DHQ: Digital Humanities Quarterly
2019
Volume 13 Number 3

Manual Annotation of Unsupervised Models: Close and Distant Reading of Politics on Reddit

Christoph Aurnhammer <aurnhammer_at_coli_dot_uni-saarland_dot_de>, Department of Language Sciene and Technology, Saarland University
Iris Cuppen
Inge van de Ven
Menno van Zaanen <menno_dot_vanzaanen_at_nwu_dot_ac_dot_za>, South African Centre for Digital Language Resources, North-West University, Potchefstroom, South Africa

Abstract

This article offers a methodological contribution to manually-assisted topic modeling. With the availability of vast amounts of (online) texts, performing full scale literary analysis using a close reading approach is not practically feasible. The set of alternatives proposed by Franco Moretti (2000) under the umbrella term of “distant reading” aims to show broad patterns that can be found throughout the entire text collection. After a survey of literary-critical practices that combine close and distant reading methods, we use manual annotations of a thread on Reddit, both to evaluate an LDA model, and to provide information that topic modeling lacks. We also make a case for applying these reading techniques that originate in literary reading more broadly to online, non-literary contexts. Given a large collection of posts from a Reddit thread, we compare a manual, close reading analysis against an automatic, computational distant reading approach based on topic modeling using LDA. For each text in the collection, we label the contents, effectively clustering related texts. Next, we evaluate the similarity of the respective outcomes of the two approaches. Our results show that the computational content/topic-based labeling partially overlaps with the manual annotation. However, the close reading approach not only identifies texts with similar content, but also those with similar function. The differences in annotation approaches require rethinking the purpose of computational techniques in reading analysis. Thus, we present a model that could be valuable for scholars who have a small amount of manual annotation that could be used to tune an unsupervised model of a larger dataset.

Introduction

The management of ever-vaster amounts of information that bombard us daily is one of the most important challenges we have been facing since the broad availability of the Internet. Technological developments during this time have drastically changed our abilities to access, process, and transfer information. In particular, the popularity of Web 2.0, which places human interaction and collaboration at its center, has led to massive amounts of available data. Additionally, other types of data are made available, including the contents of individual books that are integrated into bigger and bigger networks of texts, usurped by the continuously expanding structures of online databases. Google, as well as non-profit organizations such as Project Gutenberg, the Million Book Project and the Internet Archive, carry out large-scale projects to scan and upload the contents of whole libraries at a time. Kevin Kelly, co-founder of *Wired* magazine rejoices: “[o]nce books are digital, books seep out of their bindings and weave themselves together. The collective intelligence of a library allows us to see things we can’t see in a single, isolated book” [Kelly 2012]. So far, Google has scanned and made available over 30 million books. It would take a human an estimated twenty thousand years to read such a vast collection at the reasonable pace of two hundred words per minute, without interruptions for food or sleep [Aiden and Michel 2013, 47]. 1

Projects that make available great amounts of data, however, pose what Matthew Wilkens calls a “problem of 2 abundance”:

> We don’t read any faster than we ever did, even as the quantity of text produced grows larger by the year. If we need to read books in order to extract information from them and if we need to have read things in common in order to talk about them, we’re going to spend most of our time dealing with a relatively small set of texts. … each of us reads only a truly minuscule fraction of contemporary fiction (on the order of 0.1 percent, often much less). … we need to decide what to ignore. [Wilkens 2011, 250]

The field of Digital Humanities (DH) aims to offer methodological innovations to solve this problem of abundance. Since 2000, many have followed Franco Moretti's provocative call for distant reading. Moretti deemed close reading “a theological exercise” and urged humanists to “read less”. According to some big data theorists, the act of sampling is “an artifact of a period of information scarcity, a product of the natural constraints on interacting with information in an analog era” [Mayer-Schönberger and Cukier 2013, 16–17]. With rich metadata and computational pattern matching, difficult texts like *The Making of Americans*, are now *not-readable* in important new ways [Don et al. 2007-08] (see also [Kirschenbaum 2007]). 3

Els Stronks has opined we should make the gap between close and distant as wide as possible by further developing distant reading techniques, and at the same time making our close readings more precise and skillful in order to interpret the results gained by computational analysis [Stronks 2013, 207]. Juxtaposition of the two, however, is mostly a polemical issue or rhetorical strategy. According to Jeremy Rosen (2011), Wilkens makes it seem like we have to choose between methods. Distant reading is not a replacement but a supplement or alternative to traditional close reading practices. On the contrary, “[t]he availability of voluminous electronic data, one might conclude, makes it even more necessary to cultivate the faculty of analyzing data closely and critically” [Rosen 2011]. Indeed, in his study (2013) of changes in the geographic imagination of American fiction around the Civil War, Wilkens exclusively makes use of distant reading techniques. Yet, close and distant reading are by no means antithetical [Moretti 2009] [Earhart 2015]. 4

Moretti's argument in “Slaughterhouse of Literature” (2000) for instance, as one of the first applications of his distant reading methods, relies entirely on manual annotation of 108 detective stories, read by human beings in a graduate seminar (if not exactly “close reading,” reading for clues involves at least skimming the text). Many other works that are usually taken as representative of distant reading rely extensively on manual annotation or close readings of illustrative literary passages. So contrary to Moretti’s somewhat provocative proposition of distant reading as alternative to close reading, his work typifies the importance of human reading and tabulation. Yohei Igarashi [Igarashi 2015] has shown how the interrelatedness between close and non-close reading predates the digital humanities, as it occurs at least since educational word lists of the early twentieth century. 5

A brief survey of how different scales of analyses are connected in actual literary-critical practice will show us that indeed, close and distant reading were never mutually exclusive. In DH, this work has mostly been carried out in literary history — see, for instance, a recent study on race, religion, and the US novel [So 2019]. Stephen Ramsay, in his book *Reading Machines* (2011), offers different methods to engage in what he calls “algorithmic criticism,” [Ramsay 2011] criticism derived from algorithmic manipulation of text. Ramsay holds that we should not take close reading to be diametrically opposed to computational, data-driven approaches, since there are important similarities between the two. Both methods are interpretive, in the sense that they transform the original text into something else. “The critic who endeavors to put forth a ‘reading,’ puts forth not the text, but a new text in which the data has been paraphrased, elaborated, selected, truncated, and transduced” [Ramsay 2011, 16]. But the same is true for a computer-driven distant reading where the original text is converted to information by an algorithm. In a recent issue of *PMLA*, moreover, several scholars reflect on ways to nuance Moretti’s statements on “not reading,” effectively reducing the distance between the two textual approaches [Booth 2017] [Drucker 2017] [Piper 2017]. 6

For a good overview of papers combining close and distant reading from 2005-2015, see [Jänicke et al. 2015]. They argue that most papers that combine the two usually follow the “Information Seeking Mantra”: “Overview first, zoom & filter, details on demand” [Shneiderman 1996]. The output of a distant reading is then an overview of the data that 7 highlights potentially interesting patterns for close reading. They call this most common form of connecting scales of analysis “top-down”: first, a distant view on the textual data is shown, and later the details come into view. One can think of the reading of Christina Rossetti in Ted Underwood’s “The Longue Durée of Literary Prestige” (2016). Underwood collected two samples of English-language poetry from 1820–1919: one from volumes reviewed in prominent periodicals, and one of an obscure author, randomly chosen from a large digital library. Looking at 360 reviewed and 360 random volumes, his study assessed the strength of the relationship between poetic language and reception. The approach taken was to first look at broad patterns, and then zoom in and read a few revealing passages. Likewise, in a recent study of novelty in modernist texts [McGrath 2018], scholars first measure intratextual novelty and then use close reading of a sample to test the measurements, scores, and graphs. Another [Lee et al. 2018] analyzes textual scale as a structuring principle of geographical, spatial scale in studying the pre-modern world, by not-reading tens of thousands of Renaissance books.

Stanford Lit Lab pamphlet 4 #heuser2012 traces macroscopic changes in the British novel during the nineteenth century. It signals two interrelated transformations in novelistic language during this period: a systemic concretization of language and a change in the social spaces of the novel. The authors research quantifiable features such as word usage, adopting a “dialogic approach that oscillates between the historical and the semantic, between empirical word frequencies that reveal the historical trends of words and semantic taxonomies that help us identify the meaning and content of those trends” #heuser2012. Through their “hypothesis-testing mode of interpretation” #heuser2012, they make sure their results are semantically and culturally interpretable. This way, they offer a possible answer to what Alan Liu has called the “meaning problem” at the heart of the digital humanities: to determine the relation between “quantitative interpretation and humanly meaningful qualitative interpretation” [Liu 2013, 414]. As a certain measure of sampling when researching literary history is unavoidable, the issue of canon vesus archive is of import here [Algee-Hewitt et al. 2016]. In digital research, a researcher often works according to a process that Jo Guldi [Guldi 2018] calls winnowing: careful culling and fine-tuning of the algorithm to get rid of false positives or messy data, towards cleaner data and clearer results. In that respect, it is in fact not so far removed from more traditional approaches such as close reading the single text. 8

Therefore, we follow this trend in mixing qualitative and quantitative methods and using them to analyze corpora that solicit readings that zoom in and out between part and whole. But rather than following the top-down approach or the “Information Seeking Mantra,” in our mixed method, manual annotation is not of a sample that follows from the overview produced by the distant reading, but a method that comes before the LDA analyses, to evaluate its workings and to reflect on its strengths and weaknesses. 9

We believe that this mostly literary-historical body of work since Moretti, especially concerning the “great unread” [Moretti 2000a, 227] has value when it comes to the current “information overload” that we exemplify here using a case study from web platform Reddit. Here, we propose a transfer from the literary to non-literary informational contexts. What is interesting for our purpose here is not some presumed binary between close and distant reading. It is, rather, the collective recognition, since Moretti, that literary history is not a clearly demarcated, well-mapped, and exhaustible field, but an “uncharted expanse” [Moretti 2000] whose macroscopic shape we cannot fully know. This is true, albeit in a different, non-historical sense, for online web forums. By using a mixed method of manual annotation and LDA, we propose a means to confront this other, contemporary version of the “great unread” [Cohen 1999, 23]. Other studies #jockers2016 rely and build on the readings and value judgments of other professional readers to select their corpus. In “testing” and “correcting” previous gender and genre classifications of their peers, such studies have a strong sense of tradition that studies of a contemporary phenomenon such as Reddit threads obviously lack. 10

As announced, we propose a strategy for using manual annotation to evaluate and supplement an LDA model. Our analysis incorporates local annotation in a distant reading. The investigation requires the development of computational tools (e.g., topic identification, summarization) that deal with large amounts of documents. This should provide large patterns in the dataset that enable a fine-grained analysis of interesting parts of the data. Additionally, in-depth, qualitative inspection of the performance of the computational analyses is essential, in order to evaluate the computational approach. 11

In this article, we investigate whether we can analyze a large document collection from a distant reading perspective, indicating the different semantics of the texts within the entire collection. This analysis is compared against a manual, close reading analysis, to be able to evaluate the performance of the computational approach. If the computational technique behaves similarly to the manual approach, we can effectively use a distant reading technique to complement the close reading analysis. Summarizing, our approach entails a comparative analysis of the two reading methodologies: a close reading with manual annotation on the one hand, and a distant reading with topic modeling on the other. In order to explore both close and distant reading methods, we begin by annotating the data both manually and computationally. This way, we can reflect on and distillate the most valuable properties from both approaches. At the same time we can evaluate the suitability of generic computational techniques to high-quality manual analysis. 12

In order to explore the possible strategies that are located between the extremes of distant and close reading, we focus on a selected discussion thread of the popular online forum Reddit. Applying reading strategies from a literary studies context to this non-literary environment provides us a valuable insight into the merits of these strategies in an age of digital information. Furthermore, the Reddit thread is large enough to identify higher level patterns and manageable enough for manual analysis. Additionally, the posts within the thread, which pose suitable units of information, are expected to have different semantic content. 13

We will first give an overview of the close and distant reading methods employed on a conceptual level, and then outline how we operationalized them in this research. Following this, results of the comparison between the results found using the two methods are presented and discussed in the context of efficiently combining close and distant reading. This shows in how far we can use computational techniques as a pre-processing phase to relieve us from a large part of the labor-intensive work, enabling close reading of the interesting parts of huge data collections, an informed choice based on the results of the computational analysis. Lastly, we summarise the results and propose next steps. 14

Background

Close Reading

Close reading is an umbrella term for an assortment of reading strategies characterized by devout and detailed attention to the meaning and composition of art works. The approach was made famous by the New Critics, a group of Anglo-American literary scholars including Cleanth Brooks, William K. Wimsatt, and Monroe C. Beardsley. Inspired by I.A. Richards (author of *Practical Criticism*, 1929), Matthew Arnold, and T.S. Eliot, these scholars experienced their heyday of academic fame in the forties and fifties of the last century. Going against contemporary practices that, in their view, overvalued historical context and biographical information, the New Critics suggested that literary scholars should investigate the text itself. They wrote extensively on certain contemporary fallacies of literary analysis, for instance, letting your own emotions factor into the interpretation (the “affective fallacy” [Wimsatt and Beardsley 1949]) or writing about authorial intentions (the “intentional fallacy” [Wimsatt and Beardsley 1946]). Another practice they attacked was the paraphrasing of the contents or message of a work (the “heresy of paraphrase” [Brooks 1947]).[1] Instead, this school propagated the careful examination of evidence offered by the text itself: images, symbols, and metaphors as part of a larger structure that gives the text its unity and meaning. Of particular interest to the close reader were devices that create ambiguities, paradoxes, irony, and other forms of tension within the text. Moving from close to distant reading, the level of analysis shifts from details within single texts to categories of many texts. 15

Distant Reading

Distant reading is the practice of aggregating and processing information about, or content in, large bodies of texts without the necessity of a human reader to read these texts [Drucker 2013]. Distant reading corresponds to quantifying, computational reading methods and was introduced by Franco Moretti with the intention to identify the bigger picture in large collections of textual data that close reading cannot uncover. “Reading” is outsourced to a computer: it is in fact a form of data mining that allows information in (e.g., subjects, places, actors) or about (e.g., author, title, date, number of pages) the text to be processed and analyzed. The latter are called metadata: data about the data. Natural language processing can analyze the contents of “practically unreadably” large corpora of texts, while with data mining we can 16 expose patterns or summarize on a scale that is beyond human capacity.

In his book *Distant Reading* (2013) Franco Moretti introduces the term polemically in explicit opposition to close reading, which, to his mind, fails to uncover the true scope of literature. Moretti is founder of the Stanford Literary Lab that seeks to confront literary “problems” by scientific means – computational modeling, hypothesis-testing, automatic text processing, algorithmic criticism, and quantitative analysis. The Lab’s first pamphlet suggested that literary genres “possess distinctive features at every possible scale of analysis” and that there are formal aspects of literature that people, unaided, cannot detect [Allison et al. 2014, 8]. The second pamphlet used network theory to re-envision plots #moretti2011. Since then, Jockers (2013) has further developed distant reading in what he has called “macroanalysis”, a new approach to the literary reading and study designed for exploring digital texts in large quantities. Using computational tools to retrieve keywords, key phrases, and linguistic patterns across thousands of digital texts in databases allows researchers to attain quantifiable evidence on how literary trends have evolved over time and geographically. It can also determine what social, cultural, and historical connections exist between individual authors, texts, and genres [Jockers 2013]. In this article, we seek to combine such a macro-scaled method with close reading and manual annotation, and apply it to a non-literary dataset. 17

Methodology

Dataset

The text corpus used in the research described in this article comes from Reddit, a social content aggregation website and the self-styled “front page of the internet”. Functioning as “a bulletin of user-submitted text, links, photos, and videos ” [Duggan and Smith 2013, 2], it is a message board wherein users submit content and discuss this content in different communities. The website is further referred to as a “social voting site” [Gilbert 2013] as users (often referred to as redditors) vote submitted content up or down, sending the submissions with most upvotes to the Front page, i.e., the home page of reddit.com. Content on Reddit is organized in communities, so called subreddits. The nearly 900,000 communities of Reddit are organized around different topics like Technology, WorldNews, Music, Gaming, or PoliticalDiscussion. A single subreddit can be reached via, for example, *reddit.com/r/PoliticalDiscussion*. The single posts in one community are referred to as submissions. A submission usually contains a link, embedded images, gifs, or videos and may also contain a text written by the posting redditor. It is of special interest for textual analysis that other redditors can comment on a submission and reply to other comments. The comments are organized in a thread, a tree-like structure that allows for following the discussion chronologically and with regard to content. From any comment (every branch of the tree) further comments can emerge that yet again may receive comments (as more branches sprouting from the prior branch). Furthermore, redditors can make use of basic text formatting functions, such as quoting text of prior comments or highlighting. The text of the original submission and the comments on this submission are the primary source of information of this research. 18

The thread that formed the basis of our dataset was submitted on January 19th of 2017, and posed the following question: *Should the Democrats nominate a celebrity in 2020? What would be the pros and cons?*[2] Furthermore, the users added certain sub-questions, like: *Would celebrity power help or hurt a presidential nominee in the next election?* and *If this plan actually goes forward, who would be the best choice?*. Our dataset consists of 449 (461 including deleted comments) responses to these questions. Since the thread is still open, comments have been added after we collected and investigated our dataset. These new comments are not included in our research. 19

While investigations using Reddit data often aim at uncovering trends across single discussions and even across discussion forums (subreddits) (cf. [Zhang 2017]), the highest level of distant reading analysis in this research is on the level of a single discussion thread. This restricts the dataset to a size that can still be analysed by means of close reading and also constrains the comments to relate, even if only peripherally, to the one question or topic that initiated the discussion thread. 20

A common phenomenon in discussions in general is that the sub-questions emerge and that the focus of the topics shifts to new content. This illustrates that a technique is needed that can zoom in from the whole-thread level to 21 prevalent topics that group the single posts into content categories. These single posts, on the other hand, can not only be members of topical (content) groups, but also point towards groups related to discourse function in Reddit discussions [Zhang 2017].

The hierarchical structure of Reddit discussions, ranging from single comments to a whole thread, matches our goal to contrast low-level close reading with high-level distant reading. For a scholarly reading of a discussion forum, neither isolated comments nor a high level view on the discussion thread as a whole are sufficient. Scholars need to know about topical groupings that provide a frame of reference for single comments. While grouping of comments can be achieved by manual labeling, a distant reading approach is promising as a method that is faster, that in principle scales up to larger discussion threads, and that may remove some human biases during the annotation process. 22

Yet, the role of irony, emotion, and humor in this thread warrants a close reading approach. The pervasiveness of irony and ironic detachment in contemporary (online) culture has been described by Ian Bogost as the "escape from having to choose between earnestness and disdain," [Bogost 2016, 59]. Poe's Law, an adage of Internet culture, states that online, it's impossible to know who's joking and who's being serious. In *The Ambivalent Internet* (2017), Whitney Phillips and Ryan Milner show how digital communications, e.g., through GIFs, memes, and videos, are operationalized to fundamentally destabilize the worldviews of others. It is almost impossible to determine when an ironic posture is adopted. Humor, irony, and role playing are central to the behaviors in online environments and communities like those on reddit. A manual annotation is to be expected to detect more of this humor and irony than a topic model. Therefore, we have chosen to call the level of human annotation "close reading", which traditionally attends to tone and style as well as content of the message. 23

Close reading

While the initial question of the Reddit thread regards viewpoints on the pros and cons of a future celebrity president and the names of potential Democrat candidates, the thread soon developed into a more complex conversation in which different "new" questions were discussed as well. In order to grasp these associative developments inside the discussion and the mechanics of a forum like Reddit, we choose to employ a hypothesis-free form of close reading, which means we first started looking for patterns in the material of the discussion in a rather open-ended way, without explicitly framing our horizon of expectation, or what we were expecting to find, beforehand. We chose this approach since our aim was first and foremost to comparatively analyze methodologies, and only secondarily, to find an answer to the question posed in the thread, regarding celebrities in politics. 24

This process took place in a bottom-up fashion: two human annotators analyzed the posts in the thread, and chose a word or phrase to summarize each post. During this process, they developed a collection of labels that were assigned to the posts. Based on this reading, fifteen classes of posts were identified and color-coded based on the classes. Ten of them turned out to be related to three different underlying questions that we formulated based on the classes. We describe these at more length under "report of the close reading analysis". After manually annotating the 449 posts separately, we compared the lists of the outcome and slightly modified the categories to accommodate both our findings. Upon comparison, we discovered there was a high degree of congruence between the categories assigned to the posts in both annotations. 25

Distant reading

Parallel to the close reading based, manual annotations, we approach the Reddit thread from a distant reading perspective. Specifically, we adopt a data-driven ideal of distant reading, i.e. we want to start analysing without any human insight into the textual dataset. This results in two requirements on a distant reading method: First, the approach needs to be unsupervised, i.e. not relying on any prior labeling of the data. Second, the fact that Reddit threads are open-ended in the topics they comprise, the number of topics in the shape of clusters resulting from the distant reading algorithm needs to be variable. At least, there needs to be a possibility to model a range of clusters in a computationally feasible manner. These two requirements can be understood as equivalent to the hypothesis-free aspects (of the close reading approach) in our distant reading method. 26

One technique that fulfills both requirements is Latent Dirichlet Allocation (LDA) topic modeling [Blei et al. 2003]. This method allows for the automatic grouping of text documents according to latent content categories. These topics underlying a text corpus are modelled in a completely unsupervised manner, meaning that the algorithm does not know which texts belong together (for example, according to some human labeling) beforehand. This aspect of unsupervised LDA thus serves a data-driven ideal of distant reading. For each topic, LDA creates a different language model, as the underlying assumption is that different topics require different words and constructions. In our research, we expect that LDA is suitable as texts with different topics are expected to use a different "language". 27

Importantly, the number of topics resulting from LDA is a parameter that is set manually by the user. While determining an adequate number of topics for a dataset is often a problematic challenge for which no definite solutions are agreed on, the variability of the resulting number of cluster matches the second requirement to our desired distant reading technique. Because it is unknown how many topics are to be expected from a discussion thread, a pass through a range of numbers of LDA topics is necessary. Investigating a whole range of numbers of topics may additionally reveal several different levels of topical granularity that can be captured using LDA. 28

LDA has already been used for text analysis in the area of Digital Humanities. For example, Emmery and van Zaanen (2015) investigated the application of LDA to identify changes in the number of comments on news articles on the topic of online security before and after the Snowden revelation in 2013. 29

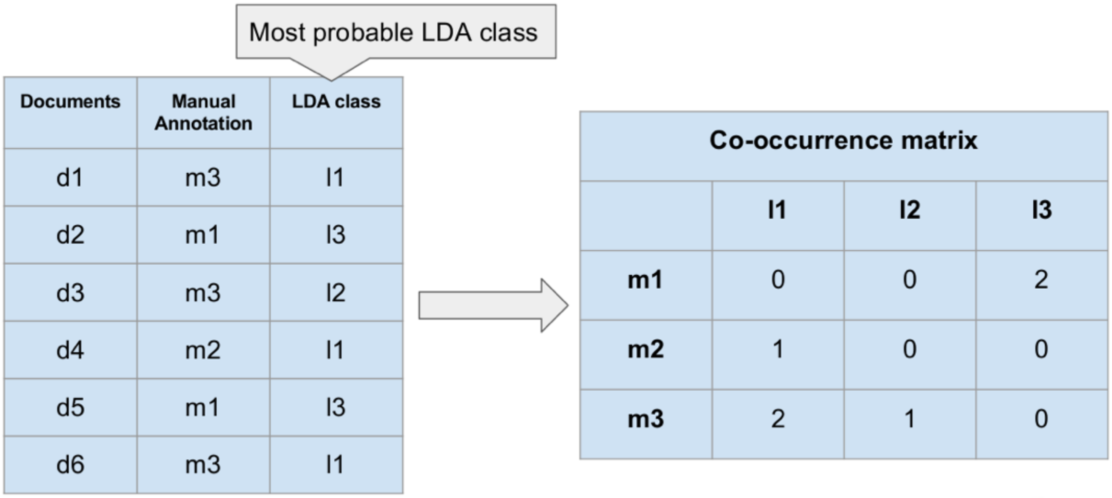
Lastly, we choose LDA as counterpart to close reading, since both approaches focus on content of the texts (posts within the Reddit thread). We strengthen this perspective by removing stop words (which are mostly function words) from the textual data, before modeling topics with LDA. 30

As mentioned above, deciding on the optimal number of topics to be modeled is problematic. In the case of a Reddit thread, new topics unfold as the discussion proceeds, which means that the number of topics depends on the size of the Reddit thread. It is therefore difficult to make an informed decision about the number of topics for an LDA model, based on manual inspection of comments alone. To resolve this problem, we propose a possibility to determine an adequate number of topics for a set of already present manual annotations. 31

During the annotation process, each post (the original thread submission question or a comment on it) in the Reddit thread is treated as a separate document (Figure 1, left). For each document, the LDA topic with the highest probability is selected. This is certainly a limiting decision, but we find that the per-document probability distributions usually strongly favor one single topic with a very high probability, while the probabilities for the other topics are close to zero. 32

Comparison of the approaches

After analyzing the documents, they are thus represented by two labels, one from two distinct sets of annotations each: a close reading annotation and a (single) LDA topic. Based on this information, we now aim to identify LDA topics and manual annotations that express similar concepts. Going through the list of documents, co-occurrences of the two annotations are counted in a matrix (Figure 1, right, LDA topics in columns, manual annotation classes in rows). 33



Most probable LDA class

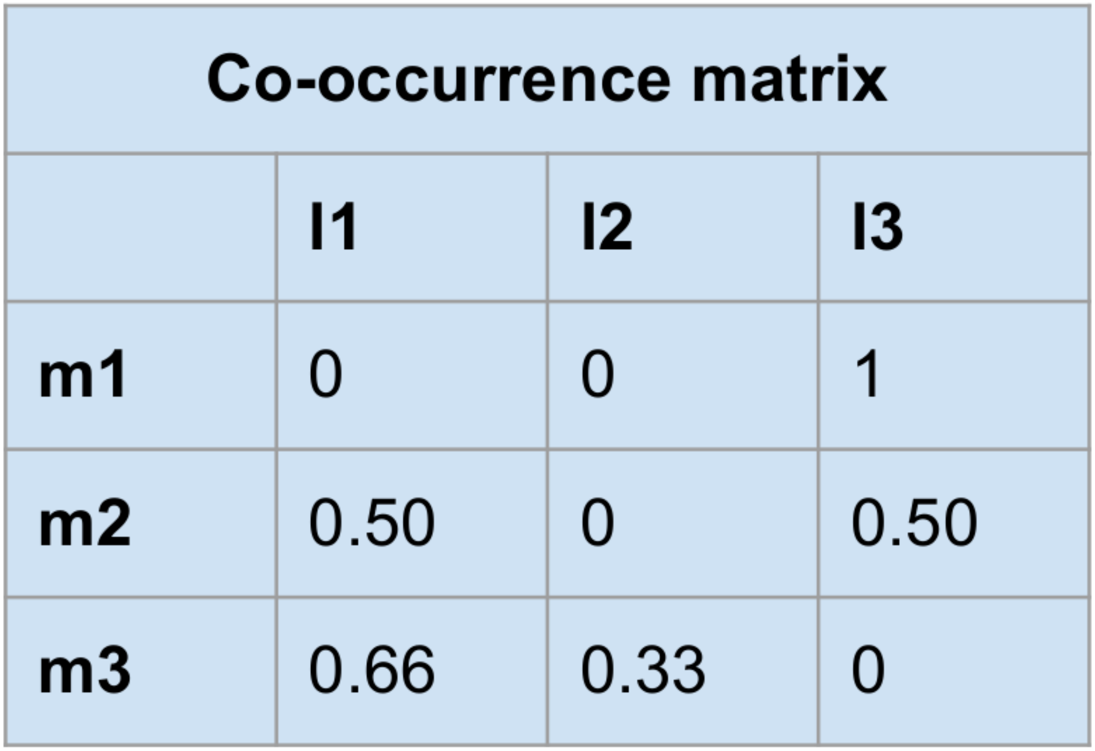
Documents	Manual Annotation	LDA class
d1	m3	l1
d2	m1	l3
d3	m3	l2
d4	m2	l1
d5	m1	l3
d6	m3	l1

Co-occurrence matrix

	l1	l2	l3
m1	0	0	2
m2	1	0	0
m3	2	1	0

Figure 1. Co-occurrences of manual annotation classes and LDA classes are counted in a matrix. Only the most probable LDA class per document is taken into account.

Given the co-occurrence matrix, the aim is to find, for each LDA topic, the manual annotation class that is best represented by the topic. In order to do so, the matrix is passed columnwise, and the manual annotation class in the row with the highest count is selected as corresponding best to the LDA class of that column. At this point we introduce an optional step of normalisation, applied on the co-occurrence matrix (Figure 2). Collecting absolute counts in the matrix may give an unfair advantage to annotation tags with high class support. During normalisation, the counts are divided by the class support of the manual annotation class of this row, resulting in fractions instead of absolute numbers. The steps described in the rest of this article always make use of both the absolute counts and the normalised counts matrix separately. 34

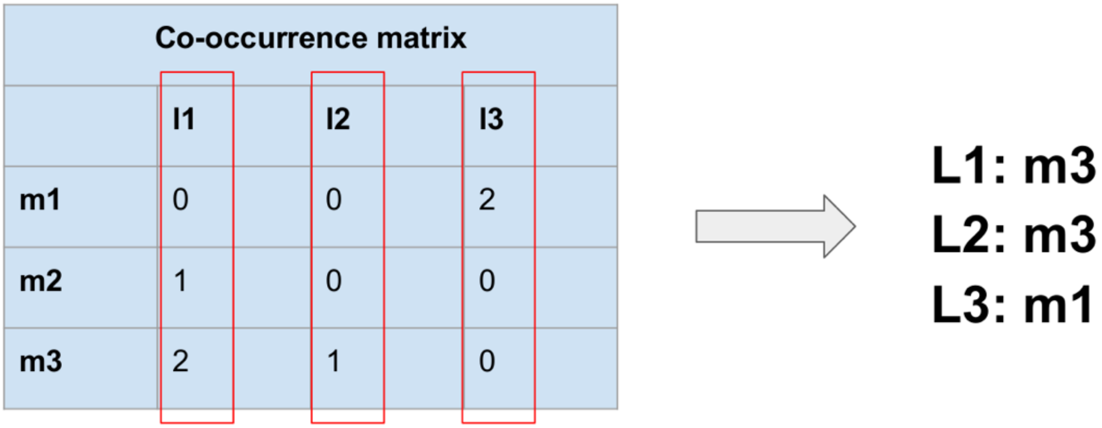


Co-occurrence matrix

	l1	l2	l3
m1	0	0	1
m2	0.50	0	0.50
m3	0.66	0.33	0

Figure 2. Normalising co-occurrence counts, by dividing by manual class support.

For each manual/LDA pair of classes with highest co-occurrence, we generate a mapping from LDA classes to manual 35 annotation classes (Figure 3). The mapping expresses classes of annotations that, from now on, we regard as corresponding to each other. Note that, as in the example of Figure 3, not necessarily all manual annotation classes are covered by the mapping.



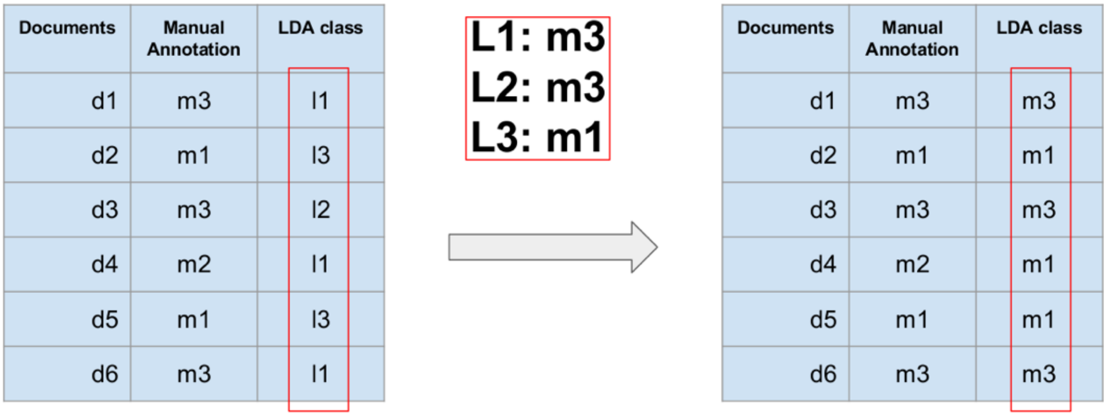
Co-occurrence matrix

	l1	l2	l3
m1	0	0	2
m2	1	0	0
m3	2	1	0

L1: m3
L2: m3
L3: m1

Figure 3. Using the highest value per column in the co-occurrence matrix (absolute counts shown), a mapping from LDA class to close reading class is created.

36 Using the mapping, the LDA classes are translated into manual annotation classes (Figure 4). These can be regarded as the computer's "guess" of the manual class that was assigned by a human. In the original list of documents, each document is now represented by two annotations that are drawn from one *common* pool of possible annotations. In machine learning terms, the manual annotations are treated as gold standard (or true labels) and the mapped LDA classes are regarded as predictions. To measure the extent to which the predictions and the gold standard overlap, we calculate the accuracy.



Documents	Manual Annotation	LDA class
d1	m3	l1
d2	m1	l3
d3	m3	l2
d4	m2	l1
d5	m1	l3
d6	m3	l1

L1: m3
L2: m3
L3: m1

Documents	Manual Annotation	LDA class
d1	m3	m3
d2	m1	m1
d3	m3	m3
d4	m2	m1
d5	m1	m1
d6	m3	m3

Figure 4. Using the mapping, the per document LDA annotation is transformed to manual classes.

37 For a single topic model, we have now obtained a measure of overlap between close reading based, manual annotations and distant reading LDA topics. In order to find the optimal topic model, several topic models with varying numbers of topics can be computed and the overlap accuracy can be compared. For the present study, we produced topic models with a number of topics ranging from 1 topic to *N* topics, with an *N* equal to the number of documents in the collection. We expect a topic model with as many topics as are documents in the corpus to be of low expressive power. Similarly, a topic model with only one LDA class cannot adequately express several manual annotation classes.

38 For the Reddit thread on the question *Should the Democrats nominate a celebrity in 2020? What would be the pros and cons?*, overlap accuracy is calculated for LDA models with 1 to 461 topics. The topic model with the best fit to the manual annotations is defined as the topic model corresponding to the highest accuracy value.

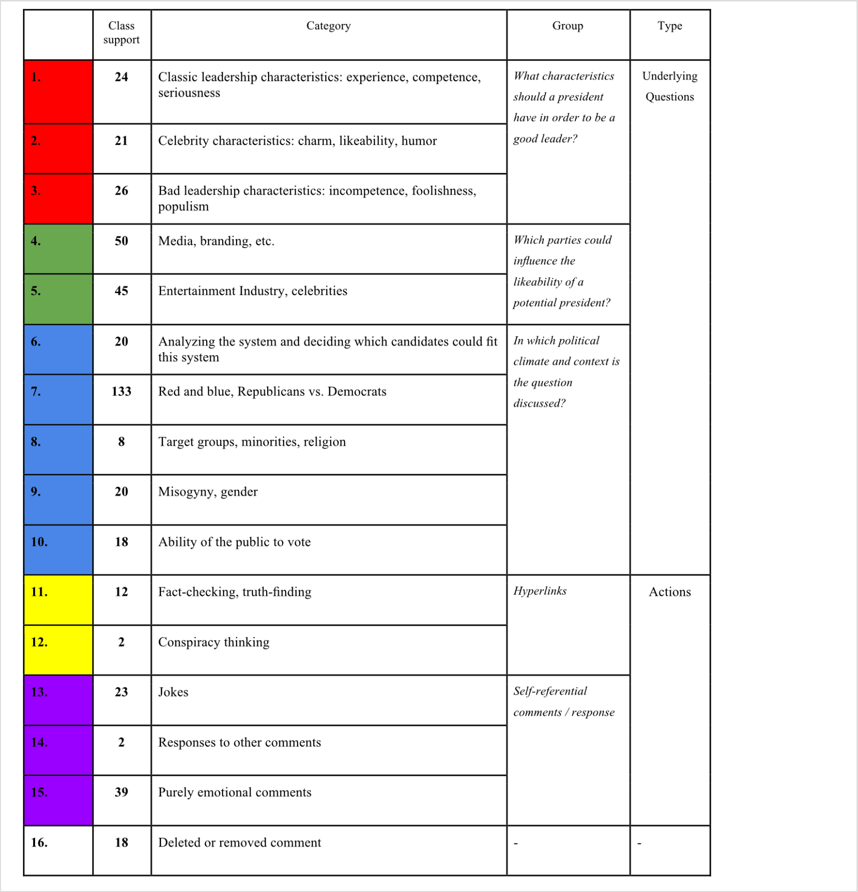
Note that increasing the number of LDA topics leads to a higher likelihood of high accuracy. Imagine the situation in which each document receives a unique LDA class. This allows for a perfect accuracy, but the predictive power is extremely low as no document is comparable according to the LDA classes. To resolve this issue, a second possibility to determine the optimal number of topics is also investigated. Here, we reverse the mapping step and transform manual classes to LDA classes. The steps described above stay the same, except the co-occurrence matrix is passed row-wise instead of column-wise[3]. 39

As seen in the results, both the forward mapping and the reverse mapping result in a perfect solution: Mapping LDA classes to manual classes leads to increasingly high accuracy with an increasing number of topics. Mapping manual classes to LDA classes, in turn, leads to complete overlap for a single LDA class, because all manual classes are correctly mapped to that single LDA class. The maximum accuracy values in the two scenarios are misleading because neither having a single LDA topic nor having hundreds is useful for the purpose of distant reading. In our approach we exploit this phenomenon by aiming for the right balance between the two perfect solutions. We do so by calculating the absolute difference between forward and reverse accuracy for each number of topics. Subsequently, we select the number of topics with the lowest absolute difference in the accuracies from the forward and reverse mapping. 40

Results

Report of the close reading analysis

Based on the close reading strategy described above, we identify fifteen classes of posts (Table 1). Ten of them are related to three different underlying questions that we formulate based on the classes: *What characteristics should a president have in order to be a good leader?* (Red, class 1-3), *Which parties could influence the likeability of a potential president?* (Green, class 4-5), *In which political climate and context is the question discussed?* (Blue, class 6-10). The other five are classified as “actions” that are divided over two groups, hyperlinks to other sources (Yellow, class 11-12) and (self-)referential comments/responses (Purple, class 13-15). 41



	Class support	Category	Group	Type
🟥 **1.**	**24**	Classic leadership characteristics: experience, competence, seriousness	*What characteristics should a president have in order to be a good leader?*	Underlying Questions
🟥 **2.**	**21**	Celebrity characteristics: charm, likeability, humor		
🟥 **3.**	**26**	Bad leadership characteristics: incompetence, foolishness, populism		
🟩 **4.**	**50**	Media, branding, etc.	*Which parties could influence the likeability of a potential president?*	
🟩 **5.**	**45**	Entertainment Industry, celebrities		
🟦 **6.**	**20**	Analyzing the system and deciding which candidates could fit this system	*In which political climate and context is the question discussed?*	
🟦 **7.**	**133**	Red and blue, Republicans vs. Democrats		
🟦 **8.**	**8**	Target groups, minorities, religion		
🟦 **9.**	**20**	Misogyny, gender		
🟦 **10.**	**18**	Ability of the public to vote		
🟨 **11.**	**12**	Fact-checking, truth-finding	*Hyperlinks*	Actions
🟨 **12.**	**2**	Conspiracy thinking		
🟪 **13.**	**23**	Jokes	*Self-referential comments / response*	
🟪 **14.**	**2**	Responses to other comments		
🟪 **15.**	**39**	Purely emotional comments		
16.	**18**	Deleted or removed comment	-	-

Table 1. Close reading annotation classes.

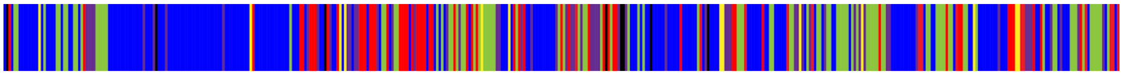


Figure 5. Linear data-visualization of the Reddit thread.

By assigning colors to the classes that relate to the same question or action, we can distinguish the variety of discussions inside one subreddit and see how these discussions develop throughout the thread (Figure 5). From left to right, the first 130 comments mostly discuss the political climate (Blue) (e.g., *“Democrats didn’t show up. Republican turnout fairly steady.”* (post 63)). These comments seem to create a context for the discussion as a whole. Thereafter, the characteristics of an ideal president are discussed (Red) (e.g., *“I say charisma goes a long way, part of being president is to inspire people. Dems should go young and the candidate should be well spoken (Obama) and inspiring. Leave the dynasties, get someone younger, inspire people to come out and vote for them.”* (post 120)). Near the end, the influence of media and celebrity culture are set against the former discussions (Green) (e.g., *“Do you think he would’ve won the popular vote if not for that video”* (post 17)). The “actions” of hyperlinking (Yellow), joking (e.g., *“We’ll get him a box. Worked for Napoleon.”* (post 239)) and (emotional) responding (e.g., *“I’ll concede that that’s an excellent point.”* (post 57)) (Purple) occur equally frequent throughout the whole subreddit. The variety of questions discussed in the subreddit does not seem to occur randomly, but in clusters: We can see a certain development in how the colors follow each other up. 42

Report of the distant reading analysis

Following application of the LDA topic modeling technique and the process to assign LDA topics to manual topics, we measure the overlap between manual and computational annotation. Accuracy measures are obtained for a varying number of topics in the LDA model. Figure 6 plots the overlap accuracy over the number of topics. It can be observed 43 that the accuracy keeps increasing with higher numbers of topics in the LDA models. The reason for this is, that with more LDA topics, more and more topics get assigned to only one document. During the mapping, the single occurrence of an LDA class is necessarily converted to the correct manual annotation class. This creates the aforementioned perfect solution to the mapping problem. However, the model with the highest overlap accuracy has low generalising power. Such a detailed topic model moves away from the desired distant reading perspective, which would group the comments into a limited set of categories.

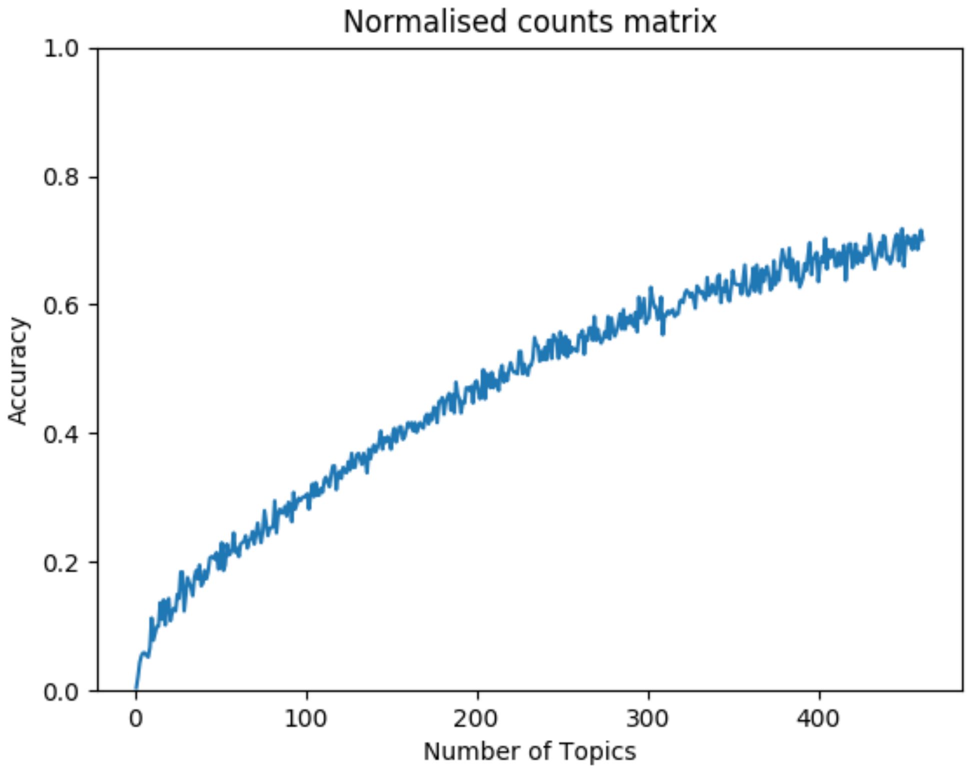


Figure 6. Evolution of overlap accuracy over the number of topics used in the LDA model.

In order to resolve the problem of overspecification, we look at a combination of the forward and reverse accuracy. As outlined above, we balance the two LDA to manual assignment and manual to LDA assignment scores and inspect where the two curves intersect. The intersection is defined as the number of topics with the lowest absolute difference in accuracy. Figures 7 and 8 depicts forward and reverse accuracies over number of topics based on absolute and normalised co-occurrence counts.

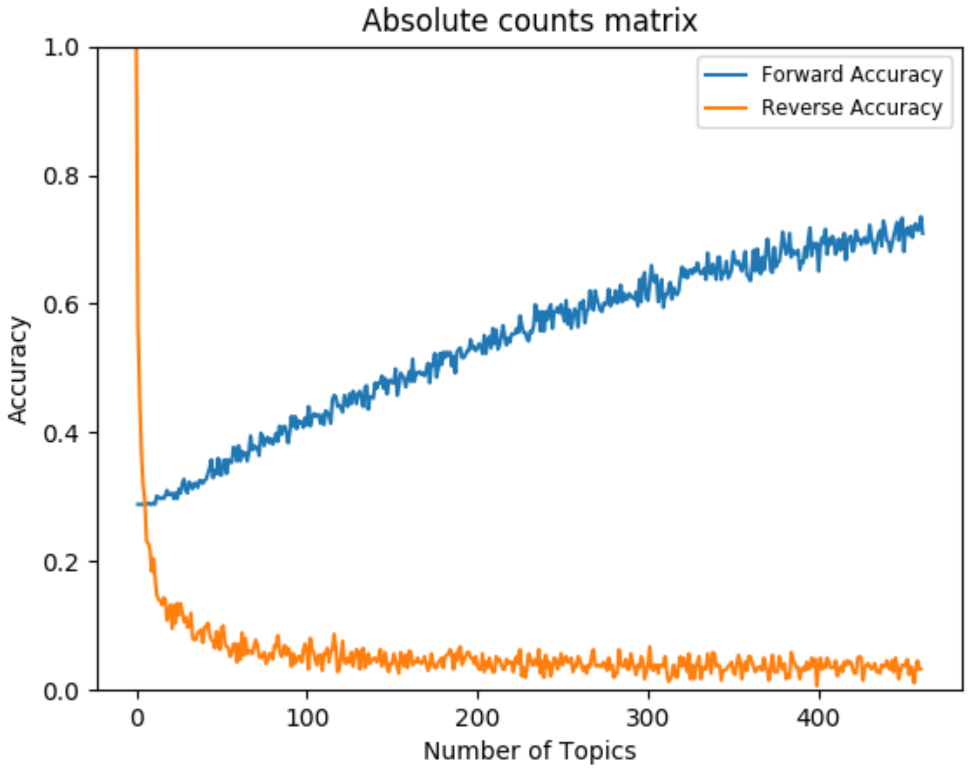


Figure 7.

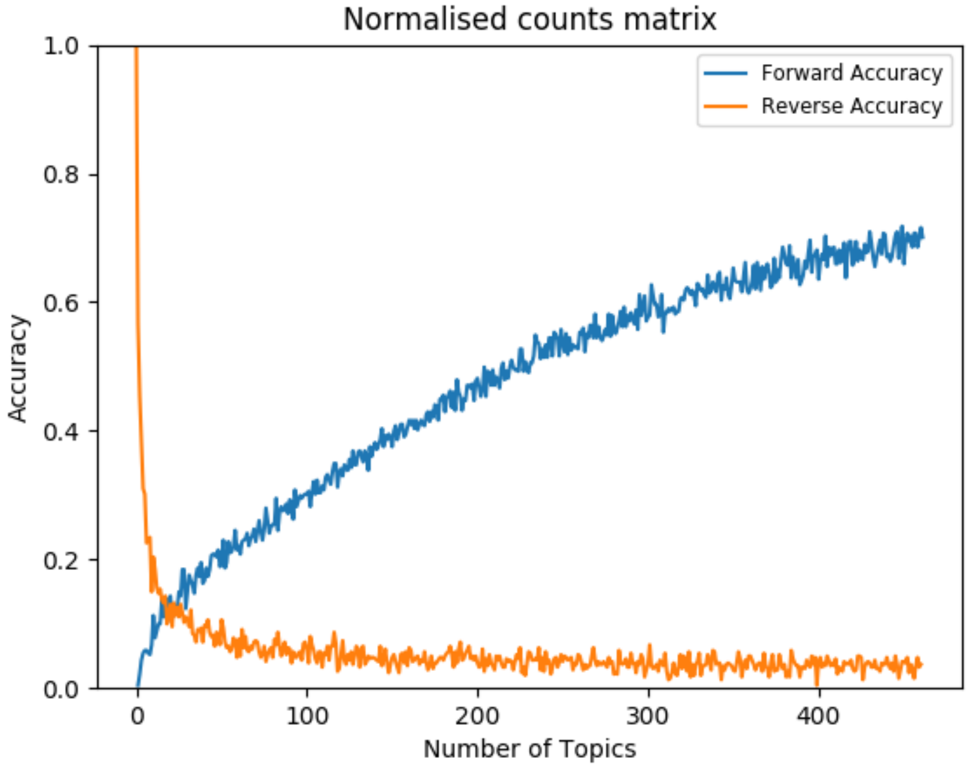


Figure 8. Evolution of the forward (LDA to manual class) and reverse mapping accuracies.

To evaluate how well the LDA model overlaps with the manual annotations it is of fundamental importance to compare the performance against a baseline. As baseline we use a random assignment of the manual annotations to the documents. We explicitly decide not to make use of another common baseline, the majority class baseline. The reason for this is that we aim to cover all classes instead of just a single one and random assignment better reflects this goal. The random assignment labels are then compared to the actual manual assignment. This process is bootstrapped and averaged over 1000 trials. The accuracy measures derived from the random assignment and the LDA topic models are presented in Table 2. Regardless of whether the co-occurrence matrix is normalised for class support the LDA classes lead to a higher overlap with the manual annotations than the random assignment. The overlap between the LDA classes and the manual annotations is thus above chance level. Comparing the non-normalised model to the normalised model it can be observed that normalisation increases recall by covering relatively many manual classes to the mapping even when a small number of topics is modeled.

46 For the Reddit thread case study, the described process results in a topic model with 6 topics that has the highest overlap with the gold standard annotations. However, the measures based on absolute counts are still strongly biased towards the majority class. Since we intend to cover all classes as well as possible, the accuracy based on absolute counts is too optimistic for our purpose. The corresponding model using normalised counts may better generalise to all classes. According to the model normalised for class support, 23 topics best represent the manual annotation.

	Accuracy	Number of Topics
Baseline (Random Assignment)	6.03%	-
Forward + Reverse (absolute counts matrix)	29.17%	6
Forward + Reverse (normalised counts matrix)	13.01%	23

Table 2. Baseline accuracy (random assignment) and maximum accuracies based on the proposed overlap measures with their corresponding number of topics.

Discussion

47 The manual annotation revealed that we did not only identify several opinions or find information that regards the question that the thread set out from; in addition, we found another class of posts that we choose to describe along the lines of *discourse function*: for instance, humor, hyperlinks, or direct responses (e.g., “yes”, “no”, “indeed”). These are elements that fulfill a certain function within the larger collective discourse, without being reducible to an answer to the main question, an opinion or a piece of information. The posts that fall within the function classes do not correspond to specific topics, so it is likely that these elements can only be identified by close reading. In particular, posts that deal with humor and irony are hard to identify computationally, which underlines the importance of human annotation. The bottom-up process of manual annotation and our hypothesis-free form of close reading unraveled underlying questions and contexts that we can use to aid computational strategy.

48 The manual annotation (close reading), as made visible by the color coding, seems to point to an associative structure between the comments. So far, our close reading approach regards the posts in a linear (chronological) order, while our distant reading methodology regards the posts without any order or structure. However, Reddit also organizes posts in hierarchical structure. What further differentiates Reddit from a real-time information network like Twitter is that the community curates the stream of content themselves. Items that they consider to be of value are “upvoted”, and those deemed unworthy are “downvoted”. This way the position of root comments within one single thread, i.e., comments that directly reply to the original submission, is determined by Reddit’s voting system. We did not investigate this in the current research, but it will certainly influence the required reading strategy for this specific environment. Note that the current distant reading approach cannot easily incorporate the inherent hierarchical information of the dataset.

49 By supplementing our close reading with a distant reading methodology, we can give the reader an overview of the relations between the discussions. The idea of studying the hierarchical structure, which with the present method

remains un(der)explored, will be the next step in our research. Examining this hierarchical structure as an extra dimension is necessary in order to come to a more comprehensive take on the information we encounter on this online platform.

Second, there is a problem with the hypothesis-free approach, as employed in our study. A predetermined research question aids in keeping the set of close reading annotation types controlled, i.e., a small number of unique topics as well as topics that stay within concise conceptual borders. Still, there are many questions on the close reading side, leading to different collections of close reading classes and we need to determine which of those correspond to the identified distant reading classes. In other words, it is as of yet unclear how generalizable this approach is. 50

We have proposed a way to find a mapping from LDA to manually assigned classes. The analysis of the close and distant reading demonstrates that LDA is indeed a possibility to generate a computational model of close reading annotations. However it needs to be clearly stated that the overlap between the two sets of labels, as measured by accuracy, is relatively low. This may, first, be due to the rather high number of 15 classes of comments on a relatively small dataset (461 comments). Second, some of the close reading based classes describe function rather than content. LDA topic modeling is designed to primarily capture content information and may thus not be able to accurately capture functional classes. 51

Still, the proposed approach leads to an overlap above chance level and future work should try to further increase the overlap. If we can find an LDA model that overlaps well with a set of annotations it is possible to generalise from the LDA model to new, unseen data. From this point on, less manual annotation would be needed, which would benefit the literary analysis by extending the size of the data. Conclusions can then be drawn on the grounds of a larger dataset. 52

Finally, the research was carried out by a diverse group of researchers, both from a computational and a cultural studies background. The interdisciplinary character of the research group is essential for this type of research, but also introduces challenges. The terminology used in the two fields is not exactly the same, which leads to a confusion of tongues. For example, clustering texts is not the same as close reading, even though the resulting information may be used as the basis for close reading. Furthermore, close reading entails much more than human annotation of text samples. We believe that for a successful application of distant reading approaches in a close reading context, this language or terminology boundary between the disciplines needs to be crossed. 53

Conclusion

We have argued that close and distant reading, from the beginning of the latter, have gone hand in hand, other than Moretti’s polemic introduction to the topic would suggest. Studies that use distant reading methods and combine it with close reading most often follow the information seeking mantra or top down approach (overview first, zoom & filter, details on demand). As an alternative, we presented a study in which we used manual annotation not as a sample that follows from the overview produced by the distant reading, but a method that comes before, or runs parallel to, the LDA analyses, to evaluate its workings and to reflect on its strengths and weaknesses. We used close reading of posts and manual annotation to fill in the gaps, to evaluate and to reflect upon the LDA distant reading. 54

We see that close reading may reveal interesting detailed knowledge that distant reading may not find, while distant reading allows for the identification of large-scale patterns by analyzing vast collections of text that cannot be seen when only analyzing small amounts. Reading the same corpora at different scales using differently tuned digital instruments can then be more illuminating than either close or distant reading of their own accord. 55

Close reading digs for complexity, opacity, irony, and ambiguity: values that stand to be reappraised in a time when we encounter vast bodies of information through multiple platforms, and when, moreover, we tend to overemphasize transparency and immediacy when processing this information. Even though we live in a time that tends to privilege the “full picture” and distrust sampling, we contend with Johanna Drucker that meaning-making is precisely based on editing, focus, and finitude: “Editing towards meaning is a fundamental skill of human survival, through the selection of pertinent information, which accumulates in a significant pattern” [Drucker 1997, 109]. Yet, another indispensable skill in a time of information overload is the ability to skim, to filter out, and to ignore the inessential. 56

Therefore, we undertook a first, explorative step in this paper towards a mixed methodology where the input of a close reading drives the analysis of the distant reading: an analysis that incorporates local annotation in a "distant" analysis. We have investigated and evaluated the use of the content-oriented LDA technique to identify clusters of texts in a way that simulates a close reading approach. Automatic analysis allows for a motivated selection of documents that should be considered for a deeper close reading analysis. Regarding distant reading methodologies, we find that our current approach that relies on LDA topic modeling cannot fully replace manual annotations altogether. Still, it may be valuable to extend from already present manual annotations. Future work will focus on human-in-the-loop approaches, such as labeled LDA [Ramage et al. 2009], which can steer the clustering method based on prior human-proposed knowledge. LDA topic modeling could then be used to find content-based clusters with a higher density of keywords, questions, and topics. Once clustered, the groups can be submitted to a closer (close reading) analysis. Our further research will explore and evaluate this possibility for Reddit discussions. 57

This work was supported by the Institute for Advanced Study in Budapest. 58

Notes

[1] [Jockers 2013, 6] Other seminal works to mention in this respect are Cleanth Brooks and Robert Penn Warren's *Understanding Poetry* (1938) and Laurence Perrine's *Sound and Sense* (1956), which together come close to "an orthodoxy of close reading" [Culler 2010, 22].

[2] https://redd.it/5oy1sz

[3] It is equally possible to simply transpose the input matrix and leave all other steps the same.

Works Cited

Aiden and Michel 2013 Aiden, E., and J. Michel. *Uncharted: Big Data as a Lens on Human Culture*. New York: Penguin, 2013.

Algee-Hewitt et al. 2016 Algee-Hewitt, M., S. Allison, M. Gemma, R. Heuser, F. Moretti, and H. Walser. "Canon/Archive. Large-scale Dynamics in the Literary Field." *Literary Lab Pamphlet 11*, 2016.

Allington et al. 2016 Allington, D., Brouillette, S., and D. Golumbia. "Neoliberal Tools (and Archives): A Political History of Digital Humanities." *LA Review of Books*, May 1, 2016, pp. 1-5.

Allison et al. 2014 Allison, S., et al. "Quantitative Formalism: An Experiment." *Stanford Literary Lab, Pamphlet 1*, 15 Jan. 2011. Web. 16 April 2014.

Blei et al. 2003 Blei, D.M., Ng, A.Y., and M.I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* , vol. 3, no. Jan., 2003, pp. 993-1022.

Bogost 2016 Bogost, I., 2016. *Play Anything. The Pleasure of Limits, the Uses of Boredom, and the Secret of Games*. New York: Basic Books, 2016.

Booth 2017 Booth, A. "Mid-range Reading: Not a Manifesto." *PMLA* 132.2, 2017. 620–627.

Brooks 1947 Brooks, C. *The Well-wrought Urn: Studies in the Structure of Poetry*. San Diego: Harcourt, 1947.

Cohen 1999 Cohen, M. *The Sentimental Education of the Novel*. Princeton: Princeton University Press, 1999.

Compagnon 2014 Compagnon, A. The Resistance to Interpretation. *New Literary History,* vol. 45, no. 2,2014, pp. 271- 80.

Culler 2010 Culler, J. "The Closeness of Close Reading." *ADE Bulletin* vol. 149, no. -, 2010, pp. 20-25.

Don et al. 2007-08 Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B., Plaisant, C. "Discovering interesting usage patterns in text collections: integrating text mining with visualization." HCIL Technical Report, 2007-08.

Drucker 1997 Drucker, J. "The Self-Conscious Codex: Artists' Books and Electronic Media." *SubStance*, vol. 26, no. 1, 1997, pp. 93-112.

Drucker 2013 Drucker, J. *Intro to Digital Humanities*, Sept. 2013. Web. 10 Aug. 2015.

Drucker 2017 Drucker, J. "Why Distant Reading Isn't." *PMLA* 132.3, 2017. 628–35.

Duggan and Smith 2013 Duggan, M., and A. Smith. “6% of online adults are reddit users.” *Pew Internet & American Life Project*, vol. 3, no. Jul., 2013, pp. 1-10.

Earhart 2015 Earhart, A. *Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies*. Ann Arbor: University of Michigan Press, 2015.

Emmery and van Zaanen 2015 Emmery, C., and M. van Zaanen. “Modelling Discussion Topics to Improve News Article Tagging”. Presented at the 2nd Digital Humanities Benelux Conference (DHBenelux 2015), Antwerp, Belgium. 9 June 2015.

Gilbert 2013 Gilbert, E. “Widespread underprovision on Reddit.” *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013.

Guldi 2018 Guldi, J. “Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora,” *Journal of Cultural Analytics*. December 20, 2018.

Heyman 2015 Heyman, S. “Google Books: A complex and controversial experiment”. *New York Times*, *28*, 2015.

Igarashi 2015 Igarashi, Y. “Statistical Analysis at the Birth of Close Reading.” *New Literary History*, Vol. 46, no. 3, 2015. pp. 485-504.

Jockers 2013 Jockers, M. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press, 2013.

Jänicke et al. 2015 Jänicke, S., G. Franzini, M. F. Cheema, and G. Scheuermann. “On close and distant reading in digital humanities: A survey and future challenges.” *Eurographics Conference on Visualization (EuroVis)-STARs. The Eurographics Association*, 2015.

Kelly 2012 Kelly, K. “Scan This Book!” *The New York Times Magazine*, 14 May 2006. Web. 25 April 2012.

Kirschenbaum 2007 Kirschenbaum. M. “The Remaking of Reading: Data Mining and the Digital Humanities.” NF Symposium on Next Generaion of Data Mining and Cyber-Enabled Discovery for Innovation. 11 October 2007.

Lee et al. 2018 Lee, J.J., B. Greteman, J. Lee, and D. Eichmann, “Linked Reading: Digital Historicism and Early Modern Discourses of Race around Shakespeare's Othello,” *Journal of Cultural Analytics*. Jan. 25, 2018.

Liu 2013 Liu, A. “The Meaning of the Digital Humanities,” *PMLA* vol. 128, no. 2, 2013. pp. 409-23.

Manderino 2015 Manderino, M., “Reading and Understanding in the Digital Age. A look at the critical need for close reading of digital and multimodal texts.” *Reading Today*, Jan/Feb. 2015, pp. 22-23.

Mayer-Schönberger and Cukier 2013 Mayer-Schönberger, V., and K. Cukier. *Big Data: A revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt, 2013.

McGrath 2018 McGrath, L., D. Higgins, & A. Hintze, “Measuring Modernist Novelty,” *Journal of Cultural Analytics*. November 9, 2018.

Moretti 2000 Moretti, F. “Conjectures on World Literature.” *New Left Review*, vol. 1 no.-, 2000, pp. 54-68.

Moretti 2000a Moretti, F. “The Slaughterhouse of Literature.” *MLQ: Modern Language Quarterly*, vol. 61 no. 1, 2000a. pp. 207-227.

Moretti 2009 Moretti, F. “Critical Response: II. ‘Relatively Blunt’.” *Critical Inquiry* vol 36, no. 1, 2009, pp. 159-71.

Moretti 2013 Moretti, F. *Distant Reading*. London: Verso, 2013.

Phillips and Milner 2017 Phillips, W., and R. Milner. *The Ambivalent Internet: Mischief, Oddity and Antagonism Online.* Cambridge: Polity, 2017.

Piper 2017 Piper. A. “Think Small: On Literary Modelling.” PMLA 132.2, 2017, 613-619.

Ramage et al. 2009 Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009, August). Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, vol. 1*, no. -, pp. 248-256. Association for Computational Linguistics.

Ramsay 2011 Ramsay, S. *Reading Machines: Toward an Algorithmic Criticism*. U of Illinois P, 2011.

Richards 1929 Richards, I.A. 1929. *Practical Criticism: A Study of Literary Judgment*. London: Routledge, 2014.

Rosen 2011 Rosen, J. “Combining Close and Distant, or the Utility of Genre Analysis: A Response to Matthew Wilkens’s

‘Contemporary Fiction by the Numbers’,” in *Post45*, December 3, 2011.

Shneiderman 1996 Shneiderman, B. “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations In Visual Languages”, *Proceedings*, 1996, pp. 336–343.

So 2019 So, R.J., H. Long, and Y. Zhu, “Race, Writing, and Computation: Racial Difference and the US Novel, 1880–2000,” *Journal of Cultural Analytics*. January 11, 2019.

Stronks 2013 Stronks, E. “De afstand tussen *close* en *distant*. Methoden en vraagstellingen in computationeel letterkundig onderzoek.” *Tijdschrift voor Nederlandse Taal- en Letterkunde*, vol. 129, no. 4, 2013, pp. 205-14.

Underwood 2016 Underwood, T. “The Longue Durée of Literary Prestige”. *Modern Language Quarterly* vol 77, no. 3, 2016. pp. 321-44.

Vaidhyanathan 2011 Vaidhyanathan, S. *The Googlization of Everything*. Berkeley and Los Angeles: University of California Press, 2011.

Wilkens 2011 Wilkens, M. “Canons, Close Reading, and the Evolution of Method.” Debates in the Digital Humanities. Ed. Matthew K. Gold. Minneapolis: University of Minnesota Press, 2011, pp. 249-58.

Wilkens 2013 Wilkens, M. “The Geographic Imagination of Civil War-Era American Fiction,” in *American Literary History* vol 25, no. 4, 2013. pp. 803-840.

Wimsatt and Beardsley 1946 Wimsatt, W.K., & M. Beardsley. “The intentional fallacy.” *Sewanee Review*, vol. 54, no. -, 1946, pp. 468-88.

Wimsatt and Beardsley 1949 Wimsatt, W.K., & M. Beardsley. “The affective fallacy.” *Sewanee Review*, vol. 57, no. 1, 1949, pp. 31-55.

Zhang 2017 Zhang, A.X., B. Culbertson, and P. Paritosh. “Characterizing online discussion using coarse discourse sequences.” *Proceedings of the Eleventh International Conference on Web and Social Media. AAAI Press*. 2017.