The Queer Gap in Cultural Analytics

Kent K. Chang

Forthcoming in *Debates in Digital Humanities 2023*, edited by Matthew K. Gold and Lauren Klein. University of Minnesota Press, 2023.

In the past few years, digital humanities researchers working with texts have demonstrated the efficacy of their methods in answering historical questions pertinent to literature and culture, often with substantial engagement in both theory and computation: through quantitative methods, we see how fictional genres consolidate (Underwood, *Distant Horizons* 34–67), how expressions of gender mutate (Cheng), and how systemic oppression has materialized in the landscape of literary publication (So and Roland). Such research falls under the rubric of cultural analytics, the intersection of cultural studies and data science, which has gained prominence in digital humanities in the past few years. While this work has become increasingly attentive to questions that arise from issues of social difference, there remains a gap in the scarce number of attempts to incorporate computational methods for studying queer cultures.

Studying queer cultures through computational methods, or queer cultural analytics, can be valuable in several ways. First, as we will see below, queer theory, and queer studies more generally, affords cultural analytics (CA) a repertoire of new inspirations. Second, queer CA can engender further debates and justifications for claims about the queerness of datasets. While critics of data science have rightly observed that sociotechnical systems powered by machine learning routinely reinforce bias, oppression, and injustice, queer CA has the potential to further mitigate the implicit marginalization from the origin by taking into account the historical and cultural contingencies that made it possible to *queer*. Third, fields neighboring digital humanities that have stronger connections to computer science, such as computational social science and human-computer interaction, exhibit an increasing interest in incorporating humanistic scholarship to better ground their research design and account for their findings. Queer CA has the potential to model and further advocate for a more meaningful intervention in technical disciplines.

But queer CA must first exist. To address the present queer gap in cultural analytics, I would like to propose a research agenda that belongs equally to queer studies and cultural analytics. I envision a computational study of queer cultures that can enrich our understanding of queer cultural histories and memories.

My coinage of "queer cultural analytics"—as opposed to "queer computational humanities" or "queer quantitative literary studies"—seeks to foreground the central role of queer culture in this line of inquiry and queer cultural artifacts as its primary object of study, as opposed to the (computational or quantitative) methodology deployed.[1] Queer cultural analytics can be understood as implicitly rejecting any opposition between quantitative and qualitative work. It exhibits a multivalent inbetween-ness: in terms of subject matter, it works in the intersection of cultural studies and data science (that, again, intersects with machine learning, human-computer interaction, and natural language processing); in terms of research output, it results in argument-driven essays that pursue humanistic questions while maintaining scientific rigor. Indeed, queer CA is in and of itself is a multipart and multivalent process. In what follows, I lay out a conceptual model that describes this process in terms of six concentric circles: question formulation, dataset construction, conceptualization, operationalization, interpretation, and presentation. By iterating through these circles of inquiry, it is possible to achieve a queer CA that allows us to investigate empirical questions pertaining to queer cultures.

THE POSTSTRUCTURAL TROUBLE

For those who are interested in queer theory but remain skeptical of cultural analytics, it's counterproductive to dwell on CA's use of categories (e.g., male/female) for the sake of computation. The conflation between the *misuse* of categories and categories per se is a reductive characterization of how most CA practitioners understand the role of categories in their work.[2] It is an indisputable fact that historically, data, statistics—and more broadly, technology as a means for categorization—have been used or associated with the marginalization and oppression of queer people.[3] However, that should not be taken to mean that technology or categorization is inherently reductive. Here we might draw a lesson from queer theory and the poststructural thinking that gave rise to it: at face value, neither identity structure nor structuralism was inherently bad. What is problematic, as explicitly announced in key texts such as Judith Butler's *Gender Trouble* and Jacques Derrida's *Of Grammatology*, are the phenomena of phallogocentrism (Butler) and logocentrism (Derrida). These concepts, which describe an implicit gesture of privileging one thing over the other (e.g., male over female, speech over writing), are the result of what some tired binary oppositions have given rise to.[4] Gender binarism limits our thinking about gender, for example, hence the need for deconstruction. One might recall Butler's famous remark: "Since I was sixteen, being a lesbian is what I've been," which demonstrates how the act of subversion, not the denouncement of category, is what frees our thinking ("Imitation and Gender Insubordination" 311). In articulating her performative ontology, Butler not so much eliminates the identity category of

1. I am aware that "cultural analytics" is associated with specific researchers (Andrew Piper and Lev Manovich) and journal (*Journal of Cultural Analytics*), but my nomenclature serves to reiterate the research agenda, and I do not intend to exclude other computational work within DH and the humanities.

2. For a relevant discussion on categories and computational methods in the context of race, see So and Roland (esp. 62–64).

3. See Gaboury on the political implications of binarism in technology, Spade and Rohlfs for examples of an unjust use of statistics, and Bianco for a reflection on the notion of tools in DH.

4. Or, Saussure's model of differential system; see Descombes 75–109.

lesbian but problematizes its presumed stability: one cannot simply *be* lesbian; one can only *be being* lesbian. In short, the root of the problem is not the idea of the binary itself, but what the binary is used for.

Machine learning must be similarly understood—and interrogated—not because of its reliance on categorization per se, but how those categories are used. Many current machine learning techniques involve classifying patterns on the basis of signals gleaned over the so-called "learning" process. Some classifications can be useful and enhance our understanding of the literary past or of identity categories.[5] Some, on the other hand, are the results of algorithmic bias and can have profound implications on individuals and groups, often discussed in terms of allocational harm and demographic bias (Hovy).[6] While both rely on the same technique, the two lines of work have a different nature and different stakes: the former probes into our understanding of genre, which is a literary-historical/literary-critical inquiry. When done inadequately, it can produce stereotypes—the "bad science," and "bad descriptions" that Laura Mandell speaks of (4). The latter tackles problems pertinent to decision theory writ large, with often immediate consequences on actual lives. While both cases can have profound impact, we should not conflate these two distinct lines of inquiry and uncritically reject algorithms and empirical inquiries altogether.

We must remember that numbers have to be read and interpreted, and this process of intellectual mediation allows us to address limitations of our data and methods. At this point in the analytical process, theory is quite useful. I will argue, however, that we should also have theory in mind *before we start*, especially theory that has helped us articulate and engage queer people and queer culture. Mandell has argued with respect to cultural analytics that any stable classification of gender is fundamentally anti-Bulterian and is compatible with neither queer theory nor poststructuralism. The problem with such work, she argues, in addition to its uncritical embrace of stable gender categories, is that it stays within the comfortable confines of formalism that easily lends itself to large-scale analysis. Even in the projects that Mandell cites as exemplary, theoretical explorations are most often mentioned in the concluding discussions as "future work." In line with Mandell, I want to emphasize that this view of the role of theory only *after* computation limits both our understanding of gender in the cultural past and our imagination of what can be done in CA. This might be where we are, but we can go further.

Driven in part by the generative potential of theory, queer CA aspires to be people- and cultural-centric. Ideally, this would mitigate the fiction of a disciplinary divide. Discussion on this topic has been disproportionately focused on disciplinary practices and methods: the efficacy or limitations of close reading has been the locus of several debates in literary studies, for instance, the *Critical Inquiry* forum following Nan Z. Da's "Computational Case against Computational Literary Studies" is a notable example. Since then, we have seen computational work in literary and cultural studies presented as an "augmented humanities" (Algee-Hewitt) or a "computational hermeneutics" (Underwood, "Machine Learning and Human Perspective"),

5. See Brown and Mandell for a discussion on this in CA.
6. Notable examples include Buolamwini and Gebru; for more background, see Noble.

as well as arguments that the discipline, naturally pluralistic, should have both (Kramnick). These discussions, while having helped find a better footing for quantitative methods in a traditionally qualitative discipline, have neglected an important aspect of what binds together these forms of qualitative and quantitative research. For both, the ultimate goal is to learn more about culture. In the case of learning more about queer culture, any method—close reading, digital programming, or a combination of both—may have the potential to reach this goal. An adherence to false binaries—"computation OR reading; numbers OR words; statistics OR critical thinking" (Algee-Hewitt, "Criticism, Augmented")—would shut down queer CA before it could even start.

Queer CA could queer divides not only within disciplines but between them. For example, technical fields such as machine learning and natural language processing (NLP) now have the tendency to incorporate research done in traditionally non-technical fields.[7] Due to the rigidity and constraints of the genre of the conference proceeding, however, such fields often struggle to engage humanistic scholarship in a substantial and elaborated manner. Put another way, researchers in those spaces are required to write about algorithms and data, but not necessarily the societies and cultures from which they emerge.[8] In contrast, humanistic essays that touch upon queer cultures, as discussed above, are agnostic if not hostile to algorithms and statistics. Queer CA encourages researchers across multiple disciplines to locate the common ground between humanistic and quantitative traditions. It encourages an embrace and understanding of statistical patterns and technical classifications together with the multiplicity and complexity afforded by humanistic approaches. The model proposed below offers one possible way in which this might be achieved.

HOW TO DO QUEER CULTURAL ANALYTICS

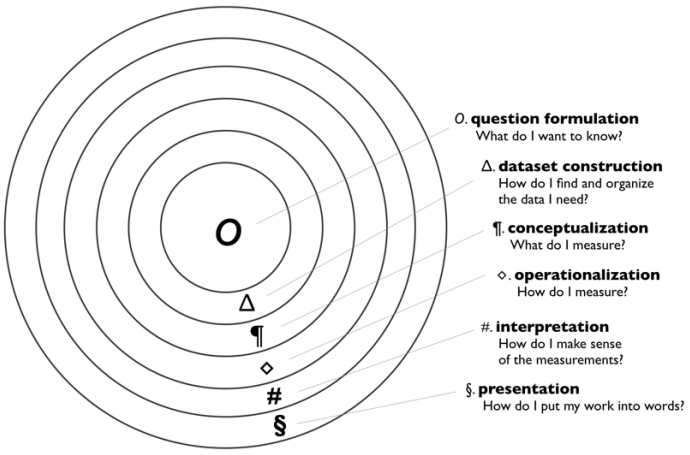

Figure 7.1. Concentric circles for queer CA. My choice of the symbols for each circle is entirely arbitrary, and this is mainly to avoid numbering each in sequence. In doing so, I wish to highlight the nonlinear nature of CA research: we always move back and forth between those circles.

7. See e.g. Wallach and Connolly.
8. See Geoff Hinton's interview with *Wired* for a related discussion.

In order to advance the prospect of queer CA, I propose a model of concentric circles of inquiry (figure 7.1). The metaphor of concentric circles aims to complicate existing accounts of methodologies in CA (and adjacent fields) and bring its sociotechnical nature to the foreground[9] My use of random symbols in lieu of the conventional numbers or bullet points is deliberate because I want to emphasize that queer CA is never linear (as entailed by the choice of a sequential numbers) and that not every stage can be presumed to be equal (as entailed by bullet points on the same level). In terms of research process, queer CA is inevitably iterative:[10] you can rarely build a dataset that will answer your research question from the start, your operationalization of your research question may lead to results that are entirely uninterpretable, or your presentation of results may fail to engage the audience that is interested in or influenced by your project. Any number of such things can and will happen, and you have to go back to earlier stages. Everyone working in CA has likely experienced similar frustrations, and I feel obliged to make this explicit in my framework for queer CA: you never just go from step 1 to step 2 to step 3, resulting in a paper at the final step. The process can be messy, and the eventual paper, if there is one at all, will not necessarily represent such messiness. This nonlinearity also characterizes the larger process of knowledge production that I envision for queer CA: it is situated, neither dualistic nor futuristic,[11] and no one can play, in Donna Haraway's words, "the god trick of seeing everything from nowhere" (Haraway 581). Queer CA does not actively or purposefully work toward discovering a singular truth about the queer cultural past from an all-knowing, subjugating vantage point. This explicit messiness, represented by a random set of symbols, aims to metaphorically counter the fetish of rationality and objectivity, which feminist approaches to data culture have taught us to approach with skepticism in the past decades.

With all that said, there remains an enclosure relationship between those iterative stages of queer CA, and it would be impossible to conceive of them as entirely nonsequential: if you fail to construct the dataset, for instance, you are unlikely to reach the presentation stage at all. At the same time, in writing about the project, one still needs to establish a logical flow to stay coherent and avoid nonsequiturs. To adequately capture this, I imagine those stages as concentric circles, inspired by Andrew Piper's nested layers for literary modeling and Gadamer's notion of hermeneutic circles (Piper; Gadamer). Queer CA, however, is not ultimately about modeling. The origin of the circles constitutes the origin of the project—we want to learn something about queer cultures. We start from there, and we will move between circles as we discover what the data—the partial representation at hand—can teach us about the unbounded totality of queer cultures. To this end, I advance three theses for *doing* queer CA.

9. See Nguyen et al. for an example.
10. On the iterative nature of modeling in CA, see So, "All Models Are Wrong."
11. For more context on the notion of futurity in data culture, see Zeffiro.

Thesis 1: Queer CA is to be driven by questions pertinent to queer cultures, not by existing archives and algorithms.

At the center of the circles (*O*) is the intellectual impulse driving the queer CA project: what is it exactly that we want to learn about the queer cultural past, which will inform our research questions? More specifically: what are the questions pertinent to queer cultures that will require us to leverage the interpretative possibilities enabled by quantitative evidence? This stage is at the innermost of the concentric circle, which means that no matter how far a scholar is in the project, they have to constantly ask themselves if they deviate from the impulse that originally drove the research project. Queer CA should be queer-culture-centric and inquiry-driven, as opposed to simply *data*-driven.

We can begin by considering the common grounds between existing CA and queer theory projects, which would then require us to further discuss the poststructural trouble described above: first, the data-scientific approaches undergirded by CA can seem incompatible with the ethics of knowledge production in queer and feminist approaches to data and the archive; second, queer theory has been anti-identarian and anti-empiricism since its inception, whereas CA routinely makes use of identity categories as the organizing principle for its data as well as metadata (that is, data describing data). This tension propels us to conceive of a *queer CA archive* that must serve the two goals that are, on the surface, conflicting:[12] a queer *CA* (with a stress on CA) archive would be an ontologically stable whole that can be used to derive reproducible quantitative evidence, which would in turn teach us about queer cultures; but, at the same time, a *queer* CA archive (with a stress on queer) has to be situated, contingent, and affective. Only when at least those two goals are met can researchers speak of their *queer dataset* as part of the larger archive and articulate its queerness.

This is no easy task, although I believe a queer CA archive will have the following defining features: it will be cultural-centric. Being cultural-centric means the archive will speak to queer cultural memory and history,[13] which can be validated with theory and relevant scholarship in queer studies. More simply put, a queer CA archive will offer an adequate representation of cultural artifacts that are integral to the collective cultural past of queer people. This jives with what David Halperin imagined, back in 1991, as the "cultural poetics of desire," which he defines as "the processes whereby sexual desires are constructed, mass-produced, and distributed among the various members of human living-groups" (40). If one were to pursue this in queer CA, scholarship on queer cultural sensibility could offer theoretical resources. Studying gay male sensibility, for instance, means studying, in Jack Babuscio's words, "a creative energy . . . a

12. In the context of queer CA, as a first step, we can think of an *archive* as generalized body of work pertinent to queer cultures. But in the context of queer studies, *archive* can mean "a theory of cultural relevance, a construction of collective memory, and a complex record of queer activity" (Halberstam*, In a Queer Time and Place* 169). For a discussion on the notion and use of *archive* in queer studies, see Anjali Arondekar, Ann Cvetkovich, Christina B. Hanhardt, Regina Kunzel, Tavia Nyong'o, Juana María Rodríguez, Susan Stryker, Daniel Marshall, Kevin P. Murphy, and Zeb Tortorici.

13. See Doan for an articulation of the distinction between memory and history in queer studies.

perception of the world which is colored, shaped, directed, and defined by the fact of one's gayness. . . . the nature of the specific set of circumstances in which, historically, we have found ourselves" (19). In more concrete terms, creators of sensibility, ranging from Oscar Wilde to Idina Menzel, are often "figures of identification" (Clum 168); through them, numerous *I*'s realize their queerness and become part of the *they*. Instead of asking simply what is gay or what is gay about X, ask what makes X gay, and how do we come to know it. In other words, instead of dwelling on the ontology of queerness, queer CA can probe into the epistemology of queer sensibility.

In this formulation, queer is neither a category nor an anti-identity construct; it instead signifies the resonance of a shared culture, its history and memory. This involves a subtle shift from queer subjectivity to queer cultural sensibility: the constitution of a queer subject, according to queer theory,[14] is largely understood as a first-person phenomenon; how to study this in CA does not seem clear or even necessary. When focused instead on cultural sensibility, researchers can study the collective aesthetic experience, and it would make more sense to look for empirical patterns that are computationally identifiable. Importantly, queerness can then be said to have been defined *a priori* because it is grounded in existing studies. For example, Jack Halberstam, among others, names an "excessively small archive" that includes "Tennessee Williams, Virginia Woolf, Bette Midler, Andy Warhol, Henry James, Jean Genet, Broadway musicals" (Halberstam, *Queer Art of Failure* 109)—the list goes on, and a queer CA project could potentially start there and consider how to, for example , arrange and transform those the works of these authors and their associated metadata in a machine-readable format, and at the same time, ensure that the cultural resonance of the authors isn't diminished in the digital sphere; describe the stylistic or thematic change has taken place in Broadway musicals since Stonewall; juxtapose similar cultural archives from different national or regional traditions to shed light on them individually—just to name a few possibilities.

When queerness is not defined solely as a first-person phenomenon but as that which encapsulates a cultural sensibility, other prospects may arise for queer CA. A notable one might be to enable different computational approaches to intersectionality that may at first seem counter-intuitive. Intersectionality has been identified as important yet challenging in CA;[15] when one looks to NLP and HCI research, work on intersectionality tends to work with fixed identity categories informed by metadata or inferred through linguistic features (e.g., use of a select list of key words, of pronouns) in the data.[16] This box-ticking approach, where an author or a data point is treated as part of a pre-defined identity group by matching associated criteria in the research design, poses a greater challenge for queer CA: since queer is by definition contingent and discursive, researchers would have to articulate the possibility for, say, queering the classification boundaries in machine learning, and queer theory is not necessarily amenable to such an undertaking. In contrast, researchers in queer CA can leverage scholarly work on,

14. For a background on prominent approaches to queer subjectivity, see Ruti 13–43.
15. See the concluding section in Kraicer and Piper, for instance.
16. See e.g. Schlesinger et al. and Jiang and Fellbaum.

say, queer African-American archives to guide them as they build their datasets[17]—theory *before*, not after, computation.

The above discussion demonstrates that, in fact, the dataset construction stage (Δ), is so close to *O*, the question formulation stage, that researchers may find themselves constantly oscillating between them, especially at the initial stages of a research project. It is also important to keep in mind that as we construct our dataset, or corpus, we must address our selection bias. The problems of representativeness that Katherine Bode has identified (in terms of "ontological gap", 97) can be far worse in a queer CA. For example, on HathiTrust, one can find plenty of works by Oscar Wilde and some by William Inge, but no sight of work by Maria Irene Fornes or Tony Kushner, which can potentially exclude research questions pertinent to BIPOC queer cultures. More archival, activist, and digitization work is necessary, although that is beyond the scope of this essay. In addition, metadata requires more attention. Queer cultural texts have often been subject to varying degrees of censorship, rendering an adequate understanding of publication history paramount.[18] Even canonical authors like Wilde may require careful handling: the text of the uncensored edition of his *Dorian Gray*, for instance, is significantly different from the original 1890 magazine edition,[19] so it is inadequate to include merely author name and title in the metadata for a project that involves text mining Wilde's work. With that said, how long one should stay in this circle of dataset construction, working to increase representation and refine metadata, will eventually depend on individual projects. This is why I insist on a concentricity model: one can always move on to other circles, and then decide how well the dataset addresses the original question.

Thesis 2: Queer CA is a quest for effective proxies for operationalization, which necessitates competence in both quantitative and qualitative traditions

Conceptualization (circle ¶) is often described as a process of translation between abstract theoretical terms and concrete conceptual anchors (Piper 653), which may not be easily applicable in queer CA; *queerness* is defined not by what it is, but by what it is not, and in this light, even measuring it can seem odd. It may be more useful to think of conceptualizing as finding proxies: ideal proxies are measurable concepts that help us answer our research question. In the operationalization circle (◊), we develop measures for those concepts for computation. To understand how conceptualization and operationalization work together at a higher level, consider this example from Jennifer Quist: the researcher wants to know the political and cultural biases behind the Nobel Prize for Literature (*research question*) and the proxy is the profile of an ideal prizewinner (*concept*). Then, she operationalizes the concept by developing variables, each a feature of laureates, derived from her reading of the official statements from the Swedish Academy; when individual laureates have one of those six features, the laureate earns one point, and the total score shows how "ideal" this particular prizewinner is (Quist). Through this example, this process can be understood as figuring what to count (conceptualization ¶)

17. Migraine-George and Currier, for example.
18. For a recent discussion on the sociological implications of censorship, see Lubin.
19. See Wilde and Frankel.

and then how to count (operationalization ◊). Because of its inherent focus on queer culture, however, queer CA involves working with non-trivial constraints in this process. Suppose we want to build a dataset from the cultural archive Halberstam describes; the possible cultural texts that exhibit queer sensibility are already limited, and the actual texts that are available in the digital format can only be more scarce. The implications are significant: if machine learning is to be adopted, there might simply not be enough training data. The challenge for queer CA, then, is to find different kind of patterns and operationalize creatively, and acknowledge and account for necessary compromises.

It is crucial to remember that if we want to explicate queer cultural patterns, machine learning (or other popular methods in DH) are not our only options; queer CA doesn't have to be the sum of queer topic models and word embeddings. When used well, merely counting a carefully selected words or phrases (n-grams) can answer interesting questions. For example, in her work on the art-historical concept of the medium, Anna Shechtman looks into the frequency of pertinent terms such as "oil on canvas" and "mixed media" in the pages of *ARTNews* from 1950 to 2000 (Shechtman). In queer CA, figuring out the appropriate terms in an analogous experiment would require us to look back to the history of queer studies. Its lessons could inform the interpretation of, say, historical trends of most frequent collocations visualized through computational methods. If the methods are more complex and involve statistical inference, a potentially useful framework for queer CA might be to examine the sociotechnical implications of predictive models in terms of contestability, or the ability to contest algorithmic decisions (Kluttz et al.). Contestability allows for critical engagement, which is of particular importance for this kind of research that is people- and culture-centric. This applies to both the humanistic and the technical sides of a project: if your machine learning algorithms create models that make any prediction integral to your argument (say, in this given passage of a novel, you have a model that predicts that the queer character experiences a lack of agency),[20] can you tease out why that is the case and explain, if at all, how you can counter the prediction? Can you identify the potential gaps in existing theories, if any, and show what in the findings seems at odds with them? When the focus of the project is to learn more about a certain aspect of queer cultures, both the computational methods and the theoretical bases engaged should be contestable.

Thesis 3: queer CA should constantly queer CA.

The key potential of queer CA materializes in the remaining circles: to queer its own discipline with adequate reasoning and justifications. The interpretation circle (#) can be the place to raise such questions as: what counts as scholarship and evidence in queer CA? In what ways can or should queer CA deviate from regular norms in research practices? Just as Queer OS welcomes crashes (Barnett et al. 54) and good describers embrace "stray details" (Marcus et al. 11), ideally, queer CA will not shy away from either obscure or contestable patterns. Queer CA need not have to be solely about successful experiments.

20. This speculative example to explain what I mean by *prediction* is inspired by Sap.

The last circle, presentation (§), is concerned with the communication of the project as a whole. Here, difficult might arise in the presentation of the technical details (like algorithms) as they relate to the central cultural inquiry. Benjamin Schmidt has argued that digital humanists should learn about the transformations that any algorithms they use can bring about, since many algorithms are in and of themselves irrelevant (Schmidt). Indeed, learning what happens under the hood after you apply the *sort()* method to a list in Python, for instance, doesn't make you a better digital humanist. But since transformations happen at the level of data, or texts, this general line of reasoning implies that it is data that sits at the center of the project, which is certainly not the case for my formulation of queer CA. Queer CA is not *just* data-driven; the human problem—not texts or any sort of data for that matter—occupies the common center of my concentric circles.

In the context of queer CA, I argue that in communicating the technical details of their projects, practitioners of queer CA should focus instead on the heuristics, or the intuitive processes, behind the algorithms they adopt, not merely on texts and their transformations: what task does the algorithm seek to complete, and what steps are taken to complete it? To use Word2Vec, for instance, then requires a basic understanding of relevant language models (skip-gram, continuous bag of words) and distributed representation, and what kind of problems in the NLP tradition those new models were designed to solve. And since those are predictive models, an intuitive understanding of typical training methods (negative sampling or hierarchical softmax), is essential. Researchers in queer CA don't necessarily need to know exactly how stochastic gradient descent processes update vectors, but some understanding about distributed representation and its relevant history in NLP, or how it relates to cosine distance, would nevertheless help them articulate how this operationalization method speaks to their research question. Grasping the thought process behind such problem-solving algorithms reflects more than an attempt to reduce the skepticism surrounding black boxes. Without an intuition of what happens in the algorithm, researchers can struggle to move between those circles of queer CA. It's true that you can grab the gensim port of Google's Word2Vec toolkit,[21] read through a set of tutorials, load pretrained models, and play around with word embeddings. If you are lucky, you might have something interesting to say. But how does Word2Vec help them operationalize, say, semantic difference in their own project? How do they ensure the operationalization is stable and yields robust patterns? If one believes computational methods enhance interpretative practices, it is responsible to have some idea of what is actually happening under the hood. This is ethical, as it is practical. Algorithms, as well as their limitations, should empower—not restrict—researchers to move freely between those concentric circles.

It can seem challenging to articulate models and algorithms with regard to queer cultures because, as with finding proxies (thesis 2), it calls for rather specific technical and humanistic sophistication. Here, I echo Ted Underwood's call for a broader institutional change (Underwood, *Distant Horizons* 161–65): like distant reading, the training required for queer CA does not fit easily into existing curricular and department structures. Indeed, the first

21. See https://radimrehurek.com/gensim/models/word2vec.html.

radical change that queer CA might aim for may be a pedagogical, not literary-critical, one. Imagine in the most ambitious world, queer cultural analysts would have working knowledge of the philosophical and critical tradition—the Hegelian notion of recognition, the Deleuzian notion of event, say—behind queer theory. They would also know that topic modeling is a mixed membership model for clustering, why the Dirichlet distribution is useful for it, and in addition to that, would know that in Python, dictionary lookup is $O(1)$. Such knowledge would allow a research working in queer CA to create "multi-disciplinary project[s]—a bridge between the humanities and quantitative social science, belonging equally to both" (Underwood, "A Broader Purpose"). Such researchers would become truly comfortable across the circular research processes I have described and would have a grasp of both the relevant queer scholarship and the heuristics behind the computational methods being deployed.

Recall that one of the earliest alliances between queer people and DH was the American Studies Association roundtable, "Transformative Mediations? Ethnic and Queer Studies and the Politics of the Digital" as well as the ensuing collaborative efforts of #transformDH. Queer CA must continue to exhibit this desire to transform, and remain vocal: what has not been built that should be? And what has been egregiously ignored but should be explored instead? After all, the rebellious impulse behind queer theory—and often DH—requires its practitioners to stay creative, to continuously reimagine what a person, a discipline, an institution can do, despite their marginalized status.

In cultural analytics, there have been numbers. And it is my hope that there will be queers.[22] And more queers, until most of this essay feels incredibly conservative and obsolete.

BIBLIOGRAPHY

Algee-Hewitt, Mark. "Criticism, Augmented." *In the Moment*, Apr. 2019. critinq.wordpress.com/2019/04/01/computational-literary-studies-participant-forum-responses/.

Arondekar, Anjali, Ann Cvetkovich, Christina B. Hanhardt, Regina Kunzel, Tavia Nyong'o, Juana María Rodríguez, Susan Stryker, Daniel Marshall, Kevin P. Murphy, and Zeb Tortorici. "Queering Archives: A Roundtable Discussion." *Radical History Review*, no. 122, 2015, pp. 211–231.

Babuscio, Jack. "Camp and the Gay Sensibility." *Camp Grounds: Style and Homosexuality*, edited by David Bergman, University of Massachusetts Press, 1993.

Barnett, Fiona, et al. "QueerOS: A User's Manual." *Debates in the Digital Humanities 2016*, University of Minnesota Press, 2016, pp. 50–59.

Bianco, Jamie "Skye." "Man and His Tool, Again? Queer and Feminist Notes on Practices in the Digital Humanities and Object Orientations Everywhere." *Digital Humanities Quarterly*, vol. 9, no. 2, 2015, p. n.p.

22. The deliberate use of *queers* here aims to invoke and commemorate the activists and academic work done since the 1960s; it does not have any pejorative connotation.

Bode, Katherine. "Why You Can't Model Away Bias." *Modern Language Quarterly*, vol. 81, no. 1, Mar. 2020, pp. 95–124.
Brown, Susan, and Laura Mandell. "The Identity Issue." *Journal of Cultural Analytics*, vol. 1, no. 1, Department of Languages, Literatures, and Cultures, Feb. 2018, culturalanalytics.org/article/11036-the-identity-issue.
Butler, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge, 1990.
———. "Imitation and Gender Insubordination." 1990. *The Lesbian and Gay Studies Reader*, edited by Henry Abelove et al., Routledge, 1993, pp. 307–20.
Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, vol. 81, PMLR, 2018, pp. 77–91, proceedings.mlr.press/v81/buolamwini18a.html.
Cheng, Jonathan. "Fleshing Out Models of Gender in English-Language Novels (1850—2000)." *Journal of Cultural Analytics*, Jan. 2020, culturalanalytics.org/article/11652-fleshing-out-models-of-gender-in-english-language-novels-1850–2000.
Connolly, Randy. "Why Computing Belongs within the Social Sciences." *Communications of the ACM*, vol. 63, no. 8, July 2020, pp. 54–59.
Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry*, vol. 45, no. 3, Mar. 2019, pp. 601–39.
Denton, Emily, et al. "Bringing the People Back In: Contesting Benchmark Machine Learning Datasets." *ArXiv:2007.07399 [Cs]*, July 2020, arxiv.org/abs/2007.07399.
Doan, Laura. "Queer History / Queer Memory: The Case of Alan Turing." *GLQ: A Journal of Lesbian and Gay Studies*, vol. 23, no. 1, 2017, pp. 113–36.
Gaboury, Jacob. "Becoming NULL: Queer Relations in the Excluded Middle." *Women & Performance: A Journal of Feminist Theory*, vol. 28, no. 2, May 2018, pp. 143–58.
Gadamer, Hans-Georg. *Truth and Method*. Translated by Joel Weinsheimer and Donald G. Marshall, Bloomsbury, 2013.
Halberstam, J. Jack. *In a Queer Time and Place: Transgender Bodies, Subcultural Lives*. New York University Press, 2005.
———. *The Queer Art of Failure*. Duke University Press, 2011.
Halperin, David M. *How to Do the History of Homosexuality*. Univ. of Chicago Press, 2002.
Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies*, vol. 14, nol. 3, pp. 575–599.
Hinton, Geoff. *Google's AI Guru Wants Computers to Think More Like Brains*. Interview by Tom Simonite, 12 Dec. 2018, www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/.
Hovy, Dirk, and Shannon L. Spruit. "The Social Impact of Natural Language Processing." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Association for Computational Linguistics, 2016, pp. 591–98.
Jiang, May, and Christiane Fellbaum. "Interdependencies of Gender and Race in Contextualized Word Embeddings." *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, Association for Computational Linguistics, 2020, pp.

17–25.
Kluttz, Daniel N., et al. "Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions." *After the Digital Tornado: Networks, Algorithms, Humanity*, edited by Kevin Werbach, Cambridge University Press, 2020, pp. 137–52.
Kraicer, Eve, and Andrew Piper. "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction." *Journal of Cultural Analytics*, 2019, culturalanalytics.org/article/11055-social-characters-the-hierarchy-of-gender-in-contemporary-english-language-fiction.
Kramnick, Jonathan. "Criticism and Truth." *Critical Inquiry*, vol. 47, no. 2, Jan. 2021, pp. 218–40.
Lubin, Joan. "Queer Formula." *Post45*, no. 5: Formalism Unbound, Part 2, Jan. 2021, post45.org/2021/01/lubin-queer-formula/.
Migraine-George, Thérèse, and Ashley Currier. "Querying Queer African Archives: Methods and Movements." *WSQ: Women's Studies Quarterly*, vol. 44, no. 3–4, 2016, pp. 190–207.
Nguyen, Dong, et al. "How We Do Things with Words: Analyzing Text as Social and Cultural Data." *Frontiers in Artificial Intelligence*, vol. 3, 2020, p. 62.
Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018.
Piper, Andrew. "Think Small: On Literary Modeling." *PMLA*, vol. 132, no. 3, May 2017, pp. 651–58.
Quist, Jennifer. "Laurelled Lives: The Swedish Academy's Praise for Its Prizewinners." *New Left Review*, vol. 104, no. March—April, 2017, pp. 93–106.
Ruti, Mari. *The Ethics of Opting out: Queer Theory's Defiant Subjects*. Columbia University Press, 2017.
Sap, Maarten, et al. "Connotation Frames of Power and Agency in Modern Films." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2017, pp. 2329–2334.
Schlesinger, Ari, et al. "Intersectional HCI: Engaging Identity through Gender, Race, and Class." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 2017, pp. 5412–27.
Schmidt, Benjamin M. "Do Digital Humanists Need to Understand Algorithms?" *Debates in the Digital Humanities 2016*, University of Minnesota Press, 2016, pp. 546–55.
Shechtman, Anna. "The Medium Concept." *Representations*, vol. 150, no. 1, May 2020, pp. 61–90.
So, Richard Jean. "'All Models Are Wrong.'" *PMLA*, vol. 132, no. 3, May 2017, pp. 668–73.
So, Richard Jean, and Edwin Roland. "Race and Distant Reading." *PMLA*, vol. 135, no. 1, Jan. 2020, pp. 59–73.
Spade, Dean, and Rohlfs, Rori. "Legal Equality, Gay Numbers and the (After?) Math of Eugenics." *S&F Online*, Vol. 13, No. 2, Spring 2016, sfonline.barnard.edu/navigating-neoliberalism-in-the-academy-nonprofits-and-beyond/dean-spade-rori-rohlfs-legal-equality-gay-numbers-and-the-aftermath-of-eugenics/0/.

Underwood, Ted. "A Broader Purpose." *The Stone and the Shell*, Jan. 2018, tedunderwood.com/2018/01/04/a-broader-purpose/.

———. *Distant Horizons: Digital Evidence and Literary Change*. The University of Chicago Press, 2019.

———. "Machine Learning and Human Perspective." *PMLA*, vol. 135, no. 1, Jan. 2020, pp. 92–109.

Wallach, Hanna. "Computational Social Science ≠ Computer Science + Social Data." *Communications of the ACM*, vol. 61, no. 3, Feb. 2018, pp. 42–44.

Wilde, Oscar, and Nicholas Frankel. *The Uncensored Picture of Dorian Gray*. Belknap Press of Harvard University Press, 2011.

Zeffiro, Andrea. "Towards a Queer Futurity of Data." *Journal of Cultural Analytics*, May 2019. culturalanalytics.org/article/11050-towards-a-queer-futurity-of-data.