Thrust Area 2: Fundamental physical data acquisition and analysis

Alfred Hero
Depts of EECS, BME, Statistics
University of Michigan

CONSORTIUM FOR VERIFICATION TECHNOLOGY

NNSA
National Nuclear Security Administration

Thrust II personnel

- Alfred Hero (UM EECS/BME/STATS): Event correlation and anomaly detection
- John Fisher (MIT CSAIL): dynamic graphical models
- Lawrence Carin (Duke ECE): compressive sensing for high-dimensional data
- Sara Pozzi/Shaun Clarke (UM NERS): physics of fission
- John Mattingly (NCSU NE): High-throughput radiation detection systems

Thrust II personnel (ctd)

- Funded by CVT
  - Elizabeth Hou (UM STATS): CVT Fellow (poster)
  - Charles Sosa (UM NERS), CVT Fellow (poster)
  - David Carlson (Duke ECE), CVT Fellow
  - Sue Zheng (MIT CSAIL), CVT Fellow*
  - Yassin Yilmaz (UM EECS): post-doctoral fellow (poster)
  - Taposh Banerjee (UM EECS): post-doctoral fellow (poster)
  - Angela Di Fulvio (UM NERS), post-doctoral fellow funded
  - Xuejun Liao (Duke ECE), post-doctoral fellow
  - Oren Freifeld (MIT CSAIL), post-doctoral fellow*

- Funded from non-CVT sources
  - Tony Van (UM STATS): M.S. student (poster)
  - Matthew Marcath (UM NERS), Ph.D candidate (poster)
  - Tony Shin (UM NERS), M.S. student
  - Steve Ward (UM NERS), M.S. student

Thrust II Kickoff Presentations

Oral presentations

- "Fundamental physical data acquisition and analysis," Al Hero (UM)
- "Convolutional dictionary learning and feature design," Larry Carin (Duke)
- "Graphical models for query-driven analysis of multimodal data," John Fisher (MIT)
- "Correlations in prompt neutrons and gamma rays from fission," Shaun Clarke (UM)
- "Data compression and analysis methods for high-throughput Detector Systems," John Mattingly (NCSU)

Also see our Thrust II poster presentations

Multi-layered data acquisition

Standoff verification layer

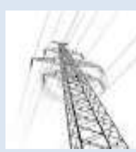



Context info
- ISR
- Public data
- Shipping/Rcv

Sensing modes
- Reports
- Newsfeeds
- Event tagging
- Social media
- Blogs
- Microblogs

On site verification layer

Inputs → Facility → Outputs




Utilities usage

Sensing modes
- Power meter
- Flow rates
- Webcams

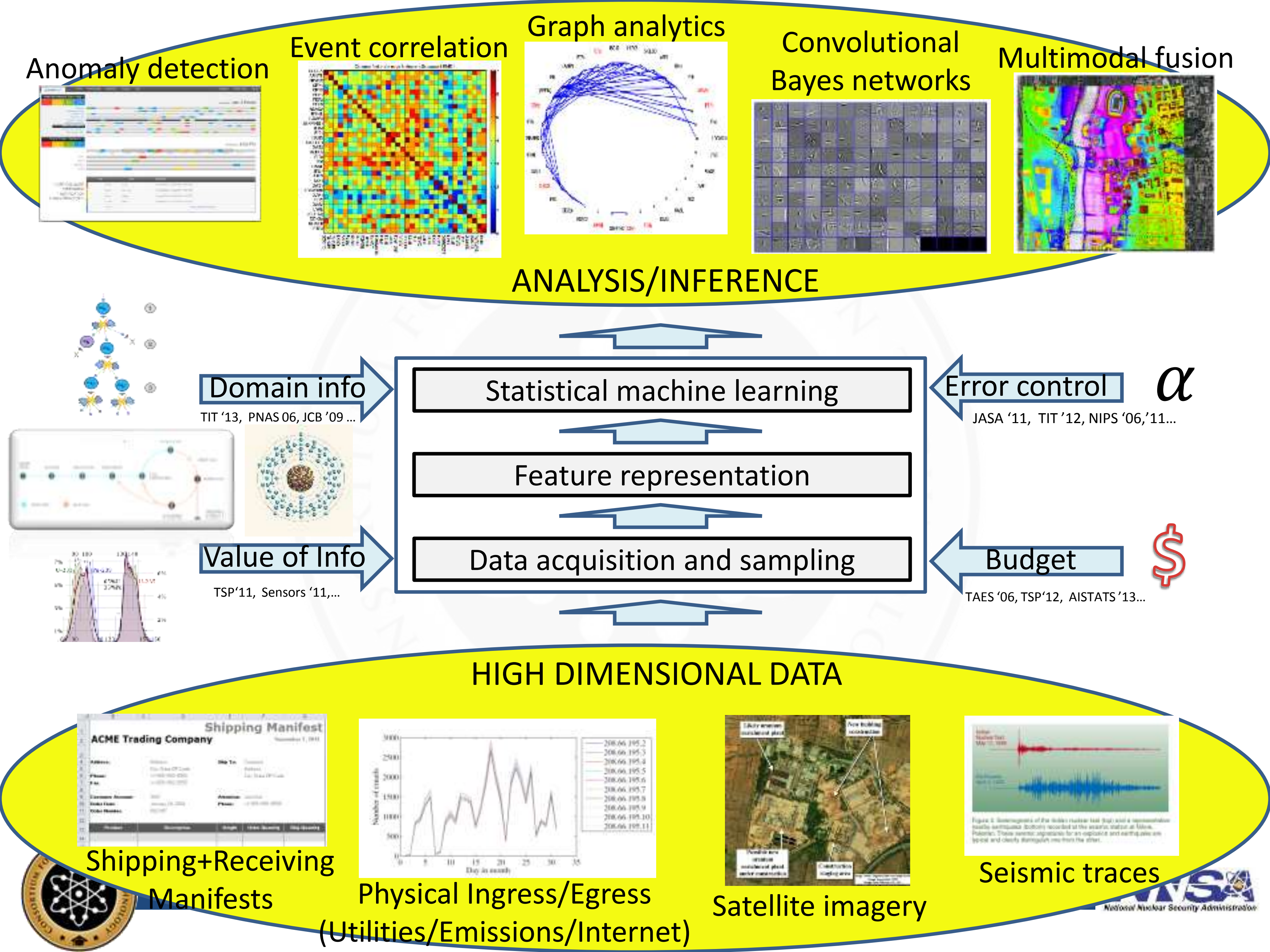
Site visits


Emissions

Sensing modes
- Neutron det
- γ imaging
- Chemical det

MASINT
- Satellite RS
- Flyby sensors
- Earthbased RS

Sensing modes:
- EO/SAR/IR
- Hyperspectral
- Radio Freq.
- X-ray/gamma
- Infrasound
- Seismic



Anomaly detection
Event correlation
Graph analytics
Convolutional
Bayes networks
Multimodal fusion
ANALYSIS/INFERENCE
Domain info
TIT '13, PNAS 06, JCB '09 ...
Statistical machine learning
Error control
JASA '11, TIT '12, NIPS '06,'11...
α
Feature representation
Value of Info
TSP'11, Sensors '11,...
Data acquisition and sampling
Budget
$
TAES '06, TSP'12, AISTATS '13...
HIGH DIMENSIONAL DATA
Shipping Manifest
ACME Trading Company
Shipping+Receiving
Manifests
Physical Ingress/Egress
(Utilities/Emissions/Internet)
Satellite imagery
Seismic traces
National Nuclear Security Administration

Event correlation and anomaly detection

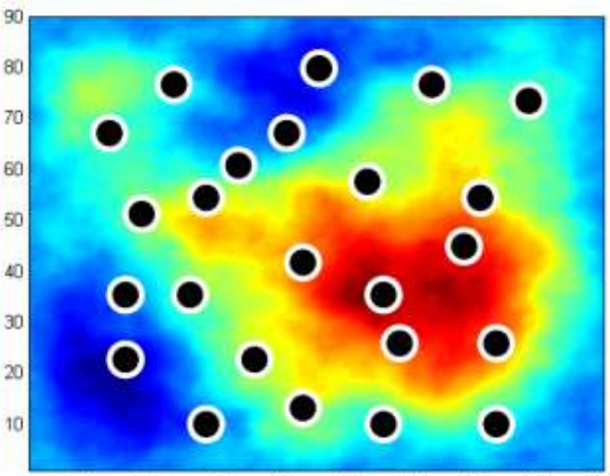
- Challenges
  - Sensors are highly distributed and asynchronous
    - Large standoff: satellite EO/IR imaging, SAR, RF, seismic, ISR
    - On-site: utility monitoring, surveillance, radionuclide detectors, emissions, outflows
  - Information sources are diverse
    - Video, images, waveforms, text
  - Event correlation at different time/space scales
  - Incipient changes may be barely detectable

Event correlation and anomaly detection

- Elements of our approach
  - Statistical hierarchical modeling of heterogeneous event streams
  - Correlation mining with constraints on communication/computation/timeliness
  - Fundamental performance limits and benchmarks
- Application areas
  - Human-aided anomaly detection
  - Event-driven compressive sampling
  - Quickest change detection
  - Distributed event correlation
- See our poster today for details on these areas

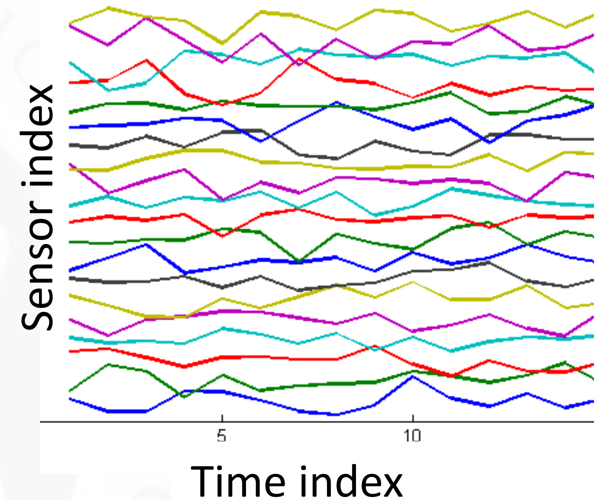
Correlation mining

Network of sensors measures spatio-temporal random field



20 sensors in a random field

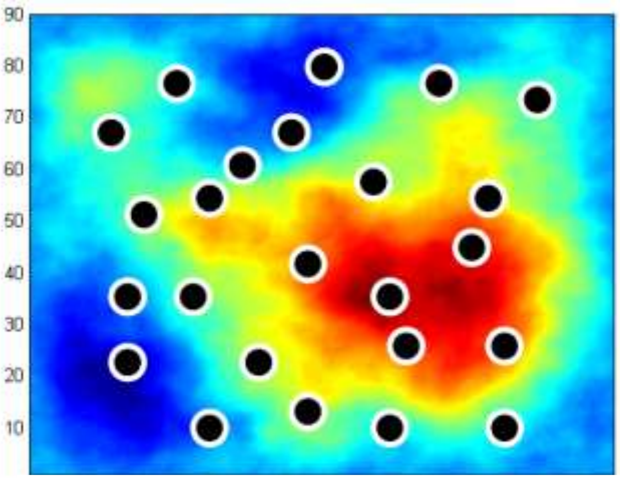
Data streams



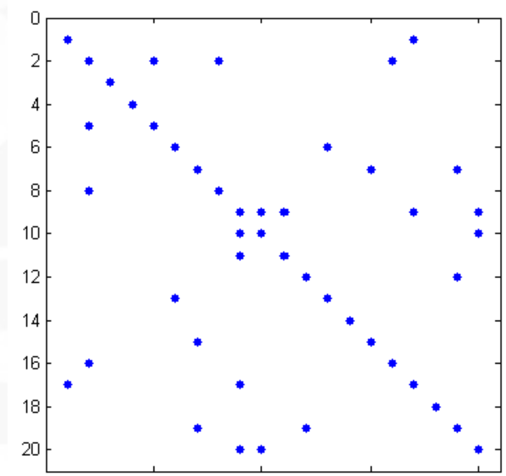
- Are any of the streams correlated over space or time?
- Are there interesting patterns of correlation?
- Have these patterns changed recently wrt a baseline?
- How much data is required to answer these questions?

The problem of false alarms

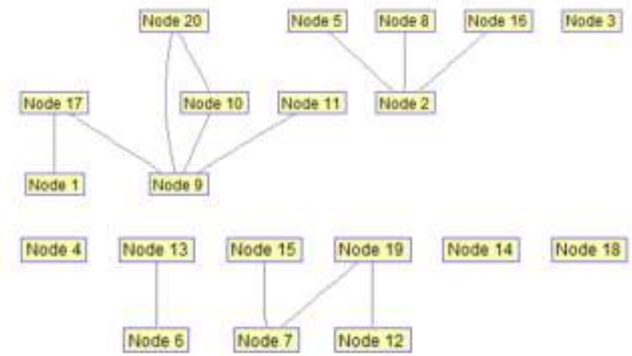
Network of sensors measures spatio-temporal random field



20 sensors in a random field



Thresholded correlation



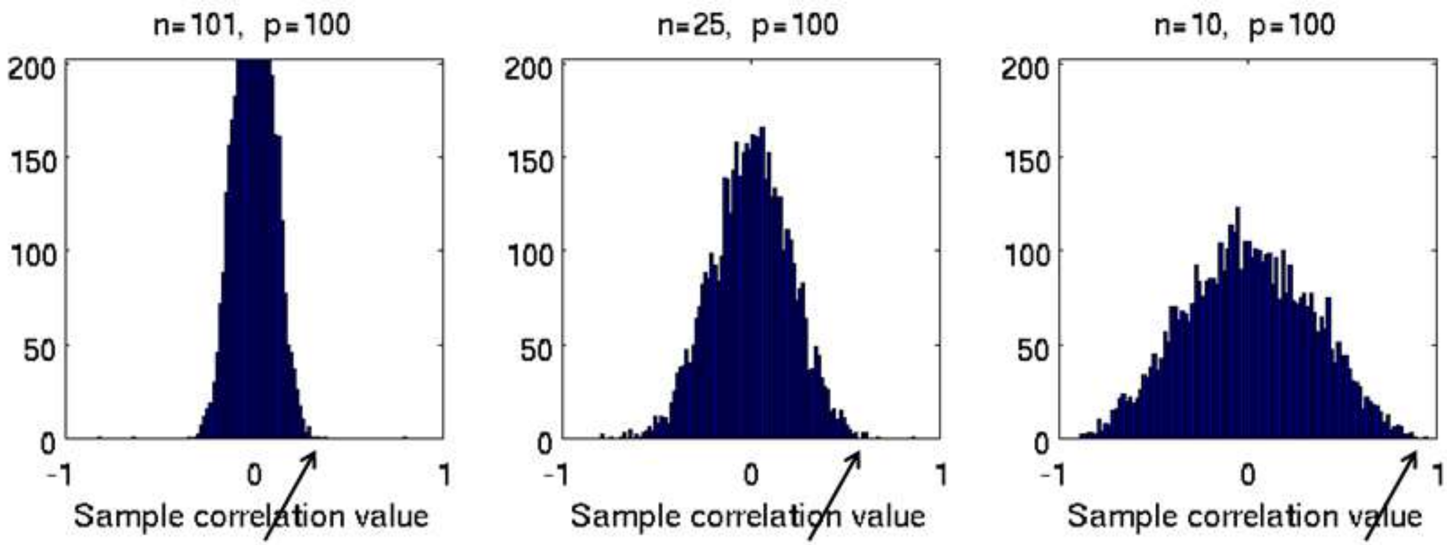
Correlation network

- Event detection: a pattern of correlation between sensors exceeds a threshold ρ
- Question: What is minimum required number n of samples to correlate information from p different sensors?
- Answer: Can determine from critical phase transition threshold [1]

[1] A.O. Hero and B. Rajaratnam, "Large Scale Correlation Screening," JASA, 2011

The problem of false alarms

- When correlation matrix is sparse there is phase transition



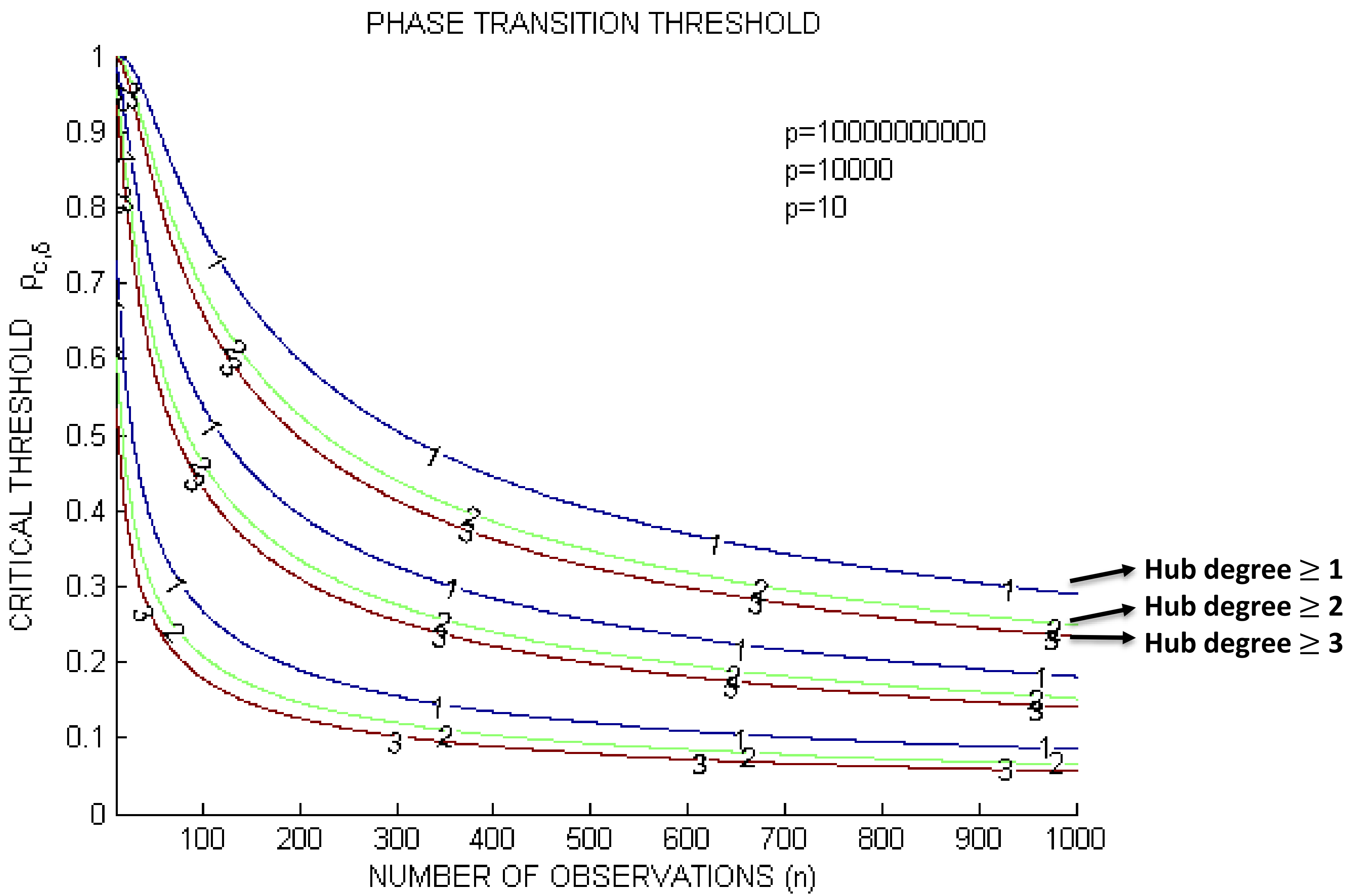
$\rho_c = \pm 0.34$ $\rho_c = \pm 0.63$ $\rho_c = \pm 0.89$

- Phase transition encountered as decrease the threshold ρ
- Critical phase transition threshold ρ_c increases in n and p [1]

$$\rho_c = \sqrt{1 - c_n (p-1)^{-2/(n-4)}}$$

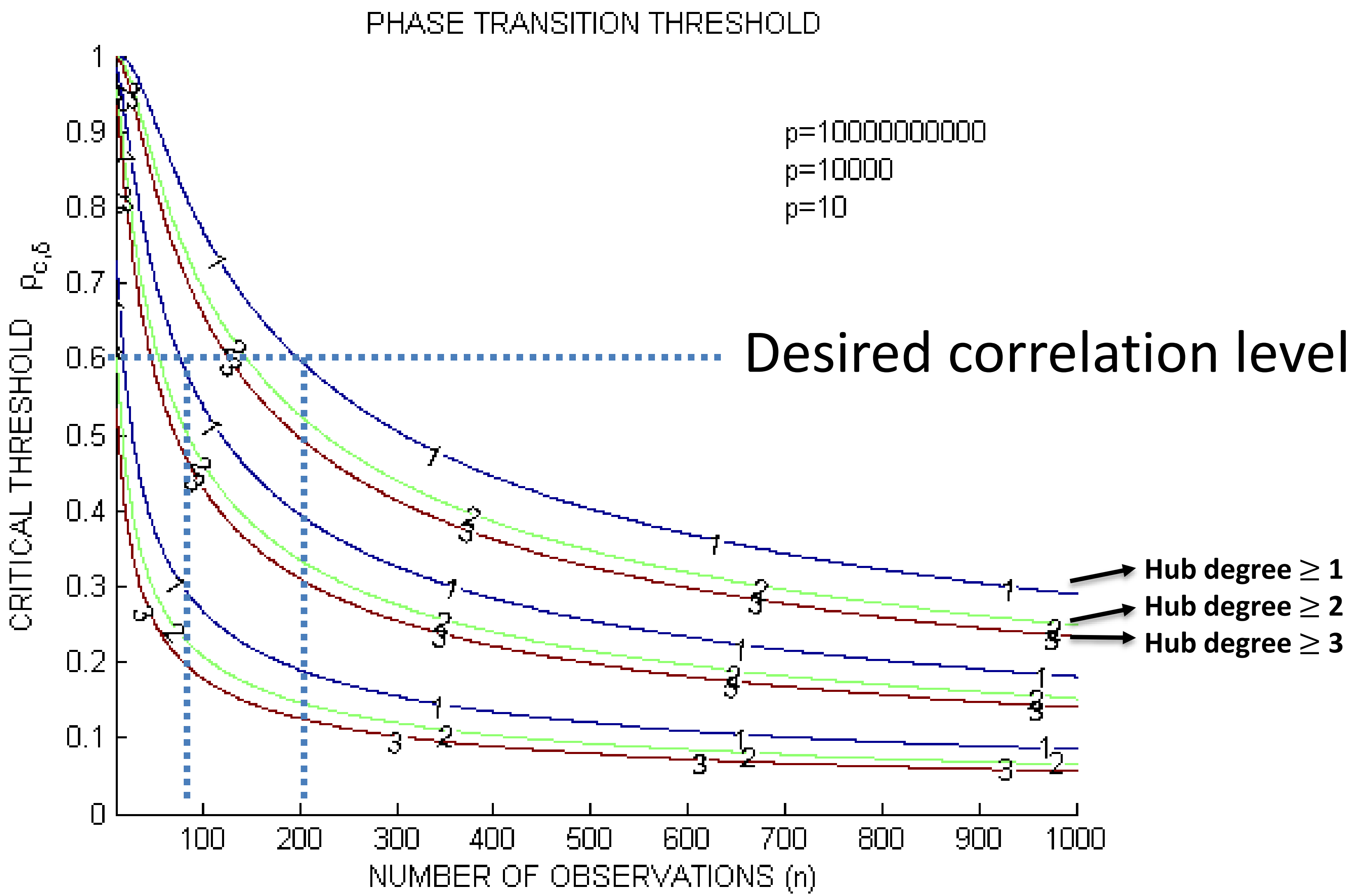
[2] A.O. Hero and B. Rajaratnam, "Large Scale Correlation Screening," JASA, 2011

Phase transition chart



[3] A.O. Hero and B. Rajaratnam, "Hub screening," IEEE Trans. Info Theory, 2012

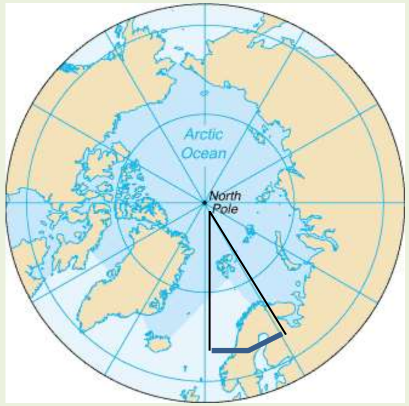
Phase transition chart



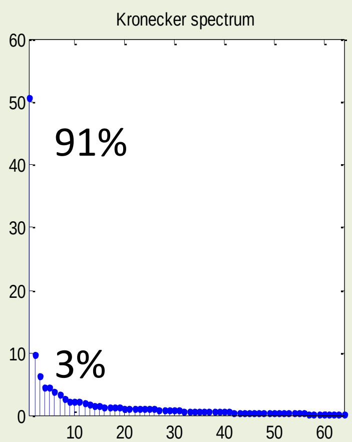
[3] A.O. Hero and B. Rajaratnam, "Hub screening," IEEE Trans. Info Theory, 2012

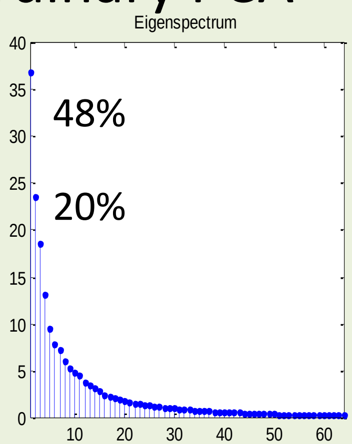
Spatio-temporal correlation mining

- Wind speed data (1948-2012)
  - 100 stations 10x10 grid
  - 2-day time windows (8 sample snapshots)
  - Period: 2001 to 2007
  - p=100, q=8, n=224



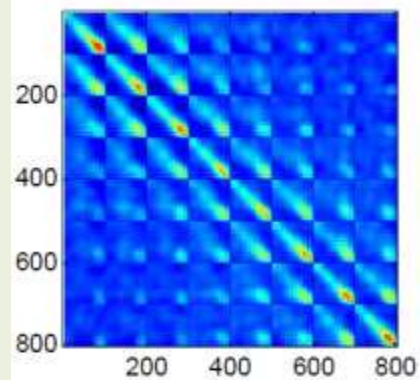
- Kronecker vs ordinary PCA



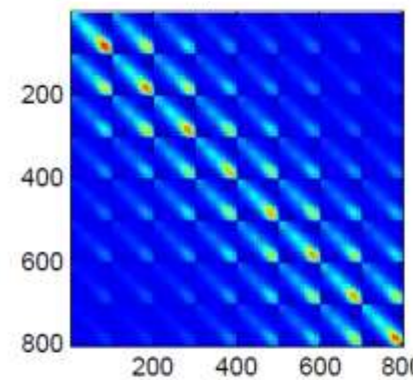


Kronecker PCA: $C = \sum_i A_i \otimes B_i$

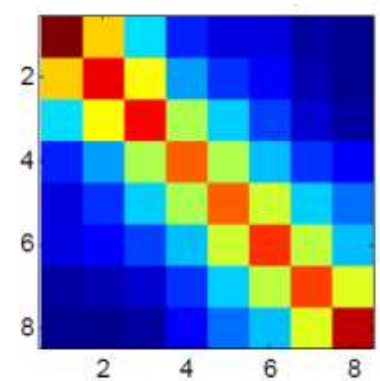
Sample cov C



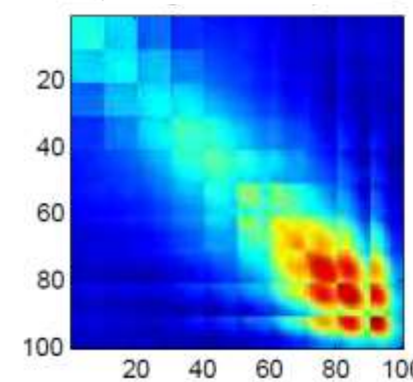
K-PCA Approx



A_1 (Temporal)



B_1 (Spatial)



[3] Tsiligkaridis and Hero, "Cov Estim...Kronecker Product Expansions," IEEE SP 2013

Conclusions

- Analysis team brings expertise from the areas of
  - Statistical machine learning and graphical models
  - Anomaly detection, quickest detection and correlation mining
  - Compressive sensing and dictionary learning
  - Physical models and their simulation
- Fundamental limits and algorithms and models are equally important.
- See our poster today:

"Event correlation and anomaly detection," Elizabeth Hou, Yasin Yilmaz, Tony Van, Taposh Banerjee, Al Hero