DOI: 10.1111/cgf.12093
Eurographics Conference on Visualization (EuroVis) 2013
B. Preim, P. Rheingans, and H. Theisel
(Guest Editors)

Volume 32 (2013), Number 3

Visual Analysis of Set Relations in a Graph

Panpan Xu[1], Fan Du[2], Nan Cao[3], Conglei Shi[1], Hong Zhou[4]and Huamin Qu[1]

[1]Hong Kong University of Science and Technology, Hong Kong, China
[2]Zhejiang University, China
[3] IBM T. J. Watson Research Center, USA
[4]Shenzhen University, China

Abstract

Many applications can be modeled as a graph with additional attributes attached to the nodes. For example, a graph can be used to model the relationship of people in a social media website or a bibliographical dataset. Meanwhile, additional information is often available, such as the topics people are interested in and the music they listen to. Based on this additional information, different set relationships may exist among people. Revealing the set relationships in a network can help people gain social insight and better understand their roles within a community. In this paper, we present a visualization system for exploring set relations in a graph. Our system is designed to reveal three different relationships simultaneously: the social relationship of people, the set relationship among people's items of interest, and the similarity relationship of the items. We propose two novel visualization designs: a) a glyph-based visualization to reveal people's set relationships in the context of their social networks; b) an integration of visual links and a contour map to show people and their items of interest which are clustered into different groups. The effectiveness of the designs has been demonstrated by the case studies on two representative datasets including one from a social music service website and another from an academic collaboration network.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentations]: User Interfaces—Graphical user interfaces (GUI)

1. Introduction

Set relations appear in various contexts in data analysis. In this paper, we focus on the visual representation of set relations in a social network where each person (node) is related to a set of items. This study is motivated by the increasing popularity of social websites such as Twitter and Facebook where people take an interest in different topics and share their favorite readings or music. Meanwhile, other datasets such as bibliographic database can also be described in the same way as it contains academic collaboration networks and researchers have different ranges of research interests.

This kind of data poses many interesting problems for visualization. In particular, we intend to develop effective visual means for 1) studying the correlation between set relations and topological distances in the social network; and 2) observing the distribution and overlaps of the sets with respect to the clusters of the items. The methods can be used to observe the effect of homophily, i.e., "similarity breeds connection" or "birds of a feather flock together" [MSLC01]. Meanwhile, visualizing the set relations over the item clusters can reveal the "traces" of individuals or groups of people in the information context where the items (such as topics or music) are organized according to their semantic relations (i.e. similarity). In summary, the set relations can be observed from two complementary perspectives, in the context of a social network and in the context of item clusters.

Many visualization techniques for set relations have been developed. Existing set visualizations can be roughly divided into two categories depending on the most important relation it intends to depict: (1) those that utilize spatial positioning to encode some primary dimension (such as geographical location) of the items, where the set relations are sketched on top of the visualization with enclosing contours or continuous lines. This includes Bub-

ble Sets [CPC09], LineSets [ARRC11] and most recently Kelp diagrams [DvKSW12]; (2) those that emphasize set relations such as overlappings and subset/superset relations while other semantic relations between the items are not taken into account. This category of visualizations includes Venn diagram and Euler diagram [SAA09] [RD10]. To the best of our knowledge, there have been no previous visual means for visualizing set relations in the context of a social network.

In this paper, we propose visual designs depicting the set relations in a social network from two perspectives: 1) a node-link view with nodes represented by glyphs showing correlations between the social distances and the set relations (subset / superset / overlap); and 2) a visual design delineating sets on a substrate visualization (e.g. contour map) which shows the clusters of all the items given their similarities, and allows viewers to observe the distribution of the items in the sets. We apply the above techniques to two real datasets. These include one from a social music service website and another from an academic collaboration network.

In summary, the major contributions of this paper are:

- a glyph design that facilitates the analysis of homophily in social networks where each node corresponds to a set of items;
- a visual design and layout method for set relation visualization with respect to clusters of items;
- two case studies based on real datasets that demonstrate how the use of the above two visualizations can lead to interesting findings in some application domains including online social networks and academic collaboration networks.

The rest of the paper is organized as follows. We first introduce the related work in Section 2. The overview of our system is provided in Section 3, followed by the details of the visualization design in Section 4 and the implementation in Section 5. We then present the experiment results on two datasets and discuss the limitations of our method in Section 6. Finally, in Section 7, we conclude the paper and suggest some future research directions.

2. Related Work

Our work draws on research in several categories. In this section, we first review the current existing visualization techniques for set relations. Then we discuss some recent research works on the visual analysis of graphs.

2.1. Set Relationship Visualization

The representations of set relationships have been studied from the very early days. Euler diagram and Venn diagram have been used extensively. This problem has also received attentions from visualization researchers.

Collins et al. [CPC09] presented Bubble Sets which uses bubble-like shapes to connect items belonging to the same set. The Bubble Sets approach is especially effective to reveal set relations over items which have fixed layouts like maps. Similar to the Bubble Sets, the LineSets method [ARRC11] uses smooth lines to connect items in the same set where the items also have fixed layouts. Dinkla et al. [DvKSW12] developed the Kelp diagram, which is a most recent algorithm for set visualization over preallocated items. The algorithm employs edge routing in order to avoid misleading crossovers or wrong set item inclusions.

Another line of set visualization techniques does not assume a fixed layout of the items. The items are spatially grouped such that the relations between sets are more recognizable. Simonetto et al. [SAA09] and Stapleton et al. [SRHZ11] developed fully automatic methods to generate Euler-like diagrams for the visualization of overlapping sets. Riche and Dwyer [RD10] hierarchically organized the intersecting sets such that the Euler diagrams can be more easily drawn.

Moreover, sets can be interpreted as hyperedges in a hypergraph, where the items are the vertices and each hyperedge could consist of multiple vertices. Methods for drawing hypergraphs have been studied in the graph drawing community [JP87] [Mäk90] [BE01]. Researchers have also investigated the existence of various types of support that would be applied in the drawing of hypergraphs. These include planar [KKS09], path-based [BCPS12], and cactus support [BCPS11].

In this paper, we propose a composited visual design for visualizing set relations with respect to the clusters of items in the data. We also propose methods to address the trade-off between the geometrical simplicity of the visual links denoting each set and the preservation of the item locations with respect to their corresponding clusters.

2.2. Visual Analysis of Social Networks

Many of the current visual analysis systems focus on depicting the topologies of the graph structures [vLKS*11] through node-link diagrams, adjacency matrices, or combinations of the two. Some integrate statistic information to enable more effective visual detection [WFC*06] [PS08] [BCD*10] [KMSH12]. Many of the statistics are node metrics derived from either local topological properties (e.g. node degree, clustering coefficient) or global structures (e.g. betweenness centrality). The metrics are directly encoded as visual attributes such as spatial position or color.

One of the future directions for graph visualization as noted in a recent survey [vLKS*11] is the integration of various data types in the visual analysis of graphs. Research efforts have focused on the visualization of heterogeneous relations (graph involving multiple types of nodes and relations) [CSL*10] [DRRD12], and graphs with node attributes [SA06] [Wat06]. We identify that in real social network data,

each person could be associated with a set of items. For example, in academic collaboration networks, each researcher is interested in different topics, and in online social networks, each user could be affiliated to a wide range of items such as the music he listens to and the movies he watches. In this kind of social network, it would be of interest to study the effect of homophily [MSLC01], which means that social links tend to exist among persons with similar characteristics. In this paper, we propose a novel glyph based visual representation for studying the correlation between the set relations and social distances.

3. Overview

In this section we present our research problems, and give an overview of the visual designs in our system.

3.1. Research Problems

We design the visualization such that they can be used to visually analyze the set relations in a social network. In particular, we identify the following research questions:

- Does the distribution of people's interests tend to be localized in a social network? Can we observe the effect of homophily? For example, in an academic collaboration network, Do researchers who never collaborate have different research interests? In an online social network, do friendship links correlate to similar interests?
- For several persons (or groups), can we observe the distributions and overlaps of their interests with respect to the clusters of items? Combining the set relations with the cluster information could be beneficial since the item clusters could provide contextual information for analyzing set relations. For example, we can get some sense of which clusters contain a lot of set intersections and in which clusters the items belong uniquely to some sets. Moreover, when there is little overlap between the interests of two persons, it is possible that items from different sets could belong to the same clusters, thus suggesting an implicit relation between the sets of interests of two persons?

3.2. Basic Idea

We propose two visual designs to address the research questions mentioned above. These include a glyph design, which, when combined with the node-link diagram, can reveal people's set relationship in the context of the social network (Fig.1 left), and an integration of visual links and a contour map to show the items of interest for several people with respect to the clusters of items (See Fig.1 right).

The glyph design can be integrated to a node-link view of the social network, and this can provide an overview of the set relations in the social network. Each glyph encodes the set relations of a person with all others, together with their distances in the graph. The glyph design shows the correlation between social distances and set relations and enables visual identification of communities with localized interests. The design is explained in detail in Section 4.1.

From the overview, several persons can be selected to further examine the items of interest of each person and their intersections with respect to the clusters of items. Here a visual design with two layers is employed. In the background layer, a contour map is used to show the clusters of the items. Visual links connecting items in the same sets are drawn on top of the contour map. Based on this design, it is possible to visually correlate the set relations depicted on the foreground to the item clusters as revealed by the contour map.

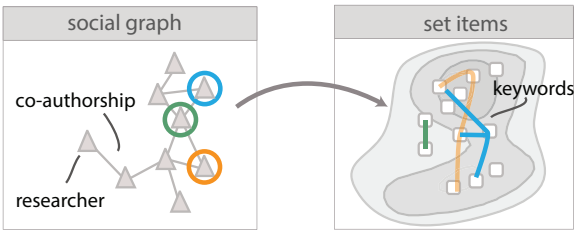


Figure 1: *An overview of the framework for analyzing set relations. Left: glyphs integrated with a node-link diagram. Right: set visualization over item clusters. The example of publication dataset is used for illustration. Three nodes are selected on the left, and the items in the sets are depicted by visual links on the left.*

4. Visual Design

In this section, we describe our visual designs in detail and the reasons that we choose these designs.

4.1. Glyph Design

For each node i, we draw a glyph encoding the information of its item set overlap and social distances to all the other nodes. We use an asymmetric measure to compute the overlap of S_i to S_j by $|S_i \cap S_j|/|S_i|$, while other measures such as the Jaccard coefficient can also be used. The social distances from node i to the other nodes are the shortest path distances on the graph in our implementation, while other distance metrics can also be applied.

There are several approaches to design the glyph that can integrate the overlap and distance information. They are different in their visual clutterness and information loss. Scatterplot (Fig. 2(a)), for example, incurs no loss in information but will become harder to read when the scatterplots for different nodes are drawn on the same screen. On the other hand, the correlation between the overlap and the distance can be computed and encoded with node size or color intensity. This introduces less visual clutter, though details

could be lost. Fig. 2(b) illustrates an alternative design. It is a gray-scale stacked histogram. The depth of the shade encodes the amount of overlap. The deeper the shade the more the overlap. In the histogram, each bar is composed of several segments. Each segment corresponds to the group of nodes which are at the same distance from node i and have a similar amount of set overlap with node i. The segments are vertically arranged by their shade. The height of each segment is proportional to the percentage of members in this group to all members with the same distance from i. The height can be scaled by S_i, the number of items in the set, as illustrated in the case study figures in Section 6. The stacked histogram can be further replaced with a stacked graph such that the shapes would have simpler geometry, as illustrated in Fig. 2(c).

Color hue as another visual channel is utilized to indicate whether a person's item set is larger than most of its neighbors in the graph. Diverging hues are assigned to each of the nodes, to encode the information as a binary value: red color represents that the size of the corresponding item set is larger than half of the neighbor nodes, and blue color represents the opposite. With this color encoding, the nodes with fewer items and are potentially subsets of the others can be easily identified. These nodes have blue hues and are more darkly shaded.

Overall, the resulting glyph features a composition of visual channels including intensity of color, hue, size and shape. After integrating the glyph to the node-link diagram, the visualization could reveal interesting patterns concerning groups of nodes and outliers. Some of our findings after applying the glyph view to real social network datasets include: 1) densely connected communities with highly localized distribution of interests; 2) subgroups in the social network where the existence of social links are highly correlated to their interest overlaps, demonstrating the phenomenon of homophily; 3) persons having interest overlap with distant nodes. These will be discussed in detail in Section 6.

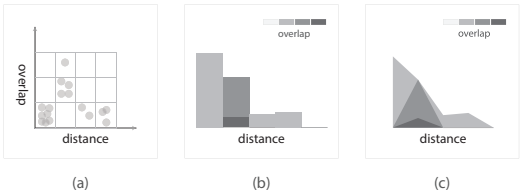


Figure 2: *Design choices for glyphs encoding set relations and social distances for each node to all the other nodes in the social graph. (a)scatterplot; (b) grayscale stacked histogram; (c) stacked graph.*

4.2. Set Visualization

Contour map, which is a density based visualization, can be used to summarize the overall distributions of the items, and recently it has been employed in many visualization systems [CSL*10] [ZBDS12]. Contour map gives a global context on top of which the sets will be overlaid. The advantage of using the density based visualization is that it can maintain scalability when there is a large number of items and in general it would be easier to detect the correlation between set relations and the global distributions of items given that the overall distribution is summarized in a concise way.

How to visually group the items in the same set is another design choice to make. There are several user tasks to be supported, including basic set relation reading tasks: (T1) find the items in a set; (T2) identify the sets that an item belongs to; and (T3) identify the set intersections. Moreover, combining the set relations with the item cluster information could enable the user tasks inculding: (T4) identify the distribution of items in a set with respect to the item clusters; and (T5) find implicit overlap between the sets.

Several options to visually group the items and to support the above tasks have been considered. Color (hue) and shape are some visual channels that can be utilized to visually group items, although they could be less effective than direct visual linking [SWS*11]. For visually linking items, one choice is to use enclosing contours, such as BubbleSets [CPC09]. This method would interfere with the underlying density based visualization though. We choose to use spanning tree like shapes to visually connect the items in the same sets, which is a generalization of the denotation used in Lineset [ARRC11] and similar to the ones in the Kelp diagram [DvKSW12]. Color-coded concentric circles as in Lineset and Kelp diagram are also used to indicate the set membership for each item (T2).

Using visual links to connect related items faces the challenge of scalability with respect to the numbers and distributions of items in the sets. Branchings and zig-zags of the visual links could make it harder to read the items in the sets (i.e., task T1 and T4) and could incur visual clutter for the detection of meaningful relations between the sets (i.e., task T3 and T5). For the visualization to be effective, it is therefore desirable to have smooth visual links with a small number of branches such that it takes less effort to visually follow the paths connecting the items [War00]. Moving the items slightly away from their original positions is a strategy that can be applied to simplify the geometry of the visual links. Meanwhile, for users to effectively perform task T4 and T5, the positions of the items in the original MDS layout should be preserved as much as possible. Therefore a trade-off should be made between two general aesthetic criteria in the visual link layout algorithm: (A1) the simplicity of the visual links; (A2) the preservation of original item locations. The details of the layout algorithm will be explained in Section 5.

5. Implementation

In this section, we describe the implementation of the view depicting set relations over item clusters.

5.1. Contour Map Construction

The input to this step is the entire set of items and the similarity between pairs. Multidimensional scaling (MDS) is employed to assign each item a position in the plane. MDS arranges items with a higher degree of similarity between each other in close proximity, and groups of similar items form visual clusters. Based on this initial layout, kernel density estimation (KDE) can be used to derive a smooth representation of the item distribution on the plane. Isocontours are traced by following points of the same density. The contours are then filled with transparent color and the additive blending effect creates regions that are more shaded indicating clusters of items. The resulting contour map is a visual abstraction of the discrete distribution of the items in the MDS layout.

5.2. Visual Link Layout

The contour map derived from the previous step serves as a reference map, based on which the visual links representing the sets are drawn. As discussed in Section 4, we intend to utilize the flexibility to move the items slightly away from their initial positions in the MDS layout, such that less visual clutter is introduced when drawing the links. The items should not move too drastically such that one could still infer the cluster that the item belongs to from the underlying contour map.

The following layout algorithm is implemented to draw visual links corresponding to several selected sets $S_i, i \in \{1,...,m\}$. The algorithm consists of two phases:

1. Formation and simplification of a spanning tree connecting all the items in $\bigcup_{i\in\{1,...,m\}} S_i$, illustrated in Fig. 3;
2. Generation of visual links for individual sets based on the "backbone" spanning tree obtained in the first step, illustrated in Fig 4.

The motivation of the first phase is to derive a reference backbone to draw the visual links for the individual sets. The following are desirable properties for the backbone spanning tree: small number of branches, smoothness in each segment, and small distortion to the item positions in the initial layout. The rationale is that the backbone should be simple in its geometry such that the visual links generated in the second phase are not cluttered (i.e., criterion A1). Meanwhile, the items should not be moved too drastically in case the contour map becomes invalid for showing the cluster that the items belong to (i.e., criterion A2). Our implementation applies the following steps to form and simplify the spanning tree:

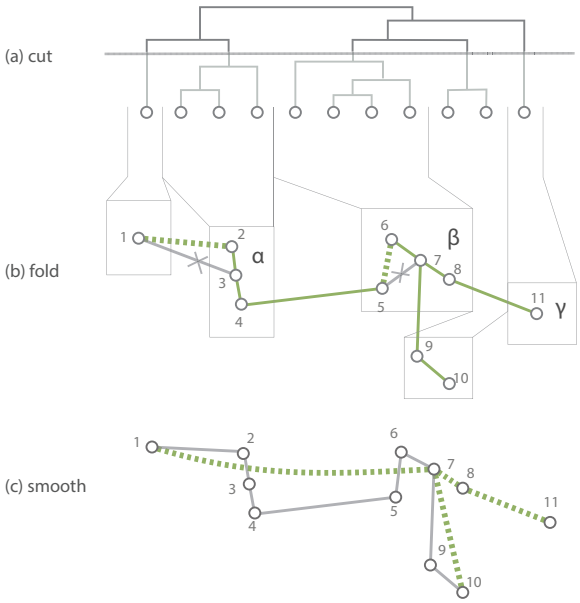


Figure 3: *Formation and simplification of a backbone spanning tree: (a) hierarchical agglomerative clustering of the items based on their euclidean distances in the initial MDS layout; (b) the nodes are grouped based on a cut on the hierarchy, the spanning tree are simplified (by reducing branches) based on the grouping, where the dashed lines are new edges; (c) the resulting branches on the spanning tree are straightened by moving the items, where the dashed lines indicate an approximate shape of the backbone spanning tree.*

1. Perform hierarchical agglomerative clustering (HAC) for all the items in the selected sets based on their Euclidean distance after MDS layout. A dendrogram showing the result of the clustering is visualized in Fig. 3(a).
2. Construct a spanning tree for the items based on the hierarchical cluster. An edge will be included in the spanning tree if it is the shortest link connecting two sibling clusters. A resulting spanning tree is shown in Fig. 3(b).
3. Perform a cut on the dendrogram at a given height. Consequently, the items will be partitioned into groups by their spatial closeness (see Fig. 3(a) and (b)).
4. Fold branches on the spanning tree (see Fig. 3(b)) based on the grouping of the items. Some edges are not on the path for connecting different groups of items and could be removed from the spanning tree. As in the figure, $e(5,7)$ is on the path for connecting group α and γ while $e(6,7)$ is not on any path connecting different groups. Therefore $e(6,7)$ could be deleted from the spanning tree. After deleting $e(6,7)$, $n(6)$ is left unconnected. The strategy we have taken is to merge $n(6)$ to the nearest edge, which is $e(5,7)$. This is accomplished by deleting $e(5,7)$ and adding $e(5,6)$, $e(6,7)$ to the tree. The result of this step is

a spanning tree which has less branches but still approximates the original shape of it.

5. Smooth the segments on the spanning tree (Fig. 3(c)). In this step, the items are moved from their original positions such that there would be less zigzags in the shape of each segment. A force-directed method could be applied in this case, with items both drawn towards their original positions and by the spring force exerted by nearby edges.

Based on the backbone spanning tree, we draw the visual links connecting individual sets in the second phase. Our method first divides the tree into segments, and routes links for each set on the segments. For each set, links from different segments are connected to form continuous visual links for each set. On each segment, there are items from different sets. To depict the item membership effectively (i.e., task T1) and to better reveal set intersections (i.e., task T3), we apply a strategy for drawing the visual links on each segment. As illustrated in Fig. 4, each set is assigned a lane parallel to the segment. To draw the visual link for a set, we scan through all the items one-by-one in the segment, and route the link to pass through: 1) the central lane if the item on the segment is common to multiple sets; 2) the lane assigned to the set if the item does not belong to the set; 3) the lane assigned to the set if the item belongs exclusively to the set. In the 3rd case, the item is also moved to the lane corresponding to the set.

Fig. 5 illustrates the intermediate results of our algorithm, and the final output, for the visualization of three sets.

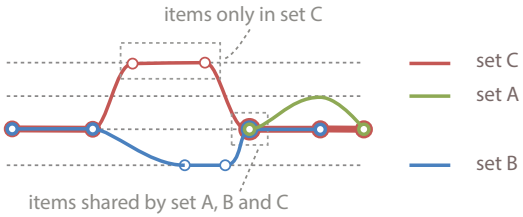


Figure 4: *The simplified and smoothed spanning tree are divided into segments (where there are no branches). In each segment, continuous lines are used to connect the items in the same sets.*

6. Results and Discussions

We select two real datasets from different application domains for case study. One is an online social network (Last.fm) where each person is interested in a set of music artists, and another is a professional social network (bibliographical databases) where each person is related to a set of research topic related keywords. The visualizations in the case study are implemented in Java with the Prefuse† information visualization toolkit.

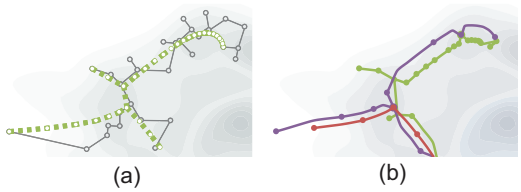


Figure 5: *The intermediate and the final results of the layout algorithm: (a) the initial spanning tree connecting all the items and the backbone spanning tree after the first phase of the algorithm is performed; (b) the visual links for individual sets drawn based on the backbone spanning tree.*

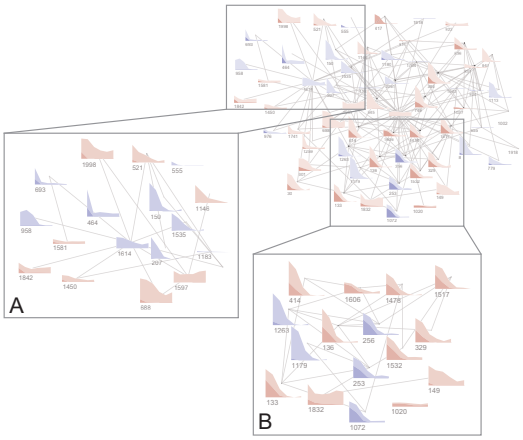


Figure 6: *Subgraphs from the last.fm social network with glyphs. There are two observations: 1) in general, a larger distance in the social network implies less overlap in interests; and 2) two groups of persons marked by the rectangles exhibit different properties on the amount of interest overlaps. The effect of homophily is more noticeable in group B.*

6.1. Last.fm

Last.fm [las12] is a social music service website which maintains a catalog of artists, albums, and tracks. Users of the website can listen to music, setup personal profiles of the artists they like, and add other users as friends. Last.fm also provides a web-service API, which can be used for querying information of the users and the artists. The available information includes the tags, the number of listeners per artist, and the similarity between two given artists.

In the case study, we used the last.fm data released in

† http://prefuse.org/

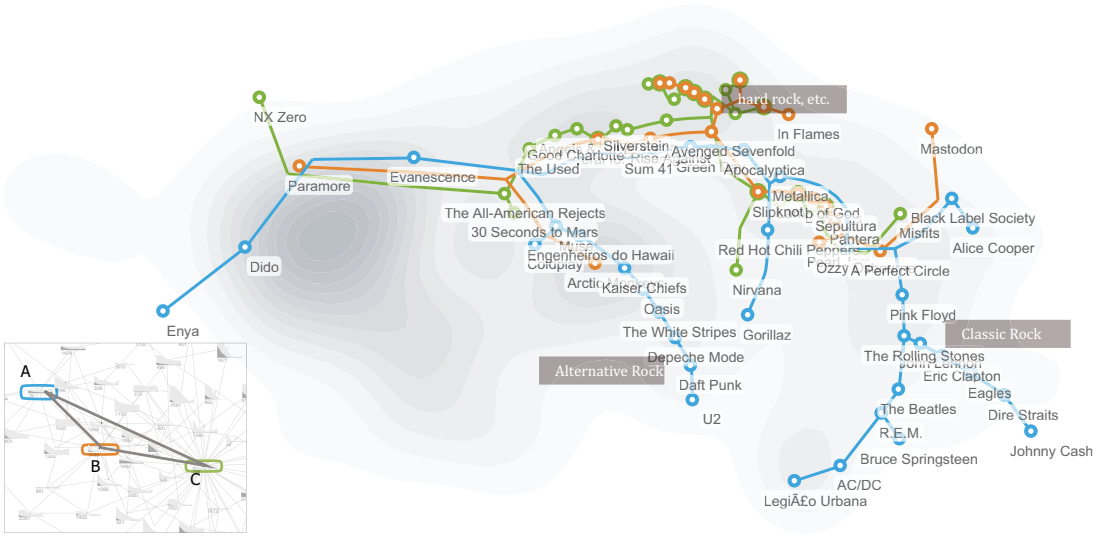


Figure 7: *Interest overlap of close neighbors in the last.fm social network. The three persons are mutually connected to each other. However, one person has drastic difference in his interest from the other two. The others have a lot of interest overlaps in several clusters of similar music artists.*

[CBK11] which contains the friendship graph of the users (the names of the users have been anonymized) and user listening history information. We selected the most popular 500 artists from the dataset according to the total listening counts. The similarity between artists is obtained through the web service API (other measures are also available for music artists). Artists form clusters based on the similarity information.

In last.fm, each person is related to a set of artists that he/she likes. We applied the glyph design to analyze the overlap of interests between persons in the context of the social network. Fig. 6 shows a subgraph extracted from the friendship graph, where each node is drawn as a glyph summarizing the information about their interest overlap and social distance to all the other nodes. It can be observed that in general, there is negative correlation between the degree of interest overlap and social distance, as most of the glyphs have a skewed shape with peaks leaning towards smaller social distances. The use of different shades enhances the visual cue for the detection of negative correlations. Groups with different levels of locality in interest distribution can also be observed.

Fig. 7 demonstrates the overlapping interests of three persons in close proximity in the social graph. As revealed in the underlying contour map, there are some major clusters of similar artists. Person A does not have much overlap with B and C even when they are in close proximity. In particular, it seems that only A is interested in clusters of music artists tagged as "classic rock" and "alternative rock" on the last.fm website. Persons B and C have a lot of overlaps over clusters of artists tagged as "hard rock" and "pop punk". Moreover, for persons B and C, even in some clusters, there are not many explicit overlaps between them, the parallel visual links for each set connect a lot of items from B and C and visually indicate implicit overlapping over the clusters.

6.2. Academic Publication Data

For academic publication data, we use the InfoVis 2004 contest dataset [FGP04]. We applied our method to the largest connected component in the co-authorship network of the information visualization community which consists of 69 nodes and 156 edges. A topic-modeling technique based on latent Dirichlet allocation (LDA) is used to extract topic related keywords from paper abstracts. We use the implementation in the MALLET toolkit [McC02] for topic modeling. The similarity between two keywords is derived from two sources: 1) their co-appearance in the same topic; and 2) co-citations of the corresponding publications. Take words that often appear in information visualization literature for example, "graph" and "layout" would be more related than "graph" and "architecture" or "graph" and "pipeline".

In Fig. 8, we show the results of applying the glyph view to the co-authorship graph in the publication data. The results also demonstrate the negative correlation of social distance to interest overlap. There are also some tightly knitted communities where the locality of interest distribution is relatively high, that is, the amount of overlap is extremely high within short distances (typically 1 or 2). We identify examples of such communities in the figure.

The divergent color encoding with red and blue on the

Figure 8: *Interest overlap on the academic collaboration network. Two observations are: 1) there are strongly connected communities with a higher level of locality of keyword distribution and 2) there are some outlier nodes having many keyword overlaps even with distant nodes.*

nodes assists the detection of potential subset/superset relations. The keywords of the nodes in the tight-knit community in Fig. 8, are likely to be covered by many of their nearby nodes, as they are blue-colored (thus their interest is smaller than many of the neighborhoods) and more shaded at smaller distances (thus the degree of overlap is high in the neighborhood). The pattern is more evident in the academic collaboration networks in comparison with the last.fm social network (Fig. 6).

Fig. 9(a) shows the set of keywords for three authors on the contour map. The three authors are Peter Pirolli, George W. Furnas and Marti Hearst. From the figure, we can see that there are keywords in same clusters that belongs uniquely to Peter Pirolli, including “retrieve”, “web” in one topic cluster, and “spreadsheet”,“multivariate”, “mining” in the other. Marti Hearst has overlap with Peter Pirolli on the keywords “dynamic”, “tree”, “support”, “interact” while George W. Furnas has overlap with Peter Pirolli on the keywords “graphic”, “interface”, “design”. The three persons are at different locations in the social network, with Marti Hearst having direct collaboration links to Peter Pirolli and George W. Furnas being farther away from both. Our method applies MDS to assign each item a position in the plane. As there are different MDS algorithms and different parameters might lead to different layouts, to test the stability of our method with regard to different MDS layouts, we adjust the parameters used in our MDS method and generate another result (see Fig. 9(b)). From the figure, we can clearly see that the visualizations from our method are stable.

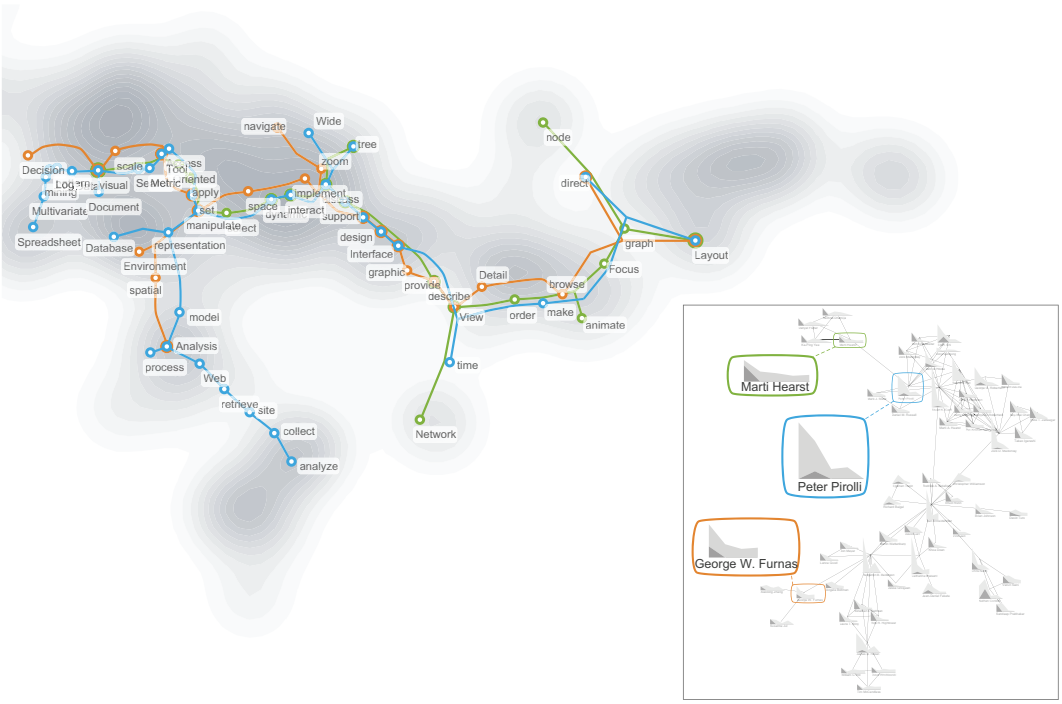
6.3. Discussions

Our methods have some drawbacks. Our method can well reveal the relationship of a small number of sets in a graph. When there are more sets, the chance of the lines crossing each other even when there is no set intersection is likely to increase and the relations among these sets might be too complicated for users to understand. It is possible to develop a more intelligent layout algorithm which makes better assignment of the lanes to the sets to minimize the crossings and further reduce visual clutter. A more plausible way is to adopt the visualization mantra, "overview first, zoom and filter, and then details on demand", in our system. An aggregation view of all the relationships is displayed first. Users can then zoom into a subset and identify a few interesting nodes, and our method can be applied to reveal the set relationships of the selected nodes. After that, users can zoom out again and select another subset of the nodes and repeat the whole process. With this iterative exploration approach, the relations among multiple nodes can be revealed. Another drawback of our method is that adding glyphs to node-link diagrams could cause visual clutter, especially when the social network is very dense. We believe rich user interactions and intelligent layout algorithms can alleviate this problem. For the visual link layout algorithm, one limitation of the current method is that since the items belonging to different sets are juxtaposed on the segments, the resulting visual links would contain a lot of zig-zags. This problem can be partially resolved by changing the order of the items on the segments.

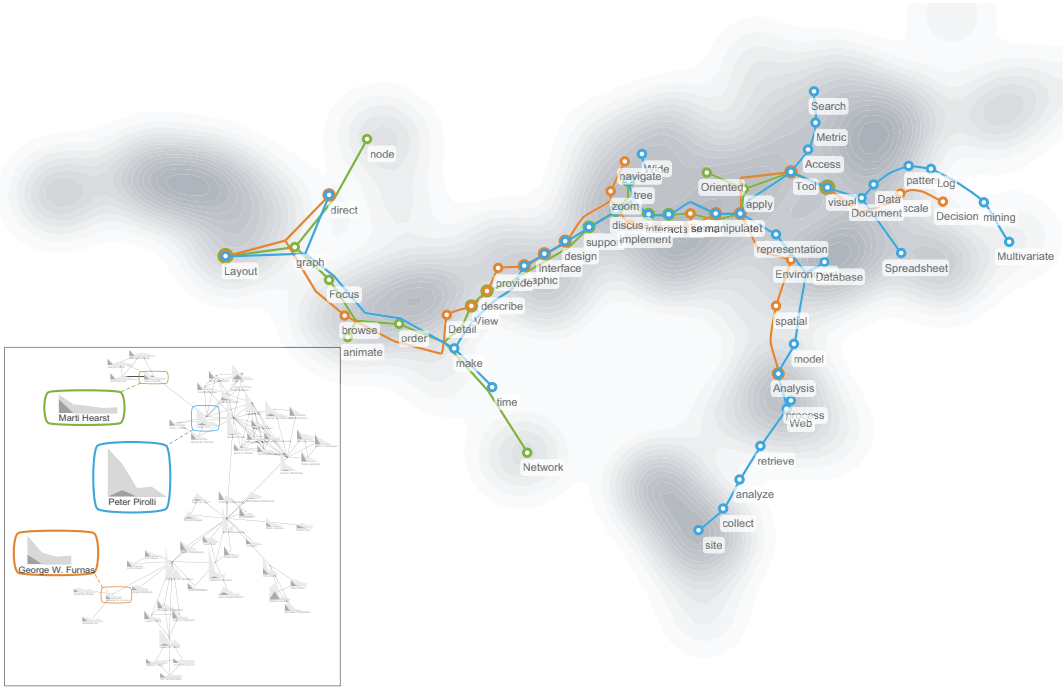
7. Conclusions and Future Work

In this paper, we proposed two visual designs for the visual analysis of set relations in a graph. The visual designs can be applied to analyze the set relation in the context of the social graph and the set relations in the context of item clusters. We applied our designs to two real datasets and discussed the findings. In the future, we intend to apply our designs to other kinds of datasets and also conduct more studies to evaluate their effectiveness. For the layout algorithm, various improvements could be made. For example, the branches of the spanning tree could be folded to further reduce visual clutter. When there are more sets, an algorithm different from our current approach for assigning and drawing the visual links on different lanes could be considered.

Acknowledgements We would like to thank Shauna Dalton for proofreading and the anonymous reviewers for their valuable and constructive comments. Hong Zhou is supported by Foundation for Distinguished Young Talents in Higher Education of Guangdong, China (No. LYM11113) and NSF of China Project (No. 61103055). Huamin Qu is supported by a grant from MSRA. Fan Du contributed to this paper during his visit to the Hong Kong University of Science and Technology.



(a)



(b)

Figure 9: *Interest overlap of three authors in an academic collaboration network. Three authors (i.e., Peter Pirolli, George W. Furnas and Marti Hearst) are selected from the social network. They are at different social distances from each other. (a) and (b) show the results with two different MDS layouts respectively.*

References

[ARRC11] ALPER B., RICHE N. H., RAMOS G., CZERWINSKI M.: Design study of LineSets, a novel set visualization technique. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2259–67. 2, 4

[BCD*10] BEZERIANOS A., CHEVALIER F., DRAGICEVIC P., ELMQVIST N., FEKETE J.-D.: Graphdice: A system for exploring multivariate social networks. *Computer Graphic Forum 29*, 3 (2010), 863–872. 2

[BCPS11] BRANDES U., CORNELSEN S., PAMPEL B., SALLABERRY A.: Blocks of hypergraphs. In *Combinatorial Algorithms*, Iliopoulos C., Smyth W., (Eds.), vol. 6460 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 201–211. 2

[BCPS12] BRANDES U., CORNELSEN S., PAMPEL B., SALLABERRY A.: Path-based supports for hypergraphs. *Journal of Discrete Algorithms 14* (2012), 248–261. 2

[BE01] BERTAULT F., EADES P.: Drawing hypergraphs in the subset standard (short demo paper). In *Graph Drawing*, Marks J., (Ed.), vol. 1984 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 164–169. 2

[CBK11] CANTADOR I., BRUSILOVSKY P., KUFLIK T.: 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proceedings of the 5th ACM conference on Recommender systems* (New York, NY, USA, 2011), RecSys 2011, ACM, pp. 387–388. 7

[CPC09] COLLINS C., PENN G., CARPENDALE M. S. T.: Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1009–1016. 2, 4

[CSL*10] CAO N., SUN J., LIN Y.-R., GOTZ D., LIU S., QU H.: Facetatlas: Multifaceted visualization for rich text corpora. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (nov.-dec. 2010), 1172 –1181. 2, 4

[DRRD12] DÖRK M., RICHE N. H., RAMOS G., DUMAIS S. T.: Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2709–2718. 2

[DvKSW12] DINKLA K., VAN KREVELD M. J., SPECKMANN B., WESTENBERG M. A.: Kelp diagrams: Point set membership visualization. *Computer Graphics Forum 31*, 3pt1 (2012), 875–884. 2, 4

[FGP04] FEKETE J.-D., GRINSTEIN G., PLAISANT C.: Ieee infovis 2004 contest, the history of infovis. In *www.cs.umd.edu/hcil/iv04contest* (2004). 7

[JP87] JOHNSON D. S., POLLAK H. O.: Hypergraph planarity and the complexity of drawing venn diagrams. *Journal of Graph Theory 11*, 3 (1987), 309–325. 2

[KKS09] KAUFMANN M., KREVELD M., SPECKMANN B.: Graph drawing. Springer-Verlag, Berlin, Heidelberg, 2009, ch. Subdivision Drawings of Hypergraphs, pp. 396–407. 2

[KMSH12] KAIRAM S., MACLEAN D., SAVVA M., HEER J.: Graphprism: compact visualization of network structure. In *AVI* (2012), pp. 498–505. 2

[las12] Last.fm, Dec. 2012. 6

[Mäk90] MÄKINEN E.: How to draw a hypergraph. *International Journal of Computer Mathematics 34*, 3-4 (1990), 177–185. 2

[McC02] MCCALLUM A. K.: Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002. 7

[MSLC01] MCPHERSON M., SMITH-LOVIN L., COOK J. M.: Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 1 (2001), 415–444. 1, 3

[PS08] PERER A., SHNEIDERMAN B.: Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2008), ACM, pp. 265–274. 2

[RD10] RICHE N., DWYER T.: Untangling euler diagrams. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 1090 –1099. 2

[SA06] SHNEIDERMAN B., ARIS A.: Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (sept.-oct. 2006), 733 –740. 2

[SAA09] SIMONETTO P., AUBER D., ARCHAMBAULT D.: Fully automatic visualisation of overlapping sets. *Computer Graphics Forum 28*, 3 (2009), 967–974. 2

[SRHZ11] STAPLETON G., RODGERS P., HOWSE J., ZHANG L.: Inductively generating euler diagrams. *IEEE Transactions on Visualization and Computer Graphics 17*, 1 (2011), 88–100. 2

[SWS*11] STEINBERGER M., WALDNER M., STREIT M., LEX A., SCHMALSTIEG D.: Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2249–2258. 4

[vLKS*11] VON LANDESBERGER T., KUIJPER A., SCHRECK T., KOHLHAMMER J., VAN WIJK J., FEKETE J.-D., FELLNER D.: Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum 30*, 6 (2011), 1719–1749. 2

[War00] WARE C.: *Information Visualization: Perception for Design (Interactive Technologies)*, 1st ed. Morgan Kaufmann, Feb. 2000. 4

[Wat06] WATTENBERG M.: Visual exploration of multivariate graphs. In *CHI* (2006), pp. 811–819. 2

[WFC*06] WONG P. C., FOOTE H., CHIN G., MACKEY P., PERRINE K.: Graph signatures for visual analytics. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1399–413. 2

[ZBDS12] ZINSMAIER M., BRANDES U., DEUSSEN O., STROBELT H.: Interactive level-of-detail rendering of large graphs. *IEEE Transaction on Visualization and Computer Graphics 18*, 12 (2012), 2486–2495. 4