

Academic Network Explorer: Making Sense of Your Research through Interaction

Zipeng Liu
zipeng@cs.ubc.ca

November 27, 2015

1 Task Abstraction

I have re-thought about this section as it needs major improvement in the proposal. Here is an outline of domain independent task abstraction using wordings from the textbook. Entity and object are the same concept and interchangeable.

- Locate (target known, location unknown) a seed entity by text searching
- Select one or multiple entities of the same kind (either paper, author, ...) as central entities
- Explore ambient entities around a (or a group of) central entity:
 - Select the type of ambient objects
 - Present the distribution of attributes of potential ambient objects
 - Aggregate (bin) some attributes like #citations of paper, #papers of author
 - Juxtapose multiple views of distributions
 - Filter items by one or multiple attributes
 - Select relationships between filters
 - Modify previous filters
- Collapse (Reduce) ambient entities
- Group (Reduce) similar entities on demand
- (OPTIONAL) Annotate on objects and actions so that it can be presented to others later; save system snapshots so that a presenter can reconstruct the images he creates before

2 Project Progress

Generally speaking, I am keeping up with the plan that I made in the proposal.

2.1 Data and Server

For the data processing, I have converted the raw data file into a MongoDB database, created necessary indices and tuned it to the best performance that I could get within limited time. For the backend server, I wrote most of the APIs with Express.js, simply tested them, but according to my previous experience, there will be modifications and new demands on the API, which would not cost much time to do given that I have set up the framework and written many helper functions (building blocks of the backend).

2.2 Network Layout

For the interface and interaction design, I believe I thought through how to layout the node-link network, which accounted as the most important design work I left from the proposal.

I do not use any of the common layout algorithm in visualization because mine is an “incremental” network. The nodes in the network are conceptualized as central or ambient nodes although some might play both roles. The basic design rationale is to draw ambient around the central node. Other design considerations include symmetry, density, alignment. Inspired by the HOLA paper [1], I first drew some networks by hand and intuition, and derived the empirical rules for positioning nodes around a central node.

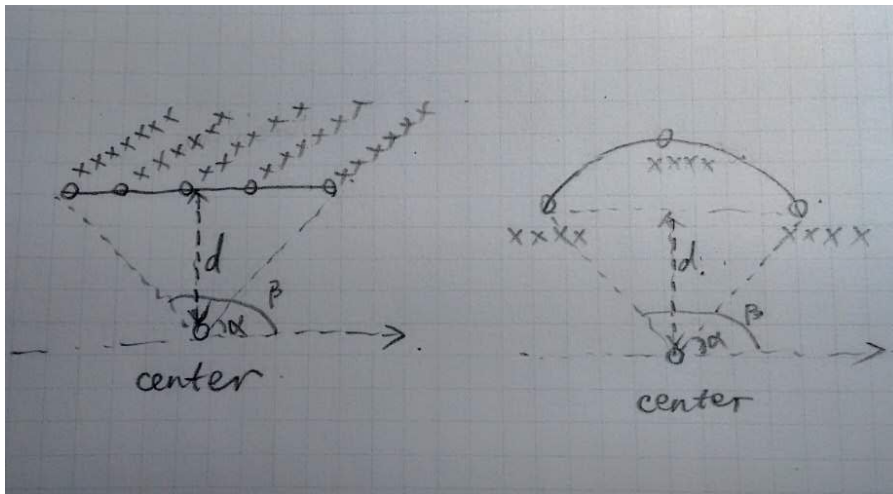


Figure 1: Specification of auxiliary line or arc for positioning ambient nodes around a center. The orthogonal straight line or arc is described by a triple (α, β, d) . Label direction is also shown.

Suppose several objects denoted as a set X are expanded around a central object A . X would be positioned along whether an arc or a orthogonal straight line according to: number of objects in X , need for long text label (they look better when aligned), need for showing order. In a polar coordinates where the central objects is the pole, position of the line or arc could be described by a triple: start angle (α) , end angle (β) , distance from pole to the line or arc (d) , as illustrated in Figure 1. As we want to make it compact to reserve space for

future expansion, we aim to minimize the distance as ambient objects would be positioned close to the center but still need to keep some space between objects and labels to avoid overlapping. Finally, the nodes are evenly spread out along the line and arc with text label written next to it, whose direction is either 0, 45 degree, 135 degree.

2.3 Problems and Challenges

I realized that there might be **too much features** in ANEX to implement within two weeks. Instead of implementing each pair wise relationship expansion, I will pick some important ones such as papers to authors, papers to citing papers, etc. Relations like venue and affiliation might be put off. The **optional task** in the task abstraction requires a fair amount of effort, but I will try to get it done because it would help greatly with disseminating knowledge generating with ANEX.

Besides, I am suspicious about the quality of this dataset. It was scraped from several digital libraries which might use different representations. **Inconsistency and deficiency** was detected during preliminary data processing and validation. The format of venue and affiliation name is extremely diverse. Different entities (authors) turned out to be the same and should be merged. It misses important entities like there is no term "visualization". However, I would stick to this one since I can still use it to illustrate my design, which should also work on other scholarly dataset.

3 Related Work

This project is related to a few previous studies and systems spanning from different areas in visualization.

3.1 Literature Analysis

Starting from decades ago, before the "birth" of visualization, researchers have studied how to display paper citation data to augment writing reports on history of science [2]. Later when digital collection of papers became available, Chaomei Chen, interested in research trends, detected "research fronts" using burst detection and betweenness analysis, and then visualized them in a clustered node-link graph [3]. John Stasko's group from Georgia Tech visualized papers published in IEEE Information Visualization Conference from 1995 to 2012 [4]. Their CiteVis system could show citations and references of papers in details, and rankings by citations as well. They got some interesting findings and patterns in visualization publications. They also tried different methods such as CiteMatrix, CiteList. SurVis [5] worked on carefully surveyed literature collection in order to disseminate literature. Users such as survey authors could structure their references with the powerful selector interaction. There are also literature browsers like Treevis.net [6], Timeviz.net [7] to disseminate visualization literature. Unlike any of these previous studies, ANEX takes a different angle and focuses on visualizing the network indicated by citation, co-authorship and other relationships.

3.2 Faceted Data

Scholarly dataset is also regarded as typical faceted data, which people have come up with creative ways to explore. Preliminary research on faceted data focused on browsing and searching [8, 9]. Marian Dork et. al proposed faceted information space [10], which used pivot interaction to enable strolling in the space. Their succinct design and slick transition inspired ANEX a lot. Jian Zhao et. al built PivotSlice [11] to easily browse multiple facets and find correlations between them. They treated facets as sets and supported expressive set manipulations through rich interaction techniques. PivotSlice was better in understanding overview of facets and relations in between, but lack of details in individual objects, which ANEX provides. Also, Keshif [12] took facets as sets to support filtering and comparison in a highly interactive, versatile document browser.

3.3 Network and Set

As we think our dataset as network data, there is work related from network visualization. Detangler [13] addressed cohesion problems of multiplex network by constructing a substrate and catalyst layer and enabling “leapfrog” between them. Kieffer et. al studied how human produced elegant orthogonal networks by hand and derived guidelines to automate such process [1]. The layout algorithm we come up with was greatly inspired by them.

As we pointed out in the data abstraction (pretending that this is the final report :)), part of the data could be also visualized as sets [14]. We also leverage techniques from sets such as union and intersection operation to select our expanding interests (central objects), multiple linked views to visualize distributions of attributes [12], and rankings of attributes [15]. These techniques help users understand what are the ambient objects and filter exactly what they want.

References

- [1] S. Kieffer, T. Dwyer, K. Marriott, and M. Wybrow. Hola: Human-like orthogonal network layout. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):349–358, Jan 2016.
- [2] Eugene Garfield, Irving H Sher, and Richard J Torpie. The use of citation data in writing the history of science. Technical report, Institute for Scientific Information, 1964.
- [3] Chaomei Chen. Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *journal of the american society for information science and technology*, 57(3):359–377, 2006.
- [4] John Stasko, Jaegul Choo, Yi Han, Mengdie Hu, Hannah Pileggi, R Sadanaand, and Charles D Stolper. Citevis: Exploring conference paper citation data visually. *Posters of IEEE InfoVis*, 2013.

- [5] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with survivis. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):180–189, Jan 2016.
- [6] H.-J. Schulz. Treevis.net: A tree visualization reference. *Computer Graphics and Applications, IEEE*, 31(6):11–15, Nov 2011.
- [7] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.
- [8] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408. ACM, 2003.
- [9] m.c. schraefel, Max Wilson, Alistair Russell, and Daniel A. Smith. mSpace: Improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49, 2006.
- [10] Marian Dork, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. Pivotpaths: Strolling through faceted information spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2709–2718, 2012.
- [11] Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan. Interactive exploration of implicit and explicit relations in faceted datasets. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2080–2089, 2013.
- [12] M.A. Yalcin, N. Elmqvist, and B.B. Bederson. Aggreset: Rich and scalable set exploration using visualizations of element aggregations. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):688–697, Jan 2016.
- [13] Benjamin Renoust, Guy Melançon, and Tamara Munzner. Detangler: Visual analytics for multiplex networks. *Comput. Graph. Forum*, 34(3):321–330, 2015.
- [14] Bilal Alsallakh, Luana Micalef, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Peter Rodgers. Visualizing sets and set-typed data : State-of-the-art and future challenges. In *Eurographics Conference on Visualization (EuroVis)*.
- [15] Alexander Lex, Nils Gehlenborg, Hendrik Strobel, Romain Vuillemot, and Hanspeter Pfister. UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '14)*, 20(12):1983–1992, 2014.