SAND2008-4804P

Flexible Data Analysis and Visualization with Titan

2008 Sandia Workshop on Data Mining and Data Analysis

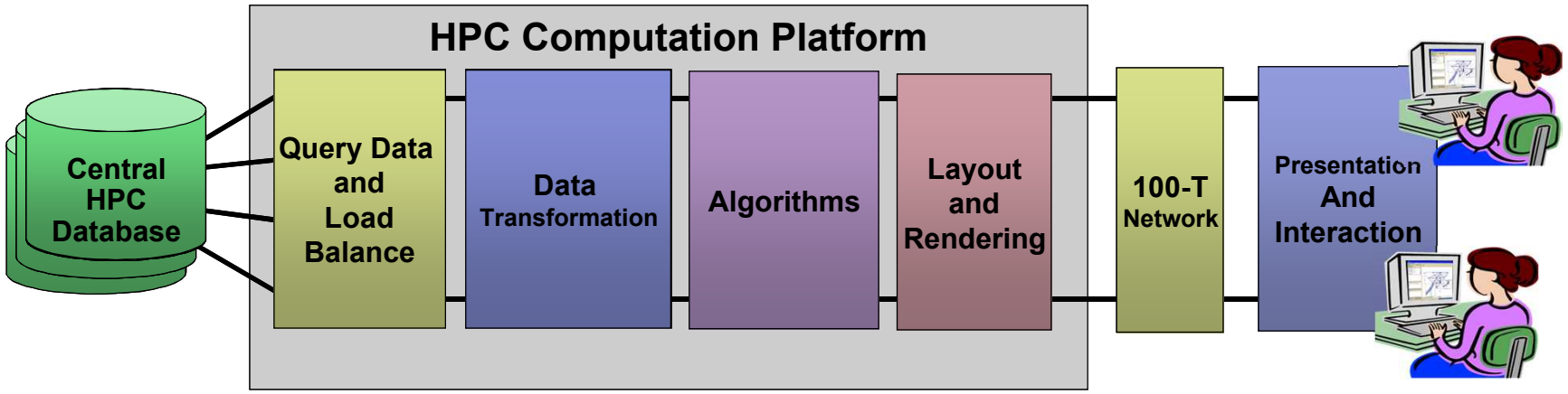
July 22, 2008

Timothy M. Shead (1424)

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Background

Assertion: real-world informatics problems are tough …

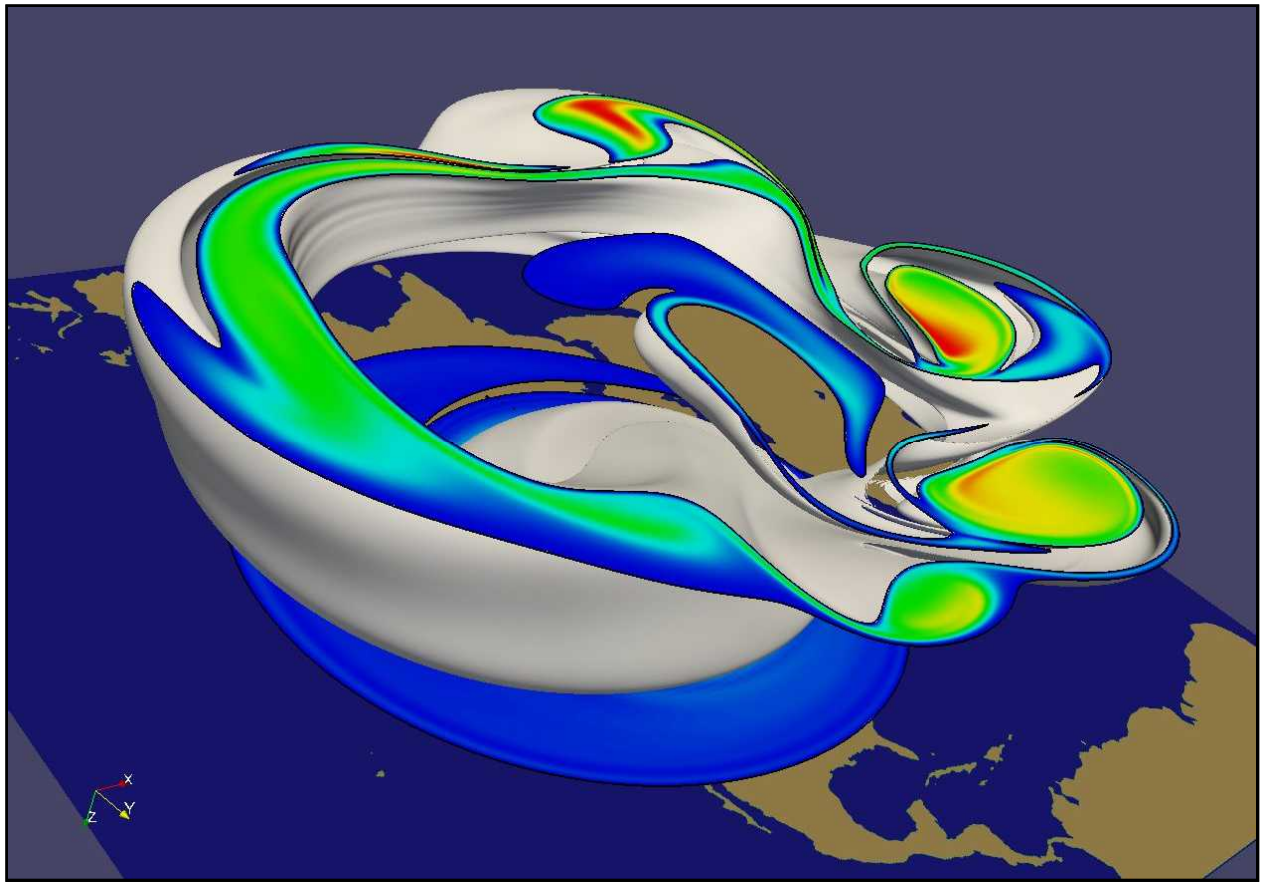


- ***Centralized terabyte databases***
- ***Millions of entities / billions of relationships***
- ***HPC servers to process data***
- ***Delivery to the desktop (client / server) is a "must-have"***
- ***Require reasonable response times (< 10 seconds)***
- ***Require reasonable frame rates (> 10 frames/sec)***
- ***Require task-specific applications***
- ***Don't reinvent the wheel***

Sandia National Laboratories

Polar Vortex Breakdown

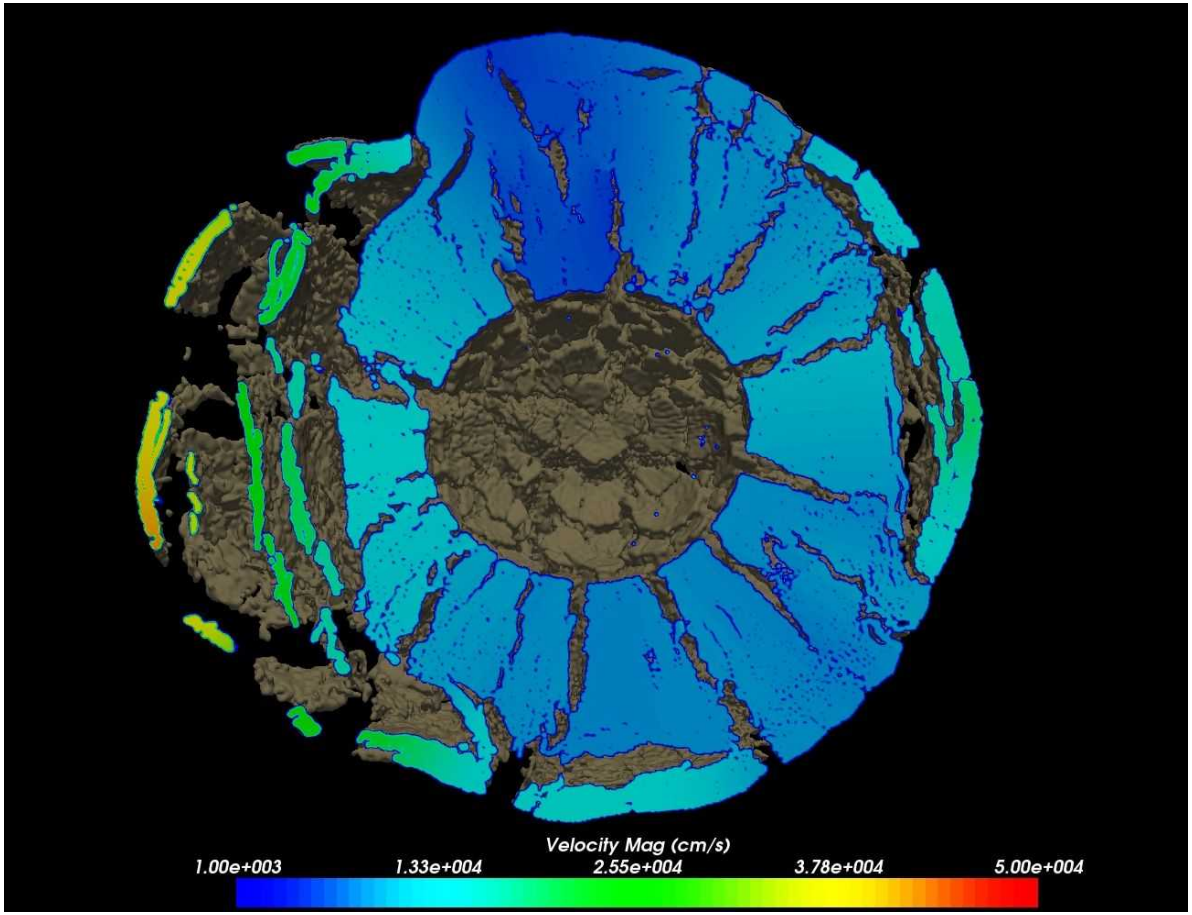
SEAM Climate Modeling, 1 terabyte on-disk, 1 billion cells (500 million cells visualized)



Sandia National Laboratories

Golevka Asteroid vs. 10 Megaton Explosion

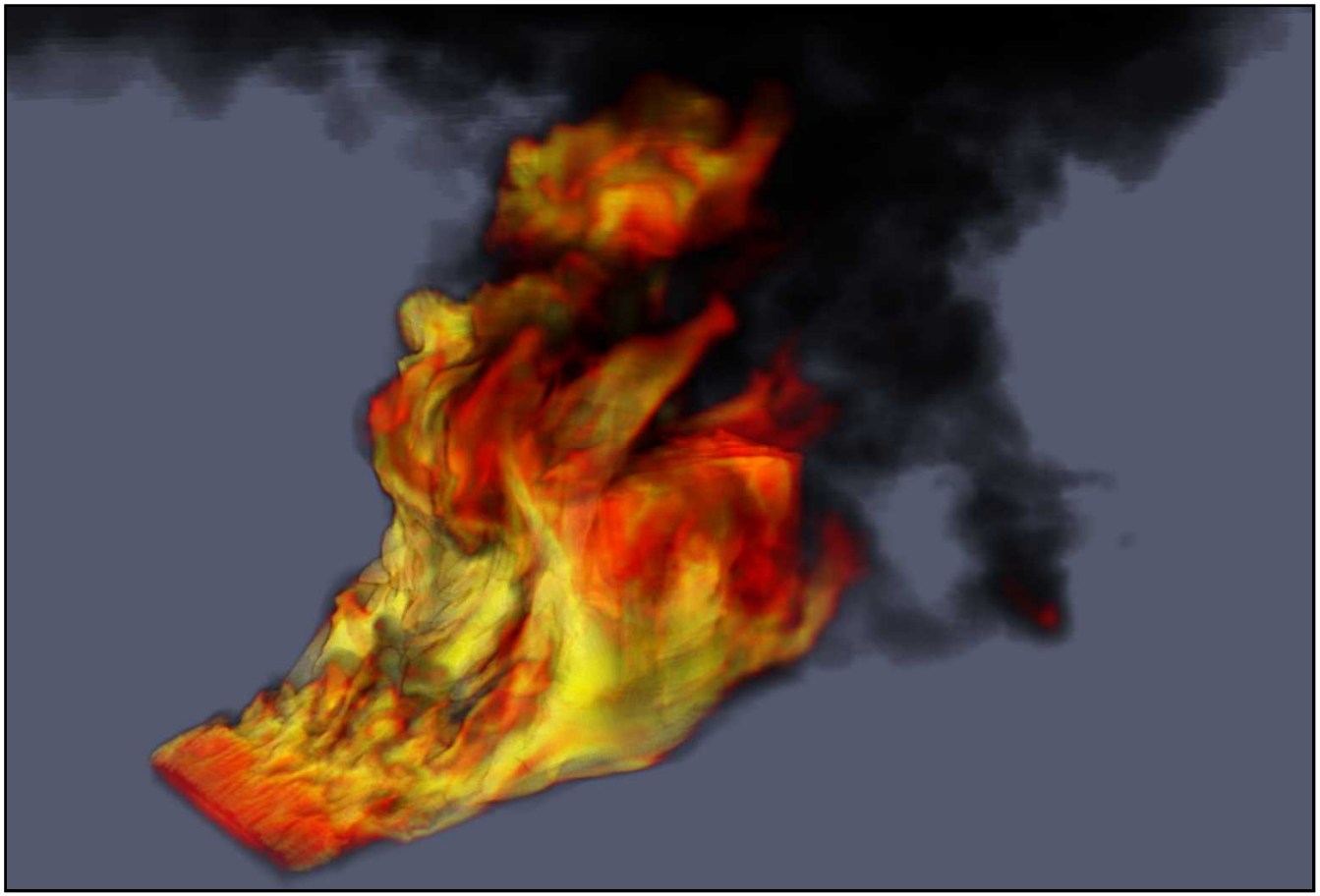
CTH shock physics, over 1 billion cells



Sandia National Laboratories

Objects-in-Crosswind Fire

Coupled SIERRA/Fuego/Syrinx/Calore, 10 million unstructured hexahedra



Sandia National Laboratories

The Titan Toolkit

Data Structures
- Table
- Tree
- DAG
- Directed Graph
- Undirected Graph
- Sparse / Dense n-way Array

Database Drivers
- MySQL
- Postgres
- Oracle
- SQLite
- ODBC / Netezza

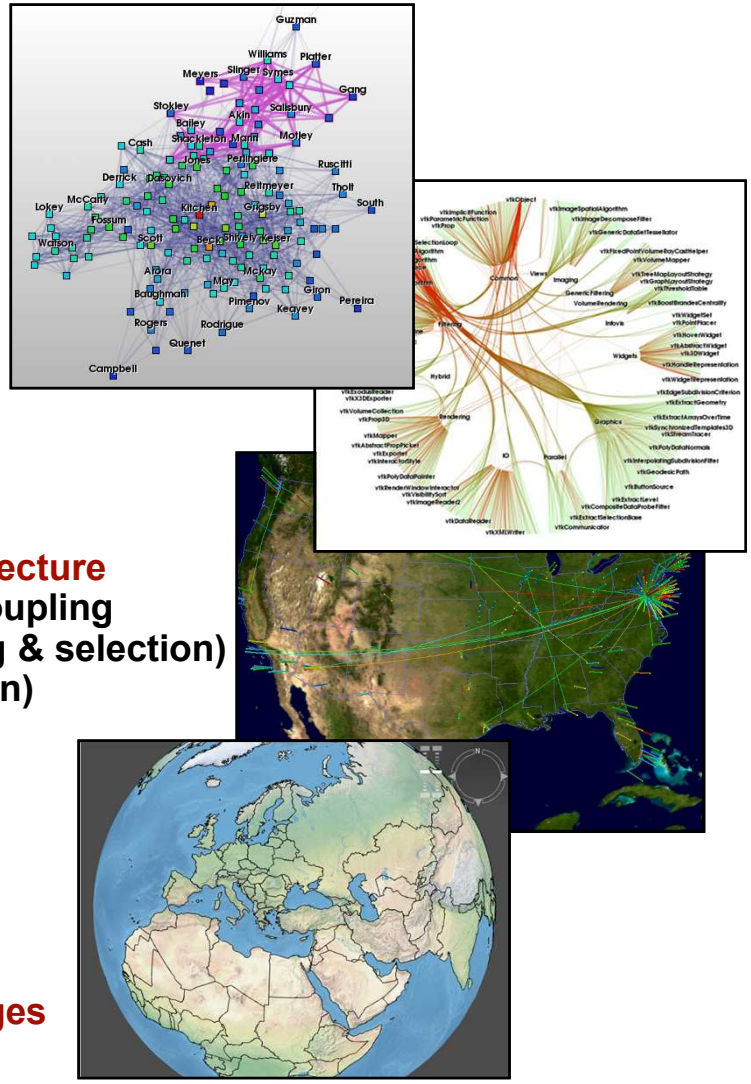
Readers
- Dimacs
- DOT
- GXL
- Chaco
- XML
- Tulip
- DelimitedText
- FixedWidth
- ISI, RIS

MATLAB Integration

BGL Graph Algorithms
- Breadth First Search
- Connected Components
- Biconnected Components
- Brandes Centrality

Linear-Time Graph Layouts
- GSpace
- Hierarchical
- Clustered
- Three tree-based variants

Representation / View architecture
- Supports scivis / infovis coupling
- Representations (rendering & selection)
- Views (window & interaction)
- Render View (scivis)
- Graph View
- Tree Layout View
- Tree Map View
- Hierarchical Graph View
- Geographic View

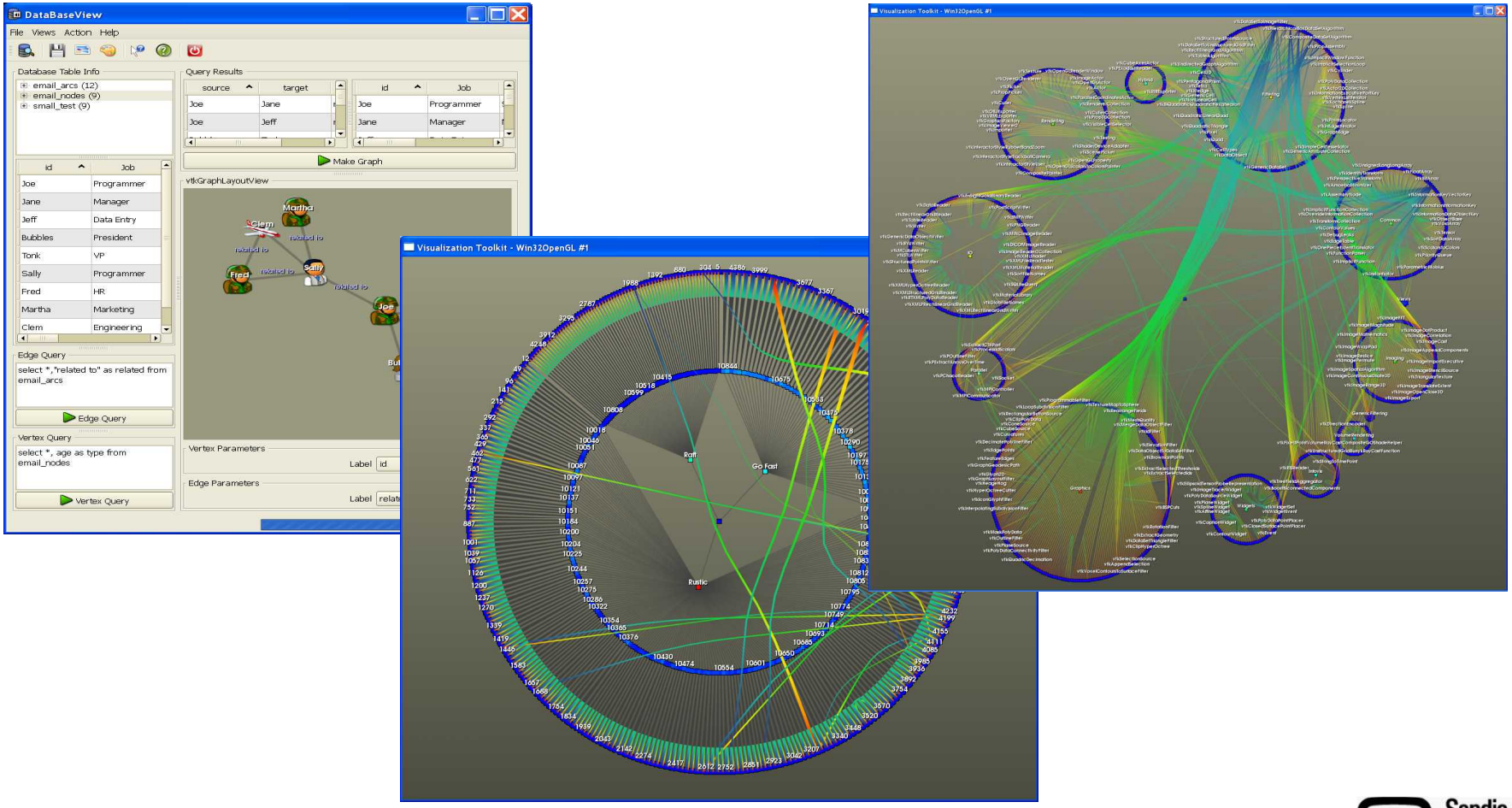


Multiple Platforms / Languages
- Windows, Linux, OSX
- Many HPC platforms
- Write Titan components in C++
- Use Titan components in C++, Python, TCL, Java
- Use Titan components as “plugins”

Sandia National Laboratories

Titan Basics

- ***Semantic Graph data structures and representations***
- ***Hierarchical algorithms, representations, and interactions***



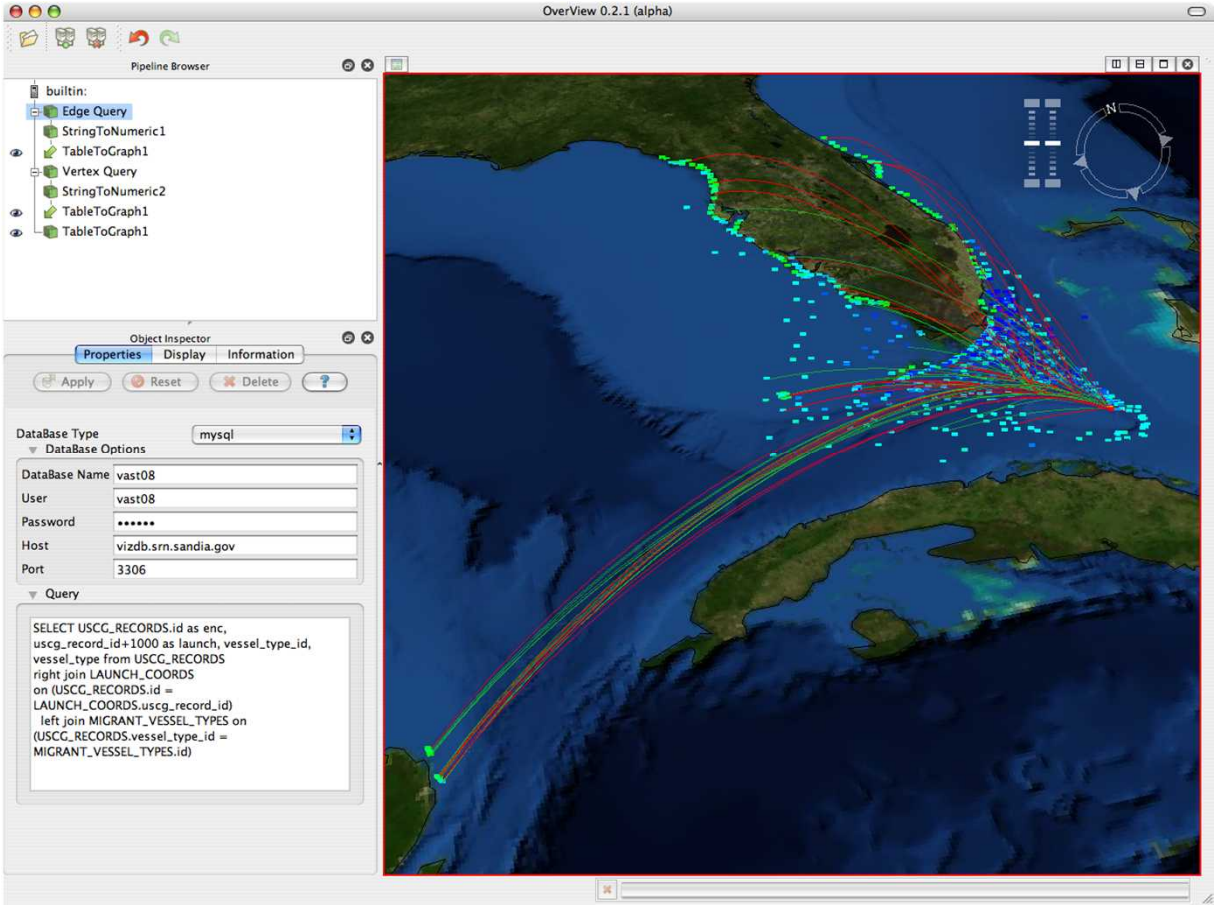
Sandia National Laboratories

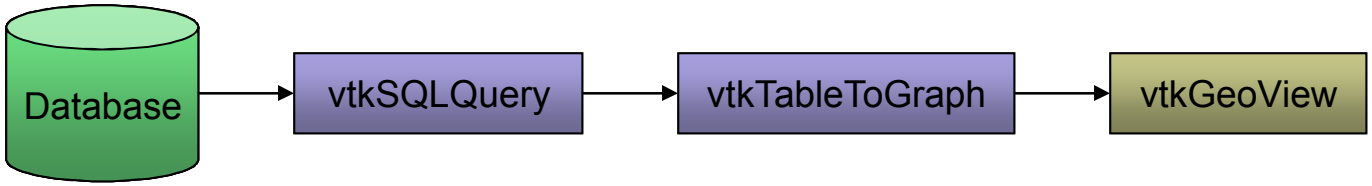
Georeferenced Semantic Graphs

David Thompson (8963), Kitware

Sandia National Laboratories

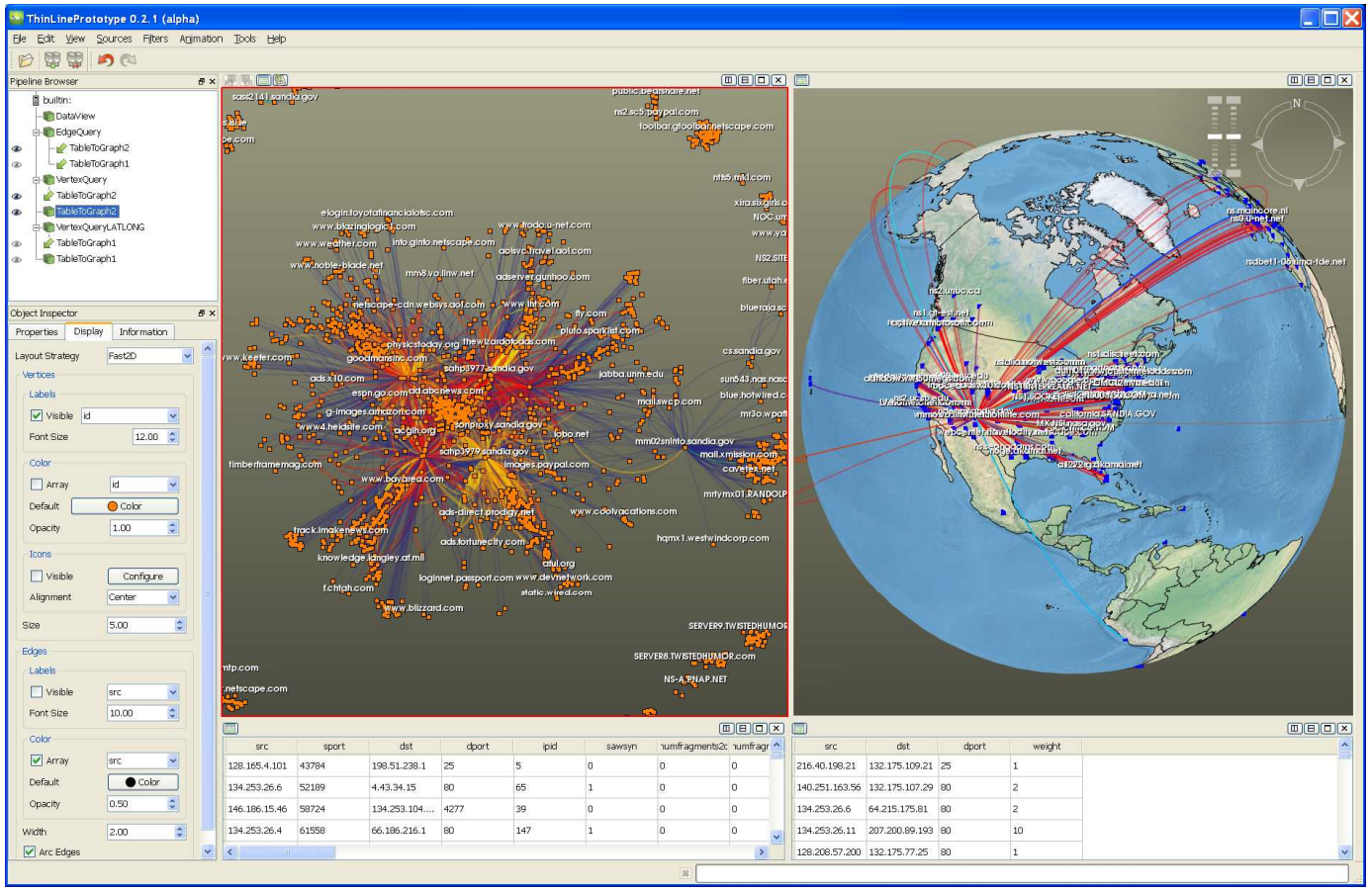
Simulated Coast Guard Interceptions

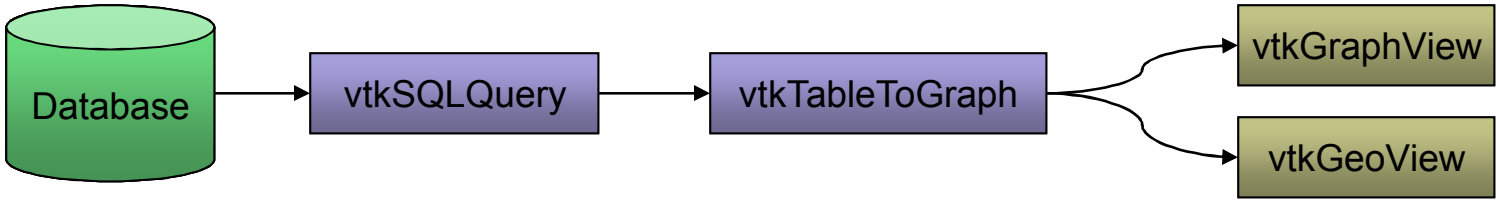




Sandia National Laboratories

Network Packets





Sandia National Laboratories

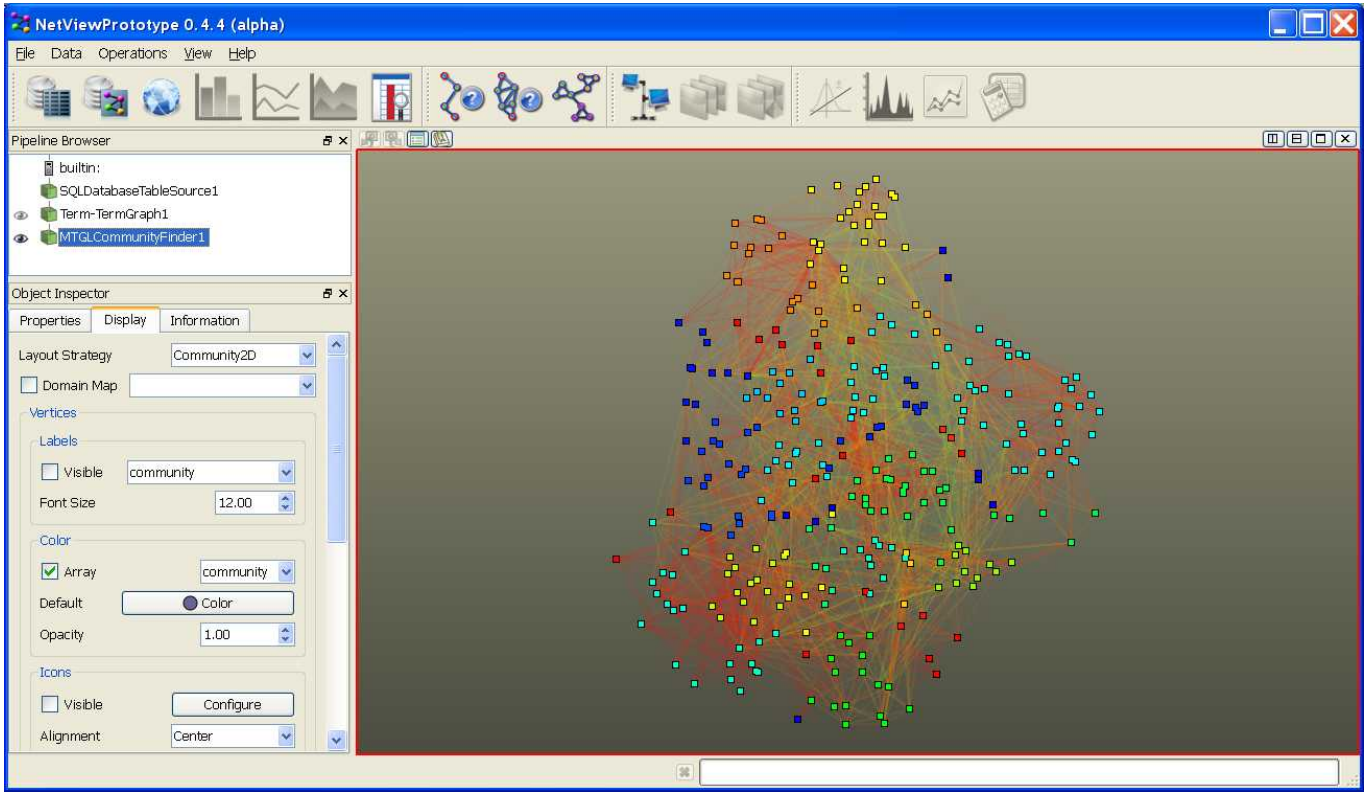
Multithreaded Graph Analysis - MTGL

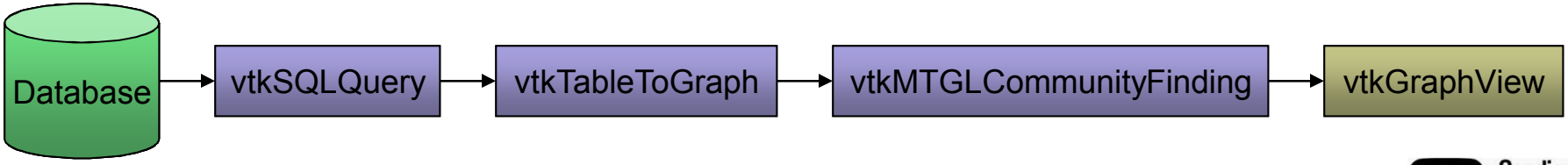
Jon Berry (1416), Vitus Leung (1415), Will McLendon (1423)

Sandia National Laboratories

MTGL Community Finding

Identifies communities within the graph based on their topological relationships. ***Solves a sequence of linear programs using a Lagrangian relaxation process using linear time and space.***

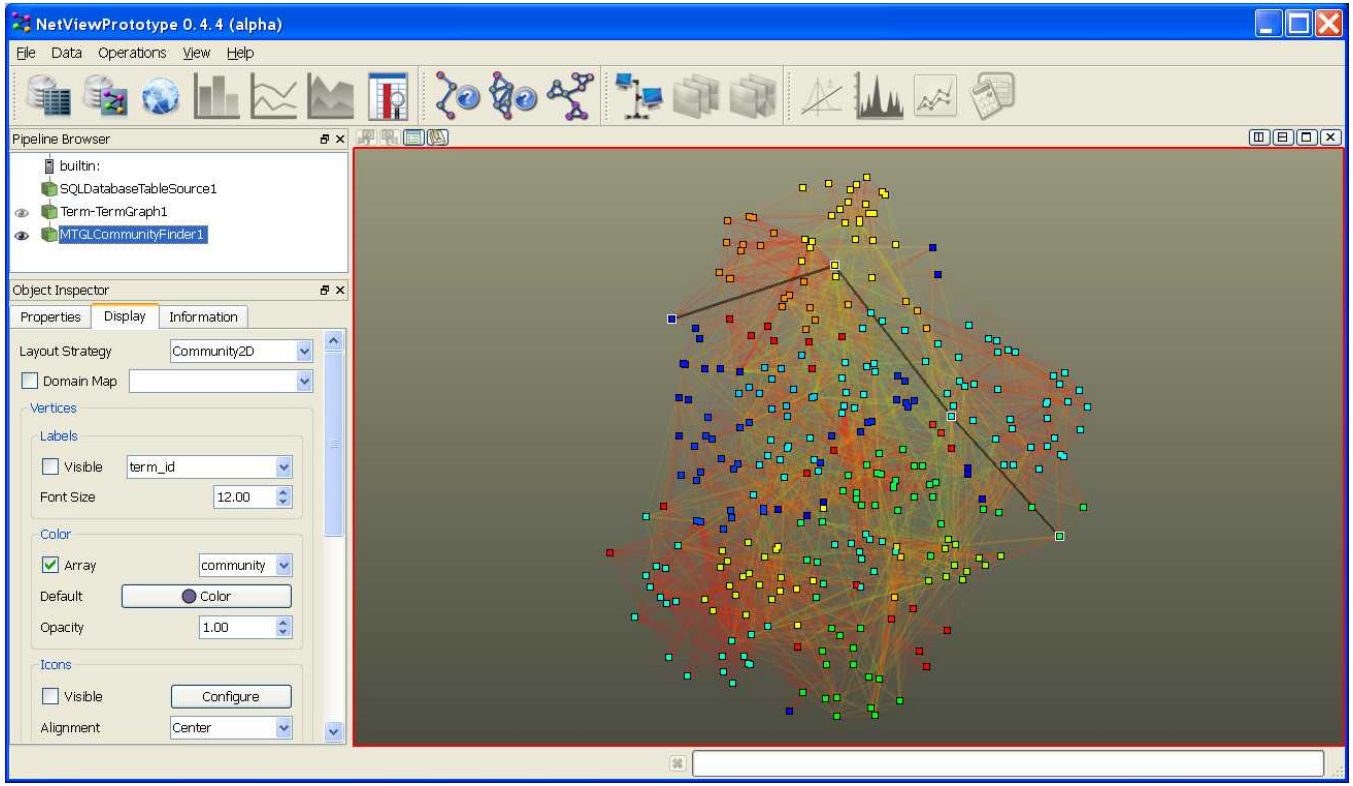


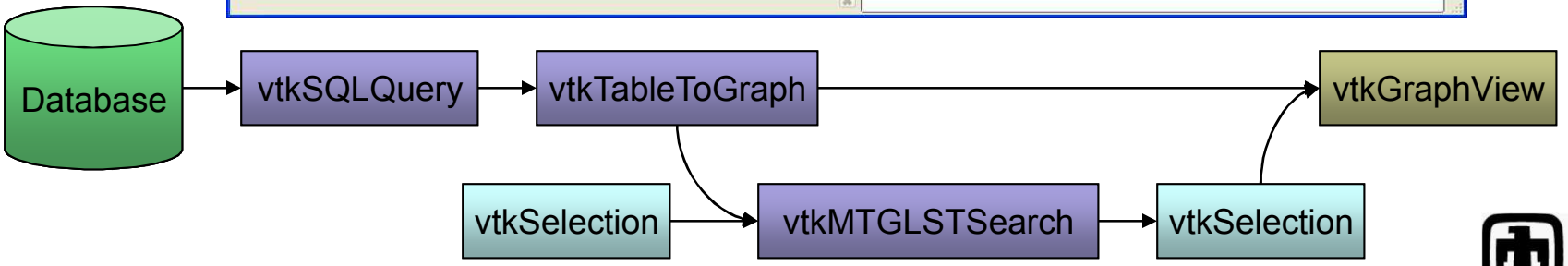


Sandia National Laboratories

MTGL ST Search

Finds all shortest paths between two nodes in a graph. Utilizes a breadth-first search from both S and T.

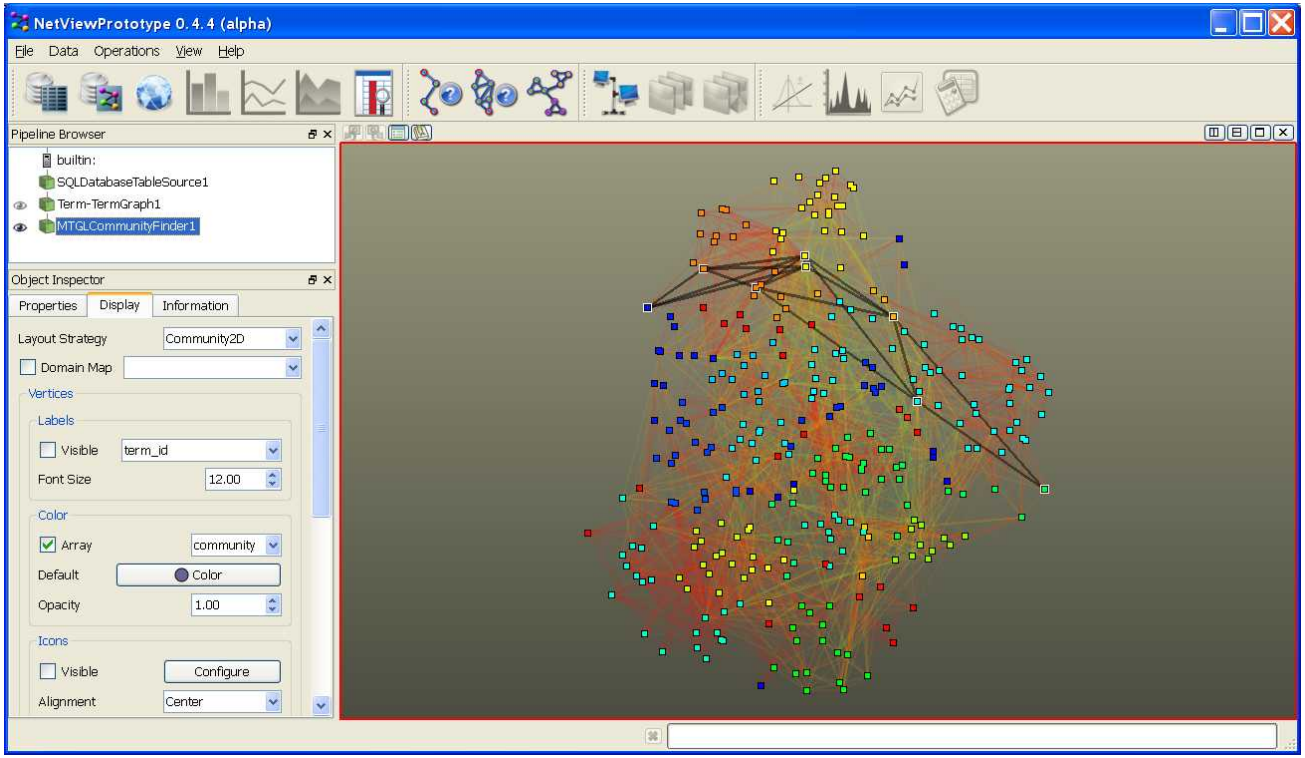


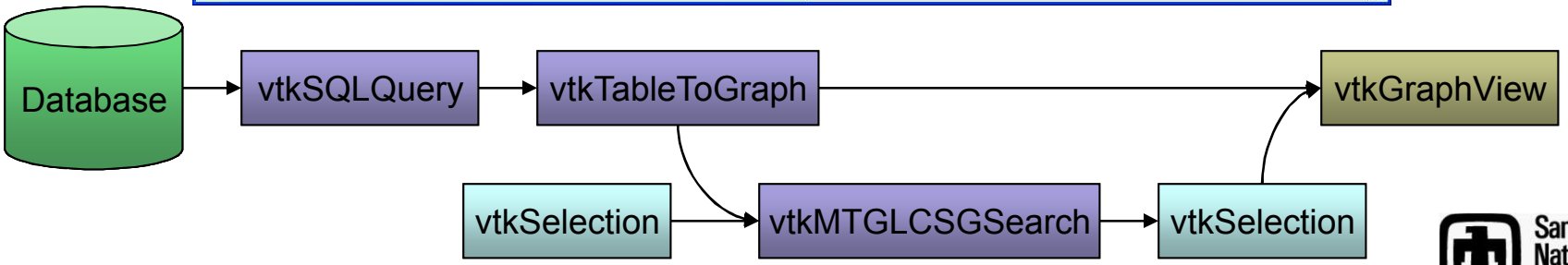


Sandia National Laboratories

MTGL Connection Sub-Graph

Finds a sub-graph between two nodes in a graph, avoiding highly connected vertices and long paths. Models the graph as an electrical circuit, extracting paths which maximize current between S and T.

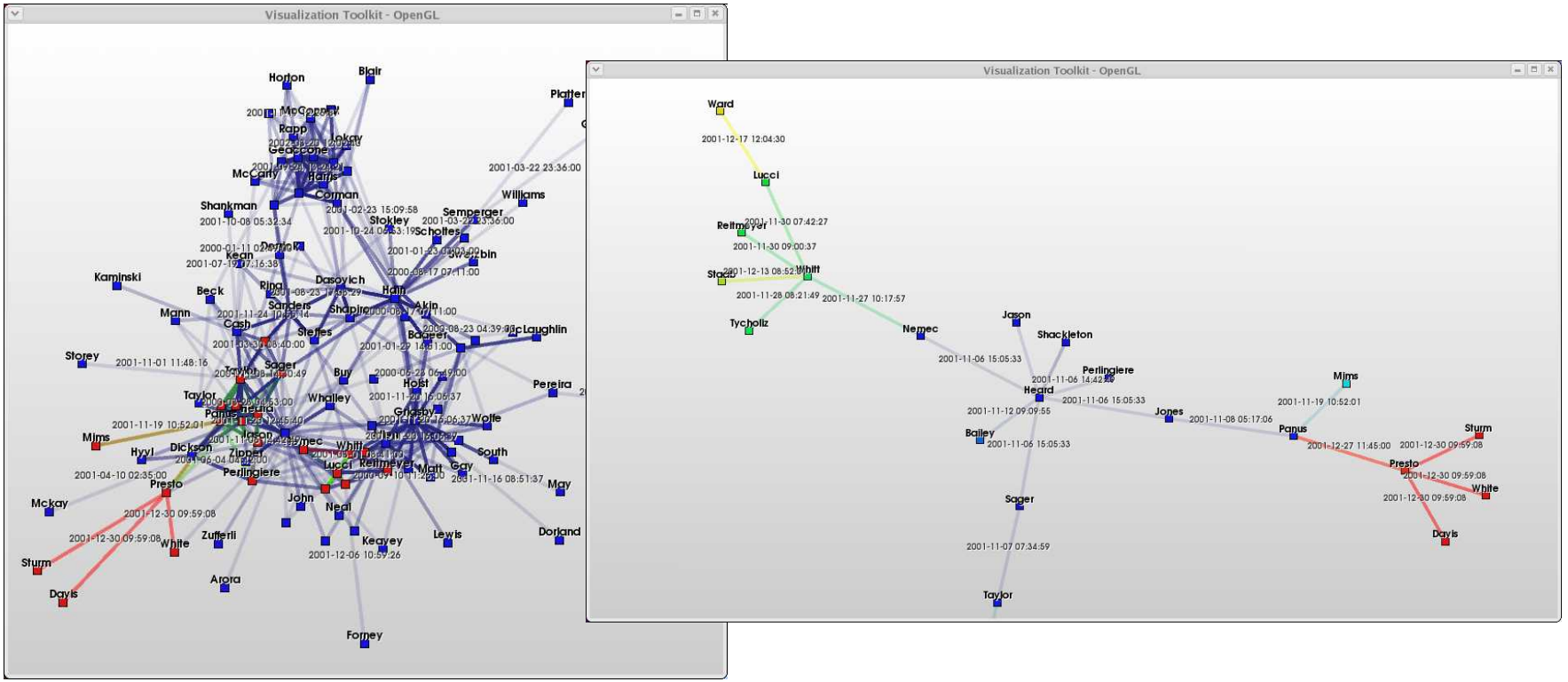


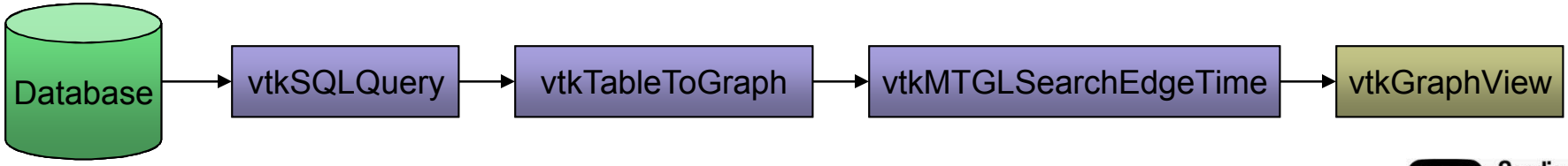


Sandia National Laboratories

MTGL Temporal Graph Search

Performs a "filtered" search from a starting event, only following paths that occurred prior-to or after the original event.

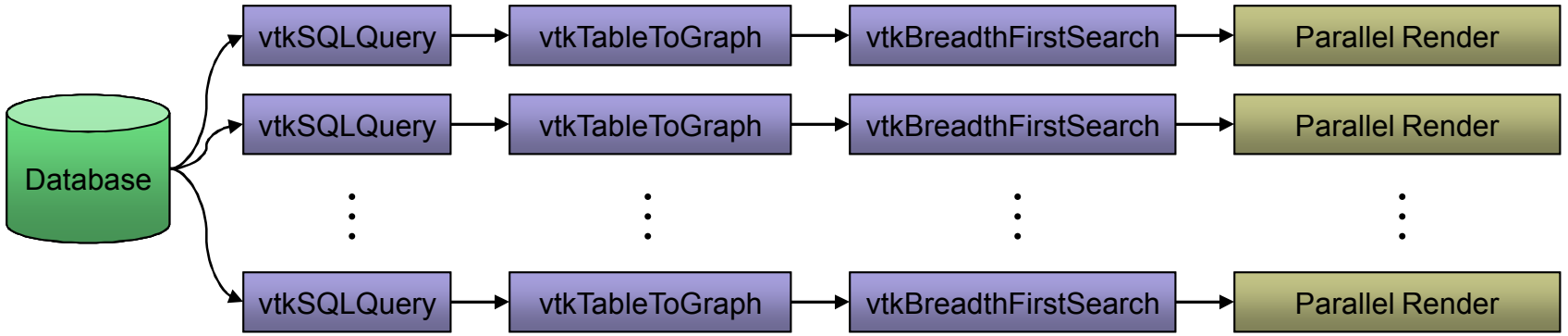




Sandia National Laboratories

Parallel Graph Analysis - PBGL

PBGL: Parallel Boost Graph Library – http://www.osl.iu.edu/research/pbgl
Andrew Lumsdaine, Douglas Gregor (Indiana University)



Current milestone: running a BFS on a random graph containing 50M vertices and 500M edges on 80 nodes.

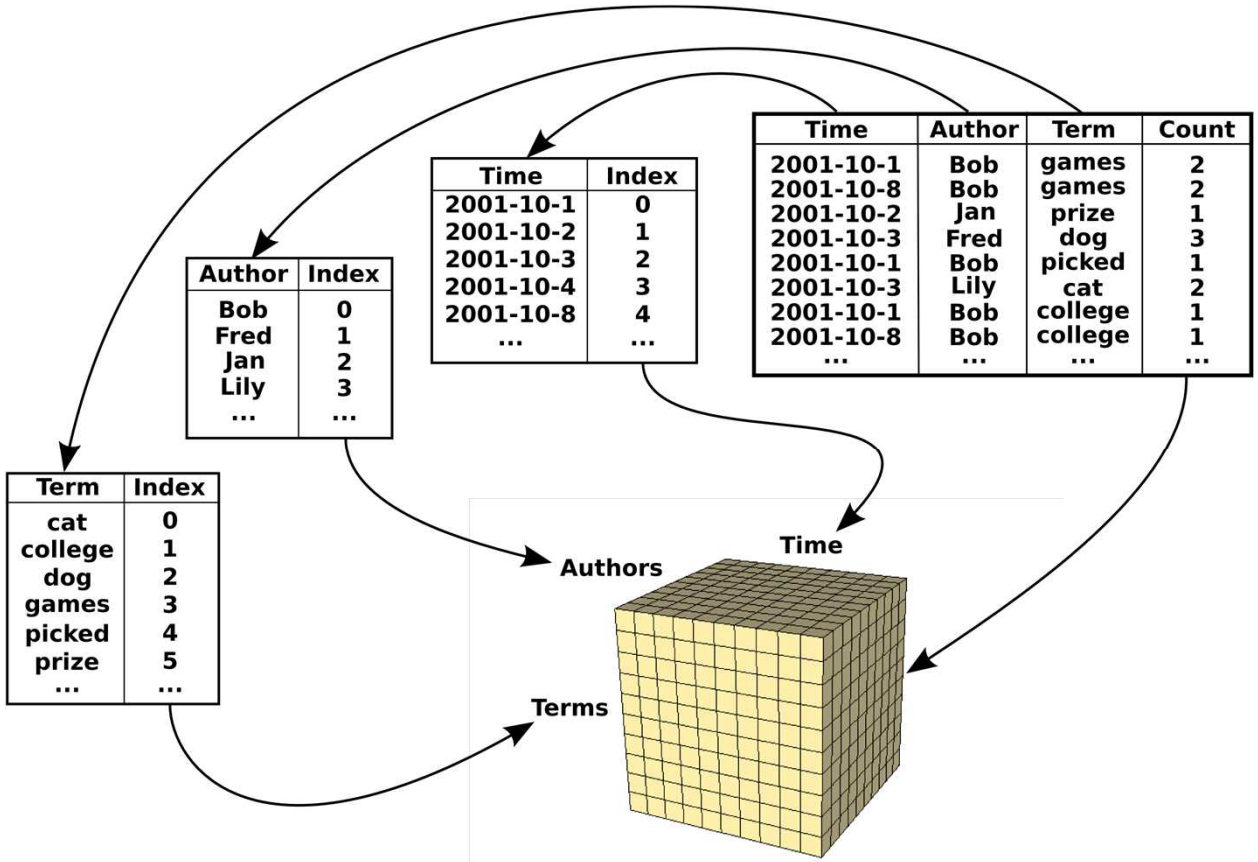
Sandia National Laboratories

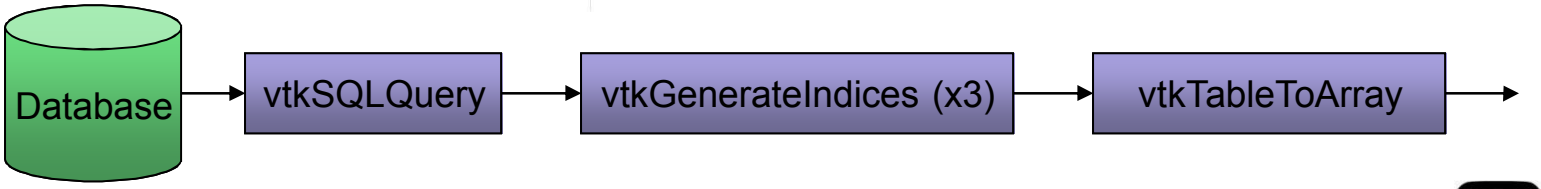
Multidimensional Analysis

Brett Bader (1415), Tammy Kolda (8962), Mark Sears (1435)

Sandia National Laboratories

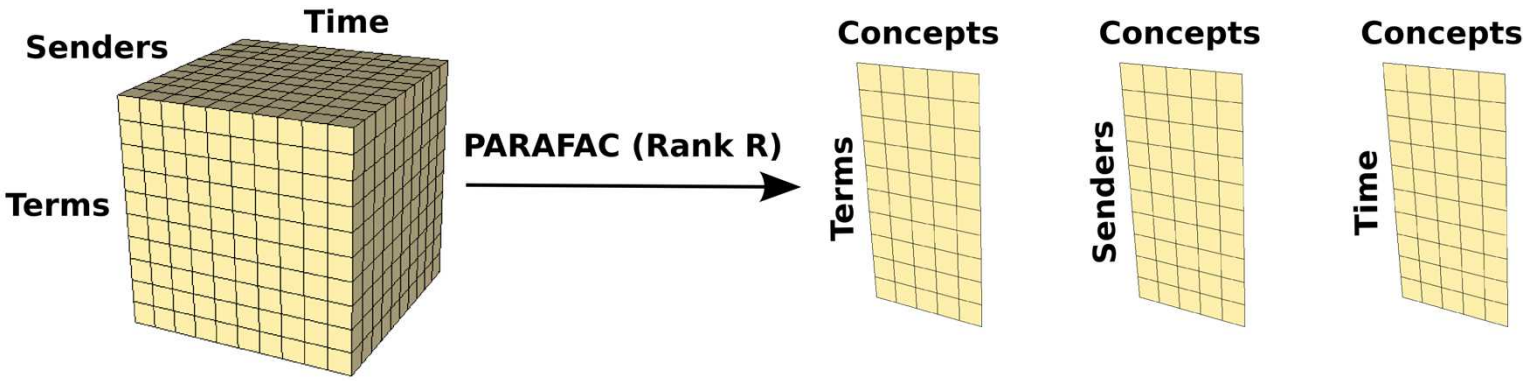
Tensor Creation

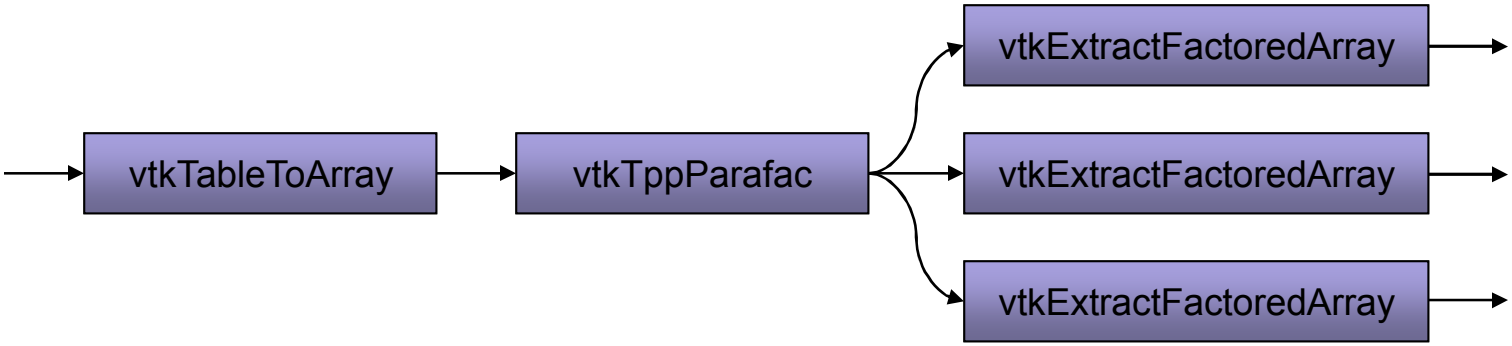




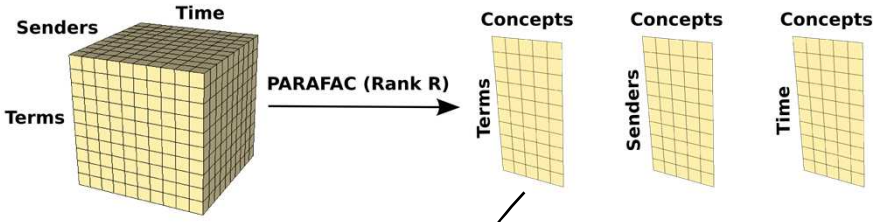
Sandia National Laboratories

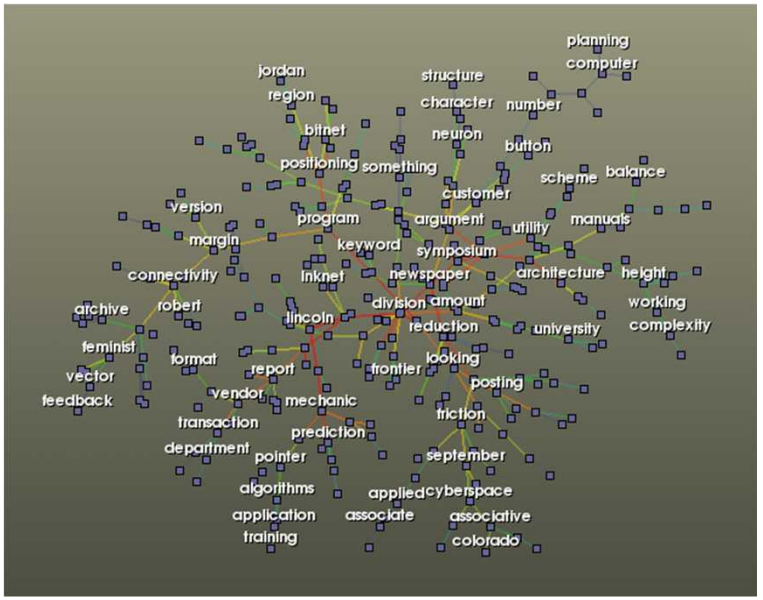
Tensor Reduction

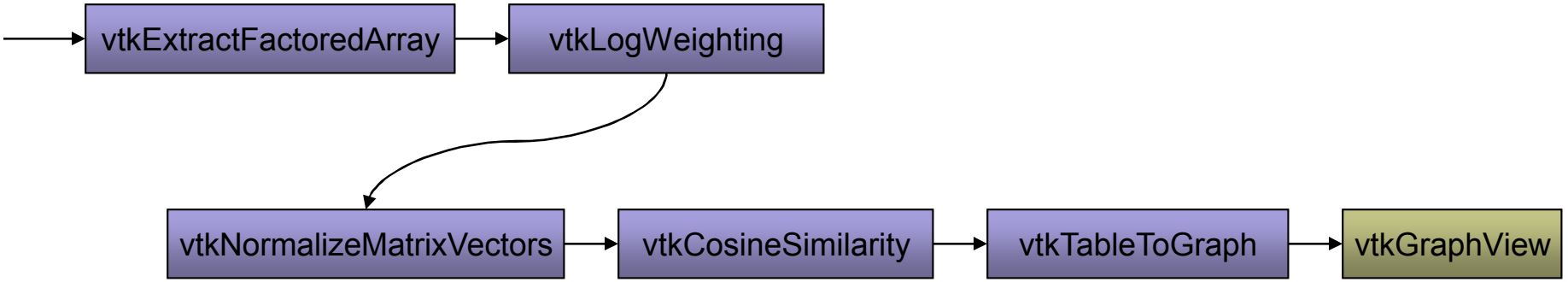




Term-term Similarity Graphs

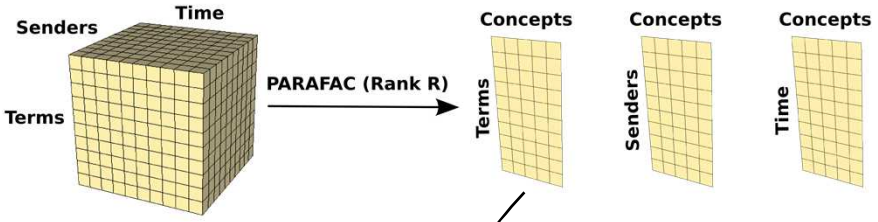


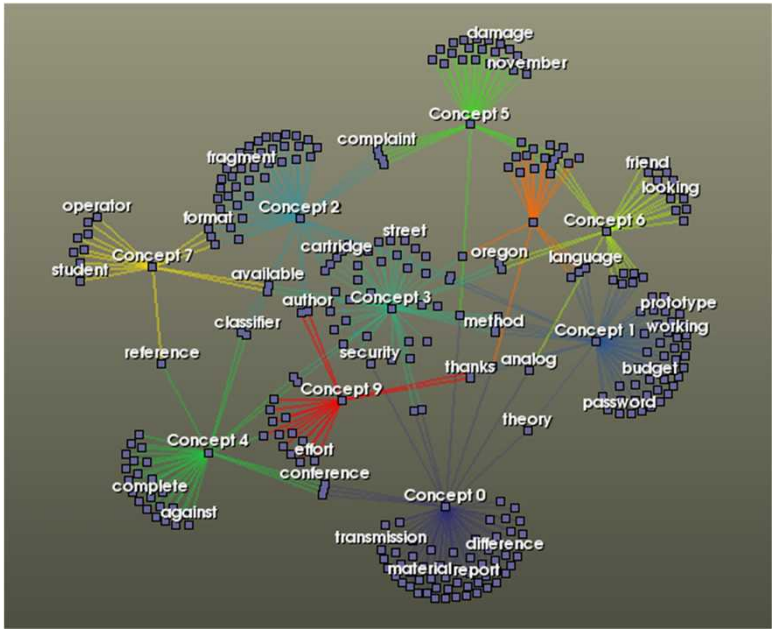




Sandia National Laboratories

Concept-term Similarity Graphs

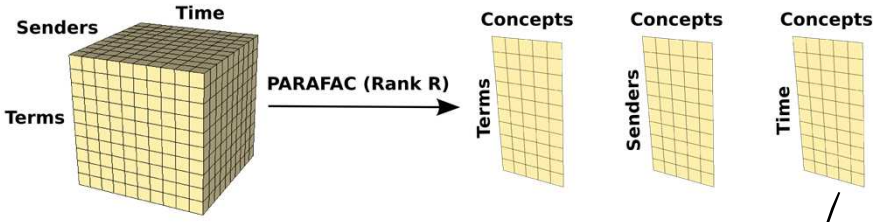


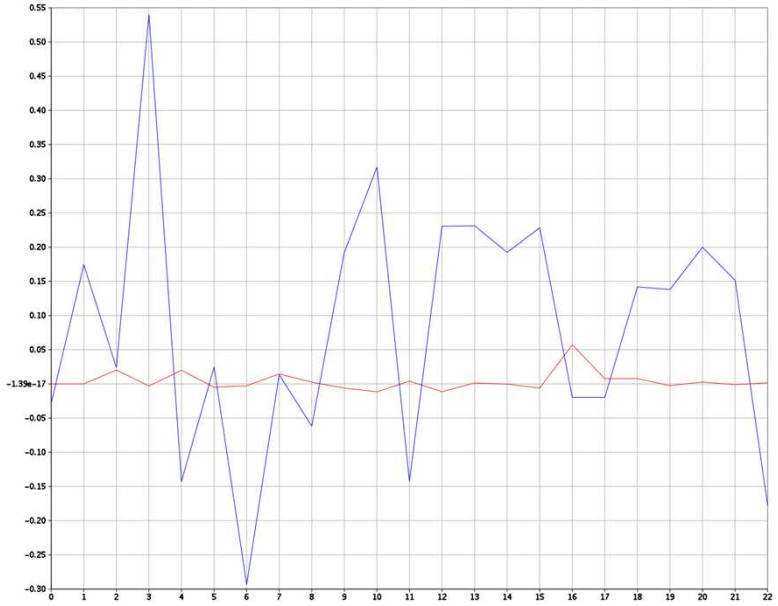


vtkExtractFactoredArray → vtkAdjacencyMatrixToEdgeTable → vtkTableToGraph → vtkGraphView

Sandia National Laboratories

Concept-Activity Over Time

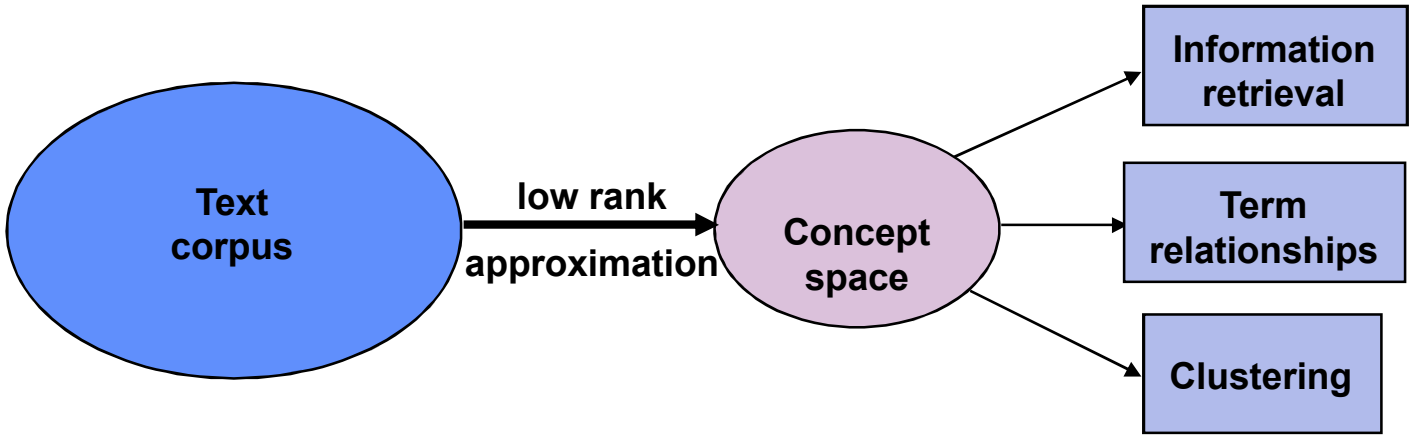


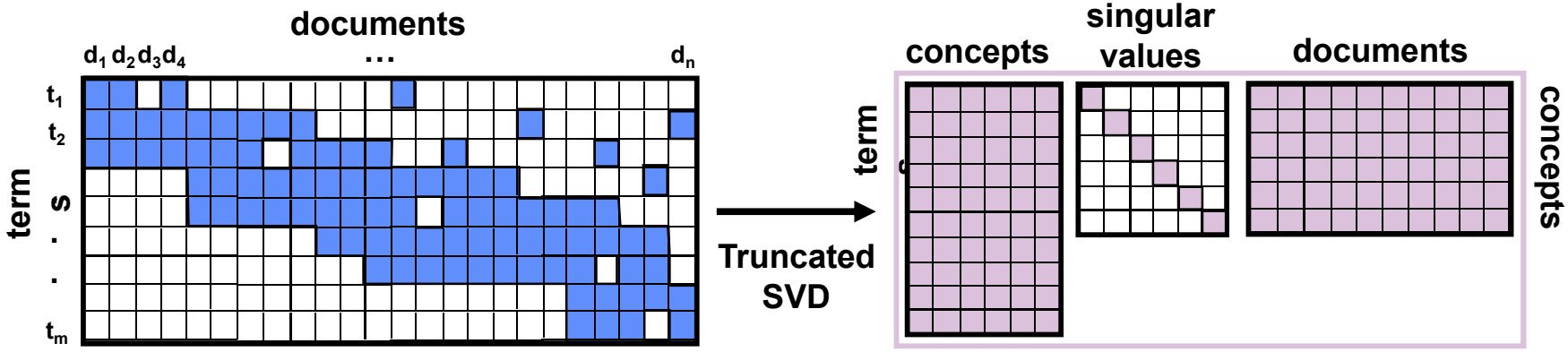


Sandia National Laboratories

Unstructured Text Analysis - ParaText™

Danny Dunlavy (1415), Pat Crossno (1424), Tim Shead (1424)





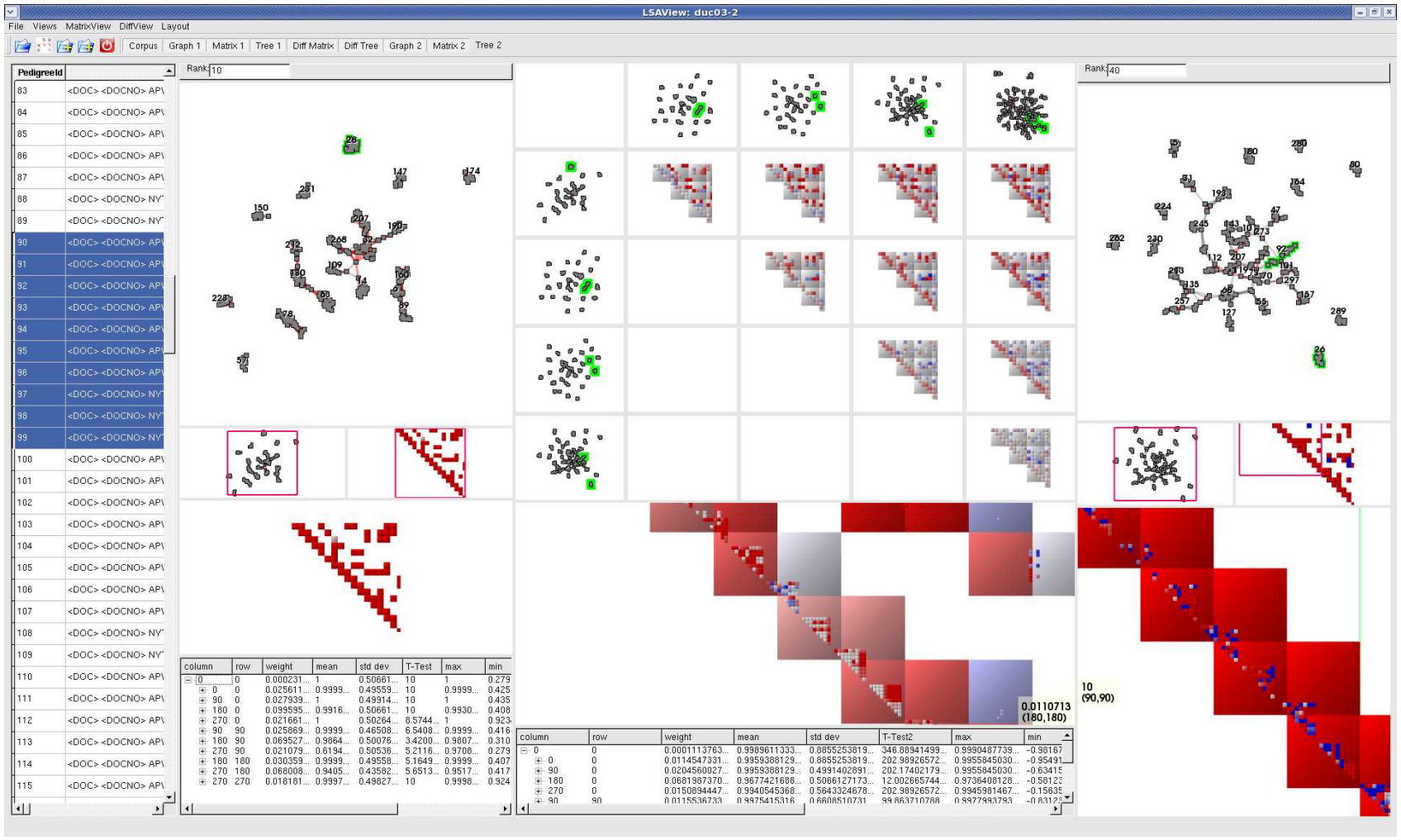
Sandia National Laboratories

Parameterization Analysis - LSAView

Pat Crossno (1424), Danny Dunlavy (1415)

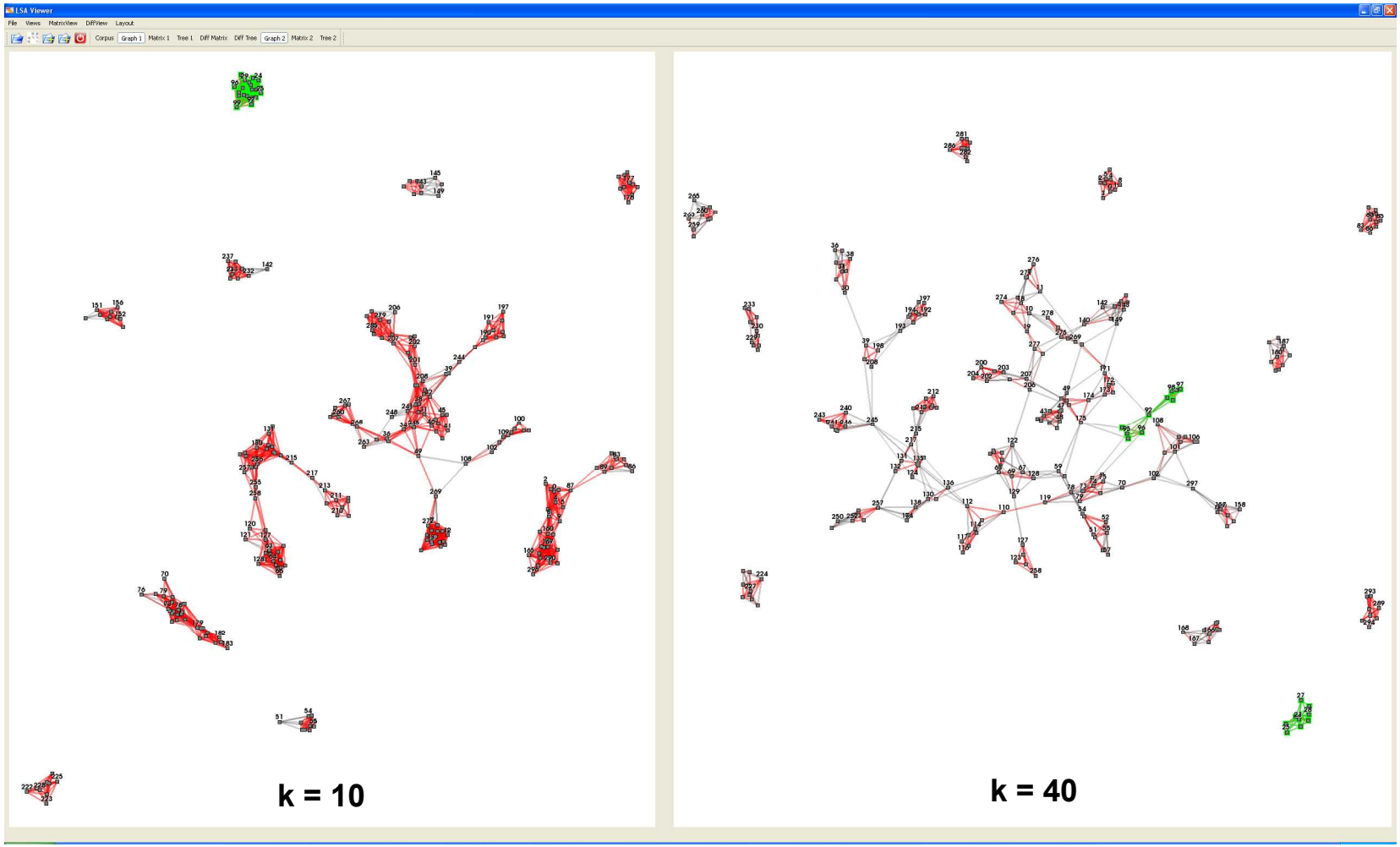
Sandia National Laboratories

Parameterization Analysis - LSAView



Sandia National Laboratories

Parameterization Analysis - LSAView



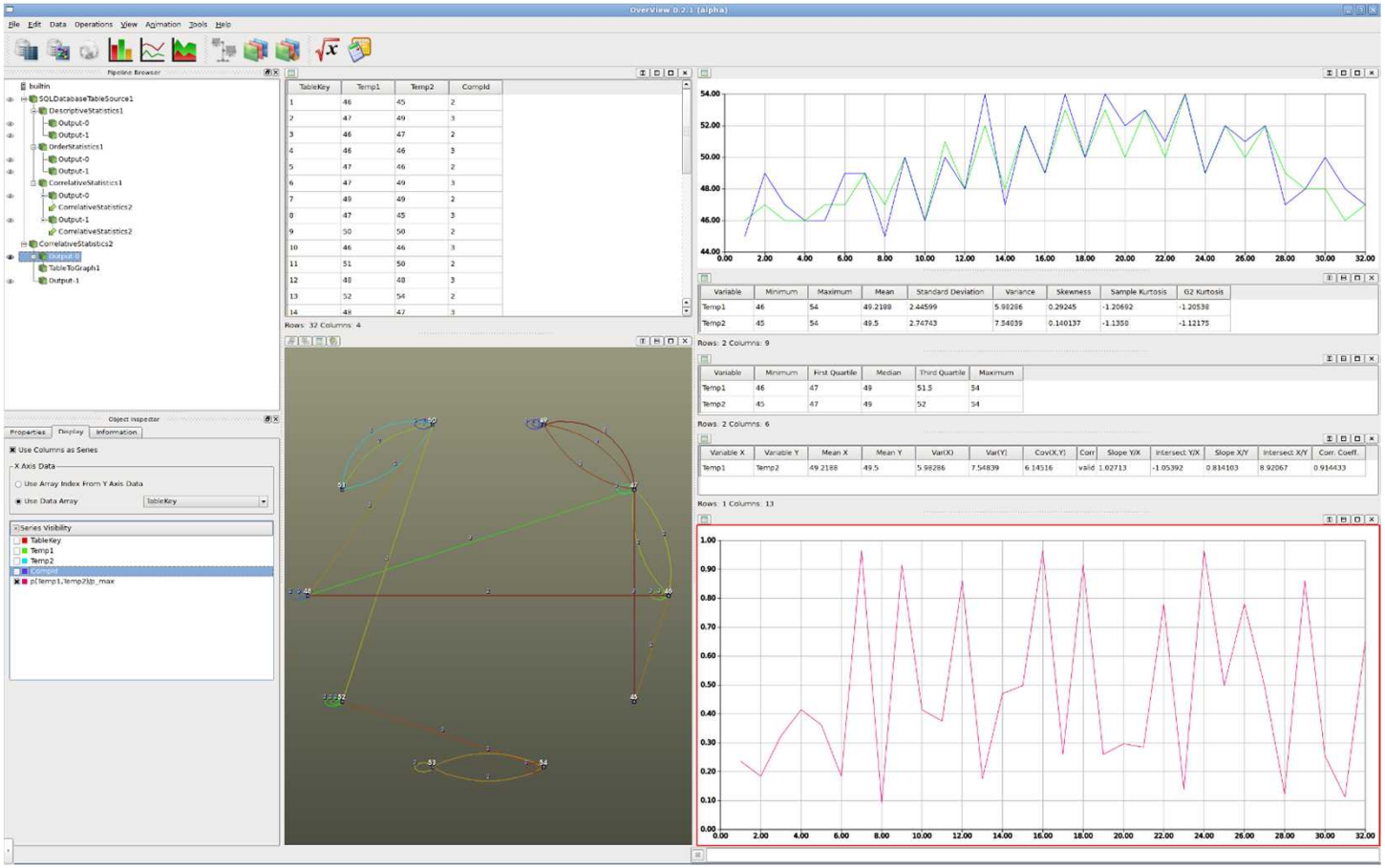
Sandia National Laboratories

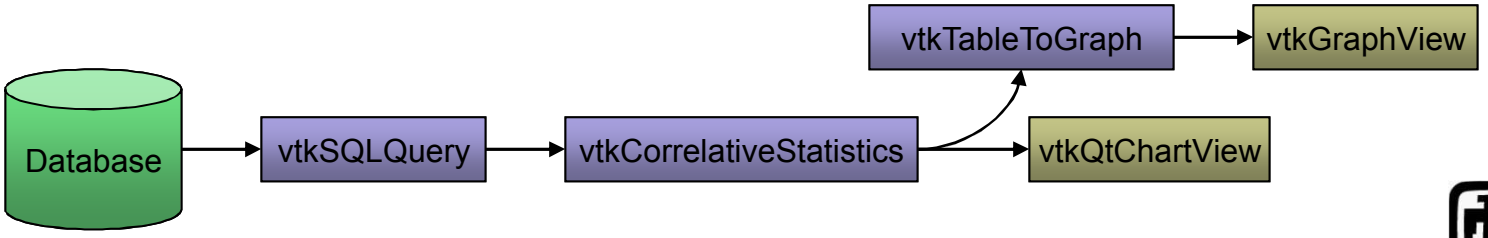
Statistical Analysis

OVIS Team

Sandia National Laboratories

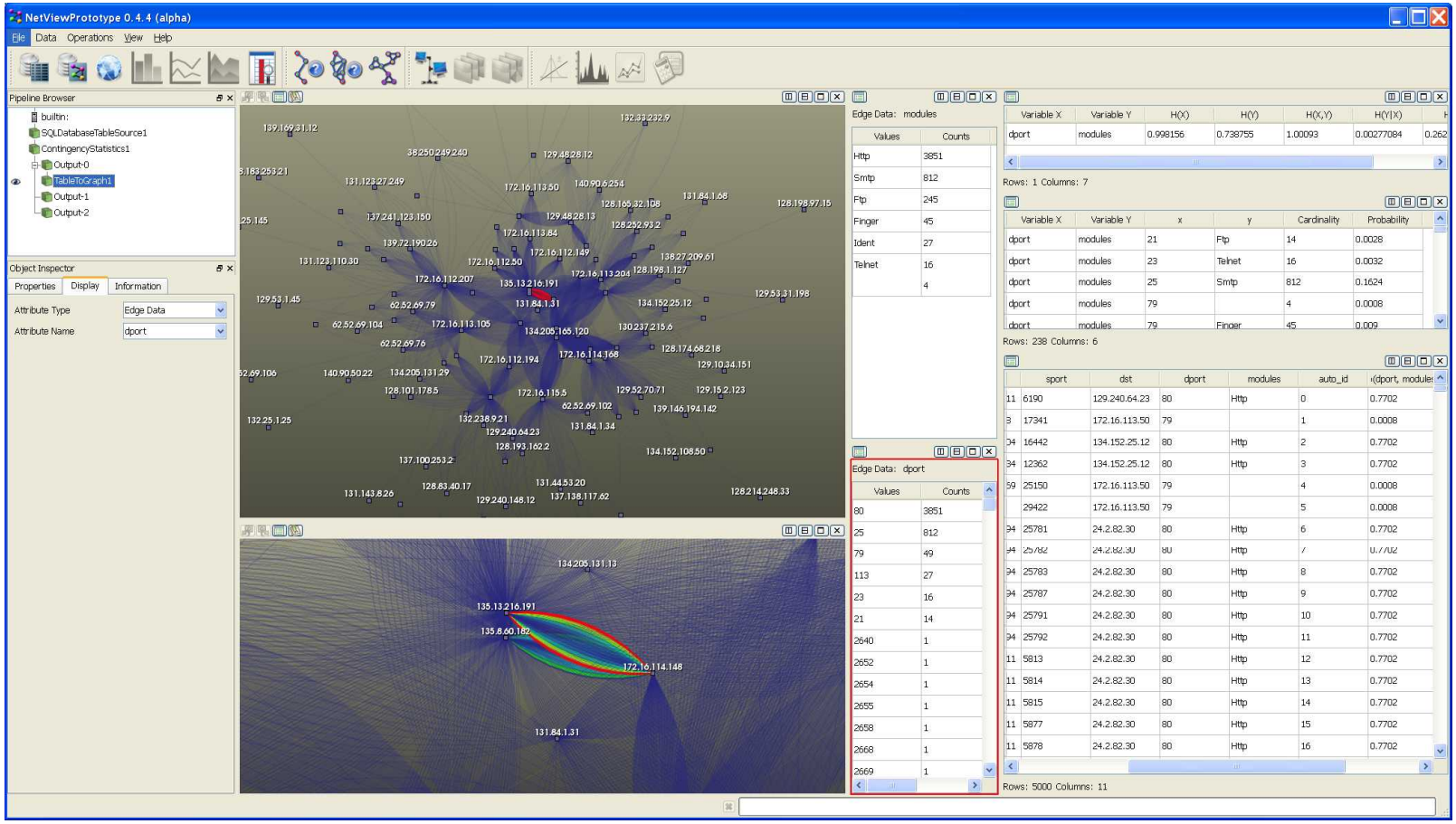
Descriptive, Order and Correlative Statistics

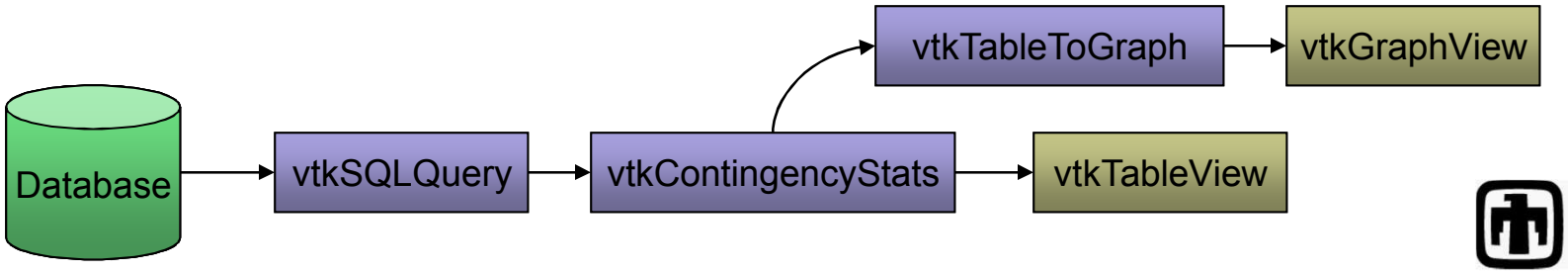




Sandia National Laboratories

Contingency Statistics



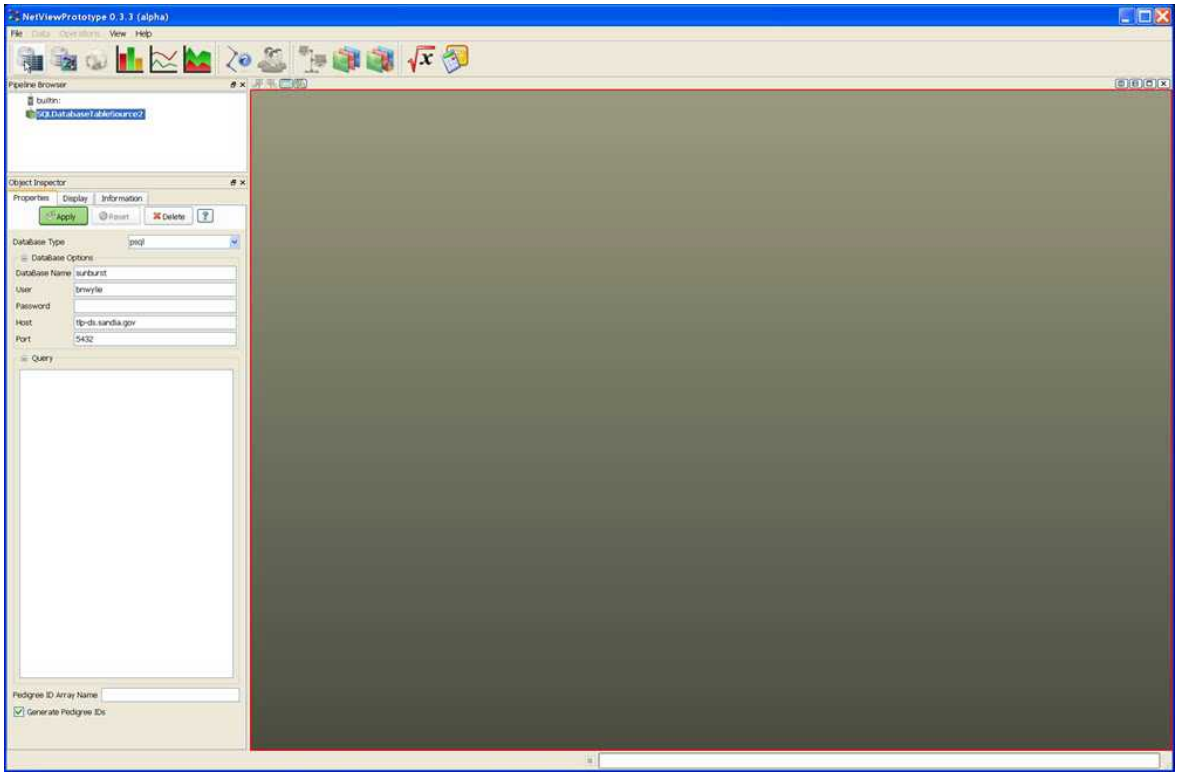


Sandia National Laboratories

Titan Resources

http://vizrd.srn.sandia.gov/vizwiki/Titan

Brian Wylie (1424), Tim Shead (1424)



Playback: 4x normal speed

Sandia National Laboratories