Chapter 4: The Thin, Long Tail of Citation Frequency

> Newcomers to this dialogue are sometimes unfamiliar with the heteroglossia of our heritage, with the voices that have created the conversation upon which we continue to build. For example, as composition and rhetoric matures, who was speaking? (Phillips, Greenberg, & Gibson, 1993, p. 443)

Inventorying an Epistemic Court

Nearly two decades ago, Donna Burns Phillips, Ruth Greenberg, and Sharon Gibson (1993) inquired into rhetoric and composition/writing studies' (RCWS) maturation using methods of counting and sorting to distinguish various subsets of aggregate data drawn from *College Composition and Communication* (*CCC*). Phillips, Greenberg, and Gibson's project provided an early snapshot of what Janice Lauer (1984) described 10 years before in a *Rhetoric Review* essay titled "Composition Studies: Dappled Discipline" as an *epistemic court*, or a locus through which disciplinary knowledge circulates, attracts attention, and gains its status. Although the Phillips, Greenberg, and Gibson (1993) study was limited to a single journal, their compilation stands as an early investigation into broad-scope data from *CCC*, which included the most frequently cited authors and works, the journal's material forms, and the journal's editors since March 1950. Their report cannot be considered a comprehensive, evidence-based Zeitgeist of the disciplinary activity spanning more than 40 years into the early 1990s; however, the inventory provided a data-driven statement against which competing perspectives on disciplinary activity could be compared. For how it was cast as both exploratory and data-driven—much like the work that follows—it must be regarded as an early instance of post-positivistic data science (Kitchin, 2014) and an agent of new and emerging big data epistemology that distinguished itself from the proof-oriented noetic trappings of a purer empiricism.

In an effort to update and contribute further to the ongoing inventorying of RCWS' epistemic court, this chapter adopts a similar exigency to that heeded by Phillips, Greenberg, and Gibson (1993) as it relies upon quantitative data, bibliometrics, and graphing as a means of thinly describing the changing nature of "who [has been] speaking" over the past 25 years, according to citation frequencies in *CCC*. Essentially, I contend that graphs, as a form of

distant reading (Moretti, 2007, 2013) and an instantiation of thin description (Love, 2010, 2013), help us to know continually unfolding tensions among specialization, the interdisciplinary reach of RCWS, and the challenges these present to newcomers to the scholarly conversation.

Suspending judgment from the outset about the consequences of specialization and interdisciplinary borrowing as generally positive or negative, this chapter seeks 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 will say more about the studies using graphs and relevant quantitative methods to understand journals and the fields sponsoring them. Doing so highlights the three basic principles of distant reading and thin description elaborated in Chapter Two. Graphs operate as data-driven visual models; their visual force is in translating a collection of data into a comprehensible figure. Graphs deliberately alter scale, and as such, they aggregate patterns linking details and nonobvious phenomena otherwise at risk of passing unobserved. As a function of their systematic compilation of replicable data, graphed patterns may empirically corroborate local, tacitly felt impressions about changing disciplinary conditions. In the second half of this chapter, I adopt as an exploratory framework Chris Anderson's (2004, 2008) work on long tails (Pareto distributions) and present graphs based on a compilation of 16,726 citations in 491 journal articles[9] published in *CCC* over 25 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 RCWS. For example, Maureen Daly Goggin's (2000) well-known history of the field, *Authoring A Discipline: Scholarly Journals and the Post-World War II Emergence of Rhetoric and Composition*, presented eight graphs, each designed to render apprehensible some data set aggregated manually from

9   To reduce anomalies, two issues of the journal—61.1 and 61.2—were restricted only to 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 25-year collection. In effect, an extra year's worth of articles were published online with these two issues.

the nine journals at the center of her study.[10] Goggin's study is one notable example where graphs have been applied similarly to the way Moretti has used 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 accounted 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 (assumed binary) gender of contributors to the journals. Goggin also used 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 its reliance on graphs, Goggin's disciplinarily innovative work provided strong examples of graphing methods that, because they translated a collection of data to a visual figure, thereby established a new scale of engagement, rendering recognizable patterns of disciplinary activity that would otherwise be difficult to discern, particularly for newcomers to the field.

This use of graphing to engage with data at a new scale deliberately adjusts the level of detail at which we ordinarily experience texts, and as such it reaffirms database, scale, and pattern as key motives for exploring disciplinarity this way. Experientially, reading tends to be a local, direct encounter, typically involving (or demanding) an identifiable, focal text. While there are sure to be exceptions (e.g., a bibliographic essay pursues a similar purpose: synthesis by reduction, the full text falls away and in its place stands a proxy, a textual double), 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 operate as thin descriptions at a scale only slightly closer to the content of the article than the title and author alone (for more on article abstracts, see Chapter Two). Readers rely on these devices to make quick decisions about whether to read a particular article or not, but reading the journal through these

---

10   Goggin gathered data from *College Composition and Communication, College English, Research in the Teaching of English, Rhetoric Society Quarterly, Freshman English News (Composition Studies), JAC, Rhetoric Review, Pre/Text,* and *Written Communication.*

devices alone is not quite the same as reading a scholarly journal in the common sense of the activity.[11]

Reading across a series of journal articles gathered around a specific 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 25 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.
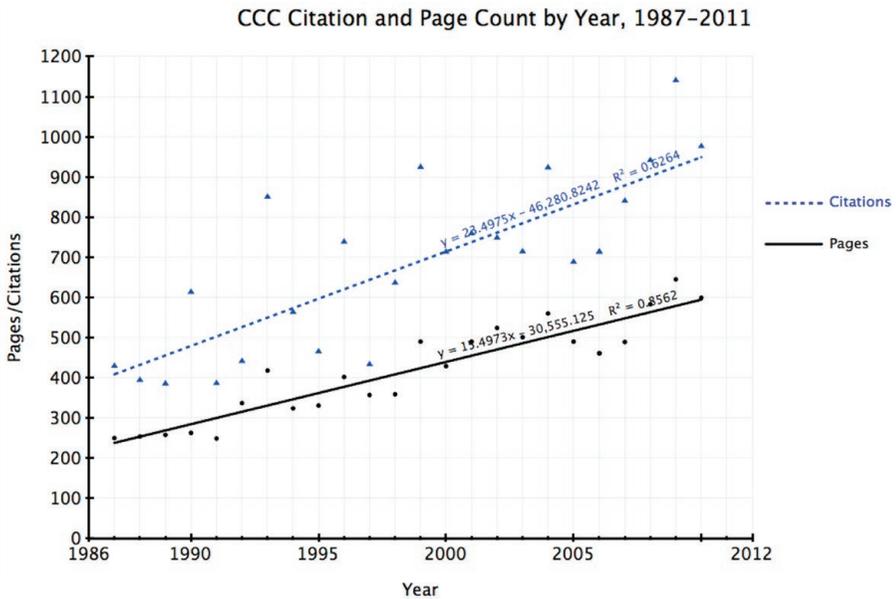

Figure 15. A plot graph indicative of page count and citation count by year in CCC, between 1987 and 2011. Trend lines for page count (solid) and citation count (dotted) indicate the gradual but steady increase in these basic features of the journal.

Figure 15 employs graphing to present growth patterns in the number of citations and the number of pages in *CCC* over 25 years. This graph il-

11   Malcolm Gladwell's (2007) *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.

lustrates just one form of knowledge available to us in exploratory quantification, in the distant reading and thin description that comes of counting, recording tallies, and plotting coordinate points. The figure overlays 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 25 years sampled, articles published in *CCC* have grown longer and they have also have come to draw more extensively on source material appropriate to include in a references list.

Besides suggesting a gradual inflation in the page count and number of sources in scholarly articles published in *CCC* over the 25 years sampled, Figure 15 also elicits questions. This inventive and generative capacity constitutes 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 25 years? Do other journals exhibit similar trends over the 25 years sampled? What about journals in other fields? How have alternative length publishing models emerged as an implicit response? For instance, *Present Tense* publishes medium-form scholarship online and *Intermezzo* (2016) is a series that corrects in the other direction, inviting "essays that are too long for journal publication but too short to be a monograph." How high might page counts and citation counts ascend 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, editorial board members, to reviewers, writers, and readers? Absent distant reading and thin description methods these questions would warrant hunches and speculation, but we would be unable to present the pattern as compellingly. The graphed pattern refreshes the questions with vivid presentational force.

Related quantitative studies foreground the promise of graphing techniques informed by distant reading and thin description for RCWS scholars as depends upon the systematic archivization of reusable, interoperable, field-wide data sets. Phillips, Greenberg, and Gibson (1993) 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 (2000) interest in elucidating patterns and the aims that justify their pursuit: chronicling the discipline's genesis. Phillips, Greenberg, and Gibson (1993) 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) risen, and who, at 15-year intervals, has been cited most frequently. They speculated, 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. (p. 451)

Whether or not this speculation holds as an enduring theory is less important than is the way these methods catalyze questions heuristically and begin to provide a means of addressing such questions more systematically than has been established to date. The demonstrable force of graphs renews the points Phillips, Greenberg, and Gibson raised concerning the field's development and maturation, the growing demand for theoretically and methodologically sophisticated 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 reinvigorate these questions and give us different ways to grasp nonobvious trends.

Bibliometric Methods and Techniques

With a few key distinctions, the methods and techniques I used to compile the frequency of citations appearing in *CCC* from 1987–2011 are similar to those applied by Phillips, Greenberg, and Gibson. Phillips, Greenberg, and Gibson (1993) recorded references appearing in any piece of work published in *CCC* between 1950 and 1993, including reviews, interchanges, and features unique to an 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 publishing process 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; for more on this change, see Chapter One). 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.[12]

I prepared the list by gathering all the works cited for each article in a single spreadsheet.[13] Because individual works cited entries often include mul-

---

12   Thirty of the articles do not use any formal citations whatsoever.

13   *CCC Online Archive*, an online resource sponsored jointly by NCTE and Syracuse University from 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.

tiple authors (or clipped lists of authors denoted with *et alia*), the citation list required extensive smoothing, which I handled manually. I coded each bibliographic entry in the list so that listings with editors, *et alia*, hyphens (repeated reference to a single author), and nonstandard authorship could be sorted apart from author listings.[14] I removed the citation entries for editors, replaced the hyphen placeholders with the full names of the appropriate author(s), and replaced the "et al." with the names of all authors collaborating on a given work. Finally, using text-matching algorithms and manual proof-reading, I double-checked the list to ensure correct spelling and name formatting. These alterations to the comprehensive works cited list resulted in an expanded roster of authors whose names appeared with every instance of a publication associated with their names.[15]

The labor involved in coding a collection of several thousand citations is due for acknowledgement. This was slow, detail-oriented work, carried out over several months and later updated to extend the data set into another half decade. And yet, the methods scale, as is demonstrated in Eric Detweiler's (2015) "'/' and '-'?: An Empirical Consideration of the Relationship Between 'Rhetoric' and 'Composition,'" which sampled two years (2001–2002 to 2011–2012) a decade apart to inform an analysis of shifting citation practices in *College Composition and Communication* and *Rhetoric Society Quarterly*. In another example, Joe Torok's (2013) "Visualizing *Present Tense*: Graphing and Mapping a Corner of the Discipline" studied citations in the first three years of articles published in *Present Tense* to investigate whether and to what extent medium-form scholarship (i.e., shorter articles) reflected distinctive patterns related to the scope of sources cited therein. While both were rigorous and substantial projects concerned with citation, the scope of these studies attests to the scalability of distant reading and thin description—and to the usefulness of projects inquiring into network sense at different scales.

With each name-reference assigned to a single slot in the comprehensive listing, various tabulations were possible; the 16,726 works cited entries be-

---

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

15   In their 2006 study of three decades of footnotes in *Critical Inquiry*, Anne Stevens and Jay Williams began with a selective (rather than organic or comprehensive) list. Explaining their methods, they noted, "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)" (p. 212). With their preliminary list, they then worked page by page through the 30-year archive of the journal, counting each appearance of a name on the list.

came a list of 19,477 name-references.[16] This also means that 2,751 name references surfaced from secondary, tertiary, and *n*-ary authors. Developed out of this expanded data set, Table 1 shows the top 103 scholars sorted in descending order by the number of references made to them in *CCC* articles between 1987 and 2011.

Table 1. The 103 most frequently cited authors in *CCC* from 1987–2011.

	# out of 19,477 references	Name (461 articles)
1	145	Linda Flower (66 articles)
2	133	Peter lbow (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 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

16   Each author listed in association with multi-authored 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).

	# out of 19,477 references	Name (461 articles)
35	43	Janet Emig, Michel Foucault, Henry Giroux, Gesa Kirsch, Geneva Smitherman
40	42	Kenneth Bruffee, David Russell
42	40	Deborah Brandt, Paulo Freire, Richard Haswell
45	39	Lil Brannon
46	38	Bruce Horner, Charles Knoblauch
48	37	Nancy Sommers, United States
50	35	James Britton, Glynda Hull, Mary Louise Pratt
53	34	Linda Brodkey, Elizabeth Flynn, Gail Hawisher, Ira Shor
57	33	Thomas Newkirk
58	32	Ann Berthoff
59	31	Susan Jarratt, Walter Ong, James Porter, Patricia A. Sullivan
63	30	Carol Berkenkotter, Pierre Bourdieu, Victor Villanueva
66	29	Sarah Freedman, Lucille McCarthy, Louise Wetherbee Phelps
69	28	Albert Kitzhaber, Carolyn Miller
71	27	Aristotle, James Paul Gee, Diana George, Gerald Graff, George Hillocks, Jr., Brian Huot
77	26	Janice Lauer, Richard Ohmann, Susan Wells
80	25	Judith Butler, Peter Mortensen, Stephen Witte
83	24	Pat Belanoff, Robert Brooke, Keith Gilyard, Anne Herrington, Bruce Herzberg, Gunther Kress, Ken Macrorie, Greg Myers, Adrienne Rich, Joseph Williams
93	23	John Ackerman, Chris Anson, Arthur Applebee, Ellen Barton, Jacques Derrida, Michael Halloran, Susan McLeod, Richard Miller, Kurt Spellmeyer, Brian Street

The simple tabulation evokes many questions worthy of exploring more deeply in the contexts of disciplinary formation, scholarly influence, 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 do well-established scholars overshadow new scholars in such a listing as this? What are some of the intriguing juxtapositions where positions in the list are shared? This latter question is a tangential one, but one worth considering for its inventive richness in a course that in-

troduces graduate students to the field. Wondering about coincidental pairings is germane to a practice I think of as *heuretic discipliniography*, or writing and re-writing the field by exploring the enigmatic intersections across different scholars' work as well as the associated pedagogical, theoretical, and methodological approaches advanced thereby.

The single, comprehensive list in Table 1 is suggestive in its own right, but it tends to occlude temporal variation: the changing tide of citation practices at lesser increments within this 25-year period. In the interest of beginning to see into this variation, consider an alternative table (Table 2) developed out of the same data set.

Table 2. Top 10 most frequently cited authors in *CCC* from 1987 to 2011, by five-year interval.

1987–1991	C	A	1992–1996	C	A	1997–2001	C	A
Total references: 2,755			Total references: 3,595			Total references: 3,881		
Total articles: 79			Total articles: 102			Total articles: 93		
Flower, Linda	56	22	Bizzell, Patricia	43	29	^Berlin, James A.	37	29
Hayes, John	41	16	^Elbow, Peter	43	30	Bartholomae, David	35	25
Lunsford, Andrea	27	18	Rose, Mike	37	22	Bizzell, Patricia	31	19
Connors, Robert	25	14	^Dewey, John	32	4	Faigley, Lester	30	22
Bizzell, Patricia	23	16	Flower, Linda	32	17	Elbow, Peter	28	18
Rose, Mike	23	14	Bartholomae, David	30	24	Connors, Robert	26	18
Faigley, Lester	22	14	Lunsford, Andrea	26	19	^Crowley, Sharon	25	15
Bartholomae, David	20	19	^Shaughnessy, Mina	25	15	^Miller, Susan	24	19
Berlin, James A.	20	14	Faigley, Lester	24	21	Lunsford, Andrea	22	15
Britton, James	20	9	Connors, Robert	23	18	Rose, Mike	20	14
			Out of top 10 from previous five years			**Out of top 10 from previous five years**		
			Britton, James			Dewey, John		
			Hayes, John			Shaughnessy, Mina		
			Berlin, James A.			Flower, Linda		

2002–2006	C	A	2007–2011	C	A	Totals, 1987–2011	C	A
Total references: 4,289			Total references: 4,957			Total references: 19,477		
Total articles: 93			Total articles: 94			Total articles: 461		
^Burke, Kenneth	31	9	^Yancey, Kathleen Blake	26	19	Flower, Linda	145	66
^Flower, Linda	28	12	Cushman, Ellen	25	13	Elbow, Peter	133	85
^Smitherman, Geneva	28	10	Elbow, Peter	23	13	Bizzell, Patricia	118	82
^Trimbur, John	28	17	^Russell, David	23	13	Bartholomae, David	112	93
^Bakhtin, Mikhail	23	10	^Bazerman, Charles	22	17	Berlin, James A.	111	90
^Cushman, Ellen	22	12	^Gere, Anne Ruggles	21	19	Connors, Robert	110	78
Bartholomae, David	20	18	Selfe, Cynthia	21	10	Lunsford, Andrea	102	74
Elbow, Peter	20	11	^Haswell, Richard	20	12	Faigley, Lester	101	79
^hooks, bell	20	16	Burke, Kenneth	19	11	Rose, Mike	96	64
^Hopkins, Edwin	20	2	^United States	18	8	Trimbur, John	77	57
^Selfe, Cynthia	20	7						

Out of top 10 from previous five years	**Out of top 10 from previous five years**	
Crowley, Sharon	Flower, Linda	
Berlin, James A.	Smitherman, Geneva	
Bizzell, Patricia	Trimbur, John	
Faigley, Lester	Bakhtin, Mikhail	
Connors, Robert	Bartholomae, David	
Lunsford, Andrea	hooks, bell	
Rose, Mike	Hopkins, Edwin	
Miller, Susan		

Notes: 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 10 for the previous five-year period. This table bears direct correspondence to similar tables appearing in Stevens and Williams's (2006) work with *Critical Inquiry* and in Phillips, Greenberg, and Gibson's (1993) work with *CCC* from 1950–1993.

What does Table 2 *do*? Certainly, it provides compelling quantitative evidence for trends and patterns in citation practices, and it foregrounds the temporal subsets within the 25-year sample by adjusting the timescale. 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 lends itself 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 most recent five-year period. In other words, we can see that Linda Flower was cited in 22 of the 79 articles (i.e., 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 (i.e., 9.7%). Further lines of inquiry include examining the lists with attention to gender, sexuality, class, 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. Admittedly, 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 RCWS and, perhaps, other fields where such work has not yet been done.

Based on the approach modeled so far, we can begin to see how quantitative studies of citations spark insights and advance questions concerning the ways citation practices change.[17] 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.[18] The list simply affirms one fairly narrow kind of presence. In other words, citation listings lack volume; they do not report whether a single source greatly influenced (and appeared repeatedly, throughout) in, say, more than 10 (or more) pages of an article or whether,

17   For a critical discussion of citation practices, see Howard Tinberg's (2006) "In the Land of the Cited," which addresses a concern that two-year college faculty tend to be obscured in such work. Tinberg made a case for more and more diverse citation practices.

18   Assessing these limitations may generate further research projects. For example, although the study featured in this chapter provides a cursory introduction to what graphing can offer, researchers in rhetoric and composition, computational rhetorics, natural language processing (NLP), and 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.

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 acknowledgement of related limitations, Phillips, Greenberg, and Gibson (1993) wrote 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" (p. 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.[19] Further, a bibliometric methodology privileges presence as a function of publication, although there are many other kinds of professional and interpersonal presence essential to disciplinary stabilization, including mentorship, conversation, and the writing and circulation of what Louise Wetherbee Phelps (2016) has called *humble genres*.

In their discussion of methods used to study the journal's authors and works cited quantitatively, Phillips, Greenberg, and Gibson (1993) wrote,

> 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. (p. 454)

So while quantitative studies of authors cited in a well-known journal

---

19   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, such as electronic texts that exist in many copies whose precise differentiations become muddled like email threads. Consequently, I have preferred to sort by author name. Consider, as an example, Franco Moretti's (2007) *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 differently skewed number of citations.

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 kind 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 Tables 1 and 2 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 RCWS, would claim to know *none* of these figures or the impact of their work? And yet, the top 103 authors cited over the 25 years sampled in *CCC* or the top 10 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 data set, 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 data sets significant to the field's formation, is only beginning to take shape (Chamberlain, 2016; Detweiler, 2015; Miller, 2014; Miller et al., 2016). To make sense of the comprehensive record of citation within *CCC* in this 25-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 25-year period—those who, by the record of citation frequency, registered a singular appearance. For this question, another series of graphs proves insightful, enabling 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. (Anderson, 2008 p. 40)

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 transformed with the rise of digital commerce. The article distinguished between spatially constrained traditional retailers and their comparably abundant online counterparts. According to Anderson, the typical Borders[20] bookstore, for example, carried 100,000 titles, but its leading online competitor, Amazon.com, offered a vastly deeper (i.e., longer) selection to consumers: more than 3.7 million titles. Anderson's research, which he expanded into a monograph, centered on the idea of the long tail: the uncommon products and specialized interests that online markets can support. *The Long Tail* (2008) 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 cited 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 and is also known as a power law called the Pareto distribution[21], which uses graphed patterns to show the distribution of power in an activity or phenomenon. In his article-length work on the long tail, Anderson (2004) provided a version of the infographic in Figure 16 to illustrate.

Here, music is the focal premise. Walmart, like Borders, offers a limited selection; 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, which includes less popular titles that continue to sell actively, despite ranking well beneath the threshold of popularity that justifies the entirety of Walmart'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 (2004) called the "tyranny of geography" (p. 17).

---

20   The Ann Arbor, Michigan-based company filed Chapter 11 bankruptcy and closed in 2011.

21   These laws of distribution go by many different names in economic theory. Vilfredo Pareto, a 19th-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, 2006, p. 247).

ANATOMY OF THE LONG TAIL

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in gray) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.

RHAPSODY
TOTAL INVENTORY: 735,000 songs
typical Wal-mart store: 39,000 songs

AMAZON.COM
TOTAL INVENTORY: 2.3 million books
typical Barnes & Noble store: 130,000 books

NETFLIX
TOTAL INVENTORY: 25,000 DVDs
typical Blockbuster store: 3,000 DVDs

THE NEW GROWTH MARKET:
OBSCURE PRODUCTS YOU CAN'T GET ANYWHERE BUT ONLINE

TOTAL SALES 22%

TOTAL SALES 57%

TOTAL SALES 20%

product not available in offline retail stores

Songs available at both Wal-Mart and Rhapsody

Songs available only on Rhapsody

Average number of plays per month on Rhapsody

6,100 — 2,000 — 1,000 — 0

0 — 39,000 — 100,000 — 200,000 — 500,000

Titles ranked by popularity

Recreated from *Wired*. Sources: Erik Brynjolfsson and Jeffrey Hu, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks.

Figure 16. Anatomy of the long tail. A recreated Wired infographic depicting the distribution of songs available in traditional retail spaces, such as Walmart, as compared with online retail sites, such as Rhapsody.

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 (2008) also acknowledged that his research has opened up to even broader possibilities for the long tail as an apparatus for exploratory and 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. (p. 11)

He arrived at an expanded view of the long tail, one that recognized that its application reaches beyond economics to other cultural phenomena. Assuming a similarly broad view of the long tail, I contend that it serves generatively 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 25-year sample. The top-ranking author-citations in *CCC* between 2007–2011 are less than they were for the same period of time 20 years earlier. The frequency of the citation set's highest circulating figures is

more spare in later years than in earlier ones in this data set. But graphing all the author-citations does more than confirm what we already know about those few at the top. The graphed citation frequency distribution 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. The long tail accounts for how citations have scattered and dispersed. Once more, graphing functions as a form of distant reading and thin description—a means of engaging with large-scale data at multiple scales to notice nonobvious patterns. 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 RCWS.

Following the Long Tail's Thinness: The Names Invoked Just Once

> It 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. (Odell, 2006, p. 149)

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 inhabiting the long tail of *CCC* citations. These six figures share this distinction with 5,761 other names referenced just once (out of 8,035 unique name references in the 25-year period in question). Another 986 names appear in the cited works just twice, which leaves 1,287 names (i.e., 16.0%) that appear in *CCC* citations three or more times within the 25-year sample. By assigning these figures to a simple graph, we can see that they follow a power law, meaning relatively few names rank highly in citation frequency (see Tables 1 and 2 above for specific references) while more than 80% of the names register a momentary appearance, usually appearing in a single article. Translated into a graph, the 25-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 103 author-citations shown in Figure 17. Attending to the immense shelf of the less-frequent citations in the data set demands a more comprehensive view—a more *distant* view, that is. We must step back even farther than did Phillips, Greenberg, and Gibson (1993) 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 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.
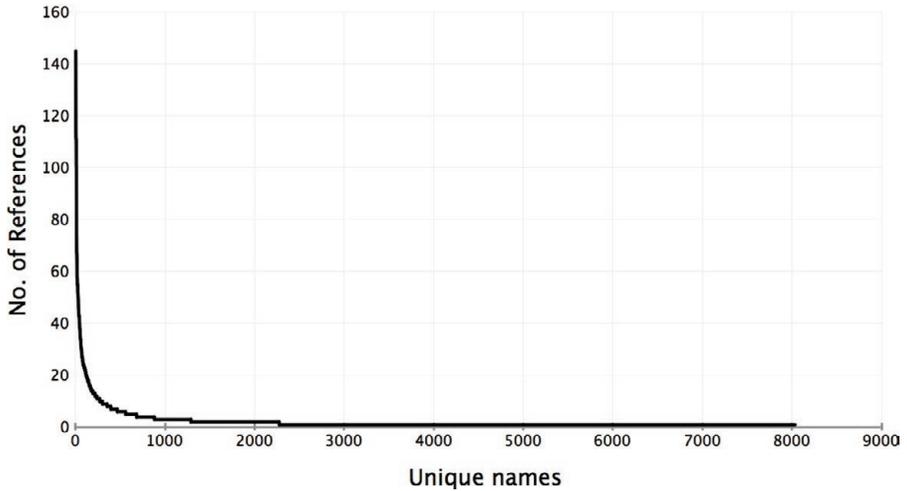
Figure 17. *References to unique names in CCC works cited from 1987–2011.*
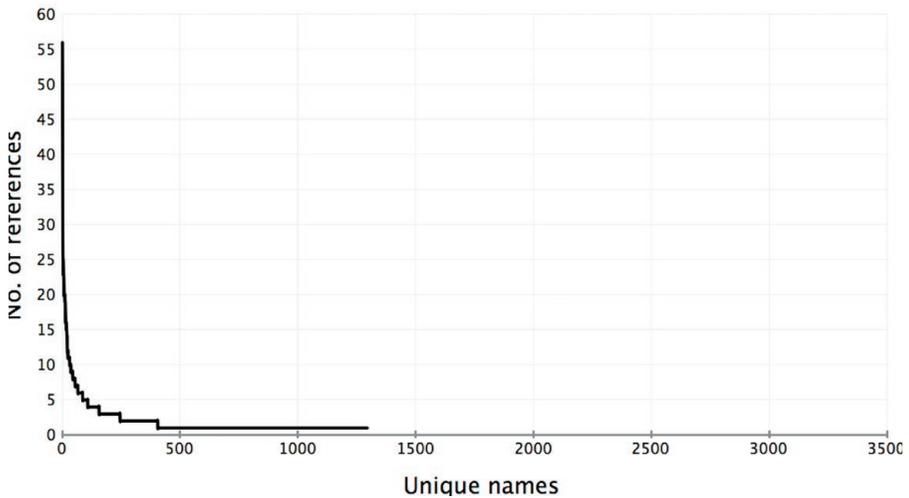
Figure 18. *Citation frequency in CCC, 1987–1991.*
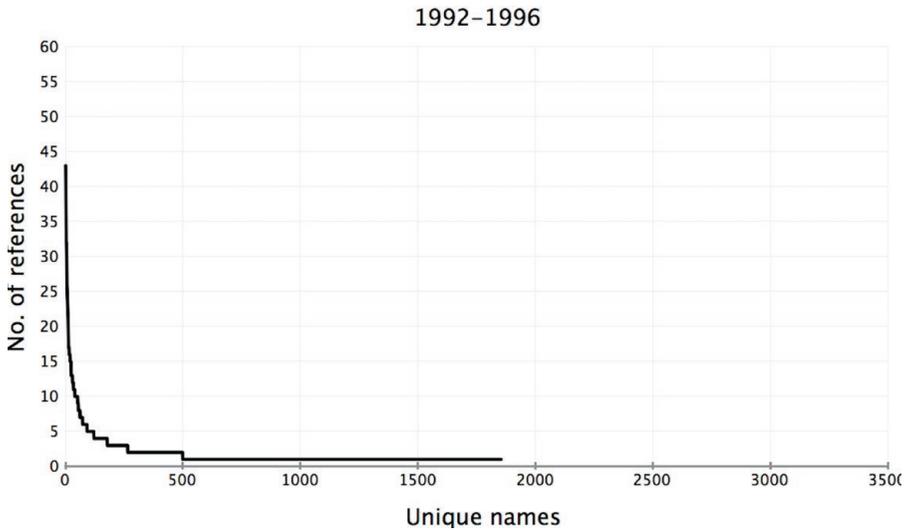
1992–1996

Figure 19. Citation frequency in CCC, 1992–1996.
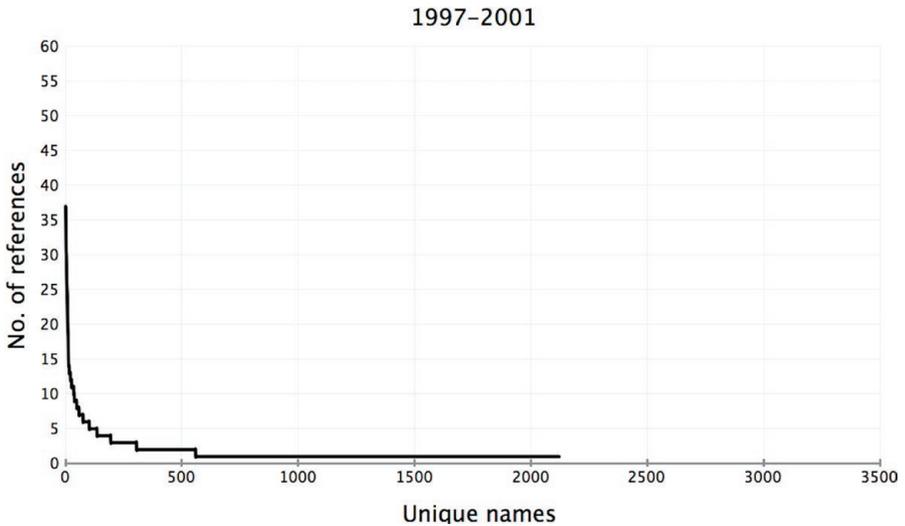
1997–2001

Figure 20. Citation frequency in CCC, 1997–2001.

Keeping in mind this more general thread of inquiry into the maturation of the field foregrounds the incremental development seen in Figure 17: How and at what rates did unique references grow? When 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? To answer these questions, consider a more nuanced series of graphs, each displaying a five-year data sub-set (much as Table 2 did). Figures

18–22 are also available online at https://wac.colostate.edu/docs/books/network/citationfrequency.gif as an animated GIF that loops to show the declining head of the curve (left) as contrasted with the elongation of the tail (right).
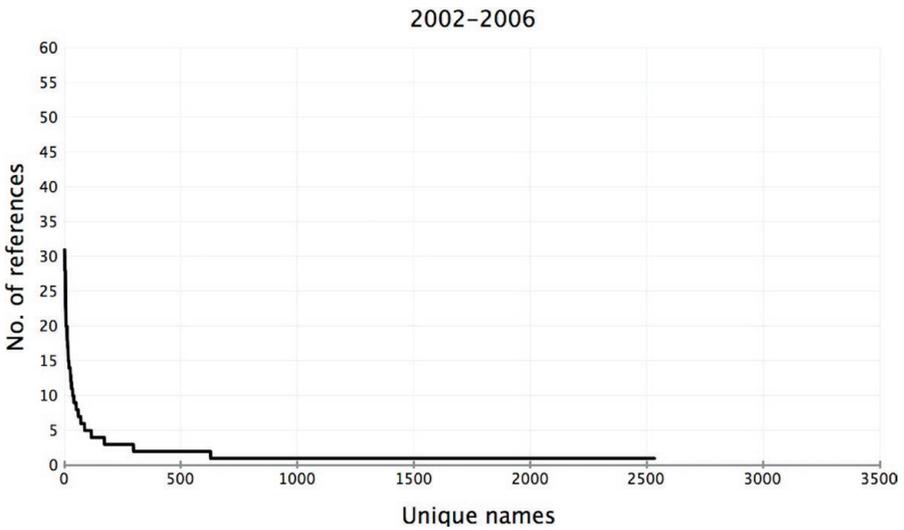

*Figure 21. Citation frequency in CCC, 2002–2006.*
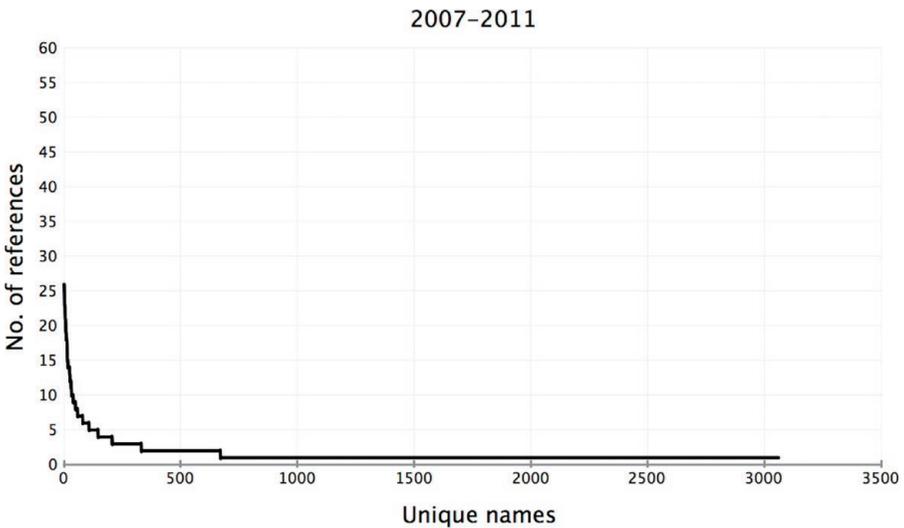
Figure 22. Citation frequency in CCC, 2007–2011.

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 draw upon. Across 25 years of citation activity in *CCC*, the long tail has grown longer, indeed, while the head has dwindled over time.

Figures 18–22 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 perceive these gradual transformations—shifts so subtle that it is easy, at the scale of a career, to disregard. The methods featured here are also 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.). Thus, we can use distant reading and thin description methods to understand with more granularity factors affecting citation distribution. Furthermore, although the graphs I have produced reflect the full data set, it is possible to use these methods to isolate and compare smaller segments of the data. Separating sub-sets 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 5 or 10 years of their careers, to work by alums of specific graduate programs, or by scholars whose research focuses on a specialized area. The methodology is considerably more dynamic and robust than what this necessarily limited introduction of it can feature.[22]

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 (2006) in his 1977 CCCC chair's address (p. 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

22   By applying a classification scheme similar to the 14 cluster areas used by the Conference on College Composition and Communication to categorize conference presentations, the full data set 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.

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., Elvis Presley, as well as disciplinary figures, like Mary P. Hiatt, who won the 1978 Braddock Award). Burke's parlor is nowadays full and teeming, more crowded than ever before. Even while the head of the distribution stands tall—an indication of the recurrence of fairly regular, recognizable names (e.g., Linda Flower, Patricia Bizzell, Peter Elbow, etc.)—the long tail tests the limits of comprehension and memory. Although we do not at this time have data from all of the major journals to investigate this fully, the changing shape of the graphed distribution reiterates more emphatically a question only hinted at in Tables 1–2, but one nevertheless crucial to the idea of a common disciplinary domain: How flat can the citation distribution become before it is no longer plausible to speak of a discipline?[23]

To clearly and responsibly engage with this complicated, shifting expanse, we need the full spectrum of data, not only the list of the most frequent appearing names. The full distribution is required if we are to examine the relationship between what has happened at the head of the distribution and what has happened furthest from it, in the long tail. From graphs, from one exercise in visual epistemology, 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 (Fig. 23), the patterns become still more vivid; the visual model more concisely conveys a shift in citation practices. In the first period, from 1987 to 1991, there were 2,755 citations. Using two criteria, (a) the number of citations made to figures at the head of the curve and (b) the number of citations to unduplicated figures in the long tail, we can create the percentage-based bar graph shown in Figure 23. In the first five-year period, then, 16% of the citations referred to figures in the top 20 and just more than 32% 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%, and the tail approached 50%. The trend continues into the last five-year period sampled. 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 20 years earlier (56 references).

---

23   This is not only a question for RCWS to consider; this method for graphing citation rates ought to generalize, suggesting its usefulness for other journals and in other fields, as well.

Head and Tail by Percentage
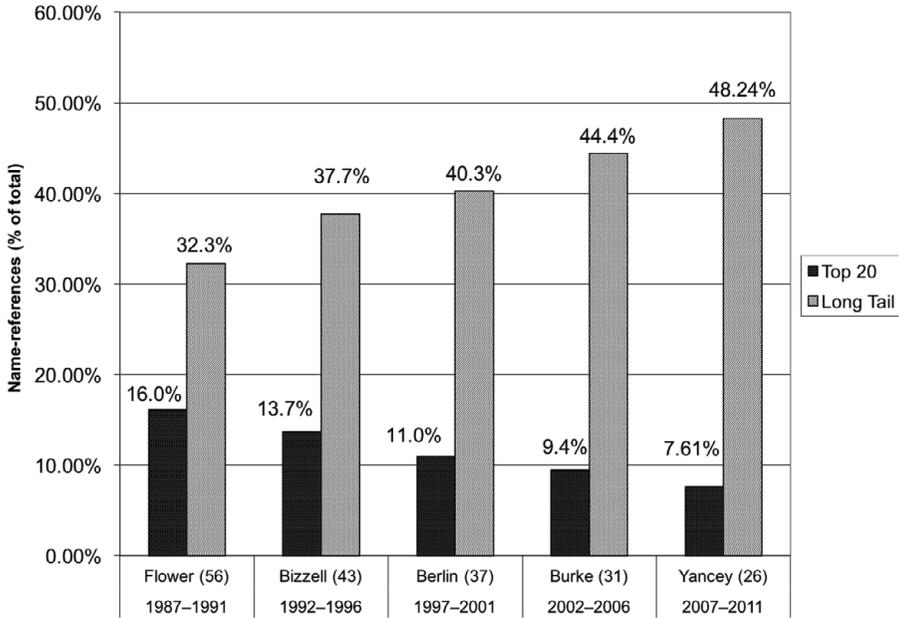
Figure 23. Percentage-based comparisons of the number of citations in a given five-year period at the head of the curve (i.e., citations by the top 20 figures) and in the long tail (i.e., those figures cited just once).

Thus, graphs underpinned with citation data assert themselves as a suggestive form of knowledge. As graphs condense, reduce, and render distant and thin 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 we might suspend judgment while simply granting credence to new forms of evidence (i.e., these models and abstracting practices), using this credence to flex and strengthen ("corroborate," n.d.). 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.

The Heads and Tails of Disciplinary Density

> Long live the dappled discipline. (Miles et al., 2008, p. 511)

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 (2010) study of disciplinary patterns is particularly helpful as one interpretive framework for these inventorying efforts. Drawing on David Kolb's research, MacDonald (2010) emphasized the importance of examining disciplinary materials to understand a field's general approach to problems, which she distinguishes as compact or diffuse (p. 22). Fields with compact problem orientations tend to align with the sciences as they enroll "assimilators" (p. 26) 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" (p. 26) whose attention to problems may be more singularly and discretely focused. Tabulating and graphing one journal's citation distribution over 25 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 re-situated in relation to a shifting compact–diffuse orientation.

Graphing provides a limited, partial read-out 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 (2004) and Richard Fulkerson (2005), 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 RCWS 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 (qtd. in Odell, 2006). Lee Odell (2006) mentioned Selzer's golden-age reference in

an afterword to Odell's 1986 CCCC Chairs' Address, which was republished in 2006. In this afterword, Odell recounted the optimism that resonated in his keynote; he remembered that Richard Larson argued back, answering Odell's 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" (p. 152). Twenty years following the delivery of his upbeat keynote address on the then-maturing discipline, Odell admitted a far more cautious, reserved attitude. By the end of the afterword, however, he turned 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. (2006, pp. 154–155)

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 RCWS. 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) an agreed upon concept of disciplinarity *in general*.[24] 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 to specialize, even as they risk of turning

---

24   Anderson (2008) acknowledged that long tail distributions adhere to a fractal pattern, according to which the curve and the tail incorporate smaller sub-distributions 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.

away from shared disciplinary perspectives (individually or in groups, as divergers or as assimilators), are beyond the scope of this study. 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 and thin description methods—we are better able to have a sense of the dynamic networks that continuously proliferate across our disciplinary materials and practices.

A changing disciplinary density is not a condition for us to solve; nonetheless, it demands a certain reckoning, particularly for visibility initiatives, graduate education, and professional development. For instance, the questions listed earlier in the chapter regarding how citations change and impact the making of the discipline remain unanswered. But, even though we cannot muster answers to those questions, we can with renewed conviction accept what David Foster described in 1988 as an "invitation to an intellectual pluralism" (p. 39), within which we can embrace these abstracting practices and the insights and questions they might productively open up for us.