

## Le projet Graph ton Anthologie Grecque (GAG)

# Restitution d'une expérience de Hackathon

Antoine Fauchié, Marianne Reboul, Enzo Poggio, Lena Krause, Margot Mellet

Publié le 25-11-2024

http://sens-public.org/articles/1728



#### Résumé

Cet article restitue une expérience de travail collaboratif mené dans le cadre d'un hackathon (lors des journées d'études Navigations anthologiques de la CRCEN à l'Université de Montréal) en présentant les membres du groupe de travail, en détaillant les processus de réflexion et les étapes de réalisation. Les différents participants du groupe de travail ont procédé à l'exploration de l'Anthologie grecque pour produire des visualisations des données de ce corpus nommés graphes. Ces graphes sont autant la structuration de différents niveaux de perspectives pour représenter l'Anthologie que le reflet de la diversité des profils et expertises des membres du groupe de travail.

#### **Abstract**

This article presents an experience of collaborative work during a hackathon (at the CRCEN's *Navigations anthologiques* conference at the Université de Montréal), introducing members of the working group and detailing the discussion and development phases. The members of the working group explored the Greek Anthology to produce visualisations of the corpus' data, called *graphs*. These graphs represent both the different levels of perspective used to map the Anthology and the diversity of the profiles and expertise of the working group members.

Mot-clés: graphique, écriture collaborative, humanités numériques Keywords: graph, collaborative writing, digital humanities

## Table des matières

Présentation de l'équipe
Antoine
Enzo
Lena
Margot
Marianne
Chauffe intellectuelle
Dialogues entre idéation et réalisation
L'Anthologie comme corpus
Idéation entre dessin et design
L'API 21
Organisation synergique
Restructuration des données
L'espace visuel
Documentation en synchrone
Fin, renouveau, et documentation posthume
Bibliographie

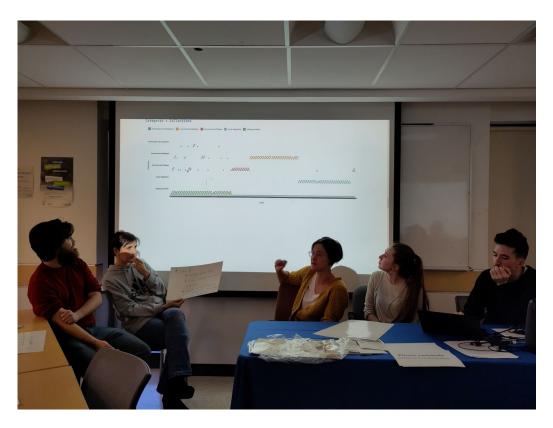
# Le projet Graph ton Anthologie Grecque (GAG)

Antoine Fauchié Marianne Reboul Enzo Poggio Lena Krause Margot Mellet

Cet article procède pour une grande part d'un travail de restitution d'un projet collaboratif mené lors du Hackathon des journées d'études des 27 au 29 octobre 2022 Navigations anthologiques (CRCEN 2022). Ce Hackathon avait pour objet le développement de prototypes faisant usage de l'API de l'Anthologie Grecque, et l'exploration d'éditorialisation de quelques centaines d'épigrammes et leurs multiples versions. Par oubli d'étapes, égarements d'archives, ou nostalgies d'un collectif, cette restitution comporte une dimension de romancisation qui ne demeure pas moins fidèle à l'expérience passée. Nous présentons ici le fruit d'un travail collaboratif, tout d'abord en décrivant le profil de ses membres, en expliquant les motivations éclectiques puis en explicitant les différentes étapes de la journée. Cet article constitue le compte rendu d'une exploration de quelques heures dans la manipulation de données.

## Présentation de l'équipe

L'équipe est constituée de plusieurs profils, capables d'intervenir sur des dimensions variées d'un projet en humanités numériques : définition d'un cadre théorique, contextualisation d'un projet, formulation des besoins, énumération des options techniques, réalisation de prototypes, rédaction d'une documentation et présentation de résultats. Voici ces profils détaillés.



 $\label{eq:figure 1} \textbf{Figure 1} - \textbf{Pr\'esentation de Graphe ton Anthologie Grecque}$ 

#### Antoine



Figure 2 - Antoine

Coordonnateur de projets à la Chaire de recherche du Canada sur les écritures numériques, et accessoirement doctorant en littérature, les processus mis en place pour la plateforme de l'Anthologie Grecque m'intéressent particulièrement pour les questions d'édition et de publication. Ce hackathon était une belle opportunité de découvrir et de parcourir les contenus de la plateforme, et de voir comment cet accès global et visuel peut être pensé par un collectif, pour lui-même et pour les autres.

#### Enzo

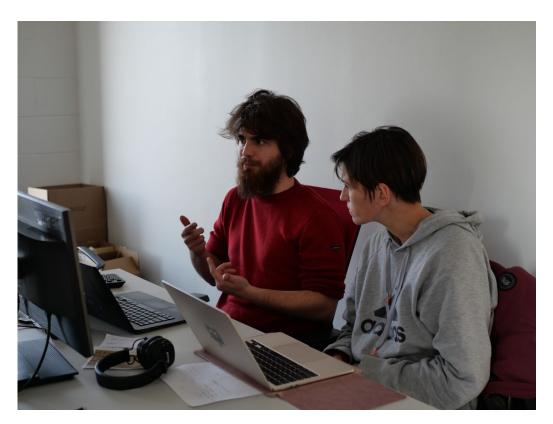


Figure 3 – Enzo et Marianne

Étant le gestionnaire d'une équipe d'intelligence artificielle à Heyday By Hootsuite, je fus intrigué par la proposition de mon amie Mathilde Verstraete de participer à cet Hackathon. Je connaissais le projet AG pour avoir assisté à une présentation succincte de la part de Marcello Vitali-Rosati lors d'une précédente rencontre, alors que je lui racontais mon cursus en TALN à l'Université de Genève. Interpellé par la liberté de sujet, je saisissais l'occasion de proposer mes compétences en traitement de données, pour en apprendre davantage sur ce corpus et rencontrer le monde académique des digital classics.

#### Lena



Figure 4 - Lena

Je travaille comme responsable du laboratoire Ouvroir d'histoire de l'art et de muséologie numérique à l'Université de Montréal, ce qui m'amène à côtoyer quotidennement les membres de la CRCEN qui développent et maintiennent l'Anthologie Grecque. Formée en histoire de l'art et en informatique pour les sciences humaines (Université de Genève; Université de Montréal), je n'avais aucune connaissance préalable à propos de l'anthologie. J'étais particulièrement motivée à participer au hackathon car je venais de participer à deux jours de conférences sur la visualisation de données (Observable Insights 2022) et j'avais très envie de mettre en pratique ce que j'y ai appris, comme l'utilisation de la nouvelle librairie Plot par exemple.

## Margot



Figure 5 - Margot

Ancienne coordonnatrice du projet de l'édition numérique collaborative de l'Anthologie Grecque, je m'efforce de suivre les développements du projet notamment présentés dans le cadre de ces journées d'étude. Toujours présente à la CRCEN, j'ai participé au hackathon dans l'espoir de travailler avec différents profils et de pouvoir également renouer avec une culture classique qui a fait partie de mon cursus.

#### Marianne



Figure 6 - Marianne

Passionnée de grec ancien et de programmation, je suis maîtresse de conférences à l'École Normale Supérieure de Lyon en Humanités Numériques (j'ai une formation traditionnelle de Lettres Classiques). J'ai été amenée à parler de mes recherches lors des journées d'études organisées dans le cadre de l'Anthologie Grecque. J'ai pu participer au hackathon pour découvrir et manipuler davantage l'API des Anthologies, et l'expérience, rare, tant pour ce qui est de l'ouverture du sujet que sur le format de réalisation, m'a permis de renouer avec une ambiance de travail intense et dans d'excellentes conditions.

#### Chauffe intellectuelle

À l'origine du travail effectué par l'équipe aussi nommée le groupe 4 se trouve une recherche de complémentarité entre les différentes expertises et intérêts des membres le composant (voir *Présentation des membres* ci-dessus). La réalisation finale est le produit des compétences apportées par les profils de chacun des membres de l'équipe : elle a impliqué l'extraction et la restructuration des données, des visualisations versionnées, une documentation produite et publiée parallèlement aux expériences réalisées. La rencontre des profils a mené à une synergie commune que retranscrit également cet article.

Une image dessinée (cf. image ci-dessous) au tout début de la période de brassage de cerveaux du groupe a permis à juste titre d'illustrer l'idéal de collaboration synergique visé non seulement par l'équipe mais plus largement par la communauté du colloque :

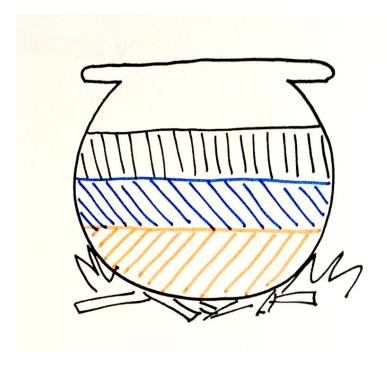


Figure 7 — La Marmite de la collaboration - dans une marmite mise à chauffer, différents éléments aux diverses aspérités et caractéristiques sont réunis. Ils forment, sur une vue en coupe, des strates de connaissances.

Ce que l'image ne montre pas directement est ce qui advient des différentes strates de connaissances une fois que l'espace qui les contient est rendu brûlant. Le travail de « chauffe intellectuelle » permet à ce qui était distinct de se mêler. Ce qui était auparavant un ensemble de composants ne partageant pas ou très peu de points communs dans leur culture se modifie pour faire émerger une complémentarité complice qui ne permet (presque) plus de distinguer les ingrédients originaux. Si le dernier stade de cette métamorphose, l'idée d'une complémentarité effaçant les oppositions disciplinaires pour fonder un collectif en action, s'est retrouvé notamment énoncé dans le cadre des échanges du colloque Navigations anthologiques, et évoque une dynamique plus générale des Humanités numériques, elle est l'amorce du projet du groupe 4 (sans lui être exclusive puisque d'autres groupes du Hackathon ont pu l'adresser ou la performer indirectement).

Dans le cadre d'un projet délimité dans le temps comme dans l'espace de ses développements (le Hackathon se déroulait au sein de l'Université de Montréal sur une durée de 8 heures), le groupe 4 avait pour objectif commun de faire s'articuler ses différents profils pour permettre un alliage d'idées (ou un maillage de connaissances) et une interopérabilité des pratiques. Si les deux perspectives sont classiquement distinguées dans les Humanités numériques tout en déterminant la particularité d'un champ disciplinaire, il était important que dans l'articulation, la théorie et la technique ne soient pas cloisonnées et que chaque membre puisse contribuer, selon ses compétences, aux deux perspectives. L'un des premiers défis qui s'est alors exprimé était moins le fait de s'accorder sur la théorie du projet (que veut-on réaliser? quelle idée souhaite-t-on démontrer?) que sur un modèle technique (comment va-t-on réaliser ensemble? comment mettre à contribution chacune des expertises pour former une proposition cohérente?) suffisamment souple et ouvert pour permettre à chacun·e des participant·e·s de contribuer avec ses pratiques numériques, d'apporter sa perspective sur l'objet d'étude et de la mettre en résonnance avec les autres approches.

C'est en gardant en tête ces objectifs de complémentarité et d'interopérabilité qu'a été discuté le sujet de la réalisation. Au départ nourri de projets ambitieux (comme la représentation des liens sémantiques entre les différentes épigrammes en créant des *embeddings*<sup>1</sup>, les efforts du groupe se sont ensuite

<sup>1.</sup> Les *embeddings* sont des représentations sémantiques des mots qui capturent leur sens et leurs relations sémantiques avec d'autres mots.

concentrés sur un projet plus modeste mais qui serait réalisable dans le temps donné. Parmi les différentes propositions de réalisation discutées, deux projets potentiels ont été identifiés et formulés lors du *brainstorming*:

- la fabrique d'un livre épigrammatique à la volée : selon des critères spécifiques et avec des parcours de lecture adaptatifs ;
- la visualisation ou le *graphe* du corpus : sous la forme de graphiques, visualiser et structurer des liens anthologiques selon des critères thématiques, littéraires et/ou philologiques.

Le premier projet des livres épigrammatiques évoquait dans sa structure et sa visée didactique le projet POP (Plateforme Ouverte des Parcours d'imaginaires), projet initié par la CRCEN en 2018 qui propose des parcours de lecture thématiques de l'Anthologie grecque. Parce qu'il impliquait une lecture plus affinée du corpus, le projet POP était orienté vers des objectifs de diffusion et de personnalisation des navigations. Cette première perspective n'a pas été retenue justement parce qu'elle aurait nécessité un travail plus approfondi de la structure de l'Anthologie. Pour nous permettre de fournir un livrable à l'issue du hackathon, nous avons conservé le deuxième sujet, visant à explorer les données anthologiques, objectif qui représentait également davantage l'ensemble des profils et les intérêts du groupe. L'horizon du hackathon pour le groupe 4 se dessinait donc comme la lecture de l'Anthologie Grecque au travers des relations entre les épigrammes et au fil des livres qui la constituent.

## Dialogues entre idéation et réalisation

La perspective de construire un prototype basé sur les données de l'Anthologie Grecque a suscité beaucoup d'envie de la part de l'équipe, comme une occasion de plonger dans un travail éditorial de longue haleine, et comme une opportunité de visualiser ces textes d'une manière inédite. Pour mieux appréhender cet ensemble de textes, il est nécessaire de comprendre l'articulation entre le projet initial et sa version numérique contemporaine. À partir de cette description et du fonctionnement de l'API, nous pouvons définir des champs exploratoires pour imaginer un nouvel objet éditorial.

#### L'Anthologie comme corpus

Si le livrable final a été réalisé de manière concrète par plusieurs mains techniques, sa modélisation en amont repose sur une considération précise du corpus de travail et s'inscrit ainsi dans la lignée des recherches déjà produites en cours à la CRCEN. Les différentes recherches qui ont été produites autour du projet d'édition de l'Anthologie Grecque, pour documenter un avancement ou expliciter une démarche, mettent chacune en perspective une même caractéristique de l'Anthologie qui est autant à l'origine de sa richesse qu'elle est la cause de sa complexité : l'Anthologie grecque est en réalité moins une œuvre littéraire qu'un corpus regroupant plus de 4 000 épigrammes (soit des fragments généralement courts) (Mellet et Verstraete 2024). Par son histoire (la multiplication des manuscrits et des compilations changeant sans cesse l'ordre, la nature et la succession des épigrammes) mais également par son principe de rassemblement de fragments distincts (par l'époque, l'auteur, le thème, le style ou même le modèle épigrammatique), l'Anthologie défie une notion d'œuvre en tant qu'objet délimitable et échappe ainsi à des principes d'édition savante (mue par la recherche d'une *vérité* du texte et l'établissement d'une unique version) (Vitali-Rosati et al. 2021). Cette perspective sur un objet littéraire est au fondement d'un projet dont la composante éditoriale souhaite retranscrire l'hétérogénéité et la multitude d'un ensemble culturel.

La notion de corpus à la différence du principe d'œuvre permet d'embrasser autant de phénomène de disparités, de pluralité que de contradiction au sein d'un collectif de textes sans lui enlever toute possibilité d'édition ou de structuration définies. Ce que cette approche implique cependant est de développer des outils de lecture, d'analyse et plus largement d'édition qui prennent la mesure du collectif sans l'unifier ou lui imposer une cohérence mais en valorisant les dynamiques dialogiques entre les fragments (l'édition des mots-clefs thématiques est un exemple). La logique anthologique dans cette perspective se fait donc par entités (terme utilisé dans l'API du projet) qui demeurent en écho sans être inter-dépendants.

#### Idéation entre dessin et design

Fondé sur cette idée de l'Anthologie comme corpus et donc comme ensemble dialogique, le projet du groupe 4 visait à refaire le lien avec un imaginaire qui se trouve justement à l'origine de l'objet de travail : celui de la couronne (ἀνθολογία ou florilège en latin) qui réunit et tisse collectivement une série

d'éléments représentatifs, comme des fleurs d'une culture (celle de la littérature grecque pour le cas de l'Anthologie Grecque) (Vitali-Rosati et al. 2020). Pouvoir montrer les liens au sein du corpus, et surtout les dynamiques d'inclusion et d'exclusion (est-ce que dans un livre où est représenté le thème de l'amour, le thème de la mort est également représenté? le thème de l'argent en-est-il exclu? est-ce que ce lien se fait au sein des mêmes épigrammes ou dans une vision plus diffuse au sein du livre?) qui parcourent le corpus et participent sans doute à établir une cohésion anthologique tout en démarquant cette anthologie d'autres structures narratives anthologiques. Ces visualisations (dont une qui correspond à un graphe et qui a donné le titre au livrable) devaient permettre, telles que décrites au cours du brainstorming de l'équipe, de structurer des liens entre les données de l'Anthologie qui n'étaient ni directement visibles sur les plateformes du projet déjà existantes ni directement perceptibles par la lecture individuelle.

#### [définition] Visualisation de données

La visualisation de données vise à produire une représentation graphique de données. Pour créer une visualisation de données, il faut typiquement travailler avec des données sérielles, des éléments d'un jeu de données à étudier/comparer/analyser et choisir un critère qu'on va associer à une caractéristique graphique. Voici des exemples :

- Histogramme ou diagramme en barre : montrer graphiquement le nombre d'épigrammes (quantité, à associer à la longueur d'une barre) pour chaque livre (une barre par livre).
- Scatterplot ou diagramme de dispersion : inscrire les mots-clefs des épigrammes d'un livre sur dans un système de coordonnées cartésiennes. En abscisse, l'axe des x, se trouve le déroulement "chronologique" des épigrammes d'un livre. Et en ordonnée, l'axe des y, se trouve la liste de mots-clefs. Pour chaque épigramme est généré un point à son emplacement (x) placé à la hauteur du mot-clef concerné (y). On peut ainsi visualiser dans un espace les passages où les même mots-clefs sont employés, le nombre d'occurences de chacun des mots-clefs (lecture horizontale sur la hauteur en y d'un mot-clef), ainsi que les mots-clef associés à un passage (lecture verticale à partir de l'emplacement en x du passage).

Les visualisations ou graphes en réseau (graph network) sont produites à partir de données structurées en nœuds et arrêtes. Les nœuds sont les points sur le réseau, et les arêtes les liens entre les points. À la différence d'un diagramme de dispersion par exemple, le placement des nœuds dans l'espace graphique n'est pas porteur de sens. Il est possible de réorganiser un réseau sans changer sa signification intrinsèque, toutefois son organisation spatiale influence la lecture et de fait l'interprétation du graphique.

Les données avec lesquelles nous avons souhaité travailler sont les suivantes :

- l'Anthologie Grecque en tant qu'un ensemble de livres;
- les **livres** qui sont des ensembles d'épigrammes organisés de façon linéaire;
- les épigrammes qui sont des entités textuelles disponibles en plusieurs traductions et auxquelles sont attribuées des mots-clefs;
- les mots-clefs (16 catégories, 859 mots-clefs) qui sont associés aux épigrammes selon une catégorisation (ensemble cohérent et thématique de mots-clés).

Il faut noter que les textes présents dans l'Anthologie Grecque sont nombreux : plus de 4 000 épigrammes avec parfois cinq traductions pour chacune. L'enjeu pour nous était donc de proposer une nouvelle façon de découvrir et de prendre la mesure de cette Anthologie. Le projet a été une première fois présenté à l'ensemble des participant·e·s du colloque avec pour objectif d'établir une lecture graphique *et* anthologique.

Deux approches complémentaires ont été proposées :

1. appréhender les livres qui composent l'Anthologie : quels peuvent être les liens entre ces livres qui se composent chacun d'un nombre variable d'épigrammes et qui représentent une culture épigrammatique particulière? L'objectif était de visualiser les connexions entre les différents livres de l'Anthologie grâce aux mots-clefs qui sont associés à chaque épigramme (épigrammes qui composent les livres). En partant du nombre de mots-clefs que peuvent partager plusieurs livres, nous proposons de donner une visualisation graphique des relations entre ces livres de l'Anthologie Grecque, à travers la qualification des épigrammes.

2. saisir l'évolution du sens des épigrammes entre elles : comment visualiser l'évolution du contenu des différentes épigrammes? Ces textes étant très nombreux, il est difficile de percevoir facilement et visuellement ce foisonnement sémantique même si l'architecture proposée par plusieurs compilateurs en différents livres y aura donné une orientation particulière. Nous utilisons les mots-clefs associés aux épigrammes, et nous analysons plus particulièrement la variation de ceux-ci à l'intérieur d'un même livre. Les mots-clefs étant catégorisés (regroupement des termes selon des ensembles lexicaux organisés), les éditeurs et les éditrices ont une pratique relativement uniformisée d'attribution de mots-clefs selon des catégories prédéfinies (époque, structuration poétique, genre littéraire, thématiques principales etc.). Il s'agit donc de visualiser les épigrammes d'un même livre selon les mots-clefs qui leur sont attribués.

Orienté par la sélection de types graphiques qui correspondaient aux approches, le projet a en partie été réalisé via la plateforme Observable qui donne accès aux codes des visualisations de données. Observable permet la collaboration en temps réel sur le code. Comme elle est hébergée en ligne, les visualisations produites sont accessibles et visibles par tou·te·s, et peuvent être intégrées dans d'autres contenus web (en utilisant des iframes) comme dans cela est le cas dans cet article.

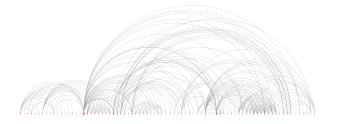
#### [définiton] Notebook

Un notebook est un environnement de programmation dans lequel le code est structuré en cellules que l'on peut exécuter (ou run) individuellement. Contrairement à la programmation en scripts, le code est ici pensé pour être segmenté, ce qui invite à une organisation différente des contenus. Il est possible d'ajouter des cellules de texte pour inclure des commentaires ou des explications sur le code : cette fonctionnalité, qui aurait pu servir à documenter le travail de l'équipe, n'a que peu été investie, la documentation a été produite en parallèle de la conception sur une note collaborative (un HedgeDoc). Dans un notebook, les cellules peuvent également être convoquées à l'intérieur d'autres cellules et ainsi produire une chaîne d'actions et interactions au sein de l'ensemble.

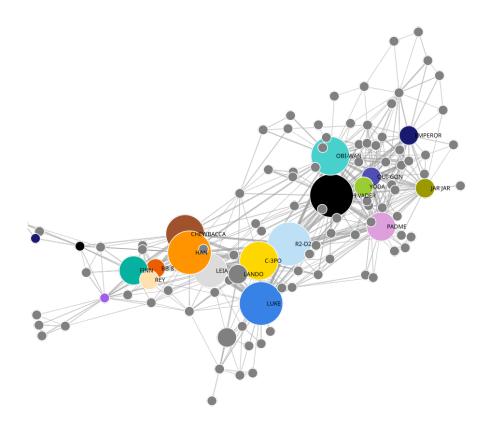
Les notebooks Observable sont spécialisés pour la visualisation de données. Ils intègrent de nombreuses fonctionnalités qui facilitent et accompagnent la création de visualisations, notamment l'intégration des librairies Plot et D3.js. Plot, développée à l'interne (in-house), prédéfinit des formats de visualisation pour lequels on doit simplement préciser la source de données et les paramètres commes les propriétés concernées, les couleurs à y associer etc. D3.js est une librairie de code qui dispose de nombreuses fonctions pour créer des visualisations de données "sur mesure". Observable inclut également des sélecteurs typiquement utilisés pour faciliter les interactions avec les visualisations, tels que des boutons ou des sélecteurs d'intervalles.

En parcourant les exemples dans Observable (qui a probablement été effectué avec une recherche par mot-clef dans l'outil de recherche intégré), nous avons identifié plusieurs formes visuelles qui pouvaient convenir et être adaptées aux données de l'API de l'Anthologie :

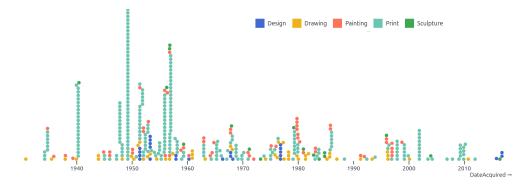
— une visualisation chronologique avec des arcs (Staats 2018) : chaque point représente un personnage des *Misérables*. Ils sont ordonnés par ordre d'apparition. Les arcs représentent les interactions entre les personnages.



— une visualisation en graphe qui montre le réseau des personnages dans STAR WARS (Thaden 2018). Chaque point est un personnage et les liens représentent leur co-occurences dans une scène. Les couleurs sont utilisées pour mettre en valeur les personnages principaux.

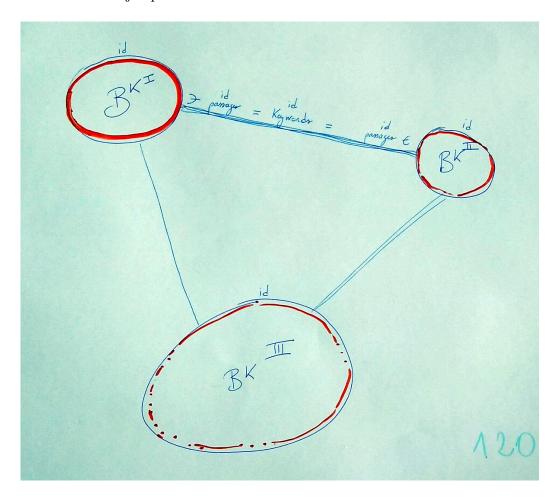


— un diagramme de points *Scatterplot* (Buffa 2022) produite avec la Librairie Plot. Dans cet exemple, les points sont placés en fonction de leur valeur sur l'abscisse (x), et s'empilent les uns sur les autres selon l'affluence par années. La couleur est associée à une autre variable, et premet d'exprimer, dans le cas de ce graphique sur les œuvres d'art du MoMA, le médium de chaque œuvre de la collection.

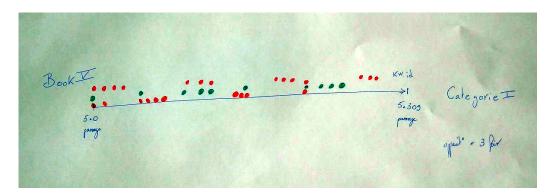


Pour chacune de ces formes graphiques, il nous semblait possibe de remplacer les données utilisées par celles de l'API de l'Anthologie. Nous avons donc commencé par faire un *fork* (une copie pour la réédition du code) de ces graphiques pour étudier plus en détail le code et notamment contrôler le format des données utilisé pour créer la visualisation. Des premières modifications ont été effectuées pour observer si et comment pourraient être intégrées nos données. C'est à cette étape qu'a commencé le travail de conception des graphiques, un travail qui exige un certain niveau de connaissances sur les données de l'API ainsi qu'une approche *design* sur les formes visuelles à produire.

Deux modèles graphiques se sont précisés, au fil du travail collaboratif, et ont été retravaillés jusqu'à leurs versions finales.



Dans ce premier dessin, nous avons utilisé une visualisation en réseau, fondée sur des nœuds et des arcs. Les nœuds sont des livres, et les liens sont des passages qui ont le même mot-clef. Ainsi, des livres qui auraient des mots-clefs en commun se trouvent reliés, et l'épaisseur du trait peut indiquer l'intensité de ces thématiques communes.



Ce second modèle graphique présente un livre comme une série ordonnée de passages. L'abscisse (x) représente les identifiants des passages, tandis que l'ordonnée (y) représente tous les mots-clefs d'un livre. Ainsi, on représente la présence de mots-clefs par passage. Ce modèle visualise, sur le principe d'une lecture distante, les thématiques d'un livre par la présence et la fréquence des mots-clefs qui sont associés à chaque passage. Une lecture plus rapprochée invite à faire des corrélations entre les co-occurrences ou les utilisations récurrentes de mots-clefs. L'association d'une couleur à chaque mot-clef, et plus largement le recours à une colorimétrie variée, en plus de son emplacement sur l'abscisse, facilite leur distinction visuelle et la lecture d'ensemble.

En parallèle de ce travail d'idéation graphique a été effectué un travail de récupération et de restructuration des données de l'Anthologie à partir de l'API pour permettre l'injection des données dans les structures visuelles.

#### L'API

Lors du hackathon, chaque groupe était invité à récupérer (ou *fetch*) les données de l'API de la plateforme principale du projet dans l'idée de pouvoir les restructurer et travailler diverses visualisations. La documentation fournie dans le cadre de cette rencontre conseillait de procéder à la création d'un dépôt (ou *dump*) en local – sur les machines des utilisateur·rice·s – de toutes les

épigrammes de l'Anthologie (décrites dans l'API comme des passages). Ce fonctionnement, qui pourrait sembler peu intuitif dans le cadre d'hackathons classiques dans les milieux des sciences informatiques, s'explique par le caractère essentiellement interdisciplinaire de l'évènement. Connaissant la flexibilité de l'API du projet mais également ses limites, les organisateurs trices ont préféré à la solution plus commune de donner accès aux participant es au corpus au travers de requêtes, l'alternative du dump de l'API. Cela permettait davantage de liberté, d'autonomie mais également de respecter l'esprit d'exploration des données qui était à l'origine de la rencontre.

#### [définition] API

```
anthologyData = ▼Object {
  passages: "https://anthologiagraeca.org/api/passages/"
 books: "https://anthologiagraeca.org/api/books/"
 scholia: "https://anthologiagraeca.org/api/scholia/"
 texts: "https://anthologiagraeca.org/api/texts/"
 media: "https://anthologiagraeca.org/api/media/"
 manuscripts: "https://anthologiagraeca.org/api/manuscripts/"
 keywords: "https://anthologiagraeca.org/api/keywords/"
  keyword categories: "https://anthologiagraeca.org/api/keyword categories/"
 languages: "https://anthologiagraeca.org/api/languages/"
 descriptions: "https://anthologiagraeca.org/api/descriptions/"
 cities: "https://anthologiagraeca.org/api/cities/"
 authors: "https://anthologiagraeca.org/api/authors/'
 comments: "https://anthologiagraeca.org/api/comments/"
 editions: "https://anthologiagraeca.org/api/editions/"
 alignments: "https://anthologiagraeca.org/api/alignments/"
  passage_internal_references: "https://anthologiagraeca.org/api/passage_internal_references/"
```

Cette visualisation de l'API du projet permet de voir comment cette dernière a été structurée : elle contient beaucoup de liens (URL) qui pointent vers une section interne de l'API. Cette organisation est logique au niveau informatique, puisqu'une machine peut aisément parcourir les liens et ainsi naviguer dans les contenus. Toutefois, en tant qu'utilisateur·rice·s de l'API lors du hackathon, cette structuration signifiait un travail plus long pour récupérer les données.

Dans l'exemple du passage ci-dessous, le premier élément du résultat est retourné par la requête https://anthologiagraeca.org/api/passages, il contient en valeurs son identifiant (id), le texte en plusieurs langues (texts[n].text)), tandis que les mots-clefs (keywords) contiennent uniquement l'URL.

```
exemplePassage = ▼Object {
 book: ▶ Object {url: "https://anthologiagraeca.org/api/books/9/", number: 1}
 fragment: 1
 sub_fragment:
 url: "https://anthologiagraeca.org/api/passages/urn:cts:greekLit:tlg7000.tlg001.ag:1.1/"
 manuscripts: ▼Array(1) [
   0: "http://digi.ub.uni-heidelberg.de/iiif/2/cpgraec23%...449838187696,6.205747605164515/full/0/default.jpg"
 texts: ▶ Array(4) [Object, Object, Object]
 authors: ▶ Array(1) [Object]
 cities: ▶ Array(0) []
 keywords: ▼Array(2) [
   0: "https://anthologiagraeca.org/api/keywords/116/"
   1: "https://anthologiagraeca.org/api/keywords/200/"
 scholia: ▶ Array(2) [Object, Object]
 comments: ▶ Array(1) ["https://anthologiagraeca.org/api/comments/670/"]
 external references: ▶ Array(0) []
 internal references: ▶ Array(0) []
 media: ▶ Array(0) []
```

Le travail effectué à partir d'un miroir de l'API du projet nous a permis d'acquérir une connaissance approfondie mais également critique du projet. L'API incarne, dans la dimension technique du projet, une facette importante voire primordiale. Ainsi, la parcourir et l'explorer dans le cadre d'une réalisation éphémère a permis de découvrir le projet comme objet technique : c'est-à-dire d'observer comment l'idée de corpus, essentielle à la perspective de recherche, avait été techniquement implémentée au sein d'une structuration des données. En revanche, cette même exploration a amené le développement d'un regard critique sur les choix de structuration des données qui avaient été faits. La réunion d'expertises et cultures diverses autour de cet objet a permis de souligner les failles techniques du projet (dont certaines avaient déjà été identifiées par les chercheurs du projet) et de questionner plus généralement les choix d'implémentation technique. Cette perspective du travail collaboratif fait du hackathon non seulement un moment ludique et une occasion d'extension d'un projet déjà mûr, mais également un cadre pour établir un dialogue critique avec un objet culturel et technique.

Par la clarté de sa structure, l'API du projet est simple de navigation à partir de son *endpoint* /api. En revanche, l'accès à certaines données plus particulières se fait par l'endpoint /api/passages/ donc sans passer par les

endpoints proposés, ce qui crée de la redondance. Par exemple, le endpoint /api/book n'a en tant que tel pas été utilisé dans le cadre du projet du groupe pour accèder aux épigrammes d'un livre. Il s'est avéré, dans le travail de restructuration de l'API, que cet endpoint relevait davantage d'un filtre pour trier les épigrammes : api/passages/?book\_number=1 donne les épigrammes du livre 1. Si la navigation dans les données de l'API s'en trouve accrue, il demeure une redondance de par le fonctionnement des URL retournées et attribuées dans chaque objet. Dans une approche restfull de l'API, la structure des données est faite de manière à permettre aux clients de spécifier les données qu'ils souhaitent récupérer, afin d'éviter le transfert d'informations inutiles.

La structuration des mots-clefs nous a particulièrement intéressé dans la mesure où c'est un des types de données que nous allions travailler pour proposer des visualisations qui regroupent les passages en fonction de leurs thématiques et qui profilent les livres de l'anthologie par thématiques. Par l'étude de la structure de l'API, nous avons remarqué en suivant les URL d'un mot-clef comme présenté ci-dessous (https://anthologiagraeca.org/api/keywords/116/) que les mots-clefs de l'API sont renseignés (names) et catégorisés (categories) en plusieurs langues. Comme la liste des passages associés au mot-clef est également présente, il aurait donc été possible d'inverser la logique et d'utiliser l'URL https://anthologiagraeca.org/api/keywords/ pour parcourir les mots-clefs et les associer au fur et à mesure à des passages. Nous avons cependant décidé de centrer le travail sur les passages, ceux-ci étant les éléments constitutifs de l'anthologie. Pour obtenir les mots-clefs liés à un passage, il faut donc faire autant de requêtes qu'il y a de mots-clefs. Pour les obtenir pour l'ensemble des passages, il est donc nécessaire d'écrire un script pour parcourir la liste des passages et repérer les valeurs sélectionnées pour chaque mot-clef. Nous avions envisagé dans un premier temps de procéder uniquement à un frontend pour rendre visibles les different plots (comme montré dans ce notebook observable HQ); cependant la diversité technologique de notre groupe nous a réorienté vers une extraction et une tranformation en python.

## Organisation synergique

Les décisions prises par le groupe ont été motivées pour correspondre à l'objectif synergique de la collaboration et faire se rejoindre les différents

profils techniques. Dans l'espace de travail du groupe 4, les environnements techniques se sont rapidemment structurés (l'espace API et l'utilisation de Python avec Enzo et Marianne; l'espace Observable avec Lena; l'espace HTML de la documentation avec Margot et Antoine), pour ensuite dialoguer et articuler les différentes productions (restructurations des données, conceptions graphiques, documentation) en un commun.

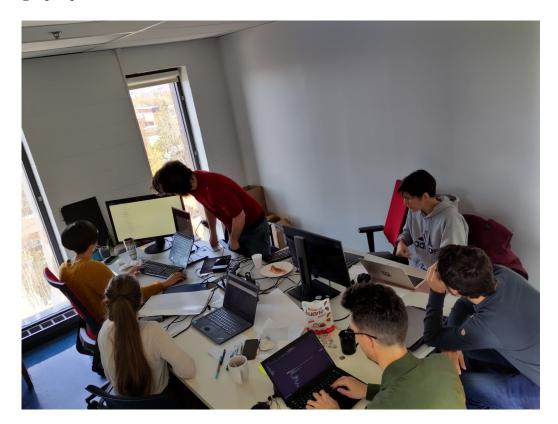


Figure 8 - Team 4 en ébullition

#### Restructuration des données

Pour le groupe de pythonistes — qui correspondrait dans une équipe au ba-ckend —, l'extraction et la transformation des données (voir ce jeu de données dumpall) a permis de faire émerger une notion importante. Afin de procéder à l'affichage et au traitement des données, un temps de sémantisation était nécessaire. Le groupe devait donc s'accorder sur une sémantique commune :

avoir ensemble une même définition de ce qui constitue un *passage*, un mot-clef, etc.

Coordonné au backend, l'espace de travail de manipulation et de visualisation des données en cours d'extraction et de restructuration, le frontend dialogue avec l'espace de documentation en synchrone pour définir le livrable et au mieux expliciter les besoins pour sa production. Dans le temps imparti, les objectifs du projet se reprécisent pour tendre vers quelque chose de plus modeste mais surtout de réalisable, réflexion qui impacte également le backend pour orienter la restructuration de manière efficace. Le graphe des liens entre épigrammes est alors envisagé à partir de données et fonctionnalités déjà existantes au sein du corpus : les mots-clefs. Les visualisations telles que documentées permettront de rendre compte 1. de la répartition des mots-clefs au sein de l'Anthologie, soit de voir quel livre a le plus de mots-clefs, de représenter les liens entre les livres ayant les mêmes mots-clefs; 2. de la répartition des motsclefs au sein d'un livre en particulier, soit de voir l'évolution d'un mot-clef au fil des épigrammes d'un même livre, de remarquer quel mot-clef y est le plus présent et comment il s'articule avec les autres (par phénomène de relation ou d'exclusion).

La donnée filtre est donc le mot-clef tandis que la donnée d'environnement est le livre (qu'il s'agisse d'un livre en particulier ou de l'ensemble des livres du corpus anthologique). Pour parvenir à la manipulation de ces données dans les structures visuelles, chaque livre devrait être représenté par le *backend*:

```
"nb_epi": 0, # nombre d'épigrame par livre
"list_id_epi": [], # la liste des ids des épigrams
"nb_kw": 0, # nombre de keyword l'intérieur des épigrame du dit livre
"list_id_kw": [], # la liste des keywords (qui s'avera
# etre la liste des URL des KW)
"name": 'undefined', # Le nom du livre
"number": 0, # le numéro du livre
"url": '', # l'url du livre
}
```

#### L'espace visuel

Suite à la phase d'idéation avec les graphiques sélectionnés, il nous fallait comprendre le fonctionnement détaillé du code et du format de données utilisé dans chaque visualisation, afin de l'adapter avec les données de l'Anthologie.

Pour la visualisation en graphe, il fallait transformer chaque livre en un nœud (node) et des liens (links) pour les relations thématiques par mot-clef. Il nous a fallu plusieurs allers-retours pour arriver au format susmentionné : les nœuds sont des livres (nom, numéro, url) pour lesquels on renseigne les épigrammes et les mot-clefs. Les liens sont décrits par une source et une cible (target), les deux nœuds à relier, ainsi que par une valeur numérique qui indique la force de lien, c'est-à-dire la quantité de mot-clefs en commun.

```
{
  source: 1 // livre source
  target: 4 // livre cible
  value: 3 // quantité de mot-cles en commun
}
```

Pour la visualisation de mot-clef (celle qui deviendra «Livre 5»), nous avons créé un jeu de données fictif pour vérifier si le format correspondait bien au résultat désiré.

```
{
  name: "chat"
  ordre: "5.0"
  cat:"domestique"
  url:"https://anthologiagraeca.org/passages/urn:cts:greekLit:
tlg7000.tlg001.ag:5.3/"
}
{
 name: "chien"
  ordre: "5.23"
  cat:"domestique"
  url:"https://anthologiagraeca.org/passages/urn:cts:greekLit:
tlg7000.tlg001.ag:5.3/"
}
{
 name: "baleine"
```

```
ordre:"5.117"
cat:"océan"
url:"https://anthologiagraeca.org/passages/urn:cts:greekLit:
tlg7000.tlg001.ag:5.3/"
} {
    name:"cheval"
    ordre:"5.250"
    cat:"loisir"
    url:"https://anthologiagraeca.org/passages/urn:cts:greekLit:
tlg7000.tlg001.ag:5.3/"
}

domestique
loisir
océan

domestique
loisir
océan

domestique
epigrammes
```

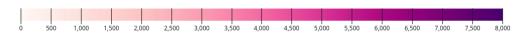
Figure 9 – Visualisation avec les données test

Une fois le format des données défini, nous avons continué à travailler sur le rendu visuel souhaité en attendant la préparation des données par l'équipe des *pythonistes*. Une bonne compréhension de l'Anthologie était notamment nécessaire pour s'assurer que les variables représentées graphiquement sont pertinentes.

#### Graphe des livres

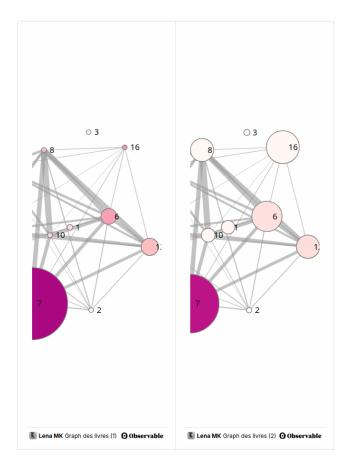
La première visualisation était chargée de montrer les liens thématiques entre les livres de l'Anthologie et donc de représenter des phénomènes de relation et d'exclusion au sein du corpus. Les livres constituent dans ce graphique les nœuds (16 livres donc 16 nœuds) et les liens ou *edges* les relations thématiques.

#### Nombre de mot-clés



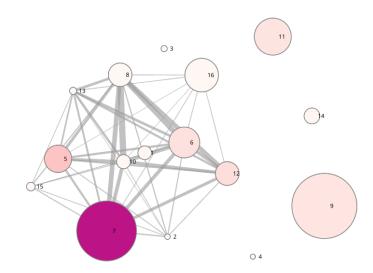
La représentation graphique avec deux variables, les tailles des nœuds et leur couleur, a mené à plusieurs hésitations et confusions au sein de l'équipe. Nous avons d'ailleurs changé de notebook en cours de route.

## Comparaison des deux graphes réalisés



Dans le Graph des livres (1), la taille des nœuds dépend du nombre de motclefs et l'intensité de la couleur reflète le nombre d'épigrammes dans le livre. À l'inverse, Graph des livres (2) représente le nombre d'épigrammes dans la taille des noeuds et le nombre de mot-clefs dans la couleur.

Nous avons décidé qu'il était plus adéquat que la taille des nœuds représente le nombres d'épigrammes par livres : plus un nœud est grand, plus le livre contient d'épigrammes. La coloration des nœuds se fait selon un dégradé (allant du blanc au violet en passant par le rose) et représente pour chaque livre le nombre de mots-clefs présents. Dans le graphique produit, le livre 7 a donc beaucoup d'épigrammes et beaucoup de mots-clefs : ce qui pourrait sembler évident mais des cas comme le livre 16 (qui comporte beaucoup d'épigrammes mais peu de mots-cles) démontrent qu'il n'y pas une corréalation systématique entre les deux aspects. Les edges représentent les liens entre les livres à partir de thématiques communes : plus un edge sera épais, plus les deux livres qu'il relie auront de mots-clefs en communs. Si la plus grande partie des livres de l'Anthologie partagent des thématiques communes, les livres 3, 11, 14, 4 et 9 apparaissent comme des cas à part, sans pour autant former un groupe thématique autonome (et cela sans que le nombre de mots-clefs constitue un facteur puisque les livres 11 et 9 ont un nombre de thématique relativement moyen). Bien entendu les visualisations produites dans le cadre du projet du groupe 4, parce qu'elles reposent sur les mots-clefs qui proviennent d'un travail d'édition en cours, sont dépositaires d'un état d'avancement du projet Anthologie grecque.



Loin de représenter un état thématique du corpus anthologique, ce graphique permet d'accèder à un portrait de l'édition qui en a été faite : dans l'enrichissement thématique, certains livres ont été plus nourris que d'autres (les livres 5 et 7 ont été des cas d'édition du projet). Cette première visualisation a engendré une seconde qui permet justement d'avoir une analyse à l'intérieur de chaque livre.

#### Livre 5

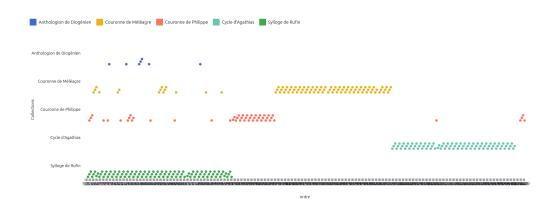
La deuxième visualisation s'attache à représenter l'évolution des mots-clefs au sein d'un livre toutes catégories de mots-clefs confondus (le projet d'édition a distingué des catégories de mots-clefs comme les formes métriques, les époques, les motifs, etc.) et en sélectionnant une catégorie en particulier. Le livre 5, qui comporte un groupe d'épigrammes relativement petit (comme montré par le premier graphe) mais avec une belle richesse thématique, a été choisi pour les premières explorations. Le livre, consacré à l'amour et l'expression

érotique, comporte plusieurs catégories thématiques qui se succèdent comme en témoigne un premier visuel :



Le survol de chaque point affiche des détails, le clic renvoie directement à l'épigramme dans l'Anthologia Graeca.

La lecture du profil thématique du livre étant complexe à établir, un autre visuel s'est consacré à la représentation de catégorie thématique précise au sein du livre 5. Le corpus anthologique se composant comme la réunion de plusieurs sources, il était intéressant de représenter pour le cas particulier du livre 5 les différentes collections représentées et surtout leurs répartitions tout au long du livre. Pour ce faire, la catégorie "Collections", qui informe sur la provenance des épigrammes, a été investie.



Quatre grandes collections se distinguent de manière claire au sein du livre 5 : le Sylloge de Rufin au début du livre, suivi par la Couronne de Philippe, la Couronne de Méléagre et le Cycle d'Agathias. On remarque donc un collage plutôt explicite dans la composition du livre, bien que le début demeure parsemé d'ajouts ponctuels (dont l'Anthologion de Diogénien).

Par exemple, de l'épigramme 5.133 appartenant à la Couronne de Philippe :

```
"Ωμος1' ἐγώ, δύο νύκτας ἀφ' Ἡδυλίου, Κυθέρεια,
σὸν κράτος, ἡσυχάσειν ὡς δοκέω δ', ἐγέλας,. τοὐμὸν ἐπισταμένη
τάλανος κακὸν οὐ γὰρ ὑποίσω
τὴν ἑτέρην, ὅρκους δ' εἰς ἀνέμους τίθεμαι.
αἰροῦμαι δ' ἀσεβεῖν κείνης χάριν, ἢ τὰ σὰ τηρῶν
ὅρκι' ἀποθνήσκειν, πότνι', ὑπ' εὐσεβίης.
```

#### (Paton 1918)

J'ai juré, Cythérée, par ta puissance, que deux nuits loin d'Hédylion je resterais tranquille; mais tu riais, j'imagine, connaissant la maladie du malheureux que je suis; car je ne tiendrai pas la seconde nuit et je livre aux vents mes serments. J'ai mieux commettre une impiété pour l'amour d'elle que d'observer la foi que je t'ai jurée, auguste déesse, et mourir de piété.

(Waltz 2003)

on passe à l'épigramme 5.134 qui appartient à la Couronne de Méléagre :

Κεκροπὶ ἡαῖνε λάγυνε πολύδροσον ἰκμάδα Βάκχου, ἡαῖνε: δροσιζέσθω συμβολικὴ πρόποσις. σιγάσθω Ζήνων ὁ σοφὸς κύκνος, ἄ τε Κλεάνθους μοῦσα μέλοι δ' ἡμῖν ὁ γλυκύπικρος ἔρως.

(Paton 1918)

Verse, flacon du pays de Cécrops, en rosée abondante le jus de Bacchus, verse, et que tous les convives portent ensemble la même santé. Silence à Zénon, le cygne de l'école, ainsi qu'à la muse de Cléanthe; ne nous occupons plus que d'Érôs le doux-amer.

(Waltz 2003)

cycle qui finit avec l'épigramme 5.215 puisque l'épigramme 5.216 inaugure le Cycle d'Agathias :

εὶ φιλέεις, μὴ πάμπαν ὑποκλασθέντα χαλάσσης θυμὸν ὀλισθηρῆς ἔμπλεον ἱκεσίης: ἀλλά τι καὶ φρονέοις στεγανώτερον, ὅσσον ἐρύσσαι ὀφρύας, ὅσσον ἰδεῖν βλέμματι φειδομένω.

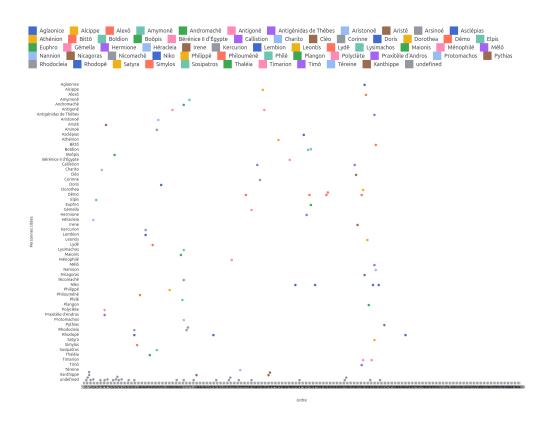
ἔργον γάρ τι γυναιξὶν ὑπερφιάλους ἀθερίζειν καὶ κατακαγχάζειν τῶν ἄγαν οἰκτροτάτων. κεῖνος δ' ἐστὶν ἄριστος ἐρωτικός, ὃς τάδε μίξει οἶκτον ἔχων ὀλίγη ξυνὸν ἀγηνορίη,

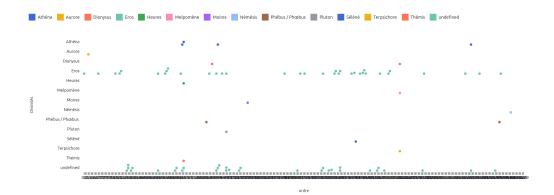
(Paton 1918)

Si tu aimes, fais que ton cœur ne se laisse jamais aller, effondré, plein de très humbles supplications; au contraire, montre dans tes sentiments quelque réserve, tiens le sourcil haut, mais tout en regardant d'un œil condescendant. Car c'est le propre des femmes de se désintéresser des orgueilleux, comme de faire des gorges chaudes des gens trop pleurnicheurs. Celui-là est un amant parfait qui sait allier à un caractère sensible un brin de fierté.

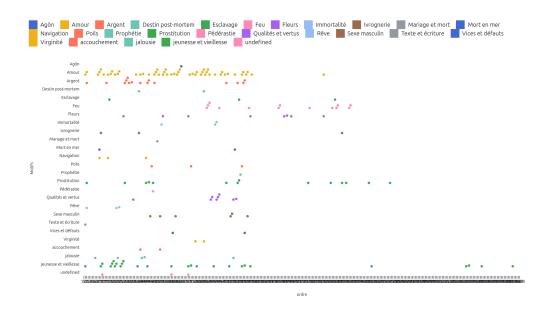
(Waltz 2003)

L'expérience a été réitérée pour les cas des catégories "Personnes cités", "Divinités", et "Motifs".





Sans trop d'étonnement, c'est principalement Eros qui est cité au long du livre 5.



De manière semblable, le thème de l'amour est majoritairement présent dans le livre, accompagné par celui de la jeunesse et vieillesse, mais on peut remarquer que ce thème et celui de l'argent semble s'exclure. Enfin si le début du livre est très thématisé (surtout par le thème principal), sa suite est moins claire et se partage entre les thématiques de la prostitution et du feu.

#### Documentation en synchrone

Réaliser un projet en si peu de temps, même à l'état de prototype, nécessite de garder des traces du chemin parcouru. D'une part pour pouvoir raconter ce qui s'est déroulé – c'est l'objet de cet article – mais aussi et surtout pour pouvoir communiquer au sein d'une équipe – entre personnes ou avec soimême. Une documentation est un système de mémorisation rassemblant un certain nombre de données utiles pour l'utilisation d'un projet, son évolution et sa maintenance. Cette tâche est loin d'être anodine, y compris dans un contexte comme celui du Hackathon – par définition court et intense. Il s'agit de relever toute information pertinente, de la contextualiser et de la rendre accessible dans l'immédiat et dans le futur.

Dès les premières ébauches, Antoine et Margot se sont chargés de collecter les informations nécessaires à une bonne compréhension des objectifs. Cet effort de formalisation avait un autre but que de faciliter la communication au sein du groupe : être capable de présenter le projet à l'issue de la journée du Hackathon. Et c'est là une spécificité importante de cet acte éditorial bien particulier, il faut en effet que le projet soit compréhensible autant par ses membres que par des personnes extérieures, autant aujourd'hui que demain.

Ainsi, en parallèle de la conception, la documentation a été réalisée en dialoguant continuellement avec les avancées du groupe. Cet objet textuel et structuré a permis autant d'archiver des états d'avancement du travail de visualisation que d'expliquer les objectifs de la recherche (les hypothèses). Il constitue, avec cet article, une trace de ce travail, comme un *commun* mis à disposition.

### Fin, renouveau, et documentation posthume

Nous avons souhaité proposer une vision différente de l'Anthologie Grecque en soumettant au lectorat plusieurs possibilités de visualisation des données, au regard de l'apport de chaque membre de l'équipe, avec son propre bagage académique et scientifique. Les visualisations proposées sont le reflet de la diversité, voire des divergences, impliquées par les compétences de chacun des membres du groupe, et ont permis de mettre au jour différents niveaux de perspective pour représenter l'Anthologie. Le croisement évolutif de ces niveaux de perspective nous a amenés à nuancer et à prendre de la distance vis-à-vis de chacune des méthodes employées, sans qu'une seule et unique méthode n'ait prépondérance sur les autres : à objets d'études différents nous avons dû appliquer des méthodes de représentations différentes, sans pour autant perdre de vue la logique globale du corpus. Enfin, grâce au versionnement automatique intégré aux notebooks Observable et à l'occasion d'écriture collaborative de cet article, nous avons également pu faire une documentation posthume des graphiques obtenus au cours de l'expérience. De la même manière que lors d'une fouille archéologique, nous avons pu remonter dans le temps et capturer les états d'avancement d'un projet autant au niveau technique de traitement des données, visuel de conceptions des graphiques ou éditorial d'organisation d'une expérience de collaboration.

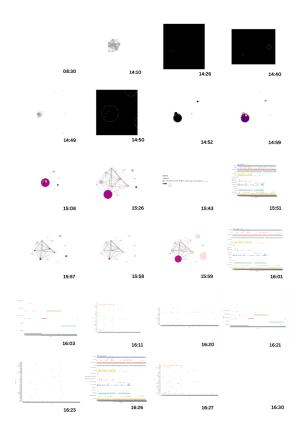


Figure 10 – Décomposition des différentes étapes de visualisation

## **Bibliographie**

Buffa, Paul. 2022. « Paul's Observable Tips & Tricks ». Observable.

Mellet, Margot, et Mathilde Verstraete. 2024. « Passés et Présents Anthologiques. Le Projet d'édition Numérique Collaborative de l'Anthologie Grecque ». In Communautés et Pratiques d'écritures Des Patrimoines et Des Mémoires. Intelligences Numériques. Presses univeristaires de Paris Nanterre.

Paton, William Roger. 1918. *The Greek Anthology*. Vol. IV. Loeb Classical Library. London-Cambridge: Harvard University Press.

Staats, Brian. 2018. « Arc Diagram ». Observable.

Thaden, Paul. 2018. « D3 Use-The-Force-Directed Graph ». Observable.

- Vitali-Rosati, Marcello, Margot Mellet, Servanne Monjour, Antoine Fauchié, Timothée Guicherd, David Larlet, et Enrico Agostini-Marchese. 2021. « L'épopée numérique de l'Anthologie grecque : entre questions épistémologiques, modèles techniques et dynamiques collaboratives ». Sens public, n° SP1596 (juillet). http://sens-public.org/articles/1603/.
- Vitali-Rosati, Marcello, Servanne Monjour, Joana Casenave, Elsa Bouchard, et Margot Mellet. 2020. « Editorializing the Greek Anthology: The palatin manuscript as a collective imaginary ». *Digital Humanities Quarterly* 014 (1).
- Waltz, Pierre. 2003. Anthologie grecque. t. II. Édité par Jean Irigoin. Vol. II. Budé. Paris : Les Belles Lettres.