FAN-FICTION AND AO3 FREE-FORM TAGGING PRACTICE: INNOVATING OPEN-SOURCE TOOLS FOR TAG NETWORK VISUALIZATION AND ANALYSIS

ELIAS BESTARD LORIGADOS

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTERS OF SCIENCE

GRADUATE PROGRAM IN FACULTY OF ELECTRICAL ENGINEERING &
COMPUTER SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

2022

Abstract

The thesis focuses on obtaining, representing, analyzing, and visualizing tag data within the Archive of Our Own (AO3) fan fiction archive, with a focus on the "Additional Tags" category. The work was undertaken in four components. First, network nomenclature was defined and formally-defined network representations were developed for the AO3 Additional Tag structure. Second, computational techniques were developed to scrape the data from AO3, both in its current state and in its archival state (from the Wayback Machine) and to store it in a format suitable for subsequent processing. Third, techniques were developed to use the scraped data in order to computationally instantiate and visualize the network structures that were developed in the first component. Finally, in the fourth component, a set of network analysis techniques were developed, which were used to extract quantitative characteristics of the networks, to identify relevant changes over time, and to identify co-creation actions between different users in the AO3 community.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Tagging and Tagging Practices

The use of tags became popular among websites associated with Web 2.0. Currently, tags are used in multiple ways, typically focusing on the use of keywords in the classification, identification, marking, and/or flagging of many different items of user-generated content, such as images, websites, videos, posts, and other media components (Lee and Chun, 2007). Tagging practices are used by many different platforms and communities on the Internet. We can find them used in social networks, online communities, blog sites, and research papers.

There are many different tagging systems, most of which currently use one or a combination of tagging practices. Tagging practices can be divided into two styles: the use of predefined tags or the use of free-form tags. Pre-defined tags are those tags drawn from a controlled language (i.e., a preset vocabulary), as determined by the tagging system. Pre-defined tags may also be referred to as 'canonical' tags or as 'meta-tags' in different communities. Free-form tags are tags that are created by users and can be (almost) anything the creator decides to use. The use of free-form tags is one of the ways in which a folksonomy[1] can emerge.

Each tagging approach has its pros and cons with respect to different criteria. For pre-defined tags, one con is that the approach is strict, as it obligates users to select a tag from among predetermined options. This may present a challenge for users when none of the predefined tags are aligned with their preferences. But, in return, the approach facilitates the visibility of the tagged items through simplified searching and filtering. The free-form

[1] A *folksonomy* is a user-generated system of classifying and organizing content by metadata, such as tags.

tagging approach, on the other hand, is more flexible than pre-defined tags. Free tagging gives freedom to the users to define their tags. This privilege, however, might affect the visibility of the tagged items. The search engines need to be more complex to make visible the tagged items and to make visible other similar items.

While surfing on the WWW, we can find countless communities within different platforms organized around shared common interests. Fan-fiction is one popular interest all over the world. Fan-fiction refers to the practice where fans (not professionals) create new plots based on popular works while freely changing or using their concepts to create new content (Flaherty, 2022). Fan-fiction communities are a growing section of the WWW; the communities allow their members to create and share fictional work freely. It has been shown that there has been and continues to be massive growth in these communities, both in the number of stories and in authors per year (De Kosnik, 2016). Each fan-fiction community platform has its structure and functionality, but no matter the community, each platform needs a system of organization for the fan-works and needs to provide a search engine that allows members to find fan-works easily. Tagging is a key part of the organization and structure of any platform, which provides different identifiers for each fan-work. Every fan-fiction community implements a tagging system, whether it is its own customized system or it re-uses an already-implemented system. Having a system for tagging their works makes the fan-works more organized and accessible.

1.2 AO3 and AO3's Tagging Structure

Archive of Our Own (AO3) is a non-profit open-source repository for fan-fiction and other fan-works (Archive of Our Own, n.d.[c]). By 2017 it was one of the biggest multi-fandom archives containing over 3 million fan-works uploaded (Duchastel de Montrouge and Baljko, 2017). A year later, they reported to have reached 4 million fan works (Archive of Our Own, n.d.[a]), and currently (dated to Jun 2022) contains over 9 million fan-works (Archive of Our Own, n.d.[c]).

Since the time of its early development, AO3 sought to establish a solid tagging structure in order to improve the accessibility of the fan-works. This archive employs a unique set of tagging practices and allows multiple forms of tagging (Duchastel de Montrouge and Baljko, 2017). The AO3 tagging system provisions for contributors to tag their fan-works under several different categories, which are provided in table 1.1.

AO3 Tag Categories

- The fandom(s) to which the fan-work belongs
- The Character(s) featured in the fan-work
- The Relationship(s) within the fan-work
- Language of the fan-work
- Warning(s) to alert the readers of certain topics in the content
- Rating to measure the intensity of a work's content (General Audience, Teen And Up Audiences, Mature, Explicit, or The work was not given any rating)
- Category for the type of relationships the fan-work contains
- **Additional Tags** are additional free tags the creator wants to add

Table 1.1: Categories of tags within AO3 (Archive of Our Own, n.d.[i]). A fan-work contributor can add one or more free tags under each category. The "Additional Tags" category (emphasis added) is the focus of this research project.

Each of the categories in table 1.1 serves its own function on the AO3 site. The categories are used during the deposit process (e.g., when a fan-work is added to the archive) and create structure in the archive. The "Additional Tags" category is the focus of this research project. This tagging category, in particular, illustrates demonstrating the folksonomy of the AO3 community, which is achieved, in part, through an archive wrangling process. Fan-works and tags, as well as the resultant tag archive structure, are all considered cultural products and have been the subject of ongoing investigation and analysis by several different groups of scholars (De Kosnik, 2016). Information about tags is of interest to scholars in fields such as Media Studies, Communication Studies and Science and Technology Studies. Scholars in this field analyze and investigate the behaviour of the community and the tag wranglers' and contributors' practical politics around the classification of contested concepts, such as race, gender roles, heteronormativity, homophobia, and ableism.

There are three aspects of the "Additional Tags" category, each described below, that combine to create the conditions for interesting community-level phenomena: site visitor expertise (section 1.2.1), contributors' tagging strategy (section 1.2.2), and wranglers' behaviour (section 1.2.3).

1.2.1 AO3's Tagging System for Site Visitors

1.2.1.1 Searching for Fan-works and Automatic Tag Completion

The "Additional Tags" category is a common way for site visitors to find works. When they use the `Advanced Search→Works` interface and begin to type terms, the AO3 site will provide automatic tag completion. This is shown in Figure 1.1 for the example of searching for the tag *Disability*.

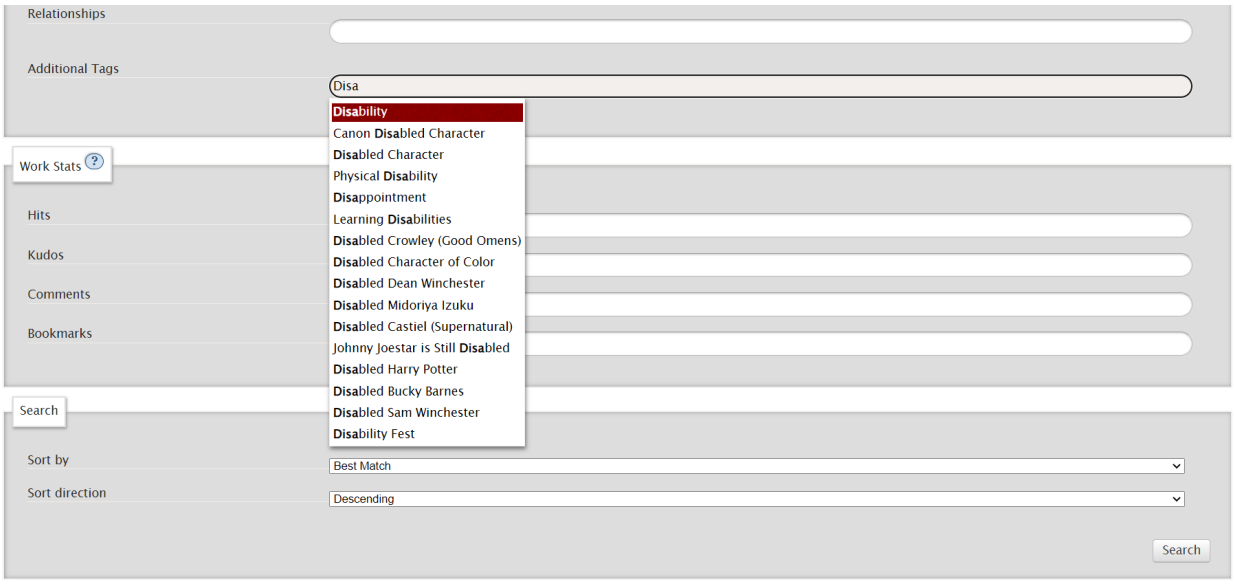


Figure 1.1: Automatic Tag Completion of AO3. `Advanced Search→Works` while searching works with the Additional Tag *Disability*. Free-tags do not appear among the possible completions.

In AO3, however, automatic tag completion is provided only for *some* of the tags that have been previously used in the category. We will refer to these as "recognized" tags for the time being. (This is a simplification since there is further nuance to this, which will be explained in Section 2.1.) In the AO3 systems, there are previously-used tags in the "Additional Tags" category, but they will not get formally "recognized" and thus will not appear as part of the automatic tag completion. Only certain users know about this "recognized" (vs not "recognized") tag characteristic of the AO3 search interface. Fan-fic contributors know that only certain users know about this distinction as well, and design their tags accordingly. For instance, some contributors may deliberately use free-tags and avoid the use of recognized tags, since they wish to resist the association with other recognized tags. Thus, this distinction has a political aspect. The political aspects of this formal "recognition" will also be explained

further in section 1.3.

1.2.1.2 Viewing Fan-works and Tag Display

Fan-work can be viewed either in a summary format (in the `Works Listing` view) or in a longer format (in the `Individual Work` view). In either case, when a user views a particular fan-work, all of the "Additional Tags" are shown alongside the fan-work's entry and include both "recognized" tags and free-tags without distinction. Figure 1.2 provides an example of a fan-work in summary format (in the `Works Listing` view), with highlighted annotations provided to illustrate the different types of tags in the "Additional Tags" category.

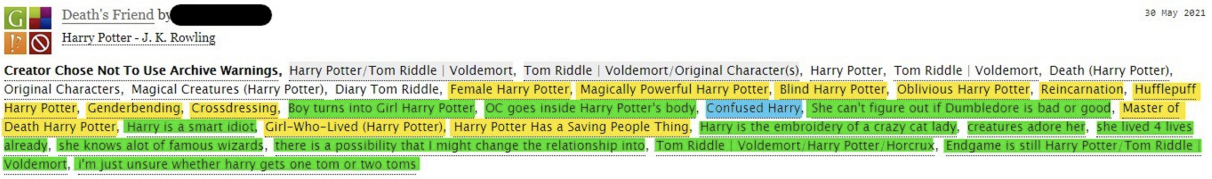


Figure 1.2: Summary of the "Death's Friend" fan-work (an excerpt from the `Works Listing` view). The fan-work was created on May 30th, 2021, was contributed by *[anonymized contributor]*, and appears in under the "Harry Potter- J.K. Rowling" fandom. These attributes were as of the capture date of October 10th, 2021 (but may have changed since then). The "Additional Tags" are listed below the fan-work title and fandom and include both "recognized" tags and free-tags without distinction. Highlighted has been used here to indicate whether the tags are "recognized" (yellow or blue) or are free-tags (green).

1.2.1.3 Browsing Fan-works by Tags

When the user interacts with the AO3 interface for browsing the fan-works, they may choose the `Browse→Tags` or the `Browse→Works` views. In the `Browse→Tags` view, only the "recognized" tags are listed. Free-tags are not included.

1.2.2 Contributors and Tags

The "Additional Tags" category in AO3 allows a fanfic contributor, when adding a new work to the archive, to make use of several options: to re-use a tag that has already been "recognized" and/or to use one or more tags of their creation (which is called "free tagging"). It is theoretically possible for multiple contributors to independently create the same new

free-tags and be unaware of each other's free-tags. At this point of maturity in the AO3 platform, free-tags are often long and idiosyncratic, so the likelihood of concurrent creation is not high but it does occur. For free tags, AO3 allows a contributor to use a freely chosen text string as a tag. Multiple free tags can be used. A free tag can be any text string with up to 100 characters, including spaces, special characters, and punctuation signs (except the comma, which is considered to be a tag separator). Special characters can include characters from most languages other than English, numbers, and emojis. For instance, both "(Trixie is a House-Elf)" and "Harry Potter Epilogue What Epilogue | EWE" are valid tags used in the "Harry Potter- J.K. Rowling" fandom.

As mentioned above, many fan-fic contributors have knowledge that tags can get formally "recognized" by the AO3 site (and thus become candidates for tag completion, become included in the listings in the browse views, and generally become more visible). Fan-fic contributors have knowledge that tags, as part of being formally "recognized", become associated and connected with other tags. Some fan-fic contributors have issues with how certain tags are employed by other contributors to identify their works. For instance, some contributors may deliberately avoid the use of recognized tags, since they wish to resist the association with other recognized tags. Contributors design their tags accordingly, both in favour of using recognized tags and in favour of unrecognized tags.

1.2.3 Tag Wranglers and Tags

Free-form tagging inevitably leads to a proliferation of tags. Although contributors may use "recognized" tags, it is not necessarily the case that they will do so (and indeed may deliberately avoid this in favour of other free tags). AO3 created a system for dealing with the proliferation of free-form tags under the "Additional Tags" category, using an approach called "Tag Wrangling". To implement its tagging system, AO3 has introduced different social roles for members of the community: wranglers, contributors, and consumers. Members can take on different roles while interacting in this community. It is the wrangler that converts the "unrecognized" tags into "recognized" ones, and who may place them in parent or child relationships to other "recognized" tags. The wrangling process is explained in more detail in section 2.1.1.

1.3 The Politics of AO3 Tagging Choice

There are several motivations at play in the design and choice of tags under the "Additional Tag" category by contributors. One aspect of the choice is what the contributor wants to flag as important and, more generally, the signals the contributor wants to employ to the community of potential readers. Creators often want to promote the visibility of their stories and thus choose tags accordingly (e.g., tags that are anticipated to be known and used by readers when searching the archive). Many contributors are motivated by visibility. Steps taken to increase fan-fic visibility can lead to different types of outcomes, such as increasing a contributor's status, and is thus a kind of political action.

On the other hand, contributors may have motivations other than visibility. For instance, creators may instead use tags that, while not widely known or unlikely to be used in searches, serve other purposes, such as signalling the creator's ideology or highlighting an important detail in the story. An example from "Harry Potter- J.K. Rowling" fandom of this is: "more of a what-would've-happened-if-snape-didn't-just-mope-and-be-mean-after-lily's-death-study", which is highlighting some thoughts the author saw as important but unlikely to appear in searches from common readers. If enough fan-fic contributors make use of a tag that highlights a particular detail such as this, then the tag wrangling process will pick this up and the free-form tag will become "recognized". This is another kind of political action.

Another example of politics while creating a fan-work and tagging it can be seen in what Messina (2021) defined as "Critical Fans" and their work. They defined critical fans as "fans who challenge harmful systems of power — such as racism, misogyny, heteronormativity, homophobia, and ableism — in their everyday fan engagements. These fan engagements include writing/reading fanfiction, posting on social media ..." (Messina, 2021). It is from this perspective that provides the basis for the observations noted earlier in this section: that Fan-fic contributors see tags, and their systems of recognition and visualizing, as having power, and that power can and should be challenged. Critical fans have issue with certain tags, how they are employed and the content of the fan-works that employ them. For instance, some contributors may deliberately avoid the use of recognized tags, since they wish to resist the association with them. This can lead to a seemingly counterintuitive phenomenon. If, in the course of processing one of these ideological tags, a wrangler recognizes it and associates it with another tag that is felt to be problematic, then the contributor community may move *away* from the use of that tag. How contributors design their tags is a component of their fan engagement.

The proliferation of tags in the content of critical fans (and in the content of fans who are less critical or not critical), and the actions of the wranglers in processing these tags, are reflective of actions of power in terms of contested concepts in the archive. The proliferation of them might affect the interaction of other users with respect to the usage of the tags.

1.4 Tag Information as a Data Source

While there is a body of prior work looking at fan-fiction communities, very little of it has focused on capturing the tag information automatically, representing it structurally, and presenting it in the form of information visualizations. For example, a basic tree structure has been used by Duchastel de Montrouge and Baljko (2017), as shown in figure 1.3. This visualization was used to show the "recognized" tag Disability as the root of a tree-like structure, with its "child" tags shown as well.

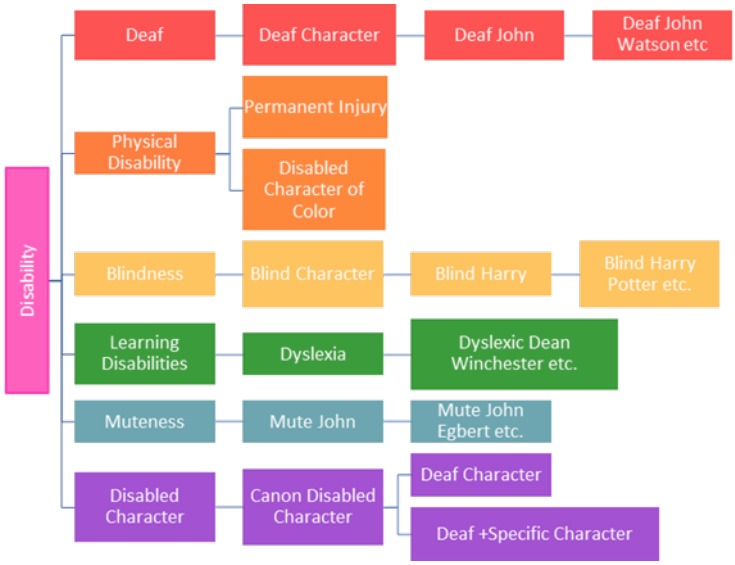


Figure 1.3: The "recognized" tag *Disability* is shown in relationship to several sub-tags, as published in (Duchastel de Montrouge and Baljko, 2019)

In addition to the tree structure visualization, word cloud visualizations have also been employed. These visualizations center a "recognized" tag and other related tags are shown, where the font size of the tag is proportional to the prevalence of use of the tag. For example, figure 1.4 shows a word cloud visualization of the tags that the wranglers have connected with the "recognized" tag Disability. The wranglers have deemed that all of these tags represent

the same concept the tag *Disability* and have linked other tags to it. It is common for the wranglers to find misspellings, as well as different capitalizations of certain words or phrases among tags (e.g, this can be seen in Figure 1.4 with "Physical" and "physical"). To mitigate the impact of this type of variation in tags, the wranglers will deem they all to "represent" the same tag.

Figure 1.4: Word cloud for the tag *Disability* from the AO3 tag structure data, as published in (Duchastel de Montrouge and Baljko, 2019).

1.5 Research Questions

In the current scholarship, the tag information in AO3 has not yet been formalized into network structures. It is highly intuitive to model the tags as nodes and the node relationships as edges, but to do so requires addressing several challenges. In this research project, we set out to tackle these challenges.

An overarching goal is to develop knowledge outcomes and knowledge products that can be taken up in an interdisciplinary manner, meaning that the results should be accessible and of utility to researchers outside of computer science. In particular, the results should be available to those who study fan-works, tags, and the resultant tag archive structures as cultural products. Thus, the tools developed in this project should be openly available and usable after the completion of this research project.

The starting point of this project is that there is information to be gained from the AO3 tags — and the "Additional Tags" in particular — and this information can be represented as a network. Thus:

Q1: (a) Which technique(s) should be used to build a network from the AO3 tagging information (either as captured "live" from the site and/or from archival information)? Which structural properties can be visualized and extracted from this network representation?

(b) How can the actions that users, contributors, and wranglers perform in AO3 be also expressed in terms of operations on the network representation?

(c) What do the structural properties of the network representation tell us about the AO3 'Additional Tags' category?

We understand that the AO3 tags are not static: they change over time, by the actions of both the users and the tag wranglers. There is a day-by-day or even hour-by-hour dynamism. It may be too ambitious to capture the full dynamics in this project, so we might wish to simplify the task to capture 'information snapshots' at specific points in time. The difference between two snapshots will capture the changes that have happened to the tag information between two points in time. Thus:

Q2: (a) How can we capture 'information snapshots' at one or more points in time?

(b) Given two information snapshots, which technique(s) could be used to compare them with one another, in terms of change over time?

(c) What characterizes the major changes to the AO3 tag information over the past years of activity?

We know that the behaviour of the contributors has an impact on the tasks that the wranglers need to perform. We also know that the wranglers' actions have an impact on the users of AO3 (both readers and contributors). This means that the tag information is co-created by the wranglers and the users via bilateral paths of influence: through the tag information over time, the wranglers impact users and then the users impact wranglers (and vice-versa). We conjecture that there are multiple bilateral paths of influence. Thus:

Q3: (a) Which actions by fan-fic contributors impact the wranglers (and, conversely, which actions by wranglers impact fan-fic contributors) and how can these actions be illustrated through operations on the network?

(b) Which technique(s) could be used to detect bilateral paths of influence that may have occurred over the past years of AO3 activity?

1.6 Overview of Thesis

This thesis is organized as follows:

- Chapter 2 sets the ground for the basic definitions covered in the research. It provides a summary of a literature review of related work around the AO3 tagging system, its structure, how to represent it and how to mine quantitative information from it. With the benefit of additional concepts and nomenclature, the research questions are restated to make use of specific formal definitions and specific network structures.
- Chapter 3 describes the methodology that was followed in order to respond to the research questions. Details of the design and implementation of the necessary software tools are provided.
- Chapter 4 presents the answers to the research question. The data used in the answers were derived using the software tools, so this chapter also shows how to the tools were used to get the desired results.
- Chapter 5 provides a conclusion to the research project and describes potential future work.

Chapter 2

Background and Related Work

This chapter sets out the foundation and background of the project. Section 2.1 explains AO3 tagging system, focusing mostly on the tag wrangling process and the AO3 tagging structure. Through these explanations, the section establishes core nomenclature used in the subsequent chapters. Additionally, section 2.2 provides a summary of work related to this project from the research literature, including other work related to AO3 and its tagging system and structure, work related to network analyses of different online communities' tagging systems, work related to tools and techniques to obtain quantitative results and visualizations of tag structures and findings. Finally with the core nomenclature and concepts established, section 2.3 provides a refined version of the research questions first articulated in section 1.5. The questions can be re-articulated more precisely once they can make use of specific nomenclature and established concepts.

2.1 The AO3 Tagging System

As discussed previously in section 1.2.2, when a contributor deposits a fan-work into the AO3 archive, they have the option to use any tags they wish under the "Additional Tags" category: to re-use a tag that has already been "recognized" or to instead use their own tag, which is called "free tagging". The tagging system, with all its structure and process of composition, will now be explained in more detail.

From the non-wrangler point of view, this tagging system can be unclear and can appear to be a black box. But this system allows contributors to create their tags and/or to use existing ones, and to do so without concern about tagging proliferation. The wrangling system makes use of a continuous process that is changing the archive tagging system structure on

an on-going process over time. As well, wranglers or moderators may refine the tag wrangling system itself on an ongoing basis.

2.1.1 Tag Wrangling Guidelines

AO3 has two main user roles that involve direct interaction with the tag system, contributors (creators or authors), and wranglers. Contributors are creating, updating and removing fan-works, which they tag according to their will. On the AO3 platform, the tags entered by contributors under the "Additional Tags" category are periodically reviewed by tag wranglers. Tag wrangling is performed by volunteers ("wranglers") following different policies described in the community to organize the freedom of tagging as a unique technique (Duchastel de Montrouge and Baljko, 2019).

Tag wranglers, when finding free-form tags, can potentially perform one of several different operations. They assess each free-form tag based on the criteria in table 2.1.

Basic wrangling guidelines
(i) is this free-form tag is likely to be used by other contributors?
(ii) has this free-form tag been used in at least three different works?
(iii) has this free-form tag been assigned by three different users?

Table 2.1: Base rules of the tag wrangling guidelines to recognize a free-form tag (Archive of Our Own, n.d.[f])

If a free-form tag meets these criteria, the wrangler can change the "status" of the tag. In section 1.2, this was described as "recognizing" the tag, but it is actually more nuanced than this. Tag wranglers may choose either to "canonize" the tag or to "synonymize" it (or to "syn" it, for short).

2.1.2 Illustration of a Canonical Tag

Before discussing to actions of "canonizing" and "synonymizing" tags, we will introduce some nomenclature for the sake of clarity.

The AO3 site provides a `Search`→`Tags` interface. There are several options within the interface, as shown in Figure 2.1.

If the user provides a tag name and then triggers the search, AO3 will display a list of candidate matches. If the user further specifies the "canonical" option (shown in Figure X as a radio button), only the wrangled tags will be returned as results. Suppose the user enters

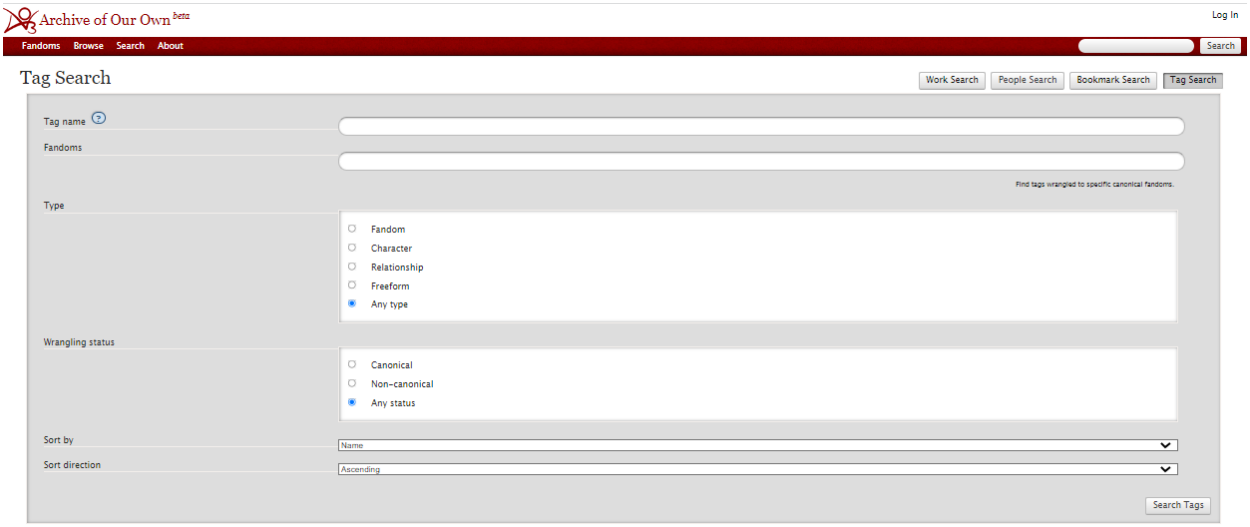


Figure 2.1: `Search`→`Tags` Interface with options to filter the search given a type, and wrangling status (Archive of Our Own, n.d.[d])

"Deaf" as the tag and selects the "canonical" option. AO3 will return a large set of results, all of which contain the term 'Deaf'. From the list of results, suppose the user selects the exact-match entry 'Deaf'. Then AO3 will display a hierarchical structure of tags with 'Deaf' at the root, such as what is shown in Figure 2.2.

This example in Figure 2.2 shows, for the recognized tag Deaf, several sections. One section is for "Parent tags (more general)", which is for displaying the list of categories or groups the tag belongs (mostly a Fandom). Another section is for "Tags with the same meaning:", which contains a list of tags where categorized to carry the same concept as the current tag (this is described further in section 2.1.4.2). Another section is for "Metatags", which contains a hierarchical structure of the predecessor (parent) tags of the current tag (multiple parents are permitted). Last, and most importantly, is the section for "Subtags", which contains a hierarchically-organized set of successor (child) tags of the current tag.

AO3 Data for the Deaf Canonical Tag

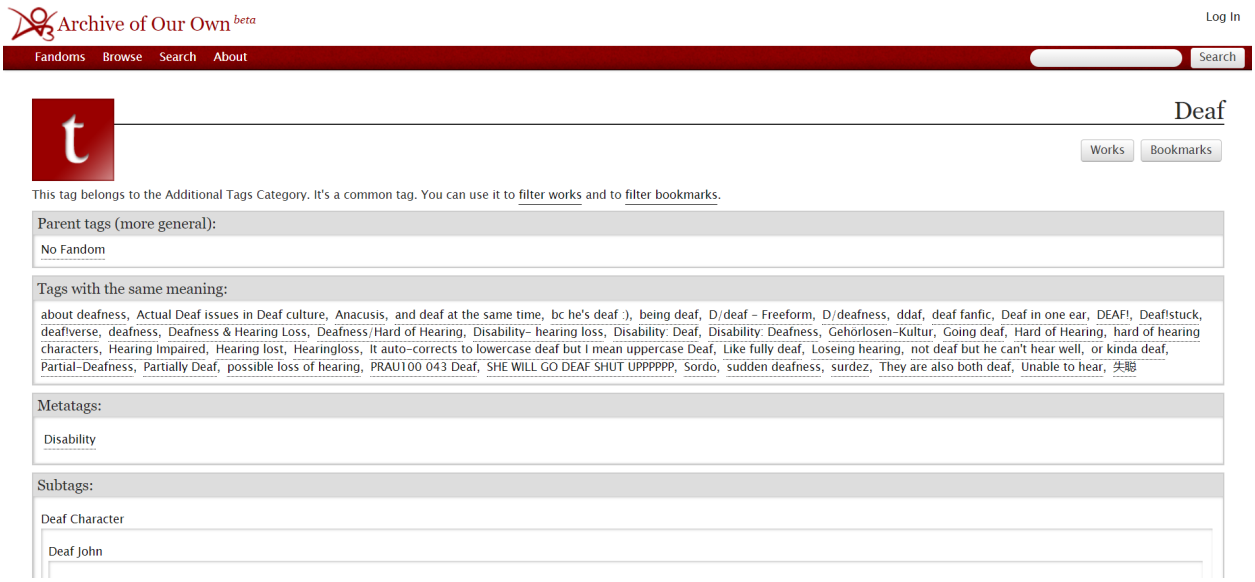


Figure 2.2: Example of accessing to a tag's interface through `Search`→`Tags`→Deaf as a canonical tag and access to it. In this example, the tag 'Deaf' is illustrated. The example illustrates Deaf as a recognized tag with only one predecessor Disability and a section of its hierarchical subtag structure composed by a directed successor "Deaf Character" and a second level successor "Deaf John".

2.1.3 Conjectured Network Structure: The Additional Tag Archive Structure (ATAS)

The AO3 hierarchical display of tags in "Subtags" section that is shown in Figure 2.2 indicates an underlying network structure. It is clear that AO3 makes use of some underlying structure to organize "recognized" tags under the "Additional Tags" category, but the details are not published. We envision the underlying network structure of AO3 as being composed of nodes (also sometimes referred to as vertices) and edges (also also sometimes referred to as links). We cannot access the AO3 structure directly, but we are provided with certain information that is displayed on the AO3 site via the search and browse functions (and can be extracted from the html of the displayed pages). Since the network has not been formally defined previously, we must define it to match the inferred network structure of the AO3 hierarchical display of tags.

We will build a network structure — which we will term the Additional Tag Archive Structure (or "ATA structure" or "ATAS", for short) — to correspond to AO3's underlying network structure. Since the details are not published, nor can we access the structure

directly, the form and structure of ATAS is a conjecture. However, the form and structure of ATAS can be validated through rationale. And we will populate and analyse the ATAS as a directed and undirected network.

We envision the ATAS as having no nodes initially, just as the AO3 structure must have been initially empty.

It is also important to clarify that AO3 also has tag archive structures for tag categories other than the "Additional Tags" category, such as the structure for the "Character" and "Relationship" categories, but these will not be modelled in this thesis project. The construction and ongoing evolution of the ATAS for "Additional Tags" will be the sole focus for the remainder of this background discussion.

In our model, nodes and edges get added to the ATAS as per the wrangler actions. As described in section 2.1.1, as per the AO3 documentation, when a wrangler determines that a free-form tag meets the criteria for formal "recognition", they may choose either to synonymize (to "syn") the tag or to canonize the tag. We now define these operations in terms of operations on the ATAS.

2.1.4 Synonymizing and Canonizing Tags

2.1.4.1 Canonizing a Tag, Part I

When a wrangler chooses to "canonize" a tag, this means that the wrangler has deemed that the tag represents a significant new concept that is not already represented by another canonical tag. Canonizing is used when a free-form tag captures something new in the community and has become sufficiently popular in the archive. Thus, for a free-form tag that meets the criteria for recognition, the wrangler will choose to canonize it if they feel that its concept not already represented in AO3's underlying network structure.

For the ATAS, we define canonizing to correspond to the operation of creating a new node in ATAS.

Because of the complexity of the AO3 hierarchical display of tags, we will build the ATAS structure iteratively. In this first iteration, we tackle the simple case of canonizing a tag for the ATAS. The nodes and edge types are listed in table 2.2.

2.1.4.2 Synonymizing a Tag

Because AO3 has been in used for many years, there is already a well-established network of canonical tags under the "Additional Tags" category. These canonical tags are used to

ATAS Node and Edge Types, v.1	
Node type(s):	*canonical node*
Edge type(s):	

Table 2.2: ATAS nodes and edges types (version 1).

indicate concepts connected to "things like genre of the work, fandom concepts contained in the work, the community the work was created for, when the work takes place in relation to the source material, or fandom-specific locations or concepts" (Archive of Our Own, n.d.[g]). It is very common for contributors to create new tags that are highly similar to canonical tags. "Generally speaking, the most unambiguous, accessible form of a tag is made canonical and all other tags referring to the same concept are linked to it as synonymous tags" (Archive of Our Own, n.d.[g]). Thus, for a free-form tag that meets the criteria for recognition as per table 2.1, the wrangler will choose to "syn" it if they feel that its concept is already represented in the AO3 tag structure as a canonical tag.

Thus, we need a way to represent the synonomous tags in the ATAS. To do so, we define each canonical node to have associated with it a non-empty set of synonyms (aka a 'syn-set'). Thus, the syn-set is represented by a different type of node than canonical tags (aka a 'syn node'). The association between a canonical node and its syn-set node is represented by an ongoing/incoming edge (or undirected edge) representing the *syn relationship*. We need to instantiate a rule that a canonical node can have only one associated syn-set node. We need to instantiate an additional rule that all syn-sets have a 'representative' tag, which is the 'canonical'. For convenience, we will assume that the syn-set excludes the canonical tag itself, and that the canonical tag needs be taken in conjuction with its syn-set. Thus, in tag wrangling processes, when a free-form tag is synonymized to a canonical tag, this is equivalent to the tag being added to the canonical node's syn-set node.

Table 2.3 shows the types of nodes and type of edges defined at this point for the ATAS.

ATAS Node and Edge Types, v.2	
Node types:	*canonical node*, *syn-set node*
Edge types:	*syn relationship*

Table 2.3: ATAS nodes and edges types (version 2).

2.1.4.3 Canonizing a Tag, Part II

We previously defined the wrangler action of canonizing a tag to correspond to the operation of creating a new canonical node in ATAS (section 2.1.4.1). We also previously defined Synonymizing to correspond to the operation of adding to the syn-syn (section 2.1.4.2). Thus, the creation of a new canonical node entails the creation of a new syn-set node (of initial size 1, to contain the canonical tag itself), since future synonymizing operations will require a place to hold the synonomized tags.

As well, the wrangler must also choose a position for the new node in the ATAS, to reflect the desired parent-child relationships. Canonical tags get positioned into a parent-child relationship in the network by the wrangler. For this, the wranglers are asked "check to see if there's already a canonical tag for a different form of the same word, or some other tag that the new one should be a synonym of (or possible a subtag)" (Archive of Our Own, n.d.[f]). These edges represent *hypernym–hyponym relationship*. In AO3, the root of the network structure is called a 'meta'-tag, and behaves as a canonical tag. The parent-child relationship means that canonical tags get organized in hierarchies, which are felt to make searching easier. In the parent-child relationship, the parent acts as the meta-tag of the child tag node, even though both are, by definition, canonical tags. In addition, any synonymized tag is treated as equivalent to its corresponding canonical tag during searching. In searching, the syn group is effective 'merged with' the canonical tag.

Table 2.4 shows the type of nodes and type of edges now defined in the ATAS. It now includes the *hyperonym–hyponym* relationship between two canonical tags.

ATAS Node and Edge Types, v.3	
Node types:	*tag node*, *syn-set node*
Edge types:	*syn relationship*, *hyperonym–hyponym relationship*

Table 2.4: ATAS nodes and edges types (version 3).

2.1.4.4 Not Synonymizing, Not Canonizing a Tag

The existence of a free-form tag does not imply that a canonization or synonymization of it will necessarily take place at any point in the future. But for such free-form tags, there is the possibility of both operations if the tag, over time, meets the criteria. Free-form tags can be thought of as being represented by nodes, but these nodes do not have a place in the ATAS (at least, not until they get synonymized or canonized, if ever).

2.1.4.5 The AO3 Interfaces and the ATAS

As described in section 2.1.3, we defined ATAS as a way to model the underlying network structure in AO3.

The Additional Tag Archive Structure (ATAS) has been constructed to correspond to several functionalities in the AO3 archive. One functionality is tag completion. Suppose a contributor uploads and is tagging their fan-work as part of the upload. During that operation. while they are typing a tag, the AO3 platform uses its underlying network structure to provide tag completion suggestions (e.g., for a given prefix of characters, multiple tags are proposed). Contributors are not obligated to select from among the possible completions. They may instead enter their own tag. Also, when community members search for fan-works by tag, tag completion suggestions are provided. Correspondingly, the ATAS is also capable of providing this functionality.

2.1.4.6 The t-Rooted Additional Tag Archive Structure (t-RATAS)

AO3 maintains all "Additional Tags" for all fandoms somehow. We can envision that this enormous network contains all canonical and synned tags, and has been made handy for searching and for navigating the archive easily. This can be modeled using the ATAS in principle. And seeing that one canonical tag can belong to different fandoms and can have a large list of elements, we can see that the ATAS will, in fact, correspond to a complex network that can contain cycles and long paths in its structure.

In this project, we will focus on subnetworks of the general ATAS, since we wish to examine specific meta-tags. We will next develop the concept of a subnetwork with a particular root node.

As mentioned before, we will be defining a cut for the general ATAS as a way to focus our study to particular Additional Tags. This subnetwork will be defined by its root node, which will be a meta-tag node t and will be the origin of the network. This t-Rooted-ATAS (t-RATAS) will contain the origin node or root, as well as all of the network structures of its subtags (successors), its direct meta-tags, and its syn-set.

Figure 2.3 shows a hand-built visualization of a partial t-RATAS, specifically a subset of the Deaf-RATAS (Archive of Our Own, n.d.[b]). We can see in this Figure 2.3 that the network structure of the subtags of Deaf and its syn-set are shown. The term "Rooted Additional Tag Archive Structure" (RATAS) will be used in this document when discussing the structure in general, whereas, when specifying a root or origin t for a RATAS, the

structure will be referred to by the term t-RATAS.

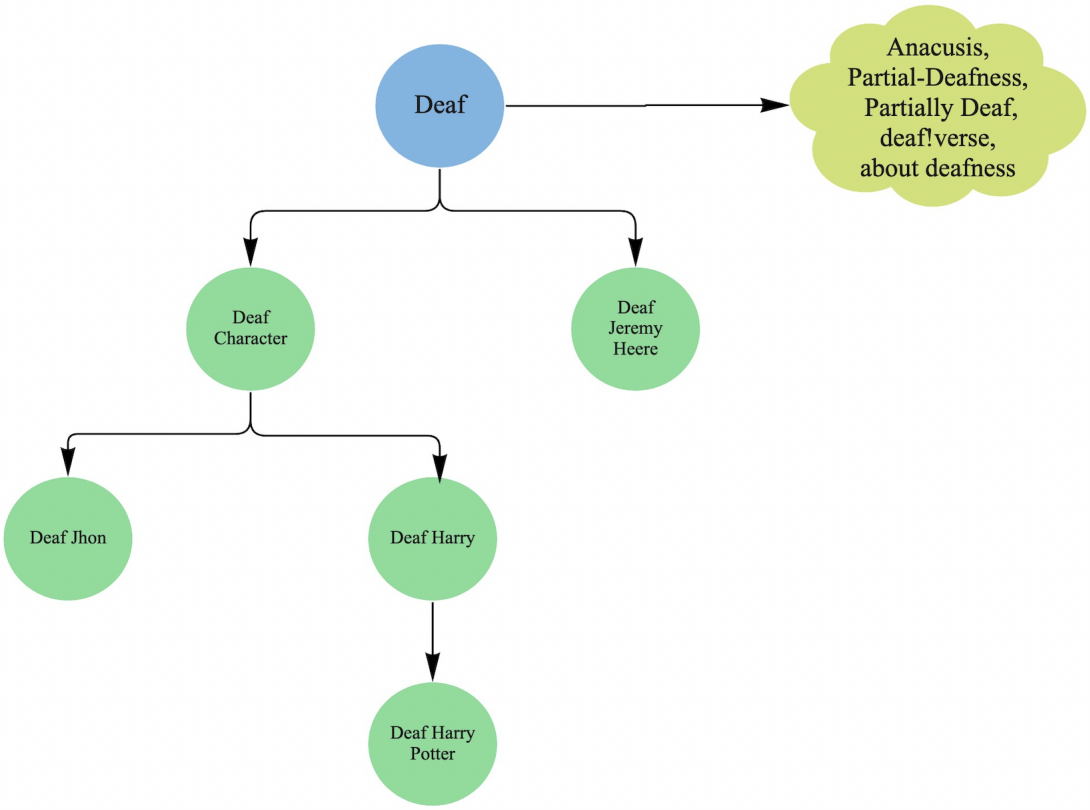


Figure 2.3: Deaf-RATAS example. Green nodes represent the canonical nodes in the subtag structure of Deaf, and yellow cloud-shaped node represent the Deaf syn-set. Example taken from a subset of the actual Deaf-RATAS in the archive

2.1.4.7 Basic t-RATAS, Expanded t-RATAS, and Fully Expanded t-RATAS

Now that we have defined the t-RATAS structure, we can identify several expansions and extensions. To distinguish among the base case and the extensions, we will establish the term **Basic t-RATAS** as the base case.

For instance, we can define an **Expanded t-RATAS**, which is an expanded version of the Basic t-RATAS which also includes all the syn-sets for all canonical nodes (and not only the root) involved in the Basic t-RATAS. This information can be accessed using the search

engine in AO3 for each canonical node.

Both the Basic t-RATAS and Expanded t-RATAS will consist of tags that have been 'wrangled', meaning that all of the tags began, at some earlier point, free-form tags and then have been recognized (e.g., canonized or synonymized into the AO3 network structure). Neither of these structures will contain unrecognized tags (aka free-form tags) because such tags, by definition, haven't been drawn into the tag structure. Free-form tags remain outside of the tag structure either because they have not been deemed to satisfy the inclusion criteria (see Table 2.1) or that they simply have not been processed by wranglers yet.

It is of interest, however, to get a sense of the slate of free-form tags that are being used to signal the concept of the root tag t within the AO3 community. Fan-work contributors are highly innovative and users continually devise new free-form tags. If and when there is sufficient uptake of new free-form tags in other works and by other users or if *it is deemed that such uptake is likely in the future*, then the new free-form tag will be wrangled into the tag structure. We can refer to these as 'candidate' free-form tags: they are related to root tag t but not (yet) wrangled. It is challenging to get a picture of this slate of these 'candidate' free-form tags.

One technique to capture the slate of 'candidate' free-form tags is via keyword matching. For instance, if a free-form tag uses the string of the term t within it, it is likely that a wrangler, at some point in the near future, will wrangle the tag into a syn set. It is possible to find, via keyword search, all such 'candidate' free-form tags.

To represent these 'candidate' free-form tags, we define a **Fully Expanded *t*-RATAS**, which consists of a particular Expanded t-RATAS with the addition of all of the 'candidate' free-form nodes related to the root t of the t-RATAS. Figure 2.4 shows a partial example of the Fully Expanded Deaf-RATAS, where all the disconnected nodes (shown in the grey colour) are these 'candidate' free-form tags (as they appear in the search of non-canonical tags of Deaf). The Fully Expanded t-RATAS in the Figure also shows all (non-trivial) syn-set for each canonical node.

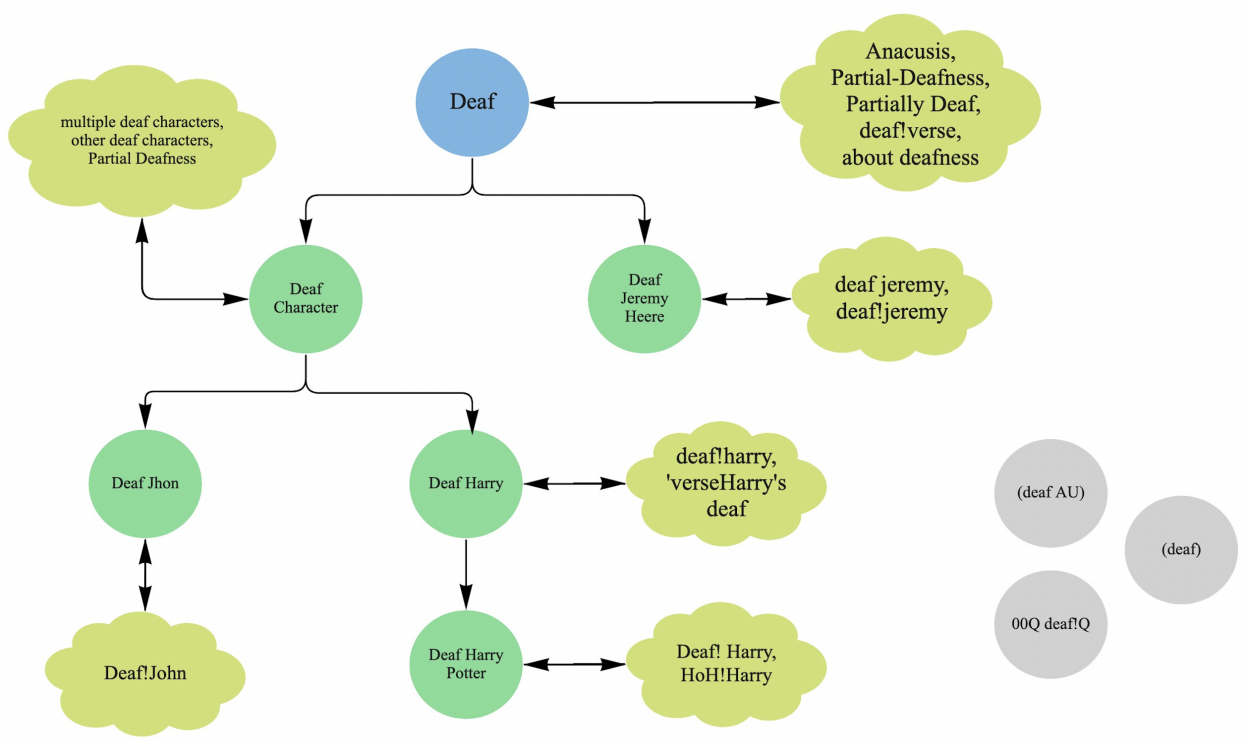


Figure 2.4: Fully Expanded Deaf-RATAS example. Green nodes represent the canonical nodes in the subtag structure of Deaf, yellow cloud-shaped node represent their syn-sets, and gray isolated nodes represent the free-form nodes. Example taken from a subset of the actual Deaf-RATAS in the archive.

2.1.4.8 The *t*-Story-Set

We define a Story-Set to be a set of fan-works. Of particular interest are Story-Sets in which all of the fan-works have been tagged with a specific tag. In a Story-Set, each fan-work will also include all of its metadata (and of particular interest, all of its Additional Tags). Story-sets can be scraped from AO3 as the collection of stories as result of browsing the fan-works of a specific tag, t. Figure 2.5 shows the collection of stories resulted by browsing Deaf.

It is straightforward to create a story set that consists of the collection of all the fan-works that have been tagged with a specific single tag t. We will refer to it t-Story-Set when specifying a t (e.g., Deaf-Story-Set).

Figure 2.5: Collection of fan-works matching the Deaf tag resulting in 542 pages of a max of 20 stories each.

2.1.4.9 Wayback Machine and AO3

It is possible to see prior snapshots of AO3 through the past years. This is possible using the non profit website *Wayback Machine* (Wayback Machine, n.d.). This site allows us to access snapshots of the archive that have been made throughout the past years. The snapshots might be limited and full access to certain URLs might not be available (e.g., when accessing certain URLs that are no longer available, the site redirects you to a close snapshot in time). Even despite these limitations, the Wayback Machine gives us with the freedom and opportunity of study the development of different RATAS over the years.

2.2 Related Work

2.2.1 Archive of Our Own (AO3) and Other Platforms

Different studies have highlighted the need for quantitative research methods for fan-fic communities because of the lack of computational tools to property mine and study different properties in the data. A study conducted by Duchastel de Montrouge and Baljko (2017) show the importance of a quantitative study, and the need for computational tools for mining the tag system in AO3, more specifically around disability tag. This leads to a need for computational tools to formalize the tag data using network structures (such as using the

Rooted Additional Tag Archive Structure, proposed in the previous section), and to visualize and study the properties of the structures, especially over time. A further study examined how the disability tag is used in the AO3 tag system and analyzed the tagging structure of AO3 around the "Disability" tag (Duchastel de Montrouge and Baljko, 2019). This analysis illustrated AO3's unique tagging system, the use of different tags in the "Additional Tag" category, and the tagging structure. The study also examined the process of tagging in AO3, as well as a selection of wrangling practices.

Dalton (2012) studied, through surveys, how users identified with AO3 and used its combination of tagging methods while searching. The study obtained positive results from the participants and reported that most of them made use of several different searching methods and found them to be effective. Therefore, Dalton (2012) concluded that AO3's "behind the scenes" wrangling process is working for users.

Black (2021) used surveys to study the relevance of the usage of tags to readers in AO3. The findings highlighted that the tags of fan-works usually factor in the users' decisions about whether to read a fan-fic or not. The study also pointed out the generally positive user experience of the AO3 tagging system.

Gupta et al. (2010) published a survey paper summarizing several techniques to study tagging systems. They analyzed diverse topics in the study of tagging, such as tag searching and tag visualizing. The survey paper concludes with suggestions for future work in the field.

Bischoff et al. (2008) studied and classified sample tags in order to understand the type of tags used in different resources and to obtain quantitative results from their usage. The study analyzed the collaborative tagging system of three platforms that were large at the time of the study, including Del.icio.us, Last.fm, and Flickr.

2.2.2 Network Analysis

2.2.2.1 Folksonomies as Co-Occurance Graph

Price (2019) performed an empirical study to reveal how "curated-folksonomies" were being used in AO3 as a collaboration between contributors and wranglers through tag analysis and interviews. Price executed a comparison on the AO3 tag system focusing on the co-occurrence graphs using a pre-wrangle and post-wrangle datasets in specific fandoms for specific characters (Price, 2019). They highlighted the practice of tag wrangling to not affect the original meaning of the tags while the pre- and post-wrangled datasets maintained remarked similarities. The study stated AO3's tag wrangling system to be largely effective,

having achieved a method for linking synonymous tags.

Price and Robinson (2021) extended the previous work performing a comparative case of study between AO3, Etsy and Tumblr to explore their classification and tagging systems, information gatekeeping, between others. They highlighted the different usage each community gives to tags and each classified them. They found the usage of tag by fans in several ways but for classification. For example, Tumblr usage of tags as meta-commentary while AO3 tag wrangling helps to organize the structure and the use of tags.

2.2.2.2 Folksonomy as Tripartite Graph

A. Chakraborty, Ghosh, and Ganguly (2012) proposed an algorithm to detect overlapping communities in folksonomies modelled as tripartite hypergraphs. They studied folksonomies that can be represented as tripartite graphs — i.e., user-resouce-tag — in platforms such as Delicious where users have multiple interests in or messages about the same resource (i.e., resources are often tagged with different tags). Instead of working on projections of the graph, they converted the hypergraph into its corresponding line graph, allowing to use community detection algorithms for unipartite graphs to obtain overlapping communities.

Capocci and Caldarelli (2008) analyzed the online collaborative tagging system "CiteULike". They represented the tagging system as a tripartite graph and performed a network analysis techniques around the clusters within the website. The study found that clustering coefficients can lead to the identification of semantic patterns.

Qassimi, Hafidi, Qazdar, et al. (2021) modeled the folksonomy of annotated books as a tripartite graph, user-resource-tag, and presented a graph based context-aware recommender system. They incorporated spectral clustering and the evaluation showed positive results on the tagging system used.

2.2.2.3 Folksonomy and Community Detection

In addition to last study, T. Chakraborty et al. (2017) conducted a survey of the start-of-the-art metrics used for the detection and evaluation of communities in networks. They used synthetic and real-world networks for a comparison of the metrics for measuring community structure. They found that the most popular and accepted metric of community analysis is Newman-Girvan's modularity. Also, their experiment's results showed that permanence and extended modularity density are most appropriate in measuring the quality of a community structure.

Complementary to these studies of overlapping communities, Crampes and Plantié (2014) presented a method that unifies community detection for three type of networks, while, at the same time, the method merges partitioned and overlapping communities.

Hotho et al. (2006) introduced a new algorithm, FolkRank, which was a new search algorithm for folksonomies that mine the structure of folksonomies detecting communities within them. On top of this, Shen and Wu (2005) studied folksonomies as networks from the WWW, indicating that the network displays both small-world and scale-free properties.

Additionally, Jäschke et al. (2007) studied tag recommendation in folksonomies. They evaluated and compared two recommendation algorithms, a user-based collaborative filtering, and a graph-based recommender (on top of FolkRank). Their study showed better results using the graph-based strategy, while FolkRank outperformed other approaches.

Shen and Wu (2005) analyzed folksonomies as complex networks, using Del.icio.us as a case study. They considered a Folksonomy Network to be a graph in which the nodes are tags, and nodes are connected only if their corresponding tags belong to the same piece of content. Then, they investigated this network for small-world and scale-free network properties. The results showed that the Folksonomy Network displayed both scale-free and small-world properties even though the whole structure was not measure and keeps growing/evolving.

Leskovec, Rajaraman, and Ullman (2020, Chapter 10) covered techniques for the study of large scale social networks on the Internet. They explained different algorithms and techniques, such as community detection. They also covered "simrank", to discover similarities between the nodes in the network and to be used to detect communities within the similar nodes. Even though the focus of this project is not specifically a social network, the main properties and ideas of this source are still relevant, as they can be used to mine interesting findings in the RATAS.

2.2.3 Future Directions

An interesting area of work related to folksonomies and AO3 tagging system's structure could involve the mix of different study approaches, such combining sentiment analysis on folksonomies to improve the tagging system and the search engine. There are several examples of this kind of work. Etemadikhou, Azimzadeh, and Karamatfar (2020) studied how to address tag ambiguity in folksonomies, and developed an approach to improve sentiment analysis by focusing on user interest with several models of word embedding. Xie et al. (2016) introduced a framework to apply sentiment analysis into search in folksonomies. They

proposed two sentiment-based ranking techniques and compared them with standard methods. The study confirmed the improvement of applying sentiment analysis to the search through the use of tag-user profiles. This is applicable to the present study, since it is known that one of the uses of the "Additional Tag" category within the tagging structure of AO3 is for the expression of creator opinion, stance, or critical position. There is the potential to extrapolate this approach of Xie et al. (2016) towards the analysis of the use of the "Additional Tags" category as a type of conversation. Related to this is work by Hoque and Carenini (2016), who proposed a text analysis system to support the collection of conversations which can obtained from the sequence of a story. Hoque, Setlur, et al. (2018) studied how to apply the principles of linguistic pragmatics to visual analytics of conversations. This idea, combined with approaches to studying the content of the fan-fic stories, might bring new ways to visualize and study AO3 tag structure in combination with AO3's corpus of fan-fic stories.

Additionally, the prospect of mining the entire ATAS as a large network is daunting, and even more if this were to include the whole of all of AO3 tagging structure, including all of the tagging categories. For large networks, Heidari and Papagelis (2020) presented a random walk method studying the evolution of a large networks and identifying the "right time to obtain a new representation of the evolving network that balances accuracy and time performance." Methods such as this will be of great potential utility for the future study of large snapshots of the main ATAS in AO3.

2.2.4 AO3, Scrapers, Data Representation and Visualization Tools

As a complement to the body of work that focuses of network analysis, there are other studies that focus on techniques for the visualization of tags, both a given point in time and their evolution over time. Dubinko et al. (2007) developed a technique to visualize the evolution of tags in Flickr, an image-sharing community online. They introduced dynamic visualizations of the evolution of the images' tags, as well as algorithms and data structures to effectively generate these dynamic visualizations.

Archive of Our Own (n.d.[e]) states that the organization does not provide an API nor direct database access to the current state of the tagging information on the site and has not stated intentions to provide one. Still, they allows the scraping of the site subject to certain conditions (e.g., to take place at specific hours, avoid overwhelming the system), and has the agent string containing the word "bot".

Price (2019) used `SocSciBot` 4.1 to scrape the data in their study, and `NodeXL`, and

Gephi to clean the data, to build the network structure, and to perform the network analyses. Thelwall (2009) introduced SocSciBot, a crawler made by University of Wolverhampton, UK, for link analysis directed to social sciences and humanities. Smith et al. (2010) developed a free tool for social network analysis, which was directed to Marketing and Research applications. NodeXL has been updated and commercialized as its Pro version (Smith et al., 2016). In addition to NodeXL, Bastian, Heymann, and Jacomy (2009) introduced Gephi as an open source tool for visualization and analysis of graph and networks.

Messina (2021), who studies fanfiction and the critical creating methods of authors in challenging socially prevailing beliefs, focuses on AO3. Their study scraped the metadata from the fan-fic stories on AO3 using an third party open source tool, which facilitated the process of obtaining of the metadata of the stories.

Related to this work is an open-source software application, DaScra , which was developed within the PiET Research Lab at York University (unpublished). The tool extracts data from AO3 by selected fandom. For all the fan-works in the specified fandom, it scrapes all of the meta-data, including all of the different tags from each of the categories, bookmarks, category, number of comments, hits, kudos, date of publication, number of words, rating, stats and other attributes. The tool stores the scraped meta-data in a file using the CSV file format.

When studying network representation tools and their use for analysis, NetworkX shows great popularity in different studies. Ladd et al. (2017) offered a guide on different utilities of NetworkX through a network analysis example. They performed the different steps while performing an analysis project while demonstrating the study of different properties. Hossen et al. (2022) used NetworkX as the main tool while performing network analysis applying graph theory study while designing a marketing network. They described different methodologies to create their networks, which can bring interesting points of view to the development of tools as well as to the use of graph theory for future work on the study of AO3 tagging system.

2.3 Research Questions: Revised

Q1: The starting point of this project is that there is information to be gained from the AO3 tags — and the "Additional Tags" in particular — and this information can be represented as a network.

(a) How to build t-RATAS and t-Story-Set for different tags t from the AO3 interfaces?

(b) How can the actions that contributors, and wranglers perform in AO3 be represented in terms of operations on the t-RATAS network representation?

(c) Which structural properties can be extracted and visualized from the t-RATAS structures? What can these properties tell us about the AO3 'Additional Tags' category (at a given point in time)?

Q2: We understand that the AO3 tags are not static: they change over time, by the actions of both the users and the tag wranglers. There is a day-by-day or even hour-by-hour dynamism. It may be too ambitious to capture the full dynamics, so we might wish to simplify the task to capturing 'information snapshots' at specific points in time. Then, the difference between two snapshots would capture the changes that have happened to the tag information between two points in time.

(a) How can we capture a snapshot of the t-RATAS at different points in time?

(b) Given two t-RATAS snapshots, which technique(s) can be developed to compare them with one another, in order to express change over time?

(c) What can t-RATAS tell us about the major changes to the AO3 tag structure over the past years of activity?

Q3: We know that the behaviour of the contributors has an impact on the tasks that the wranglers need to perform. We also know that the wranglers' actions have an impact on the users of AO3 (both readers and contributors). This means that the tag information is co-created by the wranglers and the users via bilateral paths of influence: through the tag information over time, the wranglers impact users and then the users impact wranglers (and vice-versa). Our conjecture is that there are multiple bilateral paths of influence.

(a) Which actions by fan-fic contributors impact the wranglers (and, conversely, which actions by wranglers impact fan-fic contributors) and how can these actions be illustrated through operations on the network?

(b) Which technique(s) could be used to detect bilateral paths of influence that may have occurred over the past years of AO3 activity?

2.4 Summary

This Chapter explained the overview of the tag wrangling process in AO3 while defining the different tags on the Additional Tags category in its tagging system structure. It set the ground of the network representation of the Additional Tag category with its structure, type of nodes and edges. Additionally, it is introduced the Additional Tag Archive Structure (ATAS) as the most general structure of the Additional Tags category and it was defined a sub-network of this as a t-RATAS which will be defined over a origin tag. This network can be defined and built using the AO3 interface and the wayback machine for different snapshot over the past years. It is discussed a literature review around the studies on AO3, how to perform network analysis on its tagging structures, and which tools exist currently to scrape and visualize AO3 data. Finally, it is revised the research questions defined in Sec 1.5 accordingly to the definitions of the network structures and the studies of the state of the art.

Chapter 3

Methodology

This chapter describes the methodology that was developed and employed in order to develop answers to the research questions that are posed in this thesis project:

Q1 (a) How to build t-RATAS and t-Story-Set for different tags t from the AO3 interfaces?

(b) How can the actions that contributors, and wranglers perform in AO3 be represented in terms of operations on the t-RATAS network representation?

(c) Which structural properties can be extracted and visualized from the t-RATAS structures? What can these properties tell us about the AO3 'Additional Tags' category (at a given point in time)?

Q2 (a) How can we capture a snapshot of the t-RATAS at different points in time?

(b) Given two t-RATAS snapshots, which technique(s) can be developed to compare them with one another, in order to express change over time?

(c) What can t-RATAS tell us about the major changes to the AO3 tag structure over the past years of activity?

Q3 (a) Which actions by fan-fic contributors impact the wranglers (and, conversely, which actions by wranglers impact fan-fic contributors) and how can these actions be illustrated through operations on the network?

(b) Which technique(s) could be used to detect bilateral paths of influence that may have occurred over the past years of AO3 activity?

3.1 Approach

Rather than tackle each of the three research questions separately, we developed an approach to work towards the answers in a coordinated manner. The approach consists of the following sequence of objectives:

(i) **Formal Representation:** Develop a formal, network-based, static representation of the t-ATAS and other new structures of interest. These network structures include the t-RATAS, Expanded t-RATAS, Fully Expanded t RATAS, and networks to represent the Story-Set. Static here means that the representation captures a snapshot of the AO3 data in time.

(ii) **Action Representation, Individual:** Enumerate the actions that can be performed by contributors and wranglers on the networks as per the AO3 wrangler guidelines, and express those action in terms of operations on the formal representations that were developed to under objective (i) Describe these actions in terms of their direct and indirect power to impact the t- ATAS (i.e., the political aspects of these actions).

(iii) **Action Representation, Sequences:** Identify sequences of interleaved actions by contributors and wranglers that are indicative of ATAS co-creation.

(iv) **Scraping:** Develop computational techniques that can scrape the data from AO3 that is necessary so that the networks developed under objective (i) can be built. Ensure the techniques operate on the current version of the AO3 repository and, to the greatest degree possible, on the archival versions of the AO3 repository (i.e., as per section 2.1.4.9). The techniques should be scalable (for potential further development) and should be flexible (i.e., be able to operate upon any specified subset of the AO3 repository). The techniques should serve to provide a snapshot of the AO3 data for a given point in time (i.e., as suitable for the creation of static versions of the network structures).

(v) **Network Building and Visualizing:** Develop computational techniques to build the network structures from the data that gets scraped using the computational techniques under objective (iv) and to visualize the network structures.

(vi) **Network Analysis, Static:** Develop network analysis techniques of the static network structures to derive quantitative characteristics that would be useful for fan-work

scholars.

(vii) **Network Analysis, Diffing:** Develop network analysis techniques that are capable of detecting and analyzing the differences between two static versions of the same t-RATAS. Use this technique as a way to characterize network dynamics (i.e., to represent a change over time to the network, such as a result of tag wrangling or free-tagging by a fan-fic contributor). Express the differences in terms of the different possible contributor and wrangler actions, as developed under objective (ii).

(viii) **Network Analysis, Diffing in Sequence:** Develop network analysis techniques to identify interactions between contributors and wranglers that are mediated by the tag archive structure, expressed in terms of the action sequences that were identified in objective (iii). Extract quantitative data about this co-creation and identify how the tag wranglers influence the network over time and to identify their focus points.

The specific steps outlined above were arranged into 4 phases of work, as shown in Table 3.1.

Phase	**Objective**	**Approach**
Phase 1	Objective (i), **Formal Representation**	Sections: 3.2.1 to 3.2.3.2
	Objective (ii), **Action Representation, Individual**	Section 3.2.4
	Objective (iii), **Action Representation, Sequences**	Section 3.2.5
Phase 2	Objective (iv), **Scraping**	Section 3.3
Phase 3	Objective (v), **Network Building, Visualizing**	Section 3.4
Phase 4	Objective (vi), **Network Analysis, Static**	Section 3.5.1
	Objective (vii), **Network Analysis, Diffing**	Section 3.5.2
	Objective (viii), **Network Analysis, Diffing in Sequence**	Section 3.5.3

Table 3.1: The phases of the research project.

3.2 Phase 1: Representation Development

3.2.1 The Additional Tag Archive Structure (ATAS)

As previously discussed in section 2.1.3, we developed several conjectures about the nature of the underlying network structures within the AO3 repository, specifically the *Additional Tag Archive Structure* (ATAS). Although the AO3 documentation does not include any mention

of the ATAS, we can infer its existence, characterize its structure through inference, and represent it as networks through the following definitions.

3.2.1.1 Definition: The Additional Tag Archive Structure (ATAS)

Let the **Additional Tag Archive Structure** be represented by network G as follows:

$$G = (V, E), \text{ where } V(G) \text{ represents all canonical tags and their syn-sets in AO3, and}$$
$$\text{where } E(G) \text{ contains all directed edge from } v_i \text{ to } v_j$$
$$\text{iff } v_i \text{ is a so-called meta-tag of } v_j, \forall v_i, v_j \in V(G)$$

3.2.2 The t-Rooted Additional Tag Archive Structure (t-RATAS)

As previously conjectured in section 2.1.3, the entire ATAS structure is extremely large and complex. As described in section 2.1.4.6, we expect it to be challenging to fully extract and model it, at least in terms of a first step. Researchers interested in the AO3 tagging system and its behaviour within the AO3-community are most likely to be focused on specific social topics within the archive. In section 2.1.4.6, we described the concept of a cut of this network, a subnetwork to be defined by its root node t, a network we later (section 2.1.4.7) referred to as a 'Basic' t-RATAS . In that section, we described expansions and extensions of this rooted network, in terms of including syn-sets and associated 'candidate' free-form tags: 'Expanded' and 'Fully Expanded', respectively.

3.2.2.1 Definition: The Basic t-RATAS

Suppose we identify a canonical tag t as a focus tag.
Let the sub-network $C_t = (V, E)$ represent the **t-RATAS**, which is the subnetwork of the

Additional Tag Archive Structure that is rooted at the node that corresponds to a tag t.

$C_t = (V, E)$ is an induced sub network from G, with edges $E(C_t)$ such as: (eqn:3.2.1)

$V(C_t)$ will be composed of all successors of t (with their respective RATAS), its direct predecessors from G, and the syn-set of t.

$E(C_t)$ will be composed by the edges in G connecting the nodes $V(C_t)$ which are directed by the relationship meta-tag to subtag.

$$C_T = \bigcup_{t_i \in T} C_{t_i}, \text{ the induced sub network of the ATAS containing all} \quad \text{(eqn:3.2.2)}$$

C_{t_i} for a given T, a set of canonical tags.

3.2.2.2 Definition: The Expanded t-RATAS

Given a Basic t-RATAS, C_t, and equation (eqn:3.2.2), we define the **Expanded t-RATAS** X as $X = C_{T=V(C_t)}$ where:

$$T = V(C_t) \text{ iff } v_i \text{ canonical, for all canonical nodes in } C_t. \quad \text{(eqn:3.2.3)}$$

3.2.2.3 Definition: The Fully Expanded t-RATAS

First, we define the set of t-FreeForm tags to be all free-form tags which appear on the search of a tag t after applying the filter 'non-canonical'.

For a given Expanded t-RATAS, X, and the set of t-FreeForm for the same t, we can define the **Fully Expanded t-RATAS** $Z = (V, E)$ as follows:

$$V(Z) = V(X) \bigcup t\text{-FreeForm tags} \quad \text{(eqn:3.2.4)}$$
$$E(Z) = E(X)$$

As the definition of these networks, a node represents a tag, and an edge represents

a connection from one node to another with the meta-tag/subtag relationship. It can be interpreted as a directed network if we take into account the direction from the metatag to its subtag. This direction can also be ignored and considered an undirected network. In case we can be see the relationship as the link between both tags to go from one to the other unidirectional. Both approaches bring different perspectives and results to analyze.

3.2.3 Story-Sets

We would like to examine sets of stories, as defined in section 2.1.4.8, with their Additional Tags. These sets are needed in order to study how the wrangling process co-occurs with the uploading of stories to AO3 by fan-work creators, the different political acts behind each user's interaction with the archive, and the qualitative measures for a given social topic in AO3. These sets of stories allow the creation of new datasets and different network structures for further analysis and provide quantitative information of interest to the scholars lacking useful tools to get them easily.

3.2.3.1 Definition: The t-Story-Set

To define a t-Story-Sets, let us first develop some preliminaries. Let s_i represent a fan-work (aka a 'story'). Now, given a tag t, let S_t represent the set of stories returned by AO3 when t is used as a search term. Thus, a t-Story-Set corresponds to $S_t = \{s_1, s_2, \cdots\}$.

3.2.3.2 Definition: Auxiliary Network to Connect Stories via Tags Relationships

We define the following auxiliary network for subsequent use:

$$N_T = (S_T, E_T), \text{ where} \qquad \text{(eqn:3.2.5)}$$

T represents a set tags $T = \{t_1, \cdots, t_n\}$

S_T represents a set of stories, with all $t \in T$

E_T consists of all edges $e_{i,j} = \langle s_i, s_j \rangle$ such that $t_k \in T_{s_i}$ and $t_k \in T_{s_j}$ $\forall t_k \in T$

the weight of the edge $e_{i,j} = \langle s_i, s_j \rangle$ is determined by the

number of (unique) tag relationships between the stories s_i and s_j

Equation (eqn:3.2.5) defines an auxiliary network in which the nodes are stories (e.g., the stories within a story-set $S = \{s_1, s_2, \cdots\}$). Nodes are connected by virtue of the stories

sharing the use of one or more tags (e.g., the tags within a given tag sets T).

3.2.3.3 Definition: Induced Tagsets from a Story-Set

Each story s_i has associated with it (under the "Additional Tags" category) a set of free-form tags F_{s_i}, a set of canonical tags C_{s_i}, and a set of syn tags Y_{s_i}.

$$\text{Let } T_S = F_S \cup C_S \cup Y_S, \text{ where:} \tag{eqn:3.2.6}$$

$$F_S = \bigcup_{s_i \in S} F_{s_i}, \text{the set of free-form tags for a Story-Set } S \tag{eqn:3.2.7}$$

$$C_S = \bigcup_{s_i \in S} C_{s_i}, \text{the set of canonical tags for a Story-Set } S \tag{eqn:3.2.8}$$

$$Y_S = \bigcup_{s_i \in S} Y_{s_i}, \text{the set of syn tags for a Story-Set } S \tag{eqn:3.2.9}$$

Thus, we now have different tagsets that can be used: F_S, C_S, Y_S, and T_S, as defined in (eqn:3.2.6), (eqn:3.2.7), (eqn:3.2.8), and (eqn:3.2.9).

3.2.3.4 Definition: Free-Form Tag Network for S and t-RATAS (Fully Expanded)

Suppose we have a given ***t*-Story-Set** $S = \{s_1, s_2, \cdots\}$, obtained from the result of browsing the fan-works of a tag t, and a **Fully Expanded t-RATAS** $Z = (V, E)$, built on the same origin tag t. We can now build a network which consists of nodes that correspond to the stories in S and their edges representing shared use of the *free-form* tags in Z.

$$\text{Let } N_{F'_S} = (S, E_{F'_S}) \text{ be the Free-Form Tag Network for } S \text{ and } Z, \tag{eqn:3.2.10}$$
$$\text{as defined using definition (eqn:3.2.5), where}$$
$$F'_S \text{ is the subset of free-form tags in S that also belong to } V(Z); \text{ and}$$
$$F'_S \subseteq F_S \text{ where } \forall t \in F'_S,\ t \in V(Z) \text{ and } t \text{ is a free-form node in } Z.$$

3.2.3.5 Definition: Canonical Tag Network for S and t-RATAS (Fully Expanded)

Suppose we have a given **Story-Set** $S = \{s_1, s_2, \cdots\}$ and a **Fully Expanded t-RATAS** $Z = (V, E)$. We can now build a network which consists of nodes that correspond to the

stories in S and their edges representing shared use of the *canonical* tags in Z.

Let $N_{C'_S} = (S, E_{C'_S})$ be the Canonical Tag Network for S and Z, (eqn:3.2.11)
as defined using definition (eqn:3.2.5), where
$T = C'_S$ is the subset of canonical tags in S that also belong to $V(Z)$; and
$C'_S \subseteq C_S$ where $\forall t \in C'_S,\ t \in V(Z)$ and t is a canonical node in Z.

3.2.3.6 Definition: Syn Tag Network for S and t-RATAS (Fully Expanded)

Suppose we have a given **Story-Set** $S = \{s_1, s_2, \cdots\}$ and a **Fully Expanded t-RATAS** $Z = (V, E)$. We can now build a network which consists of nodes that correspond to the stories in S and their edges representing shared use of the *syn* tags in Z.

Let $N_{Y'_S} = (S, E_{Y'_S})$ be the Syn Tag Network for S and Z, (eqn:3.2.12)
as defined using definition (eqn:3.2.5), where
$T = Y'_S$ is the subset of canonical tags in S that also belong to $V(Z)$; and
$Y'_S \subseteq Y_S$ where $\forall t \in Y'_S,\ t \in V(Z)$ and t is a syn node in Z.

3.2.3.7 Definition: Full Tag Network for S and t-RATAS (Fully Expanded)

Suppose we have a given **Story-Set** $S = \{s_1, s_2, \cdots\}$ and a **Fully Expanded t-RATAS** $Z = (V, E)$. We can now build a network which consists of nodes that correspond to the stories in S and their edges representing shared use of any tags in Z.

Let $N_{T'_S} = (S, E_{T'_S})$ be the Full Tag Network for S and Z, (eqn:3.2.13)
as defined using definition (eqn:3.2.5), where
$T = T'_S$ is the subset of tags in S that also belong to $V(Z)$; and
$T'_S \subseteq T_S$ where $\forall t \in T'_S,\ t \in V(Z)$ and t is a node in Z.

3.2.4 Action Representation, Individual

3.2.4.1 Wranglers

Tag wranglers impact the AO3 tag information in two different ways: they canonize or synonymize tags.[2] Since these acts have an impact on the tag information and only tag wranglers can bring about these impacts, these acts can be seen as political acts (or acts having power).

We can model these acts via operations on the ATAS (and specifically, the relevant section of it, the RATAS). These actions can be made visible through the connection of the Story-Set and the RATAS over time, which serve to create new connections between fan-works.

Below is a list of mappings between wrangler actions and network operations and between Fan-work Creators and network operations. In principle, the network operations could be automatically detected in dynamic network datasets, which could lead to the automatic detection of wrangler and creator actions. The mappings are listed below as a first step towards this analysis technique (which is discussed further in section 3.5.2.

Given the ATAS, $C = (V(C), E(C))$, the nodes v_i, $v_j \in V(C)$, and a tag t, we can define the actions of wranglers as follows:

(i) **Canonizing tag** t:

- Inserting t into the ATAS canonical node set: $t \cup V(C)$
- Connecting t with its predecessor and successors if any: $(\langle v_i, t\rangle, \langle t, v_j\rangle) \cup E(C)$
- Creating the t syn-set which, at first, will contain only itself: $\{t\}$

(ii) **Synonymizing the tag** t:

- Inserting t into the ATAS syn nodes: $t \cup V(C)$
- Connecting t with its canonical tag v_i syn-set: $(\langle v_i, t\rangle, \langle t, v_i\rangle) \cup E(C)$

3.2.4.2 Fan-work Creators

Similarly, creators contribute to AO3 tagging structure as the source of tags of the archive. Their decision of creating, modifying and eliminating stories and with them their tags indirectly influence the decision of the wrangler and with it the structure of the tagging system. A creator possible actions with AO3 tags:

[2]Wranglers may also move or remove nodes from the AO3 network, but we will not model these actions.

- Creating new tags which means the insertion of isolated nodes (free-form nodes) into the Fully Expanded t-RATAS
- Using existing canonical, or syn tags which impact Induced Tagsets and might affect on future wranglers actions
- Using existing free-form tags which impacts Induced Tagsets pushing them to be processed by wranglers
- Update a fan-work changing, deleting, or adding a tag which impacts Induced Tagsets towards possible changes in the ATAS

3.2.5 Action Representation, Sequences

The actions of creators (e.g., tagging new stories, updating already created ones and deleting from them previously-added Additional Tags) are reflected in the main ATAS as updates, creation and deletion of tags. These actions are also performed by wranglers as they canonize or synonymize, depending on the previous tags in the ATAS and the popularity of new tags. Thus, there is an interaction (e.g., an indirect collaboration) between the acts of wranglers and the acts of the creators. This interaction is reflected in the evolution of the ATAS with the different actions both users perform.

In addition, the sequence of actions by creating new stories using existing tags, or any of the wranglers/cretor actions beforementioned will also affect the network representation of the story-set combined with the RATAS. Focusing on the creators operation, an update of a fan-work tags might change connections in these representations over time, deleting or adding a new connection between fan-works.

This let us analyze how the actions of creator create a sequence of possible actions of the wranglers to do in the ATAS:

- Creating new free-form tags opens the potential of a canonization of synonymization by wranglers in the future.
- Using existing free-form tags opens the potential of a canonization of synonymization by wranglers in the future
- Using existing canonical, or syn tags make existent tags “stronger” in the ATAS popularizing them to their use by other creators

- Update a fan-work changing, deleting, or adding a tag indirect interaction with the ATAS as some tags were already canonized or synonymized and their meaning in the ATAS

Similarly, the actions of the wrangler bring new actions to the creators while deciding which tag to use or not. Creators may try to tag the stories bringing more visibility to them and then they will use more canonical tags instead of creating new ones. Still, they might use existing or new free-form tags trying to make them popular and with them all the stories tagged with them. This is a complex indirect relationship between wranglers and creators that affects the development of the ATAS.

In principle, these sequences of network operations could be automatically detected in dynamic network datasets, which could lead to the automatic detection of wrangler and creator interaction. The sequences have been described above as the first step toward this analysis technique (which is discussed further in section 3.5.3).

3.3 Phase 2: Scraping

The objective of this phase was set out at the beginning of this chapter:

Objective (iv) **Scraping:** Develop computational techniques that can scrape the data from AO3 that is necessary so that the networks can be built. Ensure the techniques operate on the current version of the AO3 repository and to the greatest degree possible for archival versions of the AO3 repository (e.g., as per section 2.1.4.9). The techniques should be scalable for it further development improvement and should flexibly operate upon specified subsets of the AO3 repository. The techniques should serve to provide a snapshot of the data for a given point in time (i.e., data suitable for the creation of static versions of the network structures).

3.3.1 Approach

Archive of Our Own (n.d.[h]) states the lack of any APIs or pre-existing open-source projects that could be used to mine the information needed from the archive for this study which leads the project to develop and implement its own scraper to access the data following the guidelines defined. We intend this to become an open-source tool that will mine and scrape data from AO3, mainly focusing on the Additional Tags information of the fan-works and the ATAS.

The workflow of this phase is illustrated in Figure 3.1. It consists of two main modules which will both scrape the AO3 website and generate datasets that will be usable in further analyses. These modules in the Data Scraping Workflow are: "Story-Set Scraping", and "Tag-Structure Scraping". Each module is independent of the other and generates its own kind of dataset. Each of the two types of datasets are used for building the Story-Set and the different versions of the RATAS structures, as defined in Section 3.2, visualizing them and mining useful insights to describe and further study.

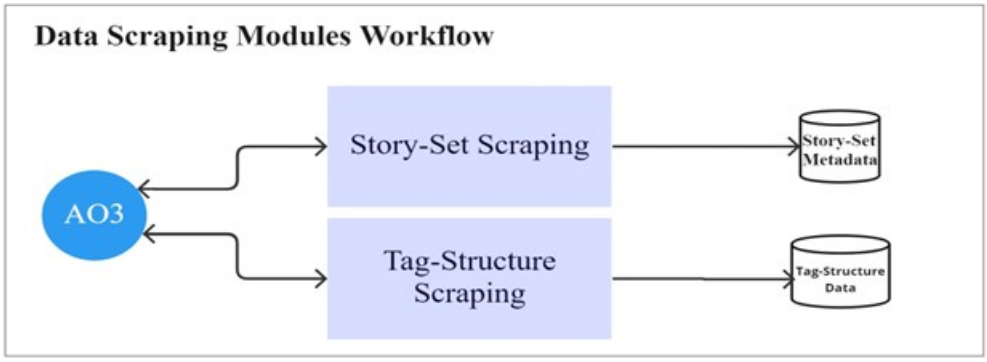


Figure 3.1: Schematic that provides an overview of the data scraping modules Workflow. The schematic illustrates the flow of data between each of two modules and AO3, as necessary to obtain all required data and to produce the processed datasets. There are two modules: Phase 1a "Story-Set Scraping", Phase 1b "Tag-Structure Scraping"

Each module has its own flow of interaction with AO3 (which is each module's main input or source of data). This workflow is as follows:

(i) **The "Story-Set Scraping" module**: accesses the AO3 website to obtain the data of different batches of fan-works. This module generates a flow of data in/on-going between the module and AO3 while accumulating all the metadata to generate an external file which will contain either a pre-defined list of fan-works or all the story-set obtained as result of the search of an specific tag of interest.

(ii) **The "Tag-Structure Scraping" module**: similar to the module above, sends queries to AO3 in batches. This module creates a flow of data while accumulates in order to store the RATAS information of different tags.

3.3.2 Phase 2a: Story-Set Scraping

3.3.2.1 Requirements

1. **Core Functionality**: The Story-Set Scraping module was envisioned as being in charge of getting a set of stories with their metadata from the AO3 archive, to construct t-Story-Sets, as per section 3.2.3.1.

2. **Input**: It should receive as input a focus tag of any type (canonical, synonymized, free-from), tag t.

3. **Output**: Its output should consist of a story-set metadata as a dataset stored in a file (in a format such that the data can be used to populate a network structure). From the metadata, the induced tagset, as defined in section 3.2.3.3, can be obtained.

4. **Other considerations**: When scraping data from the Archive of Our Own (n.d.[e]) site, the code must follow the access guidelines stated by AO3, which are: the agent should delay each request to the site, and has the agent string contains the word "bot".

3.3.2.2 Development

A tool was developed using `Python` as the main language. This has been named `ao3_analyser`. The interface of `ao3_analyser` is a Common Line Interface (CLI) and its full documentation is provided in Appendix A.

The functionality for this Story-Set Scraping module was implemented as part of the `ao3_analyser` tool, via the `DaScra` flag. The module's functionality is accessed via the CLI with the command `-ds ARG, -DaScra ARG`. It receives one argument, which indicates the tag t, as per the requirements above.

This module also allows optional parameter to set the name of the output file, either with or without a path: `-o ARG, -OutputName ARG` or `-op ARG, -OutputPath ARG` (for more detail, see Appendix A). In case these parameters are not used the module uses a defaul name and default path.

The implementation of this module was on top of `DaScra` after additional modifications. It is implemented in `NodeJS` using `puppeteer` to access the AO3 website and scrape the metadata for the stories. The module itself has its CLI which facilitates its interaction.

`DaScra` receives as main input a tag t and it scrapes all of the fan-works as per the requirements above.

AO3 has adopted a convention such that the site organizes all fan-works that contains the tag "TAG_NAME" in several pages. This can be accessed directly through the URL `https://archiveofourown.org/tags/TAG_NAME/works`, or can be reach under the `Browse`→`Tags` and manually search for the tag of interest. Figure 2.5 shows the result page of browsing all fan-works in "Deaf" tag at `https://archiveofourown.org/tags/Deaf/works`. Thus, the module accesses the URL in batches with a waiting time in between queries. The module may not be able to retrieve all fan-works when the set is extremely large (e.g., larger than $100,000$) due to a limitation within AO3, which only shows the first $100,000$ fan-work. In order to access to more stories it is needed to start using the different filtering features they provide.

For each fan-work the module extracts, it gets its metadata as shown in appendix B.2: Title, Additional Tags, Archive Warning, Author, Bookmarks, Category, Chapters, Characters, Comments, Fandom, Hits, Kudos, Language, Rating, Relationship, Series, Part, SourceURL, Updated, Words. As noted previously, the category Additional Tags is the main focus of this project, but the data from the other categories is also scraped in case of future utility. Finally, `DaScra` saves all the information in the form of a file with the name `.xlsx` Excel file.

`DaScra`, as independent tool, was expanded to be able to implement a pause feature, which causes the module to temporarily stop issuing queries for a period of two minutes after a number of 20 queries. This feature is necessary because of the guidelines for scraping the website. Also, the module originally only scraped a maximum of 1000 stories before stopping. In order to mine all stories showed in a search, it was necessary to extend it and get page by page all the stories. With this, after a number of queries it was necesary to stop and wait before continuing to not overwhelm the site. Furthermore, it was implemented in `DaScra` tool a CLI for its own interaction aside form the CLI of `ao3_analyser`. Then a feature to continue mining data from where it left which is perfect to extend already mined datasets.

3.3.2.3 Limitations

This module still presents some limitations. As mentioned above, the module has a limitation when the set of fan-works to be scraped is extremely large. This is because AO3 only shows a maximum of $100,000$ fan-works for a specific tag search, divided into 500 different pages. This implies that by default it will only appear the most recent $100,000$ fan-work, but the site offers different filtering features to access the rest of the stories. This module can extend its functionality by implementing filtering for specific different periods of time. Large sets could be sub-divided into a sequence of smaller retrieval tasks, which would allow for scraping

of all the works related to a tag, from the beginning of the archive.

The module originally scraps a number of stories (1000) as before mentioned. It was modified with an external python script to be called continuously until all the stories were mined. The script assures call `DaScra` continuously until a number of stories (passed by a parameter) are scraped starting where it left. This way we assure the completeness of mining all possible stories for a study using the implemented script.

Another limitation while scraping the data concerns the dynamic property of the archive. Popular tags, such as the Fandom tag 'Super Natural (2005 TV)', are continuously growing, which means the information in the archive could be changing during the scraping process, leading to artefacts in the data (e.g., the data keeps moving while the bot needs to wait certain minutes to not overwhelm the archive site). This issue is not likely to be resolved without access to a live `API` to access the underlying databases directly (as opposed to the served webpages).

3.3.2.4 Testing/Validation

The module was initially built on the previous working `DaScra` and then updated to new features and then revised ensuring its correctness. For this, a manual search was made finding a tag with a small number of stories which was scraped and compared to the results. Then we concluded it correctness mining the story-set of a given tag.

This functionality of this module was tested by comparing its output to a hand-constructed dataset (e.g., a manual search was performed for a tag and the matching stories were manually scraped). The output of the module matched the manually-build output dataset. Additional spot-checks were performed. This is an admittedly heuristic approach. A more robust testing process is suggested in the future work section.

3.3.3 Phase 2b: Tag-Structure Scraping

3.3.3.1 Requirements

1. **Core Functionality**:

   (a) The Tag-Structure Scraping module should obtain all the information that is necessary to build a t-RATAS from AO3's website.

   (b) The Tag-Structure Scraping module should obtain all the information that is necessary to build a t-RATAS from the Wayback Machine.

2. **Input**: It should receive as input a tag t, which represents the tag of interest.

3. **Output**: Its output should consist of a file that contains all of the information to build the Basic t-RATAS, the Expanded t-RATAS, and the Fully Expanded t-RATAS. The file of the format should be suitable as input for modules to be used subsequently (to build, analyze, and visualize the network)

4. **Other considerations**:

   (a) When scraping data from the AO3 site, the code must follow the access guidelines as stated by Archive of Our Own (n.d.[e]). It should be delayed each request to the site, and has the agent string containing the word "bot".

   (b) Given that the tool user may be interested to scrape the data for multiple values of t, the module should offer the possibility to scrape the data for multiple values in parallel (during the same crawl over the AO3 website).

3.3.3.2 Development

The tool developed for this module is called `TagScra`. It is accessed via the main CLI of `ao3_analyser` with the command `-ts [ARG1], -TagScra [ARG1`. It receives the tag t as a string.

To address requirement 4b, the module was expanded to allow the possibility of multiple tags to be provided as a comma-delimited list: via the command `-ts [ARG1, ARG2, ...], -TagScra [ARG1, ARG2, ...]`.

This module was implemented in `NodeJS` using `puppeteer` to access the AO3 website and scrape the required information.

The functionality that implements requirement 1a will be described first. `TagScra` accesses the URL `https://archiveofourown.org/tags/TAG_NAME` on the AO3 site for each tag in the input. An example of this was shown in Figure 2.2 accessing the tag "Deaf".

First, the module iterates through all the listed tags and accesses `https://archiveofourown.org/tags/` plus the tag to obtain its information. Figure 2.2 shows a partial example of the information accessed for the tag "Deaf" which is classified to either canonical, syn, or free-form tag and its information is scraped. Depending on the type of the tag the information vary and it is displayed accordingly. If it is a syn tag it is shown and then scraped the canonical tag to which the tag was synonymized to. If it is a free-form it does not have additional information needed and then it is only identified

as free-form. Lastly, in case it is a canonical tag, it contains hierarchically ordered all the subtags, all its metatags and all the syn-set as Figure 2.2 shows. `TagScra` gets all this information, which defines a RATAS for each canonical tag mined,and stores it in a `JSON` file as shown in appendix B.1.

With this, we are able to mine any RATAS and get it's expanded version. This is done while iterating for each canonical tag in the origianl RATAS and scraping its syn-set and metatags. To complement these options for a list of tags, `TagScra` is able to mine all non-canonical tags of each canonical tag in the list. This allows complementing an Expanded RATAS to build a Fully Expanded RATAS where all tags, either canonical, syn or free-form tags, related to the same search of a focus tag are represented in a Network structure.

The Basic *t*-RATAS and the Expanded and Fully Expanded Variants

Aside from the main command, `-ExpandRATAS .JSON` can be used to scrape the Expanded RATAS of a Basic RATAS already mined. Similarly, we can get the data to generate the Full Expanded RATAS by first mining all the non-canonical tags connected to the origin of the RATAS. This is possible with the argument `-NonCanonical ARG1` which receives one argument as the focus tag for which the related non-canonical tags should be obtained. It generates a `.txt` file with a list of all non-canonical tags separated by comma. Then we can build the `Full Expanded RATAS` using this file and a `JSON` file containing a Expanded RATAS. This is possible through the CLI with the command `-ExpandFullRATAS .JSON .TXT` which adds all the free-form tags in the `.txt` file as isolated nodes in the Expanded RATAS.

This module also receives secondary optional parameters to set the output file name and path as defined with `DaScra`.

The module builds the Fully Expanded t-RATAS relying on string matching with term t. However, there may be free-form tags that are related to t that do not use the tag string. An example of this is shown in the upcoming chapter in Figure 4.4 on p. 76, which shows the canonical tag *Deaf* and the synned tag *Anacusis*. This approach may exclude many different tags that are semantically related to each other but do not have string matching.

The *t*-RATAS `JSON` File Structure

The same `JSON` file format is used for all of the different versions of the RATAS. The file stores, as the main keys, the different tags were scraped as roots of the RATAS where each

of them store their information. Each key stores the tag type (e.g., canonical_tag, syn_tag, or freeform_tag), the parent tag which is the fandom which it belongs to as "parent_tag", a list of its syn-set and metatags (which will be empty in case the tag is not canonical) as "synned_tags" and "metatags" respectively, the canonical tag (in case the tag is a syn tag) the tag is synonymized to, and a hierarchical substructure with the subtags and their respective list of successors as "subtags". This last property is composed by a list of all direct subtags as strings if they do not have successors or a subsequent dictionary structure where the direct subtag is the main key and its successors are stored following the same idea recursevely.

An example of the structure of a RATAS is shown in Figure 3.2 exemplifying the form of each type of tags based on Fully Expanded Deaf-RATAS shown in Figure 2.4. We can see a canonical tag *Deaf* storing its information. Its main difference are the properties "syn-tag", "metatag", and "subtags" properties storing its syn-set and hierarchical structure. The "subtags" properties is composed by two direct subtags ("Deaf Character" and "Deaf Jeremy Heere"). The subtag "Deaf Character" contains a list of successors displaying "Deaf Jhon" as the first one and "Deaf Jeremy Heere" that does not contain any successors as it is a string. Additionally, the next tag stored is "Anacusis" which represent the structure of a syn tag. Its main difference is the property "canonical_tag" which stores the canonical tag the syn tag was synonymized to, "Deaf" in this case. Lastly, "*deaf*" represents a isolated node of a non-recognized tag. These types of tag are added to the structure only when building the Full Expanded RATAS. They have all the properties empty as they do not have any.

```
{
    ''Deaf'': {
        ''type'': ''canonical_tag'',
        ''parent_tags'': [''No Fandom''],
        ''synned_tags'': [''Anacusis'', ''Partial-Deafiness'',
           ...],
        ''subtags'': [
            {''Deaf Character'':[''Deaf Jhon'', ...]},
            ''Deaf Jeremy Heere'',  ...],
        ''metatags'': []
    },
    ''Anacusis'': {
        canonical_tag:[''Deaf''],
        ''type'': ''syn_tag'',''parent_tags'': [''No Fandom''],
        ''synned_tags'': [], ''subtags'': [], ''metatags'': []
    },
    ''(deaf)'': {
        ''type'': ''freeform_tag'',''parent_tags'': [''No
           Fandom''],
        ''synned_tags'': [], ''subtags'': [], ''metatags'': []
    }

}
```

Figure 3.2: Structure of a JSON RATAS main tags, shown for the Fully Expanded Deaf-RATAS shown in Figure 2.4. The file stores three main tags, one for each type of tags in the Additional Tag category.

3.3.3.3 Wayback Machine

To implement the functionality for requirement 1b, a secondary parameter was added to the main command: `--AO3_version YEAR`. The parameter `YEAR` receives the year for which the Wayback Machine should be accessed. This parameter allows us to choose an version of the archive that is older than what appears on the AO3 site, using the Wayback Machine, provided there is a snapshot for the given year. This way we can generate old versions of the RATAS for a certain focused tag and compare its evolution over time.

For any AO3 URL of the form `https://archiveofourown.org/tags/TAG_NAME`, the wayback machine will archive a number of snapshots.

Because of the limitations of the Wayback Machine, we manually pre-fetched datasets for the *Disability* tag for the years 2012-2018 and 2020-2021 (2019 is not present in the archive).

Then we set up `TagScra` to have an argument that controls the years of the pre-fetched data to be obtained.

3.3.3.4 Limitations

The module that builds the Fully Expanded t-RATAS relies on string matching with term t. Since there may be free-form tags that are related to t that do not use the tag string, this approach may exclude many different tags that are semantically related to each other but do not have string matching.

The Wayback Machine has several limitations. It does not contain AO3 archives for all prior years. It can be unstable and can take a significant amount of time to load. When using the Wayback Machine, one need to be aware of redirecting edges for specific paths.

Because of the decision to pre-fetch data from the Wayback machine, the `--AO3_version YEAR` option is guaranteed to work only with the tag Disability and only for certain years. In future work, functionality should be extended to any tag and to implement controlling when encountering missing data.

Each RATAS tends to be broad with a large number of children and not too many levels of dependencies. But mining the whole ATAS in order to get the Full Tag Network for the whole Story-Set is not only be computationally expensive and time-consuming, but extremely large (any visualization and analysis will consist of hundreds of thousands of nodes and edges). For this reason, the demonstration of the techniques will be limited to the *Disability* tag.

3.3.3.5 Testing/Validation

For testing this module, different small RATAS were mined from the archive. These were then compared to the AO3 website data (in html representation) to check that the hierarchical structure was reproduced exactly. The output of the module matched as the structure was vision the hierarchical layout of the RATAS. Additional spot-checks were performed. This is an admittedly heuristic approach. A more robust testing process is suggested in the future work section.

3.4 Phase 3: Network Building and Visualizing

The objective of this phase was set out at the beginning of this chapter:

Objective (v) **Network Building and Visualizing:** Develop computational techniques to build the static network structures from the data that gets scraped using the computational techniques as developed to satisfy objective (iv), Develop computational techniques to visualize the network structures.

3.4.1 Approach

The workflow of this module is shown in Figure 3.3 below. It consists of two main modules which make use of the scraped data from the previous step (described in section 3.3). The modules build all versions of the RATAS, as defined in equations (eqn:3.2.1), (eqn:3.2.3), (eqn:3.2.4), and the Connected Story-Set & RATAS networks, as defined in equations (eqn:3.2.11), (eqn:3.2.12), (eqn:3.2.10), and (eqn:3.2.13). These modules are: "Building Computational Networks" and "Visualizing Networks".

The module "Visualizing Networks" depends on the "Building Computational Networks" to create the visualizations of the networks built while "Building Computational Networks" constantly reads the scraped data to build the network representations. This workflow is as follows:

(i) **The "Building Computational Networks" module**: builds the different network structures, as defined in equations (eqn:3.2.1), (eqn:3.2.3), (eqn:3.2.4), from the scraped data in the previous phase. These networks are either subjected to a subsequent analysis module or visualized (as the main instruction of the program).

(ii) **The "Visualizing Networks" module**: receives the computational network representation built by the "Building Computational Networks" module to generate its proper visualizations highlighting different properties as the main instruction of `ao3_analyser` sent.

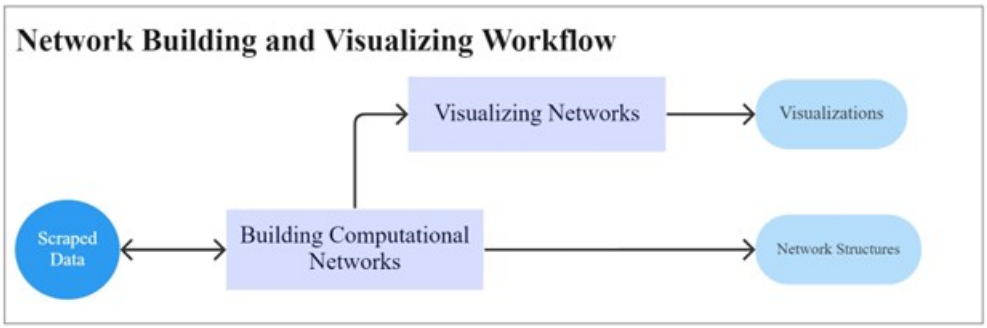


Figure 3.3: Schematic that provides an overview of the network building and visualizing modules Workflow. The schematic illustrates the flow of data between each of two modules and the data generated in Phase 1 3.3, as necessary to build all network structures either for their visualization of analysis in the upcomming phase 3.5. There are two modules: "Building Computational Networks", and "Visualizing Networks".

3.4.2 Building Computational Networks

3.4.2.1 Requirements

1. **Core Functionality**: A module which builds the following networks (which were defined in section 3.2):
   (a) Basic t-RATAS (for various values of t), as per equation (eqn:3.2.1)
   (b) Expanded t-RATAS (for various values of t)as per equation (eqn:3.2.3)
   (c) Fully Expanded t-RATAS (for various values of t)as per equation (eqn:3.2.4)
   (d) Story-Sets Connected with t-RATAS (for various combinations of values of t and stories), Free Tag Network, as per equation (eqn:3.2.10)
   (e) Story-Sets Connected with t-RATAS (for various combinations of values of t and stories), Canonical Tag Network, as per equation (eqn:3.2.11)
   (f) Story-Sets Connected with t-RATAS (for various combinations of values of t and stories), Synonymized Tag Network, as per equation (eqn:3.2.12)
   (g) Story-Sets Connected with t-RATAS (for various combinations of values of t and stories), Full Tag Network, as per equation (eqn:3.2.13)
2. **Input**: the scraped data as generated in phase 3.3.

3. **Output**: Computational network instantiations of the specified structures from section 3.2.

4. **Other considerations**: When building the networks, they must provide the main properties for their analysis and visualizations with a generic instantiation for their visualization and analysis.

3.4.2.2 Development

The functionality for the Building Computational Network module was implemented as part of the `ao3_analyser` tool, via the `NetBuilder` flag. The module's functionality is accessed indirectly through the "Visualizing Networks" or "Network Analysis" modules.

The `NetBuilder` was developed in `Python` using the library `NetworkX` for the instantiation of the network representations and their further analysis. The library `NetworkX` was chosen because it "is one of the most widely used networks or graph programming packages in python. `NetworkX` is the reference library for Python's network algorithms" (Hossen et al., 2022).

The module uses the library `json` for reading the scraped data with the RATAS information (as generated by the `TagScra` module). It also uses `pandas` for reading and managing the Story-Sets dataset, which gets scraped using the `DaScra` module.

`NetBuilder` builds and populates each network as an instance of a `Graph` data structure (as provided by the `NetworkX` library). The `Graph` class is convenient because of the facilities of the `NetBuilder` library for managing networks, as well as the integration of `NetBuilder` with other different libraries for visualizing and analyzing networks. It builds each different network using the proper inputs as follows:

To create and populate a Basic, Expanded, and Full Expanded RATAS:

The module receives a `JSON` file with the structure of the RATAS, as defined in section 3.3.3. This file structure will be process by the same method for any version of the RATAS.

An instance of a `Graph` data structure (as provided by the `NetworkX` library) is generated. Each node has its classification depending on the type of tag. The canonical nodes have attached a connection node which is connected to all syn nodes belonging to its syn-set in case the syn-set is conform with more nodes than the trivial element (the canonical tag itself). In Algorithm 1, the pseudo-code for building the network is provided.

Algorithm 1 RATAS Generation

Require: `JSON` RATAS file
  $ratas\_json \leftarrow$ read `JSON`
  $network \leftarrow$ new NetworkX DiGraph
  for $t \leftarrow key\_tag$ in $ratas\_json$ **do**
    $network \leftarrow t$, **if** $t \notin network$
    $network[t].type \leftarrow t.node\_type$
    if $t.type == canonical$ **then**
      $network[t] \leftarrow ratas\_json[t].metatags$ & t-incoming edges
      $network[t] \leftarrow ratas\_json[t].sub\text{-}network\ structures$ & t-outgoing edges
    else
      if $t.type == syn$ **then** ▷ Set t to its canonical tag Syn-Set
        $network[t].canonical.syn\_set \leftarrow t$
      end if
      if $t.type == free\text{-}form$ **then** ▷ Set t as isolated node
        $network \leftarrow t$
      end if
    end if
    Set $network[t] \leftarrow \{type, title, group, color\}$ ▷ Set node's properties
  end for

To create and populate a Story-Set & RATAS networks (all variants):

The module receives a `JSON` file that contains the structure of the RATAS (as defined in section 3.3.3) and receives a `xlsx` file that contains a Story-Set (as defined in section 3.3.2). Then the module builds the RATAS (as previously mentioned) and loads the provided Story-Set, in order to start building the network (as defined in equations (eqn:3.2.10), (eqn:3.2.11), (eqn:3.2.12), or (eqn:3.2.13), in section 3.2.3.2).

An instance of a `Graph` data structure (as provided by the `NetworkX` library) is generated by the module. Each node has its classification depending on the type of tag. As per section 2.1.4.1 defines the syn-set, the computational network structure handled them as follows: the canonical nodes are connected to a special syn-node. Then the syn-node gets is connected to tag nodes using the syn relationship. In Algorithm 1, the pseudo-code for building the network is provided.

Similarly, for the creation of Story-Sets & RATAS networks, the module iterates through all stories, creating the nodes and connecting them if they share an specific type of tag from the Additional Tag category, as defined in section 3.2.3.2. Depending on the type of the tag in the RATAS, each of the different four RATAS & Story-Set networks is built. In Algorithm 2,

the pseudocode for building the network is provided.

Algorithm 2 Story-Set & RATAS Networks Generation

Require: `JSON` RATAS file
Require: `xlsx` Story-Set file
  $ratas\_json \leftarrow$ read `JSON`
  $story\_set \leftarrow$ read `Story-Set.xlsx`
  $network \leftarrow$ new NetworkX DiGraph
  $network.nodes \leftarrow S_{story\_set}$ ▷ Story-Set $S = \{s_1, \ldots, s_i\}$
  for $t \leftarrow additional\_tag$ in $T_{story_{set}}$ **do** ▷ Set of Additional Tags in the Story-Set
    if $t \in T_{s_i}$ & $t \in T_{s_j}$ **then**
      $network \leftarrow \langle s_i, s_j \rangle$ ▷ Depending t node type int the RATAS the network we are building
    end if
  end for

3.4.2.3 Testing/Validation

The module was tested for each type of network using the datasets obtained for the previous testing processes (section 3.3.2.4, and section 3.3.3.5). Each of the files was used as input to the module and the networks were built and visualized using `NetworkX` visualization features. Then, a visual comparison with the actual structures in the AO3 website was performed to ensure its correctness. With this, we can conclude that creating and instantiating the networks based on these files is correct. This is an admittedly heuristic approach. A more robust testing process is suggested in the future work section.

3.4.3 Generating Visualizations

3.4.3.1 Requirements

1. **Core Functionality**: To visualize the `Graph` instances, as generated by The Building Computational Network module.

2. **Input**: A `Graph` instance.

3. **Output**: Dynamic screen-based visualizations highlighting the different properties of the network.

4. **Other considerations**: The module should function consistently over all of the different network types. There are many different types of networks, and for simplicity, only one visualization module will be developed.

3.4.3.2 Development

The functionality for the Generating Visualizations module was implemented as part of the `ao3_analyser` tool, via the `NetVisualizer` flag. The module's functionality is accessed via the main CLI of `ao3_analyser` through several commands, depending the network to visualize:

- `--GenerateNetwork JSON` or `-gn JSON` to generate the RATAS visualizations. The module receives a RATAS `JSON` file as input.
- To generate the Story-Set & RATAS networks, The module receives a Story-Set `xslx` dataset and a RATAS `JSON` file with the commands:
  - `--GenerateFreeFormTag XLSX JSON` to generate the Free-Form Tag network (as defined in (eqn:3.2.10))
  - `--GenerateSynTag XLSX JSON` to generate the Syn Tag network (as defined in (eqn:3.2.12))
  - `--GenerateCanonicalTag XLSX JSON` to generate the Canonical Tag network (as defined in (eqn:3.2.11))
  - `--GenerateFullTag XLSX JSON` to generate the Full Tag network (as defined in (eqn:3.2.13))

The module `NetVisualizer` was developed in `Python` using `Pyvis` as the main library for visualization purposes. `NetworkX` integrates with `Pyvis` easily, which makes the network visualization creation in `Pyvis` without any modification. It also recognize the properties in the `Graph` instances and displays them without need to modify them in the visualization manually.

The module is also accessed during the Network Analysis Phase module to generate different visualizations. This will be discussed in section 3.5 next.

For any of the invocations, the module generates an `html` file, as per the `Pyvis` library. The file offers a dynamic visualization of the network, meaning that you can hover the nodes to visualize their main properties (in/out degree, tag title), select and move around the nodes,

zoom-in and zoom-out. Depending on which visualization was generated, different nodes and edges will be highlighted differently.

To visualize a Basic RATAS, Expanded RATAS or Fully Expanded RATAS:

The static visualization of a RATAS, the module is accessed through the main CLI with the argument `--GenerateNetwork JSON` (or `-gn JSON` as a simplified form). It receives one argument which is a file path for a `JSON` file with the RATAS to be visualized.

The main differences among the visualizations of three variants were described in detail in section 3.3.3. In summary, the main difference between the Basic t-RATAS and the Expanded t-RATAS concerns the appearance of the synsets of the canonical nodes and the potential presence of new metatags (i.e., those that did not belong to the Basic RATAS). The Full Expanded RATAS will vary from the Expanded t-RATAS because of the appearance of the isolated nodes (representing free-form tags).

Visualization Design for Basic RATAS, Expanded RATAS or Fully Expanded RATAS:

The choice of colours and node shapes for these visualizations are as listed in table 3.2.

Legend: Node Colours/Shapes for RATAS Visualizations

Basic RATAS	
Root Tag (Focus Node):	round, blue
Direct Successor of Root:	round, orange
Non-Direct Successor of Root:	round, yellow
Syn-set of Focus Node:	square, magenta
Syn Tag Nodes:	triangle, magenta

Expanded RATAS , Fully Expanded RATAS	
Root Tag (Focus Node):	round, blue
Direct Children of Root:	round, blue
Non-Direct Child of Root:	round, blue
Metatag no related to the Root:	round, red
Syn-set of Focus Node:	square, magenta
Synned Tag Nodes:	triangle, magenta
Free-tags (associated with root)*:	round, grey, disconnected

**these nodes only appear in the Fully Expanded RATAs, not the Expanded RATAS*

Table 3.2: Legend of colours and shapes for nodes in the RATAS visualizations.

To visualize a Story-Set Connected with t-RATAS (all variants):

To generate the visualizations of a Story-Set & RATAS networks, the module is accessed by either: `--GenerateFreeFormTag XLSX JSON`, `--GenerateCanonicalTag XLSX JSON`, `--GenerateSynTag XLSX JSON`, or `--GenerateFullTag XLSX JSON` for the different networks. These visualizations uses all circular node shapes, and differentiate the edges in the Full Tag Network as listed in Table 3.3 depending the type of tag used.

3.4.3.3 Limitations

The visualizations can become very crowded when displaying large networks. Even though the dynamism of the visualization allows the user to focus on certain areas, this feature in some cases is not adequate.

The selection of a visualization library comes with pros and cons. `Pyvis` has great integration with `NetworkX` and has flexibility with network layout, but lacks in certain

Legend: Node Colours/Shapes for Story-Set Connected with t-RATAS Visualizations

Full Tag Network	
Syn Tag Network edge:	green
Canonical Tag Network edges:	red
Free-Form Tag Network edges:	blue

Table 3.3: Legend of colours and shapes for nodes in the Story-Set Connected with t-RATAS visualizations.

aspects. For instance, it does not implement legend generation, which is needed to easily signal the different aspects of the visualization (the legends in the Figures in this thesis were hand-built). The loading time is also significant, and is load proportionate with the size of the network. Also, there is the vulnerablility that it will become obsolete, which will imply a future re-structuring of the whole module should `Pyvis` no longer be functional (e.g., through lack of maintaince).

3.5 Phase 4: Network Analysis

The purpose of this last phase is to develop network analysis techniques as per the objectives set out in table 3.1. This includes: objective (vi), which is to extract quantitative characteristics of static network structures, objective (vii), which is to develop a technique to analyze the differences between two static networks, and objective (viii), which is to develop a technique to analyze a sequence of networks.

3.5.1 Static Analysis

Objective (vi) was to develop network analysis techniques of the static network structures to derive quantitative characteristics that would be useful for fan-work scholars. In anticipation of their needs, we decided to generate reports with a large variety of quantitative characteristics. We recognize that characteristics of the t-RATAS are likely to be of interest.. The structures of the Story-Set & RATAS, as defined in (eqn:3.2.10), (eqn:3.2.11), (eqn:3.2.12), (eqn:3.2.13), bring further relevant information to report. Because of the scope of the study, we have left the study depth of these structures to future work.

Below we list the properties to be included in the analysis report. These properties were selected because of their potential interest for scholars looking at a static t-RATAS. Each

property is accompanied by a 'descriptor', which is used in the output reports. Next to each characteristic, there is an indication of the t-RATAS variant(s) that is/are relevant (e.g., Basic, Expanded, or Fully-Expanded). There is also an indication of whether the properties of t-RATAS were obtained treating it as a undirected or directed network.

(i) **Total number of nodes**. Descriptor: total_nodes

(ii) **Total number of edges**. Descriptor: total_edges

(iii) **Total number of canonical/syn/free-form nodes**. Descriptor: total_canonical, total_syn, total_freeform

(iv) **Total number of non-trivial Syn-Sets** as defined in section 2.1.4.7. Descriptor: total_canonical_syn_connection_nodes

(v) **Total number of canonical nodes with a relatively large out-degree**. Descriptor: big_nodes

(vi) **List of canonical nodes with a relatively large out-degree**; and their respective in/out-degree. Descriptor of worksheet: large_outdegree_node

(vii) **List of canonical nodes with a relatively large in-degree**; and their respective in-degree and list of predecessors. Descriptor of worksheet: large_indegree_node

(viii) **List of canonical nodes with syn-sets**; and their respective syn-sets. Descriptor of worksheet: canonical_node_with_synset

(ix) **Whether the RATAS is strongly or weakly connected**. Descriptor: is_strongly_-connected, is_weakly_connected

(x) **The diameter and radius**. Descriptor: diameter_directed(undirected), radius_-directed(undirected)

(xi) **Precense of cycles and the nodes involves**. Descriptor: has_cycles

(xii) **Node degree (per node)**; properties of the distribution of values over all nodes (e.g., power-law distribution).

(xiii) **Center of the RATAS**. Descriptor: center_RATAS_directed(undirected)

(xiv) **Numbers k for which the network has different components of k-connectivity**. Descriptor: k_components

3.5.1.1 Development

A module was developed for `NetAnalyser` which contains different methods to obtain the properties described above. This module was implemented mainly in `Python` using `NertowrkX` as the main library.

In order to get the static properties of a RATAS, the command `--GenerateRATASProperties ARG` is used, which receives as an argument the path of a `JSON` RATAS file. The obtained properties are saved in a `xlsx` file with five different sheets as follows: (I) general network information (general_information), (II) network properties (net_properties), (III) list of all 'large' canonical nodes (large_outdegree_node), (IV) list of all syn-sets (canonical_nodes_with_synset), and (V) list of all canonical nodes with multiple successors (large_indegree_node). These are illustrated in Table C.1 below.A complete example for each worksheet is provided in Appendix C.1.

Worksheet (I) Template: General Network Information
big_nodes
total_nodes
total_edges
total_canonical_leafs
total_freeform
total_syn
total_canonical
total_canonical_syn_connection_nodes

Worksheet (II) Template: Network Properties
is_strongly_connected
is_weakly_connected
diameter_directed(undirected)
radius_directed(undirected)
center_RATAS_directed(undirected)
has_cycles
k_components

Worksheet (III) Template:
one row per each large_outdegree_node: tag string; out_degree; in_degree

Worksheet (IV) Template:
one row per each canonical_node_with_synset: tag string; Count; SynTags

Worksheet (V) Template:
one row per each large_indegree_node: tag string; in_degree; Metatags

Table 3.4: Template for Sheets I, II, III, IV, and V in the Static Analysis Report.

3.5.2 Diffing

This section specifies the network analysis techniques developed to detect and analyze the differences between two static versions of the same t-RATAS.

As described in objective (ii), this technique will be used as a way to characterize network dynamics in terms of changes over time to the network, such as a result of tag wrangling or free-tagging by a fan-fic contributor.

We seek to express the differences between two networks in terms of the different possible contributor and wrangler actions, as developed under objective (ii).

3.5.2.1 Requirements

1. **Core Functionality**: A module which develop network analysis techniques to detect and analyze the differences between two static versions of the same t-RATAS
2. **Input**: It should receive two different snapshots in time of the same t-RATAS
3. **Output**: Visualizations of the differences and similarities between the two snapshots received.
4. **Other considerations**: When performing the techniques to analyse the differences, it should consider the main properties of the network structure and highlight nodes of interest that can lead to actions of creation.

3.5.2.2 Development

The module `NetAnalyser` contains, additionally, the methods to obtain the differences between two snapshots of a RATAS mainly focussing on the presence of each additional tag. It is accessed by the main CLI for each specific analysis. It is accessed with the command `--GenTagDiff ARG1 ARG2` which receives as argument two paths of the `JSON` RATAS files to compare. This module was implemented mainly in `Python` using `NetworkX` as the main library. After obtaining the information, it is saved in a `CSV` file.

The generated file stores all the tags present in each network, whether they are present in each network, and if so, their classification. This allows us to compute and to store the political acts of wrangling that occurred from one snapshot to another (e.g., synonymizing, canonizing, creation of new free-form tag, or updating or deletion of existing tags as defined

in section 3.2.4). An example of the generated report is shown in Appendix C.1.1 section, Table C.6.

The module also computes multiple pairwise-comparisons, namely the tag differences between multiple snapshots of the same RATAS against the most recent snapshot. This functionality can be accessed by the command `--GenTagDiffDir ARG` which receives one argument as the directory where all `JSON` RATAS are stored. Each version is compared with its subsequent one in the directory to capture the evolution over time. An example of the generated report is shown in Appendix C.1.1 section, Table C.8.

Given two snapshots of the same RATAS, let's assume that V_{older} is the older snapshot and V_{newer} is the newer. The difference between (V_{older}, V_{newer}) is defined as the set of nodes and edges in V_{older} that are not in V_{newer} (i.e., the intersection of nodes and edges between both networks is subtracted from the first network). Computing diff(V_{older}, V_{newer}) results in the identification of all the elements that were lost over time (e.g., nodes from V_{older} that were not kept in V_{newer}). On the other hand, diff(V_{newer}, V_{older}) results in identification of all the new nodes and edges (e.g., new in V_{newer}, as in not present in V_{older}).

The similarities between two networks (G, H) is defined as the intersection of both set of nodes and edges. Then we define similarity (G, H) as the intersection of nodes in both networks $(V(G) \cap V(H))$ and the intersection of edges $(E(G) \cap E(H))$.

This module is also accessed in combination with `NetVisualizer` module, when deriving the visualization of the differences and similarities between two snapshots of a RATAS. The visualization highlights the differences between two RATAS and is accessed through the main `CLI` with the argument `--GenerateDifferences ARG1 ARG2` or `-gd ARG1 ARG2` as is simplified form. It receives two arguments, one for each file path of the `JSON` files to compare. Similarly, to get the similarities of two RATAS, the argument `--GenerateSimilarities ARG1 ARG2` (or `-gs ARG1 ARG2` as is simplified form) can be used. It receives two arguments, one for each file path of the `JSON` files to compare.

After calling the module to compute the differences between two RATAS snapshots, two visualizations are generated one for each difference (V_{older}, V_{newer}) and (V_{newer}, V_{older}).

In these visualization, the common nodes and edges (the ones which will be removed) are highlighted with red and the ones that will remain after the subtraction are shown in blue. This is shown in two visualizations, difference (V_{older}, V_{newer}) is illustrated on V_{older} structure while difference (V_{newer}, V_{older}) is visualized on V_{newer} nodes and edges.

When the base is an earlier version of the t-RATAS, the visualization illustrates the future direction of the t-RATAS in terms of what will be added and taken away from the specified

starting point.

When the base is an later version of the t-RATAS, the visualization illustrates the elements of the network that have been stable over the specified time period.

Visualization Design for Snapshot Differences and Snapshot Similarities:

The choice of colours for these visualizations are as listed in Table 3.5.

Legend: Node/Edge Colours for the Snapshot Differences and Snapshot Similarities Visualizations

Snapshot Differences, Diff(V_{older}, V_{newer})	
nodes/edges present in V_{older} that are **still** present in V_{newer}:	red
nodes/edges present in V_{older} and are **no longer** present in V_{newer}:	blue

Snapshot Differences, Diff(V_{newer}, V_{older})	
nodes/edges present in V_{newer} that were **also** present in V_{older}:	red
nodes/edges present in V_{newer} and were **not also** present in V_{older}:	blue

Snapshot Similarities
unique colours are used to identify the nodes that are common to both snapshots only the nodes with a sufficiently high out-degree are shown

Table 3.5: Legend of Node/Edge Colours for the Snapshot Differences and Snapshot Similarities Visualizations.

When generating the similarities of two networks, each visualization uses common color to highlight all corresponding nodes with an out-degree greater than 5. The colour gray is used for the rest of the nodes. This helps to highlight easily the bigger nodes that are shared across both visualizations.

3.5.2.3 Limitations

These sections are limited by a number of properties within a network, as well as the main differences between two or more networks. On-going interaction with scholars studying the AO3 wrangling process will enrich our understanding of the salient differences and will produce more meaningful interpretations and further analyses.

3.5.3 Diffing in Sequence

This section focuses on analyzing and identifying possible interactions between contributors and wranglers mediated by the RATAS and the RATAS story-sets. It strives to develop techniques to spot acts of collaboration/co-creation between wranglers and contributors in the different structures. It also aims to quantify basic properties of the story-set in relationship with a RATAS.

For this section, instead of implementing any specific module, hand analysis will be employed. A case study focusing on the Disability-RATAS, in relation to its set-of stories, will be used. The results will be discussed in detail in the next chapter.

The approach will first explore the Disability-Story-Set and the additional tags in it, analyzing the numbers and the frequency of usage of different tags. By analyzing the stories with the RATAS, we can create new columns containing the Disability-related tags from the Additional Tags in the Story-Set and then classify them into canonical, syn, or free-form tags. We can get the counts and relative frequencies of the various types of Disability-related tags and usages, based on the Disability RATAS. In addition, based on the all snapshots of the Disability-RATAS, we can study in detail the Story-Set over the years and determine any interesting quantitative results as the usage over the years within the Story-Set gets scraped. Finally, we can focus on finding, over time, the different behaviours that show co-creation in the archive structure such as:

- Typo correction
- Syn identification
- Canonized or something
- Deletion from the contributors
- Update from the contributors

We can also use the methodology described in section 3.5.2 and analyze all different snapshots of the Disability-RATAS. The analysis of these political acts on a defined story-set helps to spot several tags which might reflect co-creation and collaboration between wranglers and creators in AO3.

3.5.3.1 Limitations

This type of analysis will be restrained by the nature of scraped story-set and RATAS. As mentioned in section 3.3.2.3, AO3 tends to evolve quickly due to the high volume of collaborations and contributions by the community. This is also reflected in the story-sets, which are difficult the capture, creating further challenges in trying to spot political acts using the story-sets and the RATAS.

3.6 Chapter Summary

This chapter provided an overview of the work performed in order to achieve eight objectives. The list of objectives were developed in order to create a methodology to develop answers to the research questions identified at the outset. In summary, formal definitions were developed for all the representations of the networks and story sets needed to study AO3 (including RATAS and Story-Set & RATAS networks). Next, formalisms were develop to describe the fan-fic contributor and wrangling processes. This involved enumerating the actions that can be performed by contributors and wranglers on the different structures defined and related these network operations to the political aspects they represent (both as individual aspects and aspects of co-creation). Then, the development of a tool was described that scrapes the data from AO3, builds the RATAS and Story-Set & RATAS networks, generates their visualizations and analyses them. For the analysis, three main aspects were described: the analysis of a single RATAS as a static representation, the analysis of the different snapshots of the same RATAS, and the analysis of a sequence of them in relationship with the story-set (looking for acts of co-creation). We can summarize the culmination of these steps as a ready-to-use tool that obtains quantitative insights from the AO3 website, generates visualizations of various tag structures, and generates various results from their analysis.

Chapter 4

Findings

This chapter describes the answers to the research questions posed at the end of Chapter 2. In generating these answers, the methodology described in Chapter 3 was followed. As part of the methodology, we have focused on the study of how the concept of *Disability* is tagged within AO3. As discussed in Chapter 2, the study of the *Disability* tag in the AO3 tagging architecture has been of particular interest to scholars in other fields. We focused on this concept to demonstrate the results of using the computational approaches and tools for the research questions listed below in each section of this chapter:

Q1 (a) How to build t-RATAS and t-Story-Set for different tags t from the AO3 interfaces?

(b) How can the actions that contributors, and wranglers perform in AO3 be represented in terms of operations on the t-RATAS network representation?

(c) Which structural properties can be extracted and visualized from the t-RATAS structures? What can these properties tell us about the AO3 'Additional Tags' category (at a given point in time)?

Q2 (a) How can we capture a snapshot of the t-RATAS at different points in time?

(b) Given two t-RATAS snapshots, which technique(s) can be developed to compare them with one another, in order to express change over time?

(c) What can t-RATAS tell us about the major changes to the AO3 tag structure over the past years of activity?

Q3 (a) Which actions by fan-fic contributors impact the wranglers (and, conversely, which actions by wranglers impact fan-fic contributors) and how can these actions be illustrated through operations on the network?

(b) Which technique(s) could be used to detect bilateral paths of influence that may have occurred over the past years of AO3 activity?

4.1 Findings: RATAS and their Structural Properties

In this section, we answer the first cluster of the three research question clusters. We focus on the following:

> Q1c: Which structural properties can be extracted and visualized from the t-RATAS structures? What can these properties tell us about the AO3 'Additional Tags' category (at a given point in time)?

Answering this question requires us to complete two preliminary steps. First, we must be able to scrape the necessary data to build a t-RATAS and we must be able to build its computational network structure, as covered in objective (iv) and objective (v) stated in Section 3.3 and Section 3.4 respectively. This corresponds to question Q1a below and is the topic of section 4.1.1.

Q1a: How to build t-RATAS and t-Story-Set for different tags t from the AO3 interfaces?

Then, we will be in position to perform a network analysis of a t-RATAS, analyzing its structure and properties as covered in objective (vi). This corresponds to question q1b below and is the topic of section 4.1.2.

Q1b: How can the actions that contributors, and wranglers perform in AO3 be represented in terms of operations on the t-RATAS network representation?

Once these two steps are completed, the results can be synthesized and an overall answer can be provided (section 4.1.3).

4.1.1 Scraping Data, Creating and Populating the Networks

Q1a: How to build t-RATAS and t-Story-Set for different tags t from the AO3 interfaces?

4.1.1.1 Scraping Data for the t-RATAS Networks

To build a t-RATAS from the AO3 Additional Tag information, we used the technique developed and described in section 3.3.3. This module `TagScra` was used to generates all data needed to construct the Basic, Expanded, and Fully Expanded Disability-RATAS. For each of the three network variants, we need to invoke slightly different variants of the tool. Each invocation is described below.

4.1.1.2 Constructing and Visualizing the Basic Disability-RATAS

To construct the **Basic Disability-RATAS**, we invoked `TagScra` as described in section 3.3.3 as follows:

```
> python .\ao3_analyser.py -ts Disability -o rata_disability
```

In this invocation, the parameter `-ts` invokes the module `TagScra` with the single tag *Disability* (although it could have been provided with a list of tags). Then, the `-o` parameter is provided with the name specification for the `JSON` file which will used to store the scraped data that will be generated. We next used the resultant `JSON` in the next step.

To construct and visualize the Basic t-RATAS, we used the `NetVisualizer` module defined and passed it as an argument the `rata\_disability.json` file generated using the prompt:

```
> python .\ao3_analyser.py -gn rata_disability.json
```

The result of this invocation is the visualization shown in Figure 4.1. This visualization shows the Basic Disability-RATAS, which has the *Disability* tag as its root (the focus tag, shown in blue). It contains only the child tags of the focus tag. The focus node is the only node that has its syn-set displayed.

The choice of colours and node shapes for the visualizations was listed in Table 3.2, and is repeated here (Table 4.1) for convenience:

Legend: Node Colours/Shapes for Basic RATAS Visualizations

Root Tag (Disability):	round, blue
Direct Successor of Root:	round, orange
Non-Direct Successor of Root:	round, yellow
Syn-set of Focus Node:	square, magenta
Syn Tag Nodes:	triangle, magenta

Table 4.1: Legend of colours and shapes for nodes in the RATAS visualizations.

The visualization in Figure 4.1 illustrates the actions performed by wranglers and the structural properties of the Basic Disability-RATAS network, which are discussed in more detail in section 4.1.2 and section 4.1.3.

Comprehension of the visualization of a basic t-RATAS is not immediate, particularly for those most familiar with the AO3 website representation of the tag structure. The two views — the AO3 website view and the network visualization — are shown in alignment in Figure 4.2. These network visualization can be hard to read due to the large number of nodes. This is illustrated in Figure 4.2(a), which shows only a part of the AO3 hierarchy for the *Disability* tag and in Figure 4.2(b), which is network visualization counterpart, shown in whole (which is composed by more than 500 nodes).

Figure 4.2 also illustrates the concept of information density (i.e., the amount of information per unit of space). In the same amount of space, Figure 4.2(a) shows a only portion of the AO3 webpage display of the subtags of *Disability*, whereas Figure 4.2(b) shows the entire tag structure.

The network visualization in Figure 4.2(b) affords interactivity, such as zoom-in/out and 'tool-tip' display, meaning that when the user hovers over a node, a textbox appears with information such as the name of the node, its connections and general properties like the type of node, and its syn-set.

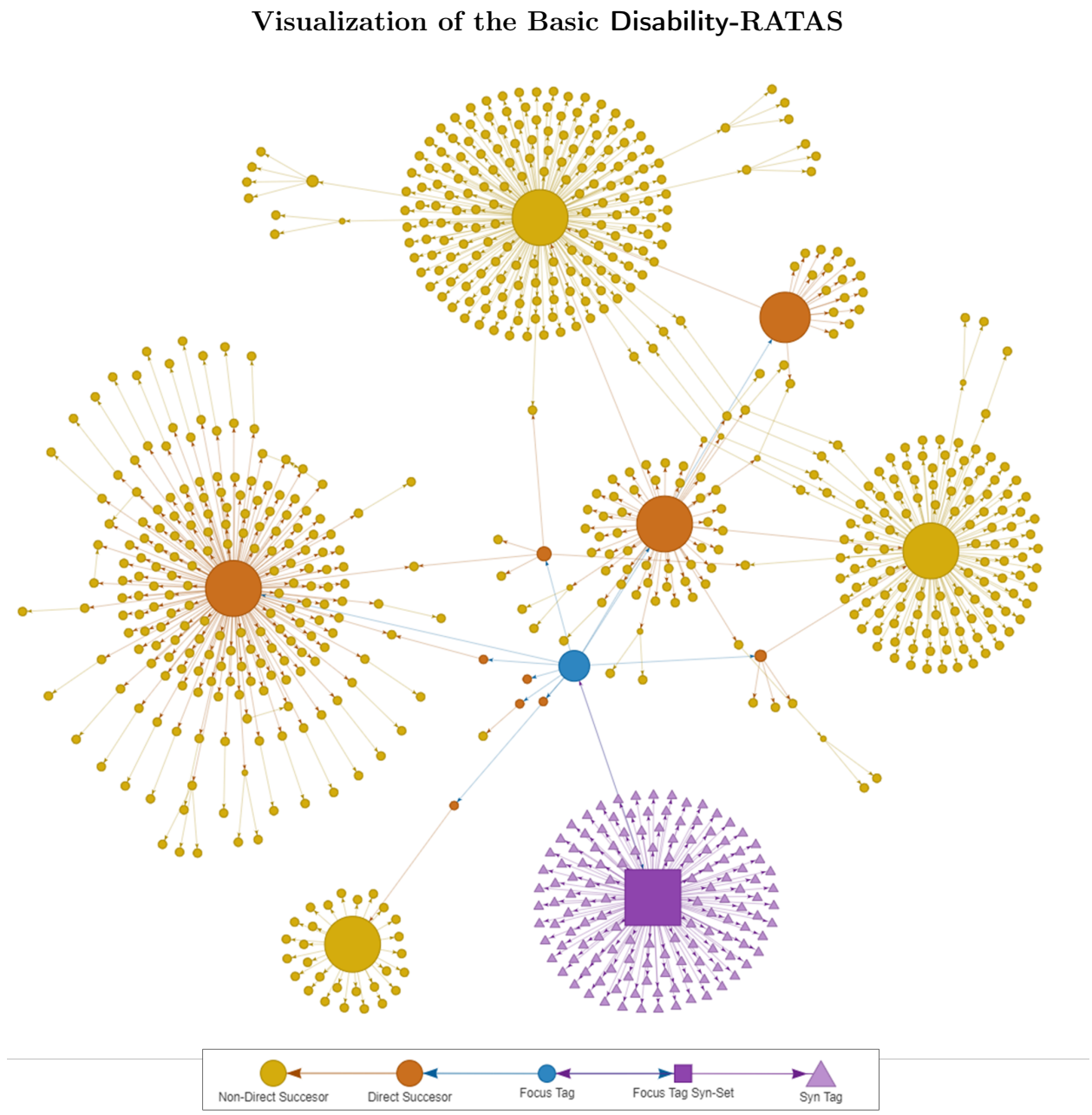
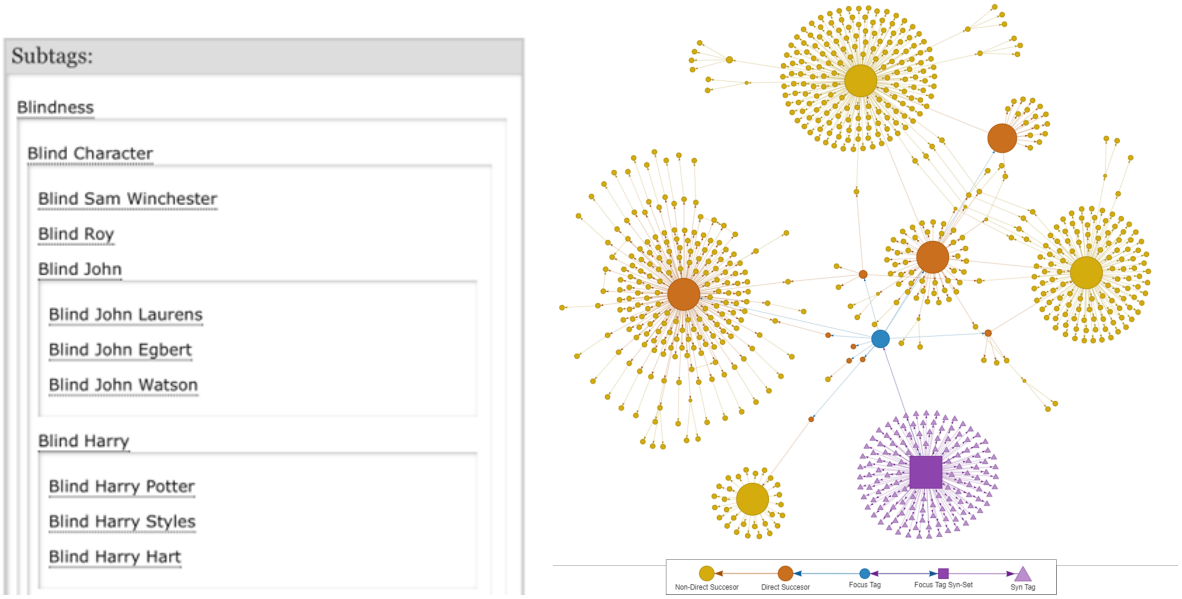


Figure 4.1: Visualization of the Basic Disability-RATAS. The tag Disability is the root of the network (the focus tag, shown as a circular blue node). The full legend is provided in Table 4.1.



(a) (b)

Figure 4.2: (a) A part of the hierarchy for the *Disability* on the AO3 website; (b) Disability-RATAS network visualization.

4.1.1.3 Constructing and Visualizing the Expanded Disability-RATAS

In order to expand the Basic Disability-RATAS, we scrape additional data through the use of a further invocation of `TagScra`:

```
> python .\ao3_analyser.py --ExpandRATAS .\rata_disability.json
    -o rata_disability_expanded
```

Note that the module requires as input the data for the Basic t-RATAS. The module will scrape the syn-sets and metatags for all of the canonical tags as in the Basic variant. The parameter `--ExpandRATAS` indicates this to `TagScra`. This invocation generates another `JSON` file, `rata_disability_expanded.json` as the output containing the additional information about the syn-sets.

To visualize the Expanded Disability-RATAS, we call the `NetVisualizer` module in a similar way as the example before.

```
>python .\ao3_analyser.py  -gn  rata_disability_expanded.json
```

Figure 4.3 shows the Expanded Disability-RATAS, which is an expanded version of Figure 4.1. The choice of colours and node shapes for the visualizations was listed in Table 3.2, and is repeated here (Table 4.2) for convenience:

Legend: Node Colours/Shapes for Expanded RATAS Visualizations

Root Tag (Disability):	round, blue
Direct Children of Root:	round, blue
Non-Direct Child of Root:	round, blue
Metatag not related to the Root:	round, red
Syn-set of Focus Node:	square, magenta
synonymized Tag Nodes:	triangle, magenta

Table 4.2: Legend of colours and shapes for nodes in the RATAS visualizations.

The visualization in Figure 4.3 illustrates the actions performed by wranglers and the structural properties of the Expanded Disability-RATAS network, which are discussed in more detail in section 4.1.2 and section 4.1.3.

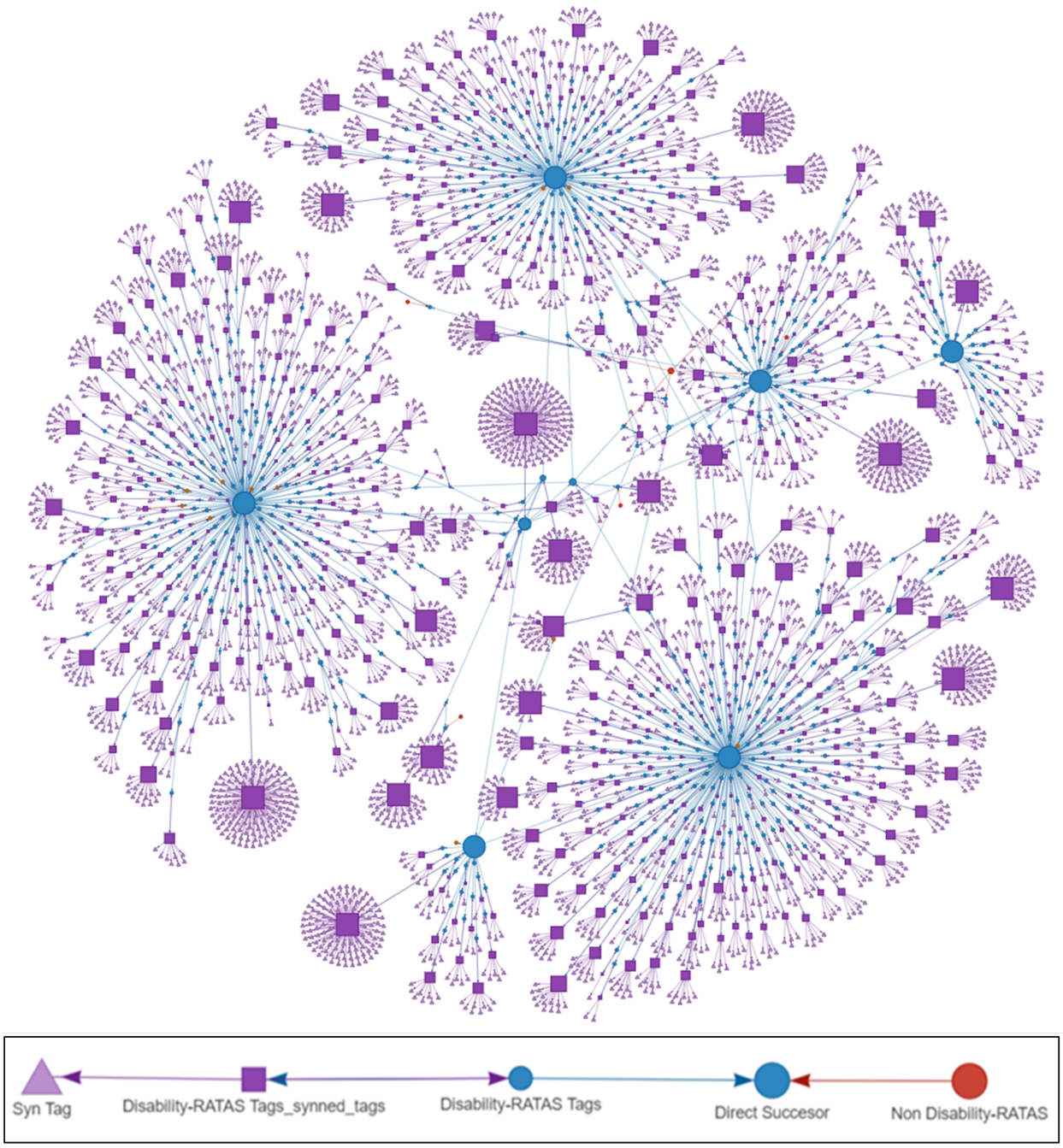


Figure 4.3: Visualization of the Expanded Disability-RATAS. The root of the network is the tag Disability. The full legend is provided in Table 4.2.

4.1.1.4 Constructing and Visualizing the Fully Expanded Disability-RATAS

Finally, in order to construct the **Fully Expanded Disability-RATAS**, we need to additionally scrape for all of the non-canonical tags related to *Disability* in AO3.

To do this, we use the `TagScra` module using the following invocation:

```
> python .\ao3_analyser.py -ts Disability --NonCanonical
    -o non_canonical_disability
```

Then we combine the obtained file with the file from the Expanded Disability-RATAS file to generate a fully expand version. We use the `ExpandFullRATAS` argument with the `TagScra` module as follows:

```
> python .\ao3_analyser.py --ExpandFullRATAS rata_disability_expanded.json
    non_canonical_disability.txt  -o rata_disability_current_expanded_full
```

Then the visualization can be generated via an invocation of `NetVisualizer` to generate the visualization.

```
> python .\ao3_analyser.py -gn rata_disability_current_expanded_full.json
```

The visualization of this fully expanded version is shown in Figure 4.4. The visualization is quite similar to the visualization of the Expanded Disability-RATAS in Figure 4.3, but for the addition of the free-form nodes that are linked to Disability. These free-form nodes are shown as isolated nodes. The choice of colours and node shapes for the visualizations was listed in Table 3.2, and is repeated here (Table 4.3) for convenience:

Legend: Node Colours/Shapes for Fully Expanded RATAS Visualizations

Root Tag (Disability):	round, blue
Direct Children of Root:	round, blue
Non-Direct Child of Root:	round, blue
Metatag not related to the Root:	round, red
Syn-set of Focus Node:	square, magenta
synonymized Tag Nodes:	triangle, magenta
Free-tags (associated with root):	round, grey, disconnected

Table 4.3: Legend of colours and shapes for nodes in the RATAS visualizations.

The visualization in Figure 4.4 illustrates the actions performed by wranglers and users and also illustrates the structural properties of the Expanded Disability-RATAS network. These aspects are discussed in more detail in section 4.1.2 and section 4.1.3.

Visualization of the Fully Expanded Disability-RATAS

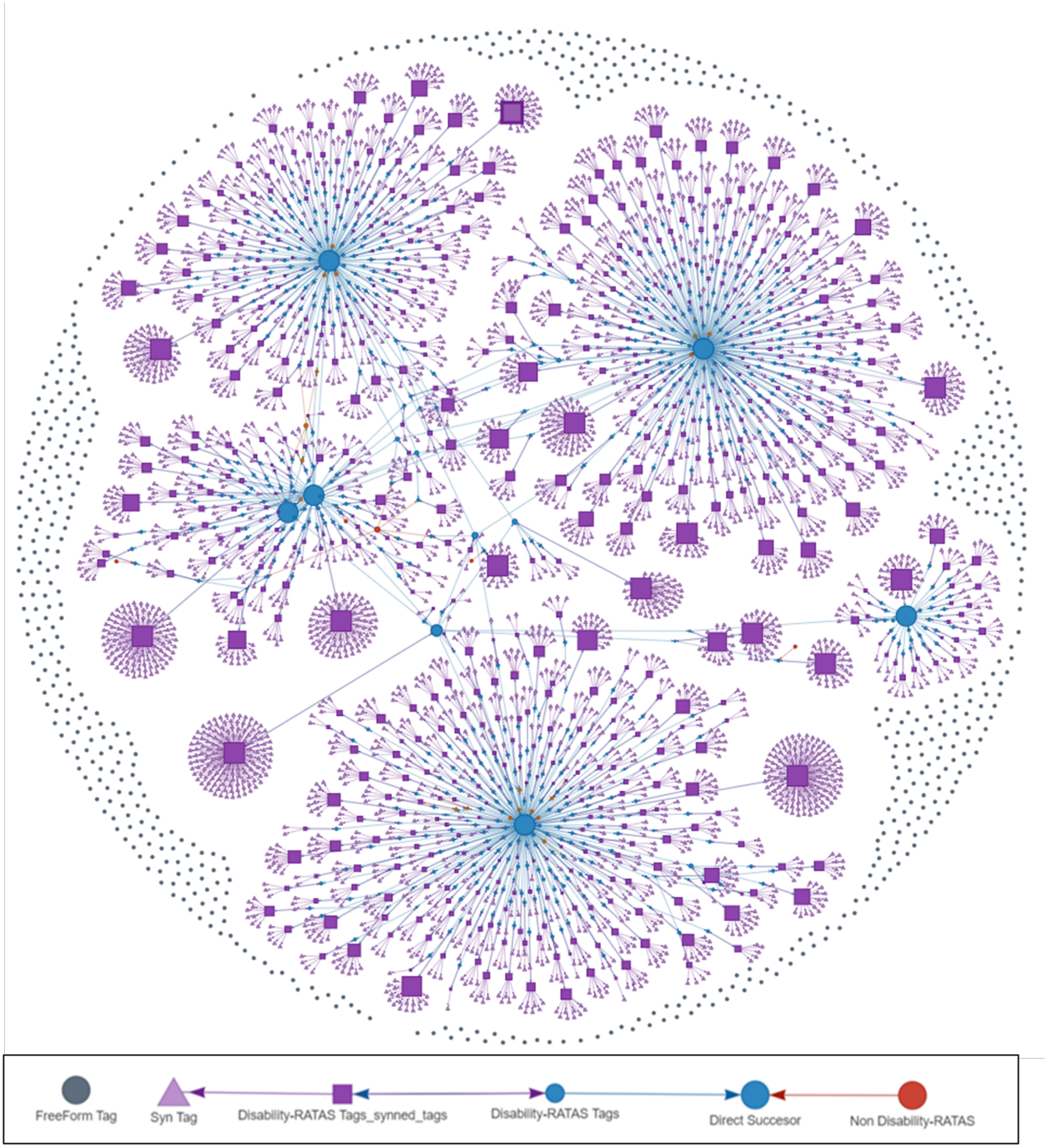


Figure 4.4: Visualization of the Fully Expanded Disability-RATAS. The root of the network is the tag Disability. The full legend is provided in Table 4.3.

4.1.2 Contributor and Wrangler Actions in the Disability-RATAS

Q1b: How can the actions that contributors, and wranglers perform in AO3 be represented in terms of operations on the t-RATAS network representation?

The representation of t-RATAS by definition are static, not dynamic, and thus are just snapshots in time. Even so, the networks provide the evidence of the actions that contributors and wranglers have performed in AO3.

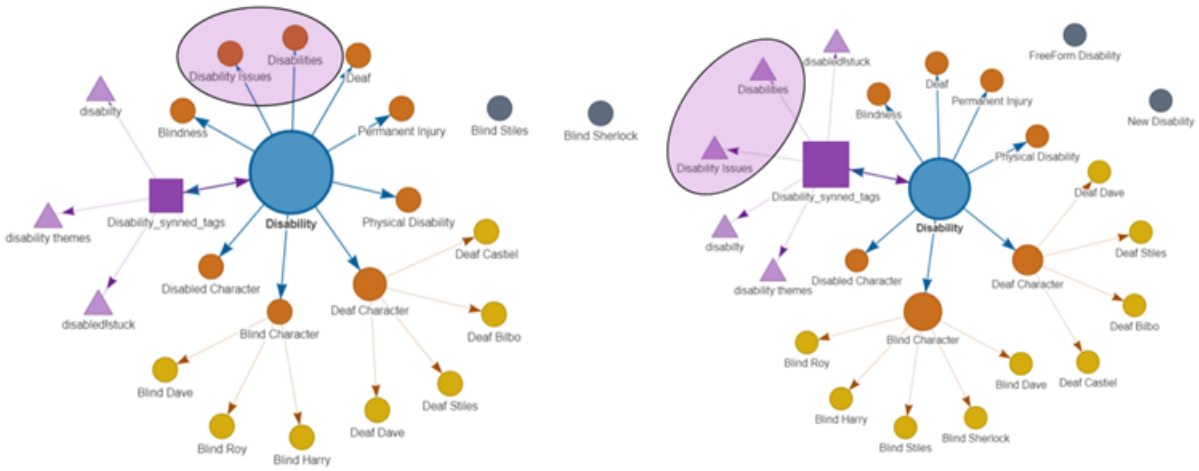
4.1.2.1 Wranglers Actions

Section 3.2.4 identified the different actions that a wrangler can perform on the AO3 tag hierarchy. These actions can be identified even when analysing a static t-RATAS.

The canonization of tags corresponds to the addition of the t-RATAS canonical nodes. When a wranger synonymizes a tag (either a canonical or free-form tag), this corresponds to an insertion of a node into a syn-set. This can be seen in the Expanded t-RATAS. The synonymization of a canonical tag can be identified in the movement or relocation of a canonical node within the ATAS to a syn-set, and the synonymization of a free-form node is seen as the addition of a node to a syn-set.

4.1.2.1.1 Synonymizing a canonical tag

Figure 4.5 illustrates the wranglers action of "synonymizing" a tag via the comparison of a before-vs-after snapshots of the Fully Extended Disability-RATAS. In this figure, the visualizations have been manually cropped and hand-annotated in order to focus in on the key aspects. The full dataset is analyzed in subsequent sections. The choice of colours and node shapes for these visualizations was listed in Table 3.2. In the figure, circled in magenta, are the two nodes that correspond to the two tags Disability Issues and Disabilities. In Figure 4.5(a), those two tags exist as canonical child tags of Disability in the Disability-RATAS. Then, in Figure 4.5(b), those two tags are now present in the syn-set of Disability. The transition corresponds to the wrangler taking the two canonical tags Disability Issues and Disabilities which were children of Disability and then synonimizing them to their former parent tag.



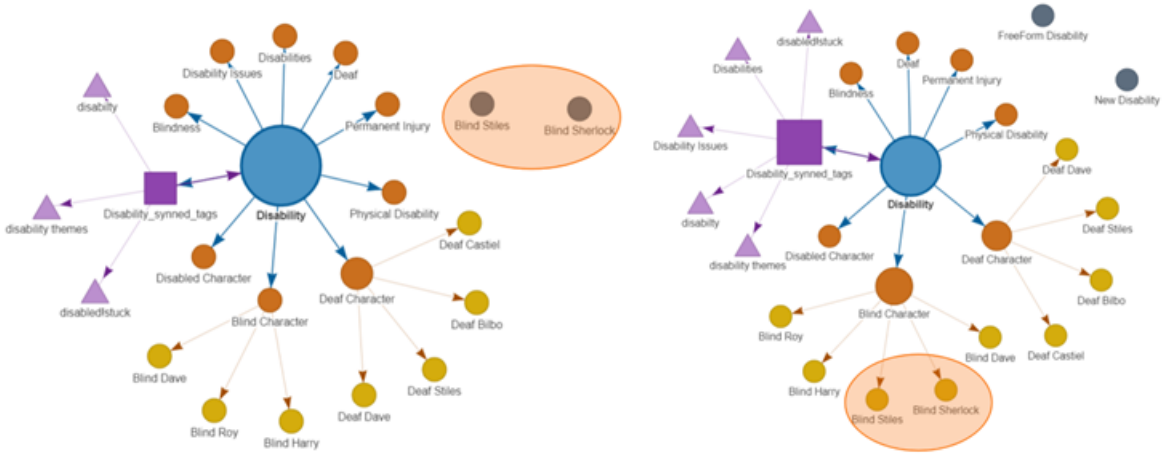
(a) State of Disability-RATAS (excerpt) prior to synomization action by wrangler

(b) State of Disability-RATAS (excerpt) following synomization action by wrangler

Figure 4.5: Illustration of the wranglers action of "syonymizing" a tag, shown via the comparison of a before-vs-after snapshot. (a) In the first snapshot, there two tags Disability Issues and Disabilities (circled in magenta) that exist as canonical child tags of Disability in the Disability-RATAS; (b) Subsequently, those tags are now present in the syn-set of Disability (circled in magenta).

4.1.2.1.2 Canonizing a Free-Form tag

Figure 4.6 illustrates the wranglers action of "canonizing" a tag via the comparison of a before-vs-after snapshot of the Fully Extended Disability-RATAS. In this figure, the visualizations have been manually cropped and hand-annotated in order to focus in on the key aspects. In the figure, circled in orange, are two nodes. In Figure 4.6(a), there two tags Blind Stiles and Blind Sherlock (circled in orange) that exist as free-form tags in the Fully Extended Disability-RATAS. Then, in Figure 4.5(b), those tags are now present in the subtag structure of Disability as successors of Blind Character (circled in orange).



(a) State of Disability-RATAS (excerpt) prior to the canonizing action by wrangler

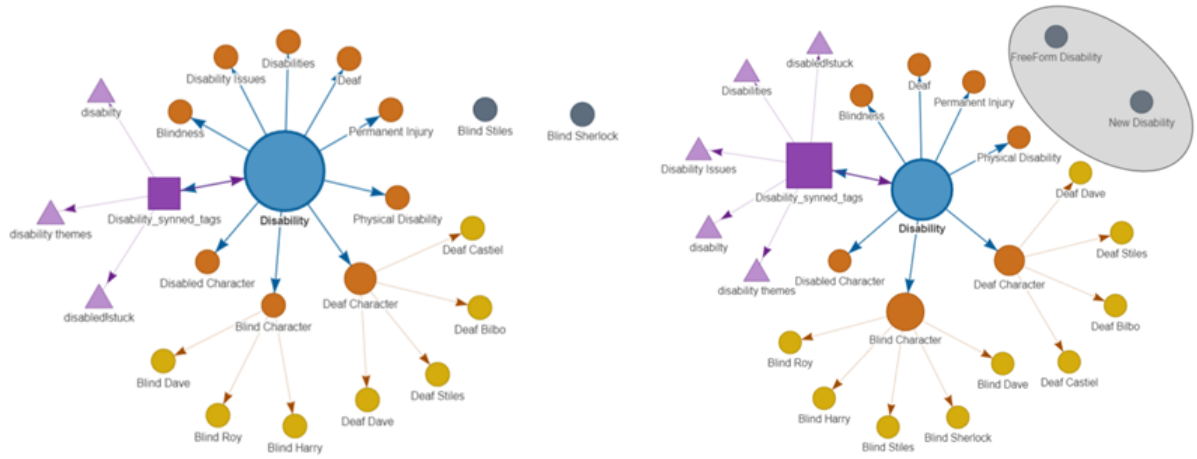
(b) State of Disability-RATAS (excerpt) following the canonizing action by wrangler

Figure 4.6: Illustration of the wranglers action of "canonization" a tag via comparison of a before-vs-after snapshot. (a) First, there two tags Blind Stiles and Blind Sherlock (circled in orange) that exist as free-form tags (isolated nodes) in the Disability-RATAS; (b) Subsequently, those tags are now present as successors of Blind Character (circled in orange).

4.1.2.2 Fan-fic Contributor Actions

When a fan-fic contributor uploads a new work and tags it with a (canonical or syn or free-form tag), no RATAS gets affected directly. If they tag their work with a free-form tag, this is evidenced by the appearance of new free-form nodes (isolated nodes) in the Fully Expanded t-RATAS. If they tag using canonical or synonymized tags, this can be shown using the Story-Set & t-RATAS Networks as described in section 4.3.2.4.

Figure 4.7 illustrates the creators action of "additing" a tag via the comparison of a before-vs-after snapshot of the Fully Extended Disability-RATAS. In this figure, the visualizations have been manually cropped and hand-annotated in order to focus in on the key aspects. In the figure, circled in gray, are two nodes. In Figure 4.7(a), there is the absence of two tags FreeForm Disability and New Disability. Then, in Figure 4.7(b), those tags are present as isolated nodes (circled in gray) after their creation.



(a) State of Disability-RATAS (excerpt) prior to the addition of new free-form tags by the contributor

(b) State of Disability-RATAS (excerpt) following the addition of new free-form tags by the contributor

Figure 4.7: Illustration of the creators action of "addition" a tag via comparison of a before-vs-after snapshot. (b) Appearance of two new isolated nodes (free-form tags) FreeForm Disability and New Disability (circled in gray).

4.1.3 Structural Properties of the Disability-RATAS

> Q1c: Which structural properties can be extracted and visualized from the t-RATAS structures? What can these properties tell us about the AO3 'Additional Tags' category (at a given point in time)?

4.1.3.1 Data Generation

As described in section 3.5.1, there are several network properties of interest to obtain from all three versions of the t-RATAS. To do so, we generated a report of the network properties of the Fully Expanded Disability-RATAS, using the invocation of `NetAnalyser` as follows:

```
>python .\ao3_analyser.py --GenerateRATASProperties
    rata_disability_expanded_full.json -o .\report_disability_expanded_full
```

A `.xlsx` file was generated containing various network properties, as per section 3.5.1.

4.1.3.2 Results

The network properties are organized on different worksheets, as per the description in section 3.5.1. Two worksheets are provided below, in Table 4.4 and Table 4.5. Table 4.5 shows the properties obtained from the Expanded Disability-RATAS rather than the Fully Expanded Disability-RATAS in order to, since our analysis will initially ignore the isolated nodes. The full reports are quite large and are provided in Appendix C.1.

Properties	Values
large_nodes	Disability, Blindness, Muteness, ...
total_nodes	5069
total_edges	5091
total_canonical_leafs	31
total_freeform	636
total_synonymized	3139
total_canonical	667
total_synsets	627

Table 4.4: Worksheet (I): General Network Information, as generated for the Fully Expanded Disability-RATAS.

Properties	Values
is_strongly_connected	FALSE
is_weakly_connected	TRUE
diameter_undirected	11
radius_undirected	6
center_RATAS_undirected	['Disability', 'Disabled Character']
has_cycles	...
k_components	[2, 1]
2_components	...

Table 4.5: Worksheet (II): Network Properties, as generated for the Expanded Disability-RATAS.

4.1.3.3 Data Analysis

4.1.3.3.1 General Concept Prevalence

The network information can tell us the general prevalance of the Disability tag is as a topic in AO3. We can see from Table 4.4 that, out of the total 5069 nodes in the network, 4442 (15.06%) nodes representing canonical tags, 636 (14.32%) nodes represent free-form tags, and

3139 (70.67%) nodes represent syn tags. These syn tags are spread out of 627 syn-sets, . This clearly shows an important difference among the amounts of canonical, free-form, and syn tags. This also could indicate a effective wrangling process that follow the guidelines, that instead of expanding the network, the wranglers are synonymizing as much as possible. On the other hand, these properties shows similar amount of free-form tags and canonical tags. This shows how the community tends to create new tags in the archive even when many syn alternatives exist. This will be flagged to point to further study in Chapter 5.

In addition, Table 4.4 shows that there are more than 4442 stories captured from the entire AO3 archive which contain any tag that relates the the topic of *Disability*. We note that there may be other fan-works not captured in the Fully Expanded Disability-RATAS (see section 3.3.3.4). Thus, it is likely that there is more than 4442 fan-fics in AO3 tagged with disability-related tags.

We can state this cause of the appearance of each tag in at least one fan-work, plus canonical tags are used in several fan-works in order to be canonized.

These results do not tell us the relative prevalence of *Disability*, as compared to other tags. For this, more of the AO3 site needs to be scraped in order to obtain the necessary information, which is left to further studies.

4.1.3.3.2 Network Shape

When we look at the visualization of the Basic Disability-RATAS in Figure 4.1, we can see in its structure the appearance of multiple predecessors in several nodes. Then, it is directly implied that the Basic Disability-RATAS (as well as they extended versions) is not a tree. This finding moves us away from having a tree perspective in the *Disability* structure. It is a known fact that a tag can have multiple meta-tags, but it does not imply that needs to happen for a specific t-RATAS. If this is the case, then the network will follow unique properties of its own.

4.1.3.3.3 Connection to other t-RATAS

After expanding the Basic Disability-RATAS to the Expanded Disability-RATAS, we can see significant growth in the expanded version. It grew from 783 nodes up to 4433 nodes, which is a total of 3650 additional nodes. The difference is completely due to syn tags, with only a small number of new canonical tags (in the form of new meta-tags of different nodes). This shows that different canonical nodes in the Basic Disability-RATAS are revealed to have

multiple predecessors once the network is Expanded. These nodes are of interest for future study since they represent connection points from the Disability-RATAS to other t-RATAS. These nodes are represented in red in Figure 4.3. One example is the tag Wheelchairs which connects as a meta-tag of five canonical tags in the Basic Disability-RATAS.

4.1.3.3.4 Syn-Set Analsysis

The syn-sets of the Disability-RATAS is of potential interest. Worksheet (I) in Table 4.5 shows us that the canonical tag Disability reported the largest syn-set with 133 syn tags, followed by Muteness with 121 syn tags. There were 75 tags that had only one tag in their syn-sets. There were 40 canonical tags that had no syn tags. The static analysis report lists the canonical nodes and their syn sets in Worksheet (IV). Worksheet (IV) for the static analysis of Disability-RATAS in provided in Table C.4 in Appendix C.1. Lastly, Worksheet (I) shows us that there are 636 free-form tags, which represent future candidate canonical or syn nodes, pending future wrangler action.

4.1.3.3.5 Network Connectivity

Worksheet (II) of the Static Analysis Report, shown in Table 4.5, shows the properties obtained from the Expanded Disability-RATAS. As mentioned previously, this structure is analyzed rather than the Fully Expanded Disability-RATAS in order to focus the analysis on the connected network structure (thus we wish to exclude the isolated nodes).

As explained in section 3.5.1, the network properties are mined for both the undirected and directed versions of the network, to generate different points of view.

Analyzing the properties in Worksheet (II) in Table 4.5, we can see that the diameter — the maximum number of nodes to be traversed between two nodes — of the Expanded Disability-RATAS is 11. Having a low diameter with respect to the number of nodes says that there are a large number of edges that are "uniformly" positioned so the largest distance from two nodes is not too far. This also constrains the girth of the network, setting smaller lengthened cycles if they exist.

In addition, the radius and center tell us the bridges of the Expanded Disability-RATAS. We refer to bridges as being the edges in the network that are essential to connect two components (or nodes) in the network. The results showed that Disability and Disabled Character are at the center of the graph, having a minimum eccentricity of six. They both are six nodes apart from any node in the network.

Studying the maximal subnetworks that have, at least, node connectivity k-strong, we can see the number of edges that need to be removed in this sub-section of the network to be disconnected. There are certain components in the Expanded Disability-RATA that are attached enough to be disconnected only when at least two edges are removed. This might bring different types of analysis on these components depending on the research in question highlighting those nodes.

4.1.3.3.6 Discovery via Exploration of the Visualizations

The static visualizations of the network, with its interactivity, is a great tool that allows scholars to zoom in and see the specifics of certain areas of the network or certain nodes. The visualization also gives a general idea about is the relationships among canonical tags, syn tags and free-form tags. Additionally, for a given t-RATAS, scholars can contrast the network visualization and the analysis reports as a way to hone in on areas for more specific analysis. The analysis using both visuals and reports generated can help to find nodes of interest to focus in the visualization, or the other way to obtain actual properties of the network that for a more technical audience can provide more deep meaning to visualize the network.

Future interdisciplinary collaboration with other scholars will enrich the results giving more meaning to different properties and pointing out different directions to take in the Network Analysis. Knowing which nodes could be of interest, which type of component structures will scholar will be looking forward to, etc. will make important directions in the study to improve its results.

4.1.3.4 Limitations

In doing this analysis, we highlight (as also noted in several subsections in Chapter 3) that there are limitations within the analysis and generation of data using the tool provided, such as:

- As proof of concept of the prototype, the software tools only scraped certain t-RATAS (specially, various versions of Disability-RATAS). Although this limits the study, it does maintain our focuses on the disability discource, as a topic of interest. Adding more RATAS to the study would add more information and would then likely generate further interesting properties to mine (e.g., to examine the generalizability of the findings, such as the six-degree separation between the center of different RATAS, the distances

between different focus tags as topics of interest, the accuracy of degree distribution within the ATAS more generally, etc.).

- Scraping multiple RATAS does take a considerable time, which may become a constaint. AO3, as one of its terms of service, requires scraper bots to wait a certain time in between queries. Each node in the RATAS represents one query, and then there will be one additional query for each canonical node in its structure for expanding the RATAS and additional queries to obtain its non-canonical tags. As we could see, one RATAS itself won't be a problem to obtain, but its expansion could potentially take several minutes. This is likely to be the case for networks large enough as the one studied with the Disability tag as its root.

- Visualizing large Networks can be a bit challenging when there is significant crowding due to a large number of nodes and edges. Creating dynamic visualizations that allow zooming in and out to get different properties is expected to help. The use of other tools will potentially also be helpful to visualize `NetworkX` networks, but still do not overcome the crowded visualizations.

4.2 Findings: Comparison of RATAS' Snapshots

In this section, we answer the second cluster of the three research question clusters, with a focus on the following question within the cluster:

Q2c: What can t-RATAS tell us about the major changes to the AO3 tag structure over the past years of activity?

In order to generate findings for Q2c, we will make use of the techniques described in section 3.5.2. These techniques provide the means of getting multiple snapshots of the t-RATAS, each snapshot from different points in time. These techniques also provide a way to visualize their differences and similarities and to express their differences through user behaviours, as was the target of objective (vii). To answer Q2c, we will first answer the following two questions as preamble::

Q2a: How can we capture a snapshot of the t-RATAS at different points in time?

Q2b: Given two t-RATAS snapshots, which technique(s) can be developed to compare them with one another, in order to express change over time?

4.2.1 Snapshot Capture

Q2c: How can we capture a snapshot of the t-RATAS at different points in time?

As described in section 3.3.3.3, we can capture different snapshots of a t-RATAS with `TagScra`. We can make use of the current version of AO3 or we can specify the use of the Wayback Machine as an argument. Through this, we can access the current version and many different versions of previous years of the archive.

To build the Disability-RATAS, we can use the module as described in section 3.3.3.3. We provide, as an argument, the archival year that is desired to obtain a snapshot of the *Disability* tag structure. As also described in that section, there is set of possible archival years that have been implemented. This generates the `JSON` file of the Disability-RATAS of that specified archival year. The specific command-line invocations is shown below, first to obtain the archival version dated to 2013, and second to obtain its current version dated 2022:

```
>python .\ao3_analyser.py -ts Disability -o rata_disability_2013 -ao3_v 2013

>python .\ao3_analyser.py -ts Disability -o rata_disability_current
```

Using this module, we capture snapshots of all of the Basic t-RATAS of interest.

Additionally, we can extend the Basic Disability-RATAS to its extended and fully extended versions using the module `TagScra` as showed in section 3.3.3. The specific comman-line invocation specify the module to extend the `JSON` RATAS obtained using the path of the file as an argument:

```
>python .\ao3_analyser.py --ExpandRATAS .\rata_disability_current.json
    -o rata_disability_current_expanded
```

In order to get the Fully Expanded Disability-RATAS, we firstly need to obtain all the non-canonical tags that match with *Disability*. The specific comman-line invocations specify the module to obtain them and followed by the invocation to fully expand the Expanded Disability-RATAS using the scraped files:

```
>python.exe .\ao3_analyser.py -ts Disability --NonCanonical
    -o non_canonical_disability

>python.exe .\ao3_analyser.py --ExpandFullRATAS
    .\rata_disability_current_expanded.json
    .\non_canonical_disability.txt  -o rata_disability_current_expanded_full
```

4.2.2 Snapshot Comparison

> Q2b: Given two t-RATAS snapshots, which technique(s) can be developed to compare them with one another, in order to express change over time?

To demonstrate the different comparison techniques (e.g., the snapshot difference visualization, the snapshot similarities visualization), we have generated two different snapshots of the Basic Disability-RATAS, one from the current 2022 version and one from 2013. Having these two different snapshots of the same Basic t-RATAS allows us to proceed with the comparative analysis and comparative visualization.

4.2.2.1 Producing the Snapshot Difference Visualizations

First, to generate the visualization of the differences between two snapshots, we can use the technique that is implemented via the module `--GenerateDifferences` (or `-gd`), as described in Section 3.5.2. We used that technique to generate an analysis of differences between 2013 and the current version of the Disability-RATAS. The invocation is shown below:

```
>python .\ao3_analyser.py -gd .\rata_disability_2013.json
    .\rata_disability_current.json
```

As described in section 3.5.2, the invocation of `-gd A.json B.json` produces two visualizations: (i) Network A, with colour highlighting of the nodes and edges (which in Network A that are also in Network B and which are not), and (ii) Network B, with highlighting the nodes and edges (which in Network B that are also in Network A and which are not). Both of these are shown in Figure 4.8 and Figure 4.9, respectively.

4.2.2.2 Interpreting a Snapshot Difference Visualization

This sub-section illustrates the interpretation of Snapshot Difference Visualizations, using Figure 4.8 and Figure 4.9 as two examples.

In Figure 4.8, a visualization is shown of the snapshot difference between Disability-RATAS 2013 as the base, with Disability-RATAS 2022 shown in relation. In Figure 4.9, a visualization is shown of the snapshot difference between Disability-RATAS 2022 as the base, with Disability-RATAS 2013 shown in relation.

The choice of colours for the nodes/edges for the visualizations was listed in Table 3.5, and is repeated here (Table 4.6, Table 4.7) for convenience:

Legend: Node/Edge Colours for the Snapshot Differences Visualizations

Diff(Disability-RATAS 2013, Disability-RATAS 2022)	
nodes/edges present in 2013 that are **still** present in 2022:	red
nodes/edges present in 2013 and are **no longer** present in 2022:	blue

Table 4.6: Legend of Node/Edge Colours for the Snapshot Differences Visualizations between Disability-RATAS 2013, as the base, and Disability-RATAS 2022, as visualized in Figure 4.8

Legend: Node/Edge Colours for the Snapshot Differences Visualizations

Diff(Disability-RATAS 2022, Disability-RATAS 2013)	
nodes/edges present in 2022 that were **also** present in 2013:	red
nodes/edges present in 2022 and were **not also** present in 2013:	blue

Table 4.7: Legend of Node/Edge Colours for the Snapshot Differences Visualizations between Disability-RATAS 2022, as the base, and Disability-RATAS 2013, as visualized in Figure 4.9

Snapshot Difference Visualization,
Base: Disability-RATAS 2013,
Disability-RATAS 2022 in Relation

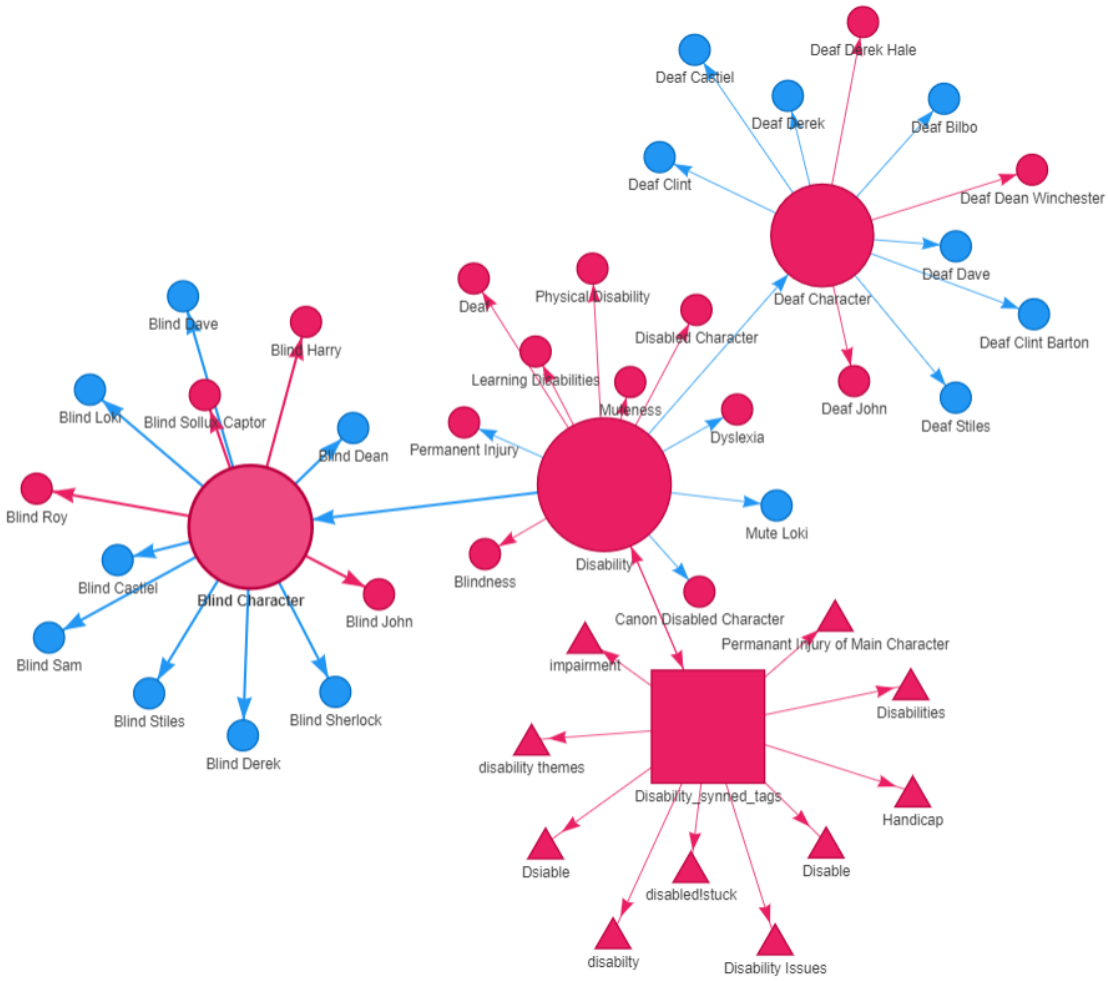


Figure 4.8: Snapshot Difference Visualization of Basic Disability-RATAS, Base: Disability-RATAS 2013, Relative to Disability-RATAS 2022. In this visualization, red indicates nodes and edges that are present in the base t-RATAS and are **still** present in the later t-RATAS. Blue indicates nodes and edges that are present in the base t-RATAS and are **no longer** present in the later t-RATAS.

Snapshot Difference Visualization,
Base: Disability-RATAS 2022, Disability-RATAS 2013 in Relation

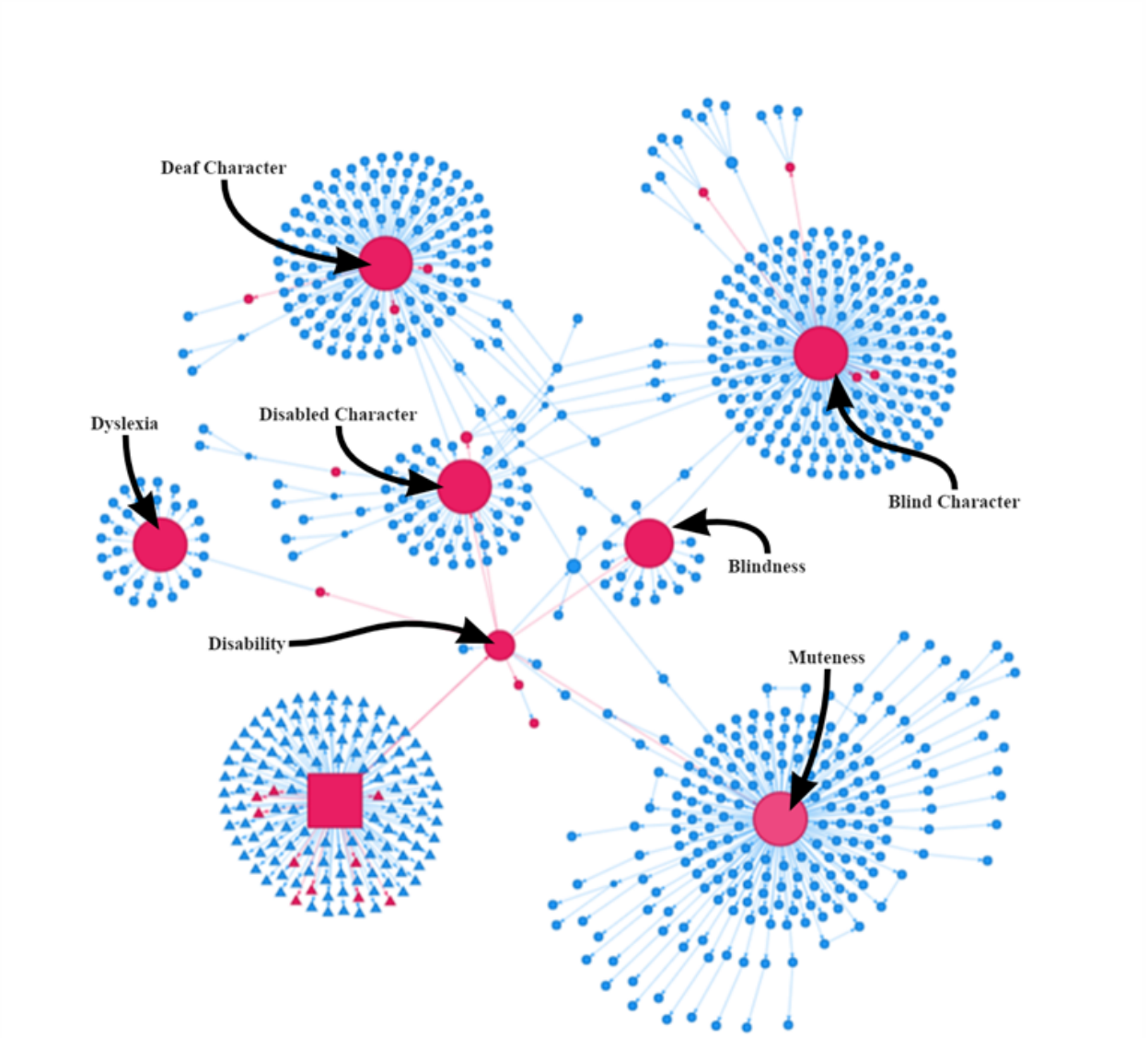


Figure 4.9: Snapshot Difference Visualization of Basic Disability-RATAS, Base: Disability-RATAS 2022, Relative to Disability-RATAS 2013. In this visualization, red indicates nodes and edges that are present in the base t-RATAS and **were also** present in the earlier t-RATAS. Blue indicates nodes and edges that are present in the base t-RATAS and were **not** present in the earlier t-RATAS.

4.2.2.3 Interpreting a Snapshot Difference Visualization for Fully Expanded t-RATAS

To further analyze the difference between the snapshots, we ran the module again, this time on to generate a visualization of the differences between the current Fully Expanded

Disability-RATAS (2022) and the Basic Disability-RATAS (2013). The invocation is as follows:

```
>python .\ao3_analyser.py -gd .\rata_disability_2013.json
    .\rata_disability_current_expanded_full.json
```

The results of this invocation are shown in Figure 4.10 and Figure 4.11. Similar as the previous illustrations, Figure 4.10 indicates the differences between Basic Disability-RATAS (2013) and Fully Expanded Disability-RATAS (2022), with the 2013 version the base.

The choice of colours for the nodes/edges for the visualizations was listed in Table 3.5, and is repeated here (Table 4.8) for convenience:

Legend: Node/Edge Colours for the Snapshot Differences Visualizations

Diff(Disability-RATAS 2013, Fully Expanded Disability-RATAS 2022)	
nodes/edges present in 2013 that are **still** present in 2022:	red
nodes/edges present in 2013 and are **no longer** present in 2022:	blue

Table 4.8: Legend of Node/Edge Colours for the Snapshot Differences Visualizations between Disability-RATAS 2013, as the base, and Fully Extended Disability-RATAS 2022, as visualized in Figure 4.10

As per the table above, red indicates nodes and edges that are present in the Basic Disability-RATAS (2013) and are **still** present in current fully expanded structure. Blue indicates nodes and edges that are present in Basic Disability-RATAS (2013) and are **no longer** present in the later version. Then we can see the lost of several edges and only two nodes that can be identified easily using the visualization. This visualization brings interesting points to analyse as many edges were removed while their nodes still remain.

When comparing Figure 4.10 with Figure 4.8, which visualized the difference with Basic Disability-RATAS (2022), we can see only two nodes are blue when comparing with the Fully Expanded Disability-RATAS (2022) instead of 16 as in Figure 4.8. This shows that 14 out of 16 of this nodes were not removed, instead they represent syn tags or free-form tags currently, as the Basic Disability-RATAS (2022) does not capture.

Furthermore, when analysing the next visualization, Figure 4.11 and similar to Figure 4.9 red nodes and edges represent the nodes that were present in 2013 while blue indicates nodes and edges represent new additions to the structure. This visual shows clearly the growth of the structure over the years.

The choice of colours for the nodes/edges for the visualizations was listed in Table 3.5, and is repeated here (Table 4.9) for convenience:

Legend: Node/Edge Colours for the Snapshot Differences Visualizations

Diff(Fully Expanded Disability-RATAS 2022, Disability-RATAS 2013)
nodes/edges present in 2022 that were **also** present in 2013: red
nodes/edges present in 2022 and were **not also** present in 2013: blue

Table 4.9: Legend of Node/Edge Colours for the Snapshot Differences Visualizations between Fully Extended Disability-RATAS 2022, as the base, and Disability-RATAS 2013, as visualized in Figure 4.11

Snapshot Difference Visualization,
Base: Basic Disability-RATAS 2013 in Relation,
Fully Expanded Disability-RATAS 2022

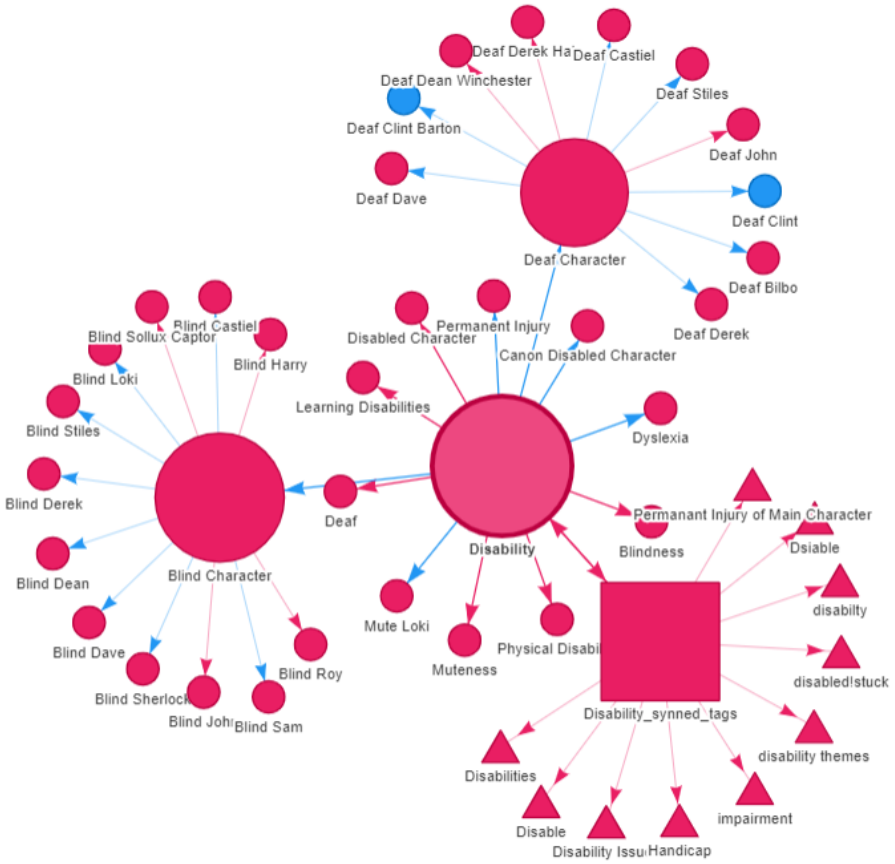


Figure 4.10: Disability-RATAS 2013, shown relative to the Fully Expanded Disability-RATAS 2022. The red nodes and edges are those which are still present as of 2022. The blue nodes and edges are those that are no longer present as of 2022 (e.g., deletions or changes to another RATAS).

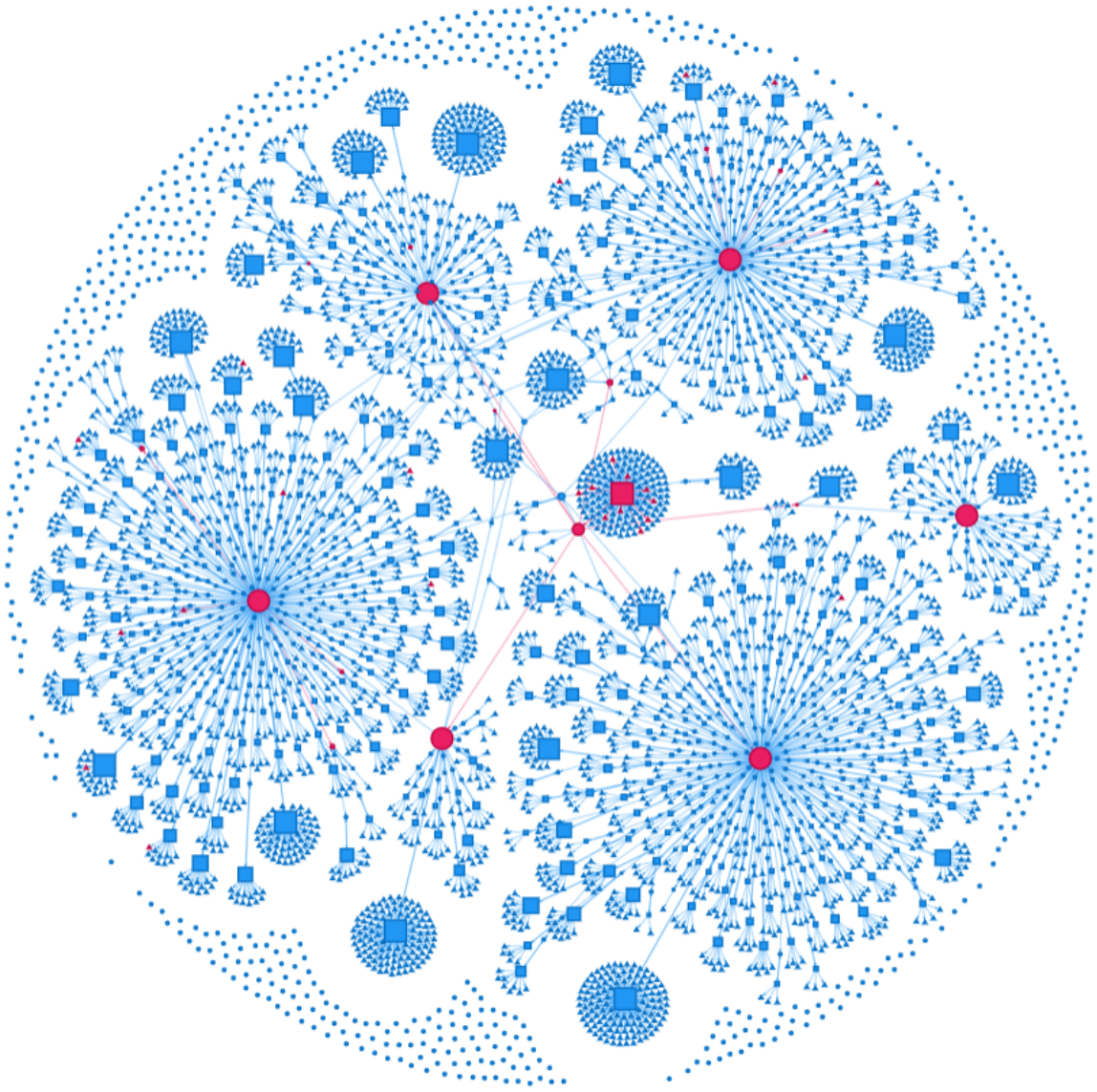


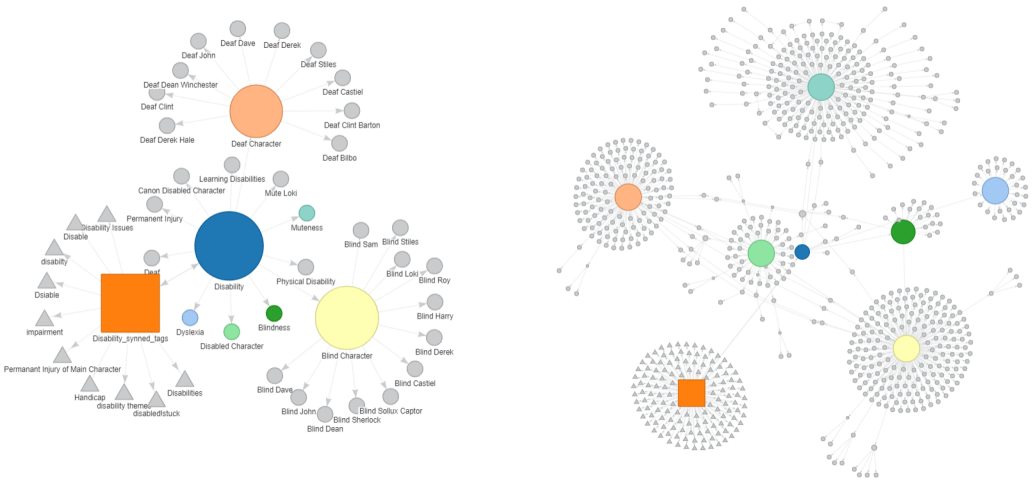
Figure 4.11: Visualization of the Fully Expanded Disability-RATAS (dated to 2022), shown relative to the Basic Disability-RATAS (dated to 2013). The red nodes and edges are those that were also present as of 2013. The blue nodes and edges are those that were not present as of 2013 (additions).

4.2.2.4 Examining Similarities Through Visualization

Similarly, we can compute and generate the visualizations highlighting the similarities of the two snapshots in a way that allows us to easily align the two snapshots. In this visualization mode, unique colours are used to identify the nodes that are common to both snapshots. The decision was made to visualize only the nodes with a sufficiently high out-degree. This will help to visualize how those nodes changed from one version to another.

This functionality was invoked using `-gs` or `--GenerateSimilarities` instead of `-gd` as described in section 3.5.2 as follows. Its output is shown in Figure 4.12:

```
>python .\ao3_analyser.py -gs .\rata_disability_2013.json
    .\rata_disability_current.json
```



(a) (b)

Figure 4.12: Visualization of the similarities between Disability-RATAS dated to 2013 and 2022. Nodes with relatively higher out-degree are highlighted with the same colors in both networks.

4.2.2.5 Examining Differences Through Extracted Data

In order to obtain quantitative results in the form of a dataset, we use the functionality as described in section 3.5.2. Continuing the study of Disability-RATAS, we scraped the version for all years availables in the Waytime Machine (2013, 2014, 2015, 2016, 2017, 2018, 2021, 2022). We then used the module to performed all sequential comparisons, generating a `.csv` file. This file, as explained in section 3.5.2, contains all tags (one per row) in all the

Disability-RATAS and then their presence in each of the snapshots. Each row is labelled with a 'type' (node addition, removal, synonymization). Through this column, we see the listing of the "political" actions taken on the networks over time. Then, the dataset stores the political actions are done by the wrangler from one snapshot to the next one. A full view of this dataset is shown at Table C.8.

Table 4.10 provides an overview of the partial results comparing the Disability-RATAS dated 2013 to the Disability-RATAS dated to 2022.

To generate this dataset, we used the module `--GenTagDiff`, which receives two arguments: one for each RATAS `JSON` file to compare (assuming the first argument is the most current version).

```
>python .\ao3_analyser.py --GenTagDiff
    .\rata_disability_current.json .\rata_disability_2013.json
    -o reports_diff_disability_current_2013
```

Tags	2013_presence	current_presence	2013_type	current_type	Action
mute ashton	FALSE	TRUE		syn_tag	addition
Deaf Clint	TRUE	FALSE	canonical_tag		removed
Mute Loki	TRUE	TRUE	canonical_tag	syn_tag	synonymized
disabilty	TRUE	TRUE	syn_tag	syn_tag	
⋮	⋮	⋮	⋮	⋮	⋮

Table 4.10: Comparison of tag usage between Disability-RATAS from 2013 and its current version. Table C.6 shows the full table.

After analyzing the full version of Table 4.10 (which is shown in Appendix C.1.1), we can conclude a consistent growth over the years in the Disability-RATAS.

4.2.2.5.1 Longitudinal Growth of Disability-RATAS by Type of Tag

Figure 4.13 provides a chart of the counts of each type of tag in the Disability-RATAS over the series of years 2014–2022. From this graph, we can see a constant growth over the years, tending to a sharp increase over the past four years. The number of canonical tags is increasing at a faster rate than the increase in synonymized tags.

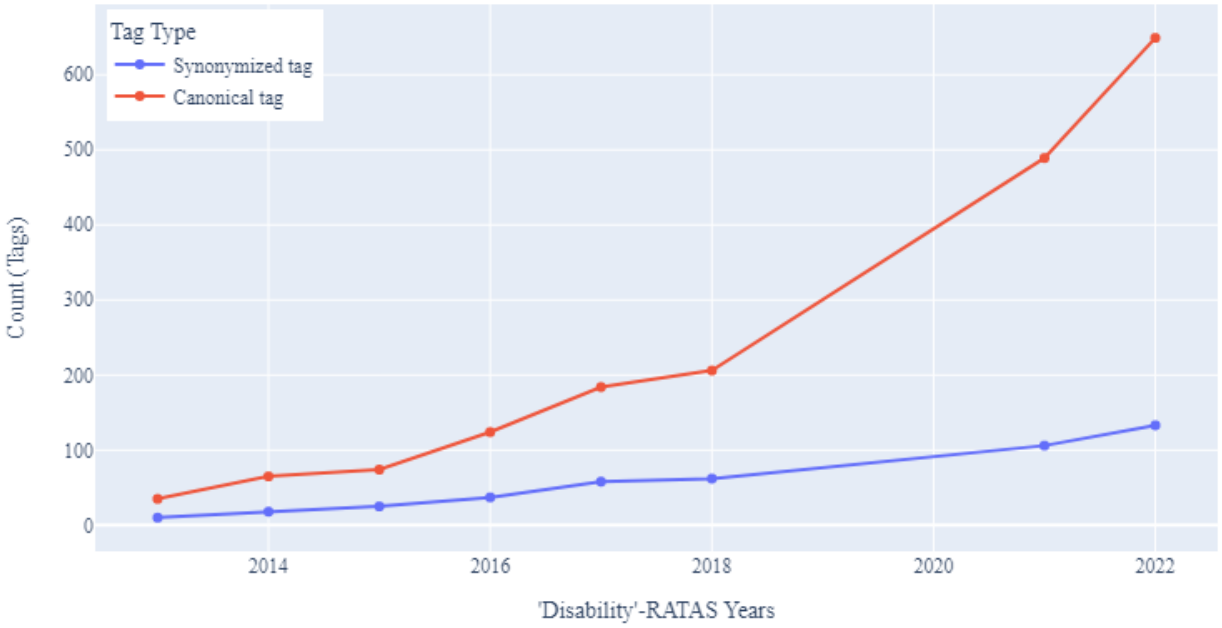


Figure 4.13: Line graph demonstrating the count of tags within the Disability-RATAS over the years 2014–2022, broken down by type (synonymized tag, canonical tag).

4.2.2.5.2 Longitudinal Trends in Wrangler Political Actions on Disability-RATAS by Type of Ranger Action

Figure 4.14 shows the number of political acts performed per year, obtained by analyzing the series of snapshots of Basic Disability-RATAS over the years (obtained through the Wayback Machine). The graph shows that the act of canonization has been the most popular wrangler action. It also shows that there has been some deletion of tags in the archive and a small number of synonymizations.

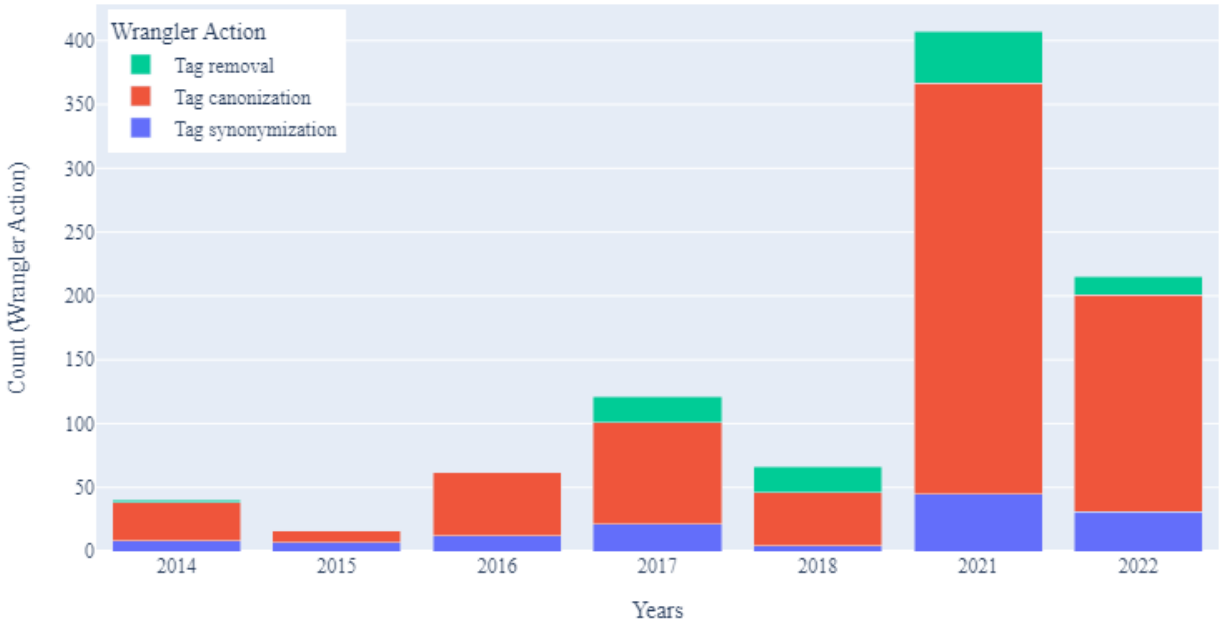


Figure 4.14: Stack-bar chart demonstrating the count of political acts performed by wranglers over the years 2014–2022 on the Disability-RATAS. Political acts include tag canonization, synonymization, and removal.

4.2.3 Discussion: Changes Over Time

Q2c: What can *t*-RATAS tell us about the major changes to the AO3 tag structure over the past years of activity?

The discussion is broken into different sections according to the generated data and visualizations.

4.2.3.1 Difference and Similarity Visualizations

Two types of visualization techniques have been developed and applied in order to highlight similarities and differences when comparing nodes and edges: one to highlight the operation of subtraction (the difference visualization) and the other to highlight the similarities among structure over time (the similarity visualization).

While analyzing the differences between two snapshots of the same RATAS, as shown in

Figure 4.8 and Figure 4.9, we can easily perceive two main factors: the loss of nodes and edges over time, and the gain of nodes and edges over time. Firstly, analyzing the changes from 2013 to 2022, focusing on Figure 4.8, we can see the loss (the blue nodes) of many tags (16 tags) that are no longer in the Basic Disability-RATAS (2022). This highlights that these tags were either removed from the archive or they were moved to another RATAS. Both of these two possibilities are of interest for further investigation, to discover what happened and why.

4.2.3.1.1 Tracking Specific Tags

Analyzing the similarities between two snapshots, as shown in Figure 4.12(a), allows us to see the nodes with a relatively high out-degree highlighted with the same colour in both versions. This allows us to visually perceive the changes that occur certain nodes over time. We can track their growth and their position in the network over time. In this similarity visualization, we are able to follow four nodes: Blindness, Dyslexia, Muteness, and Disabled Character, all of which started out as canonical leaf nodes in 2013 (Figure 4.12(a)). They hold their own subtag hierarchy as metatags (Figure 4.12(b)). By 2022, we can see that some of those nodes are now interconnected, such as Blindness, Disabled Character, and Muteness with Disability commonly shared by all of them.

In sum, the visualizations with their dynamism, allow the analyst to spot major changes in the structure of the Disability-RATAS over the years. They proved to highlight interesting nodes and edges in the studied RATAS that show the dynamic activity of the users in AO3, bringing new questions to further analysis. Their use can contribute to analyzing the generated data and focusing on specific areas of the networks. This tool in combination with the reminded data to discuss has the potential to further exploit and analyze together.

4.2.3.2 Extracted Data

4.2.3.2.1 Network Growth: New Nodes

As analyzed before, starting from Figure 4.9, which shows the Basic Disability-RATAS in 2022, we can see that there are several new nodes in 2022 that were not present in 2013. These new nodes are represented by the blue color. Then we can see where and how the old nodes have grown, becoming main components with their own successors and even with more than one predecessor. An example is seen with Blind Character tag which, in 2013, is a direct predecessor of both Disability as well as Blindness. Then, in 2022, Blind Character remains as predecessor

of Disability but also is the direct predecessor of Blindness and of Disabled Character. Similarly, Deaf Character got linked to Disabled Character through the years. Furthermore, we can back up these observations with the analysis of the data generated when comparing the use of tags over the years in the Disability-RATAS, as examined in section 4.2.2.5. The results of this data, shown in Figure 4.13, reflect how the number of both syn and canonical tags increment over the years in the Basic Disability-RATAS. We can see that canonical tags tend to increase faster than the syn tags in this Basic version of the structure and mostly within recent years. The RATAS in 2013 was composed of 43 nodes and currently contains 783 nodes, where 83% of them are canonical tags and the rest syn tags.

4.2.3.2.2 Network Contraction: Conjecturing about the Loss or Absorption of Nodes into Synsets

A side-by-side comparison of two variants of snapshot difference visualizations can be used to reveal "missing" nodes.

First, we examine a difference visualization with an older version as a base, shown relative to a more recent version. For example, consider Figure 4.8, which shows a Snapshot Difference Visualization with a base of Basic Disability-RATAS 2013, and with Basic Disability-RATAS 2022 shown in comparison. Next to this, we compare a variant, shown in Figure 4.10, which shows a Snapshot Difference Visualization with a base of a Basic Disability-RATAS 2013, and with a Fully Expanded Disability-RATAS 2022 shown in comparison.

In both Figure 4.8 and Figure 4.10, the baseline for comparison is the same: the Basic Disability-RATAS 2013. The only difference between the two figures is the RATAS used in comparison, one is Basic, the other is Fully Expanded. Recall that a Fully Expanded RATAS is the Basic RATAS, plus all syn-set for each canonical node in the basic version plus all freeform tags (isolated nodes) related to the root as defined in (eqn:3.2.4).

In Figure 4.8, there are 16 canonical nodes shown in blue. The colour blue means that these tags were present in the 2013 version of the Basic RATAS, but are no longer present in the 2022 version of the Basic RATAS. Let's call them "missing" nodes. An analyst might then wonder about the 2022 status of these nodes: where did those tags go? There are a number of possibilities:

1. Did these nodes get synonymized to another canonical node within the Disability-RATAS?

2. Did these nodes get moved to another RATAS (e.g., not present in the Disability-RATAS but present in a different t-RATAS)?

3. Were these removed completely from the AO3 tagging structure (e.g., not present in *any* RATAS)?

Possibility #1 (the nodes got synonymized to another canonical node within the Disability-RATAS) and Possibility #2 (the nodes got moved to another RATAS) can be investigated by looking at Figure 4.10. The possibility of a movement to another RATAS might bring questions to the table about which RATAS was and how does it relate to the Disability tag? Then we see that 14 of the 16 "missing" nodes are shown in red in Figure 4.10, meaning that the nodes are present in 2022 Fully Expanded Disability-RATAS. However, for each one of these 14 nodes, the edge connecting the node is blue, meaning that the relationship between the parent node and child node is no longer present in the Disability-RATAS. Thus, we can infer that those tags have been moved to a RATAS other than the Disability-RATAS. For example, the tag Mute Loki, as of 2022 fully extended version, is now synonymized to the tag Mute Loki (Marvel).

We can explain the repositioning of 14 of the 16 "missing" nodes, but what about the remaining two nodes? The two nodes are Deaf Clint Barton and Deaf Clint. In Figure 4.10, these two nodes are blue and their connecting edges are blue as well. Possibility #3 (the nodes got removed completely from the AO3 tagging structure) can be investigated by manually searching for the these nodes. For example, Deaf Clint Barton is now an isolated metatag with the parent Fandom tags of Marvel Cinematic Universe and Avengers (Comics).

4.2.3.2.3 Political Actions in the Structures

Figure 4.12 clearly shows how nodes corresponding to canonical tags such as Disabled Character, Blindness, Dyslexia, and Muteness have grown in the Disability-RATAS over the years. These nodes in 2013 were leaves, whereas in the 2022 version they significantly grew through the insertion of nodes while canonizing and the expansion of their syn-sets while synonymizing. The visualization easily helps to identify these similarities and then allows for interaction.

Studying the differences of all the Disability-RATAS, we can see how tags have been added, in which year and, then, if this happened via the wrangling process or contributor free-form tag introduction. There are some cases of tags that have been synonymized and then removed from the RATAS, only to be added in later versions as canonical tags. Figure 4.14 also shows the number of different political acts have been performed by wranglers over the

years. We can highlight an increase over all of types over time. The data obtained show that canonization is the action by wranglers that is performed most often, and mostly in recent years. On the other hand, the deletion of tags from Disability-RATAS also has shown an increase and, as shown before, brings forward new questions about these tags, to see exactly what happened to them. The deletion action can also be linked to contributors as well, through their deletion of their stories or their updating of them (e.g., and removing or changing tags). This dataset will allow further study with observations like this. Then this will also be linked to the co-creation and collaboration between authors and wranglers.

4.2.3.3 Limitations

It is necessary to highlight that the analysis of the comparison and generation of the tag report is performed using the Basic version of the *t*-RATAS networks. The module `TagScra` needs to be expanded to consistently mine the data needed for the Expanded and Fully Expanded versions for the past years. This will allow us to see whether it is possible to improve the results and insights from the analysis. Still, the derived results demonstrate the utility of the developed tool.

The Wayback Machine snapshots were incomplete (i.e., some URLs which were needed to access certain canonical tags were missing, which were needed in order to expand the networks for older versions). This limits the results, since we were not in a position to do comparisons of all the fully expanded versions of the Disability-RATAS over all of the different snapshots. There is information loss about the full slate of all of the political acts that have been performed over the years. It is not possible to obtain all the tags created, nor all of the synonymizations, as there is no access to the syn-sets of all canonical nodes in the older versions of the RATAS. Not being able to scrape the full expanded versions of the *t*-RATAS for all of the prior years compromises the tracking of changes over time for a great number of tags.

In the future, improvements in `TagScra` could be made to do extended scraping at least for the canonical tags for which the Wayback Machine does happen to capture (and to indicate more explicitly the canonical tags for which the Wayback Machine does not happen to capture and thus for which there is information loss). Additionally, the can be handled better prospectively, since new datasets (different snapshots) for specific topics can be built now, proactively for the future use. The module is already set up for this: it can be set up to scrape data periodically (e.g., once a month), to obtain data on an on-going basis of certain RATAS, to capture their full evolution.

4.3 Findings: Co-creation in AO3, Study of Story-Set and RATAS

In this section, we focus on the third cluster of research questions, with a focus on Q3b:

> Q3b: Which technique(s) could be used to detect bilateral paths of influence that may have occurred over the past years of AO3 activity?

4.3.1 Demonstration of the Technique

In order to generate the results for Q3b, we will first make use of the techniques developed in section 3.5.3, which function to answer Q3a below:

> Q3a: Which actions by fan-fic contributors impact the wranglers (and, conversely, which actions by wranglers impact fan-fic contributors) and how can these actions be illustrated through operations on the network?

In this section, we discuss the identification of actions that will constitute different paths of influence: the influence of fan-fic contributors on wranglers, and the influence of wranglers on fan-fic contributors. In the next section (section 4.3.2), we will combine multiple paths, in order to formulate bilateral paths (e.g., when each of fan-fic contributors and wrangler influence one another, which serves to co-create the ATAS).

In section 3.5.3, we defined certain operations in the network that illustrate different bilateral paths of influence in the RATAS. When analyzing the changes over time in a RATAS we can spot certain operations that show co-creation in the evolution of the network such as the deletion and addition of tags by creators which directly affects wranglers' decisions, as discussed in the previous section.

In this section, we combine both the Expanded Disability-RATAS (containing all the syn-sets related to the tag Disability) and the Disability-Story-Set (as mined with `DaScra`). In combining these, we will investigate any features that reflect this co-creation.

Section 3.3.2 described the technique for scraping story-sets. Then, using the Expanded and Fully Expanded versions of Disability-RATAS (already described in the prior sections),

we studied how the story's tags and Additional Tags relate to the contributor usage, to see the actions that indicate co-creation and collaboration between wranglers and authors. In order to mine the Disability-Story-Set, we used `DaScra` with the input:

```
>python .\ao3_analyser.py -ds 'Disability' -o story_set_disability
```

This invocation generates the Disability-Story-Set (as per section 3.2.3.1) (i.e., the set of stories returned by AO3 when Disability is used as a search term). From this, we generated a Disability-Story-Set that contained 45444 fan-fic stories.

4.3.1.1 Anomalous Non-Use of Disability-RATAS tags

From the Disability-Story-Set, we can extract the induced tagset (as defined in section 3.2.3.3 and described in section 3.3.2) by extracting the tags under the "Additional Tags" category in the metadata that was obtained for the story set. Once we have this data, we can proceed to classify all of these extracted tags with respect to their category in the Disability-RATAS network (as either canonical, syn, free-form). If any extracted tag does not appear in the Disability-RATAS network, this signals an anomaly. We found six that were anomalous.

This analysis highlights these six specific stories, for which none of their tags related to the Disability-RATAS, even though the were part of the Disability-Story-Set.

- Story Number 1: contains the tag except the mutism that's the entire fic, which did not appear in the Disability-RATAS at the time of Story-Set extraction. Later, this tag became synonymized to Muteness. This shows the latency between the contributor's free-tagging and the wrangler processing. This shows how fast the archive gets updated with deposited fan-fics and the lag in wrangler's behaviour. The story was updated on 27 Apr 2022 and, by May 2022, the tag appeared as synonymized.
- Story Number 2: was not in English and was not further analyzed.
- Story Number 3: contains the tag Deaf Seungmin which was synonymized to Deaf Kim Seungmin but was not captured in the Disability-RATAS when mining it.
- Story Numbers 4, 5, and 6: contain the additional tag Selectively Mute Runner Five which, similarly, was synonymized to selective mute Runner 5 but was not captured in the RATAS at the time of Story-Set mining.

These results show the evolution of tags within the "Additional Tags" category and behaviour of tag wrangling. However, these results are not too specific to the co-creation. Even so, this analysis revealed some anomalies that leads to another topic of interest about the recency of updates of tags in the stories and their reflection in the RATAS.

4.3.1.2 Anomalous Gaps in the Induced Tagset of Disability-Story-Set

We also performed a different analysis using the induced tagset from the Disability-Story-Set. We isolated the nodes in in the Fully Expanded Disability-RATAS which were leaves. These 'leaf tags' are the tags that were canonized by a wrangler and they not positioned to be a meta-tag (i.e., they were not parent node and thus did not group other child tags). We would expect every leaf tag to be in the induced tagset from Disability-Story-Set (after all, they got canonized by wranglers who reviewed the contributor's free-tags; if they do not appear, how did they get canonized?). We then checked to see whether this was the case. We found three leaf tags did not appear in the induced tagset from Disability-Story-Set. Their analysis is as follows:

- The tag Mute Michael Clifford currently has no work attached to it. This could mean that after the tag was canonized (as the successor of Mute Michael), its stories were updated or deleted. This clearly shows how contributors directly affect the RATAS and impacts its objective, which is to structure the archive, categorize the stories and facilitate the search for them. It is reflected in the wranglers' future action since they canonized a tag that is not in use anymore.

- The tags Blind Albedo (Genshin Impact) and Deaf Kim Seunmin not only are leaf tags, but also currently have syn-sets. But these tags presently do not have any works attached to them. This also brings the possibility that these tags formerly had attached works, but those works are now deleted or had their tags updated.

This analysis reveals some tags of interestm reflecting how the actions creators affect the meaning (and coherence) of the AO3 tag structure and, with it, the future actions of the wrangler. Similarly, the action of wranglers, canonizing or synonymizing tags, can affect the creator's behaviour to update or delete already-used tags.

Figure 4.15 shows the presence of the tags from the Fully Expanded Disability-RATAS in the Induced Tagset of Diability-Story-Set. This is a visualization of the usage of the RATAS tags within the Story-Set. Blue nodes represent the presence of the tag in the induced tagset

of Diability-Story-Set. Orange nodes show the absence of the tag in the induced tagset of Diability-Story-Set. This immediately highlights several tags for further analysis, to examine the possibility of bilateral paths, as shown before in their appearance in the archive. For example, we can see many syn tags with no appearance in the story set, which might have been caused by initial misspellings by fan-fic contributors that they subsequently corrected after the tag was synonymized by a wrangler.

Presence of Tags from the Fully Expanded Disability-RATAS in the Induced Tagset of Diability-Story-Set

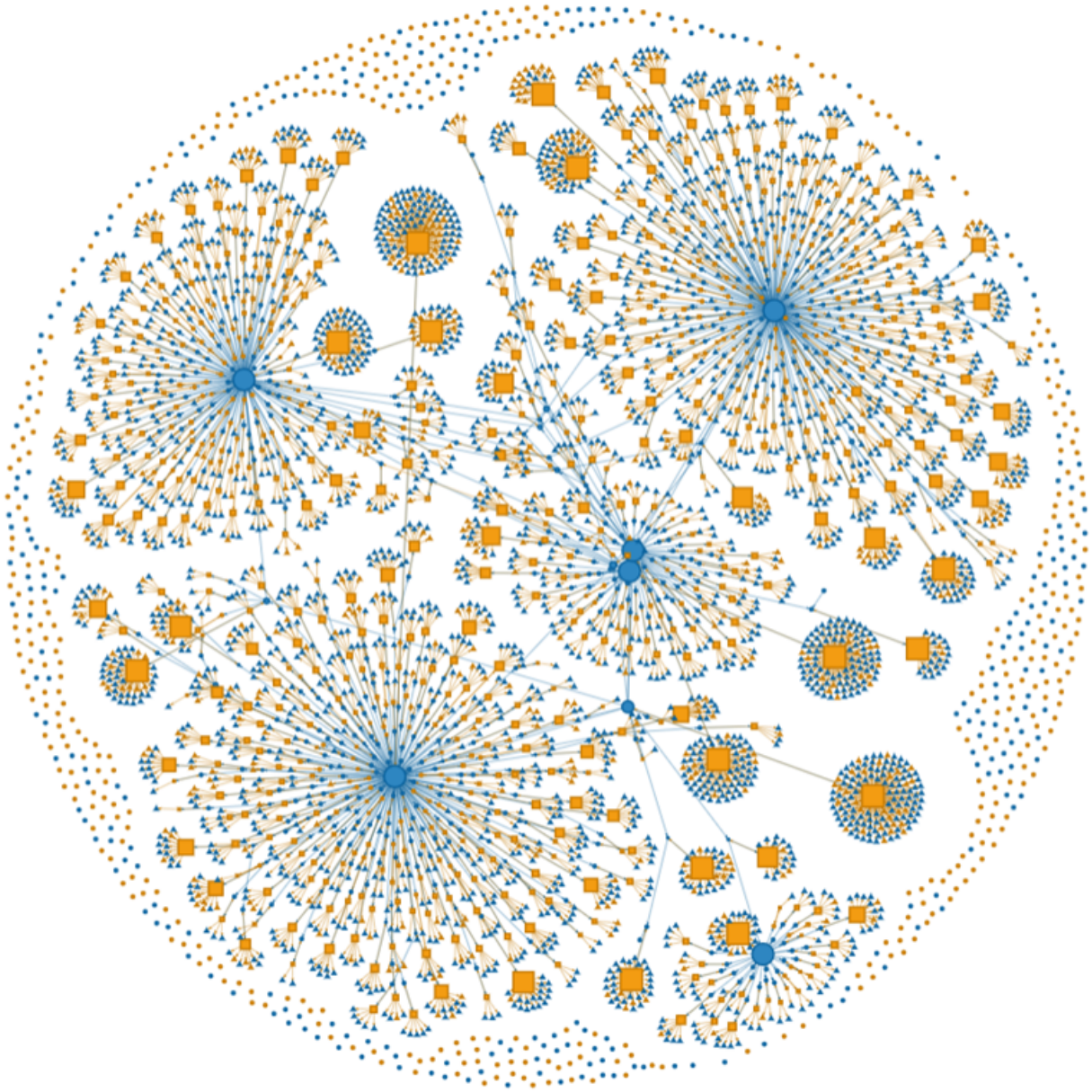


Figure 4.15: Usage of the tags from Fully Expanded Disability-RATAS. Blue nodes represent the presence of the tag in the induced tagset of Diability-Story-Set. Orange nodes represent the absence of the tag in the induced tagset of Diability-Story-Set

4.3.2 Discussion: Dynamical Properties

Having mined the induced tagset of Disability-Story-Set and the Disability-RATAS, we then investigated the ratio of Disability-related tags usage by type and by years.

4.3.2.1 Overall Usage of Additional Tags

Figure 4.16 shows a histogram of the number of tags in the "Additional Tags" category per story. Out of all stories, 13.537 are tagged with between five to nine tags. The maximum number of tags a story has is 443. Most of the stories use some in between 7-20 tags, with the upper fence of the box plot equal to 37.

Histogram, Number of tags in the "Additional Tags" category per story

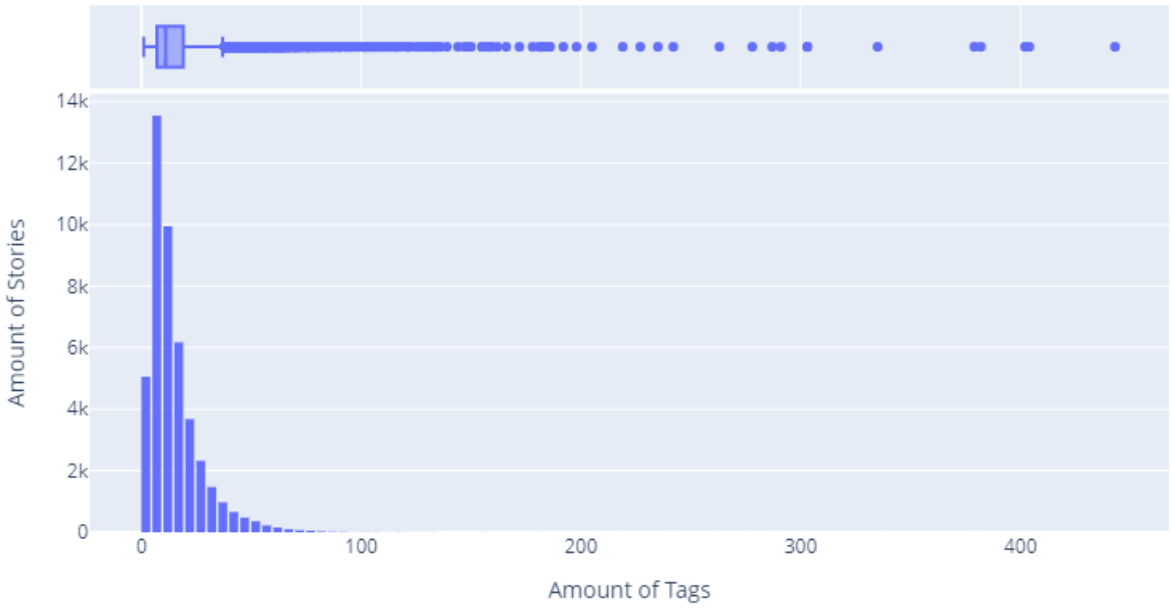


Figure 4.16: A graph showing a histogram of the number of tags in the "Additional Tags" category per story, with most stories being tagged with between five to nine tags.

4.3.2.2 Overall Usage of Tags Related to Disability

An analysis of the use of tags from the Disability-RATAS on a per-story basis for the year 2022 can be seen in Figure 4.17. The graph details the appearance of these tags in the stories within the Diability-Story-Set. Aside from the six stories with no disability tags (already described in section 4.3.1.1), there were 35,100 stories that used only one tag from the Disability-RATAS. At the high end, there were four stories that used 11 tags from the Disability-RATAS. The ratio of Disability-RATAS tags to all tags more generally in the induced tagset (i.e., tags that belong to any RATAS) was only 1.86%. Of this 1.86% of Disability-related tags, 82.3% were canonical tags, 17% were syn tags and the rest, 0.67%, were free-form tags.

Histogram of the use of the tags from the Disability-RATAS Tag on a per story basis, 2022

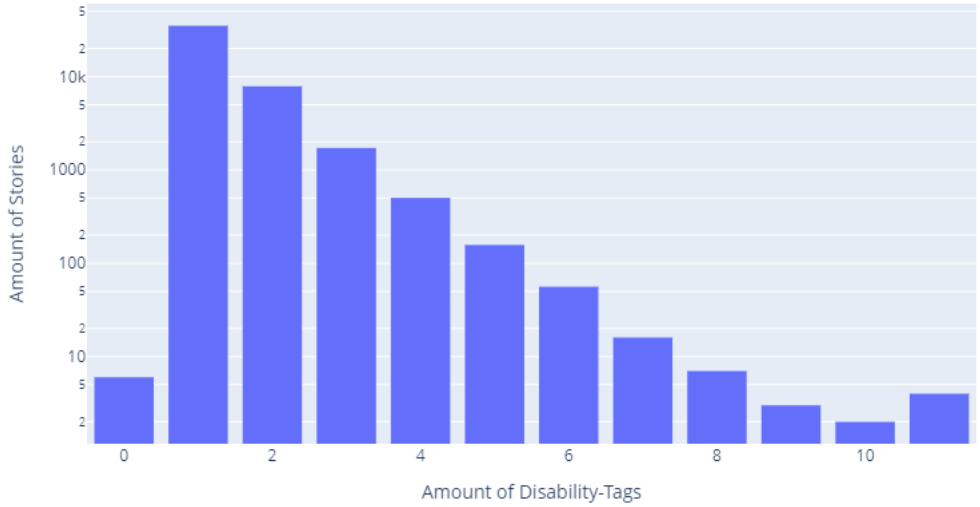


Figure 4.17: Histogram of the use of the tags from the Disability-RATAS Tag on a per story basis.

4.3.2.3 Overall Usage of Tags Related to Disability Over Time

An analysis was performed of the use of tags from the Disability-RATAS on a per-story-set basis over multiple years. Figure 4.18 shows the number of tags from the Disability-RATAS used by the Disability-Story-Set on a per-year basis. The graph shows an exponential increment over time and shows how authors are increasingly using Disability-related tags and are increasingly contributing to the RATAS development. This re-state the development of the Disability-RATAS as it is visible over the years.

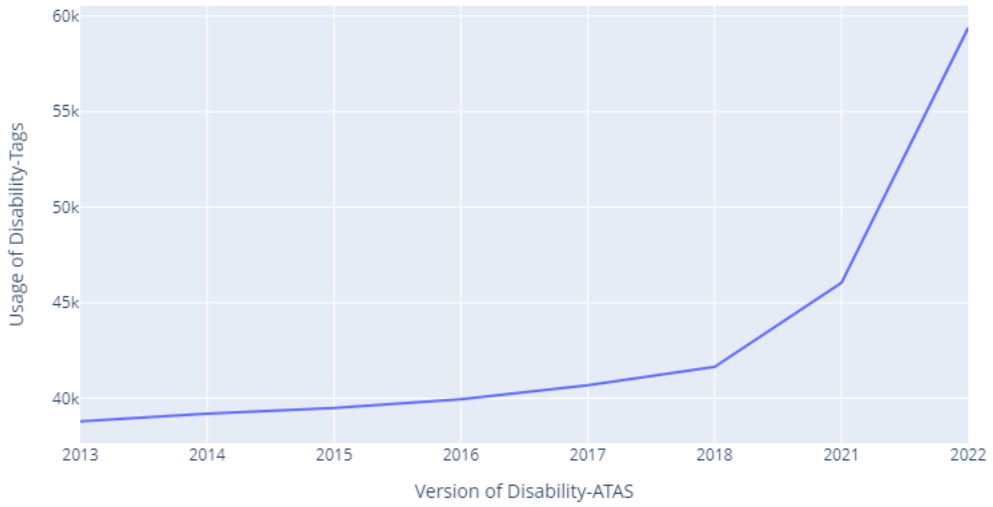


Figure 4.18: Number of tags from Disability-RATAS used by the Disability-Story-Set on a per-year basis.

4.3.2.4 Story-Set & Disability-RATAS Networks

As defined in section 3.2.3.2, we can develop different network structures to analyze the co-creation over time of the wranglers and creators with respect to a particular topic of interest. Section 3.2.3 described the definition of different networks with stories as nodes and edges representing the relationship of sharing a tag of a specific classification regarding a node. The analysis of how all the stories connect over time and seeing how the networks expand can provide new insights for further investigation. Section 3.4.2.2 showed how to generate and visualize the Story-Set & Disability-RATAS Networks. This section also discusses possible future work to expand the analysis with the tool implemented.

Firstly, in order to generate the visualizations of these networks (as defined in section 3.4.3), we use the module `NetVisualizer` as follows:

```
>python .\ao3_analyser.py --GenerateFullTagNetwork
    ./rata_disability_current_expanded_full.json
    ./story_set_disability.xlsx
```

This invocation generates the Story-Set & Disability-RATAS Full Tag Network (described specifically in (eqn:3.2.13)). This network was built with a subset of the Story-Set because of

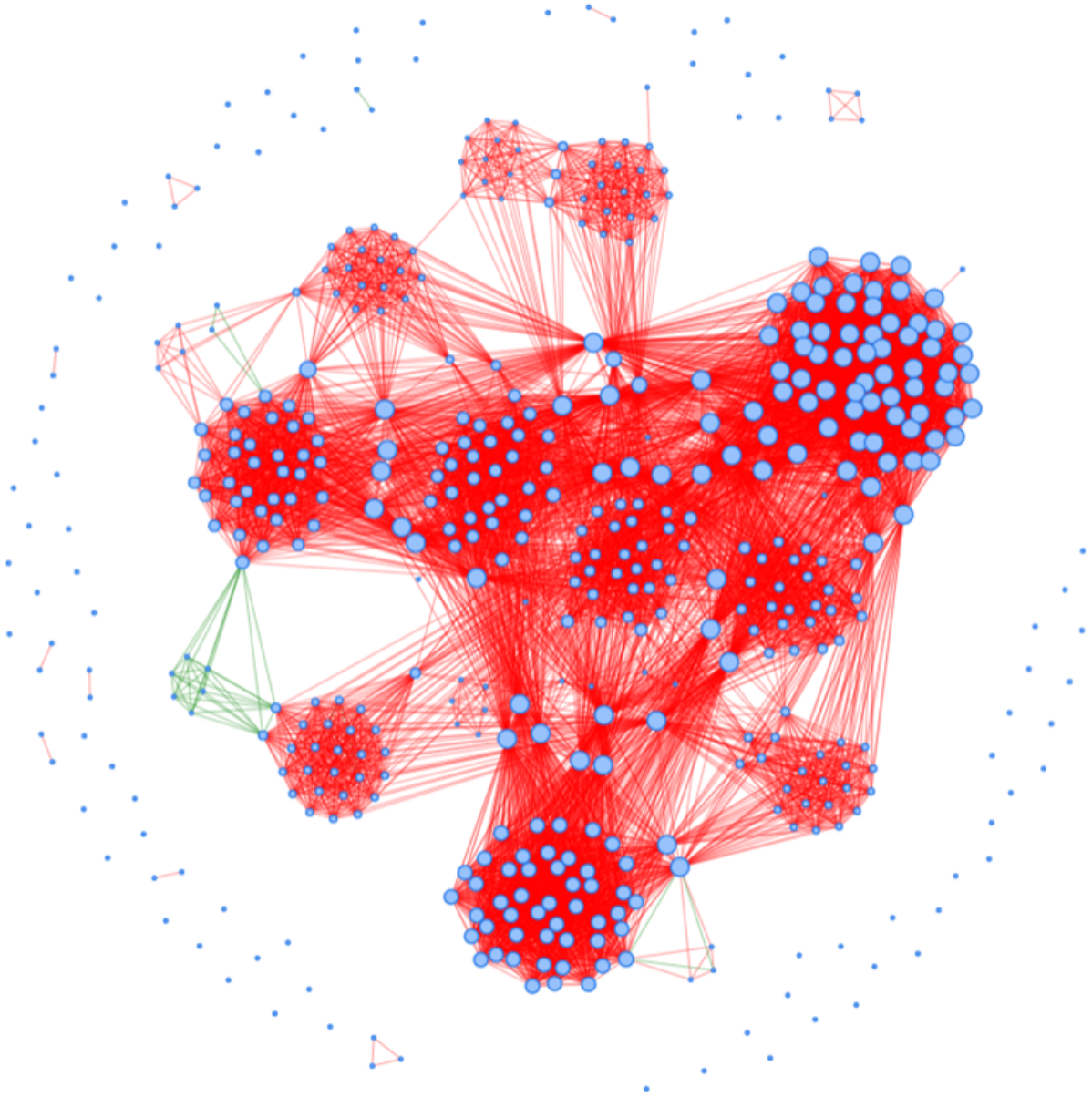


Figure 4.19: Disability-Story-Set & Disability-RATAS Full Tag Network. Green edges connect nodes (stories) that share a syn tag from Disability-RATAS. Red edges connect nodes that share a canonical tag from Disability-RATAS

.

the large number of stories and their vast usage of tags. For this, the subset was defined as the most recent 499 stories. Figure 4.19 shows a visualization of this network. It is composed of 499 nodes (stories) and 9528 edges (indicating the common use of a Disability-related tag).

Similarly, we can generate the Story-Set & Disability-RATAS Canonical Tag Network. We generated this network using the Basic Disability-RATAS from 2013 The network built from the 2013 Disability-RATAS was provided with the the Disability-Story-Set from 2013 and before.

It was built with the following command prompt:

```
>python .\ao3_analyser.py --GenerateCanonicalTagNetwork
    ./rata_disability_2013.json
    ./story_set_disability.xlsx
```

The generated network contains 343 nodes and 8999 edges. Edges in this network mean that two stories share a canonical tag from Disability-RATAS (2013). A visualization of this network is shown in Figure 4.21.

We then generate the Story-Set & Disability-RATAS Canonical Tag Network using the Fully Expanded Disability-RATAS from 2022 and the 2022 Disability-Story-Set. It was built with the following command prompt:

```
>python .\ao3_analyser.py --GenerateCanonicalTagNetwork
    ./rata_disability_current_expanded_full.json
    ./story_set_disability.xlsx
```

The generated network is composed of 499 nodes and 9489 edges, as shown in Figure 4.20. We can see an increment in the connectivity over the years, even though these visualizations are made on subsets from the story-set. Also, when we compare Figure 4.21 and Figure 4.20, we see that the 2013 version contains less stories (nodes) and fewer isolated nodes, whereas the 2022 version has more stories and isolated nodes.

For completeness, we generated, in a similar fasion, the Disability-Story-Set & Disability-RATAS Syn Tag and Free-Form Tag Networks, but these are omitted for brevity. These visualization (using the same sub-set of stories) do not show much connectivity (or not even any connectivity for the Free Form Tag Network). These visualizatons showed connectivy for different subsets of stories for which there were more appearances of the free-form tags and syn tags. It was shown before free-form tags have the lowest representation in the Story-Set, as creators tend to use more canonical/syn tags while tagging.

All these networks bring to the table promising future work. The study over time of them and their properties can bring new points of view to the analysis, as they explicitly represent

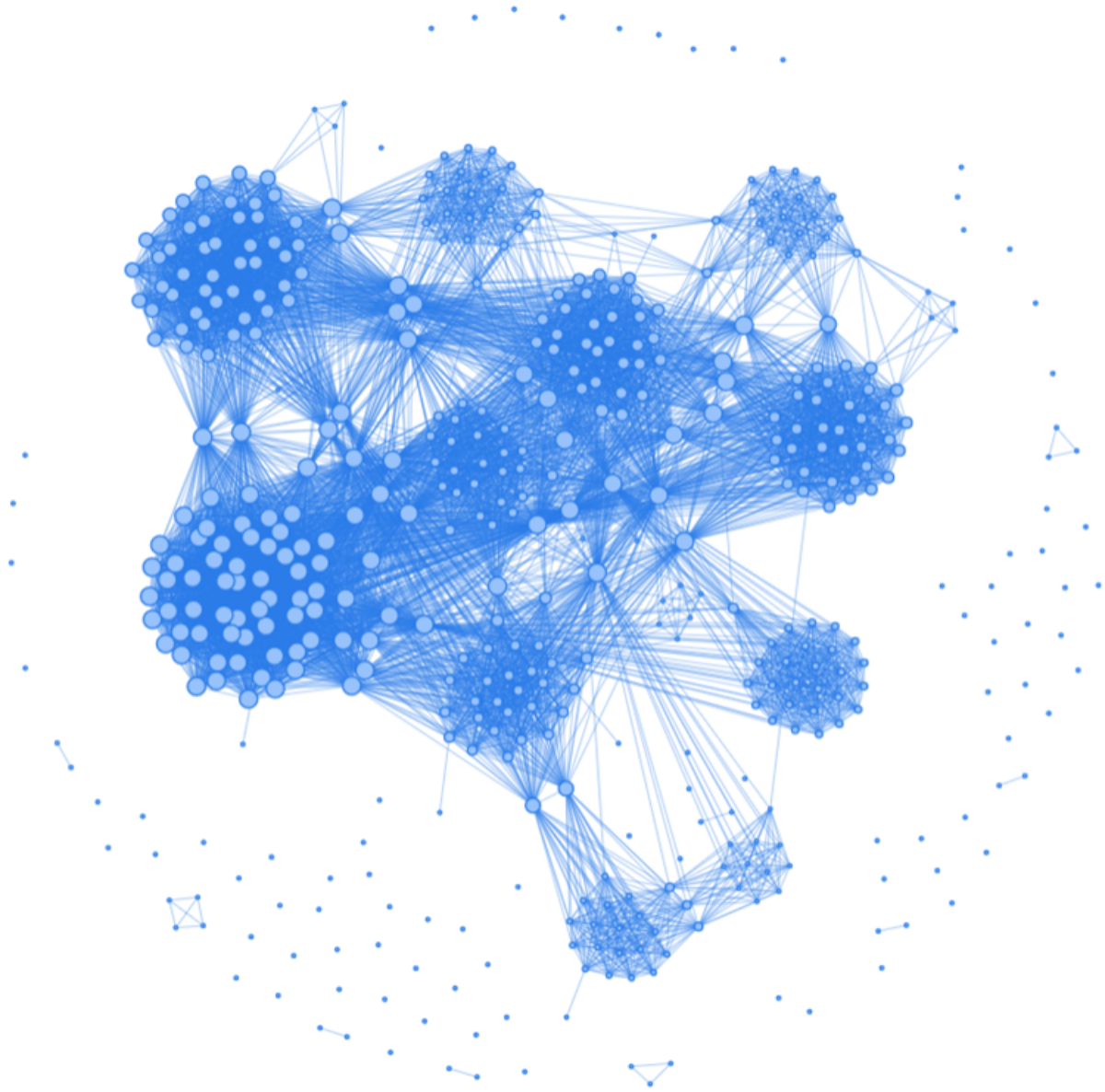


Figure 4.20: Disability-Story-Set 2022 & Disability-RATAS 2022 Canonical Tag Network. Blue edges connect nodes that share a canonical tag from Disability-RATAS.

the connections of stories by the wrangling process. This means that two stories will connect to each other in one specific network when the stories acquired a tag in common, either by addition from the creator or by the wrangling process.

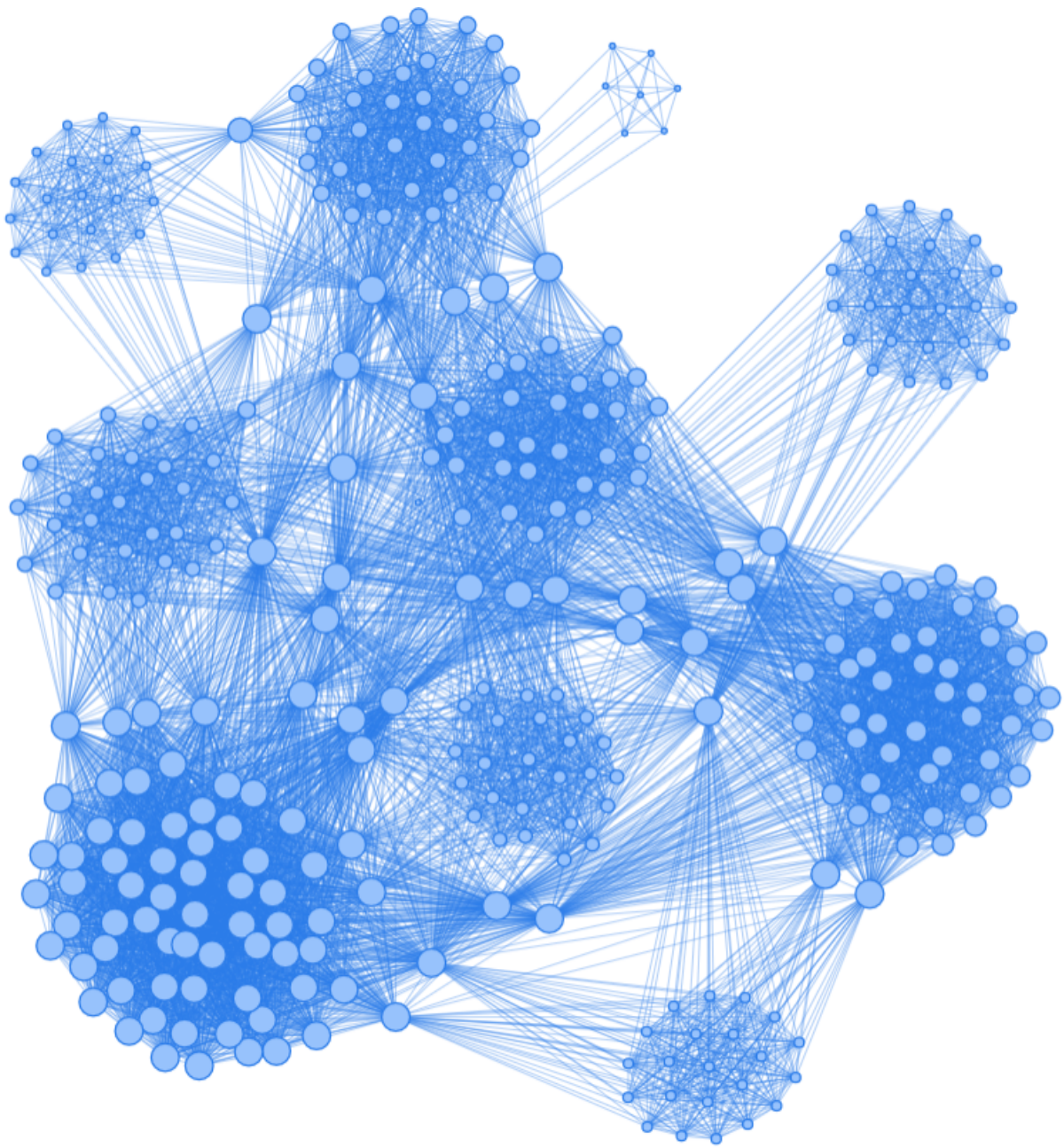


Figure 4.21: Disability-Story-Set 2013 & Disability-RATAS 2013 Canonical Tag Network. Blue edges connect nodes that share a canonical tag from Disability-RATAS.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

The motivation for this thesis project was the interest of tagging to scholars in fields such as Media Studies, Communication Studies and Science and Technology Studies. Scholars in these field analyze and investigate community behaviour (including tag wranglers, fan-fic contributors and others), with a focus on the practical politics around the classification of contested concepts, such as race, gender roles, heteronormativity, homophobia, and ableism. The thesis focused on AO3 and its tagging system, with its unique structure. Chapter 1 defined the goal of the thesis project, to develop knowledge outcomes and knowledge products that can be taken up in an interdisciplinary manner, with a goal that the results be accessible and of utility to researchers outside of Computer Science — researchers who study fan-works, tags, and the resultant tag archive structures as cultural products. Chapter 1 concluded by setting the following three main questions to direct the thesis research: How to obtain information from AO3 and represent it to further analysis; How to capture this information from different points in time and study these snapshots comparatively; and How to spot actions of co-creation and collaboration between the different types of users (contributors, wranglers, readers) in the AO3 archive.

Chapter 2 sets out to provide the necessary background needed for the thesis project. It provided in-depth detail of the AO3 tagging system and the core concepts of interest in relation to the AO3 archive and website. The tag structure and the tag wrangling process were articulated formally, in terms of network structures and network operations (as a necessary precursor for their computational representation). Chapter 2 provided a synthesis of the related literature, which covered the topics of network analysis and its application on AO3 or

different folksonomies, data extraction and visualization tools and general studies around AO3 and other similar platforms. From this synthesis, ideas for possible techniques for the quantitative analysis of tag networks were generated. Then, network analysis emerged as the main analysis technique for studying the structure of AO3's tags. Once the nomenclature and concepts from the research literature had been established, the chapter concludes with a re-articulation of the research questions from Chapter 1 in a more formal and precise manner.

Chapter 3 described the methodology used to pursue the three categories of research questions, which entailed scraping data from the AO3 website, building different network structures that represent the tagging system, and developing network visualization and analysis techniques. To develop the methodology, a set of eight objectives were identified as being instrumental to the research questions. The work was staged into three main phases, in order to tackle these objectives.

- Phase 1, corresponding to section 3.2, sets the formal nomenclature and network definitions of several different structures of interest to be constructed using the data to be scraped from AO3. It also entailed defining, in terms of network operations, the actions (both individually and in sequence) that are performed by creators and wranglers as part of the tagging practices in AO3. This work sought to establish the connection between the politics of user actions (in the sense of users having power to impact the tag structure in AO3) and specific network operations. This work also sought to represent formally the co-creation and collaboration processes that occur among the users (fan-fic contributors and wranglers) in the archive.

- Phase 2, corresponding to section 3.3, developed the computational techniques to scrape the necessary data for the network structures, both from the current AO3 archive and from snapshots accessed through the Wayback Machine. This involved processing and storing several different types of data in different datasets, in order to support different types of visualizations and analyses. Two modules of the main tool were developed: `DaScra`, reported in section 3.3.2, and `TagScra`, reported in section 3.3.3.

- Phase 3, which corresponds to section 3.4, developed the computational techniques to build and visualize the network structures that were formally defined in Phase 1, using the datasets that can be obtained using the Phase 2 techniques. This phase entailed the development of two modules: `NetBuilder` — which is in charge of building the network structures —, and `NetVisualizer` — which uses the computational network representation and created visualization of them.

- Phase 4, which corresponds to section 3.5, described the development of the different techniques to analyze and mine quantitative results from Phase 2 datasets and the Phase 3 computational network representations. In this phase, there are three main focii of study: the study of the network as a static entity (section 3.5.1); the comparative analysis of networks over time time (section 3.5.2); and, lastly the investigation of co-creation and collaboration among the different AO3 users (fan-fic contributors and wranglers) (section 3.5.3). To support these analyses, a module called `NetAnalizer` was developed, which, combined with the previous modules, is able to extract quantitative results, to visually represent points of comparison, and to create and store new datasets highlighting substructures in the network and properties of interest for further investigation.

Chapter 4 described the application of the developed computational techniques for a specific component of the AO3 tag structure: the structure concerned with the tag Disability. The chapter described the different structural properties and actions that users perform in AO3, as extracted from the AO3 site and then visualized using the different RATAS formulations. The chapter presented an analysis of the Disability-RATAS development over the years, showing the usefulness of the developed tools and techniques to compare two or more snapshots of a given t-RATAS. Lastly, the chapter articulated, in terms of network operations, the paths of co-creation and collaboration among wranglers and creators in AO3. The chapter illustrated how to use the tools and different techniques to identify anomalies in the network structure using the Story-Set & RATAS Networks, useful for identifying political acts over the years. Chapter 4 provided answers to each of the research questions set out in section 2.3, using an approach of meeting the objectives defined in section 3.1. This chapter demonstrates the successful application of the project methodology, resulting in the implementation of an extensible tool that achieves the main goals of the present study.

5.2 Limitations and Future Work

Throughout the development and application of the methodology used in this thesis project, limitations have been noted. In this section, these limitations are summaries and discussed, with an emphasis on proposed future work to overcome these limitations and to extend the work done in this research project.

5.2.1 Phase 1: Representation Development

This phase focused on the development of the nomenclature for the AO3 tagging system and its tagging structures (see section 3.2). As previously mentioned, there are no formal definitions provided by the developers of AO3, nor any full access to the tag structures. Our definitions and representations were developed deductively. Thus, these definitions and representations can be improved with the benefit of further research about AO3 and with additional sources of information about AO3, either from the Organization of Transformative Works (the host of AO3) or from other researchers. We also defined all networks around the Additional Tag category, but we can appreciate the potential importance of other tag categories, such as Character, Relationship, and Fandom. The existence of tag structures from all of the different categories will make the whole archive tagging structure more complicated and bigger. This study only focused on the Additional Tag category, but we recommend future work on the formal definition of the whole AO3 tagging structure in general, including all fo the different categories of tags.

5.2.2 Phase 2: Scraping

This phase described the implementation and workflow of two modules, `DaScra` and `TagScra` in section 3.3. In general, there are several avenues of future work and many possible extensions, and specifically in terms of making use of the Wayback Machine. This future work is described below:

DaScra

Section 3.3.2 mentioned limitations on the implementation and methodology of `DaScra`. This module is in charge of obtaining and storing the metadata for a set of stories. `DaScra` at present operates only on the current version of the AO3 archive, as accessed via the AO3 website. It can be extended. We propose to develop a connection to the Wayback Machine. This will allow for the scraping of fan-fic story sets from archival datasets (Wayback Machine snapshots). This data will show the fan-fic stories as they were at earlier points in time. This will afford the possibility to highlight all the changes a contributor has done in terms of story tagging over time. Analysis in this project has demonstrated this practice among contributors. This functionality will also make possible the other forms of analysis, for looking for co-creation and collaboration among the users in AO3 and will produce a more complete

illustration of this emergent phenomena. In this project, the Story-Set & RATAS networks were constructed and provided the basis for this type of analysis. With longitudinal data, it will become possible to compare the Story-Set & RATAS networks over the years. This will allow for any bilateral path of contribution in the archive between wranglers and contributors to be identified more easily.

Another limitation described in section 3.3.2 was the functionality of AO3 in terms of obtaining the stories corresponding to specific search terms. AO3 only shows 100000 stories in its search results. When searching popular tags, this threshold is often overcome by high numbers of search results. For example, one of the most popular Fandom in the archive is "Super Natural (2005 TV)" which, as of this moment, contains 254457 fan-works (e.g., 2.5 times larger than the largest possible search results). In order to successfully scrape all the stories from this tag, the module `DaScra` would need to be modified (e.g., by adding some filters in the process of scraping the data, such as doing a year by year search). Thus, a proposed modification would be to to scrape any story set year by year, applying filtering to it, as AO3 searching allows. This way, we can successfully obtain all the metadata for all fan-work of a tag search in the archive at a time, irrespective of search set size constraints. An additional extension would be to allow the further population of partial datasets (e.g., incremental scraping: to continue scraping from where it has already scraped and left off).

TagScra

A main limitation of this module concerns its interaction with AO3 through the Wayback Machine. Even though it is possible to obtain a snapshot of a t-RATAS from a point back in time, it is not fully implemented for all years, nor for the search of any tag. We noted fragile functionality while scraping the expanded version of a RATAS and the non-canonical tags to its full expanded version. Further research about the functionality of the Wayback Machine is needed in order to successfully scrape this data. It was observed that the Wayback Machine does not contain all the information of the archive for all years. While searching specific sites within the Wayback Machine, if they are missing, the site tends to redirect to a previous or following point in time. We propose to further study this behaviour in order to implement refined functionality, such as being able to at least scrape partial versions of the structures of interest and to signal more explicitly to the `TagScra` user when the Wayback Machine is making an out-of-year substitution.

Additionally, we also propose to use `TagScra` for the routine generation of datasets on an on-going basis. It is possible to periodically use the module as it is and start capturing

different snapshots of the different structures of interest for a certain topic. This will allow more future studies of the scraped data for specific topics and reduce the future reliance on the Wayback Machine.

Furthermore, the computational tools and the performed study focused on Basic t-RATAS, its expanded version and its fully expanded version for single values of t. However, it would be be possible to combine different tags of interest and add more iterations while their expansion to obtain broader levels of interconnections. Given the access to the main ATAS has not to be built, and it seems to be large enough to practically be scraped, it can be approximated with the use of `TagScra`. For future work, we recommend adding more levels of expansion to a certain Expanded RATAS while also adding more focus tags to their structures. This will enrich future studies of the AO3 Tagging structure with more accurate properties.

5.2.3 Phase 3: Network Building and Visualizing

Phase 3, which was described in section 3.4, focused on building the computational network representation of the networks, as defined in Phase 1, and generating the visualizations of these networks. Two modules were implemented: `NetBuilder` and `NetVisualizer`. Future work for each of them will be covered in detail below.

NetBuilder

This module, described in section 3.4.2.1, is in charge of building the network structures (to afford further analysis) and for the visualization of the networks. We propose to further expand this module, by adding in modules to create different properties of interest in the networks. These can bring new points of view to the network analysis. Currently, we are focused on building the Directed Network for the RATAS, and the possibility was also mentioned of non-directed networks as a representation. The Story-Set & RATAS Networks can also be expanded by adding weights to the edges, such as to indicate the usage of the tags shared among fan-works. Additionally, we can add different types of edges to this network, indicating other relationships among two nodes (fan-works) when they share more than one type of tag. Future studies around the possibilities of extending the network representation of all the defined networks will be triggered by other questions that arise in the network analysis stage.

The module builds the Fully Expanded t-RATAS relying on string matching with term t. Then there may be free-form tags that are related to t that do not use the tag string e.g.,

“Deaf" and “Anacusis", which is a syn tag, as shown in figure 4.4. This approache exclude many different tags that semantically relate to each other but not syntactically.

NetVisualizer

The module, described in section 3.4.3, is in charge of generating the visualizations of the networks. It was developed using `Python` and `Pyvis`. A limitation of this library is that it does not support the creation of a legend. We can expand the code in order to create legends manually, which will create more usability to the visualizations. At present, any legends must be constructed manually. We also recommend further investigation of which properties or structures of the networks are of most interest to researchers, in order to develop visualizations that best highlight the aspects of interest. Changes in the layout and adding more flexibility can be great features in the future. For example, time-lapse could be added to visualize the changes in a given network over the years (reflecting longitudinal data sets). Another feature of interest could be to develop better approaches for the comparison of two snapshots of the same network. At present, this is accomplished by using a pair of network visualizations (A compared to B, and B compared to A: for example, see Figure 4.8 and Figure 4.9. Two visualizations could be combined into one, allowing the comparison in the same visualization instead of two different ones.

5.2.4 Phase 4: Network Analysis

Phase 4, described in section 3.5, focused on the analysis of the different network structures, both via report generation and via visualization. This work has many avenues for furure work, particularly in terms of additional network analysis techniques.

The `NetAnalizer` module could be extended to investigate many more properties of networks, such as power-law degree distribution, six-degree separation, the strength of weak ties, and shrinking diameter. It also could be extended to focus on different connected components or sub-networks within the networks. We also flag for future work the study of the mathematical models they represent. Prior research literature reports on the presence of scale-free and small-world models in different tagging structures on the internet (Shen and Wu, 2005), other than AO3. These properties are not yet investigated for AO3. Additionally, future work could investigate the utility of community detection algorithms and link prediction, which might open new opportunities to develop AO3 functionality. Link prediction could potentially be useful to improve the recommendation of tags to fan-fic contributors and to

assist wranglers in their tasks (e.g., recommendations for performing wrangling certain action on specific tags). Community detection also has the potential to enrich and expand the studies of the AO3 tagging system. Locating communities of tags while studying the RATAS will spot certain patterns or substructures of interest to study by researchers. It also can enrich any future recommendation system.

The are many different algorithms to measure and detect communities in networks, as summarized in section 2.2.2.3 of the literature review. This work was done in connection to folksonomies other than AO3. There is much future work in the application of these techniques to AO3. Many prior works employed Newman-Girvan's modularity as one of the most popular approaches. This measure and other connected measures are not reported in the current analysis. This project has as its focus, indirect community-level interaction through political acts of tagging. THis stands in contrast to direct community-level interaction. The further investigation of community-level interaction, through both indirect and direct means, is an obvious direction for future work. This work could leverage the structures defined within this study, and future work could focus on deep analysis on community detection on different RATAS and also in the Story-Set & RATAS networks.

In addition, this research project focused primarily on the the RATAS and their versions. The Story-Set & RATAS Networks were investigated and analyzed, but they were not deeply studied and represent many directions for future work. We have left this the study of these networks for future work, such as their static, diffing analysis, looking for the development over time of the whole archive. Also, more deep study is needed to look for signs of co-creation, focusing on their changes over time. Investigation of their static and dynamic properties have the potential to generate new forms of evidence of these bilateral patterns in the political actions by wranglers and contributors.

All these directions for future work, and many other directions not yet identified, are possible using this current work as its basis. But any future work must be done in connection to the interests of the researchers studying AO3 in fields such as Media Studies, Communication Studies and Science and Technology Studies. To support this connection, ongoing and future work must be conducted to establish interdisciplinary collaborations.

Appendix A

Code Documentation

In order to interact with `ao3_analyser` it was implemented a CLI. We can see its help with all its arguments in the following example:

```
usage: ao3_analyser [-h]
    [-ts TAGSCRA [TAGSCRA ...]] [-ao3_v AO3VERSION]
    [--NonCanonical] [--ExpandRATAS EXPANDRATAS]
    [--ExpandFullRATAS EXPANDFULLRATAS EXPANDFULLRATAS]
    [-ds DASCRA] [--StoryNumber STORYNUMBER] [-o OUTPUTNAME]
    [-op OUTPUTPATH] [-gn GENERATENETWORK]
    [--GenerateFullTagNetwork GENERATEFULLTAGNETWORK
        GENERATEFULLTAGNETWORK]
    [--GenerateSynTagNetwork GENERATESYNTAGNETWORK
        GENERATESYNTAGNETWORK]
    [--GenerateCanonicalTagNetwork GENERATECANONICALTAGNETWORK
        GENERATECANONICALTAGNETWORK]
    [--GenerateFreeFormTagNetwork GENERATEFREEFORMTAGNETWORK
        GENERATEFREEFORMTAGNETWORK]
    [--StorySetYear STORYSETYEAR]
    [-gs GENERATESIMILARITES GENERATESIMILARITES]
    [-gd GENERATEDIFFERENCES GENERATEDIFFERENCES]
    [--GenerateRATASProperties GENERATERATASPROPERTIES]
    [--GenTagDiff GENTAGDIFF GENTAGDIFF]
    [--GenTagDiffDir GENTAGDIFFDIR]
```

```
optional arguments:
  -h, --help            show this help message and exit
  -ts TAGSCRA [TAGSCRA ...], --TagScra TAGSCRA [TAGSCRA ...]
            Call TagScra Module with a list of focus tags as
            arguments to mine. Generate a JSON file with the
            RATAS.
  -ao3_v AO3VERSION, --AO3Version AO3VERSION
            TagScra argument specifies the version (year) for
            which TagScra accesses AO3 using the Wayback Machine.
  --NonCanonical
            TagScra argument that mines and generates all non
            canonical tags related to the focus tag received as
            an argument.
  --ExpandRATAS EXPANDRATAS
            Call TagScra to expand a RATAS. It receives a RATAS
            JSON file mining all canonical tags for its expansion.
            It outputs a RATAS JSON file with the Expanded RATAS.
  --ExpandFullRATAS EXPANDFULLRATAS EXPANDFULLRATAS
            Call TagScra to fully expand a RATAS. It receives an
            Expanded RATAS JSON file and a TXT with all non-
            canonical tags. It outputs a JSON file with the Fully
            Expanded RATAS.
  -ds DASCRA, --DaScra DASCRA
            Calls DaScra Module with the Tag to mine all its fan-
            works and generates an xlsx file with the Story-Set
            metadata.
  --StoryNumber STORYNUMBER
            DaScra argument to set the number of fan-work to scrap
  -o OUTPUTNAME, --OutputName OUTPUTNAME
            Set the name of the file generated file
  -op OUTPUTPATH, --OutputPath OUTPUTPATH
            Set the path to store the generated file
  -gn GENERATENETWORK, --GenerateNetwork GENERATENETWORK
```

```
          Build a RATA and generates its visualization. It receives
          a RATAS JSON file. Outputs and shows the visualization
          of the RATAS.
--GenerateFullTagNetwork GENERATEFULLTAGNETWORK, GENERATEFULLTAGNETWORK
          Build the Story-Set & RATAS Full Tag Network and
          generates its visualization. It receives a RATAS JSON
          file and a Story-Set. Outputs and shows the
          visualization of the Full Tag Network.
--GenerateSynTagNetwork GENERATESYNTAGNETWORK, GENERATESYNTAGNETWORK
          Build the Story-Set & RATAS Syn Tag Network and
          generates its visualization. It receives a RATAS JSON
          file and a Story-Set. Outputs and shows the
          visualization of the Syn Tag Network.
--GenerateCanonicalTagNetwork GENERATECANONICALTAGNETWORK,
GENERATECANONICALTAGNETWORK
          Build the Story-Set & RATAS Canonical Tag Network and
          generates its visualization. It receives a RATAS JSON
          file and a Story-Set. Outputs and shows the
          visualization of the Canonical Tag Network.
--GenerateFreeFormTagNetwork GENERATEFREEFORMTAGNETWORK,
GENERATEFREEFORMTAGNETWORK
          Build the Story-Set & RATAS Free-Form Tag Network and
          generates its visualization. It receives a RATAS JSON
          file and a Story-Set. Outputs and shows the
          visualization of the Free-Form Tag Network.
--StorySetYear STORYSETYEAR
          Set the year of the Story-Set & RATAS Networks. It
          filters the Story-Set for all the fan-works before the
          year input as an argument.
-gs GENERATESIMILARITES, GENERATESIMILARITES,
  --GenerateSimilarites GENERATESIMILARITES, GENERATESIMILARITES
          Generates the visualizations highlighting the
          similarities between two RATAS. It receives two RATAS
          JSON files assuming the first is the older version.
```

```
-gd GENERATEDIFFERENCES GENERATEDIFFERENCES,
  --GenerateDifferences GENERATEDIFFERENCES, GENERATEDIFFERENCES
          Generates the visualizations highlighting the
          differences between two RATAS. It receives two RATAS
          JSON files assuming the first is the older version.
--GenerateRATASProperties GENERATERATASPROPERTIES
          Generates a report with the main properties of a
          snapshot of a RATAS. It receives a RATAS JSON files.
          It outputs an xslx file with different information about
          the RATAS.
--GenTagDiff GENTAGDIFF GENTAGDIFF
          Generates a report with the differences between two RATAS.
          It receives two RATAS JSON files assuming the first
          one is the older version. It outputs an xslx file with
          the differences between the RATAS.
--GenTagDiffDir GENTAGDIFFDIR
          Generates a report with the differences between a set
          of RATAS. It receives a directory full of RATAS JSON
          files assuming they are sorted from older to the most
          recent version. It outputs an xslx file with the
          differences between the RATAS.
```

Appendix B

Example of Scraped Data

B.1 JSON RATAS

Example of part of the `JSON` file generated after scraped Cored Disability-RATAS dated to 2013 using `TagScra` plus Wayback Machine:

```
{
    "Disability": {
        "type": "canonical_tag", "parent_tags": ["No Fandom"],
        "synned_tags": [
        "Disabilities", "Disability Issues",
            "disability themes", "disabilty", "Disable",
            "disabled!stuck", "Dsiable", "Handicap",
            "impairment","Permanant Injury of Main Character"
            ],
        "subtags": [
            "Deaf", "Blindness", "Mute Loki"
            { "Deaf Character": [
                    "Deaf Dave", "Deaf Clint",
                    "Deaf Clint Barton", "Deaf Derek",
                    "Deaf Derek Hale", "Deaf John",
                    "Deaf Bilbo", "Deaf Castiel",
                    "Deaf Dean Winchester", "Deaf Stiles"]},
            "Physical Disability", "Permanent Injury",
            "Disabled Character", "Canon Disabled Character",
            "Learning Disabilities", "Dyslexia","Muteness",
            { "Blind Character": [
                    "Blind Stiles", "Blind Sherlock",
                    "Blind Dave", "Blind Roy",
                    "Blind Harry", "Blind John",
                    "Blind Sam", "Blind Castiel",
                    "Blind Loki", "Blind Dean",
                    "Blind Derek", "Blind Sollux Captor"] },
        ],"metatags": [""]
    }
}
```

Figure B.1: Disability-RATAS dated to 2013.

B.2 Story-Set Metadata

Example of the Story-Set scraped using the “Disability" tag for it definition on `DaScra`. It is displayed the first 100 stories and split into two tables for clarity: Table B.1 and Table B.2.

Title	AdditionalTags	ArchiveWarning	Author	Bookmarks	Category	Chapters	Characters	Comments	Fandom
Ice-Creamed	[Vaginal Sex, Vaginal Fingering, Penis In Vagin...	No Archive Warnings Apply	RevenantIceCream	NaN	F/M	1/1	Neopolitan (RWBY), Oscar Pine	NaN	RWBY
Bounty hunting is a complicated profession	[Alternate Universe - Canon Divergence, Mutenes...	Choose Not To Use Archive Warnings	Sandzlib	123	F/M	51/?	Din Djarin, Reader, Cara Dune, Omera (Star Wars...	513	The Mandalorian (TV)
The Light In Your Eyes	[Domestic Fluff, Light Angst, Blindness, Summer...	Choose Not To Use Archive Warnings	kyeomsbb	17	M/M	27/?	Kim Mingyu (SEVENTEEN), Jeon Wonwoo, Wen Jun Hu...	17	SEVENTEEN (Band)
We never lost control	[Post-Arcane: League of Legends Season 01, Angs...	Choose Not To Use Archive Warnings	Hamlette	58	Gen, M/M	12/?	Viktor (League of Legends), Jinx (League of Leg...	90	Arcane: League of Legends (Cartoon 2021)
Blind Silence	[Mentioned Clay | Dream (Video Blogging RPF), h...	Choose Not To Use Archive Warnings	satetown	1	Gen	1/1	TommyInnit (Video Blogging RPF), Wilbur Soot, J...	NaN	Minecraft (Video Game)
Harry Potter and the Sanctuary Wards	[Canonical Child Abuse, Child Neglect, Harry Po...	Graphic Depictions Of Violence, Rape/Non-Con	charlaine2124	46	Multi	11/30	Harry Potter, Draco Malfoy, Severus Snape, Pans...	79	Harry Potter - J. K. Rowling
The Cards We're Dealt	[Slow Burn, Slow Build, Fluff, Hurt*s*Comfort, ...	No Archive Warnings Apply	ImaLittleInsane	24	F/M	120/?	Garrus Vakarian, Female Shepard (Mass Effect), ...	413	Mass Effect Trilogy
The Eye of the storm	[Unrequited Love, Canon Disabled Character, One...	No Archive Warnings Apply	Deilmo	5	M/M	3/?	Silco (Arcane: League of Legends), Viktor (Leag...	2	Arcane: League of Legends (Cartoon 2021)
The Grand Piano	[Smut, PWP, ok there is some P, Hurt*s*Comfort,...	No Archive Warnings Apply	allisonbucky	NaN	F/M	1/1	NaN	NaN	Daredevil (TV)
World We Must Defend.	[Crossover, Action*s*Adventure, Mystery, Other ...	No Archive Warnings Apply	Seychelles	4	Gen	3/?	Kaburagi T. Kotetsu, Barnaby "Bunny" Brooks Jr....	3	Tiger & Bunny
Stitch By Stitch	[Other Additional Tags to Be Added, Slow To Upd...	Graphic Depictions Of Violence	Rozegholst	22	No category	8/?	TommyInnit (Video Blogging RPF), Phil Watson (V...	5	Dream SMP - Fandom
Impossible Things	[Post-Reichenbach, Sherlock Being Sherlock, Myc...	No Archive Warnings Apply	mom2boys	11	M/M	22/?	Enola Holmes, Sherlock Holmes, John Watson, Ori...	61	Sherlock (TV)
The Adventure Begins with, a yes	[Happy Ending, Slow Burn, Sibling Rivalry, Whee...	No Archive Warnings Apply	Chaoticgood_Magizoologist	2	F/M, F/F, Multi, M/M	5/?	Harry Potter, Original Characters, Nymphadora T...	2	Harry Potter - J. K. Rowling
Hoping to inspire.	[Chronic Illness, Physical Disability, Wheelcha...	Choose Not To Use Archive Warnings	IceShiro	3	M/M	5/5	Simon (Young Royals), Wilhelm (Young Royals), E...	8	Young Royals (TV 2021)
Be my Guest	[Rizzles, Sleepovers, Sleepy Cuddles, Morning C...	No Archive Warnings Apply	JulieVerne	38	F/F	102/?	Maura Isles, Jane Rizzoli, Bass Isles, Jo Frida...	375	Rizzoli & Isles
thoughts on the aftereffects of delta-particle ...	[Pining, Touch-Starved, Intimacy, Non-Sexual In...	No Archive Warnings Apply	winter_writes	NaN	Other	1/1	Leto Atreides II, Hwi Noree, Nyshae (Dune), Kie...	NaN	Dune Series - Frank Herbert
Immovable meets Unstoppable	[Alternate Universe, Coming of Age, i guess*q**...	No Archive Warnings Apply	Multi_Elle	7	M/M	5/5	Chwe Hansol | Vernon, Boo Seungkwan, Yoon Jeong...	4	SEVENTEEN (Band)
Spiral	[Canon Disabled Character, Disability, Canon-Ty...	Graphic Depictions Of Violence, Major Character...	the_stars_in_our_eyes	4	F/M, Gen, M/M, Multi	11/?	Eddie Brock, Venom Symbiote (Marvel), Cletus Ka...	50	Venom (Comics)
Wisteria	[Knifeplay, Bloodplay, No Beta We Die Like 1*d*...	Graphic Depictions Of Violence	WarCriminalEnjoyer (jinchiten)	25	F/F	42/?	Yotsuyu goe Brutus, Warrior of Light (Final Fan...	217	Final Fantasy XIV
we are going to start transing genders, be scared	[Agender Ranboo (Video Blogging RPF), Trans Tob...	Choose Not To Use Archive Warnings	The_Neverland_System	2	Gen, M/M, Multi, Other	3/?	Clay | Dream (Video Blogging RPF), GeorgeNotFou...	NaN	Dream SMP - Fandom
Finding Love Forever	[Smut, Romance, Anal Sex, Drama, Angst, Byun Ba...	No Archive Warnings Apply	Ladyli1987	3	M/M	3/30	Byun Baekhyun, Park Chanyeol, Oh Sehun, Kim Jon...	NaN	EXO (Band)
Quiet...Too Quiet	[Kidnapping, Missing Persons, Injury, Amputatio...	No Archive Warnings Apply	HuneyBuzz	NaN	No category	23/?	Cherry Cookie (Cookie Run), Gumball Cookie (Coo...	10	Cookie Run (Video Game)
Lost & Found (Korok Edition)	[Green (Linked Universe)-centric, Selectively M...	No Archive Warnings Apply	ooopo123	6	Gen	1/1	Four (Linked Universe), Wild (Linked Universe),...	3	The Legend of Zelda & Related Fandoms
Sunflower [podfic]	[Post-Episode: s15e18 Despair, Castiel Saves Hi...	No Archive Warnings Apply	NerdyNerdenstein	1	M/M	1/1	Castiel (Supernatural), Dean Winchester, Jack K...	NaN	Supernatural (TV 2005)
The Iron Giant	[Team as Family, Found Family, Adoption, Foster...	Choose Not To Use Archive Warnings	ArdenDrifter	69	Multi, F/M, Gen, M/M	60/?	Jamie Tartt, Roy Kent, Keeley Jones, Ted Lasso,...	7774	Ted Lasso (TV)
No Worse A Fate	[Daniel*s*Jack in an asexual relationship, Disa...	Major Character Death	See_U_At_Th3_Crossroads	1	M/M, Other	16/26	Samantha "Sam" Carter, Cameron Mitchell (Stargate)	70	Stargate SG-1
Growing Pains	[trans!remus, Trans Male Character, pregnant tr...	No Archive Warnings Apply	NaN	5	M/M	1/1	Remus Lupin, Sirius Black, James Potter, Harry ...	2	Harry Potter - J. K. Rowling
The silent Prince	[Alternate Universe - Royalty, Modern Royalty, ...	Choose Not To Use Archive Warnings	Avrina	1	F/M, M/M	7/?	Original Characters, Original Female Character(...	6	No Fandom
Muted Love	[Mute Sasuke, mental health, Alternate Universe...	No Archive Warnings Apply	tachiequillsluv	15	F/M, M/M, Multi	26/28	Uzumaki Naruto, Uchiha Sasuke, Uchiha Itachi, H...	11	Naruto
What Might be Found	[Aloy (mentioned) - Freeform, Erend (mentioned)...	No Archive Warnings Apply	spite_cadet	1	F/M	1/1	Kotallo (Horizon: Forbidden West)	6	Horizon: Zero Dawn (Video Game)
So Many Things to Say	[Foster Care, Foster Care AU, Teen Peter Parker...	No Archive Warnings Apply	happyaspie	50	No category	3/?	Peter Parker, Tony Stark, Pepper Potts, Ned Lee...	69	Marvel Cinematic Universe
In the Dark, Part 1	[Major Character Injury, Blindness, Protective ...	No Archive Warnings Apply	Lexie_Raven	18	F/F, F/M, Gen	15/18	Kara Danvers, Alex Danvers, Lena Luthor, Maggie...	40	Supergirl (TV 2015)
And If My Lover Will Not Hear It	[this is basically "what if the ritual in scotl...	No Archive Warnings Apply	elleditchhorror	8	M/M	1/1	Martin Blackwood, Jonathan "Jon" Sims | The Arc...	9	The Magnus Archives (Podcast)
Narcotic	[Sickfic, Sick Character, Character Death, Impl...	Major Character Death	a home for the lost (Delilah_Dust)	NaN	M/M	5/5	Adeline Kane Wilson, Joseph Wilson, Elliot Knig...	8	Teen Titans (Animated Series)
In Perfect Technicolor	[Alternate Universe - Historical, San Francisco...	Choose Not To Use Archive Warnings	crisptrepidation	4	F/F, M/M	9/?	Sam Evans, Artie Abrams, Quinn Fabray, Santana ...	28	Glee
All the Quiet Nights You Bear	[Fluff and Angst, Angst, Emotional, Disabled Ch...	No Archive Warnings Apply	regenderate	9	F/F	11/29	Thirteenth Doctor, Rose Tyler, Yasmin Khan, Naj...	49	Doctor Who (2005)
A Sick-Ass Kong Vault Straight Through the Godd...	[Alternate Universe - Human, Humanstuck, Mute D...	No Archive Warnings Apply	godtiermeme	8	M/M, Gen	14/?	Karkat Vantas, Dave Strider, John Egbert, Kanay...	82	Homestuck
Cigarettes, Daydreams	[Hero Worship, Internalized Homophobia, Sibling...	Graphic Depictions Of Violence, Major Character...	busyyhead	3	M/M	8/?	Optimus Prime, Original Male Character(s), Will...	12	Transformers - All Media Types
Looking for a Kiss	[Alternate Universe - Modern Setting, Alternate...	No Archive Warnings Apply	agrajag	17	M/M	1/1	Stede Bonnet, Blackbeard | Edward Teach, Lucius...	6	Our Flag Means Death (TV)
Child of Lecter	[Canon-Typical Violence, Child Abuse, Child Neg...	Graphic Depictions Of Violence, Major Character...	TanaChan	342	F/M, M/M	14/?	Will Graham, Hannibal Lecter, Harry Potter, Ver...	298	Hannibal (TV)
A conspiracy of barely functional people (or: h...	[Friends to Lovers, Idiots to still idiots, Hur...	No Archive Warnings Apply	sulkybender	6	M/M	1/1	Sokka (Avatar), Zuko (Avatar), Katara (Avatar),...	16	Avatar: The Last Airbender
Children of Mine	[Childhood Friends, Friendship, Whump, Muteness...	Graphic Depictions Of Violence	TMS33	5	F/M, Gen	1/?	Ymir Fritz, Eren Yeager, Mikasa Ackerman, Armin...	2	Shingeki no Kyojin | Attack on Titan
All is fair in Love and Chaos	[Alternate Universe - Modern Setting, Original ...	No Archive Warnings Apply	Guewsy	2	F/M	66/?	Kaz Brekker, Matthias Helvar, Inej Ghafa, Jordi...	2	Six of Crows Series - Leigh Bardugo
Friendship bracelets	[Childhood Friends, Developing Friendships, Fou...	Choose Not To Use Archive Warnings	blackbatfan	24	Gen, Other	5/10	Cassandra Cain, Stephanie Brown, Tim Drake, Duk...	28	Batman - All Media Types
Orange is a Burning Color	[Fluff and Angst, Deaf Character]	Choose Not To Use Archive Warnings	Ace_is_gay	NaN	F/F, F/M	7/?	Liam Stewart, Cole Stewart, Ruby Daly, Charles ...	NaN	The Darkest Minds Series - Alexandra Bracken
Learning New Strengths	[Alternate Universe - College*s*University, Dea...	Choose Not To Use Archive Warnings	ERERIWORLD	1	M/M	8/?	Levi Ackerman, Armin Arlert, Eren Yeager, Erwin...	24	Shingeki no Kyojin | Attack on Titan
Coming Out Of His Cage (And He's Not Doing Just...	[Foster Care, Adoption, Adopted Sibling Relatio...	Graphic Depictions Of Violence	radio_chemicals	9	F/F, F/M, Gen, M/M, Multi, Other	36/36	Original Characters	17	Original Work
The Boys Who Wouldn't Grow Up	[Alternate Universe, swap au, Alternate Univers...	Choose Not To Use Archive Warnings	MissLavenderLady	3	Gen, Multi	6/?	David (Lost Boys), Michael Emerson (Lost Boys),...	8	The Lost Boys (Movies)
A Different Sort of Impossible	[Hogwarts Eighth Year, Harry Potter Epilogue Wh...	Graphic Depictions Of Violence	TeaInACoffeeMug	192	F/M	24/?	Hermione Granger, Draco Malfoy, Blaise Zabini, ...	634	Harry Potter - J. K. Rowling
Unstoppable Forces and Immovable Objects	[Essek Thelyss Needs a Hug, Relationship Negoti...	No Archive Warnings Apply	koschei	18	M/M	6/?	Essek Thelyss, Caleb Widogast	31	Critical Role (Web Series)
All We've Lost	[Developing Relationship, Developing Friendship...	Choose Not To Use Archive Warnings	myjokerbias	6	M/M	18/?	Kanda Yuu, Allen Walker, Lavi (D.Gray-man), Len...	NaN	D.Gray-man
dreaming about forever (only gets you so far)	[Alternate Universe - Human, Mistaken Identity,...	No Archive Warnings Apply	FlyofDragon	3	M/M	9/23	Alex Manes, Kyle Valenti, Max Evans, Maria DeLu...	34	Roswell New Mexico (TV 2019)
Disc-ord	[Alternate Universe - Superheroes*s*Superpowers...	Graphic Depictions Of Violence	Dark_Side_S1lver	6	Gen	3/?	Wilbur Soot, Technoblade (Video Blogging RPF), ...	1	Minecraft (Video Game)
To Become a Dragon's Hearth	[Interspecies Romance, Developing Relationship,...	Graphic Depictions Of Violence	JasperUnbound	NaN	F/M	1/1	Caim (Drag-On Dragoon), Red Dragon | Angelus, V...	2	Drag-On Dragoon | Drakengard
Alone	[chaos-company's Angstpril 2022, Day 21, "Don't...	No Archive Warnings Apply	Lif61 (UltimateFandomTrash)	NaN	F/M	1/1	Kanan Jarrus, Hera Syndulla	NaN	Star Wars: Rebels
Please just let me go (don't let me go)	[Ranboo-centric (Video Blogging RPF), Winged Ra...	Graphic Depictions Of Violence, Rape/Non-Con, U...	Bossisded	12	M/M	5/?	Ranboo (Video Blogging RPF), Ranboo's Father (V...	42	Minecraft (Video Game)
delirium.	[Regulus Black-centric, Alternate Universe - Mu...	Choose Not To Use Archive Warnings	SagaEllen	7	M/M	6/6	Regulus Black, Sirius Black, James Potter, Remu...	6	Harry Potter - J. K. Rowling
What Was Needed	[Violence, Child Abuse, Slavery, Kuruta survivo...	Graphic Depictions Of Violence	Serenechaos	342	Gen	61/?	Kurapika (Hunter X Hunter), Genei Ryodan | Phan...	942	Hunter X Hunter
Throuple Pairings one shots	[Hurt*s*Comfort, Suicide Attempt, Suicidal Thou...	Choose Not To Use Archive Warnings	PurpleMagicMusic	1	F/F, F/M, M/M	5/?	NaN	NaN	9-1-1 (TV)
Old Bones Aflame	[Alternate Universe - Canon Divergence, Jungles...	Choose Not To Use Archive Warnings	TalesOfOnyxBats	23	Gen	22/?	Azula (Avatar), Hama (Avatar)	165	Avatar: The Last Airbender
Jersey on my mind	[POV Daryl Dixon, Daryl Dixon Smut, Alexandria ...	Graphic Depictions Of Violence, Major Character...	stonecoldjerseyfox	31	F/M	49/?	Daryl Dixon, Aaron (Walking Dead), Morgan Jones...	35	The Walking Dead (TV)
Like a Crateful of Grenades	[Avengers Tower, Fake*s*Pretend Relationship, A...	No Archive Warnings Apply	luulapants	120	M/M	5/7	Matt Murdock, Frank Castle, Natasha Romanov (Ma...	236	Daredevil (TV)
Your silent words	[Violence, Alternate Universe - Boarding School...	Graphic Depictions Of Violence, Underage	Callybal	5	M/M	1/?	Will Graham, Hannibal Lecter	3	Hannibal (TV)
To Be Loved	[Alternate Universe - Modern Setting, Strippers...	No Archive Warnings Apply	its_icarus	2	M/M, Multi	5/?	Aphrodite (Hades Video Game), Hypnos (Hades Vid...	2	Hades (Video Game 2018)
And If You Were Drowned At Sea	[Jordie Rietveld Lives, Kaz Brekker is Bad at F...	Choose Not To Use Archive Warnings	Freshwaterseas	5	Gen	3/23	Kaz Brekker, Jordie Rietveld, Jesper Fahey, Wyl...	34	Six of Crows Series - Leigh Bardugo
Through the Sparks, the Flame	[Mute Upgraded Connor | RK900, Feral Upgraded C...	Choose Not To Use Archive Warnings	PennyFifer	64	M/M	10/15	Gavin Reed, Upgraded Connor | RK900, Connor (De...	76	Detroit: Become Human (Video Game)
The Dragon Speaker	[Action*s*Adventure, Alternate Universe - Fanta...	Graphic Depictions Of Violence	Wishfulthinking1979	11	F/M	25/30	Firmus Piett, Maximilian Veers, The Lady, Anaki...	342	Star Wars - All Media Types
Broken Bones [Spring - Season One]	[Slice of Life, Past Character Death, Post-Trau...	Choose Not To Use Archive Warnings	5Castaways5	15	F/M	21/21	Harvey (Stardew Valley), Robin (Stardew Valley)...	52	Stardew Valley (Video Game)
Five minus one	[Alternate Universe - Fantasy, Fantasy, Mystery...	Major Character Death, Graphic Depictions Of Vi...	Fremioda	1	F/M, M/M, Multi	14/?	Eren Yeager, Eren Yeager's Parents, Mikasa Acke...	2	Shingeki no Kyojin | Attack on Titan
Within his sight	[Serial Killer Michael Myers, Blind Character, ...	Graphic Depictions Of Violence, Major Character...	Whistle_78	11	F/M	97/?	Michael Myers, Original Female Character(s), Dr...	29	Halloween (Movies 1978-2002)
I Had To Save You	[Alternate Universe - High School, Murder, Atte...	Graphic Depictions Of Violence, Rape/Non-Con, U...	Jayfur08	1	F/F, F/M, Gen	7/?	Catherine of Aragon, Catherine Parr, Katherine ...	38	Six - Marlow/Moss
Into the Wild - Thistledown AU	[Alternate Universe - Canon Divergence, Fix-It ...	Choose Not To Use Archive Warnings	rainfoxarts	6	No category	24/27	Firestar (Warriors), Smudge (Warriors), Oakhear...	12	Warriors - Erin Hunter
Bad Faith	[Emotional Hurt*s*Comfort, Time Travel Fix-It, ...	Graphic Depictions Of Violence	Henna_Star00	135	No category	7/?	Draco Malfoy, Harry Potter, Hermione Granger, S...	182	Harry Potter - J. K. Rowling
blinded by (fraternal) love	[Clay | Dream and Toby Smith | Tubbo are Siblin...	Graphic Depictions Of Violence	raccoon_bubbles_killer	7	Gen	7/10	Clay | Dream (Video Blogging RPF), Toby Smith |...	22	Minecraft (Video Game)
The Bastard Saint Of The Barrel	[Grisha Kaz Brekker, Jordie Rietveld Lives, Kaz...	No Archive Warnings Apply	Spikey44	82	F/M, M/M, Multi	34/?	Kaz Brekker, Inej Ghafa, Nina Zenik, Matthias H...	536	Six of Crows Series - Leigh Bardugo
The Emperor That Burned and the God That Remain...	[Fire Emblem: Three Houses Spoilers, Alternate ...	No Archive Warnings Apply	AutumnVine	136	F/F	18/?	My Unit | Byleth, Edelgard von Hresvelg, Rhea (...	363	Fire Emblem: Fuukasetsugetsu | Fire Emblem: Thr...
For Your Ears Only	[Moving Out, College, selective mutism, Sign La...	No Archive Warnings Apply	shewasagaystripper	5	M/M	1/3	John Deacon, Freddie Mercury, Brian May, Roger ...	3	Queen (Band)
Fighting For Control	[Moon is named Moonyrise, Moon is Not Evil (Fiv...	Graphic Depictions Of Violence, No Archive Warn...	Rose_Demon	17	F/M, Other	5/20	Insert is Reader, Sunnydrop - Character, Moonyd...	13	Fnaf: SB
let down your hair and i'll let down my gun	[Slow Burn, Physical Disability, Post-Traumatic...	Graphic Depictions Of Violence, No Archive Warn...	solivaganteros	NaN	F/M	1/52	Ajay Ghale, Noore Najjar, Bhadra (Far Cry), Yum...	NaN	Far Cry 4
Aurachnid - Whumptober 2021	[Whumptober 2021, Suspension, Whump, Hurt*s*Com...	Graphic Depictions Of Violence	amphetamine_keen	NaN	F/F, Gen	10/31	Original Characters, Original Female Character(...	NaN	Original Work
Kidza's Adventures in the Amazing World of Deat...	[Crack Treated Seriously, Angst, Heavy Angst, I...	Graphic Depictions Of Violence	GrassMud	6	No category	9/?	Clay | Dream (Video Blogging RPF), Technoblade ...	23	Video Blogging RPF
crystalline reflections over rose flavoured ice	[Post-Traumatic Stress Disorder - PTSD, They*s*...	No Archive Warnings Apply	kvrobas	11	Gen	2/?	Galacta Knight, Kirby (Kirby), Meta Knight, mos...	14	Kirby (Video Games)
Breaths We Dare to Take	[warning for a LOT of heavy tags, Terminal Illn...	Graphic Depictions Of Violence	void_spells	17	Multi	5/7	Jayce (League of Legends), Viktor (League of Le...	44	Arcane: League of Legends (Cartoon 2021)
30 Days, 30 Stories: Star Trek Drabbles	[Writing Prompt, 30 day writing challenge, Tumb...	No Archive Warnings Apply	sourirenoire	3	M/M, F/M, F/F	26/30	Spock (Star Trek), Leonard "Bones" McCoy, James...	16	Star Trek
Kaleidoscopic Grangers	[Trans Character, Trans Female Character, Trans...	Major Character Death	AdmiralPegasus	507	Gen, F/F, F/M	364/376	Harry Potter, Hermione Granger, Ron Weasley, Ru...	3945	Harry Potter - J. K. Rowling
Blindsided	[former Sith!Obi-Wan, Former Sith!Maul, Serial ...	Choose Not To Use Archive Warnings	SLWalker	83	M/M	83/105	Obi-Wan Kenobi, Darth Maul, Original Characters...	2808	Star Wars - All Media Types
Shrimp	[Physical Disability]	Choose Not To Use Archive Warnings	Akert	NaN	No category	13/19	Samuel Drake, Nathan Drake, Elena Fisher, Victo...	NaN	Uncharted (Video Games)
Unrequited Love	[Alternate Universe - Arranged Marriage, Dubiou...	Choose Not To Use Archive Warnings	Galennox	NaN	M/M	1/1	Martin "Rekkles" Larsson, Marcin "Jankos" Janko...	NaN	League of Legends RPF
Broke	[Trauma, Hurt*s*Comfort, Modern AU, Physical Di...	No Archive Warnings Apply	nagato	NaN	M/M	2/?	Nagato | Pain, Uchiha Obito, Konan (Naruto), Ya...	4	Naruto
Avalanche	[Healer, Sigvi, Antoriel, Slow Build, Eventual ...	Choose Not To Use Archive Warnings	Viss_and_Pinegar	16	F/M	41/?	Original Altmer Character(s), Altmer (Elder Scr...	287	Elder Scrolls V: Skyrim
Transformers watching bumblebee	[bumblebee aus, bumblebee is a bean, bumblebee ...	Graphic Depictions Of Violence, Major Character...	mia_kimono2	NaN	Other, Gen	2/?	Bumblebee (Transformers), Optimus Prime, Grimlo...	NaN	Transformers: Prime
Salvation, Restoration, Liberation	[Werewolf Draco Malfoy, Werewolf Mates, Post-Wa...	Choose Not To Use Archive Warnings	Orchid_Haven	155	F/M, Gen	6/?	Hermione Granger, Draco Malfoy, Harry Potter, R...	118	Harry Potter - J. K. Rowling
Fallacy of Yesterday	[Other Additional Tags to Be Added, Winged Char...	Choose Not To Use Archive Warnings, Graphic Dep...	AfterArtist	7	No category	2/?	Charles | Grian, Ryan | GoodTimesWithScar, Oliv...	4	Hermitcraft RPF
Seeing You	[lots of miya twins fighting, samu wins all of ...	No Archive Warnings Apply	prosesareread	51	M/M	10/?	Udai Tenma, Kita Shinsuke, Ojiro Aran, Suna Rin...	74	Haikyuu!!
Anything?	[Alternate Universe - Zombie Apocalypse, Post-A...	No Archive Warnings Apply	Notmypickle	1	M/M	1/1	Lance (Voltron), Keith (Voltron), Kolivan (Volt...	NaN	Voltron: Legendary Defender
Now you see me	[Badass!Reader, Agent!Reader, Slow Burn, Phil C...	Graphic Depictions Of Violence	Desaboca	4	F/M	7/21	Reader, Nick Fury, Phil Coulson, James "Bucky" ...	6	Marvel Cinematic Universe
Can't Save Them All	[Hybrids, idek, I'm Bad At Tagging, Rescue, Goo...	Choose Not To Use Archive Warnings, No Archive ...	NaN	25	No category	7/?	Technoblade (Video Blogging RPF), Wilbur Soot, ...	22	Minecraft (Video Game)
Green	[Green Eyes, Tired Evan "Buck" Buckley, Develop...	Graphic Depictions Of Violence, Major Character...	ThirteenRedVampireBites	4	Gen, M/M	2/2	Evan "Buck" Buckley, Eddie Diaz (9-1-1 TV), Hen...	1	9-1-1 (TV)
The Piano's Missing Keys	[Alternate Universe - Foster Family, Flowers, T...	No Archive Warnings Apply	AnonymousLife	254	Gen	21/26	TommyInnit (Video Blogging RPF), Sam | Awesamdu...	218	Video Blogging RPF
Love Looks Not With the Eyes, But With the Heart	[Hurt Crosshair (Star Wars: The Bad Batch), Cro...	No Archive Warnings Apply	Yatzstar	14	Gen	6/?	Crosshair (Star Wars: The Bad Batch), Omega (St...	86	Star Wars: The Bad Batch (Cartoon)
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table B.1: First part of the Disability Story-Set. Each observation describes the “Title", “Additional Tags", “Author", “Bookmarks", “Category", “Chapter", “Characters", “Comments", and “Fandom" of a fan-work in the Story-Set.

Title	Hits	Kudos	Language	Rating	Relationship	Series	Part	SourceURL	Updated	Words
Ice-Creamed	0	NaN	English	Mature	Neopolitan/Oscar Pine	NaN	NaN	https://archiveofourown.org/works/38647278	2022-04-27	1,049
Bounty hunting is a complicated profession	26318	916	English	Mature	Din Djarin/Reader, Din Djarin/You, Din Djarin &...	Bounty hunting is a complicated profession	1.0	https://archiveofourown.org/works/29137065	2022-04-27	110,256
The Light In Your Eyes	2214	89	English	Not Rated	Jeon Wonwoo/Kim Mingyu	NaN	NaN	https://archiveofourown.org/works/35876410	2022-04-27	28,532
We never lost control	5837	389	English	Mature	Jinx & Viktor (League of Legends), Jinx & all h...	NaN	NaN	https://archiveofourown.org/works/36378655	2022-04-27	15,672
Blind Silence	25	1	English	Not Rated	Wilbur Soot & TommyInnit, Jschlatt & TommyInnit...	crimeboys	7.0	https://archiveofourown.org/works/38646114	2022-04-27	1,346
Harry Potter and the Sanctuary Wards	3798	170	English	Mature	Marcus Flint/Percy Weasley/Oliver Wood, Other S...	Sanctuary	1.0	https://archiveofourown.org/works/37141234	2022-04-27	51,266
The Cards We're Dealt	11653	123	English	Explicit	Female Shepard/Garrus Vakarian	NaN	NaN	https://archiveofourown.org/works/29812818	2022-04-27	607,773
The Eye of the storm	299	35	English	Mature	Silco (Arcane: League of Legends)/Viktor (Leagu...	NaN	NaN	https://archiveofourown.org/works/38246215	2022-04-27	12,370
The Grand Piano	18	NaN	English	Explicit	Matt Murdock/Karen Page	NaN	NaN	https://archiveofourown.org/works/38644557	2022-04-27	5,016
World We Must Defend.	79	7	English	Teen And Up Audiences	Kaburagi Kaede & Kaburagi T. Kototsu	NaN	NaN	https://archiveofourown.org/works/38130961	2022-04-27	6,574
Stitch By Stitch	1277	137	English	Teen And Up Audiences	Toby Smith | Tubbo & TommyInnit, Piglin (Minecr...	NaN	NaN	https://archiveofourown.org/works/38311231	2022-04-27	2,343
Impossible Things	1713	61	English	Mature	Sherlock Holmes/John Watson, Enola Holmes & Myc...	The Esme Variations	5.0	https://archiveofourown.org/works/33361696	2022-04-27	40,178
The Adventure Begins with, a yes	220	3	English	Teen And Up Audiences	Luna Lovegood/Harry Potter, Nymphadora Tonks/Ch...	NaN	NaN	https://archiveofourown.org/works/38313061	2022-04-27	39,114
Hoping to inspire.	1688	47	English	Not Rated	Simon/Wilhelm (Young Royals)	NaN	NaN	https://archiveofourown.org/works/37956574	2022-04-27	881
Be my Guest	34744	631	English	General Audiences	Maura Isles/Jane Rizzoli, Maura Isles & Jane Ri...	NaN	NaN	https://archiveofourown.org/works/36433231	2022-04-27	95,503
thoughts on the aftereffects of delta-particle ...	1	NaN	English	Teen And Up Audiences	Leto Atreides II/Hwi Noree	NaN	NaN	https://archiveofourown.org/works/38641995	2022-04-27	2,258
Immovable meets Unstoppable	368	24	English	Teen And Up Audiences	Boo Seungkwan/Chwe Hansol | Vernon, Minor or Ba...	NaN	NaN	https://archiveofourown.org/works/36978652	2022-04-27	17,676
Spiral	841	34	English	Mature	Eddie Brock/Venom Symbiote, Eddie Brock & Venom...	In The Night	4.0	https://archiveofourown.org/works/38096788	2022-04-27	26,421
Wisteria	5893	121	English	Explicit	Yotsuyu goe Brutus/Warrior of Light	NaN	NaN	https://archiveofourown.org/works/32867794	2022-04-27	115,219
we are going to start transsing genders, be scared	311	30	English	Teen And Up Audiences	Alexis | Quackity & Clay | Dream & GeorgeNotFou...	NaN	NaN	https://archiveofourown.org/works/38334631	2022-04-27	809
Finding Love Forever	152	8	English	Explicit	Byun Baekhyun/Park Chanyeol	NaN	NaN	https://archiveofourown.org/works/38122933	2022-04-27	3,431
Quiet...Too Quiet	788	38	English	Not Rated	Cherry Cookie & Gumball Cookie (Cookie Run), De...	NaN	NaN	https://archiveofourown.org/works/37148662	2022-04-27	3,956
Lost & Found (Korok Edition)	230	59	English	General Audiences	Four & Wild (Linked Universe), Blue & Green & R...	Wild Encounters (The Chaim Meets Wild 1v1)	5.0	https://archiveofourown.org/works/38641254	2022-04-27	1,397
Sunflower [podfic]	72	4	English	Teen And Up Audiences	Castiel/Dean Winchester	NaN	NaN	https://archiveofourown.org/works/38641041	2022-04-27	12
The Iron Giant	25196	542	English	Teen And Up Audiences	Keeley Jones/Roy Kent/Jamie Tartt, Roy Kent & J...	NaN	NaN	https://archiveofourown.org/works/34116817	2022-04-27	366,386
No Worse A Fate	587	18	English	Mature	Daniel Jackson/Jack O'Neill	NaN	NaN	https://archiveofourown.org/works/37725463	2022-04-27	29,459
Growing Pains	138	17	English	Teen And Up Audiences	Sirius Black/Remus Lupin, James Potter/Lily Eva...	NaN	NaN	https://archiveofourown.org/works/38462596	2022-04-27	6,229
The silent Prince	101	3	English	Mature	Original Female Character/Original Male Charact...	The Amaldean Crown	2.0	https://archiveofourown.org/works/36676141	2022-04-27	10,929
MuteoLove	3334	95	English	Teen And Up Audiences	Uchiha Sasuke/Uzumaki Naruto, Sai/Uzumaki Narut...	NaN	NaN	https://archiveofourown.org/works/25715272	2022-04-27	47,088
What Might be Found	195	39	English	Mature	Aloy/Kotallo (Horizon: Zero Dawn)	NaN	NaN	https://archiveofourown.org/works/38638443	2022-04-27	1,430
So Many Things to Say	3579	336	English	Teen And Up Audiences	Peter Parker & Tony Stark, Ned Leeds & Peter Pa...	Spider-man Stories: Not Otherwise Specified	20.0	https://archiveofourown.org/works/38382517	2022-04-27	11,309
In the Dark, Part 1	3043	72	English	Teen And Up Audiences	Alex Danvers & Kara Danvers, Alex Danvers/Maggi...	In the Dark	1.0	https://archiveofourown.org/works/37627942	2022-04-27	32,614
And If My Lover Will Not Hear It	244	57	English	Teen And Up Audiences	Martin Blackwood/Jonathan "Jon" Sims | The Arch...	NaN	NaN	https://archiveofourown.org/works/38637258	2022-04-27	6,923
Narcotic	13	67	English	General Audiences	Kyd Wykkyd/Joseph Wilson, Adeline Kane Wilson &...	NaN	NaN	https://archiveofourown.org/works/38633817	2022-04-27	8,521
In Perfect Technicolor	683	42	English	Mature	Artie Abrams/Sam Evans, Santana Lopez/Brittany ...	NaN	NaN	https://archiveofourown.org/works/29826528	2022-04-27	29,607
All the Quiet Nights You Bear	1261	62	English	General Audiences	Thirteenth Doctor/Rose Tyler, Thirteenth Doctor...	And We're Not Out of the Tunnel	2.0	https://archiveofourown.org/works/37525801	2022-04-27	31,192
A Sick-Ass Kong Vault Straight Through the Godd...	974	100	English	Teen And Up Audiences	Dave Strider/Karkat Vantas, John Egbert & Dave ...	NaN	NaN	https://archiveofourown.org/works/38386291	2022-04-27	56,492
Cigarettes, Daydreams	508	22	English	Explicit	Optimus Prime/Original Male Character(s), Willi...	NaN	NaN	https://archiveofourown.org/works/36961390	2022-04-26	30,698
Looking for a Kiss	586	95	English	Explicit	Blackbeard | Edward Teach/Stede Bonnet	NaN	NaN	https://archiveofourown.org/works/38635662	2022-04-27	10,172
Child of Lecter	17431	1008	English	Explicit	Will Graham/Hannibal Lecter, Petunia Evans Durs...	NaN	NaN	https://archiveofourown.org/works/37452481	2022-04-26	41,076
A conspiracy of barely functional people (or: h...	385	65	English	Teen And Up Audiences	Sokka/Zuko (Avatar)	NaN	NaN	https://archiveofourown.org/works/38634891	2022-04-26	2,433
Children of Mine	110	14	English	Mature	Ymir Fritz & Eren Yeager, Mikasa Ackerman/Eren ...	NaN	NaN	https://archiveofourown.org/works/38633880	2022-04-27	8,200
All is fair in Love and Chaos	1216	18	English	General Audiences	Matthias Helvar/Nina Zenik, Inej Ghafa & Nina Z...	NaN	NaN	https://archiveofourown.org/works/38338435	2022-04-27	79,543
Friendship bracelets	1279	97	English	General Audiences	Stephanie Brown & Cassandra Cain, Cassandra Cai...	NaN	NaN	https://archiveofourown.org/works/37466716	2022-04-26	17,092
Orange is a Burning Color	10	NaN	English	Not Rated	Ruby Daly/Liam Stewart, Ruby Daly & Suzume "Zu"...	Orange	1.0	https://archiveofourown.org/works/38615076	2022-04-26	15,614
Learning New Strengths	823	40	English	Teen And Up Audiences	Levi Ackerman/Armin Arlert	NaN	NaN	https://archiveofourown.org/works/36885187	2022-04-26	16,559
Coming Out Of His Cage (And He's Not Doing Just...	2202	68	English	Teen And Up Audiences	NaN	***ALL VERSIONS*** Coming Out Of His Cage (And ...	1.0	https://archiveofourown.org/works/34327549	2022-04-26	49,771
The Boys Who Wouldn't Grow Up	283	23	English	Teen And Up Audiences	David/Michael Emerson (Lost Boys), David/Maria ...	NaN	NaN	https://archiveofourown.org/works/38239663	2022-04-26	16,028
A Different Sort of Impossible	21391	602	English	Mature	Hermione Granger/Draco Malfoy	NaN	NaN	https://archiveofourown.org/works/36058690	2022-04-26	128,331
Unstoppable Forces and Immovable Objects	2372	193	English	Explicit	Essek Thelyss/Caleb Widogast	NaN	NaN	https://archiveofourown.org/works/38405557	2022-04-26	13,534
All We've Lost	572	29	English	Mature	Kanda Yuu/Allen Walker, Kanda Yuu & Lavi & Lena...	NaN	NaN	https://archiveofourown.org/works/38198035	2022-04-26	48,251
dreaming about forever (only gets you so far)	1075	36	English	Teen And Up Audiences	Michael Guerin/Alex Manes, mentioned Max Evans/...	NaN	NaN	https://archiveofourown.org/works/38045185	2022-04-26	12,862
Diso-ord	110	20	English	General Audiences	NaN	NaN	NaN	https://archiveofourown.org/works/38627394	2022-04-26	5,255
To Become a Dragon's Hearth	20	2	English	Explicit	Cairn/Red Dragon | Angehus	NaN	NaN	https://archiveofourown.org/works/38570904	2022-04-26	15,215
Alone	40	4	English	Teen And Up Audiences	Kanan Jarrus/Hera Syndulla	Angstpril 2022	21.0	https://archiveofourown.org/works/38629770	2022-04-26	974
Please just let me go (don't let me go)	2062	119	English	Explicit	Ranboo & Toby Smith | Tubbo, Ranboo & Toby Smit...	NaN	NaN	https://archiveofourown.org/works/38099383	2022-04-26	4,068
delirium.	764	61	English	Teen And Up Audiences	Regulus Black/James Potter, Sirius Black/Remus ...	NaN	NaN	https://archiveofourown.org/works/38418727	2022-04-26	11,294
What Was Needed	43189	1281	English	Mature	NaN	NaN	NaN	https://archiveofourown.org/works/12457293	2022-04-26	429,015
Through the Pairings one shots	420	7	English	Not Rated	NaN	NaN	NaN	https://archiveofourown.org/works/38401555	2022-04-26	14,615
Old Bones Aflame	2463	123	English	Teen And Up Audiences	NaN	NaN	NaN	https://archiveofourown.org/works/35295772	2022-04-26	33,953
Jersey on my mind	5826	123	English	Mature	Daryl Dixon & Original Female Character(s), Dar...	NaN	NaN	https://archiveofourown.org/works/25935532	2022-04-26	140,861
Like a Crateful of Grenades	6408	680	English	Explicit	Frank Castle/Matt Murdock	NaN	NaN	https://archiveofourown.org/works/37590463	2022-04-26	18,437
Your silent words	191	22	English	Explicit	Will Graham/Hannibal Lecter	NaN	NaN	https://archiveofourown.org/works/38627187	2022-04-26	1,769
To Be Loved	240	19	English	Mature	Achilles/Patroclus (Hades Video Game), Achilles...	NaN	NaN	https://archiveofourown.org/works/38510635	2022-04-26	12,702
And If You Were Drowned At Sea	615	50	English	Teen And Up Audiences	Kaz Brekker & Jordie Rietveld, Kaz Brekker & Je...	NaN	NaN	https://archiveofourown.org/works/38524720	2022-04-26	4,263
Through the Sparks, the Flame	3156	287	English	Mature	Upgraded Connor | RK900 & Gavin Reed, Upgraded ...	NaN	NaN	https://archiveofourown.org/works/35981911	2022-04-26	26,197
The Dragon Speaker	2300	98	English	General Audiences	Firmus Piett & Maximilian Veers, Firmus Piett a...	NaN	NaN	https://archiveofourown.org/works/37469116	2022-04-26	95,647
Broken Bones [Spring - Season One]	1630	69	English	Teen And Up Audiences	Harvey/Female Player (Stardew Valley), Harvey/P...	Stardew Valley: Bluebell Farm [Harvey and Noelle]	1.0	https://archiveofourown.org/works/36278845	2022-04-26	37,253
Five minus one	841	17	English	Mature	Mikasa Ackerman/Eren Yeager, Mikasa Ackerman & ...	NaN	NaN	https://archiveofourown.org/works/37229551	2022-04-26	14,708
Within his sight	3182	80	English	Explicit	Original female character/Michael Myers	NaN	NaN	https://archiveofourown.org/works/34954858	2022-04-26	310,006
I Had To Save You	618	48	English	Mature	Anne Boleyn/Catherine of Aragon, Anne of Cleves...	How To Get Away With Murder	1.0	https://archiveofourown.org/works/38149051	2022-04-26	13,273
Into the Wild - Thistledown AU	1562	35	English	Teen And Up Audiences	NaN	Thistledown AU - Prophecies Begin arc	1.0	https://archiveofourown.org/works/36883726	2022-04-26	69,281
Bad Faith	15487	810	English	Teen And Up Audiences	Draco Malfoy & Narcissa Black Malfoy, Draco Mal...	NaN	NaN	https://archiveofourown.org/works/35730984	2022-04-26	38,178
blinded by (fraternal) love	865	114	English	Teen And Up Audiences	Clay | Dream & Toby Smith | Tubbo, Clay | Dream...	dream, brother to f-cking everyone	3.0	https://archiveofourown.org/works/36089026	2022-04-26	10,576
The Bastard Saint Of The Barrel	15528	594	English	Teen And Up Audiences	Kaz Brekker/Inej Ghafa, Jesper Fahey/Wylan Van ...	NaN	NaN	https://archiveofourown.org/works/34345318	2022-04-26	111,100
The Emperor That Burned and the God That Remain...	22102	928	English	Mature	Edelgard von Hresvelg/My Unit | Byleth, Edelgar...	NaN	NaN	https://archiveofourown.org/works/27654734	2022-04-26	74,331
For Your Ears Only	81	13	English	General Audiences	John Deacon/Freddie Mercury	NaN	NaN	https://archiveofourown.org/works/38623743	2022-04-26	6,687
Fighting For Control	1056	74	English	Explicit	Sunnydrop/reader, Moonydrop/Reader, Moonymrise/R...	Taming the Virus	2.0	https://archiveofourown.org/works/38147767	2022-04-26	11,388
let down your hair and i'll let down my gun	6	NaN	English	Mature	Ajay Ghale/Noore Najjar, Bhadra & Ajay Ghale, Y...	NaN	NaN	https://archiveofourown.org/works/38623389	2022-04-26	2,551
Aurachnid - Whumptober 2021	116	NaN	English	Mature	Matt Murdock & Original Character(s), Original ...	The Adventures of Aurachnid: Atomic Disaster	4.0	https://archiveofourown.org/works/35537014	2022-04-26	34,080
Kidza's Adventures in the Amazing World of Deat...	2089	50	English	Teen And Up Audiences	Clay | Dream/Technoblade (Video Blogging RPF), ...	NaN	NaN	https://archiveofourown.org/works/37420981	2022-04-26	7,770
crystalline reflections over rose flavoured ice	621	100	English	General Audiences	galacta knight & kirby, Kirby & Meta Knight	NaN	NaN	https://archiveofourown.org/works/38339932	2022-04-26	4,449
Breaths We Dare to Take	1558	106	English	Teen And Up Audiences	Jayce/Viktor (League of Legends), Jayce (League...	The Monster You Created	1.0	https://archiveofourown.org/works/37361518	2022-04-26	17,428
30 Days, 30 Stories: Star Trek Drabbles	479	27	English	Mature	James T. Kirk/Spock, Amanda Grayson/Sarek, T'Po...	NaN	NaN	https://archiveofourown.org/works/38052646	2022-04-26	2,998
Kaleidoscopic Grangers	225551	3071	English	Teen And Up Audiences	Ariadne Granger/Ginny Weasley, Hermione Granger...	NaN	NaN	https://archiveofourown.org/works/24597805	2022-04-26	1,389,149
Blindsided	16989	535	English	Explicit	Obi-Wan Kenobi/Darth Maul	Undersong	3.0	https://archiveofourown.org/works/34460494	2022-04-26	97,222
Shrimp	120	1	English	Teen And Up Audiences	NaN	NaN	NaN	https://archiveofourown.org/works/38328628	2022-04-26	61,674
Unrequited Love	90	10	English	Mature	Marcin "Jankos" Jankowski/Martin "Rekkles" Larson	NaN	NaN	https://archiveofourown.org/works/38621481	2022-04-26	5,722
Broke	57	5	English	Teen And Up Audiences	Nagato | Pain/Uchiha Obito	NaN	NaN	https://archiveofourown.org/works/38550945	2022-04-26	3,488
Avalanche	2952	111	English	Mature	Altmer/Nord(s) (Elder Scrolls), Thalmor (Elder ...	Avalanche	2.0	https://archiveofourown.org/works/34606852	2022-04-26	121,814
Transformers watching bumblebee	96	5	English	General Audiences	Bumblebee & Optimus Prime, Bumblebee & Charlie ...	NaN	NaN	https://archiveofourown.org/works/38620140	2022-04-26	4,772
Salvation, Restoration, Liberation	10691	489	English	Explicit	Hermione Granger/Draco Malfoy, Remus Lupin/Nymp...	NaN	NaN	https://archiveofourown.org/works/33987706	2022-04-26	44,818
Fallacy of Yesterday	417	51	English	Not Rated	Oliver Brotherhood & Charles | Grian & impulseS...	NaN	NaN	https://archiveofourown.org/works/38160976	2022-04-26	3,157
Seeing You	4867	319	English	Teen And Up Audiences	Akaashi Keiji/Miya Osamu, miya osamu/akaashi ke...	NaN	NaN	https://archiveofourown.org/works/29190336	2022-04-26	33,966
Anything?	176	19	English	Mature	Keith/Lance (Voltron)	NaN	NaN	https://archiveofourown.org/works/38617929	2022-04-26	13,318
Now you see me	395	20	English	Teen And Up Audiences	James "Bucky" Barnes/Reader, Avengers Team & Re...	NaN	NaN	https://archiveofourown.org/works/38230012	2022-04-26	12,280
Can't Save Them All	1460	117	English	Mature	Wilbur Soot & Technoblade & TommyInnit & Phil W...	NaN	NaN	https://archiveofourown.org/works/37477072	2022-04-26	10,946
Green	2042	75	English	General Audiences	pre-Evan "Buck" Buckley/Eddie Diaz (9-1-1 TV), ...	FireFam Conflagration Ficlets AKA Firehouse Fic...	7.0	https://archiveofourown.org/works/38575839	2022-04-26	5,722
The Piano's Missing Keys	25941	992	English	Teen And Up Audiences	Wilbur Soot & Technoblade & TommyInnit & Phil W...	NaN	NaN	https://archiveofourown.org/works/35288320	2022-04-26	74,465
Love Looks Not With the Eyes, But With the Heart	2417	147	English	Teen And Up Audiences	Crosshair & Omega (Star Wars: The Bad Batch), C...	Dad Batch	9.0	https://archiveofourown.org/works/37594141	2022-04-26	25,486
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table B.2: Second part of the Disability Story-Set. Each observation describes the “Title", “Hits", “Kudos", “Rating", “Relationship", “Series", “Part", “SourceURL", and “Updated" (last date updated) of a fan-work in the Story-Set.

Appendix C

Generated Reports

C.1 Format of the Static Analysis Report

As explained in section 3.5.1, the static analysis module generates output data. An example analysis, as generated for the Extended Disability-RATAS, is provided. It is composed of five different worksheets each storing the general information of the RATAS, the network properties, the canonical nodes with relative large out-degree, all the syn-sets, and all canonical nodes with multiple successors.

Properties	Values
big_nodes	Disability, Blindness, Disabled Character, Mute...
total_nodes	4433
total_edges	5091
total_canonical_leafs	31
total_freeform	0
total_syn	3139
total_canonical	667
total_canonical_syn_connection_nodes	627

Table C.1: Worksheet (I): General Network Information, generated for the Extended Disability-RATAS.

Properties	Values
is_strongly_connected	False
is_weakly_connected	True
diameter_undirected	11
radius_undirected	6
center_RATAS_undirected	['Disability', 'Disabled Character']
has_cycles	[('Blind Sam Winchester', 'Blind Sam Winchester...
k_components	dict_keys([2, 1])
2_components	[{'Mute Clay | Dream (Video Blogging RPF)', 'Mu...

Table C.2: Worksheet (II): Network Properties, generated for the Extended Disability-RATAS. Stores if the network is strongly/weakly connected, its diameter, radius, center, cycles, and k-connected components with k for which the network as maximal sub-network that has, at least, node connectivity k-strongly.

Node	out_degree	in_degree
Disability	11	1
Blindness	21	2
Disabled Character	55	2
Muteness	175	2
Deaf	5	2
Disabled Eddie Kaspbrak	6	2
Blind Character	186	3
Deaf Character	129	3
Dyslexia	29	2
Dyslexic Reggie Peters (Julie and The Phantoms)...	8	1

Table C.3: Worksheet (III): List of all Large Canonical Nodes, generated for Extended Disability-RATAS.

Node	Count	SynTags
Disability	133	#disabled,(spoiler) disability,Acquired Disabli...
Blindness	103	Accidental Blindness,and blind,and blindness,AU...
Disabled Character	78	acknowledge of character's disability,also they...
Muteness	121	also he mute,also: mute main player sort of cha...
Deaf	43	about deafness,Actual Deaf issues in Deaf cultu...
Physical Disability	31	another character has a physical disability but...
Learning Disabilities	17	hinted learning disability,Implied Learning Dis...
Disabled Richie Tozier	1	Wheelchair User Richie Tozier
Disabled Eddie Kaspbrak	3	bc we need more disabled eddie content,Eddie us...
Disabled Anduin Wrynn	2	this author remembers anduin is canonically dis...
Disabled Ashton Greymoore	1	Disabled Ashton (Critical Role)
Blind Character	53	(the major character injury is him being blind)...
Blind Tony Stark	8	also tony is blind,blind Tony,blind!Tony,Blind!...
Blind Steve Rogers	3	Blind Steve,blind!steve,Steve is blind
Blind Will Graham	3	blind will,blind!Will,Will is blind
Blind Uchiha Sasuke	3	Blind Sasuke,blind! sasuke,Blind!Sasuke
Blind Bucky Barnes	5	Blind bucky,Blind! Bucky,blind!bucky,blind!Buck...
Blind Sirius Black	3	Blind Sirius,Blind! Sirius,blind!Sirius
Blind Astrid Hofferson	3	Blind Astrid,Blind Astrid AU,Blind!Astrid
Blind Merlin (Merlin)	5	blind Merlin,blind merlin headcanon,Blind!Merli...
Blind Niall Horan	3	Blind Niall,Blind!Niall,Niall is blind
Blind Liam Payne	1	blind!Liam
Blind Senju Tobirama	8	Blind Senju Tobirama (Naruto),Blind Tobirama,bl...
Blind Hatake Kakashi	5	Blind Kakashi,Blind!Kakashi,Kakashi es más cieg...
Blind Crowley (Supernatural)	1	Blinded!Crowley
Blind Gojo Satoru	4	Blind Gojo,Gojo is blind,gojo is going blind,Go...
Blind Roxanne Wolf (Five Nights at Freddy's)	2	Blind Roxanne,blind roxy
Canon Disabled Character	16	(Temporary) Canon Disabled Character,Additional...
Deaf Character	53	(D)deaf/HoH Character(s),Alternate Universe - D...
Disabled Character of Color	1	Disabled Chromatic Character
Disabled Dean Winchester	15	Character Affected: Dean,Character Affected: De...
Disabled Sam Winchester	5	Character Affected: Sam,Character Affected: Sam...
Disabled Harry	1	Harry is disabled
Disabled Marco Bott	2	disabled!marco,Disabled!Marco Bott
Disabled Barry Allen	3	Barry is a disabled former soldier who uses a c...
Disabled Remus Lupin	5	Disabled Remus,Disabled!Remus Lupin,Remus has a...
Disabled Eren Yeager	3	Disabled Eren,disabled! eren,disabled!eren
Disabled Hinata Shouyou	3	Disabled Hinata,Disabled! Hinata,disabled!Hinata
Disabled Derek Hale	3	Derek hale is disabled,Derek is disabled,Disabl...
Disabled Steve Rogers	1	Steve is a disabled person
Disabled Lance (Voltron)	2	Disabled Lance,Disabled Lance (Lance)
Disabled Hermione Granger	1	Disabled Hermione
Disabled Midoriya Izuku	8	Disabled Deku,disabled izuku,disabled!midoriya ...
Disabled Stiles Stilinski	4	disability!stiles,Disabled Stiles,disabled!Stil...
⋮	⋮	⋮

Table C.4: Worksheet (IV): List of All Syn-Sets, generated for Extended *"Disability"*-RATAS. syn-sets. All canonical tags with syn-sets different from the trivial syn-set, the number of syn tags, and the syn-set.

Node	in_degree	Metatags
Blindness	2	Disability,Blindness_synned_tags
Disabled Character	2	Disability,Disabled Character_synned_tags
Muteness	2	Disability,Muteness_synned_tags
Deaf	2	Disability,Deaf_synned_tags
Physical Disability	2	Disability,Physical Disability_synned_tags
Learning Disabilities	2	Disability,Learning Disabilities_synned_tags
Disabled Richie Tozier	2	Disability,Disabled Richie Tozier_synned_tags
Disabled Eddie Kaspbrak	2	Disability,Disabled Eddie Kaspbrak_synned_tags
Disabled Anduin Wrynn	2	Disability,Disabled Anduin Wrynn_synned_tags
Disabled Ashton Greymoore	2	Disability,Disabled Ashton Greymoore_synned_tags
Blind Character	3	Blindness,Disabled Character,Blind Character_sy...
Blind Tony Stark	2	Blindness,Blind Tony Stark_synned_tags
Blind Steve Rogers	2	Blindness,Blind Steve Rogers_synned_tags
Blind Will Graham	2	Blindness,Blind Will Graham_synned_tags
Blind Uchiha Sasuke	2	Blindness,Blind Uchiha Sasuke_synned_tags
Blind Bucky Barnes	2	Blindness,Blind Bucky Barnes_synned_tags
Blind Sirius Black	2	Blindness,Blind Sirius Black_synned_tags
Blind James T. Kirk	3	Blindness,Disabled James T. Kirk,Disabled James...
Blind Astrid Hofferson	2	Blindness,Blind Astrid Hofferson_synned_tags
Blind Merlin (Merlin)	2	Blindness,Blind Merlin (Merlin)_synned_tags
Blind Niall Horan	2	Blindness,Blind Niall Horan_synned_tags
Blind Liam Payne	2	Blindness,Blind Liam Payne_synned_tags
Blind Senju Tobirama	2	Blindness,Blind Senju Tobirama_synned_tags
Blind Hatake Kakashi	2	Blindness,Blind Hatake Kakashi_synned_tags
Blind Crowley (Supernatural)	2	Blindness,Blind Crowley (Supernatural)_synned_tags
Blind Gojo Satoru	2	Blindness,Blind Gojo Satoru_synned_tags
Blind Roxanne Wolf (Five Nights at Freddy's)	2	Blindness,Blind Roxanne Wolf (Five Nights at Fr...
Canon Disabled Character	2	Disabled Character,Canon Disabled Character_syn...
Deaf Character	3	Disabled Character,Deaf,Deaf Character_synned_tags
Disabled Character of Color	3	Disabled Character,Character(s) of Color,Disabl...
Disabled Dean Winchester	2	Disabled Character,Disabled Dean Winchester_syn...
Disabled Sam Winchester	2	Disabled Character,Disabled Sam Winchester_synn...
Disabled Harry	2	Disabled Character,Disabled Harry_synned_tags
Disabled Marco Bott	2	Disabled Character,Disabled Marco Bott_synned_tags
Disabled Barry Allen	2	Disabled Character,Disabled Barry Allen_synned_...
Disabled Remus Lupin	2	Disabled Character,Disabled Remus Lupin_synned_...
Disabled Eren Yeager	2	Disabled Character,Disabled Eren Yeager_synned_...
Disabled Hinata Shouyou	2	Disabled Character,Disabled Hinata Shouyou_synn...
Disabled Derek Hale	2	Disabled Character,Disabled Derek Hale_synned_tags
Disabled Steve Rogers	2	Disabled Character,Disabled Steve Rogers_synned...
Disabled Lance (Voltron)	2	Disabled Character,Disabled Lance (Voltron)_syn...
Disabled Hermione Granger	2	Disabled Character,Disabled Hermione Granger_sy...
Disabled Midoriya Izuku	2	Disabled Character,Disabled Midoriya Izuku_synn...
Disabled Stiles Stilinski	2	Disabled Character,Disabled Stiles Stilinski_sy...
Disabled Castiel (Supernatural)	2	Disabled Character,Disabled Castiel (Supernatur...
⋮	⋮	⋮

Table C.5: Worksheet (V): List of all canonical nodes with multiple successors, generated for Extended *"Disability"*-RATAS.

C.1.1 Diffing Report

Tags	rata_disability_2013_precence	rata_disability_current_expanded_full_presence	rata_disability_2013_type	rata_disability_current_expanded_full_type	Action
mute ashton	False	True	NaN	synonymized_tag	addition
Blind!Tim au	False	True	NaN	synonymized_tag	addition
Blind Sean McLoughlin	False	True	NaN	canonical_tag	addition
Deaf Tony Stark_synonymized_tags	False	True	NaN	connection_node	addition
Oh also Hobi is blind	False	True	NaN	synonymized_tag	addition
Mute Severus Snape	False	True	NaN	canonical_tag	addition
disability visibility	False	True	NaN	freeform_tag	addition
Deaf Tony Stark	False	True	NaN	canonical_tag	addition
cieco	False	True	NaN	synonymized_tag	addition
Blind!Adrien	False	True	NaN	synonymized_tag	addition
Disabled Kim Taehyung | V_synonymized_tags	False	True	NaN	connection_node	addition
Disabled TommyInnit (Video Blogging RPF)_synonymized...	False	True	NaN	connection_node	addition
Dyslexic Dewey Duck (Disney)_synonymized_tags	False	True	NaN	connection_node	addition
Blindizuku	False	True	NaN	synonymized_tag	addition
Blind Bill Cipher_synonymized_tags	False	True	NaN	connection_node	addition
disabled Essek	False	True	NaN	synonymized_tag	addition
disability and recovery	False	True	NaN	freeform_tag	addition
He's blinded	False	True	NaN	synonymized_tag	addition
Disabled Jeon Jungkook	False	True	NaN	canonical_tag	addition
deaf!zayn	False	True	NaN	synonymized_tag	addition
Mute Draco Malfoy_synonymized_tags	False	True	NaN	connection_node	addition
Mention of Disability	False	True	NaN	freeform_tag	addition
This will treat Remus kind of like he's disabled	False	True	NaN	synonymized_tag	addition
deaf iwaizumi	False	True	NaN	synonymized_tag	addition
Scott is Blind	False	True	NaN	synonymized_tag	addition
Mute Harry Styles_synonymized_tags	False	True	NaN	connection_node	addition
jjk is blind	False	True	NaN	synonymized_tag	addition
Deaf!Jon	False	True	NaN	synonymized_tag	addition
Disability Issues	True	True	synonymized_tag	synonymized_tag	NaN
The one where Derek is deaf	False	True	NaN	synonymized_tag	addition
Dyslexic Scout	False	True	NaN	synonymized_tag	addition
Disabled Jonathan "Jon" Sims | The Archivist_sy...	False	True	NaN	connection_node	addition
Blind!Lumine AU	False	True	NaN	synonymized_tag	addition
Disabled WWX	False	True	NaN	synonymized_tag	addition
Visually Impaired Dick Grayson	False	True	NaN	synonymized_tag	addition
Blind Floris | Fundy	False	True	NaN	canonical_tag	addition
disabled harry aswell	False	True	NaN	synonymized_tag	addition
Deaf Victoria (Cats)_synonymized_tags	False	True	NaN	connection_node	addition
deaf!carmilla deaf carmilla	False	True	NaN	synonymized_tag	addition
Deafened Castiel	False	True	NaN	synonymized_tag	addition
due to disability disadvantages	False	True	NaN	freeform_tag	addition
tae is blind	False	True	NaN	synonymized_tag	addition
Major warning for a character getting severly h...	False	True	NaN	synonymized_tag	addition
Blind!Connor	False	True	NaN	synonymized_tag	addition
mute dick	False	True	NaN	synonymized_tag	addition
Blind Soldier: 76 | Jack Morrison_synonymized_tags	False	True	NaN	connection_node	addition
also: mute main player sort of character	False	True	NaN	synonymized_tag	addition
references to past disability	False	True	NaN	freeform_tag	addition
Selectively Mute Characters	False	True	NaN	synonymized_tag	addition
Mute Shinso	False	True	NaN	synonymized_tag	addition
Deaf Hermione Granger	False	True	NaN	canonical_tag	addition
liam is mute	False	True	NaN	synonymized_tag	addition
⋮	⋮	⋮	⋮	⋮	⋮

Table C.6: Tag usage difference between the Disability-RATAS snapshots from 2013 and 2022.

C.1.2 Diffing Report Multiple Snapshots

Unnamed: 0	rata_disability_2021_type	rata_disability_current_presence	rata_disability_current_type	FROM_rata_disability_2013_TO___Action	FROM_rata_disability_2014_TO___Action	FROM_rata_disability_2015_TO___Action	FROM_rata_disability_2016_TO___Action	FROM_rata_disability_2017_TO___Action	FROM_rata_disability_2018_TO___Action	FROM_rata_disability_2021_TO_rata_disability_current_Action
0	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
1	NaN	False	NaN	NaN	NaN	canonized	NaN	NaN	removed	NaN
2	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
3	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
4	NaN	False	NaN	NaN	NaN	NaN	canonized	NaN	removed	NaN
5	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	canonized	NaN	NaN
6	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	canonized	NaN	NaN
7	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
8	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
9	canonical_tag	True	canonical_tag	NaN	NaN	canonized	NaN	NaN	NaN	NaN
10	NaN	False	NaN	NaN	NaN	NaN	NaN	removed	NaN	NaN
11	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
12	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	canonical_tag	True	canonical_tag	canonized	NaN	NaN	NaN	NaN	NaN	NaN
14	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
15	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
16	NaN	False	NaN	NaN	NaN	canonized	removed	NaN	NaN	NaN
17	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
18	NaN	False	NaN	NaN	NaN	NaN	canonized	NaN	removed	NaN
19	canonical_tag	True	canonical_tag	NaN	canonized	NaN	NaN	NaN	NaN	NaN
20	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
21	syn_tag	True	syn_tag	NaN	synonymized	NaN	NaN	NaN	NaN	NaN
22	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
23	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
24	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
25	NaN	False	NaN	NaN	NaN	NaN	canonized	NaN	removed	NaN
26	syn_tag	True	syn_tag	synonymized	NaN	NaN	NaN	NaN	NaN	NaN
27	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
28	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
29	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
30	NaN	False	NaN	canonized	NaN	NaN	removed	NaN	NaN	NaN
31	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
32	syn_tag	True	syn_tag	NaN	NaN	synonymized	NaN	NaN	NaN	NaN
33	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
34	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
35	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
36	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
37	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	canonized	NaN	NaN
38	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
39	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
40	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
41	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
42	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
43	syn_tag	True	syn_tag	NaN	NaN	NaN	synonymized	NaN	NaN	NaN
44	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
45	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
46	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
47	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
48	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
49	NaN	False	NaN	NaN	NaN	NaN	NaN	canonized	removed	NaN
50	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
51	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
52	NaN	False	NaN	canonized	NaN	NaN	NaN	NaN	removed	NaN
53	NaN	False	NaN	canonized	NaN	NaN	NaN	NaN	removed	NaN
54	canonical_tag	True	canonical_tag	NaN	NaN	canonized	NaN	NaN	NaN	NaN
55	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
56	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
57	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
58	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
59	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
60	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
61	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
62	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
63	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
64	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
65	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
66	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
67	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
68	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
69	syn_tag	True	syn_tag	NaN	synonymized	NaN	NaN	NaN	NaN	NaN
70	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	NaN
71	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
72	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
73	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
74	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
75	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	NaN
76	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
77	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
78	syn_tag	True	syn_tag	NaN	NaN	NaN	NaN	NaN	synonymized	NaN
79	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	NaN
80	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
81	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
82	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
83	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
84	NaN	False	NaN	NaN	NaN	NaN	canonized	NaN	removed	NaN
85	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
86	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized
87	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
88	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
89	NaN	False	NaN	removed	NaN	NaN	NaN	NaN	NaN	NaN
90	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
91	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
92	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
93	NaN	True	syn_tag	NaN	NaN	NaN	NaN	NaN	NaN	synonymized
94	NaN	True	syn_tag	NaN	NaN	NaN	NaN	NaN	NaN	synonymized
95	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	canonized	NaN
96	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
97	canonical_tag	True	canonical_tag	NaN	NaN	NaN	NaN	canonized	NaN	NaN
98	canonical_tag	True	canonical_tag	NaN	NaN	NaN	canonized	NaN	NaN	NaN
99	NaN	True	canonical_tag	NaN	NaN	NaN	NaN	NaN	NaN	canonized

Table C.7: Second part of the difference in sequence of all Disability-RATAS scraped using the Wayback Machine (2013, 2014, 2015, 2016, 2017, 2018, 2021, 2022).

Unnamed: 0	Tags	rata_disability_2013_presence	rata_disability_2013_type	rata_disability_2014_presence	rata_disability_2014_type	rata_disability_2015_presence	rata_disability_2015_type	rata_disability_2016_presence	rata_disability_2016_type	rata_disability_2017_presence	rata_disability_2017_type	rata_disability_2018_presence	rata_disability_2018_type	rata_disability_2021_presence
0	Blind Will Graham	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
1	Mute Junior	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	False
2	Mute Hannibal Lecter	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
3	Disabled Ryan | GoodTimesWithScar	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
4	Selectively Mute Castiel	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	False
5	Deaf Alec Lightwood	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True
6	Deaf Phil Lester	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True
7	Blind Han Jisung | Han	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
8	Deaf Todoroki Shouto	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
9	Disabled Eren Yeager	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
10	Blind Dean	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	False	NaN	False
11	Blind Takami Keigo | Hawks	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
12	Permanent Injury	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
13	Mute Louis	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
14	Blind Eren Yeager	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
15	Blind Liam Payne	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
16	Deaf Peter	False	NaN	False	NaN	False	NaN	True	canonical_tag	False	NaN	False	NaN	False
17	Mute Harry Hart	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
18	Blind Karhat	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	False
19	Blind Scott	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
20	Mute Levi Ackerman	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
21	prompt - disabilities	False	NaN	False	NaN	True	syn_tag	True	syn_tag	True	syn_tag	True	syn_tag	True
22	Disabled Crowley (Good Omens)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
23	description of disability	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
24	Selectively Mute Mirabel Madrigal	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
25	Mute Luke	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	False
26	disabilities	False	NaN	True	syn_tag	True	syn_tag	True	syn_tag	True	syn_tag	True	syn_tag	True
27	Mute Six (Little Nightmares)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
28	Deaf Logic | Logan Sanders	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
29	Disabled Anduin Wrynn	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
30	Mute Dave	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	False	NaN	False	NaN	False
31	Mute Keith (Voltron)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
32	Kalectwo	False	NaN	False	NaN	False	NaN	True	syn_tag	True	syn_tag	True	syn_tag	True
33	Disabled Jonathan "Jon" Sims | The Archivist	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
34	Mute Kim Seokjin | Jin	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
35	Thinking about disability	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
36	Blind Iwairumi Hajime	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
37	Deaf Alexander Hamilton	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True
38	Blind Min Yoongi | Suga	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
39	Mute Niall Horan	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
40	Blind Todoroki Shouto	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
41	Selectively Mute Jung Hoseok | J-Hope	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
42	Disabled Inquisitor (Dragon Age)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
43	discussions of disability	False	NaN	False	NaN	False	NaN	False	NaN	True	syn_tag	True	syn_tag	True
44	Blind Sherlock Holmes	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
45	Mute Wild (Linked Universe)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
46	Blind Grantaire (Les Misérables)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
47	Deaf Oswald Cobblepot	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
48	Disabled Geralt z Rivii | Geralt of Rivia	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
49	Blind Adrien Agreste	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	False
50	Mute Pramipto Argentum	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
51	Blind James T"d" Kirk	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
52	Blind Louis	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	False
53	Mute Niall	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	False
54	Mute Nico di Angelo	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
55	Deaf Isak Valtersen	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
56	Deaf Remus Lupin	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
57	Blind Dan Howell	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
58	Mute Joseph Wilson	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
59	Selectively Mute Persona 5 Protagonist	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
60	Deaf Jung Hoseok | J-Hope	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
61	Selectively Mute Victor Grantz | Postman	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
62	mentions to disability	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
63	Blind Marinette Dupain-Cheng	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
64	Deaf Miyuki Kazuya	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
65	With added disabilities	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
66	Living With a Disability	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
67	Selectively Mute Tommyinnit (Video Blogging RPF)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
68	Blind Geralt z Rivii | Geralt of Rivia	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
69	includes disabilities	False	NaN	False	NaN	True	syn_tag	True	syn_tag	True	syn_tag	True	syn_tag	True
70	Learning Disabilities	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
71	Deaf Lee Felix (Stray Kids)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
72	Dyslexic Evan "Buck" Buckley	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
73	Deaf Adora (She-Ra)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
74	Mute Creativity | Roman "Princey" Sanders	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
75	Disability	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
76	Mute Aziraphale (Good Omens)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
77	Deaf Carmilla	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
78	niepełnosprawność	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
79	Blindness	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True	canonical_tag	True
80	Disabled Lance (Voltron)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
81	Blind Derek Hale	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
82	Blind Anxiety | Virgil Sanders	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
83	Disabled Marcy Wu	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
84	Blind V (Mystic Messenger)	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	False
85	Mute Stiles Stilinski	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
86	Selectively Mute Technoblade (Video Blogging RPF)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
87	Mute Ashton Irwin	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
88	Selectively Mute Hunter (Bloodborne)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
89	Deaf Clint	True	canonical_tag	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
90	Selectively Mute Jeon Jungkook	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
91	Dyslexic Jaime Lannister	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
92	Deaf Eddie Kaspbrak	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
93	All three girls are disabled	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
94	mild disability	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False
95	Deaf Louis Tomlinson	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True
96	Blind Dave Strider	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
97	Blind Ian Gallagher	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True
98	Blind Tachibana Makoto	False	NaN	False	NaN	False	NaN	False	NaN	True	canonical_tag	True	canonical_tag	True
99	Disabled Catra (She-Ra)	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False	NaN	False

Table C.8: First part of the difference in sequence of all Disability-RATAS scraped using the Wayback Machine (2013, 2014, 2015, 2016, 2017, 2018, 2021, 2022).

Appendix D

Glossary

1. Archive: a collection of historical documents or records providing information about a place, institution, or group of people.
2. Canonical tags: we refer to canonical tags as already defined tags in the AO3 community.
3. Commons (information/knowledge): an information system, such as a physical library or online community, that exists to produce, conserve, and preserve information for current and future generations.
4. Fan-works: it refers to the stories (works) created by contributors (authors) in AO3.
5. Folksonomy: a user-generated system of classifying and organizing online content into different categories by the use of metadata such as electronic tags.
6. Free tags: we refer to free tags as any word choice the author of a fan work selects to tag its fan-work that has not been canonized or synonymized yet. Tags on the Archive of Our Own (AO3) may be up to 100 characters long and can include characters from most languages, numbers, spaces, and some punctuation but not a comma. The comma is used to split tags.
7. Meta-tags: refers to the predecessor-successor relationship of canonical tags while a meta-tag is the predecessor of a canonical tag. Meta-tags are treated as canonical tags.
8. Taxonomy: is the practice and science of categorization or classification. A taxonomy is a scheme of classification, especially a hierarchical classification, in which things are organized into groups or types.

Bibliography

Archive of Our Own (n.d.[a]). *AO3 Reaches 4 Million Fanworks! | Archive of Our Own*. URL: https://archiveofourown.org/admin_posts/10833 (visited on 06/13/2022).

— (n.d.[b]). *Deaf | Archive of Our Own*. URL: https://archiveofourown.org/tags/Deaf (visited on 06/14/2022).

— (n.d.[c]). *Home | Archive of Our Own*. URL: https://archiveofourown.org/ (visited on 11/04/2021).

— (n.d.[d]). *Search Tags | Archive of Our Own*. URL: https://archiveofourown.org/tags/search (visited on 06/30/2022).

— (n.d.[e]). *Selective data dump for fan statisticians | Archive of Our Own*. URL: https://archiveofourown.org/admin_posts/18804 (visited on 04/11/2022).

— (n.d.[f]). *Show Wrangling Guideline! | Archive of Our Own*. URL: https://archiveofourown.org/wrangling_guidelines/11 (visited on 06/01/2022).

— (n.d.[g]). *Tags FAQ | Archive of Our Own*. URL: https://archiveofourown.org/faq/tags?language_+id=en&language_id=en (visited on 06/01/2022).

— (n.d.[h]). *Terms of Service (TOS) FAQ Home | Archive of Our Own*. URL: https://archiveofourown.org/tos_faq (visited on 05/10/2022).

— (n.d.[i]). *Tutorial: Posting a Work on AO3 FAQ | Archive of Our Own*. URL: https://archiveofourown.org/faq/tutorial-posting-a-work-on-ao3?language_id=en#pwtcategory (visited on 06/13/2022).

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy (2009). *Gephi: An Open Source Software for Exploring and Manipulating Networks*. URL: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

Bischoff, Kerstin et al. (2008). "Can All Tags Be Used for Search?" In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. CIKM '08. Napa Valley, California, USA: Association for Computing Machinery, pp. 193–202. ISBN:

9781595939913. DOI: 10.1145/1458082.1458112. URL: https://doi.org/10.1145/1458082.1458112.

Black, Erin (2021). *"There's A Tag for That": An Exploratory Study of Tag Functions in the Archive of Our Own.*

Capocci, Andrea and Guido Caldarelli (2008). "Folksonomies and clustering in the collaborative system CiteULike". In: *Journal of Physics A: Mathematical and Theoretical* 41.22. ISBN: 1751-8121 Publisher: IOP Publishing, p. 224016.

Chakraborty, Abhijnan, Saptarshi Ghosh, and Niloy Ganguly (2012). "Detecting overlapping communities in folksonomies". In: *Proceedings of the 23rd ACM conference on Hypertext and social media*, pp. 213–218.

Chakraborty, Tanmoy et al. (Nov. 2017). "Metrics for Community Analysis: A Survey". In: *ACM Computing Surveys* 50.4, pp. 1–37. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3091106.

Crampes, Michel and Michel Plantié (Feb. 2014). "A unified community detection, visualization and analysis method". In: *Advances in Complex Systems* 17.01. Publisher: World Scientific Publishing Co., p. 1450001. ISSN: 0219-5259. DOI: 10.1142/S0219525914500015.

Dalton, Kelly L. (2012). "Searching the archive of our own: The usefulness of the tagging structure". ISBN: 1267853999. PhD thesis. The University of Wisconsin-Milwaukee.

De Kosnik, Abigail (2016). *Rogue archives: Digital cultural memory and media fandom.* English. OCLC: 1164477789. M.I.T. Press. ISBN: 978-0-262-33676-5.

Dubinko, Micah et al. (Aug. 2007). "Visualizing tags over time". In: *ACM Transactions on the Web* 1.2. ISSN: 1559-1131. DOI: 10.1145/1255438.1255439.

Duchastel de Montrouge, Cath and Melanie Baljko (May 2017). *Datascraping for disability: Analyzing disability tags on an online multi-fandoms fanfiction archive*. English. Conference Presentation. Niagara Falls, Ontario, Canada.

— (June 2019). *"I'm not sure whether to tag this as disability or not": The Politics of Tagging Disability on AO3*. English. Conference Presentation. University of British Columbia, Vancouver, B.C.

Etemadikhou, Ghazale, Fatemeh Azimzadeh, and Abdolsamad Karamatfar (2020). "Resolving Ambiguity Using Word Embeddings for Personalized Information Retrieval in Folksonomy Systems". In: *International Journal of Web Research* 3.2, pp. 26–30. ISSN: 2645-4335. DOI: 10.22133/ijwr.2021.254896.1079. eprint:

http://ijwr.usc.ac.ir/article_128964_cfd180f6eff410577eeb06c38b32ec23.pdf. URL: http://ijwr.usc.ac.ir/article_128964.html.

Flaherty, Ruth (Apr. 2022). "Fair Dealing in a Fandemic: How Pastiche can be used to Clarify the Position of User-Generated Content". en. In: *European Journal of Law and Technology* 13.1. ISSN: 2042-115X.

Gephi - The Open Graph Viz Platform (2022). URL: https://gephi.org/ (visited on 06/13/2022).

Gupta, Manish et al. (Nov. 2010). "Survey on Social Tagging Techniques". In: *SIGKDD Explor. Newsl.* 12.1, pp. 58–72. ISSN: 1931-0145. DOI: 10.1145/1882471.1882480. URL: https://doi.org/10.1145/1882471.1882480.

Heidari, Farzaneh and Manos Papagelis (Mar. 2020). "Evolving network representation learning based on random walks". In: *Applied Network Science* 5.1, p. 18. ISSN: 2364-8228. DOI: 10.1007/s41109-020-00257-3. URL: https://doi.org/10.1007/s41109-020-00257-3.

Hoque, Enamul and Giuseppe Carenini (2016). "MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations". In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. IUI '16. Sonoma, California, USA: Association for Computing Machinery, pp. 96–107. ISBN: 9781450341370.

Hoque, Enamul, Vidya Setlur, et al. (Jan. 2018). "Applying Pragmatics Principles for Interaction with Visual Analytics". en. In: *IEEE Transactions on Visualization and Computer Graphics* 24.1, pp. 309–318. ISSN: 1077-2626. DOI: 10.1109/TVCG.2017.2744684. URL: http://ieeexplore.ieee.org/document/8019833/ (visited on 06/20/2022).

Hossen, Md. Shahed et al. (Mar. 2022). "The Application of Graph Theory for The Networking of Rural Agriculture Products Marketing". In: *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, pp. 1–6. DOI: 10.1109/IC3IOT53935.2022.9767980.

Hotho, Andreas et al. (2006). "Information Retrieval in Folksonomies: Search and Ranking". In: *The Semantic Web: Research and Applications*. Ed. by York Sure and John Domingue. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 411–426. ISBN: 978-3-540-34545-9.

Jäschke, Robert et al. (2007). "Tag Recommendations in Folksonomies". In: *Knowledge Discovery in Databases: PKDD 2007*. Ed. by Joost N. Kok et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 506–514. ISBN: 978-3-540-74976-9.

Ladd, John et al. (2017). "Exploring and analyzing network data with Python". In: *Programming Historian* 6.

Lee, Sigma On Kee and Andy Hon Wai Chun (2007). "Automatic tag recommendation for the web 2.0 blogosphere using collaborative tagging and hybrid ANN semantic structures". In: *ACOS* 7, pp. 88–93.

Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman (2020). *Mining of massive data sets*. Cambridge university press.

Messina, Cara Marta (2021). "The Critical Fan Toolkit: Fanfiction Genres, Ideologies, and Pedagogies". PhD thesis. Northeastern University.

Price, Ludi (2019). "Fandom, Folksonomies and Creativity: the case of the Archive of Our Own". In: *The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization*. Ergon-Verlag, pp. 11–37. ISBN: 978-3-95650-549-2.

Price, Ludi and Lyn Robinson (Jan. 2021). "Tag analysis as a tool for investigating information behaviour: comparing fan-tagging on Tumblr, Archive of Our Own and Etsy". In: *Journal of Documentation* 77.2, pp. 320–358. ISSN: 0022-0418. DOI: 10.1108/JD-05-2020-0089.

Qassimi, Sara, Meriem Hafidi, Aimad Qazdar, et al. (2021). "Towards a folksonomy graph-based context-aware recommender system of annotated books". In: *Journal of Big Data* 8.1, pp. 1–17.

Shen, Kaikai and Lide Wu (2005). "Folksonomy as a complex network". In: *arXiv preprint cs/0509072*.

Smith, Mark et al. (2010). "NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007". In: *Social Media Research Foundation*. URL: https://www.smrfoundation.org.

— (June 2016). *NodeXL: network analysis & insights as easy as pie charts*. en-US. URL: https://nodexl.com (visited on 06/13/2022).

Thelwall, Michael (2009). "Web Crawling: SocSciBot". In: *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. Cham: Springer International Publishing, pp. 71–79. ISBN: 978-3-031-02261-6. DOI: 10.1007/978-3-031-02261-6_6. URL: https://doi.org/10.1007/978-3-031-02261-6_6.

University of Wolverhampton (n.d.). *SocSciBot: Link crawler for the social sciences*. URL: http://socscibot.wlv.ac.uk/ (visited on 06/13/2022).

Wayback Machine (n.d.). *Internet Archive: Wayback Machine*. URL: https://archive.org/web/ (visited on 06/01/2022).

Xie, Haoran et al. (2016). "Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy". In: *Information Processing & Management* 52.1, pp. 61–72.