



Lesson6:
Modeling the Web as a graph
Unit4:
Topology of the Web graph

Rene Pickhardt

Introduction to Web Science Part 2
Emerging Web Properties

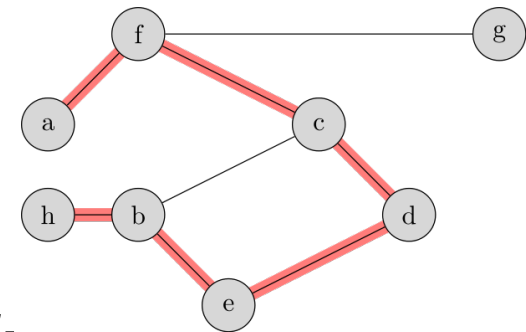


Completing this unit you should

- Understand the notion of a path in a (directed) graph
- Know that shortest paths between nodes need not be unique
- Understand the notion of a strongly connected component
- Know about the diameter of a graph
- Be aware of the bow tie structure of the Web

Paths in a graph

- An ordered set $P = \{v_1, \dots, v_n\}$ is called a path through a graph if:
 - $(v_i, v_{i+1}) \in E, 1 \leq i < n$
- Example:
 - $\{a, f, c, d, e, b, h\}$ is a path
 - $\{a, c, b\} \subset \{a, f, c, d, e, b, h\}$ is not a path

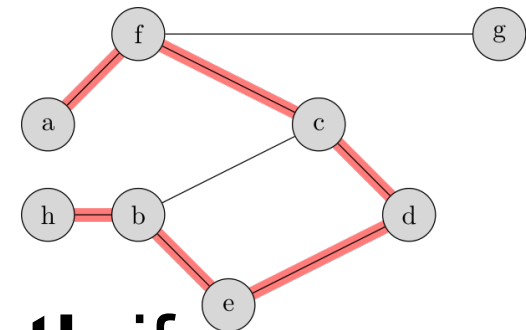


Shortest paths in a graph

- A path $P = \{v_1, \dots, v_n\}$ between v_1 and v_n is called a **shortest path** if:
 - No path P' between v_1 and v_n exists such that

$$|P'| < |P|$$

- Example:
 - $\{a, f, c, d, e, b, h\}$ is a path but not a shortest one
 - $\{a, f, c, b, h\}$ is a shortest path



Useful information on shortest paths

- Shortest paths need not to be unique
- Useful algorithms
 - Breadth first search
 - Unweighted graphs
 - Dijkstra Algorithm
 - Weighted graphs (positive weights)
 - Bellman Ford Algorithm
 - negative weights
 - Floyd Warshall
 - All pair shortest paths

(Strongly)Connected components

- A Set of vertices C such that

- $\forall u, v \in C \exists P \subset C$

- P is a path

- $\nexists v' \in V - C : \exists c \in C : \exists P$

- Such that P is a path between c and v'

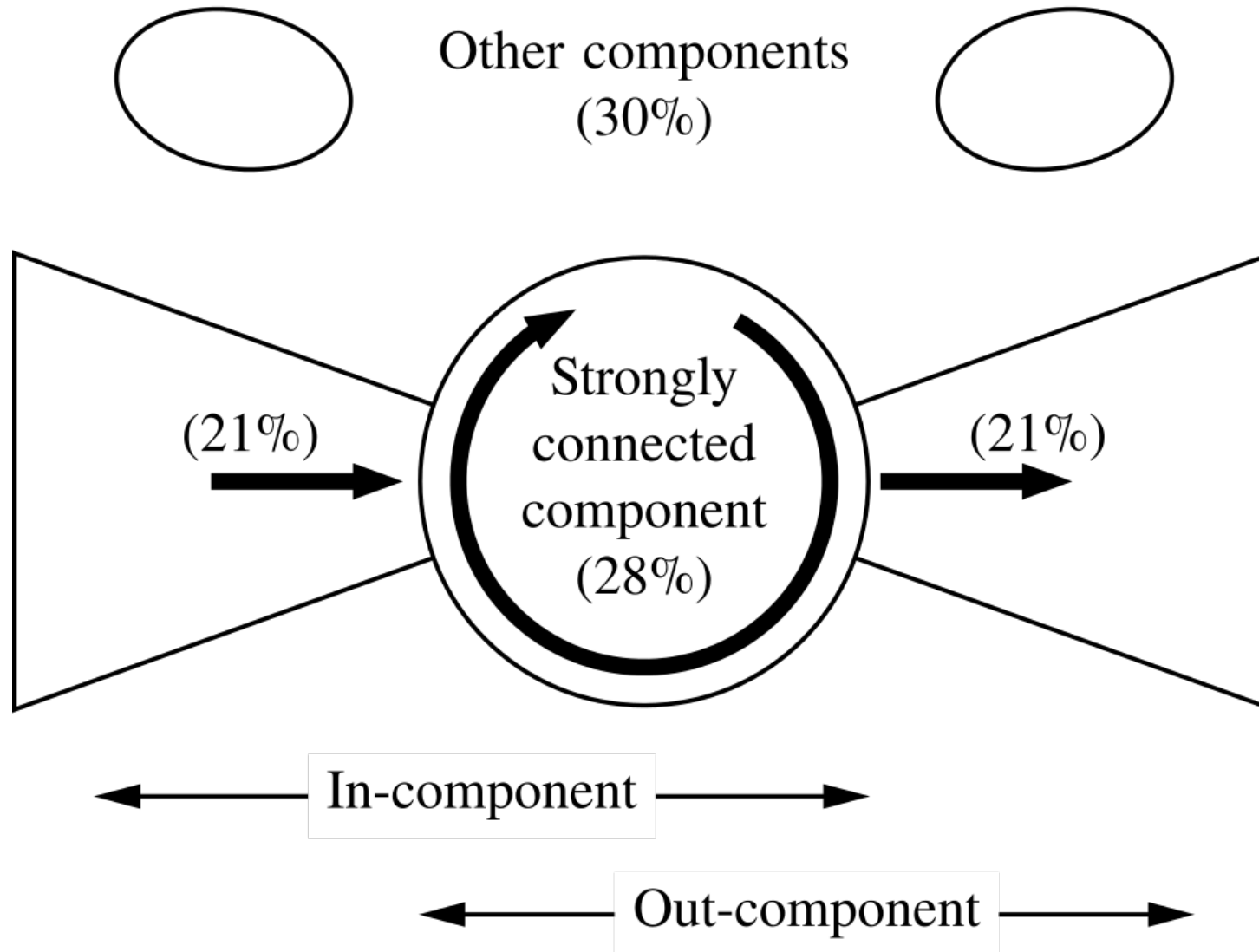
is called strongly connected component in a directed and **connected component** in an **undirected graph**



Strongly connected components

- How many strongly connected components exist on Simple English Wikipedia?
 - It is a good exercise to count them on the html dump we provided
- Size of the largest connected component
 - 54'524 vertices
- The largest strongly connected component is usually the part of the web we experience

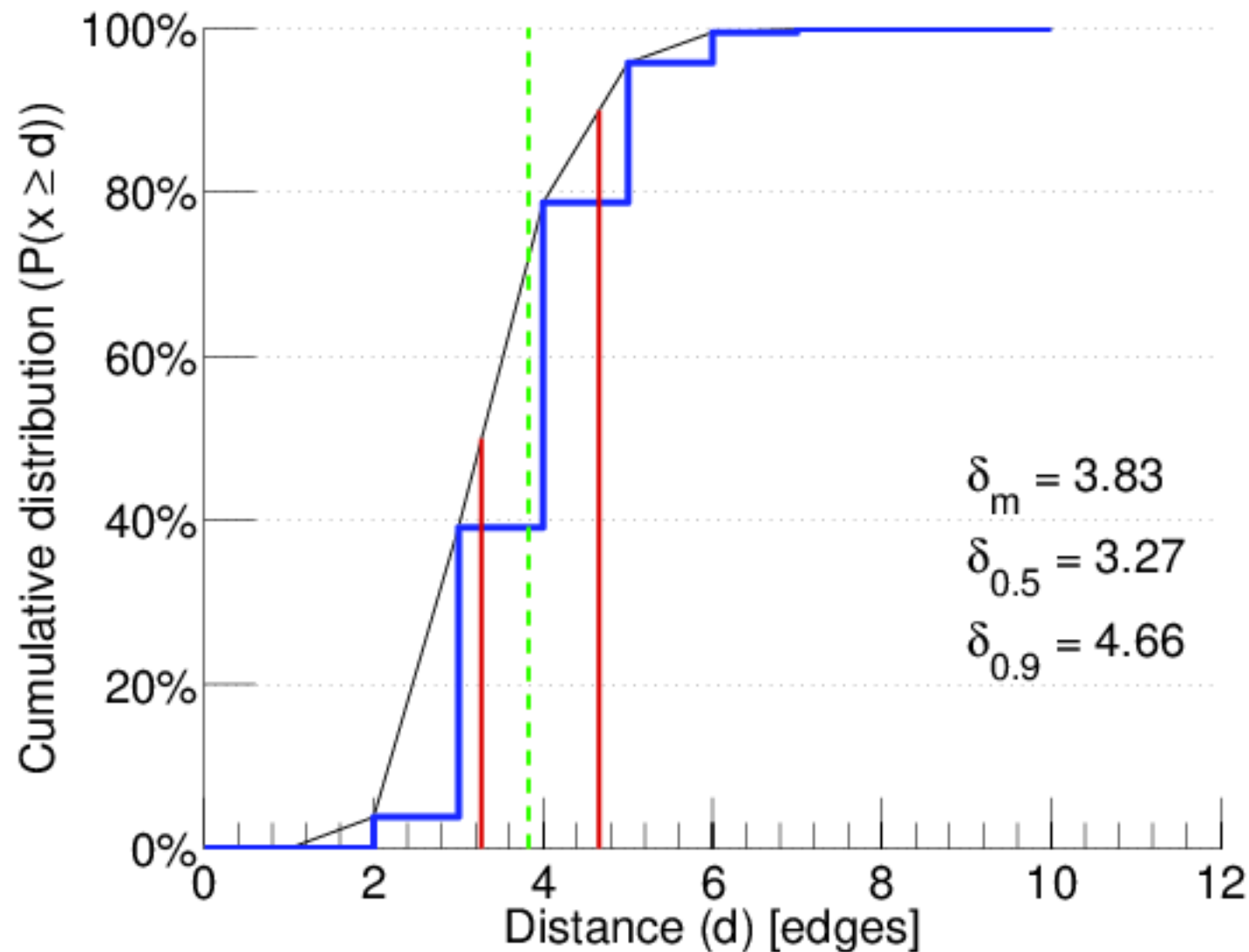
Bow tie structure of the Web



Diameter: Describing the strongly connected component (SCC)

- Each pair of nodes in a SCC is connected by a path
- There exists a shortest path between each pair of nodes
- **The length of the longest shortest path between any two nodes within a SCC is called its diameter.**
- How large do you expect the diameter of Simple Wiki's SCC?

Distance distribution of nodes from the SCC



Statistical measures for graphs that you should know

- Size and Volume
- Degree distributions
- ~~Clustering coefficients~~
- (effective) Diameter
- (Strongly) connected components
- Bow Tie structure
- ~~Assortativity~~
- Lorenz curve, Gini coefficient
- ...



Thank you for your attention!



Contact:

Rene Pickhardt
Institute for Web Science and Technologies
Universität Koblenz-Landau
rpickhardt@uni-koblenz.de

WeST 
People and Knowledge Networks

Pictures:

- <https://commons.wikimedia.org/wiki/File:6n-graf.svg> By User:AzaToth (Image:6n-graf.png similar input data) [Public domain], via Wikimedia Commons
- <https://pixabay.com/en/lightbulb-electric-light-bright-idea-31254/> Public domain
- <https://commons.wikimedia.org/wiki/File:Simple-bipartite-graph.svg> By MistWiz (Own work) [Public domain], via Wikimedia Commons
- <https://commons.wikimedia.org/wiki/File:CPT-Graphs-undirected-weighted-loop-multiedge.svg> By Pluke (Own work) [CC0], via Wikimedia Commons
- <https://commons.wikimedia.org/wiki/File:Matrix.svg> By Lakeworks (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY-SA 4.0-3.0-2.5-2.0-1.0 (<http://creativecommons.org/licenses/by-sa/4.0-3.0-2.5-2.0-1.0>)], via Wikimedia Commons
- https://commons.wikimedia.org/wiki/File:Ego_network.png By Houl0078 (Own work) [CC0], via Wikimedia Commons
- <https://commons.wikimedia.org/wiki/File:Path-graph-theory.svg> By MartinThoma (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY 3.0 (<http://creativecommons.org/licenses/by/3.0>)], via Wikimedia Commons
- https://commons.wikimedia.org/wiki/File:%22Bow-tie%22_diagram_of_components_in_a_directed_network_SVG.svg By Limaner (Own work) [CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0>)], via Wikimedia Commons
- https://commons.wikimedia.org/wiki/File:Economics_Gini_coefficient2.svg By Reidpath [Public domain], via Wikimedia Commons
- <http://konekt.uni-koblenz.de/statistics/volume>
- <https://www.flickr.com/photos/dannysullivan/369540238> via flickr CC-BY 2.0 by Danny Sullivan
- Graph statistics via <http://konekt.uni-koblenz.de/networks/link-dynamic-simplewiki>

- https://commons.wikimedia.org/wiki/File:WWW_logo_by_Robert_Cailliau.svg public domain
- <https://commons.wikimedia.org/wiki/File:WorldWideWeb.1.png> public domain
- https://en.wikipedia.org/w/index.php?title=World_Wide_Web&oldid=692731715 (definition of world wide web)
- https://commons.wikimedia.org/wiki/File:Visualization_of_the_world_wide_web_common_crawl_2012.png By Sebastian Schelter [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons
- <https://commons.wikimedia.org/wiki/File:WorldWideWebAroundWikipedia.png> Chris 73 / Wikimedia Commons [GFDL 1.3 (www.gnu.org/licenses/fdl-1.3.html) or CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons
- https://commons.wikimedia.org/wiki/File:Word_Cloud_Readers_Survey.jpg By Manipande (Own work) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons
- <https://www.flickr.com/photos/valerieebb/290711738> via Flickr and Valerie Everett
- <https://commons.wikimedia.org/wiki/File:Inner-product-angle.png> CC-BY-SA by CSTAR & Oleg Alexandrov
- https://commons.wikimedia.org/wiki/File:Synoptic_word-for-word.png By Alecmconroy (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY 3.0 (<http://creativecommons.org/licenses/by/3.0>)], via Wikimedia Commons