Abstract

This paper uses stylometry – computational analysis of writing style – to explore how the Trollope family's personal and professional interactions with one another may have influenced their writing styles. In particular, works by Frances Milton, Frances Eleanor, Thomas Adolphus, and Anthony (referred to collectively as 'the authorial Trollopes'), as well as some works by Charles Dickens, comprise the corpus. Dickens, initially used as a control in the analysis, in actuality emerged either as a potential influence on the authorial Trollopes given his stringent editorial practices, or as being influenced by the Trollopes himself. Working from the underlying assumption that authorial writing style reflects individuality and may change in accordance with interpersonal relationships as well as writerly expectations of genre, this paper shows that the authorial Trollopes' writing styles, while still distinct, are at times remarkably similar to one another. In doing so, this paper highlights the literary and historical value of those less renowned Trollopes, supporting more extensive study of family members other than Anthony. More broadly, this paper is a concept study to demonstrate the value of stylometry to Victorianists studying textual networks.

The software used to conduct this research was the R package 'stylo'.

Keywords

Acknowledgements

Tangling and Untangling the Trollopes

A Stylometric Analysis of Frances Milton Trollope, Frances Eleanor Trollope, Anthony Trollope, Thomas Adolphus Trollope, and Charles Dickens

Introduction

'Trollope' is one of the first names that comes to mind when thinking of Victorian literature. Though it is now most commonly associated with Anthony, the name Trollope was first introduced to the public through his mother, Frances Milton Trollope (1780-1863). She turned to writing to support her family after her husband's questionable financial decisions left them penniless. Despite beginning her authorial career considerably later in life than Anthony, she still managed to produce 34 novels and numerous works of non-fiction. Biographer Victoria Glendenning (xi) underlines Frances Milton's celebrity, writing that although Anthony has since overshadowed her, "his mother was the famous one in the family … *seriously* famous." Yet Frances Milton's outspokenness on matters of gender and politics courted a large number of highly critical responses. One contemporary critic (Blanchard 417) commented that "no other author of the present day has been at once so much read, so much admired, and so much abused." While she is not so well known today, Frances Milton was familiar to many Victorian readers.

Nearly all of Frances Milton's children inherited her authorial talent. Of the four who reached adulthood, only Henry (1811-1834) went unpublished. Frances Milton's eldest, Thomas Adolphus (1810-1892), published novels, travel writing and foreign correspondence, and historical texts. His sister Cecilia (1816-1849) published one novel.[i] The most famous of Frances Milton's children, Anthony, published novels, travel writing, and histories. Moreover, such authorial talent also extended to partners and children. Thomas' first wife, Theodosia Garrow Trollope (1816-1865), was a respected poet and journalist; his second wife, Frances Eleanor Trollope (1835-1913), wrote novels – four of which were serialized in

Dickens's *All the Year Round* – as well as works of non-fiction that included a biography of her mother-in-law. The final Trollope publishing in the nineteenth century was Anthony's son, Henry Merivale Trollope (1846-1926).[ii]

This paper focuses on the works of Frances Milton, Frances Eleanor, Thomas Adolphus, and Anthony (referred to collectively as 'the authorial Trollopes'). These Trollopes were personally and professionally connected in various ways. There were, for example, acknowledged authorial influences throughout the Trollope family. Anthony notes in his autobiography (79-80) that the publication of his first novel stemmed from his mother's established relationship with reputable publishers. What is more, a notice in an 1887 issue of *The New York Times* (3) observes that Anthony Trollope was "indebted to 'T. A.' for the plot of 'Dr. Thorne,' one of his most successful stories." Anthony also addresses this indebtedness in his autobiography (154) when he writes that "it was the only occasion in which I have had recourse to some other source than my own brains for the thread of a story." Another 1896 article ("MRS. TROLLOPE." 489) reviewing Frances Eleanor's biography of her mother-in-law notes that Frances Milton "was not above taking a suggestion for a story from her eldest son. Thomas Adolphus Trollope gave her the plot and the title of one of her most successful novels, 'Petticoat Government.'" In turn, it has been argued that *Petticoat Government* was a key influence for Anthony Trollope's *Chronicles of Barchester* series (Sadleir 157). Thomas Adolphus and his mother had an especially close personal connection, and often close physical proximities resulting from prolonged co-travel to international destinations such as Italy. In his autobiography, Thomas Adolphus (*What I Remember, Volume II* 357) describes the death of his mother, following her gradual loss of memory and physical deterioration, as one of "the two greatest sorrows I had ever known" (the other being the death of his first wife). "It is very common for a mother and daughter to live during many years of life together in as close companionship as I lived with my mother, but it is not common for a son

to do so," he explains (*What I Remember, Volume II* 328-9). "During many years, and many, many journeyings, and more *tête-à-tête* walks, and yet more of *tête-à-tête* home hours, we were inseparable companions and friends." The multi-layered interconnectedness of the authorial Trollopes prompted us to question whether that interconnectedness influenced their individual writing styles. More broadly, we questioned if family could function as a category for thinking about authorial style. While being related does not promise stylistic similarity, close relationships like those between the authorial Trollopes may encourage textual coproduction, however subconsciously and implicit.[iii]

The authorial Trollopes clearly engaged in influential interactions with one another that impacted the stylistic elements of their works. In this paper, we present a stylometric analysis of the authorial Trollopes' published texts to explore the extent of their relationships' impact on their authorial styles. We do not dispute the authorship of any of the works included in our corpus, but instead draw attention to the vast potential of using computational tools for greater understanding of authorial networks and interpersonal relationships that may influence text production. Anthony was far from the only prolific Trollope of his day, and we believe that the other authorial Trollopes are worthy of heightened attention. More broadly, though, this paper is a concept study to demonstrate the value of stylometry to Victorianists considering textual networks. The methods used for this study may be applied to virtually any other textual corpus, and we encourage curious scholars to adapt our methods to suit their own corpora. In particular, we encourage the study of other authorial families and networks known to be interconnected, including the Brontës, Marryats, and Lyttons. Though this paper is an exercise in distant reading, its findings or those of a similar study could be used to direct future analyses that use close reading methods.

Stylometry and Computational Literary Analysis

Stylometry refers to the practice of distinguishing a writer's authorial fingerprint by applying a quantitative approach to literary analysis. Although stylometry predates computational technologies, modern stylometry is almost always executed through computational analyses of writing style: that is, the author's conscious and subconscious use of words and syntactical structures. David Holmes traces the origins of stylometry to Augustus de Morgan's observation that some authors prefer using longer words than others in 1851. However, as Adam Pawłowski and Artur Pacewicz note, it was Wincenty Lutosławski who coined and defined the term 'stylometry' in the late nineteenth century while chronologizing Plato's writings. Some other notable stylometric analyses include those conducted on the American Federalist Papers in the 1960s (Mosteller and Wallace), the works of Jane Austen in the 1980s (Burrows, *Computation*), and the twelfth-century Latin works of Hildegard of Bingen in the 2010s (Kestemont, Moens, and Deploige). Stylometry is an effective method for considering linguistic relationships between individuals because it focuses not on the semantic content of individual works but word usage patterns across works. Stylometry may illuminate channels of editorial and personal influence otherwise undetected by readers. It may also support both close readings of primary texts and historical studies of authorial networks. Stylometry is especially useful for a large corpus like that of the authorial Trollopes; computational stylometry allows for quick identification of potential similarities between works, facilitating targeted close reading and interpersonal network analysis.

Of particular importance to a stylometric analysis is an author's use of function words, which are those words that serve grammatical purposes rather than serving as informative content in themselves (the latter are, appropriately, called 'content words'). Function words include such parts of speech as articles (the, a), conjunctions (and, or, but,

so), and pronouns (I, he, his, she, her). Because function words do not typically establish narratives, their importance is often overlooked. There is, however, a profound difference between 'John gave Mary *a* book' and 'John gave Mary *the* book'. The latter implies that John is giving Mary a particular book, and that the reader likely has some knowledge of that work, while the former assumes no knowledge. In his *Secret Life of Pronouns*, James Pennebaker argues that one's use of function words characterizes one's writing style and, more generally, reflects one's personal disposition and lived experience. The assumption that function words distinguish a writer's personal 'style' underpins this study.

Stylometry could be considered one means for what Franco Moretti calls 'distant reading': computational analysis of substantial textual corpora, as opposed to close reading of only a few works. In his 2013 book about distant reading, Moretti (*Distant Reading* 48-49) explains the distinctive value of distant reading, particularly for studies of noncanonical literature:

> [T]he trouble with close reading (in all of its incarnations, from the new criticism to deconstruction) is that it necessarily depends on an extremely small canon. This may have become an unconscious and invisible premise by now, but it is an iron one nonetheless: you invest so much in individual texts *only* if you think that very few of them really matter. Otherwise, it doesn't make sense. And if you want to look beyond the canon … close reading will not do it. It's not designed to do it, it's designed to do the opposite. At bottom, it's a theological exercise – very solemn treatment of very few texts taken very seriously – whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how *not* to read them. Distant reading: where distance, let me repeat it, *is a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems.

Distant reading – or, as Moretti puts it, *not* reading – offers an alternative, and often complementary, textual viewpoint to close reading. Indeed, computational tools may identify recurrent or unique features that go otherwise unnoticed by the human reader. Once these features are identified, though, it is up to the researcher to analyse them (Heuser and Le-Khac). Computers can tell us *what* is happening; we must be the ones to discern *why*. Recent

criticism of distant reading – and computational literary analysis more broadly – has focused on humanities scholars' misinterpretations of statistical data (Da) and unchecked enthusiasm for technological 'hype' (Kirsch), and has even argued against the use of computational tools for textual study altogether (Marche). We hold, however, that statistical data are never evidential ends, but flags that draw attention to potentially overgrown trails worth exploring. As Michael Stubbs writes in a short introduction to computational methods for literary analysis, "[s]imply identifying quantitative features of text does not lead automatically to results of literary interest, since there are always nonlinguistic factors, historical, cultural and psychological, and since there is always an intuitive leap from objective textual facts to subjective literary interpretation" (61). This is to say, computational literary analysis may offer and facilitate fresh perspectives on age-old questions: perspectives that need not discredit previous work, but build upon it using new technologies. Using computational methods, we may check our claims, lend credence to conclusions, or establish new avenues for research altogether. The applications of computational methods are vast, as exemplified by the broad reach of projects by such organizations as the Stanford Literary Lab ("Stanford Literary Lab"), led by Mark Algee-Hewitt, a scholar of eighteenth-century literature. Victorianists have also embraced distant reading for fresh perspectives of popular nineteenth-century authors. Studies have reviewed uses of repetition (Gemma, Glorieux, and Ganascia), gender representations in publishing (Bode), and poetry publication and style (Houston), to list only a few. Genie Babb even argues, citing John Theodore Merz, that the methodologies emerging in the "statistical [nineteenth] century" established the basis for the methodologies of today's computational text analysis (Babb). Pioneer of computational stylometry John Burrows initially developed computational methods of authorship attribution and style scrutiny using seventeenth-century texts, but has since applied these methods to nineteenth-century texts by such authors as Jane Austen (*Computation*) and Henry James ("Questions of

Authorship"). The Centre for Literary and Linguistic Computing (CLLC) at Australia's University of Newcastle, founded by Burrows in 1989, has conducted numerous studies of Victorian periodicals following from Burrows' work. Victorian periodicals are well suited to computational analysis both because they are in the public domain and because of the frequency with which they included anonymous authors – authors whose names might still be revealed. Further, projects like *The Wellesley Index*, *Dickens Journals Online*, and *The Periodical Trollopes* provide valuable indices of authorial attributions and accessibly formatted periodical texts. It is worth noting, though, that there are still many periodicals that are not found in such indices and even more texts that have not yet been digitized (Leary; Nicholson). We have worked with what is currently available, but are optimistic that continued digitization efforts will amplify the accuracy and value of computational methods of literary analysis.

Corpus and Methodology

Our corpus comprised 201 public-domain text files with works attributed to Anthony (56), Frances Eleanor (26), Frances Milton (52), Thomas Adolphus (46), and Charles Dickens (21). Dickens can be considered a member of the authorial Trollopes' larger network, with his relationship with Frances Eleanor being especially noteworthy. Frances Eleanor's first works were published in *All the Year Round* under Dickens' editorship. The two were also personally acquainted through Frances Eleanor's sister Ellen Ternan, who is widely considered to have been Dickens' mistress (Slater). Although Dickens was initially included as a control author, we opted to include him in the below discussion about authorial interplay given his strong influence on the Trollopes' authorial experiences, as well as the number of previous analyses concerning his work (e.g. "CLiC Dickens").

We used Wilkie Collins, Margaret Oliphant, and Walter Scott as controls to confirm the accuracy of our methods (outlined below); each of these control authors was represented by 21 public-domain texts. Scott, Collins, and Oliphant were selected as control authors as we believe their textual output – comprising novels, journalism, and book-length works of non-fiction – to be similar to those of our subjects. Moreover, the periods in which these control authors worked correlate with those of the authorial Trollopes. Periodization is important due to differences in style that scholars have identified as chronological or generational trends (Hewitt; Broughton and Kingstone). Franco Moretti, for one, traces stylistic trends in dominant – "hegemonic" – genres, finding that "with the exception of the turbulence of 1790-1810, a rather regular changing of the guard takes place, where half a dozen genres quickly leave the scene, as many move in, and then remain in place for twenty-five years or so" (*Graphs, Maps, Trees* 18). We used control authors working in roughly the same periods as the authorial Trollopes to mitigate the possibility of period-based stylistic trends affecting results. Though this study does not explicitly address questions of generational influence, such questions could drive future studies within which the methods used here may be applied.

The selection of the authorial Trollopes as the primary subjects for this study was motivated by a belief that their relationships with one another may have informed the development of their writing styles. More specifically, this study emerged from contemplation of how Thomas Adolphus' editorial work for his mother might have influenced either writer's style or, conspiratorially, whether Thomas Adolphus may have claimed authorship of his mother's unpublished works following her death. Yet computational analysis is limited in its ability to determine the extent of interpersonal influence not least because, as Robert Douglas-Fairhurst (8) notes, not all influences are strictly literary. Douglas-Fairhurst considers existential concerns that might have informed

Tennyson's poetry; we might ponder the influence of shared anecdotes and conversations on our subjects' works. Indeed, a future study comparing authors' non-literary writing (e.g. letters and journals) to their novels might further understanding of intertextual influence. One such study, led by one of this paper's authors, is currently underway (Dumbill, "Fiction").

Our texts were acquired from two online textual archives: Project Gutenberg (www.gutenberg.org) and the Internet Archive (www.archive.org). Paratextual material – e.g. transcription notices, tables of contents, indices, and appendices – was removed. Within the bodies of the texts, quotations and passages in languages other than English were retained as such sections were not believed to be so prominent in any of the corpus texts to strongly influence result accuracy. Text bodies remained wholly untouched from their transcribed or transposed states. We opted to maintain volumization where possible, as volumes were published separately and could therefore (however unlikely) be subject to stylistic differences. For this reason, one 'book' may actually be represented in multiple files that reflect independently-published volumes. A complete list of texts comprising our corpus, including links to the texts we used, is available online (citation removed for peer review).

To conduct a general stylometric analysis of the authorial Trollopes' writing styles, we used the R package 'stylo', which was developed specifically for such uses. stylo is a powerful package developed for the R programming environment by members of the Computational Stylistics Group: Maciej Eder, Jan Rybicki, and Mike Kestemont. It is a suite of tools to facilitate stylistic analyses and investigate authorship attribution claims through the use of visualizations generated through supervised and unsupervised machine learning methods. A comprehensive list of stylo's functions is available on the online R Archive Network (Eder, Rybicki, and Kestemont, "Package 'stylo'"). The function used for this study was stylo(). We offer an explanation of the ways this function was used for this study so that others may adapt our methods for their own research. The stylo package is a particularly user-

friendly and powerful option for stylometric analysis, and we have chosen to use it here so that readers may be inspired to conduct their own such studies on other authors.

The stylo() function is, as one may expect, the main function of the stylo package. It is not the place of this paper to review its functionality in depth; even the package's developers recognize the function's complexity. The Computational Stylistics Group (Eder, Rybicki, and Kestemont, "'Stylo'") describes stylo() as a general tool for performing "a variety of stylometric analyses from multivariate statistics to assess and visualize stylistic similarities between input texts," which depend upon a most-frequent-word (MFW) list for the entire corpus. Using the MFW list, stylo() applies a range of statistical procedures (cluster analysis, multidimensional scaling, or principal components analysis) to calculate 'distances' between texts: the 'closer' the two texts, the more alike their word usages. We have named each text file in the corpus using a categorical modifier (e.g. 'fetrollope' to refer to Frances Eleanor), followed by an underscore and abbreviated title (e.g. 'charming1' to refer to the first volume of *A Charming Fellow*). stylo assigns a colour to each modifier so that users may more easily distinguish authorial clusters. Note that stylo does not readily permit colour customization.

As more skeptical readers might already have surmised, there are numerous limitations to this study that should be noted. The clearest is the corpus sources themselves: Project Gutenberg and the Internet Archive. Scholars of Victorian literature are fortunate to have a wealth of digitized texts freely available, and these texts may be subject to computational scrutiny leading to new perspectives about them. The digitized versions of these texts, though, are of mixed quality. Project Gutenberg is a laudable effort to share work within the public-domain, but the quality of texts – their faithfulness to printed editions that are generally perceived as more definitive – may vary according to the ability of their crowdsourced volunteer transcribers. Although we believe most of Project Gutenberg's texts

to be of high quality, given the number of texts comprising our study's corpus we were unable to thoroughly read all of those included. For those texts unavailable through Project Gutenberg, we referred to the Internet Archive's digitized book collection. However, given that the Internet Archive depends upon optical character recognition (OCR) to transcribe its scanned books rather than human transcribers, these texts are riddled with transposition errors. Fortunately, while a clear corpus is always ideal, studies have shown that the methods used in this study are reliable even in instances with substantial amounts of noise (e.g. misspellings) (Eder, "Mind your corpus"). The visualizations presented in the next section of this paper confirm stylo's ability to correctly group texts according to author, even with transcription errors.

It is important to note at this point that a stylometric analysis is always relational. Stylometric methods cannot conclusively reveal the correct author of a text; these methods can only compare authors within a corpus to reveal trends in word usage. In this paper, we have compiled a sizable reference corpus from open access digital repositories, but this corpus does not fully represent the entire oeuvre of any of the authors included. For example, only the second volume of Frances Milton's *The Three Cousins* (1847) could be found using our textual sources, leaving the first and third volumes wanting. The omission of texts from these repositories restricts this study, as the inclusion of more texts may have contributed to more reliable classification results. Further, the canon of works by Frances Eleanor has still not been conclusively defined.

Visualising Textual Relationships

Figure 1: A cluster analysis of the authorial Trollopes, along with all of the control authors included in our corpus. An enlarged version of this image is available at (citation removed for peer review).

Figure 2 A comparative visualization for the 1000 most frequent words throughout the corpus texts by the authorial Trollopes and Charles Dickens, generating using multidimensional scaling (MDS).

The reliability of the stylo package as it has been applied to our corpus specifically is shown in the above cluster analysis (all authors, Figure 1), multidimensional scaling (MDS) (Trollopes and Dickens, Figure 2) and principal components analysis (PCA) (Trollopes and Dickens, Figure 3) visualizations, which correctly group nearly all of the authors in the corpus according to each of their texts' 1,000 most frequently used words. Exceptions to these correct groupings are Frances Eleanor's *Frances* [Milton] *Trollope: Her Life and Literary Work* (fetrollope_frances1 and fetrollope_frances2) and some of Dickens' non-fiction works (dickens_childshistory, dickens_uncommercial, and dickens_americannotes), all appearing within Thomas Adolphus' cluster in the MDS and PCA plots. While largely distinct in both the MDS and PCA plots, Thomas Adolphus, Frances Eleanor, Frances Milton, and Dickens do momentarily blend to form one large clump in each of these plots, suggesting substantial stylistic similarities across some of their works.

The MDS and PCA plots show similar conclusions derived from different statistical methods for analysing the same data collected from each of the texts. Both are included here to validate stylo's reliability. In the former, MDS is used to map how similar individual points are to one another on a distance matrix. In this case, each point represents a text file. Through MDS, the relationships between text files are formalized through equated mathematical distance measures: in this case, John Burrows' Delta (explained in Evert et al.). Thus, two texts that are more stylistically similar another appear closer together in the image, while those texts with greater stylistic differences are further away. In the latter, PCA (explained in Binongo and Smith) condenses the dimensionality of the dataset by combining each text's individual variables (single words) into compound variables (groups of words). stylo then calculates the variance – deviation from the corpus' mean – of each text,

represented by its compound variable. Thus, those texts that typify our corpus appear in the middle of the PCA matrix (at the '0' mark of each axis). The further a text is away from this point, the more distinct it is. In the PCA plot shown in Figure 3, Thomas Adolphus' *A Siren* (ttrollope_siren) sits in the middle. Figure 3 reflects the principal components for 25.3% (15% + 10.3%) of individual variables. This percentage is admittedly low, but similar results are seen in Figure 2's MDS plot; stylo reliably clusters texts exemplifying similar styles. If the number of most frequently used words analysed is reduced to 100 (MDS in Figure 4; PCA in Figure 5), the principal components reflect 45.4% (33.1% + 12.3%) of the variables. The larger number of words in the sample, though, leads to the greater inclusion of content words as well as function words, which may contribute to more accurate clustering.

Figure 4 A comparative visualization for the 100 most frequent words throughout the corpus texts by the authorial Trollopes and Charles Dickens, generating using MDS.

Figure 5 A PCA of the 100 most frequent words throughout the corpus texts by the authorial Trollopes and Charles Dickens.

We chose to consider our corpus' 1,000 most frequently used words. Although similar results were garnered with just the 100 most common words, as shown in Figures 4 and 5, authorial clusters are easier to distinguish with the increased number of words. A smaller number of most common words does make for faster computation, though, and some studies indicate that one may need only 50-100 most common words to reliably distinguish between authors (Evert et al. ii5). We opted for 1,000 words over 100 because we prioritized the readability of visualizations over speed of computation. A list of the 1,000 words used in our analysis of the authorial Trollopes and Charles Dickens is available online (citation removed for peer review). Unsurprisingly, many of the most frequently used words are function words (e.g. the, to, of, and, a). As the list continues, more content words are included. Some of these

words are unique to certain books (e.g. "florence" and "george"), while others are less specific (e.g. "lady" and "young").

Discussion of Visualizations

Questions related to volumization are made apparent by Figure 1's cluster analysis. In a cluster analysis, similar texts are grouped; the further the grouped branches are to the right, the more similar the texts. Taking Frances Milton's *Charles Chesterfield* as an example, there seem to be differences – stylistic, or even substantive – between the volumes. Though the three volumes are clustered, the third volume appears divergent from the first two. There are similarities between all of the volumes, but the first two volumes are more similar to each other than they are to the final volume. The cluster analysis could be analogized as a family tree; in the case of *Charles Chesterfield*, the first two volumes could be twins, with the third volume being their sibling. Such a lack of stylistic homogeneity supports the view of multi-part texts representing "renewed and sustained relationships" (Hughes and Lund 145). This is especially true of texts published in multi-volume form and even more so for serial texts including *Charles Chesterfield*, which originally appeared in *The New Monthly Magazine* between July 1840 and November 1841. The extended period of writing associated with serialized fiction presumably affected the way in which an author wrote; the serial text was (and still is) alive and malleable. The production of Elizabeth Gaskell's *North and South* (1855) is a frequently cited example of this reflexivity (Womack). Stylometry, then, can lead us to question evolutions in writing style as well as the more commonly commented-on changes to plot. Only two examples from our corpus show the second volume of a work being distinct from the first and third. These are Oliphant's *At His Gates* and Frances Milton's *Uncle Walter*. The consistent pairing of the second volume with at least one of the other volumes may result from narrative convention, similar to the three-act dramatic

structure, in which the second volume is expected to function as a point of continuity between the first volume's exposition and the third volume's climax and resolution.

The above MDS and PCA plots (Figures 2 and 3, respectively) both show that by considering each of the text's 1,000 most frequently used words trends emerge to distinguish authorial writing styles. This is why all of authors included in the corpus appear in their own distinct clusters. There are, however, exceptions: some authors appear close to one another, with their works nestled within other authors' clusters. One such example, as expected, is that of Frances Eleanor and Dickens. As Frances Eleanor published four novels in Dickens' periodical *All the Year Round*, including her first two, this nearness may be explained by Dickens' involved editorial policies. That is, Frances Eleanor's literary style may have been informed by a close working relationship with Dickens in her early writing days. While her works form distinct clusters in all of the visualizations above, Frances Eleanor is identified as being stylistically close to Dickens as shown by their clusters' adjacency. Such closeness suggests Dickens' sustained influence on Frances Eleanor's writing throughout her authorial career.

Though editorial relationships may be discerned through computational analysis, they do not always emerge clearly. For instance, Frances Eleanor's *Sacristan's Household* (fetrollope_sacristan1 and fetrollope_sacristan2)*,* which was serialized in *Saint Paul's Magazine* under Anthony Trollope's editorship, appears in the visualizations closer to Dickens' and Thomas Adolphus' works than Anthony's. This proximity may indicate that Dickens' influence as Frances Eleanor's first editor was stronger than that of the editor under whom she was currently writing. The distinctiveness of Frances Eleanor's and Anthony's writing styles are exemplified by, for example, the distance between Frances Eleanor's *Sacristan* and Anthony's *Phineas Finn* (atrollope_phineasfinn). The publication of these two texts in *Saint Paul's* overlapped, with *Sacristan* appearing from October 1867 to May 1869

and *Phineas Finn* from July 1868 to June 1869. Despite the concurrent publication of these works, though, stylo identifies great stylistic difference. That Frances Eleanor's and Anthony's texts do not overlap may prompt questions about the differences between Dickens' and Anthony's approaches to editorship. Dickens was an infamously hands-on editor (Fitzgerald) who envisaged the pieces in *Household Words* as seeming to have been written by a single entity – a "shadow" (Dickens 5: 619). Anthony's approach appears to have been more relaxed, without such desire to establish a consistent voice. He writes of a tendency to "give way on behalf of some literary aspirant whose work did not represent itself to me as being good" (*An Autobiography* 125). Editorial influence is also not limited to the personal interactions at the time of writing, such as Frances Eleanor's close relationship with Dickens at the time her work was published under his editorship. Influence may be longer-lasting, impacting literary style later in these writers' careers. Other texts written after Dickens' death in 1870 appear just as closely clustered with his work as those subject to his editorial scrutiny. For example, *A Charming Fellow* (fetrollope_charming[1-3]) appeared in *All the Year Round* in 1876 under the editorship of Dickens's son, Charley. Other texts written after Dickens' death and published in the *Graphic*, such as the 1888 *That Unfortunate Marriage* (fetrollope_unfortunate[1-2]) and the 1892 *That Wild Wheel* (fetrollope_wildwheel), likewise overlap significantly with Dickens' work. That is, the similarities between Frances Eleanor's and Dickens' work cannot be solely attributed to his personal involvement in their production as editor but can be at least loosely traced in works edited by others. Such analyses are also useful in dispelling the misattribution of a lesser-known writer's work to her more widely-recognized relations. For example, Frances Eleanor's 1866 *Kätchen's Caprices* (not included here) has been routinely attributed to Anthony Trollope since Michael Sadleir's inclusion of the story in his *Trollope: A Bibliography* (229). A stylometric analysis of *Kätchen's Caprices* may affirm the work's literary style as that of Frances Eleanor.

A similarly close working relationship may account for overlaps in the clusters comprising the works of Frances Milton and Thomas Adolphus. Thomas Adolphus served as a kind of research assistant to Frances Milton (e.g. Trollope, *What I Remember, Volume II* 8). Inversely, Frances Milton occasionally served as Thomas Adolphus' editor – a role that landed her on the title page of some of his work (e.g. Trollope, *A Summer in Western France*). Additionally, the two had a particularly close personal relationship, with Thomas often traveling with his mother, recalling that "my presence and companionship were necessary to her" (*What I Remember, Volume I* 355). By the late 1830s both mother and son were living in Florence and, as had been the case in London and Paris, their home became a meeting place for the English literary community. Visitors included Charles Dickens, whom Thomas Adolphus (*What I Remember, Volume II* 110) describes as a "pretty-boy looking sort of figure", Elizabeth Barrett and Robert Browning, and George Eliot. Given Dickens' proximity to Frances Milton and Thomas Adolphus in the above visualizations, one could argue that Dickens' presence, which so clearly influenced the works of Frances Eleanor, may have also influenced the writing styles of Frances Milton and Thomas Adolphus – or, alternatively, that Dickens' writing style may have been influenced by his hosts. Stylometric analyses can only suggest the influence of one author upon another, though, and cannot prove the direction of influence.

Worth mentioning is the apparent influence of genre on stylo's placement of works in the above visualizations. Both volumes of Anthony's *North America* (atrollope_northamerica1 and atrollope_northamerica2), as well as his other non-fiction works included in the corpus (e.g. atrollope_autobiography, atrollope_lordpalmerston), appear in the MDS and PCA plots as distinct from his works of fiction. Indeed, Anthony's non-fiction works lie closer to the works of his brother than to his own. This positioning suggests that Anthony's writing style for his non-fiction works differs significantly from that

of his fiction works; perhaps he referred to his brother's works as guides for venturing into the realm of non-fiction. Dickens' non-fiction *A Child's History of England* (dickens_childshistory), *American Notes* (dickens_americannotes)*, and *Uncommercial Traveller* (dickens_uncommercial) are likewise separated from their actual author and appear more stylistically similar to the works of Thomas Adolphus. There is, it would seem, a distinct similarity of both Anthony's non-fiction writing style to that of his kin. There also appears to be a noteworthy likeness of Dickens' non-fiction style to that of Thomas Adolphus, with whom he engaged in regular correspondence and whose rapportage was widely read (letters throughout Trollope, *What I Remember, Volume II*).

Additionally, the visualizations show some of the authorial clusters remarkably close to one another, at times overlapping. As one example, Dickens sits between Frances Milton and Frances Eleanor. There are also those texts that appear altogether removed from their author's cluster: most clearly, Frances Milton's travel-related works – *Belgium and Western Germany in 1833* (fm_belgium), *Domestic Manners of the Americans* (fm_domestic), *Paris and the Parisians in 1835* (fm_parisians1 and fmtrollope_parisians2), and *Vienna and the Austrians* (fmtrollope_vienna1 and fmtrollope_vienna2) – which appear on the bottom right of Figures 2 and 3. Outliers' locations in the visualizations suggest differences in writing style to other books by the same authors. What we found most interesting, though, were those textual outliers far outside their attributed author's cluster, finding their homes deep within the clusters of other authors. These outliers include Charles Dickens' *Child's History* (dickens_childshistory), written from 1851 to 1853, and Frances Eleanor's 1895 biography of Frances Milton *Frances Trollope: Her Life and Literary Work* (fetrollope_frances1 and fetrollope_frances2), which have both found their homes within Thomas Adolphus' cluster. Our stylometric analysis has identified seemingly significant stylistic similarities in these

instances; future research involving close reading and pointed historical study may elucidate or disprove such identifications.

Conclusion

In this paper, we have used computational methods to drive a discussion about the social relationships that may have informed the production of the authorial Trollopes' works. A computational stylometric analysis of our corpus – comprising 201 public-domain texts with authorship attributed to Anthony Trollope, Frances Eleanor Trollope, Frances Milton Trollope, Thomas Adolphus Trollope, and Charles Dickens – shows that some texts may have been the products of complex personal and professional relationships between these individuals. What began as an effort to visualize the clusters of similar writing styles employed by the Trollope family has since become an examination into interpersonal networks – particularly those of the family unit – and their influence on writing style. The writing styles employed by the authorial Trollopes appear to reflect the complicated social networks underpinning the wider Victorian literary marketplace. While each Trollope has a distinctive voice, computational analysis of word usage suggests that some works were influenced by working relationships, whether formal or informal, between the family members.

We encourage other literary scholars to use such computational methods to reconsider literary networks in new ways afforded by digital tools. In particular, deeper analysis of where the authorial Trollopes fit within the intertextual web of Victorian fiction and non-fiction would contribute to more thorough understandings of what we consider a vastly understudied writing family; our previous studies (Dumbill, *Vanishing*) indicate that there is a basis for such research. This research would be complemented by close readings of those texts focused on here, with close reading distinguishing clear areas of stylistic or semantic

eccentricity. Our emphasis herein has been on a kind of computational distant reading to elucidate an authorial network, but distant and close reading may be applied in tandem for a more complete picture of the relationships that may inform processes of text production. Indeed, distant reading may be enhanced by interdisciplinary collaboration and adaptation that encourages use of computational tools not just for validation of traditional literary analysis, but also for prompting new questions – and, subsequently, discoveries – altogether (Hammond). Just as distant reading can inform our understanding of texts, preliminary close reading and literary analysis has informed our provocations here. However, we hope that this is just the beginning of deeper study of *all* of the authorial Trollopes. Attention to readers' experiences of the Trollopes' work was outside the scope of the project presented here, but would enhance future distant reading projects if such material is available (Dewitt).

Above all else, we hope that this paper has sparked an interest in the application of stylometry to textual scholarship. Although a stylometric analysis cannot necessarily yield any conclusive results regarding authorship attribution, it *can* prompt scholars of literature to reconsider familiar texts in new ways, and can provide quantifiable support or opposition to extant arguments.[iv] R's stylo package offers means for fresh perspective, and extensive instructional literature about this package is freely available. Indeed, we believe that stylo could be an appropriate starting point for those scholars hoping to integrate computational methods into their own practices to better understand authorial networks and complement close reading of literary texts and historical accounts.

Works Cited

"Adolphus Trollope's Return | From the London Truth." *The New York Times*, 18 July 1887, p. 3.

Babb, Genie. "Victorian roots and branches: 'The statistical century' as foundation to the digital humanities." *Literature Compass* vol. 15, 2018, pp. 1-14.

Binongo, José Nilo G., and M. W. A. Smith. "The Application of Principal Component Analysis to Stylometry." *Literary and Linguistic Computing*, vol. 14, no. 4, 1999, pp. 445-465.

Blanchard, Laman. "Memoir of Mrs Trollope." *The New Monthly Magazine*, vol. 55, no. 219, March 1839, p. 417.

Bode, Katherine. "Graphically Gendered: A Quantitative Study of the Relationships between Australian Novels and Gender from the 1830s to the 1930s." *Australian Feminist Studies*, vol. 23, no. 58, 2008, pp. 435-450.

Broughton, Trev and Helen Kingstone. "Generations." *Victorian Literature and Culture*, vol. 46, nos. 3-4, 2018, pp. 709-715.

Burrows, John F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford University Press, 1987.

---. "Questions of Authorship: Attribution and Beyond. A Lecture Delivered on the Occasion of the Roberto Busa Award." *Computers and the Humanities*, vol. 37, 2003, pp. 5-32.

"Centre for Literary and Linguistic Computing" The *University of Newcastle*, https://www.newcastle.edu.au/research-and-innovation/centre/education-arts/cllc/research. Accessed 7 January 2019.

"CLiC Dickens." *The University of Birmingham*, https://www.birmingham.ac.uk/schools/edacs/departments/englishlanguage/research/projects/clic/index.aspx. Accessed September 2019.

Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry*, vol. 45, 2019, pp. 601-639.

Dewitt, Anne. "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press." *Victorian Periodicals Review*, vol. 48, no. 2, 2015, pp. 161-182.

Dickens, Charles. *The Letters of Charles Dickens*. Pilgrim Edition, edited by Madeleine House, Graham Storey, and Kathleen Tillotson. Clarendon Press, 1965-2002. 12 vols.

Dickens Journals Online, https://www.djo.org.uk. Accessed 26 February 2021.

Douglas-Fairhurst, Robert. *Victorian Afterlives. The Shaping of Influence in Nineteenth-Century Literature*. Oxford University Press, 2002.

Dumbill, Eleanor. "Fiction, Non-Fiction, and Family Writing: Investigating the Boundaries of Literary Genre." *Textual Cultures*, forthcoming.

---. *Vanished Authors and Invisible Women: The Reputations and Networks of George Eliot, Frances Milton Trollope, and Frances Eleanor Trollope*. Doctoral Thesis. Loughborough University, 2021.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Package 'stylo'." *The Comprehensive R Archive Network*, 26 May 2018, https://cran.r-project.org/web/packages/stylo/stylo.pdf. Accessed 1 September 2018.

---. "'Stylo': a package for stylometric analyses." *Computational Stylists Group*, 19 June 2017, https://sites.google.com/site/computationalstylistics/stylo/stylo_howto.pdf. Accessed 1 September 2018.

Eder, Maciej. "Mind your corpus: systematic errors in authorship attribution." *Literary and Linguistic Computing*, vol. 28, no. 4, 2013, pp. 603-614.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Shöch, and Thorsten Vitt. "Understanding and explaining Delta measures for authorship attribution." *Digital Scholarship in the Humanities*, vol. 32., Supplement 2, 2017, pp. ii4-ii16.

Fitzgerald, Percy. "Charles Dickens in the Editor's Chair." *The Gentleman's Magazine*, vol. 250, 1881, pp. 725-742.

Gemma, Marissa, Frédéric Glorieux, and Jean-Gabriel Ganascia. "Operationalizing the colloquial style: Repetition in 19th-century American fiction." *Digital Scholarship in the Humanities*, vol. 32, no. 2, 2017, pp. 312-335.

Glendenning, Victoria. "Foreword." *Fanny Trollope: A Remarkable Life*, by Teresa Ransom. St. Martin's Press, 1995, pp. xi-xiii.

Hammond, Adam. "The double bind of validation: distant reading and the digital humanities' 'trough of disillusionment'." *Literature Compass*, vol. 14, no. 8, 2017, https://doi.org/10.1111/lic3.12402. Accessed 21 December 2020.

Heuser, Ryan and Long Le-Khac. "Learning to Read Data: Bringing out the Humanistic in the Digital Humanities." *Victorian Studies*, vol. 54, no. 1, 2011, pp. 79-86.

Hewitt, Martin. "Victorian Generations." *Professor Martin Hewitt. Victorian Manchester and more,* https://profmartinhewitt.com/2015/05/31/victorian-generations. Accessed 20 September 2019.

Holmes, David I. "The Evolution of Stylometry in Humanities Research." *Literary and Linguistic Computing*, vol. 13, no. 3, 1998, pp. 111-117.

Houston, Natalie M. "Toward a Computational Analysis of Victorian Poetics." *Victorian Studies*, vol. 56, no. 3, 2014, pp. 498-510.

Hughes, Linda K. and Michael Lund. "Textual/Sexual Pleasure and Serial Publication." *Literature in the Marketplace*, edited by John O. Jordan and Robert L. Patten. Cambridge University Press, 2003, pp. 143-164.

Kelly, Helena, "Ellen Ternan's Connections with Rochester: Some New Material." *Dickens Quarterly*, vol. 38, no. 38, pp. 83-87.

Kestemont, Mike, Sara Moens, and Jeroen Deploige. "Collaborative authorship in the twelfth century: A stylometric study of Hildegard of Bingen and Guibert of Gembloux." *Digital Scholarship in the Humanities*, vol. 30, no. 2, 2015, pp. 199-224.

Kirsch, Adam. "Technology Is Taking Over English Departments." *The New Republic*, 2014, https://newrepublic.com/article/117428/limits-digital-humanities-adam-kirsch. Accessed 21 December 2020.

Leary, Patrick. "Googling the Victorians." *Journal of Victorian Culture*, vol. 10, no. 1, 2005, pp. 72-86.

Marche, Stephen. "Literature Is not Data: Against Digital Humanities." *The Los Angeles Review of Books*, 2012, https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities. Accessed 21 December 2020.

Moretti, Franco. *Distant Reading*. Verso, 2013.

---. *Graphs, Maps, Trees.* Verso, 2007.

Mosteller, Frederick, and David L Wallace. *Inference and Disputed Authorship*. Addison-Wesley, 1964.

"MRS. TROLLOPE." *Supplement to THE SPEAKER* 13, 2 May 1896, pp. 488-489.

Nicholson, Bob. "Counting Culture; or, How to Read Victorian Newspapers from a Distance." *Journal of Victorian Culture*, vol. 17, no. 2, 2012, pp. 238-246.

Pennebaker, James W. *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury Press, 2011.

Pawłowski, Adam and Artur Pacewicz. "Wincenty Lutosławski (1863–1954): Philosophe, helléniste ou fondateur sous-estimé de la stylométrie?", *Historiographia Linguistica*, vol. 31, no, 2/3, 2004, pp. 423-447.

The Periodical Trollopes, https://periodicaltrollopes.omeka.net. Accessed 26 February 2021.

Sadleir, Michael. *Trollope: A Bibliography.* Constable. 1928.

---. *Trollope: A Commentary*. Constable, 1945.

Slater, Michael. *The Great Charles Dickens Scandal.* Yale University Press, 2012.

Stanford Literary Lab, https://litlab.stanford.edu. Accessed 2 August 2019.

Stubbs, Michael. "Quantitative methods in literary linguistics." *The Cambridge Handbook of Stylistics*, edited by Peter Stockwell and Sara Whiteley. Cambridge University Press, 2015, pp. 46-62.

Trollope, Anthony. *An Autobiography*. William Blackwood and Sons, 1883. 2 vols.

Trollope, Thomas Adolphus. *A Summer in Western France*, edited by Frances [Milton] Trollope. Henry Colburn, Publisher, 1841. 2 vols.

---. *What I Remember, Volume 1*. Richard Bentley and Son, 1887,

---. *What I Remember, Volume II*. Richard Bentley and Son, 1887.

The Wellesley Index to Victorian Periodicals, 1824 – 1900, http://wellesley.chadwyck.co.uk. Accessed 26 February 2021.

Womack, Elizabeth Coggin. "Anticipated Ends, Atonement, and the Serialization of Gaskell's *North and South*" *Dickens Studies Annual 48,* 2017, pp. 231-252.

[i] Cecilia Trollope's novel, *Chollerton: A Tale of Our Own Times*, appeared anonymously in 1846, framed as being "by a lady".

[ii] An author named Frank Trollope published eleven works between 1865 and 1880 but few biographical details are known and he does not seem to have been related to the other Trollopes. The modern romantic novelist Joanna Trollope is related but has been excluded here for the purposes of periodization.

[iii] Helena Kelly similarly argues that Charles Dickens' purported relationship with Ellen Ternan may have influenced his characters.

[iv] One use of the methods described in our study might be to interrogate Harold Bloom's conclusion that Wilde's "Ballad of Reading Gaol" is a poor reproduction of Coleridge's "Rime of the Ancient Mariner".