CA Journal of Cultural Analytics

June 17, 2021

Can We Map Culture?

Ted Underwood, Richard Jean So

Ted Underwood, University of Illinois, Urbana-Champaign
Richard Jean So, McGill University

Peer-Reviewer: Michael Gavin, Simon DeDeo
Data Repository: 10.7910/DVN/J89DNM

ABSTRACT

Images that convert culture into physical space have a durable appeal, and numbers make it possible to literalize a spatial representation of culture by measuring the "distances" between cultural artifacts. But do cultural relationships really behave like physical distance? There are good reasons to think the analogy is imperfect, and a number of alternative geometries have been proposed—extending, in a few cases, to a systematic distinction between the mathematics of "embodied experience" and "epistemic experience" (Chang and DeDeo 2020). We test several proposed alternatives to spatial metrics against ground truth implicit in human behavior. While it is sometimes possible to improve on distance metrics, we do not yet find evidence that the information-theoretical measures recommended as appropriate for epistemic questions are generally preferable in the cultural domain.

One popular way to visualize cultural categories is to convert them into physical space.[1] The best-known example may be the map of Disneyland, which organizes imaginative genres as neighborhoods called Fantasyland, Tomorrowland, Adventureland, and Frontierland. Other artists have mapped film genres and scientific genealogies, and the internet is now a popular subject for cultural cartography: in figure 1, Randall Munroe represents internet platforms as a fractured archipelago.[2]

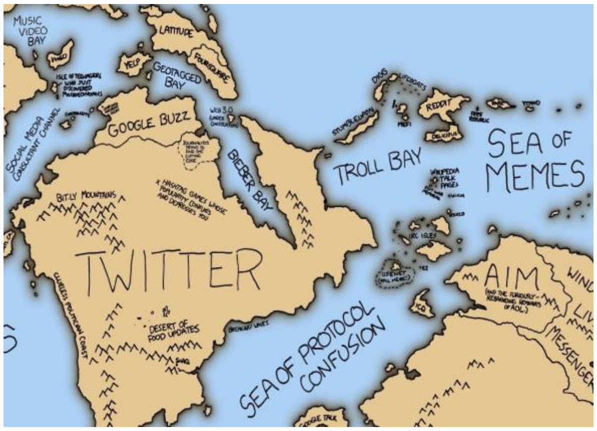

Figure 1: Detail from Randall Munroe, "Online Communities 2," XKCD, 2010.

Journal of Cultural Analytics 6 (2021): 32-51. doi: 10.22148/001c.24911

A map of mountains and coastlines has obvious comic potential: the website Pirate Bay can become a physical bay. But in the twenty-first century, fanciful maps of culture are also competing with visualizations that make a more serious claim to represent culture spatially. A project called Every Noise at Once, for instance, has measured the stylistic distances between genres of music in order to represent each genre as a point on a plane. “Turkish classical” ends up close to “Bulgarian folk,” while “doomcore” is close to “brostep.” Computational analysis gives this project a new kind of foundation, but the underlying strategy is not alien to older forms of cultural sociology. Pierre Bourdieu, for instance, similarly diagrammed cultural production as a “field” (French: *champ*) organized by two Cartesian axes.[3]

Instead of taking genre categories as units, some researchers plot individual works and infer genres from the groupings that emerge. When Benjamin Schmidt reduced the textual differences between books to two dimensions, he found broad clusters that loosely align with our generic concepts of poetry, drama, and fiction (figure 2).[4]

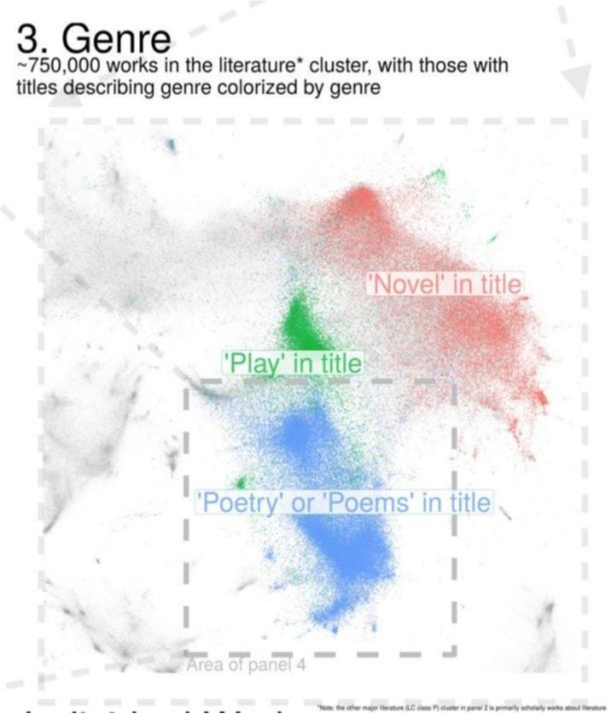

Figure 2: Detail from figure 1 in Benjamin Schmidt, “Stable Random Projection.”

In short, quantitative approaches to culture rely heavily on the conceit that cultural artifacts are positioned in a field analogous to physical space. But how seriously should we take the metaphors implicit in these visualizations? Does culture really behave like space? If not, how are cultural relationships distorted when we convert them into spatial ones?

These are not new questions for digital humanists. They emerge from a conversation about the limits of spatial representation that can be traced back at least as far as Johanna Drucker's 2011 suggestion that humanists need to replace the "realist models of knowledge" implied by standard maps with "a non-standard map that expresses the constructedness of space."[5] What may be new in the last few years is a growing confidence that we can describe the shortcomings of spatial metaphors mathematically, and find mathematical alternatives to them.

Consider, for instance, an article that appeared in this journal a few months ago: "Divergence and the Complexity of Difference in Text and Culture," by Kent Chang and Simon DeDeo.[6] The authors of this piece suggest that humanists have a lot to gain by letting go of the mathematical assumptions that govern spatial embodiment. For instance, one of our basic intuitions about space is that points are related by a function we call "distance." Distance can be measured in several different ways—as a straight line between two points, say, or as the jagged path you actually need to drive on a grid of streets. But all of these distance metrics conform to four simple intuitions:

1. The distance from a point to itself is zero.
2. All other distances are greater than zero.
3. The distance from A to B is the same as the distance from B to A (the principle of symmetry.)
4. And the direct distance from A to C must be equal to or less than the indirect path that goes from A to B and then from B to C. This is the triangle inequality, or as Chang and DeDeo more vividly express it, the no-shortcut rule.

These intuitions about distance are flexible; they can hold true even in spaces that are curved or torus-shaped.[7]

But there are also ways to measure difference that don’t obey the spatial logic of distance. Information theory describes the difference between two probability distributions with a measure called Kullback-Leibler divergence, which violates both the principle of symmetry and the no-shortcut rule. Chang and DeDeo argue that these transgressions are valuable for humanists. In breaking the rules that govern distance, KL divergence creates a measure more appropriate for “epistemic” relationships, like novelty or surprise. Breaking the no-shortcut rule makes it easier to describe the bridging effect of a transitional passage, for example, since the path from passage A, through B, to passage C can actually be shorter than the direct path from A to C.[8] Chang and DeDeo also suggest that the asymmetry of KL divergence allows it to describe situations where one cultural category partially “encloses” another. The KL divergence is smaller when we move from a relatively uniform distribution to a concentrated one than it would be going the other direction. So, using this measure of difference, a work that distributes its attention broadly across different topics can be more similar to its narrowly focused contemporaries than any of them are to it.[9]

These are important insights, and we probably ought to bear them in mind when surveying maps of culture. There is, of course, nothing wrong with projecting cultural categories onto a two-dimensional plane. All models simplify the things they represent: that’s why models are useful.[10] But Chang and DeDeo correctly remind us that the simplification entailed in mapping culture goes deeper than the obvious compression of many dimensions into two. Cultural relationships don’t have to obey a spatial logic at all.

On the other hand, measures of distance may not be quite as dominant in the computational humanities as “Divergence and the Complexity of Difference” implies in setting up this opposition. Most humanists who use mathematics are already using measures that violate some of the rules Chang and DeDeo describe. Cosine distance, for instance, is the measure most commonly used in quantitative studies of literary culture. If we envision texts as points in a high-dimensional space where each dimension measures the presence of a particular feature (a word or topic), Euclidean distance is simply the length of a line drawn directly from text A to text B. Cosine distance, by contrast, considers lines drawn to both A and B from the origin of a coordinate system, and measures the angle between the lines. Since the cosine of the angle θ is a measure of similarity, researchers typically take 1 - cos(θ) as "cosine distance." This measure has been shown to work better than Euclidean distance for many text-mining tasks.[11] But it does violate the no-shortcut rule. If we define three sample vectors,

A = [0.25, 0.25, 0.25, 0.25]

B = [0.2, 0.3, 0.3, 0.2]

C = [0.1,. 0.4, 0.4, 0.1]

the cosine distance from A to C is 0.1425. But the distance from A to B is .0194, and from B to C is .0583. Adding those two distances produces .0777. So the path going through B is a shortcut that reduces the distance from A to C by almost half. Although we speak about cosine "distance," the measure is not strictly a distance metric.[12]

Cosine distance does obey the rule of symmetry: the angle from A to B is the same as the angle from B to A. But other measures of difference used in recent humanistic scholarship violate even the rule of symmetry. For instance, some scholars have recommended measuring the difference between cultural categories by asking whether a model trained to distinguish category A (from some background population) can also distinguish category B (from the same population).[13]

One advantage of this approach is that predictive models are tailored to a particular category, and attend only to the features that consistently characterize it. Universal measures of distance and divergence are less discriminating. A measure of divergence between word frequencies, for instance, might give a lot of weight to a common word like *that.* Subtle variations in the frequency of *that* could increase the distance between country-house mysteries and hard-boiled detective novels, even if the word was not a reliable signal of either genre. But a model specifically trained to recognize a genre will attend mostly to a shorter list of words (say, *murder*, *door*, and *investigate*) that reliably distinguish the genre from other works of fiction. If the same features that distinguish country-house mysteries from the general run of fiction turn out to distinguish hard-boiled detective novels, we have strong grounds for describing the categories as similar—no matter how different they are when it comes to *that, vicar,* and *orange.*

In many ways predictive models are a good fit for our intuitions about cultural difference. Using them, however, does force us to jettison intuitions about symmetry. It is quite common for a model trained on set A to have difficulty recognizing set B, even when a model trained on B can recognize A. This asymmetry emerges fundamentally from the same logic of enclosure that Chang and DeDeo describe in information theory.[14] A generalized model that spreads its attention across many different features can often recognize categories that are in some sense "contained" by the one it was trained to distinguish—even if those works were not actually in its training set. But the logic doesn't work in reverse: a model trained on a narrow category will not necessarily recognize its affinity to all aspects of a broader category. In practice, for instance, a model of country-house mysteries can recognize Raymond Chandler better than a model trained on hardboiled detectives can recognize Agatha Christie; the hardboiled genre seems to be more specialized.[15]

Like Chang and DeDeo, we can offer an array of reasons for believing that the mathematics of predictive modeling are suited, in principle, to humanistic topics and ways of knowing. Underwood has argued elsewhere that the radical flexibility of predictive models makes them a good fit for the perspectival character of cultural difference.[16] Cultural categories are not just containers for objects. They imply a particular experience of the world—which means that the dots on a map of contemporary music genres, for instance, may also be vantage points from which different observers survey the musical landscape and see it differently. In a landscape of this kind, distance should arguably be defined relative to a vantage point. Information theory takes a few steps toward this relativistic goal, but predictive models go further. For instance, KL divergence presumes that the variables most strongly represented in a text are the most significant facts about it. A predictive model of cultural difference, by contrast, can decide that some features central to the difference between A and B are entirely irrelevant to the contrast between C and D (even if those features are strongly represented in C and D).

But this argument from first principles—however persuasive it may be for mathematicians—will not do much to convince humanists. Nor do we think humanists should be convinced. The main problem confronting mathematical models of culture is not that researchers have forgotten to reflect on first principles, but that it often remains unclear whether a model has been confirmed. Consensus is elusive because the phenomena to be modeled are themselves contested and poorly defined. For instance, we casually mentioned above that predictive models treat the country-house mystery as a category that partially encloses grittier kinds of detective fiction, implying that Raymond Chandler is more "specialized" than Agatha Christie. Is this an intuitive result, confirming our model? Or is it just wrong? Readers probably won't agree. Similarly, Chang and DeDeo point to a thought experiment where *War and Peace* turns out to be a more encompassing book than *Anna Karenina*. Maybe this is intuitive, since one of the novels, after all, does describe both war *and* peace. But we suspect it would be easy to find a reader of Russian literature who would argue fiercely that in an important sense *Anna Karenina* encompasses more of the heights and depths of human experience.

To address this difficulty, computational approaches to culture will have to build consensus slowly, on an empirical foundation. Ground truth about human judgment will have to be gathered and hypotheses will have to be preregistered before anyone can argue convincingly that one model is better than another.

Researchers have been proceeding this way long enough now that we are beginning to make some headway on the basic premises of the field. For a long time, it seemed doubtful that it was even possible to draw meaningful conclusions from a mathematical representation of culture. A bag-of-words representation of literary texts, in particular, seemed intuitively deficient to most observers. Critiques of computational literary study still direct much of their scorn at the obvious folly of merely "counting words."[17]

But experiments with preregistered hypotheses have gradually demonstrated that, for instance, human judgments about the differences between genres do correlate strongly and consistently with the differences between lexical representations of works. José Calvo Tello has tested this thesis by comparing textual models to agreement between bibliographies. He gets a strong result (Spearman's $\rho = .76$).[18] Other researchers have compared the strength of similarity between book texts in various genres to the strength of similarity between their reviews (n= 14, r = .87, p

< .001).[19] By itself, neither test would be definitive. But together, and in concert with other evidence, these experiments are starting to add up. Lexical models do provide a solid foundation, at least for broad historical conclusions.

Arguments about the mathematical structure of cultural difference will have to advance in a similar way. It is certainly plausible to say that cultural relationships have a non-Euclidean and perspectival logic. Johanna Drucker suggested as much ten years ago, and many people agree in principle. But humanists are not likely to put faith in any particular model of cultural geometry until the model has been extensively tested against ground truth.

That will take time; we don't expect to close the case in one article. But in the pages that follow, we will glance briefly at two experiments testing the hypotheses that Chang and DeDeo (and we ourselves) have advanced. Can non-spatial measures of cultural difference—drawn from information theory or from predictive modeling—significantly improve on distance measurement? To spoil the suspense: our answer is "Yes, in certain cases where there's a specific need for an alternative measure. But the evidence doesn't yet support a general opposition between 'spatial' and 'cognitive' measures of difference, or demonstrate that the latter are more appropriate for culture."

Experiment 1: The social proximity of genres in libraries

Several recent articles in cultural analytics, including works by the present authors, compare the texts of books or poems to draw conclusions about the affinities or differences between genres. But genres are also social practices, and they might overlap in ways that aren't legible in the text. For instance, fans of science fiction are often also fans of fantasy, and this might remain true—at least in theory—even if the two genres were textually very distinct. So how do we know that measures of textual similarity and difference have any relation at all to the extra-textual life of genres?

We don't yet have rich structured data about historical reading practices. But library catalogs do give us some indirect evidence about pairs of genre categories that human readers perceive as closely overlapping or as remote and incompatible. We get this evidence from the fortunate fact that a single volume can carry any number of genre tags or subject headings. This overlap allows us to estimate the cultural proximity of categories by asking which categories tend to be assigned to the same books. For instance, *The Chronicles of Narnia* may be tagged “Fantasy,” but also “Christian Fiction.” If those tags co-occur commonly, we can infer they are more compatible than less common pairings like “Christian Fiction” and “Erotica.”

This evidence permits us to stage a simple experiment, asking which measures of textual similarity correspond most closely to extra-textual evidence about generic similarity.[20] We can measure the social overlap between categories in the library as pointwise mutual information. Then we can compare the social proximity of pairs of categories to various measures of similarity between groups of texts. (When comparing two groups of texts we will, obviously, need to exclude books that bear both genre tags from the sets being compared. Otherwise textual similarity would correlate with the overlap between categories for self-evident circular reasons.)

In measuring textual similarity, we can start with familiar metrics, like cosine distance on tfidf-transformed word frequencies or topic vectors. (Euclidean distance was also measured in this experiment, although it is not shown in figure 3 because it always performed worse than cosine distance.) In addition, to test the value of predictive modeling, we can train models to distinguish each genre category from a random sample of books in the library, and measure the difference between categories A and B by using a model of A to identify books in B, and vice versa. You could call this a measure of “mutual misrecognition”: how much information do we lose in applying each category to the other? To be precise, we measure the difference between A and B the as the Fisher’s z-transformed Spearman correlation between the predictions of the two models:

$$\begin{aligned} misrec(A,B) = -\frac{1}{2}\Big(&\operatorname{arctanh}\big(\operatorname{spearman}\big(P(a|ModelA), P(a|ModelB)\big)\big) \\ &+ \operatorname{arctanh}\big(\operatorname{spearman}\big(P(b|ModelB), P(b|ModelA)\big)\big)\Big) \end{aligned}$$

Here we average predictions made in both directions to render the measure symmetric (“mutual”), but researchers who want an asymmetric measure of misrecognition can get that by using the first half of the equation.

Of these various methods of measuring textual similarity, the last had the strongest correlation (*n* = 491, *r* = .45) to the practices of library catalogers.

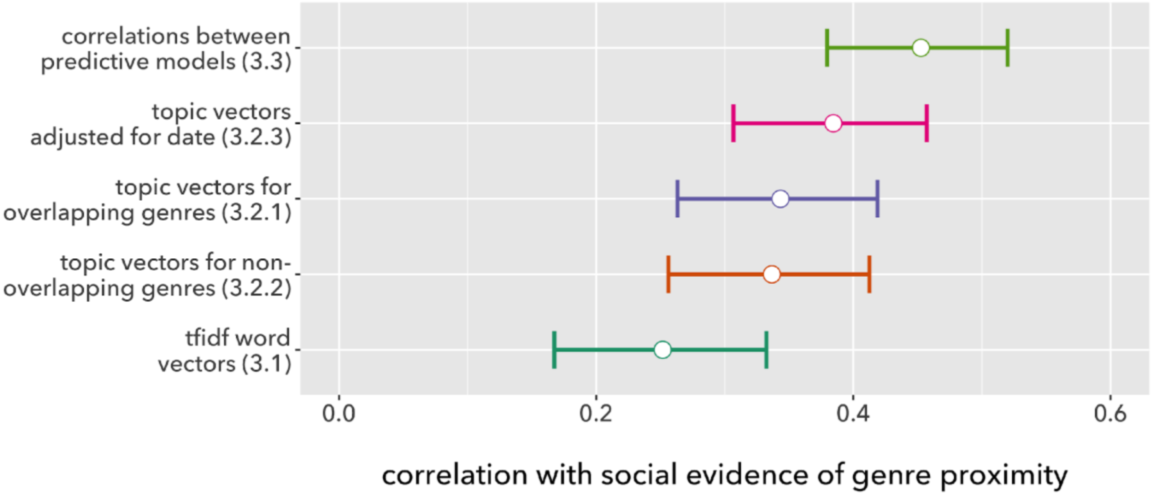

Figure 3: Pearson correlation between textual similarity and evidence of social proximity. Comparisons were made on repeated subsamples of genres; error bars cover a 95% confidence interval. From Ted Underwood, "The Historical Significance of Textual Distances."

Is a correlation of .45 good enough? Is it significantly better than the correlations we observed measuring cosine distances between topic vectors (which ranged from .33 to .38)? This is a complex question. We should start by observing that this is not exactly a test against known ground truth. It was plausible to hypothesize that the behavior of library catalogers would correlate with the textual similarities between genres. But it was not self-evident, and it may not be safe to judge a metric by its ability to produce a new and contested result. Also, while uncertainty is modeled in figure 3 through bootstrap resampling, it would be more informative to test these metrics on entirely different categories.

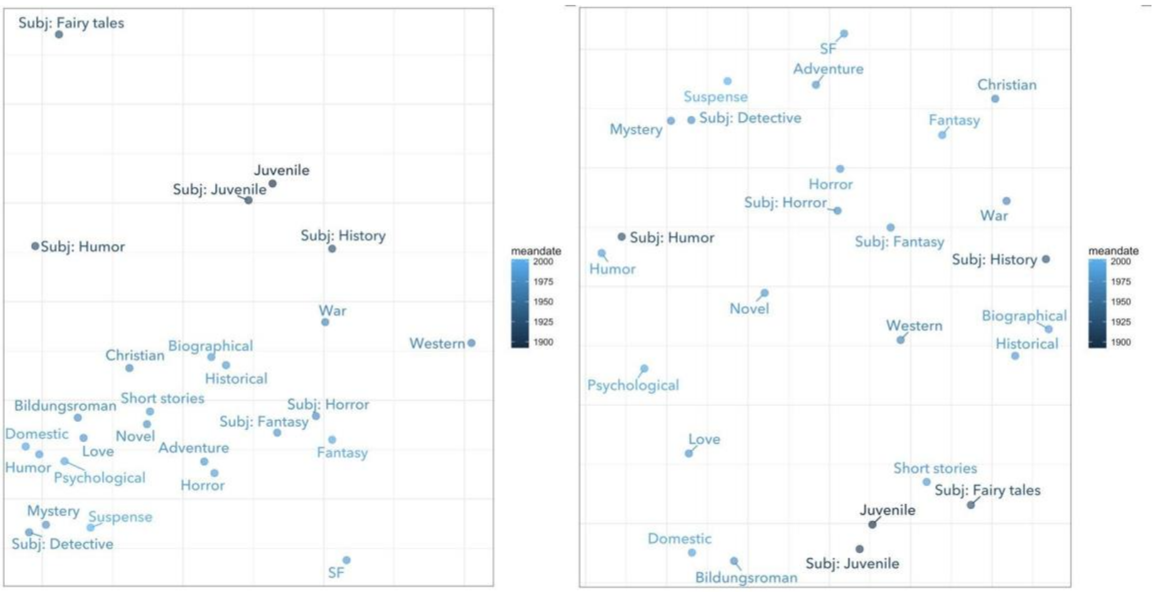

Figure 4: Distances inferred from topic vectors (left) and predictive models (right), reduced to two dimensions through multidimensional scaling.

On the other hand, we have a solid theory to explain the superior performance of predictive modeling. Predictive models performed well in this experiment because they were able to tune out aspects of generalized lexical distance that weren’t specific to the categories being contrasted. For instance, some genre and subject headings are older on average than others. English diction changes across time, so a generalized measure of lexical distance tends to represent all the older categories as quite distinct from all the newer ones. (See the left side of figure 4.) But these contingent differences of vintage aren’t central to human intuitions about the similarity of literary categories, and are not reflected in the choices of library catalogers. So methods of comparison performed better if they were able to factor publication date out of the comparison. Predictive models did this automatically, because the random contrast set in each model was selected to match the positive category’s distribution across time. (To give topic vectors a fighting chance, we also ran a test where topic vectors were corrected for date.) But predictive models might also be tuning out many other dimensions of the corpus analogous to publication date—lexical features that introduce noise for generalized distance measures but that aren’t actually relevant to the literary categories contrasted in this experiment. This tightness of focus probably explains the superior performance of a relativistic approach.

And yet, even if we believe that the differences of accuracy reported in figure 3 are robust and supported by a clear theoretical rationale, these modest differences probably won't justify abandoning a simple, handy metric like cosine distance. To be blunt, training a model for every genre is a pain. The strategy only works if you have a reasonably large number of texts in each category. Moreover, the measure of difference we have called "misrecognition" is a complicated three-cornered concept which involves not only two categories of texts but a background against which they are both discriminated. Hidden pitfalls may lurk in the selection of that background. If you do have abundant data, and if you need to measure difference in a way that corresponds as closely as possible to human intuition, this experiment suggests that it is worth exploring a relativistic, predictive approach. But a simple distance measurement is likely to work nearly as well.

Experiment 2: Innovation in fiction

Our second experiment is designed to test whether divergence measures borrowed from information theory provide a better picture of historical change than distance metrics. Most of the experimental design is borrowed from an essay about French Revolutionary debate by Alexander T. J. Barron, Jenny Huang, Rebecca L. Spang, and Simon DeDeo.[21] The authors apply information theory to understand "the emergence and persistence of word-use patterns in over 40,000 speeches" made in the National Constituent Assembly during the Revolution (1787-1794). They start by topic modeling the collection so each speech can be represented as a topic vector. Then they evaluate each speech by measuring its average KL divergence from the speeches that precede it in a time window *w*; they call this value "novelty." They also measure the KL divergence from each speech to the speeches that follow it within time *w*, and call this "transience." In both cases, the asymmetry of KL divergence is aligned with the passage of time. The authors suggest that this allows them to measure not just distance but the epistemic "surprise" of new patterns.

Generally speaking, speeches with high novelty also have high transience: unusual innovations are also quickly abandoned. But this is true only as a generalization: there are always some speeches that do innovate in durable ways. A speech of this kind diverges further from the past than the future will diverge from it. The authors quantify this tendency by subtracting transience from novelty, and call the difference "resonance." They also validate this measurement against their narrative understanding of history. Observed patterns of resonance align, for instance, with the committee structure of the Assembly.

We applied the same methods to 39,784 English-language works of fiction spread across the nineteenth and twentieth centuries (1800-2009), in order to identify books that were literally "ahead of their time"—closer to the future, on average, than to the past. We measure this temporal asymmetry using the same methods Barron et al. use to measure "resonance," but we prefer to call the measured quantity "precocity." The rationale for this choice will emerge more fully as we analyze results. But the key is that "resonance" might make it sound as if we're measuring the effect a text had on its audience, which is (as we will see) a debatable inference—at least at the scale we are investigating.

To establish the validity of this method, we began by pre-registering several hypotheses about categories of books we would expect to be unusually precocious.[22] For instance, we posited that precocity would correlate with the (log-transformed) number of copies of a book preserved in digital libraries. (Books that anticipated the future seem likely to be preserved by posterity.) We also hypothesized that works of fiction reviewed in prominent venues (1850-1950) would have higher than average precocity, and we posited the same thing about a specific group of twenty volumes that we selected by using our knowledge of literary history to identify works that have been widely imitated (*Jane Eyre*, *Native Son*) or that shaped an emerging genre (*The Time Machine*). Finally, we created lists of works in Norton anthologies or widely discussed by scholars in JSTOR, in order to ask whether canonical works are generally ahead of their time.

This experiment involved longer texts and covered a much longer timeline than the seven-year archive in Barron et al. We created a topic model with 440 topics—a relatively high number—to ensure we were capturing the granularity of this large collection. We also used a time window of 25 years, which is vastly longer than the (roughly) two-day window in Barron et al.[23] But we found the experimental strategy those researchers used to understand political debate equally effective for literary change. Most of our hypotheses were strongly confirmed. The list of books we specifically identified as influential had precocity, on average, more than a full standard deviation above the mean (*n* = 20, Cohen's *d* = 1.14, *p* < .005). Widely-reviewed novels were about half a standard deviation above the mean (*n* = 541, Cohen's *d* = 0.482, *p* < .00001). And the log-transformed number of copies preserved in a library correlated with a volume's precocity weakly but significantly (*n* = 38,794, *r* = .101, *p* << .00001).[24] Although we didn't preregister this hypothesis, bestsellers also turn out to be precocious—at least up to 1950. The effect fades thereafter.[25]

This evidence doesn't necessarily tell us much about causation. For instance, a book could be have been preserved in libraries because it was important and influential (in which case, the book is the cause). Or it could have become influential because libraries bought a lot of copies (librarians are the cause). Or, in certain cases, the book could have exerted no influence on readers at the time, becoming famous and widely purchased only *after* literary practice happened to move in its direction (the preferences of posterity are the cause). All we know, technically, is that temporal displacement correlates with certain forms of distinction and prominence. We should also note that a couple of our preregistered hypotheses were not confirmed. In particular, American works anthologized by Norton and Heath didn't display precocity significantly greater than average (although non-American works did).

But on the whole this experiment confirms the value of the methods in Barron et al. Comparing a volume to its own future and past produced a measure of temporal asymmetry that aligned very strongly with our expectation that influential and widely-reviewed volumes were in some sense "ahead of their time." The confirmation was strong enough that we now feel justified in interpreting exceptions to the pattern as puzzles in need of explanation. For instance, why were bestsellers (up to 1950) more precocious than currently canonized works of American literature? This is a surprising finding: book historians, such as John Sutherland, have defined the "ideological" purpose of the bestseller to be largely reactionary or conservative in nature: to express and feed the "needs of the reading public," which includes "consolidating prejudice."[26]

But the aspect of experimental design we validated most strongly is simply the plan of comparing a text to its own past and future. We did not see significant evidence that information-theoretic measures in particular were vital to the success of this experiment. On the contrary, we ran the same backward and forward measurements using cosine distance on topic vectors, and got nearly the same results we got with KL divergence. All the same hypotheses were confirmed. Some effect sizes were slightly smaller when measured with cosine distance. The preregistered list of 20 books was Cohen's *d* = 1.04 instead of 1.14. But several other effects were slightly larger. The list of 571 prominently-reviewed books was Cohen's *d* = 0.514, using cosine distance, instead of 0.482, using KL divergence. These small differences, in either direction, are easily swamped by choices about data cleaning and normalization, or even by random variation in bootstrap sampling; see figure 5.[27]

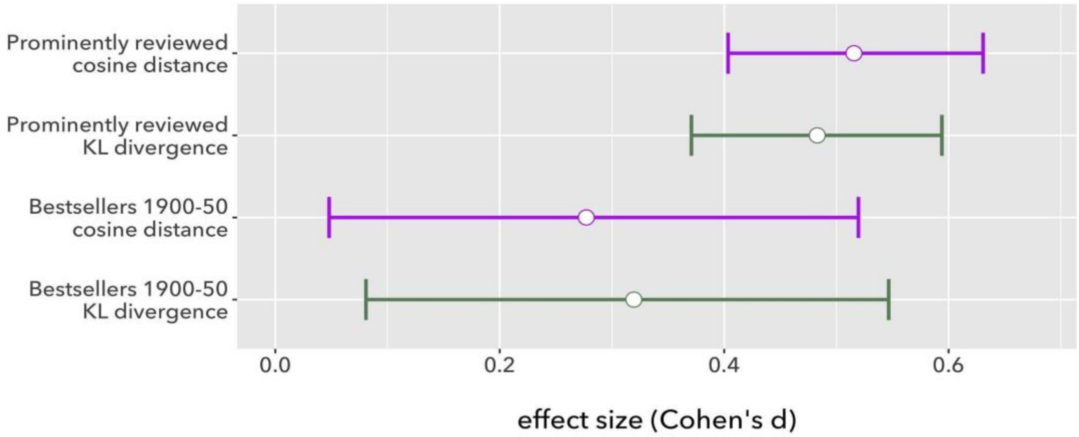

Figure 5: Effect sizes (Cohen's d) produced by contrasting the "precocity" of bestsellers 1900-1950, or prominently reviewed volumes 1850-1950, to a randomly selected contrast set. 95% confidence intervals are plotted with bootstrap resampling of the data. The differences between cosine distance and KLD are not large or consistent.

This inconclusive evidence is by no means an argument against using KL divergence. It is still possible that the estimates provided by information theory are better than those provided by spatial distance measures. Since we don't know the true effect size for any of the treatment categories tested, we can't evaluate accuracy. We can only report that the differences between these methods appear to be small.

But this is a case where negative evidence—the absence of a gap between the metrics—should shape our interpretation of results. Barron et al. argue that KLD is an appropriate metric for "rhetorical influence" because it is equivalent to "surprise." It can measure "the extent to which the expectations of an optimal learner, trained on one pattern, are violated by later patterns." This is a reasonable justification for using the metric to study historical change: if the differences between earlier and later texts were shaped by changes of breadth (entropy), KLD would be capable of capturing them.

However, we should be cautious about assuming that we have measured something directional (like “rhetorical influence” or “surprise”) simply because we have used a metric that would be *capable* of detecting a directional effect. In this case, it turns out that symmetric, non-directional measures of distance produce very nearly the same result. To be sure, KLD and cosine distance can produce different results when individual pairs of texts are compared—especially if we compare broad distributions to concentrated ones. For certain questions—e.g., questions about epistemic enclosure—KLD is clearly the right measure.[28] But we don’t yet see any evidence that it is generally preferable for questions about historical change. Perhaps cultural history is not in practice primarily expressed through changes of entropy? In any case, the differences between KLD and cosine distance seem to cancel out when groups of books are compared along the axes of interest in this experiment. The effects we actually measured seem functionally equivalent to distance—which is why, instead of talking about “surprise” or “influence,” we have characterized the measured effect simply as a forward displacement in time relative to other books published in the same year. “Precocity” seems a particularly appropriate word for this displacement, because books by young authors have a great deal of it. (They are, for obvious reasons, more similar to the future than the past.) But the same thing is true of works by prominent authors: those books are also similar to the future, and their authors seem in effect “younger” than they really are.

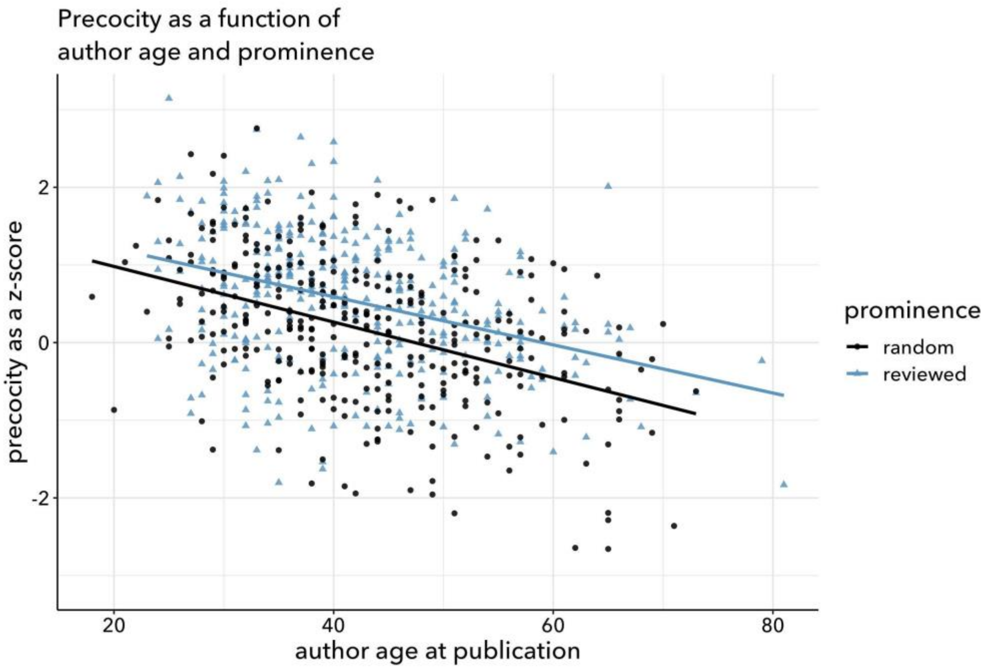

Figure 6: 694 volumes of fiction, 1850-1950. The reviewed volumes were selected from reviews in prominent periodicals; the random volumes were selected at random from HathiTrust. Younger writers always resemble the future, but prominent writers resemble the future more than might be expected for their age.

Setting quantitative evidence aside for a moment to speculate about underlying historical processes, we tend to suspect that Barron, Huang, Spang and DeDeo are right about the significance of the pattern they observe. Speeches with “resonance” in the National Constituent Assembly probably did have greater than average influence on their listeners. Although the time lags involved are much longer, we tend to suspect that the “precocity” we have measured in the history of fiction is also, at bottom, a symptom of a causal process. One needn’t believe that any single volume changed the world to surmise that reviewed and best-selling books represent, collectively, the leading edge of a wave of change. And a wave has causal agency—even if no single atom of water is terribly clear about its own role in the process. But in both of these cases, our suppositions about causality rest less on the mathematical structure of the metric used to measure a pattern than on familiarity with illustrative historical examples.

Conclusion

To answer the question in our title, then: yes, we can map culture. It is true that spatial measures of distance are not a perfect model of cultural difference. In at least a few experiments, other measures align demonstrably better with human behavior. But all models are simplifications, and the simplifications entailed in a spatial model of literary culture do not appear to be enormous or especially deceptive. The experiments reported above show that cosine "distances" measured on topic distributions (which are really spatial, even if not strictly distance metrics) usually provide a reasonable first approximation to other measures—and sometimes an indistinguishable approximation. Moreover, computational humanists already understand (and commonly use) alternative measurement strategies not grounded in spatial metaphors in cases where they are genuinely needed. There is at this point little danger that a few maps will lull the field into a naïvely realist, standpoint-free view of culture. We know that culture has a non-Euclidean geometry.

What researchers don't yet understand is the exact shape of this geometry. Perspectival models are often measurably better than standpoint-free models. But there are many different ways to acknowledge perspective. Chang and DeDeo offer principled reasons to believe that information-theoretical measures of divergence are a good fit for epistemic questions, and this seems likely to be true in cases where, say, relations between sets and supersets are at issue. But empirical evidence is still a little sparse for some application domains where these measures have been recommended. Although the arrow of time is definitely asymmetric, for instance, we don't know yet that the asymmetry of KL divergence really provides improved leverage on historical change.

The interpretive dilemma explored in this essay is likely to become more acute in the next few years. As deep neural models multiply, computational humanists will have no shortage of options for measuring difference. Some researchers are already using pre-trained transformers to measure the probability of the next event in a story, for instance, producing an estimate of "narrative flow."[29] Many new strategies will have a principled mathematical foundation. But the results they produce may not always signify what the math implies they should signify. To truly understand what we're seeing, we will need experiments that test proliferating computational metrics against evidence implicit in human behavior or explicit in human judgment.

References

[1] Some data for the second experiment in this paper was drawn from earlier collaborations with Sabrina Lee, Jessica Mercado, and Jordan Sellers. We also gratefully acknowledge editorial suggestions from Andrew Piper, Simon DeDeo, and an anonymous reviewer, all of which substantially changed our argument.

[2] Randall Munroe, "Online Communities 2," *XKCD,* October 6, 2010, https://xkcd.com/802/. For a map of physics (circa 1933), see Bernard H. Porter, "Being a Map of Physics," courtesy of Maine State Library and Mark Melnicove, in *Places & Spaces: Mapping Science,* edited by Katy Börner and Samuel Mills, http://scimaps.org/mapdetail/being_a_map_of_physi_171.

[3] Glen McDonald, "Every Noise at Once," accessed November 28, 2020, http://everynoise.com. Pierre Bourdieu, "The Field of Cultural Production, or: The Economic World Reversed," in *The Field of Cultural Production: Essays on Art and Literature,* ed. Randal Johnson (New York: Columbia University Press, 1993), 29-73.

[4] Benjamin Schmidt, "Stable Random Projection: Lightweight, General-Purpose Dimensionality Reduction for Digital Libraries," *Journal of Cultural Analytics* (September 30, 2018): 19. DOI: 10.22148/16.025.

[5] Johanna Drucker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5.1 (2011), http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

[6] Kent Chang and Simon DeDeo, "Divergence and the Complexity of Difference in Text and Culture," *Journal of Cultural Analytics* (October 7, 2020). DOI: 10.22148/001c.17585.

[7] Maurice Fréchet, "Les espaces abstraits topologiquement affines," *Acta Mathematica* 47.1-2 (1926): 25-52.

[8] Chang and DeDeo 10, 20-22.

[9] Chang and DeDeo, 12-19.

[10] Richard Jean So, "All Models are Wrong," *PMLA* 132.3 (2017): DOI: 10.1632/pmla.2017.132.3.668.

[11] Stefan Evert et al., "Understanding and Explaining Delta Measures for Authorship Attribution," *Digital Scholarship in the Humanities* 32 (2017), DOI: 10.1093/llc/fqx023.

[12] There are ways to convert cosine distance into a distance metric. But in practice, digital humanists don't.

[13] Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: University of Chicago Press, 2019), 41-47.

[14] Chang and DeDeo, "Divergence and the Complexity of Difference," 12-19.

[15] The models underlying this claim are available in Ted Underwood, "Data and Code to Support Distant Horizons" (Version v1.1, 2018). Zenodo. DOI: 10.5281/zenodo.1207277 . See particularly https://github.com/tedunderwood/horizon/tree/master/chapter2/modeloutput.

[16] Ted Underwood, "Machine Learning and Human Perspective," *PMLA* 135.1 (January 2020).

[17] Nan Z. Da, "The Computational Case against Computational Literary Studies," *Critical Inquiry* 45 (2019): 607.

[18] J. Calvo Tello. "Genre Classification in Spanish Novels: A Hard Task for Humans and Machines?" EADH 2018, https://eadh2018.exordo.com/programme/presentation/82.

[19] Kent Chang et al., “Book Reviews and the Consolidation of Genre,” DH 2020, DOI: 10.17613/02q2-1v27.

[20] Most of this work was previously published in Ted Underwood, “The Historical Significance of Textual Distances,” *Proceedings of the Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature,* Santa Fe, New Mexico, August 25, 2018, https://www.aclweb.org/anthology/W18-4507.

[21] Alexander T. J. Barron, Jenny Huang, Rebecca L. Spang, and Simon DeDeo, “Individuals, institutions, and innovation in the debates of the French Revolution,” *PNAS* 115.18 (2018). DOI: 10.1073/pnas.1717729115.

[22] William E. Underwood and Richard Jean So. “Temporal Asymmetry in the History of Fiction.” OSF (June 19, 2018): osf.io/zuq9a.

[23] The data for this model was drawn from HathiTrust. Boris Capitanu et al. (2016). The HathiTrust Research Center Extracted Feature Dataset (1.0) [Dataset]. HathiTrust Research Center, http://dx.doi.org/10.13012/J8X63JT3. Resources we used to analyze it include Andrew Kachites McCallum, "MALLET: A Machine Learning for Language Toolkit." 2002. http://mallet.cs.umass.edu, and Pauli Virtanen et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

[24] The code used in this analysis is documented in https://github.com/tedunderwood/asymmetry, as well as in the Journal of Cultural Analytics Dataverse.

[25] The 1950 dividing line is likely not incidental; book and literary historians have argued that the category of the “bestselling novel,” at least in the United States, took on a new character in the postwar period with the rise of industry conglomeration. Before 1950, bestseller lists still contained a number of novels that could be described as “excellent,” such as Ernest Hemingway’s *From Whom the Bell Tolls*, that also happened to sell well. After 1950, likely an effect of conglomeration and publishers focusing on promoting a shrinking number of books intended largely to sell, but not necessarily demonstrate literary excellence, bestseller lists became more heavily populated by books such as Mario Puzo’s *The Godfather* and Judith Krantz’s *Scruples*. For recent discussions of this effect, see Dan Sinykin’s forthcoming *The Conglomerate Era* (Columbia University Press, 2021) and Richard Jean So, *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (Columbia University Press, 2020).

[26] John Sutherland, *Bestsellers: Popular Fiction of the 1970s* (Routledge, 1981), 34.

[27] We normalized topic vectors for date and nationality, for instance. This was important because edge effects in topic modeling otherwise compress distances at the end of a long timeline. Also, American books are unevenly distributed across time.

[28] For a good example of this sort of inquiry, see Stefania Degaetano-Ortlieb and Elke Teich, “Toward an Optimal Code for Communication: The Case of Scientific English,” *Corpus Linguistics and Linguistic Theory* (2019). DOI: 10.1515/cllt-2018-0088.

[29] Maarten Sap, et al., “Recollection versus Imagination: Exploring Human Memory and Cognition via Neural Language Models,” Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, https://www.aclweb.org/anthology/2020.acl-main.178