



HAL
open science

Paohvis, un outil pour l'exploration visuelle d'hypergraphes dynamiques

Laurent Beauguitte

► **To cite this version:**

Laurent Beauguitte. Paohvis, un outil pour l'exploration visuelle d'hypergraphes dynamiques. 2020.
halshs-02570627

HAL Id: halshs-02570627

<https://shs.hal.science/halshs-02570627v1>

Preprint submitted on 12 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Paohvis, un outil pour l'exploration visuelle d'hypergraphes dynamiques

Paohvis, a tool for dynamic hypergraphs visual exploration

Laurent Beauguitte, UMR Géographie-cités - Mai 2020

Résumé

Cet article présente le logiciel *Paohvis* destiné à l'exploration visuelle des hypergraphes dynamiques. Il présente ses principales fonctionnalités, intérêts et limites.

Abstract

This paper presents the software *Paohvis* designed to visually explore dynamic hypergraphs. It presents its main functionalities, interests and drawbacks.

Mots clés

Paohvis ; Visualisation ; Hypergraphe ; Dynamique

Keywords

Paohvis ; Visualization ; Hypergraph ; Dynamic

Introduction

Les outils disponibles pour analyser et/ou visualiser des données relationnelles n'ont jamais été aussi nombreux, qu'ils s'agissent d'applications en ligne, de logiciels à installer localement ou de modules s'insérant dans des écosystèmes logiciels plus vastes (modules disponibles dans R par exemple). Mais cette abondance d'outils va de pair avec une certaine uniformité des traitements et surtout des représentations graphiques proposées, la représentation dite nœud-lien étant souvent la seule disponible.

Dans ce contexte, l'application¹ *Paohvis* développée depuis 2018 par l'équipe *Aviz* de l'INRIA mérite l'attention dans la mesure où elle présente des options originales, tant en ce qui concerne la modélisation des données (hypergraphe dynamique) que leur visualisation (*biofabric*). Après avoir rappelé ce que désignent ces deux termes, l'article présente l'outil, ses principales fonctionnalités, les données qu'il permet d'analyser et enfin

¹Dans ce texte, les termes logiciel et outil sont considérés comme synonymes ; le terme plateforme désigne l'application disponible en ligne à l'adresse <http://www.di.uniba.it/~buono/paohvis/paoh.html>. Seule la plateforme en ligne a été testée pour l'écriture de cet article.

ses intérêts et limites². Pour approfondir, le lectorat est invité à consulter les articles des personnes ayant développé l’outil (Valdivia *et al.*, 2018 et 2019) ainsi que le site web correspondant (<https://www.aviz.fr/Research/paohvis>).

1. Deux options originales : hypergraphe et *biofabric*

Si l’équipe *Aviz* travaille depuis des années sur la visualisation de données relationnelles, les outils développés à ce jour³ privilégiaient les représentations matricielles ou mixtes et la modélisation des données sous forme de graphe. L’outil *Paohvis* suppose une modélisation des données sous forme d’hypergraphe et recourt à la visualisation *biofabric*.

1.1. Un format délaissé : l’hypergraphe

Dans un graphe simple (ni boucle ni liens multiples), un lien relie nécessairement deux sommets ; dans un hypergraphe comprenant n sommets, un hyperlien comprend de 1 à n sommets. Cette catégorie était considérée par le mathématicien Claude Berge comme un moyen de généraliser et de simplifier les concepts utilisés en théorie des graphes (Berge, 1973, pp. viii-ix⁴). Or, elle a été peu reprise par la suite : on peut noter à titre d’exemple que le terme même d’hypergraphe est absent d’un manuel récent de théorie des graphes (Wallis, 2007) ou que la volumineuse encyclopédie d’analyse de réseaux sociaux parue en 2014 ne comprend pas d’entrée « Hypergraph » (Alhadjj & Rokne, 2014). Or, plusieurs types de données couramment mobilisées en analyse de réseaux gagneraient à être considérés comme des hypergraphes – nous y reviendrons lorsque nous aborderons le format de données nécessaire à *Paohvis*.

1.2 La visualisation *biofabric*

Dans un article paru en 2012, William Longabaugh propose un format de visualisation qu’il présente comme particulièrement adapté pour les grands réseaux. L’outil qu’il présente, nommé *BioFabric*, place les sommets en ligne et les liens en colonne : chaque ligne correspond à un sommet, chaque lien est représenté par une ligne verticale reliant deux sommets. Si l’auteur avance dans son article différents avantages par rapport à la visualisation lien-nœud et matricielle, ce type de visualisation reste peu employé aujourd’hui, notamment parce que rares sont les logiciels proposant cette option.

² Les personnes testant l’outil sont vivement incitées à transmettre leurs remarques à l’équipe *Aviz* (paohvis@inria.fr).

³ Voir la page <https://www.aviz.fr/Research/Projects>.

⁴ Justifiant l’intérêt de l’hypergraphe dans la préface de l’édition américaine de son ouvrage, il écrit : « by using hypergraphs, one deals with generalizations of familiar concepts. Thus, hypergraphs can be used to simplify as well as to generalize ».

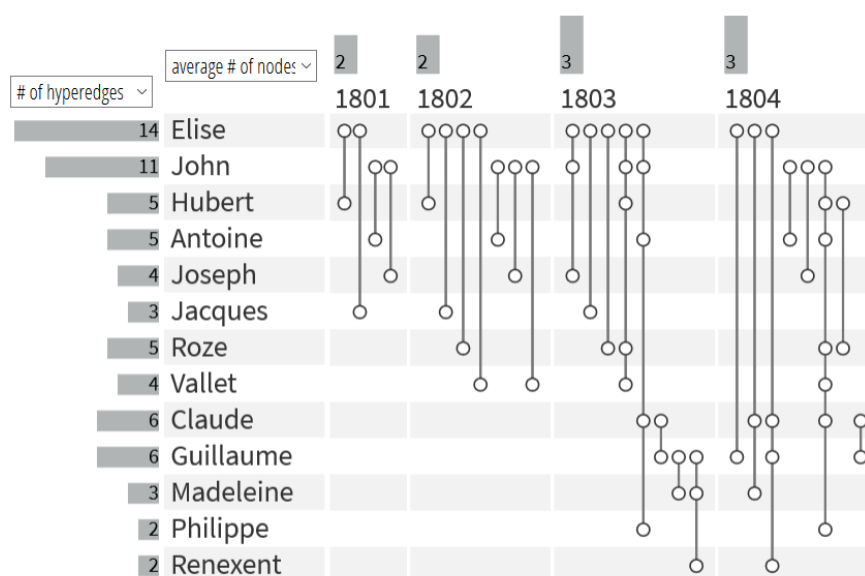
2. Paohvis : interface, données et manipulations

Lorsqu'on accède à l'application, le jeu de données affiché concerne des contrats marchands : chaque contrat, daté, est considéré comme un hyperlien, les personnes contractantes étant les sommets de l'hyperlien.

2.1. Interface et menu

L'interface de *Paohvis* comprend une fenêtre graphique occupant la majeure partie de l'écran surmontée d'une barre d'outils. Le nom du jeu de données représenté est indiqué en haut à gauche de la barre d'outils. La plupart des options proposées portent des noms explicites : charger un jeu de données (*Add new dataset*) ; adapter la hauteur de la fenêtre graphique au nombre de sommets (*Fit nodes*) ou la largeur au nombre de périodes temporelles (*time slot width*) ; restreindre la période temporelle étudiée (*go to time slot*).

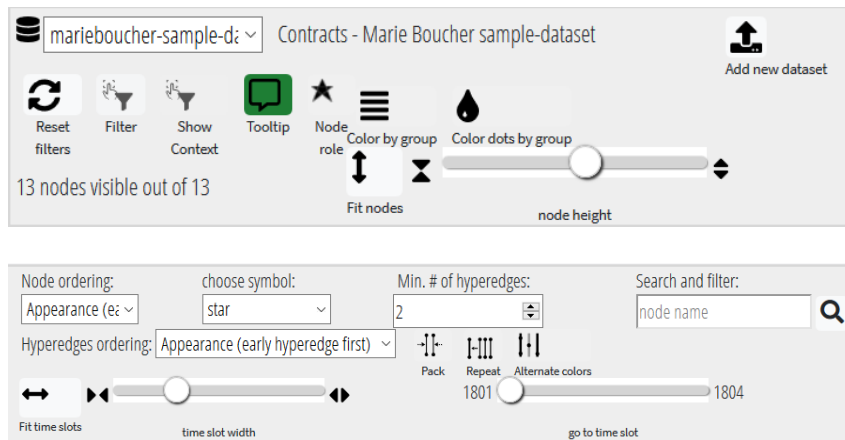
FIGURE 1 – Fenêtre graphique de *Paohvis*



Les sommets sont disposés en ligne ; les hyperliens en colonnes et classés par ordre chronologique. Cliquer sur un sommet, un hyperlien une année permet de mettre en évidence les hyperliens et sommets concernés.

Il est possible de filtrer les sommets affichés, de colorer les sommets en fonction de leur rôle (*Node role*) ou d'une variable catégorielle (*Color by group*). Il est également possible d'afficher un attribut de l'hyperlien (*Tooltip*). L'ordre d'affichage des sommets et des hyperliens peut être modifié (*Node ordering* et *Hyperedges ordering*) tout comme le symbole utilisé pour les sommets (*choose symbol*).

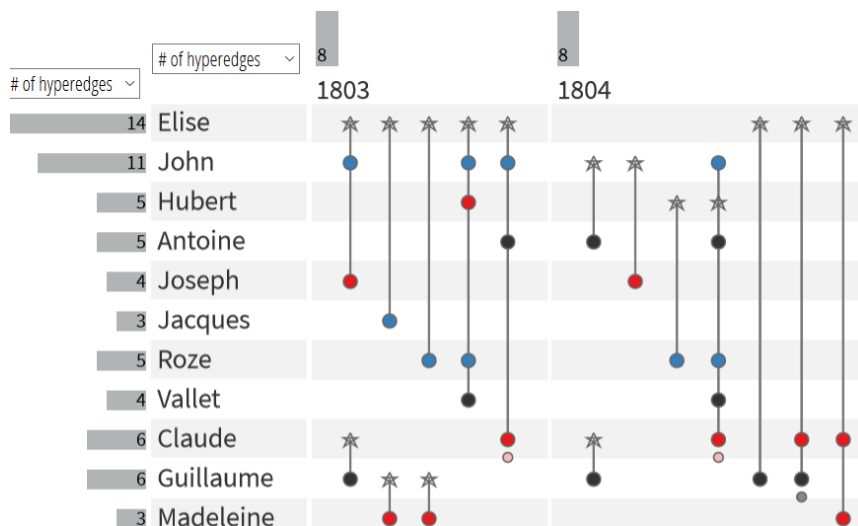
FIGURE 2 – Barre d’outils de *Paohvis*



L’option *Pack* réordonne les hyperliens afin de produire une figure plus compacte : l’option *Repeat* permet d’agréger les hyperliens comprenant les mêmes sommets à la même date (le lien représenté est alors plus épais).

Filtrer les sommets en fonction du degré est une option classique pour produire des figures lisibles, cela est possible ici avec la commande *Min. # of hyperedges*. Mais, afin que l’on n’oublie pas que des sommets non représentés après filtrage existent dans les données d’origine, des cercles plus petits apparaissent entre les lignes (ce que l’équipe d’*Aviz* a nommé des *trips*).

FIGURE 3 – Sommets centraux et *drips*



Sur cette figure, seuls les sommets (colorés en fonction de leur catégorie et ayant un signe différent selon leur rôle) présents dans plus de deux hyperliens sont représentés. Les cercles de petite taille

situés entre les lignes correspondent à des acteurs présents dans les données mais ne satisfaisant pas ce critère.

2.2 Format de données et champs d'application

L'un des obstacles principaux à l'adoption d'un nouvel outil est le temps parfois long nécessaire à la mise en forme des données afin de satisfaire aux spécifications logicielles de l'outil. Dans le cas de *Paohvis*, le format de données demande *a priori* un investissement léger. Le fichier doit être en .csv, séparateur virgule, encodé en utf-8 et sans nom de colonne.

Les trois premières colonnes doivent contenir les informations suivantes :

- identifiant (numérique ou chaîne de caractères) de l'hyperlien ;
- nom du sommet (chaîne de caractères) ;
- date (numérique ou chaîne de caractères).

La quatrième colonne, optionnelle, peut contenir une information sur l'hyperlien (chaîne de caractères) ; la cinquième, optionnelle également, une variable catégorielle portant sur le sommet.

Le fait de devoir répéter l'information sur le sommet à chaque ligne permet de gérer les changements de catégorie possibles d'une période à une autre. Si les hyperliens correspondent à des articles scientifiques et les sommets à des auteur.e.s, il est aisé de gérer les mobilités institutionnelles.

Si l'équipe d'*Aviz* donne comme exemples des documents datés (contrats marchands, articles scientifiques), le terme document peut être compris dans un sens très large et une grande partie des données représentées à l'aide de graphes bipartis pourraient sans doute être modélisées sous forme d'hypergraphes dynamiques. Je donnerai un seul exemple : celui des *featurings* entre artistes rap. On peut bien sûr construire un graphe où un lien orienté entre l'artiste *a* et *b* indique que le premier a invité le second sur un morceau. Garder une perspective dynamique est possible en indiquant un attribut date au lien créé. Mais deux problèmes apparaissent : celui de la représentation visuelle du caractère dynamique de ce graphe et surtout celui des *featurings* impliquant plusieurs artistes. Si l'artiste *a* invite *b* et *c* sur un morceau, par construction aucun lien n'apparaît entre *b* et *c* alors que tous deux sont présents sur le même morceau. Si l'on considère les artistes comme des sommets et les morceaux comme des hyperliens, les deux problèmes sont résolus. Bien d'autres données (participation à des colloques, composition de conseils d'administration, présence de personnages dans des épisodes de séries, etc.) gagneraient sans doute à être modélisées ainsi.

3. Bilan conclusif

L'outil *Paohvis* est facile à prendre en main et la mise en forme des données peu coûteuse ; les visualisations produites sont lisibles et permettent d'explorer aisément ses

données. Le recours à l'hypergraphe permet une perte d'information moindre que d'autres choix de modélisation des données. Signalons également la qualité de la documentation (en anglais exclusivement). Deux réserves tout de même (qui ont été transmises à l'équipe développant l'outil) : il est dommage que le volet visualisation ne soit pas adossé à un volet analyse statistique, trop peu de mesures sont disponibles ici alors qu'elles pourraient permettre de compléter l'exploration visuelle. Enfin, ne pas pouvoir sauvegarder les visualisations produites en format vectoriel est un peu regrettable.

Références

- Alhajj R. & Rokne J. (dir.) (2014). *Encyclopedia of Social Network Analysis and Mining*. New-York : Springer.
- Berge, C. (1973). *Graphs and Hypergraphs*. New-York : Elsevier.
- Longabaugh, W. J. (2012). Combing the hairball with BioFabric: a new approach for visualization of large networks. *BMC bioinformatics*, 13(1), 275.
- Valdivia, P., Buono, P., Plaisant, C., Dufournaud, N. and Fekete, J-D. (2019). [Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization](#). *IEEE Transactions on Visualization and Computer Graphics*.
- Valdivia, P., Buono, P., Plaisant, C., Fekete, J-D. and Dufournaud, N. (2018), [Using Dynamic Hypergraphs to Reveal the Evolution of the Business Network of a 17th Century French Woman Merchant](#). *Proceeding of the VIS4DH Workshop*, (pp. 1-5).
- Wallis W.D. (2007). *A Beginner's Guide to Graph Theory*. Boston : Birkäuser.