

Proceedings from the Document Academy

Volume 11
Issue 1 *Proceedings from the 2024 FanLIS
Symposium*

Article 6

2024

Utilizing Metadata from Heterogeneous Sources within the Framework of the JVMG and GOLEM Projects to Identify Patterns in Anime-based Fandoms on AO3

Zoltan Kacsuk
Stuttgart Media University, kacsuk@hdm-stuttgart.de

Xiaoyan Yang
University of Groningen, xiaoyan.yang@rug.nl

Saskia Dreßler
Stuttgart Media University, University of Tübingen, sd129@hdm-stuttgart.de

Federico Pianzola
University of Groningen, f.pianzola@rug.nl
Follow this and additional works at: <https://ideaexchange.uakron.edu/docam>



Martin Roth
Part of the [Digital Humanities Commons](#)
Stuttgart Media University, Ritsumeikan University, roth1003@fc.ritsumei.ac.jp
Please take a moment to share how this work helps you [through this survey](#). Your feedback will be important as we plan further development of our repository.

Recommended Citation

Kacsuk, Zoltan; Yang, Xiaoyan; Dreßler, Saskia; Pianzola, Federico; and Roth, Martin (2024) "Utilizing Metadata from Heterogeneous Sources within the Framework of the JVMG and GOLEM Projects to Identify Patterns in Anime-based Fandoms on AO3," *Proceedings from the Document Academy*: Vol. 11 : Iss. 1 , Article 6.

DOI: <https://doi.org/10.35492/docam/11/1/6>

Available at: <https://ideaexchange.uakron.edu/docam/vol11/iss1/6>

This Conference Proceeding is brought to you for free and open access by University of Akron Press Managed at IdeaExchange@Uakron, the institutional repository of The University of Akron in Akron, Ohio, USA. It has been accepted for inclusion in Proceedings from the Document Academy by an authorized administrator of IdeaExchange@Uakron. For more information, please contact mjon@uakron.edu, uapress@uakron.edu.

Introduction

The fact that online fan communities build high-quality large databases and inventories on both source and fan works as well as archives for the latter has now been well documented (see for example De Kosnik (2021), Pfeffer & Roth (2020), Yin, Aragon, Evans & Davis (2017)). Both databases on source materials and archives of fan works employ a range of different data organization approaches in order to facilitate finding works that have certain characteristics or are similar to other works in some regard (Johnson, 2014, Price, 2019). A number of studies have drawn attention to the similarity these practices and solutions represent vis-à-vis library and information science principles and practices as well as drawing attention to some of the unique aspects of this information behaviour (Price, 2019, Price & Robinson, 2021, Bourlai, 2018, Bullard, 2019). These online resources constitute an outstanding asset for studying not only fan information behaviour (De Kosnik, 2021, Price, 2019, Price & Robinson, 2021, Bourlai, 2018, Price & Robinson, 2017, Hill & Pecoskie, 2017), but also wider questions in relation to fandom and fan production, source and fan works, as well as, for example, reading practices and communities and literary evolution (De Kosnik, 2021, Kiryakos & Pfeffer, 2021, Pianzola, 2020, Pianzola, Acerbi, & Rebora, 2020, Pianzola, 2021, Moore, 2023).

This is the primary reason the Japanese Visual Media Graph (JVMG) project – funded by the German Research Foundation’s (Deutsche Forschungsgemeinschaft, DFG) e-Research Technologies program – and the Graphs and Ontologies for Literary Evolution Models (GOLEM) project – funded by the European Commission’s ERC Starting Grant – are both premised on working with data and archives compiled by online fan communities to help build research infrastructures for researchers working with popular Japanese visual media and narrative and fiction respectively. Both teams are developing knowledge graphs¹ in our respective fields, and due to the overlap in research interests between the two projects, as well as the affordances of linked data, we are working together on connecting the data to enable research on questions that lie at the intersection of these two domains.

As a first step in this joint effort we attempt to identify patterns in fully or in-part anime-based fandoms on Archive of Our Own (AO3). In this we follow previous quantitative explorations (Yin, et al. 2017, Moore, 2023), which have attempted to engage with available metadata on fanfiction at scale in order to provide context for qualitative and/or more in-depth analyses of fanfiction content, production and consumption as well as the social interactions surrounding them by painting a

¹ Knowledge graphs store information in a graph structure, often as linked data in RDF format.

more granular overall picture of the landscape of fanfiction as it is found on some of the most popular online archives. However, we also move beyond the scope of these prior work by incorporating metadata on the source materials as well in our analysis, thereby helping to establish a connection in these type of large-scale quantitative explorations to one of the original questions at the root of fan studies research (Radway, 1991[1984], Bacon-Smith, 1992, Jenkins, 1992), namely the connection of fan readings and production to the source works themselves.

The JVMG and GOLEM projects

One of the initial ideas behind the creation of the Japanese Visual Media Graph was that even researchers, who are experts in their domain, often rely on online databases created by fan communities to look up details about works of manga and anime. The amount of high quality data in these online resources is quite astounding, and potentially combining them into one unified knowledge graph would provide a very powerful research tool. Thus, the aim of the JVMG project is to work together with online enthusiast communities directly – or through following the guidelines they have set out for using their data – in order to combine their databases into an integrated knowledge graph that can be used for research on Japanese anime, manga, video games and related domains (Pfeffer & Roth 2020).

The JVMG knowledge graph currently has four main community sources (AnimeClick, The Visual Novel Database, Anime Characters Database (ACDB), and the publicly available dump of all anime titles from AniDB) and two further data sources (Wikidata and Media-Arts Database). The data in the JVMG project is made available under a CC BY-NC-SA 4.0 license.

The Graphs and Ontologies for Literary Evolution Models project aims to explore the transmission of story traits across different languages through cross-cultural analysis of fiction focusing on two primary research questions. What linguistic, stylistic, and thematic elements in stories gain popularity across different languages? What makes a story widely read, and what do readers value most in a story? To answer these questions the project combines computational literary studies with cultural evolution to investigate how stories evolve and spread across six languages: English, Spanish, Italian, Chinese, Korean, and Indonesian. This multilingual approach allows the GOLEM project to delve into how cultural and linguistic factors influence the popularity and evolution of fictional narratives across a range of different cultures and language domains.

Three types of data resources will be integrated into the GOLEM knowledge graph. First, fanfiction archives, which include Archive of Our Own (AO3) as the

primary platform, as well as Fanfiction.net, Asianfanfics, EFP, and Lofter. Second, original fiction platforms, such as Wattpad, WebNovel, megustaescribir, and Naver Series. Third, canon information sources, including the Japanese Visual Media Graph, Wikipedia, and Fandom.net. The GOLEM knowledge graph will also be made publicly accessible as soon as possible.

Scope and data alignment

Within the JVMG project we follow the Tiny Use Case (TUC) workflow methodology for research and development originally pioneered in the diggr (Databased Infrastructure for Global Games Culture Research) research project² and inspired by Agile software development practices (Freybe et al., 2019). The aim of each TUC is to work on a clearly defined research task that can be accomplished in a limited time frame, with the results and lessons learned throughout the process of a given TUC then feeding back into the development process of the JVMG knowledge graph. In this collaboration between the JVMG and GOLEM projects we decided to adopt this same incremental approach to achieving interoperability between our knowledge graphs, and to explore the opportunities opened up by combining the two.

Thus, we decided to focus on fully or in-part anime based fandoms in the AO3 data in the GOLEM knowledge graph, as anime is the most important medium for the global circulation of Japanese popular culture, and as such it has both the highest number of corresponding fandoms in the international fanfiction archive AO3 from among our domains of interest as well as a relatively wide metadata coverage in the JVMG knowledge graph. We started by examining all fandoms under the general tag “Anime & Manga”, of which there were 4,001 in the GOLEM dataset. First, we removed 3,202 fandoms that had less than 50 stories (leaving 799 fandoms), as these would be most likely too small for any meaningful longitudinal analysis. Second, we removed 145 fandoms that had no relationship with anime (leaving 654 fandoms) – typically these were fandoms that were based only on works of manga/manhwa/manhua or in some cases video games. A further 120 fandoms were removed from the scope of this analysis due to technical issues in relation to data retrieval – one of the important feedbacks from this TUC for the infrastructure development side of our projects – which made their contents unreliable for analysis. In this way we ended up with a final population of 534 fully or in-part anime based fandoms and almost 1,290,000 corresponding stories from the GOLEM knowledge graph.

² <https://diggr.link/>

Next, these fandoms were first automatically matched with anime works in the JVMG data, and then we conducted assisted manual matching to finalize the matching process. The assisted manual matching consisted of a pregenerated ordering of all anime works in the JVMG knowledge graph based on ngram similarity for each fandom separately, which allowed for a relatively fast manual checking of whether every work had been appropriately matched with each fandom. Which brings us to one of the most important challenges that the data alignment process between the two datasets entailed, namely the fact that official AO3 fandom tags have highly varying levels of aggregation in relation to the source materials they relate to.

In order to be able to properly describe this challenge of the heterogeneity of official AO3 fandom tags, we need to briefly also highlight the complicated nature of talking about the supra-work level. *Multimedia/transmedia/media franchise* – or just franchise for short – is the expression most commonly employed in the English speaking world, while *media mix* is the term native to the Japanese production system (Steinberg 2012).³ We also have the concept of *superwork*, developed in library and information science to describe these types of entities (Kiryakos & Sugimoto, 2018, Lee et al. 2018, Zevio, Mihara, & Sugimoto, 2018). And there are, of course, a range of further concepts and frameworks, which all try to somehow capture this supra-work level to which individual works are related in one way or another, such as, for example, the concept of *transmedial worlds* by Tosca and Klastrop (2019). In the following we will be using the term franchise to refer to this supra-work level, not as a way to disregard the complexity of the entity being tackled here, but rather as a shorthand for these and further potential interpretations without taking a more explicit position on this question for our present discussion.

Turning our attention to official AO3 fandom tags, the simplest case are fandoms that correspond to a single franchise like *Naruto*. Then, there are fandoms that seem to go beyond a single franchise such as *Gundam & Related Fandoms*, where *Gundam* is not the title of any single *Gundam* work, but rather the name of the franchise itself. As well as fandoms that are either a contents based slice (like *Universal Century Gundam* focusing on a particular timeline of *Gundam*) or a media specific subset (such as the self-explanatory *Attack on Titan (Movies)*) or both of a given franchise (for example *Saint Seiya (Manga & Anime 1986)*). Finally, there are work level fandoms (such as *Yu-Gi-Oh! Sevens (Anime)*), which correspond to a particular realization of a given story in a specific medium. These latter are the only fandoms that have one-to-one corresponding anime works in

³ For an analysis of the similarities and differences between transmedia franchise and media mix, see Kopylova (2016).

the JVMG data by definition. For all other fandoms there are often multiple corresponding anime works. Indeed, almost 60% of all analyzed fandoms were matched with more than one anime work.

There were 1,734 anime works that were matched with the analyzed fandoms from the JVMG knowledge graph. For each of these works we manually checked and added missing information on publication dates, genre, ACDB content rating and AnimeClick user rating. Genre information came from both ACDB genre tags (106 different values) and AnimeClick genre categories (52 possible values). In order to create a more parsimonious and unified genre category system, we created the JVMG genre variable (30 distinct values) with all ACDB genre tags and AnimeClick genre categories assigned to a single JVMG genre. It is important to note here that not only were there potentially more than one data sources in relation to genre within the JVMG knowledge graph for a single work, but due to the frequent one to many matches between fandoms and anime works there were also often multiple different values for content rating and user rating – not to mention genre categories – for a single fandom, which also made the analysis a bit more challenging, as explained in the next section.

Analysis

Having aligned the 534 fully or in-part anime based fandoms in the GOLEM data with their corresponding anime releases from the JVMG knowledge graph, and having made sure that we had as complete a metadata dataset as possible for the anime works, we proceeded to explore the AO3 fanfiction publication numbers and distributions through the lens of the added metadata on their respective anime source works. It is important to keep in mind that a part of the fandoms also had source materials in other media.

We decided to explore the data from multiple angles to have a better overall picture of the relationship between fanfiction publication numbers and patterns and the corresponding anime releases. First, we checked to see whether there are any correspondences between the various metadata about the source works in the JVMG data and the number of stories per fandom. Second, we examined the available data with a more complex data modeling approach to capture the temporal effect of anime releases on fanfiction publication numbers. Third, we created individual graphs of all fandoms with the fanfiction publication numbers over time overlayed with the release times of the source anime. These graphs were then manually analyzed and sorted to find potential patterns in the correspondences between fanfiction publication timelines and anime release dates.

1) Correspondences between metadata on source works and the number of stories per fandom

We found that there was a weak positive correlation (0.22) with the number of anime releases that belonged to a fandom and its number of published fanfiction. This was in line with our expectation that the release of new source works should increase the number of fanfiction stories. Interestingly, we found no real correlation in fanfiction story numbers with AnimeClick user rating nor with ACDB content rating. The first being a measure of how well liked the anime release is on the AnimeClick website, and the second being an indication of what audience the content is appropriate for (with values *Everyone*, *E10+*, *Teen*, *Mature*, *Adult*). Since a large number of fandoms had multiple anime works matched with them, both these correlations were explored for average, maximum and minimum values to account for the potentially differing AnimeClick user rating and ACDB content rating⁴ values for the different anime releases. However, beyond slight variations (see Table 1 below), there was no major difference in the overall results based on these three ways of calculating the correlation values, as they were all quite close to showing almost no to extremely weak correlations. In summary, while new releases corresponded to more fan activity, neither the user rating or the age appropriateness seems to be an influential factor on overall fanfiction publication numbers.

	mean	maximum	minimum
<i>AnimeClick user rating</i>	-0.001	0.048	-0.059
<i>ACDB content rating</i>	0.024	0.059	-0.007

Table 1.: Correlation of fanfiction story numbers with AnimeClick user rating and ACDB content rating

There are some differences according to the genre categories of the source anime, most notable the shōnen genre label being the most impactful for mean number of stories per fandom, meaning that shōnen fandoms tended to be more productive on average (see Table 2 below). Shōnen is a demographic genre label specific to Japanese visual media meaning younger boys, most well known through the manga magazine *Weekly Shōnen Jump* and the many global hits that have emerged from its pages like *Dragon Ball*, *Slam Dunk*, *Naruto*, *One Piece* or *My Hero Academia*. The predominance of the shōnen genre persisted even when removing the top four fandoms – all of which are shōnen and which together make up almost 40% of all analyzed stories – from the analysis to control for outliers.

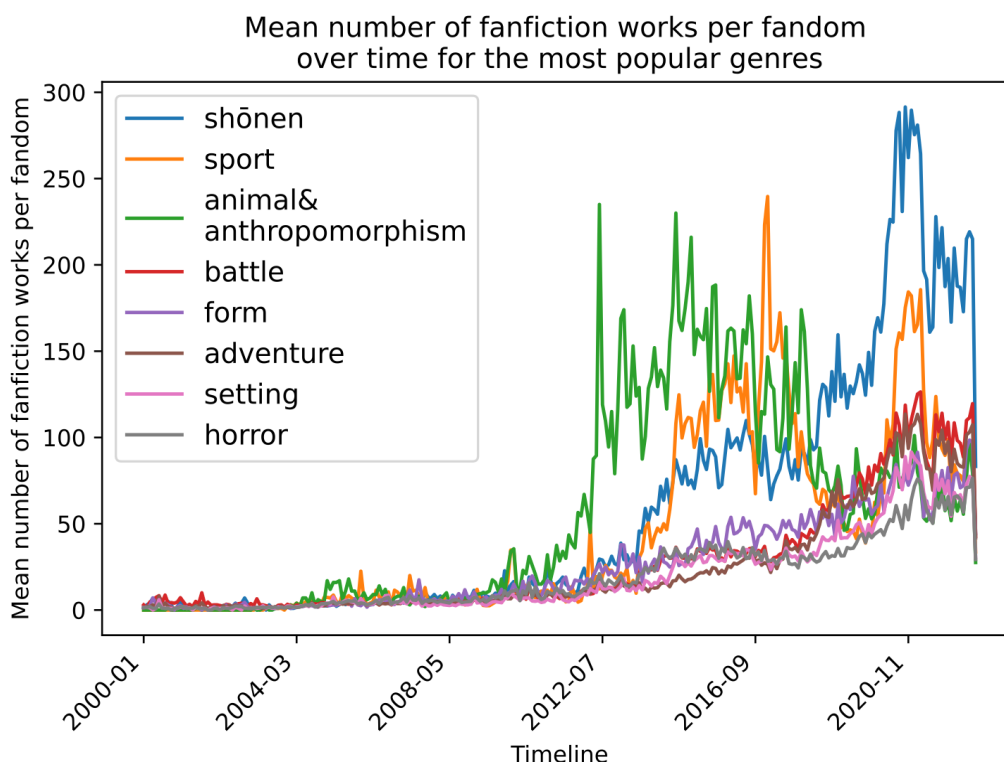
⁴ For ACDB content rating the categorical values were translated to a scale going from one to five to be able to calculate correlation values.

JVMG genre	Total number of fanfiction stories	Number of fandoms	Mean number of stories per fandom
<i>shōnen</i>	757577	73	10377.8
<i>sport</i>	245968	39	6306.9
<i>animal & antropomorphism</i>	63486	11	5771.5
<i>battle</i>	690594	148	4666.2
<i>form*</i>	219053	50	4381.1
<i>adventure</i>	647489	158	4098.0
<i>setting*</i>	774030	256	3023.6
<i>horror</i>	462430	153	3022.4
<i>action</i>	873658	297	2941.6
<i>drama</i>	464037	165	2812.3
<i>comedy</i>	681781	259	2632.4
<i>fantasy</i>	907199	349	2599.4
<i>science-fiction</i>	236428	103	2295.4
<i>suspense</i>	316216	147	2151.1

Table 2.: All JVMG genre categories with more than a thousand mean number of stories per fandom

* Form and setting are “artificial” JVMG genre categories collecting all genre tags and categories that reference some kind of formal or setting related element.

Graph 1 below shows the distribution over time of the average number of fanfiction stories per fandom published on AO3 by genre category of the source material for the top genre categories according to this view. As already discussed above, *shōnen* is the genre category most strongly related to high levels of output per fandom, which we can see in the way *shōnen* starts to overshadow every other genre category from around 2018, which is in part due to the story output of the *My Hero Academia* fandom (responsible for an astounding close to 17% of all analyzed stories). Interesting to note, however, is the pronounced lead of the animal and anthropomorphism category before this era of *shōnen* dominance, as well as the spikes in the sports category. Since there are very few fandoms that belong to the animal and anthropomorphism category, the large numbers in fanfiction production for *Hetalia: Axis Powers* drives the rise of this category in that particular period. Similarly the category sport receives huge bursts of popularity from series like *Yuri!!! on Ice* (spiking around December 2016) and *Haikyuu!!* (responsible for the spike in the sports genre roughly between mid-2020 and mid-2021), underlining how fanfiction production can often be driven by certain fandoms central at the time. It is worth noting that both *shōnen* and sports narratives often focus on groups of male characters, with a similar narrative structure focusing on competition, fighting, and characters wanting to become better/stronger – with quite a few works sharing both genre labels.



Graph 1.: Mean number of fanfiction works per fandom over time for the most popular genres according to this variable

2) Fixed effects modeling for examining the impact of anime releases on fanfiction publication numbers

Next up we used fixed effects modeling to explore whether we can find a quantifiable effect of anime releases on fanfiction publication numbers. We explored a range of different model settings, described in detail below. The inclusion of the lagged effect in some of the models (2, 4, 6, 8) aims to account for the fact that fanfiction production might not take off immediately as the anime is released, but maybe only during the following month. In a similar way the time window was introduced in some models (3, 4, 7, 8) to account for the fact that the surge in fanfiction output due to a new anime release might not concentrate in a single month but rather could be spread out over multiple months. We also examined the results of fixed effect modeling for a smaller subset of fandoms (Models 5-8 below), where the number of stories was at least a thousand in order to provide more robust longitudinal data for the model.

The fixed effects model basic setting:

$$Count_{ft} = a_1 + b_1 * Publish_flag_{ft} + c_1 * AnimeClick_rating_mean_f + d_1 * t + e_{ft}$$

Count_{ft}: the number of new AO3 fanfiction creations for each franchise (*f*) in time period (year-month) (*t*).

Publish_flag_{ft}: whether franchise *f* had new anime releases in time period (year-month) *t*, where a value from 1 to 3 indicates the number of releases, and 0 indicates no releases.

AnimeClick_rating_mean_f: the mean audience rating on AnimeClick for franchise *f*, serving as a control for fixed effects of the franchise.

t: the fixed effects of time (year-month).

e_{ft}: the random disturbance affecting **Count**.

Model 1 uses a fixed effect model with the basic setting for 534 franchises.

Model 2 uses *Publish_flag_lag1* instead of *Publish_flag* for regression analysis for 534 franchises. *Publish_flag_lag1* is the one-period lagged version of *Publish_flag* based on time.

Model 3 uses *Publish_flag_window3* instead of *Publish_flag* for regression analysis for 534 franchises. Assuming each anime release has a promotional peak period of three months, a three month window starting from the release month is created for each release to obtain *Publish_flag_window3*.

Model 4 uses *Publish_flag_lag1_window3* instead of *Publish_flag* for regression analysis for 534 franchises. *Publish_flag_lag1_window3* combines the two methods above by first applying the one-period lag and then creating a three month window.

Model 5 uses a fixed effect model with the basic setting for the top 104 franchises with at least a thousand fanfiction stories.

Model 6 uses *Publish_flag_lag1* instead of *Publish_flag* for regression analysis for the top 104 franchises with at least a thousand fanfiction stories.

Model 7 uses *Publish_flag_window3* instead of *Publish_flag* for regression analysis for the top 104 franchises with at least a thousand fanfiction stories.

Model 8 uses *Publish_flag_lag1_window3* instead of *Publish_flag* for regression analysis for the top 104 franchises with at least a thousand fanfiction stories.

Models for all 534 franchises				
	Model 1	Model 2	Model 3	Model 4
<i>Fixed Effects of Anime Releases</i>				
<i>Publish_flag</i>	13.06***			
<i>Publish_flag_lag1</i>		14.92***		
<i>Publish_flag_wind</i>			16.37***	

<i>ow3</i>				
<i>Publish_flag_lag1_window3</i>				20.36***
<i>Fixed Effects of Time</i>				
From 2015-02 the <i>t</i> (time) variable starts to show significant influence on the fanfiction <i>Count</i>				
<i>Model Fitting Statistics</i>				
<i>R-Squared</i>	0.023	0.023	0.024	0.024
<i>Adj. R-Squared</i>	0.017	0.017	0.018	0.018
<i>F-statistic</i>	9.779 on 273 and 112879 DF***	9.801 on 273 and 112879 DF***	10.021 on 273 and 112879 DF***	10.19 on 273 and 112879 DF***
Models for top 104 franchises				
	Model 5	Model 6	Model 7	Model 8
<i>Fixed Effects of Anime Releases</i>				
<i>Publish_flag</i>	16.12 (p = 0.07)*			
<i>Publish_flag_lag1</i>		20.34 (p = 0.026)**		
<i>Publish_flag_window3</i>			24.77 (p = 0.07)*	
<i>Publish_flag_lag1_window3</i>				34.62***
<i>Fixed Effects of Time</i>				
From 2015-02 the <i>t</i> (time) variable starts to show significant influence on the fanfiction <i>Count</i>				
<i>Model Fitting Statistics</i>				
<i>R-Squared</i>	0.083	0.083	0.084	0.0845
<i>Adj. R-Squared</i>	0.071	0.071	0.072	0.072
<i>F-statistic</i>	9.334 on 273 and 28015 DF***	9.341 on 273 and 28015 DF***	9.401 on 273 and 28015 DF***	9.476 on 273 and 28015 DF***

Table 3.: Results for fixed effects modeling for the impact of anime releases on fanfiction publication numbers

* Result not statistically significant.

** Result only statistically significant at a five percent threshold.

*** Result statistically significant for all common thresholds (p < 0.001).

Overall, across the various models tested, the presence of new anime releases consistently demonstrated a significant positive effect on fanfiction creation at the time of or following right after the release date, however, there are a number of important caveats to consider.

In Model 1, which utilized the basic fixed effect model for 534 franchises, the coefficient of the *Publish_flag* variable is 13.06 (statistically significant for all common thresholds), meaning that on average a given fandom will have 13 more stories the month a new related anime is released. This positive effect is further

supported by Models 2, 3, and 4, which introduce variations in the *Publish_flag* variable to capture different temporal effects such as lagged and windowed effects already explained above. We chose these two adjustments because there is often a time lag of approximately one month before fanfiction production really starts taking off, and a common publicity period for an anime release is usually three months. Notably, Model 4, which combines both lagged and windowed effects, demonstrates the highest coefficient (20.36, statistically significant for all common thresholds), suggesting that considering both temporal dimensions yields the most comprehensive understanding of the impact of anime releases on fanfiction creation. But in general, all these four models have very low adjusted R-squared scores, which suggests that there are most likely other variables at play that have an effect on the production of fanfiction, which would need to be included in the model settings.

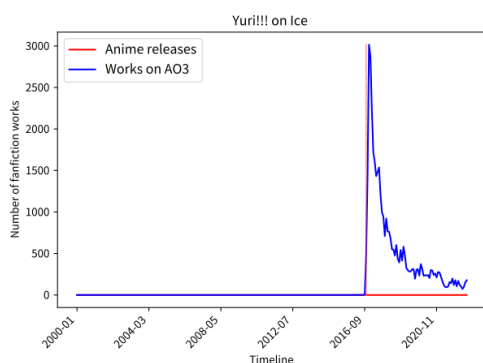
Furthermore, the analysis reveals interesting insights when focusing on only the top 104 franchises with at least a thousand fanfiction stories. Model 5, which applies a basic fixed effect model to this subset, exhibits a better fitting compared to the previous four models, yet the effect of anime release is not statistically significant. Moreover, Models 6, 7, and 8, which apply variations of lagged and windowed effects to the top franchises, further emphasize this trend. Particularly, Model 8, which combines both lagged and windowed effects for the top franchises, demonstrates the highest coefficient out of all the models (34.62, statistically significant for all common thresholds), highlighting the heightened impact of anime releases on fan engagement within the most prominent fandoms for a three month window following the release date. However, the still very low adjusted R-squared scores nevertheless suggest that further variables need to be considered to be able to provide a more complete explanation for the way fanfiction production numbers change over time.

3) Manual examination of graphs of fanfiction publication numbers over time together with the release times of the source anime

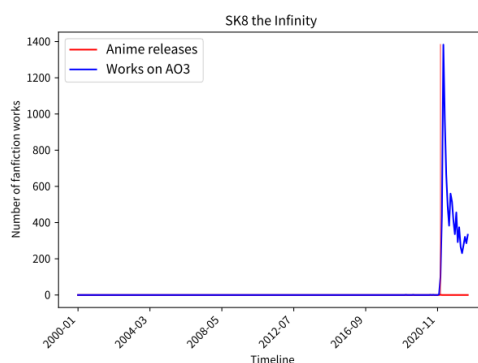
Beyond working with all the data together in the above detailed ways, we have also explored the top four hundred fandoms (based on fanfiction story numbers) as individual publication timelines together with their anime release dates. We found a large variation in patterns, with the two most numerous and clearly identifiable being: 1) A sudden peak in fanfiction publications right after the anime release followed by a decline in interest. Examples of this pattern were often original sports anime with only one season released so far, for example *Yuri!!! on Ice* (see graph 2 below), *SK8 the Infinity* (see graph 3 below), but fandoms based on other genres, such as the organized crime and gang themed *91 Days* and *Banana Fish* also conformed to this pattern. 2) The anime was released

much earlier than the fanfiction stories' publication dates. The *Digimon Frontier*, *Monster*, and *Trinity Blood* fandoms are all examples of this pattern.

(Note: In all the graphs below the blue trend lines indicate the development of the fanfiction publication outputs, and the red vertical lines the anime release dates.)



Graph 2.: *Yuri!!! on Ice*



Graph 3.: *SK8 the Infinity*

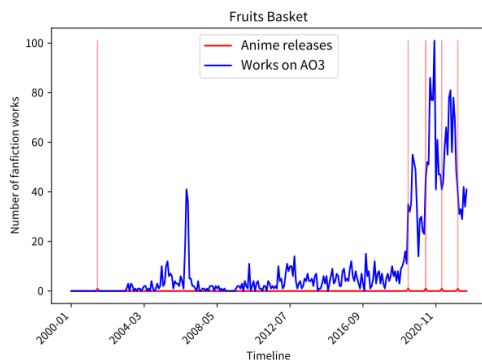
Examples of further clearly visible, however, far less numerous patterns included:

- 3) Fandoms with multiple anime releases and a constantly increasing trend of fanfiction publication numbers. Fandoms corresponding to blockbuster franchises like *Naruto* and *One Piece* demonstrated this pattern.
- 4) Fandoms with comparable fanfiction publication activity before the anime release to that of the fanfiction output following it (e.g. *Wild Adapter*, *Saiyuki Gaiden*).
- 5) Fandoms with multiple anime releases that have a pattern of fanfiction production that more or less conforms to our expectation of anime releases triggering and/or rekindling interest in the fandom. *Fruits Basket* (see graph 4 below) and *Jujutsu Kaisen* (see graph 5 below) are two examples of this pattern.

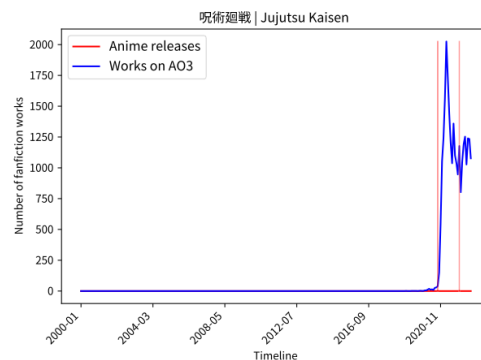
For *Fruits Basket* the first anime release came way before the official launch of AO3 in November 2009, however, looking at the later anime releases, we can see how interest in the fandom experiences a revitalization and growth period after each release, which might be due to a mix of both nostalgia and an influx of new viewers/readers (a point which would require further research to explore and potentially verify). One supporting argument that we do have from our current exploration and data, that this is indeed the case, is that this type of pattern seems to be more prevalent for older fandoms. One more interesting point here is the very obvious publication spike in 2006. This is due to a larger number of stories from August and September that year, which are most likely related to either a final plot development of the story or an anticipation of its end, which happened in November that year. We would, of course, need to engage with the stories

themselves to be able to explore what could have been the real root cause of this publication spike.

As for *Jujutsu Kaisen*, we can see that there was only minor activity in the fandom prior to the release of the first anime (especially in comparison with the scale of interest that came after). The second anime release on the graph is the anime movie *Jujutsu Kaisen 0*, which seems to have revitalized the interest in the fandom to a certain degree.

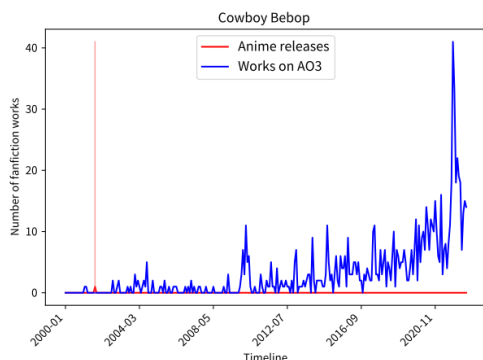


Graph 4.: *Fruits Basket*

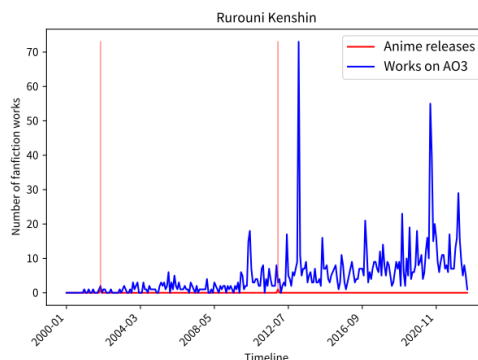


Graph 5.: *Jujutsu Kaisen*

There are also multiple fandoms that have patterns that do not fit any of the above listed ones, and also many which do not conform to our expectations of anime releases providing a driving force for fanfiction production. One reason for this is that it is not just anime releases, but also, for example, holiday seasons and the increase in the general popularity of AO3 that affect fanfiction publication numbers over time, as well as other media releases related to the fandoms. Graph 6 and 7 provide two such examples. On the left, on graph 6 we can see how a large spike in fanfiction production for *Cowboy Bebop* coincides with the November 2021 release of the Netflix live-action adaptation of the anime. On the right, on graph 7 we have the example of the *Rurouni Kenshin* fandom, where the largest spike in fanfiction publication numbers could be the result of a video game released on August 30, 2012.



Graph 6.: *Cowboy Bebop*



Graph 7.: *Rurouni Kenshin*

Summary

This paper has provided an overview of the first step in the data alignment and joint research efforts between the JVMG and GOLEM projects. Focusing on fully or in-part anime based fandoms in the AO3 data in the GOLEM knowledge graph we connected 534 fandoms with 1,734 anime works from the JVMG knowledge graph. We then proceeded to analyze the relationship between the fanfiction production numbers as well as timelines and the various metadata available on the corresponding anime works, such as release date, genre, content rating and user rating. We found a weak positive correlation between the number of anime works connected to a fandom and the fanfiction production numbers thereof. Furthermore, the shōnen genre category proved to have the strongest effect on the average story counts of fandoms. Using fixed effects modeling we were able to identify a positive effect of anime releases on fanfiction publication numbers at the time of or following right after the release. However, it is clear from the very low overall model strengths that there must also be other variables responsible for explaining the change over time in fanfiction publication numbers. Finally, we manually examined individual graphs of the top four hundred fandoms' fanfiction publication numbers over time overlayed with the corresponding anime releases. Despite there being a large heterogeneity in the patterns found in these graphs, we were able to identify a number of distinct types, as well as groups of fandoms for which the invigorating effect of anime releases on fanfiction publication numbers could be clearly seen. We hope that these examinations have demonstrated that enriching the metadata on fanfiction archives like AO3 with information on the respective source works can help fill in gaps and offer a more comprehensive picture in understanding the shape that fan activities take over time. We look forward to further detailing this picture with our future joint work on characters, that are so central to much of fanfiction, between the JVMG and GOLEM projects.

Statement on funding

The Japanese Visual Media Graph (JVMG) project is funded by the German Research Council (Deutsche Forschungsgemeinschaft, DFG). Martin Roth is one of the project leads, and Saskia Dreßler and Zoltan Kacsuk are project members of the JVMG project. The Graphs and Ontologies for Literary Evolution Models (GOLEM) project is funded by the European Commission's ERC Starting Grant. Federico Pianzola is principle investigator and Xiaoyan Yang is a project member on the GOLEM project.

References

- Bacon-Smith, C. (1992). *Enterprising women: Television fandom and the creation of popular myth*. Philadelphia, PA: University of Pennsylvania Press.
- Bourlai, E. E. (2018). 'Comments in tags, please!': Tagging practices on Tumblr. *Discourse, Context & Media*, 22, 46-56.
- Bullard, J. (2019). Curated folksonomies: Three implementations of structure through human judgment. *Knowledge Organization*, 45(8), 643-652.
- De Kosnik, A. (2021). *Rogue archives: Digital cultural memory and media fandom*. Cambridge, MA: MIT Press.
- Freybe, K., Rämisch, F., & Hoffmann, T. (2019). With small steps to the big picture: A method and tool negotiation workflow. In S. Krauwer & D. Fišer (Eds.), *Proceedings of the Twin Talks Workshop at DHN 2019 (CEUR Vol-2365)* (pp. 13-24). Aachen: CEUR Workshop Proceedings. http://ceur-ws.org/Vol-2365/03-TwinTalks-DHN2019_paper_3.pdf
- Hill, H., & Pecoskie, J. J. L. (2017). Information activities as serious leisure within the fanfiction community. *Journal of Documentation*, 73(5), 843-857. <https://doi.org/10.1108/JD-04-2016-0045>
- Jenkins, H. (1992). *Textual poachers: Television fans & participatory culture*. London: Routledge.
- Johnson, S. F. (2014). Fan fiction metadata creation and utilization within fan fiction archives: Three primary models. *Transformative Works and Cultures*, 17. <http://dx.doi.org/10.3983/twc.2014.0578>
- Kiryakos, S., & Pfeffer, M. (2021). Exploring the research utility of fan-created data in the Japanese visual media domain. In *International Conference on Asian Digital Libraries* (pp. 210-218). Cham: Springer International Publishing.
- Kiryakos, S., & Sugimoto, S. (2018). The representation of a multimedia franchise as a single entity: Contrasting existing bibliographic entities with web-based superwork portrayals. *LIBRES: Library and Information Science Research Electronic Journal*, 28(2), 40-57.
- Kopylova, O. (2016). *Media mix as adaptation: With Maeda Mahiro's Gankutsuō as an example* (Ph.D. dissertation). Kyoto: Kyoto Seika University.

- Lee, J. H., Jett, J., Cho, H., Windleharth, T., Disher, T., Kiryakos, S., & Sugimoto, S. (2018). Reconceptualizing superwork for improved access to popular cultural objects. *Proceedings of the Association for Information Science and Technology*, 55(1), 274-281.
- Moore, L. (2023). Fanfiction today: An analysis of publishing trends on Archive of Our Own. *SLIS Connecting*, 12(1), article 10. <https://aquila.usm.edu/sliscconnecting/vol12/iss1/10>
- Pfeffer, M., & Roth, M. (2020). Japanese Visual Media Graph: Providing researchers with data from enthusiast communities. In K. Golub et al. (Eds.), *2019 Proceedings of the International Conference on Dublin Core and Metadata Applications* (pp. 136-141). Seoul: Dublin Core Metadata Initiative (DCMI).
- Pianzola, F. (2020). Linked-Potter: An example of ontology for the study of the evolution of literature and reading communities. In B. Hodošček (Ed.), *JADH2020 Proceedings of the 10th Conference of the Japanese Association of Digital Humanities "A New Decade in Digital Scholarship: Microcosms and Hubs"* (pp. 28-32). Osaka: Graduate School of Language and Culture, Osaka University. <https://jadh2020.lang.osaka-u.ac.jp/programme/longpaper/pianzola-linked.html>
- Pianzola, F. (2021). *Digital social reading: Sharing fiction in the 21st century*. <https://doi.org/10.1162/ba67f642.a0d97dee>
- Pianzola, F., Acerbi, A., & Rebora, S. (2020). Cultural accumulation and improvement in online fan fiction. *CEUR Workshop Proceedings*, 2723, 2-11.
- Price, L. (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* (pp. 11-37). Baden-Baden: Ergon Verlag.
- Price, L., & Robinson, L. (2017). 'Being in a knowledge space': Information behaviour of cult media fan communities. *Journal of Information Science*, 43(5), 649-664. <https://doi.org/10.1177/0165551516658821>
- Price, L., & Robinson, L. (2021). Tag analysis as a tool for investigating information behaviour: Comparing fan-tagging on Tumblr, Archive of Our Own and Etsy. *Journal of Documentation*, 77(2), 320-358.
- Radway, J. A. (1991[1984]). *Reading the romance: Women, patriarchy, and popular literature* (2nd ed.). London: University of North Carolina Press.
- Steinberg, M. (2012). *Anime's media mix: Franchising toys and characters in Japan*. Minneapolis, MN: University of Minnesota Press.
- Tosca, S., & Klastrup, L. (2019). *Transmedial worlds in everyday life: Networked reception, social media, and fictional worlds*. New York, NY: Routledge.
- Yin, K., Aragon, C., Evans, S., & Davis, K. (2017). Where no one has gone before: A meta-dataset of the world's largest fanfiction repository. In

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6106-6110). <http://dx.doi.org/10.1145/3025453.3025720>

Zevio, S., Mihara, T., & Sugimoto, S. (2018). Identifying transmedia works from user-generated knowledge bases: Japanese pop culture study case. In *The Fifth International Workshop on Practical Application of Ontology for Semantic Data Engineering (PAOS2018)*.