Spectrum of Communication Networks

Alexander Tsiatas
26 August 2010

Joint work with Iraj Saniee, Onuttom Narayan and Matthew Andrews
Large networks

- Massive graphs representing entities and their relationships
  - Social networks
  - Biological networks
  - Transportation networks
  - Communication networks

- Intractability at such large scale

- Practical information from mathematical properties
  - Network bottlenecks, reliability, capacity
Mathematics and networks

- **Degree distribution**
  - Not very useful for bottlenecks

- **Expansion and isoperimetric (Cheeger) ratio**
  - For a set of nodes $S$, how many neighbors does $S$ have as a percentage of $S$'s size?
  - “Surface area to volume”

- **Spectral gap**
  - First non-zero eigenvalue of normalized Laplacian matrix (more later…)

- **Global network curvature or hyperbolicity**
  - (Gromov) For every 3 nodes A,B,C in the network, the shortest A-B path is close to the shortest B-C and A-C paths.

$$h(S) = \frac{e(S, \bar{S})}{vol(S)}$$
The spectral gap

- We use the normalized Laplacian Matrix:
  - \( L_{uu} = 1 \)
  - \( L_{uv} = \frac{-1}{\sqrt{d_u d_v}} \) if there is an edge \((u,v)\)
  - \( L_{uv} = 0 \) otherwise

- Eigenvalues:
  - 0 for each connected component
  - \( 0 < \lambda < 2 \) for the rest
  - Spectral gap: smallest nonzero eigenvalue
  - Eigenvalue close to zero: graph is ‘almost’ disconnected

- Cheeger’s inequality: relationship between Cheeger ratio (and therefore expansion) and spectral gap

\[
2h \geq \lambda \geq \frac{h^2}{2} \quad h = \min_{|S| \leq |\bar{S}|} h(S)
\]
Global negative curvature

• Gromov’s hyperbolicity condition (thin triangles):
  – For any 3 nodes ABC, draw the shortest A-B, B-C, and A-C paths
  – Each path lies within distance < delta of the other two

• “Flat” networks: Euclidean grids
  – No bottlenecks, poor expansion

• “Curved” networks: Trees, hyperbolic grids
  – Bottlenecks and good expansion
Global negative curvature

- Evidence for a relationship between hyperbolicity and bottlenecks in some sense
  - Route 1 unit between every pair of nodes on shortest paths
  - Load at core scales at $N^2$ for hyperbolic graphs, as opposed to $N^{1.5}$ for flat graphs
  - Congestion persists even when shortest paths are not used

- Evidence for hyperbolicity in real-world networks

$H(5,4)$
Expansion in the infinite and the finite

- Are (regular) trees good expanders?
- Conventional wisdom
  - \textbf{Yes}, exponential growth from root to leaves
- Cheeger ratio
  - \textbf{Yes} for infinite trees \((d-2)\) [Lyons et al.]
  - \textbf{No} for finite trees, just cut off one branch!
- Spectral gap
  - \textbf{Yes} for infinite trees, \(1 - 2\sqrt{(d-1)/d}\) [Friedman]
  - \textbf{No} for finite trees, tends to zero as the tree gets larger
  - Spectral gap for successively larger trees does not approach gap for the infinite tree!
- Real networks are finite, but we still want to capture basic properties…
- Real networks are large, but we can use more manageable subsets
From the infinite to the finite

• Problem: the boundary of a network
  – Infinite trees have a nonzero spectral gap, but that vanishes when the tree is finite
• Solution: use *Dirichlet eigenvalues*
  – Restrict the normalized Laplacian matrix to the rows and columns corresponding to non-boundary nodes
  – View the graph as a truncation of an infinite graph
  – Leads to a local Cheeger inequality
• Mathematically: for regular trees, the Dirichlet spectral gap for successively larger trees converges to the true spectral gap for infinite trees!
• Evidence that using Dirichlet eigenvalues provides insight into networks and their properties
  – Eliminating the boundary also gives more meaningful information in many cases – bottlenecks in the core of networks rather than near the boundary
Cuts and bottlenecks in networks

- Common problem: find a bottleneck in a network
- Formally: find a partition or cut that has low Cheeger ratio
  - “Sparse cut”, “good cut”, “bad cut”
- Bottlenecks determine network capacity and reliability
- NP-hard, but many robust approximate and heuristic algorithms
  - $k$-means
  - Affinity propagation
  - We will use spectral clustering
Problems with finding good (bad) cuts

- Balance
- Sparsity
  - Using Cheeger ratio as a metric helps
- Leskovec et al. 2008: “Bag of whiskers”
  - There will be several eigenvalues close to zero, representing the different whiskers.
  - Using Dirichlet eigenvalues can help avoid this because it avoids the boundary.
Rocketfuel datasets

- Subsets of Internet communication networks, 2002-2003
- Size
  - 121-10152 nodes
  - 456-28638 edges
- Created using freely available tools (ping, traceroute, etc.)
- Drawback: cannot see network topology past a certain point
  - Effect: an artificial boundary of degree-1 nodes
## Spectral gap in Rocketfuel data

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Nodes</th>
<th>Edges</th>
<th>Traditional gap</th>
<th>Dirichlet gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1221</td>
<td>2998</td>
<td>3806</td>
<td>0.003856095</td>
<td>0.07616252</td>
</tr>
<tr>
<td>1239</td>
<td>8341</td>
<td>14025</td>
<td>0.015935241</td>
<td>0.03585049</td>
</tr>
<tr>
<td>1755</td>
<td>605</td>
<td>1035</td>
<td>0.008963442</td>
<td>0.09585257</td>
</tr>
<tr>
<td>2914</td>
<td>7102</td>
<td>12291</td>
<td>0.001184411</td>
<td>0.04621173</td>
</tr>
<tr>
<td>3257</td>
<td>855</td>
<td>1173</td>
<td>0.010453303</td>
<td>0.04738221</td>
</tr>
<tr>
<td>3356</td>
<td>3447</td>
<td>9390</td>
<td>*</td>
<td>0.05083495</td>
</tr>
<tr>
<td>3967</td>
<td>895</td>
<td>2070</td>
<td>0.007994186</td>
<td>0.03365842</td>
</tr>
<tr>
<td>4755</td>
<td>121</td>
<td>228</td>
<td>0.035707392</td>
<td>0.06300309</td>
</tr>
<tr>
<td>6461</td>
<td>2720</td>
<td>3824</td>
<td>0.006395883</td>
<td>0.1103682</td>
</tr>
<tr>
<td>7018</td>
<td>10152</td>
<td>14319</td>
<td>*</td>
<td>0.09531827</td>
</tr>
</tbody>
</table>
Cuts in Rocketfuel data

Dirichlet cut

Traditional cut
Cuts in Rocketfuel data

- Dirichlet cut
- Traditional cut
Cuts in Rocketfuel data

Dirichlet cut

Traditional cut
Further study

• Rocketfuel datasets
  – Look at the behavior of the spectral gap with larger and larger subsets
  – Directly compare negative curvature and spectral gap
  – Classifying networks based on hyperbolicity and other properties

• Hyperbolic grids
  – Compute spectral gap (Cheeger ratio known)
Questions?