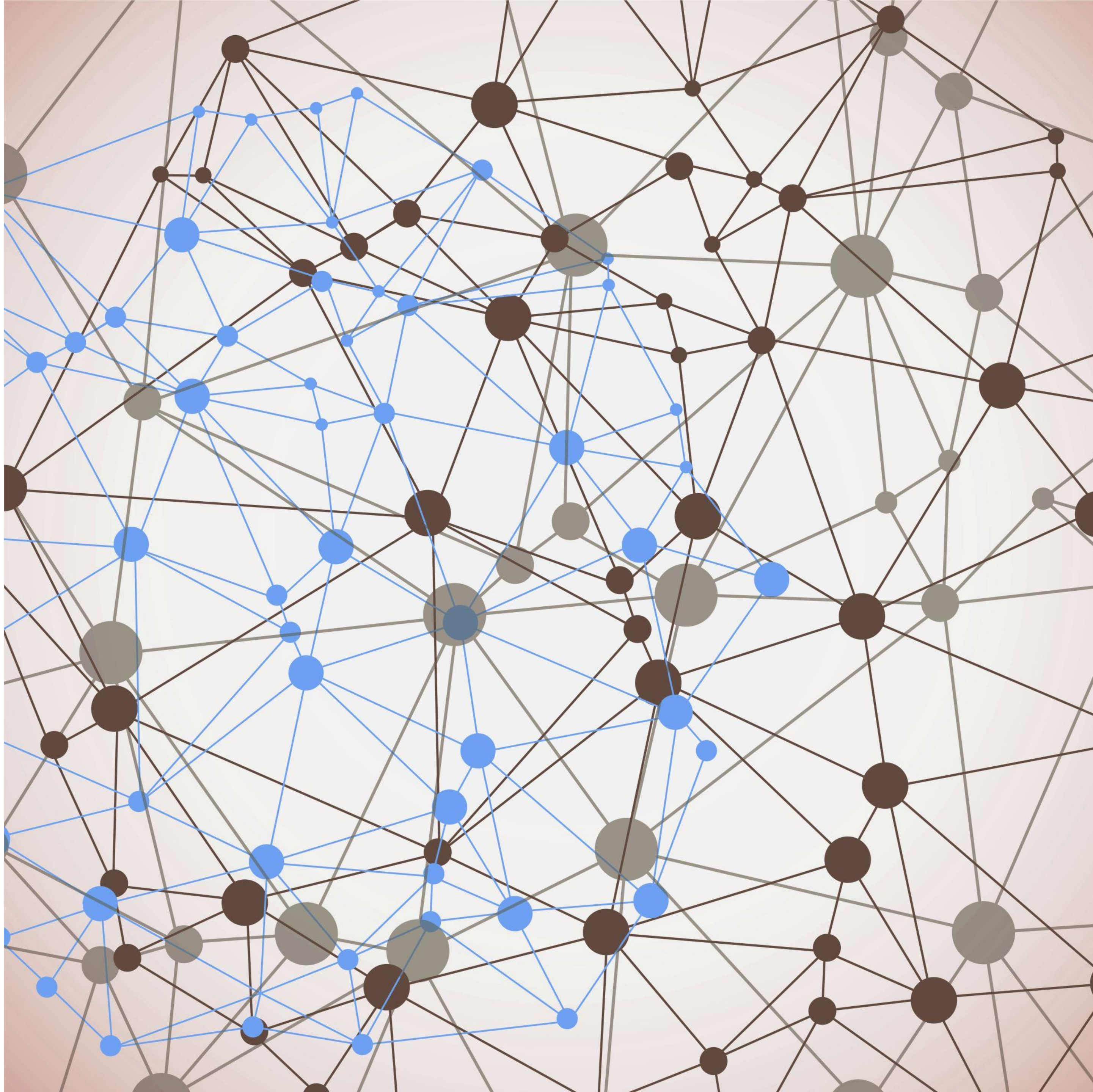
Who is the most influential physicist?

Varghese Jacob

CID : 01716176

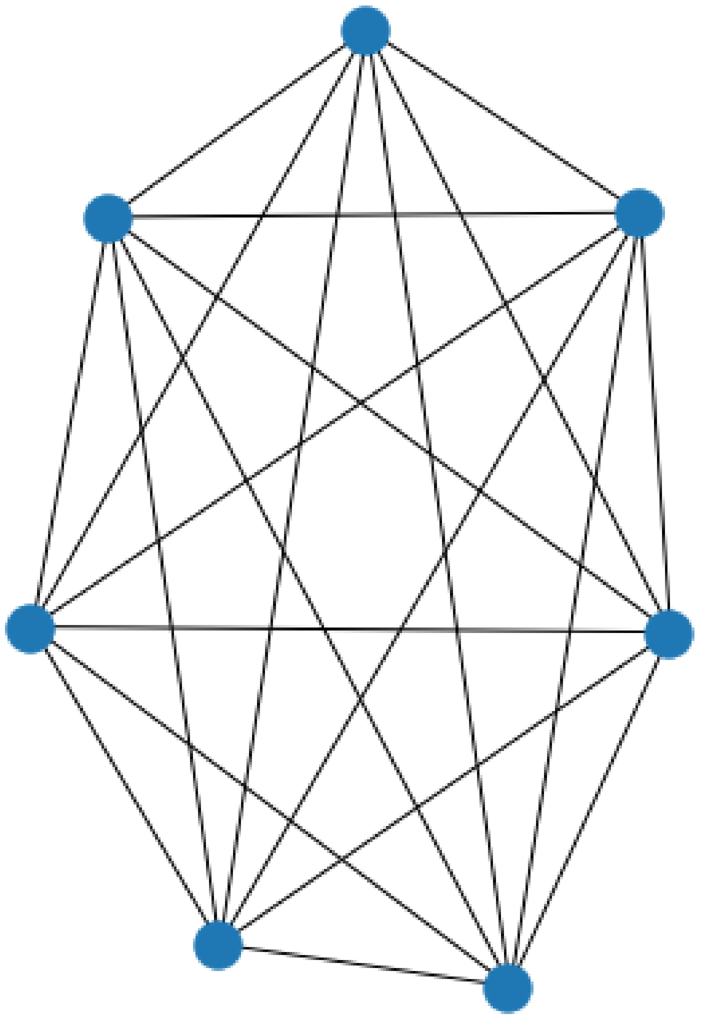


What is a network?

- Consists of objects, referred to as nodes, and the connections between them which are referred to as links or edges

- The edges can be symmetric or directed, depending on what the network is a representation of

- As the number of nodes increases, they can start to exhibit some complex characteristics and trends can be identified

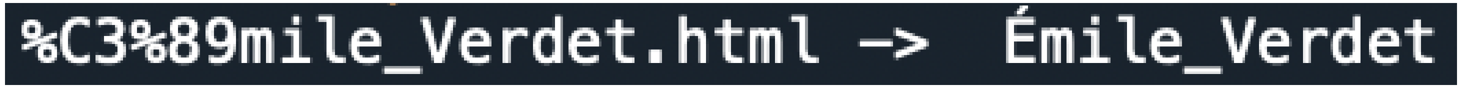


Forming the network

- To obtain measures of how influential different physicists are, we must form a network and analyze each node

- By searching through the filenames of over 1000 physicist's downloaded Wikipedia pages, we can extract the name of each physicist and form a node.

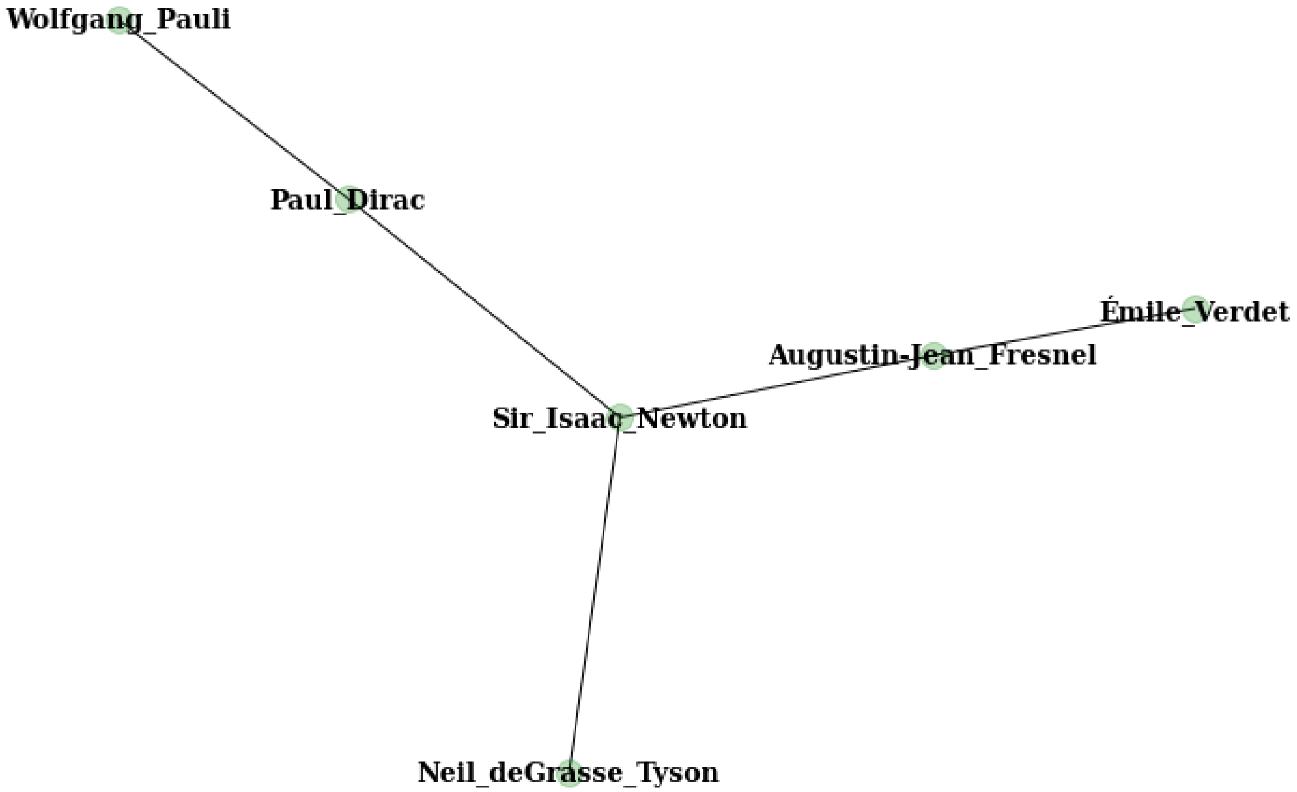
- This mostly involved simply removing the '.html' extension but in some cases where the name has an accent, the character had to be encoded from the URL-escaped character to the accented character in UTF-8



```
%C3%89mile_Verdet.html ->  Émile_Verdet
```

Forming the network

- By searching through individual Wikipedia pages, we can find the hyperlinks present and form the edges of the graph when the name of another physicist is contained in one of the hyperlinks, e.g. '/wiki/Augustin-Jean_Fresnel' found in Émile Verdet's page

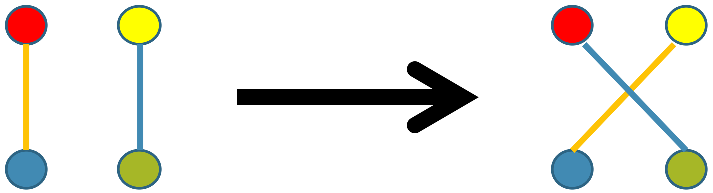


Introducing noise

- As the data we are analysing is crowd-sourced, uncertainty is minimized by utilizing a large sample size, however, we still need a way to quantify it

- A way to do this is to introduce some random noise and see how the identified trends are affected

- By selecting a percentage of existing edges at random and rearranging them, the effect of the slight perturbation can be compared to other instances of the same method, and become subject to standard statistical methods



Centrality measures

- To answer the question at hand, we need to quantify how important each node is in the network
- In network analysis, the importance of a node can be quantified in terms of centrality relative to the entire network
- The indicators that we will be investigating are degree centrality, closeness centrality and betweenness centrality
- Degree centrality is simply the degree of a node normalized by a factor of n-1 where n is the number of nodes in the network

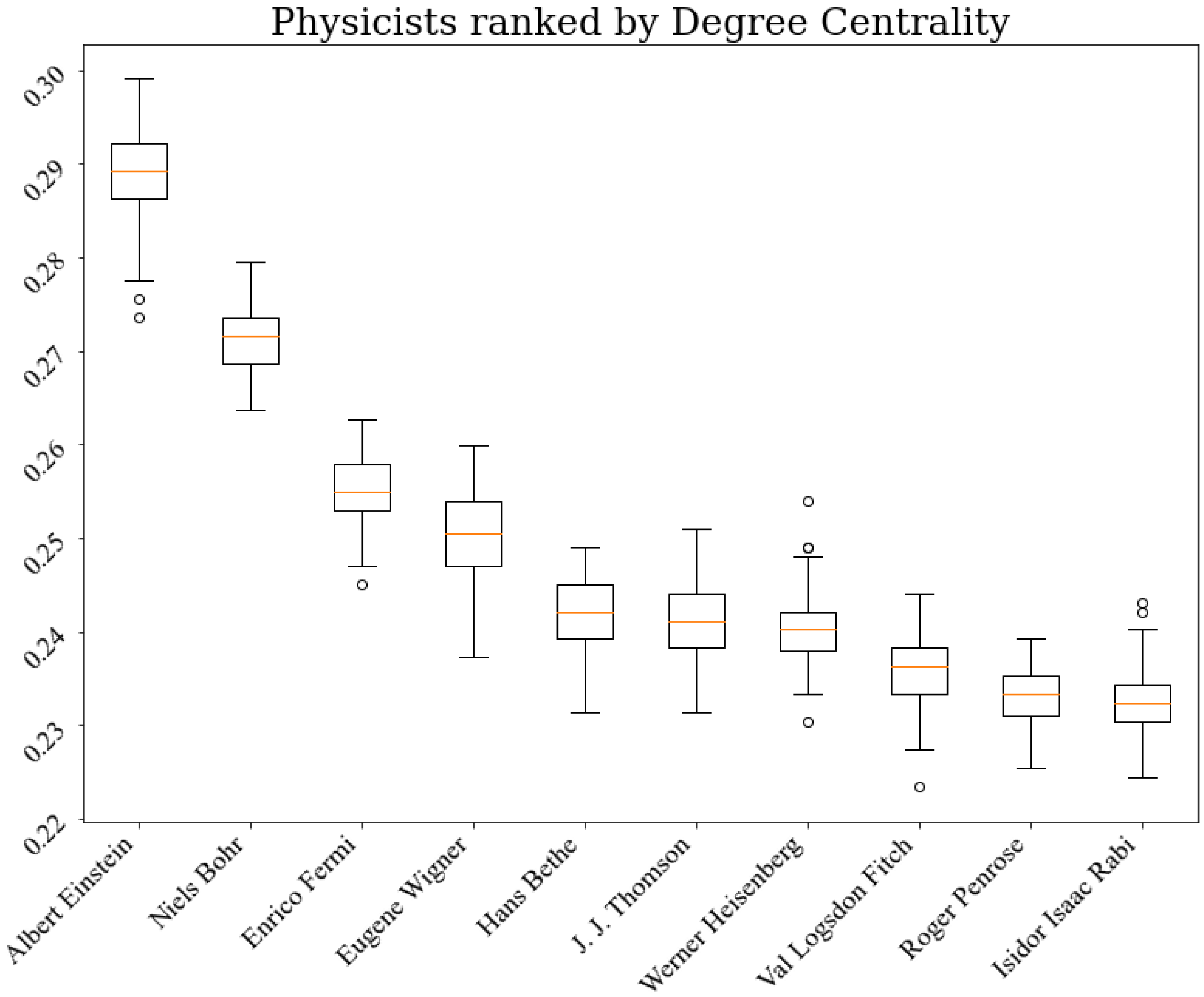
Centrality measures

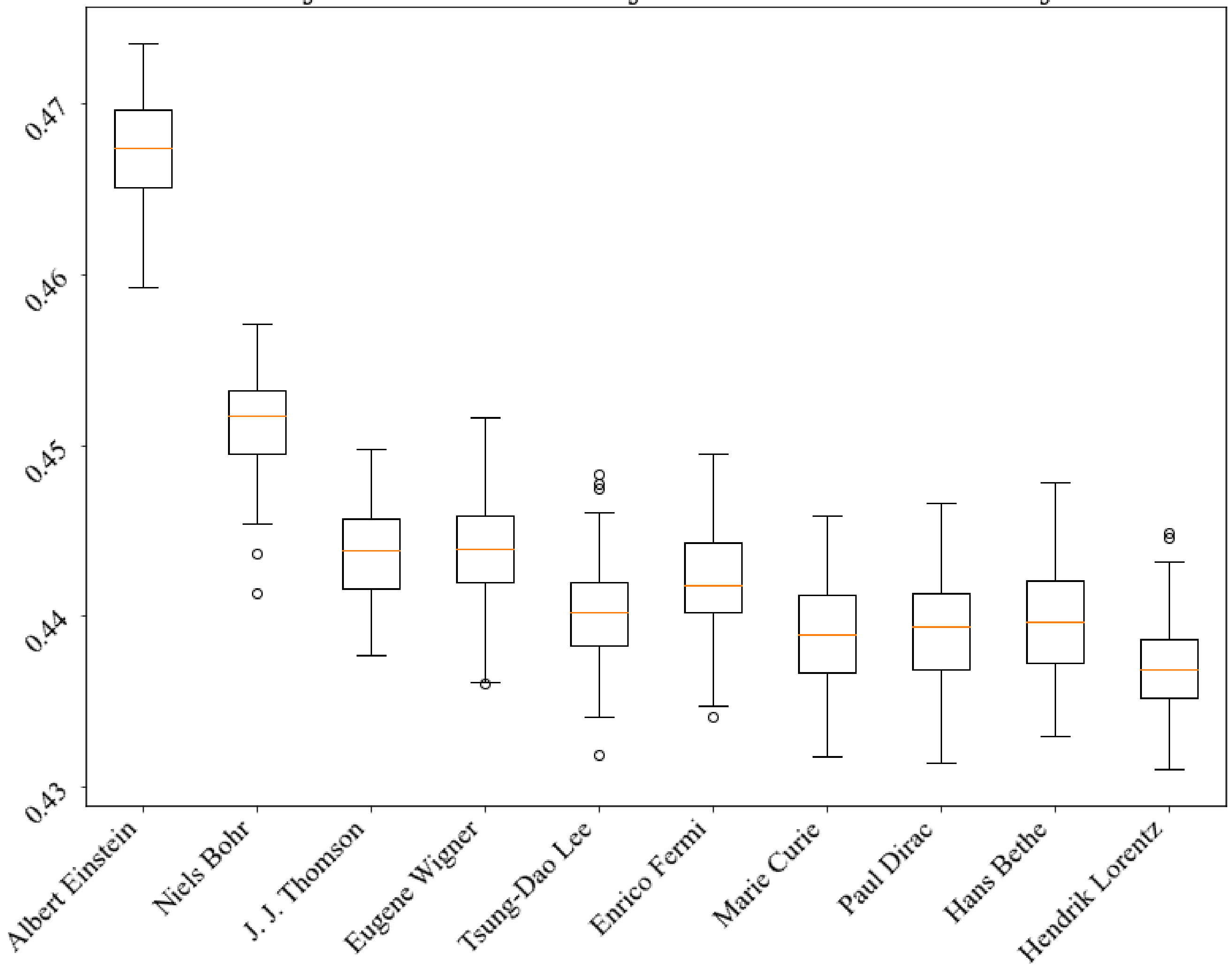
- Closeness centrality of a node, *u*, is the inverse of the average length of the shortest paths to all other nodes in a network

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$$

- Betweenness centrality of a node, *u*, is the sum of the ratio of shortest paths between all pairs of nodes that pass through *u*

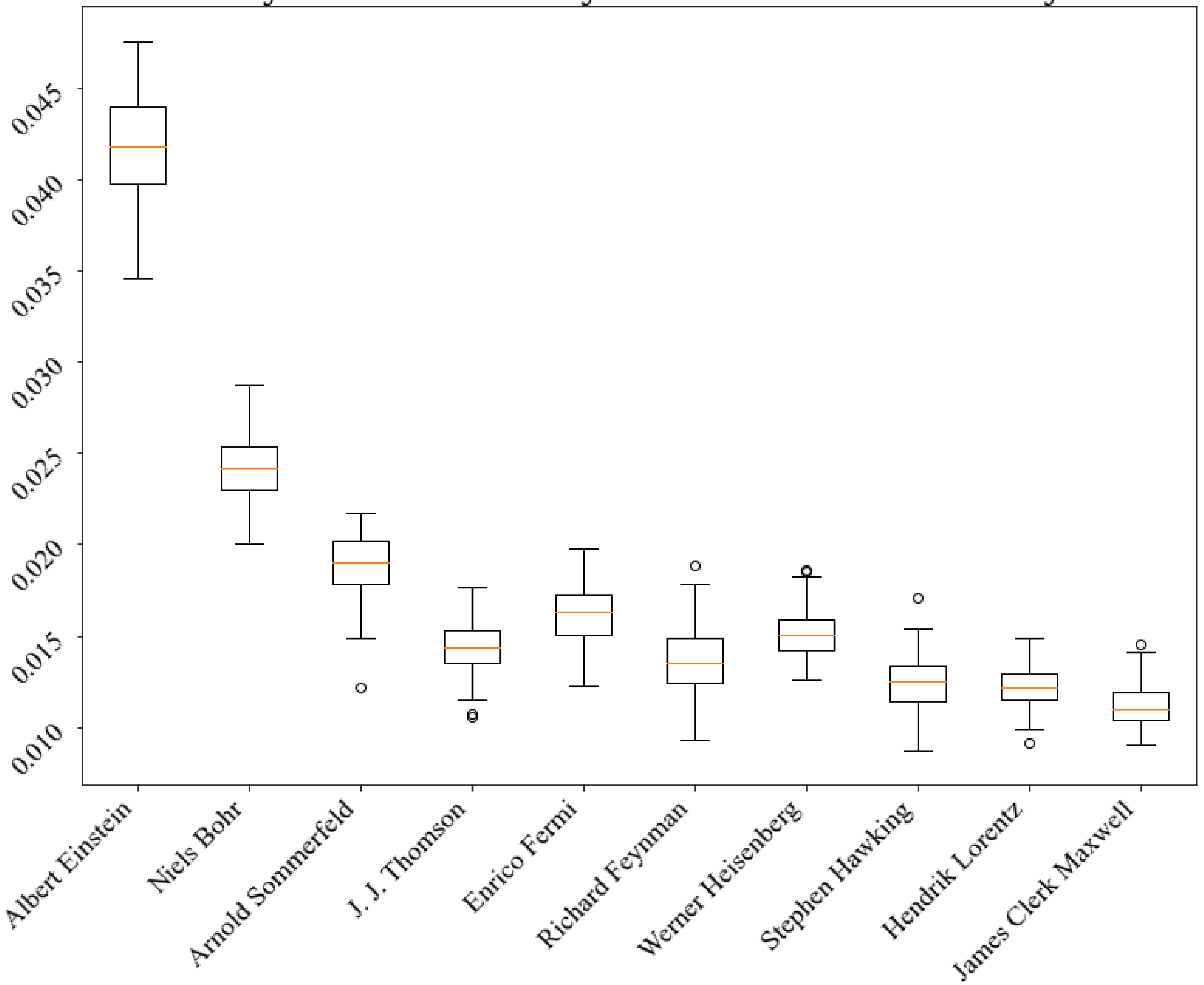
$$B(u) = \sum_{x,y \in N} \frac{\sigma(x,y|u)}{\sigma(x,y)}$$

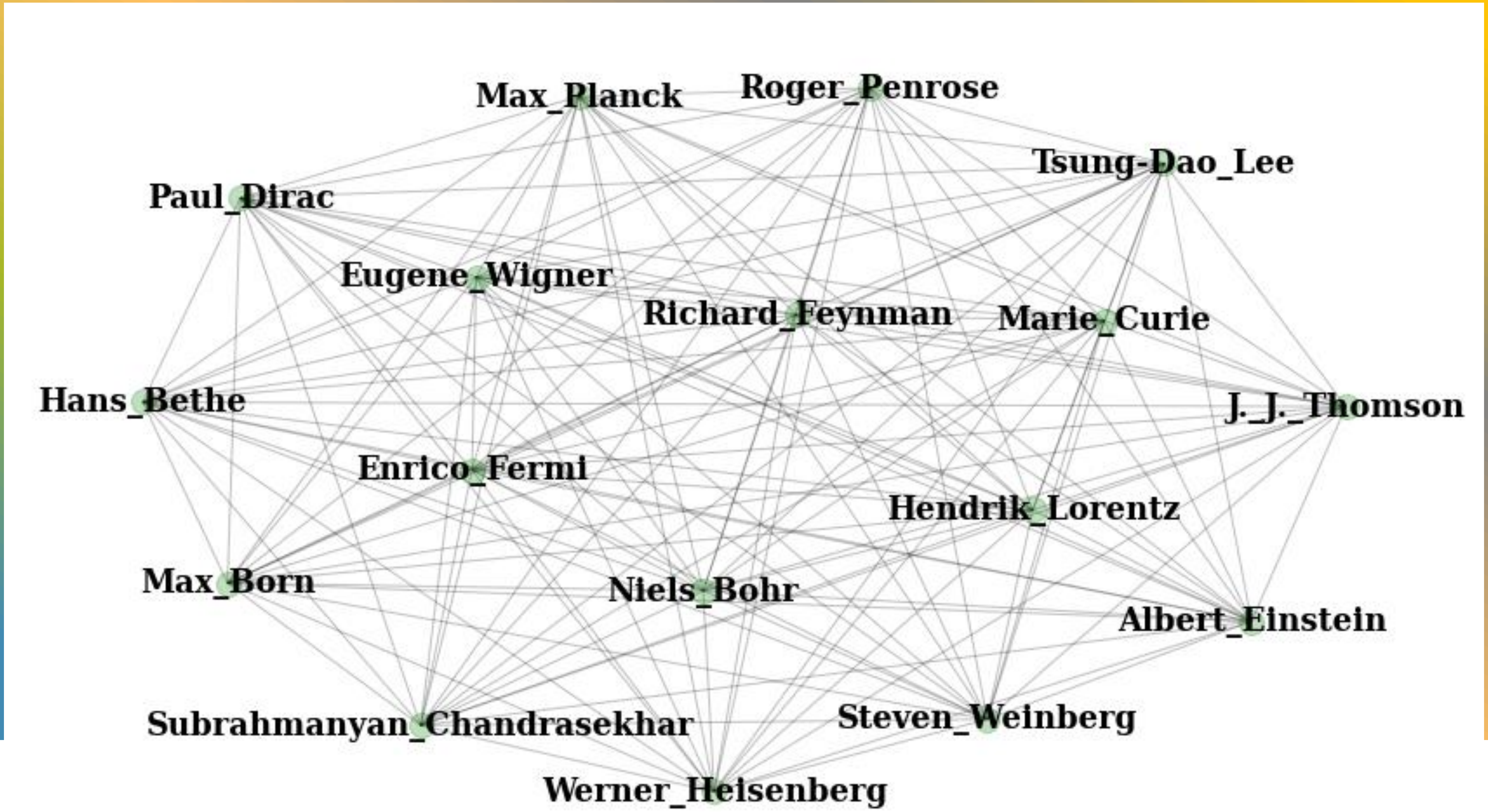

Physicists ranked by Degree Centrality
0.30
0.29
0.28
0.27
0.26
0.25
0.24
0.23
0.22
Albert Einstein
Niels Bohr
Enrico Fermi
Eugene Wigner
Hans Bethe
J. J. Thomson
Werner Heisenberg
Val Logsdon Fitch
Roger Penrose
Isidor Isaac Rabi



Physicists ranked by Closeness Centrality

Physicists ranked by Betweenness Centrality




Max_Planck
Roger_Penrose
Tsung-Dao_Lee
Paul_Dirac
Eugene_Wigner
Richard_Feynman
Marie_Curie
Hans_Bethe
J._J._Thomson
Enrico_Fermi
Hendrik_Lorentz
Max_Born
Niels_Bohr
Albert_Einstein
Subrahmanyan_Chandrasekhar
Steven_Weinberg
Werner_Heisenberg

Conclusion and Further work

- Formed a network of physicists by searching through over 1000 Wikipedia pages
- Calculated values for centrality measures of each physicist and gained evidence to suggest Albert Einstein is the most influential physicist
- Can try using the python package Natural Language Toolkit to identify details such as, field of physics, duration of career and whether a link should be directed