

Master 1
Bio-Informatique et Génomique
UE Principes de programmation et d'algorithmique



Projet 2016–2017

Hans CERIL

23 décembre 2016

Sommaire

Titre	1
Sommaire	2
1 Introduction	3
2 Extraction des données	3
3 Génération d'un graphe	3
4 Visualisation et description des graphes obtenus	4
4.1 Graphe d'interaction entre les plateformes	4
4.2 Graphe d'interactions des plateformes en fonction des axes	5
4.3 Graphe d'interaction des plateformes en fonction des régions	6
4.4 Graphe d'interaction des plateformes en fonction des villes	7
5 Conclusion	7
Bibliographie	8

1 Introduction

Biogenouest a été créé en 2002, aujourd'hui il fédère sur les deux régions de la Bretagne et des Pays de la Loire. C'est près d'environ 70 unités de recherche que compte cette organisation, il possède un partenariat privilégié avec le pôle recherche de l'enseignement supérieur.

Dans une logique de complémentarité et de mutualisation interrégionale, Biogenouest coordonne 34 plateformes autour de 6 axes technologiques :

- Génomique
- Protéomique
- Exploration fonctionnelle
- Bio-imagerie
- Analyse structurale et métabolique et bio-informatique

Ce projet devrait être normalement réalisé dans le cadre d'un hackathon, il permet d'établir d'une part, les interactions entre les 34 plateformes en fonction des différents paramètres, et d'autre part de représenter et d'analyser ces interactions. Au cours de cette présentation, l'objectif sera de produire une ou plusieurs représentations synthétiques des données mettant en évidence les analyses.

2 Extraction des données

Au cours de ce Hackaton, deux fichiers nous ont été attribués, c.à.d. soit un fichier «meta.txt» et un fichier «interac.txt», chacun de ces fichiers possède des spécificités particulières.

Fichier «interac.txt» : On a un fichier tabulé possédant des informations décrivant les interactions entre plateforme. Trois types d'interactions sont possibles :

- Pas d'interaction
- interaction possible
- interaction existante

Fichier «meta.txt» : fichier tabulé possédant des informations sur les caractéristiques de la plateforme, soit des informations sur :

- les axes
- les régions
- les villes

La difficulté de ce projet était, la lecture des données afin de pouvoir les traiter et la réalisation des différents graphes.

La lecture des données a été réalisée à partir d'un module « CSV » (Comma Separated Values). Le format CSV est un format de fichier courant pour l'échange de données tabulaires entre les applications. C'est un format texte dans lequel les données sont enregistrées ligne par ligne, et où les valeurs sont séparées par une virgule.

3 Génération d'un graphe

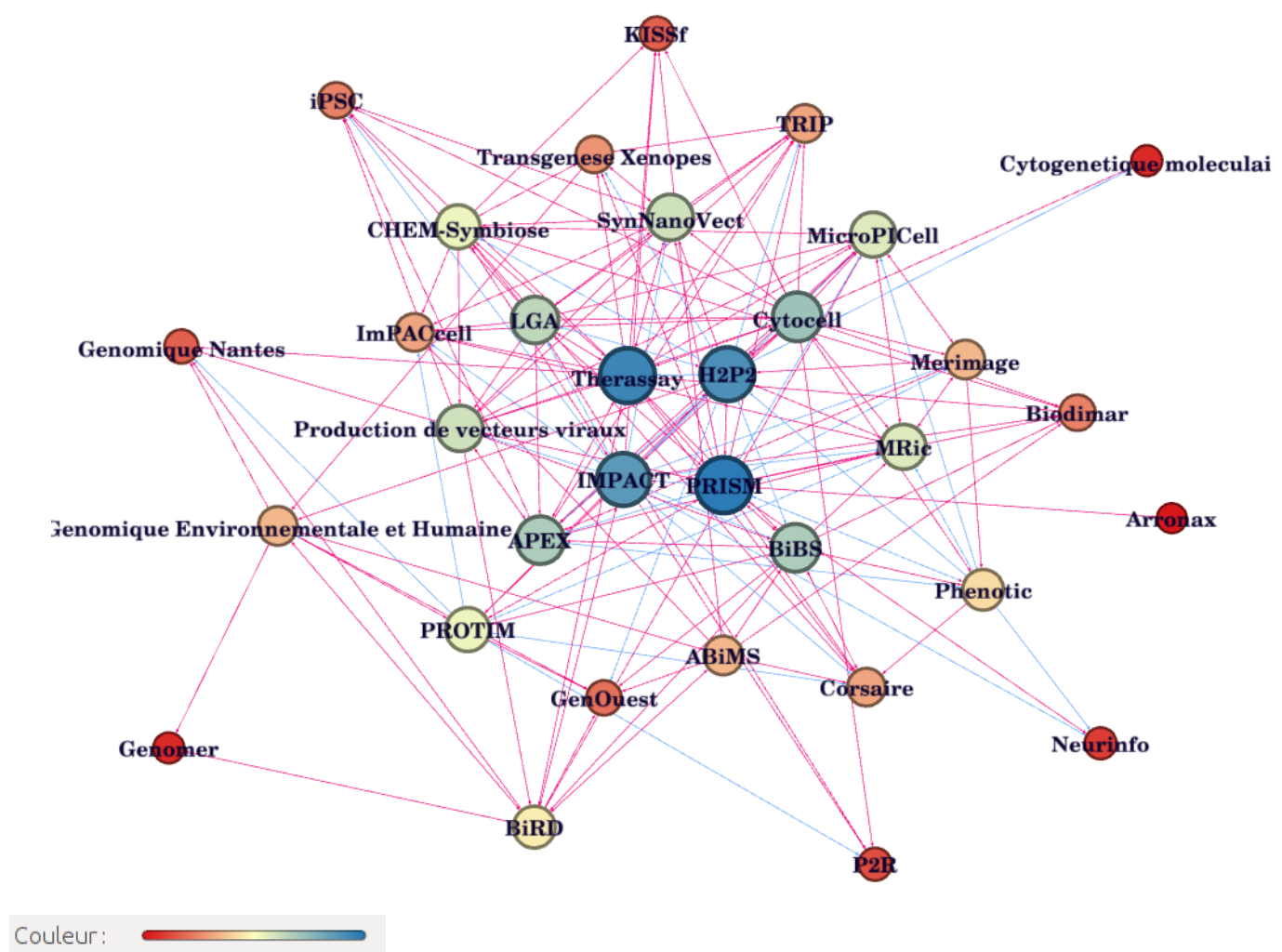
Pour générer les différents graphes, nous avons utilisé la bibliothèque Networkx qui est un logiciel libre, généralement utilisé pour l'étude des graphes et des réseaux. Afin de réaliser ce projet, nous avons écrit un script en python utilisant les commandes du module Networkx, ces commandes ont été exécutées, à l'intérieur de ce même script est apparu un graphe, en suite les résultats graphiques obtenus en format «gexf», ont été exportés pour être utilisés dans un environnement logiciel autre que sera «GEPHI».

Gephi est un logiciel de visualisation et d'analyse des graphes. Ce logiciel est organisé en 3 onglets :

- l'onglet «data laboratory» permet de voir et de modifier les nœuds et les arcs du graphe ainsi que leurs propriétés.
- l'onglet «overview» permet d'effectuer l'analyse et la mise en forme d'une grande partie du graphe (application d'un algorithme de positionnement, couleur et taille des nœuds et des arcs).
- L'onglet «preview» permet de d'améliorer la visualisation et générer une belle image.

4 Visualisation et description des graphes obtenus

4.1 Graphe d'interaction entre les plateformes

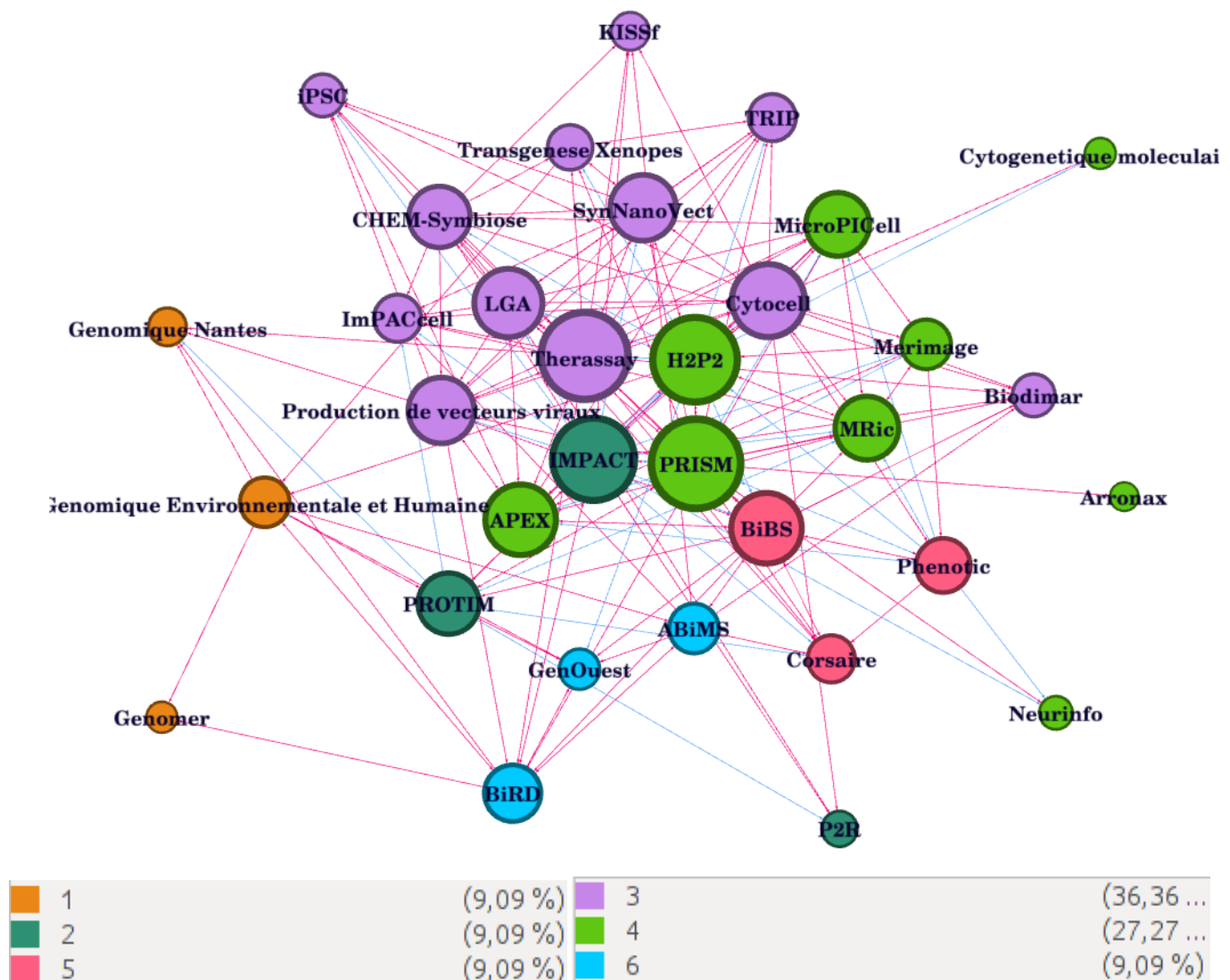


Ce graphe a été réalisé à partir des données du fichier «interact.txt», puisqu'il prend en compte uniquement les différents types d'interactions. L'épaisseur des nœuds est donc fonction du nombre d'interaction que le graphe possède. On a 2 types d'interaction :

- Interaction prouvée : 166 éléments soit 81,77 pourcent des interactions totaux.
- Interaction possible : 37 éléments soit 18,23 pourcent des interactions totaux.

Concernant l'algorithme utilisé pour la spatialisation, il était plus judicieux d'utiliser l'algorithme de «Ajustement des labels», puisqu'il était plus simple à utiliser. La théorie de « centralité » des graphes a trouvé ici toute son utilité. En théorie, les indicateurs centralité des graphes sont des mesures qui sont censées capturer la notion vague d'importance dans un graphe, en identifiant les nœuds les plus significatifs. Ici les nœuds les plus significatifs sont en bleu et se retrouvent au milieu. Ils sont donc considérés comme les plus significatifs du fait qu'ils possèdent plus d'interactions avec les autres plateformes. Tandis que les nœuds rouges sont considérés comme les moins significatifs puisqu'ils possèdent moins d'interactions avec les autres plateformes.

4.2 Graphe d'interactions des plateformes en fonction des axes

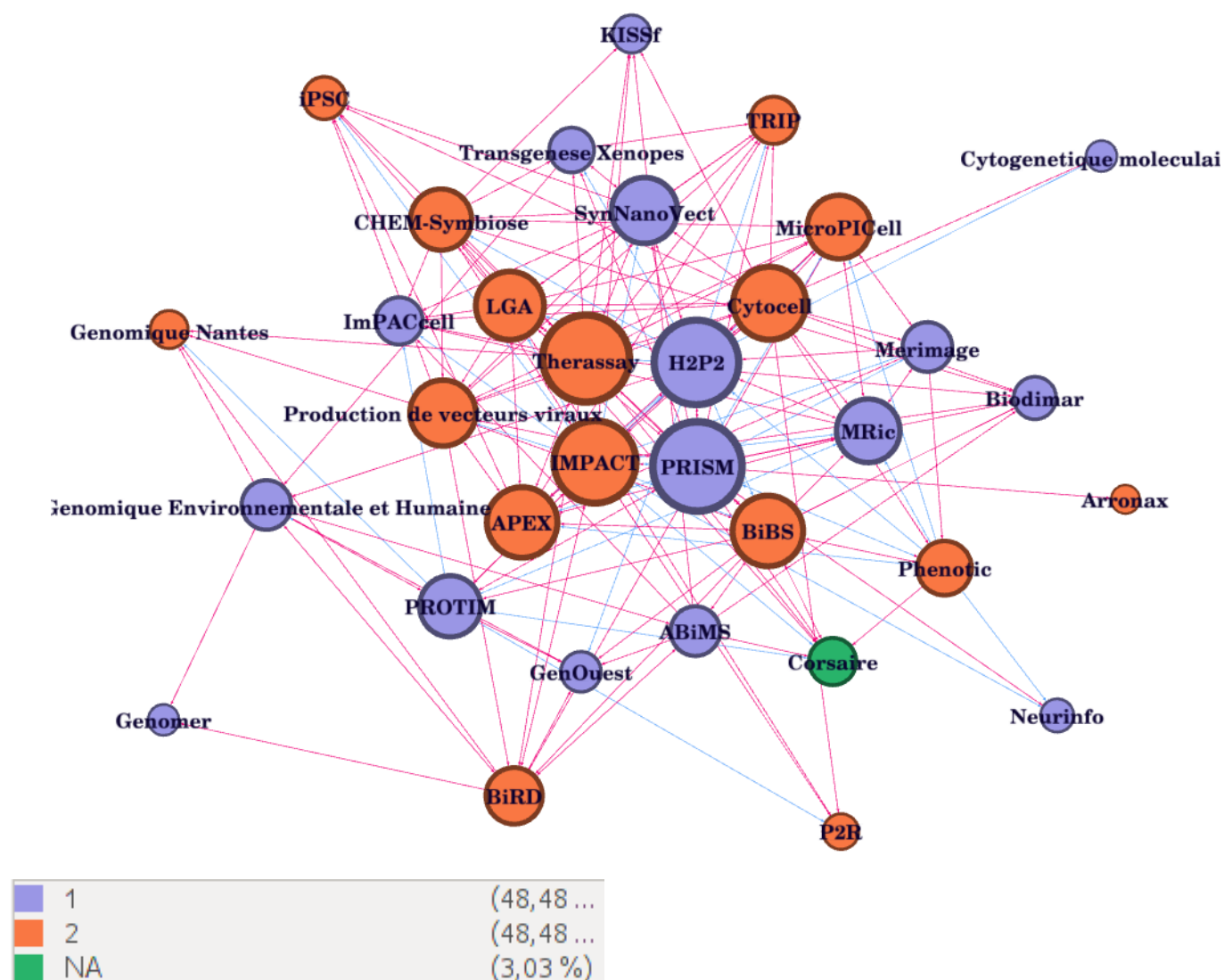


On a plusieurs axes différents correspondant à des couleurs différents. Chaque couleur correspond à des axes spécifiques :

- 1 : Génomique
- 2 : Protéomique

- 3 : Exploration Fonctionnelle
- 4 : Axe Bio-imagerie
- 5 : Axe Analyse structurale et métabolomique
- 6 : Bioinformatique

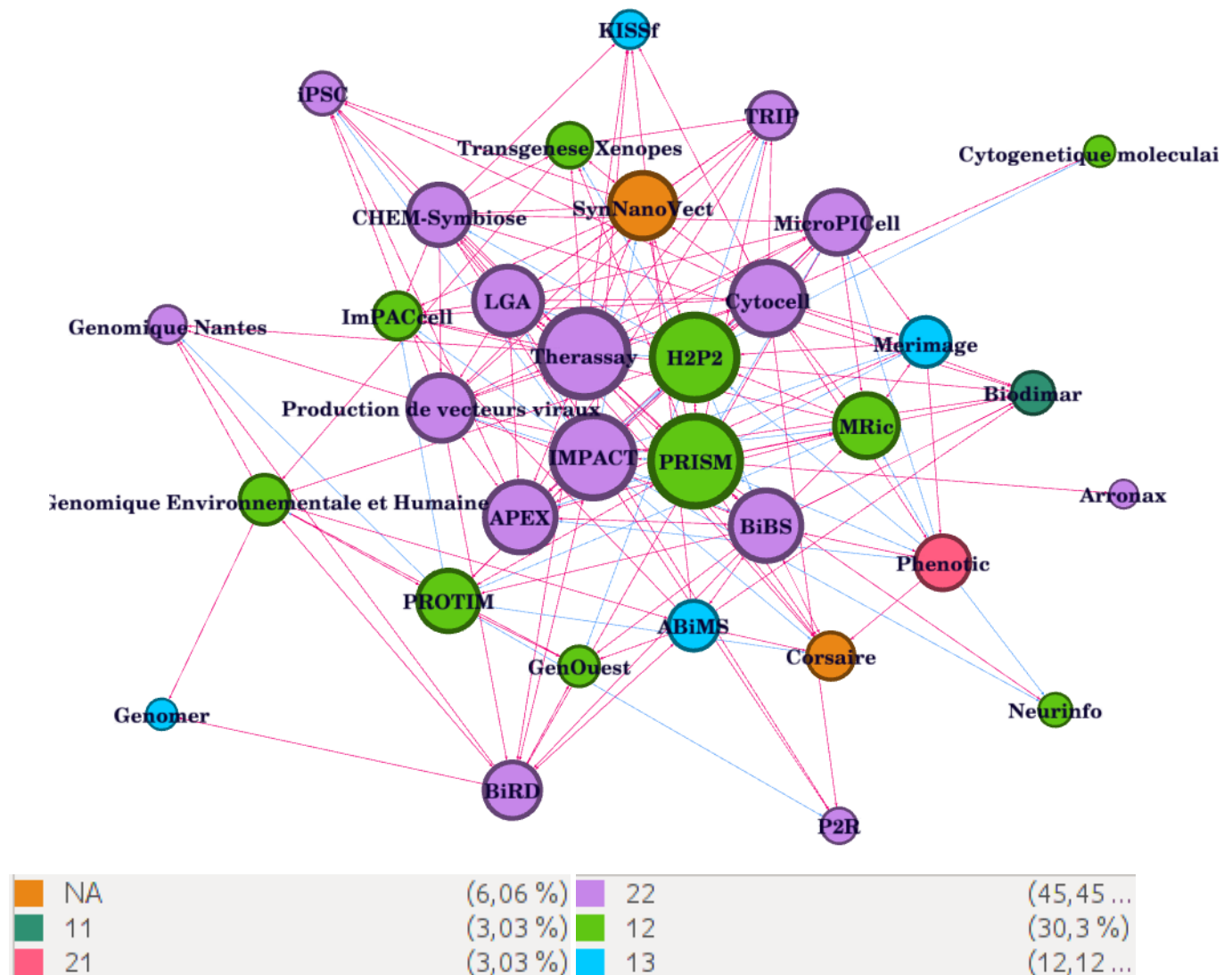
4.3 Graphe d'interaction des plateformes en fonction des régions



Chaque couleur des nœuds est fonction des régions soit :

- 1 : Régions de la Bretagne
- 2 : Région des Pays de la Loire

4.4 Graphe d'interaction des plateformes en fonction des villes



Chaque couleur des nœuds représente une ville particulière :

- 11 : Roscanvel
- 21 : Angers
- 22 : Nantes
- 12 : Rennes
- 13 : Roscoff

5 Conclusion

Ce projet a été pour moi une source de connaissance, car il m'a permis de découvrir de nombreux modules en python tel que le module Csv pour la lecture des données tabulées ainsi que le module Networkx pour la formation de graphes. La grande nouveauté a été l'utilisation de Géphi qui m'a permis de visualiser mes graphes sous différents algorithmes et de les adapter en fonction de mes besoins.

Bibliographie

<https://gephi.org/>

<http://www.biogenouest.org/>

<https://docs.python.org/2/library/csv.html>

<https://gephi.org/>