Digital Humanities Across Borders

Class 17: Networks (part 2)

Reminder on terminology

Reminder on terminology

- Nodes: things in your network

Reminder on terminology

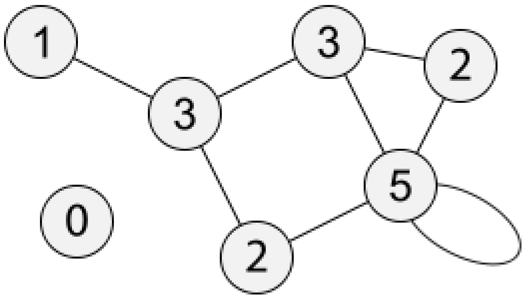
- Nodes: things in your network
- Edges: lines between things in your network

Reminder on terminology

- Nodes: things in your network
- Edges: lines between things in your network
- Directed network: property connecting your nodes only works one direction (e.g. *person* **is author of** *book*)

Reminder on terminology

- Nodes: things in your network
- Edges: lines between things in your network
- Directed network: property connecting your nodes only works one direction (e.g. *person* **is author of** *book*)
- Degree: the number of edges that connect to a node (self-connections count twice!)



Degree centrality

Degree centrality

- Works best if you have all the nodes and edges (significantly impacted by sampling)

Degree centrality

- Works best if you have all the nodes and edges (significantly impacted by sampling)
- Gets more complicated if you have a bimodal network; less meaningful the more kinds of relationships there are

Degree centrality

- Works best if you have all the nodes and edges (significantly impacted by sampling)
- Gets more complicated if you have a bimodal network; less meaningful the more kinds of relationships there are
- "Power law properties" in degree: only a few nodes have high degree
    - Most things work this way (city population, word frequency, citation networks, etc.)

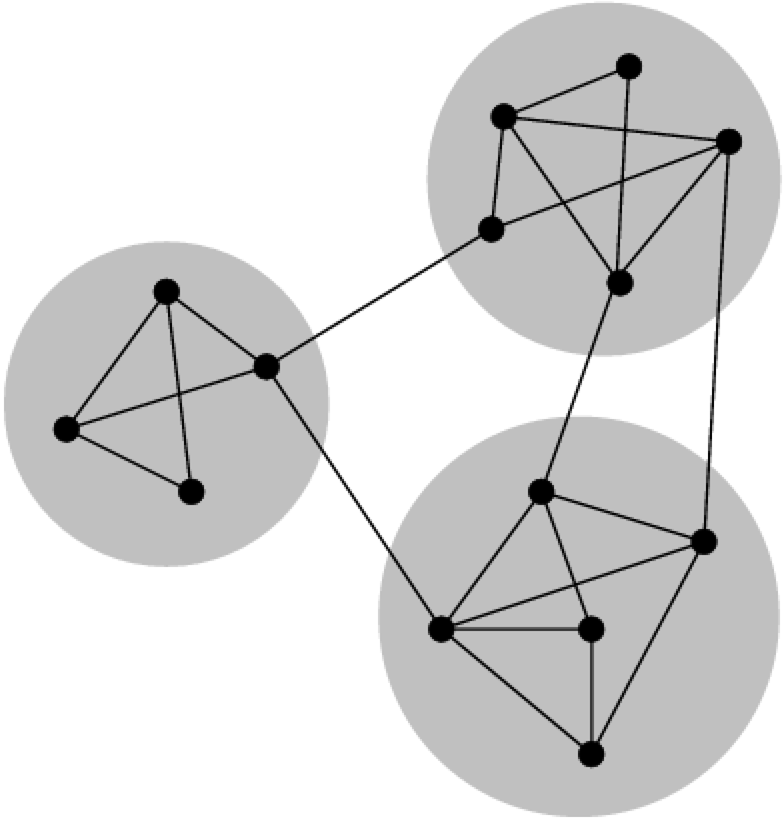
Degree centrality

- Works best if you have all the nodes and edges (significantly impacted by sampling)
- Gets more complicated if you have a bimodal network; less meaningful the more kinds of relationships there are
- "Power law properties" in degree: only a few nodes have high degree
    - Most things work this way (city population, word frequency, citation networks, etc.)
- Degree + edge weight = **strength**

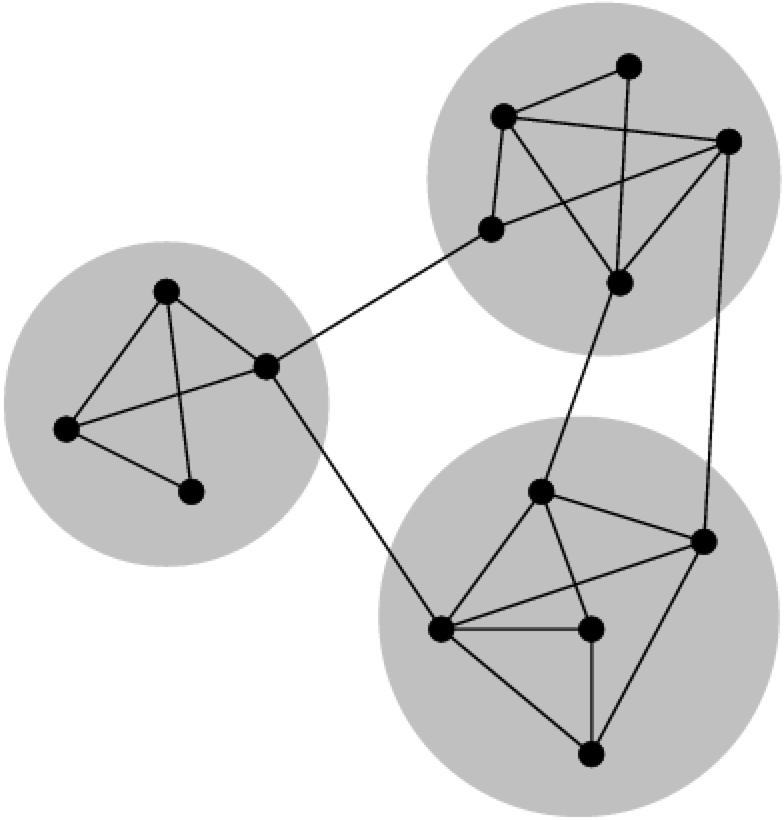
Degree centrality

- Works best if you have all the nodes and edges (significantly impacted by sampling)
- Gets more complicated if you have a bimodal network; less meaningful the more kinds of relationships there are
- "Power law properties" in degree: only a few nodes have high degree
  - Most things work this way (city population, word frequency, citation networks, etc.)
- Degree + edge weight = **strength**
- Can look at in-degree and out-degree for directed networks

Modularity/Communities

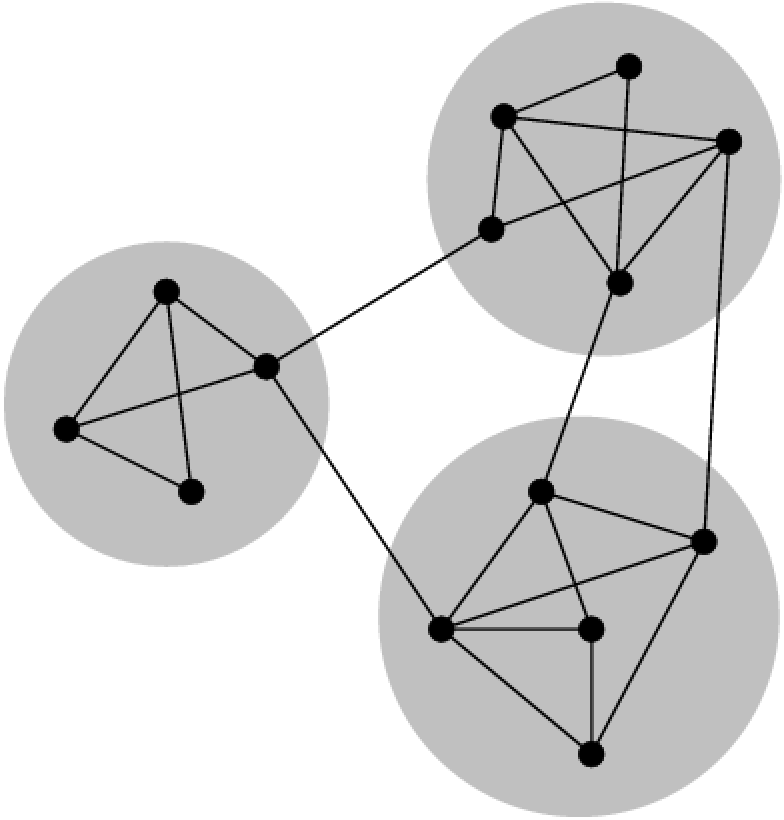


Modularity/Communities



- Have more in-connections than out-connections
- But you should subtract the expected number in an equivalent network with random edges
- Modularity algorithms draw circles around groups of nodes to get the highest possible score

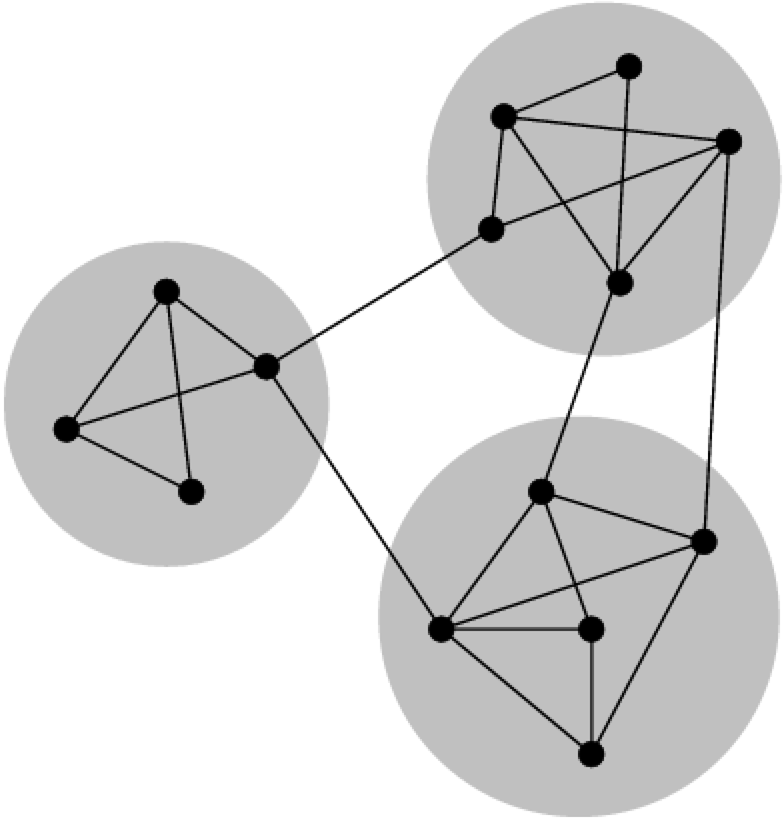
Modularity/Communities



- Have more in-connections than out-connections
- But you should subtract the expected number in an equivalent network with random edges
- Modularity algorithms draw circles around groups of nodes to get the highest possible score

Modularity algorithms don't count directed edges correctly.

Modularity/Communities



- Have more in-connections than out-connections
- But you should subtract the expected number in an equivalent network with random edges
- Modularity algorithms draw circles around groups of nodes to get the highest possible score

Modularity algorithms don't count directed edges correctly.

Including yourself (e.g. when grouping Twitter followers) can also skew the results.

Structure doesn't indicate movement

Structure doesn't indicate movement

- Correspondence networks only show people who wrote letters, obscuring local connections

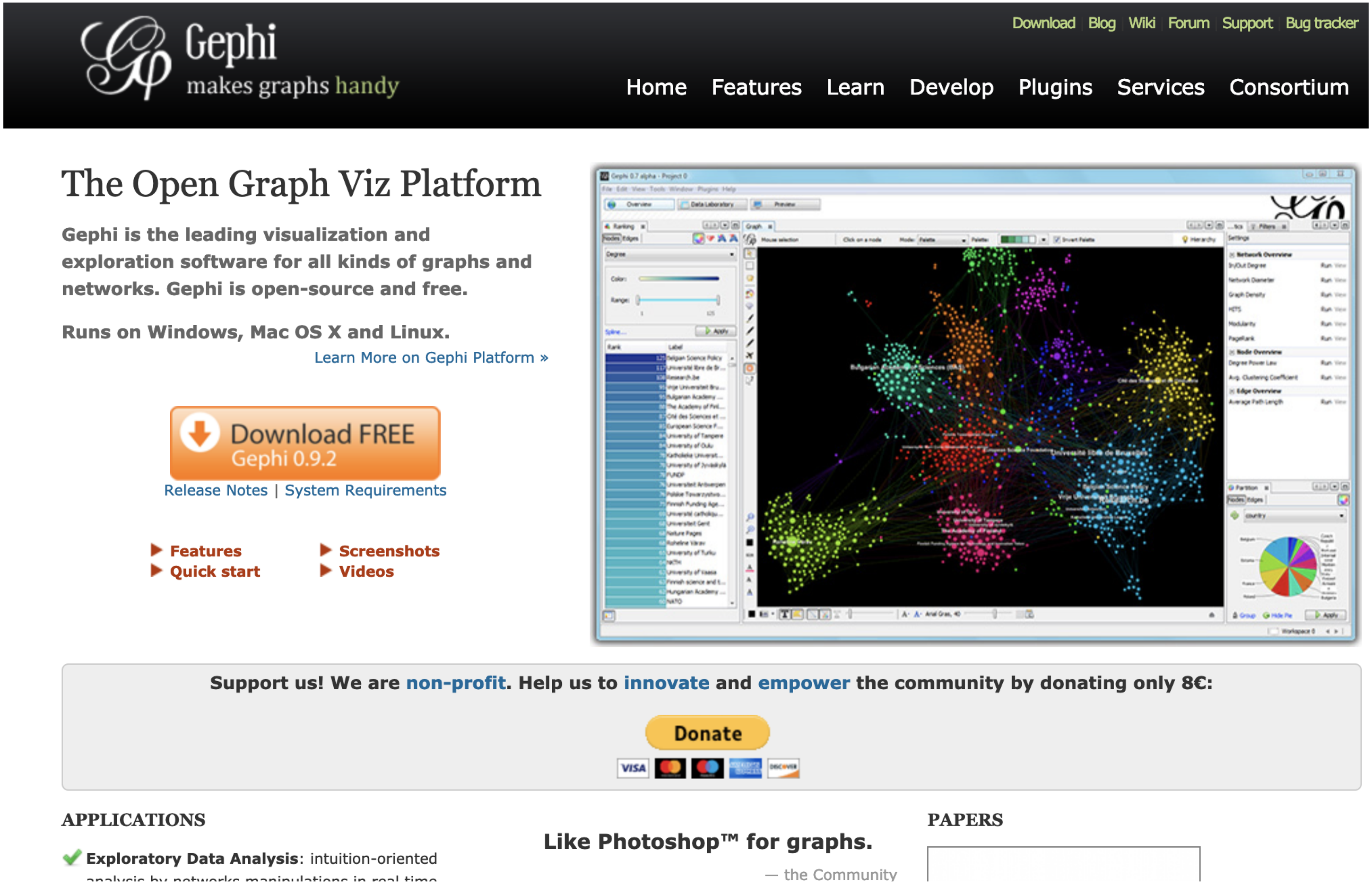
Structure doesn't indicate movement

- Correspondence networks only show people who wrote letters, obscuring local connections
- Family members probably got different kinds of correspondence than professional colleagues

Structure doesn't indicate movement

- Correspondence networks only show people who wrote letters, obscuring local connections
- Family members probably got different kinds of correspondence than professional colleagues
- Network centrality doesn't necessarily mean that a person is a key channel of information.

Gephi



Widely-used but brittle; have to have specific headers; UTF-8 w/ BOM (byte-order mark) isn't recognized; (I find it) very unintuitive to use

Meet Cytoscape

https://cytoscape.org/

Geared towards molecular biologists, but also provides core network analysis tools that cover what we need (with less headache).

Other options

webstart download demo

visual social networks

forum manual about

Visone

https://archaeologicalnetworks.wordpress.com/resources/

(h/t Shawn Graham)

Sci2

https://sci2.cns.iu.edu/user/index.php

"Fairly buggy... specializes in ingesting data of various formats and turning them into networks for analysis and visualization"

(Scott Weingart, 2013)

NODEXL

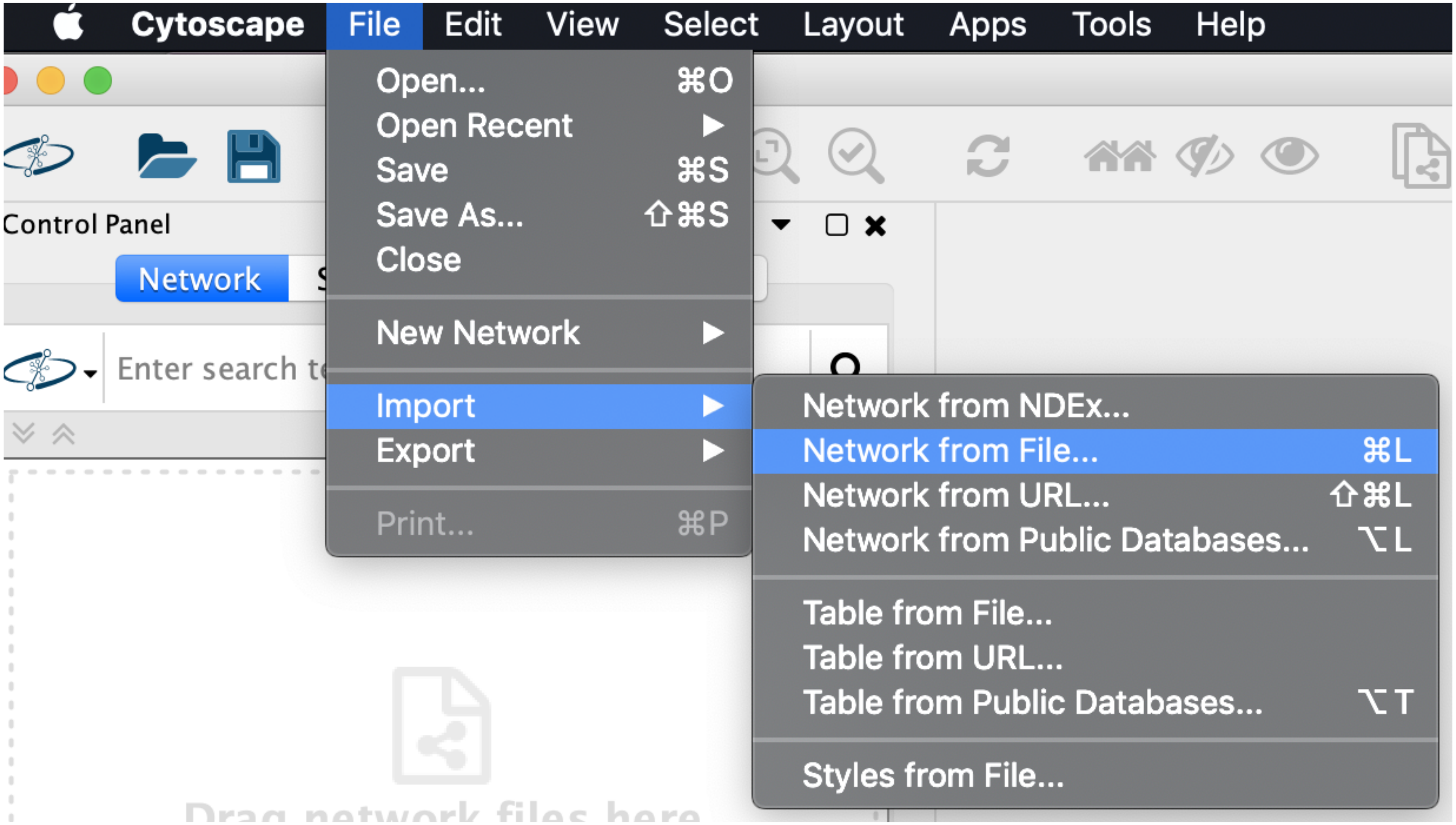
A project from the socialmedia RESEARCH FOUNDATION

NodeXL

https://www.smrfoundation.org/nodexl/

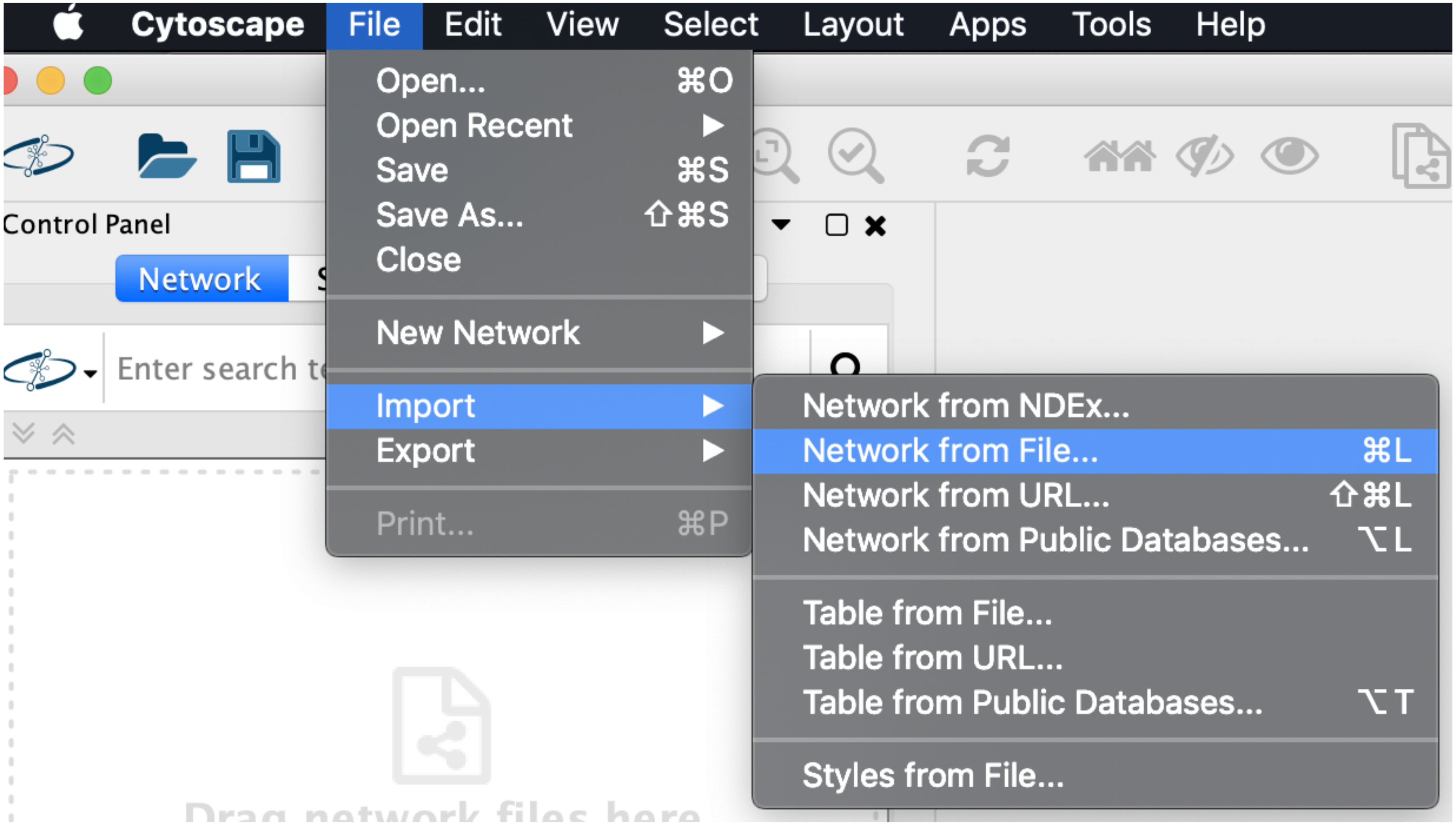
Excel plugin, Windows-only, need $40/yr license to do much

Getting data into Cytoscape



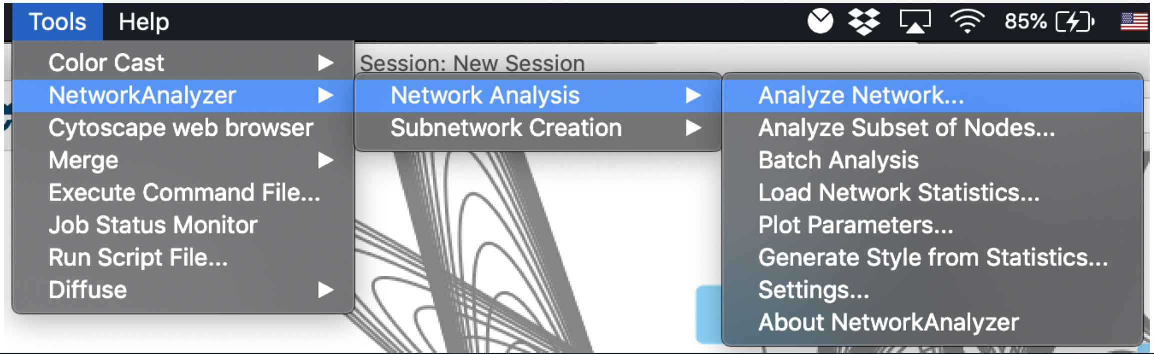
Choose the Harry Potter network for your language (from last week) or another file of network information that you're working with.

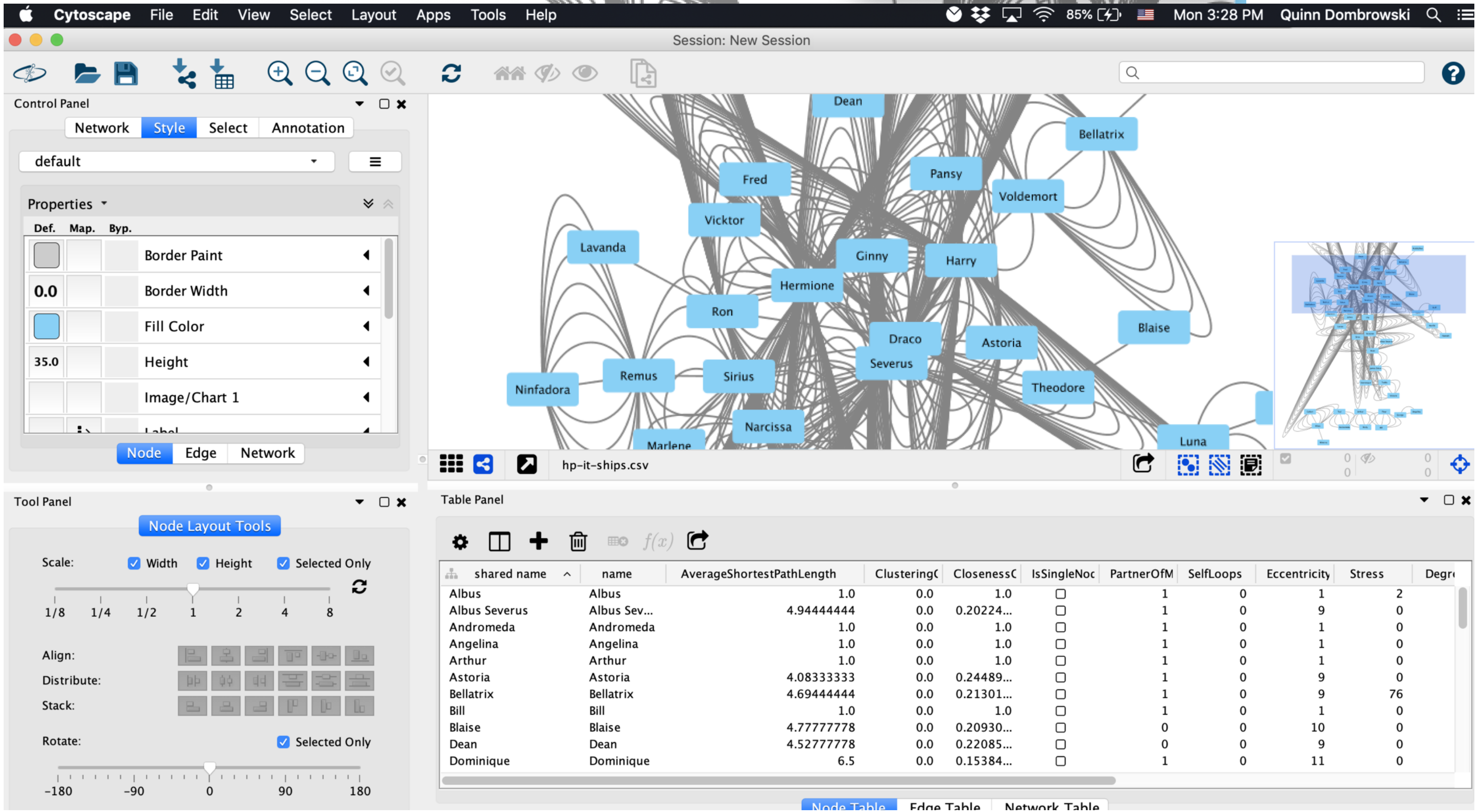
Getting data into Cytoscape



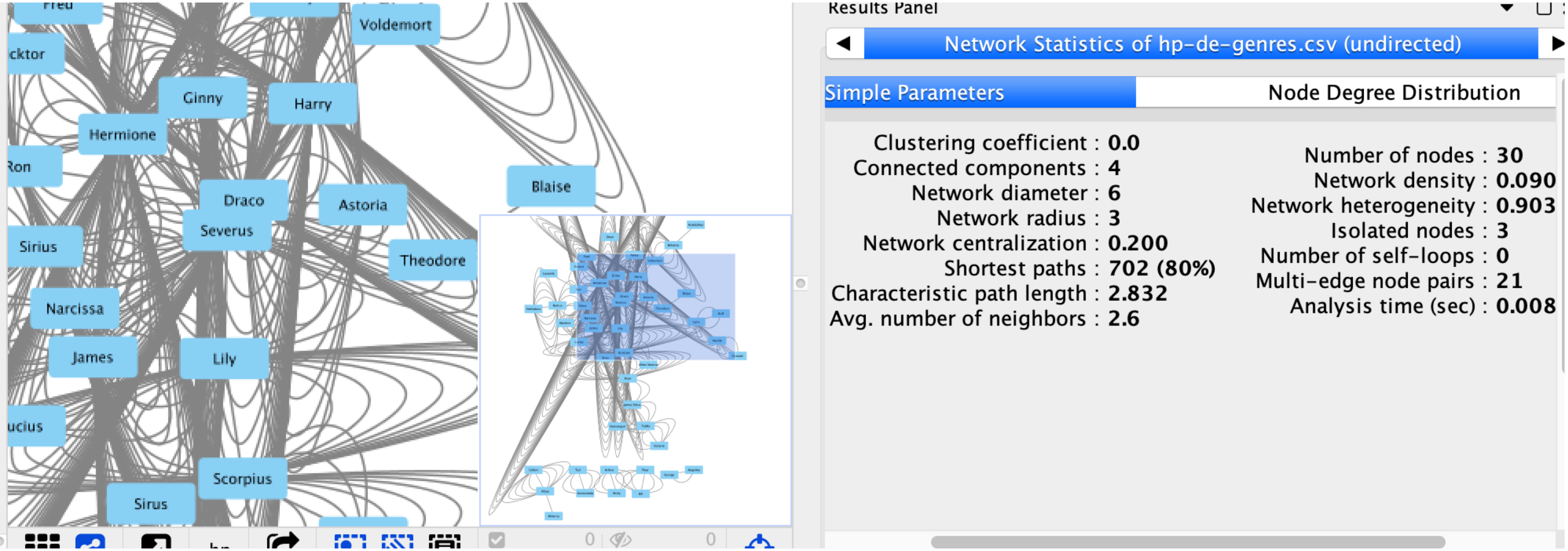
Choose the Harry Potter network for your language (from last week) or another file of network information that you're working with. **Each column must have a header.**

Analyzing network



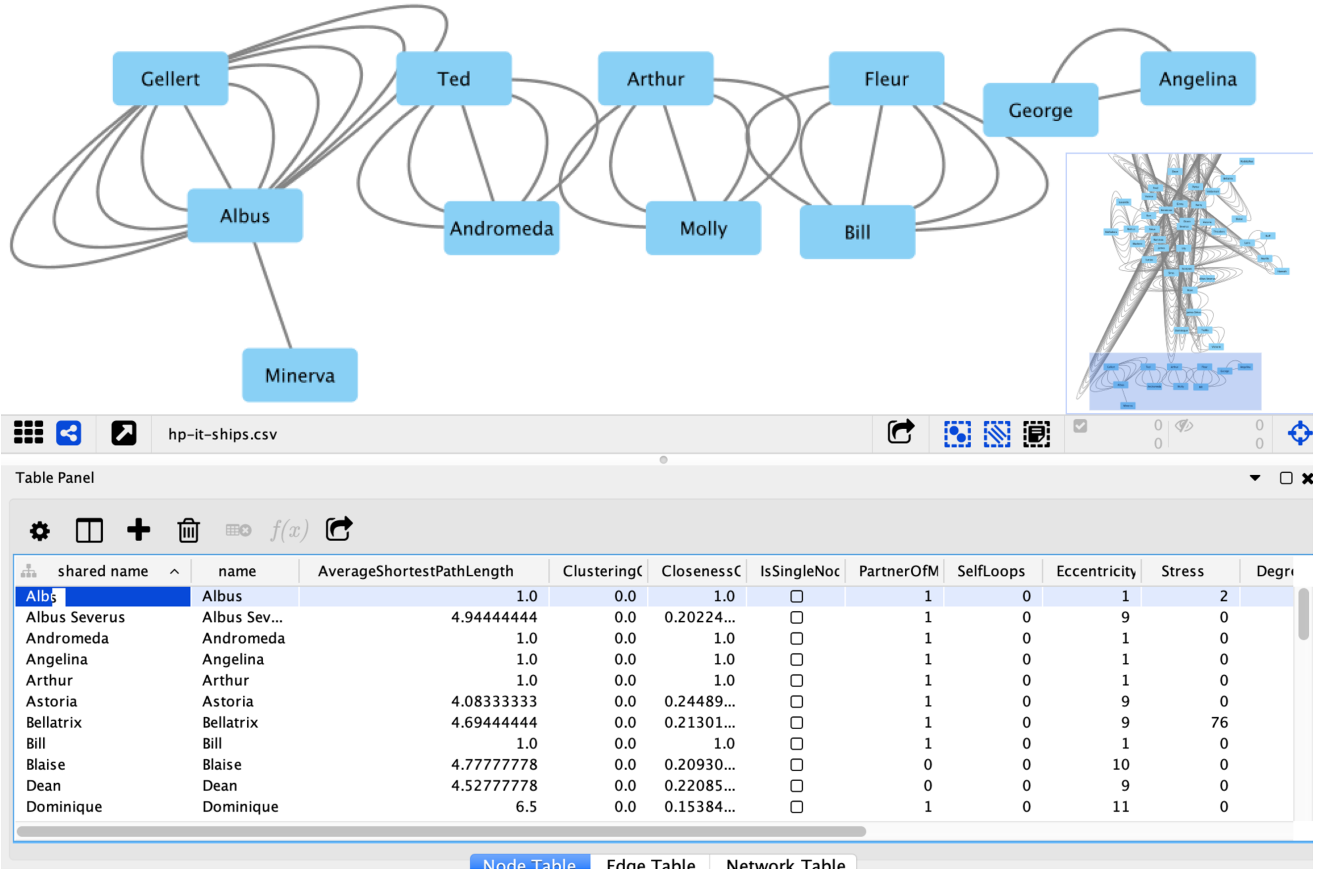


Overall network stats



Average shortest path

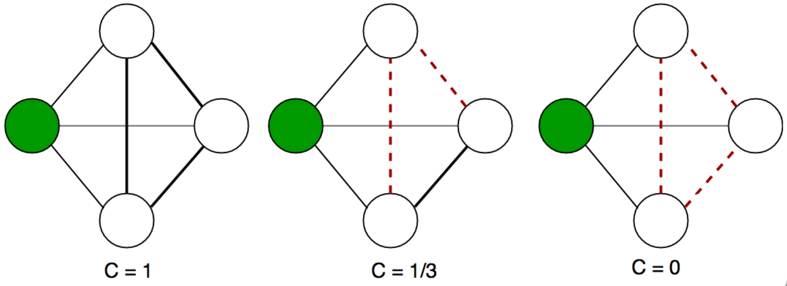
On average, how many edges it takes to get from that node to any other node it's connected to (however indirectly)

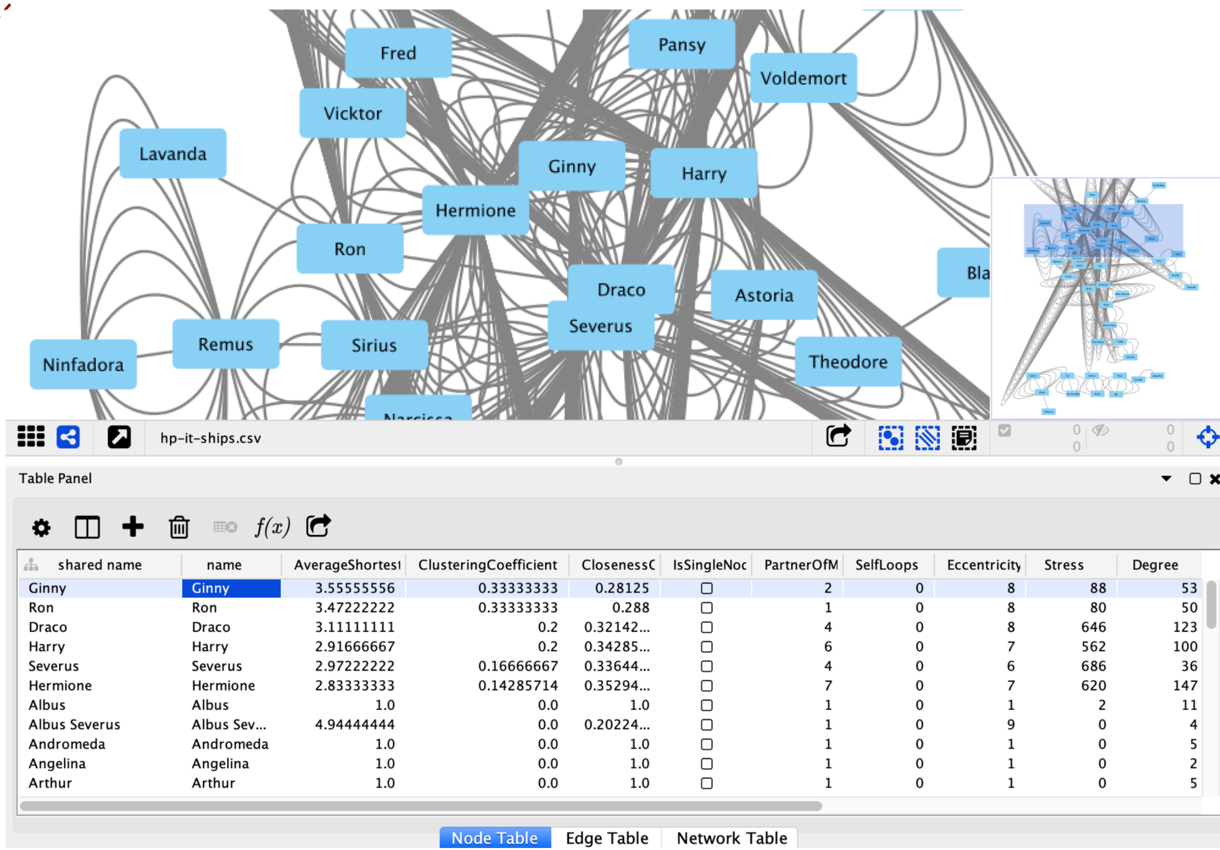


shared name	name	AverageShortestPathLength	Clustering(	Closeness(	IsSingleNoc	PartnerOfM	SelfLoops	Eccentricity	Stress	Degr
Albs	Albus	1.0	0.0	1.0	☐	1	0	1	2	
Albus Severus	Albus Sev...	4.94444444	0.0	0.20224...	☐	1	0	9	0	
Andromeda	Andromeda	1.0	0.0	1.0	☐	1	0	1	0	
Angelina	Angelina	1.0	0.0	1.0	☐	1	0	1	0	
Arthur	Arthur	1.0	0.0	1.0	☐	1	0	1	0	
Astoria	Astoria	4.08333333	0.0	0.24489...	☐	1	0	9	0	
Bellatrix	Bellatrix	4.69444444	0.0	0.21301...	☐	1	0	9	76	
Bill	Bill	1.0	0.0	1.0	☐	1	0	1	0	
Blaise	Blaise	4.77777778	0.0	0.20930...	☐	0	0	10	0	
Dean	Dean	4.52777778	0.0	0.22085...	☐	0	0	9	0	
Dominique	Dominique	6.5	0.0	0.15384...	☐	1	0	11	0	

Clustering coefficients

(Local) clustering coefficient: of the nodes connected to the current node, how likely are they to be connected to one another?



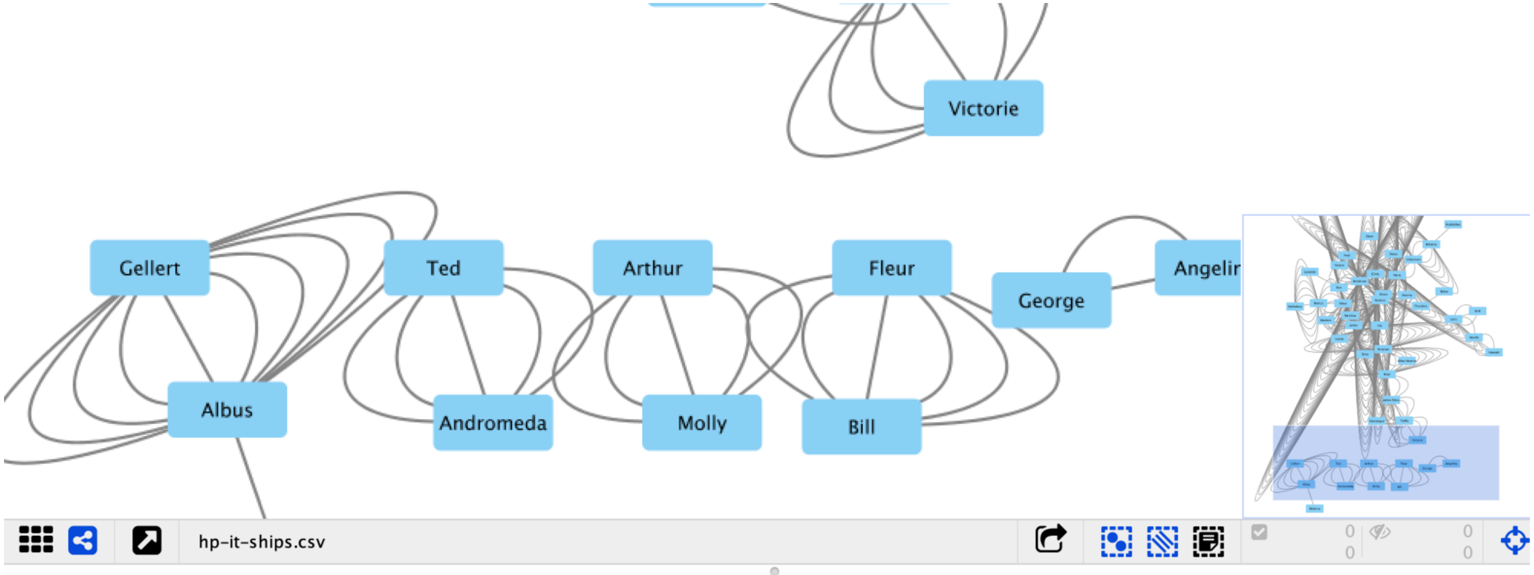


shared name	name	AverageShortest	ClusteringCoefficient	ClosenessC	IsSingleNoc	PartnerOfM	SelfLoops	Eccentricity	Stress	Degree
Ginny	Ginny	3.55555556	0.33333333	0.28125	☐	2	0	8	88	53
Ron	Ron	3.47222222	0.33333333	0.288	☐	1	0	8	80	50
Draco	Draco	3.11111111	0.2	0.32142...	☐	4	0	8	646	123
Harry	Harry	2.91666667	0.2	0.34285...	☐	6	0	7	562	100
Severus	Severus	2.97222222	0.16666667	0.33644...	☐	4	0	6	686	36
Hermione	Hermione	2.83333333	0.14285714	0.35294...	☐	7	0	7	620	147
Albus	Albus	1.0	0.0	1.0	☐	1	0	1	2	11
Albus Severus	Albus Sev...	4.94444444	0.0	0.20224...	☐	1	0	9	0	4
Andromeda	Andromeda	1.0	0.0	1.0	☐	1	0	1	0	5
Angelina	Angelina	1.0	0.0	1.0	☐	1	0	1	0	2
Arthur	Arthur	1.0	0.0	1.0	☐	1	0	1	0	5

Closeness centrality

Add up all the shortest paths, divide by number of connections.

(As a result, it's highest for isolates.)



shared name	name	AverageShortest	ClusteringCoeff	ClosenessCentrality	IsSingleNoc	PartnerOfM	SelfLoops	Eccentricity	Stress	Degree
Severus	Severus	2.97222222	0.16666667	0.3364486	☐	4	0	6	686	3
Harry	Harry	2.91666667	0.2	0.34285714	☐	6	0	7	562	1(
Hermione	Hermione	2.83333333	0.14285714	0.35294118	☐	7	0	7	620	14
Gellert	Gellert	1.5	0.0	0.66666667	☐	1	0	2	0	1
Minerva	Minerva	1.5	0.0	0.66666667	☐	0	0	2	0	
Albus	Albus	1.0	0.0	1.0	☐	1	0	1	2	1
Andromeda	Andromeda	1.0	0.0	1.0	☐	1	0	1	0	
Angelina	Angelina	1.0	0.0	1.0	☐	1	0	1	0	
Arthur	Arthur	1.0	0.0	1.0	☐	1	0	1	0	
Bill	Bill	1.0	0.0	1.0	☐	1	0	1	0	
Fleur	Fleur	1.0	0.0	1.0	☐	1	0	1	0	

Is single node

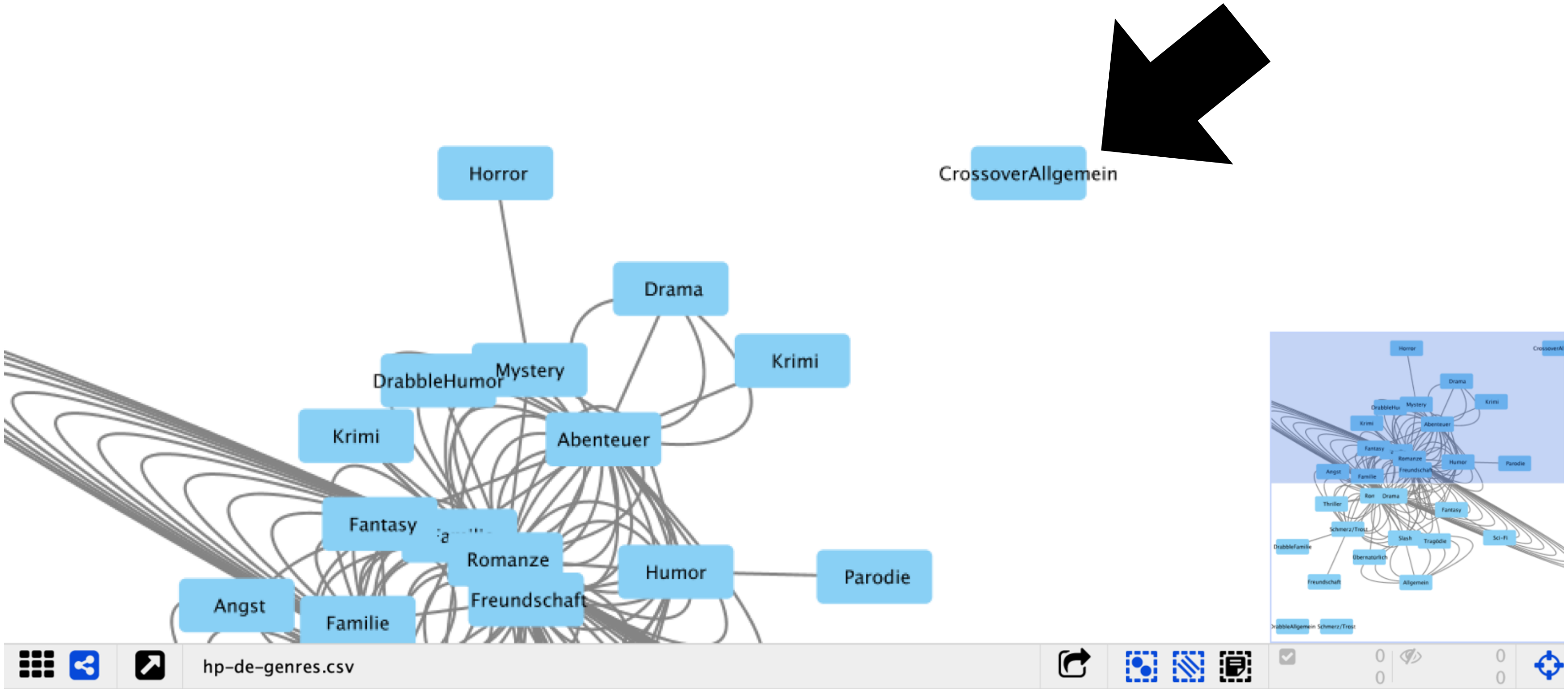
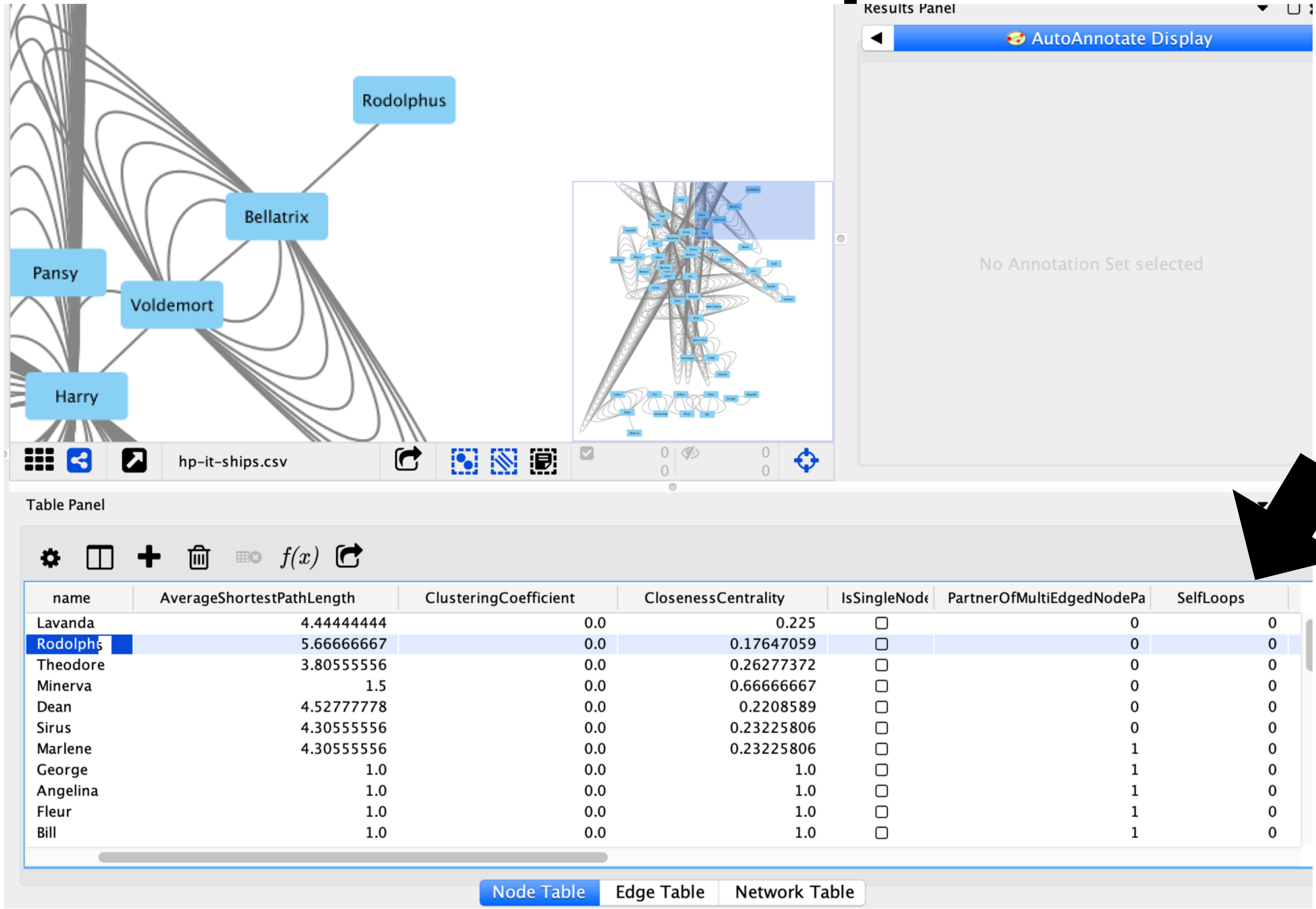


Table Panel

shared name	name	AverageSh(	Clustering(	Closeness(	IsSingleNoc	PartnerOfM	SelfLoops	Eccentricity	Stress	Degree	Betweennes
Abenteuer	Abenteuer	2.34615...	0.0	0.42622...	☐	4	0	4	440	25	0.19837...
Drama	Drama	3.30769...	0.0	0.30232...	☐	1	0	5	0	4	0.0
Fantasy	Fantasy	2.61538...	0.0	0.38235...	☐	1	0	4	66	6	0.01736...
Fantasy	Fantasy	2.65384...	0.0	0.37681...	☐	1	0	5	168	4	0.08899...
CrossoverAllgemein	Crossover...	0.0	0.0	0.0	☑	0	0	0	0	0	0.0
Humor	Humor	2.57692...	0.0	0.38805...	☐	2	0	4	158	12	0.08861...
Familie	Familie	2.38461...	0.0	0.41935...	☐	3	0	4	206	16	0.05670...
Schmerz/Trost	Schmerz/...	2.46153...	0.0	0.40625	☐	2	0	4	342	15	0.17063...
Angst	Angst	2.69230...	0.0	0.37142...	☐	0	0	4	24	2	0.00427...
Familie	Familie	2.65384...	0.0	0.37681...	☐	1	0	5	60	6	0.01995...
Sci-Fi	Sci-Fi	3.61538...	0.0	0.27659...	☐	0	0	6	0	1	0.0

Node Table | Edge Table | Network Table

Partner of multi-edged node pairs



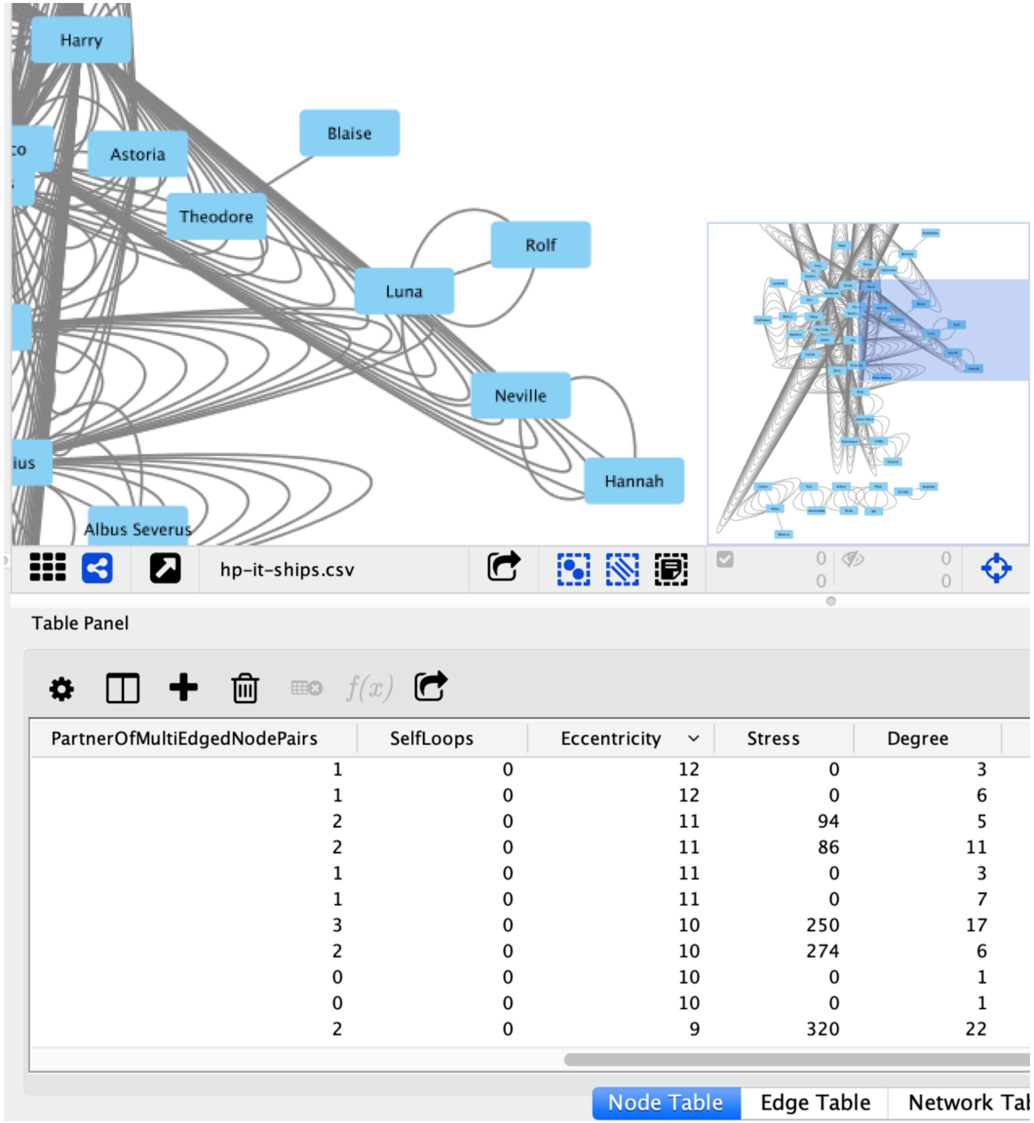
name	AverageShortestPathLength	ClusteringCoefficient	ClosenessCentrality	IsSingleNod	PartnerOfMultiEdgedNodePa	SelfLoops
Lavanda	4.44444444	0.0	0.225	☐	0	0
Rodolphs	5.66666667	0.0	0.17647059	☐	0	0
Theodore	3.80555556	0.0	0.26277372	☐	0	0
Minerva	1.5	0.0	0.66666667	☐	0	0
Dean	4.52777778	0.0	0.2208589	☐	0	0
Sirus	4.30555556	0.0	0.23225806	☐	0	0
Marlene	4.30555556	0.0	0.23225806	☐	1	0
George	1.0	0.0	1.0	☐	1	0
Angelina	1.0	0.0	1.0	☐	1	0
Fleur	1.0	0.0	1.0	☐	1	0
Bill	1.0	0.0	1.0	☐	1	0

Self-loops =

self-explanatory

Eccentricity

The opposite of the shortest path (but non-infinite).

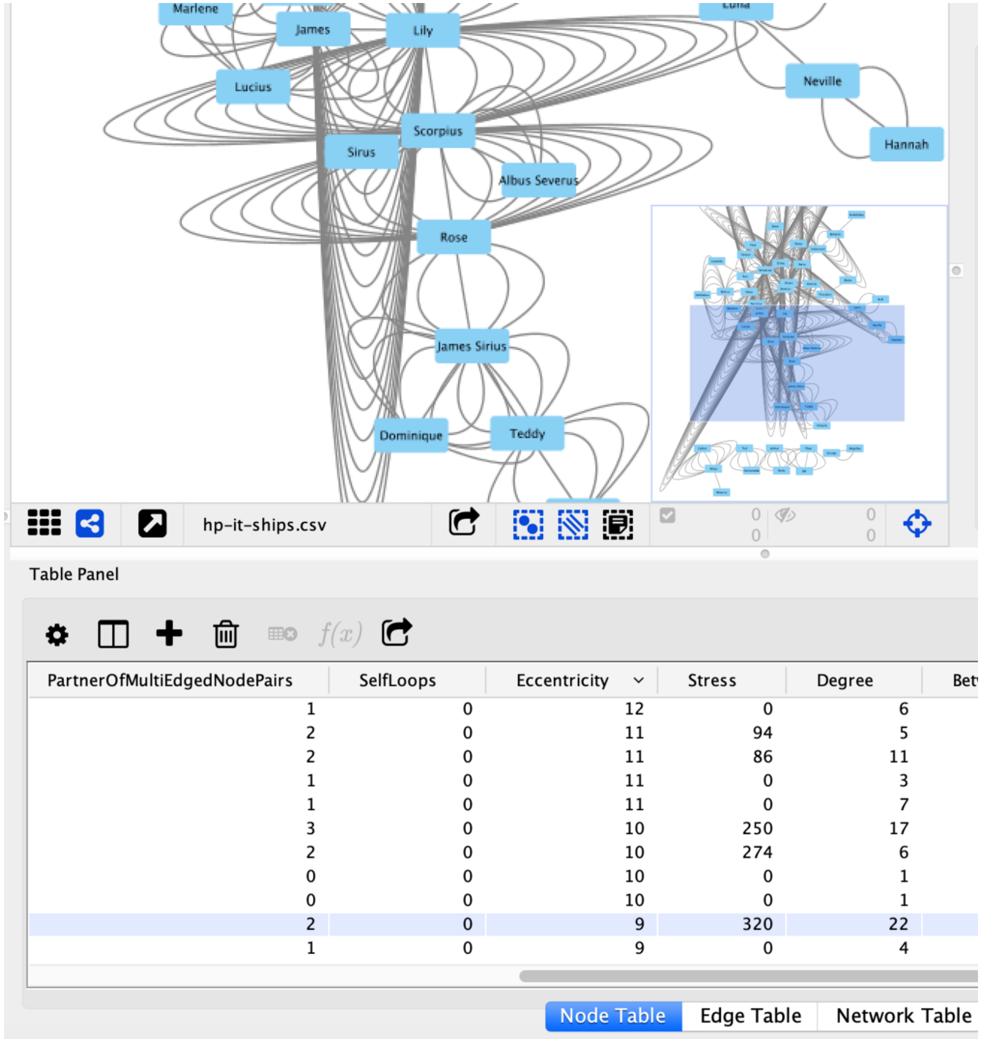


PartnerOfMultiEdgedNodePairs	SelfLoops	Eccentricity	Stress	Degree
1	0	12	0	3
1	0	12	0	6
2	0	11	94	5
2	0	11	86	11
1	0	11	0	3
1	0	11	0	7
3	0	10	250	17
2	0	10	274	6
0	0	10	0	1
0	0	10	0	1
2	0	9	320	22

Nodes on the periphery, but still connected, get high eccentricity (Hannah).

Stress

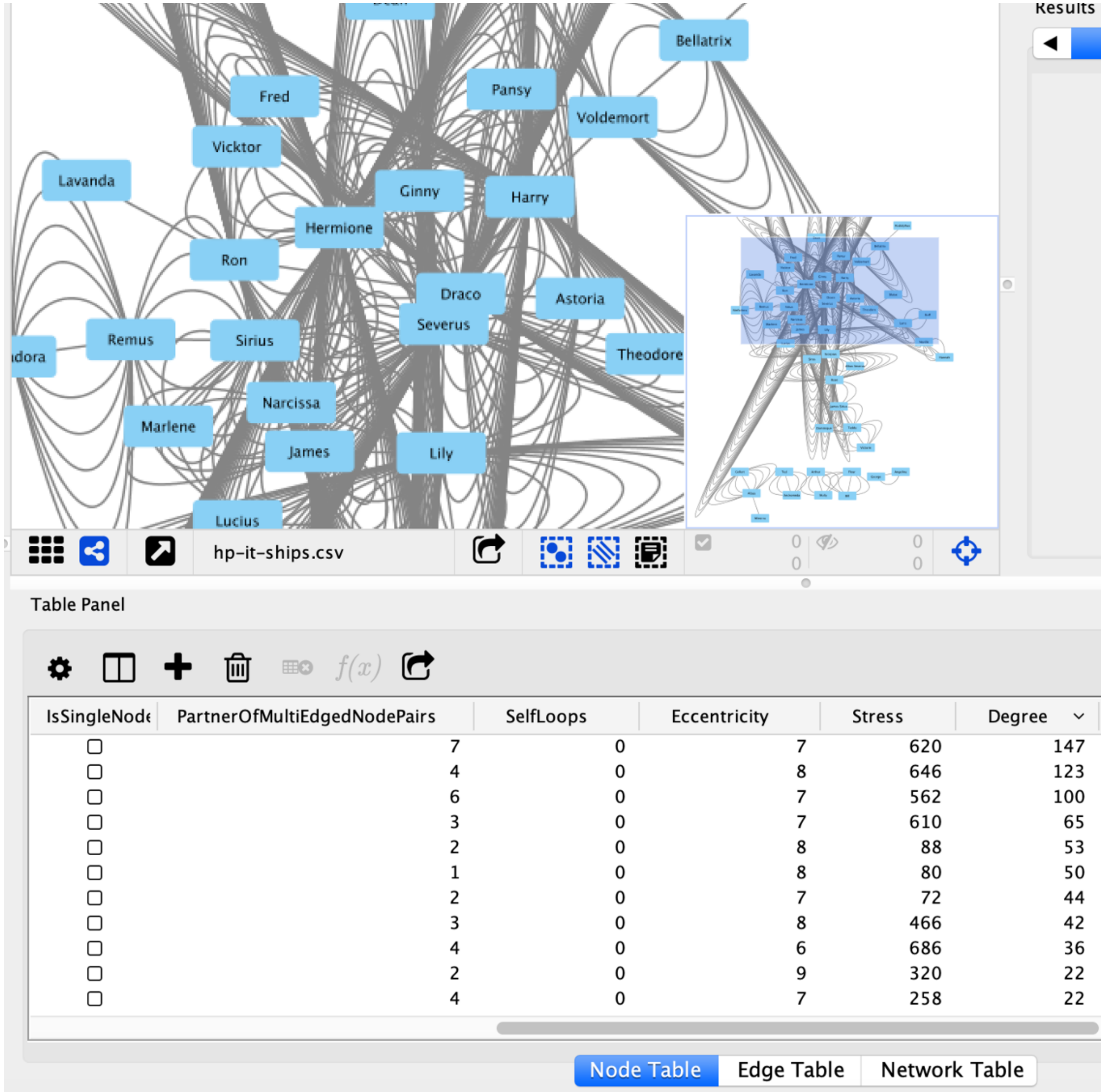
Number of shortest paths passing through the node



Nodes that are central, or that are the only connection to more than one node on the edge of the network get high stress (e.g. Rose).

Degree

Number of nodes with edges to the node.



IsSingleNode	PartnerOfMultiEdgedNodePairs	SelfLoops	Eccentricity	Stress	Degree
☐	7	0	7	620	147
☐	4	0	8	646	123
☐	6	0	7	562	100
☐	3	0	7	610	65
☐	2	0	8	88	53
☐	1	0	8	80	50
☐	2	0	7	72	44
☐	3	0	8	466	42
☐	4	0	6	686	36
☐	2	0	9	320	22
☐	4	0	7	258	22

Hermione

Draco

Harry

Betweenness centrality

of shortest paths that go through the node (stress) divided by total number of shortest paths.

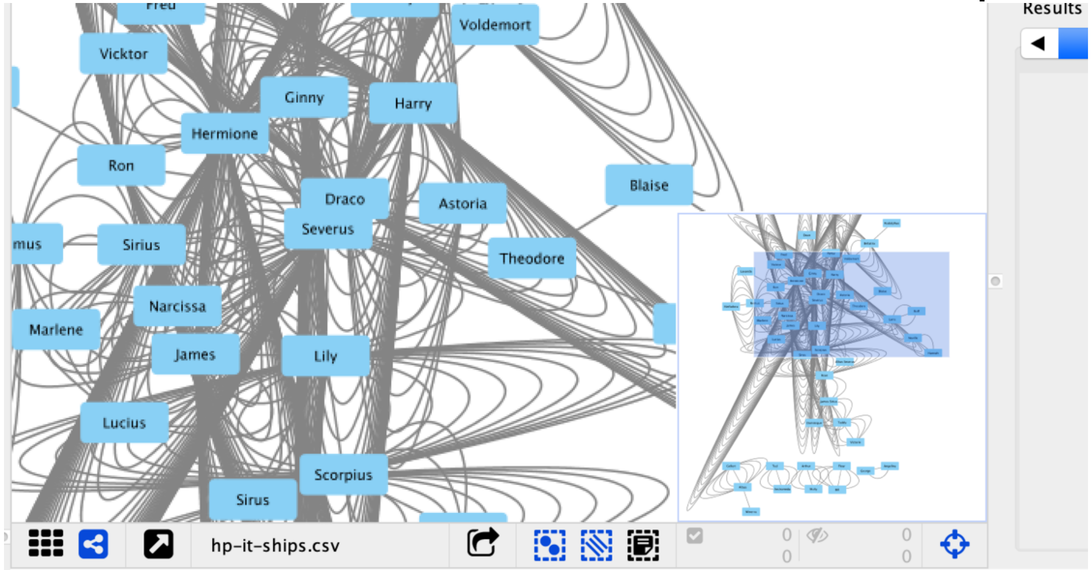


Table Panel

Eccentricity	Stress	Degree	BetweennessCentrality	NeighborhoodConnectivity
1	2	11	1.0	1.0
6	686	36	0.39867725	4.75
7	620	147	0.38068783	3.57142857
7	610	65	0.37645503	2.5
8	646	123	0.35820106	3.83333333
7	562	100	0.33544974	3.83333333
8	466	42	0.29365079	2.33333333
9	438	3	0.25238095	3.33333333
9	320	22	0.2031746	3.0
7	258	22	0.18703704	3.0
10	250	17	0.16031746	1.66666667

Node Table | Edge Table | Network Table

Indicates which nodes serve as a bridge in the network.

(Albus)

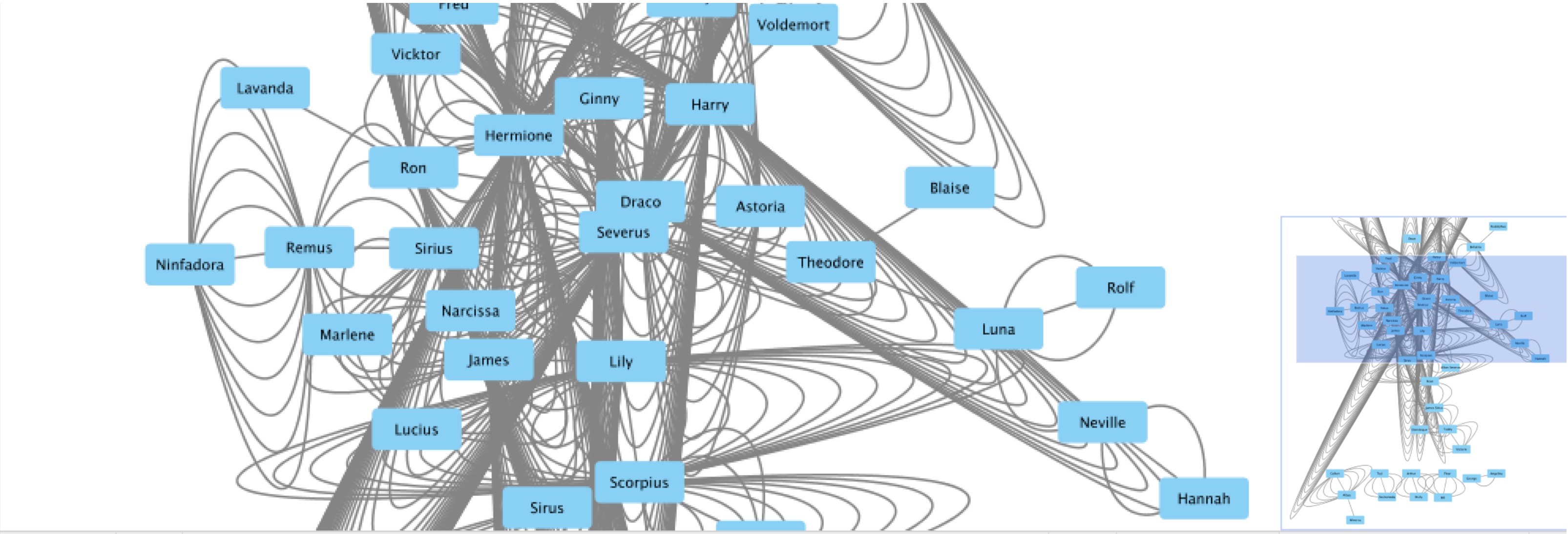
Severus

Hermione

Lily

Neighborhood connectivity

How many nodes are the current node's connections themselves connected to?



hp-it-ships.csv

Table Panel

shared name	name	AverageShortestPathLength	ClusteringCoefficient	NeighborhoodConnectivity	ClosenessCentrality	IsSi
Albus	Albus	1.0	0.0	1.0	1.0	
Severus	Severus	2.97222222	0.16666667	4.75	0.3364486	
Hermione	Hermione	2.83333333	0.14285714	3.57142857	0.35294118	
Ly	Lily	3.33333333	0.0	2.5	0.3	
Draco	Draco	3.11111111	0.2	3.83333333	0.32142857	
Harry	Harry	2.91666667	0.2	3.83333333	0.34285714	
Scorpius	Scorpius	3.97222222	0.0	2.33333333	0.25174825	
Theodore	Theodore	3.80555556	0.0	3.33333333	0.26277372	
Rose	Rose	4.72222222	0.0	3.0	0.21176471	
Sirius	Sirius	3.33333333	0.0	3.0	0.3	
James Sirius	James Sirius	5.52777778	0.0	1.66666667	0.18090452	

Lily is connected to Severus and Scorpius, who aren't connected to each other (clustering is 0), but **are** connected to other nodes.

Radiality

Diameter is the maximum eccentricity of any node in the graph (greatest distance between any pair of nodes).

Radiality is closeness centrality (average shortest path length) subtracted from diameter, plus one.

Radiality = diameter - closeness centrality + 1