Lesson6:
Modeling the Web as a graph
Unit3:
Descriptive statistics for the Web graph

Rene Pickhardt

Introduction to Web Science Part 2

Emerging Web Properties

WeST
People and Knowledge Networks

Institute for Web Science and Technologies · University of Koblenz-Landau, Germany

Completing this unit you should

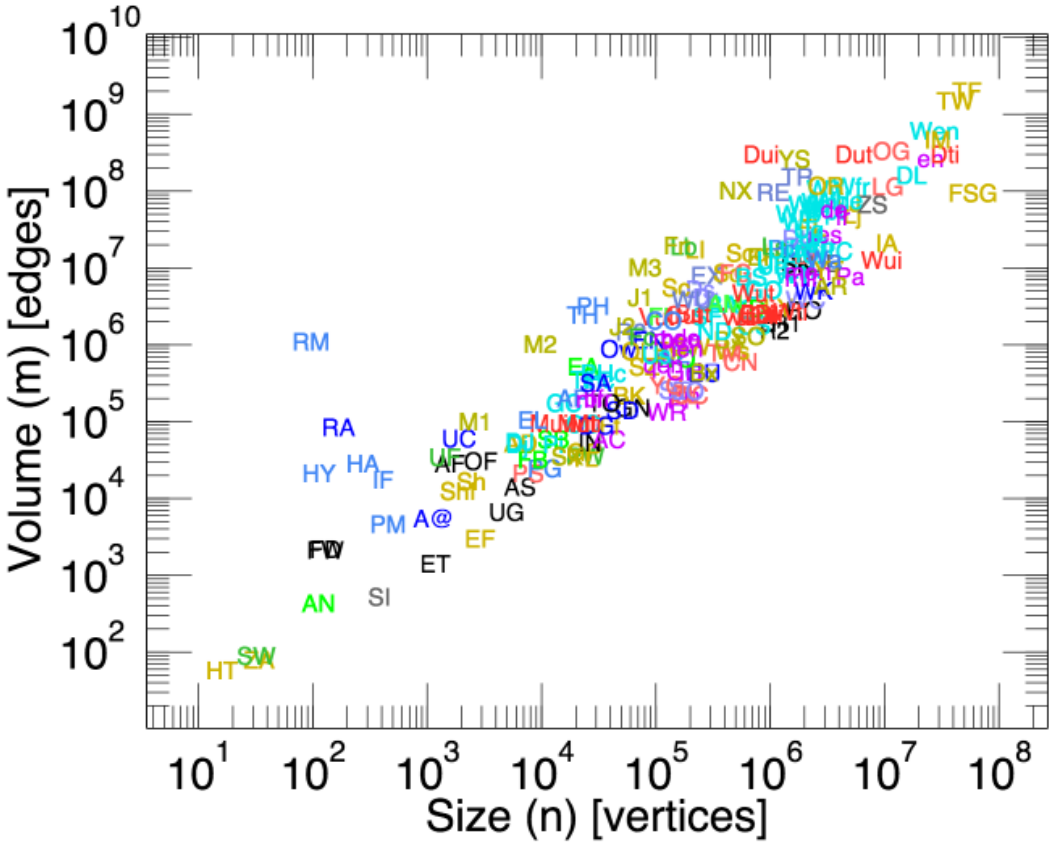
- Know terms like Size and (unique) volume
- Be able to count the in and out degree of web pages
- Have an idea what kind of law (in & out) degree distributions follow
- Know that degree is not distributed in a fair way
- Know that the Gini coefficient can be used to measure fairness

Some basic statistics about Simple English

- **Size** (Number of vertices)
  - 100‘312
- **Volume** (Number of edges)
  - 1‘627‘472
- **Unique volume** (Number of unique edges)
  - 746‘086

Empirical connection between Size and volume in over 100 networks

- There seems to be a connection
- The larger the size the larger the volume
- Beware the log log plot



Neighbors, in degree and out degree

$$In(v) = \{u \in V | \exists e \in E : e = (u, v)\}$$

$$indeg(v) = |In(v)|$$

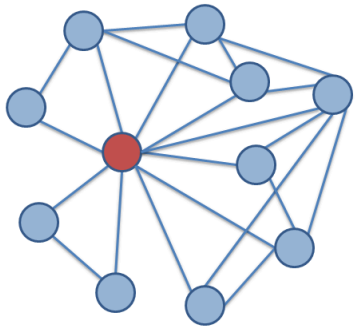
$$Out(v) = \{u \in V | \exists e \in E : e = (v, u)\}$$

$$outdeg(v) = |out(v)|$$

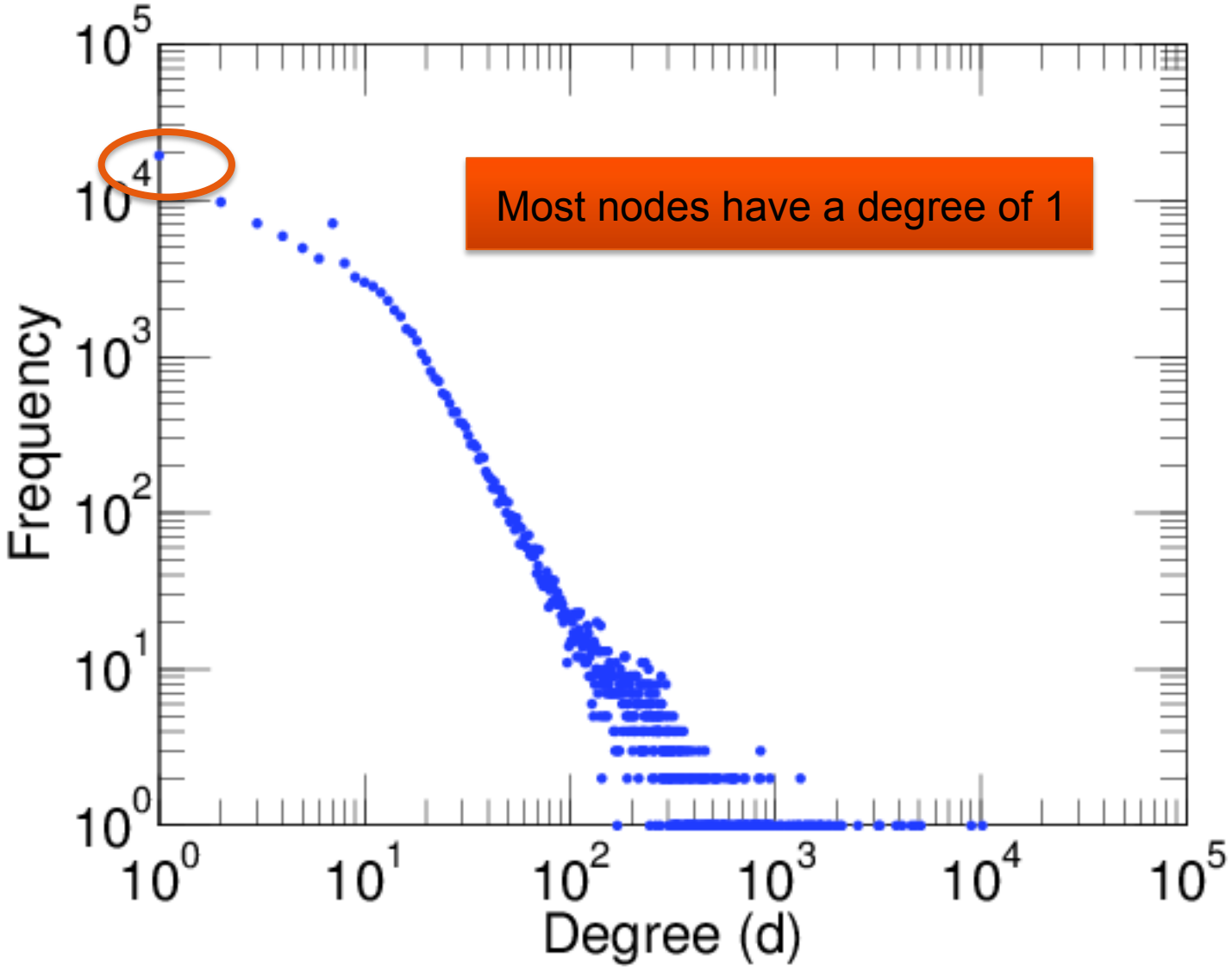
$$Neighbors(v) = In(v) \cup Out(v)$$

$$deg(v) = |Neighbours(v)|$$

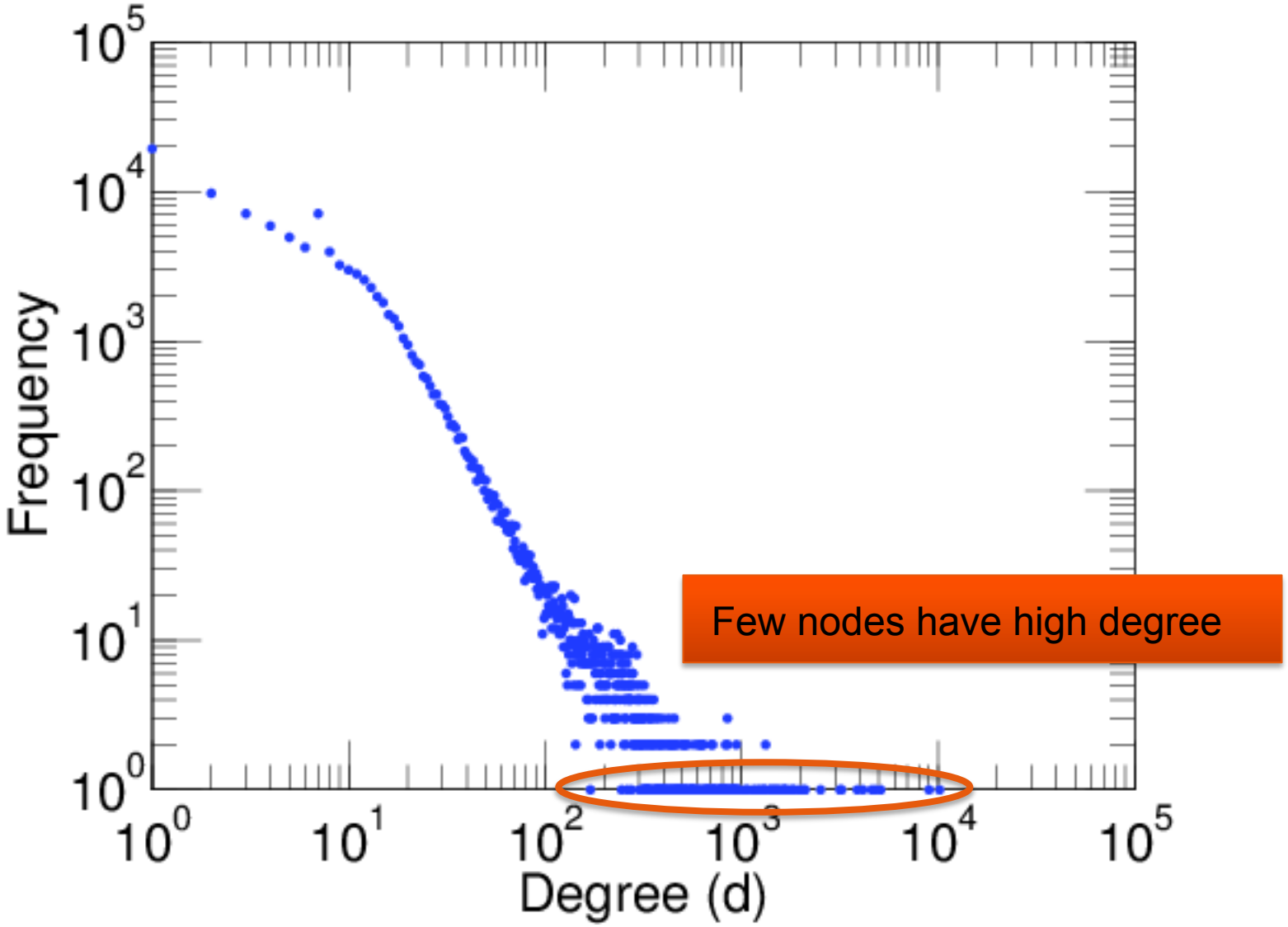
$$deg(v) \leq indeg(v) + outdeg(v)$$



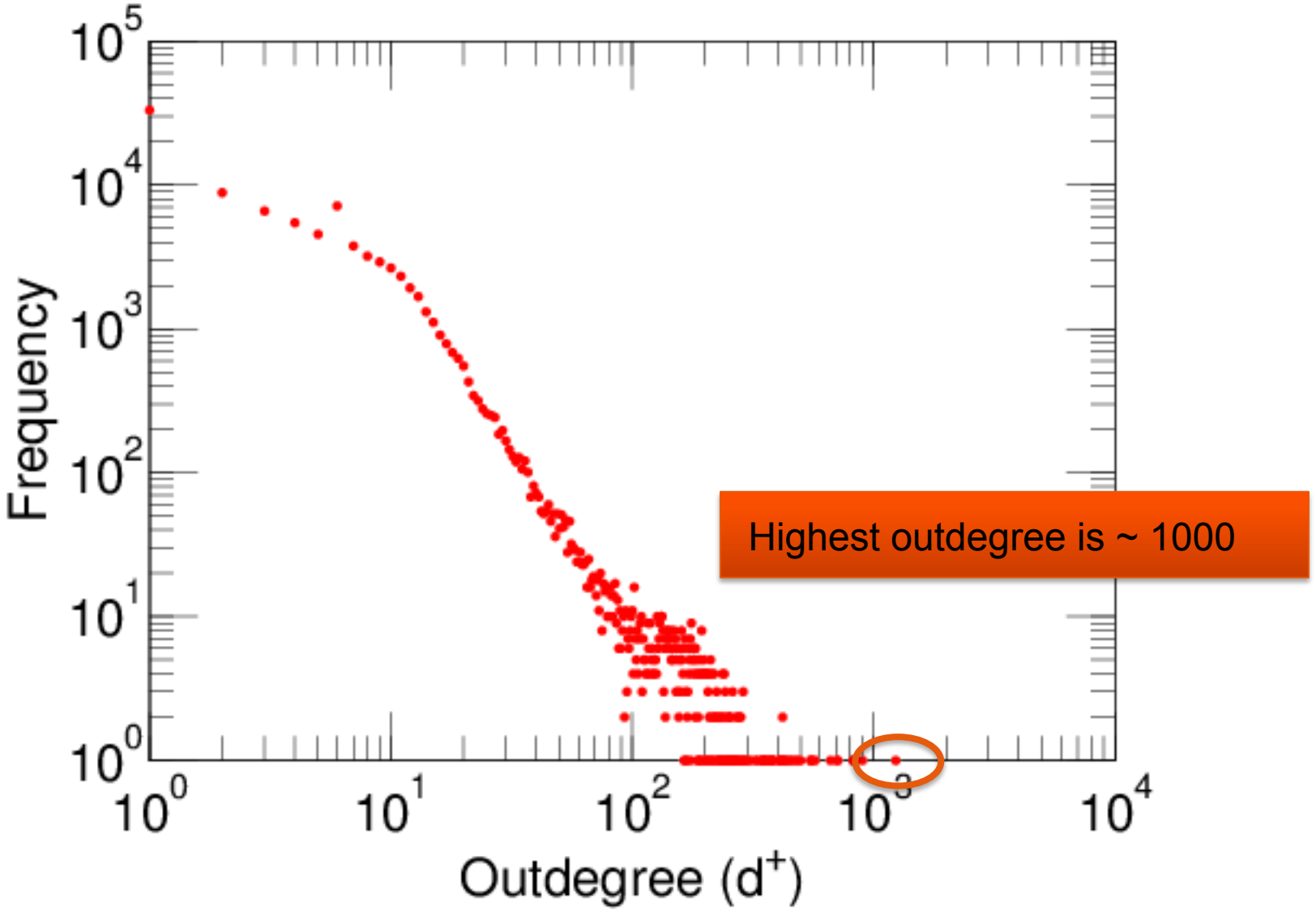
Degree distribution for Simple English Wiki



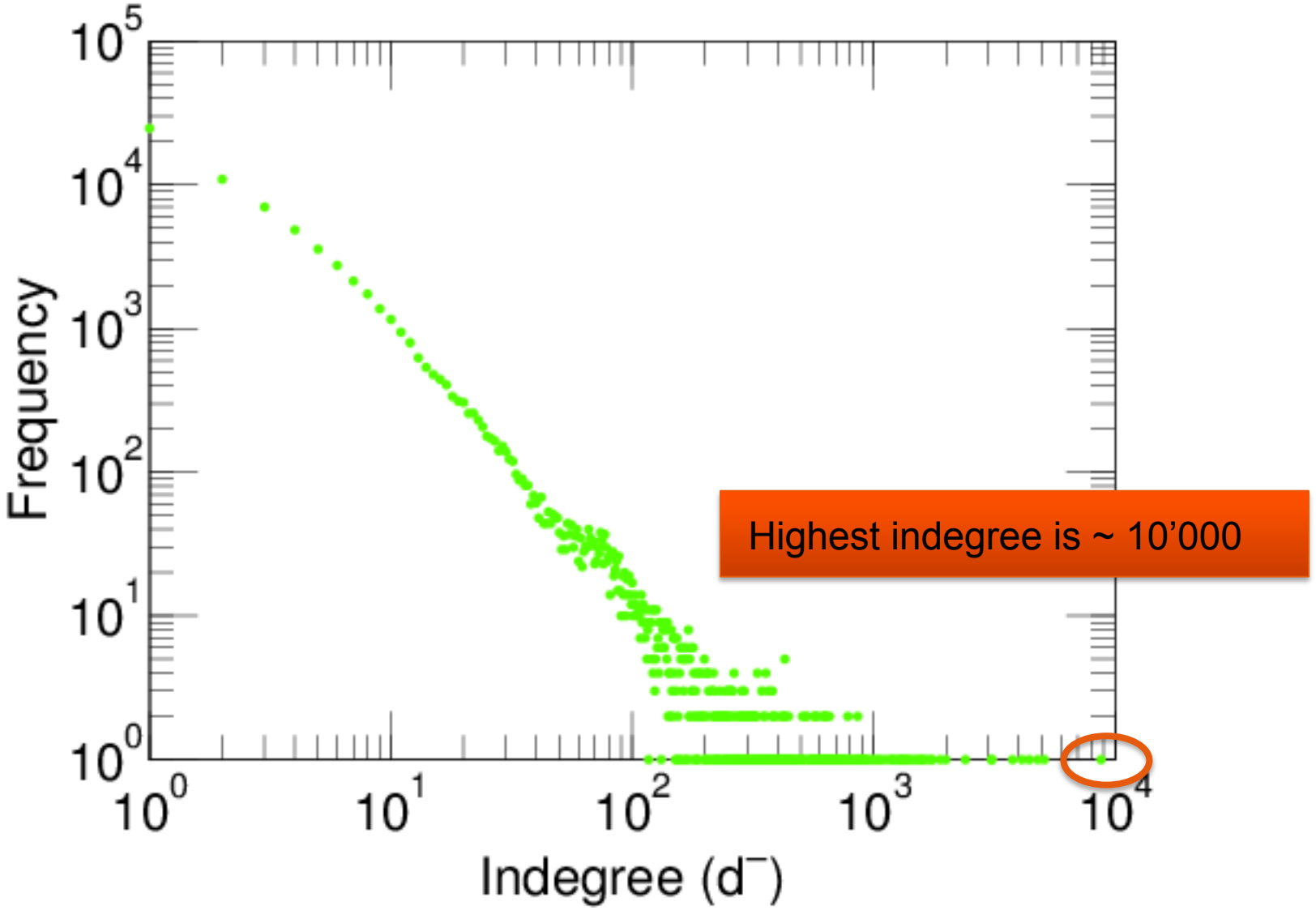
Degree distribution for Simple English Wiki



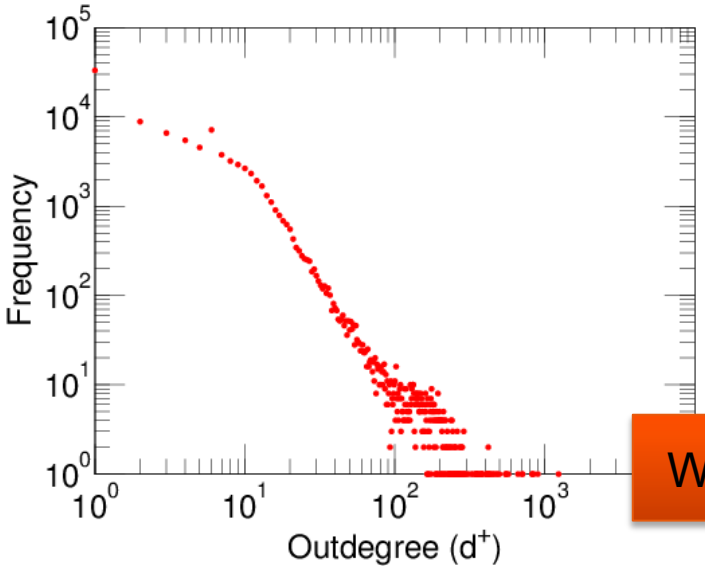
Out degree distribution

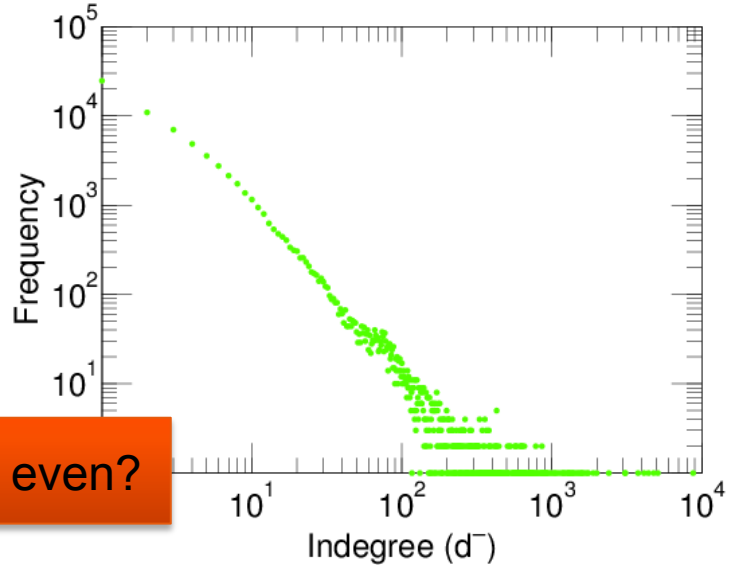


In degree distribution

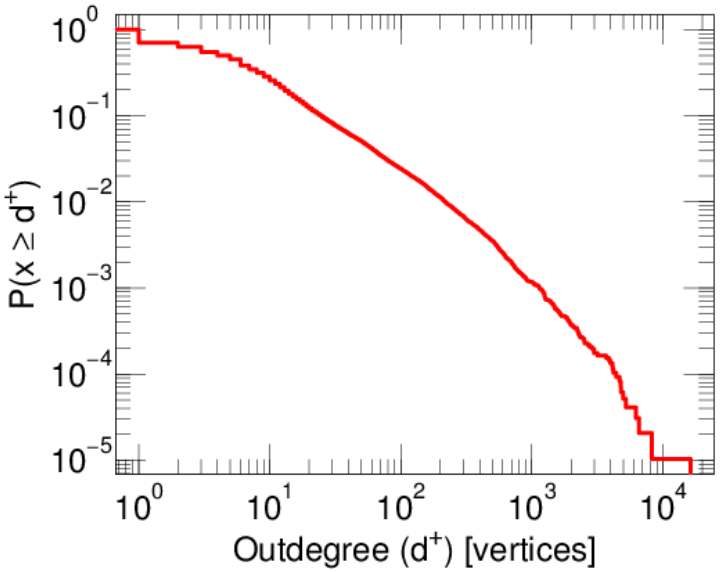


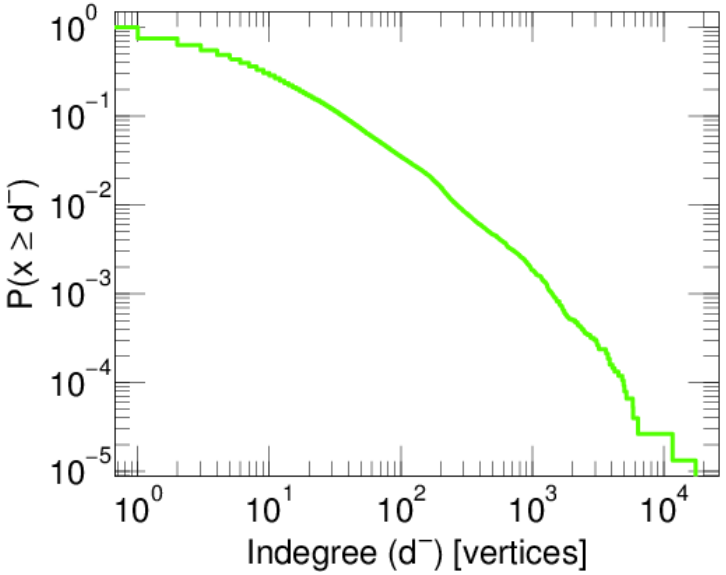
Comparing in and out degree distributions I





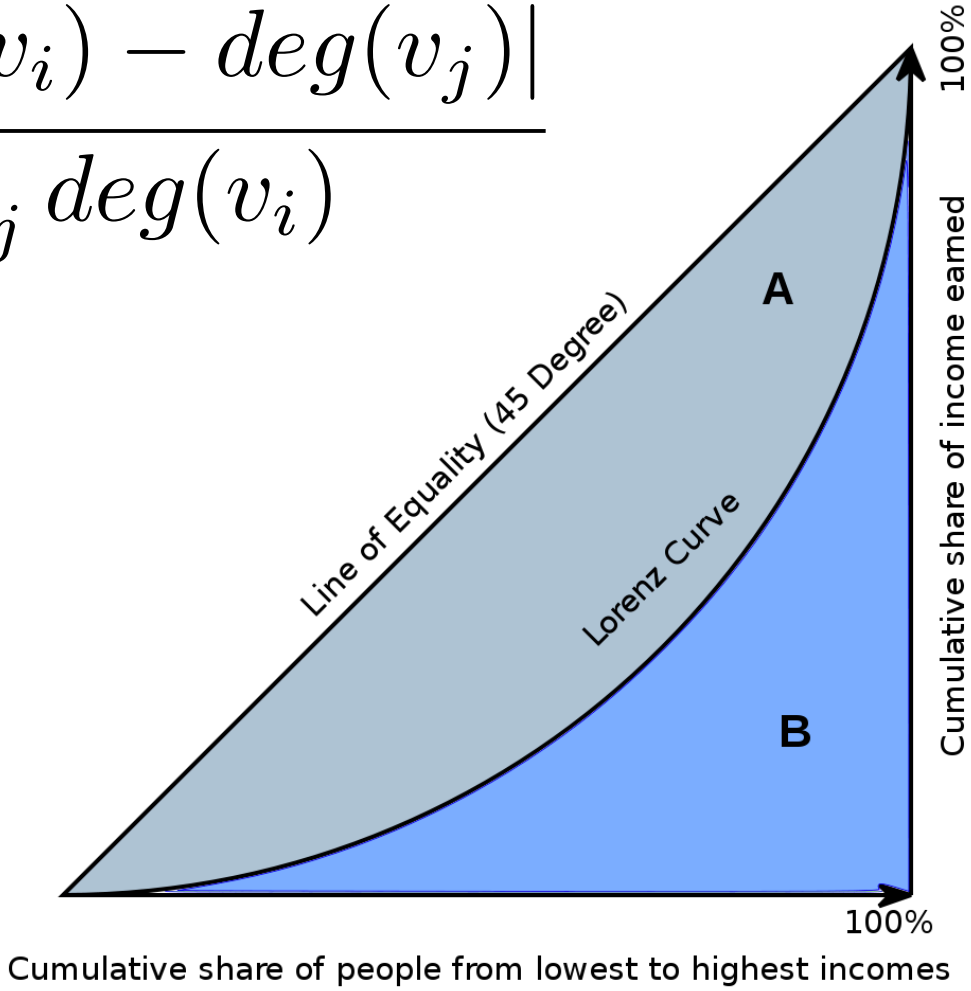
Which one is more even?



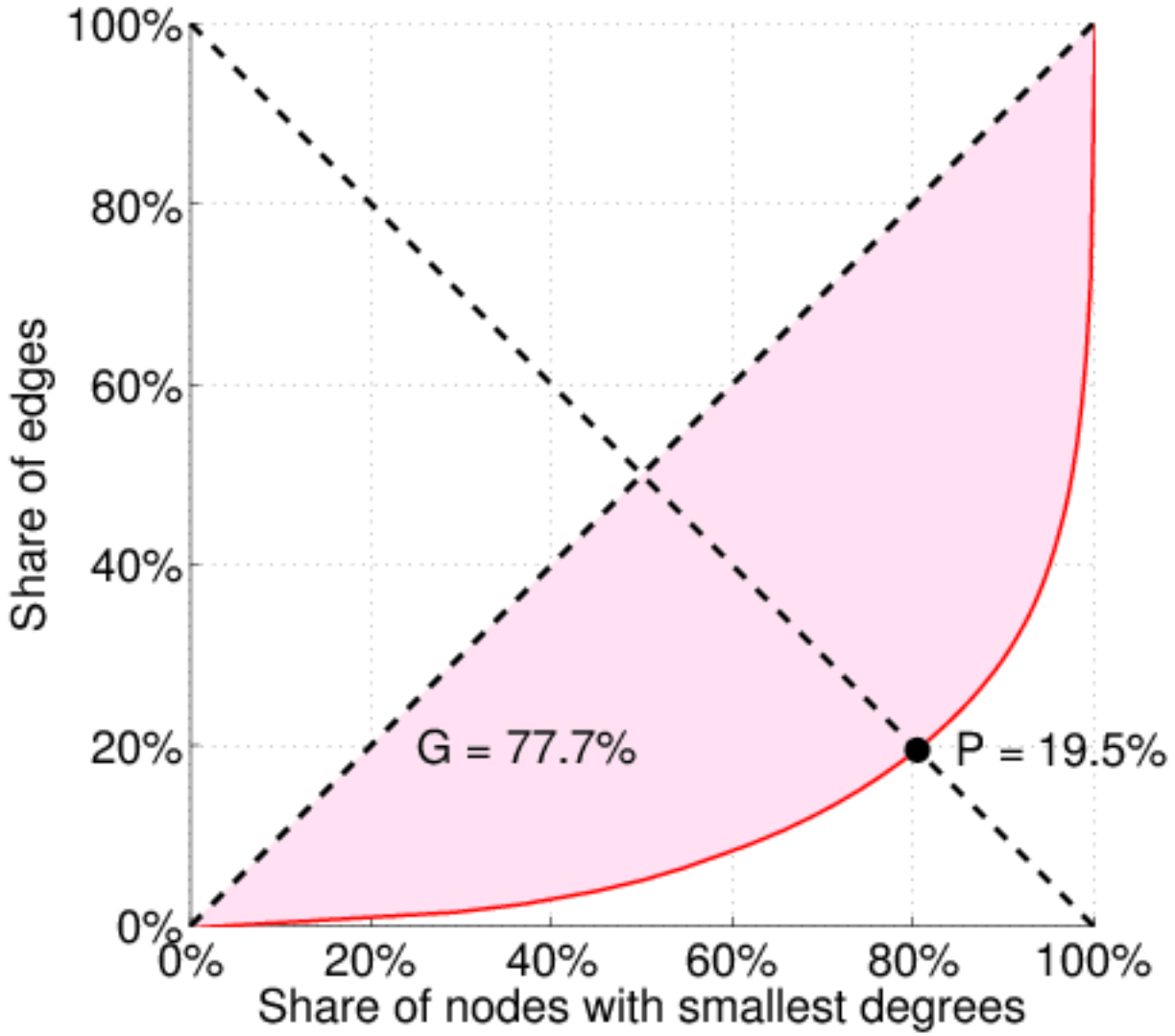


Use the Gini coefficient to describe Fairness

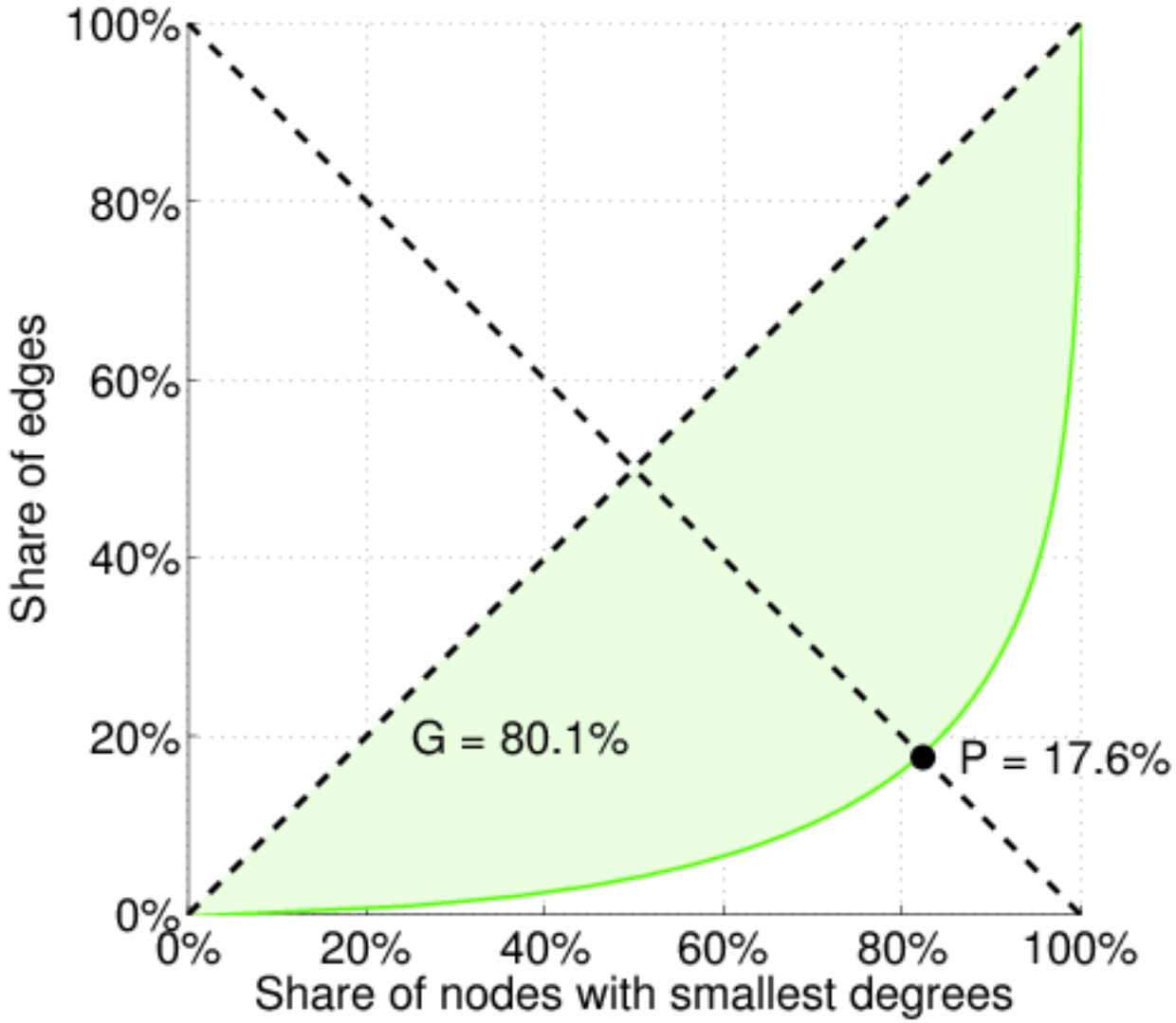
- $G = \frac{\sum_i \sum_j |deg(v_i) - deg(v_j)|}{2 \sum_i \sum_j deg(v_i)}$
- $G = A/(A+B)$
- $G \in [0, 1]$
  - 0 = fair
  - 1 = unfair



Out degree lorenz curve



In degree lorenz curve



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 simlar 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://konect.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://konect.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/valeriebb/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