

## Tilburg University

### Digital Hermeneutics and Media Literacy

van de Ven, Inge; van Nuenen, Tom

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

van de Ven, I., & van Nuenen, T. (2020). *Digital Hermeneutics and Media Literacy: Scaled Readings of The Red Pill*. (Tilburg Papers in Culture Studies; No. 241).

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Digital Hermeneutics and Media Literacy: Scaled Readings of The Red Pill ✓✓

by

Inge van de Ven & Tom van Nuenen

Tilburg University / Kings College London

I.G.M.vdVen@tilburguniversity.edu ✓✓

April 2020



This work is licensed under a  
Creative Commons Attribution-NonCommercial 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

# **Digital Hermeneutics and Media Literacy:**

## **Scaled Readings of The Red Pill**

*Inge van de Ven & Tom van Nuenen*

### **Online Challenges for Media Literacy**

In Europe and the United States, trust in the media has been steadily eroding since the 1960s (Bialik and Matsa, 2017). Americans increasingly suspect mainstream media of bias and are driven to find more objective news sources (Gallup and Knight, 2018). Academics, traditional news outlets, and bureaucratic news sources alike are faced with an erosion of trust from the public. Doubts about the accuracy of information are rendered yet more pressing due to the increasing worry that internet platforms act as radicalization pathways (Munn, 2019). Recommendation algorithms on platforms such as YouTube steer users toward ‘edgier’ content (Danks and London, 2019), and ‘clickbait’ articles on online news media – aimed at generating online advertising revenue at the expense of quality or accuracy – remain widely popular (Chen et al., 2014). Fringe online extremism has become mainstream (Nagle, 2017), with discussion forums such as the infamous politically incorrect (‘/pol/’) board on 4CHAN or The Red Pill on Reddit causing headlines due to the behavior of their members. These communities are born, in part, out of a struggle to find one’s identity and make sense of the world around oneself in a culture marked by distrust in authoritative sources of information.

In this context, the importance of the burgeoning field of media literacy education has been stressed, as a field that could provide training for students to become critical and avid readers who are able to distinguish truths from falsehoods. Media literacy education often centers on teaching students how to critically question sources, and raise doubt regarding the sender's motivation. However, in Western culture today, there is by no means a universal consensus on the reliability of major news outlets, scientific publications, and academic experts. In certain communities, the 'liberal' media like *The New York Times* are held in contempt due to their biases. As danah boyd has argued (2017), a well-intended emphasis on fact checking on the part of journalistic media and educators as a solution to misinformation might have the inadvertent effect of suggesting that in complex global, socio-political issues, there is always a singular truth, or one legitimate worldview out there, waiting to be found (boyd, 2017; 2018). Moreover, discovering this single truth becomes a responsibility that lies with the individual media user. The crisis we face today, boyd argues, takes place not at the level of facts, of *what* is true, but of epistemology: how we *know* whether something is true. Indeed, the often-used true/false dichotomy fails to render the way in which enunciations are solidified by the work of all sorts of actors (Latour, 2005); 'facts' are built by a complex work of 'truth-grounding' (Lynch, 2017).

What we need is the ability to understand epistemologies that differ from our own, requiring a cultural transformation of 'how we make sense of information, whom we trust, and how we understand our own role in grappling with information' (boyd, 2017). As educators in the humanities and social sciences, we face the challenge of understanding the epistemological differences between groups in society, instead of asking our students to adapt and repeat certain

accepted ideas on what is ideologically right or wrong. We need university educators who are able to teach across epistemologies.

### **Digital Hermeneutics and pedagogy**

To work towards filling this gap, we introduce an educational approach for digital hermeneutics. In past years, different authors have discussed the potential role of hermeneutics in reflecting on the role of (digital) technology in mediating between human subjects and the world (Mallery et al., 1986; Floyd et al., 1992, Capurro, 1992, 2010; Ihde, 1998; Ramsay, 2010; Gerbaudo, 2016; Romele et al., 2018). Rafael Capurro, for instance, used the term ‘digital hermeneutics’ to insist that hermeneutics has to engage with the challenges offered by digital technology; it must develop an ‘understanding [of] the foundations of digital technology and its interplay with human existence’ (2010: 37). Much like Romele et al, we regard digital hermeneutics primarily as a methodological issue, ‘demonstrating how dealing with digital methods and digital objects for approaching specific entities such as political opinion is still a form of interpretation, no matter how automated and quantified these methods are’ (Romele et al., 2018: 4). Our approach also relates to Paolo Gerbaudo’s concept of ‘data hermeneutics’ (2016) as a methodological reply to the anti-interpretative ideology of contemporary ‘dataism’. Gerbaudo stresses the need to identify procedures to select samples from datasets, so that they can be analysed in more depth. This, he argues, should be based on two processes: qualitative sampling procedures to reduce the size of (social media) datasets, and the development of a ‘data close reading’ that may help interpretation in relation to individual narratives, dialogical motivations and social worldviews.

Building on these points, we develop a range of pedagogical strategies for the interpretation of online culture to tackle the urgent challenges of media literacy. Digital hermeneutics combines a training in (and reflection on) interpretation with the use of computational methods and tools. Inspired by the dialogical hermeneutics set out by Hans-Georg Gadamer in *Truth and Method* (2004 [1960]), we stress the importance of interpretation as a *dialogical* process – especially in the face of sentiments described, including the increasing weight in Western culture of doubt, skepticism, and the overvaluing of independent truth-finding that Boyd warns against.

Gadamer's perspective on the interpretation of art commences from the insight that there is an 'insuperable difference between the interpreter and the author' (2004: 296). He famously understood the interpretive enterprise as a dialogue, or productive conversation with the text. Following Martin Heidegger in his decommissioning of the subject-object paradigm, Gadamer notes that the engagement with artistic representations needs to be a fundamentally reflexive exercise. In contrast to the positivist Enlightenment tradition in which subjectivity has to be left 'at the door' when commencing the analysis, Gadamer urges us to understand the existentialist tenet that prejudices are a function of our deep involvement and convergence with the world – and that they are necessary for any productive interpretative act. The only way to draw our prejudices into view, he suggests, is by their provocation when a text addresses us in its strangeness or unintelligibility (ibid.: 198).

By extent, a proper hermeneutic attempt amounts to a fusion of horizons: a convergence of vantage points of reader and text. The locus of hermeneutics is a space of vacillation, an in-between (ibid.: 295). We are familiar with the object because it stands in a tradition, and yet its cultural or historical strangeness can never be suspended. In fact, 'the circle of whole and part is not dissolved

in perfect understanding but, on the contrary, is most fully realized' (ibid.: 293). The way in which we should familiarize ourselves with the 'other' should, then, not be to transplant it into an instantly recognizable horizon but to recognize its distance and confront ourselves with its strangeness, in the hope that we may 'understand in a different way, if we understand at all' (ibid.: 296).

The value of the dialogical perspective lies in this attempt to understand the other's perspective without wanting to reduce it to one's own or vice versa, and counters the idea that there is only one truth or one explanation. Whereas Gadamer's main focus is on the *historical* gap between text and interpreter, we aim to demonstrate that his theory is just as vital for bridging ideological or epistemological differences in a contemporary context. One question that boyd (2018) believes is valuable for teachers to explore with students is: 'Why do people from different worldviews interpret the same piece of information differently?' Updating the hermeneutic circle for digital humanities, we teach students to analyze online corpora in a circular motion that vacillates between the big data ('N=all') perspective of the whole, and a close reading of the part or the sample. In this article, we propose such a model on four different levels: contextual reading, distant reading, hyperreading, and close reading.

### **The Red Pill on Reddit**

Reddit is a web platform for social news aggregation, web content rating, and discussion. Its members can submit content such as text posts, pictures, or direct links, all of which are organized by distinct message boards, curated by interest communities, called 'subreddits'. Content is organized and hierarchically ordered based on user voting: links with the highest votes are shown at the top of the home page. As such, reddit serves as a platform for multiple, linked topical

discussion forums (Meraz, 2012), as well as a network for shared identity as a diverse online community (Papacharissi, 2010). Reddit's popularity since 2006 has skyrocketed: According to website ranking platform Alexa, in 2016 the platform ranked in the Top 10 most-visited sites in the United States and in the Top 50 most-visited sites globally (Ibid.: 57).

One such subreddit, The Red Pill (r/theredpill, henceforth TRP) contains a number of loosely-associated blogs on masculinity and personal philosophy for men. TRP can be seen as part of a specific community of practice (Eckert and McConnell-Ginet, 1992; Wenger, 1998) of Pick-Up Artists (PUAs). PUAs are males who seek to be successful at 'seducing' women, chiefly by how they manage their talk-in-interaction (Hambling-Jones and Merrison, 2012). TRP belongs to the online Manosphere, which includes groups such as pickup artists, father's rights activists, involuntary celibates ('incels'), and men going their own way (MGTOW). These groups share a belief system that Blais and Dupuis-Deri (2012) have called 'masculinism', which holds that society is ruled by feminine ideas and values, that this fact is repressed by feminists and politically correct 'social justice warriors,' and that men must protect themselves against a 'misandrist' culture (Marwick and Lewis, 2017). TRP is also connected to the Alt-Right, an alliance of far right, traditionalist Christian, disenfranchised 'geeks', and pickup artists (Kelly, 2017), which have been argued to find each other in a shared rejection of human equality in the philosophy of Identitarianism (Lyons, 2017).

TRP defines itself as a forum for the 'discussion of sexual strategy in a culture increasingly lacking a positive identity for men' (Watson, 2016). Its name is a reference to the 1999 film *The Matrix*: 'swallowing the pill,' in the community's parlance, denotes the acceptance of an alternative social framework, in which men, not women, have been socially disenfranchised in the



west. A conversion to the 'red pill' is experienced as an awakening to feminism's evils and 'misandry,' that allows the members to take charge of their own lives (Ging, 2017). At the time of writing, TRP is hosting around 300,000 users.

TRP has sparked controversy since its conception as it focuses heavily on anti-establishment perspectives on matters like abuse and rape. In response, reddit placed the subreddit in 'quarantine' in September 2018, a method intended for 'communities that, while not prohibited by the Content Policy, average redditors may nevertheless find highly offensive or upsetting'.<sup>1</sup> The procedure also targets communities that 'may be dedicated to promoting hoaxes ... that warrant additional scrutiny, as there are some things that are either verifiable or falsifiable and not seriously up for debate.' The argument of verifiability, to us, is especially relevant in the abovementioned context of epistemic uncertainty: the qualification of 'verifiability' is, after all, precisely the rhetorical strategy taken by many Men's Right's Activists (MRA), who often turn to facts and figures about things such as male suicide rates, workplace fatalities and high-risk jobs, and military conscription (CBC News, 2016). Instead of offsetting these facts with different ones, we argue that hermeneutic methods of the humanities are especially effective at revealing the inherent tensions within these epistemic frameworks.

In the last years, several studies have been published that attempt to (sub-)categorize the most important themes and topics in the Manosphere's discourses. Marwick and Caplan (2018) carry out a critical discourse analysis of the term 'misandry,' in order to understand how MRA use language to construct their identity as fundamentally oppositional to feminism and to reinforce binary gender systems. Gotell and Dutton (2016) perform a discourse analysis of popular MRA websites to reveal a set of claims (e.g. sexual and domestic violence are gender-neutral problems;

feminists are responsible for negating male experiences of victimization; rape culture is a moral panic spread by feminism). Ging (2017) identifies key categories and features of the manosphere and theorizes its masculinities based on these categories. Schmitz and Kazyak (2016) use open coding methodology, followed by closed coding, on fifty articles from twelve websites, examining the MRAs' online rhetoric with which they argue for the social superiority of men, and arguing that these rhetorical strategies constitute a backlash against feminism and gender equality. Moutford (2018) delves deeper into one of these websites, Return of Kings (returnofkings.com), using topic modeling to replicate the coded themes identified in this earlier study.

These studies offer a fruitful overview for further research in their labelling of the topics that the Manosphere engages with. However, they do not really analyze tone and style, especially the role of emotions. Most importantly for our purposes here, the aim of such studies seldom is to *understand* the authors on the discussion forum. They start from a preconceived and latent frame of condemnation that, we argue, may hinder an understanding of TRP's horizon or epistemology. They ask questions such as 'what strategies do online MRA groups utilize to delegitimize feminism and the goals of gender equality?' and 'are the sampled MRA groups antithetical to feminism and the goals of gender equality?' (Schmitz and Kazyak, 2016). Rather than writing an article with the purpose of critiquing TRP, we have chosen this community because it exemplifies the kind of environment that a specific group of people turn to in order to find out the 'truth' for themselves; to make sense of the world. Such forums facilitate the exploration of questions of gender, race, and other topics in radical ways: to ask the questions that mainstream media and schools will not allow. Several colleges in the United States and Canada, in fact, have men's rights groups represented on their campuses.

We argue that many studies on platform-based discourses seem to presuppose a certain logical consistency and uniformity, where we rather see a pluriform discursivity, and focus on revealing the tensions in TRP's discourse that destabilize logical consistency. While we can view this subreddit in broad terms as a 'discourse community' – a group of people sharing a set of discourses, understood as basic values and assumptions, and ways of communicating about those goals (Swales, 2011) – we need to emphasize the role of interpretation *within* this community, and the complex dynamics of drafting and negotiating knowledge. Assuming ideological structure should not override concerns for the disparate uptakes and discussions that take place within that structure. This is all the more pressing as it is the metonymical function of a subreddit that is typically addressed in regulatory practices – for instance, when Reddit administrators block or 'quarantine' certain communities. We by no means want to understate the potential societal ramifications of the discourses on platforms like this, yet we believe it crucial to render visible how users come to certain interpretations.

In doing so, we need to take into account the pervasiveness of irony and ironic detachment in contemporary (online) culture, which has been widely discussed. Ian Bogost writes about irony as ubiquitous response in contemporary times, especially online. He defines it as the 'escape from having to choose between earnestness and disdain', a protection mechanism against risk, disappointment or affliction (2016: 59). It is a response of fractality, fixing previous representations in a new voice while never truly engaging with any. In the words of philosopher Timothy Morton: 'Anything you can do, I can do meta' (2012: 88). Terms such as 'weaponised irony' and 'irony poisoning' capture the politically fraught consequence of this attitude, in which ironic content, when sufficiently shared, can become viewed as sincere.

## **Teaching Digital Hermeneutics in class**

We developed our materials on the basis of courses, modules and guest lectures we designed and taught in the last three years. Within the scope of our courses, we primarily made use of methods from Natural Language Processing (NLP). As a field, NLP aims to allow computers to process text and to identify meaningful subjects and associations. Its potentials to infer discursive regularity, topics, or sentiments from unstructured textual data are significant (Jacobi et al., 2016). In the classroom, we worked with Jupyter, an open-source web application that allows one to create and share Python documents (notebooks) that contain live code, equations, visualizations and narrative text.

There are three primary ways to contribute on reddit: post a submission, post a comment, or vote on a submission or comment. The corpus that was gathered, after removing empty fields, or fields shorter than 100 characters, includes 42,712 posts and 1,738,979 comments. Data was collected through reddit's application programming interface (API) that allows access to submission, comment and user data, making use of the 'timesearch' package<sup>ii</sup>. The data for this paper was gathered in September 2018. The final corpus consists of a total of 170,663,495 words. While this does not constitute 'big data' in the sense that it is not big enough to cause problems for typical computational methods, it is still large and unwieldy enough to warrant the use of such methods.

Within our classroom setting, we focus on four interpretative steps on different scales. Our scales and related tools are as follows<sup>iii</sup>:

1. **Contextual reading.** What is the contextual horizon against which we can understand the linguistic particularity of this social group? *Tool: tf and tf-idf*
2. **Distant reading.** How can we find posts and comments within our corpus, given its particularities as found in step 1, that show a significant degree of discursive tension? *Tools: Topic modeling; Word Embeddings*
3. **Hyperreading.** How can we read the posts and comments that are representative of the inherent discursive tensions for this particular social group? *Tool: Concordance views*
4. **Close reading.** How can we analyse inherent and internal tensions, conflicts, and irony that we found through the previous methods? *Close reading of telling case*

***Figure 1. Moving between close and distant reading***

The overall oscillation thus ranges from distant reading as a means of filtering relevant results for a particular explorative research question about the discursive specificity of this community, to a close reading of individual posts. Beyond the framing of an aggregate “public opinion” demonstrated by these individual voices (Entman and Herbst, 2001), we argue that close reading needs to include confrontation with strangeness. As such, it involves examining particularities of style as well as the role of emotion, ambiguity, and irony, all of which demonstrate inherent instabilities and tensions in the community. In short, we propose to move from ‘external’ context (step 1) via the ‘internal’ context of the corpus (2), to the ‘local’ context (3), to the singular text. The resulting close reading then leads to a new iteration of the circle, moving back to the totality

of the corpus. We now turn to a summary of these four scales, in order to demonstrate what kinds of knowledge can be gleaned from our proposed method.

### **Scale 1: Contextual reading**

As noted, TRP maintains a highly specialized language. A glossary can be found on the platform's front page, and it includes terms such as *alpha* ('Somebody who displays high value, or traits that are sexually attractive to women'), *hypergamy* ('The instinctual urge for women to seek out the best alpha available'), *SMV* (Sexual Market Value), *LMR* ('Last Minute Resistance'), and *LTR* ('Long-Term Relationship').<sup>iv</sup> As such, we might assume that computational stylistics can pick up on the discursive specificity of this subreddit, especially when compared to related communities. Students were asked to find subreddits that share concerns, topics or interests with TRP. In our contextual reading, TRP was compared with the contents of the following other subreddits:<sup>v</sup>

- *r/dating\_advice*: a community for exchanging dating tips and advice;
- *r/mgtow*: acronym for 'Men Going Their Own Way', part of the 'manosphere', cautioning men against serious romantic relationships with women, especially marriage;
- *r/seduction*: community for dating and pick-up artistry;
- *r/mensrights*: community that seeks to promote honest discourse in regards to male issues.

These were considered comparable in the sense that they deal with romance, dating, and relationships in general. Two of them, *r/mgtow* and *r/mensrights*, are communities within the Men's Rights movement. We use these comparative corpora as we want to find distinctive terms for TRP both within the Men's Rights movement, and in the context of dating. Since we are

interested in the difference between these corpora in terms of their topical content, we chose to lemmatize the corpora and filter them for nouns.<sup>vi</sup>

We then applied tf-idf, a simple and often-used technique to trace significant terms. Tf-idf stands for ‘Term Frequency—Inverse Data Frequency’, and is used in text analysis to find differences in textual corpora. In Gerbaudo’s (2016) terms, it is a form of *sample for top*, as it focuses on the messages that can be considered as the most visible or important within a collection. Term Frequency helpfully gives us the frequency of the word in each document in the corpus. It is the ratio of the number of times the word appears in a document compared to the total number of words in that document. It increases as the number of occurrences of that word within the document increases. Each document has its own tf. Inverse Data Frequency (idf), then, is used to calculate the weight of rare words across all documents in the corpus. The words that occur rarely in the corpus have a high idf score. It is given by the equation below. Combining these two we come up with the tf-idf score (w) for a word in a document in the corpus. Our ‘contextual reading’ consisted simply of calculating tf-idf frequencies for all our corpora, in order to reflect how important words are to each subreddit.

These scores allow us to measure how isolated or specialized a community is, by looking at the proportions of actual in-words. In doing so, we can formulate an alternative Red Pill glossary by gathering all the words that do not appear with any regularity in the other subreddits. After removing results that are not so much about topical content but rather forms of metadata (e.g. popular users such as ‘Archwinger’ or other subreddit names such as ‘mrp’ [married red pill]), some of the highest tf-idf scores are for the following words, alongside their tf-idf weights. We

also asked students to look for short explanations of the terms found in this step, the results of which are shown below.

1. bluepill 0.26 (remaining oblivious to social inequalities experienced by TRP members)
2. lmr 0.225 (last-minute resistance)
3. solipsism 0.211 (the view that the self is all that can be known to exist)
4. powertalk 0.127 (language used to get something instead of conveying information)
5. smp 0.126 (sexual marketplace)
6. daygame 0.123 (picking up women during the daytime)
7. preselection 0.113 (interest in men that other women are attracted to)
8. ioi 0.107 (indicators of Interest)
9. rsd 0.091 (dating coach company)
10. lssw 0.086 (Local Sexy Single Woman)
11. fux 0.073 (part of 'Alpha fux, Beta Bux')
12. asd 0.069 ('Anti-Slut Defense: A female defense mechanism to ensure that others can will not, or should not, label her as a slut)
13. wingman 0.065
14. alphaness 0.048 (the quality of being an alpha)
15. patriarch 0.047
16. carb 0.047
17. ldr 0.045 (Long-Distance Relationship)
18. unicorn 0.044 (Third woman in monogamous relationship)
19. dgaf 0.043 (Don't Give A Fuck)
20. bodyfat 0.043
21. bodyweight 0.043
22. jiu 0.042 (part of Brazilian Jiu Jitsu)
23. dhv 0.04 (Display of Higher Value)
24. amog 0.039 (Alpha Male of Group)
25. betaness 0.038

As we see, this step teases out a number of telling words and abbreviations, such as 'lmr', 'powertalk', 'smp', 'daygame', and 'asd'. These are all expressions of the specialized language on the subreddit. This language is further characterized by essentialist constructions, and forms part of the community's social and psychology theorems in which individual traits or personality are disregarded in favor of a general behavioral analysis of traits considered inherent to the female gender. The term 'solipsism', for instance, is defined in the subreddit as 'the female's tendency to frame everything she experiences or witnesses in terms of herself and her own needs –



*personalizing it* – even when such personalization would not make contextual sense.’ LMR, an often-used term in the pick-up community, stands for ‘last-minute resistance’, which is the moment a woman revokes her consent to sexual activity, usually just before the event.

Of course, these words would vary strongly if we used more of such comparative corpora, and doing so would improve the accuracy of this list. Yet, as many of these terms are highly specialized, and occur in the glossary for TRP, the method can be considered relatively successful. It produces valuable output that students can further investigate: for instance, one student chose to look into the preference of TRP members to engage in Brazilian Jiu Jitsu. Their conclusion was that the sport is instructive to a part of the community’s mindset: jiu jitsu, according to these members, has the physical and mental benefits of martial arts training, yet also has a very low injury risk. Their argument was that this approach demonstrates the ‘gamified’ ideology of these members in their dating practices, in which ‘min-maxing’ risk versus profit is key. The central value of this contextual reading, thus, is to take these terms into consideration as we move to our distant reading.

## **Scale 2: Distant Reading**

For our distant reading, and moving towards an ‘internal’ context of the singular text, students engaged in the common NLP approach of topic modeling, which enables us to trace semantic patterns in our corpus. In the terminology of Gerbaudo (2016), we can consider it a *sample for top* approach, as it yields themes and topics that are considered by the algorithm to be most relevant in the dataset. However, we should not stop there: our focus is especially on controversial topics, in order to analyze the contrasting perspectives by community members. One approach that some

students took before running their models was to sort the subreddit's posts based on their 'controversial' status, which reddit calculates based on the aggregated amount of *both* up- and downvotes for each post.

Topic modeling programs automatically extract topics from texts, taking a single text or corpus and searching for patterns in the use of words, attempting to inject semantic meaning into vocabulary (McCallum, 2002). The topic model that was built made use of Latent Dirichlet Allocation in the widely used Scikit-learn package for Python (Pedregosa et al., 2011). The assumption behind this machine learning technique is that documents consist of multiple topics, which are considered as 'hidden variables' that reflect the thematic structure of a collection. A topic model is built without semantic assumptions on the part of the researcher: the technique is 'unsupervised' and finds relationships between words without knowing what these words mean. This makes them relatively easy to build methods for exploratory textual research. Yet, as we have already built an interpretative frame of topics that seem relevant for this particular knowledge community, the output of these topic models cannot be viewed as in an interpretative vacuum. This is not as much about statistical verification as it is about interpreting the results of these findings in the light of those of Step 1, and taking into account *internal* contextual topics. To us, topics that are interesting are those that offer such context to the themes we have identified as particular to TRP.

We filtered our corpus for nouns in order to focus on the topical content of our corpus. As we are interested in the most-relevant *thread* (a post and its related comments), we also filter our dataset by joining the csv tables for posts and comments to only include those posts that include comments – resulting in a corpus with 61.967.597 tokens. Next, we need to determine how to

decide what amount of topics is ideal. In topic modeling, the number of assumed topics is chosen from the outset. Determining how many topics make for an accurate model can be tested by calculating coherence scores. The measure students used, called ‘c\_v’, indicate the relative distance between words within a topic (i.e., how often they appear together in documents we are analyzing).

The initial coherence score, on a topic model with 20 topics, was 0.41. As such, we taught students to implement what is called the ‘elbow method’: this basically means building a number of different topic models with different topic numbers, and comparing the coherence scores for each model. In doing so, we found the optimal amount of topics to be 40, with a coherence score of 0.4703.

***Figure 2. Coherence scores for TRP topic models***

Yet, the statistically ideal number of topics does not necessarily equate to the most productive topics for close analysis. Topic models offer a form of what Gadamer called the ‘fore-conception of completeness’. Reading such models always produces a revisable presupposition in the reader: that what is to be understood constitutes something that is understandable, something that is constituted as a coherent, and therefore meaningful, whole (Gadamer, 2004: 294). The different words in the topic are ought to make some intuitive sense with regards to the topics in the corpus. Students were then divided in groups, each of which was assigned one of the most coherent topic models (with 20, 30, and 40 topics, respectively), and were

asked to pick out the most interesting topics. These groups were then combined in two bigger groups, in order to discuss which of the selected topics were most interesting. Finally, the two bigger groups had the same discussion, until a choice was made about the most ideal topic model. Based on these discussions, we chose the model of 20 topics.

Then, as we are interested in the internal tensions of our community, the students' analyses had to remain limited to a particular topic whose contents open up productive avenues of interpretation. Students were tasked with exploring a single topic of their interest, examining the words and documents related to this topic. In the class discussions, the following 5 topics of our topic model were most popular. For the purpose of this example, we will look further into the "game" topic.

**Topic: dating strategy (1.4% of tokens)**

contact attention eye dog phone validation conversation voice facebook number text interest language face body talk person confidence look approach

**Topic: game (9.5% of tokens)**

beta alpha status smv male test game attraction money provider value behavior testosterone age personality confidence level wall fact female

**Topic: culture (7.7% of tokens)**

society feminism world power government system state war pill problem culture country gender equality fact part everyone value male reason

**Topic: rape (2.1% of tokens)**

rape law court case evidence police crime victim lawyer story person drunk culture judge fact state system claim definition word

**Topic: depression (3.2% of tokens)**

anger love phase depression post mind everything anxiety self pain year part person ego experience brain trp fuck feeling advice

### **Step 3: Hyperreading**

The third step is to read, in a more traditional fashion, sentences surrounding terms that we have found ('local context'). Through the topic model we have built, we can make a selection based on criteria that take the entire corpus into account. This relates to Gerbaudo's suggestion to engage in "data close reading" by reading posts as rows in a dataset (2016). The first step we engaged in was to find the thread, i.e. a post and its related comments, which is most distinctive for the 'game' topic that we were interested in exploring more closely. This topic includes words such as 'beta', 'alpha', 'status' and 'smv', and as such seems indicative of the community's focus on dating as a gamified practice in which social and psychological theorizing plays a large role. Students used concordances, a basic feature of the NLTK package, to skim sentences and compare and contrast the use of certain words.

Hyperreading is an umbrella term for non-linear, screen-based, and computer-assisted modes of reading, including search queries, skimming, and scanning (Sosnoski, 1999; Guillory, 2008). In scholarly reading, John Guillory (2008) distinguishes browsing (the open approach of the lay reader to informational texts), scanning (the more directed approach of the scholar to archival materials), and skimming, which he relates to a surplus of publications of scholarly books. Whereas scholars like Maryanne Wolf (2018; Wolf and Gottwalt, 2016) and N. Katherine Hayles (2012) have juxtaposed hyperreading to close reading and argued that the latter practice is deteriorating under the influence of digital media, we see hyperreading as a potentially creative mode of reading large bodies of online text, and as such it can be a useful pre-stage for close reading. Hyperreading allows us to intuitively and associatively trace our own interests and key words in a non-linear fashion across the dataset, and thus helps us identify passages that contain a

large number of meaningful features, and to select these for close reading. In Gerbaudo's terms, it can be considered a form of *random sampling*, as it involves concentrating on a particular section of the text that is found through a high-level (and not necessarily chronologically ordered) overview of all the instances of a term.

With the students in our classes, we practiced a computer-assisted form of hyperreading by making use of concordances. Concordancing offers quite a basic view of the dataset. While rarely used for analytical purposes, this mode of non-linear 'reading' a term across the database – what Gerbaudo calls 'reading posts as rows in a dataset' – offers an important perspective. For instance, in the example above we see 'beta' used in terms of the psychological theories produced by this community ('there are a shit ton of beta male providers', 'my current GF has a beta orbiter'). Students were asked to keep track of the recurring themes and concerns that pique their interest, and that they wished to track more closely in the next step. It is no longer about statistical evaluations of salience here, but about what we might call interpretative centroids – which is essentially what humanities practices of close reading entail. However, the concordance view allows them a greater reach, to see which word senses, connections, and contexts are relevant throughout a larger body of texts. Below is an example for the term 'beta', which we have been tracking throughout our reading.

***Figure 3. Concordances for 'beta' in TRP***

From tracing the word 'beta' across the corpus in its different lexical environments through our hyperreading assisted by concordances, it becomes clear that the word is always used in a negative

sense. Users express that they ‘used to be a beta’ before they took the red pill and started to see the error of their ways (‘I could sense that I was engaging in more beta behavior than I should have. But I’m so glad I have the red pill perspective’) or they proclaim they still are ‘beta’ and ask for advice on how to change this. Beta is connected to having to provide for a female who does not respect you: ‘I’m just some poor beta fag who takes care of her’; ‘beta provider who didnt realize his potential in time’; ‘Just like the cuckoo, these women have given it up to the tattoo artists, the drug dealers, the bad boys and now want a comfortable ride with an unsuspecting beta’. It means losing the control and power to the female: ‘stripped of his livelihood, his best years, by being a blue pill beta man and allowing the feminine imperative’.

Most importantly, the word beta in the majority of mentions is coupled to words describing unattractiveness: ‘Joe Average Beta Male is about as desirable to women as a cold bowl of oatmeal’. In order to change your status from beta to alpha male, it is imperative to stop trying so hard and being so nice: “‘You’re TOO NICE - Don’t let that side of you show”, i.e., stop being such a beta’; ‘just being away makes me seem less beta to her’; ‘women reward Beta Males (read: “nice guys”) with nothing’; ‘I can no longer believe that sensitivity, in almost any sense of the word, is appropriate behavior for a man if he wants to be attractive. The word is practically synonymous with beta’. The alpha/beta divide is a strategy to stabilize the messy and complicated world of dating and romance, and TRP members use it to create structure and order. Based on this hyperreading, their world views seem rather consistent. In order to discover internal tensions and fissures, close reading is indispensable. Hyperreading led us to a promising sample for close reading, as it helped us discover this passage with a high ‘density’ of the characteristic words that we here traced to their context in the dataset.

#### **Scale 4: Close Reading**

Arriving at the close reading, students were asked to select the thread in TRP that they deemed most representative of the topic ‘game’ (i.e., the documents that include the highest amount of words for a certain topic), based on their hyperreading of certain individual words within this topic - notably beta, alpha, game, female, attract\* and age(d). This led to the following ‘telling’ passage that was used for close reading (see Appendix A). It can be considered a form of *zoom-in sampling* (Gerbaudo, 2016), as it involves concentrating on a particular ‘passage’ in the collection that is considered particularly significant. This passage is in fact a set of comments responding to a post on TRP, by a woman who voices her grievances after an experience with a man of her own age who left her for a considerably younger woman. Interestingly, the section itself can be considered a report of an interpretation or even a collective close reading, as multiple users try to co-construct the ‘true’ or deeper meaning of what the woman has written: ‘What she *thinks* she is saying ..... What we *hear* her saying’ [our emphasis]. This is a great example of Anthony Giddens’ conception of ‘double hermeneutics’, which refers to the research approach of an interpretation of already existing interpretations, or the scholarly interpretation of lay conceptions (1984: 20). On the level of close reading, we analyze these comments in terms of rhetoric, voice of address, tone, imagery (metaphors, symbols), rhythm/meter, rhyme, structure, and tensions (ambiguities, paradoxes, irony, sarcasm).

When we look at writing style, there are some notable typos and grammar errors (*appearantly, those year, favourble, inconvienent, opinonated*, etc.). which create the impression that the comments have been written in a hurry. The tone is emotional and aggravated, with ample



use of swear words (*bich* [sic], *cunt*). This is reinforced by the question- and exclamation marks and abbreviations (*yo*, *gf*, *grrl*, *WTF?*) which are common shorthand used on online fora. As are interjections of descriptions of mimicry, lending the writing the right ‘tone’ to guide the reader’s interpretation: e.g. *\*Eyeroll\** to invoke indignation or sarcasm.

Other rhetorical characteristics include a lack of punctuation marks [*its*], and the run-on sentence:

*It’s not that women don’t have options but her pool of alphas and badboys has diminished and to think she has to settle for beta bob makes them sad which is were the bitching starts from they all knew this day would come but didn’t think it would come because of a favourble society and illogical reasoning lie in your beds cunts.*

Digressions like these, often used in postmodern literary writing for a satiric effect, invoke a sense of irrationality and emotion, or even the disintegration of a character’s logical reasoning (think of Roberto Bolaño’s *2666* or David Foster Wallace’s *Infinite Jest*). In the present context, a reader could construe this as ironic, considering the fact that the author precisely points out that women like the one he responds to use ‘illogical reasoning.’ Arguably the strongest and most revealing indicator of personal sentiment and involvement is to be found in the laughing outburst. Possibly intended to convey disbelief and a sense of detachment, the sheer amount of HA’s expresses quite the opposite effect. It effectuates a tension between the statement and the affect that is invoked. This made us wonder about the stakes these users have in disproving the female writer’s point: why are they so passionate about the topic? Such questions obviously go against the authorial intention.

Non-literal expressions include ‘beta bob’ which refers to a submissive male, one who is ‘providing resources or validation to others, women (and perhaps men).’ This term is juxtaposed to ‘alphas an bad boys,’ the socially dominant (TRP’s glossary defines ‘alpha’ as ‘Somebody who displays high value, or traits that are sexually attractive to women’). The alliterating b’s make for an expression with staying power, ready to be picked up and reiterated, thus creating new TRP vocabulary.

Other style figures include the rhetorical question.

*Why would I want to date a woman who is more volatile, more intimidating, more questioning, more pressing, more complex, more damaged, more opinionated and more womanly (matronly)? Intimidating, damaged, opinionated? Are these supposed to be virtues? ... WTF?*

Questions like these are obviously not meant to elicit answers, but rather to make a forceful point. The qualities listed are meant to be negative ones, and no well-thinking male would choose them over the other part of the binaries: ‘less’ is obviously more. Parallelism and repetition (‘more..., more... more...’; and later: ‘less... less..., less...’ ) further generate rhetorical momentum by effectuating rhythm.

There is a high degree of TRP’s specialized vocabulary at work in this text, marked by a tendency to refer to women in an objectifying way, like ‘best before date’ (the age by which a woman’s sexual ‘market value’ diminishes) and ‘to next’ (to break off all interaction and move on to the next female). This stylistic attribute of the corpus leads us to consider the logic in the community’s language use. It is a discourse that seems highly analytical and rationalistic in its discussion of a decisively *intersubjective* field of interactions (seduction, dating, relationships). It

is a discourse, moreover, that ‘gamifies’ the activity of dating by positing rule-based systems and procedures (like the aforementioned ‘tests’) that determine one’s degree of ‘success’ in the field of seduction and dating. Not coincidentally, in the community’s parlance, ‘the game’ typically refers to the act of ‘picking up’ strangers.

Among the imagery used is the metaphor *harpy*. In Greek and Roman mythology this connotes a female monster shaped like a vulture with a human head, who steals food from her victims’ mouths. In a more current context according to Urban Dictionary, the term signifies a ‘women [sic] who draws a man into her grasp by pleasing the victims biggest desire only to destroy all that makes him what he is.’<sup>vii</sup> Obviously, trying to ‘cage a harpy,’ as this poster extends the metaphor, would be a waste of one’s time. This metaphor seems puzzling when applied to the female author under consideration. After all, this is a woman who decidedly does *not* get what she wants. On the contrary, she gets abandoned because of her age. This is an example of what in psychology and media studies is called confirmation bias (Nickerson, 1998). Subconsciously wanting to justify their beliefs, believers then pay more attention to confirming rather than disconfirming evidence, and so strengthening their initial belief becomes a self-fulfilling prophecy. Internal evidence of reliability is typically interpreted based on prior trust. By interpreting the story of a woman who is on the losing end of the dating game, by still assigning to her the role of man-eating aggressor, the poster interprets her story in such a way that it fits into his pre-existing belief system. The image of the harpy is also striking as it completely opposes the overall tone and message of the rest of the post. The image of the vulture collides with the words used for female attributes in the rest of the post, which for the most part cast woman as an object: ‘There is no love, no fidelity, no connection, and no wife or pleasant womanly company just cheap pussy.’ The

writers seem to simultaneously view women as predators and objects, in a tension which is never resolved.

Whereas women are thus presented in an ambivalent way, men are seen as a homogenous category. The authors write on behalf of men (or TRP men, that is not specified) as a group: ‘What *we* hear her saying’ ... ‘*we* have more important things to do.’ (our emphasis). The first person plural indicates that they believe their opinions on these matters to be backed up by a community. As performative language, this creates a sense of belonging or even collective identity in a world without ‘love, ... fidelity, .. connection, [or] pleasant womanly company.’ In passing, TRP is thus called upon as an antidote to the hostile world that these posters invoke, in an imaginative, even literary way.

### **Conclusion: coming full circle**

In this article, we have aimed to develop strategies of reading and analysis that mediate between the familiar and the unfamiliar, part and whole, filter bubble and seemingly objective data. We proposed an educational model on four different levels: contextual reading, distant reading, hyperreading, and close reading, describing a circular oscillation between part and whole of the dataset.

On the first scale, ‘contextual reading,’ we determined the discursive specificity of the subreddit under study by comparing it to four related subreddits on romance, dating, and relationships. Using tf-idf, we were able to reflect on how important words are to each subreddit. This gave us a number of telling words and abbreviations that are expressions of TRP’s specialized

language, and in which our students identified essentialist constructions. Students were able to unearth a ‘gamified’ ideology, which centralizes ‘min-maxing’ risk versus profit.

For our distant reading, and moving towards an ‘internal’ context of the singular text, students engaged in two common NLP approaches, topic modeling and word embeddings, in order to trace semantic patterns in our corpus. In groups, students were assigned the most coherent of the resulting topic models, and picked out the most promising one, consisting of 20 topics. The analysis was limited to a particular topic whose contents open up productive avenues of interpretation

Hyperreading allowed us to read the sentences that surrounded the terms that we found on the previous scale: to trace the words back to their ‘local context’. The first step we engaged in was to find the thread, i.e. a post and its related comments, which is most distinctive for the ‘game’ topic that we have chosen to explore. We then used concordances to skim sentences and compare and contrast the use of certain words. With our students, we practiced a computer-assisted form of hyperreading by making use of concordances. The concordance view allows them a greater reach, to see which word senses, connections, and contexts are relevant throughout a larger body of texts. allowed us to intuitively and associatively trace our interests and key words in a non-linear fashion across the dataset, and thus to identify passages that contain a large number of meaningful features. Such a passage was then selected for close reading.

From our close reading of the ‘telling’ sample identified in the hyperreading, the creation of a world that is hostile to men came to the fore, as well as certain tensions in the way women were presented both as objects and subjects, victims and predators. There was also proof of a certain form of confirmation bias in the posters, who were primed to read the original post in light

of this pre-existing logic whereby women mistreat men by definition, and that they deserve it when they are discarded for a younger lover. On the scale of close reading, students thus undertook important steps toward reconstructing the horizon of some members of this community, and entering into a dialogue with their viewpoints. Then, in order to come full circle and emphasize that the work of interpretation is never complete, we asked them to pick another topic guided by the outcome of their close reading, and let their new insights frame the new cycle of interpretation. After having analyzed how the topic ‘game’ was operationalized, they could for instance understand other topics such as ‘depression’ in light of these insights.

This approach, we hope, can be a modest answer to dana boyd’s assertion that the predicaments marking the digital public sphere, outlined at the beginning of this article, require more than practices of labeling and fact checking. They necessitate a cultural change in ‘how we make sense of information, whom we trust, and how we understand our own role in grappling with information’ (boyd, 2017). We need to teach students how others think and where the differences in perspective lie, what constitutes different horizons. From an educational point of view, this means building the capacity to see and understand someone else’s perspective while holding our own. We believe that updating dialogical hermeneutics for digital humanities offers a much-needed intersubjective potential that puts weight at attempts to understand others’ viewpoints.

## References

- [Allan 2016] Allan, J.A. “Phallic Affect, or Why Men’s Rights Activists Have Feelings,” *Men and Masculinities* 19.1 (2016): 22-41.
- [Allan 1972] Barthes, Roland. *Mythologies*. Noonday, New York (1972).  
<https://doi.org/10.2307/40099594>.
- [Bialik and Matsa 2017] Bialik, K, and K.E. Matsa, “Key trends in social and digital news media.” *Pew Research Center* (2017): 4-10.
- [Blais and Dupuis-Déri] Blais, M. and F. Dupuis-Déri, “Masculinism and the Antifeminist Countermovement.” *Social Movement Studies* 11.1 (2012): 21–39.
- [Bogost 2016] Bogost, I. *Play Anything*. Basic Books, New York City (2016).
- [boyd 2017] boyd, d. “Did Media Literacy Backfire?” *Data and Society* (2017). Available at: <https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>
- [boyd 2018] boyd, d. “You Think You Want Media Literacy... Do You?” *Data and Society* (2018). Available at <https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>
- [Capurro 1992] Capurro, R. “Informatics and Hermeneutics,” *Software Development and Reality Construction*. Springer-Verlag, Berlin (1992).
- [Capurro 2010] Capurro, R. “Digital Hermeneutics: An Outline,” *AI and Society* 25 (2010): 35–42. <https://doi.org/10.1007/s00146-009-0255-9>.
- [CBC News 2016] CBC News (2016). Mayfair Theatre cancels showing of men’s rights documentary The Red Pill | CBC News. Available at: <https://www.cbc.ca/news/canada/ottawa/the-red-pill-mayfair-theatre-ottawa-screening-cancelled-1.3876947>
- [Chen, Conroy, and Rubin 2015] Chen, N., Conroy, J., and V. L. Rubin. “Misleading Online Content : Recognizing Clickbait as ‘ False News .’” In *WMDD’15*, 15–19 (2015).
- [Collobert 2011] Collobert, R., Weston, J., Bottou, L., Karlen, M., et al., “Natural Language Processing (almost) from Scratch,” *Journal of machine learning research* 12 (2011): 2493–2537.
- [Danks and London 2017] Danks, D. and A. J. London, “Algorithmic Bias in Autonomous Systems: A Taxonomy of Algorithmic Bias,” *26th International Joint Conference on Artificial Intelligence* (2017) 4691–4697. DOI: 10.24963/ijcai.2017/654.
- [Eckert and McConnell-Ginnet 1992] Eckert, P. and S. McConnell-Ginet. “Communities of practice,” *Locating Power, Proceedings of the 1992 Berkeley Women and Language Conference* Berkeley, USA (1992): 89–99.
- [Entmann and Herbst 2001] Entman, R. M. and S. Herbst, “Reframing public opinion as we have known it,” *Mediated politics: Communication in the Future of Democracy*. Cambridge University Press, Cambridge (2001): 203-225.
- [Floyd et al. 1992] Floyd, C., R. Budde, H. Züllighoven, and R. Keil-Slawik. *Software Development and Reality Construction*. Springer-Verlag, Berlin (1992).
- [Gadamer 2004] Gadamer, H.G. *Truth and Method*, 2nd ed. Continuum, London and New York (2004 [orig. 1960]).

- [Gallup and Knight 2018] Gallup and Knight, “American Views: Trust, Media and Democracy,” *Knight Foundation Report* (2018). Available at: <https://knightfoundation.org/reports/american-views-trust-media-and-democracy>
- [Gerbaudo 2016] Gerbaudo, P. “From Data Analytics to Data Hermeneutics: The Continuing Relevance of Interpretive Approaches,” *Digital Culture and Society* 2.2 (2016). <https://doi.org/10.14361/dcs-2016-0207>.
- [Giddens 1984] Giddens, A. *The constitution of society: Outline of the theory of Structuration*. University of California Press, Berkeley (1984).
- [Gill 2007] Gill, R. “Postfeminist media culture. Elements of a sensibility.” *European Journal of Cultural Studies* 10.2 (2007): 147-166.
- [Ging 2017] Ging, D. “Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere,” *Men and Masculinities* 20.10 (2017): 1-20.
- [Guillory 2008] Guillory, J. “How Scholars Read,” *ADE Bulletin* 146 (2008): 8–17.
- [Hambling-Jones and Merrison 2012] Hambling-Jones, O., and A. J. Merrison, “Inequity in the Pursuit of Intimacy: An Analysis of British Pick-up Artist Interactions,” *Journal of Pragmatics* 44.9 (2012): 1115–27. DOI: 10.1016/j.pragma.2012.04.008.
- [Hayles 2012] Hayles, N. K. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: The University of Chicago Press (2012).
- [Ihde 1998] Ihde, D. *Expanding Hermeneutics: Visualism in Science*. Evanston: Northeastern University Press (1998).
- [Jacobi, Atteveldt and Welbers 2016] Jacobi, C., Van Atteveldt, W. and Welbers, K. “Quantitative analysis of large amounts of journalistic texts using topic modelling,” *Digital Journalism* 4.1 (2016): 89–106. DOI: 10.1080/21670811.2015.1093271.
- [Jockers 2013] Jockers, M. “‘Secret’ Recipe for Topic Modeling Themes.” (2013). Available at <http://www.matthewjockers.net/2013/04/12/secret-recipe-for-topic-modeling-themes/>
- [Latour 2005] Latour, B. *Reassembling the Social An Introduction to Actor-Network Theory*. Oxford University Press, Oxford (2005).
- [Lynch 2017] Lynch, M. “Post-Truth, Alt-Facts, and Asymmetric Controversies,” *First 100 Days* (2017). Available at <http://first100days.stsprogram.org/2017/02/06/post-truth-alt-facts-and-asymmetric-controversies-part-i/>
- [Lyon 2017] Lyon, M. “Ctrl-Alt-Delete: The Origins and Ideology of the Alternative Right,” *Political Research* 20.1 (2017). Available at <https://www.politicalresearch.org/2017/01/20/ctrl-alt-delete-report-on-the-alternative-right>
- [Mallery, Hurwitz, and Duffy 1986] Mallery, J. C., R. Hurwitz, and G. Duffy, “Hermeneutics : From Textual Explication to Computer Understanding?” *A.I. Memo*, no. 871 (1986).
- [Massanari 2015] Massanari, A. L. *Participatory Culture, Community, and Play: Learning from Reddit*. New York: Peter Lang, 2015.
- [Marwick and boyd 2011] Marwick, A., and d. boyd, “I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience,” *New Media and Society* 13:1 (2011): 114–133. DOI: 10.1177/1461444810365313.
- [Marwick and Caplan 2018] Marwick, A., and R. Caplan, “Drinking male tears: language, the manosphere, and networked harassment,” *Feminist Media Studies* 18.4 (2018): 543-559.



- [Marwick and Lewis 2017] Marwick, A., and R. Lewis, "Media Manipulation and Disinformation Online," *Data and Society* (2017). Available at: [https://datasociety.net/pubs/oh/DataAndSociety\\_MediaManipulationAndDisinformationOnline.pdf](https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf)
- [McCallum 2002] K. McCallum, A. *Mallet: A Machine Learning for Language Toolkit*, 2002. Available online at: <http://mallet.cs.umass.edu>
- [Mikolov et al. 2013] Mikolov, T., Sutskever, I. Chen, K., Corrado, G., and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems* (2013): 3111-3119.
- [Morton 2012] Morton, T. *Realist Magic*. University of Michigan/Open Humanities Press, Ann Arbor (2012).
- [Moutford 2018] Moutford, J. "Topic Modeling the Red Pill," *Social Sciences* 7.42 (2018): 1-16.
- [Nagle 2017] Nagle, A. *Kill all normies. Online Culture wars from 4chan and Tumblr to Trump and the Alt-right*. Zero Books, Winchester, UK and Washington (2017).
- [Pedregosa et al. 2011] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion et al., "Scikit-learn: Machine Learning in Python," *JMLR* 12 (2011): 2825-2830.
- [Ramsay 2010] Ramsay, S. "The Hermeneutics of Screwing Around; or What You Do with a Million Books," In *Playing with Technology in History Conference*, Niagara-on-the-Lake, 1-9 (2010).
- [Reddit 2018] Reddit, "Quarantined Subreddits" (2018). Available at: <https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/quarantined-subreddits>
- [Romele, Severom Furia 2018] Romele, A., M. Severo, and P. Furia, "Digital Hermeneutics: From Interpreting with Machines to Interpretational Machines," *AI and Society: Knowledge, Culture and Communication*. Springer, New York (2018).
- [Sahlgren, Holst, and Kanerva] Sahlgren, M., A. Holst and P. Kanerva, "Permutations as a means to encode order in word space," *The 30th Annual Meeting of the Cognitive Science Society*, Washington DC, USA (2008).
- [Schmitz and Kazyak] Schmitz R. M., and E. Kazyak, "Masculinities in Cyberspace: An Analysis of Portrayals of Manhood in Men's Rights Activist Websites," *Social Sciences* 5.18 (2016): 1-16.
- [Sosnoski 1999] Sosnoski, J. J. "Hyper-readers and their reading engines." *Passions, Pedagogies, and Twenty-First Century Technologies* (1999): 161-77.
- [Swales 1993] Swales, J. M. "Discourse community and the evaluation of written text," *Georgetown University Round Table on Languages and Linguistics (GURT) 1992: Language, Communication, and Social Meaning* (1993): 316.
- [Viner 2016] Viner, K. "The Post-Truth: How Technology Disrupted the Truth." *The Guardian* (2016).. <https://www.theguardian.com/media/2016/jul/12/how-technology-disrupted-the-truth>
- [Watson 1998] Watson, Z. "Red Pill Men and Women, Reddit, And The Cult of Gender," *Inverse* (2016). Available at: <https://www.inverse.com/article/15832-red-pill-men-and-women-reddit-and-the-cult-of-gender>

- [Wenger 1998] Wenger, E. "Communities of practice: Learning as a social system," *Systems Thinker* 9.5 (1998): 2-3.
- [Wolf 2016] Wolf, M. *Tales of literacy for the 21st century*. Oxford University Press, Oxford (2016).
- [Wolf 2018] Wolf, M. *Reader, come home: The reading brain in a digital world*. Harper, New York City (2018).



## Endnotes

---

<sup>i</sup> Due to what reddit flags as ‘shocking or highly offensive content’, there are a number of ethical issues to consider when spending time in the classroom working with datasets such as these. In particular, we need to deal with the question whether suspending judgement about these kinds of provocative and misogynous communities mean to normalize them. These issues were regularly discussed in class, with several students opting to devote their attention to different and less offensive subreddits.

<sup>ii</sup> Available at <https://github.com/voussoir/timesearch> (accessed 28 September 2019)

<sup>iii</sup> All Jupyter Notebooks used for our approach are available via GitHub: see [REDACTED]

<sup>iv</sup> See [https://www.reddit.com/r/TheRedPill/comments/2zckqu/updated\\_glossary\\_of\\_terms\\_and\\_acronyms](https://www.reddit.com/r/TheRedPill/comments/2zckqu/updated_glossary_of_terms_and_acronyms) (accessed 19 September 2019)

<sup>v</sup> Definitions for these subreddits were taken from their respective “About” sections.

<sup>vi</sup> Word counts for these corpora after filtering are as follows: r/theredpill contains 3,964,468 words, r/seduction contains 4,932,148 words, r/mgtow contains 2,733,560 words, and r/mensrights contains 1,785,502 words.

<sup>vii</sup> ‘Harpy.’ Urban Dictionary. <http://www.urbandictionary.com/define.php?term=harpy>. Consulted 21-04-2017.