Bio notes

• Education in economics, management, history of science (Ph.D.)

• Turned to digital methods for research.
  – data visualization, network analysis, natural language processing, web applications and more.
  – Member of the “Gephi Community Support” team
  – Gephi certified trainer
    http://www.clementlevallois.net

• Contact, feedback welcome: on twitter @seinecle or www.clementlevallois.net
THE GEPHI TEAM
Gephi

- Created in 2008 by a core team of 4 French computing engineers inspired by a professor.

The initial Gephi Team
- Mathieu Bastian / www
- Sebastien Heymann / www
- Julian Bilcke / www
- Mathieu Jacomy
- Franck Ghittal / www
Gephi

- In 2016,
  - 11 core developers
  - 15 developers of plugins
  - 4 Google Summer of Code
  - One Java Duke Award

- **1.5 million downloads since 2013**

- Localization is available in **French, Spanish, Japanese, Russian, Polish, Brazilian Portuguese, Chinese** and **Czech**.

- About 1,600 academic papers citing Gephi

- Very active Facebook group (**https://www.facebook.com/groups/gephi/**)
Gephi

• A software written in Java for Mac, PC and Linux

• A headless version (the Gephi Toolkit) to generate graphs automatically via API

• An ecosystem of plugins and related tools

• A commitment to open source, quality, and freedom to use.
The future of Gephi

• In version beta 0.9 since Dec. 20, 2015
  – Close to the final, stable version
  – Comparison with version 0.8:
    • Quicker
    • Cleaner code, easier to maintain and evolve
    • Accomodates larger graphs
    • Time / dynamic graphs handled differently
A note on the slides

Signals an important feature, or an error to avoid.
SETUP
Check list

- Make sure you have the latest version of Gephi installed

- In the menu, choose Help -> Check for updates

- You need a mouse with a scrolling wheel, to zoom ( )

- In the menu of Gephi, go to Tools -> Options -> Visualization -> OpenGL
  There, increasing the anti-aliasing factor can improve the quality of the view on your screen.

- We are going to download some stuff. Check your internet connection.

- In case you manipulate large networks, Gephi might need to draw on more memory on your laptop. Keep this link in your bookmarks when comes the time to increase the memory that is made available to Gephi:
  - https://gephi.org/users/install/
Adding plugins

• In the menu, go to Tools -> Plugins -> Available plugins

• Select with the tick box the following plugins: Circular Layout, OpenOrd Layout, alphabeticalsorter, Force Atlas 3D, SigmaExporter

• Click install and follow the instructions.

Note: You will remark there is a “Plugins” item in the menu as well, which is empty. That’s a secret feature. Or just a very misleading UI design.
A note on the terminology

A “node” (also called a vertex, plural vertices).

An “edge” (or a tie, or a link...)

Y was born in ...
Y’s marital status is ...
Y’s latitude is ...
Y’s longitude is ...

Nodes attributes
(NB: edges can have attributes too!)

A directed network (the direction of the relation matters)

An undirected network (the direction of the relation does not matter)

A weighted network (the edges have a “strength” represented by a numerical value)
PRACTICE A WORKFLOW IN GEPHI

From A to Z
Steps

• Import the network

• Spatialize the network

• Visualize attributes of your data

• Visualize attributes created by Gephi

• Export a visualization as a picture or pdf
The dataset

• Co-occurrence network from Les Miserables, by Victor Hugo

• 2 characters are connected if they are mentioned close to each other in the text

• Download and unzip this file

  www.clementlevallois.net/gephi/tuto/miserables.zip

Source: Dataset originally provided by The Stanford GraphBase,
http://www.ctan.org/tex-archive/support/graphbase/
1. Importing the dataset in Gephi
2. Import Report

This report provides information on the file you just opened.

- **Source:** Miserables.gexf

### Issues Tab

- **Nodes**
  - Default edge type set as UNDIRECTED: INFO
  - GEXF version 1.2: INFO

- **Auto-scale**
  - Check box for auto-scaling

- **Create missing nodes**
  - Options: New graph, Append Graph, Time frame

- **Dynamic Graph:** no
- **Hierarchical Graph:** no

![Import Report](image)

- **Number of nodes found:** 77
- **Number of edges found:** 254

Should an edge A-B be seen actually as A->B? You decide it here.

Often, networks files are simply a list of nodes (ex: A, B and C), and a list of edges connecting these nodes (ex: A –B, B-C). What if Gephi finds an edge A-D in the file? Since D is not present in the list of nodes, you are given here a choice to add it now. If not, the edge A-D will be ignored.

If you have a time-stamped graph. Importing the same graph with different time-stamps creates a “time evolving” (dynamic) graph.

If your graph is a meta-graph?

*we will ignore hierarchical graphs in this tutorial*

If you want to merge the graph with one that is already open in Gephi:

Scales nodes size. Don’t select if your network file specifies node sizes!
3. The graph, opened
4. Saving a Gephi project

• Gephi can save projects under the .gephi extension.
  – These files can only be opened in Gephi, and contain all information about your project. **Click File -> Save.**

⚠️ There are a number of reports of this files to be unstable, so **always also export your network** as indicated just below.

• The bare network itself can be exported as a file, in different formats.
  – gexf and graphML are two network formats which can store attributes for nodes and edges.

=> **Click File -> Export -> Graph File** to save your network.
5. Mouse controls

- Left click: accomplishes the action you previously selected
- Right click: move the graph around ("panning")
- Scrolling wheel: zooming in and out

"Alt" key + keep left click down while on the graph: moves the graph in the 3\text{rd} dimension (gives an effect of depth and perspective)
6. Rearranging windows, tabs, panels

My experience: don’t use this “Configure Window” menu.

Just drag, drop and resize the tabs with the mouse as you find most convenient.

Try to maximize the space for the main tab – the one where the graph is displayed!
7. Switching on labels

- Please refer to the cheat sheet in the appendix of this slide deck.

- Notice the panel at the bottom of the screen, and its “Labels” tab.

- Edges and nodes can have labels but in most use cases, only nodes have them.
8. What a layout is

- A layout is simply a procedure to position nodes on the screen according to some logic.
- Many procedures exist – you could create your own if you’d wish
9. Example of a simple layout

- Select the “Circular Layout” in the drop down menu of layouts.
- Don’t change any setting and click “run”
10. Example of a simple layout (continued)

The “circular layout” applies the following procedure: “spread nodes on a circle, at an equal distance from each other.”

Can we distribute nodes according to different procedures, yielding more insights?

- Select “Force Atlas 2” in the drop down menu of the layouts.
- Change the scaling parameter to “200” (not necessary, actually...)
- Run it!
12. Force Atlas 2 – the result

The result is that densely connected nodes tend to be grouped in the same regions of the screen.

This property is very useful to interpret a graph:
- groups of nodes can be spotted immediately,
- Relations between groups are also visible.

Recent works have demonstrated analytically this result (that with certain kinds of layouts, densely connected groups of nodes find a visual translation in a 2D plane). See Noack (2009) and Waltman, van Eck and Noyons (2010) in the reference list.
13. How do force-based layouts work

• “Nodes repulse each other like magnets, while edges attract their nodes, like springs”

• “These forces create a movement that converges to a balanced state.”

• Force Atlas 2 specificity:

“[We] tweak the repulsion [“magnet”] force so that poorly connected nodes and very connected nodes repulse less”

(see Jacomy et al. 2014 in the reference list).

- **Force Atlas and Force Atlas 2**
  - Force Atlas 2 has just been explained, the differences with Force Atlas are explained in the next slide

- **Fruchterman-Rheingold**
  - The classic one. Nodes that are connected attract each other

- **Kamada – Kawai**
  - More elaborated: the attraction force between any 2 nodes is proportional to the length of the shortest path joining them

- **Yifan Hu**
  - Gathers nodes in groups, then applies a force-based logic to the groups
Increase the speed
Each « core » of a computer processor has 2 threads. For a computer with a dual core, choose a thread number of 4.

Pushes to the periphery the nodes with many outbound links
Interesting only when the network is directed. Exemple for a network of webpages: websites with many outbound links, but few inbound links, could end up at the center of the network. But when this option is selected they are pushed to the periphery.

Increases the repulsion force
Repulsion force between unconnected nodes decreases with distance. When this option is selected, the repulsion force remains constant – leading to a network which is more expanded in space.

Prevents nodes from overlapping
Very convenient to keep the network readable. Choose and apply the layout « Label Adjust » (not shown here) to make that node labels as well don’t overlap.

Controls the strength of the links connecting nodes
If the links have weights, makes these weights overall stronger or weaker

Degree of approximation
Increasing this factor increases speed (if the option above is selected), but at the cost of flickering of node positions if the value is too high.

Increases dramatically the speed
By approximating the calculus of the repulsing force, instead of computing it exactly at each step.
15. Force Atlas or Force Atlas 2?

• Force Atlas 2 is a reworked version of Force Atlas.
  – Its implementation has been reworked to make it faster
  – The algorithm is faster thanks to multithreading

• Parameters are slightly different:
  – The force of attraction and force of repulsion in « Force Atlas » have been merged in a single parameter (« scale ») in « Force Atlas 2 ».
16. A way to misread a layout

• Are the nodes in the center more central?

This slide is based on Scott Weingart’s blog post: « Networks Demystified 8: When Networks are Inappropriate » – See refs at the end.
17. The trap, explained

• Las Vegas does appear as relatively central on the map: lots of connections to other cities, and in the middle of the South West region.

• Travelers often go to Vegas, stay a bit, then go back home.

• So Las Vegas is not a connector between airports (« hub »). It is not central in the sense of a favorite place to transit from one place to another.

• As Scott Weingart says, that’s more of a « sink ».
18. Local minima: another way force-based layouts can be misleading

• The layout can « remain stuck » in a position which does not represent the most definitive equilibrium of forces.

• Solution:
  – Keep this possibility in mind: for large graphs (~200+ nodes or more), the precise position of an individual node must be critically interpreted. Being in the neighborhood of another node might not reflect belonging to the same community (especially if they are not connected, and if the community detection algo places them in different communities).
  – However, the positioning of groups / communities of nodes remains a robust indicator of who sits in the same connectivity region.
  – I personally start with a very expanded network (large value of the « scale » parameter in Force Atlas 2) before shrinking the network to a decent size, in order to give maximum space/freedom to the nods to move around and end up in a global minima.
19. Trying out Force Atlas 3D

20. OpenOrd for very large graphs?

- My personal opinion:
  - The logic of OpenOrd can’t be explained easily
    If you’d like to try:
    Available here: http://www.researchgate.net/publication/253087985_OpenOrd_an_open-source_toolbox_for_large_graph_layout
  - The resulting layout is not the best
    • Communities are not clearly separated
    • Compare with Force Atlas 2 to see.
  - And Force Atlas 2 is actually quicker on a network of 5,000 nodes and 24,000 links.

➤ So I always recommend to use Force Atlas 2 instead.
21. Making nodes bigger

• Please refer to the cheat sheet ② in appendix to this slide deck.

• **Right**-click on the icon marked “size” and set the number to 15 or higher. Validate.

• Then **left**-click on the same “size” icon.
22. Visualizing attributes in your data

• Nodes in our example network have one attribute: the gender of the character in “Les Miserables”.

• We can overlay this information on the network by color-coding the nodes according to the values of this attribute.

• Note: we have 2 values for this attribute (male or female), but this could be many more.
23. Partitioning or ranking?

- Partitioning is this:

Works for attributes that classify nodes in categories: “male or female”, “East, West, South or North”, “country of residence”. Each category is represented by a different color, as above.

- Ranking is this:

or this:

Works for attributes that are gradual, not categorical (age, etc... any numerical attribute).
The graduation is represented visually either by bigger sizes, or by changes in shades of color (from light to dark, or from light colors to warm colors, usually).
24. Example of partitioning in Gephi 0.8.2

- In the “partition” tab, **hit the refresh button** to make sure Gephi loads all available attributes.

- In the drop down menu, we then select “Gender” which is anyway the only available attribute available for partition in our case.

- **Click on “apply”!**

  Note that if we had had a categorical attribute for an edge (relation between two characters: friend, enemy or neutral?), we could have colorized edges according to the same principle.

- **Note that if we wanted to group all nodes from the same category as one big node, we would have gone here. (that’s hierarchical graphs)**
24. Example of partitioning in Gephi 0.9

- In the “Appearance” select “Gender” which is anyway the only available attribute available for partition in our case.
- Modify the color of each group individually by clicking on each colored square.
- Or choose among preselected palette of colors, defined to use colors easy to distinguish.
- Click on “apply”!

⚠️ If you had a numerical attribute (not distinct groups, but nodes with varying ages, scores, temperatures... then you’d have to choose “ranking” here.
25. Visualizing new attributes with Gephi

- Gephi can compute metrics, such as “how central is each node” in the network.

The key insight is: these metrics are attributes to nodes (or edges), and can in turn be visualized just as any other attribute.

- Visualizing these metrics (instead of reading them in a tabular form as is usually done) stimulates / speeds up insight.
26. Detecting communities & visualizing them

- Go to the statistics panel and run the Modularity algorithm.

- This algo identifies relatively densely connected groups of nodes in the network.

- As a result, each node ends up being attributed to one of these groups (called “class” here).

- The same way we have colorized nodes by their genders, we can colorize them by the group (“class”) they belong to.

- Go to the Appearance panel, select “modularity class”, and “apply”!

Note: by clicking on the colored squared next to each category in the partition tab, you can change and choose the colors you prefer.
27. Communities and “modularity”

• Modularity (Newman 2004)
  – Number of links in each community minus the number of links in the same groups in a graph where the links would have been redistributed randomly.

  – Finding communities in a graph = define groups in a manner that the modularity score is the highest.
28. Visualization of communities: the result
29. Tips to clean up the appearance

• Apply the layout “Label adjust”
  – To clean up overlapping labels
  – Find it in the drop down menu of the layouts

• Switch to a black background?
  – Check the bottom left corner of cheat sheet ②

• Decrease the thickness of edges?
  – A slider makes it easy
30. Gephi: what’s the output?

- The next slide shows how to export your viz to a picture file. Is it what Gephi is for, ultimately?

- In my view, the value added is before all in the insights you gain in the process of exploring the viz interactively, and iteratively, *in Gephi*.

- Exporting the view to a file or a webpage meets different needs, such as communicating the insights you gained.

31. Export your visualization!

- Directly from the overview panel with the screenshot option (see cheat sheet \( \textcircled{2} \)).
- Or as a web page! With the SigmaExporter plugin.
  - Do File -> Export -> Sigma.js template and then simply upload the folder to a server
- Or in the preview panel (see cheat sheet \( \textcircled{5} \)).

Why do the views in the overview and preview panels don’t exactly match?

- Visualizing a graph in real time as in the overview panel requires a technology (OpenGL) which is incompatible with an export in the pdf or svg file formats.
- The preview panel provides this bridge towards pdf and svg, by using a different technology to render the graph.
- This difference in technology means that some visual features in the overview panel don’t translate exactly to the preview mode.
32. Les Miserables network just before export to pdf
33. To adjust labels, add a legend:

Inkscape

Adobe Illustrator
34. We’ve learned how to...

• Configure Gephi.
• Read an import report
• Use the Gephi interface
• Interpret a layout
• Understand the logic of force based algos
• Be careful about misinterpretations in reading graphs.
• Use the ranking, partitioning, and metrics panels
• Export a network as a picture
Next step:
importing data into Gephi
The Overview Panel in **Gephi 0.9**

Where all the functions are available to explore the network visually.

**Appearance**
To change the size and color of the network according to some values.
- Example: to color all nodes representing males in yellow, and all nodes representing females in green.
- Use the “Unique” tab to color / resize all nodes and edges uniformly (all in green, all in a given size...)

**Statistics**
These metrics are recorded, and can then be used to be displayed on the graph. Ex: compute the centrality of nodes. Then use the Appearance panel to make central nodes bigger.

**Layout**
Changes how the network is spread on the screen.
- Select one of the layouts in the drop-down menu and apply it to see how the position of nodes is changed.

**Partition**
(tab hidden here)

**Filters**
To hide or display only part of the network.
- The “library” and its folders contain the filters. For example, filter out nodes which have less than 3 edges. Or filter out edges which have a weight above some value.
- Drag and drop the filter you choose in the “queries” window. Several filters can be combined (ex: filter out male individuals that have less than 3 connections to others).

**General settings and options for the look of edges and labels. Access them by clicking on the small arrow**

This box of settings can be shown / hidden with the little arrow circled in red on the top right. Here, you can set if nodes are visible in 2D or 3D, what is the default color of edges, etc. The “labels” tab is particularly useful: should they be displayed or not, and at which size.
The Overview Panel in **Gephi 0.8.2**

Where all the functions are available to explore the network visually.

**Partition**
(tab hidden here)
For categorical attributes.
Example: to color all nodes representing males in yellow, and all nodes representing females in green.

**Ranking**
For numerical attributes.
Example: the older the individual, the bigger the node. Works with gradients of colors too (ex: the older the indiv., the pinker it is).

**Layout**
Changes how the network is spread on the screen. Select one of the layouts in the drop-down menu and apply it to see how the position of nodes is changed.

**Statistics**
(tab hidden here)
Computes metrics on the network. These metrics are recorded, and can then be used to be displayed on the graph. Ex: compute the centrality of nodes. Then use the ranking panel to make central nodes, bigger.

**Filters**
To hide or display only part of the network. The “library” and its folders contain the filters. For example, filter out nodes which have less than 3 edges. Or filter out edges which have a weight above some value.

Drag and drop the filter you choose in the “queries” window. Several filters can be combined (ex: filter out male indiv. that have less than 3 connections to others).

**General settings for the appearance of nodes, edges and labels**
This box of settings can be shown / hidden with the little arrow circled in red on the top right. Here, you can set if nodes are visible in 2D or 3D, what is the default color of edges, etc. The “labels” tab is particularly useful: should they be displayed or not, and at which size.
Focus on the icons of the Overview panel

Functions which are less frequently used have been grayed out.

Gephi Cheat Sheet

Switch background color! (useful for prints)

Export a screenshot.

Click the arrow for resolution settings

How to memorize all these icons??

All these controls are also available with a more explicit description in the panel here.

Once you know these controls well, the icons are a quick way to access them.

Select individual nodes
Select a region
Move individual nodes
Change the color of a node
Change the size of a node
Change the color of a node and its neighbors
Add one new node
Add one new edge
Finds shortest path between 2 nodes
Change the color of a node and its neighbors – edge weight aware.
Display the information of the selected node

Resize the graph to the size of the window.
Changes the color of all nodes to a unique color. Right click to define the color.
Resets the size of all nodes. Right click to define this size.
Colors all labels in a unique color
??
Resets the size of all labels.

More label settings:

1. The 3 buttons on the left:
   - Label size – should it track the node size?
   - Label color – should it track the node color?
   - Label font

2. The slider: Label size

3. The 2 buttons on the right:
   - Label color
   - Text to be displayed in the label

From left to right, switch on or off the display of:
- Nodes labels
- Hulls (not implemented yet)
- Edges
- Edge of the color of the source node
- Edge labels

Change edge thickness
The Data Laboratory

Where the numerical and textual data for nodes and edges can be examined and modified.

Import function
Opens a dialog window to import nodes and edges from a csv file into Gephi.

To switch between views of nodes & edges

3 default columns for nodes
- **Node**: simply a copy of the label column (or the id if there are no labels).
- **Id**: the unique identifier of the node.
- **Label**: the name of the edge which will be displayed next to it if we choose to.

Extra columns
Each node can have extra information, besides its id and label.
This extra info is written in additional columns.
Example: here, each node is characterized by a number, recorded in a column we choose to call "Modularity class".

Columns can contain numbers, text or booleans (true / false).

Helper functions to quickly edit columns
To switch between views of nodes & edges

6 default columns for edges
- **Source and Target:** the two connected nodes forming the edge.
- **Type:** Is the direction of the edge meaningful?
- **Id:** the unique identifier of the edge.
- **Label:** the name of the edge which will be displayed next to it if we choose to.
- **Weight:** how "strong" is the tie between the two nodes forming the edge? This is a numerical value.

**Import function**
Opens a dialog window to import nodes and edges from a csv file into Gephi.

**Extra columns**
Each edge can have extra information, besides its id and label, type and weight.

For example here, I added a column to characterize the connection between the 2 characters of the Miserables: friends or enemies in the novel?

**Helper functions to quickly edit columns**

**The Data Laboratory**
Where the numerical and textual for nodes and edges can be examined and modified.
The Preview panel

Where you make final adjustments before exporting your visualization to an image file (PDF, SVG or PNG)

1. Setting the parameters

Load or save parameters

2. Hit refresh!

After changing a parameter, you must hit "refresh" to see the effects.

3. Export to a picture file format
References


Flight patterns
http://www.aaronkoblin.com/work/flightpatterns/

Scott Weingart on pitfalls of reading networks:
http://www.scottbot.net/HIAL/?p=39600

Force-based algorithms for network layouts:


Y. F. Hu, Efficient and high quality force-directed graph drawing. The Mathematica Journal, 10 (37-71), 2005. (high resolution version (30 MB))