

LargeNetVis: Visual Exploration of Large Temporal Networks Based on Community Taxonomies

Claudio D. G. Linhares*, Jean R. Ponciano*, Diogenes S. Pedro, Luis E. C. Rocha, Agma J. M. Traina, and Jorge Poco

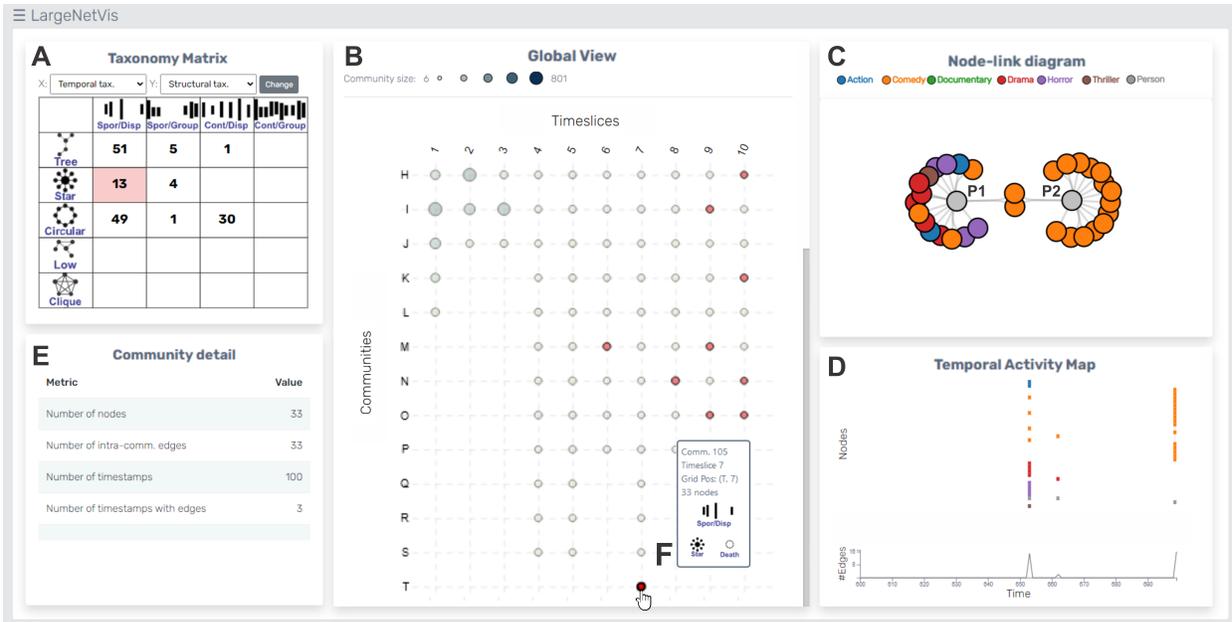


Fig. 1. LargeNetVis is a web-based visual analytics system designed to support the visual exploration of temporal networks with up to a few thousand nodes and timestamps. The system is composed of four linked visual components. While the *Taxonomy Matrix* (A) and *Global View* (B) enable global-level analysis, the *node-link diagram* (C) and the *Temporal Activity Map* (D) enable structural and temporal local-level analysis, respectively. The system also provides a panel with the network, community, or node numerical information (*Community detail* (E)). A tooltip with extra information is shown when a community is hovered over in *Global View* (F).

Abstract— Temporal (or time-evolving) networks are commonly used to model complex systems and the evolution of their components throughout time. Although these networks can be analyzed by different means, visual analytics stands out as an effective way for a pre-analysis before doing quantitative/statistical analyses to identify patterns, anomalies, and other behaviors in the data, thus leading to new insights and better decision-making. However, the large number of nodes, edges, and/or timestamps in many real-world networks may lead to polluted layouts that make the analysis inefficient or even infeasible. In this paper, we propose LargeNetVis, a web-based visual analytics system designed to assist in analyzing small and large temporal networks. It successfully achieves this goal by leveraging three taxonomies focused on network communities to guide the visual exploration process. The system is composed of four interactive visual components: the first (*Taxonomy Matrix*) presents a summary of the network characteristics, the second (*Global View*) gives an overview of the network evolution, the third (a *node-link diagram*) enables community- and node-level structural analysis, and the fourth (a *Temporal Activity Map – TAM*) shows the community- and node-level activity under a temporal perspective.

Index Terms—Information Visualization, Interactive Visualizations, Human-Computer Interaction, Electronic Health Records.

1 INTRODUCTION

Graphs are mathematical objects used to map the relation (edges) of different elements (nodes). Complex networks use graphs to model real-world events by studying the structure and dynamics of complex systems [24]. Moreover, when the dynamics of these networks are mapped through the time information, they are called temporal (or time-evolving) networks [38]. Information visualization techniques help in

analyzing temporal networks as they allow exploratory analysis and also enhance the identification of structural and temporal patterns that involve individuals and groups of elements. Visualization layouts have been proposed in the last decades for these purposes, from timeline-based approaches and animations to hybrid visualizations [12].

In our context, visual scalability refers to the ability of a layout to work properly and adapt to small and large networks. This capability is particularly important considering the number of real-world networks regularly created with the acquisition of big data from different application domains, including online social media, biological, and others. A crucial definition is, therefore, *what a large network is*. Such definition depends on the type of study and the number of elements in the network. Non-temporal networks with millions of nodes and edges are commonly analyzed in statistical studies, and these networks are often referred to as large networks [70]. For visualization purposes, on the other hand, this definition may vary according to the task and aspects such as edge density and visual clutter. To exemplify how abstract this definition is in this context, while Mi et al. [46] consider large (non-temporal) networks as those with millions of nodes, a more recent

* The first two authors contributed equally to this work.

• C. Linhares, D. Pedro and A. Traina are with Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, Brazil. E-mails: {claudiodgl, diogenes.pedro}@usp.br, {agma}@icmc.usp.br.
 • Luis E. C. Rocha is with Department of Economics & Department of Physics and Astronomy, Ghent University, Belgium. E-mail: luis.rocha@ugent.be.
 • J. Ponciano and J. Poco are with School of Applied Mathematics, Fundação Getulio Vargas, Rio de Janeiro, Brazil. E-mails: {jean.ponciano, jorge.poco}@fgv.br.

survey evaluated the size of networks in 124 visualization papers and classified networks with more than 200 nodes as very large ones [75].

We consider in this work large temporal networks as those with a few thousand nodes or timestamps. This decision was made based on the trade-off between the spread of nodes and edges throughout time, which may reduce the number of elements simultaneously shown depending on the layout (especially when adopting timeslicing [51]), and the increased complexity that the temporal dimension brings to timeline-based and animated visualizations (*e.g.*, mental map preservation). Although layouts and visualization systems designed to handle temporal networks work well with relatively small networks, most of them are ineffective used when dealing with large ones. The main reason for that is the high level of visual clutter in the layouts, which occurs depending on the network density and dynamics and is mainly caused by overlaps of nodes, edges, or groups [12]. Layouts that suffer from this issue include the traditional node-link diagram (animated or not), alluvial diagrams [56], GraphFlow [20], and the *Massive Sequence View* (MSV) representation [68]. Since each layout can benefit from a set of features, different methods have been proposed to improve layout readability and thus lead to new insights into the data. Examples include methods focused on node positioning [68], node and edge sampling [2, 77], timeslicing [51, 71], and representation simplification [1, 63, 76]. Even considering recent advances in layouts, algorithms for clutter reduction, and interactive visualization systems, enabling compelling visual exploration of large temporal networks remains a challenging and open problem [12].

Meaningful groups of nodes are frequently observed in real-world temporal networks, for example, those composed of nodes that interact more often between themselves than with nodes from other groups, the so-called communities [27]. Communities can represent, *e.g.*, groups of friends, co-authors, brain regions, similar proteins in a biological network, and others. In this paper, we propose LargeNetVis, a web-based visual analytics system created to assist experts and practitioners in analyzing small and large temporal networks. The system relies on network community detection to guide users in finding and inspecting regions of interest that may contain potentially relevant network elements. More specifically, LargeNetVis leverages three community taxonomies to allow users to understand these communities’ structural, temporal, and evolutionary patterns. Furthermore, users can explore communities with characteristics of interest and analyze their elements’ behaviors through the four linked views provided by the system.

In summary, our main contributions are: i) Integration of three taxonomies for communities in temporal networks, each accounting for a particular and relevant type of pattern; ii) LargeNetVis, a visual analytics system with multiple linked views that leverages the mentioned taxonomies to enhance the exploration of large temporal networks — the usefulness and effectiveness of the system were validated through a user study with 14 participants not involved in the system’s development; and iii) Two usage scenarios showing relevant and non-trivial patterns found in large and real-world networks.

2 RELATED WORK

Different strategies for modeling temporal changes and identifying patterns can be employed, *e.g.*, with the use of graphlets and motifs [18, 35, 39]. In this study, we leverage the temporal network community structure to identify patterns and enhance the visual analysis. This section discusses existing community taxonomies, aspects related to the visualization, and available systems.

2.1 Network community taxonomies

Different taxonomies have been proposed for network visualization. They categorize the network elements according to different criteria, *e.g.*, clusters or dimensions [17, 57], thus optimizing the process of finding elements with characteristics of interest. LargeNetVis leverages three taxonomies focused on different aspects of community categorization — the community’s structural, temporal, and evolutionary behaviors — to guide the network exploration. The categories considered by these taxonomies have also been analyzed by previous works [23, 40, 43, 49, 50, 56, 72] and represent the core of our approach.

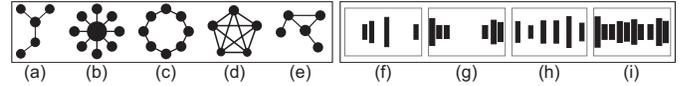


Fig. 2. Structural (a-e) and temporal (f-i) taxonomies. Vertical bars in the temporal taxonomy represent edges (y-axis) over time (x-axis).

Structural. A useful manner of understanding a network’s overall topology is categorizing its communities according to their structural patterns, such as stars, cliques, circulars, and others. For example, stars (see definition below) are commonly found on social networks. Different names for the same categories have been used in previous studies, such as clique and full, star and egocentric, loop and circular, among others [40, 72]. Based on Chenhui et al.’s work [40], we use five types of community structure classification: Tree, Star, Circular, Clique, and Low-connectivity (Fig. 2(a-e)). The *Tree* topology (Fig. 2(a)) has no cycles and differs from the *Star* topology (Fig. 2(b)), which corresponds to a central node concentrating most of the connections, with the peripheral/children nodes usually with only one connection. In the *Circular* case (Fig. 2(c)), every node has only two connections. The *Clique* (Fig. 2(d)) represents a case where all nodes are connected with each other (complete graph). Cases where none of the previous categories are suitable are called *Low-connectivity* ones (Fig. 2(e)). In real-world networks, a single community may contain more than one type of structure (details in Sec. 7) or be formed by variations of these types (*e.g.*, a circular formation with some additional edges linking non-adjacent nodes). By similarity, we categorize communities from this latter case as their reference structure (as circular in the previous example, for instance).

Temporal. Another important aspect concerns the overall temporal behavior of a community’s internal edges (intra-community edges), including their temporal activity (*i.e.*, how frequent they occur over time) and distribution (*i.e.*, how dispersed they are from each other). A previous study focused on a classification suitable for these temporal patterns, called Temporal taxonomy [43]. The first part of this taxonomy is the frequency, *i.e.*, if the intra-community edges are so frequent over time that the community is categorized as having *Continuous* activity, or if there is occasional activity and the edges occur in a *Sporadic* fashion. The taxonomy also considers dispersion information, *i.e.*, if the edges of consecutive timestamps are close to each other (*Grouped*) or if they are sparse and, consequently, *Dispersed*. By combining frequency and dispersion, this taxonomy creates a temporal classification that comprises four combinations, as illustrated in Fig. 2(f-i). In the first two combinations (Fig. 2(f-g)), communities are categorized as Sporadic and Dispersed (Fig. 2(f)), *i.e.*, there are no dense regions and at least one timestamp has no edges, or Sporadic and Grouped (Fig. 2(g)), when all edges are in a dense region but not continuous over time (we can think of as peaks of seasonal events). On the other hand, edges may occur frequently enough to be considered Continuous, and they can be grouped (Fig. 2(i)) or not (Fig. 2(h)).

Evolution. The final aspect that we consider is the network community evolution, which is essential to understand the dynamics and lifecycle of these groups of nodes, with events such as birth and growth. We consider the community taxonomy proposed by Pereira et al. [50], which includes six categories of evolutionary events: Birth, Death, Grow, Contract, Split and Merge. The first appearance of a community is classified as *Birth*, while disappearance is considered as *Death*. A community may maintain most of its nodes and *Grow* with the addition of new ones or *Contract* with the loss of nodes. A community may also *Split* into two new ones or *Merge*, leading to a new community.

2.2 Visualization of temporal networks

For temporal network visualization, layouts may be divided into timeline-based and animation techniques [12]. While animation suffers from mental map preservation in timestamps with high variation in the number of elements shown, timeline-based layouts usually lack screen space for a high number of nodes and timestamps. Among the animated layouts, the node-link diagram is still one of the most preferred representations [12]. It is used, *e.g.*, to visualize online networks, where nodes and edges arrive in a real-time manner [19]. While it

can suffer from scalability limitations when applied to networks with several nodes and edges per timestamp, representation simplification (e.g., super nodes [63]) may attenuate this problem. Despite strengths and limitations, both timeline and animations are efficient for temporal tasks. Timeline representations, however, are better for tasks that depend on the analysis of more than two timestamps [12].

A few studies try to enable visual analyses of non-temporal networks with millions of nodes [42, 46]. These studies, however, have focused only on general rather than local structures. Their visualizations lack user validation and are not suitable for temporal networks. In the context of temporal hypergraphs, *i.e.*, temporal networks where a single edge (called hyperedge) can connect more than two nodes, Valdivia et al. [67] proposed a timeline-based visualization that contains an aggregation of hyperedges and was tested in hypergraphs with up to 500 nodes. Also focusing on temporal hypergraphs, Fischer et al. [25] proposed a matrix-based visualization containing filtering, ordering, and interaction techniques. Temporal hypergraphs are significantly different from regular temporal networks, especially regarding structural patterns. Our focus is on the latter type of network.

Some studies combine temporal and structural analysis with summarization methods without considering community taxonomies [76]. The use of a network community taxonomy to summarize the visualization was initially proposed by Li et al. [40]. They focused on non-temporal networks and used the Structural taxonomy described in Sec. 2.1. Later, Wang et al. [72] extended this approach to temporal networks. They proposed a pipeline to group nodes according to common characteristics (e.g., via community detection), applying optional sampling, classifying the groups according to the Structural taxonomy, and providing a system with multiple views containing animation-based visualization, network statistics, and parallel coordinates. While Wang et al. demonstrated the importance of using a network community taxonomy to simplify temporal network visualization, they focused only on the structural aspect of the network and did not validate their pipeline and system with a user study.

Timeline-based techniques can use timeslices to represent the network time dimension. In our context, a timeslice is defined as a short observation period (e.g., 1-hour or 1-day length) and comprises all nodes and edges that occur in that period. Note that each timeslice includes several timestamps of the network. Although timeslicing methods are commonly used with discrete timestamps, they fail to reliably represent networks with continuous real-valued time (the so-called continuous (or event-based) dynamic networks) [33, 61, 62]. A set of visualization techniques have been proposed for this type of network, such as the DynNoSlice [61, 62], which employs space-time cubes to visualize the network’s events, and MultiDynNoS [5], which presents a multilevel approach to reduce running time when visualizing these networks. LargeNetVis is designed for networks with discrete timestamps.

2.3 Comparison with network visualization systems

We compared different systems for network visualization using the corresponding papers’ descriptions, system images, case studies, comprehensive surveys (e.g., [11, 12, 69]), and external videos, when available. Related to network visualization, there is a variety of systems and papers that propose techniques to enhance pattern identification [6, 9, 10, 16, 22, 25, 28, 31, 37, 41, 45, 54, 58, 68], in which we applied a selection criterion to compare. We filtered in systems that have animated or timeline-based representations, high relevance and similarity with our system, and a high number of paper citations and/or downloads. Matching the previous criteria, we selected six systems.

Table 1 summarizes seven systems considering 8 relevant aspects inspired by [60]. The *grouping* category (1) is related to node clustering or community detection, and *sampling* (2) is whether the system provides node, edge, or group sampling. The temporal dimension can be considered through *animation* (3) and/or *timeline* (4). We focus on the latter due to its advantage in temporal tasks that depend on more than two timestamps [12]. Likewise, a system may enable analysis at a global scale (*overview*) (5) and/or offer views dedicated to analyzing particular groups of nodes or time intervals of interest (*local view*) (6).

Table 1. Comparison of network visualization systems according to: (1) grouping; (2) sampling; (3) animation; (4) timeline; (5) overview; (6) local view; (7) visual scalability; (8) multiple views.

	1	2	3	4	5	6	7	8
Gephi [9]	✓	✓	✓		✓	✓	✓	
Pajek [10]	✓		✓		✓	✓	✓	
DyNetViewer [41]	✓	✓	✓		✓	✓		
DyNetVis [44]	✓		✓	✓	✓	✓	✓	
DyEgoVis [28]	✓			✓	✓	✓		✓
Wang et al.’s [72]	✓	✓	✓		✓		✓	✓
LargeNetVis (Ours)	✓	✓		✓	✓	✓	✓	✓

We consider that a system offers *visual scalability* (7) if it works for small and large networks — we considered the existing case and user studies to assess this criterion. Finally, the *multiple views* category (8) refers to whether the system provides different views on a single screen. Notably, DyNetVis [44] is the only system among the compared ones that provide both timeline and animations, while only Wang et al.’s [72], DyEgoVis [28] and ours have multiple views; these three are also the only web-based systems in the table. At last, the only system from Table 1 validated through a user study is ours (Sec. 6).

3 DESIGN TASKS

Different task taxonomies have been proposed to guide the design and evaluation of layouts and systems for temporal network visualization [3, 4, 7, 36]. As we are interested in enabling the analysis of structural and time-evolving patterns at both global and local granularity levels (*i.e.*, network, groups, and nodes), we designed our views and interactions to meet tasks considering the taxonomies proposed by Bach et al. [7] (taxonomy 1), and by Ahn et al. [3] (taxonomy 2). Taxonomy 1 considers three categories of low-level tasks, named *temporal* (when), *topological* (where), and *behavioral* (what), that may be combined to obtain compound tasks. According to this taxonomy, a system must have the following characteristics to support tasks from these categories [7]:

- T1:** allow easy identification and reaching of specific time steps (when);
- T2:** allow easy identification, situation, and tracking of elements (e.g., nodes or groups) with particular properties (where);
- T3:** allow the understanding of the nature of changes that affect the elements (what).

Focused on aspects related to the network evolution, taxonomy 2 describes tasks related to the three dimensions *entity* (an element of interest, such as node, edge, group, network), *property* (a characteristic of an element, given by centrality measures or domain properties, for example), and *temporal feature* (an event that causes changes in the state of an element, such as growth or death). Although all three dimensions are somehow related to categories also foreseen by taxonomy 1, the last dimension provides more details about existing types of temporal events, four of which are also observed in the community evolution taxonomy described in Sec. 2.1 (birth/death and growth/contraction). Since the community evolution taxonomy considers merge/split events in addition to the four mentioned types, we restricted our task T3 to consider all six types of events from this taxonomy. Finally, our system was also designed to support:

- T4:** identification of two other types of events from taxonomy 2’s third dimension: single occurrences (*i.e.*, the addition or deletion of an entity) and *peak or valley* events (*i.e.*, abrupt increases or decreases of an element property).

4 LARGENETVIS

This section describes LargeNetVis’s workflow and visual components, which were developed to support our design tasks as depicted in Table 2.

4.1 Workflow

Fig. 3 describes the system’s workflow. First, users provide a temporal network of interest and its respective node metadata (optional). Then, they can apply one of the node and edge sampling methods offered

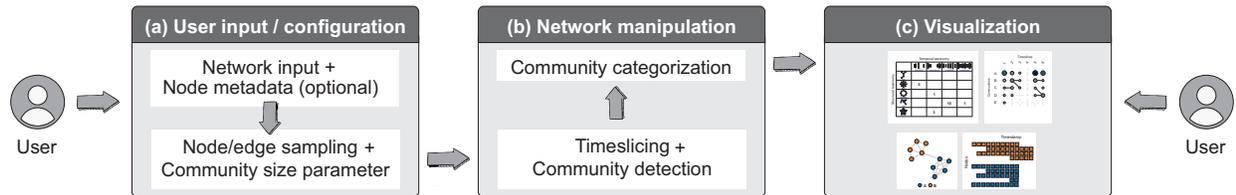


Fig. 3. LargeNetVis’s workflow. (a) The user provides a network of interest and the corresponding node metadata (optional). He/She can also apply node/edge sampling and adjust the minimum community size parametrization. (b) The system divides the network into timeslices of equal length and then detects network communities inside each of them. These communities are then categorized according to three taxonomies from the literature. (c) Finally, users explore communities with characteristics of interest and their nodes and edges through the four linked visual components.

by the system, including random node/edge sampling and Snowball sampling [64], a method suitable for large networks. Other methods, e.g., EOD [77], could be easily incorporated into the system. Users are also free to report the minimum number of nodes that a network community must have in order not to be ignored by the visualization — the community detection procedure is a further step that will be explained later on. Very small communities are often perceived in large networks, and they can pollute the layout with irrelevant information.

The next step is to divide the network data according to a user-defined number of timeslices (Fig. 3(b)). To assist users in choosing a proper number of timeslices for analysis, LargeNetVis runs the nonuniform timeslicing method proposed by Ponciano et al. [51] with all edges having the same importance. By considering local edge density, the method defines the best timeslice length for each sliding window of timestamps. LargeNetVis then suggests a range of potentially suitable numbers of timeslices based on how many timeslices are required to meet from the minimum to the maximum length returned by Ponciano et al.’s algorithm. Within this range, the default value meets the average timeslice length, as recommended for global pattern identification [51]. Note that the choice of the number of timeslices strongly affects the quality of results and response time of the remaining workflow steps, consequently affecting the visualization and pattern identification.

For each timeslice, network communities are detected using the *Instant Optimal* strategy [53], which prioritizes the quality of the community structure rather than temporal consistency. In this strategy, communities are detected separately for each timeslice and the overlap between communities from consecutive timeslices is computed. Any non-overlapping community detection algorithm could be used in this step. The LargeNetVis current version employs Louvain [14], a greedy and agglomerative algorithm based on modularity optimization that finds the best network partition by establishing groups (communities) with densely connected nodes. Although modularity optimization faces the issue of resolution limit, which may lead small and meaningful communities to be clustered into larger ones [26, 66], Louvain is still one of the most popular and recommended algorithms [27]. Also, its computational complexity is linear on the number of edges, which is an essential advantage in our case. We point the reader to the survey by Fortunato and Hric [27] for a broader understanding of network community detection algorithms. Other algorithms (e.g., Infomap [55]) could be incorporated into the system with little effort. Different algorithms would return different communities; the detection quality directly affects the remaining steps of our workflow and, ultimately, the visualization and pattern identification.

The system then categorizes each community of each timeslice using the three taxonomies described in Sec. 2.1 (Structural, Temporal, and Evolution). Finally, users visually explore the network data under different perspectives and granularity levels through the four linked visual components offered by the system (Fig. 3(c)). Respecting the well-established visual information-seeking mantra “*overview first, zoom and filter, then details-on-demand*” [59], these four components have different roles in the network analysis. Two of them provide an overview of the network structure and evolution (Sec. 4.2), and two enable local-level analysis (Sec. 4.3). In addition to the four mentioned components, the system also shows a useful numerical information panel with network, community, or node information, depending on the granularity level of analysis. By default, this panel is named *Network detail* panel, and it shows the overall number of nodes, edges, times-

Table 2. Visual components of LargeNetVis and the four tasks.

	T1	T2	T3	T4
Taxonomy Matrix		✓	✓	
Global View	✓	✓	✓	
Node-link diagram		✓		
Temporal Activity Map	✓	✓		✓

tamps, and communities detected, as well as the mean modularity value returned by the detection procedure.

4.2 Summary views

Taxonomy Matrix. This view summarizes the general network structure, temporal distribution, and evolution by showing the number of communities that correspond to each combination of categories from the two user-selected taxonomies (tasks T2, T3). Users can freely combine taxonomies two-by-two, including a taxonomy with itself — in this case, the main diagonal would show how the network is characterized according to that taxonomy alone. Fig. 4(a) illustrates this view: in the example, three communities are *Cliques* (structural taxonomy) and *Sporadic and Grouped* (temporal taxonomy).

Once the user has chosen two taxonomies for analysis, he/she can select a particular category combination of interest by clicking on its corresponding matrix cell. It is also possible to select multiple combinations, a row or a column, and even the whole matrix (by clicking on the top-left cell).

An *alternative* that we have considered for this view would be the adoption of bar charts depicting the number of communities that meet each category combination. The high number of possible combinations (there would be 30 bars when considering structural and evolution taxonomies), allied to the fact that the bar chart would emphasize the most frequent combinations while one may be interested in analyzing the less frequent ones, led us to adopt the described Taxonomy Matrix.

Global View. The communities that belong to the selected cell(s) in the Taxonomy Matrix are highlighted in Global View, a view focused on the community evolution behavior (Fig. 4(b)). This way, users can locate communities with characteristics of interest (T2) and the timeslices they occur (T1), and also know the type of temporal changes affecting them (T3). Global View can be thought of as a grid where columns represent timeslices and rows represent communities, which are depicted as circles with varying colors and sizes, visual attributes that can map information of interest (for example, the community size or average node degree). The current version of LargeNetVis adopts redundant coding [73], mapping the community size on both color and circle’s size (Fig. 4(b)).

As illustrated in Fig. 4(b), Global View enables analyses that rely on the temporal distribution of communities, including the identification of idle periods (see timeslice #4) and periods with many or few communities (compare timeslices #1 and #3, for instance). The evolutionary behaviors are visually represented as links with varying thickness between a “from” community (timeslice #X) and a “to” community (timeslice #(X+1)). The design of the links enables quick identification of splits and merges (Fig. 4(b-I)), growth and contraction (Fig. 4(b-II)), and preservation (Fig. 4(b-III)), which occurs when the number of nodes in a community does not change (T3). Linking communities

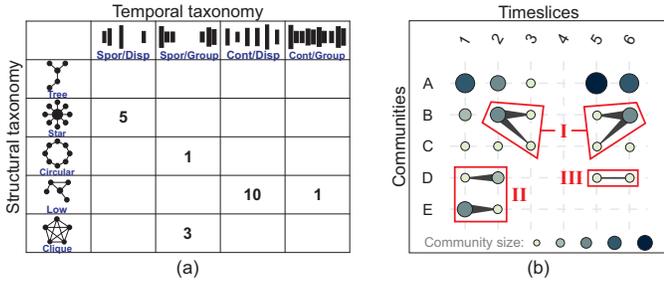


Fig. 4. Visual components of LargeNetVis that enable global analysis: (a) Taxonomy Matrix; (b) Global View. The evolutionary behaviors are: (I) split/merge; (II) grow/contract; and (III) preserve.

from different timeslices improves mental map preservation.

To avoid long links and overlaps, LargeNetVis implements a simple but efficient strategy that minimizes the link length. For the first timeslice (grid column), communities are positioned following the order they appear, *i.e.*, the order they are returned by the community detection algorithm. For each remaining timeslice, we position “to” communities such that the Euclidean distance between them and their corresponding “from” communities is minimized; communities not linked so far go to the remaining positions in the order they appear. In the end, we obtain minimal lengths for all links but merge-related diagonals (cases where there are two “from” for a single “to”). To solve these cases, the “from” community that is far from its “to” (diagonal link) switches places with the non-linked community closest to the other “from”. We provide the algorithm in the supp. material (Sec. A). Note that other positioning strategies could have been adopted. The impact of different strategies in the layout is discussed in Sec. 7.

Users can zoom in/out and pan to explore particular regions of interest. When a circle is hovered over, the system shows a tooltip informing the timeslice number, the number of nodes in that community, the community position in the grid, and the three taxonomy categories that characterize that community (see Fig. 1(F)). When a link is hovered over, it is shown a tooltip with information about the communities and timeslices involved, the communities’ sizes, and the type of event. Once the user has found a community of interest to analyze, he/she clicks on the corresponding circle and is redirected to the remaining two views (node-link diagram and Temporal Activity Map – TAM), which will be described in the next section. Details of the clicked community, including its number of nodes, edges, and timestamps with edges, are also shown in the numerical information panel, which is now referred to as *Community detail* panel (see Fig. 1(E)). If the user clicks on a circle that refers to a community categorized differently from those selected in the Taxonomy Matrix, the matrix cell representing this new community is automatically selected and the previously selected cells are deselected. Consequently, all communities in Global View that fit the newly selected categorization become highlighted (and only them). With this functionality, we help users to find communities that share the same properties as the ones they consider interesting. Note also that Global View enables exploratory visual analysis, *i.e.*, users can go directly to this view without exploring Taxonomy Matrix first.

We have considered *alternative design choices* for Global View. First, we adopted a timeline-based approach instead of an animation because timelines are better for tasks involving more than two time steps [12], an important advantage for identifying active/inactive time intervals (*e.g.*, days and nights) (T1) and tracking a community of interest throughout the observation period (T2). After this decision, we considered the adoption of traditional alluvial (or Sankey) diagrams for this view. We discarded them after preliminary tests due to the amount of screen space they require for networks with many timeslices or communities per timeslice. Inspired by visualizations employed for node-level analysis [67, 68], we created a compact representation that uses circles over time to indicate community activity and evolution. We also decided to map the community size using redundant code (circle’s size and color) to enhance the identification of small/large communities (T2).

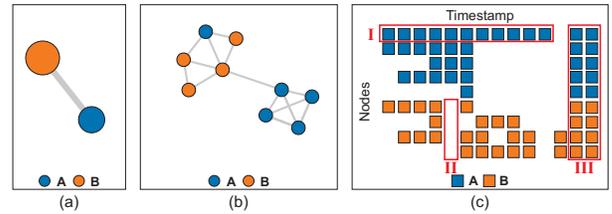


Fig. 5. Visual components of LargeNetVis that enable local analysis: (a) (Super)node-link diagram; (b) Node-link diagram; (c) TAM.

4.3 Local views

Node-link diagram. After finding a community of interest, users may want to analyze it from a more detailed perspective, *e.g.*, looking for intra-community nodes that share common properties (*e.g.*, same metadata) or studying a node’s interaction dynamics. LargeNetVis enables local-level analyses of these types (T2) through the traditional node-link diagram and a TAM [44]. Both views are simultaneously loaded when the user clicks on a circle in Global View.

Our node-link diagram, illustrated in Fig. 5(b), shows the intra-community nodes and edges in an aggregated fashion, *i.e.*, by showing all of them at once regardless of the times of the edges. If categorical node metadata is provided (*e.g.*, student’s class or employee’s department), the system enhances the analysis by associating different colors to nodes with different metadata values. Users are free to change any default color by using a color picker. The node positioning is defined such that nodes more frequently connected are close to each other (spring-force algorithm [15]) and edge bundling is optionally applied to reduce clutter [34]. The impact of different positioning methods is discussed in Sec. 7. As occurs with Global View, users can zoom in, zoom out and pan for more detailed analysis or manual node repositioning. When a node (circle) is hovered over, the system shows a tooltip informing the node id and its metadata value (if any). If a circle is clicked, the corresponding node is highlighted in TAM (described below) and the numerical information panel (now named *Node detail* panel) exhibits this node’s id, metadata value, normalized degree value, approximated and normalized betweenness centrality value (using 25% of the nodes for computation), and closeness value. These centrality values are computed for the timeslice under analysis.

To enhance the analysis of large communities, the system also offers a multilevel node-link diagram for those communities with more than a user-defined number of nodes. In these cases, LargeNetVis shows, by default, a summarized representation where nodes (referred to as “supernodes”) represent sub-communities and “superedges” indicate the presence of inter-sub-community edges (Fig. 5(a)). The algorithms used for “supernode” positioning and sub-community detection [43] are the same as before (spring-force and Louvain, respectively). The color and size of a “supernode” (circle) map the predominant color among the sub-community’s nodes and the number of nodes inside it, respectively. The “superedge” thickness maps the number of inter-sub-community edges between the two involved “supernodes”. Users can freely navigate between summarized and non-summarized node-link diagrams. When a “supernode” is clicked, its nodes are highlighted in both non-summarized diagram and TAM.

We performed preliminary tests to compare our node-link diagram with an *alternative approach* (the adjacency matrix visualization [13]). We chose the node-link diagram as it was easier to visualize the inherent structure of the communities (T2) using the diagram than with the matrix (recall Fig. 2(a-e)).

Temporal Activity Map. Besides the well-established node-link diagram, LargeNetVis includes the Temporal Activity Map (TAM), a layout successfully used for the visual exploration of networks from different domains [44]. TAM is a timeline-based visualization strategy focused on showing active nodes (*i.e.*, connected to others) over time. The vertical and horizontal axes represent the nodes and timestamps, respectively, and a colored square is drawn whenever the corresponding node is active in the network. The edges are hidden in this layout. Our system exhibits TAM at the same time as the node-link diagram and uses the same colors for node metadata (T2). For coherence, colors

changed through the mentioned color picker are also reflected in this layout. As illustrated in Fig. 5(c), active nodes are represented by fixed-size squares that form “blocks” in the visualization that enable the identification of patterns related to active/inactive nodes and groups over time as well as intervals with or without activity (T1, T4). Some examples of patterns are shown in Fig. 5(c): the presence of a node active in several consecutive timestamps (Fig. 5(c-I)), absence of all nodes with label “B” in a particular timestamp (Fig. 5(c-II)), and all nodes of the network active in the last two timestamps (Fig. 5(c-III)).

In our case, LargeNetVis provides the TAM layout for the community being analyzed and also shows a line chart depicting the number of edges per timestamp in that community. Nodes are positioned according to their metadata, and nodes with the same metadata are positioned in the order they appear. We discuss the impact of using other strategies for node positioning in Sec. 7. When a square is hovered over in TAM, the system shows a tooltip with the respective timestamp, the node id, and node metadata. By clicking on a square, the corresponding node is highlighted in the node-link diagram. We studied some *alternatives* before adopting TAM, for example, the use of an animated node-link diagram instead of a static node-link diagram associated with the TAM. We chose the second option because timeline-based approaches are preferable for analysis that depend on more than two timestamps [12], for example during peak or valley analysis (T4).

4.4 Scalability

Some procedures in the system’s workflow make LargeNetVis suitable for exploring large networks. First, node/edge sampling methods may be used as an initial attempt to reduce network data (Fig. 3(a)). Second, LargeNetVis detects communities throughout time and focuses on them (*i.e.*, on their nodes and edges) to guide users to regions of interest in the network (Fig. 3(b)). Although this guidance is helpful when exploring large networks, the number of interesting communities may remain too large and additional guidance may be required; LargeNetVis efficiently addresses these cases by also providing community categorization (Fig. 3(b)). A potential side effect of focusing on the community structure is the discarding of inter-community edges and the consequent impact on the study of the overall structure and evolution of the whole network. Following previous work [43], we consider this acceptable since inter-community edges are not as frequent as the intra-community ones [27] and nodes connected by them would create groups not as meaningful as the actual communities.

To attenuate visual scalability issues in Global View depending on the number of timeslices being used and communities detected, a default zoom-to-fit is employed, so the representation of the smallest communities is reduced to a few pixels in the worst-case scenario. Further efforts include from zoom-in/out, pan, and scrolling, to the selection of communities from categories of interest using the Taxonomy Matrix and the filtering of communities with a minimum desired number of nodes. Using a temporal network with 50,514 nodes and 108,132 edges (Twitter network [48]), supp. material (Sec. B) shows a case where Global View contains 2,033 communities and 100 timeslices.

The visual exploration of a community of interest (Fig. 3(c)) may also be impaired depending on its number of nodes and edges. To attenuate this scalability issue, besides spring-force node positioning and edge bundling, LargeNetVis shows by default a summarized node-link diagram when the number of nodes in the community is higher than a user-defined threshold (Sec. 4.3). In our supplemental material (Sec. B), we show an example of a summarized representation of a community with 3,886 nodes and 5,035 edges. Free navigation between summarized and non-summarized diagrams is allowed and both views are coordinated between themselves and with TAM, thus optimizing the process of finding and analyzing nodes (or groups) of interest in all views. Finally, we have also analyzed the computational scalability of LargeNetVis by (i) collecting user feedback regarding the overall system’s response time, and (ii) measuring the running times of the network manipulation procedures (Fig. 3(b)) for all networks and configurations discussed throughout Sections 5 and 6. LargeNetVis is fast and its response time was well-received by our user study participants. We will provide details about this analysis in Sec. 6.5.

4.5 Implementation

LargeNetVis was built with a client-server architecture and is available at <https://github.com/claودیodgl/LargeNetVis>. All views were implemented using Javascript and the D3 library [21]. The server side uses Python and employs some popular frameworks and libraries, such as Flask [30], NetworkX [47], and others.

5 USAGE SCENARIOS

We present two usage scenarios to show the LargeNetVis capabilities.

5.1 MovieLens network

The network we consider in this usage scenario (MovieLens network) was built from the *MovieLens 1M dataset* [32, 74], which contains data gathered from the same-name rating website (<https://movielens.org/>). The network is composed of 9,940 nodes (3,900 users and 6,040 movies) and 1,000,209 edges (movie ratings on a 5-star scale) distributed in 1038 timestamps (days). It is a bipartite network where edges are defined as “users rating movies”. We consider this network large because of its number of nodes and timestamps. Regardless of the number of timeslices tested in our experiments, all communities detected are born and die during the same timeslice (T3). The evolutionary behaviors growth/contraction and merge/split were not identified. We believe this occurs because there is probably no user rating the same movies more than once.

As stated in Sec. 4.2, LargeNetVis allows users to freely explore communities directly from Global View, a resource that may lead to new insights and unexpected pattern discovery. The presented usage scenario begins with such an exploratory analysis (Fig. 1). We first filtered the MovieLens network to consider only negative ratings (*i.e.*, with 1 or 2 stars, out of 5) for the top-6 most rated movie genres (action, comedy, documentary, drama, horror, thriller). Our focus was on those movies that belong to a single category among the mentioned ones, which resulted in 6,358 nodes and 47,809 edges distributed in 1,002 timestamps. Then, we adopted 10 timeslices to represent this observation period (value empirically chosen among the suggested range) and discarded communities with less than 5 nodes.

With this network configuration, we quickly identified on Global View a timeslice (#7) with more communities than the others (T1). This timeslice has 10 more communities than the timeslices with the least number of communities (#2 and #3, see Fig. 1(B)) and one more than the timeslices containing the second-highest number of communities (#4 and #5). We then became interested in exploring this additional community from timeslice #7 (Fig. 1(F)). We clicked on this community’s circle on Global View and the system informed us, through the Taxonomy Matrix cell highlighting, that this community is categorized as *star* and *sporadic/dispersed*, characteristics that represent 13 communities in the network (see Fig. 1(A)) (T2). We also noticed that this community contains an equal number of nodes and edges (33), and only 3% of its timestamps have activity (Fig. 1(E)).

Moving further from global to local analysis, LargeNetVis shows the node-link diagram and TAM for any community selected on Global View. As we see in the node-link diagram from Fig. 1(C), our community of interest is composed of two individuals that rated several movies negatively. While one of these individuals rated movies from different genres (person P1), another rated only comedies (person P2). Note also that two movies were rated by both of them: *House Arrest* (1996) and *Life with Mikey* (1993). By looking at the TAM layout, we see another behavior shared by both individuals: except for two movies rated by P1 some days later (a drama and a comedy), each individual took only one day to rate all movies (Fig. 1(D)) (T1,T4).

Fig. 1(C) illustrates a scenario where a single individual rates several movies; however, the inverse is also possible. By focusing the analysis on drama movies that were positively rated (*i.e.*, with 4 or 5 stars, out of 5) and considering again 10 timeslices and communities with at least 5 nodes, we end up having 5,875 nodes and 72,695 edges distributed in 1,038 timestamps (days). The network under this configuration has only 1 community categorized as *star* and *continuous/dispersed*, which attracted our attention when looking at the Taxonomy Matrix (Fig. 6(A)) (T2). This community has 112 nodes (9 movies and 103 users) and

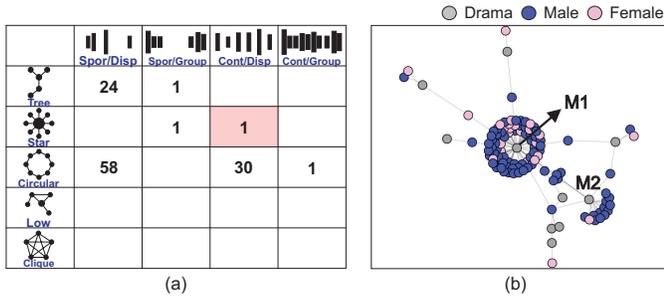


Fig. 6. The only *star* and *continuous/dispersed* community when considering drama movies positively rated. It was found thanks to the Tax. Matrix (a) and depicts two movies M1 and M2 rated by several people (b).

contains two movies that were positively rated by a lot of people, mainly by males (Fig. 6(b)). The most rated one (M1) is *One Flew Over the Cuckoo's Nest* (1975), winner of 5 Oscars and top-rated movie #18 of all times on IMDb¹; the second movie (M2) is *The Hurricane* (1999), nominated for an Oscar in 2000².

5.2 Primary School network

The Primary School network [29] is composed of interactions recorded during two days between students and teachers of a primary school in Lyon, France. There are 242 nodes and 125,773 edges distributed in 5,846 timestamps (each comprising a 20-sec interval). The network considers five school grades (first to fifth) and two classes (A and B) per grade, in a total of 10 classes. Each class has an assigned teacher. We consider this network a large one due to its number of timestamps.

By adopting 26 timeslices and communities with at least three nodes to analyze this network, we see that most communities are categorized as *clique* by the structural taxonomy (100 of 116) and that there are no *stars* in the network (Fig. 7, Tax. Matrix). Besides, there are 12 *continuous/grouped* communities while *sporadic/grouped* is the less observed temporal categorization, with only one community (T2). Note that the presence of communities with these characteristics is the opposite of what we observed in the MovieLens network (Fig. 1(A)), where there are no *continuous/grouped* communities or *cliques* (MovieLens is bipartite, after all).

When analyzing the first day of school (timeslices #1 to #7) using Global View, we see at least one growth (Fig. 7(I)), one contraction (Fig. 7(II)), and two merges (Fig. 7(III-IV)) involving some of the 68 *clique* and *continuous/dispersed* communities (T1, T2, T3). As shown in Fig. 7, these four events occur between communities that represent same-grade classes (e.g., IV is a merge between classes 3A (purple) and 3B (brown)). This is expected since same-grade students interact more often among themselves than with students from other grades [29]. Finally, the merged community from IV is highly active over time except for a small time interval (T4) (see TAM in Fig. 7), a behavior that does not occur as frequently in MovieLens as in this network (recall Fig. 1(D)). This discrepancy may explain why the Primary School have most communities with *continuous* activity while MovieLens have most communities having *sporadic* activity (compare the Taxonomy Matrices from Fig. 7 and Fig. 1(A)).

6 USER STUDY

We also performed a user evaluation to collect feedback about the system's usefulness and usability, and ideas for improvement.

6.1 Network data sets

The participants of the user study employed LargeNetVis to analyze two different networks. The first one is the Primary School network presented in Sec. 5.2. The second network, Sexual contacts [52], describes sexual encounters between sex-workers (sellers) and clients (buyers). This network is composed of 12,157 nodes, 34,060 edges, and 1,000 timestamps (each timestamp comprises a 1-day interval). We consider it a large network mainly because of its number of nodes.

¹<https://www.imdb.com/title/tt0073486/>. Accessed: 2022-03-22.

²<https://www.imdb.com/title/tt0174856/>. Accessed: 2022-03-22.

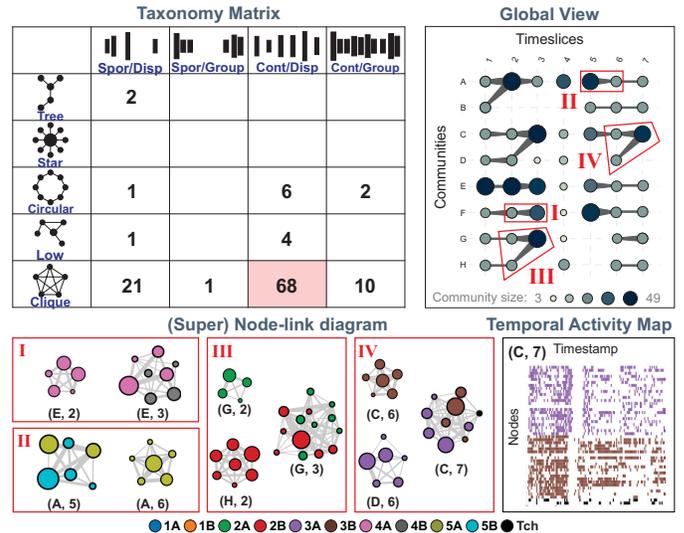


Fig. 7. Four evolutionary behaviors involving some of the 68 *clique* and *continuous/grouped* communities in the Primary School network: (I) a growth, (II) a contraction, (III, IV) two merges. The summarized node-link diagram is shown.

6.2 Procedure

Our study was conducted both online and remotely, with each participant answering the questions at different times and using their personal computers, so we did not have any control over the environment of the study. We provided each participant with a video containing instructions about the system's basic concepts and functionalities. The video was 7 minutes long and continuously available during the experiment.

6.3 Online questionnaire

The questionnaire was written in Brazilian Portuguese, in which all participants were fluent. We divided the questions in (i) background and experience; (ii) four basic questions to test the user's comprehension; (iii) six advanced questions using both networks (two multiple-choice and four open questions); (iv) Likert-scale-based questions to evaluate the user preference of the system, along with open questions to justify their choices; and (v) a mix of multiple-choice and open questions to collect the users' feedback about the system. The complete description of the questions and expected answers are available in the supplemental material (Sec. C). This questionnaire structure was based on similar user studies evaluating layouts or systems [8, 65].

The questions were divided aimed at evaluating the functionalities, layouts, and perceptions of the participants under different tasks. First, we assessed the comprehension of the basic functionalities of LargeNetVis using four basic multiple-choice questions about the Primary School (B1 – B4). Then, we asked the participants to re-open the network under a different configuration (timeslices and filters), which led to different layouts used to answer the first three advanced questions (A1 – A3), composed of one multiple-choice and two open questions. Finally, the last three questions (A4 – A6, one multiple-choice and two open questions) are related to the Sexual contacts network. In the advanced questions, the first open question for each network (A1 and A4) was focused on guiding the participants to specific patterns, and the last questions (A3 and A6) encouraged them to perform a free exploration of the system.

6.4 Participants

The experiment recruited 14 participants, which participated voluntarily in the experiment. They are professors, postdocs, and graduate students with a background in Computer Science. We asked about their previous experience working in the Visualization and Network science fields: most have published articles or advised students in at least one of the fields. One participant also has extensive experience working with visualization in a number of international companies. We also asked if

they were aware of any visual difficulties, such as color blindness, but no one indicated limitations. They had no access to our usage scenarios.

6.5 Results

Participants' background. We divided the participants into non-overlapping groups based on their self-description of experience using a scale with None, Basic, Intermediate, and Advanced knowledge for each field (Visualization and Network science). We considered a participant as a *Specialist* if he/she has advanced knowledge in both fields (we have 3 specialists). If he/she has advanced knowledge in only one field, we included him/her in the *Advanced* group (3 individuals). Participants with an intermediate knowledge in at least one of the two fields were included in the *Intermediate* group (8 individuals). These three groups were sufficient to categorize all 14 participants.

The participants spent 50m34s on average to answer the questionnaire and perform exploratory analyses. Considering the analyzed groups, the time spent answering the questionnaire increased based on the experience, with specialists spending almost one hour. We assume that participants with more knowledge of the fields spent more time assessing technical aspects of the layouts and analyzing their findings.

Questionnaire answers. All participants answered the four basic questions (B1 – B4) correctly. Two of our advanced multi-choice questions (A2 and A5) aimed at validating the layouts. In question A2 (about Global View), we provided a 5-Likert scale asking the participants their agreement level on the ease of identifying communities that merge, with all participants providing positive answers (43% agreed and 57% strongly agreed). In question A5 (about the node-link diagram and TAM), they had to describe the structural pattern and temporal distribution of a particular community. All participants marked the correct answer. Having all correct answers to our multi-choice questions, we validated that the participants understood and correctly performed the intended tasks for all considered networks and configurations.

The open questions A1, A3, A4, and A6 were formulated to enable exploratory tasks. Questions A1 and A4 (one from each network) were designed to guide the participants in exploring specific patterns. A1 asked participants to indicate communities that suggest the occurrence of a lunch break in the primary school; we obtained 92% of correct answers. Question A4 asked the participants to explore a given community using both the node-link diagram and TAM, and describe eventual patterns found. Again, 92% of the participants indicated the expected answer, which was the relationship between one seller and several buyers, with connections spread over time.

Questions A3 and A6 asked participants to freely explore the layouts from both networks and describe relevant patterns eventually found, which should be different from those described in other questions. In this case, the participants were free to write their personal findings. Our objective with these questions was not to evaluate correct or wrong answers, but to understand the overall perception of the users with the visualizations and the system's capability of enabling pattern identification. In question A3 (about the primary school), 72% of the participants answered that they could identify at least one pattern not previously described in other questions. One pattern mentioned by 30% of participants refers to the presence of the biggest network community on the second day of the network. This community comprises connections among students from six different grades and three teachers, which was very unusual for this network. We discuss this and other patterns mentioned by the participants in the supplemental material (Sec. D).

In question A6 (about the Sexual contacts network), 57% of the participants answered they could identify at least one pattern not previously described in other questions. The majority of the answers were related to the rarity of a seller's relationship with multiple buyers, for cases of small (Fig. 8(a-b)) and large communities (Fig. 8(c)). Also, one participant extended this analysis (Fig. 8(b)) by highlighting an interesting pattern in the TAM layout which shows that most encounters are 1-1 except for one involving 3 individuals on the same day. A participant also pointed out that *"the proportion of star-type communities is much smaller than tree and circular, where inadequate protection can cause the spread of sexually transmitted infections"*. About the Tree type of structural pattern (Fig. 8(d)), a participant highlighted that *"many large*

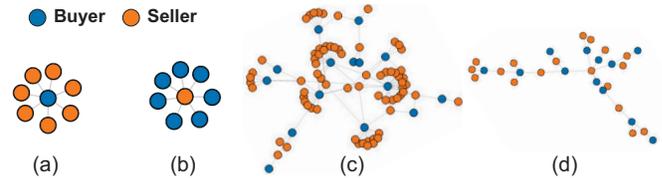


Fig. 8. Sexual contacts: some patterns identified through question A6.

communities are of the Tree type structure, which shows that they are communities formed with few connections, in which a buyer can buy from multiple sellers, but there won't be many other buyers buying from these same sellers". This participant's observation is in agreement with this network's Assortativity value, which indicates a tendency of active buyers interact with less active sellers and vice-versa [52].

LargeNetVis performance. We used two 5-point Likert-scale questionnaires to measure the participants' preferences about LargeNetVis and specifically about each layout used (Fig. 9(a,b)). We evaluated four main aspects of the system: whether the system offers visual scalability (*i.e.*, whether it worked well for both networks) (LQ1), is fast (LQ2), is useful (LQ3), and is intuitive and easy to use (LQ4). Regarding LQ1, three participants (two with intermediate and one with advanced experiences) were not sure about the system's capability on this matter (Fig. 9(a), LQ1). The major concern was that the system was not tested using networks with 1 million nodes or so. In contrast, the specialists and other participants positively evaluated the system, indicating *"a good balance between the number of elements in global and local views"*, and also that *"it was possible to understand relationships in the same way in both networks"*.

The other three aspects (Fig. 9(a), LQ2 – LQ4) were positively evaluated by all participants. Some interesting aspects highlighted by the participants about LQ2 were that *"responses to commands are immediate and instantaneous"* and *"the interaction worked very fast, it just took a while to load the sexual network at first"*. To obtain a quantitative assessment, we also measured the average running times for the workflow procedures depicted in Fig. 3(b), and we show our results in the supp. material (Sec. E). The system spent a maximum of 12.66 sec. to run all of these procedures for any network and number of timeslices discussed throughout our usage scenarios and user study. Both the user feedback and quantitative evaluation demonstrate the well-versed response time and computational scalability of LargeNetVis.

Regarding the system's usefulness (Fig. 9(a), LQ3), participants claimed that the system *"is very useful for dealing with unknown data in an exploration stage, such as observing the behavior of temporal data"*, and *"this type of tool is vital for analyzing the evolution of communities in a temporal network"*. Regarding intuitiveness and ease of use (Fig. 9(a), LQ4), participants argued that *"it is easy to use based on the graphical interface proposed and is quite intuitive for those who know networks or graphs"* and *"I found it simple to use, being easy to switch between instants of time and follow the changes in relationships"*.

Feedback on individual layouts. Fig. 9(b) shows the participants' assessments of the quality of each layout, more specifically on whether the layout is useful and helped in the network analysis. The first evaluated layout was the Taxonomy Matrix, the only one to receive negative reviews. The critique of the two participants in this case was that they did not use this layout since the taxonomy information was also available under interaction in the tooltips shown in Global View. TAM also received two neutral assessments, justified by the participants who affirmed that *"to completely understand TAM it is necessary to have a better knowledge of the network"*, and that *"TAM needs better interaction techniques, such as the selection of a node in the node-link diagram and its automatic selection in the TAM layout"*. We implemented this feature after the user study. Besides those neutral and negative assessments, participants considered that *"the integration of all layouts facilitated the analysis and each brought a different aspect of the network being analyzed"*. Another participant even complemented this idea by explaining that *"the Taxonomy Matrix allows a quick interpretation of communities, the node-link diagram is very useful to analyze the cases separately, and TAM adds the missing temporal*

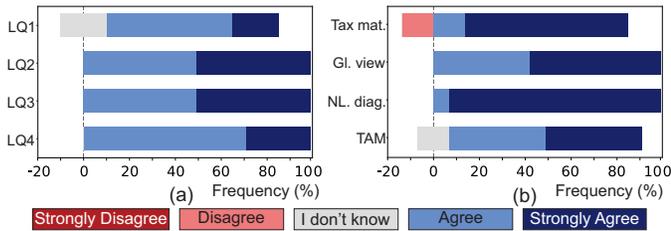


Fig. 9. Participants’ answers related to (a) LargeNetVis and (b) individual layouts. The bar length is the percentage of respondents who chose a specific Likert level. Questions from (a) concern the system’s visual scalability (LQ1), speed (LQ2), usefulness (LQ3), and intuitiveness and ease of use (LQ4). Questions from (b) concern whether each layout is useful and helps in the network analysis.

information”. Overall, the participants preferred Global View and the node-link diagram over the others.

Among the participants, 28% mentioned other systems that allow similar analysis, such as Gephi, Graphviz, Stanford Network Analysis Platform (SNAP), and Python libraries (e.g., NetworkX). According to them, LargeNetVis is preferred because “it is already a specific and ready-to-use tool for network analysis”, “it doesn’t mess up the graphs as much and it is more intuitive and easy to use” and “it is a low-code tool; users only need to format their input data and the visualization is ready to use”. On the other hand, disadvantages of our system include “the lack of flexibility in positioning views and exporting filtered data” and “it could be possible to use the views as part of other systems”.

We also asked the participants to evaluate the efficiency of the three taxonomies. In general, participants had very positive answers, claiming that “the taxonomies provide more objectively analyses, working as a filter of the network” and also citing an example: “if I am interested in a certain structure, I discard the others and optimize time and efforts”. A participant argued that the Taxonomy Matrix “shows to the user the pre-extracted and already categorized patterns, allowing less experienced users to analyze the data more directly and easily”.

Suggestions for improvement. Some participants suggested implementing edge weights to work with weighted networks, selecting multiple items in the layouts, improving the interaction between the node-link diagram and TAM (implemented after the user study), and others. The previous version of the system only exhibited network-level numerical information, and the most interesting suggestion we received in that sense was to optimize the numerical information panel by changing its information according to the granularity of analysis (network, community, or node-level information).

7 DISCUSSION AND LIMITATIONS

Although LargeNetVis is a useful tool for exploring large temporal networks, some aspects may affect the analysis and pattern identification.

Number of nodes, edges, and timestamps. Exploring networks with more than a few thousand nodes and timestamps (i.e., very large networks) could lead to a number of communities (or sizes) so elevated that the analysis would be impaired even considering LargeNetVis’s current capabilities (e.g., the multilevel node-link diagram). We believe LargeNetVis is an encouraging first step towards the analysis of such very large networks, and we now intend to extend the system by incorporating more sophisticated filters (e.g., based on centrality values) and multilevel approaches in the other views (TAM and Global View).

Node and community positioning. As studied by previous work [15, 43, 44, 68], node positioning strongly affects overlaps, visual clutter, and pattern identification. As an example, identifying merges in Global View or interesting groups of nodes in the node-link diagram and TAM would be easier or harder depending on the adopted positioning. Even though the positioning algorithms already incorporated into the system have been sufficient for the tested cases, we now intend to include more sophisticated algorithms, e.g., *Community-based node ordering* [43] for TAM and ad-hoc solutions specific for each type of structure.

Multiple categorizations of the same community. LargeNetVis’s current version associates each community with a single category of

the structural taxonomy. In real-world networks, however, the same community may contain more than one type of structure. One such example is the community shown in Fig. 8(c), which contains both *star* and *tree*. We now intend to adapt the system to consider these cases.

Network characteristics. Some taxonomies’ categories may not be perceived depending on the network characteristics and dynamics. One such example is the absence of cliques in the bipartite network MovieLens (Sec. 5.1), a type of community that would also not be identified in unipartite networks with low number of triangles. We have evaluated LargeNetVis using both unipartite and bipartite networks to show its usefulness for analyzing networks with different characteristics. LargeNetVis is designed for networks with discrete timestamps, for which timeslicing strategies are suitable [62]. Also, the system adopts uniform (equally distributed) timeslices instead of non-uniform and data driven ones. In future work, we intend to investigate whether data driven timeslices would improve our results.

Information loss. LargeNetVis offers a set of sampling methods that can be used to discard nodes and edges from the analysis. It also focuses on the network communities, discarding inter-community edges and making it infeasible the analysis of the overall structure and evolution of the whole network (recall Sec. 4.4). Although evaluating the performance of sampling methods is outside the scope of our study, we now intend to measure how much information is lost — and how relevant they are — when inter-community edges are ignored.

Visual Improvements. Based on reviewers’ comments and participants’ feedback, we added some new features to the system after the user study (adaptive numerical information panel, varying link thickness in Global View, multilevel node-link diagram, edge bundling, and coordination between node-link and TAM). These features were not used by the participants and did not directly affect the reported results. We now intend to improve multiple community selections in Global View and thoroughly evaluate all these new features.

8 CONCLUSION

In this paper, we presented LargeNetVis, a web-based visual analytics system that assists experts and practitioners in analyzing small and large temporal networks, i.e., networks varying from a few nodes and timestamps to a few thousand of these elements. LargeNetVis enhances the analysis and pattern identification by leveraging the network community structure and three taxonomies, each accounting for a particular and relevant type of pattern (structural, temporal, and evolutionary). Given a temporal network and its node metadata, the system first divides the network into timeslices of equal length. Then, it detects network communities inside each timeslice. These communities are then categorized using three taxonomies from the literature. Finally, users can explore communities with characteristics of interest and their nodes and edges through the four linked views offered by the system, two focused on global level analysis and two focused on a local level. We validated LargeNetVis’s usefulness and effectiveness through two usage scenarios and a user study with 14 participants.

ACKNOWLEDGMENTS

This work was supported by grants #2020/10049-0, #2020/07200-9, and #2016/17078-0 from São Paulo Research Foundation (FAPESP), grant #E-26/201.424/2021 from Carlos Chagas Filho Foundation for Research Support of Rio de Janeiro State (FAPERJ), CNPq, CAPES, and by the School of Applied Mathematics at Fundação Getulio Vargas (FGV). The authors thank Luis R. Pereira for providing an optimized version of his source-code [50].

REFERENCES

- [1] B. Adhikari, Y. Zhang, A. Bharadwaj, and B. A. Prakash. Condensing temporal networks using propagation. In *Proceedings of the 2017 SIAM International Conference on Data Mining (SDM)*, pp. 417–425, 2017.
- [2] N. K. Ahmed, J. Neville, and R. Kompella. Network sampling: From static to streaming graphs. *ACM Transactions on Knowledge Discovery from Data*, 8(2):7:1–7:56, June 2013. doi: 10.1145/2601438

- [3] J.-w. Ahn, C. Plaisant, and B. Shneiderman. A task taxonomy for network evolution analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):365–376, 2014.
- [4] D. Archambault and H. C. Purchase. The "map" in the mental map: Experimental results in dynamic graph drawing. *Int. J. Hum.-Comput. Stud.*, 71(11):1044–1055, nov 2013.
- [5] A. Arleo, S. Miksch, and D. Archambault. A Multilevel Approach for Event-Based Dynamic Graph Drawing. In M. Agus, C. Garth, and A. Kerren, eds., *EuroVis 2021 - Short Papers*. The Eurographics Association, 2021. doi: 10.2312/evs.20211063
- [6] D. Auber, D. Archambault, R. Bourqui, M. Delest, J. Dubois, A. Lambert, P. Mary, M. Mathiaut, G. Melançon, B. Pinaud, B. Renoust, and J. Vallet. *Tulip 5*, pp. 3185–3212. Springer New York, New York, NY, 2018. doi: 10.1007/978-1-4939-7131-2_315
- [7] B. Bach, E. Pietriga, and J.-D. Fekete. Graphdiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2014.
- [8] S. K. Badam, Z. Liu, and N. Elmqvist. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):661–671, 2019.
- [9] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*, 2009.
- [10] V. Batagelj and A. Mrvar. Pajek-program for large network analysis. *Connect*, 21:47–57, 01 1998.
- [11] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The State of the Art in Visualizing Dynamic Graphs. In R. Borgo, R. Maciejewski, and I. Viola, eds., *EuroVis - STARS*. The Eurographics Association, 2014.
- [12] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum*, 36(1):133–159, 2016.
- [13] M. Behrisch, B. Bach, N. Henry Riche, T. Schreck, and J.-D. Fekete. Matrix reordering methods for table and network visualization. *Comput. Graph. Forum*, 35(3):693–716, 2016. doi: 10.1111/cgf.12935
- [14] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, October 2008.
- [15] U. Brandes. *Force-Directed Graph Drawing*, pp. 1–6. Springer US, Boston, MA, 2008.
- [16] M. Burch, C. Vehlow, F. Beck, S. Diehl, and D. Weiskopf. Parallel edge splatting for scalable dynamic graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2344–2353, 2011.
- [17] V. Cavaller. Dimensional taxonomy of data visualization: A proposal from communication sciences tackling complexity. *Frontiers in Research Metrics and Analytics*, 6, 2021.
- [18] A. Clementi, M. Di Ianni, G. Gambosi, E. Natale, and R. Silvestri. Distributed community detection in dynamic graphs. *Theoretical Computer Science*, 584:19–41, 2015. Special Issue on Structural Information and Communication Complexity. doi: 10.1016/j.tcs.2014.11.026
- [19] T. Crnovrsanin, Shilpika, S. Chandrasegaran, and K.-L. Ma. Staged animation strategies for online dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):539–549, 2021. doi: 10.1109/TVCG.2020.3030385
- [20] W. Cui, X. Wang, S. Liu, N. H. Riche, T. M. Madhyastha, K. L. Ma, and B. Guo. Let it flow: A static method for exploring dynamic graphs. In *2014 IEEE Pacific Visualization Symposium*, pp. 121–128, 2014. doi: 10.1109/PacificVis.2014.48
- [21] D3.js. D3.js - Data-Driven Documents. <https://d3js.org>. Accessed: 2022-03-09.
- [22] T. N. Dang, N. Pendar, and A. G. Forbes. Timearcs: Visualizing fluctuations in dynamic networks. *Computer Graphics Forum*, 35(3):61–69, 2016.
- [23] C. Dunne and B. Shneiderman. Motif simplification: Improving network visualization readability with fan, connector, and clique glyphs. CHI '13, p. 3247–3256. Association for Computing Machinery, New York, NY, USA, 2013.
- [24] E. Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Inc., USA, 2011.
- [25] M. T. Fischer, D. Arya, D. Streeb, D. Seebacher, D. A. Keim, and M. Worring. Visual analytics for temporal hypergraph model exploration. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):550–560, 2021.
- [26] S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007. doi: 10.1073/pnas.0605965104
- [27] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659(Supplement C):1 – 44, 2016.
- [28] K. Fu, T. Mao, Y. Wang, D. Lin, Y. Zhang, and X. Sun. Dyegovis: Visual exploration of dynamic ego-network evolution. *Applied Sciences*, 11(5), 2021. doi: 10.3390/app11052399
- [29] V. Gemmetto, A. Barrat, and C. Cattuto. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC infectious diseases*, 14(1):695, Dec. 2014.
- [30] M. Grinberg. *Flask web development: developing web applications with Python*. "O'Reilly Media, Inc.", 2018.
- [31] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using networkx. In G. Varoquaux, T. Vaught, and J. Millman, eds., *Proceedings of the 7th Python in Science Conference*, pp. 11 – 15. Pasadena, CA USA, 2008.
- [32] F. M. Harper and J. A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), dec 2015.
- [33] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012. Temporal Networks. doi: 10.1016/j.physrep.2012.03.001
- [34] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009. doi: 10.1111/j.1467-8659.2009.01450.x
- [35] Y. Hulovatyy, H. Chen, and T. Milenković. Exploring the structure and function of temporal networks with dynamic graphlets. *Bioinformatics*, 31(12):i171–i180, 06 2015. doi: 10.1093/bioinformatics/btv227
- [36] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 21(10):1160–1172, 2015.
- [37] A. Kyrola, G. Blelloch, and C. Guestrin. Graphchi: Large-scale graph computation on just a pc. In *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation, OSDI'12*, pp. 31–46. USENIX Association, Berkeley, CA, USA, 2012.
- [38] S. Lehmann. *Fundamental Structures in Temporal Communication Networks*, pp. 25–48. Springer International Publishing, Cham, 2019.
- [39] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05*, p. 177–187. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1081870.1081893
- [40] C. Li, G. Baciuc, and Y. Wang. Module-based visualization of large-scale graph network data. *Journal of Visualization*, 20(2):205–215, May 2017.
- [41] M. Li, J. Yang, F.-X. Wu, Y. Pan, and J. Wang. DyNetViewer: a Cytoscape app for dynamic network construction, analysis and visualization. *Bioinformatics*, 34(9):1597–1599, 12 2017.
- [42] Z. J. Lin, N. Cao, H. Tong, F. Wang, U. Kang, and D. H. Chau. Demonstrating interactive multi-resolution large graph exploration. *2013 IEEE 13th International Conference on Data Mining Workshops*, pp. 1097–1100, 2013.
- [43] C. D. G. Linhares, J. R. Ponciano, F. S. F. Pereira, L. E. C. Rocha, J. G. S. Paiva, and B. A. N. Travençolo. A scalable node ordering strategy based on community structure for enhanced temporal network visualization. *Computers & Graphics*, 84:185 – 198, 2019.
- [44] C. D. G. Linhares, B. A. N. Travençolo, J. G. S. Paiva, and L. E. C. Rocha. DyNetVis: a system for visualization of dynamic networks. In *Proceedings of the Symposium on Applied Computing, SAC '17*, pp. 187–194. ACM, Marrakech, Morocco, 2017.
- [45] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. Pregel: A system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, SIGMOD '10*, pp. 135–146. ACM, New York, NY, USA, 2010.
- [46] P. Mi, M. Sun, M. Masiane, Y. Cao, and C. North. Interactive graph layout of a million nodes. *Informatics*, 3(4), 2016.
- [47] NetworkX developers. NetworkX – network analysis in Python. <https://networkx.org/>. Accessed: 2022-03-09.
- [48] F. S. Pereira, S. d. Amo, and J. Gama. Detecting events in evolving social networks through node centrality analysis. In *Workshop on Large-scale Learning from Data Streams in Evolving Environments of ECML PKDD*, pp. 83–93, 2016.
- [49] L. R. Pereira, R. J. Lopes, and J. Louçã. A taxonomy of community

- lifecycle events in temporal networks. In *2019 4th World Conference on Complex Systems (WCCS)*, pp. 1–5, 2019. doi: 10.1109/ICoCS.2019.8930777
- [50] L. R. Pereira, R. J. Lopes, and J. Louçã. Community identity in a temporal network: A taxonomy proposal. *Ecological Complexity*, 45:100904, 2021.
- [51] J. R. Ponciano, C. D. Linhares, E. R. Faria, and B. A. Travençolo. An online and nonuniform timeslicing method for network visualisation. *Computers & Graphics*, 97:170–182, 2021.
- [52] L. E. C. Rocha, F. Liljeros, and P. Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Computational Biology*, 7(3):e1001109, 03 2011.
- [53] G. Rossetti and R. Cazabet. Community discovery in dynamic networks: A survey. *ACM Comput. Surv.*, 51(2), feb 2018.
- [54] R. A. Rossi and N. K. Ahmed. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [55] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [56] M. Rosvall and C. T. Bergstrom. Mapping change in large networks. *PLoS one*, 5(1):e8694–e8694, Jan 2010. 20111700[pmid].
- [57] B. Saket, P. Simonetto, and S. Kobourov. Group-Level Graph Visualization Taxonomy. In N. Elmqvist, M. Hlawitschka, and J. Kennedy, eds., *EuroVis - Short Papers*. The Eurographics Association, 2014.
- [58] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, Nov. 2003.
- [59] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Symposium on Visual Languages*, pp. 336–343. IEEE, USA, 1996.
- [60] B. Shneiderman, C. Plaisant, and B. W. Hesse. Improving healthcare with interactive visualization. *Computer*, 46(5):58–66, 2013.
- [61] P. Simonetto, D. Archambault, and S. Kobourov. Drawing dynamic graphs without timeslices, 2017. doi: 10.48550/ARXIV.1709.00372
- [62] P. Simonetto, D. Archambault, and S. Kobourov. Event-based dynamic graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 26(7):2373–2386, 2020. doi: 10.1109/TVCG.2018.2886901
- [63] N. Stanley, R. Kwitt, M. Niethammer, and P. J. Mucha. Compressing networks with super nodes. *Scientific Reports*, 8(1):10892, Jul 2018.
- [64] A. D. Stivala, J. H. Koskinen, D. A. Rolls, P. Wang, and G. L. Robins. Snowball sampling for estimating exponential random graph models for large networks. *Social Networks*, 47:167–188, 2016.
- [65] N. Sultanum, D. Singh, M. Brudno, and F. Chevalier. Doccurate: A curation-based approach for clinical text visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(01):142–151, jan 2019.
- [66] V. A. Traag, L. Waltman, and N. J. Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.
- [67] P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):1–13, 2021.
- [68] S. van den Elzen, D. Holten, J. Blaas, and J. van Wijk. Dynamic network visualization with extended massive sequence views. *IEEE Transactions on Visualization and Computer Graphics*, 20:1087–1099, 2014.
- [69] C. Vehlou, F. Beck, and D. Weiskopf. Visualizing group structures in graphs: A survey. *Comput. Graph. Forum*, 36(6):201–225, sep 2017.
- [70] P. Wang, J. C. S. Lui, B. Ribeiro, D. Towsley, J. Zhao, and X. Guan. Efficiently estimating motif statistics of large networks. *ACM Trans. Knowl. Discov. Data*, 9(2), sep 2014.
- [71] Y. Wang, D. Archambault, H. Haleem, T. Moeller, Y. Wu, and H. Qu. Nonuniform timeslicing of dynamic graphs based on visual complexity. In *2019 IEEE Visualization Conference (VIS)*, pp. 1–5. IEEE, 2019.
- [72] Y. Wang, G. Baci, and C. Li. A layout-based classification method for visualizing time-varying graphs. *ACM Trans. Knowl. Discov. Data*, 15(4), mar 2021.
- [73] C. Ware. *Information Visualization: Perception for Design*, p. 514. Interactive Technologies. Morgan Kaufmann, Boston, 2013.
- [74] Y. Yao. Collection and streaming of graph datasets. <http://www.eecs.wsu.edu/~yyao/StreamingGraphs.html>, jul 2012. Accessed: 2022-03-22.
- [75] V. Yoghoudjian, D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu. Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Visual Informatics*, 2(4):264 – 282, 2018.
- [76] V. Yoghoudjian, T. Dwyer, K. Klein, K. Marriott, and M. Wybrow. Graph thumbnails: Identifying and comparing multiple graphs at a glance. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3081–3095, dec 2018. doi: 10.1109/TVCG.2018.2790961
- [77] Y. Zhao, Y. She, W. Chen, Y. Lu, J. Xia, W. Chen, J. Liu, and F. Zhou. EOD edge sampling for visualizing dynamic network via massive sequence view. *IEEE Access*, 6:53006–53018, 2018.