Lecture 11: Review of nonlinear geometric graph modeling

Radu Balan

Department of Mathematics, AMSC, CSCAMM and NWC
University of Maryland, College Park, MD

May 2, 2017
Main Problem

Given a weighted graph $G = (\mathcal{V}, W)$ with $n$ nodes, find a dimension $d$ and a set of $n$ points $\{y_1, \cdots, y_n\} \subset \mathbb{R}^d$ such that $W_{i,j} = \varphi(\|y_i - y_j\|)$ for some monotonically decreasing function $\varphi$. Additionally test how a random graph model explains the data.
Main Problem

Given a weighted graph $G = (V, W)$ with $n$ nodes, find a dimension $d$ and a set of $n$ points $\{y_1, \cdots, y_n\} \subset \mathbb{R}^d$ such that $W_{i,j} = \varphi(\|y_i - y_j\|)$ for some monotonically decreasing function $\varphi$. Additionally test how a random graph model explains the data.

Thus we look for a dimension $d > 0$ and a set of points $\{y_1, y_2, \cdots, y_n\} \subset \mathbb{R}^d$ so that all $d_{i,j} = \|y_i - y_j\|$’s are compatible with weighted graph.

Typical weight functions:

1. **Exponential model**: $\varphi(t) = Ce^{-t^2}$, for some $C > 0$.

2. **Power law**: $\varphi(t) = \frac{C}{t^p}$, for some $C > 0$ and $p > 0$. 
Analysis

Three studies need to be done:

1. **Random graph hypothesis**: Sort edges by weight: from the largest weight to the smallest weight. Then compare sample statistics of 3-cliques, 4-cliques and spectral gap to the expected ones for $G_{n,p}$ and $\Gamma_{n,m}$.
Analysis

Three studies need to be done:

1. **Random graph hypothesis**: Sort edges by weight: from the largest weight to the smallest weight. Then compare sample statistics of 3-cliques, 4-cliques and spectral gap to the expected ones for $G_{n,p}$ and $\Gamma_{n,m}$.

2. **SDP optimization approach**: Solve for the Gram matrix $G$ that optimizes a Semi-Definite Program; Find its effective rank and then perform the SVD of $G = Y^T Y$ to find the geometric graph.
Three studies need to be done:

1. **Random graph hypothesis**: Sort edges by weight: from the largest weight to the smallest weight. Then compare sample statistics of 3-cliques, 4-cliques and spectral gap to the expected ones for $G_{n,p}$ and $\Gamma_{n,m}$.

2. **SDP optimization approach**: Solve for the Gram matrix $G$ that optimizes a Semi-Definite Program; Find its effective rank and then perform the SVD of $G = Y^T Y$ to find the geometric graph.

3. **Laplacian eigenmaps**: The geometric graph is obtained by solving the bottom $d+1$ eigenproblems for the normalized symmetric Laplacian $\tilde{\Delta} = I - D^{-1/2} WD^{-1/2}$.
Distribution of Cliques

Expected Values

Let $X_q$ denote the number of $q$-cliques in a random graph $G$. Then the expectation of $X_q$ in $\mathcal{G}_{n,p}$ class is

$$\mathbb{E}[X_q] = \binom{n}{q} p^{q(q-1)/2}$$

The expectation of $X_q$ in the class $\Gamma^{n,m}$ is approximated by the above formula for $p = \frac{2m}{n(n-1)}$:

$$\mathbb{E}[X_q] \approx \binom{n}{q} \left( \frac{2m}{n(n-1)} \right)^{q(q-1)/2} \sim \theta q \frac{m^{q(q-1)/2}}{n^{q(q-2)}}$$

$$\mathbb{E}[X_3] \sim \theta \frac{m^3}{n^3} , \quad \mathbb{E}[X_4] \sim \theta \frac{m^6}{n^8}$$
3-Cliques and 4-cliques
Thresholds

Theorem

Let \( m = m(n) \) be the number of edges in \( \Gamma^{n,m} \).

1. If \( m \gg n \) (i.e. \( \lim_{n \to \infty} \frac{m}{n} = \infty \)) then
   \( \lim_{n \to \infty} \text{Prob}[G \in \Gamma^{n,m} \text{ has a 3-clique}] \to 1 \).

2. If \( m \ll n \) (i.e. \( \lim_{n \to \infty} \frac{m}{n} = 0 \)) then
   \( \lim_{n \to \infty} \text{Prob}[G \in \Gamma^{n,m} \text{ has a 3-clique}] \to 0 \).

Theorem

Let \( m = m(n) \) be the number of edges in \( \Gamma^{n,m} \).

1. If \( m \gg n^{4/3} \) (i.e. \( \lim_{n \to \infty} \frac{m}{n^{4/3}} = \infty \)) then
   \( \lim_{n \to \infty} \text{Prob}[G \in \Gamma^{n,m} \text{ has a 4-clique}] \to 1 \).

2. If \( m \ll n^{4/3} \) (i.e. \( \lim_{n \to \infty} \frac{m}{n^{4/3}} = 0 \)) then
   \( \lim_{n \to \infty} \text{Prob}[G \in \Gamma^{n,m} \text{ has a 4-clique}] \to 0 \).
3-Cliques and 4-Cliques
Behavior at the threshold

In general we obtain a "coarse threshold". Recall a Poisson process $X$ with parameter $\lambda$ has p.m.f. $\text{Prob}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$.

**Theorem**

In $G_{n,p}$,

1. For $p = \frac{c}{n}$, $X_3$ is asymptotically Poisson with parameter $\lambda = c^3/6$.
2. For $p = \frac{c}{n^{2/3}}$, $X_4$ is asymptotically Poisson with parameter $\lambda = c^6/24$. 
3-Cliques and 4-Cliques
Behavior at the threshold

In general we obtain a "coarse threshold". Recall a Poisson process $X$ with parameter $\lambda$ has p.m.f. $\text{Prob}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$.

Theorem

In $G_{n,p}$,

1. For $p = \frac{c}{n}$, $X_3$ is asymptotically Poisson with parameter $\lambda = \frac{c^3}{6}$.
2. For $p = \frac{c}{n^{2/3}}$, $X_4$ is asymptotically Poisson with parameter $\lambda = \frac{c^6}{24}$.

Theorem

In $\Gamma^{n,m}$,

1. For $m = cn$, $X_3$ is asymptotically Poisson with parameter $\lambda = \frac{4c^3}{3}$.
2. For $m = cn^{4/3}$, $X_4$ is asymptotically Poisson with parameter $\lambda = \frac{8c^6}{3}$. 
What do we know about the set of eigenvalues of these matrices for a graph $G$ with $n$ vertices?

1. $\Delta = \Delta^T \geq 0$ and hence its eigenvalues are non-negative real numbers.
2. $\text{eigs}(\tilde{\Delta}) = \text{eigs}(L) \subset [0, 2]$.
3. 0 is always an eigenvalue and its multiplicity equals the number of connected components of $G$,

$$\dim \ker(\Delta) = \dim \ker(L) = \dim \ker(\tilde{\Delta}) = \# \text{connected components}.$$
What do we know about the set of eigenvalues of these matrices for a graph $G$ with $n$ vertices?

1. $\Delta = \Delta^T \geq 0$ and hence its eigenvalues are non-negative real numbers.
2. $\text{eigs}(\tilde{\Delta}) = \text{eigs}(L) \subset [0, 2]$.
3. $0$ is always an eigenvalue and its multiplicity equals the number of connected components of $G$,

$$\dim \ker(\Delta) = \dim \ker(L) = \dim \ker(\tilde{\Delta}) = \# \text{connected components}.$$ 

Let $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ be the eigenvalues of $\tilde{\Delta}$. Denote

$$\lambda(G) = \max_{1 \leq i \leq n-1} |1 - \lambda_i|.$$ 

Note $\sum_{i=1}^{n-1} \lambda_i = \text{trace}(\tilde{\Delta}) = n$. Hence the average eigenvalue is about 1. $\lambda(G)$ is called the absolute gap and measures the spread of eigenvalues away from 1.
The spectral absolute gap $\lambda(G)$

The main result in [9]) says that for connected graphs w/h.p.:

$$\lambda_1 \geq 1 - \frac{C}{\sqrt{\text{Average Degree}}} = 1 - \frac{C}{\sqrt{p(n-1)}} = 1 - C\sqrt{\frac{n}{2m}}.$$ 

**Theorem (For class $\mathcal{G}_{n,p}$)**

*Fix $\delta > 0$ and let $p > \left(\frac{1}{2} + \delta\right)\log(n)/n$. Let $d = p(n-1)$ denote the expected degree of a vertex. Let $\tilde{G}$ be the giant component of the Erdős-Rényi graph. For every fixed $\varepsilon > 0$, there is a constant $C = C(\delta, \varepsilon)$, so that

$$\lambda(\tilde{G}) \leq \frac{C}{\sqrt{d}}$$

with probability at least $1 - Cn \exp(-(2 - \varepsilon)d) - C \exp(-d^{1/4}\log(n))$.\*

Connectivity threshold: $p \sim \frac{\log(n)}{n}$.\*
The spectral absolute gap
\( \lambda(G) \)

The main result in [9] says that for connected graphs w/h.p.:

\[
\lambda_1 \geq 1 - \frac{C}{\sqrt{\text{Average Degree}}} = 1 - \frac{C}{\sqrt{p(n-1)}} = 1 - C\sqrt{\frac{n}{2m}}.
\]

**Theorem (For class \( \Gamma^{n,m} \))**

Fix \( \delta > 0 \) and let \( m > \frac{1}{2} (\frac{1}{2} + \delta) n \log(n) \). Let \( d = \frac{2m}{n} \) denote the expected degree of a vertex. Let \( \tilde{G} \) be the giant component of the Erdős-Rényi graph. For every fixed \( \varepsilon > 0 \), there is a constant \( C = C(\delta, \varepsilon) \), so that

\[
\lambda(\tilde{G}) \leq \frac{C}{\sqrt{d}}
\]

with probability at least \( 1 - Cn \exp(-(2 - \varepsilon)d) - C \exp(-d^{1/4} \log(n)) \).

Connectivity threshold: \( m \sim \frac{1}{2} n \log(n) \).
Isometric Embeddings with Partial Data
Linear constraints

Given any set of vectors \( \{y_1, \cdots, y_n\} \) and their associated matrix \( Y = [y_1|\cdots|y_n] \) their invariant to the action of the rigid transformations (translations, rotations, and reflections) is the Gram matrix of the centered system:

\[
G = \left( I - \frac{1}{n} 1_n 1_n^T \right) Y^T Y \left( I - \frac{1}{n} 1_n 1_n^T \right) =: L Y^T Y L, \quad L = I - \frac{1}{n} 1_n 1_n^T.
\]

On the other hand, the distance between points \( i \) and \( j \) can be computed by:

\[
d^2_{i,j} = \|y_i - y_j\|^2 = G_{i,i} - G_{i,j} + G_{j,j} - G_{j,i} = e_{ij}^T Ge_{ij}
\]

where

\[
e_{ij} = \delta_i - \delta_j = [0 \cdots 0 1 \cdots -1 0 \cdots 0]^T
\]

where 1 is on position \( i \), \( -1 \) is on position \( j \), and 0 everywhere else.
Almost Isometric Embeddings with Partial Data
The SDP Problem

Reference [10] proposes to find the matrix $G$ by solving the following Semi-Definite Program:

$$
\begin{align*}
\min_{G} & \quad \text{trace}(G) \\
G &= G^T \geq 0 \\
|\langle Ge_{ij}, e_{ij} \rangle - \tilde{d}_{i,j}^2| &\leq \varepsilon , \quad (i,j) \in \Theta
\end{align*}
$$

where $\tilde{d}_{i,j}^2$ are noisy estimates $d_{i,j}$ and $\varepsilon$ is the maximum noise level. The trace promotes low rank in this optimization. However, this is basically a feasibility problem: Decrease $\varepsilon$ to the minimum value where a feasible solution exists. With probability 1 that is unique.

How to do this: Use CVX with Matlab.
Algorithm

Input: Symmetric $n \times n$ Gram matrix $G$.

1. Compute the eigendecomposition of $G$, $G = Q\Lambda Q^T$ with diagonal of $\Lambda$ sorted in a descending order;

2. Determine the number $d$ of significant positive eigenvalues;

3. Partition

   $$Q = [Q_1 \quad Q_2], \text{ and } \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$$

   where $Q_1$ contains the first $d$ columns of $Q$, and $\Lambda_1$ is the $d \times d$ diagonal matrix of significant positive eigenvalues of $G$.

4. Compute:

   $$Y = \Lambda_1^{1/2} Q_1^T$$

Output: Dimension $d$ and $d \times n$ matrix $Y$ of vectors $Y = [y_1 | \cdots | y_n]$
Nearly Isometric Embeddings with Partial Data

Stability to Noise

[10] proves the following stability result in the case of partial measurements. Here we denote \( \Theta_r = \{(i, j), \|y_i - y_j\| \leq r\} \) the set of all pairs of points at distance at most \( r \).

**Theorem**

Let \( \{y_1, \cdots, y_n\} \) be \( n \) nodes distributed uniformly at random in the hypercube \([-0.5, 0.5]^d\). Further, assume that we are given noisy measurement of all distances in \( \Theta_r \) for some \( r \geq 10\sqrt{d(\log(n)/n)^{1/d}} \) and the induced geometric graph of edges is connected. Let \( \tilde{d}_{i,j}^2 = d_{i,j}^2 + \nu_{i,j} \) with \( |\nu_{i,j}| \leq \varepsilon \). Then with high probability, the error distance between the estimated \( \hat{Y} = [\hat{y}_1, \cdots |\hat{y}_n] \) returned by the SDP-based algorithm and the correct coordinate matrix \( Y = [y_1 | \cdots | y_n] \) is upper bounded as

\[
\|L\hat{Y}^T\hat{Y}L - LY^TYL\|_1 \leq C_1(nr^d)^{5/4} \frac{\varepsilon}{r^4}.
\]
Optimization Criterion

Assume $G = (V, W)$ is a undirected weighted graph with $n$ nodes and weight matrix $W$. We interpret $W_{i,j}$ as the similarity between nodes $i$ and $j$. The larger the weight the more similar the nodes, and the closer they are in a geometric graph embedding. Thus we look for a dimension $d > 0$ and a set of points \{y_1, y_2, \cdots, y_n\} $\subset \mathbb{R}^d$ so that $d_{i,j} = \|y_i - y_j\|$’s is small for large weight $W_{i,j}$. This means we want to minimize

$$J(y_1, y_2, \cdots, y_n) = \sum_{1 \leq i, j \leq n} W_{i,j}\|y_i - y_j\|^2,$$

To avoid trivial solution $Y = 0$ we impose a normalization condition:

$$YDY^T = I_d.$$
The Optimization Problem

Combining the criterion with the constraint:

\[
(LE) : \begin{align*}
\text{minimize} & \quad \text{trace} \left\{ Y \Delta Y^T \right\} \\
\text{subject to} & \quad YDY^T = I_d
\end{align*}
\]

we obtained the Laplacian Eigenmap problem.

Good news: The optimizer $Y$ is obtained by solving an eigenproblem.
Algorithm (Laplacian Eigenmaps)

Input: Weight matrix $W$, target dimension $d$

1. **Construct the diagonal matrix** $D = \text{diag}(D_{ii})_{1 \leq i \leq n}$, where $D_{ii} = \sum_{k=1}^{n} W_{i,k}$.

2. **Construct the normalized Laplacian** $\tilde{\Delta} = I - D^{-1/2}WD^{-1/2}$.

3. **Compute the bottom** $d + 1$ **eigenvectors** $e_1, \cdots, e_{d+1}$, $\tilde{\Delta}e_k = \lambda_k e_k$, $0 = \lambda_1 \cdots \lambda_{d+1}$. 
Algorithm (Laplacian Eigenmaps - cont’d)

4. Construct the $d \times n$ matrix $Y$,

$Y = \begin{bmatrix} e_2 \\ \vdots \\ e_{d+1} \end{bmatrix} D^{-1/2}$

5. The new geometric graph is obtained by converting the columns of $Y$ into $n$ $d$-dimensional vectors:

$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = Y$

Output: Set of points $\{y_1, y_2, \cdots, y_n\} \subset \mathbb{R}^d$. 
References


S. Boyd, L. Vandenberghe, Convex Optimization, available online at: http://stanford.edu/boyd/cvxbook/

F. Chung, Spectral Graph Theory, AMS 1997.


P. Erdös, A. Rényi, On The Evolution of Random Graphs


