GEOMETRIC GRAPHICAL MODELS IN MACHINE LEARNING

UCLA

Andrea Bertozzi

University of California, Los Angeles

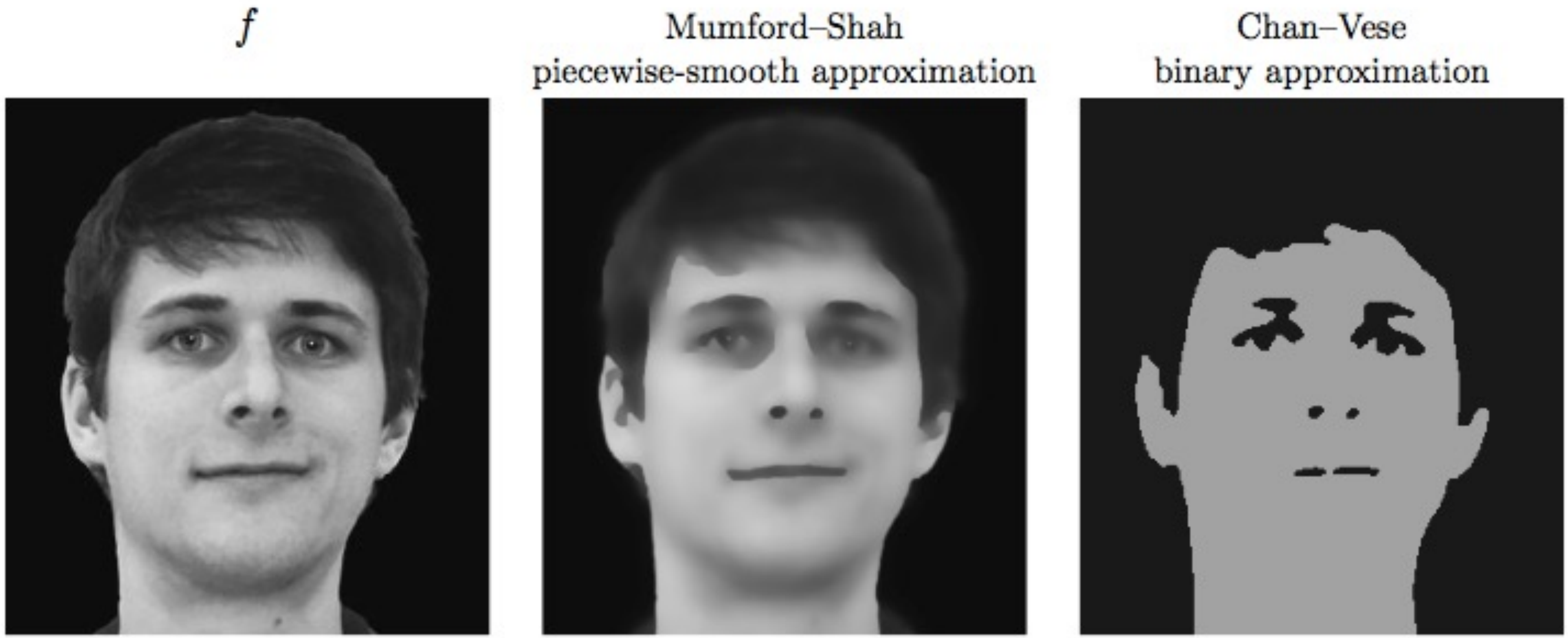
Thanks to DARPA, NSF, ONR, AFOSR, and NIJ for support.

FORMER STUDENTS AND POSTDOCS

- Tijana Kostic (Microsoft),
- Cristina Garcia (LANL),
- Justin Sunu (CGU),
- Huiyi Hu (Google) ,
- Ekaterina Murkerjev (Michigan State U),
- Zhaoyi Meng (Google),
- Joseph Woodworth (Google)
- Y. van Gennip (Univ. Delft),
- B. Osting (Univ. Utah),
- N. Guillen (U Mass Amherst)
- X. Luo (Google)
- Z. Boyd (UNC Chapel Hill)
- G. Iyer (Google)

Motivation from Image Segmentation……..

An Example (from IPOL P. Getreuer 2012)



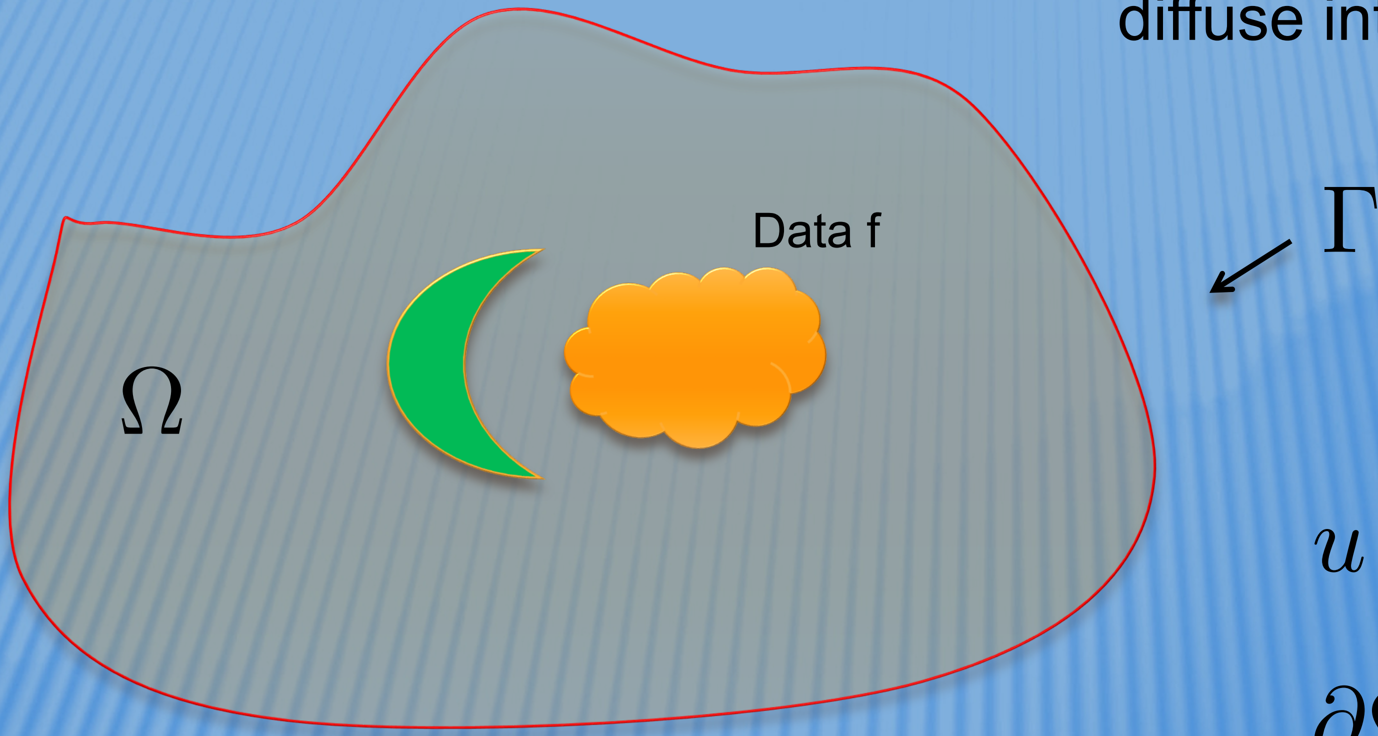
Mumford-Shah

$$E(u,\Gamma) = \int_{R^2} (u-f)^2 dx + \mu \int_{R^2-\Gamma} |\nabla u|^2 dx + \nu|\Gamma|$$

Chan-Vese

$$E(C_1,C_2,\Gamma) = \int_{\Gamma_{in}} (f-C_1)^2 + \int_{\Gamma_{out}} (f-C_2)^2 + \nu|\Gamma|$$

Total Variation, isoperimetric problems, and diffuse interfaces



$$u = \chi_\Omega$$

$$\partial\Omega = \Gamma$$

$$|\Gamma| = \int |\nabla u| \equiv |u|_{TV}$$

$$\sim \frac{\epsilon}{2}\int |\nabla u|^2 + \frac{1}{\epsilon}\int W(u)dx \equiv GL_\epsilon(u)$$

Ginzburg-Landau functional

DIFFUSE INTERFACE EQUATIONS AND THEIR SHARP INTERFACE LIMIT

Gradient descent of GL function: $$u_t = \epsilon \Delta u - \frac{1}{\epsilon} W'(u)$$

Allen-Cahn equation. Famous in materials science. Now useful for data science.

$\epsilon \to 0$ Motion by Mean Curvature

MBO SCHEME (1992)

Merriman, Bence, Osher

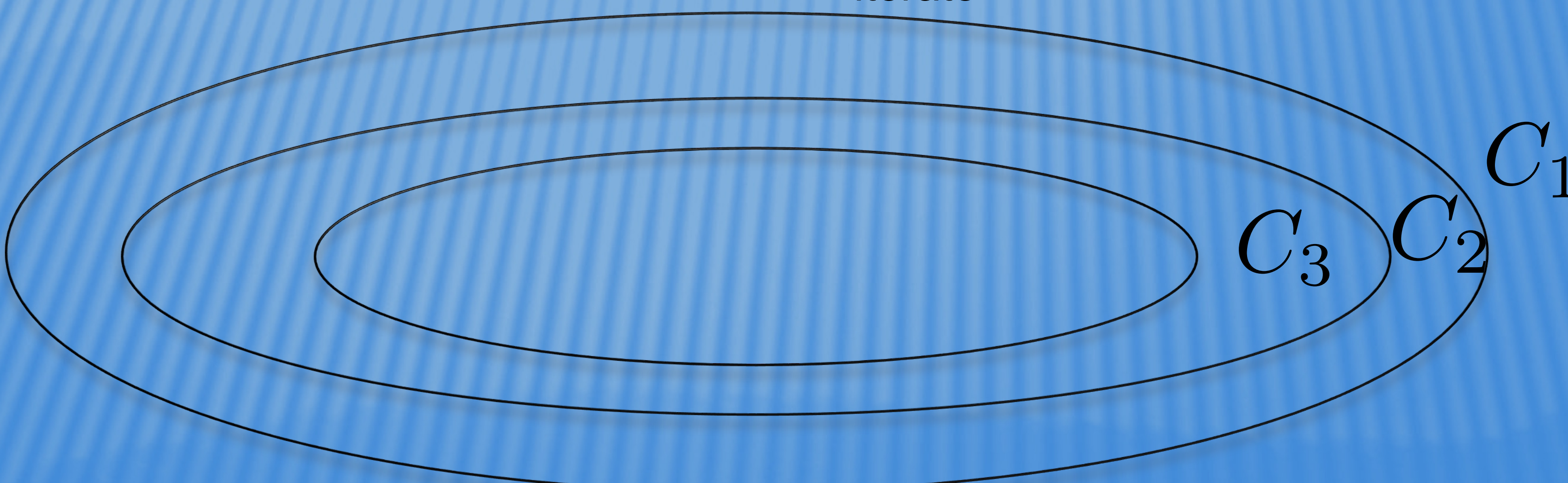
$$\begin{cases} u_t - \Delta u = 0 & \text{in } (0,+\infty)\times\mathbb{R}^N, \\ u(0,x) = \begin{cases} 1, & x \in C_0, \\ -1, & x\in \mathbb{R}^N \setminus C_0. \end{cases} \end{cases}$$

$$C_1 = \{x \in \mathbb{R}^N \mid u(h,x) \geq 0\}.$$

Threshold

Heat equation

iterate



Extended to Piecewise Constant Mumford-Shah Model by Esedoglu-Tsai 2006

FROM
EUCLIDEAN SPACE TO SIMILARITY GRAPHS
FOR LARGE DATA

- Minimal surface problem
- Laplace operator
- Pseudo-spectral methods
- Fast Fourier Transform
- Uses all the modes

- Graph mincut problem
- Graph Laplacian
- Projection to eigensubspace of graph Laplacian
- Nystrom extension/Rayleigh-Chebyshev
- Often only needs a small percentage of spectral modes.

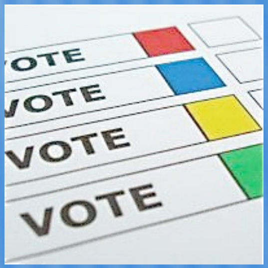
WEIGHTED GRAPHS FOR "BIG DATA"

$$w(x, y) = \exp(-||x - y||^2/\tau)$$

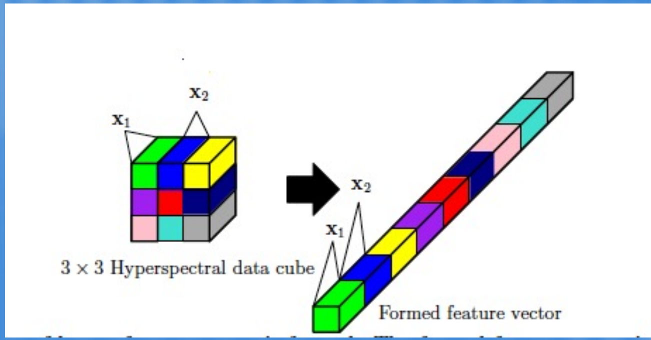
In a typical application we have data supported on the graph, possibly high dimensional. The above weights represent comparison of the data.

Examples include:

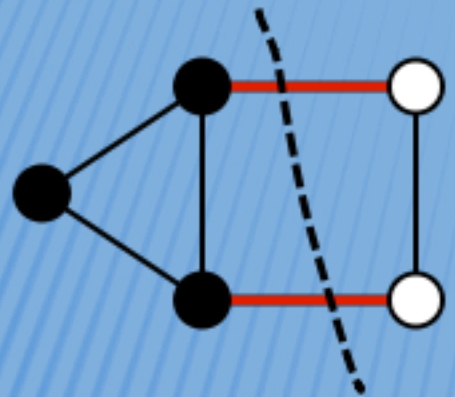
 voting records of US Congress – each person has a vote vector associated with them.



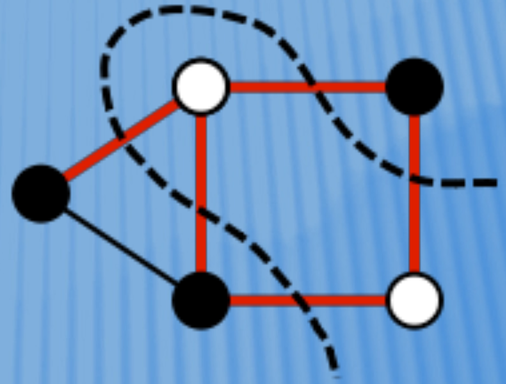
Nonlocal means image processing – each pixel has a pixel neighborhood that can be compared with nearby and far away pixels.



GRAPH CUTS AND TOTAL VARIATION



Minimum cut



Maximum cut

Total Variation of function *f* defined on nodes of a weighted graph:

$$\sum \omega_{ij} |f_i - f_j|$$

Min cut problems can be reformulated as a total variation minimization problem for binary/multivalued functions defined on the nodes of the graph.

DIFFUSE INTERFACE METHODS ON GRAPHS

Bertozzi and Flenner MMS 2012. SIGEST 2016

Arjuna Flenner
GE research

$$L(\nu,\mu)=\begin{cases} d(\nu) & \text{if } \nu=\mu, \\ -w(\nu,\mu) & \text{otherwise.} \end{cases}$$

$$\langle u, Lu\rangle = \frac{1}{2}\sum_{\mu,\nu\in V} w(\nu,\mu)(u(\nu)-u(\mu))^2$$

$$L_s = D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}.$$

$$\boxed{E(u)=\frac{\epsilon}{2}\langle u, L_s u\rangle + \frac{1}{4\epsilon}\sum_{z\in Z}(u^2(z)-1)^2+\sum_{z\in Z}\frac{\lambda(z)}{2}(u(z)-u_0(z))^2.}$$

CONVERGENCE OF GRAPH GL FUNCTIONAL

van Gennip and ALB Adv. Diff. Eq. 2012

$$f_\varepsilon(u) := \chi \sum_{i,j=1}^{m} \omega_{ij}(u_i - u_j)^2 + \frac{1}{\varepsilon}\sum_{i=1}^{m} W(u_i),$$

$$\frac{1}{2}\|\nabla u\|_{\mathcal{E}}^2 = \frac{1}{4}\sum_{i,j\in I_m} \omega_{ij}(u_i - u_j)^2.$$

Yves
Van Gennip
TU Delft

Theorem 3.1 (Γ-convergence). *$f_\varepsilon \xrightarrow{\Gamma} f_0$ as $\varepsilon \to 0$, where*

$$f_0(u) := \begin{cases} \chi \sum_{i,j\in I_m} \omega_{ij}|u_i - u_j| & \text{if } u \in \mathcal{V}^b, \\ +\infty & \text{otherwise} \end{cases} = \begin{cases} 2\chi TV_{a1}(u) & \text{if } u \in \mathcal{V}^b, \\ +\infty & \text{otherwise.} \end{cases}$$

REMOVE THE DIFFUSE INTERFACE: MBO SCHEME ON GRAPHS

Merkurjev, Kostic, and ALB, SIIMS 2013

- 1) propagation by graph heat equation + forcing term

$$\frac{\partial z}{\partial t} = -L_s z - C_1 \lambda(x)(z - z_0)$$

- 2) thresholding

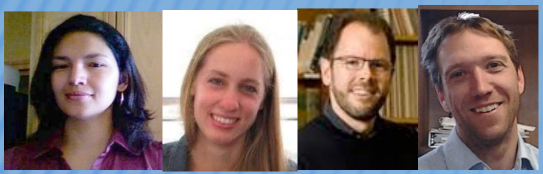
$$u^{n+1}(x) = \begin{cases} 1, & \text{if } y(x) \geq 0 \\ -1, & \text{if } y(x) < 0 \end{cases}$$

- Simple! And often converges in just a few iterations (e.g. 4 for MNIST dataset)

ALGORITHM

- I) Create a graph from the data, choose a weight function and then create the symmetric graph Laplacian.
- II) Calculate the eigenvectors and eigenvalues of the symmetric graph Laplacian. *It is only necessary to calculate a portion of the eigenvectors*.*
- III) Initialize u.
- IV) Iterate the two-step scheme described above until a stopping criterion is satisfied.
- *Fast linear algebra routines are necessary – either Raleigh-Chebyshev procedure or Nystrom extension.

GENERALIZATION MULTICLASS MACHINE LEARNING PROBLEMS (MBO)



Garcia, Merkurjev, Bertozzi, Percus, Flenner, *IEEE TPAMI, 2014*

Semi-supervised learning

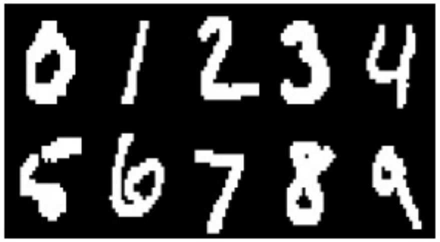


Fig. 4: Examples of digits from the MNIST data base

$$E(u) = \frac{\epsilon}{2}\langle \boldsymbol{u}, \boldsymbol{L_s u}\rangle + \frac{1}{2\epsilon}\sum_{i\in V}\prod_{j=1}^{K}\frac{1}{4}\|\vec{u}_i - \vec{s}_j\|_{L_1}^2 + \sum_{i\in V}\frac{\lambda_i}{2}\left\|\vec{u}_i - \vec{u}_i^{\,0}\right\|^2, \qquad (13)$$

where

$$\boldsymbol{u} = \begin{bmatrix} \vec{u}_1 \\ \dots \\ \vec{u}_{N_D} \end{bmatrix} \text{ with } \vec{u}_i = [(u_i)_1, \dots, (u_i)_K]$$

$$\langle \boldsymbol{u}, \boldsymbol{L_s u}\rangle = \mathrm{trace}(\boldsymbol{u}^T \boldsymbol{L_s u})$$

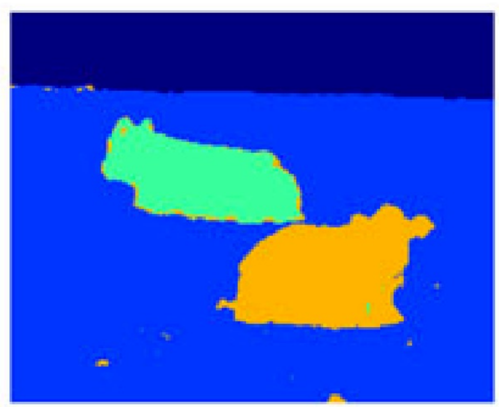
$$\|\vec{u}_i - \vec{s}_j\|_{L_1} = \sum_{m=1}^{K} |(u_i)_m - \delta_{jm}|$$

Instead of double well we have N-class well with Minima on a simplex in N-dimensions

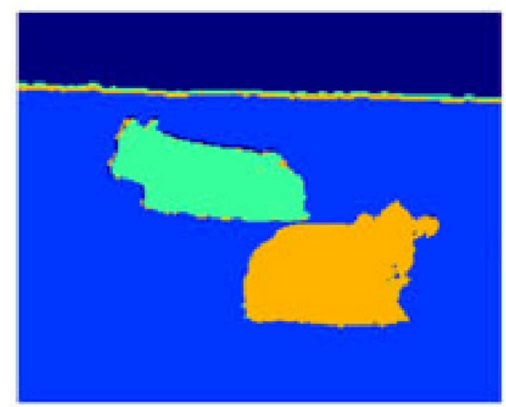
IMAGE LABELLING

(a) Original Image

(b) Labeled Data



(b) Multiclass GL



(c) Multiclass MBO

MBO SCHEME ON GRAPHS - MULTICLASS

1. Heat equation with forcing term:

$$\frac{\mathbf{U}^{n+\frac{1}{2}} - \mathbf{U}^n}{dt} = -\mathbf{L_s}\mathbf{U}^n - \boldsymbol{\mu}(\mathbf{U}^n - \hat{\mathbf{U}}). \qquad (25)$$

2. Thresholding:

$$\mathbf{u}_i^{n+1} = \boldsymbol{e}_k, \qquad (26)$$

$$\mathbf{U}^{n+\frac{1}{2}} = \mathbf{B}^{-1}\left[\mathbf{U}^n - dt\,\boldsymbol{\mu}(\mathbf{U}^n - \hat{\mathbf{U}})\right], \qquad (27)$$

where

$$\mathbf{B} = \mathbf{I} + dt\,\mathbf{L_s}. \qquad (28)$$

As before, we use the eigendecomposition $\mathbf{L}_s = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T$ to write

$$\mathbf{B} = \mathbf{X}(\mathbf{I} + dt\,\boldsymbol{\Lambda})\mathbf{X}^T, \qquad (29)$$

which we approximate using the first N_e eigenfunctions. The initialization procedure and the stopping criterion are the same as in the previous section.

MNIST DATABASE

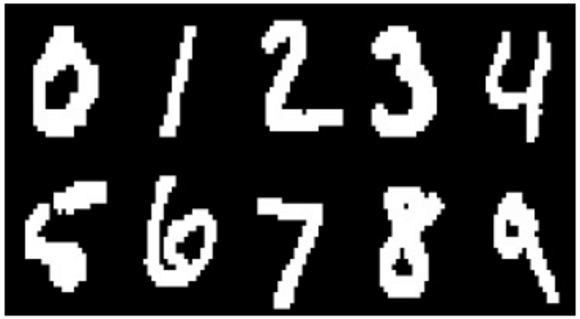


Fig. 6: Examples of digits from the MNIST data base

MNIST

Method	Accuracy
p-Laplacian	87.1%
multicut norm. 1-cut	87.64%
Cheeger cut	88.2%
linear classifiers	88%
nonlinear classifiers	96.4%-96.7%
k-NN	95.0%- 97.17%
boosted stumps	92.3%- 98.74%
neutral/convolution nets	95.3-99.65%
SVM	98.6%-99.32%
multiclass GL	96.8%
MBO reduction	96.91%

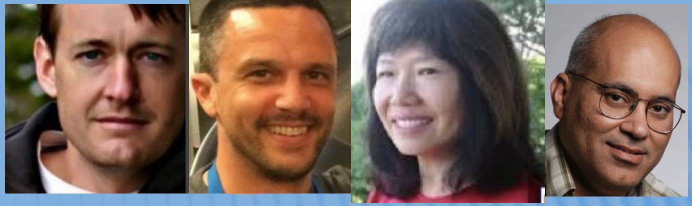
We use local rescaled graph as in Zelnik-Manor&Perona

Comparisons
Semi-supervised learning
Vs Supervised learning

We do semi-supervised with only 3.6% of the digits as the Known data.

Supervised uses 60000 digits for training and tests on 10000 digits.

NYSTROM EXTENSION

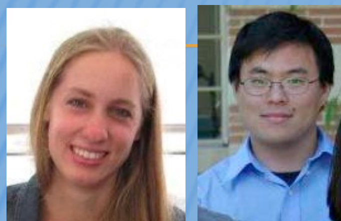


Fowlkes Belongie Chung and Malik, IEEE T. PAMI 2004.

$$W = \begin{pmatrix} W_{XX} & W_{XY} \\ W_{YX} & W_{YY} \end{pmatrix}, \quad W \sim \begin{pmatrix} W_{XX} \\ W_{YX} \end{pmatrix} W_{XX}^{-1} \begin{pmatrix} W_{XX} & W_{XY} \end{pmatrix}.$$

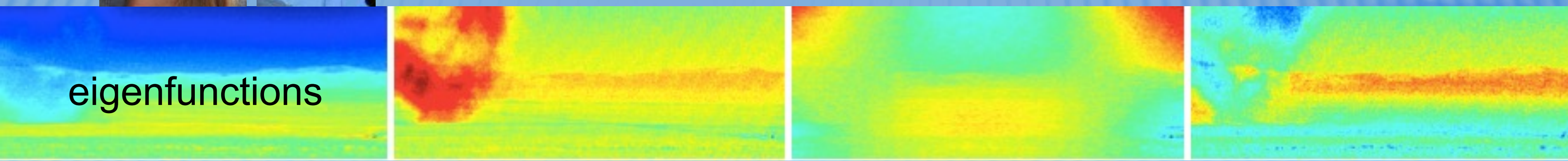
Computing W_{XX}, $W_{XY} = W_{YX}^T$ requires only $(|X| \cdot (|X| + |Y|)$ computations versus $(|X| + |Y|)^2$ for the whole similarity matrix. The method approximates W_{YY} by $W_{YX} W_{XX}^{-1} W_{XY}$ and the error is determined by how much the rows of W_{XY} span the rows of W_{YY}.

HYPERSPECTRAL VIDEO SEGMENTATION – SEMI SUPERVISED

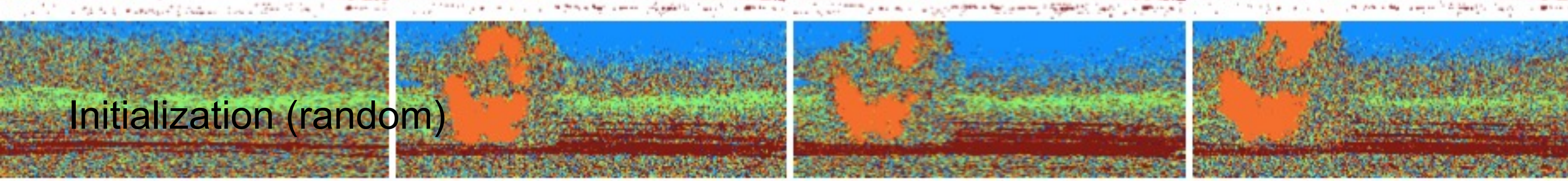


Merkurjev, Sunu, and Bertozzi, 2014, ICIP Paris 2014

Eigenfunctions computed using Nystrom



"ground truth obtained from thresholding eigenfunctions; random initialization otherwise



Four class hyperspectral pixel segmentation of gas plume, ground, mountain, and sky

THEORETICAL CONNECTION BETWEEN MBO AND GRAPH TV

3.1 Mumford-Shah MBO scheme

We first introduce a "diffuse operator" $\Gamma_\tau = e^{-\tau \mathbf{L}}$, where $\mathbf{L}$ is the graph Laplacian defined above and τ is a time step size. The operator Γ_τ is analogous to the diffuse operator $e^{-\tau \Delta}$ of the heat equation in PDE (continuous space). It satisfies the following properties.

Proposition 1. *Firstly, Γ_τ is strictly positive definite, i.e. $\langle f, \Gamma_\tau f \rangle > 0$ for any $f \in \mathbb{K}$, $f \neq 0$. Secondly, Γ_τ conserves the mass, i.e. $\langle 1, \Gamma_\tau f \rangle = \langle 1, f \rangle$. At last, the quantity $\frac{1}{2\tau}\langle 1 - f, \Gamma_\tau f \rangle$ approximates $\frac{1}{2}|f|_{TV}$, for any $f \in \mathbb{B}$.*

Use cosine angle for graph weights

$$w_{ij} = exp\{-\frac{(1 - \frac{\langle v_i, v_j \rangle}{\|v_i\| \|v_j\|})^2}{2\sigma^2}\}$$

MUMFORD-SHAH GRAPH MBO SCHEME

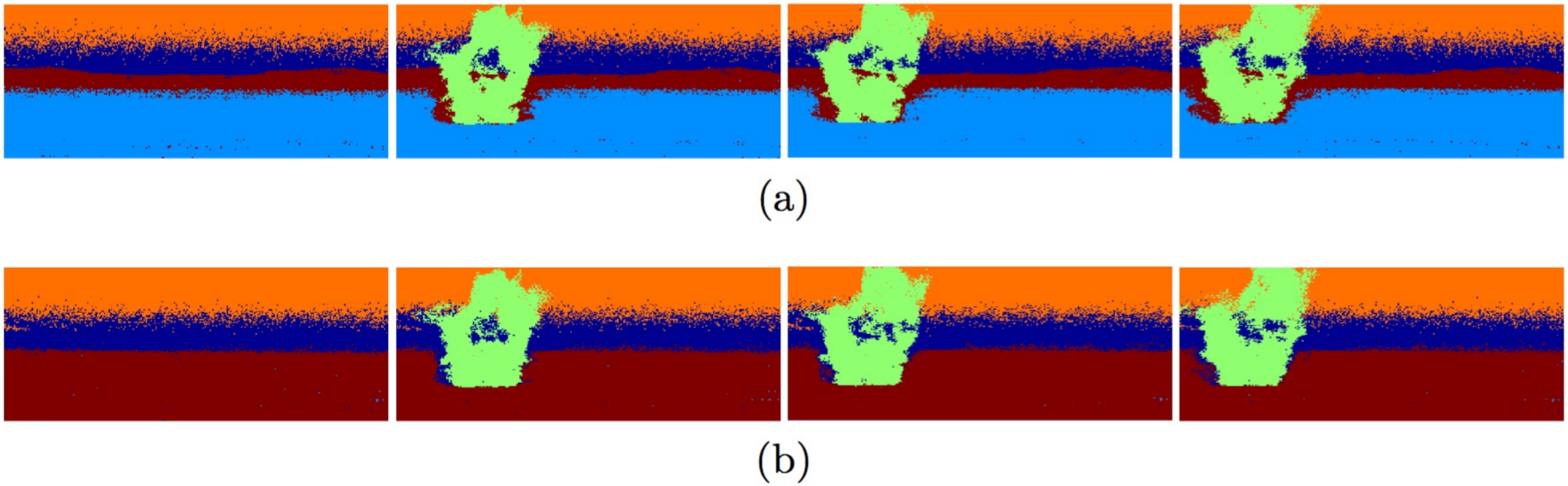
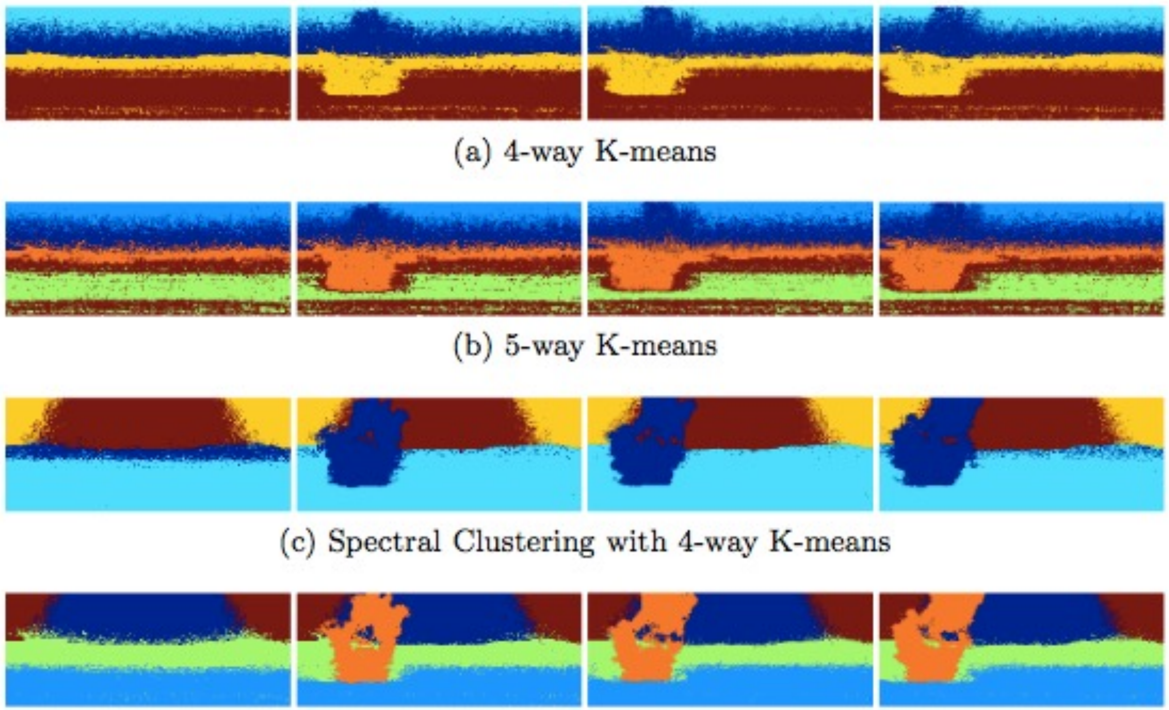


Fig. 2: The segmentation results obtained by the Mumford-Shah MBO scheme, on a background frame plus the frames 72-77. Shown in (a) and (b) are segmentation outcomes obtained with different initializations. The visualization of the segmentations only includes the first four frames.

RECALL COMPARISON TO KMEANS AND SPECTRAL CLUSTERING - UNSUPERVISED

EMMCVPR 2015 Hu, Sunu, and ALB

K-means
And
Spectral
Clustering



(a) 4-way K-means

(b) 5-way K-means

(c) Spectral Clustering with 4-way K-means

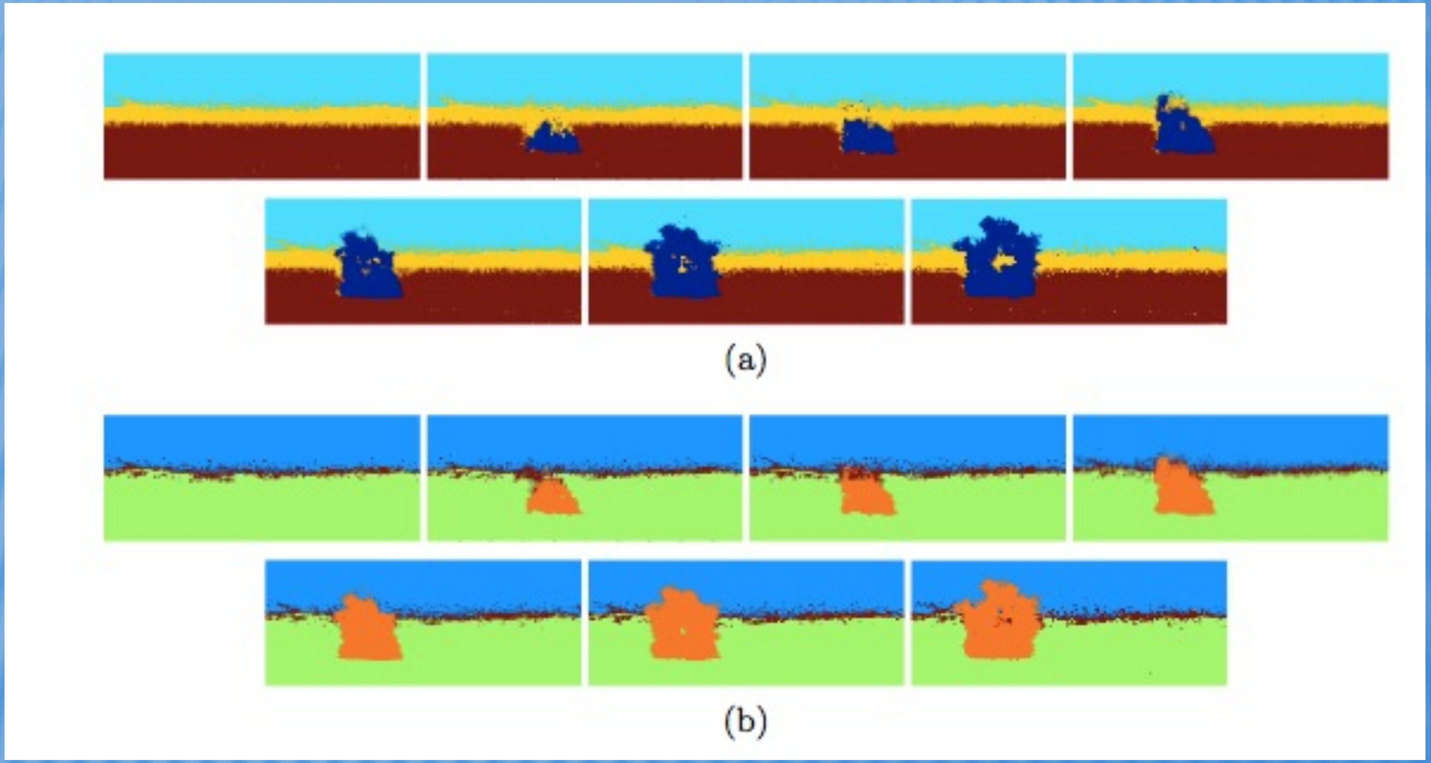
(d) Spectral Clustering with 5-way K-means

C-V SEGMENTATION ON GRAPHS USING MBO SCHEME FOR UNSUPERVISED CLUSTERING OF HYPERSPECTRAL PIXELS

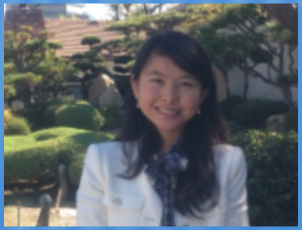
Multiclass MBO with different Initializations.

7 video frames
280K pixels

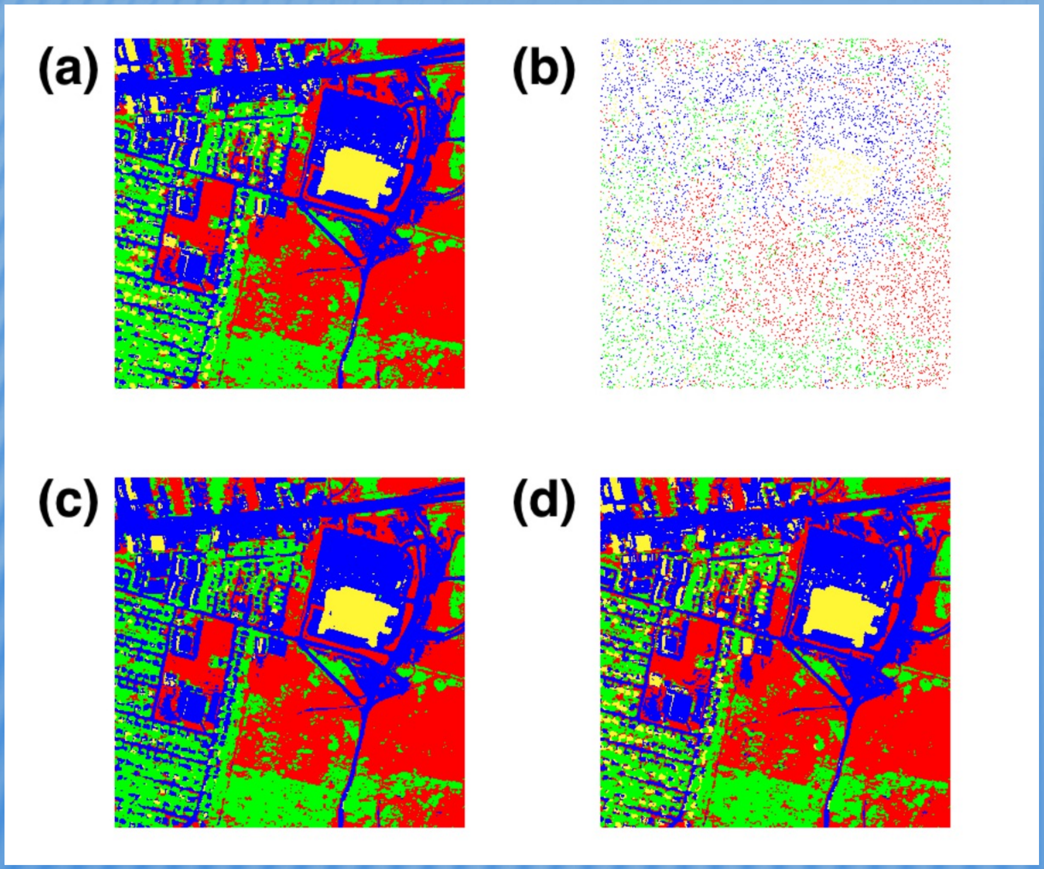
Each pixel is 128 dimensions

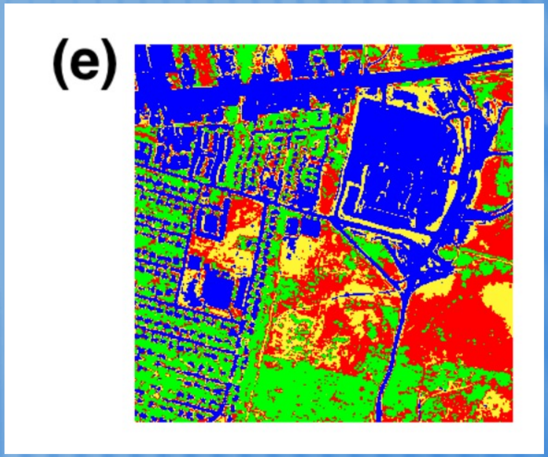


EMMCVPR 2015 Hu, Sunu, and ALB



FOUR CLASS "URBAN" CLASSIFICATION





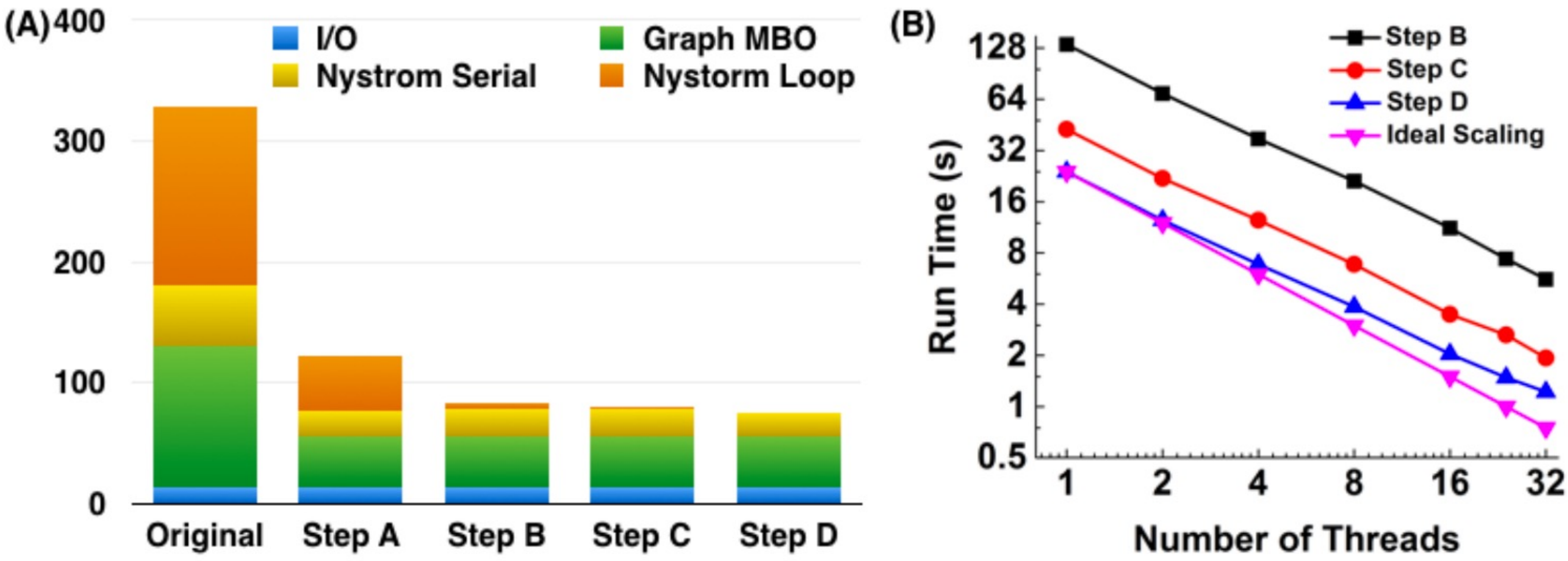
(a) Ground Truth
(b) Training data (10% random)
(c) Semi-supervised graph cut
(d) Unsupervised graph cut
(e) Spectral clustering

Zhaoyi Meng, Ekaterina Merkurjev, Alice Koniges, Andrea L Bertozzi, Hyperspectral Video Analysis Using Graph Clustering Methods, IPOL 2017.

Ground truth from http://www.escience.cn/people/feiyunZHU/Dataset GT.html

PARALLELIZATION – EXASCALE READY PLATFORM CORI AT LBNL NERSC ~ 300 HYPERSPECTRAL VIDEO FRAMES ~ 13M PIXELS

Meng et al, IWOMP 2016.



(A): The run time of different optimization steps. Step A: parallelizing the inner j-loop and BLAS3 optimization on Graph MBO. Step B: parallelizing the outer j-loop. Step C: normalizing and forming all Zis to Xmat. Step D: using uniform sampling and chunked Y matrices. (B): The scaling results of the OpenMP parallelization of the Nystr¨om loop. The black line with squares, the red line with circles and the blue line with triangles show the scaling results of step B, C and D respectively. The pink line with upside down triangles shows the ideal scaling.

MODIFIED CHEEGER CUT AND RATIO CUT METHODS. – EKATERINA MERKURJEV, ANDREA BERTOZZI, XIAORAN YAN, AND KRISTINA LERMAN, INVERSE PROBLEMS 2017

Variants of MBO and GL functional for these binary cut problems

$$\mathrm{cut}(S, \bar{S}) = \sum_{x \in S, y \in \bar{S}} w(x, y).$$

$$d(x) = \sum_{y \in V} w(x, y).$$

Ratio Cut

$$\mathrm{RatioCut}(S, \bar{S}) = \mathrm{cut}(S, \bar{S}) \left(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \right)$$

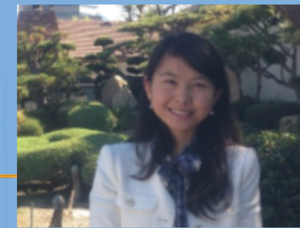
Normalized Cut

$$N\mathrm{cut}(S, \bar{S}) = \mathrm{cut}(S, \bar{S}) \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})} \right)$$

Cheeger Cut

$$h(S, \bar{S}) = \frac{\mathrm{cut}(S, \bar{S})}{\min(|S|, |\bar{S}|)}.$$

COMMUNITY DETECTION – MODULARITY OPTIMIZATION



Joint work with Huiyi Hu,Thomas Laurent, and Mason Porter SIAP 2013.

$$\text{Modularity: } Q = \frac{1}{2m}\sum_{ij}(w_{ij} - \gamma P_{ij})\delta(g_i, g_j)$$

Newman, Girvan, *Phys. Rev. E 2004*.

The modularity of a partition measures the fraction of total edge weight within each community minus the edge weight expected if edges were placed randomly using some null model.

$[w_{ij}]$ is graph adjacency matrix
P is probability nullmodel (Newman-Girvan) $P_{ij}=k_ik_j/2m$
$k_i = sum_j\ w_{ij}$ (strength of the node)
Gamma is the resolution parameter
g_i is group assignment
2m is total volume of the graph = $sum_i\ k_i = sum_{ij}\ w_{ij}$

This is an optimization (max) problem. Combinatorially complex – optimize over all possible group assignments. Very expensive computationally.

BIPARTITION OF A GRAPH

Given a subset A of nodes on the graph define

Vol(A) = sum $_{i \text{ in } A}$ k_i

Then maximizing Q is equivalent to minimizing

$$Cut(A, A^c) - \frac{\gamma}{2m} \text{vol}A \cdot \text{vol}A^c$$

Given a binary function on the graph f taking values +1, -1 define A to be the set where f=1, we can define:

◊ $|f|_{TV} = \frac{1}{2}\sum_{i,j} w_{ij}|f_i - f_j| = 2Cut(A, A^c)$ (Total Variation on Graph);
◊ $\|f\|^2_{L_2} = \sum_i k_i(f_i)^2$ (L_2 norm);
◊ $m_2(f) = (\sum_i k_i f_j)/2m$ (Mean).

EQUIVALENCE TO L1 COMPRESSIVE SENSING

Thus modularity optimization restricted to two groups is equivalent to

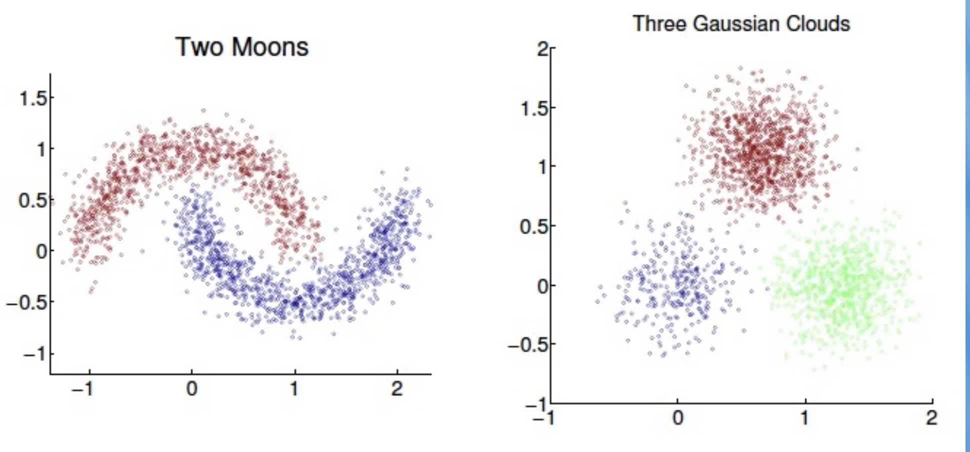
$$\text{Minimize}_{\{f:G\to\{\pm 1\}\}} |f|_{TV} - \frac{\gamma}{2}\|f - m_2(f)\|_{L_2}^2$$

This generalizes to n class optimization quite naturally

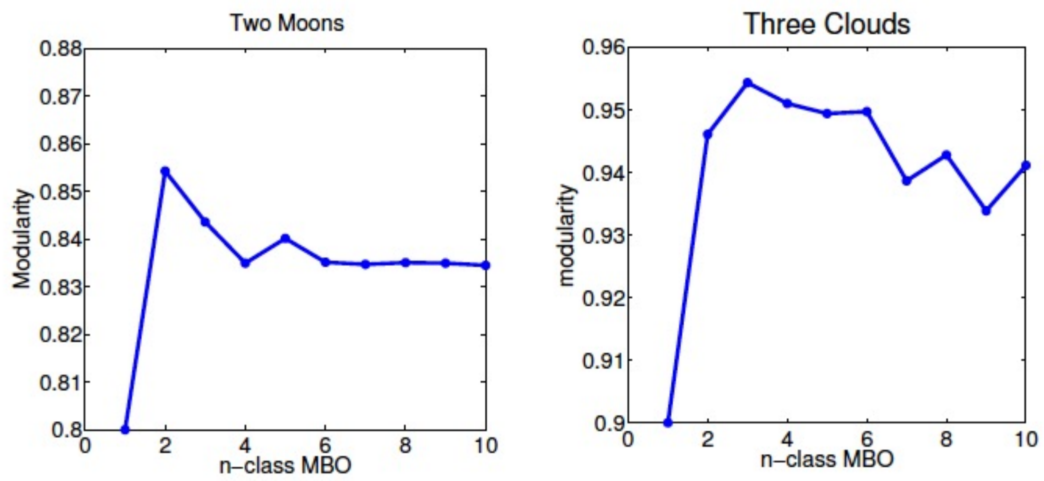
$$\text{Minimize}_{\{f:G\to V^n\}} \; E(f) := |f|_{TV} - \gamma\|f - m_2(f)\|_{L_2}^2$$

Because the TV minimization problem involves functions with values on the simplex we can directly use the MBO scheme to solve this problem.

MODULARITY OPTIMIZATION MOONS AND CLOUDS



(a) (b)



(a) (b)

MNIST DIGIT CLASSIFICATION USING MODULARITY – UNSUPERVISED

Binary segmentation of 4 and 9:
13782 handwritten digits. Graph created based on similarity score between each digit. Weighted graph with 194816 connections.

	N_c	Q	NMI	Purity	Time (seconds)
GenLouvain	2	0.9305	0.85	0.975	110 s
Modularity MBO ($\hat{n} = 2$)	2	0.9316	0.85	0.977	11 s
Multi-$\hat{n}$ MM ($\hat{n} \in \{2, 3, \ldots, 10\}$)	2	0.9316	0.85	0.977	25 s
Spectral Clustering (k-Means)	2	NA	0.003	0.534	1.5 s

TABLE 4.2

Full multiclass Segmentation of all 70K digits

TABLE 4.3
Computation times and network diagnostics for partitions of the MNIST 70k data set.

	N_c	Q	NMI	Purity	Time (s)
GenLouvain	11	0.93	0.92	0.97	10900
Multi-$\hat{n}$ MM ($\hat{n} \in \{2, 3, \ldots, 20\}$)	11	0.93	0.89	0.96	290 / 212 *
Modularity MBO 3% GT ($\hat{n} = 10$)	10	0.92	0.95	0.96	94.5 / 16.5 *

* *Calculated with the RC procedure.*

11 digits because there are two classes for the digit 1 ; with a flag and without a flag.

SIMPLIFIED ENERGY LANDSCAPE FOR MODULARITY USING TOTAL VARIATION

Z. Boyd, E. Bae, X. C. Tai, and A. L. Bertozzi, SIAM J. Appl. Math. 2018

Standard formulation:

$$\underset{\hat{n}\in\mathbb{N},\{A_\ell\}_{\ell=1}^{\hat{n}}\in\Pi(G)}{\operatorname{argmax}} \sum_{\ell=1}^{\hat{n}} \sum_{ij\in A_\ell} w_{ij} - \gamma\frac{k_i k_j}{2m}, \tag{8}$$

Balanced cut (I):

$$\underset{\hat{n}\in\mathbb{N},\{A_\ell\}_{\ell=1}^{\hat{n}}\in\Pi(G)}{\operatorname{argmin}} \sum_{\ell=1}^{\hat{n}} \left(\operatorname{Cut}(A_\ell, A_\ell^c) + \frac{\gamma}{2m}(\operatorname{vol} A_\ell)^2\right), \tag{9}$$

Balanced cut (II):

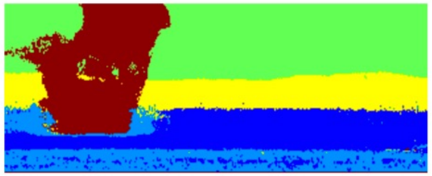
$$\underset{\hat{n}\in\mathbb{N},\{A_\ell\}_{\ell=1}^{\hat{n}}\in\Pi(G)}{\operatorname{argmin}} \sum_{\ell=1}^{\hat{n}} \left(\operatorname{Cut}(A_\ell, A_\ell^c) + \frac{\gamma}{2m}\left(\operatorname{vol} A_\ell - \frac{2m}{\hat{n}}\right)^2\right) + \gamma\frac{2m}{\hat{n}}, \tag{10}$$

Balanced TV (I):

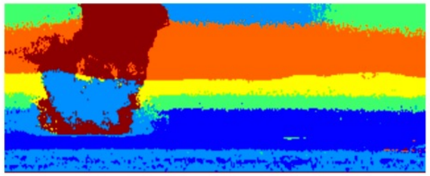
$$\underset{\hat{n}\in\mathbb{N},u\in\Pi(G)}{\operatorname{argmin}} |u|_{TV} + \frac{\gamma}{2m}\left|\left|k^T u\right|\right|_2^2, \tag{11}$$

Balanced TV (II):

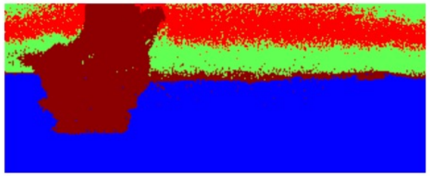
$$\underset{\hat{n}\in\mathbb{N},u\in\Pi(G)}{\operatorname{argmin}} |u|_{TV} + \frac{\gamma}{2m}\left|\left|k^T u - \frac{2m}{\hat{n}}\right|\right|_2^2 + \gamma\frac{2m}{\hat{n}}. \tag{12}$$



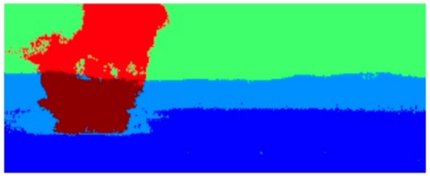
Balanced TV



Spectral Clustering



NLTV [28]



GenLouvain

STOCHASTIC BLOCK MODELS ARE A DISCRETE SURFACE TENSION

J. Nonlin. Sci. 2019 – Z. Boyd, M. A. Porter and A. L. Bertozzi

Assume adjacency matrix elements A_{ij}

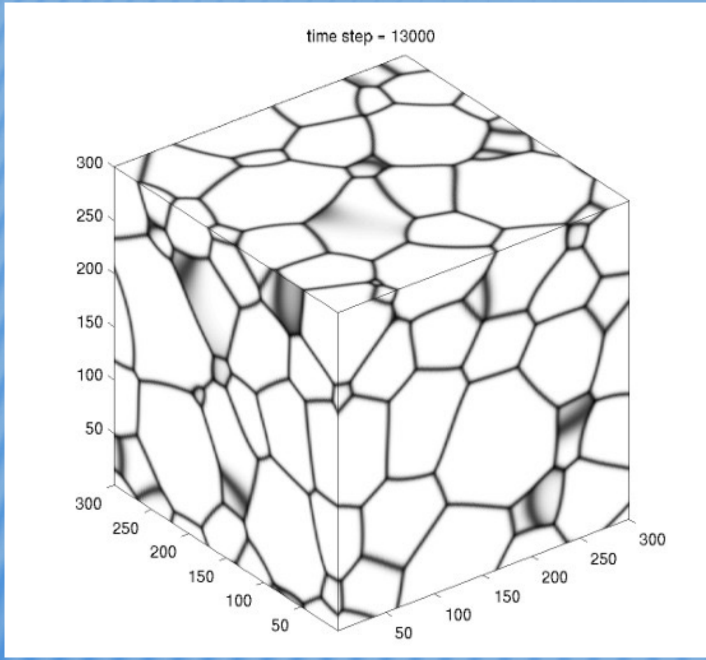
are Poisson-distributed with parameter $\omega_{g_i g_j} \dfrac{k_i k_j}{2m}$

Where m is the number of edges in the network.
This is called the PLANTED PARTITION MODEL. Newman (2016) shows that Modularity optimization is equivalent to Maximum Likelihood Estimation for SBMs for this class of models. The formula is:

$$\operatorname*{argmax}_{g,\omega} \sum_{i,j} \left[A_{ij} \log(\omega_{g_i g_j}) - \omega_{g_i g_j} \frac{k_i k_j}{2m} \right] .$$

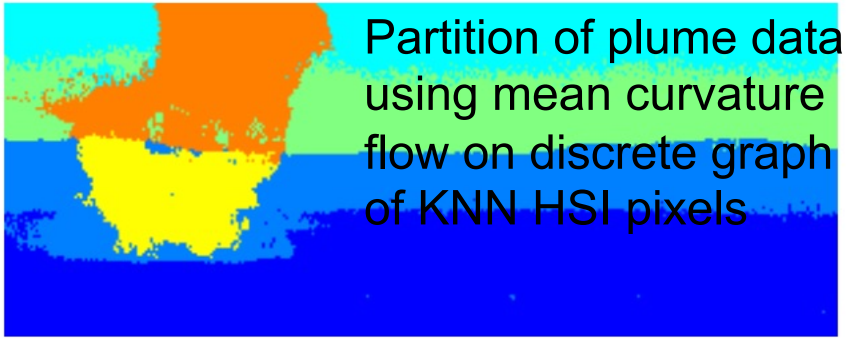
We show that this is a discrete analogue of a surface tension problem and adapt and algorithm of Esedoglu and Otto to the discrete case. Here we also have to optimize for the values of the surface tensions between classes.

SURFACE ENERGIES IN THE CONTINUUM



$$\sum_{\alpha,\beta} \sigma_{\alpha\beta} \text{Area}(\Gamma_{\alpha\beta}) \,,$$

$$\text{Area}_{\alpha\beta} = |u^{\alpha}|_{\text{TV}} + |u^{\beta}|_{\text{TV}} - |u^{\alpha} + u^{\beta}|_{\text{TV}} \,.$$



Partition of plume data using mean curvature flow on discrete graph of KNN HSI pixels

PROPOSITION 3.1. *Maximizing the likelihood of the parameters g and ω in the degree-corrected SBM is equivalent to minimizing*

$$\sum_{\alpha,\beta} \left[W_{\alpha\beta} \text{Cut}_{g,A}(\alpha,\beta) + e^{-W_{\alpha\beta}} \frac{\text{vol}_{g,A}(\alpha)\,\text{vol}_{g,A}(\beta)}{2m} \right] , \tag{6}$$

where $\text{Cut}_{g,A}(\alpha,\beta) = \sum_{\substack{g_i=\alpha \\ g_j=\beta}} A_{ij}$, $\text{vol}_{g,A}(\alpha) = \sum_{g_i=\alpha} k_i$, *and* $W_{\alpha\beta} = -\log \omega_{\alpha\beta}$.

RESULTS OF SBM MODEL USING CURVATURE BASED ALGORITHMS

		PP	LFR	MS	Caltech	Princeton	Penn. St.	Plume
	Nodes	16,000	1,000	10,230	762	6,575	41,536	284,481
	Edges	2.9×10^5	9.8×10^3	1.0×10^5	16,651	293,307	1,362,220	2,723,840
	Communities	10	40	10	8	4	8	5
Score	MCF	0	0	0	−0.16	−0.02	−0.56	−1.41
	AC	0	0	0	0.21	0.58	−0.04	−1.23
	MBO	0	0	0	0.53	1.12	0.40	−1.21
	KL	0.28	0.03	0.04	−0.16	0.11	−0.55	−1.38
	Reference	0	0	0	0	0	0	0

TABLE 1

	PP	LFR	MS	Caltech	Princeton	Penn. State	Plume
MCF	5.36	17.71	3.47	1.39	1.46	38.91	77.91
AC	5.37	26.27	7.28	8.84	480.4	3853	268.7
MBO	4.27	11.05	1.73	0.67	7.43	382.31	270.0
KL	16,566	176	5,117	20	662	95,603	980,520

TABLE 2
Computation times (in seconds).

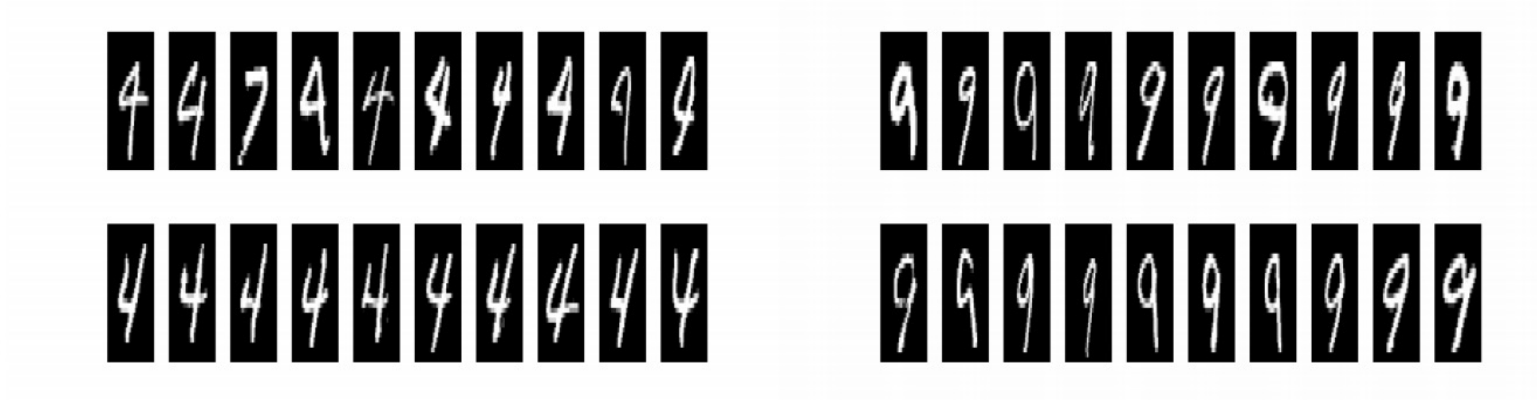
KL = Kernighan and Lin

Preprint available at arXiv:1806.02485v1

UNCERTAINTY QUANTIFICATION

With X. Luo, A. Stuart, and K. Zygalakis (SIAM UQ 2018)

Probit classifier; level set method for Bayesian inverse problems both extended to graphs from Euclidean space; generalize Ginzburg-Landau to Bayesian setting.



(a) Fours in MNIST (b) Nines in MNIST

FIG. 5. *"Hard to classify" vs "easy to classify" nodes in the MNIST* (4, 9) *dataset under the probit model. Here the digit "4" is labeled +1 and "9" is labeled -1. The top (bottom) row of the left column corresponds to images that have the lowest (highest) values of s_j^l defined in* (8) *among images that have ground truth labels "4". The right column is organized in the same way for images with ground truth labels* 9 *except the top row now corresponds to the highest values of s_j^l. Higher s_j^l indicates higher confidence that image j is a* 4 *and not a "*9*", hence the top row could be interpreted as images that are "hard to classify" by the current model, and vice versa for the bottom row. The graph is constructed as in Section* 5, *and $\gamma = 0.1$, $\beta = 0.3$.*

PAPERS

- A. L. Bertozzi and A. Flenner, *Multiscale Modeling and Simulation*, 10(3), 2012.
- Tijana Kostic and Andrea Bertozzi, *J. Sci Comp.*, 2012
- Y van Gennip and ALB *Adv. Diff. Eq*. 2012
- H. Hu, Y. van Gennip, B. Hunter, A.L. Bertozzi, M.A. Porter, I*EEE ICDM'12*, 2012.
- Y. van Gennip et al *SIAP* (spectral clustering gang data) 2013
- E. Merkurjev, T. Kostic, and A. L. Bertozzi, *SIAM J. Imag. Proc. 2013.*
- Huiyi Hu, Thomas Laurent, Mason A. Porter, Andrea L. Bertozzi, *SIAM J. Appl. Math., 2013.*
- C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner and A. G. Percus,, *IEEE Trans. PAMI 2014*
- Y. van Gennip, N. Guillen, B. Osting, and A. L. Bertozzi, Mean curvature, threshold dynamics, and phase field theory on finite graphs, *Milan J. Math. 2014.*
- Huiyi Hu, Justin Sunu, and Andrea L. Bertozzi, Multi-class Graph Mumford-Shah Model for Plume Detection using the MBO scheme, *Proc. EMMCVPR* Hong Kong 2015, pp. 209-222.
- E. Merkurjev, E. Bae, A. L. Bertozzi, X-C. Tai, Global binary optimization on graphs for classification of high dimensional data, *JMIV* 2015.
- X. Luo and A. L. Bertozzi, Convergence Analysis of the Graph Allen-Cahn Scheme, J Stat. Phys. 2017.
- E. Merkurjev, A. Bertozzi, X. Yan, and K. Lerman, Modified Cheeger and Ratio Cut Methods Using the Ginzburg-Landau Functional for Classification of High-Dimensional Data, Inverse Problems 2017.
- Bertozzi and Flenner – SIGEST 2016.
- Z. Meng, E. Merkurjev, A. Koniges, and A. L. Bertozzi, Hyperspectral Image Classification Using Graph Clustering Methods, Image Processing Online (with code), 7 (2017), pp. 218-245, 2017.
- Z Meng et al IWOMP 2016
- Z. Boyd, E. Bae, X.C. Tai, and A. L. Bertozzi, Simplified energy landscape for modularity using total variation, SIAM J. Appl. Math 78(5), 2018.
- A. L. Bertozzi, X. Luo, A. M. Stuart, and K. C. Zygalakis, Uncertainty Quantification in the Classification of High Dimensional Data, SIAM UQ, 2018.
- A. L. Bertozzi, Proc. ICM Rio de Janeiro, 2018
- Z. Boyd, M. A. Porter and A. L. Bertozzi, J. Nonlin. Sci., 2019