Tamkang University
淡江大學

Tamkang University

Social Computing and Big Data Analytics

社群運算與大數據分析

Measurements and Tools of Social Network Analysis

(社會網絡分析量測與工具)

1042SCBDA13

MIS MBA (M2226) (8628)

Wed, 8,9, (15:10-17:00) (B309)

Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

http://mail. tku.edu.tw/myday/

2016-05-25

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2016/02/17	Course Orientation for Social Computing and Big Data Analytics (社群運算與大數據分析課程介紹)
2	2016/02/24	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data (資料科學與大數據分析： 探索、分析、視覺化與呈現資料)
3	2016/03/02	Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem (大數據基礎：MapReduce典範、 Hadoop與Spark生態系統)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
4	2016/03/09	Big Data Processing Platforms with SMACK: Spark, Mesos, Akka, Cassandra and Kafka (大數據處理平台SMACK：Spark, Mesos, Akka, Cassandra, Kafka)
5	2016/03/16	Big Data Analytics with Numpy in Python (Python Numpy 大數據分析)
6	2016/03/23	Finance Big Data Analytics with Pandas in Python (Python Pandas 財務大數據分析)
7	2016/03/30	Text Mining Techniques and Natural Language Processing (文字探勘分析技術與自然語言處理)
8	2016/04/06	Off-campus study (教學行政觀摩日)

課程大綱 (Syllabus)

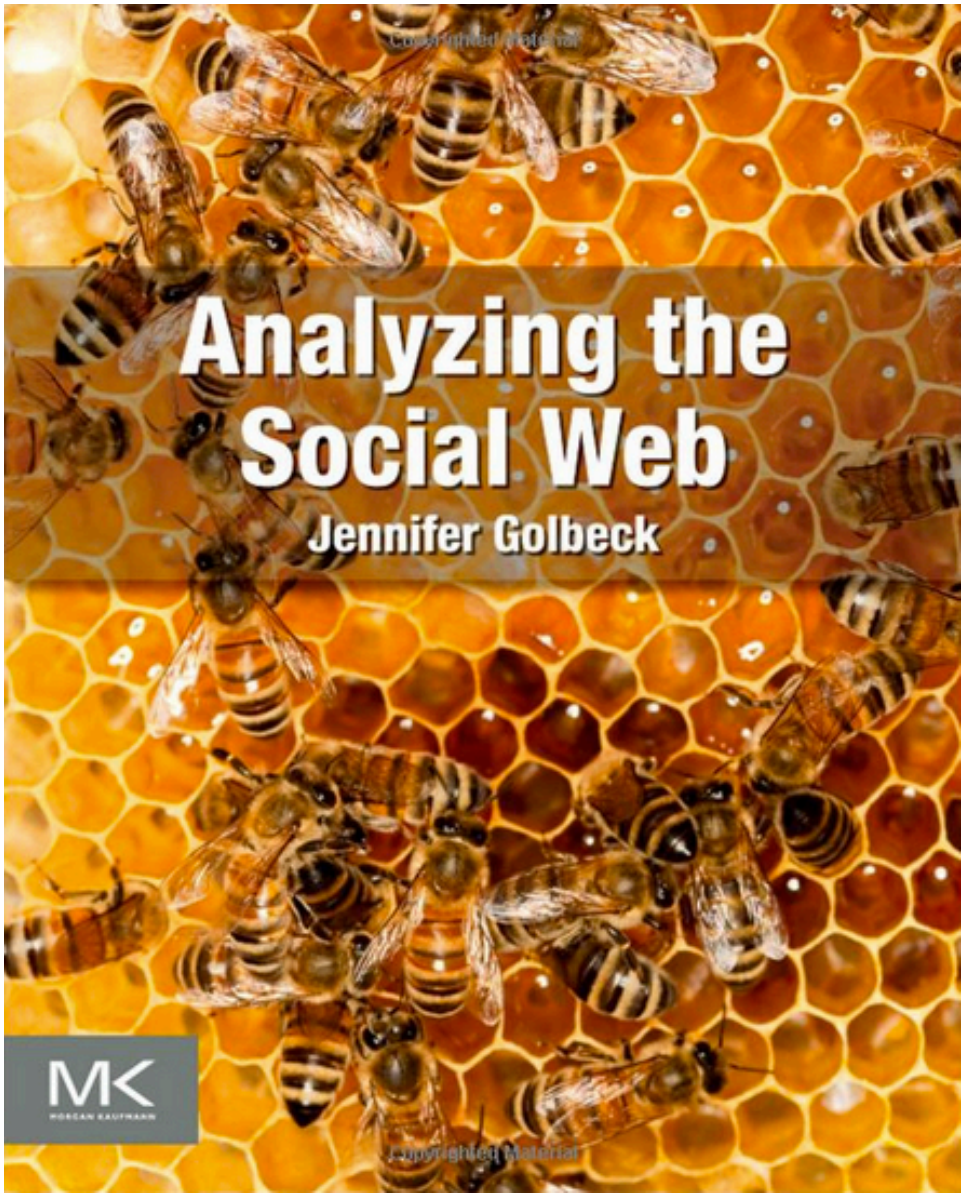
週次 (Week)	日期 (Date)	內容 (Subject/Topics)
9	2016/04/13	Social Media Marketing Analytics (社群媒體行銷分析)
10	2016/04/20	期中報告 (Midterm Project Report)
11	2016/04/27	Deep Learning with Theano and Keras in Python (Python Theano 和 Keras 深度學習)
12	2016/05/04	Deep Learning with Google TensorFlow (Google TensorFlow 深度學習)
13	2016/05/11	Sentiment Analysis on Social Media with Deep Learning (深度學習社群媒體情感分析)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
14	2016/05/18	Social Network Analysis (社會網絡分析)
15	2016/05/25	Measurements and Tools of Social Network Analysis (社會網絡分析量測與工具)
16	2016/06/01	Final Project Presentation I (期末報告 I)
17	2016/06/08	Practice of Big Data Analytics in Teradata (Teradata 大數據分析實務) [Invited Speaker: Irene Chen (陳維君), Consultant, Teradata]
18	2016/06/15	Final Project Presentation II (期末報告 II)

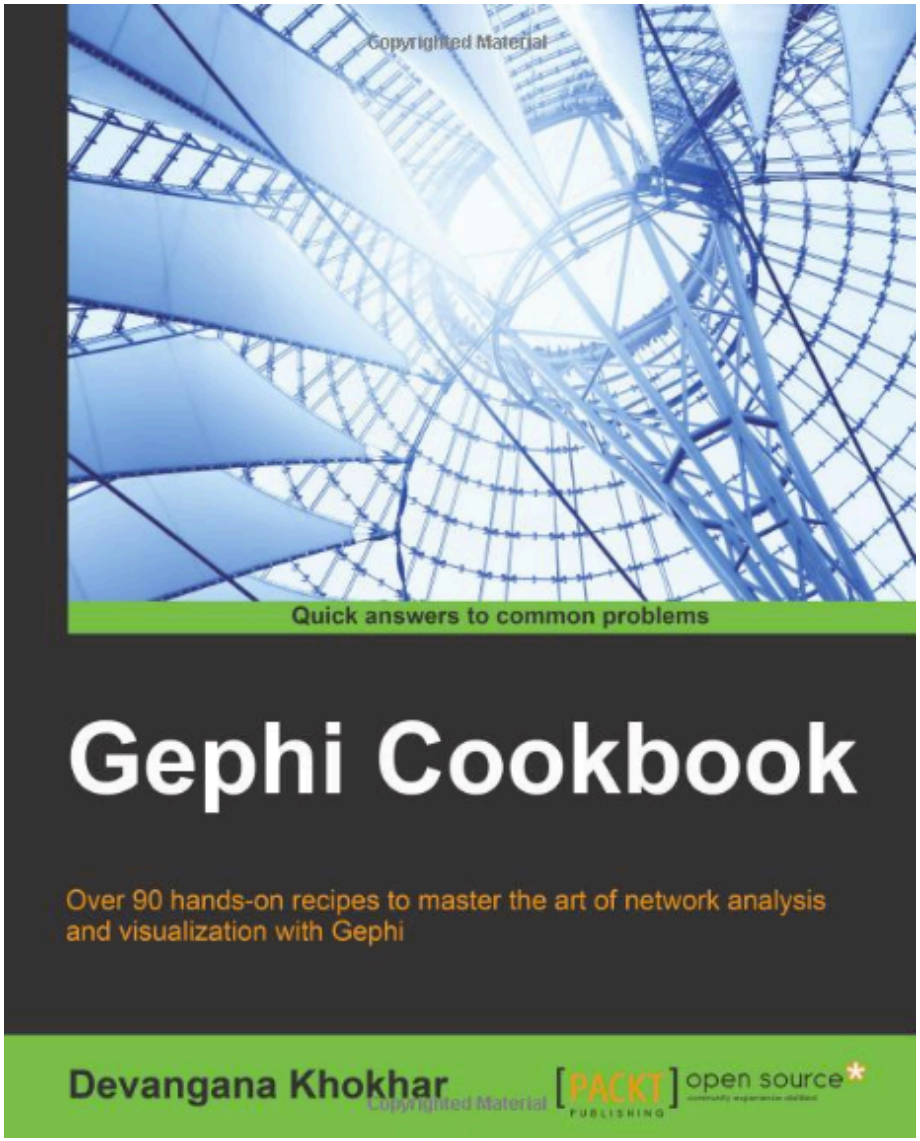
Measurements and Tools of Social Network Analysis

Jennifer Golbeck (2013), **Analyzing the Social Web**, Morgan Kaufmann



Source: http://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311

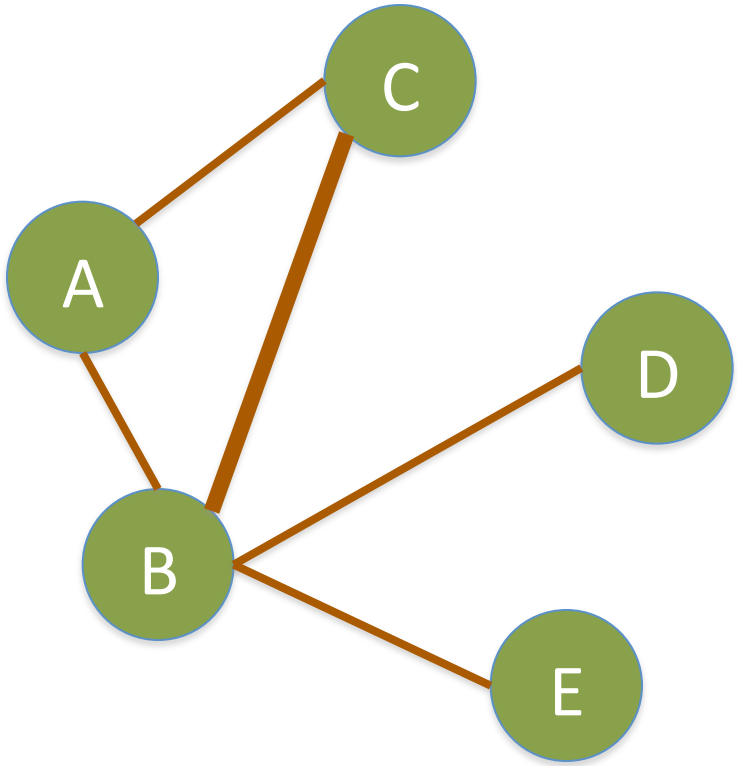
Devangana Khokhar (2015), Gephi Cookbook, Packt Publishing



Source: http://www.amazon.com/Gephi-Cookbook-Devangana-Khokhar/dp/1783987405

Graph Theory

Graph



Graph

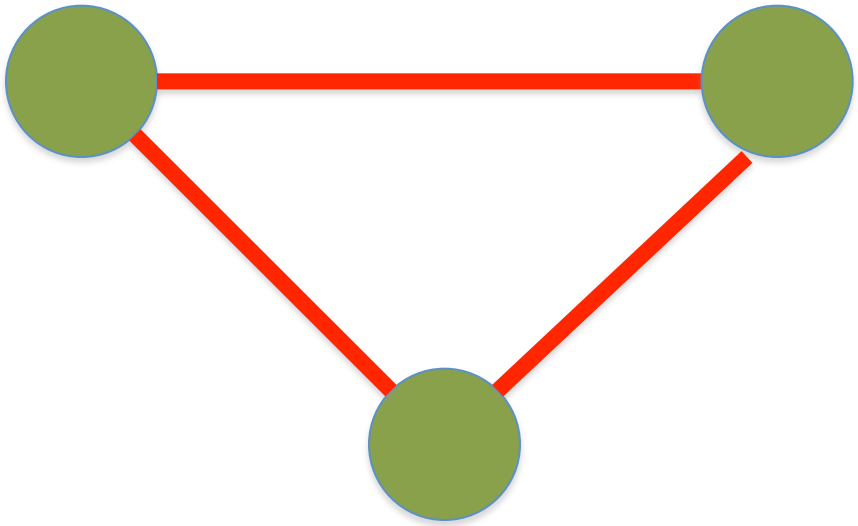
g = (V, E)

Vertex (Node)

Vertices (Nodes)

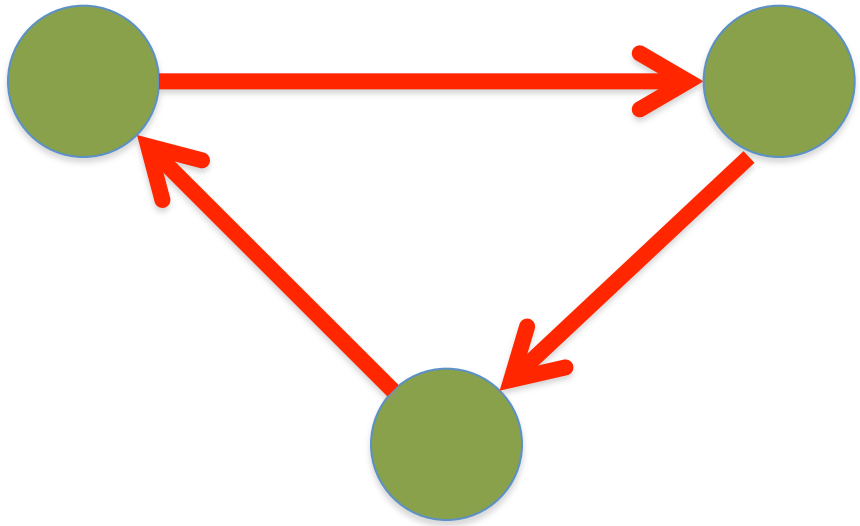
Edge

Edges

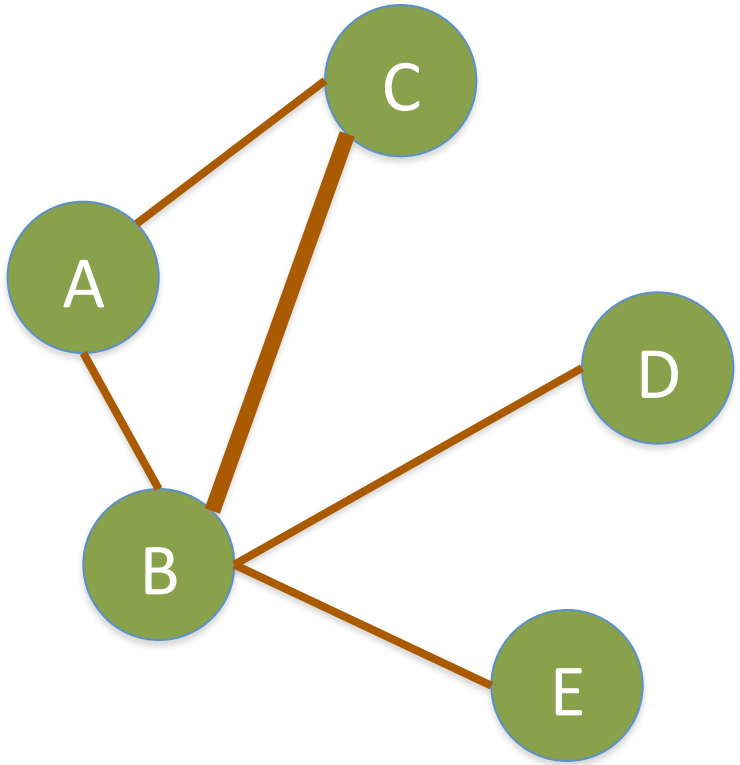


Arc

Arcs

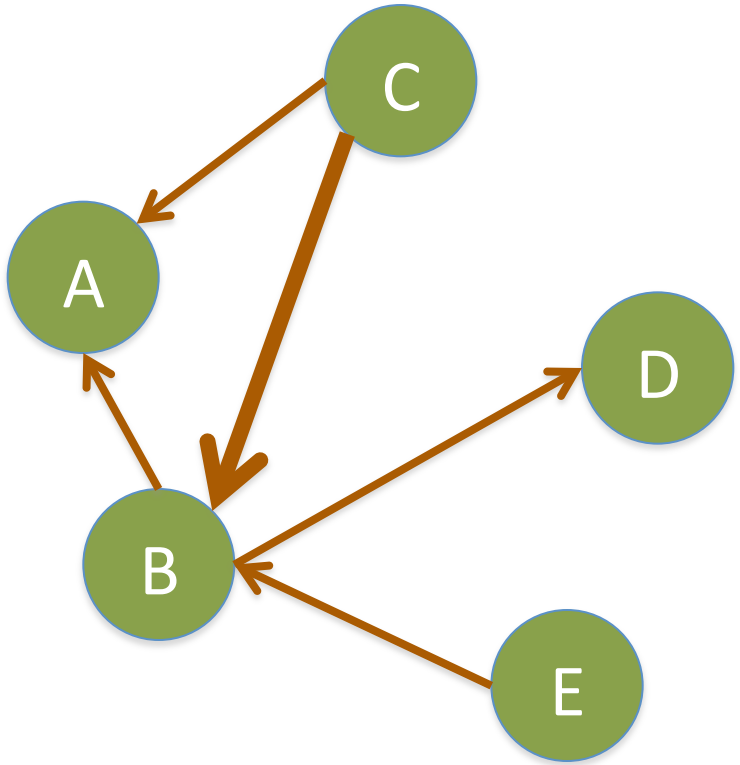


Undirected Graph



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

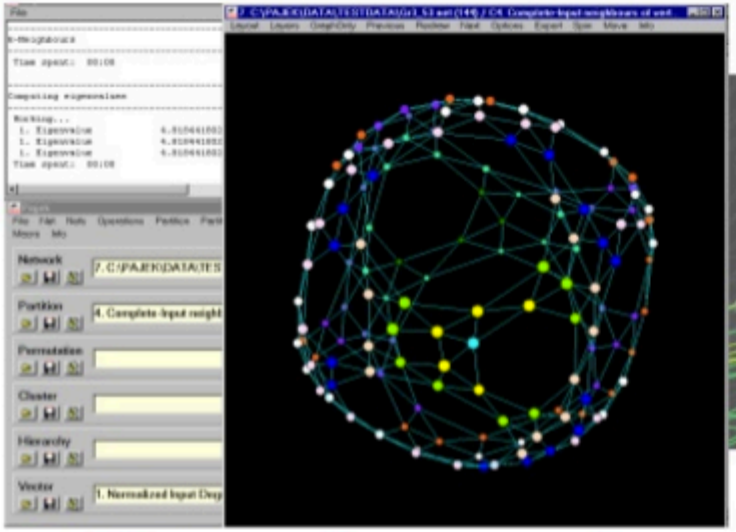
Directed Graph

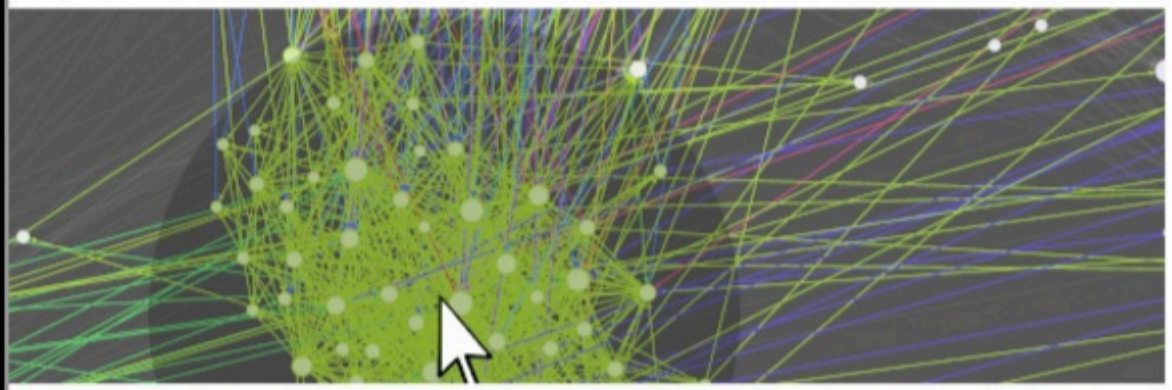


Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Measurements of Social Network Analysis

Exploratory Network Analysis





1 see the network

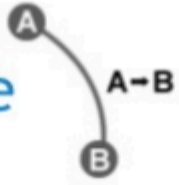
1st graph viz tool: Pajek (1996)
Vladimir Batagelj, Andrej Mrvar

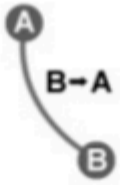
2 interact in real time

Gephi prototype (2008)
group, filter, compute metrics...

3 build a visual language

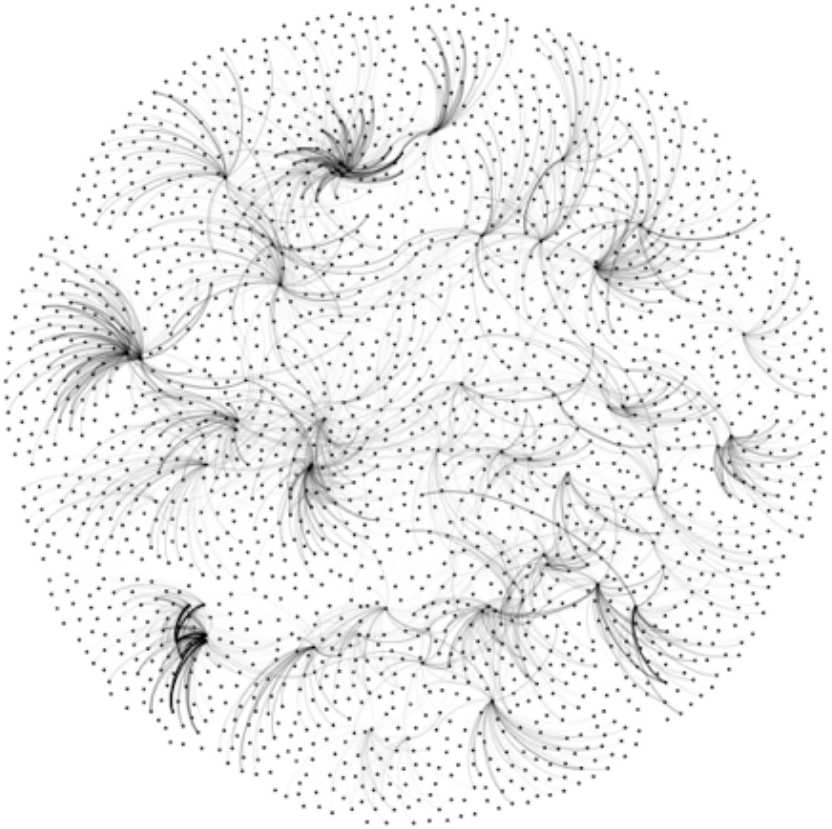
size by rank, color by partition,
label, curved edges, thickness...

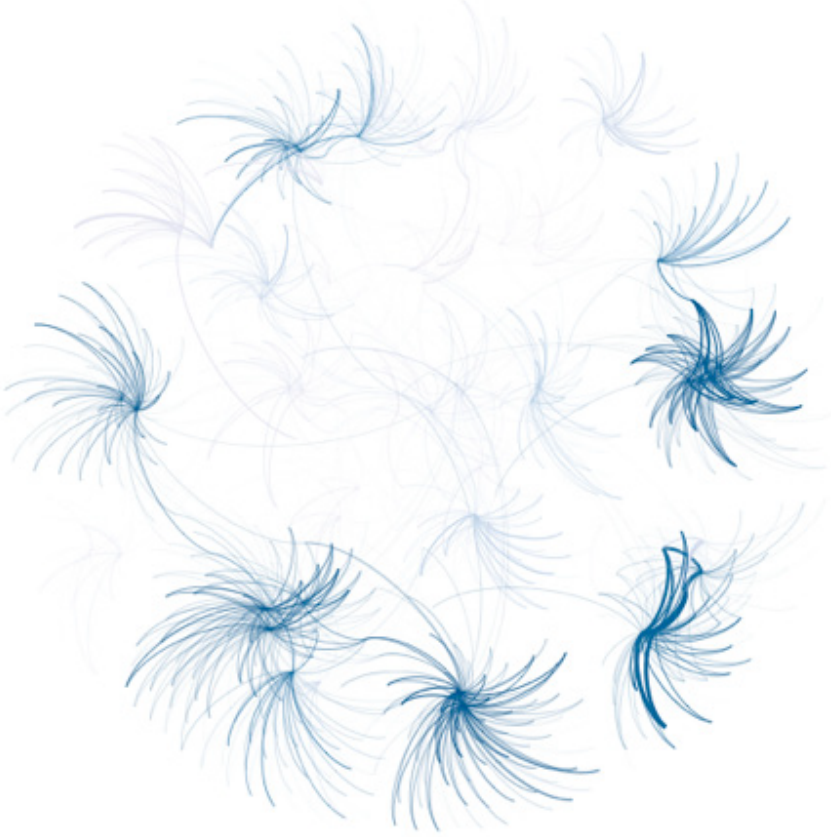




Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Looking for a "Simple Small Truth"? What Data Visualization Should Do?





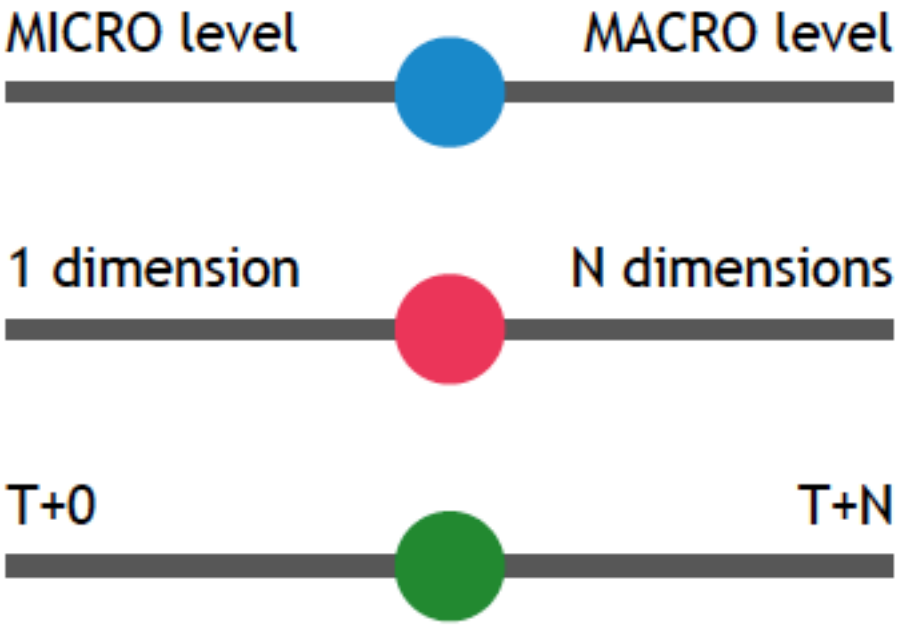
1. Make complex things **simple**
2. Extract **small** information from large data
3. Present **truth**, do not deceive

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Measurements

Looking for Orderness in Data

Make varying 3 cursors simultaneously to extract meaningful patterns



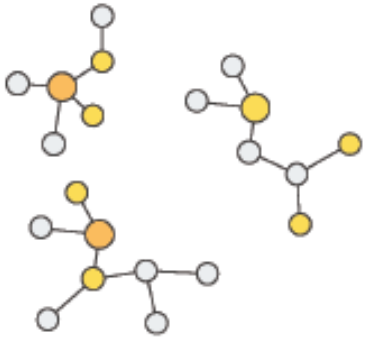
at different levels

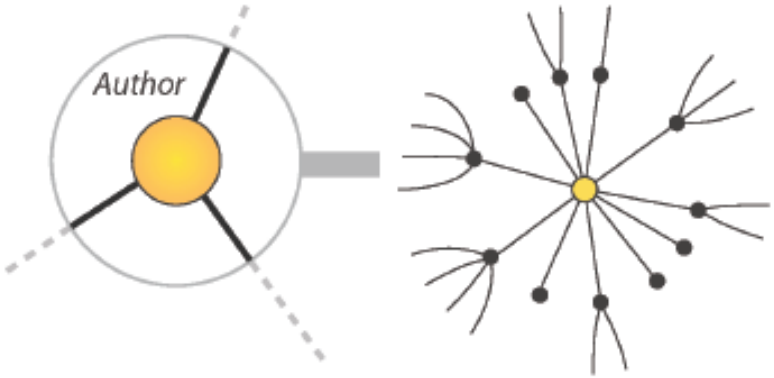
on multiple dimensions

at time scale

Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

“Zoom” cursor on Quantitative Data





Global
- connectivity
- density
- centralization

Local
- communities
- bridges between communities
- local centers vs periphery

Individual
- centrality
- distances
- neighborhood
- location
- local authority vs hub

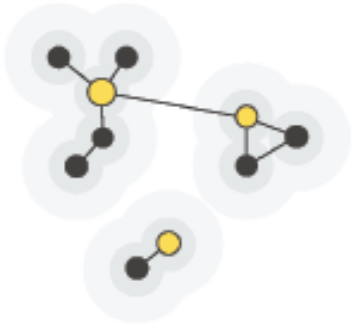
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

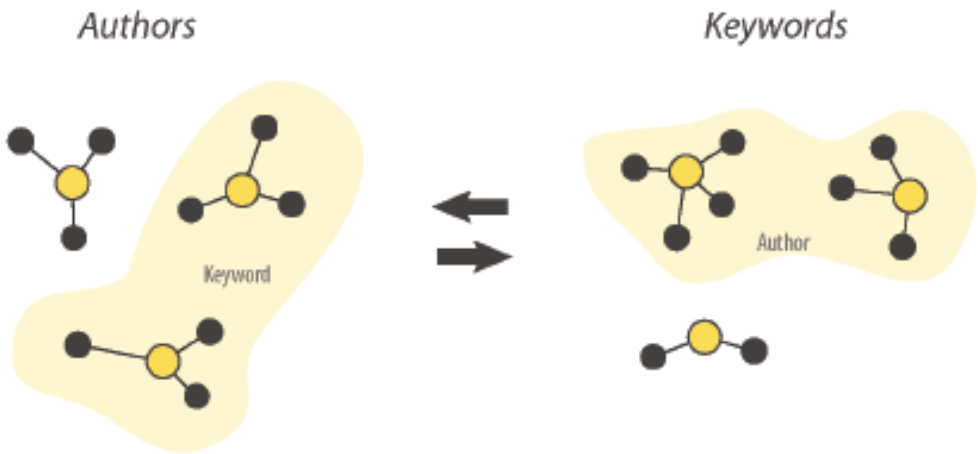
“Crossing” cursor on Quantitative Data

Social
- who with whom
- communities
- brokerage
- influence and power
- homophily

Semantic
- topics
- thematic clusters

Geographic
- spatial phenomena





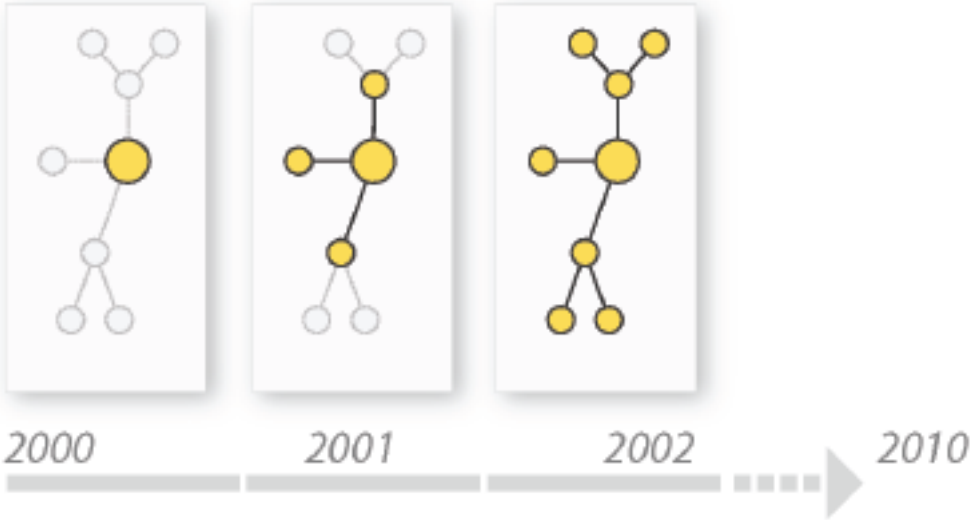
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

“Timeline” cursor on Temporal Data

Evolution of social ties

Evolution of communities

Evolution of topics



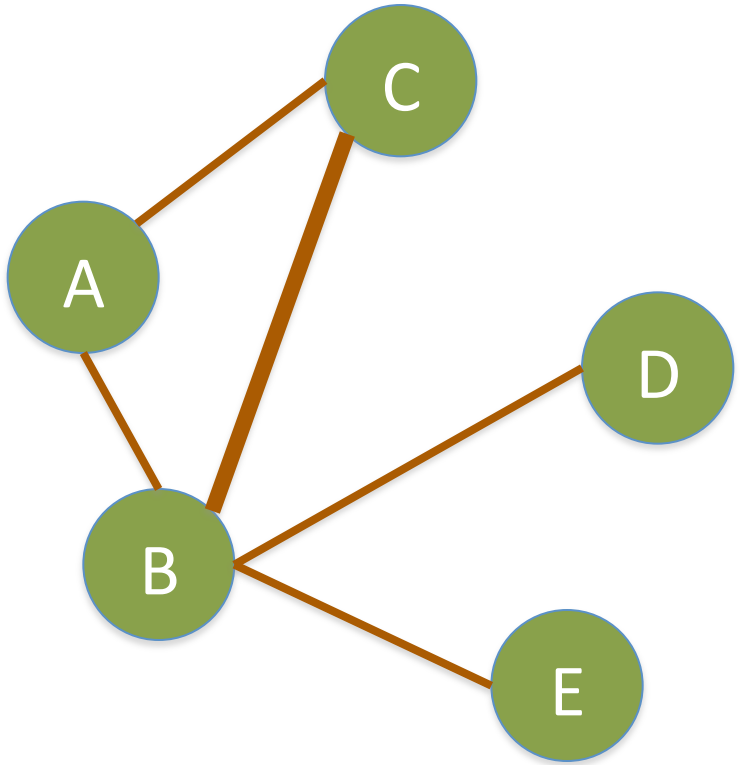
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

SNA Guideline

# nodes	
1 - 100	lists + edges in bonus, focus on qualitative data
100 - 1,000	How attributes explain the structure? • easy to read, "obvious" patterns • focus on entities (in context) • metrics are tools to describe the graph (centrality, bridging...) • links help to build and interpret categories of entities challenge: mix attribute crossing and connectivity
1,000 - 50,000	How the structure explains attributes? • hard to read, problem of "**hidden signals**": track patterns with various layouts and filtering • focus on structures • metrics are tools to build the graph (cosine similarity...) • categories help to understand the structure challenge: pattern recognition
> 50,000	require high computational power

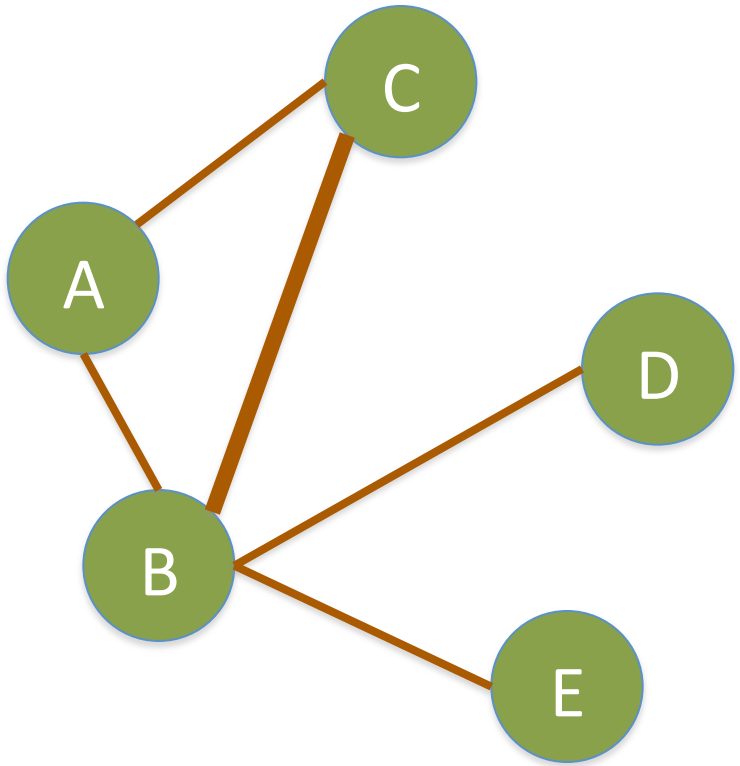
Source: http://sebastien.pro/gephi-icwsm-tutorial.pdf

Degree



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Degree



A: 2

B: 4

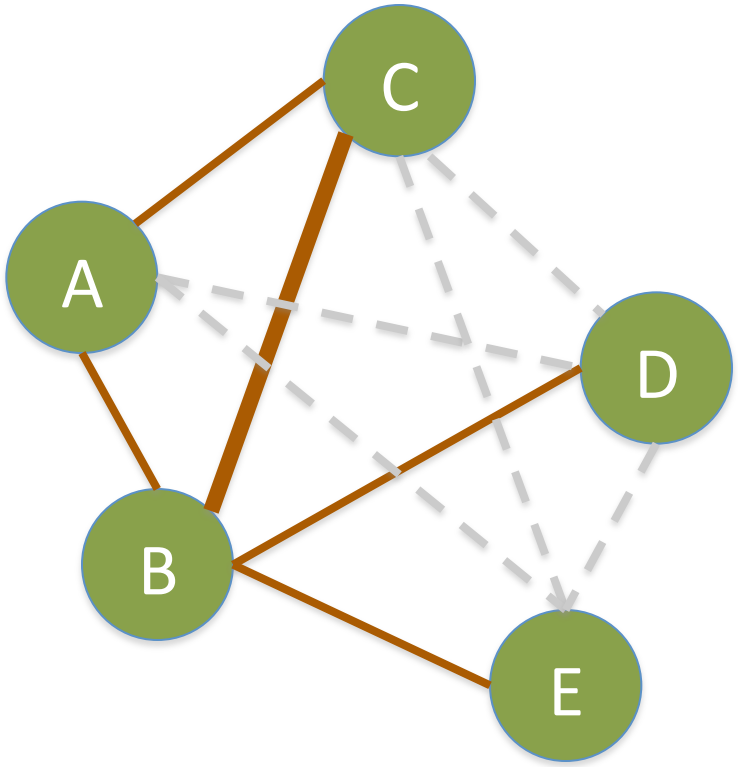
C: 2

D:1

E: 1

Source: https://www.youtube.com/watch?v=89mxOdwPfxA

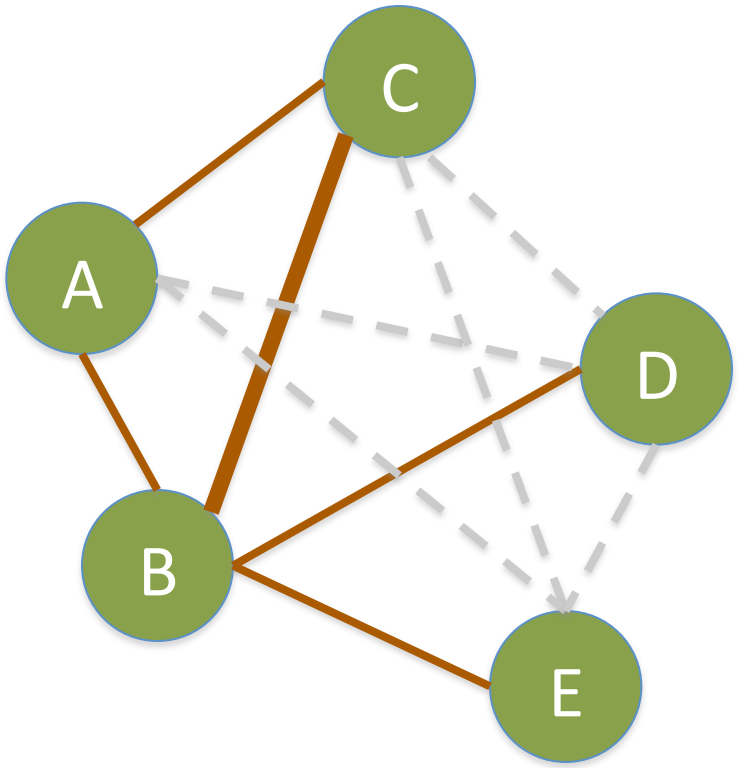
Density



Source: https://www.youtube.com/watch?v=89mxOdwPfxA

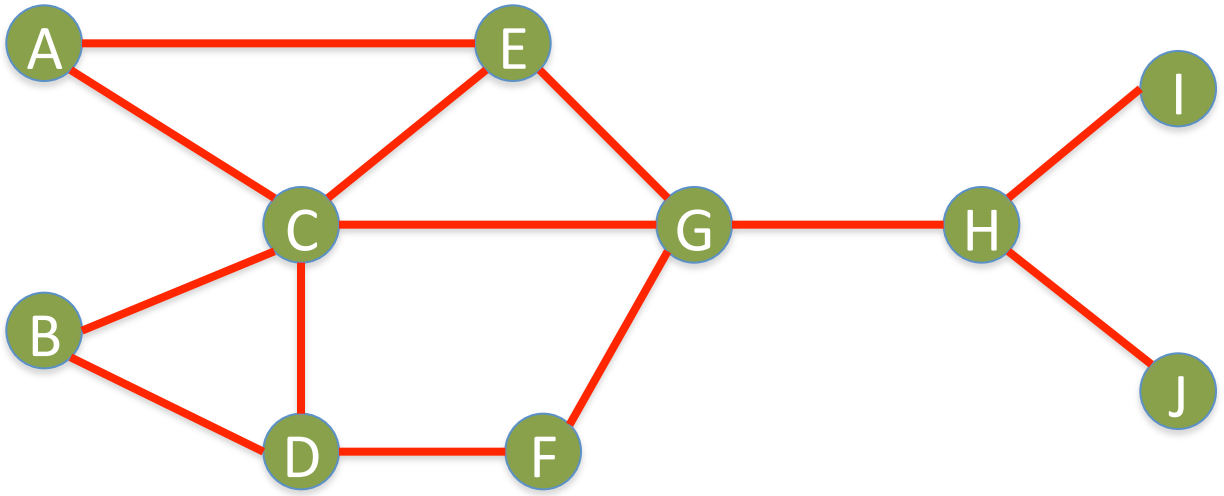
Density

Edges (Links): 5
Total Possible Edges: 10
Density: 5/10 = 0.5



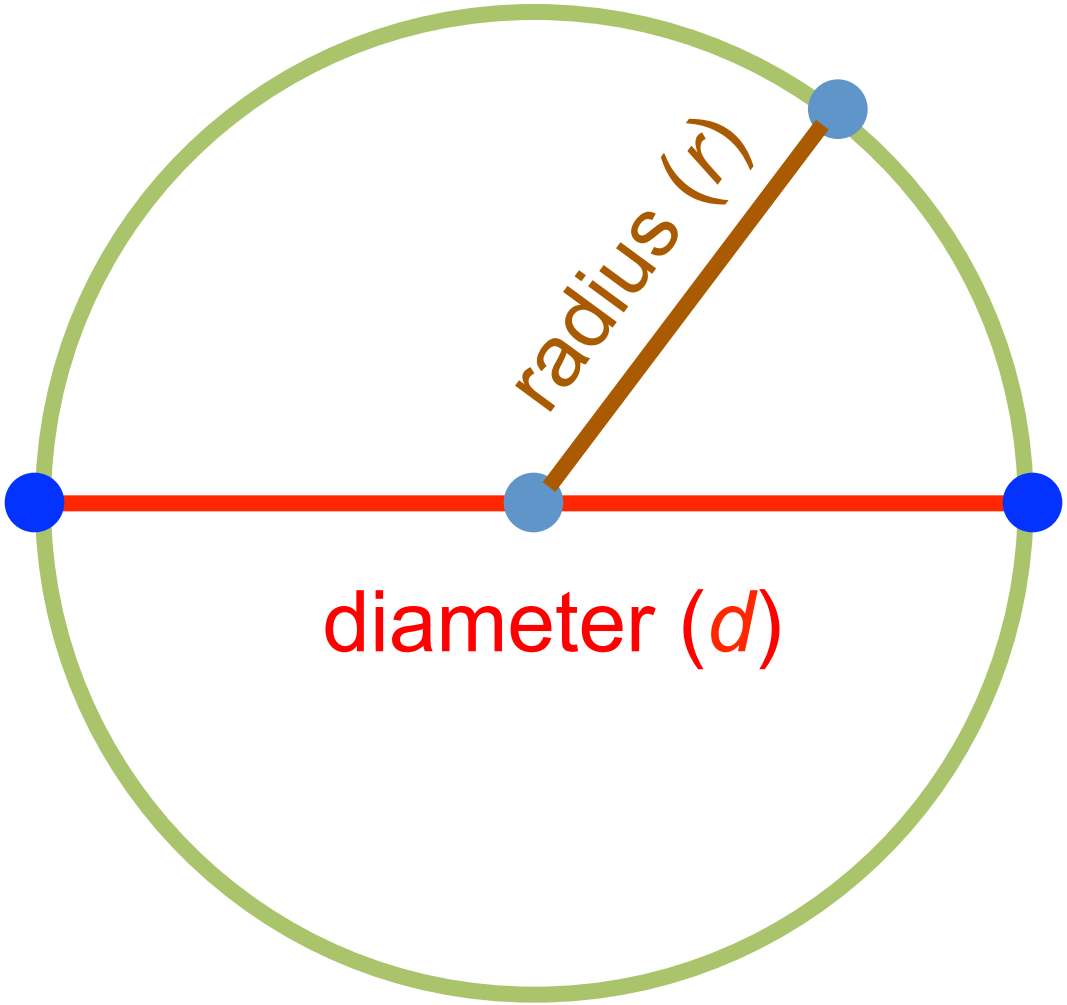
Source: https://www.youtube.com/watch?v=89mxOdwPfxA

Density

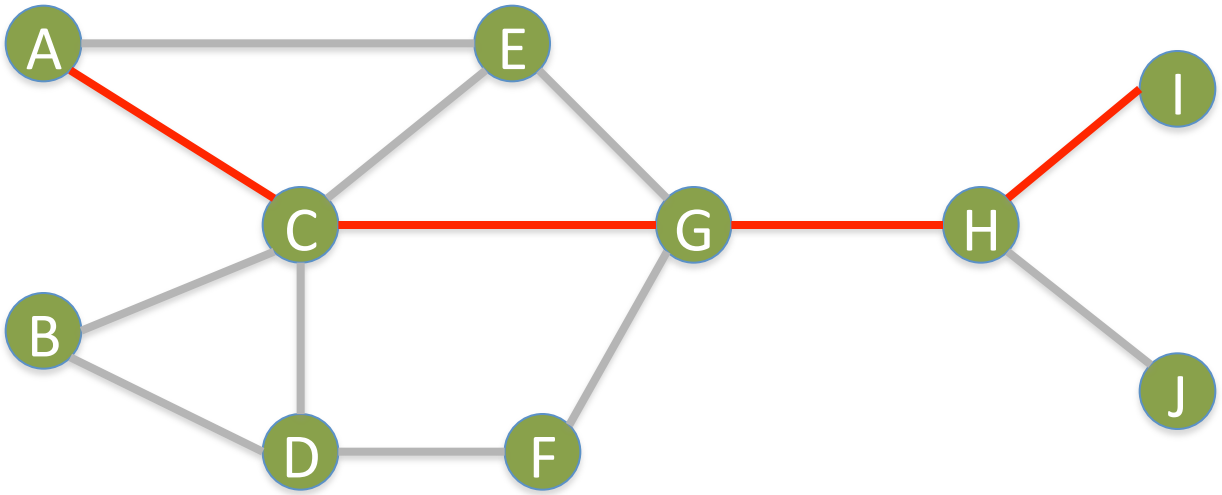


Nodes (n): 10
Edges (Links): 13
Total Possible Edges: (n * (n-1)) / 2 = (10 * 9) / 2 = 45
Density: 13/45 = 0.29

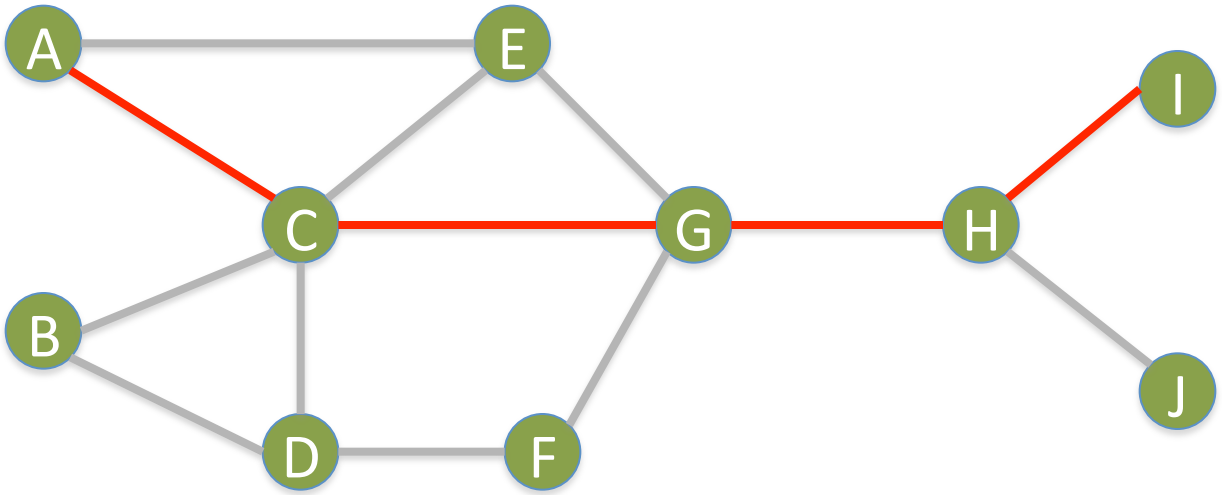
Diameter



Diameter

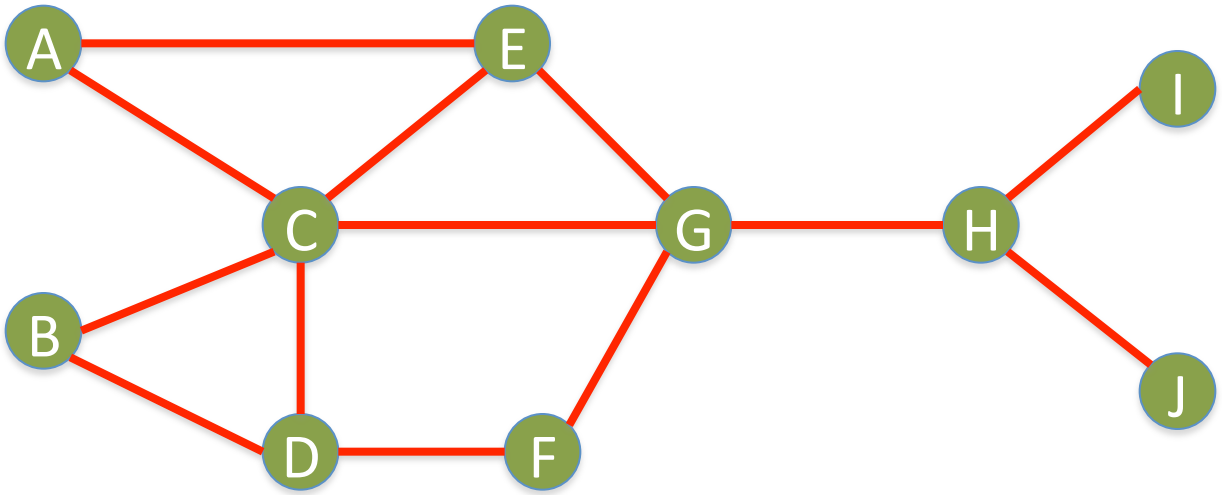


Diameter
Geodesic Path (Shortest Path)



A→ I : Diameter = 4

Which Node is Most Important?



Centrality

- Important or prominent actors are those that are linked or involved with other actors extensively.
- A person with extensive contacts (links) or communications with many other people in the organization is considered more important than a person with relatively fewer contacts.
- The links can also be called **ties**.
  A central actor is one involved in many ties.

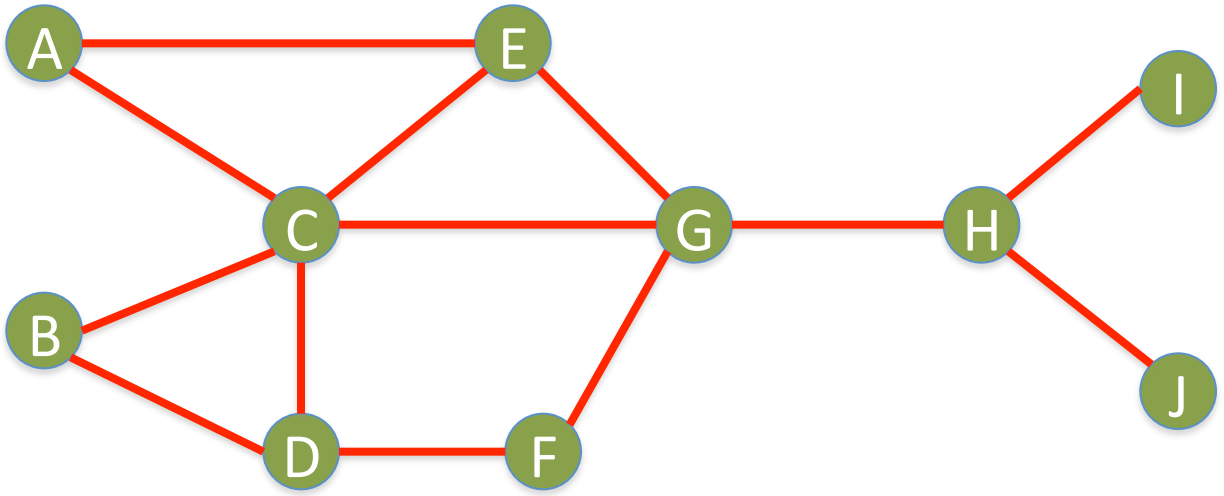
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data"

Social Network Analysis (SNA)

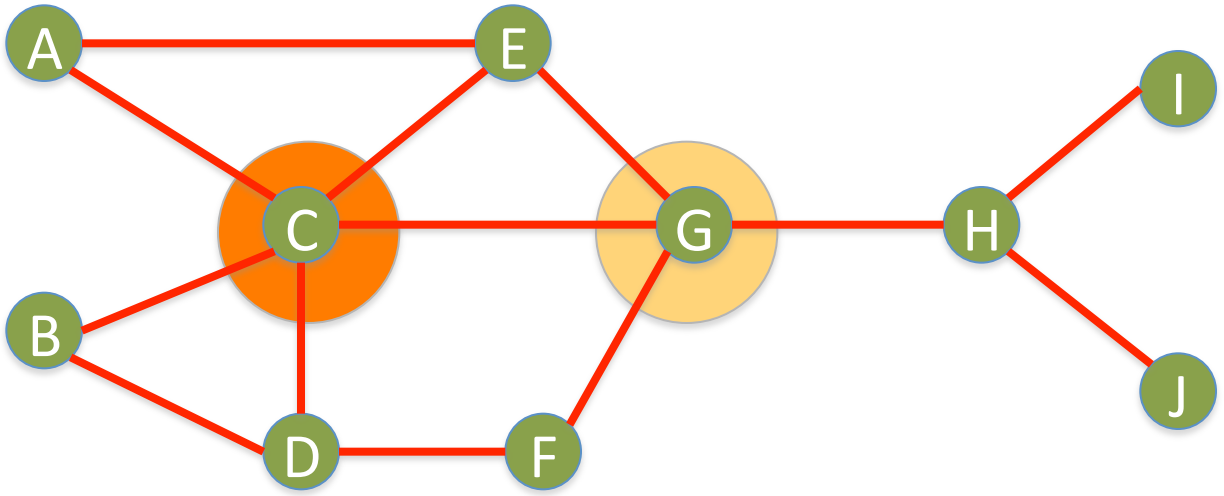
- Degree Centrality
- Betweenness Centrality
- Closeness Centrality

Degree Centrality

Social Network Analysis: Degree Centrality



Social Network Analysis: Degree Centrality



Node	Score	Standardized Score
A	2	2/10 = 0.2
B	2	2/10 = 0.2
C	**5**	**5/10 = 0.5**
D	3	3/10 = 0.3
E	3	3/10 = 0.3
F	2	2/10 = 0.2
G	**4**	**4/10 = 0.4**
H	3	3/10 = 0.3
I	1	1/10 = 0.1
J	1	1/10 = 0.1

Betweenness Centrality

Betweenness centrality:

Connectivity

Number of shortest paths going through the actor

Betweenness Centrality

$$C_B(i) = \sum_{j<k} g_{ik}(i) / g_{jk}$$

Where g_{jk} = the number of shortest paths connecting jk
$g_{jk}(i)$ = the number that actor i is on.

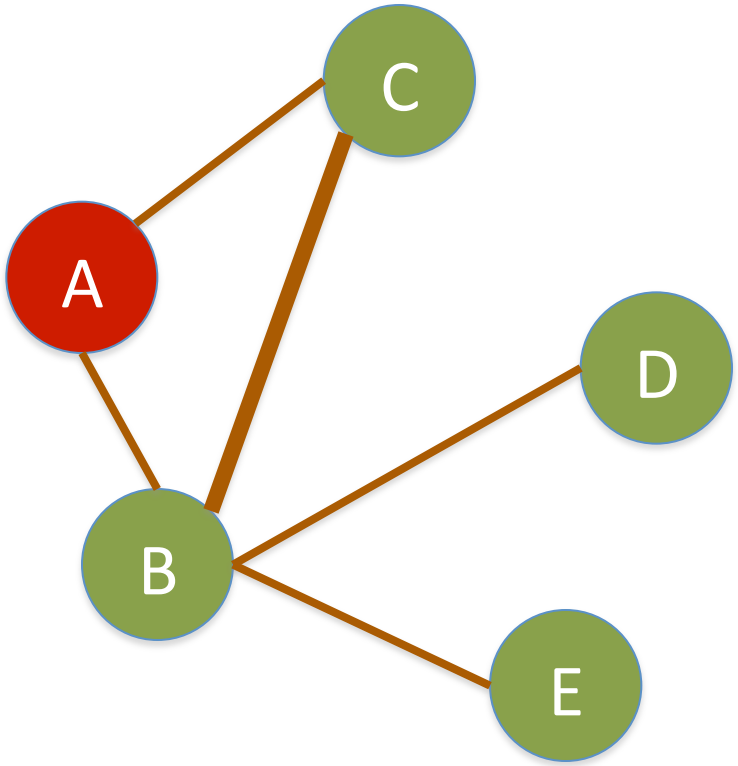
Normalized Betweenness Centrality

$$C'_B(i) = C_B(i) / [(n-1)(n-2)/2]$$

Number of pairs of vertices excluding the vertex itself

Source: https://www.youtube.com/watch?v=RXohUeNCJiU

Betweenness Centrality

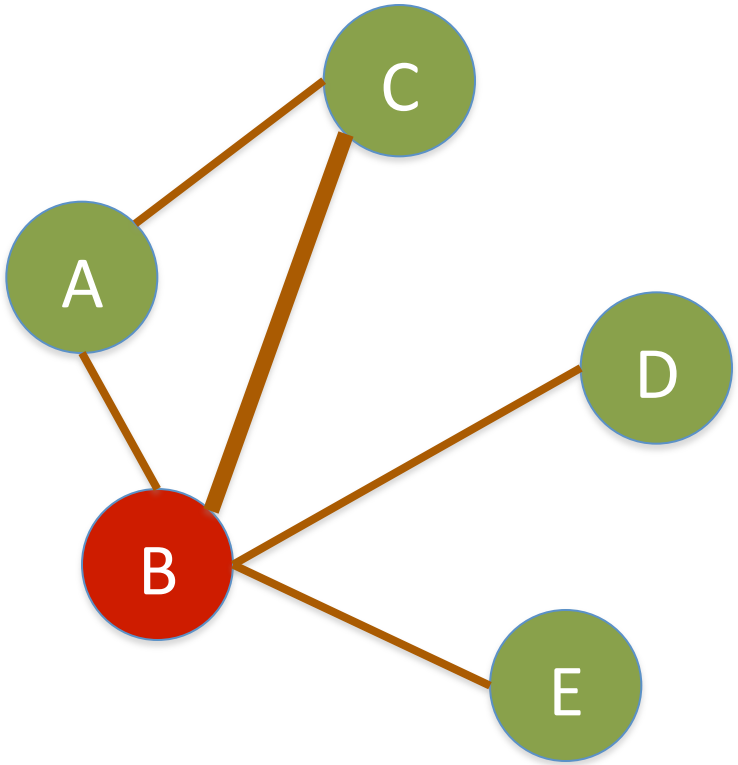


A:
B→C: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
C→D: 0/1 = 0
C→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

A: Betweenness Centrality = 0

Betweenness Centrality

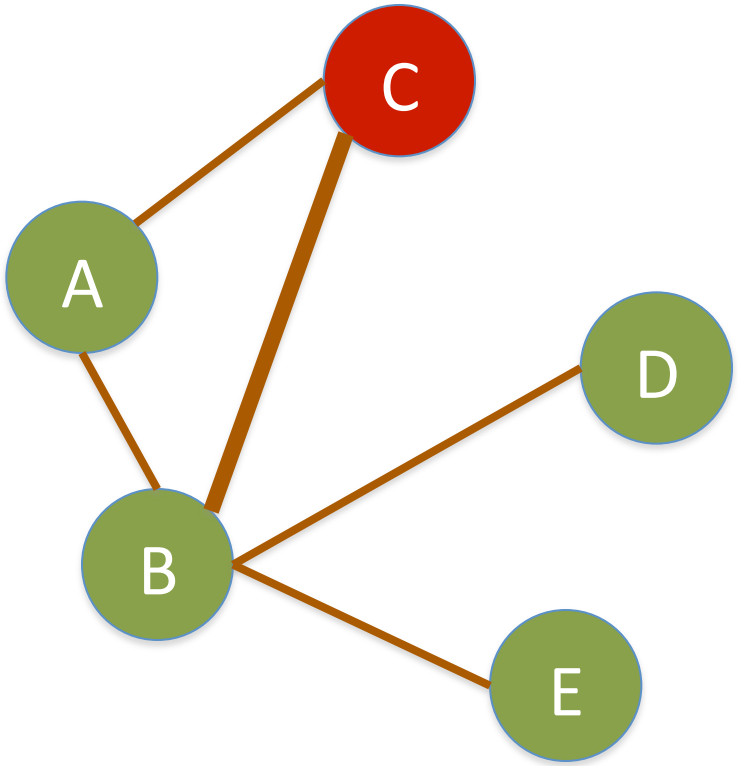


B:
A→C: 0/1 = 0
A→D: 1/1 = 1
A→E: 1/1 = 1
C→D: 1/1 = 1
C→E: 1/1 = 1
D→E: 1/1 = 1

Total: 5

B: Betweenness Centrality = 5

Betweenness Centrality

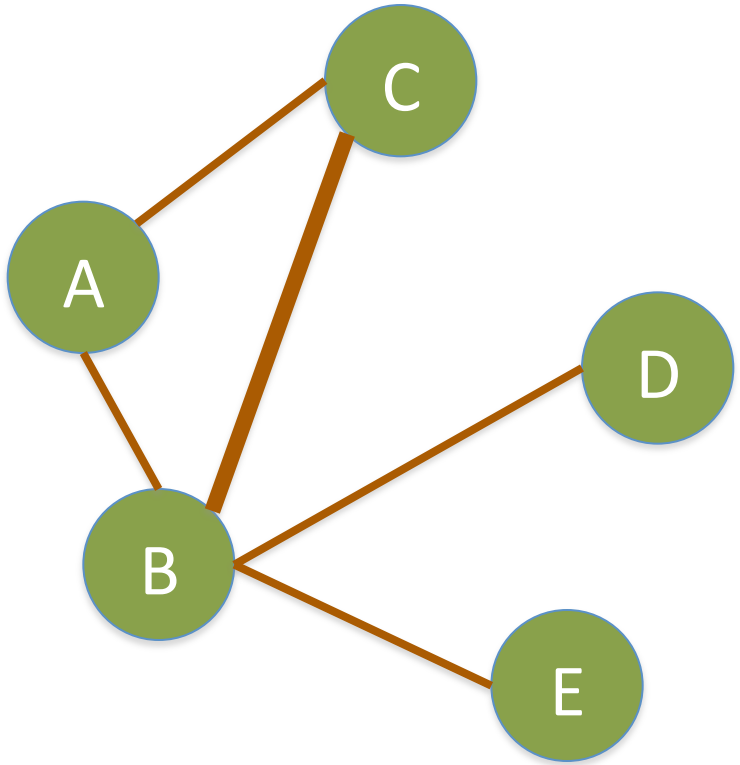


C:
A→B: 0/1 = 0
A→D: 0/1 = 0
A→E: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

C: Betweenness Centrality = 0

Betweenness Centrality



A: 0

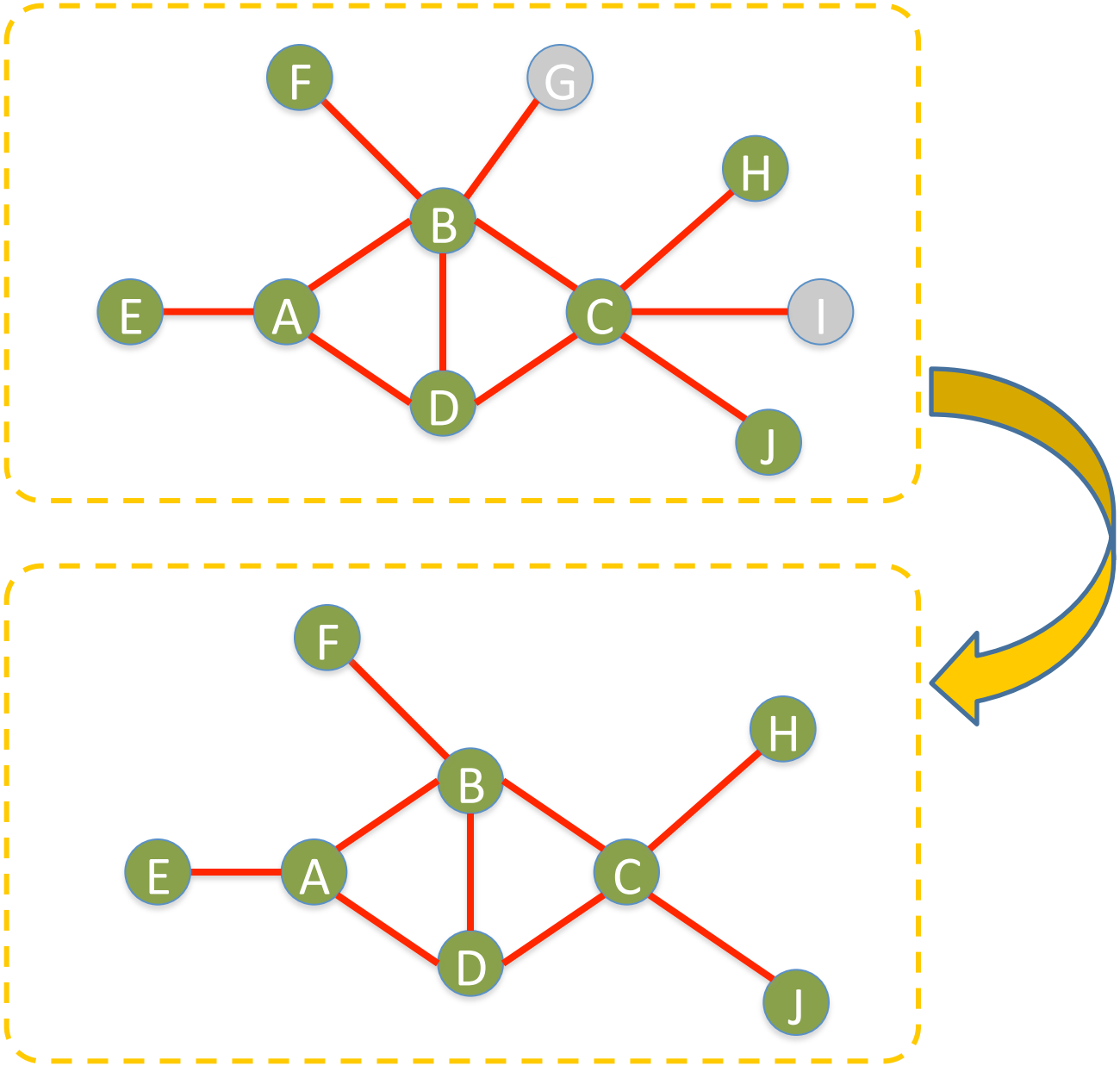
B: 5

C: 0

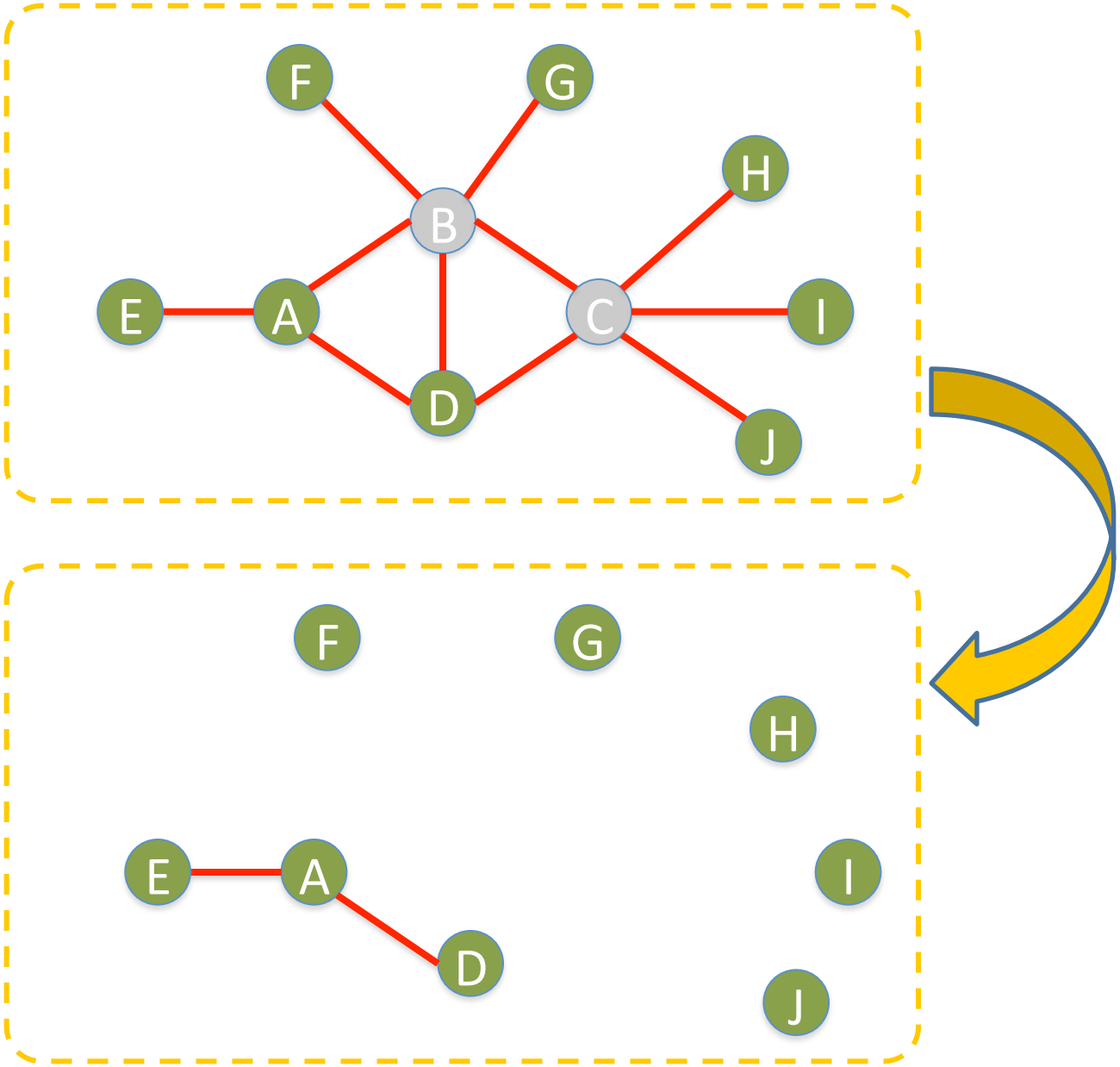
D: 0

E: 0

Which Node is Most Important?

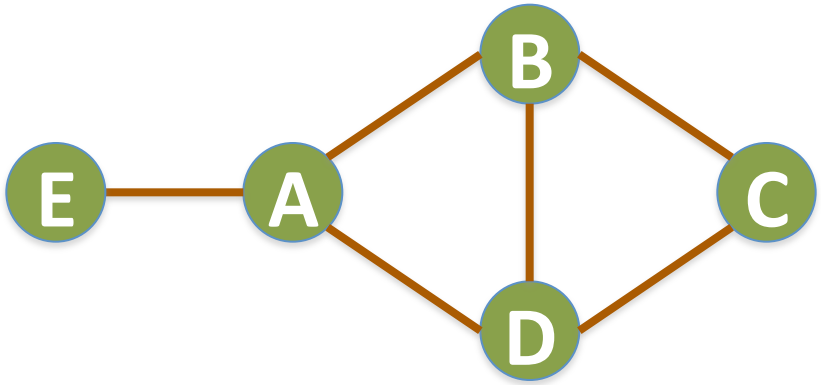


Which Node is Most Important?

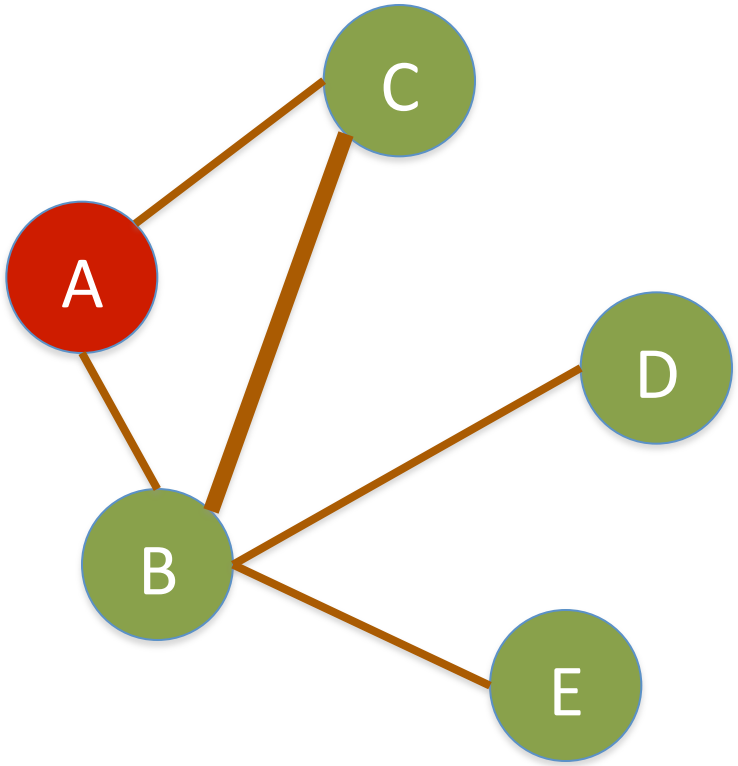


Betweenness Centrality

$$C_B(i) = \sum_{j<k} g_{ik}(i) / g_{jk}$$



Betweenness Centrality



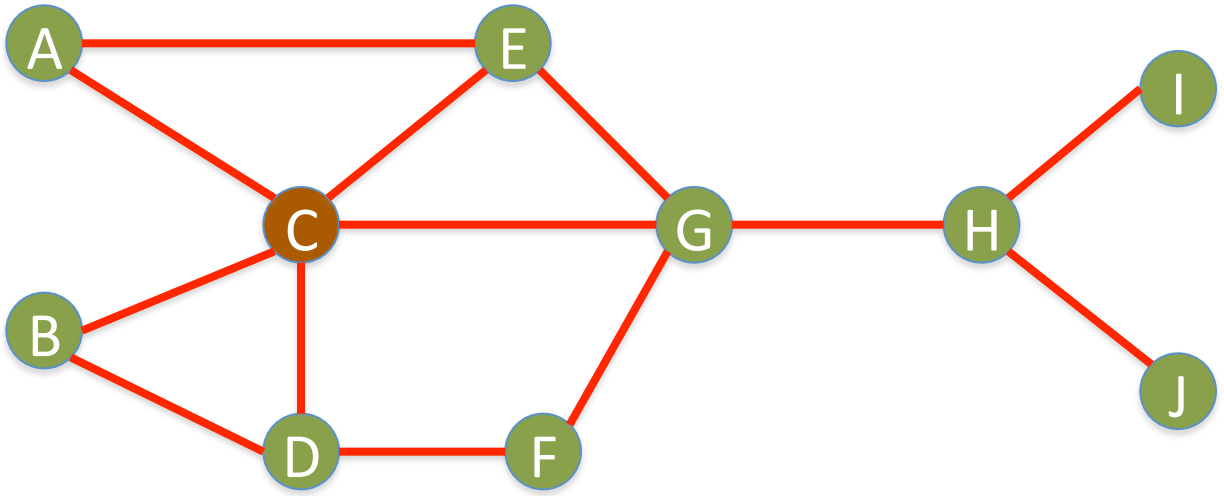
A:
B→C: 0/1 = 0
B→D: 0/1 = 0
B→E: 0/1 = 0
C→D: 0/1 = 0
C→E: 0/1 = 0
D→E: 0/1 = 0

Total: 0

A: Betweenness Centrality = 0

Closeness Centrality

Social Network Analysis: Closeness Centrality

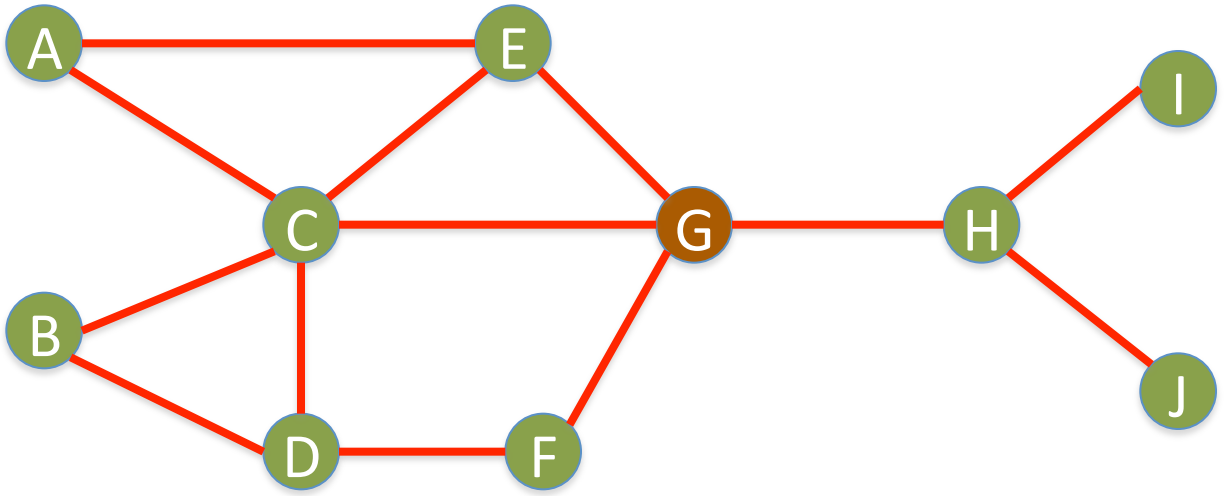


C→A: 1
C→B: 1
C→D: 1
C→E: 1
C→F: 2
C→G: 1
C→H: 2
C→I: 3
C→J: 3

Total=15

C: Closeness Centrality = 15/9 = 1.67

Social Network Analysis: Closeness Centrality

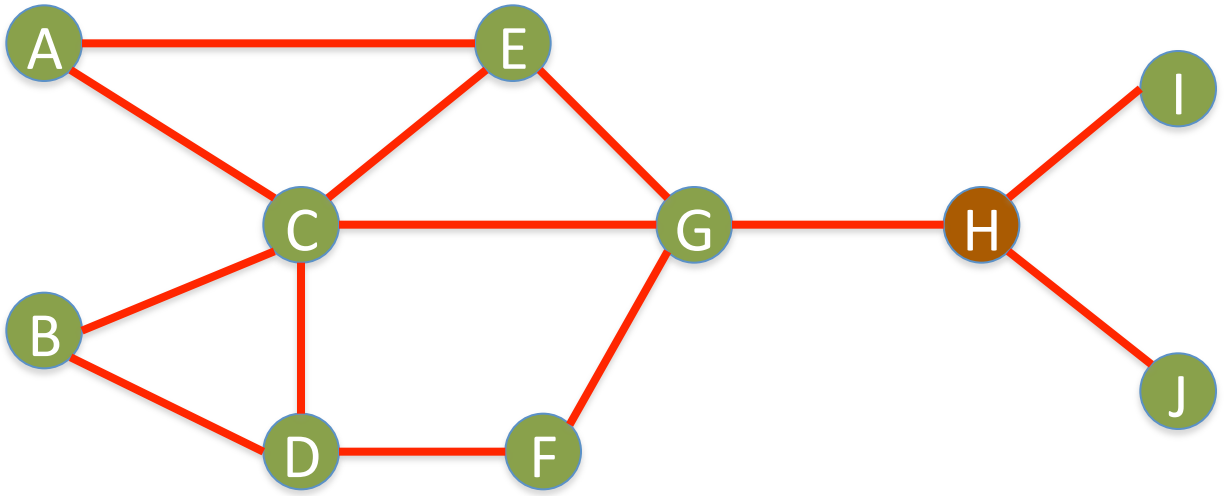


G→A: 2
G→B: 2
G→C: 1
G→D: 2
G→E: 1
G→F: 1
G→H: 1
G→I: 2
G→J: 2

Total=14

G: Closeness Centrality = 14/9 = 1.56

Social Network Analysis: Closeness Centrality

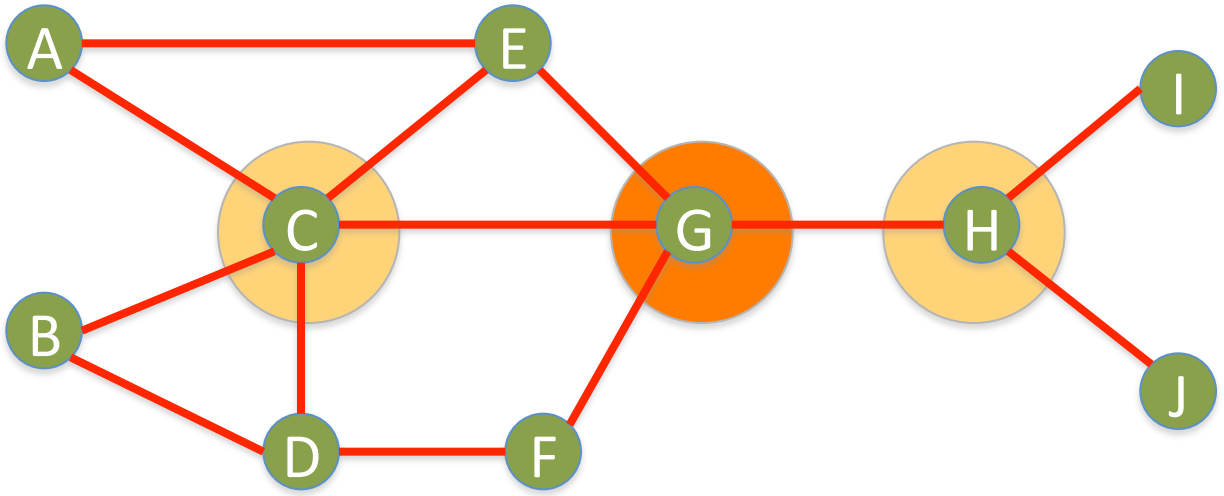


H→A: 3
H→B: 3
H→C: 2
H→D: 2
H→E: 2
H→F: 2
H→G: 1
H→I: 1
H→J: 1

Total=17

H: Closeness Centrality = 17/9 = 1.89

Social Network Analysis: Closeness Centrality



G: Closeness Centrality = 14/9 = 1.56 1

C: Closeness Centrality = 15/9 = 1.67 2

H: Closeness Centrality = 17/9 = 1.89 3

Eigenvector centrality:

Importance of a node
depends on
the importance of its neighbors

Social Network Analysis: Closeness Centrality

Sum of the reciprocal distances

$$C_C(p_k) = \sum_{i=1}^{n} d(p_i, p_k)^{-1}$$

where $d(p_j, p_k)$ is the geodesic distance (shortest paths) linking p_j, p_k

Social Network Analysis: Betweenness Centrality

$$C_B(p_k) = \sum_{i<j}^{n} \frac{g_{ij}(p_k)}{g_{ij}}; \quad i \neq j \neq k$$

where g_{ij} is the geodesic distance (shortest paths) linking p_i and p_j and $g_{ij}(p_k)$ is the geodesic distance linking p_i and p_j that contains p_k.

Source: Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, *6*(3), 403-412.

Social Network Analysis: Degree Centrality

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k)$$

where $a(p_i, p_k) = 1$ if and only if p_i and p_k are connected by a line
0 otherwise

$$C'_D(p_k) = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1}$$

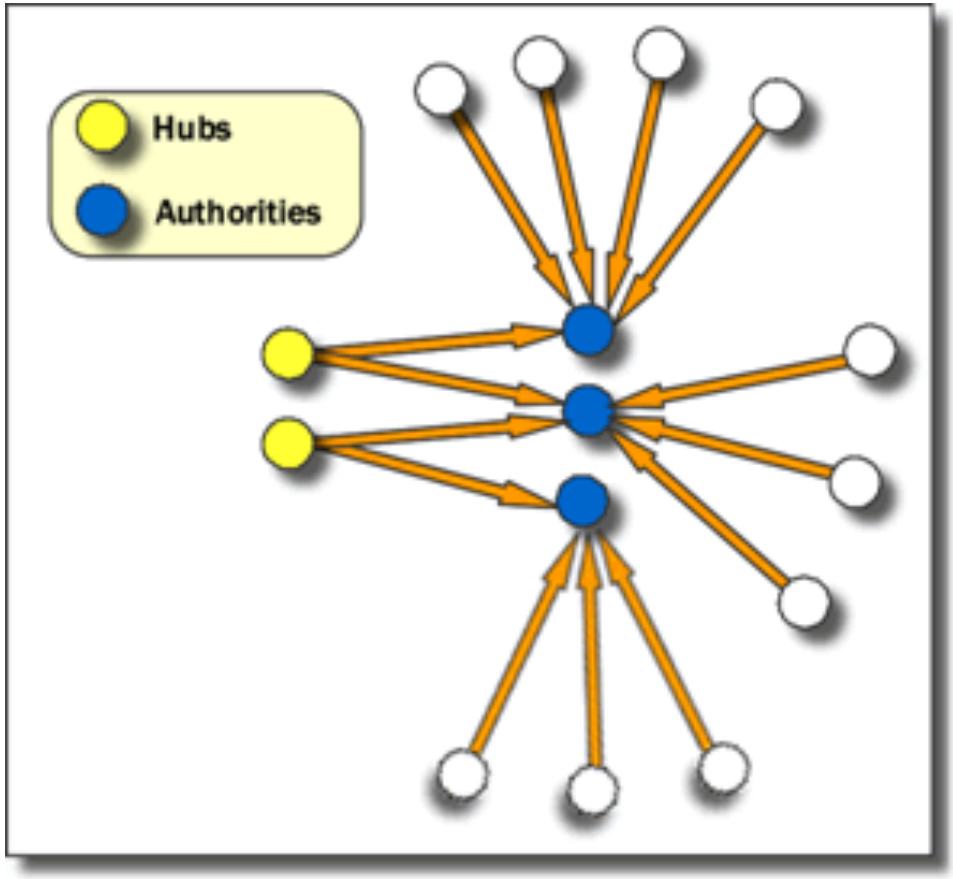
Social Networks, 1 (1978/79) 215–239

Centrality in Social Networks
Conceptual Clarification

Linton C. Freeman

*Lehigh University**

Social Network Analysis: Hub and Authority



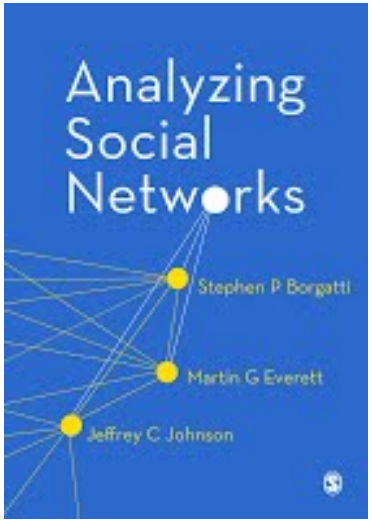
Hubs are entities that point to a relatively large number of authorities. They are essentially the mutually reinforcing analogues to authorities. Authorities point to high hubs. Hubs point to high authorities. You cannot have one without the other.

Source: http://www.fmsasg.com/SocialNetworkAnalysis/

Tools of Social Network Analysis

Social Network Analysis (SNA) Tools

- NetworkX
- igraph
- Gephi
- UCINet
- Pajek



Tools of Social Network Analysis

- Focused Desktop Tools
  - –Gephi
  - –NodeXL
  - –Cytoscape
  - –Ucinet
  - –Pajek

Source: https://www.researchgate.net/post/Which_open_source_software_is_best_for_network_data_analysis

Tools of Social Network Analysis

- Developer Tools
  - –NetworkX
  - –iGraph
  - –SNAP
  - –sigma.js

Source: https://www.researchgate.net/post/Which_open_source_software_is_best_for_network_data_analysis

Gephi

Gephi
makes graphs handy

Download Blog Wiki Forum Support Bug tracker

Home Features Learn Develop Plugins Services Consortium

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

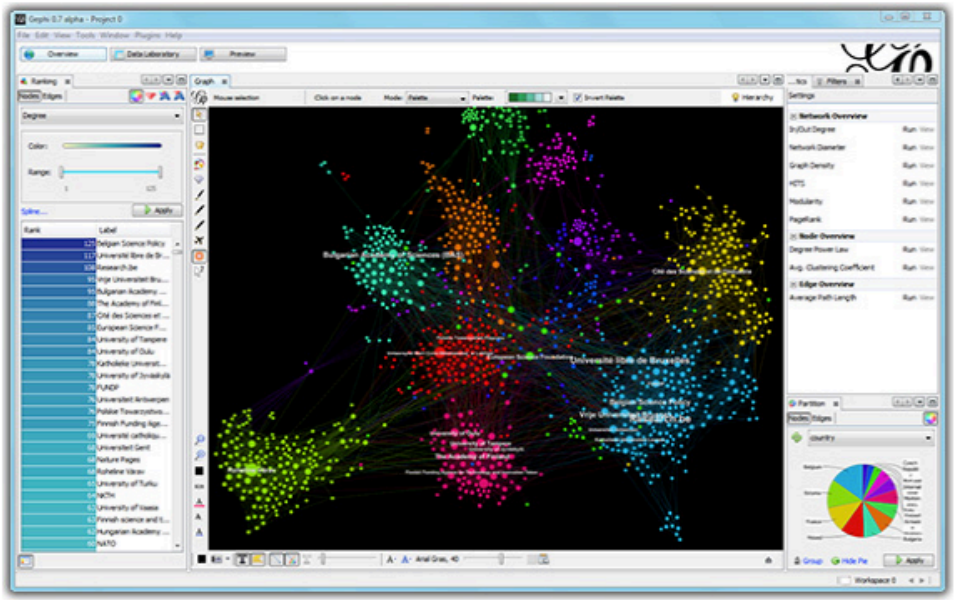
Runs on Windows, Mac OS X and Linux.

Learn More on Gephi Platform »

Download FREE
Gephi 0.9.1

Release Notes | System Requirements

- Features
- Quick start
- Screenshots
- Videos



Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:

Donate

APPLICATIONS

- **Exploratory Data Analysis**: intuition-oriented analysis by networks manipulations in real time.
- **Link Analysis**: revealing the underlying structures of associations between objects.
- **Social Network Analysis**: easy creation of social

Like Photoshop™ for graphs.

— the Community

LATEST NEWS

Gephi updates with 0.9.1 version

PAPERS

https://gephi.org/

UCINET

UCINET Software

Search this site

Home | Download | Versions | FAQs & Tips | Buy!

Navigation

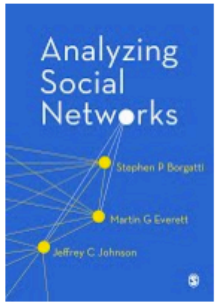
- Home
- Datasets
- Download
- History
- How to Use
  - QuickStart guide
  - Users group
  - FAQs & Tips
  - Tutorials
  - Spanish Resources
  - Workshops
  - More ...
- Purchasing
  - Info
  - Buy now
  - Trial Version
- Versions
  - Recent
  - Jan 2012 only
  - 2011 and before
  - Fix List

Links

- Analytic Technologies
  - NetDraw
  - E-Net
- LINKS Center
  - Workshop
- Steve Borgatti

New! UCINET-oriented book on social network analysis now available! See details.

UCINET 6 for Windows is a software package for the analysis of social network data. It was developed by Lin Freeman, Martin Everett and Steve Borgatti. It comes with the NetDraw network visualization tool.



If you use the software, please cite it. Here is a sample citation:

- **Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.**

For customer support (e.g., ordering info, billing etc) contact roberta@analytictech.com. For tech support join the users group or contact support@analytictech.com. We prefer you try the users group first since the answer to your question may benefit others.

Download and/or Purchase

- The program can be downloaded and used for free for 90 days. In addition, students can purchase the downloaded program for $40. Faculty and government can purchase the downloaded program for $150, and all others pay $250. Site licenses and extremely generous volume discounts are available.
- Note that all purchases are provided as electronic downloads. If necessary you can order a CD from us for an exorbitant fee, but there is no reason to do this. Purchasers of the software are welcome to burn their own CDs at will. They are also free to download the program to all of their computers.
- For more details, including questions about taxes, shipping costs, payment methods, etc., please visit the Order Info page.

News

Week-long workshop on SNA
The LINKS Center at the University of Kentucky is offering its annual 1-week summer workshop on social network analysis June 6-10, 2016 on the University of Kentucky campus ...
Posted Mar 15, 2016, 12:54 PM by Steve Borgatti

Showing posts **1 - 1** of **9**.
View more »

Current Version

Version 6.614 | 22 May 2016 Changed Network|Compare aggregate proximity matrices|partition to be able to handle missing valuesChanged the CLI's IPF routine to default to treating diagonal values

https://sites.google.com/site/ucinetsoftware/home

Pajek

Networks / Pajek

Program for Large Network Analysis

In January 2008 this page was replaced by Pajek Wiki.

Pajek runs on Windows and is free for noncommercial use.

DOWNLOAD Pajek

Data: test networks, GPHs, GEDs, PDB files.

Screenshots; History; Manual (pdf); Papers/presentations; Applications; in News; Examples: SVG, PDF.

How to ? English / Slovene / Japanese (problems with IE - download and use Acrobat reader).
Pajek nicely runs on Linux via Wine, **Converting Excel/text into Pajek format**.
Pajek to SVG animation. **WoS to Pajek**.

Slides from **NICTA workshop**, Sydney, Australia, June 14-17, 2005.
Slides from **workshop at GD'05**, Limerick, Ireland, Sept 11-14, 2005.
Pajek workshop at **XXVIII Sunbelt Conference**, St. Pete Beach, Florida, USA, January 22-27, 2008: slides.
Network analysis course at **ECPR Summer School in Methods and Techniques**, Ljubljana, Slovenia, July 30 - August 16, 2008.

W. de Nooy, A. Mrvar, V. Batagelj: ***Exploratory Social Network Analysis with Pajek***, CUP, January 2005; ESNA page.
P. Doreian, V. Batagelj, A. Ferligoj: ***Generalized Blockmodeling***, CUP, November 2004.

Chapter about Pajek: V. Batagelj, A. Mrvar: ***Pajek - Analysis and Visualization of Large Networks***.
in Jünger, M., Mutzel, P., (Eds.) *Graph Drawing Software*. Springer, Berlin 2003. p. 77-103 / Amazon.

An improved version of the paper presented at Sunbelt'97 was published in Connections 21(1998)2, 47-57 - V. Batagelj, A. Mrvar: ***Pajek - Program for Large Network Analysis*** (PDF; PRISON.KIN).

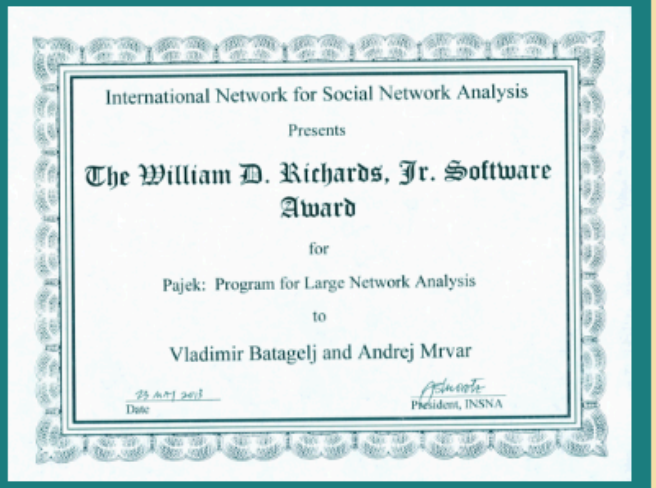
Our layouts for ***Graph-Drawing Competitions***: GD95, GD96, GD97, GD98, GD99, GD00, GD01 and GD05.

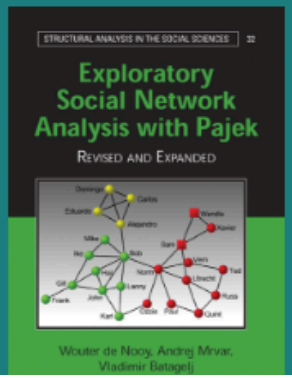
http://vlado.fmf.uni-lj.si/pub/networks/pajek/

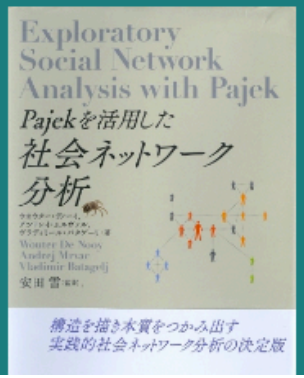
Pajek

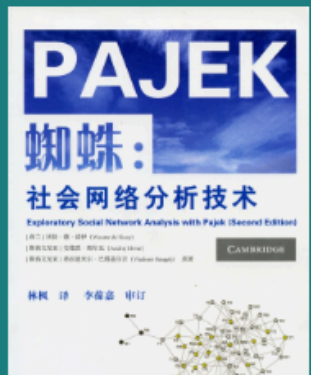
Pajek: analysis and visualization of large networks

	Ver.	32 bit	64 bit
May 10, 2016	4.10	Web Start Pajek Pajek Install Shield Install-Zip Portable	Web Start Pajek Pajek Install Shield Install-Zip Portable
March 1, 2016	4.09	Install Shield Install-Zip Portable	Install Shield Install-Zip Portable
Sept. 25, 2011	2.05	zip	zip
Pajek mailing list		Datasets	









http://mrvar.fdv.uni-lj.si/pajek/

NodeXL

CodePlex Project Hosting for Open Source Software | Register | Sign In | Search all projects

NODEXL Network Graphs
The Social Media Research Foundation

NodeXL: Network Overview, Discovery and Exploration for Excel

HOME | SOURCE CODE | DOWNLOADS | DOCUMENTATION | DISCUSSIONS | ISSUES | PEOPLE | LICENSE

Page Info | Change History (all pages) ★ Follow (365) | Subscribe

socialmedia RESEARCH FOUNDATION
OPEN TOOLS, OPEN DATA, OPEN SCHOLARSHIP FOR SOCIAL MEDIA

Donate

Search Wiki & Documentation

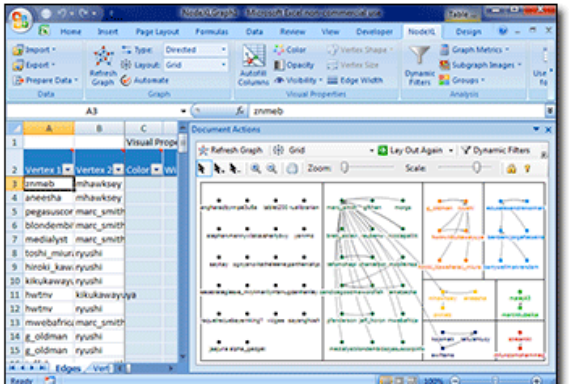
download

CURRENT	NodeXL Basic Excel Template 2014
DATE	Thu Jan 23, 2014 at 7:00 AM
STATUS	Beta
DOWNLOADS	213,252
RATING	★★½☆☆ 17 ratings

Review this release

NodeXL Basic is a free, open-source template for Microsoft® Excel® 2007, 2010, 2013 and 2016 that makes it easy to explore network graphs. With NodeXL, you can enter a network edge list in a worksheet, click a button and see your graph, all in the familiar environment of the Excel window.

NodeXL Pro offers additional features that extend NodeXL Basic, providing easy access to social media network data streams, advanced network metrics, and text and sentiment analysis, and



MOST HELPFUL REVIEWS

★★★☆☆ Install on Windows 10 Pro 64-bit I gives an error message: 'Customized functionality in this application will not work because the c... (more)

https://nodexl.codeplex.com/

Cytoscape

http://www.cytoscape.org/

NetworkX

NetworkX

NetworkX Home | Documentation | Download | Developer (Github)

High-productivity software for complex networks

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Documentation
all documentation

Examples
using the library

Reference
all functions and methods

Features

- Python language data structures for graphs, digraphs, and multigraphs.
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g. text, images, XML records)
- Edges can hold arbitrary data (e.g. weights, time-series)
- Open source BSD license
- Well tested: more than 1800 unit tests, >90% code coverage
- Additional benefits from Python: fast prototyping, easy to teach, multi-platform

Versions

Latest Release

networkx-1.11
30 January 2016
downloads | docs | pdf

Development

2.0dev
github | docs | pdf
build passing
coverage 94%

Contact

Mailing list
Issue tracker

NetworkX Home | Documentation | Download | Developer (Github)

https://networkx.github.io/

igraph

http://igraph.org/redirect.html

SNAP

By *Jure Leskovec* STANFORD UNIVERSITY

- SNAP for C++
- SNAP for Python
- SNAP Datasets
- What's new
- People
- Papers
- Citing SNAP
- Links
- About
- Contact us

Open positions

Open research positions in **SNAP** group are available here.

Stanford Network Analysis Project

SNAP for C++: Stanford Network Analysis Platform

Stanford **N**etwork **A**nalysis **P**latform (**SNAP**) is a general purpose network analysis and graph mining library. It is written in C++ and easily scales to massive networks with hundreds of millions of nodes, and billions of edges. It efficiently manipulates large graphs, calculates structural properties, generates regular and random graphs, and supports attributes on nodes and edges. SNAP is also available through the NodeXL which is a graphical front-end that integrates network analysis into Microsoft Office and Excel.

Snap.py: SNAP for Python

Snap.py is a Python interface for SNAP. It provides performance benefits of SNAP, combined with flexibility of Python. Most of the SNAP C++ functionality is available via Snap.py in Python.

Stanford Large Network Dataset Collection

A collection of more than 50 large network datasets from tens of thousands of nodes and edges to tens of millions of nodes and edges. In includes social networks, web graphs, road networks, internet networks, citation networks, collaboration networks, and communication networks.

Tutorials

Tutorials on using SNAP, on methods to analyze large network data, on ways how to think about networks and how to model them at the level of network structure, and on methods to study evolution and dynamics of diffusion and cascading behavior in networks.

- Tutorial on **Large Scale Network Analytics with SNAP** will be held at WWW-15 conference, Florence, Italy, May 18, 2015. More info.

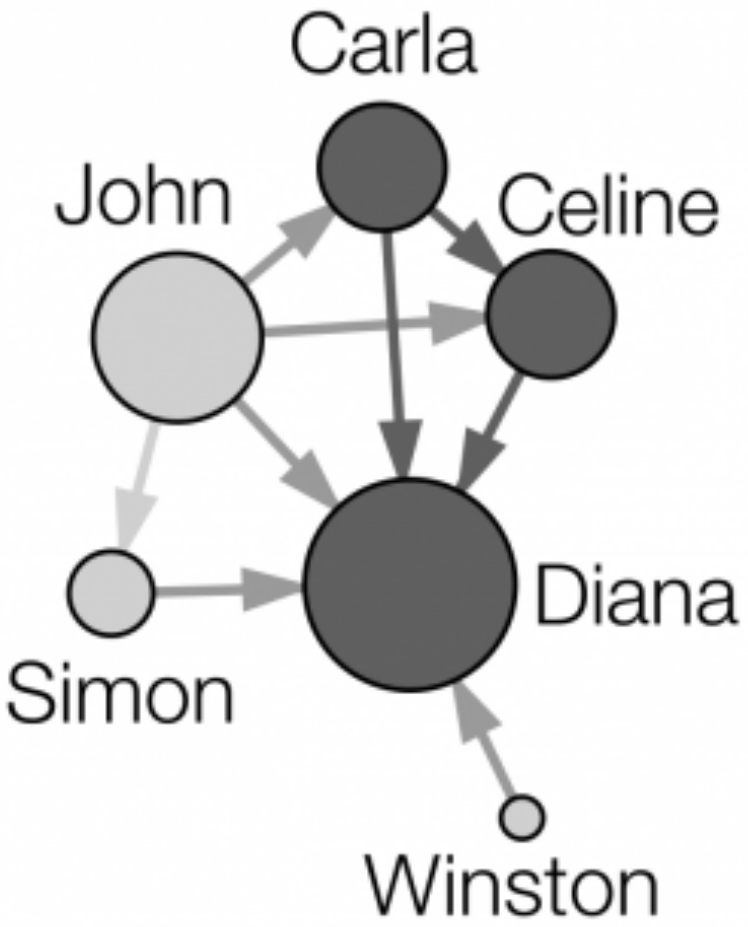
http://snap.stanford.edu/

sigma.js

http://sigmajs.org/

Gephi: Social Network Analysis and Visualization

Network Analysis and Visualization with Gephi



Nodes

Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2

Edges

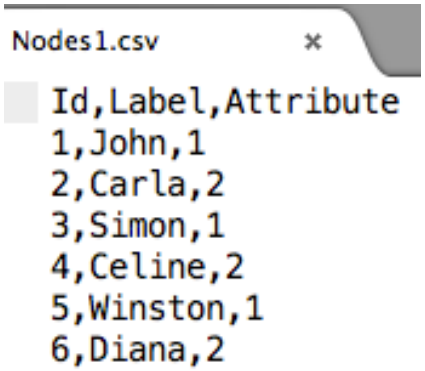
Source,Target
1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6

Source: http://www.martingrandjean.ch/gephi-introduction/

Nodes and Edges
CSV Text Data for Gephi

Nodes1.csv

```
Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2
```



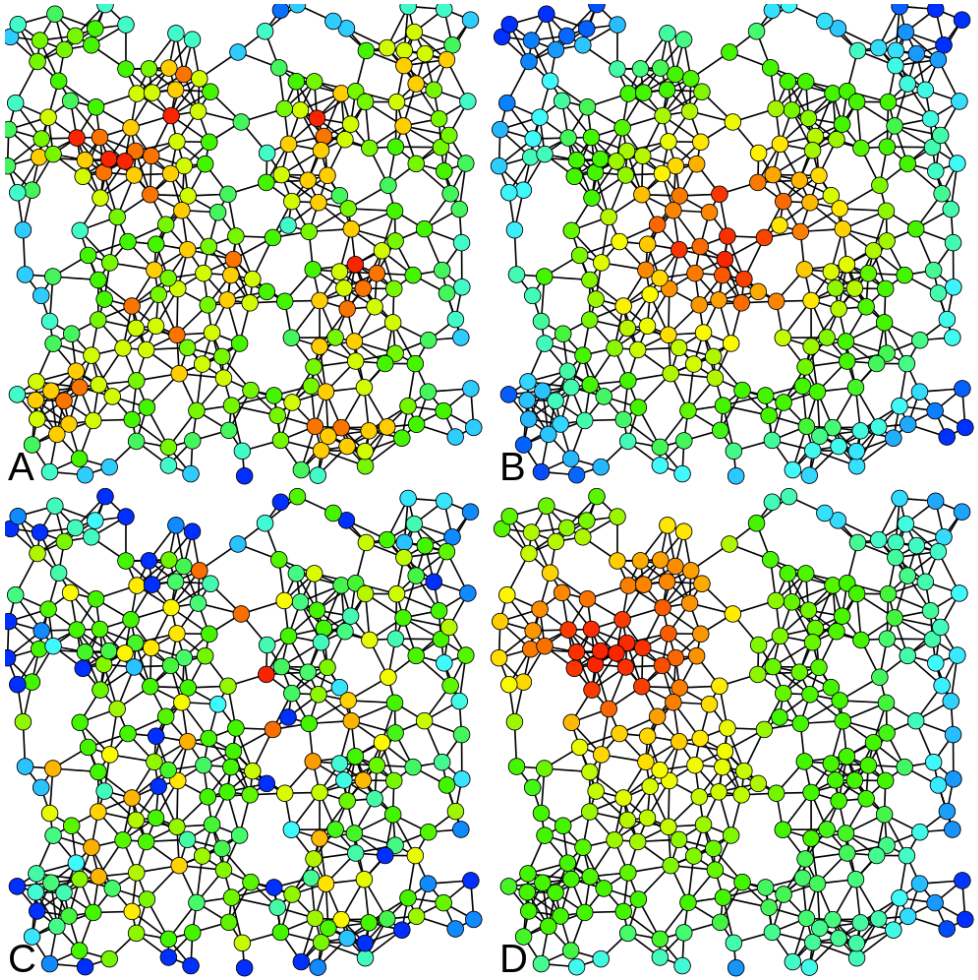
Edges1.csv

```
Source,Target
1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6
```

Source: http://www.martingrandjean.ch/gephi-introduction/

A = Degree centrality
number of connexions

B = Closeness centrality
closeness to the entire network

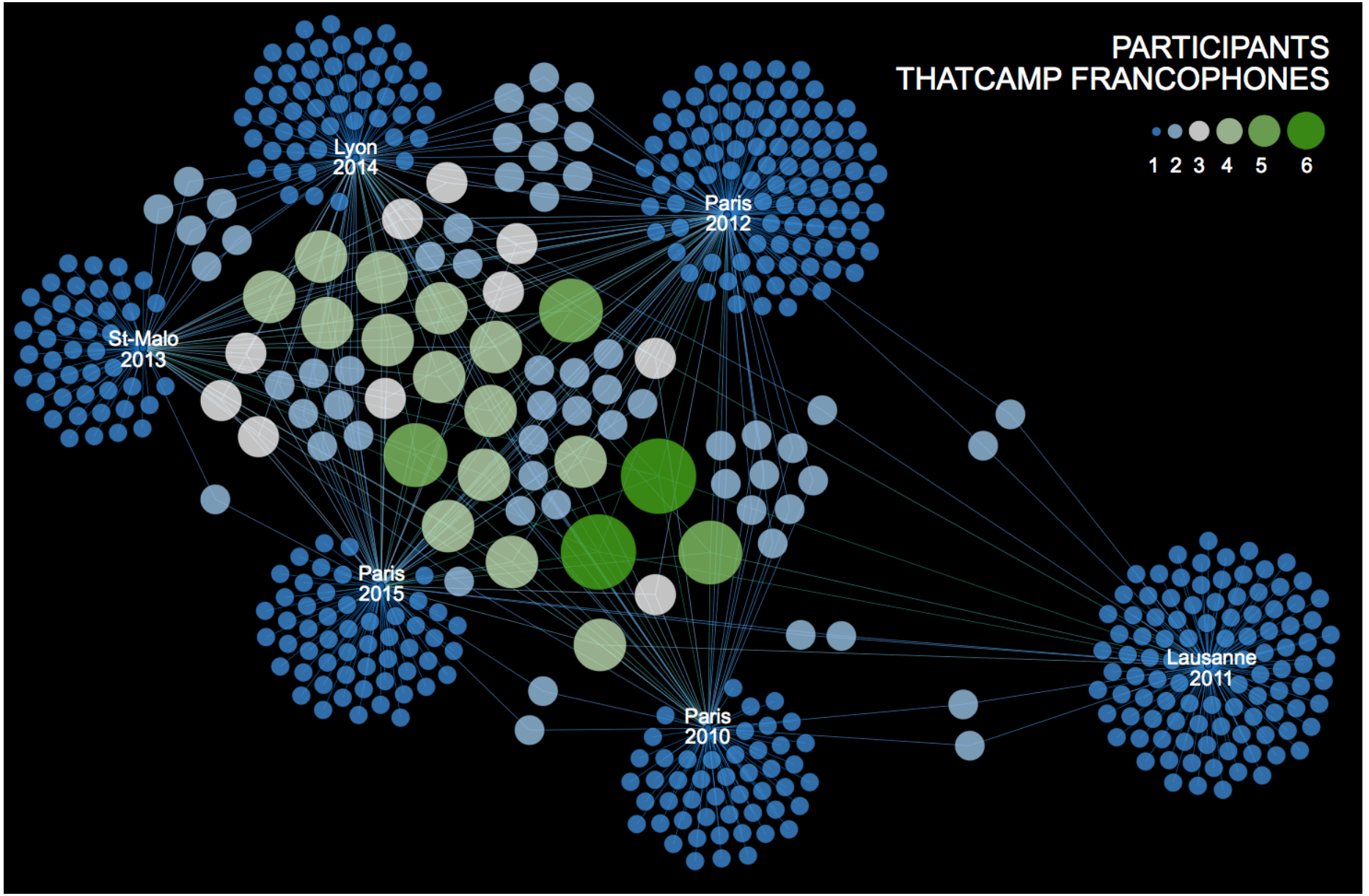


C = Betweenness centrality
bridges nodes

D = Eigenvector centrality
connection to well-connected nodes

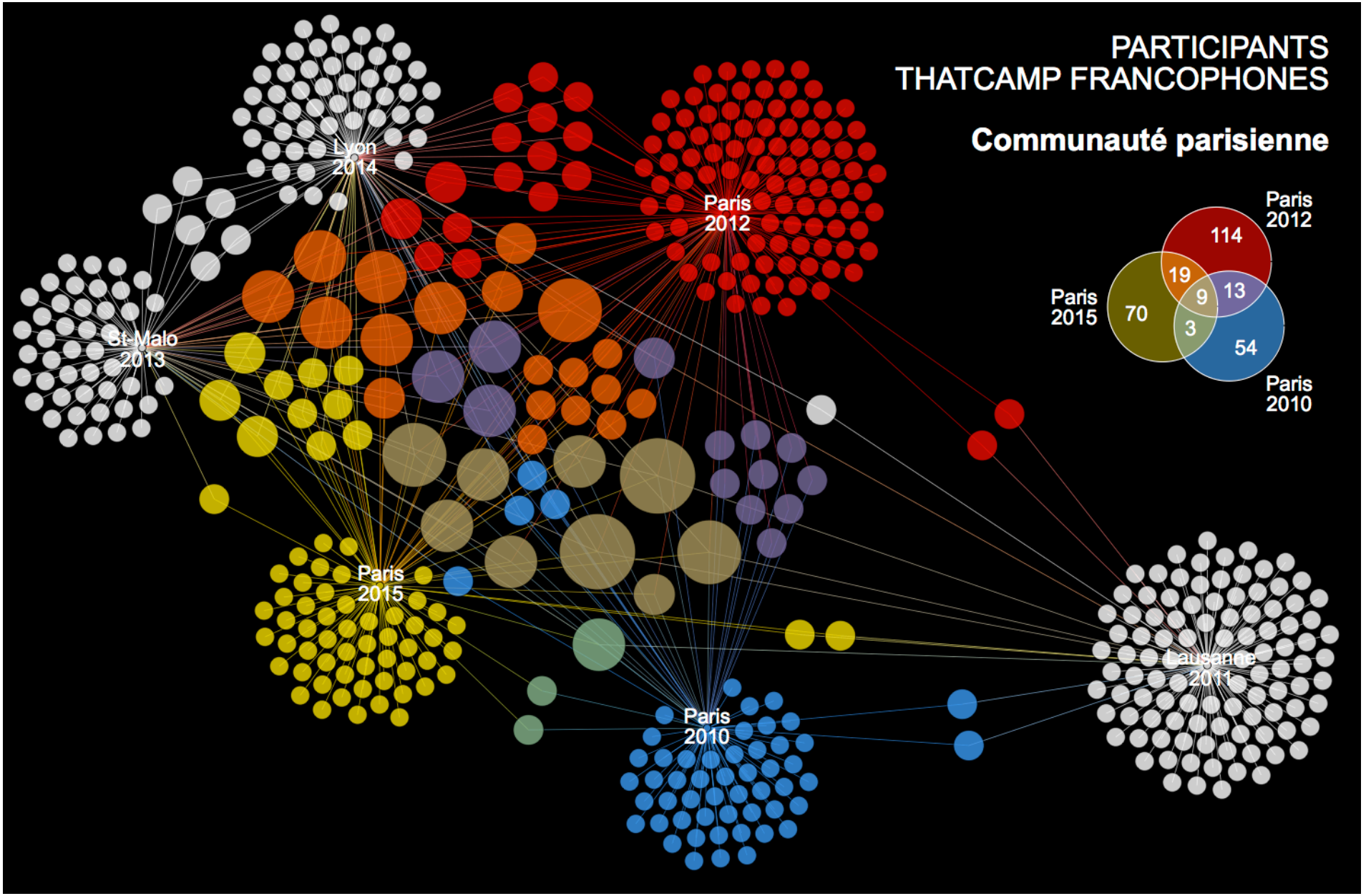
Source: http://www.martingrandjean.ch/gephi-introduction/

Conference Participants



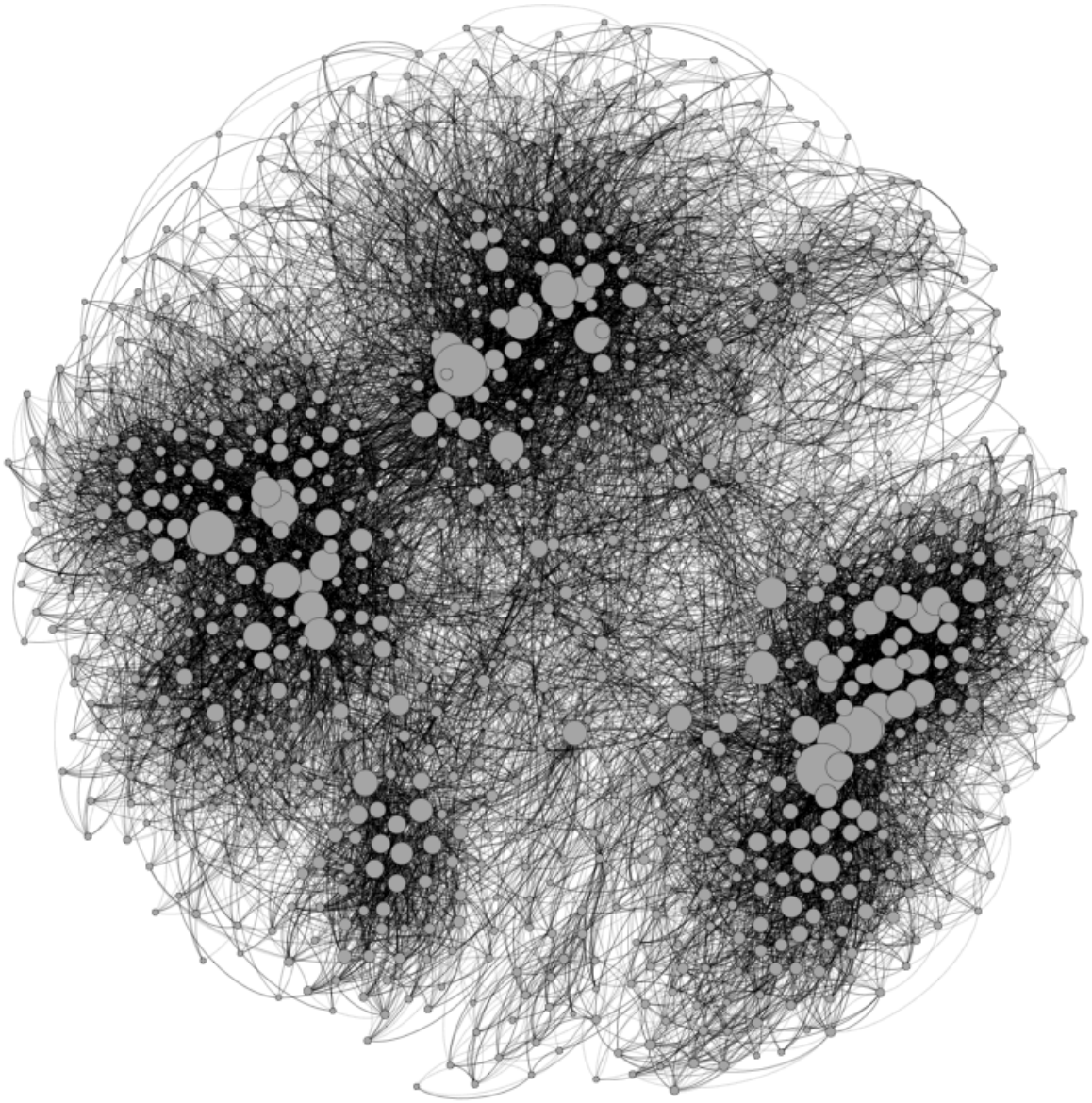
Source: http://www.martingrandjean.ch/analyse-de-reseau-thatcamp-et-communaute-des-humanites-numeriques-francophones/

Conference Participants



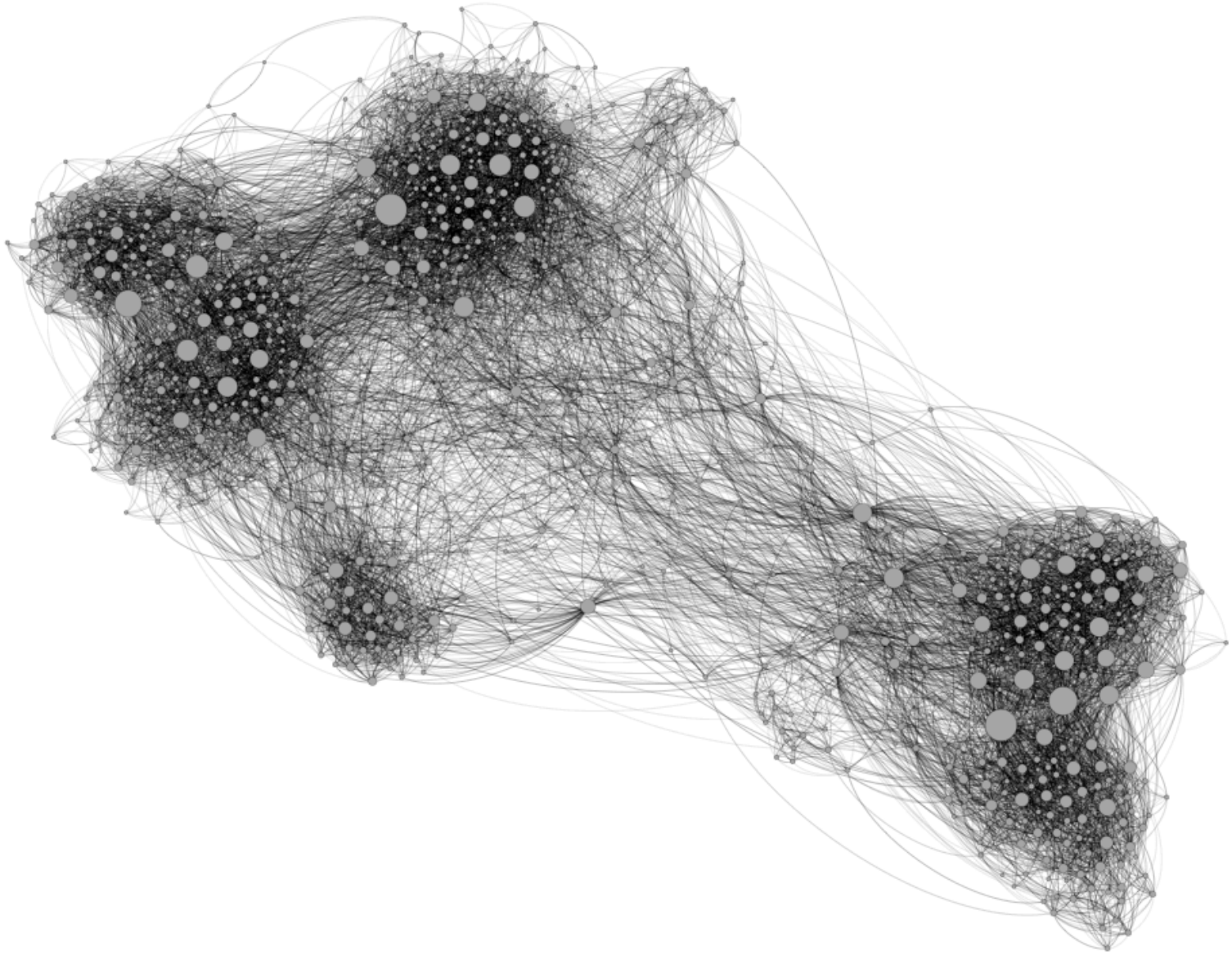
Source: http://www.martingrandjean.ch/analyse-de-reseau-thatcamp-et-communaute-des-humanites-numeriques-francophones/

Fruchterman Reingold



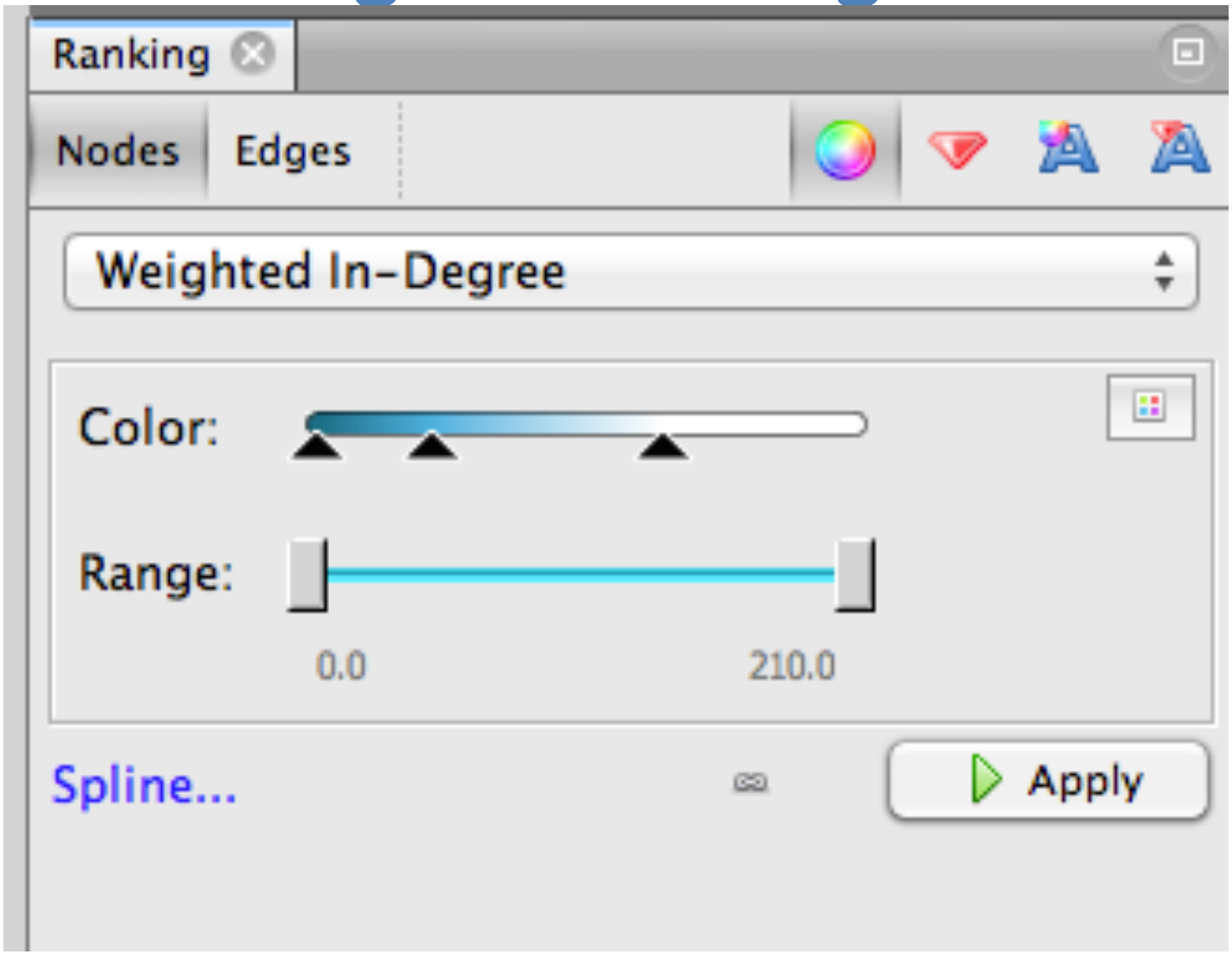
Source: http://www.martingrandjean.ch/gephi-introduction/

Force Atlas 2



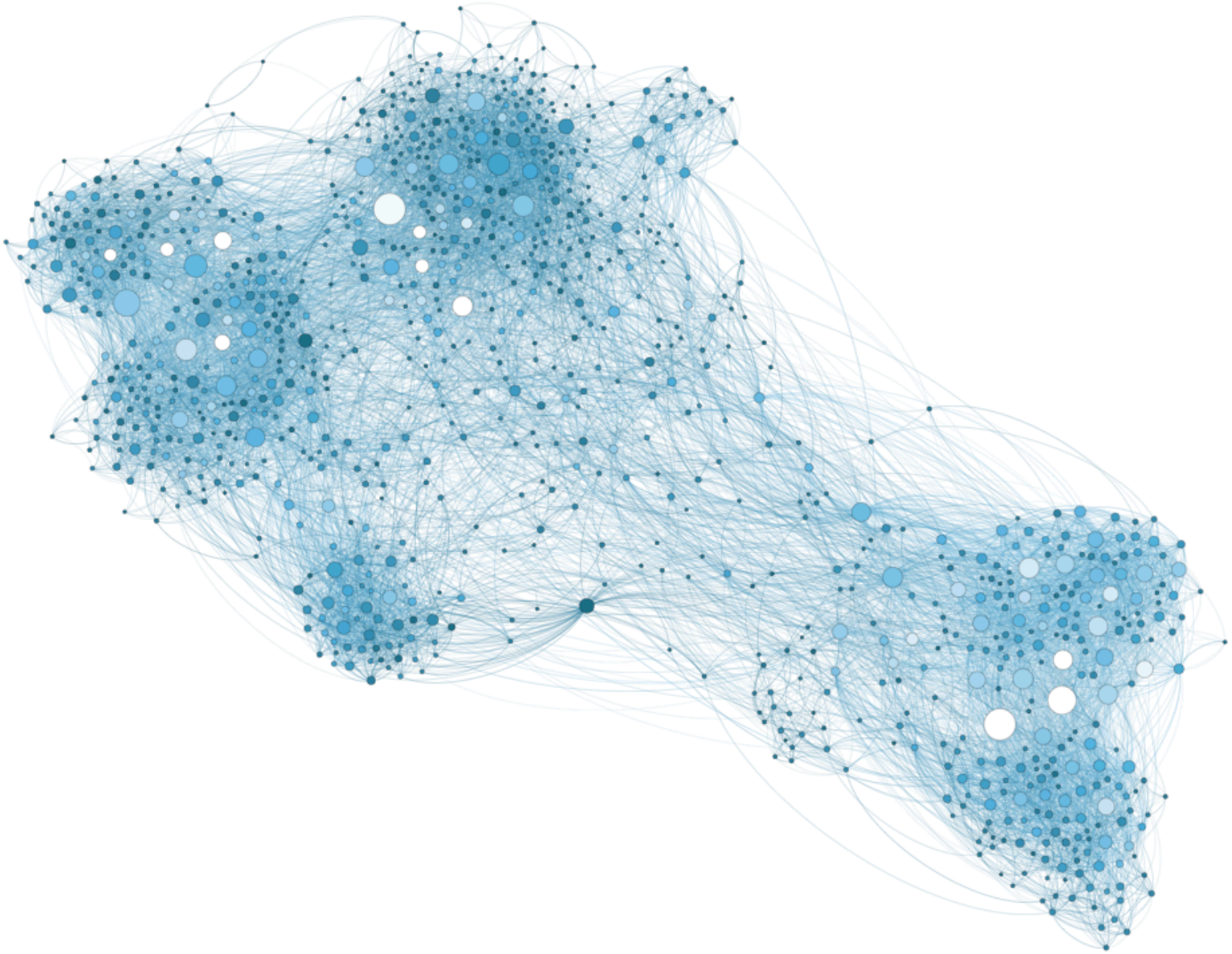
Source: http://www.martingrandjean.ch/gephi-introduction/

Nodes' color
Weighted In-Degree



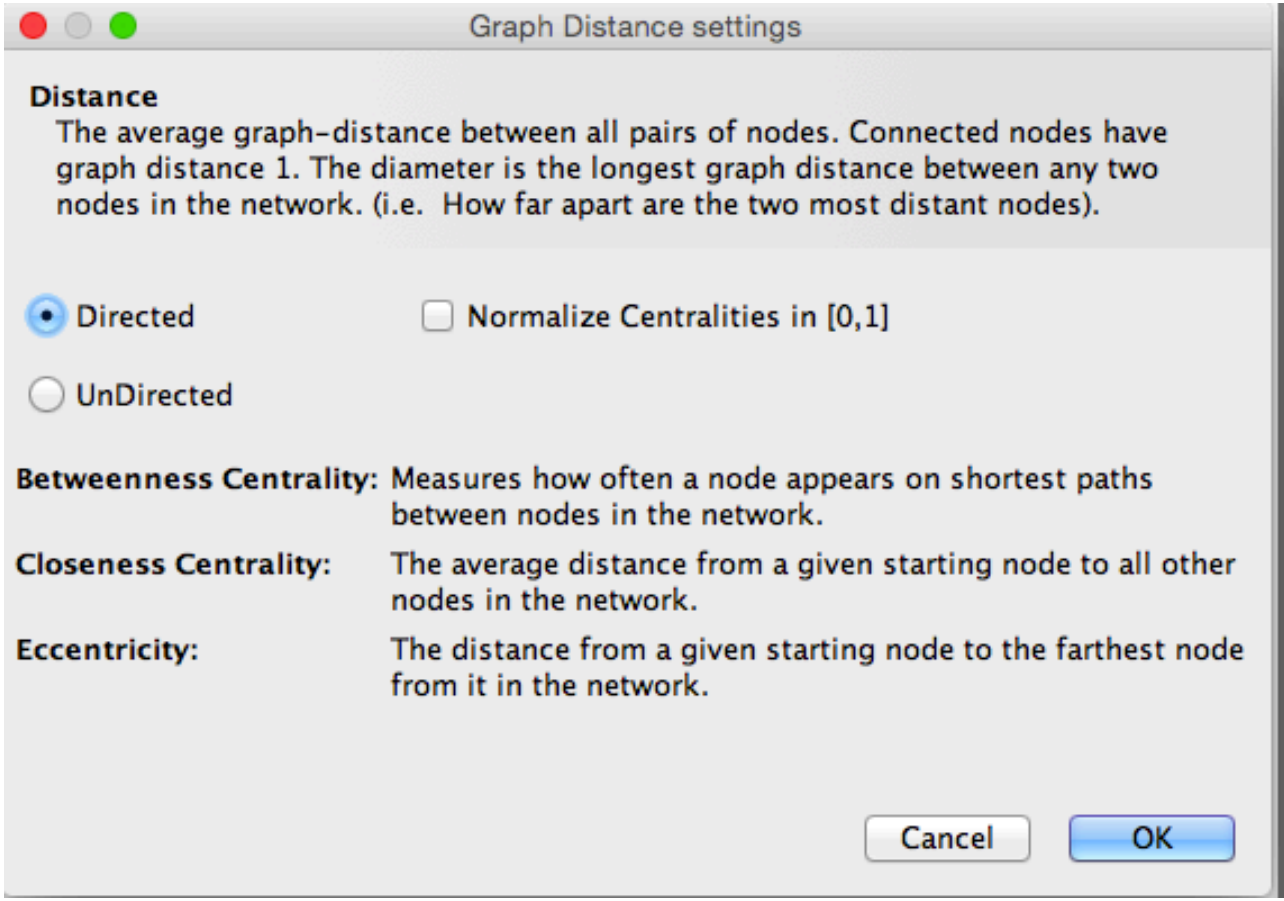
Source: http://www.martingrandjean.ch/gephi-introduction/

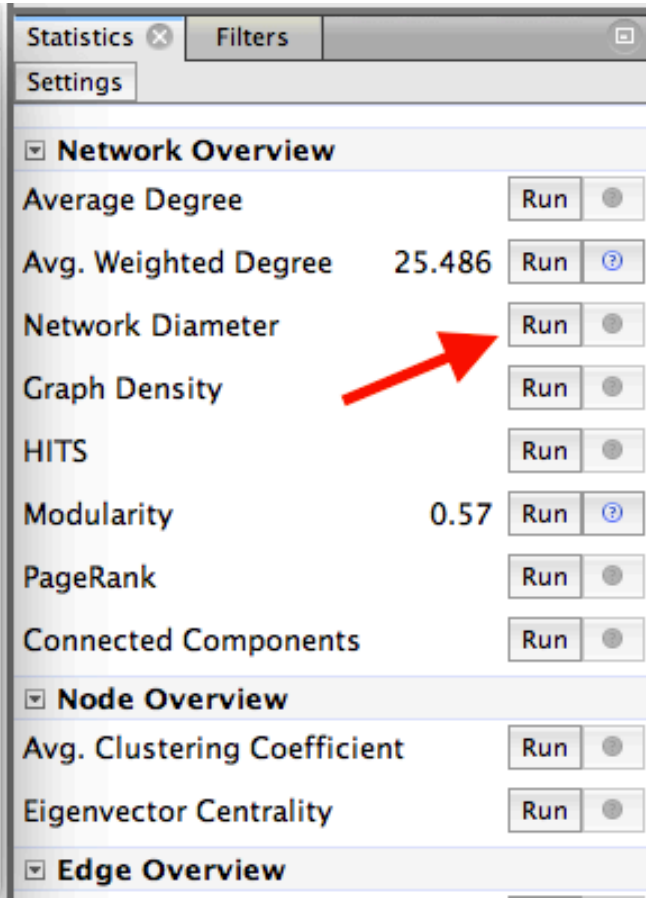
Weighted In-Degree



Source: http://www.martingrandjean.ch/gephi-introduction/

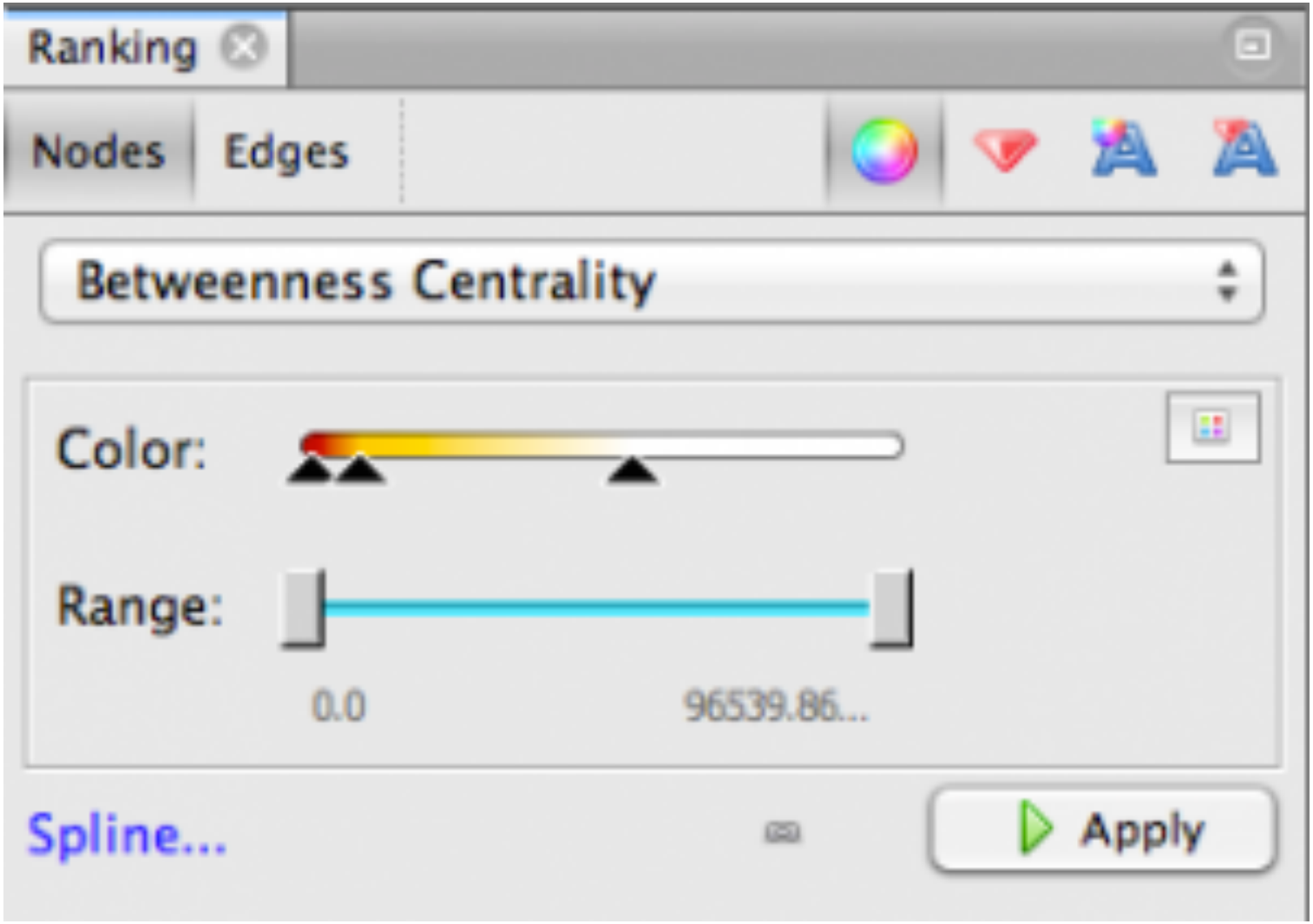
Network Diameter
Betweenness Centrality
Closeness Centrality





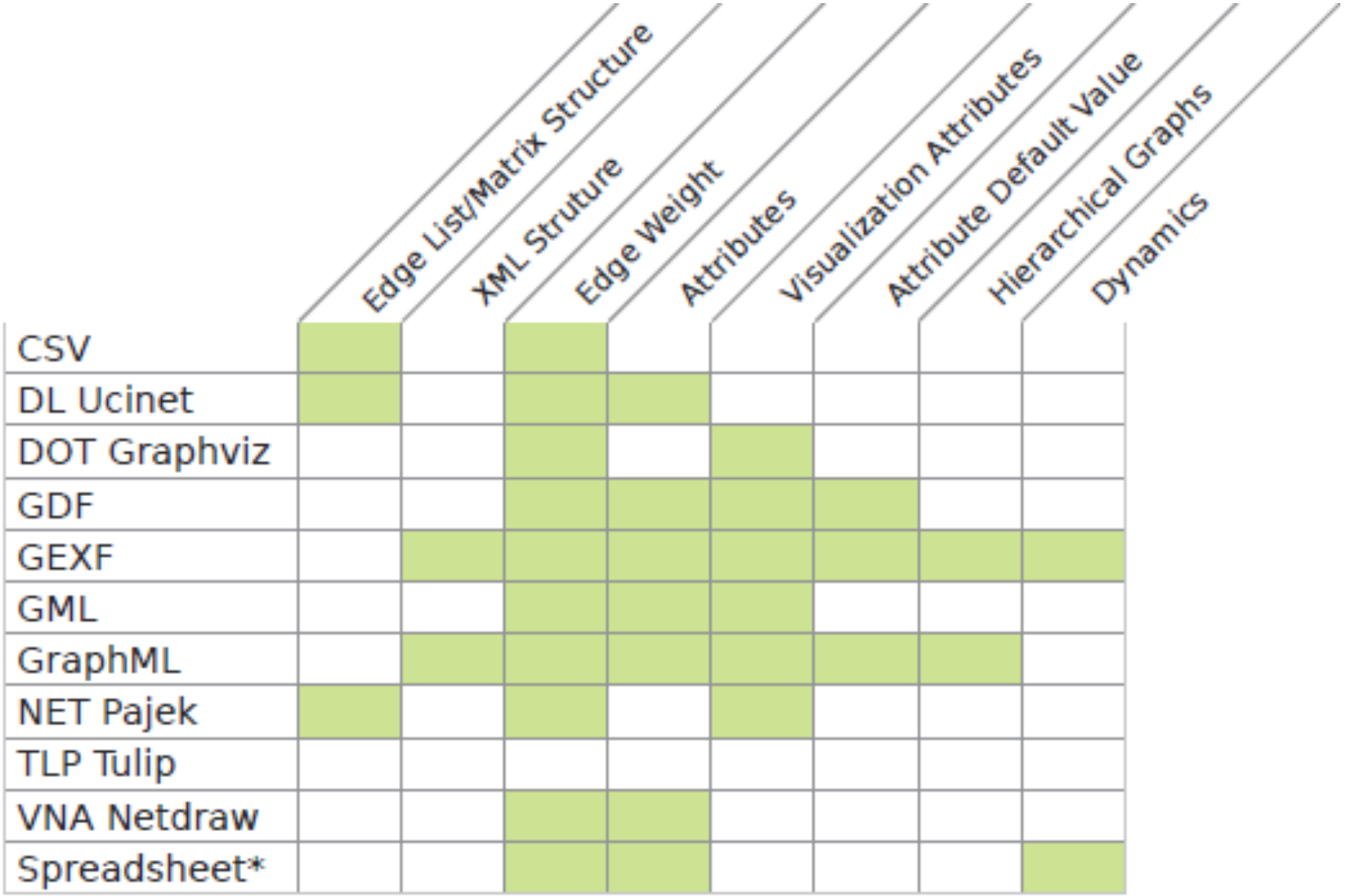
Source: http://www.martingrandjean.ch/gephi-introduction/

Nodes' color
Betweenness Centrality



Source: http://www.martingrandjean.ch/gephi-introduction/

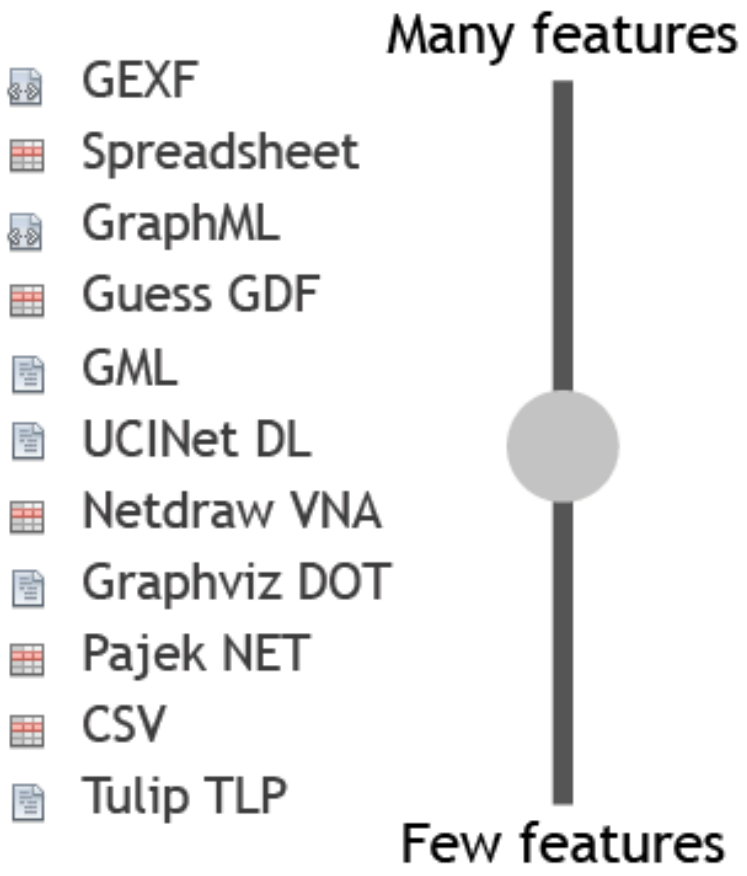
Gephi Supported Graph Formats



Source: https://gephi.org/users/supported-graph-formats/

Gephi Supported Graph Formats

Do you need...



File Type
- XML
- Tabular
- Text

Source: https://gephi.org/users/supported-graph-formats/

Gephi

Gephi
makes graphs handy

Download Blog Wiki Forum Support Bug tracker

Home Features Learn Develop Plugins Services Consortium

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

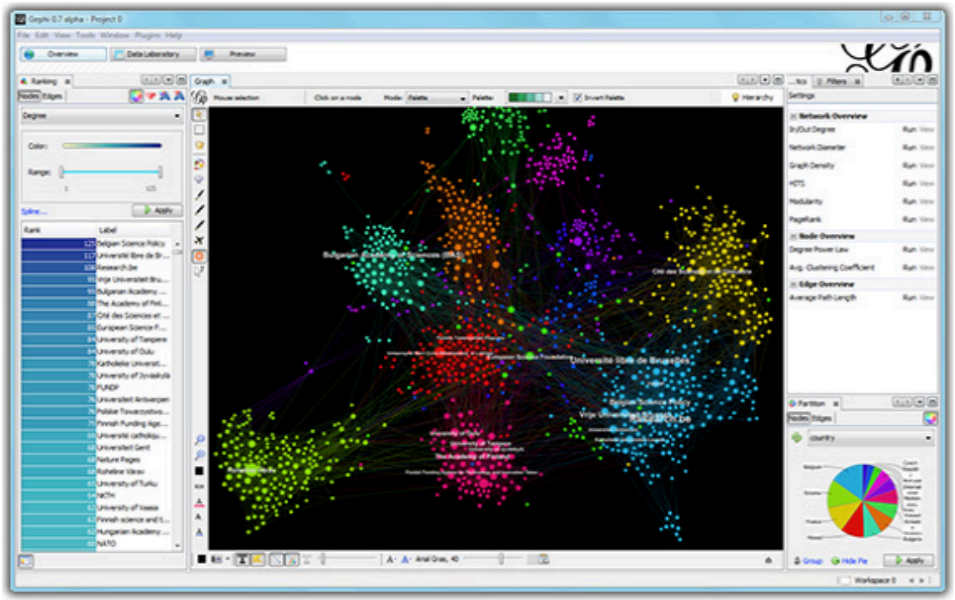
Runs on Windows, Mac OS X and Linux.

Learn More on Gephi Platform »

Download FREE
Gephi 0.9.1

Release Notes | System Requirements

- Features
- Quick start
- Screenshots
- Videos



Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:

Donate

APPLICATIONS

- **Exploratory Data Analysis**: intuition-oriented analysis by networks manipulations in real time.
- **Link Analysis**: revealing the underlying structures of associations between objects.
- **Social Network Analysis**: easy creation of social

Like Photoshop™ for graphs.

— the Community

LATEST NEWS

Gephi updates with 0.9.1 version

PAPERS

https://gephi.org/

Download Gephi

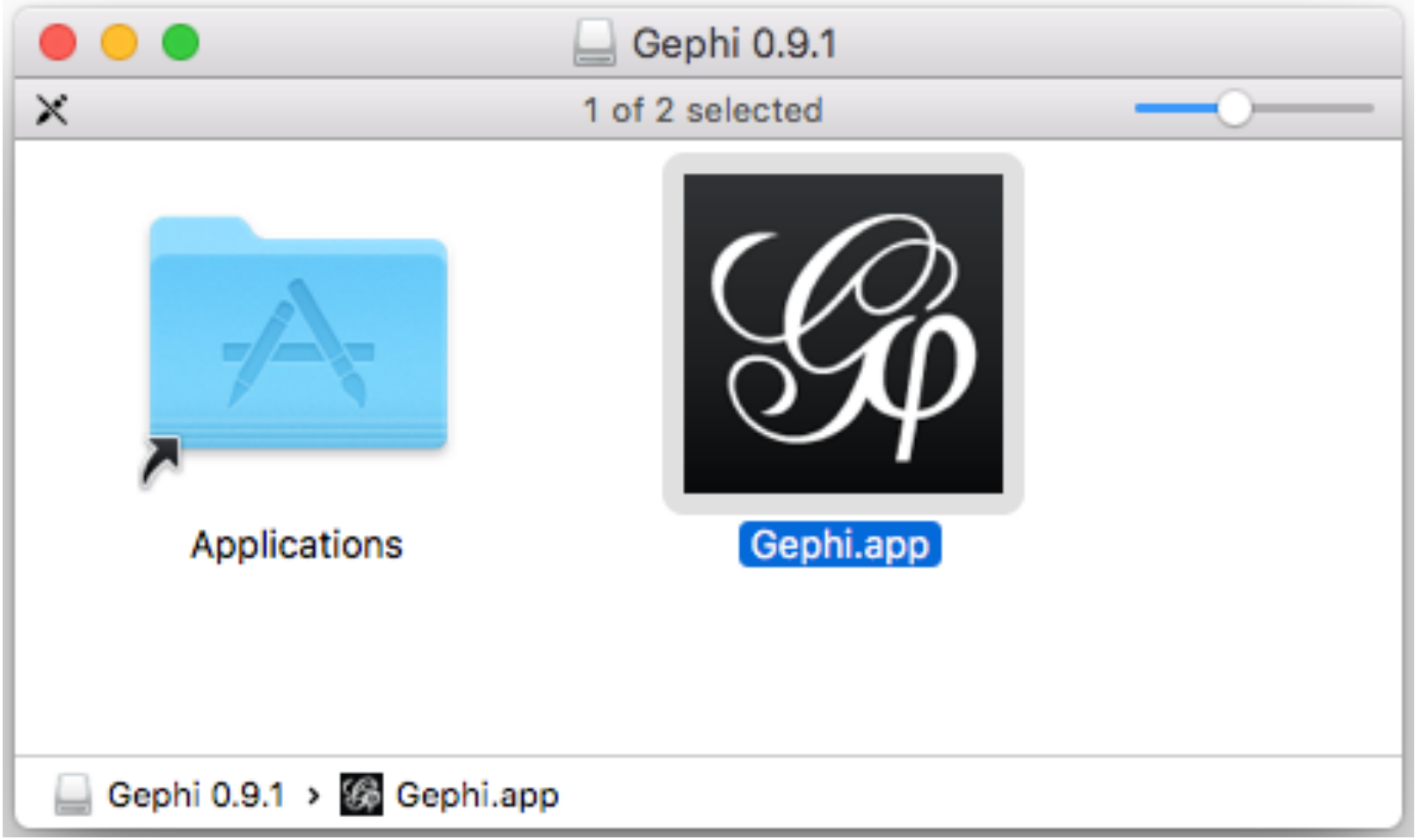
https://gephi.org/users/download/

Download Gephi

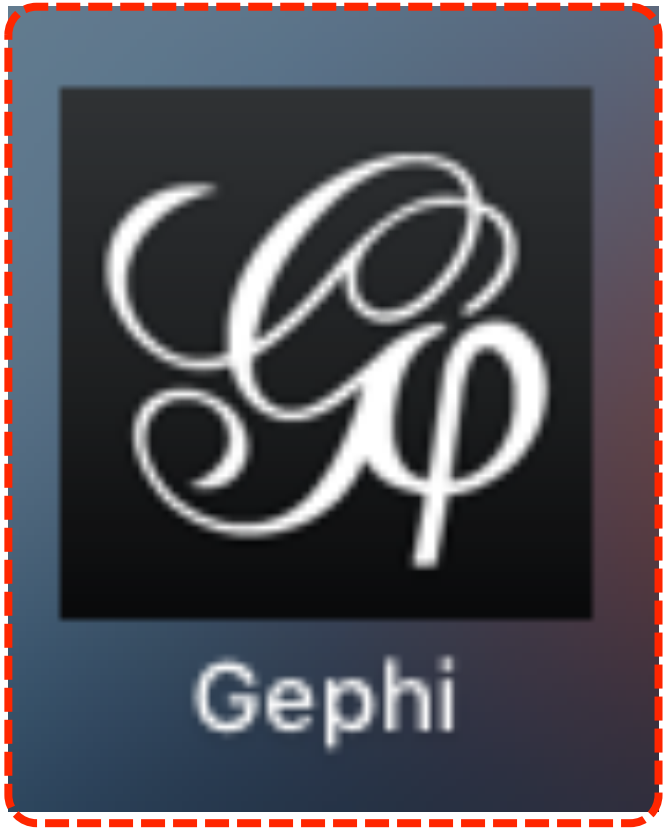
gephi-0.9.1-macos.dmg

Disk Image - 121.1 MB

Gephi 0.9.1

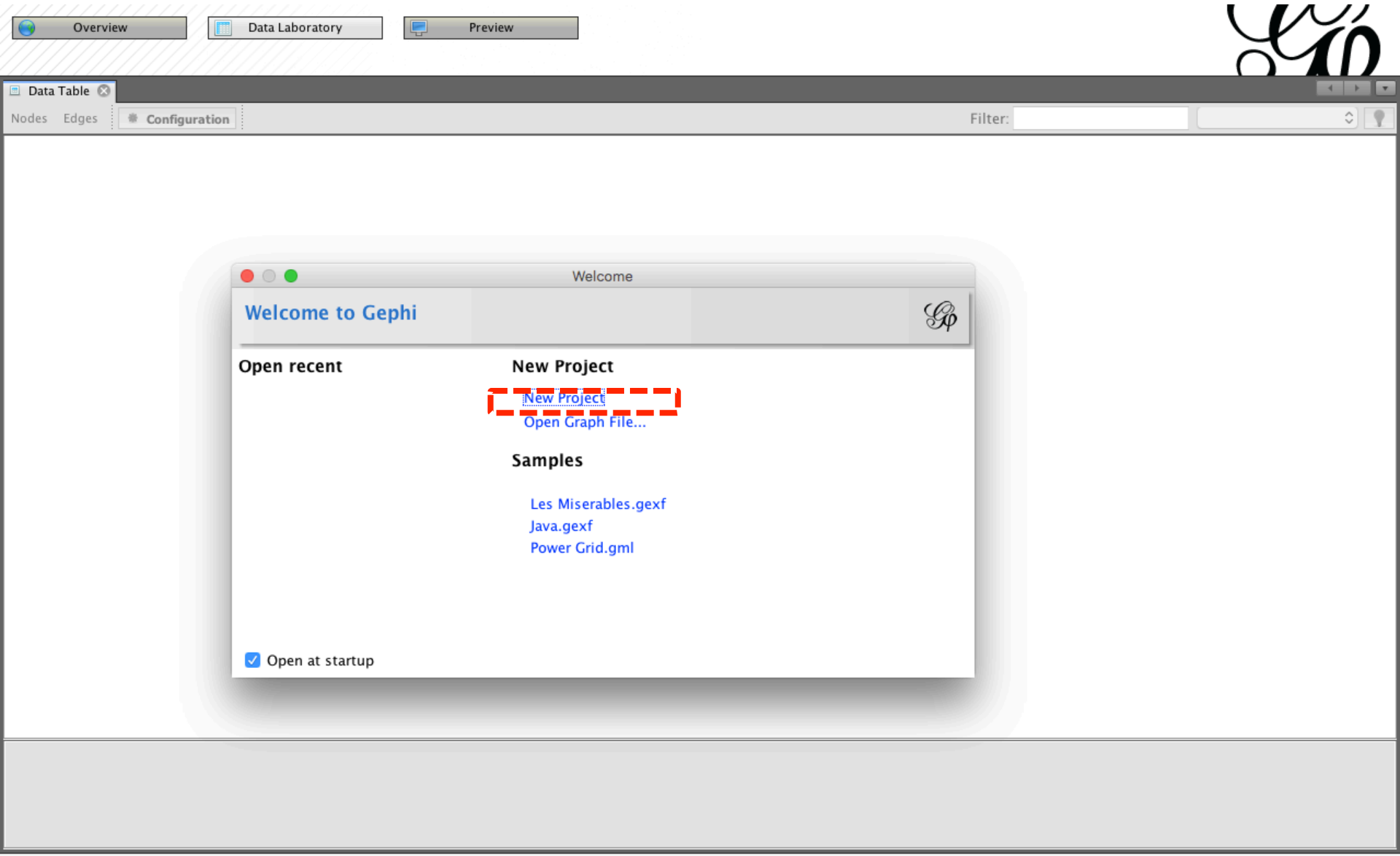

Gephi 0.9.1
1 of 2 selected
Applications
Gephi.app
Gephi 0.9.1 › Gephi.app

Gephi

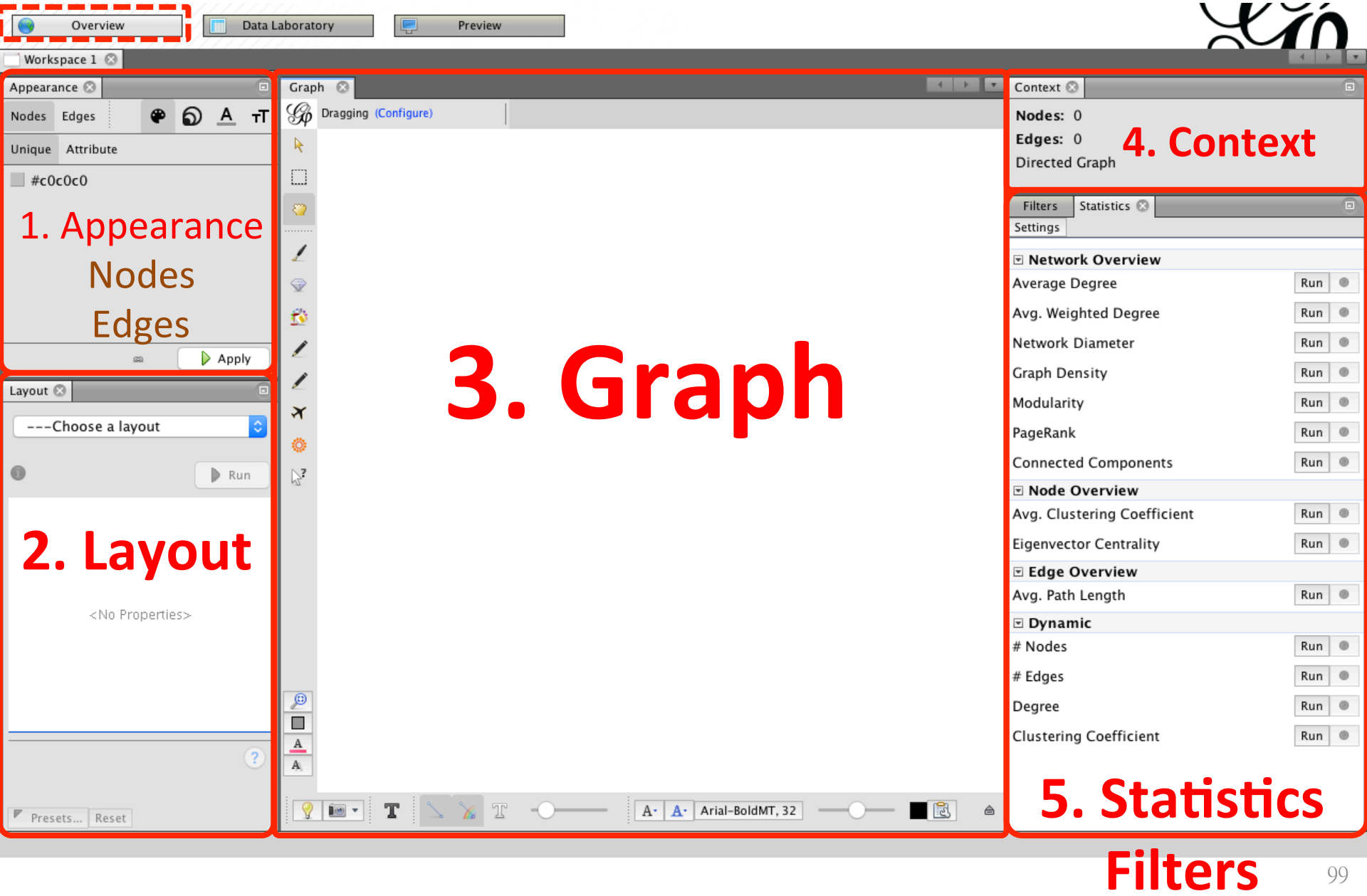


Gephi:
New Project
Import
Nodes1.csv and
Edges1.csv
to Gephi

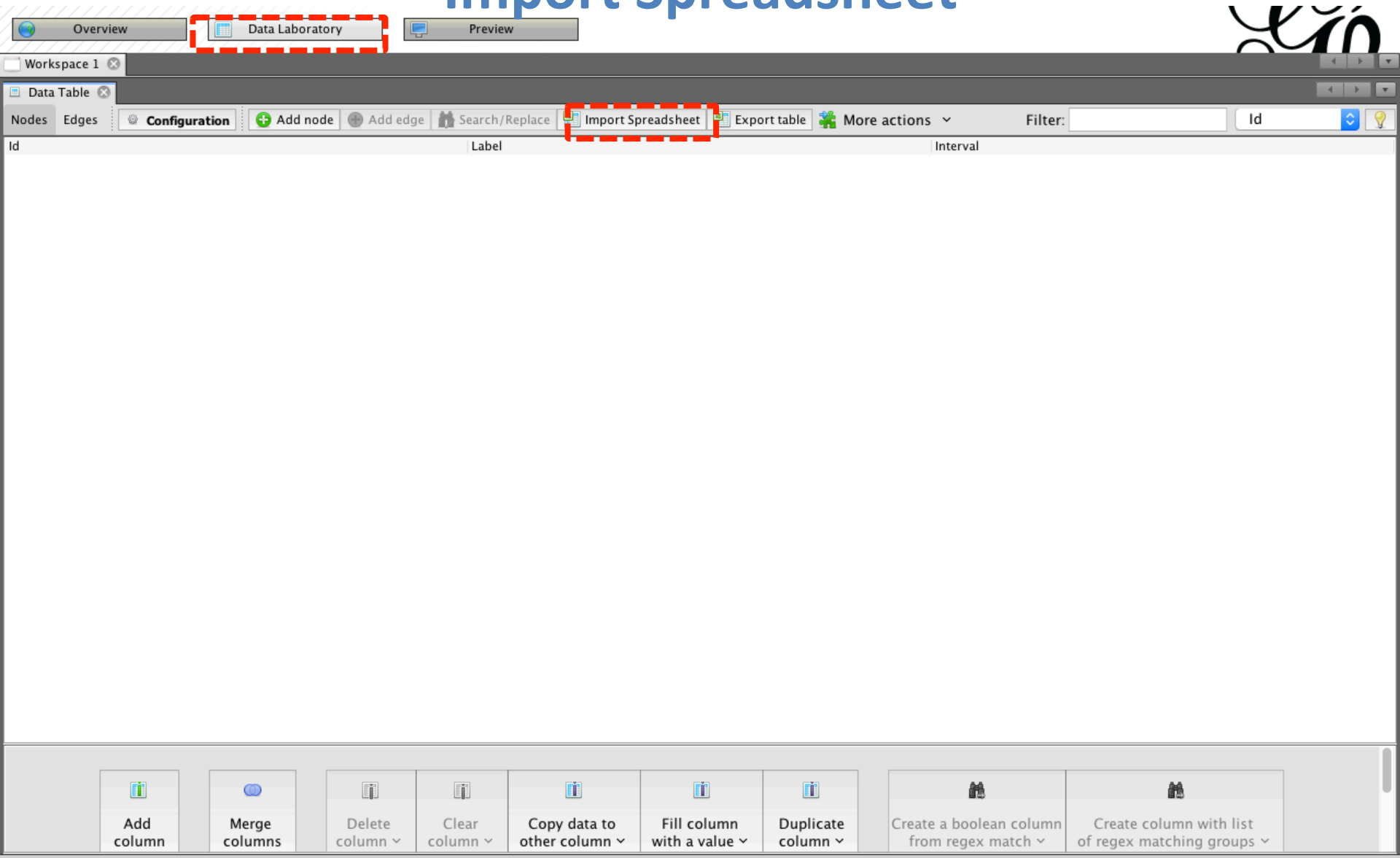
Gephi New Project



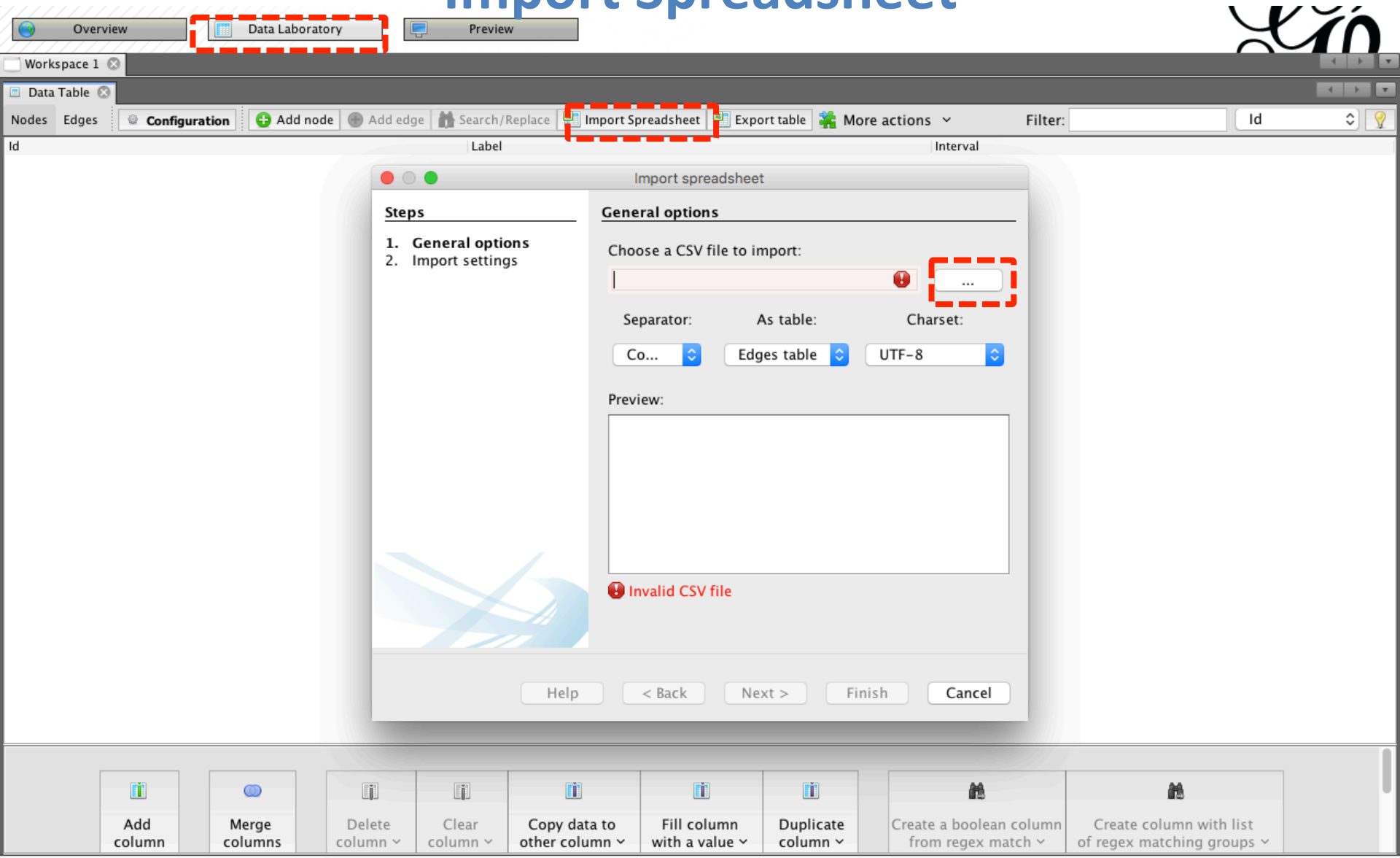
Gephi Overview



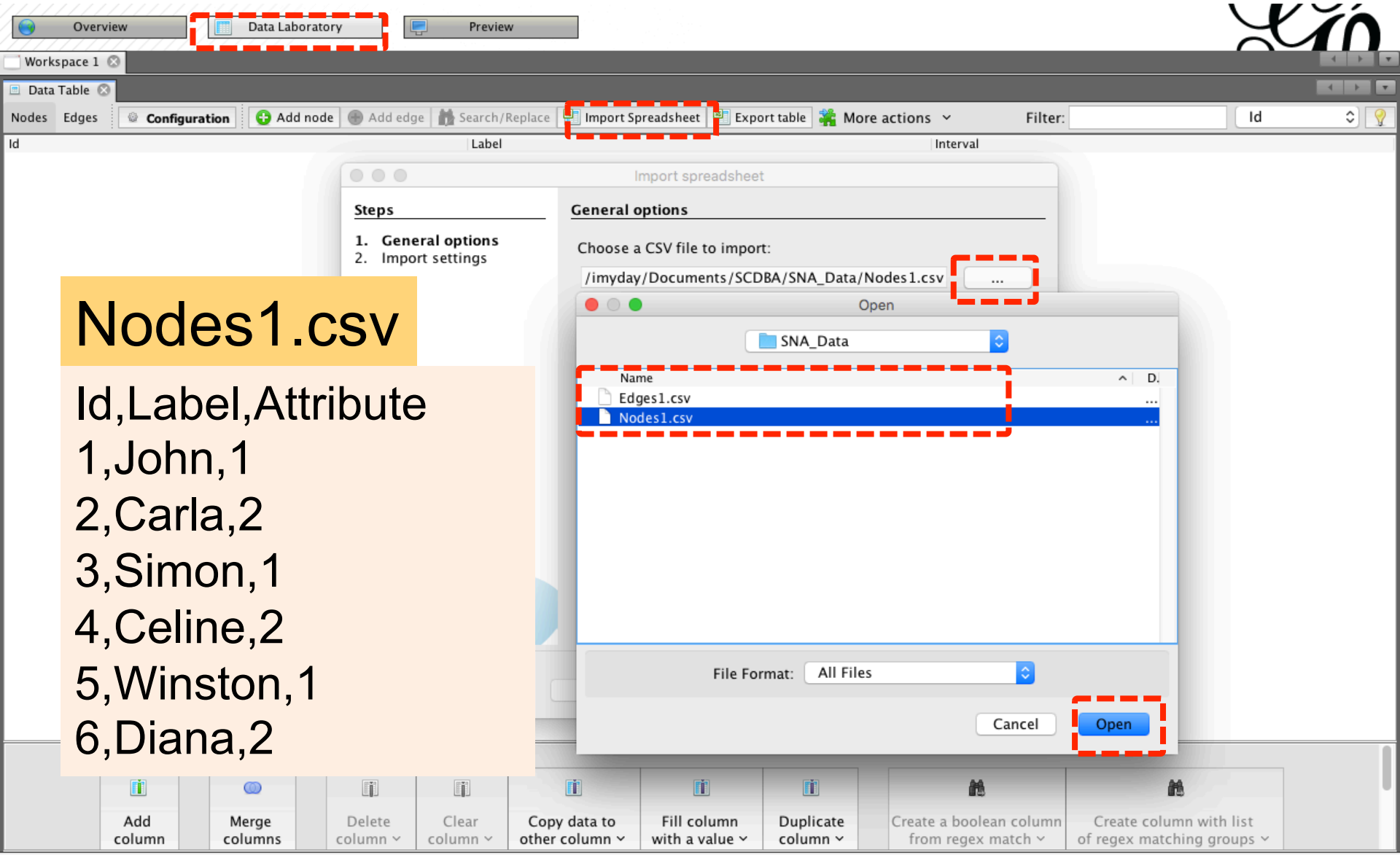
Gephi Data Laboratory:
Import Spreadsheet



Gephi Data Laboratory:
Import Spreadsheet



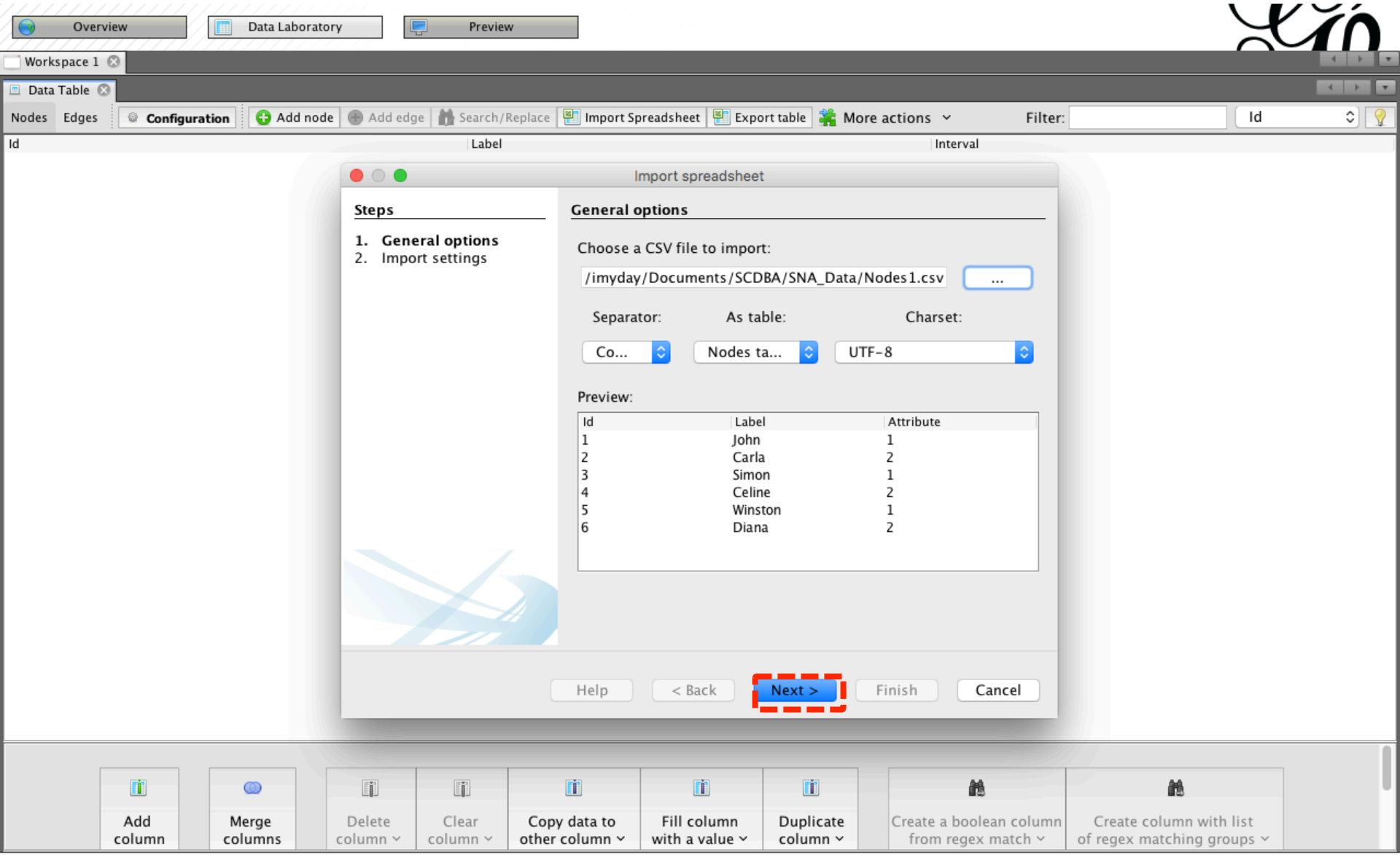
Import Nodes1.csv to Gephi



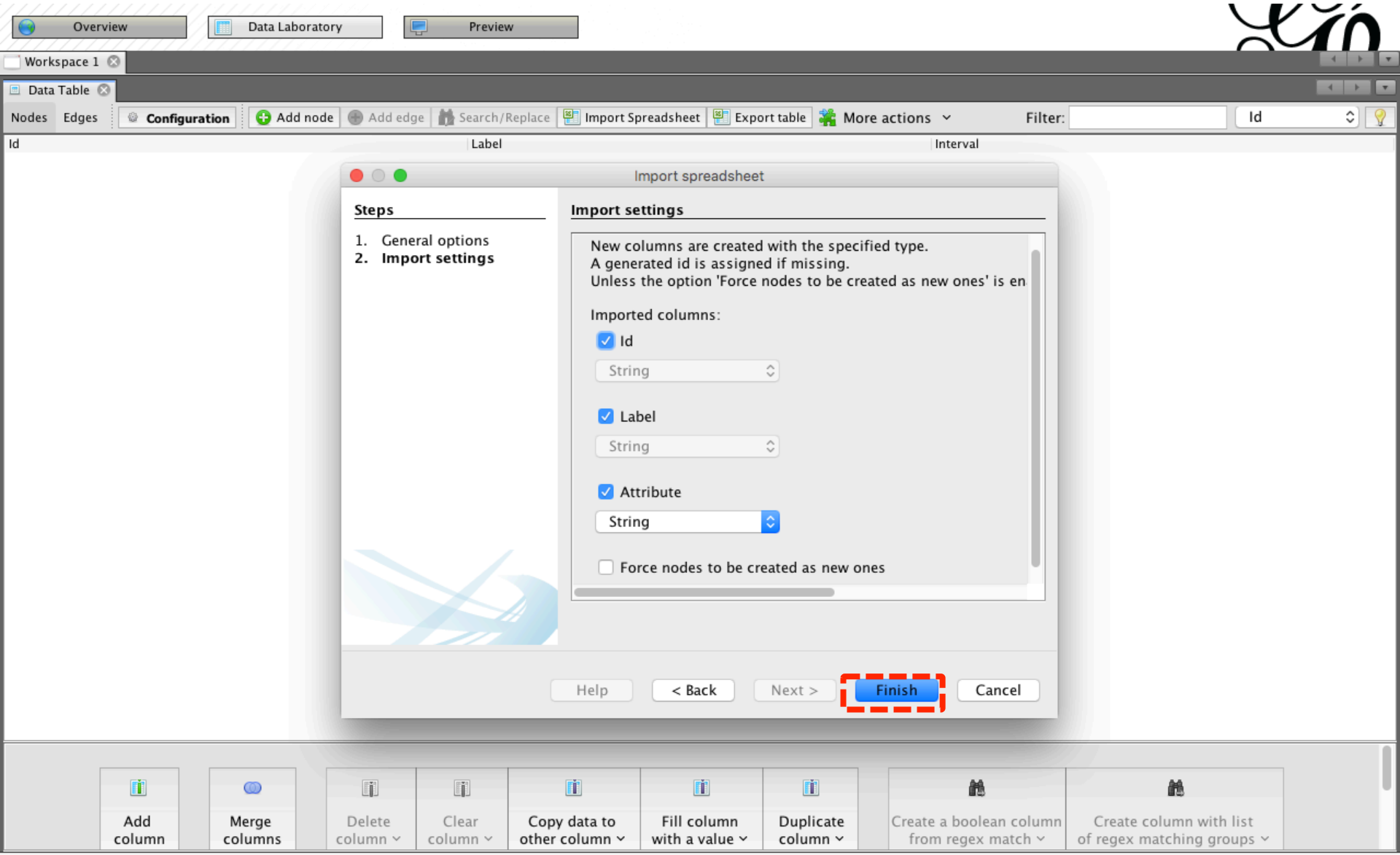
Nodes1.csv

```
Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2
```

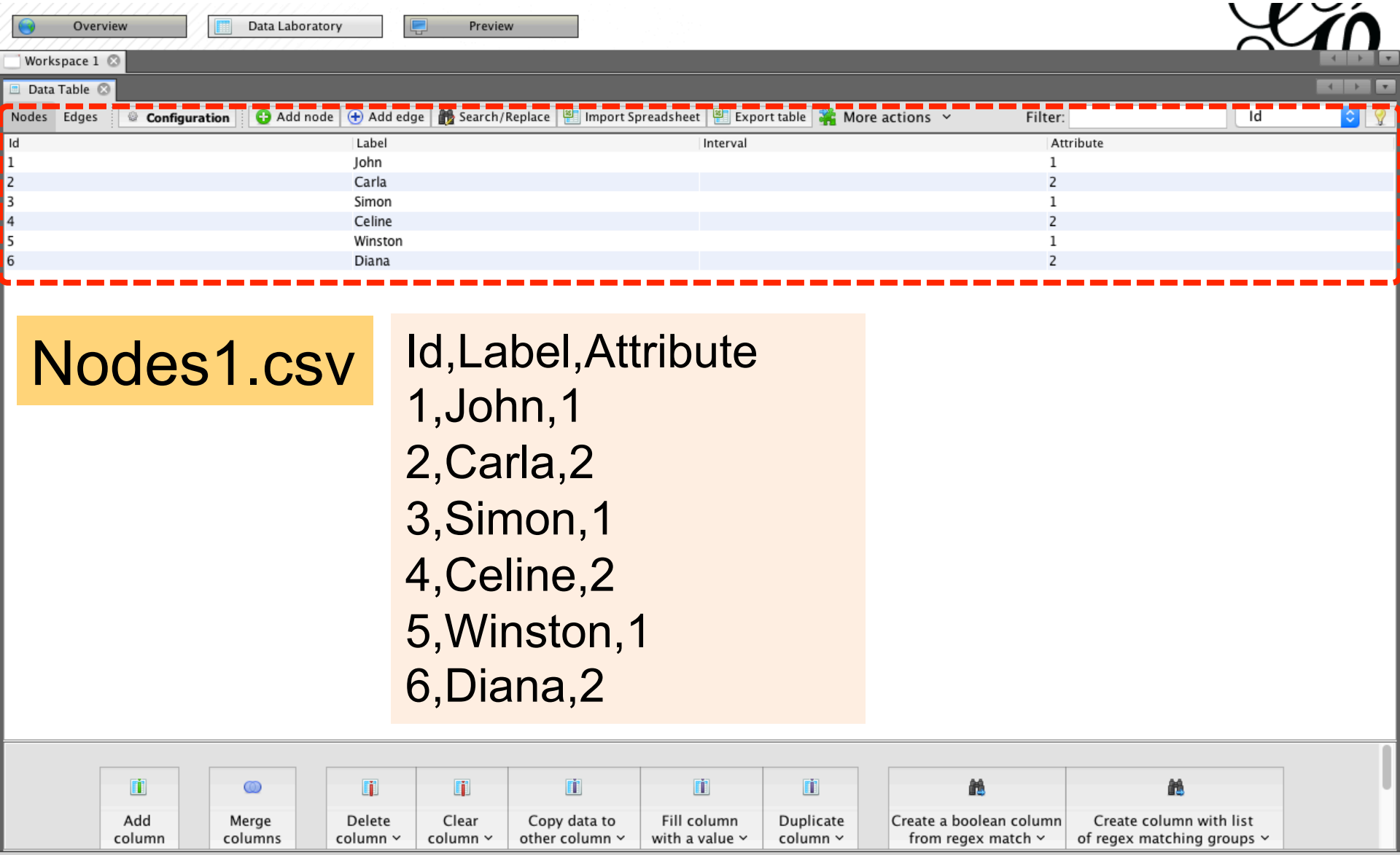
Import Nodes1.csv to Gephi



Import Nodes1.csv to Gephi



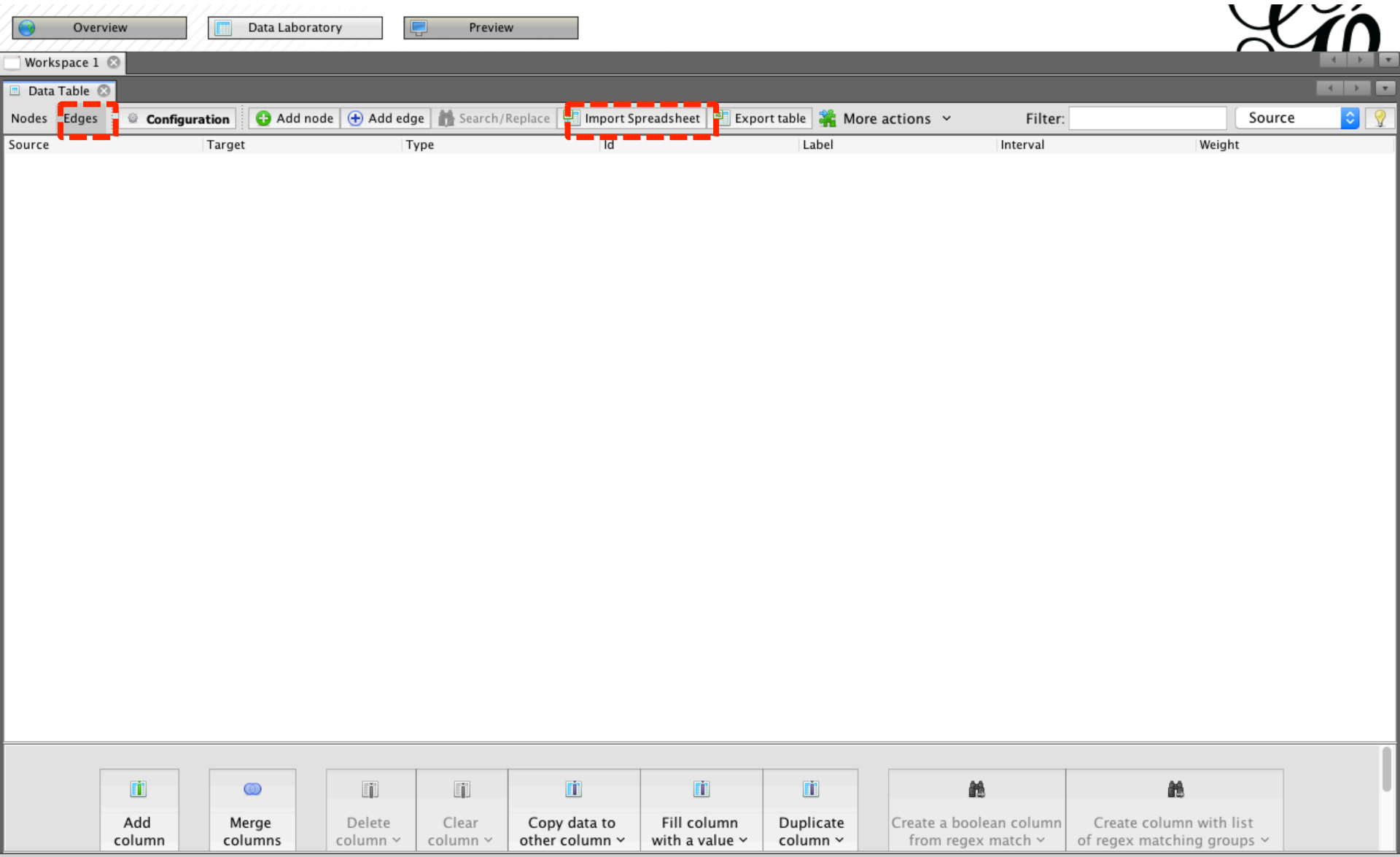
Import Nodes1.csv to Gephi



Nodes1.csv

```
Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2
```

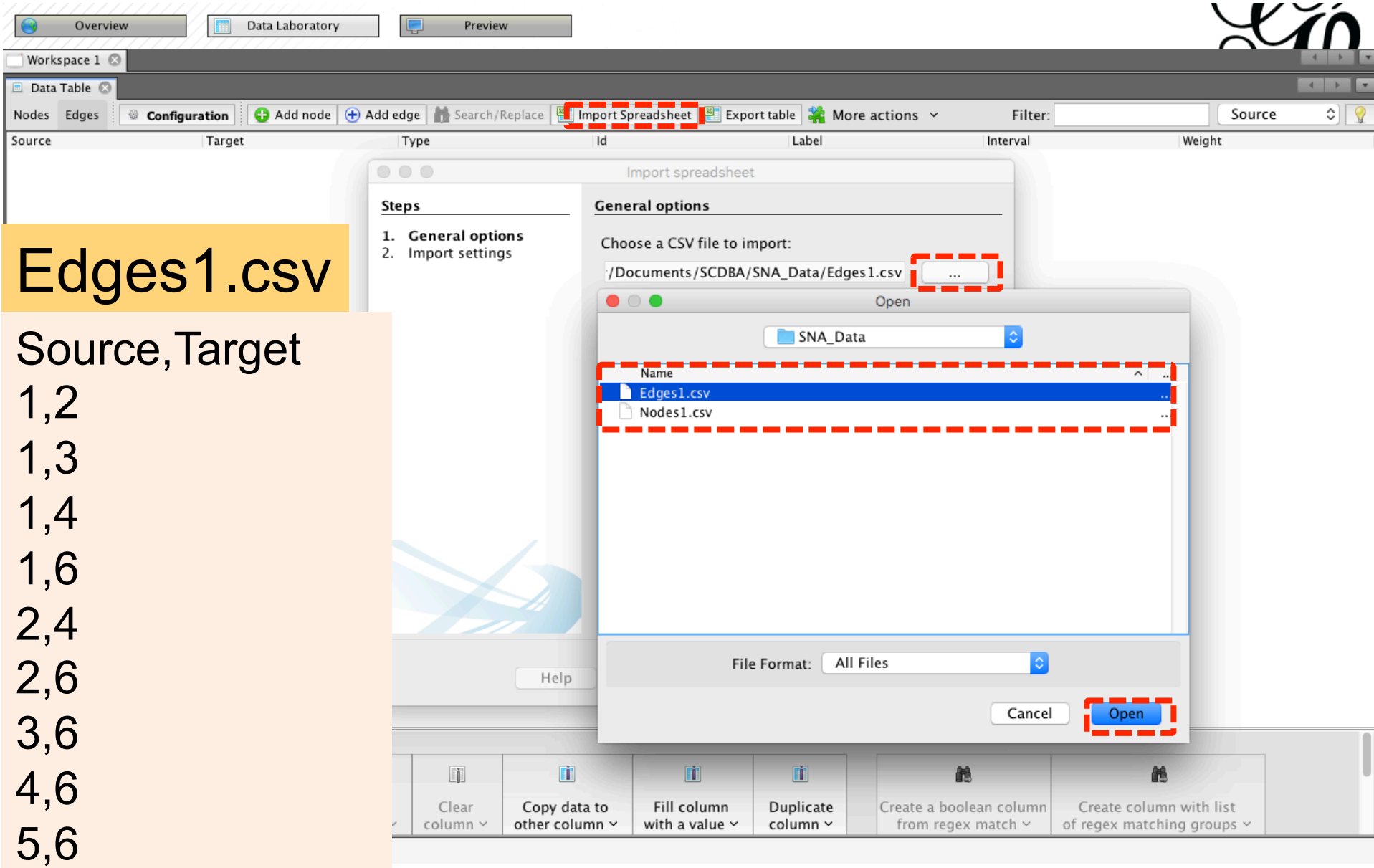
Import Edges1.csv to Gephi



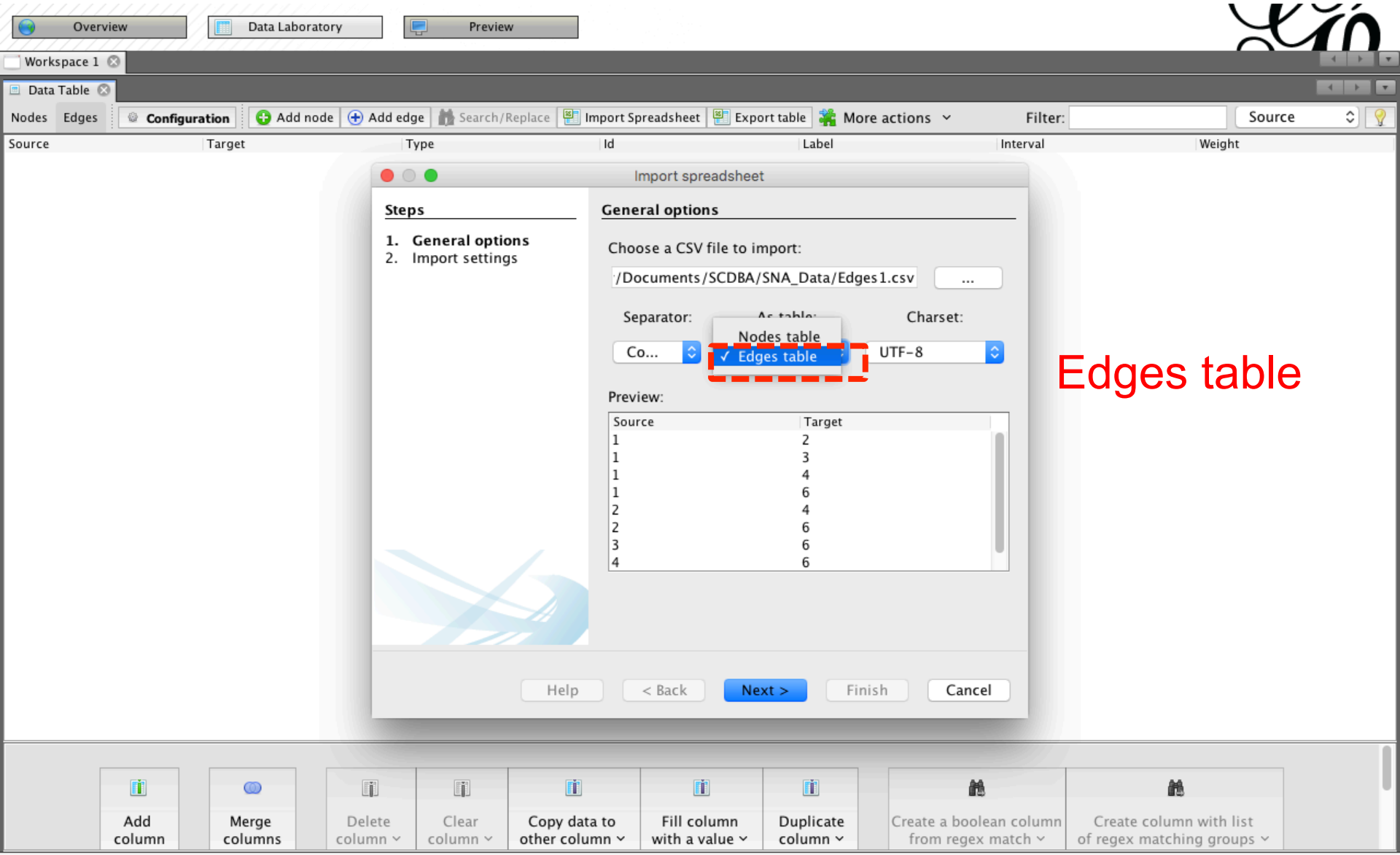
Import Edges1.csv to Gephi

Edges1.csv

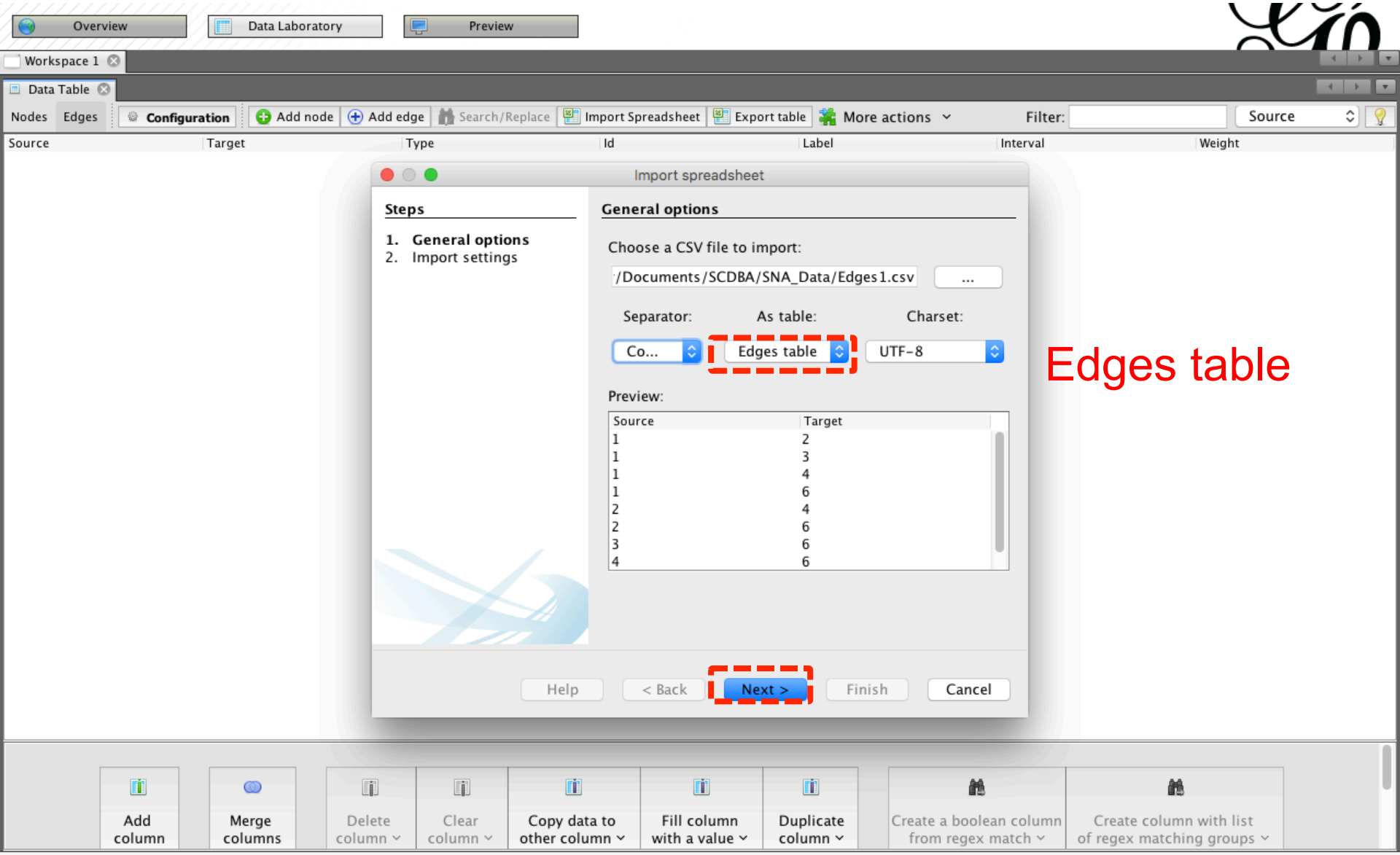
Source,Target
1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6



Import Edges1.csv to Gephi

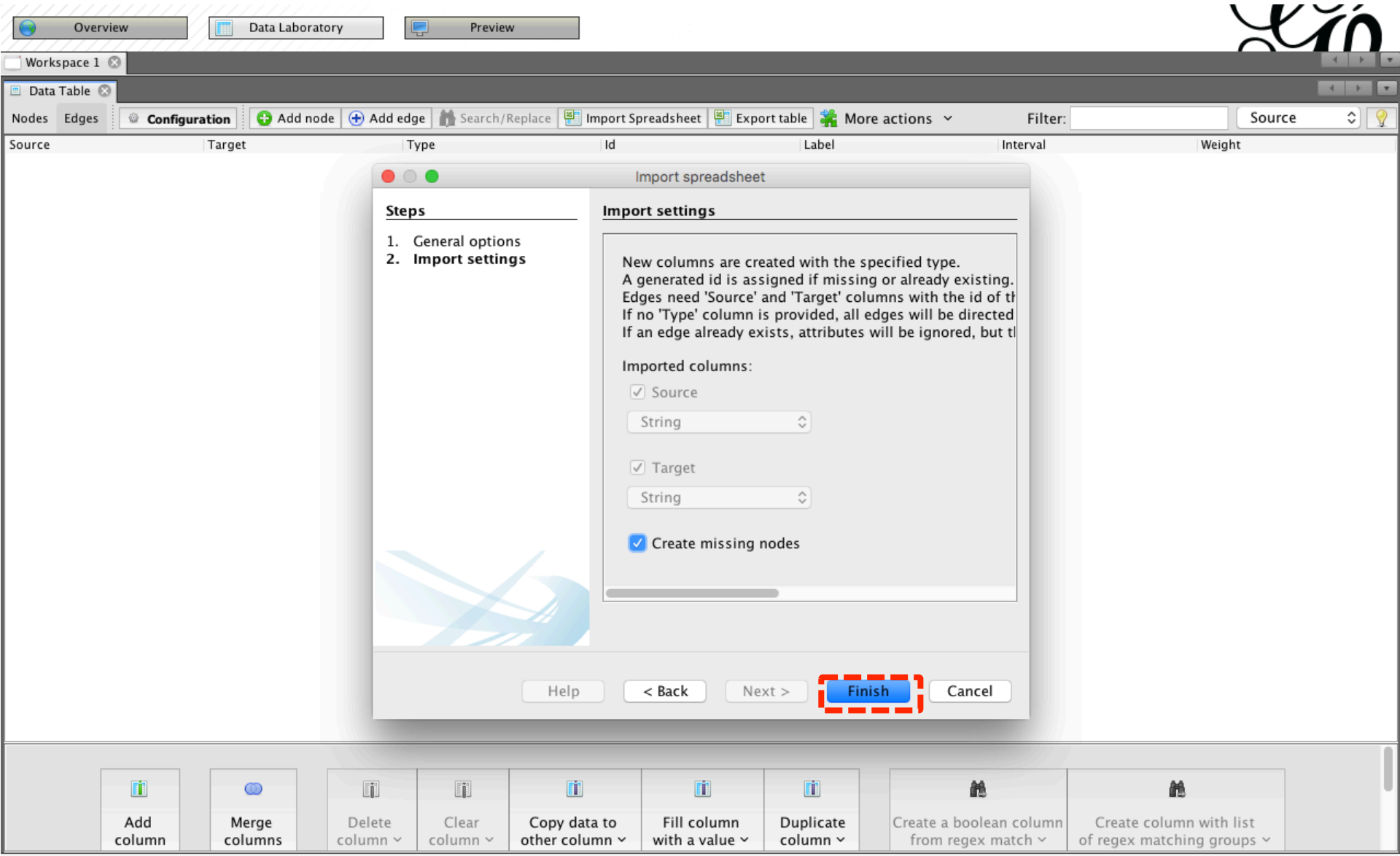


Import Edges1.csv to Gephi

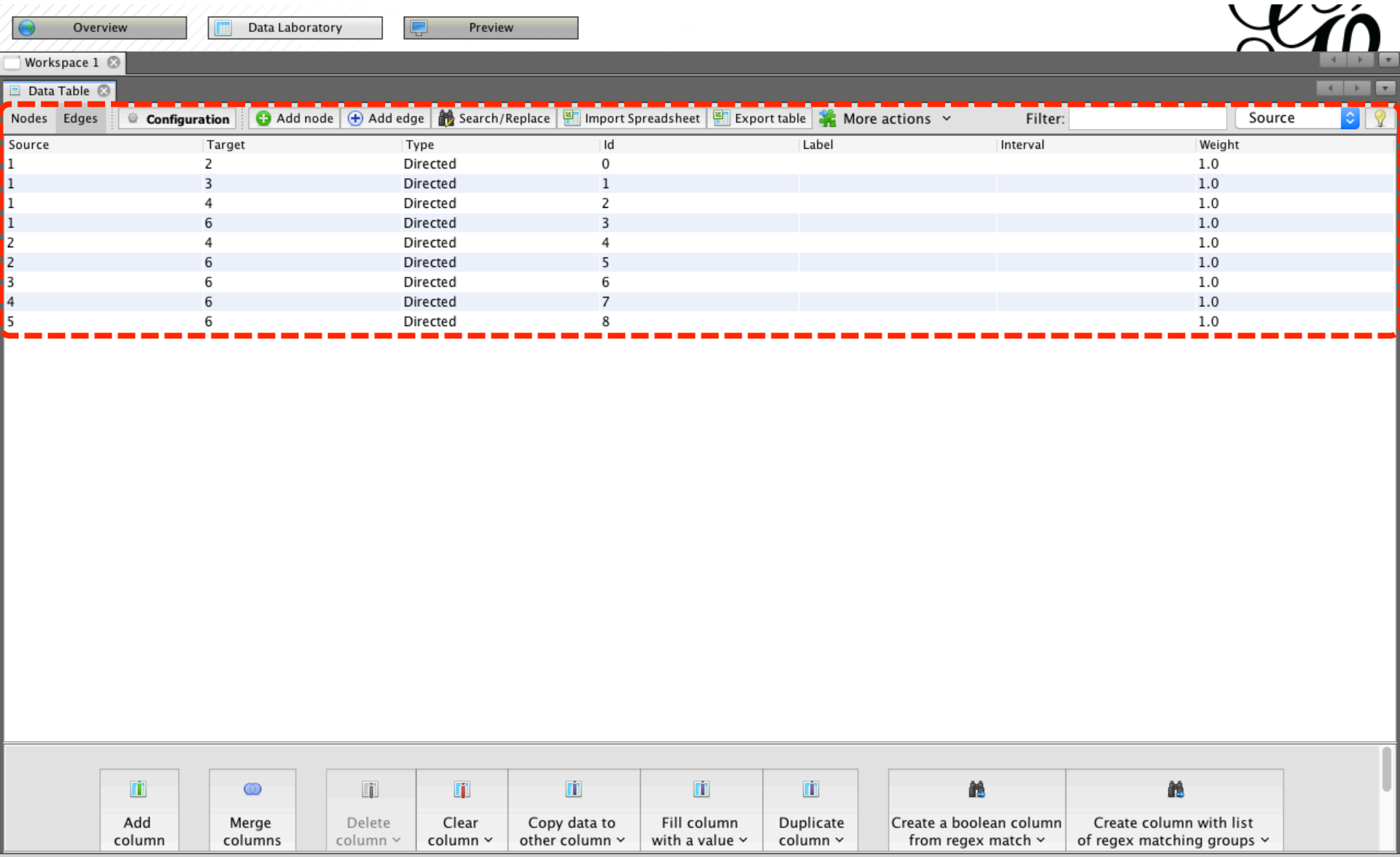


Edges table

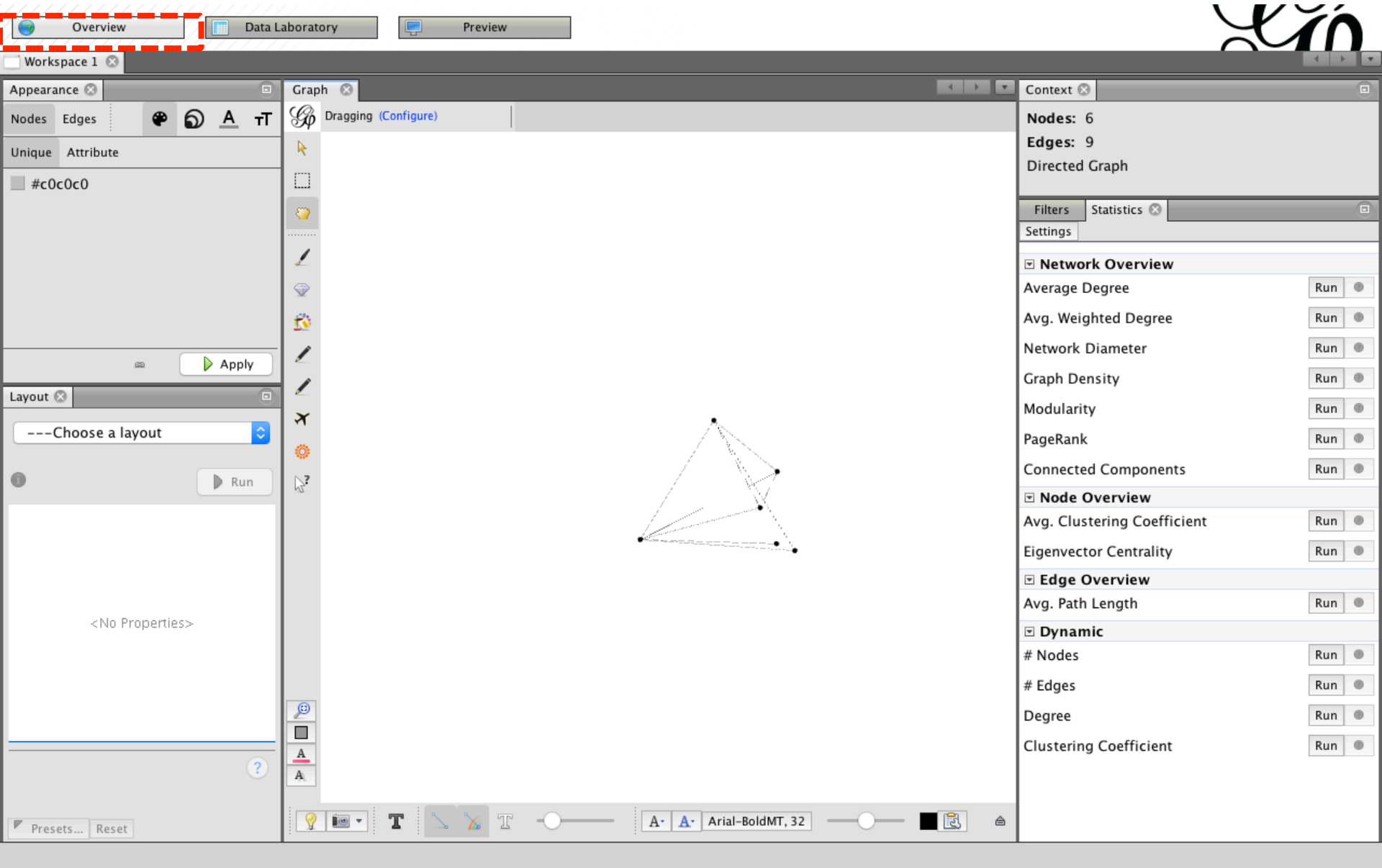
Import Edges1.csv to Gephi



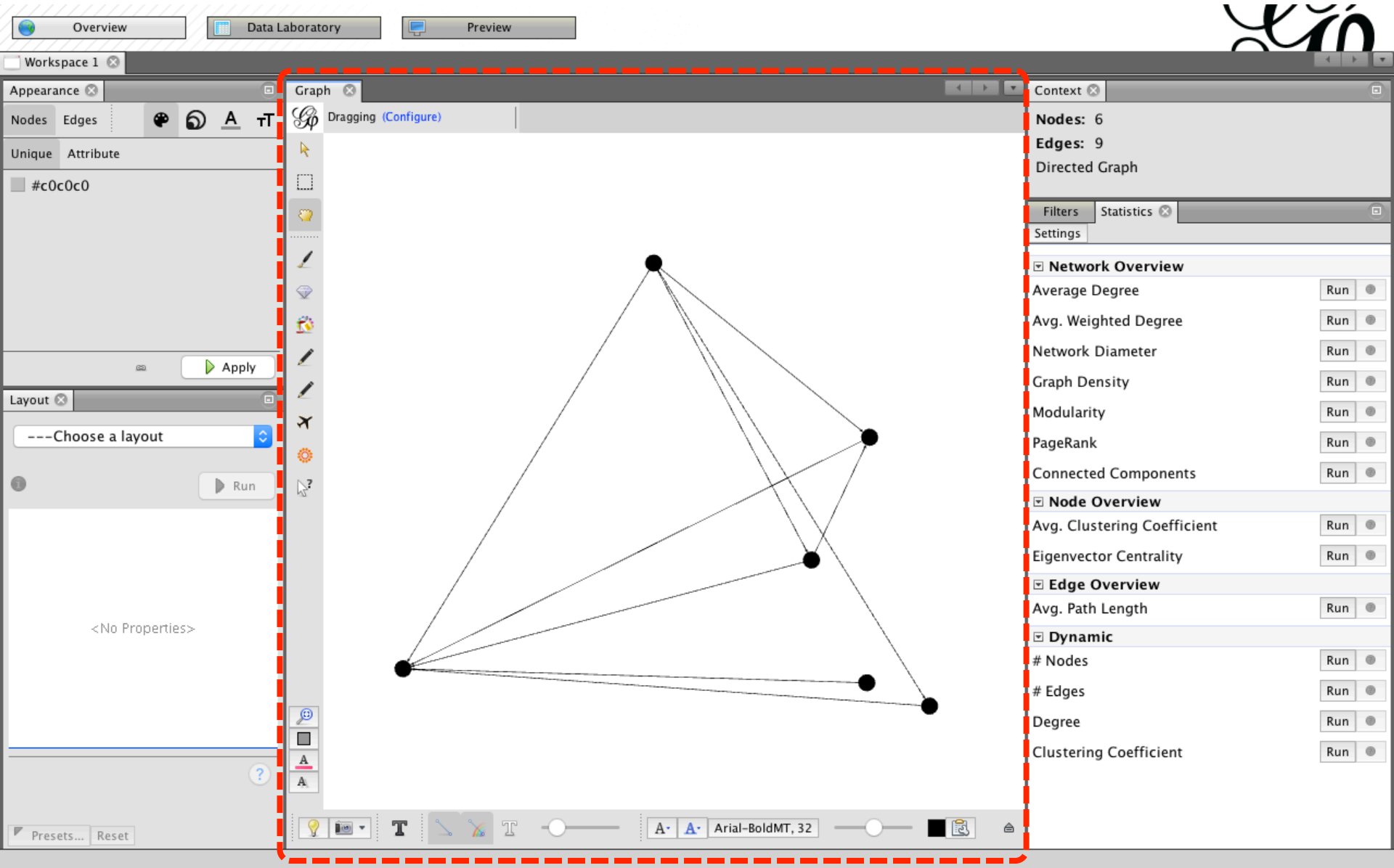
Import Edges1.csv to Gephi



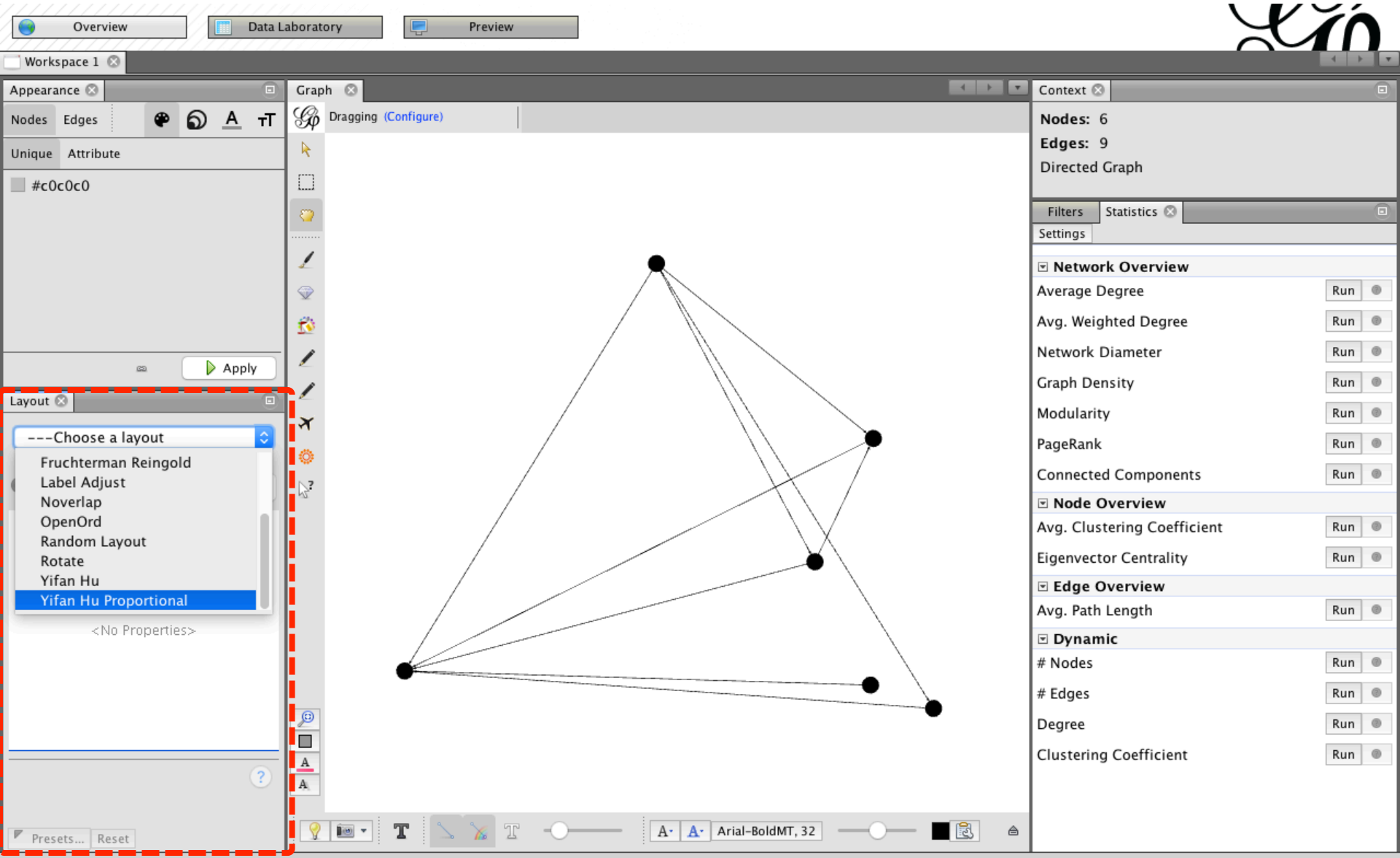
Gephi Overview



Gephi Overview: Graph

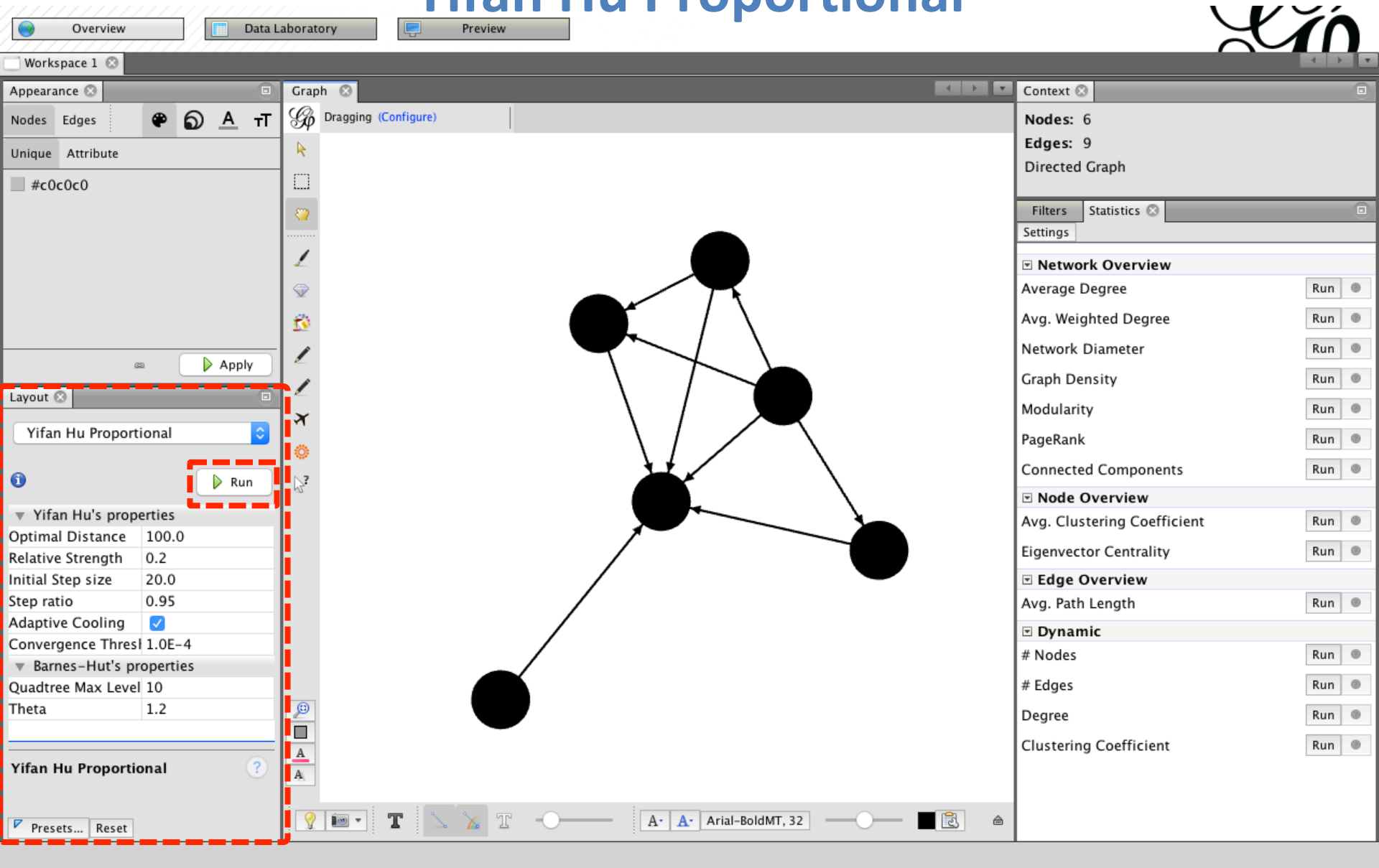


Gephi Overview: Layout



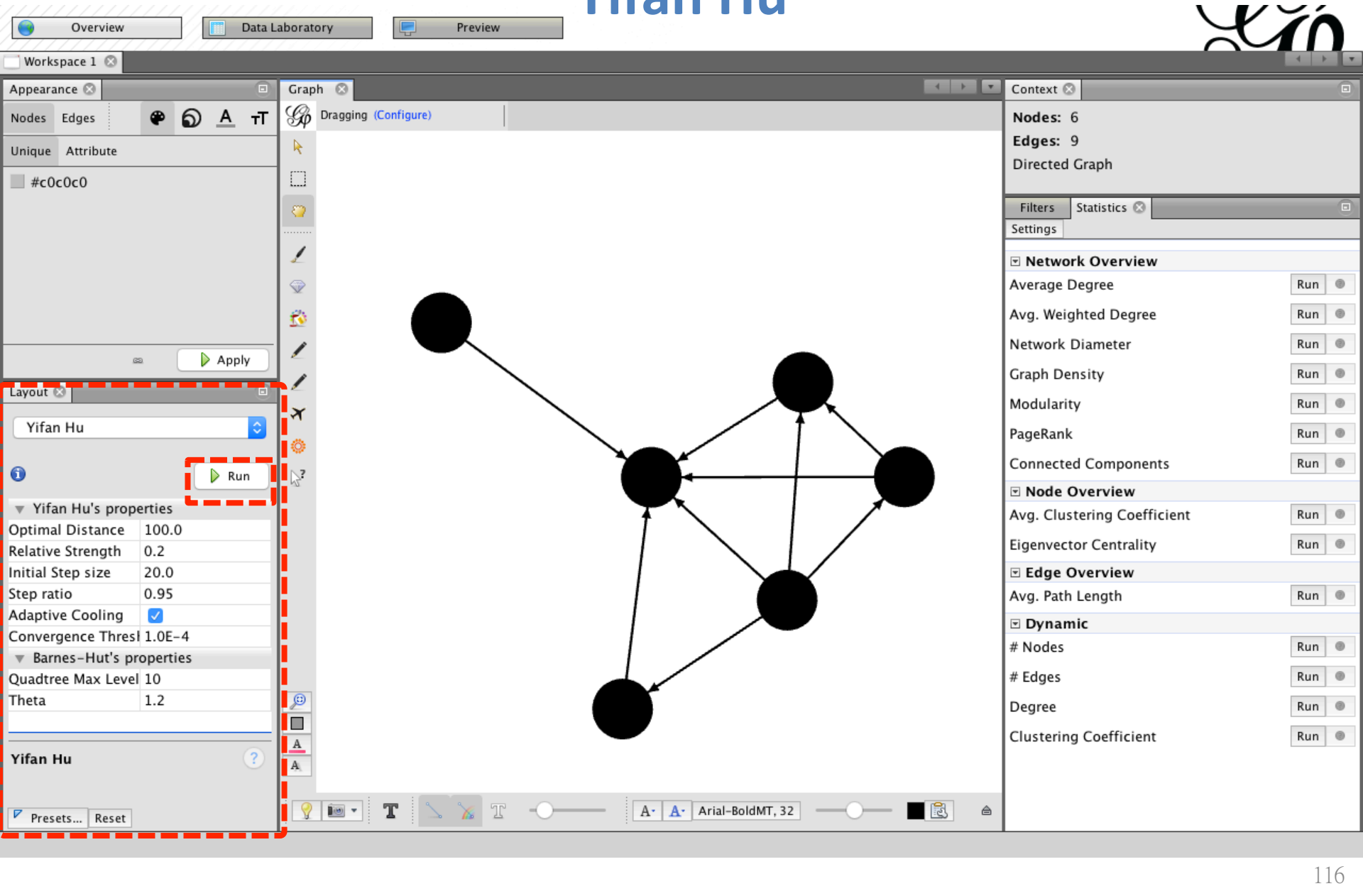
Gephi Overview: Layout

Yifan Hu Proportional

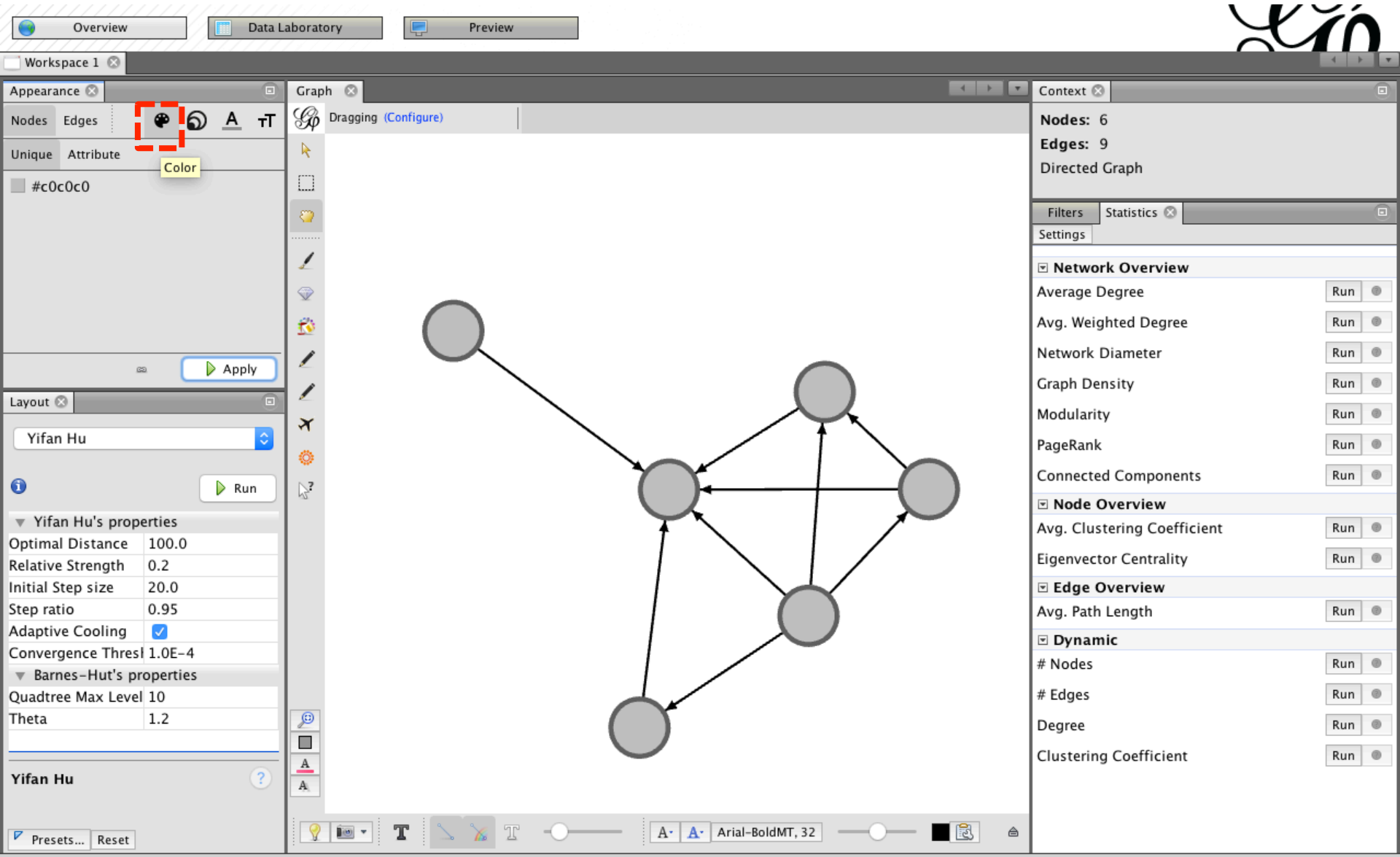


Gephi Overview: Layout

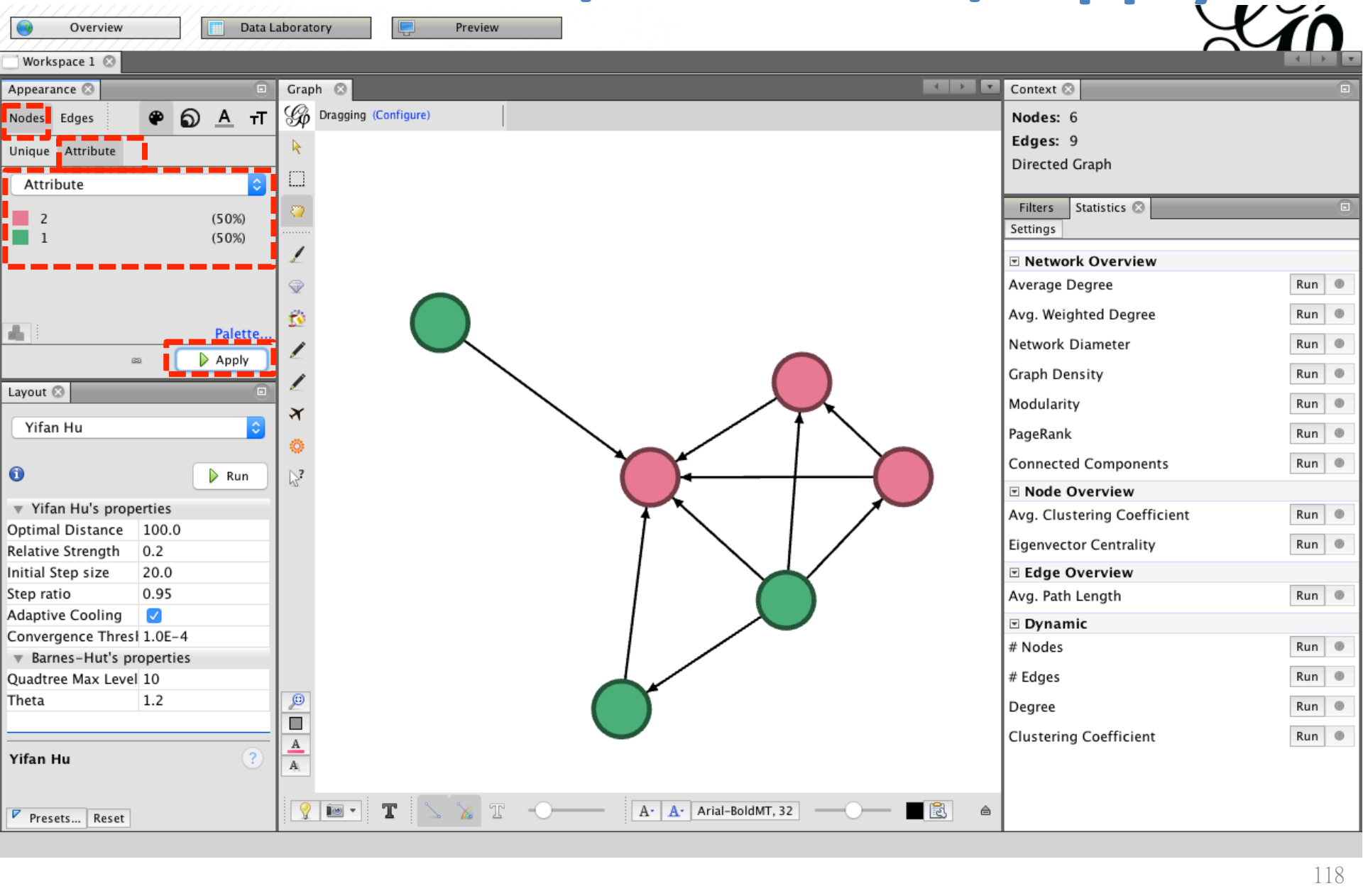
Yifan Hu



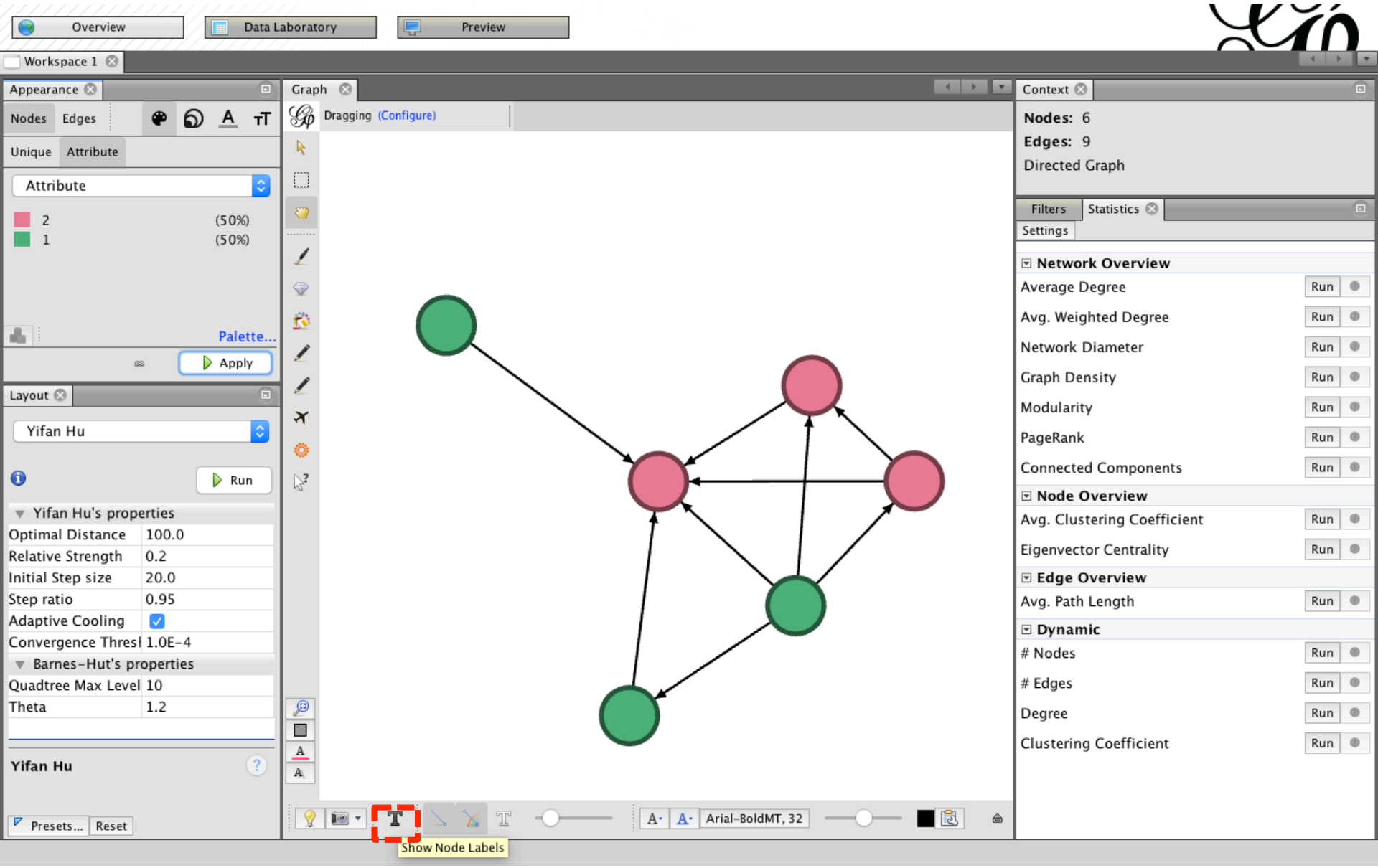
Appearance: Nodes Color



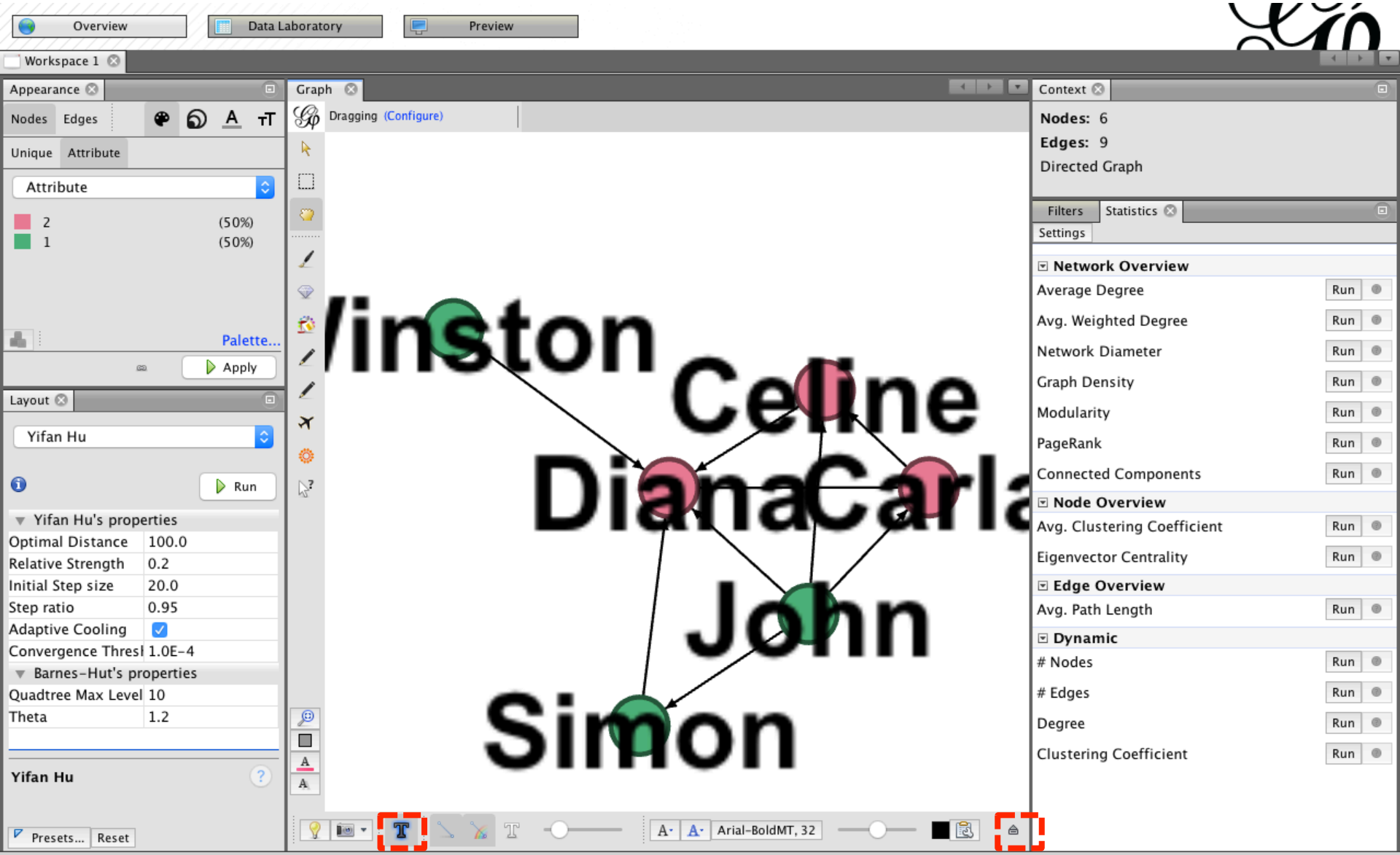
Nodes Color / Attribute / Apply



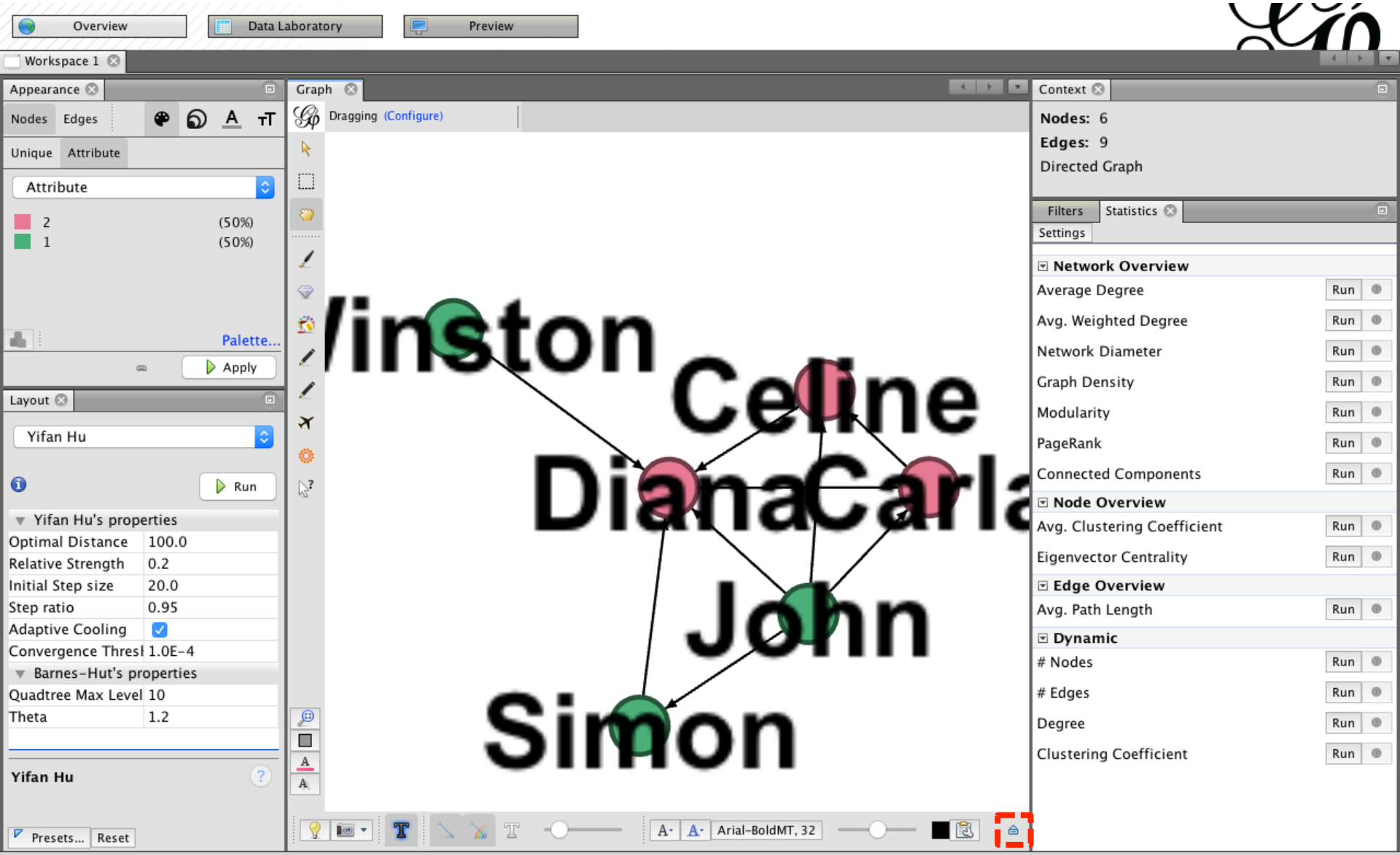
Show Node Labels



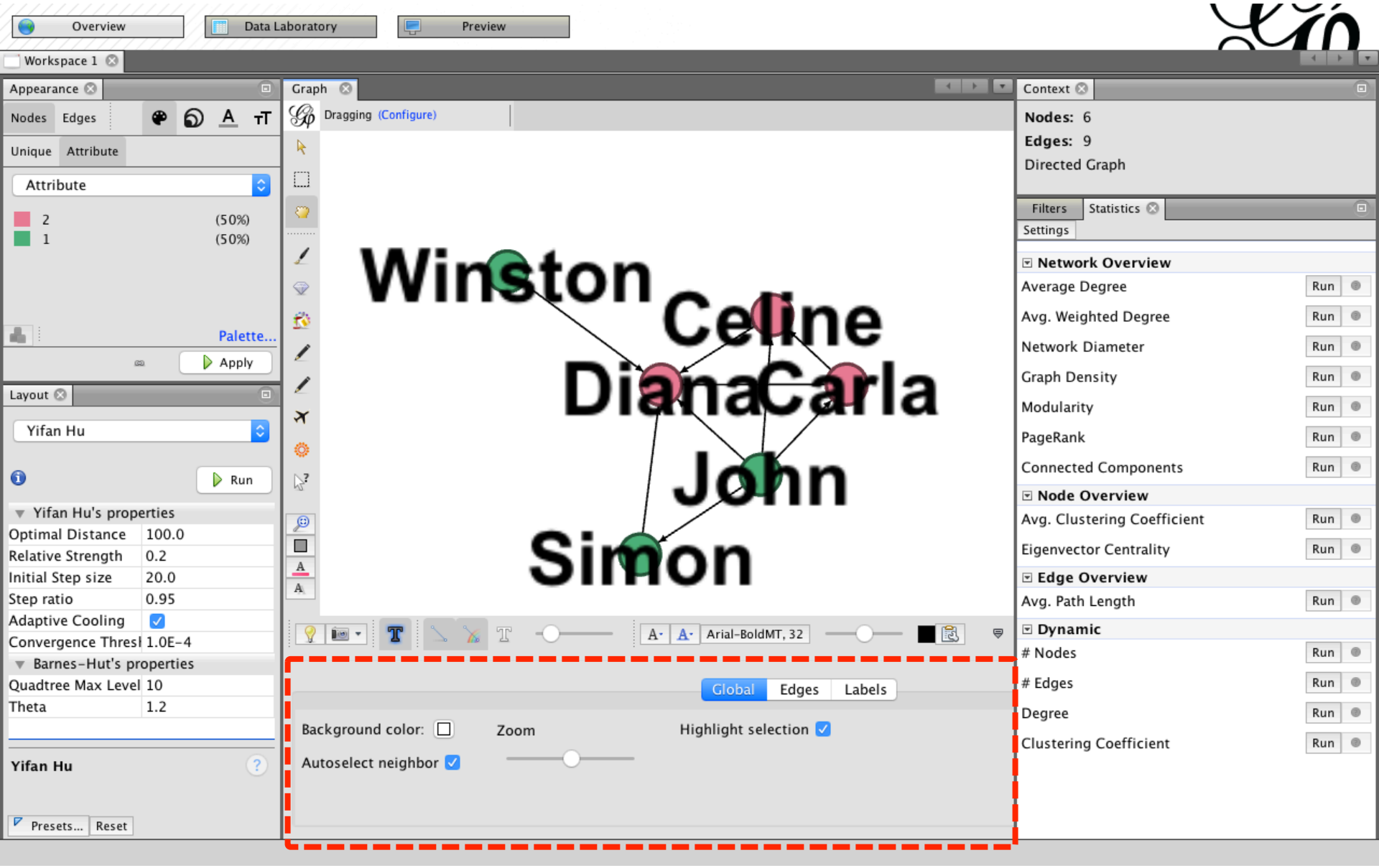
Show Node Labels



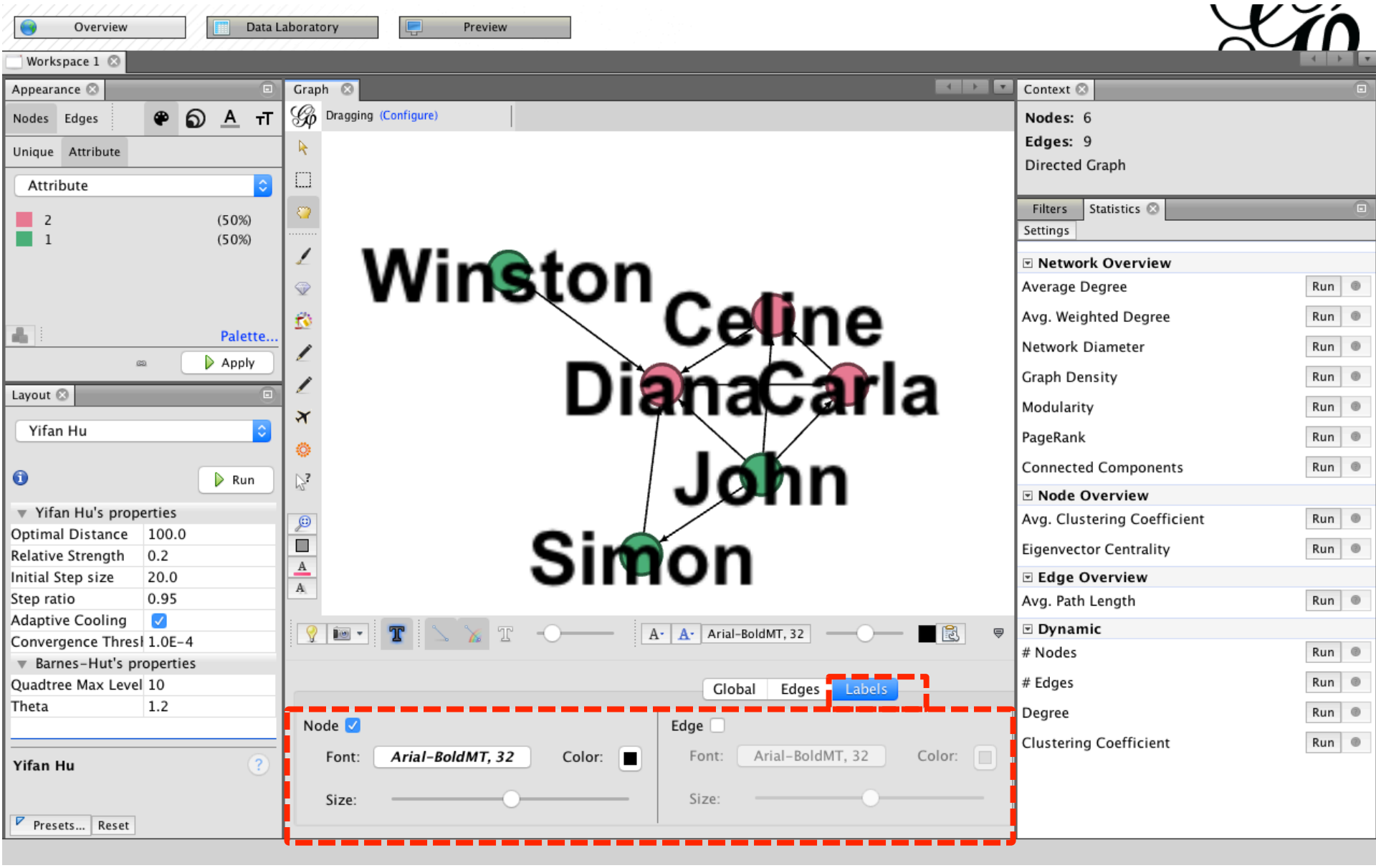
Show Labels


Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
Attribute
Attribute
2 (50%)
1 (50%)
Palette...
Apply
Layout
Yifan Hu
Run
Yifan Hu's properties
Optimal Distance 100.0
Relative Strength 0.2
Initial Step size 20.0
Step ratio 0.95
Adaptive Cooling
Convergence Thresl 1.0E-4
Barnes-Hut's properties
Quadtree Max Level 10
Theta 1.2
Yifan Hu
Presets...
Reset
Graph
Dragging (Configure)
Winston
Celine
Diana
Carla
John
Simon
Arial-BoldMT, 32
Context
Nodes: 6
Edges: 9
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density 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

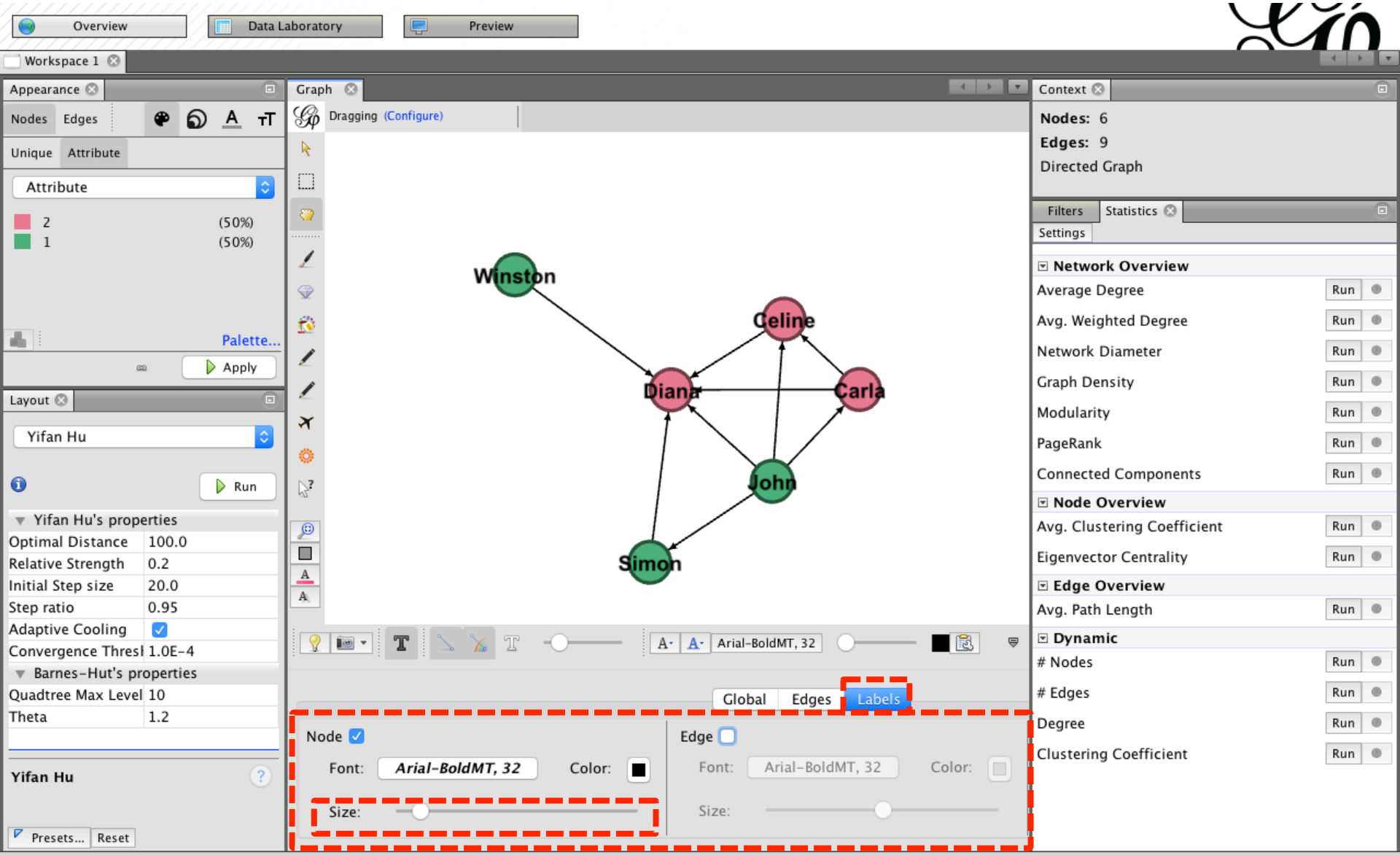
Global Edges Labels



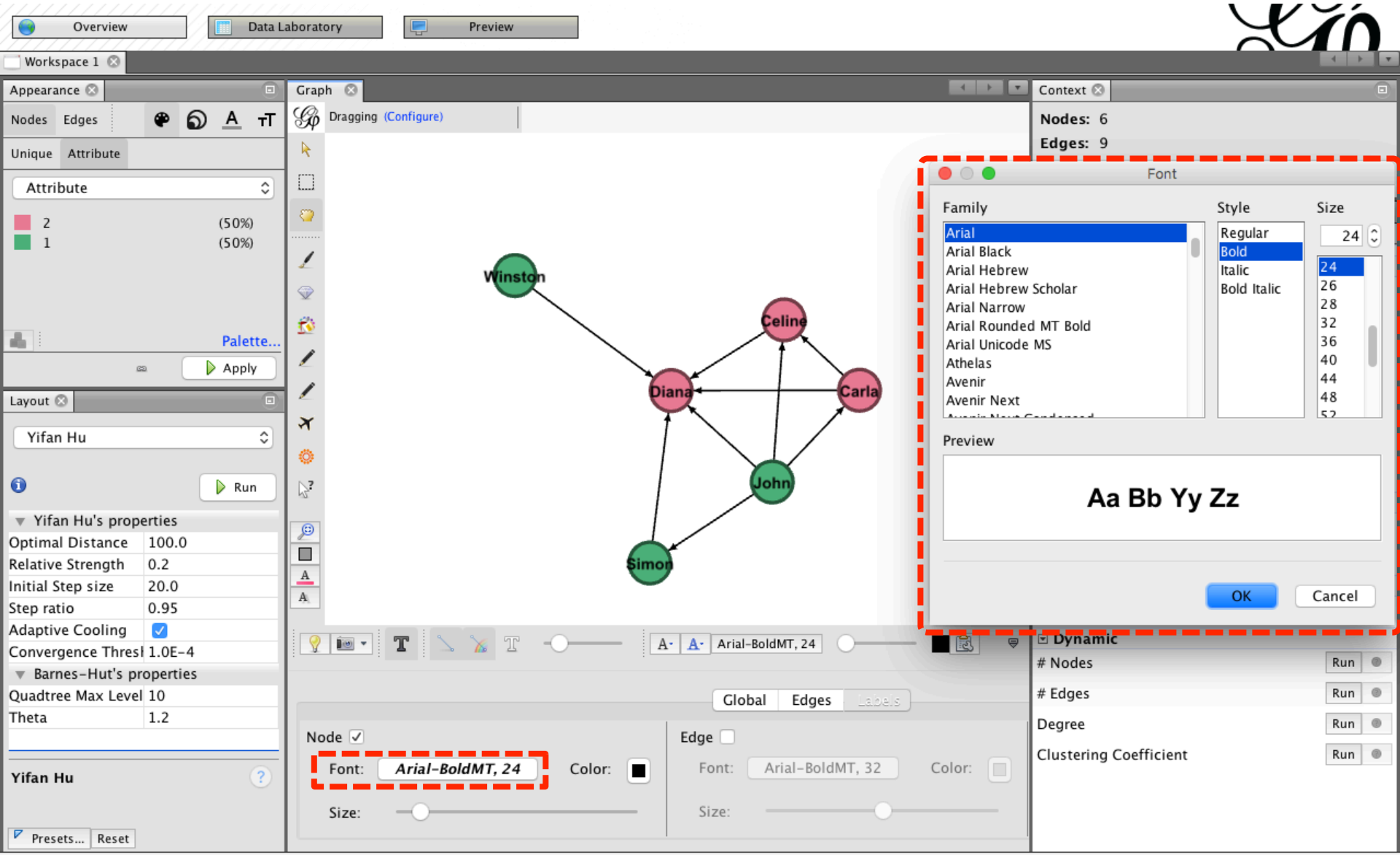
Labels


Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
Attribute
Attribute
2 (50%)
1 (50%)
Palette...
Apply
Layout
Yifan Hu
Run
Yifan Hu's properties
Optimal Distance 100.0
Relative Strength 0.2
Initial Step size 20.0
Step ratio 0.95
Adaptive Cooling
Convergence Thresl 1.0E-4
Barnes-Hut's properties
Quadtree Max Level 10
Theta 1.2
Yifan Hu
Presets...
Reset
Graph
Dragging (Configure)
Winston
Celine
Diana
Carla
John
Simon
Arial-BoldMT, 32
Global
Edges
Labels
Node
Font: Arial-BoldMT, 32
Color:
Size:
Edge
Font: Arial-BoldMT, 32
Color:
Size:
Context
Nodes: 6
Edges: 9
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density 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

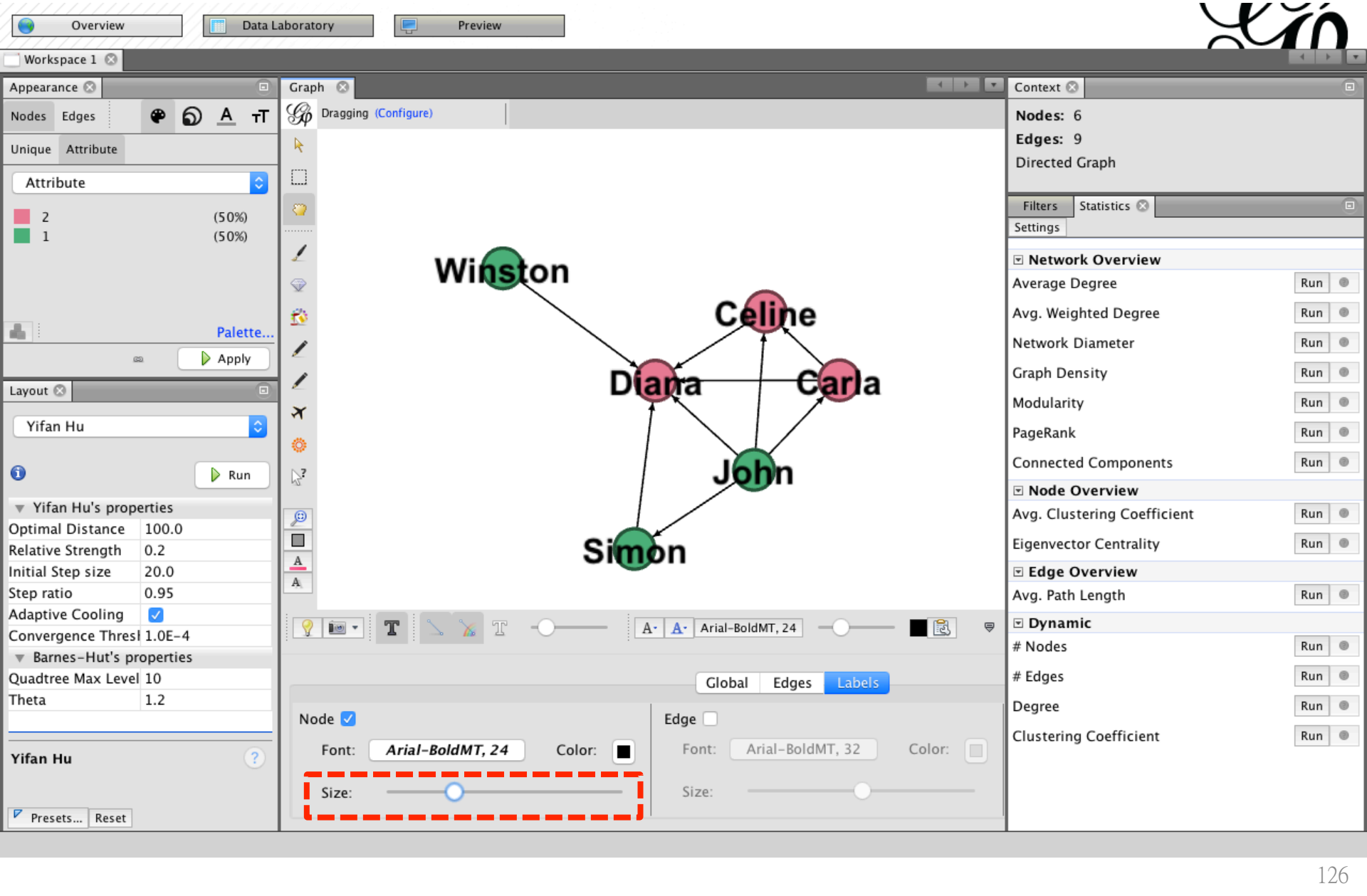
Labels Node Size



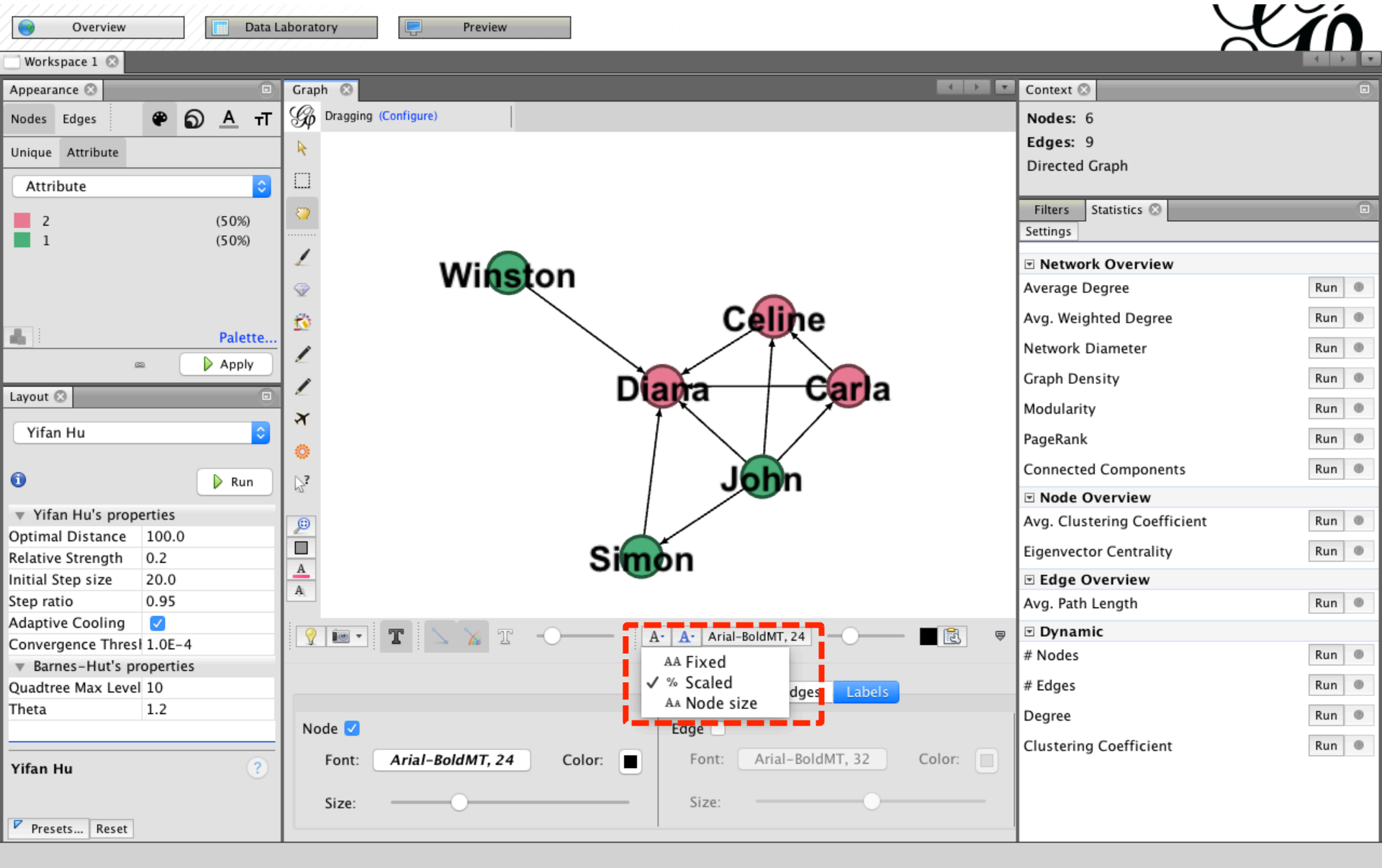
Labels Node Font Size



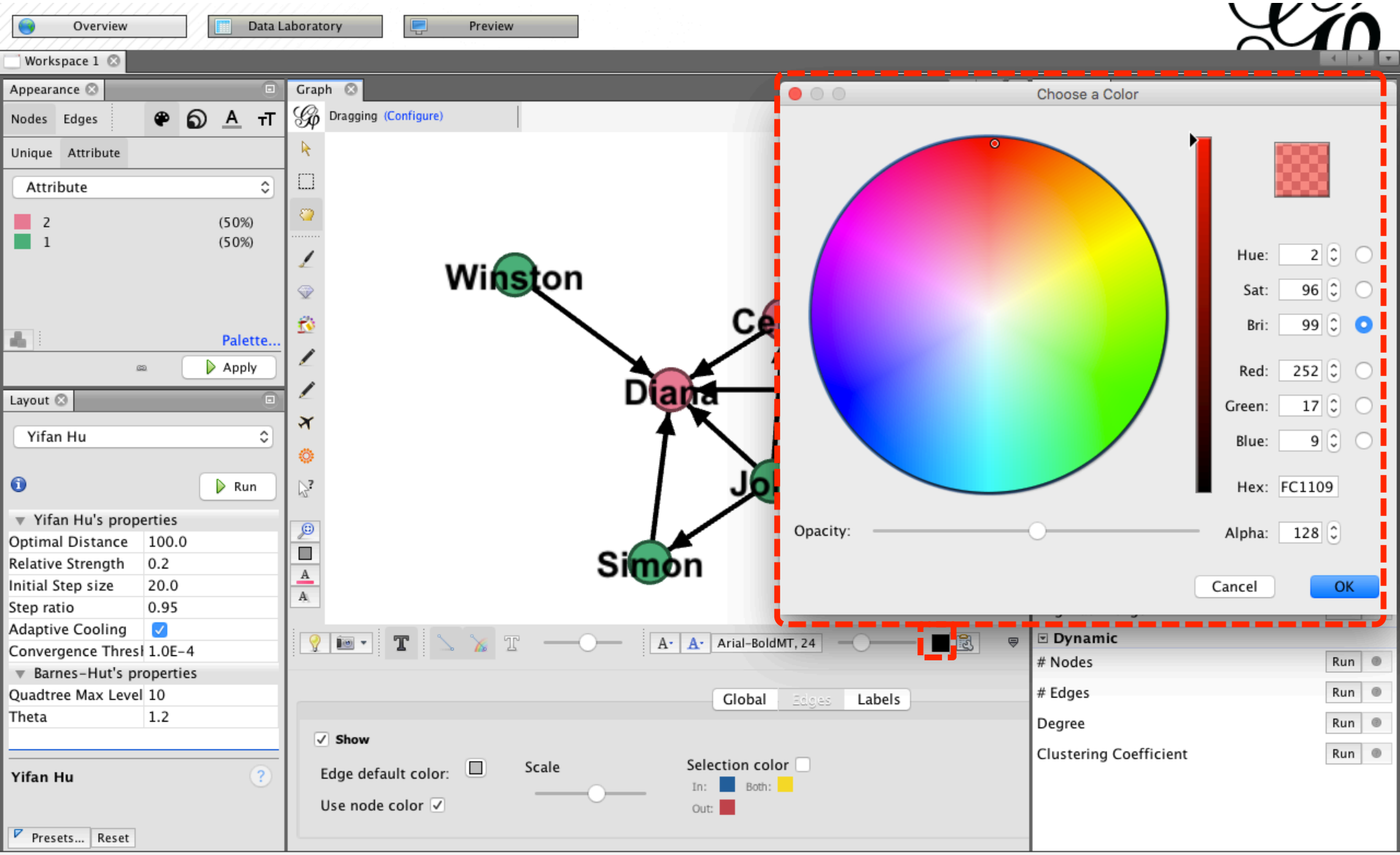
Labels Node Size



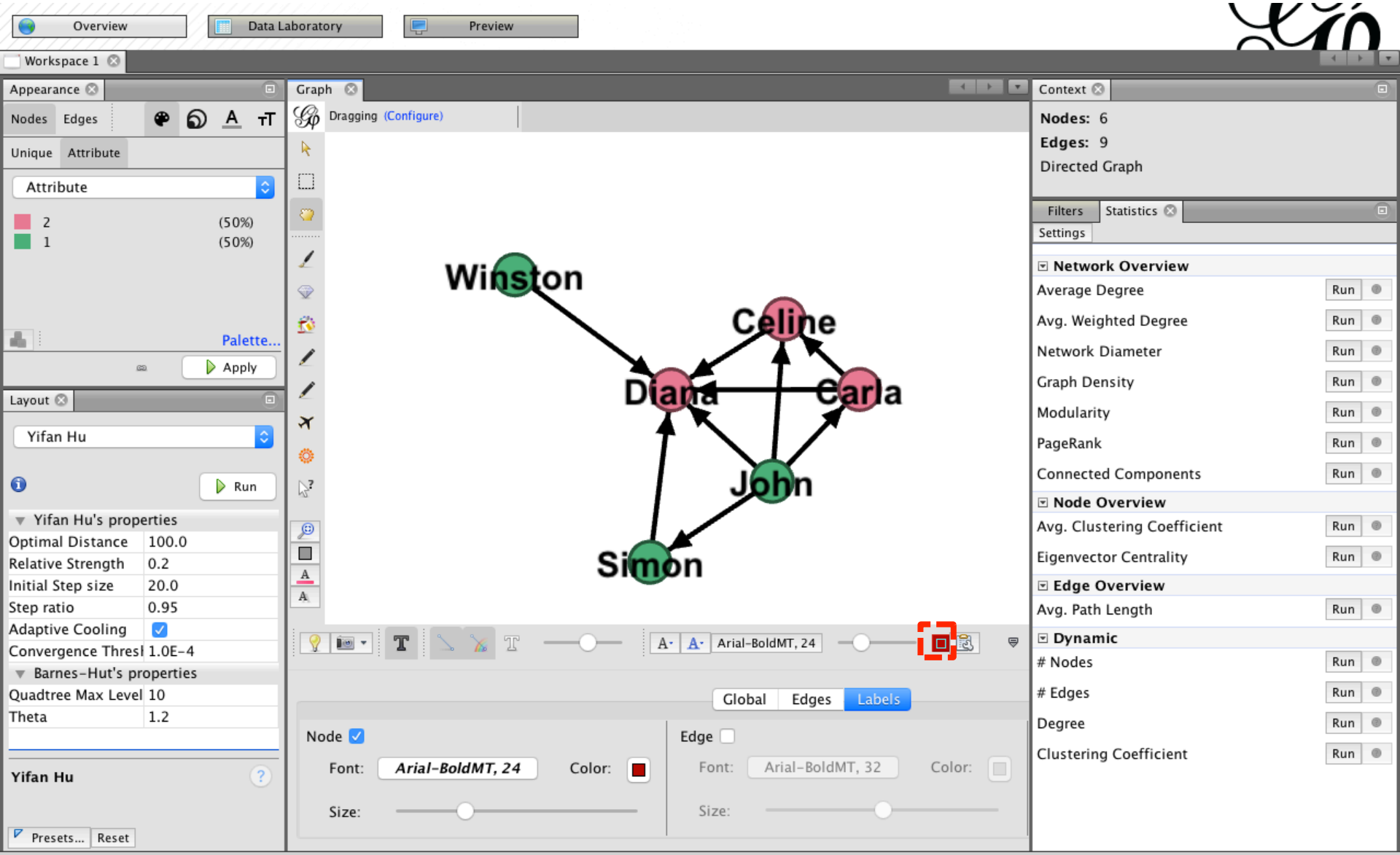
Labels Scaled



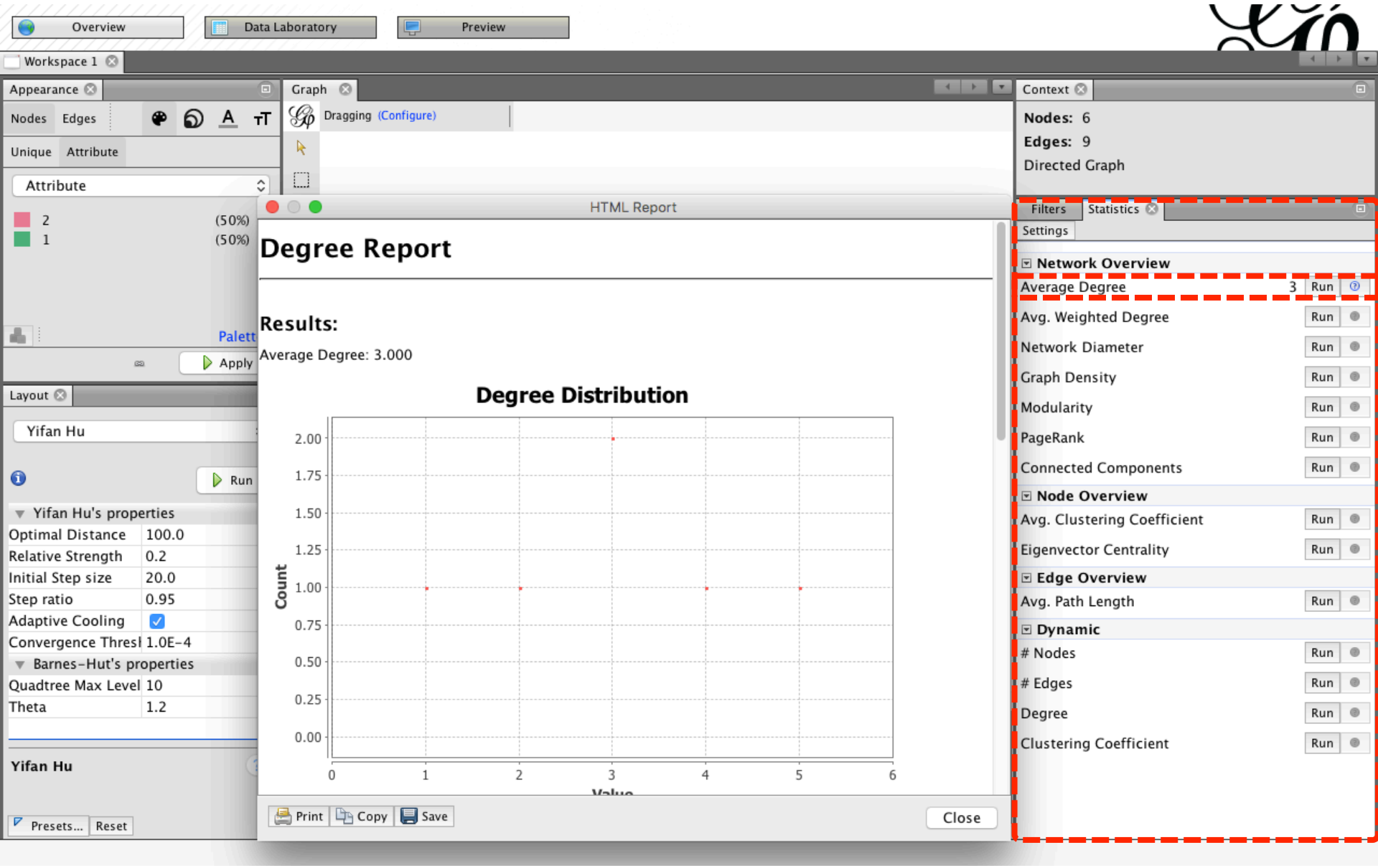
Labels Color



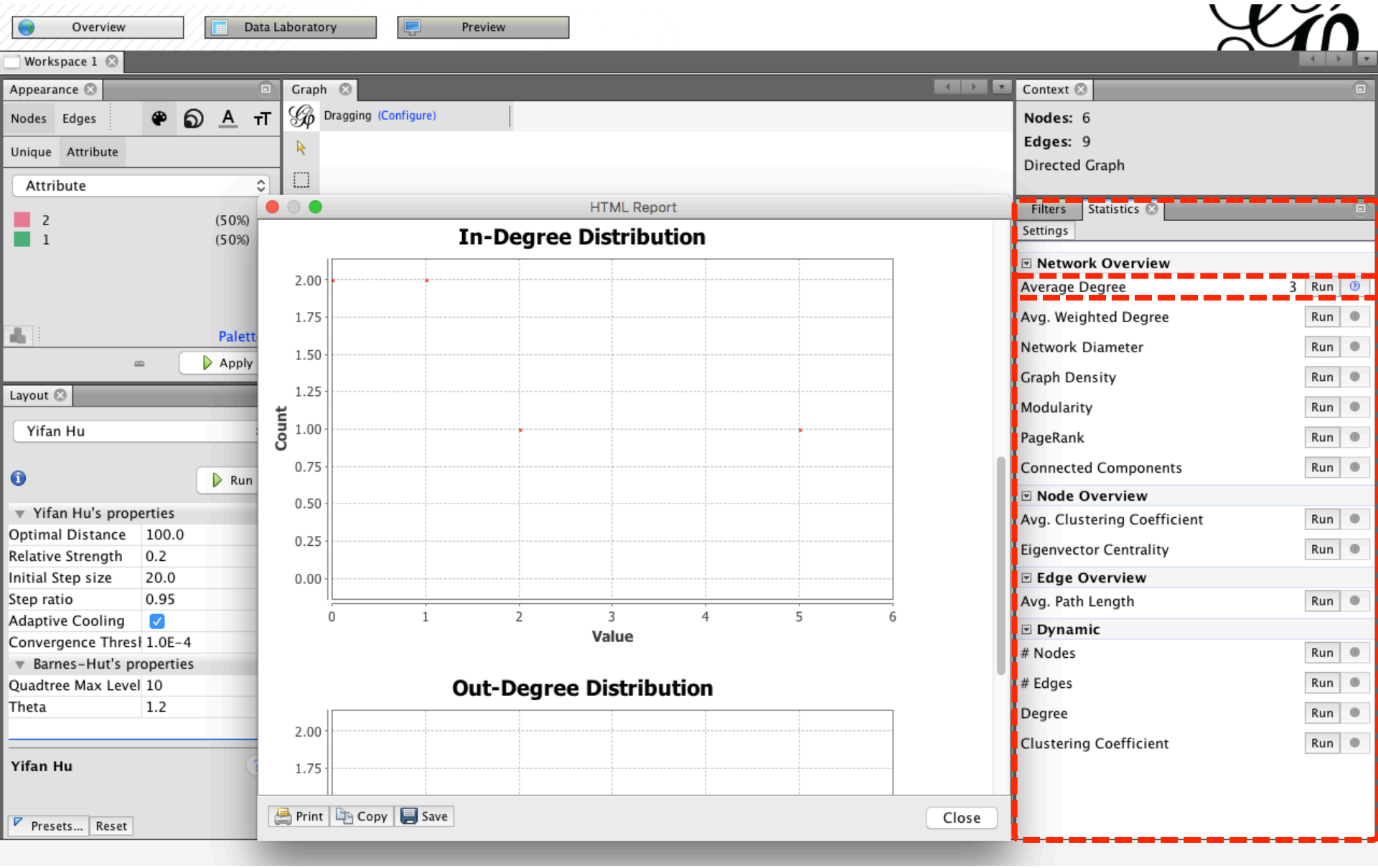
Labels Color



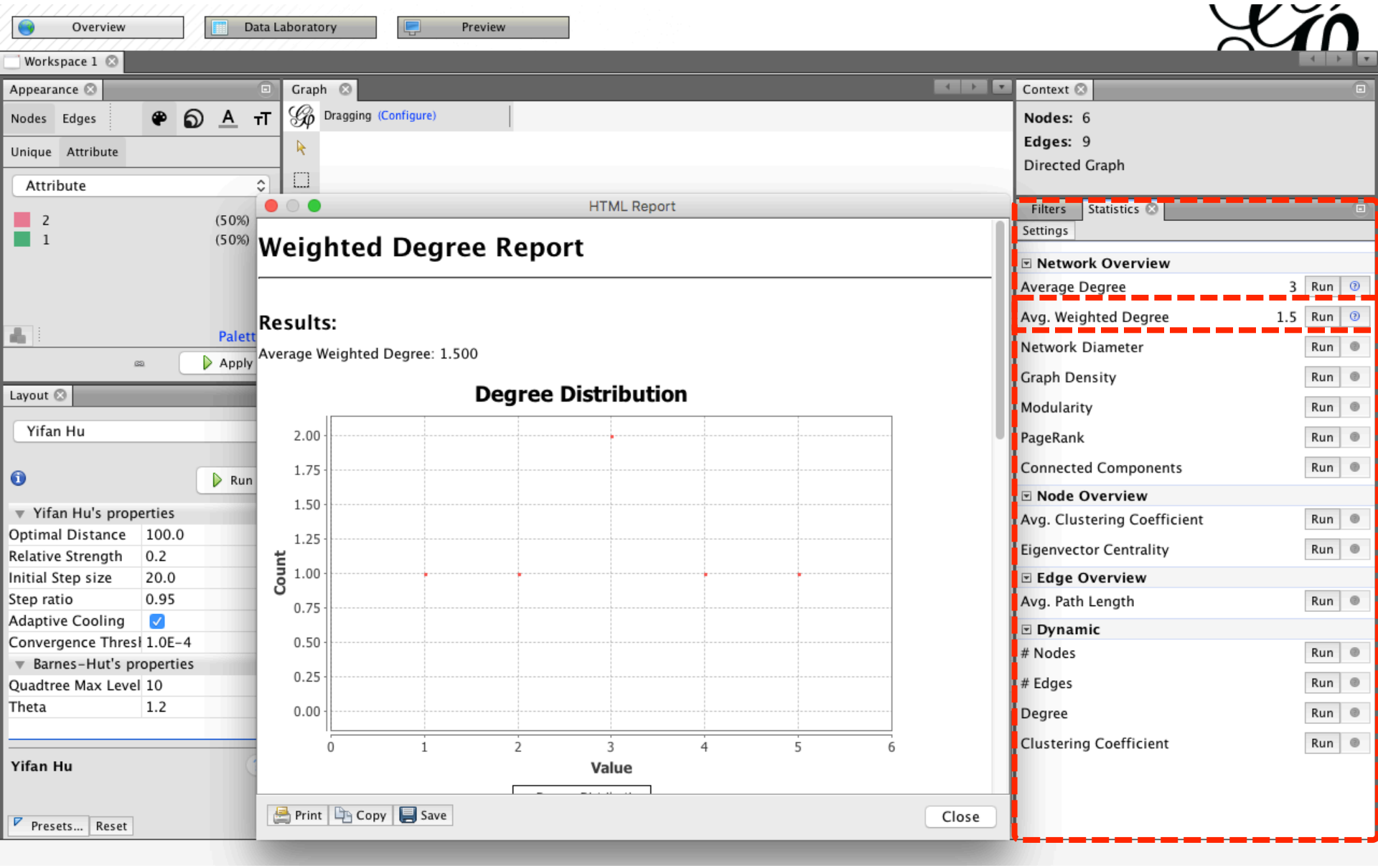
Gephi Statistics: Average Degree



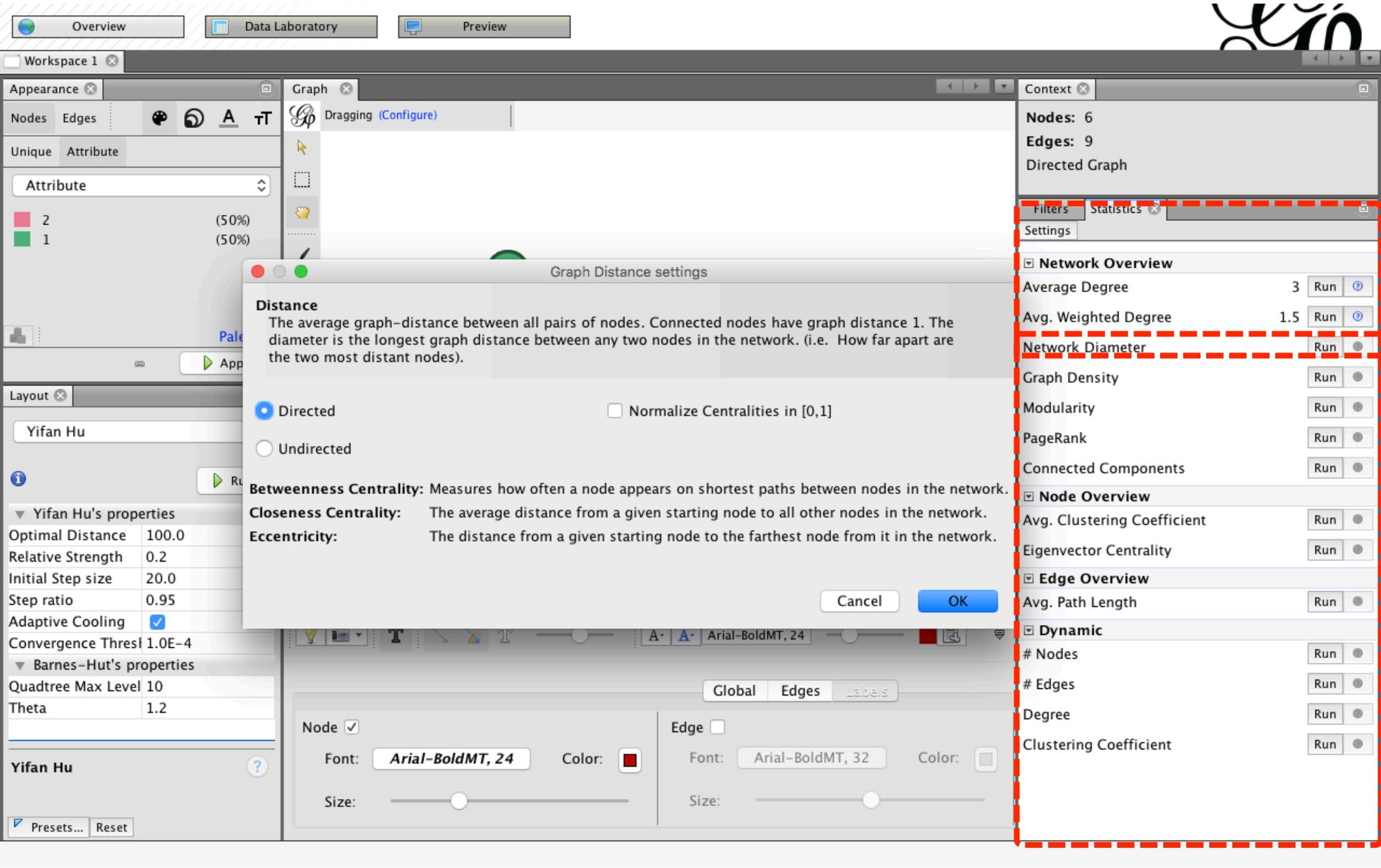
Gephi Statistics: Average Degree



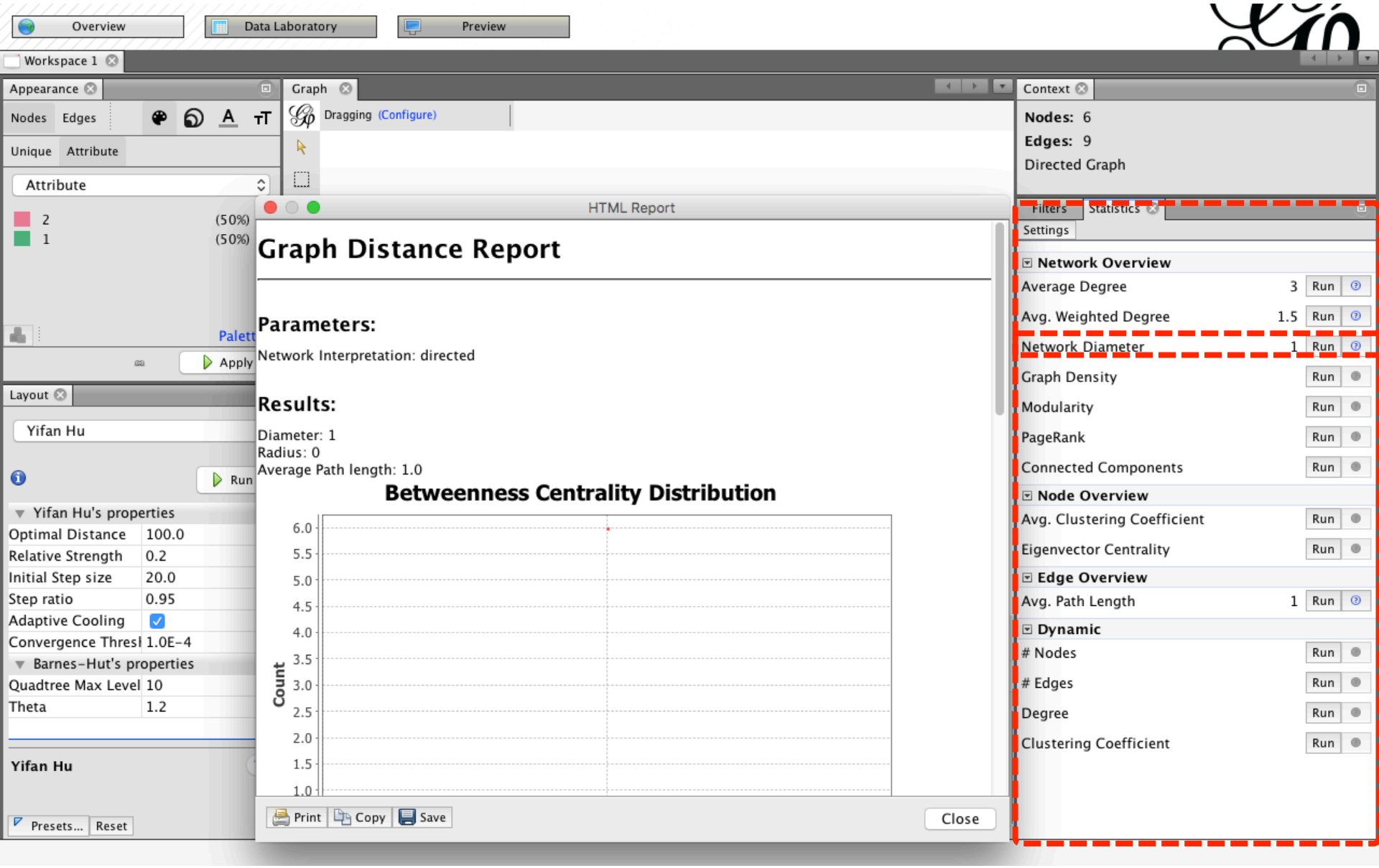
Gephi Statistics: Avg. Weighted Degree



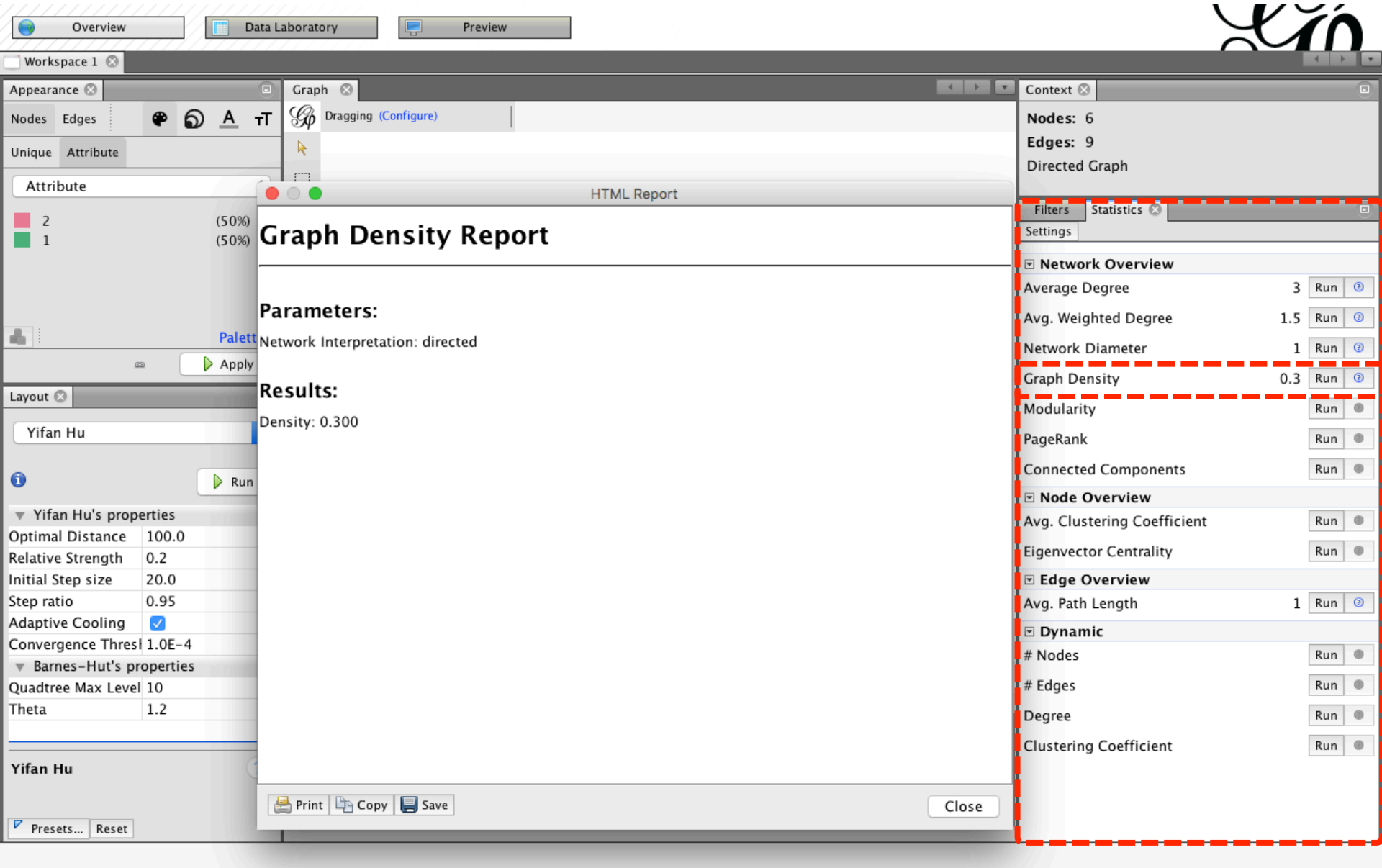
Gephi Statistics: Network Diameter



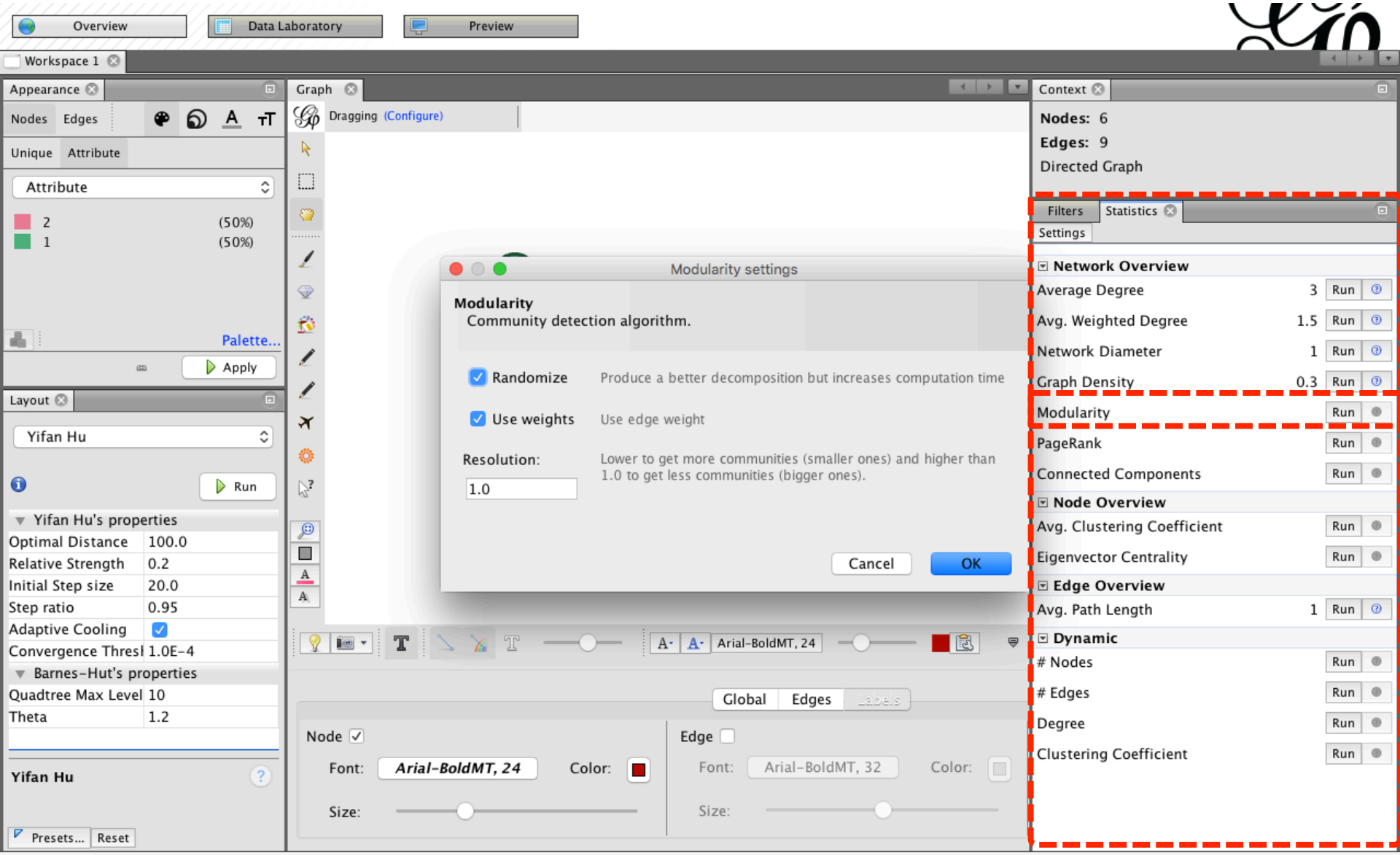
Gephi Statistics: Network Diameter



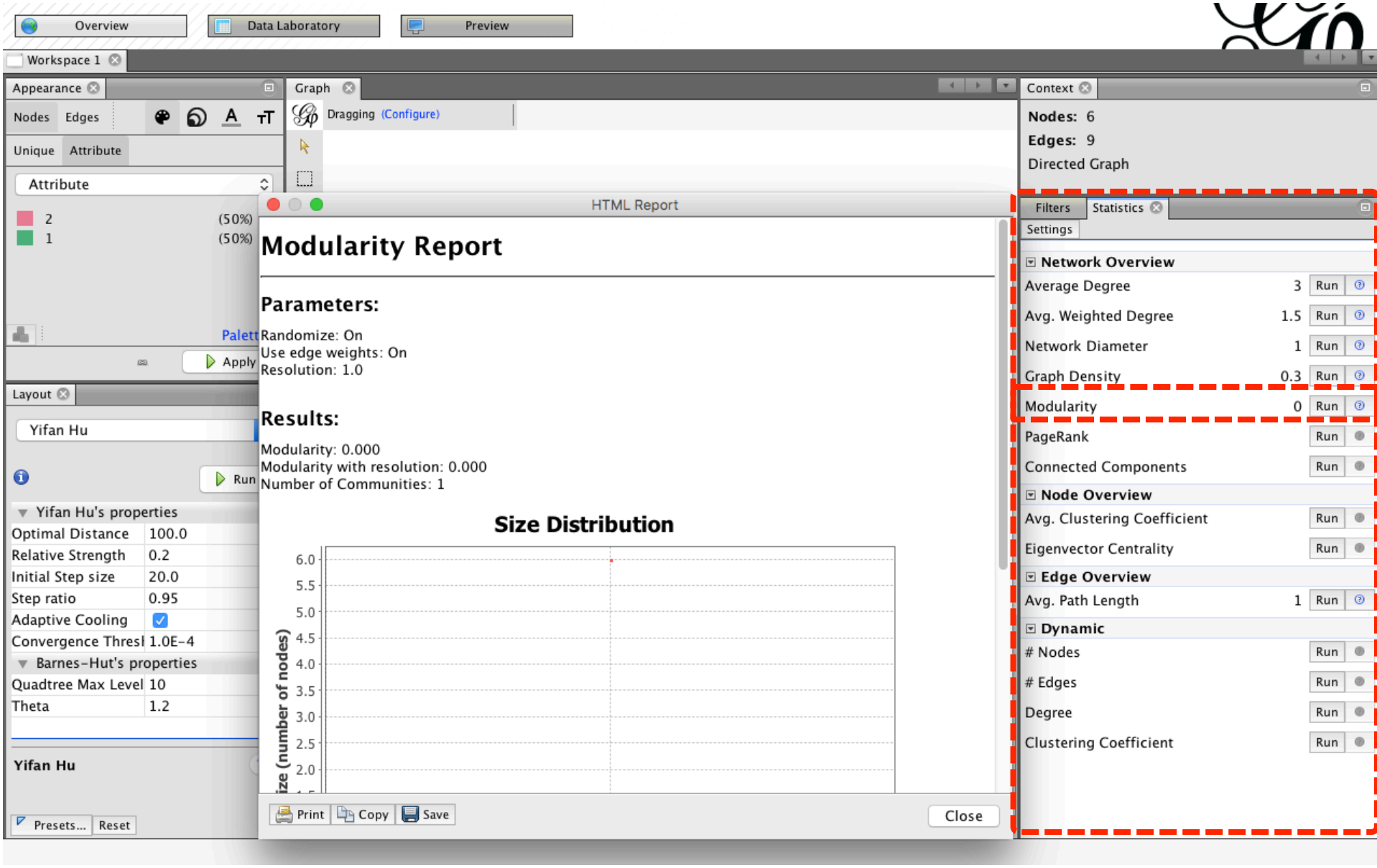
Gephi Statistics: Graph Density



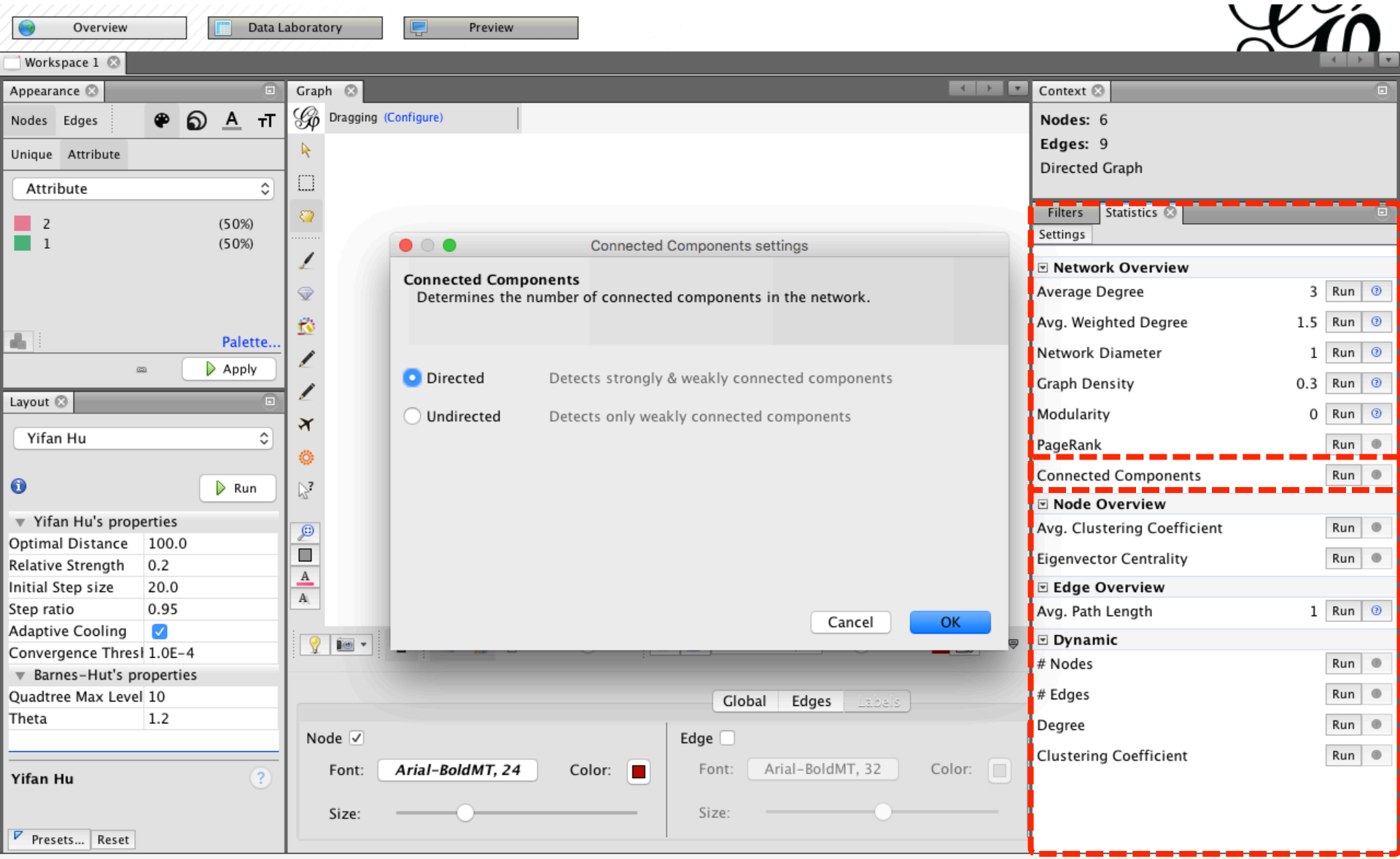
Gephi Statistics: Modularity



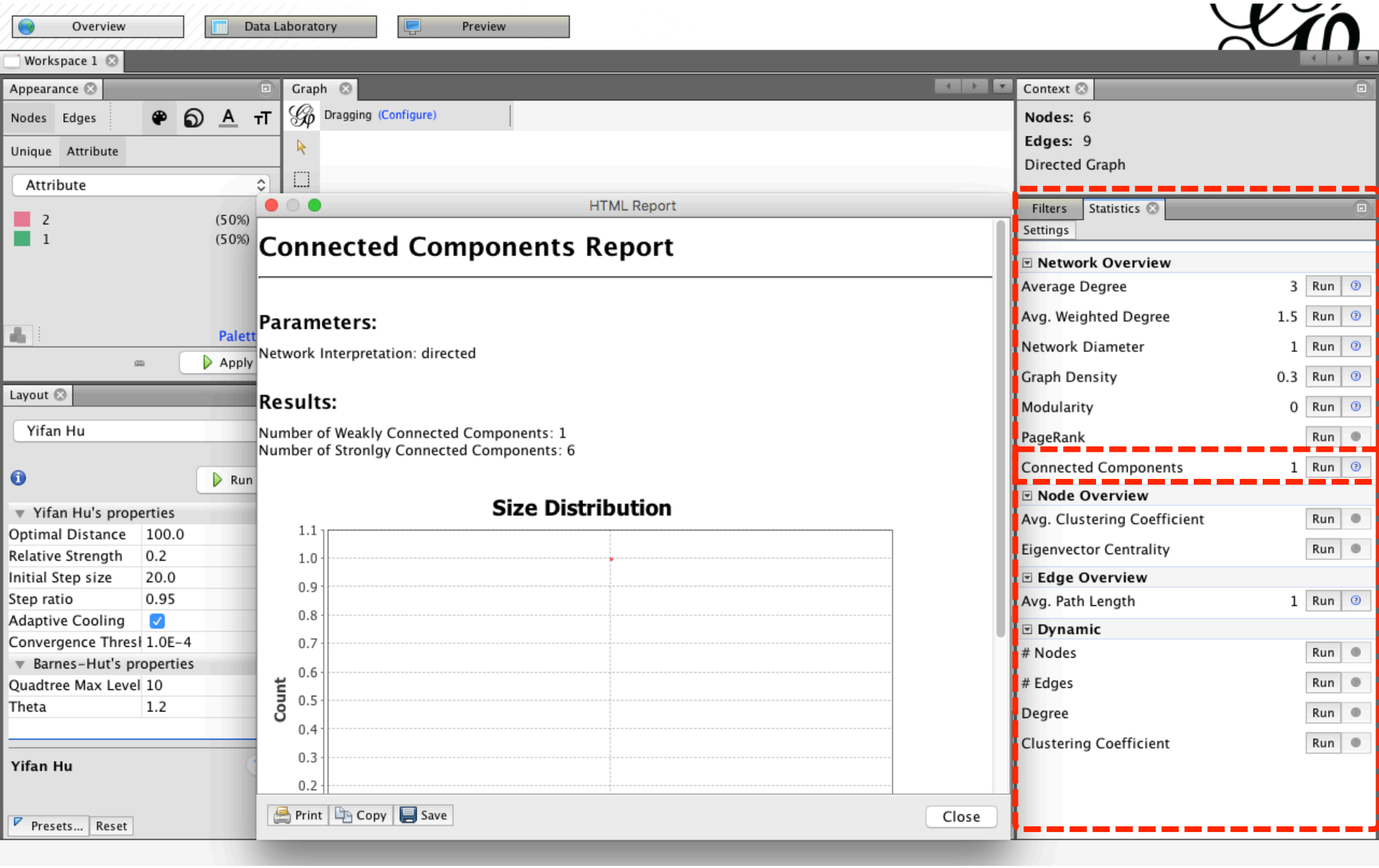
Gephi Statistics: Modularity



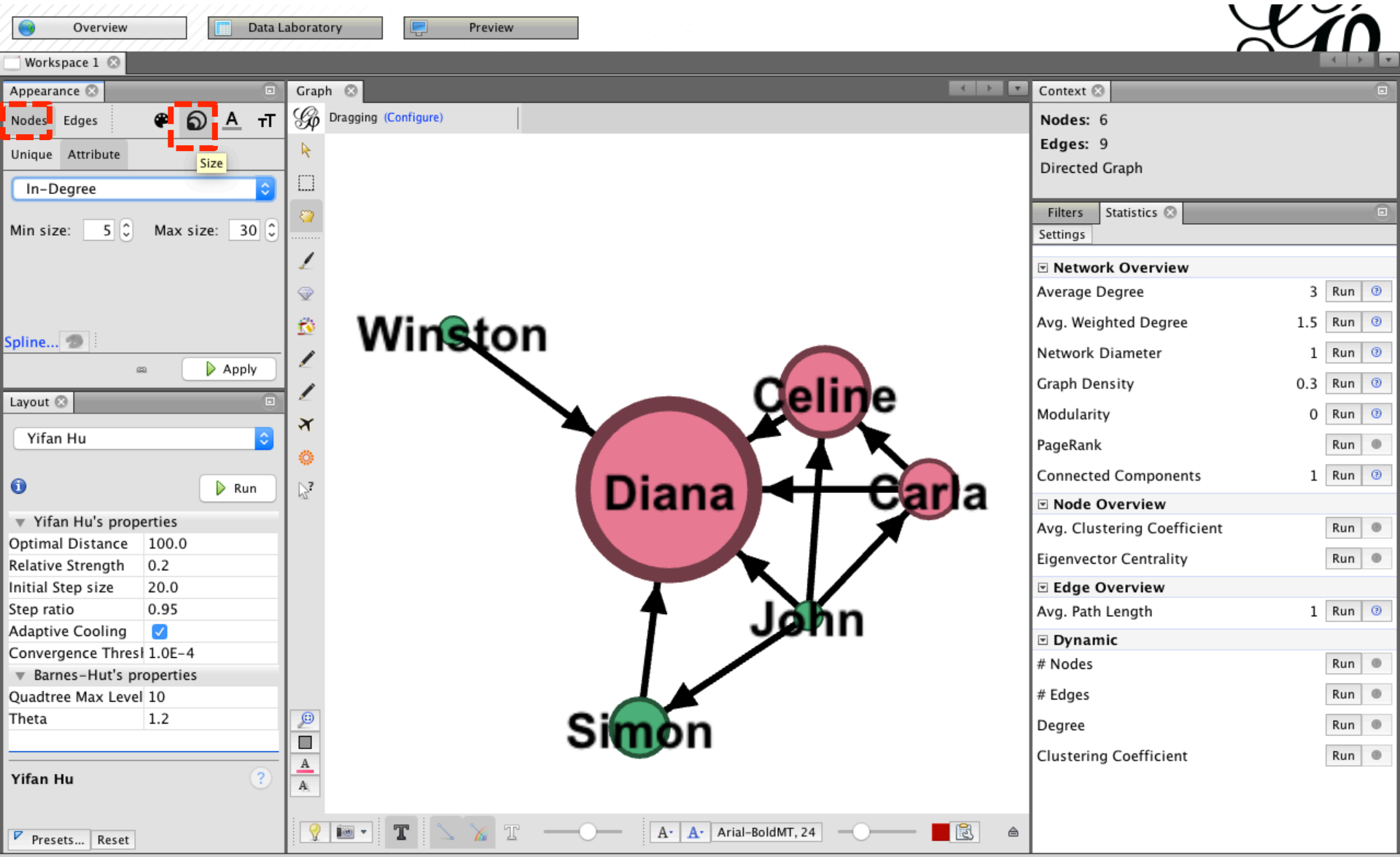
Gephi Statistics: Connected Components



Gephi Statistics: Connected Components

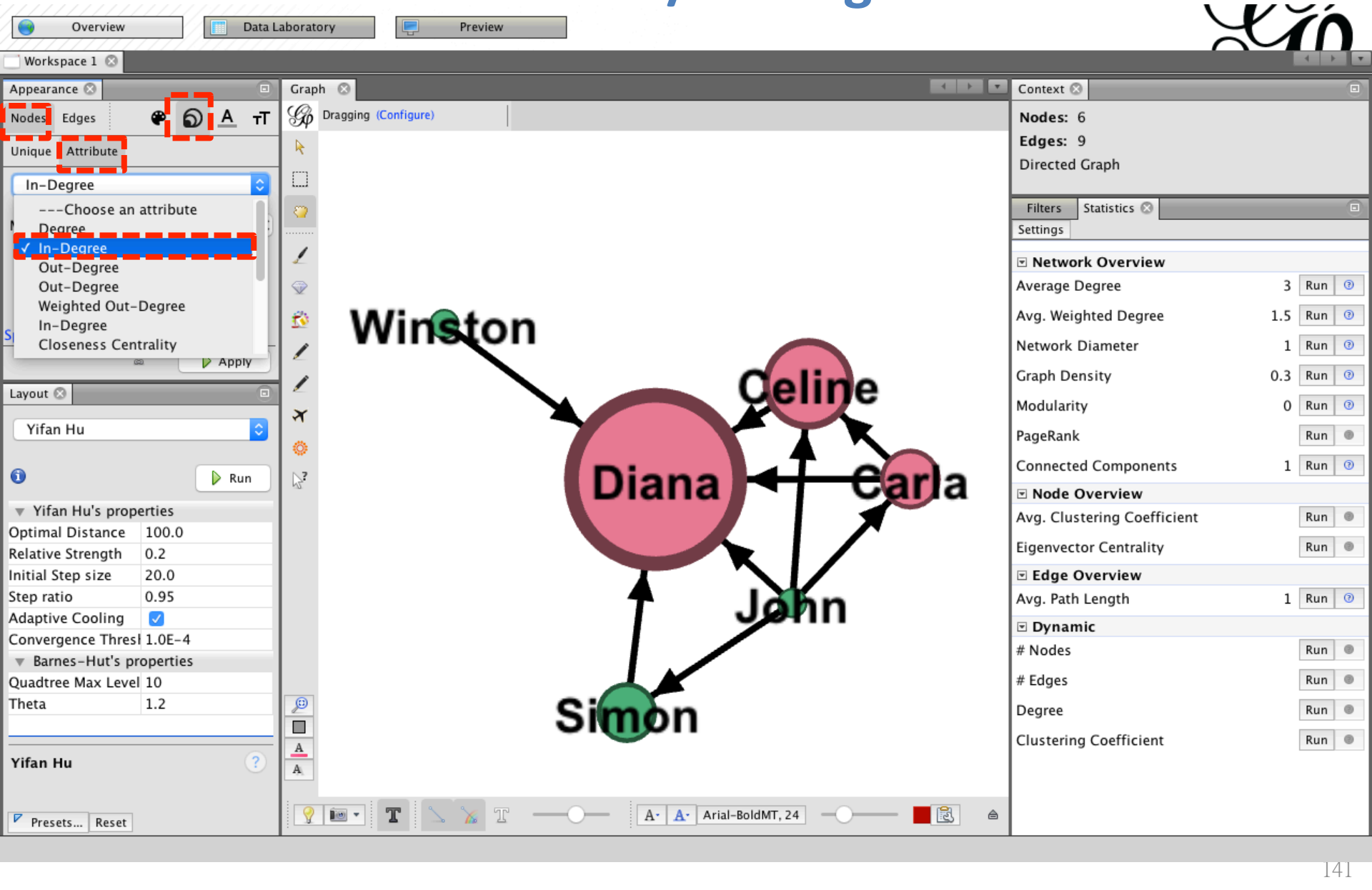


Appearance Nodes Size



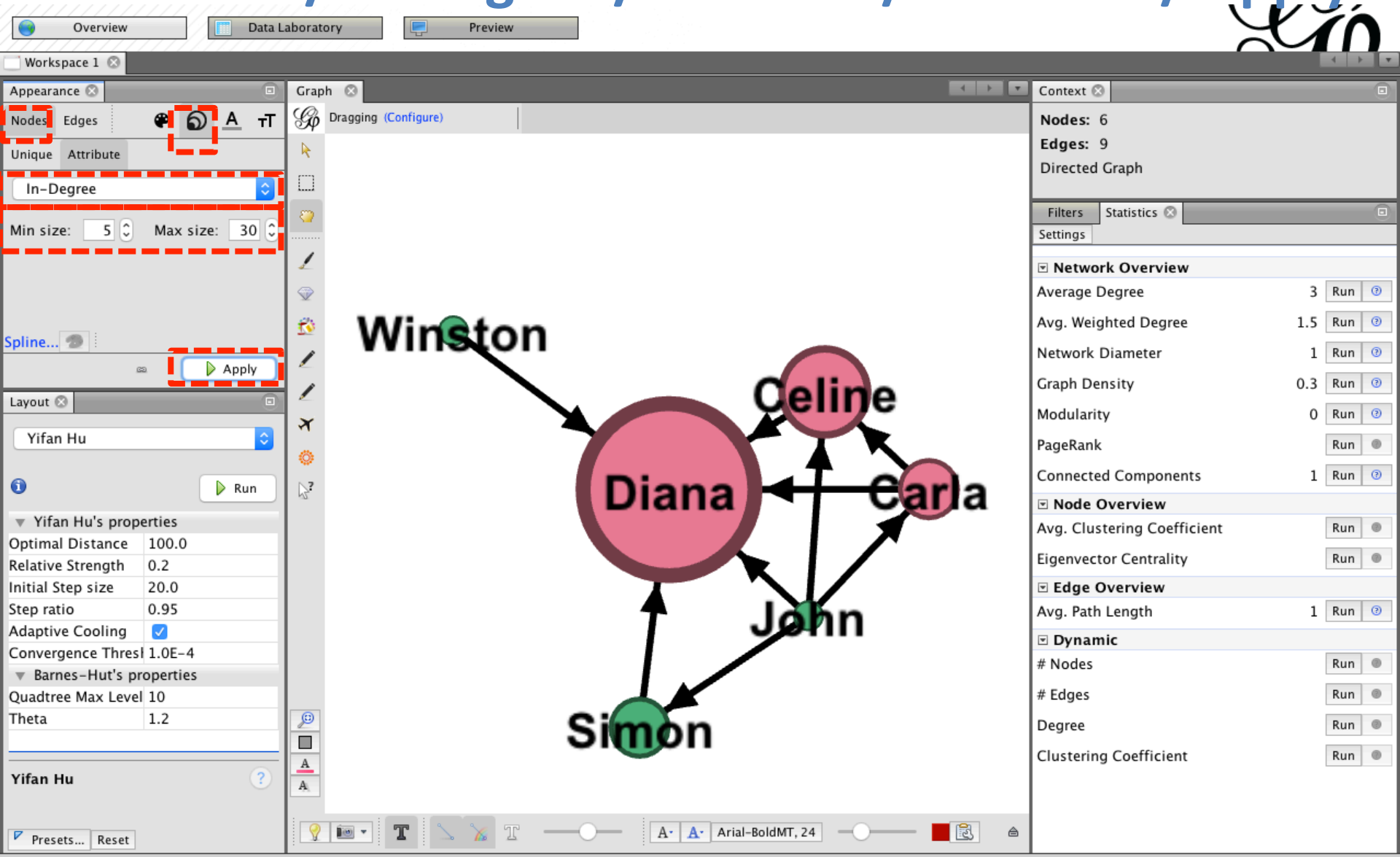
Appearance Nodes Size

Attribute / In-Degree



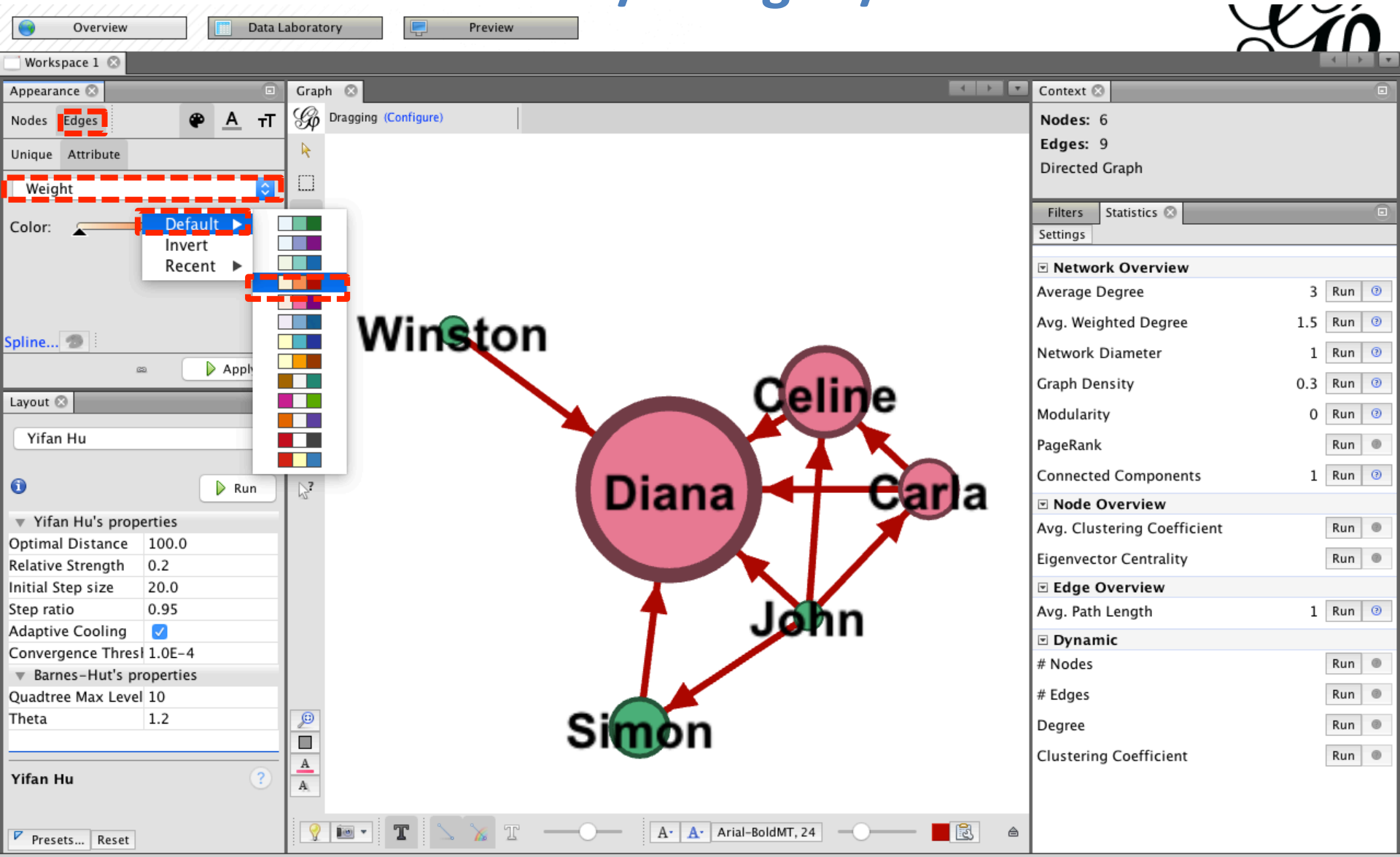
Appearance Nodes Size

Attribute / In-Degree / Min size / Max size / Apply



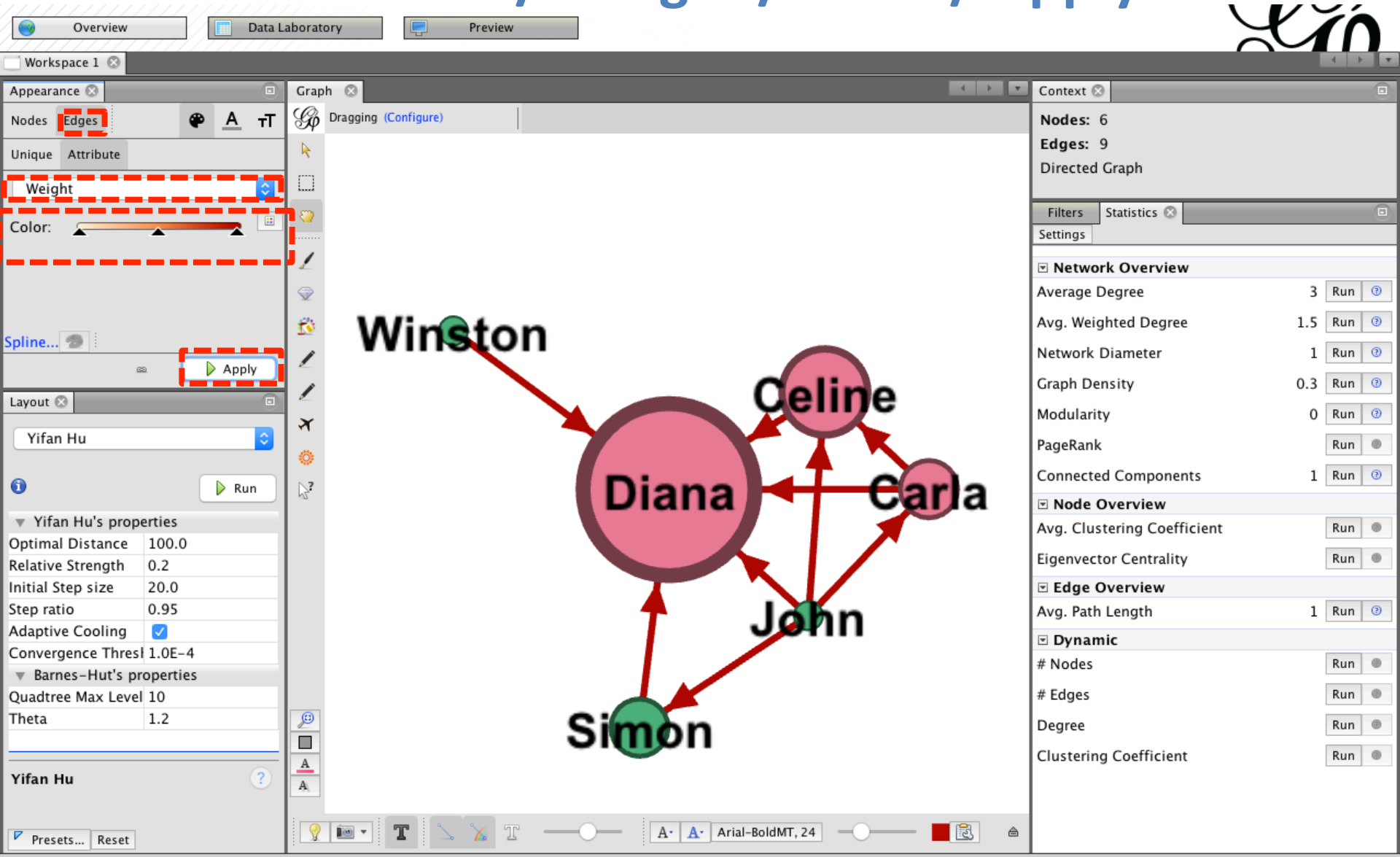
Appearance Edges

Attribute / Weight / Color

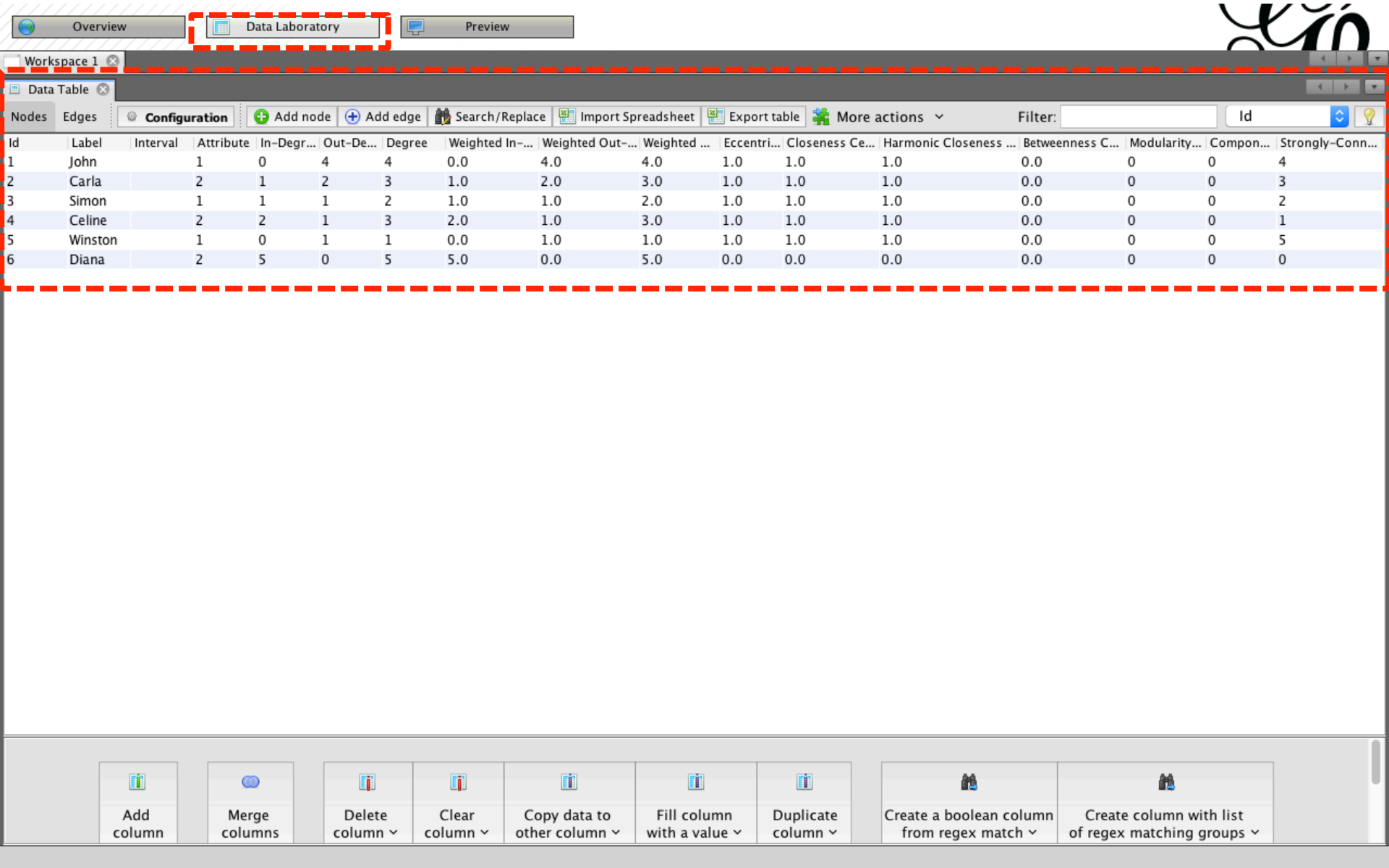


Appearance Edges

Attribute / Weight / Color / Apply

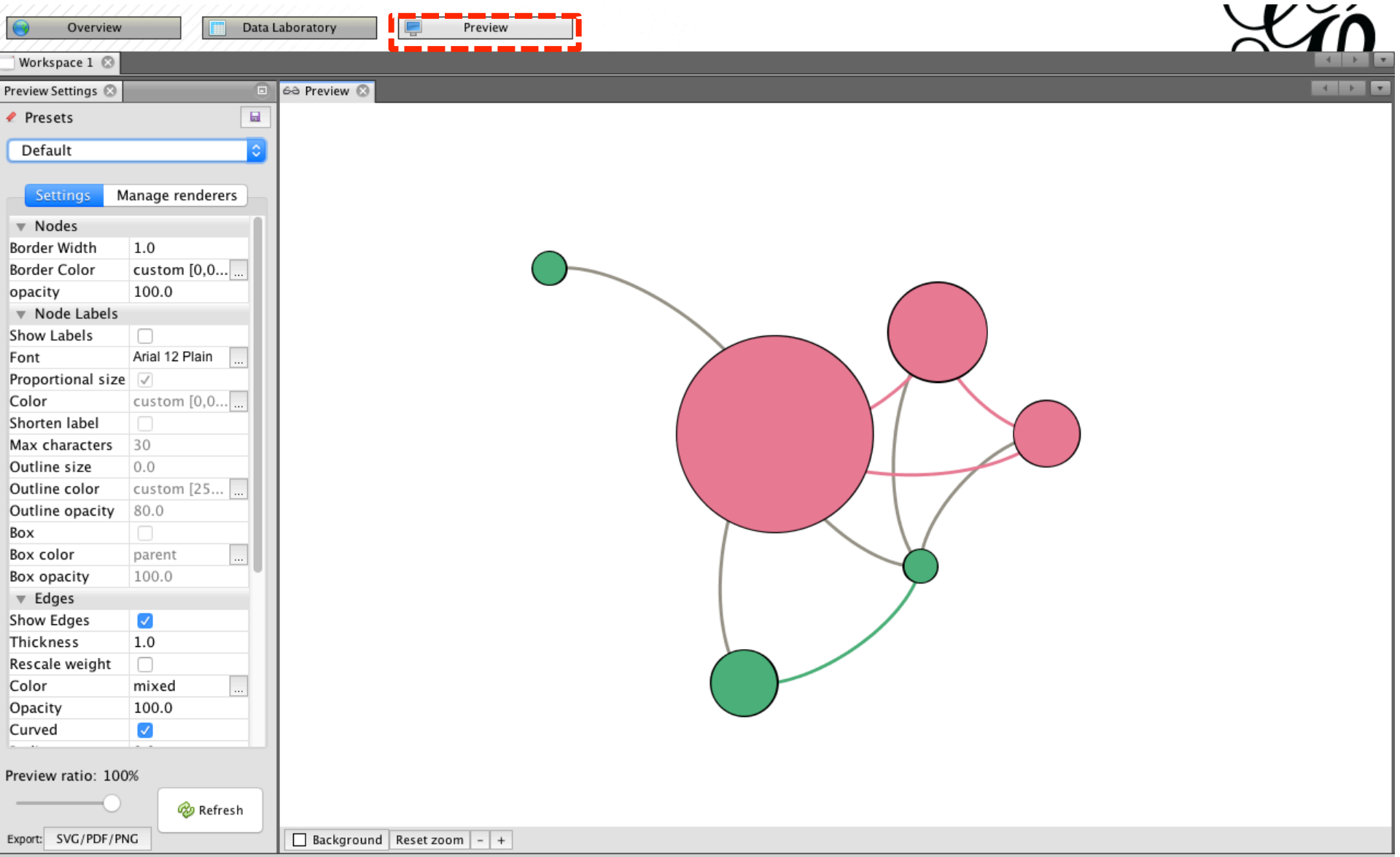


Gephi Data Laboratory

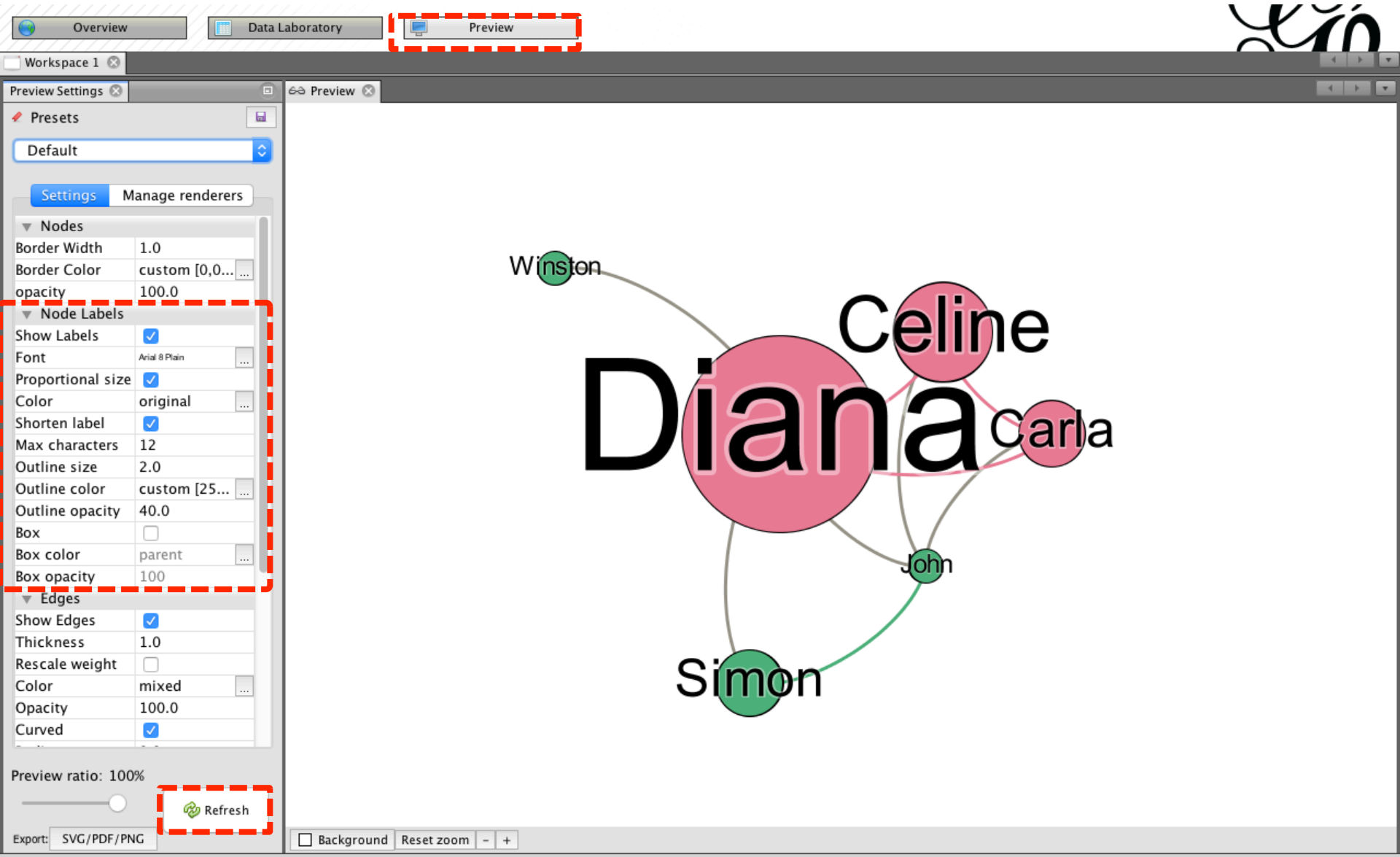


Id	Label	Interval	Attribute	In-Degr...	Out-De...	Degree	Weighted In-...	Weighted Out-...	Weighted ...	Eccentri...	Closeness Ce...	Harmonic Closeness ...	Betweenness C...	Modularity...	Compon...	Strongly-Conn...
1	John		1	0	4	4	0.0	4.0	4.0	1.0	1.0	1.0	0.0	0	0	4
2	Carla		2	1	2	3	1.0	2.0	3.0	1.0	1.0	1.0	0.0	0	0	3
3	Simon		1	1	1	2	1.0	1.0	2.0	1.0	1.0	1.0	0.0	0	0	2
4	Celine		2	2	1	3	2.0	1.0	3.0	1.0	1.0	1.0	0.0	0	0	1
5	Winston		1	0	1	1	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0	0	5
6	Diana		2	5	0	5	5.0	0.0	5.0	0.0	0.0	0.0	0.0	0	0	0

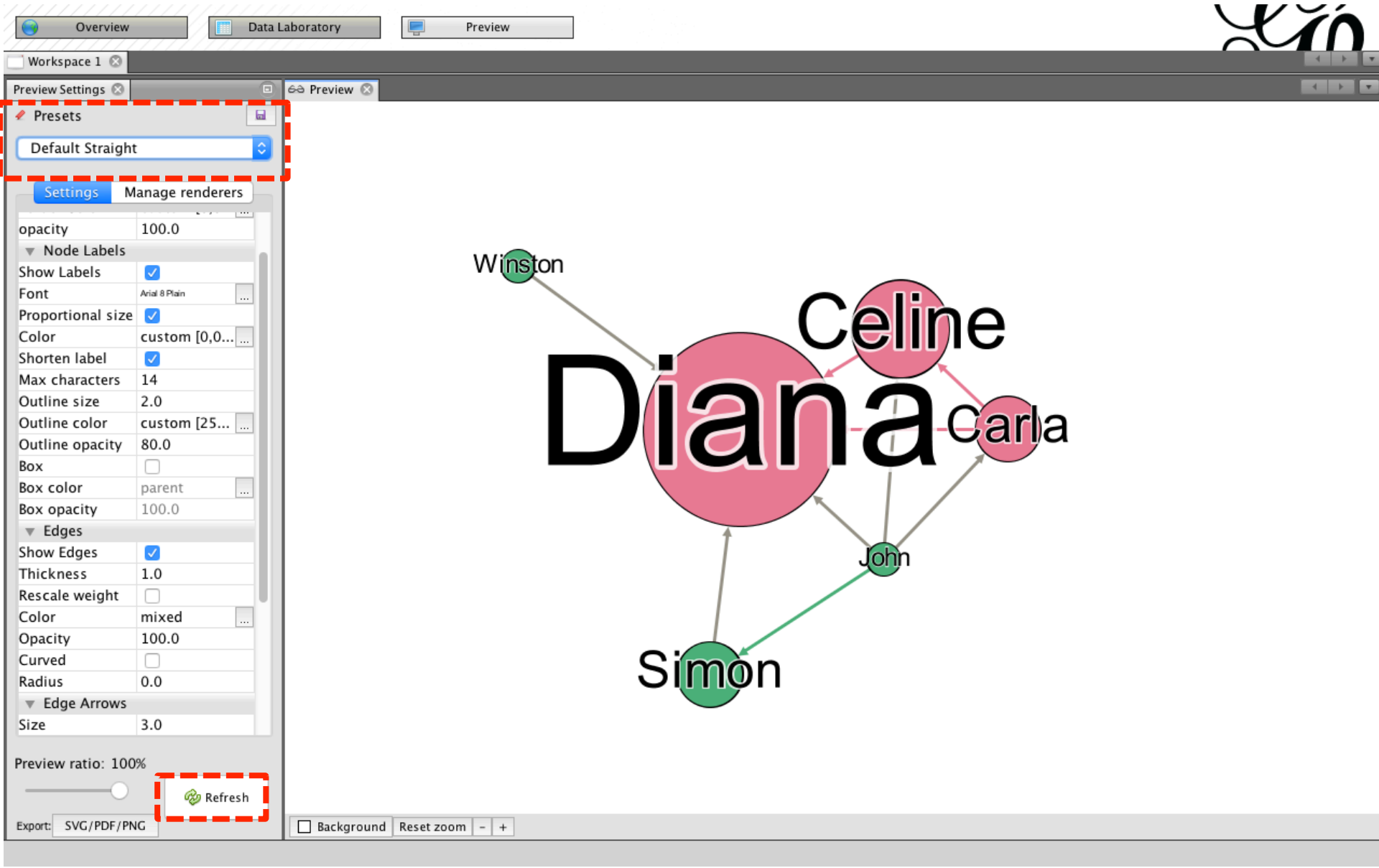
Gephi Preview



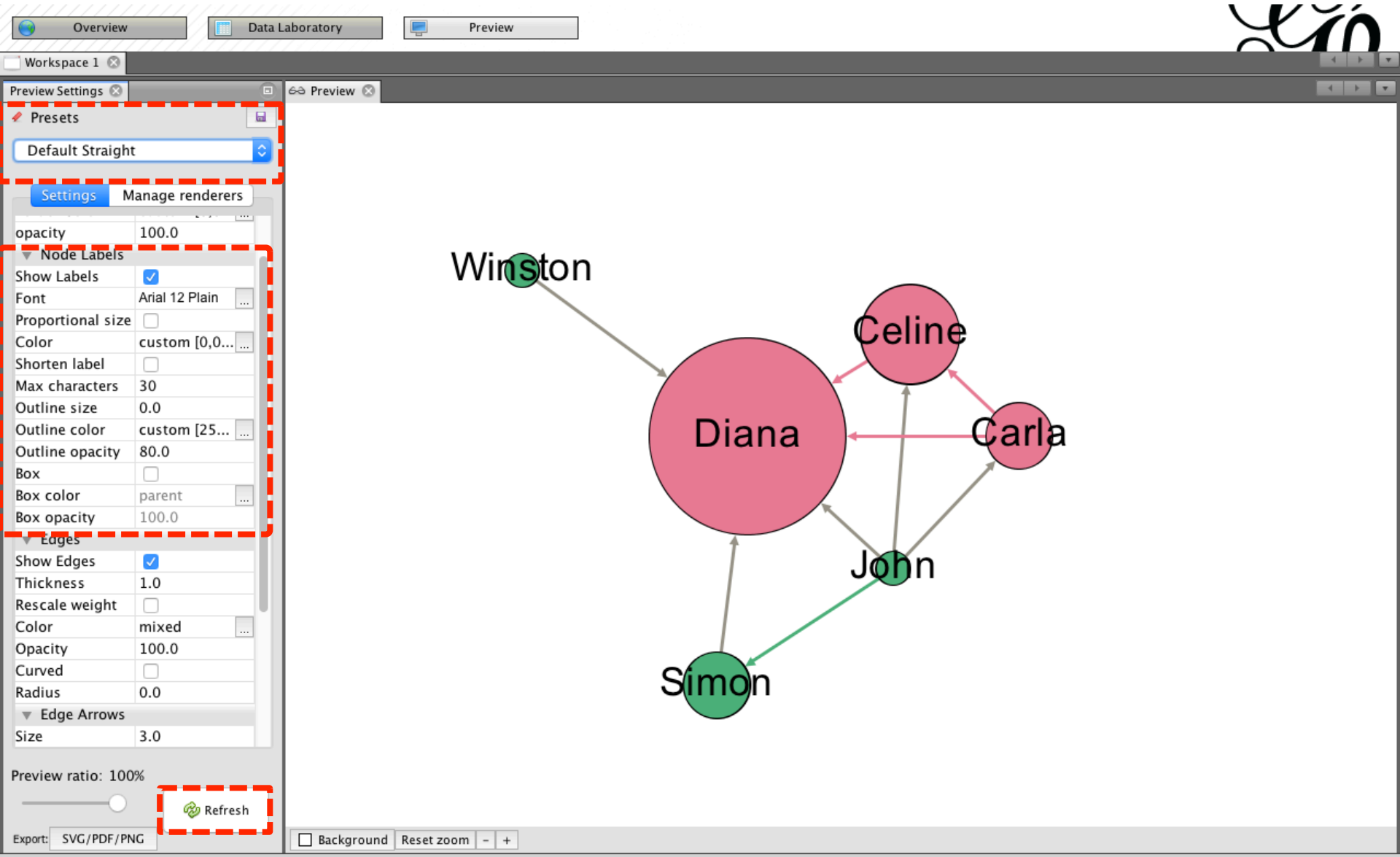
Gephi Preview: Show Labels



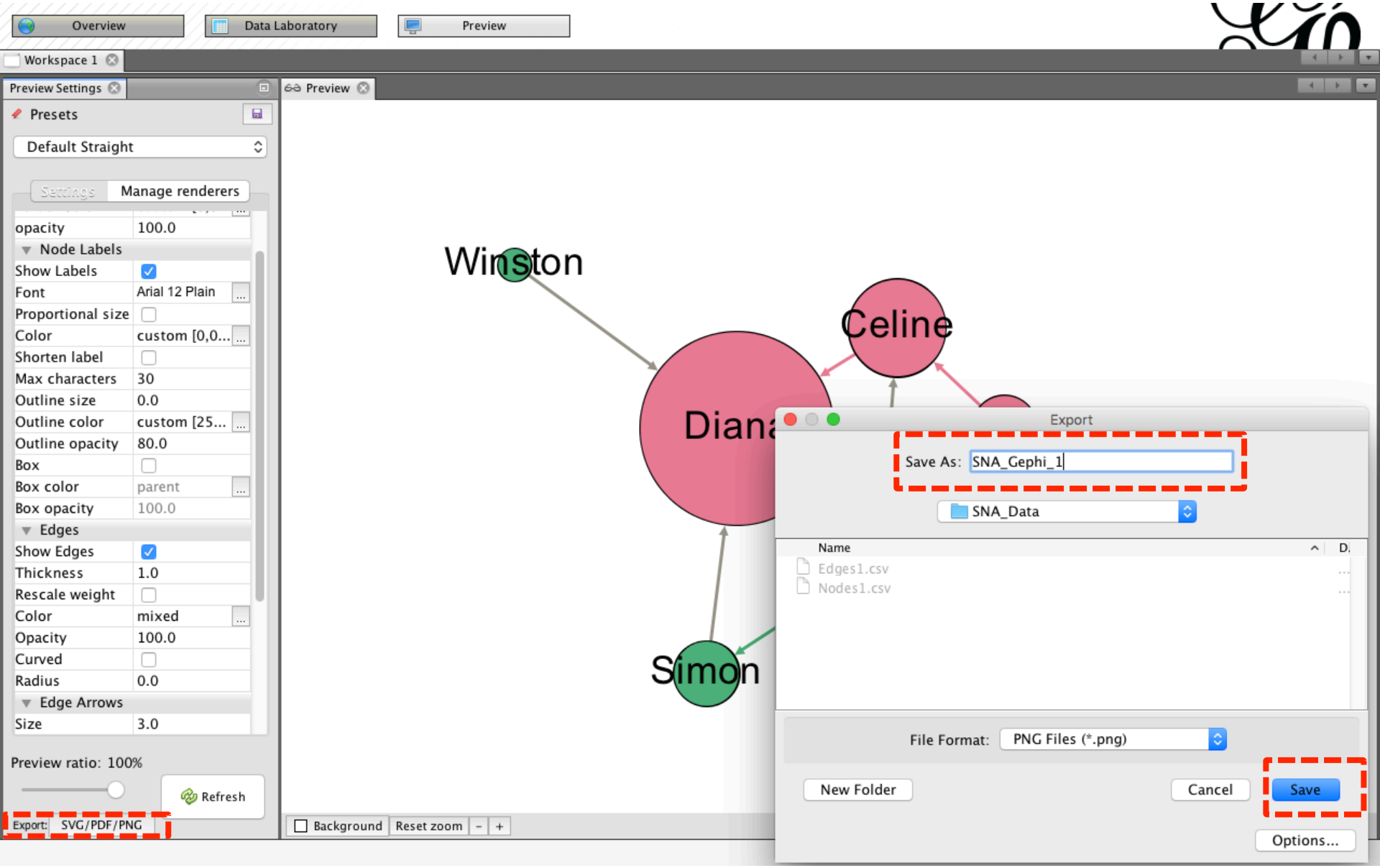
Gephi Preview: Default Straight



Gephi Preview: Default Straight

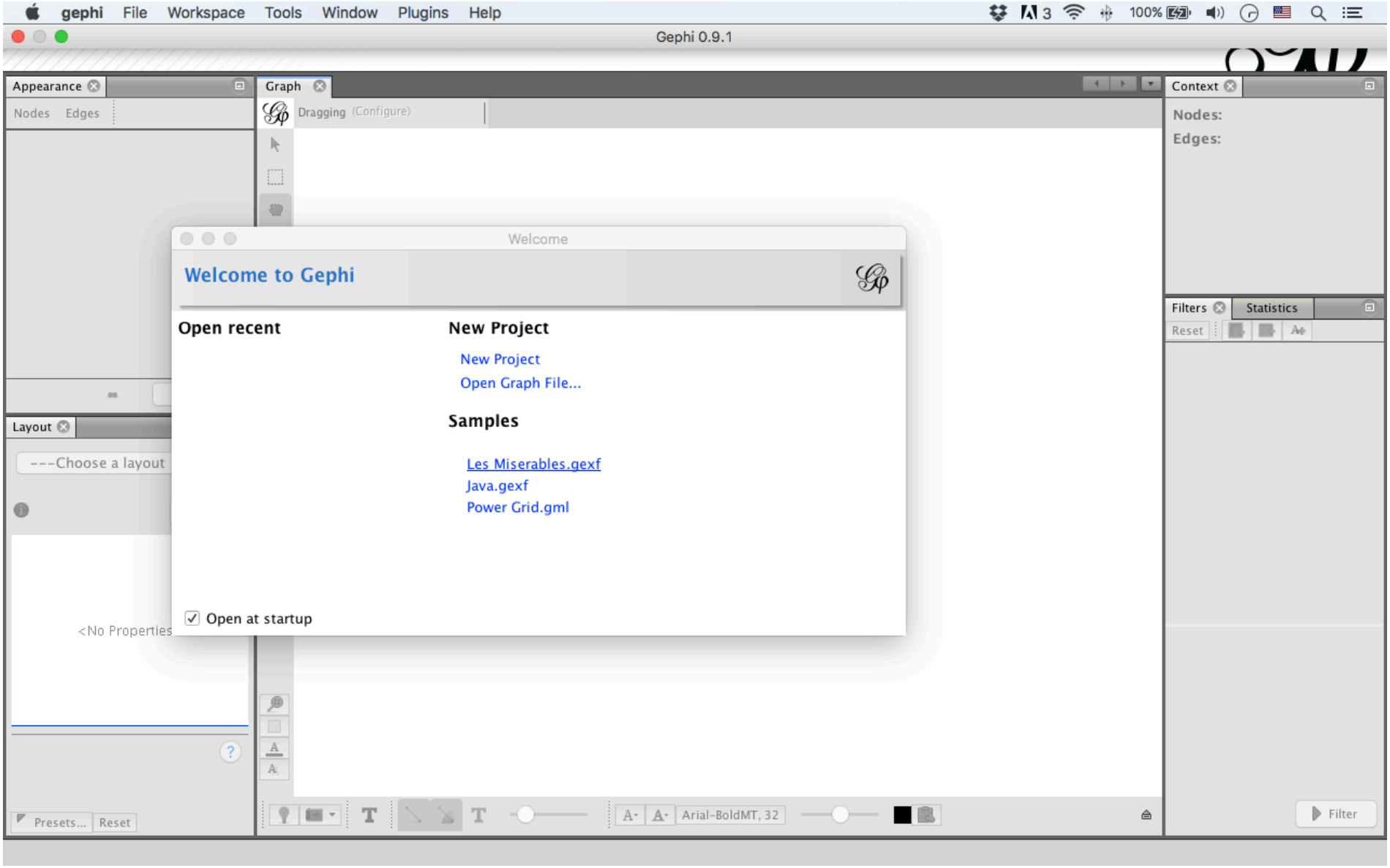


Gephi Preview: Export SVG/PDF/PNG

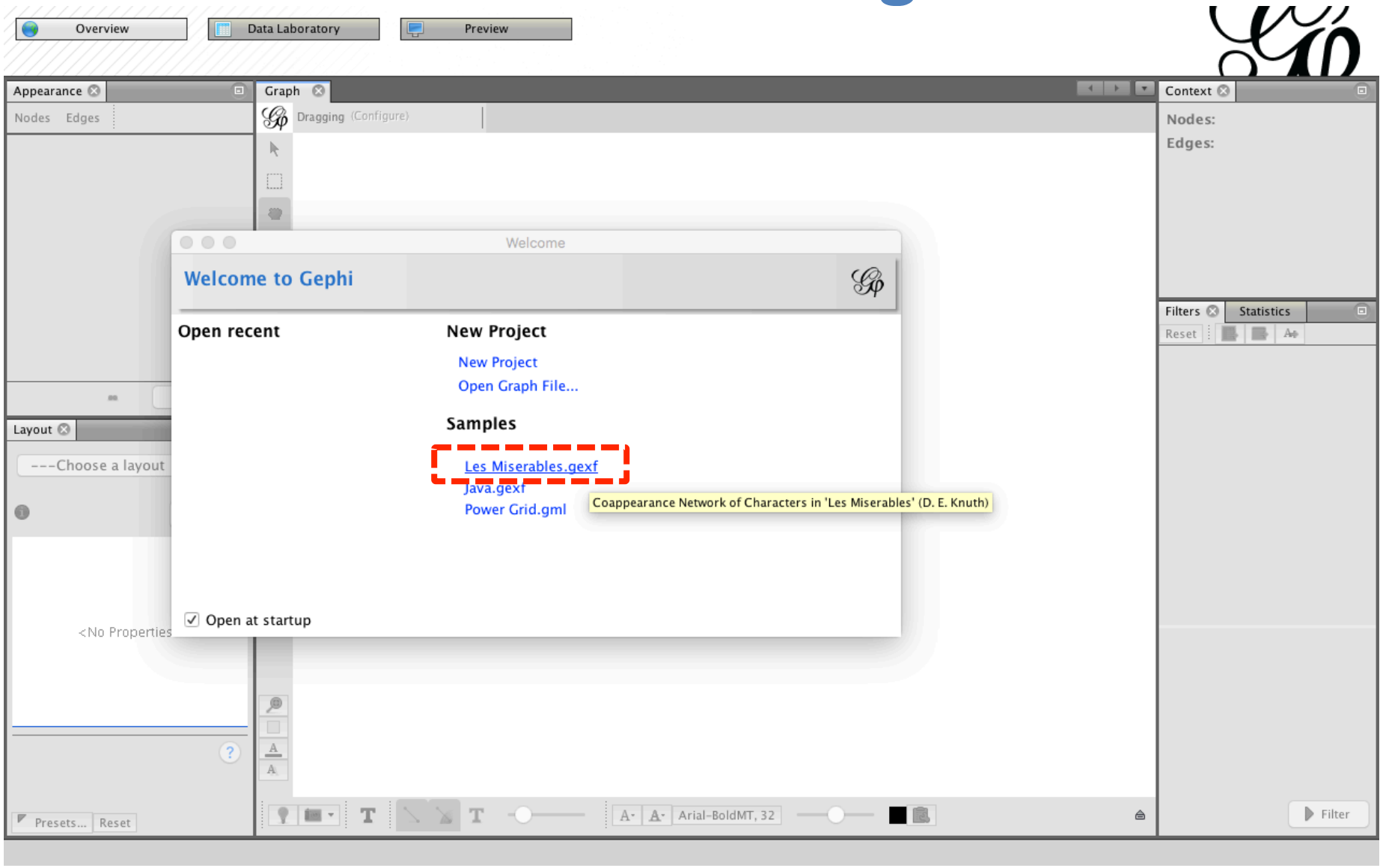


Open Gephi Samples

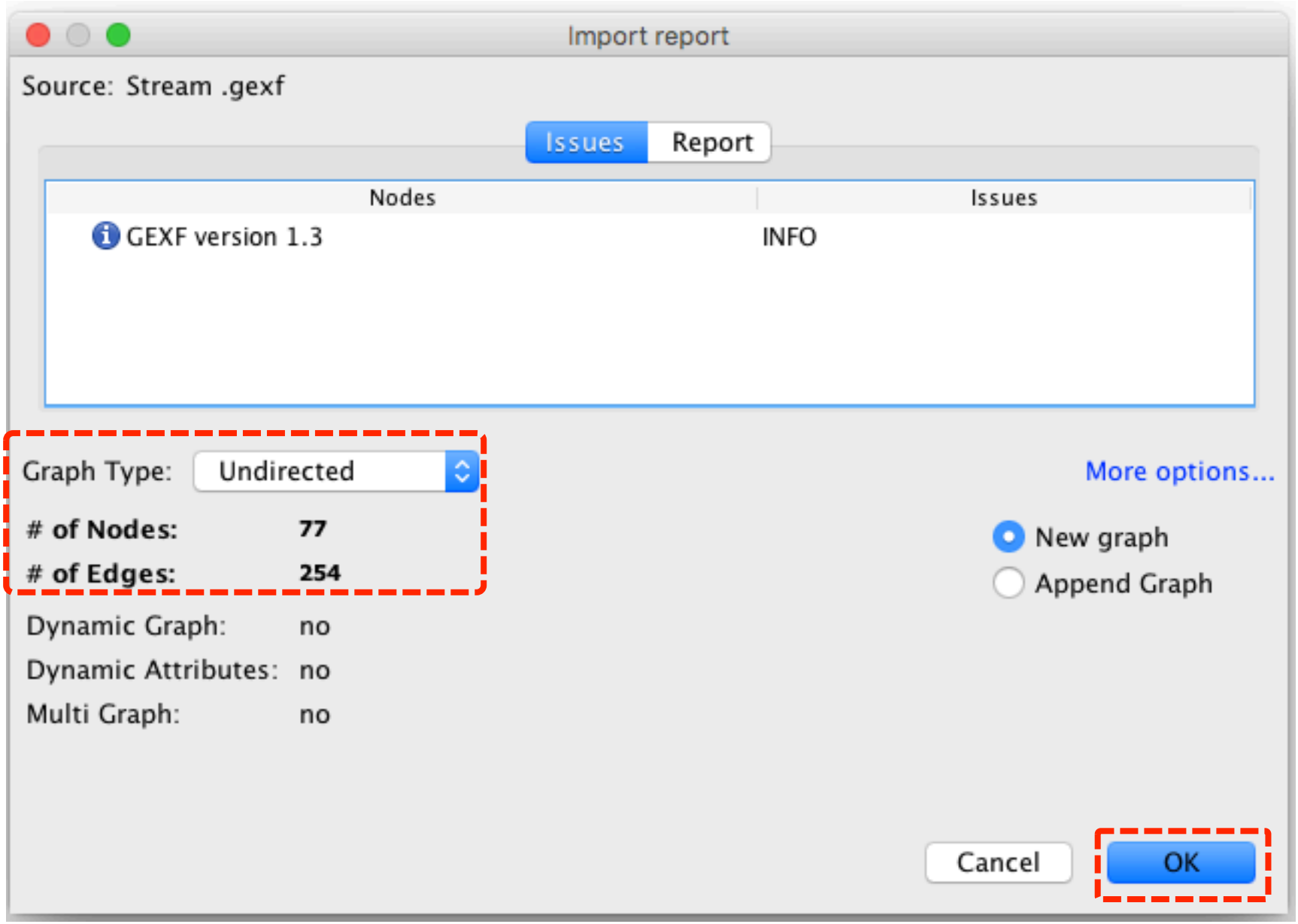
Gephi Samples



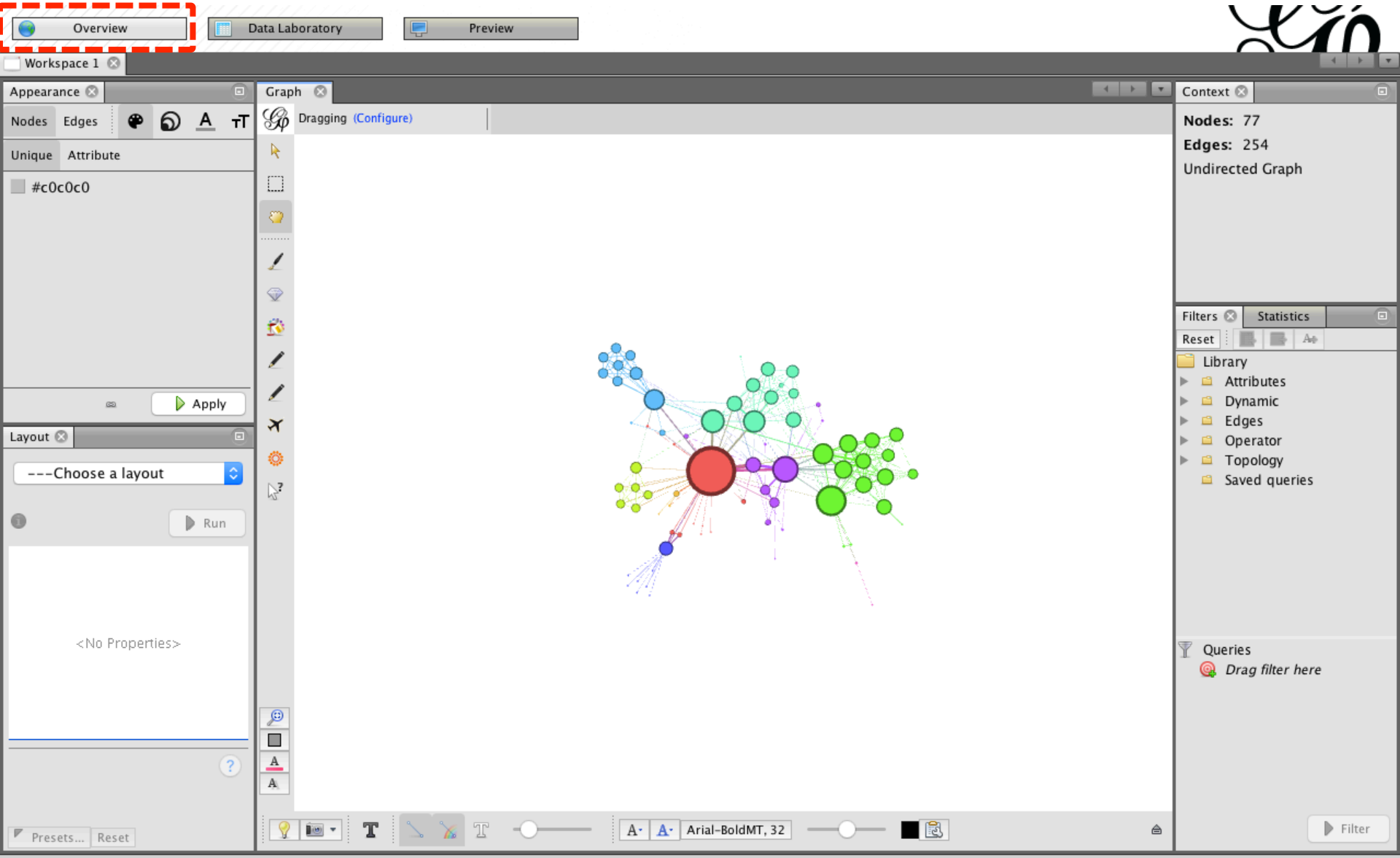
Gephi Samples
Les Miserables.gexf



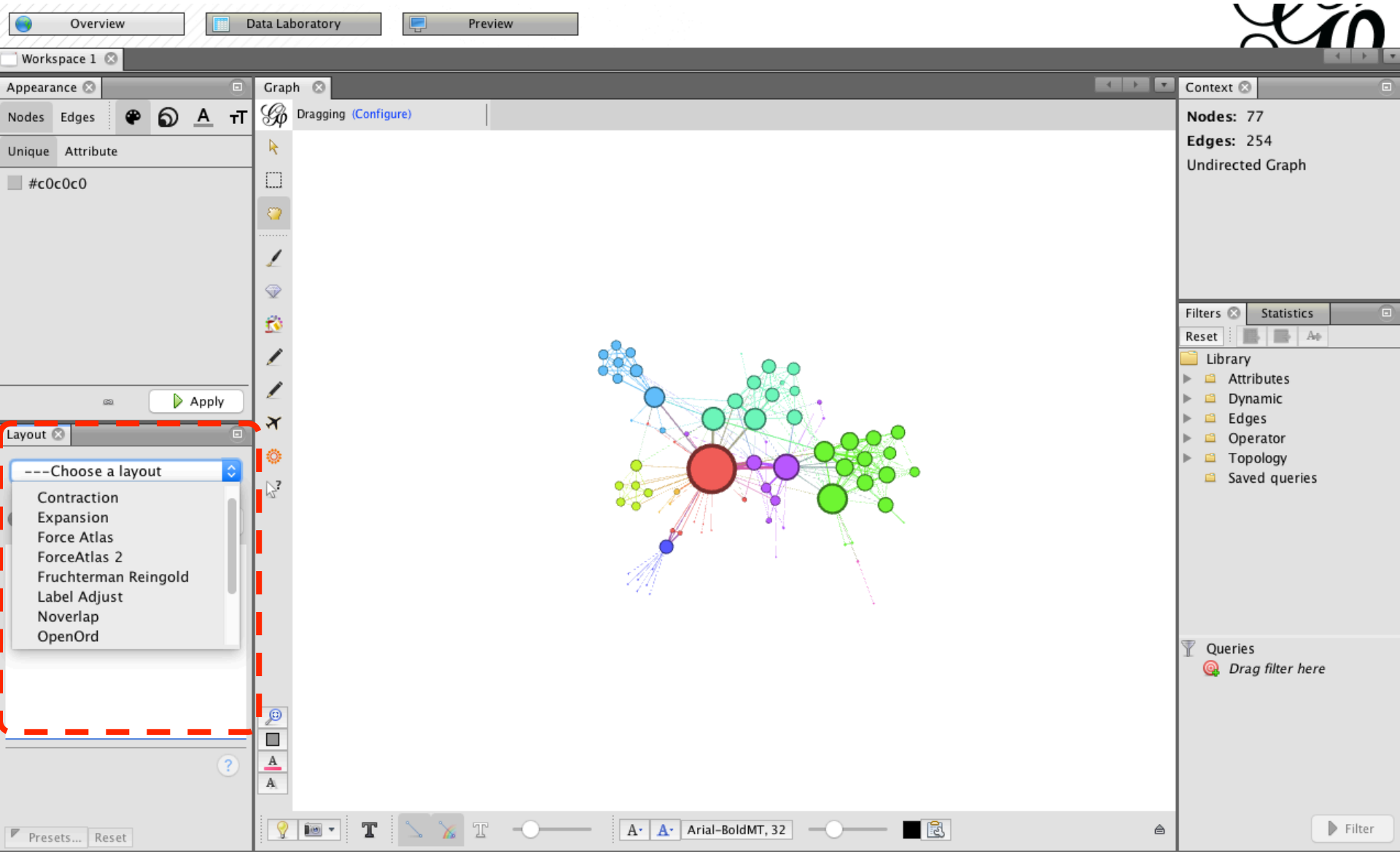
Gephi Import Report



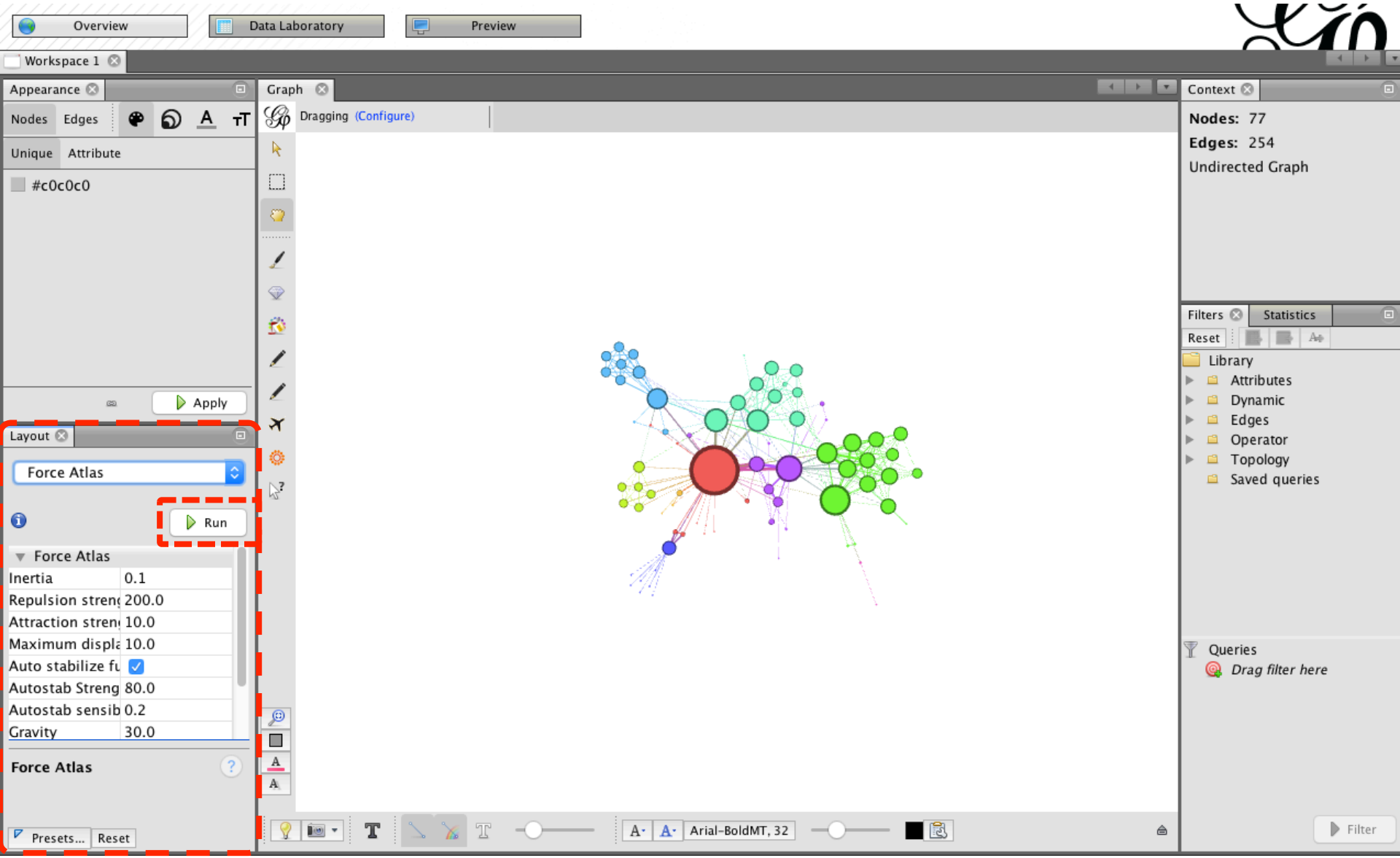
Gephi Overview



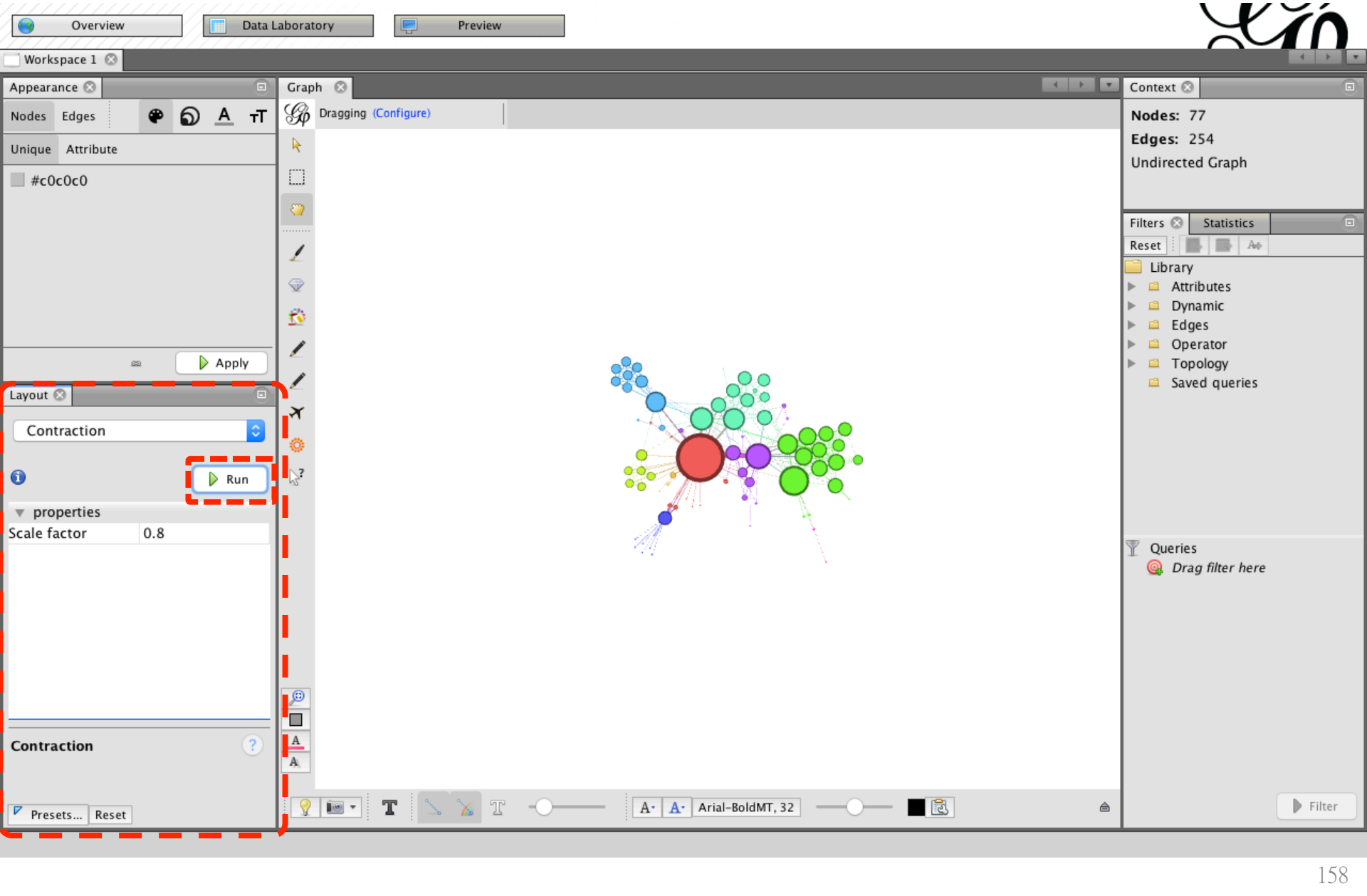
Gephi Layout



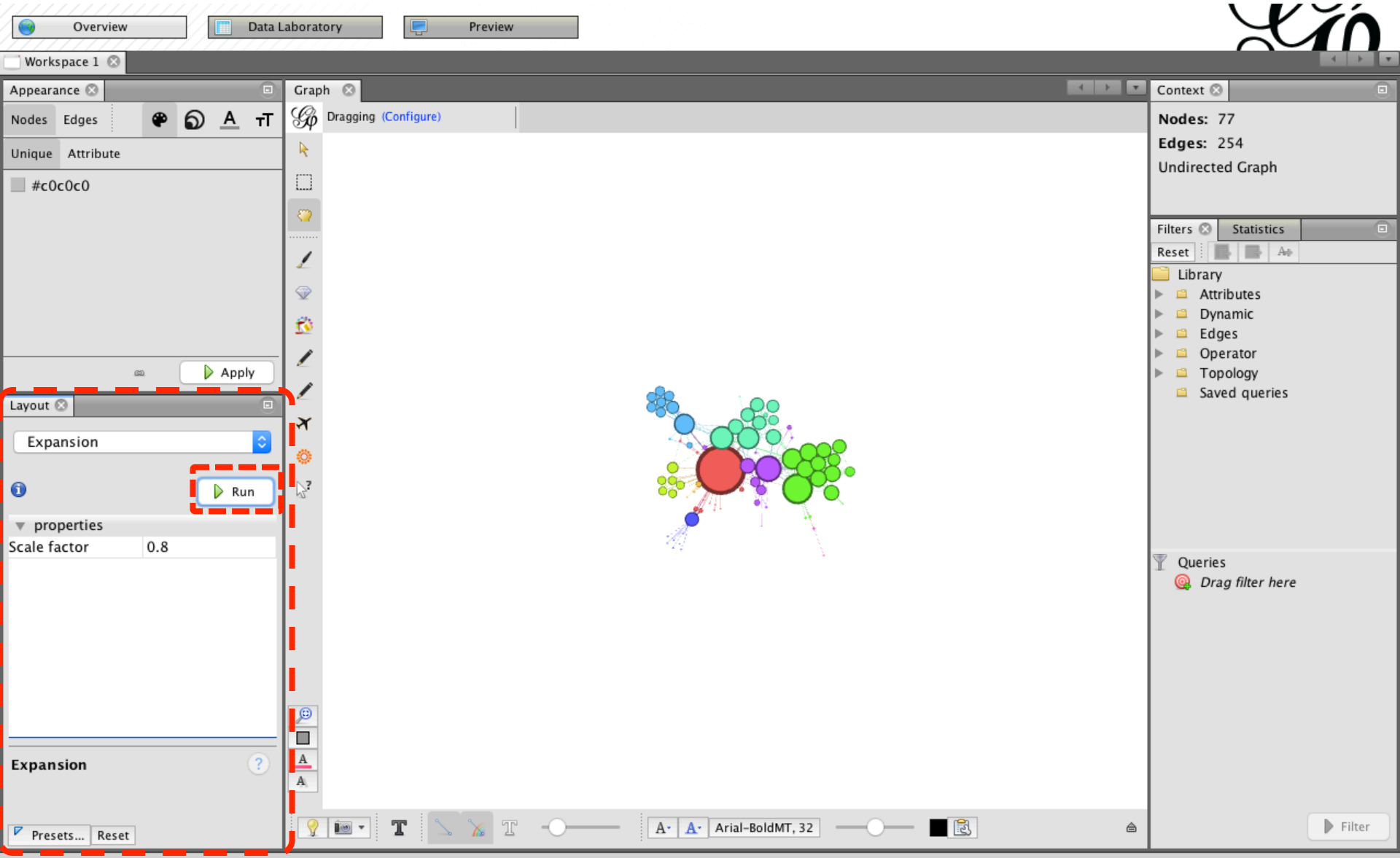
Gephi Layout: Force Atlas



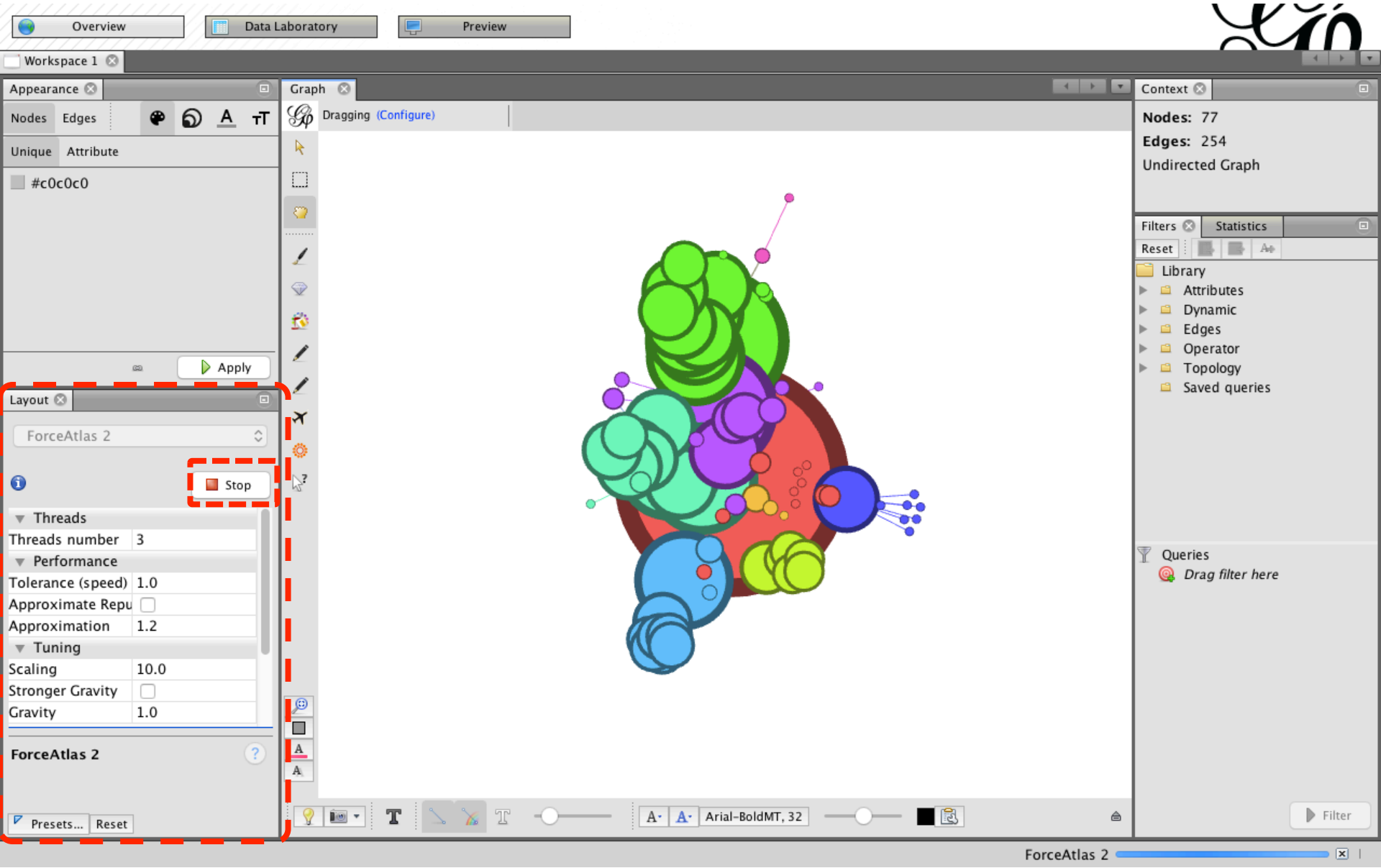
Gephi Layout: Contraction



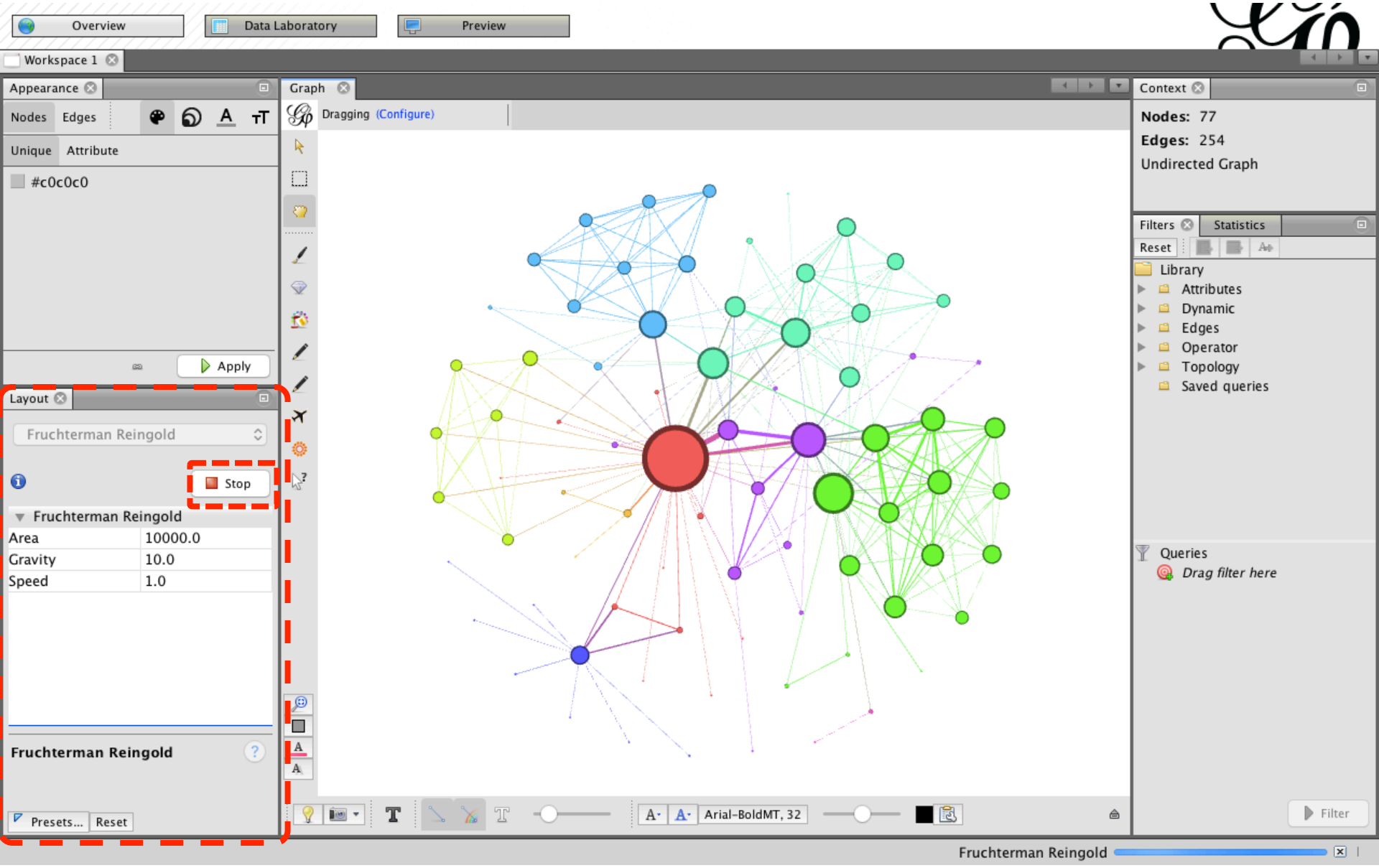
Gephi Layout: Expansion



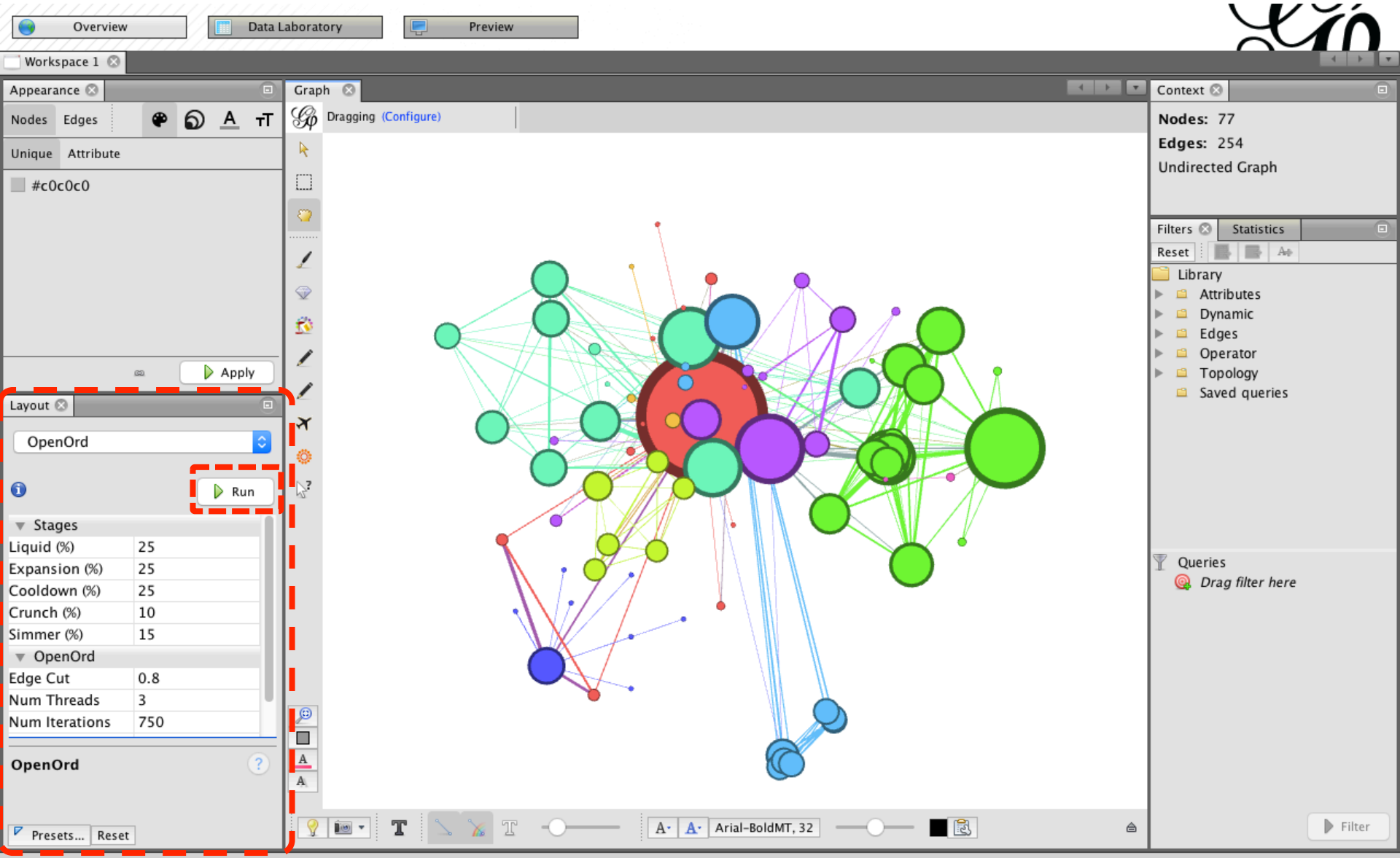
Gephi Layout: ForceAtlas 2



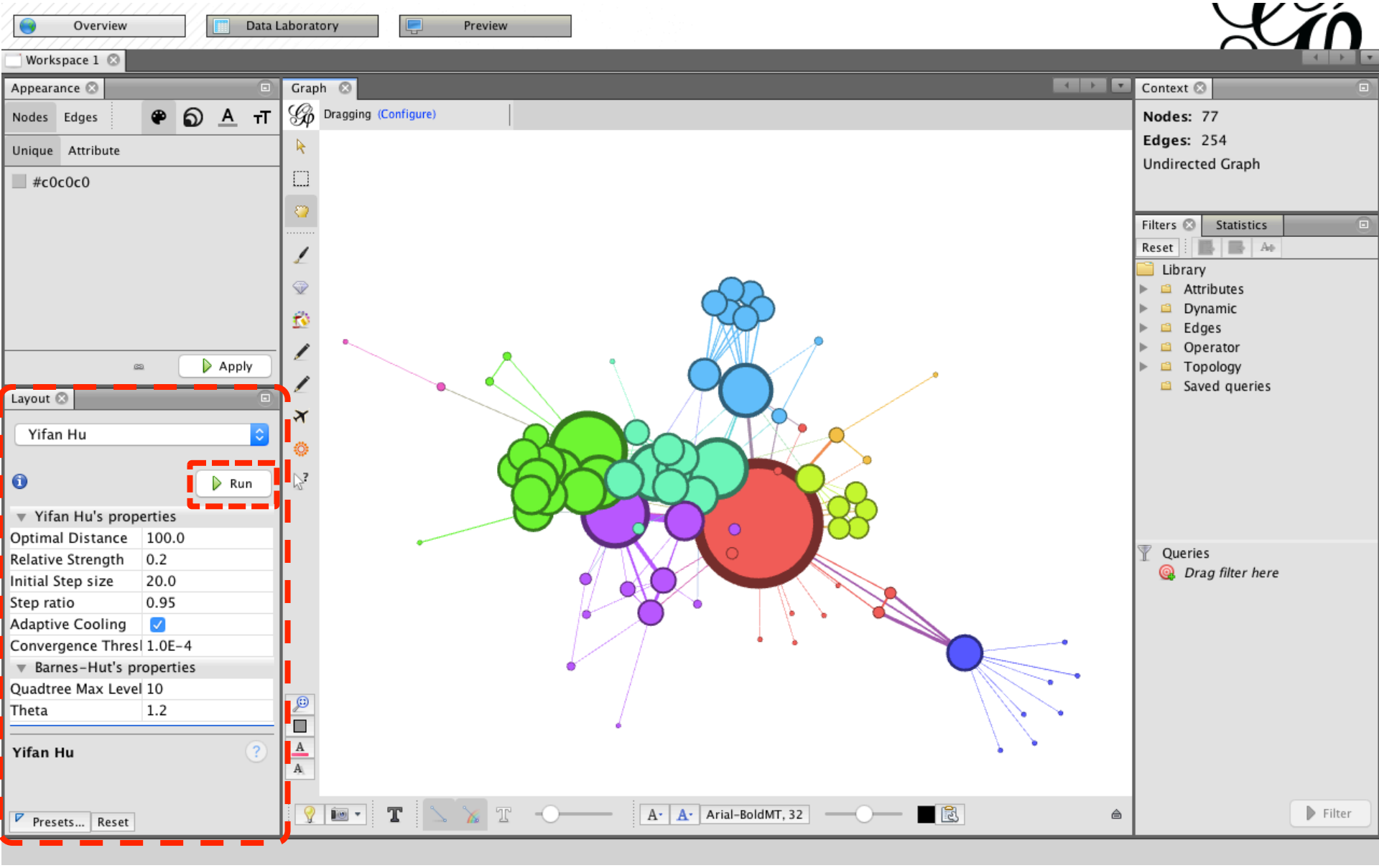
Gephi Layout: Fruchterman Reingold



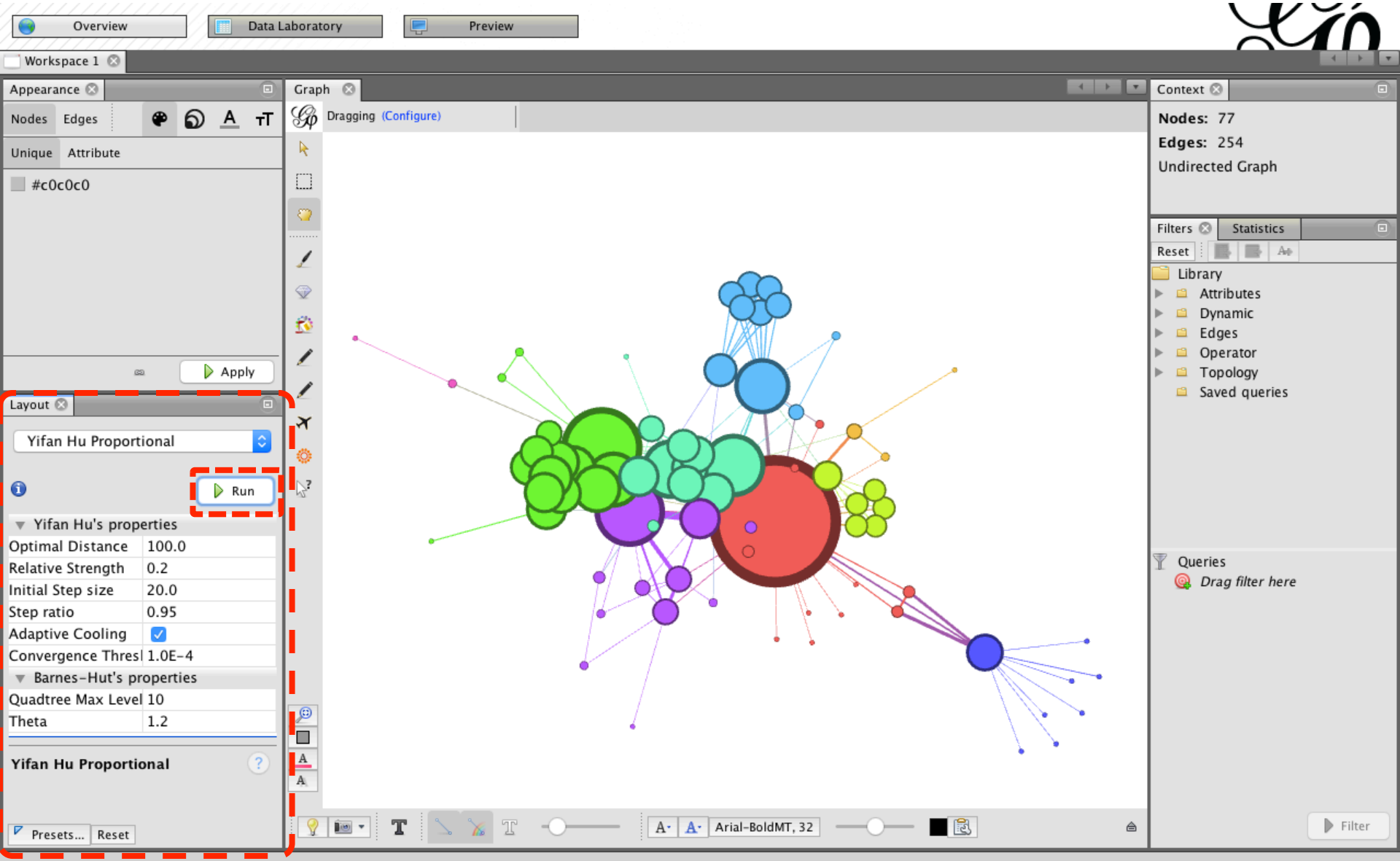
Gephi Layout: OpenOrd



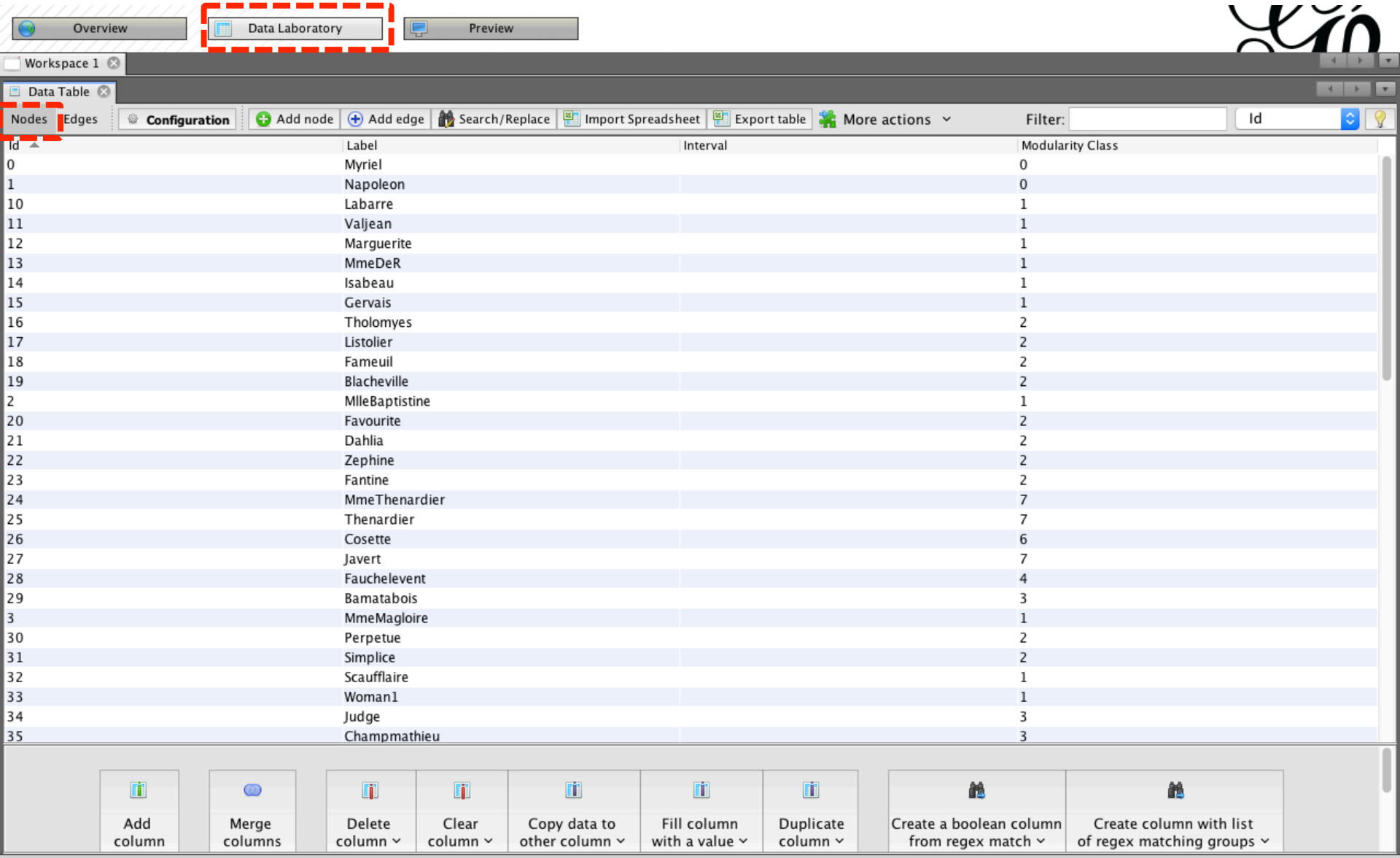
Gephi Layout: Yifan Hu



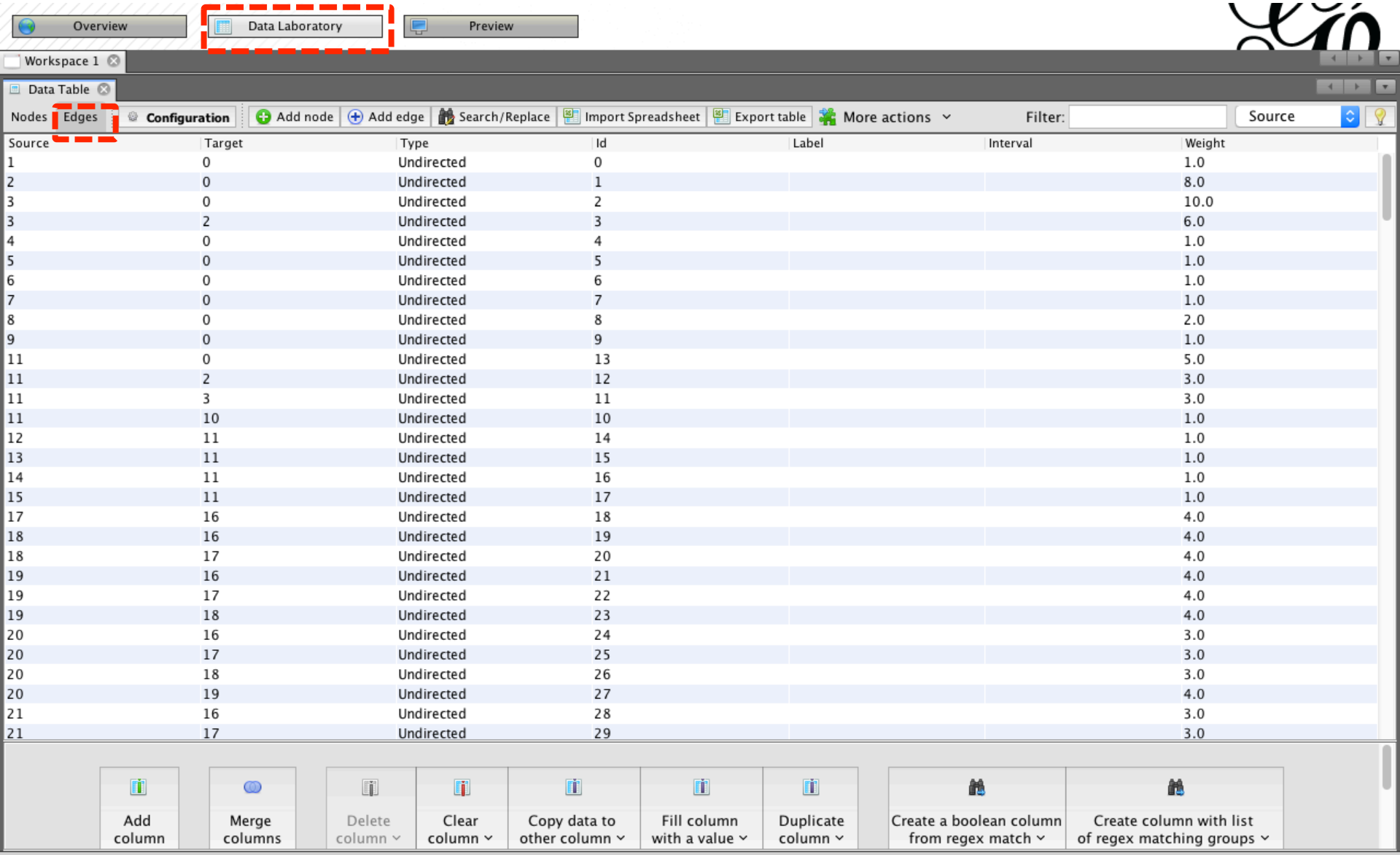
Gephi Layout: Yifan Hu Proportional



Gephi Data Laboratory: Nodes

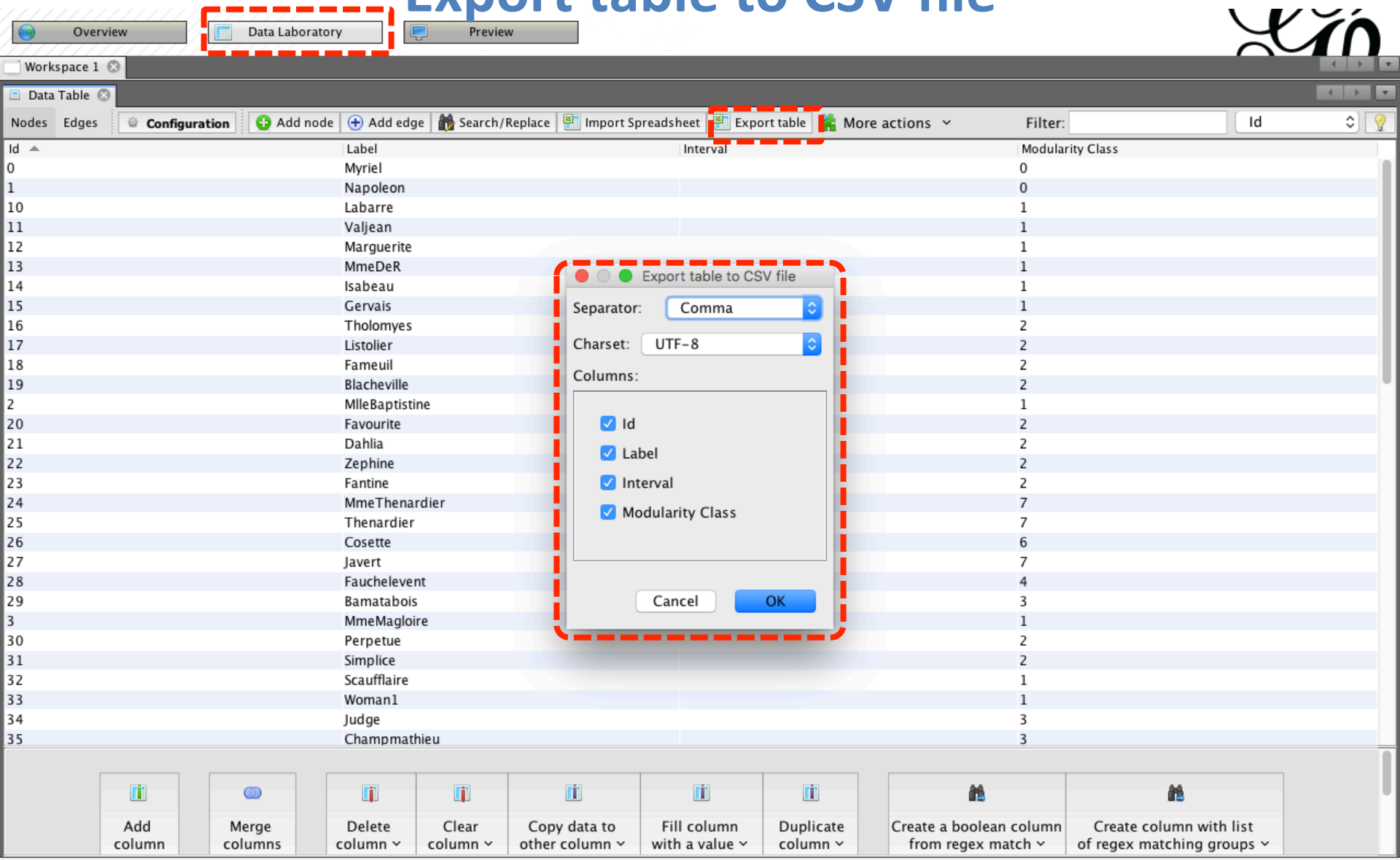


Gephi Data Laboratory: Edges

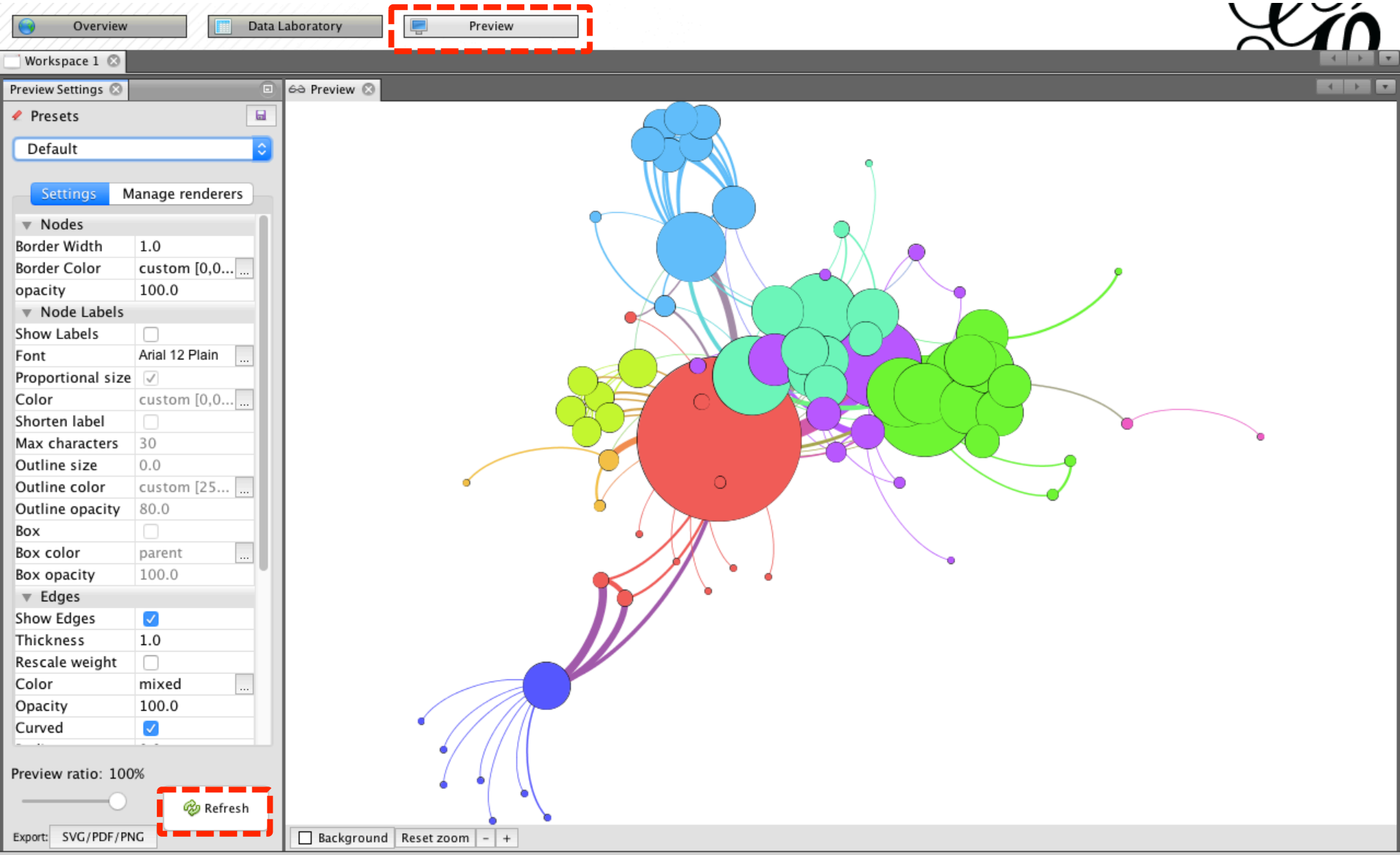


Source	Target	Type	Id	Label	Interval	Weight
1	0	Undirected	0			1.0
2	0	Undirected	1			8.0
3	0	Undirected	2			10.0
3	2	Undirected	3			6.0
4	0	Undirected	4			1.0
5	0	Undirected	5			1.0
6	0	Undirected	6			1.0
7	0	Undirected	7			1.0
8	0	Undirected	8			2.0
9	0	Undirected	9			1.0
11	0	Undirected	13			5.0
11	2	Undirected	12			3.0
11	3	Undirected	11			3.0
11	10	Undirected	10			1.0
12	11	Undirected	14			1.0
13	11	Undirected	15			1.0
14	11	Undirected	16			1.0
15	11	Undirected	17			1.0
17	16	Undirected	18			4.0
18	16	Undirected	19			4.0
18	17	Undirected	20			4.0
19	16	Undirected	21			4.0
19	17	Undirected	22			4.0
19	18	Undirected	23			4.0
20	16	Undirected	24			3.0
20	17	Undirected	25			3.0
20	18	Undirected	26			3.0
20	19	Undirected	27			4.0
21	16	Undirected	28			3.0
21	17	Undirected	29			3.0

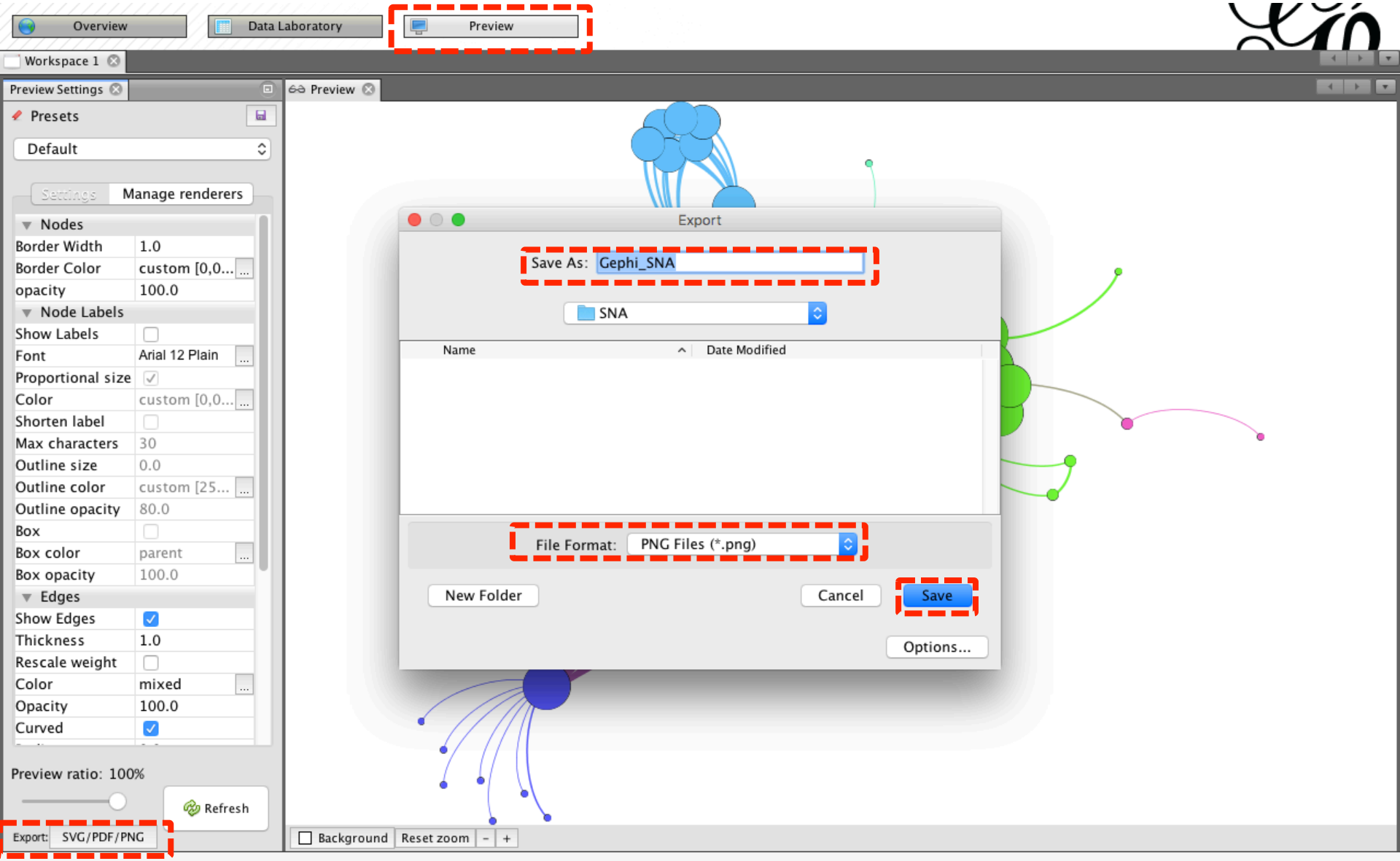
Gephi Data Laboratory:
Export table to CSV file



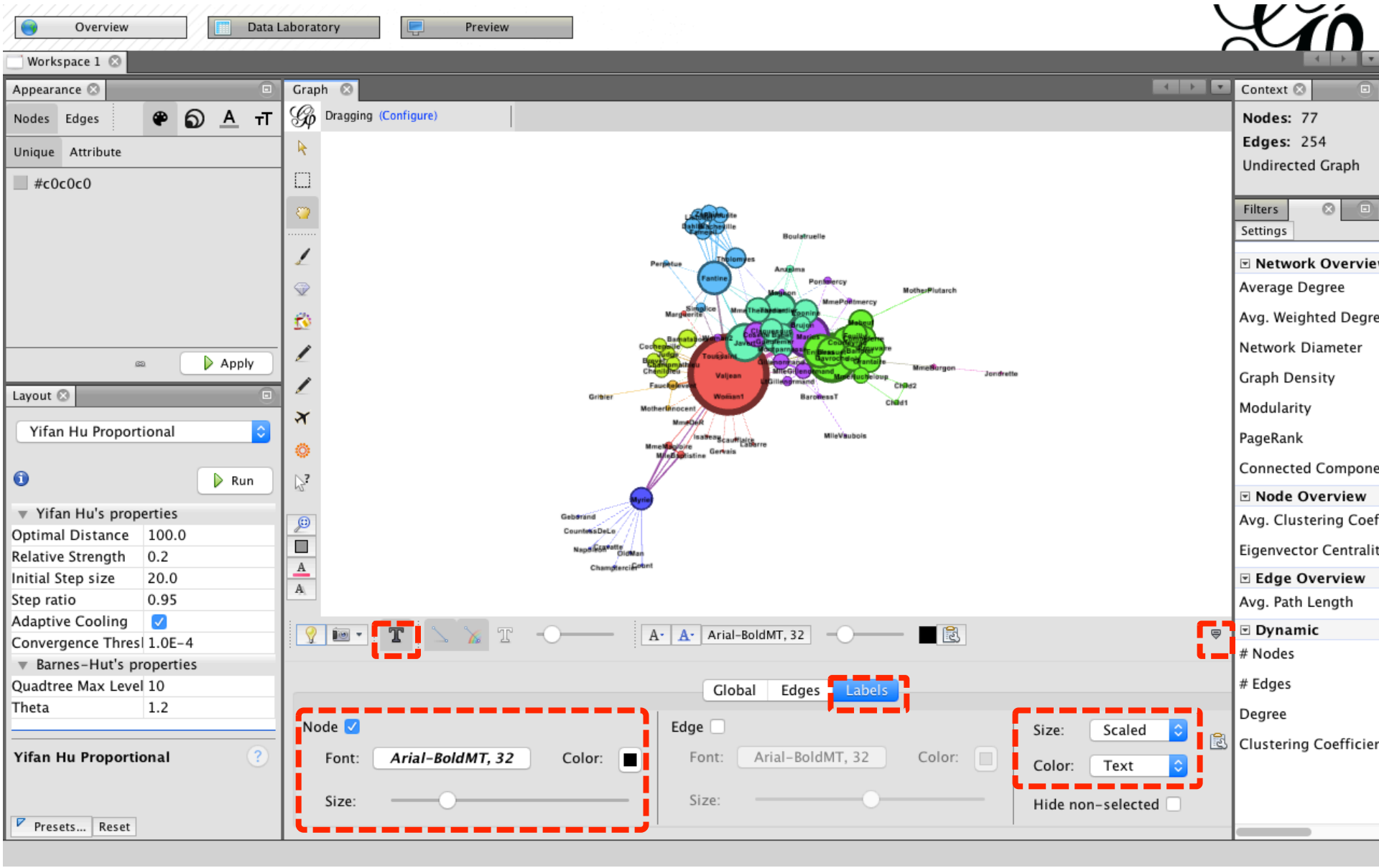
Gephi Preview



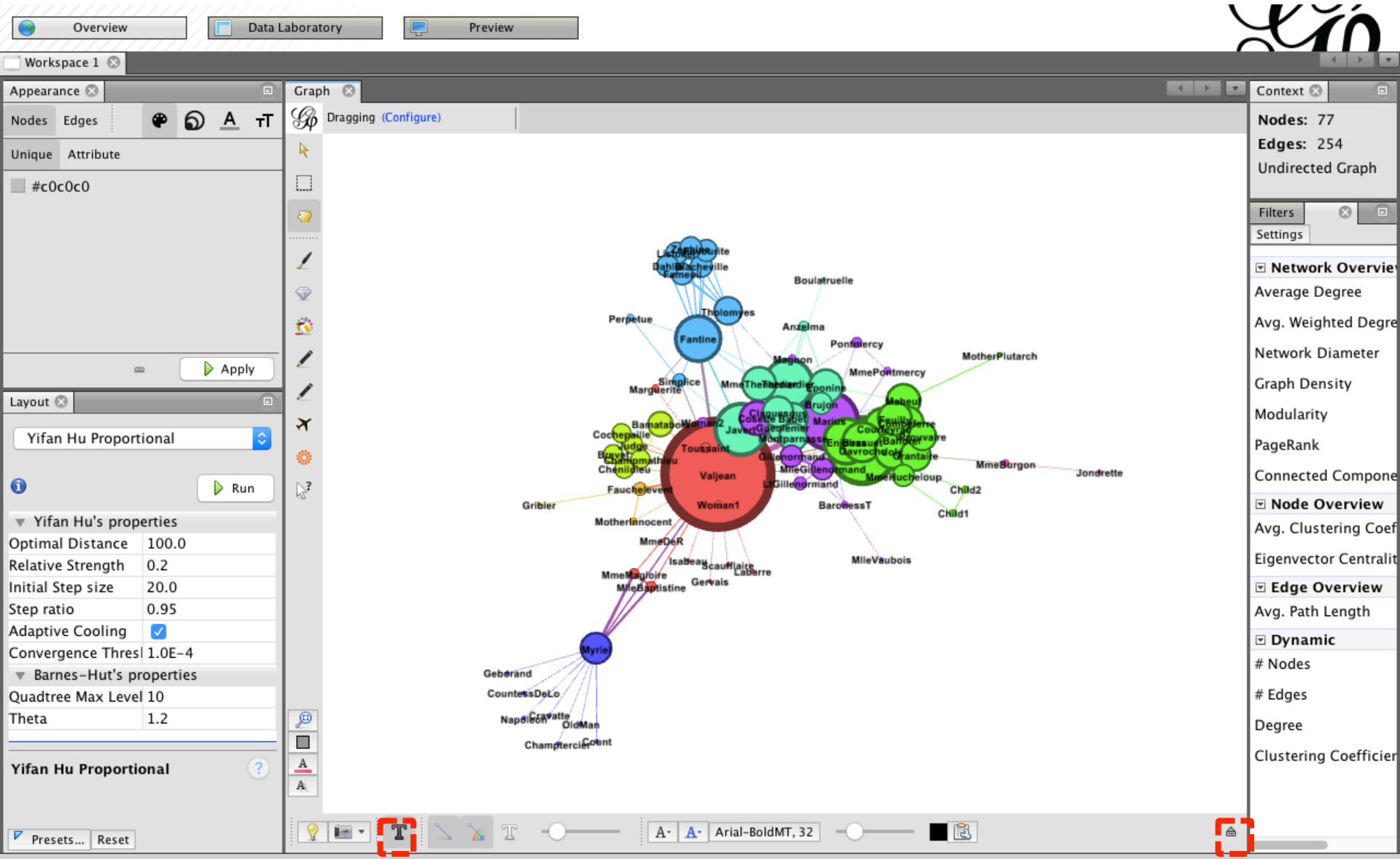
Gephi Preview: Export SVG/PDF/PNG



Gephi Overview: Text Labels



Gephi Overview: Text Labels



Comparison of Social Network Analysis (SNA) Tools

General Comparison of SNA Tools

Software	NETWORKX	IGRAPH	GEPHI	PAJEK
TYPE	LIBRARY	LIBRARY	STAND ALONE	STAND ALONE
PLATFORM	PYTHON	PYTHON\R\C LIBRARY	WINDOWS	WINDOWS
COMPUTATIONAL TIME	FAST	FAST	FAST	MEDIUM
NO. OF NODES	1 MILLION	1 MILLION	0.15 MILLION	1 MILLION

Naheed Akhtar (2014)

Source: Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)

Network Types Supported by SNA Tools

Graph type	Networkx	IGraph	Gephi	Pajek
1-Mode network	Yes	Yes	Yes	Yes
2-Mode network Graph	Yes	Yes	Yes	Yes
Multirelational network Graph	No	No	No	Yes
Temporarily network Graph	Yes	No	No	Yes

Naheed Akhtar (2014)

Source: Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)

Graph Layout Supported by SNA Tools

Layout	Networkx	IGraph	Pajek	Gephi
Circular layout	Yes	Yes	Yes	Yes
Random layout	Yes	Yes	Yes	No
Spectral layout	Yes	No	No	No
Spring layout	Yes	Yes	Ye s	Yes
Graphviz layout	Yes	No	No	No
Kamanda kawai	No	Yes	Yes	No
Fruchterman reingold	No	Yes	Yes	No
Force Atlas layout	No	No	Yes	No

Naheed Akhtar (2014)

Source: Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)

Execution Time for SNA Features

SNA Features	Networkx	IGraph	Gephi	Pajek
Load time	54.67 sec.	3.707 sec	29 sec	3 sec
Degree centrality	58.57 sec	6.199 sec	4 sec	2 sec
Graph degree	60.87 sec	6.22 sec	4 sec	2 sec
Page rank	120.78 sec	9.81 sec	10 sec	No
Hits	57.23 sec	15.43	8 sec	No
Cliques	66.98 sec	9.35 sec	Na	No
Density	58.94 sec	3.302 sec	4 sec	No
Modularity	81 .4 sec	9 sec	30 sec	6 sec
Network diameter	35 sec	3.51 sec	120 sec	No
Core	65.84 sec	6.532 sec	No	1 sec
Cohesion	No	8.943 sec	No	No
Clustering coefficient	3303.99 sec	1800 sec	1200 sec	108 sec
Hub	76.57	5.831 sec	3 sec	No
Authority	Array is to big	6.783 sec	3 sec	No

Naheed Akhtar (2014)

Source: Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)

Comparative analysis of Social Networking Analysis tools

Software	Pajek	Gephi	Social Network Visualizer	Netlytic	Graphviz
Version	1.26	0.7 alpha	1.56 Beta	Tier 1,2,3	2.38.0
Type	Stand-alone software	Stand-alone software	Stand-alone software	Stand-alone software	Stand-alone software
Platform	Windows	Java	Windows	Windows	Windows
License	Free	GNU GPL	Free	Tier 1,2 (Free) Tier 3 (CS)	Free
Expectable Computing Time	Fast(C)	Medium(JAVA)	Fast(C)	Medium(JAVA)	Fast(C)
Tractable number of nodes	500000 nodes	150000 nodes	100000 nodes	300000 nodes	1400000 nodes
Time to load 10^5 nodes and 10^6 edges	24 seconds	40 seconds	46 seconds	50 seconds	34 seconds
File formats					
GML	No	Yes	Yes	Yes	No
Pajek(.net)	No	Important Only	No	No	No
GraphML	Export only	Yes	Yes	Yes	No
DL	Yes	Yes	Yes	Yes	No
GEXF	No	Yes	Yes	Yes	No
Graph types					
Two-mode graphs	Yes	No	No	No	Yes
Multi-relational graphs	Yes	No	No	Yes	Yes
Temporality	Yes	No	No	Yes	Yes
Visualization layouts					
FruchtermanReingold	Yes	Yes	Yes	Yes	No
Kamada Kawai	Yes	Yes	No	No	Yes
Other spring layouts	No	Yes	Yes	No	Yes
Indicators					
Degree centrality	Yes	Yes	Yes	Yes	Yes
Betweenness centrality	Yes	Yes	Yes	Yes	Yes
Closeness centrality	Yes	Yes	Yes	Yes	Yes
Dyad census	No	No	No	No	No
Triad census	Yes	No	No	No	No
HITS	No	Yes	Yes	No	No
Page Rank	No	Yes	Yes	Yes	No
Clustering Algorithms					
Edge Betweenness	No	No	No	Yes	No
Walktrap	No	No	No	Yes	No
Spinglass	No	No	No	Yes	No
Dendogram Display	Yes	Yes	Yes	Yes	Yes

Source: Agrawal, H., Thakur, A., Slathia, R., & Sumangali, K. (2015). A Comparative Analysis of Social Networking Analysis Tools. J Inform Tech Softw Eng, 5(157), 2.

Comparative analysis of Social Networking Analysis tools

Software	**Pajek**	**Gephi**
Version	1.26	0.7 alpha
Type	Stand-alone software	Stand-alone software
Platform	Windows	Java
License	Free	GNU GPL
Expectable Computing Time	Fast(C)	Medium(JAVA)
Tractable number of nodes	500000 nodes	150000 nodes
Time to load 10^5 nodes and 10^6 edges	24 seconds	40 seconds
File formats		
GML	No	Yes
Pajek(.net)	No	Important Only
GraphML	Export only	Yes
DL	Yes	Yes
GEXF	No	Yes
Graph types		
Two-mode graphs	Yes	No
Multi-relational graphs	Yes	No
Temporality	Yes	No

Source: Agrawal, H., Thakur, A., Slathia, R., & Sumangali, K. (2015). A Comparative Analysis of Social Networking Analysis Tools. J Inform Tech Softw Eng, 5(157), 2.

Comparative analysis of Social Networking Analysis tools

Software	Pajek	Gephi
Visualization layouts		
FruchtermanReingold	Yes	Yes
Kamada Kawai	Yes	Yes
Other spring layouts	No	Yes
Indicators		
Degree centrality	Yes	Yes
Betweenness centrality	Yes	Yes
Closeness centrality	Yes	Yes
Dyad census	No	No
Triad census	Yes	No
HITS	No	Yes
Page Rank	No	Yes
Clustering Algorithms		
Edge Betweenness	No	No
Walktrap	No	No
Spinglass	No	No
Dendogram Display	Yes	Yes

Source: Agrawal, H., Thakur, A., Slathia, R., & Sumangali, K. (2015). A Comparative Analysis of Social Networking Analysis Tools. J Inform Tech Softw Eng, 5(157), 2.

References

- Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” 2nd Edition, Springer. http://www.cs.uic.edu/~liub/WebMiningBook.html
- Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann. http://analyzingthesocialweb.com/course-materials.shtml
- Devangana Khokhar (2015), Gephi Cookbook, Packt Publishing
- Bastian M., Heymann S., Jacomy M. (2009), “Gephi: an open source software for exploring and manipulating networks”, International AAAI Conference on Weblogs and Social Media.
- Agrawal, H., Thakur, A., Slathia, R., & Sumangali, K. (2015). A Comparative Analysis of Social Networking Analysis Tools. J Inform Tech Softw Eng, 5(157), 2.
- Naheed Akhtar (2014), "Social network analysis tools." In 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)