Visual Analytics for Multimodal Social Network Analysis: A Design Study with Social Scientists

Sohaib Ghani, *Student member, IEEE*, Bum Chul Kwon, *Student member, IEEE*,
Seungyoon Lee, Ji Soo Yi, *Member, IEEE*, and Niklas Elmqvist, *Senior member, IEEE*

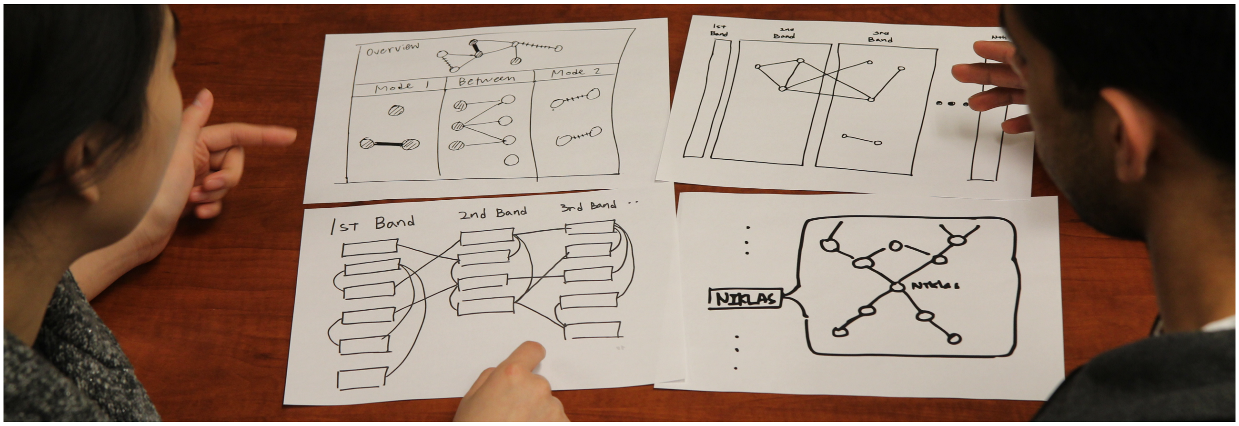


Fig. 1. Sketches for visualizing multimodal social networks developed collaboratively by social scientists and visual analytics experts. Early designs (top) partition graphs, whereas latter ones (bottom left) use vertical bands while maintaining compatibility with node-link diagrams (bottom right).

Abstract—Social network analysis (SNA) is becoming increasingly concerned not only with actors and their relations, but also with distinguishing between different types of such entities. For example, social scientists may want to investigate asymmetric relations in organizations with strict chains of command, or incorporate non-actors such as conferences and projects when analyzing co-authorship patterns. Multimodal social networks are those where actors and relations belong to different types, or *modes*, and multimodal social network analysis (mSNA) is accordingly SNA for such networks. In this paper, we present a design study that we conducted with several social scientist collaborators on how to support mSNA using visual analytics tools. Based on an open-ended, formative design process, we devised a visual representation called *parallel node-link bands* (PNLBs) that splits modes into separate bands and renders connections between adjacent ones, similar to the list view in Jigsaw. We then used the tool in a qualitative evaluation involving five social scientists whose feedback informed a second design phase that incorporated additional network metrics. Finally, we conducted a second qualitative evaluation with our social scientist collaborators that provided further insights on the utility of the PNLBs representation and the potential of visual analytics for mSNA.

Index Terms—Design study, user-centered design, node-link diagrams, multimodal graphs, interaction, qualitative evaluation.

1 Introduction

Social network analysis (SNA) [43] is the collective name for a family of methods used to analyze sets of social actors connected by relations. SNA has become increasingly important due to modern information technologies that allow humans to connect and relate in entirely new and easily observable ways. As a case in point, social media websites, such as Facebook, Twitter, and LinkedIn, include hundreds of millions of users with various types of relations between them. The scale and complexity of these massive networks put increasing demands on software support for computation, statistics, and decision making. Visual analytics is increasingly being used for this purpose, and several new systems have been proposed that merge visual representations and network statistics to aid social network analysis, including Gephi [3], NodeXL [27, 47], GraphDice [7], TimeMatrix [54], and GUESS [1].

However, in many real-world settings, the networks consist of not one but several different types, or *modes*, of nodes. Examples include co-authorship networks that contain not just authors, but also the venues they attend and the journals they publish in; organizational charts that contain employees as well as the departments they belong to; and information retrieval processes that involve both databases and the people who access them. Consequently, these *multimodal* (cf. unimodal) social networks also have multiple types of edges depending on whether the edge is connecting nodes of the same mode (within-mode, such as representing friendship among people) or different modes (between-mode, such as employee affiliations with departments, or employees accessing databases). However, while unimodal network visualization is prevalent, as evidenced by the examples above, few techniques exist for visualizing multimodal graphs. Further, in the field of social science, a standard theoretical framework for analyzing multimodal social networks, especially those that involve more than two modes, has not yet been established.

- *S. Ghani and N. Elmqvist are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. E-mail: {sghani, elm}@purdue.edu*
- *B. C. Kwon and J. S. Yi are with the School of Industrial Engineering, Purdue University, West Lafayette, IN, USA. E-mail: {kwonb, yij}@purdue.edu*
- *S. Lee is with the Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA. E-mail: seungyoon@purdue.edu*

Submitted to IEEE VAST 2013. Do not redistribute.

In this paper, we present a design study on the use of visual analytics to aid social scientists in conducting multimodal social network analysis (mSNA). Retrospectively,[1] we largely follow the methodological framework proposed by Sedlmair et al. [44]. As with any design study, a major obstacle is often to articulate user (social scientists in our case) needs and requirements (*low task clarity* [44]). To aid in this process, we included one of our social scientist collaborators as an active participant in the project and a co-author of this paper from its very beginning. Nevertheless, this low task clarity was exacerbated by the fact that multimodal social networks are not yet a well-established concept even in social science, so our collaborators had a difficult time defining requirements and features. We therefore decided to conduct an iterative design process that included (1) an initial open-ended formative phase, (2) a visualization design phase, (3) an interim qualitative evaluation with five social scientists, (4) another design iteration incorporating additional analytics support, and (5) a second qualitative evaluation with our domain experts validating our changes. Due to the low task clarity of mSNA, our visual analytics tool—MMGRAPH—played an interesting role throughout this process: it served not only as a tool for answering questions that our domain experts had about their data, but it also became a living prototype that helped them see the potential in visual analytics and aided them in coming up with new requests for future iterations of the tool.

Our contributions in this paper include the following: (i) characterizing the problems and abstracting some potential tasks for multimodal social network analysis (mSNA); (ii) proposing a visual analytics tool for mSNA (MMGRAPH) refined through multiple iterations of design and evaluation with social scientists; and (iii) sharing several lessons learned through this iterative design process that are specific to the domain of social science and social network analysis. We want to emphasize that this research project is a design study, and thus clearly *problem-driven* and not *technique-driven*. Thus, even though we ended up deriving some unique representation and interaction techniques (e.g., the open-sesame interaction in Section 5.2), demonstrating the novelty of these techniques is *not* the emphasis of this paper.

2 BACKGROUND

Social networks are universal to the human condition. They can be used to model our relations to friends and relatives; our connections to groups and organizations; and the social structure of our very own neighborhood, state, or nation. Researchers have long studied these types of social networks using network theory and techniques through what is collectively known as *social network analysis* (SNA) [34]. These techniques enable the examination of social phenomena such as whether friendship leads to the contagion of obesity (or vice versa), what causes radicalization and terrorist group formation in extremist communities, and the nature of user-to-user interactions on social media platforms. Below we review cross-disciplinary research within both social science and visualization that this work touches.

2.1 Definitions

We define a *multimodal graph* G as the traditional ordered pair $G = (V, E)$ comprised of a set of vertices V and edges E, but where vertices can be partitioned using a modality equivalence relation $\sim_{mod}$. This modality relation $\sim_{mod}$ is defined using the notion of *vertex type*, and the equivalence classes (partitions) defined by relation are called *modes*. We can further define a modality relation for edges based on a tuple of the modes of the two vertices an edge connects.

For example, consider a simple biological food chain network where V is the set of all entities in the network, and E are their linkages. We can define the equivalence classes (modes), such as "plants" (green), "prey animals" (red), "predators" (blue), and "habitats" (yellow), and define the modality relation $\sim_{mod}$ to organize vertices based on which category they belong.

While multimodal graphs are prevalent in many domains, this paper is primarily motivated by SNA [43]. We define *multimodal social network analysis* (mSNA) as SNA in the presence of *multimodal social networks*, i.e., where the social actors can be partitioned into modes (not all of them necessarily sentient or even living). Given this definition, the food chain above can be regarded as a multimodal social network (even if "biological network" is perhaps closer at hand).

This particular example represents one type of multimodal network where different modes represent the attributes of a single type of entity (e.g., species). The relationships represent species eating other species. Another major source of multimodal networks is multivariate tabular data, where there are multiple types of entities that have relationships with each other, as well as attributes of those entities [39]. As the modes represent different types of entities, the types of links also vary across different pairs of modes. For instance, in the dataset we work with in the current paper, PIs "conduct" projects, program managers "award" projects, and PIs "belong to" institutions.

2.2 Social Network Analysis

With the advent of communication technologies (e.g., social media websites and collaborative knowledge tools) and large computational platforms (e.g., sensor networks and data collection tools), a massive amount of social network data is being collected today. Much effort has been made to analyze such data, leading to several interesting applications of network theories to social phenomena (e.g., [25, 37]).

However, to the best of our knowledge, investigation of social networks has largely dealt with unimodal social networks (e.g., friendship relations among friends and co-authorship relations among authors) and there has been an increasing need to expand our formulation to address relations between multiple types of nodes [14]. Such expansion enables a wide variety of questions to be examined beyond those from the realm of unimodal networks. For example, dynamics involving friendship can be better explained by considering the way people participate in events or affiliate with social organizations. Co-authorship among authors can be better understood when looking at the projects they work on as well as the institutions they are employed by. These additional types of nodes may have substantial effects on the formation and dissolution of social ties, and therefore allow us to investigate more complex social interaction among multiple types of entities.

The simplest configuration of multimodal networks is *two-mode networks* that consist of two modes—also called affiliation networks in the context of groups and members [9, 51]. A large part of existing literature on two-mode networks has dealt with bipartite graphs, defined as a graph with nodes in two distinct sets, and links only between nodes of one set and nodes of the other [36]. These links can be considered as between-mode ties (as opposed to within-mode ties).

A review of the relevant literature (e.g., [4, 9, 18]) suggests several representative measures for multimodal networks:

- Q1 Measures at the individual node level:
  - Q1a Centrality: Which nodes are central in terms of between-mode ties? (e.g., degree centrality: which nodes in mode A are tied to the most nodes in mode B?; betweenness centrality: which nodes in mode A are most central in terms of bridging otherwise disconnected nodes in mode B?)
  - Q1b Positions and roles: Which nodes occupy similar positions and roles? (e.g., structural equivalence: which nodes in mode A are similar in terms of their ties to nodes in B?)
  - Q1c Attributes: Which nodal attributes impact between-mode ties? (e.g., how does the attribute of nodes in mode A impact their ties to nodes in mode B?)
- Q2 Measures at the global network level:
  - Q2a Density: How dense is the network, between and within modes? (What is the number of between-mode and within-mode links divided by the maximum number of possible links between-mode and within-mode, respectively?)
  - Q2b Centralization: How centralized is the network in between-mode ties? (Are nodes in one mode similar in terms of number of ties they have to another mode, or are ties unequally distributed with a few dominating nodes?)

[1] Sedlmair et al. [44] was published after this design study started.

Q2c Subgroups: Are there visible substructures in the network? (Which nodes in mode A can be clustered based on being connected to the same or similar nodes in mode B?)

Interpretation of these measures vary depending on the source of multimodal data. In particular, when multimodal networks are derived from tabular data, the nodes are of different levels or entities, and the types of links depend on which modes are being examined, therefore requiring a context-specific application of the measures.

So far, a standard approach to analyzing multimodal networks has been to transform them into unimodal social networks either through projection or through separation. For example, the ties between manufacturers and users of a product are transformed (or projected) to ties among manufacturers established in case of common users [10]. In other cases, a manufacturer's network and a user's network are divided and analyzed separately, or combined into a unimodal network with the modes treated as node attributes. While these conversions are done for simplicity as well as for utilizing existing unimodal social network analysis (uSNA) tools, a rich set of information is lost during this process [36]. First of all, networks produced through this transformation do not represent a direct relation between actors of different modes, but an indirect relation induced by their common affiliation to a set of events [4]. Second, by removing one set of nodes from the data or combining the nodes, attribute information associated with each of the modes cannot be simultaneously considered [46].

Given these limitations, we believe that a social network analysis tool that presents the subtleties and complexities of multimodal social networks can help social scientists gain useful new insights about what they would like to learn from such multimodal networks.

2.3 Multimodal Network Visualization

Our review of graph visualization [20, 21, 30] has identified only a few studies on multimodal networks. We summarize these below:

2.3.1 Compound Network Visualization

A common approach for visualizing a multimodal graph is to treat it as a unimodal graph, with different colors or shapes distinguishing between types of modes and links. Such *compound network visualizations* are found in several visualization systems (e.g., [3, 8, 46]). However, this approach confounds all modes within the same view, and the resulting visual complexity can be high. Many techniques have been studied to overcome such visual complexity from large graph [50]. Despite such techniques, only certain ties or nodes can be shown at any point in time, but this naturally results in data being omitted.

2.3.2 Eliminating Modes

Another approach eliminates modes by *projecting* [39] nodes based on connections to a particular mode. This retains the connectivity structure, yet reduces the number of nodes (by mode). For example, if authors A and B write paper X together, the ties (A-X) and (B-X) can be merged (*projected*) into (A-B). While this approach can lower the overall complexity, it comes at the cost of information loss. For example, (A-W), (B-W), and (C-W) will have the same merged network as (A-X), (B-X), (B-Y), (C-Y), (C-Z), and (A-Z) even though in the former case, three authors collectively wrote a single paper, while in the latter case, three pairs of authors wrote three different papers.

2.3.3 Linked Network Visualization

A third approach is to use multiple views, each of which renders a different mode of the graph separately (*linked network visualization*). Between-mode ties are visualized using visual links or brushing (when nodes are selected in one view, corresponding nodes in another view are highlighted). VisLink [13], semantic substrates [2, 45], the list view in Jigsaw [48], and SmallWorlds [24] are examples of this idea.

More specifically, all four of these examples provide distinct planes that can be used for different mode networks, and then show connections between the planes using graphical links. However, while VisLink views are often less cluttered than compound network visualizations, the view can still be visually complex due to overlapping between-mode and within-mode ties. When brushing is used, the clutter is reduced. However, only a partial set of between-mode ties can be shown in this case. Semantic substrates, the list view in Jigsaw, and SmallWorlds are less cluttered, but this is mainly because they do not show within-mode ties. When there are clear hierarchical structures between the nodes, TreeNetViz [23], which uses a radial space-filling visual design, is another effective method. However, not all multimodal networks have such an existing hierarchical structure. Finally, the recent GraphTrail [16] tool visualizes and aggregates attributes associated with nodes and edges. The multimodal networks are called *heterogeneous networks* in their work, but involve different types of nodes and edges. However, their focus is on building visual queries using these graphs, not on general visual analytics for mSNA.

2.4 Reducing Data and Visual Complexity

The problem of visualizing multimodal social network is similar to "the curse of dimensionality" [6], which is a problem occurring when multidimensional (not multimodal) data are projected onto a display. Such projection easily generates clutter, distortion, and ensuing confusion. To make matters worse, visualizing a unimodal social network often consumes two spatial dimensions while visualizing unidimensional data consumes only one spatial dimension. Thus, it is not possible to directly borrow ideas from multidimensional visualization to address the challenges of multimodal network visualization. However, we can draw upon lessons learned from previous work for overcoming the curse of dimensionality. A review reveals the following strategies:

Divide and Conquer. Based on subdividing a problem into smaller components until each component is small enough to easily solve, this is one of the core strategies in many sub-disciplines of computer science, and the same principle has been applied to visualize multidimensional data. However, divide and conquer only visualizes a subset of data at a time, making the global view difficult to understand. Examples include scatterplot matrices (SPLOMs) [26], graph exploration with degree-of-interest [49], and Worlds within Worlds [19].

Distortion. In order to show the overall picture more effectively, some techniques distort the orthogonal relationships between dimensions, thereby gaining compactness by sacrificing familiarity. Examples include parallel coordinates [31], star coordinates [33], and Flexible LINked Axes (FLINA) [12].

Compression. When the amount of data and the number of dimensions surpass a certain level, the information may be drastically compressed using meta-dimensional information to show the overview at the cost of information loss. Examples include Principal Component Analysis [26], multidimensional scaling [15], and Scagnostics [53].

Metaphor. When compression and distortion cause a visualization to become difficult to understand, some metaphors that are readily detectable (e.g., human faces) or understandable (e.g., a magnet metaphor) can help users deal with the complexity. Examples include Chernoff faces [11] and Dust & Magnet [55].

One interesting pattern common to these strategies is the trade-off between different elements of the visualization. If one wants to show more data or attributes either through distortion or compression strategies, the resulting visualization becomes visually complex. If one would like to lower complexity through the divide and conquer strategy, the amount of data shown in a single view will decrease. The key issue is striking a balance between these two factors.

3 Overview: Visual Analytics for mSNA

The goal of this study was to support our social scientist colleagues in performing multimodal social network analysis. For this purpose, we recruited a social scientist with professional interests in mSNA as an active collaborator and co-author for this project (the third co-author A3). While existing literature presents theoretical and analytical frameworks for two-mode networks, those are yet to be fully expanded to multimodal ones. Therefore, we decided to employ an exploratory and user-driven design process with the following stages:

I. **Early design:** formative sketching, brainstorming, prototyping, and requirements gathering;

II. **Iterative tool development:** progressively refining our visual analytics tool based on domain expert feedback;

III. **Formative evaluation:** qualitative evaluation of our visual analytics tool with five social scientists;

IV. **Iterative tool refinement:** creating additional features based on feedback from the formative evaluation; and

V. **Summative evaluation:** qualitative evaluation of the current state of our visual analytics tool using domain experts.

4 PHASE I: EARLY DESIGN

The early design process consisted of brainstorming, sketching (Figure 1), and reviewing existing work in the domain. We built an early prototype with a sample data set (NSF funding award data), so that A3 could make sense of the effectiveness of visual analytics. Based on A3's input, this low-fidelity prototype was a compound node-link diagram with color-coding to convey mode information. The prototype supported zooming and panning as well as interactive graph layout. Its introduction helped us derive specific and contextualized questions that social scientists might want to answer using the tool. Some of these questions fall within the range of research questions discussed in Section 2.2; some do not (e.g., q5), yet add utility to the tool:

- q1 Who are the most successful investigators? Which institutions are they from? (Q1a and Q1c)
- q2 Who are the collaborators of a particular investigator? Have they collaborated on multiple projects? (Q1b and Q2c)
- q3 What are the overall patterns and trends in collaboration and funding? (Q2a, Q2b, and Q2c)
- q4 Which program manager awarded most projects? (Q1a)
- q5 How can I find programs with specific subjects or contents?

These questions helped prioritize features to be implemented and worked as test cases to verify whether features have a true purpose in designing the initial prototype. However, at the same time, these tasks tended to be mere extensions of tasks existing for unimodal social network analysis. We found that it was quite challenging for our domain experts to come up with such questions this early in the design process. These difficulties are further discussed in Section 9.4.

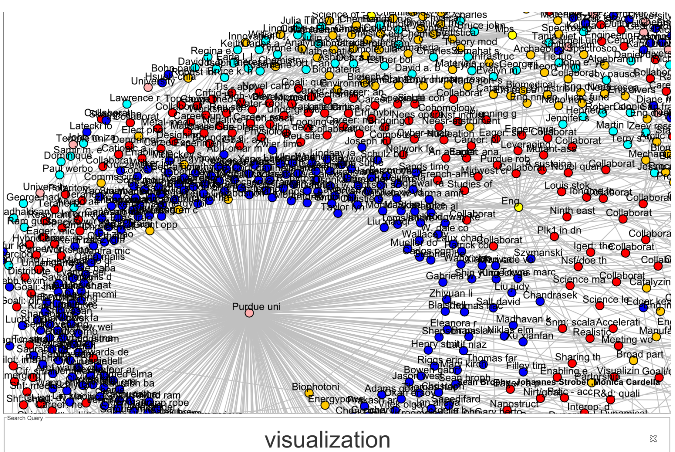


Fig. 2. Compound network visualization in MMGraph. This view is used in parallel with the PNLBs view (Figure 3) on a second monitor.

5 PHASE II: ITERATIVE TOOL DEVELOPMENT (MMGRAPH)

Starting from our low-fidelity prototype from the early design phase (above), we then built an initial visual analytics tool for mSNA that we call MMGRAPH. MMGraph is a Java application built using the Piccolo library [5]. The tool loads multimodal graphs using the GraphML format where a specific node attribute `type` is used to convey the mode of each vertex. The initial visualization we provided was a standard compound network visualization (see Section 2.3.1) with color-coding to visualize mode (Figure 2).

5.1 Parallel Node-Link Bands

One of our earliest findings from the iterative design process was that the compound network visualization was not appropriate for multimodal social network analysis. While it was familiar to A3, it also caused high visual complexity and made distinguishing between within-mode and between-mode ties difficult. Our conclusion was that a more structured organization of the visual space was necessary.

Based on our design process (above), we therefore added[2] *parallel node-link bands* (PNLBs) (Figure 3) to MMGraph. PNLBs use visual node and link marks partitioned into separate *bands* based on modes to minimize visual clutter, yet which arranges the bands in parallel to maintain cohesiveness. This design actively omits all edges except for those between bands that have been placed adjacent to each other on the visual space (neighboring between-mode network), which improves scalability without affecting data accuracy. In other words, all the edges between nodes within a band (within-mode network) and all the edges between nodes that belong to non-neighboring bands (non-neighboring between-mode network) are hidden. The bands can be reordered to expose other between-mode networks. The technique draws inspiration from the list view in the investigative analytics tool Jigsaw [48], where entities of a particular type are arranged in separate lists and relations between entities are drawn as lines connecting them.

The motivation behind this visual design is to impose structure on the multimodal graph without sacrificing the familiar node-link representation while organizing relations between entities in a logical manner. Based on our discussion on strategies for dimensionality reduction (Section 2.4), our approach uses a combination of divide and conquer (not showing both within-mode and between-mode network simultaneously) and distortion (change the node-link layout to the parallel node-like layout) to achieve an efficient visual representation.

Given that one of our central design requirements is to maintain familiarity for social scientists, it is worth taking a step back to evaluate how PNLBs Correlate to standard node-link diagrams. One important advantage in this respect is that PNLBs retain the same node-link diagram format, even if the layout is different and no longer as free-form and organic. On the other hand, this layout still follows the same basic idea as the conventional visual layout of bipartite graphs familiar to most social scientists. The remaining hurdle is how to communicate the fact that PNLBs hide all edges except for those between adjacent bands as well as those within the same band.

5.2 Design Space

We now explore the free parameters in the PNLBs design space to fully map out the utility of the technique and support the user tasks.

Node Representation. The node representations in each band should remain similar to those in a node-link diagram, but can be augmented with additional visual variables. We use the border color for each band to encode mode and also add the name of the mode to the header part of the band. In addition, for multimodal graphs where textual data has been associated with some of the nodes, we provided a word cloud popup interaction that shows a frequency-based summary of the keywords associated with the node under the mouse cursor (Figure 4(a)). We use a simple tf-idf [32] mechanism to extract these keywords. A related approach is to provide an ego-network popup that instead shows the direct neighbors of the node under the mouse cursor as a node-link diagram, with the node itself at the center and the neighbors in a radial layout (Figure 4(b)).

Edge Representation. The display space between adjacent bands is reserved for rendering the between-mode edges that span the neighboring node bands. We use a edge representation based on direct lines connecting the two closest parts of the nodes in adjacent bands. For these between-mode networks, edge representations easily become cluttered due to many edge crossings. Thus, clutter reduction techniques [17], such as edge bundling [28] and illustrative parallel coordinates [41], could be utilized. For within-mode networks, i.e.,

[2]This means that MMGraph has both the compound network visualization and PNLBs, which are shown in parallel on dual displays.

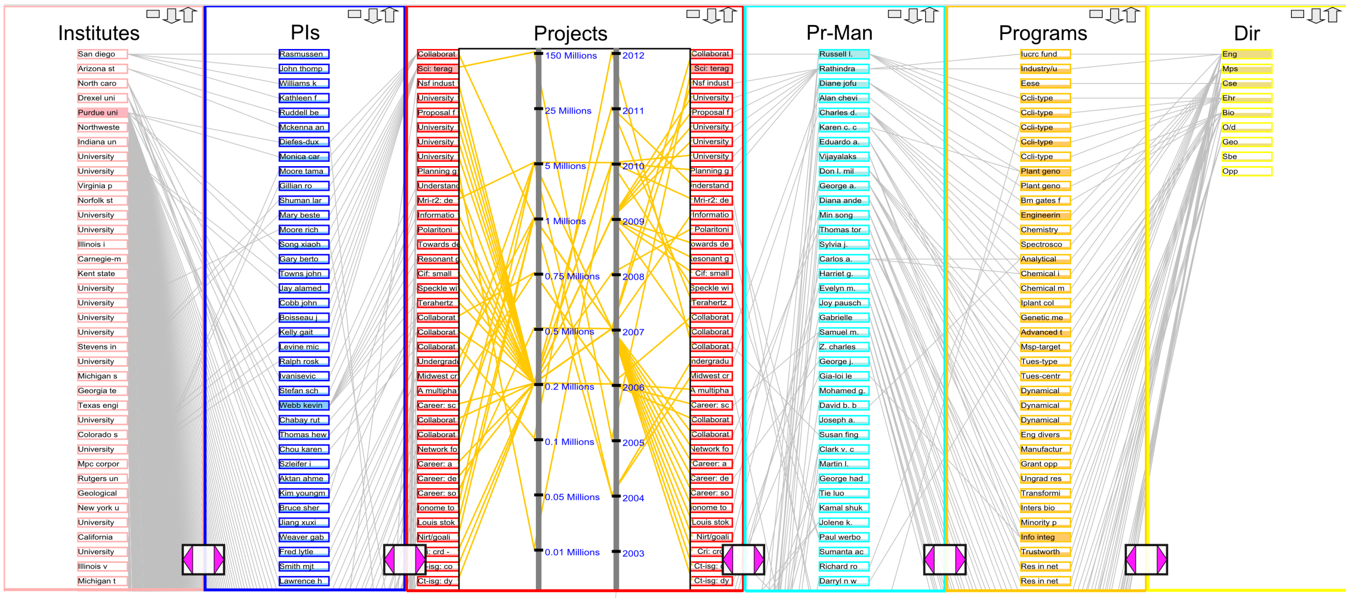
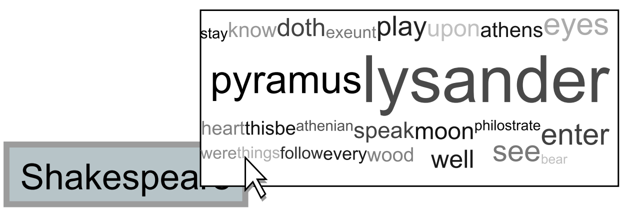
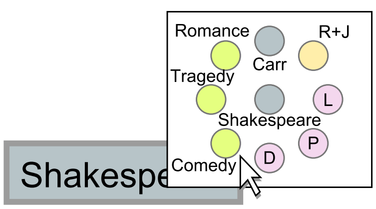


Fig. 3. Parallel node-link bands (PNLBs) being used to visualize multimodal NSF funding data consisting of Institutions, PIs (and Co-PIs), Projects, program managers (Pr-Man), NSF programs (Programs), and NSF directorates (Dir). Color glyphs inside nodes represent degree centrality.



(a) Word cloud of node contents.



(b) An ego-network popup

Fig. 4. Exploring additional visual representations for node attributes.

those that connect nodes within the same band, several design alternatives exist. One approach is to use arcs—similar to arc diagrams [52] and MatLink [29]—since the nodes reside in the same band. However, we found that these approaches make a narrow vertical band too cluttered. Another alternative is to provide an ego-network that only shows the within-mode network of a particular node (Figure 5), which we implemented in MMGraph as the "within-network" view. The within-network view is slightly different from the ego-network popup, which shows all the neighboring nodes regardless of bands that it belongs to without showing any edges. Though the within-network view only shows neighboring nodes in the within-mode network, the between-mode edges between the neighboring nodes and the nodes in the adjacent bands are also shown. This provides additional insights, such as "how are my collaborators associated with other projects and institutes?" as shown in Figure 7(b).

Within-Mode Sorting. Organizing nodes within a particular band is akin to graph layout on a single graphical axis (the vertical axis). The order of nodes inside the stack for each band is a free parameter, and can be controlled in several different ways (exposed to the user):

- **Node attributes:** The user may sort nodes based on node attributes such as name, age, or income.
- **Edge attributes:** Edge attributes, such as time or weight, or the number of edges to a particular node (i.e., its total degree, or its degree within a specific mode) can be used for sorting.
- **Connectivity:** A common ordering is to reorder a band by its connection to nodes in another band. We support this by a "bring-to-top" command that reorders all bands to bring all the neighbors of a particular node to the top of their bands.

Out of these three approaches, sorting by edge attributes, particularly by between-mode centrality (implemented as "connection to the right" or "connection to the left" in the system) could be relevant to mSNA. These features can easily answer questions such as "Who is the most successful grant writer? (Q1a, q1)," because it is equivalent to finding "Which node in the PI mode has the highest degree centrality based on its ties to the projects mode."

Open Sesame. To close the loop between PNLBs and parallel coordinate plots, we propose a band extension mode where a selected band is parted in half using an animation by invoking what we call an "open-sesame" interaction. The parting animation then exposes a parallel coordinate display for multivariate data in the space between the labels. This display would be used to represent multivariate node data, such as time stamps, quantities, and ordinal values. Figure 3 shows a screenshot from our implementation where the mode "Projects" has been expanded in this way. This combination of PNLBs and parallel coordinates, where parallel coordinates reveal attribute values inside bands on user request, is matched with Q1c. This approach is also more powerful than other encoding approaches because it can present multiple attributes of nodes in a mode at the same time, and it can be used in conjunction with existing interaction techniques developed for parallel coordinates. For example, filtering over multiple dimensions (e.g., time and grant amounts in Figure 3) turned out to be a powerful way to select a set of nodes and edges out of the global network.

Additional Interaction and Navigation. Beyond the above interactions, our iterative design process caused us to add several additional interaction and navigation techniques to the MMGraph tool: highlighting (by hovering over a node), brushing (by selecting one or several nodes), searching (by entering queries in a text box), panning, and zooming (inspired by TableLens [42] as shown in Figure 6).

6 PHASE III: FORMATIVE QUALITATIVE EVALUATION

As part our iterative design process, we conducted a qualitative study to evaluate our prototype implementation for mSNA tasks.

6.1 Method

We invited five domain experts (4 graduate students and 1 faculty member from our university's school of communication) as study participants. All participants had been professionally analyzing social networks for more than one year (mean = 2.7 years). They reported that they had experience in network data, such as Facebook friends networks, terrorist networks, donation networks, and authorship networks. All of them had experience with the SNA tool UCINET [8].

An experimenter, the second author of this paper who was not involved directly in developing the system, administered the experiments for all five participants. At the beginning of each experiment, the experimenter described the experimental procedure and tool for ten minutes. For the next forty minutes, each participant was asked to use the tool to answer four mSNA tasks which are explained in Section 6.3. During the task, each participant was encouraged to think aloud. The experimenter also engaged participants in conversation by asking questions. After a participant completed all the tasks, the experimenter interviewed the participant about the experiment and the tool for ten minutes. We collected audio and screen recordings of the experiments. Each experiment lasted around one hour.

6.2 Dataset

Because of its general interest to scientists in the United States, we decided to use a multimodal dataset derived from funding data from the U.S. National Science Foundation (NSF). NSF provides a publicly available database of awarded grants that dates back several decades. The NSF award search[3] allows for advanced queries and saving search results as tabular data. Using this tabular data as source, we built a multimodal graph using a process similar to Liu et al. [39], and stored it as GraphML with a node attribute encoding the mode.

Since the entire NSF funding dataset is large in size (some 330,000 awards), we narrowed it down by specifying the awardee organization as Purdue University and awarded years as 2003 to 2012. Despite only selecting Purdue, the dataset includes a total of 95 institutions because many projects have external partners. Our final dataset includes six modes: 315 projects, 205 programs, 507 PIs and co-PIs (PIs henceforth), 95 institutions, 160 program managers (Pr-Man), and 9 directorates (henceforth Dir). We merged PIs and Co-PIs into a single mode (PI) to avoid duplicating individuals. This also yielded a within-mode network that connected investigators with their collaborators.

The NSF dataset is organized as follows: A PI is affiliated with one or more institutions. A PI collaborates with zero or more PIs (this is the only within-mode network in this NSF dataset). A PI is involved in one or more projects. Each project is awarded by one program manager (Pr-Man). Each project belongs to one or more programs. A program manager belongs to one or more programs. A program belongs to a directorate (Dir).

[3] http://www.nsf.gov/awardsearch/

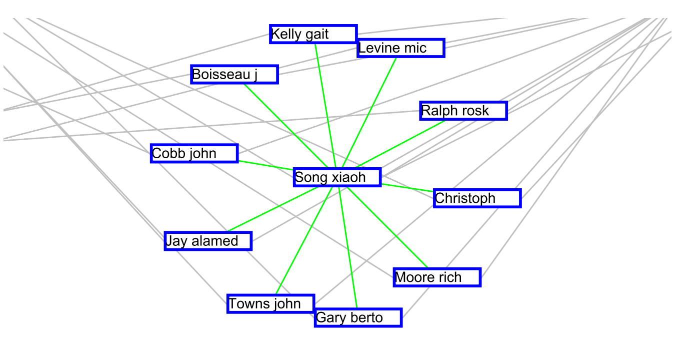


Fig. 5. Within-mode network view for studying relations inside a band.

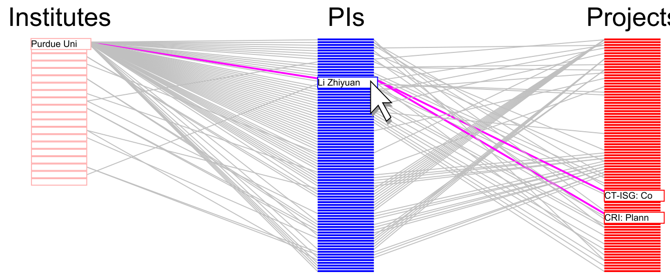


Fig. 6. Data abstraction mode inspired by the TableLens [42].

6.3 Tasks

The tasks in the experiment included the following questions:

- T1 Who is the most successful PI in terms of number of awards?
- T2 Are there PIs who have been awarded several grants together?
- T3 Do some program managers often award grants to the same PIs?
- T4 Which are projects that have been awarded more than $70M, and what are their commonalities?

While some of these tasks could certainly be solved using computational means, perhaps even more efficiently than using visual analytics, we wanted to include a broad spectrum of tasks to reflect how an mSNA tool is used in realistic settings. Even though we are strong believers in visual analytics augmenting existing (computational) tools, it is also true that switching to another application in mid-analysis may break the user's flow. For this reason, our tasks were motivated by the graph task taxonomy proposed by Lee et al [38]. Thus, task T1 could be answered by analyzing the between-mode network between PI and Project, whereas T2 required analyzing the within-mode network of the PI mode. To answer T3, participants would have to connect the indirect relationship between Program Manager and PI. Finally, T4 was an open-ended task that required participants to delve into network data by formulating their own hypotheses and testing them using the visual analytics tool. Based on our argument above, we started from basic tasks (T1 and T2) that could be answered with a simple interaction, and gradually exposed participants to more difficult tasks. We also let participants pursue any interesting serendipitous observations found throughout the analysis process and report on their findings.

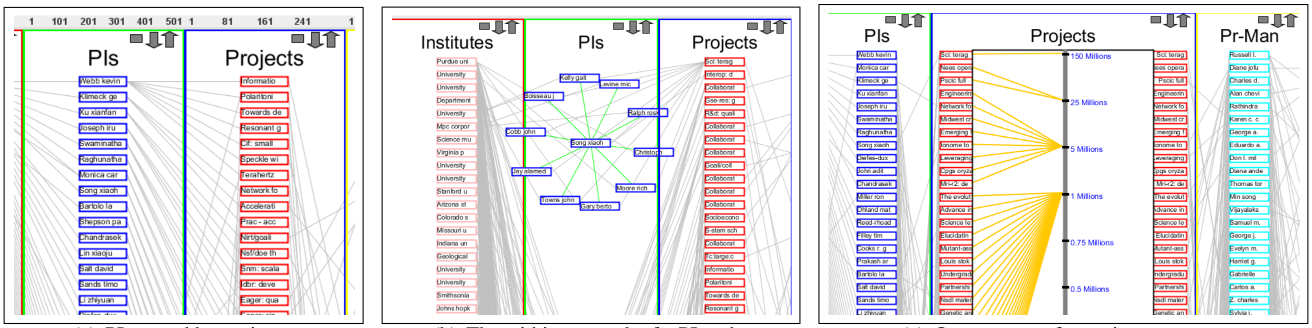
6.4 Results

In general, participants successfully completed all tasks. T1 and T2 were completed without major problems, and using similar approaches. To complete T1, participants sorted the PIs mode by connection to the Projects mode (the degree centrality of the between-mode network between the PI and Project modes), which shows that Kevin Webb (a professor of electrical and computer engineering) has seven projects (Figure 7(a)). T2 does not have a definite answer, but participants found that they could sort the PIs mode by within-mode connections (sort by the degree within a specific mode) to find PIs who have high collaborations with others and check whether some of them were repeated collaborators using the within-network view (Figure 7(b)).

In contrast, participants' approaches varied when solving T3. Some used brushing to see the relationship between PIs, Projects, and Program Manager to see if there were any common occurrences over multiple projects. Others simply used ego-networks to see if the same pair of PI and program manager appeared at the same time.

Since T4 is a more open-ended task, the usages of PNLBs were more diverse. Initially, all the participants successfully used the open-sesame interaction to find the project awarded more than $70M (see Figure 7(c)). Subsequently, some used word clouds over projects; others moved all of the associated PIs and program managers to the top of the band to see their relationships. Some even investigated beyond the specific scope of T4 and searched for what other projects the PIs of the 70 million-dollar project have worked on and explored that data.

All participants stated that they preferred the PNLBs view over the compound view for virtually all mSNA tasks. In particular, they liked



(a) PIs sorted by projects. (b) The within-network of a PI node. (c) Open-sesame for project amounts.

Fig. 7. Screenshots of the PNLBs view highlighting several user interactions during the user study.

the idea of "dissecting complicated datasets into multiple bimodal sets," so that they could focus on two adjacent bands at a time. They noted that this tool was easy to use for first-time users, and that it took less than an hour to be proficient at using it. One participant also stated that "patterns and insights found in this view are more digestible due to its structure [than the compound network visualization]." However, several participants stressed that having the mutually interlinked compound and PNLBs views allowed for transferring details from multimodal networks to the network overview and vice versa.

Participants also felt that some functions in the PNLBs view were particularly helpful for mSNA tasks. For example, they all seemed to enjoy how interactions often brought about new hypotheses and easily supported exploring them. One participant used the open-sesame interaction to determine the project with the largest grant amount. She then used the word cloud to view the details of that project. This made her curious if other projects with similar keywords had been awarded. She queried the keyword and found another project, and viewed the ego-network of that project. Several PIs were common to both projects, and the participant then viewed their within-network. She wondered if these projects were granted by the same program manager, and so brought all their program managers within view. Such long sequences of actions were well-supported by our tool, something which the participants informed us that they appreciated.

7 Phase IV: Iterative MMGraph Refinement

The purpose of our formative evaluation was to evoke feedback on the MMGraph tool, which to that point had only been guided by A3, our social scientist co-author. Below we review this feedback and then discuss the concrete changes we made to MMGraph in response.

7.1 Formative Feedback

All participants expressed a desire to be able to use the tool with their own social networks, such as authorship networks, donation networks, terrorist networks, and organizational networks. One participant, who has an interest in terrorist networks, believed that the tool could be helpful for finding patterns in terrorist networks by studying the relationship between sponsors, operators, and cell leaders. Another participant, who studies donation networks, stated that the tool could be used to identify whom she should target for raising certain types of donations by analyzing previous interactions. She hoped to use open-sesame interaction to view the key characteristics of potential donors.

A few weaknesses of the PNLBs view were also uncovered. Some participants disliked the fact that the compound network visualization was underused, and some wanted to use the view as a canvas where they could filter out the nodes they did not want to study. Others proposed clustering the nodes in the compound view based on the projected network or other network metrics. Most participants wanted to be able to see connections beyond the most adjacent modes in the PNLBs view. They wanted to have an ego-network starting from a node in one band to all nodes in all bands, which would provide a summary for a particular node of interest. They also proposed streamlining the interface by replacing menus with toolbars and dialog boxes to make the exploration faster, more effortless, and more discoverable.

Perhaps the most significant feedback, echoed by several participants, was requests for integrating traditional network metrics into the tool. In other words, if MMGraph to that point had emphasized visual aspects of multimodal graphs, this feedback effectively highlighted the need for computational metrics and analytics components in the tool.

7.2 MMGraph Refinements

We participant feedback to make additional improvements to MMGraph, including several user interface and interaction refinements. Participants also noted that the node representation could be utilized to visualize particular network and data attributes. We implemented this new feature as water-level color glyphs (Figure 8). User-selected attribute values are normalized and used to fill the node.

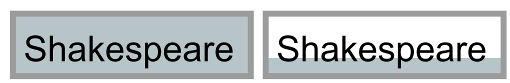


Fig. 8. Water-level color glyphs representing a node value.

Beyond such minor changes, the most significant feature we added during this second development phase was the ability to order, filter, and visualize nodes based on network metrics. Of course, the fact that most of these metrics are defined for unimodal and not multimodal networks meant that we had to derive these definitions ourselves:

- **Multimodal degree centrality:** the number of edges for a vertex that connect other vertices either within the same mode (*within-mode degree centrality*), or to vertices only in other modes (*between-mode degree centrality*). Can also be defined for a specific mode, i.e., the degree centrality for a vertex to mode A.

  This is a measure for how well-connected a vertex is to vertices in one or more modes in the multimodal network.

- **Multimodal betweenness centrality:** the multimodal betweenness centrality of a vertex in one mode is equal to the number of shortest paths from all vertices to all others in other modes that pass through that vertex. Can also be defined for specific modes.

  The measure captures how important a vertex is in connecting vertices in other modes. For example, in a co-authorship network, papers with high multimodal betweenness centrality connects many authors who otherwise have not collaborated.

- **Multimodal closeness centrality:** the sum of the shortest paths from a certain vertex in one mode to all the other vertices in other modes. Again, can be similarly defined for specific modes.

  This measure should be interpreted as the distance of the vertex to all other vertices in their modes. For example, in a database-user network, a database with low closeness centrality would quickly disseminate information to all users.

8 Phase V: Summative Qualitative Evaluation

Finally, to validate our improvements and new analytics capabilities in MMGraph, we conducted a second qualitative evaluation. We used the same NSF dataset as in the previous evaluation (Section 6.2).

8.1 Method

We recruited three domain experts: one faculty member (P1) and two graduate students (P2 and P3) from the same school of Communication to learn how our improved MMGraph supported mSNA. P1 had also participated in the previous formative qualitative evaluation, but P2 and P3 had never seen MMGraph before. Instead of providing pre-created tasks as in Section 6.3, in this phase, we allowed participants to freely study the NSF network, generate questions, and solve them using MMGraph. The purpose for this phase was to observe how the newly added features (the node attribute visualization in Figure 8 and additional multimodal network metrics) were used and, more importantly, what kinds of other tasks participants would like to perform.

8.2 Results

Overall, all participants successfully understood how to use the tool, did not encounter major problems, and were positive about its capabilities. The open-ended evaluation design worked well; all participants used the tool to come up with interesting network questions. In fact, this allowed us to both validate MMGraph's new features introduced in Phase IV as well as continue to study its general utility for mSNA.

8.2.1 Validation of New Features

The newly added multimodal network metrics (Section 7.2) were appreciated by our participants. For example, P2 wanted to determine the researcher at Purdue University who had been awarded the most NSF grants with PIs from other institutions. The participant found the answer by sorting the PI band by multimodal betweenness centrality. Similarly, P3 wanted to find program managers who provide grants to the largest number of PIs, or, in other words, program manager who accepted a variety of different PIs. Again, sorting program managers by multimodal betweenness centrality with respect to PI yielded this information. A participant noted that sorting a band by centrality measures can show how closeness, betweenness, or degree can impact the tie density to other bands using MMGraph.

The fact that MMGraph now also encodes other node attributes (such as funding dates) using color glyphs also helped participants quickly make sense of the dataset and generate interesting questions to explore further. Earlier in the session, P1 found that a certain rotating program manager had awarded grants to more than 10 different projects in the participant's own research area. This was an entirely new insight for P1, and the participant was curious about the availability of the program manager since such rotating program managers tend to stay no longer than a few years. Surveying the node attributes, the participant performed the open- sesame interaction to see which dates the grants had been awarded by the particular program manager. It turned out that the program manager had granted all those projects in 2008 and 2009. This led the participant to conclude that the program manager had finished his or her tenure at the NSF, and was likely not a good contact for future grant proposals.

8.2.2 Impacts of MMGraph on mSNA

Beyond validating the new features added to MMGraph, we were also able to more clearly observe how MMGraph impacts mSNA. Our participants articulated this more clearly in this phase probably because we did not give specific tasks here. Particularly, we had an interesting discussion with P1, who mentioned that a social network analyst often starts his or her investigation with a particular node instead of the whole network. The node is selected based on various network metrics, such as centrality. After learning about the node, the analyst gradually expands the scope of investigation to connected nodes.

In contrast, in our evaluation sessions, participants demonstrated that they conducted a more structured, step-by-step investigation as follows: First, participants used using PNLBs to gain an overview of the between-mode relationships across all modes. The overview led participants to successfully exclude uninteresting between-mode networks (e.g., having a low number of edges drawn between the two modes). Second, after the overview, participants investigated the mode-to-mode relationships instead of delving into a single node. Using multimodal social network metrics, participants were able to sort a band of interest and understand the correlation between the band with others. Participants were also able to easily investigate the correlation between the within-mode network of one mode and the between-mode network with another mode. Third, after finding an interesting phenomenon between modes, participants investigated details of interesting nodes by using the open-sesame interaction and/or the popup view for word clouds. These three steps were repeated iteratively throughout the mSNA process.

9 Discussion

Our two qualitative evaluations raised many interesting points and insights; we highlight the important ones below.

9.1 Benefits of PNLBs on mSNA

Throughout the design study, we learned that MMGraph provides several benefits over common mSNA approaches (e.g., the compound network visualization and network metrics).

First, MMGraph, especially PNLBs, provides an effective structure for mSNA, which cannot be supported by compound network visualization. We believe that compound network visualizations contain too much information in a single view without proper abstraction. This complexity hampers social scientists from seeing the big picture, and may lead them to focus on an individual node at the outset of the investigation. In contrast, the mode-by-mode division provided by PNLBs seems to be a proper external representation that social scientists can easily understand and work with. The zoomed-out view of PNLBs worked as a useful overview of the whole multimodal network data despite not readily providing within-mode ties. Participants also effectively focused on mode-to-mode relationships. This finding is consistent with the lesson in our previous study [35], where an explicit visualization of temporal data changes the investigative analysis process to become more of a top-down process rather than bottom-up.

Second, tight integration between network metrics and visual representation provides a seamless train of analysis. As shown in the two evaluation studies, participants appreciated that various network metrics are used for ordering and encoding through glyphs instead of being simply presented as a list of numbers. We believe that this tight integration allowed our participants to iteratively ask questions and answer them immediately, using that insight in their next question. Such train of analysis could be easily broken if one has to export data in order to run statistical tests in a separate program.

9.2 Divide-and-Conquer and Details-On-Demand

The design strategy used for PNLBs is based on the divide-and-conquer approach (see Section 2.4) as well as showing relations on demand. We intended to minimize visual clutter and complexity, yet it forced users to perform mSNA tasks more efficiently.

We chose the between-mode network as the dominant visualization for PNLBs because our investigation revealed that between-mode networks tended to be most interesting to social scientists. During the evaluation sessions (Phase III and V), we learned that research questions are often related to finding nodes in a mode that play a central role in certain events (e.g., donations, terrorist plot, and idea generation), which could be answered by dissecting the multimodal network into bimodal networks and analyzing them separately.

In addition, to satisfy the needs for further investigation, we designed the PNLBs view to show the within-mode network and other node attribute information *on demand*. Since representations for the between-mode ties and bands already occupy the whole display, we devised an option to let the user view the within-mode network of each node as the within-network view (Figure 5). While using the within-mode network view, users could still preserve the overview of the compound network visualization and PNLBs. Beyond the within-network popup, we provide multivariate node attributes using a dynamically activated parallel coordinate view using the open-sesame interaction only at the explicit request of the user (Figure 7(c)).

Some may argue that such a discontinuous analysis path may harm users because they cannot see the overall picture all at once. However, in our evaluations, we found that this step-by-step approach seemed well-suited to how our users performed the sensemaking task. As participants revealed in their comments, the compound network visualization easily became too busy when the number of nodes and edges increase. PNLBs, on the other hand, force users to view networks partitioned into modes by revealing only a limited amount of connectivity information at a time. Starting from this very structured representation, users can progressively unlock new information on-demand using interaction, enabling a progressive refinement of the analysis [22].

9.3 mSNA Tasks

Throughout this design study, we gradually learned more about what kinds of tasks social scientists would like to perform during mSNA. Many tasks came from observations of how participants explored and learned about the provided NSF data using MMGraph. Other tasks were captured from the discussion regarding their own data and how they would analyze their data using MMGraph. While the following list is not exhaustive, it is an initial set of tasks for mSNA, which may contribute to increase task clarity in mSNA. Note that this list does not not include common tasks that are also found in unimodal SNA (e.g., seeing the distribution of a network metric over a particular mode), and the design of MMGraph may bias the kinds of tasks that we elicited:

- Identify a node with an extreme network metric w.r.t. another mode (e.g., in a media network with media, words, and readers as modes, P3 wanted to find the most central medium, word, and reader with respect to each other).
- Correlate within-mode and between-mode network (e.g., in the NSF dataset, participants were curious whether a successful collaboration (within-mode network) repeated on other projects (between-mode network)).
- Correlate attributes and between-mode network (e.g., in a Wikipedia edit network with authors, edits, and discussions as modes, P2 wanted to view correlation between the edit date (attribute) and the edit counts from a group of authors (between-mode network)).
- Correlate attributes and within-mode network (e.g., in a university student network, P2 wanted to view correlation of cultural background of students (attribute) with their friendships (within-mode network)).

Together with unimodal graph tasks [38], these are the types of tasks that visual analytics tools for mSNA should support.

9.4 Visual Analytics Studies with Social Scientists

We here discuss problems that our social science collaborators encountered while working with visualization researchers on this project. By their nature, visualization researchers tend to be technique-driven and are eager to come up with novel visualizations. This makes it difficult for domain experts to emphasize the problems to be solved. We found that some ideas originating from our social scientist collaborators tended to be inadvertently discouraged or downplayed by the visualization researchers in the team simply because they were not novel enough (from a visualization viewpoint) or too complex to implement.

Furthermore, the analysis of social network data in social science fields is usually driven by established theories. In this sense, the exploratory analysis of data supported by visual analytics yielded a radically different approach to sensemaking than traditional methods, which caused some difficulties to bridge. Lastly, initial paper prototypes may not be sufficient for social scientists to fully understand the significance and potential of a technique. We found that our social scientist colleagues were best able to understand when confronted by a concrete visualization (e.g., PNLBs) with a specific context (e.g., the NSF multimodal dataset) as Lloyd and Dykes [40] suggest.

10 Conclusions

We have presented a design study on applying visual analytics to multimodal social network analysis (mSNA). Our study, which involves a social scientist at the outset and was methodically conducted in five phases that each fed the next, involved initial early design, iterative tool development, a formative evaluation, a second iterative development phase, and a final summative evaluation. The resulting MMGraph tool combines both a compound network visualization as well as a view using a visualization technique called parallel node-link bands (PNLBs). Through the iterative design processes, we not only refined MMGraph but also learned what kinds of tasks social scientists would like to conduct while analyzing a multimodal network. We also learned the effectiveness of divide-and-conquer and details-on-demand in designing an effective tool for social scientists.

10.1 Limitations

Obviously, the PNLBs technique is not universally effective. Even though our evaluation studies show that the abstraction used in PNLBs is well accepted by our research participants (see Section 9.2), PNLBs may not effectively support mSNA in the following conditions: (1) when multiple combinations of between-mode networks should be considered (PNLBs basically support multiple bimodal, or pair-wise, network analysis instead of true multimodal network analysis); (2) when complex multimodal networks configurations where modes overlap each other or have hierarchical relationships (TreeNetViz [23] and its variations would be more effective in this case); and (3) when the size of a network is too large (an abstraction above individual nodes and edges should be provided in this case).

In addition, there may some limitations in our evaluation studies, such as a small number of research participants and a lack of diversity in used datasets. Further longitudinal studies with more research participants and user domain data would be necessary, using the results from our studies as a starting point and guiding example.

10.2 Future Work

The visual analytics tools and techniques designed in this design study are only a small group of methods specifically designed for such tasks, and we think that the space is wide open for further work. For example, multimodal networks often represent affiliation ties and social circles, which provide conditions for future connections. In this sense, visual analytics of longitudinal multimodal networks can expand our understanding of network dynamics. In addition, additional user studies with their own domain datasets would deepen our understanding of what kinds of tasks that they would like to conduct while dealing with multimodal network analysis.

Acknowledgments

We thank our social scientist colleagues who participated in the study as well as Sukwon Lee, Sung-Hee Kim, and Aditya Srinath who provided constructive feedback on this study. This work was partly supported by the U.S. Department of Homeland Security's VACCINE Center under award 2009-ST-061-CI0003.

References

[1] E. Adar. GUESS: a language and interface for graph exploration. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 791–800, 2006.

[2] A. Aris and B. Shneiderman. Designing semantic substrates for visual network exploration. *Information Visualization*, 6(4):281–300, 2007.

[3] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International Conference on Weblogs and Social Media*, 2009.

[4] L. Beauguitte. Basic notions on social network analysis - dealing with 2-mode networks. Technical report, Sept. 2010.

[5] B. B. Bederson, J. Grosjean, and J. Meyer. Toolkit design for interactive structured graphics. *IEEE Transactions on Software Engineering*, 30(8):535–546, 2004.

[6] R. Bellman. *Adaptive control processes*. Princeton University Press, Princeton, NJ, 1961.

[7] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J.-D. Fekete. GraphDice: A system for exploring multivariate social networks. *Computer Graphics Forum*, 29(3):863–872, 2010.
[8] S. P. Borgatti, M. Everett, and L. C. Freeman. *UCINET for Windows: software for social network analysis*. Analytic Technologies, Harvard, 2002.
[9] S. P. Borgatti and M. G. Everett. Network analysis of 2-mode data. *Social Networks*, 19(3):243–269, Aug. 1997.
[10] U. Cantner, M. Guerzoni, and A. Martinelli. Innovation and market dynamics: A two-mode network approach to user-producer relation. Maastricht, Apr. 2011.
[11] H. Chernoff. The use of faces to represent points in K-Dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, June 1973.
[12] J. H. T. Claessen and J. J. van Wijk. Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2310–2316, Dec. 2011.
[13] C. Collins and S. Carpendale. VisLink: revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1192–1199, Dec. 2007.
[14] N. Contractor, P. Monge, and P. Leonardi. Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network. *International Journal of Communication*, 5:682–720, 2011.
[15] T. F. Cox and M. A. Cox. *Multidimensional scaling*. CRC Press, 2001.
[16] C. Dunne, N. H. Riche, B. Lee, R. Metoyer, and G. Robertson. GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the ACM Conference on Human Factors in Computer Systems*, pages 1663–1672, 2012.
[17] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.
[18] K. Faust. Using correspondence analysis for joint displays of affiliation networks. In P. J. Carrington, J. Scott, and S. Wasserman, editors, *Models and methods in social network analysis*, pages 117–147. Cambridge University Press, 2005.
[19] S. K. Feiner and C. Beshers. Worlds within worlds: metaphors for exploring n-dimensional virtual worlds. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 76–83, 1990.
[20] L. C. Freeman. Visualizing social networks. *Journal of Social Structure*, 1(1), 2000.
[21] L. C. Freeman. Graphical techniques for exploring social network data. In J. Scott and S. Wassermann, editors, *Models and methods in social network analysis*. Cambridge University Press, 2004.
[22] D. Gotz, M. X. Zhou, and V. Aggarwal. Interactive visual synthesis of analytic knowledge. In *Proceedings of the IEEE Symposium on Visual Analytics Science & Technology*, pages 51–58, 2006.
[23] L. Gou and X. L. Zhang. TreeNetViz: revealing patterns of networks over tree structures. *IEEE Transactions on Visualization and Computer Graphics*, 17(6):2449–2458, Dec. 2011.
[24] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllerer. SmallWorlds: Visualizing social recommendations. *Computer Graphics Forum*, 29(3):833–842, 2010.
[25] R. Guimerà, B. Uzzi, J. Spiro, and L. A. N. Amaral. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722):697 –702, Apr. 2005.
[26] D. J. Hand, H. Mannila, and P. Smyth. *Principles of data mining*. MIT Press, Aug. 2001.
[27] D. Hansen, B. Shneiderman, and M. A. Smith. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann, 2010.
[28] J. Heinrich, Y. Luo, A. E. Kirkpatrick, H. Zhang, and D. Weiskopf. Evaluation of a bundling technique for parallel coordinates. *arXiv preprint arXiv:1109.6073*, 2011.
[29] N. Henry and J.-D. Fekete. MatLink: Enhanced matrix visualization for analyzing social networks. In *Human-Computer Interaction — INTERACT 2007*, volume 4663 of *LNCS*, pages 288–302, 2007.
[30] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, Jan./Mar. 2000.
[31] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the IEEE Conference on Visualization*, pages 361–378, 1990.
[32] S. K. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21, 1972.
[33] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, pages 107–116, 2001.
[34] D. Knoke and S. Yang. *Social network analysis*. SAGE, 2008.
[35] B. C. Kwon, W. Javed, S. Ghani, N. Elmqvist, J. S. Yi, and D. Ebert. Evaluating the role of time in investigative analysis of document collections. *IEEE Transactions on Visualization and Computer Graphics*, 18(11):1992–2004, Nov. 2012.
[36] M. Latapy, C. Magnien, and N. D. Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, Jan. 2008.
[37] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Life in the network: the coming age of computational social science. *Science*, 323(5915):721–723, 2009.
[38] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proceedings of BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 82–86, 2006.
[39] Z. Liu, S. Navathe, and J. T. Stasko. Network-based visual analysis of tabular data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 39–48, 2011.
[40] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2498–2507, 2011.
[41] K. T. McDonnell and K. Mueller. Illustrative parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1031–1038, 2008.
[42] R. Rao and S. K. Card. The Table Lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 318–322, 1994.
[43] J. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, London, UK, 1991.
[44] M. Sedlmair, M. D. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
[45] B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):733–740, 2006.
[46] L. Singh, M. Beard, L. Getoor, and M. B. Blake. Visual mining of multimodal social networks at different abstraction levels. In *Proceedings of the International Conference on Information Visualization*, pages 672–679, 2007.
[47] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave. Analyzing social media networks with NodeXL. In *Proceedings of the International Conference on Communities and Technologies*, pages 255–264, 2009.
[48] J. T. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.
[49] F. van Ham and A. Perer. "Search, show context, expand on demand": Supporting large graph exploration with degree-of-interest. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):953–960, 2009.
[50] T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. van Wijk, J.-D. Fekete, and D. Fellner. Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum*, 30(6):1719–1749, 2011.
[51] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Nov. 1994.
[52] M. Wattenberg. Arc diagrams: visualizing structure in strings. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 110–116, 2002.
[53] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Proceedings of the IEEE Symposium on Information Visualization*, page 21, 2005.
[54] J. S. Yi, N. Elmqvist, and S. Lee. TimeMatrix: analyzing temporal social networks using interactive matrix-based visualizations. *International Journal of Human-Computer Interaction*, 26(11):1031–1051, 2010.
[55] J. S. Yi, R. Melton, J. Stasko, and J. Jacko. Dust & Magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(3):239–256, June 2005.