Centre for
Data Analytics

Insight

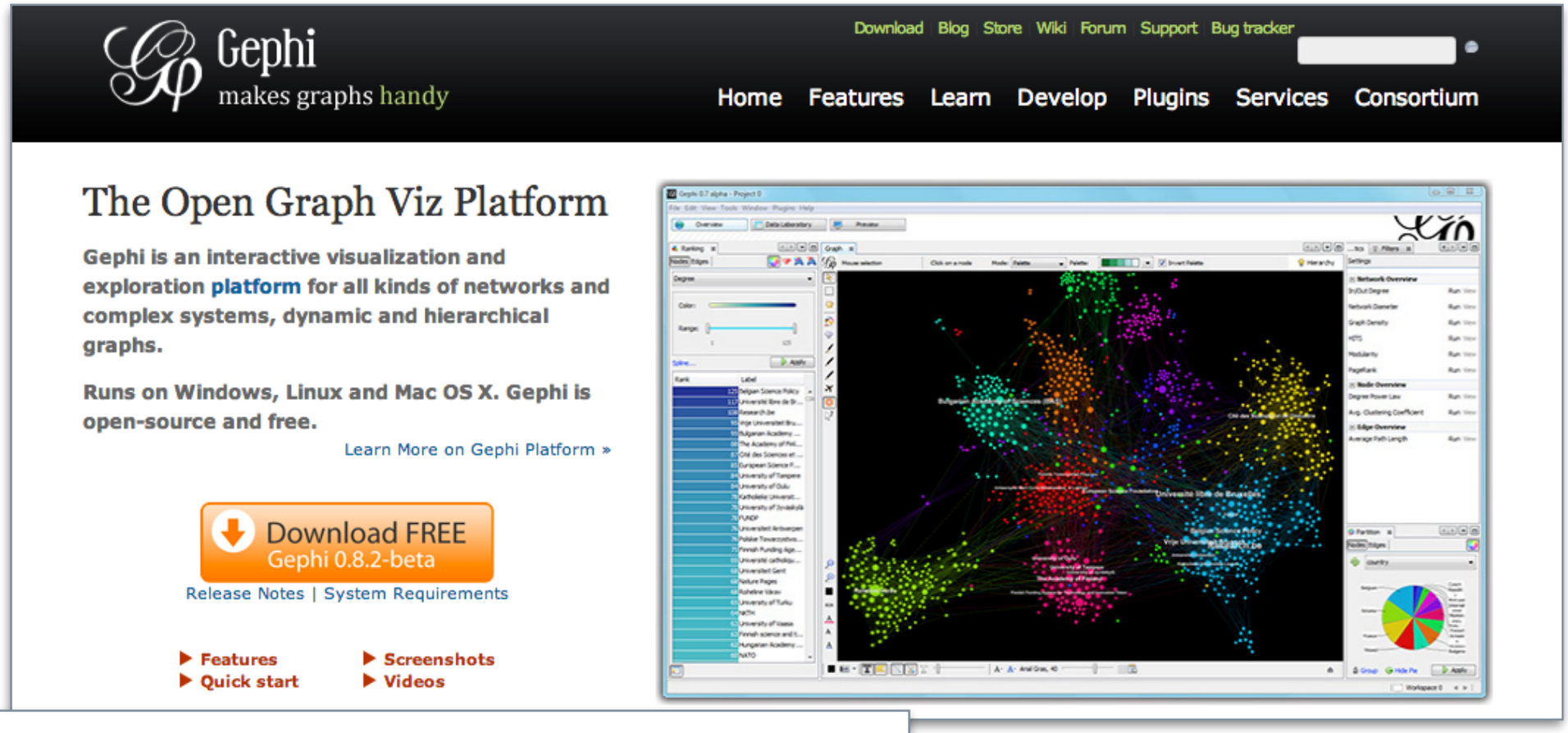
Practical Social Network Analysis With Gephi

Dr. Derek Greene
Insight @ UCD

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD DUBLIN SFI Science Foundation Ireland

Gephi - Introduction

- Gephi: Open source interactive network exploration and visualisation tool for Windows, Linux and Mac OS X.



Gephi: an **open source software** for **exploring** and **manipulating networks**.
M Bastian, S Heymann, M Jacomy - ICWSM, 2009 - aaai.org
Abstract **Gephi** is an **open source software** for graph and **network** analysis. It uses a 3D render engine to display large **networks** in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets ...
Cited by 708 Related articles All 15 versions Import into BibTeX Saved More

http://gephi.org

Input Data - Graph Formats

- First task - loading your network data into Gephi.
- Gephi supports loading and saving graphs in a number of formats.
- Simplest approach is to use comma-separated (CSV) data exported from tools such as Excel or R.

Simple binary graph
No weights or attributes

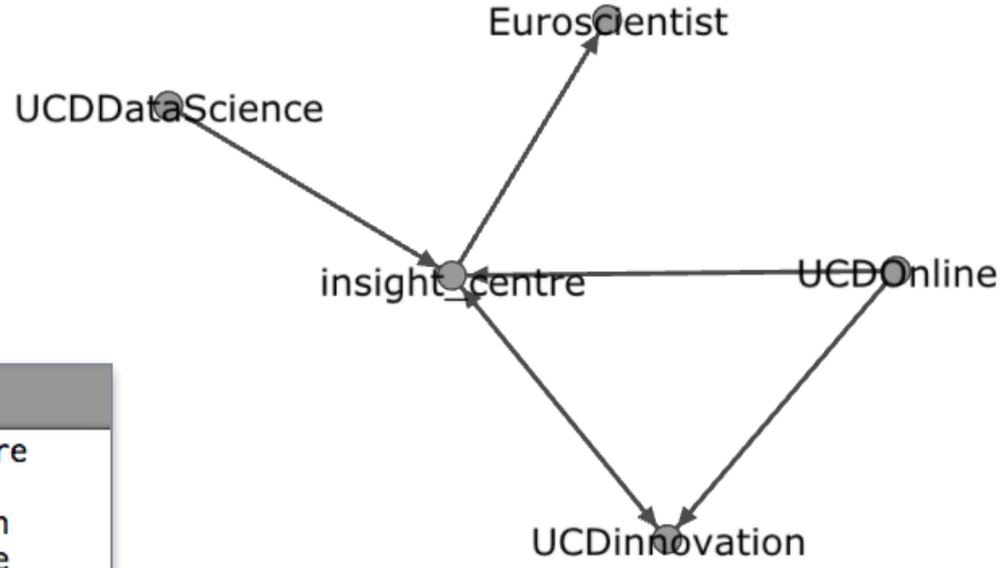
Edge From

Edge To

	A	B	C
1	UCDDataScience	insight_centre	
2	UCDOnline	insight_centre	
3	insight_centre	UCDinnovation	
4	UCDinnovation	insight_centre	
5	insight_centre	Euroscientist	
6	UCDOnline	UCDinnovation	
7			

eg_ucd.csv

```
UCDDataScience,insight_centre
UCDOnline,insight_centre
insight_centre,UCDinnovation
UCDinnovation,insight_centre
insight_centre,Euroscientist
UCDOnline,UCDinnovation
```



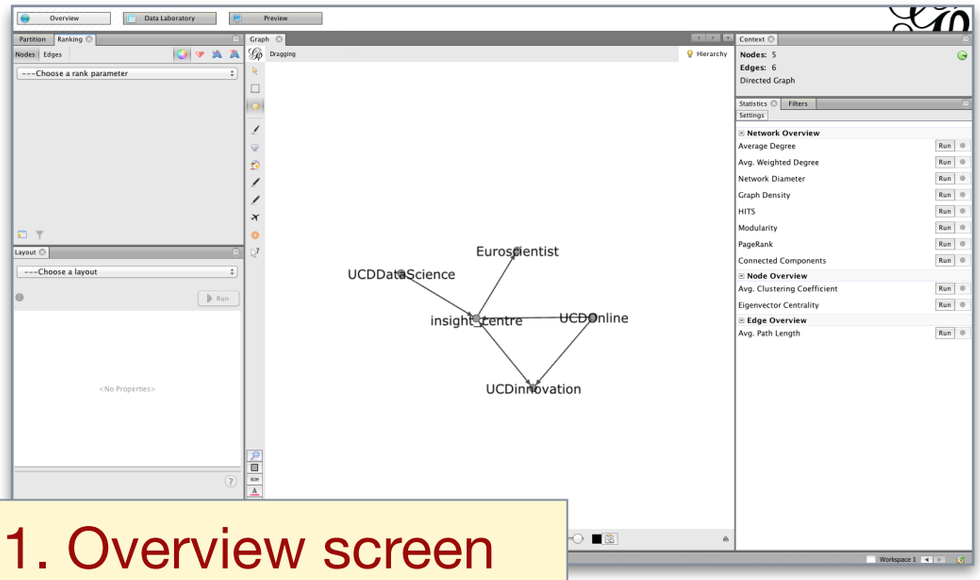
On loading, Gephi will ask whether graph is directed or undirected.

Input Data - Graph Formats

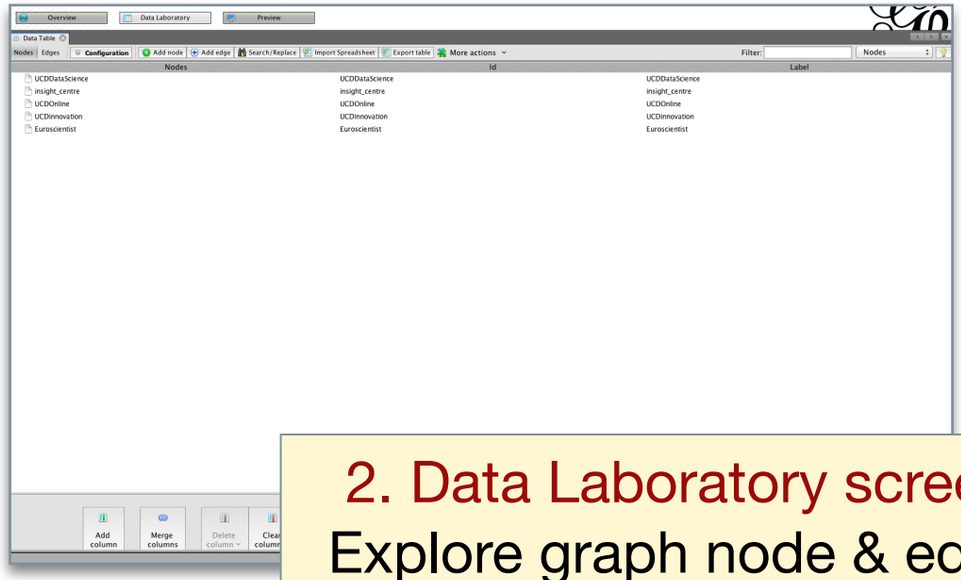
- Gephi supports more comprehensive file formats which can store node and edge attributes, together with layout and presentation information (e.g. position, size, colour etc).
- Native format is a .gephi file, which can contain multiple “workspaces” (i.e. several different graphs).
- Gephi also supports several open formats which can be used to exchange data with other tools. Common examples:
  - GEXF: XML Graph Exchange Format for complex networks. http://gexf.net
  - GraphML: XML Graph Markup language http://graphml.graphdrawing.org http://cs.brown.edu/~rt/gdhandbook/chapters/graphml.pdf
  - GML: Plain text Graph Modeling Language for describing graphs http://en.wikipedia.org/wiki/Graph_Modelling_Language

Gephi Screens

- Three key screens in Gephi, accessible from top buttons…



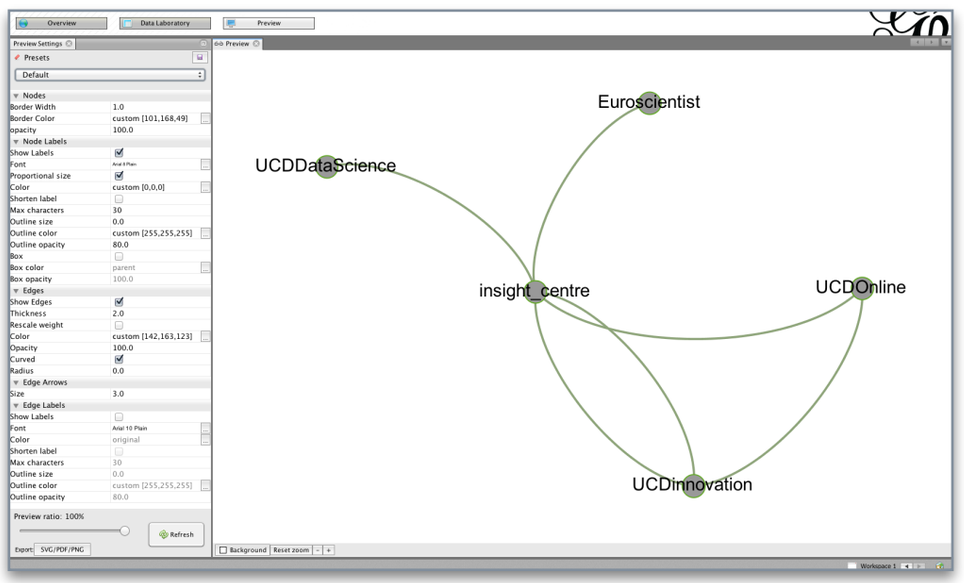
1. Overview screen
Interact with your graph in a visual way.



2. Data Laboratory screen
Explore graph node & edge attribute data in tabular form.

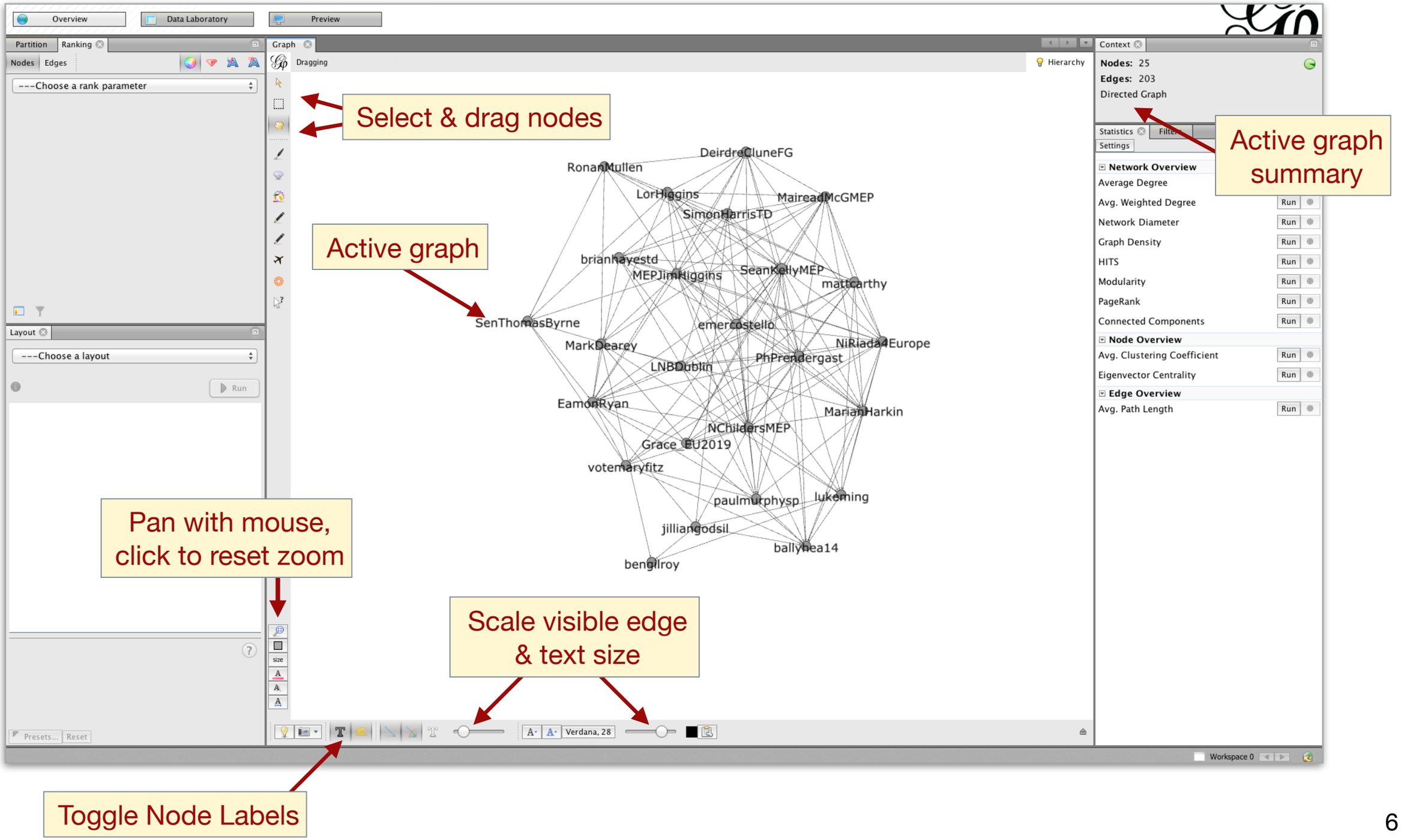
Each screen contains multiple tabs with specific functionality.

3. Preview screen
Fine-tune and render your graph for export.



Graph Overview Screen

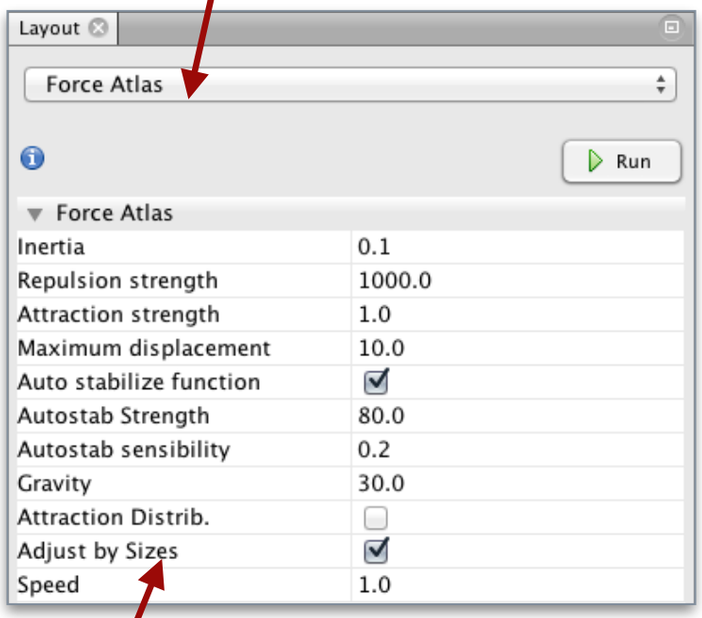
- Gephi defaults to the Overview screen - the “draft” interactive view.



Graph Layouts

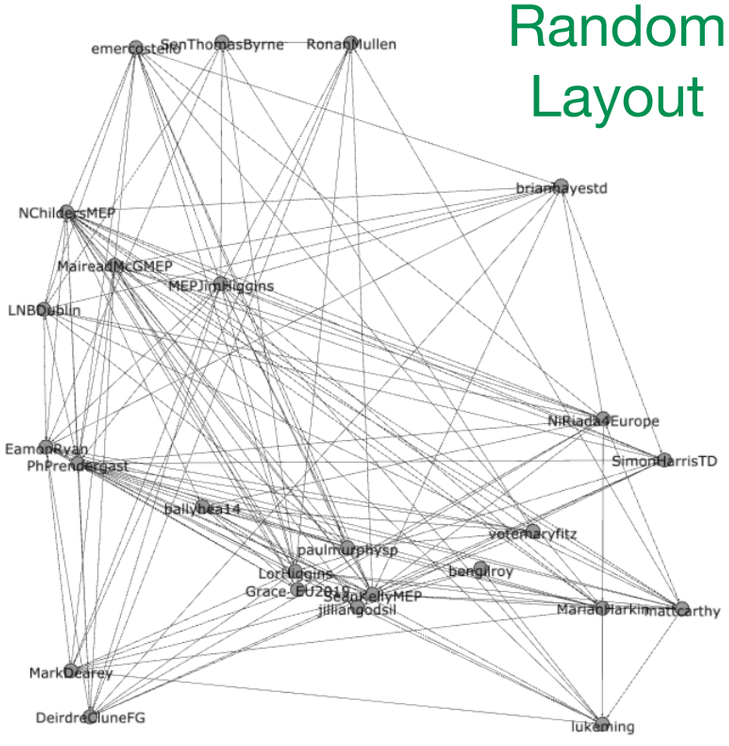
- Common first step is to apply a layout algorithm to re-position nodes in the graph so as to improve its readability and aesthetics.

Choose algorithm in "Layout" tab

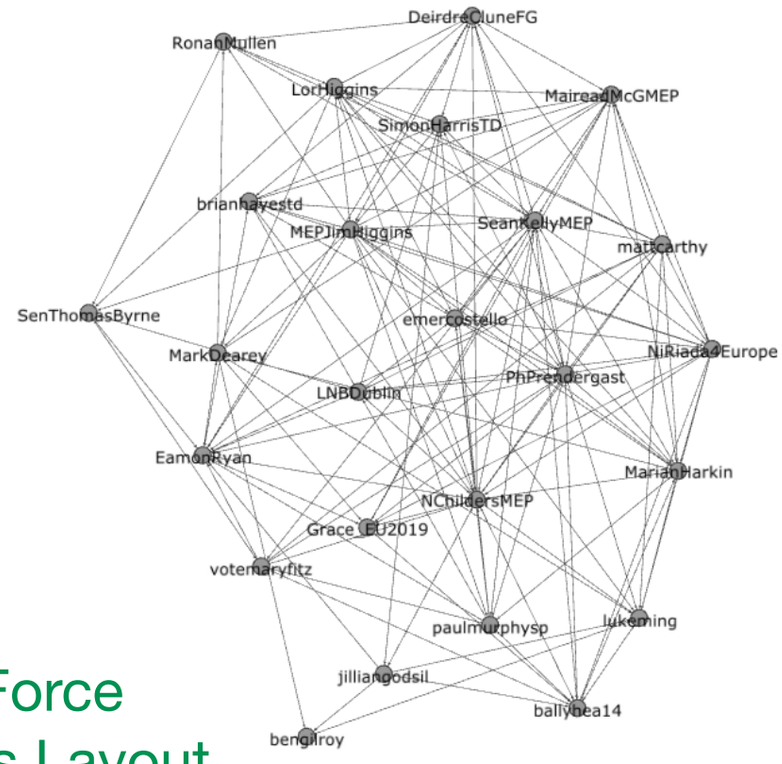


Properties to tweak algorithm

➡ Frequent choice is a "force atlas" algorithm, which tries to ensure as few edges cross as possible.



Random Layout

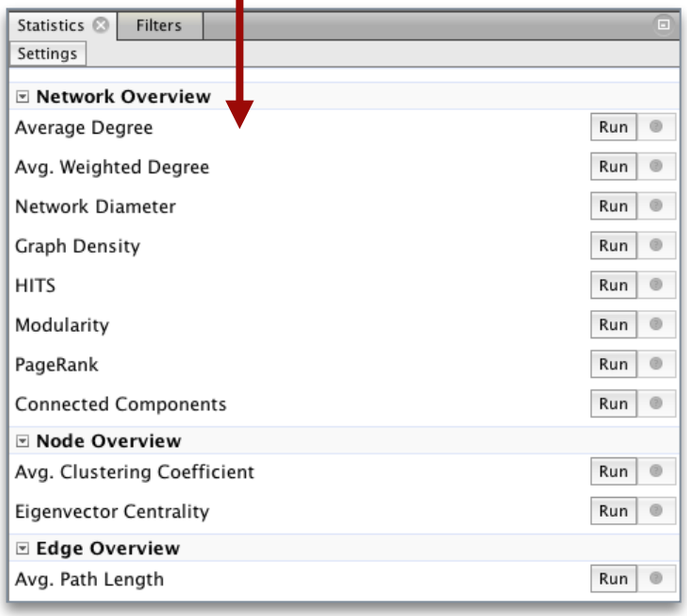


Force Atlas Layout

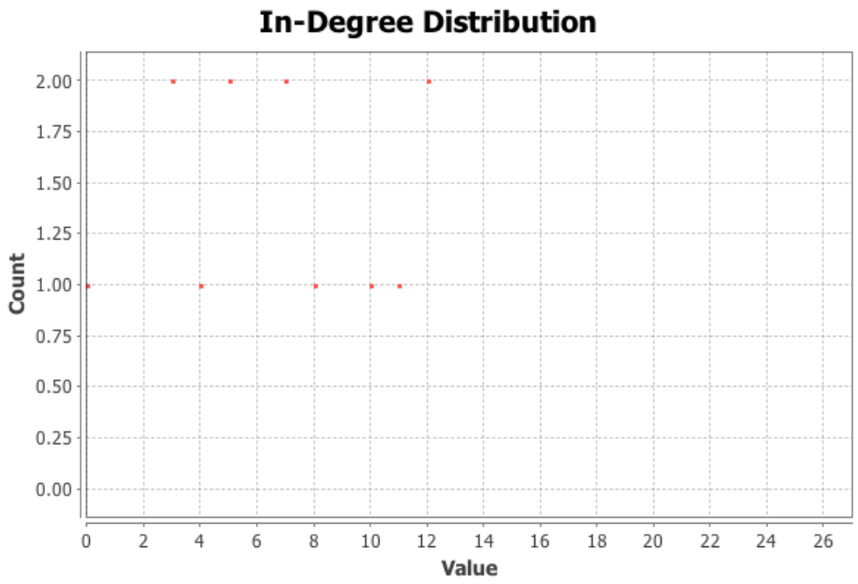
Graph Statistics

- Gephi provides a range of metrics for calculating statistics that characterise a graph and its nodes.

Run a metric from “Statistics” tab



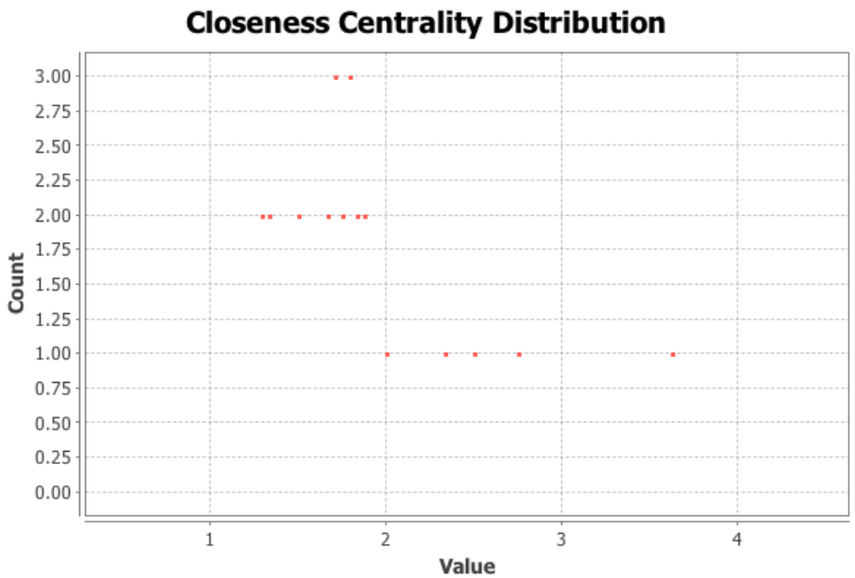
- Run “Average Degree”: Produces distribution plot of node in-degrees (user “popularity”)



- Run “Avg Path Length”: Produces a report of centrality scores for all nodes.

Results:

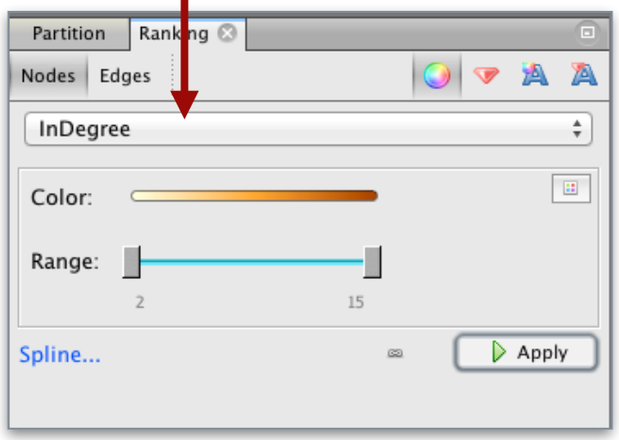
Diameter: 5
Radius: 2
Average Path length: 1.8483333333333334
Number of shortest paths: 600

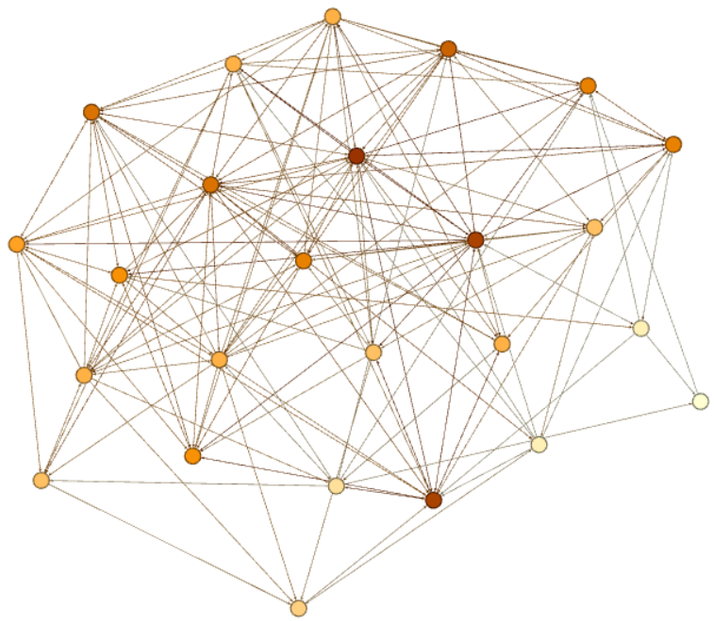


Ranking Nodes

- Nodes can be re-sized and/or coloured based on their statistics, using the “Ranking” tab.

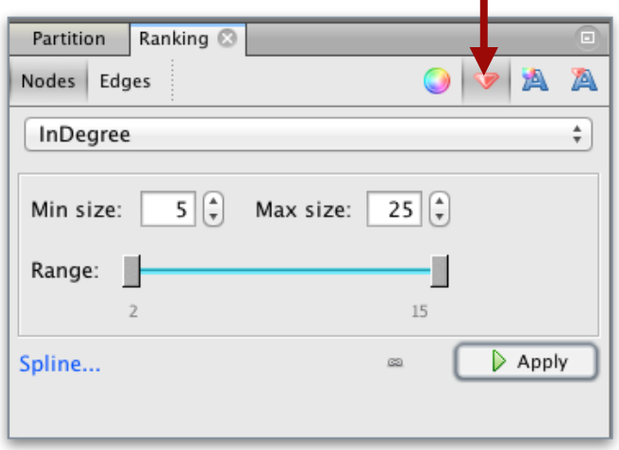
Choose a statistic to rank nodes

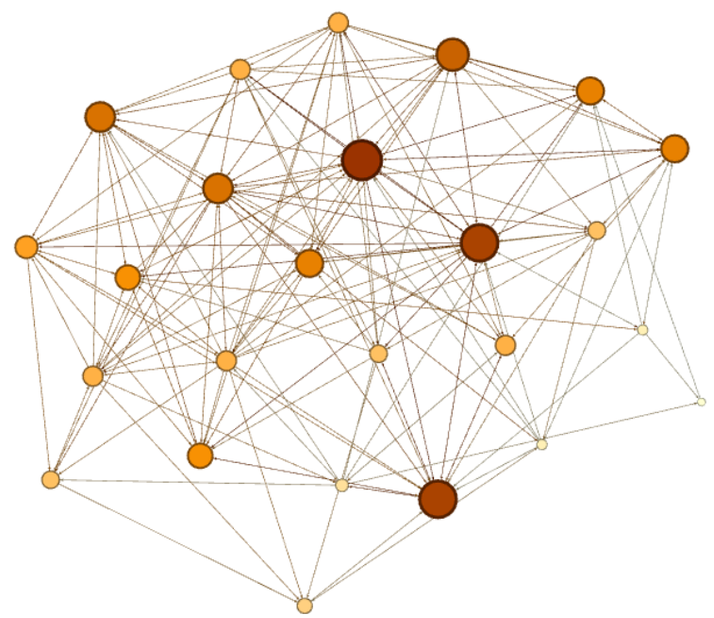




More saturated (red) colour indicates higher in-degree

Switch to resize mode



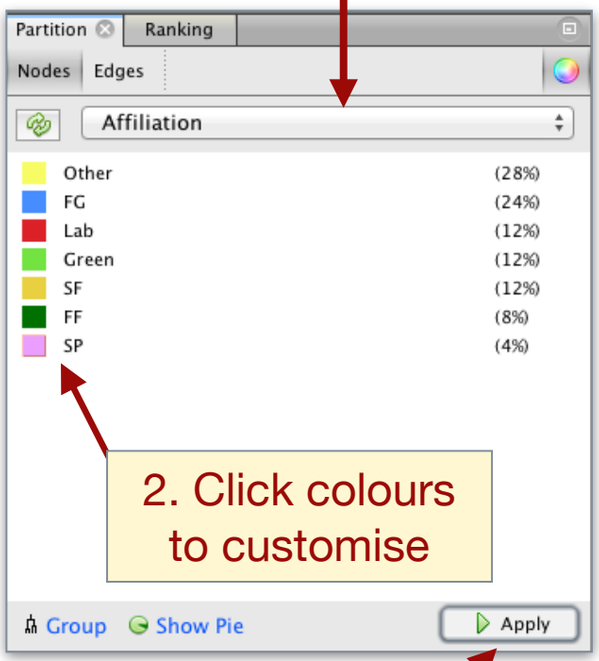


Large node size indicates higher in-degree

Graph Clustering

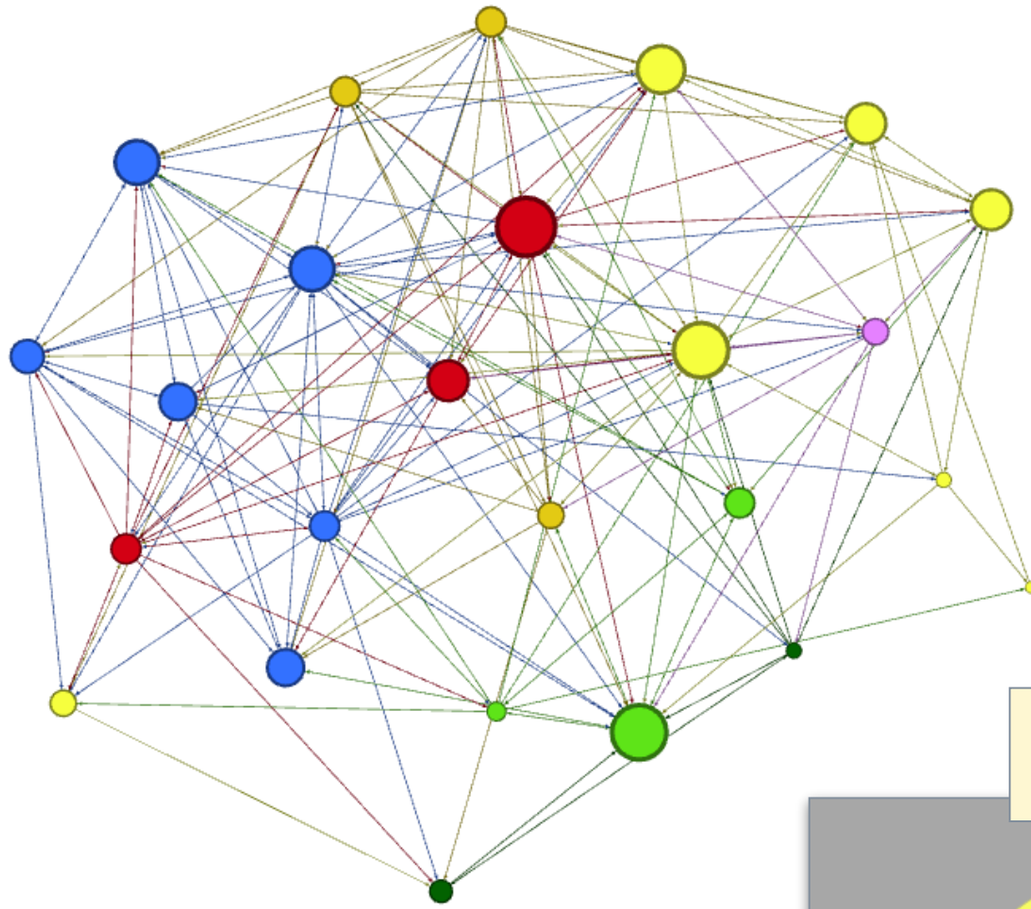
- Nodes can be “clustered” and coloured based on their attribute values (e.g location, affiliation etc) in the “Partition” tab.

1. Hit Refresh icon, then choose attribute to use to cluster nodes



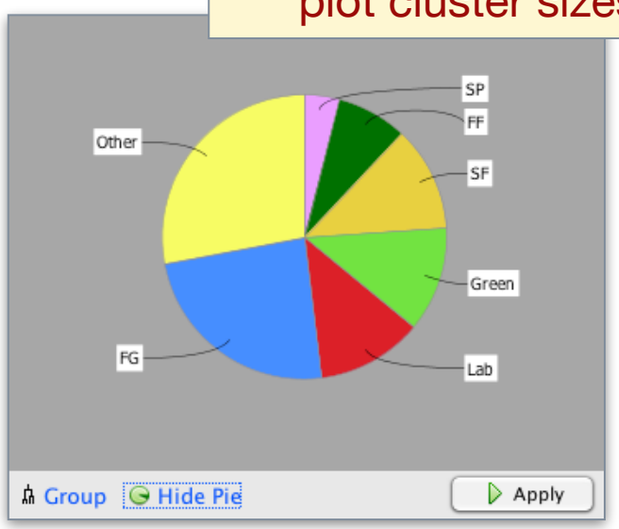
2. Click colours to customise

3. Click “Apply” to re-colour nodes



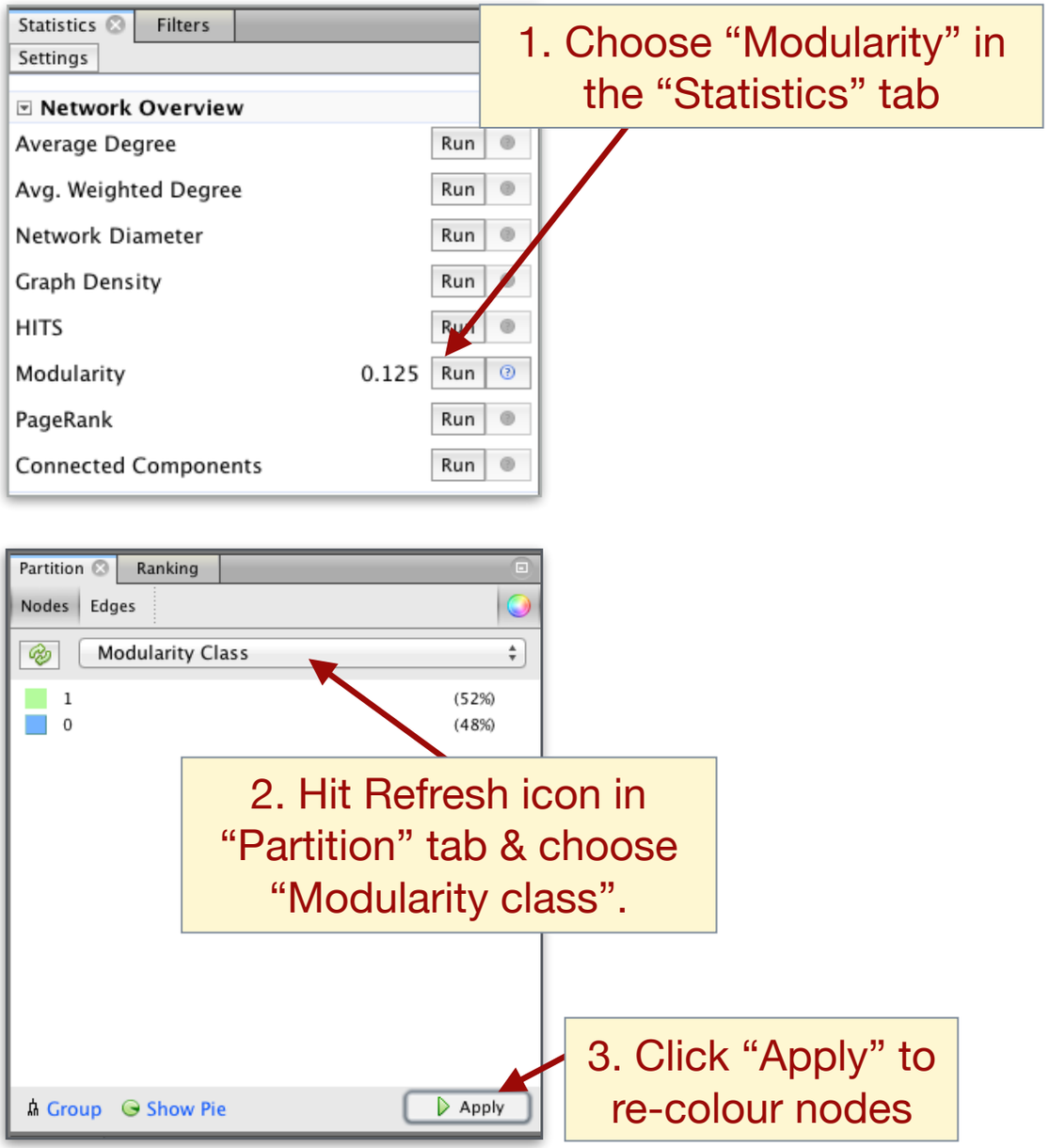
Clustering of politicians based on “affiliation” attribute

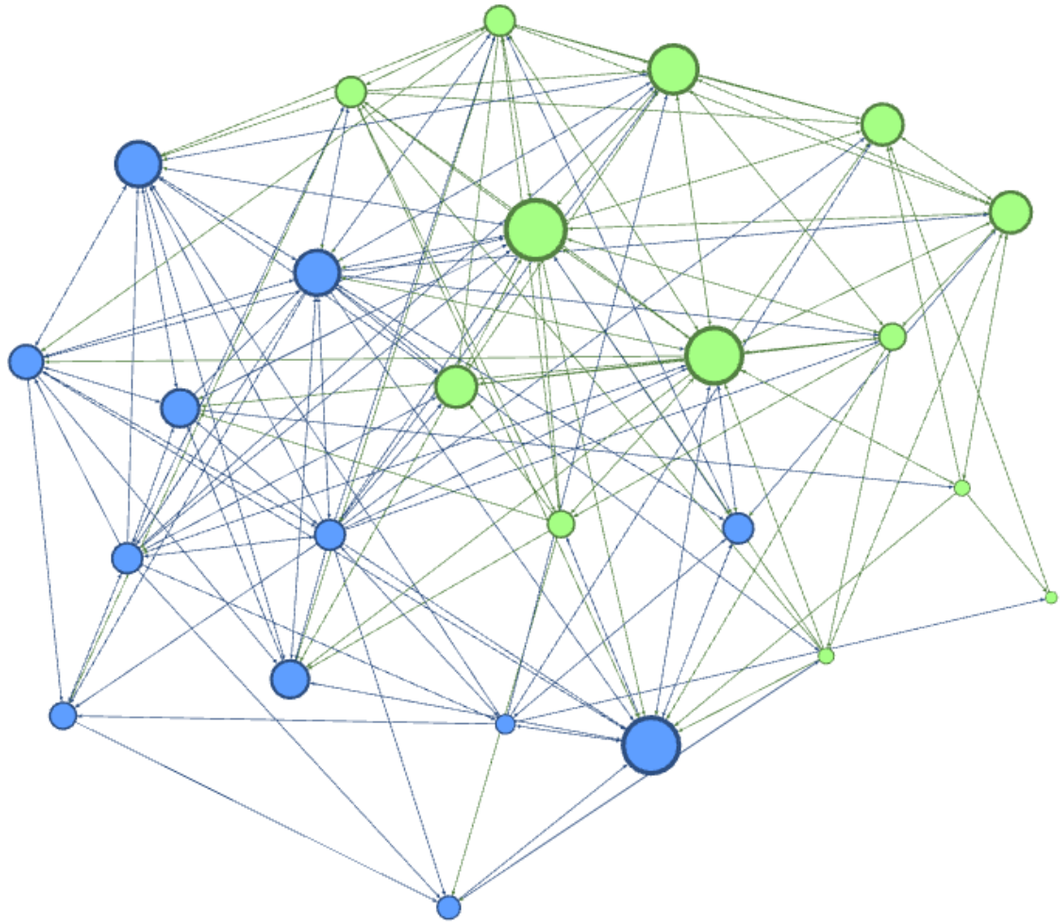
Click “Show Pie” to plot cluster sizes



Graph Clustering

- In cases where a grouping of nodes is not known apriori, we can apply cluster analysis methods to automatically detect groups in the data (e.g. communities of similar Twitter users)

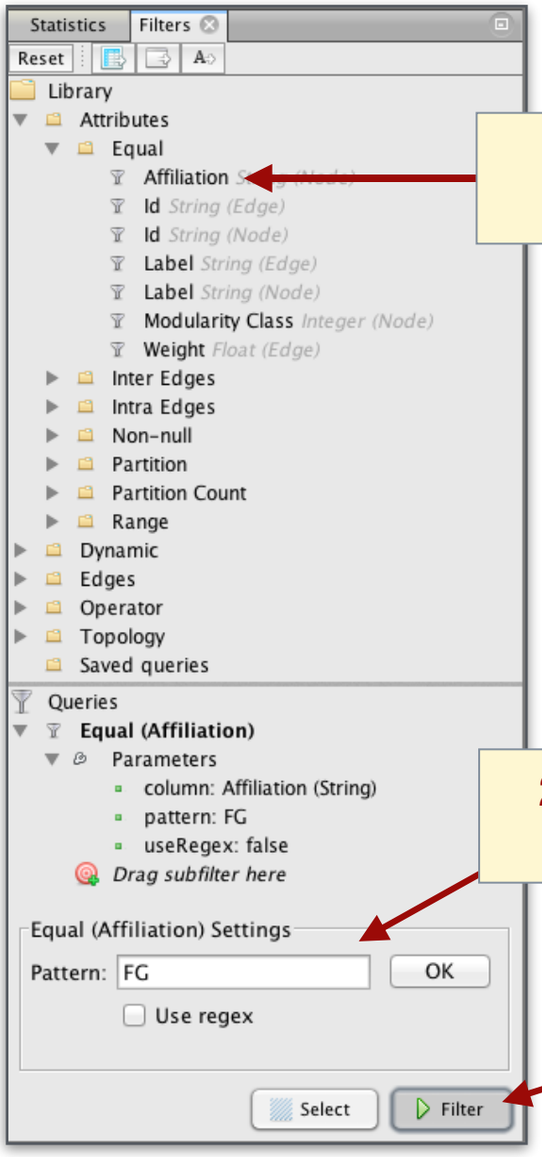




Modularity-based clustering produces 2 communities

Filtering Nodes

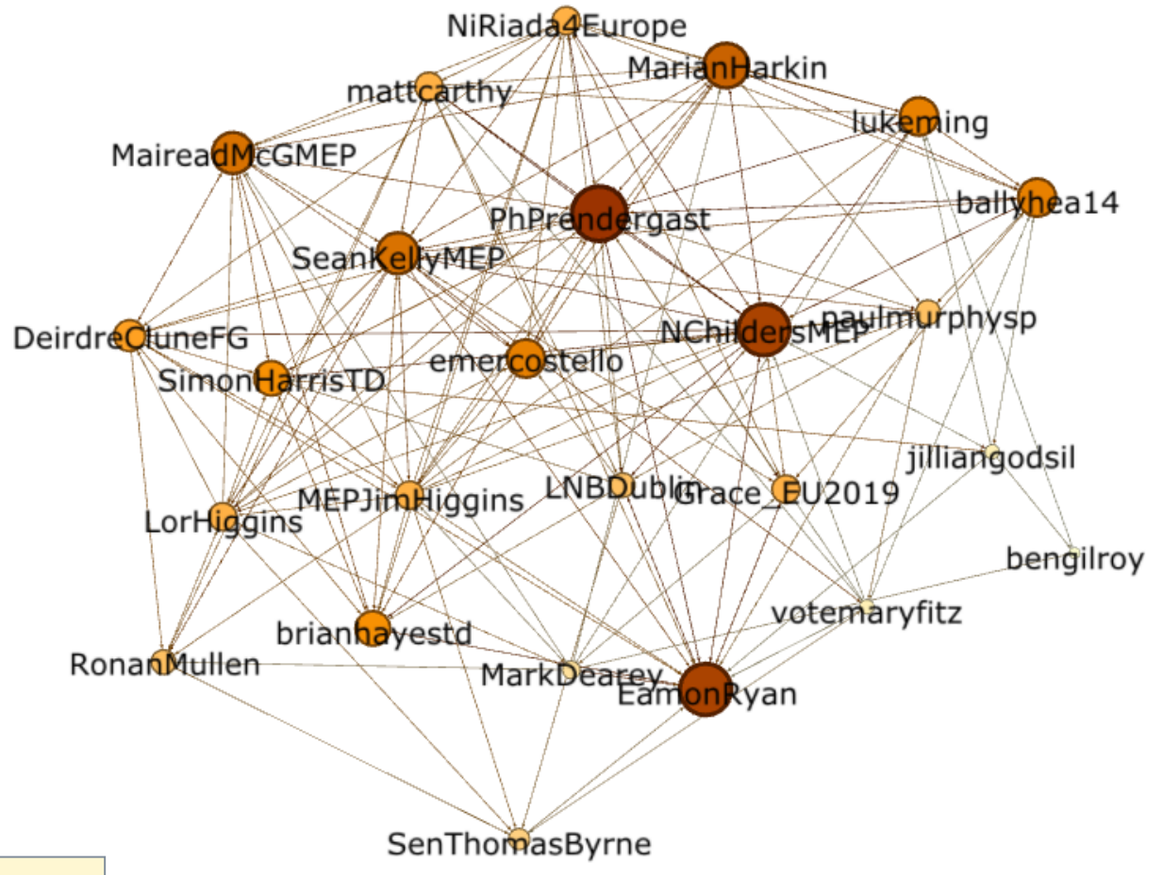
- The “Filters” tab supports complex methods to temporarily highlight or hide subsets of nodes and edges in the graph.
- Nodes can be filtered by attribute value or based on node statistics.



1. Choose the node attribute to filter on.

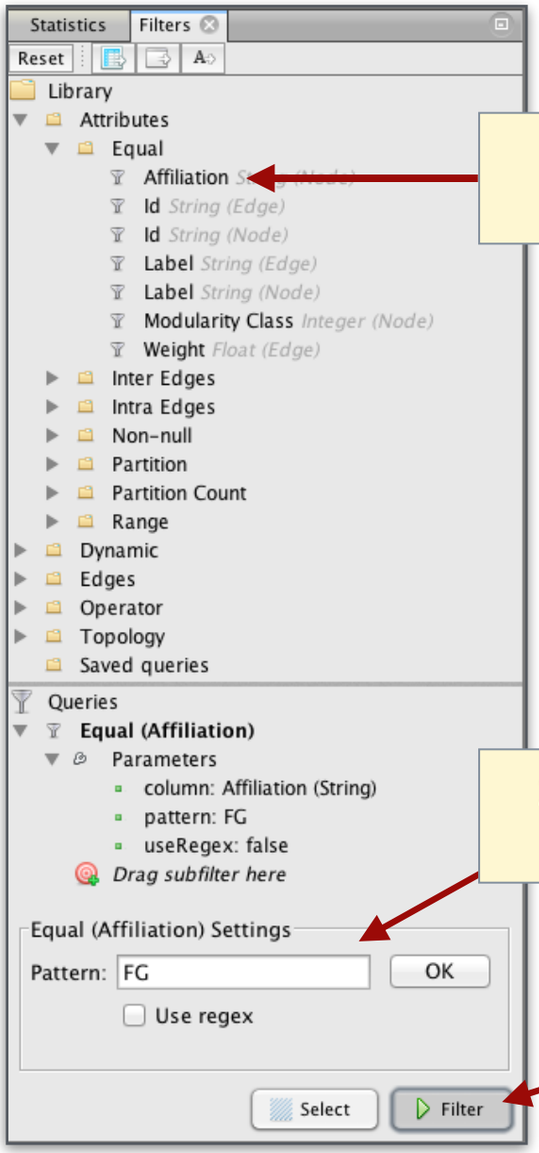
2. Choose the values to search for, click “OK”.

3. Click “Filter” to hide other nodes.



Filtering Nodes

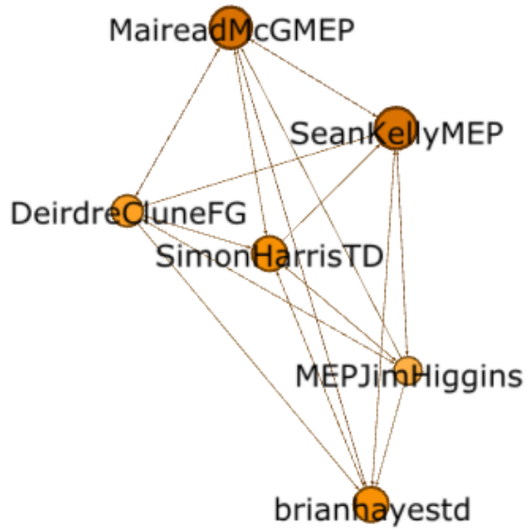
- The “Filters” tab supports complex methods to temporarily highlight or hide subsets of nodes and edges in the graph.
- Nodes can be filtered by attribute value or based on node statistics.



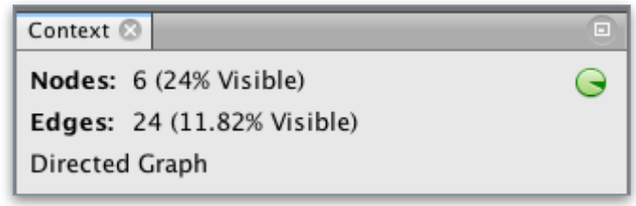
1. Choose the node attribute to filter on.

2. Choose the values to search for, click “OK”.

3. Click “Filter” to hide other nodes.



Filtered set of nodes based on affiliation = “FG”

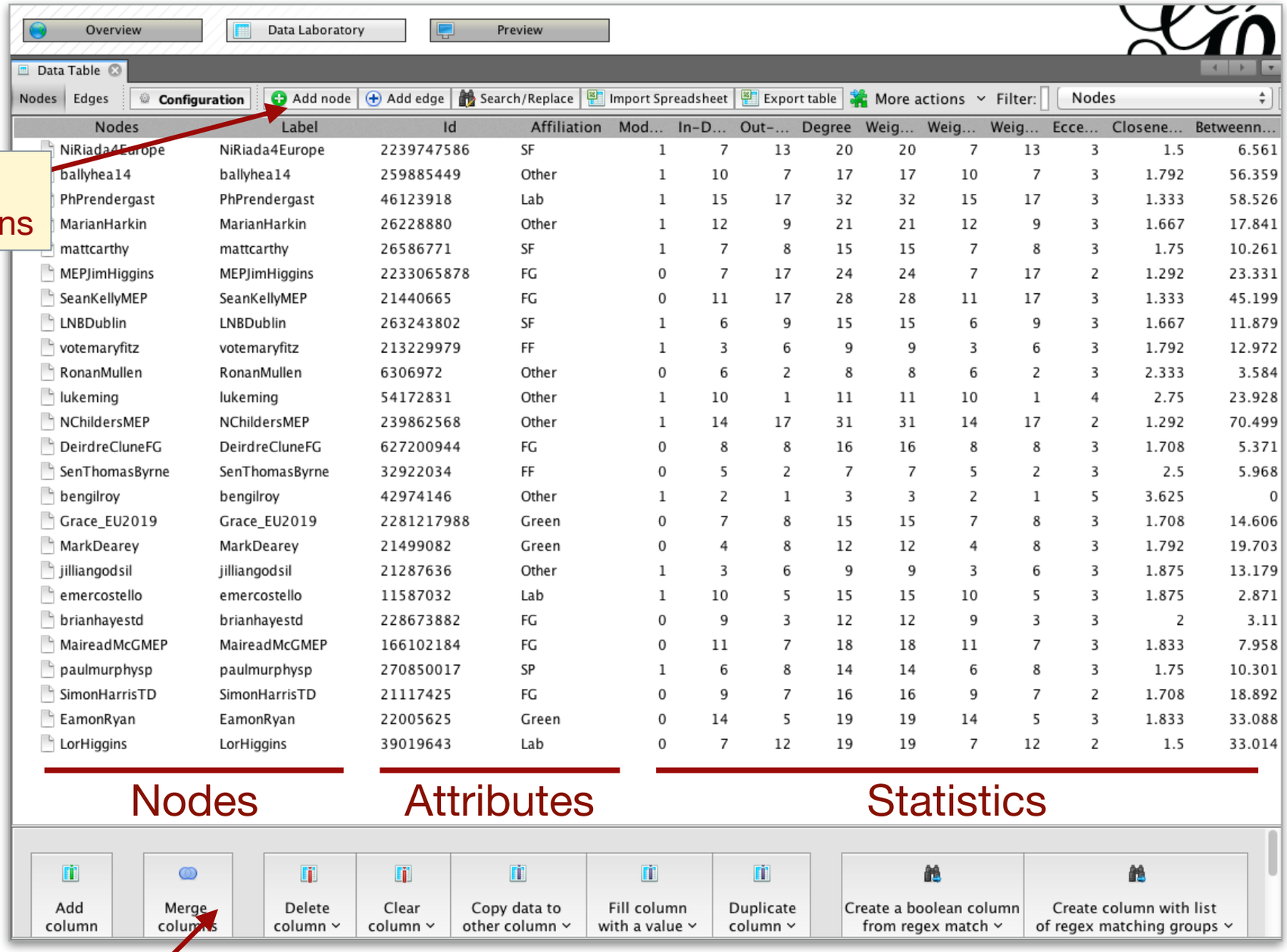


Gephi Data Laboratory Screen

- Laboratory: alternative tabular view of the same graph data.

Graph Nodes

Graph operations

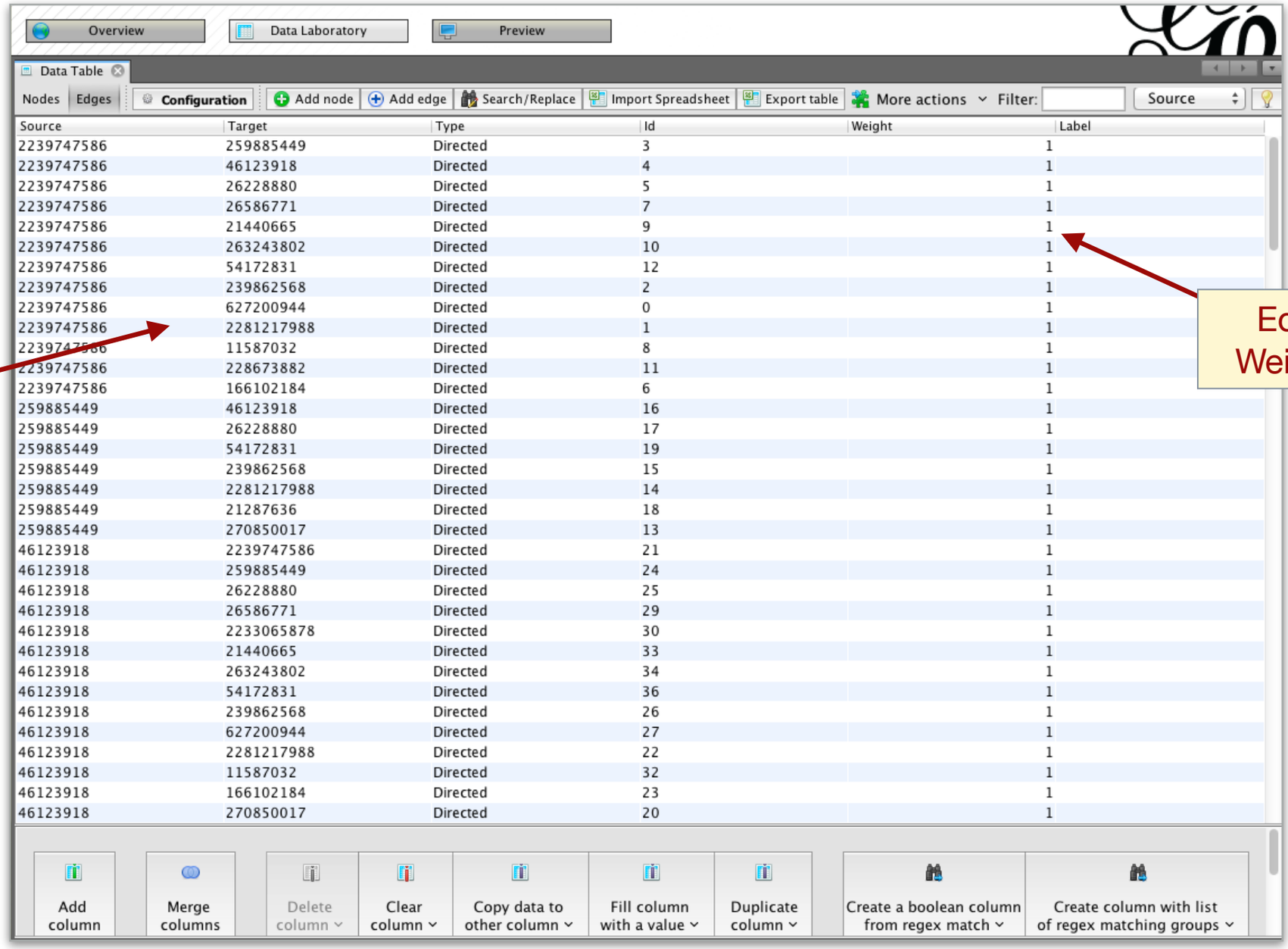


Column operations

Gephi Data Laboratory Screen

- Laboratory: alternative tabular view of the same graph data.

Graph Edges

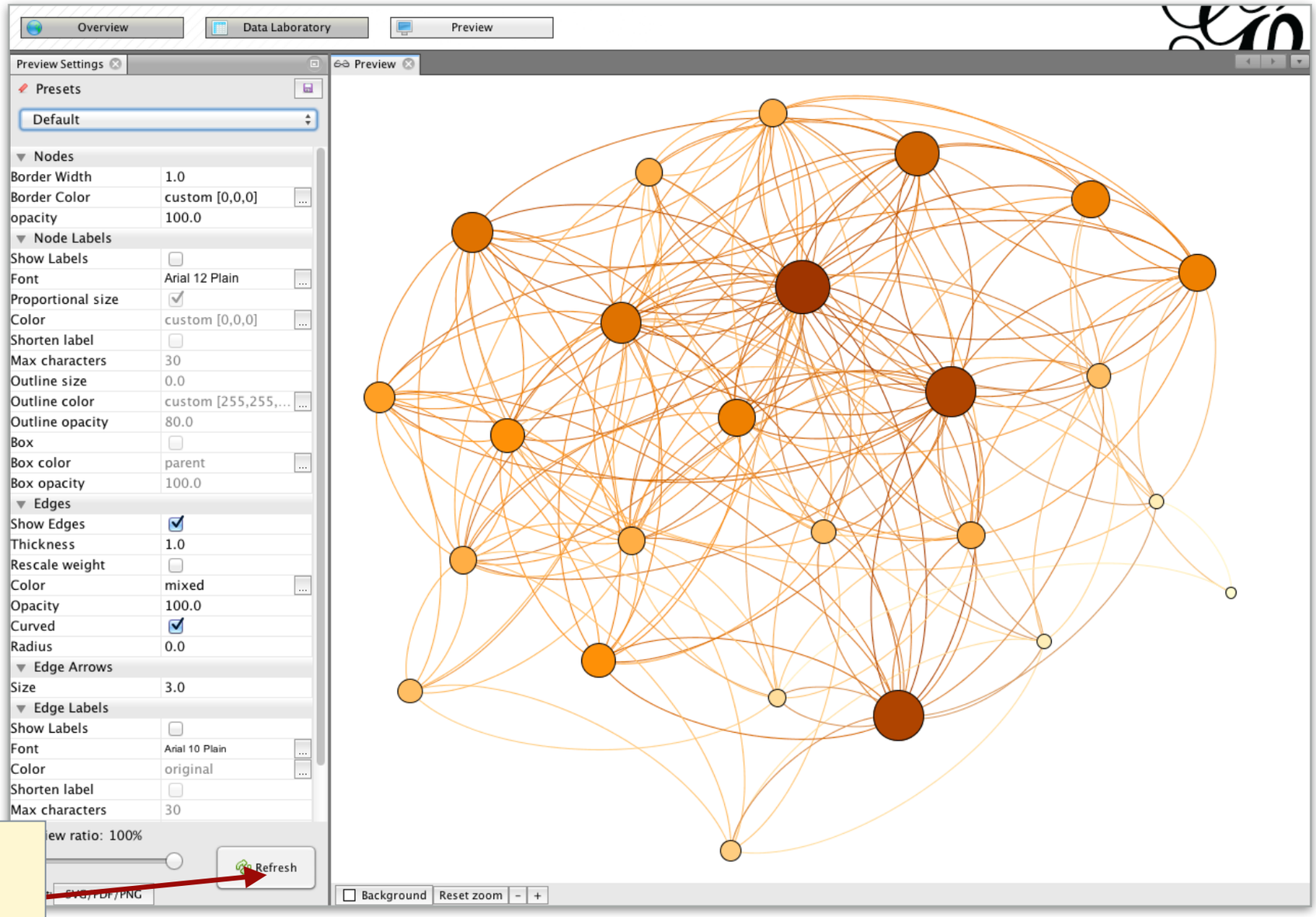


Gephi Preview Screen

- Preview: Tweak the appearance of your graph before exporting a publication quality image.

Node appearance

Edge appearance



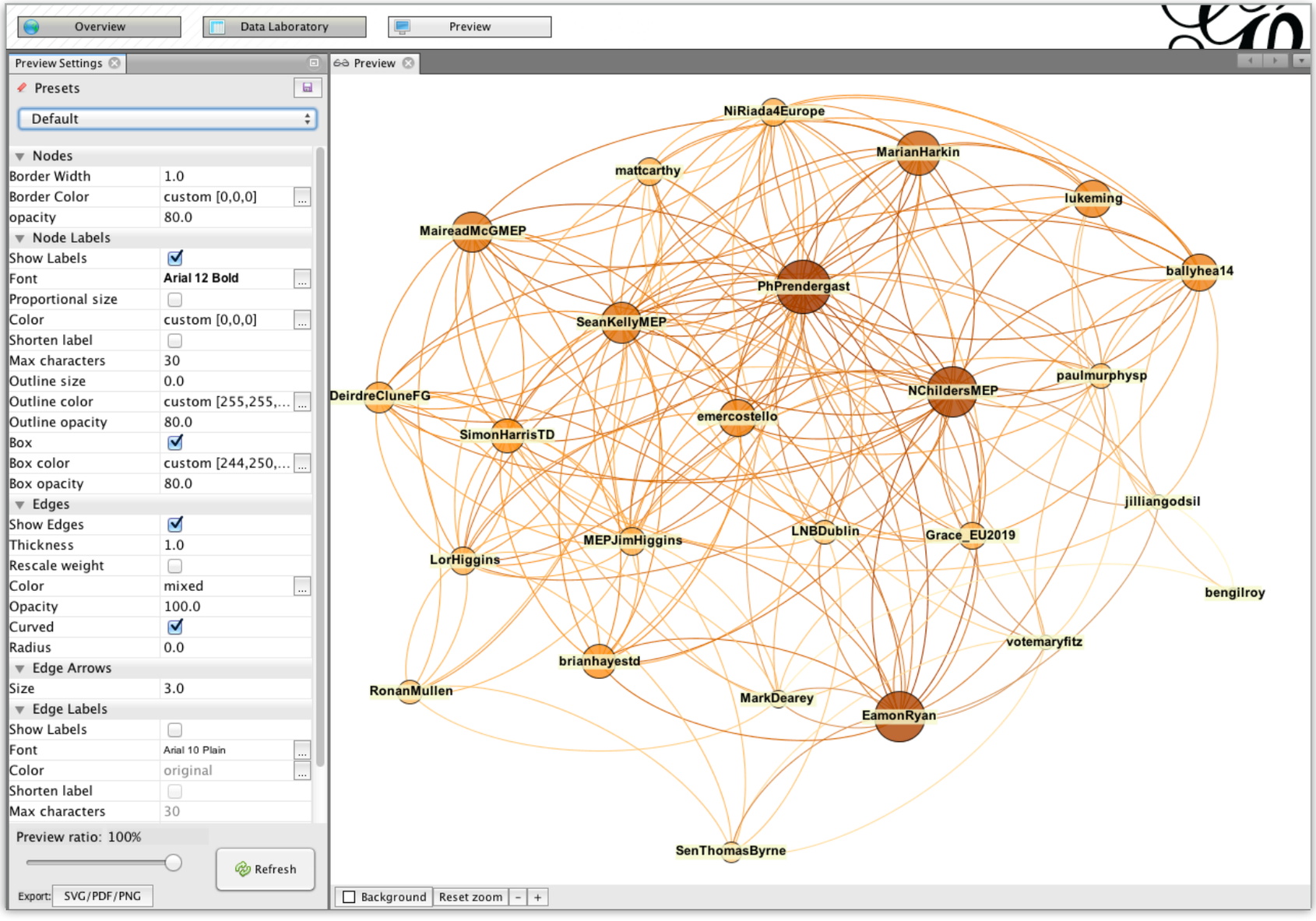
Hit “Refresh” each time to update

Use mouse or icons to move/crop graph

Gephi Preview Screen

- Preview: Tweak the appearance of your graph before exporting a publication quality image.

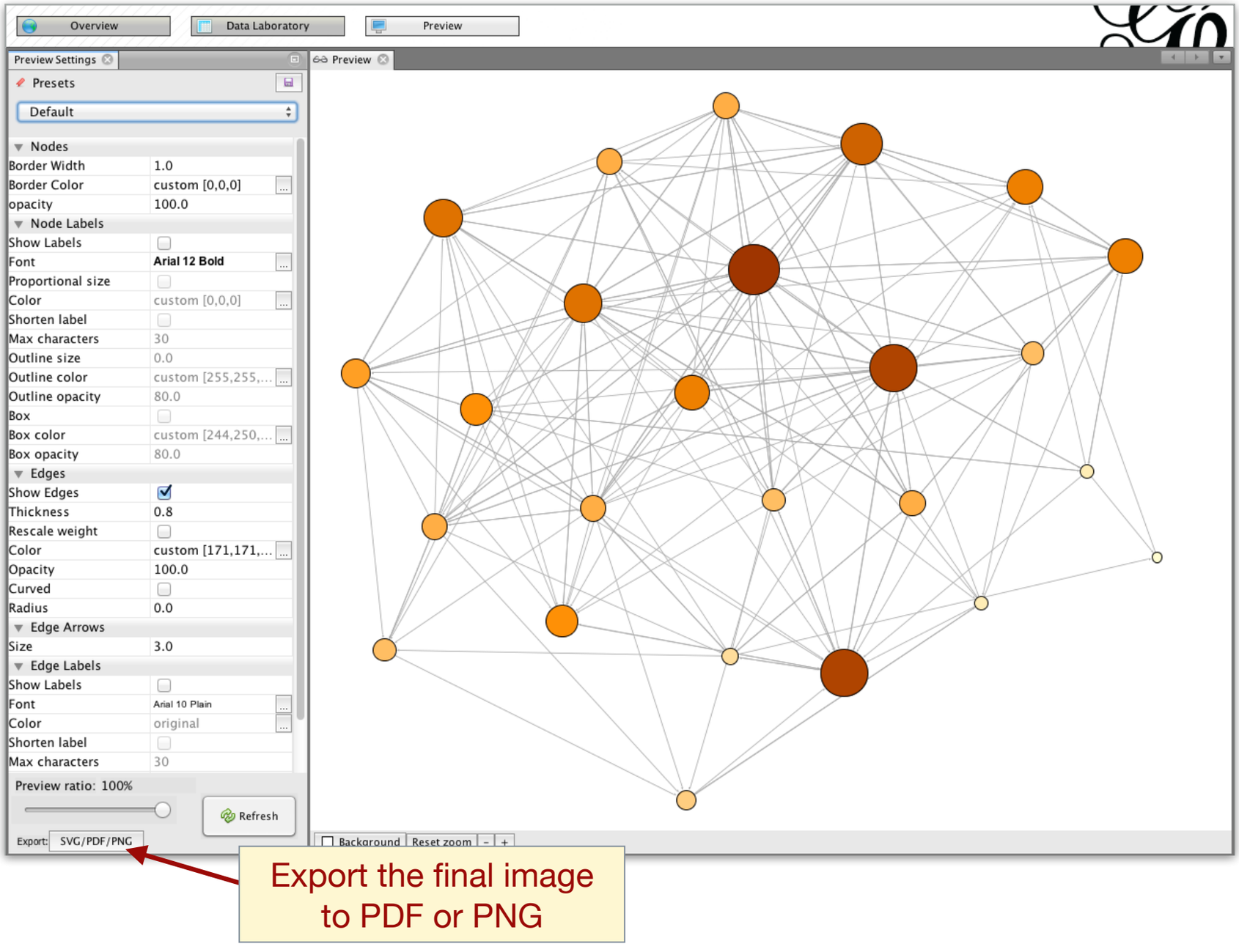
Node appearance



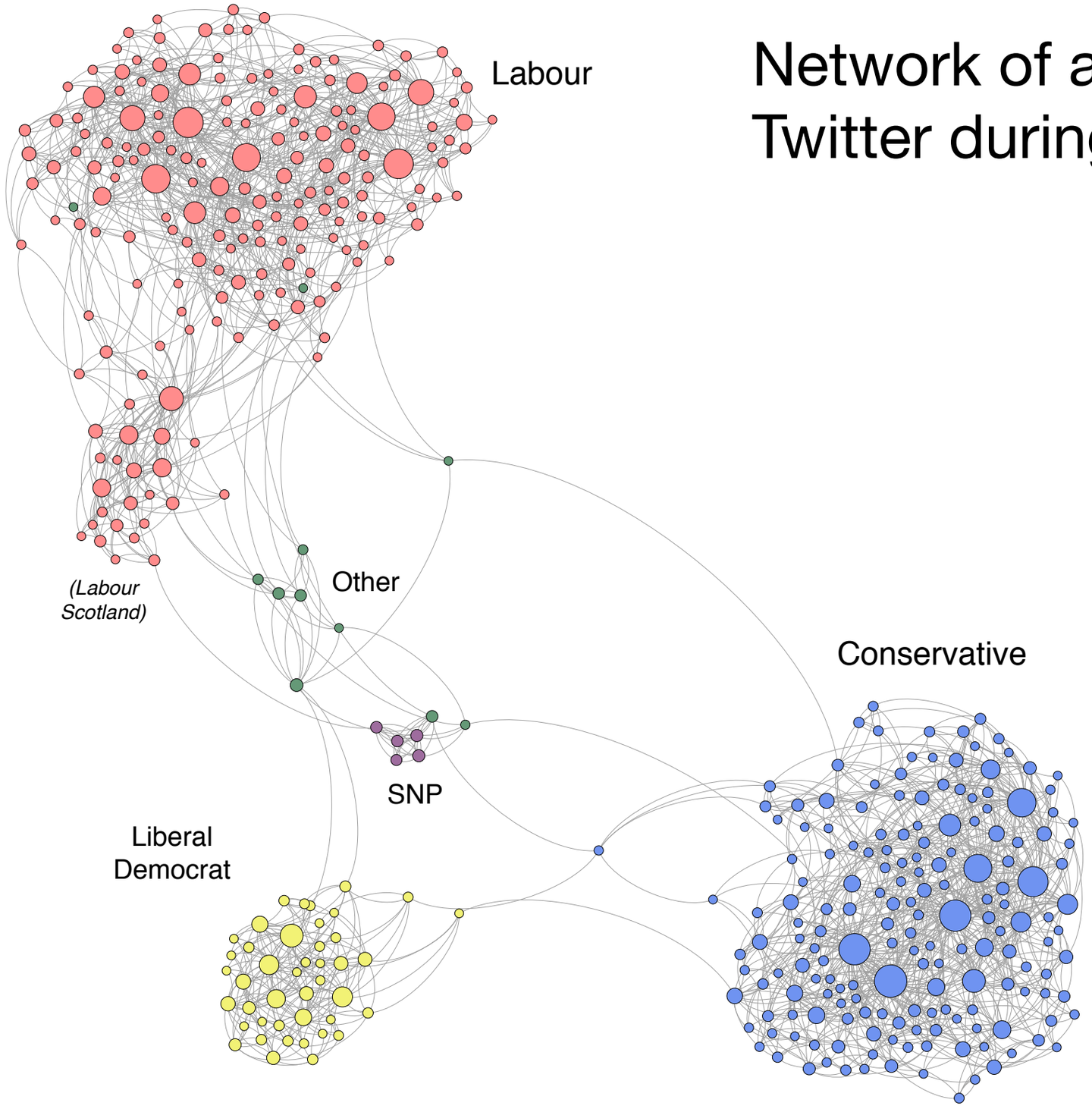
Gephi Preview Screen

- Preview: Tweak the appearance of your graph before exporting a publication quality image.

Edge appearance



Example: UK Politics

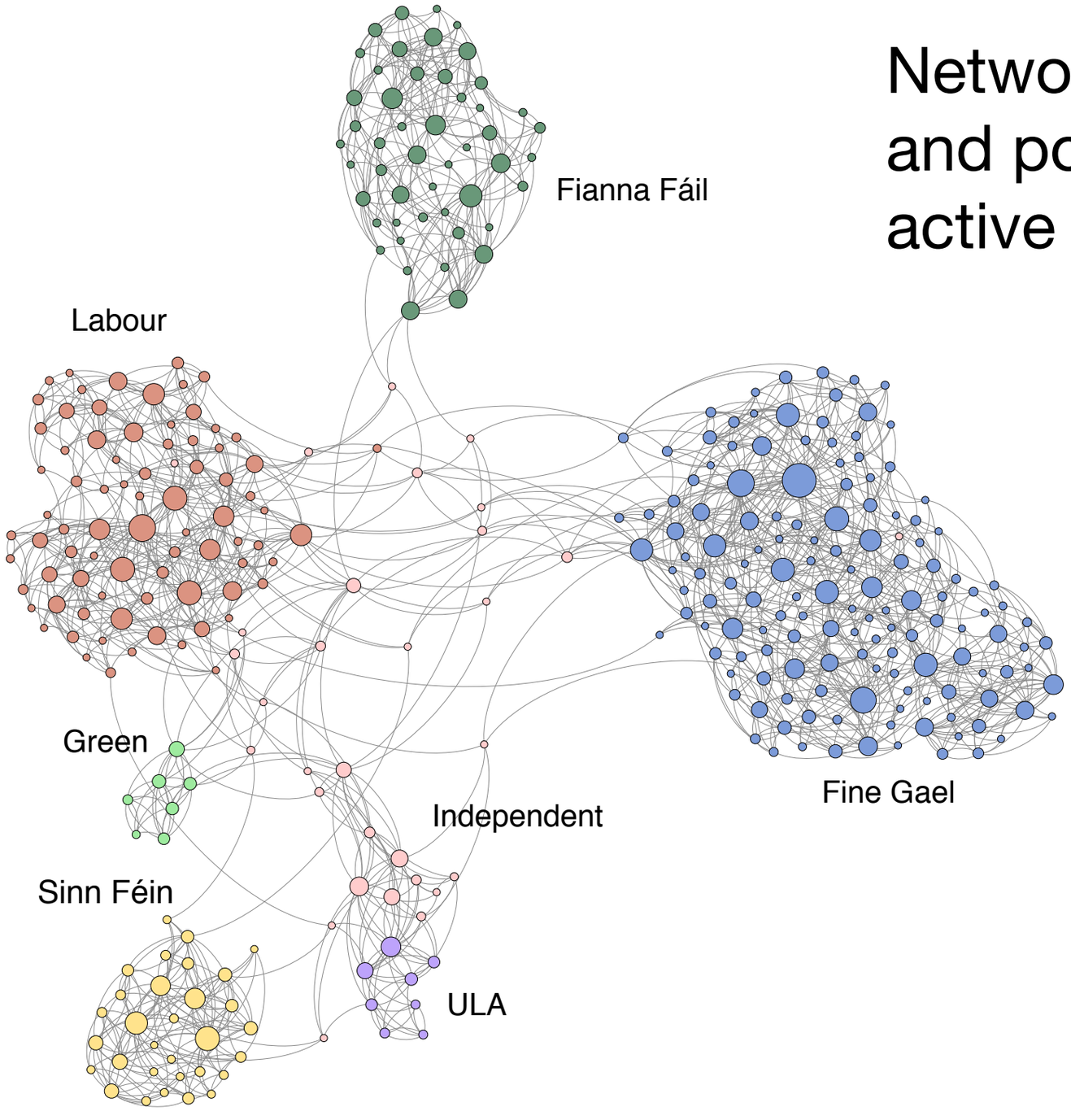


Network of all UK MPs active on Twitter during 2012.

http://mlg.ucd.ie/networks/politics-uk.html

Greene & Cunningham (2013)

Example: Irish Politics



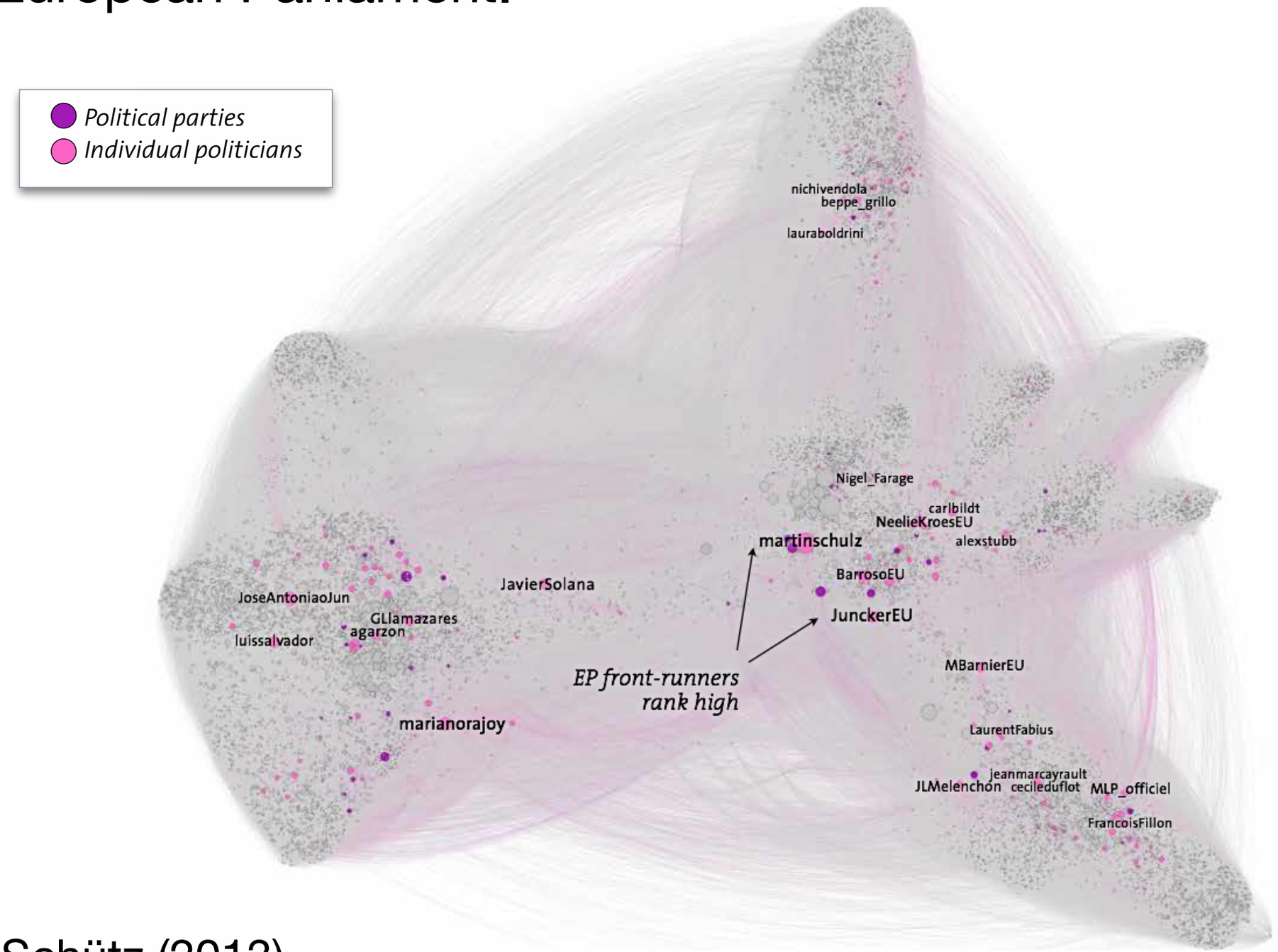
Network of 348 Irish politicians and political organisations active on Twitter in 2011-2012.

http://mlg.ucd.ie/networks/politics-ie.html

Greene & Cunningham (2013)

Example: European Political Twittersphere

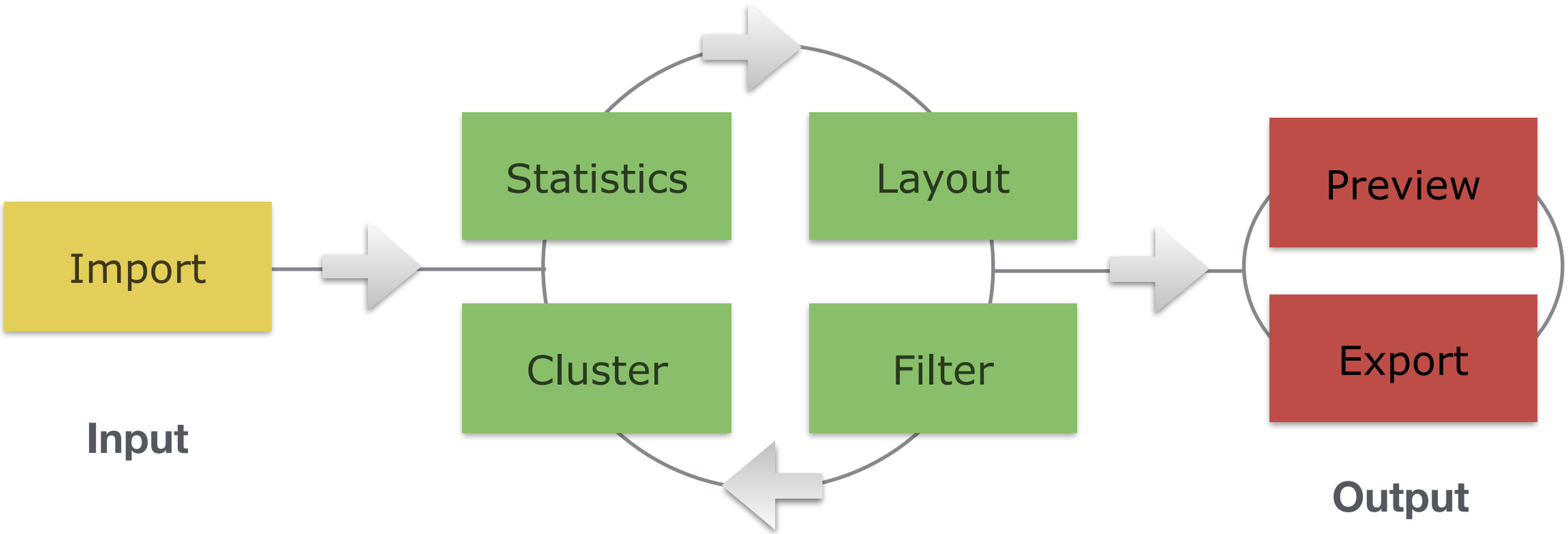
Ranking of Twitter accounts for prominent members and groups of European Parliament.



Maireder & Schütz (2013)

Conclusion

- Gephi provides a powerful workflow for exploring and visualising graphs…



Gephi: an **open source software** for **exploring** and **manipulating networks**.
M Bastian, S Heymann, M Jacomy - ICWSM, 2009 - aaai.org
Abstract **Gephi** is an **open source software** for graph and **network** analysis. It uses a 3D render engine to display large **networks** in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets ...
Cited by 708 Related articles All 15 versions Import into BibTeX Saved More

http://gephi.org

References

- M. Bastian, S. Heymann & M. Jacomy (2009). Gephi: an open source software for exploring and manipulating networks. Proc. ICWSM-09.
- A. Maireder & F. Schütz (2014). The European Political Twittersphere: Network of top users discussing the 2014 European Elections. GFK white paper.
- Greene, D. & Cunningham, P. (2013). Producing a Unified Graph Representation from Multiple Social Network Views. Proc. ACM Web Science’13.