PART I

The Digital World
Chapter 1

Abstraction, Singularity, Textuality

The Equivalence of “Close” and “Distant” Reading

In a blog post entitled “A Dataset for Distant Reading Literature in English, 1700–1922,” Ted Underwood describes as “malarkey” the “version of distant reading currently circulating in the public imagination”—namely, that it analyzes “a massive database that includes ‘everything that has been thought and said.’” He continues,

In the early days of distant reading, Franco Moretti did frame the project as a challenge to literary historians’ claims about synchronic coverage. (We only discuss a tiny number of books from any given period—what about all the rest?) But even in those early publications, Moretti acknowledged that we would only be able to represent “all the rest” through some kind of sample.

Underwood is correct in a narrow sense: Moretti engages in, and occasionally acknowledges his use of, data sampling. But it does not follow that the public imagination, or the mainstream media outlets feeding it, confected the view of distant reading as enabling direct and objective access to a comprehensive literary-historical record. Moretti’s work provides more than ample grounds for this public perception, as does Jockers’s closely related paradigm of macroanalysis. While claiming
direct and objective access to “everything,” these high-profile authors represent and explore only a very limited portion of the literary system, and do so in an abstract and often ahistorical way.

Moretti has been criticized in similar terms previously: for adopting a reductive approach to literature and associating data with comprehensive and authoritative knowledge. Those who reject any role for data in literary history maintain both of these criticisms, on the basis that data are inimical to literature, and only close reading can explore its nuance and complexity. Katie Trumpener, for instance, argues that data-based methods “violate” the “individuality” of literary works (160), while Stephen Marche insists “literature is not data. Literature is the opposite of data” (160). James English attributes such responses to the discipline’s foundationally “negative relationship” to “counting,” noting its intensification in the face of “ever more stringent quantification regimes of value and assessment,” as well as Moretti’s role in exacerbating that oppositional perspective (xii, xiii).

Those who advocate the use of data in literary studies typically deny that Moretti’s approach is needlessly reductive. Echoing Moretti’s account in Distant Reading of the method as “a little pact with the devil” (48), they acknowledge that abstracting and simplifying complex phenomena is an inevitable consequence of quantitative approaches, but one that is justified by the new forms of knowledge it enables (Love 374). Regarding Moretti’s tendency to “overestimat[e] the scientific objectivity of his analyses” (Ross np) opinion is more divided. Some who support a data-rich approach to literary history perceive Moretti’s claim to authoritative knowledge as an unfortunate side effect of his polemical intent to challenge literary history’s reliance on close reading. As Tim Burke writes, “There is no requirement to purchase the entire methodological inventory he makes available, or to throw overboard close reading or aesthetic appreciation” (41). Others ascribe a more foundational essentialism to Moretti’s work. John Frow argues that Moretti conceives of “literary history . . . as an objective account of patterns and trends” by “ignor[ing] the crucial point that these morphological categories he takes as his base units are not pre-given but are constituted in an interpretive encounter by means of an interpretive decision” (“Thinking” 142).

In my view, these critiques describe the symptoms—not the essence—of a problem, which also characterizes Jockers’s macroanaly-
sis, as well as the New Criticism’s core method of close reading. Con-
trary to prevailing opinion, distant reading and close reading are not 
opposites. These approaches are united by common neglect of textual 
scholarship: the bibliographical and editorial approaches that literary 
scholars have long depended on to negotiate the documentary record. 
Because of this neglect, like the New Critics before them, Moretti and 
Jockers cannot benefit from the critical and historical insights present-
et by editorial and bibliographical productions. As a consequence, 
both authors conceive and model literary systems in reductive ways and 
offer ahistorical arguments about the existence and interconnections 
of literary works in the past.

I

Underappreciated in commentary on distant reading and macro-
analysis is the shifting meanings of both terms. When Moretti origi-
nally proposed distant reading in “Conjectures on World Literature” 
in 2000, it was a “new critical method” for world literary studies, not 
for literary history (55). Distant reading aimed to overcome the focus 
on national canons by collating the work of multiple scholars to iden-
tify and explore “units that are much smaller or much larger than the 
text: devices, themes, tropes—or genres and systems” (57). With his 
framework of world literary studies was superseded by literary history 
(indeed, a national formation of that endeavor, focused on eighteenth-
and nineteenth-century British literature). While units smaller or larg-
er than the text were theoretical notions in his “Conjectures” essay, in 
this book they are translated into data points. A systemic approach to 
literary history also became central, with the abstract modeling devices 
of the title—graphs, maps, trees—employed to explore, and to visual-
ize, the operations of a literary history that “cannot be understood by 
stitching together separate bits of knowledge about individual cases, 
because it isn’t a sum of individual cases: it’s a collective system, that 
should be grasped as such, as a whole” (*Graphs* 4).

Computational methods and digital resources were in turn cen-
tral to Moretti’s 2013 book, *Distant Reading*, but there literary history 
was ceding ground to “the theory of literature” as the focus in “the
encounter of computation and criticism” (Moretti, “Operationalizing” 9). Although literary data remains central, the primary object of distant reading is now less often literary systems—designated social, material, and political contexts for literary development and change—than the “concepts of literary study” (1). And while Moretti previously identified the importance of literary systems in their inclusion of the “great unread” (“Conjectures” 55), these concepts of literary study (including characterization, plot, and dramatic form) are approached predominantly through formal and computational analyses of canonical literary works.¹ Jockers’s focus has remained more consistent over time. But his recent work demonstrates this same shift from literary history—his explicit concern in his 2013 book, *Macroanalysis: Digital Methods and Literary History*—to categories of literary analysis: in his case, plot and characterization.² Yet even as Moretti and Jockers have moved from a historical to a conceptual emphasis in their own work, distant reading and macroanalysis dominate—and limit—public, and much academic, perception of what data-rich literary history entails.

Pace Underwood’s defense, in their literary-historical work both Moretti and Jockers present literary data and digital collections as pre-critical, stable, and self-evident. In conceiving data and computation as providing direct and comprehensive access to the literary-historical record, they deny the critical and interpretive activities that construct that data and digital record and make them available for analysis. In Moretti’s work on literary history, literary data are consistently presented as facts rather than interpretations. Thus the first chapter of *Graphs* repeatedly references “the large mass of [literary] facts” as “ideally independent of interpretations” (3), as “data, not interpretation” (9), and as “useful because they are independent of interpretation” (30). On this basis, Moretti accords his arguments an unrealistic exactitude. For instance, his claim that bibliographical data “can tell us when Britain produced one new novel per month or week or day, or hour for that matter” (9) denies the inevitable gaps between the publishing context and the bibliographies he proposes to explore them with.³ Similarly, Moretti presents data visualization as a transparent window onto history, with the idea that “graphs, maps, and trees place the literary field literally in front of our eyes—and show us how little we still know about it” (2). The same understanding of literary data appears in *Distant Reading*, where Moretti celebrates data visualization as providing “a set of two dimensional signs . . . that can be grasped at a single
glance” (211). Such descriptions, which substitute seeing what is there for the interpretive acts involved in constructing literary data, organizing it, and ascribing a historical explanation to the results, underpin Moretti’s contention to explore “the literary field as a whole” (67).

While this view of literary data as factual and transparent has been noted—and critiqued—such assessments miss its underlying cause: Moretti’s lack of interest in the scholarly infrastructure that enables his analyses. For results derived from analog bibliographies—as in the first chapter of Graphs and his stylistic “Reflections on 7,000 Titles” in Distant Reading—parentheses and footnotes occasionally admit that comprehensive access to the facts of literary history is not achieved. For example, figure 7 in the latter study, showing the number of British novels, stops in 1836, while the other graphs extend to 1850. In a footnote, Moretti explains the discrepancy with the comment “it seems very likely that Andrew Block’s bibliography significantly overstates the number of novels published after that date” (188). Yet acknowledging that his dataset arises from a (“significantly”) flawed interpretive encounter affects neither Moretti’s rhetoric nor his subsequent analysis. The chapter still claims to “read the entire volume of the literary past” (58), and while the data are absent from figure 7, Block’s bibliography is the only source for titles published from 1836 to 1850. Moretti proceeds, in other words, by analyzing titles he knows never existed.

While literary data derived from analog bibliographies are only “ideally independent of interpretations,” Moretti regards mass-digitized collections as achieving this independence. With such collections becoming the rhetorical, if not the primary analytical, focus of Distant Reading, Moretti looks forward just “a few years,” to when “we’ll be able to search just about all novels that have ever been published and look for patterns among billions of sentences” (181). He notes that, while literary studies has previously experienced “the rise of quantitative evidence . . . without producing lasting effects, . . . this time is probably going to be different, because this time we have digital databases and automatic data retrieval” (212). While digital technologies are celebrated and foregrounded, beneath these claims lies the same disregard for the specifics of the disciplinary infrastructure that characterized Moretti’s approach to bibliographies. This attitude is apparent in an interview in which Moretti aligns digital humanities with three elements:
new, much larger archives; new, much faster research tools; and a (possible) new explanatory framework. The archives and the tools are there to stay; they are important but not intellectually exciting. What appeals to me is the prospect of a new explanatory model—a new theory and history of literature. (Moretti, “Bourgeois” np)

In presenting digital “archives” or collections as “there to stay,” Moretti disregards their status—like bibliographies—as interpretative constructs. And unlike print-based bibliographies, most digital collections are constantly changing: a dynamism with significant practical and conceptual challenges for literary history.

Asserting comprehensive access to the historical literary record is even more essential to Jockers’s notion of macroanalysis. Although he sometimes presents his method as complementing rather than replacing idiographic approaches to literature (Macroanalysis 171), foundational to macroanalysis is Jockers’s view of interpretation as methodologically defective: “Interpretation is fueled by observation, and as a method of evidence gathering, observation—both in the sciences and in the humanities—is flawed” (6). While interpretation and observation are “anecdotal and speculative,” “big data” is supposedly constituted without human involvement and thus offers “comprehensive and definitive” historical facts (31). According to Jockers, literary scholars “have the equivalent of . . . big data in the form of big [digital] libraries . . . [or] massive digital-text collections,” and these enable “investigations at a scale that reaches or approaches a point of being comprehensive. The once inaccessible ‘population’ has become accessible and is fast replacing the random and representative sample” (7–8). As Jockers says of one of Moretti’s analyses, such unprecedented and supposedly uninterrupted access to the documentary basis of literary history “leaves little room for debate” (20): a perspective that overlooks the fact that all collections are selections, made according to (implicit or explicit) arguments about value, and with varying degrees of expertise and funding.

Jockers employs a number of scientific metaphors to buttress this association of scale and comprehensive access, the most explicit being “open-pit mining or hydraulicking.” While “microanalysis” (including reading and digital searching) discovers “nuggets,” macroanalysis accesses “the deeper veins [that] lie buried beneath the mass of
gravel layered above” (9). In working with the “gravel” of literary history, employing “the trommel of computation to process, condense, deform, and analyze the deep strata from which these nuggets were born,” Jockers supposedly “unearth[s], for the first time, what these corpora really contain” (10, my italics). This metaphor not only renders literary history as concrete an entity as a mountain: all of it can be accessed and analyzed. It also conflates analysis with the achievement of complete access. The network visualizations with which Jockers presents the cumulative results of macroanalysis reinforce this view of literary data as factual and comprehensive. As chapter 5 explores in depth, because the form and meaning of most network graphs change when new nodes and edges are added, to claim that they display the structures and relationships that organized literature in the past implies that all data are available for analysis.

While Moretti occasionally acknowledges limitations in his data (before proceeding with analyses regardless), Jockers maintains that any “leap from the specific to the general” is flawed because based on interpretation (28). Only in the book’s final chapter does he admit the obvious gap between his datasets and the “population” of nineteenth-century novels, describing his largest “corpus of 3,346 texts” as “incomplete, interrupted, haphazard,” and noting, “The comprehensive work is still to be done” (172). This concession generates an awkward comparison of macroanalysis with Charles Darwin’s theory of evolution, which reinforces Jockers’s equation of knowledge with scale and comprehensive access. While both are “idea[s],” because “there are further dimensions to explore” (171), literary scholars are advantaged over evolutionary biologists “in terms of the availability of our source material” (175). In a context where bigger is better—as Jockers says elsewhere in the book, “eight is better than one, [but] eight is not eight thousand, and, thus, the study is comparatively anecdotal in nature” (25)—his “3,346 observations and 2,032,248 data points” are seemingly indicative of knowledge in and of themselves (172). Jockers concludes by admitting one impediment to macroanalysis, but it is only legal: though almost “everything has been digitized,” post-1923 publications remain (at the time he was writing) protected by copyright, leaving literary scholars dependent on legal reforms before they might realize “what can be done with a large corpus of texts” (175).

A recent collaboration of the Stanford Literary Lab, which Jockers
is no longer part of, departs in one important way from the approach to literary history, data, mass digitization, and computation I have described (Algee-Hewitt et al.). Literary Lab Pamphlet 11 closely attends to the gaps between “the published” (all literary works made public in history), “the archive” (the portion of what was published that has been preserved and is now increasingly digitized), and “the corpus” (the segment of the archive selected for a research question). Although incorrectly imagining that the “convergence of these three layers into one . . . may soon be reality,” in acknowledging that this state has not been achieved, the authors admit the constructed—and selective—nature of literary data. Yet Pamphlet 11 follows Moretti’s and Jockers’s precedent in misconstruing the nature of our disciplinary infrastructure, in that the authors presume to overcome the selections and biases of mass-digitized collections by using analog bibliographies to generate “a random sample” of what was published (2). This strategy misses the vital point that both digitized collections and analog bibliographies are derived from “the archive,” predominantly the collections of major (usually American or British) university libraries. Pamphlet 11 also replicates Moretti’s and Jockers’s approach in not publishing its datasets.

Moretti often references his sources of data—chapter 1 of Graphs, for instance, begins by listing the bibliographies it draws upon—and he advocates data sharing: “Because . . . data are ideally independent from any individual researcher, [they] can thus be shared by others, and combined in more ways than one” (5). Moretti, however, does not share his data. Jockers occasionally publishes the results of data analysis, such as the five hundred themes developed from topic modeling, presented as word clouds on his website (Jockers, “500”). But he does not provide the textual data analyzed, even at the level of word frequencies, and is significantly less open than Moretti about the composition of his datasets. Although in more recent work Jockers adjusts this approach somewhat, for research pertaining to Macroanalysis I have discovered only one instance in which he indicates the titles and authors investigated, and then, only for 106 of the total 3,346 works. These are identified almost incidentally, in reporting confusion matrices (Jockers, “Confusion”).

In Moretti’s case, one might suppose it possible to reconstruct his datasets from cited sources. But his account (in an appendix to Graphs) of creating the dataset for “British novelistic genres, 1740–1900,” high-
Abstraction, Singularity, Textuality • 25

lights why this is not feasible. There he describes his periodization as “not always explicit” in the bibliographies (31), thus evincing the role of his own—unpublished and therefore unspecified—interpretive decisions in data construction. And even if Jockers listed all the titles and authors he analyzed, it would be impossible to reconstruct the basis of his arguments without access to the textual data he uses, which are not just texts of literary works but highly prepared—or preinterpreted—selections from those texts. Neither critics nor supporters of Moretti’s and Jockers’s methods typically comment on this lack of published data. But far from an incidental oversight, this practice maintains the fiction that literary data are prior to interpretation: it removes the need either to describe the procedures for collecting, cleaning, and curating datasets or to expose the inevitably selective and limited collections resulting from that construction.

The meaning derived from a literary-historical dataset—like the interpretation of a literary work—is shaped, profoundly, by the methodological and critical frameworks through which it is approached, and by the selections and amplifications those frameworks produce. Two scholars can read the same dataset—like the same literary work—and derive different meanings. While an independent observer may be more or less convinced by the different arguments, deciding between them depends upon access to the object on which they are based. In the absence of data publication, distant reading and macroanalysis are analogous to a scholar finding a set of documents in an archive or archives, transcribing them, analyzing those transcriptions, publishing the findings, and asserting that they demonstrate a definitive new perspective on the literary field, without enabling anyone to read the transcriptions (or in Jockers’s case, without revealing the titles of most of the original documents).

II

As noted in the introduction, Moretti and Jockers have been highly influential in foregrounding data-rich models of literary systems as primary units of historical analysis. Yet in not recognizing the critical and constructed nature of the bibliographies and mass-digitized collections they use to create these models, neither author can benefit from
the historical insights that underpin and are translated through this scholarly infrastructure. The resulting models can be used to address certain questions. But they are ultimately reductive: these models do not represent the historical existence of literary works, including the ways in which they connect to produce literary systems.

In modeling literary systems, Moretti and Jockers define literary works as single entities in time and space, typically located as such by the date of first book publication and the author’s nationality. Literary works are constituted as literary systems when they share these basic features—as in “nineteenth-century” or “British” novels—with other characteristics added within that framework (such as the author’s gender or the work’s genre). This basic understanding of literary systems is evident, for instance, in Moretti’s analysis of “7,000 titles (British novels, 1740 to 1830)” in *Distant Reading* (179–210), discussed above, or of eighteenth- and nineteenth-century British novels defined in terms of the number of titles published, authors’ genders, or fictional subgenres (*Graphs* 5–9, 26–27, 28–30). It likewise underpins Jockers’s exploration in *Macroanalysis* of “758 works of Irish-American prose literature spanning 250 years” (37) or 3,346 nineteenth-century British and American novels.

Depending on the reliability of the source and the type of questions asked, datasets constructed on this basis can support insights into trends in new literary production. Jockers’s study of Irish-American prose pursues an approach manifested in other digital projects—some of my own work included (Bode, *Reading*)—of using publication data to test existing perspectives on literary history. Employing a dataset with the date of first publication, as well as “the geographic settings of the works, author gender, birthplace, age, and place of residence” (36), Jockers challenges the notion of a “lost generation” of Irish-American authors from 1900 to 1930 and proposes a likely explanation for this misperception: a predominance of eastern male authors in the canon—and hence in critical assessments—of Irish-American literature (38–48). Moretti’s work on new literary production extends beyond testing and revising particular arguments in literary history and is highly innovative in this respect. His study of titles, for instance, investigates a category of literary data that had not, as far as I know, been subjected to synoptic, stylistic analysis previously. More broadly, Moretti combines multiple bibliographies to challenge claims about
the distinctiveness of new literary production in different historical periods, as in his discussion in Graphs of gender trends in British novel publication (17–20).

But literary works are not defined by a single time and place, and collecting them together in those abstract terms does not represent the interconnections that constitute literary systems. William St Clair’s The Reading Nation in the Romantic Period aptly diagnoses the limitations of this approach. Like Moretti and Jockers, St Clair rejects what he calls the “parade of authors” convention in literary history, where canonical authors file past the commentator’s box in chronological order, taken as representative of the historical period in which they wrote. But he equally dismisses the “parliament of texts” approach, where literary works first published at a particular time, and usually by authors of a particular nationality, are understood as “debating and negotiating with one another in a kind of open parliament with all the members participating and listening” (4). As St Clair notes, literary systems frequently include “texts written or compiled long ago and far away” (3), and some literary works are inevitably more widely published, circulated, read, and referenced than others.

New domestic literary production, the basis of Moretti’s and Jockers’s datasets, is only a subsection of the literature available at any time and place. By considering only that component—without accounting for its relative importance or acknowledging that literary systems encompass other types of works—Moretti and Jockers occlude major aspects of how literature existed in the past. The date of first book publication overlooks the differing availability of literary works in the years after they are published and that first book editions are not necessarily—and for many periods are rarely—the first time works are available. Some titles are never published as books, and many literary works—whether in book or other formats—are republished, sometimes on multiple occasions. Likewise, an author’s nationality is a poor marker for the geographical existence of a literary work in a marketplace that has been globalized since at least the eighteenth century. More broadly, the construction of literary systems from the categories supplied by enumerative bibliographies—the title, the author, the date of publication, the publisher, and by association, the text—ignores the different titles, author names, dates and places of publication, and texts that occur as literary works are issued and reis-
sued, and the implications of these differences for understanding production and reception.

Even for studies that focus on new book publications, this approach to modeling literary systems ignores most differences between literary works, and hence, most dynamics of those systems. Reflecting on Moretti’s work, David A. Brewer notes that flattening the literary field “nicely undoes the monumentalizing that so often accompanies the literary canon,” but at the expense of ignoring the varied profiles and presences of works in history (162). Brewer focuses on commercial success, arguing that the popularity of different literary works at the time they are published and in subsequent generations accords them “a massively different footprint” in history, altering their influence, and hence their meaning, for readers (163). But commercial success is not the only relevant factor. As textual scholars show by exploring the material and social dimensions of literary works, multiple issues shape their meaning, extending from the documentary forms they take to the relative positions and prestige of the individuals and institutions involved in producing them (authors, publishers, editors, illustrators, booksellers, advertisers) and the interconnected systems (economic, religious, educational, legal, geopolitical) in which they circulate.9

While textual scholars such as Johanna Drucker (“Entity”), Paul Eggert (Securing), and Jerome McGann (New) thereby conceptualize literary works as events—unfolding over time and space and gaining different meanings in the relationships thereby formed—Moretti and Jockers construct literary systems as composed of singular and stable entities and imagine that this captures the complexity of such systems. In fact, because their datasets miss most historical connections between literary works, their analyses rely on basic features of new literary production to constitute both the literary phenomenon requiring explanation and the explanation for it. Macroanalysis purports to investigate “the context in which [literary] change occurs,” chiefly by analyzing words in nineteenth-century novels (156). What Jockers actually shows is the capacity of his computational method (a combination of stylistic analysis, topic modeling, and network analysis) to predict whether a work (or “bag of words” from that work) was by a man or woman, and its date of publication, genre, and national origin, from a corpus defined according to those parameters. Notwithstanding the variable accuracy of this approach for different categories,10 the methodologi-
cal demonstration is impressive for extending stylistic analysis beyond small groups of documents.

But the methodological achievement does not translate into historical insight because the study considers only an abstract amalgam of literary works. In reducing context to a few predetermined categories, Jockers is confined to stating their presence. He cannot offer any alternative influences, nor can he comment on the extent to which gender, nationality, and chronology shape literary history, except perhaps implicitly, in the proportions of titles misidentified by his models. The approach yields very general, and I would argue, self-evident statements. To give examples drawn from the conclusions to Jockers’s various chapters: “the linguistic choices an author makes are, in some notable ways, dependent upon, or entailed by, their genre choices” (104); “there are both national tendencies and extranational trends in the usage of . . . word clusters” (114); “a writer’s creativity is tempered and influenced by the past and the present, by literary ‘parents,’ and by a larger literary ecosystem” (156); and “thematic and stylistic change does occur over time” (164). The generality of these conclusions is predetermined by the dematerialized and depopulated conception of influence underpinning the analysis. The model constitutes literary works as a system based on the date of (presumably first book) publication, and any book within the dataset is understood to exert influence in a chronologically discrete manner, regardless of the actual conduits of literary influence, which require availability to readers who buy, borrow, and sometimes write literary works. Because he is modeling a diffused and generalized system, the “influence” of gender, genre, temporality, and nationality is in turn diffuse and generalized.

The inadequacy of this conception of literary systems is foregrounded when Moretti considers readers, who, as Anne DeWitt notes, “are both central to his argument and absent from his evidence” (162). Moretti takes literary data on publication and/or formal features of literary works as both expressive of and explicable by the actions of readers and the market. We can see this strategy in Moretti’s discussion of the first graph in Graphs: the “rise of the novel” across a number of national contexts (Britain, Japan, Italy, Spain, and Nigeria) at different times. Leaving aside the question of whether his graph depicts the numbers he attributes to it, Moretti ascribes the leap “from five–ten new titles per year . . . to one new novel per week” to “the horizon of
novel-reading,” the shift in the market that occurs when the novel is transformed from “an unreliable commodity” to “that great modern oxymoron of the regular novelty: the unexpected that is produced with such efficiency and punctuality that readers become unable to do without it” (5). The argument makes intuitive sense, but it presumes that only—and all—new titles by authors of particular nations were available to, and read by, only—and all—readers of those nations. The explanation, in other words, claims that publication data are both indicative of national reading habits and explicable in terms of that activity.

A similarly circuitous mode of argumentation characterizes “The Slaughterhouse of Literature” chapter in Distant Reading. Moretti proposes a framework for canon formation, wherein readers are the “butchers” of literary history who read novel A (but not B, C, D, E, F, G, H . . . ) and so keep A “alive” into the next generation, when other readers may keep it alive into the following one, and so on until eventually A becomes canonized.

Readers, not professors, make canons. (67)

Nominating formal choices as the reason readers select certain titles over others, Moretti employs the example of detective fiction and decodable clues to demonstrate this process. He identifies the presence of such clues in Arthur Conan Doyle’s fiction as the reason that author was progressively selected by generations of readers to attain his now canonical status. Again, this is an interesting but circular argument. Moretti acknowledges one of the ways in which his claims are “tautological”: “if we search the archive for one device only, and no matter how significant it may be, all we will find are inferior versions of the device, because that’s really all we are looking for” (87).

Yet the same problem—of assuming the shape of the past from that of the present—occurs at a larger scale in that Moretti assumes that authors who have a canonical status in the present were selected from the time of first publication. This argument is intrinsic to his evolutionary model, and while Moretti supports it by citing an empirical study (68), others show its falsity. St Clair, for instance, demonstrates the minute early nineteenth-century readerships of five of the “big six” Romantic male poets (excepting Byron) who form the contemporary canon (660): however that Romantic canon was formed, it was not
based on the poetry contemporaneous readers preferred. While in the earlier study Moretti aligns publication with reading (a title was published; ergo it was read), in this instance his argument requires titles to be published but not read. What determines if titles were read is whether they had decodable clues; thus, once again, a feature of the data (the presence or absence of decodable clues) is used both to indicate and explain the activities of readers.

Moretti has said, in defense of his method, that reducing literary works to one or two features is part of the “specific form of knowledge” that distant reading provides: “fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models” (Graphs 1). My argument is not against reduction and abstraction per se. While especially obvious in data-rich studies (which rely on identifying attributes that can be represented in uniform fields), reduction and abstraction characterize all analysis. Close readings do not interpret literary works as a whole but specific, extracted instances of particular, abstracted features of those works. What I am arguing is lost in Moretti’s and Jockers’s approach—especially in their definition of literary systems as analogous to first book publication by authors of a designated nationality—is precisely a historical sense of “interconnection.” Failing to acknowledge that the disciplinary infrastructure they use is made not given, and thus overlooking the historical information embedded in it, Moretti and Jockers model literary systems in terms of potentially, and certainly relatively, abstract categories of production. In the process they ignore the socially, spatially, and temporally specific and complex ways in which literary works exist and relate to one another in particular, historical contexts.

III

Although distant reading initially faced considerable resistance from literary scholars, now a common response to that paradigm is the call to integrate nondata- and data-based approaches. While Moretti originally suggested that distant reading should replace close reading, this integrated position is the one he subsequently adopted. Describing the contrast between close attention to the canon and distant exploration of the archive in terms of “too much polyphony” on the one hand and
“too much monotony” on the other, Moretti asserts, “It’s the Scylla and Charybdis of digital humanities. The day we establish an intelligible relationship between the two, a new literary landscape will come into being” (Distant 181). Other examples of this stance abound, including from Jockers (Macroanalysis 26) and scholars such as Frederick Gibbs and Daniel Cohen, who argue for the profitability of “mov[ing] seamlessly between traditional and computational approaches” or between “our beloved, traditional close reading and untested, computer-enhanced distant reading” (70).

This apparent moderation in the terms of debate belies the continuing perception of close and distant reading (or micro- and macro-analysis, or nondata- and data-based approaches) as opposites. Whether close reading is presented as less “rational” than distant reading (Moretti, Graphs 4), or more authentic and authoritative (Trumpener), or if together, the two perspectives are understood to supplement the others’ limitations, close and distant reading are conceived as antithetical in their assumptions and approaches. However, the main problems I have sought to diagnose in distant reading and macroanalysis—a disregard for textual scholarship and an assumption that literary works are stable and singular entities—are ones they share with the New Criticism and its foundational method of close reading. Distant reading and macroanalysis take the core object and premise of the New Criticism—the decontextualized text as the source of all meaning—to a conclusion rendered more abstract and extreme by the number of texts under consideration.

As is well known, the New Criticism was an early- to mid-twentieth-century movement that subordinated the historical and contextual (biographical, material, sociological) concerns of literary scholarship to “the text” itself. The critique of this movement is also well established, with the contextual focus in many subsequent forms of literary history—feminism, postcolonialism, New Historicism—explicitly rejecting the New Critical view of the text as a self-contained and self-referential aesthetic object. Despite the apparent demise of the New Criticism, the continuing centrality of close reading in literary studies, including literary history, and the rhetorical focus of such research on the text, perpetuates the earlier movement’s dismissal of textual scholarship (Cain). As Eggert (“Book”), McGann (“Note”), and others have observed, assuming that literary works are texts, and that texts
are single, stable, and self-evident entities, dismisses the documentary record’s multiplicity, and with it the critical contributions of those endeavors—bibliography and scholarly editing—dedicated to investigating and understanding that multiplicity.

Even as contemporary enactments of close reading often foreground context, the centrality and assumed singularity of the text, and the disassociation from the literary work’s complex historical existence this produces, can negatively impact the capacity of such analyses to investigate literary history. Mary Elizabeth Leighton and Lisa Surridge highlight this effect in critical discussion of Wilkie Collins’s *Moonstone*. Describing the varied interpretations that contemporary critics offer of its meaning for nineteenth-century readers—from tale of “imperialist panic” to critique of colonial domination—they note that all critics assume, first, that they are “reading the same text” as readers in the past and, second, that all past readers encountered the same text as each other. In fact, as Leighton and Surridge show, *Moonstone* “took on strikingly different forms—and hence different meanings—in different markets,” specifically in British and American serializations (207). In projecting textual singularity onto a historical period characterized by documentary multiplicity, the close readings these critics produce obscure the historical production and reception of this literary work even as they propose to emphasize that context.

Notwithstanding such instances, close readings are generally protected from the abstraction inherent in the notion of the text by the knowledge infrastructure in which they are embedded, and by a focus on particular documents. Scholarly editions provide critics with carefully historicized texts for consideration; when one is not available, the standard practice of bibliographical referencing ties discussion of the supposedly singular text to a version of the work. Moreover, because a close reading inevitably analyzes a version, any discussion of the text is contextualized by the information about the work’s history contained in the material form that the critic assesses. Distant reading and macro-analysis do not benefit from such provisions or protections; to the contrary, as this chapter has argued, these approaches negate the interpretive nature of the disciplinary infrastructure they use, as well as their own role in constructing the meaning of the data they derive from it.

While explicitly opposing close reading, the form of Moretti’s and Jockers’s arguments mirrors the New Criticism’s perception of the text
as the source of all meaning, even as the text under consideration has expanded from a single version of a literary work to a version of bibliographical and textual data derived from multiple versions of literary works. Moretti’s investigations of readers based only on data relating to first book publications enacts this view of the text as inevitably containing all that is relevant to interpreting it. In treating the literary system as a dispersed linguistic field, Jockers takes the rhetoric of text to its ultimate conclusion, proposing a literary-historical world in which there are no structures beyond the textual. “Signals” of gender, genre, or nationality, comprised entirely of word frequencies, are substituted for gender, genre, or nationality as historical and cultural constructs. Far from the opposite of close reading, the dematerialized and depopulated understanding of literature in Jockers’s work enacts the New Criticism’s neglect of context.

Whether literary histories are conducted in traditional or data-rich forms, the outcomes of analysis are inevitably tied to the object analyzed. When a gap exists between the contemporary object assessed and the historical object it supposedly represents—and when the critic is unaware or dismissive of that gap—no degree of nuance or care in the reading can supply that historical meaning. Herein lies the fundamental problem with proposing to integrate close and distant reading as the obvious way forward for research in literary history. Understood in terms of the different perspectives the two approaches offer, this strategy seems eminently sensible: data-rich analysis has the potential to explore large-scale patterns and connections in ways that nondata-rich research cannot; likewise, conventional textual analysis can provide insights into the meaning of literary works that quantitative studies cannot.

Yet in couching debate about the role of data purely in terms of method, this response maintains the focus on the mode of analysis employed and conceals the lack of an adequately historicized object to analyze. What data-rich literary history needs is an object capable of representing literary systems—as manifestations of literary works that existed and generated meaning in relation to each other in the past—while managing the documentary record’s complexity, especially as it is manifested in new digital knowledge infrastructure. The lack of such an object, not the fundamental opposition of data and literature, is the real reason it has proven so difficult, in practice if not in theory, to
integrate “traditional and computational methods” for the purposes of historical investigation (Gibbs and Cohen 70).

An appropriately historicized representation of relationships between literary works in the past would avoid the problems I have identified in Moretti’s and Jockers’s approach, which claims to represent everything—directly, comprehensively, and objectively—while exploring only a limited part of any given literary system. The difficulties with their approach are not the result of using data to investigate past literature. They occur because distant reading and macroanalysis adopt and perpetuate the disregard for textual scholarship foundational to the New Criticism, without benefiting from the institutional and infrastructural protections afforded to close reading. Given this source of the problem, the next chapter proposes a solution from textual scholarship. I argue that the field’s foundational technology of the scholarly edition supplies both the supports and constraints necessary for data-rich literary history, while providing a framework capable of extending the insights gained from engagements with emerging digital infrastructure to the broader discipline.