Diffuse interface methods on graphs: Data clustering and Gamma-limits

Yves van Gennip
joint work with Andrea Bertozzi, Jeff Brantingham, Blake Hunter

Department of Mathematics, UCLA

Research made possible by
ONR grants N000141210040 and N000141010221

Graph Exploitation Symposium
MIT Lincoln Laboratory
April 18 2012
Overview

- Graphs and data clustering
- Graph Laplacian and clustering
- Spectral clustering and street gangs
- Ginzburg-Landau for nonlinear clustering
- $\Gamma$-limits and nonlocal means
- Future work
Graphs and data clustering
Graph Laplacian and clustering
Spectral clustering and street gangs
Ginzburg-Landau for nonlinear clustering
\( \Gamma \)-limits and nonlocal means
Future work
Data points are represented by nodes in an undirected graph. Similarity is encoded in edge weights $\omega_{ij}$. Data clustering is represented by node labels $u_i$. 
The goal

Clustering: Label all the nodes with a $k$-valued label, $u_i \in \{1, \ldots, k\}$, such that an appropriate quantity is optimized.

We choose to minimize normalized cut (Shi, Malik, 2000). If $V$, the set of all nodes, is divided into $k$ disjoint clusters $C_i$:

$$\text{ratio}(C_i) := \frac{\sum_{i \in C_i, j \in V \setminus C_i} \omega_{i,j}}{\sum_{i \in C_i, j \in V} \omega_{i,j}}$$

and

$$\text{Ncut} := \sum_{i=1}^{k} \text{ratio}(C_i).$$

Normalization is to avoid small clusters.

Note: Here we prescribe the number of clusters $k$. 
Minimizing NCut is an NP-complete problem (Papadimitriou 1997).

Practical solution: Solve a relaxed version of the problem.
Overview

- Graphs and data clustering
- **Graph Laplacian and clustering**
- Spectral clustering and street gangs
- Ginzburg-Landau for nonlinear clustering
- $\Gamma$-limits and nonlocal means
- Future work
Graph Laplacians in the literature

**Unnormalized graph Laplacian**

\[(\Delta u)_i = \sum_j \omega_{ij}(u_i - u_j)\]

**Random walk graph Laplacian**

\[(\Delta u)_i = \frac{1}{d_i} \sum_j \omega_{ij}(u_i - u_j) \quad \text{with degree } d_i = \sum_j \omega_{ij}\]

**Symmetric normalized graph Laplacian**

\[(\Delta u)_i = \sum_j \frac{\omega_{ij}}{\sqrt{d_i}} \left(\frac{u_i}{\sqrt{d_i}} - \frac{u_j}{\sqrt{d_j}}\right)\]

For an overview, see e.g. Von Luxburg (2007).
Focus on the random walk Laplacian

Adjacency matrix $A$ and degree matrix $D$:

$$A = \begin{pmatrix} \omega_{11} & \ldots & \omega_{1m} \\ \omega_{21} & \ldots & \omega_{2m} \\ \vdots & \vdots & \vdots \\ \omega_{m1} & \ldots & \omega_{mm} \end{pmatrix}, \quad D = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & d_m \end{pmatrix}$$

The random walk Laplacian is then given by

$$\Delta u = (1 - D^{-1}A)u.$$
Eigenvectors as solutions to a relaxed problem


The solution to the NCut problem is given by the minimization of

\[ \text{Tr}(H'(I - D^{-1}A)H) \]

over all $D$-orthonormal matrices $H$ whose columns are indicator functions of clusters.

If we relax the condition on $H$ to allow any real-valued $D$-orthonormal matrix, then the minimization of $\text{Tr}(H'\Delta H)$ is solved by the matrix $H$ whose columns are the first $k$ eigenvectors (ordered according to increasing eigenvalue) of $\Delta = I - D^{-1}A$.

Use the first $k$ eigenvectors of the random walk Laplacian as approximations to the cluster indicator functions.
Trivial, ‘clean’, example.

\[ A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad \Delta = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{pmatrix} \]

The eigenvalues are 0, 0, 1, 1, 1, 1 with (unnormalized) eigenvectors (next page)
Trivial, ‘clean’, example, cont’d

\[ v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \]

\[ v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ -1 \\ -1 \end{pmatrix}, \quad v_5 = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}, \]
### ‘Noisy’ example

**Matrix A**

\[
A = \begin{pmatrix}
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix}
\]

**Matrix Δ**

\[
\Delta = \begin{pmatrix}
\frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & -\frac{1}{4} & 0 \\
-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
-\frac{1}{3} & -\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 \\
0 & 0 & 0 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\
-\frac{1}{5} & 0 & -\frac{1}{5} & -\frac{1}{5} & \frac{4}{5} & -\frac{1}{5} \\
0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 1
\end{pmatrix}
\]

**Eigenvalues:** 0, 0.2649, 0.8748, 1.1012, 1.2599, 1.3826. The (unnormalized) eigenvectors corresponding to 0 and 0.2649:

- **v_1**

\[
v_1 = \begin{pmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
1
\end{pmatrix}
\]

- **v_2**

\[
v_2 = \begin{pmatrix}
0.3069 \\
0.4882 \\
0.2816 \\
-0.5571 \\
-0.1742 \\
-0.4974
\end{pmatrix}
\]
How do we go from eigenvector to indicator function. Two often used options:

- Threshold the values in the eigenvector.
- Spectral clustering (Ng, Jordan, Weiss, 2002): Use eigenvectors as basis for data points. Then use $k$-means to separate clusters in this space.
Overview

- Graphs and data clustering
- Graph Laplacian and clustering
- Spectral clustering and street gangs
- Ginzburg-Landau for nonlinear clustering
- $\Gamma$-limits and nonlocal means
- Future work
Identify gang membership based on geosocial information

This work with Jeff Brantingham and Blake Hunter started out as a REU project in the summer of 2011.
Given data:

- Average location of (nonviolent) stops, with pairwise distances \( d_{i,j} \)
- (Anonymized) individuals involved in stop, with pairwise social interactions \( S_{i,j} \)

Gang affiliation is available as ground truth.

Construct a graph via the adjacency/weight matrix

\[
A_{i,j} = \alpha S_{i,j} + (1 - \alpha) e^{-d_{i,j}^2/\sigma^2},
\]

with parameters \( \alpha \) and \( \sigma \).
Eigenvectors for specific choice of $S$

Eigenvector 2  Eigenvector 3  Eigenvector 4

Hotspots!
The pies are the clusters we find, the colors indicate gang affiliation.

Pie charts made with code from Traud, Frost, Mucha, Porter (2009)
Possible explanations

- Sparse data?
- Social structures in reality do not follow gang affiliations?
- Wrong method?
Overview

- Graphs and data clustering
- Graph Laplacian and clustering
- Spectral clustering and street gangs
- Ginzburg-Landau for nonlinear clustering
- $\Gamma$-limits and nonlocal means
- Future work
The Ginzburg-Landau functional: phase separation

From materials science and image analysis:

**Ginzburg-Landau functional**

$$GL_\varepsilon(u) = \frac{\varepsilon}{2} \int |\nabla u|^2 + \frac{1}{\varepsilon} \int W(u)$$

When minimized (under some extra constraint, e.g. fixed mass or fidelity to data): phase separation

**$L^2$ gradient flow**

Allen-Cahn equation: $$u_t = \varepsilon \Delta u - \frac{1}{\varepsilon} W'(u) \text{ (+ constraint)}$$

Use graph Laplacian to formulate AC equation on graphs.
The Allen-Cahn equation on graphs

\[(u_i)_t = -\varepsilon(\Delta u)_i + \frac{1}{\varepsilon} W'(u_i) (+ \text{constraint})\]

can be seen as a more stringent relaxation of the NCut minimization problem.  
(N.B. Sign convention: \(-\Delta\) is a negative operator on graphs)

The double well term involving \(W'\) forces the solution to be close to binary.
\begin{align*}
(u_i)_t &= -\varepsilon(\Delta u)_i + \frac{1}{\varepsilon} W'(u_i) + \lambda_i (u_i - (u_{\text{orig}})_i) \\
\lambda_i &= \begin{cases} 
1 & \text{i in original} \\
0 & \text{otherwise}
\end{cases}
\end{align*}

- **Vertices**: the pixels from both images
- **Edge weight**: 
  \[ \omega_{ij} = e^{-\|x_i - x_j\|^2 / \tau} \]
  where \( \tau \) is a scale parameter and
- **Constraint**: fidelity to data in original image

- \( x_i \) is the feature vector of pixel \( i \)
Application: Data ‘inpainting’ (Bertozzi, Flenner, 2012)

\[(u_i)_t = -\varepsilon(\Delta u)_i + \frac{1}{\varepsilon} W'(u_i) + \lambda_i(u_i - (u_{\text{orig}})_i), \quad \lambda_i = \begin{cases} 1 & \text{known data} \\ 0 & \text{otherwise} \end{cases}\]

Feature vector consisting of \(n\) votes: yes (1), no (-1), or did not vote (0).
\(n \in \{8, 10, 12, 14, 16\}\)

- Vertices: 435 members of the 1984 US House of Representatives
- Edge weight:
  \(\omega_{ij} = e^{-\|x_i - x_j\|^2/\tau}\) where \(\tau\) is a scale parameter and
- \(x_i\) is the voting vector of individual \(i\)
- Constraint: fidelity to 5 individuals of known party affiliation (dem. or rep.)
\[(u_i)_t = -\varepsilon(\Delta u)_i + \frac{1}{\varepsilon} W'(u_i)\]

with mass constraint \(\sum_i u_i = 0\) (wells @ \(\pm 1\))

- Vertices: points forming two 2D moons in \(\mathbb{R}^{100}\) (with noise)
- Edge weight dependent on distance between points
- Mass constraint
Graphs and data clustering
Graph Laplacian and clustering
Spectral clustering and street gangs
Ginzburg-Landau for nonlinear clustering
$\Gamma$-limits and nonlocal means
Future work
What about small $\varepsilon$?

In the continuum case it is known that $\text{GL}_\varepsilon \Gamma$-converges to the total variation functional on binary functions $u$:

$$\text{GL}_\varepsilon(u) \xrightarrow{\Gamma} \sigma_W \int |\nabla u|, \quad \text{as } \varepsilon \to 0.$$  

This measures the interface between the two phases $u = 0$ and $u = 1$, with surface tension $\sigma_W$. (Modica, Mortola, 1977)

What happens on graphs?

$$\text{GL}_\varepsilon(u) = \frac{\varepsilon^2}{4} \sum_{ij} \omega_{ij} (u_i - u_j)^2 + \frac{1}{\varepsilon^2} \sum_i W(u_i)$$

What scaling in $\varepsilon$ to use?

By going to a graph we have lost the intrinsic length scale in the gradient. Do we want to keep $\varepsilon$ in the first term?
Why is $\Gamma$-convergence interesting?

**If $GL_\epsilon \xrightarrow{\Gamma} GL_0$ and a compactness property holds, then:**

- If $u_\epsilon$ minimizes $GL_\epsilon$ and $u_\epsilon \rightarrow u_0$, then $u_0$ minimizes $GL_0$

**Definitions, if you are interested:**

**$\Gamma$-limit of sequence of functionals**

A sequence $\{F_n\}$ $\Gamma$-converges to $F_0$ as $n \rightarrow \infty$ if, for all $u \in \text{Dom}(F)$

1. $\forall u_n \rightarrow u \liminf_{n \rightarrow \infty} F_n(u_n) \geq F(u)$ and
2. $\exists u_n \rightarrow u \limsup_{n \rightarrow \infty} F_n(u_n) \leq F(u)$.

**Compactness property**

Compactness: $F_n(u_n) < C \Rightarrow \{u_n\}$ has a convergent subsequence.
For a fixed general graph with edge weights $\omega_{ij}$ independent of $\epsilon$

\[
\frac{1}{4} \sum_{ij} \omega_{ij}(u_i - u_j)^2 + \frac{1}{\epsilon} \sum_i W(u_i) \xrightarrow{\Gamma} C_W \sum_{i,j} \omega_{ij}|u_i - u_j| \text{ as } \epsilon \to 0,
\]

where the limit functional is defined on binary functions $u$ taking values in $\{0, 1\}$. $C_W$ is a constant depending only on $W$.

- The total variation $\sum_{i,j} \omega_{ij}|u_i - u_j|$ is related to graph cuts.
- Dirichlet energy $\frac{1}{4} \sum_{ij} \omega_{ij}(u_i - u_j)^2$ does not scale with $\epsilon$. 

Graph based functional vs numerical discretization

For a regular square grid (4-regular graph) of $N \times N$ nodes on the torus $\mathbb{T}^2$ we compare

- **Graph based function with uniform edge weights** $\omega_{ij} = \frac{1}{N}$:

$$\text{GL}_g^\varepsilon(u) = N^{-1} \sum_{i,j=1}^{N} (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + \varepsilon^{-1} \sum_{i,j=1}^{N} W(u_{i,j})$$

(The double subscripts denote horizontal and vertical directions.)

- **Functional given by the discretization of the continuum GL functional**:

$$\text{GL}_d^\varepsilon(u) = \varepsilon \sum_{i,j=1}^{N} (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + \varepsilon^{-1} N^{-2} \sum_{i,j=1}^{N} W(u_{i,j})$$
By the earlier result, for fixed $N$ and $\varepsilon \to 0$:

$$GL^g_{\varepsilon}(u) \xrightarrow{\Gamma} N^{-1} \sum_{i,j=1}^{N} (|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|),$$

with $u_{i,j} \in \{0, 1\}$

Then let $N \to \infty$:

$$\ldots \xrightarrow{\Gamma} \int |u_x| + |u_y|,$$

$u(x) \in \{0, 1\}$.

Combine both limits via scaling $\varepsilon = N^{-\alpha}, N \to \infty$:

$$N^{-1} \sum_{i,j=1}^{N} (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + N^\alpha \sum_{i,j=1}^{N} W(u_{i,j}) \xrightarrow{\Gamma} \int |u_x| + |u_y|,$$

with $u(x) \in \{0, 1\}$, if $\alpha$ is large enough.
Results: Discretized functional

- Fix $\varepsilon$ and let $N \to \infty$: 

$$GL^d_\varepsilon(u) \xrightarrow{\Gamma} \frac{\varepsilon}{2} \int |\nabla u|^2 + \frac{1}{\varepsilon} \int W(u).$$

- Then, by Modica-Mortola, if $\varepsilon \to 0$: 

$$\cdots \xrightarrow{\Gamma} \hat{\sigma} W \int |\nabla u|,$$

with $u(x) \in \{0, 1\}$.

- Combine both limits via $\varepsilon = N^{-\alpha}, N \to \infty$: 

$$N^{-\alpha} \sum_{i,j=1}^{N} (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + N^{\alpha-2} \sum_{i,j=1}^{N} W(u_{i,j}),$$

with $u(x) \in \{0, 1\}$, if $\alpha$ small enough.
Nonlocal means type functional

Still consider square grid on $\mathbb{T}^2$. Given $f \in C^\infty(\mathbb{T}^2)$ create completely connected graph with weights $\omega = e^{-d^2/\sigma^2}$ where $\sigma > 0$ and $d$ compares patches of $L$ nodes by $L$ nodes sampled from $f$. Then, for binary valued $u$,

$$N^{-4} \sum_{i,j,k,l} \omega_{i,j,k,l} |u_{i,j} - u_{k,l}| \xrightarrow{\Gamma} \int_{\mathbb{T}^2} \int_{\mathbb{T}^2} \omega(x,y) |u(x) - u(y)| \, dx \, dy,$$

where

$$\omega(x,y) = e^{-\frac{4L^2}{\sigma^2}(f(x) - f(y))^2} \text{ if } \sigma, L \text{ are fixed and } N \to \infty,$$

$$\omega(x,y) = e^{-c^2 \int_{S_\ell} (f(x+z) - f(y+z))^2 \, dz} \text{ if } \sigma = N/c \text{ and } L/N \to \ell \text{ as } N \to \infty,$$

where $S_\ell = \{ z \in \mathbb{T}^2 : |z_1| + |z_2| \leq \ell \}$. 

Yves van Gennip (UCLA) Diffuse interface methods on graphs April 17–18 2012 34/36
Overview

- Graphs and data clustering
- Graph Laplacian and clustering
- Spectral clustering and street gangs
- Ginzburg-Landau for nonlinear clustering
- $\Gamma$-limits and nonlocal means
- Future work
Some future work

- Generalize the asymptotic (graph to continuum) results, *e.g.* on graphs sampled from a manifold
- Apply Ginzburg-Landau to other data clustering or image segmentation problems
- See if and how other clustering/community detection methods fit into this framework, *e.g.* the multiplex method of Mucha, Richardson, Macon, Porter, Onnela (2010)
- Look at other PDE and PDE questions translated onto graphs

Thank you for your connectivity
Some future work

- Generalize the asymptotic (graph to continuum) results, e.g. on graphs sampled from a manifold
- Apply Ginzburg-Landau to other data clustering or image segmentation problems
- See if and how other clustering/community detection methods fit into this framework, e.g. the multiplex method of Mucha, Richardson, Macon, Porter, Onnela (2010)
- Look at other PDE and PDE questions translated onto graphs

Thank you for your connectivity