

# #

## Chapter 2: Patterned Images of a Discipline: Database, Scale, Pattern

### Visualizing a Discipline

Over the last two decades, numerous researchers, artists, and software developers have produced work that shares basic qualities with distant reading and thin descriptive methods and that by demonstrative force add depth and credence to those methods' viability. Gathering a few of these projects—to notice them and to become more directly familiar with their designs—is useful for establishing a richer notion of the practices and operations involved with distant–thin methods. Of course, not all of these projects—a small sample of which are listed below—follow distant–thin methods *self-consciously*, nor do all of them identify with RCWS or with the digital humanities, but they nevertheless add up to a rising investment in alternative treatments of large, complex, and unwieldy textual collections—across a variety of domains, academic, popular, and public–civic. In each of the following examples, data sets were transformed into something different—a tag cloud, an interactive network visualization, a weighted list, a re-arranged film, and so on, each punctuating the recent history of distant reading and thin description and indicating an expanding methodological milieu for the 21st century that augments the traditionally privileged inquiry-proximities of the close and the thick.

- *Wordcount* by computer programmer and artist Jonathan Harris. Harris's (2004) website described WordCount as “an artistic experiment in the way we use language.” The project displays all the words from the British National Corpus in a horizontal sequence based on commonness, or frequency of appearance.
- Michael J. Faris's visualizations of citation networks among queer rhetorical scholarship. The in-progress (as of October 2017) project includes six different network visualizations based on nearly 250 publications.
- The serialized, iterative publications of chapters from Franco Moretti's *Graphs, Maps, Trees* (2007) as articles in the *New Left Review*, beginning in November–December of 2003 and continuing for two

subsequent issues through the summer of 2004. Later the articles were expanded and paired with additional projects based on the methodology in *Distant Reading* (2013).

- Producer Lenka Clayton's (2002) release of the film *Qaeda, Quality, Question, Quickly, Quiet*, which rearranged George W. Bush's 2002 State of the Union address by splicing the film into individual word-utterances before re-ordering them alphabetically.
- Jeremy Tirrell's (2012) mapped histories of online rhetoric and composition journals, which plotted locations where digital scholarship has been produced, weighted phrases as localizable themes in digital scholarship, and proposed as promising and possible many more maps that resonate with the disciplinary atlas I sketch in Chapter Five.
- Nathan Yau's (2010) Walmart Growth Map, which presents U.S.-based geolocations for new Walmart and Sam's Club stores as a map-based time series.
- The Writing Studies Tree, a collaborative data-gathering and visualization initiative that its developers (Miller, Licastro, & Belli, 2016) characterize as "a crowdsourced, online, open-access, interactive database of individual scholars, educational institutions, and the disciplinary movements that connect them." The tree generates interactive network visualizations (i.e., zoomable, clickable) for more than 4,568 relationships among 1,744 people and 495 institutions indicative of multivariable disciplinary genealogies (Benjamin Miller, 2017, personal communication).

Although these seven projects range from scholarly publications and software applications to more experimental and artistic installations, they represent an eclectic collection that indicates how distant reading and thin description can be used to render apprehensible nonobvious patterns intrinsic to complex and relatively large data sets. As such, they also prime—and in the case of Faris's work, Tirrell's work, and the Writing Studies Tree *carry out*—new practices that extend distant and thin methodologies to engage with the rising complexity of RCWS. That is, taken together, these projects suggest some of the ways disciplinary activity might be rendered as visually traceable, the field thereby modeled as dynamic and complex networked phenomena.

Surveying this broader collection of precursors to *network sense*, let's consider one additional example in greater detail to understand how distant-thin methods have as a contemporary development emerged amidst visual modeling and abstracting practices, and, even more precisely, how database, scale, and pattern operate as bridging concepts for such work.

### The Words Speakers Use

Looking at the number of times words have been used by speakers at each party's convention suggests the different themes the parties have highlighted.

The Republican speakers have used more terms related to terrorism and the war in Iraq, while the Democratic speakers were more likely to mention health care or jobs.

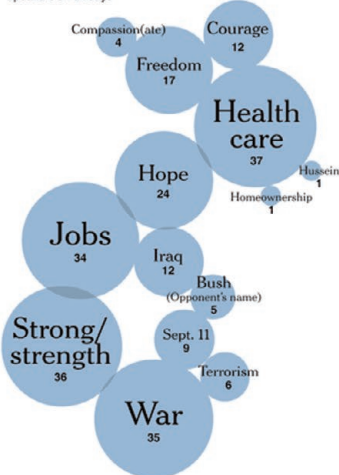
In addition, Republicans were more likely to mention the opposing candidate by name, something the Democrats rarely did.

MATTHEW ERICSON

#### Republican Convention Speakers on Monday, Tuesday and Wednesday



#### Democratic Convention Speakers on all days



SOME OF THE REPUBLICAN SPEAKERS

SOME OF THE DEMOCRATIC SPEAKERS



NUMBER OF TIMES SPEAKERS USED EACH WORD OR PHRASE	John McCain	Rudolph W. Giuliani	Arnold Schwarzenegger	Laura Bush	Zell Miller	Dick Cheney	Bill Clinton	Barack Obama	John Edwards	John Kerry
Opponent's name		12			15	13				2
Hussein	1	6		2	2	3				
Terrorism	2	21	2						2	
Religion	1	1	1							1
Afghanistan	1	2		3	2	2	1		1	
Courage	5	1			1	1	1		1	
Homeownership				3		1	1			
Freedom	7	9	4	6	12	7	2	2	2	5
Compassion(ate)			2			1				
Sept. 11	4	11		2	1	3	2		3	4
War	18	13		7	8	10	6	6	10	14
Iraq	2	4	3	4	2	3	2	3	9	4
Education		1		4		1			2	3
Strong/strength	2	3	1	4		2	5		5	8
Troops		1	2		2	2			1	4
Economy			4			2	3	1	1	3
Hope	1	4	1	4		1		13	8	4
Health care		2		1		1	5	1	9	12
Jobs				1		1	6	5	6	12

Sources: Federal News Service transcripts of speeches; Republican National Convention

The New York Times

Figure 4. The New York Times infographic "The Words Speakers Use." This infographic adopts a tabular bubble chart to highlight concentrations of convention speakers prior to the 2004 U.S. presidential election.

On Thursday, September 2, 2004, following the Democratic and Republican National Conventions, Matthew Ericson's color infographic (see Fig. 4) appeared on the front page of *The New York Times*. Dotted with an assortment of red and blue circles, the graphic presented a bubble graph to in-

dicate the “number of times words were used per 20,000 words spoken” by speakers at each of the national conventions leading up to the presidential elections later that fall. The infographic prompted considerable buzz in the days that followed—mentions of its memorable adaptation of bubble graphs to stand in for recurrent words and phrases stirred in conversations among those following the approaching election or those interested in the methodology of its rendering. The information graphic drew attention specifically for the way it aggregated key words from the lengthy transcripts, translating the debates into an abstract visual model—thin and distant—that concentrated several days of convention addresses into just a few column inches of front page space. Further, the information graphic boosted interest in what was at the time an expanding set of automated tools available online for processing samples of texts, from single paragraphs and simple lists to full chapters, monographs, and speech transcripts into a concordance of recurring terms.

“The Words Speakers Use” information graphic operates in this chapter as an anchoring example of data visualization techniques used to render variously scaled textual corpora. Similar processes—typically grouped under the heading of text mining or data mining—have been around for decades, but until recently, they were not so widely available for timely, inexpensive, everyday use. Data-mining methods involve varying degrees of complexity depending on which parts of text are to be treated. Historical precedents for data mining span the work of linguists and indexers and include the labor-intensive pre-digital initiatives to develop concordances for large collections of legalistic, literary, and religious texts. Undeniably these precedents have bearing for the methods carried out across this project. Rather than develop more deeply the history of generating keyword concordances, for now, I refer to “The Words Speakers Use” because it pinpoints one watershed moment for distant–thin methods circulated publicly as an infographic. As such, it also reinforces the connections among Franco Moretti’s distant reading, Heather Love’s thin description, and Johanna Drucker’s theorization of graphesis, which I addressed in some detail in the previous chapter, and it serves as a reference point for the concepts key to the remainder of this chapter.

The presentation of transcript-based patterns (i.e., concentrations of nouns and noun phrases, comparatively aligned to see the frequency of usage by individual convention speakers and also by the collective lineups) bears resemblance to the set of visual models and abstracting practices advanced at the nexus of distant reading and thin description—a nexus this chapter sets out to orient and to deepen conceptually by anchoring these methods to three key concepts: *database*, *scale*, and *pattern*. “The Words Speakers Use” offers a

thin–distant rendering of the national conventions, reproducing transcripts in the form of labeled bubble charts. The graphical presentation involves aesthetic choices. Applying a slight translucence, the bubbles are shaded with red or blue, corresponding to the arch-color associated with the primary political parties. Shadow-outlines of each of the figures suggest placeholders—silhouettes asserting that what matters more than the detailed physical appearances of the speakers are the utterances themselves. The wordcounts may be read for the collective contributions of *all* speakers at each convention or read for the recurrent terms in the speeches of *individual* speakers from either party. The graphic presentation is designed to amplify selected aspects of the data, such as the disproportionate number of references across party lines to topics like health care, war, and terrorism. Bubbles and the two-column alignment makes this much more than a simple listing of recurrent terms—”The Words Speakers Use” lays bare the focal terms in each party’s language choices relative to the other’s.

Ericson’s *Times* infographic influenced subsequent (also increasingly interactive) infographics following the 2011 conventions and the Republican National Convention in 2012 (see, e.g., Ericson, 2011; Ericson & Bostock, 2012). Thus considering this a noteworthy precursor to the expanding milieu of infographics, I elaborate upon this example at the outset to set up the premises that ground the chapter ahead. In this chapter, I set out from the occasion of numerous data visualization projects, including those above, to discuss selected conceptual dimensions related directly to the development of abstract visual models rendered through distant reading and thin description methods. Following a gradual but intentional build-up from *concepts* to *practices* and finally to *consequences*, I sketch the importance that each of three concepts—database, scale, and pattern—has for the *visual modeling* and *abstracting practices* that I argue offer generative heuristics for a *network sense* of the discipline. Database, scale, and pattern name three concepts vital to the advancement of distant–thin methods as well as the models and practices these methods carry out in service of network sense—a capacity for knowing as interconnected divergent aspects of the field constituted by its discourses, citations, and emplaced professional activities.

Further, the conceptual–theoretical framework for distant–thin methods insists on their functioning *heuretically*, to use Gregory Ulmer’s term, which emphasized the ways these models and practices destabilize commonplaces, opening up new questions, provoking insights, promoting speculation, and stimulating conversations that will encourage those invested in notions of disciplinarity—including newcomers in particular. This inventional emphasis is especially interested in revisiting long-established assumptions about RCWS leveled predominantly by the problematic processes identified

in Chapter One: anecdotal/experiential, local, and predominantly invested in ethnographic methods, underpinned by fixations on deep, close reading and thick description. For the complementary alternatives laid out ahead, I have included mention of heuristics to ensure that the scalable, abstract visual models assembled function as considerably more than representations. While they are representative, they are also open-ended, adaptive, and subject to updates; they are not static, final statements but concentrations of data and metadata with a high degree of tolerance for keeping up with the inevitably shifting terrain of the field—for inciting those who make them and who interact with them to anticipate, speculate, wonder, and project a generative curiosity onto a future horizon.

To deepen the emphasis on generativity across these distant–thin practices, consider again the ways thin description functions to describe, and just what describing *does*. By assigning language to experience, description makes experience accessible (albeit *partially* accessible) and durative (albeit *limitedly* durative), setting experience into rhetorical circulation. This accomplishes, strictly speaking, a mediated shifting-out, and although the shift-out is beholden to what is and what was, the description itself functions as an act of becoming for the ways it shapes a future horizon. As such, description is generative, a heuristical intervention. Description participates in setting up what is next just as much as it freezes in representational amber what was or what has been. This recognition of description as generative is influenced by Bruno Latour, who wrote (in a pseudonymous essay) that

descriptions are always in words and appear very much like semiotic commentaries on a text or like a programming language. They define actors, endow them with competences and make them do things, and evaluate the sanction of these actions very much like the narrative program of semioticians. (Johnson, 1988, p. 306)

In this context, description sets a script that participates in a gradual transformation of agency, action, and materials in the world.

Consider Ericson’s infographic once more in these terms: It represents the discourses of each convention, but, by reducing, aggregating, and simplifying those discourses, it reduces complexity *temporarily* to the semantic patterns intrinsic to each party’s platform. Thus, understood as rhetorical constructs, perhaps abstract visual models like this can, in turn, influence the future discourses they sponsor. The distant–thin infographic is both descriptive and generative. In what follows, I will address each of three concepts key to distant–thin methods—database, scale, and pattern—so that the elaboration of specific models in the second half of this project will be enriched.



6

Because the distant–thin methods demonstrated in this book are both data-driven and in service of network sense, the case for them cannot proceed responsibly without acknowledging the availability of suitable data sets and the related matter of disciplinary databases, their sporadic curation, and their perilous sustainability. The previous chapter established that numerous disciplinographic accounts have been built on testimonial and ethnographic methods. They have relied upon thick description, tended toward archival methods and local–historical cases, and adopted narrative—storytelling—as the predominant discursive mode. A well-documented attachment to thick-descriptive and narrativistic approaches among disciplinographic accounts is worth reconsidering because it cannot help but elicit the issues I address here on the status of disciplinary databases, the interdependence of database and narrative, and the importance of revitalizing what I call *databasic infrastructure* for RCWS in service of stabilizing disciplinary trajectories cognizant of newcomers, divergent stakeholders, and disciplinographic accounts capable of involving durable data sets and narrativistic impulses. I contend that, although they have been tremendously important, hyperlocal, narrative-based accounts of disciplinary emergence operate more powerfully when paired with data-based accounts. In addition to composing narrative accounts, scholars must also begin to systematically build and curate the field’s databases (e.g., program profile data, directories of programs, journal indexes, etc.).

Contemporary definitions of databases often foreground the record-like storage of electronic objects (or object-markers that stand in for physical objects, such as would be the case for an inventory spreadsheet). In “The Database as System and Cultural Form,” Christiane Paul (2007) distinguished five types of databases: hierarchical, network, relational, client/server, and object-oriented (p. 96). Hierarchical and network databases privilege predict-

6 Chapter Two’s arrangement presents as section breaks a series of five iconic graphic organizers retrieved in 2017 from *The Noun Project*: “Database” by Ed Jones, “Scale” by Oliviu Stoian, “Pattern” by Alex Fuller, “Bar Chart” by Alfredo, and “Bear” (i.e., Abstracting Practices) by Musaplated.

able structural relationships among data housed in the system; hierarchical databases adhere to tree-like logics, whereas network databases follow comparable branching but are more bush-like, allowing elements to splice from multiple higher-order elements. According to Paul, client/server databases support remote, distributed access, and object-oriented databases are usually built in anticipation of computational uses, such that the elements organized in them are readily assemblable, particularly for use with different object-oriented programming languages. Among the five types, however, relational databases have achieved the broadest contemporary favor for everyday uses because they are highly flexible, allowing content to be sorted, re-organized, and modeled with high degrees of flexibility. For a more general definition that reaches across these specific types, Lev Manovich (2007) wrote,

In computer science, database is defined as a structured collection of data. The data stored in a database are organized for fast search and retrieval by a computer and therefore a database is anything but a simple collection of items. (p. 39)

Manovich's perspective on databases is worth noting because among his best-known scholarly assertions is the argument he advanced in his 2001 book, *The Language of New Media*, that narrative and database are "natural enemies." Arguing for databases as an emergent cultural form symptomatic of the computer age, he expressed the distinction and their rivalrousness as follows:

As a cultural form, the database represents the world as a list of items, and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world. (p. 225)

Manovich's characterization of the relationship between databases and narratives as rivals responded in part to an exigency: It was a timely polemic set up to advance the status of databases, which became popular and gained cultural footing as a common computational form throughout the 1990s, in contexts of new media production. For Manovich, to create new media is to extend databases, to produce mediated handles on the data sets collected and organized in them.

In a 2007 symposium on databases published in *PMLA*, N. Katherine Hayles reframed Manovich's agonistic metaphor, casting as a more cooperative model narrative and database instead as *natural symbionts*. They depend upon one another, Hayles noted, recalling that "the great strength of database,



of course, is the ability to order vast data arrays and make them available for different kinds of queries” (p. 1604). She accounted as well for the hyperflexibility of relational databases, pointing out many ways relational databases as media objects pluralize monolithic logics for expression. Database generally follows a paratactic logic (shuffling in any direction), whereas narrative generally follows a syntactic logic (linearly sequenced, progressive). If we can get beyond the high contrast in these primary logics, narrative and database have established undercurrents of complementarity and fortuitous interdependence (e.g., data-based narratives and narrativistic data sets).

Having staged this formulation of a mutualistic symbiosis between narrative and database, we can return to any of the disciplinographic accounts authored over the past three decades to inquire into the ratio between narrative and database, to call the question, in effect, of which logic led and which logic followed in any given snapshot of the field. That is, we can shift from the Manovich–Hayles frames and use them as an analytic, a basis for inquiring into tendencies long manifest in RCWS’s accounts. With few exceptions, disciplinographies narrate, and when they have invoked data sets, those data sets have been small scale, situated within the narrative, and adapted as forms of evidence, local and temporary to serve as footings for the narrative arc of the account. Only with few exceptions have they drawn on data sets organized into databases, freely recirculating such that they might be reassembled relationally or influence follow-up accounts. This is not a problem to be solved, but an opportunity for us to acknowledge and pursue. As Grahame Weinren (2007) has written, databases can treat narratives as their operationalized objects, and, to play this out further, we should be able to envision (if not fully develop) a database of localist, ethnographic, narrative-driven disciplinographies.

The Manovich–Hayles analytic is also useful for applying to major databases in the field. Consider the Digital Archive of Literacy Narratives (DALN), a robust collection of video- and audio-recorded narratives of literacy development housed at The Ohio State University. Certainly its contents accord with the definitions of databases presented earlier, and it is also in accord with Weinren’s (2007) two conditions for databases, that “1) it is composed of smaller elements . . . and 2) it can be traversed in a multiplicity of ways” (p. 66). Similar to the previous example, this analytic primes useful inquiries into the design of a disciplinary database, its narrativistic usefulness, but also its capacity for supplying data in tune with distant–thin methods. The scope of this study prohibits a full-blown analysis of this archive, yet the argument becomes more fully apparent through this addition: Disciplinary narratives and the visibility they seek depend vitally on open, searchable, circulable databases.

This limited discussion of databases invites new thinking about their importance to the work of thin and distant methods. That is, my hope is for

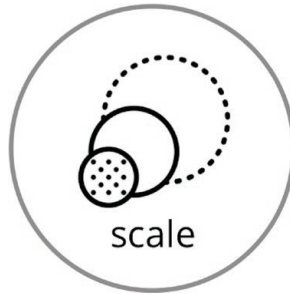
this brief discussion to function as a catalyst for a second move, from the Manovich–Hayles debate as an analytic to the Manovich–Hayles debate as a heuristic to guide and influence future thinking about the interdependence of database and narrative for disciplinography. The visualizations modeled in the book’s later chapters all rely upon datasets that, although they stem from seemingly ordinary disciplinary artifacts (e.g., full-text journal articles, citations, and geolocation), the data they abstract are not readily available, easy to locate, or pushed into wider circulation for expanded adaptations as a result of this work. Without venturing yet more deeply into the limitations of grand databases just as with grand narratives (Lyotard, 1984), and without teasing out distinctions between so-called big data and boutique data—useful and important though these are—the fundamental assumption of the importance of databases and data sets is adequate for the purposes of this book’s primary goal: the advancement of distant/thin methods in service of network sense.

Two final points of value punctuate this case for revitalized considerations of disciplinary databases in RCWS. The first is David Weinberger’s (2007) efforts to draw a correspondence between books and databases as each serves the *externalization of memory* (p. 170). Memory externalization is crucial to disciplinary visibility, and this quality of databases applies from robust efforts to establish records, such as with the Writing Studies Tree and the National Census of Writing, down to the seemingly innocuous but no less important record-keeping related to conference dates and locations, editorial rotations for journals, and geolocations of watershed events for the field—the sum of which are only sporadically, unevenly curated.

The second point of value arises out of speculation about possibilities for something like a disciplinary *database of intentions*, to use John Batelle’s (2005) phrase. Database of intentions is an idea that has shifted considerably toward marketing in recent years, but the basic premise is that we can learn much about what people seek—as individuals or as collectives—by the terms they use when they search. To return to a marketing example, the person who searches for recliner chairs on Amazon is probably interested in recliner chairs. Online advertisers take great interest in this, presenting those who have entered “recliner chairs” as a search string with a deluge of leads—ads and images positioned to entice further pursuit of the recliner chair and possibly to culminate with a purchase.

But the database of intentions, taken to a disciplinary domain, opens onto the invaluable forms of knowledge that would become available if, for instance, we openly shared the search strings entered on CompPile, a collection of scholarly records for work published in RCWS. A high concentration of searches for “assessment,” “archive,” or “queer” would indicate piqued curiosities related to those terms. And the database of intentions also reports on omissions and si-

lences. That is, if three or four years pass without a search for Kenneth Burke or Walter Ong (or any variation on their names), perhaps there is a corresponding insight into an ebb of influence, waning interest or displaced interest, or the relocation of searches to other niche databases altogether.



Having addressed the importance of disciplinary databases for the development of distant and thin methods toward a more fully realized network sense, scale is the next concept impacting these processes. Recalling the heuristic imperative of distant and thin methods, scale emphasizes that these treatments offer much more than measurement, quantification, and the presumed scientific force of data-backed assertions. Scale foregrounds as possible *traversals*, inquiry-movements that shift, whether purposefully or exploratorily, between isolated data-points and larger collectives. In an attempt to deepen the concept of scale for distant reading and thin description, here I account for the concept as it circulates commonly in geography and cartography, disciplinary domains located provisionally through Mark Monmonier's *How To Lie With Maps* (1996) and *Mapping It Out: Expository Cartography for the Humanities and Social Sciences* (1993). Read across these geographic approaches, I suggest that the concept of scale invests distant reading and thin descriptive methods as well as the visual models they generate with adaptive traversals, *scopic* change-ups that allow us—whether as makers (writers) or as interactors (readers)—to get at questions about anomaly and generalizability in the patterned phenomena such methods surface. Where distant reading and thin description methods scale, we elicit questions, as well, about comparison: How do the keyword concordances (or most frequently occurring *n*-grams) of any particular article relate to the keyword distillates of the issue or volume or year? Of the journal during a specific editor's tenure? Of the entire history of the journal? Across multiple journals in RCWS? Scale, in one sense, enables us to customize the aperture of the visual models, to move from a minute, local, and specific magnitude (*micro-*, *nano-*, and *idio-*) to high magnitudes and broadened orders (*mezo-*, *exo-*, and *macro-*) and back again, while exploring resemblances among phenomena at any other magnitude.

As commonly understood, the term *scale* relates both as a verb and a noun to accessibility across transcendable magnitudes. According to the *Oxford English Dictionary*, *scale* derives from the Old French *escaller*, which resembles staircase or ladder. Modern connotations for the verb include *climb*, *get over*, *ascend*, and *to get to or reach the top of*. Where the concept relates to metrology, or measurement, *scale* conjures associations with *proportion*: *To measure or represent (a quantity) in exact proportion to its absolute size or according to an arbitrary defined scale* (“Scale”). These selections offer a preliminary but adequate definitional perimeter—footing sufficient for capturing the most conventional attachments of meaning to the term where research methods are concerned. *Scale* applies to distant reading and thin description methods—and it is even underscored by this elaboration of the concept—because the abstract visual models demonstrated in the next three chapters, if they will be put to a wide range of rhetorical uses, must *do* what existing models have not done.

Consider indexes as another example of *scale* in relationship to distant reading and thin description. In most scholarly monographs, the index is inscribed at a fixed scale. The terms in the index are collected, arranged, and presented at the singular scale of the monograph. Following the conventions of print, indexes tend not to be scalable. That is, if we want to isolate the index for a particular chapter in the monograph, we are out of luck (or, as poor luck would have it, we have new work to do in manually sorting from the comprehensive index only the listed items whose page numbers fall in the specific chapter’s range). The smaller-scale index is hidden, camouflaged among the pile of terms and phrases appearing elsewhere in the book. Distant reading and thin description methods set out to complicate this commonplace treatment, to alter it, and, with this tactical alteration, to generate new catalysts for inquiry. Animated indexes, citation distribution graphs, and maps plotting scholarly activity—each of which will be presented in chapters that follow—are designed with scalability as an imperative, a feature.

For mapmakers and geographers, *scale* is a basic convention and even a commonplace. John Campbell’s (1991) *Introductory Cartography* defined *scale* as “the ratio between measurements taken between points [on the surface of a globe] and measurements taken between equivalent locations on the earth” (p. 24). In a strictly representationalist paradigm, *scale* initiates a referential correspondence between the map and that which it sets out, in more or less detail, to represent. Sticking to Modernist conventions for cartography, Campbell elaborated on the various ways in which *scale* is expressed on a map’s surface—by word statement (i.e., one measure represents another measure), by numerical ratio statement (i.e., 1:10,000), and by graphic *scale* (i.e., a unit-specific graphical element). One advantage of a graphic *scale* is that

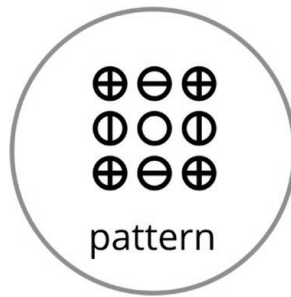
once it has been defined, it adjusts with changes to the overall size of the map due to resizing or photographic reproduction. By contrast, the information graphic in the opening section of this chapter expresses scale by word statement: “Number of times words were used per 20,000 words spoken” (translated to a numerical ratio statement, this would appear as 1:20,000). Ericson’s graphic also introduced an expression about the temporal scale with “20,000 words is about one day’s worth of speeches.”

Mark Monmonier (1996) began his book *How To Lie with Maps* with a discussion of scale that resembles Campbell’s treatment. Monmonier referred to the three scale-types identified as verbal scales, ratio scales, and graphic scales, further emphasizing the adaptive aspects of graphic scales that make them the best choice for digital and photographic reproduction. Further, Monmonier distinguished between scales as *equivalent* (as in equivalency statements, a usage he discouraged) and as *representation*. *Representational scale* is an inclusive term to identify the common three-term typology (verbal, ratio, graphic) applied to conventional cartography. A counterpart to representational scale can be found in one of Monmonier’s (1993) earlier books, *Mapping It Out*, which promoted the use of maps in tandem with expository writing in other disciplines while expressing concern at the limited uptake of mapping practices beyond geography.

While representational scales are primarily invested in establishing correspondences between real material spaces and cartographic representations of such spaces, conceptual scales apply to matters of symbolism, placement, labeling, and sizing of conceptual elements. Explaining a “scale of concepts,” Monmonier (1993) noted, “the map author needs to identify features and relationships to be shown [as well as] their relative importance” (p. 102). Monmonier’s main example is squarely in the domain of geography as he looked at the conceptual choices involved in a map developed by an anthropologist to show multiple routes between two towns. How much should the map depict the hazards impinging on any given route? This is a question of conceptual scale. A more conceptually inclusive map struggles under the burdens of excessive symbolism, which can become confusing. Conceptual inclusion and the principles of selection, reduction, and simplification pull in contrary directions. Their relationship, where distant reading and thin description is concerned, is in constant tension: a critical, or crisis, state. This is why we must engage in distant reading and thin description with an explicit, direct noticing of scale. Scalar choices have bearing on the visual models produced by these methods—particularly where they integrate cartography—as they do for any and all attempts to present spaces and concepts visually.

To convert scale from a technical specification to a practice, it may be useful to revisit Michel de Certeau’s (1988) *flâneur*, the urban pedestrian who

by footfall drifts in the city, practicing “ground level” operations and speaking in elective pathways of intensity, desire, and curiosity (p. 97). Elevated by maneuverability across scales, the flaneur’s corollary in the context of distant and thin methods is the *planeur*, the gliding, aerial, drift-quester whose perspectival perch is configurable. Planeury—as an everyday practice—combines the capacity of screens as digital viewports to quickly switch between the zoomed-in and the zoomed-out, to explore by these variations the specific and the general as interdependent, and to foreground the synechdocal (or part–whole) quality of visualization methods, throughout which scale is a foundational consideration.



Pattern, a condition of comprehensible, repeated occurrence, is the third concept useful for establishing a groundwork for distant reading and thin description in service of network sense. Pattern is as common a reference in mathematics and computation as in art, music, and design. It is an encompassing concept whose wide gulf operates with great variation. For the purposes of supplying a third conceptual foothold for the methods demonstrated by this research, however, note that pattern—as a concept—tends to arise on a continuum between the latent mathesis of all observable phenomenon on one extreme and, at the other, the generative setup on an expanding horizon for switching between sensorial noticing (i.e., forms and rhythms) and predictive speculation: In consideration of some patterned phenomenon, what next?

The first chapter of Mark Taylor’s (2003) *Moment of Complexity*, “From Grid to Network,” is noteworthy for contrasting large-scale patterns and then generalizing those from the architectural urbanscapes where he located them to other phenomena, such as emerging media. This grid-network contrastive tension allows Taylor to account for the ways that network logics productively substitute distinctive patterns (often verging on chaotic, indistinguishable non-patterns) for the modernist constraints associated with grids. To explain this shift, a contemporary turn from one logic to another, Taylor drew on the work of postmodern architect Frank Gehry, whose buildings explored “new frontiers of complexity” without “simply negat[ing] modernism and the world

it represents” (p. 41). According to Taylor, “The grid does not merely disappear but morphs into forms that are dynamic rather than rigid, organic rather than mechanical, complex rather than simple” (p. 41). Taylor noted several connections between Gehry’s expansion of architectural principles that by extension create a clearing that would allow—methodologically—for more dynamic, organic, and complex structures and the data-driven, scalable models produced by distant reading and thin description. Such methods must be considered a response to the complex and maturing interdisciplinary domain of RCWS. Distant reading and thin description advance with the presumption that it is both possible and appropriate to seek, find, and express visually patterns that implicated an ever more complex conglomeration of disciplinary materials and activities—and, furthermore, that these pursuits do not simply negate existing reading practices and deeper—thicker ways of knowing.

Gehry’s architectural response to complexity offered some precedent for thinking about models that integrate rigid, Euclidian shapes and more flexible network qualities. According to Taylor (2003), Gehry often relied on new media and digital interfaces to present his architectural plans: “With the moving images on these mobile surfaces, Gehry seems to achieve the impossible: he simultaneously sets forms in motion and gives movement form. Far from a static structure, Gehry’s building is a *complex* ongoing event” (p. 42). It’s a small but reasonable jump from Gehry’s postmodern, hybrid architecture and the ongoing orchestration of events that constitute an academic discipline. Certainly we can find examples of the discipline as “a *complex* ongoing event.” The ongoing, *live* quality of the discipline has for quite some time been a catalyst for new methods and methodologies. In 1987, Stephen North ended the introduction to *The Making of Knowledge in Composition*, noting “this was to be a new era, and it would demand new kinds of knowledge produced by new kinds of inquiry” (p. 17)—one of many statements about disciplinarity during the period of criticality from about 1986–1990 that came with new graduate programs, new tenure-track appointments, and the publication of theoretical monographs, like North’s and Phelps’s. As I have argued, the late 1980s marked a moment of criticality for RCWS—one that would demand the sharpening of methodological rigor and new devices for apprehending the vast materials of disciplinary interest. But I mention North here not only because his project is widely acknowledged as a turning point from the nascence of the discipline to an era marked by more widespread legitimacy and stability, but, as well, because his project was one that (a) attempted to demarcate the discipline as patterns of research activity and (b) clearly understood the abstract phenomena of disciplinarity as a “*complex* ongoing event.” Twenty years later, this “complex ongoing event” continues to unfold. As such, “new kinds of inquiry” must keep stride.

I am not the first to suggest the complexity of disciplinarity (i.e., of coming to terms with the roiling amalgamation of activities and materials that constitute a field), nor do I hold that distant reading and thin description methods alone will absolve all the infrastructural challenges that accompany rising disciplinary complexity. Distant reading and thin description methods, however, aid attempts to understand this complexity differently by acknowledging that patterns emerge in a wide variety of data that have gone untreated historically. In this sense, the methods, where pattern is subject to deeper consideration, are additive rather than substitutive. However provisional they are, whether long-set or occurring only briefly before shifting again, the patterns germinated by distant reading and thin description methods should be considered in concert with other accounts of the discipline. Specifically, the next three chapters elaborate patterns based on the rising and falling of keywords, shifting citation practices, and mapping program and career path locations.

To assert the importance of pattern at the juncture of disciplinary complexity and architectural infrastructure yet further, recall Steve Johnson's *Emergence*, a 2001 book that examined "the connected lives of ants, brains, cities, and software." Writing about Manchester, England, in its industrial heyday, Johnson differentiated two varieties of complexity, one that resulted from sensory overload (the abundance of a frenzied scene) and another that yielded self-organizing systems (systems in which high-order phenomena are not explicitly decided by a central authority but rather where such rules rise tacitly from below). The discipline of RCWS, perhaps like many fields, coalesces as complex, following both of Johnson's types—as abundant with stimuli and as self-organizing. In response to complexities of both varieties, Johnson noted the degree to which pattern renders durable the emerging infrastructures:

Understood in the most abstract sense, what Engels observed are *patterns* in the urban landscape, visible because they have a repeated structure that distinguishes them from the pure noise you might naturally associate with an unplanned city. They are patterns of human movement and decision-making that have been etched into the texture of city blocks, patterns that are then fed back to the Manchester residents themselves, altering their subsequent decisions . . . . A city is *a kind of pattern-amplifying machine*: its neighborhoods are a way of measuring and expressing the repeated behavior, and sharing that information with the group. Because those patterns are fed back to the community, small shifts in behavior can quickly escalate into larger movements . . . just



a few repeating *patterns* of movement, amplified into larger shapes that last for lifetimes: clusters, slums, neighborhoods. (pp. 40–41, emphasis added)

Johnson folded together dual connotations of pattern, and he read pattern through one densely populated cityscape known well for the industrial boom that accelerated it into a haphazardly planned urban center. Like Taylor, Johnson took an interest in the adequacy of architectural examples to illustrate concepts of complexity and pattern—concepts he elaborated in an effort to explain and explore the focal concept in his book, *Emergence*.

By analogy, many of Taylor and Johnson's assertions generalized favorably to the emergence and maturation of RCWS since 1987, particularly as pertain to the relationship between complexity and pattern. As the discipline has become more complex (some would say, in more derisive terms, *diffuse*), efforts to trace the outstretched lines of RCWS inquiry have been challenged by the depth, breadth, and rate of this expansion. Yet, pattern-amplifying devices are abundant:

- journals (not limited to the original nine studied by Maureen Daly Goggin [2000], but also those newer journals and others whose operations have slowed or halted altogether);
- graduate and undergraduate course descriptions and syllabi;
- monographs and edited collections;
- dissertations and theses;
- conference proceedings;
- textbooks (this one is the patterning device most often studied);
- conference keynote addresses;
- policy statements at all levels; and
- the writing circulated in social media and on listservs.

These constructs are things we have made: pattern-amplifying devices *constitutive of and constituted by* the field of RCWS. Their localizable spheres of circulation are neighborhood-like, offering “way[s] of measuring and expressing the repeated behavior, and sharing that information with the group” (Johnson, 2001, p. 41).

There are examples of studies designed to elicit patterns in some of these areas. We have pocketed studies that have surveyed and sampled nearly every one of these devices, and succeeded—by varied methods—in expressing patterns indicated by a set of materials (i.e., Richard Larson's well-known bibliographies [1988], Robert Connors's [1997] research on textbooks, and, more recently, Susan Peck MacDonald's survey of the Conference on College Composition and Communication [CCCC] program titles [2007]). Still, many of

these samples and surveys work by the singular interval; they are anchored in time—temporally bracketed in a single moment of publication, often produced by an individual or small group as a one-time scholarly project. Distant reading and thin description must proceed more slowly and more inclusively; they must culminate in a series of built *things*—collections and assemblages that aid us in articulating the patterns woven through the field. Their greatest promise for success will depend on a long-term, painstakingly detailed treatment of the abundant materials that can be entered into this realm of consideration—where the *stuff* of the discipline is rendered for patterns that can be traced anew and where the tracing done by others is iterative, ongoing, and in full view. Although it is a modest starting point, this book features distant reading and thin description practices organized around keyword frequencies, citation trends, and locative projections. These are first steps, initial inquiries that, if effective, will prompt us to wonder what other patterns we can know.



Rustling beneath the radar of most conceptual debates about abstraction is the most commonplace example of academic abstracting practices: the one-paragraph summary of a journal article. Ordinarily positioned at the beginning of the article, such abstracts serve to reduce the article into a single paragraph whose purposes are many—efficiency, superficial assessments of relevance, the provocation of memory, and so on. Although individual article abstracts are neither systematically produced nor scalable, their functions are consistent with distant reading and thin description. Reading an abstract is not an exact substitution for reading an article, but, without motivating a storm of critical objection, abstracts are written and circulated to provide an alternative to the article—a double that is selective, reductive, and simpler than the full-length article. What is so satisfying about article abstracts—and perhaps the basic justification for their unquestioned proliferation—is that they do something the article itself cannot.

Abstracts started appearing in *College Composition and Communication* (CCC) in the February 2000 issue, during Marilyn Cooper's term as editor of the journal. Most would agree that abstracts make sense, no matter which of

the uses they are put to, by readers, researchers, and those either prone to forgetting or too busy to read half of an article before deciding whether it might hold something of relevance. Abstracts, even though they are both magical (performing the article's temporary disappearance) and mundane (almost to the point of going unnoticed), set a precedent for the abstracting practices that are broadened through distant reading and thin description methods. With this in mind, the methods advanced here might be understood as a contemporary extension of *abstracting practices*—practices that, much like the speculative instruments they proliferate, make a difference in the ways the discipline is imagined.

In his brief introduction to *Graphs, Maps, Trees*, Moretti (2007) explained at different stages that his methods for visualizing text-based patterns are both abstract *and* concrete:

Finally, these three models [graphs, maps, and trees] are indeed, as the subtitle intimates, abstract. But their consequences are on the other hand extremely concrete: graphs, maps, and trees place the literary field literally in front of our eyes—and show us how little we still know about it . . . Here, the methodology of the book reveals its pragmatic ambition: for me, abstraction is not an end in itself, but a way to widen the domain of the literary historian, and enrich its internal problematic. (p. 2)

Models, on the one hand, are abstract; their consequences, on the other hand, are cast with an intensifier as “extremely concrete.” Moretti's pairing of abstract and concrete as contrastive terms is to be expected considering that they are usually treated as dichotomies. The juxtaposition of these two concepts—abstract and concrete—hearkens back to a rationalistic tradition, in which the abstract is contrasted with an empirically verifiable reality and, as well, with the predominant epistemological realms of reason and logic. Moretti indicated that these terms operate together in distant reading and, as such, they are compatible and that they apply at different stages to the three types of datavisual models his book examines. Yet, because the quotation above is one of just four occasions in *Graphs, Maps, Trees* where “abstract” is mentioned, this passage invites a deeper inquiry into the concept of abstraction. Moretti's references to *abstract* and *concrete* could be viewed as a moment where he backslides from one of the key terms in the book's subtitle, *Abstract Models for a Literary Theory*. In effect, the passage quoted at length above serves to blunt the risk-taking pursuit of abstraction; abstraction in its potential drift away from strictly rationalistic epistemology is hereby attached to a scientific agenda and tendered “as the unproblematic antithesis of the concrete”

(Berthoff, 1986, p. 230). This is not a long-lasting problem for Moretti's project because he qualified the varied processual intervals at which these descriptors apply: "concrete" applies to consequences rather than to the visualizations themselves. However, as one of the few specific references to "abstract," the passage's discussion of distant reading as abstract prompts further consideration of distant reading and thin description as abstracting practices. Distant reading methods involve the selective arrangement of data sets in ways that allow us to apprehend patterns; where data sets are too large and unwieldy, often because such data span large corpora, geographic areas, or periods of time, distant reading supports an interest in the recognition of forms and patterns, which often involve translating data by selection and reduction from nonobvious assortments into observable, suggestive patterns. Considering that distant reading methods proceed in the interest of producing scalable, abstract visual models, what does the reference to "abstract" indicate about distant reading? And why should these visual models be thought abstract?

The abstractive qualities or capacities I want to emphasize in distant reading practices amplify patterns, interconnections, and relationships among selections of data. Recognizing forms and patterns in data fosters network sense; we begin to be able to see those distributed, circulating, and non-obvious formations previously compromised by a lack of field-wide record keeping. For the models presented in Chapters Three through Five, the forms relate to keyword concordances, citation trends, and geographic distributions of scholarly activity. Further, acknowledging a build-up from database, scale, and pattern toward abstracting practices suggests that we might encounter the models as extensions of the collections of discourse and metadata they are fashioned from, thereby regarding them as they produce a *wandering resourcefulness*, similar to a quality I.A. Richards (1994) attributed to words in his work on speculative instruments.

From this, two preliminary responses become clearer as to identifying what is gained from conceiving of distant reading and thin description as abstractive: (a) These practices are compatible with interests in the recognition of form and pattern, and in many cases such forms and patterns are obfuscated amidst complexity until they are presented with varying degrees of selectivity and abridgment; (b) these methods translate collections of disciplinary materials into patterned images, rendering forms recognizable through text mining, data mining, layering, visual design, and presentation. The disciplinary materials subject to distant reading and thin description consequently offer a plentitude of renderings; they are abundant, yielding a profoundly deep, complex resourcefulness that is only partly apprehended by conventional reading practices.

Having thus far deliberated on the abstractive dimension of these meth-

ods, briefly consider the quantitative, empirical dimension—the concrete. The reason for dealing with abstraction is that it supplies counterbalance to any presumption about these methods as aspiring to advance a pure science, that—because they deal with quantification, computational processes, and data visualization—they avoid rhetorical aspects of interpretation and meaning or that they are not especially appropriate for initiating a theoretical curiosity.

When Moretti (2007) contrasted the abstract quality of the models with “extremely concrete” consequences (p. 2), he referred to the potency of such models for the way they confront us with suggestive patterns that cannot be ignored—patterns that demand some judgment as to intelligibility, leaving us to judge them pragmatically “for how they concretely change the way we work” (p. 91). Just as Ericson’s *New York Times* infographic materialized the word frequencies from the national conventions in 2004, so do distant reading and thin description typically feature concentrations of data that demand thinking through their implications. Ericson’s graphic laid bare certain rhetorical strategies (e.g., Democrats didn’t mention Bush as frequently as Republicans mentioned Kerry). These unearthed facts lend themselves to additional lines of inquiry and researchable questions. The models produced by these methods bear out some rhetorical force, especially where they materialize certain aspects of a large data set that have not been presented previously. For Moretti’s (2007) work on literary histories, such models “enrich [the] internal problematic” of that field of study (p. 2). My aim is for this work to extend his assertion and to demonstrate that such methods, too, will enrich the internal problematic of disciplinarity for RCWS.

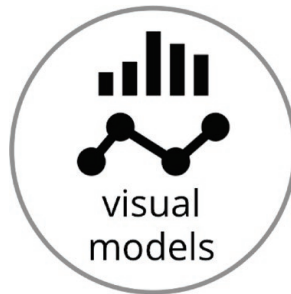
Distant reading and thin description methods and the scalable, abstract visual models developed by such methods integrate scientific and artistic aspects, abstract and concrete aspects, so much so that they are most appropriately described as hybrids. Hybridity, in this case, refers to the combinatory presence of these qualities, commonly argued to be at odds. But hybridity also addresses the inclusive attitude or disposition—the manner of distant reading pursuits—that embraces this combinatory quality while diminishing neither quantitative nor qualitative dimensions. David Foster, in a 1988 *JAC* essay, “What Are We Talking About When We Talk About Composition?” addressed this quandary, arguing for a receptivity to the hybrid epistemology that mutually values these seemingly incompatible methodological orientations in RCWS research:

As informed readers and deliberately inclusive thinkers, we must be the measure of our discipline. Science cannot claim ascendancy in any area of human knowledge, particularly in that complex blend of knowledge-streams we call composition. We must be wary of those who, uncomfortable with

the ambiguities of discourse and complacent with the quantitative, empirical perspective, would have us assume that perspective alone. As informed readers, we must juggle and juxtapose the claims of different modes of inquiry, recognizing what each contributes and what each lacks. To refuse this invitation to an intellectual pluralism, to settle in its place for a single perspective, is to invite the punishment we all hated in grade school: having to write the same sentence one hundred times. In this case, it would be “I will not know. I will not know. I will not know...” (p. 38)

Distant reading and thin description methods operate as abstracting practices; their experimental pursuit is thoroughly rhetorical, even while it embraces the middle ground where science and art converge, ripe with both ambiguities and possibilities, where, as Foster said, “we must juggle and juxtapose the claims of different modes of inquiry” (p. 38).

In the section that follows, I extend the convergence of database, scale, and pattern beyond abstracting practices to examine just how it is these methods function by way of visual models.



Michael Pemberton’s 1993 *CCC* essay, “Modeling Theory and Composing Process Models” renewed basic definitional questions about models—asking and exploring just what models *are* and considering some of the points of terminological confusion and subsequent dismissals based on efforts to eschew models due to presumptions of their mere positivism. Beginning in 1914, with Pierre Duhem, who “criticized models for their failure to be positivistic *enough*” (qtd. in Pemberton, 1993, p. 40), Pemberton recounted the development of Linda Flower and John Hayes’s cognitive process model and effectively summarized the series of strong reactions, including Marilyn Cooper and Michael Holzman’s, that it provoked, many of which “suggest[ed] that the assumptions embraced by cognitivists are far *too* positivistic” (p. 41). The perception that models are too positivistic might persist, but Pemberton’s

work thoughtfully headed off this danger and provided a starting point toward understanding that such criticisms of simple models must not dissuade further development of abstract visual models.

By revisiting selected references to modeling where cognitive science and RCWS converged in the early 1980s, Pemberton (1993) highlighted the terminological ambiguity at play in references to “model.” For instance, Pemberton explained that in their critique of protocol analysis, Cooper and Holzman used “model” and “theory” interchangeably; Flower and Hayes, too, were generous with the functional range of meanings they attributed to the term, as they claimed that their cognitive process model “is both a theory and a distillation of data” (p. 21; qtd. in Pemberton, p. 44). Pemberton deepened the concept of model by setting out from this conundrum—“this elusive and frequently shifting notion of what constitutes a ‘model’ in composition studies” (p. 44)—and by urging a more “careful” approach to models that would “guard against the urge to dismiss, preemptively, the value of a model merely because it contains imperfections” (p. 46). His article continued with an examination of the concept along two lines: models as simplifications, and models as misleading representations.

Pemberton effectively revisited the presumed-to-be clear epistemological weaknesses of models because they are simplifications and because they are (potentially) misleading representations. Where simplification is concerned, Pemberton brought aboard Michael Carter (1990) who wrote that the greatest strength of cognitive studies is that they are deliberately reductive (p. 47). This deliberate reduction occurs with distant reading, which is “not an obstacle, but a *specific form of knowledge*: fewer elements, hence a sharper sense of their overall interconnection” (Moretti, 2007, p. 1). Distant reading and thin description and the visual models produced by these methods are *deliberately reductive*. They reintroduce granularity where it had gone missing; as such, granularity of selected data sets materializes the nonobvious so that patterns might emerge, so that layers from large aggregates of data might be suggestively distinguished and so that connections and associations might be strengthened and amplified—if only temporarily or for the purposes of inquiry. How else might we work through such piles of data but by distant and thin methods and their deliberate reductions? In preparation for distant reading and thin description methods, then, still more work is due to break down the presumed epistemological weaknesses of models. Models, as the work in the next three chapters seeks to make clear, are crucial to distant reading and thin description because they function as an intermediary between dynamic data, which are too laborious to read in long form each time they are updated, and the insights generated from those data when they are presented differentially. Moreover, models help contextualize data; they establish relationships that

bridge form and dynamism; and they are particularly effective at expressing such linkages succinctly and accessibly.

A secondary risk, according to Pemberton, involves misrepresentation. Pemberton (1993) explained this danger as follows: “This risk derives, principally, from the possibility that incidental properties of the illustrative model or preferred analogy may be mistakenly attributed to the process or phenomenon it is being used to explain” (p. 49). Again, models are limited and limiting, a truism that can be asserted about *all* research—whether examined at the scale of data, model, theory, or paradigm; Pemberton reiterated a related point: “incompleteness is an unavoidable epistemological weakness common to all models and all methods of data collection” (p. 53). Pemberton’s work on models deserves such an extensive recap because his account offers a thorough and nuanced treatment of modeling theory stationed squarely in the wheelhouse of RCWS. Perhaps Pemberton’s most pertinent contribution is in his identification of models as “*partial isomorphs* of their subjects” (p. 45). As partial isomorphs, models “will be similar in form to their subjects but embody fewer of their subjects’ constituent properties,” and, consequently, “it will always be possible to find certain properties in the subject that the model will not account for” (p. 45). His acknowledgement of this quality is consistent with the general principles grounding this entire section: Models simplify and reduce, and in so doing, they also amplify, granularize, and strengthen highly suggestive associations that become traceable.

The definitional depth Pemberton contributed to this discussion of models is not intended to curtail or reign in the array of models pertinent to this research project or to the field more generally; instead, it is offered in the spirit of honing the potential of models for materializing nonobvious patterns in the disciplinary data sets considered in the following chapters. Even though he focused on Flower and Hayes’s cognitive process model, Pemberton (1993) accounted for models in general; he was not partial to visual models. In fact, he downplayed the presumption that models should be thought, by default, as visual when he wrote, in a parenthetical aside, that “although [Flower and Hayes’s cognitive process model] has a graphic presentation, a diagram is not a requisite for something to be considered a model. A model may be presented, for example, in purely textual form” (p. 49). While this inclusive gesture is appropriate to his theorization of models, it leaves us with a momentary imbalance given that I am primarily concerned here with deepening the concept of visual models. We must now consider the adjectival term in the phrase by turning to theoretical expansion of the visual in visual models. For this, Johanna Drucker’s work on graphesis will help us further align distant reading and thin description methods and visual epistemology.

Drucker’s primary aim was to present graphesis as a hybridized middle



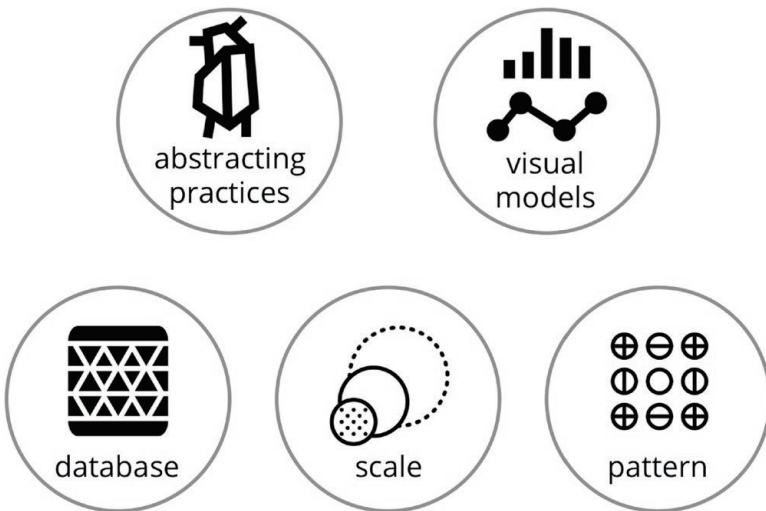
domain that appropriates the felicitous influences of the sciences and the humanities on visual epistemology and that, in so doing, achieves a thoroughly rhetorical understanding of visual knowledge production. Drucker (2010) stated this case plainly: “graphical structures are rhetorical arguments” (p. 17). In itself, an assertion like this is not surprising to those who have been thinking about and writing about design, visuality, or the rhetorics of art, comics, computer interfaces, or photography for some time. Drucker acknowledged the extensive precedents for graphesis; she conceived graphesis to be “profoundly interdisciplinary” (p. 4) and she went on to account for some of the specific ways visual epistemology is produced and circulated. Drawing on scholars in the sciences as well as the humanities, Drucker also revisited commonplaces about visual epistemology held both by scientists and by humanists, with precedents that predate print technologies: “Even before the existence of print technology, visual images served varied epistemological functions—from the representation of information in condensed, legible form, to the expression of complex states of mind and experience” (p. 2). Her presentation of graphesis as a theoretical, rhetorical intervention into commonplace thinking about visual epistemology is conducive to distant reading and thin description methods and the visual models they produce. As you will see in the following chapters, abstract visual models produced by these methods, understood in accordance with graphesis, must “be conceived as procedural, generative, emergent, as a co-dependent dynamic in which subjectivity and objectivity are related” (Drucker, 2010, p. 4). Further, Drucker explained that graphesis depends upon opening up and expanding the concept of epistemology—of what knowing involves, of what can be known, and of what coalesces where knowledge is claimed:

We have to go beyond thinking of knowledge in terms of mechanistic and static relations in which things known and things shown are assumed to be independent entities operating in an objective universe of phenomena existing in advance of their apperception. Visual epistemology is based on a more radical theory of knowledge. The radical concept of subjectivity, and of the co-dependent nature of knowledge and interpretation, have been integral to quantum physics for nearly a century and to cognitive studies for half that long. Graphesis takes these concepts as foundational and uses them to construct a theory of knowledge through attention to the graphical form of its many expressions. (p. 34)

To begin drawing distinctions between the “many expressions” of interest to graphesis, Drucker offered a loose typology: “They can work 1) through

offering a visual analogy or morphological resemblance, 2) through providing a visual image of non-visible phenomena, or 3) by providing visual conventions to structure operations or procedures” (p. 5). Although these three types are most evident in the quantitative research found in the sciences, Drucker elaborated on each in such a way that makes the typology more broadly applicable for graphesis and its “profoundly interdisciplinary” reach. Drucker detailed these distinctions so that she could break down model-types, introducing an analytical scheme to more acutely historicize the precedents for graphesis before reuniting the three strands of model-types, as her discussion would “rest on an assertion that visual epistemology must be synthesized at the intersection of humanistic and scientific concepts of knowledge” (p. 10). This intersection, I argue, must be thought of as thoroughly rhetorical. For visual epistemology and the visual models proliferated to date in RCWS, we must reconsider not only how they have adapted and evolved, but we must also understand the models rhetorically—not as aging historical statements, but as performative figures perpetually animated and ongoing, as figures that *move*, compelling assent and changing the ways the discipline is conceptualized, imagined, and enacted, so that its *future* work can be carried out.

## Visualizing a Discipline



Up to this point, this chapter has reviewed selected concepts relevant to an emerging methodology designed to visualize patterns in the emergence and maturation of RCWS. Distant reading and thin description intervene into a disciplinary problematic—one grounded in matters of highly irregular data gathering, an abundance of reading materials piling up, and predominantly

thick and narrativistic bases for claim-making about disciplinary formation—with datavisual models generated to bolster a sense of the field as an interconnected and ongoing phenomenon. Providing this conceptual groundwork, which emphasizes databases, scale, pattern, abstracting practices, and visual modeling, brings this methodology more fully into view as a generative epistemic technology whose thin, distant qualities offer handles for inquiring into complex, distributed disciplinary activities and materials: for developing a *network sense* of the field.

To recap, the chapter advances the following key principles for distant reading and thin description:

1. Distant reading and thin description *mediate* between collections of data and abstract phenomena that can be difficult to identify, such as disciplinarity.
2. Distant reading and thin description apply self-consciously at particular *scales*. They are applied at various magnitudes of measurement (from the small to the large) but flatten out complex phenomena so that we can materialize traceable networks of associations. This associative capacity makes it possible to travel between the selections of data and the complex phenomenon under investigation.
3. These methods are *visually rhetorical* in that the visual models they produce articulate potentially patterned images that function as arguments, influencing ways of thinking, and widening the perspectives available as they relate to a phenomenon, such as disciplinary emergence and maturation.
4. The visual models produced by these methods stand as *partial isomorphs* of the phenomena they depict. They induce a known degree of reduction and selection, though it is often temporary in the sense that the models are constantly reintegrated into the complexity that they are designed to help us apprehend. In this sense, the qualities of reduction and selection are, paradoxically, also expansive. By rendering a more granular, selected set of data, the selection is, by its reduction, amplified.
5. Distant reading and thin description are generative, convening a *heuretics*, in Ulmer's theorizing of the term. That is, they are above all *inventive* and highly *suggestive*. Resisting attachment to positivism or scientific proof, they function as *speculative instruments* that promote inquiry into *theoretical curiosities*; they are not wholly invested in interpretation of texts, nor are they constrained by a strictly representationalist paradigm.
6. Distant reading methods involve *visual presentation* and almost always

pair the visual models with discursive accounts that advance the matter under consideration by explanation and analysis. In this sense, the mode of presentation is *hybrid*.

This set of principles does not exhaust the full range concepts touched upon in this chapter, nor does it reduce to summary all the applications for distant reading and thin description imagined—and even enacted—by Moretti, Love, and also by Matthew Jockers (2013), whose monograph *Macro-analysis* applied distant reading and data visualization to large-scale analyses of literary texts. Certainly there remains still more conceptual groundwork to be refined if datavisual methods are to take hold and have a lasting impact on the ways we seek to understand such complex, abstract phenomena as an academic discipline. But the purpose of this chapter has remained to articulate salient concepts and to present them as a basis before proceeding with fuller demonstrations in Chapters Three through Five. It should be clear by now that visualizing a discipline involves a set of practices suited to a rapidly expanding range of applications—well beyond what have been, thus far, uses aligned with literary analysis and interpretive–hermeneutic ends. With the conceptual groundwork primed, we can now focus on the animated index, citation graphs, and maps of scholarly activity, thereby moving ahead with the development of patterned images of the discipline of rhetoric and composition/writing studies.