ILMENAU UNIVERSITY OF TECHNOLOGY

ISWCS'13

Sparsity and Low Rank for Robust Social Data Analytics and Networking

Georgios B. Giannakis

Acknowledgments: Morteza Mardani and Dr. Gonzalo Mateos
MURI Grant No. AFOSR FA9550-10-1-0567
NSF grant ECCS-1202135

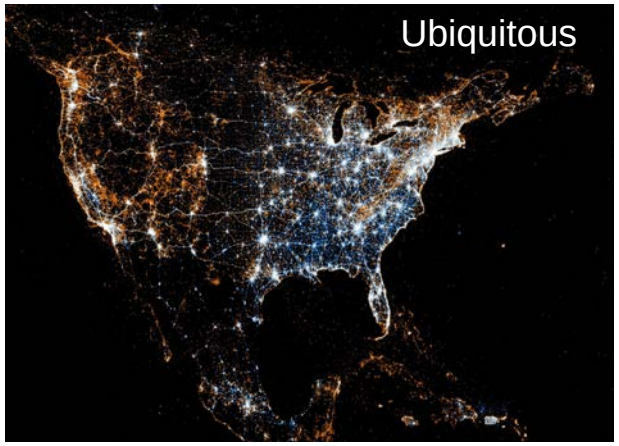
UNIVERSITY OF MINNESOTA
Driven to Discover℠

August 29, 2013
Ilmenau, Germany

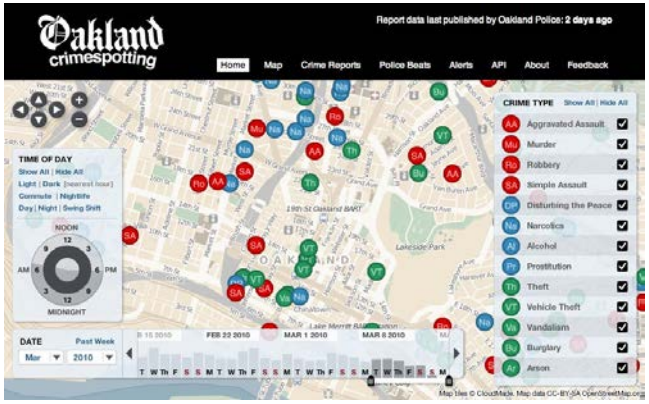
Learning from Big Data

"Data are widely available, what is scarce is the ability to extract wisdom from them"

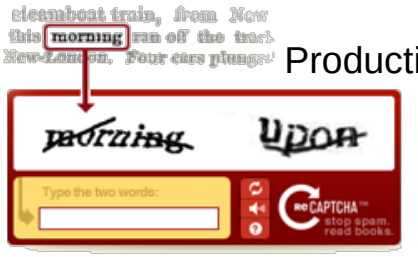
Hal Varian, Google's chief economist



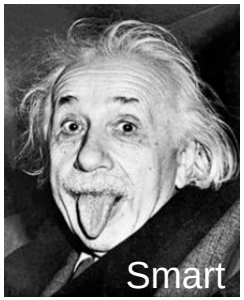
Fast

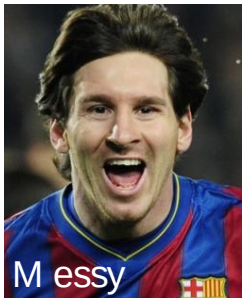


Revealing



Productive

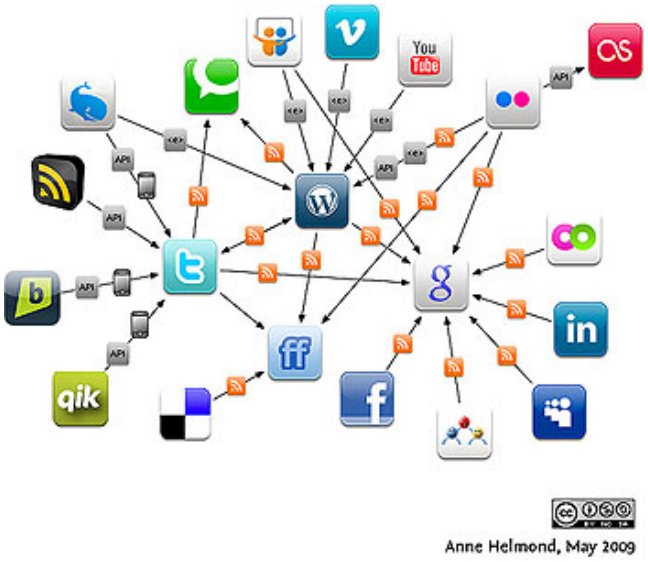




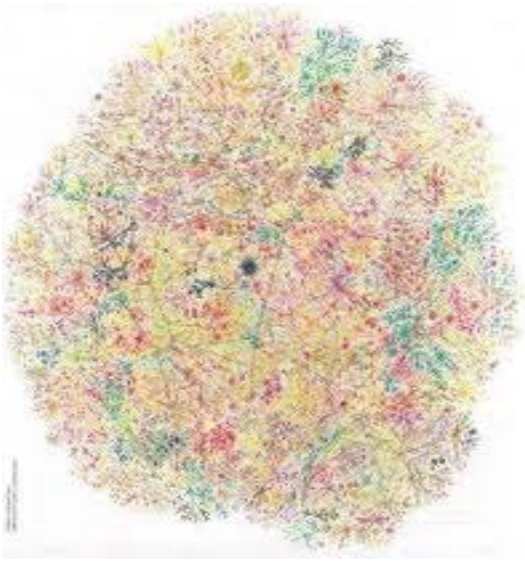
K. Cukier, "Harnessing the data deluge," Nov. 2011.

Network-science analytics

Online social media



Internet



Clean energy and grid analytics

- High-level goal: process, analyze, and learn from large pools of network data
- The means here: leverage sparsity and low rank
  - *Complexity control* through parsimonious model selection
  - *Robustness* to outliers

E. D. Kolaczyk, *Statistical Analysis of Network Data*, Springer, 2010.

Roadmap

- Sparsity control for robust principal component analysis
  - Least-trimmed squares and outlier sparsity
  - Algorithm and applications

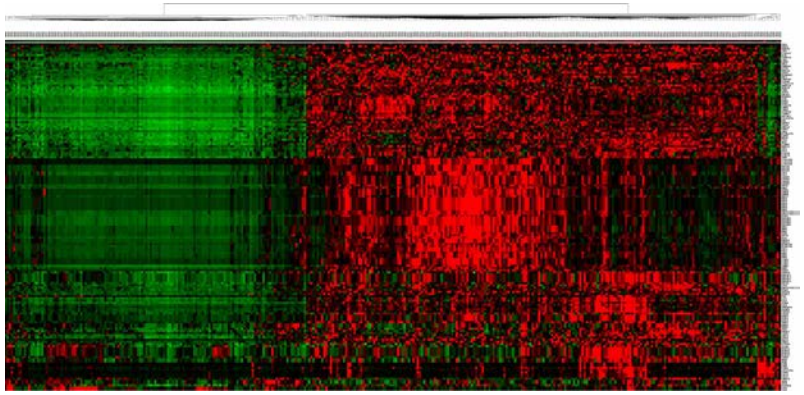
- Unveiling network anomalies via sparsity and low rank
  - Traffic modeling and identifiability
  - (De-) centralized and online algorithms
  - Numerical tests

- Conclusions and future research directions

Principal component analysis

- **Motivation:** (statistical) learning from high-dimensional data

DNA microarray



Traffic surveillance

- Principal component analysis (PCA) [Pearson' 1901]
  - Extraction of low(est)-dimensional structure
  - Applications: source (de)coding, anomaly ID, recommender systems …
  - PCA is non-robust to outliers [Huber'81], [Jolliffe'86], [Wright et al'09-12]

Objective: robustify PCA by controlling outlier sparsity

PCA formulations

- Training data $\{\mathbf{y}_t \in \mathbb{R}^p\}_{t=1}^T, \quad \hat{\mathbf{\Sigma}} = \sum_{t=1}^T \mathbf{y}_t \mathbf{y}_t'/T$

- Minimum reconstruction error
  - Compression operator $\mathbf{B} \in \mathbf{R}^{q \times p}, \; q \leq p$
  - Reconstruction operator $\mathbf{C} \in \mathbf{R}^{p \times q}$

$$\min_{\{\mathbf{C},\mathbf{B}\}} \sum_{t=1}^T \|\mathbf{y}_t - \mathbf{C}\mathbf{B}\mathbf{y}_t\|_2^2, \quad \text{s.to } \mathbf{C}'\mathbf{C} = \mathbf{I}_q$$

- Component analysis model $\mathbf{y}_t = \mathbf{C}\mathbf{w}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T$

$$\min_{\{\mathbf{C},\mathbf{w}_t\}} \sum_{t=1}^T \|\mathbf{y}_t - \mathbf{C}\mathbf{w}_t\|_2^2, \quad \text{s.to } \; \mathbf{C}'\mathbf{C} = \mathbf{I}_q$$

Solution: $\hat{\mathbf{C}} = q\text{-evecs}[\hat{\mathbf{\Sigma}}], \quad \hat{\mathbf{B}} = \hat{\mathbf{C}}', \quad \hat{\mathbf{w}}_t = \hat{\mathbf{C}}'\mathbf{y}_t$

Robustifying PCA

- Outlier variables $\{\mathbf{o}_t\}_{t=1}^T$ s.t. $\mathbf{o}_t = \begin{cases} \mathbf{x}_t \neq \mathbf{0}_p, & \mathbf{y}_t \text{ outlier} \\ \mathbf{0}_p, & \text{otherwise} \end{cases}$

$$\mathbf{y}_t = \mathbf{C}\mathbf{w}_t + \mathbf{o}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T$$

  - Nominal data obey $\mathbf{y}_t = \mathbf{C}\mathbf{w}_t + \boldsymbol{\varepsilon}_t$; outliers something else
  - Linear regression [Fuchs'99], [Giannakis et al'11]

  - Both $\{\mathbf{C}, \mathbf{w}_t\}$ and $\mathbf{O} := [\mathbf{o}_1, \ldots, \mathbf{o}_T]'$ unknown, $\mathbf{O}$ typically sparse!

- Natural (but intractable) estimator

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{O}\}} \|\mathbf{Y} - \mathbf{W}\mathbf{C}' - \mathbf{O}\|_F^2 + \lambda_0 \|\mathbf{O}\|_0, \quad \text{s.to } \mathbf{C}'\mathbf{C} = \mathbf{I}_q. \qquad \text{(P0)}$$

G. Mateos and G. B. Giannakis, ``Robust PCA as bilinear decomposition with outlier sparsity regularization,'' *IEEE Transactions on Signal Processing*, pp. 5176-5190, Oct. 2012.

Universal robustness

$$\|\mathbf{O}\|_{2,r} := \sum_{t=1}^{T} \|\mathbf{o}_t\|$$

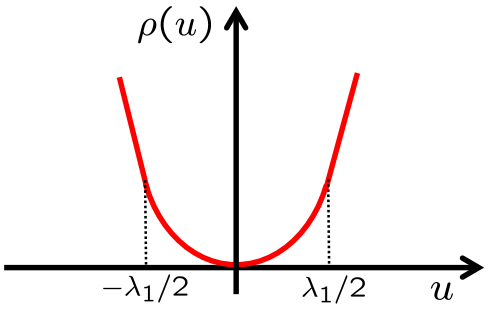
- (P0) is NP-hard $\Longrightarrow$ relax $\|\mathbf{O}\|_0$ e.g., [Tropp'06]

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{O}\}} \|\mathbf{Y}-\mathbf{W}\mathbf{C}'-\mathbf{O}\|_F^2+\lambda_1\|\mathbf{O}\|_{2,r}, \quad \text{s.to } \mathbf{C}'\mathbf{C}=\mathbf{I}_q. \qquad \text{(P1)}$$

  - Role of sparsity-controlling λ_1 is central

Q: Does (P1) yield robust estimates $\{\hat{\mathbf{C}}, \hat{\mathbf{w}}_t\}$?

A: Yap! Huber estimator is a special case



$$\min_{\{\mathbf{C},\mathbf{w}_t\}} \sum_{t=1}^{T} \rho(\mathbf{y}_t - \mathbf{C}\mathbf{w}_t)$$

$$\text{s.to } \mathbf{C}'\mathbf{C} = \mathbf{I}_q$$

$$\rho(\mathbf{r}) := \begin{cases} \|\mathbf{r}\|^2, & \|\mathbf{r}\| \le \lambda_1/2 \\ \lambda_1\|\mathbf{r}\| - \lambda_1^2/4, & \|\mathbf{r}\| > \lambda_1/2 \end{cases}.$$

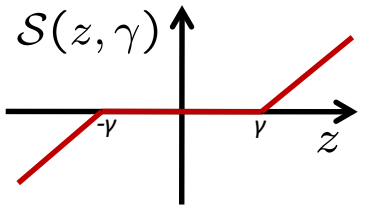
Alternating minimization

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{O}\}} \|\mathbf{Y}-\mathbf{W}\mathbf{C}'-\mathbf{O}\|_F^2+\lambda_1\|\mathbf{O}\|_{2,r}, \quad \text{s.to } \mathbf{C}'\mathbf{C}=\mathbf{I}_q. \qquad \text{(P1)}$$

Algorithm 1 : Batch AM solver

Initialize $\mathbf{O}(0)=\mathbf{0}_{T\times p}$.
for $k=1,\ldots$ **do**
  Form $\mathbf{Y}_o(k)=\mathbf{Y}-\mathbf{O}(k-1)$.
  Compute $\mathbf{U}(k)\mathbf{D}(k)\mathbf{V}(k)'=\text{svd}_q[\mathbf{Y}_o(k)]$.
  Update $\mathbf{C}(k)=\mathbf{V}(k)$ and $\mathbf{W}(k)=\mathbf{U}(k)\mathbf{D}(k)$.
  Update $\mathbf{O}(k)=\mathcal{S}\left(\mathbf{Y}-\mathbf{W}(k)\mathbf{C}'(k),\lambda_1/2\right)$.
end for

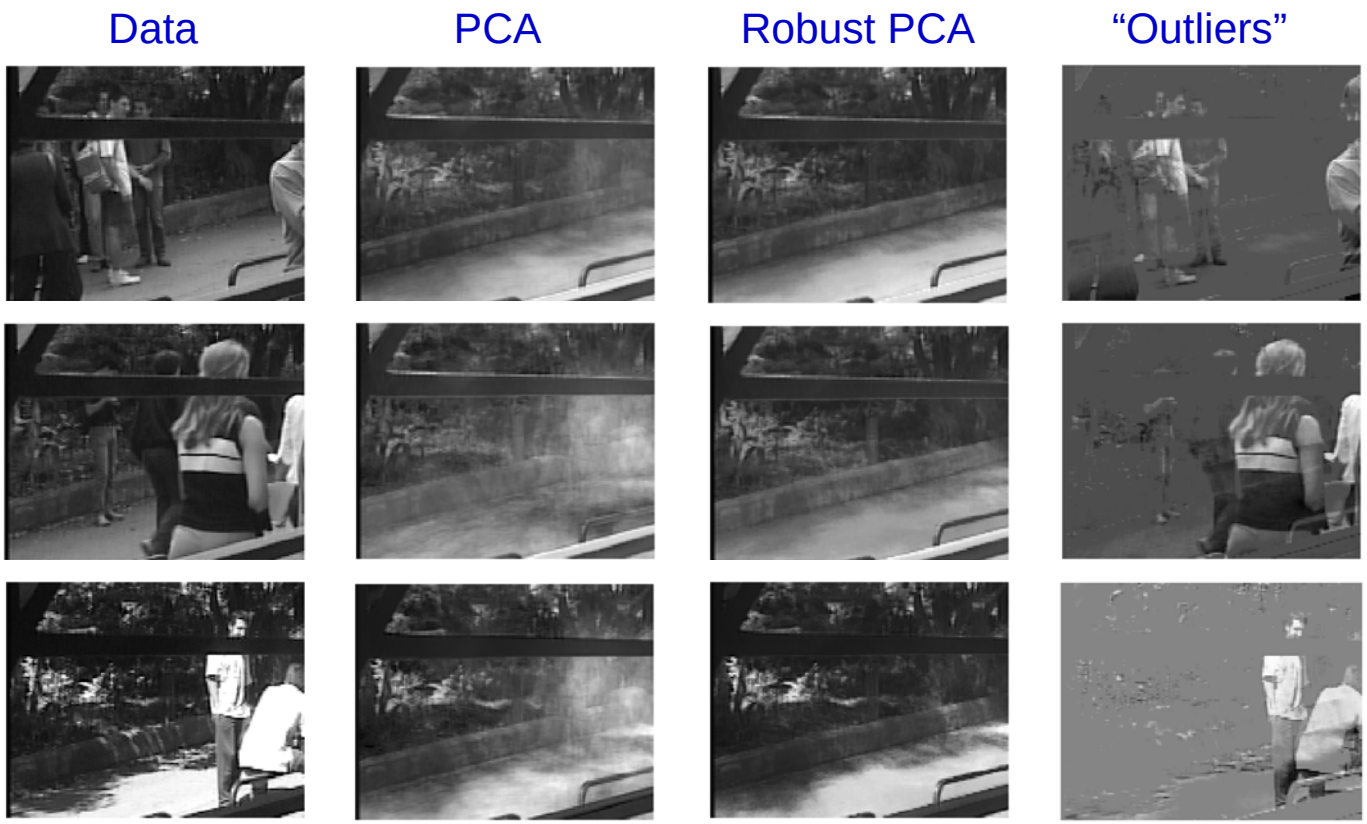
- $\mathbf{C}$ update: SVD of outlier-compensated data
- $\mathbf{O}$ update: row-wise soft-thresholding of residuals



Proposition : Algorithm 1's iterates converge to a stationary point of (P1)

Video surveillance

- Background modeling from video feeds [De la Torre-Black '01]

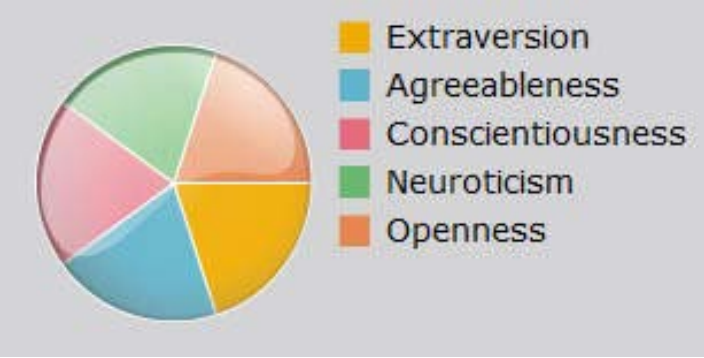


$T = 520$
$p = 19200$
$q = 10$

Data: http://www.cs.cmu.edu/~ftorre/

Big Five personality factors

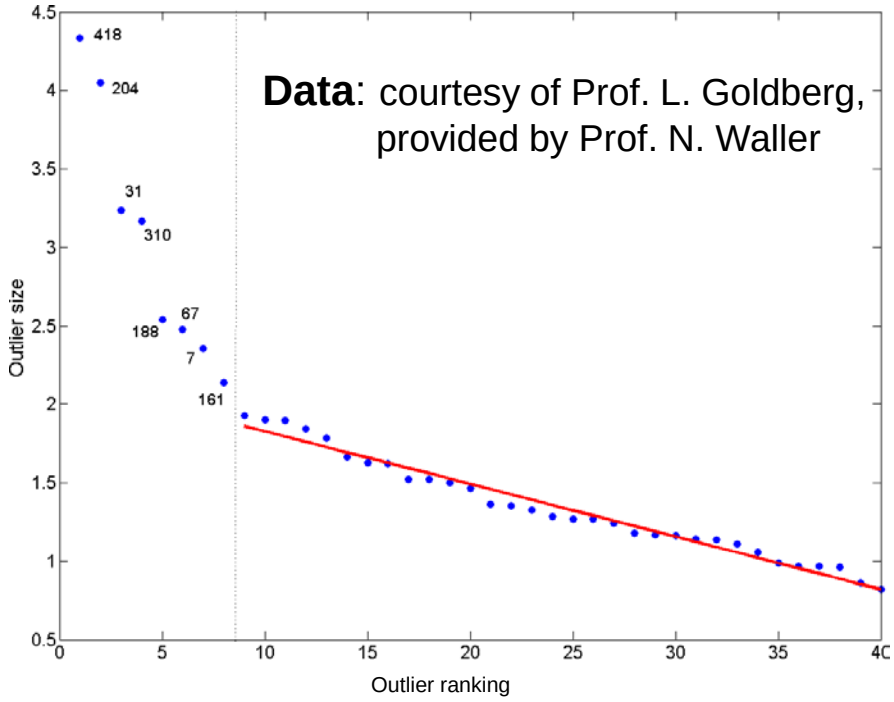
- Measure five broad dimensions of personality traits [Costa-McRae' 92]



- Big Five Inventory (BFI)
  - Short-questionnaire (44 items)
  - Rate 1-5, e.g.,
    `I see myself as someone who…
    …is talkative'
    …is full of energy'

Eugene-Springfield BFI sample (WEIRD)

$T = 437,\ p = 44,\ q = 5$



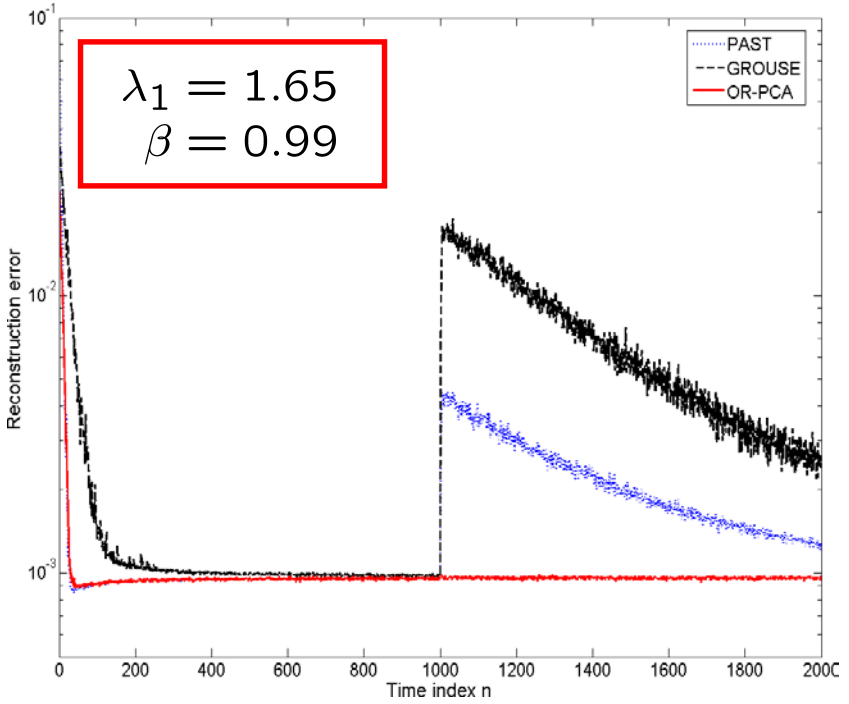
Handbook of personality: Theory and research, O. P. John, R. W. Robins, and L. A. Pervin, Eds. New York, NY: Guilford Press, 2008.

Online robust PCA

$$\min_{\{\mathbf{C},\mathbf{w}_n,\mathbf{o}_n\}} \sum_{n=1}^{N} \beta^{N-n} \left[\|\mathbf{y}_n - \mathbf{C}\mathbf{w}_n - \mathbf{o}_n\|_2^2 + \lambda_1 \|\mathbf{o}_n\|_2 \right], \quad 0 < \beta \leq 1$$

- Scalability via exponentially weighted subspace tracking
- At time n, do not re-estimate

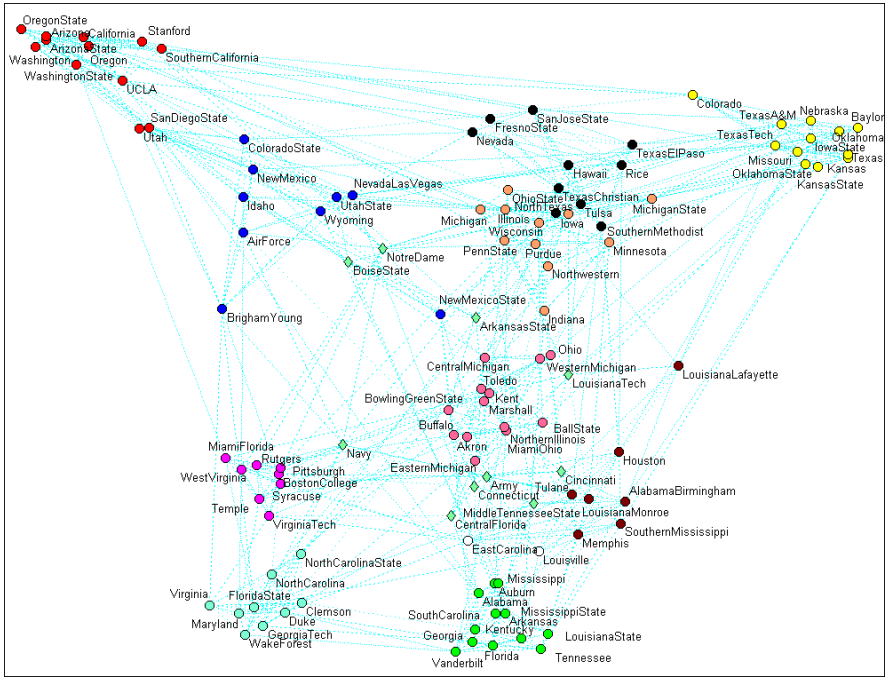
$$\mathbf{o}(n-1), \ldots, \mathbf{o}(1)$$

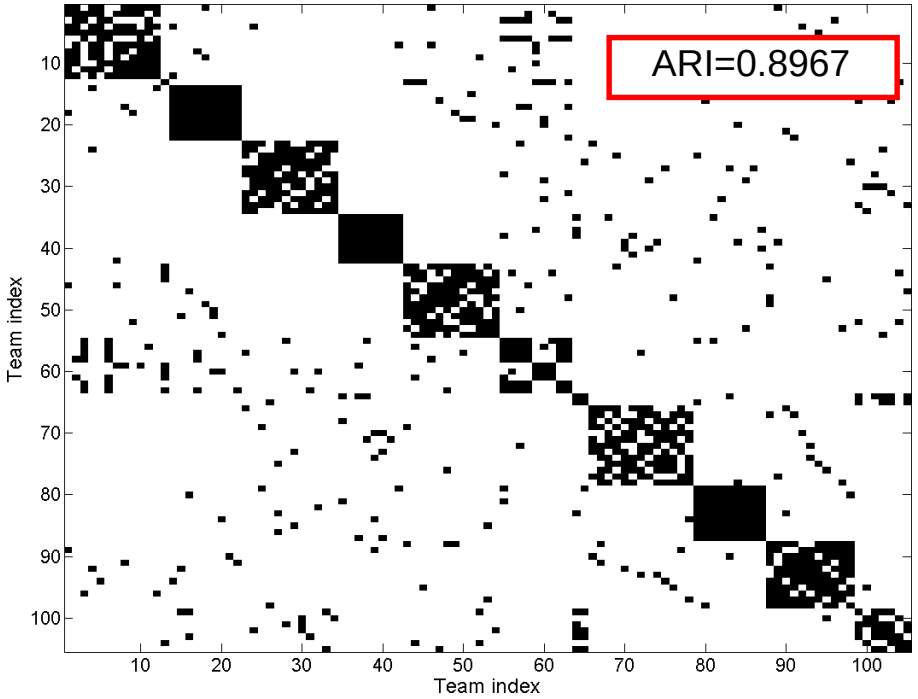


- Motivation: Real-time big data and memory limitations
- Nominal: $\mathbf{y}_n = \mathbf{C}\mathbf{w}_n + \boldsymbol{\varepsilon}_n, \quad n = 1, \ldots, 2000. \;\; p = 150, \;\; q = 5, \;\; \sigma_\varepsilon^2 = 10^{-3}.$
- Outliers: $\mathbf{y}_n \sim \mathcal{U}[-0.5, 0.5], \quad n = 1001, \ldots, 1005.$

Robust unveiling of communities

- Robust kernel PCA for identification of cohesive subgroups
- Network: NCAA football teams (vertices), Fall '00 games (edges)





- Identified exactly: Big 10, Big 12, ACC, SEC,…; Outliers: Independent teams

Data: http://www-personal.umich.edu/~mejn/netdata/

Roadmap

- Sparsity control for robust principal component analysis
  - Least-trimmed squares and outlier sparsity
  - Algorithm and applications

- Unveiling network anomalies via sparsity and low rank
  - Traffic modeling and identifiability
  - (De-) centralized and online algorithms
  - Numerical tests

- Conclusions and future research directions

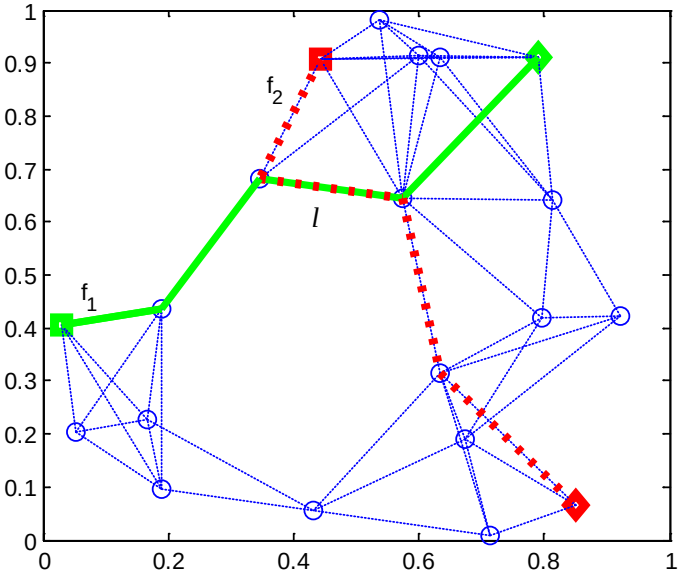
Modeling traffic anomalies

- Anomalies: changes in origin-destination (OD) flows [Lakhina et al'04]
  - Failures, congestions, DoS attacks, intrusions, flooding
- Graph $\mathcal{G}(N, L)$ with N nodes, L links, and F flows ($F >> L$); OD flow $z_{f,t}$
- Packet counts per link l and time slot t

Anomaly

$$y_{l,t} = \sum_{f=1}^{F} r_{l,f}(z_{f,t} + a_{f,t}) + v_{l,t}$$

$r_{l,f} \in \{0,1\}$



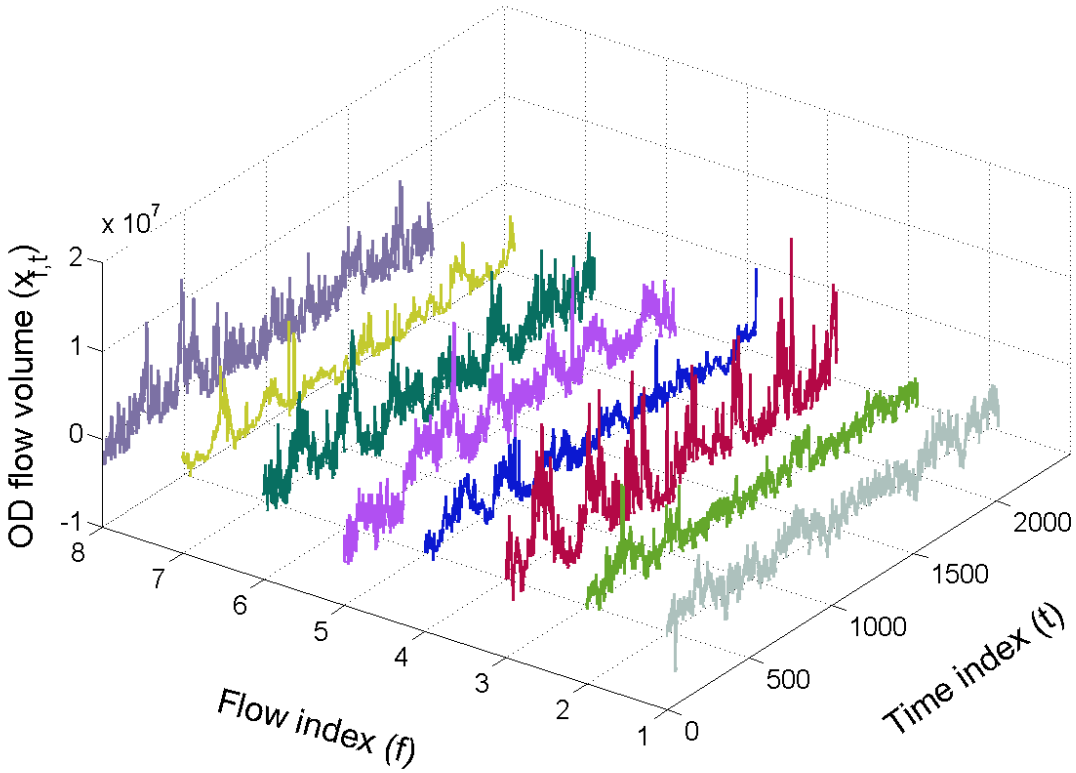
- Matrix model across T time slots: $\mathbf{Y} = \mathbf{R}(\mathbf{Z} + \mathbf{A}) + \mathbf{V}$

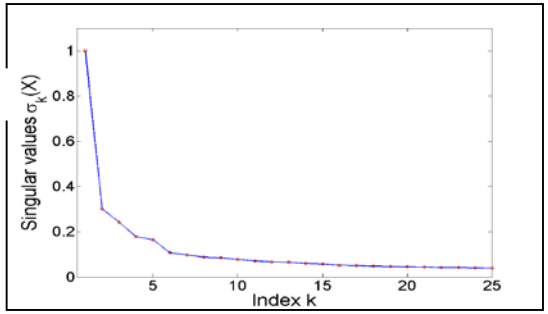
$\mathbf{Y}$: $L \times T$; $\mathbf{R}$: $L \times F$

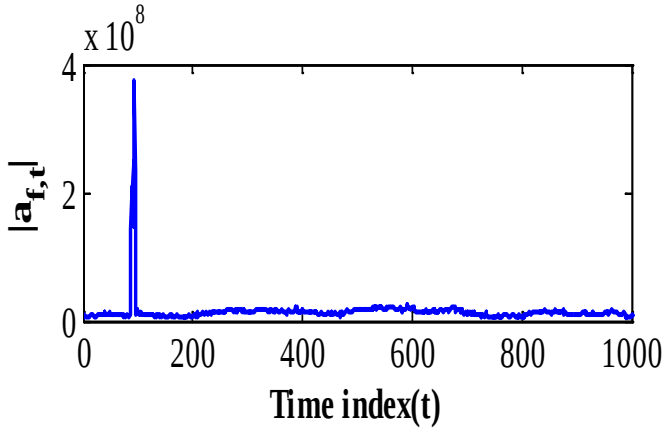
Low-rank plus sparse matrices

$$\mathbf{Y} = \mathbf{R}(\mathbf{Z} + \mathbf{A}) + \mathbf{V}$$

- **Z** has low rank, e.g., [Zhang et al'05]; **A** is sparse across time and flows







Data: http://math.bu.edu/people/kolaczyk/datasets.html

General decomposition problem

$$\mathbf{Y} = \underbrace{\mathbf{RZ}}_{:=\mathbf{X}} + \mathbf{RA} + \mathbf{V}$$

- Given $\mathbf{Y}$ and routing matrix $\mathbf{R}$, identify sparse $\mathbf{A}$ when $\mathbf{Z}$ is low rank
  - $\mathbf{R}$ fat but $\mathbf{X}$ still low rank

$$\{\hat{\mathbf{X}}, \hat{\mathbf{A}}\} = \arg \min_{\{\mathbf{X},\mathbf{A}\}} \frac{1}{2}\|\mathbf{Y} - \mathbf{X} - \mathbf{RA}\|_F^2 + \lambda_1 \|\mathbf{A}\|_1 + \lambda_* \|\mathbf{X}\|_* \qquad \text{(P1)}$$

$$\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$$

- Rank minimization with the nuclear norm, e.g., [Recht-Fazel-Parrilo'10]
  - Principal Comp. Pursuit (PCP) [Candes et al'10], [Chandrasekaran et al'11]

Challenges and importance

$$\mathbf{Y} = \mathbf{X} + \mathbf{RA} + \mathbf{V}$$

- $\mathbf{RA}$ not necessarily sparse and $\mathbf{R}$ fat $\Rightarrow$ PCP not applicable

- $\underbrace{LT}_{X} + \underbrace{FT}_{A} \quad >> \quad \underbrace{LT}_{Y}$

- Important special cases
  - ***R*** = ***I*** : matrix decomposition with PCP [Candes et al'10]
  - ***X*** = ***0*** : compressive sampling with basis pursuit [Chen et al'01]
  - ***X*** = $\boldsymbol{C}_{Lxp}\boldsymbol{W'}_{pxT}$ and ***A*** = ***0*** : PCA [Pearson 1901]
  - ***X*** = ***0, R*** = ***D*** unknown: dictionary learning [Olshausen'97]

Exact recovery

- Noise-free case

$$\mathbf{Y} = \mathbf{X}_0 + \mathbf{R}\mathbf{A}_0 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}' + \mathbf{R}\mathbf{A}_0$$
$$r = \text{rank}[\mathbf{X}_0], \;\; s = \|\mathbf{A}_0\|_0$$

$$\min_{\{\mathbf{X},\mathbf{A}\}} \|\mathbf{X}\|_* + \lambda\|\mathbf{A}\|_1 \quad \text{s.to } \mathbf{Y} = \mathbf{X} + \mathbf{R}\mathbf{A} \qquad \text{(P0)}$$

Q: Can one recover sparse $\mathbf{A}_0$ and low-rank $\mathbf{X}_0$ exactly?

A: Yes! Under certain conditions on $\{\mathbf{X}_0, \mathbf{A}_0, \mathbf{R}\}$

Theorem: Given $\mathbf{Y}$ and $\mathbf{R}$, assume every row and column of $\mathbf{A}_0$ has at most *k*<*s* non-zero entries, and $\mathbf{R}$ has full row rank. If C1)-C2) hold, then with $\lambda \in (\lambda_{\min}, \lambda_{\max})$ (P0) exactly recovers $\{\mathbf{X}_0, \mathbf{A}_0\}$

C1) $(1 - \mu(\Phi, \Omega_R))^2(1 - \delta_k(\mathbf{R})) > \alpha$

C2) $\lambda_{\min} := \beta\|\mathbf{R}'\mathbf{U}\mathbf{V}'\|_\infty < \lambda_{\max} := \sqrt{s^{-1}}[\gamma^{-1} - \mu(\Phi, \Omega_R)\sqrt{r(1 + \delta_k(\mathbf{R}))}]$

M. Mardani, G. Mateos, and G. B. Giannakis,``Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Information Theory*, 2013.

Numerical validation

- Setup

  L=105, F=210, T = 420

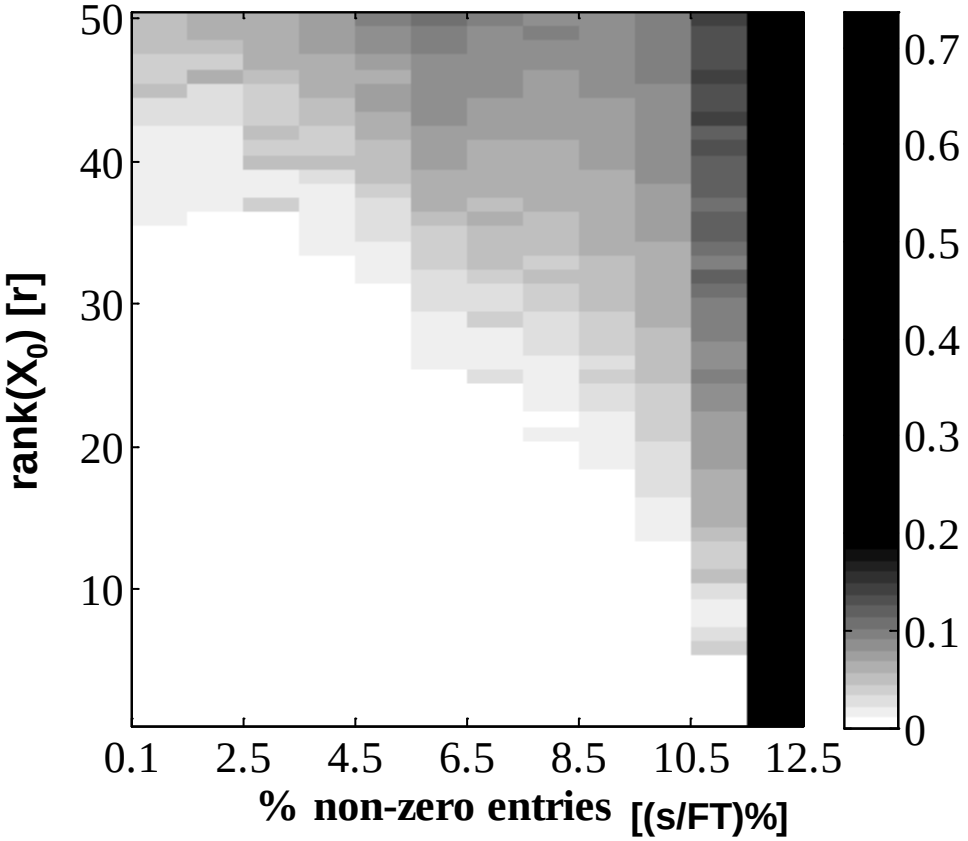
  R ~ Bernoulli(1/2)

  $\mathbf{X}_o$ = **RPQ**', **P**, **Q** ~ N(0, 1/FT)

  a_{ij} ϵ {-1,0,1} w.p. {π/*2*, *1*-π, π/*2*}

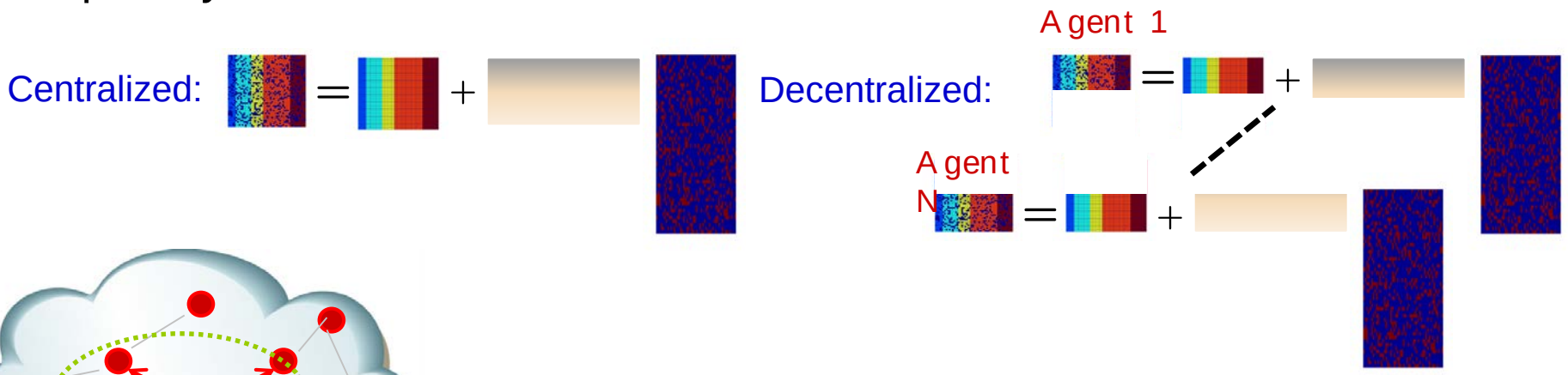
- Relative recovery error

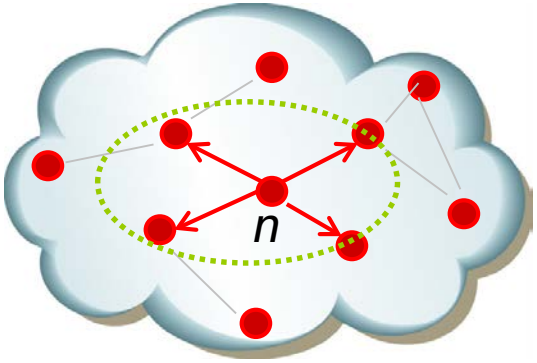
$$e = \frac{\|\hat{\mathbf{A}} - \mathbf{A}_0\|_F}{\|\mathbf{A}_0\|_F}$$



Decentralized in-network processing

- Spatially-distributed link count data





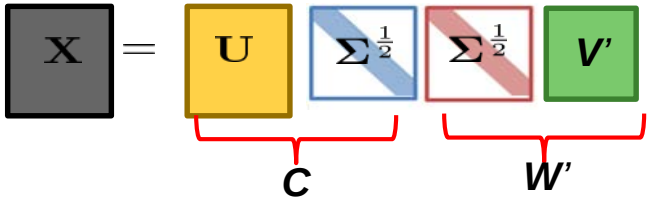
- Local processing and single-hop communications

Goal: Given local link counts per agent, unveil anomalies in a distributed fashion by leveraging low-rank of the nominal data matrix and sparsity of the outliers.

- Challenge: $\|\cdot\|_*$ not separable across rows (links/agents)

Separable regularization

- Key property



$$\|\mathbf{X}\|_* := \min_{\{\mathbf{C},\mathbf{W}\}} \frac{1}{2}\left\{\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2\right\}, \text{ s.to } \mathbf{X} = \mathbf{C}\mathbf{W}'$$

$\mathbf{C}$: $L \times \rho$, $\rho \geq \text{rank}[\mathbf{X}]$

- Separable formulation equivalent to (P1)

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{A}\}} \frac{1}{2}\|\mathbf{Y} - \mathbf{C}\mathbf{W}' - \mathbf{R}\mathbf{A}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \frac{\lambda_*}{2}\{\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2\} \quad \text{(P2)}$$

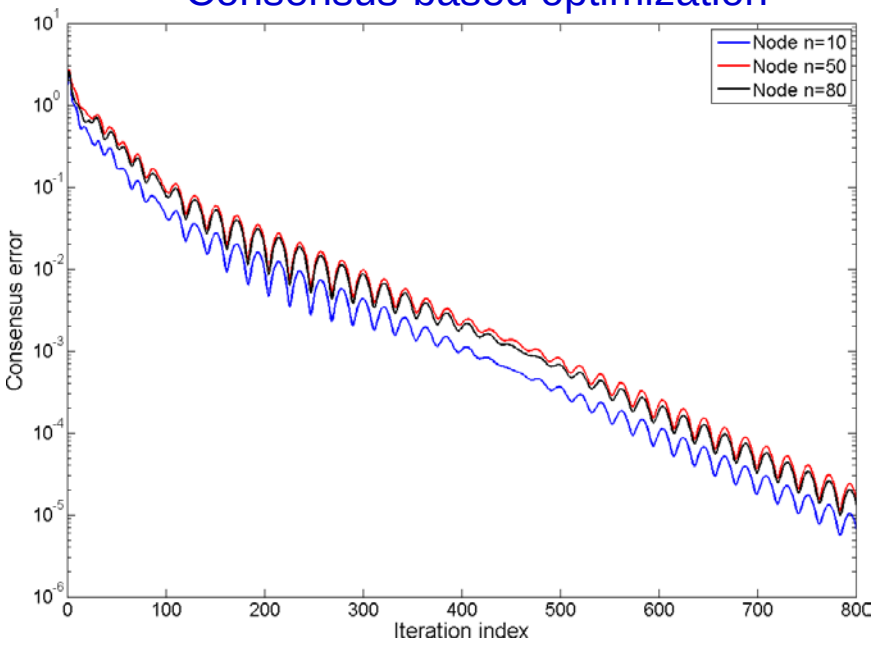
  - Nonconvex; less variables: $LT \Rightarrow \rho(L+T)$

Proposition: If $\{\bar{\mathbf{C}}, \bar{\mathbf{W}}, \bar{\mathbf{A}}\}$ stat. pt. of (P2) and $\|\mathbf{Y} - \bar{\mathbf{C}}\bar{\mathbf{W}}' - \mathbf{R}\bar{\mathbf{A}}\| \leq \lambda_*$, then $\{\hat{\mathbf{X}} := \bar{\mathbf{C}}\bar{\mathbf{W}}', \hat{\mathbf{A}} := \bar{\mathbf{A}}\}$ is a *global optimum* of (P1).

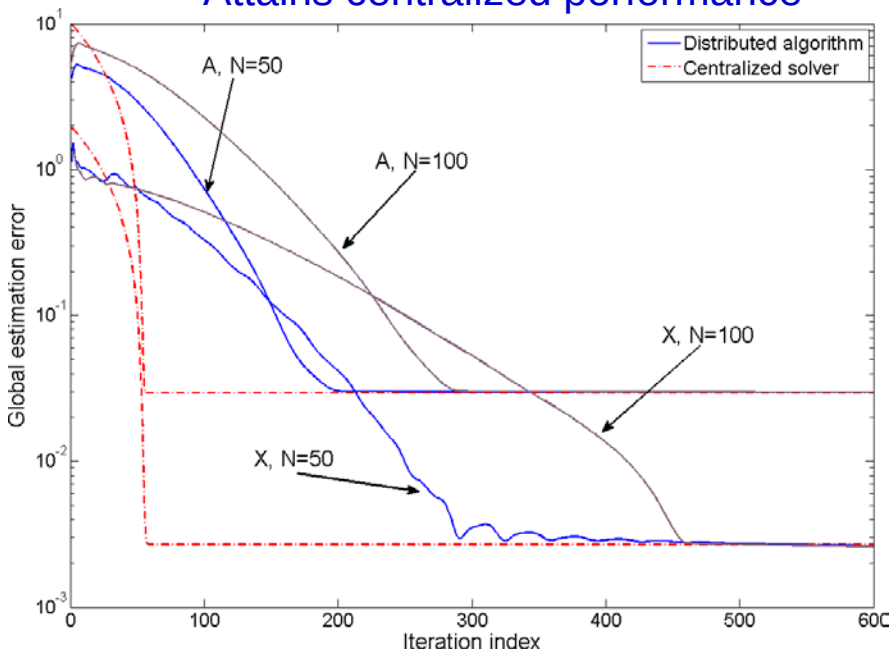
Distributed algorithm

- Alternating-direction method of multipliers (ADMM) solver for (P2)
  - Method [Glowinski-Marrocco'75], [Gabay-Mercier'76]
  - Learning over networks [Schizas-Ribeiro-Giannakis'07]

Consensus-based optimization



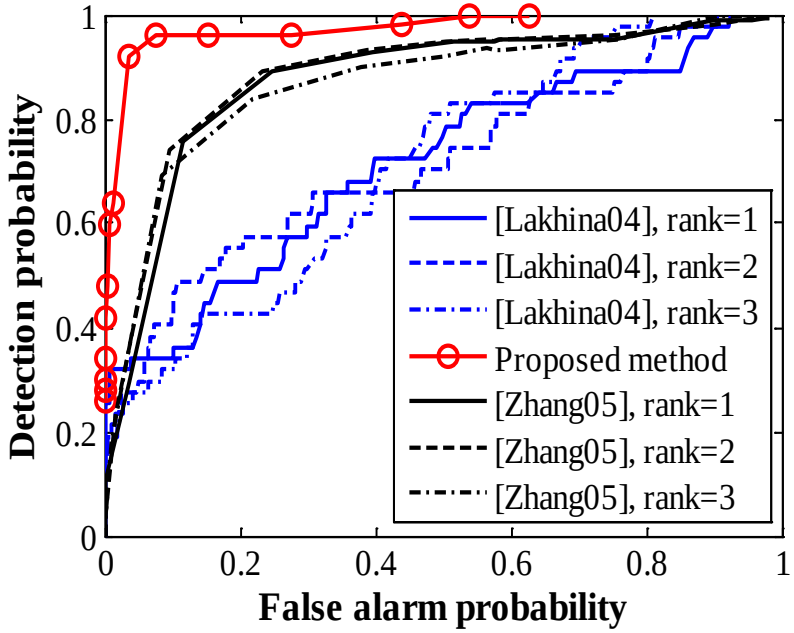
Attains centralized performance

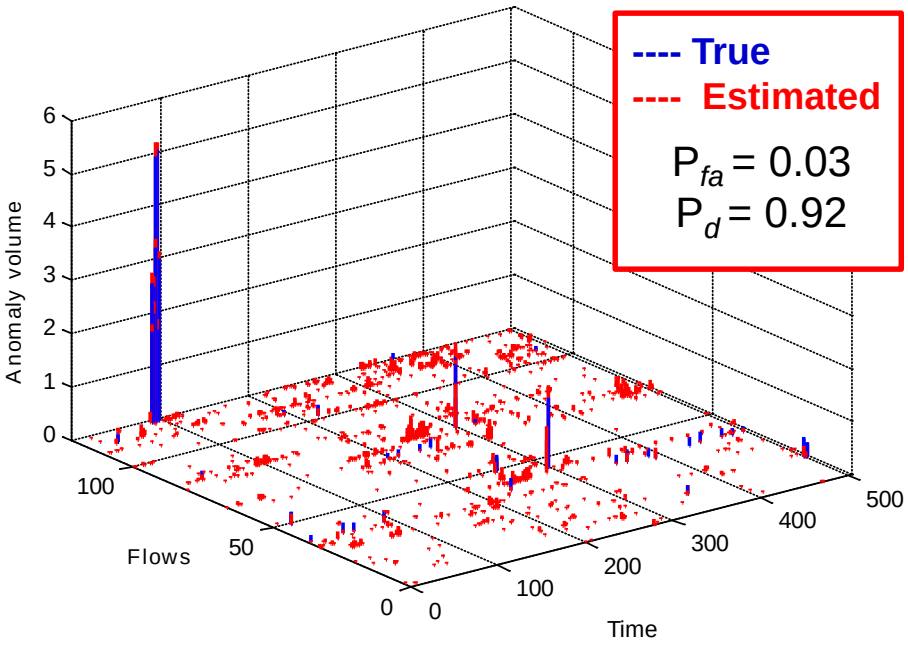


M. Mardani, G. Mateos, and G. B. Giannakis, "In-network sparsity regularized rank minimization: Algorithms and applications," *IEEE Transactions on Signal Processing*, Feb. 2013.

Internet2 data

- Real network data
  - Dec. 8-28, 2008
  - N=11, L=41, F=121, T=504





Data: http://www.cs.bu.edu/~crovella/links.html

Dynamic anomalography

- Construct an estimated map of anomalies in real time
- Streaming data model:

$$\mathcal{P}_{\Omega_t}(\mathbf{y}_t) = \mathcal{P}_{\Omega_t}(\mathbf{x}_t + \mathbf{R}_t\mathbf{a}_t + \mathbf{v}_t), \;\; t = 1, 2, \ldots \qquad \mathbf{x}_t := \mathbf{R}_t\mathbf{z}_t$$

Goal: Given $\{\mathcal{P}_{\Omega_i}(\mathbf{y}_i), \mathbf{R}_i\}_{i=1}^t$ estimate $(\mathbf{x}_t, \mathbf{a}_t)$ online when $\{\mathbf{x}_t\}$ is in a low-dimensional space and $\{\mathbf{a}_t\}$ is sparse

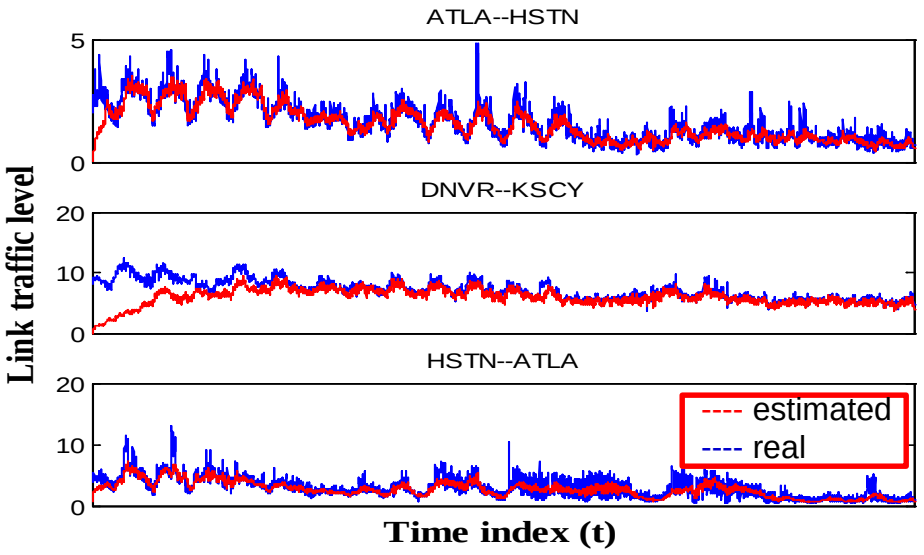
- (Robust) subspace tracking
  - Projection approximation (PAST) [Yang'95]
  - Missing data: GROUSE [Balzano et al'10], PETRELS [Chi et al'12]
  - Outliers: [Mateos-Giannakis'10], GRASTA [He et al'11]
- Compressed "outliers" challenge identifiability

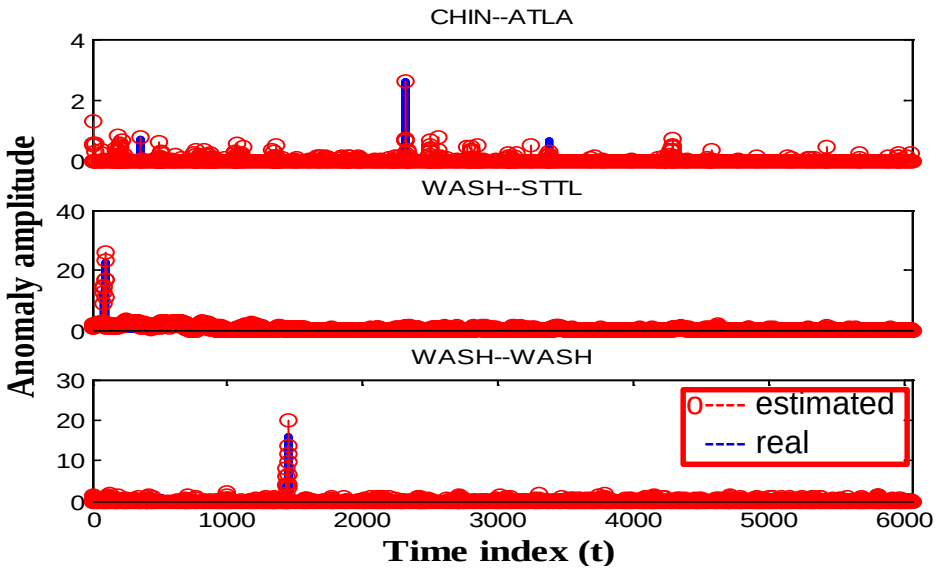
M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE Journal of Selected Topics in Signal Processing*, pp. 50-66, Feb. 2013.

Online estimator

- Challenge: $\|\cdot\|_*$ not separable across columns (time) $\Rightarrow \mathbf{x}_t = \mathbf{C}\mathbf{w}_t$
- Approach: regularized exponentially-weighted LS formulation

$$\min_{\{\mathbf{C},\mathbf{W},\mathbf{A}\}} \sum_{\tau=1}^{t} \beta^{t-\tau} \left[\frac{1}{2}\|\mathcal{P}_{\Omega_\tau}(\mathbf{y}_\tau - \mathbf{C}\mathbf{w}_\tau - \mathbf{R}_\tau\mathbf{a}_\tau)\|_2^2 + \frac{\lambda_*}{2\sum_{u=1}^{t}\beta^{t-u}}\|\mathbf{C}\|_F^2 + \frac{\lambda_*}{2}\|\mathbf{w}_\tau\|_2^2 + \lambda_1\|\mathbf{a}_\tau\|_1 \right]$$



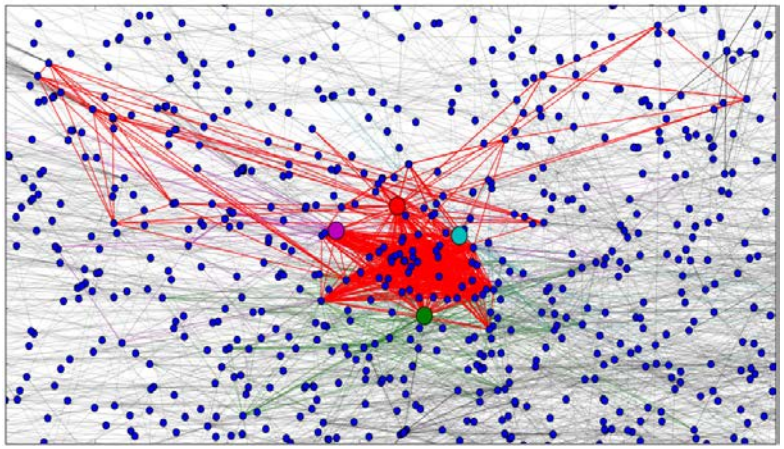


Current research directions

Broadening the scope

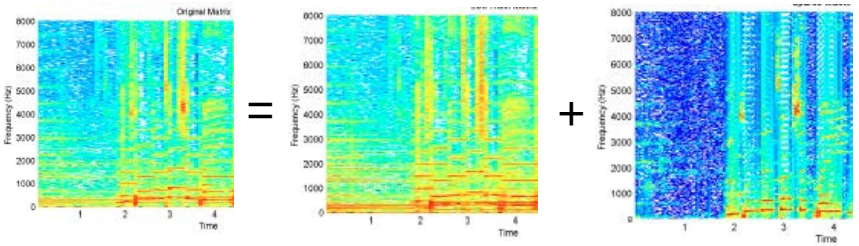
Approach: graph data, decompose the egonet feature matrix using PCP

arXiv Collaboration Network (General Relativity)



Broadening the model $\mathbf{Y} = \mathbf{X} + \mathbf{D}\mathbf{A}$

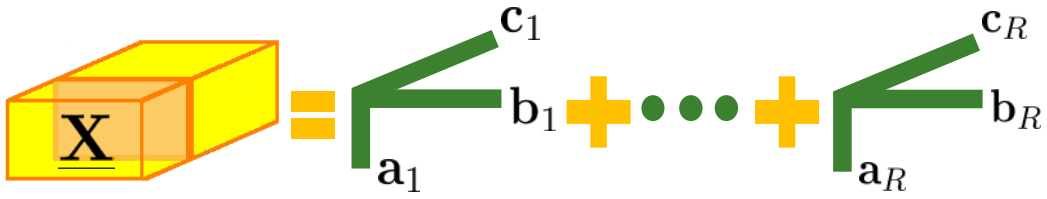
- Singing voice separation



- Face recognition

Inference for big tensor data

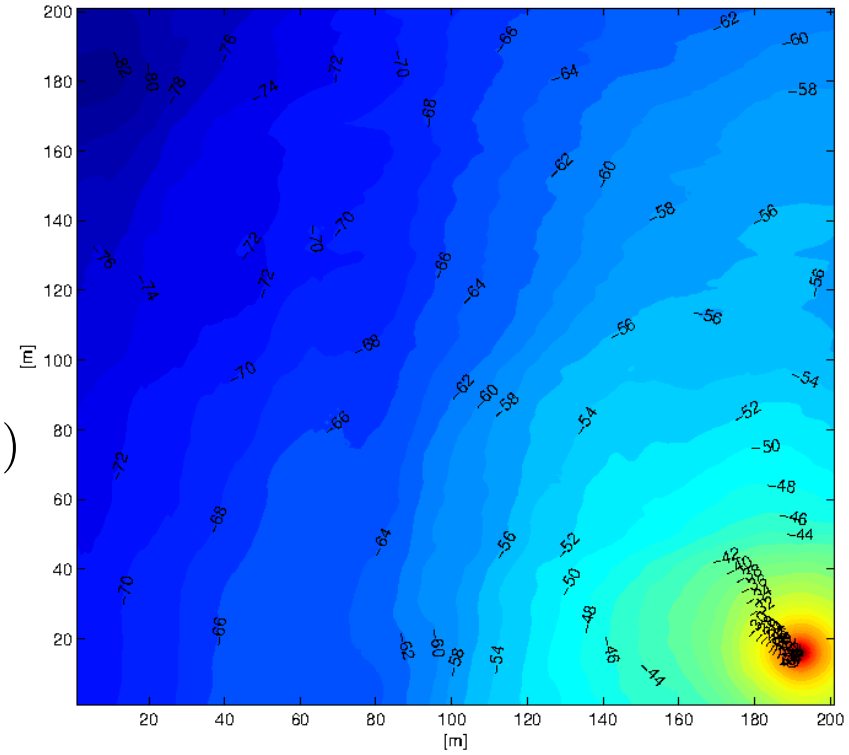
- Dynamic networks
- Forecasting of loads, renewables



Any-to-any channel acquisition

- Shadowing model-free approach
  - Slow variations in shadow fading
  - Low-rank any-to-any CG matrix $\hat{\mathbf{G}}$

Estimated channel gain (CG) map



Approach: low-rank matrix completion

$$\min_{\mathbf{C},\mathbf{W}} \|\mathcal{P}_{\mathcal{S}}(\mathbf{G} - \mathbf{C}\mathbf{W}')\|_F^2 + \lambda(\|\mathbf{C}\|_F^2 + \|\mathbf{W}\|_F^2)$$

Payoffs: global view of any-to-any CGs;
real-time propagation metrics;
efficient resource allocation

Outlook: kernel-based extrapolator for missing CR-to-PU measurements, or future time intervals

S.-J. Kim and G. B. Giannakis, "Dynamic Network Learning for Cognitive Radio Spectrum Sensing," *Proc. of Intl. Workshop on Comp. Advances in Multi-Sensor Adaptive Process.*, Saint Martin, 2013.

Concluding summary

- Robust social data analytics and networking
  - Leveraging sparsity and low rank
- Robust principal components analysis
  - Control sparsity in model residuals for robust learning
  - Identifying invalid survey protocols and communities
- Unveiling network anomalies via convex optimization
  - Reveal when and where anomalies occur
  - Exact recovery of low-rank plus compressed sparse matrices
  - Distributed/online algorithms with guaranteed performance

Thank you!