

Academic Network Explorer: Making Sense of Your Research through Interaction

Zipeng Liu
zipeng@cs.ubc.ca

November 10, 2015

1 Domain

Literature reading is an important skill for researchers across all fields, but it is never easy to find related work or make sense of them, especially for inexperienced beginners. While modern digital libraries might ease the burden of acquiring publications, it is often not feasible to read them all. Typically, there are multiple groups of researchers working on a topic all over the world and we need to understand the community if we want to be part of it. Sometimes, we want to know who are doing what and how well they do before going to a conference. These are all non trivial tasks for researchers, which requires a lot of probing and reading.

Digital libraries collect rich scholarly data, especially meta information of publications, which is now used majorly for searching and browsing. However, with interactive visualization, we could enable exploration of the academic world, and gain more insights in a more engaging and understandable way than reading words by words of documents, which is the general goal of our proposing system – Academic Network Explorer (ANEX).

2 Data Description

Our dataset includes paper information, paper citation, author information and author collaboration. There are about two million papers, eight million citations, one million authors and four million collaborations. For papers, there are attributes such as title, authors, affiliations, publication venue, year, abstraction and references; for author, there is name, affiliations, count of published papers, extracted key terms and some scholarly indices like H-index. This dataset is provided by ArnetMiner [1], which scraped and collected publications and authors in computer science for years. It can be downloaded from <https://aminer.org/billboard/AMinerNetwork>.

It is a typical network data induced by multiple relations like citation, co-authorship. It can also be regarded as set data since there are natural belongingness like authors in affiliations and papers in venues, and also artificial ones like authors who are described by some key terms. Besides, as many previous work pointed out, this is a faceted data because objects have many different aspects: papers have a few associated attributes.

3 Task Abstraction

In this project, we are going to focus on making sense of a local region, which users are interested in, of this giant academic network from multiple angles. Also, we facilitate finding related work of certain topics and provide overview of them. Notice that we are not interested in the main content of the published papers, which requires text related techniques.

ANEX should be able to answer the following questions. **The list is not exhaustive, but indicates typical basic functionalities. By integrating several tasks, users should be able to gain a mental image of their areas of interests and have a sense of general direction of what to read next.**

- Making sense of a paper
 - What is this paper about? What are the related topics?
 - What are the references of this paper? Who cited it after its publication?
 - How important is this paper in terms of related topics
 - What are the similar papers?
- Making sense of a researcher
 - What does he work on?
 - What is he famous for? What's his contribution in the field?
 - Who are his co-authors? How do they do in the field?
 - Who are the most related researchers?
 - How does his work evolve through time?
- Making sense of a venue
 - What is this venue about?
 - What is its academic impact?
 - Who published papers on it?
 - What are the most influential work?
- Making sense of a few related topics
 - What are some related or similar key terms?
 - What are the some influential (must-read) papers?
 - Who are working on it? Who are the most influential authors?
 - How does hot research questions evolve?
 - What are the first-tier publication venues?

In general, all of these questions could be generalized as exploration of **ambient objects around a centered object** whose relations might be given by citation, authorship, collaboration, belongingness and similarity of keywords, as well as **attributes of these objects** like numerical indices for rankings and temporal features.

4 Design Rationale

The nature of this problem is exploring objects locally, and hence we are not interested in providing users with overview of the whole dataset, but instead, enabling expansion from a seed through rich interaction. We will follow the general guideline “from detail to overview via selections and aggregations” coined as DOSA by Stef van den Elzen et. al [2], except that the overview in our case is more of a local context.

The task abstraction highlighted the networking feature (objects and relationships). ANEX’s exploration process is adding up pieces by pieces starting from a seed through flexible and interactive expansion, which supports specifying different objects and selecting interesting subsets given attribute distributions. As a side effect of the expansion, users can grasp the features of attributes.

5 Scenario of Use

A first year PhD student is looking for a problem to start with. His supervisor provided him a few topics in visualization that he could consider working on and asked him to read related literature on each of the topic before decision, but the professor did not give him exactly what to read. He is new to this area of research, so first of all he wants to find one or two good overview paper, and then get a sense of what challenges and problems are in these topics his supervisor provided. For the most interesting topics, he would like to find out a few famous papers in different flavors by different research groups.

With ANEX, he creates a seed topic “visualization”. He expands some highly related or co-occurring keywords with the seed topic. He also expands some highly cited authors, and their famous papers. By skimming the paper titles, he might found some popular summary work of visualization such as “The Value of Visualization” by Jarke van Wijk (see Fig. 1). He might also google the term and read the wikipedia page for vis as well as some book chapters. For each of his interested topics, he might interact with ANEX in the same way to get a sense of the research fields. After some serious reading, he might look for related or similar papers from competing research groups in that field with ANEX’s expansion interaction (see Fig. 2).

6 Personal Expertise

I have done some work on visualizing a paper’s citation during my undergraduate study, but in that project I focused on text (pdf) processing, data cleaning and trying out a scalable software architecture to put it in real usage. In this project, we are using a dataset of paper meta information and focusing on relationships between facets.

I am familiar with D3 and some other tools that we are planing to use except that I have not decided yet on a MVC framework: whether to use AngularJS or Backbone that I have experience in, or try a new one like React.

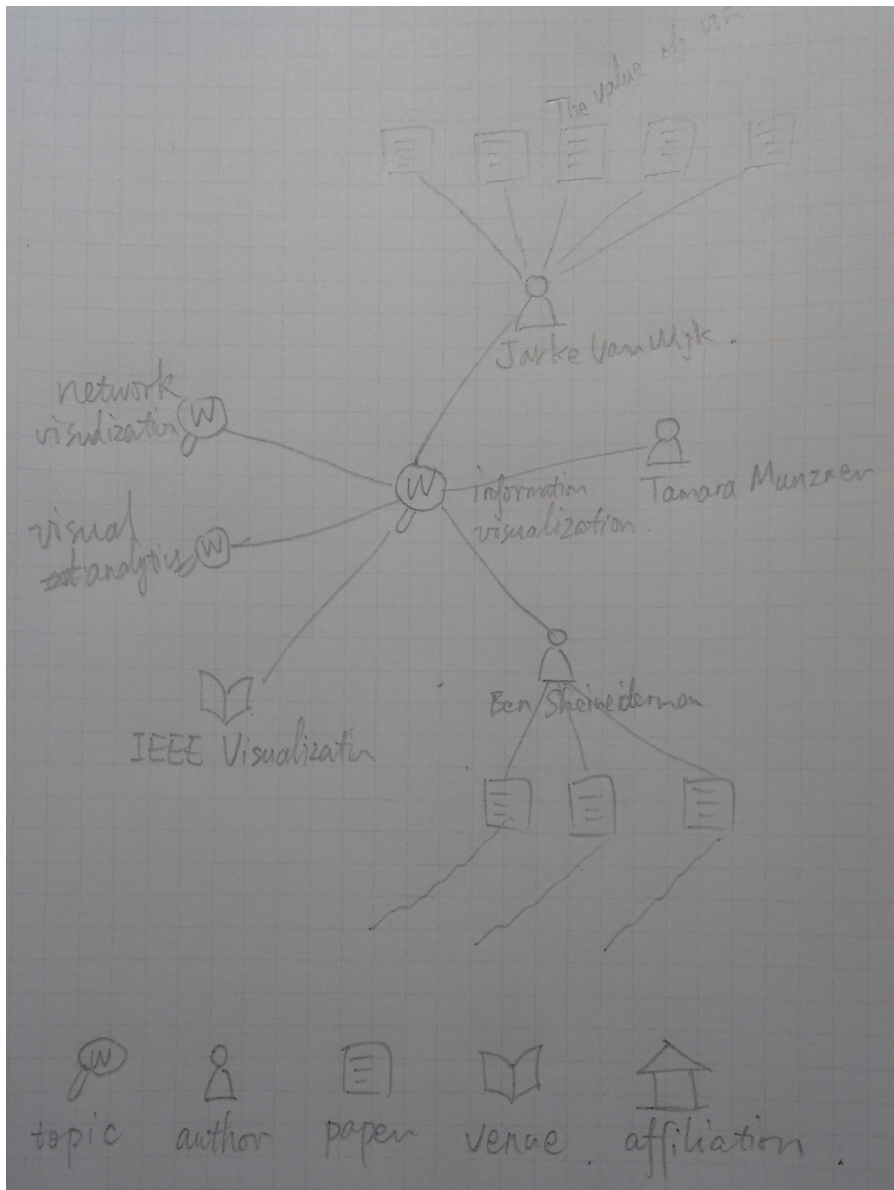


Figure 1: Design sketch showing the main view of ANEX. The network is showed in a node-link diagram, with different glyphs representing different types of objects. However, we have not decided on a layout algorithm because no simple obvious solution fits. An incremental layout is under consideration.

7 Planning Implementation

ANEX will be implemented in both backend and frontend. On the backend, there is a MongoDB database for data storage and we will provide a RESTful application interface for the frontend. For potential future deployment, we are planning to build it in Docker containers. On the frontend, we have not decided

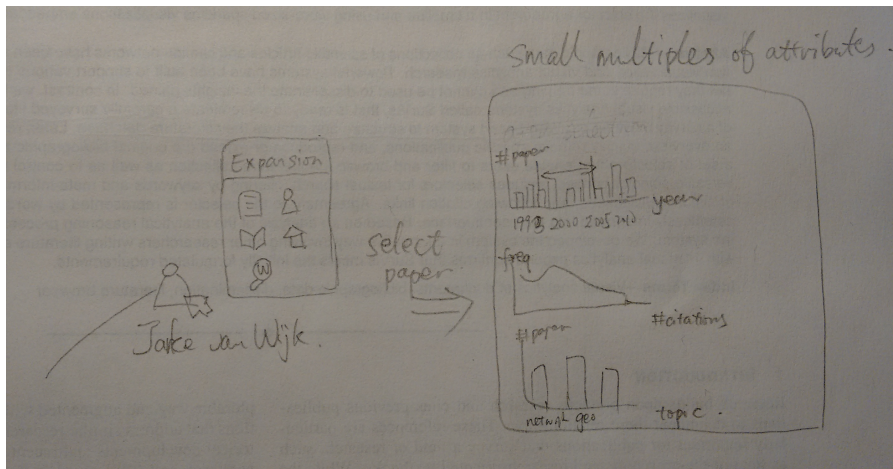


Figure 2: Illustration of ANEX’s expansion interaction. First, the user chooses an object type to expand, and then he could see distributions of attributes of possible expanding objects. He could select a subset of them from the small multiples or simply expand all. After specification of what to expand, the objects would be added to the graph or merged with existing nodes.

which MVC framework to use. It could be AngularJS or React. We are going to use D3 for manipulating SVG elements and layout.

8 Milestones and Schedule

Major Milestones are listed below:

- Nov 23 (2 weeks): complete design; setup database; implement API; write up related work section of final report.
- Dec 7 (2 weeks): finish all functions on frontend; write a few sections of final report such as introduction, task and data abstraction.
- Dec 13 (1 week): write most of report; polish interface (small tweaks); prepare presentation.
- Dec 17 (4 days): polish report.

9 Related Work

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

9.1 Literature Analysis

Starting from decades ago, before the “birth” of visualization, researchers have studied how to display paper citation data to write report on history of science [3]. Interested in research trends, Chaomei Chen detected “research fronts”

using burst detection and betweenness analysis, and visualized them in a clustered node-link graph [4]. John Stasko’s group from Georgia Tech visualized papers published in IEEE Information Visualization Conference from 1995 to 2012 [5]. Their CiteVis system could show how a paper was citing and cited in details, and rankings by citations as well. They got some interesting findings and patterns in paper citations of visualization. They also tried different methods such as CiteMatrix, CiteList. SurVis [6] 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, Timeviz.net to disseminate visualization literature.

9.2 Faceted Data

Scholarly dataset is also regarded as typical faceted data, and people have come up with creative ways to explore it. Preliminary research on faceted data focused on browsing and searching [7, 8]. Marian Dork et. al proposed faceted information space [9], which used pivot interaction to enable strolling in the space. Their succinct design and slick transition is also worth to mention. Jian Zhao et. al built PivotSlice [10] to easily browse multiple facets and find correlations between them. They treated facets as sets and supported expressive set manipulations through rich interaction techniques. Also, Keshif [11] took facets as sets to support filtering and comparison in a highly interactive , versatile document browser.

9.3 Network and Set

As we think our dataset as network data, there is work related from network visualization. Detangler [12] addressed cohesion problems of multiplex network by constructing a substrate and catalyst layer and enabling “leapfrog” between them.

As we pointed out in the task abstraction, part of the data could be also visualized as sets [13]. We are also considering borrowing some techniques from sets such as union and intersection operation to select our expanding interests, previewing in small multiples to visualize attributes [11], and ranking of attributes [14].

References

- [1] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Ar-netminer: Extraction and mining of academic social networks. In *KDD’08*, pages 990–998, 2008.
- [2] S. van den Elzen and J.J. van Wijk. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2310–2319, Dec 2014.
- [3] Eugene Garfield, Irving H Sher, and Richard J Torpie. The use of citation data in writing the history of science. Technical report, 1964.

- [4] 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.
- [5] 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.
- [6] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with survi. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):180–189, Jan 2016.
- [7] 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.
- [8] Max Wilson, Alistair Russell, Daniel A Smith, et al. mspace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49, 2006.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Benjamin Renoust, Guy Melançon, and Tamara Munzner. Detangler: Visual analytics for multiplex networks. *Comput. Graph. Forum*, 34(3):321–330, 2015.
- [13] 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)*.
- [14] 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.