Visual Exploration of a Historical Vietnamese Corpus of Captioned Drawings: A Design Study

Kailiang Fu, Tyler Gurth, David H. Laidlaw, and Cindy Anh Nguyen

Abstract—This paper presents a design study focusing on the exploratory visual analysis of a unique historical dataset, consisting of approximately 4000 visual sketches and associated captions from a primary historical book published in 1909-1910. The book, which offers insight into Vietnamese crafts and social practices, poses the challenge of extracting cultural meaning and narrative structure from thousands of drawings and multilingual captions. Our research aims to explore and evaluate the effectiveness of multiple visualization techniques in uncovering meaningful relationships within the dataset while working closely with professional historians. The main contributions of this study include refining historical research questions through task and data abstraction, combining and validating visualization techniques for historical data interpretation, and involving a focus group of historians for further evaluation, sharing generalizable insights valuable for future domain-specific visualization tools.

Index Terms— Visualization design and evaluation methods, Arts and humanities, Exploratory data analysis

1 Introduction

Our project was motivated by an encyclopedic book published in 1909-1910, representing social and material life in Vietnam at the time through a set of around 4000 images with French and Vietnamese captions. In 2009, a re-edition of the book was produced with English translations, and Vietnamese transliterations of the Nôm script [1]. We are treating this document as a set of entities, with each image and its captions representing a single entity. The historical research problem involves looking for social and cultural meaning within the dataset, from both individual entities and the relationships between them at various levels of detail.

In this study, we have utilized visualization tools to reveal significant relationships within the data. Our primary focus has been on exploration, hypothesis generation, and analysis, rather than communicating historical findings.

Collaborating closely with Dr. Cindy Nguyen, a professional historian and co-author who possesses in-depth knowledge of book history and the dataset, we investigated the effectiveness of various visualization tools, their contributions to the analysis, and their limitations. Moreover, we address open problems in the visual analysis of data like ours. We treat the data in this book as a set of entities, with each image and its captions a single entity. The historical research problem is to look for social and cultural meaning in that set of entities. Some meaning comes from individual entities, but there is also potential historical meaning in how each image relates to the others at many different levels of detail.

We sought visualization tools to help find those meaningful relationships and report which tools were helpful, how they helped, and where they were limited. In close collaboration with Cindy Nguyen, a leading scholar in the field of digital humanities and Vietnamese history, we carried out the design and evaluation process.We then conducted a focus group comprising three humanities researchers. This group further assessed the effectiveness of three visualization methods, providing a more comprehensive understanding of their applicability to historical data. We also report on open problems we found that could affect visual analysis of data like ours. ▶R4-4

This paper makes three main contributions:

1. In Sec. 3, we identify three generalizable tasks for visualization of historical data: gaining an overview, generating research questions and hypotheses, and contextualizing data. These tasks support analysis at the corpus, discrete, and relational levels, respectively.
2. In Sec. 5 and Sec. 6.1, we evaluate and compare six categories of common visualization techniques, detailing the strengths and weaknesses of each technique with respect to the three identified tasks.
3. In Sec. 6.2, we offer guidelines for future researchers to choose appropriate visualization methods based on their objectives and familiarity with the dataset.

We anticipate this paper to be primarily beneficial for humanities scholars, especially those new to visualization research. Given the unique datasets and insights from domain experts, our findings cater to their specific demands and interests in understanding historical data visually. Additionally, visualization scholars might find value in our identification of common shortcomings in current methods, offering opportunities to develop improved approaches for similar datasets. ▶R2-2

2 Related Work

In the following sections, we outline the related works of this research, drawing from the literature on visualization methods, evaluation design, and digital humanities. ▶R1-2 ▶R3-2 ▶R4-2 ▶R2-3

2.1 Visualizing and Assessing Data in Research

Visualization methods have become increasingly successful and widely used in recent years, enabling researchers to uncover patterns and improve communication with their audience. Various techniques, such as hierarchical clustering [2], force-directed graphs, minimum spanning trees (MST), shaded and seriated matrices [3] [4] [5], and dimension reduction [6] [7] cater to diverse research topics.

An MST, for instance, is employed to visualize high-dimensional molecular data while preserving global and local features [8]. t-SNE

Kailiang Fu is with Brown University. E-mail: kailiang_fu@alumni.brown.edu

Tyler Gurth is with Brown University. E-mail: tgurth21@gmail.com.

David H. Laidlaw is with Brown University. E-mail: david_laidlaw@brown.edu.

Cindy Anh Nguyen is with University of California, Los Angeles. E-mail: cnguyen@seis.ucla.edu.

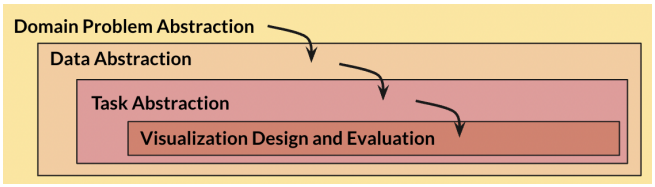


Figure 1: Four Layers of Revised Nested Model for Visualization Design

is utilized for assessing soil liquefaction in environmental studies [9], and UMAP is applied to recognize patterns in single-cell DNA data in bio-engineering [10]. These examples demonstrate the diverse applicability of visualization techniques in different research fields, which inspires our investigation into their effectiveness on visual-textual historical data.

While some art history studies have utilized deep convolutional neural networks (CNN) to cluster similar drawings [11] [12] or use image processing and color analysis to identify relationships among images and inspire conclusions [13], no universal standard exists for gauging the quality of these visualizations, as humanities research projects have diverse goals and contexts. We aim to explore the possibility of establishing generalizable tasks for visualizing complex humanities datasets.

2.2 Design Study

We have designed our study based on a modified version of Munzner's Nested Model for visualization design, and validation [14]. As illustrated in Fig. 1, we begin by abstracting the domain problem in this revised version. Following that, we extract common properties of data from the domain. We then design corpus-level, discrete-level, and relation-level tasks tailored to the domain and data characteristics. Finally, we create and evaluate various visualizations based on the task abstractions. It is important to note that the tasks are influenced by Shneiderman's principles for developing visualization systems, which are providing two levels of overview, and offering details on demand [15]. Much like our own experiment, McKenna et al's activity framework breaks visualization problems into different "activities" as a means of visual exploration. In our own experiment, each visualization method was explored on its own with its own motivations and desired outcomes. [16] Syeda et al present a different, more expedited approach to studying visualization problems: the design study "lite" methodology [17]. While a lack of time constraints moved us away from this approach, it depicts one of the many ways to come up with designs in the field of humanity-focused visualizations.

▶R1-2

Kucher et al survey a large variety of textual visualizations, similar to our own exploration of different visualization methodology. [18] However, Kucher's experiment focuses on textual information and the display of text, while our own focuses far more on the resulting sorting of images, their analysis, and greater relation within their associated culture.

As demonstrated in the design methodology suggested by Sedlmair et al [19], there are two axes to measure the contribution of design studies: task clarity and information location. By testing different visualization methods based on the same tasks, we validate the clarity of task abstraction, which is the first axis. Our study advances on the second axis, information location, by closely collaborating with expert Cindy Nguyen at each of the four design stages. This collaboration allows us to best evaluate the balance between information remaining as implicit knowledge in the expert's head and data or metadata available in a digital form that can be incorporated into the visualization.

2.3 Digital Humanities

Digital humanities use computational tools to help generate research questions, analyze data, and create visualizations. As more tools and data become available, digital humanities have become increasingly important for humanities researchers. Previous research has shown that visualization tools can be effective for identifying patterns in data [20] and presenting them in a meaningful way [11]. However, these studies have typically focused on a single visualization technique and compared it to no visualization at all. Our case study offers a more comprehensive comparison of different visualization techniques applied to the same dataset, providing researchers with a range of options to start with. In agreement with Bradley et al's statements regarding a need for strong collaboration in the digital humanities [21], our team exemplifies this core principle; our team is composed of three computer scientists (two of whom are students double majoring in art history and history) and an expert in the digital humanities field. This convergence of perspective led to a greater breadth of understanding and exploration throughout our study, as Bradley noted. Janicke et al explore the visual text analysis in digital humanities, similar to our own study. [22] More specifically, they investigate differences in close and distant readings with these analyses, just as our experts analyzed different tools we used as being effective on macro and micro levels. Some tools were better for understanding overall structure of the dataset, while others were better performing for small views of specific themes. Janicke et al explores text analysis, however, removed from image analysis. Windhager et al, in a similar manner to our own survey, survey a wide variety of different visualization techniques for cultural analysis. On a somewhat differing path, however, they play with image vectors and do not ground their study in one corpus as challenging as the ink-drawing, multi-lingual dataset we study [23].

▶R3-2

3 Abstraction

In this section we outline the domains of our contribution and the common challenges faced by historians when analyzing data (Sec. 3.1). Next, we summarize the characteristics of the dataset used in this study (Sec. 3.2). We then present the task abstraction, which aids historians in conducting multi-level analyses (Sec. 3.3).

3.1 Domain Abstraction

In this subsection we define our application domain and some of its challenges and opportunities.

▶R1-3
▶R2-4

3.1.1 Domain Definition

Our visualization tools contribute to the domain of history, a wide field of study which examines cultural, political, and economic transformations of past human society. We address two specific subfields of history, specifically book history and cultural history. Book history is a branch of history that examines the content, production, distribution, and reception of books broadly conceived as manuscripts and other printed materials. This area of study explores the social, cultural, economic, and technological aspects of books, as well as their roles in the transmission of knowledge, ideas, and information. Cultural history seeks to understand cultural practices and historic forces within the development of society.

▶R2-3

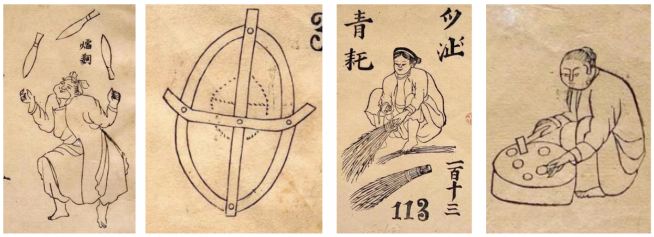
3.1.2 Domain Obstacle

Key challenges in this field encompass the extensive corpus size of certain books, cultural and linguistic obstacles, and the presence of limited or fragmented sources. The history domain requires the ability to navigate, organize, and sort vast amounts of uneven data. Visualization techniques offer new ways of investigating patterns and meaning within a complex set of historical data. We hypothesized that visualization methods applied to this specific dataset could contribute deeper understanding of the production of the book (book history) as well as Vietnamese social and material life (cultural history).

▶R4-3
▶R1-3

3.2 Data Abstraction

The data we use in this study is of great size and diverse format, reflecting common problems that historians face in their research.



Knife juggler | Body of a heater | Making brooms | Tinsmith making cups

Figure 2: Four example images from the dataset with their English captions

Page Number	Image Number	French Text	Viet Char Text
98	918	Images populaires.	Từ bi oản quả giơ tay oản quả
98	918	Images populaires.	Nuôi ong tay áo
98	918	Images populaires.	Rán sành ra mỡ
98	918	Images populaires.	Ăn cây nào rào cây ấy
98	919	Images populaires.	Kinh Châu phó hội

Figure 3: One Image Having Multiple Rows

This study's data is from the book "Technique du peuple annamite - Kỹ thuật của người annam - Mechanics and Crafts of the Vietnamese People" published in 1909-1910 and produced by a French colonial administrator Henri Oger and unnamed Vietnamese contributors. The 700-page book comprises 4,356 drawings, 4,428 captions in French, 4,428 captions in English (translated from the French captions), and 2,904 captions in Vietnamese chữ Nôm (a logographic Sinitic writing system of the Vietnamese language). As a comprehensive encyclopedia of Vietnamese material culture, it encompasses a wide range of topics such as art, literature, religion, music, fashion, food, and many others that make up the fabric of society. Dr. Nguyen and her students created an excel sheet of the French, English translation, and Vietnamese character Nôm caption, corresponding metadata, and manually labeled characteristics based on a social scientific methodology called content coding. One example of content labels we created was if the text or the drawing has gendered characteristics such as female, male, mixed, uncertain, or not applicable. Because one image can have multiple Vietnamese text captions corresponding to various parts of the image, as shown in Fig. 3, we generated an Excel sheet with 4,454 rows.

The resulting Excel sheet thus has 8 columns: "Page Number," "Image Number," "French Text,"" Viet Char Text," "English Translation," "Link to Image," and "Gendered based on Visual" as shown in Fig. 4.

3.3 Task Abstraction

To facilitate the design study, we have established a task abstraction that will assist historians in achieving multilevel objectives. These tasks will also be the primary focus of our visualization design. Task T1, T2, and T3 focus respectively on the dataset's corpus, discrete, and relational levels. By helping researchers to perform the three tasks, the tool would enable them to conduct multilevel analysis of visual and multilingual text elements. It is worth noting that these tasks are not mutually exclusive and can be performed in different orders, multiple times, and in combination with each other. For example, researchers can go back to T1 more than one time to position their research topics within the general picture.

3.3.1 T1 Gain Overview of the Dataset

Corpus Level : Given the large size and multilingual nature of our dataset, it is imperative for historians to quickly and efficiently gain an overview. But it is challenging to browse 4,454 images with three language captions and their gender characteristics. Our visualization needs to help researchers quickly gain a general understanding of the dataset. The goal is to develop an initial sense of the data, including its structure, size, scope, and main features. For example, a useful overview of the dataset is to understand general large-scale clusters to break down the dataset into manageable entry points to focus on. Another general overview of the dataset is to gain a sense of there are strong similarities within the dataset that even generate meaningful clusters. In other words, are there recognizable patterns and clusters in the dataset that merit additional investigation? Or is there not enough information within the dataset to uncover clusters of similarity?

3.3.2 T2 Generate Research Questions and Hypotheses

Discrete Level: Given the rich details encapsulated in every data entry - be it the image, multilingual captions, or metadata - historians should be able to drill down to individual entries effortlessly. By providing extensive information about each point, the tool should support the investigation of their unique attributes and features. This thorough examination can lead to the development of new research questions or hypotheses that require additional exploration.

3.3.3 T3 Contextualize Data

Relational Level: Visualizations can assist researchers see relational connections between thematic representations and embed them within a more complex social world. For example, if one is interested in how Vietnamese women are represented in the colonial period, a naive search for 'women' will only yield a limited number of results, missing captions with "midwife" or "sorceress" or other complex representations of gendered female social worlds involving objects such as clothing, medicine, and labor. Instead, our tool needs to support historians in efficiently summarizing as much relevant information as possible to facilitate their research on a specific topic.

4 Design

In this section, we discuss how we vectorize our data into a high dimensional vector space (Sec. 4.1) and how we visualize the embedded dataset based on the embedding (Sec. 4.2). Evaluations of the different methods follow in Sec. 5.

4.1 Data Embedding

In our study, we examine both textual and visual embeddings. Textual embedding, a subfield of Natural Language Processing (NLP), focuses on converting text into numerical representations called "embeddings." We opted for BERT (Bidirectional Encoder Representations from Transformers) as our textual embedding method due to its state-of-the-art performance across various NLP benchmarks, surpassing traditional techniques like bag-of-words models, TF-IDF, and word2vec. Additionally, BERT supports multiple languages, making it suitable for our trilingual dataset.

We utilize mpnet, a modified version of BERT and a top-performing sentence transformer, capable of mapping sentences and paragraphs into a 768-dimensional dense vector space. This feature is ideal for applications such as clustering and semantic search. With these embeddings, we create various visualizations to aid historical research.

Regarding visual embedding, we employ Yale DHLab's Pixplot, which uses convolutional neural networks (CNNs) to embed images. We will discuss this method further in Sec. 4.2.5.

4.2 Data Visualization

In the following sections, we outline the various visualization techniques assessed during our design study. Additionally, we present an illustrative example figure for each visualization method.

4.2.1 Distance Matrix

We calculate the cosine distance between each pair of captions based on the cosine similarity of their embedding vectors. For any two

Page Number	Image Number	French Text	Viet Char Text	English Translation	Link to Image	Gendered
1	1	Devin aveugle.	Thày bói	Blind medium.	edu/research/vis/	Male
1	2	Gestes du mendiant.	Ăn mày	Beggar's gestures.	edu/research/vis/	Male
1	3	Le cureur d'oreilles.	Lấy ráy tai	Ear cleaner.	edu/research/vis/	Male
1	4	dehors de la maison.	đường	Childbirth outside the home.	edu/research/vis/	Female
1	5	vie magique.	NA	world of magic.	edu/research/vis/	Mixed
1	6	viande.	NA	Butcher.	edu/research/vis/	NA
1	7	Rabot.	NA	Plane.	edu/research/vis/	NA

Figure 4: Excel Selection with Columns: Page Number, Image Number, French Text, Viet Char Text, English Translation, Link to Image, and Gendered based on Visual

English	French Text	Viet Char	The dog swir	Children swi	Cultured mai	Man urinatir	Filing shells f	Tool to facili	Bundled volu
The dog s	Baignade du	Tắm chó	0	0.541	0.975	1	0.91	0.914	0.947
Children	Gamins en t	Trẻ con lă	0.541	0	0.907	0.899	0.94	0.918	0.956
Cultured	Lettré écrivant des car		0.975	0.907	0	0.685	0.904	0.99	0.949
Man urin	Homme urinant en ple		1	0.899	0.685	0	0.954	0.914	0.995
Filing she	Limage des c	Dùi ốc	0.91	0.94	0.904	0.954	0	0.605	0.821
Tool to fa	Instrument pour facilit		0.914	0.918	0.99	0.914	0.605	0	0.849
Bundled	Empaquetage des volu		0.947	0.956	0.949	0.995	0.821	0.849	0
Method t	Mode de compression		0.977	0.976	0.858	0.983	0.719	0.783	0.694
Connecti	Mode de jor	Bó cặp sác	0.994	0.989	0.883	0.982	0.82	0.886	0.534
Repairing	Réparation c	Hàn chun	0.857	0.98	0.891	0.979	0.911	0.961	0.98
Doctor sc	Médecin au	Thày thuố	0.961	0.953	0.977	0.978	0.932	0.881	0.963
Students	Les étudiant	Cống Thậ	0.95	0.91	0.907	0.961	0.938	0.902	0.965
Students	Étudiants en train de c		0.909	0.877	1	0.965	0.981	0.816	0.997
Laureate	Docteur rev	Ông Tiến	0.995	0.988	0.889	0.894	0.977	0.936	0.982

Figure 5: Colored Reordered distance matrix; Darker Color Represents Higher Similarity

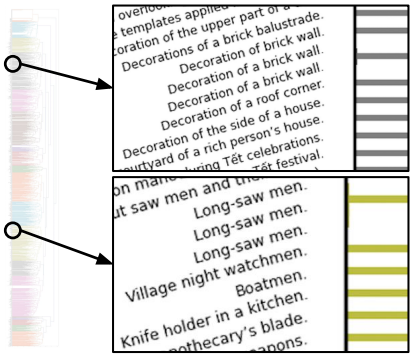


Figure 6: Hierarchical Clustering, Horizontal View

embeddings, A and B, the cosine distance is calculated by

$$\text{Cosine Distance} = 1 - \frac{(S_c(A,B)) + 1}{2}$$

$$S_c(A,B) \in [-1,1]$$

Where S_c indicates the cosine similarity between two vectors and is calculated by:

$$\text{Cosine Similarity} = S_c(A,B) = \frac{A \cdot B}{||A||||B||}$$

The formula indicates that the cosine distance between two captions ranges from 0 to 1. A score of 0 means the captions are very similar, while a score of 1 means the captions are completely dissimilar. We created three distance matrices based on this formula for three different languages. Each distance matrix has 4,454 rows and columns. Fig. 5 is a selection of the distance matrix.

To make it easier to compare similar data points, we re-ordered the matrices with the following steps derived from hierarchical clustering:

1. Compute the distance matrix for data points (rows or columns).
2. Initialize individual clusters for each data point.
3. Merge the pair of clusters with the smallest **average** distance.
4. Recalculate distances between the new cluster and the remaining ones.
5. Repeat steps 3-4 until only one cluster remains. Extract the clustering hierarchy (dendrogram).
6. Reorder the matrix such that it's row and column is the order of data points being merged in the dendrogram.

After reordering, captions with similar scores were placed closer together. We also used colors to better visualize the distance scores. Cells with lower scores (closer to 0) were given darker colors to indicate greater similarity.

4.2.2 Hierarchical Clustering

Researchers could directly analyze the dataset by using the hierarchical clustering dendrogram. To create it, we repeatedly find the most similar clusters and merge them until only one cluster is left. The order in which we merge the clusters determines the dendrogram's structure. This means that captions that are more similar to each other will be grouped together earlier on the the dendrogram. Fig. 6 shows hierarchical clustering of 4,454 captions, where similar captions are grouped further left on the graph. For instance, the top-right zoom-in view reveals that "Decoration of brick wall," "Decoration of a brick wall," and "Decoration of a brick wall" are combined into one cluster far to the left, indicating their strong similarity.

4.2.3 Force-Directed Graph

We also use the distance matrix to create force-directed graphs to visualize the dataset. We employ the D3 algorithm, which utilizes physics simulation to position nodes and edges to minimize the energy of the system [24]. By simulating an attractive and repulsive force between each pair of connected nodes, the algorithm determines the position of each node. The attractive force is larger if the distance score between any two captions is closer to 0, while the repulsive force is larger if the distance score between any two captions is closer to 1. Because putting all of the edges of a complete graph would be overwhelming, we need to set rules for choosing what edges to use. Fig. 7 shows the case of selecting edges that connects the node's top 5 similar captions based on cosine distance.

Fig. 8 shows the results of filtering out edges that connect nodes with cosine distances larger than 0.25.

4.2.4 Minimum Spanning Tree

We create a visual layout of the 'nearest neighbor' (most similar caption) of every node using a Minimum Spanning Tree (MST), where the distance between captions reflects their relative similarity. An MST is constructed by iteratively linking together the most similar remaining pair of items as long as that link does not create a loop in

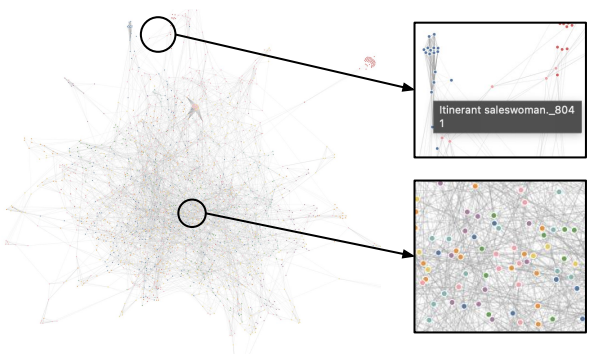


Figure 7: Force-Directed Graph that Connects Each Node with Its Most Similar Neighbors

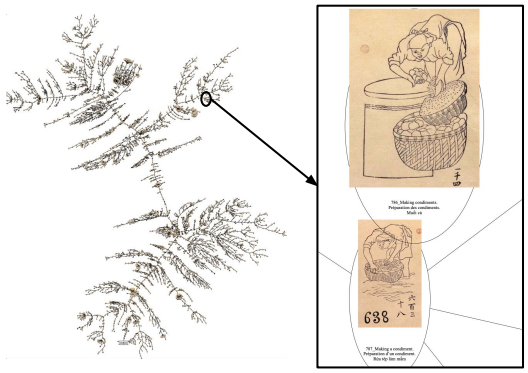


Figure 9: Default Visualization of Minimum Spanning Tree (MST) with Inset Views on "Making condiments" and "Making a condiment"

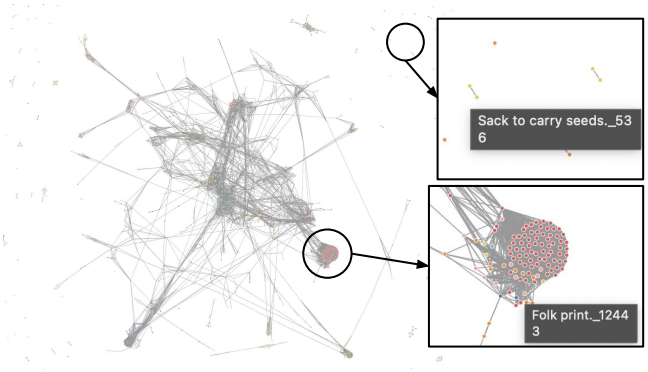


Figure 8: Cleaned Version of the Force-Directed Graph by Setting the Threshold as 0.25

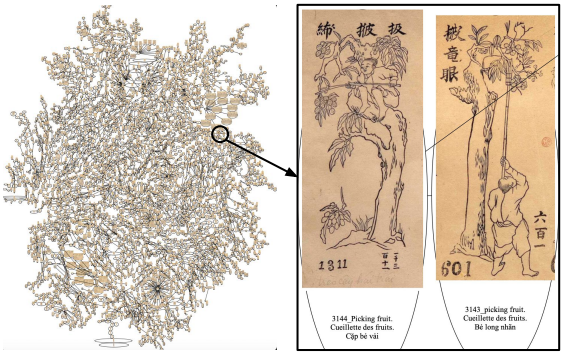


Figure 10: Re-arranged MST with PRISM Algorithm with Inset Views on Two "Picking fruit" Sketches

the resulting set of links. After determining what edges to include, we need to decide how to arrange nodes in the graph. We could either prevent overlap among edges (Fig. 9), which is the default version or prevent overlap among nodes (Fig. 10). The later graph approach used the PRISM algorithm to remove overlaps while maintaining the proximity relations between the nodes by using a stress function [25].

4.2.5 Dimension Reduction

We also explored dimension reduction techniques, including t-SNE (t-distributed Stochastic Neighbor Embedding) and UMAP (Uniform Manifold Approximation and Projection) to lay out our items in 2D. t-SNE works by mapping the input data points to a lower-dimensional space while preserving the structure and relationships between the data points as much as possible. This is done by minimizing the Kullback-Leibler divergence between the joint probabilities of the similarities between data points in the high-dimensional space and the low-dimensional space [26]. We also inputted our manual content coded label of 'gendered by visual' onto the t-SNE output in order to add a level of gendered analysis on top of the topic clusters. Fig. 11 is the t-SNE output of the dataset with the gendered manual labels.

UMAP (Uniform Manifold Approximation and Projection) is a dimensionality reduction technique similar to t-SNE. It aims to preserve the global structure of the data as well as the local neighborhood relationships between the data points. UMAP, on the other hand, uses a different mathematical formulation based on Riemannian geometry and algebraic topology [27]. We generated two UMAP projections. The first one, shown in Figure [refer to fig:umap], utilizes the 768-dimensional caption embeddings. The second visualization employs PixPlot from the Yale DHLab, which is based on the visual embeddings of images. To create this projection, PixPlot first converts the images into embeddings using a convolutional neural network (CNN) and then projects these vectors into a 2D space using UMAP. The resulting PixPlot output is presented in Figure [refer to fig:pixplot].

4.2.6 Radial Spanning Tree

In addition to algorithms that emphasize the overall structure, we also investigate a method tailored for local structures when working with smaller datasets: the radial spanning tree. In this approach, the root node is situated at the center of the tree, while the remaining nodes are arranged on the periphery in multiple layers. The tree's edges extend outwards from the center, creating a radial pattern. Fig. 14 is an example of setting "midwife massaging a woman about to give birth" as the root node with 3 layers using Reingold–Tilford "tidy" algorithm [28].

5 Results and Evaluation

Nguyen is an expert in Southeast Asian cultural history and book history, with a special emphasis on visual materials. She has dedicated over eight years to studying the image dataset. She first provided a preliminary evaluation of the results from different visualization methods with the three tasks in mind: (1) to gain an overview of the dataset, (2) identify patterns that need further investigations, and (3) to contextualize data and identify relational connections. Based on the pre-evaluation, Nguyen selected the three most insightful visualization methods for further scrutiny: distance matrix, minimum spanning tree, and dimension reduction.

Subsequently, a focus group consisting of three humanities researchers, unfamiliar with these visualization techniques, re-evaluated the three top-performing methods to further gauge their efficacy. Through this approach, we aimed to ascertain the effectiveness of our visualization methods, particularly for those new to such ▶R3-3

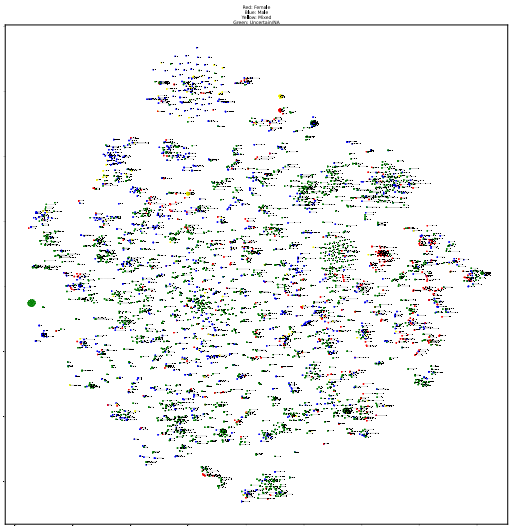


Figure 11: t-SNE Graph with Nodes Colored Based on Gender Labels

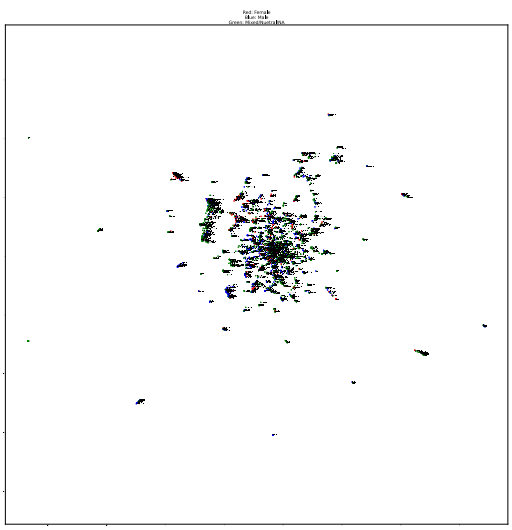


Figure 12: UMAP Graph with Nodes Colored Based on Gender Labels

techniques. The detailed evaluations of Nguyen and this focus group are elaborated upon in the following sections. ▶R4-4

5.1 Distance matrix

A distance matrix, while not the most visually engaging, offers a powerful tool for generating research queries (T2). Its primary strength lies in presenting an accurate understanding of the relationships between data points. The distance matrix can significantly aid in contextualizing words within the entire caption corpus, and its hierarchical clustering provides an overview for broader categorization (T3).

One practical application of the distance matrix is to select a caption and rank all other captions based on their similarity score. For instance, Fig. 15 demonstrates captions ranked by their similarity to "Midwife massaging a woman about to give birth". This ranking could open pathways for investigations into themes like childbirth, societal relations, and medicinal practices. Viewing the distance matrix in tools like Excel allows for precise comparisons among captions based on their numerical significance.

In the focus group evaluation, the Distance Matrix was especially lauded for its utility in achieving T2. Participant 1, scoring it a 3, remarked, “Once you select a point (term, etc.), it’s much easier to see relationships and start to ask questions.” This feedback underscores the matrix’s effectiveness in fostering inquiry and hypothesis generation. ▶R3-3

However, a potential challenge arises when considering the sheer size of the $4,454 \times 4,454$ matrix. Such detailed similarity scores among all data points might be overwhelming for some users, even with strategies like coloring and matrix reordering (T1). Nevertheless, presenting the matrix alongside relevant images and employing multilingual captions can lead to more diversified and nuanced investigations of a topic. ▶R4-4

5.2 Hierarchical Clustering

Hierarchical clustering pushed a rethinking about the type of clustering that we needed to do in the next level of analysis. For example, to cluster by objects, people, or actions, or on a more fine-grained level such as materials like ‘gold’, which then bring in a variety of different industries. Hierarchical clustering thus provided an understanding of the higher-level breakdown of topics (T1) and moved to generate research questions (T2). But it squeezes all data into a single dimension and thus loses many meaningful connections, thus negatively impacting meaningful question generation (T2) and data contextualization (T3). For example, in Fig. 6, although "Itinerant rice-making saleswomen" and "A woman’s role in the construction industry" share similarities in terms of gender, "A woman’s role in the construction industry" is positioned at the very bottom, while "Itinerant rice-making saleswomen" appears near the top of the captions.

5.3 Force-Directed Graph

Force-directed graphs use simulated forces to connect nodes with high similarities, resulting in clusters that pull together the most similar captions, especially on the rim. Each cluster on the perimeter can help researchers generate a research question. (T2) Fig. 16, for example, shows the cluster of "itinerant saleswoman" (left) and "folk prints" (right).

However, the central bulk of the graph, as shown in Fig. 17, can be messy, making it hard to reveal patterns and generate meaningful questions from entangled data points. (T1)

Achieving a balance between maintaining a clean graph and including enough edges to connect all points in Force-directed graphs can be challenging. If we opt for a clean graph, many points may not be connected. Consequently, contextualizing data points becomes difficult if two related nodes are not connected, unlike in MST where every pair of nodes is connected. On the other hand, if we aim to display all potentially useful connections, the graph could become cumbersome, entangled, extend off the screen, and slow to navigate. (T3)

5.4 Minimum Spanning Tree

For the minimum spanning tree, participants can easily identify similar captions by looking at what nodes are nearby and connected. (T3) The number of connections is limited because only the minimum number of edges to connect everything is kept, thus enabling viewers to follow a long range of relations and generating interesting topics in the exploration. It presents the size of the dataset in a visually appealing way (T1), allows data to group into meaningful clusters based on their similarities, and provides explainability suggesting why some

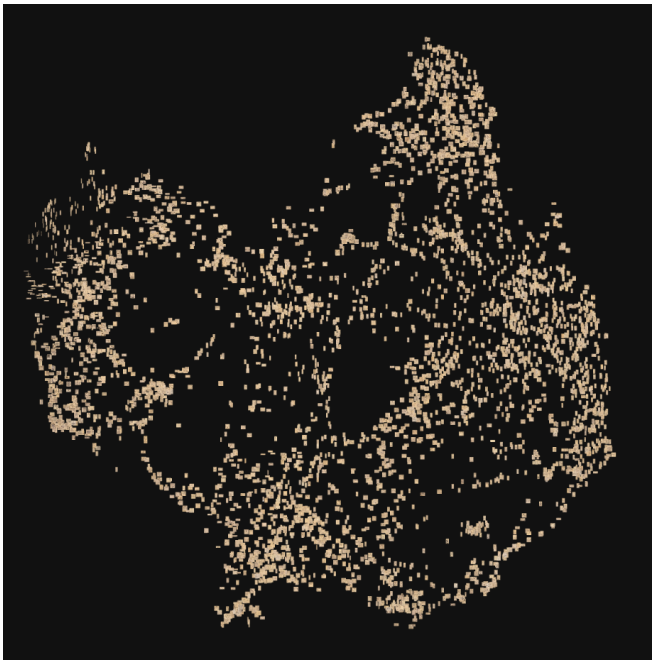


Figure 13: Pixplot Graph

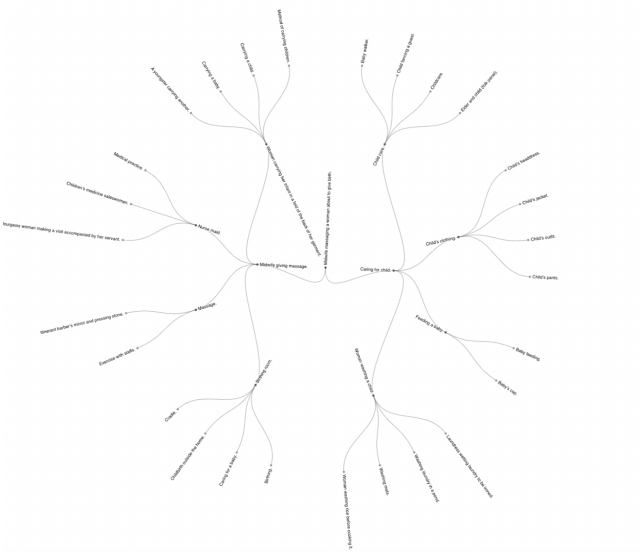


Figure 14: Radial Spanning Tree

captions are grouped together by the gradual transition of meaning along multiple edges. MST excels for T2 by helping researchers identify hidden patterns and relationships within raw data. For example, when examining the MST for the caption "Carrying a baby," it swiftly displays the most similar caption and leads to various branches. By exploring the branch leading to “A youngster carrying another”, researchers can generate questions about children’s relationships. By investigating the branch connected to "Child’s hat," researchers can pose questions about clothing styles in colonial Vietnam. This exploration method assists in generating research questions as users can select a random caption, navigate through similarities, and formulate research questions based on these potential connections. Furthermore, these connections are agnostic to other types of classifications such as person or object. Instead, they can bring together a complex social world of objects, people, and practices that are implicit in similarity of textual meaning.

The focus group shared some mixed reactions to MST. While they acknowledged its potential, especially in T2 and T3 for raising questions and observing relational ties, the expansive size of the dataset posed challenges. Tracing connections became daunting for participants. Both participants 1 and 3 awarded a score of 2 for MST’s capability to offer an overview (T1). Furthermore, while participant 3 found the juxtaposition of images and multilingual captions beneficial for immediate comparisons, participant 1 experienced difficulty processing both images and captions simultaneously. This feedback underscores the need for a more intuitive interface or guided navigation for large datasets. More training time might also impact this evaluation.

5.5 Dimension Reduction

Dimension reduction techniques aim to conserve the cosine distance between data points. Consequently, distinct caption-image pairs are likely to reside further apart. Feedback from the focus group underscored that t-SNE was especially adept at fulfilling T2 and T3.

While dimension reduction preserves distance relationships and aids in forming meaningful clusters (T1), its explainability can sometimes be ambiguous. For instance, experts might readily identify that magic-related captions form a cluster and amulet-related captions congregate together. Yet, deciphering why the caption "Driving" situates between these two clusters might be elusive, as depicted in Fig. 18.

Nonetheless, despite such occasional ambiguities, dimension reduction excels in topic generation (T2). Contrasting the MST, which emphasizes the transition of singular nodes, dimension reduction techniques maintain the relative positions among all nodes. This property births a myriad of intriguing clusters. Such clusters, revolving around themes like Buddhism, the rice industry, or sugar vendors, can potentially become focal points meriting deeper exploration. Additionally, the incorporation of manual labels from ’gendered by visual’ content coding augments another stratum of gender-based topic analysis.

In the realm of data contextualization (T3), tools like Pixplot are invaluable. They not only cluster drawings with subtle visual distinctions, hinting at possible authorships, but also serve critical research queries, like discerning if a book has multiple contributors. Such visual clustering can be pivotal. For instance, groups based on facial expressions or features might suggest varied stylistic representations of humans. Clusters segregating simple sketches from intricate folk scenes could also imply the handiwork of multiple artists in the original drawings’ creation.

Participant 3 of the focus group, notably, awarded t-SNE a high score (3) for its proficiency in T2 and T3. They emphasized its comparative advantage over MST, highlighting its less distracting visual presentation and the innate ease stemming from the spatial proximity of points, as opposed to the web of edges connecting nodes in the MST.

5.6 Radial Spanning Tree

A radial spanning tree is effective in contextualizing data. (T3) For example, Nguyen used the radial spanning tree to understand how depictions of childcare and childbirth are portrayed in a dynamic and socially embedded manner. By using the radial spanning tree, the expert could explore a particular image and see how it is connected to other contexts. Compared with cluster dendrograms, which place all leaves at the same level, radial spanning tree neatly display two directions of contextualization in 3 layers: one towards midwife and massage and the other towards caring for the child. One caption’s surrounding branches also help researchers generate new topics (T2).

However, the radial spanning tree has its limitations. It only allows for a local inspection of one caption at a time and does not enable experts to identify larger patterns or clusters. (T1)

	Midwife massaging a woman about to give birth.
Midwife massaging a woman about to give birth.	0
Midwife giving massage.	0.135
Birthing room.	0.313
Birthing room.	0.313
Woman carrying her infant in a fold of the back of her garment.	0.342
Caring for a baby.	0.358
Childbirth outside the home.	0.363
Nurse maid.	0.44
Birthing.	0.458
Birthing.	0.458

Figure 15: Top 10 Similar Captions for "Midwife massaging a woman about to give birth"

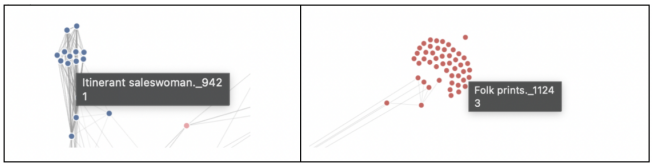


Figure 16: Two Clusters at Force-Directed Graph

6 Discussion

In this section, we comparatively evaluate the visualization methods' effectiveness in addressing our specific tasks (Sec. 6.1). We also discuss some lessons learned from this study and potential avenues for future work (Sec. 6.2).

6.1 Comparative Analysis

To comparatively evaluate the visualization methods' effectiveness in accomplishing the three tasks outlined in the Abstraction section, we asked Cindy Nguyen and the focus group to categorize all visualizations into three: "not effective," "somewhat effective," and "very effective." Tab. 1 shows the resulting matrix from Cindy Nguyen, where we denote "not effective" as 1, "somewhat effective" as 2, and "very effective" as 3.

Our focus group evaluated three visualization methods: Distance Matrix, Minimum Spanning Tree, and Dimension Reduction for tasks T1, T2, and T3. This group consisted of three humanities researchers inexperienced with visualizations and task abstraction. After introducing the tasks and visualizations, participants evaluated each method with scores and narrative explanations. Tab. 2 shows the

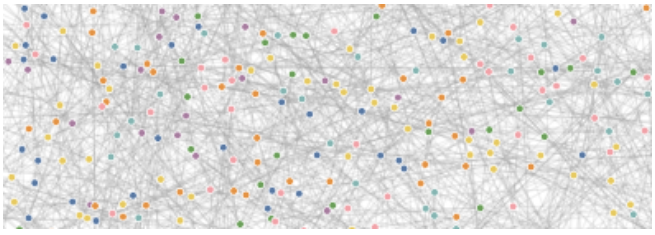


Figure 17: Center of a Force-Directed Graph

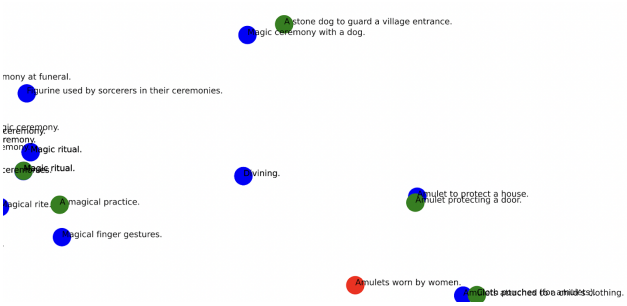


Figure 18: Part of the t-SNE Graph

Table 1: Comparison Matrix according to Expert Evaluator Cindy Nguyen (1 is not effective; 2 is somewhat effective; and 3 is very effective)

	T1 Gain Overview of the Dataset	T2 Generate Research Questions and Hypotheses	T3 Contextualize Data
Distance Matrix	1	3	3
Hierarchical Clustering	2	1	1
Force-Directed Graph	1	2	1
Minimum Spanning Tree	2	3	3
Radial Spanning Tree	1	2	3
Dimension Reduction	2	3	3

average of the ratings from the focus group evaluation. Our summary of results consisted of two of the participants' evaluation, since one of the participants did not directly answer the evaluation questions from the perspective of a researcher, and the participant was not able to complete a narrative explanation of their numerical score. This evaluation reinforces our comparative analysis, providing more granular insights into the methods' effectiveness.

6.2 Lessons Learned

The task abstraction and methods used in this study offer valuable insights for handling complex humanities and historical datasets, like collections of captioned drawings, bibliographies of titles, and historical encyclopedia entries. We discerned that three investigation levels are fitting for this problem with thousands of items, elaborating on Shneiderman's principle of two-level overviews and details on demand for larger datasets.

Table 2: Comparison Matrix according to Focus Group Averaged Score (1 is not effective; 2 is somewhat effective; and 3 is very effective)

	T1 Gain Overview of the Dataset	T2 Generate Research Questions and Hypotheses	T3 Contextualize Data
Distance Matrix	1.0	2.5	2.0
Minimum Spanning Tree	2.0	2.0	2.5
Dimension Reduction	1.5	2.0	2.5

For method usage, if researchers are uncertain about framing research questions, we recommend starting with an MST to identify intermediate connections and understand the corpus's structure. These connections reveal thematic patterns, enabling serendipitous discoveries.

Hierarchical clustering can provide a top-down view of the corpus, especially beneficial when researchers haven't delved deep into the text. The clusters inform optimal categorization and thematic description.

After understanding the corpus structure via MST and hierarchical clustering, focusing on specific entities using the distance matrix and radial layout is advised. This process refines research questions around particular themes, directing researchers to assemble comparative datasets to support their work.

Force-directed graphs, while promising, are not recommended for datasets of our scale or larger. Defining forces that outperform the MST was challenging. Any attempt to go beyond the MST often resulted in a tangled web, possibly due to the sheer number of nodes and potential edges.

The focus group's feedback provided additional lessons. Participant 1 likened the Distance Matrix to how they commence research in archives, emphasizing its potential utility in archives. They also found the t-SNE helpful in relating data points and the MST valuable in providing another layer of information. Meanwhile, Participant 3 emphasized the MST and t-SNE's prowess in data overview and hinted at the Distance Matrix's computational benefits in the research's latter stages. Overall, participants were enthusiastic about distance measures and data visualization, viewing them as methods to restructure connections in datasets and potential alternatives to traditional cataloging systems in libraries and archives.

Regarding future work, an integrated tool merging all visualization methods would be ideal, allowing a seamless transition between techniques based on tasks. Moreover, refining individual methods to boost their versatility for various tasks—like enhancing the MST with more edges or simplifying force-directed graphs for clarity—would be beneficial.

7 Conclusion

In this study, we investigated visualization tools for analyzing a historical dataset of images and captions from a 1909-1910 document on life in Vietnam. Working with a historian, we identified three key tasks for book history visualization: obtaining an overview, formulating research questions and hypotheses, and contextualizing data. We assessed six visualization techniques and discussed their advantages and disadvantages for each task. Based on our findings, we advise researchers uncertain of where to begin to utilize minimum spanning trees (MST) and hierarchical clustering to comprehend corpus structure and discover themes. Once acquainted with the overall structure, they can concentrate on specific entities using the distance matrix and radial layout, enabling them to create focused research questions and integrate supplementary materials to support their research project.

References

[1] Henri Oger. Technique du peuple annamite = mechanics and crafts of the annamese people = ky thuat cua nguoi annam. In Olivier Tessier Dinh Binh Tran Sheppard Ferguson, Philippe Le Failler, editor, EFEO; TVKHTHTPHCM; The gioi. Hanoi, Ho Chi Minh City, 2009. 3.

[2] Ying Zhao and George Karypis. Evaluation of hierarchical clustering algorithms for document datasets. In Proceedings of the Eleventh International Conference on Information and Knowledge Management, CIKM '02, page 515–524, New York, NY, USA, 2002. Association for Computing Machinery.

[3] J. Wang, B. Yu, and L. Gasser. Concept tree based clustering visualization with shaded similarity matrices. In 2002 IEEE International Conference on Data Mining, 2002. Proceedings., pages 697–700, 2002.

[4] Michael Behrisch, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck, and Jean-Daniel Fekete. Matrix reordering methods for table and network visualization. Comput. Graph. Forum, 35(3):693–716, June 2016.

[5] Nathan Gale, William C Halperin, and C Michael Costanzo. Unclassed matrix shading and optimal ordering in hierarchical cluster analysis. J. Classif., 1(1):75–92, December 1984.

[6] Alfred Inselberg and Tova Avidan. Classification and visualization for high-dimensional data. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '00, page 370–374, New York, NY, USA, 2000. Association for Computing Machinery.

[7] Alexander Platzer. Visualization of snps with t-sne. PLOS ONE, 8(2):1–6, 02 2013.

[8] Daniel Probst and Jean-Louis Reymond. Visualization of very large high-dimensional data sets as minimum spanning trees. Journal of Cheminformatics, 12(1):12, Feb 2020.

[9] Pierre Guy Atangana Njock, Shui-Long Shen, Annan Zhou, and Hai-Min Lyu. Evaluation of soil liquefaction using ai technology incorporating a coupled enn / t-sne model. Soil Dynamics and Earthquake Engineering, 130:105988, 2020.

[10] Etienne Becht, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. Evaluation of umap as an alternative to t-sne for single-cell data. bioRxiv, 2018. ▶R3-3

[11] Dominik Bönisch. The curator's machine: Clustering of museum collection data through annotation of hidden connection patterns between artworks. International Journal for Digital Art History, (5):5.20–5.35, May 2021. ▶R4-4

[12] Giovanna Castellano and Gennaro Vessio. A deep learning approach to clustering visual arts. 2021.

[13] Qiaoling Zeng, Mingu Lee, and Juhyun Eune. Digital design method of cultural heritage using ancient egyptian theological totem. Heliyon, 9(5):e15960, May 2023.

[14] Tamara Munzner. A nested model for visualization design and validation. IEEE Transactions on Visualization and Computer Graphics, 15(6):921–928, 2009.

[15] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In Proceedings 1996 IEEE Symposium on Visual Languages, pages 336–343, 1996.

[16] Sean McKenna, Dominika Mazur, James Agutter, and Miriah Meyer. Design activity framework for visualization design. IEEE Transactions on Visualization and Computer Graphics (InfoVis '14), 20(12):2191–2200, 2014.

[17] Uzma Haque Syeda, Prasanth Murali, Lisa Roe, Becca Berkey, and Michelle A. Borkin. Design study "lite" methodology: Expediting design studies and enabling the synergy of visualization pedagogy and social good. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[18] Kostiantyn Kucher and Andreas Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In 2015 IEEE Pacific Visualization Symposium (PacificVis), pages 117–121, 2015.

[19] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. IEEE Transactions on Visualization and Computer Graphics, 18(12):2431–2440, 2012.

[20] Giovanna Castellano and Gennaro Vessio. Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview. Neural Computing and Applications, 33(19):12263–12282, Oct 2021.

[21] Adam James Bradley, Mennatallah El-Assady, Katharine Coles, Eric Alexander, Min Chen, Christopher Collins, Stefan Jänicke, and David Joseph Wrisley. Visualization and the digital humanities:. IEEE Computer Graphics and Applications, 38(6):26–38, 2018.

[22] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. Visual text analysis in digital humanities. Computer Graphics Forum, 36(6):226–250, 2017.

[23] Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. Visualization of cultural heritage collection data: State of the art and future challenges. IEEE Transactions on Visualization and Computer Graphics, 25(6):2311–2330, 2019.

[24] Mike Bostock. Force-Directed graph. `https://observablehq.com/`

`@d3/force-directed-graph`, February 2023. Accessed: 2023-3-11.

[25] Hendrik Strobelt, Marc Spicker, Andreas Stoffel, Daniel Keim, and Oliver Deussen. Rolled-out wordles: A heuristic method for overlap removal of 2d data representatives. In Computer Graphics Forum, volume 31, pages 1135–1144. Wiley Online Library, 2012.

[26] Martin Wattenberg, Fernanda Viégas, and Ian Johnson. How to use t-sne effectively. Distill, 1(10):e2, 2016.

[27] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

[28] E.M. Reingold and J.S. Tilford. Tidier drawings of trees. IEEE Transactions on Software Engineering, SE-7(2):223–228, 1981.