A Visual Analytics Framework for Analyzing Parallel and Distributed Computing Applications

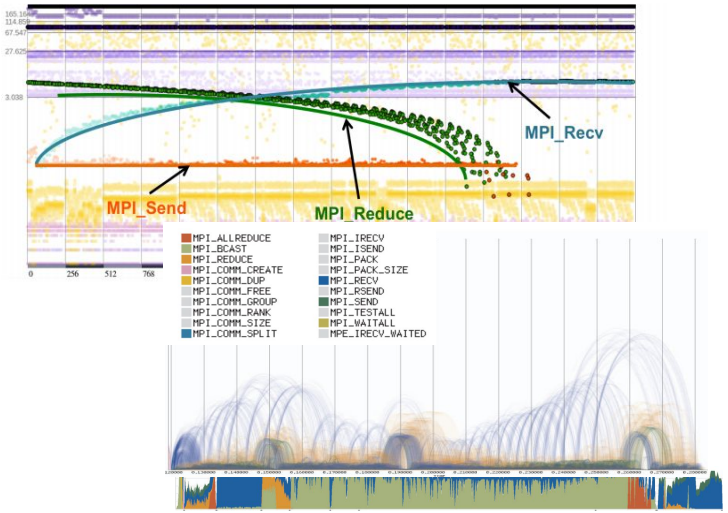
Kelvin Li[1], Takanori Fujiwara[1], Suraj P. Kesavan[1], Caitlin Ross[2], Misbah Mubarak[3], Christopher D. Carothers[2], Robert R. Ross[3], Kwan-Liu Ma[1]

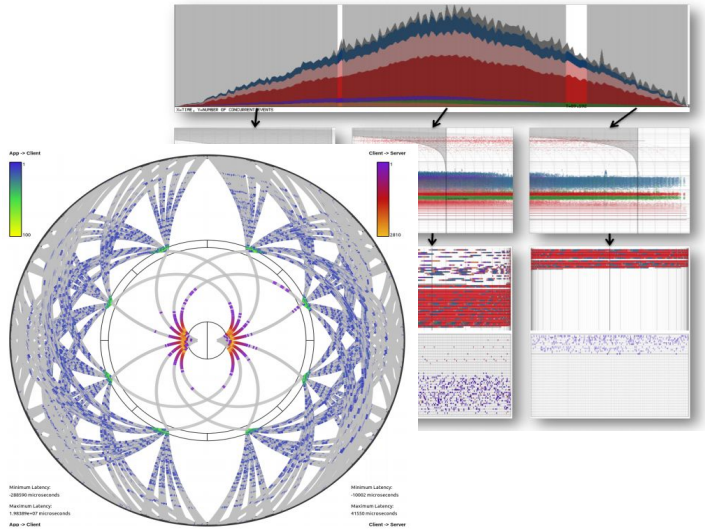
University of California, Davis (UCD)[1]
Rensselaer Polytechnic Institute, Troy (RPI)[2]
Argonne National Laboratory, Chicago (ANL)[3]

Oct. 20, 2019

vds

Visualization and Visual Analytics for HPC

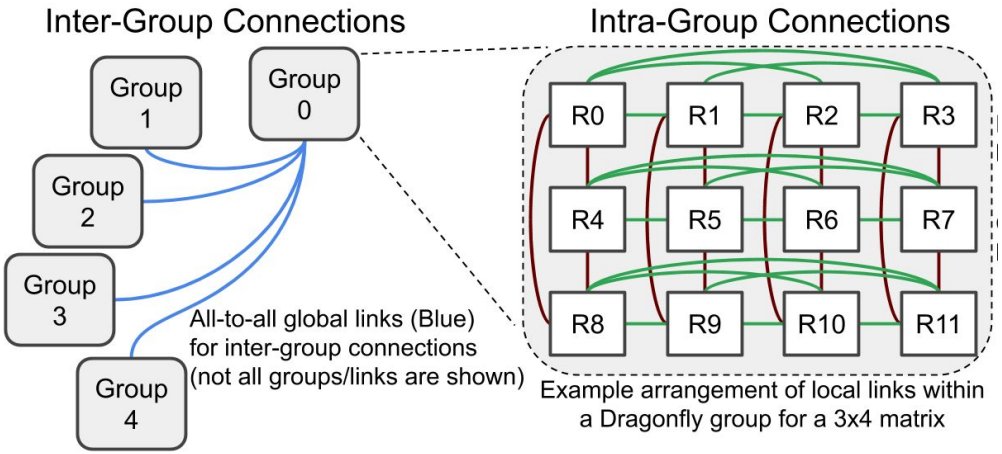




Isaacs et al. 2014: A survey on performance visualization

Advance in HPC and Supercomputing

- Exascale Computing
- More compute nodes
- Complex interconnects
  - Dragonfly (Kim et al. 2008)
  - Slimfly (Besta et al. 2014)



Argonne National Laboratory's Theta
https://www.alcf.anl.gov

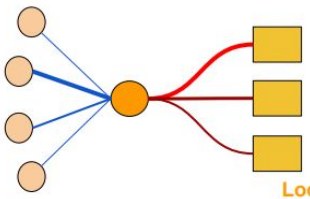
Analyzing HPC Data

Communication

Multivariate

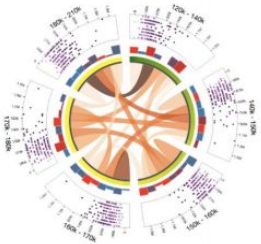
Temporal

Analyze flow at specific points



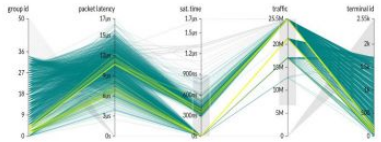
Local

Global



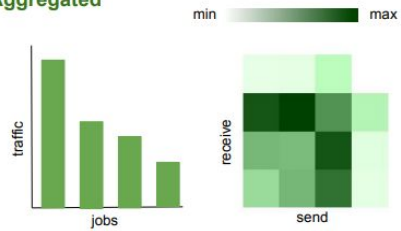
Explore connections and behaviors

Investigate details on demand



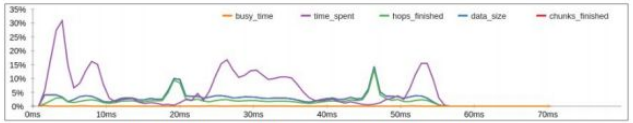
Detailed

Aggregated

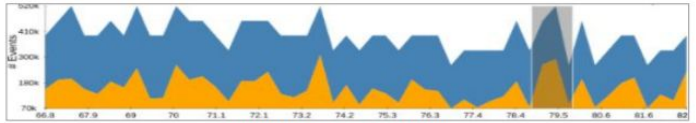


Provide summaries & statistics

Normalized | Absolute

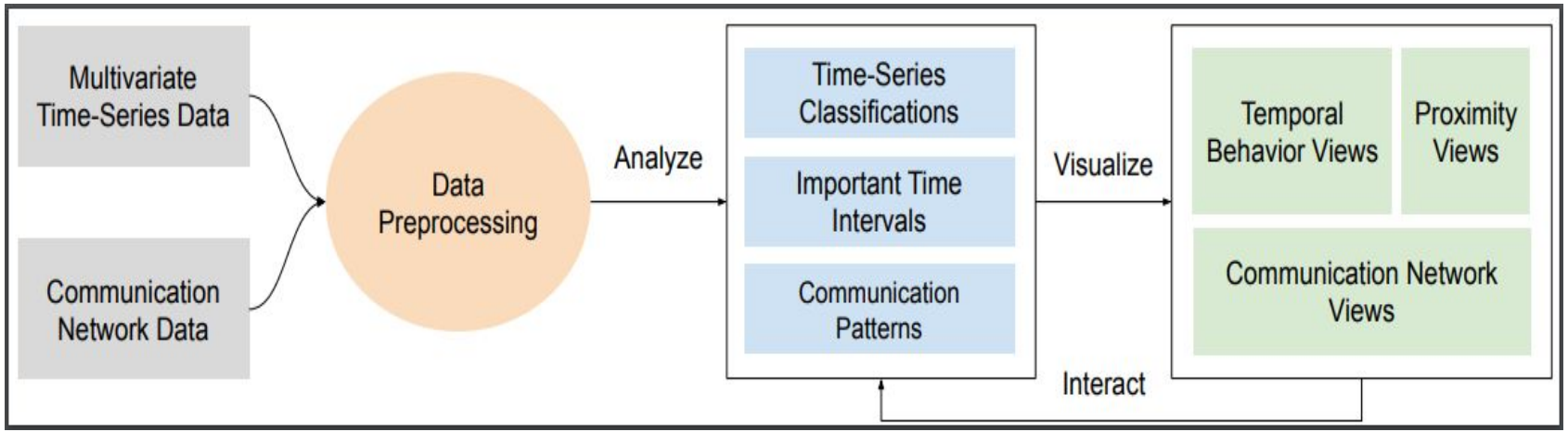


Overview temporal characteristics and behaviors



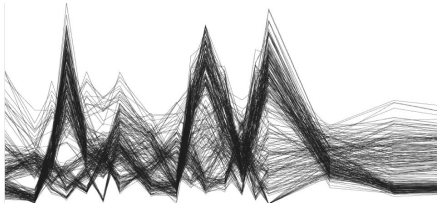
Explore temporal correlations

Overview of our Visual Analytics Framework

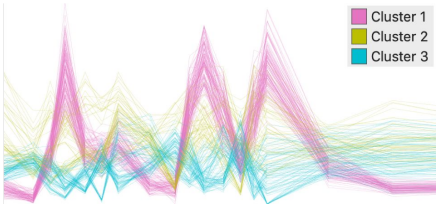


Time-Series Clustering and Classification

Without clustering



With clustering



Time-series data

Distance similarity measure

- Euclidean distance
- Dynamic Time Warping (DTW)
- Time Warp Edit Distance (TWED)

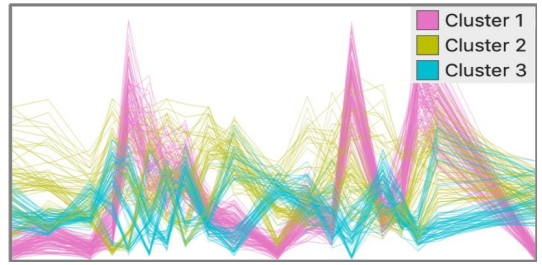
Clustering algorithms

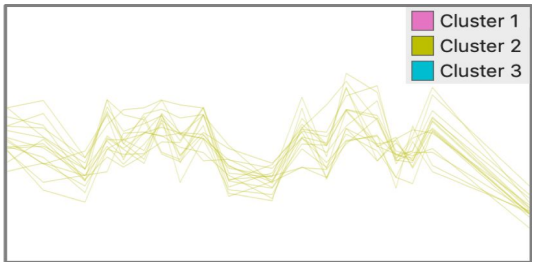
K-means[1]

K-mediods[2]

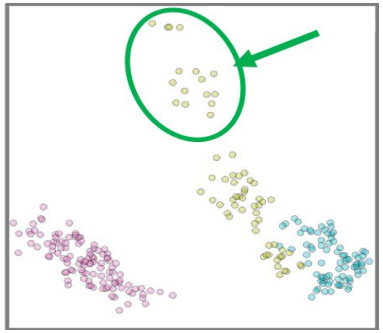
Hierarchical[3]

Time-Series Dimensionality Reduction

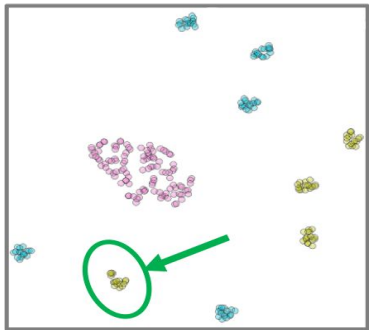


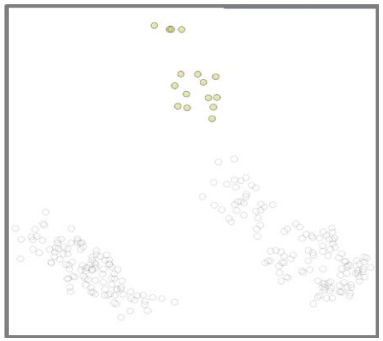


MDS

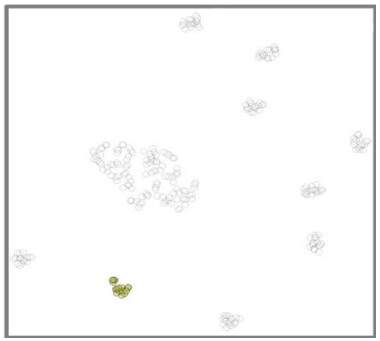


T-SNE



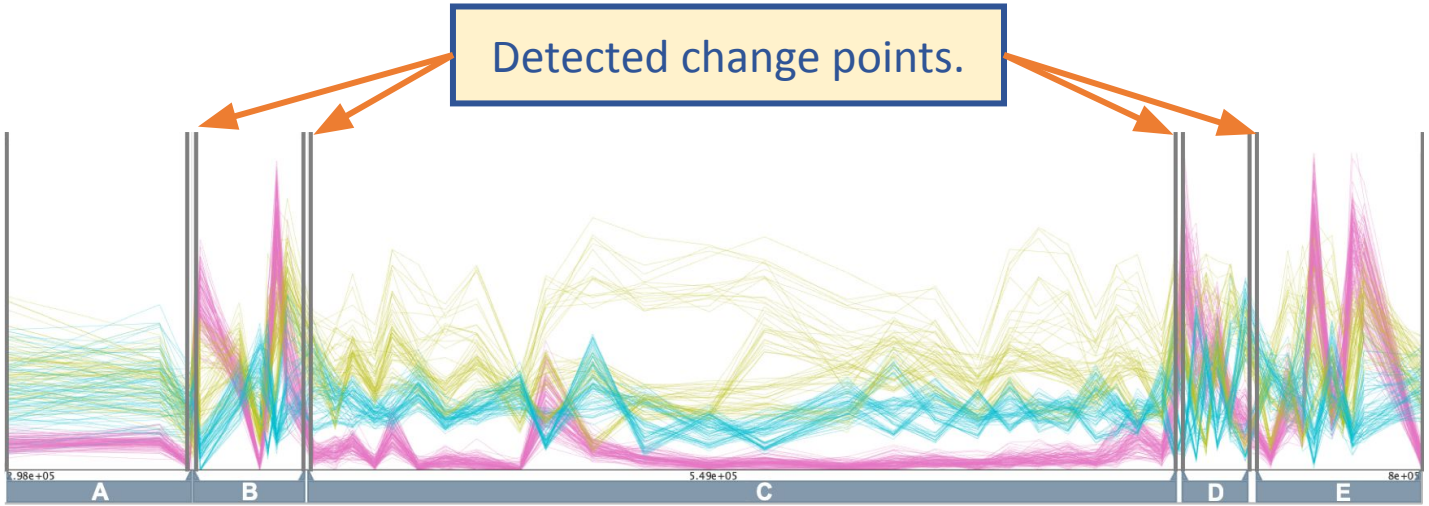


Global Structure



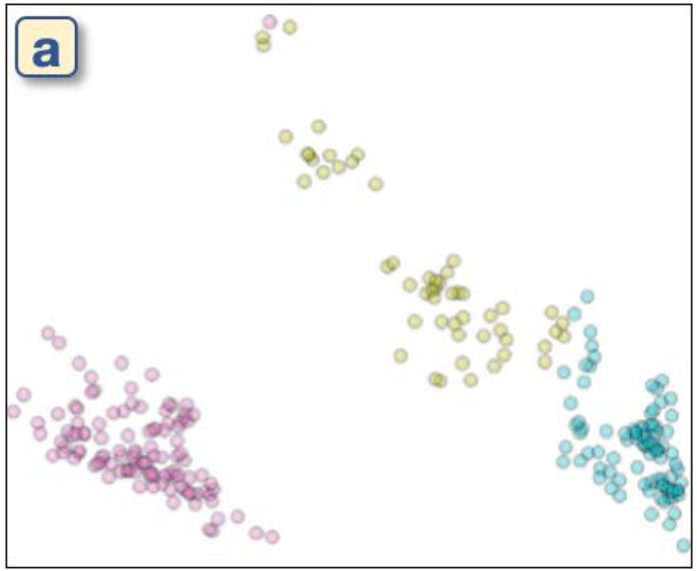
Local Structure

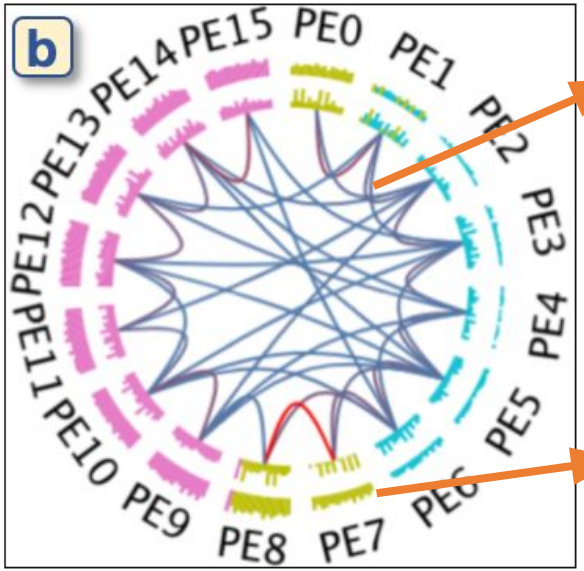
Change Point Detection



Automatic segmentation based on E-Divisive (James et al. 2015) for finding change points.

Communication Patterns





Line colors encodes communication events.

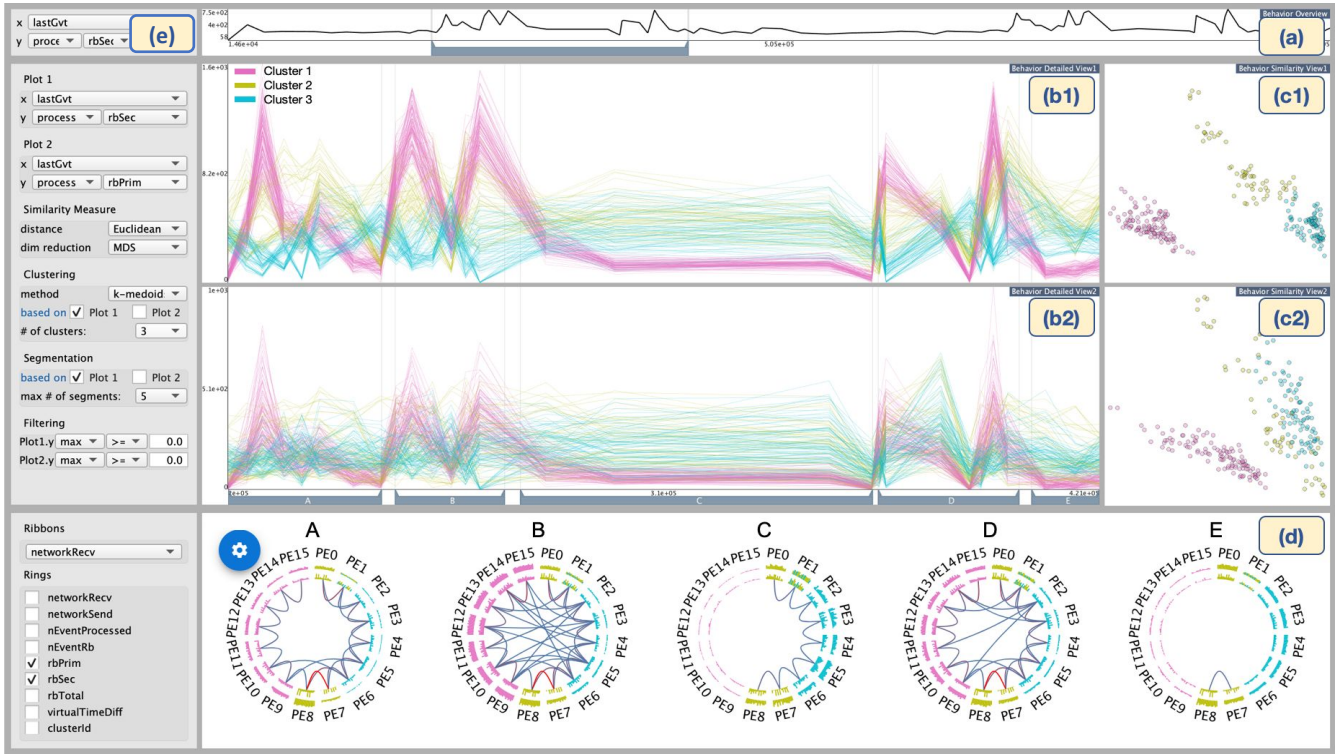
Bar heights represents metric values.

Case Study

Parallel discrete-event simulation (**PDES**) with optimistic scheduling

- Compute nodes process events individually.
- Two compute nodes only communicate on connected events.
- Events occurred in the wrong order needs to be **rollback**.

Visual Interface based on our framework for analyzing PDES



(a) See an overview of rollback behaviors.

(b1, b2) See detailed behaviors of two selected metrics.

(c1, c2) See similarities and dissimilarities among the processes.

(d) See network communication patterns for the selected intervals(A, B, C, D)

Analyzing PDES

- Quick overview of PDES behaviors and potential problems.
- Identify important stages or intervals of the PDES.
- Analyze when and where a performance bottleneck occurs.
- Explore the workload balance and similarities among compute nodes.

vds

Work in Progress

- In-situ data processing and online analysis.
- Real-time monitoring of PDES.
- Monitor and analyze different types of HPC applications.

vds

Thank You!

Acknowledgement

This research has been sponsored in part by the U.S. Department of Energy through grants DE-SC0007443, DE-SC0012610, and DE-SC0014917.