Traitement de l'information

Ophélie Fraisier

2018 – 2019

1 Différence moyenne / médiane : calculez la moyenne et la médiane des 2 variables ci-dessous à l'aide de LibreOffice.

Variable A	23	12	15	12	11	19	17	17	20	14
Variable B	16	17	17	17	18	17	17	18	17	1630

2 Analyses de base à partir d'un fichier JSON

2.1 Téléchargez le fichier JSON à l'adresse suivante : https://data.toulouse-metropole.fr/explore/dataset/elections-regionales-2015-second-tour-ville-de-toulouse-resultats/information/

2.2 Exportez à l'aide de Python les variables `nbr_listea`, `nbr_listeb`, et `nbr_listec` dans un fichier CSV.

Pour rappel, une solution simple lorsque vous souhaitez rapidement stocker des informations dans un fichier est de rediriger la sortie de votre script à l'aide de l'opérateur `>` :

```
python3 mon_script.py > my_file.txt
```

Cette commande indique au système de stocker la sortie du script `mon_script.py` dans le fichier `my_file.txt` au lieu de l'afficher dans le terminal. Attention, si le fichier `my_file.txt` existe déjà, cette commande écrasera son contenu.
Si vous voulez ajouter du contenu à la fin d'un fichier existant, utilisez plutôt l'opérateur `>>` :

```
python3 mon_script.py >> my_file.txt
```

2.3 Calculez la moyenne, médiane, écart-type et variance de chacune de ces variables.

3 Régression linéaire

3.1 À partir des données du fichier JSON de l'exercice précédent, calculez la droite de régression et le coefficient de détermination entre les variables `nombre_dinscrits` et `nombre_dexprimes`.

3.2 Idem pour `nombre_dinscrits` et `nombre_dabstentions`.

4 Nuage de mots

4.1 Récupérez les titres des articles de presse présents sur la page suivante : https://news.google.com

4.2 À l'aide d'un générateur de nuage de mots, réalisez une représentation graphique à partir de ces titres.

Quelques générateurs de nuages de mots que vous pouvez tester :
- `https://www.jasondavies.com/wordcloud/`
- `http://www.wordle.net`
- `https://worditout.com`
- `https://www.wordclouds.com`
- `https://wordart.com`

5 Graphe de co-hashtags

On appelle « co-hashtags » des hashtags apparaissant dans le même tweet. Ils sont souvent intéressant à regarder pour déterminer quels sont les tweets les plus proches en termes de contenus.

5.1 Notez dans un fichier CSV tous les co-hashtags présents dans les tweets ci-dessous (fichier CSV à 2 colonnes, chaque ligne représentant un co-hashtag).

#Buzz : Le top 5 des vidéos tournées à #Toulouse qui ont cartonné sur le #web en 2016
#Toulouse apporte son soutien à la candidature de #Paris2024 en mobilisant les #Toulousains et toutes les forces de vives du territoire !
Jolis habits d'#hiver ce matin sur le site de la météopole à #Toulouse : des #gelées blanches par -5,9°C sous abri !
A l'école Armand Leygue à #Toulouse, #Paris2024 gagne le coeur de la #Generation2024 qui rêve des Jeux #GagnonsEnsemble #RoadToLima
Très #froid en #Europe ce matin : -23°C à #Moscou, -16°C à #Munich, -10°C à #Berne, -5°C à #Toulouse, -3°C à #Madrid mais +9°C à #Lisbonne
Demain, 1er match de l'aventure #mondial2017 contre la Slovénie ! RDV à 19h30 pour entamer le combat ! ! ! #bleuetfier #prepa #toulouse
J-1 #Paris2024 poursuit sa tournée à #Toulouse & vivra #FRASLO ! Tout le pays fête le #sport ! #Handball2017 #GagnonsEnsemble #RoadToLima
Table Ronde "Les acteurs du #spatial au service du @VendeeGlobe", vendredi 06/01 à #Toulouse
Quel est, selon vous, le sportif de #Toulouse qui a marqué l'année 2016 ?Votez sur @cotetoulouse ! #Sport
Automobilistes toulousains, attention au brouillard ce matin #toulouse #trafic #meteo @tlsetrafic

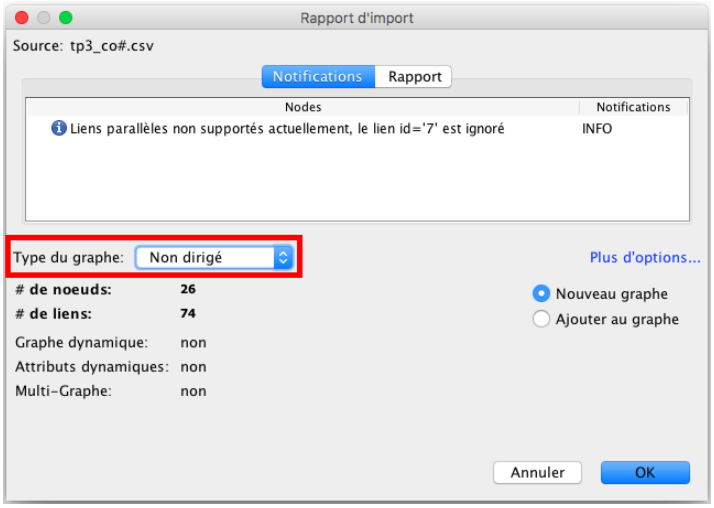
5.2 Télécharger Gephi à l'adresse suivante : https://gephi.org/users/download/

Gephi est un logiciel libre permettant la visualisation et manipulation de graphes.

5.3 Représentez le graphe de co-hashtags à l'aide de Gephi.

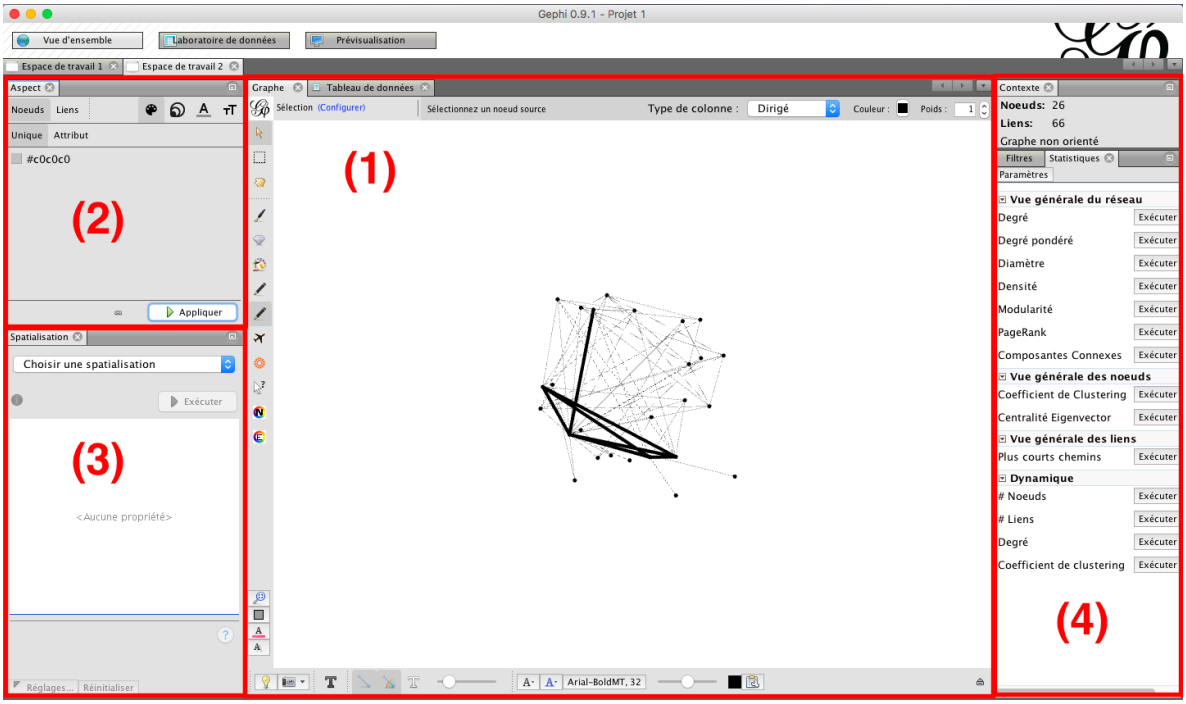
Vous pouvez directement ouvrir le fichier CSV créé précédemment dans Gephi à partir de `Fichier > Ouvrir`.

Précisez `Type de graphe : Non dirigé`



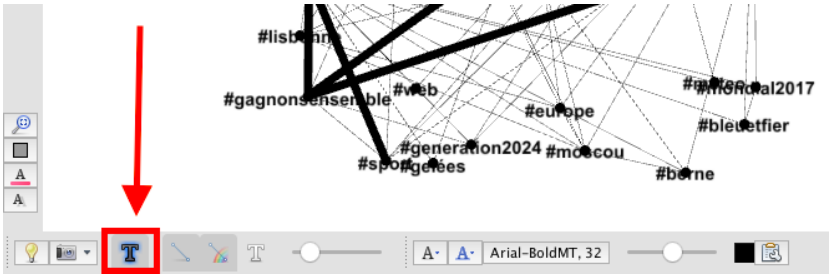
Vous arrivez sur l'interface globale, composée de 4 zones principales :

1. la zone centrale permettant d'afficher le graphe et le tableau de données,
2. le panel en haut à gauche permettant de modifier l'aspect des nœuds, liens et labels,
3. le panel en bas à gauche permettant de modifier la disposition du graphe,
4. le panel à droite permettant de calculer des métriques de graphe et de filtrer les données de manière avancée.



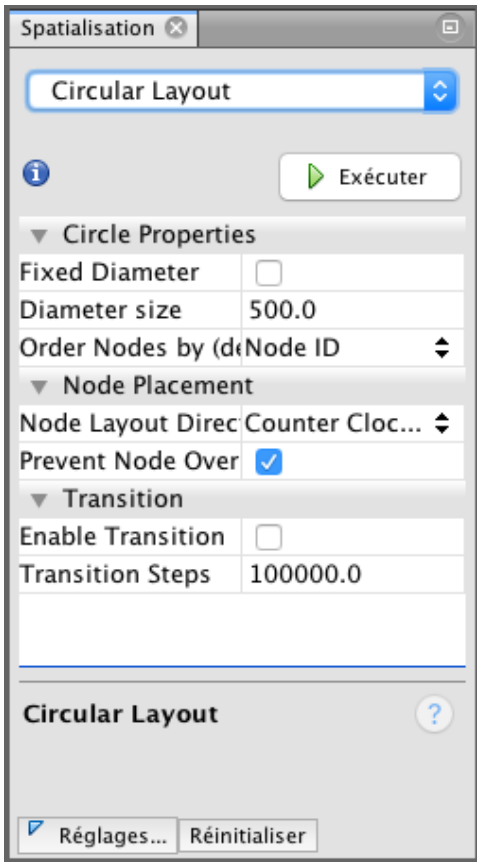
5.3.1 Affichez les labels de sommets.

Le bouton pour afficher les labels des nœuds est placé en bas de la zone d'affichage du graphe.



5.3.2 Changer la disposition du graphe.

Testez différents layouts de disposition de graphe. Si vos nœuds sont trop regroupés pour avoir un graphe lisible, pour pouvez utiliser le layout `Expansion` qui permet d'espacer les nœuds tout en gardant la structure globale.

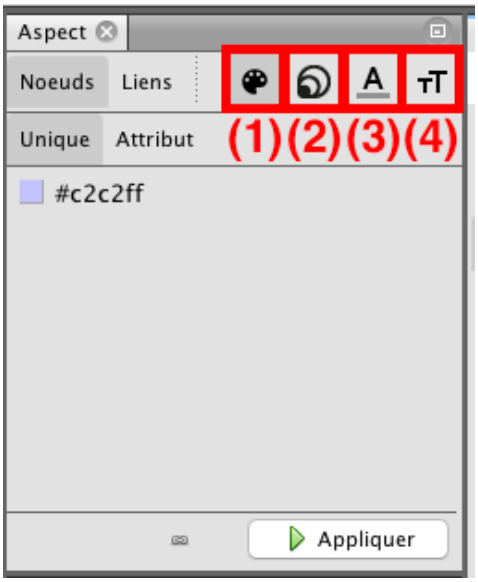


5.3.3 Changez la taille et la couleur des éléments.

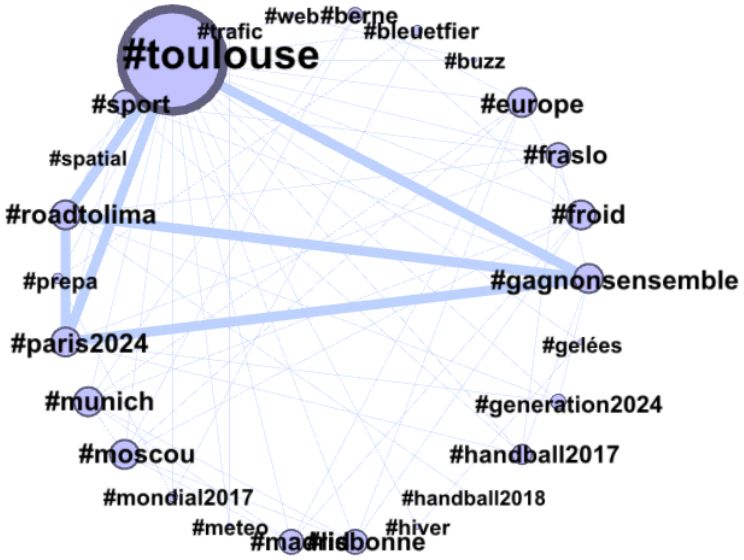
Essayez de changer la taille et la couleur des nœuds, des labels et des liens.

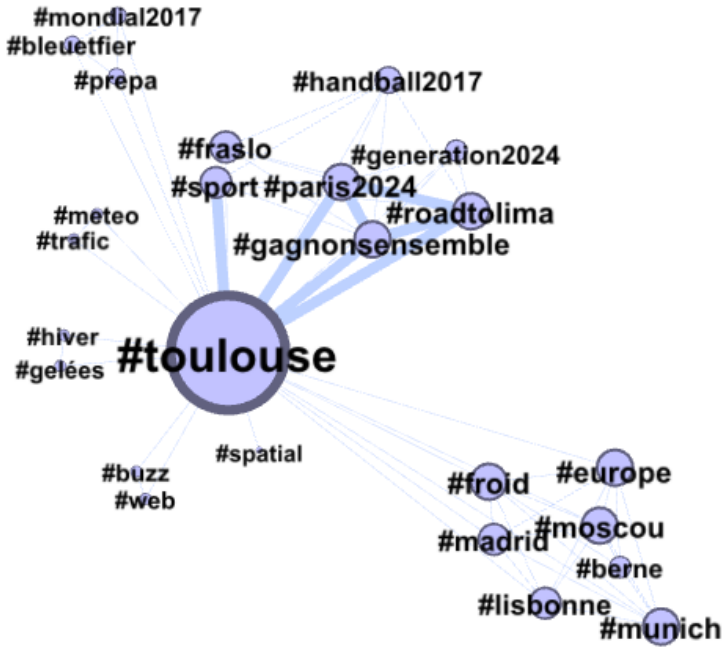
1. Couleur des nœuds ou des liens
2. Taille des nœuds
3. Couleur des labels
4. Taille des labels

Vous pouvez préciser une valeur unique de taille et de couleur par élément, mais vous pouvez aussi sélectionner un attribut.



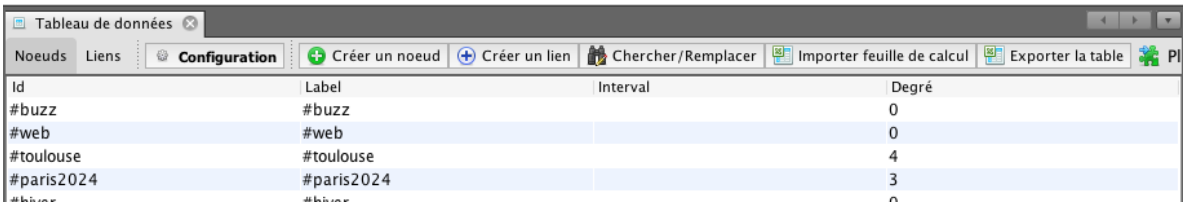
Exemple de graphes organisés à l'aide d'un layout `Circular`, puis à l'aide d'un layout `OpenOrd`, avec la taille des nœuds et des labels de nœuds proportionnelle aux degrés. Ce choix graphique permet de rapidement voir que `#toulouse` est le hashtag le plus fréquent, mais il ne permet pas de repérer facilement les différents groupes de hashtags.

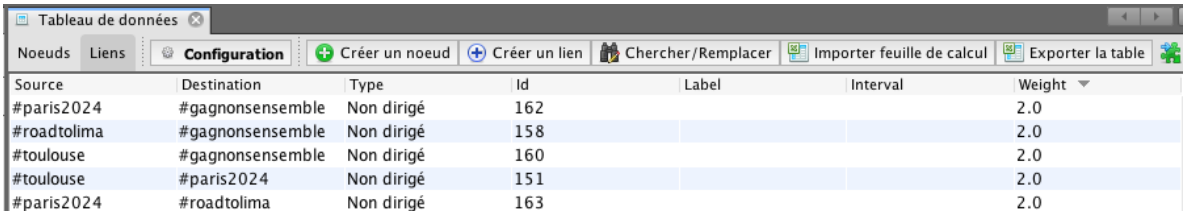




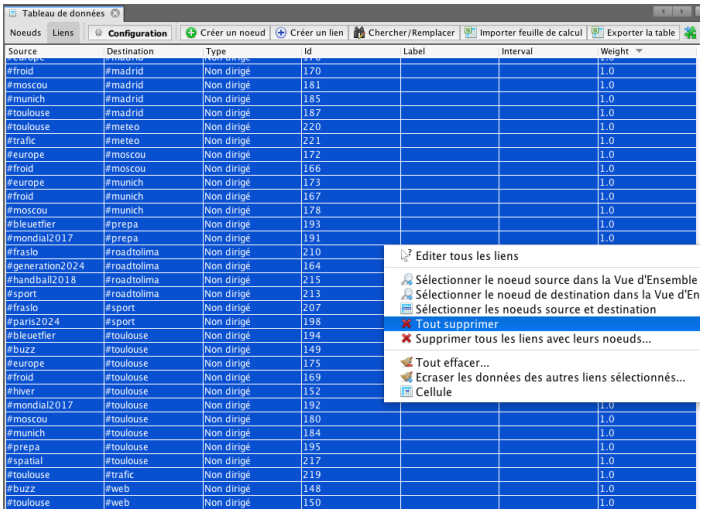
5.3.4 Filtrer les données

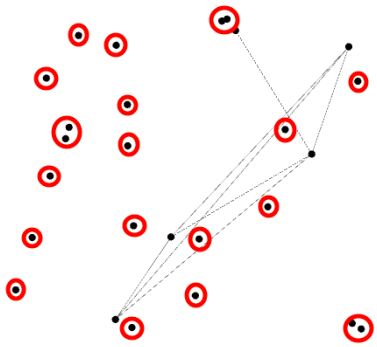
Si vous souhaitez filtrer rapidement les donnés sans passer par les filtres avancés, vous pouvez utilisez le tableau de données.



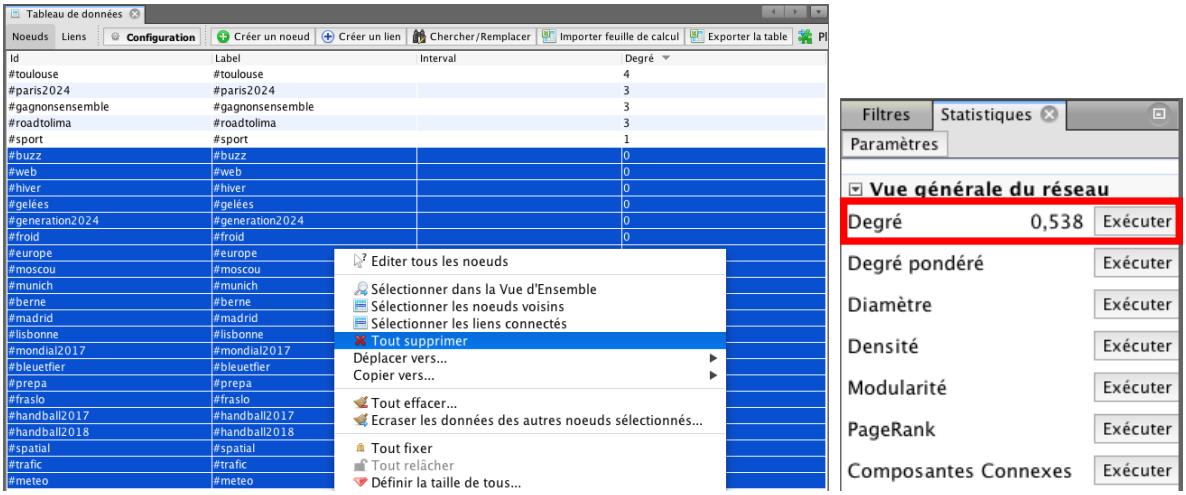


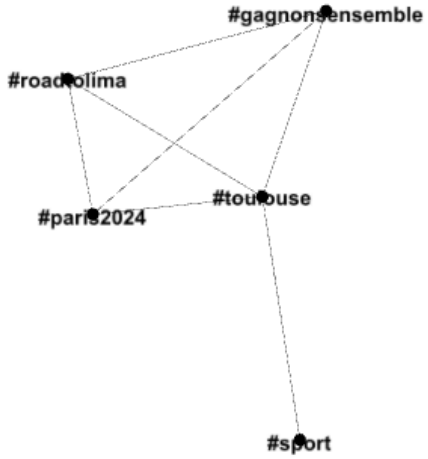
Si vous voulez par exemple supprimer les liens de poids 1, allez dans la vue `Liens` du tableau de données, sélectionner les liens que vous souhaitez supprimer, puis faites `Clic-droit > Supprimer`. Le problème avec cette démarche est que les nœds ne sont pas modifiés, et le graphe résultat contient donc de nombreux nœuds solitaires.





Pour corriger ceci, il faut également supprimer les nœuds de degré 0. Si l'attribut `Degré` n'apparaît pas dans le tableau, il faut le calculer dans le panel des métriques de droite.





5.3.5 Pour aller plus loin

Quelques tutoriels (en anglais) pour aller plus loin avec Gephi :
— Quick start : `https://gephi.org/users/quick-start/`
— Vizualization : `https://gephi.org/users/tutorial-visualization/`
— Layouts : `https://gephi.org/users/tutorial-layouts/`