March 3, 2020

Class so far:
1. least squares $\rightarrow$ objective functions subroutine
2. optimization $\rightarrow$ SGD
3. dimensionality redux/matrix fact.
   PCA $\leftrightarrow$ spectral clustering
   latent factors $\rightarrow$ node representation learning

Next: "network science"
   methods for graph data
   Graph: set of nodes and edges

Example: FB social network
<table>
<thead>
<tr>
<th>Network</th>
<th>Node</th>
<th>Edge</th>
<th>Undir/Dir</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB</td>
<td>users</td>
<td>friendship</td>
<td>undir</td>
</tr>
<tr>
<td>FB</td>
<td>users</td>
<td>message</td>
<td>dir</td>
</tr>
<tr>
<td>Cora</td>
<td>papers</td>
<td>citation</td>
<td>dir</td>
</tr>
<tr>
<td>Email</td>
<td>email addresses</td>
<td>sent email</td>
<td>dir</td>
</tr>
<tr>
<td>Coauthorship</td>
<td>scientists</td>
<td>coauthored paper</td>
<td>undir</td>
</tr>
<tr>
<td>Co-purchasing</td>
<td>co-viewing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Use relational structure:

1. Which nodes are important or central
2. How a network is organized (clustering)
3. How networks evolve?
4. How stuff spreads?
5. Predict properties about nodes?

Today: 1) Matrices associated with graphs
2) Common network structure
Adjacency matrix \( A \)

\[
A_{ij} = \begin{cases} 
1 & \text{if } i \to j \text{ in graph} \\
0 & \text{otherwise}
\end{cases}
\]

undirected: \( A = A^T \)

\[
A^k_{ij} = \# \text{ of length-}k \text{ paths from } i \text{ to } j
\]

\[
[A^{k-1} \cdot A]_{ij} = \sum_{l} A^{k-1}_{il} A_{lj}
\]

weighted: \( W_{ij} = \# \text{ of emails sent from } i \text{ to } j \)
Diagonal degree matrix $D$

\[ d = A1 \quad d_i = \sum_j A_{ij} \quad i \neq 0 \Rightarrow d_i = 2 \]

\[ D_{ii} = d_i \]

Random walk matrix $P$

\[ P = A^T D^{-1} \quad P_{ji} = A^T_{ij} \quad D^{-1} = A_{ij} / d_i \]

\[ P_{ri} = P_{ji} = P_{ki} = \frac{1}{3} \]

column stochastic
Graph Laplacian $L$ (undir. graph)

$G = (V, E)$

$\text{Cor: } L \text{ symm., pos. semidef}$

$L = D - A$

Claim: $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$

$= \sum_{(i,j) \in E} x_i^2 + x_j^2 - 2x_i x_j$

$= \sum_{i \in V} d_i x_i^2 - 2 \sum_{(i,j) \in E} x_i x_j$

$= x^T D x - 2 \sum_{1 \leq i, j \leq n} A_{ij} x_i x_j$

$= x^T (D - A) x$

$= x^T L x$
A, D, P, L

1. Matrix-vector, matrix-matrix multiplication
   \[ z = p^T y = D^{-1} A y \]
   \[ z_i = \frac{1}{d_i} \sum_j A_{ij} y_j \]

2. Solve systems of equations:
   
   PageRank: \( (I - \alpha P) x = (1-\alpha) 1 \)

3. Compute eigenvalues/eigenvectors
   
   Spectral clustering: \( L x = \lambda x \)
Basic spectral graph theory

Claim: number of connected components

= number of zero eigenvalues in \( L \)

Proof: \( c_{i}^{(s)} = \begin{cases} 1 & i \in S \\ 0 & o/w \end{cases} \)

\[
c^T L c = \sum_{(i,j) \in E} (c_i - c_j)^2 = \sum_{i,j \in S} (1-1)^2 + \sum_{i,j \notin S} (0-0)^2 = 0
\]

\( c^{(1)}, \ldots, c^{(k)} \quad c^{(i)^T} c^{(i)} = 0 \)

\[
0 = x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2 \quad x = \sum_{j} x_j c^{(i)}
\]

constant if \( i, j \) in same component
\[ c^T L c = 0 \implies L c = 0 c \]

\[ L = V \Lambda V^T \]

\[ x^T L x = x^T V \Lambda V^T x \]
\[ = y^T \Lambda y = \sum \lambda_i y_i^2 \]
\[ = 0 \]

\[ \Rightarrow y_i \neq 0 \implies \lambda_i = 0 \]

\[ V (\Lambda V^T) V^T x \]

\[ x^T L x \]
\[ \frac{x^T x}{x^T x} \]

\[ x \in \text{span}\{v_1, \ldots, v_k\} \]

\[ L v_i = 0 \]
Milgram (1969)
296 “random” people
Target person in Boston
64 successful

6 degrees of separation
diameter ≤ 6