

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Lignes du temps connectées pour l'analyse visuelle d'adaptations et de relations
entre artistes**

FRANÇOIS LÉVESQUE

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Juin 2020

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Lignes du temps connectées pour l'analyse visuelle d'adaptations et de relations
entre artistes**

présenté par **François LÉVESQUE**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Jinghui CHENG, président

Thomas HURTUT, membre et directeur de recherche

Michael MCGUFFIN, membre

DÉDICACE

*À Katerine Caron,
pour ton soutien inestimable*

REMERCIEMENTS

L'aboutissement de ce projet n'aurait jamais été possible sans le soutien continu de plusieurs personnes.

Je tiens d'abord à exprimer ma profonde gratitude envers Thomas Hurtut, mon superviseur de recherche à qui je dois un support inestimable. Sa grande patience, sa rigueur et son expertise ont poussé le projet plus loin que je ne l'aurais imaginé. Ses qualités m'ont inspiré du début à la fin de cette recherche, et continueront à m'inspirer lors de mes futurs projets.

Je tiens également à remercier le professeur Michel Gagnon qui m'a persuadé à faire une maîtrise recherche, aidé à trouver un domaine de recherche et présenté à Thomas Hurtut.

Je remercie aussi le professeur Michael McGuffin pour ses précieux conseils en début de projet. Je remercie les professeurs Michael McGuffin et Jinghui Cheng d'avoir participé à titre de membres du jury pour ce mémoire.

J'aimerais remercier la fabuleuse équipe de nos partenaires chez *Bibliothèque et Archives nationales du Québec (BAnQ)* qui ont partiellement financé ce projet et fourni le jeu de données. Je remercie tout particulièrement Jean-François Gauvin pour l'accompagnement tout au long du projet et Véronique Parenteau pour le soutien organisationnel lors des tests utilisateurs. Je remercie également Dominique Piché et Marielle St-Germain pour leur travail colossal de désambiguïsation des données liées.

Je tiens à remercier ma famille, en particulier Katerine Caron, Normand Lévesque et Marianne Lévesque, qui m'ont encouragé, écouté et soutenu pendant ce projet de recherche.

Enfin, je remercie mes amis pour leurs encouragements soutenus.

RÉSUMÉ

Le récent essor des données liées ouvertes démocratise l'accès à des jeux de données qui prennent souvent la forme de graphes, complexes à comprendre sans support visuel adapté. La visualisation de graphes est un sujet de recherche encore très actif. Bien qu'il existe aujourd'hui de nombreuses techniques de visualisation, dès qu'un graphe atteint une certaine grandeur (problème d'échelle) et une certaine quantité d'attributs (problème de densité d'informations) les techniques existantes perdent en efficacité.

Lors d'un récent projet d'innovation en valorisation des données avec BAnQ, nous avons travaillé avec un vaste jeu de données sur les adaptations musicales. L'objectif était d'arriver à représenter les relations de collaboration et d'influence entre les artistes tout en restant lisible. Les techniques de visualisation de graphes existantes ne permettaient pas de réaliser ces tâches avec suffisamment de clarté, ce qui nous a amené à proposer une nouvelle méthode de visualisation nommée MuzLink. L'objectif de ce travail de recherche est la conception, le développement et l'évaluation d'une nouvelle méthode de visualisation de graphes bipartis multivariés.

La nouvelle méthode de visualisation a d'abord nécessité la création d'un nouveau glossaire musical pour communiquer sans ambiguïté les différents rôles et relations qui existent entre artistes et chansons. La méthode de visualisation proposée est centrée sur un seul artiste. Elle s'inspire de plusieurs techniques existantes (node-link, lignes du temps et beeswarm) et les combine afin de représenter de façon explicite l'évolution des relations de collaboration et d'influence dans le temps. Cette méthode est divisée en trois lignes du temps : la première contient les chansons produites par l'artiste, la deuxième contient les chansons qui ont inspiré l'artiste et la troisième les chansons qui s'inspirent de productions de l'artiste. Chaque relation d'influence entre chansons est représentée par un lien, créant du même coup des lignes du temps connectées. Cette visualisation principale est juxtaposée à trois autres outils complémentaires qui permettent de répondre à des tâches spécifiques (listes d'artistes, notice de chanson et résumé de l'artiste). Les listes d'artistes énumèrent les artistes qui ont inspiré, collaboré ou se sont inspirés de l'artiste. La notice de chanson affiche des informations supplémentaires sur une chanson sélectionnée. Le résumé de l'artiste donne une vue globale sur les rôles occupés et les types de productions de l'artiste.

Un questionnaire a été développé pour évaluer la capacité de la méthode de visualisation à répondre aux tâches définies en début de projet. Il est composé de trois principales sections. La première est un tutoriel qui permet de se familiariser avec la méthode et notre glossaire

musical. La deuxième est un questionnaire chronométré à choix de réponses. La troisième est une exploration libre des données, suivie d'un sondage à questions ouvertes. Il a été complété par 18 participants, 9 hommes et 9 femmes, âgés de 18 à 70 ans.

Les résultats du questionnaire à choix de réponses révèlent que notre méthode de visualisation permet de répondre rapidement et avec précision aux tâches définies en début de projet. 83,3% des tâches ont été réussies avec un temps de réponse moyen de 10,1 secondes. L'analyse des résultats semble révéler que la majorité des erreurs proviennent d'une mauvaise compréhension de notre glossaire musical, notamment la différence entre «inspiration» et «influence». Dans le sondage à réponses ouvertes, les participants mentionnent être satisfaits et amusés par l'information véhiculée. La méthode permet également un bon niveau d'exploration. La majorité des participants affirment avoir découvert de nouvelles informations. Pendant la période d'exploration libre de 8 minutes, les participants ont en moyenne visité 3 artistes, interagi avec 38 artistes dans les listes d'artistes et 81 chansons sur les lignes du temps.

La principale limitation de la méthode est le niveau de zoom fixe centré sur l'artiste. Un premier prototype de visualisation présentant une vue globale de tous les artistes a été conçu. Pour l'exploration d'un large jeu de données, il serait pertinent d'évaluer sa pertinence dans un contexte exploratoire et de l'intégrer avec MuzLink. D'autre part, ajouter des fonctionnalités de filtres et de zoom pourrait grandement améliorer la découverte.

La méthode de visualisation proposée n'est pas limitée au domaine de la musique. Avec quelques modifications, nous croyons qu'elle pourrait aisément s'appliquer à des jeux de données similaires, citons à titre d'exemple les relations entre publications scientifiques et les relations entre œuvres cinématographiques. Une autre avenue intéressante serait d'explorer comment les lignes du temps connectées pourraient constituer une méthode de visualisation générale pour la représentation d'ensembles interconnectés et mutuellement exclusifs dans le temps.

ABSTRACT

The rise of Linked Open Data in the recent years is democratizing access to complex data structures, usually taking the form of very large multivariate graphs. These are typically hard to understand for humans and require the use of visualization tools. While this data structure is a very well-researched topic in data visualization, large and complex graphs — with several heterogeneous node and link attributes — are hard to visualize with existing techniques.

During a project in collaboration with the national library and archives agency of the Quebec government, we have been mandated to visualize a complex dataset of relationships between artists. Eleven key exploratory tasks were selected during the initial design process. Standard graph visualization techniques weren't suited to answers those specific tasks. Furthermore, since the dataset's structure is an unusual combination of a bipartite graph and trees, a thorough literature review found very few related or similar work. This research proposes a novel visualization method called MuzLink to efficiently explore and discover relationships between artists. The objective is to evaluate the effectiveness of this method for the selected task with a formal user study.

Based on all eleven selected tasks, a fully functional prototype was developed. The proposed solution is a multi-facet visualization tool centered around a specific artist. The first view contains 3 connected timelines: (1) the inspirations, (2) the productions, and (3) the influences. Beeswarms of songs are positioned on the timelines according to their relationship with the artist. Songs on the middle timeline are the artist's production. Songs that inspired the artist are on the top timeline. Songs inspired by the artist are on the bottom timeline. Links are drawn between songs having an influence relationships. This connected timeline allows a topological exploration of inspiration relationships around an artist's musical production. The second view shows inspired, collaborators, and influenced artists ordered by their number of implications with the current artist. Interactions between the two complementary views allow answering a larger range of tasks. The user can easily navigate between artists with a search bar and by clicking on related artists. Clicking on a song reveals more details regarding the relationships among other songs.

A formal user study was conducted with 18 participants aged between 18 and 70 years old. It is divided into three main sections. The first is an onboarding process which aims to impart the tool functionalities and our music vocabulary to the participants. Next, the participants must answer a set of 11 timed questions using MuzLink. At the end, they can freely explore

the dataset before answering an open-ended survey.

The results of the study revealed that MuzLink is efficient at answering all 11 selected tasks with speed and accuracy. Open-ended questions at the end of the study have shown that the tool is intuitive, shows an excellent level of discovery, and communicate satisfying insights. Moreover, an analysis of the activity logs showed a great level of exploration.

MuzLink's biggest weakness is the lack of general context and filters. Using Schneiderman's mantra could greatly improve the explorability of the tool by giving users a better context and more control.

The proposed method is not limited to the field of music. We believe it could be applied on similar datasets such as cinema, scientific publications, and open source projects. Moreover, future work could be done on the connected timelines as a general visualization method. We believe its rich visual language may be well suited to represent interconnected sets over time.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xii
LISTE DES FIGURES	xiii
LISTE DES SIGLES ET ABRÉVIATIONS	xvi
LISTE DES ANNEXES	xvii
CHAPITRE 1 INTRODUCTION	1
1.1 Problématique	1
1.2 Objectifs	3
1.3 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Cadre d'analyse	4
2.2 Considérations de design	5
2.2.1 Marques et canaux de communication	5
2.2.2 Interactivité	6
2.3 Techniques de visualisation usuelles	7
2.3.1 Visualisation de graphes multivariés	8
2.3.2 Visualisation d'arbres	8
2.3.3 Visualisation de lignes du temps	10
CHAPITRE 3 DÉMARCHE	12
3.1 Traitement et définitions des données	13
3.1.1 Désambiguïsation et augmentation des données	13

3.1.2	Structure de données	14
3.1.3	Glossaire	16
3.1.4	Ordres de grandeur	18
3.2	Abstraction des tâches	18
3.3	Revue de littérature	19
3.4	Design et prototypage	19
3.4.1	Prototype 1 : Représentation matricielle	19
3.4.2	Prototype 2 : Représentation centrée sur l'artiste - Graphe à bandes .	21
3.4.3	Prototype 3 : Représentation en lignes du temps connectées	23
3.4.4	Prototype 4 : Représentation en lignes par niveaux	24
3.4.5	Prototype final : MuzLink	25
3.4.6	Implémentation	25
3.5	Évaluation	27
3.5.1	Recrutement des participants	27
3.5.2	Déroulement de l'expérience	30
3.6	Pertinence de l'article	32

CHAPITRE 4 ARTICLE 1: MuzLink - CONNECTED BEESWARM TIMELINES FOR VISUAL ANALYSIS OF MUSICAL ADAPTATIONS AND ARTIST RELATIONSHIPS

4.1	Abstract	33
4.2	Introduction	33
4.3	Design Process and Task Abstraction	37
4.3.1	User-Centered Questions	39
4.3.2	Task Abstraction	40
4.4	Related Work	41
4.4.1	Multivariate Graph Visualization	41
4.4.2	Tree Visualization	42
4.4.3	Timeline Visualization	42
4.4.4	Academic Publication Visualizations	43
4.5	MuzLink	43
4.5.1	Entry Point	44
4.5.2	A Multi-Part Visualization	44
4.5.3	Interactions	48
4.5.4	Onboarding	49
4.6	Implementation	50

4.7	User Study	50
4.7.1	Experiment overview	50
4.7.2	Results	51
4.7.3	Discussion	55
4.8	Conclusion & Future Work	55
4.9	Acknowledgments	56
CHAPITRE 5 DISCUSSION GÉNÉRALE		57
5.1	Échantillon	57
5.2	Engagement utilisateur	58
5.3	Commentaires des participants	59
5.4	Vue globale et contexte	60
5.5	Limites de l'évaluation	61
CHAPITRE 6 CONCLUSION		63
6.1	Synthèse des travaux	63
6.2	Limitations de la solution proposée	64
6.3	Améliorations futures	64
RÉFÉRENCES		65
ANNEXES		70

LISTE DES TABLEAUX

Tableau 3.1	Attributs de chansons et d'artistes.	15
Tableau 3.2	Glossaire musical : il permet de mieux communiquer les particularités du jeu de données	17
Tableau 3.3	Questions pour l'analyse du jeu de données. Pour chaque question, on spécifie la structure de données (Arb : tree, G : graphe biparti) ainsi que les items (Art : artiste, C : chanson) utilisés.	20
Table 4.1	Song and artist attributes.	36
Table 4.2	Music glossary: To better communicate the dataset particularities throughout this article, we use the musical glossary below. We will often refer to these terms throughout this paper.	38
Table 4.3	Questions for the analysis of song adaptations data. For each, we detail to which layer of the data structure (T: tree, G: bipartite graph) and which part of the bipartite graph (A: artist, S: song), it pertains. . . .	40
Table 4.4	Tasks used during the controlled user study. Each task is a concrete and assessable version of a question in Table 4.3.	52

LISTE DES FIGURES

Figure 1.1	Graphe biparti et arbres. Les liens du graphe sont gris et les liens des arbres sont rouges. Les artistes sont en bleu à gauche et les chansons en orange à droite.	2
Figure 2.1	Techniques de représentation de graphes multivariés telles que présentées dans la revue de littérature de Nobre et al. [1].	9
Figure 2.2	Techniques de représentation d'arbres les plus courantes	11
Figure 3.1	Graphe biparti et arbres. Les liens du graphe sont gris et les liens des arbres sont rouges. Les artistes sont en bleu à gauche et les chansons en orange à droite.	16
Figure 3.2	Représentation en réseau multivarié de la figure 3.1. En fusionnant les deux structures de données, il est possible de générer un graphe multivarié entre artistes. Chaque lien représente une relation autour d'une chanson spécifique. Les rôles des artistes sont représentés sur les extrémités des liens.	18
Figure 3.3	Prototype 1 : Matrice des relations entre artistes	22
Figure 3.4	Prototype 2 : Représentation centrée sur l'artiste à l'aide de graphes à bandes	22
Figure 3.5	Prototype 3 : Représentation en lignes du temps connectées	24
Figure 3.6	Prototype 4 : Représentation en lignes du temps par niveaux	26
Figure 3.7	Pipeline de transformation des données, entre le format Resource Description Framework (RDF) jusqu'au déploiement du terminal consommé par MuzLink	27

Figure 4.1 MuzLink is an artist-centered interactive visualization tool that allows users to explore collaboration and inspiration relationships between artists. Here, the user is exploring French singer Jacques Brel. The tool is divided into 4 juxtaposed and interactive views. First, the Connected Timelines (1): the middle one gathers Jacques Brel’s songs, the top one gathers all songs that inspired him and the bottom one gathers all songs that were inspired by him. In a similar fashion, the Artist Lists (2) show artists who inspired, collaborated with, or were influenced by Jacques Brel. Clicking on a song — in this case "La chanson des vieux amants" — reveals its Detail Sheet (3). The Artist Overview (4) summarizes Jacques Brel’s production. Edge cases are highlighted at the top right corner (5). 34

Figure 4.2 Bipartite graph and tree representation. The graph edges are grey and the tree edges are red. The artists are in blue on the left and the songs in orange on the right. 37

Figure 4.3 Multivariate network representation of Fig. 4.2. By merging the two data structures, it is possible to create a single network between artists. Each edge represents a relationship derived from song. Artists’ roles are shown on both ends of the link. 39

Figure 4.4 Connected Timelines: Inspirations (1), Productions (2) and Influences (3) 45

Figure 4.5 Artist Lists are divided into 3 columns where each column contains related artists ordered by the number of relations. The first column lists the top inspirators, the second lists the top collaborators, and the last lists the top influenced. The user can filter the list by roles (1). Each artist on the list has a small bar chart showing the role distribution (2). 46

Figure 4.6 The Detail Sheet is a small card showing more details about the selected song: the release year, the country, the language, the artists and their roles, the source song, the number of adaptations, the number of and covers, and the full adaptation tree. 48

Figure 4.7 The Artist Overview shows the type of productions the artist made and the roles he bore over his career. 48

Figure 4.8 Score for each task during the experiment. Successful tasks are in green, failed tasks in red and skipped tasks in grey. 53

Figure 4.9	Distribution of the time it took for participants to complete each task during the experiment. Outliers are shown as dots.	54
Figure 4.10	Quantitative feedback score distribution for each aspect. Outliers are shown as dots.	54
Figure 5.1	Genre des participants	57
Figure 5.2	Niveau d'éducation des participants	58
Figure 5.3	Groupes d'âge des participants	58
Figure 5.4	Distribution des interactions enregistrées pour chaque participant lors de la période exploratoire de 8 minutes.	59
Figure 5.5	Vue globale présentant tous les artistes du jeu de données de Bibliothèque et Archives nationales du Québec (BAnQ). Figure tirée de l'article «MusX : Online Exploring and Visualizing Graph-Based Musical Adaptations» par Lévesque et al. [2]. Chaque bulle représente un artiste. Sa taille varie en fonction du nombre de productions. La position encode la direction des inspirations de l'artiste. L'axe y encode le nombre de relations d'inspirations : le nombre de fois qu'un artiste s'est inspiré d'un autre. L'axe x encode le nombre de relations d'influences : le nombre de fois qu'un artiste a été influencé par d'autres.	60

LISTE DES SIGLES ET ABRÉVIATIONS

AWS	Amazon Web Services
BAnQ	Bibliothèque et Archives nationales du Québec
CÉR	Comité d'éthique de la recherche
CSV	Comma-separated values
HTTPS	Hypertext Transfer Protocol Secure
LOC	Linked Open Data
RDA	Resource Description and Access
RDF	Resource Description Framework
SVG	Scalable Vector Graphics

LISTE DES ANNEXES

Annexe A	Cas d'utilisation	70
Annexe B	Cas d'utilisation abstraits	72
Annexe C	Annonce de recrutement	74
Annexe D	Courriel de recrutement	75
Annexe E	Autorisation d'accès aux lieux	76
Annexe F	Formulaire de consentement	77
Annexe G	Questionnaire	81
Annexe H	Résultats bruts	82
Annexe I	Réponses aux questions ouvertes	83

CHAPITRE 1 INTRODUCTION

La montée des Linked Open Data (LOC) dans le domaine culturel démocratise l'accès à des ensembles de données complexes qui prennent généralement la forme de grands graphes multivariés et multicouches. Ces jeux de données sont souvent difficiles à comprendre pour un humain sans un outil de visualisation adapté. Bien que la visualisation de graphes soit un sujet de recherche très actif, les techniques existantes perdent en efficacité lorsque les graphes sont multivariés (plusieurs attributs hétérogènes sur les noeuds ou les liens) et larges (plus de 100 items) [1].

Lors d'un projet d'innovation et de valorisation de données en partenariat avec BAnQ, nous avons travaillé avec un jeu de données particulièrement complexe répertoriant les adaptations entre chansons. L'objectif de ce projet d'innovation était de valoriser leur base de données et de propulser l'exploration et la découverte de relations, pour finalement encourager l'emprunt d'albums musicaux dans la bibliothèque. Un processus de design a permis de dégager un ensemble de questions clés auxquelles notre méthode de visualisation devait répondre. Il doit permettre de comprendre comment les artistes collaborent et s'influencent au cours de leur carrière, et de reconnaître certaines valeurs aberrantes (par exemple les chansons les plus adaptées ou les artistes les plus influents).

1.1 Problématique

Le jeu de données contient deux types d'items : des chansons et des artistes. On y compte près de 24 000 chansons et 19 000 artistes. Les chansons contiennent des métadonnées courantes comme le titre et la date de parution, mais aussi certains attributs plus remarquables :

- **Liste d'auteurs et de compositeurs** : Tous les auteurs et compositeurs qui ont participé à la production de la chanson.
- **Liste d'interprètes principaux** : Tous les interprètes qui sont considérés comme principaux interprètes. Il s'agit généralement des premiers interprètes à avoir chanté la chanson.
- **Liste d'interprètes de reprise** : Tous les interprètes qui ont fait des reprises de la chanson. Un artiste ne peut pas être à la fois interprète principal et interprète de reprise.
- **Chanson parent** : Si la chanson est une adaptation, la chanson parente représente la source de l'adaptation. Toutes les chansons dans le jeu de données ont une chanson

parente, ont plusieurs adaptations ou les deux.

Si l'on respecte l'abstraction de données de Munzer [3], le jeu de données est un mélange inhabituel entre un **graphe biparti** et des **arbres**.

Le graphe biparti est composé d'artistes d'un côté et de chansons de l'autre. Un lien entre une chanson et un artiste représente une participation. Cette participation est caractérisée par un rôle, soit celui d'auteur, de compositeur, d'interprète principal ou d'interprète de reprise. Ce lien est représenté en gris dans la Figure 1.1.

Le jeu de données contient également plusieurs arbres d'adaptations. Ces derniers sont issus des relations d'adaptations entre chansons. Chaque chanson du graphe biparti fait nécessairement partie d'un arbre. Un arbre possède au maximum 4 niveaux de générations. Les liens des arbres sont représentés en rouge dans la Figure 1.1.

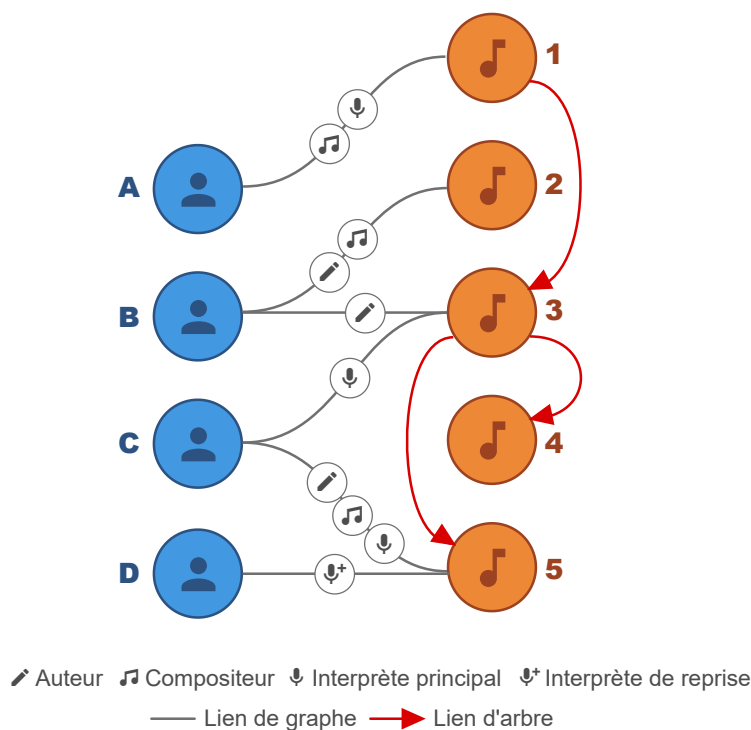


Figure 1.1 Graphe biparti et arbres. Les liens du graphe sont gris et les liens des arbres sont rouges. Les artistes sont en bleu à gauche et les chansons en orange à droite.

La double nature graphe et arbre du jeu de données rend difficile l'application des techniques de visualisation actuelles comme nous le montrerons dans le Chapitre 2. Pour les quelques techniques qui s'appliquent, de nombreuses tâches exploratoires sur les relations entre artistes ne peuvent pas être répondues.

1.2 Objectifs

L'objectif premier de ce projet de recherche est de concevoir une nouvelle méthode de visualisation qui permet l'exploration et la découverte de relations entre artistes de la base de données de BAnQ de façon précise et rapide. La méthode doit être accessible aux utilisateurs profanes et doit nécessiter un apprentissage minimum. À l'aide d'un prototype fonctionnel, nous souhaitons démontrer la performance de cette méthode à répondre aux questions avec une étude utilisateur formelle.

Ce projet de recherche constitue d'abord et avant tout une étude de cas sur un jeu de données précis. Partager en détail le processus de design et les résultats d'un tel projet pourra alimenter les recherches futures avec des jeux de données et tâches exploratoires similaires.

1.3 Plan du mémoire

Ce mémoire est un mémoire par article. Le prochain chapitre présente la revue de littérature, qui dresse un aperçu des techniques existantes en représentation de graphes, d'arbres et de lignes du temps dans la littérature. Le chapitre 3 présente la démarche de l'ensemble du travail de recherche. Le chapitre 4 contient l'article scientifique associé à ce travail de recherche. Le chapitre 5 est une discussion générale sur la méthodologie et sur les résultats en lien avec la revue de littérature. Enfin, le chapitre 6 résume le travail de recherche, présente ses limitations et propose des recommandations pour des recherches futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre est une introduction aux concepts clés en visualisation de données dans le contexte du projet de recherche. Il décrit d’abord le cadre d’analyse «what, why and how» qui a guidé le développement de MuzLink. Il présente ensuite les considérations de design de base en visualisation : le concept de marques et de canaux, et le concept d’interactivité. Il termine en présentant les techniques usuelles de représentation de graphes, d’arbres et de lignes du temps en visualisation de données. Une revue de littérature plus en profondeur sur les travaux reliés est présentée dans l’article à la section 4.4.

2.1 Cadre d’analyse

Munzner définit un cadre pour la conception de visualisations [3]. Il est divisé en trois grandes sections qui prennent la forme de questions : «what», «why» et «how».

What Il y a d’abord le «what». Il s’agit essentiellement de l’étape de compréhension et de définition des données. Quelles sont les données actuelles? Comment sont-elles organisées? Quel est l’ordre de grandeur? Nous avons déjà observé le «what» dans l’introduction — un graphe biparti composé d’arbres d’adaptations, où chaque lien et noeud possède un ensemble d’attributs hétérogènes.

Why Ensuite, il y a l’étape du «why», essentiellement à quoi sert la visualisation. Elle débute avec une identification de l’utilisateur cible, ses contraintes et son niveau de familiarité avec les outils de visualisation. Selon ses besoins, on peut alors définir et *abstraire* les tâches les plus importantes. Dans notre cas, comme présenté dans le Tableau 3.3, MuzLink doit permettre l’*exploration* et la *découverte* de relations entre artistes dans le temps. Il doit permettre de *repérer* et *comparer* des *tendances* entre artistes.

How Finalement, l’étape du «how» constitue les choix de design et d’interaction qui aboutissent à la conception d’un outil de visualisation. Elle débute avec une identification des contraintes techniques : format, accessibilité, ratio et taille d’écran, latence, etc. Des prototypes de plus en plus fidèles sont ensuite conçus de façon itérative.

Alors que le «what» et le «why» ont déjà été explorés dans l’introduction, la prochaine section présentera le «how» avec les considérations de design qui ont guidé le produit final.

2.2 Considérations de design

Cette section présente plusieurs considérations de design importantes qui ont guidé la conception de la méthode de visualisation. Elle présente d’abord les marques et canaux de communications les plus efficaces pour encoder des valeurs. Elle présente ensuite quelques indications essentielles pour l’ajout d’interactions. Finalement, elle démontre les avantages de la séparation d’aspects en vues.

2.2.1 Marques et canaux de communication

Une visualisation de données peut être formalisée comme étant composée de marques géométriques dont certaines variables stylistiques, appelées canaux, sont utilisées pour encoder de l’information à travers leurs valeurs [3]. Une marque peut être un point, une ligne ou une surface. Cette dernière peut être stylisée pour encoder une information à l’aide d’un canal. Les plus utilisés sont la position, la couleur, la forme géométrique, l’angle, la longueur, la grandeur et le volume. On distingue 5 caractéristiques quant à l’efficacité des canaux [4] :

- Discriminabilité : La facilité avec laquelle un canal peut permettre de différencier des marques.
- Associativité : L’efficacité avec laquelle un canal peut grouper des marques, même si les autres canaux changent.
- Précision : L’efficacité avec laquelle un canal peut quantifier la différence entre des marques.
- Ordonnabilité : L’efficacité avec laquelle un canal peut ordonner des marques.
- Amplitude : La quantité de valeurs différentes qu’un canal peut encoder.

Le canal de la position est le plus efficace pour les 5 caractéristiques. Il permet une grande précision, facilitant la discrimination de valeurs sur une très large amplitude. Le canal de la longueur est aussi très efficace. En deuxième dimension, l’utilisation de la surface perd en efficacité. Elle est généralement moins précise, rendant moins évidente la comparaison entre deux marques. La luminance (l’intensité de lumière) et la saturation ont une efficacité semblable. Ils sont généralement recommandés lorsque les canaux de position et de taille sont déjà utilisés. Leur précision n’est pas très grande, ils ont donc une amplitude limitée. La teinte est adaptée pour la représentation d’attributs catégorisables et non ordonnés. Ils ont une très faible amplitude, on recommande donc d’utiliser la teinte lorsque le nombre de catégories est petit (moins de 10 environ). La forme est aussi adaptée pour la représentation de catégories non ordonnées. Comme la couleur, la forme a une amplitude limitée. Cependant, elle a une moins bonne associativité : il est généralement plus facile d’associer deux marques ayant la

même couleur, que deux marques ayant la même forme, comme en témoigne l’effet visuel du «popup» décrit par Munzner [3].

2.2.2 Interactivité

Munzner définit trois types de manipulations de vues : les modifications temporelles, la sélection et la navigation [3]. Chaque modification possède des avantages et des contextes d’utilisation qui leurs sont propres. La manipulation d’une vue est essentielle pour comprendre des jeux de données qui sont trop gros ou complexes pour être compris dans leur ensemble. Une manipulation permet de simplifier la vue en mettant un aspect en valeur.

Modifications temporelles

La modification temporelle est la manipulation la plus courante. Il s’agit d’un changement dans l’encodage des canaux, d’un réarrangement, d’un triage, d’un changement d’un point de vue, etc. Ces reconfigurations permettent de montrer les données sous différents idiomes pour accomplir un plus grand nombre de tâches.

Lorsqu’une visualisation montre des séries de données de façon progressive, l’animation entre les différents états permet à l’utilisateur de garder le contexte pendant la transition entre les deux états [5].

La transition peut être automatique, ou contrôlée par l’utilisateur. Selon la degré de complexité de l’outil, il peut être possible de combiner soi-même des encodages. Cela est fréquent dans les outils de visualisation générale, comme GraphViz¹ ou Gephi².

Sélections

La sélection est une interaction fondamentale en visualisation de données. Une sélection peut se faire sur un item du jeu de données. Elle peut également se faire sur un lien, dans le cas d’un graphe ou d’un arbre. Une interaction de sélection peut aussi fonctionner sur un ensemble d’items partageant des attributs similaires. La sélection peut se faire avec un clic ou un survol de souris, permettant deux états de sélection. Avec le clavier et les autres boutons de la souris (clic de molette par exemple), le nombre d’états peut grandement augmenter. La sélection peut se produire sur un item à la fois — chaque sélection provoque la désélection de l’ancienne sélection — ou sur plusieurs items à la fois — les sélections s’accumulent. Ce comportement est souvent combiné avec un «toggle» qui permet de désélectionner en

1. <https://www.graphviz.org/>

2. <https://gephi.org/>

resélectionnant. Certains outils, selon les tâches requises, possèdent des fonctionnalités de sélection plus avancées : création de groupes de sélection, possibilité de vider des groupes, de déplacer des items entre les groupes, etc. [3]

La sélection devrait toujours donner une rétroaction visuelle, notamment changer d'apparence pour sortir du lot, afin de confirmer l'action de l'utilisateur et éviter la confusion. Cette mise en évidence doit provoquer un changement visuel assez grand pour que l'utilisateur puisse le remarquer. Lors d'une mise en évidence par changement de couleur par exemple, la nouvelle couleur doit avoir un niveau de luminance, de saturation ou une teinte suffisamment différente de l'ancienne pour être perçue. Un autre choix de mise en évidence est l'animation, en faisant osciller les items ou en ajouter des lignes pointillées animées par exemple. C'est une pratique plus rare, mais les travaux de Ware et Bobrow confirment que la méthode est souvent plus performante que les méthodes traditionnelles [6].

L'usage de la sélection est multiple. Elle peut fournir davantage de contexte : un survol peut afficher une infobulle, un clic peut afficher une notice ou révéler d'autres items reliés, etc. Dans une visualisation multifacette, elle permet de voir un item, ou un ensemble d'items, sous toutes les facettes avec aisance.

Navigations

Lorsque les jeux de données sont très grands, il est pertinent de permettre à l'utilisateur de changer de point de vue. La navigation peut être divisée en trois composantes : le zoom, la translation et la rotation [3]. Le zoom permet de contrôler le niveau de détail. Il donne la possibilité à l'utilisateur d'explorer toutes les données d'un point de vue général, puis d'explorer plus en détail certaines parties. La translation permet de naviguer dans le jeu de données lorsqu'un niveau de zoom est appliqué. Plus souvent utilisée dans un contexte en trois dimensions, la rotation permet de voir un jeu de données sous différents angles.

2.3 Techniques de visualisation usuelles

Cette section catalogue les méthodes de visualisation les plus courantes pour la représentation de graphes multivariés, d'arbres et de lignes du temps, trois structures de données reliées au jeu de données de BANQ.

2.3.1 Visualisation de graphes multivariés

La figure 2.1 — inspirée par les travaux de Nobre et al. [1] — regroupe les techniques de représentation de graphes en trois catégories : les «node-links», les matrices et les arbres implicites.

Les «node-links» constituent la représentation de graphes la plus explicite : les noeuds et les liens sont représentés directement. Ils ont une configuration topologique, lorsque les noeuds et les liens sont positionnés pour diminuer les croisements de liens. Cette représentation est la plus adaptée pour les tâches topologiques, où les relations — ou la hiérarchie — doivent être visuellement explicites. Les attributs peuvent alors être encodés directement sur les liens ou les noeuds en utilisant les canaux d’aire, de longueur ou de couleur. Si les attributs de noeuds sont ordinaux, une configuration par facettes est efficace pour grouper des noeuds entre eux. Il est également possible d’utiliser le canal de la position des axes pour encoder des attributs continus avec une configuration par positionnement.

Les matrices permettent de représenter davantage de noeuds dans un espace restreint. Ils n’explicitent pas les liens, rendent plus difficiles les tâches topologiques. Cependant, en simplifiant la vue, ils permettent de représenter davantage d’attributs. La matrice d’adjacence permet d’exposer toutes les relations entre les noeuds du graphe. Chaque case représente une relation, et peut utiliser un ou plusieurs canaux pour encoder des attributs. Lorsqu’il existe peu de liens entre les noeuds, la matrice d’adjacence contient de nombreuses cases vides. La courtepointe permet de représenter les relations entre les noeuds les plus reliés pour des graphes clairsemés. Le brio-fabric est une autre technique moins étudiée qui permet de représenter des graphes clairsemés, mais peut représenter des attributs hétérogènes sur les noeuds et les liens.

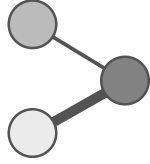
Un graphe peut être dérivé en arbre implicite pour faciliter la lecture. Outre la configuration topologique, un arbre peut être visualisé à l’aide d’une treemap ou d’une représentation radiale des feuilles. Ces techniques seront explorées dans la section suivante portant sur la visualisation d’arbres.

2.3.2 Visualisation d’arbres

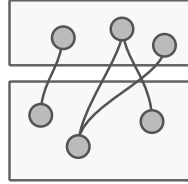
Un arbre peut-être représenté de façon explicite — en affichant directement les noeuds et les liens — ou de façon implicite — en affichant seulement les feuilles [7] (figure 2.2). La représentation explicite positionne les noeuds sur des axes parallèles, de façon radiale ou dans une configuration libre. Cette représentation est la plus efficace pour accomplir des tâches topologiques, mais consomme beaucoup d’espace lorsque la quantité de noeuds augmente.

Node-link

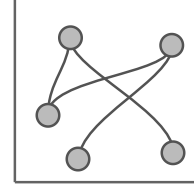
┌ Configuration **topologique** ─ ─ Configuration par **attribut** ─ ─



Encodage sur liens & nœuds

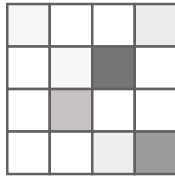


Facettes

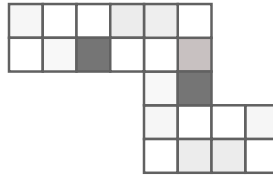


Positionnement

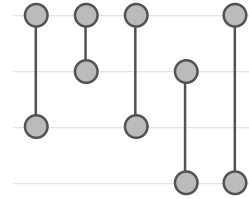
Matrices



Matrice d'adjacence

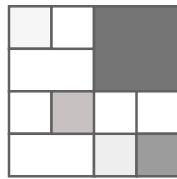


Courtepointe

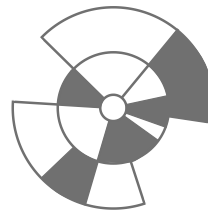


BrioFabric

Arbre implicite



Treemap



Radial

Figure 2.1 Techniques de représentation de graphes multivariés telles que présentées dans la revue de littérature de Nobre et al. [1].

Pour représenter davantage de noeuds et de niveaux, les représentations implicites sont souvent plus adaptées. Dans une treemap, les noeuds sont agrégés en boîtes imbriquées l'une dans l'autre. Bien que cette méthode ne montre pas les liens entre les noeuds, elle met clairement en évidence les catégories hiérarchiques. La représentation implicite radiale possède des avantages similaires à ceux de la treemap, mais utilise une représentation radiale.

2.3.3 Visualisation de lignes du temps

Une ligne du temps peut être représentée de façon linéaire, radiale, en spirale ou irrégulière [8]. Alors que la représentation est la plus adaptée pour représenter des items qui se succèdent chronologiquement de façon séquentielle, les représentations radiales ou en spirale sont plus adaptées pour des items qui sont périodiques (annuels ou mensuels par exemple). Les items positionnés ont un horodatage et peuvent optionnellement avoir une durée, laquelle est habituellement représentée avec le canal de longueur.

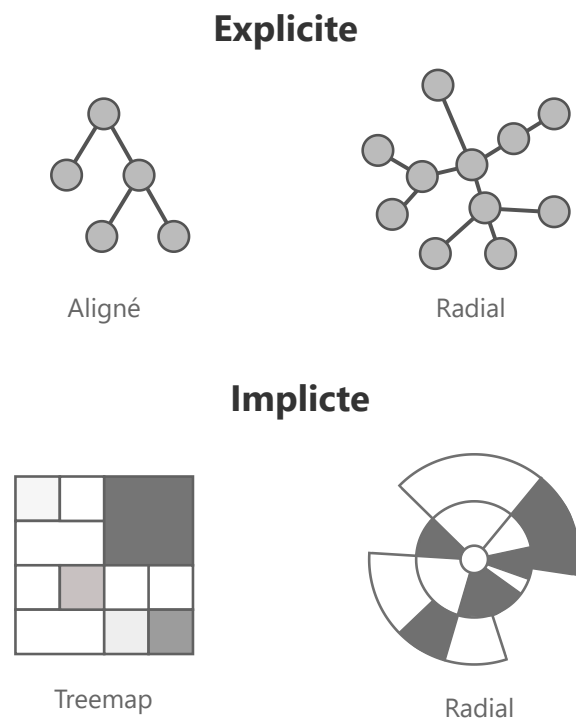


Figure 2.2 Techniques de représentation d'arbres les plus courantes

CHAPITRE 3 DÉMARCHE

Ce chapitre relate en détail les 5 étapes de la démarche du projet de recherche : définition des données, identification des tâches, revue de littérature, prototypage et évaluation. Il explique également en quoi l'article présenté au chapitre suivant fait partie intégrante de ce projet de recherche.

Définition des données Cette étape, analogue à l'étape «what» de Munzner, vise à définir et comprendre la structure de données. Elle définit également l'échelle de grandeur du jeu de données complet. Elle définit aussi une taxonomie — un vocabulaire spécifique au domaine — pour faciliter les étapes subséquentes.

Identification et abstraction des tâches Équivalente à l'étape «why» de Munzner, cette étape se concentre sur l'objectif de l'outil de visualisation, c'est-à-dire quelles tâches il permettra d'accomplir. Elle est d'abord réalisée avec des utilisateurs cibles et des experts du domaine des données. Cela permet de recenser une liste de tâches concrètes. En parallèle, une revue de littérature permet de répertorier les tâches «types» associées à certaines structures de données connues. Il s'agit ensuite de transformer et fusionner les tâches concrètes en tâches abstraites — plus larges et non spécifiques. L'objectif est de concevoir un catalogue exhaustif et abstrait de toutes les tâches pertinentes.

Revue de littérature Cette étape est une recherche de travaux similaires ou reliés. Elle permet d'abord de valider la pertinence du problème et de le situer dans le contexte scientifique actuel. Elle permet ensuite de guider la recherche de solutions. Une bonne abstraction des tâches est essentielle pour faciliter la recherche de travaux connexes issus d'autres contextes.

Design et prototypage Le prototypage, analogue à l'étape «how» de Munzner, est essentiellement un processus d'idéation itératif. Il débute avec des explorations rapides sur papier, puis évolue vers des maquettes informatisées de basse à haute fidélité. L'objectif de cette étape est la production d'un prototype fonctionnel pour procéder à la dernière étape d'évaluation. Cette étape s'appuie largement sur les étapes précédentes, en particulier les étapes de définition des données («what») et d'identification des tâches («why»). Dans un contexte de recherche, la revue de littérature permet la conception de modèles innovants pouvant contribuer à la communauté scientifique.

Évaluation Cette dernière étape consiste à évaluer l'efficacité du prototype à accomplir les tâches définies dans l'étape d'abstraction des tâches. Elle prend généralement la forme d'une étude utilisateur formelle. Les résultats compilés permettent de confirmer ou d'infirmer l'atteinte des objectifs. Elle met également en lumière les limites et faiblesses plus subtiles du prototype.

Les 5 étapes sont itératives et ne suivent pas un ordre chronologique. Selon la progression du projet de recherche, des retours en arrière sont souvent indispensables.

3.1 Traitement et définitions des données

La première étape de la démarche de recherche fut un traitement et un nettoyage des données. Ensuite, une analyse de la structure, de l'ordre de grandeur et de la sémantique du jeu de données ont grandement guidé la revue de littérature et l'abstraction des tâches aux étapes subséquentes.

3.1.1 Désambiguïsation et augmentation des données

Pour faciliter le traitement et l'analyse des données, le jeu de données a été nettoyé et désambiguïsé à l'aide de technologies du web sémantique.

Les données brutes de BAnQ ont été fournies sous un format Comma-separated values (CSV). Elles ont été modélisées dans un graphe RDF. Ce modèle a été choisi car il est facile à utiliser avec des LOC comme MusicBrainz¹ ou DBPedia² pour désambiguïser ou augmenter les données. Les deux principales entités sont les **artistes** et les **chansons**. La modélisation RDF se concentre donc sur celles-ci. Il existe des relations entre artistes et chansons (auteur, compositeur et interprète), entre chansons (chanson parente, chanson adaptée) et entre les artistes (membres d'un groupe, partenariats). Le reste des informations sont encodées avec des données littérales (dates, notes, titres, etc.), ou alors avec des liens externes vers des pages DBPedia (pays, langues, genres, artistes, etc.). Bâties dans un contexte bibliothécaire, les données sont modélisées selon le standard du RDA Registry³. Resource Description and Access (RDA) est un standard qui catalogue des normes et règles pour la description de données bibliographiques. La RDA Registry propose un vocabulaire standard qui représente des éléments RDA, des entités et des terminologies en RDF. Le processus d'ingénierie de données de ce projet de recherche est parcouru en détail dans un article dédié [2].

1. <https://musicbrainz.org/>

2. <https://wiki.dbpedia.org/>

3. <https://www.rdaregistry.info/>

L'utilisation d'un graphe RDF permet non seulement de désambiguïser et d'augmenter les données, mais permet également un partage plus rapide avec d'autres LOC. En outre, il a permis la création d'un terminal Protocol and RDF Query Language (SPARQL) pour démocratiser l'accès aux données. Comme nous le verrons dans la section 3.4.6, ce terminal est la source de données du prototype de visualisation final.

3.1.2 Structure de données

Un format abrégé de la structure des données est présenté au tableau 3.1. On retrouve deux principales entités : des **chansons** et des **artistes**. Les chansons ont un identifiant unique, un titre, une langue, un pays d'origine, une liste d'auteurs, une liste de compositeurs, une liste d'interprètes principaux et une liste d'interprètes de reprise. Un interprète principal est un chanteur qui a chanté la chanson lors de sa production originale, tandis qu'un interprète de reprise est un chanteur qui a repris une chanson sans la modifier de façon substantielle. Une chanson peut avoir un ou plusieurs auteurs, compositeurs et interprètes. Une chanson instrumentale, n'a pas d'auteur ni de compositeur. Un artiste possède un identifiant unique. Il a également un nom, une année de naissance et un pays de naissance.

Un artiste et une chanson peuvent être reliés par une implication dérivée des attributs `author_ids` (implication d'auteur), `composer_ids` (implication de compositeur), `interpreter_ids` (implication d'interprète principal) ou `cover_ids` (implication d'interprète de reprise). Ces relations créent un graphe biparti composé d'artistes et de chansons. Les liens entre ces deux entités représentent les implications caractérisées par un ou plusieurs rôles : auteur, compositeur, interprète principal ou interprète de reprise. Dans la figure 3.1, l'artiste A a interprété et composé la chanson 1. L'artiste B a écrit et composé la chanson 2 et écrit la chanson 3. L'artiste C a chanté en tant qu'interprète principal la chanson 3, et écrit, composé et chanté en tant qu'interprète principal la chanson 5. L'artiste D a repris la chanson 5.

Deux chansons peuvent également être liées ensemble avec l'attribut `parent_id`. C'est une relation d'adaptation. Comme une chanson peut être adaptée plusieurs fois, chaque chanson fait partie d'un arbre d'adaptations. Toutes les chansons faisant partie du graphe biparti font donc également partie d'un arbre d'adaptations. Dans la figure 3.1, les chansons 4 et 5 sont adaptées de la chanson 3. La chanson 3 est elle-même adaptée de la chanson 1.

À partir de cette représentation hybride de données graphe et arbre, il est possible d'observer les relations de *collaborations* entre artistes. Dans la figure 1.1, on remarque par exemple que B a collaboré en tant qu'auteur avec C pour la production de la chanson 3. Alors que les relations entre auteur, compositeur et interprète principal illustrent des relations de *collaboration*, la relation d'interprète de reprise met en lumière une *inspiration*. L'artiste D a repris la chanson

Tableau 3.1 Attributs de chansons et d'artistes.

Attribut		Type	Description
Chanson	id	texte	Identifiant unique de chanson
	title	texte	Titre de la chanson
	language	texte	Paroles de la chanson
	origin_country	string	Pays dans lequel la chanson a été produite
	author_ids	liste	Liste des auteurs impliqués dans la production de la chanson
	composer_ids	liste	Liste des compositeurs impliqués dans la production de la chanson
	interpreter_ids	liste	Liste des interprètes impliqués dans la production de la chanson
	cover_ids	liste	Liste des interprètes qui ont fait des reprises de la chanson
	parent_id	texte	Si la chanson est une adaptation, la chanson parent de laquelle elle a été adaptée
Artiste	id	texte	Identifiant unique de l'artiste
	name	texte	Nom de l'artiste
	birth_year	entier	Année de naissance de l'artiste
	birth_country	texte	Pays de naissance de l'artiste

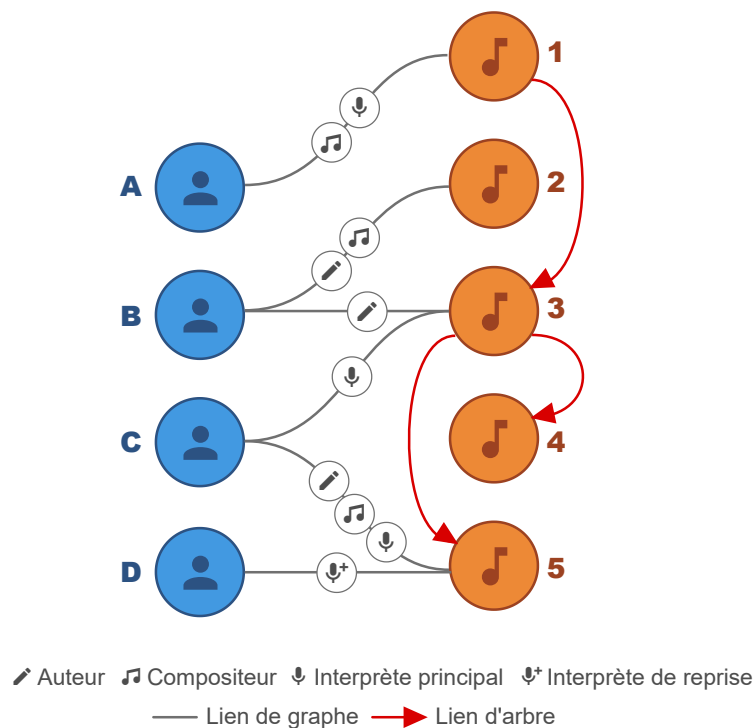


Figure 3.1 Graphe biparti et arbres. Les liens du graphe sont gris et les liens des arbres sont rouges. Les artistes sont en bleu à gauche et les chansons en orange à droite.

5. Il s'est donc *inspiré* de l'artiste C. Ces observations se limitent cependant à l'observation des liens du graphe. Si l'on observe aussi les liens des arbres (en rouge), d'autres relations d'inspiration apparaissent. Par exemple, l'artiste B s'est inspiré de A, car il a adapté la chanson 1, originalement composé et chantée par A.

Pour faciliter l'analyse des relations d'inspiration, il est possible de dériver la structure de données en faisant une abstraction des arbres d'adaptation et en montrant directement les relations entre artistes. Dans la figure 3.2, les relations de *collaboration* et d'*inspiration* directionnelles sont représentées. Cette représentation sous forme de graphe multivarié permet de mieux communiquer la complexité des relations qui existent entre artistes. Chaque lien — ou relation — est généré par une chanson (collaboration de production, adaptation ou reprise).

3.1.3 Glossaire

La définition des données a rendu essentielle la création d'un glossaire (tableau 3.2). Ce dernier permet de communiquer avec concision et précision les particularités du jeu de données.

Tableau 3.2 Glossaire musical : il permet de mieux communiquer les particularités du jeu de données

Mot	Définition
Adaptation	Une chanson adaptée d'une autre chanson. Généralement par le biais d'une réécriture/traduction des paroles ou d'un nouvel arrangement musical.
Chanson source	Chanson à l'origine d'une adaptation.
Chanson originale	Une chanson qui n'est pas une adaptation d'une autre chanson. Une <i>chanson source</i> n'est pas toujours une <i>chanson originale</i> car il est possible d'avoir plus de deux niveaux d'adaptation.
Reprise	La reprise d'une chanson, sans modification suffisamment importante pour être considérée comme une adaptation.
Collaborateur	Un artiste impliqué dans la production de la chanson. Un collaborateur peut être un auteur, un compositeur et/ou un interprète.
Interprète de reprise	Un interprète qui a chanté une reprise de chanson sans être impliqué dans sa production originale. Cet ensemble est mutuellement exclusif avec l'ensemble des <i>collaborateurs</i> .
Productions d'un artiste	Toutes les chansons dans lesquelles l'artiste a été impliqué, en tant qu'auteur, compositeur et/ou interprète. Cet ensemble comprend les reprises chantées par l'artiste.
Inspirations d'un artiste	Toutes les <i>chansons sources</i> des <i>productions de l'artiste</i> .
Influences d'un artiste	Toutes les <i>adaptations</i> ou <i>reprises</i> qui découlent des <i>productions de l'artiste</i> .

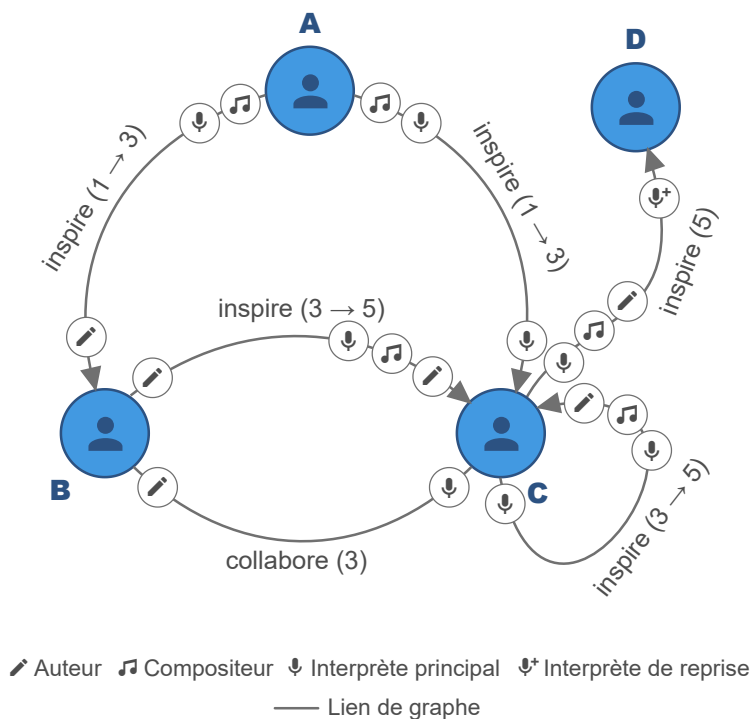


Figure 3.2 Représentation en réseau multivarié de la figure 3.1. En fusionnant les deux structures de données, il est possible de générer un graphe multivarié entre artistes. Chaque lien représente une relation autour d'une chanson spécifique. Les rôles des artistes sont représentés sur les extrémités des liens.

3.1.4 Ordres de grandeur

Le jeu de données contient près de 24 000 chansons et 19 000 artistes. Toutes les chansons font partie d'un arbre d'adaptations. Environ 10 000 sont à la racine d'un arbre, près de 13 500 sont au niveau 2, 800 au niveau 3, 72 au niveau 4 et 14 au niveau 5. Généralement, les arbres sont très petits et n'ont que 2 niveaux. L'implication d'interprète apparaît près de 43 000 fois, celle d'auteur, 32 000 fois, et celle de compositeur, 700 fois. Seulement 20 artistes ont plus de 100 productions et près de 90% en ont moins de 10.

3.2 Abstraction des tâches

Pour concevoir un outil adapté aux demandes de BANQ, nous avons défini les tâches clés que l'outil de visualisation devait être en mesure d'accomplir. En suivant un processus de design orienté utilisateur, des exemples de cas d'utilisation ont d'abord été répertoriés en collaboration avec des experts de l'institution, des libraires et des revues de littérature pertinentes. Ces cas d'utilisations sont catalogués en annexe A. Ils ont ensuite été classés sous 6

angles différents : angle paroles, angle adaptations, angle langues & adaptations, angle pays d'origine, angle spectacles et angle artistes. Ces cas sont abstraits afin d'éviter la redondance (voir annexe B).

Une analyse de ces propositions avec les parties prenantes du projet à BAnQ a mené à la sélection de 11 questions de bases auxquelles devraient être capables de répondre les utilisateurs de l'outil de visualisation (tableau 3.3). Ces questions sont pour la plupart orientées vers l'exploration et la découverte de relations entre artistes dans le temps.

En suivant l'abstraction des tâches de Munzner [3], en ce qui touche aux actions possibles, l'outil doit faciliter la *découverte* de relations entre artistes. Elle doit *présenter* l'information de façon à *divertir* les utilisateurs. La visualisation doit aussi permettre la *localisation* de chansons et d'artistes précis, tout en facilitant la *navigation* et l'*exploration* d'artistes reliés. Il devrait être aisé d'*identifier* et de *comparer* les sphères de collaborations et d'influences des artistes. En ce qui a trait aux données, la méthode de visualisation doit mettre en évidence les *données aberrantes* et les *caractéristiques* pour chaque artiste. Il doit en outre permettre de suivre la *topologie* des arbres d'adaptations dans le temps, tout en faisant ressortir les relations de *dépendance* entre artistes.

3.3 Revue de littérature

La revue de littérature qui a guidé le projet est détaillée dans l'article du prochain chapitre dans la section 4.4. Cette dernière dessine un portrait global des techniques de visualisation existantes reliées à la structure de notre jeu de données et aux tâches définies à la section 3.2. Elle expose en quoi ces techniques sont insuffisantes ou inadaptées à notre étude de cas, et motive la conception d'une méthode innovante.

3.4 Design et prototypage

Les premiers prototypes ont d'abord été produits sur papier. Les plus pertinents ont ensuite été numérisés. Cette section présente les itérations principales de design qui ont mené au prototype final présenté dans l'article à la section 4.5.

3.4.1 Prototype 1 : Représentation matricielle

Un artiste peut avoir 3 relations différentes : une relation de collaboration, une relation d'inspiration (lorsqu'il s'inspire d'un autre artiste) ou une relation d'influence (lorsqu'il inspire un autre artiste). Nous appelons aussi ces relations « niveaux de relation ». Les relations de

Tableau 3.3 Questions pour l'analyse du jeu de données. Pour chaque question, on spécifie la structure de données (Arb : tree, G : graphe biparti) ainsi que les items (Art : artiste, C : chanson) utilisés.

Question	Arb	G	Art	C
1. Qui collabore avec l' artiste ?	✓	✓	✓	
2. Qui inspire l' artiste ?	✓	✓	✓	
3. Qui est influencé par l' artiste ?	✓	✓	✓	
4. Quels rôles un artiste a joués au cours de sa carrière ?		✓		✓
5. Quels rôles les collaborateurs/inspireurs/influenceurs ont joués en relation avec l' artiste ?	✓	✓	✓	✓
6. Qui a le plus inspiré un artiste ?	✓	✓	✓	✓
7. Quel artiste s'est le plus inspiré ?	✓	✓	✓	✓
8. À quel point une chanson est influente comparée aux autres ?	✓			✓
9. Une chanson a-t-elle surtout été adaptée ou reprise ?	✓			✓
10. Qui a adapté ou repris une chanson ?	✓			✓
11. Comment une chanson a-t-elle été adaptée ou reprise au fil du temps ?	✓			✓

collaboration sont des relations de niveau 0, car elles sont sur le même niveau des arbres d'adaptations. Les relations d'inspirations sont des relations de niveau -1, car l'artiste adapte une chanson d'un niveau plus haut dans l'arbre. Inversement, les relations d'influences sont des relations de niveau +1 car la chanson de l'artiste est adaptée dans un niveau plus bas dans l'arbre. Ces trois relations sont caractérisées par un ou plusieurs rôles (auteur, compositeur ou interprète).

Le premier prototype conçu est une matrice d'adjacence $n*n$, où n représente le nombre d'artistes (figure 3.3.1). Chaque case de matrice est divisée en sous-matrices $3*3$ pour représenter toutes les combinaisons possibles entre les 3 rôles et les 3 relations possibles (figure 3.3.2). Chaque case de cette sous-matrice est plus foncée selon la quantité de chansons qui s'y retrouve.

Avantages Cette représentation permet de représenter toutes les relations possibles entre tous les artistes avec un niveau de détail élevé.

Désavantages C'est une méthode qui s'adapte mal lorsque le nombre d'artistes augmente. Montrer tous les artistes nécessite une matrice $19000 * 19000$, ce qui est difficilement lisible. De plus, très peu d'artistes ont des relations entre eux, de sorte qu'il y aurait très peu de cases remplies avec le jeu de données actuel. Elle ne permet pas non plus de représenter les chansons qui causent la fluctuation de teinte, à moins d'utiliser une interaction.

3.4.2 Prototype 2 : Représentation centrée sur l'artiste - Graphe à bandes

L'idée de représenter tous les artistes a été abandonnée pour plusieurs raisons. D'abord, la quantité d'artistes est très grande, ce qui rend difficile la représentation de tous les artistes sur l'écran. De plus, les artistes ont des relations avec un petit nombre d'artistes, les relations entre artistes sont donc très dispersées. Finalement, les questions répertoriées dans le tableau 3.3 nécessitent un grand niveau de détail dans la visualisation, lequel serait possible avec un niveau de zoom plus élevé. Comme les artistes constituent l'unité naturelle de jeu de données et que la majorité des questions sont orientées sur l'artiste, les prochains prototypes montrent tous les données associées à un seul artiste spécifique.

Le deuxième prototype présente une vue centrée sur l'artiste sous la forme d'un grand tableau. Chaque ligne représente un artiste avec une relation de collaboration, d'inspiration ou d'influence. Les artistes qui ont le plus de relations sont positionnés plus haut dans le tableau. Chaque colonne présente un rôle : auteur, compositeur ou interprète. Pour chaque chanson, un carré coloré est positionné dans le tableau selon l'artiste et la relation. Sa cou-

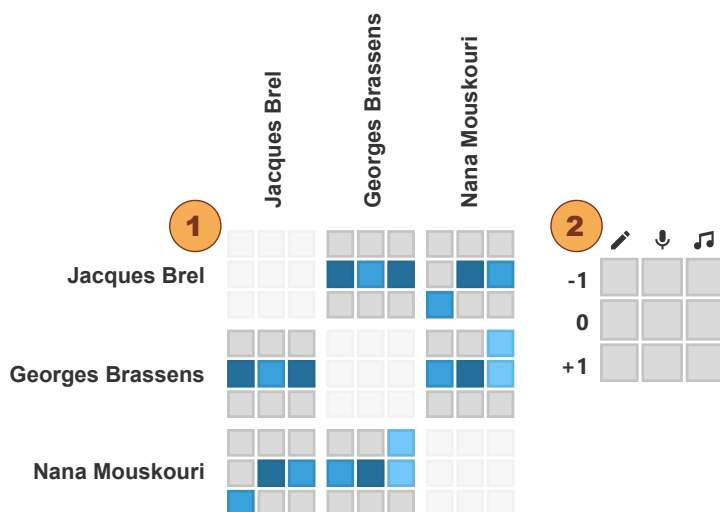


Figure 3.3 Prototype 1 : Matrice des relations entre artistes

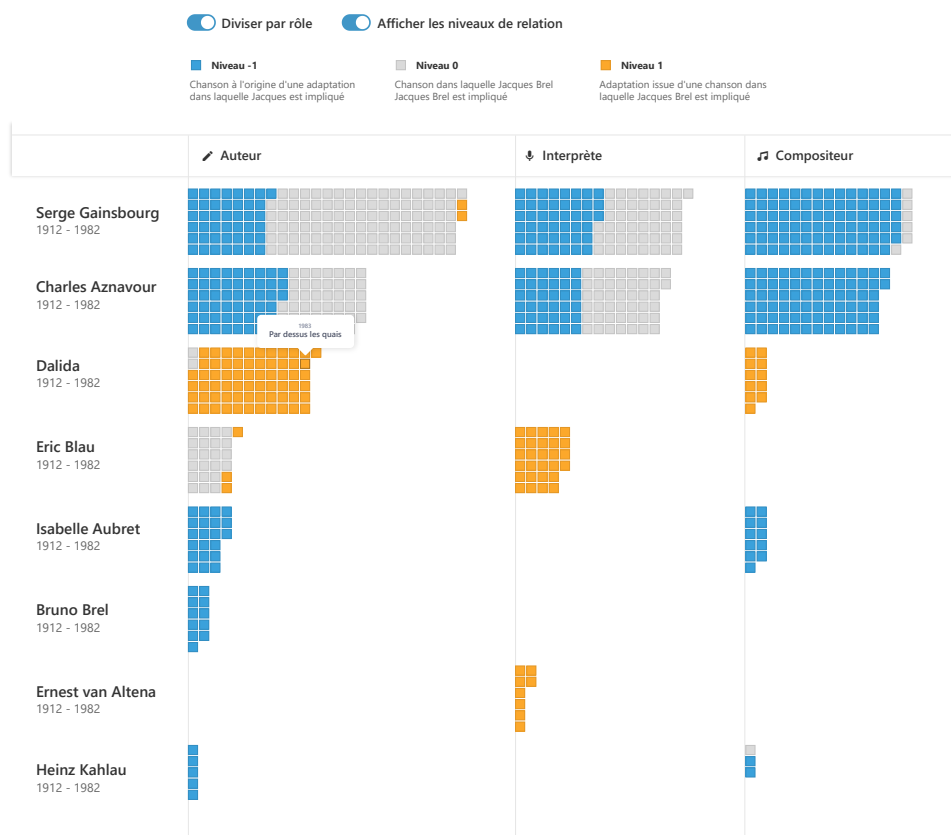


Figure 3.4 Prototype 2 : Représentation centrée sur l'artiste à l'aide de graphes à bandes

leur est modifiée selon le *niveau* de la relation : bleu pour niveau -1, gris pour niveau 0 et orange pour le niveau +1. Dans la figure 3.4, nous explorons les artistes liés à Jacques Brel. La chanson «Par dessus les quais» est positionnée sur la ligne de l'artiste Dalida vis-à-vis la colonne «Auteur» et est colorée en orange. C'est donc une chanson que Dalida a adaptée de Jacques Brel en ayant le rôle d'auteur.

Avantages Ce prototype permet de voir une grande quantité d'informations dans un petit espace. Il permet d'explorer tous les types de relations entre artistes (collaboration, inspiration et influence) en les caractérisant par rôles, tout en présentant les chansons associées.

Désavantages Le principal désavantage est que les chansons sont dédoublées dans chaque case. Par exemple, si Dalida a adapté «Par dessus les quais» en jouant le rôle d'auteur, de compositeur et d'interprète, cette chanson sera dupliquée sur les trois colonnes. Également, si une chanson a été adaptée par 6 artistes, elle sera dupliquée sur 6 lignes. Cela peut rendre la lecture difficile pour un utilisateur non habitué. Il est également difficile de comprendre la généalogie des chansons car elle est uniquement représentée avec de la couleur. Une représentation topologique — en arbre — est nettement plus efficace pour communiquer ce genre d'information. Finalement, si le nombre de lignes dans le tableau est grand, il devient difficile de voir toutes les relations d'un seul coup d'oeil.

3.4.3 Prototype 3 : Représentation en lignes du temps connectées

Le prototype 2 communiquait difficilement les relations topologiques entre chansons. Le prototype 3 résout ce problème en affichant directement les arbres d'adaptation. La visualisation est centrée sur un artiste qui est représenté sur une ligne du temps au centre. Pour chaque chanson, un disque est positionné. Les artistes en relation avec l'artiste sélectionné ont chacun une ligne du temps. S'ils ont produit des inspirations pour l'artiste, ils sont situés en haut de ce dernier. À l'inverse, s'ils ont produit des chansons influencées par des productions de l'artiste, ils seront situés en bas de ce dernier. Des liens sont dessinés entre les productions lorsqu'il existe un lien d'adaptation. Dans la figure 3.5 on observe Nana Mouskouri. La chanson «Par dessus les quais» produite par Nana Mouskouri est adaptée d'une chanson produite par Dalida.

Avantages Ce prototype a l'avantage de présenter de façon topologique toutes les relations d'inspirations et d'influences entre artistes. Pour chaque relation, il est possible de voir la chanson associée.

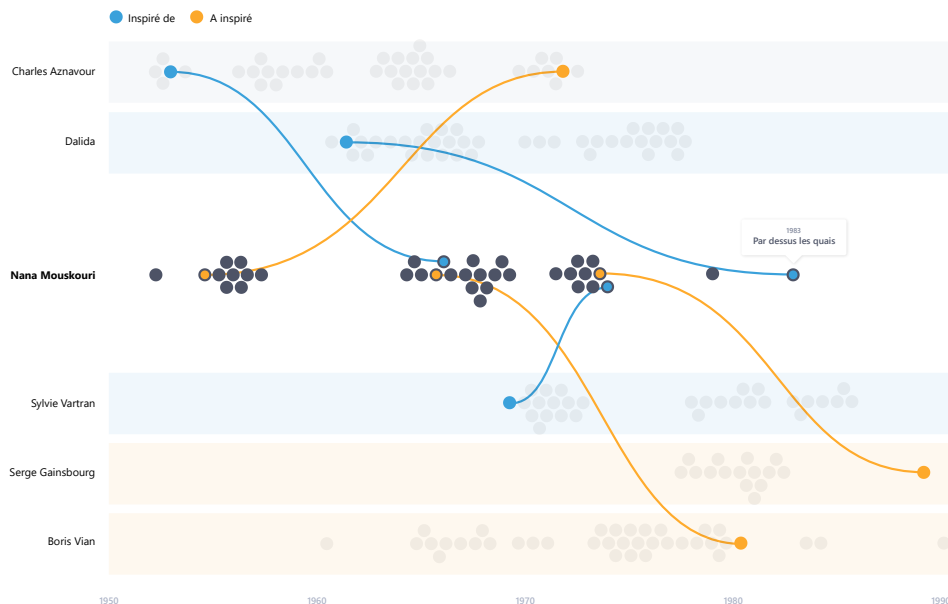


Figure 3.5 Prototype 3 : Représentation en lignes du temps connectées

Désavantages Lorsque la liste d’artistes est longue, les liens entre les chansons sont difficiles à suivre car ils deviennent très longs. De plus, si un artiste est à la fois inspirant et influencé par Jacques Brel, il sera situé soit en bas ou en haut de ce dernier, ce qui porte à confusion. Le problème des chansons dupliquées persiste, une chanson adaptée par 6 artistes va être dupliquée sur 6 lignes du temps. Alors que cette visualisation est efficace pour présenter les relations d’influence entre artistes, elle ne peut pas présenter les relations de collaboration.

3.4.4 Prototype 4 : Représentation en lignes par niveaux

Le quatrième prototype reprend l’idée des lignes du temps connectées du prototype 3. Alors que chaque artiste avait sa propre ligne du temps, dans ce nouveau modèle, il n’y a que 3 lignes du temps. La première contient les chansons qui sont des inspirations pour l’artiste. La deuxième contient les productions de l’artiste. La troisième contient les chansons qui sont des adaptations des productions de l’artiste. Cette nouvelle disposition permet de condenser toutes les chansons dans l’écran tout en évitant la duplication de chansons. Comme ce nouveau positionnement ne permet pas de présenter les artistes, il nécessite l’ajout d’une deuxième vue d’artiste dans le bas de l’écran. Cette dernière permet de lister les artistes ayant le plus inspiré (niveau -1) ou s’étant le plus inspiré (niveau +1) de l’artiste. Une interaction sur un artiste met en évidence les chansons dans lesquelles l’artiste est impliqué.

Chaque artiste est présenté avec un petit graphique à bandes qui résume les rôles joués dans ses implications avec l'artiste exploré.

Avantages Cette visualisation permet de représenter toutes les chansons reliées à un artiste — provenant de relations d'inspiration, de collaboration ou d'influence — dans un espace plus petit. D'un seul coup d'oeil, il est possible de repérer les artistes les plus influents et les plus inspirés dans le tableau du bas. La topologie des arbres d'adaptation est facile plus facile à lire, car les lignes du temps sont rapprochées.

Désavantages Ce prototype ne montre pas les artistes qui ont le plus souvent collaboré avec l'artiste exploré. Il ne fait pas non plus la différence entre adaptations et reprises.

3.4.5 Prototype final : MuzLink

Le prototype final reprend en grande partie le prototype 4 en l'ajustant pour régler ses faiblesses. D'abord, une nouvelle colonne pour les artistes ayant le plus collaboré a été ajoutée au centre du tableau d'artistes. Les lignes du temps ont été divisées en deux : la partie la plus large à gauche rassemble les productions et les adaptations, et la partie de droite affiche les reprises.

D'autres améliorations comme l'ajout d'une légende, d'un tutoriel interactif, d'une fiche de chanson interactive et d'un résumé de l'artiste permettent d'augmenter la performance de l'outil à répondre aux questions du tableau 3.3. L'article au chapitre suivant décrit en détail les choix de design de ce prototype final dans la section 4.5.

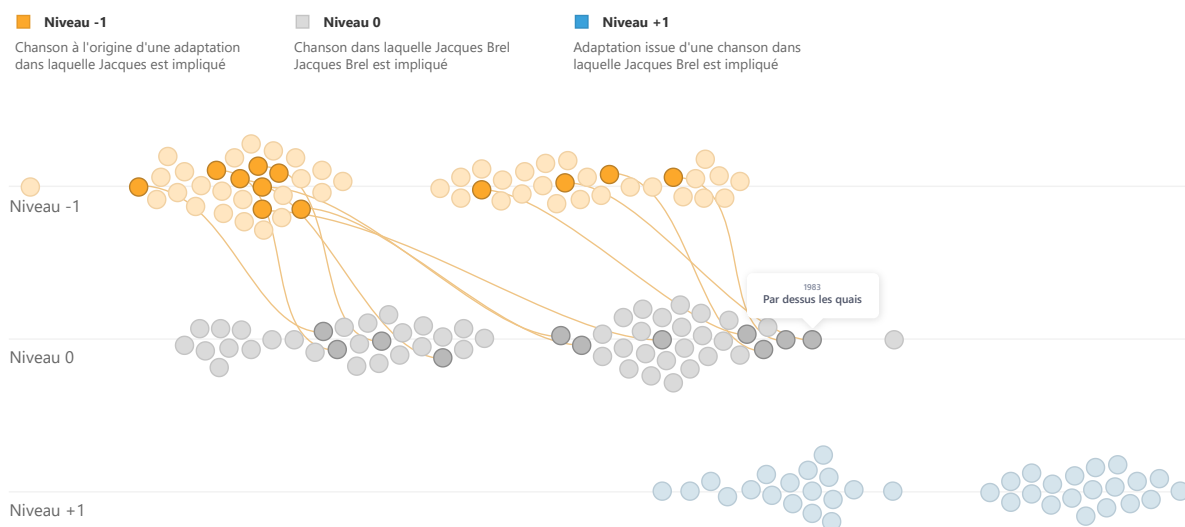
3.4.6 Implémentation

Le prototype final a été développé avec les technologies du web pour faciliter le partage. Les visualisations ont été construites en Scalable Vector Graphics (SVG). Ce format permet de créer de formes vectorielles qui s'adaptent à toutes les tailles d'écrans. La librairie d3.js⁴ a été utilisée pour faciliter la création et l'animation des formes SVG dans un contexte de visualisation de données. Les interactions ont été ajoutées avec la librairie Vue.js⁵. Le prototype final est complètement statique (sans serveur) et est hébergé sur un Content Delivery Network (CDN) à haute performance.

Le prototype effectue une requête au serveur de BANQ lors du chargement initial pour charger

4. <https://d3js.org/>

5. <https://vuejs.org/>



Top des artistes de niveau -1 ⓘ	🎵 📄 🗑️	Top des artistes de niveau +1 ⓘ	🎵 📄 🗑️
Serge Gainsbourg 32 implications		Samuel Quevillon 72 implications	
Charles Aznavour 21 implications		Heinz Kahlau 32 implications	
Dalida 21 implications		Dalida 32 implications	
Eric Blau 21 implications		Charles Aznavour 32 implications	
Isabelle Aubret 21 implications		Ernest van Altena 32 implications	
Bruno Brel 21 implications		Isabelle Aubret 32 implications	
Ernest van Altena 21 implications		Eric Blau 32 implications	

Figure 3.6 Prototype 4 : Représentation en lignes du temps par niveaux

toutes les données. Ce serveur fait alors une requête SPARQL sur le graphe RDF — hébergé sur sur GraphDB, puis transforme les données sous format CSV — plus compact — avant de les retourner au client (voir la figure 3.7). Le serveur BAnQ garde en cache les données CSV pour accélérer le temps de réponse.

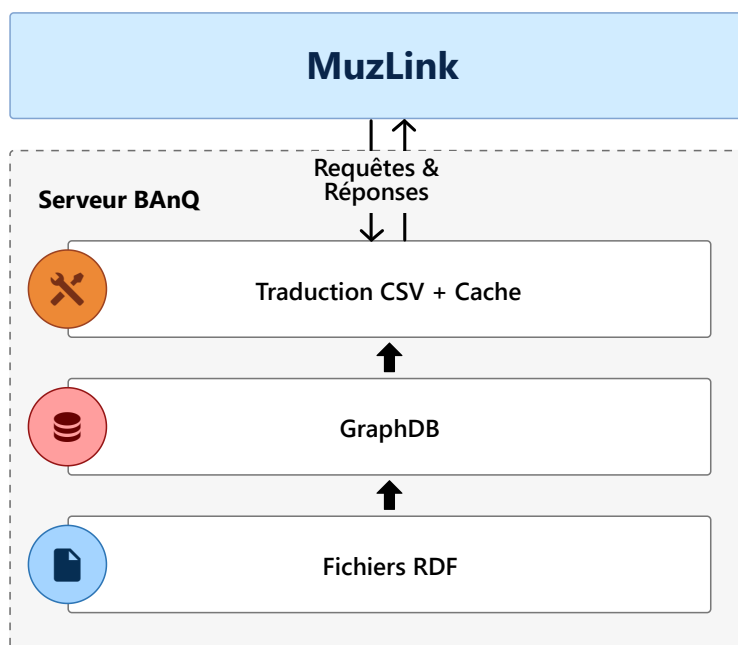


Figure 3.7 Pipeline de transformation des données, entre le format RDF jusqu’au déploiement du terminal consommé par MuzLink

3.5 Évaluation

Cette section décrit en détail l’étude utilisateur menée pour évaluer l’outil MuzLink. Elle passera en revue les différents aspects du processus qui ont mené à son approbation par le Comité d’éthique de la recherche (CÉR) de Polytechnique Montréal.

3.5.1 Recrutement des participants

Paramètres d’inclusion et d’exclusion

Les participants inclus dans l’expérience doivent avoir 18 ans et plus et être capables d’utiliser les fonctions de base d’un ordinateur (clavier et souris). Sont exclus les participants ayant des déficiences visuelles affectant la perception des couleurs. Comme la couleur joue un rôle

important dans l’outil de visualisation, une personne n’ayant pas les capacités pourrait fausser les résultats. Nous visons une répartition homme et femme équitable.

Recrutement

Notre partenaire industriel BAnQ a effectué un recrutement en ligne à l’aide d’une publication sur leur page Facebook. Le texte de cette publication est en annexe C. Cette annonce décrit brièvement l’objectif de recherche. Un lien au bas de l’annonce mène vers le formulaire d’inscription sur la plateforme de consultation publique du gouvernement du Québec (<https://consultation.quebec.ca/>). Ce formulaire demande à l’usager uniquement son nom complet ainsi que son courriel. Les coordonnées d’une personne-ressource sont ajoutées à l’annonce si le potentiel participant a des questions.

Les candidats ayant été retenus reçoivent ensuite un courriel qui contient des informations supplémentaires sur le déroulement de la séance d’évaluation. Ce courriel est disponible en annexe D. Lorsque le participant clique sur le bouton «Je veux participer», il est redirigé vers la page de recrutement de l’expérience. Tous les détails concernant l’expérience sont affichés : heure et lieu de l’expérience, coordonnées de la personne à contacter en cas de questions et informations sur l’objectif de recherche. Le candidat doit accepter le formulaire de consentement avant de s’inscrire. Le formulaire de consentement est en annexe F. La confidentialité des participants est garantie par la plateforme de consultation en accord avec leur politique de confidentialité (voir <https://consultation.quebec.ca/pages/politique>).

Compensation financière

Trois candidats tirés au sort parmi les candidats ayant participé à l’expérience ont reçu un chèque cadeau d’une valeur de 50\$ chez Renaud-Bray. Cela a motivé les participants à se déplacer dans la salle où a eu lieu l’étude.

Confidentialité des participants

Données de la phase de recrutement Lors de la phase de recrutement, nous avons recueilli les noms et courriels des participants. Seuls François Lévesque et Thomas Hurtut ont accès à ces données. Elles ont uniquement les deux utilités suivantes :

1. Envoyer un courriel d’invitation à l’expérience avec le lieu et la date aux candidats intéressés.
2. Prendre les présences le jour de l’expérience afin de tirer au sort un candidat qui a réellement participé à l’expérience. Les noms, prénoms et courriels sont définitivement

supprimés une fois ces deux actions complétées. Elles sont supprimées sur la plateforme de recrutement. Les feuilles de présence sont déchiquetées.

Données de la phase d’expérience Lors de la phase de l’expérience, les données recueillies pour l’étude sont complètement anonymes. Elles ne sont en aucune façon liées avec les noms et courriels collectés lors de la phase de recrutement. Il est donc impossible de faire un lien entre les résultats obtenus et une personne particulière.

Au début du questionnaire, pour évaluer la diversité de l’échantillon, chaque participant devra uniquement donner son genre, son groupe d’âge, ainsi que son niveau d’éducation. Ces informations sont insuffisantes pour reconnaître une personne précise. À noter qu’il est impossible de prévoir et de contrôler la diversité de notre échantillon. Les questions sont posées pour mieux évaluer les résultats obtenus.

Protection de l’anonymat Les données présentées dans les publications ne font jamais référence à des informations qui pourraient permettre de reconnaître un participant. Par ailleurs, par design, les résultats obtenus ne sont en aucune façon associés à une personne précise. Des identifiants uniques sont accordés pour différencier les différentes personnes.

Confidentialité des données

Données de la phase de recrutement Les données recueillies lors de la phase de recrutement contiennent des informations sensibles (nom et courriel). Ces dernières sont conservées sur le serveur de la plateforme de consultation publique qui respecte les plus hauts standards de sécurité (voir <https://consultation.quebec.ca/pages/politique>). Ces données auront uniquement les deux utilités suivantes :

1. Envoyer un courriel d’invitation à l’expérience avec le lieu et la date aux candidats intéressés.
2. Prendre les présences le jour de l’expérience afin de tirer au sort un candidat qui a réellement participé à l’expérience. Les noms, prénoms et courriels sont définitivement supprimés une fois ces deux actions complétées. Elles sont supprimées sur la plateforme de recrutement. Les feuilles de présence sont déchiquetées.

Données de la phase d’expérience Les données recueillies lors de l’expérience sont anonymes. Il sera impossible de faire un lien entre les résultats obtenus et une personne particulière. De plus, le questionnaire en ligne utilisera le protocole Hypertext Transfer Protocol Secure (HTTPS) pour crypter les communications et garantir l’anonymat.

Conservation des données Les données du questionnaire seront envoyées à un serveur externe avec le protocole HTTPS. Les données seront donc toutes cryptées. Les données seront conservées et cryptées sur un serveur Amazon Web Service (AWS) qui respecte de nombreux programmes de conformité reconnus mondialement, dont Cloud Security Alliance Controls (CSA), ISO 27001 (Security Management Controls) et ISO 27018 (Personal Data Protection). Seul le propriétaire (François Lévesque) peut décrypter et voir les données recueillies en s’authentifiant avec un mot de passe. Les données sont également conservées pendant 10 ans sur un serveur crypté du département de génie logiciel. L’intégrité et l’accessibilité de cette base de données seront garanties par des sauvegardes (« backups ») périodiques.

3.5.2 Déroulement de l’expérience

L’expérience a eu lieu le 8 février 2020 dans la Salle LaSerre à BAnQ. Les autorisations pour l’accès à la salle sont en annexe E.

Les participants doivent répondre à un questionnaire en ligne avec un ordinateur connecté à Internet. Ils ont tous la même version du questionnaire. Ils ne peuvent pas communiquer entre eux. Le responsable présent dans la salle où se déroule l’expérience ne doit pas avoir de comportements ou de propos qui pourraient influencer les réponses des participants. Il peut cependant assister les participants qui auront des problèmes techniques ou leur poser des questions qui n’auront aucun impact sur leurs réponses.

Le questionnaire est inspiré de travaux similaires du domaine, citons entre autres Loi et coll. [9] et Heer et Nielsen et coll. [10].

Les personnes sélectionnées doivent remplir un questionnaire disponible en ligne. Ce dernier est divisé en 6 étapes : (1) Test de vision, (2) Questions générales, (3) Tutoriel interactif, (4) Évaluation des tâches, (5) Exploration libre et (6) Questions ouvertes. Ces étapes sont décrites en détail au prochain chapitre dans la section 4.7. Quelques indications supplémentaires sont ajoutées dans les sous-sections suivantes.

Test de vision

Le test de vision chromatique d’Ishihara est utilisé pour détecter et exclure les participants qui ont des anomalies dans la perception des couleurs. Une seule réponse échouée est suffisante pour exclure un participant de l’expérience.

Parmi les 38 planches proposées par Ishihara, seulement 10 sont utilisées pour détecter les problèmes de distinction entre le bleu et le jaune (tritanomalie et tritanopie) [11]. Ce choix a été fait par contrainte de temps, car l’expérience ne dure que 30 minutes. De plus, 10 planches

peuvent être suffisantes pour détecter la grande majorité des anomalies [12].

Questions générales

Afin d'évaluer la diversité de notre échantillon de recherche, nous demandons aux participants de fournir quelques informations générales : genre, groupe d'âge et niveau d'éducation. Ces informations permettent de mieux évaluer et comprendre les résultats. Aucune information personnelle permettant d'identifier un participant n'est enregistrée afin de garantir l'anonymat de chacun.

Tutoriel interactif

Le participant doit ensuite suivre un tutoriel interactif pour comprendre comment utiliser l'outil. Une boîte de texte « pop-up » avec un court texte explique certaines fonctionnalités précises de la visualisation en pointant des centres d'intérêt (boutons, graphiques ou autre). Le participant doit cliquer sur le bouton « Suivant » pour voir la prochaine boîte de texte. Cette étape devrait durer 5 minutes au maximum. L'utilisateur peut ensuite explorer un artiste librement pendant 8 minutes pour se familiariser avec l'interface.

Évaluation des tâches

Suite au tutoriel, les participants doivent répondre à différentes questions qui nécessitent l'utilisation de l'outil de visualisation. Tous les participants répondent aux mêmes 11 questions. Ils ne peuvent pas demander l'aide d'autres participants ou du responsable de l'expérience afin de ne pas fausser les résultats. Les 11 questions sont dérivées des questions du tableau 3.3. Elles permettent donc d'évaluer directement la capacité de l'outil à atteindre les objectifs de recherche. Les questions sont listées à l'annexe G.

Questions ouvertes

Suite à l'exploration libre, le participant doit répondre à 4 questions ouvertes. Cette dernière partie du questionnaire servira à recueillir des opinions qualitatives et inédites sur l'outil. Elle donnera l'occasion au participant de communiquer ses difficultés rencontrées et/ou certains avantages inattendus lors de l'utilisation de l'outil. Les questions posées sont les suivantes :

- Avez-vous appris quelque chose pendant l'exploration ?
- Sur une échelle de 1 à 10, êtes-vous satisfait de l'information véhiculée par la visualisation ? Pour quelle(s) raison(s) avez-vous donné cette note ?

- Sur une échelle de 1 à 10, êtes-vous diverti par la visualisation ? Pour quelle(s) raison(s) avez-vous donné cette note ?
- Sur une échelle de 1 à 10, la visualisation est-elle facilement compréhensible ? Pour quelle(s) raison(s) avez-vous donné cette note ?

3.6 Pertinence de l'article

Au prochain chapitre, l'article présente la démarche complète du projet de recherche. L'introduction fait une définition des données fournies par BAnQ. La section *Design Process and Task Abstraction* expose le processus d'abstraction de tâches en collaboration avec des experts de BAnQ. La section *Related Work* recense les travaux de recherche qui ont situé la recherche dans le contexte scientifique actuel et guidé les choix de design. La section *MuzLink* rapporte et motive tous les aspects du design du prototype final en relation avec les *tâches*. La section *User Study* présente la structure de l'étude utilisateur formelle qui a été menée avec des visiteurs de BAnQ. Elle présente ensuite les résultats objectifs et quantitatifs obtenus. Une interprétation des résultats relève les forces et faiblesses de l'outil de visualisation. Elle propose également des pistes d'amélioration pour travaux futurs.

CHAPITRE 4 ARTICLE 1 : MuzLink - CONNECTED BEESWARM TIMELINES FOR VISUAL ANALYSIS OF MUSICAL ADAPTATIONS AND ARTIST RELATIONSHIPS

François Lévesque and Thomas Hurtut

Submitted to IEEE Transactions on Visualization and Graphics

4.1 Abstract

The rise of open data in the cultural domain is democratizing access to complex datasets usually presented as large multivariate and multilayered graphs. However, the exploration of such datasets is challenging for laypersons. The objective of this work is to develop and evaluate a new method for exploring and understanding such structured data. More precisely, this paper proposes MuzLink, an interactive visualization tool that allows the user to navigate, search, locate, and compare collaborative & influential relationships between musical artists through the exploration of musical adaptations. The proposed tool is based on a set of connected timelines visualizing how an artist's collaborations, inspirations, and influences evolved over time. This design study is conducted in close collaboration with BAnQ, the national library and archives agency of the Quebec government. A controlled user study, done with a group of BAnQ users, shows how the proposed approach is capable of performing a considerable set of analytical and exploratory tasks.

4.2 Introduction

With the rise of Linked Open Data (LOD) in recent years, the quantity and quality of Open Data portals have greatly increased [13]. These datasets take the form of graphs, which are easily understandable for a machine, but too complex to read and navigate for a human without a visualization tool. Although graph visualization is a well-researched topic, there are still limits to the number of attributes and nodes that can be displayed on a screen [1]. As part of an open data innovation project, *Bibliothèque et Archives Nationales du Québec (BAnQ)* — the national library and archives agency of Quebec government — has shared a previously unpublished music database, listing adaptation trees and collaboration graphs between artists, thus providing a more complex structure to the existing available databases such as *MusicBrainz* [14] and *Discogs* [15]. The underlying structure of this dataset is not limited to the field of music, but also applies for instance to cinema, scientific literature, and open source code: essentially every network of timestamped items produced by various

Jacques Brel

This visualization highlights artists related to Jacques Brel. Each bubble represents a song. They are organized into three major sections: inspirations, productions et influences.

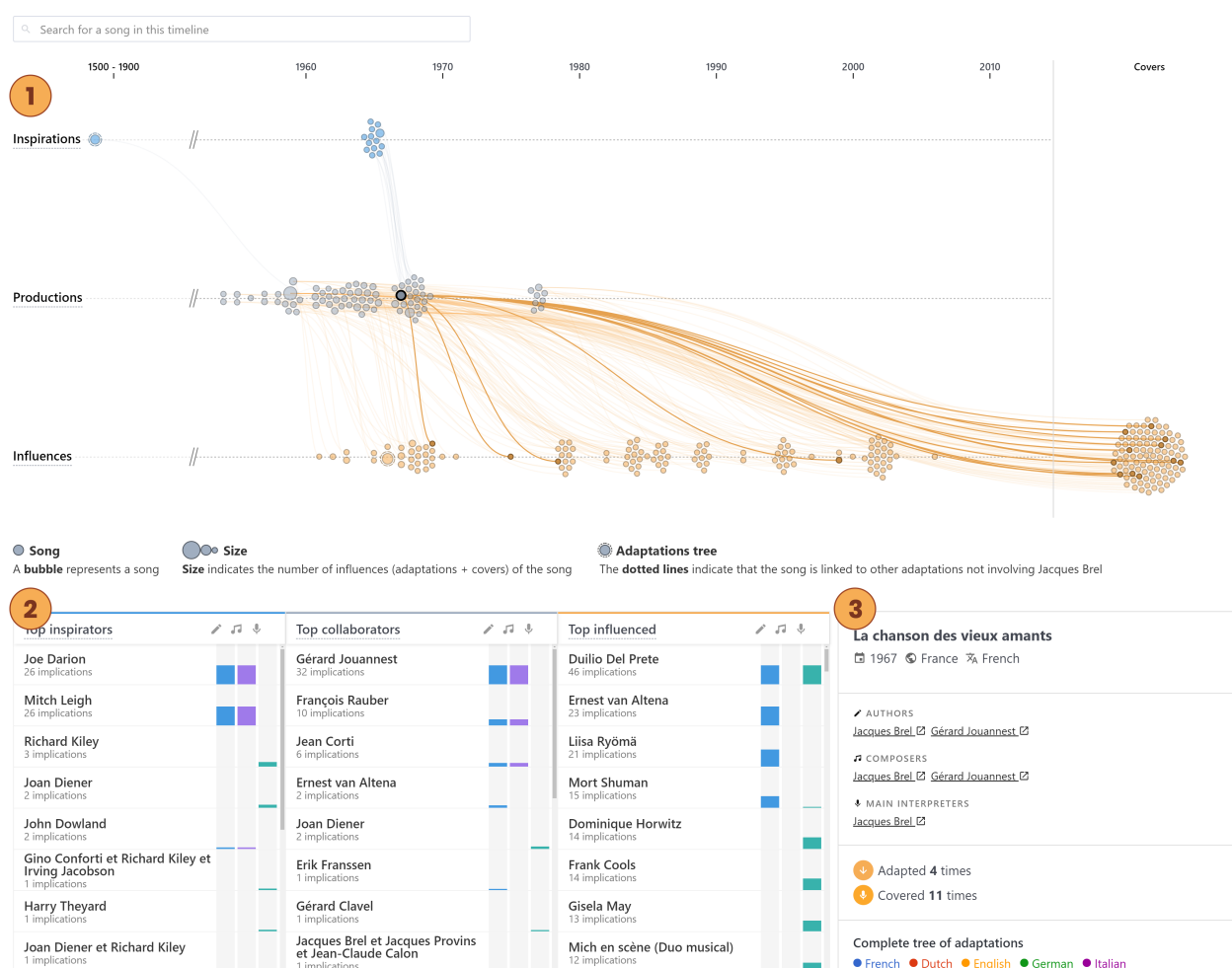


Figure 4.1 MuzLink is an artist-centered interactive visualization tool that allows users to explore collaboration and inspiration relationships between artists. Here, the user is exploring French singer Jacques Brel. The tool is divided into 4 juxtaposed and interactive views. First, the Connected Timelines (1): the middle one gathers Jacques Brel's songs, the top one gathers all songs that inspired him and the bottom one gathers all songs that were inspired by him. In a similar fashion, the Artist Lists (2) show artists who inspired, collaborated with, or were influenced by Jacques Brel. Clicking on a song — in this case "La chanson des vieux amants" — reveals its Detail Sheet (3). The Artist Overview (4) summarizes Jacques Brel's production. Edge cases are highlighted at the top right corner (5).

entities.

This design study presents a novel interactive tool to browse and analyze an artist’s collaborations and inspirations through their songs over time. The target user is the average library visitor who wishes to explore music data, discover new songs and artists, and thus might then borrow a musical album from the library collection. The proposed system targets several questions (detailed in Section 4.3), that this typical user could seek to answer. These questions were defined through a design process involving both BAnQ experts and visitors.

The BAnQ dataset Over the last decade, librarians from BAnQ built a unique dataset of around 24 000 songs and 19 000 artists. The available attributes for both songs and artists are detailed in Table 4.1. Among the available song attributes, a few stand out :

- **Author and composer IDs:** the authors and composers involved in the song production.
- **Main interpreter IDs:** the interpreters who are considered the official interpreters. Usually, the first singers who first performed the given song.
- **Cover interpreter IDs:** the interpreters who sang covers of the song. An artist can’t be both a main and a cover interpreter.
- **Parent song ID:** if the song is an *adaptation* (see Table 4.2 for our glossary), the parent song from which it was adapted. All songs in the dataset either have a parent song and/or have adaptations.

Following Munzner data abstraction [3], the BAnQ dataset is an unusual combination of two different data structures:

1. A **bipartite graph** with artist nodes on one side, and song nodes on the other. The edges symbolize the artist’s involvement in a song’s production. This involvement is characterized by a *role* (author, composer, or interpreter). Those edges are represented by the grey lines in Fig. 4.2.
2. **Trees** representing adaptation relationships between all songs. Each edge symbolizes an adaptation relationship between a *source song* and an *adaptation*. Those edges are represented by red lines in Fig. 4.2.

Furthermore, by merging the two data structures, this dataset can also be seen as a multi-variate network, where each node is an artist and each edge is a relationship (Fig. 4.3). This relationship is characterized by a role (author, composer, or interpreter) and by a relationship type (collaboration or inspiration). Unlike the collaboration, the inspiration is directed: an artist can be inspired and/or can inspire.

Table 4.1 Song and artist attributes.

Attribute		Type	Description
Song Attributes	ID	<i>string</i>	Unique song ID
	Title	<i>string</i>	Song’s title
	Language	<i>string</i>	Lyrics language
	Origin country	<i>string</i>	Country in which the song was produced
	Author IDs	<i>array</i>	A list of all authors involved in the song production
	Composer IDs	<i>array</i>	A list of all composers involved in the song production
	Main interpreter IDs	<i>array</i>	A list of all original interpreters involved in the song production
	Cover interpreter IDs	<i>array</i>	A list of all interpreters who made a cover on the song
	Parent song ID	<i>string</i>	If the song is an adaptation, the original song ID from which it was adapted
Artist Attr.	ID	<i>string</i>	Unique artist ID
	Name	<i>string</i>	Artist name
	Birth year	<i>integer</i>	Artist birth year
	Birth country	<i>string</i>	Artist birth country

Because of its complex structure, the dataset is difficult to explore and visualize. First, because its hybrid tree and graph structure makes it difficult to apply existing visualization models. Second, even with a network representation, the structure has unique characteristics — heterogeneous links according to the type of relationship and heterogeneous attributes on the ends of the links — that are just as challenging to represent with existing techniques.

Contributions We propose MuzLink, an artist-centered design based on connected timelines to enable a quicker understanding of collaboration and inspiration relationships with other artists. The design was made to facilitate the identification, the search, and the comparison of those relationships over time under an overall exploratory context. The contributions are the following:

1. We propose a visualization model called MuzLink to visualize complex hybrid graph and tree structures in an exploratory context. Specifically, the model aims to help identify and understand complex relationships — that evolve over time — between artists and songs in the musical domain.
2. We implement a fully functional and interactive web-based MuzLink prototype. The system is publicly available online. Moreover, it stays in sync with the original dataset

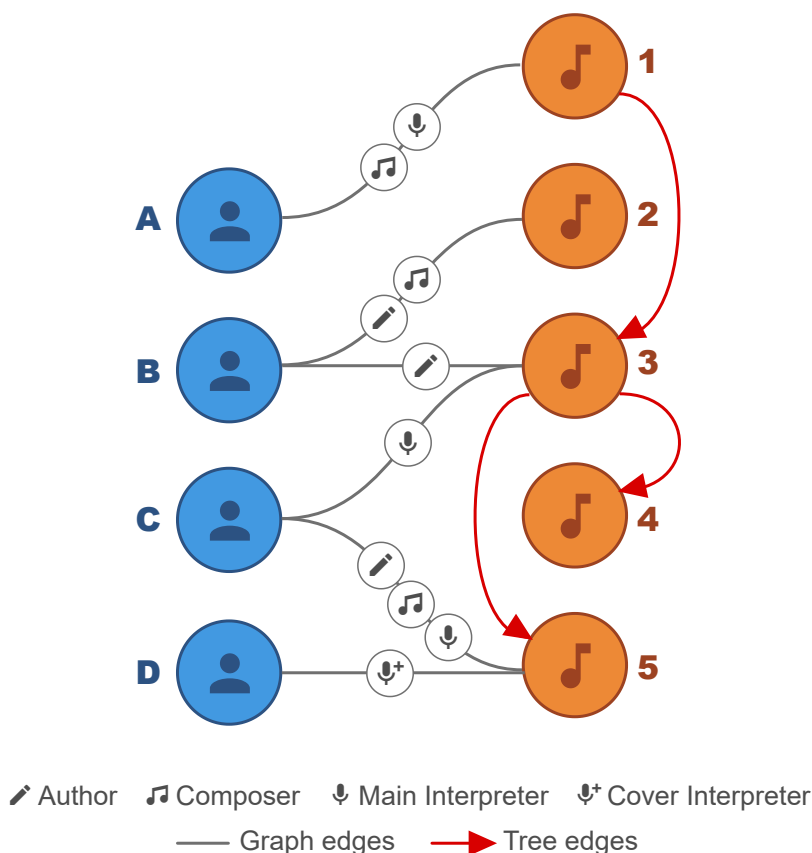


Figure 4.2 Bipartite graph and tree representation. The graph edges are grey and the tree edges are red. The artists are in blue on the left and the songs in orange on the right.

since it directly consumes BANQ’s public API.

3. We evaluate how MuzLink helps answer the identified questions (Table 4.3) by conducting a controlled user study involving 18 BANQ users. To evaluate the strengths and weaknesses of the tool, we measure the task response time, gather quantitative and qualitative feedback, and analyze the interaction logs.

4.3 Design Process and Task Abstraction

This section first describes the questions the target users of MuzLink seek to answer. They were identified during the design process conducted in collaboration with BANQ. From this set of use cases, we then derive user tasks, up to their abstract forms.

Table 4.2 Music glossary: To better communicate the dataset particularities throughout this article, we use the musical glossary below. We will often refer to these terms throughout this paper.

Word	Definition
Adaptation	A song adapted from another song. Generally through a lyric rewrite/translation or a novel musical arrangement.
Source song	The song at the origin of an adaptation.
Original song	A song which isn't an adaptation of another song. A <i>source song</i> isn't always an <i>original song</i> since more than two levels of adaptation are possible.
Cover song	A cover of a song, without enough significant modifications to be considered an adaptation.
Collaborator	An artist involved in the song's production. A collaborator may be an author, composer, and/or interpreter.
Cover interpreter	An interpreter who sang a cover a song without being involved in its original production. This set is mutually exclusive with the <i>Collaborator</i> set.
Artist's productions	All songs in which the artist was involved, as an author, composer, and/or interpreter. This set includes the covers sang by the artist.
Artist's inspirations	All <i>source songs</i> from the <i>artist's productions</i> .
Artist's influences	All <i>adaptations</i> or <i>covers</i> from the <i>artist productions</i> .

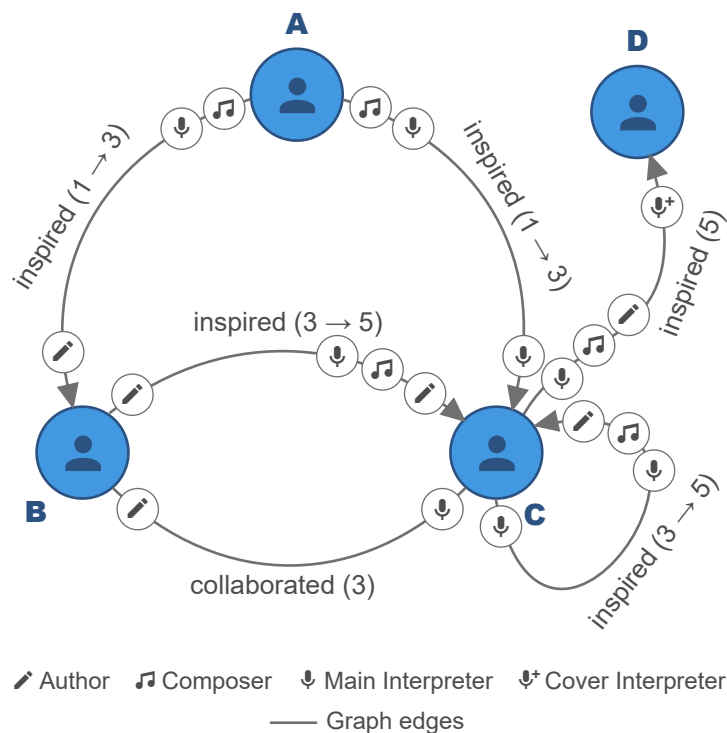


Figure 4.3 Multivariate network representation of Fig. 4.2. By merging the two data structures, it is possible to create a single network between artists. Each edge represents a relationship derived from song. Artists' roles are shown on both ends of the link.

4.3.1 User-Centered Questions

The overall target users are the BAnQ website and library visitors. Use cases were determined based on discussions, interviews, and brainstorming with BAnQ collaborators, and a literature review on tree and graph-based visualization systems [16,17], throughout the design process. The general exploratory objective from a user's perspective is to discover unsuspected relationships between artists and to analyze how these relationships evolved through time. We identified a set of 11 fundamental questions at different stages of the data analysis process. These questions are shown in Table 4.3.

Questions Q1 through Q3 seek to explore all relationships between artists — collaboration and inspiration relationships — from an artist's perspective. Question Q4 and Q5 explore the roles played in these different relationships — author, composer, and/or interpreter. Question Q6 seeks to locate the most inspired or influenced artists for a given artist. Q7 compares all artists to find the most influential or inspired. Question Q8 through Q11 focuses on single songs. While Q8 seeks to find the more influential songs, Q9, Q10, and Q11 seek to describe the influence of a song. Q9 describes the type of influence — adaptation or cover.

Table 4.3 Questions for the analysis of song adaptations data. For each, we detail to which layer of the data structure (T: tree, G: bipartite graph) and which part of the bipartite graph (A: artist, S: song), it pertains.

Question		T	G	A	S
1	Who <i>collaborates</i> with an artist ?	✓	✓	✓	
2	Who <i>inspires</i> an artist ?	✓	✓	✓	
3	Who is <i>influenced</i> by an artist ?	✓	✓	✓	
4	Which roles have an artist played over the course of his or her career		✓		✓
5	Which roles do the collaborators/inspirators/influencers play in relation to an artist ?	✓	✓	✓	✓
6	Who has most inspired or influenced an artist ?	✓	✓	✓	✓
7	Who is the most inspired or influenced artist ?	✓	✓	✓	✓
8	How influential is a song in comparison with others?	✓			✓
9	Was a song mostly adapted or covered?	✓			✓
10	Who adapted or covered a song ?	✓			✓
11	How was a song adapted or covered over time?	✓			✓

Q10 highlights the artist who produced said adaptations or covers. Q11 characterizes the influence of the song over time.

4.3.2 Task Abstraction

Task abstraction is a key step in the design process to detach the use cases from the domain-specific language [?, 3, 18]. In the following paragraphs, abstract tasks are written in italics. Starting with a known artist, the user should be able to *explore*, *discover*, and *enjoy* the sometimes surprising relationships that exist with other artists. Specifically, he or she must *browse* the distribution of his songs, inspirations, and influences over time — essentially following an artist’s career through their production and influences in a sequential manner. Using the adaptations tree’s *topology*, the user can precisely trace back the origin of a song or explore its influence over time. He should also be able to *locate* the *outliers* (the most influential songs, the oldest ones, etc.). He should also *identify* the roles played for each song. The user not only wants to *summarize* an artist’s relationships over time, but also wants to *compare* them with other artists, for example, by determining whether one artist is more influential than another. This comparison is made possible by the appearance of *features* and *trends* in the data.

4.4 Related Work

The dataset is a combination of a bipartite graph and trees (Fig. 4.2). The dataset can also be seen as a multivariate graph between artists (Fig. 4.3). Since the tasks identified are primarily concerned with exploring relationships of all types between artists, work related to multivariate graphs is more relevant than work related to bipartite graphs. As shown in Section 4.3, the temporal aspect of the data is crucial in answering some questions. As a result, our literature review focuses on multivariate graphs, trees, and timelines. We also reviewed academic publications visualizations, which have many features similar to our dataset: a bipartite graph composed of authors and publications, coupled with numerous reference trees.

4.4.1 Multivariate Graph Visualization

Multivariate graph visualizations can be divided into two main types of layouts: node-link layouts and adjacency matrices [1].

The node-link layouts allow an efficient topological exploration [1, 19]. To visualize and explore relationships between artists, Miller et al. [20] proposed Linked Jazz. It uses a force-directed node-link layout, where each node represents an artist, and each edge a relationship between two artists. The layout allows a good level of exploration, but links between artists aren't categorized by relationship type or role. This makes it difficult to understand the artist's sphere of influence and collaboration. To better communicate the nature of relationships, some visualization tools encode attributes directly on the edges or the nodes [10, 21, 22]. This is well suited when there are few attributes and the topology must stay readable. Otherwise, it is also possible to encode data on nodes via their position, either by positioning them in sets [23, 24], or by positioning nodes on the x/y axis [25, 26]. This is well suited when the attributes are more important than the topology.

Adjacency matrices are generally not well suited for large and sparse graphs like ours [27]. While they are powerful visualization tools to locate outliers and features in a dataset when attributes are ordinal, they are not suitable for exploring specific nodes [28–30].

While multivariate graphs are effective at representing artists and their relationships, they do not show individual songs and their adaptation trees, which is required to answer many questions from Table 4.3. In a nutshell, multivariate graph layouts on their own are not sufficient for our specific use case. Integrating a second visualization technique to make up for their weaknesses can greatly enhance the range of possible tasks. For example, Munzer built MizBee, a radial graph visualization to compare two genomes, combined with aggregated

chromosome and block views to highlight specific features [31]. Still in the genetic field, Seo and Shneiderman developed a tool combining node-link and matrix layouts to effectively bring to light genetic patterns [32]. As our dataset combines two structures, combining complementary visualization methods is an interesting way to increase the number of possible tasks.

4.4.2 Tree Visualization

Tree visualization layouts can be divided into two categories: explicit — via node-links — and implicit — via a treemap-like layout [7]. Node-links are generally more understandable for unfamiliar users [33], but quickly consume more screen space. Conversely, treemaps can represent a greater density of data, but are less suitable for topological tasks [34]. Yet, some techniques, such as Cushion Treemaps [35] and Squarified Treemap [36], allow for better visualization of the tree structure.

Our music dataset contains thousands of relatively small trees, each containing an average of 10 elements. As the trees are simple and the visualization is mainly addressed to laypersons, the node-link representation is the most relevant for a single tree.

For the representation of multiple trees, the literature proposes a few different techniques. Separate trees can be visualized via small multiples [37]. From this layout, associated nodes are visually linked via an edge [38–40]. While this technique allows direct relations to be traced between nodes, it does not scale well. To overcome this weakness, nodes can also be colored [41, 42]. This technique is better suited to give an overview of the relations between trees. Alternatively, node-links can also be fused to provide a natural model for hierarchical structures [43]. This method works best with a few structurally similar trees. To overcome the weaknesses of the techniques mentioned above, some works propose 3D visualization methods. By positioning the trees on parallel planes, and linking them either with edges or color, it is possible to visualize the trees in small multiples or by overlapping them with a 90-degree camera rotation [44, 45]. These techniques do not guarantee that the equivalent nodes will overlap, which is necessary to ensure readability [33].

4.4.3 Timeline Visualization

A timeline representation can be linear, radial, spiral, or irregularly shaped [8]. Among these, the linear representation is the most adapted to support the positioning of items chronologically following a sequence. Radial and spiral representations are more suited for periodic time visualization where a full rotation could represent a year or a month.

Attempts at representing graphs on timelines have already been made. EdgeMaps is a tool to explore the relationships of influence between entities [26]. It was used to represent influential relationships between philosophers. They are arranged on a timeline, and an interaction reveals directed and weighted edges from/to other entities. While this model shows the overall influences between philosophers, it doesn't show the publications through which this influence was calculated in the first place.

4.4.4 Academic Publication Visualizations

Publication datasets share many similarities with our dataset. They are composed of a bipartite graph with publications and authors instead of songs and artists. Publications are part of reference trees, just as songs are part of adaptation trees. This section presents a few design studies specifically related to the visualization of academic publications, and how they manage to facilitate exploration tasks.

Dörk et al. [46] proposed PivotPaths, an interface representing up to three facets of a complex dataset (e.g. authors, publications, keywords, categories) at the same time by agglomerating nodes from a graph into adjacent sections. A user study revealed it was a pertinent tool for exploration and focus tasks. However, it cannot display the codependent relationships between entities (e.g. influence relationships between authors).

Matejka et al. [47] developed Citeology, an interactive tool specifically focusing on the publications themselves and arranging them on a timeline. For each reference, an edge is drawn from the source to the target publication. An interaction on a specific publication highlights its associated edges. However, the visualization does not show the authors and their relationships.

Nobre et al. [48] proposed Juniper, a tool combining a tree and table view to efficiently visualize several node and edge attributes on very large and complex graphs. A user study with a publication dataset revealed that this model is well suited for focused tasks on relationships between authors and their publications. However, with its great flexibility also comes a greater learning curve, which may not be adapted to our target audience. Moreover, being a tool adapted to all types of multivariate graphs, it will hardly be able to efficiently highlight certain attributes typically associated with music.

4.5 MuzLink

Based on the target users, identified tasks, and a review of previous work, we conducted an iterative conceptualization process to design the proposed MuzLink tool. We present it in

this section, and motivate each aspect of its design.

4.5.1 Entry Point

One of the first design choices was to determine the entry point of the visualization. To be readable while showing a great level of detail, the large dataset needs to be divided into subsets. Since artists relationships are supported by songs and their musical adaptations, they constitute a natural subset. Other subsets, such as the origin country or the lyric's language, made less sense for the identified questions. Filtering the data around one specific artist provides many advantages: it results in small and readable chunks of data, which are fully suited to answer medium to low-level questions regarding a single artist's relationships and songs.

4.5.2 A Multi-Part Visualization

To efficiently answer all 11 questions from Table 4.3, the visualization was divided in 4 views: the Connected Beeswarm Timelines (Fig. 4.1.1), the Artist Lists (Fig. 4.1.2), the Detail Sheet (Fig. 4.1.3) and the Artist Overview (Fig. 4.1.4). The Connected Timelines mainly serves as a topological visualization of songs' relationships. The Artist Lists serve as an aggregation representation of related artists. The Detail Sheet shows detailed metadata regarding a selected song in the Connected Timelines. The Artist Overview gives a summary of the type of productions and the sphere of influence of the artist.

Connected Beeswarm Timelines

To allow deeper explorations, showing each item — songs and artists — on an atomic level is an important consideration. As shown by Anscombe's quartet [49], especially for large and complex datasets, a summary hides the true data structure, resulting in information loss. After filtering songs related to a given artist A, we are left with a number of small adaptation trees, where each node is a song and each edge is an adaptation relationship. Each song, represented by a small circle, is categorized according to the nature of its relationship with the artist A:

1. **An inspiration:** a song adapted or covered by artist A
2. **A production:** a song produced, or partly produced, by artist A
3. **An influence:** an adaptation or cover from one of artist A's productions

For each category above, we build an associated timeline (see Fig. 4.4). This attribute-driven grouping allows to quickly reflect the relationship nature between songs [1].

For each timeline, songs are horizontally positioned by their release year. This linear and chronological layout was chosen to better read items in a sequential order [50]. When large swaths of the timeline are empty, they are cut to maximize screen utilization [50]. In order to preserve visibility of the songs, overlapping is avoided using a collision algorithm, creating a beeswarm-like timeline [51].

In addition to the x/y position channels, each song encodes its influence level — the sum of its adaptations and covers — using the area visual channel. This channel thus adds visual weight to nodes part of a bigger sphere of influence, communicating relevant cues regarding the inspirations' evolution over time.

Since each node is part of a tree, all associated edges are added on top to effectively create a *Connected Timeline*. All edges are drawn with a low opacity and a small Bézier curve to reduce the visual clutter when the graph is dense. The node-link tree visualization is chosen for its expressive representation better suited for network topology and connectivity tasks [52]. With such a layout, it becomes difficult to distinguish individual trees, but the accumulation of edges provides rich visual cues regarding the trend and distribution of the relationships.

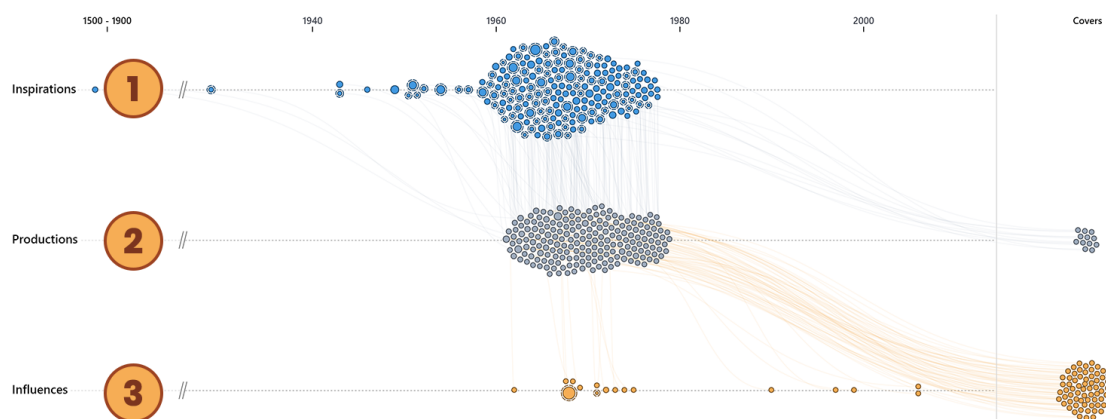


Figure 4.4 Connected Timelines: Inspirations (1), Productions (2) and Influences (3)

Sometimes a song may belong to multiple categories. For example, song Y is an adaptation of song X, and song Z is an adaptation of song Y. If an artist wrote song X and Z, then song Y is both an *inspiration* and an *influence* for the artist, and thus appears on two timelines. This duplication can be misleading and hard to interpret. Therefore, when such a case arises, the song is duplicated and a special warning appears on the top right (Fig. 4.1.5). These

notifications explain why and when special cases are present on the visualization.

Artist Lists

While the first view presents the relationships between the **songs**, the second view presents the **artists** related to the current artist. All the songs have been arranged in 3 different categories according to the nature of their relationship with the current artist. In a very similar fashion, the artists in the *Artist Lists* are arranged in 3 categories following the same semantics as shown in Fig. 4.5:

1. **Inspirators**: All artists involved in the creation of a given artist's *inspirations*.
2. **Collaborators**: All artists involved in the creation of a given artist's *productions*.
3. **Influenced**: All artists involved in the creation of a given artist's *influences*.

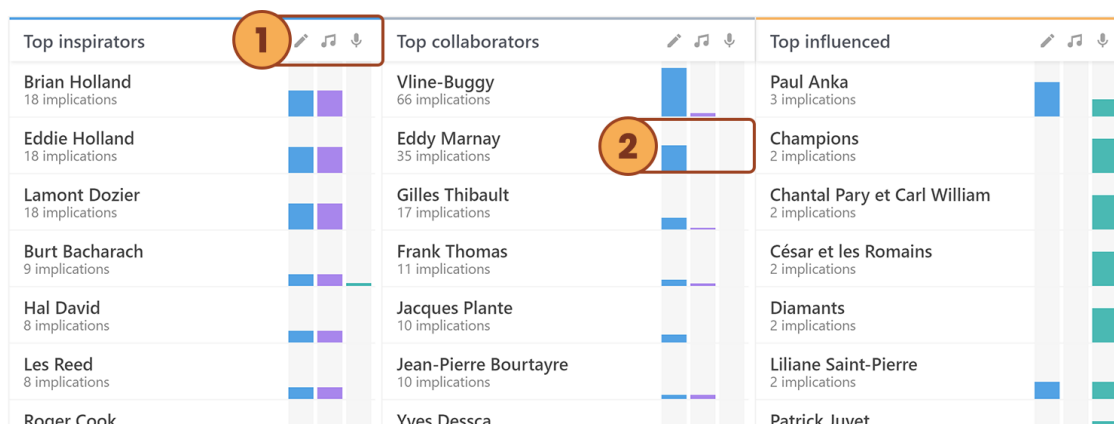


Figure 4.5 Artist Lists are divided into 3 columns where each column contains related artists ordered by the number of relations. The first column lists the top inspirators, the second lists the top collaborators, and the last lists the top influenced. The user can filter the list by roles (1). Each artist on the list has a small bar chart showing the role distribution (2).

Unlike the timelines, the lists are aligned horizontally. While a timeline naturally reads from left to right in the west, a list naturally reads from top to bottom. Each artist in the list is represented as a table row with the following data:

- His **full name**
- His **number of implications** in the category (inspirator, collaborator or influenced)
- The **distribution of the roles occupied** for all his implications (author, composer or interpreter). This distribution is visualized using small bar charts (Fig. 4.5.2). Those allow quick and precise comparison by using the length, a particularly effective visual channel [3].

An artist can belong to many lists at the same time. For each list, the artists are ranked in order of involvement, from the most involved to the least involved. This way, it is easier to scan which artists are the most significant for each category. With the small role filters at the top of each list (Fig. 4.5.1) it is possible to rank and filter by a specific role (author, composer, or interpreter).

Detail Sheet

Selecting a song on a timeline brings up its *Detail Sheet* on the bottom right corner (Fig. 4.6). This sheet displays all the song's metadata: name, release year, origin country, lyric language, authors, composers, interpreters, and source song. Additional stats regarding its sphere of influence are shown: the number of adaptations and the number of covers. At the bottom, the full adaptation tree in which the selected song is part of is displayed. This full tree offers additional insights compared to the timeline tree:

1. The timeline only displays a maximum of 3 levels per tree: one level above and one level below the artist's *productions*. The full tree allows the user to explore hidden levels for a specific tree. This is particularly useful to find the distant root *source* of the song. Conversely, it may be used to find the distant adaptations of a song.
2. Each node part of the tree uses the color channel to encode the language of the lyrics. This provides insights regarding the direction of inspiration flows between languages.
3. While timelines effectively represent inspiration flow patterns over time, their layout is less suited for the topological reading of a single tree. The full tree provides a hierarchical view that is easier to read and follow, especially when the number of nodes is large.

Artist Overview

At the top is a general overview of the artist's productions (Fig. 4.7). The first chart shows the sums of the different types of production: original songs, adaptations, and covers. The second chart shows the sums of the roles assumed for all his productions: author, composer, and interpreter. The bar chart visualization allows efficient comparison between different values [3]. Therefore, this representation makes it possible to accurately and quickly compare artists with each other.

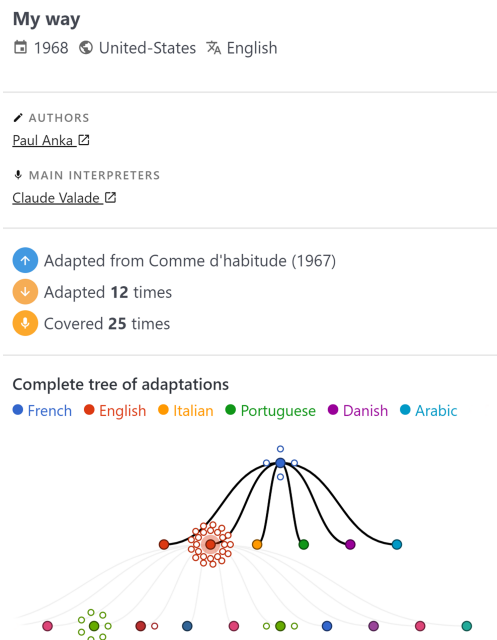


Figure 4.6 The Detail Sheet is a small card showing more details about the selected song: the release year, the country, the language, the artists and their roles, the source song, the number of adaptations, the number of and covers, and the full adaptation tree.

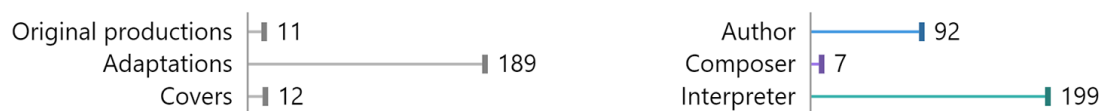


Figure 4.7 The Artist Overview shows the type of productions the artist made and the roles he bore over his career.

4.5.3 Interactions

Being a multi-part visualization, adding interactions to MuzLink was primordial to take advantage of all views and effectively connect them.

Interactivity across Views The Connected Timelines and the Artist Lists are the primary views of the tool. While they are spatially close and semantically related through the use of the same tripartite structure (inspirations, collaborations, and influences), interactions are paramount to conceptually link them and facilitate the transition from one view to another.

When the user hovers an artist from an Artist List, all songs in which the artist was involved are highlighted on the corresponding Connected Timeline. Likewise, hovering a specific chart

band from an Artist List only highlights the songs in which the artist was involved for the hovered role. This interaction reinforces the relationship between the two views by making it possible to locate an artist’s production from the Artist Lists to the Connected Timelines.

Song Context and Metadata As songs are the basic unit of visualization, it is essential to highlight their characteristics and metadata. Hovering a song on a timeline highlights the nodes and edges from the tree it is part of, revealing its context in relation to the artist and through time. Hovering also reveals a tooltip, quickly and explicitly revealing the song’s metadata: its title, release year, production type (original song, adaptation or cover), and original artists listed by roles.

A click on a song locks the tree highlight and opens the Detail Sheet (section 4.5.2). Hovering a song on the full tree highlights the same song on the Connected Timelines view, and hovering a song on the Connected Timelines view also highlights the same song on the full tree. This bidirectional interactivity helps conceptually and semantically connecting the two views while facilitating the transition between one view and another.

Search Bar The large quantities of items on the timelines make it difficult to find a specific song. A search bar above the view allows to quickly find a song by its title. A song selected from the search bar is highlighted on the Connected Timelines view and its associated Detail Sheet appears at the bottom, quickly providing its full context and metadata.

Navigation and Discoverability The visualization was designed to facilitate the discovery of relationships between artists. Therefore, it allows a fluid and natural navigation between the different views by artist. A user may navigate to another artist’s view by clicking on an Artist Lists’ item or an artist from the song’s Detail Sheet. Alternatively, the user may search a specific artist on the artist search bar at the top.

4.5.4 Onboarding

Since the dataset is complex and the interface rather original for a layperson, we designed an onboarding strategy which supports the user in learning how to use and read a new interactive visualization [53]. MuzLink’s onboarding consists of an internal — in the visualization tool itself — guided tour, integrated for the first use. The tour goes through all parts of the tool in a logical sequence, basically following a storytelling approach. This technique is an effective method for imparting knowledge [54]. Visitors can revisit the guided tour as many times as they wish.

4.6 Implementation

We propose a web-based version of MuzLink using HTML, CSS and JavaScript. It leverages the D3.js library to build all SVG visualizations and the vue.js framework for interactions. A demo is publicly available here: <https://muzlink.witify.io>. Although the tool is fully functional, it is important to note that it has been optimized for 1080p screens and the Chrome browser. The tool communicates directly with BANQ’s server. It is therefore always in sync with the latest data.

4.7 User Study

To evaluate the effectiveness of MuzLink, we conducted a formal user study to validate the initial objectives. We designed an experiment that introduces the participant to MuzLink and assigns tasks to be solved with the tool.

4.7.1 Experiment overview

The experiment, conducted in a physical room, is a fully automated quiz. Participants present themselves in the room and answer the quiz on a computer on their own. To minimize the influence on participant responses, members of the research team can only answer questions related to technical issues or questions regarding the organization and structure of the experiment.

We recruited 18 participants - 9 men and 9 women - between the ages of 24 and 74 with no particular skills or education. Of these, 10 went to university. The others have college degrees. Participants never used MuzLink before the experiment.

The objective of the experiment is to evaluate MuzLink’s performance in answering all questions from Table 4.3. It is divided in four main phases:

1. **Tutorial phase:** The participant completes the onboarding guide described in Section 4.5.4. The guided tour takes an average of 7.5 minutes. Afterwards, the participant is asked to get used to the tool by freely exploring an artist’s view for a few minutes.
2. **Task phase:** The objective is to evaluate the effectiveness of the tool in answering all questions from Table 4.3. For each question, participants complete an associated assessable task (Table 4.4). They must analyze the visualization and answer to the best of their ability from a list of options. Their reaction time is measured. They can skip a question at any time. This phase is expected to take 10 minutes on average.

3. **Exploration phase:** After answering the questions. The participant can explore the data again. This time, however, he can browse freely from one artist to another and use the artist search function. Real-time logs are generated throughout this phase to measure the participant’s level of engagement and exploration. A timer limits the duration of this phase to 10 minutes.
4. **Feedback phase:** After freely exploring the data, the participant completes an open-ended survey. First, he is asked if he has discovered any insights during the Exploration phase. Then, he is asked to rate and give qualitative feedback about his understanding, satisfaction and enjoyment of the MuzLink tool overall. At the end, the participant can optionally leave general comments.

4.7.2 Results

Task Completion Quality

The quiz is divided into 11 concrete and assessable tasks derived from the 11 questions in Table 4.3. To reflect actual usage as closely as possible, the quiz uses BAnQ’s dataset. To ensure a direct evaluation of the tool, rather than the knowledge of a particular artist, each task concerns a different artist. We carefully chose them to allow an unambiguous and assessable answer from the participant.

Figure 4.8 shows the verdicts of the participants’ answers. Figure 4.9 shows the average time required to complete each task. All tasks were successfully completed by the majority with an average response time of about 10.1 seconds.

Tasks 1, 8 and 10 took the longest time to complete. We believe it’s natural that the first task takes the longest since the participant must first become familiar with the tool. Task 8 requires comparing areas, which is less obvious as evidenced by Munzer’s work [3]. It also explains why this task is one of the most incorrectly answered. Furthermore, out of the 18 participants, 4 selected one of the adaptations of the most influential song, rather than the most influential song itself. Therefore, we believe part of the lower score is due to a bad understanding of the glossary we introduced through our tool. An interactive onboarding process may better communicate this vocabulary than the current passive guided tour. Moreover, visual cues to help differentiate *inspirations* and *influences*, such as icons and color, could be added to the labels and tooltips. Task 10 took longer since it required several steps: search a song with the search bar, select it, find its adaptations and their respective artists. Despite its greater completion time, it remains very successful.

The Task 7 had the highest failure rate, which is surprising since it’s the exact reverse of

Table 4.4 Tasks used during the controlled user study. Each task is a concrete and assessable version of a question in Table 4.3.

Question	Task
1	Who most collaborated with {artist}?
2	Who most inspired {artist}?
3	Who was most influenced by {artist}?
4	Which role {artist} carried the most?
5	{artist A} is often influenced by {artist B}, which role he or she carried the most through this relation?
6	Which artist is the most adapted and covered?
7	Which artist is the most influenced?
8	What is the most influential song?
9	How often was a song adapted and/or covered?
10	Who adapted {song}?
11	Over which period {artist} produced his or her songs?

Task 6. Task 6 asks for the most *adapted and covered* artist, while Task 7 asks for the most *influenced* artist. An analysis of the responses revealed that 5 participants chose the most adapted and covered artist on both tasks. Since the format of the question was exactly the same, we believe that many participants did not notice that the question had changed between Task 6 and Task 7. This problem could have been avoided by highlighting the difference between the questions with two different colors. Rather than asking them in sequence, it would also have been appropriate to distance them in the questionnaire to avoid confusion.

Level of Engagement

During the Exploration phase, the participant was free to explore the dataset using our tool. Real-time logs generated throughout this phase include artist and song click, artist and song searches, artist and song mouseovers, and role filter clicks. These logs were generated from the quiz web interface client and sent in real time to a database. On average, participants visited 2.8 artists, including 0.8 via the search bar. Since the exploration period took 8 minutes, the participants thus visited each artist for a little less than 3 minutes on average. Participants took the time to visit each artist in detail. They hovered 37.8 artists on the lists, 81.3 songs on the timelines, clicked on 6.8 songs, and used the role filters about 7 times. We believe these logs indicate that the tool induces a good level of exploration and interactivity.

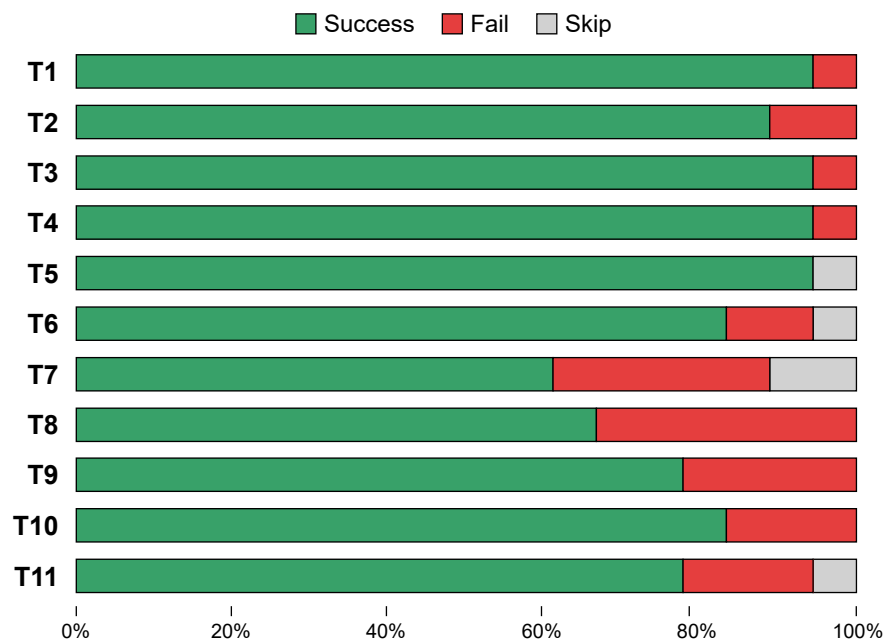


Figure 4.8 Score for each task during the experiment. Successful tasks are in green, failed tasks in red and skipped tasks in grey.

Quantitative and Qualitative Feedback

Following the Exploration phase, participants were asked if they discovered any new information. Of the 18 participants, 16 claimed to have learned new information as a result of their exploration. More specifically, many mention having discovered new unsuspected relationships between artists. Others claim to have learned the origin of several songs.

We then asked participants to give quantitative and qualitative feedback on three different aspects: understandability, satisfaction, and entertainment. Understandability measures the ease at which the participant understood the tool. Satisfaction measures the perceived level of appreciation for the tool's ability to answer questions. Entertainment measures the level of enjoyment using the tool. Each participant gave a score on a 1-10 Likert scale, where 1 represents disagreement and 10 represents agreement. They also added a comment to justify their score. Fig. 4.10 shows score distributions for each aspect. Overall, the scores are favorable in every aspect.

For the "comprehension" aspect, the vast majority of participants say that the tool is easy to understand and easy to use. Six participants pointed out that the unusual layout requires a little time to adapt, but once familiar, it becomes easy to read. Some participants were concerned not to manage to untangle the links between the timelines, but they quickly realized that the interactions were enough to answer the questions.

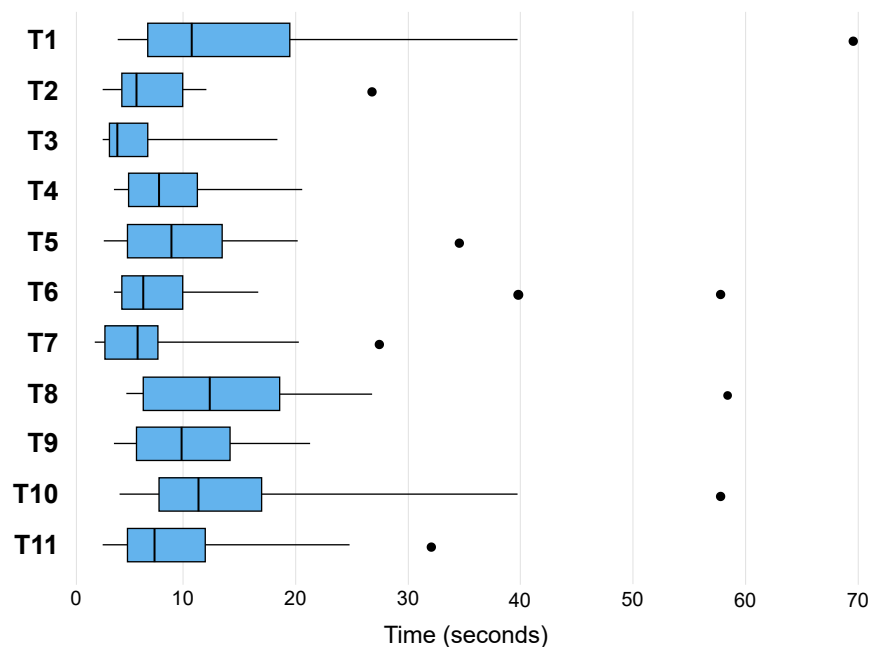


Figure 4.9 Distribution of the time it took for participants to complete each task during the experiment. Outliers are shown as dots.

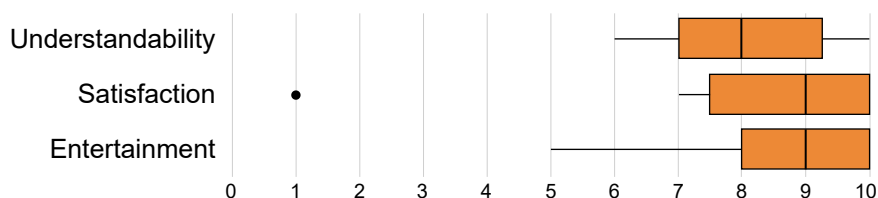


Figure 4.10 Quantitative feedback score distribution for each aspect. Outliers are shown as dots.

For the "satisfaction" aspect, the majority of the negative comments underline the incompleteness of the dataset rather than the visualization itself — which is a good sign in our case. The tool successfully propelled the discovery among the participants, who quickly understood the limitations of the dataset. Five participants underline that the tool is effective in discovering new relationships, the information is clear and quick to access.

For the "entertainment" aspect, eight participants mention being pleasantly surprised by their discoveries. The tool was effective in quickly identifying relationships between artists. Eight participants appreciated the colorful graphic aspect and the smooth transitions, which compelled them to explore. Three participants do not consider the tool enjoyable, mentioning that the visualization is very technical and not suitable for the entertainment of average people.

General comments at the end of the experiment are generally positive. Seven participants underline their enthusiasm for the tool and are excited to see the project evolve. Two participants add that it would be relevant to include filters and zoom capabilities to browse some timelines in detail. This is indeed a limitation that we will address in the discussion.

4.7.3 Discussion

The user study shows that MuzLink makes it possible to answer the questions in Table 4.3 reliably and quickly. The majority of participants showed great interest in the tool and stated that they discovered new information. Feedback and log analysis indicate that the tool is fitted for musical exploration, thus acting as a relevant entry point to borrow a musical album at BAnQ.

The tool requires a small learning curve, especially because of the newly introduced glossary. However, we believe the latter is essential to fully understand the complexity behind all artists' relationships. In future work, updating the onboarding process by making it interactive and fine-tuning the visual cues supporting our glossary could greatly lower this learning curve.

The limits of the tool are mainly associated with its fixed zoom level. For the exploration of large datasets such as ours, Schneiderman's mantra proposes an approach starting with a large overview, then leaving to the user the choice to zoom and filter and then analyze the details [55]. In future work, adding a contextual view, summarizing all data, would greatly benefit MuzLink's usage. Furthermore, adding more zoom and filter capabilities — timeline zoom, filter by music genre/language, etc. — would allow users to further explore the dataset.

4.8 Conclusion & Future Work

The inspirational, influential, and collaborative relationships between artists constitute a complex dataset, which can be described as a hybrid tree and graph structure, or as a complex multivariate graph. An extensive literature review shows existing techniques perform poorly at complete key tasks on such a dataset. In this paper, we proposed a novel approach based on connected beeswarm timelines to explore evolving relationships between artists over time. To support this objective, we conducted a controlled user study to evaluate the strengths and weaknesses of the tool. Results show that MuzLink is an enjoyable tool, effective for performing many discovery tasks in an exploratory context.

MuzLink is not a general tool and only focuses on the field of music. Yet, we believe it could

be adapted to other fields with similar datasets, such as film and scientific publications. Moreover, the connected timelines, due to their rich visual language, could be exploitable in other contexts where items from mutually exclusive sets are linked through time.

4.9 Acknowledgments

The authors deeply thank Jean-François Gauvin, Marielle St-Germain and Benoît Migneault from BAnQ for their fruitful discussions, feedback, and overall collaboration along the project described in this paper. They also thank Dominique Piché and Michel Gagnon from Polytechnique Montreal for the help on the data engineering aspects of the project, and Michael J. McGuffin from ÉTS for the helpful discussions at the beginning of the project.

CHAPITRE 5 DISCUSSION GÉNÉRALE

Ce chapitre fait un bref retour sur les résultats obtenus dans l'article. Il présente des détails supplémentaires et discute des limites de l'évaluation. Les résultats bruts du questionnaire sont également présentés en annexe H.

5.1 Échantillon

La base de l'échantillonnage est constituée des personnes qui ont aimé la page Facebook de BAnQ. La BAnQ a effectué une annonce Facebook pour recruter les participants intéressés en échange d'une carte cadeau chez Renaud-Bray. L'échantillon est donc biaisé, car il contient davantage d'individus ayant un intérêt pour la culture que la population générale. Cependant, ce biais n'a pas été jugé problématique pour notre expérience. Comme l'outil s'adresse principalement à des adeptes de musique, il est au contraire plus proche de notre utilisateur cible.

La distribution des 18 participants est très bien répartie. Exactement la moitié des participants sont des femmes, l'autre moitié sont des hommes (figure 5.1).

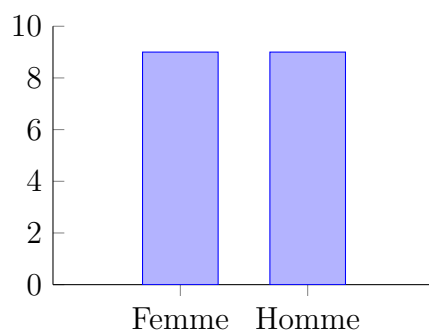


Figure 5.1 Genre des participants

7 participants ont un baccalauréat. 4 participants ont terminé des études supérieures. Les 4 autres participants ont soit un DEP ou un DEC (figure 5.2).

La distribution d'âge est presque équitable entre 25 et 64. Un seul participant a plus de 65 ans (figure 5.3).

Cette distribution diversifiée permet d'avoir un portrait plus réaliste et complet de l'efficacité de l'outil de visualisation avec la population générale.

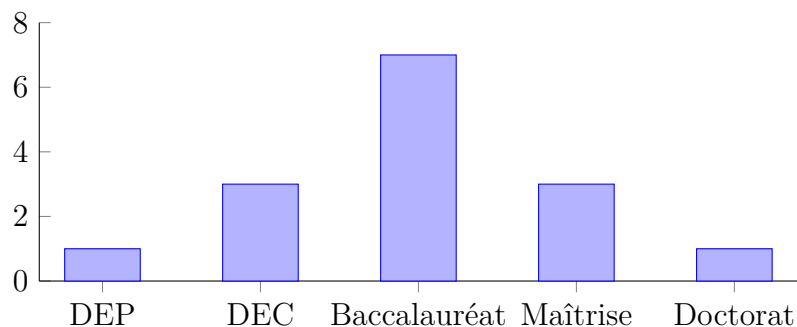


Figure 5.2 Niveau d'éducation des participants

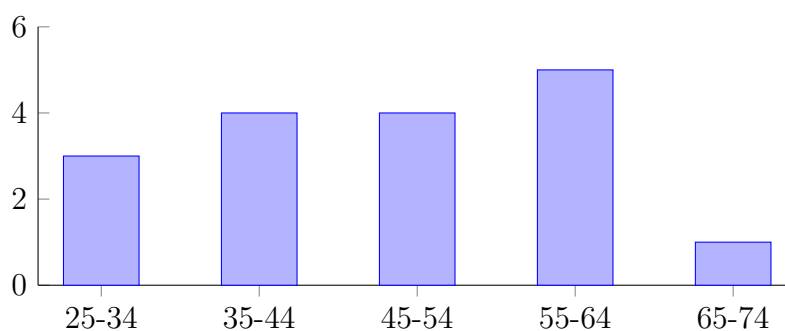


Figure 5.3 Groupes d'âge des participants

5.2 Engagement utilisateur

Lors de la période exploratoire, les interactions et actions des participants ont été enregistrées. Ces dernières ont déjà été abordées dans l'article au chapitre précédent. Cette section présente davantage de détails quant à la distribution des données. La figure 5.4 présente la répartition des interactions enregistrées par utilisateur. Les survols de souris constituent l'interaction la plus fréquente. Les participants ont en moyenne survolé près de 40 artistes et 60 chansons. Les filtres ont été utilisés près de 2 fois en moyenne. Comme en témoigne le graphique des navigations, les utilisateurs ont une tendance à visiter que deux ou trois artistes en profondeur. Ils ont plus tendance à cliquer sur les chansons sur les lignes du temps que de naviguer vers d'autres artistes. Comme l'exploration libre ne durait que 8 minutes, le participant passe ainsi en moyenne 2 à 3 minutes par artiste. Il aurait été intéressant d'augmenter la durée de la période exploratoire afin d'observer si le nombre d'artistes explorés aurait augmenté. Ces données nous permettent néanmoins de constater que les participants préfèrent explorer en profondeur un artiste plutôt que de naviguer d'un artiste à l'autre. Comme en témoignent les commentaires des participants présentés dans la section suivante, la page d'un artiste en elle-même permet un grand niveau d'exploration et de découverte.

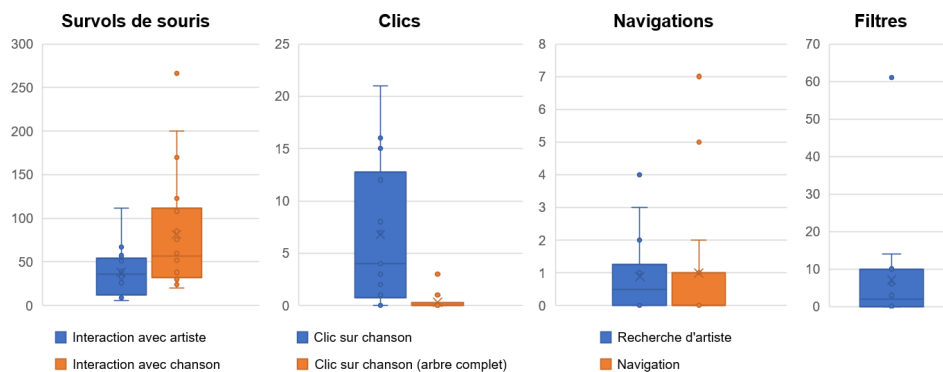


Figure 5.4 Distribution des interactions enregistrées pour chaque participant lors de la période exploratoire de 8 minutes.

5.3 Commentaires des participants

La totalité des réponses aux questions ouvertes est présentée en annexe I. Concernant l'aspect de découvrabilité, 95% des participants affirment avoir découvert de nouvelles informations en naviguant librement dans l'outil. Grâce à sa prise en main intuitive, et aux fonctionnalités de navigation fluides, la méthode permet un bon niveau de découvertes et d'explorations.

Concernant l'aspect de satisfaction — à savoir si les participants sont satisfaits de l'outil pour propulser leurs recherches — les commentaires sont encore une fois très positifs. Les participants ont donné une note moyenne de 8,3/10. Près de 55% mentionnent spécifiquement être très satisfaits par l'outil. Cependant, 11% affirment qu'ils apprennent plus facilement en lisant, l'outil n'est pas adapté à leur mode d'apprentissage. Un autre 11% affirme qu'il manque certains artistes. Cela constitue une faiblesse du jeu de données et non de la visualisation en tant que telle.

Concernant l'aspect de divertissement, les résultats sont encore une fois très probants. La note moyenne est de 8,7/10. 44% des participants mentionnent que l'outil est *intéressant*. 33% des participants font référence à la visualisation elle-même en disant qu'elle est *efficace* ou *satisfaisante*.

Les participants ont donné en moyenne une note de 8,1/10 pour l'aspect compréhension. Sur les 11 tâches du questionnaire, 78% des participants ont fait 3 erreurs ou moins. La visualisation est donc généralement bien comprise, mais nous croyons qu'elle peut être améliorée. L'incompréhension de l'outil provient en partie du glossaire utilisé : 28% des participants affirment avoir eu du mal à comprendre la signification des codes à leur première utilisation. L'incompréhension de l'outil provient également d'un manque de contrôle sur la quantité des

données. 17% mentionnent une trop grande quantité d'informations et déplorent l'absence de filtres. Les propositions mentionnées dans l'article — l'ajout d'une vue globale et de filtres, la création d'un tutoriel interactif et une meilleure communication du glossaire musical — pourront grandement aider à améliorer ce score.

5.4 Vue globale et contexte

Quelques participants ont mentionné l'absence de zoom et de vue globale. Dans le cadre du projet d'innovation avec BAnQ, une vue globale a été conçue (figure 5.5). Cette dernière a été évaluée avec des experts de l'institut [2], mais ne servait pas de point d'entrée pour l'exploration avec MuzLink. Lors de travaux futurs, une vue globale faisant partie intégrante de MuzLink pourra être conçue et évaluée afin de faciliter la prise en main et augmenter l'explorabilité.

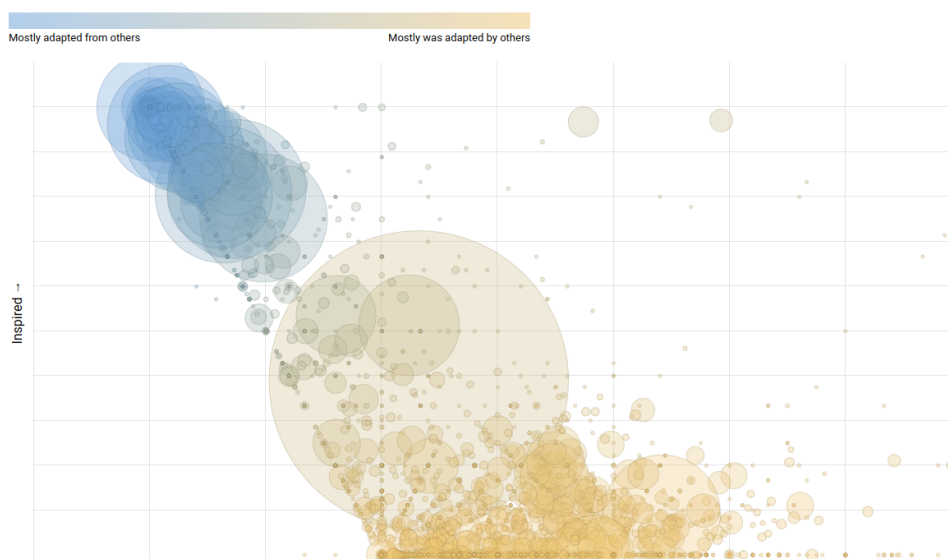


Figure 5.5 Vue globale présentant tous les artistes du jeu de données de BAnQ. Figure tirée de l'article «MusX : Online Exploring and Visualizing Graph-Based Musical Adaptations» par Lévesque et al. [2]. Chaque bulle représente un artiste. Sa taille varie en fonction du nombre de productions. La position encode la direction des inspirations de l'artiste. L'axe y encode le nombre de relations d'inspirations : le nombre de fois qu'un artiste s'est inspiré d'un autre. L'axe x encode le nombre de relations d'influences : le nombre de fois qu'un artiste a été influencé par d'autres.

5.5 Limites de l'évaluation

Les résultats de la recherche peuvent être faussés par plusieurs facteurs : la clarté des questions, la taille de l'échantillon, l'intérêt du participant ou son niveau de fatigue. Cette section présente les facteurs ayant un grand potentiel d'impact sur notre analyse.

Le questionnaire a été testé avec 2 utilisateurs volontaires avant le début de l'expérience. Cette période de test a révélé de nombreux problèmes dans la formulation des questions. La qualité du test a grandement été améliorée en réponse à leurs commentaires. Malgré cette analyse préliminaire, les questions 7 et 8 ont souvent été mal comprises des participants lors de l'expérience. Sans commentaires de leur part, il est difficile de comprendre l'origine de ces erreurs. Dans l'article, nous spéculons sur une mauvaise clarté des questions et du glossaire musicale. Un groupe de discussion permettrait de mieux comprendre l'origine de ces erreurs. Les commentaires des participants pourraient nous indiquer plus précisément si le problème provient du questionnaire lui-même, ou de la visualisation, et donneraient des pistes de solutions intéressantes pour des travaux futurs.

L'échantillon initial prévu était de 25, ce qui est standard pour des études de cas en visualisation de données. Ce nombre est cependant très variable selon la portée et la complexité de l'outil de visualisation évalué. Certaines études présentent des résultats avec seulement 3 ou 4 participants. 18 participants est donc un nombre acceptable pour une étude de ce genre. Sur l'aspect démographique, les participants sont très diversifiés. Cependant, la source des participants est unique. Tous les participants sont des personnes qui ont aimé la page Facebook de la BAnQ. Ce sont des personnes qui ont un intérêt pour participer à une expérience de recherche. Il est donc possible que l'échantillon ait un biais pour des personnes ayant une curiosité intellectuelle orientée vers la culture. En outre, la moyenne de scolarité de l'échantillon est plus élevée que la moyenne québécoise. Les gens moins scolarisés sont sous-représentés.

L'évaluation comporte un grand nombre de questions. Il est possible que les participants se démotivent ou se fatiguent au cours de l'expérience. Les dernières questions sont en moyenne plus souvent ratées que celles au début. Si la fatigue est effectivement la cause de ce taux d'échec plus grand à la fin de l'expérience, raccourcir l'expérience ou rendre aléatoire l'ordre des questions pourrait aider à diminuer ce biais.

Le mécanisme de rétroaction en fin de questionnaire est basé sur 4 aspects : la découvrabilité, le divertissement, la satisfaction et la compréhension. Les réponses des participants ont mené à de nombreuses conclusions pertinentes dans notre analyse. Cependant, comme il s'agissait de questions ouvertes, une partie des réponses manquaient de rigueur, ou déviaient de l'aspect

questionné. Il aurait été pertinent de créer des questions cadrées afin de recueillir des opinions sur des sous-aspects particuliers.

CHAPITRE 6 CONCLUSION

6.1 Synthèse des travaux

La BAnQ possède un jeu de données de relations entre artistes musicaux ayant une structure unique. Lors d'un projet d'innovation en valorisation de données, en vue d'encourager les emprunts d'albums musicaux, l'institut québécois souhaitait mettre en valeur ses données afin de propulser la découverte de relations entre artistes. Un processus de design orienté utilisateur a permis de sélectionner 11 tâches clés que l'outil de visualisation devait permettre d'accomplir. Étant une combinaison inhabituelle entre un graphe biparti et des arbres d'adaptations, la structure des données est complexe. Avec plus de 19 000 artistes et 24 000 chansons, elle est également très grande. Une revue de littérature a permis de constater que les méthodes de visualisation existantes ne permettaient pas de réaliser les tâches sélectionnées sur le jeu de données.

L'objectif du projet de recherche est de (1) concevoir une nouvelle méthode de visualisation pour accomplir des tâches d'exploration et de découverte de relations entre artistes sur le jeu de données de BAnQ, (2) de développer un prototype fonctionnel et (3) d'évaluer avec une étude formelle avec des utilisateurs.

Le prototype final, appelé MuzLink, est une visualisation à deux facettes centrée sur un artiste. La première facette, avec ses lignes du temps connectées, permet de visualiser les relations d'influences par rapport aux chansons de l'artiste de façon topologique. La deuxième facette, avec ses listes d'artistes, représente les artistes ayant le plus inspiré, collaboré ou ayant été inspirés par l'artiste. Des interactions permettent de relier les deux vues complémentaires afin d'élargir l'éventail de tâches possibles. Une barre de recherche et des liens vers d'autres artistes et d'autres chansons permettent une navigation fluide, propulsant l'exploration.

Une étude utilisateur formelle a été menée pour évaluer l'efficacité de l'outil à répondre aux 11 tâches sélectionnées en début de projet. Les résultats permettent de conclure que l'outil est adapté pour répondre aux questions de façon rapide et précise. Des questions ouvertes en fin de projet ont également révélé que la méthode de visualisation permet un excellent niveau de découvrabilité, qu'il est intuitif et qu'il communique les idées de façon satisfaisante. Une analyse des interactions enregistrées révèle que MuzLink a également un bon niveau d'explorabilité.

6.2 Limitations de la solution proposée

L'étude utilisateur a révélé quelques faiblesses. D'abord, le glossaire musical utilisé dans l'outil de visualisation ne semble pas avoir été compris de tous. Il pourrait être communiqué plus clairement, ou alors être modifié pour être plus intuitif. Rendre le tutoriel interactif pourrait également guider plus efficacement l'utilisateur. Quelques participants ont mentionné que la densité d'informations peut être difficile à lire et à comprendre. Afin de respecter le mantra de Shneiderman, l'ajout de filtres et d'une fonctionnalité de zoom pourrait aider l'utilisateur à fouiller les données avec plus de contrôle.

6.3 Améliorations futures

Pour des travaux futurs, MuzLink pourrait intégrer le mantra de Shneiderman en proposant une vue globale et générale, puis en proposant une fonctionnalité de zoom et de filtre à la demande. Cette approche a le potentiel d'améliorer la prise en main et l'explorabilité de l'outil. Une évaluation formelle avec des utilisateurs permettra de le confirmer.

Dans une optique plus large, nous croyons que les lignes du temps connectées de MuzLink proposent une méthode de visualisation qui n'est pas limitée au jeu de données de BAnQ. Une ligne du temps linéaire permet de représenter des items en séquence dans le temps. Quelques travaux passés, comme Citeology [47] et EdgeMaps [26], ajoutent des liens entre les items de la ligne du temps, permettant de visualiser l'évolution temporelle des relations entre les items d'un graphe. Les lignes du temps connectés de MuzLink ajoutent une dimension supplémentaire. Alors que les travaux passés se limitent généralement qu'à une seule ligne du temps, notre méthode propose l'utilisation de plusieurs lignes du temps parallèles. Pour des graphes où chaque noeud fait partie d'un ensemble exclusif, il est possible de tracer une ligne du temps pour chacun de ces ensembles. Les noeuds sont ensuite ajoutés sur les différentes lignes du temps selon leur ensemble respectif. Finalement, les liens dessinés entre les noeuds permettent d'analyser les relations entre les ensembles du graphe dans le temps. Il serait intéressant de tester cette méthode avec des jeux de données multiples, comme le cinéma, les publications scientifiques et les influences entre philosophes, ainsi qu'analyser les forces et les faiblesses de cette dernière à accomplir certaines tâches.

RÉFÉRENCES

- [1] C. Nobre, M. Streit, M. Meyer et A. Lex, “The state of the art in visualizing multivariate networks,” *Computer Graphics Forum*, vol. 38, p. 807–832, 2019.
- [2] F. Lévesque, M. St-Germain, D. Piché, J.-F. Gauvin, M. Gagnon et T. Hurtut, “Musx : Online exploring and visualizing graph-based musical adaptations,” dans *Proc. IEEE International Conference on Data Engineering*, Apr 2020.
- [3] T. Munzner, *Visualization analysis and design*. AK Peters Visualization Series, CRC Press, 2014.
- [4] J. BERTIN, “Semiology of graphics : diagrams, networks, maps,” 1983.
- [5] J. Heer et G. Robertson, “Animated transitions in statistical data graphics,” dans *Proc. IEEE Transactions on Visualization and Computer Graphics*, 2007.
- [6] C. Ware et R. Bobrow, “Motion to support rapid interactive queries on node-link diagrams,” *ACM Transactions on Applied Perception*, vol. 1, p. 3–18, 2004.
- [7] H. Schulz, “Treevis.net : A tree visualization reference,” *IEEE Computer Graphics and Applications*, vol. 31, n^o. 6, p. 11–15, 2011.
- [8] M. Brehmer, B. Lee, B. Bach, N. H. Riche et T. Munzner, “Timelines revisited : A design space and considerations for expressive storytelling,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, n^o. 9, p. 2151–2164, 2017.
- [9] H. Loi, T. Hurtut, R. Vergne et J. Thollot, “Programmable 2d arrangements for element texture design,” *ACM Transactions on Graphics*, vol. 36, n^o. 4, 2017.
- [10] C. B. Nielsen, S. D. Jackman, I. Birol et S. J. M. Jones, “Abyss-explorer : Visualizing genome sequence assemblies,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, n^o. 6, p. 881–888, 2009.
- [11] “Types of color blindness.” [En ligne]. Disponible : <https://nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness/types-color-blindness>
- [12] S. Ishihara *et al.*, “Tests for color blindness,” *American Journal of Ophthalmology*, vol. 1, n^o. 5, p. 376, 1918.
- [13] J. Umbrich, S. Neumaier et A. Polleres, “Quality assessment and evolution of open data portals,” dans *Proc. International Conference on Future Internet of Things and Cloud*, 2015, p. 404–411.
- [14] “Musicbrainz - the open music encyclopedia,” <https://musicbrainz.org/>.

- [15] “Discogs,” <https://www.discogs.com/>.
- [16] I. Herman, G. Melançon et M. S. Marshall, “Graph visualization and navigation in information visualization : A survey,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, n°. 1, p. 24–43, 2000.
- [17] T. Von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. van Wijk, J.-D. Fekete et D. W. Fellner, “Visual analysis of large graphs : state-of-the-art and future research challenges,” *Computer Graphics Forum*, vol. 30, n°. 6, p. 1719–1749, 2011.
- [18] H.-J. Schulz, T. Nocke, M. Heitzler et H. Schumann, “A design space of visualization tasks,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, n°. 12, p. 2366–2375, 2013.
- [19] J. Pretorius, H. C. Purchase et J. T. Stasko, *Tasks for Multivariate Network Analysis*. Springer International Publishing, 2014, p. 77–95.
- [20] M. Miller, J. Walloch et M. Pattuelli, “Visualizing linked jazz : A web-based tool for social network analysis and exploration,” *Proceedings of the American Society for Information Science and Technology*, vol. 49, 2012.
- [21] H. Neuweger, M. Persicke, S. Albaum, T. Bekel, M. Dondrup, A. Hüser, J. Winnebold, J. Schneider, J. Kalinowski et A. Goesmann, “Visualizing post genomics data-sets on customized pathway maps by prometra – aeration-dependent gene expression and metabolism of corynebacterium glutamicum as an example,” *BMC Systems Biology*, vol. 3, 2009.
- [22] T. J. Jankun-Kelly et Kwan-Liu Ma, “Moiregraphs : radial focus+context visualization and interaction for graphs with visual nodes,” dans *Proc. IEEE Symposium on Information Visualization*, 2003, p. 59–66.
- [23] B. Shneiderman et A. Aris, “Network visualization by semantic substrates,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, n°. 5, p. 733–740, 2006.
- [24] E. M. Rodrigues, N. Milic-Frayling, M. Smith, B. Shneiderman et D. Hansen, “Group-in-a-box layout for multi-faceted analysis of communities,” dans *Proc. IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*, 2011, p. 354–361.
- [25] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist et J. Fekete, “Graphdice : A system for exploring multivariate social networks,” *Computer Graphics Forum*, vol. 29, n°. 3, p. 863–872, 2010.
- [26] M. Dörk, S. Carpendale et C. Williamson, “EdgeMaps : visualizing explicit and implicit relations,” dans *Proc. Visualization and Data Analysis*, vol. 7868, 2011, p. 157 – 168.

- [27] M. Ghoniem, J.-D. Fekete et P. Castagliola, “On the readability of graphs using node-link and matrix-based representations : A controlled experiment and statistical analysis,” *Information Visualization*, vol. 4, n^o. 2, p. 114–135, 2005.
- [28] E. Kerzner, A. Lex, C. Sigulinsky, T. Urness, B. Jones, R. Marc et M. Meyer, “Graffinity : Visualizing connectivity in large graphs,” *Computer Graphics Forum*, vol. 36, n^o. 3, p. 251–260, 2017.
- [29] C. Mueller, B. Martin et A. Lumsdaine, “A comparison of vertex ordering algorithms for large graph visualization,” dans *Proc. International Asia-Pacific Symposium on Visualization*, 2007, p. 141–148.
- [30] N. Elmqvist, T. Do, H. Goodell, N. Henry et J. Fekete, “Zame : Interactive large-scale graph visualization,” dans *Proc. IEEE Pacific Visualization Symposium*, 2008, p. 215–222.
- [31] M. Meyer, T. Munzner et H. Pfister, “Mizbee : A multiscale synteny browser,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, n^o. 6, p. 897–904, 2009.
- [32] J. Seo et B. Shneiderman, “Interactively exploring hierarchical clustering results [gene identification],” *Computer*, vol. 35, n^o. 7, p. 80–86, 2002.
- [33] M. Graham et J. Kennedy, “A survey of multiple tree visualisation,” *Information Visualization*, vol. 9, p. 235–252, 2010.
- [34] Q. Nguyen et M. Huang, “Enccon : An approach to constructing interactive visualization of large hierarchical data,” *Information Visualization*, vol. 4, p. 1–21, 2005.
- [35] J. J. Van Wijk et H. van de Wetering, “Cushion treemaps : Visualization of hierarchical information,” dans *Proc. IEEE Symposium on Information Visualization (InfoViz)*, 1999.
- [36] M. Bruls, C. Huizing et J. J. van Wijk, “Squarified treemaps,” dans *Proc. Eurographics / IEEE VGTC Symposium on Visualization*, 2000.
- [37] E. R. Tufte, *The Visual Display of Quantitative Information*, 2^e éd. Graphics Press, 2001.
- [38] I. F. Cruz, W. Sunna, N. Makar et S. Bathala, “A visual tool for ontology alignment to enable geospatial interoperability,” *Journal of Visual Languages and Computing*, vol. 18, n^o. 3, p. 230–254, 2007.
- [39] N. Sheth, K. Börner, J. Baumgartner et K. Mane, “Treemap, radial tree and 3d tree visualizations,” dans *Proc. IEEE Information Visualization Conference*, 2003, p. 128–129.

- [40] W. Zainon et P. Calder, “Visualising phylogenetic trees,” dans *Proc. Conferences in Research and Practice in Information Technology*, 2006.
- [41] T. Munzner, F. Guimbretiere, S. Tasiran, L. Zhang et Y. Zhou, “Treejuxtaposer : Scalable tree comparison using focus+context with guaranteed visibility,” *ACM Transactions on Graphics*, vol. 22, p. 453–462, 2003.
- [42] D. Auber, M. Delest, J.-P. Domenger, P. Ferraro et R. Strandh, “Evat : Environment for visualization and analysis of trees,” dans *Proc. the IEEE Symposium on Information Visualization*, 2003.
- [43] G. W. Furnas et J. M. Zacks, “Multitrees : enriching and reusing hierarchical structure,” dans *Proc. Conference on Human Factors in Computing Systems*, 1994.
- [44] T. Dwyer et F. Schreiber, “Optimal leaf ordering for two and a half dimensional phylogenetic tree visualisation,” dans *Proc. Australasian Symposium on Information Visualization*, 2004.
- [45] E. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler et S. Card, “Visualizing the evolution of web ecologies.” dans *Proc. Conference on Human Factors in Computing Systems*, 1998, p. 400–407.
- [46] M. Dörk, N. Henry Riche, G. Ramo et S. Dumais, “Pivotpaths : Strolling through faceted information spaces,” dans *Proc. IEEE Symposium on Information Visualization*, 2012.
- [47] J. Matejka, T. Grossman et G. Fitzmaurice, “Citeology : Visualizing paper genealogy,” dans *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, 2012.
- [48] C. Nobre, M. Streit et A. Lex, “Juniper : A tree+table approach to multivariate graph visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, n° 1, p. 544–554, 2019.
- [49] F. J. Anscombe, “Graphs in statistical analysis,” *The American Statistician*, vol. 27, n° 1, 1973.
- [50] W. Aigner, S. Miksch, W. Müller, H. Schumann et C. Tominski, “Visualizing time-oriented data—a systematic view,” *Computers Graphics*, vol. 31, n° 3, p. 401 – 409, 2007.
- [51] A. Eklund, “R package beeswarm : The bee swarm plot, an alternative to stripchart,” <https://rdr.io/cran/beeswarm/>, 2016.
- [52] M. Okoe, R. Jianu et S. Kobourov, “Node-link or adjacency matrices : Old question, new insights,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, n° 10, p. 2940–2952, 2019.

- [53] C. Stoiber, F. Grassinger, M. Pohl, H. Stitz, M. Streit et W. Aigner, “Visualization onboarding : Learning how to read and use visualizations,” 2019.
- [54] E. M. Lidal, M. Natali, D. Patel, H. Hauser et I. Viola, “Geological storytelling,” *Computers & Graphics*, vol. 37, n°. 5, p. 445–459, 2013.
- [55] B. Shneiderman, “The eyes have it : A task by data type taxonomy for information visualizations,” dans *Proc. IEEE Symposium on Visual Languages*, 1996, p. 336–343.

ANNEXE A CAS D'UTILISATION

- Dans combien de langues distinctes la chanson Adeste Fideles a-t-elle été interprétée ? Et à partir de quand pour chacune des adaptations
- Quels sont les artistes ou auteurs-compositeurs étrangers privilégiés par les adaptateurs québécois et vice versa ?
- Combien de chansons des Beatles ont été adaptées en français ?
- Quelles chansons d'Elvis Presley sont des adaptations du français et quelles sont celles d'origine québécoise ?
- Quelle est l'origine de la pièce T'as-tu d'la bière icitte ?
- Dans quel pays a-t-on pu entendre Lindberg en anglais ?
- Quelle chanson de Félix Leclerc fut adaptée en allemand ?
- Sous quelle forme Mon Pays de Gilles Vigneault a-t-elle fait le tour de la planète ?
- Quelle fut la décennie où il y a eu le plus grand nombre d'adaptations en anglais de chansons françaises et vice versa.
- Quel est le palmarès des 10 chansons qui ont l'objet du plus grand nombre d'adaptations selon le jeu de données.
- Quelles sont toutes les chansons de Noël qui furent adaptées ?
- Quelles sont toutes les chansons tirées de films, de la radio ou de la télévision qui furent adaptées ?
- Quelles sont les comédies musicales qui furent adaptées ?
- Quelles sont les pièces classiques qui furent adaptées ?
- Quelles sont les chansons traditionnelles qui furent adaptées ?
- Quelles sont les chansons à thèmes religieux ?
- Quelles sont les chansons pour enfants qui furent adaptées ?
- Sur quelles pièces instrumentales des paroles furent-elles écrites ?
- Quelles sont les versions intermédiaires entre une chanson originale et son adaptation françaises ? Bref, quel est l'arbre généalogique d'une chanson.
- Quelles sont les adaptations qui ont connu un plus grand succès que l'original au regard des propres adaptations tirées de cette dernière ?
- Qui est le premier interprète d'une chanson (original ou adaptation) ?
- Quelles sont les chansons originales qui furent primées ?
- Quelles sont les différentes variations de titres d'une chanson ?
- Quelles sont les chansons qui sont issues ou devenues des publicités ?
- Dans quel genre musical y a-t-il le plus d'adaptations (country, blues, etc.) ? Cette

donnée n'est pas présente

- MCLC : mais si on se lie avec des sources qui contiennent cette donnée . . .
- J'aimerais obtenir les références complètes des albums qui comprennent l'adaptation par April March de "Laisse tomber les filles" (Serge Gainsbourg) (MCLC : je n'ai pas vérifié si cette adaptation figure dans les données, mais ce serait bien de pouvoir répondre à ce genre de questions en se liant à d'autres jeux de données !
- Il s'agit de Chick habit enregistré en 1996. Les détails des albums sont probablement disponibles sous Discogs.
- Quelles chansons parlent du Saguenay ?
- Quelles chansons de John Lennon ont été adaptées en français ?
- Quelles chansons écrites par un auteur québécois ont été adaptées dans une autre langue ?
- Quelles chansons ont été interprétées durant les Francofolies de 2018 ?

ANNEXE B CAS D'UTILISATION ABSTRAITS

Angle Paroles

- Quelle est l'évolution dans le temps des émotions, des thèmes, du style (richesse du vocabulaire, longueur des mots/phrases) ou de la durée des chansons ?
- Quelles sont les chansons (ou les artistes) les plus joyeux, les plus tristes ou ayant le plus de vocabulaire ?
- Quelles sont les chansons les plus courtes ou les plus longues ?
- Existe-t-il une corrélation entre la durée d'une chanson et son niveau de tristesse ?
- Existe-t-il une corrélation entre l'année de publication d'une chanson et son niveau de tristesse ?
- Est-ce que les chansons (ou les artistes) anglophones ont les mêmes thèmes que les chansons francophones ? adaptations ?
- Quel est le vocabulaire spécifique à un genre ?
- Quel est le vocabulaire spécifique à un artiste ?
- Évolution des mots les plus populaires dans le temps ?
- Quelles chansons parlent d'un certain thème ?

Angle Adaptations

- Comment la quantité d'adaptations évolue dans le temps ?
- Y-a-t-il une période dans le temps où l'on retrouve davantage d'adaptations ?
- Quelles sont les chansons (ou autorités, ou genres) les plus adaptés ?
- Quel est l'auteur qui a le plus adapté de chansons ?
- Quelles sont les émotions prépondérantes des chansons les plus adaptées ?
- Quelles sont les thèmes prépondérants des chansons les plus adaptées ?
- Quel est le niveau de vocabulaire des chansons originales avec adaptations versus les chansons originales sans adaptation ?
- Y-a-t-il une corrélation entre le niveau d'adaptation et la richesse du vocabulaire ?
- Quelle est l'origine d'une chanson précise ?

Angle Langues & Adaptations

- Dans combien de langues distinctes une chanson a-t-elle été interprétée ?
- Quelles - ou combien de - chansons d'un artiste d'une certaine langue ont été adaptées dans une autre langue ?
- Quelles chansons sont des adaptations du français et/ou d'origine québécoises ?
- Dans quel pays une chanson apparaît dans une autre langue pour la première fois ?

- Comment évolue la relation entre les langues d'origine et de destination des adaptations dans le temps ?

Angle Pays d'origine & Adaptations

- Quels sont les artistes ou auteurs-compositeurs étrangers privilégiés par les adaptateurs québécois ?
- Quelles - ou combien de - chansons d'un artiste d'une certaine origine ont été adaptées dans un autre pays ?
- Comment évolue la relation entre les pays d'origine et de destination des adaptations dans le temps ?

Angle Spectacles

- Quelles chansons ont été interprétées durant les Francofolies de 2018 ?

Angle Autorités

- Quelles sont les autorités les plus adaptés ?
- Quel est l'auteur qui a le plus adapté de chansons ?
- Évolution de la carrière d'une autorité ? (productions, thèmes, émotions, collaborations avec d'autres autorités) Certains éléments nécessitent des ajouts dans la bases (thèmes, émotions) à l'exception des catégories générales (ex. : Noël, télévision, film, etc.)
- Comment les autorités travaillent entre elles :
 - En fonction du pays (silo ?) d'origine et de travail.
 - Selon leurs rôles (interprète/compositeur/auteur)
 - Selon le genre de musique

ANNEXE C ANNONCE DE RECRUTEMENT

#BAnQ a récemment collaboré avec l'École @polymtl pour un projet d'innovation. Nous cherchons des volontaires pour participer à l'évaluation de cet outil de visualisation de données musicales, le 8 février 2020, à la Grande Bibliothèque.

Détails : <https://consultation.quebec.ca/processes/visualisation-musicale>

ANNEXE D COURRIEL DE RECRUTEMENT

Bonjour,

Vous nous avez récemment fait part de votre intérêt à participer à l'évolution du portail de BAnQ et c'est la raison pour laquelle nous vous contactons aujourd'hui.

En marge de ce projet, BAnQ a récemment collaboré avec l'École Polytechnique pour un projet d'innovation financé par le Secrétariat du Conseil du Trésor du Québec. C'est à ce titre que vous êtes convié à participer à l'évaluation d'un outil de visualisation de données musicales.

Ce projet de recherche en partenariat entre Polytechnique Montréal et BAnQ a pour objectif de proposer une méthode de visualisation pour la représentation des relations d'adaptation entre chansons et des relations entre artistes.

Trois certificats cadeaux chez Renaud-Bray, d'une valeur de 50\$ chacun, seront offerts à trois participants tirés au sort. Il est à noter que ce tirage est offert par Polytechnique et non par BAnQ.

L'évaluation aura lieu à la salle La Serre à BAnQ au 475 boulevard de Maisonneuve Est, Montréal, le 8 février 2020 à 10h30 ou 13h30 et sera d'une durée d'environ 30 minutes.

[Je veux participer](#)

Pour participer à l'évaluation, vous devez être âgé de 18 ans et plus, ne devez pas avoir de déficience visuelle affectant votre perception des formes colorées et devez être à l'aise avec l'utilisation d'un ordinateur.

Il est à noter que vous pourrez en tout temps quitter la salle si vous ne souhaitez plus participer à l'évaluation.

Pour toutes questions, adressez-vous à François Lévesque à l'adresse suivante : francois.levesque@polymtl.ca

Nous vous remercions à l'avance pour votre participation à l'évaluation de notre outil de visualisation.

Cordialement,


François Lévesque


Étudiant à la maîtrise en génie logiciel

Polytechnique Montréal

ANNEXE E AUTORISATION D'ACCÈS AUX LIEUX

Prêt d'une salle et d'équipement informatique

 Ce message a été identifié comme étant du courrier indésirable. [Courrier légitime](#)

 Parenteau, Véronique <vparenteau@banq.qc.ca>
12:22

À : 'francois.levesque@live.ca'



Bonjour,

Comme convenu, je confirme que BANQ accepte de vous prêter une salle (la Serre) ainsi que des ordinateurs portables afin de mener vos tests utilisateurs. L'horaire reste à déterminer et sera tributaire de la disponibilité de la salle.

Cordialement,

Véronique Parenteau
Bibliothécaire, développement et soutien
Direction de la médiation documentaire et numérique
Bibliothèque et Archives nationales du Québec

475, boulevard De Maisonneuve Est
Montréal (Québec) H2L 5C4
Téléphone : 514 873-1101 poste 6587
veronique.parenteau@banq.qc.ca
www.banq.qc.ca



BANQ-Avis de confidentialité

Ce courriel est une communication confidentielle et l'information qu'il contient est réservée à l'usage exclusif du destinataire. Si vous n'êtes pas le destinataire visé, vous n'avez aucun droit d'utiliser cette information, de la copier, de la distribuer ou de la diffuser. Si cette communication vous a été transmise par erreur, veuillez la détruire et nous en aviser immédiatement par courriel.

ANNEXE F FORMULAIRE DE CONSENTEMENT

Titre du projet de recherche :

Méthode de visualisation de graphes bipartis multivariés pour la représentation de relations entre artistes musicaux dans le temps (projet de recherche financé par Polytechnique Montréal)

Équipe de recherche :

François Lévesque

Candidat à la maîtrise en génial logiciel

École Polytechnique de Montréal

C.P. 6079, succ. Centre-ville

Montréal (Québec)

H3C 3A7

Tél. (514) 340-4711 poste 7109

Fax. (514) 340-5139

Adresse courriel : francois.levesque@polymtl.ca

Thomas Hurtut

École Polytechnique de Montréal

C.P. 6079, succ. Centre-ville

Montréal (Québec)

H3C 3A7

Tél. (514) 340-4711 poste 7109

Fax. (514) 340-5139

Adresse courriel : thomas.hurtut@polymtl.ca

Préambule :

Nous vous invitons à participer à un projet de recherche qui vise à évaluer l'efficacité d'une méthode de visualisation de données musicales complexes car vous répondez aux critères du participant cible auquel s'adresse cette visualisation.

Cependant, avant d'accepter de participer à ce projet et de signer ce formulaire d'information et de consentement, veuillez prendre le temps de lire, de comprendre et de considérer atten-

tivement les renseignements qui suivent. Nous vous invitons à poser toutes les questions que vous jugerez utiles au directeur de recherche ou aux autres membres de l'équipe de recherche et à leur demander de vous expliquer tout mot ou renseignement qui ne serait pas clair.

Présentation du projet de recherche et de ses objectifs :

L'objectif général du projet de recherche est de développer une méthode de visualisation inédite pour représenter plus efficacement les relations d'influence et de collaboration entre artiste, au travers des relations d'adaptation entre chansons.

Afin d'atteindre nos objectifs, nous avons développé un outil de visualisation interactif permettant d'explorer les relations qui existent entre des artistes au fil du temps. Nous avons conçu un questionnaire qui permettra d'évaluer la capacité de cet outil à accomplir certaines tâches.

Nature et durée de votre participation au projet de recherche :

Votre participation dans le cadre du présent projet consistera à répondre à une série de questions sur vos capacités de perception et de découverte avec un outil de visualisation de données. Toutes les réponses que vous fournirez seront confidentielles. Nous estimons que vous aurez besoin de 20 à 30 minutes pour compléter le questionnaire ci-joint. Si vous acceptez de participer au présent projet, veuillez simplement compléter le questionnaire ci-joint. Si vous acceptez de participer au présent projet, veuillez accepter les conditions énoncées dans ce formulaire puis remplissez le questionnaire en ligne.

Dans le cadre de ce projet, seules les personnes majeures n'ayant aucune déficience visuelle affectant la perception des couleurs seront incluses dans l'étude. De plus, vous devez être suffisamment à l'aise avec un ordinateur pour interagir avec les éléments du questionnaire fourni.

Avantages pouvant découler de votre participation au projet de recherche :

Un certificat cadeau d'une valeur de 50\$ chez Renaud-Bray sera offert à un participant choisi au hasard. Autrement, vous ne retirerez aucun bénéfice personnel de votre participation au présent projet de recherche. Toutefois, votre participation va permettre de valider une méthode de visualisation inédite, qui pourra être utilisée dans de nombreux domaines où des réseaux collaboratifs existent. Elle pourra aider et/ou inspirer d'autres chercheurs dans le milieu de la visualisation à représenter des graphes et des arbres dans le temps.

Si vous êtes intéressés par les conclusions de ce projet de recherche, nous pourrions vous envoyer une copie numérique du mémoire à la fin du projet. Il est à noter qu'il ne sera pas possible pour l'équipe de recherche de faire un lien entre vos réponses et le courriel fourni à la fin du questionnaire.

Inconvénients pouvant découler de votre participation au projet de recherche :

Votre participation au projet de recherche nécessitera entre 20 et 30 minutes de votre temps afin de répondre à un questionnaire en ligne.

Risques pouvant découler de votre participation au projet de recherche :

Le présent projet de recherche ne vous fera courir aucun risque additionnel que les risques que vous encourez dans votre vie quotidienne. Vous pourriez éventuellement ne pas savoir quoi répondre à une question. Dans une telle éventualité, sentez vous bien à l'aise de ne pas répondre à la question.

Compensation financière :

Un certificat cadeau d'une valeur de 50\$ chez Renaud-Bray sera offert à un participant choisi au hasard.

Indemnisation en cas de préjudice et droits du participant :

Si vous deviez subir quelque préjudice que ce soit par suite de votre participation à ce projet de recherche, vous ne renoncez à aucun de vos droits ni ne libérez les chercheurs, l'organisme subventionnaire ou l'établissement de leurs responsabilités légales et professionnelles.

Participation volontaire et possibilité de retrait :

Votre participation à ce projet de recherche est volontaire. Vous êtes donc libre de refuser d'y participer en ne répondant pas au questionnaire en ligne. De plus, vous pouvez à tout moment décider de vous retirer du projet sans avoir à motiver votre décision et sans risquer d'en subir de préjudice. Dans le cas où vous quitteriez le questionnaire en ligne avant la fin, seules les questions que vous aurez complétées pourront être considérées par les chercheurs.

Les chercheurs ou le comité de la recherche de l'École Polytechnique pourront retirer les participants sans leur consentement, s'ils ne respectent pas les consignes du projet de recherche ou s'il existe des raisons administratives d'abandonner le projet, notamment pour des raisons de sécurité et de faisabilité.

Confidentialité :

À la réception du questionnaire dûment rempli, le chercheur responsable du projet recueillera et consignera les réponses aux questions auxquelles vous aurez répondu. Dans le but d'assurer la confidentialité des participants, seuls les renseignements nécessaires à la bonne conduite du projet de recherche seront recueillis dans le questionnaire, tels que votre sexe, votre groupe d'âge ainsi que votre niveau d'éducation. Ces informations ne permettront pas de vous identifier. Elles seront utilisées pour évaluer la diversité de l'échantillon des participants ayant répondu au questionnaire afin de mieux comprendre les résultats obtenus. Par ailleurs, si vous

décidez de fournir votre adresse courriel à la fin du questionnaire afin de recevoir les conclusions de l'étude, cette information ne sera pas enregistrée avec vos réponses fournies. Ainsi, il ne sera pas possible de faire un lien entre votre courriel et vos réponses soumises. Tous les renseignements recueillis au cours du projet de recherche demeureront strictement confidentiels dans les limites prévues par la loi. Les données collectées par le chercheur principal seront conservées dans une base de données sécurisée par un mot de passe.

Diffusion des résultats :

Le chercheur responsable utilisera les données du projet de recherche pour les simples fins du projet de recherche. Les données du projet de recherche pourront être publiées dans des revues scientifiques ou partagées avec d'autres personnes lors de discussions scientifiques

Personnes ressource :

Si vous avez des questions concernant le projet de recherche, vous pouvez communiquer avec Thomas Hurtut au (514) 340-4711, poste 7109 ou encore par courriel à thomas.hurtut@polymtl.ca.

Si vous avez des questions concernant votre participation au projet de recherche, vous pouvez communiquer avec le président du Comité d'éthique de la recherche de l'École Polytechnique, M. Yuvin Chinniah, au (514) 340-4711, poste 2268 ou encore par courriel à yuv.chinniah@polymtl.ca

Consentement :

En répondant au questionnaire en ligne, vous acceptez de participer à ce projet de recherche aux conditions qui y sont énoncées. Veuillez garder une copie du présent document.

Accepter

Refuser

Soumettre

ANNEXE G QUESTIONNAIRE

1. Quel artiste a le plus souvent collaboré avec Dalida, tous rôles confondus, selon vous ?
2. Quel artiste a le plus souvent inspiré Dalida, tous rôles confondus, selon vous ?
3. Quel artiste a le plus souvent été influencé par Dalida, tous rôles confondus, selon vous ?
4. Quelle chanson de Jacques Brel est la plus influente selon vous ?
5. Parmi toutes les inspirations de Ginette Reno, la plus ancienne a été produite en quelle année ?
6. Sur l'ensemble de ses productions, Pierre Delanoë a-t-il été plus souvent auteur, compositeur ou interprète ?
7. Joe Dassin a adapté et repris des chansons de Pierre Delanoë. Sur l'ensemble de ces chansons, Joe Dassin a-t-il été plus souvent auteur, compositeur ou interprète ?
8. Est-ce que l'artiste Giangilberto Monti est le seul à avoir adapté Le poinçonneur des lilas de Serge Gainsbourg ?
9. La chanson Point of no return a été sélectionnée. Combien de fois a-t-elle été reprise et adaptée par d'autres artistes ?
10. Parmi ces 4 artistes, lequel influence le plus les autres (reprises + adaptations) selon vous ?
11. Parmi ces 4 artistes, lequel s'inspire le plus des autres selon vous ?

ANNEXE H RÉSULTATS BRUTS

#	Question	Correct	Incorrect	Passé	Total	Temps moyen (s)
1	Top collaborateur	17	1	0	18	15,8
2	Top inspireur	16	2	0	18	7,2
3	Top influencé	17	1	0	18	5,8
4	Rôle d'artiste	17	1	0	18	8,2
5	Rôle d'artiste 2	17	0	1	18	10,2
6	Lignes du temps : le plus inspirant	15	2	1	18	11,3
7	Lignes du temps : le plus inspiré	11	5	2	18	7,6
8	Chanson la plus influente	12	6	0	18	14,5
9	Compter adaptations et reprises	14	4	0	18	10,0
10	Chercher chanson et trouver adaptations	15	3	0	18	15,6
11	Date d'inspiration	14	3	1	18	10,1

ANNEXE I RÉPONSES AUX QUESTIONS OUVERTES

Avez-vous découvert quelque chose ?

Seules les réponses complètes sont présentées. Les erreurs d'orthographe ont été corrigées.

- U1 : Oui. De la "trivia" sur Dalida.
- U2 : Non.
- U3 : J'ai survolé le logiciel sans chercher à trouver une information en particulier. Toutefois il y a beaucoup à apprendre, les données contenues dans le logiciel sont très intéressante.
- U4 : Oui.
- U5 : Oui, c'est Claude François qui a écrit Comme d'habitude, et non l'inverse...
- U6 : L'outil est très intéressant, particulièrement pour les musicologues. J'ai hâte de voir les influences en musique classique et en musique contemporaine
- U7 : -
- U8 : Oui, toutes ces connections entre les créateurs sont très intéressantes.
- U9 : Oui, que j'ai beaucoup de choses que je ne connais pas
- U10 : Oui. L'intégration de plusieurs informations est pertinente ; on y découvre des relations que l'on n'aurait pas pu déceler aussi aisément sans cet outil.
- U11 : J'ai fait l'exploration rapidement, je ne me souviens pas des artistes mais je sais qu'il y en a que je ne connaissais pas. Avec cet outil, je pourrais aussi découvrir des chansons moins connues d'un artiste.
- U12 : L'outil est très convivial et permet d'avoir un grand aperçu sur les artistes, musiciens , auteurs, compositeurs de chanson, mais de cinéma je ne sais pas !
- U13 : Beaucoup d'influences et d'inspiration...
- U14 : Oui, les influences d'un mes artistes favoris (Michel Pagliaro), et ceux qui ont fait des reprises. Aussi appris qu'il a fait une reprise de Hey Jude.
- U15 : -
- U16 : Oui, la façon dont on peut associer différents activités connexes avec une thème particulier, dans l'expérience Claude François
- U17 : Oui, que Jacques Brel avait adapté ses propres chansons à quelques reprises.
- U18 : Oui

Êtes-vous satisfait de l'information véhiculée par la visualisation ?

Seules les réponses complètes sont présentées. Les erreurs d'orthographe ont été corrigées.

- U1 : Parce que l'interface est très conviviale, mais encore incomplète. Certains artistes plus "obscur" étaient absents.
- U2 : Parce que je préfère les livres et que de l'information inutile sur les statistiques et de trop.
- U3 : J'ai l'impression que le logiciel n'est pas très complet au niveau de la quantité de données. En cherchant des artistes plus actuels, internationaux, il n'y avait pas de fiches pour eux.
- U4 : Les données semblent très complètes.
- U5 : J'ai appris des choses, et j'en aurais appris davantage avec plus de temps. Il faut quand même prendre le temps de se familiariser un peu avec la plateforme, mais ça se fait assez bien.
- U6 : Clarté, navigabilité, concision malgré la grande densité d'information.
- U7 : Il est très instructif et agréable de voir l'information sous cet angle. On est interpellé pour découvrir des liens entre les créateurs, et découvrir d'autres oeuvres.
- U8 : La possibilité d'explorer est vraiment intéressante
- U9 : C'est personnel, j'apprends plus en pratiquant qu'en lisant et faisant une fois.
- U10 : Parce qu'on y trouve une mine d'informations grâce à une banque de données très large. J'ai été fasciné par tout ce qu'il est possible de découvrir.
- U11 : Je suis satisfait, je dirais qu'il faudrait peut-être des couleurs plus flash. Peut-être aussi des formes différentes, les cercles pour une influence, un carré pour une production originale, un polygone etc... Pour ma part je m'habituerai en utilisant l'outil, mais ça pourrait aider des personnes ayant plus de difficultés.
- U12 : ma perception de ce que j'ai eu de cet outil web. merci
- U13 : Trop d'informations nouvelles ; interface claire mais inhabituelle.
- U14 : J'ai pu répondre à plusieurs questions au sujet d'artistes que je connais à peine, impressionnant.
- U15 : C'est intéressant mais il manque peut-être un aspect visuel (pochettes ? visages ?) et audio au document pour le rendre plus attrayant ?!
- U16 : J'ai trouvé que toujours il y a des choses pour améliorer
- U17 : Pour certains artistes, le nombre de chansons est si grand qu'on a de la difficulté à naviguer. J'aurais aussi aimé pouvoir "dé-sélectionner" une chanson.
- U18 : Aspect visuel intuitif donne information rapide.

Êtes-vous divertit par la visualisation ?

Seules les réponses complètes sont présentées. Les erreurs d'orthographe ont été corrigées.

- U1 : Colorée, mais pas trop. Transitions très "smooth". Aucunement agressant à utiliser.
- U2 : -
- U3 : Voir à quel point une production en a influencé d'autres est très impressionnant. C'est intéressant voir les relations entre artiste, et le logiciel est très facile à utiliser.
- U4 : Le concept est très intéressant.
- U5 : Je trouve que les points sont petits, c'est difficile de cibler le bon. Sur un portable, ça doit être encore pire.
- U6 : Nouveau pour moi.
- U7 : Je pourrais passer des heures à faire des recherches dans ce type de tableau.
- U8 : -
- U9 : Parce que le logiciel est bien fait.
- U10 : C'est un outil de recherche (parfois pointue) d'informations, et donc n'est pas un jeu et ne me semble pas conçu pour "divertir". Oui, on peut s'amuser un peu tant qu'on y explore la nouveauté de l'interface mais les gens qui comptent l'utiliser régulièrement veulent des résultats précis, exacts, pertinents, qui permettent de faire des associations qu'ils ignoraient auparavant.
- U11 : J'ai bien aimé la visualisation, j'ai même souri en me rendant compte des liens entre chaque chansons. Une fois l'apprentissage fait, c'est intéressant de voir les influences, la grandeur de celle-ci par exemple (grosseur des cercles).
- U12 : Il y a plusieurs champs qui nous donne un large spectre de l'artiste, sa vie et sa création.
- U13 : Site de musique, sans aucun lien pour l'écouter ! Lien visuel entre la partie supérieure et inférieure peu évident. Surplus de liens vers la partie gauche (nuage de points) : difficile à distinguer visuellement ; manipulation du curseur très, trop fine.
- U14 : Je suis très visuel en général, je comprends toujours mieux les choses quand j'ai une façon de les représenter. Avec la grosseurs des points et la quantité des liens par exemple, on voit très bien l'importance des relations et des productions.
- U15 : C'est chouette mais ça demeure un peu aride (graphique) comme présentation (cela dit c'était intéressant de participer à cette expérience !)
- U16 : Les bulles sont agréables, les fils aussi, je trouve que les couleurs ne sont pas fatigants.
- U17 : C'était étonnant de voir les relations entre les oeuvres.

- U18 : Appris de nouvelles choses.

La visualisation est-elle facilement compréhensible ?

Seules les réponses complètes sont présentées. Les erreurs d'orthographe ont été corrigées.

- U1 : Quand un artiste est trop influent ou influencé etc. le nombre de petits ronds peut devenir un peu difficile à lire.
- U2 : Très compréhensible.
- U3 : -
- U4 : Je crois qu'avec un peu plus de pratique, ça doit devenir plus facile à naviguer, mais pour une première fois, c'est un peu mélangeant.
- U5 : Pas trop de détails, c'est aéré.
- U6 : -
- U7 : Une fois qu'on s'est adapté, on a de plus en plus de facilité à naviguer dans le tableau.
- U8 : Peut devenir rapidement confus quand on est pas familier avec l'interface
- U9 : Le logiciel illustre bien
- U10 : "Les traits entre les "disques" (relation entre "production" et "influences") ne sont pas visuellement assez clairs. Il faudrait pouvoir agrandir la section pour mieux voir les correspondances. Quand il y a plusieurs langues pour les adaptations, le défilement sur la même ligne n'est pas intuitif. Il faut connaître d'avance le truc pour accéder à la partie non visible de cette ligne."
- U11 : Il faut s'habituer à l'outil, ce n'est pas difficile mais il faut s'adapter et comprendre ce que l'on regarde. Par contre il devient intéressant pour une recherche musicale approfondie.
- U12 : Les contrastes des couleurs, les tableaux en bas de page sont un peu trop restreints. le choix des couleurs des picots est à revoir.
- U13 : Question d'habitude sans doute, mais la recherche en musique se fait beaucoup par genre musical, et pas toujours par auteur ou titre. Probablement plus adapté pour des chercheurs, où des recherches précises et pointues. Pour une recherche de découverte simple, l'interface me semble trop complexe. Impossibilité de zoomer dans les nuages de points, trop denses.
- U14 : La présentation est très claire, à la fois visuellement, et avec les détails par écrit et les quantités bien affichés en nombre, donc les 2 façons de voir les choses sont disponibles et facile à utiliser.
- U15 : Au début, je craignais de ne pas savoir démêler cet écheveau de liens, mais

finalement (enfin, vous verrez, selon mes résultats!!.....) je crois que j'ai réussi à apprivoiser le système....

- U16 : Ce n'était pas très difficile de comprendre.
- U17 : Parfois, les lignes qui tracent les relations entre les éléments croisent d'autres bulles et on ne sait pas exactement quelle oeuvre est liée à quelle chanson.
- U18 : Compréhension simple et efficace

Commentaires généraux

Seules les réponses complètes sont présentées. Les erreurs d'orthographe ont été corrigées.

- U1 : Excellent outil. Je lui souhaite un bel avenir !
- U2 : Non.
- U3 : Merci! J'ai beaucoup aimé tester ce logiciel et aimerais l'utiliser lorsqu'il sera plus complet.
- U4 : L'aspect esthétique est plus ou moins agréable (les petits ronds agglomérés pour représenter les chansons).
- U5 : J'aurais continué encore longtemps à explorer !
- U6 : Continuez. Superbe outil culturel.
- U7 : J'ai remarqué que les noms qui commencent par É apparaissent après les Z dans l'ordre alphabétique des collaborateurs. Et j'ai aussi remarqué qu'Alfa Rococo n'apparaît pas dans ceux qui ont fait une reprise du Poinçonneur des lilas.
- U8 : -
- U9 : Oui, est-ce que ce logiciel pourra aidé les artistes à aller chercher des redevances pour leur travail.
- U10 : Bravo pour cette initiative intéressante ; il y a un très bon potentiel. J'aimerais beaucoup qu'un tel outil existe aussi pour la musique classique et serais heureux de participer à son élaboration.
- U11 : Étant fan de toutes musiques confondues, j'aime bien cet outil. Je pourrais passer beaucoup de temps devant celui-ci et laisser aller ma curiosité. Je crois que je pourrais même perdre la notion du temps.
- U12 : J'aurai aimé qu'il y ai un peu de jeux d'images, des images dynamiques en 3D sur certaines champs au lieu que ça soit très statique.
- U13 : Est-ce que les données de ce projet seront «ouvertes», disponibles pour d'autres usages, à l'interopérabilité avec d'autres ressources. En ce moment, il n'y aucune fonctionnalité 2.0 : historique, partage, favoris, playlist, etc. L'exhaustivité des données est parfois un piège qui mène à un trop-plein d'informations. Un seul interface, une seule

navigation ne peut convenir à tous les usagers et pour tous les usages. Il manque des informations sur la constitution des données, leur agrégation, les ouvrages de références utilisés. C'est un projet de la BAnQ, de la Bibliothèque nationale et centrale du Québec, alors des informations complètes sur la constitution de cet ensemble de données est nécessaire et obligatoire. C'est le rôle et la mission de BAnQ d'expliquer et de diffuser des informations sur ses publications, comme cet ensemble de données musicales. Il n'y a aussi aucun crédit de mentionner, les membres de l'équipe, les responsables (techniques ou documentaire), etc. Sur ce plan, ce projet manque d'une présentation rigoureuse. Question : y a-t-il un calendrier et un échéancier pour l'implantation de cet outil de visualisation ?

- U14 : Non.
- U15 : Non, merci de m'avoir invitée à participer à cette expérience, j'espère que vous serez content de ce qu'elle vous aura apporté et révélé à vous, chercheur / développeur / idéateur !
- U16 : Félicitations pour l'engagement, le travail dur. L'éducation c'est une manière de contribution à la société. Bonne continuation.
- U17 : Il y a beaucoup d'éléments dans l'interface. Pour un usager "lambda", c'est quand même complexe. Peut-être que l'utilisateur pourrait choisir quelles informations l'intéressent pour pouvoir paramétrer l'affichage ? Ou pouvoir zoomer ?
- U18 : Non.