Theory and Applications of Complex Networks

Class Two

College of the Atlantic

David P. Feldman

12 September 2008

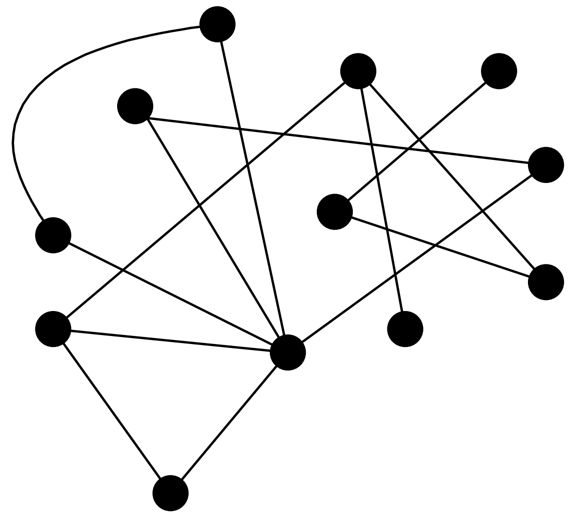
`http://hornacek.coa.edu/dave/`

1. Representing networks
2. Variations on networks
3. Basic structural properties of networks

What is a Network?

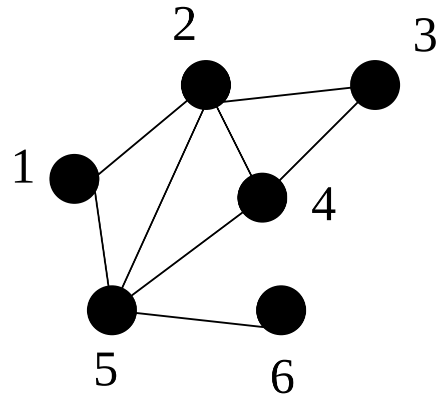
1. A collection of **nodes**

2. A collection of **edges** connecting nodes



- Let $N =$ number of nodes.
- Let $M =$ number of links or edges.
- Networks are also knows as graphs, particularly among mathematicians.

Network Representation: Adjacency Matrix

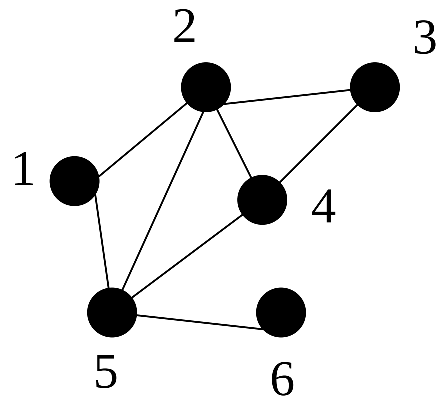


- Adjacency matrix A: $A_{ij} = 1$ if there is a link between nodes i and j. Otherwise $A_{ij} = 0$.
- For the graph shown above:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (1)$$

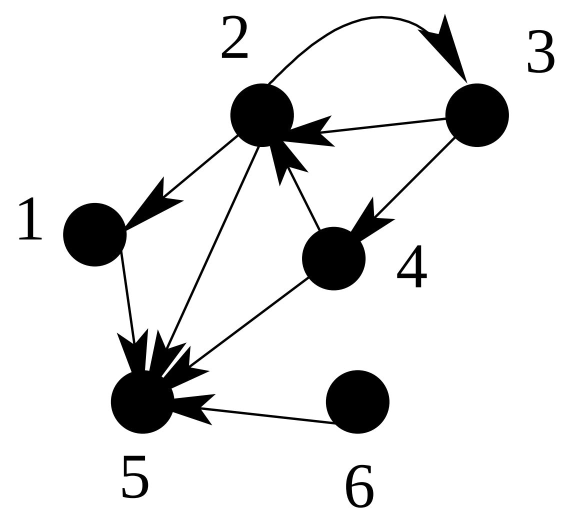
- Note that A is symmetric

Adjacency Matrix vs. Lists



- For the networks we will consider the adjacency matrix is usually *sparse*. I.e., it has lots of zeros.
- This means that it is an inefficient representation because we waste memory keeping track of a vast number of zeros.
- An alternative is to simply list the links by referring to the nodes they connect
- Ex: $(1, 2)$, $(1, 5)$, $(2, 3)$, $(2, 4)$, $(2, 5)$, $(3, 4)$, $(4, 5)$, $(5, 6)$.

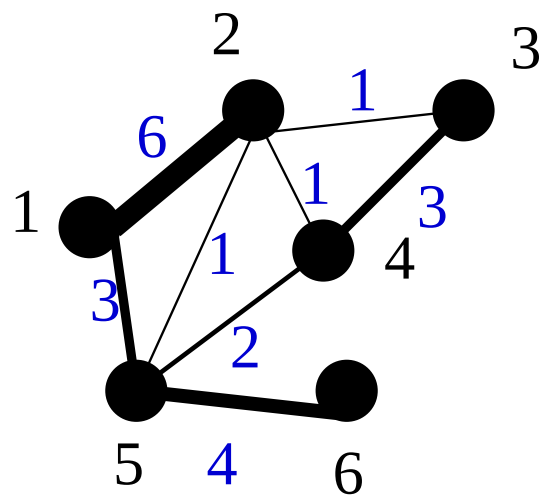
Variation: Directed Network



- Links have direction. Adjacency matrix is no longer symmetric.
-

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{2}$$

Variation: Weighted Network



- Links have weights, indicating differint strengths of connection. Adjacency matrix is no longer all 1's and 0's
-

$$A = \begin{pmatrix} 0 & 6 & 0 & 0 & 3 & 0 \\ 6 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 3 & 0 & 2 & 0 \\ 1 & 3 & 0 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & 4 & 0 \end{pmatrix} \quad (3)$$

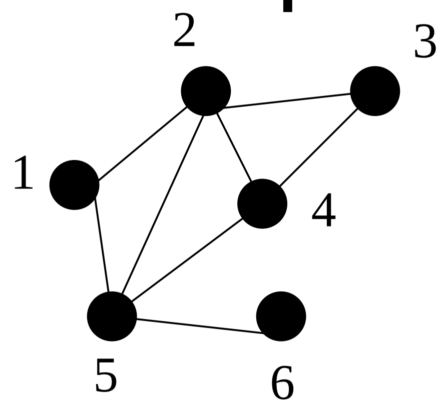
Variation: Heterogenous Networks

- These are networks where there are different kinds of nodes and/or different kinds of links.
- One common example is a graph in which there are two types of nodes, where nodes can only be connected to nodes of the other type.
- These types of networks are known as *bipartite graphs*.
- Example: A network of senators and corporations, where corporations are connected to senators via donations.

Basic Network Properties

- Given a network, what are some useful ways of describing its connectivity, organization, structure, etc?
- Today, some basic and (mostly) quite standard definitions.
- I will focus on regular networks, but most of the quantities generalize fairly naturally to directed and/or weighted networks.
- You should be aware that there is not yet a common set of notation for most of these quantities. Different authors and different communities use different conventions.

Basic Network Properties: Degree

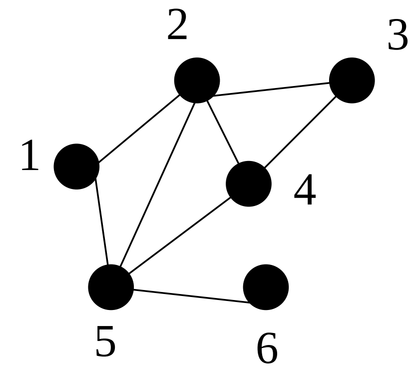


- The *degree* k of a node is the number of links connected to it.
- The degree is sometimes called *coordination number* and denoted with z. This is mostly a physics convention.
- Ex: $k_1 = 2$, $k_2 = 4$, $k_6 = 1$.
- Often we are interested in the average degree of all the nodes.
- This is often denoted k or $\langle k \rangle$. The latter is called "the expectation value of k."
- For this graph, $k = 2.67$.
- There is a hard and an easy way to calculate k.

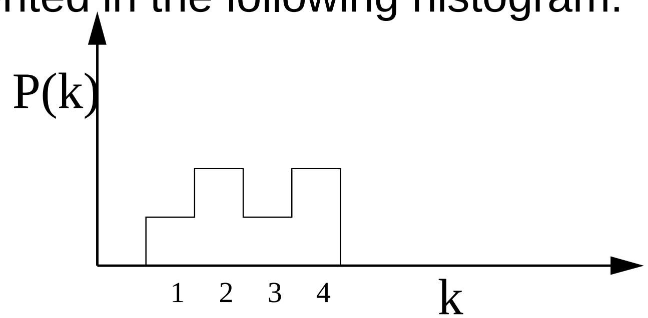
Basic Network Properties: Degree Distribution

- We are usually interested in more than just the average degree.
- Are some nodes more connected than others? How much variance is there about the mean degree?
- For that matter, is the notion of an average degree or variance even meaningful?
- These questions can be addressed by looking at the *degree distribution*.
- $P(k)$ is the probability that a randomly chosen node has degree k.

Basic Network Properties: Degree Distribution



- There is one node with degree 1, two nodes with degree 2, one node with degree 3, and two nodes with degree 4.
- This can be represented in the following histogram:



- Later in the course we will examine in considerable detail different function forms for $P(k)$ and what they tell us.

Basic Network Properties: Distance and Diameter

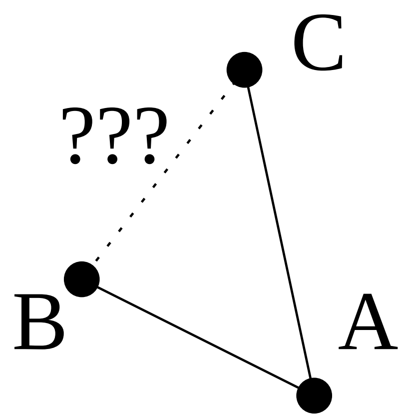
- *Distance* d_{ij} between nodes i and j
- $d_{ij} =$ # of links along shortest path connecting i and j.
- This is also denoted $d(i, j)$ or $\delta(i, j)$.
- This is sometimes referred to as the *geodesic distance*.
- The shortest path is called a *geodesic*.
- The mean distance ℓ is the average of the d_{ij}'s.
- Apparently there is not an entirely standard way to do this average; sometimes the self-distances ($d_{ii} = 0$) are included and sometimes they are not.
- For large networks it doesn't matter too much.

Basic Network Properties: Distance and Diameter

- ℓ may be thought of as a measure of the size of the network.
- The diameter d of a graph is defined to be the distance of the longest geodesic.
- $d = \max_{ij} d_{ij}$.
- The diameter is another measure of the size of the network.
- A network is said to have the “small world” property if ℓ grows no faster than the log of the number of nodes: $\ell \sim \log(N)$.
- More on small-world graphs later in the course.

Network Properties: Clustering and Transitivity

- To what extent are your friends friends with each other?
- There are two clustering measures that quantify the tendence of friends to be friends.



- There are two common ways to measure clustering.
- The following discussion closely follows Newman, “Structure and Function of Complex Networks,” 2003.

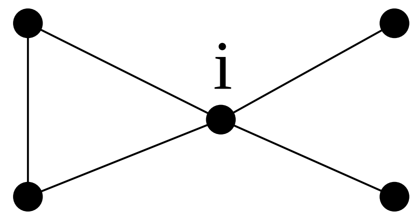
Clustering and Transitivity: Method Two

- Consider a node i of degree k.
- Let e_i denote the number of edges among i's k neighbors.
- Max # of links that could exist among these k neighbors $= \frac{1}{2}k(k-1)$.
- Think about this until it makes sense.
- The the cluster coefficient $C_i^{(2)}$ for site i is:

$$C_i^{(2)} = \frac{e_i}{\frac{1}{2}k(k-1)} = \frac{2e_i}{k(k-1)} . \tag{4}$$

- $C_i^{(2)} =$ friends among i's friends as a fraction of the total possible number of friends among i's friends.
- The average clustering coefficient is denoted $C^{(2)}$ and is defined in the natural way.

Clustering and Transitivity: Method Two: Example



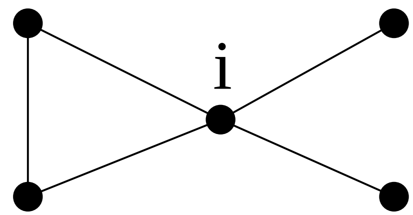
$$C_i^{(2)} = \frac{e_i}{\frac{1}{2}k(k-1)} = \frac{2e_i}{k(k-1)} \; . \qquad (5)$$

- For node i, $k = 4$.
- $e_i =$ # of edges among i's neighbors $= 1$. neighbors.
- Plugging in, we get

$$C_i(2) = \frac{2 \times 1}{4 \times 3} = \frac{1}{6} \; . \qquad (6)$$

- btw, the (2) upstairs on the C's references the fact that this is the second way of defining C. This notation isn't standard. Usually a paper only uses one of the two definitions for C and you should be careful to make sure which one it's using.

Clustering and Transitivity: Method Two Alternative

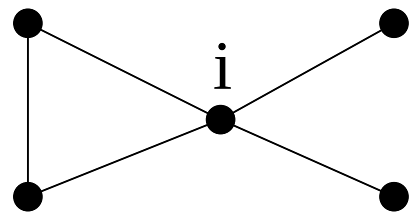


- Here is another formula for $C_i^{(2)}$:

$$C_i^{(2)} = \frac{\#\text{ of triangles connected to vertex } i}{\#\text{ of triples centered on } i} \, . \tag{7}$$

- For this example, we get $C_i^{(2)} = \frac{1}{6}$ just like before.
- It might take a few moments of quiet contemplation to convince yourself that these two formulae are equivalent.

Clustering and Transitivity: Method One



- Here is an different definition for the cluster coefficient.

$$C_{(1)} = \frac{3 \times \#\text{ of triangles}}{\#\text{ of connected triples}} . \qquad (8)$$

- For this example, we get $C^{(1)} = \frac{3 \times 1}{8}$.
- Note that this is a property of the entire network, not a single node.
- $C^{(1)} =$ fraction of transitive triples.
- I believe that this definition is more common in sociology. The previous definition is more common in physics.

Network Properties: Which Nodes are the Most Important?

- Which nodes are the most important in a network?
- What different roles might nodes play?
- How are these different roles distributed among the nodes?
- Measures of importance of a node are often called *centrality*.
- There are several different notions of centrality. Some of the following definitions are more standard than others.
- The following several slides follows Mason and Verwoerd, section four.

Network Properties: Degree Centrality

- Key Idea: An important node is involved in many interactions.
- The degree centrality of a node is simply its degree.
- Thus, under this line of reasoning, the most important node is the one with the most connections.

Network Properties: Closeness Centrality

- Key Idea: An important node is close to lots of other nodes.
- The *excentricity* of node j:

$$C_e(j) \;=\; \max_i d_{ij} \;. \tag{9}$$

- I.e., $C_e(j)$ is the distance from j to the node that is furthest away from j.
- Another interesting notion is the *center* of the graph. This is given by the set of points that are closest to everybody else:

$$\mathcal{C} \;=\; \{i : C_e(i) = \min_j C_e(j)\} \;. \tag{10}$$

- The above equation just says that the center is the middle of the graph.

Network Properties: Betweenness Centrality

- Key Idea: An important node connects lots of other nodes. I.e., an important node will be on a high proportion of paths between other nodes.
- To calculate $C_b(i)$, the betweenness centrality for node i:
  1. Consider all pairs of nodes $j, k \neq i$.
  2. Determine the shortest path between all such j, k.
  3. Then $C_b(i) =$ fraction of those paths which go through i.

Network Properties: Eigenvalue Centrality

- Key Idea: An important nodes are connected to many other important nodes
- Details later in the course.

A Few References

- Mark E.J. Newman, The Structure and function of complex networks. SIAM Review, 45(2), 167-256. 2003.
- O. Masin and M. Verwoerd, Graph Theory and Netowrks in Biology. 2006.
- Skye Bender-deMoll, Potential Human Rights Uses of Network Analysis and Mapping: A report to the Science and Human Rights Program of the AAAS, 2008.
- Duncan J. Watts, The "New" Science of Networks, Annual Review of Sociology 30:243-270, 2004.
- C. Gros, Complex and Adaptive Dynamical Systems: A primer, forthcoming Springer 2008. Chapter 1 is on networks.

- All except are available online.