# Cultural accumulation and improvement in online fan fiction

Federico Pianzola[a,b], Alberto Acerbi[c] and Simone Rebora[d,e]

[a] *University of Milan-Bicocca, Piazza dell'Ateneo nuovo 1, 20126 Milan, Italy*
[b] *Sogang University, 35 Baekbeom-ro, Daeheung-dong, Mapo-gu, Seoul, South Korea*
[c] *Brunel University London, Uxbridge, UB8 3PH, United Kingdom*
[d] *University of Verona, Lungadige Porta Vittoria 41, 37129 Verona, Italy*
[e] *University of Basel, Petersplatz 1, 4051 Basel, Switzerland*

**Abstract**

We analyse stories in Harry Potter fan fiction published on Archive of Our Own (AO3), using concepts from cultural evolution. In particular, we focus on cumulative cultural evolution, that is, the idea that cultural systems improve with time, drawing on previous innovations. In this study we examine two features of cumulative culture: accumulation and improvement. First, we show that stories in Harry Potter's fan fiction accumulate cultural traits—unique tags, in our analysis—through time, both globally and at the level of single stories. Second, more recent stories are also liked more by readers than earlier stories. Our research illustrates the potential of the combination of cultural evolution theory and digital literary studies, and it paves the way for the study of the effects of online digital media on cultural cumulation.

**Keywords**

cultural evolution, cumulative culture, Harry Potter, digital literary studies, fan fiction, literature

## 1. Introduction

In many cultural domains we can observe a progress through cumulative improvements. The efficiency of information storage, just to take a familiar example, increased through centuries with a series of innovations, in a process that continues today: a contemporary smartphone can store thousands of books. In cultural evolution theory this process is broadly described as cumulative cultural evolution [1, 2].

There are no universally agreed measures of cumulation [2], but it seems a sensible assumption that its degree differs in different cultural domains. In some domains, such as technology, we can observe clear marks of accumulation, while in others, such as arts, even if accumulation is not absent [3, 4], its scope appears limited. Many factors could explain why the degree of accumulation differs as such, but two are particularly interesting for cultural evolution. One is availability, that is, the number of possible models or cultural traits one has at disposal. While the exact details are debated [5, 6], the basic idea is that more versions of the same

*CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands*
EMAIL: federico.pianzola@unimib.it (F. Pianzola); alberto.acerbi@brunel.ac.uk (A. Acerbi); simone.rebora@univr.it (S. Rebora)
ORCID: 0000-0001-6634-121X (F. Pianzola); 0000-0001-5827-8003 (A. Acerbi); 0000-0002-1501-3774 (S. Rebora)

item available to copy from increase the probability that the item will be successfully preserved. The second is fidelity. Cultural transmission is a noisy process. Transmission chain experiments show that information transmitted orally tend to be lost and transformed [7]. In some cases, populations of individuals converge on similar cultural traits without the need of accurate passage of information, because of various constraints [8, 9]. In other cases, however, information needs to be preserved and successfully transmitted to allow the reproduction of a cultural trait [10]: modern complex technologies are preserved by virtue of supports that allows faithful transmission.

The contemporary diffusion of online digital media produces an increase of availability and fidelity in several domains [11]. A pertinent question is then what the consequences for cultural cumulation are; whether this results in more improvement, or perhaps more improvements in domains where it was limited before, because of relative lack of availability and fidelity.

To start answering this question we analyse cultural cumulation in AO3, the biggest fan fiction archive for stories in English. Previous research on fanfiction has found, for example, that authors improve their writing skills in time, widening the range of their vocabularies [12]. Here we focused on two features of cultural cumulation explicitly derived from cultural evolution theory: accumulation and improvement. Accumulation refers to an increase in the number of different cultural traits, while improvement refers to the fact that more recent traits are "better", according to some metrics. (The third feature of cumulative culture discussed in [11], ratcheting, that is, the fact the new innovations draw on previous innovations, is not analysed here).

Literary studies have theorized the historical increase in complexity (accumulation) of literature, for instance, with the introduction of formal innovations by literary modernism (e.g. stream of consciousness) and postmodernism (e.g. self-reflexivity) [13]. More problematic is the case of improvement. The quality of literary works has been traditionally judged by a restricted group of people, publishers and literary critics, who decided which texts deserved to be included in the literary canon, often based on a criterion of originality, i.e. the presence of some rhetorical or stylistic innovation in linguistic expression. This kind of institutionalised prestige has often been opposed to the popularity of bestselling fiction [14], appreciated by many because of its serial replication of known plot schemes or themes. More broadly, the issue concerns the contrast between a narrow selection of canonized works and the entirety of the archived literary production [15]. Fanfiction is an example of popular literature, produced in a context in which social exchange, collaborative work, and fidelity to the canon or to fandom tropes is the norm. In contrast, professional authors have often worked alone and pursuing originality in the history of literature. In this context, we think it is appropriate to talk about improvement in terms of cumulative culture, which focuses on the reception of cultural artefacts and their widespread popularity as a measure of improvement [2]. On the other hand, it might be the case that our model is not directly applicable to canonical literature because power relationships and prestige influence that cultural field more strongly than they do with fanfiction.

## 2. Methods

We chose to work with AO3 data because they are freely accessible, scraping of the website is allowed and supported by the developers maintaining the servers, and there is an excellent metadata system based on organized tags (Table 1). In addition, when uploading a story on

**Table 1**
AO3 stories' metadata

| Metadata | Description |
|---|---|
| Title | title of the story |
| Author | author of the story |
| URL | unique identifier of the story, assigned when the draft is created |
| Date | date of last update of the story |
| Summary | summary of the story, if written by the author |
| Language | language of the story |
| Words | number of words of the story |
| Chapters | number of chapters written and total number of chapters, if available (e.g. 1/1 or 15/?) |
| Completion | status of the story, either "completed" or "in progress" |
| Kudos | number of times someone liked the story |
| Bookmarks | number of times the story has been bookmarked |
| Hits | number of times the story has been viewed |
| Comments | number of comments left at the end of the story |
| Tag | label inserted by the author |
| Tag type | one of 7 categories: fandom, character, relationship, freeform, rating, archive warning, relationship orientation |

AO3, authors can add various kinds of tags, among which: *fandom tag*, to signal the fictional universe/es to which the story is related (e.g. "Harry Potter - J. K. Rowling"); *character tag*, to list the characters appearing in the story ("Hermione Granger"); *relationship tag*, to list the relationship/s in which the story characters participate (e.g. "Draco Malfoy/Harry Potter"); *freeform tag*, which can be related to any other aspect of the story (e.g. "POV Draco"), the fandom (e.g. "Community: daily_deviant"), or fanfiction writing's conventions (e.g. "Ron Weasley bashing").

AO3 has a system of tag wrangling, i.e. volunteers that continuously monitor newly introduced tags aggregating them with existing ones—without replacing them—when they refer to the same characters/relationships/themes. In our analysis we relied on this tag aggregation, which we implemented thanks to the creation of a linked-data knowledge base mapping all "synonym" tags used [16].

We collected all metadata of the stories tagged with the fandom tag "Harry Potter - J. K. Rowling" (217,772 stories), including all tags and information listed in Table 1. We excluded stories not in the English language, stories with less than 10 words, and stories published in 2020, obtaining a final sample of N = 196,726 stories. Fans started uploading their stories on AO3 since November 2009, when the archive became publicly accessible. However, some of the stories in AO3 are imported from previous publications or have been written in the past. In such cases, authors can backdate their stories, indicating a year earlier than the one of upload on AO3. Since we are using AO3 as a data source to study cultural accumulation in fan fiction, we want to consider the backdated year as the original date of these stories. With this procedure, the first year with more than 100 stories is the year 2002 (324 stories). We fixed this minimum threshold in order to group the stories in percentiles (see below). Overall, it should be kept in mind that data from earlier than 2010 are less representative of the Harry Potter fan fiction compared to later years, since only a portion of the stories published elsewhere has been imported. Moreover, this date adjustment can be used only for the accumulation analysis, but not for the improvement analysis, since hits and kudos started to accumulate only from the date of publication on AO3.

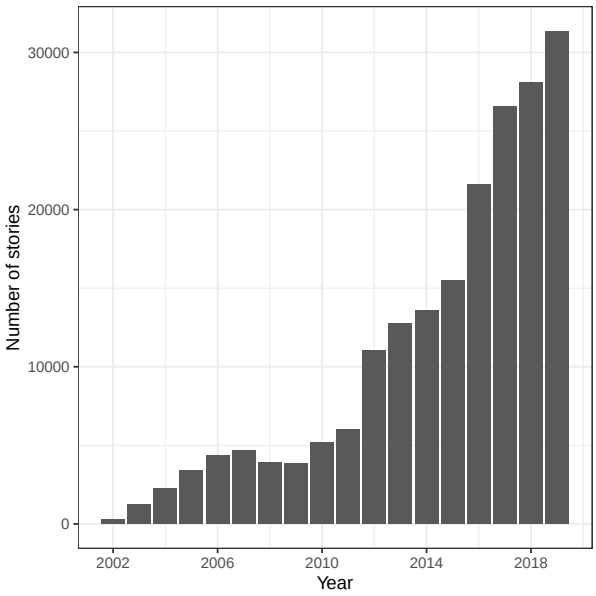


**Figure 1:** Number of stories in the AO3 archive we considered in the analysis.

## 2.1. Accumulation

In order to have accumulation, the number of cultural traits at time *t+1* should be higher than the number of cultural traits at time *t*. The hypothesis we test in our dataset is that the number of unique tags increases in time. We first test this hypothesis on the overall number of different tags, i.e. if the overall Harry Potter fandom accumulates traits (more characters, more relationships, more themes). Since the overall accumulation is associated to the parallel increase in the total number of stories (see Figure 1), we also check the average number of unique tags per story, i.e. if single stories accumulate traits (more characters, more relationships, more themes per story).

For the latter analysis, we confront stories with a similar popularity (defined by the number of kudos received) relative to the year in which they are published. There are two reasons for doing this. First, our cumulative culture hypothesis does not predict necessarily that all stories should increase the number of traits, only that the "best" recent stories should have more traits than the "best" earlier stories. Given the increase in the total number of stories, it could be that this effect is hidden when averaging on the total number of stories. Second, adding more tags can be a way to make a story more discoverable by readers with different interests and consequently increase the number of hits and the probability to receive kudos, so it makes sense to compare stories with similar popularity through years. To do this, we group the stories of each year in different percentiles according to the number of kudos and analyse the trend of the number of unique tags for the first, middle, and last decile.

## 2.2. Improvement

There is improvement when the cultural traits at time *t+1* are "better" (more effective according to some measure) than the cultural trait at time *t*. The hypothesis we test in our dataset is that the appreciation of stories increases in time, i.e. stories receive more hits and kudos.

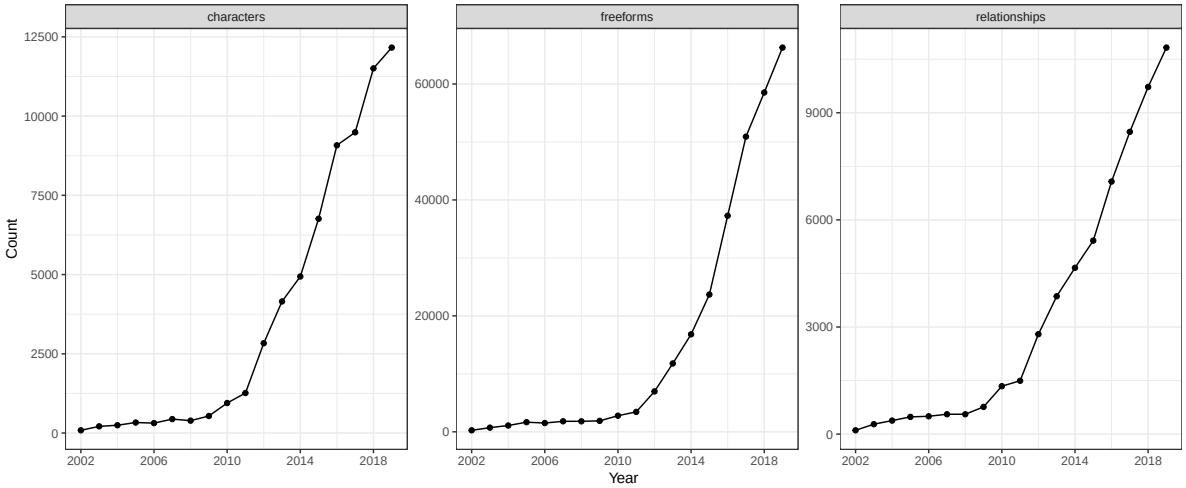


**Figure 2:** Number of unique tags per year, divided in "characters", "freeforms" and "relationships".

As we did for accumulation, we analyse both the global trend and the trend for single stories. A further issue concerns the choice of a reliable measure for stories' appreciation. For the analysis of global trend, we use three measures. The absolute number of kudos is simply the total sum of kudos received. This measure may favour very popular but less appreciated, in proportion, stories, and also favour older stories that have had more time to accumulate kudos. We thus use an additional second measure: the kudos/hits ratio. The ratio accounts for the age of a story, but it may favour "niche" stories, with very few hits and kudos. For this reasons, we finally consider a weighted measure, an engagement score $S$ computed as the true Bayesian average [17, 18] of the kudos/hits ratio, calculated as:

$$S = wK + (1 - w) * K_{av}$$

where $K$ is the kudos/hits ratio of the story, $K_{av}$ is the average of the ratio in a certain year, and $w$ is a weighting parameter, calculated as:

$$w = \frac{H}{H + H_{av}}$$

where $H$ is the number of hits for a certain story, and $H_{av}$ is the average number of hits per story for a certain year.

For single stories, we only compute the engagement score $S$, for the first, middle, and last decile.

# 3. Results

## 3.1. Accumulation

Figure 2 shows that there is accumulation of the overall number of unique tags.

We can also visualise how for each character, or group of characters, the number of freeforms and relationships tags increase, meaning that there is more diversity in stories. Figure 3 is an example for relationships that involve the three most popular characters. It shows the relative

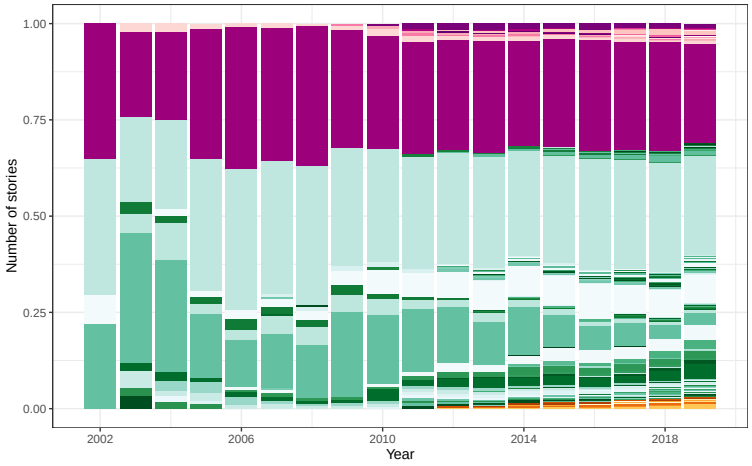


**Figure 3:** Variety and relative proportion of relationships tags for Harry Potter (green palette), Draco Malfoy (purple palette), and Hermione Granger (orange palette).

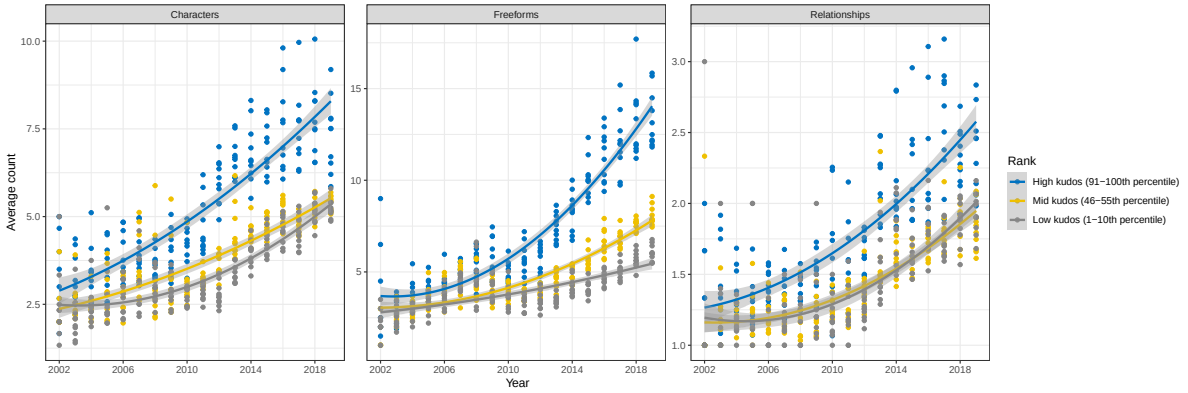


**Figure 4:** Yearly average number of character, relationship and freeform tags per story. Stories with the same popularity/quality (measured by kudos in percentiles) are compared.

proportion (on the total number of stories where the characters' relationships are tagged) of relationships tags for the three characters Harry Potter (green palette), Draco Malfoy (purple palette), and Hermione Granger (orange palette). It can be seen that diversity increases in time, since new relationships are introduced every year.

Looking at the average number of unique tags per story, Figure 4 shows that they also increase in time. This is true for stories with all levels of popularity/quality, as can be seen from the accumulation slopes of Figure 4, which are always positive from 2002 to 2019 for all three samples (low, mid, and high percentiles). However, the best stories (higher percentiles) have a faster pace of accumulation, as predicted by the cumulative culture hypothesis.

## 3.2. Improvement

Figure 5 shows the increase in time of popularity/quality in the stories. The overall number of kudos (central panel) decreases for recent stories. This highlights that older stories have had more time to accumulate kudos, and it is a widespread phenomenon, e.g. for citations

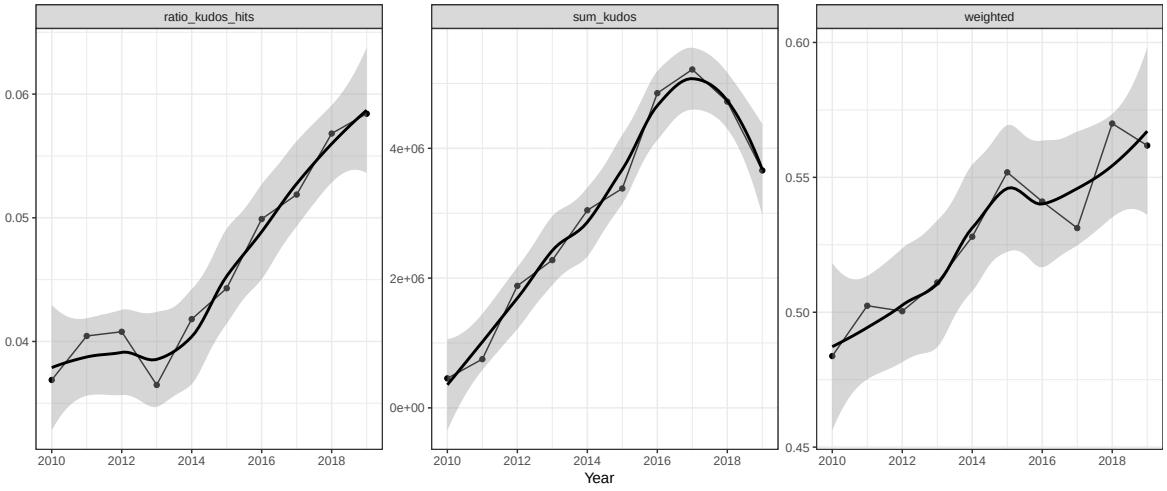


**Figure 5:** Yearly values of popularity/quality measured using the three different techniques described in the text.

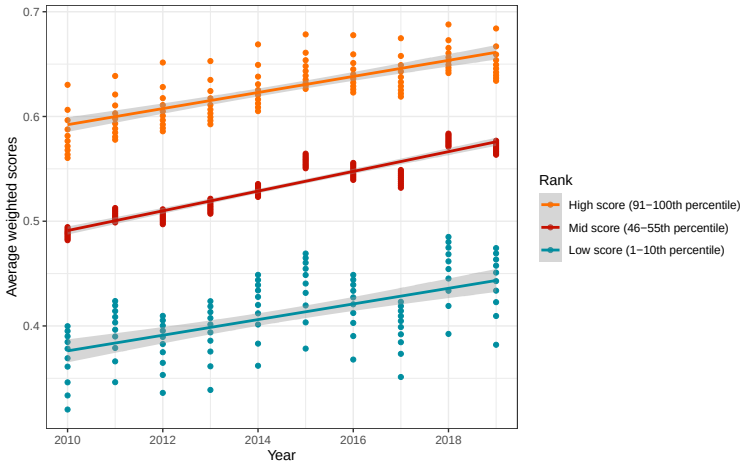


**Figure 6:** Yearly engagement score ($S$) comparing stories with the same popularity/quality (measured in percentiles).

of scientific articles [19]. When considering the ratio kudos/hits (left panel) or the weighted engagement score (right panel), readers' engagement increases in time. This is also true for single stories, for all three percentiles considered in our analysis (Figure 6).

## 4. Discussion

The results of our work in progress show that we can observe two features of cumulative culture, namely accumulation and improvement, in online fan fiction. Interestingly, the two features can be detected both at the global level, i.e. considering all stories, and when analysing single stories. While the former observation is expected, as matched by the parallel increase in the number of stories, the latter is more surprising, suggesting that individual stories became, through years, more complex and more appreciated by readers.

In future works, to provide a complete picture of cumulative culture in this domain, we plan to also analyse the "ratcheting" feature, that is, the fact that novel innovations draw on past ones. In our case, this would imply to detect, for example, that new introduced tags tend to appear in correspondence of specific previous tags, e.g. "Alternate Universe - Modern Setting" with "Alternate Universe - Canon Divergence".

In this research, and in the planned research on ratcheting, we focused on tags. The analysis of tags' frequency can be combined with techniques like text re-use detection and stylometry, to check whether older stories contain sentences that are "copied" or adapted in more recent stories (word frequency patterns), or to check text similarities concerning themes (topic modelling). Similarly, accumulation, or an increase in complexity, could be detected at the more fine-grained level of textual properties.

We are not aware of similar research for fiction published by institutionalized authors and publishers, but it would be interesting to compare cumulative cultural evolution in two similar domains, one which is supported by digital online media and one which is not. As discussed in the Introduction, an important research question is whether particular features of online digital media—increased fidelity and availability—support cultural accumulation.

Another aspect that would be interesting to take into account is the effect on fanfiction of the publication of further instalments of the original work (e.g. new books or films). An effect has been documented for fanfiction related to various fandoms, but not all of them [20]. For Harry Potter, for example, the release of the spin-off "Fantastic Beasts and Where to Find Them" in November 2016 increased the number of published stories in 2016 and 2017. However, on AO3 there was no such effect with the release of the last two films of the original series "Harry Potter and the Deathly Hallows" part 1 and part 2, in November 2010 and July 2011 respectively, because at the time a lot of Harry Potter fanfiction was being posted on Fanfiction.net, which indeed saw a huge increase in posting in those months [20].

More generally, we believe cultural evolution provides a fruitful theoretical background for digital literary studies, introducing a new range of testable hypotheses that can elucidate some of the dynamics studied by digital humanities. On the other hand, digital texts (and the sophisticated methodologies developed in literary studies and Natural Language Processing to analyse them) are an interesting and increasingly data-rich domain for cultural evolutionists. We hope our research will contribute to the further exploration of this interdisciplinary space.

## Acknowledgments

Thanks to fffinnagain, destinationtoast, Shay Guy, and the other fans who publicly shared their statistics about fanfiction.

## References

[1] J. Henrich, The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter, Princeton University Press, Princeton & Oxford, 2015.

[2] A. Mesoudi, A. Thornton, What is cumulative cultural evolution?, Proceedings of the Royal Society B: Biological Sciences 285 (2018) 20180712. doi:10.1098/rspb.2018.0712.

[3] P. Tinits, O. Sobchuk, Open-ended cumulative cultural evolution of Hollywood film crews, Evolutionary Human Sciences 2 (2020). doi:10.1017/ehs.2020.21.

[4] O. Sobchuk, P. Tinits, Cultural Attraction in Film Evolution: the Case of Anachronies, Journal of Cognition and Culture 20 (2020) 218–237. doi:10.1163/15685373-12340082.
[5] J. Henrich, Demography and Cultural Evolution: How Adaptive Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case, American Antiquity 69 (2004) 197–214. doi:10.2307/4128416.
[6] K. Vaesen, M. Collard, R. Cosgrove, W. Roebroeks, Population size does not explain past changes in cultural complexity, Proceedings of the National Academy of Sciences of the United States of America 113 (2016) E2241–2247. doi:10.1073/pnas.1520288113.
[7] J. M. Stubbersfield, E. G. Flynn, J. J. Tehrani, Cognitive evolution and the transmission of popular narratives: A literature review and application to urban legends, Evolutionary Studies in Imaginative Culture 1 (2017) 121–136. doi:10.26613/esic.1.1.20.
[8] P. Boyer, Cognitive Tracks of Cultural Inheritance: How Evolved Intuitive Ontology Governs Cultural Transmission, American Anthropologist 100 (1999) 876–889.
[9] O. Morin, How Traditions Live and Die, Oxford University Press, London & New York, 2015.
[10] A. Acerbi, A. Mesoudi, If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution, Biology & Philosophy 30 (2015) 481–503. doi:10.1007/s10539-015-9490-2.
[11] A. Acerbi, Cultural Evolution in the Digital Age, Oxford University Press, Oxford, New York, 2019.
[12] C. Aragon, K. Davis, C. Fiesler, Writers in the Secret Garden: Fanfiction, Youth, and New Forms of Mentoring, MIT Press, Cambridge, 2019.
[13] B. McHale, Postmodernist Fiction, Methuen, London, 1987.
[14] J. Porter, Popularity/Prestige, Technical Report 17, 2018. URL: https://litlab.stanford.edu/LiteraryLabPamphlet17.pdf.
[15] F. Moretti, The Slaughterhouse of Literature, Modern Language Quarterly 61 (2000) 207–228. doi:10.1215/00267929-61-1-207.
[16] F. Pianzola, fedormyskin/Linked-Potter, 2020. URL: https://github.com/fedormyskin/Linked-Potter.
[17] J. Balraj, C. Farook, Enhance Rating Algorithm for Restaurants, in: K. Arai, R. Bhatia (Eds.), Advances in Information and Communication, Springer, Cham, 2020, pp. 224–234. doi:10.1007/978-3-030-12385-7_18.
[18] ebc, How to Rank (Restaurants), 2015. URL: http://www.ebc.cat/2015/01/05/how-to-rank-restaurants/.
[19] A. G. Stacey, Robust parameterisation of ages of references in published research, Journal of Informetrics 14 (2020). doi:10.1016/j.joi.2020.101048.
[20] S. Guy, Fanfiction.net: Fandoms over time, 2015. URL: https://toastystats.tumblr.com/post/111930409603/fanfictionnet-fandoms-over-time-toasty-says.