Paths and distances

M. Mitrović Dankulov and A. Alorić

December 14, 2022

Outline

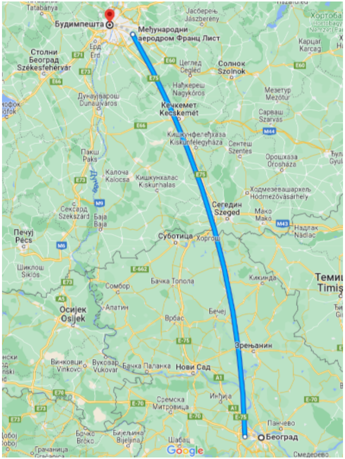
Importance of distance

- Physics: strength of the force decreases with distance

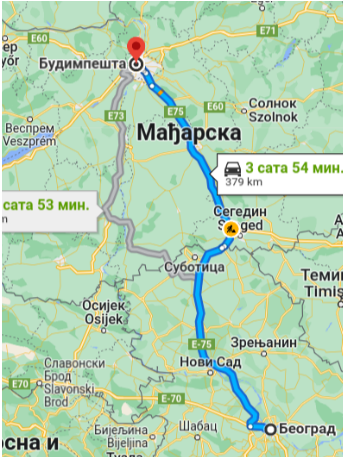
- Geography: everything is related to everything else, but near things are more related than distant things

Distance in physical world

Euclidian

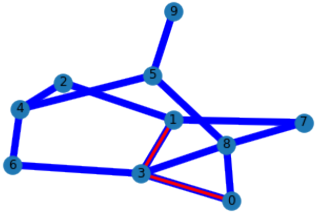


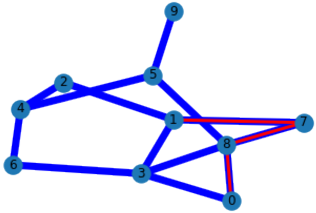
Non-Euclidian



Distance in networks

What is the distance between two webpages or two people?

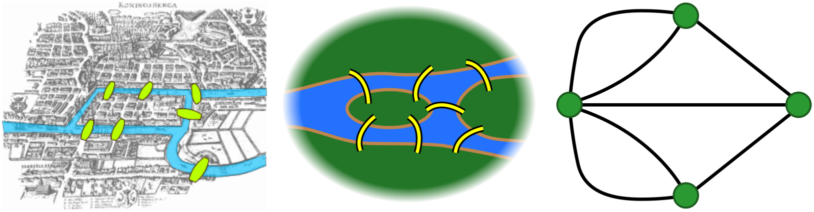




In complex networks theory distance==length of path

Everything started with a path

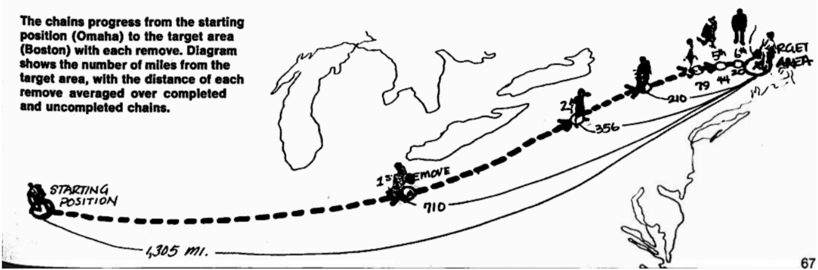
Seven Bridges of Königsberg



Euler set the foundations of network theory

It continued with paths

Small-world experiment in 1967

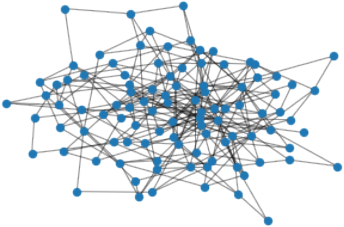


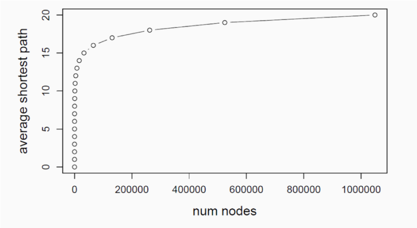
Six degrees of separation

Source: Migram's article in Psychology Today

It continued with paths

Erdos-Renyi networks - N nodes; p - probability to have a link between any pair of nodes;

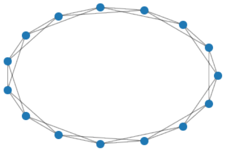




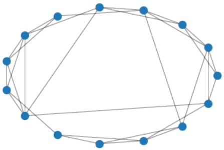
$$\langle l \rangle = \frac{log(n)}{log(p(n-1))}$$

It continued with paths

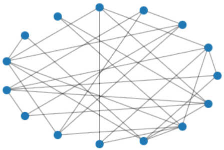
N nodes; k-neighbours; p - rewiring probability;



$T = 15$, $\langle l \rangle = 2.29$



$T = 11$, $\langle l \rangle = 2.03$

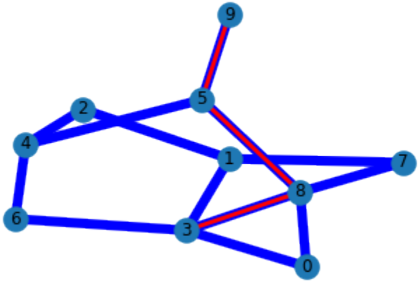


$T = 9$, $\langle l \rangle = 1.9$

Path

- Path is a route that runs along the links of the network
- Path's length l is the number of links the path contains
- Path P between nodes i_0 and i_n: $P = \{(i_0, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n)\}$
- Length of path P is $l = n$

Undirected networkss: path

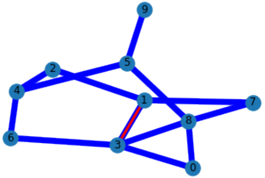


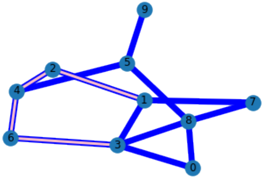
path from node 3 to node 9: $P_{3,9} = \{(3,8),(8,5),(5,9)\}$

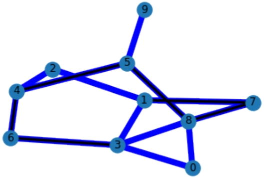
path's length $l = 3$

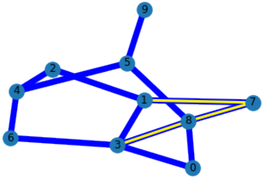
Undirected networks: paths

There is more than one path between a pair of nodes



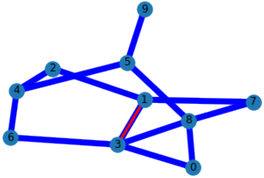


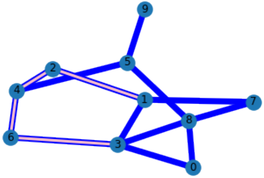




Undirected networks: paths

There is more than one path between a pair of nodes

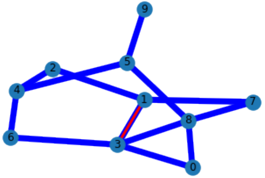


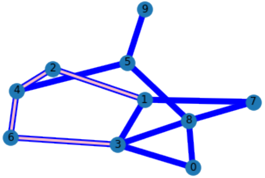


$$P_{3,1} = \{(3,1)\}$$

Undirected networks: paths

There is more than one path between a pair of nodes



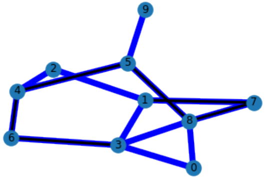


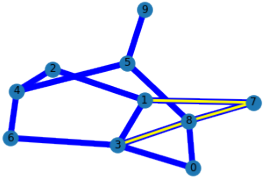
$P_{3,1} = \{(3,1)\};\ l = 1$

$P_{3,1} = \{(3,6),(6,4),(4,2),(2,1)\};\ l = 4$

Undirected networks: paths

There is more than one path between a pair of nodes

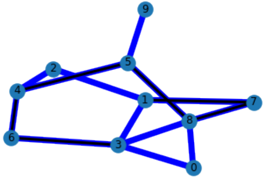


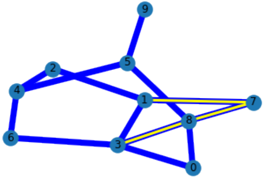


$$P_{3,1} = \{(3,6),(6,4),(4,5),(5,8),(8,7),(7,1)\}$$
$$l = 6$$

Undirected networks: paths

There is more than one path between a pair of nodes





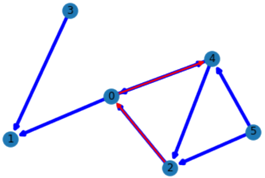
$$P_{3,1} = \{(3,6),(6,4),(4,5),(5,8),(8,7),(7,1)\}$$
$$l = 6$$

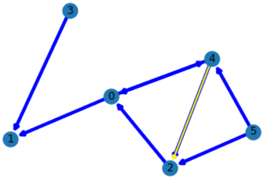
$$P_{3,1} = \{(3,8),(8,7),(7,1)\}$$
$$l = 3$$

In undirected networks $P_{1,3} \equiv P_{3,1}$

Directed networks: paths

In directed networks the existence of a path from node i to node j does not guarantee the existence of a path from j to i





$P_{2,4} = \{(2,0),(0,4)\};\ l = 2$ $\qquad$ $P_{4,2} = \{(4,2)\};\ l = 1$

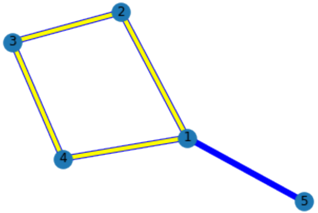
Weighted networks: paths

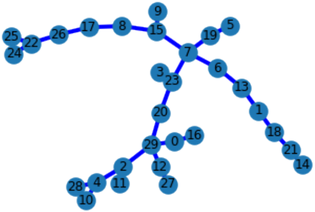
- We can have 2 paths - weighted and toplogical (unweighted) paths
- These paths are different
- If link weights are distances or strengths - weighted path $P^w_{i_0 i_n}\{(i_0, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n)\}$, $l^w_{i_0 i_n} = \sum_{(i,j)\in P^w_{i_0 i_n}} W_{ij}$

Shortest path

- Shortest path between nodes i and j is the path with the fewest number of links
- Shortest path between i and j is called distance between these nodes d_{ij}
- There can be multiple shortest paths between the same pair of nodes
- Shortest path cannot contain loops and cannot intersect itself
- undirected $d_{ij} = d_{ji}$; directed $d_{ij} \neq d_{ji}$

Path types



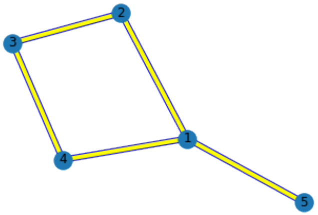


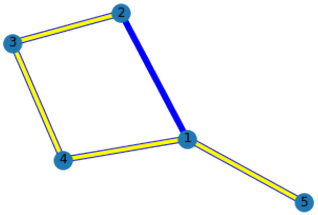
Cycle - a path with the same start and end node

Networks withour cycles - acyclic network

Connected acyclic network is a tree

Path types





Eulerian Path - a path that traverses each link exactly once

Hamiltonian Path - a path that visits each node exactly once

Average path length and diameter

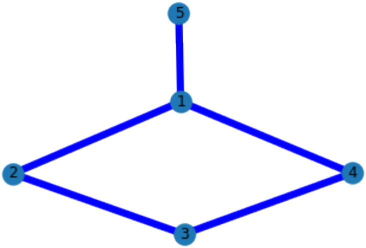
Average path length $\langle l \rangle$ - the average of the shortest paths between all pairs of nodes

$$\langle l \rangle = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i>j} d_{ij}$$

Diameter - the longest shortest path in a network, or the distance between the two furthest nodes

$$D = max_{i>j}(d_{ij})$$

Average path length and diameter



$d_{12} = 1,\ d_{13} = 2,\ d_{14} = 1,$
$d_{15} = 1,\ d_{23} = 1,\ d_{2,4} = 2,$
$d_{25} = 2,\ d_{34} = 1,\ d_{35} = 3,$
$d_{45} = 2$

$\langle l \rangle = 1.6$

$D = 3$

Diameter of disconnected network is $D = \infty$.

Number of shortest paths

Number of shortest paths N_{ij} and distance d_{ij} between nodes i and j can be calulated using adjacency matrix A

$d_{ij} = 1$: if $A_{ij} = 1$ $N_{ij} = 1$ otherwise it is 0

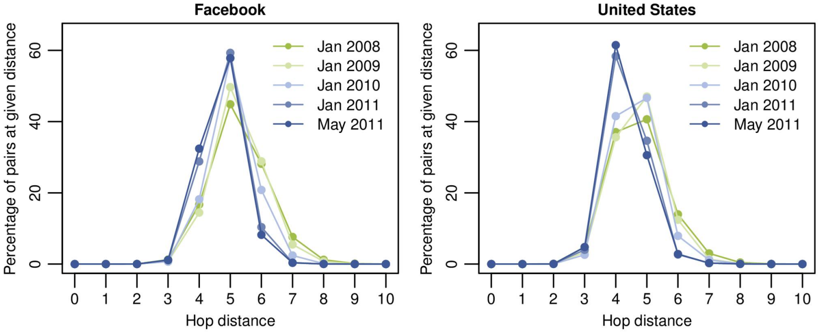
$d_{ij} = 2$: $A^2_{ij} > 1$, $N_{ij} = A^2_{ij}$

$d_{ij} = 3$: $A^3_{ij} > 1$, $N_{ij} = A^3_{ij}$

$d_{ij} = d$: $A^d_{ij} > 1$, $N_{ij} = A^d_{ij}$

Algorithm for finding shortest paths: Breadth-First Search (BFS) Algorithm. See section 2.8 in the Network Science book for details.

Shortest path distribution



Source: https://www.facebook.com/notes/10158927855913415/

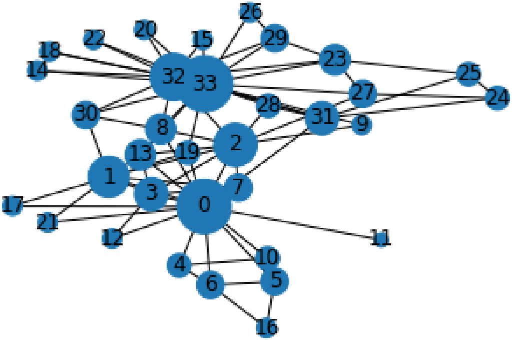
Why do we care?

- In complex networks not all nodes are equally important: airport New York is more important than some local airport
- The same is true for edges: New York - London connection
- Importance depends on the process - different centrality measures

Hubs and degree centrality

Node degree $q_i = \sum_{i=1}^{N} A_{ij}$ - degree centrality

Nodes with high number of neighbours are more important



Path based centrality measures

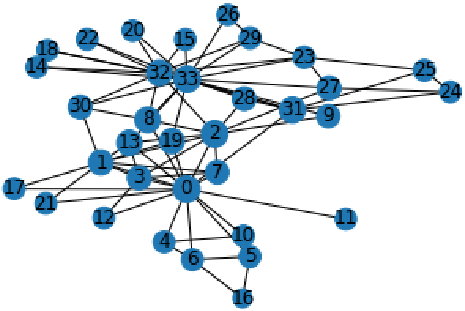
- Closeness centrality - how close is the node to other nodes
- Betweenness centrality of
  - a node - a fraction of shortest paths that pass through a node
  - an edge - a fraction of shortest paths that pass through an edge

Closeness centrality

Closeness centrality of node i - inverse of the sum of the distances of node i to all other nodes in the network

$G_i = \frac{1}{\sum_{j \neq i} d_{ij}}$

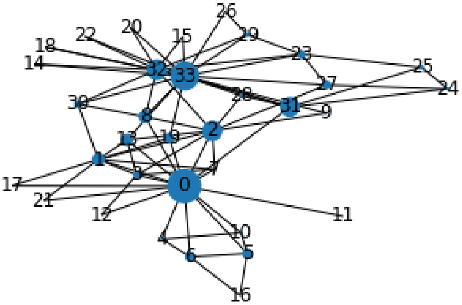
Normalized - $g_i = \frac{N-1}{\sum_{j \neq i} d_{ij}}$



Betweenness centrality of a node

σ_{hj} - number of shortest paths from h to j, and $\sigma_{hj}(i)$ the ones that pass through node i

Betweenness centrality of node i $b_i = \sum_{h \neq i, j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$



Betweenness centrality of an edge

σ_{hj} - number of shortest paths from h to j, and $\sigma_{hj}(e_{ik})$ the ones that pass through an edge e_{ik}

Betweenness centrality of an edge e_{ik} $b_{e_{ik}} = \sum_{h\neq i,k,j\neq i,k} \frac{\sigma_{hj}(e_{ik})}{\sigma_{hj}}$