# Visual Characterization of Personal Bibliographic Data Using a Botanical Tree Design

Tsai Ling Fung* Kwan-Liu Ma†

University of California, Davis

**ABSTRACT**

This paper presents a preliminary study on egocentric visualization of a bibliographic database. The visualization design is based on a botanic tree metaphor, resulting in a visually interesting and information rich depiction of one's research career in terms of publication records. The case studies reveal both the strengths and limitations of the current design.

**Index Terms:** K.8.1 [Personal Computing]: Application Packages—Graphics; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Evaluation/methodology

## 1 INTRODUCTION

The area of personal visualization emerges as all information about individuals and their activities and behaviors goes online. Personal visualization may be done for either exploring data or facilitating personal improvement. Therefore, both general public and domain experts such as sociologists and security analysts can make use of personal visualization. In this paper, we present a study of individual researchers' professional development in terms of their publications, co-authorships, and associated scientific disputes using a personal visualization design based on a botanical tree metaphor [12].

Online bibliographic databases storing rich information are important resources to scientific researchers. These databases are also of strong interest of sociologists for studying the development of a particular discipline or patterns of academic career. For example, these patterns could be of citations for peer-reviewed journal articles among the scientists working in the same field. In one study [13], an analysis of citation networks shows that scientific debates exhibit distinct spiral patterns.

Our study uses a dataset extracted from Cornell's public scientific library at arXiv.org. This dataset is unique because it contains not only the bibliographic data for published research works, but also scientific disputes among over 400,000 researchers working in physics and related fields. The disputes are in the form of published articles for either commenting a published work or responding to particular comments. Our objective is to visually characterize each individual researcher's career path and compare one's to another's. We like to observe different career paths in terms of one's publication record and co-authorships. We also like to understand the social process due to disputes as well as the implications of the resulting challenges to one's career. Our case studies show that the botanic tree like visualization can effectively bring out particular aspects of one's professional development and also facilitate comparisons.

## 2 RELATED WORK

There has been growing research on visualization of personal content. With data surrounding our lives like medical history [11], activity logs [1, 19], social network [9], etc., we also see growing demands of tools and devices for visualization and management of the data to facilitate self-development or to gain insights. Recent research done by Huang *et al.* suggests that using visual means on personal data can encourage individuals to examine their personal life and help them promote self-awareness and self-expression, and possibly improve their lives [7]. The types of insight that may be gained are characterized in [3] based on an analysis of quantified self-presentations. As daily data such as email information [17], life events [16], and related cyber activities [6] being increasingly collected into digital formats, it's helpful to map such personal digital archives into visual representations to trigger reflection or narrative and recall of a person. By the help of media, personal visualization can also be a real-time, projected self-expression [9]. To support better personal storytelling, Thudt *et al.* mention the challenges of integrating multiple data sources, personalization, self-reliance and privacy in visualizing personal data as autobiographical visualization, which might be a potential aspect of personal visualization [15].

Our work uses a more organic visualization design, largely based on *Contact Trees* [12], introduced by Sallaberry *et al.* for studying contact diaries [5], data collected for the research of individuals' social contacts and activities. Contact Trees uses a botanical tree like visualization design to effectively present multidimensional data with different tree features. Other botanical tree visualizations [2, 8] also encode high-dimension data into aesthetic tree appearances. LinkedIn Career Tree Builder is a tool using a node-link diagram laid out into a tree like visualization for highlighting career experience [22]. There are other organic visualization designs. PeopleGarden by Xiong *et al.* [21] employs a flower metaphor to visualize online social interactions. UbiFit Garden by Consolvo *et al.* [4] shows that using pleasing visualization can help people continuously improve their ideal lifestyle. More recently, Wang *et al.* also use a flower metaphor for visualizing Facebook data [18]. Their study found that more abstract or organic visualizations could be a good trigger for insight gaining, although it could take more time to interpret the visualization.

We use bibliography data in our study. Bibliography data is commonly visualized with node-link diagrams to show the relationship between authors [14, 10]. Since we are interested in studying individual researchers in an egocentric way, a botanic tree design is adequate. Wu *et al.* introduce PathWay [20], also an egocentric visualization design, to present and study patterns and trends in an individual researcher's bibliographic data over a selected period of time. In particular, they study not only patterns in scholarly contributions but also scientific disputes associated with one's academic career. We use a dataset similar to theirs. However, Pathway is a much more abstract visualization design than our botanic visualization which shows greater details of one's publication record and enables different studies.

## 3 AN EGOCENTRIC VISUALIZATION DESIGN

Our egocentric visualization design is extended from Contact Trees [12] with a particular emphasis on usability. The visual representation is a 2D tree, and its usable tree features include trunk size,

*e-mail:tlfung@ucdavis.edu

†e-mail:ma@cs.ucdavis.edu

Electronic proceedings of the IEEE VIS 2015 workshop
*Personal Visualization: Exploring Data in Everyday Life*

trunk sides, tree branches and branch sides, sticks, leaf and its size, color, and shape, fruits, and tree roots. The case studies presented here do not use all the tree features. A web-based, interactive visual interface is provided to the users to make easier mapping of a selected data attribute to a selected tree feature.

Considering the bibliographic database we intend to study, some possible tree feature mappings to use are as follows. The tree trunk presents the tree's major shape giving a view of one's overall career. Trunk size can indicate the total number of articles authored by this researcher. Trunk side can be mapped to some natural binary division of the data. For example, in most of the trees presented here, we place an article with a DOI (Digital Online Identification) number on the right, and without on the left. The other possible division is to place all conference papers on one side, and all journal papers on the other side. Tree branch is probably best to encode time information. Branches from bottom to top of the tree can represent different periods of time over one's career. Branch side can also be used to encode information. For example, we can place regular articles on the upper side of the branch, and articles that give comments to another article or responses to comments on the lower side. Each leaf may display a particular kind of collaborative relationship that the ego has. Leaf color or size can indicate quantitative information such as paper length, number of co-authors, or number of citations received. Finally, fruit can be used to highlight a particular aspect of an article. Egocentric visualizations of the bibliographic data using some of the above mappings are shown in Figure 1.

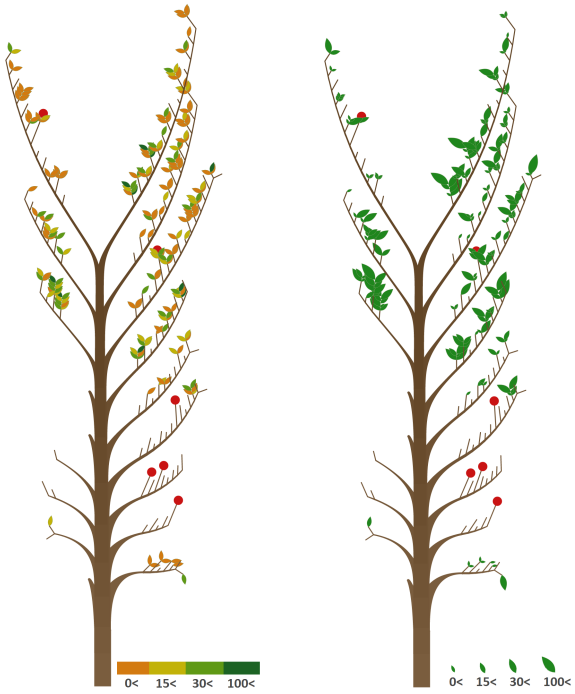


Figure 1: An egocentric visualization of one's publication record using a botanic tree metaphor. Each leaf is a co-author. The total number of publications that the co-author has is mapped to either leaf color (Left) or size (Right). An article with a fruit attached indicates somebody comments on this article.

## 4 Case Studies

We have chosen a few researchers sharing similar patterns or with contrasting patterns for our studies. The objective is to show usability and identify limitations of the current design. A more comprehensive evaluation of the botanic tree design and a sociological study of the whole database remain to be conducted.

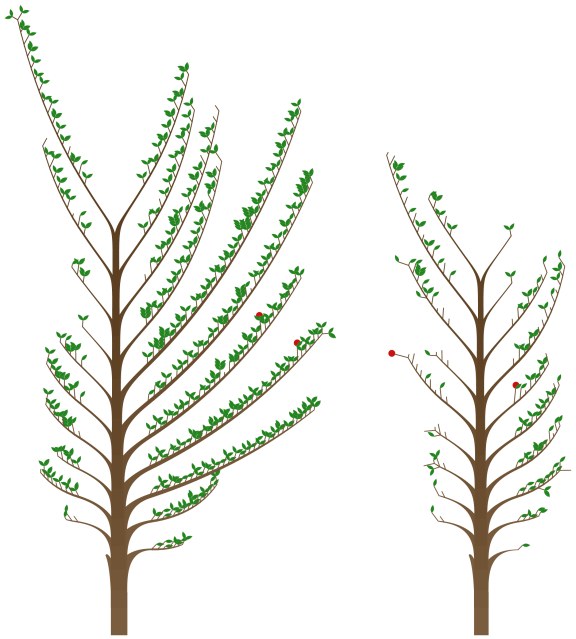

Figure 2: Two active researchers. Each branch encodes two years of publications. Going upward, the trees display their publications between 1993-2013 for the left one and 1995-2012 for the right one.

We start with two more productive researchers as depicted in Figure 2. The right side of trunk displays articles with DOI while the left side display those without. We place regular articles on the upper side of each branch, and those articles commenting others' articles or responding to comments are placed on the lower side. while both researchers look productive over their career, the one on the left has more publications than the one on the right, but only two of his articles are comments to others' works. In contrast, the researcher on the right has written ten articles to comment others' works, and he also has produced more single-author articles.

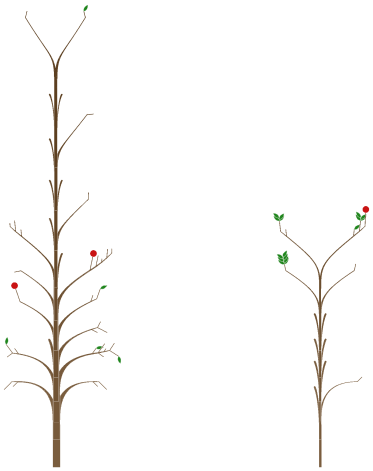

Figure 3: Two less productive researchers.

In contrast, Figure 3 displays the trees of two less productive researchers. It's interesting to see the one on the left was more active over the first half of his career, while the one on the right almost did not publish at all until much later in his career. The

one on the left has several articles commenting on others' works especially early in his career. This researcher also has very few collaborators. It's not clear if the two sets of comments he received in the middle part of his career impacted him or not.

According to [20], we know there are some researchers having very close collaborative relationships. We present two such researchers in Figure 4. As seen, their trees have similar shapes. The similarity is particular strong over the early time of their careers. We examine their similarity further by combining their trees into one as shown in Figure 5. As expected, the resulting tree become quite symmetrical and their close work relationship appears even clearer. Note that we have to change mappings. Here the upper side of each branch displays articles with DOI while the lower side shows articles without. Each leaf is still a co-author. Orange leaves are those articles commenting others' works, while green leaves are regular articles.

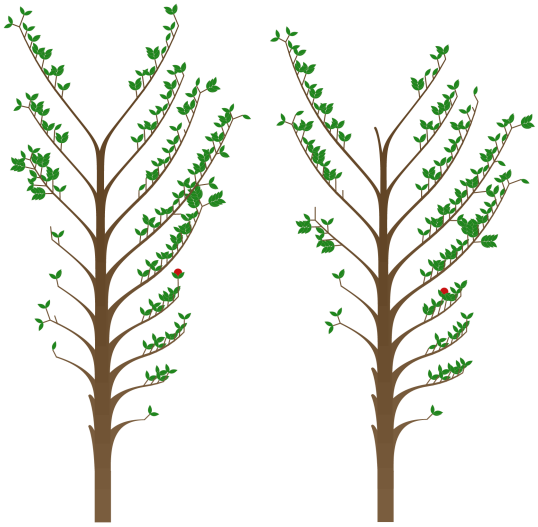

Figure 4: Two researchers with a close collaborative relationship.

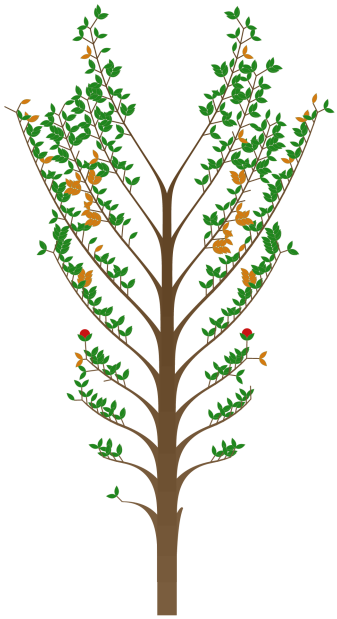

Figure 5: Visualization of the same two researchers in Figure 4 in one tree with each researcher on one side of the tree. The tree looks highly symmetrical as expected.

In the next example, we separate single-authored or first-authored articles from co-authored articles. Figure 6 shows the resulting visualization compared to a partition based on DOI. While this author clearly becomes more collaborative, he still plays significant roles in many of his publications. In particular, the number of single-authored articles do not seem to decrease as this author becomes more senior.

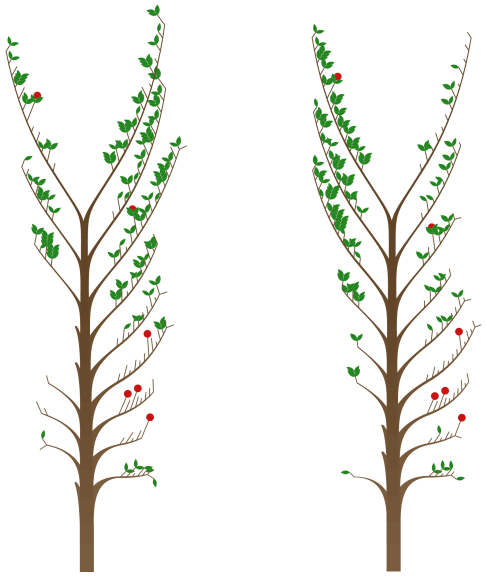

Figure 6: The same researcher in Figure 1. Left: Articles with DOI are placed on the right side of the trunk while those without DOI are on the left side. Right: Single-authored or first-authored articles are placed on the right side of the trunk while co-authored articles are on the left.

Finally, we make a visualization of a researcher's publication record in terms of article length, which is mapped to leaf size. The visualization also separates single-authored or first-authored articles from co-authored articles by trunk side. Two researchers with the identical academic period are displayed in Figure 7. The researcher on the left has largely published on his own while the one on the right has mainly published as a co-author. Both researchers have written articles of different lengths but we can see the one on the right wrote more longer articles. By the dominant color of the tree. it is clear that he has more co-authors in his publications. The other researcher has no publication with more than 10 co-authors since no dark green leaf is found in his tree. It's interesting to see both researchers become more collaborative and tend to write longer papers later in their careers.

## 5 Conclusion

The botanic tree design for personal visualization is effective in presenting multidimensional information. Our preliminary study on a bibliography database using this design suggests several directions for further studies:

- While our tree-like visualization looks appealing it is not without limitations. It is not scalable. We shall introduce more abstract levels of the tree for comparison and selection tasks. Furthermore, current tree features limits how many categories an attributes can have. For example, there might not be an attribute in the dataset that can be naturally mapped to the two sides of a tree.
- The usability of the visualization very much depends on how easy it is for the user to specify each tree feature mapping. Our system provides an interactive interface for users to specify


Figure 7: The same researcher in Figure 6 (left) compared to another researcher for the same academic period 1997-2013 (right). Leaf size indicates the length of a paper. Leaf color tells us if the article is co-authored or not and approximately how many co-authors. Single-authored or first-authored articles are placed on the right side of the trunk while co-authored articles are on the left. Articles with DOI are placed on the upper side of each branch while those without are on the lower side.

mapping of tree features. However, this interface remains to be evaluated and improved.

- It seems plausible to make image-level tree classification to find like egos.
- It is desirable to provide alternative organic visualization designs to the users. One design might match one particular type of data better.
- We shall engage sociologists for studying sociological aspects of one's academic career development.

We also like to enhance the bibliographic database by collecting other data such as citation numbers for conducting deeper analyses.

## Acknowledgements

This research has been sponsored in part by the National Science Foundation via grants NSF DRL-1323214 and NSF IIS-1320229, and UC Davis RISE program.

## References

[1] J. S. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz. Shuteye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1401–1410, 2012.

[2] E. B. Chlan and P. Rheingans. A botanically inspired high-dimensional visualization with multivariate glyphs. In *Proceedings of the Sixth Joint Eurographics - IEEE TCVG Conference on Visualization*, VISSYM'04, pages 231–236, 2004.

[3] E. K. Choe, B. Lee, and M. Schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *Computer Graphics and Applications, IEEE*, 35(4):28–37, July 2015.

[4] S. Consolvo, D. W. McDonald, and J. A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 405–414, 2009.

[5] Y.-C. Fu. Contact diaries: Building archives of actual and comprehensive personal networks. *Field Methods*, 19(2):194217, 2007.

[6] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, and N. Shkrob. Youpivot: Improving recall with contextual search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1521–1530, 2011.

[7] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.

[8] E. Kleiberg, H. van de Wetering, and J. J. Van Wijk. Botanical visualization of huge hierarchies. In *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, pages 87–, 2001.

[9] M. Leung, M. Tomitsch, and A. Vande Moere. Designing a personal visualization projection of online social identity. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1843–1848, 2011.

[10] S. Noel, C.-H. H. Chu, and V. Raghavan. Co-citation count vs correlation for influence network visualization. *Information Visualization*, 2(3):160–170, 2003.

[11] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 221–227, New York, NY, USA, 1996. ACM.

[12] A. Sallaberry, Y. chih Fu, H.-C. Ho, and K.-L. Ma. Contact Trees: A Technique for Studying Personal Network Data. *CoRR*, abs/1411.0052, 2014.

[13] U. Schwed and P. Bearman. The temporal structure of scientific consensus formation. *American Sociological Review*, 75:817840, 2010.

[14] Z. Shen, M. Ogawa, S. T. Teoh, and K.-L. Ma. Biblioviz: A system for visualizing bibliography information. In *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation - Volume 60*, APVis '06, pages 93–102, 2006.

[15] A. Thudt, D. Baur, and S. Carpendale. Autobiographical visualizations: Challenges in personal storytelling. In *In Proceedings of the DIS'14 Workshop on A Personal Perspective on Visualization and Visual Analytic*, 2014.

[16] A. Thudt, D. Baur, and S. Huron. Visual mementos: Reflecting memories with personal data. *Visualization and Computer Graphics, IEEE Transactions on*, PP(99):1–1, 2015.

[17] F. Viegas, D. Boyd, D. Nguyen, J. Potter, and J. Donath. Digital artifacts for remembering and storytelling: posthistory and social network fragments. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10 pp.–, Jan 2004.

[18] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *Computer Graphics and Applications, IEEE*, 35(4):82–93, July 2015.

[19] J. Wood. Visualizing personal progress in participatory sports cycling events. *Computer Graphics and Applications, IEEE*, 35(4):73–81, July 2015.

[20] M. Q. Y. Wu, R. Faris, and K.-L. Ma. Visual exploration of academic career paths. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 779–786, 2013.

[21] R. Xiong and J. Donath. Peoplegarden: Creating data portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 37–44, 1999.

[22] N. YAU. (2010) Build your LinkedIn career tree. [online]. Available at http://flowingdata.com/2010/11/29/build-your-linkedin-career-tree/.