

# ColTop: Visual Topic-based Analysis of Scientific Community Structure

Moataz Abdelaal

*Institute for Visualization and Interactive-Systems (VIS), University of Stuttgart  
Stuttgart, Germany  
mottazabdelfattah@gmail.com*

Florian Heimerl

*Department of Computer Sciences  
University of Wisconsin – Madison  
Madison, WI, USA  
heimerl@cs.wisc.edu*

Steffen Koch

*Institute for Visualization and Interactive-Systems (VIS), University of Stuttgart  
Stuttgart, Germany  
steffen.koch@vis.uni-stuttgart.de*

Speaker: Robert Krueger

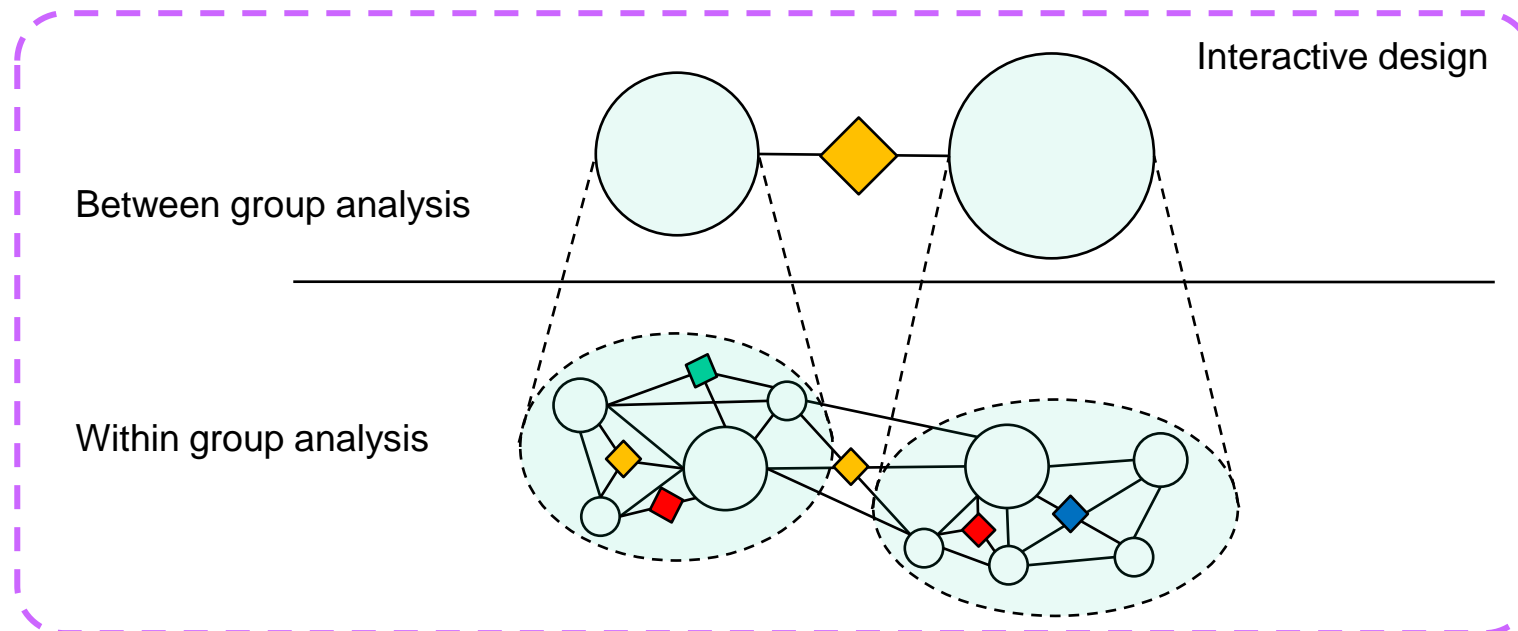
# Motivation

- A funding agent is deciding which groups or researchers to fund in a specific research community
  - Who are the key groups and authors?
  - What are the different research interests?
  - Collaboration information
    - Frequent collaborators?
    - Collaborations between / within groups?
    - Collaborations patterns?
    - Contents of the publications

} HOW  
+  
} WHAT

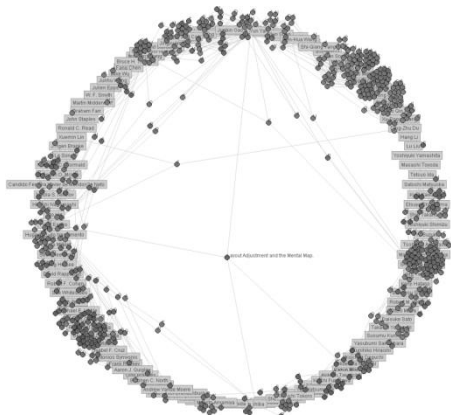
# Our Contributions

- Combining co-authors information with topic-profile information
- Supporting multi-level analysis (overview & details)
- Interactive design

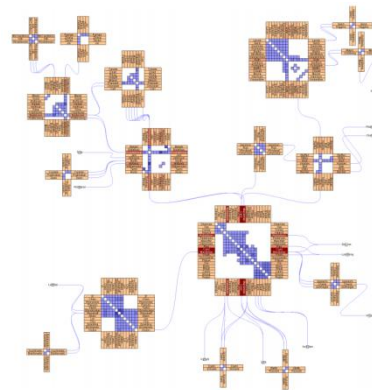


# Current Approaches

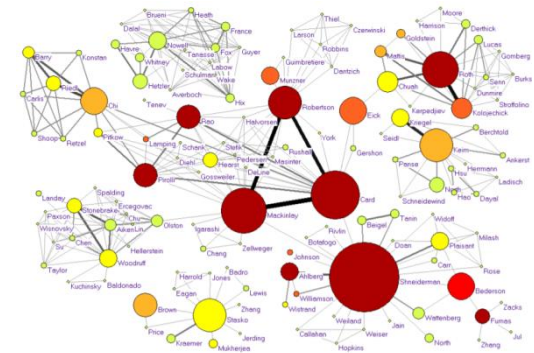
	Graph based (R. Ichise et al. 2005, K. Misue 2008)	Matrix based (Matlink 2007, NodeTriX 2007)	Others (W. Ke et al. 2004, B. Alper et al. 2011)
Combine co-author information with other meta-data information	✗	✗	✓
Analysis on multiple level	✗	✗	✗



*K. Misue 2008*



*NodeTriX (Henry et al. 2007)*



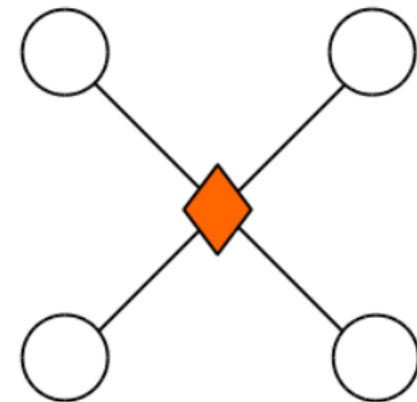
*W. Ke et al. 2004*

## Data Set

- All following examples: Vis publication data set
  - Available at: <http://vispubdata.org>
- All IEEE VIS(Week) publications from 1990-2015
- Total number of papers: 2752
- Total number of unique authors: 4890

# Visualization of Collaborations

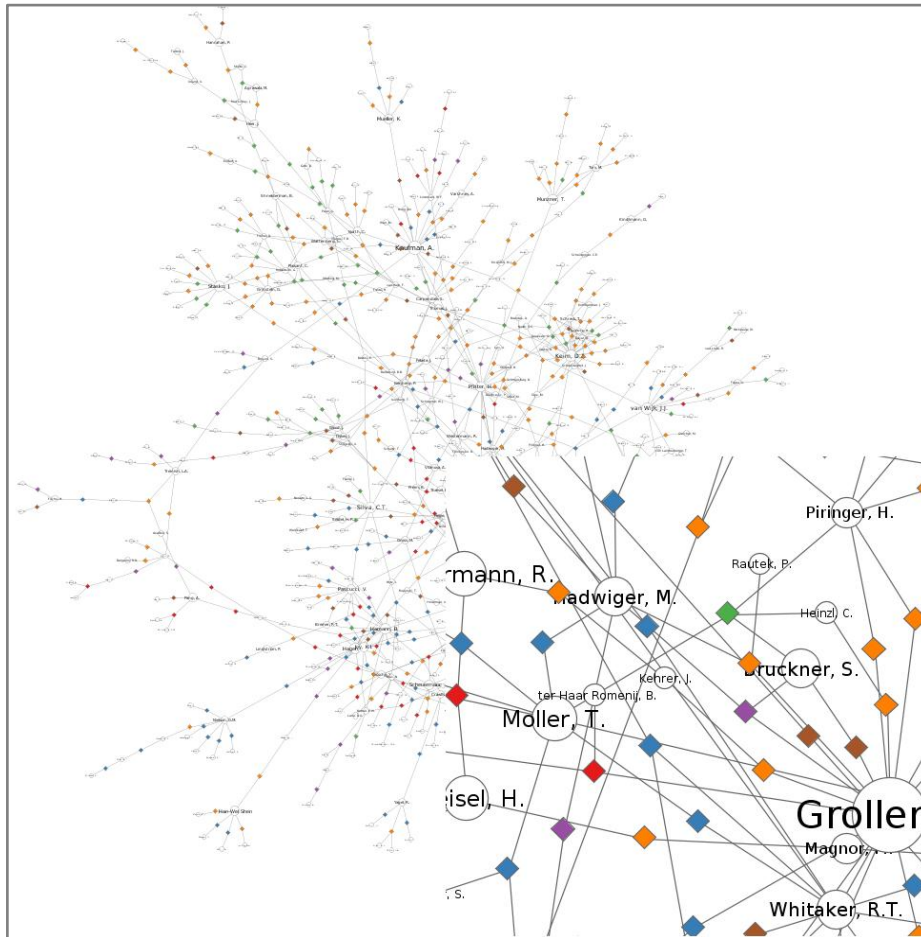
- Node-link diagram
- Each node is a
  - Research group (overview), or
  - Person (detail view)
- The size of node corresponds to num. of publications
- Force-directed layout (*Barnes et al. 1986*)
- Collaboration nodes
  - Can be activated by user
  - Collaborations prominent / less clutter
  - Diamond shape
  - Color encodes topic of publication



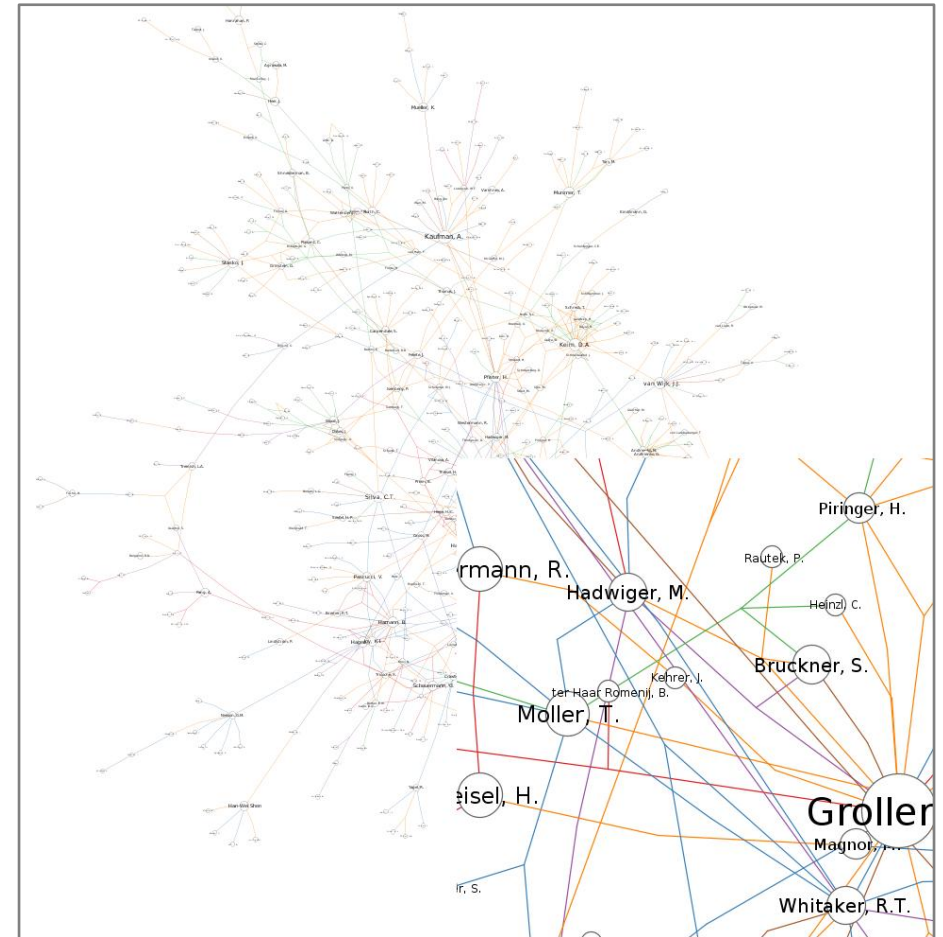
Collaborations between four groups



# Collaboration Node



Collaboration nodes are visible (the diamond shape)

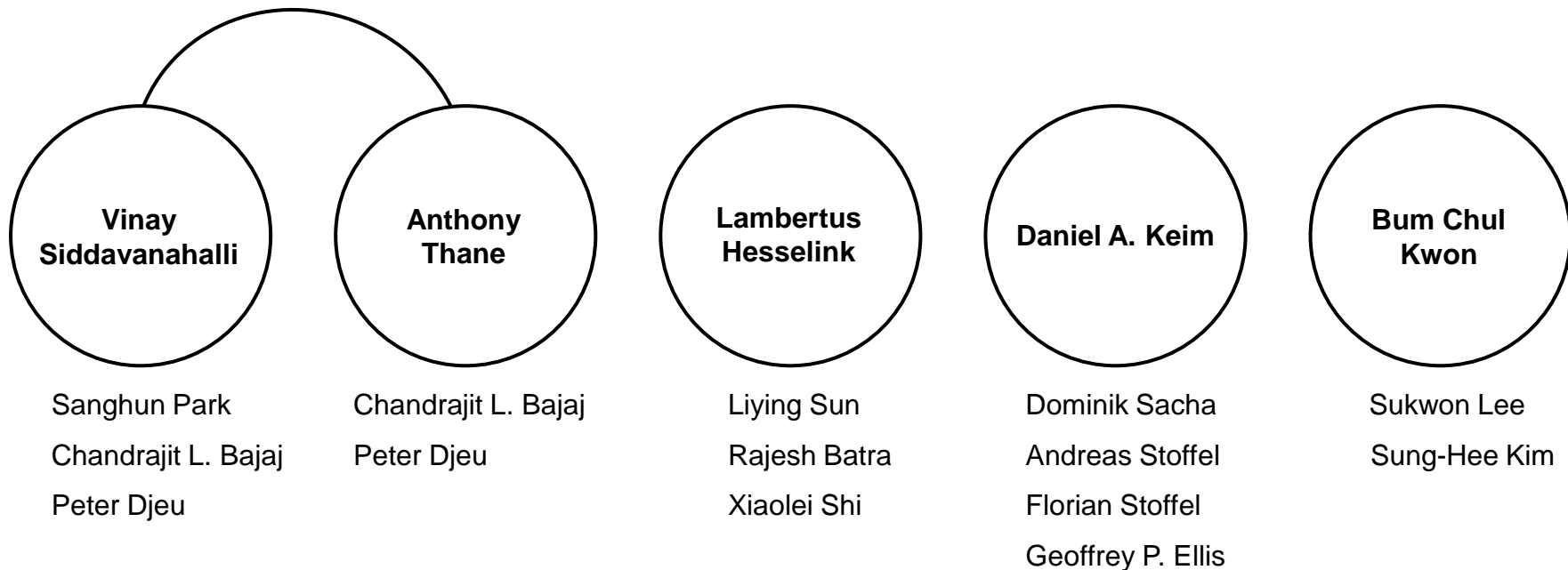


Collaboration nodes are hidden

# Creating the Overview: Grouping Strategy

## Researcher identification as groups

Year	Paper Title	Abstract	Author Names
2002	Case study: Interactive render	Adaptive me	Sanghun Park;Chandrajit L. Bajaj;Vinay Siddavanahalli
2004	TexMol: interactive visual exp	While molec	Chandrajit L. Bajaj;Peter Djeu;Vinay Siddavanahalli;Anthony Thane
2004	Topology visualization of the	An ideal visu	Liyong Sun;Rajesh Batra;Xiaolei Shi;Lambertus Hesselink
2014	Knowledge Generation Model	Visual analyt	Dominik Sacha;Andreas Stoffel;Florian Stoffel;Bum Chul Kwon;Geoffrey P. Ellis;Daniel A. Keim
2016	VLAT: Development of a Visu	The Informa	Sukwon Lee;Sung-Hee Kim;Bum Chul Kwon





## Collaboration Topics: Topic Modeling

- We use the popular LDA algorithm (applied to paper abstracts)
- Process
  - Pre-processing
  - Topic modelling
  - Assign most likely topic to each publication
- Top 10 words characterize each topic
- Each topic has a unique color

**Topic 1:** field flow feature structure vector line particle  
fluid simulation surface

**Topic 2:** volume algorithm rendering based efficient  
quality function point surface image

**Topic 3:** analytic process challenge research visual  
analyst event temporal system tool

**Topic 4:** environment interactive software system  
computer graphic interface processing display

**Topic 5:** large dimensional exploration number  
representation visual information set result space

The top 10 keywords per topic.

Topic 1: Flow Visualization,

Topic 2: Volume Visualization,

Topic 3: Visual Analytics,

Topic 4: Human Computer Interaction (HCI),

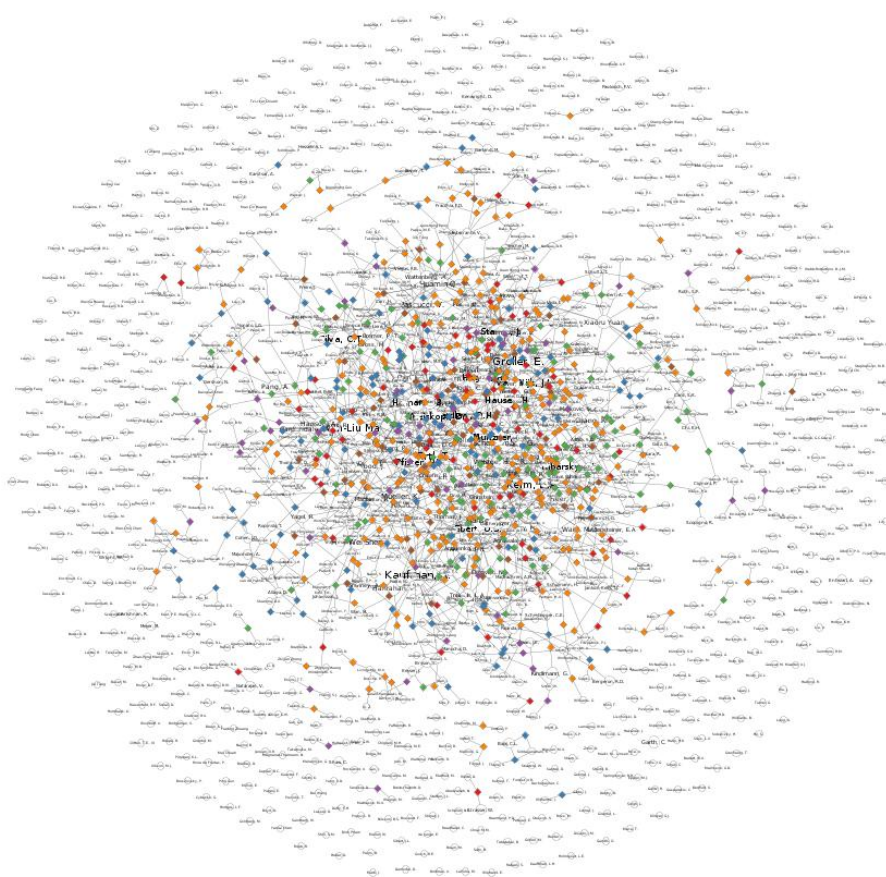
Topic 5: Information Visualization

*(One possible interpretation )*

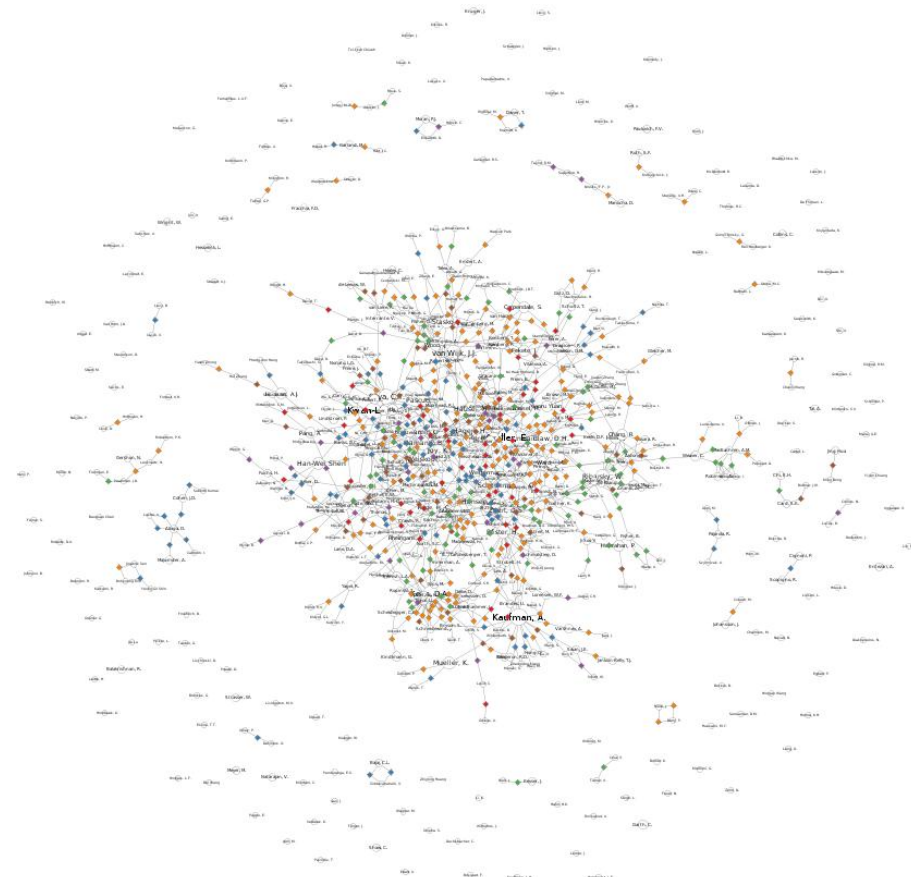
# Interaction Techniques

- Overview first
  - Zoom-out and panning
  - Highlight neighbors on hovering
  - Show/hide the collaboration nodes
- Zoom and filtering
  - By number of publications
  - By number of members
  - By topic
  - Search functionality
- Details on demand: selection
  - Tabbed panel for detailed analysis
- Details on demand: drill-down
  - See the inside view of the group
  - Relation with the upper-level

# Interaction Example: Exploration & Filtering



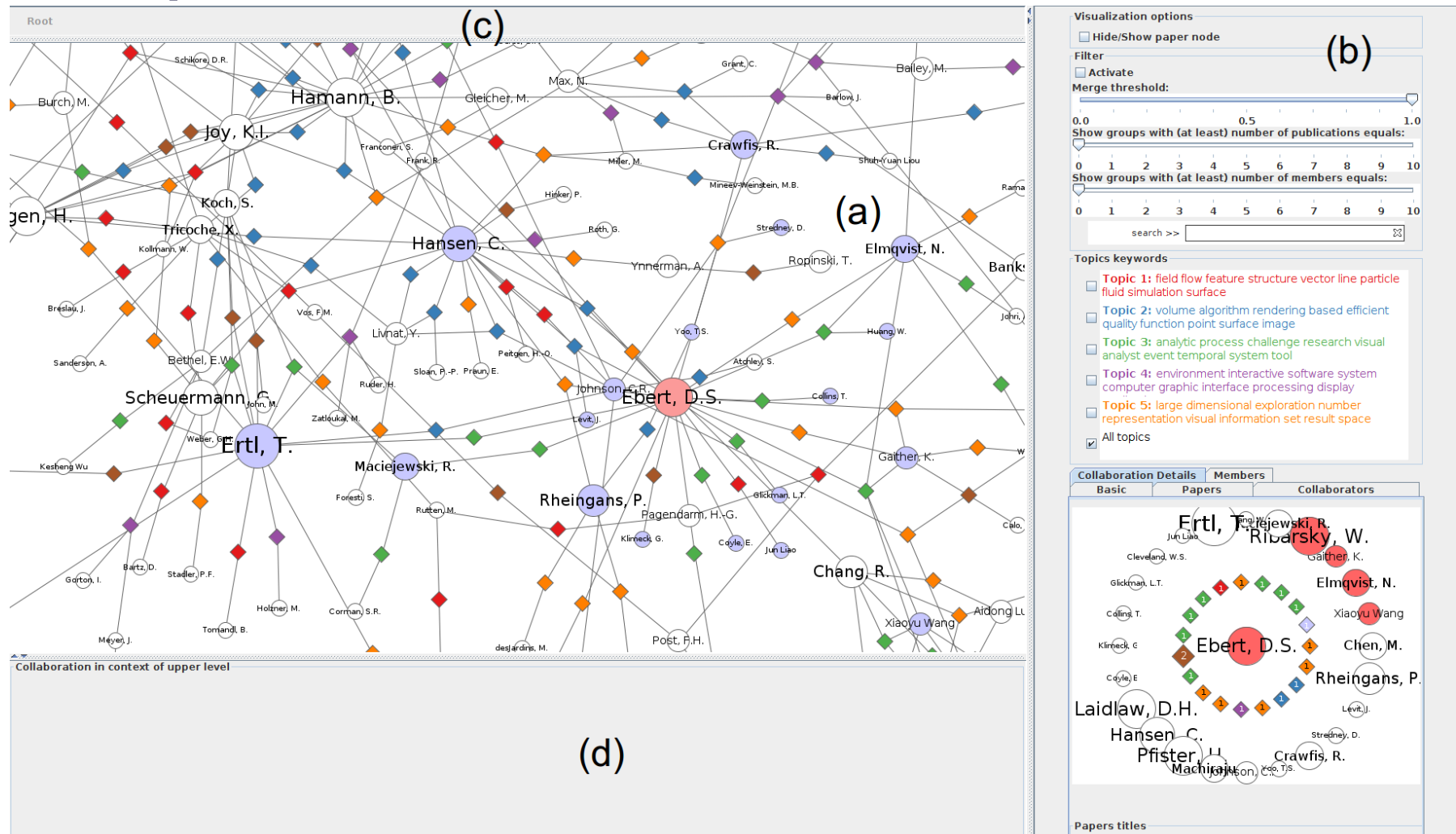
Before filtering



After filtering

- Group members  $\geq 1$
- No. publications  $\geq 2$

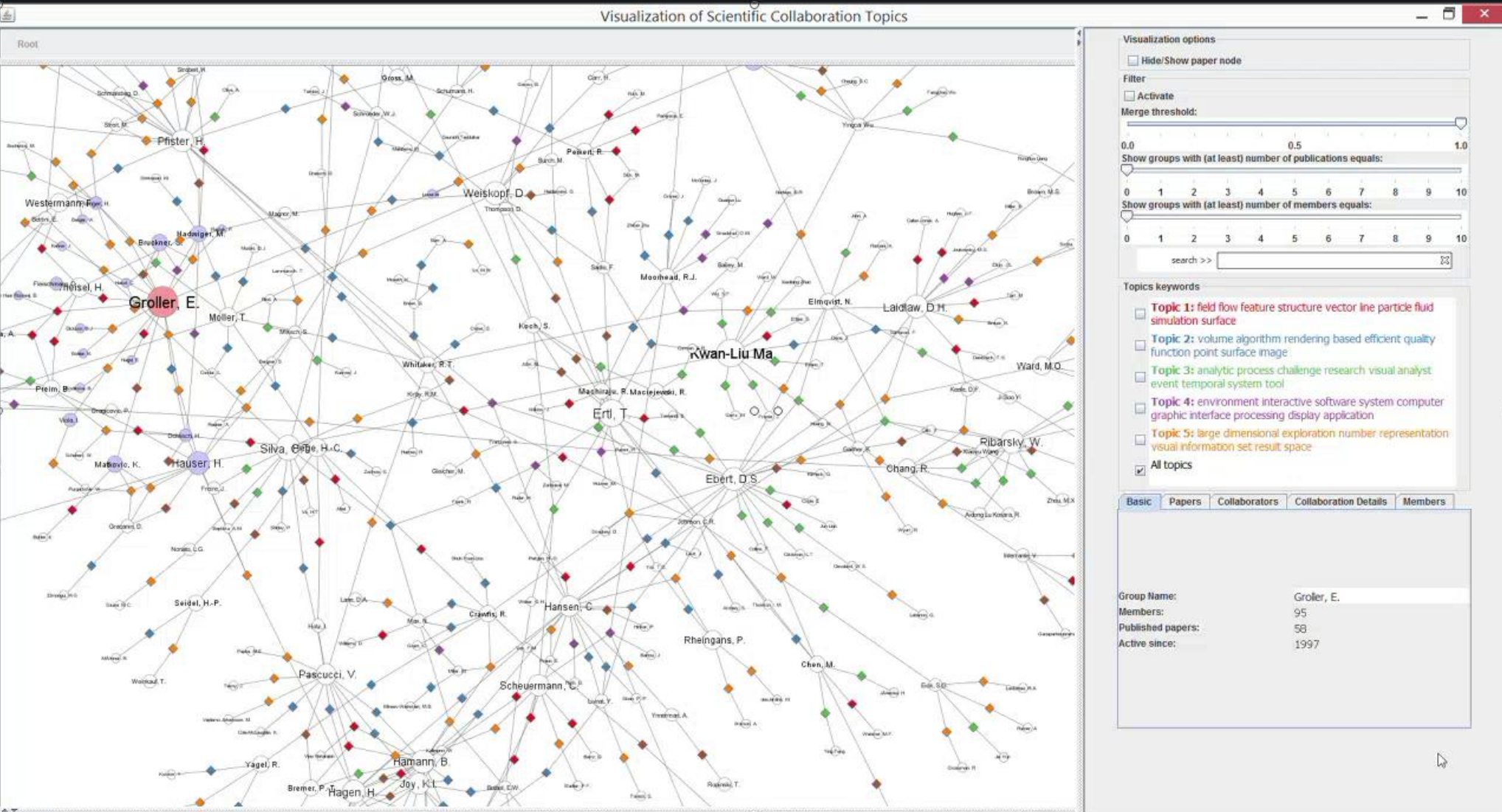
# ColTop: Main Screen



The main screen of our system. (a) The graph view shows collaboration topics between groups. (b) The interaction panel. (c) The path bar. (d) The context panel.



# Analysis Scenario



## Conclusion

- A multi-level, interactive graph visualization system that combines co-author information with topic profile information
  
- Future Work
  - Visualize the temporal dynamics and evolution of topic collaboration over time
  - Conduct a user study to further evaluate the capability our system



## References

- R. Ichise, H. Takeda, and K. Ueyama, “Community mining tool using bibliography data,” in *Ninth International Conference on Information Visualisation (IV’05)*, July 2005, pp. 953–958.
- K. Misue, “Visual analysis tool for bipartite networks,” in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2008, pp. 871–878.
- W. Ke, K. Borner, and L. Viswanath, “Major information visualization authors, papers and topics in the acm library,” in *Proceedings of the IEEE Symposium on Information Visualization*, ser. INFOVIS ’04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 216.1–. [Online]. Available: <http://dx.doi.org/10.1109/INFOVIS.2004.45>
- B. Alper, N. Riche, G. Ramos, and M. Czerwinski, “Design study of linesets, a novel set visualization technique,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2259–2267, Dec 2011.
- N. Henry and J.-D. Fekete, “Matrixexplorer: a dual-representation system to explore social networks,” *IEEE transactions on visualization and computer graphics*, vol. 12, no. 5, pp. 677–684, 2006.
- ———, “Matlink: Enhanced matrix visualization for analyzing social networks,” in *IFIP Conference on Human-Computer Interaction*. Springer, 2007, pp. 288–302.

## References

- N. Henry, J.-D. Fekete, and M. J. McGuffin, “Nodetrix: a hybrid visualization of social networks,” *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, pp. 1302–1309, 2007.
- D. Newman, A. Asuncion, P. Smyth, and M. Welling, “Distributed algorithms for topic models,” *Journal of Machine Learning Research*, vol. 10, no. Aug, pp. 1801–1828, 2009.
- A. K. McCallum, “Mallet: A machine learning for language toolkit,” 2002, <http://mallet.cs.umass.edu>.
- J. Barnes and P. Hut, “A hierarchical  $O(n \log n)$  force-calculation algorithm,” *nature*, vol. 324, no. 6096, pp. 446–449, 1986.
- J. Heer, S. K. Card, and J. Landay, “Prefuse: A toolkit for interactive information visualization,” in *ACM Human Factors in Computing Systems (CHI)*, 2005, pp. 421–430. [Online]. Available: <http://vis.stanford.edu/papers/prefuse>
- P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. D. Stolper, M. M. Sedlmair, J. Chen, T. Moller, and J. Stasko, “vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, 2017, to appear. [Online]. Available: <https://hal.inria.fr/hal-01376597>