Detangler: Visual Analytics for Multiplex Networks — **Source link** ⧉

Benjamin Renoust, Guy Melançon, Tamara Munzner

Institutions: National Institute of Informatics, University of Bordeaux, University of British Columbia

Related papers:

- Entanglement in Multiplex Networks: Understanding Group Cohesion in Homophily Networks

- Multilayer Networks

- Structural measures for multiplex networks.

- Network Analysis in the Social Sciences

- A Nested Model for Visualization Design and Validation

Detangler: Visual Analytics for Multiplex Networks

Benjamin Renoust, Guy Melançon, Tamara Munzner

HAL Id: hal-01467984
https://hal.archives-ouvertes.fr/hal-01467984

Submitted on 15 Feb 2017

Detangler: Visual Analytics for Multiplex Networks

B. Renoust[1,2] and G. Melançon[2] and T. Munzner[3]

[1]National Institute of Informatics & JFLI CNRS UMI 3527, Tokyo, Japan
[2]University of Bordeaux & LaBRI CNRS UMR 5800, Bordeaux, France
[3]University of British Columbia, Vancouver, Canada

Abstract

A multiplex network has links of different types, allowing it to express many overlapping types of relationships. A core task in network analysis is to evaluate and understand group cohesion; that is, to explain why groups of elements belong together based on the underlying structure of the network. We present Detangler, a system that supports visual analysis of group cohesion in multiplex networks through dual linked views. These views feature new data abstractions derived from the original multiplex network: the substrate network and the catalyst network. We contribute two novel techniques that allow the user to analyze the complex structure of the multiplex network without the extreme visual clutter that would result from simply showing it directly. The harmonized layout visual encoding technique provides spatial stability between the substrate and catalyst views. The pivot brushing interaction technique supports linked highlighting between the views based on computations in the underlying multiplex network to leapfrog between subsets of catalysts and substrates. We present results from the motivating application domain of annotated news documents with a usage scenario and preliminary expert feedback. A second usage scenario presents group cohesion analysis of the social network of the early American independence movement.

Categories and Subject Descriptors (according to ACM CCS): Visualization [Human-centred computing]: Visualization Techniques—Graph drawings

1. Introduction

In the exploratory visual analysis of network data, the high-level task of detecting structural patterns often involves the lower-level abstract task of understanding groups. Understanding and explaining the reasons that elements of the same group belong together can be framed as many related questions: *What makes a group a whole? Why are these entities tied together? How do different groups compare to each other? Are all groups equally homogeneous?*

Our motivation to address this problem came from working with archivists at INA, the French national institute responsible for archiving and reselling all news material broadcast within France over the past several decades. They must create groups of documents in response to requests from a diverse set of potential users including television and print journalists, documentary filmmakers, and contemporary talk show creators. These groups are based on many possible criteria, including people, places, and topics. The archivists need to validate that the proposed groups are a

good match with these requests, and thus are interested in examining the group structure in detail.

For example, in a collection of news documents such as newspaper articles and TV news excerpts, links between documents can be inferred if they share similar content. There are several different ways to group documents together: according to whether they concern the same people, or describe the same places, or are about the same topics. Each of these ways of grouping them yields a different set of link relationships. To precisely capture the full structure of the relationships between documents, the network should have multiple layers, where the links induced by sharing one kind of similarity are assigned to a different layer than those for a different similarity measure. This kind of network, where there are links of different types, is called a **multiplex network** [DDSRC*13].

The task of analyzing group cohesion is difficult with multiplex networks because they express many overlapping types of relationships simultaneously. This problem is well suited for interactive visual analysis through a combination

of interactive exploration initiated by the user and sophisticated computational support; in short, visual analytics. The combination of the data abstraction of multiplex networks and the task abstraction of group cohesion analysis arises in many domain contexts from bioinformatics to digital humanities. In this paper we present two usage scenarios: an annotated news document example from the archivists who originally piqued our interest in this problem, and a social network analysis example using data from the early American independence movement.

The first contribution of this paper is the design and implementation of Detangler, a system to support the interactive visual analysis of multiplex networks. A second contribution is the new data abstraction of two networks derived from the original multiplex network: the **substrate** network contains the entities that are the focus of exploration, and the **catalyst** network captures the connections between these entities. In the motivating example above, the substrates are documents, and they are linked through catalyst topics. The third contribution is a visual encoding technique that supports spatial stability between the views of the substrate and catalyst networks through **harmonized layout**. The fourth contribution is an interaction technique that allows users to **leapfrog** between the substrate and catalyst views by pivoting across a selection in one view to see linked highlighting in the other based on computations carried out in the underlying multiplex network.

2. Multiplex Networks

To illustrate the ideas behind multiplex networks, Figure 1(a) has the concrete example of a bibliographic dataset composed of papers, authors, and keywords that describe paper topics. A simple way to model this network would be the traditional **monoplex** network where all links are of a single type, such as the co-authorship network shown in Figure 1(b) where nodes are authors and links are papers; thus, people who have published a paper together are connected.

The more complete way to model this dataset is as a **multiplex** network, where links can have different types. Alternate names for this kind of network include *heterogeneous* [ABHR*13], *multilayer* [HD11], and *multimodal* [GKL*13]. In our bibliographic example, families of links correspond to keywords in common between specific coauthors, as shown in Figure 1(c). A straightforward way to visually encode these networks is to visually distinguish the different link types with colors, resulting in superimposed visual layers. This approach has at least two challenges in terms of scalability. The upper limit on the number of link types is around ten because of the well-known limits on the distinguishability of small regions of non-contiguous colors [Mun14][Section 10.3.1]. The upper limit for the number of nodes and links is not as crisply defined, but visual clutter can be problematic even with dozens of items; the representation does not scale to hundreds of items.

In the new data abstraction that we propose, the visual encoding and interaction takes place via two networks derived from this multiplex network, as shown in Figure 1(d). The **substrates** are the nodes in the original multiplex network, and any pair of nodes that is connected in the multiplex network is connected with a single link in the substrate network. This simple example has a fully connected substrate network, but in general any structure is possible.

The **catalysts** are created from the multiplex links, and can be considered as nodes in a new network. Catalysts are linked in this network if there are co-occurring links of the same type between two nodes in the original multiplex network, as shown in Figure 1(e). Catalysts can be understood as the reasons why substrates are linked to one another.

A single node in this catalyst network corresponds to all links of the same type in the original multiplex network, which would be a visually complex layer in the obvious representation. Our proposed abstraction thus has a similar spirit to the scagnostics abstraction for exploring scatterplot matrices [WAG05], where each node in the meta-scatterplot represents an entire scatterplot in the original dataset. It also shares the spirit of the PivotGraph representation [Wat06], where the catalyst graph acts as a visually compact roll-up of a complex structure. The full power of this dual-network abstraction is unleashed with carefully designed interaction techniques so that the user can quickly drill down by leapfrogging between different subsets of interest, as described in Section 4.

This data abstraction allows the general questions about group membership that we identified above to be expressed as more specific questions about these two new networks: *How do the substrates interact? How much do the catalysts interact? What is the shape of the catalyst network? How do these two networks fit together?*

2.1. Entanglement Measures

We build on the previous work of three entanglement measures proposed by Renoust et al. [RMV14]. One of these operates on individual catalysts: the **entanglement index** for *each* catalyst indicates how well each catalyst mixes with others. The other two operate on an entire subgraph of any size, up to the entire network. The **entanglement intensity** of an *entire subgraph* indicates how much *all* of the catalysts in that group mix together. Roughly speaking, group cohesion is more intense when substrates are all connected for the same reasons – that is, via catalysts of the same type – than if they are connected through catalysts of different types. The analysis of cohesion should also consider whether catalysts are invoked equally within a group. The **entanglement homogeneity** of an *entire subgraph* indicates how balanced the mix of *all* catalysts together is across the whole group.
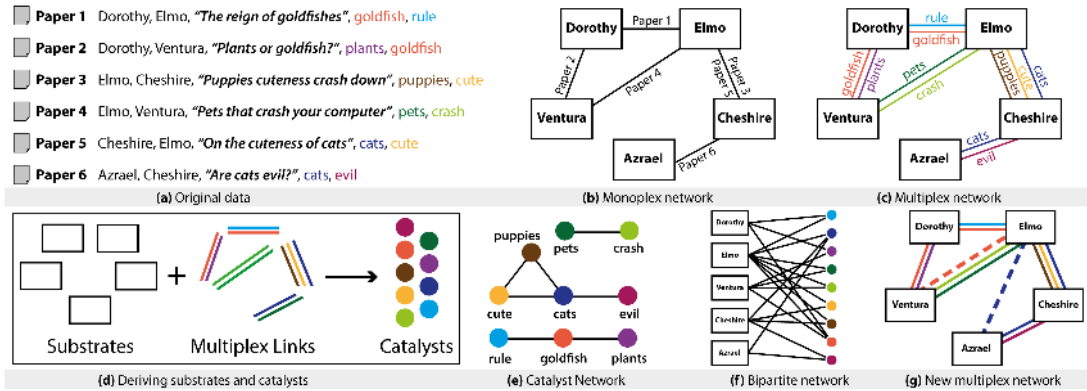
Figure 1: Network abstractions. (a) Bibliographic example with papers, authors, and keywords. (b) Monoplex network with papers as the only link type. (c) Multiplex network with multiple link types, one for each keyword. (d) Deriving substrates and catalysts. (e) Catalyst network where keywords are linked if they co-occur as multiplex links. (f) Bipartite networks handles nodes of two types, for example authors and keywords, but lose information. (b) Mapping this bipartite representation back into a multiplex network introduces incorrect links.

2.2. Comparison to Other Abstractions

Our abstraction distinguishes substrate nodes from catalyst nodes. The abstraction of a bipartite network is a well-known way to model networks with two types of nodes, and an obvious question is whether a bipartite representation would suffice for the group cohesion analysis task. It would not suffice because simply projecting a multiplex network down into a bipartite network loses crucial information. The bipartite network shown in Figure 1(f), where one node type is authors and the other node type is keywords, illustrates the problem: its multiplex equivalent is not equivalent to the original multiplex network, as shown in Figure 1(g). The multiplex network that is reconstructed from this bipartite network contains incorrect links: Elmo and Azrael have not co-authored a paper, even though they each have written papers on the topic of cats. These two representations are only equivalent in the case where the association between substrates and catalysts is fully symmetric, for example if the links correspond to similarity between nodes. Many interesting use cases have asymmetric links, including this bibliographic example.

Our abstraction also is more powerful than simply projecting the multiplex network into a weighted monoplex network. While these networks are much easier to visualize, they also lose a great deal of crucial information. Figure 2 shows an example of two different multiplex networks that project to identical substrate networks but have completely different catalyst network topologies: ring vs. star.

3. Related Work

Some previous work on multiplex network visualization uses straightforward visual encoding approaches based on layering and partitioning that do not scale. For example, Schreurs
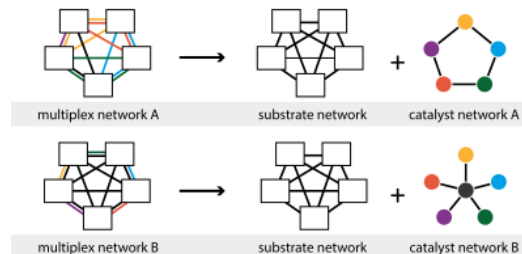

Figure 2: Weighted projections also lose information; these two multiplex networks have identical projected substrate networks but very different catalyst network topologies.

et al. [STF*13] propose distinguishing between multiplex link types with color coding, leading to the problems discussed in Section 2; the unfiltered view suffers from extreme visual clutter and the filtered view shows only a tiny fraction of the available information. De Domenico et al. [DDPA14] propose partitioning the network into multiple layers on tilted 3D planes, leading to the well-known problems associated with the use of 3D space for abstract data such as networks [Mun14][Section 6.3]. These approaches do not provide interaction support for the group cohesion analysis task.

Some previous systems focus on the task of comparison between multiplex networks; while they feature carefully designed visual encodings, they still do not support group cohesion analysis. Alper et al. [ABHR*13] propose the use of weighted projections visually encoded as both node-link views and adjacency matrix views for analyzing brain connectivity. They focus on comparing multiple groups to understand how they differ, whereas we focus on looking at one group to understand what makes it cohesive. Hascoet et al. [HD11] emphasize the combination of superimposed and

side-by-side layers, but again with a focus on comparison between groups rather than cohesion.

Riche et al. propose several scalable techniques for interacting with densely connected graphs [RDLC12], including networks with multiple link types. Our approach handles the problem at the data abstraction level by introducing the catalyst networks, where a single node corresponds to an entire visual layer with many links in the full multiplex network, rather than at the level of interaction techniques applied to that original network.

Several authors have proposed interaction techniques that rely on pivoting between different representations, as does our leapfrog interaction technique. While both NetLens [KPLB07] and GraphTrail [DHRL*12] rely on pivot-based interaction, they do not provide node-link views and thus do not support tasks that require topological understanding, such as the graph cohesion analysis task. The RelEx system [SFMB12] does support pivoting between different node-link views, but it is highly focused on a specific task of overlay network traffic optimization, and does not support the graph cohesion analysis task.

The previous work that comes closest to addressing the group cohesion analysis task is a design study from Ghani et al. [GKL*13]: this task is alluded to as an interesting open problem at the end of the paper. They focus on analyzing multiplex networks that arise in social network analysis, but emphasizing the problem of multiple node types more than of multiple link types. The multiplex network is represented a *k*-partite network, leading to the information loss problems that we characterize in Section 2; moreover, global topological understanding is not supported because node-link views are limited to local neighborhoods.

4. Interaction

The design of the Detangler visual analytics system encompasses both interaction and visual encoding techniques. We present the interaction design in this section, and then cover visual encoding in Section 5.

Detangler features interaction techniques designed to support cohesion analysis for the groups that occur within the underlying multiplex network by supporting drill-down for interactively chosen subsets of items by pivoting between the substrate and catalyst networks. We call the result of pivoting back and forth between these views a **leapfrog** interaction.

4.1. Pivoting From Substrates

Mapping subsets of items between the substrate and catalyst networks in a way that reflects the true structure of the underlying multiplex network requires careful interaction and algorithm design. The complexity of the problem is due to the fact that the association between these networks is neither

transitive nor symmetric. Catalysts are associated with multiplex network edges. Specifically, a catalyst corresponds to the combination of multiplex edges of a particular type and the nodes that they connect in the multiplex network.

Figure 3 shows an example of the pivot pipeline mapping from substrates to catalysts. The user initiates the pivot by selecting a subset of the substrate nodes. The system then finds the subgraph of the underlying multiplex network induced by those nodes; that is, finding all multiplex edges that directly connect any pair of these nodes in the multiplex network. New entanglement measures are computed based on that multiplex subgraph. The group entanglement measures of homogeneity and intensity are computed for that subgraph. Conveniently, the catalyst subgraph that corresponds to that multiplex subgraph is also extracted as a side effect of that computation without any further computational overhead. Moreover, the individual measure of the entanglement index for every single catalyst in the network is computed with respect to that multiplex subgraph, and those new values are used to update all of the node sizes in the catalyst view. When no subgraph is selected, the node sizes show entanglement index values computed with respect to the entire multiplex network.

4.2. Pivoting From Catalysts

An analogous operation of pivoting from catalysts to substrates can be defined, but it is not identical because of the asymmetry in the data abstractions. A single catalyst node corresponds to multiple edges within the multiplex network, so the new set of substrate nodes is generated by traversing these edges within the multiplex network and finding the set of nodes that they connect. Detangler supports two different selection modes, *AND* and *OR*, as shown in Figure 4. The rest of the computation is similar, including the entanglement computation.

4.3. Leapfrogging

The combination of pivoting from substrates to catalysts and then pivoting from catalysts to substrates results in a *leapfrog* interaction, as shown in Figure 5. In a leapfrog, the catalysts corresponding to the group of selected substrates is used as the new selection in the catalyst network. We then obtain a new set of substrates that are connected within the multiplex network by those catalysts. We call this loop a *leapfrog* to emphasize the asymmetric nature of back to back pivot interaction: the set of substrates selected at the end of the leapfrog is not the same as the initial set. For example, when we pivot from a group of authors to the set of keywords in their co-authored papers, we may then want to ask the follow-up question *Who else is publishing papers with these same keywords?*

Leapfrogging using the *AND* operator tends to quickly converge to the empty set where no substrates are selected
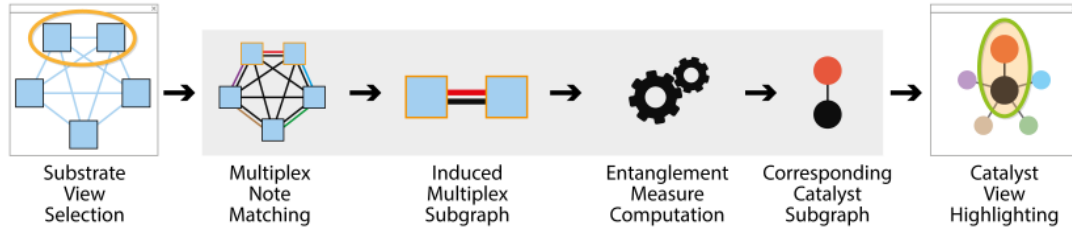
Figure 3: Pivoting from substrates to catalysts. The user selects nodes in the substrate view. Detangler retrieves the corresponding multiplex nodes and creates the induced subgraph of edges that connect them in the multiplex network. It computes the two group entanglement measures of homogeneity and intensity for this subgraph, and along the way finds the corresponding subgraph in the catalyst network. It also computes new individual entanglement index measures for each catalyst node with respect to this subgraph.
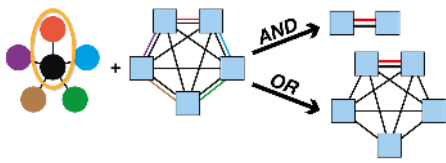

Figure 4: The two modes of pivoting from catalysts. The more restrictive *AND* operator yields only the substrate nodes connected through *all* of the designed catalysts; the more expansive *OR* mode yields the substrates nodes connected through *any* of the catalyst links.
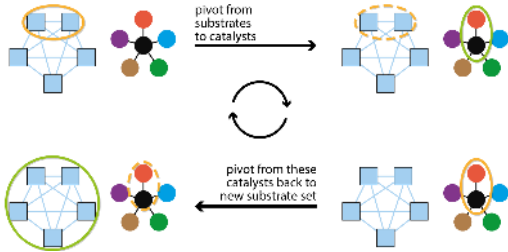

Figure 5: Leapfrogging between substrates and catalysts through back-to-back pivots.

after a small number of pivots, whereas leapfrogging with the *OR* operator converges to the whole set of all possible substrates after a few pivots.

5. Visual Encoding

The Detangler interface is shown in Figure 6. It features two main side-by-side views with coordinated node-link layouts: the substrate network on the left and the catalyst network on the right. Two supporting views are on the far left: an inter-action menu on the bottom and a detail view for the entanglement measures for the current selection on the top.

Our choice of node-link layouts was guided by the two criteria from Ghoniem et al. [GFC04]. From a task point of view, we identify topological structure analysis tasks as a building block of our motivating group cohesion analysis task. From a data point of view, both of these derived networks are considerably more sparse in terms of edge density than the underlying multiplex network, so the visual complexity is manageable. The node positions are determined by our new harmonized layout technique, as discussed in detail in Section 5.3. To emphasize the difference between the views, substrate nodes are encoded as squares and catalyst nodes as circles. Node size in the catalyst view encodes the catalyst entanglement index with respect to the current selection set.

5.1. Lasso

The two derived network views are coordinated with linked highlighting, where the pivot correspondence between views is computed as described in Section 4. The selection itself is handled using the lasso technique proposed by McGuffin and Jurisica [MJ09] that supports fluid selection of either irregularly shaped or rectangular areas. Once drawn, the lasso can be dragged as a brush. Whenever the lasso captures a new node, it is highlighted as a selected node with red fill, and a new entanglement analysis is triggered on the fly. Corresponding nodes in the other view are highlighted with red outlines for the nodes and red coloring for the links. All catalyst nodes are immediately resized according to the newly computed entanglement indices. To emphasize the highlighted items, the other items are dimmed to a transparency level of 50%. Lasso selection can occur in either view. In the discussion below, we will refer to the view used to select, in which we draw the lasso, as the *source* view; the other view is called the *target* view.

The two global entanglement measures, namely intensity and homogeneity, are directly visually encoded onto the lasso using color. These values are normalized to the same domain, ranging from 0 to 1. Figure 6b shows our sequential-segmented color ramp. The main sequence region combines three hues in order of monotonically decreasing luminance.
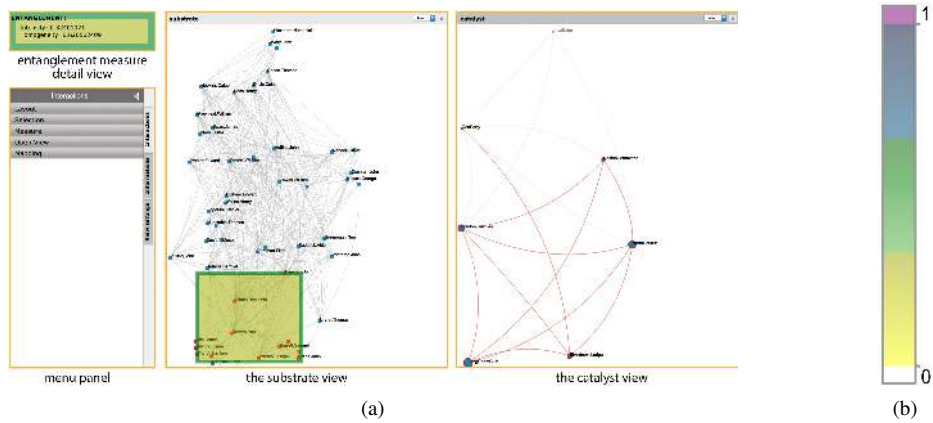
Figure 6: The Detangler interface. (a) The four views are annotated in orange and labelled. The current source is the substrate view on the left with a rectangular lasso showing the user's selection, and the target is the catalyst view on the right showing corresponding highlighting. The two secondary views are on the left: the entanglement measure detail view and the menu panel. (b) Sequential-segmented color ramp. The bottom segment is white; the main sequence has three hues with monotonically decreasing luminance; the top segment is bright purple, with discrete change of hue and luminance for maximal contrast.

At the bottom of the range, low values are shown with low salience as high-luminance yellow, for a subtle luminance change against the white background. The middle of the range is a mid-luminance green hue, and the top of the sequence is a low-luminance dark blue. The scale is segmented at the bottom to white and at the top with a discrete change of both hue and luminance to bright purple, emphasizing maximum values in a salient way compared to the dark blue values near the top of the range.

The entanglement homogeneity value is mapped to the thick perimeter of the lasso, and the entanglement intensity value is shown by the lasso's interior fill color. The entanglement view in the upper left displays these values numerically, and also encodes them as colors in its own perimeter and interior to provide a visual association between the lasso and this secondary view. Figure 6 shows a rectangular lasso, while Figures 7, 9, and 10 show irregularly shaped lassos. The lower left control panel has tabs with additional buttons that allow users to switch the catalyst correspondence operator between *AND* and *OR* modes.

5.2. Leapfrogging

Leapfrogging is performed by double-clicking on a lasso selection, turning the target nodes into the source nodes and changing the visual encoding changes accordingly. This sequence of operations requires four states that must be distinguishably visually encoded. The normal state for items that are not selected in the source view is blue nodes and light grey links. Selected nodes in the source view are red. In the target view, the highlighted state is that nodes are outlined in red and links are red, and the dimmed state reduces the other

items in that view with 50% transparency. Figure 7 shows these encodings in a leapfrog sequence example.

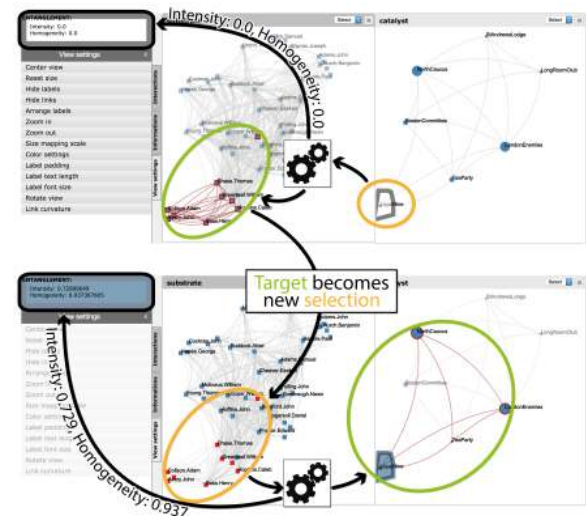Figure 7: Leapfrog sequence showing visual encoding changes between pivot steps, with Detangler screenshots rather than the schematic diagram in Figure 5.

5.3. Harmonized layout

We introduce a harmonized layout technique designed to limit the cognitive load of manipulating multiple views by providing spatial stability so that users can easily match up corresponding items between the two views. Since the two

views contain different entities, designing a layout to support this correspondence requires explicit algorithmic support. Figure 8 illustrates the algorithm. The high-level sequence of operations is:

1. Create a layout for only the catalyst network, and then remove the links between the catalyst nodes, as in Figure 8(a).
2. Divide the catalysts into two classes: low entanglement and high entanglement. The intent is that substrates should be kept close to low-entanglement catalysts that refer to a limited number of specific substrates. In contrast, because they are linked to many substrates, high-entanglement catalysts are only used to resolve ambiguities rather than as primary magnets to avoid hairballs.
3. Add substrate nodes in the layout. Anchor substrates with temporary links to all low-entanglement catalysts that they are connected to. If there are no low-entanglement catalyst to which they can be attached, connect them to all their high-entanglement catalysts, as in Figure 8(b).
4. Pin the catalyst nodes and run a layout algorithm that only moves the anchored substrates, as shown in Figure 8(c). Our implementation uses a standard force directed algorithm in both Steps 1 and 4, but other layouts are also possible.
5. Use the harmonized node positions in the layouts in the separate substrate and catalyst views. The substrate nodes share spatial stability with their significant catalysts, as shown in Figure 9.

6. Implementation

Detangler is implemented on top of two visualization frameworks. We use Tulip [AAB*14] for network and layout computations, and D3 [BOH11] in conjunction with jquery-ui as a front end for rendering the visual objects. The coordination between the two servers is built on top of a server-client architecture.

Detangler benefits from the full suite of Tulip's graph manipulation capabilities, including the ability to use different metrics for layout and coloring, and to carry out graph-theoretic grouping operations such as finding induced subgraphs. It also benefits from all of the interaction and immediate deployment capabilities of D3.

Detangler provides interactive frame rates for multiplex networks containing dozen or hundreds of catalysts and hundreds of substrates. The current scalability bottlenecks are SVG rendering and communication with the server. Achieving further algorithmic scalability would be straightforward by using a different rendering architecture that uses WebGL, such as Tulip itself. The focus with Detangler is the perceptual scalability provided by the dual-view abstraction, where a single visual object captures information that would require an entire visual layer of many objects in previous work.

Detangler is publicly available at http:

//myfoodomain.webcv.info/ (temporary URL for blind review).

7. Results

We present two different usage scenarios for group cohesion analysis with Detangler: our motivating application of annotated news documents from the INA, and an example from social network analysis. (The supplemental video and materials walk through an interactive session in more detail.) We also present results from an expert feedback session midway through our iterative refinement process.

7.1. Usage Scenario: Road Safety

Our first usage scenario is a small network of annotated INA documents that resulted from the query *road safety*. In this multiplex network, the substrates are documents, and the catalysts are annotations on these documents in the form of keywords. The set of 19 documents and 19 keywords are connected through 334 multiplex links, as shown in Figure 9.

This scenario demonstrates how the combination of harmonized layout and pivoting can answer the questions: *Is this group homogeneous? If there are subgroups within it, how cohesive are they they?*

Before the user selects any items, the entanglement detail view displays global entanglement measures of 0.52 for homogeneity and 0.33 for intensity, clearly indicating that the group is not very cohesive at all. The topology of the catalyst network shows two clusters separated by two bridge nodes: the catalysts `speed` and `accidentPrevention`. The cluster on the left consists of 9 catalyst keywords, including `prison`. Lassoing this group reaches 4 documents in the substrate view, and the blue lasso interior reflects the high entanglement values of 0.93/0.72 showing that the corresponding documents form a more cohesive subgroup. Reading these documents showed that they report on a specific news story about a bad driver condemned to prison due to subsequent offenses of exceeding the speed limit.

7.2. Usage Scenario: Paul Revere

The second usage scenario presents the 18th century social network of the American independence movement [Fis94]. In this multiplex network, the substrates are people, and the catalysts are societies. Figure 10 shows the 917 multiplex links connecting 49 people and 7 societies. Following Healy's analysis from a royalist intelligence analyst point of view [Hea13], we refer to these as terrorist cells.

This scenario demonstrates the use of leapfrogging to find people and societies that present specific behaviors. We note that `Paul Revere` is famous to most American schoolchildren, but many of these societies are not. We use Detangler
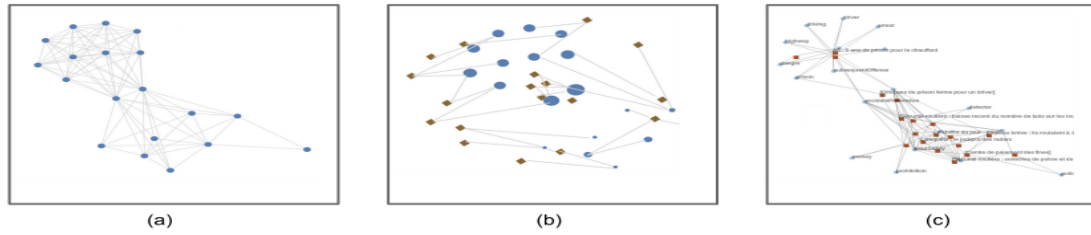
Figure 8: Harmonized layout illustration. (a) Lay out the catalyst network. (b) Remove edges. Assign substrates to catalysts according to two groups based on their entanglement measures. (c) Lay out the substrates without moving the catalysts, prior to separating catalysts and substrates into different views.
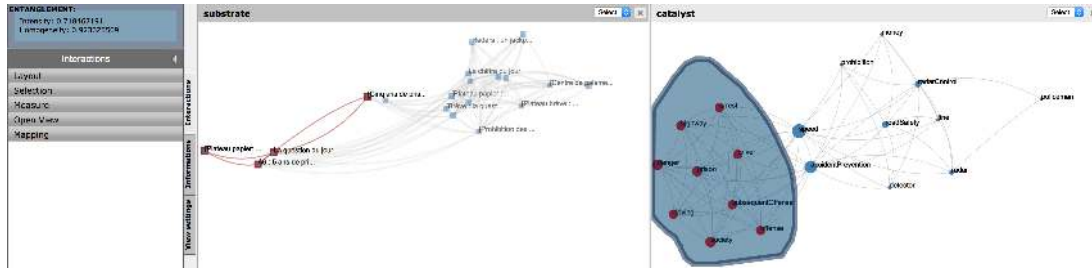

Figure 9: INA dataset of documents concerning speed and road safety. The right catalyst view shows two clusters of topics connected by the bridge nodes `accident prevention` and `speed`. The user has lassoed the cluster on the left as source nodes; the blue interior color shows high entanglement values, indicating that the group is cohesive. The left substrate view shows four documents highlighted as target nodes. The harmonized layout provides spatial stability between these two views.

to analyze the questions *Does the multiplex network structure reveal Paul Revere's importance? Does it reveal influences between these societies?*

In Figure 10(a) the user has lassoed `Paul Revere`. We can observe that he is connected to five societies. A single leapfrog from him through these societies back to the substrate network using the *OR* operator illustrates that he can reach every other person in the network – as he did in his famous night-long ride – because they are all highlighted with red outlines. Figure 10(b) instead shows the result of leapfrogging with *AND* operator: he is the only person still selected, showing that nobody else could have done this long night ride. The multiplex network structure does indeed reveal the basis for Revere's enduring fame.

7.3. Expert feedback session

We ran a preliminary experiment with expert users on an early prototype of Detangler, with a focus on qualitative impressions for formative rather than summative evaluation. We did not conduct quantitative measurements of time and error; a controlled experiment to validate Detangler's utility remains as future work.

We gathered feedback from four experts, archivists from INA, with a separate 3-hour session for each participant. We asked each participant to analyze 12 different INA document

group datasets. Each dataset was composed of between 20 and 80 substrate documents, annotated by between 20 and 150 catalyst keywords. The participants used three different interfaces: Detangler, a Microsoft Excel spreadsheet with title and keywords, and a monoplex network view of documents with edges representing keywords implemented in Tulip. All participants had extensive prior experience with spreadsheet usage and were given a 20-minute training session on the two network-based interfaces. Participants were asked to carry out 4 main tasks: find specific subgroups of documents, find outlier documents, indicate how cohesive the group of documents is, and tell the story of the document group. The experiment was thus within-subjects for tasks and interfaces, with dataset ordering counterbalanced between participants.

At the end of the session, participants rated each interface on Likert scales from 1 to 5 with respect to five criteria: *usage*, namely clarity of how to use the system; *conciseness*, namely whether the interface avoids presenting unnecessary information that leads to cognitive overload; *readability*, namely whether the participant could clearly see the important information; *interaction*, namely whether the interaction was sufficiently fluid to keep participants in a state of flow; and *understanding*, namely whether the interface allowed a good understanding of the group's composition. They were also asked to indicate how confident they felt in conducting an analysis with each tool, as an arbitrary per-
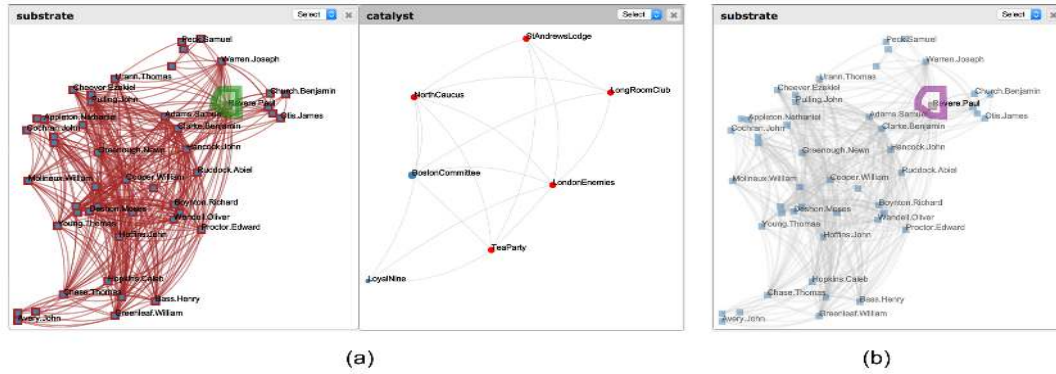
Figure 10: American independence movement dataset. The right catalyst view shows terrorist cells, and the left substrate view shows terrorists such as Paul Revere. (a) Lassoing `Paul Revere` and then leapfrogging with the *AND* operators shows that he can reach every other terrorist in the network, as he did in his famous ride. (b) In contrast, leapfrogging with the *OR* operator shows that he is the only person with connections to all others, and is thus the only person who could have done that ride.

centage. We finished with semi-structured interviews with open-ended qualitative questions.

Figure 11 shows the results, portrayed by Bertifier [PDF14]. U1 had no previous experience with visualization and was rather skeptical about it, and was the only participant who felt more confident with spreadsheets overall. Both U2 and U3 had some previous experience with network and tree visualizations, and U4 had extensive previous experience with using network visualization from a previous project; these three had more confidence analyzing with Detangler than with the other interfaces.

The Usage results provide some preliminary confirmation that multiplex network analysis with Detangler was useable. The pattern for Interaction and Understanding was low ratings for spreadsheets, medium for monoplex networks, and high for multiplex networks. We do note that the Understanding criterion is particularly biased in favor of Detangler, which was designed specifically for the objective of group cohesion analysis.

All participants gave positive feedback during the interviews on the utility of side by side network views, the ease of use of the leapfrog interaction, the informativeness of the catalyst network view, and the match of the entanglement

measures to their mental models of group cohesion. U2 and U4 noted the readability of the harmonized layout, and U4 noted that analysis with Detangler was particularly useful for the open-ended storytelling task. Quotes include (English translations):

- *U1: I did not know much about graph visualization, but even if the system is complex, I like that it is intuitive to learn.*
- *U2: I like that we can associate the documents from the terms.[...] The catalyst view summarizes well what happens in the documents.*
- *U3: It is like a word-cloud but better because it gives a sense of a hierarchy between terms.*
- *U4: The catalyst view makes it easy to isolate the different concepts concerning the group of documents.*

Some negative feedback about limitations was subsequently addressed in our iterative refinement process, including better label management (all) and a more fluid way to trigger leapfrogging (U3). Some feature requests were not addressed, including displaying more descriptive statistics (U1, U2, U3) and better workflow integration with export and history support (U4).

8. Conclusion and Future Work

We have presented Detangler, a visual analytics system for group cohesion analysis. Our novel data abstraction supports scalable analysis with the derived data of substrate and catalyst networks; they are linked together through entanglement computations on the underlying multiplex network. All links of the same type, which are traditionally represented as a full visual layer with complex structure, are rolled up into a single node in the derived catalyst network. Our harmonized layout technique provides spatial stability between the substrate and catalyst networks, and our leapfrog interaction technique connects them through asymmetric pivoting between selected subsets in each.

	SPREADSHEET				MONOPLEX				MULTIPLEX			
	U1	U2	U3	U4	U1	U2	U3	U4	U1	U2	U3	U4
USAGE	■	▬	■	■	■	■	■	■	■	■	■	■
CONCISENESS	■	▬	▬	■	■	■	■	■	■	■	■	■
READABILITY	■	▬	▬	■	▬	■	■	▬	■	■	■	■
INTERACTION					▬	■	■	▬	■	■	■	■
UNDERSTANDING					■	■	▬	■	■	■	■	■
CONFIDENCE	■	■	■	■	■	■	■	■	■	■	■	■

Figure 11: Subjective ratings and overall confidence.

An interesting open question is how to extend our approach from the analysis of *single-multiplex* to *multiple-multiplex* networks, where there is still a single set of substrates but multiple categories of catalysts, as requested by one of the expert feedback session participants. Another open problem is the management of dynamic multiplex networks, with specific interaction designed to support the exploration of time-related information.

References

[AAB*14] AUBER D., ARCHAMBAULT D., BOURQUI R., DE-LEST M., DUBOIS J., PINAUD B., LAMBERT A., MARY P., MATHIAUT M., MELANÇON G.: Tulip III. In *Encyclopedia of Social Network Analysis and Mining*, Alhajj R., Rokne J., (Eds.). Springer, 2014, pp. 2216–2240. 7

[ABHR*13] ALPER B., BACH B., HENRY RICHE N., ISEN-BERG T., FEKETE J.-D.: Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2013), ACM, pp. 483–492. 2, 3

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2301–2309. 7

[DDPA14] DE DOMENICO M., PORTER M. A., ARENAS A.: Multilayer analysis and visualization of networks. *arXiv preprint arXiv:1405.0843* (2014). 3

[DDSRC*13] DE DOMENICO M., SOLÈ-RIBALTA A., COZZO E., KIVELÄ M., MORENO Y., PORTER M. A., GÒMEZ S., ARENAS A.: Mathematical formulation of multi-layer networks. *arXiv preprint arXiv:1307.4977 physics.soc-ph* (2013). 1

[DHRL*12] DUNNE C., HENRY RICHE N., LEE B., METOYER R., ROBERTSON G.: GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2012), ACM, pp. 1663–1672. 4

[Fis94] FISCHER D. H.: *Paul Revere's ride*. Oxford University Press, 1994. 7

[GFC04] GHONIEM M., FEKETE J.-D., CASTAGLIOLA P.: A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)* (2004), IEEE, pp. 17–24. 5

[GKL*13] GHANI S., KWON B. C., LEE S., YI J. S., ELMQVIST N.: Visual analytics for multimodal social network analysis: A design study with social scientists. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2032–2041. 2, 4

[HD11] HASCOËT M., DRAGICEVIC P.: *Visual comparison of document collections using multi-layered graphs*. Tech. rep., LIRMM-00601851, 2011. 2, 3

[Hea13] HEALY K.: Using metadata to find Paul Revere. http://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/, 2013. 7

[KPLB07] KANG H., PLAISANT C., LEE B., BEDERSON B. B.: NetLens: iterative exploration of content-actor network data. *Information Visualization 6*, 1 (2007), 18–31. 4

[MJ09] MCGUFFIN M. J., JURISICA I.: Interaction techniques for selecting and manipulating subgraphs in network visualizations. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 937–944. 5

[Mun14] MUNZNER T.: *Visualization Analysis and Design*. A K Peters Visualization Series, CRC Press, 2014. 2, 3

[PDF14] PERIN C., DRAGICEVIC P., FEKETE J.: Revisiting Bertin matrices: New interactions for crafting tabular visualizations. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2082–2091. 8

[RDLC12] RICHE N. H., DWYER T., LEE B., CARPENDALE S.: Exploring the design space of interactive link curvature in network diagrams. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI)* (2012), ACM, pp. 506–513. 4

[RMV14] RENOUST B., MELANÇON G., VIAUD M.-L.: Entanglement in multiplex networks: understanding group cohesion in homophily networks. *Social Network Analysis - Community Detection and Evolution XVIII*, 274 (2014). 2

[SFMB12] SEDLMAIR M., FRANK A., MUNZNER T., BUTZ A.: RelEx: Visualization for actively changing overlay network specifications. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2729–2738. 4

[STF*13] SCHREURS B., TEPLOVS C., FERGUSON R., DE LAAT M., BUCKINGHAM SHUM S.: Visualizing social learning ties by type and topic: rationale and concept demonstrator. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (2013), ACM, pp. 33–37. 3

[WAG05] WILKINSON L., ANAND A., GROSSMAN R.: Graph-theoretic scagnostics. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)* (2005), IEEE Computer Society, pp. 157–164. 2

[Wat06] WATTENBERG M.: Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2006), ACM, pp. 811–819. 2