APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

Relaxed Linear Algebra Methods for Knowledge Discovery

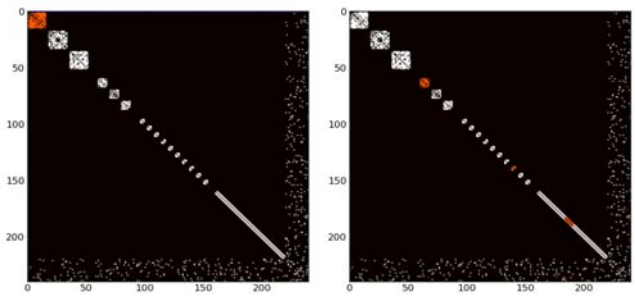
S&T Campaign: Computational Sciences
Predictive Simulation Sciences

Alexander Breuer, (410) 278-5525
Alexander.m.breuer.civ@mail.mil

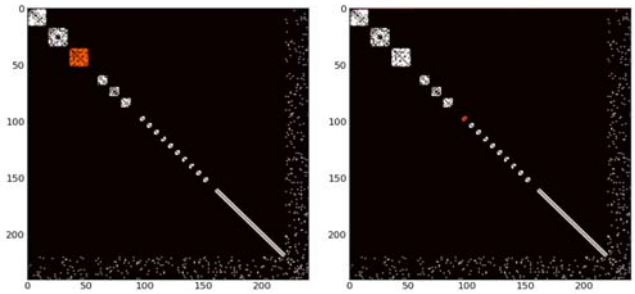
Research Objective

- **Knowledge discovery problems are frequently posed as linear algebra problems, and then solved iteratively.**
- **Current research suggests that approximate results may be as good as exact solutions, yet not reliable criteria exist to determine when a solution is sufficient for knowledge discovery.**
- **Approximate solutions may be available at *substantially* reduced costs: there may be an order of magnitude reduction in iterations.**
- **Analysis of the convergence of knowledge discovery-relevant features will enable substantial savings in solving these linear algebra problems.**

Approximate eigenvector communities



Eigenvector communities



Cliques and communities can be detected with eigenvectors (spectral clustering). Community detection using approximate (top) and exact (bottom) eigenvectors.

Challenges

- **Classical convergence norms — i.e., the 2-norm — may be too pessimistic for knowledge discovery convergence. Other norms may be more appropriate, but do not admit elegant analysis.**
- **Convergence of 1-norm results in traditional subspace methods, such as Krylov subspaces, is more difficult. Alternate, sparsity-preserving variants may be necessary.**

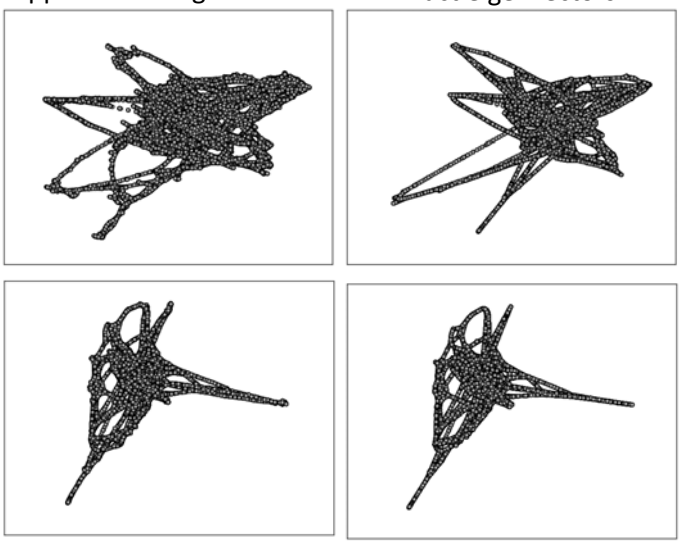
ARL Facilities and Capabilities Available to Support Collaborative Research

- **We have shown that eigenvector methods may not require subspace invariance; instead, only norm-maximization is necessary.**
- **We have shown that filtering properties of eigenvector projection methods may not be preserved in some popular approximation methods when classical start guesses are used (in preparation for submission to ETNA).**
- **We have developed noise filtering orthogonal projections that are far faster than eigenvectors.**
- ***Excalibur* (No. 26 on the top 500 supercomputer list) available for data analytics.**

Complementary Expertise/ Facilities/ Capabilities Sought in Collaboration

- **Large relational datasets for testing and evaluation are helpful.**
- **Co-developers in HPC software for data analytics are welcome. We currently use the Trilinos framework.**
- **Researchers interested in approximate results in scientific computing may find meaningful cross-pollination of ideas.**

Approximate eigenvectors | Exact eigenvectors



Approximate eigenvectors for visualizing networks: approximate eigenvectors (left) reveal much of the same structure as exact eigenvectors (right) but require an order of magnitude fewer FLOPS.

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED