# User-Generated World Literatures: a Comparison between Two Social Networks of Readers

Marco Antonio Stranisci<sup>1,\*,†</sup>, Viviana Patti<sup>1</sup> and Rossana Damiano<sup>1,†</sup>

<sup>1</sup>Dipartimento di Informatica, Università di Torino, Italy

#### Abstract

Widely employed digital archives of literary works are not necessarily neutral sources of knowledge, but rather literature ecosystems where the relevance of authors and works are shaped by communities of users. This could affect the visibility of authors belonging to historically marginalized groups of people, as it has been demonstrated by previous works on the under-representation of non-Western writers. In this work, we present an exploratory analysis on how divergent representations of works written by African and Asian writers emerge from different platforms. More specifically, we compare the reviews gathered from an Italian and an international social networks of readers, in order to highlight how authors from these continents are perceived by two different communities of users.

#### Keywords

Knowledge Graphs, Fairness, non-Western canon

## 1. Introduction

The Under-Represented Writers Knowledge Graph (URW-KG) [1] is a knowledge base of writers classified on the basis of their ethnic origin. We created such resource for exploring and mitigating the underrepresentation of non-Western writers on Wikipedia, where old forms of discrimination still exist [2, 3] despite the great opportunities that semantic web technologies unfolded for readers and writers [4].

The analysis of the data contained in the URW-KG shows that underrepresentation is not an homogeneous phenomenon, but varies depending on the archives: Wikipedia provides a more Western-centric snapshot of the World Literature [5] than other sources like Open Library and Goodeads [6]. Such differences may be explained by the different communities of users that produce contents on these platform.

In this paper, we analyze how these communities give shape to different representation of a sample of 40 non-Western writers born in Africa and Asia. To do so, we compared reviews published on Anobii, an Italian-based social network of readers, and LibraryThing, an international platform. The result of the analysis shows that, while the number of reviews

<sup>19</sup>th IRCDL (The Conference on Information and Research science Connecting to Digital and Library science), February 23–24, 2023, Bari, Italy

<sup>\*</sup>Corresponding author.

<sup>&</sup>lt;sup>†</sup>These authors contributed equally.

<sup>🛆</sup> marcoantonio.stranisci@unito.it (M. A. Stranisci); viviana.patti@unito.it (V. Patti); rossana.damiano@unito.it (R. Damiano)

<sup>© 0 2023</sup> Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

and their expressed sentiment about authors vary between the two platforms, a common Western-centric attitude of readers emerges from these communities.

The paper is structured as follow: in Section 2 the URW-KG is briefly described. Section 3 summarizes the data gathering process from Wikidata and illustrates a mitigation strategy of underrepresentation. In Section 4, the analysis of Anobii and LibraryThings communities of readers is presented. In Section 5 conclusion and future works are presented.

# 2. The Under-Represented Writers Knowledge Graph

Several digital resources providing information about literary works and authors are available online, created with different scopes and purposes. Some of them are general-purpose archives that host user-generated contents, while some others are the outcome of monographic researches curated by Digital Humanities scholars. Example of the former type are websites like Goodreads, owned by Amazon, LibraryThing, owned by AbeBooks, and Anobii, owned by the Italian publisher Mondadori, namely, social networks where users share their reads and review books. A project of the Internet Archive Foundation, Open Library is an online service for book loan where users can have an active role by adding and editing records in the archive and reviewing books. Wikidata is a Knowledge Graph linked to Wikipedia, which has a section devoted to literary works. Monographic platforms focus on specific literary genres across countries and ages. It is the case of The European Literary Text Collection<sup>1</sup> [7], a multi-lingual dataset of novels written from 1848 to 1920, DraCor<sup>2</sup> [8], a collection of plays in multiple languages, and MiMoText<sup>3</sup>, a corpus of French and German novels published from 1750 to 1799.

The Under-Represented Writers Knowledge Graph (URW-KG)<sup>4</sup>, a dataset of writers and their works [1, 9], has been created with the specific goal of studying and reducing underrepresentation. Defining underrepresentation is a hard task, as it may bring with it arbitrary distinctions and the use of inadequate terms. To provide an objective and clear definition, we adopted the term 'Transnational', which refers to people who "operated outside their own nation's boundaries, or negotiated with them" [10]. This definition focuses on the relation between a person and their country of birth, based on two axes: (i) being born in a former colony country, which follows the post-colonial theories [11]; (ii) belonging to an ethnic minorities in a Western country.<sup>5</sup>.

Our resource is built upon the Under-Represented (UR) Ontology Network, which organizes knowledge about authors and their works and aligns information gathered from different sources. The Ontology Network is composed of two main modules: the Under-Represented Writers Ontology (URW-O) [1]<sup>6</sup> and the Under-Represented Books Ontology (URB-O)<sup>7</sup>. The URW Ontology (URW-O), which represents biographical knowledge about writers, gathers biographical information into a DUL:SITUATION patterns [12], which is the setting for the author,

<sup>&</sup>lt;sup>1</sup>https://www.distant-reading.net/eltec/

<sup>&</sup>lt;sup>2</sup>https://dracor.org/

<sup>&</sup>lt;sup>3</sup>https://mimotext.github.io/

<sup>&</sup>lt;sup>4</sup>https://underrepresented.di.unito.it/

<sup>&</sup>lt;sup>5</sup>A more detailed description is present in [1]

<sup>&</sup>lt;sup>6</sup>purl.archive.org/urwriters <sup>7</sup>purl.archive.org/urbooks

their role and the other spatial and temporal information (eg: dul:TimeInterval, dul:Place) [9]. The URB Ontology (URB-O) inherits the representation of works and editions from FRBR [13] and FaBiO [14], designed for guiding the cataloguing of works according to Semantic Web principles. The PROV Ontology [15] is used to model the attribution of a book to their author (*prov:wasAttributedTo*), and to make explicit the source of knowledge from which an information is sourced (*prov:wasDerivedFrom*).

# 3. A Quantitative Assessment of Underrepresentation

Building the URW-KG was a two-step process. We first gathered from Wikidata all the entities of type Person (wd:Q5) with occupation (wdt:P106) writer (wd:Q36180), novelist (wd:Q6625963), or poet (wd:Q49757). Consistently with the definition of underrepresentation provided above, we filtered out all the people born before 1808, which marks the beginning of the Spanish-American conflict – considered as the first process of decolonization. After this first step, we obtained a KG of 194, 065 authors and 145, 375 works, that allowed us to identify a significant underrepresentation of Transnational Writers. In fact, they are only 17, 368 on Wikidata, the 9%, and the number of their works is even more represented, since they are only the 8, 380 (5.7%).

In order to understand if such underrepresentation is specific to Wikidata or if is shared by other archives, we integrated into the knowledge graph three external sources of knowledge specialized in archiving books: OpenLibrary<sup>8</sup>, Goodreads<sup>9</sup>, and Google Books<sup>10</sup>. For this integration, we considered the writers obtained from Wikidata as a constant, and gathered all their works from Open Library and Goodreads: by doing so, we could assess the difference in the number of works on Wikidata and on the external resources.

The first part of the data integration strategy consisted in collecting writers' *ids* from Open-Library and Goodreads. We first considered all the 55, 834 authors with an OpenLibrary id, retrieving all their works and editions; we thus obtained 891, 037 works and 1, 742, 252 editions. Results include a set of useful information, including language, publisher, synopsis, subjects, year of publishing, place of publishing, and ISBN. Then, in order to bypass the limitations posed by the lack of APIs from Goodreads, we scraped from Goodreads all the 9, 963 writers who do not have an OpenLibrary id (instead of considering all authors with a Goodreads identifier). Finally, using the Google Books API, we retrieved from Google Books additional information about the collected works through their ISBN. By doing so, 1, 177, 801 new items were obtained and mapped onto OpenLibrary and Goodreads.

The impact of the data integration strategy is presented in in Table 1. As it can be observed, the mapping resulted in a significant and general increase of information about writers literary production. The number of works rose from 145, 375 to 1, 371, 326 (8 times more), and information about them grew accordingly: there are 20 times more blurbs and 34 more subjects. The mapping did not only generically increased the number of works and all the correlated information, but also contributed in rebalancing the proportion between Western and Transnational

<sup>&</sup>lt;sup>8</sup>https://openlibrary.org/

<sup>%</sup>https://www.goodreads.com/

<sup>&</sup>lt;sup>10</sup>https://books.google.com/

## Table 1

Entity type	Wikidata	External	Total
Work	145,375	1,225,951	1,371,326
Edition	0	2,920,053	2,920,053
Subject	5,693	186,960	192,653
Publisher	2,609	234,281	236,890
Blurb	40,532	787, 125	827,657
		1 .	1

Number of works, editions, subjects, publishers, and blurbs on Wikidata and on external resources

works. The former, which are only the 5.7% on Wikidata, become the 10.1% in the KG after the integration.

# 4. Comparing Social Networks of Readers

The mappings between Wikidata and external resources showed that the role of communities of contributors can be crucial in reducing (or spreading) under-representation: platforms based on different groups of users result in different representations of World Literature. In this section, we deepen this intuition by providing a comparative analysis between two social networks of readers, in order to understand how and to which extent readers from different digital platforms discuss about Transnational writers.

## 4.1. Data Gathering and Processing

For our investigation, we chose two platforms where users can upload, review, and discover books: LibraryThing<sup>11</sup>, which is composed of English readers, and Anobii<sup>12</sup>, a website owned by the Italian publisher Mondadori and thus having a prominent audience of Italian people. Such a choice paves the way not only for a generic comparison between two website, but may also provide insights about the different attitudes about books and writers from a local and an international community of people.

Given the high number of writers in our KG, the absence of APIs for gathering data from Anobii and LibraryThing, and the lack of a structural mapping of authors with these archives, we performed an analysis of a subset of writers. We selected a sample of the Transnational writers who have the highest number of editions on Open Library (the topmost 20 born in Africa, and the topmost 20 born in Asia), and manually found their personal pages on the two platforms. We then gathered all their works and readers' reviews. The number of reviews from LibraryThing is 16, 758, while reviews gathered from Anobii are 12, 141. For each review we computed the Sentiment score with existing tools that reached the State of the Art for Italian [16] and English [17] Sentiment Analysis. Both tools give as output a label (positive or negative) and a score in the range 0 - 1. Scores of all reviews labeled as 'negative' were converted to

<sup>11</sup> https://www.librarything.com/

<sup>&</sup>lt;sup>12</sup>https://www.anobii.com/

negative values. Finally, we computed the average sentiment of reviews for each author. Table 2 shows results about African writers, while Table 3 shows result for Asian authors.

## 4.2. Quantitative Analysis of the Reviews

First of all, a different distribution of writers by continent can be observed in the two communities. Reviews about African authors are the 70.5% on LibraryThing, but only the 53.6% on Anobii. Differences between the number of reviews on LibraryThing and on Anobii may be also be observed at a fine grained level. Figure 1 shows the 10 authors with the highest delta between the number of reviews on LibraryThing and Anobii. Such differences may be a signal of the impact of certain authors in a local market like Italy (e.g., Tahar Ben Jelloun), compared to international trends. In this sense, widening the scope of the analysis by augmenting the number of digital archives from different countries may brings light on which representations of "literatures" emerge from readers at global and local contexts (see [18] for a similar approach to Wikipedia).

Secondly, it is worth mentioning that readers from LibraryThing and Anobii seem to show more interest for authors of Western origin, even when they are not born in Western countries. As it can be observed from the data in Tables 2 and 3, 15 out of 20 African-born authors are children of European or American parents, or belong to white minorities. The number of reviews increases this gap, since the 91.4% of reviews are about European or American descents. Proportion are inverted for Asian-born authors: only the 25% of them have European origins (however, the 42% of reviews are about them). This fact highlights the vulnerabilities of the criteria upon which the distinction between Transnational and Western writers is drawn (see Section 2), which should be integrated with other features.

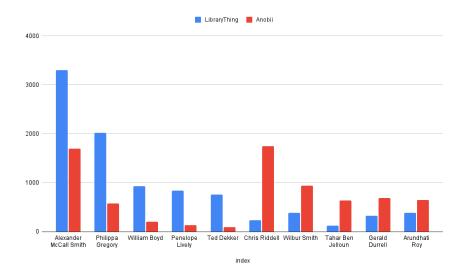


Figure 1: The 10 authors with a highest delta between Anobii and LibraryThing for number of reviews.

## Table 2

The 20 Transnational writers born in Africa with a higher number of editions in Open Library, the number of reviews received by their works on LibraryThing (LT) and Anobii, and the average sentiment expressed by reviews. Sentiment was computed only for authors with 10 or more reviews.

Writer	Country	<b>OL</b> Editions	LT Revs	LT SA	Anobii Revs	Anobii SA
Alexander McCall Smith	Zimbawe	758	3,368	0.514	1,695	0.404
Wilbur Smith	Zambia	674	393	0.379	941	0.379
Peter Dickinson	Zambia	396	280	0.317	26	0.282
Jacques Derrida	Algeria	392	78	0.181	100	-0.184
J. M. Coetzee	South Africa	357	1,093	0.120	1,003	-0.375
Seamus Heaney	Senegal	325	538	0.586	88	0.354
Chris Riddell	South Africa	323	232	0.642	1,749	0.497
Penelope Lively	Egypt	304	860	0.480	131	0.123
Chinua Achebe	Nigeria	300	528	0.401	70	-0.169
Nadine Gordimer	South Africa	292	213	0.320	202	-0.100
Wole Soyinka	Nigeria	251	87	0.244	26	0.600
Toyin Falola	Nigeria	237	4	-	0	-
Philippa Gregory	Kenia	229	2,067	0.312	576	0.286
Alain Badiou	Morocco	206	35	0.06	37	-0.184
Nawal El Sadaawi	Egypt	202	93	0.278	0	-
<b>Richard Dawkins</b>	Kenya	189	1,040	0.281	605	0.105
Simon Scarrow	Nigeria	189	133	0.681	177	0.179
Tahar Ben Jelloun	Egypt	187	640	0.494	119	-0.194
Andrée Chedid	Egypt	178	20	0.386	2	-
William Boyd	Ghana	178	948	0.364	199	0.186

### 4.3. Sentiment Analysis of the Reviews

The two tools adopted for detecting the sentiment of reviews [19, 16] are both based on a BERT-based Language Model<sup>13</sup>. However, since they are trained on different languages, they may provide as output slightly different scores. As a matter of fact, a brief look at columns 4 and 6 of Tables 2 and 3 shows that the tool for Italian language is more skewed over negative and lower values. For this reason, a general comparison of aggregated scores across platforms is not useful.

The Sentiment Analysis of reviews of books of African writers (Table 2) seems to be consistent with differences emerged from the quantitative analysis. Except from J. M. Coetzee, who obtained the most negative sentiment score both on Anobii and on LibraryThing, readers opinions significantly differ. Simon Scarrow, namely the first ranked author for sentiment with a score of 0.681 on LibraryThing, is only ranked 8th on Anobii. On the contrary, Wole Soynka obtained a score of 0.600 on Anobii while on LibraryThing he is only 16th. Other significant differences regard Chinua Achebe, Nadine Gordimer, and Tahar Ben Jelloun, who respectively obtain a score of 0.401, 0.304, and 0.494 on LibraryThing against the following negative values on Anobii: -0.169, -0.100, and -0.194.

The trend of readers' sentiment about Asian writers (Table 3) is not as simple to interpret as the previous one. This is probably due to the lowest number of reviews per author who may have had an impact on the average scores. Many authors receive comparable scores. Eg:

<sup>&</sup>lt;sup>13</sup>Bidirectional Encoder Representations from Transformers: https://github.com/google-research/bert

## Table 3

The 20 Transnational writers born in Asia with a higher number of editions in Open Library, the number of reviews received by their works on LibraryThing (LT) and Anobii, and the average sentiment expressed by reviews. Sentiment was computed only for authors with 10 or more reviews.

Writer	Country	<b>OL</b> Editions	LT Revs	LT SA	Anobii Revs	Anobii SA
Rajneesh	India	1,278	482	0.637	100	0.563
Anita Ganeri	India	1,154	123	0.851	12	0.666
Deepak Chopra	India	537	860	0.461	131	0.420
Stephanie Laurens	Sri Lanka	530	930	0.320	731	0.381
Amrita Pritam	Pakistan	428	0	-	0	-
Gerald Durrell	India	423	328	0.736	685	0.570
Salman Rushdie	India	398	997	0.288	663	0.079
Michael Ondaatje	Sri Lanka	317	695	0.321	175	0.335
Spike Milligan	India	282	104	0.456	9	-
Gulzar	Pakistan	266	6	-	0	-
Sri Chinmoy	Bangladesh	197	19	0.949	0	-
Ted Dekker	Indonesia	186	768	0.303	95	0.363
Idries Shah	India	162	145	0.456	5	-
Amin Maalouf	Lebanon	162	121	0.702	219	0.186
Zulfikar Ghose	Pakistan	158	4	-	0	-
Ruskin Bond	India	156	41	0.641	1	-
Solomon Volkov	Tajikistan	154	12	0.008	1	-
Vandana Shiva	India	154	20	0.437	51	-0.632
Pramoedya Ananta Toer	Indonesia	153	50	0.607	2	-
Arundhati Roy	India	145	389	0.368	648	0.368

Rajneesh (0.637 vs 0.563), Deepak Chopra (0.461 vs 0.420), and Stephanie Laurens (0.320 vs 0.381). Anita Ganeri, namely the author who receives the highest sentiment on Anobii (0.666) is also the 2nd ranked on LibraryThing, while Sri Chinmoy (0.949) and Solomon Volkov (0.008) are respectively the author with the best and the worst scores on LibraryThing, but these results are not comparable with Anobii, since an adequate number of reviews from this platform is missing. In general, readers' sentiment seem less controversial between the two platforms for Asian authors: the only two examples of high divergence are Amin Maalouf (0.702 on LibraryThing; 0.186 on Anobii) and Vandana Shiva, who obtained the lowest sentiment on Anobii with -0.632 while on the LibraryThing it scored 0.437.

# 5. Conclusion and Future Work

In this paper, we presented a quantitative analysis of readers' reviews about Asian and African writers on Anobii and LibraryThing. The analysis shows that differences emerge in the reception of authors on the two platforms, both quantitatively and in the term of the sentiment expressed by reviews. Such differences may be interpreted as country-specific, since Anobii is mainly frequented by Italian readers while LibraryThing is more international. However, a common aspect between the two communities of readers is that they both give more attention to Africanborn and Asian-born writers with Western origins. Future work will be devoted to widening the analysis with a higher number of platforms and writers, in order to provide tools and resources for exploring country-specific receptions of literary works.

# References

- [1] M. A. Stranisci, V. Patti, R. Damiano, Representing the under-represented: A dataset of post-colonial, and migrant writers, in: 3rd Conference on Language, Data and Knowledge, LDK 2021, volume 93, Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, 2021, pp. 1–14.
- [2] J. Adams, H. Brückner, C. Naslund, Who counts as a notable sociologist on wikipedia? gender, race, and the "professor test", Socius 5 (2019) 2378023118823946.
- [3] J. Sun, N. Peng, Men are elected, women are married: Events gender bias on wikipedia, arXiv preprint arXiv:2106.01601 (2021).
- [4] L. Nakamura, "words with friends": socially networked reading on goodreads, Pmla 128 (2013) 238–243.
- [5] I. Almond, World Literature Decentered: Beyond the "West" through Turkey, Mexico and Bengal, Routledge, 2021.
- [6] M. A. Stranisci, G. Spillo, C. Musto, V. Patti, R. Damiano, The urw-kg: a resource for tackling the underrepresentation of non-western writers, arXiv preprint arXiv:2212.13104 (2022).
- [7] C. Schöch, M. Eder, C. Odebrecht, M. Kestemont, A. Primorac, J. Tonra, K. M. Poniž, C. Kanellopoulou, Distant reading for european literary history. a cost action, Proceedings of DH2018 (2018).
- [8] F. Fischer, I. Börner, M. Göbel, A. Hechtl, C. Kittel, C. Milling, P. Trilcke, Programmable corpora: Introducing dracor, an infrastructure for the research on european drama, Digital Humanities 2019 (2019) 5.
- [9] M. A. Stranisci, V. Basile, R. Damiano, V. Patti, Mapping biographical events to odps through lexico-semantic patterns?, in: 12th Workshop on Ontology Design and Patterns, WOP 2021, volume 3011, CEUR-WS, 2021, pp. 1–12.
- [10] B. Boter, M. Rensen, G. Scott-Smith, Unhinging the National Framework: Perspectives on Transnational Life Writing, Sidestone Press, 2020.
- [11] G. C. Spivak, Can the subaltern speak?, in: Colonial discourse and post-colonial theory, Routledge, 2015, pp. 66–111.
- [12] A. Gangemi, V. Presutti, Ontology design patterns, in: Handbook on ontologies, Springer, 2009, pp. 221–243.
- [13] B. Tillett, What is frbr? a conceptual model for the bibliographic universe, The Australian Library Journal 54 (2005) 24–30.
- [14] S. Peroni, D. Shotton, Fabio and cito: ontologies for describing bibliographic resources and citations, Journal of Web Semantics 17 (2012) 33–43.
- [15] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhao, Prov-o: The prov ontology, W3C Recommendation, World Wide Web Consortium, United States, 2013.
- [16] F. Bianchi, D. Nozza, D. Hovy, Feel-it: emotion and sentiment classification for the italian language, in: The 16th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2021.
- [17] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, R. Vollgraf, FLAIR: An easy-to-use framework for state-of-the-art NLP, in: Proceedings of the 2019 Conference of the North

American Chapter of the Association for Computational Linguistics (Demonstrations), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 54–59. URL: https://aclanthology.org/N19-4010. doi:10.18653/v1/N19-4010.

- [18] C. Hube, F. Fischer, R. Jäschke, G. Lauer, M. R. Thomsen, World literature according to wikipedia: Introduction to a dbpedia-based framework, arXiv preprint arXiv:1701.00991 (2017).
- [19] S. Schweter, A. Akbik, FLERT: Document-level features for named entity recognition, 2020. arXiv:2011.06993.