CommunityExplorer: A Framework for Visualizing Collaboration Networks

Leonel Merino Dominik Seliner Mohammad Ghafari Oscar Nierstrasz

Software Composition Group, University of Bern, Switzerland
http://scg.unibe.ch

Abstract

Understanding the network of collaborations, identifying the key players, potential future collaborators, and trends in the field are very important to carry out a project successfully. In this paper, we present *CommunityExplorer*, a visualization framework that facilitates presenting, exploring, and understanding the network of collaborations at once. The framework performs data extraction, parsing, and modeling automatically. It is easy to adopt and utilizes a bigraph visualization that scales well. We demonstrate the advantage of CommunityExplorer to identify the collaboration of authors on 346 and 104 research papers published in SOTFVIS/VISSOFT and IWST communities respectively. We found that even though SOFTVIS/VISSOFT has more contributors, IWST exhibits more collaboration. We discovered that contributors in IWST are more resilient than those in SOFTVIS/VISSOFT, which are more volatile. Moreover, collaboration in IWST is concentrated in a single large group, while in SOFTVIS/VISSOFT it is spread among many tiny groups and a few medium-sized ones.

Categories and Subject Descriptors Human-centered computing [*Visualization*]: Visualization systems and tools

Keywords Collaboration network visualization, node-link diagram, Smalltalk, Pharo

1. Introduction

People within diverse communities across the world collaborate in various kinds of projects, such as producing a software system in industry or writing a paper in academia. Understanding the network of collaborations, identifying key players and potential future collaborators, and uncovering trends in the field are very important to carry out a project successfully.

The relationships between collaborators and projects, *e.g.*, authors and papers, can be depicted as a graph, or a *collaboration network*, where collaborators are nodes and projects are edges that relate them. Normally, people collaborate in projects in the same domain (*e.g.*, a database researcher would collaborate in database papers). These projects and collaborators form a *community*. Within a community there can be disjoint subgraphs formed by *groups* of collaborators who have worked together directly or indirectly. Understanding which collaboration groups have been most significant in the past, and which are currently active is an important step towards understanding the relevant literature in a domain, current research trends, and potential collaborators for one's own work.

Identifying key groups, assessing their relevance, and analyzing their evolution can be challenging tasks that require appropriate support. A suitable visualization of collaboration networks can augment human understanding by mapping properties of the data to graphical dimensions that can address these challenges. Previous research has proposed the analysis of collaboration networks based on statistical analysis (Biryukov and Dong 2010), metric analysis (Wu et al. 2009), and structure analysis (Chen et al. 2011). However, the few that have proposed graph-based visualizations (Brandes et al. 2009; Huang et al. 2008; Tymchuk et al. 2014) suffered from edge cluttering.

Researchers who want to adopt visualization for the analysis of collaboration among communities have to devote considerable effort collecting tools and preparing their data. We think that a framework that facilitates these steps can boost the use of visualization. In this paper we propose such a framework. It includes a built-in visualization that avoids edge cluttering by using a bigraph. In it, both collaborators and projects are explicitly represented as nodes connected by edges. In our experience, bigraphs yield tidy representations that can encode quantitative attributes in both the size and the intensity of the color of nodes. Although understanding collaborations can be important in

IWST '16 August 23, 2016, Prague, Czech Republic
Copyright © 2016 held by owner/author(s). Publication rights licensed to ACM.
ACM 978-1-nnnn-nnnn-n/yy/mm. . . $15.00
DOI: http://dx.doi.org/10.1145/nnnnnnn.nnnnnnn

many settings, we focus on the specific case of understanding collaborations between authors.

We demonstrate the utility of our framework through a case study to visualize the networks of collaborations among groups of 639 and 162 researchers who have published their work in SOFTVIS/VISSSOFT and IWST respectively. We conducted this study based on the complete set of publications in VISSOFT (IEEE Working Conferences and International Workshops on Software Visualization), SOFTVIS (ACM Symposium on Software Visualization) and IWST (International Workshop on Smalltalk Technologies).

We believe that the framework can help (i) collaborators looking for potential collaboration (*e.g.*, authors looking for potential co-authors), (ii) stakeholders seeking to identify key groups (*e.g.*, researchers looking for key literature), and (iii) communities seeking insight into their evolution (*e.g.*, conferences deciding to restrict or expand their topics).

Accordingly, we focus on the following research questions:

RQ1. How can visualization support users to identify collaboration groups, evaluate how relevant they are, and study how they evolve?

RQ2. How does collaboration in the SOFTVIS/VISSOFT community differ from others?

RQ3. How suitable are bigraphs (as opposed to graphs) for modeling collaboration networks?

In Section 2 we compare our approach to related work. Section 3 presents our framework and describes the visualization. Section 4 elaborates a case study in which we visualize the publications in the VISSOFT/SOFTVIS venues, and compares it to a visualization of publications in IWST. Section 5 concludes the paper and outlines future work.

2. Related Work

Relevant research tackling the problem of visualizing collaboration by using various types of graph representation have been proposed in the past. Dörk *et al.* (Dörk et al. 2012) used a tripartite graph to represent authors, titles and keywords of paper collections. Nodes representing paper titles are located in the middle section of the visualization, each of which connect through Bezier edges to their author nodes in the upper part, and to their keywords in the lower part of the visualization. Since the titles of papers occupy a large part of the space available, users can only analyze a few papers at a time. Edges that connect the few author nodes and keyword nodes normally overlap. To counteract this effect, when users select a title, the outgoing edges are highlighted. As a result, the visualization helps users to explore the relationships between authors, titles and keywords by traversing the paths that connect them.

Stasko *et al.* (Stasko et al. 2008) also used an n-partite graph that represents multiple types of data using nodes of different shape and color. In their design, they chose to expose only part of the graph and allow users to expand and collapse portions as they traverse the data. In contrast, our visualization aims to help users to view collaboration in a community as a whole. We apply a force-based layout on a bigraph representation to diminish edge overlapping, to facilitate group identification, and to encode properties to the size and the intensity of the color of nodes producing a readable representation.

Normally graph-based visualizations suffer from edge cluttering when scaled up to large data sets. Gansner *et al.* (Gansner et al. 2011) used edge bundling to alleviate cluttering and reveal high-level edge patterns in large graphs. Dunne and Shneiderman (Dunne and Shneiderman 2013) improved graph readability by a motif simplification in which common patterns of nodes and links are replaced by compact and meaningful glyphs. Both represent complementary approaches that can be included in our tool in the future.

Important research has taken place to identify collaboration groups and reveal communities. Vehlow *et al.* (Vehlow et al. 2015) established four types of visualization techniques that help users to identify group structures in graphs. These types however, focus only on visualizations that explicitly encode collaboration groups using graphical attributes, instead of visualizations such as ours that aim at discovering implicit groups through the use of multiple graph types and layouts. Hu and Shi (Hu and Shi 2015) reviewed layout algorithms and interactions for exploration proposed for large graphs. They observed there is a need of techniques to analyze evolution in constantly increasing graphs. In our approach we tackled both issues: we deal with large graphs by using bigraphs that reduce the number of edges and their cluttering, and we allow users to analyze evolution by encoding a temporal property in the color intensity of nodes.

Several tools and visualization techniques have been proposed to analyze collaboration. Huang and Huang (Huang and Huang 2006) proposed a technique that uses concentric rings to represent contributors and their relationships that focuses on the relationships of one author at a time. Brandes *et al.* (Brandes et al. 2009) visualized collaboration in Wikipedia. In their technique the size and the intensity of the colors of nodes are used to encode quantitative properties, however overlapping edges interfere with users' ability to effectively analyze relationships. Osborne *et al.* (Osborne et al. 2013) introduced the *Rexplore* tool that allows users to gather data from multiple online sources. Data are processed and visualized as a graph in which the colors of edges and sizes of nodes encode various types of relationships and the impact of contributors respectively. Rexplore, however, focuses only on a single contributor at the time. In contrast, we aim at providing an overview of collaboration of the whole community. In our approach the network is modelled as a bigraph to untangle edges, making room for a richer visualization able to encode more properties such as temporality, and finally deriving a deeper analysis.

3. The proposed framework

We introduce *CommunityExplorer*, a framework for visualizing the network of collaborations within a community. The source code, installation instructions, and full-sized pictures of the visualizations presented are available online.[1]

The framework is implemented in Pharo[2] — a Smalltalk inspired language and environment — chosen mainly because of its liveness and expressiveness. On the one hand, the live Pharo Playground (Chiş et al. 2015) allows us to continuously test our implementation: every change in the code pane can be reflected shortly after in the visualization pane. On the other hand, the expressive API of the Roassal visualization engine (Araya et al. 2013) allows us to create interactive visualizations with just a few lines of code.

The framework allows the following steps to be performed automatically:

1) *Importing* BibTex files, or a folder containing papers in PDF.

2) *Extracting* authorship fields from BibTex files automatically, and from files in PDF by defining parsing heuristics. Although parsing data that is structured (*e.g.*, BibText) is straightforward, doing so in the plain text extracted from PDF files is much more complex. In order to visualize such data, it has to be cleaned of misspelled words, and normalized to match identical entities across multiple papers (*e.g.*, author name). The framework includes two out-of-the-box heuristics for parsing the text based on the layout used in the paper. Users can also develop their own heuristics considering the specific characteristics of their data (*e.g.*, custom layout), and the framework provides feedback concerning their accuracy. Moreover, heuristics developed can be reused for visualizations of multiple communities.

3) *Cleaning* the extracted data. On the one hand, extracted data may contain misspelled words. On the other hand, some information can have different formats in multiple files. We normalize data by grouping similar strings as measured by the Jaccard index. For instance, the author of a paper named *John Doe* may appear in a different paper as *J. Doe*. We transform each extracted name to the latter form, we remove non-letter characters, and then we calculate their similarities.

4) *Modelling* data as a first-class object. Our framework populates a model with objects that contain the extracted data and that hold links to other objects in the model. Users can pose queries to the model, for instance they can ask *how many authors does a paper have on average?* as follows:
`(model papers collect:[:e | e authors size]) average.`

5) *Visualizing* collaboration networks, or customizing a visualization to their particular needs. Thanks to the expressiveness of Roassal API, the visualization can be specified as a short script (see Listing 1).

Listing 1. Roassal specification of the collaboration network visualization

```
|projects contributors|
view := RTView new.
projects := (RTBox new
                color:[:a| Color black
                      alpha:(1 − ((2015 − a year)∗0.071))];
                height:5; width:10) elementsOn: papers.
view addAll: projects.
contributors := (RTEllipse new color:[:a| a getColor])
                                  elementsOn: authors.
view addAll: contributors.
RTMetricNormalizer new elements: contributors;
                          normalizeSize: #score.
RTEdge
   buildEdgesFromObjects: papers
   from: #yourself
   toAll: #authors
   using: (RTLine new color: (Color blue alpha:0.4))
   inView: view.
RTForceBasedLayout new on: (projects, contributors).
(projects, contributors) @ RTSetEdgeAlpha.
(projects, contributors) @ RTPopup.
contributors @ (RTLabelled new textElement:[:e| e model
                                              authorName];
                             fontSize:4).
projects @ RTDraggable.
```

Typically collaboration networks are visualized as graphs (Brandes et al. 2009; Huang et al. 2008; Tymchuk et al. 2014) where projects are represented by sets of edges connecting collaborator nodes (see Figure 1, at left). However, the large number of edges that typically overlap, can make it difficult for users to obtain insights into the collaborations. We tackle this issue by modeling the collaboration network as a bigraph that contains two sets of nodes, one representing projects (*e.g.*, papers) as rectangles, and the other representing collaborators (*e.g.*, authors) as circles (see Figure 1, at right). However, visualizing a network in which the same set of contributors collaborate in multiple projects will lead to more edges when it is modeled as a bigraph rather than a graph. We hypothesize that bigraphs are more suitable for visualizing networks that expand by collaborations introduced between new contributors rather than by new collaborations introduced among existing contributors. In Section 4 we show an example of such a community.

In our visualization (shown in Figure 2) blue circles represent collaborators in the community who have collab-

[1] `http://scg.unibe.ch/research/CommunityExplorer`

[2] `http://www.pharo.org`

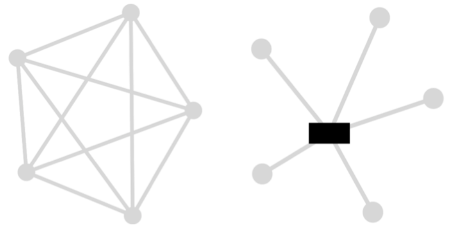

Figure 1. An example comparing visualization of a collaboration network using a graph (left) and a bigraph (right).

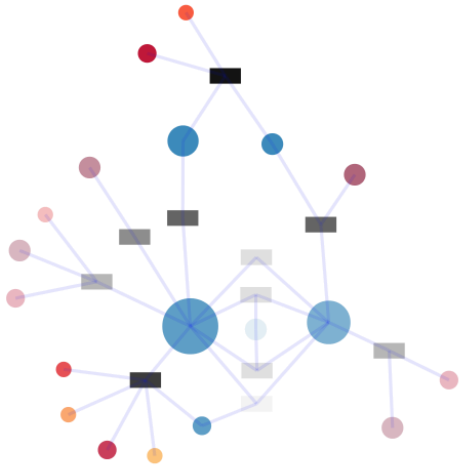

Figure 2. Visualizing a Collaboration Network.

orated in several projects, while a gradient of color from yellow to red is used for new collaborators who have collaborated in only one project. The color is used to map a qualitative attribute (*e.g.*, a role played in the project). Rectangles represent projects, and three pre-defined ratios (portrait, landscape and square) are used to encode qualitative properties of projects. The intensity of the color of rectangles can be used to encode a quantitative property of projects. Edges connect a project node to its collaborator nodes, which are distributed using a force-directed layout.

4. Case Study

In this section we demonstrate the usage of CommunityExplorer to investigate collaboration among authors in communities formed by the publications over several years of a venue. In this context, projects are papers and collaborators are authors. We elaborate on the support that the visualization provides to cope with the challenges posed in Section 1. We seek to identify groups of collaborations, evaluate their relevance, and characterize their evolution.

In our visualization the size of an author node encodes the number of papers published as a measure of the impact of the author in the community. Rectangles represent projects: those in landscape orientation correspond to conference papers, while portrait rectangles represent posters. Squares depict tool papers. The intensity of the gray in rectangles encodes the year of the publication of the paper: the older the paper the lighter the color. In this way papers and (inactive) authors start to fade away as they age. Edges connect a paper node to its author nodes, which are distributed using a force-directed layout. The visualization supports users to identify groups (disjoint subgraphs). Consequently, we revisit the research questions posed in Section 1.

4.1 RQ1. How can visualization support users to identify collaboration groups, evaluate how relevant they are, and study how they evolve?

First, we start by importing the whole collection of 346 papers published in SOFTVIS/VISSOFT in PDF format. CommunityExplorer extracts titles and authors from the files using the default parsing heuristic included. We then visualize a model containing the extracted data. The visualization (shown in Figure 3) shows the main groups. We notice the contrast between groups 1 and 2. The former represents a small group with only recent publications, while the latter exposes a medium-sized group with a similar number of papers, but mostly published in the past. We find that group 3 is somewhat homogeneous in terms of the impact of their authors in the community. The two main authors in the center can be described as being very productive. We observe that the remaining groups are composed of a main author collaborating with others with less impact. Group 5 is the largest and more productive, in which new publications are balanced between the main author (in its center) and new authors in the periphery.

Second, we assess the impact of authors in the community. We realize that the order of authors may not be useful since it may have many different meanings in different communities. Sometimes it is strictly alphabetical, sometimes in order of contribution, and sometimes the final position is reserved for supervisors. However, we hypothesize that in some groups of this community the last position is reserved for supervisors. Often supervisors guide the work of many authors, so if that hypothesis holds, in a visualization that maps author's impact as the sum of papers to a node's size we expect a reduced relative size of supervisors nodes compared to one that uses a weighted sum (that assigns more weight to authors that appear in the first positions). We visually compare collaboration networks that define an author's impact as (i) the number of published papers, or (ii) a weighted sum, calculated as $\sum_{i=1}^{N} \frac{\gamma_i}{2^{i-1}}$, where N is the highest number of authors that any paper has in the collection, and γ_i represents the number of papers on which an author appears in the ith position. Figure 4 compares the two approaches. We realize that the relative size of author nodes does not change significantly in groups A and B; while in groups C and D, which do not consider the order of authors, the relative size of nodes is more disproportionate. Although further analysis is required to draw a conclusion,

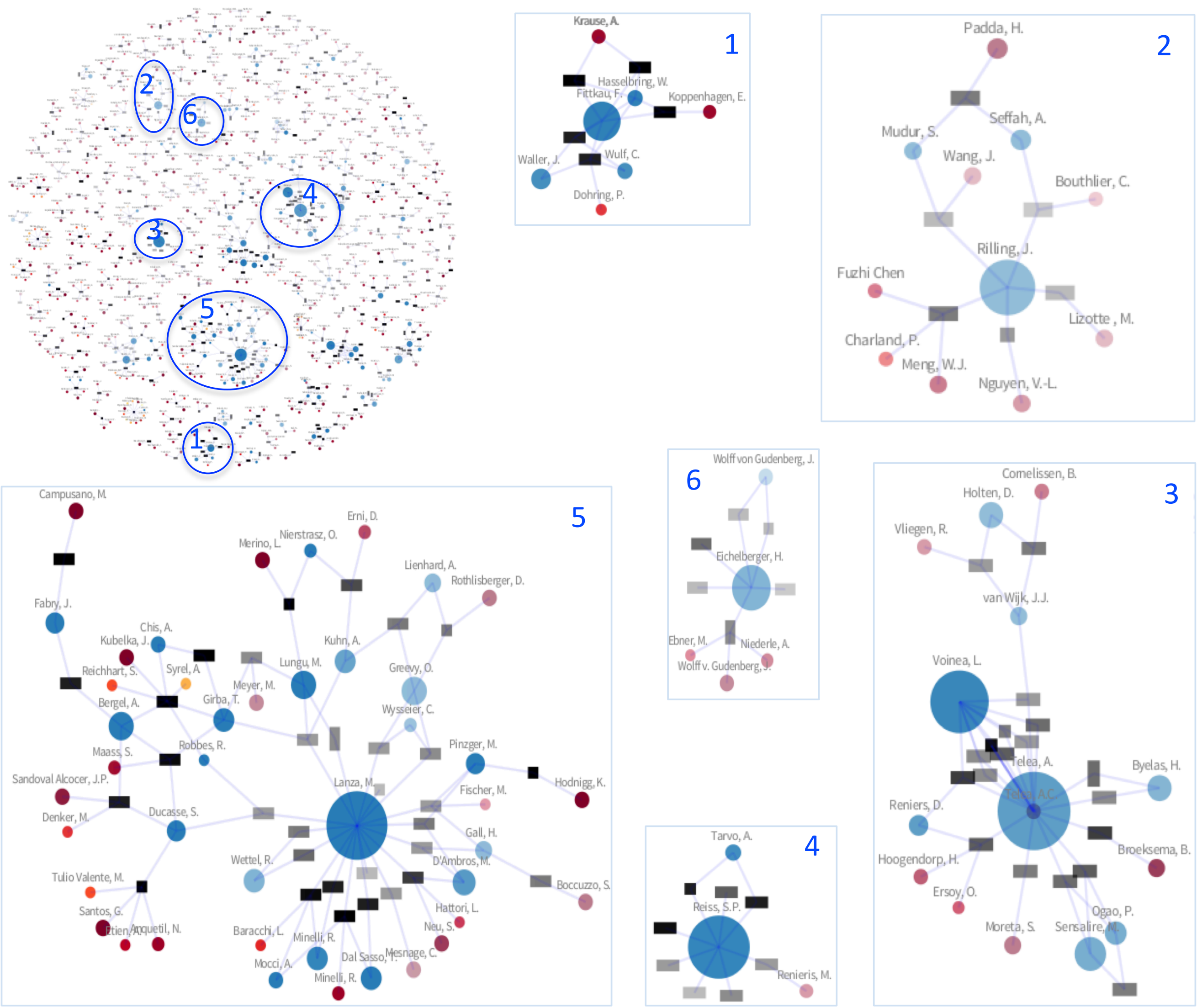

Figure 3. Exploring the collection of papers of VISSOFT/SOFTVIS conferences.

the results show how the visualization provides hints for that analysis.

Third, we analyze the evolution of groups. To this end, we classify the period of activity of groups into three categories: (i) resilient groups depicted by a gradient from light gray to black in the publication nodes (see Figure 3, groups 3, 4 and 5), (ii) aging groups that at some point of time stopped publishing (see Figure 3 groups 2 and 6) depicted by groups that only contain paper nodes in gray tones (not black), and (iii) new groups that have gained high impact in a few years, such as in Figure 3 group 1 (paper nodes mostly black). Figure 5 shows the evolution of two opposite groups. The one at the top expands over time, connecting to other groups and exposing a strong publication rate; although some parts of the network start to fade showing lack of publications in general, author nodes have strong colors showing the resilience of the group. The group at the bottom grows until 2008 and since then its colors have started to fade away.

4.2 RQ2. How does collaboration in the SOFTVIS/VISSOFT community differ from others?

We wonder whether the topology of the SOFTVIS/VISSOFT network is common or whether it exhibits peculiarities. We want to assess if the characteristics that we found in that community are present in another network as well. Thus we compare the network with another community: IWST.[3] We gather the complete set of 104 papers published in the 11 editions from 2003 to 2015. We followed the same steps of the workflow described in Section 3 and produced the visualizations shown in Figure 6 that correspond to SOFT-

[3] International Workshop on Smalltalk Technologies

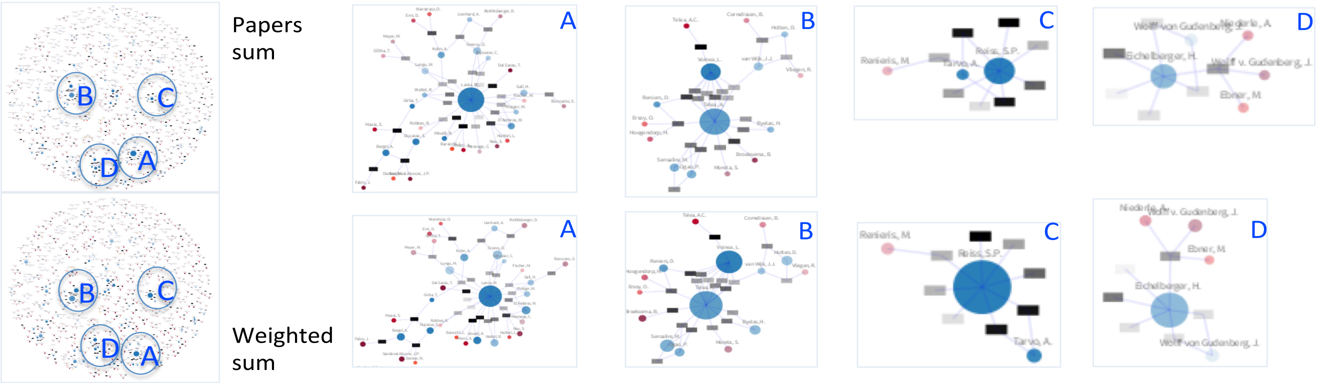

Figure 4. Comparing impact of authors (four groups) encoded in the size of circles using (top) number of papers, and (bottom) weighted sum.

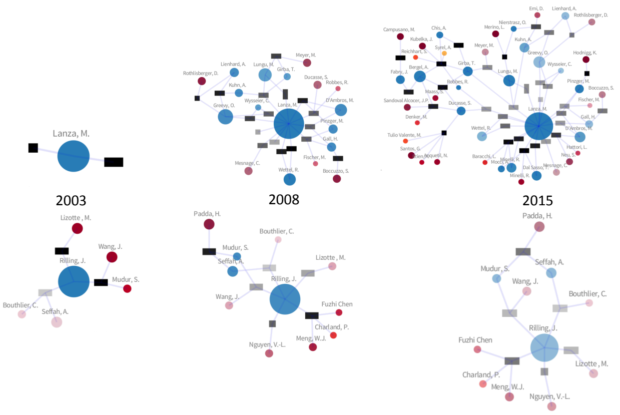

Figure 5. Evolution of two collaboration groups in the VISSOFT community. On top we have a group that organically grows over time. Strong colors show that the group is continuously publishing papers in the community. In the bottom we see a group that stopped growing in 2008. Colors start to fade out denoting an ageing group.

VIS/VISSOFT community (left) and IWST (right). In the IWST community we observe more collaboration than in SOFTVIS/VISSOFT. Beyond noticing that the IWST network is smaller than that of SOFTVIS/VISSOFT, we recognize that they expose different topologies. We can see that author nodes are much more connected in IWST than in SOFTVIS/VISSOFT. In the IWST community we find a large main collaboration group in the centre of the visualization that includes most of the nodes in the network. In contrast, in SOFTVIS/VISSOFT we find several medium size groups and a huge number of small collaborations. In IWST, contributors are more resilient than those in SOFTVIS/VISSOFT, who are more volatile. We note that in IWST blue nodes are predominant (authors with several publications), meanwhile in SOFTVIS/VISSOFT we observe a mix of blue with the gradient reddish tones.

Then we focus on the IWST network to analyze the structure of the main group that we found. We identify a main contributor at the centre that is surrounded by papers on which he has collaborated. Connected to those papers are his main co-authors (large blue nodes), which are surrounded by paper nodes that connect with their co-authors as well. Although in the visualization the position of nodes is not fixed, we perceive that in most cases the proximity of author nodes provides a hint of co-authorship. The more collaborations authors have, the closer they get. Authors at the centre denote collaborations with multiple groups.

4.3 RQ3. How suitable are bigraphs (as opposed to graphs) for modeling collaboration networks?

We have hypothesized that modelling collaboration networks as bigraphs can help us to avoid edge overlapping. Figure 7 shows the collaboration network of the SOFTVIS/VISSOFT (left) and IWST (right) communities modelled as graphs. Although we realize that the graph-based network provides a good notion of the size and number of groups, we perceive that in the bigraph-based (shown in Figure 6) network nodes and edges are better identified. The explicit representation of papers in the bigraph facilitates the identification of relevant authors that naturally move apart from their co-authors. We analyse the number of edges and whether they overlap in both graph types. Programmatically querying the visualization to retrieve the number of edges that cross is a fairly easy task in Roassal since each element in the visualization is a first-class object that can be queried. In Table 1 we summarize the results. We found that the number of edges in the graph doubled the edges in the bigraph, and that edge overlapping in the bigraph is one fifth of the overlapping in the graph, both using the default configuration of the force-based layout provided by Roassal.[4]

[4] charge := -30. gravity := 0.1. friction := 0.9 (among other parameters)

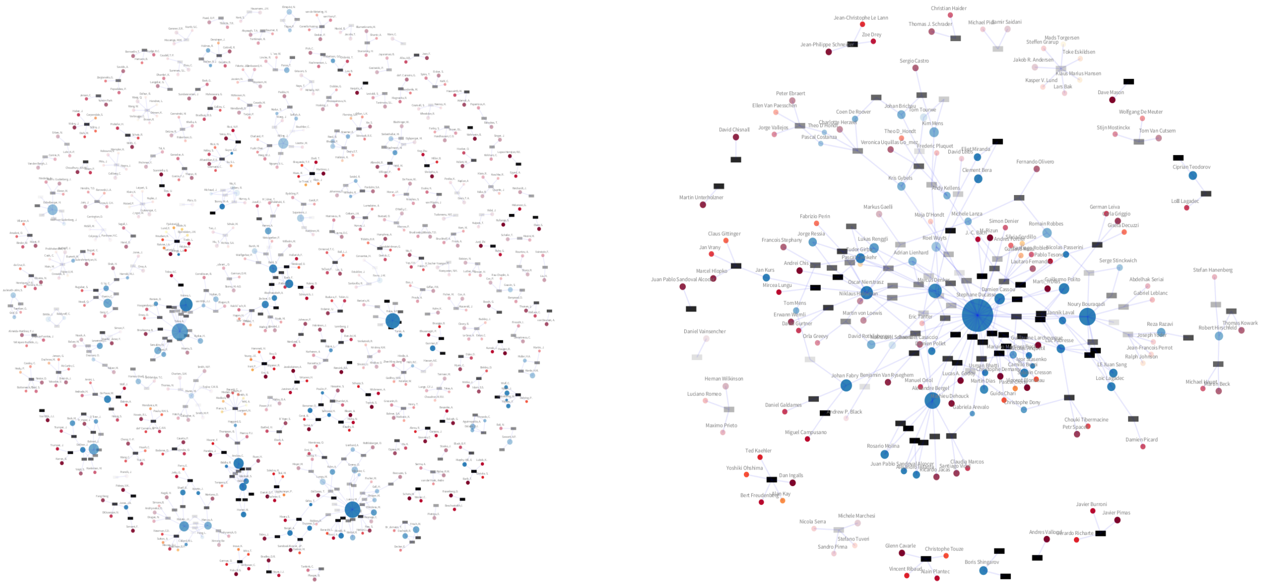

Figure 6. Comparing collaboration networks of SOFTVIS/VISSOFT (left) and IWST (right) modelled as bigraphs.

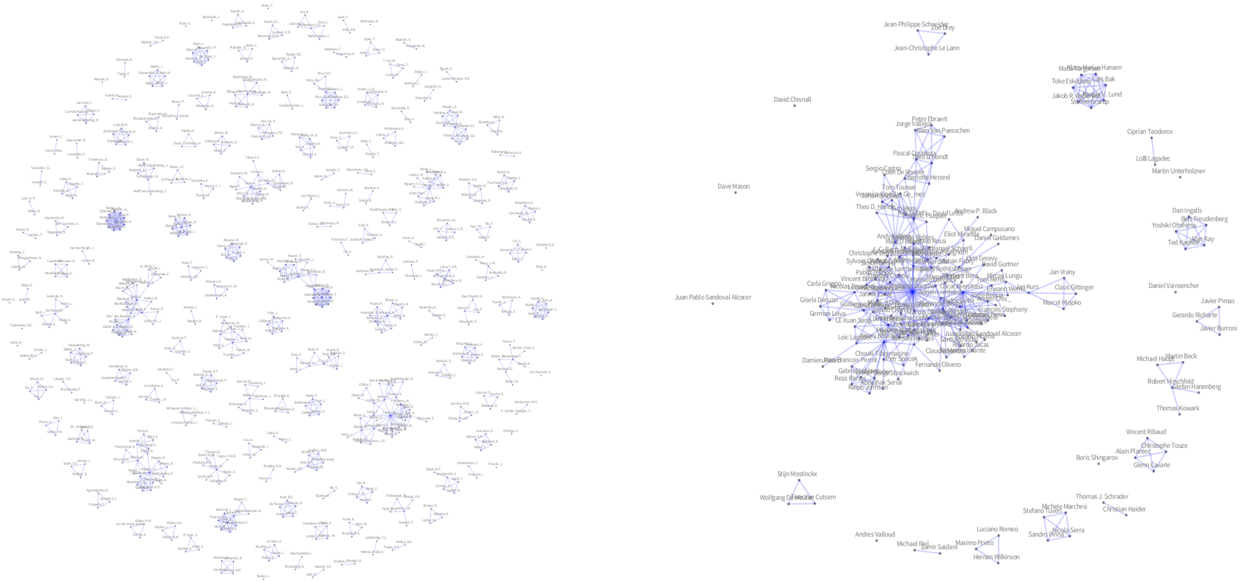

Figure 7. Comparing collaboration networks of SOFTVIS/VISSOFT (left) and IWST (right) modelled as a graph.

The number of edge crossings does not, however, convey the magnitude of the gain since that number depends on the total number of edges in the visualization that can potentially cross. In Figure 8 we compare the normalized values of edge crossings using for each type of graph in both communities. Although the chart shows limited benefit (10%) for the visualization of the SOFTVIS/VISSOFT community, it shows a big impact for the visualization of the IWST community.

Table 1. Characteristics of collaboration networks of VISSOFT/SOFTVIS and IWST communities modelled as a graph and as a bigraph.

Community	Type	Nodes	Edges	Overlap
SOFTVIS /VISSOFT	Graph	639	1852	6653
	Bigraph	954	876	1401
IWST	Graph	162	670	8973
	Bigraph	264	317	1185

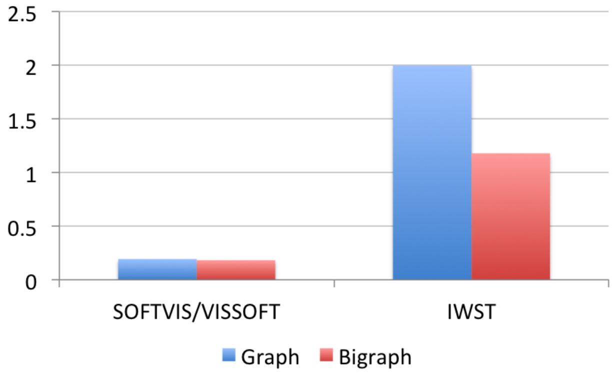

Figure 8. Comparing edge-crossing between collaboration network modelled as a graph and as a bigraph in IWST and SOFTVIS/VISSOFT communities.

5. Conclusion and Future Work

In this paper we have proposed the CommunityExplorer, a framework for visualization of collaboration networks. The framework introduced a visualization that models the collaboration network as a bigraph. As a result, we reduced overlapping edges to make room for encoding properties in the size and in the intensity of the color of nodes. We demonstrated its utility through a case study visualizing the publications of the VISSOFT/SOFTVIS community and contrasting the results to the visualization of collaboration in IWST community. We showed how the visualization allows us to identify groups, to assess their relevance and to analyze their evolution.

In the future we plan to expand the framework by designing new visualization techniques to increase the properties that can be mapped, experimenting with parameters of the layout to improve the readability of the visualization, and investigating visualizations better suited to compare collaboration among different communities.

Acknowledgments

We gratefully acknowledge the financial support of the Swiss National Science Foundation for the project "Agile Software Analysis" (SNSF project No. 200020-162352, Jan 1, 2016 - Dec. 30, 2018). Merino has been partially funded by CONICYT BCH/Doctorado Extranjero 72140330.

References

V. P. Araya, A. Bergel, D. Cassou, S. Ducasse, and J. Laval. Agile visualization with Roassal. In *Deep Into Pharo*, pages 209–239. Square Bracket Associates, Sept. 2013. ISBN 978-3-9523341-6-4.

M. Biryukov and C. Dong. Analysis of computer science communities based on dblp. In *Research and advanced technology for digital libraries*, pages 228–235. Springer, 2010.

U. Brandes, P. Kenis, J. Lerner, and D. van Raaij. Network analysis of collaboration structure in wikipedia. In *Proceedings of the 18th international conference on World wide web*, pages 731–740. ACM, 2009.

H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles. Collabseer: a search engine for collaboration discovery. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, pages 231–240. ACM, 2011.

A. Chiş, T. Gîrba, O. Nierstrasz, and A. Syrel. The Moldable Inspector. In *Proceedings of the 2015 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, Onward! 2015, pages 44–60, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3688-8. doi: 10.1145/2814228.2814234. URL `http://scg.unibe.ch/archive/papers/Chis15a-MoldableInspector.pdf`.

M. Dörk, N. H. Riche, G. Ramos, and S. Dumais. Pivotpaths: Strolling through faceted information spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2709–2718, 2012.

C. Dunne and B. Shneiderman. Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3247–3256. ACM, 2013.

E. R. Gansner, Y. Hu, S. North, and C. Scheidegger. Multilevel agglomerative edge bundling for visualizing large graphs. In *Visualization Symposium (PacificVis), 2011 IEEE Pacific*, pages 187–194. IEEE, 2011.

Y. Hu and L. Shi. Visualizing large graphs. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(2):115–136, 2015.

J. Huang, Z. Zhuang, J. Li, and C. L. Giles. Collaboration over time: characterizing and modeling network evolution. In *Proceedings of the 2008 international conference on web search and data mining*, pages 107–116. ACM, 2008.

T.-H. Huang and M. L. Huang. Analysis and visualization of co-authorship networks for understanding academic collaboration and knowledge domain of individual researchers. In *Computer Graphics, Imaging and Visualisation, 2006 International Conference on*, pages 18–23. IEEE, 2006.

F. Osborne, E. Motta, and P. Mulholland. Exploring scholarly data with Rexplore. In *The Semantic Web–ISWC 2013*, pages 460–477. Springer, 2013.

J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

Y. Tymchuk, A. Mocci, and M. Lanza. Collaboration in open-source projects: myth or reality? In *Proceedings of the 11th working conference on mining software repositories*, pages 304–307. ACM, 2014.

C. Vehlow, F. Beck, and D. Weiskopf. The state of the art in visualizing group structures in graphs. In *Eurographics Conference on Visualization (EuroVis)-STARs*, pages 21–40, 2015.

B. Wu, F. Zhao, S. Yang, L. Suo, and H. Tian. Characterizing the evolution of collaboration network. In *Proceedings of the 2nd ACM workshop on Social web search and mining*, pages 33–40. ACM, 2009.