Gephi

makes graph handy

20171226

放款資訊科 馮書昭

Outline

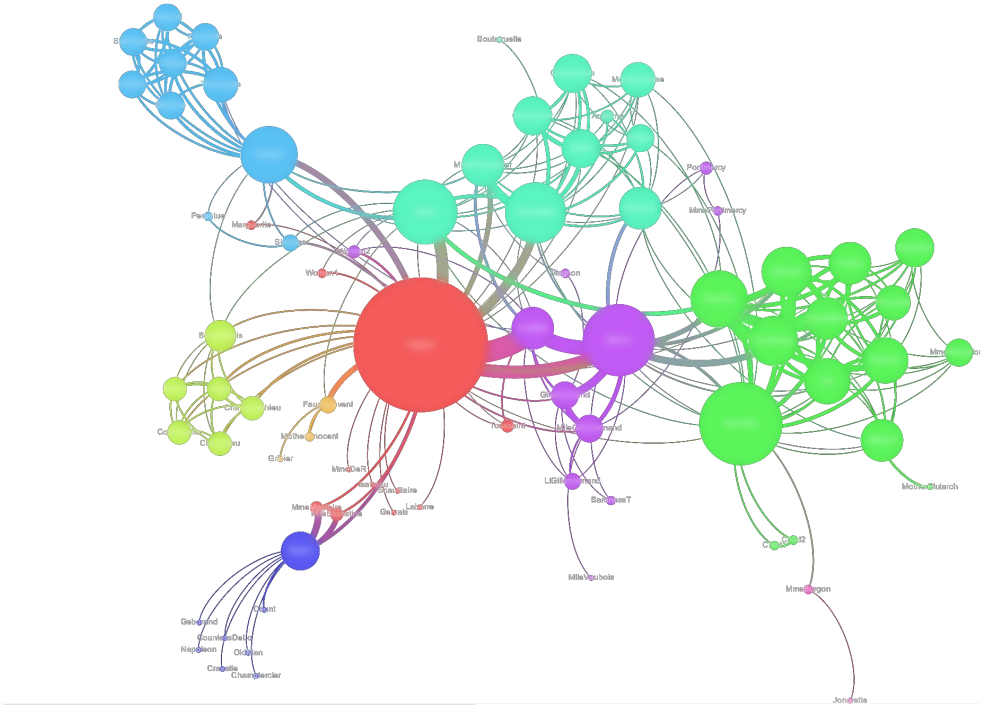
- 下載安裝
- Gephi簡介
- 圖論(Graph)、網絡分析
- Gephi實作

下載安裝

- 主程式
  - 有Windows/Mac OS/ Linux版本
  - 會需要用到Java
    - Mac有內建，Windows/Linux需要下載
  - 補充：
    - 系統需求
    - 簡單錯誤處理

Gephi 簡介

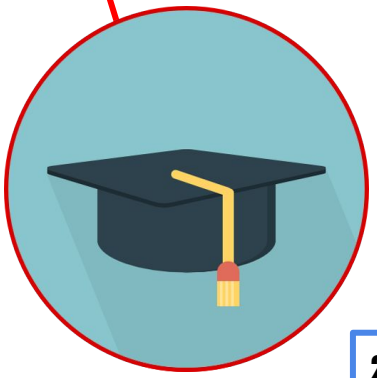
- 一個以Java、OpenGL撰寫的開源軟體
- 目的為網絡分析(Network Analysis)及視覺化(Visulization)



Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: an open source software for exploring and manipulating networks." *Icwsm* 8 (2009): 361-362.

Gephi 極簡史

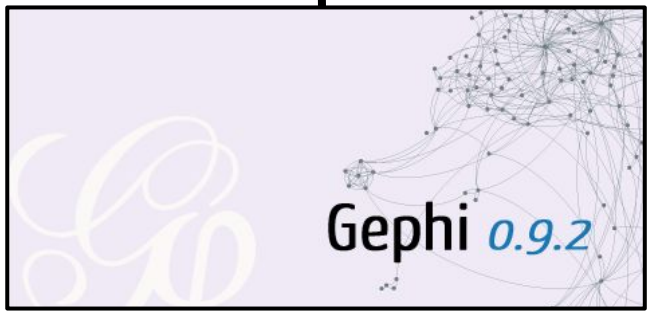
起源於法國康比涅科技大學的學生專案



2008/7/31釋放出第一版 0.6alpha1

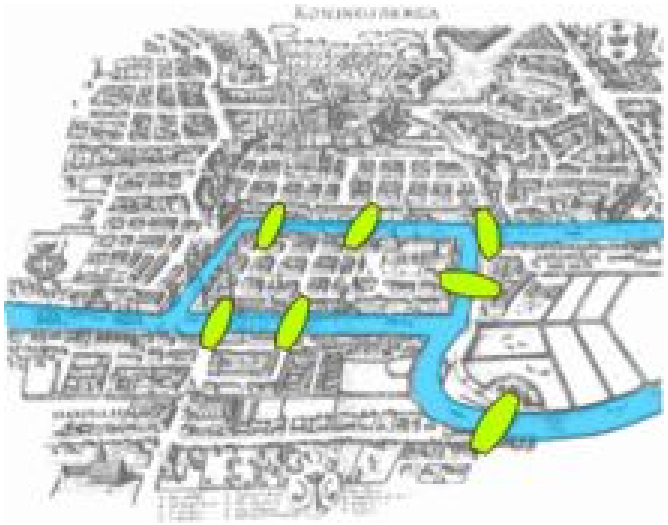
2010，Gephi Consortium成立，意在
支持Gephi的發展
未來版本的釋出

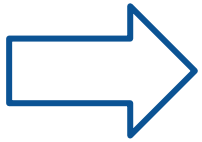
最新版：ver. 0.9.2 2017/9/26

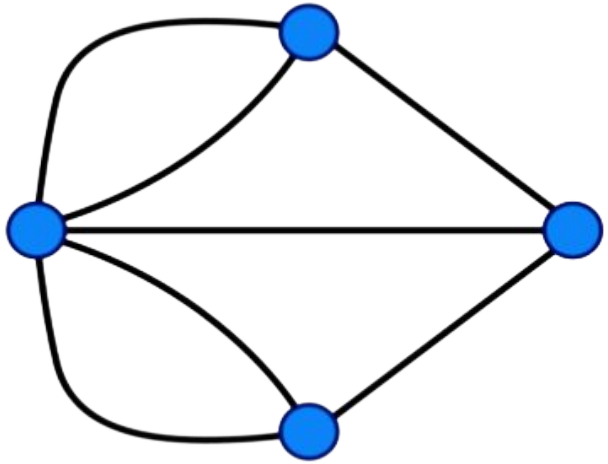


圖論(Graph Theory)

- 圖 (Graph G=(V,E))
  - 節點 (Vertexs = Vertices, V)
  - 邊 (Edges, E)



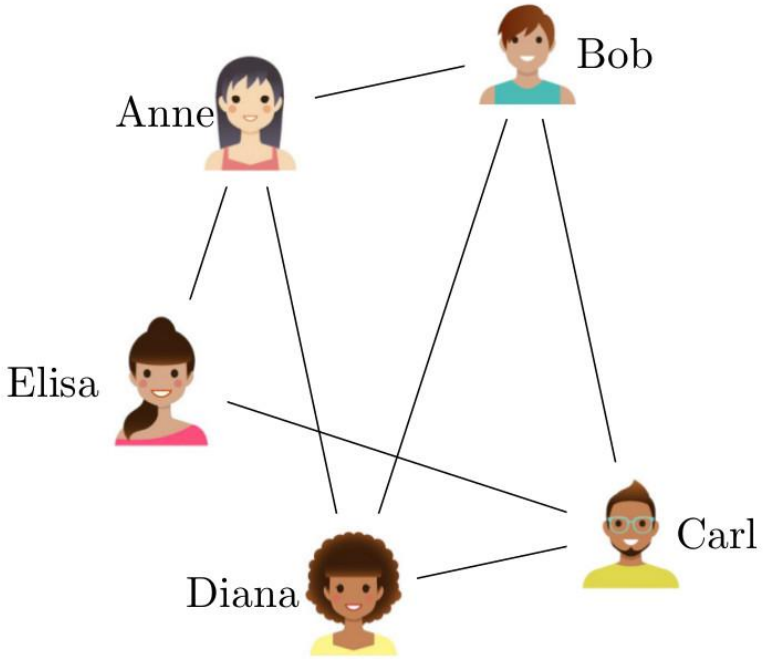




Seven Bridges of Königsberg, Leonhard Euler , 1736

網絡(Networking)

- G = (V, E) → G = (N, L) ; N = nodes, L = links
- Degree : 1 個 node所接到的link數
- Directed vs. Undirected
- Weighted vs. Unweighted
- Local(Ego) vs. Global



網絡特徵

- Global
  - Graph Density → Avg Degree
  - Path
    - Network Diameter
    - Avg. Path Length e.g. 6 degree separation
- Ego
  - Weight
  - Centrality
    - Degree Centrality
    - Closeness Centrality
    - Betweenness Centrality
- Connected Components

資料來源

- 只要確認好要判斷的個體(Nodes)以及之間的關係(Links), 都可以作為資料來源
  - <<悲慘世界>>角色出現關係
    - 個體:角色　關係:出現在同一章
  - Twitter追隨者網絡
    - 個體:使用者　關係:追隨
  - 政治人物粉絲團同溫層
    - 個體:使用者　關係:留言、點讚、分享.....
  - 八點黨的剪不斷裡還斷愛情多角關係
  - .....

Gephi實作

- Gephi支援的輸入格式
  * GEXF
  * GDF
  * GML
  * GraphML
  * Pajek NET
  * GraphViz DOT
  * CSV
  * UCINET DL
  * Tulip TPL
  * Netdraw VNA
  * Spreadsheet

其他輸入參考資料: https://gephi.org/users/supported-graph-formats/

資料集簡介

- 資料集(CSV)
  - Coappearance weighted network of characters in the novel “Les Miserables” from Victor Hugo

Node

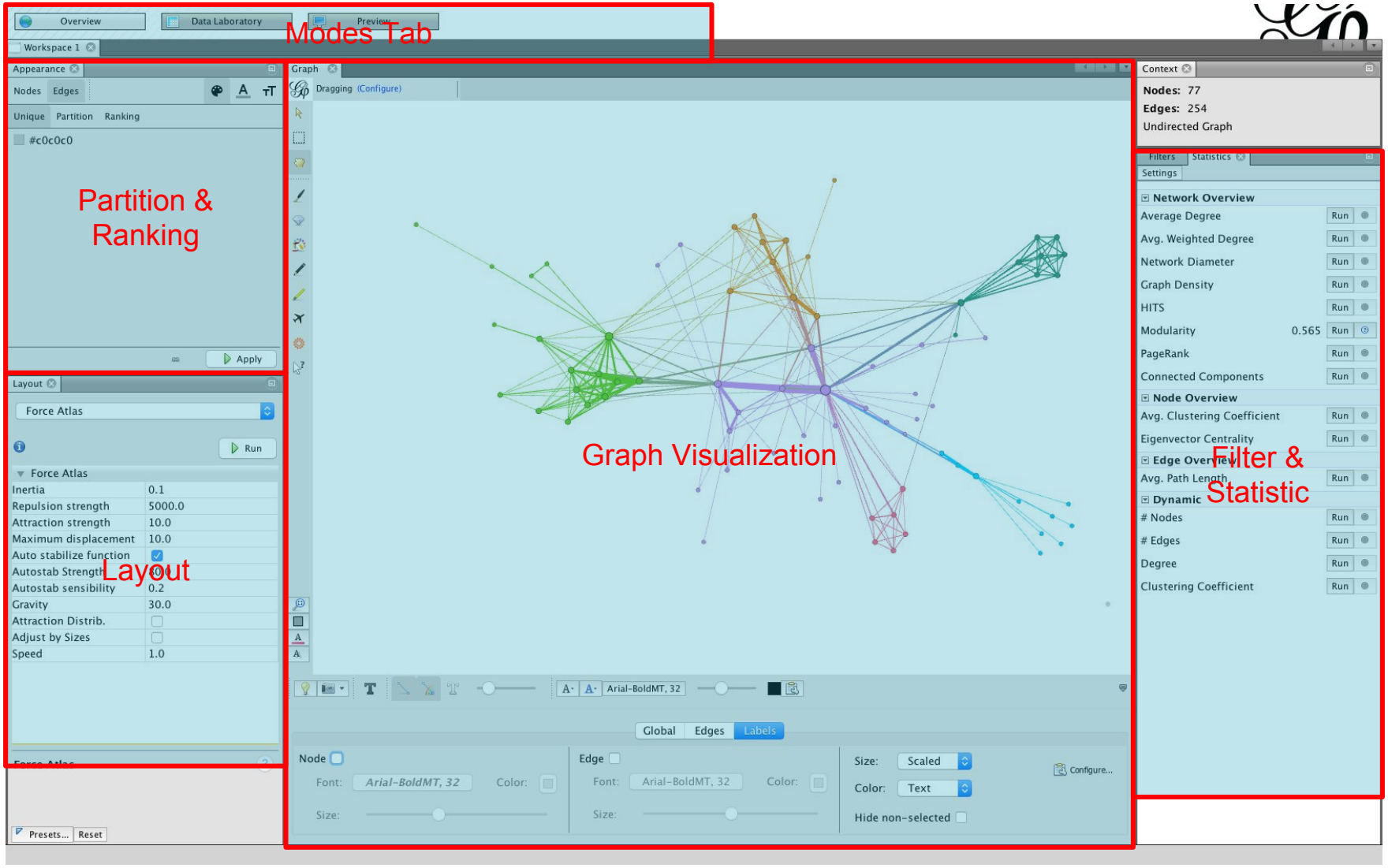
Id	Label
0	Myriel
1	Napoleon
2	MlleBaptistine
3	MmeMagloire
4	CountessDeLo
5	Geborand
6	Champtercier
7	Cravatte
8	Count
9	OldMan
10	Labarre
11	Valjean
12	Marguerite
13	MmeDeR
14	Isabeau

Edge

Source	Target	Type	Id	Label	Weight
1	0	Undirected	0	NA	1
2	0	Undirected	1	NA	8
3	0	Undirected	2	NA	10
3	2	Undirected	3	NA	6
4	0	Undirected	4	NA	1
5	0	Undirected	5	NA	1
6	0	Undirected	6	NA	1
7	0	Undirected	7	NA	1
8	0	Undirected	8	NA	2
9	0	Undirected	9	NA	1
11	0	Undirected	13	NA	5
11	2	Undirected	12	NA	3
11	3	Undirected	11	NA	3
11	10	Undirected	10	NA	1
12	11	Undirected	14	NA	1

D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley, Reading, MA (1993).

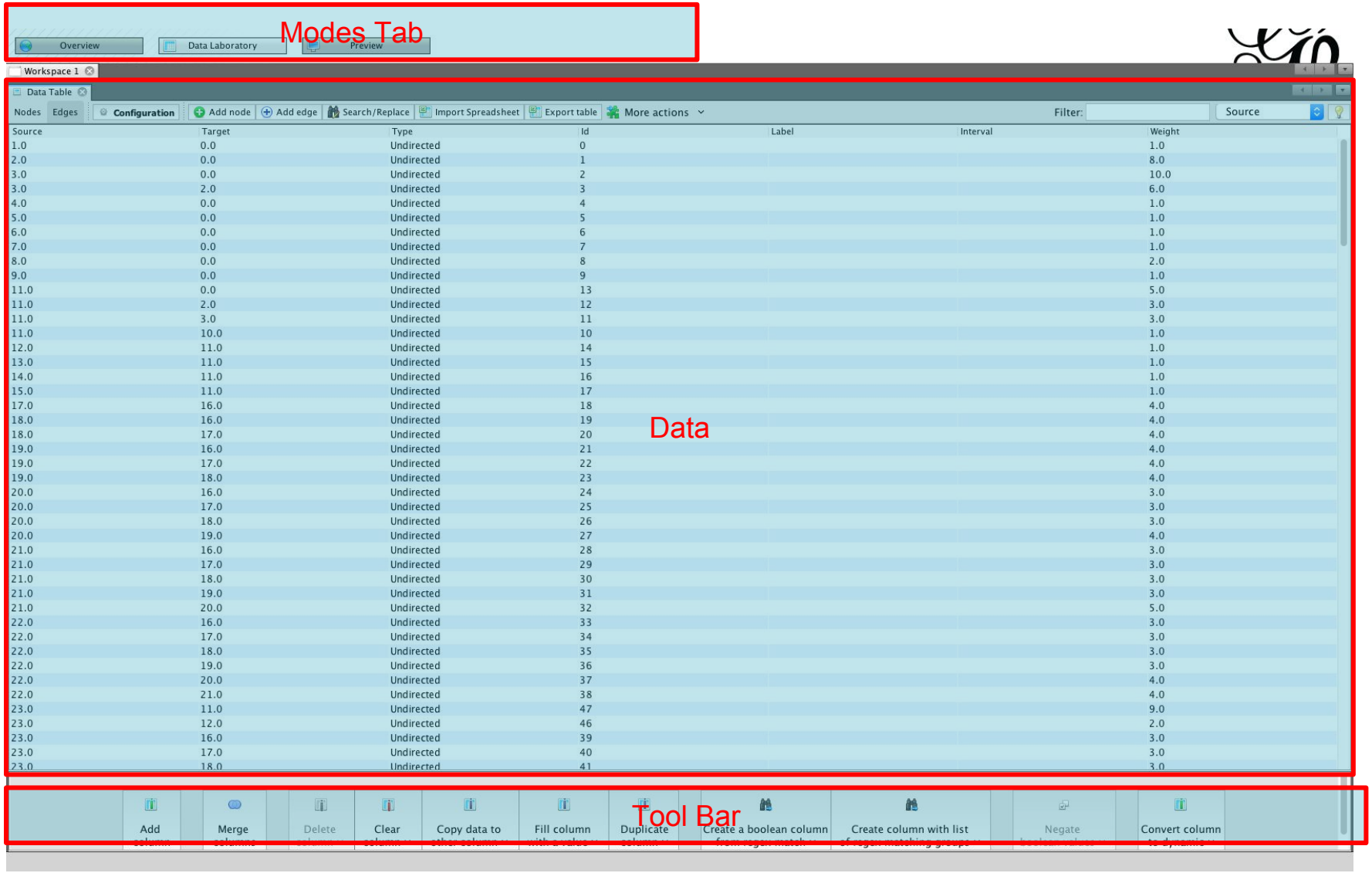
介面介紹-Overview



Overview - 最主要的視覺化處理

- Mode Tab
  - 切換三個mode的按鈕
- Graph Visulization
  - 目前圖片視覺化後會具有的屬性
- Layout
  - 控制節點分佈的演算法(key word : Force-directed graph drawing)
- Partition & Ranking
  - 依照節點特性改變節點大小或上色
- Filter & Statistic
  - 計算或篩選節點
    - 會產生 Metric

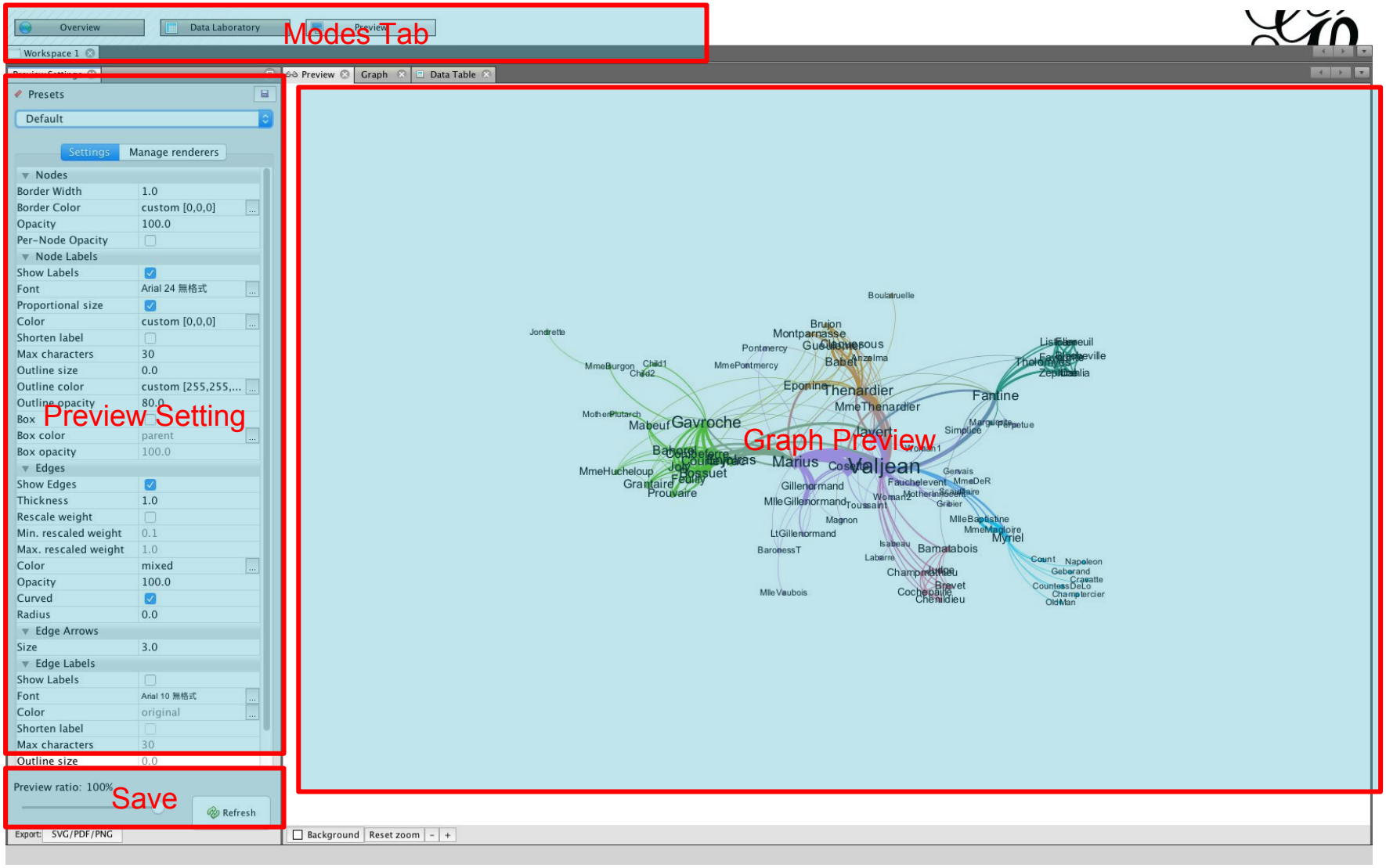
介面介紹-Data Laboratory



Data Laboratory - 資料管理

- Mode Tab
  - 切換三個mode的按鈕
- Data
  - 目前置入的資料，會隨著Filter & Statistic做改變
- Tool Bar
  - 做一些簡單的資料處理(建議在Gephi外處理)

介面介紹-Preview

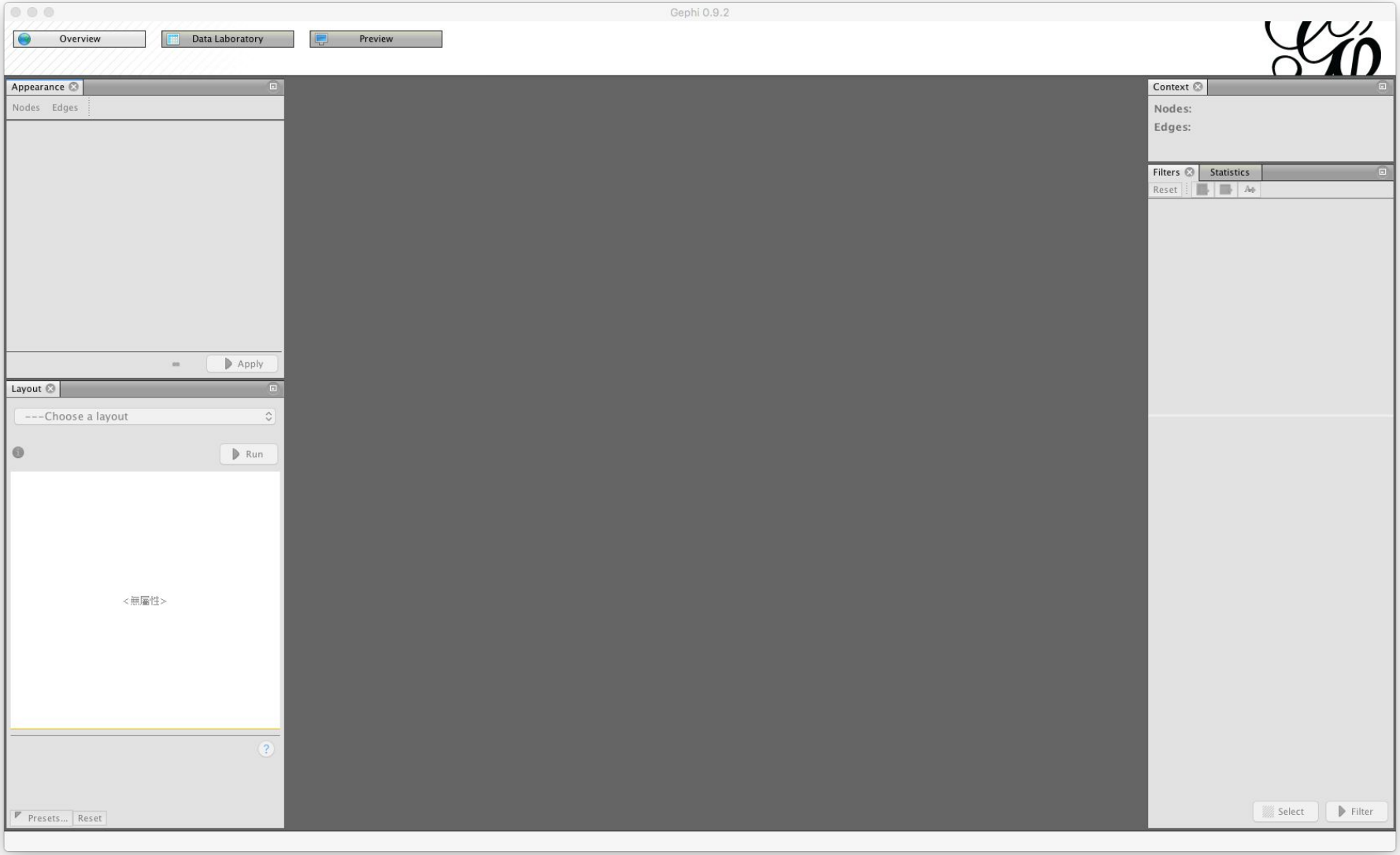


Preview - 輸出前的微調

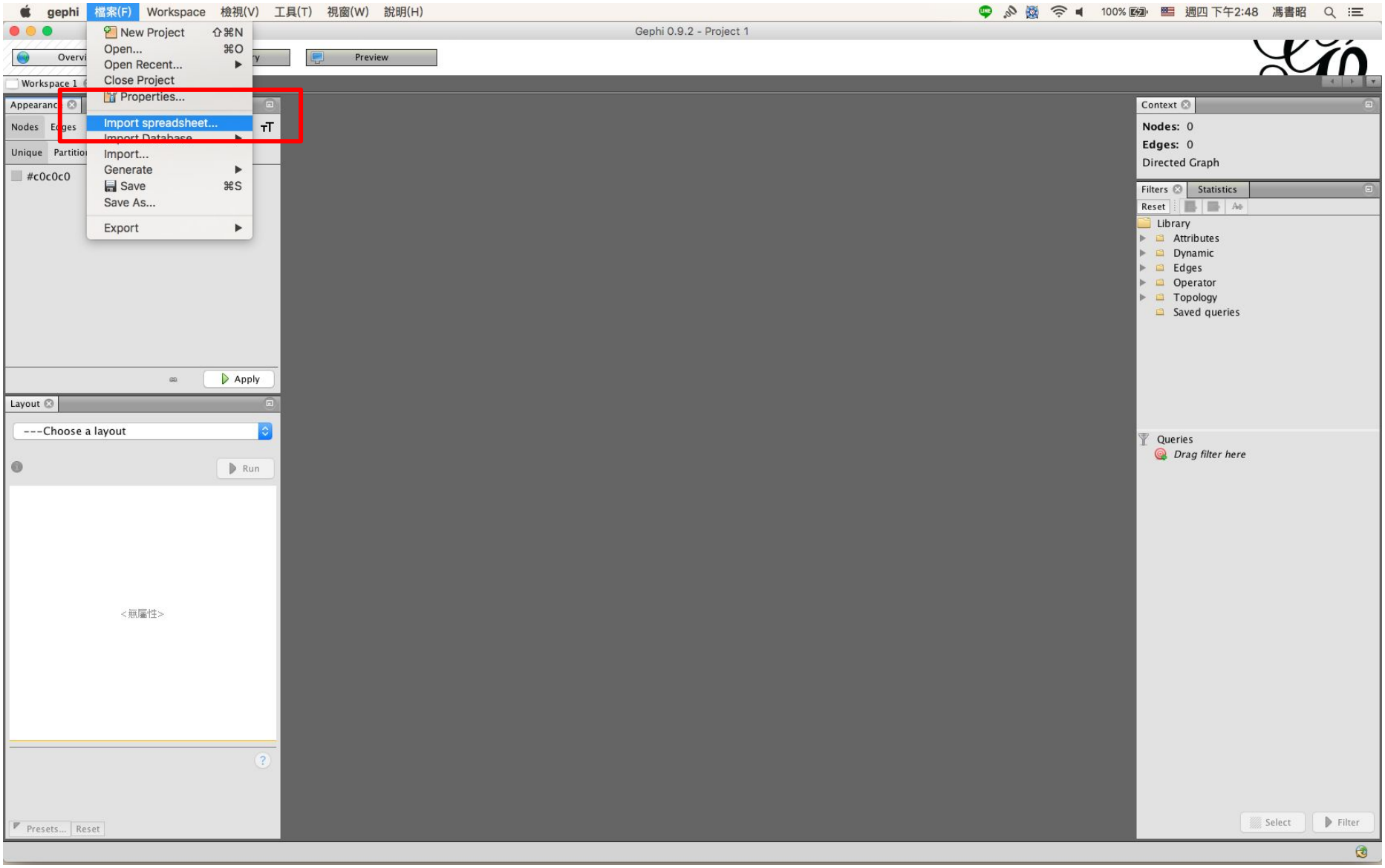
- Mode Tab
  - 切換三個mode的按鈕
- Preview Setting
  - 輸出微調
- Graph Preview
  - 預計輸出
- Save
  - 輸出影像

實作截圖

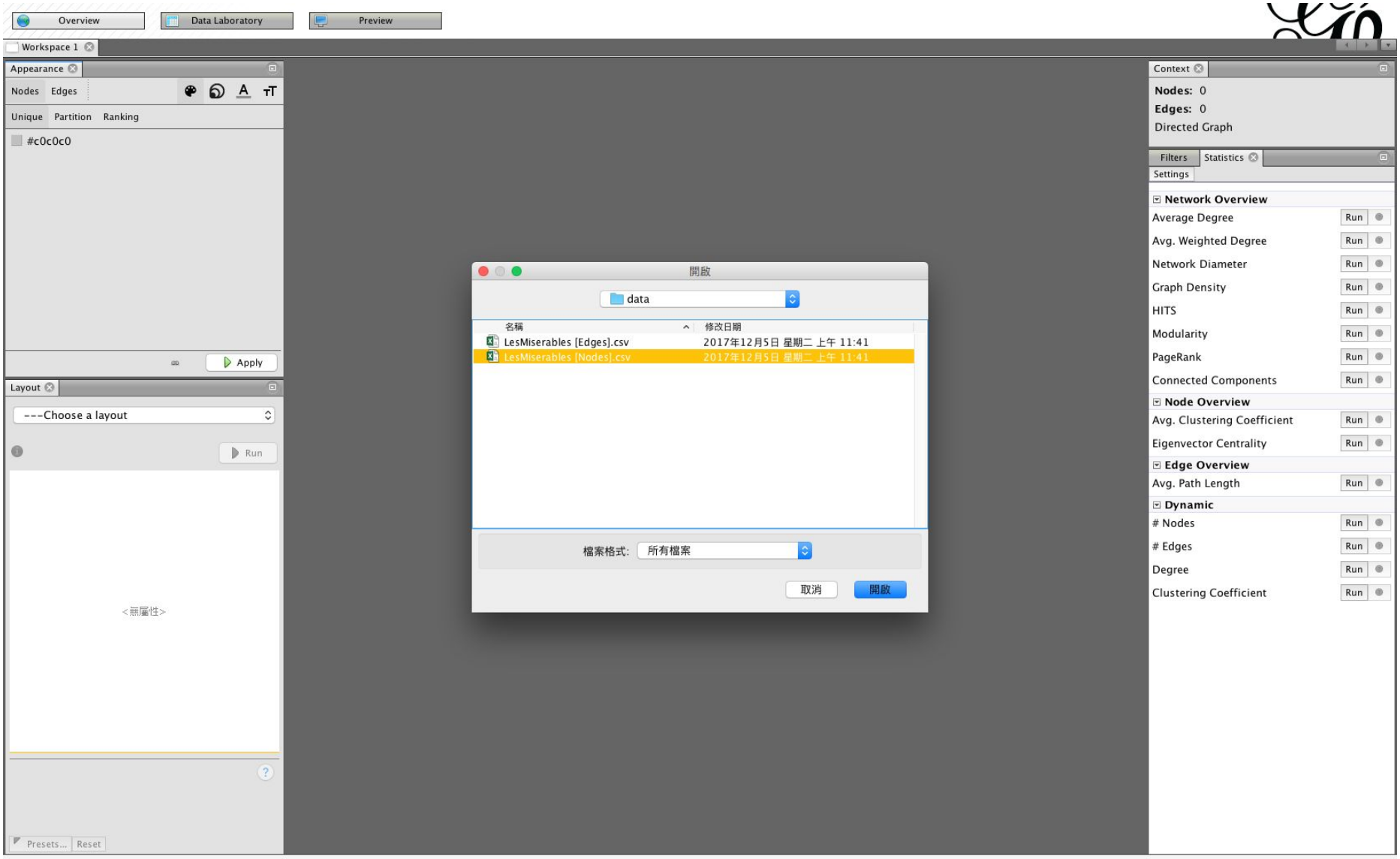
1. 開啟Gephi


Gephi 0.9.2
Overview
Data Laboratory
Preview
Appearance
Nodes
Edges
Apply
Layout
---Choose a layout
Run
<無屬性>
Presets...
Reset
Context
Nodes:
Edges:
Filters
Statistics
Reset
Select
Filter

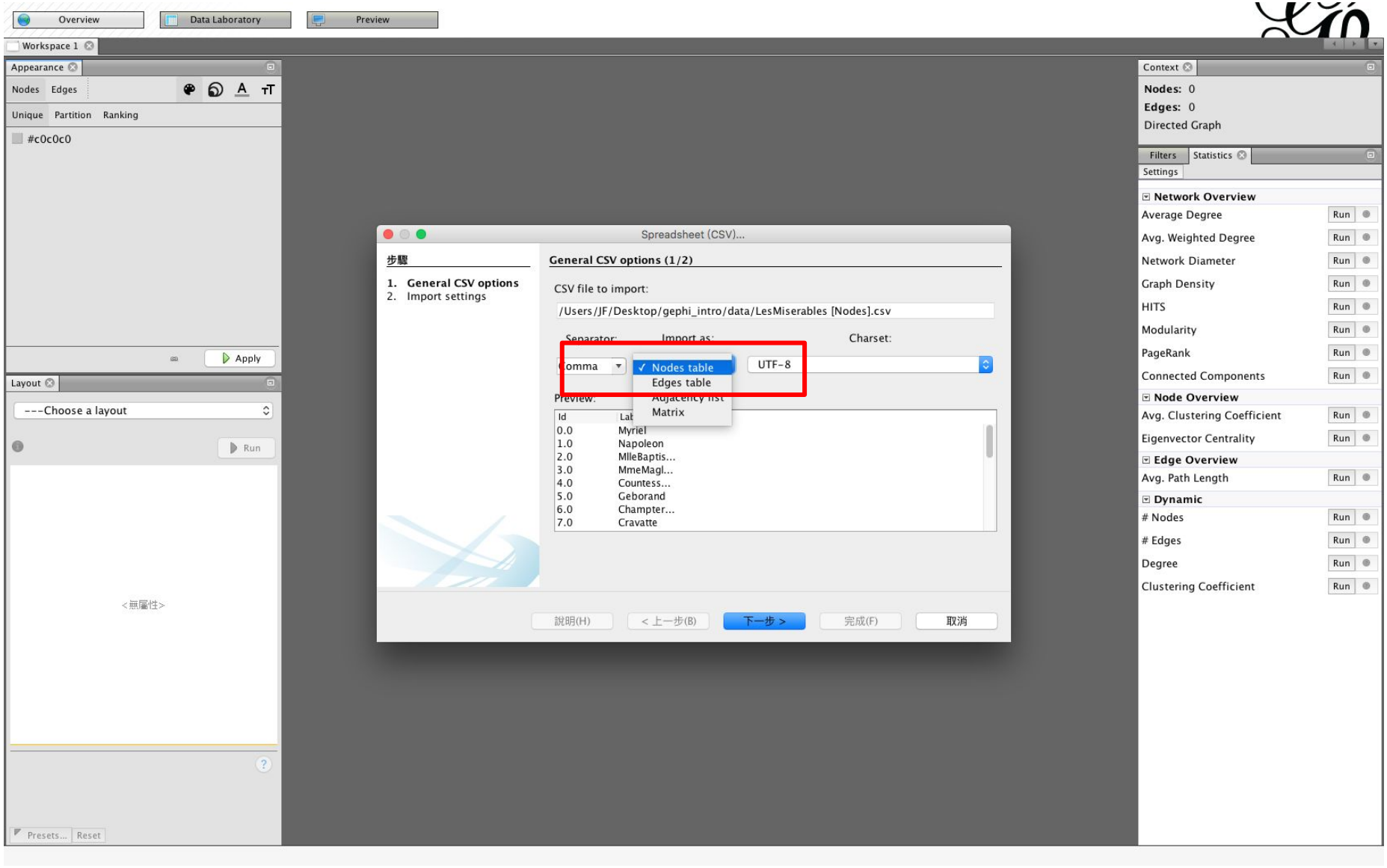
2. 開始匯入檔案


gephi
檔案(F)
Workspace
檢視(V)
工具(T)
視窗(W)
說明(H)
New Project ⇧⌘N
Open... ⌘O
Open Recent...
Close Project
Properties...
Import spreadsheet...
Import Database...
Import...
Generate
Save ⌘S
Save As...
Export
Gephi 0.9.2 - Project 1
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
#c0c0c0
Apply
Layout
---Choose a layout
Run
<無屬性>
Presets...
Reset
Context
Nodes: 0
Edges: 0
Directed Graph
Filters
Statistics
Reset
Library
Attributes
Dynamic
Edges
Operator
Topology
Saved queries
Queries
Drag filter here
Select
Filter

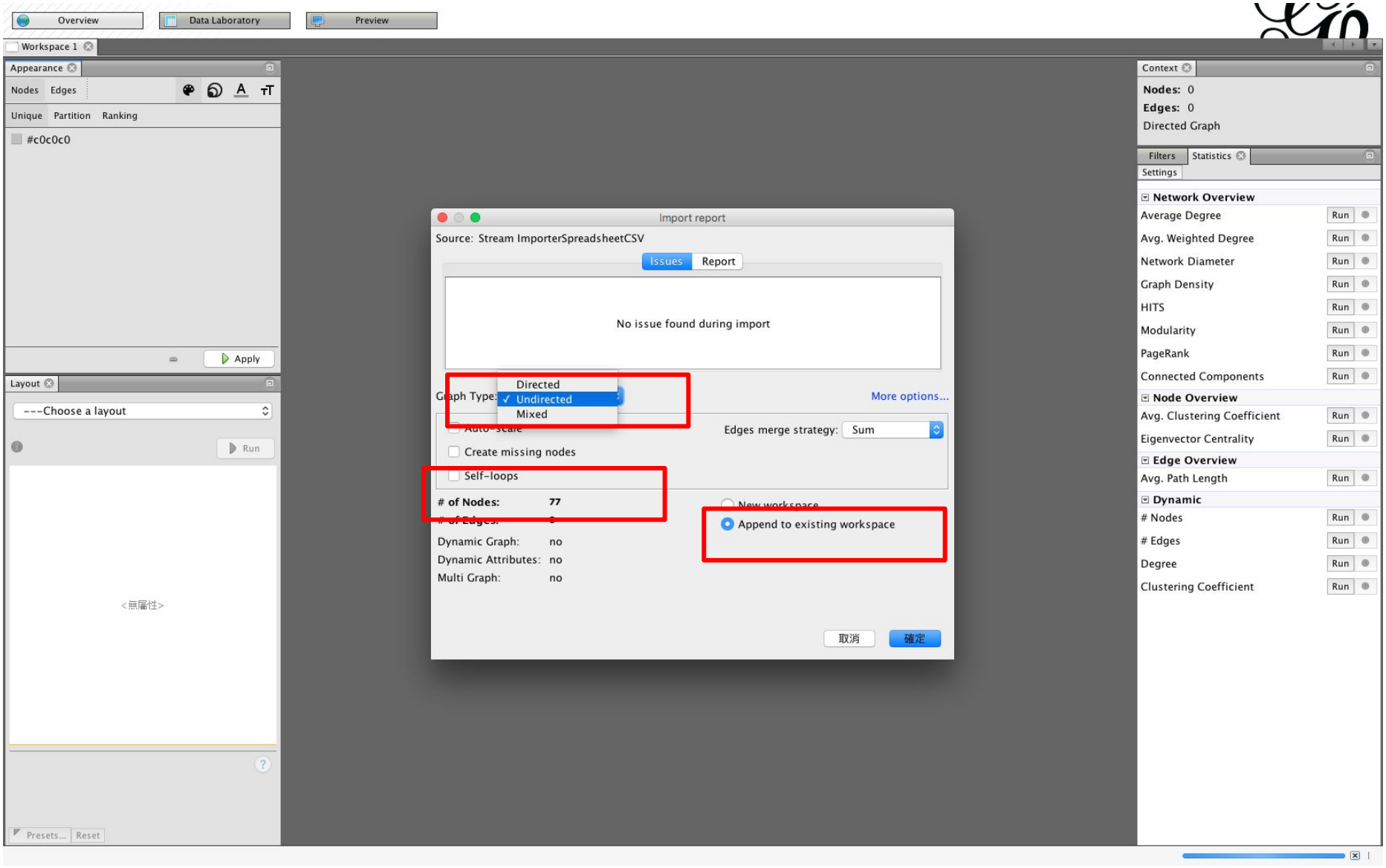
3. 先匯入Node資料


Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
Partition
Ranking
#c0c0c0
Apply
Layout
---Choose a layout
Run
<無屬性>
Presets...
Reset
開啟
data
名稱
修改日期
LesMiserables [Edges].csv
2017年12月5日 星期二 上午 11:41
LesMiserables [Nodes].csv
2017年12月5日 星期二 上午 11:41
檔案格式:
所有檔案
取消
開啟
Context
Nodes: 0
Edges: 0
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree
Avg. Weighted Degree
Network Diameter
Graph Density
HITS
Modularity
PageRank
Connected Components
Node Overview
Avg. Clustering Coefficient
Eigenvector Centrality
Edge Overview
Avg. Path Length
Dynamic
Nodes
Edges
Degree
Clustering Coefficient
Run

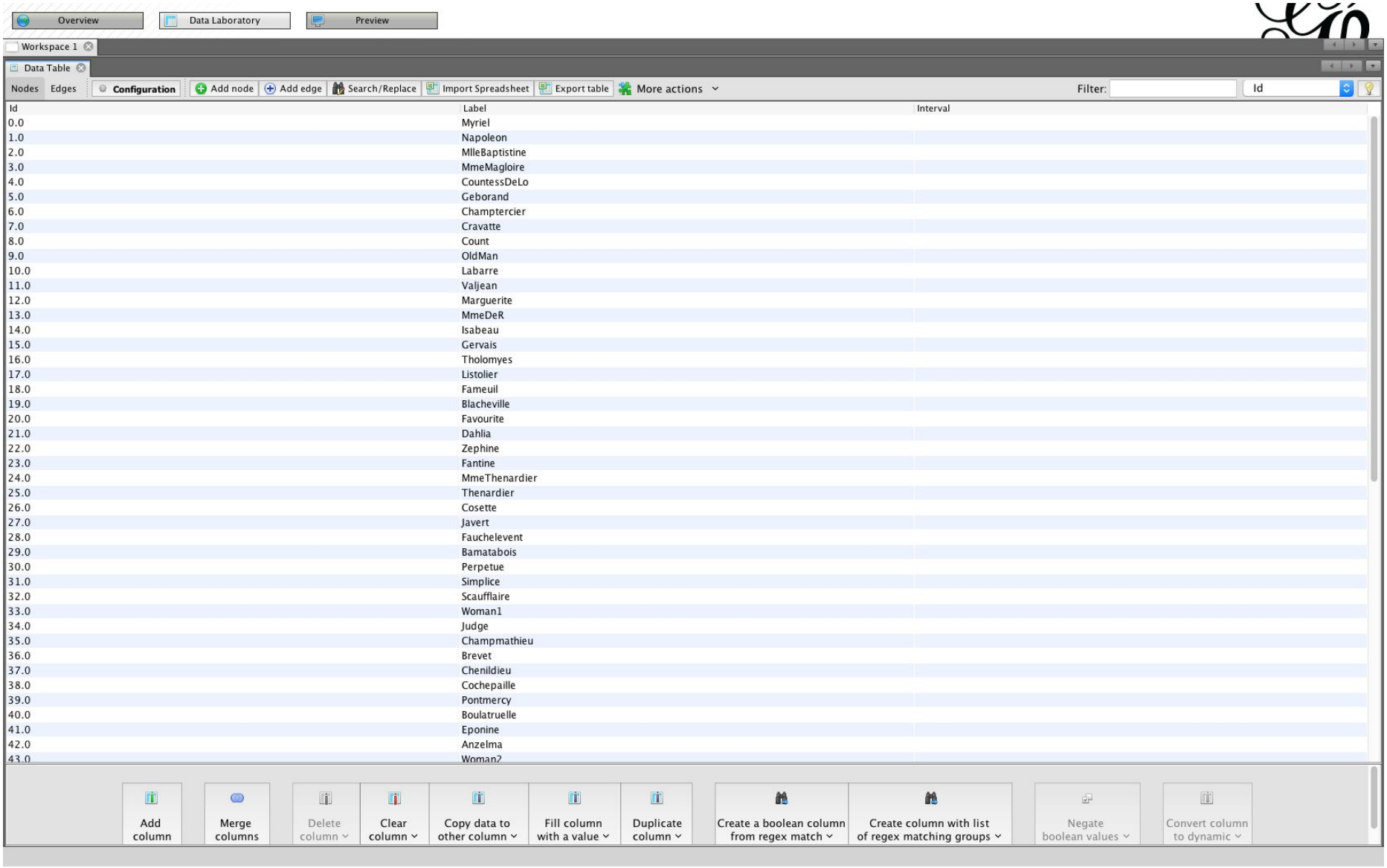
4. 類型請選擇Node


Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
Partition
Ranking
#c0c0c0
Apply
Layout
---Choose a layout
Run
<無屬性>
Presets...
Reset
Spreadsheet (CSV)...
步驟
1. General CSV options
2. Import settings
General CSV options (1/2)
CSV file to import:
/Users/JF/Desktop/gephi_intro/data/LesMiserables [Nodes].csv
Separator:
Import as:
Charset:
Comma
Nodes table
Edges table
Adjacency list
Matrix
UTF-8
Preview:
Id
0.0 Myriel
1.0 Napoleon
2.0 MlleBaptis...
3.0 MmeMagl...
4.0 Countess...
5.0 Geborand
6.0 Champter...
7.0 Cravatte
說明(H)
< 上一步(B)
下一步 >
完成(F)
取消
Context
Nodes: 0
Edges: 0
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree
Avg. Weighted Degree
Network Diameter
Graph Density
HITS
Modularity
PageRank
Connected Components
Node Overview
Avg. Clustering Coefficient
Eigenvector Centrality
Edge Overview
Avg. Path Length
Dynamic
Nodes
Edges
Degree
Clustering Coefficient
Run

5. 可再度確認節點數，這裡不新增新的工作表

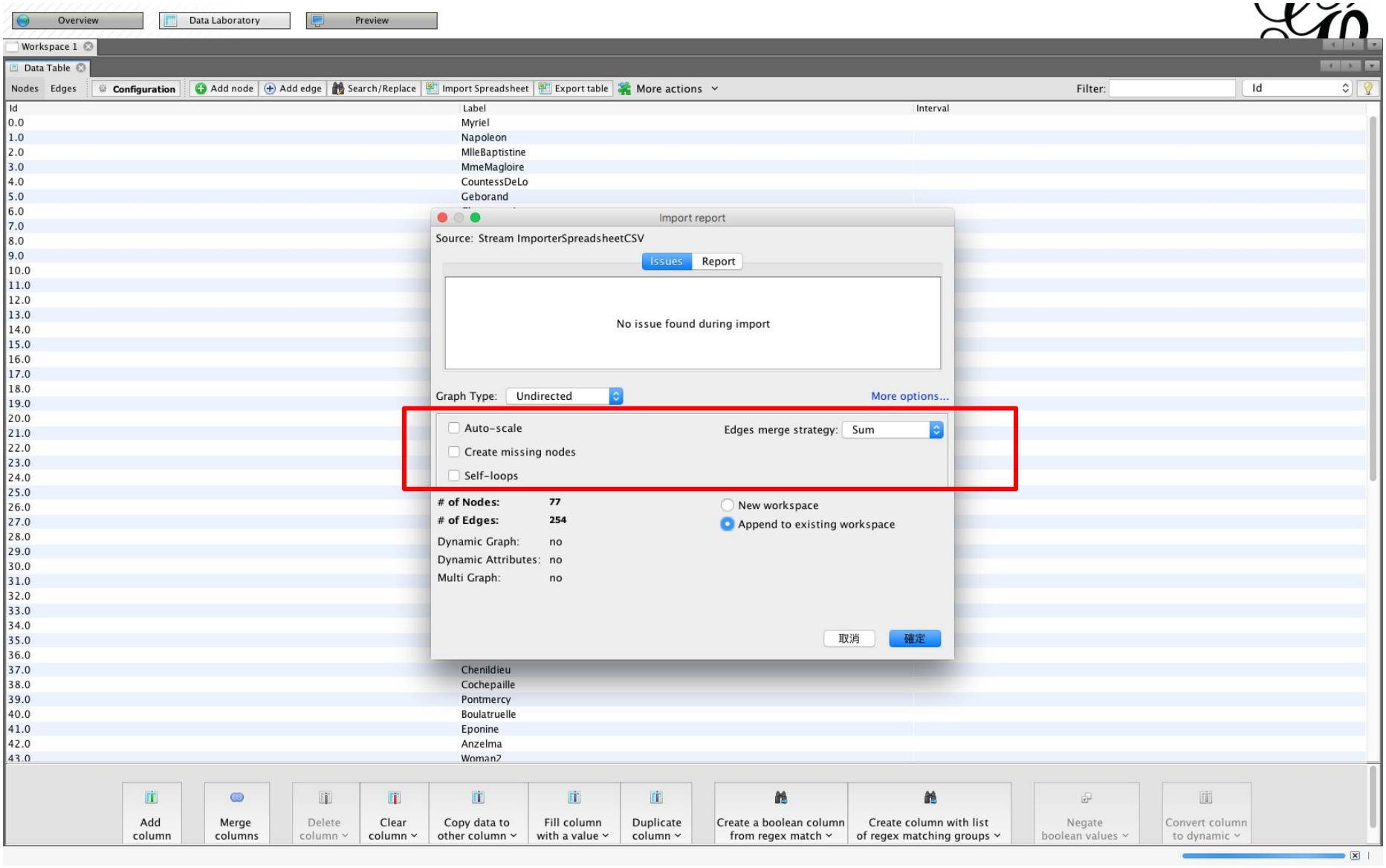


6. 匯入成功

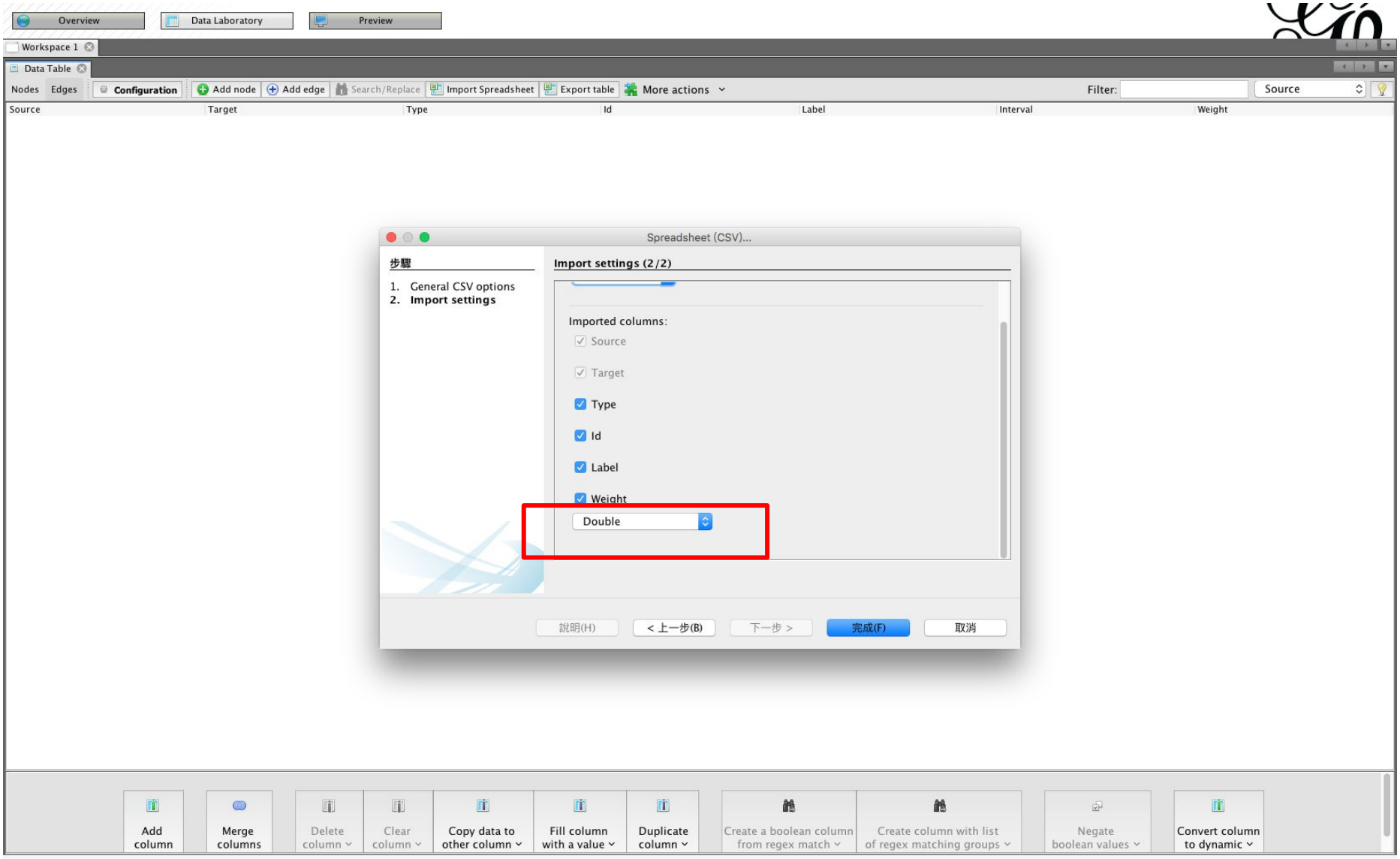


Id	Label	Interval
0.0	Myriel	
1.0	Napoleon	
2.0	MlleBaptistine	
3.0	MmeMagloire	
4.0	CountessDeLo	
5.0	Geborand	
6.0	Champtercier	
7.0	Cravatte	
8.0	Count	
9.0	OldMan	
10.0	Labarre	
11.0	Valjean	
12.0	Marguerite	
13.0	MmeDeR	
14.0	Isabeau	
15.0	Gervais	
16.0	Tholomyes	
17.0	Listolier	
18.0	Fameuil	
19.0	Blacheville	
20.0	Favourite	
21.0	Dahlia	
22.0	Zephine	
23.0	Fantine	
24.0	MmeThenardier	
25.0	Thenardier	
26.0	Cosette	
27.0	Javert	
28.0	Fauchelevent	
29.0	Bamatabois	
30.0	Perpetue	
31.0	Simplice	
32.0	Scaufflaire	
33.0	Woman1	
34.0	Judge	
35.0	Champmathieu	
36.0	Brevet	
37.0	Chenildieu	
38.0	Cochepaille	
39.0	Pontmercy	
40.0	Boulatruelle	
41.0	Eponine	
42.0	Anzelma	
43.0	Woman2	

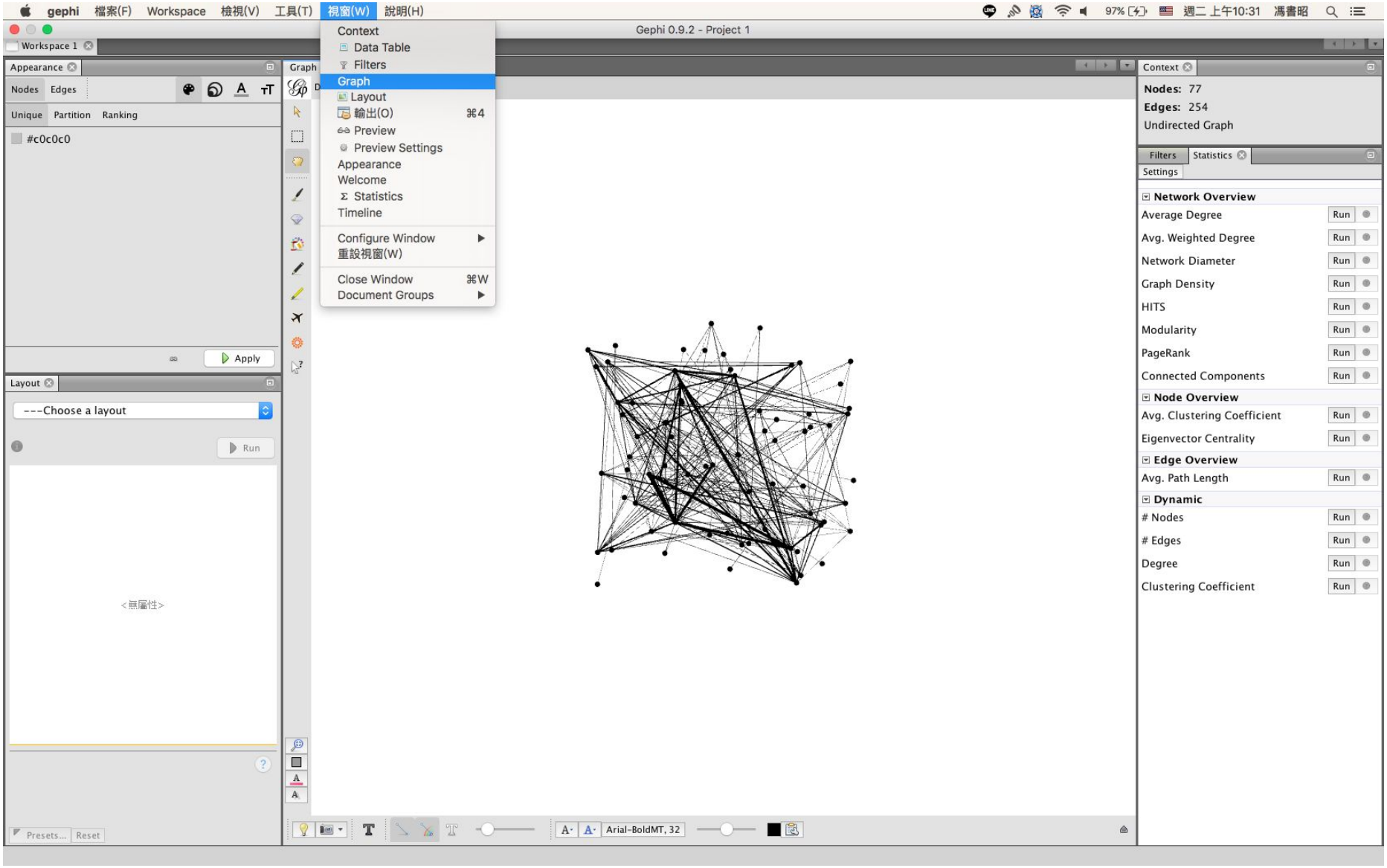
7. 同樣步驟匯入Edges, 這裡有進階選項



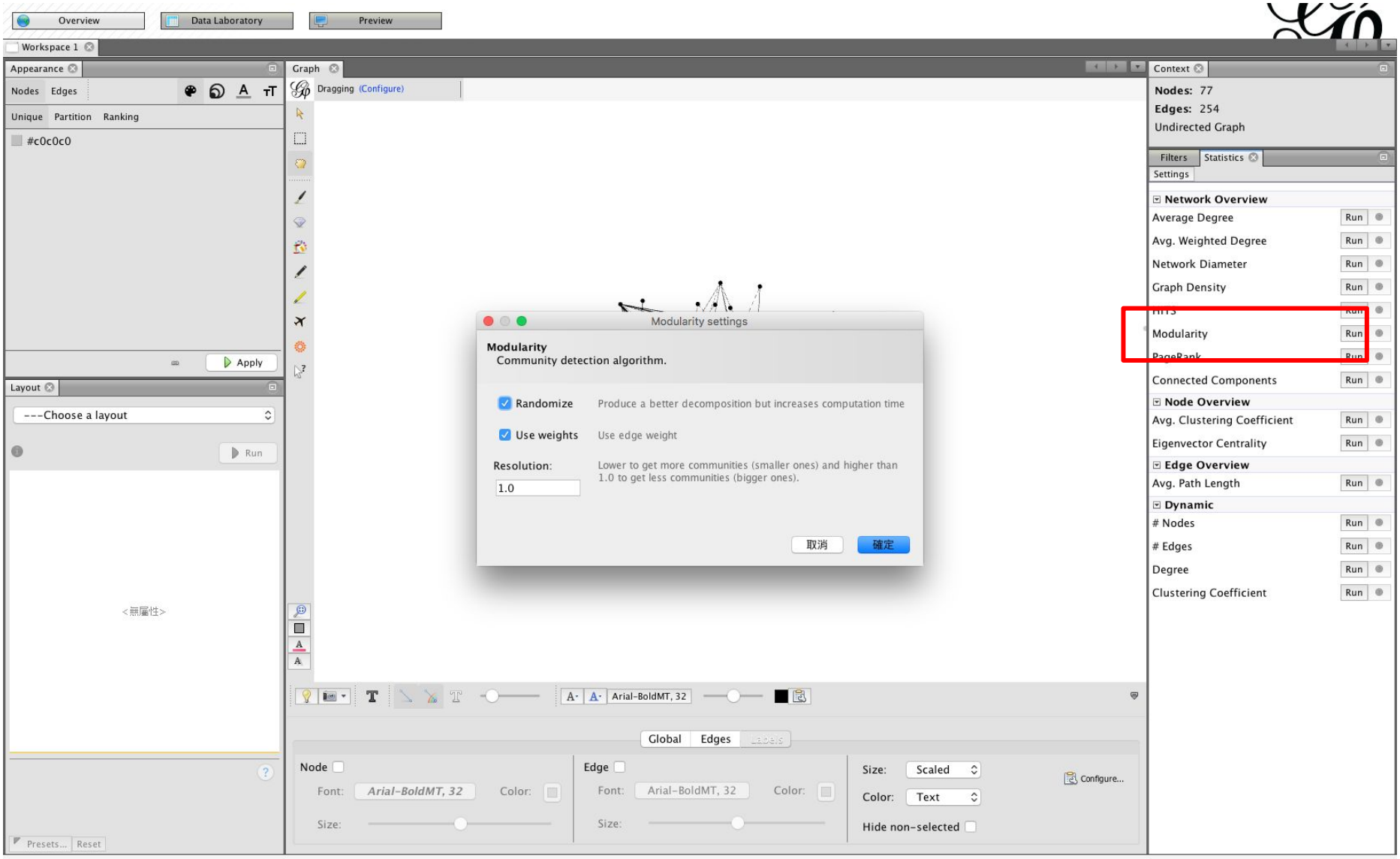
8. 可以在第二頁把錯誤的類型(e.g. 文字->數字)改回來



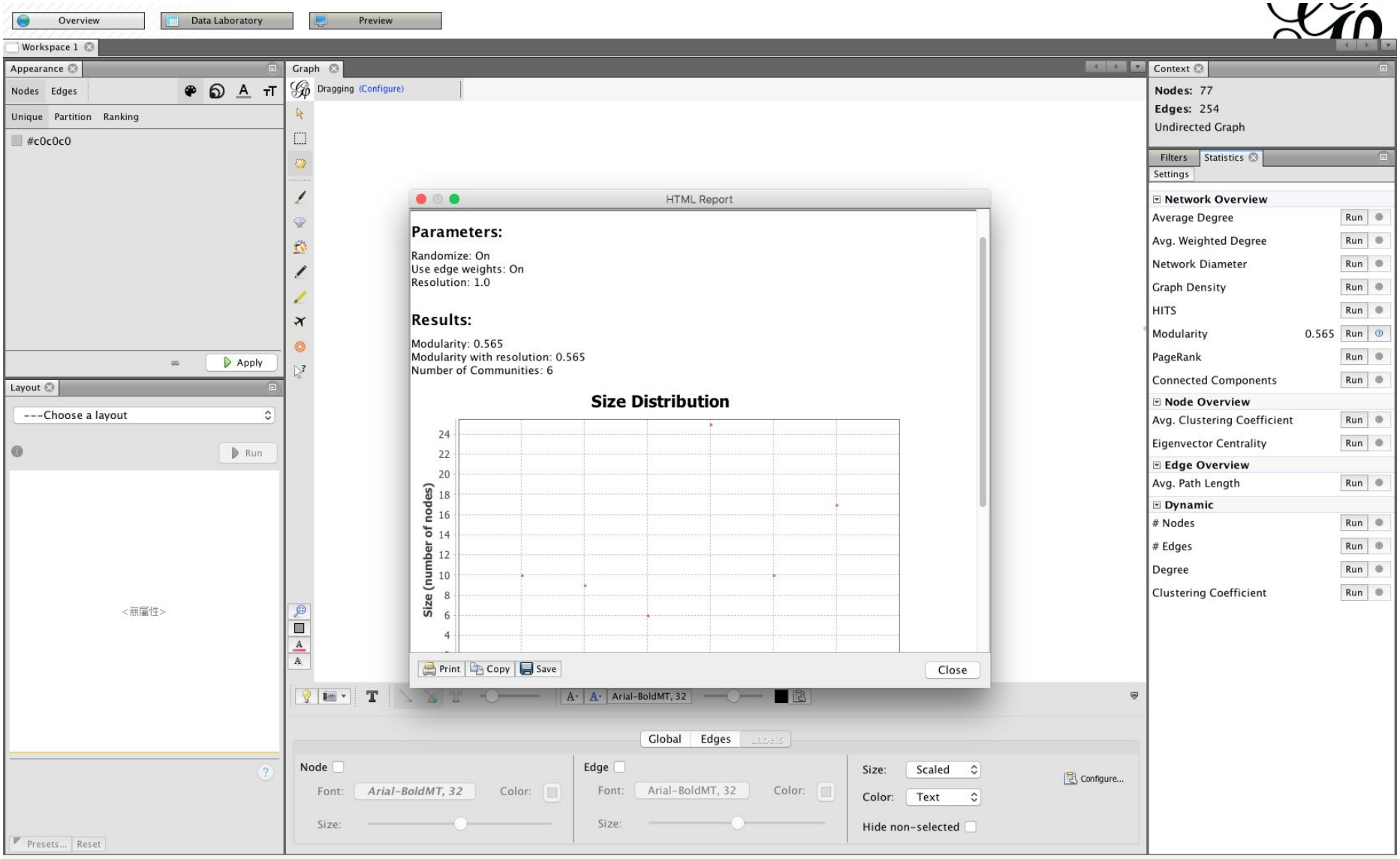
9. 都匯好後把圖表打開，發現目前點是隨機散佈的



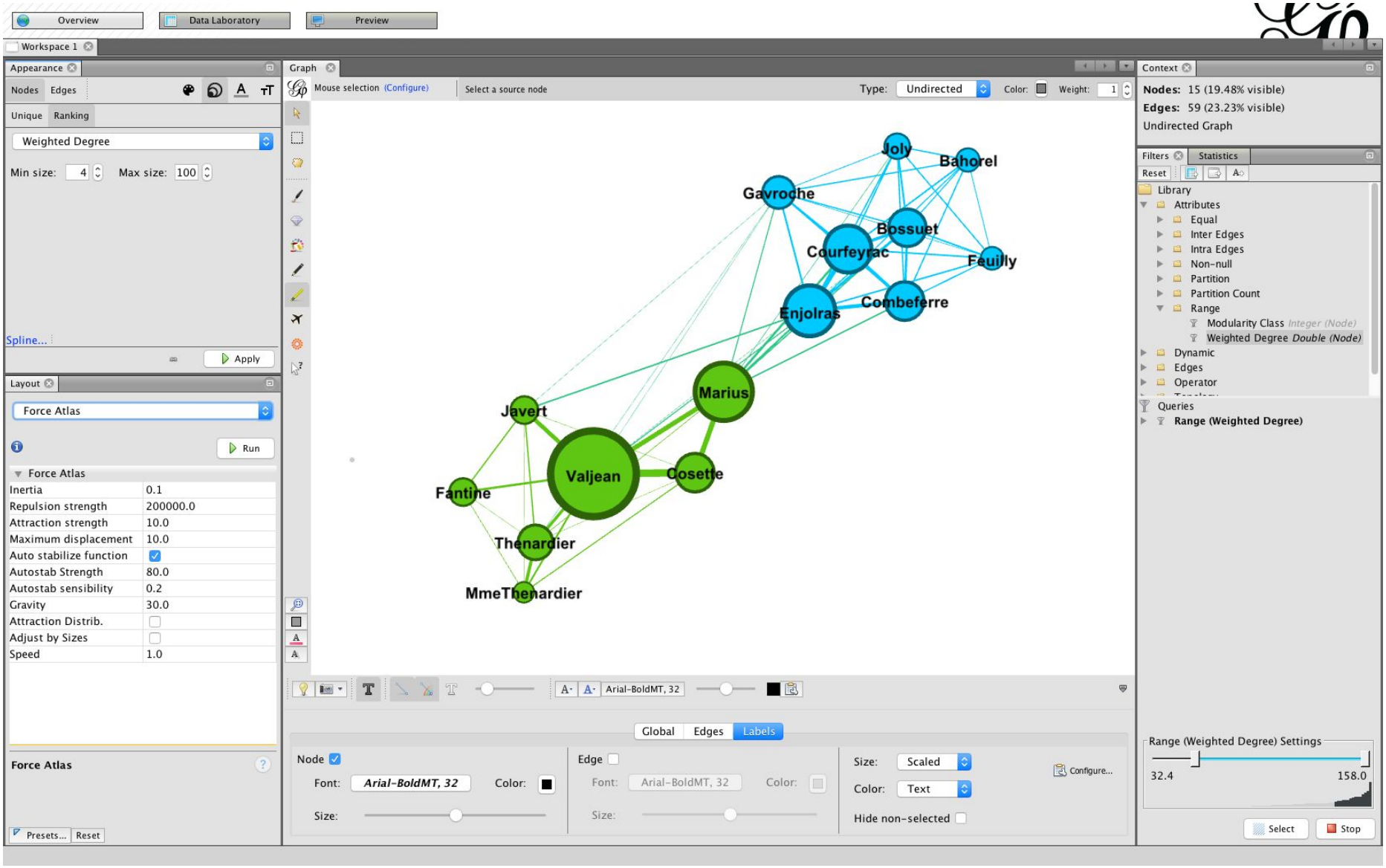
10. 可以藉由先計算Modularity來找出分群



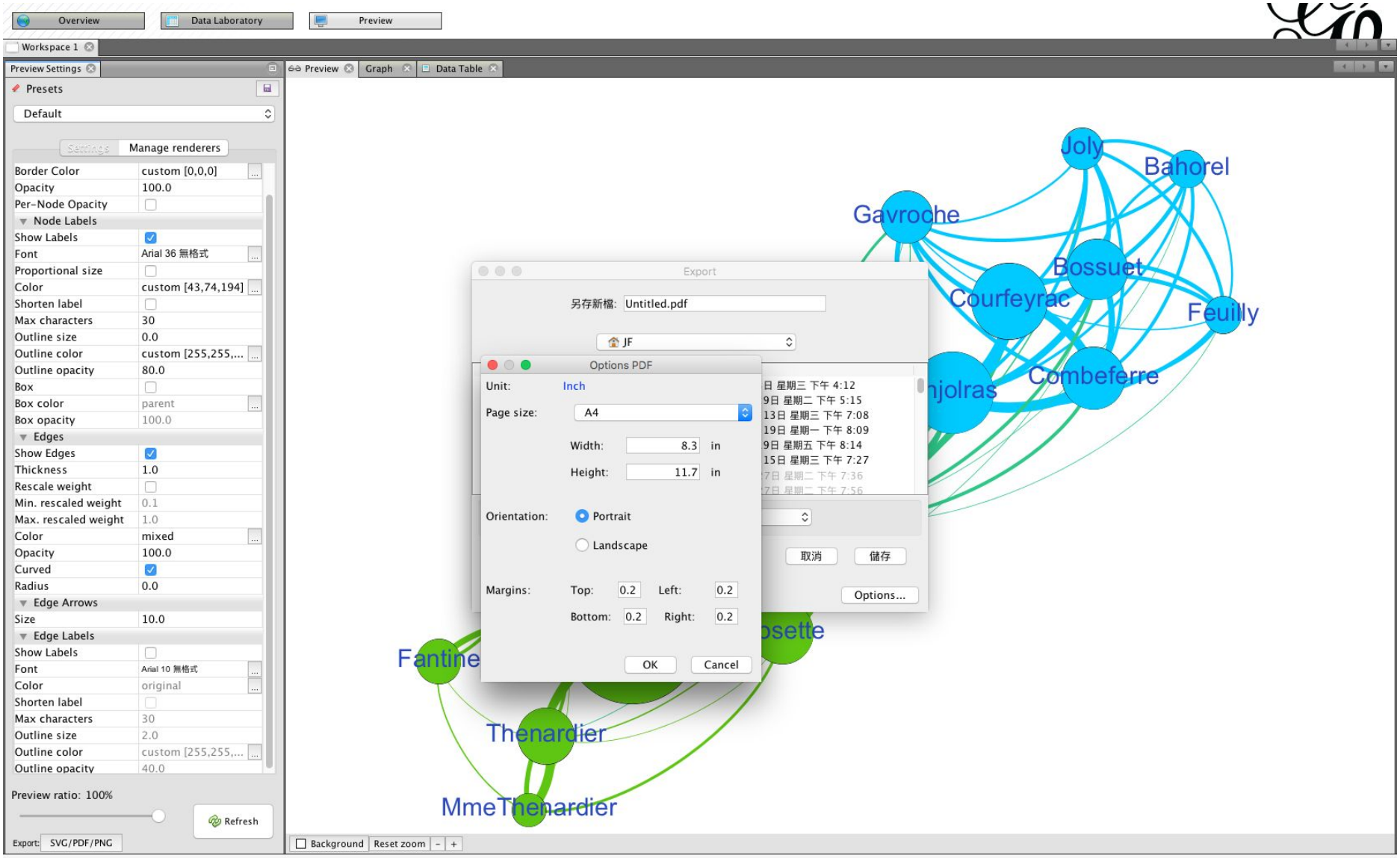
11. 會跳出統計的Metric



11. (自由發揮) 計算權重，找出最常跟別人一起出現的前15個節點，分群後依照權重、分群改變節點大小以及顏色，並用Force Atlas改變節點位置

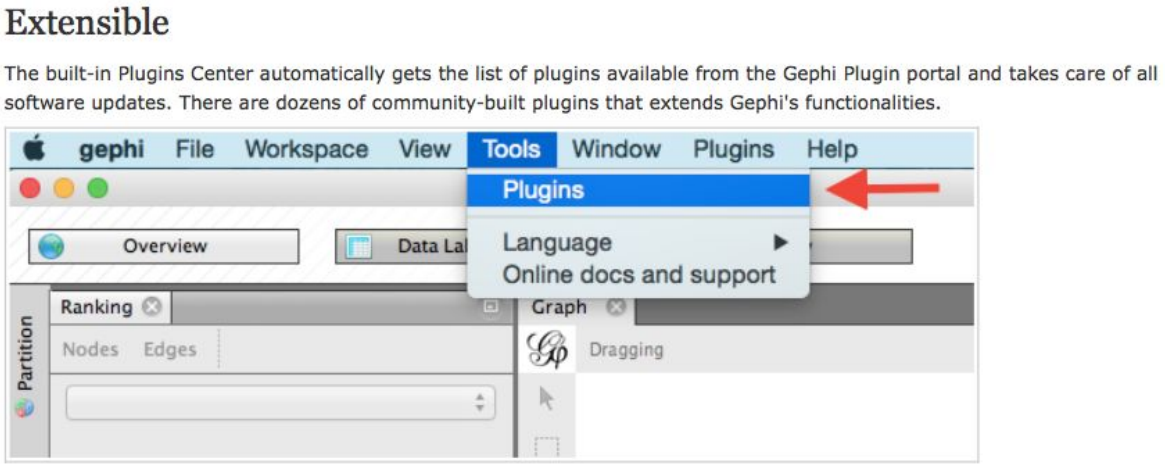


12. 輸出(建議修改輸出Option，因為Label有時候會被切掉)



補充資料

- 官網教學、引用等其他資訊
- 社群
  - fb社團
  - 論壇
- 資源
  - 內建



  - 插件(Plugins)

補充資料2

- Gephi Developer
  - Github下載點
  - Java & Netbeans(Java IDE) required
  - Splited modules communicate with API

感謝聆聽