E6885 Network Science Lecture 2: 
*Network Representations and Characteristics*

Ching-Yung Lin, Dept. of Electrical Engineering, Columbia University

September 16th, 2013
## Course Structure

<table>
<thead>
<tr>
<th>Class Date</th>
<th>Lecture</th>
<th>Topics Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/09/13</td>
<td>1</td>
<td>Overview of Network Science</td>
</tr>
<tr>
<td>09/16/13</td>
<td>2</td>
<td>Network Representation and Feature Extraction</td>
</tr>
<tr>
<td>09/23/13</td>
<td>3</td>
<td>Network Partitioning, Clustering and Visualization</td>
</tr>
<tr>
<td>09/30/13</td>
<td>4</td>
<td>Graph Database</td>
</tr>
<tr>
<td>10/07/13</td>
<td>5</td>
<td>Network Sampling, Estimation, and Modeling</td>
</tr>
<tr>
<td>10/14/13</td>
<td>6</td>
<td>Network Analysis Use Case</td>
</tr>
<tr>
<td>10/21/13</td>
<td>7</td>
<td>Network Topology Inference and Prediction</td>
</tr>
<tr>
<td>10/28/13</td>
<td>8</td>
<td>Graphical Model and Bayesian Networks</td>
</tr>
<tr>
<td>11/11/13</td>
<td>9</td>
<td>Final Project Proposal Presentation</td>
</tr>
<tr>
<td>11/18/13</td>
<td>10</td>
<td>Dynamic and Probabilistic Networks</td>
</tr>
<tr>
<td>11/25/13</td>
<td>11</td>
<td>Information Diffusion in Networks</td>
</tr>
<tr>
<td>12/02/13</td>
<td>12</td>
<td>Impact of Network Analysis</td>
</tr>
<tr>
<td>12/09/13</td>
<td>13</td>
<td>Large-Scale Network Processing System</td>
</tr>
<tr>
<td>12/16/13</td>
<td>14</td>
<td>Final Project Presentation</td>
</tr>
</tbody>
</table>
TA of the course

- Xiao-Ming Wu <xw2223>; Office Hours: Friday 2-4pm (7LE3 Schapiro Building)
The fundamental connectivity of a graph $G$ may be captured in an $N_v \times N_v$ binary symmetric matrix $A$ with entries:

$$A_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in E \\ 0, & \text{otherwise} \end{cases}$$

$A$ is called the Adjacency Matrix of $G$.
Some properties of adjacency matrix

- The row sum is equal to the degree of vertex.

\[ d_i = A_{i+} = \sum_j A_{ij} \]

- Symmetry:

\[ A_{i+} = A_{+i} \]

- Number of walks of length \( r \) from the \( r \)-th power of \( A : A^r \)

\[ A_{ij}^r \]
For directional graph

- \{i,j\} represents a directed edge from \( i \) to \( j \).

- In and Out degrees:

\[
A_{i+} = d_i^{\text{out}} \quad \quad A_{+j} = d_j^{\text{in}}
\]
Algorithms

▪ Some questions:
  – Are vertex $i$ and $j$ linked by an edge?
  – What is the degree of vertex $i$?
  – What is the shortest path(s) between vertex $i$ and $j$?
  – How many connected component does the graph have?
  – (for a directed graph,) does it have cycles or is it acyclic?
  – What is the maximal clique in a graph?
Algorithmic Complexity

- ‘Tractable’: Polynomial Time:
  \[ O(n^p) \quad p > 0 \]

- ‘Intractable’: Super-Polynomial Time, e.g.:
  \[ O(a^n) \quad a > 1 \]

The design of efficient algorithms is usually nontrivial.

Usually improved by indexing, storage, removing redundant computations, etc.
Example – Finding all vertices that are reachable from a vertex

- Breadth-First Search (BFS)
- Depth-First Search (DFS)
Some questions to ask:

- Triplets of vertices (triads) in social dynamics
- Paths and flows in graph
- Importance of individual element => how ‘central’ the corresponding vertex is in the network
- Finding communities

Characteristics of individual vertices

Characteristics of network cohesion
Centrality

- “There is certainly no unanimity on exactly what centrality is or its conceptual foundations, and there is little agreement on the procedure of its measurement.” – Freeman 1979.

- Degree (centrality)
- Closeness (centrality)
- Betweenness (centrality)
- Eigenvector (centrality)
Degree Distribution Example: Power-Law Network


\[ p_k = e^{-m} \cdot \frac{m^k}{k!} \]

Newman, Strogatz and Watts, 2001

\[ p_k = C \cdot k^{-\tau} e^{-k/\kappa} \]
Another example of complex network: Small-World Network

- Six Degree Separation:
  - adding long range link, a regular graph can be transformed into a small-world network, in which the average number of degrees between two nodes become small.

from Watts and Strogatz, 1998
Indication of ‘Small’

- A graph is ‘small’ which usually indicates the average distance between distinct vertices is ‘small’

\[
\bar{d} = \frac{1}{N_v(N_v + 1)/2} \sum_{u \neq v \in V} dist(u, v)
\]

For instance, a protein interaction network would be considered to have the small-world property, as there is an average distance of 3.68 among the 5,128 vertices in its giant component.
Some examples of Degree Distribution

- (a) scientist collaboration: biologists (circle) physicists (square), (b) collaboration of move actors, (d) network of directors of Fortune 1000 companies
Degree Distribution

Fig. 4.1 Degree distributions. Left: the router-level Internet network graph described in Section 3.5.2. Right: the network of measured interactions among proteins in *S. cerevisiae* (yeast), as of January 2007. In each plot, both $x$- and $y$-axes are in base-2 logarithmic scale.

Degree Correlations

**Fig. 4.3** Image representation of the logarithmically transformed joint degree distribution $\{\log_2 f_{d,d'}\}$, for the router-level Internet data (left) and the protein interaction data (right). Colors range from blue (low relative frequency) to red (high relative frequency), with white indicating areas with no data. Note that both x- and y-axes are also on base-2 logarithmic scales.

Conceptual Descriptions of Three Centrality Measurements

Fig. 4.4 Illustration of (b) closeness, (c) betweenness, and (d) eigenvector centrality measures on the graph in (a). Example and figures courtesy of Ulrik Brandes.
Closeness

- Closeness: A vertex is ‘close’ to the other vertices

\[ c_{CI}(v) = \frac{1}{\sum_{u \in V} \text{dist}(v,u)} \]

where \( \text{dist}(v,u) \) is the geodesic distance between vertices \( v \) and \( u \).
Betweenness

- Betweenness measures are aimed at summarizing the extent to which a vertex is located ‘between’ other pairs of vertices.

- Freeman’s definition:

\[ c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t | v)}{\sigma(s, t)} \]

- Calculation of all betweenness centralities requires
  - calculating the lengths of shortest paths among all pairs of vertices
  - Computing the summation in the above definition for each vertex
Eigenvector Centrality

- Try to capture the ‘status’, ‘prestige’, or ‘rank’.
- More central the neighbors of a vertex are, the more central the vertex itself is.

\[
c_{Ei}(v) = \alpha \sum_{\{u,v\} \in E} c_{Ei}(u)
\]

The vector \( c_{Ei} = (c_{Ei}(1), \ldots, c_{Ei}(N_v))^T \) is the solution of the eigenvalue problem:

\[
A \cdot c_{Ei} = \alpha^{-1} c_{Ei}
\]
PageRank Algorithm (Simplified)
PageRank Steps

\[ R(u) = d \sum_{v \in B_u} \frac{R(u)}{N_v} + e \]

where

\[ F_u \quad \text{The set of pages } u \text{ points to} \]

\[ B_u \quad \text{The set of pages point to } u \]

\[ N_u = |F_u| \quad \text{Number of links from } u \]

\[ d \quad \text{Normalization / damping factor} \]

\[ e = \frac{1-d}{N} \quad \text{In general, } d=0.85 \]

- Example: Simplified Initial State:
  \[ R(A) = R(B) = R(C) = R(D) = 0.25 \]

- Iterative Procedure:
  \[ R(A) = R(B) / 2 + R(C) / 1 + R(D) / 3 \]
Solution of PageRank

- The PageRank values are the entries of the dominant eigenvector of the modified adjacency matrix.

\[
\begin{bmatrix}
R(p_1) \\
R(p_2) \\
\vdots \\
R(p_N)
\end{bmatrix}
\]

where \( R \) is the solution of the equation

\[
\begin{bmatrix}
\frac{(1-d)}{N} & l(p_1, p_1) & l(p_1, p_1) & \cdots & l(p_1, p_N) \\
\frac{(1-d)}{N} & l(p_2, p_1) & \cdots & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{(1-d)}{N} & l(p_N, p_1) & \cdots & \cdots & l(p_N, p_N)
\end{bmatrix}
\begin{bmatrix}
R(p_1) \\
R(p_2) \\
\vdots \\
R(p_N)
\end{bmatrix}
\]

where \( R \) is the adjacency function \( l(p_i, p_j) = 0 \) if page \( p_j \) does not link to \( p_i \), and normalized such that for each \( j \),

\[
\sum_{i=1}^{N} l(p_i, p_j) = 1
\]
Example: Brain of an epilepsy patient

Fig. 4.10 Left: Three-dimensional reconstruction of the brain of an epilepsy patient, with EC\textit{o}G implant grid super-imposed. Right: Example of an EC\textit{o}G time series at one electrode, both for one seizure (bottom) and after smoothing (i.e., bandpass filtering) and averaging over eight such seizures (top). Preictal (blue) and ictal (red) periods are indicated between pairs of parallel lines.
Network Representation of Cortical-level Coupling

*Fig. 4.11* Network representations of cortical-level coupling between brain regions about each electrode, during preictal (left) and ictal (right) periods.
Visual Summaries of Degree and Closeness Centrality

preictal    \hspace{1cm} \text{diff}    \hspace{1cm} \text{ictal}
Visual Summaries of Betweenness Centrality and Clustering Coefficient

preictal  diff  ictal
Connectivity of Graph

- A measure related to the flow of information in the graph
- Connected $\Rightarrow$ every vertex is reachable from every other
- A connected component of a graph is a maximally connected subgraph.
- A graph usually has one dominating the others in magnitude $\Rightarrow$ giant component.
Network Cohesion

- Questions to answer:
  - Do friends of a given actor in a SN tend to be friends of one another?
  - What collections of proteins in a cell appear to work closely together?
  - Does the structure of the pages in the WWW tend to separate with respect to distinct types of content?
  - What portion of a measured Internet topology would seem to constitute the backbone?

- Definitions differ in
  - Scale
  - Local to Global
  - Explicitly (.e.g., cliques) vs implicitly (e.g. clusters)
Local Density

- A coherent subset of nodes should be locally dense.
- Cliques:

A sufficient condition for a clique of size \( n \) to exist in \( G \) is:

\[
N_e > \left( \frac{N_v^2}{2} \right) \left( \frac{n - 2}{n - 1} \right)
\]
Weakened Versions of Cliques -- Plexes

- A subgraph $H$ consisting of $m$ vertices is called $n$-plex, for $m > n$, if no vertex has degree less than $m - n$.

\[ 1\text{-plex} \rightarrow \text{No vertex is missing more than 1 of its possible } m-1 \text{ edges.} \]
Another Weakened Versions of Cliques -- Cores

- A \textit{k-core} of a graph \( G \) is a subgraph \( H \) in which all vertices have degree at least \( k \).

- Batagelj et. al., 1999. A maximal \textit{k-core} subgraph may be computed in as little as \( O( N_v + N_e) \) time.

  Computes the shell indices for every vertex in the graph

  \textit{Shell index of} \( v = \text{the largest value, say} \ c, \text{such that} \ v \text{ belongs to the} \ c\text{-core of} \ G \text{ but not its} \ (c+1)\text{-core.} \)

  For a given vertex, those neighbors with lesser degree lead to a decrease in the potential shell index of that vertex.
Density measurement

- The density of a subgraph $H = (V_H, E_H)$ is:

$$
\text{den}(H) = \frac{|E_H|}{|V_H|\left(|V_H| - 1\right)/2}
$$

Range of density

$$
0 \leq \text{den}(H) \leq 1
$$

and

$$
\text{den}(H) = (|V_H| - 1)\bar{d}(H)
$$

average degree of $H$
Use of the density measure

- Density of a graph: let $H = G$
- ‘Clustering’ of edges local to $v$: let $H = H_v$, which is the set of neighbors of a vertex $v$, and the edges between them
- **Clustering Coefficient** of a graph: The average of $\text{den}(H_v)$ over all vertices
An insight of clustering coefficient

- A triangle is a complete subgraph of order three.
- A connected triple is a subgraph of three vertices connected by two edges (regardless how the other two nodes connect).
- The local clustering coefficient can be expressed as:

\[
\text{den}(H_v) = \text{cl}(v) = \frac{\tau_\Delta(v)}{\tau_3(v)}
\]

- The clustering coefficient of G is then:

\[
\text{cl}(G) = \frac{1}{V'} \sum_{v \in V'} \text{cl}(v)
\]

Where \( V' \subseteq V \) is the set of vertices \( v \) with \( d_v \geq 2 \).
An example

\[ C_{n_1} = \frac{\text{number of triangles connected to node } n_1}{\text{number of triples centered on node } n_1} = \frac{1}{6} \]

Triangles = 1

Triples = 6
Transitivity of a graph

- A variation of the clustering coefficient \( \Rightarrow \) takes weighted average

\[
cl_T(G) = \frac{\sum_{v \in V'} \tau_3(v) cl(v)}{\sum_{v \in V'} \tau_3(v)} = \frac{3\tau_\Delta(G)}{\tau_3(G)}
\]

where

\[
\tau_\Delta(G) = \frac{1}{3} \sum_{v \in V} \tau_\Delta(v)
\]

is the number of triangles in the graph

\[
\tau_3(G) = \sum_{v \in V} \tau_3(v)
\]

is the number of connected triples

\( \Rightarrow \) The friend of your friend is also a friend of yours

Clustering coefficients have become a standard quantity for network structure analysis. But, it is important on reporting which clustering coefficients are used.
Vertex / Edge Connectivity

- If an arbitrary subset of $k$ vertices or edges is removed from a graph, is the remaining subgraph connected?

- A graph $G$ is called $k$-vertex-connected, if (1) $N_v > k$, and (2) the removal of any subset of vertices $X$ in $V$ of cardinality $|X|$ smaller than $k$ leaves a subgraph $G - X$ that is connected.

The vertex connectivity of $G$ is the largest integer such that $G$ is $k$-vertex-connected.

- Similar measurement for edge connectivity
Vertex / Edge Cut

- If the removal of a particular set of vertices in $G$ disconnects the graph, that set is called a vertex cut.

- For a given pair of vertices $(u,v)$, a $u$-$v$-cut is a partition of $V$ into two disjoint non-empty subsets, $S$ and $S'$, where $u$ is in $S$ and $v$ is in $S'$.

Minimum $u$-$v$-cut: the sum of the weights on edges connecting vertices in $S$ to vertices in $S'$ is a minimum.
Minimum cut and flow

- Find a minimum u-v-cut is an equivalent problem of maximizing a measure of flow on the edges of a derived directed graph.
Graph Partitioning

- Many uses of graph partitioning:
  - E.g., community structure in social networks

- A cohesive subset of vertices generally is taken to refer to a subset of vertices that
  - (1) are well connected among themselves, and
  - (2) are relatively well separated from the remaining vertices

- Graph partitioning algorithms typically seek a partition of the vertex set of a graph
  in such a manner that the sets $E(C_k, C_{k'})$ of edges connecting vertices in $C_k$ to
  vertices in $C_{k'}$ are relatively small in size compared to the sets $E(C_k) = E(C_k, C_{k'})$ of
  edges connecting vertices within $C_{k'}$. 
Classify the nodes

Fig. 4.6 AIDS blog network, from Chapter 1, with nodes colored according to their membership in the ‘bowtie’ decomposition: strongly connected component (yellow), in-component (blue), out-component (red), and tendrils (pink).
Example: AIDS blog network
Example: Karate Club Network
Hierarchical Clustering

- Agglomerative
- Divisive

In agglomerative algorithms, given two sets of vertices $C_1$ and $C_2$, two standard approaches to assigning a similarity value to this pair of sets is to use the maximum (called single-linkage) or the minimum (called complete linkage) of the similarity $x_{ij}$ over all pairs.

\[ x_{ij} = \frac{\left| N_{v_i} \Delta N_{v_j} \right|}{d(N_v) + d(N_v - 1)} \]

The “normalized” number of neighbors of $v_i$ and $v_j$ that are not shared.
Hierarchical Clustering Algorithms Types

- Primarily differ in [Jain et. al. 1999]:
  - (1) how they evaluate the quality of proposed clusters, and
  - (2) the algorithms by which they seek to optimize that quality.

- Agglomerative: successive coarsening of partitions through the process of merging.

- Divisive: successive refinement of partitions through the process of splitting.

- At each stage, the current candidate partition is modified in a way that minimize a specific measure of cost.

- In agglomerative methods, the least costly merge of two previously existing partition elements is executed.

- In divisive methods, it is the least costly split of a single existing partition element into two that is executed.
Hierarchical Clustering

- The resulting hierarchy typically is represented in the form of a tree, called a dendrogram.

- The measure of cost incorporated into a hierarchical clustering method used in graph partitioning should reflect our sense of what defines a ‘cohesive’ subset of vertices.

- In agglomerative algorithms, given two sets of vertices $C_1$ and $C_2$, two standard approaches to assigning a similarity value to this pair of sets is to use the maximum (called single-linkage) or the minimum (called complete linkage) of the dissimilarity $x_{ij}$ over all pairs.

- Dissimilarities for subsets of vertices were calculated from the $x_{ij}$ using the extension of Ward (1963)’s method and the lengths of the branches in the dendrogram are in relative proportion to the changes in dissimilarity.

$$x_{ij} = \frac{|N_{v_i} \Delta N_{v_j}|}{d(N_v) + d(N_v - 1)}$$

- $N_v$ is the set of neighbors of a vertex.
- $\Delta$ is the symmetric difference of two sets which is the set of elements that are in one or the other but not both.

$x_{ij}$ is the “normalized” number of neighbors of $v_i$ and $v_j$ that are not shared.
Other dissimilarity measures

- There are various other common choices of dissimilarity measures, such as:

\[ x_{ij} = \sqrt{\sum_{k \neq i, j} (A_{ik} - A_{jk})^2} \]

- Hierarchical clustering algorithms based on dissimilarities of this sort are reasonably efficient, running in \( O(N_v^2 \log N_v) \) time.
Hierarchical Clustering Example

Fig. 4.7 Hierarchical clustering of the karate club network.
Questions?