Comparative Study and Evaluation of Hybrid Visualizations of Graphs

Emilio Di Giacomo ⦿, Walter Didimo ⦿, Giuseppe Liotta ⦿ Fabrizio Montecchiani ⦿, Alessandra Tappini ⦿

Abstract—Hybrid visualizations combine different metaphors into a single network layout, in order to help humans in finding the "right way" of displaying the different portions of the network, especially when it is globally sparse and locally dense. We investigate hybrid visualizations in two complementary directions: (i) On the one hand, we evaluate the effectiveness of different hybrid visualization models through a comparative user study; (ii) On the other hand, we estimate the usefulness of an interactive visualization that integrates all the considered hybrid models together. The results of our study provide some hints about the usefulness of the different hybrid visualizations for specific tasks of analysis and indicates that integrating different hybrid models into a single visualization may offer a valuable tool of analysis.

Index Terms—Network Visualization, Hybrid Visualizations, Evaluation, User Study

✦

1 INTRODUCTION

G RAPHS are widely used to model networked data sets in a variety of application domains. Their visualization amplifies human cognition and accelerates knowledge extraction processes. Choosing which layout metaphor is more suitable for a pictorial representation of a graph is a central problem, and the heterogeneous connectivity structure of many real-world networks makes it often difficult to find a clear and unanimous answer. *Hybrid visualizations* combine different metaphors into a single network layout, in order to help humans in finding the "right way" of displaying the different portions of the network.

In this scenario, particular interest is devoted to those real-world networks that exhibit a double structural nature: they are globally sparse but locally dense, i.e., they contain *clusters* of highly connected nodes (also called *communities* in social network analysis) that are loosely connected to each other (see, e.g., [2], [3], [4]). Examples include social and financial networks [5], [6], [7], [8], as well as biological and information networks [9], [10]. The visualization of networks of this type through a unique node-link diagram is sometimes unsatisfactory, due to the visual clutter caused by the high number of edges in the dense portions of the network (see, e.g., Fig. 1(a)). One of the seminal ideas to overcome this problem is the NODETRIX hybrid visualization model introduced by Henry, Fekete, and McGuffin [11]. It adopts a node-link diagram to represent the (sparse) global structure of the network, and a matrix representation for denser subgraphs identified and selected by the user (Fig. 1(c)). After the introduction of NODETRIX, hybrid visualizations have become an emerging topic and inspired an array of both theoretical and application results [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26].

A preliminary version of this paper appeared in the proceedings of the 29th International Symposium on Graph Drawing and Network Visualization [1]; the article received the best paper award and the authors have been invited to submit an extended version to IEEE TVCG.

- *All authors work at the Department of Engineering, University of Perugia, Italy. E-mail: {emilio.digiacomo, walter.didimo, giuseppe.liotta, fabrizio.montecchiani, alessandra.tappini}@unipg.it.*

Contribution. In this paper we investigate hybrid visualizations in two complementary directions:

- We compare different hybrid visualization models through a user study that addresses two main research questions: RQ1 − "Are hybrid visualizations more effective than node-link diagrams for the visual analysis of clustered networks?"; RQ2 − "When considering specific tasks of analysis, are there differences in terms of response time or accuracy among different hybrid visualization models?". The study focuses on three models designed to work on similar types of networks: The aforementioned NODETRIX model [11]; the CHORDLINK model [15], which represents clusters as chord diagrams instead of adjacency matrices (Fig. 1(b)); and the RCI-NODETRIX model [25], a variant of NODETRIX that allows independent orderings for the matrix rows and columns to reduce inter-cluster edge crossings (Fig. 1(d)).

- We estimate the usefulness of an interactive visualization that integrates all the above hybrid models together, thus allowing the user to choose the preferred way of representing different portions of a network (Fig. 2). The evaluation is done by means of the ICE-T methodology, introduced by Wall et al. [27], which enables a quantitative measurement of the "value" of a visualization, within a framework defined by Stasko [28].

To the best of our knowledge, our study is the first that addresses research question RQ1, and that considers RQ2 for hybrid visualizations that adopt different styles to represent clusters. Our work is also motivated by open questions in [15], [25]. Namely, [15] suggests to perform a user study to compare CHORDLINK and other hybrid visualizations, and [25] asks what is the impact of reducing crossings between inter-cluster edges at the expenses of independent row/column orderings in NODETRIX. The results of our study provide some hints about the usefulness of the different hybrid visualizations for specific tasks of analysis. Also, the assessment through the ICE-T methodol-

(a) NODELINK

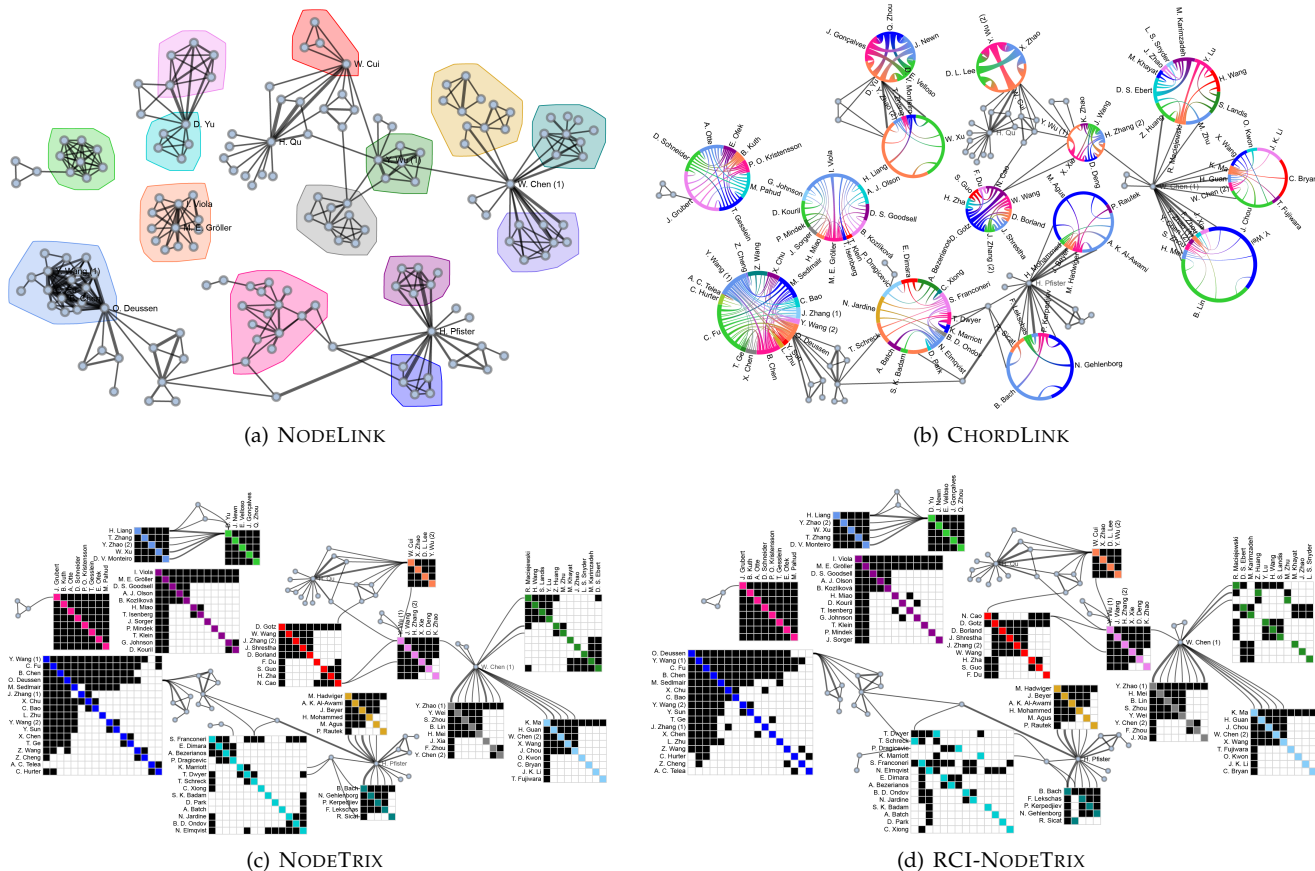(b) CHORDLINK

(c) NODETRIX

(d) RCI-NODETRIX

Fig. 1. The same clustered network with our four visualization models.

ogy indicates that integrating different hybrid models into a single visualization may offer a valuable tool of analysis. Nonetheless, as every other cognitive experimental work, our study has several limitations, which we clearly discuss and which trace the perimeter of our results.

The paper is structured as follows. Section 2 briefly surveys the scientific literature related to our work. Section 3 explains in detail the design of our user study, while Section 4 analyzes the corresponding findings. Section 5 reports the results of the ICE-T methodology. Conclusions and future research directions are discussed in Section 6. All the experimental data are available at http://mozart.diei. unipg.it/tappini/hybridUserStudy/data-extended.html.

2 RELATED WORK

Our paper pertains two main research topics: Hybrid graph visualizations and user studies in graph drawing and network visualization. We briefly survey these two topics, focusing on those aspects that are more related to our research.

Hybrid graph visualizations. Early works propose hybrid models that mix Euler/Venn diagrams and Jordan arcs to represent different types of relationships between sets of objects [29], [30]. Similar drawing styles are used to represent compound graphs, where the nodes are hierarchically grouped into clusters and edges can connect clusters other than nodes (see [31], [32], [33] for surveys on the subject). Hybrid visualizations that combine node-link and treemaps are also studied [34], [35], [36].

Our focus is on hybrid graph representations that mix different visual metaphors to visually convey both the global structure of a sparse network and its locally dense subgraphs. In this context, the NODETRIX model introduced by Henry, Fekete, and McGuffin [11] for social network analysis is one of the most cited contributions of the InfoVis conference [37]; this model is implemented in a system where the user can select (dense) portions of a node-link diagram to be represented as adjacency matrices. NODETRIX is also exploited to analyze other real-world graphs, such as ontology graphs [16] and brain networks [26].

Along the same research trajectory, Angori et al. [15] propose an alternative model, called CHORDLINK. Similarly to NODETRIX, this model is designed to work in a system where users can visually identify and select clusters on an initial node-link diagram; differently from NODETRIX, the selected cluster regions are represented as chord diagrams. CHORDLINK aims to represent all edges as geometric links and to preserve the layout outside clusters by possibly duplicating some nodes within a cluster; however, each node can appear in at most one cluster, as for NODETRIX.

The user study presented in our paper compares NODE-TRIX and CHORDLINK, as they are conceived to work on networks with similar structure and within systems with similar characteristics. Additionally, it considers the RCI-NODETRIX model [25], a variant of NODETRIX that allows independent orderings of the rows and columns in a matrix, to possibly reduce crossings between inter-cluster edges.
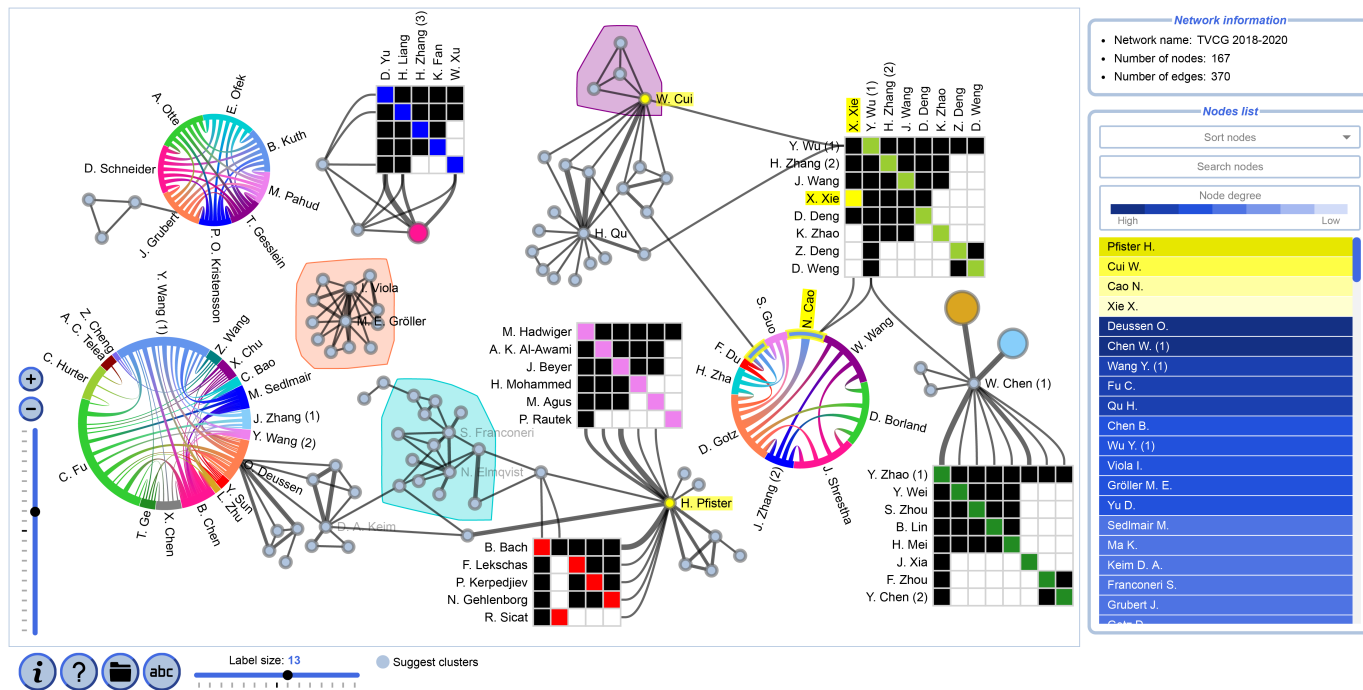
Fig. 2. Network visualization that integrates all different hybrid models considered in this article. The network represents the co-authorships on TVCG articles in the years 2018–2020. To the right of the drawing canvas, the list of node labels is shown, ordered by decreasing values of node degrees; four nodes in this list are selected and the corresponding nodes are highlighted in the network layout.

In the context of social network analysis, Henry, Bezerianos, and Fekete [22] investigate a variant of NODETRIX that considers "overlapping clusters", i.e., where a node can occur in multiple clusters at the same time. They conclude that this kind of node duplication may help in the execution of community-related tasks, but sometimes interferes with other graph readability tasks.

Batagelj et al. [17] propose a system where the user can choose to represent each cluster according to a desired drawing style. Differently from NODETRIX and CHORDLINK, the system in [17] is designed to automatically compute a set of clusters that guarantees desired properties (e.g., planarity) for the graph of clusters and adopts the orthogonal drawing style [38] (instead of a straight-line node-link diagram) to represent the outside of the clusters. Hybrid visualizations are also used for the analysis of dynamic networks (see, e.g., [39], [40], [41]). Finally, several theoretical contributions study the complexity of minimizing inter-cluster edge crossings in different hybrid visualization models [12], [13], [14], [18], [19], [20], [21], [23], [24], [25], [42], [43], [44].

User studies in graph drawing and network visualization. The evaluation of graph visualization methods and systems through the execution of cognitive user studies has an established tradition, which dates back to the late 90s [45], [46]. We discuss here the contributions that are mainly related to our study, while we refer the reader to [47] for a recent comprehensive survey on the subject.

There is a series of works that compare node-link diagrams with matrix-based representations [48], [49], [50], [51], [52], [53], [54], [55]. A common finding of these studies is that node-link diagrams have usually better performance

on topology and connectivity tasks when graphs are not too large and dense, while matrices perform better on group tasks. Our study does not aim to further compare node-link and matrix representations, but rather to investigate hybrid visualizations that integrate these two, or other types of, drawing conventions.

In the context of hybrid graph visualizations, Henry and Fekete [56] conduct a user study on MatLink, a model that combines adjacency matrices overlaid with node-link diagrams using curvature for the links. They find that MatLink outperforms the two individual metaphors (node-link diagrams and adjacency matrices) for most of the considered tasks, including path-related tasks, where matrices are usually worse than node-link. Differently from our study, [56] does not focus on the visualization of networks with clusters. Henry et al. [22] present a user study aimed to understand whether node duplication for non-disjoint clusters improves the performance of NODETRIX for some types of tasks. Since the majority of hybrid visualizations are designed to deal with disjoint clusters, our study focuses on this setting; moreover, we consider tasks that are mostly different from those addressed in [22].

As a final remark, to the best of our knowledge, our evaluation on the usefulness of integrating multiple models into a single visualization is the first attempt in this direction.

3　DESIGN OF THE COMPARATIVE USER STUDY

This section describes in detail the design of our user study. The target population consists of researchers and analysts (including practitioners, academics, and students) that make use of network visualization to accomplish tasks of analysis on real-world networks. In the following we discuss the

visualization models, the tasks and the hypotheses, the stimuli, and the experimental procedure.

3.1 Visualization Models

We evaluate four different models for the visualization of undirected clustered networks, where subsets of nodes are grouped into clusters (see Fig. 1 for an illustration). An edge connecting two nodes in the same cluster is an *intra-cluster edge*; every other edge is an *inter-cluster edge*. The models are:

– NODELINK (NL). The classical node-link model, where nodes are represented as small disks and edges are straight-line segments connecting their end-nodes. In this model, we visually highlight each cluster through a colored convex region that includes all the cluster's nodes.

– CHORDLINK (CL). The model proposed in [15], [57]. Nodes outside clusters and inter-cluster edges are drawn as in the NODELINK model. Clusters are represented as chord diagrams. A node in a cluster may have multiple copies, each represented as a colored circular arc along the circumference of the chord diagram; all copies of the same node have the same color. An intra-cluster edge is a "ribbon" connecting two of the copies representing its end-nodes.

– NODETRIX (NT). The model introduced in [11]; each cluster C of size n is represented by an $n \times n$ adjacency matrix. Nodes outside clusters and edges between them are drawn as in NODELINK. An inter-cluster edge having an end-node v in a cluster C is drawn as a curve incident to the row or to the column associated with v, on one of the sides of the matrix representing C.

– RCI-NODETRIX (RC). A variant of NODETRIX, proposed in [25], [58]. The difference with the NODETRIX model is that in each adjacency matrix, the row and the column associated with the same node may have different indices, in order to avoid some crossings between inter-cluster edges. As a consequence the matrices may not be symmetric.

Rationale. Among the various types of hybrid visualizations described in the literature, we selected NT and CL as they are designed to work similarly within visualization systems devoted to the analysis of real-world networks. We exploited the system in [15], which implements both these models in a unique interface, where the implementation of NT reflects the one given in [59] by the authors of [11]. The system in [15] allows direct support for clustered drawings in the NL model and makes it possible to create drawings in all the supported models by defining the same set of clusters on the same node-link diagram. For the purposes of our experiment, we enriched the system with the RC model.

3.2 Tasks

We defined six different tasks, listed in Table 1. The tasks of Table 1 have two attributes: LeeTax, which classifies each task according to the taxonomy by Lee et al. [60]; and AmarTax, which describes the low-level visual analytics operations needed to execute each task according to the taxonomy by Amar et al. [61].

Rationale. We designed the user study with a set of tasks that requires to explore the drawing locally and globally. Moreover, each task is easy to explain, it can be executed in a reasonably short time, and it can be easily measured. Concentrating on representative tasks is a common approach for this kind of experiments (see, e.g., [62]), which supports generalizability to more complex tasks that include these representatives as subroutines. Most of our tasks are used in previous graph visualization user studies (e.g., [46], [55], [63], [64]) and they cover all task categories of LeeTax [60], with the exception of the browsing category. We excluded the latter because it requires to interact with the visualization and we decided to avoid interaction to keep the test as simple as possible and avoid possible confounding factors.

With respect to the recent top-level task classification by Burch et al. [47], we observe that all our tasks are *interpretation tasks*. Indeed, our goal is to evaluate the differences of the considered visualization models in terms of readability, understandability, and effectiveness.

The specific chosen tasks are designed to be representative of real exploratory questions that a user formulates when analyzing a network. At the same time each task is formulated in a way that makes it possible to easily measure the user's performance. Namely: T1 refers to a classical question about whether two entities of a network are directly connected; T2 focuses on establishing the importance of a node with respect to another based on node-degree; T3 simulates a task where the user wants to establish whether two nodes are relatively close to each other in terms of theoretical distance in the network; T4 concentrates on quickly detecting a node in a portion of the network based on one of the displayed attributes; T5 reflects a task in which the user wants to establish the relevance of a cluster with respect to another in terms of their level of connectivity; T6 aims to estimate the level of connectivity between different portions of the network. About task T5, we also point out that there are two commonly used definitions for the density of a graph with n nodes and m edges: $d_1 = \frac{m}{n}$ and $d_2 = \frac{2m}{n(n-1)}$. We adopted definition d_1 for two reasons: (a) it is simpler to explain to a user; (b) according to previous research work [65], d_1 is a better descriptor of the complexity of real-world networks. Indeed, the visual perception of the density of a cluster region is affected by the number of nodes in the cluster; if a drawing contains two clusters with different sizes, the largest one may be perceived as a denser portion of the drawing, even if it has lower density according to d_2.

3.3 Hypotheses

Similarly to previous works (e.g., [22], [55]), we define our hypotheses based on tasks, structuring them according to the task categories of LeeTax.

H1: On topology-based tasks (T1, T2, T3), we expect NODELINK to have faster response time than hybrid visualizations. In contrast, we expect hybrid visualizations to have a lower error rate than NODELINK, and CHORDLINK to behave better than NODETRIX and RCI-NODETRIX.

H2: On attribute-based tasks (T4), we expect NODETRIX and RCI-NODETRIX to outperform the other two models in terms of response time and error rate.

H3: On overview tasks (T5, T6), we expect hybrid visualizations to perform better than NODELINK in terms of both

TABLE 1
Tasks used in our study.

Task	LeeTax	AmarTax
T1. Is there an edge that links the two highlighted nodes?	topology-based (adjacency)	retrieve value
T2. Which of the two highlighted nodes has higher degree?	topology-based (adjacency)	retrieve value; sort
T3. Is there a path of length at most k that connects the two highlighted nodes?	topology-based (connectivity)	retrieve value; compute derived value; filter
T4. Which of the following three node labels appear in the highlighted portion of the network?	attribute-based (on the nodes)	retrieve value; filter
T5. What is the denser* cluster between the two highlighted?	overview	filter; compute derived value; sort
T6. How many edges directly connect the two highlighted parts of the drawing?	overview	filter; compute derived value
*The cluster density is the ratio between the number of edges and the number nodes in a cluster		

response time and error rate. Among the hybrid visualizations, we expect NODETRIX and RCI-NODETRIX to be better than CHORDLINK, especially for cluster density estimation.

Rationale. About **H1**, our expectations in terms of response time are motivated by the fact that NODELINK is quite intuitive and widely used. Moreover, hybrid visualizations intrinsically require to switch from a visualization metaphor to another during the visual exploration, which may represent a cognitive effort. Concerning the error rate, we think that, by reducing the visual clutter, hybrid visualizations help to avoid ambiguities (such as edges that are almost collinear) and therefore may better support topology-based tasks. Also, since topology-based tasks are known to be harder when dealing with matrices, we expect CHORDLINK to have better performance than NODETRIX and RCI-NODETRIX in terms of error rate. About **H2**, we believe that placing labels on a matrix side is more effective than placing them around chord diagrams or near nodes in a node-link diagram. In chord diagrams labels may be harder to read due to their rotation, while in node-link diagrams they may be hidden by edges. About **H3**, we expect hybrid visualizations to behave better than NODELINK due to their capability to provide a clearer cluster representation. For tasks that require to estimate cluster density, NODETRIX and RCI-NODETRIX have the advantage that the proportion between black (edges) and white (non-edges) cells immediately conveys the density of a cluster; this estimation is more difficult in CHORDLINK, where node duplication may give the impression that a cluster is sparser than it actually is.

3.4 Stimuli

Our experimental objects are three real-word networks of small/medium size. The first one, weavers, is an animal social network with 64 nodes and 177 edges, describing the interactions of a colony of weavers in the usage of nests [66], [67]. The second one, e.coli, is a biological network with 97 nodes and 212 edges that describes transcriptional interactions in the Escherichia coli bacterium [68]. The third one, dblp, is a co-authorship network obtained from the DBLP repository [69] by searching for the keyword "network visualization" and considering only the largest connected component, which has 118 nodes and 322 edges.

For each of the four visualization models described in Section 3.1, we produced a diagram of the three networks described above. The diagrams for NODELINK are computed through the force-directed algorithm in the D3.js library [70]. Starting from these drawings, we defined some geometric clusters with the K-means-based technique described in [15]. As explained in Section 3.1, the system presented in [15] is used to compute the diagrams in the CHORDLINK, NODETRIX, and RCI-NODETRIX models with the same sets of clusters. The algorithm for NODETRIX is based on [11], [59] and uses the *leaf order method* to compute the row/column order [71]. The algorithm for RCI-NODETRIX is a variant of the one for NODETRIX, where the orders for the rows and the columns are independently computed to reduce the number of crossings between inter-cluster edges; this is done by an adaptation of the sifting algorithm for layered drawings [72], [73]. In all four diagrams of the same network, we labeled all the nodes that belong to clusters and few high-degree nodes outside the clusters. We avoid label duplication in all the drawings; this reduces visual clutter and suffices to correctly interpret the data in all models; in particular, in NODETRIX and RCI-NODETRIX we filled each cell whose row and column refer to the same node with a color distinct from black and white (these cells correspond to the main matrix diagonal in NODETRIX). Also, we use numerical id labels instead of real names to guarantee anonymity and to avoid that users could be influenced by their knowledge about the network.

Each of the 12 stimuli obtained by applying each of the 4 conditions (models) to the 3 experimental objects (networks) is used in all of the 6 tasks described in Section 3.2, for a total of $4 \times 3 \times 6 = 72$ trials. For T1, T2, and T3, we highlighted the node labels with a yellow background; to help the user to locate the nodes, we also put a red cross close to the clusters containing them. For T4 and T6, we highlighted the regions of interest by enclosing them inside a colored polygonal area. Finally, for T5 we indicated the two clusters of interest with large red labels. The trials for the network weavers can be found in the Supplemental Material.

Rationale. The visualization models that we compare are suitable for networks with up to few thousand nodes and edges, while for significantly larger networks ad-hoc tech-

niques are required that typically reduce the amount of displayed information. The choice of using networks with few hundred elements avoids an excessive burden for the participants. Namely, we wanted that each trial could be executed in a reasonable amount of time without an excessive fatigue and that the whole test could be completed in about 30 minutes. Further, since we decided to show static images (without zoom), the whole picture of the network should be displayed with a level of zoom that keeps the labels readable. Since the considered hybrid models are intended to visualize networks that are globally sparse but locally dense, we selected three networks that exhibit this structure. Moreover, we designed the specific trials so that the user was required to explore both the sparse parts of the network, represented by the node-link metaphor, and the dense parts, represented in different ways depending on the model. Finally, we remark that for each task we generated three different trials, one for each network. To mitigate the fatigue effect and keep the overall time of the experiment reasonable, we did not formulate multiple repeats of each task for the same network.

3.5 Experimental Setting and Procedure

We designed a between-subject experiment where each participant was exposed to one of the four conditions and hence to 18 trials. The users executed the test fully on-line. The questionnaire was prepared using the LimeSurvey tool (https://www.limesurvey.org/) and is structured as follows. First, some information about the user are collected, namely: gender, age, educational level, expertise in graph visualization, screen size, and possible color vision deficiency. Then, the visualization model to be assigned to the user is decided in a round robin fashion. Based on this assignment, a video tutorial is presented, followed by a training phase in which the user has to answer a trial for each task with an explanatory feedback in case of wrong answer. Next, the 18 trials are presented in random order. Finally, the user is asked for some qualitative feedback: two Likert scale questions about the aesthetic quality of the drawings and about the easiness of the questions, plus an optional free comment. While no time limit was given to complete the test, the participants were asked to answer each question as fast as they could but, at the same time, trying to be accurate. For each user, we collected the answers and the time spent on each question. We recruited the participants with announcements to the gdnet, ieee_vis, infovis mailing lists and to the computer engineering students of the universities of Perugia and Roma Tre.

The actual experiment was preceded by a pilot study with 19 participants, mostly colleagues and students in computer engineering, who are representatives of our class of target users. Based on the feedback received from the pilot study, we made some small changes to the survey. More precisely, for task T4 we increased the number of labels to be found from one to three, because we had 100% correct answers. For task T6 we changed the type of question from a single choice question (the user selects the answer from a fixed set of values) to a free text answer. We made this change because the limited number of options helped to guess the right answer (some participants reported that

when they had in mind a wrong value that was not present among the options, they selected the closest value).

Rationale. As previously explained, exposing the users to all four conditions would imply each user solving 72 trials. We believe that keeping the same level of attention in such a long experiment is difficult, and may cause many participants prematurely quitting the test. Besides such undesired fatigue effect, a within-subject design would also imply that each user sees the same experimental object 24 times, which makes it difficult to avoid the learning effect. Hence, we adopted a between-subject design, where each participant is exposed to only one condition. This choice limited the number of trials per user to 18, thus mitigating both the fatigue and the learning effect, which is further counteracted by presenting the trials in a random order. Finally, since the test includes a video tutorial and a training phase to make the user familiar with the given visualization model, an additional advantage of the between-subject design is that these phases can be focused on one model only. About the execution of the experiment, we opted for a fully on-line test for two reasons: (i) the difficulties to perform a controlled in-person experiment due to the COVID-19 pandemic; (ii) the possibility of recruiting a larger number of participants that better represent our target population, through announcements on the aforementioned mailing lists.

4 RESULTS OF THE COMPARATIVE USER STUDY

Participants. We collected questionnaires from 89 participants. We discarded seven tests for various reasons. One of the participants indicated in the free comments area that some images were not shown properly. Four participants indicated to have some color vision deficiency. Since they happened to be all assigned to the same model, we decided to discard their tests to avoid an unbalanced effect of this factor on the results of the experiment. Finally, since the experiment was fully online and thus not controlled, we discarded two tests whose total response time (i.e., the total time spent to answer the 18 trials) was an outlier. According to common practice, we consider the total response time of a test an outlier if it falls more than 1.5 times IQR below the first quartile or above the third quartile. Of the remaining 82 tests, 19 were for CHORDLINK and 21 for each of the other models. Regarding the participants, 66 (80.49%) were males, 15 (18.29%) were females, and 1 (1.22%) preferred not to answer. The majority of them (82.72%) were aged below 40. 85.37% of the participants has at least a Bachelor's degree, with 34.15% of them having a doctoral degree. 62.2% of the participants declared medium or high familiarity with graph visualization and 68.29% used a screen of size at least 15". See the Supplemental Material for detailed charts about the personal information about the participants.

First-level Analysis - Quantitative Results. We compared the performance of the four models over all data in terms of error rate and response time. For T1–T5, the error rate of a user is the ratio between the number of wrong answers and the total number of questions. Recall that there are three questions per task and that in T4 the user has to find three labels for each question. About T6, the error on a question is computed as $1 - \frac{1}{1+|u-r|}$, where u is the value given by

TABLE 2
Results for error rate (top) and response time (bottom) for each task.

	Task	Models ranked by average values	Kruskal-Wallis		Pairwise comparisons (p-value)					
			$H(3)$	p-value	NL-CL	NL-NT	NL-RC	CL-NT	CL-RC	NT-RC
Error rate	Overall	CL (0.222), RC (0.262), NT (0.280), NL (0.283)	3.042	0.385	-	-	-	-	-	-
	T1	CL (0.175), NL (0.191), RC (0.254), NT (0.349)	**8.471**	**0.037**	1.000	0.179	1.000	**0.036**	1.000	0.549
	T2	CL (0.140), NL (0.143), RC (0.191), NT (0.238)	2.419	0.490	-	-	-	-	-	-
	T3	CL (0.140), RC (0.143), NT (0.270), NL (0.333)	**10.882**	**0.012**	**0.035**	0.788	**0.024**	1.000	1.000	1.000
	T4	NT (0.048), CL (0.053), RC (0.101), NL (0.138)	6.703	0.082	-	-	-	-	-	-
	T5	NT (0.270), NL (0.302), CL (0.386), RC (0.397)	2.407	0.492	-	-	-	-	-	-
	T6	RC (0.246), NL (0.282), CL (0.290), NT (0.293)	0.478	0.924	-	-	-	-	-	-
Response time	Overall	NL (28.77), NT (35.35), RC (35.97), CL (39.81)	**8.190**	**0.042**	**0.032**	0.666	0.321	1.000	1.000	1.000
	T1	NL (16.64), NT (19.91), RC (23.47), CL (26.12)	**20.084**	**<0.001**	**0.000**	1.000	**0.031**	**0.016**	0.832	0.720
	T2	NL (25.23), CL (36.35), NT (37.63), RC (39.95)	**12.632**	**0.006**	0.058	0.061	**0.006**	1.000	1.000	1.000
	T3	NL (24.58), RC (39.56), NT (42.68), CL (47.69)	**19.533**	**<0.001**	**0.000**	**0.005**	**0.011**	1.000	1.000	1.000
	T4	NT (35.88), RC (37.04), NL (42.36), CL (50.60)	**9.793**	**0.020**	1.000	0.579	1.000	**0.018**	0.143	1.000
	T5	NL (29.71), NT (37.76), CL (42.30), RC (44.24)	3.657	0.301	-	-	-	-	-	-
	T6	RC (31.56), NL (34.07), CL (35.83), NT (38.26)	1.265	0.737	-	-	-	-	-	-

the user and r is the correct value; the error rate for T6 is the average of the errors on the three questions of the task.

By performing the Shapiro-Wilk test with significance level $\alpha=0.05$, we found that data were not normally distributed. Hence, we performed the non-parametric Kruskal–Wallis test with significance level $\alpha=0.05$, which is suitable for comparing multiple independent samples. We finally performed post-hoc pairwise comparisons by using Bonferroni corrections. (See also [74], [75].)

Table 2 summarizes the results of our analysis both for the error rate (top) and for the response time (bottom). For each task, we list the models sorted by increasing values of average error rate or response time (shown in parentheses). We report in the table the statistic (column $H(3)$) and the p-value of the Kruskal-Wallis test. Finally, for those results that are statistically significant, we report the adjusted significance for each pairwise comparison (after Bonferroni corrections). Comparisons that are statistically significant are in bold. The box-plots of the error rate and response time for all the tasks can be found in the Supplemental Material.

Per-expertise Analysis - Quantitative Results. We refined the first-level analysis described above and performed a second-level analysis aimed at understanding whether there is a difference between users that are more expert in network visualization and those that are less expert. We think that experts can take advantage of hybrid visualizations more than non-experts. Indeed, hybrid visualizations intrinsically require more effort, which could result in an obstacle to the analysis for non-expert users. To investigate this, we restrict the general analysis described above to experts and non-experts, separately. We refer to this analysis as *per-expertise analysis*.

To identify expert users, we consider two different criteria: the self-declared levels of expertise and of education; see Fig. 3. Regarding the level of expertise, we distinguish between *veterans*, namely those users who declared high or medium expertise, and *novices*, namely those who declared low or none expertise. Concerning the educational level, we distinguish between *seniors*, namely those users with a Mas-
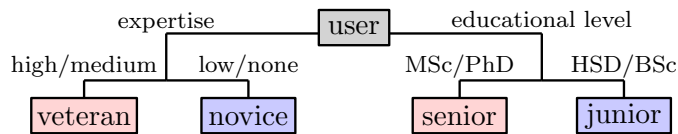

Fig. 3. Users' classification for the per-expertise analysis of the results.

ter's or a doctoral degree, and *juniors*, namely those users with at most a Bachelor's degree. Since all the recruited users have a background in computer science and some familiarity with network visualization, the educational level can be considered a good indicator of the users' expertise. This is also confirmed by the fact that 80% of users with a Master's or a doctoral degree declare a high or medium expertise and 73% of those who declare a high or medium expertise have a Master's or a doctoral degree.

Tables 3 and 4 summarize the results of our analysis for expert users, namely veterans and seniors. We report in the Supplemental Material the charts showing the average error rate and the average response time for novices and juniors.

Motivated by the lack of statistical significance in the results for non-experts (both novices and juniors), we conducted an additional independent experiment with another set of non-expert users. We recruited 41 students from the Bachelor's course in computer engineering of the University of Perugia, who executed the same test described above, with exactly the same environment and procedure as for the first experiment. This new experiment provided some evidence of the null hypothesis for this class of users, as the only significant result (p-value $= 0.016$) concerns task T2, for which NODELINK has a lower average error rate than NODETRIX.

Qualitative Results. At the end of the test, we presented to the users the following questions: (F1) *"How much do you like the diagrams you have seen?"* and (F2) *"How easy did you find answering the test questions?"*. The answers, given in a 5-point Likert scale, are summarized in the Supplemental Material, where we also report the answer distributions as

TABLE 3
Results for error rate (top) and response time (bottom) for each task considering only veterans.

	Task	Models ranked by average values	Kruskal-Wallis		Pairwise comparisons (p-value)					
			H(3)	p-value	NL-CL	NL-NT	NL-RC	CL-NT	CL-RC	NT-RC
Error rate	Overall	CL (0.120), RC (0.197), NL (0.228), NT (0.288)	10.346	0.016	0.386	1.000	1.000	0.008	0.423	1.000
	T1	CL (0.056), NL (0.067), RC (0.103), NT (0.292)	22.240	<0.001	1.000	0.002	1.000	0.000	1.000	0.005
	T2	CL (0.056), NL (0.100), RC (0.103), NT (0.292)	10.693	0.014	1.000	0.199	1.000	0.023	1.000	0.77
	T3	CL (0.056), RC (0.128), NT (0.292), NL (0.333)	11.141	0.011	0.015	1.000	0.106	0.184	1.000	0.940
	T4	CL (0.017), NT (0.044), RC (0.062), NL (0.160)	6.517	0.089	-	-	-	-	-	-
	T5	NL (0.233), CL (0.250), NT (0.271), RC (0.308)	0.824	0.844	-	-	-	-	-	-
	T6	CL (0.175), NL (0.200), RC (0.223), NT (0.256)	1.802	0.614	-	-	-	-	-	-
Response time	Overall	NL (23.94), NT (33.11), RC (35.17), CL (44.35)	14.509	0.002	0.001	0.419	0.104	0.136	0.803	1.000
	T1	NL (13.40), NT (20.52), RC (24.80), CL (27.05)	21.646	<0.001	0.000	0.113	0.003	0.069	1.000	0.947
	T2	NL (19.49), NT (36.82), RC (40.40), CL (40.66)	17.271	<0.001	0.003	0.034	0.001	1.000	1.000	1.000
	T3	NL (20.40), NT (39.30), RC (39.96), CL (55.26)	17.531	<0.001	0.000	0.027	0.024	0.643	1.000	1.000
	T4	NT (34.72), RC (35.99), NL (43.85), CL (53.20)	11.771	0.008	1.000	0.359	1.000	0.008	0.085	1.000
	T5	NL (20.09), NT (31.46), RC (39.55), CL (53.80)	9.078	0.028	0.027	1.000	0.958	0.132	0.707	1.000
	T6	NL (26.44), RC (30.34), NT (35.84), CL (36.15)	3.536	0.316	-	-	-	-	-	-

TABLE 4
Results for error rate (top) and response time (bottom) for each task considering only seniors.

	Task	Models ranked by average values	Kruskal-Wallis		Pairwise comparisons (p-value)					
			H(3)	p-value	NL-CL	NL-NT	NL-RC	CL-NT	CL-RC	NT-RC
Error rate	Overall	CL (0.104), RC (0.218), NT (0.278), NL (0.289)	9.786	0.020	0.043	1.000	1.000	0.027	0.381	1.000
	T1	CL (0.083), NL (0.167), RC (0.214), NT (0.405)	13.006	0.005	1.000	0.077	1.000	0.008	1.000	0.063
	T2	CL (0.083), NL (0.167), RC (0.119), NT (0.214)	3.247	0.355	-	-	-	-	-	-
	T3	CL (0.042), RC (0.119), NT (0.262), NL (0.333)	8.657	0.034	0.046	1.000	0.135	0.575	1.000	1.000
	T4	CL (0.013), NT (0.043), RC (0.107), NL (0.190)	11.714	0.008	0.012	0.071	1.000	1.000	0.183	0.948
	T5	CL (0.167), NL (0.267), NT (0.310), RC (0.333)	1.927	0.588	-	-	-	-	-	-
	T6	CL (0.150), RC (0.193), NL (0.230), NT (0.236)	1.546	0.672	-	-	-	-	-	-
Response time	Overall	NT (20.20), NL (29.58), RC (36.38), CL (40.04)	5.394	0.145	-	-	-	-	-	-
	T1	NL (14.73), NT (20.20), RC (24.13), CL (25.72)	14.676	0.002	0.005	0.530	0.010	0.270	1.000	0.709
	T2	NL (25.18), CL (33.97), NT (38.61), RC (44.39)	9.682	0.021	0.614	0.290	0.011	1.000	1.000	1.000
	T3	NL (24.87), RC (42.35), NT (42.98), CL (46.59)	10.475	0.015	0.024	0.073	0.058	1.000	1.000	1.000
	T4	RC (36.64), NT (36.70), NL (46.22), CL (49.02)	6.635	0.074	-	-	-	-	-	-
	T5	NL (30.12), RC (37.96), NT (40.22), CL (56.18)	0.297	0.961	-	-	-	-	-	-
	T6	RC (32.82), NL (36.38), CL (28.78), NT (36.45)	1.265	0.737	-	-	-	-	-	-

box-plots; we assigned a score from 1 (lowest) to 5 (highest) to each answer. While there is no statistically significant difference among the models, about (F1) NODELINK received the highest percentage of strongly negative appreciations and NODETRIX received the highest percentage of strongly positive appreciations, although with high variance. About (F2), the easiness of answering was judged medium on average for all the models. Moreover, concerning (F1) experts (veterans in particular) prefer hybrid visualizations rather than NODELINK, while there is no evident difference between the models for non-experts (both novices and juniors). About (F2), both for experts and non-experts the answers are in line with the general case.

In what follows we report, for each model, a summary of the main free comments posted by the participants at the end of the study.

NODELINK: All comments point out that the visual clutter caused by dense portions of the network makes the execution of some tasks difficult. This is coherent with the motivation behind the introduction of hybrid graph visualizations.

CHORDLINK: The comments point out that node duplication affects the drawing readability and makes it difficult to perform tasks related to cluster density. This is coherent with our rationale about Hypothesis **H3**, i.e., node duplication may give the impression that a cluster is sparser than it actually is. Other comments highlight that cluster regions represented by circles with small diameter make intra-cluster edges difficult to distinguish in some cases.

NODETRIX: The main comments report a difficulty in reading inter-cluster edges and their incidence to the matrices.

RCI-NODETRIX: The main comment here is that it is somewhat counter-intuitive that the matrices do not use the same row and column order, and this has a negative impact on following paths in the network.

We further report that two participants found our definition of density (denoted as d_1 in Section 3.2) less intuitive

than the alternative one (denoted as d_2 in Section 3.2).

Discussion. The following highlights are a summary of our results. We first discuss the results per hypothesis and, for each of them, we describe the results about the first-level analysis and about the per-expertise analysis. We then consider the results over all tasks. We conclude the discussion by looking at the data from different perspectives.

– Hypothesis **H1** is largely supported by the results in terms of response time and partially supported in terms of error rate. *First-level analysis.* NODELINK is faster (with statistical significance) than: CHORDLINK and RCI-NODETRIX for task T1; RCI-NODETRIX for task T2; all hybrid models for T3. The slower response time of RCI-NODETRIX with respect to NODELINK for all the topology-based tasks seems to confirm the difficulty pointed out by some participants about dealing with non-symmetric matrices. In terms of error rate, while there is no statistically significant difference for task T2, we observe that for task T3 both CHORDLINK and RCI-NODETRIX yield better accuracy than NODELINK, and for task T1 CHORDLINK behaves better than NODE-TRIX. One may wonder why the same behavior is not observable between CHORDLINK and RCI-NODETRIX on task T1; our interpretation is that this might depend on the smaller number of crossings that RCI-NODETRIX usually causes between edges that are incident to the matrices with respect to NODETRIX. *Per-expertise analysis.* In terms of response time, for both veterans and seniors the results for T1 and T3 are essentially confirmed, while for veterans we have additional statistical significance; in particular, NODELINK is faster than all hybrid models. The data support **H1** in terms of error rate also in the smaller group of expert users. Namely, the better performance of CHORDLINK with respect to NODETRIX is confirmed for task T1 and it is additionally observed on T2 for veterans. Also, the better accuracy of CHORDLINK with respect to NODELINK for T3 is confirmed.

– The data provide some evidence to support hypothesis **H2**. *First-level analysis.* The two models based on matrices seem to lead to faster response time than the other two models, with statistical significance when comparing NODETRIX and CHORDLINK. In terms of error rate, we do not observe any statistically significant difference that supports or disproves our hypothesis. The high accuracy achieved with all models seems to reveal that this task is indeed generally easy. *Per-expertise analysis.* In terms of response time, the results for veterans in the per-expertise analysis confirm what we observed in the first-level analysis. Concerning the error rate, we mention that for seniors CHORDLINK has better performance than NODELINK in the per-expertise analysis.

– Hypothesis **H3** is not supported by our results. *First-level analysis.* We do not observe any statistically significant difference among the four models. *Per-expertise analysis.* The only statistically significant result related to **H3** is NODELINK being faster than CHORDLINK on T5 for veterans, which contrasts our hypothesis.

– We now evaluate the four visualization models on over all tasks. *First-level analysis.* The general analysis shows that NODELINK outperforms CHORDLINK in terms of response time with statistical significance. *Per-expertise analysis.* The per-expertise analysis for veterans confirms what observed

in the first-level analysis in terms of response time. Further, concerning error rate CHORDLINK outperforms NODETRIX both for veterans and seniors and CHORDLINK outperforms NODELINK for seniors. These results suggest that NODELINK is faster than CHORDLINK, which however is more accurate than NODETRIX.

– Concerning the comparison between expert and non-expert users, the results of the per-expertise analysis partially support our idea that hybrid visualizations are more suited for expert users than for non-experts. Namely, veterans and seniors achieve better accuracy with CHORDLINK than with NODELINK on some tasks. Also, the only statistically significant result of the second experiment with students (concerning T2) indicates that non-experts have better performance with NODELINK.

– To better understand the difference between expert and non-expert users, we further analyzed the data from a different perspective: for each visualization model we compared the performance of the experts and of the non-experts. The results of this analysis, called *per-model analysis*, are reported in the Supplemental Material. They show that, not surprisingly, the expert users (both veterans and seniors) have better accuracy than non-experts on several tasks. This is particularly evident for CHORDLINK. On the other hand, novices are faster than veterans with CHORDLINK. Our interpretation of this observed phenomenon is that novices found the CHORDLINK model more tricky to use and thus they tend to guess the right answer.

– There are other interesting questions that are suggested by the data collected from our experiments. For example, one may wonder whether there are visualization models that are more hindered by a screen of small size, or what is the impact of the network size on the performance of the users with the different models. These analyses did not lead to statistical significant results, but we report in the Supplemental Material the charts showing the average error rate and the average response time for each visualization model.

Limitations. We conclude by discussing the limits of our study. The choice of not allowing interaction implied to use networks of small/medium size that fit into the screen window; also, it required to have a set of predefined clusters that the user cannot change. On the other hand, a non-interactive environment facilitated the execution of an on-line test; we believe that enabling visual interaction for the considered models would require a different study design, preferably based on a controlled experiment. Further, interactions may introduce confounding factors and it is difficult to design interaction features that are fair to all models.

The networks used for the comparative study have similar characteristics in terms of size, density, and cluster structure. Hence, our results should not be generalized to networks that have significantly different characteristics.

The number of tasks was limited to six, which is in line with many previous studies. Although some works use a larger number of tasks (see, e.g., [55]), we believe that more tasks may cause long execution times and a high fatigue effect for the users, which may result in less reliable data.

As pointed out at the end of Section 3.4, to keep the experiment affordable for the user we decided to avoid multiple repeats for each task. We believe that the potential

impact of our choice on the quality of the results is less relevant than the one caused by an eccessive fatigue effect, which might produce unreliable answers from the user. To better evaluate the implications of this aspect, it would be interesting in the future to design new experiments that consider the execution of the same type of task multiple times, without increasing significantly the total number of stimuli. For example, one can think of a between-subject experiment that partitions the types of tasks over the different groups of users, as we did for the visualization models. Such an experiment would require a larger number of participants.

About the interpretation of the results, we remark that for tasks T1–T5 the error rate of a user is computed as the ratio between the number of wrong answers and the total number of questions, while for T6 the answers are evaluated on a fuzzy rating scale. An alternative approach could be redesigning the experiment so that all the tasks allow for a fuzzy rating scale evaluation. Depending on the type of task, this may lead to an interpretation of the results that better reflects how network visualizations are used in practice.

Finally, the visualization models that we compare may be sensitive to the specific algorithms used to produce the drawings. This justifies further investigation with different layout algorithms.

Findings and guidelines. Keeping in mind the discussed limitations, we conclude this section by summarizing the major findings of our experiment.

- When a user aims to get insights on the structural properties of a globally sparse but locally dense network (e.g., connectivity level between nodes or node centrality), node-link diagrams may lead to faster analysis than hybrid visualizations.
- Still about getting insights on the structural properties of the network, if one wants the analysis to be more accurate even if slower, using hybrid visualizations (in particular CHORDLINK) is recommendable. This is even more evident for expert users.
- When the analysis requires to find nodes with specific attributes, our results suggest that matrix-based visualizations (in particular NODETRIX) are recommendable, both for experts and for non-expert users.

5 ICE-T EVALUATION

The comparative study presented in the previous section does not identify a hybrid model that is clearly superior to the others for the execution of both topology-based, attribute-based, and overview tasks, independently of the user's expertise. Also, as highlighted in the previous section, one of the limits of our comparative study is the lack of user interaction. This makes it natural to evaluate the effectiveness of a system that integrates different types of hybrid models and that offers a more powerful interactive interface to the user. In particular, besides the possibility of getting an initial set of automatically computed clusters (like in the comparative study), the user should be able to manually modify these clusters and decide how to visualize them, based on her ability to interpret a certain type of diagram rather than another, and depending on the size and connectivity level of each cluster. This flexibility could be considered an additional element of complexity for the user. However, we believe that, after a suitable training phase, the higher degree of freedom that it offers with respect to using a unique model for clusters representation, can help the user to get more insights from the visualization.

Following the considerations above, we implemented an interactive Web-based environment that allows users to visualize a network by arbitrarily combining all hybrid models presented in this paper; the system is available at http://mozart.diei.unipg.it/tappini/ChordLink/ and it is shown in Fig. 2. Among typical zooming and panning operations, this system provides facilities to: (i) manually or automatically select clusters and decide the representation of each cluster, including the possibility of collapsing it into a single (bigger) node; (ii) quickly change the type of visualization for a cluster; (iii) drag vertices into a cluster or move vertices of a cluster out of it; (iv) customize the set of node labels that are displayed, either by applying specific predefined policies or by acting manually on each single label. Additionally, the system has a panel to the right of the drawing canvas, which lists the labels of all nodes in the network, ordered alphabetically or by decreasing values of node degrees, depending on the user's preference. One can interact with this list, searching for a specific label or selecting a label to highlight the corresponding node in the layout.

We conducted an evaluation of this visualization environment by means of the ICE-T methodology proposed by Wall et al. [27], which enables a quantitative measurement of a visualization within a framework defined by Stasko [28]. According to this framework, the *value* of a visualization is a linear combination of four *components*, each referring to a specific ability of the visualization: (1) *Insight* – Ability to spur and discover insights and/or insightful questions about the data; (2) *Time* – Ability to minimize the total time needed to answer a wide variety of questions about the data; (3) *Essence* – Ability to convey an overall essence or takeaway sense of the data; (4) *Confidence* – Ability to generate confidence, knowledge, and trust about the data, its domain and context.

The methodology of Wall et al. introduces a hierarchical extension of the original value framework of Stasko. Each of the four components comprises one to three *guidelines*, capturing the core-concepts of the component; each guideline contains one to three *heuristics*, i.e., an actionable and rate-able statements that reflect how the visualization achieves that guideline. These heuristics must be individually rated by visualization experts on a 7-point rating scale from 1 (strongly disagree) to 7 (strongly agree), or NA-not applicable. This rating is collected using a survey, available at http://visvalue.org, with a total of 21 heuristics.

Following the ICE-T methodology (which suggests using five ore more experts), we recruited 6 participants in the field of graph drawing and network visualization, 4 from academy and 2 from industry. For each participant we established an individual remote session (over MS Skype or MS Teams) of about 30 minutes, in which we: (i) introduced the visualization environment to the participant; (ii) invited her to train with the environment under our supervision; (iii) asked the participant to continue interacting with the environment and completing the evaluation offline. For this last phase, we provided the participant with a collaboration

ICE-T		Components			
		Insight	Time	Essence	Confidence
Participants	P1	5,61	5,67	5,50	5,67
	P2	6,56	6,25	7,00	6,50
	P3	5,94	5,75	5,75	5,00
	P4	6,50	6,42	6,50	5,17
	P5	6,83	6,58	6,50	5,50
	P6	5,56	3,67	5,50	5,25
Statistics	avg	6,17	5,72	6,13	5,52
	std	0,49	0,98	0,57	0,49

Fig. 4. Participants' ratings. The color mapping is red (1) to green (7) with white (4) being neutral. The table shows the overall strength of each visualization with respect to each of the four components. The values in a column show how the different raters scored a visualization with respect to a specific component. For each component we also report the average and the standard deviation.

network representing co-authorships in IEEE TVCG articles in the years 2018–2020, and we suggested some potential general tasks (local and global) to perform, leaving the participant the freedom to expand the analysis at her discretion. The evaluation process ended when the participant sent the filled ICE-T survey to the evaluators; we also collected some post-evaluation feedback to complement their ratings.

The results of the evaluation are reported in the table of Fig. 4. Following the methodology in [27], for each participant we determined the score of a component by first computing the score of each guideline as the average of the scores of its heuristics, and then by averaging over the scores of all component's guidelines. All components received a cumulative average score higher than 5, which, according to the indications in [27], is considered a positive evaluation. More in details, the strength of the visualization is particularly evident in terms of the components Insight and Essence, which received a cumulative average score above 6, with a general consensus across all participants. In terms of Time, the average score of one participant is slightly below the neutrality threshold. Looking at the scores of the single heuristics for this component, the participant disagrees that the visualization supports using different data attributes to reorganize the visualization's appearance and indicates that the visualization does not sufficiently support smooth transitions between different levels of detail in viewing the data. We feel that this second aspect could be improved by designing suitable morphing techniques to pass from a level of detail to another; this opens to a future research direction, mostly unexplored in the context of hybrid visualizations. We finally remark that, in terms of Confidence, one participant considered NA the heuristic that estimates whether the visualization helps understand data quality; this is coherent with the fact that our visualization is not specifically tailored to address this issue.

6 CONCLUSIONS AND FUTURE RESEARCH

In this paper we have investigated hybrid visualizations in comparison with the classical node-link representations. The study has covered two complementary directions: (i) On the one hand, we evaluated the effectiveness of different hybrid visualization models through a comparative user study; (ii) On the other hand, we estimated the usefulness of an

interactive visualization that integrates all the considered hybrid models through an ICE-T evaluation.

Concerning direction (i), as an answer to RQ1, the results of the comparative study suggest that hybrid visualizations may help to overcome some limits of node-link diagrams in accurately executing topology-based tasks on globally sparse but locally dense networks, at the expenses of the execution time. About RQ2, we could not conclude that any of the considered hybrid models is superior; however, for some tasks, we observed better accuracy with CHORDLINK and faster execution with NODETRIX. These findings are more evident when focused on expert users, while a focus on non-experts does not reveal specific advantages when using a model rather than another. We also remark that our comparative study has some limitations and should not be generalized to settings significantly different from ours. In particular, in addition to considering networks with structural properties that are different from those used in our study, another possible extension can include interaction features and, as a consequence, tasks in the browsing category [60]. Also, the users of our experiment are researchers and analysts that make use of network visualization to accomplish tasks of analysis on real-world networks. It is natural to ask to what extent our findings hold true for users unfamiliar with any network visualization technology.

Concerning direction (ii), the general outcome of the ICE-T evaluation is rather positive, especially in terms of insight and essence of the visualization. We also learned that the visualization does not sufficiently support smooth transitions between different levels of detail in viewing the data. This last remark suggests a novel research direction, namely the study of efficient and effective morphing techniques to pass from a level of detail to another in the context of hybrid visualizations.

We conclude this paper by mentioning three additional research directions that stem from our work.

- We described a between-subject experiment. It would be interesting to design a similar experiment in which the visualization models are a within-subject factor.
- We investigated visualizations in which clusters are spatially separated. One could extend the experimental study to visualization models that allow the spatial overlap of clusters (see, e.g., [76], [77]).
- The outcome obtained with the ICE-T evaluation results from integrating the node-link model with different hybrid visualization models. Comparing our interactive system to one that allows the interaction with just one model (in particular node-link) is another direction for future work.

REFERENCES

[1] E. Di Giacomo, W. Didimo, F. Montecchiani, and A. Tappini, "A user study on hybrid graph visualizations," in *Graph Drawing and Network Visualization*, H. C. Purchase and I. Rutter, Eds. Cham: Springer International Publishing, 2021, pp. 21–38.

[2] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[3] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[4] M. A. Porter, J.-P. Onnela, and P. J. Mucha, "Communities in networks," *Notices of the American Math. Soc.*, vol. 56, pp. 1082–1097, 1164–1166, 2009.

[5] P. Bedi and C. Sharma, "Community detection in social networks," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 6, no. 3, pp. 115–135, 2016.

[6] W. Didimo, G. Liotta, and F. Montecchiani, "Network visualization for financial crime detection," *J. Vis. Lang. Comput.*, vol. 25, no. 4, pp. 433–451, 2014.

[7] J. Onnela, K. Kaski, and J. Kertész, "Clustering and information in correlation based financial networks," *The Europ. Phys. J. B-Cond. Matter and Complex Syst.*, vol. 38, no. 2, pp. 353–362, 2004.

[8] H. Wu, J. He, Y. Pei, and X. Long, "Finding research community in collaboration network with expertise profiling," in *ICIC (1)*, ser. LNCS, vol. 6215. Springer, 2010, pp. 337–344.

[9] G. W. Flake, S. Lawrence, C. L. Giles, and F. Coetzee, "Self-organization and identification of web communities," *IEEE Computer*, vol. 35, no. 3, pp. 66–71, 2002.

[10] H. Mahmoud, F. Masulli, S. Rovetta, and G. Russo, "Community detection in protein-protein interaction networks using spectral and graph approaches," in *CIBB*, ser. LNCS, vol. 8452. Springer, 2013, pp. 62–75.

[11] N. Henry, J. Fekete, and M. J. McGuffin, "NodeTrix: A hybrid visualization of social networks," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1302–1309, 2007.

[12] P. Angelini, G. Da Lozzo, G. Di Battista, F. Frati, M. Patrignani, and I. Rutter, "Intersection-link representations of graphs," *J. of Graph Algorithms and Applications*, vol. 21, no. 4, pp. 731–755, 2017.

[13] P. Angelini, P. Eades, S. Hong, K. Klein, S. G. Kobourov, G. Liotta, A. Navarra, and A. Tappini, "Turning cliques into paths to achieve planarity," in *Graph Drawing*, ser. LNCS, vol. 11282. Springer, 2018, pp. 67–74.

[14] ——, "Graph planarity by replacing cliques with paths," *Algorithms*, vol. 13, no. 8, p. 194, 2020.

[15] L. Angori, W. Didimo, F. Montecchiani, D. Pagliuca, and A. Tappini, "Hybrid graph visualizations with ChordLink: Algorithms, experiments, and applications," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 2, pp. 1288–1300, 2022.

[16] B. Bach, E. Pietriga, and I. Liccardi, "Visualizing populated ontologies with OntoTrix," *Int. J. Semantic Web Inf. Syst.*, vol. 9, no. 4, pp. 17–40, 2013.

[17] V. Batagelj, F. Brandenburg, W. Didimo, G. Liotta, P. Palladino, and M. Patrignani, "Visual analysis of large graphs using (X,Y)-Clustering and hybrid visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 11, pp. 1587–1598, 2011.

[18] G. Da Lozzo, G. Di Battista, F. Frati, and M. Patrignani, "Computing NodeTrix representations of clustered graphs," *Journal of Graph Algorithms and Applications*, vol. 22, no. 2, pp. 139–176, 2018.

[19] E. Di Giacomo, W. J. Lenhart, G. Liotta, T. W. Randolph, and A. Tappini, "(k,p)-planarity: A relaxation of hybrid planarity," in *WALCOM: Algorithms and Computation*, ser. LNCS, vol. 11355. Springer, 2019, pp. 148–159.

[20] ——, "(k,p)-planarity: A relaxation of hybrid planarity," *Theor. Comput. Sci.*, vol. 896, pp. 19–30, 2021.

[21] E. Di Giacomo, G. Liotta, M. Patrignani, I. Rutter, and A. Tappini, "NodeTrix planarity testing with small clusters," *Algorithmica*, vol. 81, no. 9, pp. 3464–3493, 2019.

[22] N. Henry, A. Bezerianos, and J. Fekete, "Improving the readability of clustered social networks using node duplication," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1317–1324, 2008.

[23] P. Kindermann, J. Sauer, and A. Wolff, "The computational complexity of the ChordLink model," in *38th European Workshop on Computational Geometry (EuroCG 2022)*, 2022.

[24] G. Liotta, I. Rutter, and A. Tappini, "Parameterized complexity of graph planarity with restricted cyclic orders," in *Proceedings of WG 2022*, pp. 383–397.

[25] ——, "Simultaneous FPQ-ordering and hybrid planarity testing," *Theoretical Computer Science*, vol. 874, pp. 59–79, 2021.

[26] X. Yang, L. Shi, M. Daianu, H. Tong, Q. Liu, and P. M. Thompson, "Blockwise human brain network visual comparison using Node-Trix representation," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 181–190, 2017.

[27] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko, "A heuristic approach to value-driven evaluation of visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 491–500, 2018.

[28] J. Stasko, "Value-driven evaluation of visualizations," in *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, 2014, pp. 46–53.

[29] D. Harel, "On visual formalisms," *Commun. ACM*, vol. 31, no. 5, pp. 514–530, 1988.

[30] G. Sindre, B. Gulla, and H. G. Jokstad, "Onion graphs: Asthetics and layout," in *VL*. IEEE Computer Society, 1993, pp. 287–291.

[31] U. Dogrusöz, E. Giral, A. Cetintas, A. Civril, and E. Demir, "A layout algorithm for undirected compound graphs," *Inf. Sci.*, vol. 179, no. 7, pp. 980–994, 2009.

[32] M. Kaufmann and D. Wagner, Eds., *Drawing Graphs, Methods and Models (the book grow out of a Dagstuhl Seminar, April 1999)*, ser. LNCS, vol. 2025. Springer, 2001.

[33] K. Sugiyama, *Graph Drawing and Applications for Software and Knowledge Engineers*, ser. Series on Software Engineering and Knowledge Engineering. WorldScientific, 2002, vol. 11.

[34] W. Didimo and F. Montecchiani, "Fast layout computation of clustered networks: Algorithmic advances and experimental analysis," *Inf. Sci.*, vol. 260, pp. 185–199, 2014.

[35] C. Muelder and K. Ma, "A treemap based method for rapid layout of large graphs," in *PacificVis*. IEEE Comp. Soc., 2008, pp. 231–238.

[36] S. Zhao, M. J. McGuffin, and M. H. Chignell, "Elastic hierarchies: Combining treemaps and node-link diagrams," in *INFOVIS*. IEEE Computer Society, 2005, pp. 57–64.

[37] "CiteVis: Visualizing citations among InfoVis conference papers," http://www.cc.gatech.edu/gvu/ii/citevis.

[38] C. A. Duncan and M. T. Goodrich, "Planar orthogonal and polyline drawing algorithms," in *Handbook of Graph Drawing and Visualization*. Chapman and Hall/CRC, 2013, pp. 223–246.

[39] F. Beck, M. Burch, S. Diehl, and D. Weiskopf, "A taxonomy and survey of dynamic graph visualization," *Comput. Graph. Forum*, vol. 36, no. 1, pp. 133–159, 2017.

[40] S. Hadlak, H. Schulz, and H. Schumann, "In situ exploration of large dynamic networks," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2334–2343, 2011.

[41] S. Rufiange and M. J. McGuffin, "Diffani: Visualizing dynamic graphs with a hybrid of difference maps and animation," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2556–2565, 2013.

[42] P. Angelini and G. Da Lozzo, "Beyond clustered planar graphs," in *Beyond Planar Graphs, Communications of NII Shonan Meetings*, S. Hong and T. Tokuyama, Eds. Springer, 2020, pp. 211–235.

[43] J. J. Besa Vial, G. Da Lozzo, and M. T. Goodrich, "Computing k-modal embeddings of planar digraphs," in *ESA*, ser. LIPIcs, vol. 144. Dagstuhl Publishing, 2019, pp. 19:1–19:16.

[44] E. Di Giacomo, G. Liotta, M. Patrignani, and A. Tappini, "NodeTrix planarity testing with small clusters," in *Proc. 25th International Symposium on Graph Drawing and Network Visualization (GD '17)*, ser. LNCS, F. Frati and K. Ma, Eds., vol. 10692, 2018, pp. 479–491.

[45] H. C. Purchase, R. F. Cohen, and M. I. James, "An experimental study of the basis for graph drawing algorithms," *ACM J. Exp. Algorithmics*, vol. 2, p. 4, 1997.

[46] H. C. Purchase, "Performance of layout algorithms: Comprehension, not computation," *J. Vis. Lang. Comput.*, vol. 9, no. 6, pp. 647–657, 1998.

[47] M. Burch, W. Huang, M. Wakefield, H. C. Purchase, D. Weiskopf, and J. Hua, "The state of the art in empirical user evaluation of graph visualizations," *IEEE Access*, vol. 9, pp. 4173–4198, 2021.

[48] A. Abuthawabeh, F. Beck, D. Zeckzer, and S. Diehl, "Finding structures in multi-type code couplings with node-link and matrix visualizations," in *VISSOFT*. IEEE Comp. Society, 2013, pp. 1–10.

[49] B. Alper, B. Bach, N. H. Riche, T. Isenberg, and J. Fekete, "Weighted graph comparison techniques for brain connectivity analysis," in *CHI*. ACM, 2013, pp. 483–492.

[50] J. Christensen, J. Bae, B. Watson, and M. Rappa, "Understanding which graph depictions are best for viewers," in *Smart Graphics*, ser. LNCS, vol. 8698. Springer, 2014, pp. 174–177.

[51] W. Didimo, E. M. Kornaropoulos, F. Montecchiani, and I. G. Tol-lis, "A visualization framework and user studies for overloaded orthogonal drawings," *Comput. Graph. Forum*, vol. 37, no. 1, pp. 288–300, 2018.

[52] M. Ghoniem, J. Fekete, and P. Castagliola, "On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis," *Inf. Vis.*, vol. 4, no. 2, pp. 114–135, 2005.

[53] R. Keller, C. M. Eckert, and P. J. Clarkson, "Matrices or node-link diagrams: which visual representation is better for visualising connectivity models?" *Inf. Vis.*, vol. 5, no. 1, pp. 62–76, 2006.

[54] M. Okoe and R. Jianu, "GraphUnit: Evaluating interactive graph visualizations using crowdsourcing," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 451–460, 2015.

[55] M. Okoe, R. Jianu, and S. G. Kobourov, "Node-link or adjacency matrices: Old question, new insights," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 10, pp. 2940–2952, 2019.

[56] N. Henry and J. Fekete, "MatLink: Enhanced matrix visualization for analyzing social networks," in *INTERACT (2)*, ser. Lecture Notes in Computer Science, vol. 4663. Springer, 2007, pp. 288–302.

[57] L. Angori, W. Didimo, F. Montecchiani, D. Pagliuca, and A. Tap-pini, "ChordLink: A new hybrid visualization model," in *Graph Drawing*, ser. LNCS, vol. 11904. Springer, 2019, pp. 276–290.

[58] G. Liotta, I. Rutter, and A. Tappini, "Simultaneous FPQ-ordering and hybrid planarity testing," in *SOFSEM*, ser. LNCS, vol. 12011. Springer, 2020, pp. 617–626.

[59] "NodeTrix Javascript," https://github.com/IRT-SystemX/nodetrix.

[60] B. Lee, C. Plaisant, C. S. Parr, J. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proceedings of BELIV 2006*, 2006, pp. 1–5.

[61] R. A. Amar, J. Eagan, and J. T. Stasko, "Low-level components of analytic activity in information visualization," in *IEEE Symposium on Information Visualization (InfoVis 2005)*, 2005, pp. 111–117.

[62] H. C. Purchase, *Experimental Human-Computer Interaction - A Practical Guide with Visual Examples*. Cambridge Univ. Press, 2012.

[63] H. C. Purchase, J. Hamer, M. Nöllenburg, and S. G. Kobourov, "On the usability of lombardi graph drawings," in *Graph Drawing - 20th International Symposium, GD 2012*, ser. LNCS, W. Didimo and M. Patrignani, Eds., vol. 7704. Springer, 2012, pp. 451–462.

[64] B. Saket, P. Simonetto, S. G. Kobourov, and K. Börner, "Node, node-link, and node-link-group diagrams: An evaluation," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2231–2240, 2014.

[65] G. Melançon, "Just how dense are dense graphs in the real world? A methodological note," in *Proceedings of BELIV 2006*, 2006, p. 1–7.

[66] R. E. van Dijk, J. C. Kaden, A. Argüelles-Ticó, D. A. Dawson, T. Burke, and B. J. Hatchwell, "Cooperative investment in public goods is kin directed in communal nests of social birds," *Ecology Letters*, vol. 17, no. 9, pp. 1141–1148, 2014.

[67] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI*, 2015. [Online]. Available: http://networkrepository.com

[68] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proceedings of the National Academy of Sciences*, vol. 100, no. 21, pp. 11 980–11 985, 2003.

[69] M. Ley, "The DBLP computer science bibliography." [Online]. Available: https://dblp.uni-trier.de

[70] M. Bostock, V. Ogievetsky, and J. Heer, "D³ Data-Driven Documents," *IEEE TVCG*, vol. 17, no. 12, pp. 2301–2309, 2011.

[71] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola, "Fast optimal leaf ordering for hierarchical clustering," vol. 17, no. suppl_1, 06 2001, pp. S22–S29.

[72] C. Bachmaier, H. Buchner, M. Forster, and S.-H. Hong, "Crossing minimization in extended level drawings of graphs," *Discrete Applied Mathematics*, vol. 158, no. 3, pp. 159–179, 2010.

[73] C. Matuszewski, R. Schönfeld, and P. Molitor, "Using sifting for *k*-layer straightline crossing minimization," in *Graph Drawing, 7th International Symposium, GD'99*, 1999, pp. 217–224.

[74] W. J. Conover, *Practical nonparametric statistics*, ser. Wiley series in probability and mathematical statistics. Wiley, 1980.

[75] H. C. Thode, *Testing for normality*. Marcel Dekker, 2002.

[76] W. Didimo, F. Giordano, and G. Liotta, "Overlapping cluster planarity," *J. Graph Algor. Appl.*, vol. 12, no. 3, pp. 267–291, 2008.

[77] H. Omote and K. Sugiyama, "Force-directed drawing method for intersecting clustered graphs," in *APVIS*. IEEE Computer Society, 2007, pp. 85–92.

Emilio Di Giacomo received the PhD in Information Engineering from the University of Perugia in 2003. He is associate professor at the University of Perugia. His research interests include graph drawing, network visualization, and algorithm design. He collected about 120 scientific publications in the above areas. He is publication editor of the J. of Graph Algorithms and Appl., and Steering Committee member of the Graph Drawing and Network Visualization Symposium.

Walter Didimo received the PhD in Computer Engineering from the University of Rome "La Sapienza" in 2000. He is associate professor at the University of Perugia. His research interests include graph drawing, network visualization, and algorithm design. He collected more than 150 scientific publications in the above areas. He is associate editor of the IEEE Access journal and served in the Steering Committee of the Symposium on Graph Drawing and Network Visualization.

Giuseppe Liotta received his PhD degree from Sapienza, University of Rome. He is a professor of computer science at the University of Perugia. His research interests include algorithm design, computational geometry, network visualization, and graph drawing. On these topics he edited special issues, wrote surveys papers, book chapters, and published about 300 research papers. He currently serves and the editor in chief of Computer Science Review and of the J. of Graph Algorithms and Applications.

Fabrizio Montecchiani received the PhD in Information Engineering in 2014 from the University of Perugia, where he currently works as associate professor. His research interests include graph drawing, computational geometry, visual analytics, and Big Data algorithms. He collected more than 100 scientific publications in the above areas. He has been guest editor of the J. of Graph Algorithms and Applications, and he has been a PC member of international conferences such as EuroCG, GD, MFCS, and WG.

Alessandra Tappini received the PhD in Information Engineering from the University of Perugia in 2020. She is currently a postdoctoral researcher at the same university. Her research interests include graph drawing, graph algorithms, and network visualization. During her PhD program, she spent some months in the Algorithmics group of the University of Tübingen. She has been a PC member of the 30th International Symposium on Graph Drawing and Network Visualization (GD 2022).