Analyzing the impact of science funding programs on the evolution of research fields

William Ribarsky

University of North Carolina at Charlotte

A Visual Analytics Approach

A more comprehensive approach to reveal scientific impact & innovation.

- General data-driven methods for revealing trends, events, relationships, possible cause & effect.
- Interaction is essential for putting the human analyst in the loop.
- Tools that can be put in the hands of impact or policy analysts and program managers.

VIS CENTER

Visual Analytics Tools

Tools embodying general methods for comprehensive analysis:

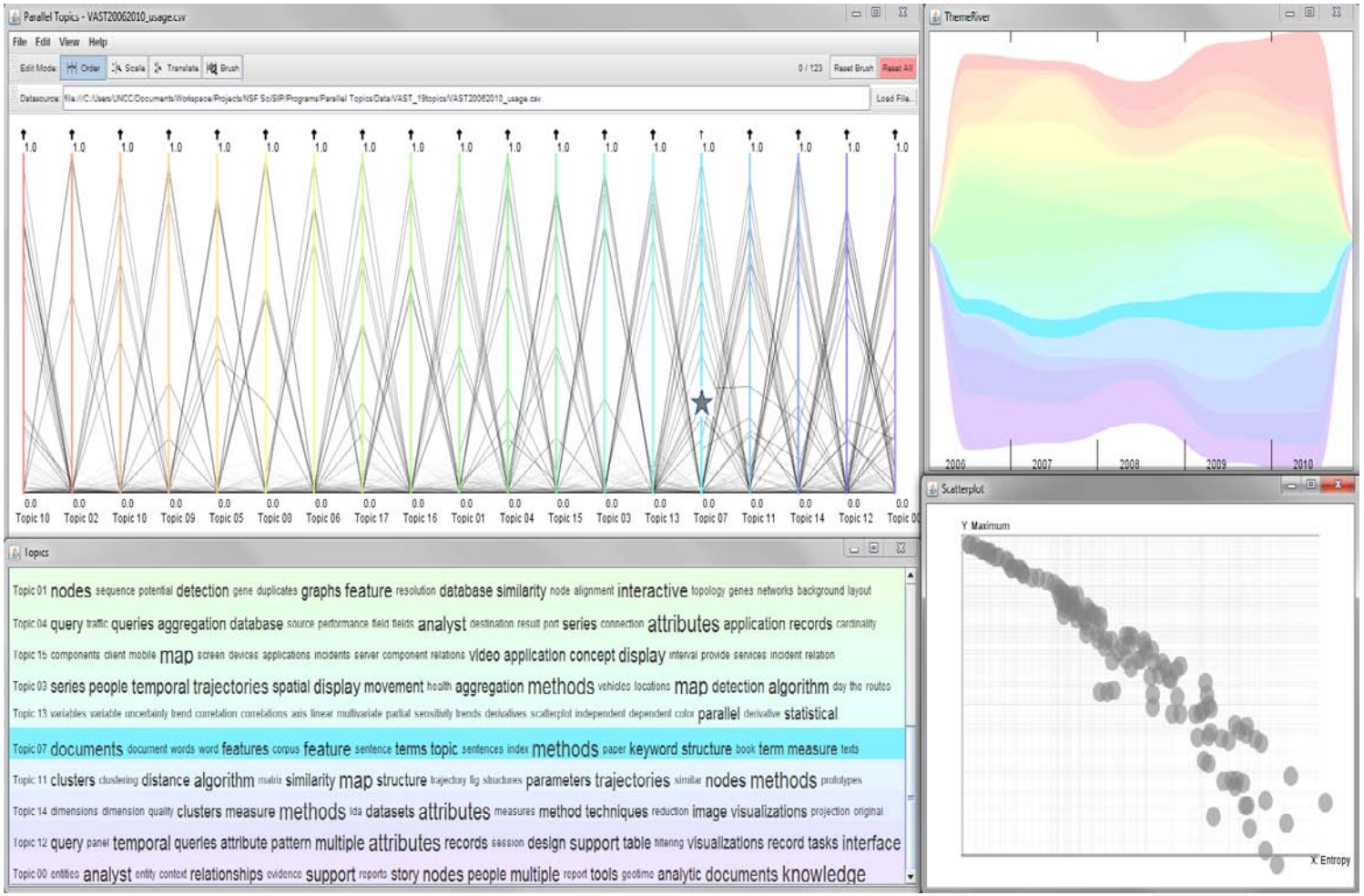
- Exploration, discovery, and general trends (Example: ParallelTopics, LeadLine)
- Deeper, detailed analysis (Example: Lead-Lag)
- Forecasting impact and decision support for science policy (Example: Knowledge Encapsulation Framework and dynamic Bayesian nets model)

VIS CENTER

ParallelTopics

Topic streams over time

Distribution of proposals over topics

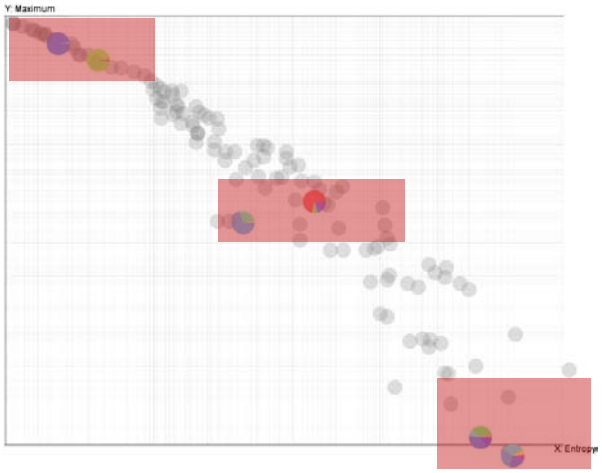


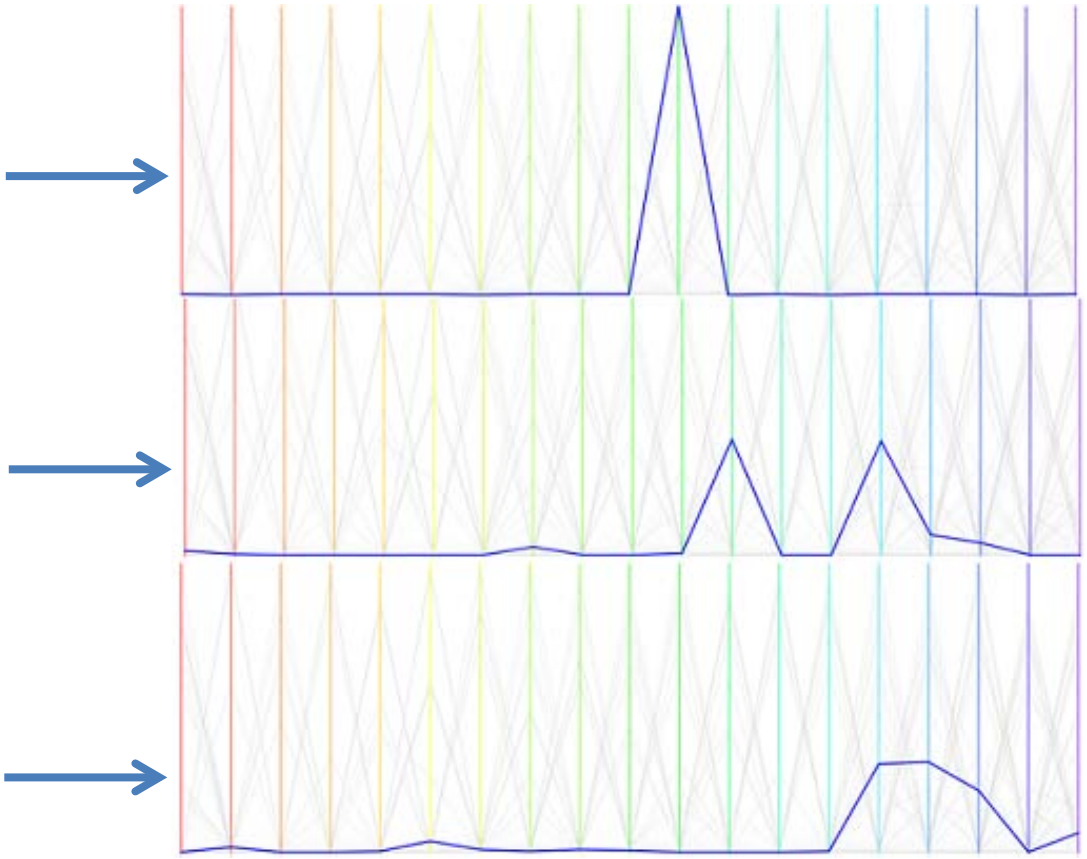
Key word lists for each topic

Distribution of proposals from single topic to multiple topic

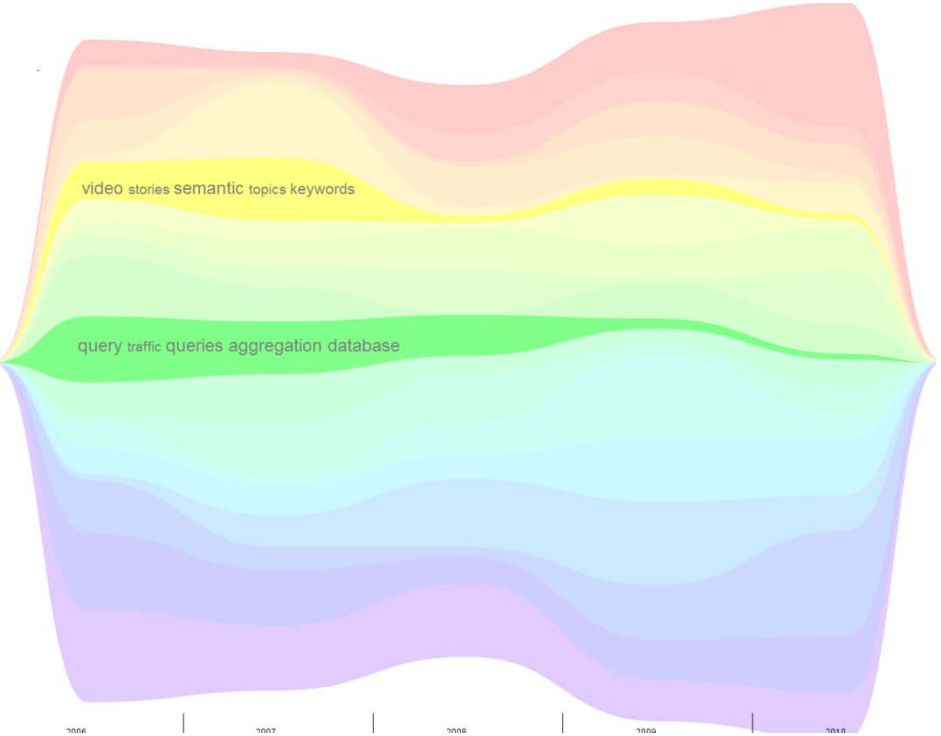
VIS CENTER

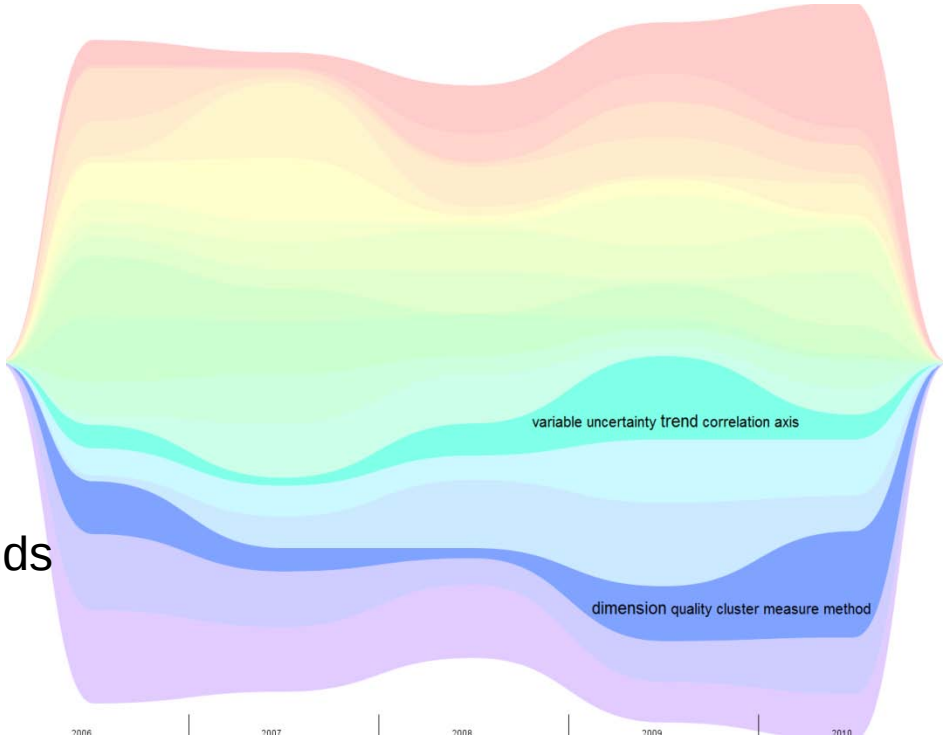
ParallelTopics





Topic Trends in ParallelTopics

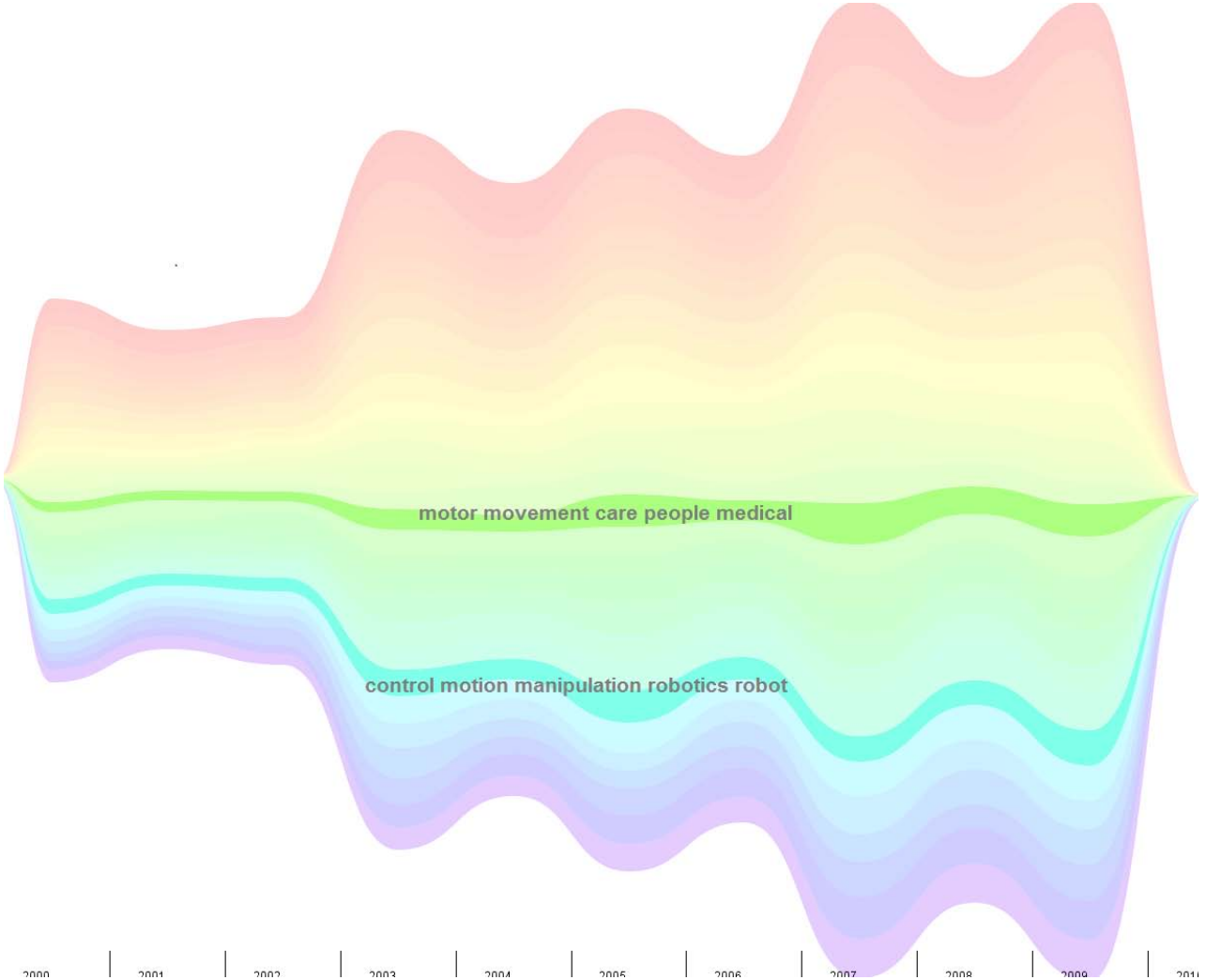




Visual Analytics research paper trends (2006-2010)

VIS CENTER

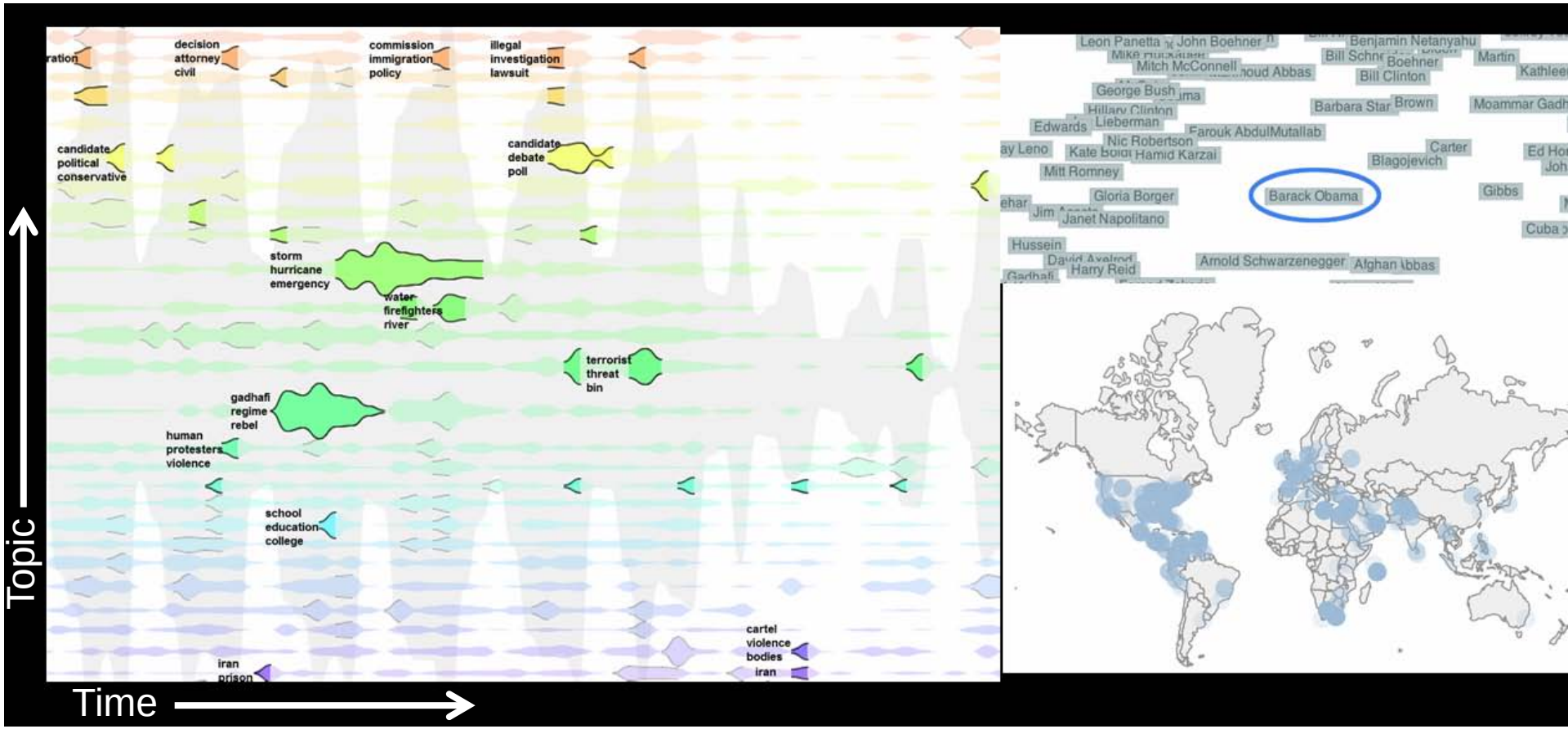
Topic Trends in ParallelTopics



Overview of awarded proposals in NSF Information & Intelligent Systems Directorate from 2000 to 2010.

LeadLine

Entity View



Event View

Geo View

Lead-Lag

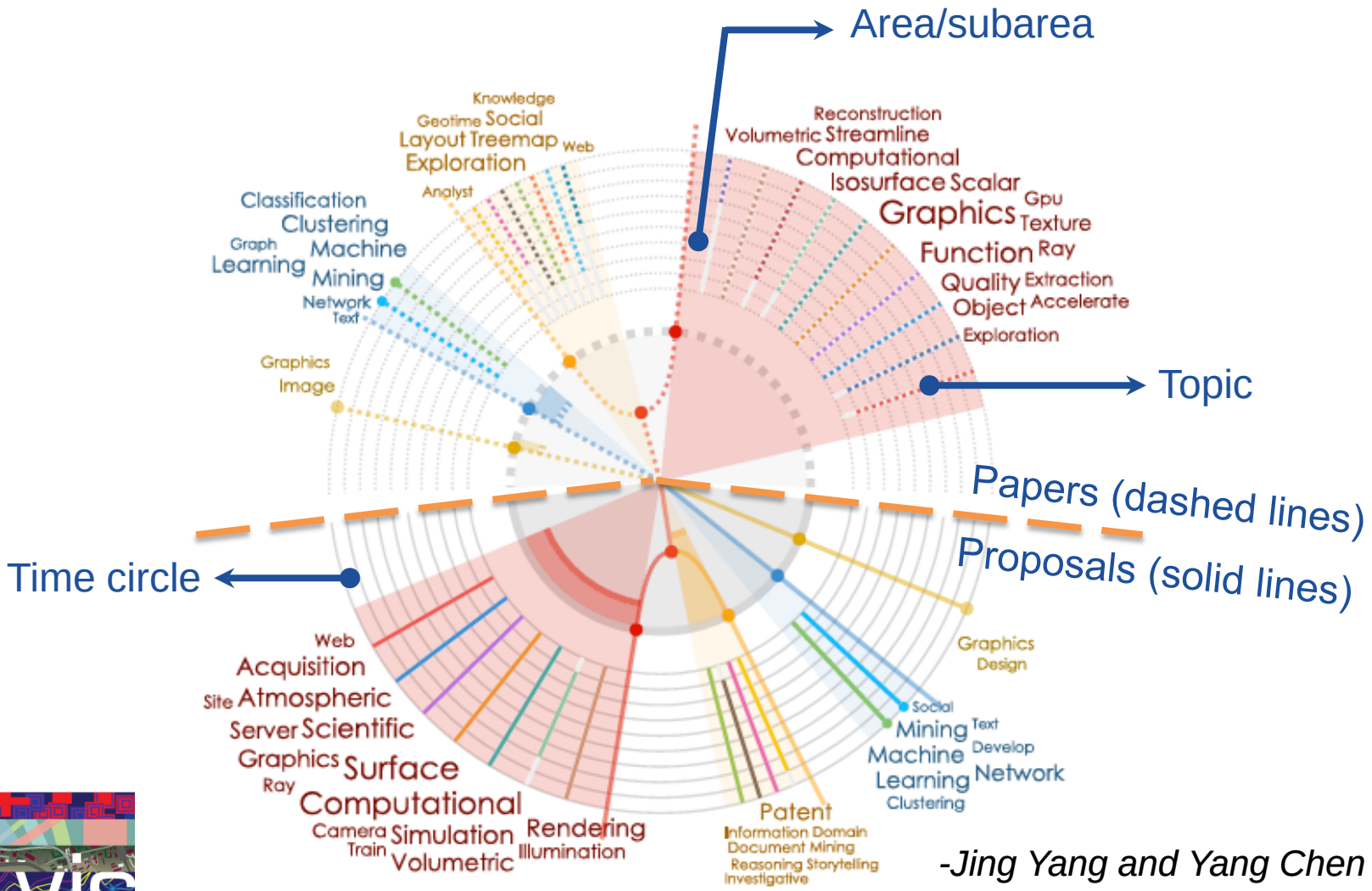
Investigate the lead-lag relations between funding and research (Collaboration of Jing Yang with Dr. Shixia Liu and her group at Microsoft Research Asia)

- Does funding lead/lag research in a given research field or topic?
- How do the relationships between funding and research evolve over time?
- What are the most important proposals/papers that shape the lead of funding/research?
- Has consequences for:
  - Scientific investment assessment
  - Funding allocation and program establishment
  - Review panel organization

Methodology

- Compare two or more corpora
- Perform topic modeling on the combined corpora to identify research areas, subareas, and topics.
- Conduct lead-lag analysis on multiple levels of detail
  - Overall lead
  - Dynamic change of lead over time
- Visualize the results
  - Radial tree: overall lead-lag
  - Twisted-ladder: temporal change
- Example
  - NSF funded grant abstracts (2000-2009)
    - 1K grant abstracts (IIS, CCF, and CNS)
  - Abstract of papers (2000-2009)
    - 5k conference papers (Vis, InfoVis, KDD, SIGGRAPH, etc.)

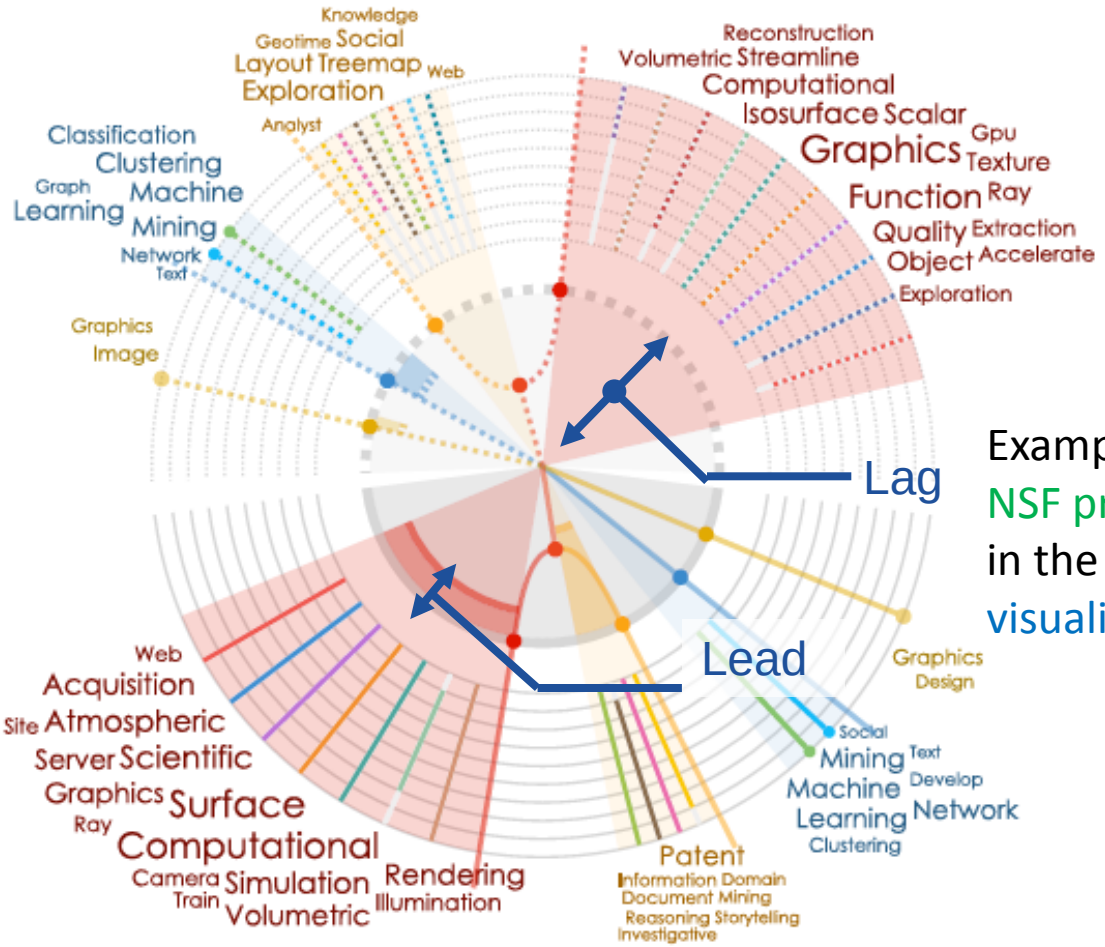
Lead-Lag
NSF vs. Papers: Overall leads



-Jing Yang and Yang Chen

VIS CENTER

NSF proposals vs. Papers: Overall leads

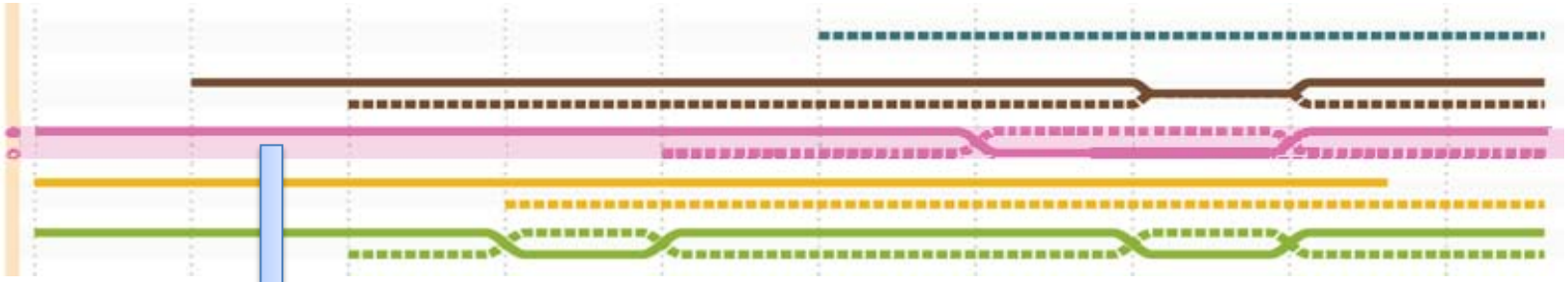


Example:
NSF proposals lead papers in the subarea of scientific visualization

VIS CENTER

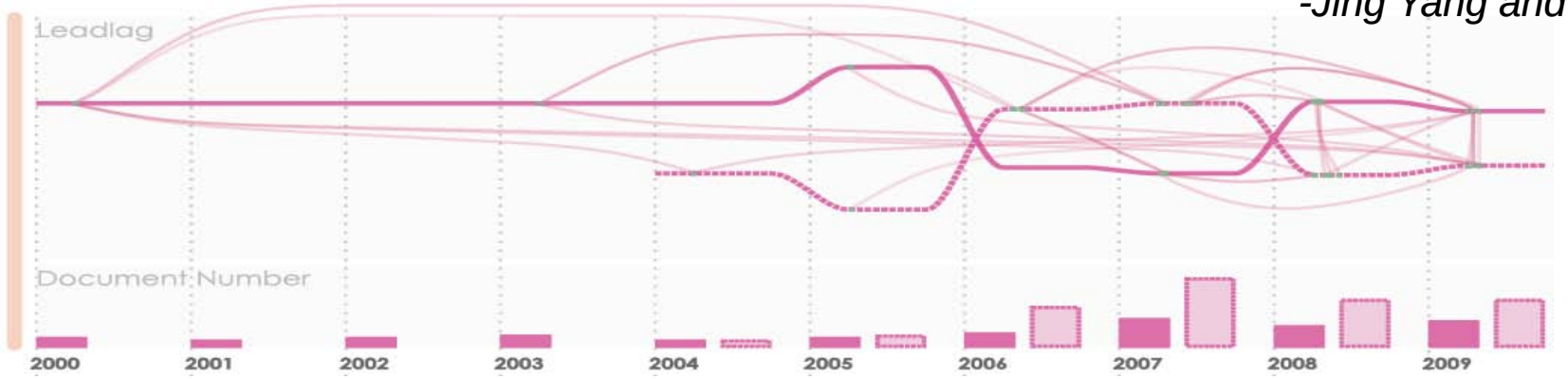
NSF Proposals vs. Papers: Local changes

Subarea of information visualization NSF leads Paper



A topic of "analytic/knowledge/reasoning". Paper (dash) leads in 2006 and 2007. NSF (solid) leads over the rest of the years.

-Jing Yang and Yang Chen

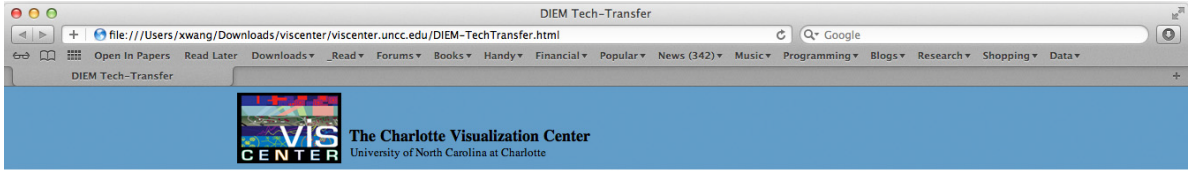


Document alignment between NSF and paper on the topic from 2000 to 2009.

Questions?

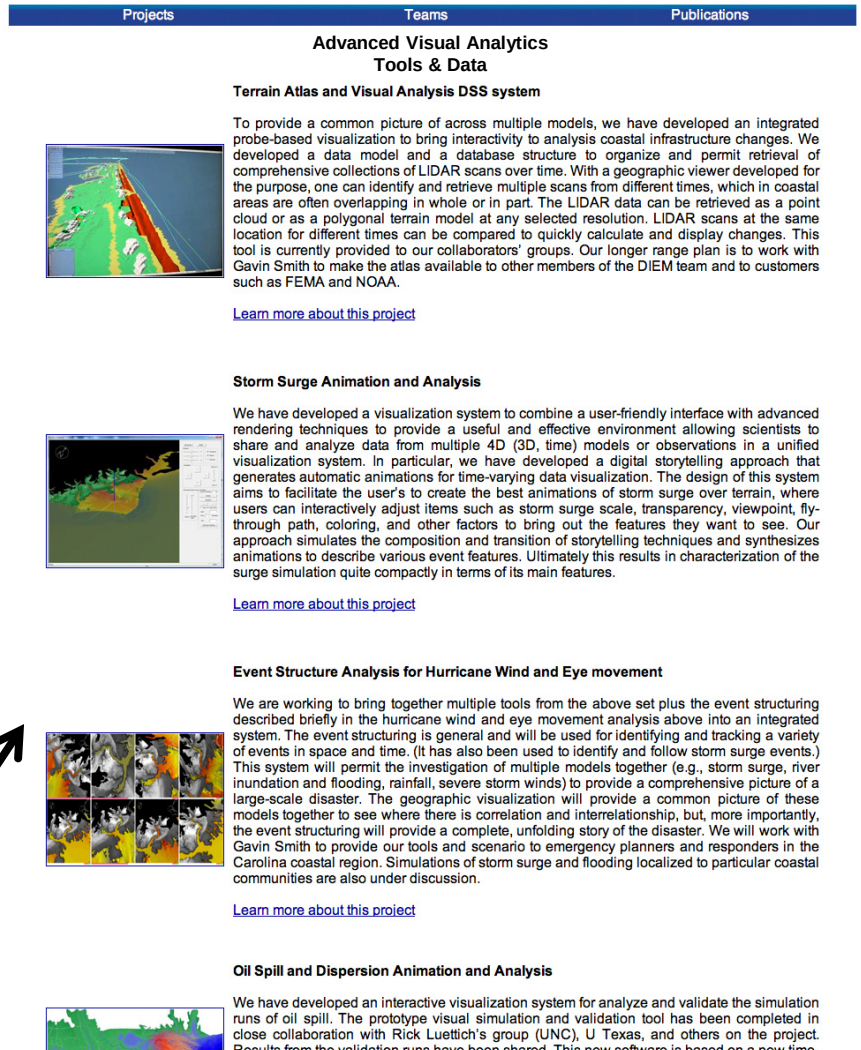
www.viscenter.uncc.edu

Products --- Tools & Data Website

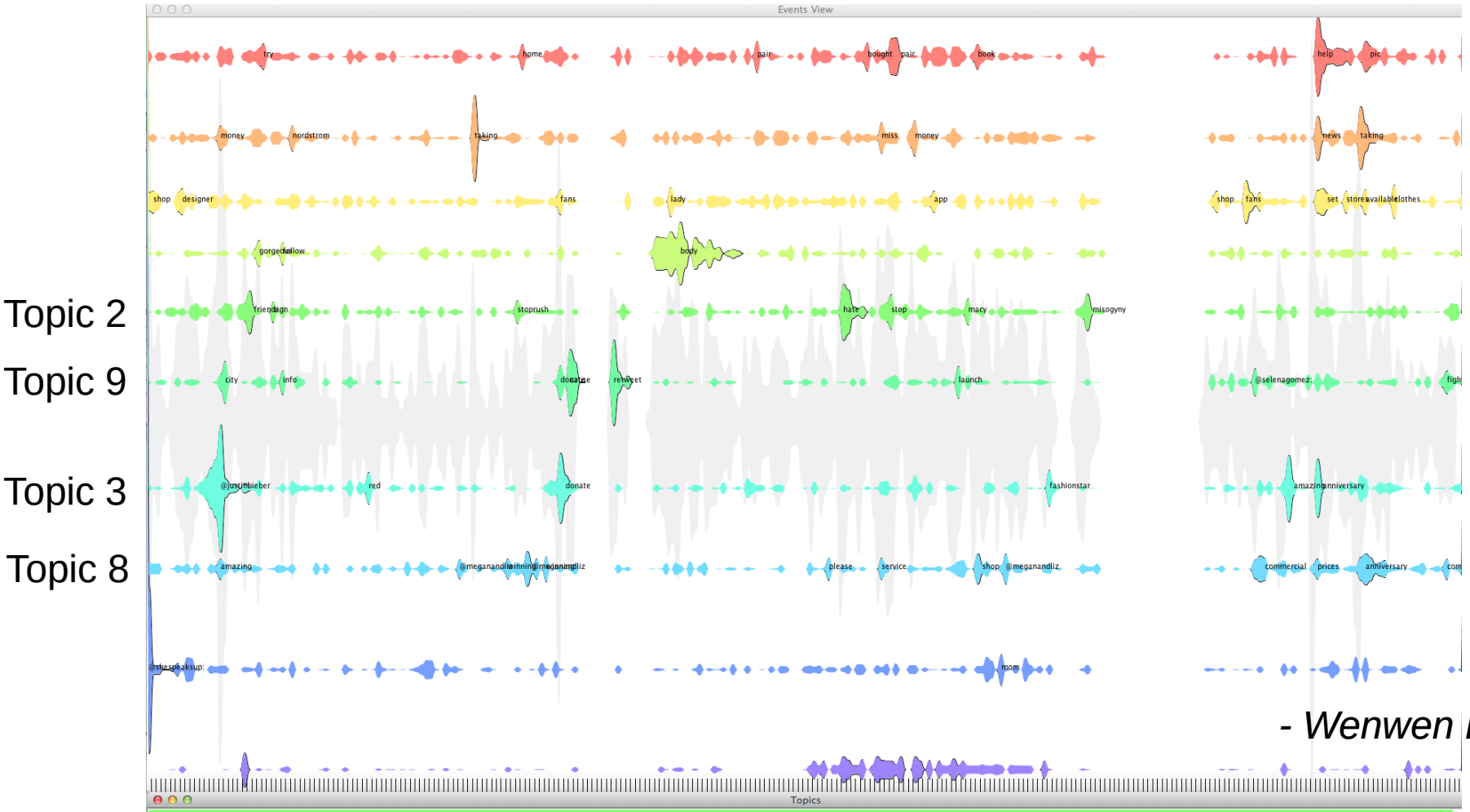


- We are developing and will soon launch a Tools + Data Website via www.viscenter.uncc.edu
- This site will contain downloads of the above tools plus others with descriptions of their capabilities and how to use them.
- Some publicly available data will also be available.
- Downloads will be by permission.

Example from another project

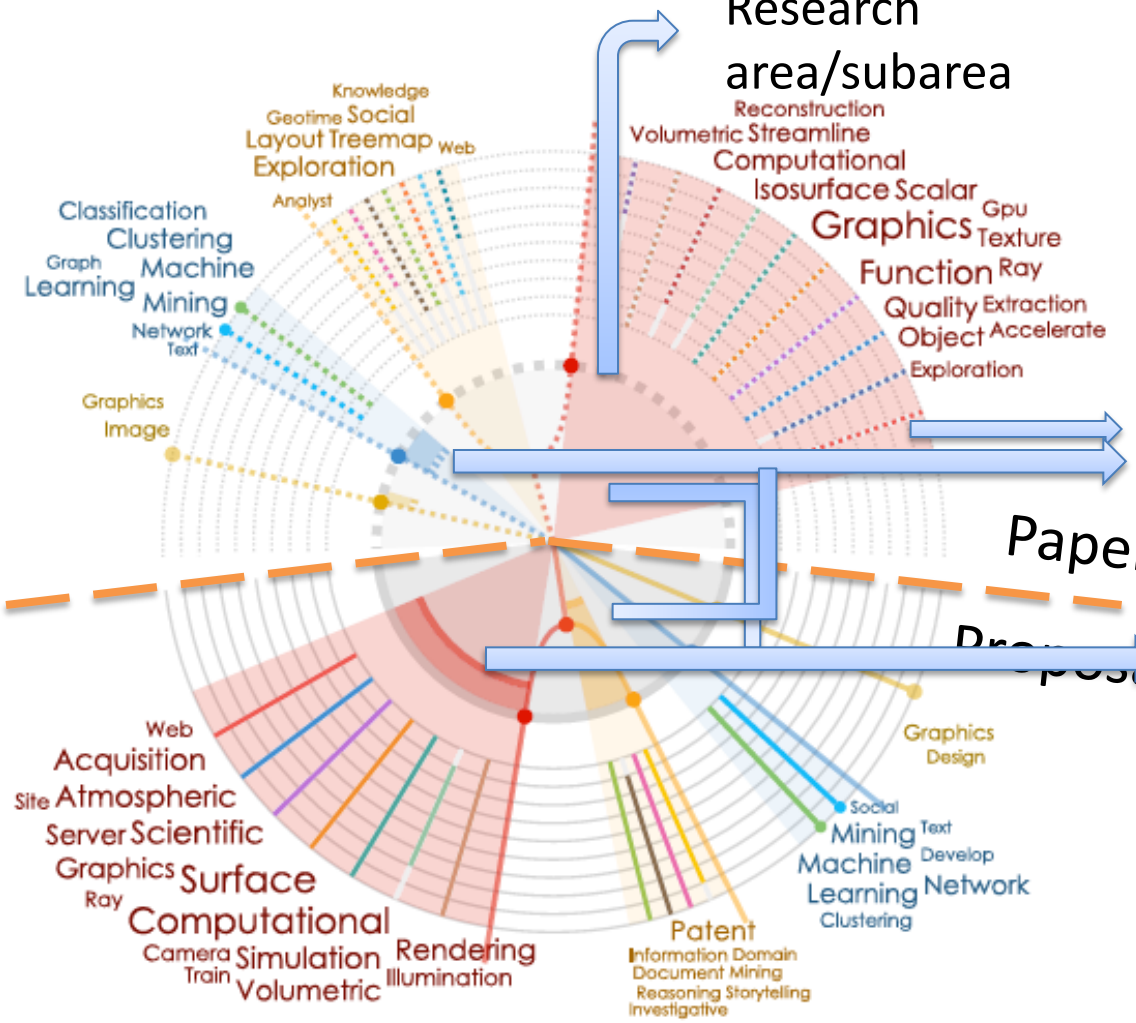


Know Your Customers (Know Their Customers)



- Wenwen Dou

Lead-Lag
NSF vs. Papers: Overall leads



Paper leads NSF proposal in the area of Research topic data mining

NSF proposal leads paper in the subarea of scientific visualization

-Jing Yang and Yang Chen