

## Big Data et Big Graphs : Visualisation

Maxime Lefrançois

De nombreux jeux de données dans le Big Data ont un modèle de graphes. Dans ce TP vous prendrez en main un logiciel de visualisation de graphes nommé Gephi, qui facilite l'analyse et l'interprétation de données sous forme de graphes.

Vous travaillerez en binôme, et m'enverrez votre rapport par email :

- à [maxime.lefrancois@emse.fr](mailto:maxime.lefrancois@emse.fr), avant Vendredi 18, 20:00.
- l'objet de votre mail commencera par [Gephi2016]
- les deux membres du groupe doivent être parmi les destinataires de ce mail.
- en pièce jointe : un répertoire bien organisé et zippé, qui contient le résultat de votre travail : votre rapport, et les captures d'écran de vos visualisation au format SVG, nommées de manière significative, et décrites dans votre rapport.

Slides introductives : <http://www.lamsade.dauphine.fr/~poc/spoc/HamamacheKheddouci.pdf>

### Tutoriel (2h environ)

Suivre le tutoriel à l'adresse : <http://www.martingrandjean.ch/gephi-introduction/>

Cheatsheet:

- interface : [http://www.clementvallois.net/gephi/tuto/en/gephi\\_cheat%20sheets\\_en.pdf](http://www.clementvallois.net/gephi/tuto/en/gephi_cheat%20sheets_en.pdf)
- Layout : <https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>

Attention :

- section 2.1:
  - Gephi fonctionne avec le Java installé sur vos machines ;
  - demander explicitement l'interface en Anglais (important pour s'y retrouver)
  - Avant de chercher à installer un plugin, il faut mettre à jour le système (icone 'planète' en bas à droite de l'écran)
  - Le plugin Noverlap est installé par défaut
  - Le plugin MultiModeNetworksProjection s'appelle maintenant MultimodeNetworksTransformationPlugin

### Twitter

Installez le plugin TwitterStreamingImporter. Prenez-le en main pendant ce TP, et choisissez un cas d'étude que vous avaliserez et interprétez dans votre rapport.

Ce plugin nécessite d'avoir une application Twitter. Créer votre compte Twitter (si ce n'est pas déjà fait) Et créez dans la foulée une application Twitter <https://apps.twitter.com/>.

- (exemple pour le formulaire :

- Name : BigData2016Visulisation-votrenom
- Description : Utilisation de Gephi et du plugin Twitter Streaming Importer pour un TP de Big Data
- Website : <http://www.mines-stetienne.fr/>
- Callback URL : <ne pas remplir>

## Facebook

Voici trois posts qui décrivent comment extraire ses données depuis Facebook à l'aide de l'application Netvizz, et comment les visualiser avec Gephi. Prenez-le en main pendant ce TP, et choisissez un cas d'étude que vous aaliserez et interprétez dans votre rapport.

- <http://allthingsgraphed.com/2014/08/28/facebook-friends-network/>
- <https://msalganik.wordpress.com/2015/02/12/visualizing-your-personal-network-on-facebook/>
- <https://persuasionradio.wordpress.com/2010/05/06/using-netvizz-gephi-to-analyze-a-facebook-network/>

## Evaluation

**Objectif** : Vous choisirez, analyserez et interprétez deux jeux de données de votre choix, et décrirez cela dans trois sections différentes de votre rapport. Présentez chaque section comme si c'était un post de blog (je vous encourage d'ailleurs à aller au bout et publier votre travail sur internet, ce qui pourrait être très bénéfique pour votre CV.)

**Contraintes** : les jeux de données analysés doivent être différents pour chaque groupe.

**Grille d'évaluation** : Montrez-moi que vous avez compris et utilisé les différentes fonctionnalités de Gephi, notamment les layout et les statistiques pour les graphes. Il est bienvenu d'utiliser le plugin MultimodeNetworksTransformationPlugin pour analyser un jeu de données à partir d'une visualisation sous un angle différent.

**Pour les bon développeurs** : On peut importer dans Gephi des données dans de nombreux format. Voir cette page web qui recense différents formats supportés par Gephi, et un exemple de type de données que vous pourriez récupérer sur Internet pour visualiser et interpréter.

<https://github.com/gephi/gephi/wiki/Datasets> . Ceux qui sont bon programmeurs pourraient choisir de développer leur propre jeu de données à partir de jeux de données existants, notamment obtenues à partir des données ouvertes (<http://data.grandlyon.com/> ou <https://www.data.gouv.fr> ), Ou des données du Web des données liées (cf. plugin Gephi SemanticWebImport). La notation tiendra compte du compromis réalisé pour développer vous-même un jeu de données. **Votre « post de blog » doit permettre au lecteur de reproduire votre démarche.**