Cultural cartography with word embeddings

Dustin S. Stoltz [a], Marshall A. Taylor [b],*

[a] *Department of Sociology and Anthropology, Lehigh University, Bethlehem, PA, USA*
[b] *Department of Sociology, New Mexico State University, Las Cruces, NM, USA*

A B S T R A C T

Using the frequency of keywords is a classic approach in the formal analysis of text, but has the drawback of glossing over the relationality of word meanings. Word embedding models overcome this problem by constructing a standardized and continuous "meaning space" where words are assigned a location based on relations of similarity to other words depending on how they are used in natural language samples. We show how word embeddings are commensurate with prevailing theories of meaning in sociology and can be put to the task of interpretation via two kinds of navigation. First, one can hold terms constant and measure how the embedding space moves around them—much like astronomers measured the changing of celestial bodies with the seasons. Second, one can also hold the embedding space constant and see how documents or authors move relative to it—just as ships use the stars on a given night to determine their location. Using the empirical case of immigration discourse in the United States, we demonstrate the merits of these two broad strategies for advancing important topics in cultural theory, including social marking, media fields, echo chambers, and cultural diffusion and change more broadly.

Meaning is often central to cultural analysis (Mohr et al., 2020, p. 2; Spillman, 2020, p. 1). As texts offer widely-available and unobtrusive sources of "meaning in the wild," formal text analysis has steadily produced a suite of tools for studying how meaning is articulated by individuals, groups, and organizations (Mohr, 1998; Mohr & Duquenne, 1997; Mohr & Lee, 2000). Many of these tools, however, are based on token counting, which is generally a poor fit with the relational theories of meaning they accompany.

The meaning of a text (e.g., a book, article, or social media comment) is, typically, established by the extent it references certain concepts or entities (Weber, 1984), often by *counting* certain words (Namenwirth & Weber, 2016; Weber, 1990). Lee and Martin (2014) refer to this process as *cultural cartography* in that, like a topographic map of terrain, it selectively simplifies texts in useful ways. The main problem with counting approaches is not that word order or subtlety is lost, or that certain words are selected as representative (cf. Breiger, Wagner-Pacifici, & Mohr, 2018; Mohr, Wagner-Pacifici, & Breiger, 2015). Indeed, "it is precisely because of their impoverishment that maps are useful" (Lee & Martin, 2014, p. 12). *Rather, procedures based on counting tokens are appropriate for discrete measures (*i.e., either-or*) but not graded measures (*i.e., more or less*).* The spread of populist ideologies across political campaigns, the prevalence of diversity rhetoric among management consultants, or the relative difference between labor discussions in Germany, Iceland, and the United Kingdom are all cases where graded measures are more appropriate than discrete measures for analyzing

meaning. Each case refers to generic ideas that can be more and less present or discourses that can be more and less alike. Concerns about the *magnitude* of conceptual engagement and similarity are central to cultural analysis; but count-based measures are ill-suited for measuring magnitudes.

The alternative we propose remains within the spirit of Lee and Martin's cultural cartography in that we aim to simplify texts in faithful ways while preserving the graded, relational meanings of words (Kirchner & Mohr, 2010; Mohr, 1998). Word embedding models offer a means to do just this. These models allow us to substitute the comparison of frequencies with the comparison of distances by providing standardized maps of meaning space in which semantic relations are realized as spatial relations. Importantly, these models do not "understand" meaning, nor do these procedures substitute for interpretation (Chakrabarti & Frye, 2017; Nelson, 2020; Popping, 2012).[1] Instead, they "condense information to facilitate an intersubjectively valid interpretation" (Lee & Martin, 2014, p. 1).

In what follows, we overview word embedding models as a means of mapping meaning space. We then organize the various applications of embeddings in cultural analysis into two broad kinds of "navigation." The first we refer to as *variable embedding space* methods, which involve holding terms constant to measure how the meaning space moves around them. Much like astronomers measured the changing relative locations of celestial bodies with the seasons, these methods measure the changing relative locations of words. More technically, these methods entail splitting a corpus by a covariate—e.g., time periods or authors—and training word embeddings on each subset of the corpus. Second, *fixed embedding space* methods use the same map of meaning space to measure how documents[2] or authors[3] change in relation to it. Here, just as ships determine their location relative to stars in the night sky, these methods measure the relative position of authors or documents *vis-à-vis* a single set of word embeddings. Documents or authors are represented as nebulae or clouds of locations, and measuring similarities to other documents, authors, words, or concepts becomes a transportation problem. We illustrate both approaches using the case of immigration discourse and its evolution in the United States.

1. Mapping meaning space

1.1. Language modeling and relational theories of meaning

There are two strategies for representing words as numbers: discrete and distributed. In the simplest discrete approach, each unique word is represented by a binary vector where each unique word in the corpus is assigned to a position in the vector (i.e., one-hot encoding). Comparing whether two terms (tokens) are the same entails measuring whether their term (type) vectors have a "1″ in the same position. Words are either precisely the same, or they are not.[4] The relationship between the numbers standing in for terms, however, do not correspond to the graded relationship between the meanings of terms (Smith, 2019, p. 2). *Two terms that mean similar things will be just as different as two terms that mean dissimilar things.* The second strategy overcomes this problem by creating "distributed" representations of terms' meanings. Word embedding models induce such "distributed" representations by summarizing the co-occurrences of words in natural language corpora (Lenci, 2018).

Chomsky famously asserted that corpus linguistics and the statistical study of text was a dead-end (Chomsky, 2002, pp. 15–20; R. Harris, 1995, pp. 96–98),[5] and yet the kernel of the "distributional hypothesis" (Sahlgren, 2008) is found in the work of Chomsky's dissertation advisor, Harris. Drawing on Bloomfield and Sapir in particular, Harris argued (1954, p. 156) "difference of meaning correlates with difference of distribution" (see also Joos, 1950).[6] Similarly, Firth—influenced by Malinowski (Firth, 1935; Rose, 1980; Young, 2011)[7]—stated that words' meanings can be inferred from their "habitual collocations":

> …a text in such established usage may contain sentences such as 'Don't be such an ass!', 'You silly ass!', 'What an ass he is!' In these examples, the word ass is in familiar and habitual company, commonly collocated with *you silly-, he is a silly-, don't be such an-*. You shall know a word by the company it keeps! One of the meanings of ass is its habitual collocation with such other words as those above quoted. (Firth, 1957, p. 11, emphasis in original)

---

[1] It is a category error to say word embedding models "understand" meaning (Bender & Koller, 2020; Glenberg & Mehta, 2009), and an example of the "symbol grounding" problem (Harnad, 1990; Lizardo, 2016).

[2] We use "document" to refer to any aggregate of text: books, articles, chapters, paragraphs, social media comments, etc.

[3] We use "author" to refer to individuals, groups, communities, or organizations responsible for producing texts.

[4] Each unique word (type) could be assigned an integer at random, where word (tokens) are the same if their assigned integers are precisely equal. The problem remains: words are either exactly the same, or not.

[5] Among many examples: "It must be recognized that the notion 'probability of a sentence' is an entirely useless one, under any known interpretation of this term" (Chomsky, 1969, p. 57).

[6] Saussure is often ritualistically cited to motivate such a "relational" approach to language, but his work largely ignored semantics to focus on phonology (Norris, 1985, p. 62; Stoltz, 2019).

[7] Along with Firth, Wittgenstein ([1953] 2009, pp. 80, 109) is often cited as positing a similar theory. He, like Firth, was likely influenced by Malinowski (Gellner, 1998, p. 149).

This basic hypothesis was supported by early statistical analyses (Henley, 1969; Rubenstein & Goodenough, 1965), leading Miller and Charles (1991, p. 24) to foreshadow word embeddings: "the general idea is to consolidate various kinds of information about a word's contexts into a single representation that characterises those contexts." The goal, then, is to assign each word a *single vector* such that the "gradedness in distributional representations correlates with gradedness in semantic phenomena" (Boleda, 2020, p. 228).[8]

Inferring a word's meaning by summarizing its (linguistic) context aligns with relational theories of meaning—pioneered by John Mohr (Kirchner & Mohr, 2010; Mohr, 1998, 2000), along with Pierre Bourdieu, Ron Breiger, Ann Mische, Harrison White, Viviana Zelizer and many others (Mische, 2011; Pachucki & Breiger, 2010; Zelizer, 2012)—as well as pragmatic, embodied, connectionist, and practice-theoretic approaches to cultural learning (Arseniev-Koehler & Foster, 2020; Ellis, 2019; Erk, 2016; Glenberg & Robertson, 2000; Ignatow, 2007, 2016; Landauer & Dumais, 1997; Lizardo, Sepulvado, Stoltz, & Taylor, 2019; Strauss & Quinn, 1997).

Accordingly, social scientists are beginning to use word embeddings (e.g., Boutyline, Arseniev-Koehler, & Cornell, 2020; Hofstra et al., 2020; Jones, Amin, Kim, & Skiena, 2020; Linzhuo, Lingfei, & Evans, 2020; Ornaghi, Ash, & Chen, 2019; van Loon & Freese, 2019). This early work largely builds on the fact that word embeddings mirror the stereotypical racial, ethnic, and gender related biases found in the texts they are trained on (Brunet, Alkalay-Houlihan, Anderson, & Zemel, 2018; Caliskan, Bryson, & Narayanan, 2017; Lewis & Lupyan, 2020).[9] While this can be a concern for some downstream applications—indeed, leading some researchers to attempt to "debias" word embeddings (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016a; Gonen & Goldberg, 2019)—this is a strength for those studying these associations as features of the social world. Rather than distortions in the semantic space, these associations reflect the contours of cultural formations. Some key terms relating to word embeddings and their definitions are presented in Table 1.

2. Word embeddings: Some technical details

2.1. The term-Co-occurrence matrix

The simplest distributional model begins by assigning each unique term in a corpus a vector. Each unique term is also a "position" in that vector indicating co-occurrences. Each entry in a term's vector is the frequency it occurs "next to" the term corresponding to that dimension within a given window, say five terms on either side.[10] The result is a target-term by context-term matrix, and the cells indicate the frequency a term (row) co-occurs with another term (column) within that window (see also Lee & Martin, 2014, pp. 15–16). This matrix is often called a *term-co-occurrence* or *term-context matrix* (TCM), and formalizes Firth's "habitual collocations" (1957, p. 11).

2.2. Dimension reduction

There are two drawbacks to using the TCM alone. First, using larger corpora, each term's vector would be long (as long as the number of unique terms in the corpus) and sparse (with many zeros indicating two words never co-occur). Second, and more importantly, two target terms are considered similar to the extent they share an exact matching set of co-occurring terms.

For both these reasons, dimension reduction is applied to the TCM (Lenci, 2018, p. 157; Wong, Ziarko, & Wong, 1985). The TCM is reduced not by selecting "columns" (i.e., feature selection), but rather by finding a few latent patterns summarizing the information in the matrix (i.e., feature extraction). This is motivated by the idea that "co-occurrences collected from corpora are noisy data that hide more abstract semantic structures" (Lenci, 2018, p. 157).

To summarize, at their most basic, word embedding models involve creating a TCM and reducing its dimensions. The n-dimensional TCM is "reduced" to at least n-1 latent dimensions and thus those original elements have been *embedded* into a "lower-dimensional" space. Advances in embedding models primarily improve the "tuning" of this low-dimensional mapping (see Appendix B). This notion forms the basis of "embedding" methods in computer science and is a mathematical feature common to many matrix factorization tools familiar to social scientists—e.g., factor analysis and correspondence analysis.[11] In comparison to other kinds of dimensional analysis, word embeddings have a relatively high number of dimensions because, in part, relational meanings of words are *intransitive* (word A can be similar to word B, and word C can be similar to word B, but this does not necessarily entail word C being similar to A). This

---

[8] It is important to emphasize words' meanings can be inferred from their linguistic contexts, and thus "difference of meaning correlates with difference of distribution" (Z. Harris, 1954, p. 156 emphasis added). This is contrary to problematic neo-Saussurean formulations in sociology (Stoltz, 2019), wherein words' meanings are said to be entirely constituted by their linguistic context alone. See Bender and Koller (2020, p. 7) for a similar critique of Wittgensteinian formulations: "the slogan 'meaning is use'... refers not to 'use' as 'distribution in a text corpus' but rather that language is used in the real world to convey communicative intents to real people."

[9] Although an open discussion, these models are found to accurately reflect co-occurrence patterns in corpora (Ethayarajh et al., 2019, p. 9), as opposed to "exaggerating" associations. In particular, researchers have validated these associations against traditional techniques, such as surveys and implicit association tests (Joseph & Morgan, 2020; Kozlowski et al., 2019).

[10] Goldberg (2016, pp. 367–369) provides a more detailed discussion: "the size of the sliding window has a strong effect on the resulting vector similarities. Larger windows tend to produce more topical similarities (i.e."dog", "bark" and "leash" will be grouped together, as well as "walked", "run" and "walking"), while smaller windows tend to produce more functional and syntactic similarities (i.e."Poodle", "Pitbull", "Rottweiler", or "walking", "running", "approaching")."

[11] Methods falling under the labels "topic modeling" and "latent semantic analysis" involve applying dimension reduction on a DTM (as opposed to the TCM). The dimensions of embedding matrices are not "interpreted" directly (cf. Bodell et al., 2019), as one might interpret "topics" or "principal components," but rather the relations between elements' locations as defined by those dimensions.

Table 1
Definitions.

Word Embeddings	Vector representations of words in an ($n < N$)-dimensional space (see Vector Space Model below), where N is the total number of words in a corpus and vectors are dense (i.e. no zero) and usually consisting of real numbers.
Document-Term Matrix (DTM)	A matrix where documents are rows and terms are columns (or vice versa), and a cell entry is a numerical representation of the jth word in the ith document.
Term Co-Occurrence/Context Matrix (TCM)	A symmetric N-by-N matrix, where N is the number of terms in a corpus, with cells numerically representing the extent the jth term tends to appear in the ith term's context window.
Relative Frequency	The raw count of the jth term in the ith document divided by that document's total word count.
Word Representation	A discrete number or vector of numbers standing in place of a unique string of characters.
Vector Space Model	A representation of words or documents as locations in a "meaning-space" in which semantic relations are understood as geometric relations.
Semantic Direction	A vector extracted from an embedding space by subtracting juxtaposing terms (or terms), pointing away from the term (or terms) subtracted.
Semantic Region	A vector corresponding to an area of the embedding space built by averaging several related terms (i.e. a semantic centroid) or clustering the embedding space (e.g., atom topic modeling).

necessitates higher dimensionality to prevent distortions.

2.3. From semantic relations to spatial relations

Next, and key for our argument, this matrix can be interpreted *geometrically*, such that the row vectors designate terms' locations in a continuous space (usually Euclidean). As geometric relations correspond to semantic relations, we can use operations from linear algebra to extract meaningful associations (Erk, 2012). Mainly, distance measures *semantic similarity* between two words, but also adding, subtracting, and averaging vectors produce intuitive results—sometimes referred to as relation induction, relation extraction, or semantic projection—useful for the cultural analyst (e.g., Arseniev-Koehler & Foster, 2020; Kozlowski, Taddy, & Evans, 2019; Nelson, 2021; Taylor & Stoltz, 2020). For example, subtracting "man" from "king" and adding "woman" produces a vector near "queen," or, adding "south" to "Africa" will be near the intersection of the two terms' co-occurrence contexts and thus surrounded by words related to "South Africa." While there are other kinds of relations one could extract (see the Discussion), we focus on similarity.

2.4. Pretrained embeddings and the selection of training corpora

Selecting corpora with which word embeddings (hereafter just *embeddings*) will be trained is an important consideration. Pretrained embeddings (see Table 2) are estimated using large-scale, widely-representative, "naturally occurring" corpora. For example, fastText embeddings are trained on several billion web pages from the Common Crawl (Joulin et al., 2016). More targeted pretrained embeddings also exist, such as those trained on the Corpus of Historical American English separately for each decade from the 1810s to the 2000s (Davies, 2012).

We contend (along with Spirling & Rodriguez, 2019) that researchers should *default* to using pretrained sources, and use corpus-trained if required by the research question.[12] The computational resources and time required to accurately train these models can be borne once (Lazer & Radford, 2017, p. 33), and pretrained embeddings enhance comparability across studies. There are, of course, limitations to pretrained embeddings (cf. Bender, Gebru, McMillan-Major, & Shmitchell, 2021). In particular, they may not be available for a specific context or language community, or use training corpora that systematically omit relevant populations.

Corpus-trained (or locally-trained) embeddings, by contrast, are word vectors trained exclusively on the researcher's corpus, or a broader corpus comprising the researcher's specific corpus (e.g., exploring the works of Marx using embeddings trained on all 19th century social science books). In general, a researcher will use pretrained embeddings if interested in analyzing how their documents relate to widely shared semantic associations within a given language community (i.e., "mainstream" culture). A researcher may want to train their own embeddings if they instead want to analyze any semantic idiosyncrasies within their targeted collection (e.g., Nelson, 2021), are attempting to tap into "counter-cultural" meanings, or if pretrained embeddings are not available for a specific language community, context, or population.

3. Navigating meaning space

Interpretation proceeds by measuring how units of analysis are related to each other by analogy to space. We can interpret, say, what it means for "immigration" to be a certain distance from "school" when we compare this to its distance from "family." That is, a *fixed waypoint* must be defined from which we gain our perspective on the relative distances of other points. Accordingly, we divide methods based on what unit of analysis is "fixed" (see Table 3).

In our first group of methods, the terms are fixed while the relations between them are allowed to vary. We refer to this as *variable embedding space methods*. We accomplish this by subsetting our corpus by a covariate, usually time or author, training multiple sets of embeddings and measuring the differences in the relative location of key terms within each of these spaces—much like early

---

[12] Furthermore, pretrained embeddings can often be "retrofitted" to highlight certain lexical relations (e.g., Faruqui et al., 2014).

Table 2

Example of commonly used pretrained embeddings.

	Corpora	Tokens	Vectors	Dimensions
fastText	Wikipedia 2017, UMBC webbase corpus, statmt.org news	16 billion	1 million	300
	Common Crawl	600 billion	2 million	300
GloVe	Wikipedia 2014, English Gigaword 5th Edition	6 billion	400 thousand	50 to 300
	Common Crawl	840 billion	2.2 million	300
	Twitter	27 billion	1.2 million	25 to 200
word2vec	Google News dataset	100 billion	3 million	300

Table 3

Uses of Variable and Fixed Embedding Spaces.

Embedding space	Points of reference Terms	Semantic directions or semantic regions	Documents or authors
Variable A corpus is subset by a covariate (typically by time or author), and embeddings are trained on each subset.	How does the distance between two (or more) terms change from one estimated embedding space to another?	How does the distance between terms and semantic direction or regions change from one estimated embedding space to another?	
Fixed An embedding is trained on a single corpus. Authors or documents are defined as the aggregate of locations of the words associated with them.	How does the distance between documents (or authors) and terms in a single embedding space differ by document covariates?	How does the distance between documents (or authors) and semantic directions or regions in a single embedding space differ by document covariates?	How does the distance between documents (or authors) in a single embedding space differ by document covariates?

astronomers measured how the positions of celestial bodies changed across the seasons. In our second group—which we call *fixed embedding space methods*—the embedding space is held constant while measuring how documents or authors differ in relation to each other. We accomplish this by measuring the relative locations of documents or authors defined as aggregates of terms in a single set of embeddings—just as ships use the stars at a given time to determine their location.

3.1. Variable embedding space

Kulkarni, Al-Rfou, Perozzi, and Skiena (2015) offered one of the first studies to fix terms and compare their changing relative position in embeddings trained on text from different time periods. They empirically demonstrate the well-known discursive shift over the 20th century where the word "gay" changed from being located beside "cheerful" and "frolicsome" to being near "lesbian" and "bisexual." Similarly, Garg, Schiebinger, Jurafsky, and Zou (2018) track changes in gender and ethnic biases in English over the same time period by comparing the changing distances between gender- and ethnicity-related terms and a list of adjectives and occupational terms (see also Jones et al., 2020; Kozlowski et al., 2019). This involves a "time-lapse" approach, training on texts from different time periods and measuring changes in the meaning space between periods.

We could subset a corpus by variables other than time, and train separate models on each subset to measure how the meaning of terms varies across, e.g., individuals, communities, or organizations. For example, Bonikowski, Feinstein, and Bock (2019) divide presidential candidates' speeches by individual candidates to find the 50 nearest neighbors of two focal terms derived from each candidates' embeddings. As each model is trained on separate candidates, the associations are specific to the candidate. Similarly, Zannettou, Finkelstein, Bradlyn, and Blackburn (2018) compare models trained on text from different communities: the 4-chan board /pol/ and Gab (see also An, Kwak, Posegga, & Jungherr, 2019; Rho, Mark, & Mazmanian, 2018; Schild et al., 2020).

There is a caveat to comparing embeddings trained on separate corpora. Matrices may need to be "aligned" (Hamilton, Leskovec, & Jurafsky, 2016, p. 4; Kulkarni et al., 2015, p. 4) because the random processes involved render non-unique solutions even when training on the same corpora with the same methods (cf. Orlikowski, Hartung, & Cimiano, 2018; Tahmasebi, Borin, & Jatowt, 2018). Simply put, alignment involves rotating and scaling of two or more matrices, preserving the distances between terms within each embedding while approximately aligning the terms across embeddings (Artetxe, Labaka, & Agirre, 2016; Ruder, Vulić, & Søgaard, 2019).

3.1.1. Measuring the distance between terms over time

To demonstrate variable embedding methods, we first show the simplest use, which is measuring how the distance between two (or more) words differs over time. Specifically, we measure how the meaning of "immigration" has shifted in American English using embeddings trained on the Corpus of Historical American English for each decade from 1880[13] to 2000 (Davies, 2012; Hamilton et al., 2016).[14]

---

[13] The first valid word vector for "immigration" is in 1880 and thus this is our starting point.
[14] https://nlp.stanford.edu/projects/histwords/

Here, we take the cosine similarity between the vector of "immigration," on one hand, and the vectors for "job," "crime," "family," and "school," on the other, for each decade. In the plot (see Fig. 1), immigration was strongly associated with "job" at the end of the 19th century, and while "school," "family," and "job" slowly increased in the 20th century, immigration has grown even more associated with "crime." While "immigration" had its lowest observed cosine similarity with "crime" in the 1890s (0.076, on a scale of -1 to 1, where larger positive values indicate more similarities in discursive context), it reached its peak similarity in the 1990s (0.344), with the nearest other similarity with "immigration" in that decade being school (0.087). These temporal trends reflect a change in the meaning space surrounding immigration in American English over nearly one and half centuries. Culture researchers might then assess the extent to which variance in these discursive shifts for a population over time are explained by period- or cohort-level effects—a linguistic extension of the cultural change research carried out mostly using survey data (cf. Hamilton et al., 2016; Kiley & Vaisey, 2020; Vaisey & Lizardo, 2016).

3.1.2. Measuring the distance between terms and cultural dimensions

Next, we extract a semantic direction pointing toward a pole of the cultural dimensions of race, social class, and morality. This involves identifying sets of juxtaposing terms for each cultural dimension—understood as generic binary oppositions that "individuals use in everyday life to classify agents and objects in the world" (Kozlowski et al., 2019, p. 911). For example, for social class, this would be affluence vs poverty, rich vs poor, and so on. The vector for "poverty" is subtracted from "affluence." This is repeated through the set of class-related pairs, averaging the results. The final vector represents one pole of the cultural dimension of affluence.[15] The affluence, race, and morality dimensions were constructed using term pairs taken from Kozlowski et al. (2019), pp. 935–937.[16]

We measure the changing position of the term "immigrant" as well as "citizen" relative to these cultural dimensions in each decade. The *y*-axis of Fig. 2 (both panels) shows that, regardless of decade, "citizens" (yellow triangles) tend to be closer to the "white" pole of the race dimension, whereas "immigrants" (purple dots) tend to be closer to the "black" pole. In the *x*-axis of the left panel, we see that "citizens" in 2000 is closer to the "high class" pole of the affluence dimension than is "immigrants" in 2000. In the *x*-axis of the right panel, we see that "citizen" has clustered near the "good" pole of the morality dimension throughout the decades, whereas "immigrants" have generally moved closer to "good" over the decades, with 1980, 1990, and 2000 being the closest to "good."

These associations could be interpreted in one of two (non-mutually exclusive) ways. First, engagement with the "citizen," "white," "high class," and "good" poles of their respective cultural dimensions may be correlated because—unlike "immigrant," "black," "low class," and "bad"—these concepts have been largely *absent* in U.S. public discourse over time. This interpretation suggests that these "discursively absent" poles have been remarkably stable *unmarked* (and therefore taken-for-granted and normative) categories for citizenship, race, social class, and morality discourses in the United States (Brekhus, 1998).[17] The second interpretation is that the temporal stability in how the immigrant-citizen dimension correlates with these other dimensions reflects durable symbolic boundaries in the U.S. This corroborates prior research finding the "immigrant-as-nonwhite"[18] vs. "citizen-as-white" boundary has been an enduring cultural structure in U.S. public discourse and attitudes (Mora & Paschel, 2020; Sáenz & Douglas, 2015).

3.2. Fixed embedding space

A straightforward use of a fixed embedding space is measuring the relationship between key terms or cultural dimensions. Arseniev-Koehler and Foster (2020), for example, train a model on over a hundred thousand *New York Times* articles and measure the distance between terms related to "obesity" and key cultural dimensions: gender, morality, health, and socioeconomic status (see also Nelson, 2021). An analyst can also use fixed embeddings to compare the relations between documents or authors, in particular determining the general *semantic similarity* between documents or authors. These similarities can be used for a variety of ends such as comparing document revisions, text classification, measuring content change, building semantic networks, or studying content diffusion (e.g., Ahlgren & Colliander, 2009; Berry & Taylor, 2017; Strang & Dokshin, 2019; Teplitskiy, 2016; Zhang & Pan, 2019).

3.2.1. Document-by-document distances

Prior social scientific research estimating document similarities often relied on discrete word representations. For example, Bail (2012) used plagiarism detection software (Bloomfield, 2008) to compare press releases by civil society organizations to media coverage. Similarly, Grimmer (2010) used the same software to compare press releases from Senate offices to media coverage. This software, after lemmatizing two documents, searches for exact matches in strings of six words. This technique, then, treats words as either the same or not, therefore not accounting for the graded, relational meaning of words.

---

[15] There are several procedures for deriving a "semantic direction" from an embedding space (Arseniev-Koehler & Foster, 2020, pp. 18–9; Bolukbasi et al., 2016b, pp. 42–3; Boutyline et al., 2020; Ethayarajh et al., 2019; Kozlowski et al., 2019, p. 943 fn 8; Larsen et al., 2015, p. 5; Taylor & Stoltz, 2020).

[16] The embeddings did not have the following for all time periods: advantaged, propertied, sumptuous, swanky, ritzy, uncorrupt, pureness, necessitous; skint, penurious, unmonied, unprosperous, moneyless, transgressive, knavish, afro.

[17] A more in-depth analysis with the "marked-unmarked" framework might benefit more from a keyword absence/presence method, since a category's "unmarked" status is, by definition, signified by its absence.

[18] A more systematic analysis would either construct a "nonwhite-white" semantic direction (as opposed to "black-white"), or construct several semantic directions with "white" as one pole and as series of non-white racial/ethnic categories as the opposite poles. These steps would be necessary to more confidently link this finding to the "immigrant-as-nonwhite" symbolic boundary interpretation.
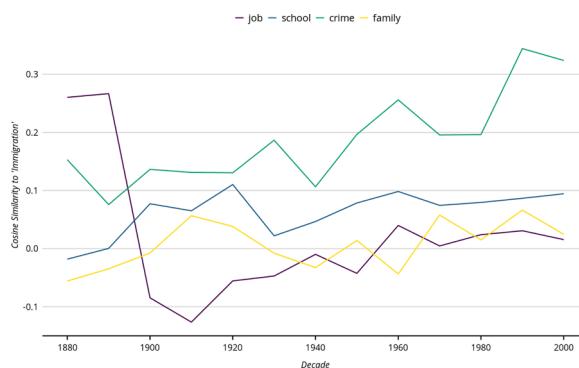
Fig. 1. Cosine Similarity of 'Immigration' and Key Terms by Decade, 1880 to 2000.
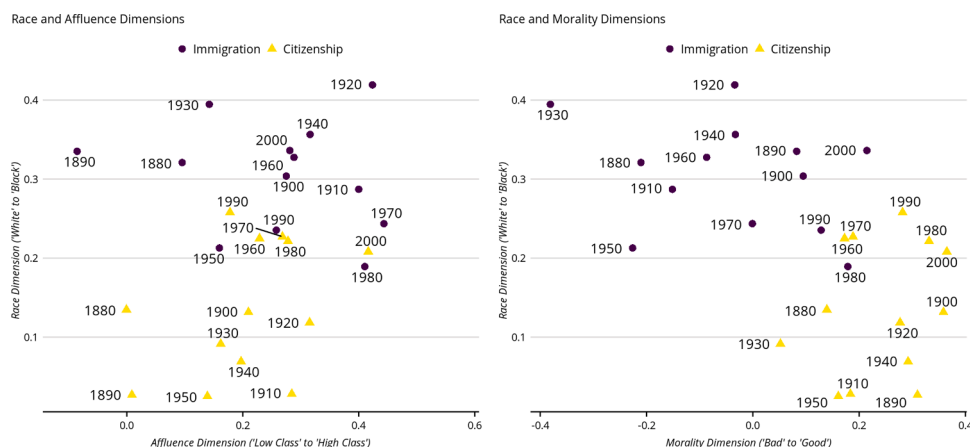

Fig. 2. 'Immigrant' and 'Citizen' on Key Cultural Dimensions, 1880 to 2000.

Farrell's (2016) method is even closer in spirit to the fixed embeddings approach. To compare the similarity of writings by climate contrarian organizations with news outlets and political offices, Farrell reduced the dimensionality of the DTM by applying SVD, and comparing the cosine similarity of the resulting document vectors (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). While this formed the backbone of information retrieval for decades, recent research demonstrates that similarity measures using embeddings, specifically Word Mover's Distance (Kusner, Sun, Kolkin, & Weinberger, 2015),[19] outperform this procedure on various tests, including plagiarism detection (Tashu & Horváth, 2018).

With Word Mover's Distance (WMD), documents are represented not just by counts of unique terms, but also the semantic relations between those terms. A document becomes a cloud of locations in the embedding space. Determining similarity is treated as a transportation problem where the "cost" of moving one document's cloud of locations to another is equivalent to the semantic similarity between two documents. The DTM provides the "amounts" to be moved and the embedding matrix provides the "distances" that these amounts are moved. The result is a document-by-document similarity matrix, or an author-by-author similarity matrix if similarities are averaged or each document has a single author: for example, Pomeroy, Dasandi, and Mikhaylov (2019) use this to measure the relationship between nation-states based on the similarity of their United Nations speeches.

To demonstrate WMD (using LC-RWMD[20]), we find the document-by-document similarity matrix between a preprocessed (see Appendix A) corpus of nearly 200,000 U.S. news articles[21] and a preprocessed corpus of press releases published by immigration-focus

---

[19] One alternative to WMD, and related algorithms, involves taking the average vector of all the words in a document (sometimes called Word Centroid Distance), and comparing these resulting vectors using cosine similarity (Berry & Taylor, 2017; Lix et al., 2020), which is likely well suited for smaller documents, such as social media comments.

[20] Transportation problem uses Earth's Mover's Distance (Rubner et al., 1998) to compare multidimensional distributions. Many have found computationally efficient solutions (Atasu et al., 2017; Tithi & Petrini, 2020; Werner & Laber, 2019; Wu et al., 2018). Our method incorporates one: Linear Complexity Relaxed Word Mover's Distance (LC-RWMD).

[21] "All the News" corpus is a collection of 204,135 news articles from 18 U.S. news organizations, mostly from 2013 to early 2018 (Thompson, 2018).

advocacy organizations.[22] We collected a total of 976 press releases: from two far-right organizations, the Center for Immigration Studies ($N = 155$) and the Federation for American Immigration Reform ($N = 377$), as well as two left-wing organizations, National Network for Immigrant and Refugee Rights ($N = 117$) and Women's Refugee Commission ($N = 327$). The publication date of the press releases ranged from 1998 to 2020 (median of 2015), but they were "pooled" together and compared to news articles across time points ranging from January 2016 to June 2017. We select three news organizations considered more right-leaning (*Breitbart*, Fox News, and *National Review*) and three more left-leaning (*Talking Points Memo, New York Times*, and *Buzzfeed News*) for our demonstration.

We find each news organization's similarity to each advocacy organization by the average of the similarity between their respective news articles per year and the press releases pooled by advocacy organization. We show similarities for all articles during this time frame, and also a subset of articles referencing "immigration" or "immigrants" at least once during this time frame ($N = 58,758$ for the full corpus and $N = 7,991$ for the subset), and plotted their average similarities to the right-wing (red) and left-wing (blue) press releases over time (see Fig. 3).

For left-leaning news organizations, similarity between articles and press releases increases from mid-2016 and peaks in early 2017. When right-leaning news organizations write about "immigration" they are significantly more similar to the press releases from right-wing advocacy organizations than from left-wing advocacy organizations. However, when left-leaning news outlets write about immigration, they are equally similar to both left- and right-wing advocacy organizations, and are also more similar to press releases from either side of the political spectrum, save for early 2016.

This finding aligns with work on echo chambers and political polarization in U.S. society—an important area of contemporary cultural sociology—showing liberals are more likely than conservatives to be comfortable with ambiguity and change, while conservatives shifted more rightward than their liberal counterparts shifted leftward (Bail et al., 2018; Jost et al., 2007). Specifically, this is evidence that asymmetric polarization (Grossmann & Hopkins, 2016) might be occurring within immigration discourse. If the temporal ordering of these data were such that the press releases were consistently published before the sampled news articles, this question of asymmetric polarization could also be framed as a question of cultural diffusion (Bail, 2016).

3.2.2. Document-by-concept distances

3.2.2.1. Measuring conceptual engagement. Concept Mover's Distance (CMD) (Stoltz & Taylor, 2019; Taylor & Stoltz, 2020) quantifies the extent a document "engages with" a theoretically-motivated "focal" concept. To do so, we hold the embedding space constant while measuring the position of documents relative to it, specifically measuring each document's distance from a specific "region" of the embedding space. This can be used to explore, for example, the linear association between the concept of "death" and actual body counts in Shakespeare's First Folio (Stoltz & Taylor, 2019). Like WMD, CMD relies on a "transportation problem" logic to measure the distance between each document and at least one "pseudo-document."[23]

In the simplest case, this pseudo-document consists of a single word associated with a single vector in the embeddings. Measuring engagement with more "specific" concepts—e.g., "liberal politics" or "conservative politics" instead of simply "politics"—is accomplished by adding relevant terms to the pseudo-document to create *compound concepts*: for example, by using a pseudo-document consisting of "politics" *and* "liberal" to measure "liberal politics."

Consider again the "All the News" corpus. We use CMD (here with pretrained embeddings), to measure the extent all news articles from January 2012 to March 2018 engaged the concept "immigration." We also measured engagement with the following specified concepts: "immigration + job," "immigration + school," "immigration + crime," and "immigration + family."

Fig. 4 shows the smoothed engagement time trends for each concept, averaged by month. According to the plot, engagement with "immigration" and its related compound concepts have followed a similar time trend from 2012 to mid-2018. The "immigration + school" compound concept peaked in late-2014 to mid-2015, around the time the Obama administration's planned extensions to the Deferred Action for Childhood Arrivals (DACA) policy (school requirements were a key element for program eligibility) in late 2014 that were in court battles throughout 2015 and 2016. One such planned extension was the Deferred Action for Parents of Americans and Lawful Permanent Residents (DAPA), aligning with the spike for "immigration + family" during the same time. Whatever the mechanism, this spike in engagement suggests immigration within the contexts of schooling and family seemed to be prevalent modes of media immigration discourse in late 2014 to 2015. Media immigration discourse of all stripes rises in mid-2016 to early-2017—coinciding with Donald Trump winning the Republican primary and the U.S. presidency.

Fig. 5 plots just the monthly average engagement with "immigration" (i.e., the dark purple trend line in Fig. 4, with a lower smoothing factor). The figure shows that media engagement rises predictably with relevant sociopolitical events: e.g., introduction DACA policy, the DACA expansion, and Trump's "immigrants-as-rapists" presidential running announcement speech on June 15, 2015. This predictable finding is in line with work on heteronomy in media fields, where fields of journalistic production are known to be highly responsive to shifts and pressures from political (and economic) fields (Benson, 1999; Bourdieu, 2005).

3.2.2.2. Measuring engagement with binary concepts. Opposing words—i.e. binary concepts—are likely to be near one another within

---

[22] We collected the press releases by building a scraper using the "rvest" R package (Wickham, 2019).
[23] One could measure distance to a larger semantic region by averaging several word vectors forming a "semantic centroid," or using atom topic modeling (Arora et al., 2018; Arseniev-Koehler et al., 2020). In both cases the resulting vector will be in the same dimensional space as the embeddings and can easily be used with the CMD/WMD solver.

Breitbart, Fox News, National Review

Talking Points Memo, New York Times, Buzzfeed News
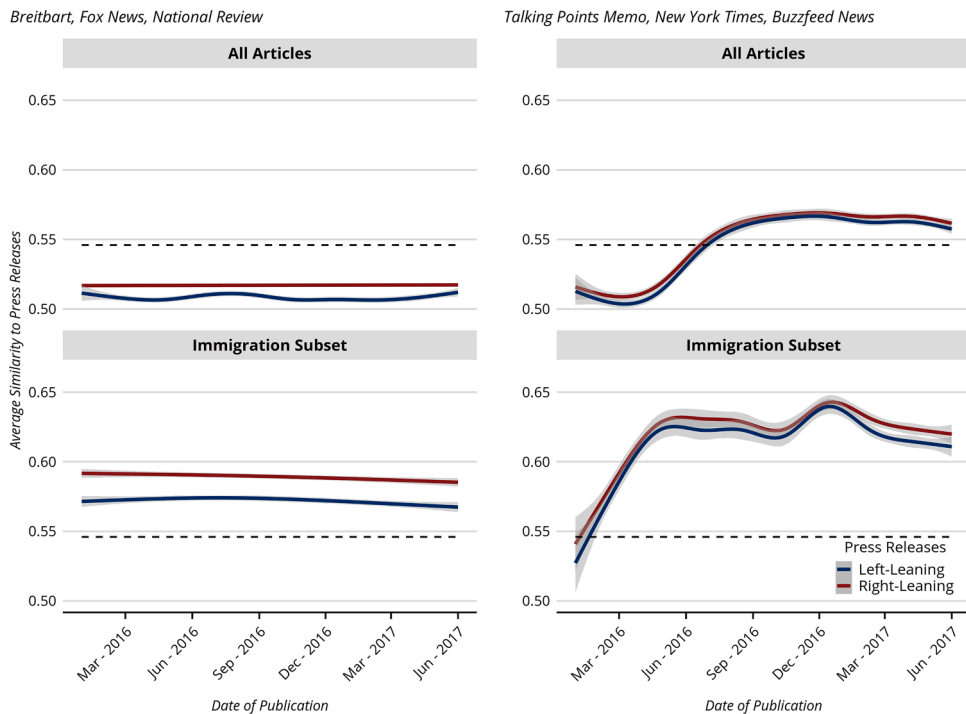
Fig. 3. News Articles' Similarity to Press Releases (with WMD). *Note*: Top panels include all articles in the dataset. Bottom panels include only articles referencing "immigration" or "immigrants" at least once. The dashed line represents the mean similarity to the press releases for all news articles, irrespective of political lean.
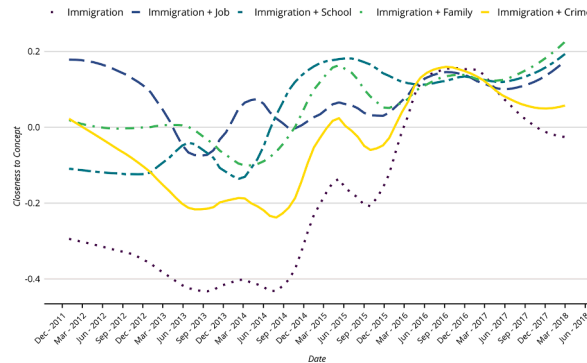
Fig. 4. News Articles' Conceptual Engagement Over Time (with CMD).

the embeddings space because they are used in similar contexts and they often co-occur (Deese, 1966; Justeson & Katz, 1991; Miller & Charles, 1991, pp. 25–6). As such, oppositions such as "good" and "evil" occupy similar positions in any adequately-trained corpus because they are used in similar ways and mutually oriented toward a shared meaning (Boutyline, 2017; A. Goldberg, 2011). For example, a researcher may be interested in examining engagement with "evil," but a document that engages strictly with "good" would still be highly ranked because the distance of any given word in that document is roughly equidistant to "good" and "evil" within the embeddings space.

To address this, we can use "semantic directions" as described previously (Taylor & Stoltz, 2020). An analyst can measure engagement with one "pole" of a binary by (1) extracting a direction in the meaning space pointing toward a pole of the binary opposition, (2) adding the estimate as a row vector to the embeddings matrix, and (3) adding a pseudo-document to the corpus consisting of only a single reference to that estimated vector. From there, using CMD is the same: it quantifies the cost to move all words in an observed document to the estimated vector—in effect, measuring engagement with one pole of a binary concept as opposed to the other pole (e.g., "evil" *as opposed to* "good").

Consider again the "All the News" corpus. Following the same procedure outlined previously, we constructed two cultural dimensions: an "immigrant" dimension and a "race" dimension, with larger positive values indicating more engagement with the
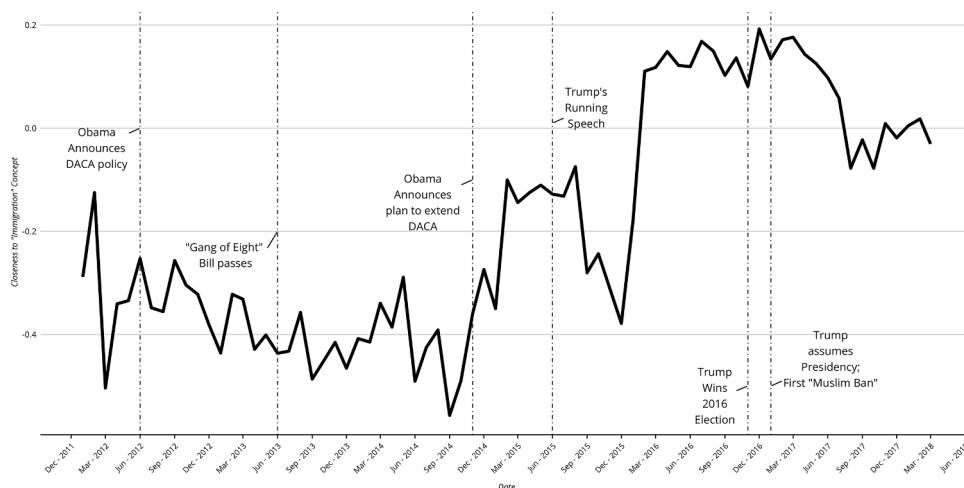
Fig. 5. News Articles' Conceptual Engagement and Key Events (with CMD).

"immigrant" pole as opposed to a "citizen" pole and more engagement with the "black" pole as opposed to the "white" pole, respectively. The immigrant-citizen dimension was constructed using the pairs listed in Table 4, and is understood as measuring a persistent symbolic structure rather than a legal distinction (cf. Beaman, 2016; Jaworsky, 2013).[24]

Smoothed media engagement time trends with the respective poles of these cultural dimensions (averaged by month-year) are shown in Fig. 6 top panel; the bottom panel shows how much each month-year's average engagement deviates from the previous month-year's engagement. Engagement with immigration and race cultural dimensions in U.S. news media (specifically, between March 2013 and September 2017) appears to follow similar trends: when news media engage "immigrant" more relative to "citizen" in their discourse, so too do they generally engage "black" more relative "white." This trend is also in line with the above analysis using historical embeddings from 1880 to 2000, again pointing to the remarkable temporal stability of "citizen" and "white" as the unmarked social categories of U.S. public discourses on cultural citizenship and race (Brekhus, 1998) and/or the historical persistence of the "immigrant-as-nonwhite" vs. "citizen-as-white" symbolic boundary (Mora & Paschel, 2020; Sáenz & Douglas, 2015).

4. Extensions

4.1. Synthesized embedding spaces

While distinct, variable and fixed embedding space methods could be combined—what one might call *synthesized embedding space* methods. Synthesized approaches would involve simultaneously assessing the changing structure of embedding spaces as a function of some external variable—most likely time—*and also* how documents and authors relate to one another or to a concept vis-à-vis some fixed comparable points across the embedding spaces. For example, if we had a sample of English-language news articles from the past century, we could measure how news organizations spanning several decades engage with a concept or cultural dimension using decade-specific embeddings.

4.2. Beyond "Similarity"

Our demonstration, as well as much current work using embeddings, focuses on semantic *similarity*, which allows us to explore relations of polysemy, synonymy, antonymy, as well as metaphor and metonymy (Erk, 2012). However, there are other relations to be explored: specifically, scale (e.g., good, greater, greatest), classification (i.e., entailment and part-whole relations like hypernymy, hyponymy, meronymy, or holonymy) or object qualities (Fulda, Ricks, Murdoch, & Wingate, 2017; Grand, Blank, Pereira, & Fedorenko, 2018). Most pretrained embeddings already encode these relations, to some extent, and can be extracted with simple post-processing (cf. Fulda et al., 2017; Levy, Remus, Biemann, & Dagan, 2015).

For example, one can use "Hearst patterns" (Hearst, 1992)—"such as," "like," "including"—to extract hypernymy/hyponymy (Baroni, Bernardi, Do, & Shan, 2012; Roller & Erk, 2016). Similarly, Fu et al. (2014) propose a post-processing method for learning a linear projection for hierarchical relations using known hypernym/hyponym pairs (see also Roller, Erk, & Boleda, 2014). Other approaches modify the training phase to better incorporate hierarchical information. For example, Le, Roller, Papaxanthos, Kiela, and

---

[24] See footnote 22. While we use "citizen" as the pole opposite "immigrant" to highlight the symbolic structure through which immigrants have been discursively othered in the U.S., citizen is not an antonym for immigrant in any formal way since, obviously, an immigrant can certainly be a legal citizen.

Table 4
Term pairs for immigration-citizenship cultural dimension.

Immigrants	Citizens
Immigration	Citizenship
Immigrant	Citizen
Foreign	Domestic
Foreigner	Native
Outsider	Insider
Stranger	Local
Alien	Resident
Foreigner	Resident
Alien	Native
Immigrant	Local
Foreign	Familiar
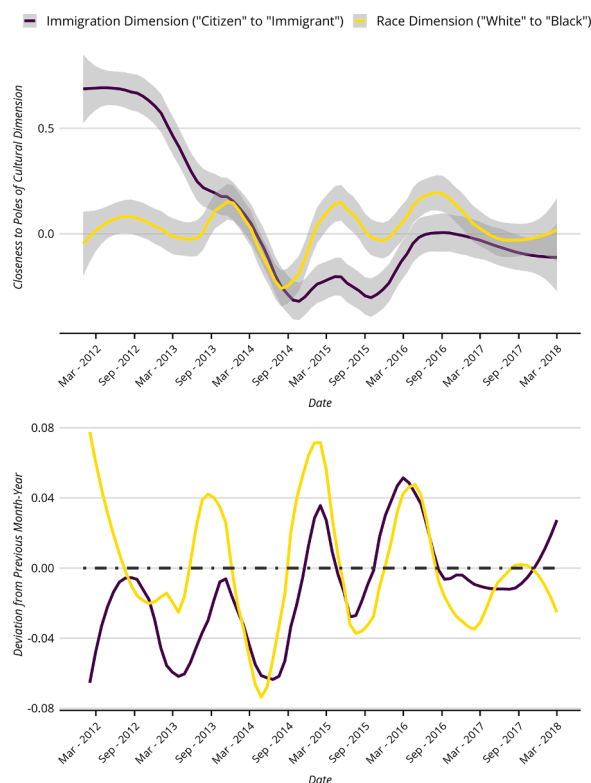
Fig. 6. News Articles' Engagement with Key Cultural Dimensions (with CMD).

Nickel (2019) use "hyperbolic" space (Nickel & Kiela, 2017) rather than Euclidean space (see also Kruszewski, Paperno, & Baroni, 2015; Weeds, Clarke, Reffin, Weir, & Keller, 2014). Hyperbolic embeddings with Hearst patterns (Le et al., 2019) might be particularly useful for scholars of culture working with conceptual metaphor theory (Hart-Brinson, 2016; Kharchenkova, 2018; Rotolo, 2020), since models designed to infer transitive hyponym-hypernym relations (e.g., inferring that "conifer *is a* plant" by way of including "conifer *is a* tree" and "tree *is a* plant" in the input Hearst matrix) might also be used to capture the source-target structures behind metaphorical reasoning (e.g., "country *is a* family" (Lakoff, 2010)) that are rarely explicit in written language.

4.3. Concluding Remarks

Formal text analysis in sociology was pioneered by scholars like John Mohr, Wendy Griswold, Karen Cerulo, and Kathleen Carley, and even as it continues to evolve with unprecedented access to large quantities of texts and computational power, the methods of this early era continue to be productive. Although we note limitations, this is not to say count based methods are no longer useful. Rather, embedding methods should be used alongside, as a welcome addition to the cultural analyst's toolbox.

The variety of options available to cultural analysts interested in computational text analysis signals an exciting time. From word counts and dictionary methods to embeddings and other unsupervised learning algorithms, computational text analysis is quickly

becoming as institutionalized as ethnography, interviews, and historical-comparative methods. Looking back, it is quite fascinating to see that John Mohr saw this advent coming with such clarity as far back as at least 1998:

> . . . [I]t is probably worth pointing out that we are just now entering what must surely be the golden age of textual analysis. What sets this moment in history apart is the incredible proliferation of on-line and on-disk textual materials. Previously, scholars who were interested in doing some form of content analysis were compelled to spend huge amounts of time readying their texts for analysis. Now one can easily sit at one's desk and more or less instantaneously summon up a fantastic array of cultural texts in electronic form. (Mohr, 1998, p. 366)

We are certainly amid the golden age of text analysis in the social sciences—just as John predicted and helped ensure.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.poetic.2021.101567.

References

Ahlgren, P., & Colliander, C. (2009). Document–document similarity approaches and science mapping. *Journal of Informetrics, 3*(1), 49–63.

An, J., Kwak, H., Posegga, O., & Jungherr, A. (2019). Political discussions in homogeneous and cross-cutting communication spaces. *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 68–79.

Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2018). Linear algebraic structure of word senses, with applications to polysemy. *Transactions of the Association for Computational Linguistics, 6*, 483–495.

Arseniev-Koehler, A., Cochran, S., Mays, V., Chang, K.-W., & Foster, J. G. (2020). Discourses of death. 10.31235/osf.io/nkyaq.

Arseniev-Koehler, A., & Foster, J. (2020). Machine learning as a model for cultural learning. In arXiv [cs.CY]. arXiv. http://arxiv.org/abs/2003.12133.

Artetxe, M., Labaka, G., & Agirre, E. (2016). Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2289–2294.

Atasu, K., Parnell, T., Dunner, C., Sifalakis, M., Pozidis, H., Vasileiadis, V., Vlachos, M., Berrospi, C., & Labbi, A. (2017). Linear-complexity relaxed word Mover's distance with GPU acceleration. In 2017 IEEE International Conference on Big Data. 10.1109/bigdata.2017.8258005.

Bail, C. A. (2016). *Terrified*. Princeton University Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences of the United States of America, 115*(37), 9216–9221.

Bail, C. A. (2012). The fringe effect civil society organizations and the evolution of media discourse about Islam since the September 11th attacks. *American Sociological Review, 77*(6), 855–879.

Baroni, M., Bernardi, R., Do, N.-Q., & Shan, C.-C. (2012). Entailment above the word level in distributional semantics. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 23–32.

Beaman, J. (2016). Citizenship as cultural. *Sociology Compass, 10*(10), 849–857.

Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Bender, E., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. https://openreview.net/forum?id=GKTvAcb12b.

Benson, R. (1999). Field theory in comparative context. *Theory and Society, 28*(3), 463–498.

Berry, G., & Taylor, S. (2017). Discussion quality diffuses in the digital public square. Proceedings of the 26th International Conference on World Wide Web. https://dl.acm.org/ 10.1145/3038912.3052666.

Bloomfield, L. (2008). WCopyFind. http://plagiarism.phys.virginia.edu/Wsoftware.html.

Bodell, M. H., Arvidsson, M., & Magnusson, M. (2019). Interpretable word embeddings via informative priors. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1909.01459.

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Applied Linguistics*. 10.1146/annurev-linguistics-011619-030303.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016a). Man is to computer programmer as woman is to homemaker?. *Advances in Neural Information Processing Systems* (pp. 4349–4357). Curran Associates.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016b). Quantifying and reducing stereotypes in word embeddings. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1606.06121.

Bonikowski, B., Feinstein, Y., & Bock, S. (2019). The polarization of nationalist cleavages and the 2016 US presidential election. https://osf.io/pmg95/download.

Bourdieu, P. (2005). The political field, the social science field, and the journalistic field. *Bourdieu and the Journalistic Field, 29*, 47.

Boutyline, A. (2017). Improving the measurement of shared cultural schemas with correlational class analysis. In Sociological Science (Vol. 4, pp. 353–393). 10.15195/v4.a15.

Boutyline, A., Arseniev-Koehler, A., & Cornell, D. (2020). School, studying, and smarts. https://osf.io/preprints/socarxiv/bukdg.

Breiger, R. L., Wagner-Pacifici, R., & Mohr, J. (2018). Capturing distinctions while mining text data. *Poetics, 68*, 104–119.

Brekhus, W. (1998). A sociology of the unmarked. *Sociological Theory, 16*(1).

Brunet, M.-E., Alkalay-Houlihan, C., Anderson, A., & Zemel, R. (2018). Understanding the origins of bias in word embeddings. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1810.03611.

Caliskan, A., Bryson, J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186.

Chakrabarti, P., & Frye, M. (2017). A mixed-methods framework for analyzing text data. *Demographic Research, 37*, 1351–1382.

Chomsky, N. (1969). Quine's empirical assumptions. *Words and Objections* (pp. 53–68). Springer.

Chomsky, N. (2002). *Syntactic Structures*. Walter de Gruyter.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora, 7*(2), 121–157.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

Deese, J. (1966). *The Structure of Associations in Language and Thought*. Johns Hopkins University Press.

Ellis, N. (2019). Essentials of a theory of language cognition. *Modern Language Journal, 103*, 39–60.

Erk, K. (2012). Vector space models of word meaning and phrase meaning. *Language and Linguistics Compass*. https://onlinelibrary.wiley.com/ 10.1002/lnco.362.

Erk, K. (2016). What do you know about an alligator when you know the company it keeps? *Semantics and Pragmatics, 9*, 11–17.

Ethayarajh, K., Duvenaud, D., & Hirst, G. (2019). Understanding undesirable word embedding associations. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1908.06361.

Farrell, J. (2016). Network structure and influence of the climate change counter-movement. *Nature Climate Change, 6*(4), 370–374.

Faruqui, M., Dodge, J., Jauhar, S., Dyer, C., Hovy, E., & Smith, N. (2014). Retrofitting word vectors to semantic lexicons. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1411.4166.

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society, 34*(1), 36–73.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

Fu, R., Guo, J., Qin, B., Che, W., Wang, H., & Liu, T. (2014). Learning semantic hierarchies via word embeddings. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1199–1209.

Fulda, N., Ricks, D., Murdoch, B., & Wingate, D. (2017). What can you do with a rock? In arXiv [cs.AI]. arXiv. http://arxiv.org/abs/1703.03429.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America, 115*(16), E3635–E3644.

Gellner, E. (1998). *Language and solitude*. Cambridge University Press.

Glenberg, A., & Mehta, S. (2009). Constraints on covariation. *Italian Journal of Linguistics*.

Glenberg, A., & Robertson, D. (2000). Symbol grounding and meaning. *Journal of Memory and Language, 43*(3), 379–401.

Goldberg, A. (2011). Mapping shared understandings using relational class analysis. *The American Journal of Sociology, 116*(5), 1397–1436.

Goldberg, Y. (2016). A primer on neural network models for natural language processing. *The Journal of Artificial Intelligence Research, 57*, 345–420.

Gonen, H., & Goldberg, Y. (2019). Lipstick on a Pig. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1903.03862.

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2018). Semantic projection. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1802.01241.

Grimmer, J. (2010). A bayesian hierarchical topic model for political texts. *Political Analysis, 18*(1), 1–35.

Grossmann, M., & Hopkins, D. A. (2016). *Asymmetric Politics*. Oxford University Press.

Hamilton, W., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1605.09096.

Harnad, S. (1990). The symbol grounding problem. *Physica D. Nonlinear Phenomena, 42*(1), 335–346.

Harris, R. (1995). *The Linguistics Wars*. Oxford University Press.

Harris, Z. (1954). *Distributional Structure. Word, 10*(2-3), 146–162.

Hart-Brinson, P. (2016). The Social imagination of homosexuality and the rise of same-sex marriage in the United States. *Socius, 2*, Article 2378023116630555.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. Proceedings of the 14th Conference on Computational Linguistics, 539–545.

Henley, N. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 176–184.

Hofstra, B., Kulkarni, V. V., Galvez, S. M.-N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–Innovation paradox in science. *Proceedings of the National Academy of Sciences of the United States of America, 117*(17), 9284–9291.

Ignatow, G. (2007). Theories of embodied knowledge. *Journal for the Theory of Social Behaviour, 37*(2), 115–135.

Ignatow, G. (2016). Theoretical foundations for digital text analysis. *Journal for the Theory of Social Behaviour, 46*(1), 104–120.

Jaworsky, B. N. (2013). Immigrants, Aliens and Americans. *American Journal of Cultural Sociology, 1*(2), 221–253.

Jones, J., Amin, M. R., Kim, J., & Skiena, S. (2020). Stereotypical gender associations in language have decreased over time. *Sociological Science, 7*, 1–35.

Joos, M. (1950). Description of language design. *The Journal of the Acoustical Society of America, 22*(6), 701–707.

Joseph, K., & Morgan, J. H. (2020). When do word Embeddings accurately reflect surveys on our beliefs about people? In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/2004.12043.

Jost, J. T., Napier, J. L., Thorisdottir, H., Gosling, S. D., Palfai, T. P., & Ostafin, B. (2007). Are needs to manage uncertainty and threat associated with political conservatism or ideological extremity? *Personality & Social Psychology Bulletin, 33*(7), 989–1007.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). FastText.zip: Compressing text classification models. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1612.03651.

Justeson, J., & Katz, S. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics, 17*(1), 1–19.

Kharchenkova, S. (2018). The market metaphors. *Poetics, 71*, 71–82.

Kiley, K., & Vaisey, S. (2020). Measuring stability and change in personal culture using panel data. *American Sociological Review, 85*(3), 477–506.

Kirchner, C., & Mohr, J. (2010). Meanings and relations. *Poetics,, 38*(6), 555–566.

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture. *American Sociological Review, 84*(5), 905–949.

Kruszewski, G., Paperno, D., & Baroni, M. (2015). Deriving Boolean structures from distributional vectors. *Transactions of the Association for Computational Linguistics, 3*, 375–388.

Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. Proceedings of the 24th International Conference on World Wide Web, 625–635.

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015). From word Embeddings to document distances. *International Conference on Machine Learning*, 957–966.

Lakoff, G. (2010). *Moral Politics*. University of Chicago Press.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem. *Psychological Review, 104*(2), 211.

Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1512.09300.

Lazer, D., & Radford, J. (2017). Data ex machina. 10.1146/annurev-soc-060116-053457.

Le, M., Roller, S., Papaxanthos, L., Kiela, D., & Nickel, M. (2019). Inferring concept hierarchies from text corpora via hyperbolic embeddings. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1902.00913.

Lee, M., & Martin, J. L. (2014). Coding, counting and cultural cartography. *American Journal of Cultural Sociology, 3*(1), 1–33.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics, 4*(1), 151–171.

Levy, O., Remus, S., Biemann, C., & Dagan, I. (2015). Do supervised distributional methods really learn lexical inference relations? Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics, 970–976.

Lewis, M., & Lupyan, G. (2020). What are we learning from language? In psyarxiv.com. https://psyarxiv.com/7qd3g/download?format=pdf.

Linzhuo, L., Lingfei, W., & Evans, J. (2020). Social centralization and semantic collapse. *Poetics, 101428*.

Lix, K., Goldberg, A., Srivastava, S., & Valentine, M. (2020). Timing differences.

Lizardo, O. (2016). Cultural symbols and cultural power. *Qualitative Sociology, 39*(2), 199–204.

Lizardo, O., Sepulvado, B., Stoltz, D. S., & Taylor, M. A. (2019). What can cognitive neuroscience do for cultural sociology? *American Journal of Cultural Sociology*. 10.1057/s41290-019-00077-8.

Miller, G., & Charles, W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*(1), 1–28.

Mische, A. (2011). Relational sociology, culture, and agency. In The Sage Handbook of Social Network Analysis.

Mohr, J. (1998). Measuring meaning structures. *Annual Review of Sociology, 24*(1), 345–370.

Mohr, J. (2000). Introduction: Structures, institutions, and cultural analysis. *Poetics, 27*(2-3), 57–68.

Mohr, J., Bail, C., Frye, M., Lena, J., Lizardo, O., McDonnell, T., Mische, A., Tavory, I., & Wherry, F. (2020). *Measuring Culture*. Columbia University Press.

Mohr, J., & Duquenne, V. (1997). The duality of culture and practice. *Theory and Society, 26*(2), 305–356.

Mohr, J., & Lee, H. (2000). From affirmative action to outreach. *Poetics, 28*(1), 47–71.

Mohr, J., Wagner-Pacifici, R., & Breiger, R. L. (2015). Toward a computational hermeneutics. *Big Data & Society, 2*(2), Article 2053951715613809.

Mora, G. C., & Paschel, T. S. (2020). Antiblackness as a logic for anti-immigrant resentment. *Sociological Forum, 94*, 991.

Namenwirth, J. Z., & Weber, R. P. (2016). *Dynamics of Culture*. Routledge.

Nelson, L. (2020). Computational grounded theory. *Sociological Methods & Research, 49*(1), 3–42.

Nelson, L. (2021). Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century U.S. South. *Poetics, 101539*.

Nickel, M., & Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems 30* (pp. 6338–6347). Curran Associates, Inc.

Norris, C. (1985). *The Contest of Faculties*. New York: Methuen.

Orlikowski, M., Hartung, M., & Cimiano, P. (2018). Learning diachronic analogies to analyze concept change. Proceedings of the Second Joint. https://www.aclweb.org/anthology/W18-4501.pdf.

Ornaghi, A., Ash, E., & Chen, D. (2019). Stereotypes in high stake decisions. *Center for Law & Economics Working Paper Series, 2.* https://www.nber.org/~dlchen/papers/Implicit_Bias_in_the_Judiciary.pdf.

Pachucki, M., & Breiger, R. (2010). Cultural holes. *Annual Review of Sociology, 36*(1), 205–224.

Pomeroy, C., Dasandi, N., & Mikhaylov, S. J. (2019). Multiplex communities and the emergence of international conflict. *PloS One, 14*(10), Article e0223040.

Popping, R. (2012). Qualitative decisions in quantitative text analysis research. *Sociological Methodology, 42*(1), 88–90.

Rho, E. H. R., Mark, G., & Mazmanian, M. (2018). Fostering civil discourse online. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–28.

Roller, S., & Erk, K. (2016). Relations such as Hypernymy: Identifying and exploiting hearst patterns in distributional vectors for lexical entailment. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1605.05433.

Roller, S., Erk, K., & Boleda, G. (2014). Inclusive yet selective. Proceedings of Coling 2014, the 25th International Conference on Ccomputational linguistics, 1025–1036.

Rose, D. (1980). Malinowski's influence on wittgenstein on the matter of use in language. *Journal of the History of the Behavioral Sciences, 16*(2), 145–149.

Rotolo, M. (2020). Religion imagined. *Sociological Forum, 35*(1), 167–188.

Rubenstein, H., & Goodenough, J. (1965). Contextual correlates of synonymy. 10.1145/365628.365657.

Rubner, Y., Tomasi, C., & Guibas, L. (1998). A metric for distributions with applications to image databases. Sixth International Conference on Computer Vision, 59–66.

Ruder, S., Vulić, I., & Søgaard, A. (2019). A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research, 65,* 569–631.

Sáenz, R., & Douglas, K. M. (2015). A call for the racialization of immigration studies. *Sociology of Race and Ethnicity, 1*(1), 166–180. https://doi.org/10.1177/2332649214559287.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics, 20*(1), 33–53.

Schild, L., Ling, C., Blackburn, J., Stringhini, G., Zhang, Y., & Zannettou, S. (2020). "Go eat a bat, chang!": An early look on the emergence of Sinophobic behavior on web communities in the face of COVID-19. In arXiv [cs.SI]. arXiv. http://arxiv.org/abs/2004.04046.

Smith, N. A. (2019). Contextual Word Representations. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1902.06006.

Spillman, L. (2020). *What is Cultural Sociology?* John Wiley & Sons.

Spirling, A., & Rodriguez, P. (2019). Word Embeddings. Working paper.

Stoltz, D. (2019). Becoming a dominant misinterpreted source. *Journal of Classical Sociology*. 10.1177/1468795X19896056.

Stoltz, D., & Taylor, M. (2019). Concept mover's distance. *Journal of Computational Social Science, 2*(2), 293–313.

Strang, D., & Dokshin, F. (2019). Peer review and the production of scholarly knowledge. In *The Production of Managerial Knowledge and Organizational Theory, 59* pp. 103–121). Emerald Publishing Limited.

Strauss, C., & Quinn, N. (1997). *A Cognitive Theory of Cultural Meaning*. Cambridge University Press.

Tahmasebi, N., Borin, L., & Jatowt, A. (2018). Survey of computational approaches to lexical semantic change. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1811.06278.

Tashu, T., & Horváth, T. (2018). *Pair-Wise: Automatic Essay Evaluation using Word Mover's Distance* (pp. 59–66). CSEDU.

Taylor, M., & Stoltz, D. (2020). Integrating semantic directions with concept Mover's distance to measure binary concept engagement. *Journal of Computational Social Science*. https://link.springer.com/article/10.1007/s42001-020-00075-8.

Teplitskiy, M. (2016). Frame search and *Re*-search. *The American Sociologist, 47*(2), 264–288.

Thompson, A. (2018). "All the news" Corpus [Data set]. https://components.one/datasets/all-the-news-articles-dataset/.

Tithi, J., & Petrini, F. (2020). An efficient shared-memory parallel Sinkhorn-Knopp algorithm to compute the word Mover's distance. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/2005.06727.

Vaisey, S., & Lizardo, O. (2016). Cultural fragmentation or acquired dispositions? *Socius, 2,* Article 2378023116669726.

van Loon, A., & Freese, J. (2019). Towards a more computational affect control theory. 10.31235/osf.io/r7ewx.

Weber, R. P. (1984). Computer-aided content analysis. *Qualitative Sociology, 7*(1), 126–147.

Weber, R. P. (1990). *Basic Content Analysis*. SAGE.

Weeds, J., Clarke, D., Reffin, J., Weir, D., & Keller, B. (2014). Learning to distinguish hypernyms and co-hyponyms. Proceedings of Coling 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2249–2259.

Werner, M., & Laber, E. (2019). Speeding up Word Mover's Distance and its variants via properties of distances between embeddings. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1912.00509.

Wickham, H. (2019). rvest. https://cran.r-project.org/web/packages/rvest/index.html.

Wong, S. K. M., Ziarko, W., & Wong, P. (1985). Generalized vector spaces model in information retrieval. Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 18–25.

Wu, L., Yen, I. E. H., Xu, K., Xu, F., Balakrishnan, A., Chen, P.-Y., Ravikumar, P., & Witbrock, M. (2018). Word Mover's Embedding. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1811.01713.

Young, M. (2011). Malinowski's last word on the anthropological approach to language. *Pragmatics, 21*(1), 1–22.

Zannettou, S., Finkelstein, J., Bradlyn, B., & Blackburn, J. (2018). A quantitative approach to understanding online antisemitism. In arXiv [cs.CY]. arXiv. http://arxiv.org/abs/1809.01644.

Zelizer, V. (2012). How I became a relational economic sociologist and what does that mean? *Politics & Society, 40*(2), 145–174.

Zhang, H., & Pan, J. (2019). *CASM. Sociological Methodology, 49*(1), 1–57.

Dustin S. Stoltz is an Assistant Professor of Sociology at Lehigh University. His primary fields of inquiry are economic sociology, cultural sociology, stratification, computational social science, and cognitive social science. His work appears in *Sociological Theory, Sociological Science, Sociological Forum, Poetics, Socio-Economic Review, Journal of Computational Social Science, Socius, The American Sociologist, Journal for the Theory of Social Behavior, Journal of Classical Sociology, American Journal of Cultural Sociology*, and the *European Journal of Social Theory,* among no other outlets.

Marshall A. Taylor is an Assistant Professor of Sociology at New Mexico State University. His primary areas of interest are cultural sociology, cognitive social science, computational social science, sociological theory, and political sociology. His work appears in *Sociological Theory, Sociological Science, Socio-Economic Review, Poetics, American Journal of Cultural Sociology, Journal of Computational Social Science, Socius, Stata Journal, Journal of Classical Sociology*, and *Deviant Behavior*, among other outlets.