

Frequently Cited Passages Across Time: New Methods for Studying the Critical Reception of Texts

Jonathan Reeve, Sierra Eckert, Milan Terlunen

Department of English and Comparative Literature, Columbia University

Research Questions

We use text reuse detection to identify frequently cited passages from George Eliot's novel *Middlemarch* in a corpus of over 4,000 critical articles (1950-2015) obtained from JSTOR that contain the word "Middlemarch." We attempt to answer questions like:

- What passages of the novel are critics most interested in quoting? How do these critical favorites change over time?
- Can we detect a life-cycle for certain passages which rise and fall in popularity over time?
- What are the linguistic properties of quotations, as compared with non-quotations?
- What makes us, as readers, choose to quote one passage instead of another? Can this be connected to major movements within the field (e.g. feminist criticism, deconstruction), institutional structures (eg. particular journals) or even to a specific critic?

Methods

We use a custom-built text reuse detection algorithm, text-matcher, that uses a techniques of approximate text matching in a two-pass process to identify similar passages between two texts. The first pass uses sequence matching to find a single matching trigram between the two texts. The algorithm then compares Levenshtein distances. We then visualize these results at two different scales: first, through a heat map visualizng citations at the book and chapter level and second, through a text browser, visualizing citations at the sentence and paragraph level. We plot these results at several different scales, revealing several striking patterns discussed in the following sections.

Frequently Quoted Chapters

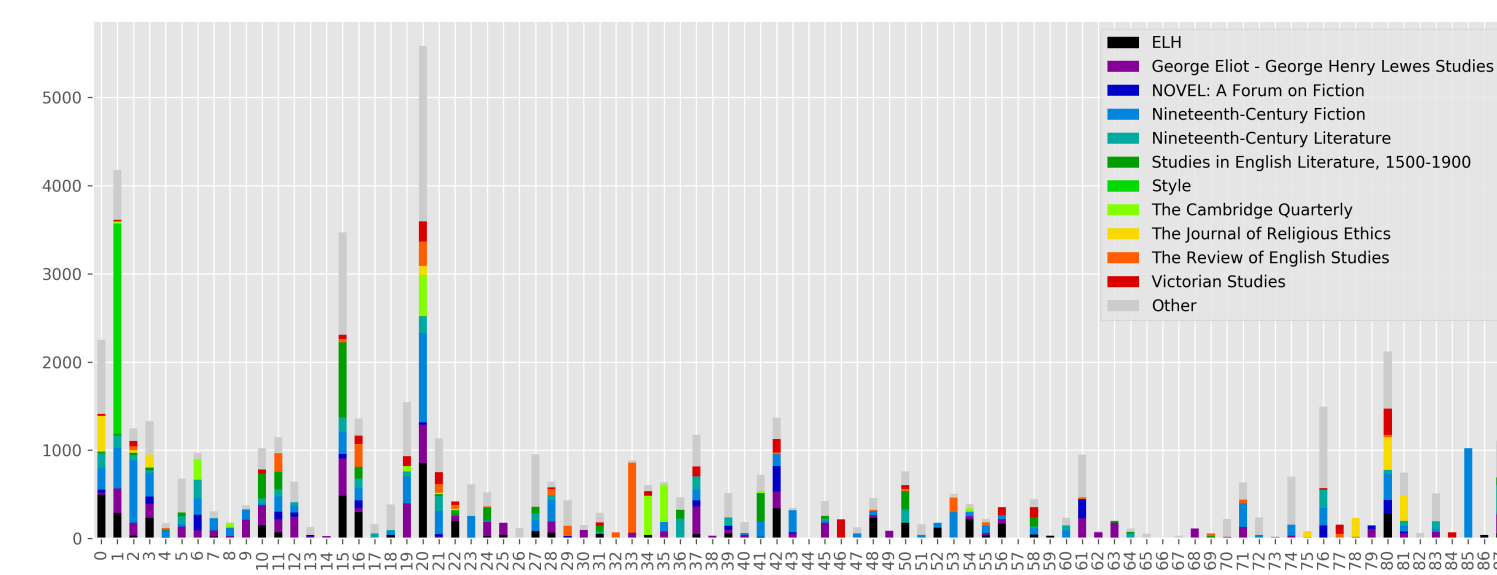


Figure 1: Quoted words by chapter, subdivided by journal.

Journals play a role in determining the relative number of citations of a given chapter, *Style*, for instance, contributes heavily to the number of Chapter 1 citations. Critical attention, across journals, tends to favor the first half of the novel.

Literary Critical Specialization

Literary critics writing in the specialist journal *George Eliot / George Henry Lewes Studies* quote different passages from their counterparts publishing in more generalist journals. A positive score here indicates a chapter is quoted more by specialists:

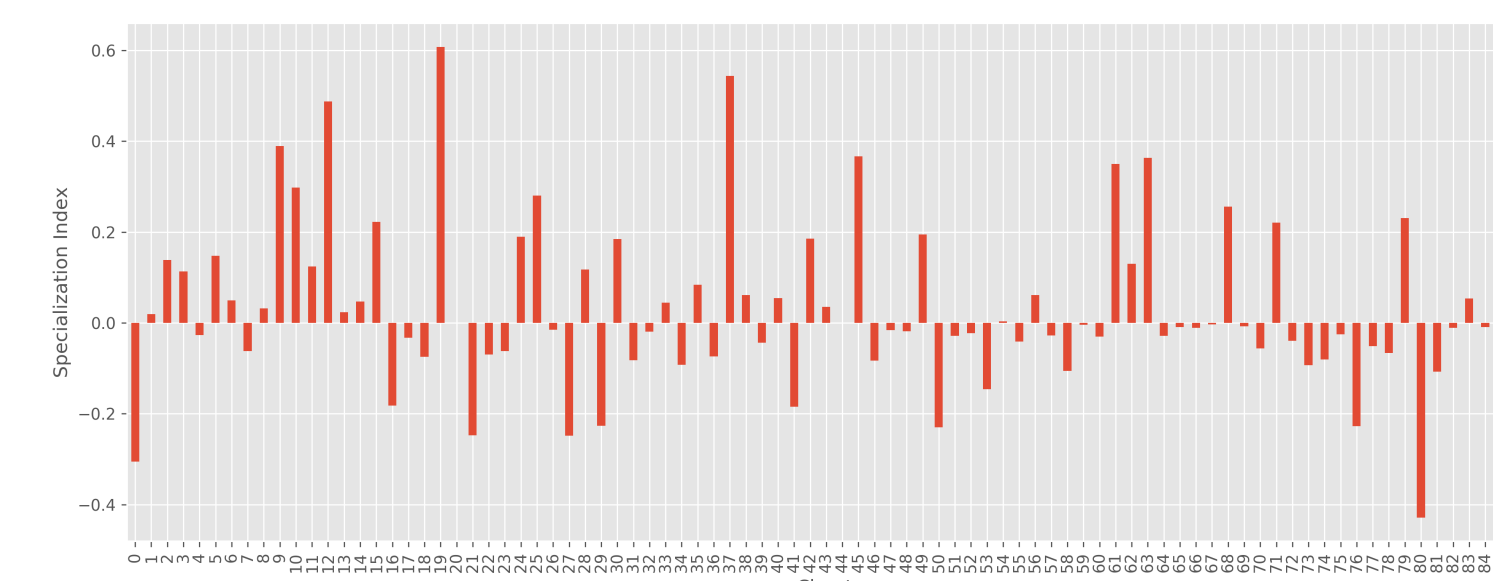


Figure 2: Chapters quoted, specialists vs. non-specialists.

Properties of Quotations

Figure 5 shows characteristic parts of speech of quotations (positive), and non-quotations (negative).



Figure 3: Distinctive parts of speech of quotations

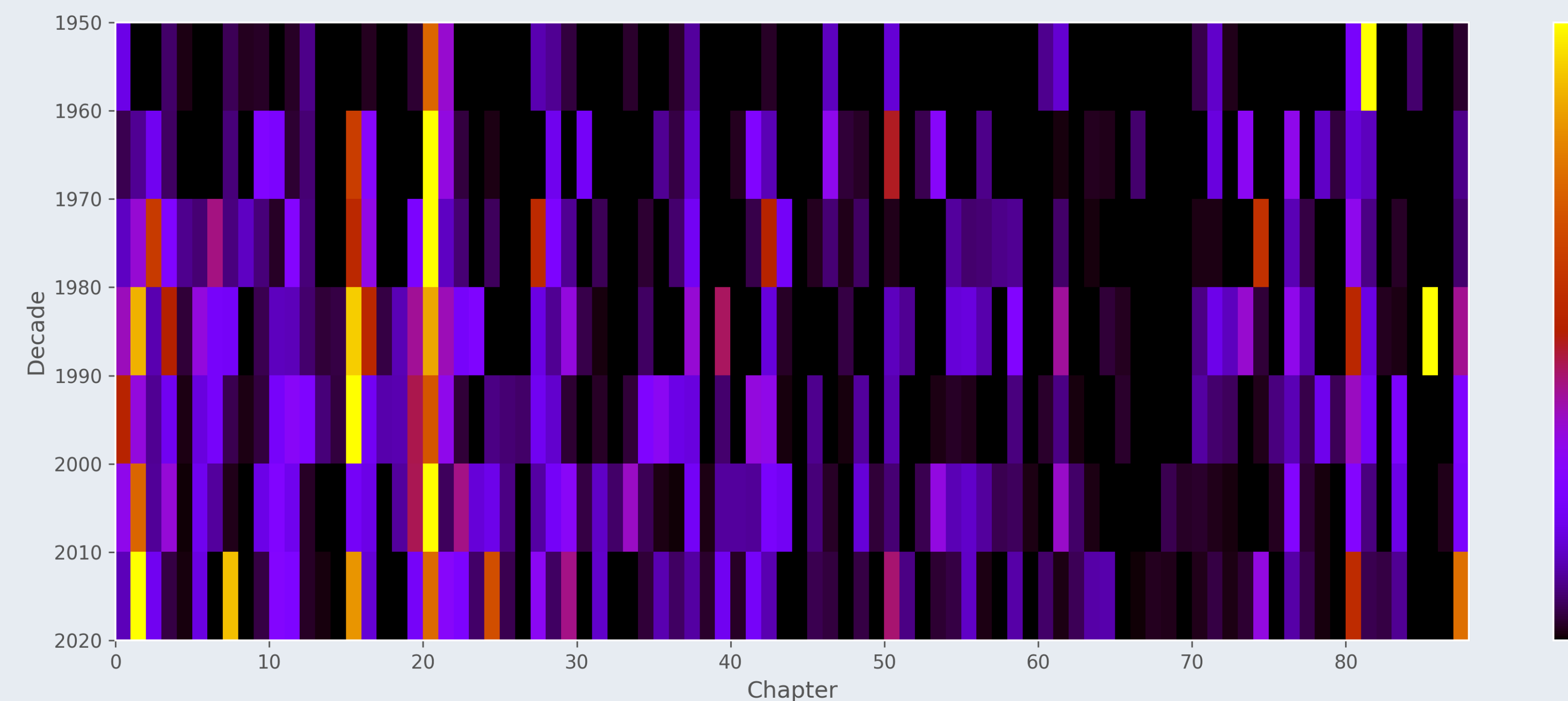
Table 1 shows most distinctive words of non-quotes (negative, above), and quotes (positive, below):

Word	Freq. Diff.
say	-0.006397
lydgate	-0.004499
mr.	-0.003395
bulstrode	-0.002948
fred	-0.002781
rosamond	-0.002616
eye	0.001024
live	0.001042
consciousness	0.001054
great	0.001080
new	0.001134
woman	0.001154
world	0.001883
light	0.002183
life	0.002417

Table 1: (Non-)Distinctive words of quotations

Middlemarch: Shifts in Critical Attention over Time

Viewed diachronically, trends arise in critical quotations of the novel. Critics' favorite chapters to excerpt shift from Chapter 81 in the 1950s to Chapter 20 in the 60s-70s, and to Chapter 15 in the 80s-90s.



From the Critical Heatmap Edition: Most Quoted Passage

is not unusual. I nat element of tragedy which lies in the very fact of frequency, has not yet wrought itself into the coarse emotion of mankind; and perhaps our frames could hardly bear much of it. If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heart beat, and we should die of that roar which lies on the other side of silence. As it

Links and Contacts

- Project page: github.com/lit-mod-viz/middlemarch-critical-histories
- Jonathan Reeve: jonathan.reeve@columbia.edu
- Sierra Eckert: sierra.eckert@columbia.edu
- Milan Terlunen: m.terlunen@columbia.edu