the use of data concepts. This aligns with Calzada Prado and Marzal's (2013) definition of using data, emphasizing synthesizing and representing data results for inquiry purposes and applications. Building on these previous definitions, we conceptualized data literacy as a set of skills of creating/modifying data, extracting information from the dataset, finding relationships in the data, categorizing data, using data, and data implications.

Data visualization and visualization literacy are also crucial components of any definition of data literacy (Deahl, 2014; Rubin, 2020). Visualizations enhance students' comprehension and reasoning of complex phenomena by offering contextually-rich information for explanation and argumentation (Roberts et al., 2014; Shreiner & Guzdial, 2022). Researchers designing visualizations to support students in extracting and interpreting data patterns have found that visualization literacy facilitates data literacy development (Alper et al., 2017; Bishop et al., 2020).

One common example of a visualization that helps learners see the relationships between ideas in data is concept maps, which typically depict ideas as circles connected together by lines (Cañas, 2003; Schwendimann, 2015). However, data literacy demands more than identifying relationships as in concept maps, involving perceiving patterns in large-scale complex datasets that frequently mix quantitative and qualitative information and are presented in both static and dynamic forms (Cramer et al., 2015). Network visualizations utilize graph theory, dynamic system theory, and statistical analysis to represent patterns in complex large datasets for users to solve data-intensive problems in the real world (Cramer et al., 2015). Network visualization tools typically include nodes representing individual data points that are connected to each other by edges/lines representing a relationship (Figure 1). Statistical information about nodes and edges provides some built-in datacentric features of automated data visualization (e.g., auto-sizing nodes), which can make some hard-to-notice relationships and patterns more visible (Ahnert et al., 2020). In short, network visualizations are powerful vehicles for students to learn analytical and computational skills for data analysis (Cramer et al., 2015) and leverage those skills to solve real-world problems using large-scale data (Pastor-Satorras & Vespignani, 2001). The present study hypothesized that functionalities of network visualization tools could facilitate students to recognize relationships and patterns in the dataset.

Many network visualization tools for data analysis rely on systematically and automatically generated datasets. For instance, studies have explored visualizations showing global scientific collaboration based on the Web of Science database (Börner et al., 2003). However, we recognize that for students to represent their interests, it might be valuable to support a more fluid data creation process where students enter their own data and that acknowledges the subjectivity that captures students' choices, not objectively complete sets of information. To help us think about this process, we drew on Drucker's (2015) concept of *capta,* which delineates a difference between *data as* information "given" to researchers as downloaded and uncontested and *capta* as the process of negotiation, interpretation, and contestation that is implicated when unstructured sources are transformed into structured, defined fields. For example, we did not pre-define what interests or hobbies students could link themselves to. Instead,

students in the study drew on their own interpretation of what "interests/hobbies" meant to them and negotiated and reasoned with each other about what it meant to create a "hobby" node and then attach an edge to that node. The "traditional" definitions of network node-and-edge data provided an epistemic framework within which students could reason about, critique, and transform information into structured data.

Our goal was to combine these ideas about creating and negotiating data with the previous set of skills around data creation and interpretation and the necessity to view both individual data points and the aggregate patterns within the datasets. In the context of network visualization, this means moving between interpretations of a node and an entire network, and we view it as a network-specific parallel of data literacy practices comparing single-data-point perspectives with aggregate dataset patterns. Table I summarizes literature sources of each subset of data literacy skills and manifestations in network visualization practices that we focused on in this study.

Personally meaningful data literacy research

Interpreting data visualizations can be intimidating without context/content understanding (Shreiner & Guzdial, 2022). Most statistics classes present data with minimal context; however, data derives meanings from contexts and interpreters (Acker & Bowler, 2018). Research in information visualization has shown the advantages of utilizing students' experience to enhance visualization understanding (Bae et al., 2022). Work that situates students' data reasoning within personally meaningful data generally falls into two categories: student-created data and pre-existing data of potential interest to students.

Student-created data

The first category positions students as data creators. Lee et al. (2021) note the value of students' direct experiences with data collection for enhancing data interpretation skills. Van Wart et al. (2020) showed that integrating student-contributed data (e.g., photos and drawings) into a participatory digital mapping tool allowed them to leverage community knowledge for understanding data valences and applications through practices, including sampling data and visualizing results. Similarly, students in Stornaiuolo (2020), who constructed meaningful data narratives about their habits and media uses, came to view data as a versatile tool with contextual relevance to address issues of personal interests and support learning in personally meaningful manners. Additionally, students in Lee & Drake (2013) understood the sensitives of different measures of the center using their recess physical movement data. Bergner et al. (2021) developed dancer students' statistical reasoning of variance and periodicity using their own dance movement. Building on these studies, this study aimed to help learners forge similar connections to their data by providing opportunities to represent and explore their interests in network visualizations.

Pre-existing data of potential interest to students

While using student-created data is valuable, the ability to make sense of second-hand data is crucial to authentic data literacy practices (Duschl, 2008). Second-hand data can also be chosen and presented in personally relevant ways, a process referred to as personalization strategies (Robert & Lyons, 2020). Many studies personalize data activities by either selecting data relevant to students' interests or combining second-hand datasets with data that students themselves provide. Khan (2020) examined how students' use of open large-scale datasets to model family autobiographies enhanced their thinking about aggregate social structures in relationship to individual family histories and understanding of complex comparative logic and bias in data representations. The integration of open datasets with personal stories is also evident in Lopez et al. (2021), which involved students analyzing data patterns and linking them to issues of nutrition and climate change in local communities. Similarly, students in Calabrese Barton et al. (2021) engaged in the learning of big data and small data divide by looking for connections between their local context and large datasets in predicting the impacts of COVID-19 on their community. The present study built on personalization strategies, easing initial challenges posed by unfamiliar data by providing learners with opportunities to control, change, and interpret the data in personally meaningful ways. We worked closely with collaborating teachers to ensure all second-hand data provided to students was locally meaningful for them based on their interests and/or ongoing classroom activities.

Method

The current study is part of a larger project called Visualizing Funds of Identity (VFOI), which intends to leverage network visualization to help students understand more about themselves and their community as they hone their fundamental data literacy (c.f., Authors, 2023). The present analysis focuses on students' explorations of data literacy. The participants were twenty-two fifth and sixth graders from the Midwestern United States who had not received data literacy instruction before, nor had their teachers covered data literacy concepts outside this

implementation. Due to absences, seventeen students completed both pre- and post-tests. This study uses pre/post-test performance and classroom video data to explore how students' in-situ understanding of network-data literacy emerged from classroom activities. We view the network visualization competencies as tied to and building on broader data literacies and, therefore, focused pre/post-tests on networks to keep them brief and approachable.

Design

The VFOI project utilizes Net.Create (Authors, 2018, URL) an open-source network visualization tool for learners to simultaneously enter, represent, and view complex data as a network or table (Authors, 2021a; Authors, 2021b). Figure 1 is the identity network created in this study, depicting students' identities in the form of interests and connections to each other (for a dynamic version, see URL). Nodes in this network represent individual students, locations, hobbies, and significant people, while edges represent relationships such as "likes/interested in." These are both sized and placed using a force-direction algorithm common to many network visualization tools. Nodes with many connections, or a high degree-centrality measure, are automatically larger and more central in the network. If two nodes have more than one shared edge, the edge will be thicker and draw the nodes closer. Thus, a node's size and position in a network visualization communicate its relative importance. Net.Create differs from other network analysis tools in its capacity for simultaneous data entry and real-time adjustment of the visualization for all users currently viewing the network. We coordinated with two teachers to identify locally meaningful and relevant ideas, around which we designed six network visualization activities to engage students in a cycle of data creation, modification, and exploration of networks in physical and digital forms to support their data literacy development.

The curriculum began with students constructing a physical network using yarn to connect themselves to peers who shared a similar interest. Each student acted as a node representing themselves, and the yarn symbolized edges connecting students with overlapping interests. To create the edges, one student answered a question about themselves (e.g., interests) and then passed the ball of yarn to classmates who raised hands to indicate similar answers/experiences. This activity was designed to enable students to observe connections with their peers by constructing a physical network that represents them, thus preparing them for subsequent activities using digital networks. Activity 2 involved students in pairs/trios to create new nodes/edges describing themselves in a researcher-built Net.Create yarn network. Facilitators provided instructions on node/edge creation to the class before grouping students. This activity introduced basic network concepts and vocabulary through creating/modifying network visualization, fostering students' recognition of networks in making relationships more salient.

Activity 3 and 4 used a researcher-built Net.Create network (Figure 2) about the author Kelly Barnhill, whose novel, *The Girl Who Drank the Moon*, was students' mandatory reading. This network contained data from Barnhill's recent Twitter posts. In activity 3, student dyads categorized nodes from the network that were printed on individual pieces of paper without referencing the digital network. We expected prior knowledge would impact students' categorization and inference-making about the network topic. Activity 4 asked students to categorize nodes in the network by editing node attributes (Figure 2), where they could also see peers' categories. They were also encouraged to reference brief node descriptions we added in Net.Create, addressing confusion we had overheard in activity 3. We conjectured that viewing nodes and connections in Net.Create would impact students' categorization and inference-making. These activities were designed to be complementary in fostering students' appreciation of network affordances in supporting categorizations and inference-making.

Activity 5 presented students with a pre-made network modeling the integration of entities in the chicken industry (Figure 3). This topic was chosen because students participated in a designing and managing fictional farms project. Teachers had developed this cross-curricular activity because of the school's proximity to rural farm areas and active locally sourced food community. Student dyads explored the network and voted (by editing each node's attribute) for the three nodes that they felt were "most important" for operating a successful chicken farm. Students were free to establish their own criteria to assess 'importance' and were asked to explain their reasoning. This activity was intended to motivate students to explore and understand entities' connections represented in the network, particularly the way node size represented centrality/multiple connections as a key normative approach to determining the importance of nodes.

Activity 6 engaged students in a researcher-designed board game modeling a social media network (Figure 4). Students were divided into two groups: a fictional social media company and fictional social media members. The company group received a Net.Create graph containing data resembling what a real social media company might access, including who 'liked' what content and who 'followed' whom. Using this information, they aim to share 'content cards' like a cat video with a chosen user to get the content seen by the most users. The user's task was to share/ignore content based on their profile card. Importantly, the user profile card had information not in the network given to the company. This activity aimed to motivate students to use various network features to

understand how big companies use social media users' information for advertising and develop an awareness of how this might be valuable in shaping their social media self-representation.

Analysis

We analyzed students' pre/post-tests for the first research question. The pre/post-test includes a sample network (Figure 5) depicting imaginary individuals' interests and connections, accompanied by fourteen open-ended questions (Figure 6) to evaluate students' understanding of network visualization concepts. We used an emergent, open-coding process to analyze students' pre/post-tests to assess how our implementation changed students' understanding of network visualizations. We took this approach to characterize students' understanding of key network science ideas (e.g., edges, connectedness, centrality) and *how* they talked about each of these ideas. Initially, two coders independently open-coded students' responses on pre/post-tests. One coder identified higher-level concepts that were captured in students' responses. For example, *network components* indicated when a student mentioned nodes or edges, *network features* indicated connectedness, and *network implications* indicated instances where students reasoned about how the network could be used. Concurrently, another coder focused on how students talked about these broader concepts. For example, rather than *connectedness*, their codes noted that the student *talked about a direct connection to sports* or *talked about a series of connections (x follows y, who posted z)*. After this initial coding process, the two coders developed a consensus coding scheme that linked higher-level concept codes to subcodes that described how a concept was talked about in students' answers. Both coders independently applied this coding scheme to the pre/post-tests. They excluded questions if two or fewer students answered meaningfully across the pre/post-tests (e.g., a response other than "idk"). The initial percent agreement was calculated for each code and subcode (n=29). The agreement was above 90% for all codes except *talking about connections broadly* (82.4%) and *network implications* (87.6%). The coders reviewed each of these codes' purpose and resolved the initial discrepancies, resulting in a final agreement of 100%. Next, we looked at each student's answers and assigned one point if a student demonstrated understanding of each of our coded concepts in any question. We summed students' scores for each concept, allowing us to determine how many students understood the concept. Due to the small sample size and the non-normality of the data, we did not conduct statistical analyses and instead provided rich descriptions of patterns in students' emergent understanding.

To answer our second research question, we reviewed classroom video data of all six activities in sequence to understand how students' data literacy emerged and was mediated across network activities. The first author content-logged videos of each activity (Erikson, 2006). These videos included whole classroom activities and discussions, and small group interactions that were recorded through screen recording and/or 360 cameras. In activity 1, we focused on how the entire class constructed the yarn network to assess its mediation of students' early data practice. In activity 2, we focused on small groups and whole-class discussions to examine how Net.Create meditated students' data creation and exploration of data. We prioritized whole class debriefs for activities 3 & 4 because students were asked to share and explain the categories they had chosen, allowing us to understand their reasoning and how the activity mediated their categorization. Similarly, we analyzed students' discussion in activity 5 to understand how using Net.Create to votes with peers mediated their reasoning about the importance of nodes. In activity 6, we analyzed two rounds of discussions along with students' gameplaying mentioned in the discussion to understand the mediational process of their evolving understanding. The first author identified hypothetical general patterns about students' emergent understanding and how this understanding was mediated for each activity based on the combination of logs and initial Interaction Analysis (IA; Jordan & Henderson, 1995). Next, the first author selected several video clips for each day that had optimal audio quality as representatives of the observed patterns and shared them with the research team. The research team re-watched the videos from activity 2 together, using the set of skills in our data literacy definition as codes to identify students' data literacy practice. Table I summarizes how each skill was identified within the network activities, along with illustrative examples that were iteratively refined throughout the data analysis to capture students' data literacy practices across activities comprehensively. After this initial coding process, the research team split up the remaining selected clips to code individually and then shared ideas around the coding with the whole team. In particular, we selected moments when codes were first applied (indicating emergence) and those with the highest code frequency (implying productive interactions) in each activity for collective data analysis sessions where we further verified code applications and summarized findings into narratives. The findings section provides narrative syntheses of our interpretations of students' data literacy development and how that was mediated by network activity.

Findings

Research Question 1: Did students show an improved understanding of network visualization of data in post-tests compared to the pre-tests?

Our analysis suggested that students' understanding of several network visualization concepts improved after the implementation (Table II), and the collaborating teachers also remarked that they felt students had learned a great deal. More students showed an understanding of the seven concepts (network components, connections, edges, node content, network's overall shape, network implications, and centrality) on the post-test than on the pre-test. And the ways in which they utilized these concepts shifted from simplistic to more complex understanding.

[Table II]

Network components were coded when students talked about what could be displayed in network visualization. On the pre-test, four students mentioned at least one network component, while on the post-test, thirteen students did. Additionally, more students discussed people and things being part of networks on the post-test. For example, in response to "write what you think a network is," on the pre-test, one student wrote, "something online that you can get info off of," and on the post-test, they said, "a network is usually something online that shows how people might be connected to certain things."

Post-test results also show students' improved abilities to identify and reason about connections in a network. On the pre-test, ten students identified connections in the sample network, with the majority noticing direct connections (a single edge connecting two nodes). For example, one question asked, "Based on the network above, who do you think would be friends with Luca in real life?" On the pre-test, students most often selected a person who was directly connected to Luca (e.g., "I think Alexis because Luca follows Alexis"). On the post-test, more students identified connections (17), and some students used more complex reasoning about overlapping connections (e.g., "Alexis and Marisol, they are all connected to Mexico City and have the thicker line") and indirect connections.

Eight students on the post-test (versus four on the pre-test) reasoned about edges. Most striking was the increase in referring edge thickness (indicated connections between two nodes). On the pre-test, students focused on networks as something found on the internet, whereas on the post-test, more students reasoned about the provided network's implications. Finally, more students on the post-test claimed who in the provided network was the most popular or had the most influence using ideas of centrality. Students primarily thought that having more connections meant that a person was more popular, but a couple looked at connections to only other people (ignoring interest and content) to make that decision.

Research Question 2: How did network-visualization-centered curriculum mediate students' ability to understand data literacy concepts and practices?

Activity 1: Identity network in yarn activity

This activity supported *data creation* in the physical space and early engagement with recognizing relationships within data. As each student added themselves to the network, they created a new "edge" in the network and, thus, a new piece of data. In creating this data, students unexpectedly justified why they should be connected to the previous students through stories about themselves/their families. For example, one student connected to another because their family was from the same place. Instead of simply stating the same origin, he explained the family history of parents and grandparents. Thus, an outcome of this activity was the students' rich descriptions of the data –not only their own answers to a question but a nuanced description of their connections to others. Students' creation of data was mediated by forming physical connections with the yarn and through engaging in storytelling. This process may also reflect an emergent understanding of how to *find relationships in the data by* focusing on how two people (nodes) were connected. At the end of the activity, students began to categorize the data briefly in their sense-making. When asked what they noticed about the yarn network, one student said that "a lot of people have pets," a claim that required them to view both dog and cat nodes' edges and synthesize them.

Activity 2: Identity network in Net.Create

This activity facilitated students' *data creation* using Net.Create, whose affordances mediated students' increasingly complex data reasoning. Students' *data creation* often began with the practice of *extracting information* by searching for nodes of themselves or group members, asking questions like "Where am I," and "Let me find mine," although they were not required to. Locating personal nodes was normally followed by connecting them to other nodes, like the local city. This shared connection was frequently used as an initial *data creation* move. As students

created edges, prompts in Net.Create (Figure 7) asked them to choose a connection type (i.e., 'important to'). Edge-type prompts mediated students' *data creation by* making salient *relationships in data* and providing chances for them to *modify the data*. For example, in a dyad where one student primarily entered the data while another observed, we noticed that the student who entered the data was asked to *modify* the connection type between the observing student and the "sister" node from "likes/interested in" to "important to." *Data creation* using Net.Create fostered students attending to the node-edge pairs, which is key to understanding how network visualization represents relationships. The process of creating connections between nodes supported students' articulation of their interests in ways that helped them align with the existing data and identifications of overlaps/connections with their peers.

In addition to *finding relationships* between two nodes, Net.Create's simultaneous and automated (i.e., nodes' sizes and position adjustment) data entry function helped students notice network features and others' contributions, which supported their exploration of relationships across the entire dataset. Students had a bird's-eye view of the entire network in Net.Create, which was hard to achieve (if not possible) in the yarn activity. This bird's-eye view fostered a more global perspective of patterns of the entire dataset. As students explored the visualization (e.g., zooming, dragging, etc.) to *extract information* from the network, many expressed excitements that the network's shape looked like a "constellation" or "web." Students also noticed that some nodes' sizes were relatively bigger, "yours are like a big node," some nodes were positioned in the middle, "I'm like the main part, in the middle of everything," and some nodes were on the margin, "yours are just over there." Noticing those *relationship features* of nodes' size and positions could be a stepping-stone for more advanced data reasoning (e.g., making inferences) in later activities. Therefore, Net.Create visualization meditated students shift from a granular connection between their nodes and a target node towards aggregate ingestion of relationships within the network data.

Students' recognition of relationships was also evident in the debrief, in which facilitators' prompts continued to mediate students to move beyond *extracting information* from individual nodes toward *finding relationships in the dataset*. One student responded that many people were connected to food and their state. Students also displayed an emergent understanding of *using* network visualization data, "somebody looking at the network could learn all the things that you have in common and don't have in common."

Activity 3: Author network in paper activity

This activity supported students' skills in *extracting information* and *categorizing data*. Student-created categories include "agreed on/liked," "politics," "book stuff," and "miscellaneous." Their explanations suggested that prior knowledge about familiar node content and peer discussions mediated their initial categories creation. For example, categories of "politics" and "book stuff" could be mediated by their prior exposure to similar topics and readings. The category of "agreed-on/liked" may reflect a pair's shared personal way of data viewing. We acknowledge that students were unfamiliar with some nodes' content, but they were encouraged to ask questions, which sparked productive discussions about social issues (e.g., equity and equality).

Additionally, facilitators prompted students to *use the data* and their *categorizations* to draw inferences about the author and the book's theme. Although various prompts (e.g., what 'pops up' to you when sorting those nodes?) were used, students appeared to struggle with making broader inferences. However, they were able to draw simple *data implications* based on prior knowledge about familiar nodes. For example, upon being asked what books they might recommend to someone who liked *The Girl Who Drank the Moon,* students generated several books related to specific nodes that they were familiar with, e.g., Kelly Barnhill's other two books and *Frankenstein* whose author, Mary Shelly, was a node in the network (Figure 2). We interpreted students' data practice here as *information extraction* on the specific node content, *categorization,* and simple *data implications* (e.g., book recommendations) using prior knowledge/experiences pertaining to specific node content.

Students' diverse prior experiences with node content generated various ways of categorizing data. For example, the Minneapolis node was grouped into different categories, and students' justification revealed their varied understanding of node inclusion in the network: "She was born here," "She has family there," and "She wrote the book there." Facilitators validated all *categorizations* when solely considering the node alone and subsequently directed students' attention to the Minneapolis node and its connections in Net.Create, in which it was connected to both Kelly Barnhill and Politics in the network (Figure 2). The facilitator elaborated on the connections between the three nodes, as many political issues that Barnhill discussed were at the national level, but some of them pertained to Minneapolis exclusively. Guiding students to read connections of nodes in the digital network and comparing it to the categorization experience without the network reference potentially mediated their gradual recognition that a single focus on individual nodes' content could limit their inference-making of the overall story, which provided a foundation for the next activity.

Activity 4: Author network in Net.Create

Students' reflection on categorization in Net.Create suggested network visualization and its functions mediated *data categorization.* Students demonstrated an improved understanding of using the network to draw inferences. Compared to the single response about Barnhill's book theme during activity 3, more students answered and identified the possible themes as [human] rights, fantasy, fiction, and politics. Students' comparison of two categorization methods (paper activity versus Net.Create) revealed that their better understanding of nodes because of Net.Create mediated their inference-making about the book themes. One student named two categories that were used during the paper activity: categorized Kelly Barnhill and Mary Shelly together as authors and categorized rights as important things. The students stated that their categorization in the paper activity was primarily based on their understanding of specific nodes' content, and the lack of context about nodes made the categorization harder than the categorization in Net.Create, where they could access node descriptions and connections. The same challenges were explained by another student who used Minneapolis as an example. She mentioned that their team categorized the Minneapolis node as a place in activity 3 because they did not know its connections to politics and Kelly Barnhill. However, categorization using Net.Create was easier because, "with notes, you can, it's kind of easier because you can kind of see how it's like more grouped together." The student also confirmed with the facilitator that they *used the relationships* in the network during their *categorization* process. Here, we interpret students' data practice as *finding relationships* in the data (e.g., how nodes were grouped together), *data categorization*, and *using the data* to make inferences about the book theme. Those data practices were mediated by the node descriptions that researchers provided and the visual representation of the network in Net.Create.

Activity 5: Chicken industry network

This activity continued to limit students to viewing the network rather than editing to support their focus on *using the data* and *finding relationships in the data* via network visualization. These skills require inferences, making them more cognitively complex than creating, modifying, or extracting, and the discussion showed that students displayed these more complex skills. For the most part, students focused on what they already knew over what was degree-central in the network visualization (i.e., a larger node means more connections). While the network was built to show corporate integration, most students focused on what was needed to support chickens on their hypothetical farms; as one student explained, "To have a chicken farm, you actually need chickens." Indeed, the most voted-for nodes were "chicks" and "chicken feed," both things that directly 'lead to' sustaining a population of birds in a common-sense fashion. Figure 3 shows how most of the nodes with higher votes were directly connected to either adult chickens or chicks. Importantly, though, at least some students were using the data contained within the network and moving beyond reasoning based on their prior assumptions. As one student explained, "All of the things connected to [a node] probably make it happen."

Beyond *using the data*, there was also evidence of students *finding relationships in the data*. When asked if having the network helped them understand chicken farming, one student replied, "I feel like if the circle node has more connections, then you can tell it is very important." This student, at least, understood the relational notion of degree and used this to inform their decision to vote for the node "poultry corporation." Making this reasoning visible to the whole class could potentially mediate others to engage with and eventually appropriate those ideas.

This chicken industry network mediated a particular kind of sensemaking. The graph of nodes and connections became a testing ground, a tool for students to explore and understand their preconceived notions. For instance, one group of students voted for "adult chicken" three times, explaining that "you need adult chicken to get chicks." While this went against the rules of the assignment (voting for one node three times instead of three different nodes), it allowed students to express their understanding of chicken farming. This notion about chickens likely came from a community source, either from their everyday lives or previous classwork on farming, because in commercial farming, it is not your adult birds that produce the next generation.

Activity 6: Social media marketing game

This activity supported students' use of multiple network features collectively to understand how big companies use social media users' data to advertise products. We present Student O's reflections and game-playing to demonstrate her evolving understanding across two rounds of discussions and how that progressivity was mediated. O's initial round of discussion on the marketing strategy demonstrated that her reasoning might be mediated by an emergent understanding of how networks work from the previous activities. Her strategy was described as passing a content card to those with the most connections. Her initial marketing pitch was to distribute the content card "Best Sports Bloopers of the Week" to 'Gabe Green' (Figure 8), who had the most connections (degree=11). However, Green's player ignored the card because his role card indicated a dislike of sports. Still, O's marketing strategy sharing and her performance in the game showcased her ability to *extract information* and *find relationships* in the data, such as

identifying nodes with a higher number of connections. Furthermore, she was able to *use the data to draw inferences*, deciding the most influential person in the network. O's approach of evaluating a node's significance based on its number of connections might have been mediated by previous activity where students had discussed the importance of nodes' connections. However, this strategy failed to disseminate the content card to a broader audience because it only attended to connection numbers and ignored how nodes were connected (e.g., edge types and connected nodes). We interpret that O had not yet seen the specific kinds of connection as relevant for this task, but later, that was made salient by watching other students' game-playing and discussion.

In the final discussion, O reflected on Student H's marketing strategy: choosing the first person "based on what was being liked, instead of giving [the content card] to somebody randomly." H's strategy proved successful as the content card she chose was distributed to many users. During play, the teacher guided H to zoom in on the social network, and H murmured, "Who likes, ok Robert really likes sports." She continued *data extraction* by silently zooming in/out of the social network before handing Robert Red the content card 'Highlights of local sports games." H's marketing strategy was successful because Robert Red is the third most connected node (degree = 8) and enjoys sports, as indicated by his sharing of sports content (Figure 8). It was unclear what other information H was extracting from the network and considering before choosing Robert Red. However, her final pick and the accompanying murmuring evidenced her attention to Red's connections.

The teacher's guidance in finding whose interests match the content card within the network visualization mediated the effective marketing approach of identifying an appropriate person to promote a content card that matches the person's interests. How H played the game functioned as a mediational means for O to notice the importance of how and what nodes were connected in addition to the number of connections. H's performance during the game demonstrated her ability to *extract different types of information* from the network and *find relationships* (e.g., the connection between Red and sports). H was also able to *categorize* Red's connections as games and *use that information* to show Red a content card related to that interest.

Discussion

This study reported elementary students' engagement with network visualization and data literacy practices in six network-related activities. Pre/post-test analyses demonstrated students' improved understanding of network visualization. The Interaction Analysis provided nuances into how the activity mediated students' data literacy development.

This present analysis has substantiated that network visualizations are a compelling tool for supporting data literacy. We documented how network visualization features and designed network activities successfully engaged students in an iteration of developing early data literacy skills, including creating/modifying data, extracting information, finding relationships, categorizing data, using data, and data implications. Network representations enhanced students' ability to observe and understand relationships between entities. Net.Create's functionalities of automated live data representation (Ahnert et al., 2020) and multi-user data entry allowed students to collaboratively create and interpret data simultaneously (Drucker, 2015). Along with other features (edge type prompt, nodes' descriptions, etc.), the network was effective in helping students move beyond attention from the individual data point of their interests/familiarity to identify patterns of the entire dataset and interpret the patterns. Facilitators demonstrated other representations in Net.Create, including filtering out a subset of the network and table, that are potentially powerful mediators of data literacy skills and are applicable to other tools. We acknowledge that not every student shows a complicated articulation of network visualization and data literacy concepts, though they all appeared to have understood the concepts well enough to apply them in-situ and create, interpret, and negotiate their ideas in the context of the network visualization. Therefore, future implementations will aim to tease out ways to enhance network affordances to make those concepts more salient to students and support their data reasoning skills.

Although network visualization tools are powerful in supporting young students' data sense-making and interpretation, students need an entry point to engage productively with those visualizations (Roberts & Lyons, 2020). This study has demonstrated that students' locally and personally relevant resources are valuable to mediating students' data literacy practices. The identity network supported students' different ways of representing themselves, potentially mediating students' shifting between different perspectives of data viewing. The author network and the chicken industry network invited students' out-of-classroom knowledge/experiences to play a major mediational means in their data literacy development. The social media marketing game helped students to make connections to social media to practice *data implication* practices. Pre/post-tests reflected students' growing awareness of how big companies collect/archive data and data security.

Our analysis attended to students' participation in six network visualization activities and presented mediations of network visualization tools and classroom activities that supported students in recognizing

relationships of the dataset. We concluded that network visualization is an effective means for elementary students to develop data literacy skills, and students' locally/personally relevant data is an asset to students' data engagement.

References

Authors (2014) blinded.

Authors (2018) blinded.

Authors (2021a) blinded.

Authors (2021b) blinded.

Authors (2022) blinded.

Authors (2023) blinded.

Acker, A. & Bowler, L. (2018). Youth data literacy: Teen perspectives on data created with social media and mobile device ownership. *Hawaii International Conference on System Sciences 2018 (HICCS).* January 3-6, 2018.

Ahnert, R., Ahnert, S. E., Coleman, C. Nicole, & Weingart, S. (2020). *The network turn: changing perspectives in the humanities.* Cambridge, United Kingdom: Cambridge University Press.

Alper, B., Riche, N. H., Chevalier, F., Boy, J., & Sezgin, M. (2017, May). Visualization literacy at elementary school. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 5485-5497).

Bae, S. S., Vanukuru, R., Yang, R., Gyory, P., Zhou, R., Do, E. Y. L., & Szafir, D. A. (2022). Cultivating visualization literacy for children through curiosity and play. *IEEE Transactions on Visualization and Computer Graphics*, *29*(1), 257-267.

Bergner, Y., Mund, S., Chen, O., & Payne, W. (2021). Leveraging interest-driven embodied practices to build quantitative literacies: A case study using motion and audio capture from dance. *Educational Technology Research and Development*, *69*, 2013-2036.

Bertin, J. (1983). *Semiology of graphics: diagrams, networks, maps.* Madison, Wis.: University of Wisconsin Press.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational studies in mathematics*, *45*, 35-65.

Bishop, F., Zagermann, J., Pfeil, U., Sanderson, G., Reiterer, H., & Hinrichs, U. (2019). Construct-a-vis: Exploring the free-form visualization processes of children. *IEEE transactions on visualization and computer graphics*, *26*(1), 451-460.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, *37*(1), 179-255.

Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world.* Cambridge, Massachusetts: The MIT Press.

Bowler, L., Acker, A., Jeng, W. and Chi, Y. (2017). "It lives all around us": Aspects of data literacy in teen's lives. In S. Erdelez & N.K. Agarwal (Eds.), *Proceedings of the Association for Information Science and Technology* (pp. 27– 35.) Hoboken, NJ: Wiley.

Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors,34*, 535-554.

Calabrese Barton, A., Greenberg, D., Turner, C., Riter, D., Perez, M., Tasker, T., Jones, D., Herrenkohl, L. R., & Davis, E. A. (2021). Youth Critical Data Practices in the COVID-19 Multipandemic. *AERA Open*, *7*.

Cañas, A. J., Coffey, J. W., Carnot, M. J., Feltovich, P., Hoffman, R. R., Feltovich, J., & Novak, J. D. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. *Report to the Chief of Naval Education and Training*, 1-108.

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, *63*(2), 123-134.

Cramer, C., Sheetz, L., Sayama, H., Trunfio, P., Stanley, H.E., Uzzo, S. (2015). NetSci High: Bringing Network Science Research to High Schools. In: Mangioni, G., Simini, F., Uzzo, S., Wang, D. (eds) Complex Networks VI. Studies in Computational Intelligence, vol 597. Springer, Cham.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382-393.

Deahl, E. (2014). Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments. Social Science Research Network. https://doi.org/10.2139/ssrn.2445621

Drucker, J. (2015). Graphical Approaches to the Digital Humanities. *A new companion to digital humanities*, 238-250.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*(1), 268-291.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R.Miettinen, & R.-L. Punamäki (Eds.), *Perspectives on activity theory*. Cambridge University Press.

Engeström, Y., Brown, K., Engeström, R., & Koistinen, K. (1990). Organizational forgetting: An activity-theoretical perspective. In D. Middleton & D. Edwards (Eds.), *Collective remembering* (pp. 139–168). Sage Publications, Inc.

Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. In J. Green Camilli, G., &. Elmore, P. (Ed.), *Handbook of complementary methods in educational research. (3rd ed.)* (pp. 177–192). American Educational Research Association.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124-158.

Gerofsky, S. (2011). Seeing the graph vs. being the graph. In G. Stam, & M. Ishino (Eds.), Integrating gestures: The interdisciplinary nature of gesture (pp. 245–256). Amsterdam: John Benjamins Publishing Company.

Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The Mathematical Imagery Trainer: from Embodied Interaction to Conceptual Learning. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1989-1998).

Humburg, M., Craig, K., Szostalo, M., Danish, J., Hmelo-Silver, C., & McCranie, A. (2021, January). Noticing, Understanding, and Encouraging Positive Engagement with Collaborative History Learning. In *Reflecting the Past and Embracing the Future: ISLS Annual Meeting 2021 (Online Event)*.

Jiang, S., Lee, V. R., & Rosenberg, J. M. (2022). Data science education across the disciplines: Underexamined opportunities for K-12 innovation. *British Journal of Educational Technology*, *53*(5), 1073-1079.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39–103.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305-325.

Kahn, J. (2020). Learning at the intersection of self and society: The family geobiography as a context for data science education. *Journal of the learning sciences*, *29*(1), 57-80.

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664-672.

Lee, V. R., & Drake, J. (2013). Quantified recess: Design of an activity for elementary students involving analyses of their own movement data. In *Proceedings of the 12th international conference on interaction design and children* (pp. 273-276).

Lopez, M. L., Roberto, C., Rivero, E., Wilkerson, M. H., Bakal, M., & Gutiérrez, K. (2021). Curricular Reorganization in the Third Space: A Case of Consequential Reasoning around Data. In de Vries, E., Hod, Y., & Ahn, J. (Eds.), Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021. (pp. 466-473). Bochum, Germany: International Society of the Learning Sciences.

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, *86*(14), 3200.

Roberts, J., Lyons, L., Cafaro, F., & Eydt, R. (2014). Interpreting data from within: Supporting human-data interaction in museum exhibits through perspective-taking. Proceedings of the 2014 Conference on Interaction Design and Children, Aarhus, Denmark.

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know?. *Journal of the Learning Sciences, 29*(1), 154-164.

Rubin, A., & Mokros, J. (2018). Data Clubs for Middle School Youth: Engaging Young People in Data Science. In *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10), Kyoto, Japan.*

Roberts, J., & Lyons, L. (2020). Examining spontaneous perspective taking and fluid self-to-data relationships in informal open-ended data exploration. *Journal of the learning sciences*, *29*(1), 32-56.

Schwendimann, B. A. (2015). Concept maps as versatile tools to integrate complex ideas: From kindergarten to higher and professional education. *Knowledge Management & E-Learning: An International Journal, 7*(1), 73-99.

Shreiner, T. L., & Guzdial, M. (2022). The information won't just sink in: Helping teachers provide technology-assisted data literacy instruction in social studies. *British Journal of Educational Technology*, *53*(5), 1134-1158.

Stornaiuolo, A. (2020). Authoring Data Stories in a Media Makerspace: Adolescents Developing Critical Data Literacies. *Journal of the Learning Sciences*, *29*(1), 81–103.

Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space. *Journal of the Learning Sciences*, *29*(1), 127–153.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, *21*(1), 14-23.

Wertsch, J. V. (1981). The concept of activity in Soviet psychology: An introduction. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 3-36). Sharpe.

Wilkerson, M., Lanouette, K. A., Shareff, R. L., Erickson, T., Bulalacao, N., Heller, J., ... Reichsman, F. (2018). Data transformations: Restructuring data for inquiry in a simulation and data analysis environment. In J. Kay & R. Luckin (Eds.), *Rethinking learning in the digital age: Making the learning sciences count.* Proceedings of the 13th International Conference of the Learning Sciences (ICLS 2018) (pp. 1383–1384). London, England: ISLS.
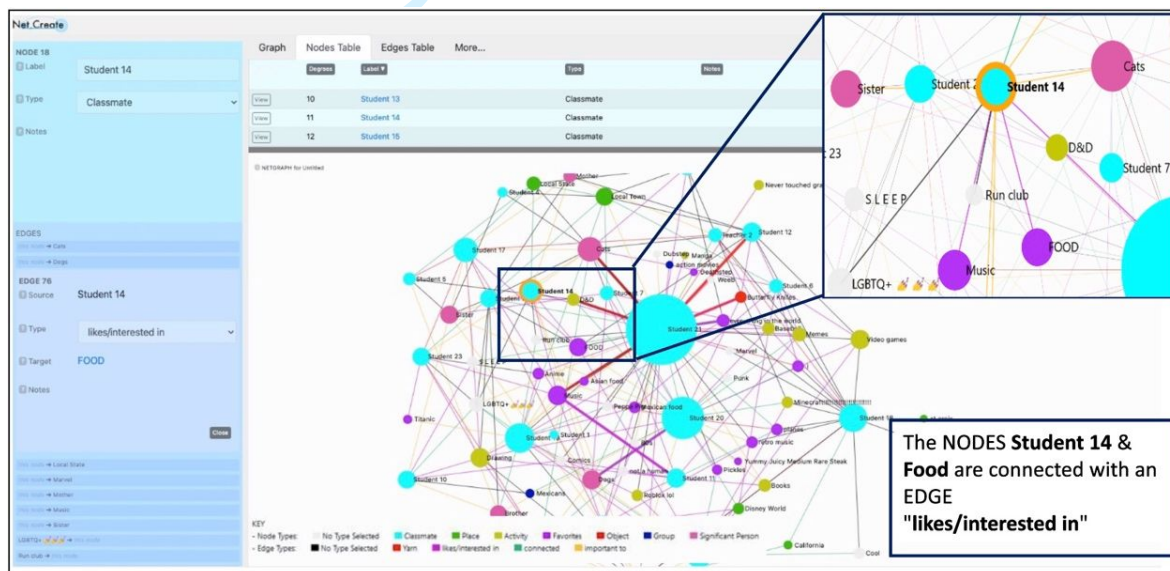

Figure 1. Students' identity network in Net.Create with an expanded view of sample nodes and edges.
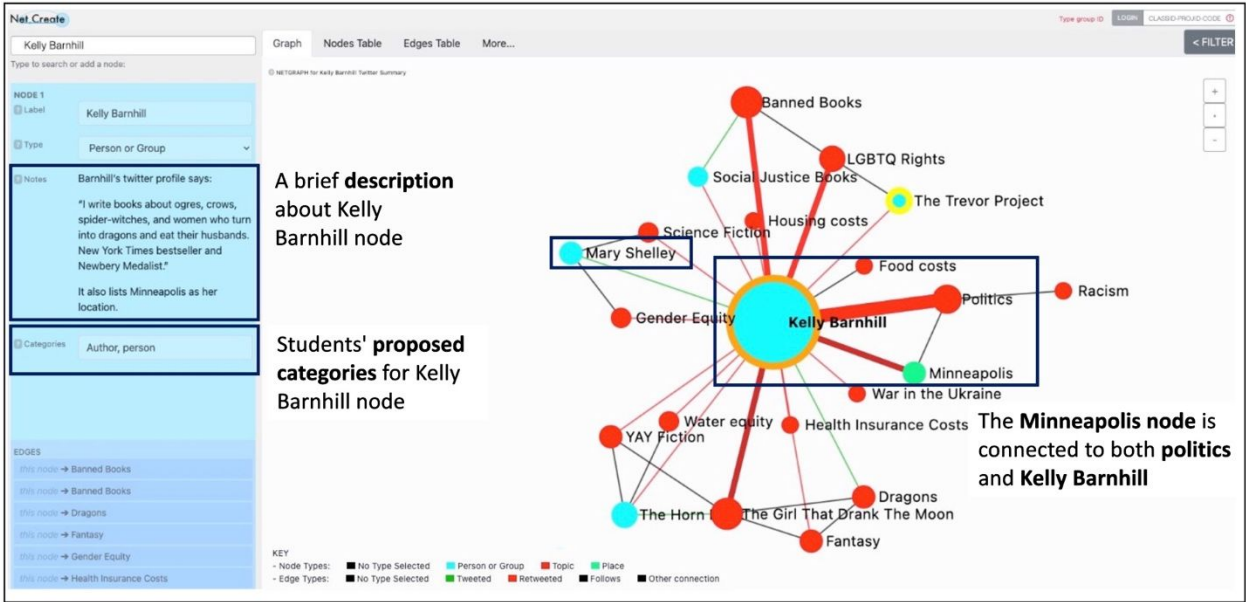
Figure 2. Net.Create's author network highlighting Barnhill, Shelley, and Minneapolis nodes in the network and Barnhill's node information
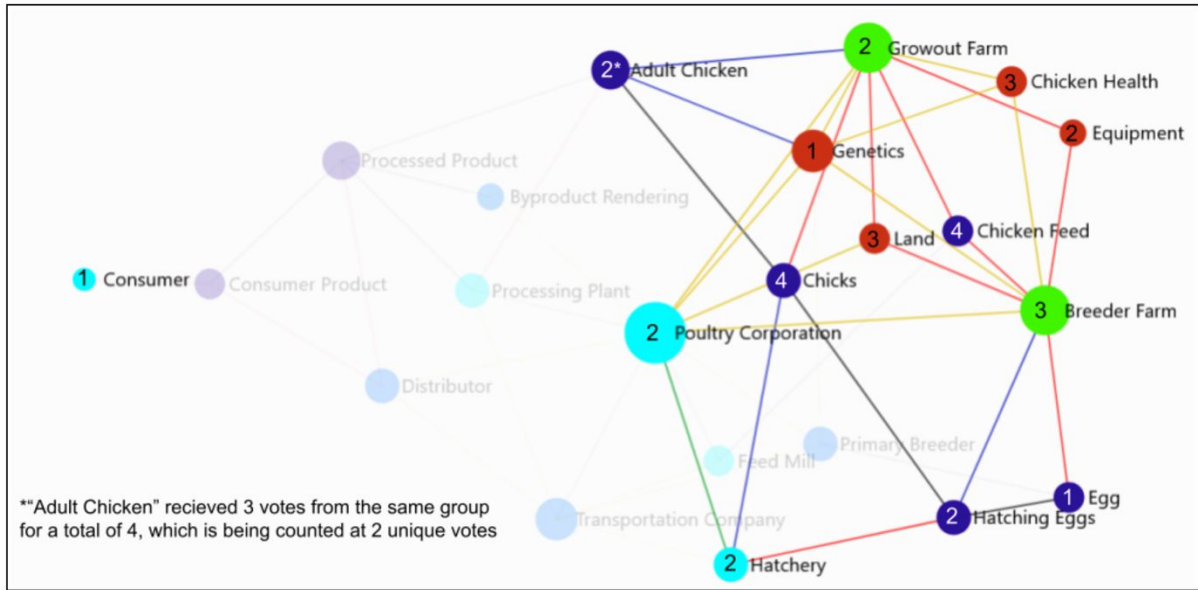
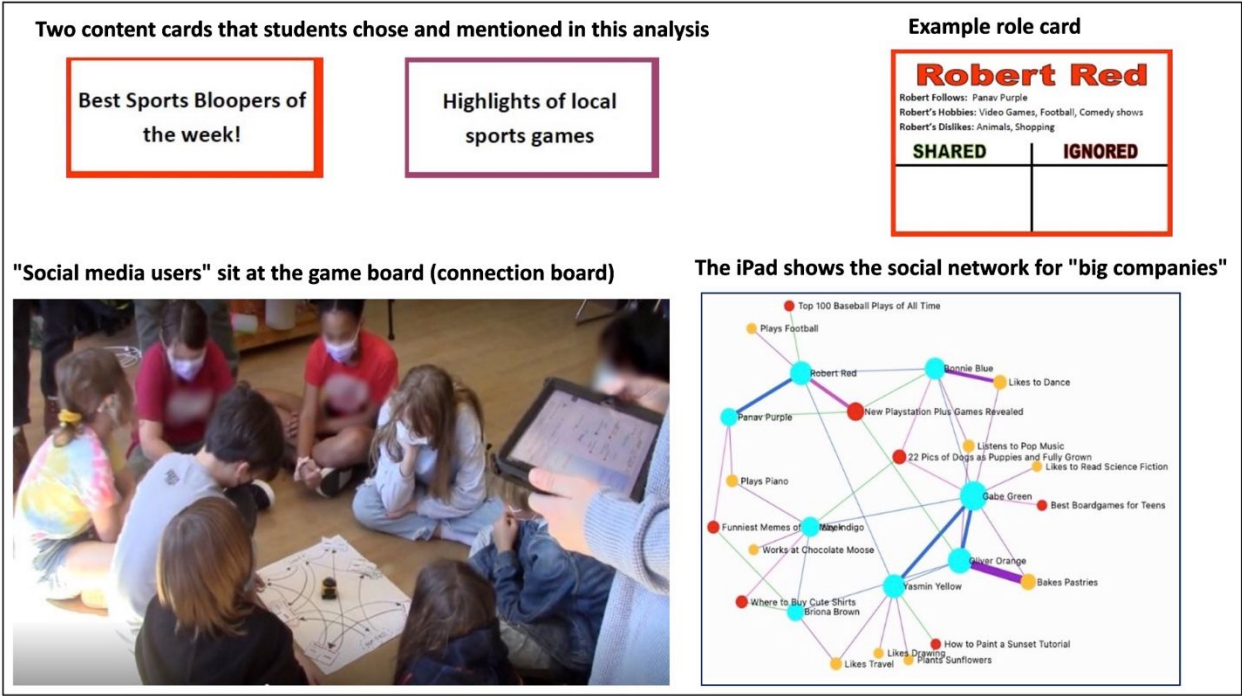Figure 3. The chicken industry network shows how many votes each node received.

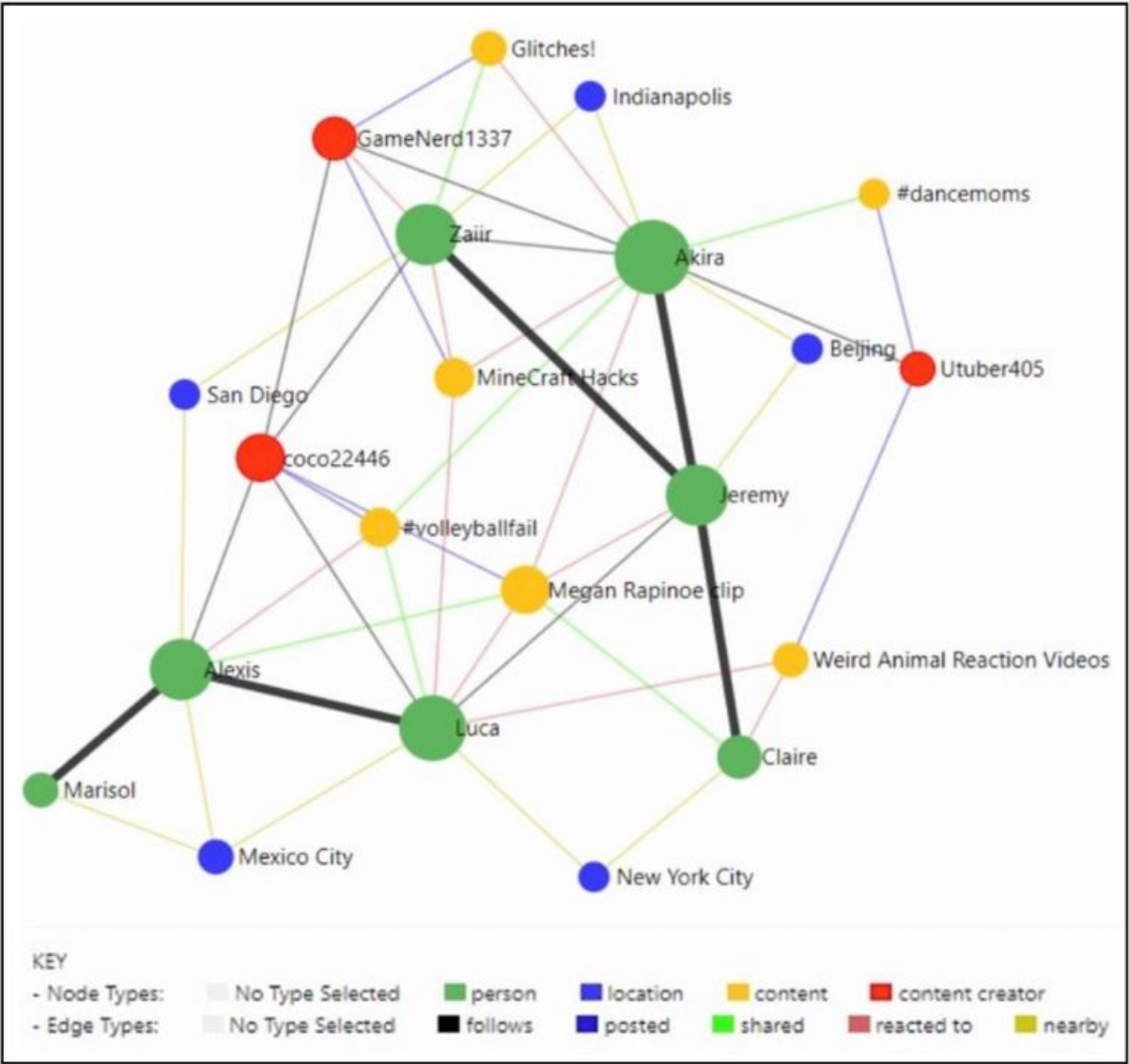Figure 4. The social media marketing game context.

Figure 5. The sample network in pre/post-tests.

1. In a few sentences, write what you think a network is.
2. Based on the network above, who would you guess is friends with Luca in real life? Why them?
3. Suppose a celebrity posted a sports video on social media. Based on the network above, who do you think would find out about this? Why?
4. Suppose Zaiir asked for recommendations on new content to follow; what would you recommend based on the network above? Why?
5. What does it mean that Luca and Claire are NOT connected to each other in the network, but are both connected to New York City?
6. Based on the network above, who would be able to see Alexis' post if their profile is viewable only by friends? How about friends of friends?
7. Who is probably the most popular person in the network above? Why?
8. What is the most popular topic in the network above? Why?
   a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~
9. Who has the most influence? Why?
10. Suppose GameNerd1337 posted a new video on social media. Who would likely find out about it first based on the network above, Claire or Luca?
11. ~~What kind of data would you need to build this kind of network?~~
12. What could a corporation like Meta or TikTok do, having the information in this network?
13. ~~How could Utuber405 extend their influence in the network above?~~
   a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~
14. ~~Suppose a new creator, AntiVaxx12, starts posting videos claiming that vaccines are ineffective. Who in the network is likely to see these videos if GameNerd1337 follows AntiVaxx12?~~
   a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~

Questions with strikethroughs were excluded from analysis because two or fewer students answered meaningfully (e.g., a response other than "idk") across the pre/post-tests

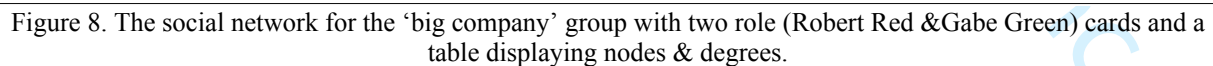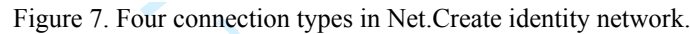Figure 6. Open-ended questions in pre/post-tests

Figure 7. Four connection types in Net.Create identity network.

Figure 8. The social network for the 'big company' group with two role (Robert Red & Gabe Green) cards and a table displaying nodes & degrees.

Table I. Data literacy codes used in the video analysis.

Codes	Illustrative examples in Net.Create	Literature source
Create data	when students create nodes, create edges, and the discussion accompanying the creation.	This is inspired by *Capta* (Drucker, 2015), which described the process of students collaboratively creating and interpreting data.
Modify data	when students suggest changes about edges and nodes to the person who edits the network, including instances when the suggestion is declined/accepted and the actual editing changes.	This is from *Capta* (Drucker, 2015).
Extract information	when students extract information about a single data point, which includes: 1. describe pre-existing nodes (e.g., specific content-wise) 2. discuss nodes/edges details 3. scroll up/down to find what information there is in the network	This is an elementary skill in the Friel et al. (2001) framework. It refers to the extraction of elementary information about a single data point/value to an explicit question.
Use the data	when students recognize existing information in the network and then using that information to continue to make connections to the pre-existing nodes or add non-existing nodes. Using existing information to understand the topic and make inferences: • guessing book themes • inferring significant factors for the chicken industry • choosing a marketing influencer	We simplified and adapted Calzada Prado and Marzal's (2013) definition of using data to our context as students use existing information to make the decision on expanding the network or making inferences about the network topic.
Data implication	when students talk about how to use the data in the network for real-world problems.	We adapted the "move beyond data" in Curcio (1987) to our context as the skills of extending and inferring from representations to answer questions. To distinguish it from the code "use the data," answering questions beyond the present network is key.
Categorize the data	when students talk about nodes' categorization, grouping nodes, and naming categories. ● using their understanding of specific node content to categorize data. ● categorizing data based on the node connections in the network. ● noticing/showing nodes that are in the same category	This is from the classifier perspectives to data in Konold et al. (2015), in which students group cases with similar values as a unit.

| Find relationships in data | when students focus on nodes connections/edges types, nodes/edges sizes (e.g., recognizing a node is bigger than another, an edge is thicker than another), and nodes position (e.g., some are in the center). | This is an intermediate skill in the Friel et al. (2001) framework. It refers to interpolating and finding relationships in data on a graph. For example, identification of trends (Wainer, 1992), local/global comparison of graph features and focus on more than a single specifier (Carswell, 1992), combining and compiling data to discover relationships (Bertin, 1983). |

Table II. Network visualization concepts and how students talked about them on pre/post-tests.

	Number (Percent) of Students on Pre-test	Number (Percent) of Students on Post-test
Network Components	**4 (23.5)**	**13 (76.5)**
• People	1 (5.9)	9 (52.9)
• Things	2 (11.8)	9 (52.9)
• Data	1 (5.9)	2 (11.8)
Connections	**10 (58.8)**	**17 (100)**
• Connections broadly	1 (5.9)	3 (17.7)
• Connections between people and things	0 (0)	3 (17.7)
• Connections between things	1 (5.9)	2 (11.8)
• Connections between people	0 (0)	4 (23.5)
• Direct connections	6 (35.3)	13 (76.5)
• Overlapping connections	1 (5.9)	2 (11.6)
• Indirect connections	0 (0)	4 (23.5)
• Number of connections	0 (0)	2 (11.8)
Edges	**4 (23.5)**	**8 (47.1)**
• Edge type	4 (23.5)	5 (29.4)
• Edge thickness	0 (0)	5 (29.4)
Node Content	**3 (17.7)**	**8 (47.1)**
Overall Shape of the Network	**2 (11.8)**	**4 (23.5)**
• Graph	0 (0)	1 (5.9)

• Chart	0 (0)	2 (11.8)
• Web	2 (11.8)	2 (11.8)
Network Implications	**9 (52.9)**	**11 (64.7)**
• Social Media	1 (5.9)	0 (0)
• Internet/Technology	8 (47.1)	6 (35.3)
• Used Network to reason about things not in the network	2 (11.8)	7 (41.2)
Centrality	**1 (5.9)**	**6 (35.3)**
• Based on node size	0 (0)	1 (5.9)
• Based on number of connections	0 (0)	3 (17.7)
• Based on number of connections to people only	1 (5.9)	2 (11.8)

Figure 1. Students' identity network in Net.Create with an expanded view of sample nodes and edges
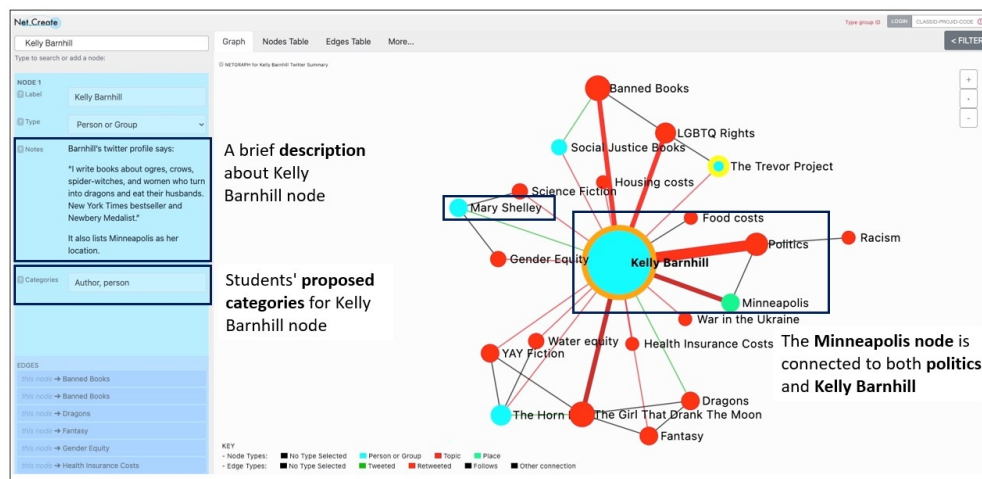
338x190mm (96 x 96 DPI)

Figure 2. Net.Create's author network highlighting Barnhill, Shelley, and Minneapolis nodes in the network and Barnhill's node information
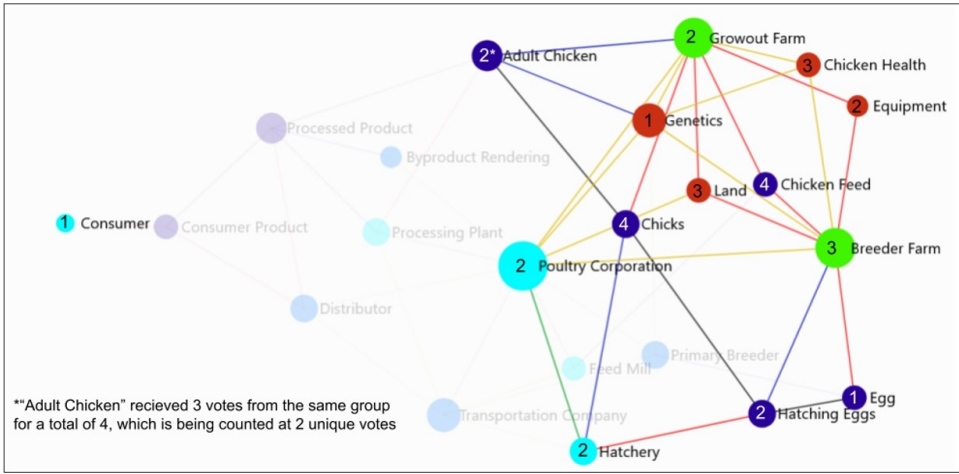
338x190mm (96 x 96 DPI)

Figure 3. The chicken industry network shows how many votes each node received
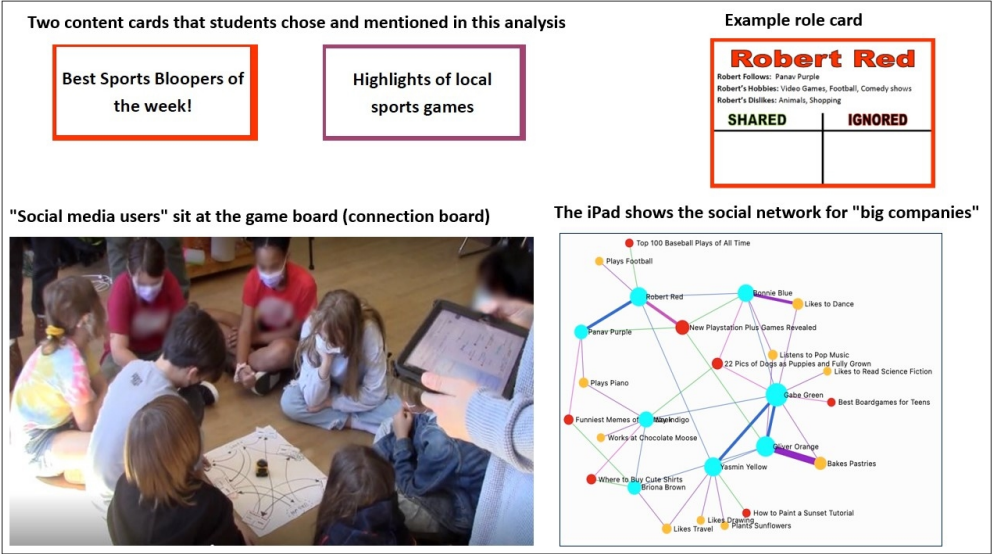
338x190mm (96 x 96 DPI)

Figure 4. The social media marketing game context

338x190mm (96 x 96 DPI)
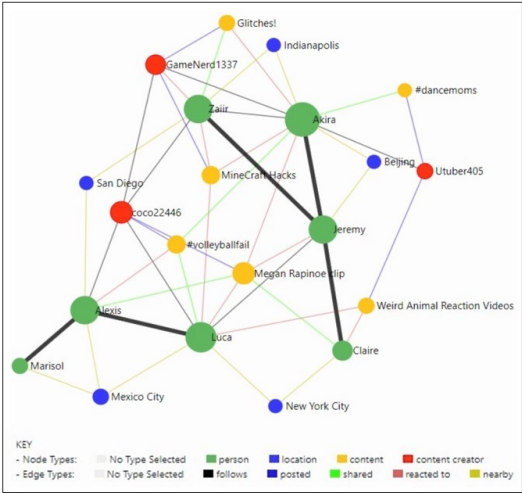
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 5. The sample network in pre/post-tests

338x190mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

1. In a few sentences, write what you think a network is.
2. Based on the network above, who would you guess is friends with Luca in real life? Why them?
3. Suppose a celebrity posted a sports video on social media. Based on the network above, who do you think would find out about this? Why?
4. Suppose Zaiir asked for recommendations on new content to follow; what would you recommend based on the network above? Why?
5. What does it mean that Luca and Claire are NOT connected to each other in the network, but are both connected to New York City?
6. Based on the network above, who would be able to see Alexis' post if their profile is viewable only by friends? How about friends of friends?
7. Who is probably the most popular person in the network above? Why?
8. What is the most popular topic in the network above? Why?
   a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~
9. Who has the most influence? Why?
10. Suppose GameNerd1337 posted a new video on social media. Who would likely find out about it first based on the network above, Claire or Luca?
11. ~~What kind of data would you need to build this kind of network?~~
12. What could a corporation like Meta or TikTok do, having the information in this network?
13. ~~How could Utuber405 extend their influence in the network above?~~
    a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~
14. ~~Suppose a new creator, AntiVaxx12, starts posting videos claiming that vaccines are ineffective. Who in the network is likely to see these videos if GameNerd1337 follows AntiVaxx12?~~
    a. ~~Is there any way you could reorganize how the information is presented to help you answer this kind of question? How?~~

Questions with strikethroughs were excluded from analysis because two or fewer students answered meaningfully (e.g., a response other than "idk") across the pre/post-tests

Figure 6. Open-ended questions in pre/post-tests

350x376mm (72 x 72 DPI)

41
42
43
44
45
46
47
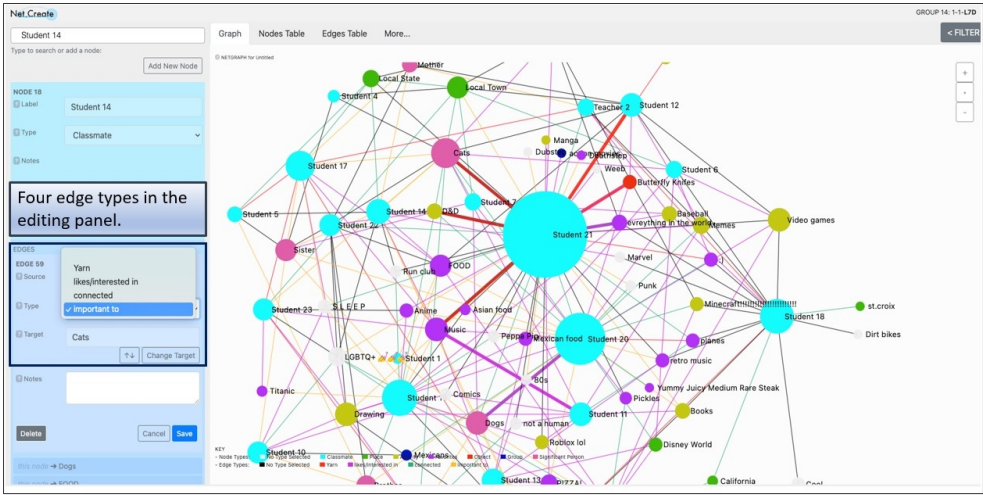48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 7. Four connection types in Net.Create identity network
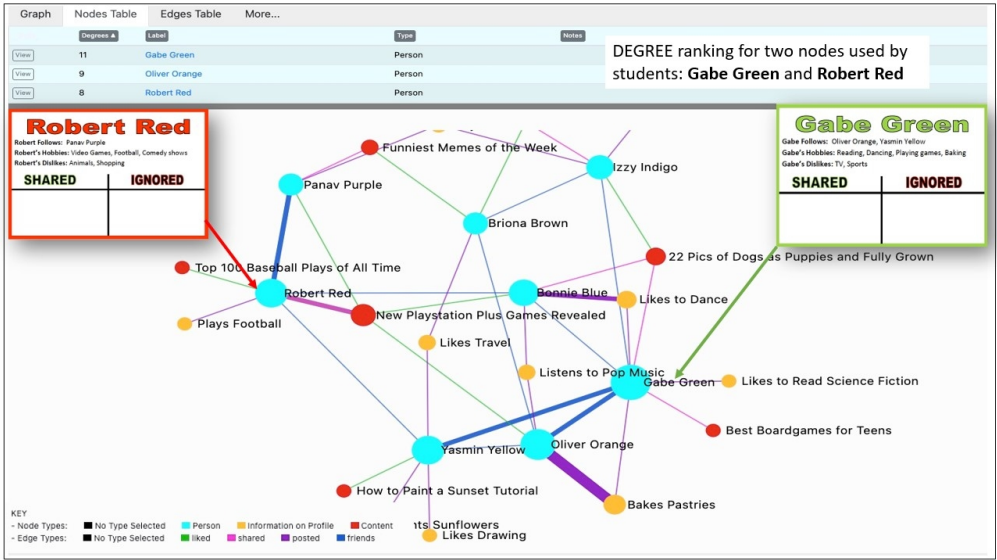
338x190mm (96 x 96 DPI)

Figure 8. The social network for the 'big company' group with two role (Robert Red &Gabe Green) cards and a table displaying nodes & degrees

338x190mm (96 x 96 DPI)

Table I. Data literacy codes used in the video analysis.

Codes	Illustrative examples in Net.Create	Literature source
Create data	when students create nodes, create edges, and the discussion accompanying the creation.	This is inspired by *Capta* (Drucker, 2015), which described the process of students collaboratively creating and interpreting data.
Modify data	when students suggest changes about edges and nodes to the person who edits the network, including instances when the suggestion is declined/accepted and the actual editing changes.	This is from *Capta* (Drucker, 2015).
Extract information	when students extract information about a single data point, which includes: 1. describe pre-existing nodes (e.g., specific content-wise) 2. discuss nodes/edges details 3. scroll up/down to find what information there is in the network	This is an elementary skill in the Friel et al. (2001) framework. It refers to the extraction of elementary information about a single data point/value to an explicit question.
Use the data	when students recognize existing information in the network and then using that information to continue to make connections to the pre-existing nodes or add non-existing nodes. Using existing information to understand the topic and make inferences: • guessing book themes • inferring significant factors for the chicken industry • choosing a marketing influencer	We simplified and adapted Calzada Prado and Marzal's (2013) definition of using data to our context as students use existing information to make the decision on expanding the network or making inferences about the network topic.
Data implication	when students talk about how to use the data in the network for real-world problems.	We adapted the "move beyond data" in Curcio (1987) to our context as the skills of extending and inferring from representations to answer questions. To distinguish it from the code "use the data," answering questions beyond the present network is key.
Categorize the data	when students talk about nodes' categorization, grouping nodes, and naming categories. • using their understanding of specific node content to categorize data. • categorizing data based on the node connections in the network. • noticing/showing nodes that are in the same category	This is from the classifier perspectives to data in Konold et al. (2015), in which students group cases with similar values as a unit.

| **Find relationships in data** | when students focus on nodes connections/edges types, nodes/edges sizes (e.g., recognizing a node is bigger than another, an edge is thicker than another), and nodes position (e.g., some are in the center). | This is an intermediate skill in the Friel et al. (2001) framework. It refers to interpolating and finding relationships in data on a graph. For example, identification of trends (Wainer, 1992), local/global comparison of graph features and focus on more than a single specifier (Carswell, 1992), combining and compiling data to discover relationships (Bertin, 1983). |

Table II. Network visualization concepts and how students talked about them on pre/post-tests.

	Number (Percent) of Students on Pre-test	Number (Percent) of Students on Post-test
Network Components	**4 (23.5)**	**13 (76.5)**
• People	1 (5.9)	9 (52.9)
• Things	2 (11.8)	9 (52.9)
• Data	1 (5.9)	2 (11.8)
Connections	**10 (58.8)**	**17 (100)**
• Connections broadly	1 (5.9)	3 (17.7)
• Connections between people and things	0 (0)	3 (17.7)
• Connections between things	1 (5.9)	2 (11.8)
• Connections between people	0 (0)	4 (23.5)
• Direct connections	6 (35.3)	13 (76.5)
• Overlapping connections	1 (5.9)	2 (11.6)
• Indirect connections	0 (0)	4 (23.5)
• Number of connections	0 (0)	2 (11.8)
Edges	**4 (23.5)**	**8 (47.1)**
• Edge type	4 (23.5)	5 (29.4)
• Edge thickness	0 (0)	5 (29.4)
Node Content	**3 (17.7)**	**8 (47.1)**
Overall Shape of the Network	**2 (11.8)**	**4 (23.5)**
• Graph	0 (0)	1 (5.9)
• Chart	0 (0)	2 (11.8)
• Web	2 (11.8)	2 (11.8)
Network Implications	**9 (52.9)**	**11 (64.7)**
• Social Media	1 (5.9)	0 (0)
• Internet/Technology	8 (47.1)	6 (35.3)
• Used Network to reason about things not in the network	2 (11.8)	7 (41.2)
Centrality	**1 (5.9)**	**6 (35.3)**
• Based on node size	0 (0)	1 (5.9)

• Based on number of connections	0 (0)	3 (17.7)
• Based on number of connections to people only	1 (5.9)	2 (11.8)