Network Science

Barabási: Ch. 2 — Graph Theory — Lecture 2

Joao Meidanis

University of Campinas, Brazil

September 26, 2020

Summary

Brief Statistics Review

Average, moments, standard deviation

For a sample of N values $x_1, x_2, \ldots, x_N$:

- Average (mean):

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- The n^{th} moment:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i^n$$

- Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2}$$

Distributions

For a sample of N values $x_1, x_2, \ldots, x_N$:

- Distribution:

$$p_x = \frac{1}{N} \sum_{i=1}^{N} \delta(x, x_i)$$

where the Kronecker δ is defined as

$$\delta(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$$

- We have:

$$\sum_x p_x = 1$$

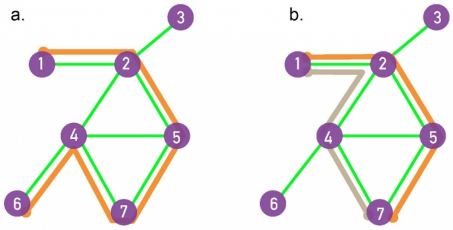
- Continuous case (density function f):

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Paths and Distances

Paths and Length

- Physical distance usually irrelevant in networks:
  - a webpage can link to others very far away
  - two neighbors may not know each other
- Definition: a **path** is a route following network links (some texts require distinct nodes)
- Path **length**: number of links traversed



Shortest Paths, Distance, Diameter

- **Shortest path** from i to j: smallest number of links
- d_{ij} = **distance** from i to j = length of a shortest path from i to j
- Undireted network: $d_{ij} = d_{ji}$
- Directed network: often $d_{ij} \neq d_{ji}$
- Directed network: existence of $i \to j$ path **does not guarantee** existence of $j \to i$ path
- Computing distances:
  - powers of adjacency matrix — **good to know**
  - BFS (breadth first search) algorithm — **fast** — **good to run**
- d_{max} = **diameter** = maximum distance in network
- Average distance (connected graph):

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} = \frac{1}{2L_{max}} \sum_{i \neq j} d_{ij}$$

Number of Paths

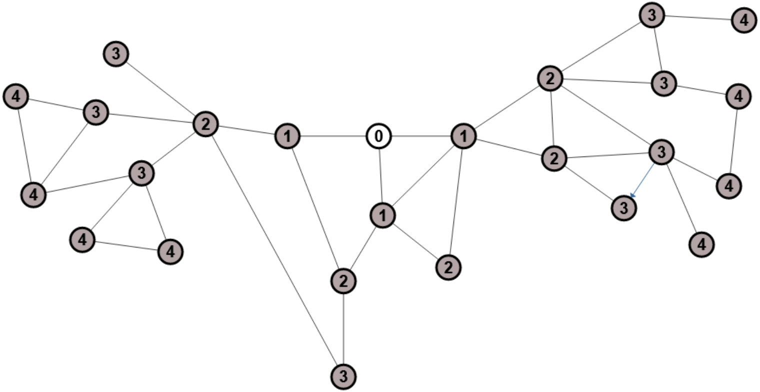
- $N_{ij}^{(k)}$ = number of length-k paths from i to j
- Can be computed from adjacency matrix A_{ij}
- There is a link from i to j if and only if $A_{ij} = 1$
- Then $N_{ij}^{(1)} = A_{ij}$
- There is a length-2 path from i to j if and only if there is k such that $A_{ik}A_{kj} = 1$
- The number of such paths is $N_{ij}^{(2)} = \sum_k A_{ik}A_{kj} = A_{ij}^2$
- And so on. In general

$$N_{ij}^{(k)} = A_{ij}^k$$

Breadth First Search (BFS)

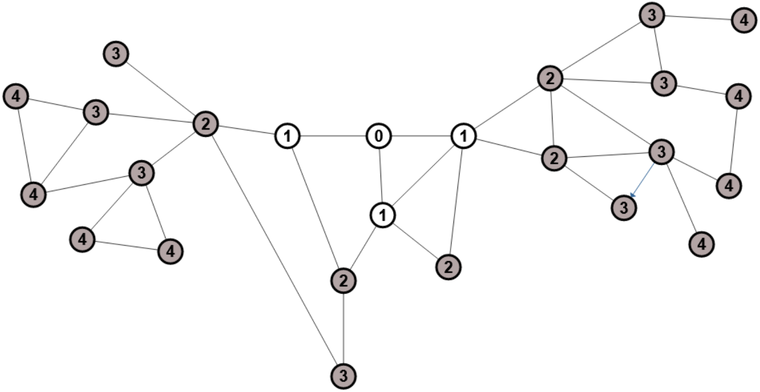
Breadth First Search (BFS)

algorithm: step 0

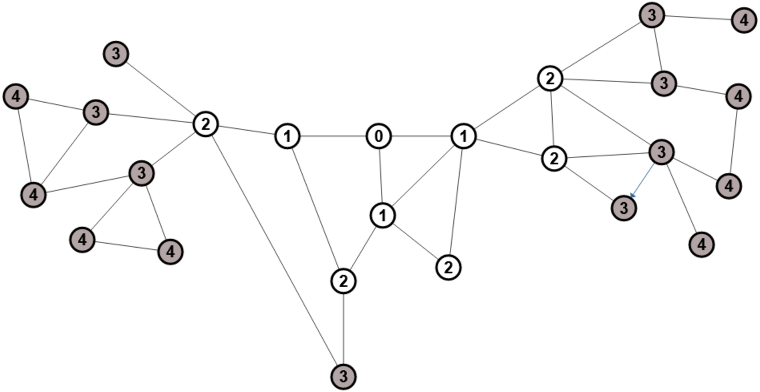


Breadth First Search (BFS)

algorithm: step 1

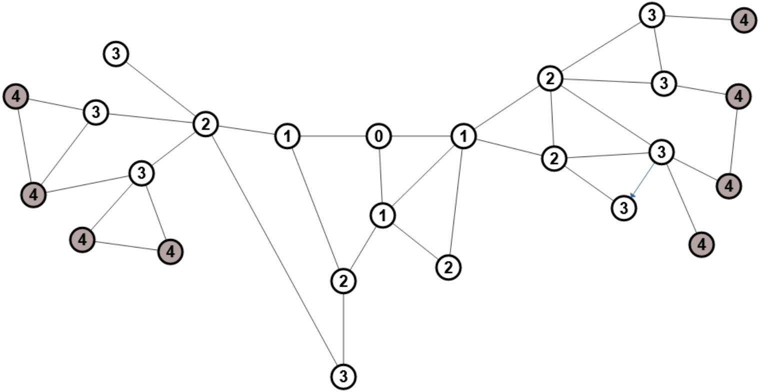


algorithm: step 2



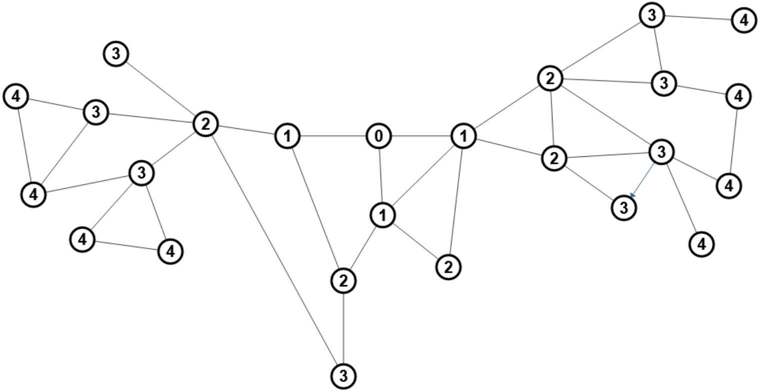
Breadth First Search (BFS)

algorithm: step 3



Breadth First Search (BFS)

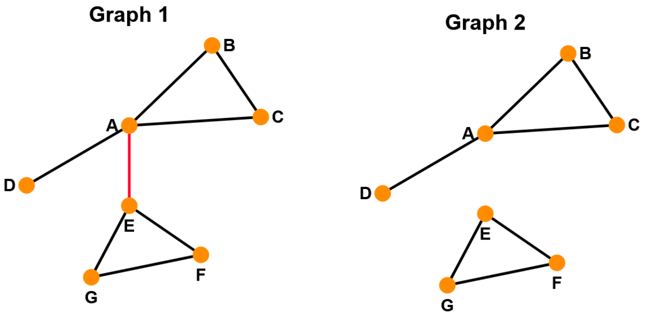
algorithm: step 4



Connectivity

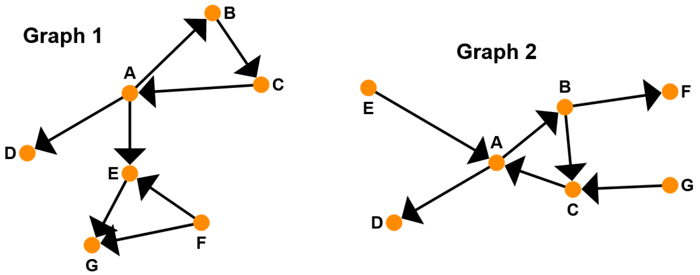
Connectivity for Undirected Graphs

- **Connected graph**: any two nodes can be joined by a path
- **Disconnected graph**: two or more **connected components**
- **Giant component**: the largest connected component
- **Isolates**: the other connected components
- **Bridge**: link whose removal increases the number of components



Connectivity for Directed Graphs

- **Strongly Connected** graph: has paths back and forth from every node to every other node (e.g., AB path and BA path)
- **Weakly connected** graph: connected if we disregard link orientations
- **Strongly connected** components: can be identified; sometimes a single node
- **In-component**: nodes that reach a s.c.c.
- **Out-component**: nodes reachable from a s.c.c.



Clustering coefficients

Clustering coefficient

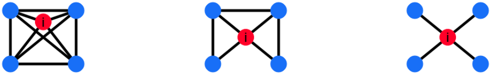
- What fraction of the possible links exist among my neighbors?

$$C_i = \frac{2L_i}{k_i(k_i - 1)},$$

where:

- L_i = number of links between node i's neighbors
- k_i = degree of node i

$$C_i \in [0, 1]$$



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

Clustering coefficient for the entire network

- Average clustering coefficient

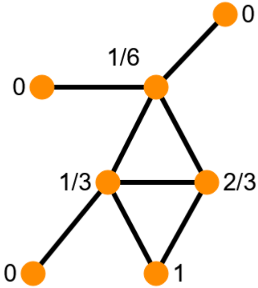
$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

- Global clustering coefficient

$$C_\Delta = \frac{3 \times \#\text{Triangles}}{\#\text{Connected Triplets}}$$

- **connected triplet**: path ABC, but ABC and CBA are considered to be the same triplet.
- a triangle contributes 3 triplets to the denominator
- a path ABC without link AC contributes 1 triplet to the denominator
- both $\langle C \rangle, C_\Delta \in [0, 1]$, not necessarily equal

Clustering coefficients: Example



$$\langle C \rangle = \frac{13}{42} \sim 0.310$$

$$C_{\Delta} = \frac{6}{16} = 0.375$$