



**HAL**  
open science

## On the Popularity of Classical Music Composers on Community-Driven Platforms

Ioannis Petros Samiotis, Andrea Mauri, Chirstoph Lofi, Alessandro Bozzon

► **To cite this version:**

Ioannis Petros Samiotis, Andrea Mauri, Chirstoph Lofi, Alessandro Bozzon. On the Popularity of Classical Music Composers on Community-Driven Platforms. International Conference on Web Engineering (ICWE 2023), Jun 2023, Alicante, Spain. pp.327-335, 10.1007/978-3-031-34444-2\_24 . hal-04157062

**HAL Id: hal-04157062**

**<https://hal.science/hal-04157062v1>**

Submitted on 10 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the Popularity of Classical Music Composers on Community-Driven Platforms

Ioannis Petros Samiotis<sup>1</sup>[0000-0001-6518-0688], Andrea  
Mauri<sup>2</sup>[0000-0002-1263-4575], Christoph Lofi<sup>1</sup>[0000-0001-5641-5510], and  
Alessandro Bozzon<sup>1</sup>[0000-0002-3300-2913]

<sup>1</sup> TU Delft, Delft, The Netherlands

[i.p.samiotis@tudelft.nl](mailto:i.p.samiotis@tudelft.nl), [c.lofi@tudelft.nl](mailto:c.lofi@tudelft.nl), [a.bozzon@tudelft.nl](mailto:a.bozzon@tudelft.nl)

<sup>2</sup> Université Claude Bernard Lyon 1, Lyon, France

[andrea.mauri@univ-lyon1.fr](mailto:andrea.mauri@univ-lyon1.fr)

**Abstract.** Traditionally, the popularity of classical music composers is approximated through commercial figures like album releases, record sales, or live performances. However, commercial factors only provide one piece of the overall picture. The success of community-driven platforms has profoundly changed how people consume and interact with music, and, consequently, our understanding of what popularity is. People discuss their favourite artists, archive knowledge regarding them and share their work through multimedia platforms. In this paper, we investigate how data from these platforms can provide a more comprehensive view on popularity and engagement regarding the long-tail of classical music composers. We combine album release data provided by MusicBrainz, the commitment of people in maintaining the composers' Wikipedia pages and user engagement in classical music videos on YouTube. Our analysis provides a complementary multi-faceted view on community engagement and urges future research to expand on user-generated content for a more diverse expression of popularity in the music domain.

**Keywords:** social media · user-generated content · music · popularity · web crawling

## 1 Introduction

Popularity is a desired achievement for artists, as it promotes their work, facilitates the interaction with the broader audience, and ultimately affects how people will remember them in years to come. This holds especially true in classical music, which includes centuries-long catalog of works created and reinterpreted by a long list of artists. Access to those works and their artists is essential in preserving parts of historic and cultural heritage, as classical music has fundamentally influenced western music throughout history.

Recording labels have pioneered the methods to capture audio performances, preserving a plethora of classical music pieces of the old or contemporary composers. Inevitably though, these recordings are skewed towards already established and well studied composers and their works can be found performed by various artists.

However, popularity is not only related to talent, but also former success [1] and “*the need of consumers to consume the same art as others*” [2]. When people are exposed to art, share it with others, or consume media about it, they create what is called “consumption capita” [3]. These activities have become much easier to perform with the widespread adoption of Web technologies and the availability of community-driven platforms. People from around the world with different cultural backgrounds can openly share their knowledge, experiences, and recordings with others in the ever expanding publicly-accessible knowledge bases and social media communities.

For this reason we believe that by tapping into user-generated content, we can find interesting insights on user-engagement and popularity of classical composers online. To that end, we analyze user-generated data on Wikipedia and YouTube compared to album releases retrieved from MusicBrainz<sup>3</sup>, to investigate our hypothesis that *User-engagement on community-driven platforms develops composer popularity differently than the album release trends in classical music*. Our findings can be used to indicate to what extent those platforms potentially hold content and information about composers who have zero to low number of official album releases.

## 2 Related Works

Research on the multilingual online encyclopedia Wikipedia<sup>4</sup>, has shown that it is possible to predict real-world opinions and popularity through analyzing user interaction on the platform. For example, Mestyán et al [4] presented a model using Wikipedia’s user-generated content that accurately predicts movie box office success, while another study by Wei et al [5] uses user interaction on Wikipedia to predict stock market values.

YouTube<sup>5</sup> is a video-sharing, social media platform, which encourages a user-content-user interaction [6] and hosts a staggering amount of videos across a wide variety of categories. While research work has mostly focused on network dynamics [7–9] and opinion mining [10–12], there have been studies on popularity on the platform. Chatzopoulou et al. [13] studied the correlation between views, number of comments and ratings and popularity, while in [14] it is found that popularity on YouTube follows the conclusions of studies on “consumption capita” [15]. Related to our study, the work of Cayari [16] studies how YouTube has fundamentally affected musical art forms and has essentially changed the way people listen, share and consume music. This holds especially true for classical music, where with a quick search on the platform, it is possible to find several results with multiple compilations of classical music works, live footage of concerts, interpretations of compositions, and educational material.

Our study is inspired by the work of Bellogin et al [17], the first comparing music artist popularity from different Web and music services. The main difference with our work is that we look into user-engagement regarding music

<sup>3</sup> <https://musicbrainz.org/>

<sup>4</sup> <https://en.wikipedia.org/wiki/Wikipedia>

<sup>5</sup> <https://www.youtube.com/>

on generic platforms that host music information among other categories. In our study, we also focus on genre-specific content on those generic platforms, rather than overall music consumption trends. Related to classical music, Schedl et al. [18] studied the online engagement of fans of classical music on Twitter and Last.fm finding that classical music is under-represented on social media, as classical music enthusiasts seem averse sharing content on both of the studied platforms. In this study we argue that we need further analysis on community-driven platforms to be able to assess the extent those platforms contain domain specific information.

### 3 Data Collection

We started with quantifying user engagement on Wikipedia and YouTube. We defined a list of classical music composers which we are considering for analysis in different platforms and selected a reliable source of knowledge about album releases. In this section, we outline our data collection process. The data retrieval and scripts, as well as the resulting dataset are available here<sup>6</sup>.

**DBpedia:** to select the composers, we retrieved a list of their names from various classical music periods from DBpedia<sup>7</sup> using the Virtuoso SPARQL Query Interface<sup>8</sup>. We used the name contained in the URI as an unambiguous representation of a composers' name, as there could be multiple entries with the same name, but different people. We queried the composer names, using the following Yago entities available for classical music periods in DBpedia: WikicatBaroqueComposers, WikicatClassicalComposers, WikicatClassical-periodComposers, WikicatRomanticComposers, Wikicat18th-centuryClassicalComposer, Wikicat19th-centuryClassicalComposers, Wikicat20th-centuryClassicalComposers, Wikicat-21stcenturyClassicalComposers. We finally collected 5928 different composers, distributed as follows: 1126 baroque, 1025 romantic, 725 classical, 96 from eighteenth century, 345 from nineteenth century, 3155 from twenty century and 1383 from twenty one century. Obviously, a composer can belong to multiple periods.

**Wikipedia:** we used Wikipedia APIs<sup>9</sup> to retrieve each composer's page and related information. To ensure we retrieve the page related to the composer, we use as query the name as it appears in the "About" on DBpedia. For instance, to retrieve the page of *Alexander Müller* we use *Alexander\_Müller\_(composer)*. This also holds true for composers having the same name, like *Johann Strauss I* and his son *Johann Strauss II*. We follow the same procedure for a composer's full name on MusicBrainz and YouTube. More specifically, we retrieved: 1) the number of edits a page has received, 2) the number of users who edited the page, 3) the number of languages the page is translated to, 4) the size of the page (in KB), and 5) how many sections it contains. We use these data as a proxy to measure how many people engage on Wikipedia with a composer's entry. They show us the amount of users maintaining a page, how much are they committed

<sup>6</sup> [https://github.com/ipsamiotis/classical\\_popularity](https://github.com/ipsamiotis/classical_popularity)

<sup>7</sup> <https://wiki.dbpedia.org/>

<sup>8</sup> <https://dbpedia.org/sparql>

<sup>9</sup> [https://www.mediawiki.org/wiki/API:Main\\_page](https://www.mediawiki.org/wiki/API:Main_page)

in keeping the information correct and updated (number of edits), accessible to many people (languages) and complete (page length and number of sections).

**MusicBrainz:** Musicbrainz is a community-maintained music encyclopedia that collects metadata about music artists and their released works (albums, recordings and more). We used their up-to-date and proven reliable information on *Album Releases* per classical music composer. We retrieved this using a Python wrapper to its API<sup>10</sup> and in our analysis we refer to them as *Album Releases*.

**YouTube:** we gathered videos from YouTube using youtube-dl<sup>11</sup> based on the queries: “composer name” + “music” and “composer name” + “live music”, resulting in a total of 184,019 video entries. The retrieved videos contained a high amount of non-relevant entries. We cleaned the dataset by discarding all videos where the composer name was not present in either the title, in one of the tags or in the description of the video. This way, we decreased the number of videos relevant to our study to 69,261. Since users also view videos and react to them by leaving a like, dislike or a comment, we also retrieved the number of: 1) likes, 2) dislikes, 3) views, 4) comments, 5) unique uploaders and 6) duration of videos. From these engagement data, we further calculated the average number of likes, dislikes, views, video duration and comments per composer.

In the final dataset we collected, we find 97% of the composers have a Wikipedia page, 87% of the composers have at least one video on Youtube, while only 70% of them have albums on MusicBrainz. Also, 63% of composers are in common to all the platforms, with Wikipedia and YouTube having the biggest overlap (99%), while MusicBrainz lacks 31% of composers present in Wikipedia.

## 4 Data Analysis

We first analyse if the data we use as a proxy for user engagement on Wikipedia and YouTube is correlated. To that end, we calculated the Spearman correlation [19] of the popularity ranking of composers each data property was generating. This will give some insights into how comparable user engagement is between platforms. To investigate if user-engagement-driven popularity exhibit differences compared to the album releases, we calculate the similarity between those trends. Using the Jaccard similarity index [20], we are able to observe differences between the different popularity rankings and album releases, finding interesting results on both top-to-bottom and bottom-to-top rankings. The intuition is that by computing the Jaccard index at different level of cut-off it is possible to see how much each platform agrees on who are the top  $n$  (or bottom  $n$ ) composers.

### 4.1 Calculating Popularity per Platform

We calculate a single popularity ranking per platform based on user activity metrics as described in Section 3. These metrics are related to: 1) how much

<sup>10</sup> <https://python-musicbrainzngs.readthedocs.io/en/v0.7.1/>

<sup>11</sup> [https://pypi.org/project/youtube\\_dl/](https://pypi.org/project/youtube_dl/)

the people engage with the platform in the context of classical music and 2) what kind of data they create. Their connection to popularity is based on the fact that people who are interested on a composer and are active users of those platforms, they will either create data about them, interact with data others created and/or discuss about them with other users. All these factors are an “online version” of those described on “consumption capita”, which indicates the reasons why artists become famous among people, apart from their talent [3]. We first calculate the correlation of those metrics, following the methodology of [13], to find if the engagement metrics we chose are correlated with each other, to be later considered part of a platform-wide “popularity ranking”. This popularity ranking, as discussed before, is being used as a proxy of real popularity of composers on those platforms.

**Wikipedia:** we compute the correlation of the the metrics described in Section 3 using the Spearman coefficient. As shown in Table 1, most of the different metrics exhibit high correlation (greater than 0.5), with the least correlated being the number of languages with the page size and number of sections metadata.

		Wikipedia				
		Sections	Languages	Revisions	Users	
Pages Size	0.82	0.38	0.72	0.69		
Sections	1	0.38	0.63	0.61		
Languages		1	0.49	0.67		
Revisions			1	0.87		
		Youtube				
		Views	Likes	Dislikes	Duration	Comments
Videos	0.86	0.85	0.81	0.89	0.05	
Views	1	0.98	0.94	0.81	0.07	
Likes		1	0.93	0.81	0.07	
Dislikes			1	0.76	0.06	
Duration				1	0.06	

**Table 1.** Spearman correlation the metrics of Wikipedia and Youtube

**Youtube:** we performed the same analysis on the data gathered from YouTube, finding that they are all highly correlated - as shown in Table 1 - except the number of comments. This could be explained by limitations on the number of comments which could be accessed by the APIs, so we excluded them from the final rankings. The others are consistent with the results reported in [13].

To compare the popularity of composers on Wikipedia, YouTube and album releases, we needed a single ranking per platform. To achieve this, we ranked the composers based on each metric per platform. If a composer is not present on a platform, the values of his or her platform metrics are set to 0 (e.g., for composers not present on MusicBrainz, we set their album releases number to 0). In this way we obtain list of the same length and we can compute correlation coefficients. Then, we aggregated the different rankings into a single one for each

platform using the average ranking aggregation method. We then compared them using the Spearman correlation coefficient.

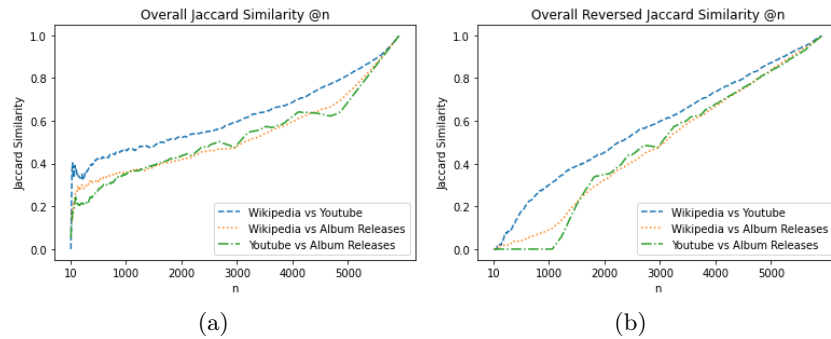
	Wikipedia	MuiscBrainz
YouTube	0.42	0.34
Wikipedia	1	0.36

**Table 2.** Spearman correlation between the platform rankings and album releases.

As Table 2 shows, all platforms are positively correlated, with YouTube and Wikipedia exhibiting a stronger correlation to each other’s popularity rankings.

## 4.2 Popularity on Platforms and Album Releases

We then calculate the similarity between the popularity rankings derived from the community-driven platforms and the rankings obtained by looking at composers’ *Album Releases*. Such comparison can inform us on the degree to which online communities engage in data creation, compared to the official recordings produced by the industry shown in the *Album Releases*. We compute similarity at different ranks, which intuitively can be encoded as “how much different rankings agree” with each other. To that end, we used the Jaccard similarity index. We observe - in Figure 1a - that Wikipedia’s and YouTube’s popularity rankings are more similar to each other, than compared to the *Album Releases*. While this was expected, considering the fact that they are both community-driven platforms (see Table 2), this is still notable as both platforms have strong differences in scope and purpose. Wikipedia’s ranking is closer to those of *Album Releases*, when compared to YouTube. This results actually is more in line with related work in [4,5], that shown user-engagement metrics can predict real-world popularity to a certain extent.



**Fig. 1.** Jaccard similarity at different ranking cut-offs with normal (a) and reverse ranking (b)

We compute the Jaccard similarity also looking at the bottom of the ranking, that is by looking at the long-tail of each platform. Following a reverse ordering on popularity rankings, starting from most “obscure” composers, the

uniqueness of YouTube’s ranking becomes more evident, as we see in Figure 1b. The popularity of the long-tail composers on YouTube is completely dissimilar to the *Album Releases*, as it stays at 0 until we consider 1000 composers. This dissimilarity means that for composers with 0 released albums, the users still engage in data creation on YouTube. This is very encouraging for future works that need to find information regarding classical music composers, as YouTube contains video artefacts and user-engagement metadata for those composers who don’t have registries with official album releases.

In the same way, Wikipedia, noticeably less severely, has low similarity with *Album Releases* as well for the first 1000 composers. This means that even composers with low-count of official album releases, there are still Wikipedia entries with stored information, generated in its entirety by the community of the platform. These findings strengthen our hypothesis that community-driven platforms behave similarly to each other and differently to *Album Releases*, potentially holding user-generated information for composers with low number of official album releases.

## 5 Conclusion

In this work we investigated to what extent popularity of classical music composers on generic community-driven online platforms, follows official album releases as registered in MusicBrainz. We found that Wikipedia and YouTube although they share similarities with each other regarding composers’ popularity rankings, they differ on the long-tail of popularity, based on album releases. We discovered that Wikipedia’s popularity ranking follows more closely the ranking found in album releases, following results from similar studies on the platform. This comes in contrast with the popularity rankings as witnessed on YouTube, which don’t follow as closely those of album releases. For example on YouTube, there are many composers who have no entry for an album release on MusicBrainz but they still garner a notable number of followers who engage with their work. This could reflect the more democratized and diverse manner of information creation on YouTube, compared to the more rule-based and expert driven one on Wikipedia. Due to the similarity of rankings between Wikipedia and YouTube, we find that complementary studies of different platforms could assist to a more holistic overview of published corpora, especially in classical music. Therefore, community-driven online platforms show potential in preserving information regarding composers and their works that are under-represented in the official recorded canon.

**Acknowledgements** This work is partially supported by the European Commission under the TROMPA project (H2020 770376). We thank Cynthia C.S. Liem for her insights on popularity in music and contributions in discussions around the topic.

## References

1. Glenn M MacDonald. The economics of rising stars. *The American Economic Review*, pages 155–166, 1988.



2. Moshe Adler. Stardom and talent. *The American economic review*, 75(1):208–212, 1985.
3. George J Stigler and Gary S Becker. De gustibus non est disputandum. *The american economic review*, 67(2):76–90, 1977.
4. Márton Mestyán, Taha Yasseri, and János Kertész. Early prediction of movie box office success based on wikipedia activity big data. *PloS one*, 8(8), 2013.
5. Pengyu Wei and Ning Wang. Wikipedia and stock return: Wikipedia usage pattern helps to predict the individual stock movement. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 591–594, 2016.
6. Mirjam Wattenhofer, Roger Wattenhofer, and Zack Zhu. The youtube social network. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
7. Stefan Siersdorfer, Sergiu Chelaru, Jose San Pedro, Ismail Sengor Altingovde, and Wolfgang Nejdl. Analyzing and mining comments and comment ratings on the social web. *ACM Transactions on the Web (TWEB)*, 8(3):17, 2014.
8. John Paolillo, Sharad Ghule, and Brian Harper. A network view of social media platform history: Social structure, dynamics and content on youtube. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
9. Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. How useful are your comments?: analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*, pages 891–900. ACM, 2010.
10. Claudia Orellana-Rodriguez, Ernesto Diaz-Aviles, and Wolfgang Nejdl. Mining emotions in short films: user comments or crowdsourcing? In *Proceedings of the 22nd International Conference on World Wide Web*, pages 69–70. ACM, 2013.
11. Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, and Katja Filippova. Multi-lingual opinion mining on youtube. *Information Processing & Management*, 52(1):46–60, 2016.
12. Aliaksei Severyn, Olga Uryupina, Barbara Plank, Alessandro Moschitti, and Katja Filippova. Opinion mining on youtube. 2014.
13. Gloria Chatzopoulou, Cheng Sheng, and Michalis Faloutsos. A first step towards understanding popularity in youtube. In *2010 INFOCOM IEEE Conference on Computer Communications Workshops*, pages 1–6. IEEE, 2010.
14. Oliver Budzinski and Sophia Gaenssle. The economics of social media stars: an empirical investigation of stardom, popularity, and success on youtube. *Ilmenau Economics Discussion Papers*, 21(112), 2018.
15. Moshe Adler. Stardom and talent. *Handbook of the Economics of Art and Culture*, 1:895–906, 2006.
16. Christopher Cayari. The youtube effect: How youtube has provided new ways to consume, create, and share music. *International Journal of Education & the Arts*, 12(6):n6, 2011.
17. Alejandro Bellogín, Arjen P de Vries, and Jiyin He. Artist popularity: Do web and social music services agree? In *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.
18. Markus Schedl and Marko Tkalčič. Genre-based analysis of social media data on music listening behavior: are fans of classical music really averse to social media? In *Proceedings of the First International Workshop on Internet-Scale Multimedia Management*, pages 9–13, 2014.
19. Leann Myers and Maria J Sirois. Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences*, 12, 2004.
20. Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.