



Gephi

makes graph handy

20171226

放款資訊科 馮書昭

Outline

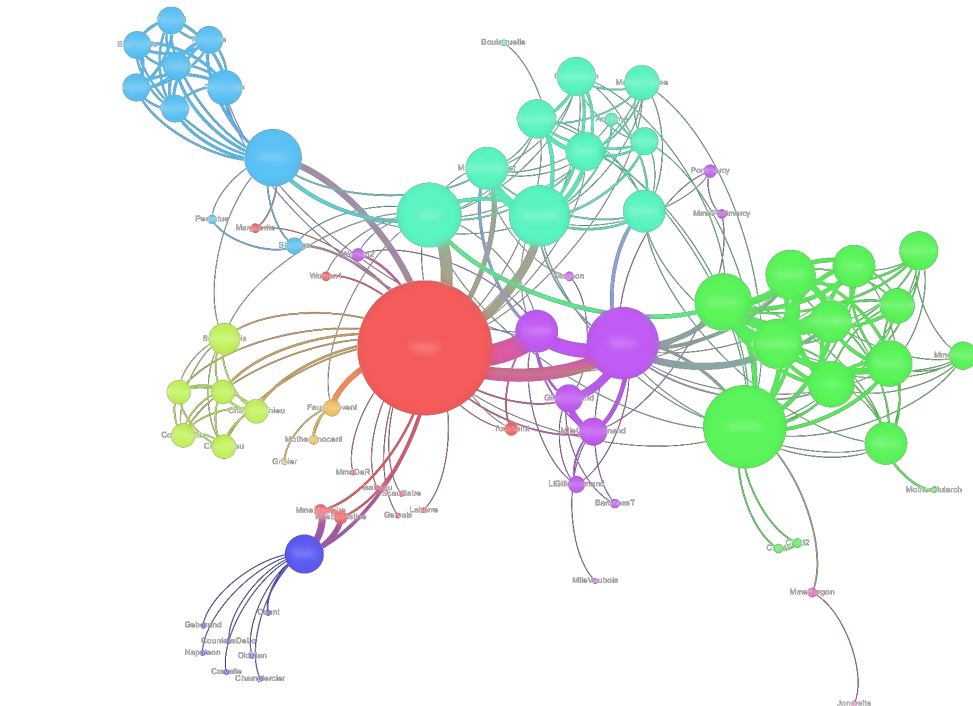
- 下載安裝
- Gephi簡介
- 圖論(Graph)、網絡分析
- Gephi實作

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(Visualization)



Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: an open source software for exploring and manipulating networks." *lcsm* 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

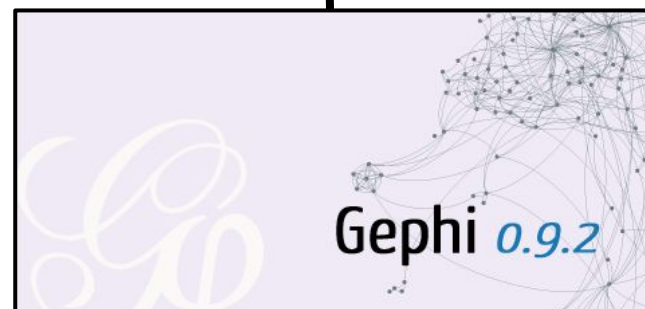


2008/7/31 釋放出第一版 0.6alpha1



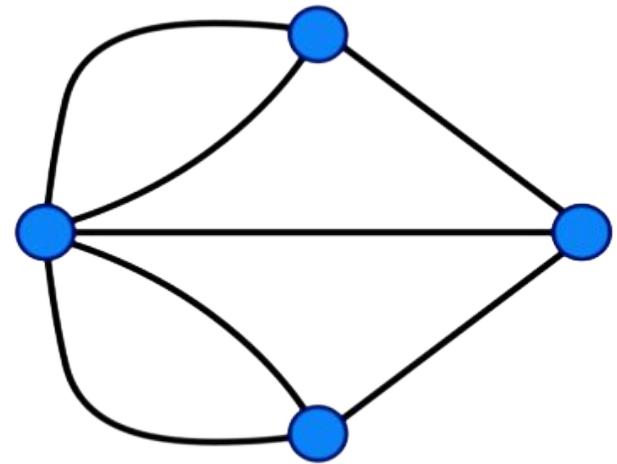
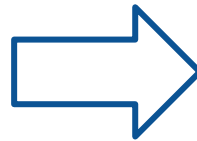
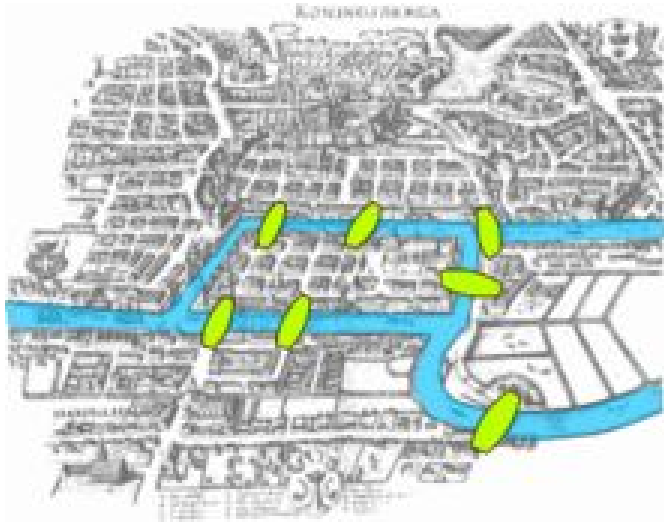
2010, Gephi Consortium 成立, 意在支持 Gephi 的發展
未來版本的釋出

最新版: ver. 0.9.2 2017/9/26



圖論(Graph Theory)

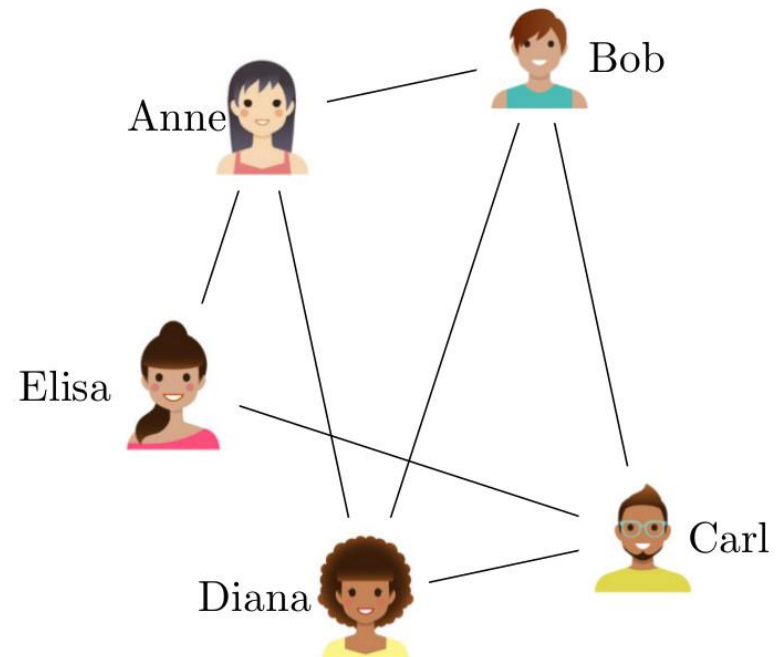
- 圖 (Graph $G=(V,E)$)
 - 節點 (Vertexts = Vertices, V)
 - 邊 (Edges, E)



Seven Brigs of Königsberg, Leonhard Euler , 1736

網絡(Networking)

- $G = (V, E) \rightarrow G = (N, L)$; 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

介面介紹-Overview

The screenshot displays the Gephi software interface, which is used for network visualization. The interface is divided into several panels:

- Modes Tab:** Located at the top, it includes tabs for Overview, Data Laboratory, and Preview. The Overview tab is currently selected.
- Appearance Panel:** On the left, it allows users to customize the look of nodes and edges. It includes sections for Nodes (Unique, Partition, Ranking) and Edges, with a color picker set to #c0c0c0.
- Layout Panel:** Below the Appearance panel, it shows the selected layout (Force Atlas) and a list of parameters for the Force Atlas layout, such as Inertia, Repulsion strength, Attraction strength, Maximum displacement, Auto stabilize function, Autostab Strength, Autostab sensibility, Gravity, Attraction Distrib., Adjust by Sizes, and Speed.
- Graph Visualization:** The central area displays a complex network graph with nodes and edges. The nodes are colored in shades of green, yellow, and purple, and the edges are colored in shades of green and purple.
- Context Panel:** On the right, it provides summary statistics for the graph, including Nodes (77), Edges (254), and Undirected Graph.
- Filters and Statistics Panel:** Below the Context panel, it offers various filters and statistics for the graph. The Network Overview section includes metrics like Average Degree, Avg. Weighted Degree, Network Diameter, Graph Density, HITS, Modularity (0.565), PageRank, and Connected Components. The Node Overview section includes Avg. Clustering Coefficient and Eigenvector Centrality. The Edge Overview section includes Avg. Path Length. The Dynamic section includes # Nodes, # Edges, Degree, and Clustering Coefficient.
- Node and Edge Properties Panel:** At the bottom, it allows users to configure the font, size, and color of nodes and edges. It includes a 'Global' tab and a 'Labels' tab.

Overview - 最主要的視覺化處理

- Mode Tab
 - 切換三個mode的按鈕
- Graph Visualization
 - 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

Modes Tab



Workspace 1

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
1.0	0.0	Undirected	0			1.0
2.0	0.0	Undirected	1			8.0
3.0	0.0	Undirected	2			10.0
3.0	2.0	Undirected	3			6.0
4.0	0.0	Undirected	4			1.0
5.0	0.0	Undirected	5			1.0
6.0	0.0	Undirected	6			1.0
7.0	0.0	Undirected	7			1.0
8.0	0.0	Undirected	8			2.0
9.0	0.0	Undirected	9			1.0
11.0	0.0	Undirected	13			5.0
11.0	2.0	Undirected	12			3.0
11.0	3.0	Undirected	11			3.0
11.0	10.0	Undirected	10			1.0
12.0	11.0	Undirected	14			1.0
13.0	11.0	Undirected	15			1.0
14.0	11.0	Undirected	16			1.0
15.0	11.0	Undirected	17			1.0
17.0	16.0	Undirected	18			4.0
18.0	16.0	Undirected	19			4.0
18.0	17.0	Undirected	20			4.0
19.0	16.0	Undirected	21			4.0
19.0	17.0	Undirected	22			4.0
19.0	18.0	Undirected	23			4.0
20.0	16.0	Undirected	24			3.0
20.0	17.0	Undirected	25			3.0
20.0	18.0	Undirected	26			3.0
20.0	19.0	Undirected	27			4.0
21.0	16.0	Undirected	28			3.0
21.0	17.0	Undirected	29			3.0
21.0	18.0	Undirected	30			3.0
21.0	19.0	Undirected	31			3.0
21.0	20.0	Undirected	32			5.0
22.0	16.0	Undirected	33			3.0
22.0	17.0	Undirected	34			3.0
22.0	18.0	Undirected	35			3.0
22.0	19.0	Undirected	36			3.0
22.0	20.0	Undirected	37			4.0
22.0	21.0	Undirected	38			4.0
23.0	11.0	Undirected	47			9.0
23.0	12.0	Undirected	46			2.0
23.0	16.0	Undirected	39			3.0
23.0	17.0	Undirected	40			3.0
23.0	18.0	Undirected	41			3.0

Data

Tool Bar

Add column Merge columns Delete column Clear column Copy data to column Fill column with values Duplicate column Create a boolean column from regematch Create column with list of regematching groups Negate Convert column to dynamic

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

Overview Data Laboratory **Modes Tab** Preview

Workspace 1

Preview Graph Data Table

Preview Setting

Graph Preview

Save

Export: SVG/PDF/PNG

Background Reset zoom - +

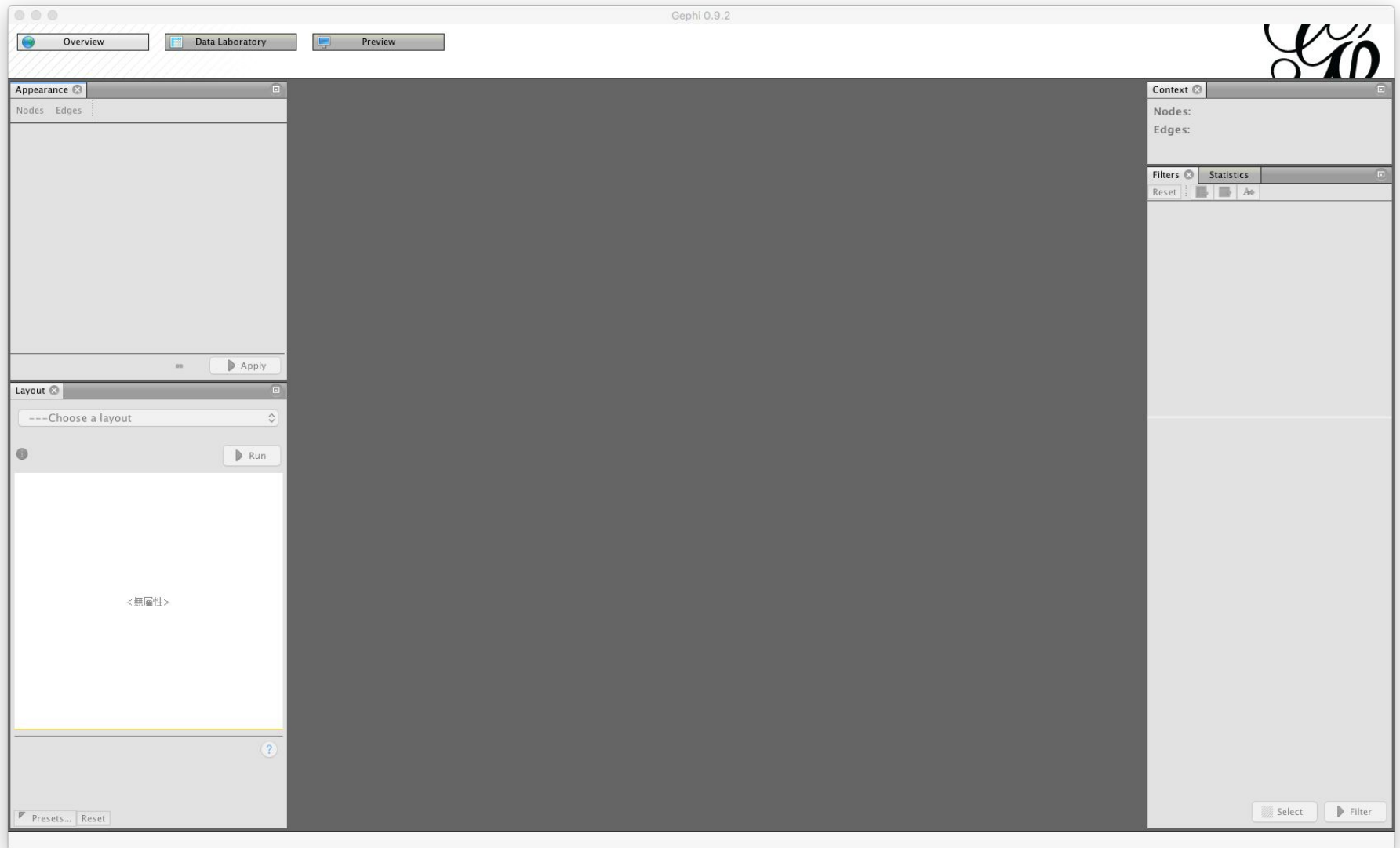
The screenshot displays the 'Preview' tab of the Data Laboratory software. On the left, a 'Preview Setting' panel is visible, containing sections for 'Nodes' (Border Width, Border Color, Opacity, Per-Node Opacity, Node Labels) and 'Edges' (Show Edges, Thickness, Rescale weight, Min. rescaled weight, Max. rescaled weight, Color, Opacity, Curved, Radius, Edge Arrows, Edge Labels). The 'Nodes' section is currently expanded. The main area shows a 'Graph Preview' of a network graph with numerous nodes and edges. The nodes are labeled with names, and the edges are colored and weighted. The interface includes a top navigation bar with 'Overview', 'Data Laboratory', and 'Preview' tabs. A 'Workspace 1' dropdown is also present. At the bottom, there is an 'Export' section with options for 'SVG/PDF/PNG' and a 'Background' checkbox. A 'Reset zoom' button and zoom controls are also visible.

Preview - 輸出前的微調

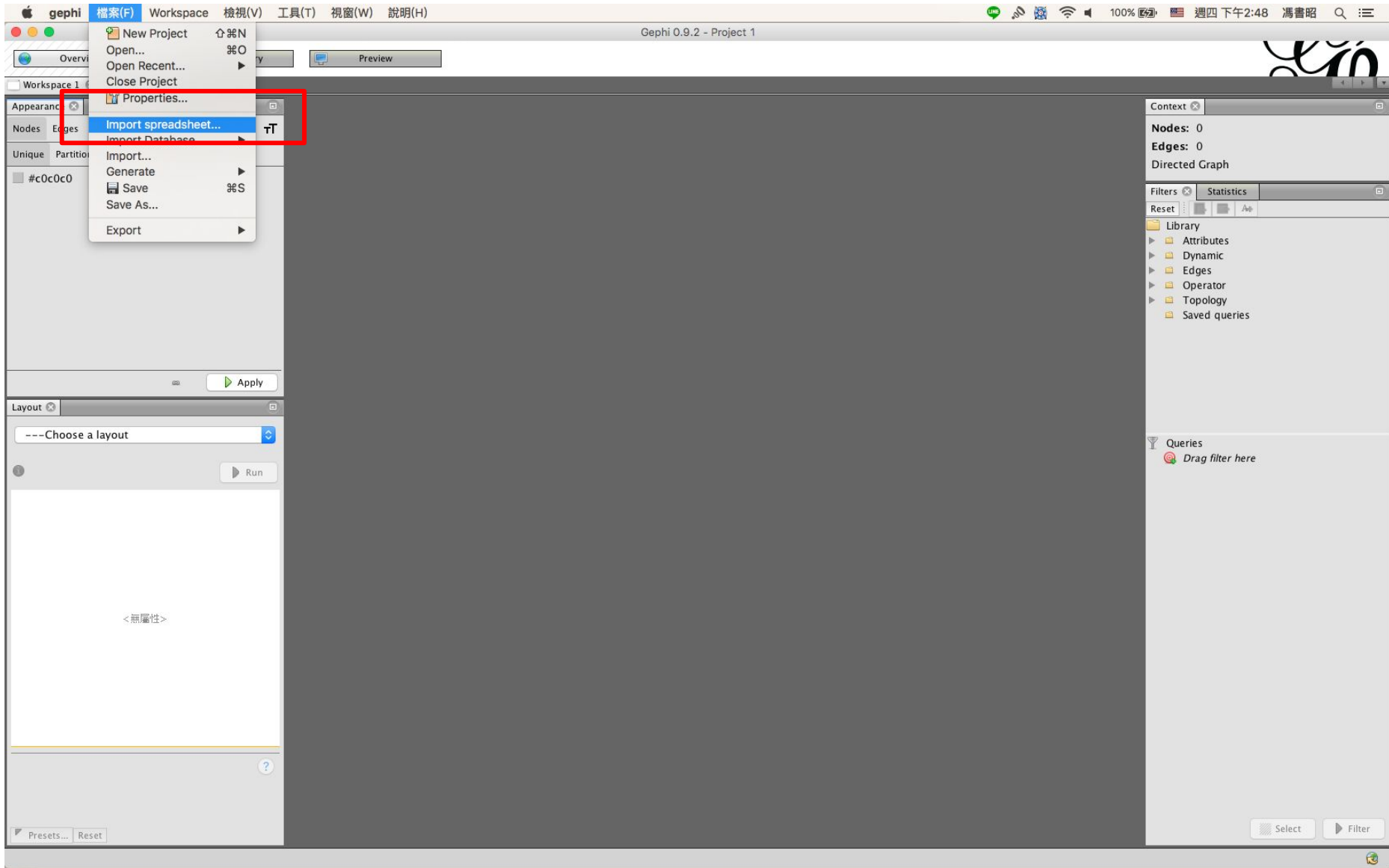
- Mode Tab
 - 切換三個mode的按鈕
- Preview Setting
 - 輸出微調
- Graph Preview
 - 預計輸出
- Save
 - 輸出影像

實作截圖

1. 開啟Gephi



2. 開始匯入檔案



3. 先匯入Node資料

The screenshot displays a network analysis software interface with a central workspace and several side panels. A file selection dialog is open in the center, showing a list of files in a 'data' folder. The dialog has a title bar with '開啟' (Open) and a search bar with 'data'. The file list has two columns: '名稱' (Name) and '修改日期' (Modification Date). The file 'LesMiserables [Nodes].csv' is selected and highlighted in yellow. Below the file list, there is a '檔案格式' (File Format) dropdown menu set to '所有檔案' (All Files). At the bottom of the dialog are '取消' (Cancel) and '開啟' (Open) buttons.

The software interface includes the following panels:

- Overview**: Contains 'Appearance' and 'Layout' tabs. The 'Appearance' tab shows 'Nodes' and 'Edges' settings, including a color picker set to '#c0c0c0'. The 'Layout' tab shows a 'Choose a layout' dropdown and a 'Run' button.
- Data Laboratory**: A panel for data management.
- Preview**: A panel for previewing data.
- Context**: A panel showing graph statistics: 'Nodes: 0', 'Edges: 0', and 'Directed Graph'.
- Filters**: A panel for filtering data.
- Statistics**: A panel for network statistics, including 'Network Overview' (Average Degree, Avg. Weighted Degree, Network Diameter, Graph Density, HITS, Modularity, PageRank, Connected Components) and 'Node Overview' (Avg. Clustering Coefficient, Eigenvector Centrality).
- Edge Overview**: A panel for edge statistics (Avg. Path Length).
- Dynamic**: A panel for dynamic statistics (# Nodes, # Edges, Degree, Clustering Coefficient).

4. 類型請選擇Node

The screenshot shows the Gephi 0.9.2 interface with the 'Spreadsheet (CSV)...' dialog box open. The dialog is for importing a CSV file. The 'General CSV options (1/2)' tab is active. The 'CSV file to import' field shows the path '/Users/JF/Desktop/gephi_intro/data/LesMiserables [Nodes].csv'. The 'Separator' is set to 'Comma'. The 'Import as' dropdown menu is open, showing 'Nodes table' (selected) and 'Edges table'. The 'Charset' is set to 'UTF-8'. A preview table is visible below the dropdown.

Id	Lat	Matrix
0.0	Myriel	
1.0	Napoleon	
2.0	MlleBaptis...	
3.0	MmeMagl...	
4.0	Countess...	
5.0	Geborand	
6.0	Champter...	
7.0	Cravatte	

5. 可再度確認節點數, 這裡不新增新的工作表

The screenshot shows the Gephi software interface with the 'Import report' dialog box open. The dialog box is titled 'Import report' and has a 'Source: Stream ImporterSpreadsheetCSV'. It contains a 'Report' tab and an 'Issues' tab. The 'Report' tab shows 'No issue found during import'. The 'Issues' tab is highlighted with a red box. Below the report, there are configuration options for the graph type, edges merge strategy, and workspace options. The 'Graph Type' dropdown is set to 'Undirected' (highlighted with a red box). The 'Edges merge strategy' is set to 'Sum'. The 'Create missing nodes' checkbox is checked. The 'Self-loops' checkbox is unchecked. The '# of Nodes' is 77 (highlighted with a red box). The '# of Edges' is 0. The 'Dynamic Graph' is 'no'. The 'Dynamic Attributes' are 'no'. The 'Multi Graph' is 'no'. The 'New workspace' radio button is selected (highlighted with a red box). The 'Append to existing workspace' radio button is also selected. At the bottom of the dialog box are '取消' (Cancel) and '確定' (OK) buttons.

Workspace 1

Appearance

Nodes Edges

Unique Partition Ranking

#c0c0c0

Layout

---Choose a layout

Run

<無屬性>

Presets... Reset

Import report

Source: Stream ImporterSpreadsheetCSV

Issues Report

No issue found during import

Graph Type: Directed Undirected Mixed

Auto-scale

Edges merge strategy: Sum

Create missing nodes

Self-loops

of Nodes: 77

of Edges: 0

Dynamic Graph: no

Dynamic Attributes: no

Multi Graph: no

New workspace

Append to existing workspace

取消 確定

Context

Nodes: 0

Edges: 0

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

6. 匯入成功

OverviewData LaboratoryPreview

Workspace 1

Data Table

NodesEdgesConfigurationAdd nodeAdd edgeSearch/ReplaceImport SpreadsheetExport tableMore actions

Filter:Id

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	Fauchelevant	
29.0	Bamatabois	
30.0	Perpetue	
31.0	Simplice	
32.0	Scaufflaire	
33.0	Woman1	
34.0	Judge	
35.0	Champfmathieu	
36.0	Brevet	
37.0	Chenildieu	
38.0	Cocheapille	
39.0	Pontmercy	
40.0	Boulatruelle	
41.0	Eponine	
42.0	Anzelma	
43.0	Woman2	

Add columnMerge columnsDelete columnClear columnCopy data to other columnFill column with a valueDuplicate columnCreate a boolean column from regex matchCreate column with list of regex matching groupsNegate boolean valuesConvert column to dynamic

7. 同樣步驟匯入Edges, 這裡有進階選項

Workspace 1

Data Table

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

Filter: Id

Id Label Interval

0.0 Myriel

1.0 Napoleon

2.0 MlleBaptistine

3.0 MmeMagloire

4.0 CountessDeLo

5.0 Geborand

6.0

7.0

8.0

9.0

10.0

11.0

12.0

13.0

14.0

15.0

16.0

17.0

18.0

19.0

20.0

21.0

22.0

23.0

24.0

25.0

26.0

27.0

28.0

29.0

30.0

31.0

32.0

33.0

34.0

35.0

36.0

37.0

38.0

39.0

40.0

41.0

42.0

43.0

Chenildieu

Cochepaille

Pontmercy

Boulatruelle

Eponine

Anzelma

Woman?

Import report

Source: Stream ImporterSpreadsheetCSV

Issues Report

No issue found during import

Graph Type: Undirected More options...

☒ Auto-scale

☐ Create missing nodes

☐ Self-loops

Edges merge strategy: Sum

of Nodes: 77

of Edges: 254

Dynamic Graph: no

Dynamic Attributes: no

Multi Graph: no

☐ New workspace

☒ Append to existing workspace

取消 確定

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column Create a boolean column from regex match Create column with list of regex matching groups Negate boolean values Convert column to dynamic

8. 可以在第二頁把錯誤的類型(e.g. 文字->數字)改回來

The screenshot shows the Data Laboratory interface with the 'Import settings (2/2)' dialog box open. The dialog box has a sidebar with steps: 1. General CSV options, 2. Import settings. The main area is titled 'Import settings (2/2)' and contains a list of 'Imported columns' with checkboxes: Source, Target, Type, Id, Label, and Weight. The 'Type' column is checked and has a dropdown menu set to 'Double'. This dropdown menu is highlighted by a red rectangle. At the bottom of the dialog box are buttons: 說明(H), < 上一步(B), 下一步 >, 完成(F), and 取消.

Workspace 1

Overview Data Laboratory Preview

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

Spreadsheet (CSV)...

步驟

1. General CSV options
2. Import settings

Import settings (2/2)

Imported columns:

- ☒ Source
- ☒ Target
- ☒ Type
- ☒ Id
- ☒ Label
- ☒ Weight

Double

說明(H) < 上一步(B) 下一步 > 完成(F) 取消

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column Create a boolean column from regex match Create column with list of regex matching groups Negate boolean values Convert column to dynamic

9. 都匯好後把圖表打開，發現目前點是隨機散佈的

The screenshot displays the Gephi 0.9.2 software interface. The main window shows a network graph with 77 nodes and 254 edges, which are currently randomly distributed. The 'Graph' menu is open, showing options like Context, Data Table, Filters, Graph, Layout, 輸出(O), Preview, Preview Settings, Appearance, Welcome, Statistics, and Timeline. The 'Layout' window on the left allows selecting a layout. The right sidebar provides network statistics and overview panels.

Context

- Context
- Data Table
- Filters
- Graph**
- Layout
- 輸出(O)
- Preview
- Preview Settings
- Appearance
- Welcome
- Statistics
- Timeline
- Configure Window
- 重設視窗(W)
- Close Window
- Document Groups

Network Overview

Metric	Run
Average Degree	Run
Avg. Weighted Degree	Run
Network Diameter	Run
Graph Density	Run
HITS	Run
Modularity	Run
PageRank	Run
Connected Components	Run

Node Overview

Metric	Run
Avg. Clustering Coefficient	Run
Eigenvector Centrality	Run

Edge Overview

Metric	Run
Avg. Path Length	Run

Dynamic

Metric	Run
# Nodes	Run
# Edges	Run
Degree	Run
Clustering Coefficient	Run

10. 可以藉由先計算Modularity來找出分群

The screenshot displays a network visualization software interface. A central dialog box titled "Modularity settings" is open, showing the "Modularity" community detection algorithm. The dialog includes checkboxes for "Randomize" and "Use weights", both of which are checked. A "Resolution" slider is set to 1.0. The dialog also contains a description of the algorithm and buttons for "取消" (Cancel) and "確定" (OK).

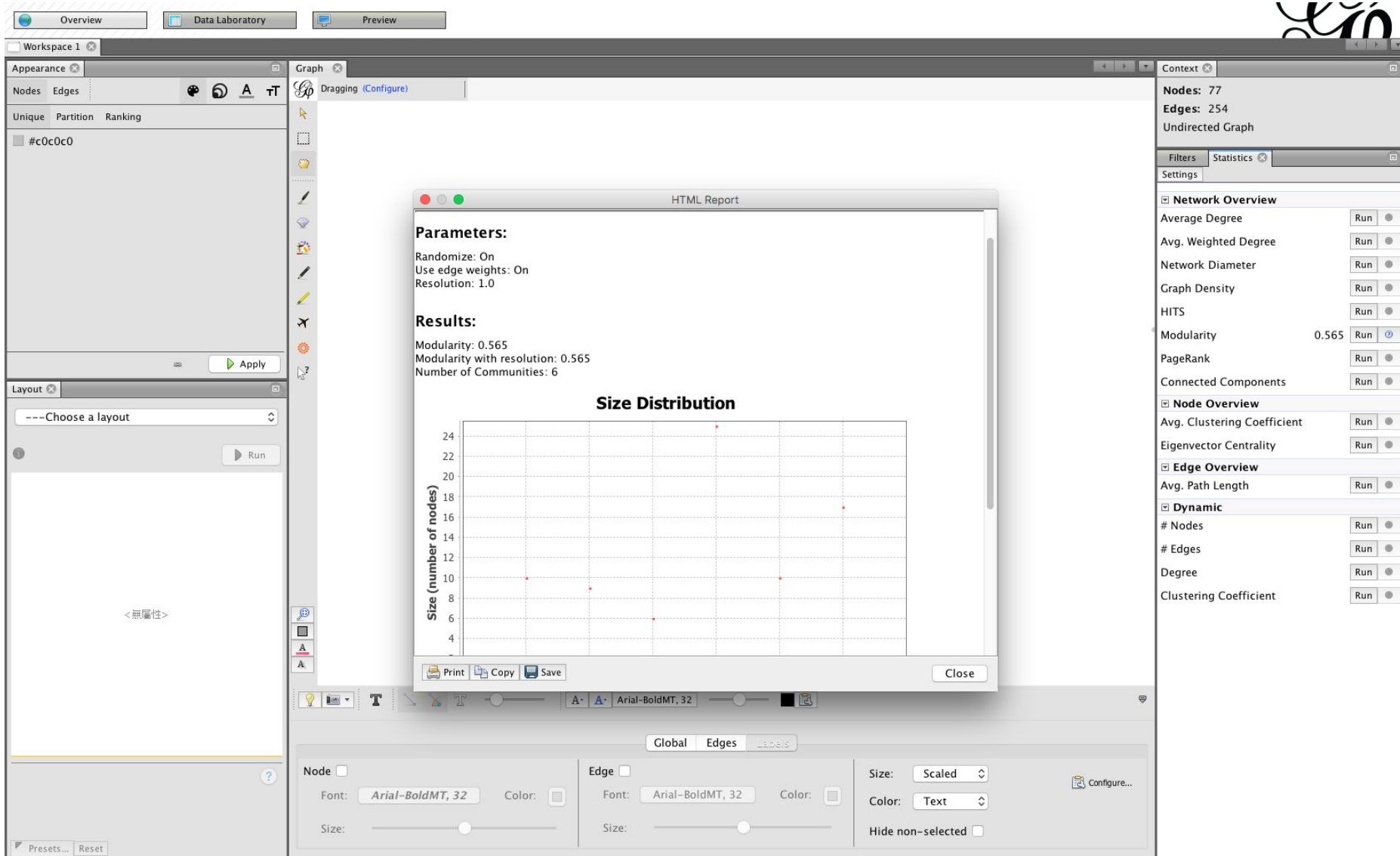
The background interface includes several panels:

- Appearance:** Contains tabs for "Nodes" and "Edges". The "Nodes" tab is active, showing a color selection area with the hex code "#c0c0c0".
- Layout:** Contains a dropdown menu labeled "Choose a layout" and a "Run" button.
- Graph:** The main workspace area, currently showing a small network graph.
- Context:** Contains a "Statistics" tab with a list of network metrics and their corresponding "Run" buttons. The "Modularity" metric is highlighted with a red box.

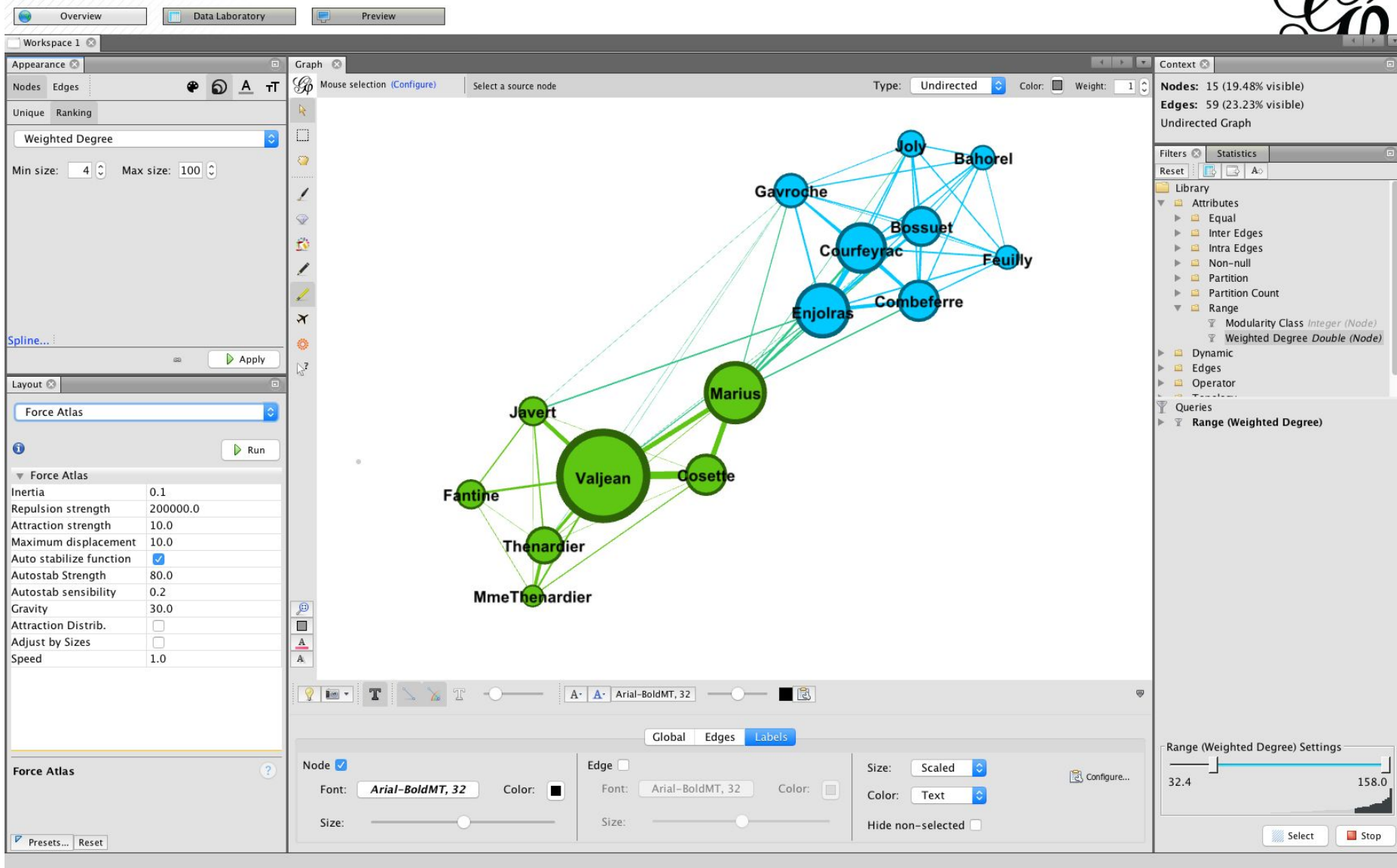
The "Context" panel lists the following metrics:

- Nodes: 77
- Edges: 254
- Undirected Graph
- Filters
- Statistics
- Settings
- Network Overview
 - Average Degree
 - Avg. Weighted Degree
 - Network Diameter
 - Graph Density
 - Modularity (highlighted)
 - PageRank
- Connected Components
- Node Overview
 - Avg. Clustering Coefficient
 - Eigenvector Centrality
- Edge Overview
 - Avg. Path Length
- Dynamic
 - # Nodes
 - # Edges
 - Degree
 - Clustering Coefficient

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)

The screenshot displays a network visualization software interface. On the left, the 'Preview Settings' panel is open, showing various rendering options for nodes and edges. The 'Node Labels' section is expanded, showing settings for 'Show Labels' (checked), 'Font' (Arial 36 無格式), 'Proportional size' (unchecked), 'Color' (custom [43,74,194]), 'Shorten label' (unchecked), 'Max characters' (30), 'Outline size' (0.0), 'Outline color' (custom [255,255,...]), 'Outline opacity' (80.0), 'Box' (unchecked), 'Box color' (parent), 'Box opacity' (100.0), 'Edges' (checked), 'Thickness' (1.0), 'Rescale weight' (unchecked), 'Min. rescaled weight' (0.1), 'Max. rescaled weight' (1.0), 'Color' (mixed), 'Opacity' (100.0), 'Curved' (checked), 'Radius' (0.0), 'Edge Arrows' (unchecked), 'Size' (10.0), 'Edge Labels' (unchecked), 'Font' (Arial 10 無格式), 'Color' (original), 'Shorten label' (unchecked), 'Max characters' (30), 'Outline size' (2.0), 'Outline color' (custom [255,255,...]), and 'Outline opacity' (40.0). The 'Preview ratio' is set to 100%. The 'Export' button is visible at the bottom left.

In the center, an 'Export' dialog box is open, showing '另存新檔: Untitled.pdf' and 'JF' as the file format. Below it, an 'Options PDF' dialog box is open, showing 'Unit: Inch', 'Page size: A4', 'Width: 8.3 in', 'Height: 11.7 in', 'Orientation: Portrait' (selected), and 'Margins: Top: 0.2, Left: 0.2, Bottom: 0.2, Right: 0.2'. The 'OK' and 'Cancel' buttons are at the bottom.

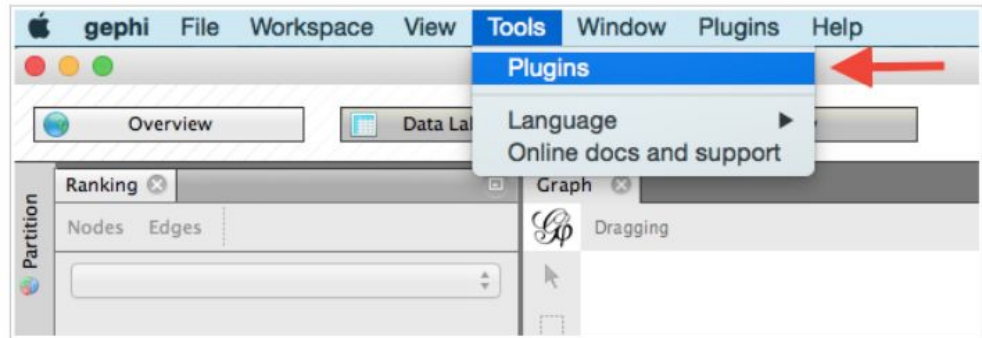
On the right, a network graph is visible, showing nodes labeled with names like Joly, Bahorel, Gavroche, Courfeyrac, Bossuet, Feuilly, Combeferre, Jolras, Fantine, Thenardier, and Mme Thenardier, connected by edges. A list of dates and times is visible on the right side of the graph area.

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)

Extensible

The built-in Plugins Center automatically gets the list of plugins available from the Gephi Plugin portal and takes care of all software updates. There are dozens of community-built plugins that extends Gephi's functionalities.



補充資料2

- Gephi Developer

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

 Graph API

 Layout API

 Attributes API

 Statistics API

 Import API

 Export API

 Tools API

 Filters API

 Generator API

 Project API

 LongTask API

感謝聆聽