Structural Patterns in Complex Networks
Spectral Analysis

Ernesto Estrada
University of Strathclyde, Glasgow, UK

www.estradalab.org
“A complex system is a system composed of interconnected parts that as a whole exhibit one or more properties not obvious from the properties of the individual parts.”

✓ Emergence
✓ Agent-based
✓ Self-organised
✓ Short-range interactions
✓ Networked
Complex networks are the structural skeletons of complex systems

\[(A)_{ij} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ have a direct link} \\
0 & \text{otherwise}
\end{cases}\]
**Complex Networks in Nature**

- **Inter-species**
- **Inter-cellular**
- **Inter-molecular**
- **Inter-atomic**
Friendship networks
Collaboration networks
Sexual networks
Buyer-seller networks...
“What she is not, I can easily perceive – what she is I fear it is impossible to say.”

Edgar Allan Poe
Network Structure: Universality

Small-Worldness

Scale-freeness

What can we learn by studying the region around a node?
Protein Essentaility

Yeast

Protein-protein Interactions

Essential: because if knocked out the cell dies.

Are the most ‘central’ proteins the essential ones?
**Node Centrality**

**Degree**

\[ DC(p) = \sum_{q} A_{pq} \]

**Closeness**

\[ CC(p) = \frac{n - 1}{\sum_{q} d(p, q)} \]

**Betweenness**

\[ BC(p) = \sum_{i} \sum_{j} \frac{\rho(i, p, j)}{\rho(i, j)} \]

Percentage of Essential Proteins

Space for improvement?

<table>
<thead>
<tr>
<th></th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC(i)</td>
<td>50.0</td>
</tr>
<tr>
<td>DC(i)</td>
<td>46.7</td>
</tr>
<tr>
<td>BC(i)</td>
<td>40.0</td>
</tr>
<tr>
<td>Random</td>
<td>22.7</td>
</tr>
</tbody>
</table>

% of essential proteins vs # of links
Strategy: Use All Possible Routes

\[ CW_2(i) = 3 \]

\[ CW_3(i) = 2 \]

\[ CW_4(i) = 15 \]

Equal to the degree of the node

Transitivity
Subgraph Centrality: The Concept


\[
EE_i = \sum_{k=0}^{\infty} c_k \left( \text{# of CWs in } k \text{ steps starting at } i \right)
\]

\[
= \sum_{k=0}^{\infty} c_k \left( A^k \right)_{ii}
\]

\[
EE_i = \sum_{k=0}^{\infty} \frac{\left( A^k \right)_{ii}}{k!} = \left( e^A \right)_{ii}
\]

Eigenvalues of $A$: $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N$

Eigenvector of $\lambda_\mu$: $\phi_\mu = (\phi_\mu(1), \phi_\mu(2), \cdots)$

$$EE_i = \sum_{j=1}^{n} \left[\phi_j(i)\right]^2 e^{\lambda_j}$$

Bipartivity: Estrada and Rodríguez-Velázquez: Phys. Rev. E 72, 2005, 046105

$$EE(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) + \sum_{j=1}^{N} [\gamma_j(i)]^2 \cosh(\lambda_j)$$
Subgraph Centrality & Protein Essentiality

Estrada, E., Proteomics 6, 35 (2006)

![Graph showing percentage of essential proteins for different subgraph centrality measures: EE(i) = 63.3%, CC(i) = 50.0%, DC(i) = 46.7%, BC(i) = 40.0%, Random = 22.7%]
Networks with Homogeneous Link Weights

A Physical Analogy


\[ \beta = \frac{1}{k_B T} \]

Inverse temperature
A Physical Analogy


GAS

SOLID

The temperature can be considered as an external “stress”, e.g., social agitation, physiological stress or an economical situation.
A Statistical-mechanical Approach


\[
EE(G, \beta) = \sum_{j=1}^{n} e^{\beta \lambda_j}
\]

Partition function

\[
p_j = \frac{e^{\beta \lambda_j}}{\sum_{j=1}^{n} e^{\beta \lambda_j}}
\]

Probability

\[
F(G, \beta) = H - TS = -\beta^{-1} \ln EE
\]

Helmholtz Free Energy

\[
H(G, \beta) = -\sum_j \lambda_j p_j
\]

Total Energy

\[
S = -k_B \sum_j \left[ p_j (\beta \lambda_j - \ln EE) \right]
\]

Entropy
Global Structure of Networks

Modular OR Homogeneous?
Let $|S| \leq 1/2|V|$ and $|\partial(S)|$ represents the number of edges with exactly one endpoint in $S$.

**Edge Expansion (Isoperimetric Number)**

$$h(G) = \min_{1 \leq |S| \leq \frac{n}{2}} \frac{|\partial(S)|}{|S|}$$

A good expansion (homogeneous) network is one for which $h(G) = O(1)$. That is, ‘what you see locally is what you get globally’.

**Can we identify them?**
Good Expansion and Graph Spectra

\[ \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \]
\[ \frac{\lambda_1 - \lambda_2}{2} \leq h(G) \leq \sqrt{2\lambda_1 (\lambda_1 - \lambda_2)} \]


Large spectral gap \((\lambda_1 - \lambda_2)\) implies high expansion

\[ \lambda_1 - \lambda_2 = 11.147 \]
\[ \lambda_1 - \lambda_2 = 2.766 \]
How large the spectral gap should be?

$\lambda_1 - \lambda_2 = 8.714$

$\lambda_1 - \lambda_2 = 5.559$

Homogeneous?

Modular?
Spectral Scaling Method

Let:

\[ EE_{\text{odd}}(i) = [\gamma_1(i)]^2 \sinh(\lambda_1) + \sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) \]

Homogeneous network: that for which \( \lambda_1 - \lambda_2 \) is large enough as to consider that:

\[ [\gamma_1(i)]^2 \sinh(\lambda_1) \gg \sum_{j=2}^{N} [\gamma_j(i)]^2 \sinh(\lambda_j) \]

Then:

\[ EE_{\text{odd}}(i) \approx [\gamma_1(i)]^2 \sinh(\lambda_1) \]

Estrada, E. *Europhys. Lett.* 73, 2006, 649
Then:

\[
\log[y_1(i)] \approx \log A + \eta \log[EE_{odd}(i)]
\]

\[
\eta = 0.5
\]

\[
A = [\sinh(\lambda_1)]^{-0.5}
\]

\[
r = 0.99999
\]

Estrada, E. Europhys. Lett. 73, 2006, 649
Good Expansion and Modularity

E. Coli Metabolic Network

Perron-Frobenius Eigenvector

Subgraph Centrality
Universal Classes of Networks

\[
\left[\gamma_1(i)\right]^2 \sinh(\lambda_1) \equiv EE_{odd}(i)
\]

CLASS I

\[
\left[\gamma_1(i)\right]^2 \sinh(\lambda_1) \leq EE_{odd}(i), i \in V
\]

CLASS II

\[
\left[\gamma_1(i)\right]^2 \sinh(\lambda_1) \geq EE_{odd}(i), i \in V
\]

CLASS III

CLASS IV
Universal Classes of Real-World Networks

Estrada, E. Phys. Rev. E 75, 2007, 016103
Example: Protein Residue Networks

Protein Network

Spectral Scaling
How Protein Residue Networks are Built?

\[ A_{ij} = \begin{cases} 
H(r_c - r_{ij}) & i \neq j \\
0 & i = j 
\end{cases} \]

\[ H(x > 0) = 1 \]
\[ H(x \leq 0) = 0 \]
A Universal Class for Protein Networks?

E. Estrada, Biophys. J. 98, 2010, 890-900

595 proteins
Class IV Proteins are Small

E. Estrada, Biophys. J. 98, 2010, 890-900
Holes in Proteins

CHORDLESS 15-CYCLE

CAVITY = BINDING SITE

NODE VS. AMINO ACIDS IN BINDING SITES

30 33 40 43 44 60 64 67 68 71 95 96 101 103 108 140
27 30 33 40 43 59 60 62 64 67 71 92 95 96 101

E. Estrada, Biophys. J. 98, 2010, 890-900
Spectral Scaling at Different Resolutions

$r_c = 5.0\,\text{Å}$

$r_c = 10.0\,\text{Å}$

$r_c = 7.0\,\text{Å}$

$r_c = 15.0\,\text{Å}$
Can we Identify Network Communities?  
A Communicability Approach

\[ G_{pq} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} \] (# of walks in \( n \) steps)

\[ = \sum_{n=0}^{\infty} \frac{A^n_{pq}}{n!} = (e^A)_{pq} \]

\[ G_{pq} = (e^A)_{pq} = \sum_{\mu=1}^{N} e^{\lambda_{\mu}} \phi_{\mu}(p)\phi_{\mu}(q) \]

\[ G_{pq} = \left( e^A \right)_{pq} = \sum_{\mu=1}^{N} e^{\lambda_{\mu}} \phi_{\mu}(p)\phi_{\mu}(q) \]

\[ G_{pq} = \phi_1(p)\phi_1(q)e^{\lambda_1} + \sum_{j \geq 2}^{++} \phi_j(p)\phi_j(q)e^{\lambda_j} + \sum_{j \geq 2}^{--} \phi_j(p)\phi_j(q)e^{\lambda_j} \]

\[ + \sum_{j \geq 2}^{+-} \phi_j(p)\phi_j(q)e^{\lambda_j} + \sum_{j \geq 2}^{-+} \phi_j(p)\phi_j(q)e^{\lambda_j} \]
$\varphi_1(p)$ for $\lambda_1 = 3.734644$

$\varphi_2(p)$ for $\lambda_2 = 2.952552$

$\varphi_3(p)$ for $\lambda_3 = 2.651093$
Eigenvector Expansion

\[ G_{pq} = \phi_1(p)\phi_1(q)e^{\lambda_1} + \sum_{j \geq 2}^{++} \phi_j(p)\phi_j(q)e^{\lambda_j} + \sum_{j \geq 2}^{--} \phi_j(p)\phi_j(q)e^{\lambda_j} \]

+ \sum_{j \geq 2}^{+-} \phi_j(p)\phi_j(q)e^{\lambda_j} + \sum_{j \geq 2}^{-+} \phi_j(p)\phi_j(q)e^{\lambda_j} \]

A. Translational motion
B. Intra-cluster communicability
C. Inter-cluster communicability
Community: group of nodes for which the **intracluster** communicability is larger than the **intercluster** one.

\[ \Delta G_{pq} = \sum_{j \geq 2} \varphi_j(p)\varphi_j(q)e^{\lambda_j} + \sum_{j \geq 2} \varphi_j(p)\varphi_j(q)e^{\lambda_j} \]
Social Communities

Karate Club

Communicability graph
Social Communities: All-cliques
Social Communities: Similarity
Bottlenose dolphins

Network
Thermal Green's function

\[ G_{pq}(\beta) = \left( e^{\beta A} \right)_{pq} = \sum_{\mu=1}^{N} e^{\beta \lambda_{\mu}} \varphi_{\mu}(p)\varphi_{\mu}(q) \]

\[ \Delta G_{pq}(\beta) = \sum_{\mu=2}^{N} e^{\beta \lambda_{\mu}} \varphi_{\mu}(p)\varphi_{\mu}(q) \]

\[ = \left( \sum^{++} + \sum^{--} - \sum^{+-} - \sum^{-+} \right) e^{\beta \lambda_{\mu}} |\varphi_{\mu}(p)\varphi_{\mu}(q)| \]
Variable Temperature

$\beta = 0.0105$

$\beta = 20.00$
Social Network Example

$\beta = 1.0$
Social Network Example

\[ \beta = 0.3 \]
$\beta = 0.2$
Social Network Example

$$\beta = 0.1$$
Negative temperatures

Communicability as a Classifier

Healthy

Months after Stroke
DTI exploits the fact that water diffusion in the brain is anisotropic; in particular, diffusion along tracts is greater than diffusion across them.
9 subjects - suffered subcortical strokes;

10 controls.

DTI computes all connections between all voxels.
Discrimination Analysis


\[ W = \tilde{A} = D^{-1/2} AD^{-1/2} \]

```
(a) Raw Data

(b) Communicability

(c) healthy

(d) stroked
```
Communicability in damaged brains

Red: Stroke  Blue: Decreased communicability

Red: Stroke  Green: Increased communicability

Images courtesy of Dr. Jonathan J. Crofts.
The End
THANK YOU!

More info at:

www.estradalab.org