(a) **Title**

TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations

(b) Author Names

a. Ji Soo Yi
   School of Industrial Engineering, Purdue University
b. Niklas Elmqvist
   School of Electrical and Computer Engineering, Purdue University
c. Seungyoon Lee
   Department of Communication, Purdue University

(c) **Running Head**

TimeMatrix: Temporal Social Network Visualization

(d) **Acknowledgements**

None

(e) **Corresponding Author**

Ji Soo Yi
315 N. Grant Street
West Lafayette, IN 47907
Email: yij@purdue.edu
Phone: +1-765-496-7213
Fax: +1-765-494-1299

(f) Abstract

Visualization plays a crucial role in understanding dynamic social networks at many different levels (i.e., group, subgroup, and individual). Node-link-based visualization techniques are currently widely used for these tasks and have been demonstrated to be effective, but we found that they also have limitations in representing temporal changes, particularly at the individual and subgroup levels. To overcome these limitations, we present a new network visualization technique, called "TimeMatrix," based on a matrix representation. Interaction techniques, such as overlay controls, a temporal range slider, semantic zooming, and integrated network statistical measures, support analysts in studying temporal social networks. To validate our design, we present a user study involving three social scientists analyzing inter-organizational collaboration data. The study demonstrates how TimeMatrix may help analysts gain insights about the temporal aspects of network data that can be subsequently tested with network analytic methods.

(g) Body Text

1. Introduction

A large number of network visualization techniques have been developed and employed for representing social networks since Moreno's sociogram was first introduced in the 1930s (Scott, 2000). They are now increasingly being used by social scientists for examining social phenomena ranging from interpersonal relationships, group behavior, intra- and inter-organizational collaboration, and international flows (Borgatti et al., 2009). Most visualization techniques for social networks are based on node-link diagrams (Henry et al., 2007a), where nodes represent actors and links represent connections between those actors. These graphical images of networks are now providing analysts with quick and intuitive ways of understanding social network data where traditional graph analytical methods fail (Freeman, 2000; McGrath et al., 2003).

With the rapid development of computing technologies in the recent decade, collecting longitudinal social network data has become easier than ever before. These dynamically updated social networks provide researchers with even greater opportunities to reveal insights about network change over time as well as the antecedents and outcomes of those network changes (Monge & Contractor, 2003; Powell et al., 2005; Snijders, 2001). Node-link diagrams have also evolved to meet these needs for temporally changing sequences. These techniques primarily take advantage of animation (e.g., Moody et al., 2005), multiple snapshots (or small multiples) (e.g., Tufte, 1990), and three-dimensional views (e.g., Brandes & Corman, 2003; Kapler et al., 2008).

However, there are several limitations for these modified node-link visualizations. Mainly, the edges of node-link visualizations are generally too narrow and long to visually encode additional metrics, such as time. Investigating changes over a long period using animation or small multiples is challenging because our memory and the available screen real estate are limited.

To overcome these limitations, we propose a new approach to visualizing temporal social network data through a matrix-based visual representation, called "TimeMatrix." The visualization is based on temporally aggregated glyphs displaying data in limited screen space, called "TimeCells." These cells also incorporate major statistical measures of social networks (e.g., degree centrality). We have implemented a prototype of TimeMatrix and present results from a user study where the tool was employed by three expert social scientists for analyzing a social network on inter-organizational collaboration activities across 20 years.

2. Background

2.1. Temporal Social Network Analysis

The need for visualizing temporal social networks is a result of the increasing attention to *emergent networks* in social sciences (Monge & Contractor, 2003). As opposed to formal networks that are based on a pre-imposed and static structure, emergent networks capture the flexible and dynamic characteristics of linkages. These linkages have their own life cycle in which they are formed, maintained, dissolved, and reformed over time (Monge & Contractor, 2003). Collectively, these linkages lead to the emergence of social structure from previously chaotic and random states (Kontopoulos, 1993; McKelvey, 1997).

Considering multiple levels in analyzing social networks has been increasingly emphasized, where networks are analyzed at (1) the nodal and dyad levels, (2) the subgroup level, and (3) the global level of entire network (Brass et al., 2004; Contractor et al., 2006). Based on this framework, major inquiries about temporal aspects of networks can be distinguished into three levels as well. At the nodal and dyad levels, the questions center on investigating the changes in the position of nodes as well as capturing the dynamics of tie addition and deletion. At the subgroup level, changes in the patterns of local interactions among groups are of central concern. Especially, when these groups are formed based on attribute similarity, the question relates to the role of node attributes in the processes of network change. At the global level, the key question concerns macro-patterns of structural changes in an entire network (Robins et al., 2005).

More specifically, there are a number of theoretical and empirical questions that can be asked at each of these three levels. First, at the nodal and dyad levels, studies focus on various mechanisms of network evolution to examine the emergence, growth, and dissolution of nodes and ties (Powell et al., 2005). When focusing on nodes, the main question is identifying the central nodes and how they change over time. With regard to ties, an intriguing question is examining the patterns for how new ties are formed. For example, Barabasi and Crandall (Barabasi & Crandall, 2003) suggest preferential attachment as the main driver of tie formation. Powell et al. (Powell et al., 2005) suggest and test four mechanisms in which connections are made: accumulative advantage (richer-get-richer rule), homophily (similarity in traits), follow-the-trend (the dominant choices of others), and multi-connectivity (a pursuit for diversity and multiplexity). In other words, these different organizing principles lead to the changes of social structure over time. Further, studies have paid attention to how ties are retained over time. In particular, the most prominent mechanism concerns the role of inertial forces, in which past ties predict future ties (Granovetter, 1985; Gulati & Gargiulo, 1999). In this sense, studies have examined how networks at earlier points in time predict current networks. This idea also leads to the mechanisms of tie deletion, in which questions about the types of ties that are more likely to weaken and disappear over time are addressed (Burt, 2000, 2002). In all of these areas, a visualization that shows how ties appear, evolve, and disappear can provide useful insights for building initial predictions before network summary measures are obtained.

Networks are often decomposed into partitions so that internal structures can be analyzed (Batagelj et al., 1999). At this subgroup level, the processes by which nodes come together and form dense communities are important aspects of network evolution (Backstrom et al., 2006; Coleman, 1990; Palla et al., 2007). How do subgroups, or collections of nodes, grow over time? How do ties within and between nodes in different subgroups change over time? While some approaches to subgroups are based on connectivity structure (Batagelj et al., 1999), subgroups can also be identified based on the similarities among node attributes. Theories of homophily support this idea (Mcpherson et al., 2001). For example, organizations of the same type or location might form dense ties within themselves or have consistent patterns of connectivity.

At the global level, studies pay attention to the overall change in the network topologies over time (Provan et al., 2007). For example, a network at an earlier point of time might resemble a centralized structure with a dense core and a sparse periphery. Over time, the network might become more decentralized and diversified, representing a cluster structure with a number of visible subgroups. In addition, the global network structure might resemble a small world structure or a scale-free network. These topologies have important implications for the overall efficiency and sustainability of the network.

A variety of statistical measures and analytical methods have been developed to examine and model network structure, and further, longitudinal changes in the structure. Representative measures used in network analysis are summarized in Table 1, and the details of the definition and calculation these measures can be found in (Monge & Contractor, 2003; Provan et al., 2007; Wasserman & Faust, 1994). Table 1 suggests several interesting aspects of these measures. First, some measures are used to capture the characteristics of vertices, while others are for edges. Among the various measures of edges, some are primarily dedicated to addressing the dyad level characteristics, while others are for the subgroup or the global levels. While some measures of individual node attributes remain static, most of the other measures of both vertices and edges can change over time during the course of network evolution. In other words, in order to fully understand the dynamic patterns of social network change, these various measures of vertices and edges at multiple levels need to be examined across temporal spans. At the initial stage of exploring networks, in particular before specific hypotheses can be established, making choices about which measures are to be employed in the analysis could be a challenging task.

[Table 1 will be placed here.]

Visualization could be instrumental in meeting this challenge. Using network visualization, analysts could gain initial insights about dynamic networks and recognize patterns of changes, which could subsequently be tested and proven with confirmatory network statistical methods.

Network changes can often be analyzed by identifying the temporal associations between the attributes of vertices and edges. As noted above, some attributes are static over time (e.g., gender and ethnicity) while other attributes change over time (e.g., degree centrality). Examining the relationships between these variables can answer interesting questions about network dynamics. For example, if an analyst is curious whether a well-connected member in a social networking website attracts more ties than a less connected member does, the analyst would examine the relationship between degree centrality at one point of time and the growth of ties in subsequent time periods. In another example, studies of health-related behaviors deal with the reciprocal interaction between nodes and ties. In other words, are people who have obese friends likely to become obese over time? Or, are people with obesity likely to become friends with each other over time? As these examples show, some research questions involve examining relationships between edge-attributes and edge-attributes, while some address relationships between vertexattributes and edge-attributes. Visual representations that can depict both of these attributes can strongly support the examination of network change.

In summary, to help analysts investigate changes in social networks, we believe that visualization techniques for temporal social network analysis (TSNA) should support the following tasks:

- *Task 1 – Analysis of temporal changes at the global level.* As discussed previously, investigation of dynamic social networks can be performed at three levels (i.e., the individual, subgroup, and global levels). Understanding of global level changes over time is mainly examining changes in the topological structure.
- *Task 2 –* Analysis of temporal changes at the subgroup level. Investigation of subgroup level depends on how the subgroups are formulated. There are two major approaches: one is *aggregation based on connectivity (Task 2a)* (Batagelj et al., 1999), and the other is *aggregation based on node attributes (Task 2b)* (Mcpherson et al., 2001). For example, suppose that we are investigating friendship networks in a high school classroom. The former approach is investigating different cliques and understanding what causes the formation of these cliques; the latter approach is categorizing students depending on their demographic attributes such as ethnicity and investigating how ethnicity influence friendship ties. In both cases, subgroups will change over time.
- *Task 3 - Analysis of temporal associations among nodal and dyad level attributes*. At the nodal and dyad levels, changes of a network can be understood by examining how *node attributes (Task 3a)* and *edge attributes (Task 3b)* (see Table 1) change over time. Further, the temporal associations between these attributes can be examined (*Task 3c*). For example, there are three combinations: edge vs. node, edge vs. edge, and node vs. node. In particular, in the case of a multiplex network when there are multiple types of edges,

the network can be examined by investigating the temporal association between different types of edges. An example of different types of edges will be discussed in Section 4.

Note that these three types of tasks can be performed in a simultaneous manner. For example, while investigating temporal changes at the global level (*Task 1*), analysts may find interesting patterns at the subgroup level (*Tasks 2a and 2b*). While examining the subgroups (*Tasks 2a and 2b*), the analysts may want to examine whether the subgroups were formed based on any particular node attributes (*Task 3a*) or edge attributes (*Task 3b*).

2.2. Node-Link Based Representations

A number of network visualization tools, such as Pajek (Batagelj & Mrvar, 2005) and KrackPlot (Krackhardt et al., 1994), have been developed and used in analyzing social network data (Freeman, 2000). Most of them employ traditional node-link representations except for a few that use matrix representations (e.g., Abello & van Ham, 2004; Henry et al., 2007b) or a combination of node-link and matrix representations (e.g., Henry & Fekete, 2007; Henry et al., 2007a). In 2007, Henry et al. surveyed two websites (Social Network Analysis Repository[1] and the Visual Complexity Website[2]) that archive network visualization techniques and found that 103 out of 107 are node-link-based representations (Henry et al., 2007a). The advantages of node-link representation have been found in its representation being more intuitively understood and therefore better supporting user tasks such as clustering and path finding (Lee et al., 2006). An example of node-link representation is shown in Fig. 1.

[Fig. 1 will be placed here.]

There are three major approaches adopted to support temporal investigation in a node-link based representation. The most common approach is by showing successive snapshots or by us-

1 http://www.insna.org/software
2 http://www.visualcomplexity.com

ing animation (Moody et al., 2005). By passively seeing or actively interacting with multiple snapshots of a network over time, analysts can observe changes of social network. A variation of the animation approach could be using the idea of small multiples (Tufte, 1990) to show several snapshots of the graph at different points in time. Other techniques (e.g., Brandes & Corman, 2003; Kapler et al., 2008) use three-dimensional graphics, in which two dimensions are used for presenting a node-link diagram and the other dimension is used for the temporal values. For example, Configurable Spaces (Kapler et al., 2008) use the vertical axis as the time dimension and provide users with interaction techniques, so that they can explore different time intervals. These approaches often reveal interesting topological changes over time at the global and subgroup levels (*Tasks 1 and 2*), which provide analysts with insights on how clusters are formulated.

Several aggregation techniques proposed for node-link diagrams are instrumental to understanding the subgroup level of investigation. For example, using metanodes (e.g., Abello et al., 2006; Auber, 2003), multiple nodes are aggregated into metanodes, and the aggregated metanodes could serve as the representation of subgroups or the global network (Elmqvist & Fekete, 2009). Depending on the aggregation of nodes, analysts could create subgroups based on network topology (*Task 2a*) or similarity in node and edge attributes (*Task 2b*).

While the strength of node-link-based approaches lies in the above tasks, there are limitations that stem from the innate nature of node-link diagrams. First, adding temporal dimensions is difficult because screen real estate is consumed by presenting two-dimensional node-link diagrams and second, edges are unsuitable for conveying much data beyond connectivity. Since node-link-diagrams consume two-dimensional screen real estate to present network, it is cumbersome to incorporate the temporal dimension into spatial representation. The animation approach relies on human memory, and the small multiple approach consumes screen real estate. The three-dimensional approach may be helpful in resolving this issue, but a perspective view caused by the three-dimensional approach causes distortion. Due to these limitations, analyzing the changes in various statistical measures of edges (*Tasks 3a and 3b*) solely using node-link-diagrams is challenging. In addition, since edges are often represented with lines, encoding temporal statistics and attributes in these edges (see Table 1) is difficult. They are narrow; their lengths vary; and they overlap with each other. If an analyst wants to compare temporal statistics of edges (*Task 3c*), node-link diagrams are thus not optimal.

2.3. Matrix Representations

Adjacency matrix representation can effectively supplement node-link diagrams, particularly for large and/or dense graphs (Ghoniem et al., 2005). First of all, in an adjacency matrix, cells that represent edges are uniform and square, therefore allowing the embedding of additional information. We were also inspired by Matrix Zoom (Abello & van Ham, 2004) and the Zoomable Adjacency Matrix Explorer (ZAME) (Elmqvist et al., 2008), both of which allow the exploration of a large-scale graph on different zoom levels. The latter, ZAME, is of particular relevance since it uses glyphs to render aggregated information for edge cells. ZAME supports several different types of glyphs (color shades, histograms, bands, etc) which could be used to embed relatively complex information in a limited space, some even for showing changes over time.

However, complex glyphs may not be visible when the whole visualization system is zoomed out to see the overall trends of a large-scale graph. Semantic zooming (Bederson & Hollan, 1994) alleviates this issue since we can use different visual representations depending on the amount of screen real estate allocated to a particular glyph. For example, as the user zooms out from a detailed view to an overview, we can dynamically change the visual representation of each cell from smooth histogram to step histogram, to average, and finally to color shade.

However, at the same time, we should acknowledge that matrix-based techniques have some drawbacks as well. Matrices are generally inefficient for presenting sparse networks since they waste space for isolated or sparsely connected nodes. Some commonly found users tasks are difficult (e.g., path finding) to perform using matrix-based techniques (Lee et al., 2006). However, existing work has already addressed these issues, including techniques such as NodeTrix (Henry et al., 2007a), MatLink (Henry & Fekete, 2007), and path visualization (Shen & Ma, 2007) that all overcome the innate issues of matrix-based visualization techniques.

3. TimeMatrix

We propose TimeMatrix, a matrix-based graph visualization technique designed to support temporal social network analysis. By leveraging the scalability of matrix representations, we are able to easily and efficiently support large and dense social networks. The representation, especially the TimeCells (Section 3.1), lends itself well to displaying temporal data and statistical information on edges and nodes on the surface of the glyphs representing nodes and edges in the graph (*Tasks 3a-c*). Corresponding interaction techniques, such as semantic zooming (Section 3.2), aggregation (Section 3.3), and node reordering (Section 3.6), allow for investigating a network at multiple granularity levels and layouts (*Tasks 2a-b*). Finally, overlays (Section 3.4) and filters (Section 3.5) allow for comparing temporal data and statistical information.

[Fig. 2 will be placed here.]

3.1. TimeCell

The primary visual component in the TimeMatrix technique is the TimeCell, a visual aggregate that displays temporal information associated with a node or edge as a composite glyph (similar to glyphs in ZAME (Elmqvist et al., 2008)). The temporal information spans the whole time period for the graph, and may represent one node or edge, or a whole aggregated hierarchy.

The visual representation for a TimeCell depends on the data to visualize. Fig. 2 (a) shows an example of a TimeCell bar chart glyph. In this case, the horizontal (X) axis represents the time dimension, and the vertical (Y) axis represents any vertex- or edge-based attributes (e.g., the number of edges during the time frame). This helps present individual level temporal statistics of edges on a single cell (*Task 3b*). Since the size of each cell tends to be very small, only simple glyph visualizations can be used, such as two-tone pseudo-coloring (Saito et al., 2005), stacked graphs (Byron & Wattenberg, 2008), or 2-band horizon charts (Few, 2008).

TimeMatrix incorporates TimeCell glyphs for not only edges, but also for nodes on both the rows and column headers of the matrix representation as shown in Fig. 2 (b). By juxtaposing temporal information for both nodes and edges in this way, we aim to help users see the temporal associations between multiple attributes in a single, integrated view (*Task 3a*). This also can be helpful in visualizing evolving graphs whose nodes are appearing and disappearing over time.

3.2. Semantic Zooming

TimeCells are typically only allocated a few pixels of space on a screen in a zoomed-in view. When the screen allocation becomes too small, simple bar charts cannot be drawn on a TimeCell. We use semantic zooming to overcome this issue. As shown in Fig. 3 (a), when less than 100 pixels (10 pixels high and 10 pixels wide) are allotted for a TimeCell, bar charts in a TimeCell are replaced with a color shade glyph. The color is determined by the overall values of a TimeCell, and it will not show any temporal changes. However, combined with a range slider—discussed in the following section—a user can see temporal changes over time by identifying the changes in the shades of the glyphs. When less than a pixel is allotted for a TimeCell, variation of encoded color is difficult to see. Thus, only binary status of a TimeCell will be presented. In other words, a TimeCell will show an assigned color only when there is any edge between two vertices. While this representation does not show any temporal data, it provides a high-level overview of the global (*Task 1*) and subgroup (*Task 2*) network structure for the TimeMatrix.

[Fig. 3 will be placed here.]

3.3. Aggregation

The TimeMatrix matrix is a *hierarchically aggregated* (Elmqvist & Fekete, 2009) visual structure where the representation of nodes and edges (called *visual aggregates*) can show more than one node or edge in the underlying graph dataset. To facilitate this, the visual node and edge aggregates show summaries of the data as TimeCells as shown in Fig. 3 (b) (discussed above).

Hierarchical aggregation is useful for providing an overview of a dataset, even for a very large-scale dataset. In addition, aggregation is useful when researchers are interested in combining several semantically grouped nodes or edges into a single entity. This entity represents the whole group and allows examination of its attributes at the aggregate level. Because the aggregation is hierarchical, the user can iteratively group together nodes and edges to form a hierarchy. In other words, an aggregation can not only contain other nodes but also other aggregations, so a user may first form the most obvious aggregation and gradually make larger aggregates by combining smaller aggregates. The user can also split an aggregate into its constituent parts. This feature supports *Tasks 2a and 2b* even though the topological structure shown in a matrix view may be less obvious than that shown in a node-link diagram.

3.4. Overlays

To cope with the large amount of labels, statistics, and visual representations that can be associated with each TimeCell for both nodes and edges, TimeMatrix uses an *overlay system* as shown in Fig. 3 (c). This approach is similar to plastic transparencies for overhead projectors, where each of the above metrics is drawn on its own transparency (*overlay*), and a control mechanism is used to manage the order, color, translucency, and visibility of the transparencies on the virtual overhead projector (the actual visual display). The rationale for this mechanism is to provide a highly configurable display where the user can adapt the visual representation for a particular task and reduce the amount of visual clutter that is currently not relevant.

Another use for the overlay system is to visualize different types of edges. For example, when a pair of nodes is involved in two types of network activities (e.g., emails and phone calls in a friendship network) at different time points, the two distinct histograms along with the time range indicated in the x-axis can show the changes in patterns over time. Since different overlays can have different ranges in the X- and Y-axis, the ranges used by TimeCells are designed to include the total ranges of different overlays. For example, if the ranges of two overlays are 1990 – 1996 and 1993 – 2000, all TimeCells will use the range of 1990 – 2000.

The overlay feature is particularly useful when comparing different measures of nodes or edges. For example, if the analyst wants to understand the relationship between degree centrality and betweenness centrality, the analyst can add two layers on top of the nodes to show the temporal association between the two measures. Therefore, this feature facilitates *Task 3c*.

Layers are managed using a control box that builds on the familiar layer concept of image manipulation applications, such as Adobe® PhotoShop® and GIMP. This box provides a list of overlays, where each overlay can be toggled and its transparency and color can be changed.

3.5. Filtering

A range slider (Fekete, 2004) as shown in Fig. 3(d) is used to filter out specific nodes or edges from the main view. Since both nodes and edges may have multiple attributes, a range slider can be associated with these various node or edge attributes. When a range slider is used with an edge attribute, edges that have values falling outside of the boundary of the range slider will be greyed out (when TimeCell shows bar charts) or removed (when TimeCell shows color shades or pixels). When a range slider is used with a node attribute, all of the edges associated with the filtered nodes will be greyed out or disappear. However, filtered nodes are not hidden in order for the layout of the matrix visualization to remain constant.

As the filtering range is changed, the visualization is updated. The filtering range can be changed not only by moving one of the slider thumbs, but also by moving the slider button itself. This is particularly useful when analyzing global changes in a certain time interval (*Task 1*).

3.6. Node Reordering

To help understand the relationships between attributes of nodes and edges, nodes in the TimeMatrix can be reordered, for example using sorting. This feature is similar to the block modeling techniques that have been developed in previous matrix-based representations (Batagelj et al., 1999). For example, when nodes in TimeMatrix are sorted by gender, TimeCells can be clearly clustered into four categories, male-to-male, male-to-female, female-to-male, and female-to-female relationships (assuming that the underlying graph is directed). By investigating each cluster, temporal changes of relationships depending on gender similarities or differences could be easily identified (*Task 2b*). A stable sorting algorithm is used to preserve any existing order.

Besides sorting, it is possible to utilize external statistical tools to generate proper clusters for rearranging the matrix. To support this use case, TimeMatrix can import a node permutation through a file and reorder its nodes accordingly.

3.7. User Interface

The TimeMatrix prototype (Fig. 5 (a)) consists of the main view (on the left side) and a control menu (on the right side). Most of the visual exploration is performed in the main view. As is common in other matrix-based visualization techniques, the vertices are shown on both the left and top borders of the main view. Edges are shown as cells where the vertices intersect. We also use the space on the top left corner for overview, and the currently shown area in the main view will be indicated in the overview. A user can aggregate edges by lassoing while pressing the middle mouse button, and split an aggregated edge by clicking the middle mouse button.

The control menu (right) has three panels: the overlay panel, the sorting and aggregation panel, and the temporal filtering panel. Using the overlay panel, a user can toggle different overlays as well as change their color and transparency. The sorting and aggregation panel allows a user to sort or aggregate nodes based on a particular attribute of nodes, selected from a dropdown list. The temporal filtering panel also has a dropdown list to select the attribute that the range slider is based on. However, the range slider is mostly used with a time variable.

4. User Study

We recruited three researchers (two graduate students and one faculty member; two female and one male) who had experience in social network analysis using node-link-based visualization tools. Each participated in an approximately two-hour-long semi-structured interview. One experimenter introduced the features of TimeMatrix to participants in a step-by-step manner. After a set of relevant feature was demonstrated, each participant was asked to conduct a set of social network analysis tasks. The participant responses (e.g., feedback, questions, and suggestions) during the interview were transcribed and analyzed using TAMSAnalyzer[3], a text analysis markup system. Henceforth, the three participants will be labelled as P1, P2, and P3.

The dataset is based on the records of inter-organizational collaboration activities, which focus on initiatives to use Information and Communication Technologies (ICTs) for various development goals in developing countries. Inter-organizational networks were categorized into two

[3] http://tamsys.sourceforge.net/

types of organizational collaboration: first, joint implementation of development projects and second, knowledge-sharing through participation in global forums and conferences. The data represents a total of 730 unique organizations that participated in project implementation or knowledge sharing between 1987 and 2008. Each node also has three different attributes: region (i.e., Asia, Africa, Europe, Latin America and the Caribbean, North America, Oceania), organization type (i.e., governmental, intergovernmental, nongovernmental, and private/for-profit), and geographic scope (i.e., international, regional, and national).

Throughout the interview, participants were asked to achieve two goals: 1) to examine the changing patterns of collaboration of the two types (development projects and knowledge sharing) over time and 2) to investigate the role of organizations of different types, regions, and geographic scopes in the collaboration activities. To help participants achieve these overall goals, the interviewer demonstrated the features of TimeMatrix and asked the participant to perform a set of directed tasks. The tasks were derived from the three levels as proposed in Section 3: the nodal/dyad (Task 3), subgroup (Task 2), and global levels (Task 1):

First, to promote nodal/dyad level analysis, participants were asked to find central actors, any interesting patterns in individual TimeCells and row/column headers, and temporal changes. Obviously, all participants extensively used zooming and panning to explore different parts of TimeMatrix in order to become familiar with the tool. P2 and P3 successfully identified central actors—World Bank, International Development Research Centre (IDRC), and United States Agency for International Development (USAID)—using degree centrality shown on row/column headers (see Fig. 4a). P2 especially appreciated statistical measure overlaid on nodes and said "Showing node attributes is the best feature, [it] gets you into the analysis of the data." However, P1 did not seem to understand the feature and failed to identify main actors. In addition, participants noticed two interesting patterns from temporal trends in individual TimeCells: 1) every collaboration project started with development project followed by knowledge sharing (P1-3); and 2) a number of collaborative projects peaked in the 1990s and dropped off in the later years (see Fig. 4b) (P2-3). The first pattern surprised participants since it is somewhat counter-intuitive. P3 said, “If asked before [this session], I would have seen knowledge sharing happening first and then formalizing into projects in organizational ties, but the surprising thing is that the reverse is happening.” Interestingly, participants quickly noticed this pattern via the overlay tool that represents the temporal patterns of two types of ties (e.g., blue for joint implementation; red for knowledge-sharing) on a single chart.

[Fig. 4 will be placed about here.]

Second, to promote subgroup level analysis, participants were asked to find dense subgroups and to identify patterns from aggregated TimeCells (see Section 3.3). Participants appreciated that TimeMatrix allows them to aggregate nodes by their attributes (i.e., regions, types, and geo-scope) because this feature helped them see the overall trends of different groups (P2-3) and compare “collaboration *within* groups” with “collaboration *between* groups” (P1) (see Fig. 5a and Fig. 6). Participants reported several insights while they investigated the data set. All three identified that organizations in different regions and types have different collaboration patterns (e.g., “Relationships in Latin America are developing over time, but relationships in Europe have been pretty consistent” and “Geographically bound organizations are more likely to work with each other.”). P2 also reported interesting temporal patterns using the aggregation feature by saying “Similar to what I was thinking, it all exploded in the mid 90s. […, However,] There are exceptions. In certain cases, it starts with knowledge sharing, or sometimes there is no knowledge sharing at all. In Africa and Latin America, it starts with knowledge.” However, participants were uncomfortable with identifying dense subgroups probably due to their unfamiliarity with TimeMatrix. P3 said, "Can I see density? Not sure how to see this. [I am] used to looking at this from a graph standpoint. I actually don't know." P2 also pointed out the ordering of the nodes is very crucial for understanding the overall structure and clusters. Even though TimeMatrix allows users to reorder the nodes according to attributes associated with nodes, P2 also pointed out that the order of nodes within each block matters in understanding the trends because it is very difficult to identify cliques if they are spatially distant.

[Fig. 5 will be placed about here.]

[Fig. 6 will be placed about here.]

Third, to promote global level analysis, participants were asked to identify changes of network structure and collaboration patterns over time. P3 actively zoomed out the view to see all of the nodes on a single screen and used the range slider to see patterns in creation and deletion of ties. P3 commented, "I could already demonstrate how to use [TimeMatrix] to someone else, definitely the time slider [meaning the range slider] and the navigation. The time slider is beautifully intuitive." However, all participants had a difficult time seeing global network structure changes, such as the change of centrality: "None seen (P1)," "Hard to say. Depends on what we are looking at (P2)," and "Clustering and centrality is more evident in the node-link diagrams (P3)." We expected that the cross patterns, block patterns, and intermediate patterns would help participants obtain initial insights about the extent of centralization and the existence of subgroups in the network, but we found that these patterns are not quite intuitive, at least for the three participants recruited in this particular study.

After accomplishing these tasks, each participant was asked to estimate the accuracy of his or her analyses. The responses were "About 80% confidence in answers (P1)," "About 60-70%

confidence (P2)," and "Absolute confidence in the tool (P3)," which indicates that participants were able to rely on this tool. P3 further demonstrated interest in the tool by asking when they could get a copy of the tool. In order to explicitly compare TimeMatrix to node-link-based visualization, the interviewer showed them the node-link-diagrams of the same data set (see Fig. 1) and asked participants to contrast TimeMatrix to node-link diagrams. They reported that node-link diagrams were much more intuitive and easier for understanding density (P1, P3), centrality (P1-3), clustering (P3), growth (P3), and linkage (P3) of social networks.

Suggestions for the future development of TimeMatrix were collected as well. The most common suggestion was to resolve usability issues, such as showing labels properly (P2, P3), increasing the performance of the system (P1), providing a separate overview window (P1), and adding grids between TimeCells to make distinctions between cells more visible (P3). Participants also requested integration with other network analysis tools (e.g., UCINET) (P1), node-link-based visualization tools (P1, P3), and a text-based query system (P2).

5. Discussion

Our user study showed that the TimeMatrix could be useful for understanding a temporally changing social network. Although the matrix-based representation may be less intuitive than node-link representations due to the presentation of the topology, we found that many analytic tasks, especially in the nodal, dyad, and sub-group levels, were well-supported by TimeMatrix.

In addition, we found that temporal alignment is important in analyzing temporal social networks. Since time is a universal measure, any statistical measures with temporal aspects (e.g., edge counts between pairs of nodes and degree centrality measures of nodes) should be temporally aligned, so that a user can make sense of the sequences of activities and implications. This appears to be especially important when different overlays are merged together on a single

TimeCell. In other words, when different overlays are drawn on a single TimeCell, they should use the same time span so that two different temporal statistics or measures are comparable (see Fig. 5). In addition, among the different TimeCells, the temporal range slider was helpful to support temporal alignment. The initial purpose of the range slider was to be merely a filter mechanism. However, it quickly became obvious that the range slider is a very convenient tool for comparing different TimeCells that are spatially apart. As demonstrated in a case study (see Fig. 6), a user can compare values in different TimeCells without cluttering screen with labels or hovering mouse cursors to see tooltips of different elements. Here, the usage of the range slider is very similar to the water level feature of SeeIT[4].

However, challenges still exist. One of the biggest challenges in our approach is showing a large-scale temporal social network. The amount of information that is necessary for social network analysis is typically large, and therefore, presenting all the data is difficult on limited screen estate. We initially speculated that semantic zooming might resolve this issue but found that the information loss (especially temporal information) is too large for users to be able to find interesting patterns. The hierarchical aggregation played a major role in overcoming this limitation. By sacrificing the information regarding topological structure, users can see temporal information even in the fully zoomed-out view, which turned out to be difficult to be achieved in other methods (e.g., analysis with node-link-based visualization tool or statistical methods). Interestingly, the user can investigate the detailed topological structure after making sense of the overview, as demonstrated in the user study. By manually changing aggregation levels, users can also see the detailed topological structure when it is needed.

[4] *http://www.advizorsolutions.com/*

In summary, various representation and interaction techniques introduced in this paper support two high level tasks. First, through semantic zooming and aggregation, users can make smooth trade-offs between overview and details. By sacrificing temporal information (via semantic zooming) and topological structure (via aggregating), users can see an overview of even a complex and large network structure. Users also can choose the proper level of abstraction by manually choosing the boundary of aggregation, and the hierarchical structure of aggregation also supports users to gradually change the level of abstraction. Second, through overlays and range sliders, users can align different types of temporal information, so that identifying potential causal relationships between various statistics and attributes becomes more efficient.

6. Conclusions

We have presented TimeMatrix, a matrix-based visualization technique designed for temporal social network analysis that supports several tasks that are often difficult to perform in node-link-based visualization techniques. More specifically, the contributions of this paper are as follows:

- Categorizing visual analytic tasks in temporal social network analysis (*Tasks 1, 2, and 3*);
- Proposing an adjacency-matrix-based visual representation (TimeMatrix) for visualizing temporal graphs which complement node-link temporal graph visualization techniques; and
- Supplementing TimeMatrix with interaction techniques supporting highly interactive visual exploration of real-world social networks across multiple levels of analysis.

To validate our design, we have presented a small user study with three subject-matter expert users involving a temporal social network consisting of more than 730 nodes and more than 22,105 individual edges. The user study showed that TimeMatrix supports many of the analysis tasks that we categorized in Section 2.1. However, it also reveals some of limitations of TimeMatrix that we would like to resolve in the future (see below).

7. Future Work

The current study is a mere first step toward an effective dynamic network analysis tool, and its limitations should be clearly noted. First, the design of TimeMatrix is not based on an exhaustive review of potential visualization methods. Instead, TimeMatrix was designed as a backlash against the dominance of node-link-based visualizations in spite of their limitations in presenting temporal aspects of social network. In order to derive a better design, an exhaustive review of the design space should be conducted in the future. More specifically, the design of the TimeMatrix technique uncovered a range of interesting visualization concepts to be further explored, including hierarchical aggregation, glyph-based visual aggregates, and overlay mechanisms for managing visual clutter. Ultimately, combining node-link diagrams and TimeMatrix may be a better compromise, as shown by prior studies (e.g., Henry & Fekete, 2007; Henry et al., 2007a) as well as suggested by three participants.

Second, the design of TimeMatrix is not based on comprehensive understanding of underlying cognitive aspects of social network analysts. Through interaction with social network scientists and analysts, we learned that their cognitive activities while investigating temporal aspects of network are quite complex: They collect various information from multiple sources, such as temporal changes of network structure, causality between different temporal information, network statistics, and their prior knowledge. Thus, more in-depth cognitive task analysis or contextual inquiry should be conducted to better understand their cognitive activities. Detailed knowledge of cognitive activities will be essential for evaluating different design alternatives.

Third, in order to prove the effectiveness of TimeMatrix and understand how social network analysts utilize its different operations, more comprehensive and large-scale studies are needed. However, since the ultimate success of TimeMatrix is only measured by how effectively social network analysts gain insights from the tool, long-term and multifaceted evaluation techniques would be necessary, such as multi-dimensional in-depth long-term case studies (MILCs) (Shneiderman & Plaisant, 2006) and insight-based evaluation (Saraiya et al., 2004).

8. References

Abello, J., & van Ham, F. (2004). *Matrix Zoom: A Visual Interface to Semi-External Graphs.* Paper presented at the IEEE Symposium on Information Visualization, 183–190.

Abello, J., van Ham, F., & Krishnan, N. (2006). ASK-Graph View: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics, 12*(5), 669-676.

Auber, D. (2003). Tulip-a huge graph visualization framework. In M. Jünger & P. Mutzel (Eds.), *Graph Drawing Software, Mathematics and Visualization* (pp. 105–126): Springer-Verlag.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). *Group formation in large social networks: membership, growth, and evolution.* Paper presented at the The 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 44-54.

Barabasi, A. L., & Crandall, R. E. (2003). Linked: the new science of networks. *American Journal of Physics, 71*, 409-410.

Batagelj, V., & Mrvar, A. (2005). PAJEK: Program for analysis and visualization of large networks. *Ljubljana, Slovenia: University of Ljubljana*.

Batagelj, V., Mrvar, A., & Zaveršnik. (1999). *Partitioning approach to visualization of large graphs.* Paper presented at the Graph Drawing '99, 90-97.

Bederson, B. B., & Hollan, J. D. (1994, Nov 2-4 1994). *Pad++: A zooming graphical interface for exploring alternate interface physics.* Paper presented at the The ACM Symposium on User Interface Software and Technology, Marina del Rey, CA, USA, 17-26.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science, 323*(5916), 892-895.

Brandes, U., & Corman, S. R. (2003). Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization, 2*(1), 40-50.

Brass, D. J., Galaskiewicz, J., Greve, H. R., & Tsai, W. (2004). Taking stock of networks and organizations: A multilevel perspective. *Academy of Management Journal, 47*(6), 795-817.

Burt, R. S. (2000). Decay functions. *Social Networks, 22*(1), 1-28.

Burt, R. S. (2002). Bridge decay. *Social Networks, 24*(4), 333-363.

Byron, L., & Wattenberg, M. (2008). Stacked Graphs–Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics, 14*(6), 1245-1252.

Coleman, J. S. (1990). *Foundations of Social Theory*: Harvard University Press.

Contractor, N. S., Wasserman, S., & Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review, 31*(3), 681-681.

Elmqvist, N., Do, T. N., Goodell, H., Henry, N., & Fekete, J. D. (2008). *ZAME: Interactive Large-Scale Graph Visualization.* Paper presented at the IEEE Pacific Visualization Symposium, 215-222.

Elmqvist, N., & Fekete, J.-D. (2009). Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics*.

Fekete, J.-D. (2004). *The infovis toolkit.* Paper presented at the IEEE Symposium on Information Visualization, Austin, Texas, 167-174.

Few, S. (2008, Jun/Jul). Time on the Horizon. *Visual Business Intelligence Newsletter*.

Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure, 1*(1).

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

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology, 91*(3), 481-510.

Gulati, R., & Gargiulo, M. (1999). Where Do Interorganizational Networks Come From? 1. *American Journal of Sociology, 104*(5), 1398-1438.

Henry, N., & Fekete, J.-D. (2007). *MatLink: Enhanced Matrix Visualization for Analyzing Social Networks.* Paper presented at the INTERACT 2007, 288-302.

Henry, N., Fekete, J.-D., & McGuffin, M. J. (2007a). NodeTrix: A Hybrid Visualization of Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 1302-1309.

Henry, N., Goodell, H., Elmqvist, N., & Fekete, J.-D. (2007b). 20 Years of Four HCI Conferences: A Visual Exploration. *International Journal of Human-Computer Interaction, 23*(3), 239-285.

Kapler, T., Eccles, R., Harper, R., & Wright, W. (2008). *Configurable Spaces: Temporal Analysis in Diagrammatic Contexts.* Paper presented at the IEEE Symposium on Visual Analytics Scinece and Technology, Columbus, Ohio USA, 43-50.

Kontopoulos, K. M. (1993). *The Logics of Social Structure*: Cambridge University Press.

Krackhardt, D., Blythe, J., & McGrath, C. (1994). KrackPlot 3.0: An improved network drawing program. *Connections*, *17*(2), 53-55.

Lee, B., Plaisant, C., Parr, C. S., Fekete, J.-D., & Henry, N. (2006). *Task taxonomy for graph visualization.* Paper presented at the the AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, Venice, Italy, 1-5.

McGrath, C., Krackhardt, D., & Blythe, J. (2003). Visualizing Complexity in Networks: Seeing Both the Forest and the Trees. *Connections*, *25*(1), 37-47.

McKelvey, B. (1997). Quasi-natural Organization Science. *Organization Science, 8,* 351-380.

Mcpherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27,* 415-444.

Monge, P. R., & Contractor, N. S. (2003). *Theories of Communication Networks*: Oxford University Press, USA.

Moody, J., McFarland, D., & Bender-deMoll, S. (2005). Dynamic Network Visualization. *American Journal of Sociology, 110*(4), 1206-1241.

Palla, G., Barabasi, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*(7136), 664 - 667.

Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences 1. *American Journal of Sociology*, *110*(4), 1132-1205.

Provan, K. G., Fish, A., & Sydow, J. (2007). Interorganizational Networks at the Network Level: A Review of the Empirical Literature on Whole Networks. *Journal of Management, 33*(3), 479-516.

Robins, G., Pattison, P., & Woolcock, J. (2005). Small and Other Worlds: Global Network Structures from Local Processes 1. *American Journal of Sociology, 110*(4), 894-936.

Saito, T., Miyamura, H. N., Yamamoto, M., Saito, H., Hoshiya, Y., & Kaseda, T. (2005). *Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data.* Paper presented at the the 2005 IEEE Symposium on Information Visualization, 173-180.

Saraiya, P., North, C., & Duca, K. (2004, October 10-12, 2004). *Evaluation of microarray visualization tools for biological insight.* Paper presented at the The IEEE Symposium on Information Visualization, Austin, Texas, USA, 1-8.

Scott, J. (2000). *Social Network Analysis: A Handbook*: Sage.

Shen, Z., & Ma, K.-L. (2007). *Path Visualization for Adjacency Matrices*. Paper presented at the Eurographics/ IEEE-VGTC Symposium on Visualization.

Shneiderman, B., & Plaisant, C. (2006). *Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies.* Paper presented at the Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, Venice, Italy, 1-7.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological methodology, 31*(1), 361-395.

Tufte, E. R. (1990). *Envisioning information*: Graphics Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*: Cambridge University Press.

Table 1. Measures of interests in social network analysis (Monge & Contractor, 2003; Provan et al., 2007; Wasserman & Faust, 1994).

Types	Measures	Temporal	Examples
Nodes	Centrality	yes	Degree centrality; closeness centrality; betweenness centrality; Eigenvector centrality
	Attributes	yes or no	Demographic characteristics of actors such as gender and ethnicity (static); behavioral characteristics of actors such as smoking and number of friends (temporal)
	Positions and roles	yes	Star; liaison; brokerage; gatekeeper; isolate
	Life cycle	yes	Emergence; disappearance
Edges	Dyad-level	yes	Direction; reciprocity; frequency; strength; multiplexity; stability
	Subgroup-level	yes	Size; density; component; clique; clustering
	Global-level	yes	Centralization; core-periphery; reachability; small-world network
	Life cycle	yes	Emergence; growth; dissolution

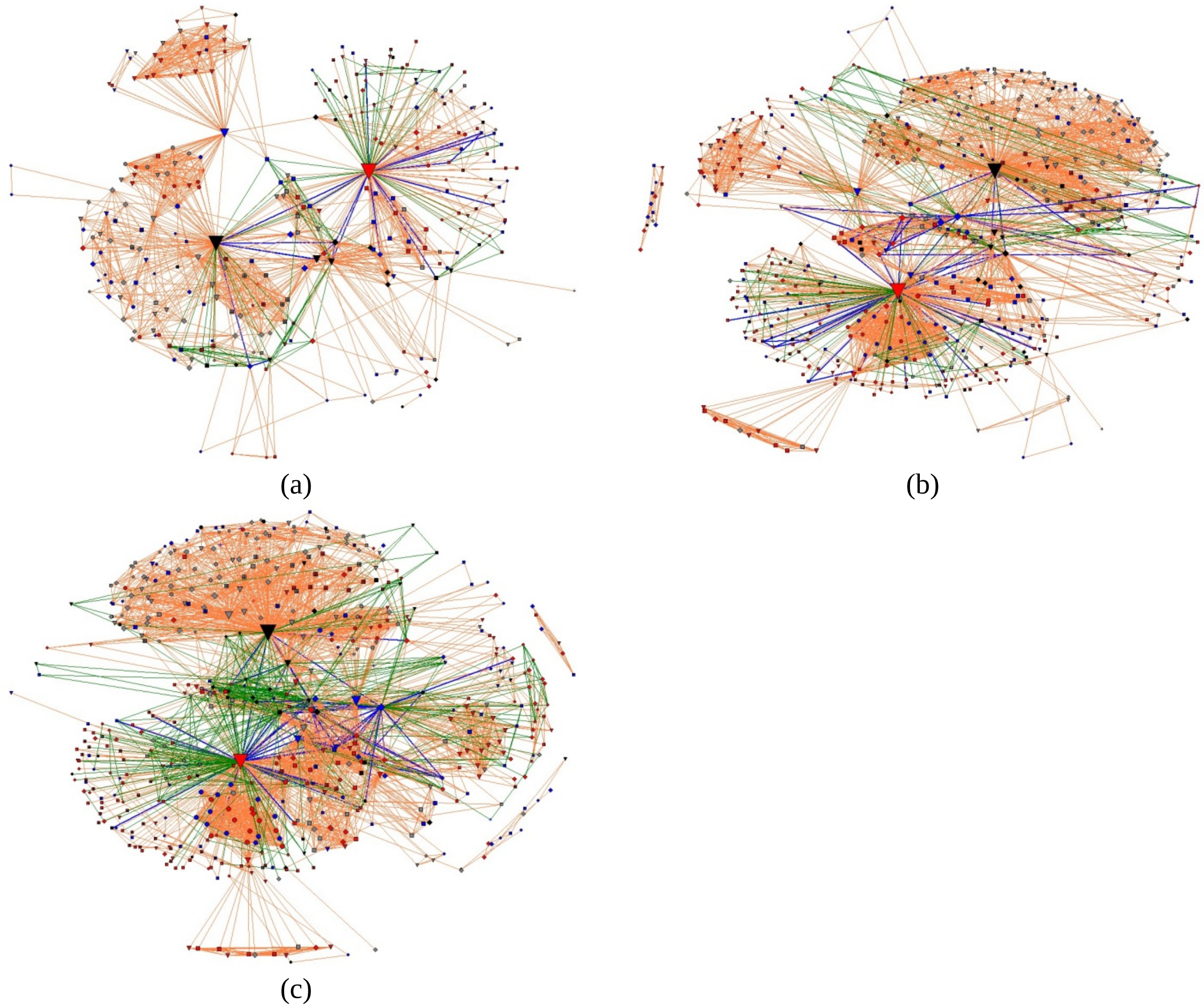
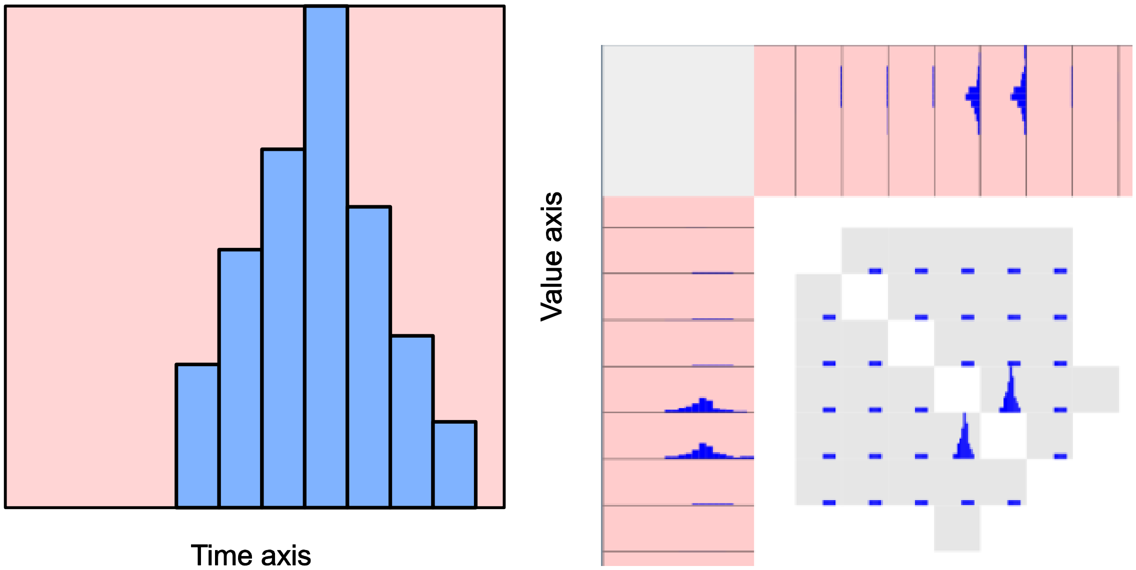


Fig. 1. A node-link representation of inter-organizational collaboration activities: (a) 1997-1999; (b) 2000-2002; and (c) 2003-2005.



(a) (b)

Fig. 2 (a) TimeCell visual representation showing edge count for a pair of nodes over a period of time, and (b) a screen shot of TimeMatrix using TimeCells for both edges and vertices.

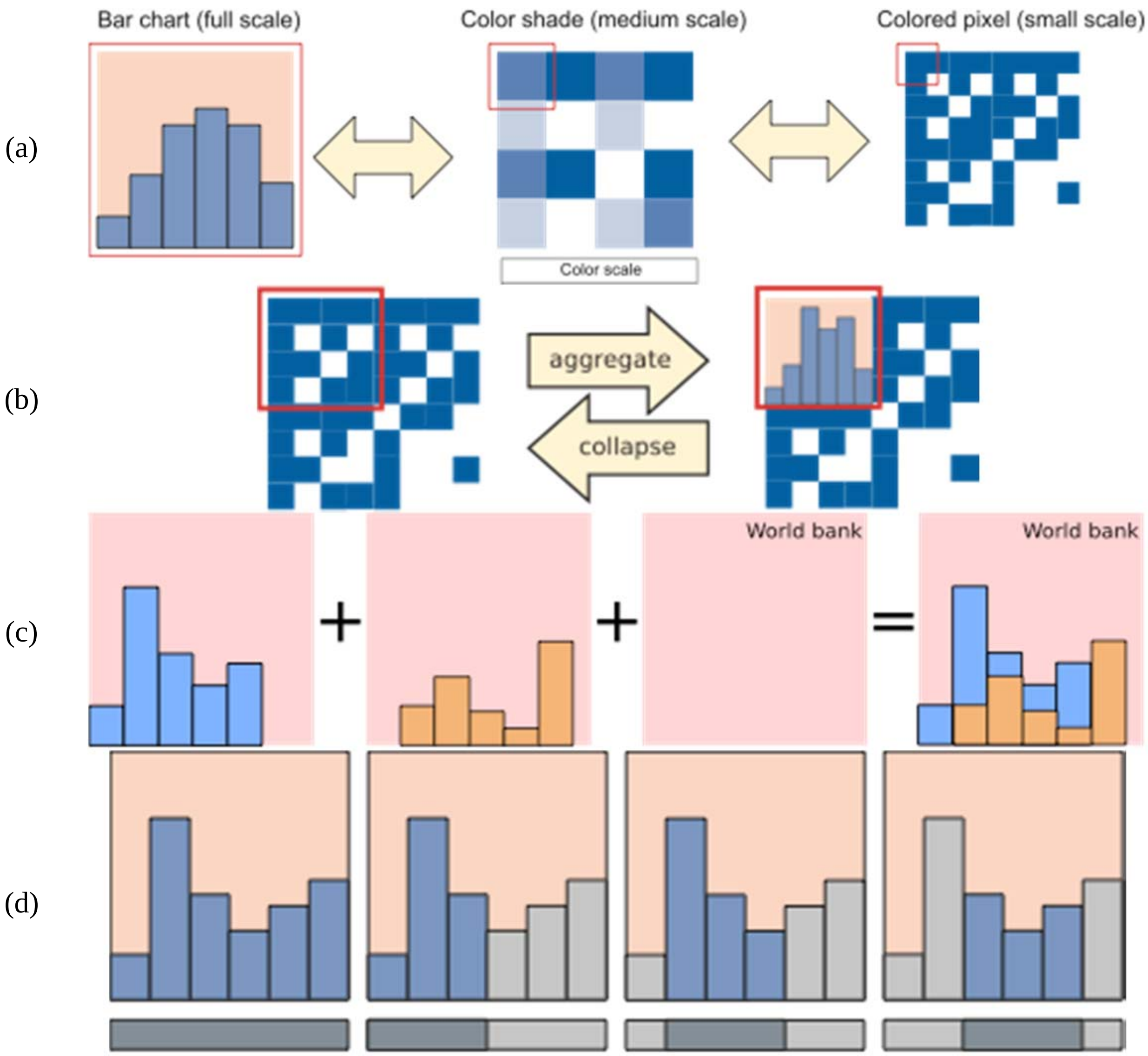


Fig. 3 Interaction Techniques for TimeMatrix: (a) Semantic zooming; (b) Aggregation and collapsing; (c) Overlays; and (d) Range Slider.

(a)

(b)

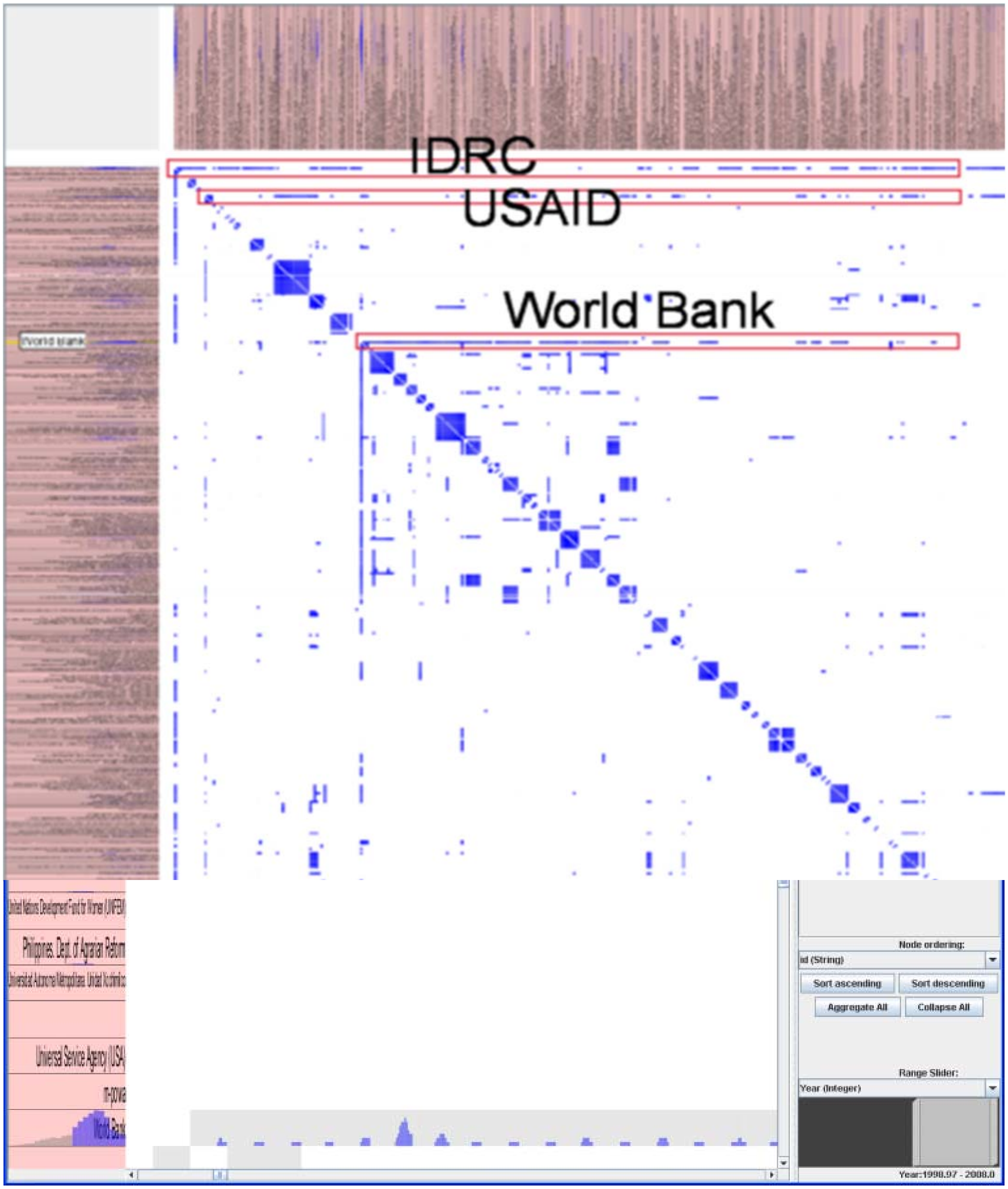


Fig. 4. Central actors in the of inter-organizational collaboration network in (a) zoomed-out view and (b) zoomed-in view.

(a)

(b)

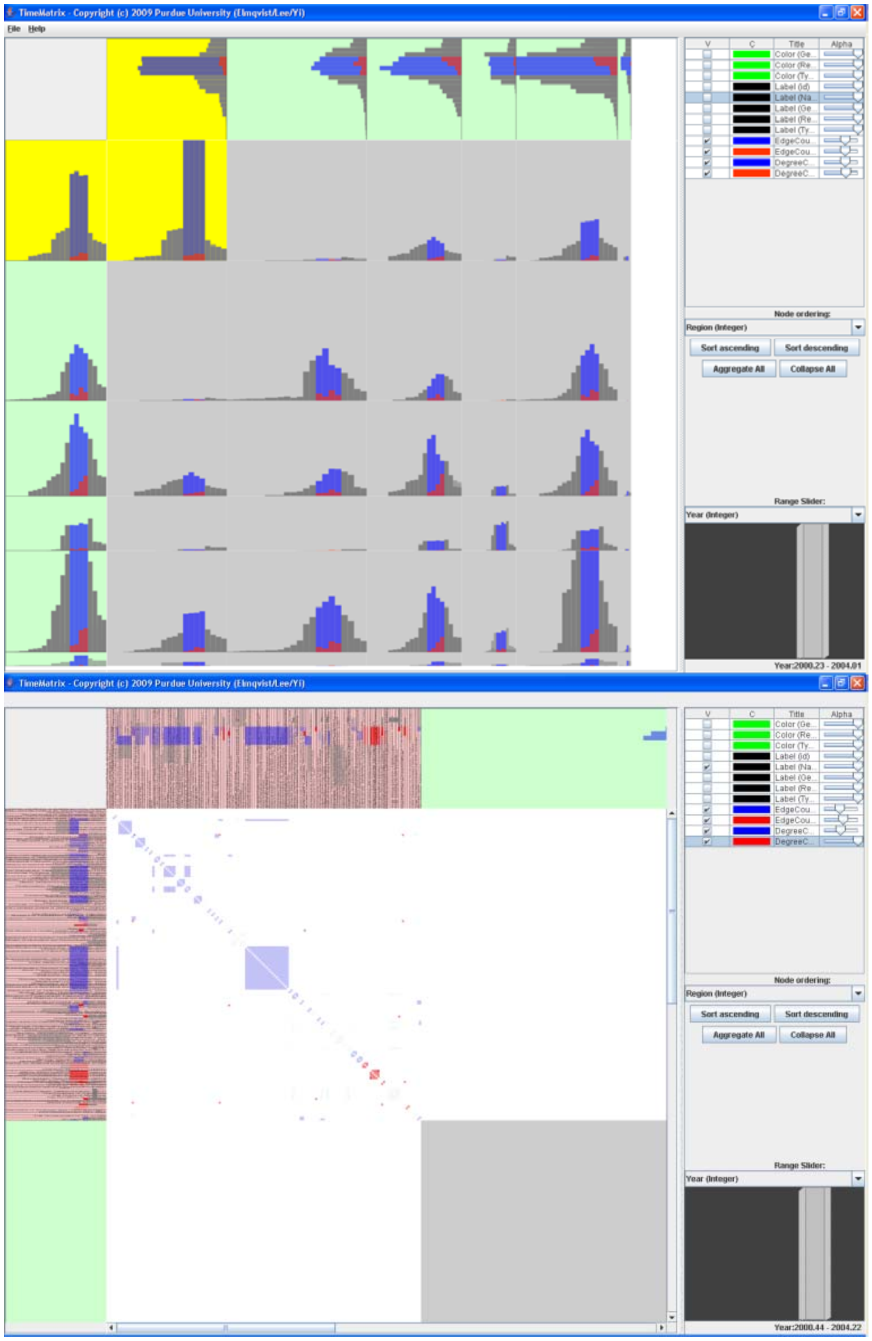


Fig. 5. (a) Aggregated TimeCells by regions (the aggregated nodes represent Asia, Africa, Europe, Latin America and the Caribbean, North America, and Oceania from the left) and (b) collapsed TimeCells representing the Asia region.

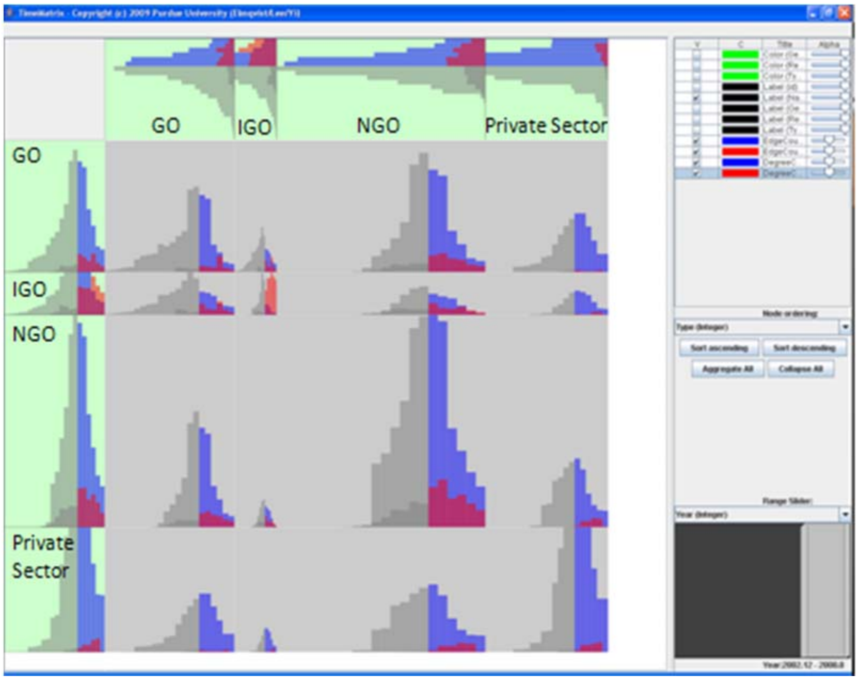


Fig. 6. Aggregated TimeCells by Organization Type.