Who Wrote Wuthering Heights?

Published version: https://doi.org/10.1093/llc/fqaa031

Despite the profile and critical acclaim of his three siblings, Patrick Branwell Brontë has largely been neglected by literary history. Commonly known as Branwell Brontë, he was the Brontë family's only son, the fourth born of six children. While there has been some consideration of his writing (FitzGerald 2013), such appreciation remains relatively scant. This might be attributed to Branwell's oeuvre, modest in comparison to those of his sisters, Charlotte, Anne and Emily. Branwell is recognised as having had some ability as a writer, and indeed, his literary career was "not without achievement" (Donovan 2016, 225), but ultimately, many critics concur that he "wasted his talent" (Donovan 2016, 214). Branwell struggled with alcoholism and addiction throughout his life (Barker 1997, 179), becoming a constant source of concern for his father and siblings (Lock 2017). Despite his troubles, Branwell seemed to enjoy an intermittent but stable relationship with his sisters, and from an early age the four Brontés formed a creative collaboration (Butcher 2019, 4) which would persist throughout much of their writing lives (Malfait and Demoor 2015; Braxton 2019).

In the wake of Branwell's death in 1848, his friends William Deardon and Francis Grundy began promoting the rumour that it was in fact he and not Emily, also deceased at this point, who had written *Wuthering Heights* (Mellor 1993, 190). The contention first circulated in 1867 when an anonymous article in *People's Magazine* expressed doubt that Emily, "a timid and retiring female", could have written such a coarse and savage novel (Willis 1947, 157). Responding to the article through a letter published in the *Halifax Guardian*, Deardon—under the pseudonym William Oakendale—claimed to have recognised similarities between the novel and draft materials by Branwell which he had seen long before *Wuthering Heights* was published. Having detailed Deardon's allegations in her seminal essay on the matter, Irene Cooper Willis roundly dismantles his account. Willis' essay was published in 1947 and seems to be the final substantial and expert word on the matter.

If one digs extensively enough, they will find a selection of scholarly works which seriously examine this claim, though most of these are either decades old, penned by students as part of their graduate studies, or published under imprints of dubious reputation. The general consensus among scholars is that Emily Brontë wrote *Wuthering Heights* and there is very little evidence to suggest otherwise. Most other claims to the contrary have come from enthusiasts, amateur critics who contend that there is an element of truth to the rumours. In 2014, Chris Firth, who self-published *Branwell Bronte's Tale: Who Wrote Wuthering Heights?* through Amazon's CreateSpace platform (Firth 2014), remarked in a local periodical how he believes that Branwell wrote *Wuthering Heights* (Knights 2014). In the article, Firth calls on scholars to use "computer tests" to examine his hypothesis, a call to which this essay is responding.[1]

Stylometric Analysis

Using the Stylo package for R (Eder, Kestemont, and Rybicki 2013), this paper analyses the authorial fingerprint in *Wuthering Heights* to determine what contribution, if any, Branwell, or anybody else, might have made to writing the novel. Stylometry is a statistical technique which indicates likely authorship, forming an "impression" of how a particular author writes by counting the frequency of words across sample texts. While the specific techniques differ across the iterative stages of this study, the analysis is always conducted using the 100 most frequent words[2] from the chosen samples, with the

similarity between styles measured using Support Vector Machine classification, Burrows' Delta and Cosine Delta (Burrows 2002; Eder 2016; Evert et al. 2017).

What Firth fails to recognise is that a lack of samples from all likely candidates makes using a computer-assisted analysis to address this particular question something of an issue. He stated in 2014 that scholars were waiting on further samples of Branwell's surviving prose with which to test the authorship of *Wuthering Heights* (Knights 2014), but the issue is not Branwell, for whom we have the drafts of an unfinished novel, "and the weary are at rest" (Brontë 2015),[3] the limitation in this instance is Emily. This complication has long been faced by literary historians and biographers: "…the personality of the subject of our attention and speculations, the reserved and private figure of Emily Brontë herself" (Pykett 1989, 1). Aside from the novel, which in this context we must consider disputed, we have very few surviving examples of Emily's writing other than her poetry. Unfortunately, poetry is unsuited to measuring the authorial fingerprints in fiction.[4] A robust stylometric analysis requires two things: the disputed text and *suitable* samples from *each* of the authorial candidates. We can measure Branwell's style against *Wuthering Heights* using "and the weary are at rest", but we have no equivalent test sample for Emily.[5]

Some letters and diary notes by Emily do exist, but in addition to the aforementioned issues of genre, a paucity of volume makes them unsuitable as statistical samples. Gathering what materials we could from the letters and diaries readily available,[6] we compiled a sample of some 1,200 words. Recent work by Maciej Eder has shown that as little as 2,000 words may be a sufficient amount of text for stylometric analysis (2017), a significant revision on previous approximations of 5,000 words (Eder 2015). Indeed, Eder has also suggested, and one of the authors of this paper has found from their own experience working with textual fragments, that one can even go lower, perhaps to 1,000 words and below depending on the texts in question.[7] Thus, while an imperfect sample, our first test assessed the authorship of *Wuthering Heights* using "and the weary are at rest" alongside Emily's small selection of letters and diary notes.

Using rolling stylometry (Eder 2016) with Burrow's Delta, Branwell's authorial fingerprint does emerge as marginally stronger than Emily's (see Fig. 1),[8] but conducting the same analysis using the SVM classifier,[9] which Eder shows is a more appropriate measure in this instance, reveals that Emily's fingerprint is the most prominent, but there are still sections of the novel which are closer to Branwell's style (see Fig. 2).

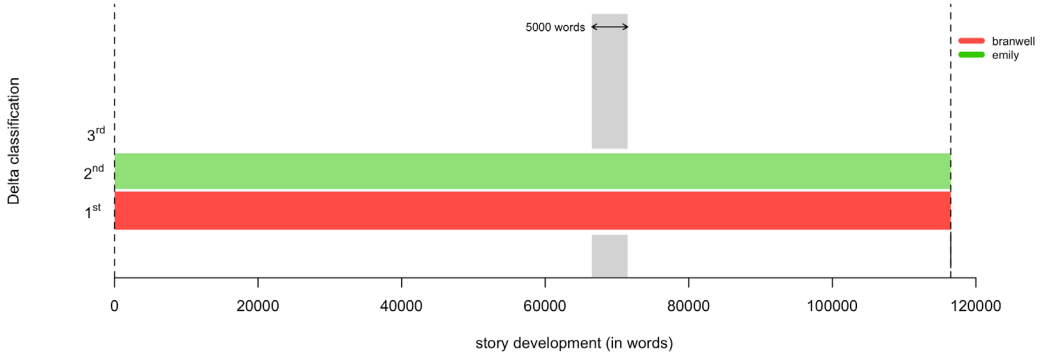

Figure 1. Wuthering Heights assessed using Eder's rolling stylometry (Delta, 100 MFW)

Figure 2. Wuthering Heights assessed using Eder's rolling stylometry (SVM, 100 MFW)

The sensationalists among us might be tempted to take these results at face value and claim that Branwell is the true author of *Wuthering Heights*, or at the very least, a significant contributor. But we must remember the limitations in Emily's writing sample; not only is it questionable in terms of length, it is at a distinct disadvantage to that of her brother's, which is fiction and thus more likely to be generically more similar in style to *Wuthering Heights*.

Adding Anne and Charlotte as additional candidates provides a more useful test, and one which essentially dismisses Branwell's claim to the novel. This is because Charlotte's style emerges as the dominant signal,[10] followed by Anne, with Branwell ranked as the third most likely candidate, results which are consistent across tests using Burrows' Delta and SVM classification (see Figs. 3 and 4).

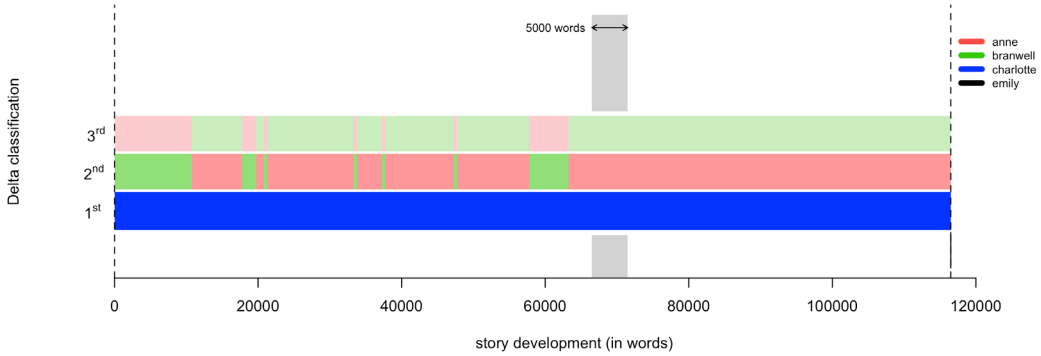

Figure 3. Wuthering Heights assessed using Eder's rolling stylometry (Burrows' Delta, 100 MFW)

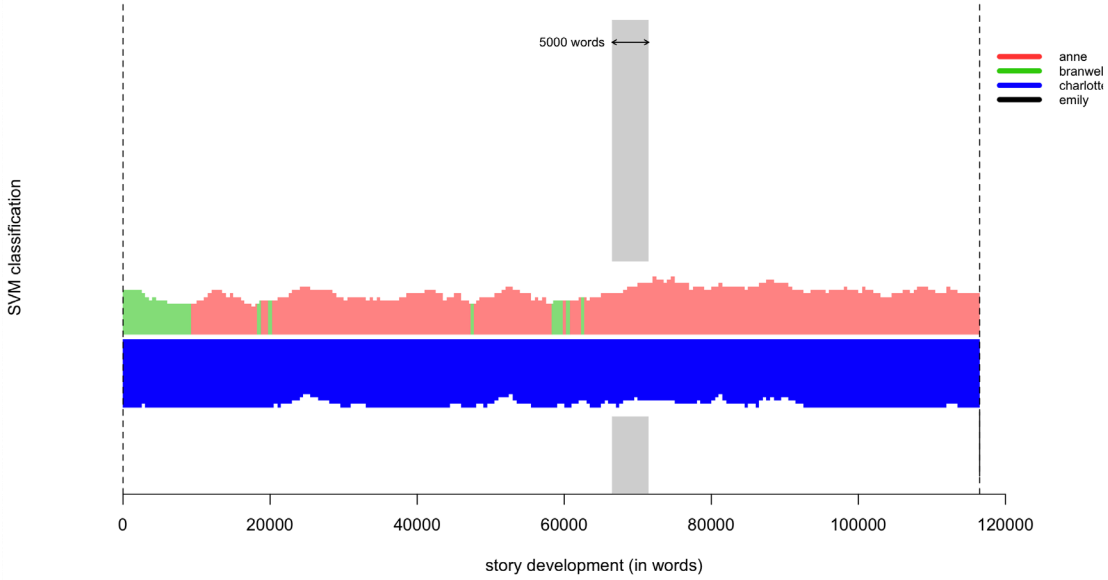

Figure 4. Wuthering Heights assessed using Eder's rolling stylometry (SVM, 100 MFW)

Comparing these four tests shows Branwell's signature is only evident in the novel when an unreliable dataset is used; when all candidates and their works of fiction are featured, his style fades into the background to the point where it is negligible. In other words, Branwell's style only shows up in *Wuthering Heights* when competing with work by Emily which cannot be considered suitable for analysing the likely authorship of a piece of fiction. When the other two sisters, for whom we do have more suitable writing samples, are featured in the analysis, Branwell emerges as the *least likely* to have contributed to the writing of *Wuthering Heights*. And so, at this point, though with further results still to present, one can very reasonably conclude that Branwell did not write *Wuthering Heights*.

A cluster analysis of the fiction penned by the Brontës (see Fig. 5) gives a clear picture of who wrote *Wuthering Heights*—it was Emily.

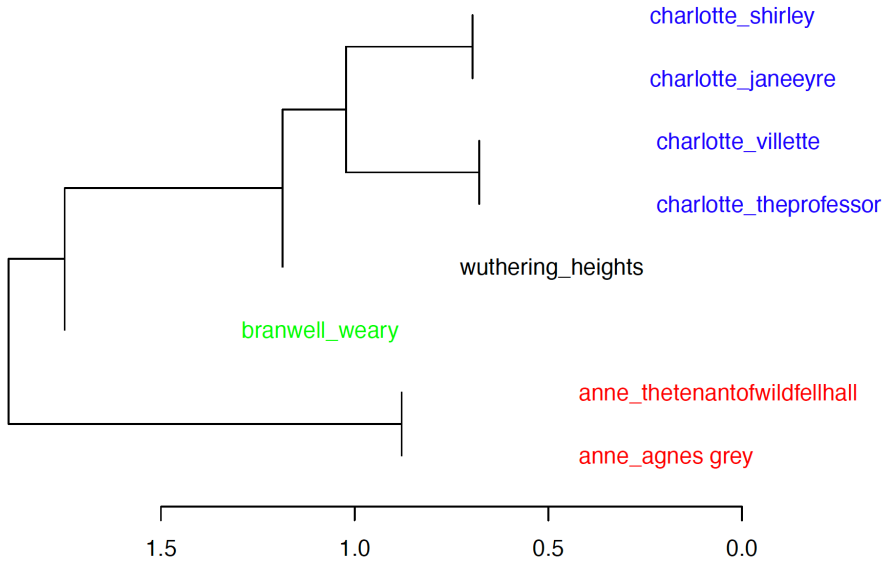

Figure 5. Cluster analysis of the Brontë siblings' fiction (Delta, 100 MFW)[11]

That the novel essentially stands alone within the cluster tells us that the authorial signature in the novel is not that of Branwell, Charlotte or Anne, and thus the most reasonable conclusion is that it was written by Emily, as most people have always recognised.

But it is not quite "case closed". If someone other than Emily is the true author of *Wuthering Heights*, these limited findings show that it is most likely Charlotte, a theory which has also has gathered a modicum of traction among Brontë enthusiasts (Carter 2011; Ostrowski 2013). It may be unsurprising that the authors of this paper also reject the idea that Charlotte wrote the novel, but do think it interesting that this study shows quantitative evidence of the Brontë family's creative collective that has been explored by various critics (Van Der Meer 2008; Malfait and Demoor 2015; Braxton 2019; Butcher 2019).

Creative Collaborations between the Brontës[12]

As indicated in this essay's notes,[10] the 1847 first edition has been used in the hope of diminishing the influence of Charlotte's authorial fingerprint, who edited the posthumous 1850 second edition. However, repeating the previous cluster analysis using Cosine instead of Burrows' Delta (Evert et al. 2017) produces a slightly different result, placing *Wuthering Heights* and Branwell's "and the weary are at rest" in a cluster with Charlotte's novels, with Anne's authorial fingering remaining the most distinct within the Brontë canon (see Fig. 6).

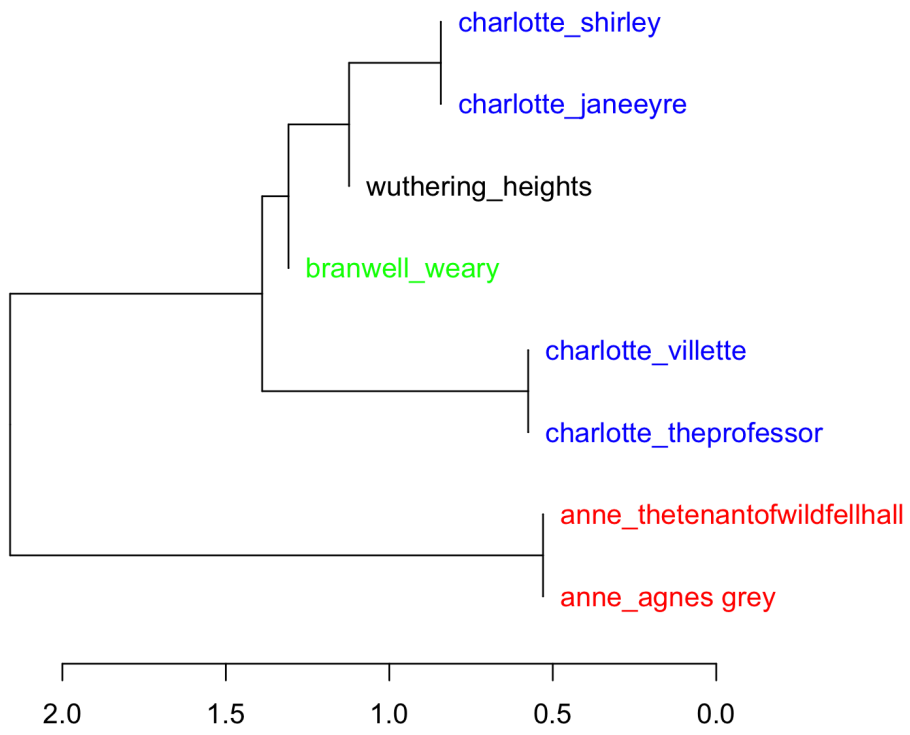

Figure 6. Cluster analysis of the Brontë siblings' fiction (Cosine Delta, 100 MFW)

This new cluster shows evidence of some shared creative influence among the siblings. It is possible that up until now this study has understated the significance of their collaborations,[13] and that many of the results produced can simply be attributed to

creative interactions which began in their formative years and continued into adulthood.[14]

Collaboration between the Brontës is important to this study's purposes in that it offers another approach to assessing the authorship of *Wuthering Heights*. Conducting a further cluster analysis that includes a relatively comparative sample of British fiction from beyond the Brontë family reveals something that has been absent from the previous tests: in a broader context, *all of the Brontës* are stylometrically similar (see Fig. 7).[15]

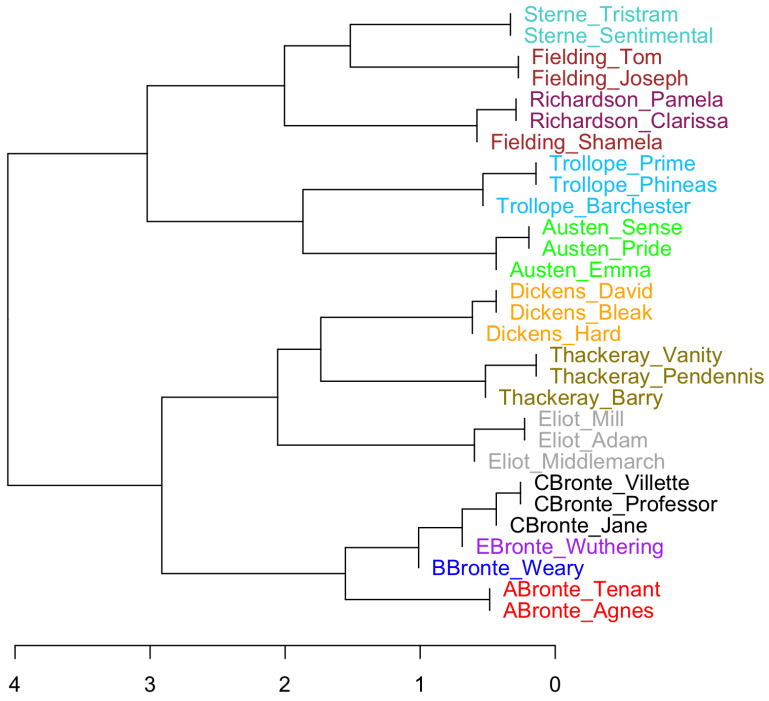

Figure 7. Cluster analysis of a small collection of British fiction (Cosine Delta, 100 MFW)

The results from the previous clusters are replicated, in that Anne's style remains the outlier, with Branwell and Emily "in between" Charlotte and Anne. But it is remarkable that the Brontës cluster as a familial unit when added to a study with other British writers. One would not anticipate members of the same family having a stylistic fingerprint so similar that they form their own cluster when assessed in the context of their wider national canon. These findings prove that the Brontës had a significant influence on each other's style, such that, at a macro-level, they gather together as a unique cluster.[16]

In the context of this particular study, these findings suggest that any proximity between *Wuthering Heights* and the style of either Branwell or Charlotte can be attributed to this unusual, markedly measurable affinity that the Brontës all share as writers. This particular cluster (see Fig 7.) can be taken as quantitative evidence that the Brontës were so influential to each other's style that a computer will not always distinguish much stylistic difference between their work. Knowing this, interpreting any of the findings in this study as proof that someone other than Emily wrote *Wuthering Heights* ignores the interaction between each of the analyses presented. Perhaps some of these findings show that Branwell and Charlotte did have an influence on Emily's writing, but when taken as a whole, the results in this study paint a picture of influence and collaboration, not mistaken authorial identity.

The Unmistakable Air of Masculinity

In those few limited tests that can be conducted with this imperfect dataset, one can confidently draw the conclusion that Branwell did not write *Wuthering Heights*, and that, as most scholars and critics have always suspected, Emily is its author. At the very best, Branwell might be said to have contributed some inspiration, exposing Emily to the sorts of afflictions and obsessions that emanate as themes in the novel (Mellor 1993, 191). Branwell's personal traits and mannerism seem to match those of Heathcliff, so perhaps Emily's brother was more of an unwitting participant in the development of *Wuthering Heights*. But there is a deeper issue here, one which Willis called out in the forties and might benefit from some re-articulation: the authorship of *Wuthering Heights* would never have been contested had Emily Brontë been a man.

The non-mystery mystery of who wrote *Wuthering Heights* is a timely reminder of those gender biases which have long held women writers back:

> As to the suggestion made when *Wuthering Heights* was first published, and repeated since, that no woman's hand could have penned *Wuthering Heights*, that over every page there hangs an "unmistakable air of masculinity", this hardly requires to be answered nowadays, with the knowledge that we have of the inter-mixture of so-called male and female qualities in most people. It has become a commonplace that artists do not often produce work that is identifiable with the personalities they appear to have and that qualities which in the artist's everyday personality are un-noticeable or even seem to be lacking burst forth in his or her artistic work. (Willis 1947, 168)

It is encouraging to see Emily's claim to the novel has not been questioned more widely, that "Victorian constructions of femininity" and the author's persona as the unmarried "virginal daughter" of a clergyman, incapable of producing such a text, seem to have dissipated within contemporary criticism.

Yet it remains unacceptable that one can Google "Who Wrote Wuthering Heights?" and see her life's great literary accomplishment questioned by amateur critics and literary sleuths. In 2004, Chairman of the Brontë Society, Robert Barnard, publicly decried this theory "which has been going an awful long time", citing what amounts to misogyny, the reality that there are people who believe "a woman could not or should not write a book of such greatness", as the motivations for the rumours (*Halifax Courier*, 2004). Such unfounded positions should not be allowed to pass without response, and publications like *Keighley News*, however limited in circulation, should be confronted for publishing sensationalist pieces which do little but further serve the marginalisation of women from canon—it is bad enough that Victorian sentiments forced Emily to first publish under a male pseudonym without her legacy being further denigrated by continuing biases.

The issue is not entirely one of misogyny, of course, as it is sometimes Charlotte who is put forward as the novel's true author, but as demonstrated in this essay, stylometric analyses of these contentions do not lend credence to any reality other than that which is most widely accepted, that *Wuthering Heights* was written by Emily Brontë. Charlotte writes this herself in her preface to the second edition, remarking upon the tension between Emily's perceived personality and the contents of the novel, affirming her sister's authorship and reminding us that a person as they are seen in public can be at odds with the products of their own imagination. It would be nice if those proponents of Charlotte's claim to *Wuthering Heights* would accept the word of the author they are so eager to laud.

As a closing addendum, it is important that we reclaim the promise of computer-assisted critical techniques from those who would cite such practices for misplaced, personal agendas. As scholars of the digital humanities, we should not allow commentators, evidently invested in *Wuthering Heights* being recognised as the work of someone other than Emily, to use our discipline and its techniques as something of an intellectual threat. The article in *Keighley News* almost suggests that stylometry will eventually come for Emily Brontë, as though the purpose of our field is to engage in "gotcha" criticism. Authorship attribution is not about re-attributing the work of one to another, but adding a different kind of quantitative evidence to the complex questions of culture that can be addressed through the dynamics of style and influence. Computer-assisted analyses—when *properly implemented*—will do little to serve interpretive fantasies and the intentional disempowerment of legitimate authorial voices.

References

Barker, Juliet. 1997. *The Brontës: A Life in Letters*. Viking.

Braxton, Kimberley Jayne. 2019. 'The Brontës' Writing Community: Family, Partnership and Creative Collaboration'. Doctoral, Keele University. United Kingdom. http://eprints.keele.ac.uk/6083/.

Brontë, Branwell. 2015. 'And the Weary Are at Rest'. In *The Works of Patrick Branwell Brontë: Volume 3, 1837-1848*, edited by Victor A. Neufeldt, 420–66. Routledge.

Burrows, John. 2002. '"Delta": A Measure of Stylistic Difference and a Guide to Likely Authorship'. *Literary and Linguistic Computing* 17 (3): 267–87.

Butcher, Emma. 2019. *The Brontës and War: Fantasy and Conflict in Charlotte and Branwell Brontë's Youthful Writings*. Palgrave Macmillan.

Carter, Michele. 2011. *Charlotte Brontë's Thunder: The Truth Behind the Brontë Sisters' Genius*. Boomer Publications.

Donovan, Julie. 2016. 'The Poetry and Verse of Branwell Brontë'. In *A Companion to the Brontës*, edited by Diane Long Hoeveler and Deborah Denenholz Morse, 213–27. John Wiley & Sons.

Eder, Maciej. 2015. 'Does Size Matter? Authorship Attribution, Small Samples, Big Problem'. *Literary and Linguistic Computing* 30 (2): 167–82. https://doi.org/10.1093/llc/fqt066.

———. 2016. 'Rolling Stylometry'. *Digital Scholarship in the Humanities* 31 (3): 457–69. https://doi.org/10.1093/llc/fqv010.

———. 2017. 'Short Samples in Authorship Attribution: A New Approach'. In *DH2017 Conference Abstracts*. Montréal. https://dh2017.adho.org/abstracts/341/341.pdf.

Eder, Maciej, Mike Kestemont, and Jan Rybicki. 2013. 'Stylometry with R: A Suite of Tools'. In *Digital Humanities 2013: Conference Abstracts*, 487–89. University of Nebraska–Lincoln.

'Emily Bronte's Letters and Diary Papers'. 2005. Academic.Brooklyn.Cuny.Edu. 2005. http://academic.brooklyn.cuny.edu/english/melani/novel_19c/wuthering/diary_papers.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 2017. 'Understanding and Explaining Delta Measures for Authorship Attribution'. *Digital Scholarship in the Humanities* 32 (suppl_2): ii4–16. https://doi.org/10.1093/llc/fqx023.

Firth, Chris. 2014. *Branwell Bronte's Tale: Who Wrote Wuthering Heights?* Createspace Independent Publishing.

FitzGerald, Sally. 2013. 'Painful Life, Azrael, The Weary, and Dr Wheelhouse: The Diverse Legacy of Branwell Brontë'. *Brontë Studies*, November. https://doi.org/10.1179/147489306X108487.

Halifax Courier. 2004. '"Branwell, Not Emily, Wrote Wuthering Heights"', 2004. https://www.halifaxcourier.co.uk/news/branwell-not-emily-wrote-wuthering-heights-1-1957738.

Hoover, David L. 2009. 'Word Frequency, Statistical Stylistics and Authorship Attribution'. In *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*, edited by Dawn Archer, 35–51. Ashgate.

Hoover, David L., and Shervin Hess. 2009. 'An Exercise in Non-Ideal Authorship Attribution: The Mysterious Maria Ward'. *Literary and Linguistic Computing* 24 (4): 467–89. https://doi.org/10.1093/llc/fqp027.

Knights, David. 2014. 'Brontë Writer Asks Boffins for Help Proving Branwell Wrote Wuthering Heights'. *Keighley News*, 2014. https://www.keighleynews.co.uk/news/11629763.bronte-writer-asks-boffins-for-help-proving-branwell-wrote-wuthering-heights/.

Lock, Pam. 2017. 'Brontës under the Influence: The Legacy of Branwell's Drinking'. The Conversation. 2017. http://theconversation.com/brontes-under-the-influence-the-legacy-of-branwells-drinking-85649.

Malfait, Olivia, and Marysa Demoor. 2015. 'Sibling Collaboration and Literary After-Life: The Case of the Brontës'. *Brontë Studies* 40 (3): 187–200. https://doi.org/10.1179/1474893215Z.000000000148.

Mellor, Anne K. 1993. *Romanticism and Gender*. Routledge.

Ostrowski, Eryk. 2013. *Charlotte Brontë i Jej Siostry Śpiące*. MG.

O'Sullivan, James. 2014. 'Finn's Hotel and the Joycean Canon'. *Genetic Joyce Studies*, no. 14. http://www.geneticjoycestudies.org/GJS14/GJS14_OSullivan.htm.

Pykett, Lyn. 1989. *Emily Brontë*. Rowman & Littlefield.

Van Der Meer, Carolyne. 2008. 'Branwell Brontë'. In *The Literary Encyclopedia*. https://www.litencyc.com/php/speople.php?rec=true&UID=581.

Weidman, Sean G., and James O'Sullivan. 2018. 'The Limits of Distinctive Words: Re-Evaluating Literature's Gender Marker Debate'. *Digital Scholarship in the Humanities* 33 (2): 374–390. https://doi.org/10.1093/llc/fqx017.

Willis, Irene Cooper. 1947. 'The Authorship of "Wuthering Heights"'. *The Trollopian* 2 (3): 157–68. https://doi.org/10.2307/3044519.

Notes

[1] We appreciate that Firth might not be considered a professional critic or scholar and so do not want to be seen to be unduly fuelling his assertions by engaging with them in a published essay. Nor do we want to be seen to be assailing him unnecessarily, but as a competent adult and former recipient of an Arts Council Writer's Award (*Halifax Courier*, 2004) who has publicly claimed that stylometry will, without any real expertise in the subject, disprove Emily Brontë's authorship of *Wuthering Heights*, we think it important to offer him, as well as the long deceased Deardon and anyone who might share their position, a robust and hopefully final response on this matter.

[2] The authors of this paper have consistently used no more than 100 most frequent words because they subscribe to the theoretical view that results become less indicative of authorial fingerprint as the number of features is increased. When stylometry is conducted using a small sample of high-frequency words, typically function words, the analysis is conducted using words which are "especially resistant to intentional authorial

manipulation" (Hoover 2009, 35), and thus suited to determining subconscious authorial fingerprints rather than content distinct to the particular narrative.

[3] The length of this draft manuscript is 25,529 words.

[4] There is little to be gained in extending the tests to include the poetry written by Branwell and Emily. As with the letters and diary notes, the style that one encounters in poetry is very different to that found in fiction. To confirm this position, we ran the tests anyway, and in all instances the files containing poetry clustered together, away from the fiction, regardless of the author. We replicated these findings with other authors who have written both poetry and fiction, and again, the authorial signal was lost to the styles of form: the poetry always clustered with the poetry, the fiction with the fiction. As far as authorship attribution is concerned, there is nothing to be gained from combining poetry and fiction into one corpus.

[5] Readers may be interested in studies which attempt to overcome the challenges of non-ideal authorship attribution, such as that by Hoover and Hess (2009).

[6] The following source, maintained at the City University of New York's Brooklyn College, was used:
http://academic.brooklyn.cuny.edu/english/melani/novel_19c/wuthering/diary_papers.

[7] While not explicitly discussed in the published article, O'Sullivan's stylometric analysis of James Joyce's "Finn's Hotel" vignettes found that, in certain contexts, the fragments would cluster relatively consistently despite some of them being very short (O'Sullivan 2014).

[8] The visualisation presented lists authorial signatures in terms of most likely candidate, with the most likely being "1st". The x-axis represents the disputed text, accounting for any variation in signature across 5,000-word segments.

[9] A good account of the difference between SVM and Delta can be found in Eder's work on rolling stylometry, where he explains that, unlike SVM, "Delta does not combine individual training samples into averaged profiles for each class" (2016, 460).

[10] This might be unsurprising had we used the 1850 second edition, published posthumously and edited by Charlotte, though we doubt the extent to which an editor's fingerprint would dominate that of the original author. For our analysis we used the authoritative Clarendon edition, based on Emily's original 1847 first edition.

[11] The authors also completed a Bootstrap Consensus analysis (100–1,000 MFWs) which replicated these results.

[12] With thanks to Jan Rybicki for suggesting the addition of this section, as well as the inspiration and dataset necessary for the broader analysis of British fiction (see Fig. 7).

[13] With thanks to Mary O'Connell and Emma Butcher for providing some helpful hints and suggestions on the subject of the Brontë's collaborations.

[14] We appreciate that there is much scope here for the dangerous speculation that this paper seeks to subvert, that an alternative interpretation of the cluster presented in Figure 6 might be that Charlotte did more than just edit *Wuthering Heights* and did in fact write the novel in its entirety. Like the Branwell theory, there are very few of what one might consider robust scholarly resources in support of this position. Our rolling classifications do show that Charlotte's fingerprint is the most dominant of the siblings (see Figs. 3 and 4), but again, the significance of these findings is diminished by the absence of further samples of Emily's writing. One could also conclude from these results that Charlotte also wrote Branwell's unfinished manuscript, that she wrote everything except that attributed to Anne, though a more informed reading would be that she simply influenced the work of her brother and sister, and perhaps, they influenced her in return. It would be quite a leap to read this analysis as proof that Charlotte actually wrote any bit of *Wuthering Heights*, but we are sure that there are those

who will pick this one set of results out of the many produced in this paper—all of which should be considered *together*—and bend it to their critical purposes.

[15] The authors of this paper are somewhat wary of stylometric analyses where the selection of authors is based on problematic notions like canonicity or perceived similarity, an issue which O'Sullivan has also raised in other studies (Weidman and O'Sullivan 2018, 378). Whenever a study is reliant on the subjective selection of counter samples, as is the case here, the argument can always be made that an entirely different selection of authors might yield different results. Choosing a small selection of British authors to compare to the Brontës feels a little too random, but it is appreciated that in this instance there is merit to doing so as it demonstrates something particularly peculiar about stylistic affinities between the Brontës that could not otherwise be illustrated.

[16] A further interesting study would be to test the degree to which this trend is replicated across other literary families, or if there is in fact something very particular about the ways in which the Brontës developed something of a shared style.