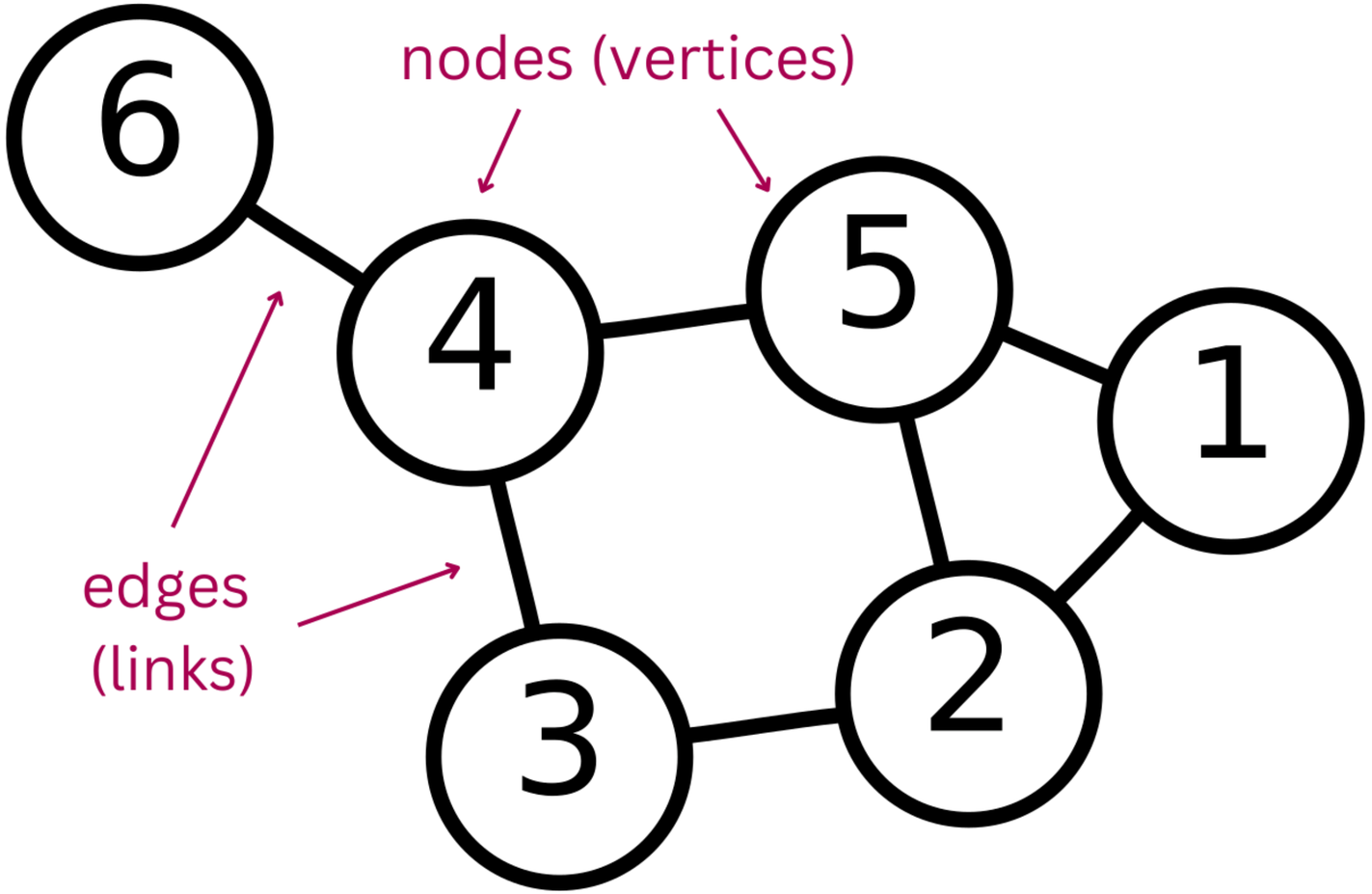
Ollscoil na Gaillimhe
University of Galway

Network Analysis with Gephi

Oksana Dereza
Digital Library Developer
oksana.dereza@universityofgalway.ie

University *of* Galway.ie

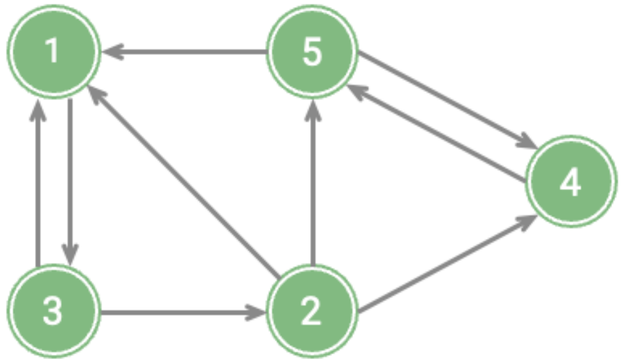
What is a graph?



What types of graphs are there?

Ollscoil na Gaillimhe
University of Galway

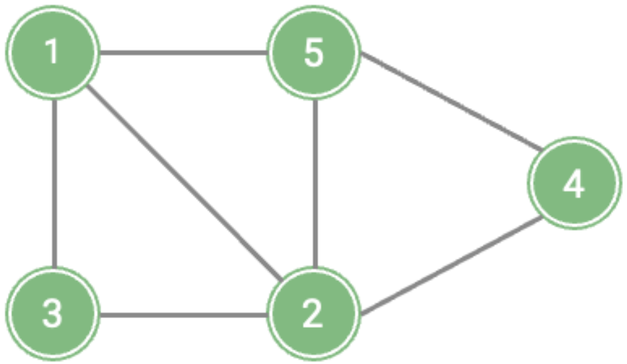
Directed (connections as arrows)



Directed graph

(Edges with direction)

Undirected (connections as lines)



Undirected graph

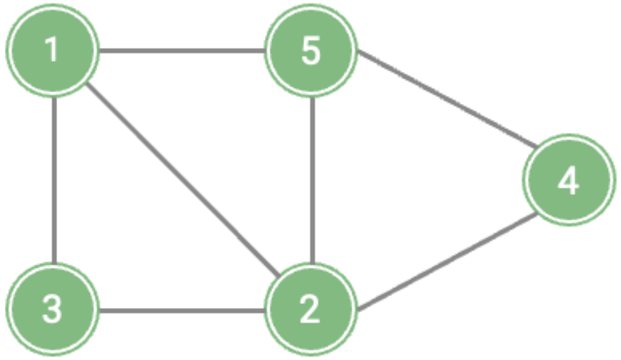
(Edges without direction)

Image source: Hello Algo

What types of graphs are there?

Ollscoil na Gaillimhe
University of Galway

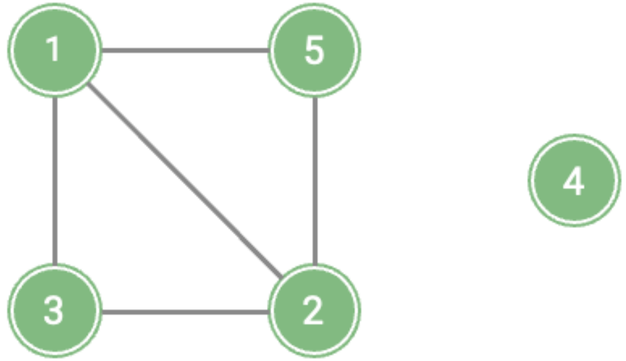
Connected (all nodes are connected)



Connected graph

(All vertices are reachable)

Disconnected (some nodes are isolated from the main graph)



Disconnected graph

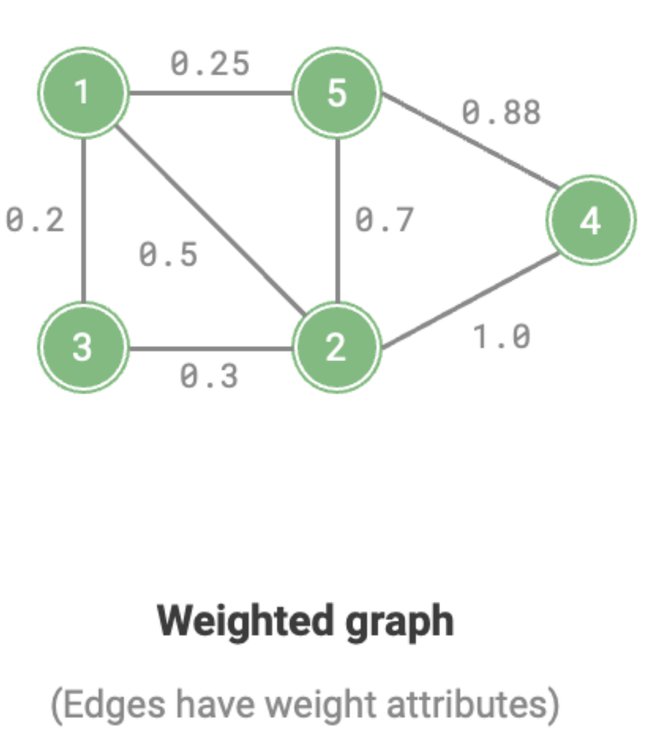
(Some vertices are unreachable)

Image source: Hello Algo

What types of graphs are there?

Ollscoil na Gaillimhe
University of Galway

Weighted (connections have numerical values)



Unweighted (connections don't have numerical values)

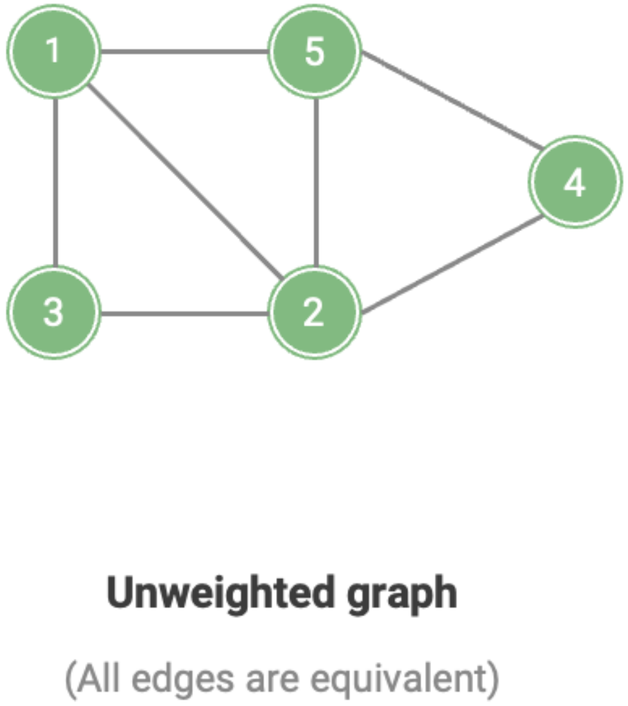


Image source: Hello Algo

What types of graphs are there?

Bipartite: there are two subsets of nodes, and every edge connects a node from subset A to a node from subset B

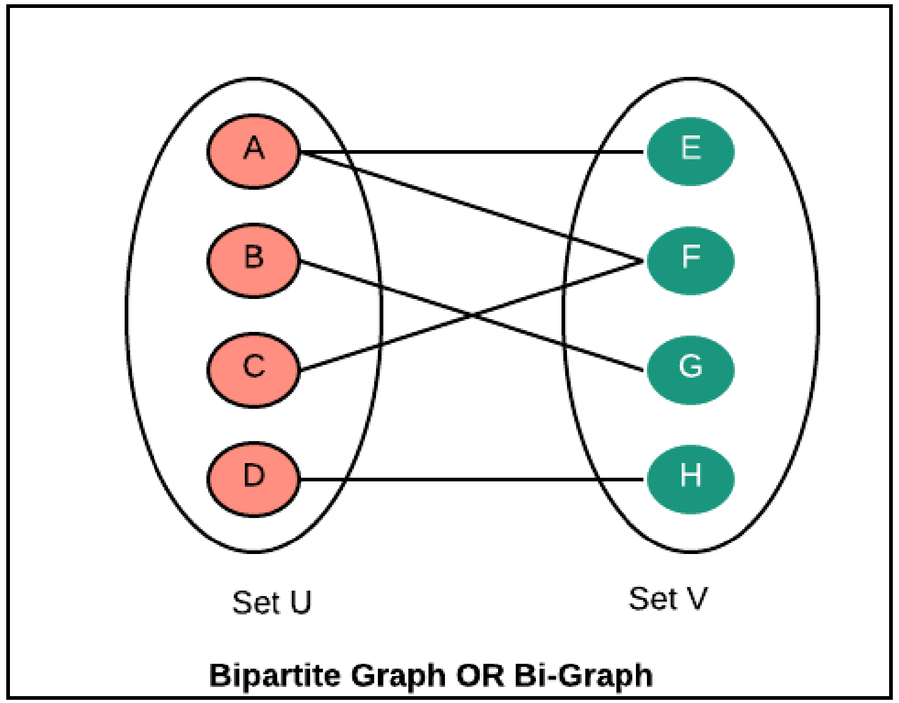


Image source: Tutorial Horizon

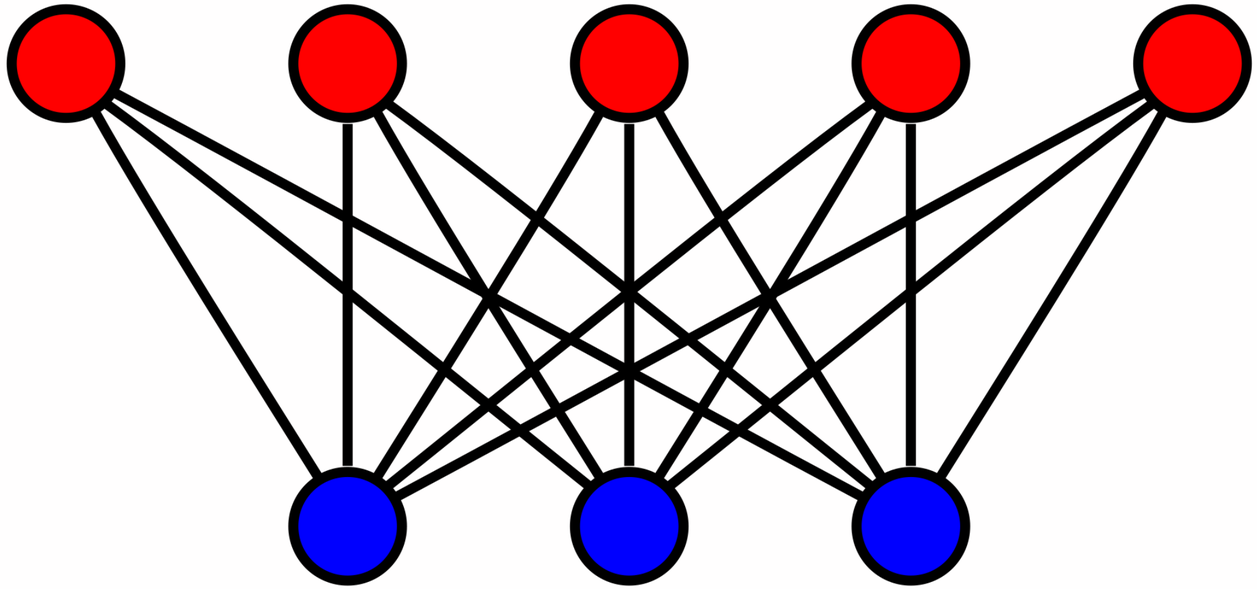


Image source: Wikimedia

What types of graphs are there?

Ollscoil na Gaillimhe
University of Galway

Complete: every node is connected to every other node, but there are no self-loops

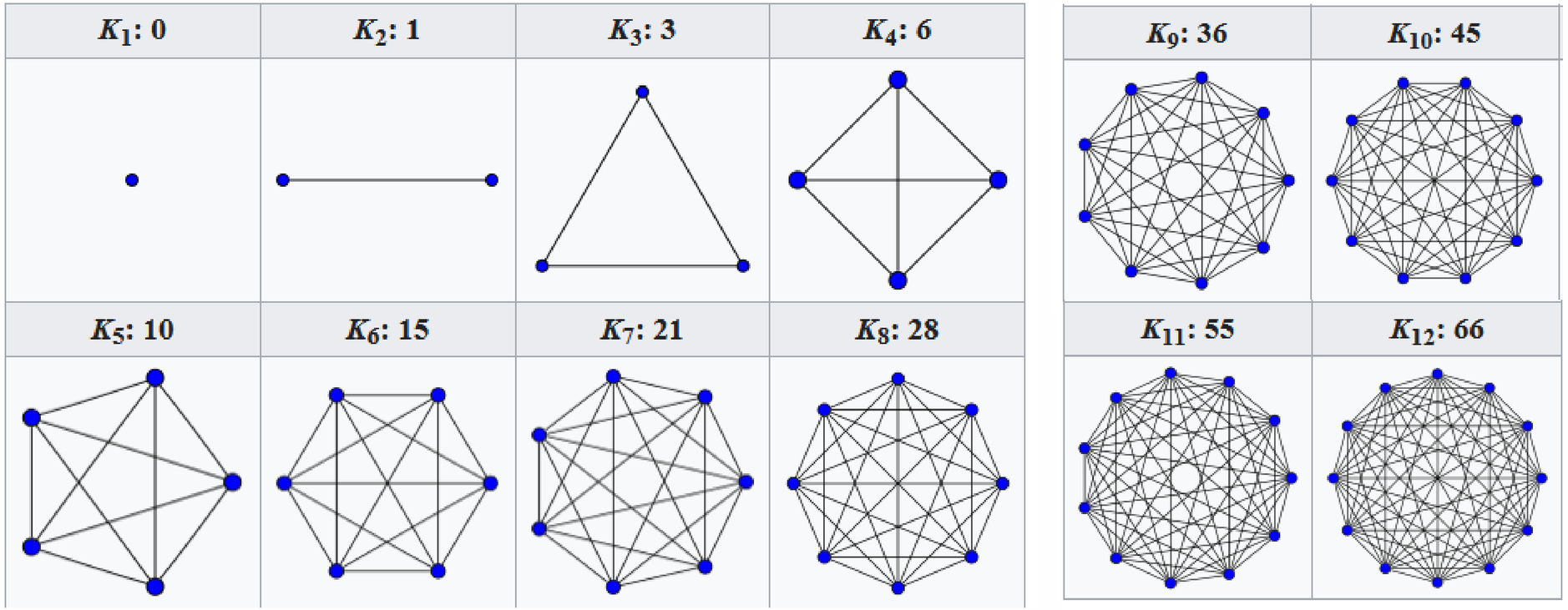


Image source: Wikimedia

What types of graphs are there?

Ollscoil na Gaillimhe
University of Galway

Small world: each node has few direct neighbours, but its neighbours are very likely to be connected to each other and most nodes can be reached from any other node in just a few steps

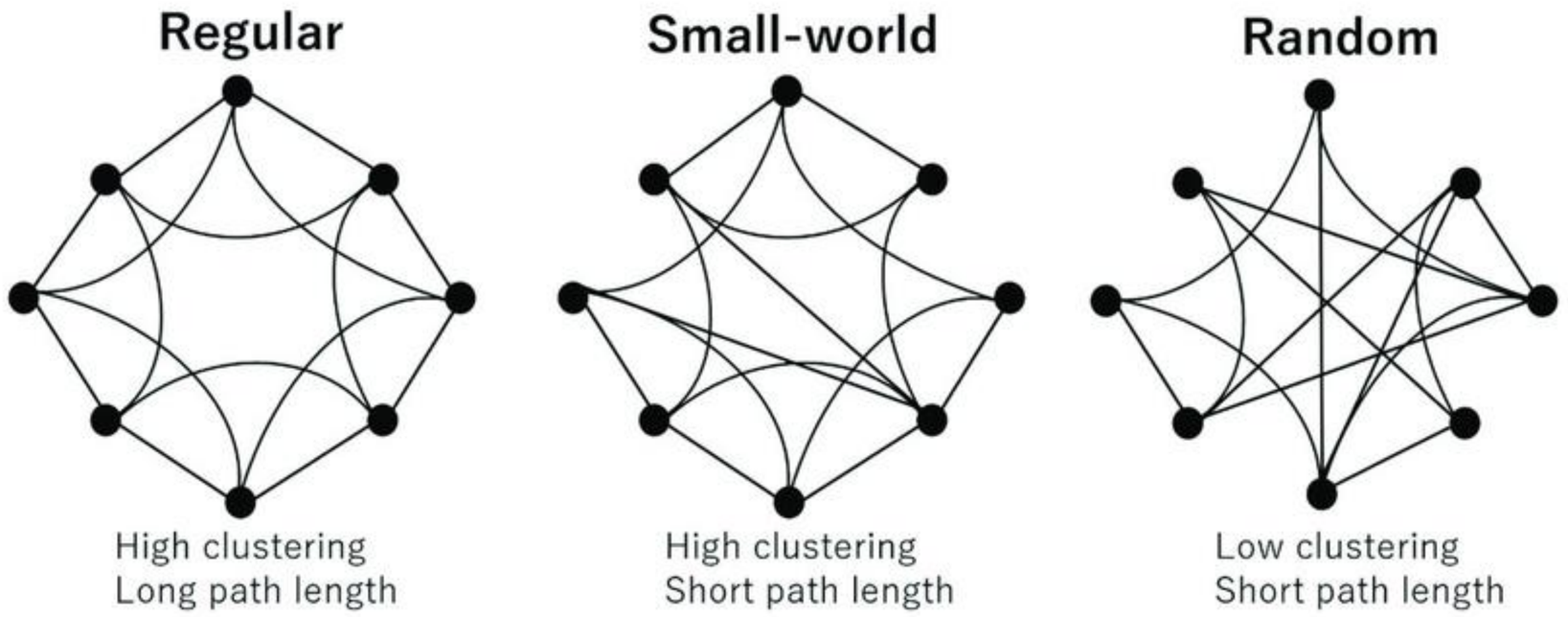


Image source: ResearchGate

Graphs in Real Life

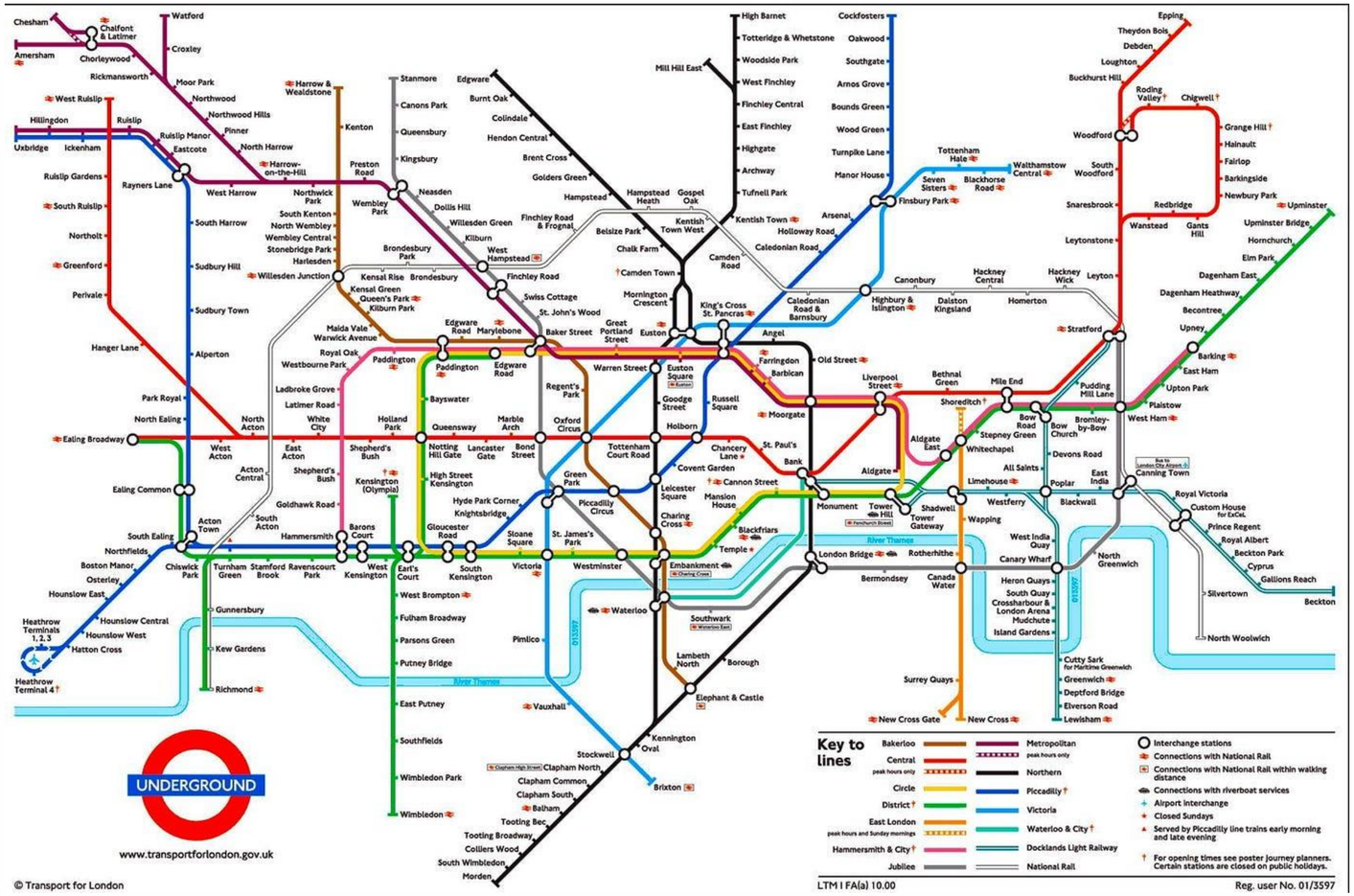


Image source: [Transport for London]

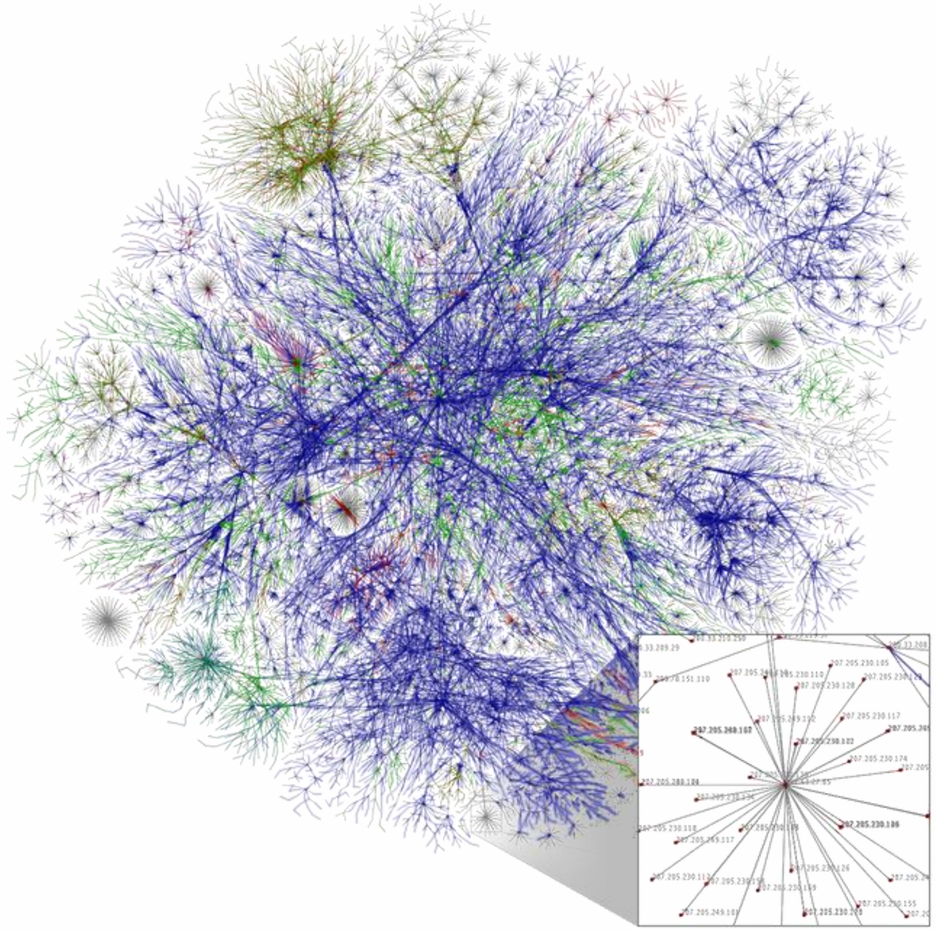


Image source: Wikimedia

Social network analysis (SNA)

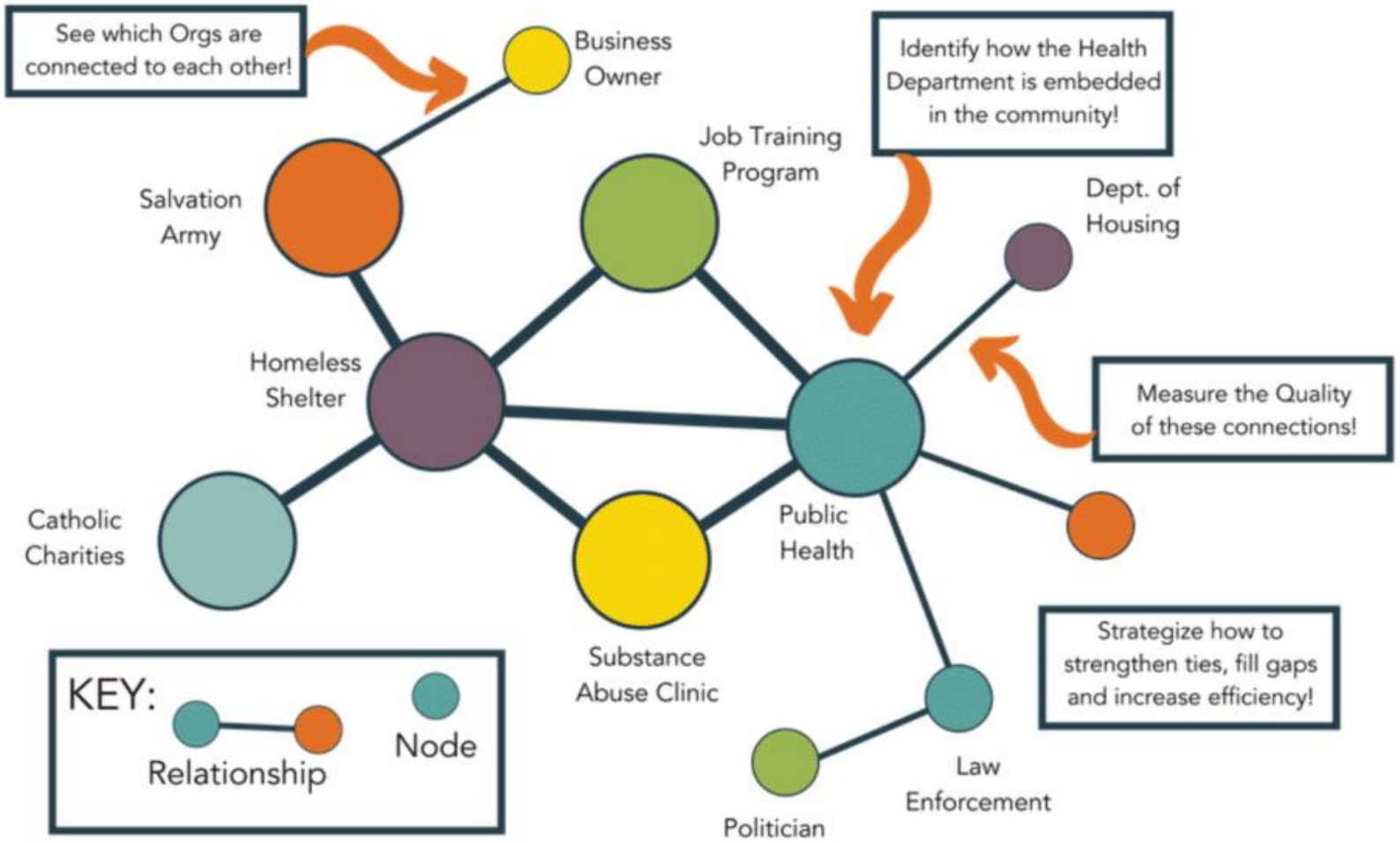


Image source: VisibleNetworksLabs

Communities

A **community** is a dense subgraph where all (or almost all) nodes are interconnected

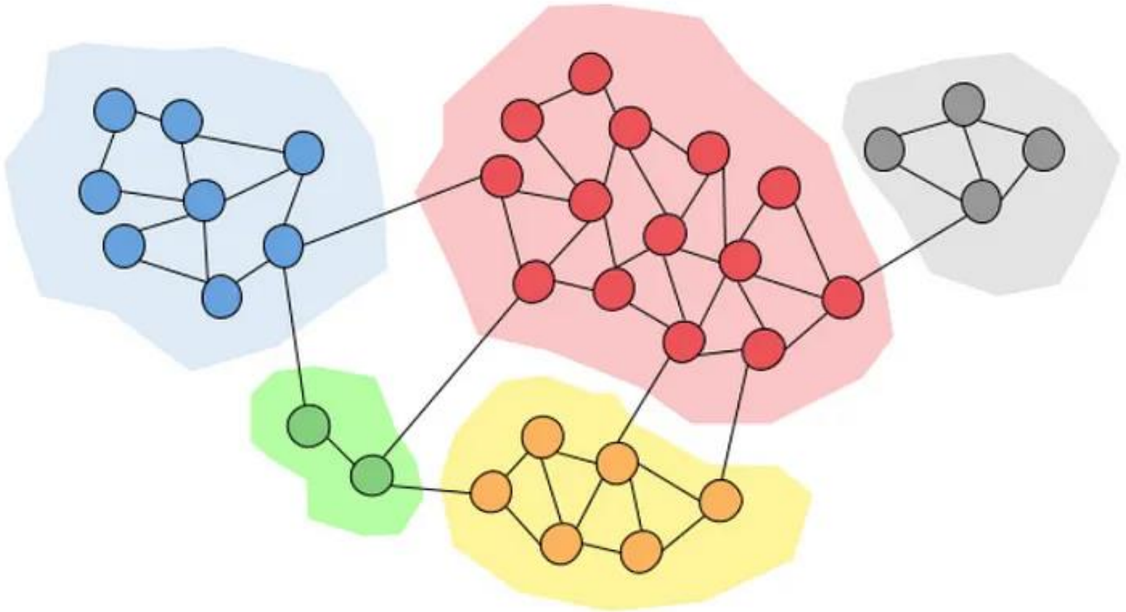


Image source: TDS Archive

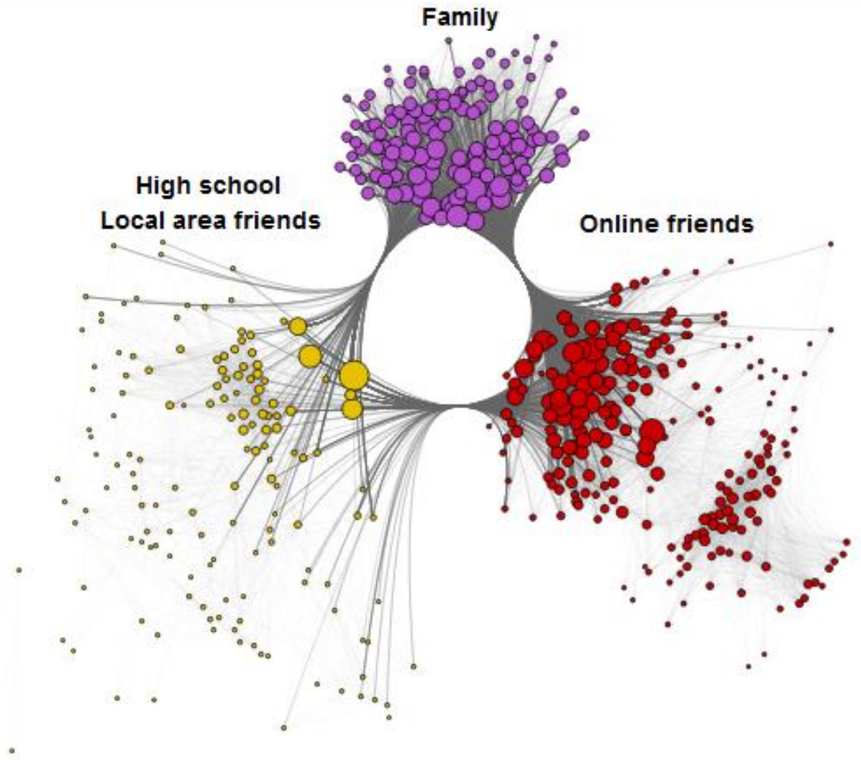


Image source: Wolfram

How important is a node?

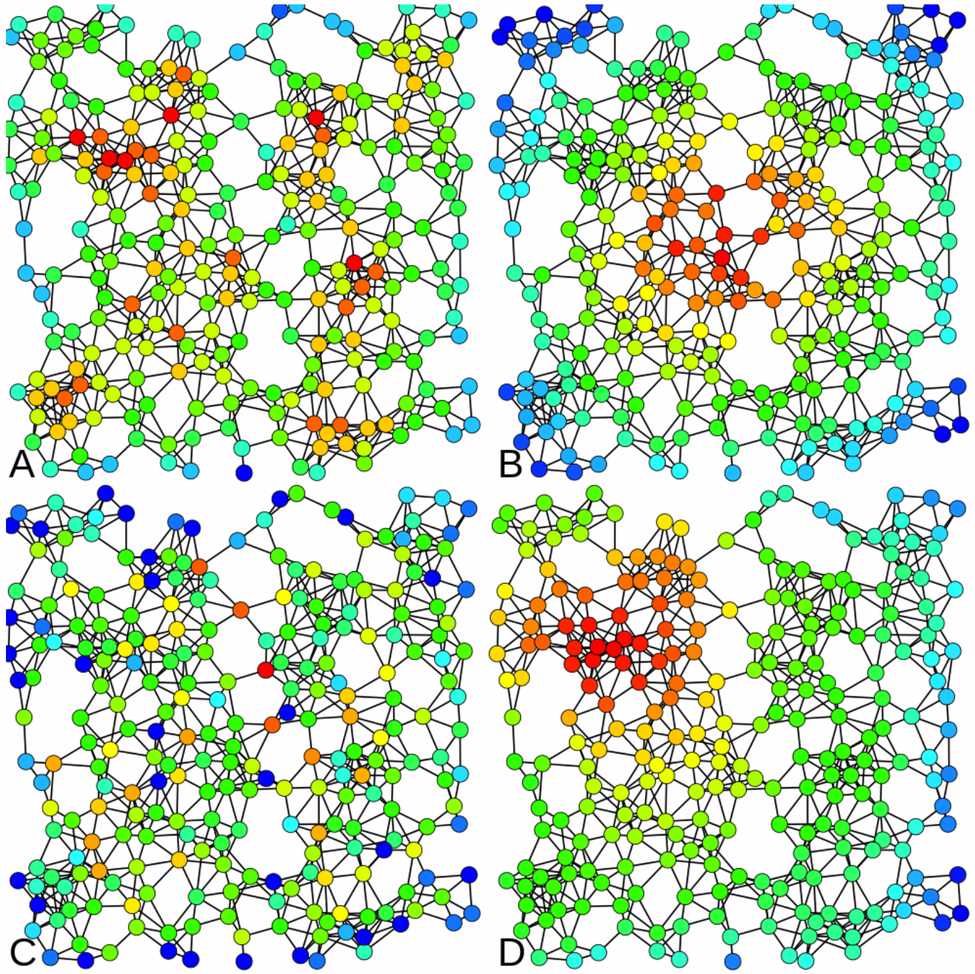


Image source: Claudio Rocchini / Wikimedia

Degree: the number of connections a node has.

Weighted degree: the number of connections of a node divided by the total number of connections in the graph.

Degree centrality: the more connections a node has, the more important it is **(A)**

Closeness centrality: the more central a node is (i.e., the shorter the path from it to all other nodes), the more important it is **(B)**

Betweenness centrality: the more often a node connects two other nodes, the more important it is **(C)**

Eigencentrality: "the more friends your friends have, the more important you are" **(D)**

Other graph metrics

Ollscoil na Gaillimhe
University of Galway

Assortativity coefficient determines with whom the "important" nodes are connected: if they are connected with other "important" nodes, the coefficient value is high, otherwise, it is low.

Clustering coefficient is the degree of interaction between a node's immediate neighbors, i.e., the probability that the node's closest neighbors are not only connected to it but also to each other.

Density is the ratio of the number of edges to the maximum possible number of edges. Communities tend to have a high clustering coefficient and high density.

Modularity measures how much denser the connections within a group are compared to the connections between groups. This metric is used to partition the graph into communities.

Graph Formats

Ollscoil na Gaillimhe
University of Galway

Graphs are usually stored in text files (.gml, .csv) or in XML files (.graphml, .gexf), where all the nodes, edges, and their attributes – for example, the name of a node or the weight of an edge – are listed.

GEXF

```
621      <node id="67" label="MotherPlutarch">
622        <attvalues>
623          <attvalue for="modularity_class" value="8"/>
624        </attvalues>
625        <viz:size value="4.0"/>
626        <viz:position x="668.9568" y="-204.65488"/>
627        <viz:color r="91" g="245" b="91"/>
628      </node>
629    </nodes>
630    <edges>
631      <edge id="0" source="1" target="0"/>
632      <edge id="1" source="2" target="0" weight="8.0"/>
633      <edge id="2" source="3" target="0" weight="10.0"/>
634      <edge id="3" source="3" target="2" weight="6.0"/>
```

CSV

```
1  Source,Target,Type,weight,book
2  Addam-Marbrand,Jaime-Lannister,Undirected,3,1
3  Addam-Marbrand,Tywin-Lannister,Undirected,6,1
4  Aegon-I-Targaryen,Daenerys-Targaryen,Undirected,5,1
5  Aegon-I-Targaryen,Eddard-Stark,Undirected,4,1
6  Aemon-Targaryen-(Maester-Aemon),Alliser-Thorne,Undirected,4,1
7  Aemon-Targaryen-(Maester-Aemon),Bowen-Marsh,Undirected,4,1
8  Aemon-Targaryen-(Maester-Aemon),Chett,Undirected,9,1
9  Aemon-Targaryen-(Maester-Aemon),Clydas,Undirected,5,1
10 Aemon-Targaryen-(Maester-Aemon),Jeor-Mormont,Undirected,13,1
11 Aemon-Targaryen-(Maester-Aemon),Jon-Snow,Undirected,34,1
12 Aemon-Targaryen-(Maester-Aemon),Samwell-Tarly,Undirected,5,1
13 Aerys-II-Targaryen,Brandon-Stark,Undirected,4,1
```

Gephi

- Free open-source software for visualising networks
- Download: https://gephi.org/
- User guide: https://gephi.org/users/
- Tutorials: https://gephi.org/users/

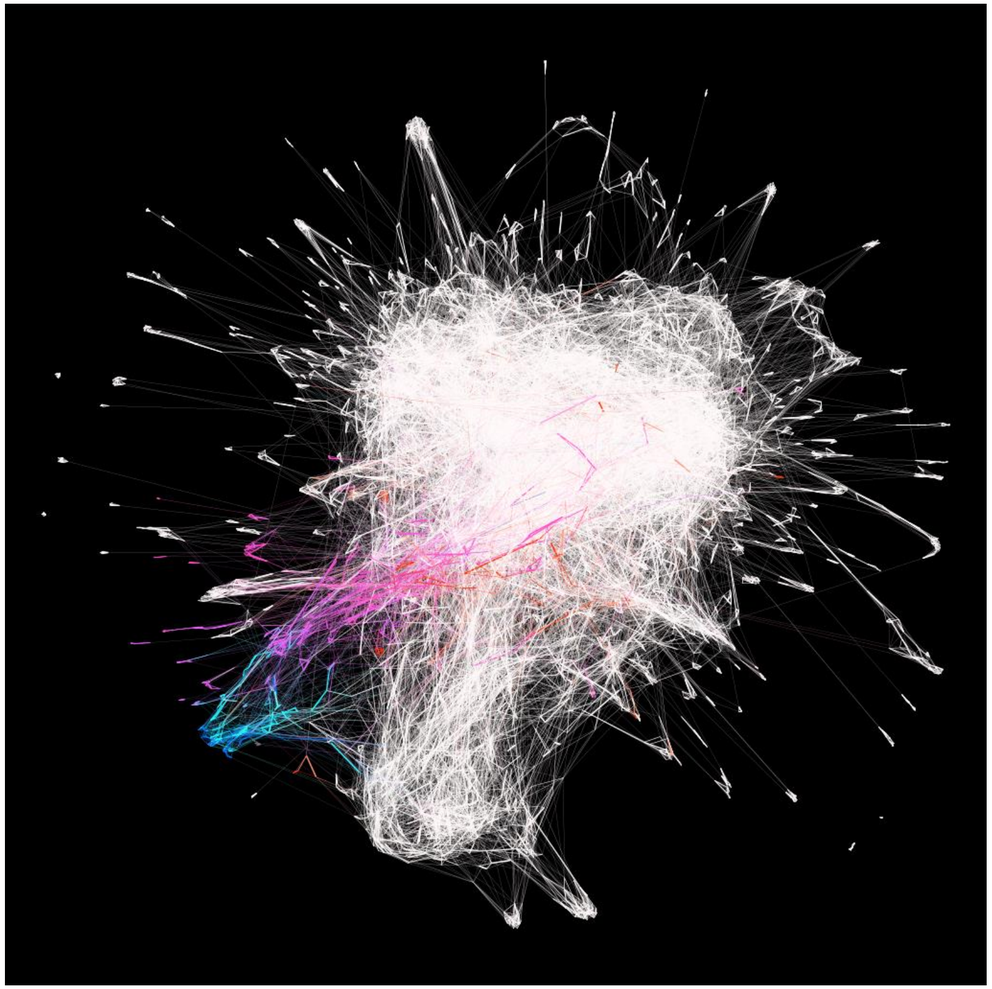
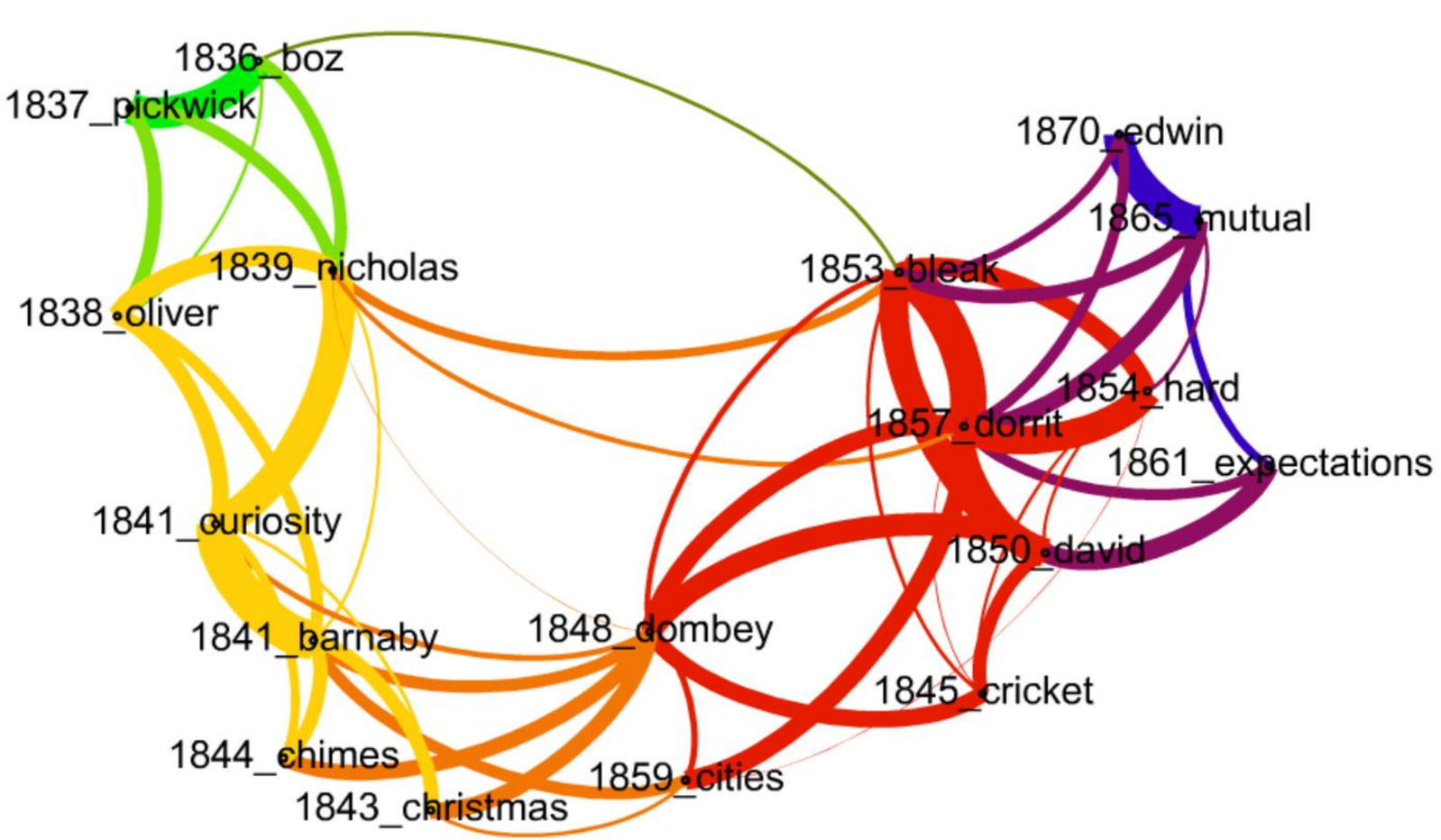


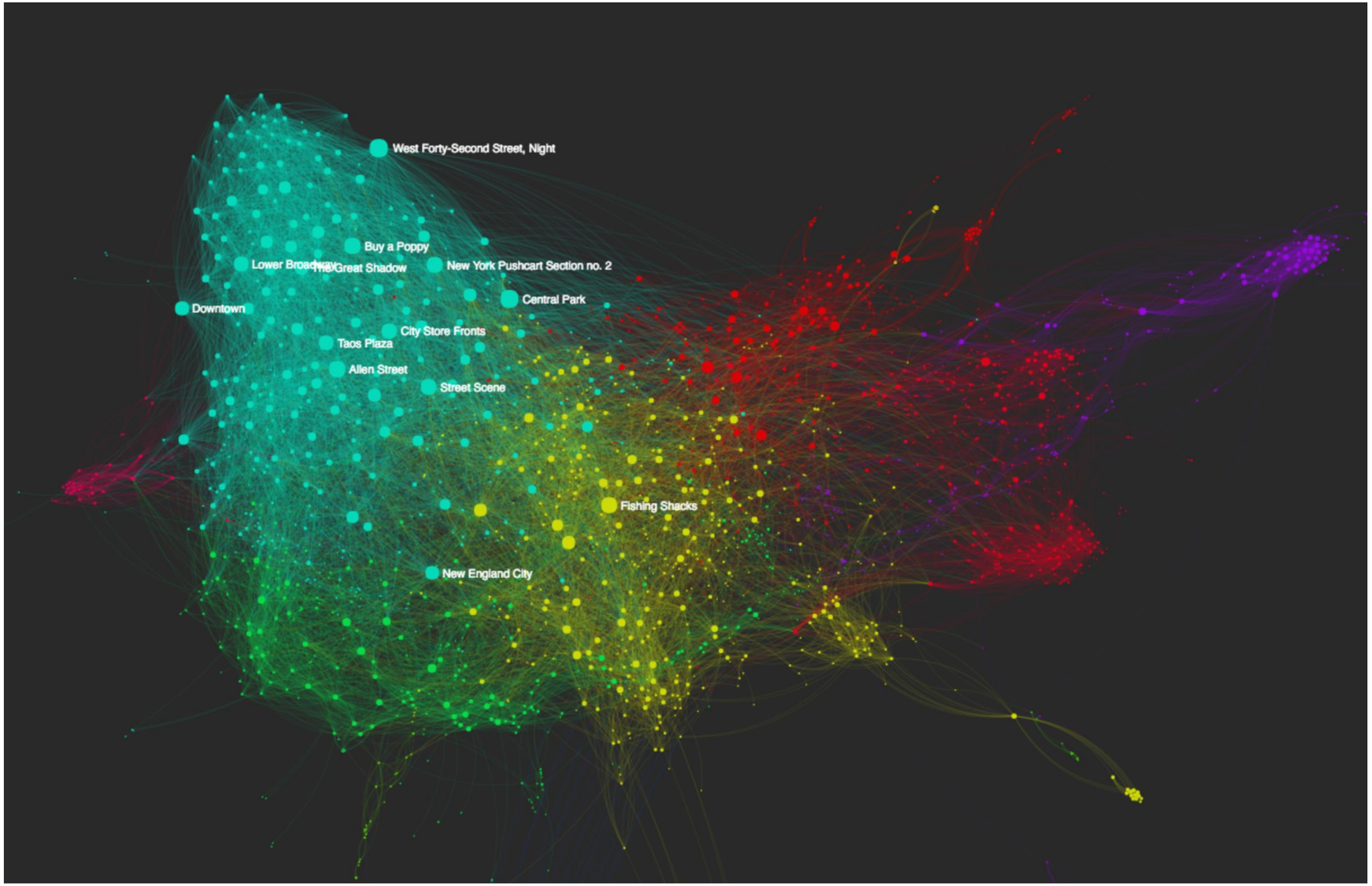
Image source: DHd 2016

Examples



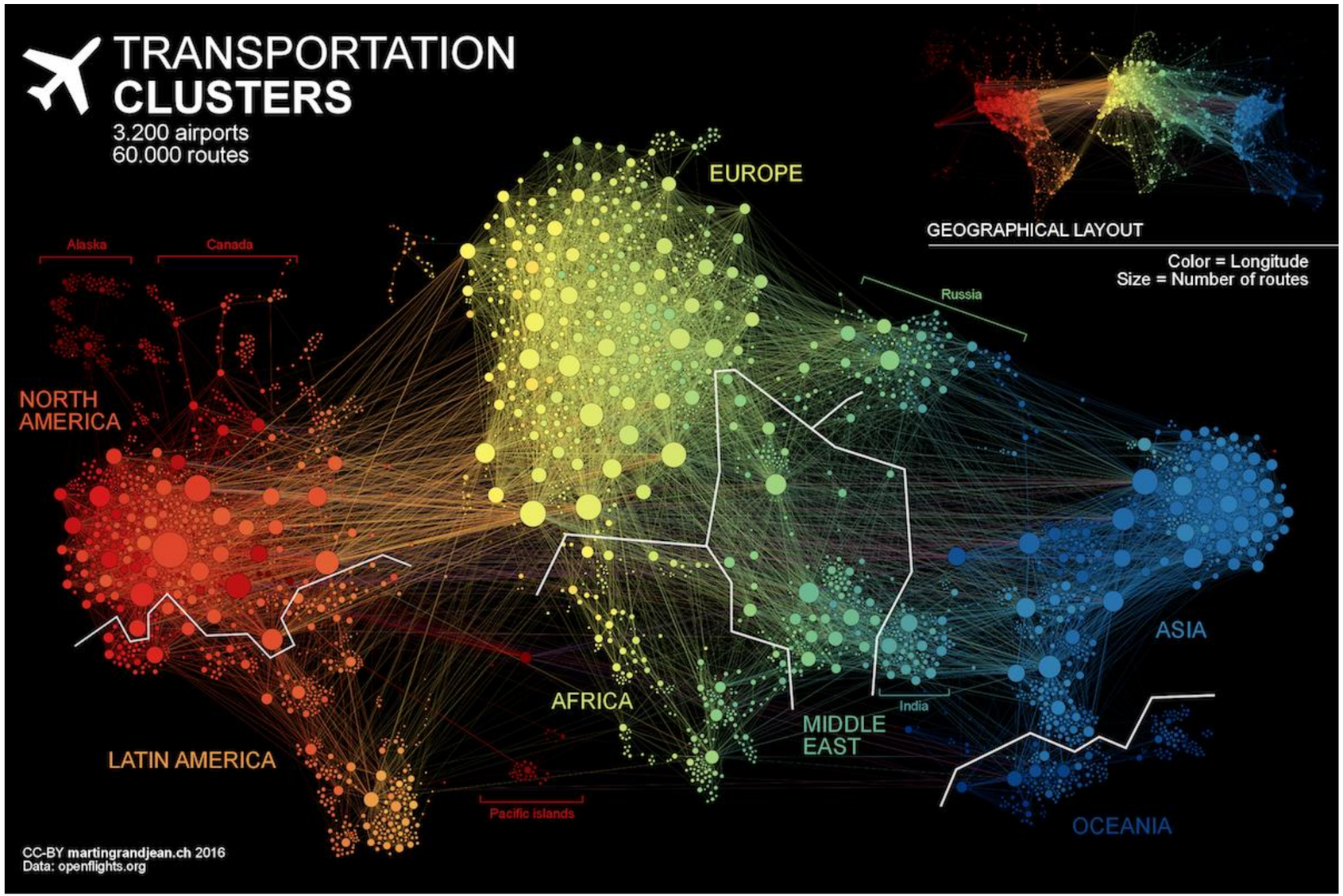
A chronology of Charles Dickens' novels built using the most frequent words in the text by Dr. Jan Rybicki.

Examples



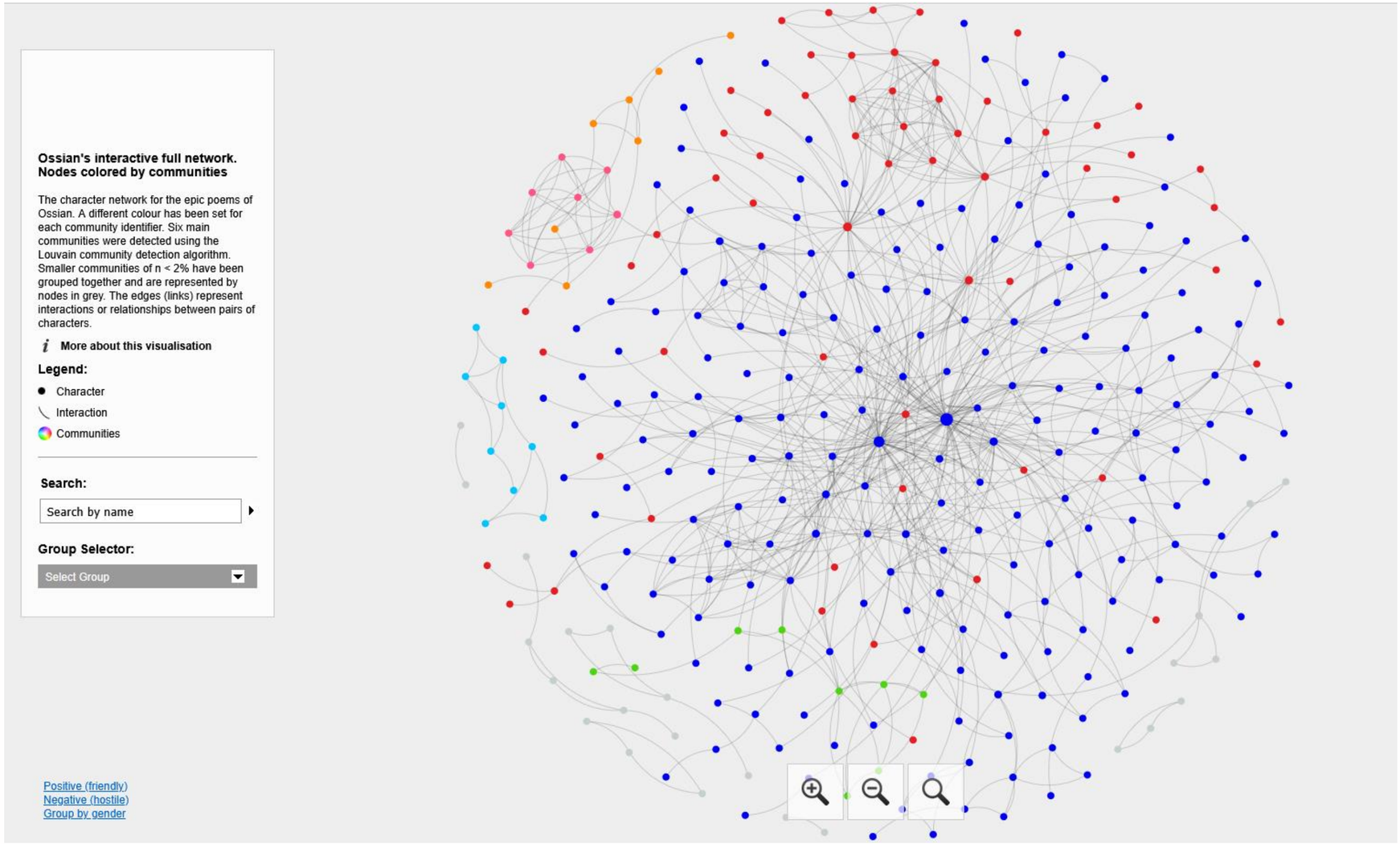
A network of thematic connections in artworks from the Smithsonian American Art Museum, made between 1900 and 1935, by Dr. Matthew Lincoln.

Examples



An air transportation network based on openflights.org data, where nodes represent airports, node sizes reflect the number of routes and colours roughly represent continents, by Dr. Martin Grandjean.

Examples



[An example of an interactive graph], created by Dr. Joseph Yose, based on the Ossian cycle of epic poems by James Macpherson.

Data Sources

Ollscoil na Gaillimhe
University of Galway

Workshop materials: https://github.com/universityofgalway-library/digital-tools-workshops/tree/main/networks-gephi

Alternatively, download data for the workshop from Kaggle

Other websites & data repositories with interesting graph data

- Network Data Repository
- Stanford Large Network Dataset Collection
- Colorado Index of Complex Networks
- University of Michigan
- Network Corpus

Ollscoil na Gaillimhe
University of Galway

Hands-on session

University *of* Galway.ie

Further Reading

1. Martin Grandjean. (2021). Introduction to Social Network Analysis: Basics and Historical Specificities. [video & pdf]
2. Andrew Beveridge. (2020). Network of Thrones. A Song of Maths and Westeros.
3. Gephi Documentation
4. Gephi tutorials (official & created by users)
5. Billy Mosse. (2020). A quick tutorial on Gephi layouts — using a Q&A StackOverflow dataset. Medium.
6. Brown University Libraries. Network Analysis with Gephi
7. Christine Egan. (2023). Social Network Analysis and Visualization with Python. Medium.
8. Mohammed Saqr, Sonsoles López-Pernas, Miguel Ángel Conde-González & Ángel Hernández-García. (2024). Social Network Analysis: A Primer, a Guide and a Tutorial in R. In: Saqr, M., López-Pernas, S. (eds) Learning Analytics Methods and Tutorials. Springer, Cham. https://doi.org/10.1007/978-3-031-54464-4_15