

*Filière Informatique*

Projet de semestre d’automne

2017 - 2018

VisuDNA-II

Rapport final

*14 janvier 2018*

*Responsables*: P. Kuonen, J. Stopanni, B. Wolf

*Externe*: D. Atlan (Phenosystem SA)

*Auteur*: S. Bouquet

Index

[1 Introduction 4](#__RefHeading___Toc216_3475013932)

[2 Résumé du projet VisuDNA 4](#__RefHeading___Toc274_3475013932)

[2.1 Fonctionnalités 4](#__RefHeading___Toc354_3475013932)

[2.1.1 Cas d’utilisation 4](#__RefHeading___Toc356_3475013932)

[2.1.2 Description 4](#__RefHeading___Toc358_3475013932)

[2.2 État de l’art et choix d’une application 4](#__RefHeading___Toc360_3475013932)

[2.3 Logiciels généraux de traitement de graphe 5](#__RefHeading___Toc362_3475013932)

[2.4 Logiciels spécialisés de traitement de graphe 5](#__RefHeading___Toc364_3475013932)

[2.5 Choix entre *G*ephiet Cytoscape  5](#__RefHeading___Toc366_3475013932)

[2.5.1 Résultat 6](#__RefHeading___Toc368_3475013932)

[2.6 Choix du format de graphe 6](#__RefHeading___Toc370_3475013932)

[3 Analyse 6](#__RefHeading___Toc276_3475013932)

[3.1 Interactome 6](#__RefHeading___Toc381_3475013932)

[3.2 Interactome augmenté 6](#__RefHeading___Toc383_3475013932)

[3.3 Cahier des charges 6](#__RefHeading___Toc278_3475013932)

[3.4 État de l’art 6](#__RefHeading___Toc280_3475013932)

[3.4.1 Logiciels trouvés 7](#__RefHeading___Toc832_3475013932)

[Gephi 7](#__RefHeading___Toc840_3475013932)

[Cytoscape 7](#__RefHeading___Toc842_3475013932)

[Pajek 7](#__RefHeading___Toc844_3475013932)

[Tulip 7](#__RefHeading___Toc846_3475013932)

[VisANT 7](#__RefHeading___Toc848_3475013932)

[CSAT 7](#__RefHeading___Toc850_3475013932)

[GAMA 7](#__RefHeading___Toc852_3475013932)

[3.4.2 Critères de comparaison 7](#__RefHeading___Toc834_3475013932)

[3.4.3 Résultats 9](#__RefHeading___Toc836_3475013932)

[3.4.4 Conclusion 9](#__RefHeading___Toc838_3475013932)

[4 Modélisation 10](#__RefHeading___Toc385_3475013932)

[4.1 Cas d’utilisation 10](#__RefHeading___Toc282_3475013932)

[4.2 Fiches descriptives 10](#__RefHeading___Toc284_3475013932)

[5 Implémentation 10](#__RefHeading___Toc286_3475013932)

[5.1 Module de transformation des nœuds en dictionnaire 10](#__RefHeading___Toc288_3475013932)

[5.2 Fusion des données sample 10](#__RefHeading___Toc290_3475013932)

[5.3 Filtre de chemins 10](#__RefHeading___Toc292_3475013932)

[6 Tests 10](#__RefHeading___Toc294_3475013932)

[7 Conclusion 10](#__RefHeading___Toc296_3475013932)

[8 Annexes 11](#__RefHeading___Toc377_3475013932)

[9 Références 12](#__RefHeading___Toc372_3475013932)

[9.1 Chapitre 2 12](#__RefHeading___Toc387_3475013932)

[9.2 Chapitre 3 12](#__RefHeading___Toc389_3475013932)

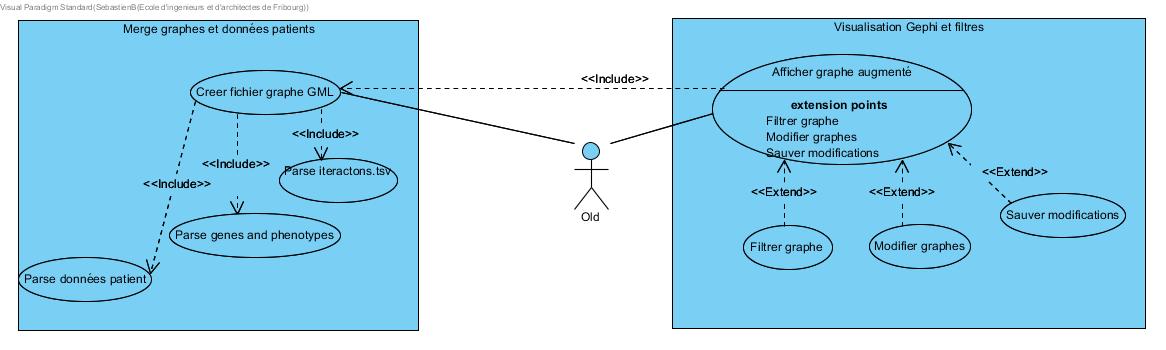
# Introduction

# Résumé du projet VisuDNA

## Fonctionnalités

### Cas d’utilisation

Dans le but de bien comprendre le travail réaliser durant le précédent projet et dû au fait qu’aucun schéma de cas d’utilisation de l’application générale n’était présent dans la documentation, une modélisation du système existant (peut-être approximative) a été réalisée a posteriori.

Figure 1 : schéma de cas d’utilisation

### Description

On s’aperçoit que le projet est composé de deux parties: la première permettant de créer un fichier de graphe au format *GraphML* en associant les données de l’interactome avec les données *sample*. On obtient ainsi un graphe augmenté qui peut être visualiser à l’aide du programme open-source Gephi.

La visualisation des graphes dans Gephi permet certaines options de filtre sur les graphes ainsi que l’ajout de modifications qui peuvent être sauvegardées dans le fichier de graphe.

## État de l’art et choix d’une application

[Sisto(2015)]

## Logiciels généraux de traitement de graphe

L’état de l’art du projet précédent a révélé deux logiciels généraux qui semblaient convenir aux buts du projet car ils permettent le traitement de graphes conséquents et possèdent des plugins de traitement de graphes d’interactome. Il s’agit de Cytoscape  et de Gephi.

A noter: les plugins existants pour Cytoscape  ne permettaient pas l’ajout d’annotation sur les noeuds ni l’utilisation de sources de données autre que sa base de données associée.

## Logiciels spécialisés de traitement de graphe

Exemple de logiciels trouvés:

• EINVis

• Navigator

• IGV

Ces outils ont reçus comme principale critique de ne pas permettre de grande liberté quant à leur utilisation et au développement possible de modifications ce qui les rend donc inintéressants également pour ce nouveau projet.

## Choix entre Gephi et Cytoscape

Ces outils ont les deux été retenus dans un premier temps car ils satisfont (ou du moins satisfaisaient) les critères suivants :

• Traitement de gros graphes [sic].

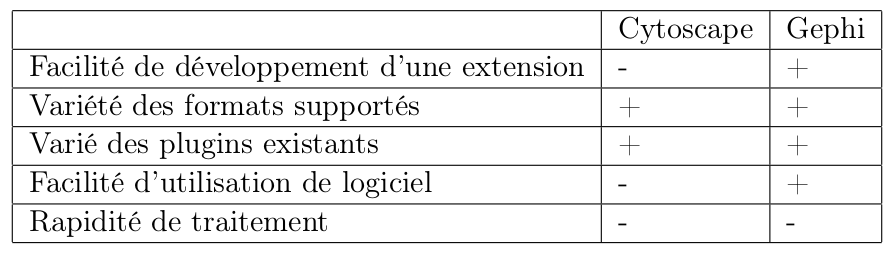
• Possibilité de modifier le layout.

• Recherches personnalisées.

• Annotation personnalisée sur les nœuds et les arrêtes.

• Coloration des nœuds.

### Résultat

Figure 2 : Avantages de *Gephi* et de *Cytoscape*

## Choix du format de graphe

Le projet précédent a retenu le format GML car il semble simple et, surtout, il est supporté par Gephi et Cytoscape . Il possédait par contre le désavantage de ne pas être basé sur du XML ce qui, a priori, n’est pas un désavantage pour ce nouveau projet.

# Analyse

## Interactome

Description d’un interactome et sa représentation sous forme de graphe.

## Interactome augmenté

Description d’un interactome augmenté et sa représentation sous forme de graphe.

## Cahier des charges

Maybe ?

## État de l’art

[Jarukasemratana&Murata(2013)]

### Logiciels trouvés

La phase de recherche a permis de découvrir six outils correspondants aux critères recherchés. Il s’agit de Gephi*,* Cytoscape *,* Pajek *,* Tulip*,* VisANT*,* CSATet GAMA.

#### Gephi

#### Cytoscape

#### Pajek

#### Tulip

#### VisANT

#### CSAT

#### GAMA

### Critères de comparaison

Ces logiciels ont été testés sur les critères suivants :

* **La communauté** : rend compte du nombre de personnes utilisant le logiciel et développant des plugins pour celui-ci.
* **L’interactivité** : détermine comment le graphe est affiché à l’écran. Par exemple si un logiciel anime un changement de layout du graphe au lieu d’afficher une barre de chargement l’interactivité est jugée supérieure.
* **L’intégrabilité**: détermine la possibilité et la facilité avec laquelle on peut intégrer un logiciel ou une partie de celui-ci à l’intérieur de notre application.
* **L’extensibilité** : détermine s’il est possible de modifier le fonctionnement du logiciel, par exemple en développant un plugin pour celui-ci.
* **La documentation** : définit si une documentation claire existe pour l’installation et le développement d’extensions, l’utilisation de l’API, etc.
* **Le support 3D** : détermine si le logiciel est capable d’afficher un graphe en trois dimensions.
* **Les performances** : détermine les performances relatives des logiciels. Ce point est uniquement une estimation des performances après une utilisation des logiciels. Aucun test de benchmark n’a été effectué.
* **Les capacités de filtre** : détermine quels filtres peuvent être appliqués sur un graphe en utilisant des fonctionnalités existantes.

Chaque logiciel a reçu une notre entre 1 et 4 pour chaque critère. Si une information n’a pas pu être trouvée, la note *0* a été attribuée. Certains critères ont été jugés plus important que d’autres, c’est pourquoi une plus forte pondération leur a été attribuée.

### Résultats



### Conclusion

A la suite de cette évaluation, trois logiciels semblent plus adaptés que les autres : il s’agit de Gephi, Cytoscape, Tulip et CSAT. Étant donnée que *Gephi* et *Cytoscape* sont très proches entre eux, nous avons décidé de préférer *Cytoscape* pour sa plus grande communauté d’utilisateur dans le domaine de la bio-informatique. Nous conservons *Tulip* parce qu’il paraît comme le logiciel le plus performant en version desktop et *CSAT* pour les intérêts d’une application web (moins dépendant de la machine utilisée, performances) et parce que ce logiciel à déjà été développer en interne de la HES-SO Fribourg ; son principal défaut, par contre, est le temps nécessaire à obtenir certaines fonctionnalités nécessaires pour ce projet que la plupart des autres logiciel possède déjà (p.ex : le déplacement manuel des nœuds d’un graphe).

# Modélisation

## Cas d’utilisation

## Fiches descriptives

# Implémentation

## Module de transformation des nœuds en dictionnaire

## Fusion des données sample

## Filtre de chemins

*// A penser: documenter l’algorithme mis en place (p.ex pseudo-code)*

# Tests

# Conclusion

# Annexes

* State\_of\_the\_art.pdf : description

# Références

## Chapitre 2

* [Sisto(2015)] M. Sisto, Rapport final (2015).
* [Sisto(2014)] M. Sisto, Cahier des charges v4 (2014).
* Gephi : cf. [*https://gephi.org/*](https://gephi.org/)
* Cytoscape : cf. [*http://www.cytoscape.org/*](http://www.cytoscape.org/)

## Chapitre 3

* [Jarukasemratana&Murata(2013)] *Recent Large Graph Visualization Tools : A Review.* Téléchargé de: <https://www.jstage.jst.go.jp/article/imt/8/4/8_944/_article>.
* [Stoppani(2013)] *Complex Systems Visualization Toolkit.* Téléchargé de: <http://garetjax.info/csat-report-v1.0.pdf>.
* Pajek : <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
* Tulip : <http://tulip.labri.fr/TulipDrupal/>
* VisANT : <http://visant.bu.edu/>
* CSAT : <http://stoppani-1.tic.hefr.ch/>
* GAMA : <http://gama-platform.org/>