Sébastien Bouquet

VisuDNA-II

Projet de semestre d’automne 2017-2018

octobre 18, 2017

**Outils de visualisation de graphes: état de l’art**

# Abstract

Ce rapport a pour but de rechercher le meilleur outil existent de visualisation de graphes, plus spécifiquement dans le domaine de la bio-informatique et de la visualisation de génomes.

# La phase de recherche a permis de découvrir six outils correspondant aux critères recherchés. Il s’agit de Gephi, Cytoscapte, Pajek, Tulip, VisANT, et le travail de J. Stoppani. Ces logiciels ont été testés sur les critères suivants :

**La communauté :** rend compte du nombre de personnes utilisant ce logiciel.

**L’interactivité :** détermine comment le graph 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’intégrabilité :** détermine la possibilité et la facilité avec laquelle on pourrait 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 plug-in.

**La documentation :** définit si une documentation claire existe pour l’installation et le développement d’extension, l’utilisation de l’API, etc.

**Le support 3D :** détermine si le logiciel est capable d’afficher un graphe en trois dimensions.

**Les performances :** détermines 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.

Chaque logiciel a reçu une notre entre 1 et 4 pour chaque critère définit. Si une information n’a pas pu être trouvée, la note ‘0’ est attribuée.

# Pajek

Logiciel développé pour Windows en C. Il permet de gérer des graphes dans l’ordre de grandeur de 500'000 nœuds. Une version spéciale du programme : Pajek-XXL existe permettant l’affichage de réseau jusqu’à 2'000'000 de nœuds et diminuant la capacité de RAM nécessaire. Puis une troisième version Pajek-3XL qui augmente le numéro des nœuds et peut gérer des graphes jusqu’à 109 nœuds.

L’interactivité avec le graphe et les options de visualisation sont par contres limitées. Ce programme utilise son propre format de fichier de graphe (NET) et permet l’exportation en fichier GraphML.

# Gephi

Programme Java, permet le développement de plug-ins. Un des meilleurs programmes testés en ce qui concerne l’interactivité et l’aspect visuel des graphes. Un autre point positif est le support de nombreux format de fichier de graphe notamment : GML, GraphML, GDF, CSV, Tulip TLP, Pajek NET, GraphViz DOT.

# Cytoscapte

Programme Java, possède de nombreuses fonctions de filtres et de lay-out pour les graphes. L’interactivité est quelque peu limitée mais l’aspect visuel des graphes est réussi. Possibilité d’ajouter des plug-ins et d’intégrer le programme en utilisant leur API. Un plug-in existe pour permettre l’utilisation de NVidia CUDA ce qui peut améliorer grandement la vitesse de calcul d’après le GPU utilisé. Ce logiciel supporte également de nombreux formats de graphe dont : SIF, NNF, GML, BioPAX, GraphML, JSON, CSV.

# Tulip

Programme en C++ utilisant la librairie OpenGL ; développer à l’origine pour Windows, il existe à présent des versions Mac et Linux (à compiler). Supporte la visualisation 3D des graphes et des réseaux jusqu’à 2'000'000 de nœuds sur un ordinateur de bureau moderne. L’aspect visuel des graphes ainsi que l’interactivité sont bons. Il supporte par contre un seul format de graphe (Tulip TLP).

# VisANT

Programme développé en Java spécifiquement pour la bio-informatique. Peut être utilisé comme application local ou comme applet Java. Base de données accessible contenant déjà des graphes. Il supporte neuf formats de fichier de graphe dont : GML, BioPAX, KGML, VisML.

# CSAT

Cette application a été plus compliquée à évaluer étant donné que contrairement aux autres logiciels, il s’agit d’une application web. Ce qui comporte également des avantages comme le fait de pouvoir améliorer la puissance de calcul en faisant du calcul parallèle. Ou de stocker les transformations effectuées sur un graphe sur serveur. GraphML

# Résultats

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Criteria** | Community | Interactivity | Integrability | Extensability | Documentation | 3D support | Performaces (estimation) | Filters capabilities | Additionnal information | Total |
| **Importance** | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |  |  |
| Gephi | 3 | 4 | 1 | 3 | 4 | 2 | 3 | 3 |  | 38 |
| Cytoscape | 4 | 2 | 2 | 4 | 3 | 2 | 3 | 3 |  | 38 |
| Pajek | 3 | 2 | 1 | 2 | 2 | 3 | 4 | 4 | Pas d'annoation sur les nœuds/edges | 31 |
| Tulip | 2 | 3 | 3 | 4 | 3 | 3 | 4 | 3 | TLB, GML, CSV exports | 38 |
| J. Stoppani | 1 | 3 | 4 | 3 | 3 | 3 | 3 | 4 |  | 38 |
| VisANT | 2 | 3 | 3 | 2 | 3 | 1 | 3 | 3 | Version Java applet (seems complicated & java desktop | 33 |

# Analyse

A la suite de cette évaluation, trois logiciels semblent plus adaptés que les autres : il s’agit de Gephi, Cytoscape et du travail de J. Stoppani. On remarque que Tulip et VisAnt restent également des options envisageables.

# Références

Jarukasemratana, S & Murata, T (2013). *Recent Large Graph Visualization Tools : A Review.* Retrieved from <https://www.jstage.jst.go.jp/article/imt/8/4/8_944/_article>.

Stoppani, J (2013). *Complex Visualization Toolkit.* Retrieved from <http://garetjax.info/csat-report-v1.0.pdf>.