

# Evolution of Writings in an Interest-oriented Online Community

February 14, 2017

## 1 Introduction

Many communities in the Internet are formed by a group of people sharing common interests. In particular, the contemporary pop culture sees the rise of a specific kind of online communities - fandoms. Fandoms, by convention, are a group of people that identify, connect and interact with each other based on a similar interest, such as a movie, a book or a band [**wiki:fandom**].

Transformative works, or in a more common term, fan works, are one typical kind of production from fandoms[**wiki:transformative work**]. These are creative works made by fans based on one or more original works, and are often centered around certain characters or story lines. For example, a story written by a contemporary fan about Sherlock Holmes in his retirement is considered a fan fiction in the Sherlock Holmes fandom. Although fan works contain multiple media types such as art, music and games, one of the most common type is creative writing—fanfictions.

This project studies the evolution of fanfictions in fandoms. Data for the study comes from the online transformative work archive site Archive of Our Own (AO3). This site allows free host for works that users upload, and categorizes them based on fandoms. It also utilizes a metadata system to store the works information, allowing for many filtering and classifying operations. Established in 2010, it has become one of the most popular transformative work archives. We collect fanfictions from 34 most popular fandoms on AO3 as of March 2016 <sup>1</sup>, containing ?? (to add) pieces of fan-

---

<sup>1</sup>from this list: <http://archiveofourown.org/media>

fictions. We then filter the fictions by language, keeping only those written in English, which leaves ?? (to add) fictions.

Besides the work texts, the metadata collected includes 23 fields that can be roughly divided into two categories. One is tags generated by the author, which describes the works content; the other is tags that are automatically generated, and describes the works other information, as well as the readers' feedback. Table 1 gives the names of these fields.

Table 1: Metadata of the writings

Content related	Non-content related
Additional Tags, Archive Warnings, Category, Characters, Fandoms, Rating, Relationship, Summary, Text, Title	Author, Bookmarks, Chapter Index, Chapters, Comments, Complete Date, Hits, Kudos, Language, Notes, Publish Date, Update Date, Word Count

We're especially interested in the Kudos field, which shows how many times a fiction has been "liked" by readers. In particular, we assume that authors aim at writing works that receive more Kudos, and therefore clicking Kudos functions as an interaction between authors and readers. So we want to focus on the relation between Kudos and certain features of the fictions.

We use language models to analyze the fictions. We start with the unigram model, turning a text into a vector containing probabilities for each unigram. The probabilities are smooth with Good Turing smoothing, implemented using the Simple Good Turing algorithm [gales1995good]. This method adds some probabilities to the previously unseen unigrams, thus avoid having unigrams of probability 0.

## 2 Results

## 3 Methods

main.bib