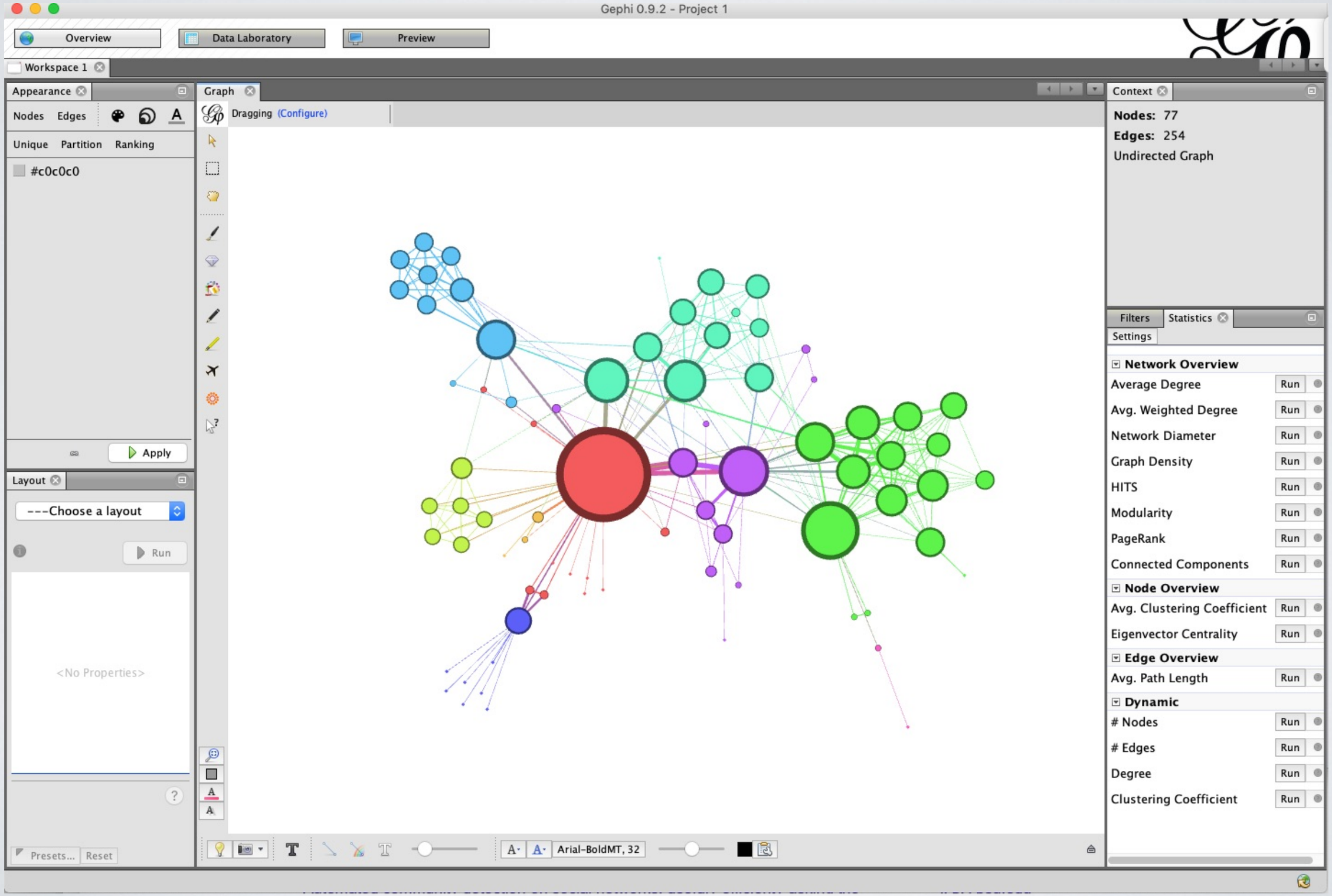
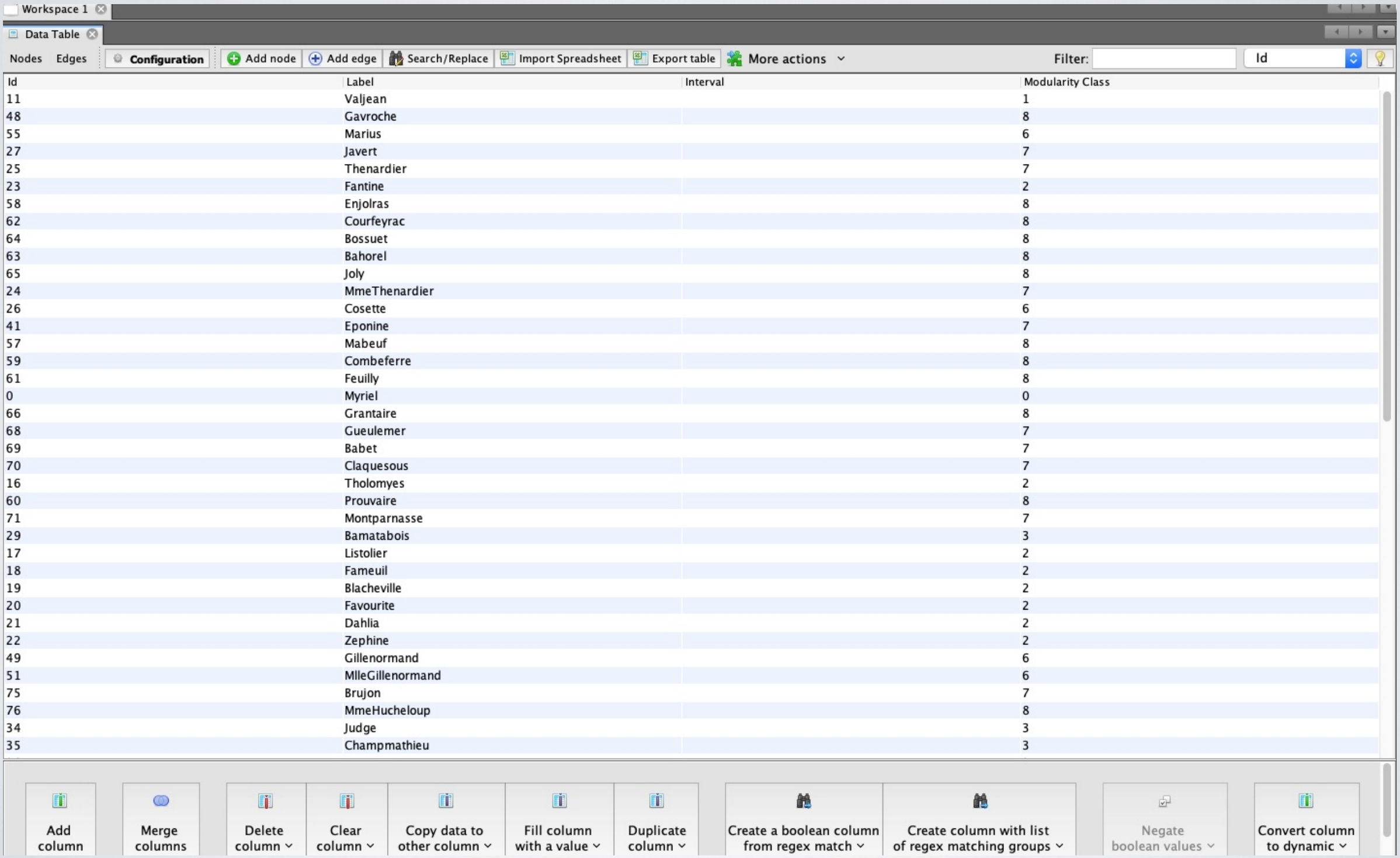
GEPHI: HANDS-ON

OVERVIEW WINDOW

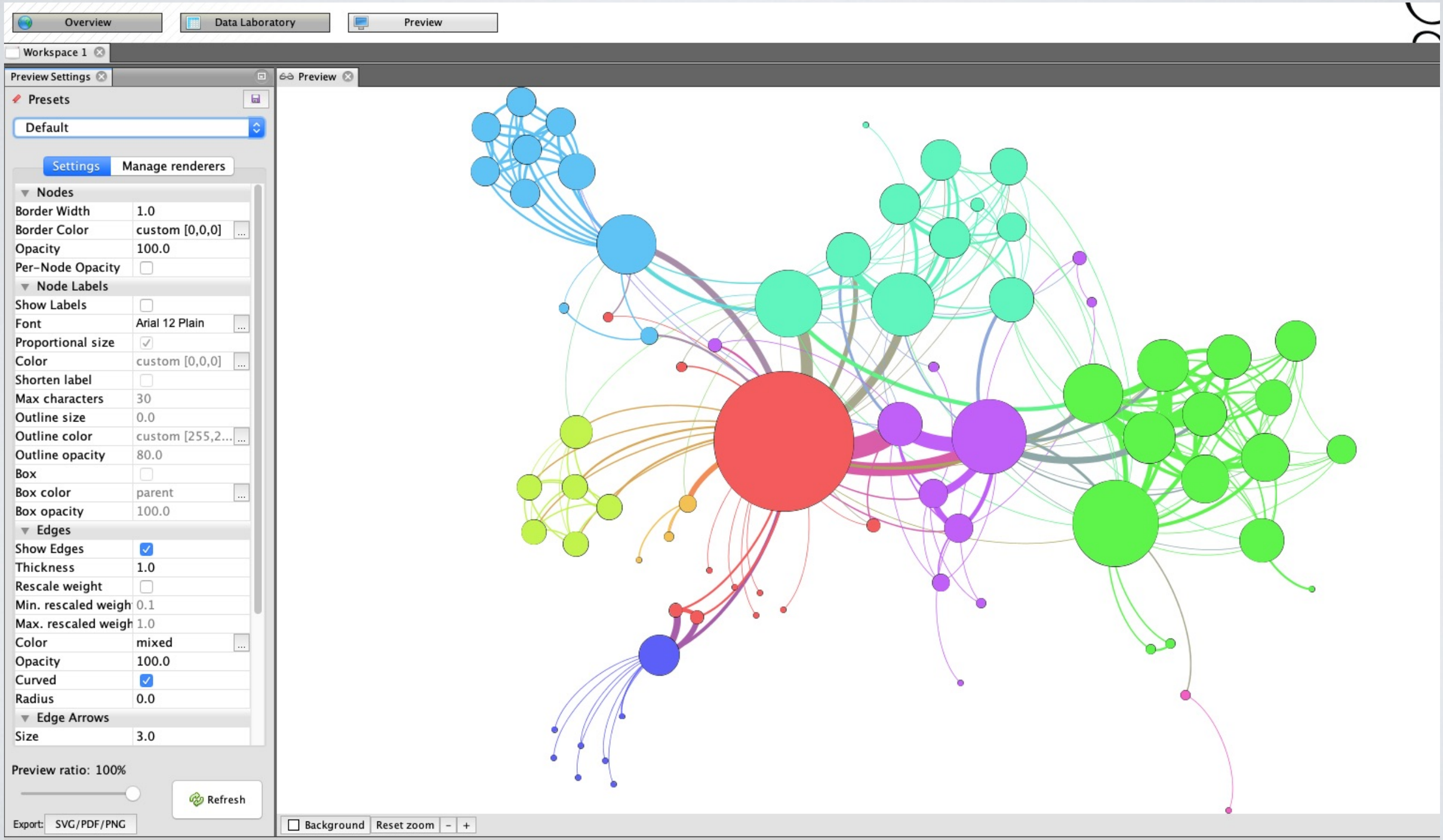


DATA LABORATORY



Id	Label	Interval	Modularity Class
11	Valjean		1
48	Gavroche		8
55	Marius		6
27	Javert		7
25	Thenardier		7
23	Fantine		2
58	Enjolras		8
62	Courfeyrac		8
64	Bossuet		8
63	Bahorel		8
65	Joly		8
24	MmeThenardier		7
26	Cosette		6
41	Eponine		7
57	Mabeuf		8
59	Combeferre		8
61	Feuilly		8
0	Myriel		0
66	Grantaire		8
68	Gueulemer		7
69	Babet		7
70	Claquesous		7
16	Tholomyes		2
60	Prouvaire		8
71	Montparnasse		7
29	Bamatabois		3
17	Listolier		2
18	Fameuil		2
19	Blacheville		2
20	Favourite		2
21	Dahlia		2
22	Zephine		2
49	Gillenormand		6
51	MlleGillenormand		6
75	Brujon		7
76	MmeHucheloup		8
34	Judge		3
35	Champmathieu		3

PREVIEW


Overview
Data Laboratory
Preview
Workspace 1
Preview Settings
Presets
Default
Settings
Manage renderers
Nodes
Border Width 1.0
Border Color custom [0,0,0]
Opacity 100.0
Per-Node Opacity
Node Labels
Show Labels
Font Arial 12 Plain
Proportional size
Color custom [0,0,0]
Shorten label
Max characters 30
Outline size 0.0
Outline color custom [255,2...
Outline opacity 80.0
Box
Box color parent
Box opacity 100.0
Edges
Show Edges
Thickness 1.0
Rescale weight
Min. rescaled weigh 0.1
Max. rescaled weigh 1.0
Color mixed
Opacity 100.0
Curved
Radius 0.0
Edge Arrows
Size 3.0
Preview ratio: 100%
Refresh
Export: SVG/PDF/PNG
Preview
Background
Reset zoom
-
+

1)IMPORT DATA

IMPORT

- Gephi import all standard graph format.

	Edge List/Matrix Structure	XML Struture	Edge Weight	Attributes	Visualization Attributes	Attribute Default Value	Hierarchical Graphs	Dynamics
CSV	■		■					
DL Ucinet	■		■	■				
DOT Graphviz			■		■			
GDF			■	■	■	■		
GEXF		■	■	■	■	■	■	■
GML			■	■	■			
GraphML		■	■	■	■	■	■	
NET Pajek	■		■		■			
TLP Tulip								
VNA Netdraw			■	■				
Spreadsheet*			■	■				■

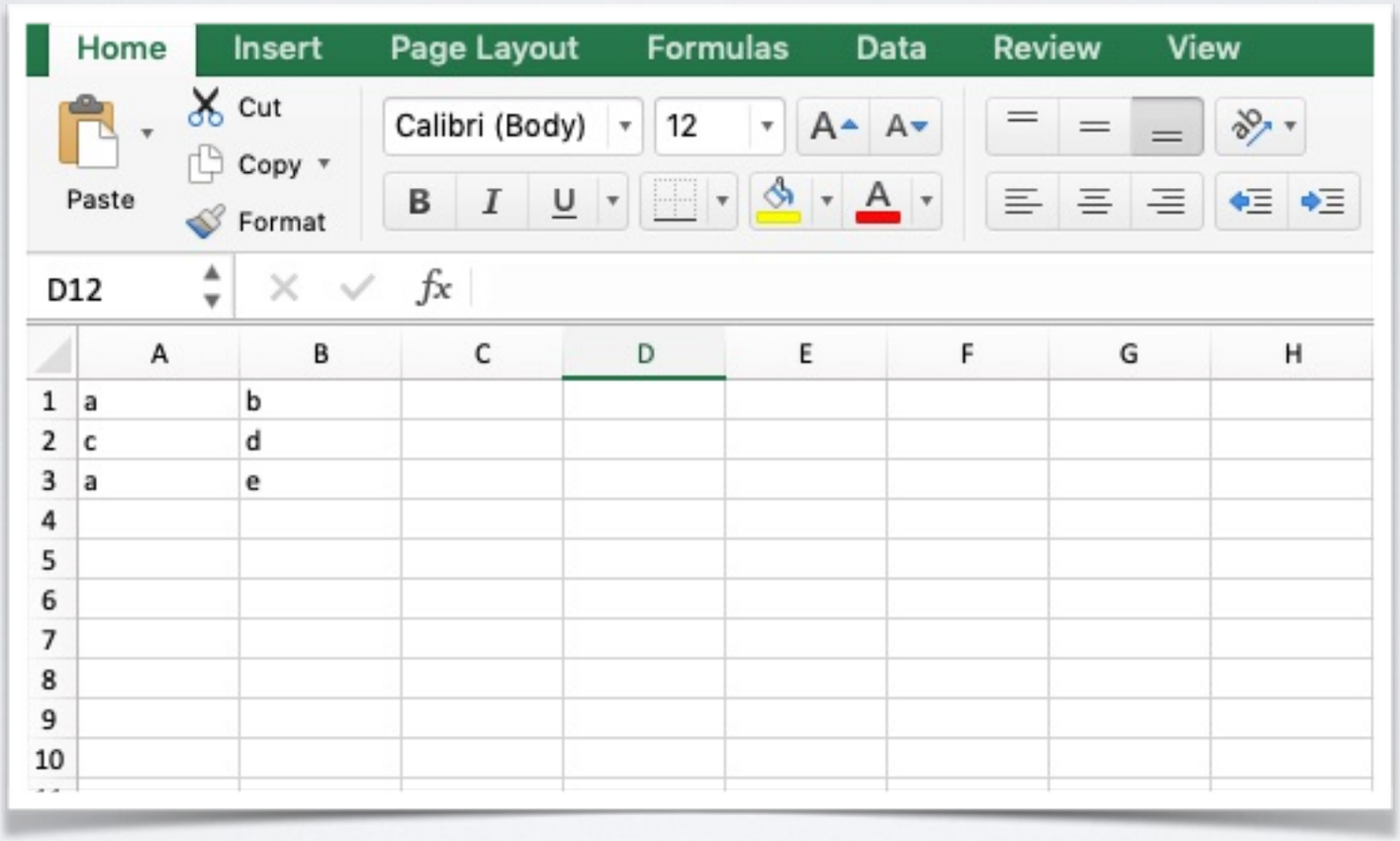
IMPORT

- Most useful for YOUR personal data
- Spreadsheet (Excel, open office cal..)
  - Matrix
  - CSV
- CSV : comma separated value (same as spreadsheet)
- GEXF (to reload a saved file)

IMPORT

- 3 ways to represent data:
  - Adjacency list
  - Edge table (+Node table)
  - Matrix

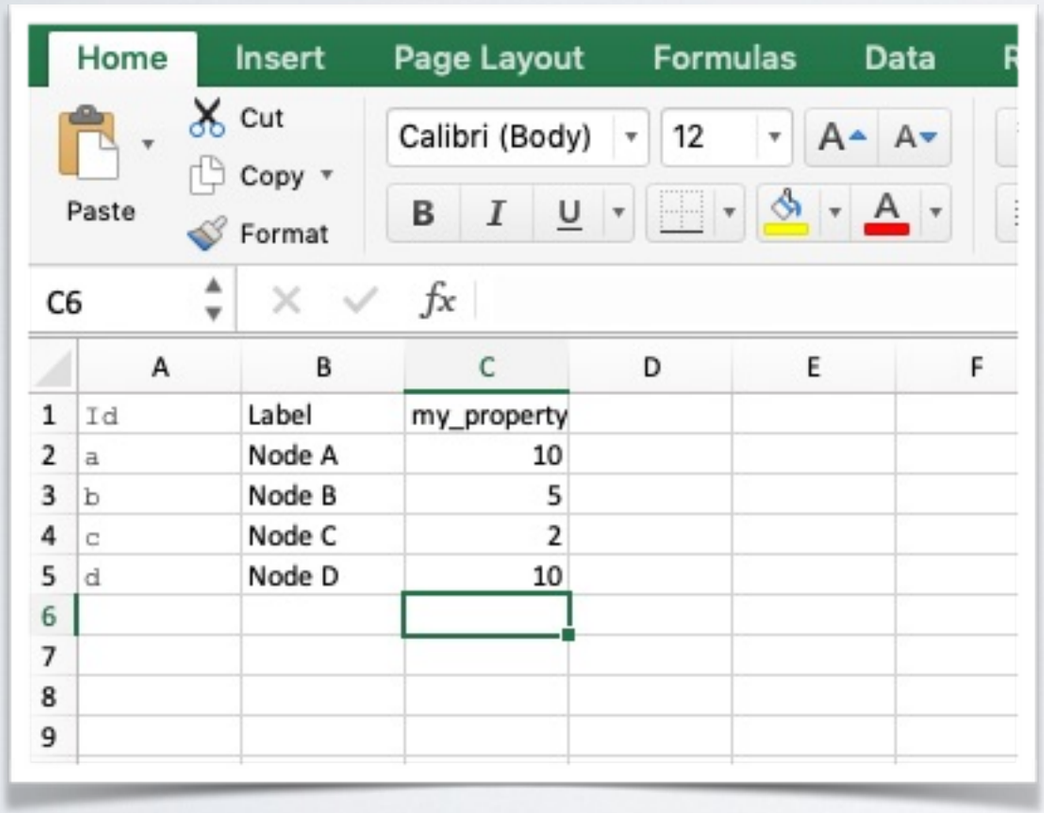
ADJACENCY LIST



	A	B	C	D	E	F	G	H
1	a	b						
2	c	d						
3	a	e						
4								
5								
6								
7								
8								
9								
10								

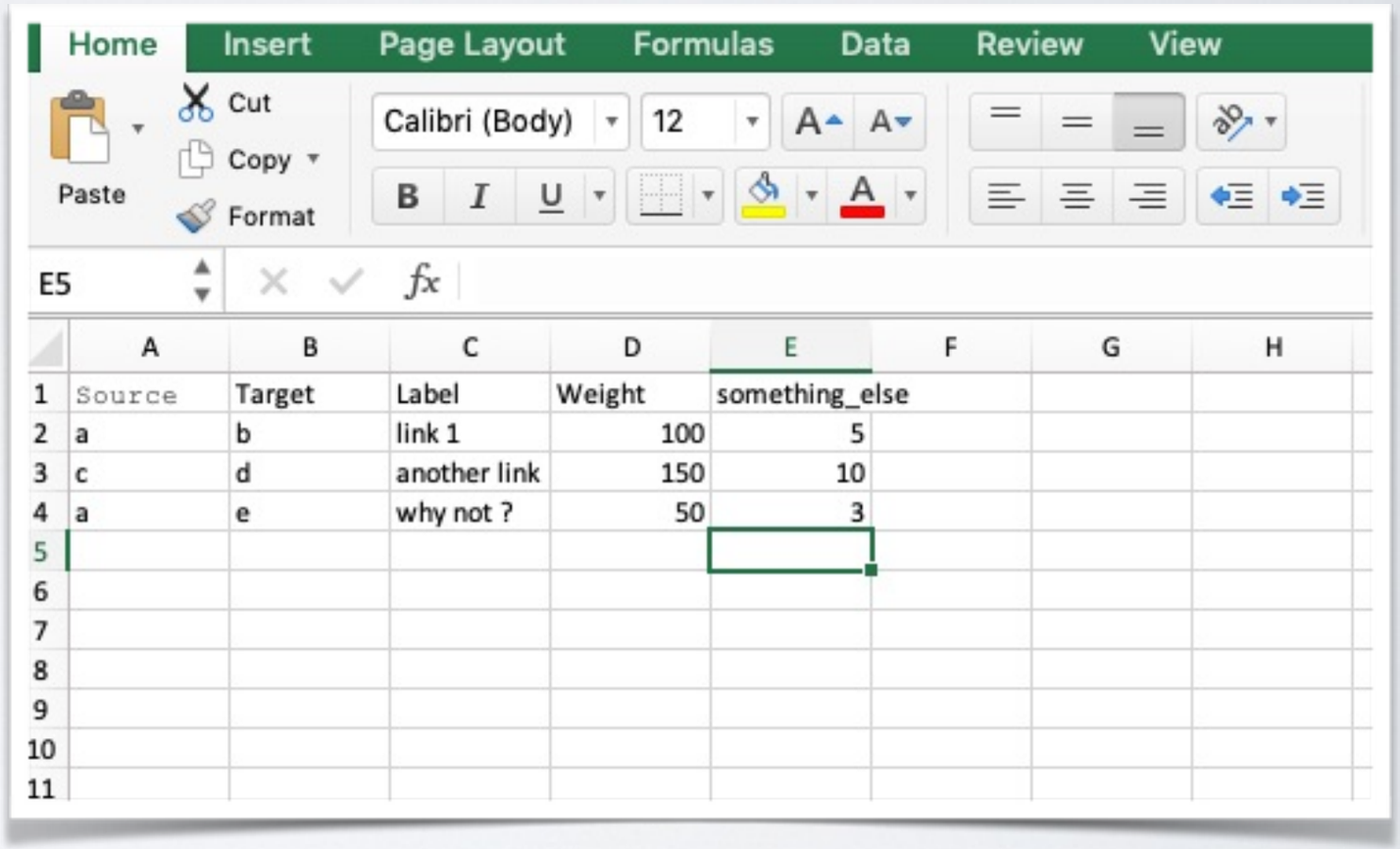
TABLES

Node table



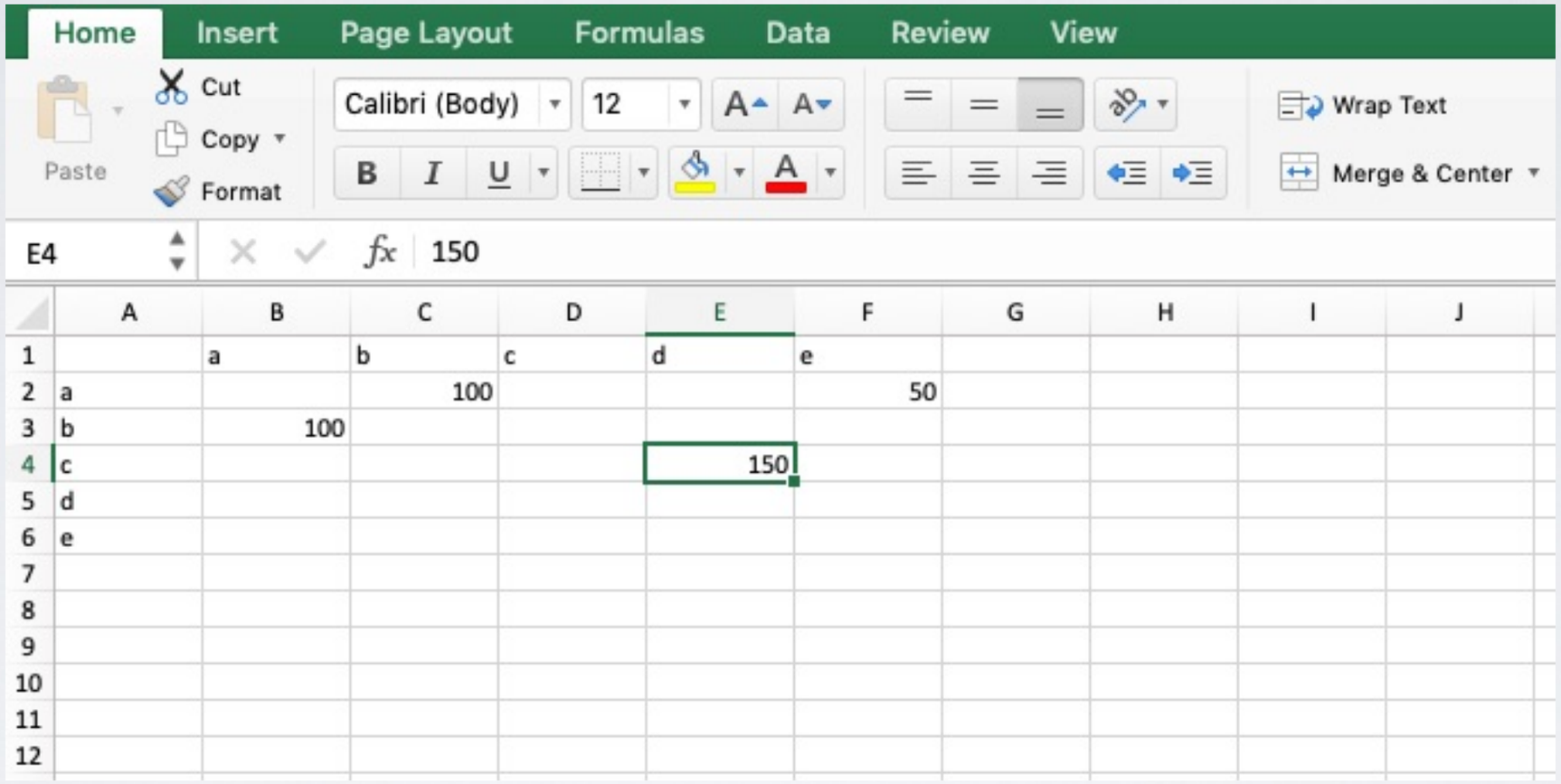
	A	B	C	D	E	F
1	Id	Label	my_property			
2	a	Node A	10			
3	b	Node B	5			
4	c	Node C	2			
5	d	Node D	10			
6						
7						
8						
9						

Edge table



	A	B	C	D	E	F	G	H
1	Source	Target	Label	Weight	something_else			
2	a	b	link 1	100	5			
3	c	d	another link	150	10			
4	a	e	why not ?	50	3			
5								
6								
7								
8								
9								
10								
11								

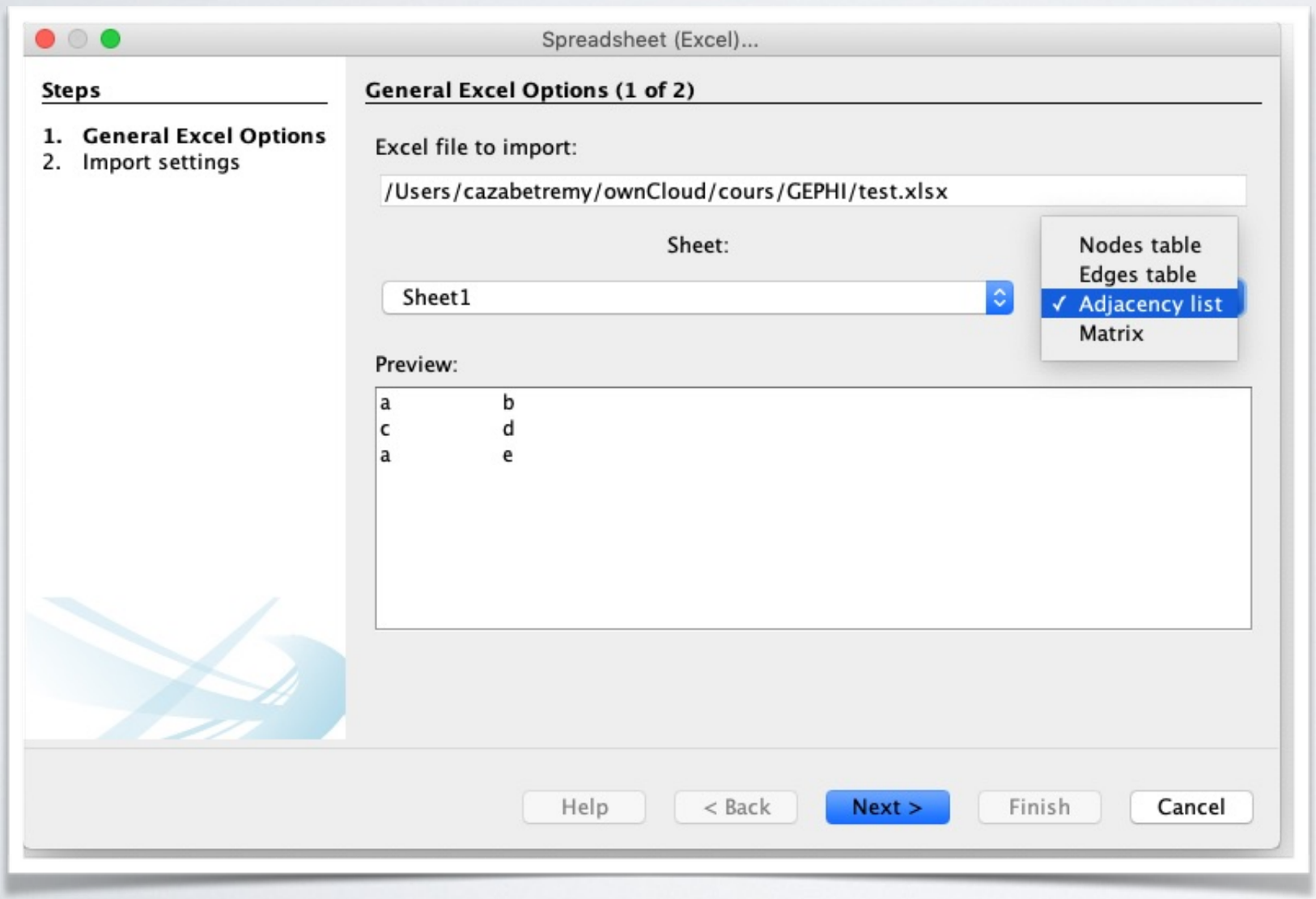
MATRIX

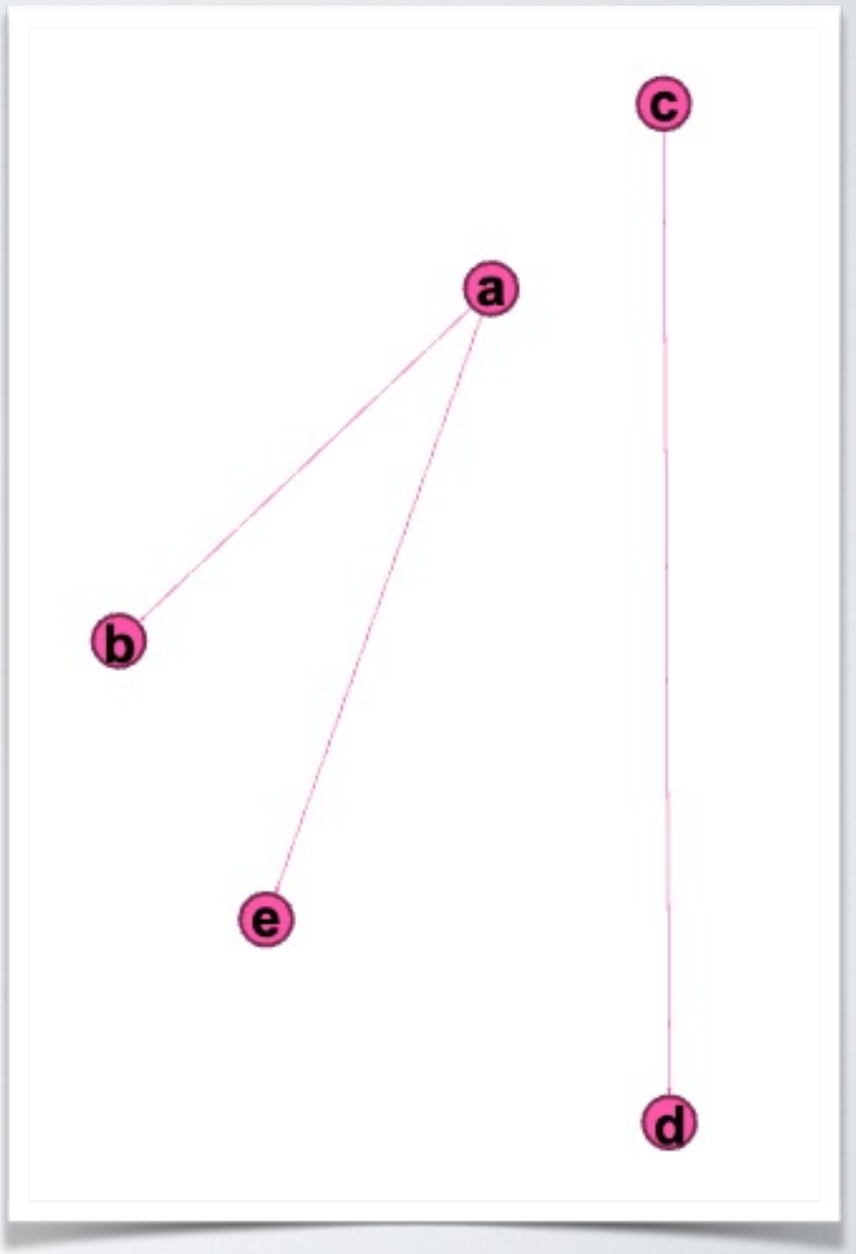


	A	B	C	D	E	F	G	H	I	J
1		a	b	c	d	e				
2	a		100			50				
3	b	100								
4	c				150					
5	d									
6	e									
7										
8										
9										
10										
11										
12										

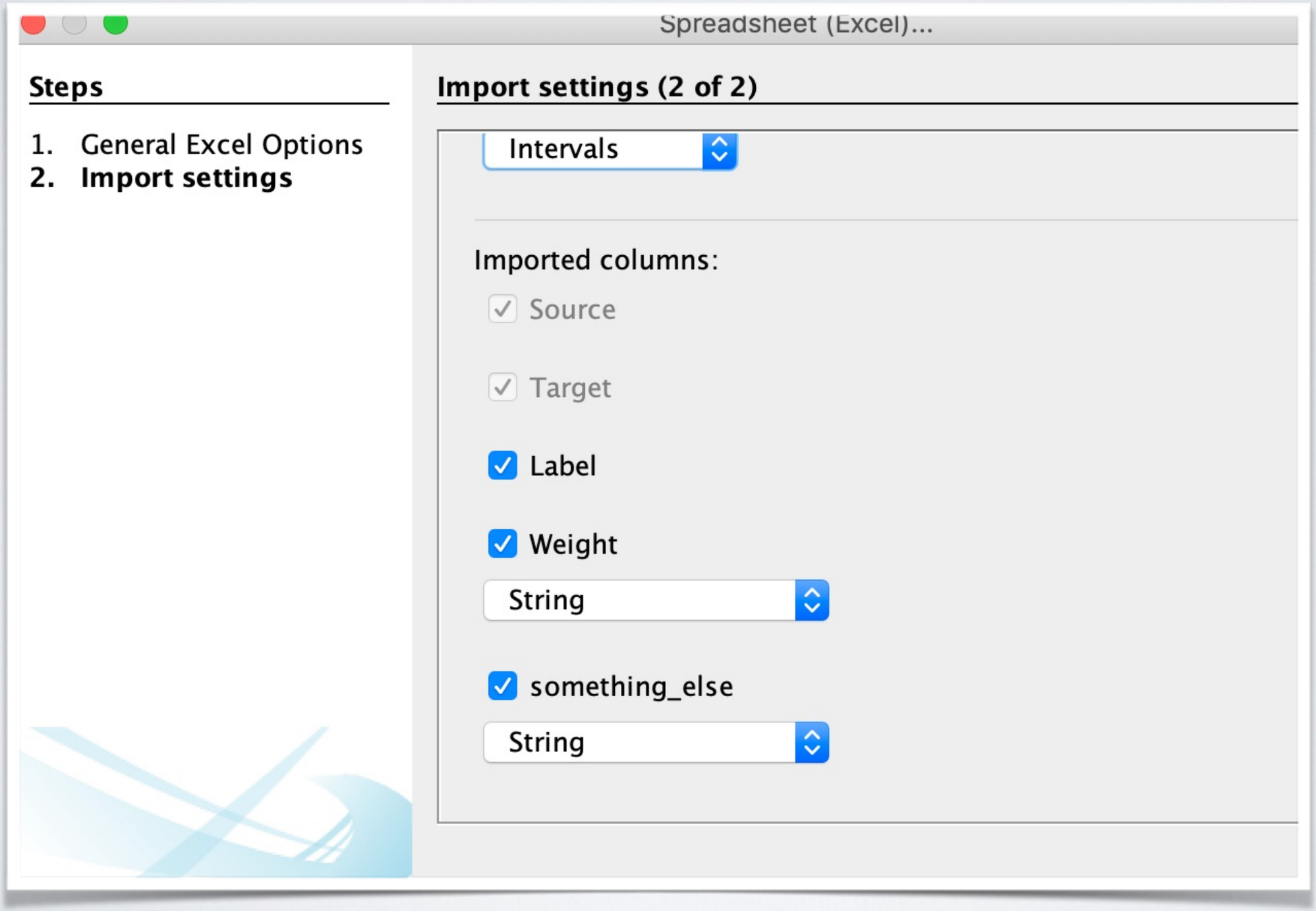
IMPORT

- Monfichier.xlsx





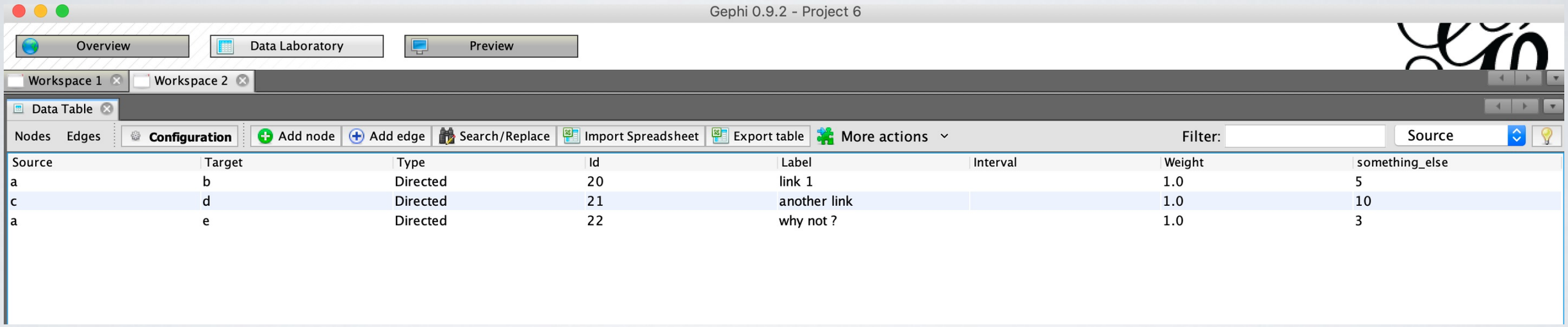
IMPORT


Spreadsheet (Excel)...
Steps
1. General Excel Options
2. Import settings
Import settings (2 of 2)
Intervals
Imported columns:
Source
Target
Label
Weight
String
something_else
String

IMPORT

- Whole graph can be loaded in a single step
  - File > Open > …
- Graph can be completed after loading
  - (For an already imported graph)
  - Data Laboratory > Import Spreadsheet
  - For instance, Node Table/Edge Table

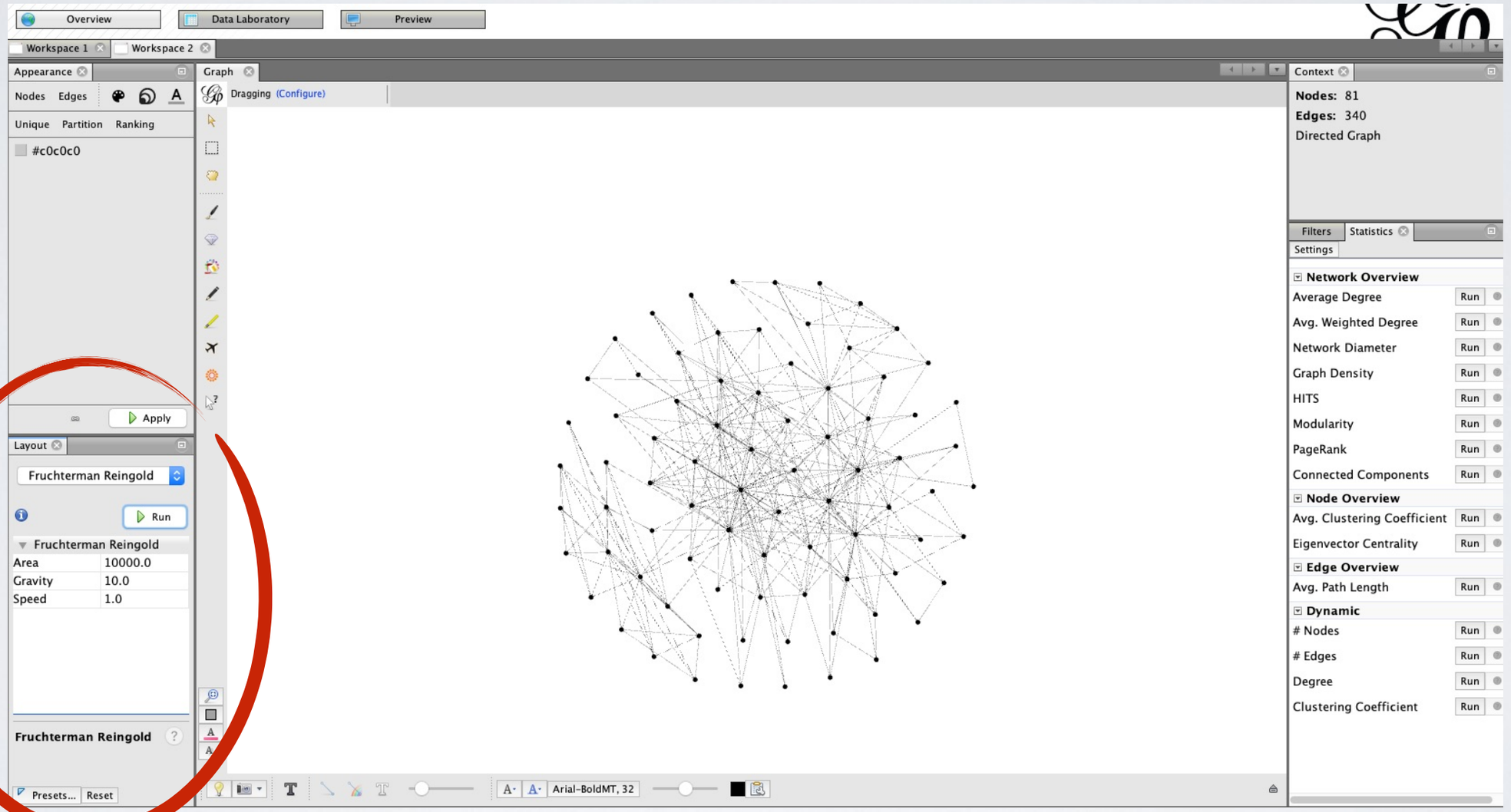
IMPORT


Gephi 0.9.2 - Project 6
Overview
Data Laboratory
Preview
Workspace 1
Workspace 2
Data Table
Nodes
Edges
Configuration
Add node
Add edge
Search/Replace
Import Spreadsheet
Export table
More actions
Filter:
Source
Source Target Type Id Label Interval Weight something_else
a b Directed 20 link 1 1.0 5
c d Directed 21 another link 1.0 10
a e Directed 22 why not ? 1.0 3

IMPORT

- One can modify the data in Gephi
  - Add node/edge
  - Create new column (data property)
  - …
- But it is better to work on the original data

GRAPH LAYOUT


Overview
Data Laboratory
Preview
Workspace 1
Workspace 2
Appearance
Nodes
Edges
Unique
Partition
Ranking
#c0c0c0
Apply
Layout
Fruchterman Reingold
Run
Fruchterman Reingold
Area 10000.0
Gravity 10.0
Speed 1.0
Fruchterman Reingold
Presets...
Reset
Graph
Dragging (Configure)
Arial-BoldMT, 32
Context
Nodes: 81
Edges: 340
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity Run
PageRank Run
Connected Components Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Nodes Run
Edges Run
Degree Run
Clustering Coefficient Run

GRAPH LAYOUT

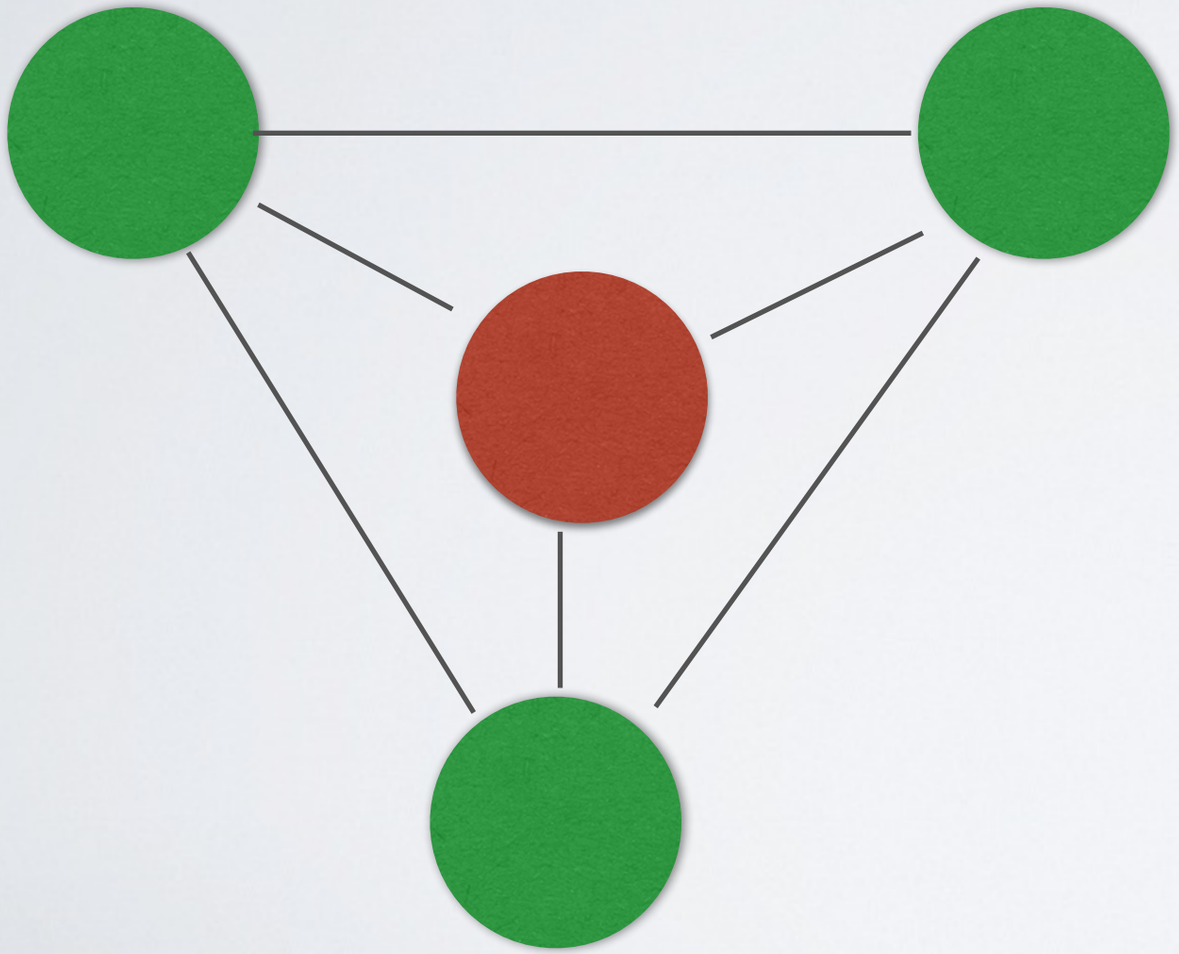
- Layout: Algorithm to automatically position nodes in the graph
  - Different algorithms are available. They provide different position of nodes.
  - Intuitive explanation of layouts: the force directed / Spring-Particule layout
    - Edges are springs that attract nodes
    - Nodes are repulsive particles
    - "let the physics do" until stabilization
    - => No guarantee on the properties of node positions.

- Beware interpretations!

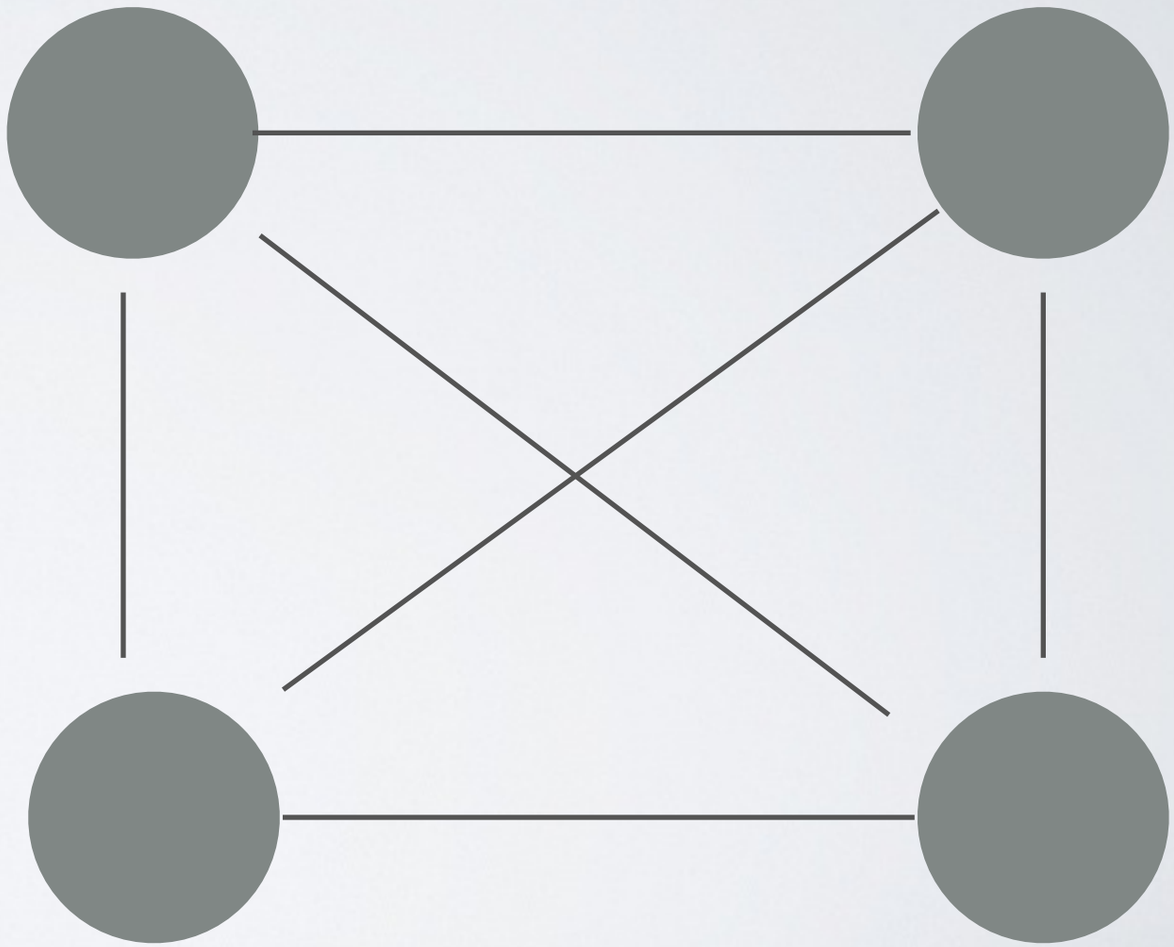
* ForceAtlas
* Fruchterman-Reingold
* YifanHu Multilevel
* OpenOrd
* ForceAtlas 2
* Circular Layout

GRAPH LAYOUT

Hierarchic/centralized



Egalitarian



Same graph…

GRAPH LAYOUT

- Rule of thumb workflow
  - 1)Try "Yifan Hu" and "Fruchterman reingold".
  - 2) Not satisfied ?
    - Go to ForceAtlas 2. Try to run and pause. Tune parameters, in particular "dissuad hubs", "prevent overlap", ..
    - Nodes are too close ? Use "Expansion" a few times
    - Nodes overlap ? Try "noverlap" one time
  - Algorithms (most) start from the **current** configuration, so the result can slowly improve (for large graphs in particular, were some parts might be "stuck" somewhere)

GRAPH LAYOUT

LET'S TRY !

SOME NETWORKS TO PLAY WITH:

COLORADO NETWORK INDEX
HTTPS://ICON.COLORADO.EDU/#!/

UCINET DATASETS
HTTPS://SITES.GOOGLE.COM/SITE/UCINETSOFTWARE/DATASETS

NETWORK REPOSITORY
HTTP://NETWORKREPOSITORY.COM

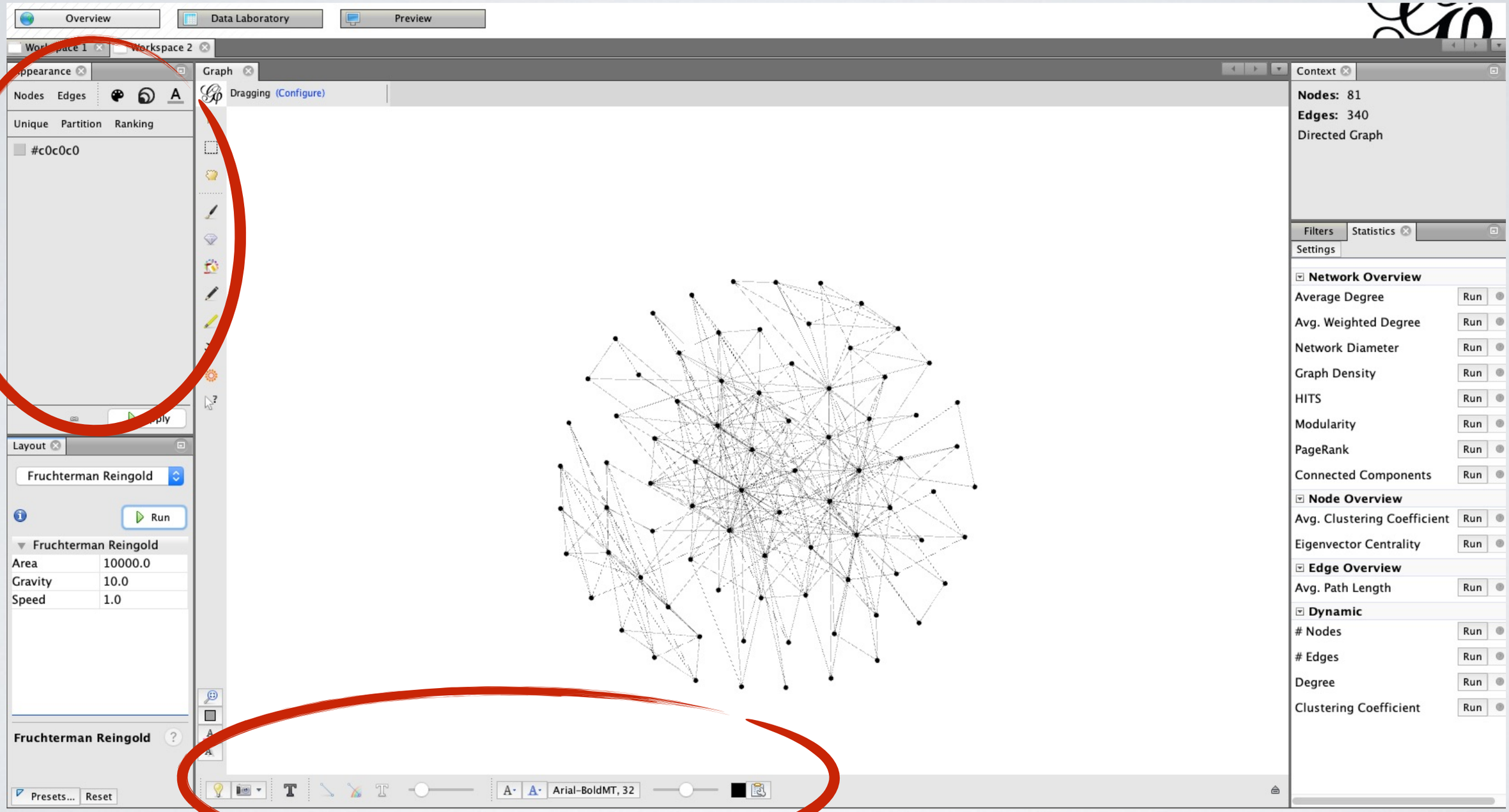
UCI NETWORK DATA REPOSITORY - UCIRVINE)
HTTPS://NETWORKDATA.ICS.UCI.EDU/INDEX.HTML

HTTP://KONECT.UNI-KOBLENZ.DE

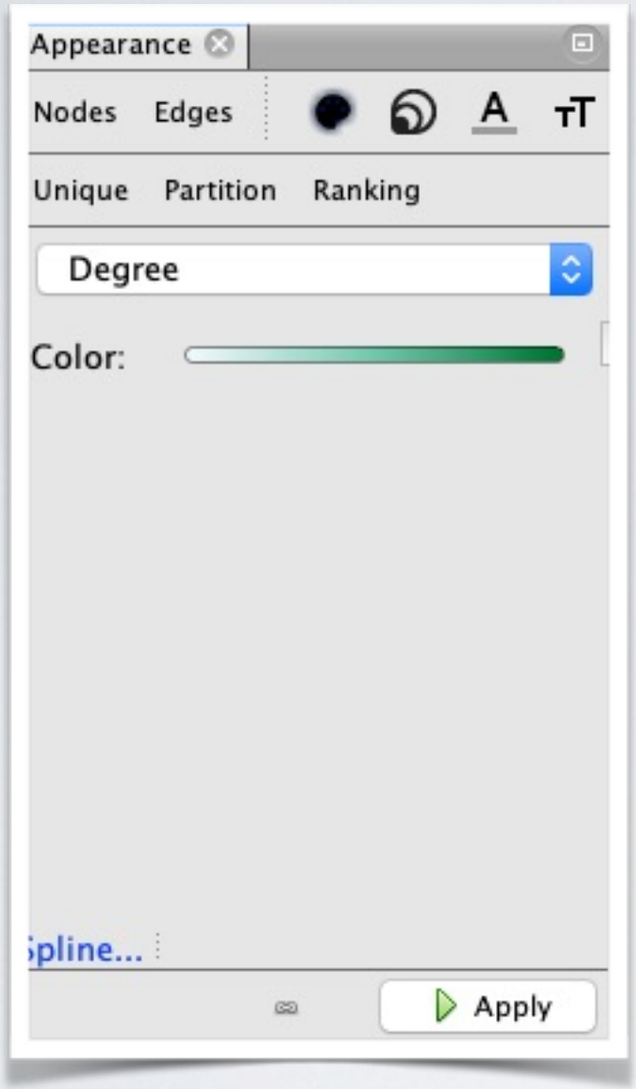
...

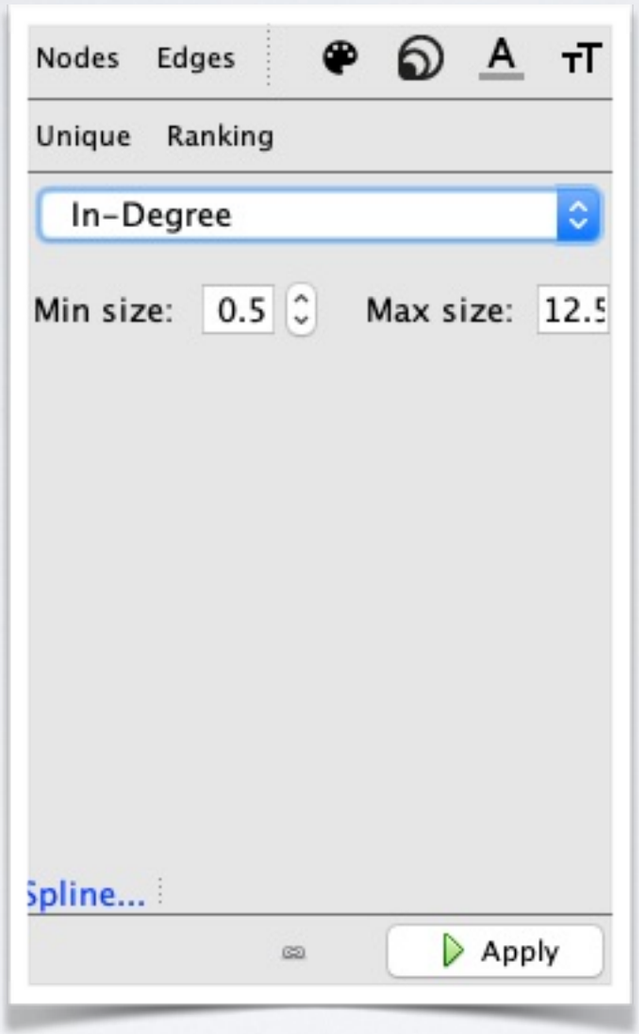
NODE DESIGN

NODE DESIGN


Overview
Data Laboratory
Preview
Workspace 2
Graph
Dragging (Configure)
Nodes Edges
Unique Partition Ranking
#c0c0c0
Layout
Fruchterman Reingold
Run
Fruchterman Reingold
Area 10000.0
Gravity 10.0
Speed 1.0
Fruchterman Reingold
Presets...
Reset
Arial-BoldMT, 32
Context
Nodes: 81
Edges: 340
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity Run
PageRank Run
Connected Components Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Nodes Run
Edges Run
Degree Run
Clustering Coefficient Run

COLORS AND SIZE





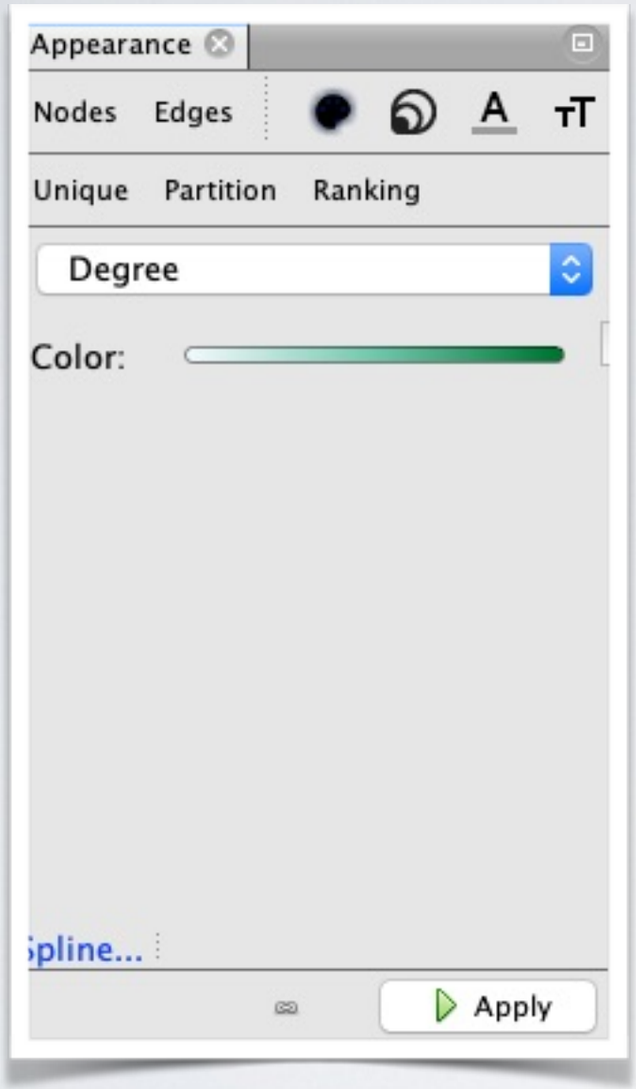
Node color, Node Size
Label color, Label Size

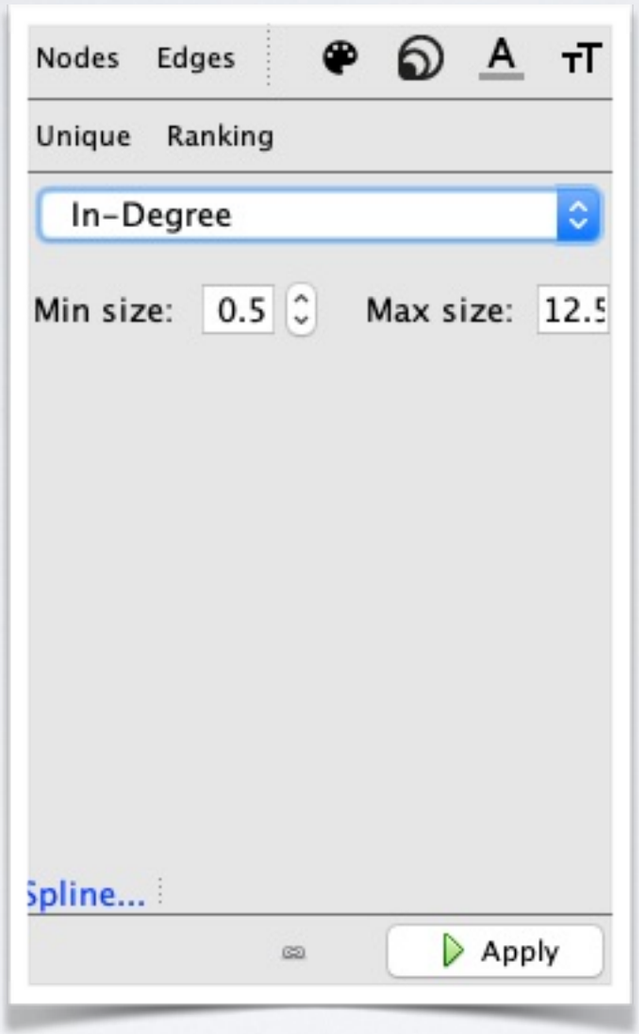
Same for all or based on
attribute
(Compute attribute first)

<u>Spline Example:</u>
Country, Size = GDP
China: 13 000 B - Madagascard: 13 B
Node 1000 times bigger

Spline => Log
log(10 000) = 3 - log(10)=1
3 times bigger

COLORS AND SIZE





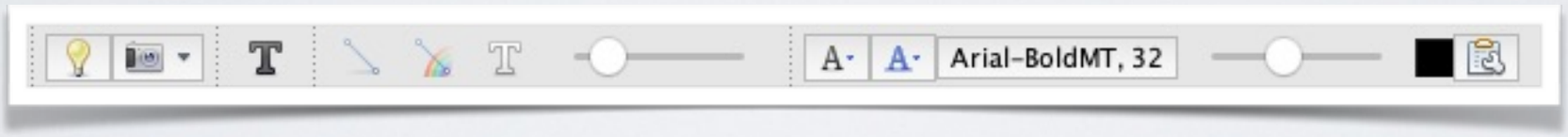
Node color, Node Size
Label color, Label Size

Same for all or based on
attribute
(Compute attribute first)

<u>Spline Example:</u>
Country, Size = GDP
China: 13 000 B - Madagascard: 13 B
Node 1000 times bigger

Spline => Log
log(10 000) = 3 - log(10)=1
3 times bigger

COLORS AND SIZE



Hide/show edges

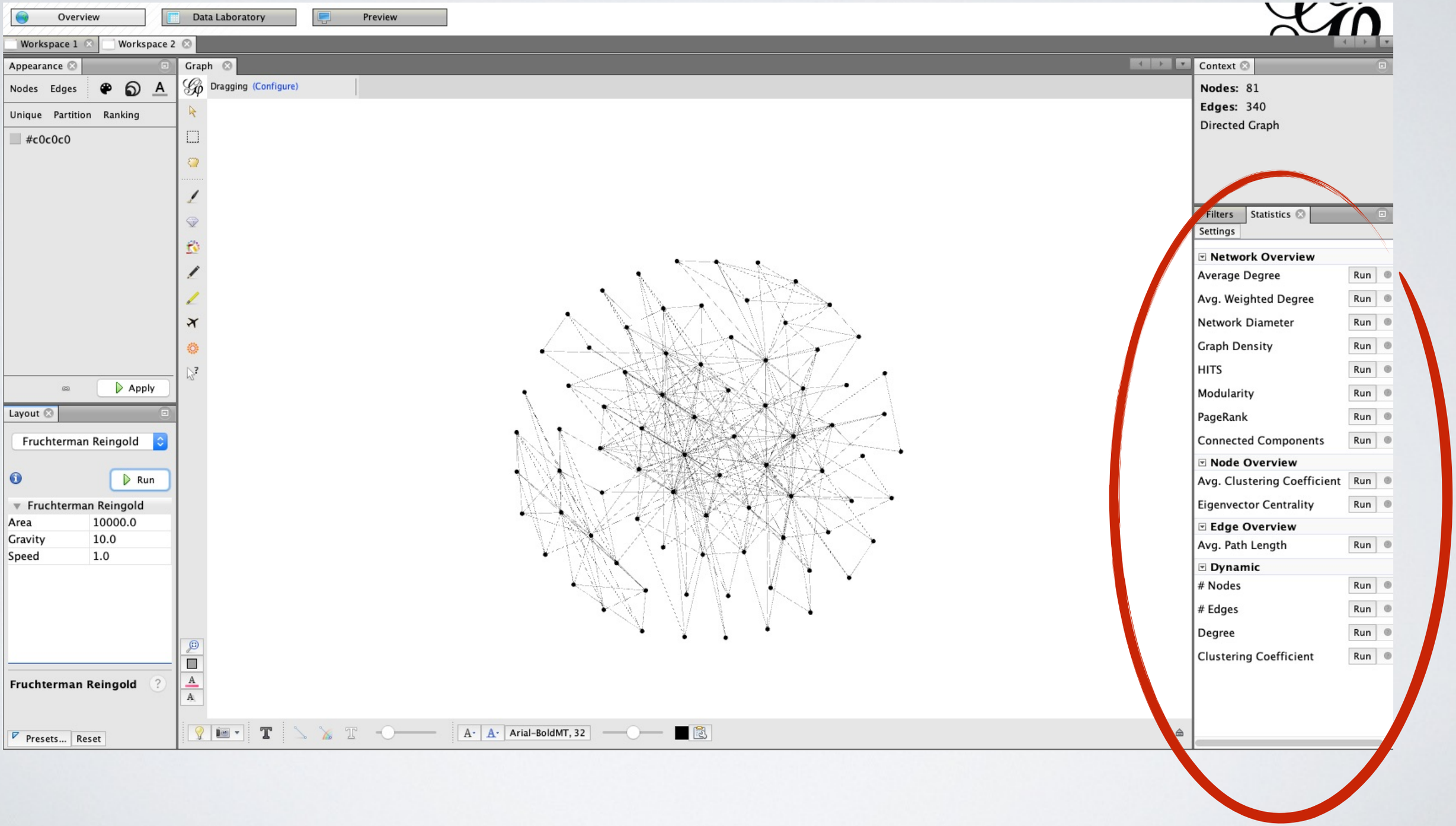
Simple tuning

Let your mouse on a button to see its role

Tip: *Size Mode*:

Text size proportional to node size

COMPUTE STATISTICS


Overview
Data Laboratory
Preview
Workspace 1
Workspace 2
Appearance
Nodes
Edges
Unique
Partition
Ranking
#c0c0c0
Apply
Layout
Fruchterman Reingold
Run
Fruchterman Reingold
Area 10000.0
Gravity 10.0
Speed 1.0
Fruchterman Reingold
Presets...
Reset
Graph
Dragging (Configure)
Arial-BoldMT, 32
Context
Nodes: 81
Edges: 340
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity Run
PageRank Run
Connected Components Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Nodes Run
Edges Run
Degree Run
Clustering Coefficient Run

Running statistics:
=>

COMPUTE STATISTICS

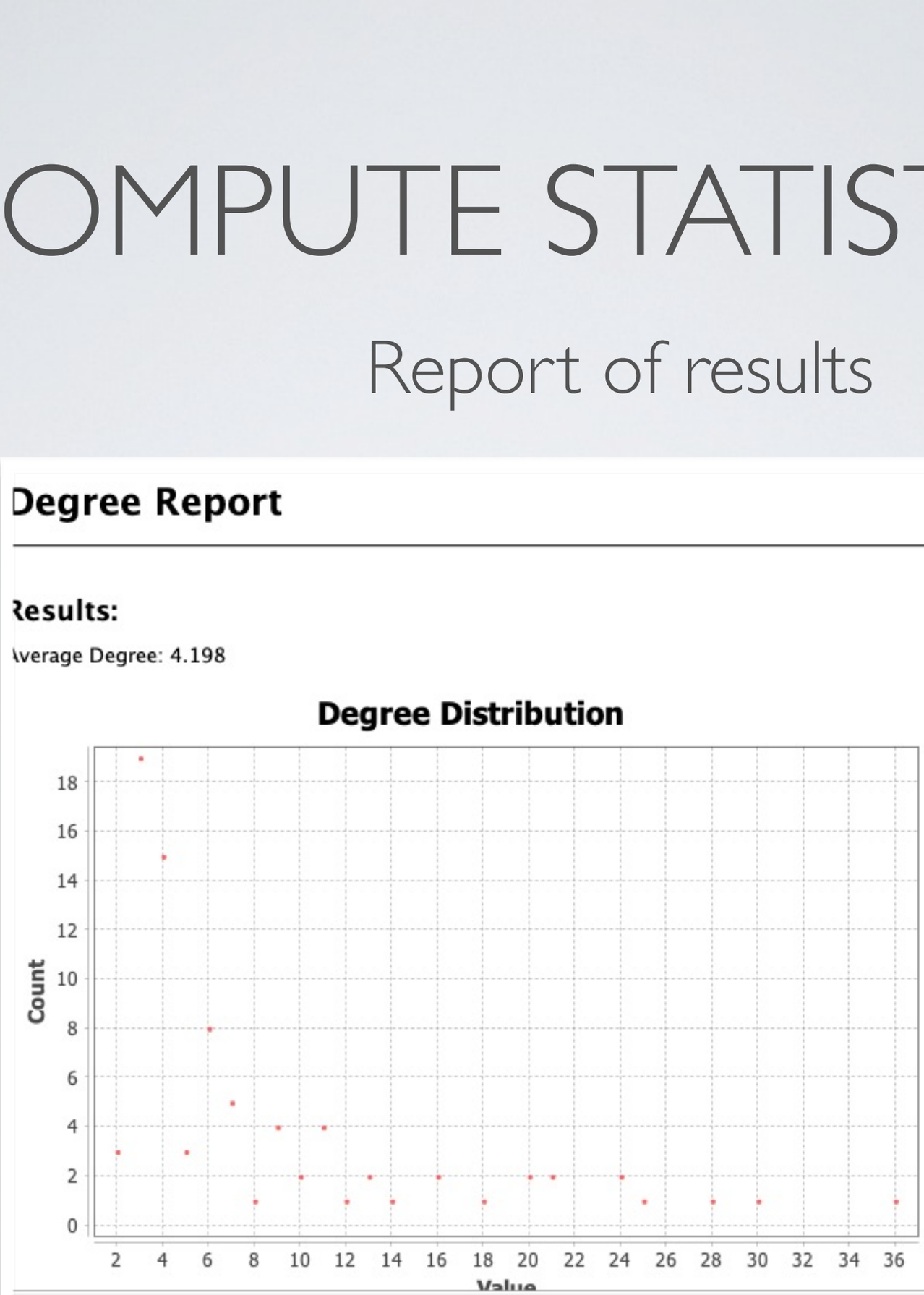
Report of results

Degree Report

Results:

Average Degree: 4.198

Degree Distribution



COMPUTE STATISTICS

Values available in Data Lab =>Export in Spreadsheet

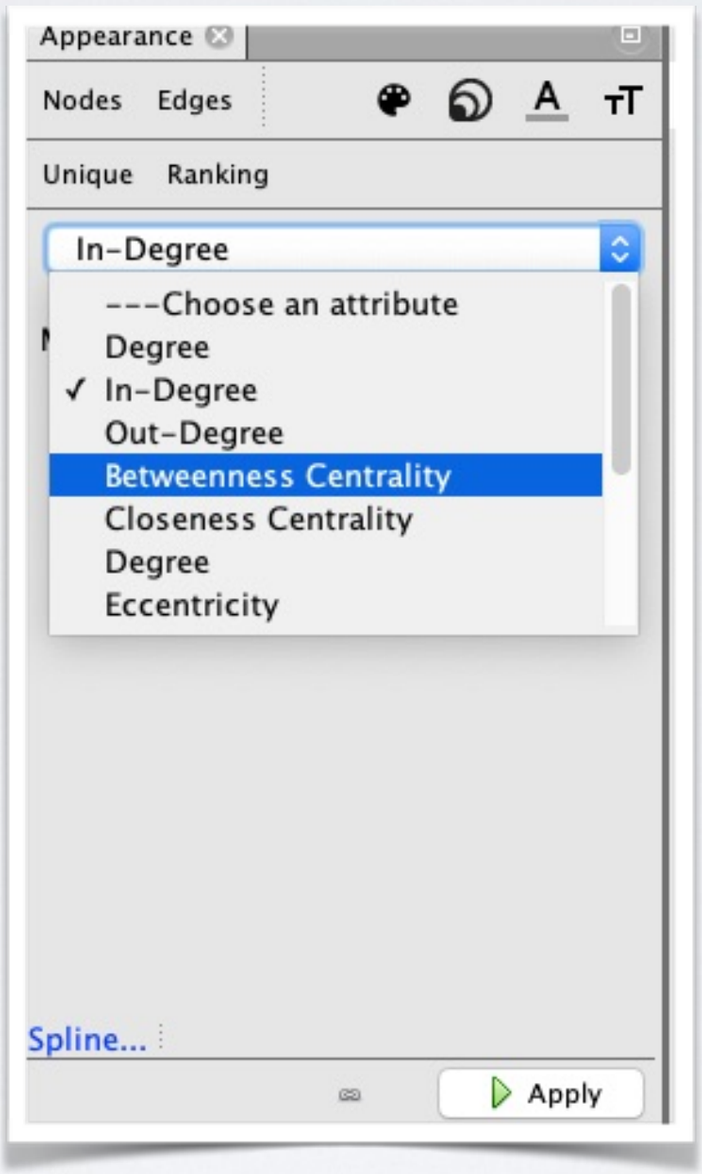
Data Table

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions

Id	Label	Betweenness Centrality	In-Degree	Degree	Closeness Centrality	Weighted In-Degree
Jaime-Lannister	Jaime-Lannister	0.0	0	14	0.528926	0.0
Aerys-II-Targaryen	Aerys-II-Targaryen	0.0	2	3	0.393939	2.0
Brynden-Tully	Brynden-Tully	14.166667	4	6	1.0	4.0
Catelyn-Stark	Catelyn-Stark	117.513095	9	28	0.653846	9.0
Cersei-Lannister	Cersei-Lannister	64.24881	10	21	0.469697	10.0
Eddard-Stark	Eddard-Stark	75.746429	3	36	0.681319	3.0
Joffrey-Baratheon	Joffrey-Baratheon	57.105952	7	20	0.612245	7.0
Jon-Arryn	Jon-Arryn	0.935714	2	11	0.509804	2.0
Lysa-Arryn	Lysa-Arryn	0.0	9	9	0.0	9.0
Petyr-Baelish	Petyr-Baelish	21.60119	9	16	0.5	9.0
Robb-Stark	Robb-Stark	122.691667	13	24	0.605263	13.0
Robert-Baratheon	Robert-Baratheon	180.70119	8	30	0.6375	8.0
Sandor-Clegane	Sandor-Clegane	6.753571	8	11	0.6	8.0
Tyrion-Lannister	Tyrion-Lannister	100.740476	8	25	0.634921	8.0
Tywin-Lannister	Tywin-Lannister	0.583333	1	9	0.496063	1.0
Gregor-Clegane	Gregor-Clegane	2.083333	6	7	1.0	6.0
Walder-Frey	Walder-Frey	0.0	3	3	0.0	3.0
Daenerys-Targaryen	Daenerys-Targaryen	0.0	0	16	0.468354	0.0
Aggo	Aggo	0.0	1	3	0.541667	1.0
Cohollo	Cohollo	0.0	2	4	0.75	2.0
Doreah	Doreah	0.0	2	4	0.75	2.0
Drogo	Drogo	8.0	2	13	0.923077	2.0
Haggo	Haggo	0.0	3	4	0.666667	3.0
Illyrio-Mopatis	Illyrio-Mopatis	0.0	4	5	1.0	4.0
Irri	Irri	2.5	4	6	1.0	4.0
Jhiqui	Jhiqui	0.0	4	4	0.0	4.0
Jhogo	Jhogo	0.0	2	4	0.571429	2.0
Jorah-Mormont	Jorah-Mormont	8.833333	4	9	1.0	4.0
Mirri-Maz-Duur	Mirri-Maz-Duur	0.0	4	4	0.0	4.0
Qotho	Qotho	2.0	5	6	1.0	5.0
Varys	Varys	4.333333	12	13	1.0	12.0
Viserys-Targaryen	Viserys-Targaryen	0.0	7	7	0.0	7.0
Alyn	Alyn	0.0	1	2	0.349593	1.0
Arya-Stark	Arya-Stark	102.235714	5	20	0.615385	5.0
Barristan-Selmy	Barristan-Selmy	0.25	6	8	0.75	6.0
Bran-Stark	Bran-Stark	53.302381	8	18	0.5625	8.0
Brandon-Stark	Brandon-Stark	0.0	1	3	0.403101	1.0
Ilyn-Payne	Ilyn-Payne	0.5	2	4	0.529412	2.0
Jon-Snow	Jon-Snow	129.207143	4	21	0.581633	4.0
Jory-Cassel	Jory-Cassel	50.285714	3	11	0.525	3.0
Loras-Tyrell	Loras-Tyrell	0.0	6	6	0.0	6.0
Luwin	Luwin	4.566667	8	10	0.5	8.0
Mordane	Mordane	3.25	5	7	0.478261	5.0

COMPUTE STATISTICS

Values available to set nodes/edges colors/size



COMPUTE STATISTICS

- Average Degree
  - => Distribution of degrees …
- Average Weighted Degree
- Network Diameter
  - =>Diameter: Longest shortest path
  - Compute for nodes:
    - Betweenness (#shortest paths)
    - Closeness (#avg distance)
    - Harmonic closeness/centrality (Closeness variant)
    - Eccentricity (Longest distance to another node)

COMPUTE STATISTICS

- HITS
  - => Hubs and authorities (Hubs = cited by authorities, authorities = citing hubs)

- Pagerank
  - => Hubs score by "vote"

- Modularity, Connected components
  - Next section, communities

COMPUTE STATISTICS

- Avg Clustering coefficient
  - =>friends of my friends are my friends
- Eigenvector Centrality
  - Variant of PageRank
- Avg. Path Length
  - => avg. between all pairs of nodes
- Dynamic
  - =>To see later

COMMUNITIES

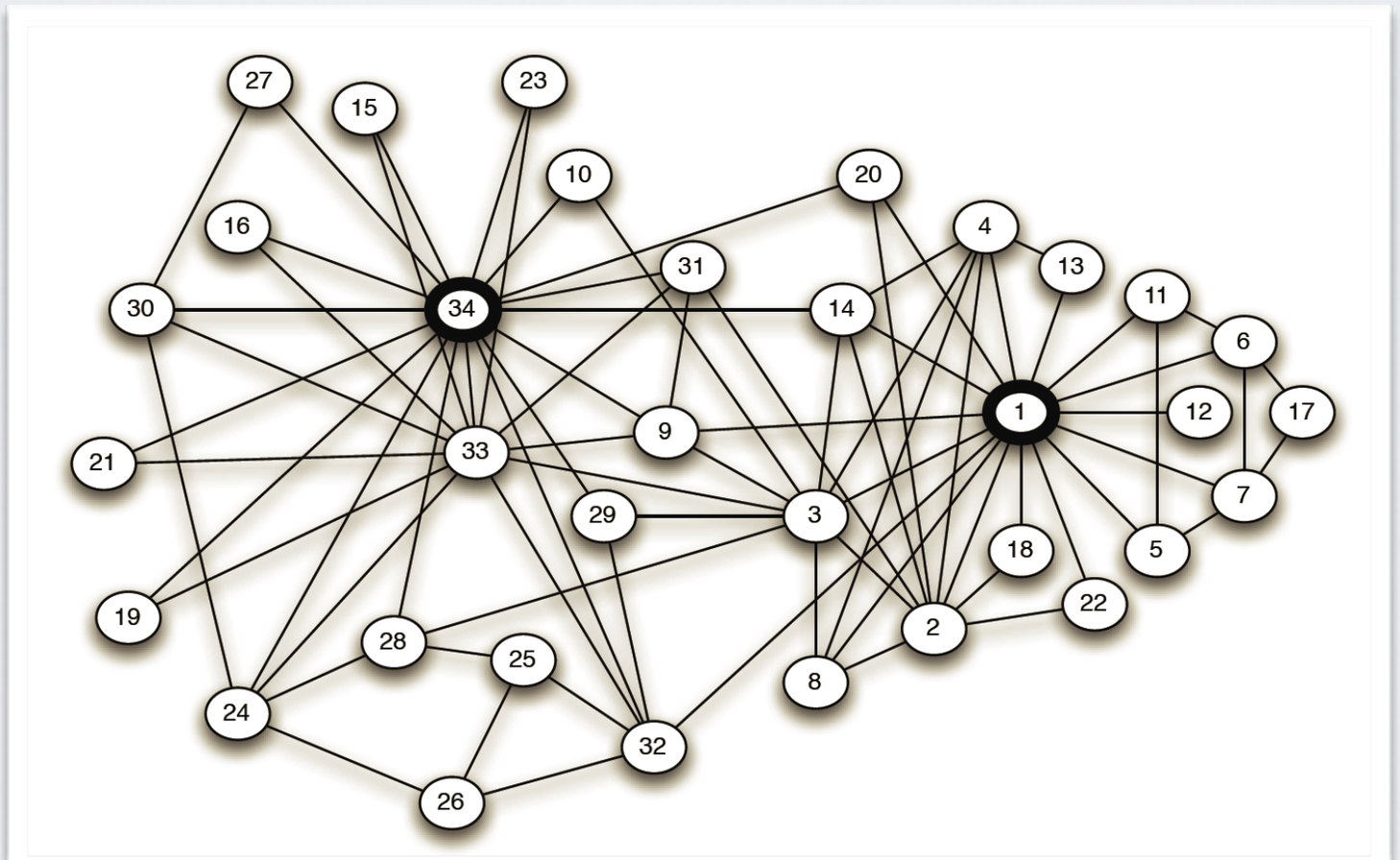
OR “GRAPH CLUSTERS”

COMMUNITY DETECTION

- One of the most active research field in network science
- 1000+ methods proposed
- No clear objective:
  - Discover groups of nodes strongly connected and weakly connected to the rest of the graph ?
  - Discover group of nodes that make sense ?
    - Core-periphery
    - Blocks (male/female in a sentimental network)
- My field of research :)

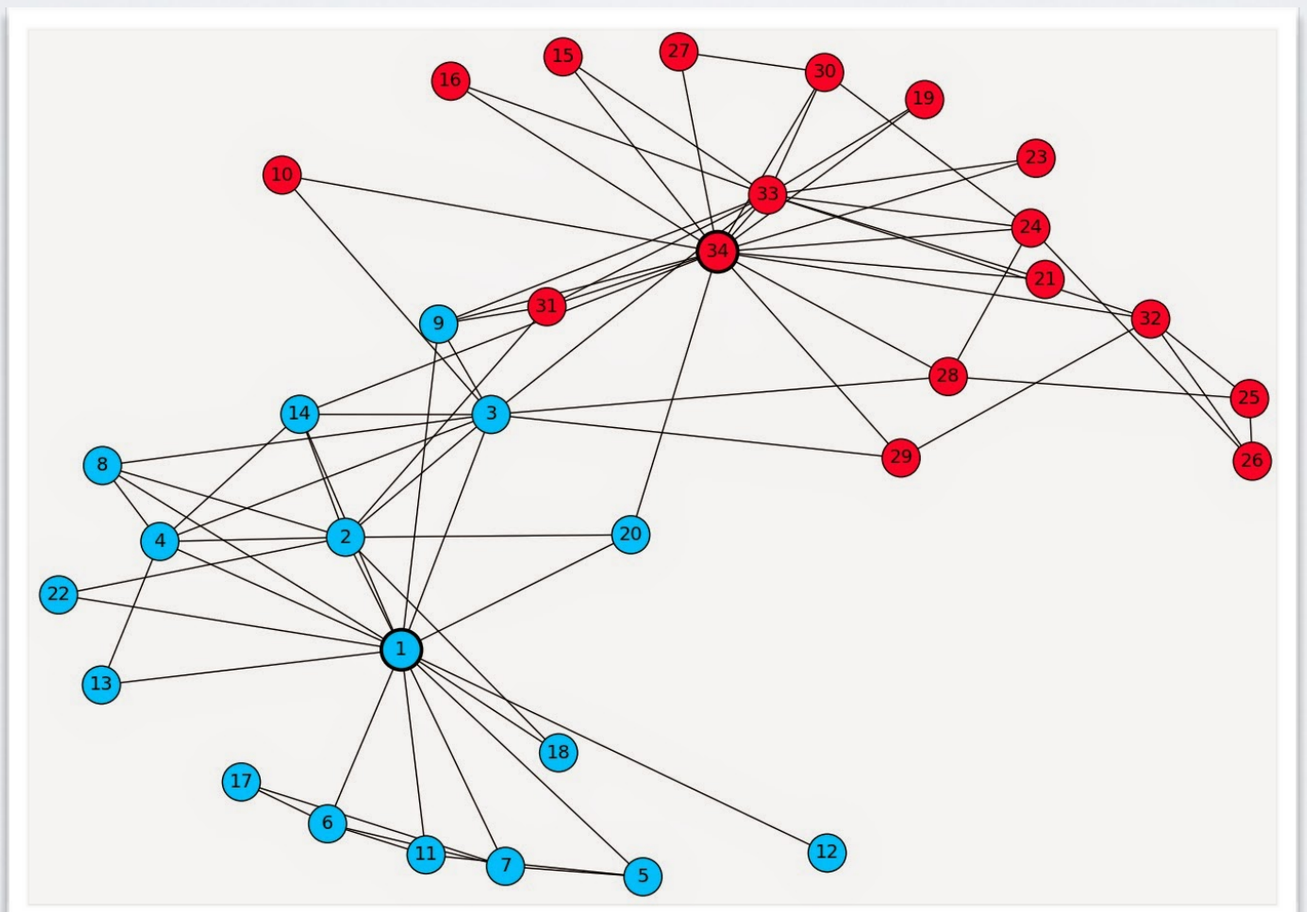
COMMUNITY DETECTION

- History: Karate Club Graph



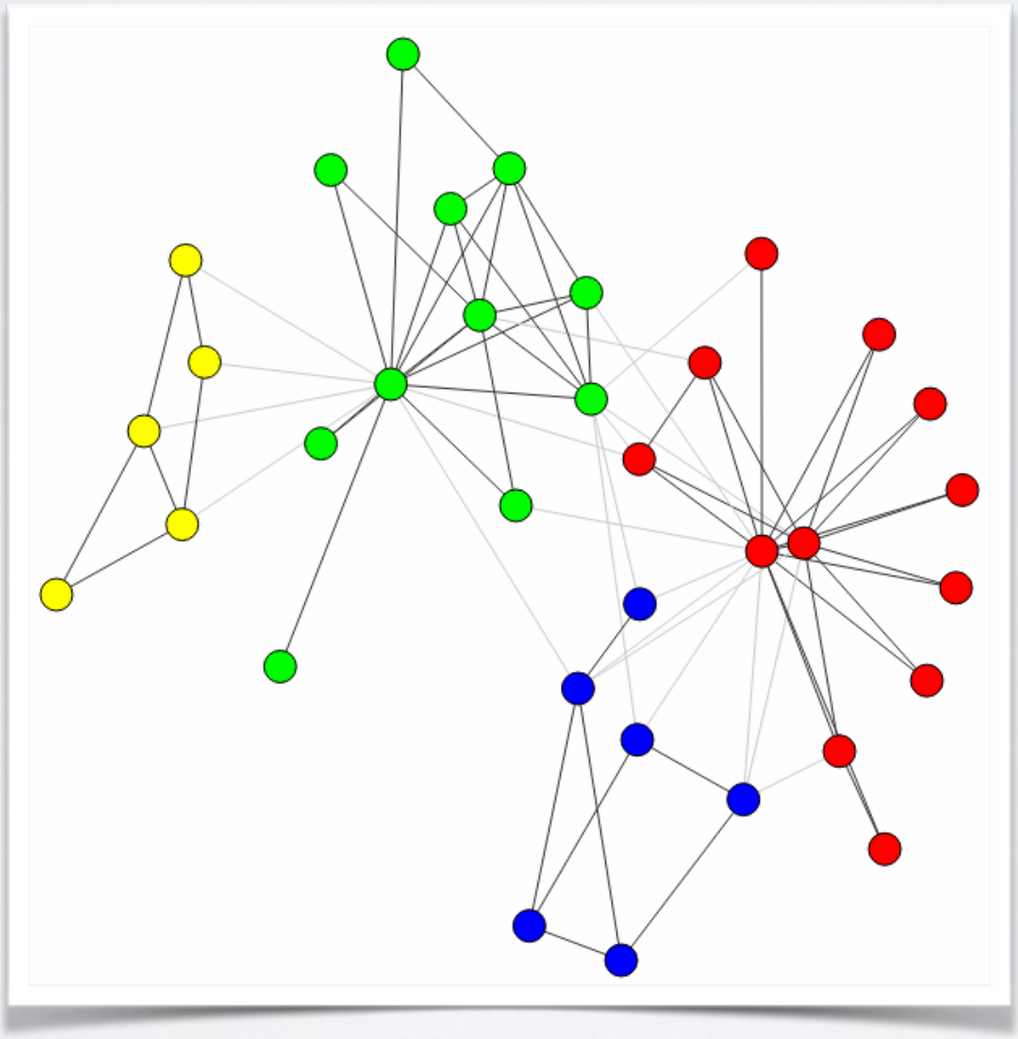
COMMUNITY DETECTION

- History: Karate Club Graph



COMMUNITY DETECTION

- History: Karate Club Graph



COMMUNITY DETECTION

- In Gephi: Louvain Algorithm
  - De facto standard (but might change in a near future…)

- Greedy optimization of modularity
  - Modularity =>
    Fraction of edges between communities in **the graph**
    - Fraction of edges between communities in a **randomized graph**
  - Greedy optimization => Fast but not exhaustive
    - Do not necessarily find the solution of **highest** modularity
    - Two runs can yield different results

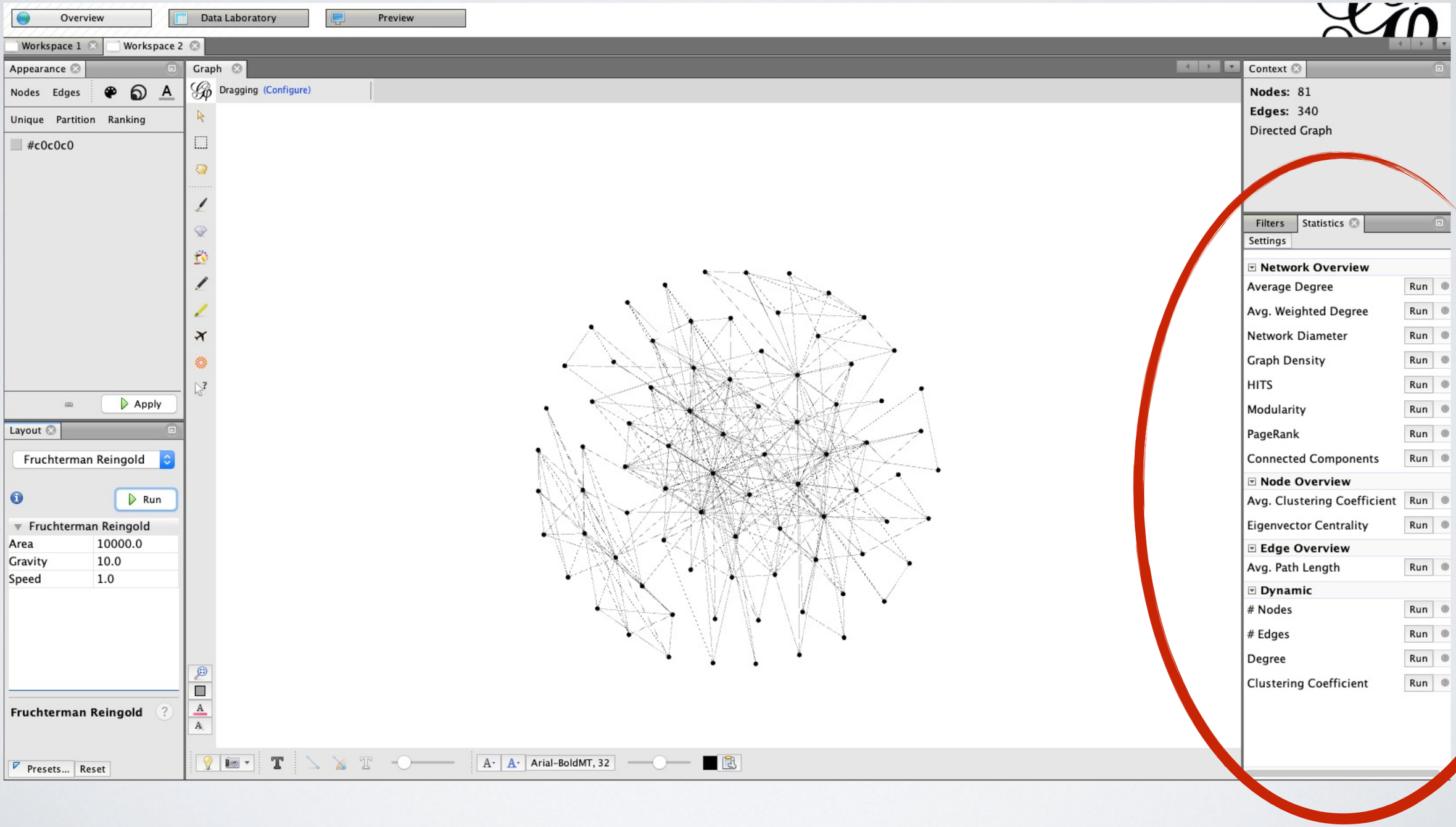
COMMUNITY DETECTION

- Most famous alternatives:
  - Stochastic block models:
    - Search from **blocks** that are **random sub-graphs**
    - Can allow to fix the number of clusters
  - Infomap:
    - Find partitions that allow to **compress** the graph : Occam razor
  - Leiden algorithm:
    - Improvement of Louvain with guarantees on community properties (connectedness..)

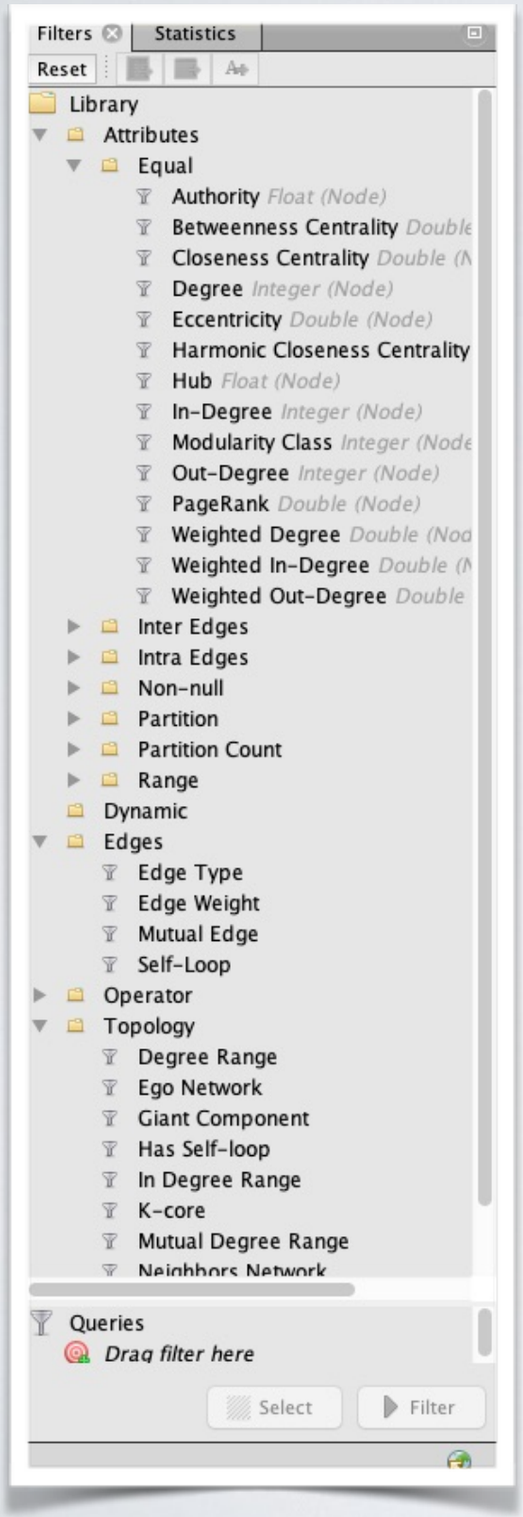
- Famous drawback of all methods: The resolution limit
  - In summary: methods cannot find small communities in large graphs

FILTERS

FILTERS


Overview
Data Laboratory
Preview
Workspace 1
Workspace 2
Appearance
Nodes
Edges
Unique
Partition
Ranking
#c0c0c0
Apply
Layout
Fruchterman Reingold
Run
Fruchterman Reingold
Area 10000.0
Gravity 10.0
Speed 1.0
Fruchterman Reingold
Presets...
Reset
Graph
Dragging (Configure)
Arial-BoldMT, 32
Context
Nodes: 81
Edges: 340
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity Run
PageRank Run
Connected Components Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Nodes Run
Edges Run
Degree Run
Clustering Coefficient Run

FILTERS



- Filter-out nodes/Edges
- Based on properties:
  - User defined:
    - Genre is : male
    - Age is: between 25 and 35
  - Computed:
    - Degree is : >10
    - PageRank is : between 0.2 and 0.5
  - Based on topology
    - Ego network at depth
    - Giant component
    - K-core (subgraph of internal degree at list k)
- Filtered graph can be exported

RANDOM TIPS

RANDOM TIPS

- From the overview, you can right-click on a node to see it in the data laboratory

- From the data lab, you can right-click on a node to see it in the overview

- The data lab has many features to discover:
  - Filtering nodes/edges,
  - Create nodes/edges
  - Group nodes
  - …

- I recommend to avoid using this.
  - 1)Build your data in excel/text file
  - 2)Compute statistics and visualize with Gephi

SAVING

- Computed statistics:
  - From the data laboratory => Export tables
    - Can be re-load with table import

- The graph topology:
  - File/Export (choose a file format)
  - With all information (colors, node location…) => Choose GEXF format

- The complete workspace (filters…)
  - File/Save

- Graph as a picture
  - File/Export/"PDF/SVG …"
    - PDF & SVG : vectors => infinite zoom, small weight
    - PNG: pixels by pixel, weight depends on size

SPECIAL COLUMNS

- Some columns in Gephi are "special", they always exist:
  - Nodes:
    - ID
    - Label <-Column used to display the label
    - Interval
  - Edges
    - Source
    - Target
    - Type <-Directed/Undirected
    - Id
    - Label <-Column used to display the label
    - Interval
    - Weight <-Column used for the width of edges

Tools to fill those columns from other ones

DYNAMICS

DYNAMIC GRAPHS

- Nodes and edges have a special column for dynamic
- Flexible format:
  - <[start1,start2);(start3,start4]>
  - [and] means "limit included"
  - (and) means "limit excluded"
    - [3,3] => only 3
    - [3,3)=>impossible
  - Accept dates YYYY/MM/DD/MM:SS, timestamps…

DYNAMIC GRAPHS

1) Your spreadsheet file

Source	Target	Label	Weight	somethin_e	Interval
a	b	link 1	100	5	<[2005,2006)>
c	d	another link	150	10	<[2005,2005];[2006,2006]>
a	e	why not ?	50	3	<[2003,2007]>

2)After import

Interval	Weight	something_else	interval
	1.	5	<[2005,2006)>
	1.	10	<[2005,2005];[2006,2006]>
	1.0	3	<[2003,2007]>

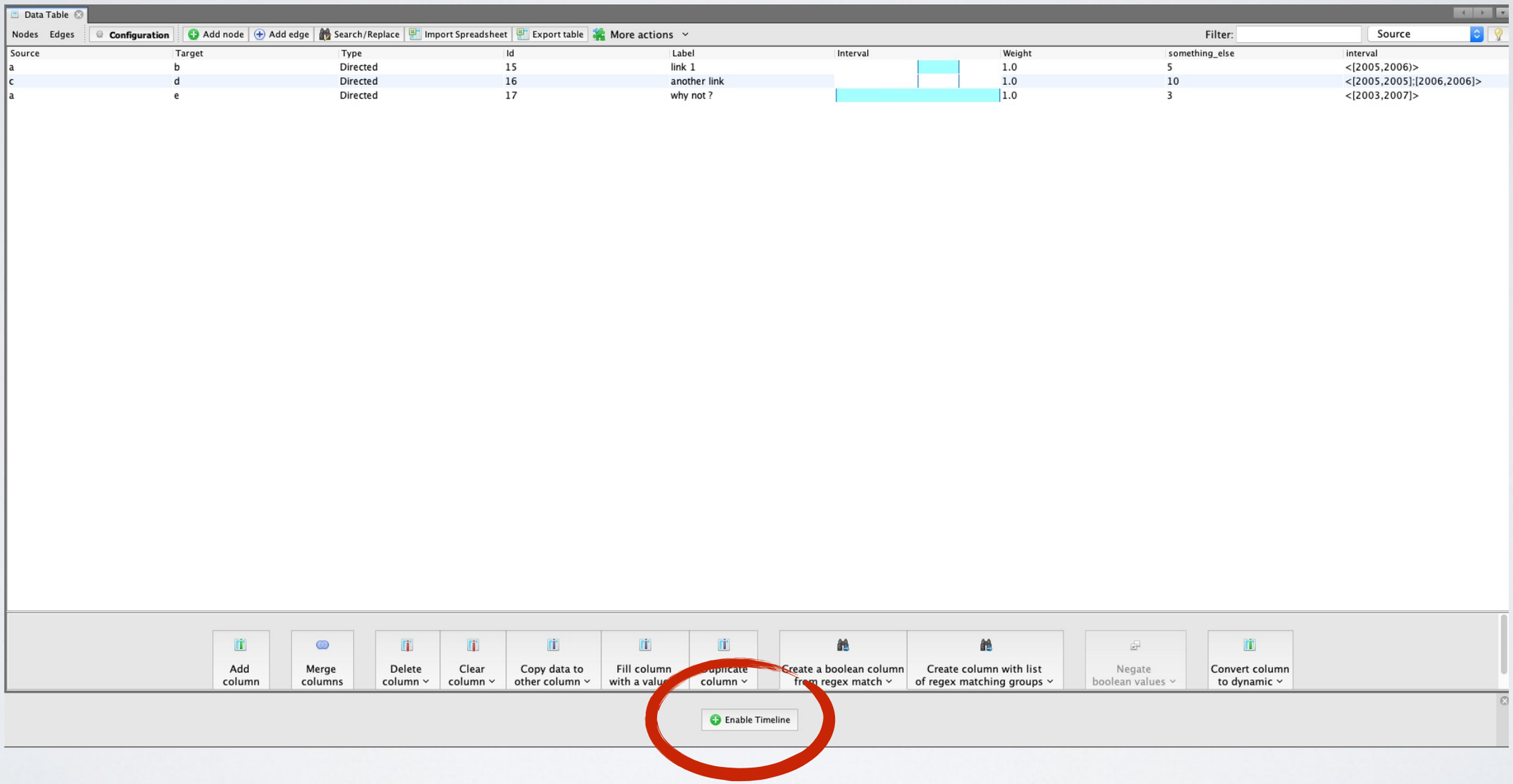
3)Copy data to other column +(Data Lab/Configuration/time Intervals as graphic)

Data Table

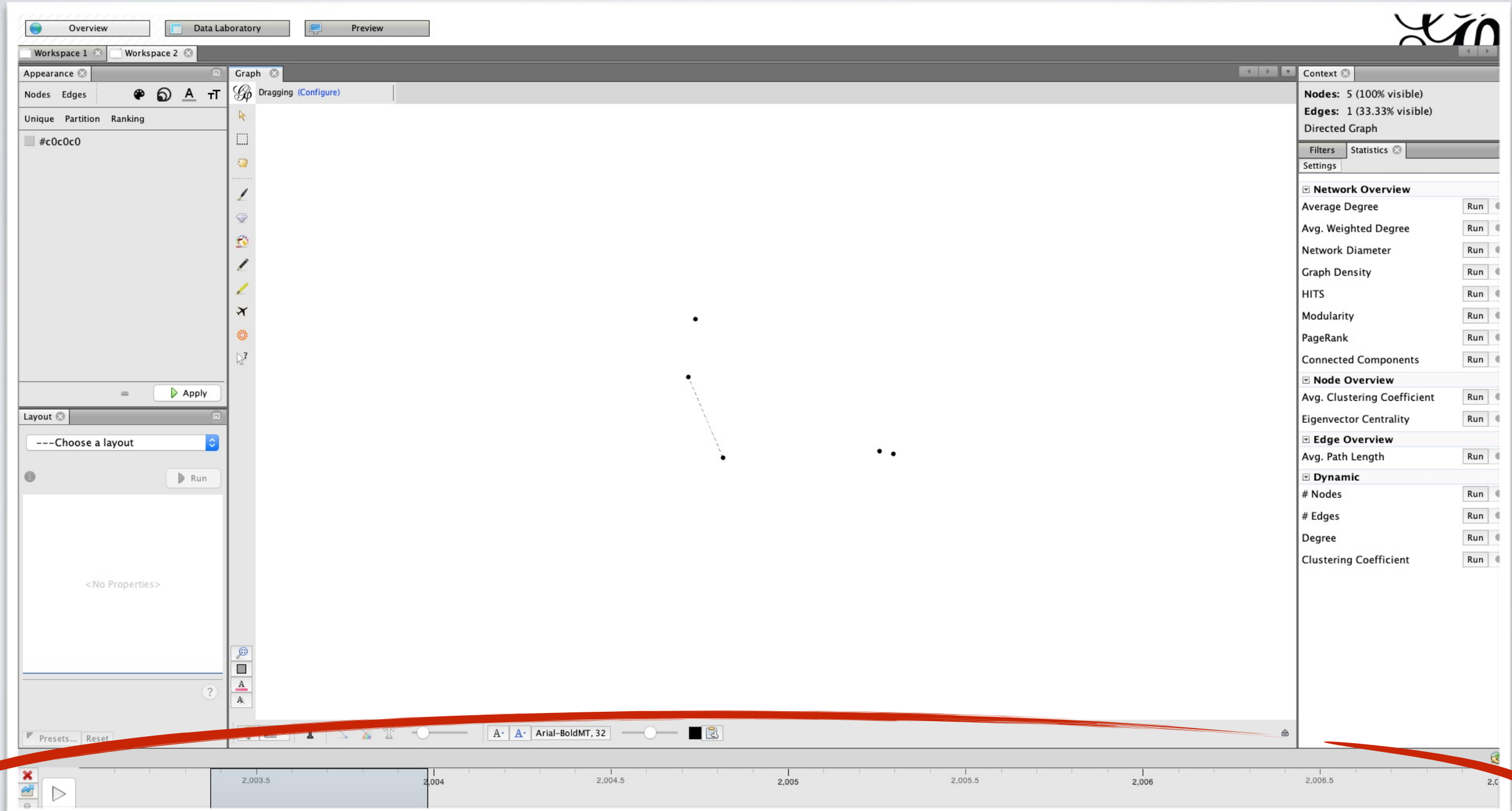
Nodes Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | Mo

Source	Target	Type	Id	Label	Interval	Weight
a	b	Undirected	12	link 1		1.0
c	d	Undirected	13	another link		1.0
a	e	Undirected	14	why not ?		1.0

DYNAMIC GRAPHS



DYNAMIC GRAPHS



PLUGINS

PLUGINS

- https://gephi.org/plugins/#/browse/type/layout
  - Or just search "Gephi plugins" in Google
- Any programer can write and propose plugins, i.e. extensions for gephi
- They add some missing features

PLUGINS

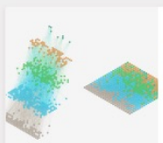
Layout

GeoLayout
3 months ago
A layout to display geocoded data

Network Splitter 3D
3 months ago
This layout can be used to split a network layout into distinct Z-Layers. E.g. after

Graphviz Layout
3 months ago
Graphiz Layout Plugin

Circular Layout
3 months ago
This plugin provides three circular layouts; one simple circle (titled "Circular Layout",



Isometric Layout
3 months ago
This layout uses an isometric perspective to visualize networks.

EventGraphLayout
3 months ago
Layout algorithm for event graphs

MdsLayout
3 months ago
This plugin takes a set of coordinates on 2 dimensions and creates a layout based on

MultiGravity ForceAtlas 2
3 months ago
Force Atlas 2 Modification with multiple sources of

Map Of Countries
3 months ago
A layout to display maps in the backgrounds of networks

Polygon Shaped Nodes
3 months ago
Allows node shapes to be changed to any polygon

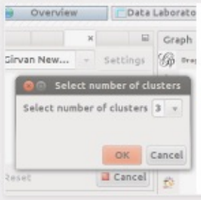
Clustering - Communities

Newman-Girvan Clustering
3 months ago
Girvan-Newman clustering algorithm

Leiden Algorithm
3 months ago
The Leiden algorithm finds well-connected communities in large scale networks.

Chinese Whispers Clustering
4 years ago

Label Propagation Clustering
5 years ago



Girvan Newman Clustering
5 years ago

Markov Cluster Algorithm (MCL)
5 years ago

Molecular Complex Detection (MCODE) Clustering
5 years ago

Preview

Image Preview
3 months ago
Image Preview

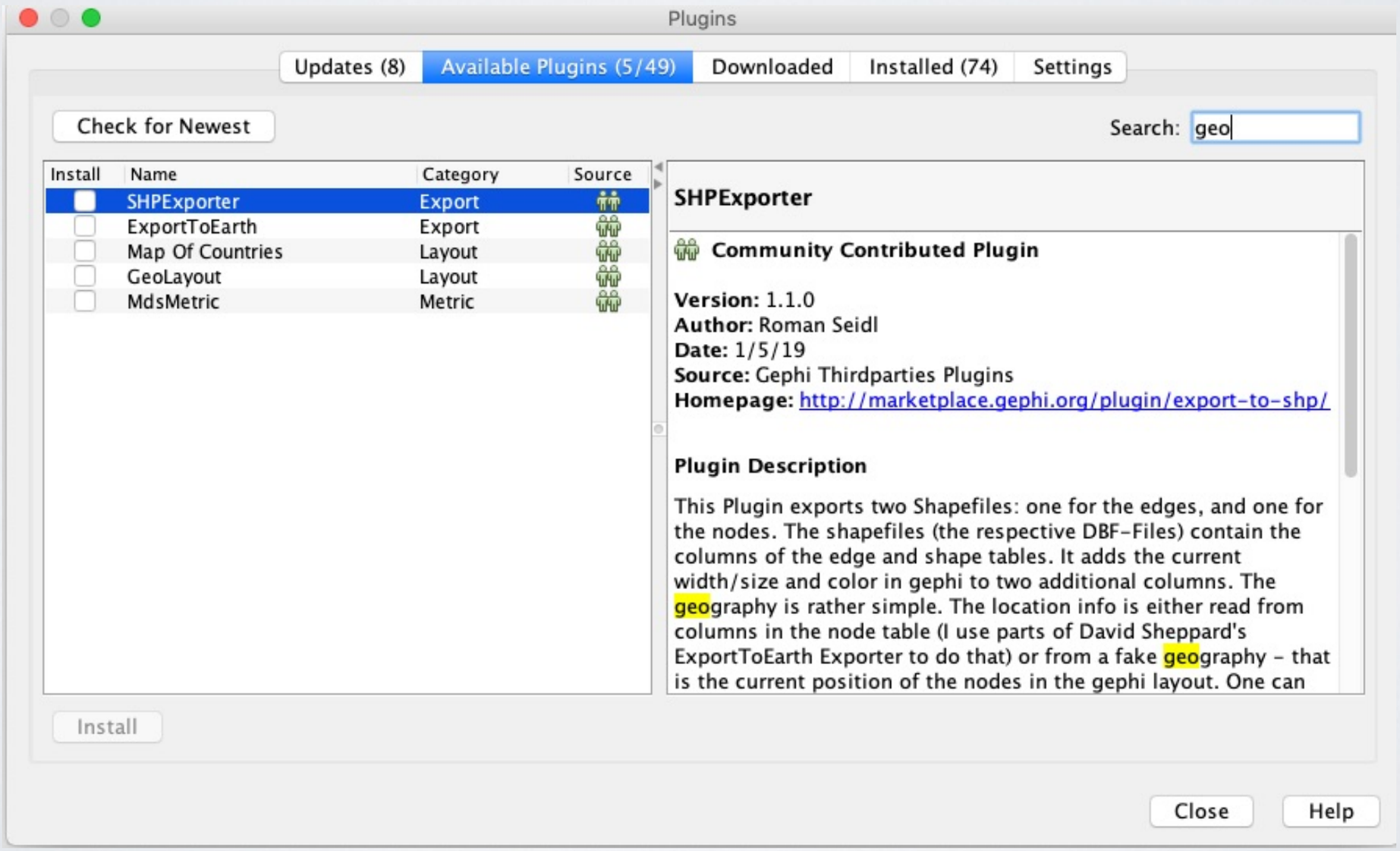
Polygon Nodes
4 years ago

Image Preview
5 years ago

Multi Colour Renderer
5 years ago

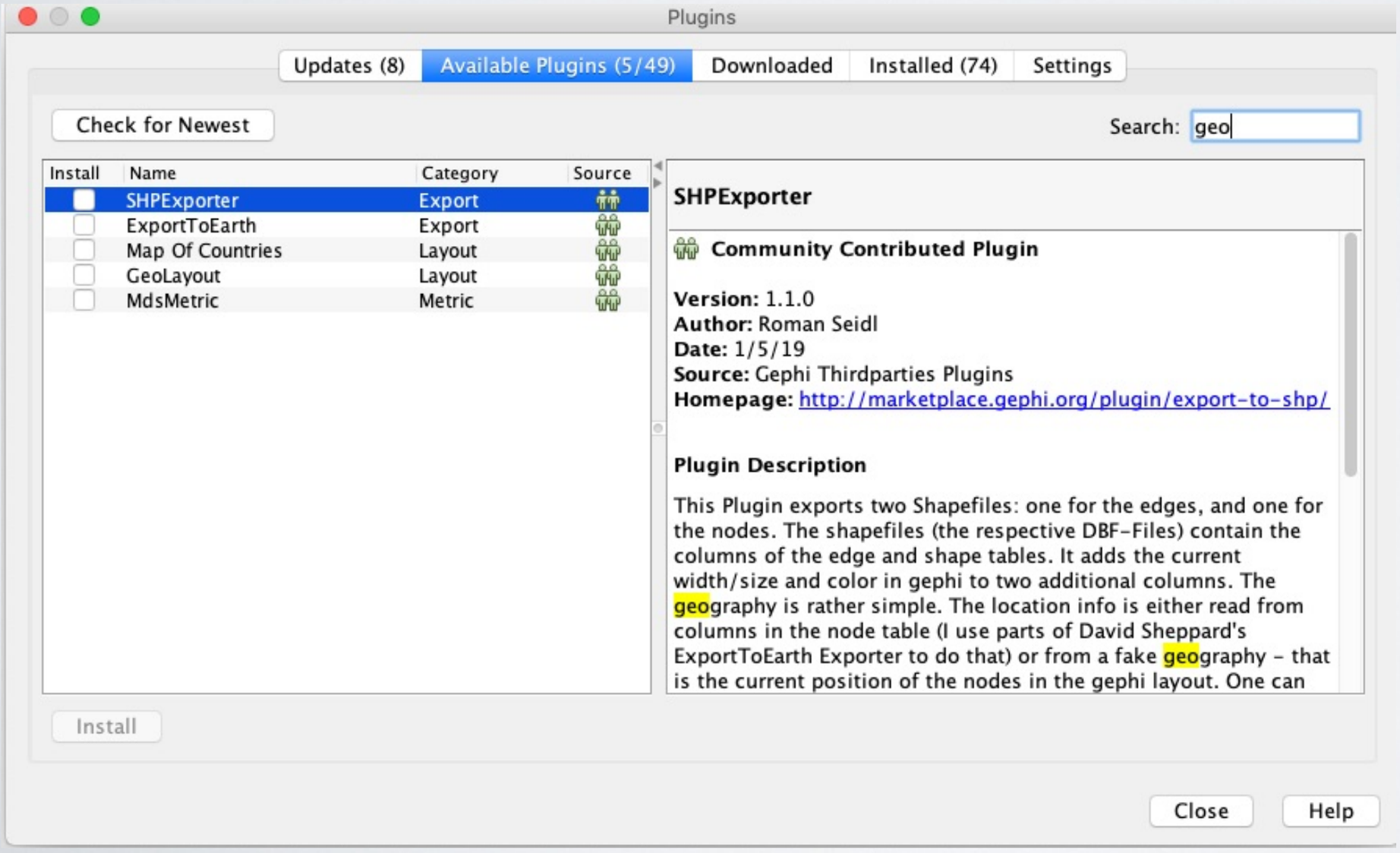
PLUGINS

- One useful example: spatial data
- Menu Tools/plugins



PLUGINS

- One useful example: spatial data
- Menu Tools/plugins



Install & Restart

PLUGINS

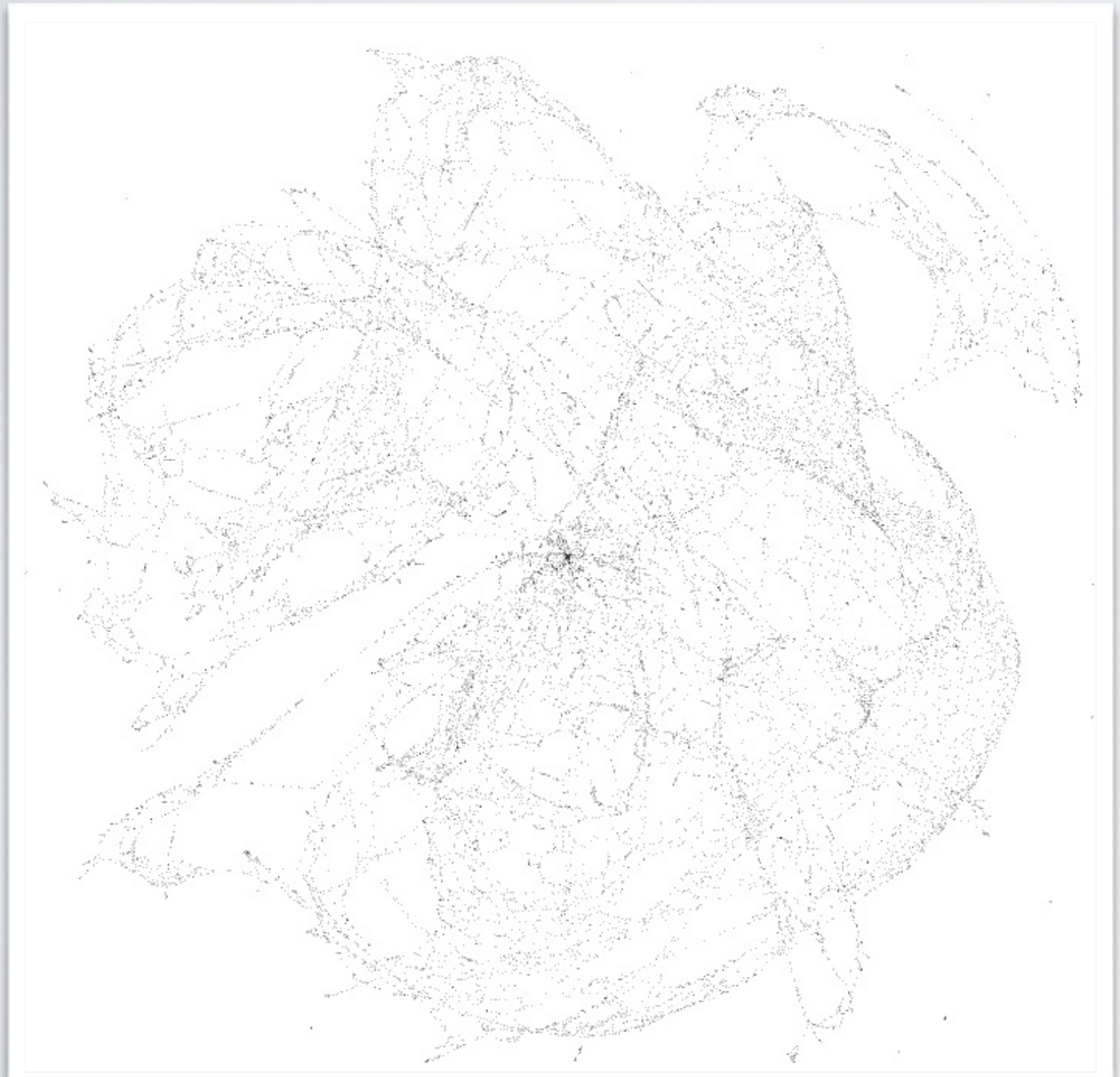
Data Table

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions

Id	Label	Interval	long	lat
94,007	94,007		545093.002768	5745916.976144
93,988	93,988		545080.017439	5745907.367803
94,005	94,005		545089.356087	5745896.358694
94,024	94,024		545100.929711	5745906.86376
83,221	83,221		539818.817157	5735003.783806
83,406	83,406		539912.430728	5735016.087779
83,394	83,394		539907.25143	5735039.474208
83,203	83,203		539813.235245	5735028.328475
92,962	92,962		544523.317524	5745737.0094
92,989	92,989		544539.021366	5745710.463221
85,691	85,691		541083.14309	5746704.063996
85,661	85,661		541079.257419	5746764.407639
73,356	73,356		536116.130192	5737095.306725
73,302	73,302		536096.368326	5737101.364769
73,331	73,331		536107.575995	5737115.613899
60,835	60,835		530908.290834	5738048.825509
60,840	60,840		530911.798513	5738064.835963
73,267	73,267		536082.994521	5737086.303785
73,250	73,250		536075.445215	5737106.73136
90,854	90,854		543384.332791	5747176.386853
90,752	90,752		543316.320898	5747227.769978
82,894	82,894		539722.990082	5738190.983088
83,043	83,043		539800.737978	5738239.868615
96,382	96,382		546311.861193	5740625.882888
96,333	96,333		546281.270276	5740485.653805
65,180	65,180		532912.207712	5743655.991137
65,242	65,242		532940.126881	5743604.327301
64,522	64,522		532768.614987	5749057.549238
64,549	64,549		532779.12833	5749078.741838

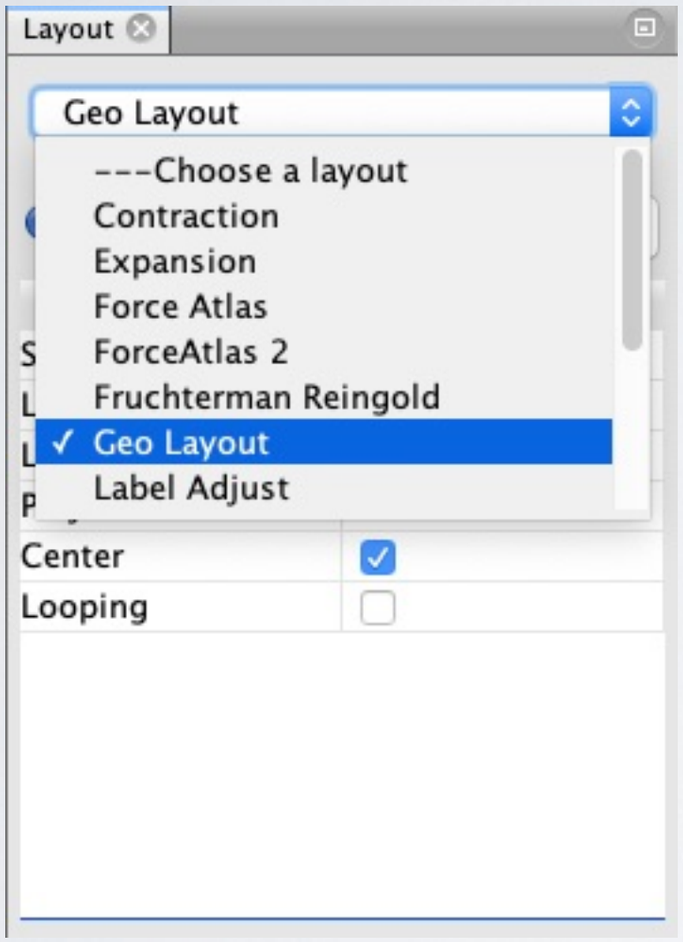
Dataset with columns for position

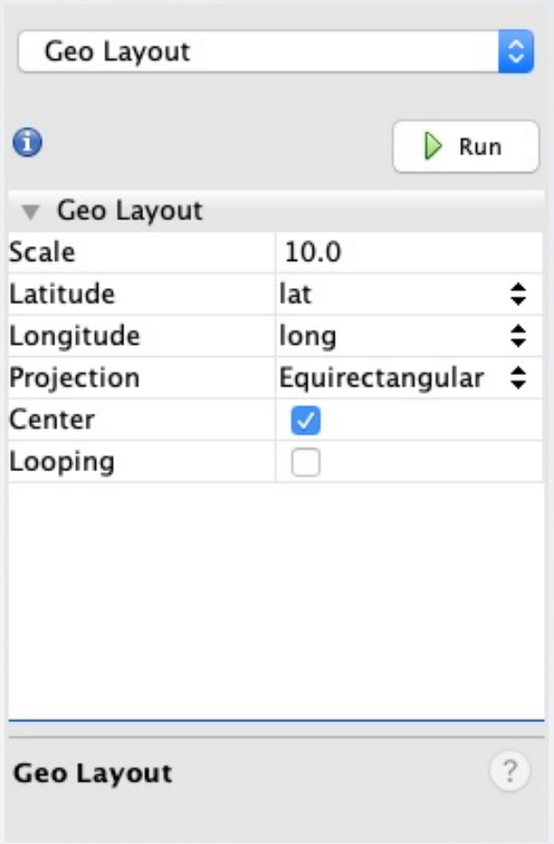
PLUGINS



Using graph layout

PLUGINS


Layout
Geo Layout
---Choose a layout
Contraction
Expansion
Force Atlas
ForceAtlas 2
Fruchterman Reingold
✓ Geo Layout
Label Adjust
Center
Looping


Geo Layout
Run
Geo Layout
Scale
10.0
Latitude
lat
Longitude
long
Projection
Equirectangular
Center
Looping
Geo Layout

PLUGINS