

Combining clutter reduction methods for temporal network visualization

Jean R. Ponciano
Faculty of Computing, Federal
University of Uberlândia, Brazil
School of Applied Mathematics,
Getulio Vargas Foundation, Brazil
jean.ponciano@fgv.br

Claudio D. G. Linhares
Institute of Mathematics and
Computer Sciences, University of São
Paulo, Brazil
claudiodgl@usp.br

Luis E. C. Rocha
Department of Economics &
Department of Physics and
Astronomy, Ghent University,
Belgium
luis.rocha@ugent.be

Elaine R. Faria
Faculty of Computing, Federal
University of Uberlândia, Brazil
elaine@ufu.br

Bruno A. N. Travençolo
Faculty of Computing, Federal
University of Uberlândia, Brazil
travençolo@gmail.com

ABSTRACT

Temporal network visualization is a powerful tool that assists users in understanding network structure and dynamics. One of the most popular visual representations in this context is the *Massive Sequence View (MSV)*, a timeline-based layout that allows the identification of patterns, anomalies, and other structures from global to local scales. MSV may suffer from visual clutter when applied to real-world networks due to the large number of nodes, edges, and/or timestamps. To enhance the analysis, several clutter reduction methods have been proposed in the literature. No study, however, has evaluated different strategies that combine methods with respect to their effectiveness in reducing visual clutter while highlighting meaningful patterns. In this paper, we combine node ordering, edge sampling, and timeslicing methods to analyze how these combinations impact layout readability and pattern identification. We consider recent applications of MSV in the context of infection dynamics to study the effect of different combinations in the visualization layout. Through two case studies with real-world networks, we demonstrate the superiority of combining at least two high-quality methods in relation to the use of a single method. We also show that edge sampling should be used as a complementary strategy, always associated with a high-performance node ordering.

CCS CONCEPTS

• **Information systems** → *Data analytics*; • **Human-centered computing** → **Visualization**;

KEYWORDS

network visualization, visual clutter, network community, infection spread dynamics

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SAC '22, April 25–29, 2022, Virtual Event,

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-8713-2/22/04...\$15.00

https://doi.org/xx.xxx/xxx_x

ACM Reference Format:

Jean R. Ponciano, Claudio D. G. Linhares, Luis E. C. Rocha, Elaine R. Faria, and Bruno A. N. Travençolo. 2022. Combining clutter reduction methods for temporal network visualization. In *The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22)*, April 25–29, 2022, Virtual Event. . ACM, New York, NY, USA, 8 pages. https://doi.org/xx.xxx/xxx_x

1 INTRODUCTION

A temporal network is a modelling framework used to represent interactions over time involving instances from a domain [17]. This type of network contains three basic components (sometimes referred as dimensions [20]): node (which represents a domain's instance), edge (which represents an interaction between two instances), and time (when an interaction occurred – represented through edge's timestamp). Examples of temporal networks include email communications, face-to-face interactions, neuronal activity, and others [9, 11].

The visual exploration of temporal networks allows one to understand the network structure and evolution, thus improving the identification of patterns and anomalies and optimizing decision making [15, 21]. There are several visual representations (also called layouts) through which users can analyze a temporal network, for example animated node-link diagrams and matrix-based layouts [14], *Massive Sequence View (MSV)* [29], *Temporal Activity Map (TAM)* [16], and others. The large number of nodes, edges, and/or timestamps in real-world networks, however, may lead to a high level of visual clutter caused by the excessive overlap of nodes and edges [13], impairing layout readability and pattern identification.

The MSV layout is a timeline-based visualization commonly adopted for tasks involving the distribution of edges over time [14, 29]. There are several methods designed to reduce visual clutter in MSV representations, from node ordering (e.g., [15, 16, 29]) to edge sampling (e.g., [22, 31]) and timeslicing (also known as temporal resolution changing – e.g., [16, 20]). It is expected that an efficient edge sampling method maintains the most relevant edges in the MSV layout. However, such edges may result in overlap depending on the adopted node ordering. In the same way, edge lengths are reduced by efficient node ordering strategies, but the amount of visual information may continue too large depending on the temporal resolution and number of edges. The combination of such methods

affects the level of visual clutter, pattern identification, and decision making. No study, however, has evaluated different strategies that combine methods concerning their effectiveness in reducing visual clutter on MSV layouts while highlighting meaningful visual patterns.

In this paper, we combine methods that employ different approaches to reduce visual clutter on MSV layouts and analyze how these combinations impact layout readability. More specifically, we evaluate the joint employment of a node ordering method (based on time of appearance, node degree, or CNO [15]), an edge sampling method (without sampling, random sampling, or SEVis [22]), and a specific temporal resolution scale (the original network resolution or Adaptive Resolution [20]). As each of these methods leverages a particular network dimension (*node*, *edge*, or *time*), our study is motivated by the following questions: “*Is the visual analysis enhanced when using layouts produced by the combination of two methods (high-performance methods for two network dimensions and a naive approach for the other)?*” and “*Is the visual analysis enhanced when using layouts produced by the combination of three methods (high-performance methods for the three network dimensions)?*”.

The paper is organized as follows. Section 2 describes fundamental concepts and related work. Section 3 presents our methodology. Section 4 presents case studies considering two real-world networks with distinct characteristics. Section 5 discusses the results and limitations. Section 6 presents the conclusion and future work.

2 BACKGROUND AND RELATED WORK

To compare different strategies that combine methods for clutter reduction (hereafter referred as “combinations”), we rely on the recent application of MSV for the analysis of infection spread dynamics [12, 13]. With this in mind, this section first describes MSV in detail. Then, we introduce the *Susceptible-Infected (SI)* infection dynamics model used to simulate infection propagation in temporal networks and show how to adapt MSV for this context. Finally, we present existing methods for reducing clutter in MSV visualizations.

2.1 Massive Sequence View (MSV)

Several layouts have been proposed over the years to represent complex (non-temporal) networks, such as node-link diagrams and matrix-based layouts [8]. For temporal networks, one of the most recommended layouts is *Massive Sequence View (MSV)* [14]. In this layout, nodes are mapped into rows, timestamps are mapped into columns, and edges are represented as vertical straight lines linking two rows at a given column (see Figures 1(a,b)). Given the large number of nodes, edges, and/or timestamps in real-world networks, MSV may suffer from visual clutter, i.e., an excessive overlap of visual elements that impairs the visualization [29].

2.2 Susceptible-Infected (SI) infection dynamics

The temporal network characteristics, for example periods of idleness and bursts of interactions, make such networks useful to study epidemics [24]. As a matter of fact, several real-world temporal social networks have been proposed to analyze infection spread dynamics in different environments and situations (e.g., schools [7] and sexual encounters [26]).

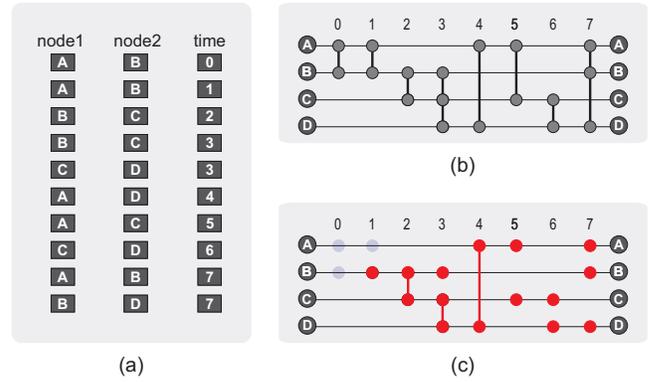


Figure 1: Visualizing temporal networks and infection spread dynamics: (a) List of edges; (b) Massive Sequence View (MSV); (c) Transmission path (TP) layout. Blue node: Susceptible. Red node: Infected. Patient zero: node B at timestamp 1.

We consider a fundamental infection dynamics model named *susceptible-infected (SI)* [1]. In this model, all nodes start susceptible, except for the patient zero that is a node chosen to be initially infected. As the network evolves, each interaction involving an infected node and a susceptible one has a probability p of infecting the susceptible node. The SI model is appropriate to study the worst-case scenarios of infection dynamics because once infected, the node never recovers [1, 24].

Given that temporal network visualization is a useful tool for analyzing infection spread dynamics, MSV can be adapted to highlight the transmission path (TP) over time [13, 23]. In this paper, we refer to this adapted MSV as TP layout. This representation shows only those edges through which the infection propagates, i.e., edges representing contacts between infected and susceptible nodes that resulted in new infections. Figure 1(c) illustrates a TP layout where the SI infection spread takes place on the network with $p = 1$. Susceptible nodes are represented with a different color and transparency. This layout facilitates identifying the infected nodes and the timings of new infections.

2.3 Methods for visual clutter reduction

We can categorize methods designed to reduce visual clutter in network visualization according to the network dimension they manipulate (i.e., *node*, *edge*, or *time*) [20]. Regarding the node dimension, there are methods to sample relevant nodes [4], to summarize the network (for example via super-nodes [28]), and to define proper node ordering in the layout [15, 16, 29]. Node ordering methods are useful in the context of MSV layouts because they reduce edge length, thus reducing overlaps and improving the perception of temporal patterns from global to local perspectives.

Besides naive strategies such as *Appearance* and *Degree*-based node ordering — which sort nodes according to the chronological order of connections and according to the ascending order of accumulated (in/out) degrees, respectively —, more elaborate strategies have been proposed. The *Optimized MSV* [29] is a hierarchical strategy that reduces clutter by minimizing both edge block overlap and standard deviation of the edges lengths. Another

strategy is *Recurrent Neighbors (RN)* [16], a recursive method that approximates nodes that are more often connected together. Not least, *Community-based node ordering (CNO)* [15] is a visual scalable and multi-level method that leverages existing node ordering and detection of network communities, i.e., groups of nodes that interact more often among themselves than between groups [5]. CNO is flexible and can employ different community detection algorithms, including the Louvain, a modularity-based method commonly adopted in the literature [2, 19].

Methods that belong to the edge dimension usually focus on removing less relevant edges, thus reducing visual information through edge sampling strategies. The *Edge Overlapping Degree (EOD)* [31], for example, is an edge sampling method designed for MSV layouts that considers overlaps between neighboring edges and their length as indicators of visual clutter. The *Streaming Edge Sampling for Network Visualization (SEVis)* [22] is an edge sampling method also suitable for streaming networks¹ and layout-agnostic. For each window of timestamps with size w_{size} over time, SEVis (i) incrementally computes the k most relevant nodes so far in the network using *Space-Saving* [18]; (ii) decides if a community detection using the window data must be performed and runs the detection if necessary; (iii) accepts the intra-community edges that have at least one relevant incident node.

Finally, regarding the time dimension, one may focus the analysis on relevant observation periods or change the network temporal resolution by grouping consecutive timestamps [16, 20, 27, 30]. This later strategy is particularly useful when dealing with temporally sparse networks and can be made through uniform timeslicing [20], i.e., by considering a global and static temporal resolution scale in which all timestamps of the network represent the same length of time (e.g., 1-hour or 1-day interval), or through non-uniform timeslicing, i.e., by having timestamps with different lengths. As an example, Wang et al. [30] proposed a non-uniform timeslicing method that changes the timestamp attribute of the network edges such that a balanced visual complexity (similar number of edges over time) is achieved. Likewise, Ponciano et al. [20] proposed a non-uniform timeslicing method (hereafter named *Adaptive Resolution (AR)*) that considers the local edge density to define the most suitable resolution scale for each window of timestamps with size w_{size} over time. AR is also suitable for streaming networks and relies on the forgetting mechanism Fading Sum [6] to discount old edges in the network. The greater the value of the fading factor (α) parameter, the more importance is given to older edges.

Some of the studies mentioned above combine two methods when generating MSV visualizations. The layouts used by the AR creators to analyze the method's performance, for example, relied on the Recurrent Neighbors node ordering [20]. CNO was similarly adopted when generating the MSV layouts to evaluate SEVis [22] and Optimized MSV was used in the evaluation of EOD [31]. To the best of our knowledge, however, no study has compared different strategies combining methods from two or three dimensions to analyze their effectiveness in terms of visual clutter reduction and pattern identification.

¹Streaming network is a type of network whose size is potentially unbounded and in which edges are continuously arriving in non-stationary distribution [20, 22].

3 METHODOLOGY

We have restricted our analysis to strategies combining a node ordering method (*Appearance, Degree, or CNO*), a timeslicing method (*original temporal resolution or AR*), and an edge sampling method (*SEVis, random, or None (original network)*). Each combination comprises one, and only one, method from each category. Except if stated otherwise, the order in which the methods of a combination are executed is defined by (i) timeslicing, (ii) node ordering, and then (iii) edge sampling. For a given temporal resolution, the tuple “ (x,y) ” will be used to refer to the application of the node ordering x followed by the edge sampling y .

After applying a combination on a given temporal network, we obtain a MSV layout that allows us to quantify the produced level of visual clutter and to identify eventual visible patterns. We consider two measures to compare the different combinations. The first measure is the *number of intersections* in the MSV layout, which refers to the sum of how many parts of each edge overlap other edges [15]. A part of an edge is defined as the region between two adjacent nodes in the layout. The smaller the number of intersections, the less visual clutter. The second is the *Kolmogorov-Smirnov* measure [22], which gives the distance (ks-d) between two empirical cumulative distribution functions. Its value varies from 0 (identical distributions) to 1 (totally different distributions). In our case, we use this measure to compare the edge distribution before and after sampling. The smaller the ks-d, the better the edge sampling.

Besides the quantitative evaluation, we also consider a qualitative visual analysis of each produced layout. We first analyze how different combinations impact visual clutter reduction and pattern identification using MSV. After that, we perform a visual exploration of pre-simulated infection spread dynamics through TP layouts generated by different combinations. We first take a temporal network and simulate epidemics using the SI model in this original network (original temporal resolution and without edge sampling). Thereafter, the combinations are executed considering all network data. Finally, we analyze how each combination affected the analysis of the transmission path and pattern identification. Examples of patterns in this context include the perception of who infected whom and when the infection happened, groups of infected nodes, and epidemic outbreak.

4 CASE STUDIES

This section presents case studies considering two real-world temporal networks with distinct characteristics: the *Museum* network [10] (a relatively small data set) and the *Sexual* network [25, 26] (a large data set). All layouts were produced with DyNetVis [12].

4.1 Methods configuration

To be consistent with previous research, we chose parameter values inline to previous studies [15, 20, 22].

- AR: fading factor equals to 0.99 ($\alpha = 0.99$) and window size of 50 ($w_{size} = 50$) or 100 timestamps ($w_{size} = 100$), depending on the analysis.
- CNO: *Louvain* [2] as network community detection method and *Recurrent Neighbors* [16] as community and node ordering strategy.

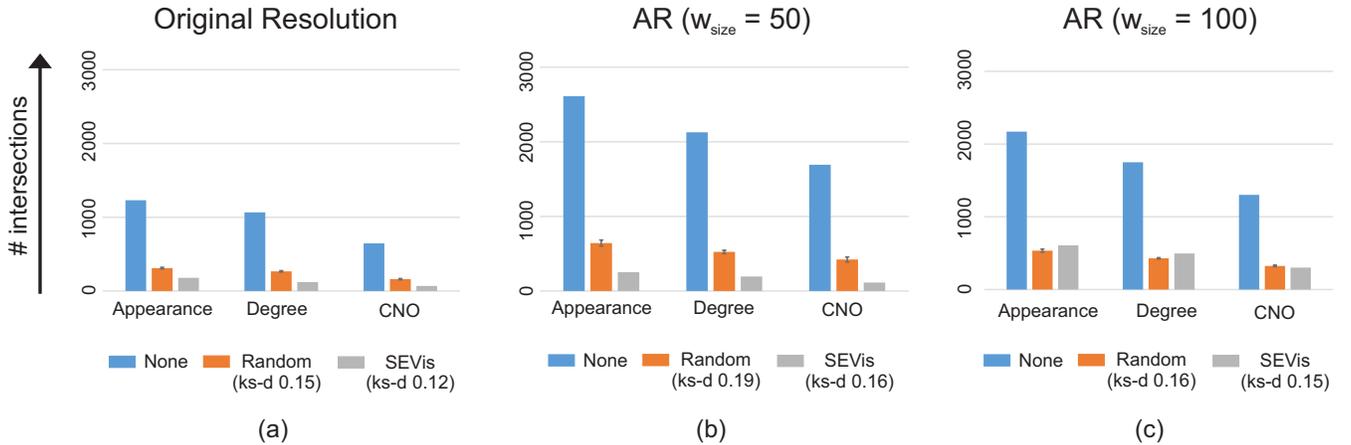


Figure 2: Number of intersections per combination for the *Museum* network. (a) Original temporal resolution. (b) AR ($w_{size} = 50$), which resulted in $avg_r = 2.44 \pm 2.29$, where avg_r refers to the average temporal resolution adopted. (c) AR ($w_{size} = 100$), which resulted in $avg_r = 2.86 \pm 1.61$. Each color refers to an edge sampling method (*None* means original network). Random sampling executed 10 times for each node ordering (black bars refer to standard deviations).

- SEVis: by default, window size of 100 timestamps ($w_{size} = 100$) and $k = 0.25|V|$ most relevant nodes considered by *Space-Saving* [18], where $|V|$ is the number of nodes in the network. Changes in these parameter values will be explicitly mentioned. The ratio to decide whether a new community detection must be executed, also using *Louvain*, is $t_r = 0.8$.
- Random edge sampling: accepts an edge with probability $p_a = 0.5$.

4.2 Museum network

The *Museum* network [10] is composed of data related to face-to-face proximity between people visiting the *Science Gallery* in Dublin, Ireland. The original network contains 72 nodes and 6,980 edges distributed in 1,312 timestamps. Each timestamp refers to a 20-second interval in the original temporal resolution (Res. 1). To analyze this network, we rely on combinations involving: original resolution and AR with $w_{size} = 50$ and $w_{size} = 100$ (temporal dimension); Appearance, Degree, and CNO node ordering (node dimension); None, Random, and SEVis edge sampling (edge dimension).

Figure 2 shows the number of intersections for each evaluated combination. The layouts adopting the original temporal resolution (Figure 2(a)) have a smaller number of intersections when compared with those from AR using $w_{size} = 50$ (Figure 2(b)) and $w_{size} = 100$ (Figure 2(c)). AR groups edges from consecutive timestamps, so the overall number of timestamps is reduced and, as a consequence, the number of intersections increases. Although the layouts from the original resolution have fewer intersections, they have more timestamps and therefore are horizontally large. This impairs the identification of global patterns, requires more screen space and breaks the user’s mental map [20].

For any timeslicing method/configuration in Figure 2, the combination *Appearance* for node ordering and *None* for edge sampling – i.e., combination (Appearance, None) –, generates the layouts with the highest levels of visual clutter (highest number of intersections), followed by (Degree, None) and (CNO, None) – which reaffirms

CNO high quality. When applying edge sampling strategies, the number of intersections is greatly reduced, as expected (see orange and gray bars in Figure 2). For the original temporal resolution (Figure 2(a)) and AR with $w_{size} = 50$ (Figure 2(b)), regardless of the node ordering, SEVis always outperforms random sampling, producing layouts that contain fewer intersections and better represent the original network (smaller ks-d). When observing layouts from AR with $w_{size} = 100$ (Figure 2(c)), however, SEVis produces more intersections than the random sampling for two node ordering methods (*Appearance* and *Degree*), i.e., SEVis maintained edges that are long when employing these ordering methods, thus increasing the number of intersections. When combined with CNO, SEVis outperforms the other combinations regardless of the temporal resolution being used: CNO reduces the number of intersections by repositioning nodes, which implies in less visual clutter and easier pattern identification, and SEVis improves the layout even more by discarding less relevant edges while preserving the characteristics of the network before sampling.

Figure 3 presents four MSV layouts generated by different combinations using AR with $w_{size} = 100$ (Figure 2(c)). This configuration was chosen because the application of SEVis along with AR with $w_{size} = 100$ resulted in more intersections than SEVis plus AR with $w_{size} = 50$ (gray bars in Figures 2(b,c)). Our goal is to evaluate SEVis under this circumstance, i.e., with such a high intersection level. The layout from (Appearance, None), shown in Figure 3(a), is dense and has too many intersections. When we apply SEVis (combination (Appearance, SEVis), Figure 3(b)), several edges of the network are discarded (1,749 out of 4,600), but the maintained edges remain too long, and so the layout readability is almost the same (although a few groups of relevant nodes are now highlighted – see red brackets). By repositioning nodes with CNO in the network without sampling (combination (CNO, None), Figure 3(c)), the analysis is greatly improved: groups of nodes are highlighted (red brackets) and a time interval that seemed to have a high level of activity with (Appearance, None) is, in fact, an interval with

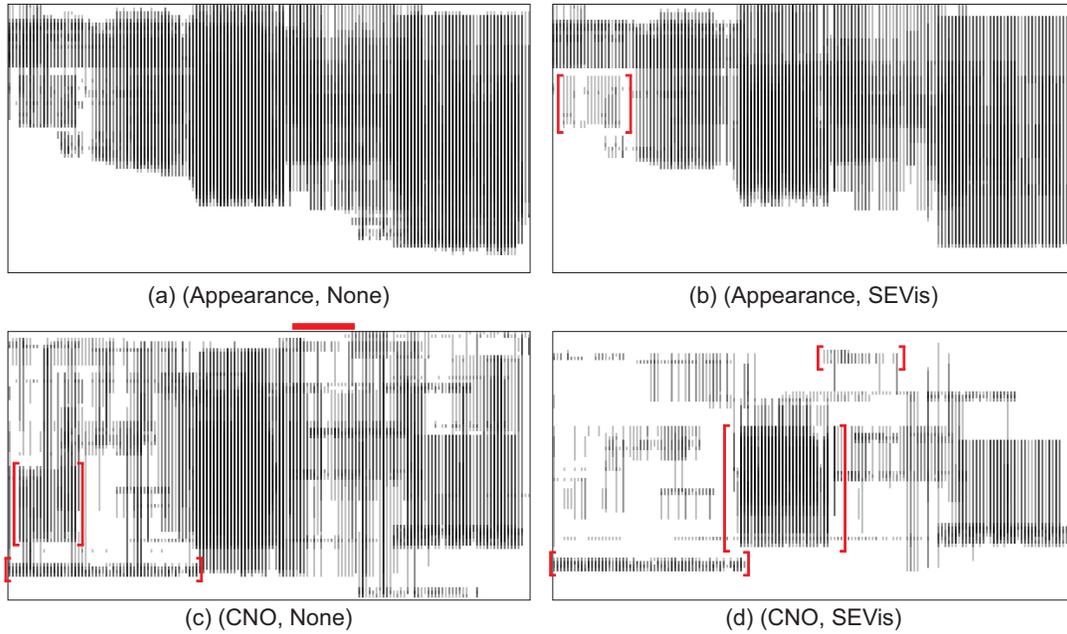


Figure 3: MSV layouts generated by different combinations for a portion of the *Museum* network using AR with $w_{size} = 100$. (a) Combination (Appearance, None). (b) (Appearance, SEVis). (c) (CNO, None). (d) (CNO, SEVis). Time interval: from $t = 100$ to $t = 769$ in the original resolution. Red bars and red brackets indicate existing patterns.

low activity (see the red bar in Figure 3(c) and the respective time interval in Figure 3(a)). Combination (CNO, SEVis), shown in Figure 3(d), produces a cleaner layout and allows quick identification of relevant groups of nodes (see red brackets).

The impact of different combinations in the TP layout is shown in Figure 4. For this evaluation, we first simulated a SI infection spread with $p = 0.1$. The patient zero is the node with the highest degree in the aggregated network (i.e., when considering all timestamps at once) in the first timestamp it appears. By comparing combination (Appearance, None) from the original temporal resolution (Figure 4(a)) with the combination (Appearance, None) from AR with $w_{size} = 100$ (Figure 4(b)), we see that AR maintains the characteristics of the edge distribution from the original resolution while reducing the number of timestamps (1,312 vs 569 for representing the entire network). This leads, for instance, to faster identification of the epidemic outbreak and other high activity periods (see black bars in Figure 4(b)).

By adopting AR with $w_{size} = 100$ and moving from (Appearance, None) to (CNO, None), as illustrated in Figures 4(b,c), the identification of groups of nodes is now possible (see arrows in Figure 4(c)). These groups, that are obtained by CNO, correspond to network communities. Since a community may represent, e.g., household members or people in workplaces, one may be interested in analyzing intra-community infection spread [3]. When applying SEVis (combination (CNO, SEVis), Figure 4(d)), only the most relevant nodes and edges are maintained. In this case, a CNO community may continue the same (see arrow in Figure 4(d)), completely disappear (see ‘X’ symbol), or change (see brackets). If we move from (Appearance, None) to (Appearance, SEVis) instead of

to (CNO, None), as shown in Figures 4(b,e), the layout becomes cleaner but the identification of groups is infeasible. Only when running CNO, such identification is possible (Figure 4(f)). Note that the node positioning is different when comparing Figures 4(d,f). After sampling, the set of edges considered by CNO is a subset of the original one; different network communities are detected and, therefore, a different node ordering is produced.

Overall, the combination of AR along with CNO (without SEVis) improved pattern identification in this network (Figure 4(c)). The application of SEVis without CNO, on the other hand, did not present the same effectiveness, even when combined with AR (Figure 4(e)). In summary, there are pairwise combinations that enhance the visual analysis, but not all of them. Not least, the combination of all three methods (AR, CNO, and SEVis) improved layout readability regardless of the order in which the node ordering and the edge sampling methods were employed (Figures 4(d,f)).

4.3 Sexual network

The *Sexual* network [25, 26] is composed of sexual encounters between sex-workers and their clients. The data was collected through posts in a public online forum. Each post evaluates the encounter, so each one comprehends an edge that connects the involved individuals (nodes) in the network. There are 12,157 nodes and 34,060 edges distributed in 1,000 timestamps.

Figure 5 shows different combinations (different TP layouts) for the *Sexual* network. For this evaluation, we applied the SI model in the original network with infection probability $p = 0.25$. The patient zero is an arbitrary node in the first timestamp it appears. All analyses with SEVis adopt $k = 0.005|V|$, i.e., SEVis considers

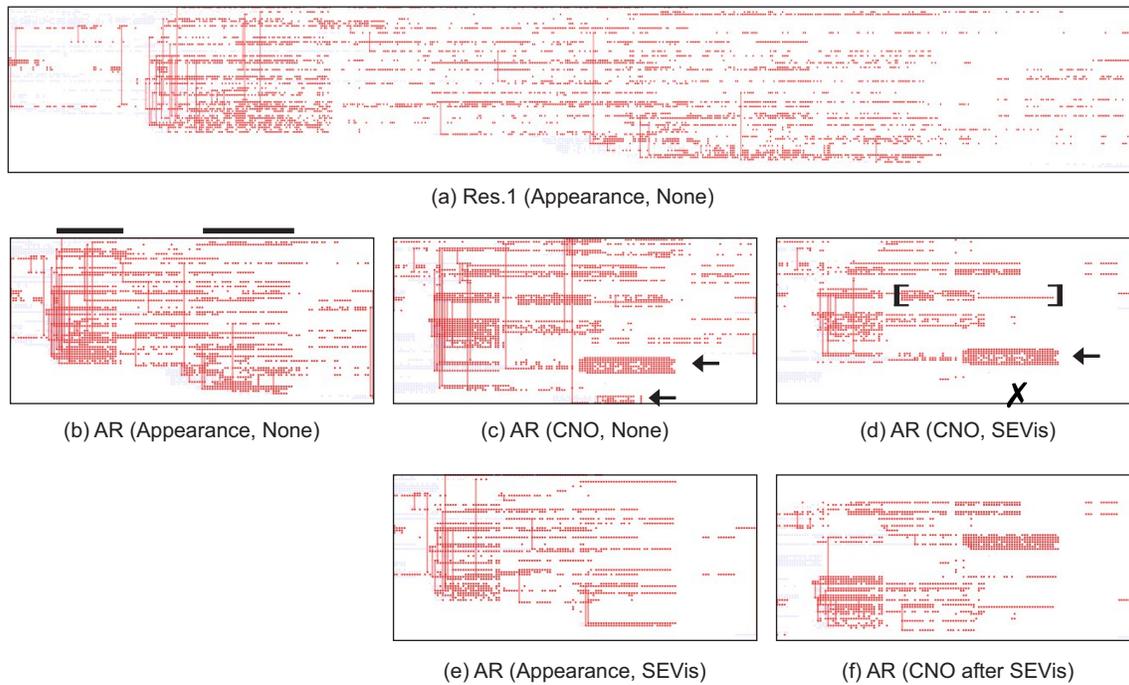


Figure 4: Different combinations in the TP layout using the *Museum* network. (a) Original resolution, combination (Appearance, None). (b-f) AR with $w_{size} = 100$. (b) (Appearance, None). (c) (CNO, None). (d) (CNO, SEVis). (e) (Appearance, SEVis). (f) CNO applied in the layout from (e), i.e., CNO after SEVis. First evaluation: sequence of combinations following sub-figures (a) → (b) → (c) → (d). Second evaluation: sub-figures (a) → (b) → (e) → (f). Time interval: from $t = 220$ to $t = 870$ in the original resolution. Black icons (bars, arrows, brackets, and 'X') highlight existing patterns.

all edges incident to the top-60 more relevant (frequent) nodes. In the layouts, the intensity of the edge color is related to the number of overlapping edges and intersections (the darker the color, the higher the number of overlaps and intersections). By comparing (Appearance, None) with (CNO, None) in the original resolution (Figures 5(a,b)), one may see that CNO greatly reduces the number of edge overlaps (lighter colors). The identification of patterns in the layout, however, remains difficult. Even when applying (CNO, SEVis), global pattern identification is not optimized (Figure 5(c)).

Because of the network density, AR with $w_{size} = 50$ and $\alpha = 0.99$ redistributed all edges in 96 timestamps. Considering that the first w_{size} timestamps used in the AR computation correspond to the method's cold start [20], 950 timestamps of the original resolution were converted in only 46 timestamps in the adaptive resolution. The layout produced by the combination (Appearance, None) considering this resolution is shown in Figure 5(d). Note that approximately the first half of the layout (left-middle) contains too few edges (cold start), while the second half (middle-right) contains all the other edges in a high edge overlap level (dark colors). So many overlaps involving long edges lead to a number of intersections so elevated that the visual analysis becomes unfeasible without complementary strategies that promote further edge reduction or removal (e.g., CNO along with SEVis).

For the adopted adaptive resolution scale, neither (Appearance, SEVis) nor (CNO, None) improves the overall layout readability (Figures 5(e,f)). Only when combining all three methods (AR plus

(CNO, SEVis)), particular groups of nodes and edges are revealed. Given the new number of timestamps (96), running SEVis with $w_{size} = 100$ means that the entire network is considered as a single window of timestamps. In this case, since both CNO and SEVis used Louvain for community detection, all CNO inter-community edges are discarded by SEVis. In addition, SEVis also discards those CNO intra-community edges that involve non-relevant nodes. In a more detailed perspective, the analysis is analogous to the one from Figures 4(c,d).

The groups of nodes and edges that were maintained in the TP layout after employing AR plus (CNO, SEVis), Figure 5(g), may be further analyzed using interactive tools, such as *zoom* and *pan* (Figure 5(h)). These tools would allow to explore, e.g., nodes inside a particular group or perform cross-comparison between groups. Groups of nodes considered as non-relevant by SEVis are discarded and therefore not visible (see 'X' symbols in Figure 5(h)). When applying SEVis using smaller windows, there is no match between the communities detected by CNO and SEVis, so edges maintained by SEVis may connect nodes positioned far from each other, thus increasing the number of intersections and polluting the layout. Figure 5(i) shows the TP layout produced by AR, CNO and SEVis using $w_{size} = 40$. Although small groups may be lost due to edge overlap (in comparison with Figure 5(g)), it is now possible to identify relevant connections between the visible groups, which facilitates contact tracing and cross-comparisons. Note that the mentioned interactive tools would benefit this analysis as well.

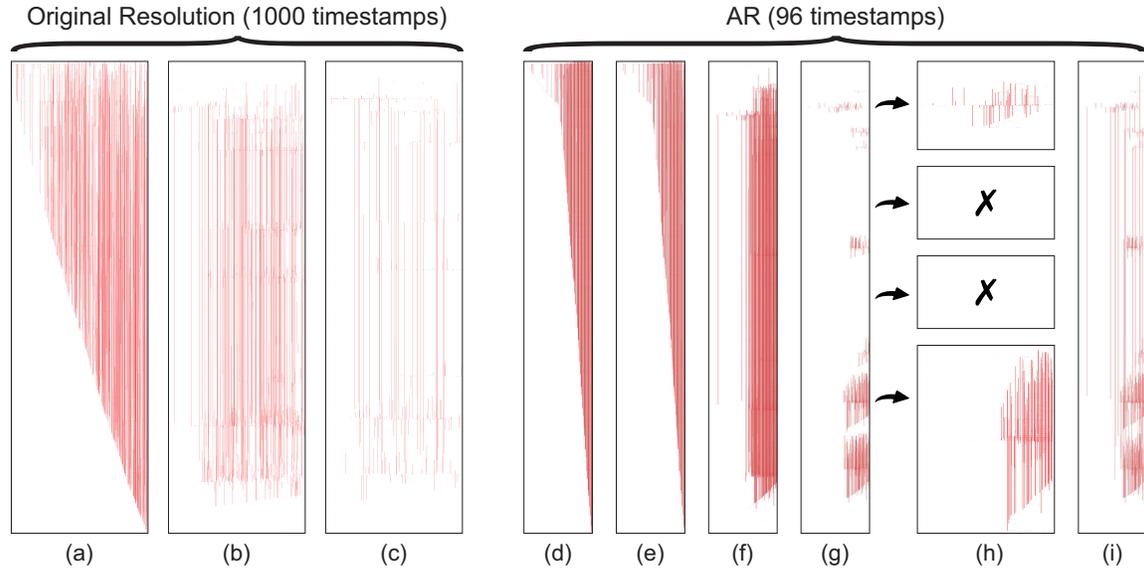


Figure 5: Impact of different combinations in an epidemics visual analysis using the *Sexual* network. (a-c) Original temporal resolution (1,000 timestamps). (a) (Appearance, None). (b) (CNO, None). (c) (CNO, SEVis $w_{size} = 100$). (d-g) AR with $w_{size} = 50$ (96 timestamps), which resulted in $avg_r = 24.52 \pm 5.55$. (d) (Appearance, None). (e) (Appearance, SEVis $w_{size} = 100$). (f) (CNO, None). (g) (CNO, SEVis $w_{size} = 100$). (h) Zooming in particular groups from (g). (i) (CNO, SEVis $w_{size} = 40$).

5 DISCUSSION AND LIMITATIONS

We have considered two real-world networks with distinct characteristics. The *Museum* network is a relatively small data set with 72 nodes and 6,980 edges (5.32 edges per timestamp on average) that contains more timestamps than nodes. The *Sexual* network is a large data set with 12,157 nodes and 34,060 edges (34.06 edges per timestamp on average) that contains more nodes than timestamps.

Testing different combinations and parameters for the evaluated methods would be too demanding. For the same reason, other tasks, combinations involving other methods, and other application sequences (e.g., executing SEVis first and then running AR) were not tested. Even though we cannot generalize our findings, we are now capable of responding to the questions presented at the beginning of this paper for the performed evaluation:

“Is the visual analysis enhanced when using layouts produced by the combination of two methods (high-performance methods for two network dimensions and a naive approach for the other)?”

Answer: It depends on the network characteristics and the combination/parameter values being considered. For the *Museum* network, the combination of AR and CNO (without SEVis) greatly improved layout readability (Figure 4(c)). The combination of AR and SEVis (without CNO and therefore using *Appearance*), on the other hand, did not present the same effectiveness (Figure 4(e)). In essence, an edge sampling method should be considered as a complementary strategy, always associated with the employment of a high performance node ordering.

“Is the visual analysis enhanced when using layouts produced by the combination of three methods (high-performance methods for the three network dimensions)?”

Answer: Yes. To a greater or lesser extent, this happened for both networks. When analyzing the *Sexual* network, only the combination of all three methods (AR, CNO, and SEVis) highlighted existing patterns (Figures 5(g-i)). The evaluated pairwise combinations were not effective (Figures 5(c,e,f)).

Lastly, contrary to CNO, which requires the aggregated network, SEVis and AR use only the more recent data through sliding windows over time. In combinations involving both CNO and SEVis, the “gap” that exists between the static CNO node ordering and SEVis dynamic execution may result in discrepancies in the network community detection, leading to long edges in the layout (recall Figure 5(i)). Running both SEVis and CNO in such a dynamic manner would naturally attenuate this situation as CNO would change the node ordering at each window [22].

6 CONCLUSION

Various methods have been proposed to reduce clutter in temporal network visualization. In particular, the *Massive Sequence View (MSV)* layout has attracted the attention of researchers who want to enhance the network visual analysis. In recent years, MSV has also been adapted and used for the analysis of dynamic processes that take place on the network, e.g., the simulation of infection spreads.

This paper compared different combinations of clutter reduction methods developed for, or applied to, MSV. We have considered in our study node ordering, edge sampling, and timeslicing methods, thus covering all three network dimensions (namely *node*, *edge*, and *time*). Through two case studies where we simulated infection spread dynamics in real-world networks with distinct sizes and characteristics, we showed how the combination of different methods affected the layout and, consequently, pattern identification

and decision making. We were able to assess the superiority of combinations involving two or more high-performance methods compared with the adoption of a single method and showed that edge sampling should be used as a complementary strategy, always associated with a high-performance node ordering. As future work, we intend to evaluate the methods under different parameters and more methods and layouts (e.g., animated node-link diagrams).

ACKNOWLEDGMENTS

This research was supported by the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq [grant number 456855/2014-9], Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES Print - Grant number 88881.311513/2018-01), and São Paulo Research Foundation (FAPESP - Grants number 2020/10049-0, 2020/07200-9, 2016/17078-0).

REFERENCES

- [1] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. 2008. *Dynamical Processes on Complex Networks*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791383>
- [2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (October 2008), P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>
- [3] Stefano Bonaccorsi, Stefania Ottaviano, Francesco De Pellegrini, Annalisa Socievole, and Piet Van Mieghem. 2014. Epidemic outbreaks in two-scale community networks. *Physical Review E* 90, 1 (2014), 012810. <https://doi.org/10.1103/PhysRevE.90.012810>
- [4] Flavio Chiericetti, Anirban Dasgupta, Ravi Kumar, Silvio Lattanzi, and Tamás Sarlós. 2016. On Sampling Nodes in a Network. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 471–481. <https://doi.org/10.1145/2872427.2883045>
- [5] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3 (2010), 75 – 174. <https://doi.org/10.1016/j.physrep.2009.11.002>
- [6] Joao Gama. 2010. *Knowledge Discovery from Data Streams* (1st ed.). Chapman & Hall/CRC.
- [7] Valerio Gemmetto, Alain Barrat, and Ciro Cattuto. 2014. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC infectious diseases* 14, 1 (Dec. 2014), 695. <https://doi.org/10.1186/PREACCEPT-6851518521414365>
- [8] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. 2005. On the Readability of Graphs Using Node-Link and Matrix-Based Representations: A Controlled Experiment and Statistical Analysis. *Information Visualization* 4 (2005), 114 – 135.
- [9] Petter Holme and Jari Saramäki. 2012. Temporal Networks. *Physics Reports* 519, 3 (October 2012), 97–125.
- [10] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. 2011. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology* 271 (2011), 166–180. <https://doi.org/10.1016/j.jtbi.2010.11.033>
- [11] P. S. Keila and D. B. Skillicorn. 2005. Structure in the Enron Email Dataset. *Computational & Mathematical Organization Theory* 11, 3 (Oct. 2005), 183–199. <https://doi.org/10.1007/s10588-005-5379-y>
- [12] Claudio DG Linhares, Jean R Ponciano, Jose Gustavo S Paiva, Luis EC Rocha, and Bruno AN Travençolo. 2020. DyNetVis-An interactive software to visualize structure and epidemics on temporal networks. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 933–936.
- [13] Claudio D. G. Linhares, Jean R. Ponciano, Jose Gustavo S. Paiva, Bruno A. N. Travençolo, and Luis E. C. Rocha. 2019. *Visualisation of Structure and Processes on Temporal Networks*. Springer International Publishing, Cham, 83–105. https://doi.org/10.1007/978-3-030-23495-9_5
- [14] Claudio D. G. Linhares, Jean R. Ponciano, Jose Gustavo S. Paiva, Bruno A. N. Travençolo, and Luis E. C. Rocha. 2021. A comparative analysis for visualizing the temporal evolution of contact networks: a user study. *Journal of Visualization* (13 Jun 2021). <https://doi.org/10.1007/s12650-021-00759-x>
- [15] Claudio D. G. Linhares, Jean R. Ponciano, Fabiola S. F. Pereira, Luis E. C. Rocha, Jose Gustavo S. Paiva, and Bruno A. N. Travençolo. 2019. A scalable node ordering strategy based on community structure for enhanced temporal network visualization. *Computers & Graphics* 84 (2019), 185 – 198. <https://doi.org/10.1016/j.cag.2019.08.006>
- [16] Claudio D. G. Linhares, Bruno A. N. Travençolo, Jose Gustavo S. Paiva, and Luis E. C. Rocha. 2017. DyNetVis: A System for Visualization of Dynamic Networks. In *Proceedings of the Symposium on Applied Computing (SAC '17)*. ACM, Marrakech, Morocco, 187–194. <https://doi.org/10.1145/3019612.3019686>
- [17] N. Masuda and R. Lambiotte. 2016. *A Guide to Temporal Networks*. World Scientific. <https://doi.org/10.1142/q0033>
- [18] Ahmed Metwally, Divyakant Agrawal, and Amr El Abbadi. 2005. Efficient computation of frequent and top-k elements in data streams. In *International Conference on Database Theory*. Springer, 398–412. https://doi.org/10.1007/978-3-540-30570-5_27
- [19] J. Mothe, K. Mkhitarian, and M. Haroutunian. 2017. Community detection: Comparison of state of the art algorithms. In *2017 Computer Science and Information Technologies (CSIT)*. IEEE, 125–129. <https://doi.org/10.1109/CSITechnol.2017.8312155>
- [20] Jean R. Ponciano, Claudio D.G. Linhares, Elaine R. Faria, and Bruno A.N. Travençolo. 2021. An online and nonuniform timeslicing method for network visualisation. *Computers & Graphics* 97 (2021), 170–182. <https://doi.org/10.1016/j.cag.2021.04.006>
- [21] Jean R. Ponciano, Claudio D. G. Linhares, Sara L. Melo, Luciano V. Lima, and Bruno A. N. Travençolo. 2020. Visual Analysis of Contact Patterns in School Environments. *Informatics in Education* 19, 3 (2020), 455–472. <https://doi.org/10.15388/infedu.2020.20>
- [22] Jean R. Ponciano, Claudio D. G. Linhares, Luis E. C. Rocha, Elaine R. Faria, and Bruno A. N. Travençolo. 2021. A streaming edge sampling method for network visualization. *Knowledge and Information Systems* (29 Apr 2021). <https://doi.org/10.1007/s10115-021-01571-7>
- [23] Jean R. Ponciano, Gabriel P. Vezono, and Claudio D.G. Linhares. 2021. Simulating and visualizing infection spread dynamics with temporal networks. In *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados*. SBC, Porto Alegre, RS, Brasil, 37–48. <https://doi.org/10.5753/sbbd.2021.17864>
- [24] L. E. C. Rocha and V. D. Blondel. 2013. Bursts of vertex activation and epidemics in evolving networks. *PLoS Computational Biology* 9 (2013), 1–9. <https://doi.org/10.1371/journal.pcbi.1002974>
- [25] Luis E. C. Rocha, Fredrik Liljeros, and Petter Holme. 2010. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proceedings of the National Academy of Sciences* 107, 13 (2010), 5706–5711. <https://doi.org/10.1073/pnas.0914080107> arXiv:<https://www.pnas.org/content/107/13/5706.full.pdf>
- [26] Luis E. C. Rocha, Fredrik Liljeros, and Petter Holme. 2011. Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLoS Comput Biol* 7, 3 (03 2011), e1001109. <https://doi.org/10.1371/journal.pcbi.1001109>
- [27] Luis E. C. Rocha, Naoki Masuda, and Petter Holme. 2017. Sampling of temporal networks: Methods and biases. *Phys. Rev. E* 96 (Nov 2017), 052302. Issue 5. <https://doi.org/10.1103/PhysRevE.96.052302>
- [28] Natalie Stanley, Roland Kwitt, Marc Niethammer, and Peter J. Mucha. 2018. Compressing Networks with Super Nodes. *Scientific Reports* 8, 1 (18 Jul 2018), 10892. <https://doi.org/10.1038/s41598-018-29174-3>
- [29] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk. 2013. Reordering Massive Sequence Views: Enabling temporal and structural analysis of dynamic networks. In *2013 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 33–40. <https://doi.org/10.1109/PacificVis.2013.659612>
- [30] Y. Wang, D. Archambault, H. Haleem, T. Moeller, Y. Wu, and H. Qu. 2019. Nonuniform Timeslicing of Dynamic Graphs Based on Visual Complexity. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 1–5. <https://doi.org/10.1109/VISUAL.2019.8933748>
- [31] Y. Zhao, Y. She, W. Chen, Y. Lu, J. Xia, W. Chen, J. Liu, and F. Zhou. 2018. EOD Edge Sampling for Visualizing Dynamic Network via Massive Sequence View. *IEEE Access* 6 (2018), 53006–53018. <https://doi.org/10.1109/ACCESS.2018.2870684>