Grasping Rhetoric and Composition by Its Long Tail: What Graphs Can Tell Us about the Field’s Changing Shape

Presented as a series of graphs, bibliographic data gathered from College Composition and Communication provides perspective useful for inquiring into the changing shape of the field as it continues to mature. In its focus on graphing, the article demonstrates an application of distant reading methods to present patterns not only reflective of the most commonly cited figures in CCC over the past twenty-five years, but also attendant to a steady increase in the breadth of infrequently cited figures.

Inventorying an Epistemic Court

Nearly two decades ago, Donna Burns Phillips, Ruth Greenberg, and Sharon Gibson inquired into rhetoric and composition’s maturation using methods of counting and sorting to distinguish various subsets of aggregate data drawn from College Composition and Communication. Phillips, Greenberg, and Gibson’s
Essentially, I contend that graphs help us see with fresh perspective continually unfolding tensions among specialization, the interdisciplinary reach of rhetoric and composition, and the challenges these present to newcomers to the scholarly conversation.

In an effort to update and contribute further to the ongoing inventorying of rhetoric and composition’s “epistemic court,” the following article adopts a similar exigency to that heeded by Phillips, Greenberg, and Gibson as it uses bibliometrics and graphing as a means of articulating different perspectives on the changing nature of “who [has been] speaking” over the past twenty-five years, according to citation frequencies in *College Composition and Communication*. Essentially, I contend that graphs, as a form of distant reading (Moretti), help us see with fresh perspective continually unfolding tensions among specialization, the interdisciplinary reach of rhetoric and composition, and the challenges these present to newcomers to the scholarly conversation.

Rather than characterizing at the outset the consequences of specialization and interdisciplinary borrowing as generally positive or negative, I seek to demonstrate how graphs can function as a productive, suasive abstracting practice that will allow us to look more carefully at what has happened to citation practices in *CCC* from 1987 to 2011. Toward this end, first I say more about the studies using graphs and relevant quantitative methods to understand journals and the fields sponsoring them. Doing so highlights two principles of distant reading that graphs foreground: 1) deliberately altering scale allows us to see aggregate patterns linking details and non-obvious phenomena, and 2) the systematic compilation of replicable data may empirically corroborate local, tacitly felt impressions about changing disciplinary conditions. The sec-
ond half of this article adopts as an exploratory framework Chris Anderson’s work on long tails (Pareto distributions) and presents graphs drawn based on a compilation of 16,726 citations in 491 journal articles published in CCC over twenty-five years. Departing from studies of citation that have focused exclusively on the most frequently referenced figures, I argue that graphing the relationship between the most frequently cited figures and the changing distribution of infrequently referenced figures produces a unique perspective on a changing disciplinary density of great relevance to specialists, generalists, and initiates alike.

**Precedents for Graphing and Quantification: Accounting for Scholarly Activity**

Graphing and the methods of quantification at their foundation have precedents in composition studies. For example, Maureen Daly Goggin’s well-known history of the field, *Authoring a Discipline: Scholarly Journals and the Post–World War II Emergence of Rhetoric and Composition*, presents eight graphs, each designed to render apprehensible some dataset aggregated (manually) from the nine journals at the center of her study. Goggin’s study is one notable example where graphs have been applied similar to the way Franco Moretti uses them to study literary genres in historical contexts: to deliberately alter the level of detail at which texts are customarily read with the aim of connecting overlooked minutiae and broader phenomena. Five of Goggin’s eight graphs account for some criterion applied to all contributing authors for nine major journals from 1950 to 1990: a pair of line graphs showing affiliations of authors to two- and four-year institutions, an area graph showing institutional affiliation (public or private, college or university), a horizontal bar graph presenting the number of contributors from departments other than English, and a two-line graph drawn according to the gender of contributors to the journals. Goggin also uses a vertical bar graph to show MLA membership by geographic region and a pair of line graphs for the percentages of conference papers published in CCC and *College English*. Although *Authoring a Discipline* is unreflective about the use of graphs, Goggin provides strong examples of graphing methods that, because they change the scale of detail, help us engage with patterns of disciplinary activity that would otherwise be difficult to discern, particularly for newcomers to the field.

A deliberate adjustment in the level of detail at which we ordinarily experience texts: this is a key motive when producing graphs as a distant reading...
method, and it is a common tactic for mediating large datasets, including scholarly corpora. Experientially, reading tends to be a local, direct encounter, typically involving (or demanding) an identifiable, focal text. While there are sure to be exceptions (a bibliographic essay, for instance, pursues a similar purpose: synthesis by reduction), there exists a default level of detail commonly associated with reading. For traditional scholarly journals, the default scale is the article, and more specifically its words, sentences, and paragraphs. Print journals already include numerous features designed to help readers assess smaller-scale units, such as the issue and article, before reading more thoroughly. A simple table of contents, for example, supports a glancing sort of distant reading at one scale, and article abstracts allow for distant reading at a scale only slightly closer to the stuff of the article than the title and author listing. Readers rely on these devices to make quick decisions about whether to read a particular article or not, but reading the journal through these devices alone is not quite the same as reading a scholarly journal in the common sense of the activity.3

Reading across a series of journal articles gathered around a particular research question, one might or might not notice variations in the lengths of the articles or patterns in the number of sources cited in each. And yet, article lengths have changed significantly over twenty-five years, as have the number of sources referenced in a given article, issue, or volume on average. Readers might notice trends related to these mundane details across a collection of research (whether it is random or more purposefully gathered), but these details are, nevertheless, transparent and easily tabulated. Graphs allow us to zoom out, to see patterns in length and citation count across a selection of articles. This illustrative exercise in distant reading renders tangible those patterns that almost certainly go unrecognized (except intuitively) when we read at the default scale, picking up a few articles at a time.

Figure 1 employs graphing methods to present growth patterns in the number of citations and the number of citations in CCC over twenty-five years. This graph illustrates just one form of knowledge available to us in quantification, in the distant reading that comes of counting, recording tallies, and plotting coordinate points. The figure presents two sets of data: the lower area accounts for page counts by year; the upper area accounts for citation counts by year. Indeed, over the past twenty-five years, articles published in CCC have
grown longer and also draw more extensively on source material appropriate to include in a list of works cited.

Besides suggesting a gradual inflation in the page count and number of sources in scholarly articles published in *CCC* over the last twenty-five years, Figure 1 also elicits questions. This inventive and generative capacity creates a heuristic with unmistakable bearing on questions about the field’s formation as well as the material and discursive bases for disciplinary maturation. Why have page counts and citation counts nearly doubled in twenty-five years? Do other journals exhibit similar trends over the last twenty-five years? What about journals in other fields? How high will annual page counts and citation counts climb before leveling off? Or before significantly altering the work entailed in reading or writing a scholarly article? What culpability in this trend do a journal’s stakeholders bear, from publishers, editors, and editorial board members to reviewers, writers, and readers? Absent distant reading methods, these questions would still warrant hunches and speculation, but we would be unable to present the pattern as compellingly. The graphed pattern renews the questions with vivid presentational force.

Figure 1. A plot graph indicative of page count and citation count by year in *CCC*, between 1987 and 2011. Trend lines for page count (solid) and citation count (dotted) indicate the gradual but steady increase in these basic features of the journal.
Graphs hold open these questions and give us different ways of grasping non-obvious trends.

Related quantitative studies foreground a second point about the potentials of distant reading and graphing techniques for the rhetoric and composition scholars: the bearing of quantitative methods on the stabilization of reusable, interoperable, field-wide datasets. Phillips, Greenberg, and Gibson presented a history of similar scope and quantified basis (i.e., the counting of citations, the listing of editors, etc.) to Goggin’s book-length project. There is a high degree of overlap between Goggin’s interest in elucidating patterns and the aims that justify their pursuit: chronicling the discipline’s genesis. Phillips, Greenberg, and Gibson relied on tables and historical narrative rather than graphs to deliver their findings; however, even by simple quantification they were able to distinguish patterns related to who has published most frequently, how citation counts have steadily (perhaps quietly) climbed, and who, at fifteen-year intervals, has been cited most frequently. They speculate, from these tallies, about the causes for the rising rates of citation: “There is a dramatic contrast in the number of citations between early and recent CCC issues, attributable to the developing body of composition scholarship, the maturing of the field, the increasing demand for theoretical grounding of pedagogical practice, and the political necessity for supporting the professionalism of the discipline” (451). Whether this speculation takes hold as a viable, long-lasting theory is less important than is the way these methods catalyze questions and begin to provide a means of addressing such questions more systematically than is otherwise available. The demonstration of graphs renews points Phillips, Greenberg, and Gibson raised concerning the field’s development and maturation, the growing demand for theoretically sophisticated composition scholarship, and the complicated politics of citation, as well as related matters, such as pressures to publish, the competitive nature of traditional publishing, and citation as a function of ethos insofar as it represents the sources one has taken into consideration. Graphs hold open these questions and give us different ways of grasping non-obvious trends.

**Methods**

With a few key distinctions, the methods I have used to compile the frequency of citations appearing in CCC from 1987 to 2011 are similar to those applied by Phillips, Greenberg, and Gibson, who recorded references appearing in any piece of work published in CCC between 1950 and 1993, including reviews, interchanges, and features unique to a particular editor (e.g., Ken Macrorie’s
“Miscellany”). I have focused exclusively on articles—items likely to have been anonymously peer-reviewed following that change to the journal in 1987 and that adhered to the roster-like listing of works cited appearing in alphabetical order at the end of the article (a convention that was introduced to the journal at nearly the same time as anonymous peer review). Thus, this study includes every citation listed in association with the 491 articles published in CCC from 1987 to 2011, amounting to a comprehensive record of 16,726 works cited entries.4

I prepared the list by gathering all of the lists of works cited for each article in a single spreadsheet.5 Because individual works-cited entries often include multiple authors (or clipped lists of authors denoted with et alia), the citation list required considerable smoothing. I coded each bibliographic entry in the list so that listings with editors, et alia, dashes (repeated reference to a single author), and nonstandard authorship could be sorted apart from author listings.6 I removed the citation entries for editors, replaced the dash placeholders with the full names of the appropriate authors, and replaced the et al. abbreviation with the names of all authors collaborating on a given work. Finally, using text-matching algorithms followed by manual proofreading, I checked the list to ensure correct spelling and name formatting. These alterations to the comprehensive works cited list resulted in an extensive roster of authors whose names appeared with every instance of a publication associated with their names.7

With each name-reference assigned to a single slot in the comprehensive listing, various tabulations were possible; the 16,726 works cited entries became a list of 19,477 name-references.8 Developed out of this dataset, Figure 2 shows the top 102 scholars sorted in descending order by the number of references made to them in CCC articles between 1987 and 2011.

The simple tabulation evokes many questions worthy of exploring more deeply in the contexts of professional development and graduate education: What is at stake in knowing or not knowing any of the figures shown here? What presences and absences are most striking? To what degree are new scholars—in such a listing as this—overshadowed by well-established ones? What are some of the intriguing juxtapositions where positions in the list are shared? This is a tangential question, but one worth considering for its inventive richness in a course that introduces graduate students to the field. Wondering about coincidental pairings is germane to a practice I think of as heuretic discipliniography, or writing and rewriting the field by exploring the intersections across different scholars’ bodies of work as well as the associated pedagogical, theoretical, and methodological approaches they mobilize.
# Out of 19,477 References
Name (461 articles)

1 145 Linda Flower (66 articles)
2 133 Peter Elbow (85 articles)
3 118 Patricia Bizzell (82 articles)
4 112 David Bartholomae (93 articles)
5 111 James A. Berlin (90 articles)
6 110 Robert Connors (78 articles)
7 102 Andrea Lunsford (74 articles)
8 101 Lester Faigley (79 articles)
9 96 Mike Rose (64 articles)
10 77 John Trimbur (57 articles)
11 73 Kenneth Burke
12 68 Sharon Crowley
13 67 Mikhail Bakhtin
14 65 Cynthia Selfe
15 62 John Hayes
16 58 Anne Ruggles Gere, Joseph Harris
18 57 Charles Bazerman, Lisa Ede
20 55 Ellen Cushman, bell hooks, Kathleen Blake Yancey
23 52 CCCC, Maxine Hairston, Stephen North
26 51 Shirley Brice Heath, Mina Shaughnessy
28 50 John Dewey, Min-Zhan Lu
30 48 Susan Miller
31 46 Marilyn Cooper, Donald Murray
33 45 Edward White
34 44 Jacqueline Jones Royster
35 43 Janet Emig, Michel Foucault, Henry Giroux, Gesa Kirsch, Geneva Smitherman
40 42 Kenneth Bruffee, David Russell
42 40 Deborah Brandt, Paul Freire, Richard Haswell

Figure 2. The 102 most frequently cited authors in CCC from 1987 to 2011.
The single, comprehensive list in Figure 2 is suggestive in its own right, but it tends to downplay temporal variation: the changing tide of citation practices at lesser increments within this twenty-five-year period. In the interest of beginning to see into this variation, consider an alternative table developed out of the same dataset (i.e., the same initial spreadsheet listing author-references).

What does Figure 3 do? Certainly it provides compelling quantitative evidence for trends and patterns in citation practices, and it foregrounds the temporal subsets within the twenty-five-year sample. In it we encounter a form of knowledge unavailable at the usual and customary scale at which journals are read—the individual article. The data as presented also lend themselves to potential analysis of the centrality of a given figure in a given period of time as well as the waning centrality of even the most frequently cited authors in the
### Figure 3. Top ten most frequently cited authors in CCC from 1987 to 2011, by five-year interval.

<table>
<thead>
<tr>
<th>Year Interval</th>
<th>Total References</th>
<th>Total Articles</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flower, Linda</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hayes, John</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lunsford, Andrea</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Connors, Robert</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizzell, Patricia</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rose, Mike</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Faigley, Lester</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bartholomae, David</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Berlin, James A.</td>
<td>20</td>
</tr>
<tr>
<td>1992–1996</td>
<td>3595</td>
<td>102</td>
<td>43</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizzell, Patricia</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rose, Mike</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dewey, John</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flower, Linda</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizzell, Patricia</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Faigley, Lester</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>28</td>
</tr>
<tr>
<td>1997–2001</td>
<td>3881</td>
<td>93</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Berlin, James A.</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Barthesolm, David</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizzell, Patricia</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Faigley, Lester</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>28</td>
</tr>
<tr>
<td>2002–2006</td>
<td>4289</td>
<td>93</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Burke, Kenneth</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flower, Linda</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Smitherman, Geneva</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trimbur, John</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bakhtin, Mikhail</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cushman, Ellen</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bartholomae, David</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>hooks, bell</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hopkins, Edwin</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Selfe, Cynthia</td>
<td>20</td>
</tr>
<tr>
<td>2007–2011</td>
<td>4957</td>
<td>94</td>
<td>26</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yancey, Kathleen Blake</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cushman, Ellen</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Russell, David</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizerman, Charles</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gere, Anne Ruggles</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Selfe, Cynthia</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Haswell, Richard</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Burke, Kenneth</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>United States</td>
<td>18</td>
</tr>
<tr>
<td>Totals, 1987–2011</td>
<td>19477</td>
<td>461</td>
<td>Flower, Linda</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Elbow, Peter</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bizzell, Patricia</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bartholomae, David</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Berlin, James A.</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Connors, Robert</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lunsford, Andrea</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Faigley, Lester</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rose, Mike</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trimbur, John</td>
<td>77</td>
</tr>
</tbody>
</table>

*Figure 3. Top ten most frequently cited authors in CCC from 1987 to 2011, by five-year interval. C: Citation count (total number of name-references). A: Article count (total number of articles in which citations appear). ^ Indicates a scholar not ranked in the top ten for the previous five-year period. This table bears direct correspondence to similar tables appearing in Stevens and Williams's work with *Critical Inquiry* and in Phillips, Greenberg, and Gibson's work with CCC from 1950 to 1993.*
most recent five-year period. In other words, we can see that Linda Flower was cited in 22 of the 79 articles (27.8%) published between 1987 and 1991; Kenneth Burke, the leading figure between 2002 and 2006, was referenced in just 9 out of 93 articles (9.7%). Further lines of inquiry include examining the lists with attention to gender, race, ethnicity, and disability; theorizing what constitutes career longevity; and exploring the relationships between bibliographic prominence and other criteria, such as national leadership roles, institutional affiliation, and areas of research. Granting all of the known limitations in what we can extrapolate about the field at large from this sample, this also suggests a change within CCC: the prominence of the top-most cited authors is gradually and relatively steadily declining.9

Simply, quantitative studies of citations spark insights and advance questions concerning the ways citation practices changed.10 Yet these methods are not without qualification. A conventional listing of citations does little to reflect the scope of the reference as it is taken up or the framing language used to introduce the source within an article itself.11 The list simply affirms presence. In other words, citation listings lack dimension; they do not report whether a single source was used extensively in, say, more than ten (or more) pages of an article or whether, on the other extreme, it was hardly mentioned at all. When aggregating a long list of references, these dimensions fall away; we are left with a basic list—a reduced, concentrated record. Also, the gathering of name-references into a single list downplays aspects of production, reception, and circulation of a source, as well as the career of the author. In their own acknowledgment of related limitations, Phillips, Greenberg, and Gibson write that we will find sharp differences between the popularity of a particular source (like Mina Shaughnessy’s Errors and Expectations) and a particular author (like Andrea Lunsford, who is cited frequently but for a wide array of different articles): “One explanation for this circumstance may be focus: Shaughnessy, for example, generally restricted her work to a single area, while Lunsford published on a variety of issues” (454). Across a given career, one author might remain highly specialized while another might shift from one area of inquiry to another, thus producing a record of scholarship more reflective of a generalist’s wanderlust.12

In their discussion of methods used to study the journal’s authors and works cited quantitatively, Phillips, Greenberg, and Gibson write:

Such quantitative measures help determine what can be considered within the community as common knowledge, and common knowledge is the power base.
Writers will construct their discourse around what their audiences can be assumed to know and accept. Researchers will see the investigative techniques as models. Initiates will ingest this core as part of the membership rite. CCCC members will rely on name recognition in the elections shaping the organization that molds the field. In sum, work associated with these names becomes the traditional paradigm, and all subsequent work moves toward its support, its enlargement, or its overthrow. (454)

So while quantitative studies of authors cited in a well-known journal may offer a reasonable indication of the “common knowledge” of the field, this approach must not appear to produce a definitive roster of influences on the discipline. Compilations drawn from lists of citations might prompt us to wonder about the kinds of knowledge formal references demand of a reader, and a wide variety of contextualizing techniques within the articles themselves are sure to help familiarize readers with those voices brought into the piece—whatever the motive. The lists presented in Figures 2 and 3 indicate frequency—a convergence, possibly, of popularity, notoriety, and influence. Lists like these are powerful indications of the “hits” in composition studies. Who, identifying themselves with the field of composition studies, would claim to know none of these figures or the impact of their work? And yet, the top 102 authors cited over the last twenty-five years in CCC or the top ten authors by five-year increment—even though they are indicative of certain currents in a disciplinary conversation—do not tell us enough about what has happened across the entire sample of name-references in the set of citations. Turning to graphs based on the entire dataset, there is more we can know from this quantitative approach. The well-known influences, after all, are likely to rank relatively high in a comparable sample of citations drawn from other journals in the discipline—although this research, like so much work with large datasets significant to the field’s formation, remains to be done. To make sense of the comprehensive record of citation within CCC in this twenty-five-year period—to corroborate the degree of dappledness, we must look not only at what has changed among the top-most cited figures. We must also come to terms with what has happened among those sources invoked infrequently in a twenty-five-year period—those who, by the record of citation frequency, registered a singular appearance.
by the record of citation frequency, registered a singular appearance. For this question, another series of graphs proves insightful, enabling a deeper inquiry into just how cacophonous the epistemic court has become.

**Too Dappled a Discipline? Graphing the Long Tail of Author Citation**

_In our fixation on star power, we cheer the salary inflation of A-listers and follow their absurd public lives with an attention that far exceeds our interest in their work. From the superstar athletes to celebrity CEOs, we ascribe disproportionate attention to the very top of the heap. We have been trained, in other words, to see the world through a hit-colored lens._

_Chris Anderson, The Long Tail_

In October 2004, Chris Anderson, an editor at _Wired Magazine_, reached out to a popular audience in his article “The Long Tail” with arguments about how economic notions of scarcity and abundance have been overhauled with digital commerce. The article distinguished between spatially constrained traditional retailers and their comparably abundant online counterparts. According to Anderson, the typical Borders bookstore, for example, carried 100,000 titles, but its leading online competitor, Amazon.com, offers a vastly deeper (i.e., larger) selection to consumers: more than 3.7 million titles. Anderson’s research, which he expanded into a full-length monograph in 2006, centers on the idea of the long tail: the niche interests that online markets now reach. *The Long Tail* is an extended inquiry into the phenomenon of these market niches—how they work, how digital circulation stimulates them, and how they have fundamentally challenged more conventional storefront economics.

Anderson didn’t come up with the idea for the long tail himself. In his work, he cites many influences on his thinking from economics and technology studies. But his timely insights and striking examples certainly have done much to popularize the concept in recent years. Conceptually, the long tail comes from statistics and graphing; it is a feature of a power law or Pareto distribution—graphed patterns that underscore the uneven distribution of some activity or phenomenon. In his article-length work on the long tail, Anderson provides the composite graph shown in Figure 4, “Anatomy of the long tail,” to illustrate.
Here, music is the focal premise. Wal-Mart, like Borders, offers a limited selection of music; even while the discount retail giant provides a large selection of hits, it simply cannot match what an online competitor, such as Rhapsody, makes available: those less popular titles that continue to sell actively, despite ranking well beneath the threshold of popularity that justifies the entirety of Wal-Mart’s stock. From left to right, the graphed distribution accounts first for the high-ranking hits commonly available on store shelves; gradually it gives way to the long tail—the rich expanse of less-popular albums and tracks that continue to sell at markedly lower rates than their counterparts at the head of the curve. The long tail’s recurrent niches are thin but extensive; thus, it represents a formidable base for economic activity untouched by conventional store-shelf retailers and what Anderson calls the “tyranny of geography” (17).

Anderson’s early work on long tails focused explicitly on these market trends; he later adapted the premise to look into patterns in media and entertainment sales. Yet, Anderson also acknowledges that his research has opened up to even broader possibilities for the long tail as an apparatus for descriptive statistics: “Seen broadly, it’s clear that the story of the Long Tail is really about
the economics of abundance—what happens when the bottlenecks that stand between supply and demand in our culture start to disappear and everything becomes available to everyone” (11). He arrived at an expanded view of the long tail, one that recognizes that its application reaches beyond economic matters to cultural phenomena. Assuming a similarly broad view of the long tail, I contend that it serves suggestively as a basis for graphing the citation-frequency data introduced earlier so that we can make sense not only of what has happened in CCC to those names mentioned most often (i.e., the hits) but also what has happened to the long tail of author citation over the last twenty-five years. The top-ranking author citations in CCC between 2007 and 2011 are fewer than they were for the same period of time twenty years earlier, yet graphing all of the author citations not only confirms what we already know about those at the top but also sheds light on what has happened to the long tail of author citation—those names appearing just once or twice in the journal’s works cited—during the same periods. Once more, graphing functions as a form of distant reading and a necessary means of engaging with large-scale data. The long tail shows how an abstract visual model potentially elicits new insights and, with its descriptive acuity, raises new questions, some of which might help explore the continuing genesis and maturation of the field.

Why the Tail Matters: The Names Invoked Just Once

[I]t is too easy to overlook elements of our history that reinforce and enrich our current work. We are too prone to let superficial differences blind us to significant connections between past and present.

Lee Odell, “Diversity and Change: Toward a Maturing Discipline”

What do Maya Angelou, Andy Rooney, Bill Gates, Queen Hatshepsut, Roger Ebert, and Elvis Presley have in common? Despite being well-known figures, each of them was cited in CCC just once between 1987 and 2011. Thus, they inhabit the long tail of CCC citation; these six figures share this distinction with 5,761 other names referenced just once (out of 8,035 unique name-references in the twenty-five-year period in question). Another 986 names appear among cited works just twice. This leaves little more than 1,287 names (16.0%) that
appear in *CCC* citations three or more times within the twenty-five-year sample. By assigning these figures to a simple graph, we can see that they follow a power law, meaning relatively few names rank high in citation frequency (see Figures 2 and 3 above for specific references) while more than 80 percent of the names register a momentary appearance, usually appearing in a single article. Rendered into a graph, the twenty-five-year data sample appears visually, converting quantitative measures into an abstract model with qualitative effects.

As a model, the long tail helps us recognize just how thin a sliver of citations are captured in the list of the top 102 author-references shown in Figure 2. Attending to the immense shelf of the less frequent citations in the dataset demands a more comprehensive view—more distant view, that is. We must step back even further than did Phillips, Greenberg, and Gibson to realize, on the one hand, the limits of a hit-driven view of citation activity and, on the other hand, the ever-fuller breadth of activity that manifests in the long tail. Certainly the figures at the top tell

Certainly the figures at the top tell *something* about citation practices and centrality in the journal’s scholarly conversation; however, the larger number of figures at the bottom indicates *something more*. It is, after all, in this long, flat expanse of unduplicated references that we can begin to assess just how broad-based the conversations (in a given journal) have grown—and just how much the centered, coherent, and familiar locus of conversation, based on citation practices, has slid.

![Graph of Author-references in CCC, 1987-2011](image)

*Figure 5. References to unique names in CCC works cited from 1987 to 2011.*
us something about citation practices and centrality in the journal’s scholarly conversation; however, the larger number of figures at the bottom indicates something more. It is, after all, in this long, flat expanse of unduplicated references that we can begin to assess just how broad-based the conversations (in a given journal) have grown—and just how much the centered, coherent, and familiar locus of conversation, based on citation practices, has slid.

Keeping in mind this more general thread of inquiry into the maturation of the field, subdividing the graph above into five-year increments—much the same as with Figure 3—foregrounds the incremental development of Figure 5: how and at what rates did it grow? Did the vertical portion at the left first spike sharply from the horizontal axis? Has the tail always been as proportionately long? Have the two ends grown at relatively consistent paces since 1987? For this, consider a more nuanced series of graphs, each displaying a five-year data subset.

The series of graphed distributions at five-year increments highlights a gradual transformation while also confirming that since 1987, even as the total number of citations climbed higher in each subsequent five-year period, the once-steep grade has flattened out considerably. As the scholarly record grows, authors have a more complex array of sources to drawn upon. Across twenty-five years of citation activity in CCC, the long tail has grown longer, indeed, while the head has dwindled over time.

Figure 6. Citation frequency in CCC, 1987–1991.
Figure 7. Citation frequency in CCC, 1992–1996.

Figure 8. Citation frequency in CCC, 1997–2001.
Figure 9. Citation frequency in CCC, 2002–2006.

Figure 10. Citation frequency in CCC, 2007–2011.
Figures 6–10 present a chronographic report on the evolution of one sample from the field’s scholarly record, and this evolution would be easy to overlook if we fixated only on the most frequently cited figures or if we relied on experiential impressions of the journal. Approaching the full record in this way allows us to apprehend these gradual transformations—shifts so subtle that it is easy, at the scale of a career, to disregard, such as the creeping publishing trends to which so much contemporary scholarship has fallen in step. Thus, we can use distant reading methods to understand with more granularity factors affecting citation distribution. Furthermore, although at this stage the graphs I have produced reflect the full dataset, it is possible to use these methods to isolate and compare smaller segments of the data. Separating subsets of the citation data would allow us to search for patterns according to many different criteria, exploring, for instance, the frequency of citation made to work by scholars within the first five or ten years of their careers, to work by alums of specific graduate programs, or by scholars whose research focuses on a specialized area.

A changing citation frequency also affects the depth and variety in what one reads. The reading problem—a problem of “keeping up with new work” acknowledged by Richard Lloyd-Jones in his 1977 CCC Chair’s address (50)—remains a contemporary challenge not only for newcomers to the discipline but also for those who have spent many years actively practicing and participating in the field themselves. Even self-described generalists, in those moments when they are again reminded of the Sisyphean demands of the field’s ongoing quality, inevitably experience (if indirectly, by felt sense) the lengthening of the long tail as a burdensome certainty: the unyielding march of time coupled with the burgeoning material resources piling up in the disciplinary commons. In economics, the long tail is sometimes called the heavy tail. The tail is, in this sense, paradoxical: an abundant, weighty expanse consisting of a highly uneven mix of sources, from the new to the forgotten to the idiosyncratic (viz., those, such as Elvis Presley, listed earlier, as well as disciplinary figures, such as 1978 Braddock Award-winning Mary P. Hiatt). Burke’s parlor is nowadays full and teeming, more crowded than ever before. Even while the head of the distribu-
From graphs, then, come new insights, new provocations, and new questions: what has changed, over time, in the relationship between the head of the curve and the long tail?

Switching to a simple bar graph, the patterns become still more vivid; the visual model more concisely depicts a shift in citation practices.

In the first period, from 1987 to 1991, there were 2,755 citations. Using two criteria—the number of citations made to names at the head of the curve and the number of citations to unduplicated names in the long tail—we can create the percentage-based bar graph shown in Figure 11. In the first five-year period, then, 16 percent of the citations referred to names in the top twenty and just more than 32 percent of the citations were in the long tail. Over the next five years, we find a slight decrease in the percentage of citations occupied by the head; one-time references in the long tail reached slightly higher. Over the next five years, again the same shift appears: the head shrunk, the tail grew. And between 2002 and 2006, the number of citations climbs to 4,289, and the trend continues: the head fell below 10 percent; the tail approached 50 percent. The trend continues into the most recent five-year period. The most frequently cited figure—Kathleen Blake Yancey—was invoked in scholarly articles roughly half as often (26 references) as Linda Flower was in the comparable period of time twenty years earlier (56 references).

Thus, graphs underpinned with citation data assert themselves as a suggestive form of knowledge. As graphs condense, reduce, and render distant the default level of detail, they make possible a more comprehensive engagement
with patterns and trends. Power law distributions apply to citation practices in all journals and all academic disciplines; with graphs we can see how those distributions change over time. Rather than proving, confirming, or validating claims about disciplinary fragmentation or, more positively, eclecticism, I prefer to cast these graphs in terms of what they allow us to corroborate. *Corroboration* is a term that, in its connotations of contingency and flexibility, suggests that we might suspend judgment while simply granting credence to new forms of evidence (i.e., these models and abstracting practices) and use this to flex and strengthen (see “Corroborate”). The word *corroborate*, with its Latin root *robur*, a root shared by *robust*, places an emphasis on the mobilizing, inventive capacity of these graphs as visual models that can do much to shape our insights into disciplinary patterns. This emphasis on corroboration also foregrounds our individual and collective agency in shaping the field rather than resorting too quickly to endist speculation, prediction making, or discourses of disciplinary crisis.

Figure 11. Percentage-based comparisons of the number of citations in a given five-year period at the head of the curve (i.e., citations by the top twenty names) and in the long tail (i.e., those names cited just once).
The Heads and Tails of Disciplinary Density

Long live the dappled discipline.
Libby Miles et al., “Thinking Vertically”

Scholarly publishing has entered a time of tremendous flux, which is precisely why we must be more systematic than we have been about inventorying an evolving epistemic court. Susan Peck MacDonald’s study of disciplinary patterns is particularly helpful as one interpretive framework for these inventorying efforts. Drawing on David Kolb’s research, MacDonald emphasizes the importance of examining disciplinary materials to understand a field’s general approach to problems, which she distinguishes as compact or diffuse (22). Fields with compact problem orientations tend to align with the sciences as they enroll “assimilators” who synthesize divergent theories and methods and bring them to bear on common problems, whereas fields with diffuse problem orientations tend to align with the humanities as they enroll “divergers” whose attention to problems may be more singularly and discretely focused. Tabulating and graphing one journal’s citation distribution over twenty-five years may help us be more fully aware of the field’s evolution while it is happening and, furthermore, realize how different scholarly outlets, such as CCC, are situated and resituated in relation to a shifting compact-diffuse orientation.

Graphing provides a partial readout of the field’s pulse with respect to compactness and diffuseness, which complicates speculation about where the field stands at any given moment and where it is headed. Implicit in recent claims about disciplinary disunity and fragmentation, such as those by David Smit and Richard Fulkerson, is an assumption about an idealized state—a relatively contained, balanced ecosystem within which disciplinary conversations about the most pressing concerns, the most viable methods, and the most promising theoretical grounding lend stability to the notion of disciplinarity. These normative visions of rhetoric and composition studies are not easy for us to pin down and examine because they operate tacitly, informed by one’s institutional and departmental location, the time period one’s career has spanned, and a large number of other factors (training, publishing activity, leadership roles, etc.). But we should, nevertheless, remain fully cognizant of subtle references to what Jack Selzer once characterized as the “golden age” of composition studies wherever it lurks as a backdrop to this or that observation about the field’s uncertain—and some have argued tenuous—future. Lee Odell mentioned Selzer’s “golden age” reference in the afterword to his 1986 CCCC
Chair’s Address, “Are We There Yet? Prospects for a Maturing Discipline,” which was recently republished in a collection compiled and edited by Duane Roen. Odell briefly recounted the optimism that resonated in his keynote address; he remembered that Richard Larson argued back, answering his optimism with a warning of fragmentation. Reflecting on the moment, Odell acknowledged that his optimism may have been premature, that “it was a mistake to disregard what Dick Larson said” (152). That is, twenty years following the delivery of his upbeat keynote address on the then maturing discipline, Odell admits a far more cautious, reserved attitude. By the end of the short piece, however, he turns again toward optimism:

Change will continue to be rapid, and progress will always be slow. But at the center of the process of change and progress we find ourselves and our students continually growing—testing, reflecting, refining our assumptions about teaching and learning. So are we there yet? Are we mature as a profession? Probably not, especially if maturity means a time of stasis, a time without change. Are we maturing as a profession? Quite possibly—at least as long as we continue to grow as professionals. And that’s cause enough for optimism. (154–55)

Disciplinary terrain is constantly shifting, perhaps at what appears to be a faster rate than in many fields due to the adaptive, dappled spirit of much of the work in rhetoric and composition studies. Depending largely on one’s vantage point—that is, on whether one looks at the head or the tail of a citation frequency distribution—the field can appear to be highly focused, with a recognizable set of shared, dedicated principles and motives, or, it can appear as a loose amalgamation of pocketed clusters and enclaves, each holding fast to a relatively unique set of interests while neglecting (mindfully or not) any concept of disciplinarity in general. The full spectrum of citation data brings to light how both vantage points—generalist and specialist—are simultaneously implicated. As specialized enclaves negotiate a shared disciplinary frame, they simultaneously contribute to the shaping of the field at higher orders of magnitude. Though they are significant for us to evaluate regularly, the divergent factors motivating compositionists (individually or in groups, as divergers or as assimilators) to
specialize even as they risk turning away from shared disciplinary perspectives are beyond the scope of this article. Yet, with the graphing methods demonstrated here, we might better understand the ways specializations and those invested in them negotiate and cohabit disciplinary scenes, such as scholarly journals. We may prefer to be upbeat or recalcitrant about the patterns suggested above, but by noticing—whether by graphing or other distant reading methods, we are better able to have a sense of the dynamic networks that continuously proliferate in our disciplinary materials and practices.

A changing disciplinary density: this is not a condition for us to solve; nonetheless, it demands a certain reckoning, particularly for graduate education and professional development. Even though we cannot, perhaps, muster answers to the questions listed earlier, we can with renewed conviction accept what David Foster described in 1988 as an “invitation to an intellectual pluralism,” within which we can embrace these abstracting practices and the insights and questions they might productively open up for us.

Acknowledgments

Many thoughtful, generous friends, colleagues, and mentors helped me shape this project. I am especially thankful for valuable comments from Laurie Gries, Jim Brown, and Jana Rosinski, and I appreciate the support and guidance I received from Kathleen Yancey and two anonymous CCC reviewers.

Notes

1. To reduce anomalies, two issues of the journal—61.1 and 61.2—were restricted to only the articles catalogued in JSTOR. In each of those issues, 19 additional articles were published online, but those articles (38 in all) have been omitted from this study because the unusual publishing cycle skews comparisons across the twenty-five year collection. In effect, an extra year’s worth of articles were published online with these two issues.


3. Malcolm Gladwell’s Blink is suggestive here. Gladwell’s work relates numerous examples of “rapid cognition,” the quick, subconscious judgments that tacitly shape our impressions of the world.

4. Thirty of the articles do not use any formal citations whatsoever.
5. **CCC Online Archive**, an online resource sponsored jointly by NCTE and Syracuse University for the years 2005–2009, was one laboratory for the development of this data. The works cited available at this site were transferred from dynamic text PDF files and through optical character recognition (OCR) processing for articles available only as static PDFs.

6. Nonstandard authorship citations included anonymous, corporate, organizational, institutional, username, listserv, and other varieties where human authors were not explicitly named.

7. In their study of three decades of footnotes in *Critical Inquiry*, Anne H. Stevens and Jay Williams began with a selective (rather than organic or comprehensive) list. Explaining their methods, they note: “To begin our investigation, the staff of *Critical Inquiry* devised a list of theorists whose work we knew had been frequently cited. (To have tabulated every author cited in every article would have required more resources than we had at hand)” (212). With their preliminary list, they then worked page by page through the thirty-year archive of the journal, counting each appearance of a name on the list.

8. Each author listed in association with multiauthored works was credited with one reference tally. That is, where Linda Flower and John Hayes (and others in certain cases) appear as authors, each of them recorded one reference tally in the overall listing. This explains why the reference count (19,477) is higher than the original number of works cited entries (16,726).

9. There are clear dangers in leaping from patterns in *CCC* to patterns applicable to the field at large. Yet, this work with citation frequency in *CCC* should suggest the value in extending these methods to other journals in composition studies and, perhaps, other fields where such work has not yet been done.

10. For a critical discussion of citation practices, see Howard Tinberg’s “In the Land of the Cited,” which addresses a concern that two-year college faculty tend to be obscured in such work. Tinberg makes a case for more and more diverse citation practices.

11. Assessing these limitations may generate further research projects. For example, although the study featured in this article provides a cursory introduction to what graphing can offer, researchers in rhetoric and composition, computational rhetorics, and natural language processing (NLP) computational linguistics could begin to examine the in-text locations where citations are brought in. Doing so would allow us to know more fully which references are subject to elaborate framing and which are subject to less. Sentiment analysis would also help us rethink the positive and negative evaluations made about sources where they appear in scholarly corpora.
12. An alternative approach could use specific titles of sources rather than author names as its primary sorting key. With tracing sources, however, comes a greater challenge due to republishing. Sources commonly appear in iterations, or as trans-cclusive texts published in many different versions. Consequently, I have preferred to sort by author name. Consider, as an example, Franco Moretti’s *Graphs, Maps, Trees*, which appeared as a series of articles in the *New Left Review* before it became a monograph. Using source titles as a primary sorting key would, in this case, reflect a skewed number of citations, assuming citations that refer to both versions of Moretti’s work.

13. Undoubtedly, other meaningful, relevant forms of knowledge circulate, even continuing in the undocumented lore about which North wrote in 1987 and continuing to play out in both formal and informal venues, from the hallways and lounges at national conferences to program listservs, syllabi, and so on.

14. Recent crowd-sourced big data projects, such as the Writing Studies Tree created by researchers at the CUNY Graduate Center, suggest some of the promising ways computing platforms may augment the crucial work of more systematically tracing the field’s genealogies (Perl et al.).


16. These laws of distribution go by many different names in economic theory. Vilfredo Pareto, a nineteenth-century Italian economist, is generally credited for coming up with the law of distribution, better known as the 80:20 Rule, which generally poses that a small percentage of a population will hold disproportionately high measures of wealth and power relative to the large percentage of a population (Ball 247).

17. The distant reading and graphing methods presented here are amenable to sorting by other criteria as well. It would be possible, for example, to determine the changing rate of reference to different forms of publication (e.g., chapters in edited collections, single-author monographs, peer-reviewed articles, online resources, etc.), though attending to these distinctions is beyond the scope of this article.

18. By applying a classification scheme similar to the fourteen cluster areas used by the Conference on College Composition and Communication to categorize conference presentations, the full dataset could be subdivided into corresponding groupings for “Language,” “Creative Writing,” “Basic Writing,” and so on. Graphs produced by this technique could suggest distinctions in the scholarship associated with these respective areas as well as the values embraced and promoted therein. These would not necessarily reflect widespread disciplinary values, but they would make accessible a view of area-specific citation patterns within *CCC* since 1987.
19. This is not only a question for rhetoric and composition to consider; this method for graphing citation rates ought to generalize, suggesting its usefulness for other journals and in other fields, as well.

20. Anderson acknowledges that long tail distributions adhere to a fractal pattern, according to which the curve and the tail incorporate smaller subdistributions within the larger one. These small niches help us account for the ways specialization perpetuates micro-patterns that are locally consistent with the larger patterns in the field. Many special interest groups articulate distinctive perspectives on the field and their relations to it. The methods introduced here might help us understand how larger-scale conceptions of disciplinarity can be negotiated with the perspectives promoted by smaller groups whose identifications with the field at large require qualification.

Works Cited


Derek Mueller

Derek Mueller is an assistant professor of written communication at Eastern Michigan University, where he teaches classes in writing, rhetoric, and technology. His articles and reviews have appeared in Computers and Composition, Kairos, Composition Forum, and JAC.