THE AUTHOR

Between Style and Substance

Mark Algee-Hewitt
The Stanford Literary Lab
Krakow 2016

The Day may arrive when the psychology of composition is unified with the science of objective evaluation, but so far they are separate. It would be convenient if the passwords of the intentional school, "sincerity," "fidelity," "spontaneity," "authenticity," "genuineness," "originality," could be equated with terms of analysis such as "integrity," "relevance," "unity," "function"; with "maturity," "subtlety," and "adequacy," and other more precise axiological terms – in short, if "expression" always meant aesthetic communication. But this is not so.

Wimsatt and Beardsley, "The Intentional Fallacy," 1946

We know now that a text is not a line of words releasing a single 'theological' meaning (the 'message' of the Author-God) but a multi-dimensional space in which a variety of writings, none of them original, blend and clash. The text is a tissue of quotations drawn from the innumerable centers of culture.

Once the Author is removed, the claim to decipher a text becomes quite futile.

Roland Barthes, *The Death of the Author*, 1967

Stylometry – the statistical analysis of literary style – does not seek to overturn traditional scholarship by literary experts and historians, rather it seeks to complement their work by providing an alternative means of investigating works of doubtful provenance. At its heart lies an assumption that authors have an ==unconscious== aspect to their style, an aspect that cannot be consciously be manipulated but which posses features that are quantifiable and that may be distinctive. By measuring and counting these features, stylometrists hope to uncover the =="characteristics"== of an author.

Holmes and Kardos, "Who was the author?" 2003

Instead, they [Mosteller and Wallace] focused on so-called *function words*, words like conjunctions, prepositions, and articles that carry little meaning by themselves (that about what "of" means) but that define relationships of syntactic or semantic functions between other ("content") worlds in the sentence. These words are therefore largely topic-independent and may serve as useful indications of an ==author's preferred way to express broad concepts== such as "ownership."

Juola, "Authorship Attribution," 2006

Content-free words are the 'syntactic glue' of a language: They are words that carry little meaning on their own but form the bridge between words that convey meaning. Their joint frequency of usage is known to provide a useful stylistic fingerprint for authorship, and thus suggests a method of comparing author styles.

> Hughes et al., "Quantitative patterns of stylistic influence in the evolution of literature," 2011

Unlike typical document classification, however, in authorship attribution one does not desire to classify documents based on document content. Instead, one wishes to perform classification based upon author signal, or "style."
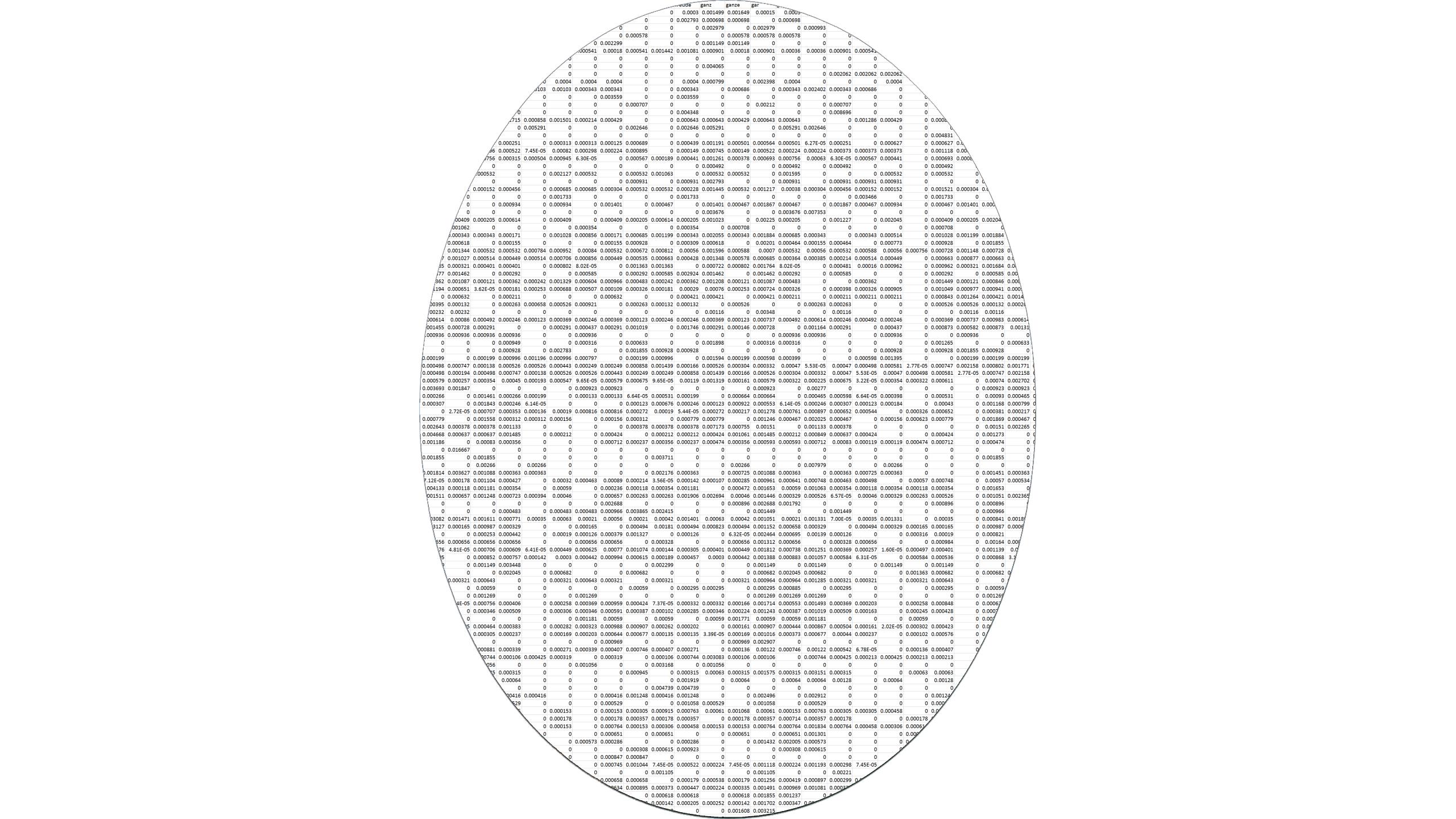
> Jockers and Witten, "A comparative study of machine learning methods for authorship attribution," 2010

[James's style] pervades his entire vocabulary, from the most frequent words in English to the rarest and most peculiarly Jamesian adverbs. But that distinctive style is not monolithic. Rather, it develops so gradually and consistently throughout his career that quantitative evidence from his use of words places his novels in almost perfectly chronological order.

Hoover, "Corpus Stylistics, Stylometry, and the Styles of Henry James," 2007

We demonstrate the effectiveness of multiple methods of stylometry in nonadversarial settings and show that authors attempting to modify their writing style can reduce the accuracy of these methods from over 95% to the level of random chance. With this we have demonstrated that current approaches to stylometry cannot be relied upon in an adversarial setting.

Brennan et al., "Adversarial Stylometry," 2012
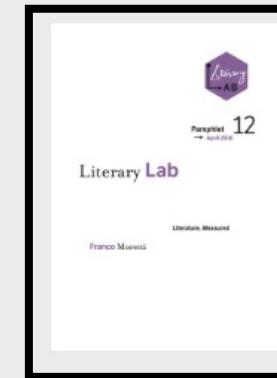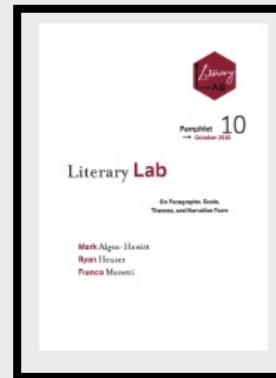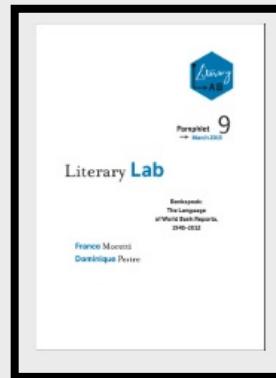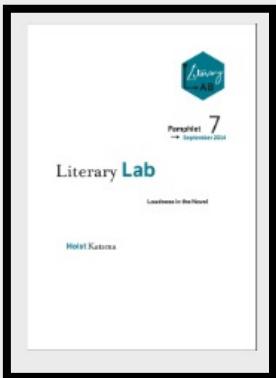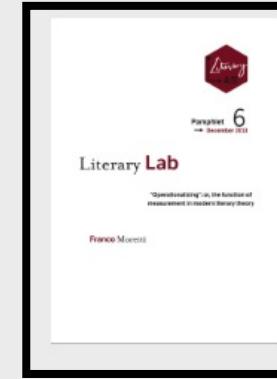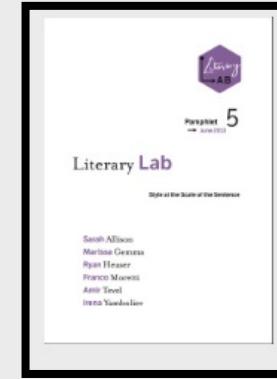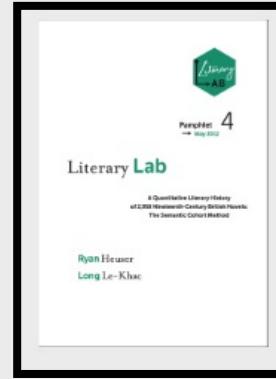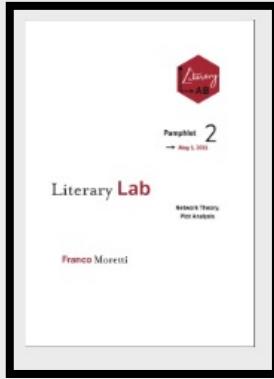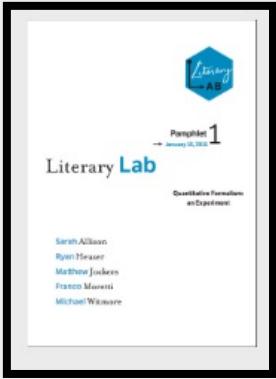
Even now, when we study the history of a concept, a literary genre, or a branch of philosophy, these concerns assume a relatively weak and secondary position in relation to the solid and fundamental role of an author and his works.

These differences indicate that an author's name is not simply an element of speech (as a subject, a complement, or an element that could be replaced by a pronoun or other parts of speech. Its presence is function in that it serves as a means of classification. A name can group together a number of texts and thus differentiate them from others. A name also establishes different forms of relationships among texts.
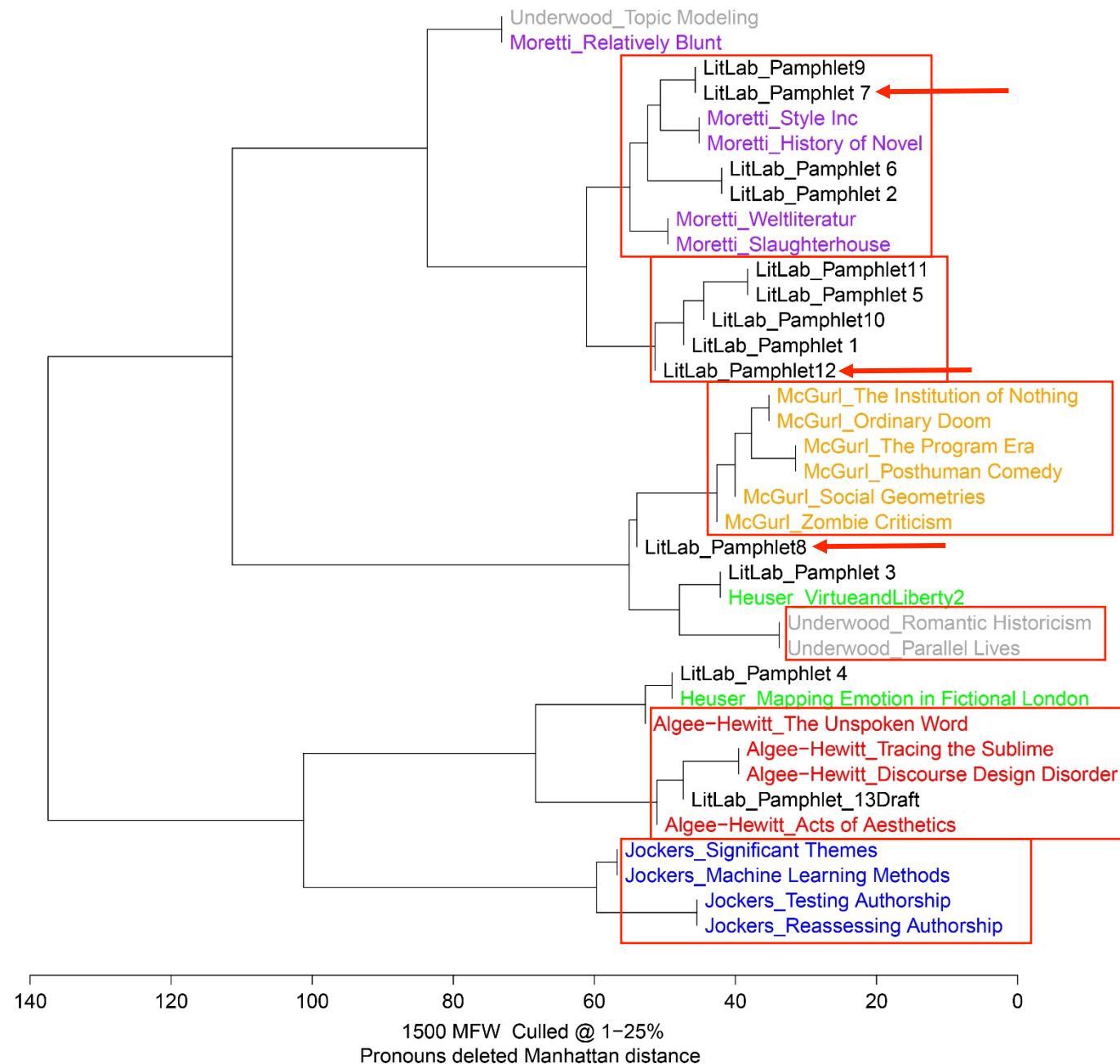
We can conclude that, unlike a proper name, which moves from the interior of a discourse to the real person outside who produce it, the name of the author remains at the contours of texts – separating one from the other, defining their form and characterizing their mode of existence.

Foucault, "What is an Author?" 1969

Pamphlets of the Stanford Literary Lab (2011-2016)

Literary Lab Pamphlet
Cluster Analysis

Underwood_Topic Modeling
Moretti_Relatively Blunt
LitLab_Pamphlet9
LitLab_Pamphlet 7
Moretti_Style Inc
Moretti_History of Novel
LitLab_Pamphlet 6
LitLab_Pamphlet 2
Moretti_Weltliteratur
Moretti_Slaughterhouse
LitLab_Pamphlet11
LitLab_Pamphlet 5
LitLab_Pamphlet10
LitLab_Pamphlet 1
LitLab_Pamphlet12
McGurl_The Institution of Nothing
McGurl_Ordinary Doom
McGurl_The Program Era
McGurl_Posthuman Comedy
McGurl_Social Geometries
McGurl_Zombie Criticism
LitLab_Pamphlet8
LitLab_Pamphlet 3
Heuser_VirtueandLiberty2
Underwood_Romantic Historicism
Underwood_Parallel Lives
LitLab_Pamphlet 4
Heuser_Mapping Emotion in Fictional London
Algee−Hewitt_The Unspoken Word
Algee−Hewitt_Tracing the Sublime
Algee−Hewitt_Discourse Design Disorder
LitLab_Pamphlet_13Draft
Algee−Hewitt_Acts of Aesthetics
Jockers_Significant Themes
Jockers_Machine Learning Methods
Jockers_Testing Authorship
Jockers_Reassessing Authorship

140   120   100   80   60   40   20   0

1500 MFW Culled @ 1−25%
Pronouns deleted Manhattan distance

Literary Lab Pamphlet

Bootstrap Consensus Tree

Underwood Mapping Emotion in Fictional London

Underwood Romantic Historicism

Underwood Parallel Lives

Heuser Virtue and Vice

LitLab Pamphlet 4

LitLab Pamphlet 3

Algee-Hewitt The Unspoken Word

Jockers Significant Themes

LitLab Pamphlet 8

Moretti Relatively Blunt

Underwood Topic Modeling

Jockers Machine Learning Methods

Jockers Testing Authorship

Jockers Reassessing Authorship

LitLab Pamphlet 13Draft

Algee-Hewitt Acts of Aesthetics

Algee-Hewitt Tracing the Sublime

Algee-Hewitt Discourse Design Disorder

McGurl Zombie Criticism

McGurl The Program Era

McGurl Posthuman Comedy

McGurl Social Geometries

McGurl The Institution of Nothing

McGurl Ordinary Doom

LitLab Pamphlet 2

LitLab Pamphlet 6

Moretti History of Novel

Moretti Style Inc

LitLab Pamphlet 7,9

LitLab Pamphlet 8

Moretti Slaughterhouse

Moretti Weltliteratur

LitLab Pamphlet12

LitLab Pamphlet10

LitLab Pamphlet 1

LitLab Pamphlet 5

LitLab Pamphlet11

50−1450 MFW Culled @ 1−25%
Pronouns deleted Manhattan distance Consensus 0.5

But the word [pamphlet] captured the euphoria of being on our own; the freedom to publish what we wanted, when and how we wanted: short, long, and even *very* long, our pamphlets never come out a minute earlier than they're ready, nor a minute later, either; and without going through the grinder of editing "styles".

A scientific essay, composed like a Mahler symphony: discordant registers that barely manage to coexist; a forward movement endlessly diverted; the easiest of melodies, followed by leaps into the unknown. I have often tried to write like this, and always failed. Then, with the pamphlets, the form has suddenly emerged.

Franco Moretti, "Literature, Measured", Pamphlet 12

Literary Lab Pamphlet
Bootstrap Consensus Tree

Underwood Parallel Lives

Underwood Romantic Historicism

Heuser Mapping Emotion in Fictional London

Heuser Virtueand Liberty2

LitLab Pamphlet 3

LitLab Pamphlet 4

Jockers Machine Learning Methods

Jockers Testing Authorship

Jockers Reassessing Authorship

Algee-Hewitt The Unspoken Word

Jockers Significant Themes

LitLab Pamphlet8

Moretti Relatively Blunt

Underwood Topic Modeling

LitLab Pamphlet 13Draft

Algee-Hewitt Acts of Aesthetics

Algee-Hewitt Tracing the Sublime

Algee-Hewitt Discourse Design Disorder

McGurl Zombie Criticism

McGurl The Program Era

McGurl Posthuman Comedy

McGurl Social Geometries

McGurl The Institution of Nothing

McGurl Ordinary Doom

LitLab Pamphlet 2

LitLab Pamphlet 6

Moretti History of Novel

Moretti Style Inc

LitLab Pamphlet 7 9

LitLab Pamphlet8

Moretti Slaughterhouse

Moretti Weltliteratur

LitLab Pamphlet12

LitLab Pamphlet10

LitLab Pamphlet 1

LitLab Pamphlet11

LitLab Pamphlet 5

50−1450 MFW Culled @ 1−25%
Pronouns deleted Manhattan distance Consensus 0.5

The graph of *Pride and Prejudice* does indeed exhibit these "waves", but with the added benefit of turning the abstract concept into a concrete image – not only a wave, but a wave with a specific length and degree.

Holst Katsma, "Loudness in the Novel," Pamphlet 7

What is done is never undone; the plot as a system of regions; the hierarchy of centrality that exists among characters; finally – and it's the most important thing of all, but also the most difficult – one can *intervene* on a model; make experiments.
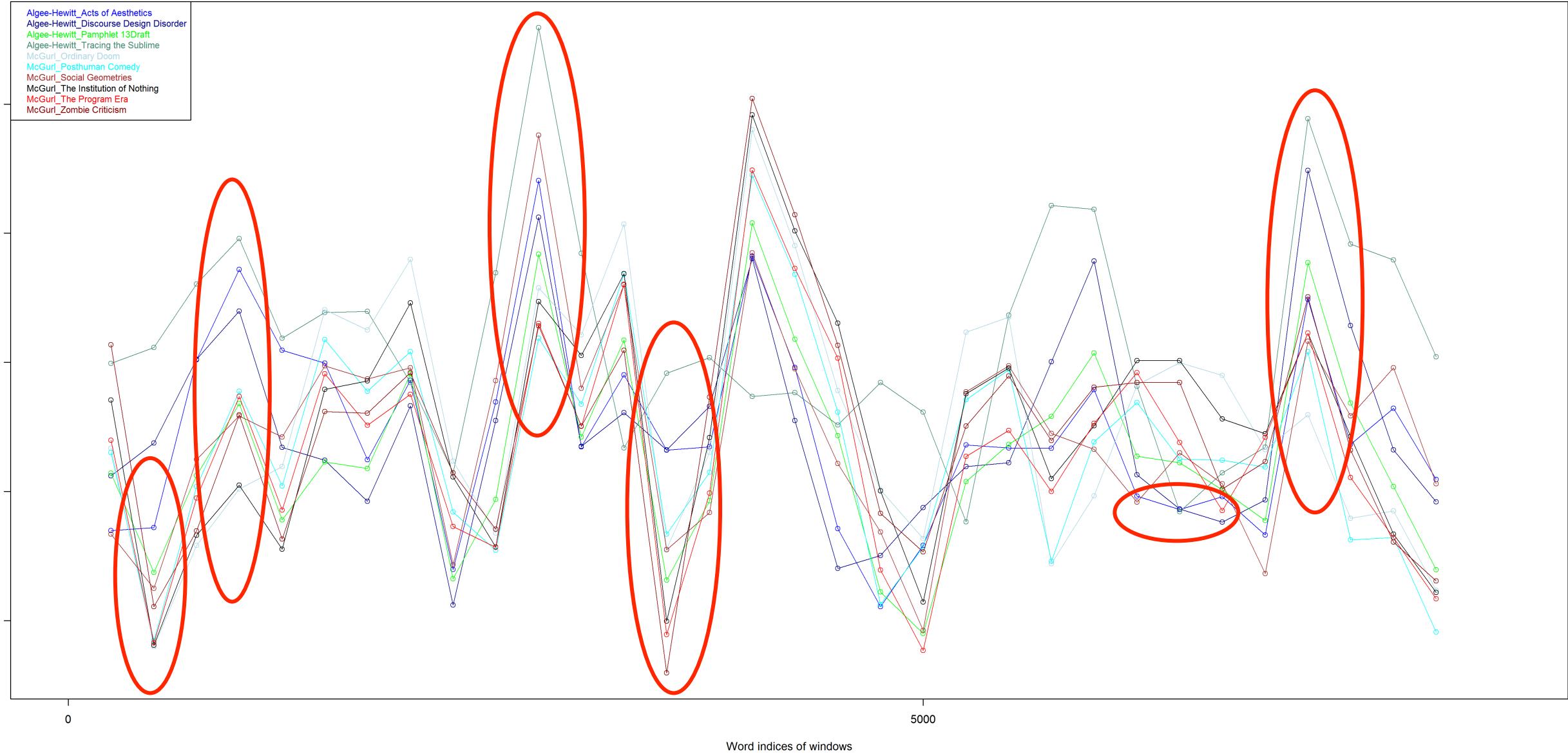
Franco Moretti, "Network Theory, Plot Analysis," Pamphlet 2

Literary Lab Pamphlet
Bootstrap Consensus Tree

Underwood Parallel Lives

Underwood Romantic Historicism

Heuser Mapping Emotion in Fictional London

Heuser Virtueand Liberty2

LitLab Pamphlet 3

LitLab Pamphlet 4

Jockers Machine Learning Methods

Jockers Testing Authorship

Jockers Reassessing Authorship

Algee–Hewitt The Unspoken Word

Jockers Significant Themes

LitLab Pamphlet8

Moretti Relatively Blunt

Underwood Topic Modeling

LitLab Pamphlet 13Draft

Algee–Hewitt Acts of Aesthetics

Algee–Hewitt Tracing the Sublime

Algee–Hewitt Discourse Design Disorder

McGurl Zombie Criticism

McGurl The Program Era

McGurl Posthuman Comedy

McGurl Social Geometries

McGurl The Institution of Nothing

McGurl Ordinary Doom

LitLab Pamphlet 2

LitLab Pamphlet 6

Moretti History of Novel

Moretti Style Inc

LitLab Pamphlet 7

LitLab Pamphlet 9

Moretti Slaughterhouse

Moretti Weltliteratur

LitLab Pamphlet12

LitLab Pamphlet10

LitLab Pamphlet 1

LitLab Pamphlet 5

LitLab Pamphlet11

50−1450 MFW Culled @ 1−25%
Pronouns deleted Manhattan distance Consensus 0.5

Rolling Delta of Literary Lab Pamphlet 8: "Between Canon and Corpus"
Co-Authors: McGurl and Algee-Hewitt

Legend:
- Algee-Hewitt_Acts of Aesthetics
- Algee-Hewitt_Discourse Design Disorder
- Algee-Hewitt_Pamphlet 13Draft
- Algee-Hewitt_Tracing the Sublime
- McGurl_Ordinary Doom
- McGurl_Posthuman Comedy
- McGurl_Social Geometries
- McGurl_The Institution of Nothing
- McGurl_The Program Era
- McGurl_Zombie Criticism

Y-axis: Delta(centroid, window)

X-axis: Word indices of windows

To the potential pitfalls of sampling, representativeness and quantifying in general, we have added the problems of ranking and valuation.

Algee-Hewitt and McGurl, "Between Cannon and Corpus," Pamphlet 8

To the concept of the son, which seemed exhausted, he added the complexities of calamity and evil.

Jorges Luis Borges, "Three Versions of Judas," 1944

Properly stated, the original question here is not "Who is the author of X?" but "Do the entries in this scatter-plot fall into any intelligible pattern?"

Burrows, Questions of Authorship, 2001