

# **Expressions of Community**

Connor Gilroy

03/16/2023

# Table of contents

<b>Preface</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
1.1 Short version . . . . .	5
1.1.1 Questions and goals . . . . .	5
1.1.2 A few theoretical clarifications . . . . .	8
1.1.3 A note on methods . . . . .	10
1.2 Long version . . . . .	11
1.2.1 Introduction . . . . .	11
1.2.2 Motivation . . . . .	14
<b>2 Density and abundance</b>	<b>24</b>
2.1 Background . . . . .	24
2.2 Motivation . . . . .	24
2.2.1 Research questions . . . . .	24
2.2.2 Short version . . . . .	25
2.2.3 Long version . . . . .	27
2.2.4 Notes . . . . .	32
2.3 Data and methods . . . . .	33
2.4 Results . . . . .	38

2.5	Interpretation . . . . .	49
2.6	Next steps . . . . .	52
2.7	Appendix: Tables . . . . .	57
2.7.1	Outcome 1: You feel you're part of the LGBT community. . . . .	57
2.7.2	Outcome 2: I don't feel I belong to anything I'd call a community.	58
<b>3</b>	<b>Geography or Gemeinschaft?</b>	<b>59</b>
3.0.1	Short version . . . . .	59
3.0.2	Long version . . . . .	62
3.1	Previous version . . . . .	79
3.1.1	Introduction . . . . .	79
3.1.2	Data and methods . . . . .	81
3.1.3	Results . . . . .	83
<b>4</b>	<b>Community talk</b>	<b>87</b>
4.0.1	Short version . . . . .	87
4.0.2	Long version . . . . .	89
<b>5</b>	<b>Conclusion</b>	<b>100</b>
	<b>References</b>	<b>101</b>

# Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

# 1 Introduction

## 1.1 Short version

### 1.1.1 Questions and goals

In this dissertation, I propose to study expressions of and experiences of community in order to better understand what community means, how community happens, and what community does. My empirical focus is on expressions of community from LGBTQ people, because community in LGBTQ contexts is highly resonant and visible rather than relegated to the background or taken for granted.

My investigation will proceed in three parts. I begin by examining **how the ordinary meaning of community is constructed through its use**. Does everyday talk about community invoke or map onto sociological understandings of *Gemeinschaft*? Are those sociological resonances more apparent in LGBTQ community discourse than in more generic or neutral social contexts? Second, I'll investigate **how individual sense of community varies across place-based contexts**. Do simple structural features of place-based contexts suffice to explain variations in belonging? Or are there differences beyond that, something more? Third, I'll investigate **the consequences of community-oriented talk in virtual communities**. Do linguistic expressions of community come from embedded or peripheral actors? Do they result in downstream

interactional consequences, or are they mere rhetoric?

This work will make two contributions: First, I aim to refine our understanding of how distinct aspects of community – shared culture, place, social ties – relate to each other; I anticipate finding that these are largely complementary rather than substitutes. Second, I aim to connect some disparate threads across existing research in order to systematically show that community can happen in at least two distinct ways: either it can be predicated on sameness in a narrow and exclusionary way, or it can be intrinsically heterogeneous and inclusive. The former is unsurprising, but the latter type of community expands the boundaries of how we might think of community as being empirically possible.

Recent historical and contemporary LGBTQ communities provide an interesting set of cases for me to draw on in order to understand the operation of community writ large, for several reasons. First, LGBTQ community cannot be taken for granted; LGBTQ people who want community together must actively seek it out (sometimes through literal migration) and build it. “Queer cultural reproduction” (Orne 2017) takes conscious choice and effort. Compared to contexts where community might be relegated to the background, like Tönnies’s stereotypical peasant village, this makes LGBTQ expressions of community highly visible and conducive to study. Second, even though the historical emergence of LGBTQ communities and identities is a relatively recent phenomenon (Bérubé 2011; D’Emilio 1992), LGBTQ communities cover the full range of things that a community can be, according to typologies of Gemeinschaft like Brint (2001)’s. Such communities can be virtual or emplaced; they can span from concrete social ties to imagined collectivities. Metaphorical concepts like constellations (Giesecking 2020) and archipelagoes (Ghaziani 2014) are used to characterize and emphasize the variety and interconnectedness of queer communities (for those two authors, specifically as they

unfold in urban spaces). This variation is real and meaningful but not infinite; Winer (2020) has shown how “gay imagined communities” possess a clear, concrete “imagined center” to which individual gay, bi, and queer men have ambivalent relations – what Winer calls “solidarity with disdain.”

Finally, LGBTQ communities can be homogeneous or diverse. Sometimes differences and tensions divide lesbian, bi and queer women from gay, bi and queer men; lesbians and gay men from bisexuals; cis LGBQ people from trans people; and queer people from people with more defined or less fluid orientations. Sometimes, LGBTQ people bridge those differences and build community together. They also can bridge or fail to bridge other kinds of social difference: race, age, class, and so on. A rich qualitative literature documents and categorizes the forms LGBTQ community can take, establishing that multiple distinct types of LGBTQ community can exist, characterized in part by their differing approaches to difference. Brown-Saracino (2017) shows how, for LBQ women in small cities, there might be a singular “lesbian community” premised on the sameness of identity, or a “hybrid” form of identity and community predicated on identities that are diverse and plural instead – or else, a third option, LBQ community can be effectively nonexistent, folding into a broader “ambient community” instead. Orne (2017) charts the “Disneyfication” of Boystown in Chicago alongside the failures of alternative queer political spaces to live up to their ideals; instead, they find the most radical and interesting potential in “sexy communities.” Together, these two ethnographies show that both ways of forming community around a shared identity – shared sameness, or shared difference – happen in various times and places. I’ll develop a way of studying expressions of community that accommodates both of these possibilities, and I’ll foreground heterogeneity when I turn to studying perceived experiences of community.

### **1.1.2 A few theoretical clarifications**

It's worth emphasizing at the outset that identity does not automatically create community. A conceit of this project is that identities like sexuality *might be* a basis for creating community, but probably only in conjunction with other features that might be shared, and active work on the part of cultural entrepreneurs. It is because community building in this context takes work that the study of expressions is so valuable.

What are those shared features that might make up a sociological definition of community? My work focuses on three: shared culture, shared ties, and shared place. I think it's possible to rank those features, in that order, based on prior research that's gone into refining the concept of community and exploring its limits. I don't plan to reevaluate the relative importance of these factors in this dissertation, but the structure-culture tension and the disputed role of place affect the operational choices I make and will shape how I interpret my findings.

Why culture over ties? The starker example is the concept of “imagined communities,” which dispense with the possibility of interaction among all, or even most, members of a given community entirely. Beyond that, however, a number of studies have found that a shared moral order (Vaisey 2007) or perception of cohesion (Boessen et al. 2014) matters more for creating a sense of community (*Gemeinschaft*) than do social interactions or ties.

Why ties over place? If community has turned out to be “liberated” rather than “lost” (Kelly et al. 2014; Wellman 1979), and if virtual community is possible (Driskell and Lyon 2002; Rheingold 2000), then that suggests that shared ties take precedence over shared place. But at the same time, these characteristics might complement rather than supplement each other. We know that the rise of virtual connections has not led to the death of distance (Spiro, Almquist, and Butts 2016), and mostly-virtual communities

can hold in-person meetups that strengthen members' sense of togetherness (Rheingold 2000). (More generally, “augmented reality” rather than “digital dualism” describes how virtual and offline experiences of community might intersect (Jurgenson 2011; Orne 2017).)

Given this complementarity, another way to put this might be that the more of these features are present to some degree, the more “prototypical” a community looks (Bruckman 2022): a tight-knit small town has all three, a group of Wikipedians has two, and a nation-state has only one. But breaking down the potential features that create a community goes some way toward explaining a key finding from qualitative studies of everyday conceptualizations of community (Levine 2017; Winer 2020), that the *referent* for “community” can be hard to pin down precisely. Is it all the people in a place, a set of people with social relations to each other, or people connected by a shared (sub)culture? This referential ambiguity partly arises from the multiple, potentially conflicting, elements of the definition of community.

But, backing up one step further, there’s one more theoretical point to clarify: what’s culture, anyway? At least as it relates to creating community, culture extends beyond moral values and beliefs, to include group styles, norms, and practices; it includes both declarative (i.e. linguistic or discursive) and nondeclarative (i.e. practical or embodied) modes (Lizardo 2017). Reproduction of queer cultures – and thereby queer communities – manifests everywhere from socialization into particular kinds of lesbian identity talk (Brown-Saracino 2017) to acquisition of a particular gay habitus (Orne 2017). I’d use those examples to deviate from Vaisey (2007) and argue that in focusing only on moral values as key to *Gemeinschaft*, he’s not expansive enough.

But what does it mean for culture to be shared? And why do I think something like “cultural density” might be better than “moral unity” (Vaisey 2007) to describe how

shared culture generates community? Culture is structured; beliefs, values, symbols, etc. fit together into systems of meaning through schemas and frames. Without delving too deeply into the cognitive details, I think certain configurations of cultural ideas and practices allow for the incorporation of diversity and variation into the framework of a single community, while others don't; again, I lean on those examples from Orne and Brown-Saracino as a sort of existence proof.

### **1.1.3 A note on methods**

I am taking a quantitative and computational approach to studying LGBTQ expressions of community. This offers a breadth and scale that complements and extends the deeply grounded research that already exists in this area. It's also a logical fit: given my theoretical framework and expectations, I need a window into how culture is structured, and I need to connect other attributes like ties, interactions, and place. Computational and digital methods – especially techniques from computational text analysis – provide this. They expand what can be done with traditional quantitative methods and enable me to build more directly on prior qualitative research. CSS methods can incorporate more insights from qualitative work into a quantitative framework, especially in terms of measuring and operationalizing culture (Mohr et al. 2020). Novel techniques like word embeddings are particularly suited for measuring how culture is structured into schemas and frames (Arseniev-Koehler and Foster 2022; Boutilier, Cornell, and Arseniev-Koehler 2021). These methods have limitations; most explicitly, they are not suited for studying the nondeclarative, embodied aspects of group culture (Lizardo 2017; Orne 2017). That is one reason for my focus on the linguistic, discursive aspects of culture and community. There will be other gaps – wouldn't it be ideal if I had the kinds of data I plan to use in chapters 1, 2, and 3 all for the *same* LGBTQ community or set

of communities? – but triangulating through multiple, partial cases is the most realistic way forward.

## 1.2 Long version

### 1.2.1 Introduction

“Community” does a lot of work out there in the social world. Why do social media companies purport to give people tools to “build community” with – and why have billions of people adopted those tools? Why do protestors demand not only that cities defund police departments, but also that they “invest in community” instead? In these varied circumstances, community is a motivation, an end, an object of desire, even an solution to social problems. I’ll pose community as a question instead. I want to investigate variations in what community means, how it happens, and what it does, by focusing particularly on the varied experiences of LGBTQ people and the communities they belong to.

At the same time, significant debates about the nature and operation of community have long taken place in sociology. What’s the relation between community and society (Tönnies [1887] 2001)? Has community been lost or saved or liberated in contemporary post-industrial societies (Hampton and Wellman 2018; Kelly et al. 2014; Wellman 1979; Wellman and Leighton 1979)? How bound is community to place, or is virtual community also possible (Baym 1994; Calhoun 1998; Driskell and Lyon 2002; Rheingold 2000)? Though many of these debates are settled, they show that community is an object of theoretical contention as well. I aim to unsettle some of what we think we know about community, by paying close attention to the margins of the phenomenon.

I’ll do this through triangulation, attacking the phenomenon from multiple comple-

mentary angles. As you read this proposal, you might see ways any one chapter could be expanded into a deeper, more focused dissertation on its own. But I'm aiming for breadth instead. In other words, I plan to shamelessly take advantage of my disciplinary home in sociology – not linguistics, not geography, not human-computer interaction. I'm not bound to study language, or place, or technology – only some facet of the social world. I can follow a social phenomenon, and a group of people for whom that phenomenon is often important, across multiple contexts. Those are my anchors.

Right now, each chapter proposal stands somewhat on its own. As they develop, I'll work to bring them closer together. I'd appreciate any thoughts you have about how to shape these projects so that they bend toward each other and align into a coherent whole. At the same time, by having distinct focal points, the chapters offer ways to bracket some of the moving parts – chapter 1 focuses squarely on discourse but remains somewhat descriptive; chapter 2 brings in representative survey measures but sacrifices some richness and detail; chapter 3 gives up that representativeness in order to study relational processes in more depth.

At a discursive level, I will examine the literal meaning and broader resonance of “community,” as part of a relational system of words and meanings. How does community vary in terms of what it means in everyday social discourse? How do expressions of community vary among social groups and social contexts? Can linguistic expressions of community and belonging be a measurement tool for something deeper? In LGBTQ contexts I expect the meaning of community to be more closely tied to identity and to Gemeinschaft than to mundane notions of place; I also expect more frequent expressions of language invoking community and belonging.

Scaling up to the level of broader contextual characteristics and down to the level of individually-reported experienced sense of community, I will next investigate the relation

between the two. How do cultural and place-based features like density and diversity shape variation in individual sense of community? I expect those features to facilitate community and belonging for LGBTQ people, especially in their connectedness to the LGBTQ community specifically, where they might not do the same for other groups of people.

Finally, at a relational level, through a case based in virtual communities; connect discursive measures from the first project with similar contextual features as in the second. How do expressions of community relate to structural and interactional dynamics like embeddedness within a group? While I expect the most embedded individuals to strongly invoke language of community and belonging in order to build community for the group, the most peripheral individuals within a network may do likewise in their own attempts to belong. LGBTQ virtual spaces provide a context wherein expressions of community are highly visible and where these processes can be observed.

This multifaceted, multi-pronged investigation matters because it interrogates who community is for and when community is possible. What are the limits, the edges, the boundaries? Who gets to experience belonging? Some framings of community are conducive to diversity, heterogeneity, and inclusion; some are narrow, homogenizing, and exclusionary. The former offer more chances to more people to have experiences of community. To the extent that community is a positive thing for individuals to experience and participate in, something worth creating and seeking out, finding out more about the unexpected and counterintuitive conditions under which belonging can emerge is a worthwhile endeavor.<sup>1</sup> Moreover, in addition to being an end in itself,

---

<sup>1</sup>Community is bad sometimes – i.e., even when an individual experiences it as a positive, the externalities of it can be quite negative. Think of white nationalist groups or men's rights activists or QAnon conspiracy theorists. One of the things people in those groups talk about gaining is a sense of belonging; many outsiders might seriously prefer that these people found some other way to achieve that experience. Granted, other people think that queer community is literally destroying the fabric of society.

community can be a precursor to various forms of collective and social action, like political mobilization. Understanding variations in the meaning, the conditions, and the consequences of community enriches our knowledge of a phenomenon that provides grounds for much of social life.

## 1.2.2 Motivation

### 1.2.2.1 Background

Community is a meso-level social phenomenon that springs into existence when certain conditions are met. In turn, it's something people might long for or become attached to, something that can motivate them to individual or collective action. There's a stable core to this phenomenon which is well understood, because community has long been one of the core sociological concepts. Community is the product of social density and cultural density, the result of both of which is the perception and reality of a group people can belong to and form attachments to. Social density in the form of relations and interactions is essential, but network cohesion alone is not sufficient for the emergence of a coherent and perceived community. Social density is often associated with physical density or emplaced co-presence (Collins 2004; Durkheim [1893] 1997), but the reality of virtual communities has shown that place, while beneficial, is not essential to the formation of community (Hampton and Wellman 2003; Rheingold 2000). What is necessary, however, is some form of shared culture. The precise terminology for cultural density varies in the literature; for instance, Vaisey (2007) refers to "moral unity" as the key to *Gemeinschaft*, while Tavory (2016) refers to "moral density"; I prefer "cultural" over "moral" because it encompasses norms, styles, and schemas that aren't necessarily moral values, but am alluding to the same domain. Similarly, I use "density" over "unity" because the latter takes for granted an assumption of sameness or uniformity,

which is not universally present in all communities.

Despite these recognizable core characteristics, community is also a notoriously broad and ambiguous concept. Partly this is because communities come in a complex taxonomy (Brint 2001), from the prototypical peasant village (Bruckman 2022; Tönnies [1887] 2001) all the way to virtual (Rheingold 2000) and imagined communities (Anderson [1983] 2016). At the margins are types of communities that call key facets of the concept into question. For instance, virtual communities are noteworthy because they unsettle the centrality of place, showing that togetherness in place isn't essential to creating a community – even if the theory of augmented reality (as opposed to digital dualism) shows that place can still play a role even for virtual communities (Jurgenson 2011; Orne 2017; Rheingold 2000). Imagined communities, similarly, challenge the centrality of social interaction (but “are imagined communities real?” is an oxymoron). While this extensibility might seem to render the notion of community so flexible as to be analytically useless, individuals prove to be quite capable of perceiving community in the social world. A useful way to account for this flexibility is to conceive of community as a fuzzy, prototype-based category (Bowker and Star 2000; Bruckman 2022; Lakoff [1987] 2008; Lakoff and Johnson [1980] 2008). Layering more of these features – shared culture, social ties, togetherness in place – together produces more recognizable, more prototypical instances of community.

One way to observe how and where the meso, group-level phenomenon of community happens is at the individual level. This measurable individual experience and feeling of belonging is called “sense of community” (Boessen et al. 2014; Frost and Meyer 2012; Mannarini and Fedi 2009; McMillan 1996; McMillan and Chavis 1986), and it generally implies the perception that a group exists and that the individual is a member. This could be a misperception – all of the typical caveats of self-reports apply! –

but I find that generally unlikely. Individual sense of community is a useful barometer for the actual reality of community for two reasons. First, the threshold for forming an actual group is low (according to Simmel (2012), three people). Second, even if something perceived as a community completely lacks something sociologists might consider a fundamental definitional characteristic, that perception is informative – and that affective, experiential dimension really can shape individuals' social actions. This is, in fact, the insight behind imagined communities; though they operate at a scale where social interaction is impossible, they are real enough to motivate people to acts they would not engage in otherwise (Anderson [1983] 2016). Sense of community is a psychological rather than sociological concept, in individuals' heads rather than out there in the world. But some social phenomena are perceptual, cultural, and distributed, even shared collective feelings or experiences; community is one of those phenomena. Accordingly, measures of sense of community are one angle I'll take to understand how community itself happens.

While *community* and *sense of community* are well-established concepts and measures, I would propose a third, called *expressions of community*. This is a language or vocabulary of belonging, encompassing both explicit invocations of the concept of community as well as implicit statements that resonate with it. With this concept, I deviate from studies that measure belonging through subcultural innovation and group conformity (e.g., Lucy and Bamman 2021). Adoption of a group discursive style is one way a group culture is produced (Eliasoph and Licherman 2003), of course, but not the same thing as expressions of community. Expressions of community are relational and potentially performative, i.e., they might be part of the process of creating community for a speaker or an audience. This potential is most evident in social discourse that is interactional, like a conversation. For instance, suppose that two members of

a group chat are debating the efficacy of vaccines, in an increasingly contentious way. A third member steps in to intervene, stating that we, as a community, have a range of opinions on many things, but we all belong together. Both contending parties moderate their debate going forward. Invoking community here is an act of repair, and achieves its intended proximate effect of reducing conflict – but for all parties involved, and for any bystanders, it potentially strengthens the very community being invoked. These attempts to invoke community are maybe not always successful in creating community. For instance, no matter that a property management company insists that its pricey apartment complex offers the chance to be part of a community alongside other amenities, this does not mean that neighbors actually talk to each other, much less feel any real sense of connection.<sup>2</sup> Observing the linguistic phenomenon and measuring its effects are distinct. Sociologists might interpret community talk as a kind of ritual (Collins 2004; Durkheim [1912] 2001), part of the work of welcoming new members, reaffirming longstanding commitments, or even repairing the group in moments of tensions. These expressions of community can be emotional, frequently bordering on trite and sentimental, but are still worth taking seriously for the impact they might have. In some circumstances, of course, they could be so formulaic and pervasive as to be bleached of meaning. In that way, part of what community talk simultaneously does is construct the meaning of community.

#### **1.2.2.2 Research questions**

Expressions of community provide the first puzzle for this research. How do they construct the meaning of community, and what connotations does that meaning entail?

---

<sup>2</sup>One way to describe unsuccessful attempts at invoking community that are particularly blatant, cynical, inauthentic, or exploitative is “community-washing”, by analogy with similar terms. I’m not alone in coining this term, which I first used privately in October 2020; a pair of British researchers came up with it independently in March 2021. <https://sifted.eu/articles/community-led-platforms/>

How does the meaning of community vary across discursive social contexts, as well as the depth and resonance of invocations of belonging? If part of what makes “community” ambiguous and hard to pin down is that it’s a folk concept as well as an academic one, then treating the meaning of community as an empirical question is one way to make progress. Seeing if that meaning *varies* systematically will also be revealing – in my case, I suspect that community talk is ultimately more shallow in generic contexts and deeper and more resonant in texts coming from marginalized groups.

A second puzzle: how can “community” accommodate variation? Structural embeddedness and interactional density feel relatively clear to conceptualize and measure, though “ties” are cultural constructs (White [1965] 2008) and interactions presumably are as well. Cultural embeddedness, as Goldberg et al. (2016) might call it, or cultural density, as I prefer, is not so clear-cut. Does cultural embeddedness mean conformity? If so, how might we make sense of community spaces or groups that contain riotous diversity, wide ranges of self-expression, and strong norms of inclusivity? I’d argue that there are communities that are “dense” along this cultural dimension without being uniform. Maybe one reason is that beliefs and values and other constituent elements of culture aren’t a hodgepodge simply stacked atop each other, they fit together in schemas and people use agency in engaging with them as a toolkit (Boutyline 2017; Goldberg 2011; Hunzaker and Valentino 2019; Swidler 1986). One simple way to make progress here might be to return to individual-level sense of community, and see whether that can be high in contexts of structural and cultural variation. Another way might be to leverage the distinction I drew above between group style, which implies a something shared or the same among group members, and linguistic expressions of community, which might provide a window into how community might exist alongside difference. Metaphorically, poking at the edges of the concept of community is how I expect to

uncover things that are interesting and unknown about how community happens and what it does in the world.

### 1.2.2.3 Case

#### Queer Counterpublics

By queer culture we mean a world-making project, where “world,” like “public,” differs from community or group because it necessarily includes more people than can be identified, more spaces than can be mapped beyond a few reference points, modes of feeling that can be learned rather than experienced as a birthright. The queer world is a space of entrances, exits, unstematized lines of acquaintance, project horizons, typifying examples, alternate routes, blockages, incommensurate geographies. World making, as much in the mode of dirty talk as of print-mediated representation, is dispersed through incommensurate registers, by definition *unrealizable* as community or identity.

– Berlant and Warner, “Sex in Public” (1998)

LGBTQ communities provide an interesting edge case for understanding the operation of community writ large. As Berlant and Warner’s notion of queer counterpublics demonstrates, queer social worlds are fluid, heterogeneous, and complex; they have unique features that exceed the bounds of what might be recognizable as community. And yet, more down-to-earth ideas of community and identity are central to the experiences of many LGBTQ people. The emergence of queer spaces and the communities that inhabit them is improbable and surprising; their continued existence is contingent and fragile (Bérubé 2011; D’Emilio 1992; Ghaziani 2014). But enough people evidently want them to make them happen – at first, in spite of stigma and oppression, and still,

in spite of normative and assimilatory pressures. Even today, many LGBTQ people want distinctive spaces in which to build queer community and culture (Pew Research Center 2013). Because queer community can't be taken for granted, it's an ideal site for observing explicit processes of community.

Several aspects of LGBTQ communities make them interesting. One unique feature is a lack of rootedness, which shapes queer life trajectories: for the most part, LGBTQ people are not born into and do not grow up in queer communities and queer spaces (Weston 1995). This is one reason that the existence and experience of queer community isn't something LGBTQ people can take for granted. I suspect that this is also one reason that surveys targeted toward LGBTQ people ask about community in detailed and explicit ways, where more general surveys often do not; the need for community is visible and salient in LGBTQ contexts. Similarly, this might be why LGBTQ people have sought out virtual communities since those spaces first came into being.

At the same time, LGBTQ identities are often strongly held and offer a strong basis for community formation. LGBTQ identity shapes attitudes toward sexuality, but also beliefs and values more broadly (Schnabel 2018). LGBTQ identities are potentially encompassing enough to constitute a subculture (Fischer 1975; Mattson 2015a) or a counterpublic (Berlant and Warner 1998), leading theorists to write about gay culture (Halperin 2012) or a way of life (Foucault 1998). Structurally speaking, the mere fact that gay/queer/LGBTQ community exists as a possible object of attachment and source of belonging is itself noteworthy; the identity itself structurally and cognitively provides a potential axis of belonging that isn't available otherwise. By contrast, the unmarked category of "heterosexual" isn't available as a source of community in the same way – referring, for instance, to "members of the heterosexual community" is nonsensical (Zerubavel 2018), even if many communities do have heteronormativity and

heterosexuality as defining or central traits (Eliasoph 1998).

Queer communities are marked by their diversity and fluidity, by a proliferation of identities, expressions, practices, and ways of being. Due to processes of social sorting, this variation is not observable everywhere; some particular LGBTQ groups and spaces are more homogeneous and exclusionary, some more diverse and welcoming. Importantly, whether they embrace difference or exclude it, queer communities are not utopic and egalitarian; they're structured by inequalities of race, class, age, gender, embodiment, etc. [CITE: numerous]. In the most optimistic assessments, queer communities and queer spaces might offer ways to complicate or bridge class (Mattson 2015b) or race (Orne 2017); a more pessimistic take might be that other social structures present a fundamental obstacle to fully experiencing or building community. Regardless, anarchic, chaotic variation is a defining feature of the overarching LGBTQ community as a whole (Brekhuis 2003; Brown-Saracino 2017; Licherman 1999; Mattson 2015b; Orne 2017). This variation might be thought to undermine community cohesion and lead to divisions, especially in the context of selective state recognition (Butler 2004a). But I would argue, and ultimately hope to show, that empirical reality is more complex than the most pessimistic predictions of queer theorists and activists – namely, that the most normative and assimilable LGBTQ people will leave the rest behind. For example, how much solidarity is there between cisgender LGBQ people and trans people, in the current moment of anti-trans activism and legislation? Far more than none, but perhaps still not enough. Licherman (1999) once argued for “flexible forms of solidarity that sustain both unities and particularities,” as essential for a multicultural democracies; I would argue that queer communities furnish an important example of that possibility. As political communities that are less stable and unitary, and more fluid or heterogeneous, they show how to facilitate both a sense of group belonging and the capacity for

collective action, while incorporating democratic notions of freedom (Butler 2004b).

I would not argue that the LGBTQ case is completely unique. In some ways, LGBTQ people might be similar to people who hold other marginalized and minoritized identities: shared experiences of social marginalization, stigma, and discrimination, or even the experience of being a numeric minority; all factors which could heighten the perceived need for community, affecting what it means and how it can be created. Those analytic features are important, but I don't aim to use them to draw analogies between, say, sexuality and race (i.e., between the experiences of LGBTQ people and people of color). A key reason researchers are reluctant to analogize in that way is that people of course simultaneously hold multiple social identities, or occupy multiple social positions, and those identities can interact in complex ways. To translate that identity-oriented frame into terms of community and belonging, it's very likely that queer people of color find their ability to experience belonging in queer spaces to be mediated by race (Held 2017), and in communities of color mediated by sexuality. In any case, stigma and exclusion in heteronormative spaces operate as a push factor in tandem with the attractive pull of queer community, to explain why and how LGBTQ people come to seek out queer spaces. Variation on both of those fronts shapes how LGBTQ individuals experience connection to and participation in LGBTQ community.

Examining different ways queer community can play out across distinct geographic, temporal, and technological settings is how I might aim to triangulate around the concept of community. For some of the projects I propose below, a generic counterpart may provide a means of comparison; in other cases, internal variation might be adequate.

#### **1.2.2.4 Method**

My methodological approach is quantitative and computational. It's oriented toward synthesis, and aimed at complementing prior work on queer communities, and about community more broadly. This prior work is primarily qualitative, and I think there's space to try something different. I'll combine techniques for text, spatial, and network data analysis with conventional methods, and draw on the advantages of quantitative methods for studying variation. This also plays to the strengths of my training. What I'm offering here by taking a computational and quantitative approach to community is a different angle – not necessarily a better one, but hopefully a complementary one, and one that I'm uniquely positioned to undertake.

# 2 Density and abundance

How place characteristics shape individual sense of community for LGBQ people

## 2.1 Background

## 2.2 Motivation

*This is an informal writeup of the analysis I did for the second chapter of my dissertation over the summer, elaborating on the results I presented at ASA 2022. Below, I've copied the research questions and theoretical motivations from my proposal and abstract.*

### 2.2.1 Research questions

*From my prospectus defense slides.*

1. What features of individuals and places are associated with individual sense of community for LGBTQ people?
  - a. Do dense places full of LGBTQ people and institutions facilitate a greater sense of connection to the LGBTQ community, or are those exactly the places

where LGBTQ community fades into the background?

- b. What about the diversity and heterogeneity of a place — can that contribute to a sense of connection or is it necessarily an obstacle?

### **2.2.2 Short version**

Shared social and cultural characteristics are how community is created from identity, but what are the actual conditions that lead to strong LGBTQ communities rather than weak or nonexistent ones? In particular, how do various forms of social difference relate to individuals' sense of connection to the LGBTQ community? Chapter 2 addresses the conditions that make LGBTQ community possible for individuals by examining variations across places as well as heterogeneity within place-based contexts. While this chapter can't focus on expressions of community in the same way as the preceding or following chapters, it offers a parallel focus on perceived experiences of community instead.

The top-level expectation is that a dense place-based context with an abundance of queer people and institutions will facilitate a correspondingly strong sense of belonging and connectedness to the LGBTQ community for individual survey respondents. However, the qualitative literature hints at some alternative possibilities: perhaps individuals' perceived sense of belonging and connectedness to the LGBTQ community does not correlate – or even runs counter to – the abundance, density, and diversity of LGBTQ people and institutions in a place. That would suggest that the perceptual experience of community might stand in for structural and demographic “facts on the ground” rather than complementing or arising from them. In other words, maybe San Francisco is like Brown-Saracino’s Ithaca, and ambient community there takes the place of LGBTQ community specifically. Winer (2020)’s key finding of solidarity-with-disdain

might have place-based limitations as well – his interviewees, after all, come only from Southern California – in smaller and more scattered contexts LGBTQ people might not be able to afford to symbolically distance themselves from the center of the community. Finally, Forstie (2020)’s study of LGBTQ communities in small cities raises the possibility that those LGBTQ communities might be less fragmented and more cohesive, especially across lines of difference, compared to communities in larger cities.

The data sources I’ll combine allow me a unique opportunity to evaluate some theories from the qualitative literature on LGBTQ communities. While Brown-Saracino (2017) gains analytic leverage by looking at places that are very similar to each other on the surface and exploring their differences, she can’t explore the full space of places where LGBTQ people live and where they might experience community (or not). But the three kinds of place-based identities she uncovers can be mapped onto community connectedness measures like those in the Generations study. Hybrid identities and traditional lesbian communities would *both* be consistent with a strong sense of belonging to the LGBTQ community; “ambient community” would be signaled by high belonging in general but low LGBTQ-connectedness specifically. The sense-of-community survey measures are informative even though they can’t distinguish what, exactly, “LGBTQ community” means in a given place or what precise forms it takes. (While the Generations survey asks about connections to “*the* LGBT community,” actual LGBTQ communities are often multiple or fragmented.)

To sort out some of these competing expectations, I’ll refine my place-based covariates to distinguish central cities of metropolitan areas from peripheral areas (and, among central cities, especially the so-called “big four” from more ordinary cities) – in addition to measures of physical density, demographic diversity, and the presence of LGBTQ third places (i.e., gay bars, which I’ll obtain from Gregg Mattson). Finally, I still want to

take Brown-Saracino's ideas about place-based narratives literally and operationalize those narratives, probably by scraping text from public-facing city websites.

A few other notes about this chapter: as I've already discussed, one of the challenges of studying experiences of community is the slipperiness of the referent – what, in a given case, is “the community”? I also highlighted how important it is to distinguish community from identity. One of the exciting aspects of the Generations survey is that it actually captures (some of) this complexity. Most critically, it asks respondents about their connection and belonging to *any* community and to *LGBTQ community specifically*. And it also includes separate questions about LGBTQ community and identity. I plan to use the two distinct measures of community connectedness as my primary regression outcomes and include individual characteristics (including sexual orientation, gender identity, race, age) as well as place-based measures as covariates that capture potentially-salient sources of heterogeneity.

In feedback, the Generation study's main prior finding (Meyer and Choi 2020) about community connectedness has come to the surface; namely, political affiliation seems to matter – the minority of respondents who are Republican express much less LGBTQ community connectedness than do Democratic respondents. Perhaps this means that Vaisey's ideas about shared moral values as the key to Gemeinschaft hold up here, in that values are how an otherwise inclusive community draws its boundaries.

Sadly, I don't think I'll have the statistical power to estimate interactions between individual and place characteristics, though it would be logical to think those interactions exist.

### 2.2.3 Long version

*Included for completeness, feel free to skim or skip to the results writeup.*

Motivating this work is the fact that an individual's sense of community is shaped by structural features of the contexts in which they are embedded (Boessen et al. 2014). Purely virtual and distributed communities (Driskell and Lyon 2002) aside, experiences of community are typically local and emplaced (Brint 2001). At the broadest level, then, place characteristics shape how and whether individuals experience a strong sense of community. Place characteristics most obviously matter for place attachment; but they also matter for other objects of belonging, like identity-based groups. While other work might examine how the meaning and interpretation of "community" varies in a qualitative sense, this project focuses on the strength or magnitude of sense of community without addressing how its meaning might vary.

LGBTQ people provide a particularly interesting case to study the strength of community. For marginalized and minority groups, the stakes of belonging are heightened in light of a history of stigma and exclusion; the place characteristics that matter might differ from those that facilitate community for the generic, unmarked majority (see Zerubavel 2018 for a discussion of markedness). For instance, whether homogeneity lead to a feeling of cohesion, or whether it is instead stifling, depends on who a person is. Because they have the potential to be unique, I'm interested in what LGBTQ experiences of community might reveal about the relationship between place and community.<sup>1</sup> Moreover, LGBTQ attachments to community are emplaced, both through the existence of archetypical institutions like gay bars (Mattson 2020) and gayborhoods (Ghaziani 2014) as well as through broader and more diffuse constellations of significant places (Giesecking 2020). By linking features of place to individual LGBTQ experiences,

---

<sup>1</sup>It is no doubt possible that ethnic, racial, religious, or political minorities might share some commonalities with sexual minorities, but I prefer to focus on one axis of difference as a starting point and not overgeneralize. An appropriately intersection analysis of the LGBTQ context will of course consider how social identities like race and gender factor into the focal relationships between place characteristics and sense of community.

I hope to uncover the conditions that facilitate sense of community. I imagine three key factors to be especially relevant: density, diversity, and place-based culture.

I expect density to be a net positive. By density, I simply mean physical density of people; since Durkheim, sociologists recognized that this physical density facilitates a social density of interactions (Durkheim [1912] 2001; Tavory 2016). I expect the densest places to entail the strongest sense of community for LGBTQ people. Cities are historically entwined with the formation of collective LGBTQ identity (D'Emilio 1992), so much so that LGBTQ studies has been critiqued for its metronormative emphasis (Halberstam 2005). More broadly, cities have been argued to facilitate either individualism (Simmel [1903] 1971) or collective life (Jacobs 1961). A main mechanism for density to contribute to a sense of community for LGBTQ people is *movement*, especially given that most LGBTQ people do not necessarily grow up around many other queer people or with access to queer spaces. This stereotypical story of the attractive force of gay urban life (D'Emilio 1992) has also been called the “great gay migration” (Weston 1995), although that's a fraught analogy to make. In any case, it's easy to imagine concentrations of LGBTQ life as an attractive force, leading queer people to concentrate around each other and to individually experience a greater sense of community as a result. But maybe queer migration is also part of what's been called the “big sort” (Bishop 2009), and those that remain are positively self-selected as well. If both rural and urban LGBTQ people have the resources and opportunities to be where they feel that they can belong, then differences between them in sense of community might be lessened (see some initial evidence below).

The relationship between community and diversity, even for LGBTQ people, seems potentially more ambivalent and heterogeneous. I mean diversity in a multidimensional social space (maybe akin to Lee et al. 2021); not just diversity in terms of sexuality,

not just ethnoracial diversity; everything that contributes to the perception that a place has many kinds of people. It is not obvious that this contributes positively to a sense of community in itself or if it just goes along with density or other traits; for instance, people belonging to social minority groups, like sexual minorities, might incidentally like more diverse places simply because those places have more people like them.

But it's possible to imagine that the opposite could hold; that there are people for whom diversity itself positively contributes to their sense of community, of belonging to an identity group and/or to a place. While I do not expect this to be universal, it is a possibility I'm intrigued by: the experience and existence of community is predicated on something being shared among people, but can shared difference be enough? There is qualitative evidence that LGBTQ people can build community across difference in various ways, even creating a shared sense of hybrid and plural identities (Brown-Saracino 2017; Orne 2017). In a different line of evidence, thorough critiques (Abascal and Baldassarri 2015; Baldassarri and Abascal 2020) have recently shown that diversity – ethnoracial diversity, specifically – is not the intrinsic obstacle to social trust that it was once asserted to be (Putnam 2007). The data I can assemble may not sort out the stakes of solidarity across difference in diverse contexts, but I can start to assess the possibilities and variations.

Of course, those two factors aren't exhaustive; material resources and institutions – especially the presence and concentrations of third places like bars and coffee shops (Oldenburg 1998) – no doubt positively contribute to a sense of community as well. Assembling all of these place variables at the zip code and other levels of geography is relatively straightforward; I can create population density and other spatial variables like diversity indices from the American Community Survey and other Census Bureau data sources. The locations of third places like (gay) bars and coffee shops might come

from web APIs like Yelp’s, which provide contemporary estimates. (For previous work, I’ve created a geocoded data set of gay bars using Web Archive pages from 2007, but 2022 is closer in time to the 2016-2019 time period of the Generations study.) Note, however, that these contextual variables are all quite broad; I don’t aim to measure the composition of individual social worlds – the frequency of interaction, the social characteristics of alters, the establishments people frequent; place characteristics only measure the background context in which those social worlds take place.

But these structural and demographic characteristics still aren’t enough; they’re also overly broad in that they don’t capture the specificity of a particular place (Gieryn 2000) – which might be more or less conducive to community. This remainder can be conceptualized as *place-based culture* (Brown-Saracino 2017). To go beyond the catchall of culture, I hope to explain some of those residual unknowns by operationalizing the narratives associated with a place. This captures the perception and feeling of a place, at least a widely experienced version of it (i.e. declarative, public culture, in terms of Lizardo (2017)’s taxonomy). For instance, it’s possible to tell a story about a place where queer people are visible and abundant, a central part of the place’s appeal. Place narratives might even mediate the relation between structural features and individual experiences of community; e.g., it’s possible to tell a story about a place where diversity is a positive good rather than an obstacle to building community. Ethnographers say these stories matter (Brown-Saracino 2017; Orne 2017), that they’re part of how community plays out differently in practice even in places that appear similar on the surface – both in what exactly “community” comes to mean in a given place, and how strong the sense of belonging is for the individuals there.

Measuring place-based cultural narratives is more ambitious than measuring structural and demographic features. To do so, I need geocoded text data that conveys

public-facing representations of a place. I can imagine two types of digital data that might contain general place-based narratives. First, official tourism-oriented websites; Brown-Saracino (2017) discusses the parallels between those official stories and the narratives her LBQ respondents personally hold, arguing that they arise from the same broad cultural source. These webpages often take the form of “Visit [place name]”, for instance <https://visitseattle.org/> or <https://www.visitithaca.com/> – both of which prominently feature LGBTQ themes to highlight their highly visible LGBTQ communities. Systematically searching for and scraping the text from those webpages<sup>2</sup> would provide a corpus of narratives about cities, towns and other local regions. A second, more geographically granular option, would be geocoded Wikipedia pages. These are less official, but more organic, and despite important systematic biases still provide critical information about what a place is known for. With either corpus, I can fit topic models to distill common and unique themes, and then use those topic prevalences as covariates in regression models.

#### 2.2.4 Notes

- The short version of the motivation comes from the revision memo for my prospectus. The longer version of my theoretical expectations comes from the full prospectus document. It may be redundant, but I included it anyway.
- After encountering the data, I’m less confident in my ability to disentangle diversity from minority abundance, so I may need to deemphasize that part of the motivation. Feel free to prove me wrong, though.
- Gregg never got around to sending me the gay bar data. Maybe that’s a problem to leave for future work, if I ever have the time and energy for collaborations again.

---

<sup>2</sup>I have used web scraping, web crawling, and API-based data collection extensively in prior work; this is one of the unique things I can contribute beyond merely reanalyzing secondary survey data.

- As you'll see below, I haven't tackled the cultural narratives piece at all. That's a unique idea; I still like it, but I don't know whether I'll have the time to implement it. => I won't, forget about it! I do think my MSA findings let me comment on whether that's even viable, so discuss that in the conclusion.

## 2.3 Data and methods

The contribution of this paper is to analyze the association between place characteristics from the ACS (at the ZCTA and MSA levels) with individual survey responses about community and belonging in wave 1 of the Generations study.

The key data source for this paper is the *Generations* study (Meyer 2020), a three-wave representative panel survey of three age cohorts of cisgender Black, white, and Latinx lesbian, gay, bisexual, and queer and other nonheterosexual (LGBQ) people in the United States. The publicly-available version of the data set includes multiple measures of sense of community and belonging as parts of two composite scales: a generic *Social Wellbeing* scale and an LGBT-specific *Community Connectedness* scale. The public data, however, do not include geographic location beyond urban/rural and Census region. The restricted portion of the data set records respondent locations at more granular geographies: state, metropolitan/micropolitan statistical area, and finally zip code. To maintain respondent privacy, these restricted data are held by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan and accessed through a secure virtual environment, with a data use agreement to maintain confidentiality. (One key provision of this agreement: I cannot disclose *which* geographic locations are actually observed in the Generations data, although I present summary statistics below.) The University of Washington IRB approved the use of the restricted data for this study.

I use four measures from wave 1 of the Generations study - both the two full scales and the first single item from each - as outcome variables to represent community and belonging. The multi-item scales are validated and robust. By contrast, single items from those scales are framed to more precisely target *belonging*. While it's important to justify deviations from standard, validated scales (Mustillo, Lizardo, and McVeigh 2018), it's also important not to over-project meaning onto latent constructs derived from survey measures or take those meanings for granted (Martin and Lembo 2020, 2021). As a compromise, I analyze both and discuss any noteworthy divergences below. These are the scales and items I analyze:

- The **LGBT Community Connectedness** scale, a 4-point scale constructed from an average of 7 items, previously validated as a cognitive/affective construct in Frost and Meyer (2012). Items touch on emotional attachment, participation, and political and collective action within “the” (singular) LGBT community.
- The **Social Wellbeing** scale, a 7-point scale constructed from an average of 15 items. This scale includes not only items that ask about respondents’ relations to community, but also to society and the world more broadly.
- The first item on the Community Connectedness scale, “You feel you’re a part of the LGBT community,” with four response options ranging from “Agree strongly” to “Disagree strongly.”
- The first item on the Social Wellbeing scale, “I don’t feel I belong to anything I’d call a community,” with seven response options ranging from “Strongly disagree” to “Strongly agree.” (*Strongly disagree* with this negative statement corresponds in direction to *agree strongly* on the LGBT Community Connectedness question, so I align my model results and figures accordingly.)

For place-based data, I draw on the US Census Bureau’s American Community Sur-

vey [CITE] via the `tidycensus` package [CITE]. At the smallest levels of geography, only pooled 5-year ACS estimates are available. I use the 2019 estimates (from 2015-2019), which overlap with wave 1 of the Generations survey (2017-2018) and also have the highest-quality same-sex couples data. I uploaded relevant covariates, for all ZCTAs and MSAs, to ICPSR's virtual data enclave, and then joined and filtered them to only the zip codes and MSAs from which respondents in the Generations study were sampled.

In this paper, I consider two focal place-based covariates from the ACS: population density and prevalence of same-sex couples. I analyze these place variables at two geographic levels available in the restricted Generations data - ZCTA and MSA. At the MSA level, I use population-weighted densities aggregated up from the zip code level, to account for the fact that MSA boundaries - derived from counties - vary hugely across different regions of the country [CITE: geographers who say using MSA population density is bad]. I focus primarily on results at the zip-code level, which turn out to be more substantively and statistically significant.

In my statistical models, I test four different functional specifications of the association between population density and community/belonging:

- **Linear.** The most straightforward specification, allowing me to detect whether denser places are associated with a stronger or weaker experience of community.
- **Quadratic.** A theoretically-informed extension of the linear specification. This allows for the possibility, discussed above, that both rural small towns and large urban environments are more conducive to community and belonging than suburban sprawl.
- **Logarithmic.** A data-driven transformation, because the distribution of ZCTA population densities is right-skewed. Taking the (base-10) log produces a more

normal distribution of the variable. (Because of this same skew, I present plots involving population density on a log scale below.) Like the linear specification, the log transformation can only be monotonically increasing or decreasing.

- **Spline.** Agnostic and flexible, using a generalized additive model (GAM) to learn a potentially nonlinear functional form from the data.

As I'll show below, while the more complex specifications sound plausible, there's no evidence to support their use.

I use only one specification for the prevalence/abundance of LGBTQ+ people: the proportion of households that are same-sex couples. There are no fine-grained geographic estimates of LGBTQ+ identity (and based on Census Bureau trends, there never will be), so partnered households are the best proxy. I considered instead using *counts* of same-sex couples, as well as total population counts, rather than constructing two variables that are essentially *rates*. (Note: Because I pull same-sex couple counts and household counts from separate ACS tables, a few zip codes have nonsensical or extreme values for percent same-sex couples, which I drop from summary statistics. These are zip codes with small populations and/or high proportions of residents in group quarters.) It's an open question whether the *proportion* or *count* of same-sex couples is theoretically more important, but this alternate (and potentially simpler) specification did not turn out to be especially promising, so I did not explore it systematically.

I control for a set of individual demographic, socioeconomic, and other identity-based characteristics from the Generations data: gender (cisgender women, cisgender men, and [some] nonbinary/genderqueer people; transgender potential respondents took the parallel TransPop survey, not the Generations survey), sexual orientation (lesbian/gay, bisexual, queer/other), age cohort (younger, middle, older), race (white, Black/African American, Latino/Hispanic), political affiliation (Republican, Democrat, independent/other),

and education (high school or less, more than high school). Some of these traits are associated with community and belonging in interesting ways I won't discuss in this paper. Individual covariates can be analyzed with the publicly available version of the Generations data set, and so they're peripheral to my central aim of analyzing the restricted geographic data. (Most notably - and unsurprisingly - the small minority of LGBTQ Republicans report low LGBT community connectedness and belonging. This finding was previously reported in a bivariate analysis [CITE], but my results show that it holds up in the presence of other controls.) Others of these controls aren't associated with one or another of the outcomes, but I retain them for consistency across models.

Beyond potentially being associated with community/belonging, these individual traits vary spatially. Race is of course central to the spatial demography of the United States, and political affiliation is markedly geographically structured as well (both regional and urban/rural divides). In terms of gender and sexuality, gay men are more likely than lesbian women to concentrate in urban centers [CITE]. Education level affects opportunities to relocate, and respondents of different ages may be at different points in their life course that affect where they choose to live (e.g., younger respondents in cities, older respondents in suburbs).

The marginal effects plots I show below present predicted values and predicted probabilities with the individual controls set at their reference categories: young white lesbian Republican women with a high school education or less. Of course, reference categories are political and theoretically important [CITE: Sasha], but the choice of reference categories does not affect my main results. (Most notably, "white" and "Republican" both shift the level of LGBT belonging and connectedness downward compared to Black/Latinx or Democrat/independent, but they do not alter the interpretations of the place covariates.)

With 4 outcomes, 2 geographic levels, and 4 functional specifications of density, my main results consist of a series of 32 statistical models – although many of these models turn out to be uninformative. I model the multi-item scale outcomes (which can take fractional values) with linear regressions, and the single-item outcomes with ordered logistic regressions. For the spline functional specification of population density, I use GAM extensions of both model types (from the mgcv package [CITE]). As a robustness check for the multi-item scale outcomes, I test two additional varying-intercept multilevel models grouped at the MSA level. This is not a fruitful approach, but if it had been, it would have been appropriate to recast many of the other models as multilevel models.

## 2.4 Results

First, I'll describe the distributions of the relevant variables, to contextualize the main statistical models that follow. Table 2.1 shows that the typical Generations respondent (47.2%) agrees that they feel a part of the LGBT community, and this corresponds closely to the overall 7-item average of 2.97 on the Community Connectedness scale. The scale is flipped, so that higher numbers represent greater connectedness, and a respondent who agrees with every item would receive a score of 3. Consistent with the LGBT-specific trend, respondents tend on average to disagree with the notion that they don't belong to any community, and this aligns with the 4.67 average response for the 15-item Social Wellbeing scale (again, aligned so that higher values represent more positive outcomes). All four measures point to a moderately positive sense of community and belonging on average, but with enough variability to attempt to model systematic differences among respondents and the places in which they live.

Table 2.4 shows the distributions of the two key place covariates for the zip codes and metropolitan areas represented in the *Generations* study. Population densities vary

Table 2.1: Outcomes

(a) Individual item outcomes

	N	%
You feel you're a part of the LGBT community.		
Agree strongly	239	17.4%
Agree	649	47.2%
Disagree	389	28.3%
Disagree strongly	99	7.2%
I don't feel I belong to anything I'd call a community.		
Strongly disagree	252	18.3%
Moderately disagree	297	21.6%
Slightly disagree	227	16.5%
Neither agree nor disagree	142	10.3%
Slightly agree	198	14.4%
Moderately agree	168	12.2%
Strongly agree	92	6.7%

Source: *Generations* study (Meyer 2020)

(a) Scale outcomes

	Mean	Std. dev.
LGBT Community Connectedness scale (1-4)	2.97	0.56
Social Wellbeing scale (1-7)	4.67	0.91

Source: *Generations* study (Meyer 2020) 39

widely across zip codes (sample SD = 5,200 individuals per sq. km), and a strong right skew is evident where a tail of zip codes are especially densely populated. The mean respondent lives in a zip code with 2,500 individuals/km<sup>2</sup>, while the median respondent lives in a zip code with only 1,000 individuals/km<sup>2</sup>. The population-weighted mean density for all populated zip codes is 1,560 individuals per sq. km, meaning that LGBQ individuals in this representative sample live in zip codes that are on average substantially denser than the American population at large. Same-sex couples are around 1.1% of the households in the average zip code represented in the study, with a slightly lower median (0.8%) and a reasonable amount of variation (SD = 1.1%). As with density, the average is higher than the population-weighted mean for all populated zip codes of roughly 0.75%. (This is shaped by the fact that 54% of populated ZCTAs recorded 0 same-sex couples in the 2015-2019 ACS time period. Because the ACS is not a census and same-sex couples are rare, many of these are not likely to be true zeroes.) On both distributions, especially percentage of same-sex couples, note that MSAs show much less variability than zip codes.

These covariate distributions shape how I model, present, and interpret my results. The distribution of population densities, in particular, informs my decision to display predicted values on a log scale. For interpreting magnitudes, you can anchor on the idea that around half of zip codes are above and below 1,000 individuals per sq. km, and around half are above and below 1% same-sex couple households

Table 2.4: Place characteristics

	Median	Mean	Std. dev.
Zip codes in the Generations study (N = 1,238)			
Population density (individuals/sq. km)	1,000	2,500	5,200

Percent same-sex couple households	0.8%	1.1%	1.1%
MSAs in the Generations study (N = 217)			
Weighted population density (individuals/sq. km)	1,200	2,050	2,540
Percent same-sex couple households	0.9%	0.88%	0.23%

Source: 5-year American Community Survey, 2019

Note: Characteristics for zip codes and metropolitan areas represented in the Generations study. Values are rounded to maintain privacy. Values are weighted by number of respondents, meaning that these are the values *experienced* by the average respondent.

Two examples drawn from the full set of zip codes will provide anchor values for those distributions and help contextualize the model results that follow. These are shown in Figure 2.1. (Remember, I can't discuss which zip codes and MSAs are actually included in the Generations study.) I've chosen these examples as quantitative outliers corresponding to culturally significant places, illustrating what a place where 10% or more of households are same-sex couples actually look like.

- Zip code 94114 encompasses the Castro, San Francisco's gayborhood and one of the most prominent gay neighborhoods in the country. 12% of households in the area are same-sex couples. The Castro is in a densely-populated residential part of the city (9,500 individuals per sq. km), and adjacent to the extremely dense downtown core of San Francisco (with densities reaching 20,000 individuals per sq. km). San Francisco is in turn the densest part of the wider Bay Area, and one of the densest major cities in the country.
- Zip codes 92262 and 92264 coincide closely with Palm Springs, a gay resort town in Southern California. 12% and 14% of households are same-sex couples, respectively. However, as the map shows, Palm Springs is far from the most densely-

populated part of the Riverside, CA, metropolitan area (population densities are 300 and 150 individuals per sq. km, respectively). In other words, it's only extreme on one place characteristic, not both dimensions. Overall densities in the area are much lower compared to the Bay Area, and more typical of the country as a whole.

These real places illustrate the plausible upper end of the range for the prevalence and abundance of LGBTQ people, proxied through same-sex couples, and two distinct points on the spectrum of population densities. In the subsequent results, moving from 1% same-sex couples and 1,000 individuals per sq. km to 10% and 10,000 individuals per sq. km is like moving from a typical zip code in the sample to a place like the Castro.

And now, dear reader, the results.

**Population density.** On its own, increasing population density in a respondent's zip code has a positive association both with belonging to the LGBT community ( $b = .13$ ), and with belonging to any community in general ( $b = .14$ ). On the latent variable scales, a shift from 1 person per sq km to 10,000 people per sq km (arbitrary extreme, but plausible, values) is predicted to move each outcome by .52 and .56 respectively. For LGBT belonging, this doesn't change the mean predicted category (still "Agree"); for generic belonging, this shifts the mean predicted category from "Neither agree nor disagree" to "Slightly disagree" (the direction of the scale is reversed). [TODO: simulate predicted probabilities and first differences with confidence intervals; what I've done here treats the coefficient and cutpoint estimates as if they had no uncertainty.]

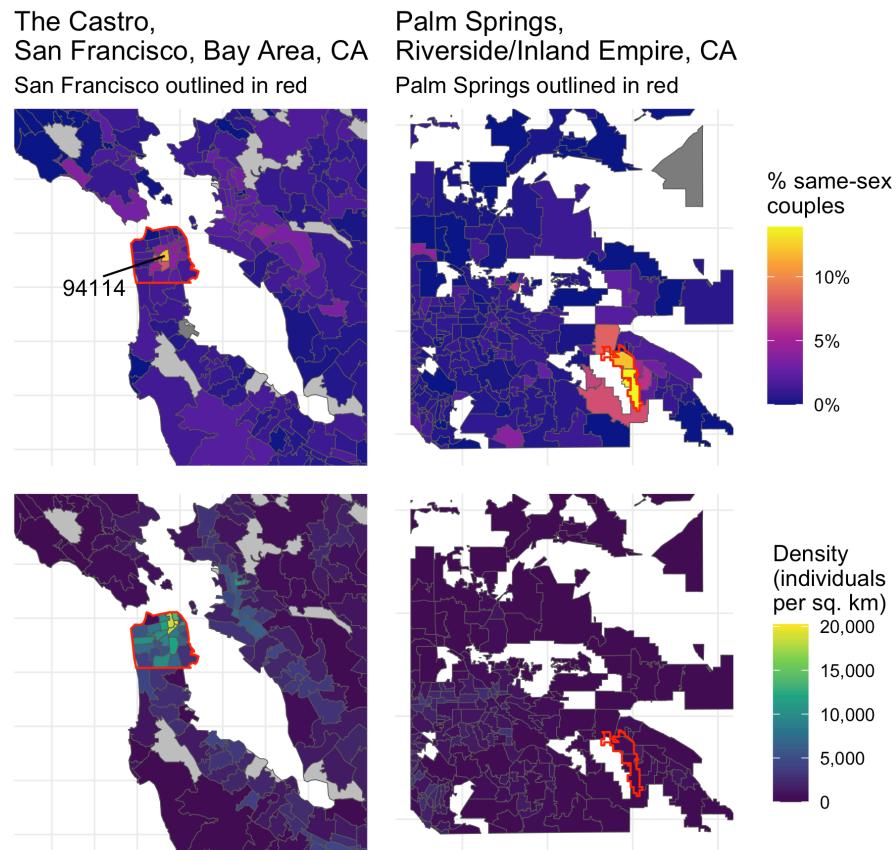


Figure 2.1: Example zip codes with high prevalence of same-sex couples: the Castro (94114) and Palm Springs (92262, 92264).

## Sense of community and zip code characteristics

Ordered logistic regressions

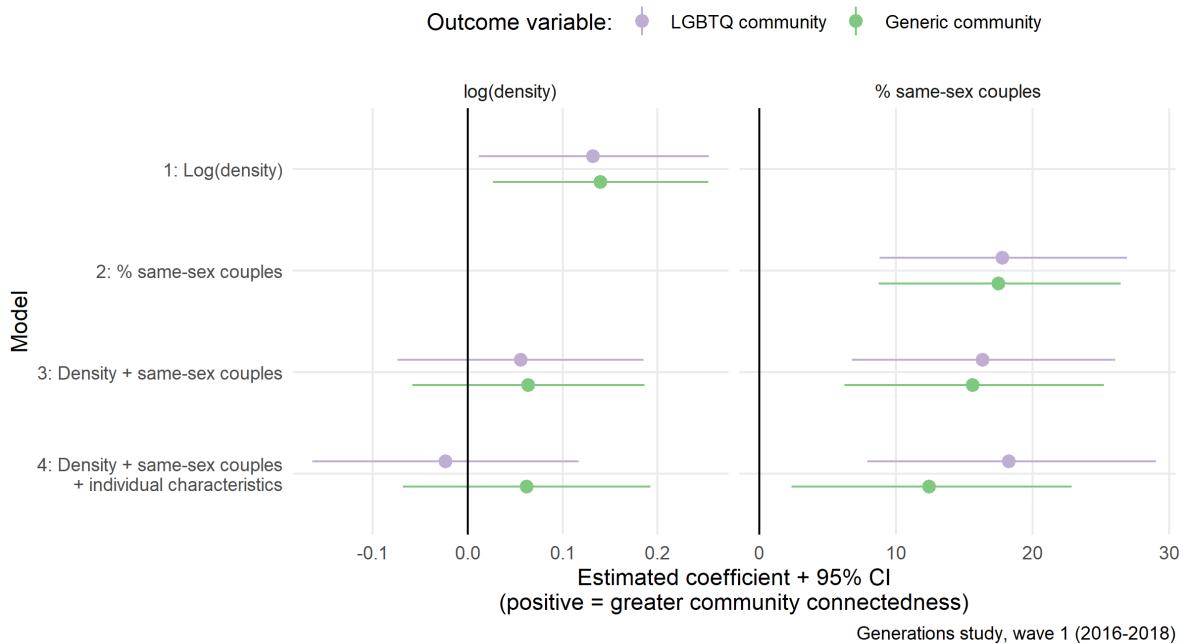


Figure 2.2: Density and prevalence of same-sex couples across 8 models.

Sense of community and individual characteristics  
Ordered logistic regressions

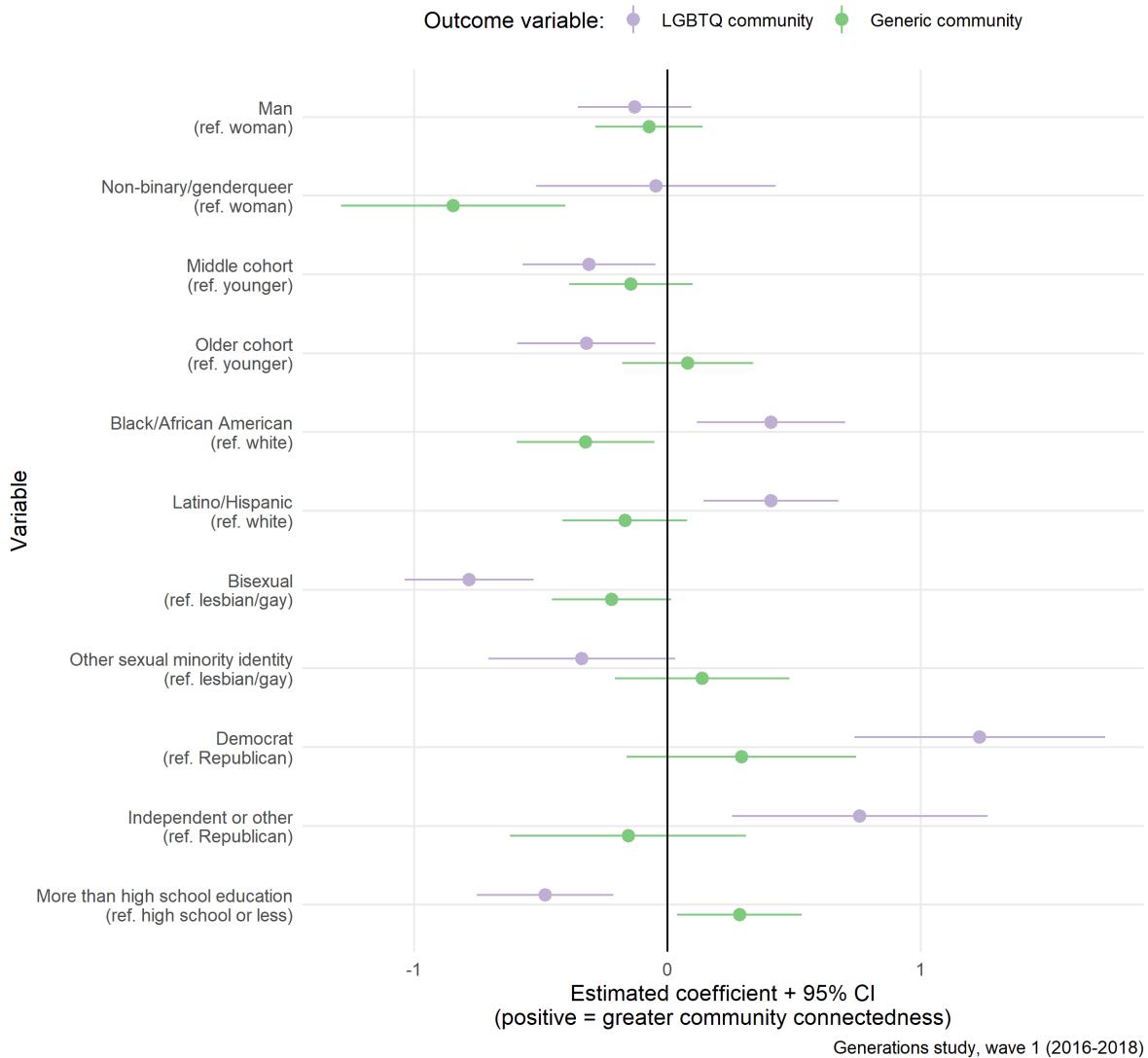
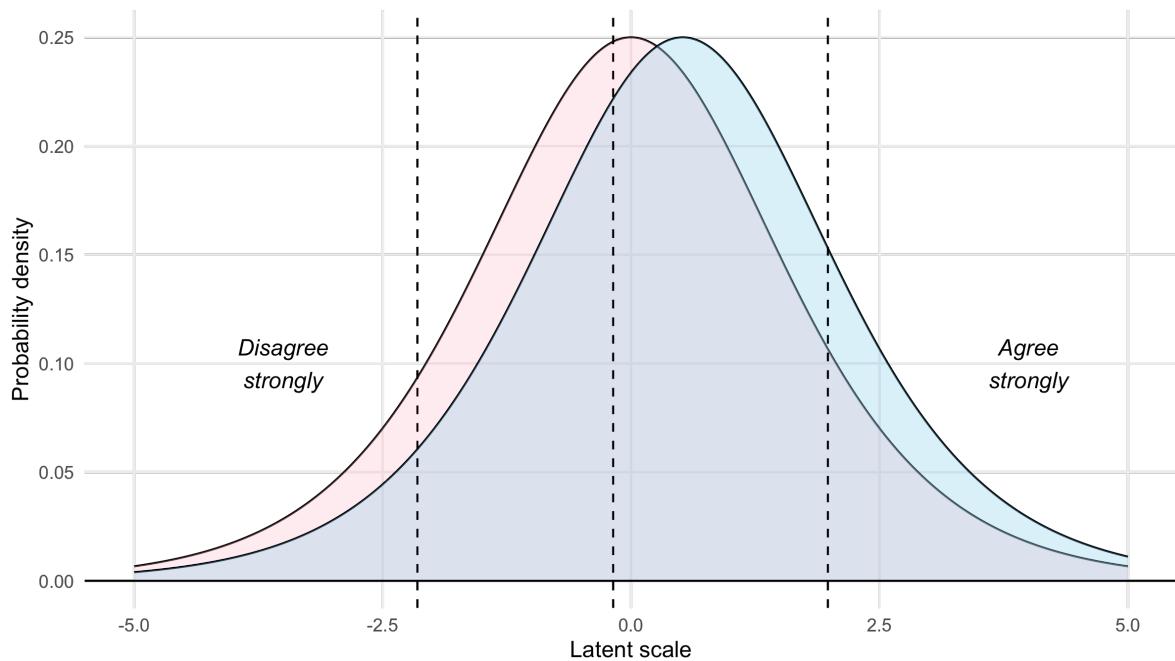


Figure 2.3: Individual control variables, 2 models.

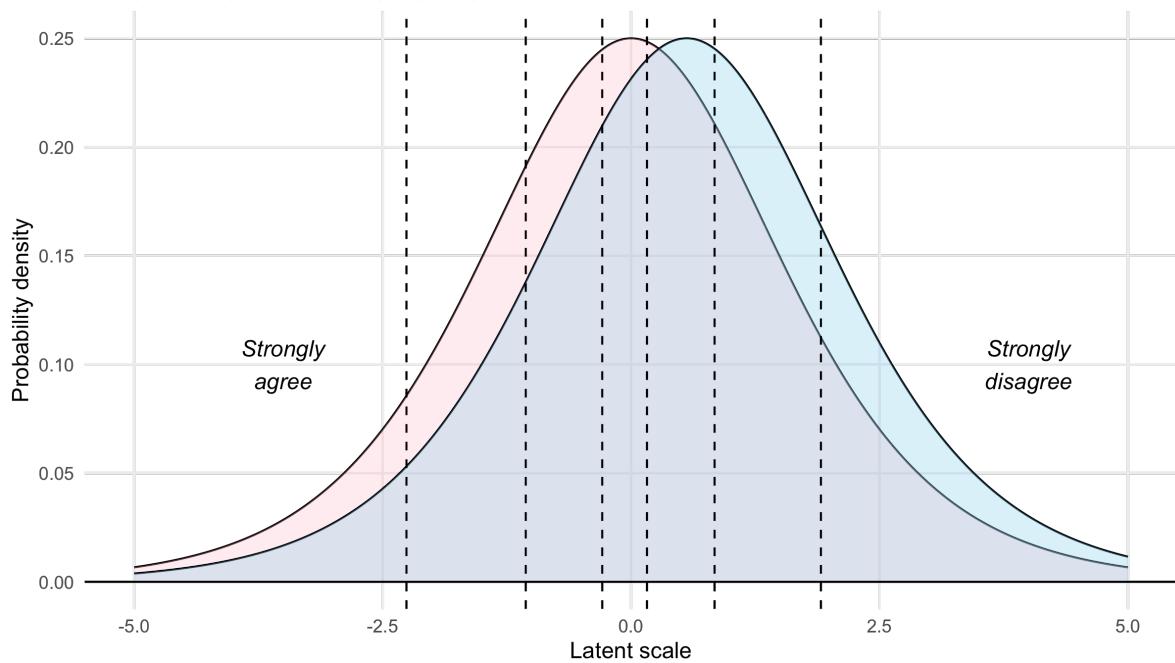
You feel you're part of the LGBT community.

Shift from 1 person/sq km to 10,000 people/sq km



I don't feel I belong to anything I'd call a community.

Shift from 1 person/sq km to 10,000 people/sq km

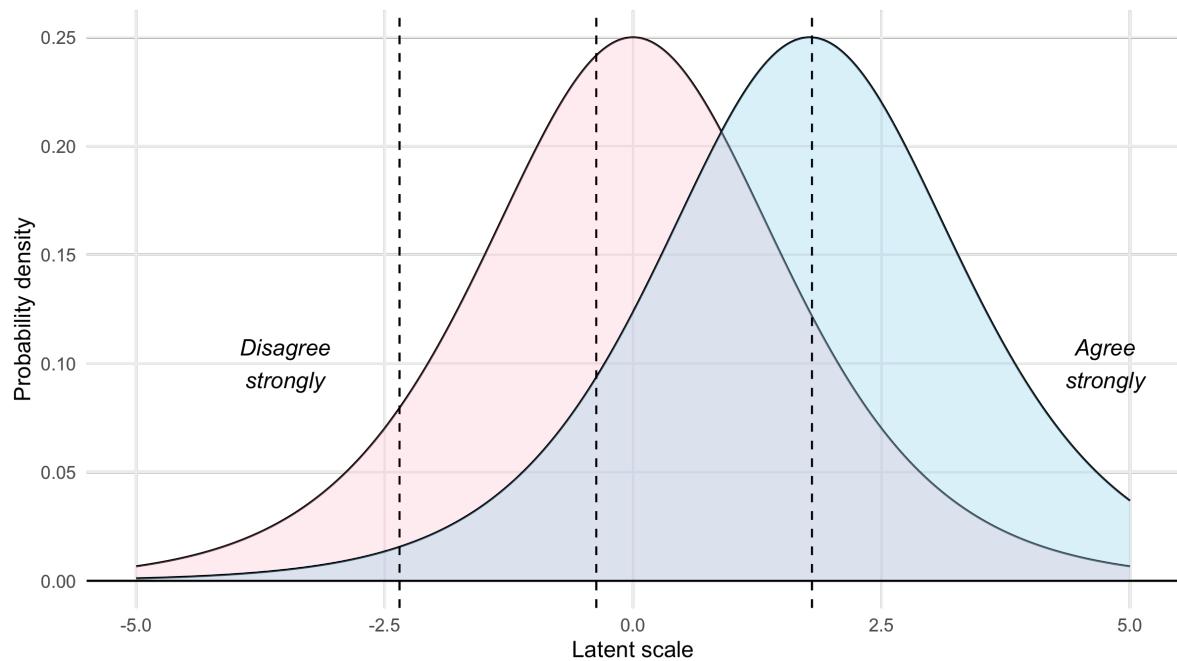


**Percent same-sex couples.** The percentage of same-sex couples in a respondent's

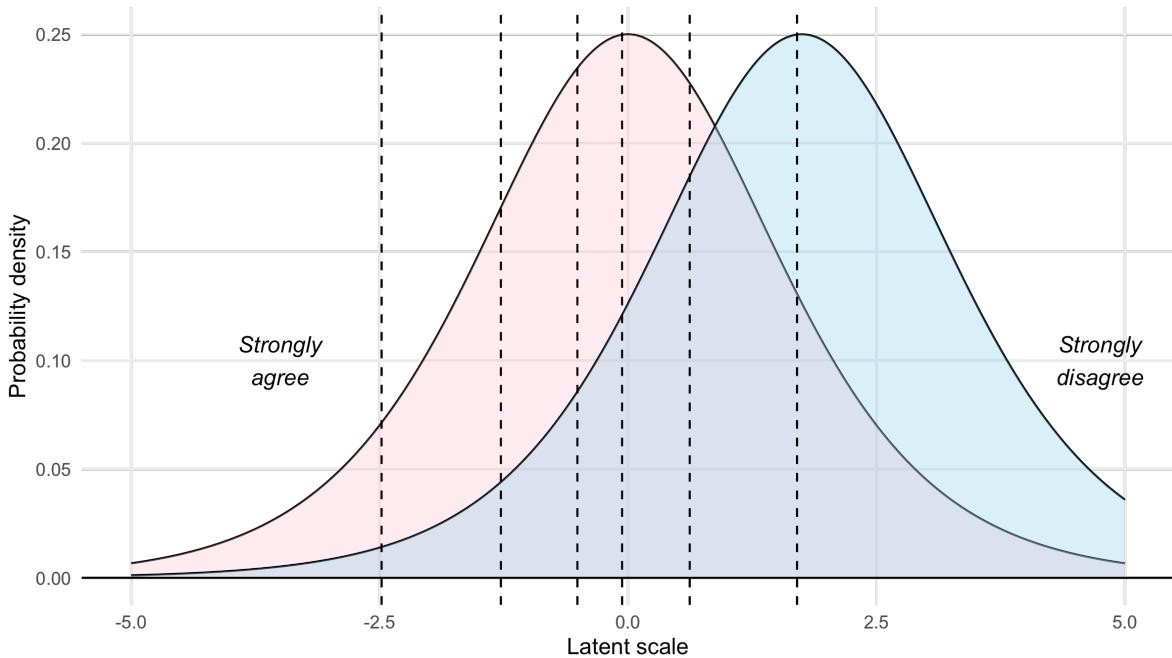
zip code is also positively associated with belonging to the LGBT community ( $b = 17.77$ ), and with belonging to any community in general ( $b = 17.49$ ). On the latent variable scales, a shift from 0% to 10% same-sex couples is predicted to move each outcome by 1.777 and 1.749 respectively. (Very few zip codes have a proportion of same-sex couple households as high as 10%; these are places like the Castro, or Provincetown, or Palm Springs.) That shift is nearly enough to move the mean predicted category for LGBT belonging to “Agree strongly.” For generic belonging, the mean category shifts from “Slightly disagree” to “Strongly disagree.” (The estimated cutpoints differ slightly from the previous models, so the mean category at baseline is not necessarily the same.)

You feel you're part of the LGBT community.

Shift from 0% same-sex couples to 10% same-sex couples



I don't feel I belong to anything I'd call a community.  
 Shift from 0% same-sex couples to 10% same-sex couples



**Both covariates together.** The estimated coefficient for density more than halves (to 0.06, for both outcomes) and is no longer statistically distinguishable from zero. By contrast, the estimated coefficient for percent same-sex couples changes little (decreasing slightly to 16.33 and 15.61 respectively). The estimated cutpoints shift somewhat but are relatively stable.

**Both covariates, with individual characteristics.** With six demographic control variables (gender, age, race, sexual identity, political affiliation, and education), the coefficients for the two place characteristics do not change much. The point estimate for the coefficient for density for the LGBT belonging outcome is actually *negative*, at -0.02, but indistinguishable from zero; for generic belonging it is 0.06 and still not statistically significant. The coefficient for percent same-sex couples is 18.26 (SE  $\pm 5.36$ ) for LGBT belonging and 12.43 (SE  $\pm 5.21$ ) for generic belonging. [TODO: is there a way to simulate or test whether the LGBT-belonging coefficient really does have a larger

magnitude?]

**Individual characteristics.** The individual controls themselves show associations with the two outcomes in some cases. Unlike with the two place characteristics, at times there are clear differences between the two outcomes, even differences in the signs of coefficients.

- Gender: no association with LGBT belonging; non-binary respondents report lower generic belonging than men or women, all else equal.
- Age: middle and older cohorts both report less LGBT belonging than the youngest cohort, all else equal. No association with generic belonging.
- Race: Black and Latinx respondents report higher LGBT belonging than white respondents, all else equal. Black respondents report less generic belonging than white respondents.
- Sexual identity: Bisexual respondents report less LGBT belonging than lesbian/gay respondents. Otherwise, no statistically significant differences.
- Political affiliation: Republicans report much lower LGBT belonging than Democrats or independents, all else equal. No statistically significant differences for generic belonging.
- Education: respondents with higher levels of education report *less* LGBT belonging but *more* generic belonging, all else equal.

## 2.5 Interpretation

It seemed like population density itself could be positively associated with community connectedness and belonging for LGB people — and it is, in a bivariate model. But that association does not hold up once other factors are taken into account. Other individual and place characteristics that are themselves associated with physical density

must mediate that association. And even in the bivariate case, there isn't a large difference in levels of belonging between the densest and least dense places.

By contrast, the prevalence of other LGBTQ people, as proxied with same-sex couples, is positively associated with community connectedness for LGB people. This remains true even when other factors are controlled for. Places with truly high proportions of same-sex couples are rare, but the models all predict that in these places respondents will report a heightened sense of community connectedness, both to the LGBT community and in general.

This means that the presence of other LGBTQ people specifically matters more than the mere concentration of people in general. (Those two factors are weakly correlated, even at the zip code level.) Given the extent that the environments of the densest urban places differ from suburban and rural locations, it is surprising that density has so little discernable association with sense of community. This is especially surprising given theory that urban life is thought to facilitate a diversity of subcultures [cite: Fischer] and that urban contexts have been key for the historical emergence of LGBTQ subcultures specifically [cite: D'Emilio, Bérubé, Orne, etc.]. But even if cities did in fact facilitate LGBTQ people coming together to form communities in the first place, cross-sectionally it is the current presence of those LGBTQ people, not the density of the place, that is highly important for experiencing a connection to community.

Individual characteristics, while not the focus of this project, are worth interpreting and contextualizing as well. (Trends for individual-level covariates are identical in models without the two place characteristics, so the incorporation of place-level covariates does not change the interpretation.)

Gender and sexual identity stand out as much for what *doesn't* matter as for what does. It is odd and noteworthy that the privileges accorded to cisgender gay men, in-

cluding their greater representation in and access to institutions like gay bars, etc., do not translate into a greater sense of being part of the LGBT community. Perhaps this is Winer's solidarity-with-disdain in action. It is less surprising that bi individuals report feeling less a part of the LGBT community than others; well-known phenomena of biphobia and/or bi erasure are likely to be at play here. Finally, while there are no differences between men and women, non-binary individuals experience less belonging *only* to community in general, not to the LGBT community specifically. It may be that LGBT spaces provide something of a refuge from a world that is otherwise rigidly organized around maintaining a gender binary – even if those spaces are often structured by gender as well. (A gender X sexual identity *interaction* might be worth investigating to more precisely tackle the question of differences in belonging between gay men, lesbians, and bi people and queer people of all genders.)

The remaining four covariates — age, race, political affiliation, and education — lead to a range of interpretations, from the unsurprising to the puzzling. It is not surprising that the youngest cohort feels more a part of the LGBT community than older cohorts, having come of age with greater societal acceptance, higher queer visibility, and more access to queer spaces (which privilege youth in a variety of ways). Nor is it surprising that political affiliation has such a strong association with LGBT belonging. The overwhelming majority of LGBQ respondents in the survey (and LGBTQ people in general) are Democrats or independents; to the extent that political values shape what the LGBT community is, it makes sense that Republican LGBT people would feel alienated from it. Next, given the realities of sexual racism [cite: Orne, Held, Stacey & Forbes, etc.], and the fact that many queer spaces – from gay bars to entire gay neighborhoods – are structured by whiteness, it would be reasonable to expect that queer people of color would feel *less* a part of the LGBT community. Actually,

however, Black and Latinx respondents report a greater sense of being a part of the LGBT community than white respondents (even though Black respondents report less belonging in general). I don't have a solid interpretation for this yet. Finally, even though LGBTQ people are more educated on average than their straight, cisgender counterparts [cite: Mittleman], LGBQ people with lower educational attainment report a greater sense of being a part of the LGBT community – but, paradoxically, less belonging overall. I might have thought that higher levels of education might offer more access to community, so I don't know how to make sense of these opposing trends. Of course, all of these individual covariates have their own correlations with each other, but the associations I've investigated so far hold up in bivariate analyses as well.

## 2.6 Next steps

I see two different ways to extend the work I've already done: deepen my existing analyses and probe them for robustness, or broaden the scope of the place-based characteristics I incorporate into my models.

For depth and robustness, I might first think more carefully about density. Am I confident that density doesn't matter? Or am I just not thinking about and operationalizing it the right way? A few concerns:

- First, the functional form of the association is potentially at issue. The log transformation is more data-driven than theoretically-driven, though it does capture the association with community connectedness better than a linear relationship. Maybe a quadratic model or something similar would better encode my theoretical expectations.
- Second, there's the issue of scale (i.e., the modifiable areal unit problem). Is

zip-code-level the right choice, or are higher-level scales more meaningful? Geographers and urban scholars advise that using MSA-level density per se isn't a valid choice, because differences in land area between counties (and therefore MSAs) are so stark between different parts of the US (the West and the rest, in particular). But the overall population or overall density people experience in an area might matter on top of or instead of the local experience in a zip code itself. Population-weighted densities get closer to what the average person experiences; it would be easiest to calculate these MSA-level values (or county-level values?) from zip codes, but feasible to start from tract or block levels instead – even though I don't know what tracts/blocks respondents actually live in. Of course, any MSA-level variables exclude the minority of respondents living outside MSA/ SA areas.

- Zip codes are, of course, an even weaker (too-large) proxy for neighborhoods than Census tracts, so if neighborhood-level density is what matters, then the lack of granularity in the survey geocoding may obscure the relationship between density and community as well. This is a problem if LGBTQ people systematically are more likely to live in the denser parts of their zip codes. (Looking at a handful of gayborhood examples – 98102/98122 in Seattle, 94114 in San Francisco, and 60657 in Chicago – I do believe that the gayborhoods systematically fall in the denser parts of those zip codes.)
- Finally, the range of the data might limit what can be observed about density. One of the challenges using data from the United States is that we don't observe more than a handful of really dense places. Even a city as dense as Seattle is mostly single-family, detached homes. If the benefits of density for community only take hold at San Francisco or New York-level densities (i.e., “walkable” densities), then

they'll be hard to observe from statistical models of nationally-representative US-based surveys. I had hoped a log transformation might address this, but it appears that it didn't. It may be that a different operationalization of urbanicity — a "big 4" or "central city" indicator — might be more relevant.

A second way to make this chapter deeper and more robust involves thinking more carefully about the prevalence/abundance of sexual minority people, and how this might relate theoretically and empirically to diversity/heterogeneity. To me, it's more intuitive that either high numbers or high proportions of LGBTQ people would be beneficial for a sense of belonging to the LGBT community than that *integration* between same-sex and different-sex couples would have that same kind of effect. At the same time, I don't want to undersell my finding. Affective attachment to LGBT community might have been independent of or even inversely related to the actual presence of other LGBTQ people; it could have been more of a symbolic than a material phenomenon. Prior research on LGBTQ people in small cities, or on solidarity with disdain in large metro areas, might have predicted a different result than what I actually found. I'm not sure how to go beyond prevalence/abundance and meaningfully account for any added impact of diversity. Empirically, zip-code-level entropy seems highly correlated with minority prevalence (I think that just mathematically how entropy works), and again the conceptual reason that MSA-level integration would have a positive effect on belonging isn't clear – maybe opposite effects on generic and LGBT-specific belonging, if we're thinking about assimilation. Other measures of diversity – racial/ethnic diversity, economic integration, etc. – might provide additional signals of a tolerant or welcoming environment, but there may be other ways of getting at those same characteristics. In addition to the conceptual and empirical issues, part of the challenge for me is methodological – I don't have a clear sense for how measures of entropy and indices of dissimilarity work,

exactly, or how it's reasonable to interpret them.

Instead of or in addition to focusing on other ways of understanding the two place covariates I've analyzed so far, I might expand the scope of place-based characteristics I include. My key innovation would be to incorporate cultural rather than demographic or physical features of places. This could be as simple as looking at residual variation between places in a multilevel model and chalking up unobserved differences to "culture," or it could involve an attempt to actually operationalize place-based culture.

My best idea so far for this is to take literally Brown-Saracino's (2017) qualitative observations about the power of place-based cultural narratives. She notes the affinity and resonance between her interviewees' stories about the places they live and the more official stories highlighted in travel guides and websites. I could search for and scrape text from those websites and then quantify the prevalence of different themes using a simple topic model. (This would resonate too with the precedent of using gay travel guides to study LGBTQ life.) Once again, the scale at which these narratives might be the most salient is unclear, but the (Census) place or city level seems most promising to start with. To frame this issue, for instance, how relevant are stories about Seattle to LGBTQ people who live in Bellevue or Tacoma?

In a similar vein – given the cultural dimension of politics, and given the importance of political affiliation at the individual level, place-based measures of political climate (e.g., 2016 Trump vote share) seem important to include. As with every place characteristic, there's a debate to be had about whether the zip code or MSA level is more salient.

Alongside adding covariates and reevaluating how I operationalize them, I might bring the individual traits further into focus by considering how they interact with place characteristics, as it's reasonable to think that they would. But, as my covariates and modeling strategies get more complex, I'll be more and more limited by the number

of respondents in the data. Of course, with only 1484 respondents, I don't observe every zip code in the country; though I do get more than one respondent per location, especially at higher levels of aggregation like MSA. (You could even imagine a spatial model estimating kind of a "risk surface" for belonging, as Jon Wakefield would put it, but that seems unnecessarily complex for little added benefit.) At some point, we'll have to feel confident that I've learned enough for one dissertation chapter, rather than trying every model specification imaginable.

## 2.7 Appendix: Tables

### 2.7.1 Outcome 1: You feel you're part of the LGBT community.

	Model 1	Model 2	Model 3	Model 4
log10(density)	0.13 *		0.06	-0.02
	(0.06)		(0.07)	(0.07)
pct_ssc		17.77 ***	16.33 ***	18.26 ***
		(4.61)	(4.90)	(5.36)
W1GENDER(2) Man				-0.13
W1GENDER(5) Non-binary/genderqueer				-0.05 (0.24)
COHORT(2) Middle				-0.31 * (0.13)
COHORT(3) Older				-0.32 * (0.14)
SCREEN_RACE(2) Black/African American				0.41 ** (0.15)
SCREEN_RACE(3) Latino/Hispanic				0.41 ** (0.14)
W1SEXMINID(2) Bisexual				-0.78 *** (0.13)
W1SEXMINID(3) Other sexual minority identity				-0.34 (0.19)
GP1(2) Democrat				1.23 *** (0.25)
GP1(3) Independent or other				0.76 ** (0.26)
GEDUC2(2) More than high school				-0.48 *** (0.14)
(4) Disagree strongly (3) Disagree	-2.15 *** (0.20)	-2.35 *** (0.11)	-2.21 *** (0.20)	-2.34 *** (0.34)
(3) Disagree (2) Agree	-0.18 (0.18)	-0.37 *** (0.07)	-0.23 (0.19)	-0.28 (0.33)
(2) Agree (1) Agree strongly	1.98 *** (0.19)	1.80 *** (0.09)	1.94 *** (0.19)	2.03 *** (0.34)
AIC	3587.52	3577.14	3578.42	3183.72
BIC	3608.73	3598.35	3604.94	3267.16
Log Likelihood	-1789.76	-1784.57	-1784.21	-1575.86
Deviance	3579.52	3569.14	3568.42	3151.72
Num. obs.	1484	1484	1484	1360

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

## 2.7.2 Outcome 2: I don't feel I belong to anything I'd call a community.

	Model 1	Model 2	Model 3	Model 4
log10(density)	0.14 *		0.06 (0.06)	0.06 (0.07)
pct_ssc		17.49 *** (4.50)	15.61 ** (4.82)	12.43 * (5.21)
W1GENDER(2) Man				-0.07 (0.11)
W1GENDER(5) Non-binary/genderqueer				-0.84 *** (0.23)
COHORT(2) Middle				-0.14 (0.12)
COHORT(3) Older				0.08 (0.13)
SCREEN_RACE(2) Black/African American				-0.32 * (0.14)
SCREEN_RACE(3) Latino/Hispanic				-0.17 (0.13)
W1SEXMINID(2) Bisexual				-0.22 (0.12)
W1SEXMINID(3) Other sexual minority identity				0.14 (0.18)
GP1(2) Democrat				0.29 (0.23)
GP1(3) Independent or other				-0.16 (0.24)
GEDUC2(2) More than high school				0.28 * (0.13)
(7) Strongly agree (6) Moderately agree	-2.26 *** (0.19)	-2.48 *** (0.11)	-2.32 *** (0.19)	-2.29 *** (0.32)
(6) Moderately agree (5) Slightly agree	-1.06 *** (0.18)	-1.28 *** (0.08)	-1.12 *** (0.18)	-1.08 *** (0.31)
(5) Slightly agree (4) Neither agree nor disagree	-0.29 (0.17)	-0.51 *** (0.07)	-0.34 * (0.17)	-0.30 (0.31)
(4) Neither agree nor disagree (3) Slightly disagree	0.16 (0.17)	-0.06 (0.07)	0.11 (0.17)	0.15 (0.31)
(3) Slightly disagree (2) Moderately disagree	0.84 *** (0.17)	0.62 *** (0.07)	0.78 *** (0.17)	0.85 ** (0.31)
(2) Moderately disagree (1) Strongly disagree	1.91 *** (0.18)	1.70 *** (0.08)	1.87 *** (0.18)	1.98 *** (0.31)
AIC	5628.47	5618.89	5619.84	5109.60
BIC	5665.59	5656.01	5662.27	5208.73
Log Likelihood	-2807.24	-2802.44	-2801.92	-2535.80
Deviance	5614.47	5604.89	5603.84	5071.60
Num. obs.	1485	1485	1485	1363

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

# 3 Geography or Gemeinschaft?

Disentangling the meaning of “community” through word embeddings

## 3.0.1 Short version

Chapter 1 asks, what can we learn about community from how the word is used in naturally-occurring language? This effort responds, in part, to a call from Levine (2017) to investigate the power and ambiguity of “community” in contexts other than local governance. My approach will be to systematically investigate connotations of community in texts generated by different groups of LGBTQ people, using word embeddings – a type of model that mathematically represent words based on the contexts in which they appear. The most basic mathematical model is an overlay of many different contexts; with some tweaks, word embedding models can compare different sets of contexts instead.

Either in the introduction or in this chapter, I plan to include more basic descriptive contextualization of “community,” e.g., keywords and N-grams over time. Though I think this is a case where underlying meanings are more of interest than surface-level words, readers will expect to see some of the latter.

As a proof of concept, I have already produced preliminary results come from a single off-the-shelf model trained on a general, generic corpus of text, meaning that there's nothing LGBTQ-specific about them – yet. Nevertheless, these initial models reveal that the spatial connotations of community are linguistically distinct from the social and cultural connotations of the word, while the latter two hang together, suggesting the existence of a geography-Gemeinschaft continuum. In other words, the theoretical distinction between shared culture and shared social ties does not appear in these generic results. (By contrast, in some of Winer's interview data, the “imagined community” and the world of everyday social ties are distinguishable.)

The empirical work I have already done illustrates the inherent breadth and ambiguity of the concept of community, but it doesn't directly address the adjacent question of valence – generally, empirically, is “community” something with purely positive connotations (Levine 2017), or can it be something more ambivalent (Winer 2020)? It does, however, suggest that the source for those sentiments is to be found on the Gemeinschaft side, not the geography side.

The preliminary word embeddings work I have already done will be further developed in the following ways:

1. Use the same basic framework to analyze texts from different groups, and compare results. This would allow me to measure variation in how community is used, and by extension, consider what community might “mean” to each. Based on feedback, I think this is the most promising direction for Chapter 1.
2. Apply some algebraic transformations and use the existing models as a tool that measures Gemeinschaft. This is how I intend to measure social and cultural expressions of community in Chapter 3, in order to examine the consequences of having a strong sense of community.

My current plan for Chapter 1 is to fit local word embedding models to informal written text, drawn from queer and trans Usenet groups (Dame-Griff 2019), and compare how the concept of community is used in these groups to the generic model I've already explored. This application is appealing for a few reasons. First, informal written text (McCulloch 2019) is the most logical source for analyzing linguistic expressions of community, because it directly shows how the meaning of community is constructed through specifically social interactions. Social interactions in Usenet groups were text-based, as opposed to increasingly the image-, sound-, and video-centric content of contemporary social media. Spoken language is more ephemeral and harder to collect and analyze; more formal written texts like news articles are further afield from the actual experience of community. (That said, LGBTQ media like *The Advocate*'s archives would be a reasonable fallback corpus.) Second, while no longer widely-used, Usenet groups are one of the earliest instances of the successful creation of vibrant virtual communities – and in fact, their existence offered critical evidence demonstrating that virtual community is possible at all. Because the two other chapters will draw on contemporary cases, having one that provides a more historical perspective would set up a more general foundation for the dissertation as a whole. For this chapter, the main source of comparative analytic leverage is the generic model I've already done a “close read” of; by comparison, I'd expect the geographic connotations of community to be minimal in a virtual context, and the *Gemeinschaft* connotations to be emphasized in an queer context. I will pay particular attention to whether the social and cultural dimensions of community are distinct in these expressions.

### **3.0.2 Long version**

“Community” – the literal word and the latent concept – is pervasive in everyday discourse. It appears in extremely varied and distinct contexts, everywhere from corporate mission statements to protest slogans. What is it about the meaning of community that makes it so amenable to these varied uses? What might it be doing when it does appear? I propose to investigate two interrelated questions: first, what is going on when people literally talk about “community”; and second, what is happening when people invoke the latent social concept that the word “community” points to, whether they literally use that word or not. I’m especially interested in how the empirical everyday meaning(s?) of “community” corresponds to or deviates from sociological understandings of *Gemeinschaft*, of community as a meso-level form of social organization. The computational method of word embeddings offers an empirical technique that is uniquely suited for investigating these questions, because – unlike other methods for computational text analysis – they move from surface-level words to underlying meanings and their relations (Arseniev-Koehler and Foster 2022; Stoltz and Taylor 2021). While the main aim of this project is descriptive and intrinsic, to provide a grounded account of the everyday social life of a social concept, it also furnishes a more precise tool for measuring when social actors implicitly or explicitly invoke community and belonging – in the *Gemeinschaft* sense – in the course of discursive action and interaction. I use a generic set of word embeddings to understand what “community” means in contemporary English-language discourse, and I apply algebraic transformations to show how the spatial and sociological connotations of the word can be disentangled. I’m still evaluating how best to use this new, resultant embedding as a measurement tool. I’ll discuss one preliminary experiment in applying that sociologically-inflected embedding for “community” to map out variation in community-oriented language in a historical set of virtual communities

called Usenet, as well as other possible social texts I might draw on instead.<sup>1</sup> At the end of this chapter proposal, I'll describe multiple alternative potential applications and what the payoffs of each might be.

Community has a power and ambiguity that render it suitable for strategic rhetorical uses in everyday discourse. Two qualitative examples illustrate the shades of meaning that “community” can take on; together, these illustrate the range of variation that I might expect to see in a computational investigation of what community means as a folk concept. First, in the context of local governance, Levine (2017) observes that “community” becomes a “floating signifier of the good,” a halo of positivity to cover the real operation of local decision-making and to provide legitimacy for action. For Levine, this is harmful; he notes the impossibility of “the community” wanting one single thing as a uniform entity, and the harms of ascribing collective representational authority to whoever can show up to participatory events (Levine 2021). In his case, the word is constantly used in a positive and justificatory light, but so flexibly as to lose coherent meaning. Second, and in contrast to that wholehearted positivity, Winer (2020) finds that his interviewees have an ambivalent and distancing relationship with the “imagined gay community,” drawing a distinction between “the community” at large and their own social circles. Rather than pure vagueness, this points to another specific rhetorical use, to critique an in-group’s flaws rather than to justify desired actions. In these accounts, “community” assumes differing valences, with slippery or counterintuitive referents, but in each case the concept does important discursive work. I do not aim to create a taxonomy of these rhetorical strategies; instead I’ll show how all of these uses together add up and contribute to the overall semantic resonance that “community” takes on.

---

<sup>1</sup>I've [experimented](#) with applying a toy version of half a dozen variations on word embedding methods to the word/concept of “community”. What I discuss here is what struck me as having the most potential.

Language, after all, is social and shared (Saussure [1916] 1972); later I'll show how that shared foundation can be a springboard for understanding local deviations.

For fully understanding the social life of a complex concept, academic definitions are insufficient on their own, but worth reviewing as an anchor for comparison. “Community” is a phenomenon sociologists have elaborated on since Tönnies ([1887] 2001); they've created taxonomies of different types of communities (Brint 2001); argued over what communities count as “real” (Driskell and Lyon 2002; Rheingold 2000); and debated about what features – shared social networks/interactions or shared cultural/moral traits – are most fundamental to the creation and experience of it (Boessen et al. 2014; Vaisey 2007). What emerges consistently is the metaphor of a tightly knit social fabric, a group of people bound together by shared ties, shared culture, and possibly shared place.

However, the issue with using academic definitions of “community” as a starting point is that a strict definitional logic of concepts and categories does not apply “out there” in the real social world; classical logic may be useful for technical jargon, but it is not how ordinary human concepts work. Instead, everyday concepts are fuzzy and prototype-based (Bowker and Star 2000; Lakoff [1987] 2008, [1987] 2008; Monk 2022; Rosch and Mervis 1975; Zerubavel 2002). One way to see this is to think about how people figure out that something *is* a community in the first place. As Bruckman (2022) argues, we might decide that a virtual community like Wikipedia is a community through mental comparison to prototypes of community like a small town. A given example of community does not have to have all of a specific set of features in order to fall under the concept; rather, concepts and categories are bundles of “intensions,” inherently fuzzy constellations of characteristics and cues. “Community,” as an everyday concept, bundles together a spread of connotations; as I will show, the most notable of these are

a geographic sense of “local place” and a sociological sense of “social group” or “object of belonging.”<sup>2</sup> Because “community” encompasses both, even the most mundane use of community to refer to local place might still invoke the sentiment and connotation of *Gemeinschaft*. That is partly where the fuzziness and ambiguity of community as a concept could come from, and also part of the concept’s discursive power. However, distinguishing these two senses is necessary in order to open up a window into where and how the underlying meaning of *Gemeinschaft* and belonging appears in different discursive contexts.

This project, then, uses an empirical, inductive, and computational approach to discover what community means and compare it to theoretical expectations derived from sociological literature. It integrates the sociological definition of *Gemeinschaft* in an iterative way, making this an abductive approach (Brandt and Timmermans 2021), rather than a purely grounded one (Nelson 2017). To examine the resonances and connotations of community in generic English discourse, word embeddings are my computational method of choice. Word embeddings are a relatively recent (Mikolov et al. 2013) computational operationalization of an old linguistic idea, called the *distributional hypothesis* (Sahlgren 2008). As Firth (1957) put it, “You shall know a word by the company it keeps.” Accordingly, these models represent words as a function of all of their immediate contexts. To give one example:

“The history of all hitherto existing **society** is the history of class struggles.”

(Marx [1848] 1972)

An embeddings model would take this sentence and learn about the semantic connotations of the word “society” from its position near “history”, “class”, and “struggle”;

---

<sup>2</sup>Importantly, having multiple connotations is not quite the same thing as the pure linguistic phenomenon of polysemy (Hamilton, Leskovec, and Jurafsky 2016), where a word has two totally separate meanings that context clearly disambiguates (e.g. “river bank” and “money in the bank”).

it might also learn linguistic features common to nouns from its position in relation to words like “is” and “of.”<sup>3</sup> Naturally, a model needs many such examples as training data, to produce a single overarching numeric representation for each word in a vocabulary. There are two main commonly-used word embeddings models – word2vec, based on a shallow neural network (Mikolov et al. 2013), and GloVe, based on cooccurrence matrix factorization (Pennington et al. 2014). (The differences are minor, so I choose one or the other based on convenience, not principle.) There are many other innovations and variations on the basic method; the most interesting is a family of models called *contextualized* word embeddings (e.g. BERT), which allow for more than one vector per word, instead of a single one – of course, these models take much more training data, and are much more complex! Simpler, older models offer a more straightforward and interpretable starting point; tackling the question of what “community” means in everyday English, in practice.

A generic model with a robust, comprehensive view of as many contexts as possible would approximate “the” meaning of every word in a language. To train general models for a given language, the most common corpora are large and publicly accessible texts from the Internet, e.g., Wikipedia pages, newswire articles, social media, or anything else that can be conveniently crawled from the web. (Historical embeddings use digitized book corpora.) These pretrained models can be used for a variety of questions and tasks, but they stand in contrast to locally trained models derived from specific corpora. Of course, meanings of words do vary – over time, over space, and by other social characteristics (Bamman, Dyer, and Smith 2014; Soni, Klein, and Eisenstein 2021). A general model trained on easily-accessible data works to the extent that meanings

---

<sup>3</sup>The size of the context “window” is an important parameter embeddings models can vary, as is the size of the vector representation that is output. For instance, 6 might be a reasonable window, 300 might be a reasonable embedding size (Pennington, Socher, and Manning 2014; Rodriguez and Spirling 2020).

are common or shared; of course, this flattens variation. Given the social characteristics of the authors of formal online texts like Wikipedia or news corpora, in terms of gender, race, education, nationality, etc. (Hargittai and Shaw 2015; Vrana, Sengupta, and Bouterse 2020), models trained on those data necessarily overrepresent hegemonic cultural viewpoints. This overrepresentation is a form of bias, but also a matter of substantive interest (Caliskan, Bryson, and Narayanan 2017; Garg et al. 2017; Jones et al. 2020).

Distinct from other text-as-data methods (e.g. keyword dictionaries or topic models), embeddings models create dense, distributed vector representations of words. In this way, word embeddings encode a relational model of meaning; they build up a system of signs (Saussure [1916] 1972), a vocabulary, in which distances (or their inverse, *similarity* measures) in a high-dimensional space can be calculated between every pair of words. This makes embeddings useful for social-science problems where meaning matters – especially where variations or changes in meaning are of interest; they've been applied fruitfully in cultural sociology (Kozlowski, Taddy, and Evans 2019) and political science (Rheault and Cochrane 2020). An embeddings model can be thought of as distilling shared, declarative public culture (Lizardo 2017); Arseniev-Koehler and Foster (2022) go even further to argue that the training process is a reasonable heuristic model for actual cultural cognition.

But the dimensions the models learn are not themselves interpretable (nor are they consistent across models).<sup>4</sup> To derive interpretable dimensions, social scientists use anchor words and simple algebra. One common approach is to construct new binary dimensions through subtraction (Kozlowski et al. 2019; Taylor and Stoltz 2020), opposing

---

<sup>4</sup>When comparing embeddings models, there are methods to *align* the embeddings spaces to facilitate comparison.

pairs of concepts that can be thought of as antonyms (e.g. rich - poor, woman - man).<sup>5</sup> This idea springs from the algebraic analogy tasks that first made word embeddings notable in NLP (e.g., king - man + woman  $\approx$  queen). While these binary oppositions have a clear basis in cultural sociology (Douglas 1966; Durkheim [1912] 2001; Saussure [1916] 1972), they are not the only possibility. They don't necessarily make sense for a concept that might bundle together multiple overlapping connotations or characteristics.

Instead, in this work, I deviate and borrow a different algebraic idea from NLP: “de-biasing” an embedding through orthogonal projection away from a target word vector (Gonen and Goldberg 2019). This approach originates in an attempt to mitigate gender bias in the words for different professions and occupations – which is undesirable for NLP tasks such as machine translation (Caliskan et al. 2017), even if it represents cultural associations or demographic facts about particular occupations that might be worth studying in themselves (Jones et al. 2020). That foundational work on this method makes it clear that it does not remove all the connotations of the undesired word – i.e., it does not fully succeed in de-biasing – but it is successful enough to use to disentangle the connotations of a concept like community.

I proceed from a standard set of word embeddings pre-trained on Wikipedia and newswire text (i.e., formal, written online English) and publicly released (Pennington et al. 2014). This is a generic set of embeddings, with all of the cultural biases that that necessarily entails. Starting with the vector for community, I select the 1000 words that are the nearest neighbors to it in the vocabulary, as measured by cosine similarity. I then decompose that subset of the embeddings space using principal components analysis, which I choose for its relative interpretability compared to other dimensionality reduction methods. I manually inspect the first several of these dimensions for poten-

---

<sup>5</sup>The mathematical reification of binary gender is one example where the shortcomings of binary axes are obvious; what if masculinity and femininity are thought of as orthogonal rather than opposed?

tial substantive interpretations of interest, and select the second dimension for further analysis as representing a contrast between geographic and sociological connotations. I average the vectors for the most extreme words on either end of that PCA dimension, then finally project the original “community” vector away from the geographic end (and, implicitly, toward the *Gemeinschaft* end), in effect “de-biasing” the new “community” vector of those connotations. I argue that this produces an embedding that means community in a purely sociological sense, rather than a spatial one.

Word embeddings encode a notion of similarity; a key way to understand what a given word means in a model is to examine the words that are most closely related to it. These are the words that would show up in similar contexts. For a robust picture of what community means, I examine and visualize the 1000 nearest neighbors, shown with example words labeled, in Figure 1. To project the 300-dimensional vectors down into the two dimensions shown in the figure, I decompose those vectors with PCA.<sup>6</sup> The first dimension (the x-axis in Figure 1) ranges from words like “if” and “we” and “not” to words like “not-for-profit,” “community-based,” “lgbt,” and “interfaith.” Based on the distributions of words along these dimensions, I label this first dimension as encoding a linguistic distinction between common, functional words and words that are more complex and substantive. While important for structuring the overall space of meaning, this distinction is not relevant for my analysis except potentially as a filtering mechanism. The second dimension (the y-axis in Figure 1), however, is more salient. Ranging from words like “town” and “located” to words like “cooperation,” “governance,” “organizations,” and “collective,” it encodes what I label a distinction between geography and *Gemeinschaft*. “Community” itself falls nearly in the middle between the two poles of

---

<sup>6</sup>The 300 dimensions of the embeddings space encode a substantial amount of subtle information that is lost with dimensionality reduction, so the proportion of variance explained by the first several dimensions is relatively low. Kozlowski et al. (2019) have shown similar results in an experiment with PCA and with explicit cultural dimensions.

this dimension.<sup>7</sup>

Drawing on each end of this geography-Gemeinschaft continuum, I select the 10 words (from the 1000-word neighborhood) that are the most extreme on either end. Figure 2 lists these two sets of words and shows that they do in fact fall into two distinct blocks – highly similar within each group, and highly distinct from the other group. By construction, the word “community” is highly similar to both groups – it quite literally bundles these two connotations together in a single concept. To produce a more robust vector measure for each underlying connotation of “geography” and “Gemeinschaft,” I average the 10 individual word vectors, as is common practice (Kozlowski et al. 2019; Waller and Anderson 2021).

Finally, using the “de-biasing” method I explained previously, I project the vector for “community” away from the averaged geography vector, creating a new concept vector I label community-as-Gemeinschaft. Figure 3 is a two-dimensional projection of this process that illustrates the results. In the figure, the x-axis represents similarity to the new community-as-Gemeinschaft vector; the y-axis represents similarity to the averaged geography words. By definition, each vector has a similarity to itself of 1, and the result of orthogonal projection is that the community-as-Gemeinschaft vector has a similarity to the geography vector of exactly 0. This has two consequences: community-as-Gemeinschaft remains very similar to the original community vector *and* to the averaged Gemeinschaft vector. An alternative approach – subtracting out the geography domain – does not result in a vector with the same properties.

The key outstanding question is what to do with what I’ve learned so far. How might I apply this novel, empirical understanding of “community” in generic English discourse to

---

<sup>7</sup>My results are robust to two checks: first, using 100 neighboring words instead of 100; second, using a separate set of embeddings trained from Twitter rather than Wikipedia data, using the same GloVe method.

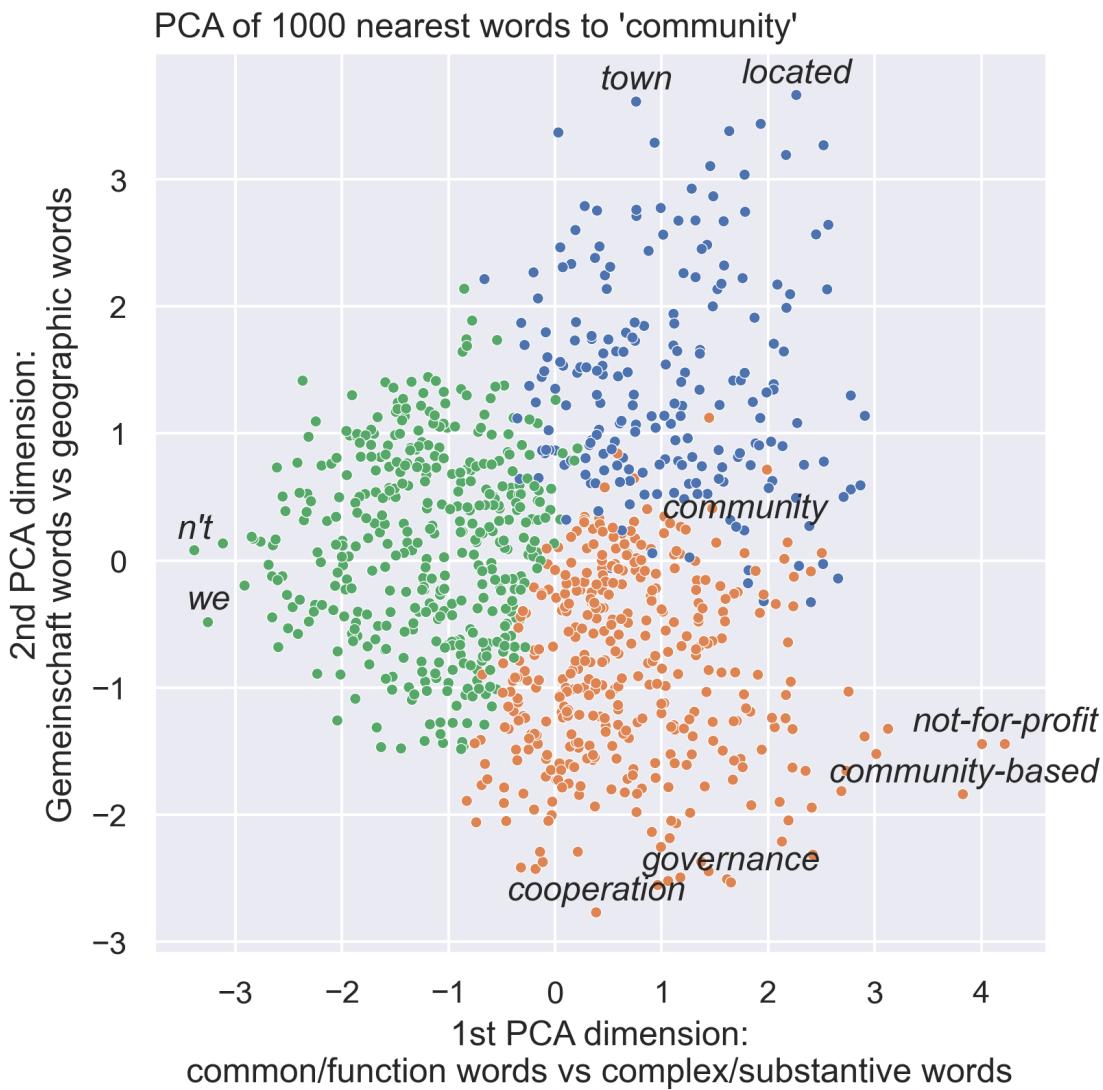


Figure 3.1: PCA decomposition of 1000 nearest word embeddings to “community”, showing the first two dimensions. While the space is continuous, k-means clustering with  $k = 3$  effectively divides it into functional words in green, geography words in blue, and Gemeinschaft words in red. Importantly, the clustering is applied to the original 300-dimension vectors, not the derived PCA results.

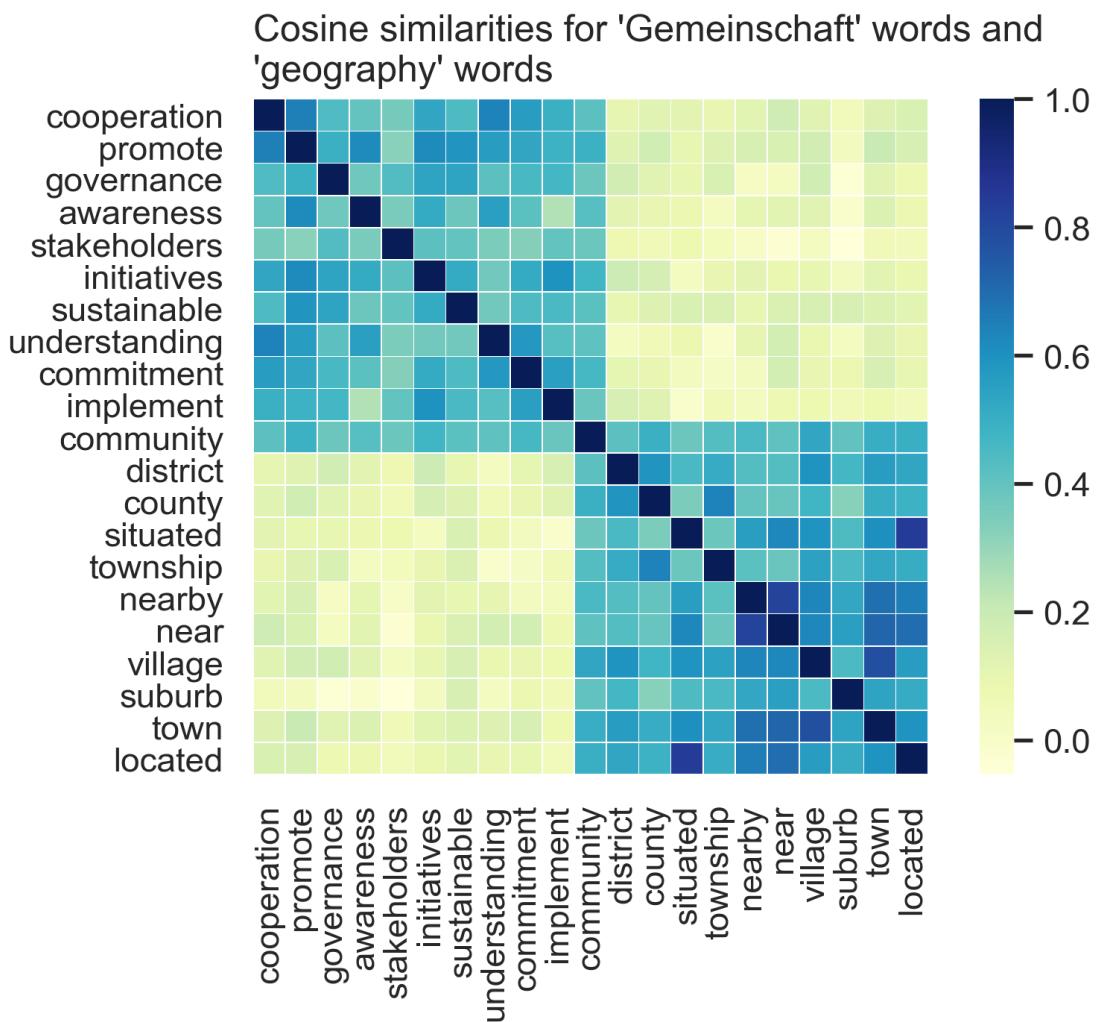


Figure 3.2: Similarities of the 10 highest and 10 lowest words along PCA dimension 2, which I have labeled geography – Gemeinschaft.

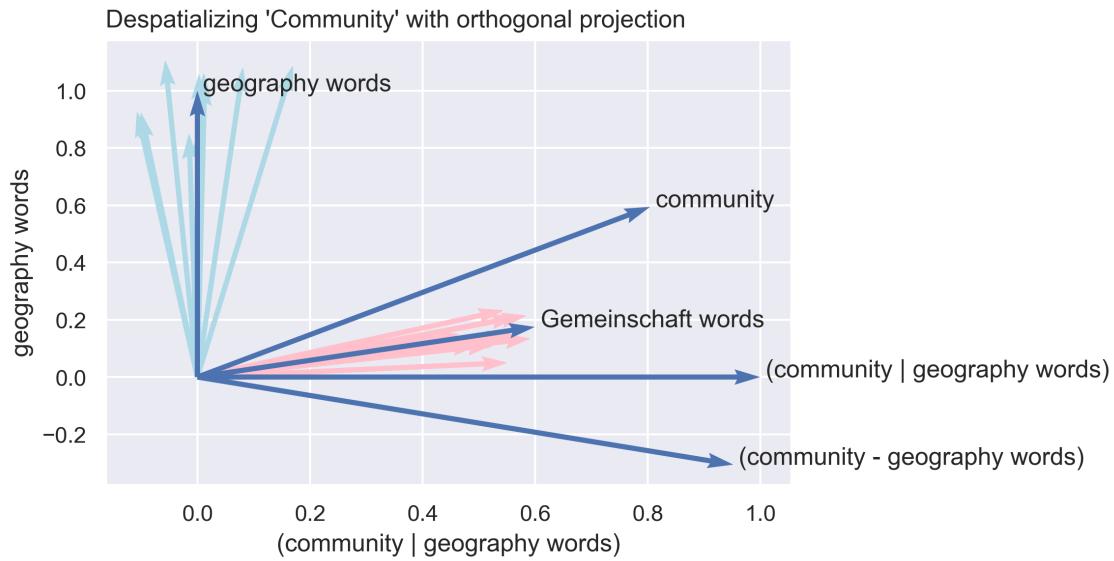


Figure 3.3: Orthogonal projection of the embedding for “community” away from the averaged vector of 10 geography-related words. The resultant embedding (community | geography) is highly similar to the averaged vector of 10 Gemeinschaft words. A binary opposition (community - geography) is shown for comparison, but is less similar to the Gemeinschaft words.

cases of sociological interest? I'll discuss several options. A first possibility is to use this derived vector as a measurement tool, to measure how much a text resonates overall with the concept of community in the *Gemeinschaft* sense. Uniquely, compared to other text-as-data methods, embeddings models uncover both explicit and implicit references, due to their relational nature. This application is what I've experimented with so far, on an easy-to-use test corpus that isn't necessarily the most compelling final choice. To move from individual words to longer texts – sentences or paragraphs or entire documents – some kind of aggregation method is necessary. Multiple such methods for summing or averaging word embeddings exists; one normalized or weighted method with advocates in sociology is called Word Mover's Distance (Kusner et al. 2015) or Concept Mover's Distance (CMD) (Stoltz and Taylor 2019). CMD calculates a distance metric between any given text and a target vector – in this case, the despatialized community vector representing community-as-*Gemeinschaft*. I applied this CMD measure to a corpus of texts from a set of virtual communities, with the idea that social text, where people are potentially building a community together through words, is where invocations of *Gemeinschaft* are most likely to appear and where they will have the most impact.

For convenience, I start by applying CMD to a public data set called 20 Newsgroups (Lang 1995); this is a collection of texts from 1993 from Usenet groups with distinct sets of topics ranging from religion and politics to computers and sports, and it's a conventional test data set for text-related machine learning classification problems (Albishre, Albathan, and Li 2015; Shalaby and Zadrozny 2017; Sia, Dalmia, and Mielke 2020, among many others; Silge and Robinson 2017; Wallach 2006). Usenet is one of the earliest virtual communities, a decentralized discussion platform that predates the internet (Rheingold 2000). Importantly, Usenet groups vary in terms of their cultures as well as in terms of their substantive content; some are friendly and welcoming and inspire

a sense of belonging in their members, while others are contentious and hostile (Baym 1994). Thus far, what I've been able to show is that there is significant between-group variation among the 20 newsgroups in how much their discussion posts resonate<sup>8</sup> with community-as-Gemeinschaft ( $R^2 = .19$ ); of course, there is substantial within-group heterogeneity as well. I haven't figured why this between-group variation exists: is it because some groups are more welcoming, and some more hostile, for instance? Is it a spurious artifact of topical differences? Ascertaining what drives that variation would take a closer read of the source texts (Nelson 2017); whether I decide to delve deeper into this particular corpus or not, it might provide a good reference case as I seriously interrogate the applicability of the CMD method.

A second possibility is, instead of using the community vector from the pretrained model on a corpus in a direct application, to fit local models to that corpus and then examine deviations from the meanings encoded in the generic, pretrained model – comparing models, in effect. To continue my example, what would “community” mean in a model that only saw it used in contexts from a place like Usenet? Given that Usenet is a virtual community, would the word have a geographic connotation at all? Or would it already be despatialized? That would show how early users of virtual platforms worked to remake and redefine community and establish their digital spaces as “real” communities; this trend might even be carried forward into contemporary virtual communities (of the sort I discuss in the Chapter 3 proposal). In different contexts, community might be more strongly tied to the idea of local place, even devoid of any sense of social organization; in still others, it might be associated more strongly with particular social identities. In any of these cases, comparing to both the overall generic embedding for “community” and the derived vector for community-as-Gemeinschaft would allow me to triangulate

---

<sup>8</sup>Stoltz and Taylor say “engage with” a concept instead of “resonate,” but I think they imply too much intentionality.

and obtain a clearer signal for how local relations of meaning are organized.<sup>9</sup>

In conjunction with either of the above approaches, I might tackle the question of valence directly: is community always inherently good, or does it shade into ambivalence in certain cases? Perhaps community has become increasingly positive over time, or perhaps it loses its affective charge as it becomes more ubiquitous – if community language has, indeed, become more prevalent. Addressing the valence and prevalence of community language requires reaching for text analysis tools beyond word embeddings. Embeddings-derived vocabularies of belonging might provide a robust set of keywords to trace in a descriptive way, and near-neighbor words in embeddings spaces can be combined with valence lexicons to measure positivity (Mendelsohn, Tsvetkov, and Jurafsky 2020). I might show, for instance, that the positive valence of community comes from the *Gemeinschaft* part of its connotation, where the geography words are more neutral. But particular methodological care must be taken to avoid spurious results (van Loon et al. 2022).

Finally, I'm not committed to using a corpus from a virtual community like Usenet or its contemporary successors; instead, as I develop the this project I'll consider which corpora might provide interesting sources of variation. The key consideration here is the correspondence between the sources of textual data and the conceptual questions of interest.

For instance, change over time is one conventional dimension of variation in social science applications of word embeddings (Kozlowski et al. 2019; Rheault and Cochrane 2020). However, my initial experiments with historical word embeddings (Hamilton et al. 2016) have so far turned up little evidence that “community” changes in any interesting

---

<sup>9</sup>In a similar vein, contextual embeddings models like BERT also offer a separate, distinct way to study local variation – but with even greater reliance on a complex pretrained black-box model derived from large data sets.

way over the course of the 20th century. It's still useful to outline multiple historical data sources for examining semantic changes over time, and note their advantages and shortcomings:

- Google Ngrams and other standard corpora used for historical embeddings have produced interesting general results, especially related to social biases and stereotypes encoded in language (Garg et al. 2017). Unfortunately, these have severe limitations in consistency post-2000, injecting bias into both descriptive frequency measures of terms and embeddings models themselves (Pechenick, Danforth, and Dodds 2015). This limits the use of these sources for studying contemporary changes that have taken place, like the rise of social media, that might have led to the fragmentation, proliferation, or some other reshaping of the meaning of community.
- Media articles from sources like the New York Times offer a different means of constructing embedding sets over time that extends from the 1980s to the 2010s (Arseniev-Koehler and Foster 2022; Mendelsohn et al. 2020), pushing temporal comparisons into an interesting time frame. However, I'm unconvinced that newspapers of record are where I will observe the most interesting invocations of community.
- By contrast to the New York Times, alternative and identity-focused media, like *The Advocate* and other long-running LGBTQ magazines, might be more salient, but would require me to assemble a corpus that has never been analyzed in this way before. Other researchers have assembled similar corpora from feminist movement literature (Nelson 2021a) and ethnic newspapers (Kim 2022) for forms of text analysis other than word embeddings; they've also trained embeddings models on 19th-century text data, namely abolitionist newspapers and narratives of lives

in the US South (Nelson 2021b; Soni et al. 2021). All of these have produced intriguing and important results, but in some cases these corpora were the bases for entire dissertations in themselves.

- Another comparison with a temporal dimension would be to take literally the notion that the concept “community” spans informal, everyday use and formal, academic use – and compare the two. In this case, I might ask what “community” would mean if a model learned its meaning inductively from sociological abstracts or social theory, and whether that meaning shifts to trace key intellectual currents like the Chicago School’s community ecology or the late 20th debates over “community lost” and the rise of virtual and identity-based communities. Kozlowski et al. (2019) have a supplemental analysis that does something like this with social class. Again, the lack of consistency in digitized corpora that aim to be generic or all-encompassing is a significant limitation in terms of what time frames can be compared.

Time isn’t the only covariate of potential interest. Social texts also offer the possibility of categorical comparisons between groups:

- Returning briefly to Usenet, 20 Newsgroups isn’t the only data source; there are historical archives of queer and trans Usenet communities (Dame-Griff 2019) that might contrast with more general topical newsgroups, and I might compare any of these to contemporary social media-based groups as well.
- Lastly, different kinds of corpora with distinct logics of production offer yet another possibility for constructing comparisons, and not necessarily temporally-based ones; for instance, corporate marketing versus political texts like speeches and activist literature. Intuitively, the latter might encode a clearer and more

authentic expression of *Gemeinschaft* into the meaning of “community” – but the tone of corporate slogans strikes me as eager to exude the positive valence that “community” generally has, and so the reverse could well be true.

Even if the data are unsettled, I have many possibilities for deploying word embeddings as an analytic and measurement tool to study variations in the meaning of and invocation of community.

## 3.1 Previous version

### 3.1.1 Introduction

Community is an ambiguous concept with both sociological and everyday uses. Understanding what is essential and what is incidental to the meaning of the term opens up the ability to analyze how much people engage with or invoke the underlying meaning of community in different social groups or social contexts. Experiences of community, participation, and belonging are important both intrinsically, as one facet of individual experiences in social groups, and extrinsically as one potential motivation for social action. Using a computational approach to text analysis, I offer an empirical understanding of how “community” is used in popular English-language discourse. I apply that understanding to a set of virtual communities to demonstrate how they vary.

In one view, community is a product of moral or cultural unity combined with social or interactional density (e.g. Tavory 2016, drawing on Durkheim). This way of thinking about community treats *Gemeinschaft* as the core element of community, and geographic proximity as incidental. As one line of evidence in favor of this view, extensive research has shown that virtual communities are real communities, in the sense that they really do instill a feeling of *Gemeinschaft* in their members (Driskell and

Lyon 2002; e.g, Rheingold 2000). Of course, geography still structures much of social life; propinquity may enable community without being synonymous with it (Spiro et al. 2016). Despite this theoretical distinction, the word “community” in everyday use may still entail geographic as well as sociological or psychological senses. Disentangling or disambiguating these senses is a core challenge for computational work. The payoff is that we might then be able to observe in real settings when people create, invoke, or experience a deep sense of community.

I use word embeddings to analyze the meaning of community. Embedding models are an ideal method for this problem, as opposed to other text-as-data techniques like descriptive frequencies of word counts or topic models, because they encode meaning relationally. I use the similarities between “community” and other words in tandem with sociological definitions of community to determine what meanings are being brought together in the vector representation of the word. Preliminarily, I rely on pretrained GloVe embeddings, originally developed from a large corpus of Wikipedia and other online data (Pennington et al. 2014).

Groups and individuals vary in the extent to which they experience a sense of community, and in the extent to which they communicate that sense of community through their language. I demonstrate this variation by adapting Concept Mover’s Distance (Stoltz and Taylor 2019) and applying it to posts from 20 Usenet newsgroups from the 1990s. These virtual communities were a precursor to present-day online and social media communities. Different Usenet groups have different cultures, and I show that average engagement with the concept of community differs between groups based on their topical focus. My initial findings are in line with my prior expectations in some cases, and diverge in others.

These analyses inform each other. The more theoretical analysis of the general dis-

cursive meaning of community is needed in order to interpret how people in different groups invoke that underlying meaning. The more empirical application to real communities not only shows the utility of the theory and method, but can allow me to refine it in an iterative process. In particular, I plan to use the theoretical dimensions I develop to difference out the geographic aspect of community in my empirical analysis.

This work makes two contributions. Substantively, it offers a new way of studying experiences of community across different groups, with potential applications in offline as well as online contexts. Methodologically, it combines prior sociological theory and text data to bring empirical clarity to a complex abstraction, with potential relevance for studies of other sociological concepts. I contribute to a growing literature adapting embeddings for sociological ends by extending the method to a concept that is rather less concrete and clear-cut. For instance, “community” is not quite like binary concepts such as social class (Kozlowski et al. 2019), health or morality (Arseniev-Koehler and Foster 2022), or conservative or liberal politics (Taylor and Stoltz 2020). Community has no one clear antonym. While one might conceptually oppose community to the individual, one could just as easily oppose community to society writ large. Disentangling the meaning of community with word embeddings requires making these sorts of theoretically-informed analytic choices.

### 3.1.2 Data and methods

Word embeddings are dense vector representations of words based on their contexts. These models were developed by natural language processing (NLP) researchers Mikolov et al. (2013), and have been imported into the social sciences, especially cultural sociology and political science (e.g. Kozlowski). I use pretrained GloVe embeddings, rather than training models locally on the Usenet data set. Generally, prior social science

researchers have found pretrained embeddings to be reasonably robust, stable, and generalizable Rodriguez and Spirling (2020), and I find similarly in supplemental analyses.

To measure engagement with community in at the level of individual Usenet posts, I use Concept Mover’s Distance (Stoltz and Taylor 2019). Concept Mover’s Distance is a sociological adaptation of Word Mover’s Distance (Kusner et al. 2015), which aggregates all of the word-level embeddings in a document and compares them to another. Because I use the Python implementation of Word Mover’s Distance in the gensim package, there are minor differences in the distance metric and algorithm; I expect the results reported below to be robust to such minor variations. I also do not standardize or invert the values I report.

Usenet is a distributed system for sharing electronic messages which predates the contemporary Internet, organized into topical groups such as alt.atheism or rec.motorcycles. Some of these groups are reported to have had a strong sense of community, while others were known for their hostility (Baym 1994; Dame-Griff 2019). This expected variation makes Usenet a compelling data source for analyzing how community-oriented different groups might be.

The 20 Newsgroups data set was compiled in 1995 and consists of nearly 20,000 messages approximately evenly distributed across those 20 groups (Lang 1995). It is a purposive sample of groups collected to represent diverse topical categories, originally intended for machine learning classification research (e.g. Dai and Le 2015) and also used as a teaching tool for topic modeling (e.g. Silge and Robinson 2017). These data are both convenient to obtain (through scikit-learn) and relevant, and so I repurpose them here for a more sociologically-oriented analysis.

### 3.1.3 Results

I analyze what, exactly, about the meaning of community is encoded in an embedding model by examining how similar words are to each other, according to the cosine similarity of their vectors. Figure 1 shows this schematically with a small set of words I selected. The axes of the figure are “similarity to ‘community’” and “similarity to ‘individual.’” “Neighborhood,” for instance, is far more similar to “community” than to “individual.”

More systematically, an initial qualitative analysis examining the nearest neighboring words to community (Figure 2) reveals both words related to geography (e.g. “local”, “area”, “residents”) and words related to *Gemeinschaft* (e.g. “organization”, “support”). PCA of a larger set of neighboring words ( $N = 100$ ) reveals similar trends, and formal clustering or dimension-creation is a logical next step.

In additional analyses, I show that the meaning of community is stable across corpora and over time. I follow Rodriguez and Stirling and confirm that the meaning of community has a moderately high correlation across embedding models pretrained on two independent data sources, Wikipedia and Twitter (Pearson correlation between cosine similarities = 0.636). I also confirm that the meaning of community is moderately stable over the course of the 20th century, using historical word embeddings pretrained with Google n-grams (Hamilton et al. 2016). These results increase my confidence that these pretrained embedding models encode a core definition of community and can be fruitfully applied to the Usenet data.<sup>10</sup>

Turning to those data, I apply Concept Mover’s Distance (CMD) to each of the posts from the 20 Usenet groups included in the data set ( $N = 18,296$  posts in total). For now, I simply measure the distance of each post from the undifferentiated concept

---

<sup>10</sup>These supplementary analyses can be viewed at <https://ccgilroy.github.io/community-discourse/>

“community,” though I plan to difference out the geographic dimension of community in subsequent refinements of the analysis.

The full distribution of CMD values in Figure 3 shows substantial overlap between Usenet groups, as well as substantial variation within groups. This is unsurprising, because each of these groups is a virtual community of some sort, and individual posts might be expected to vary in their tone and content substantially, just as individual experiences might vary.

At the same time, the different Usenet groups do in fact differ substantially and significantly in the average closeness of their posts to the concept of “community.” Figure 4 shows the predicted mean CMD for each group ( $R^2 = 0.133$ ,  $F = 147.17$ ,  $p = 0.0$ ). This simple statistical test shows that a small but substantial proportion of the variation in CMD is variation between different Usenet groups, rather than variation within groups. In some cases, this variation is in line with what we might expect topically: religion groups are more community-like, the *forsale* group is the least. In other cases, such as politics or sports, it may be that the specific norms and tone of the communities serve to generate a sense of community, regardless of how contentious or communal the topic might be.

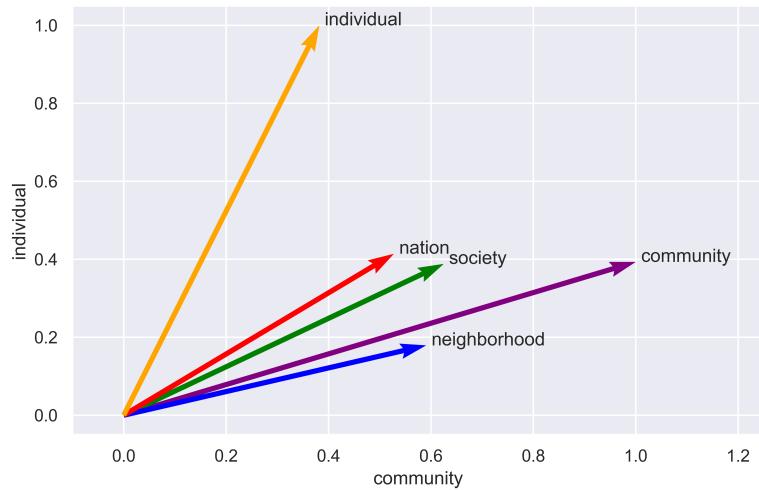


Figure 3.4: 2-dimensional projection of word similarities

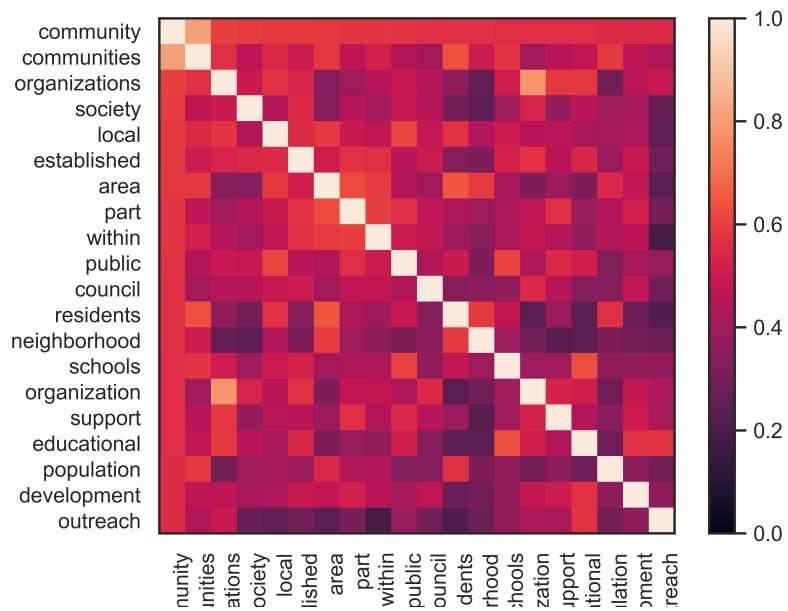


Figure 3.5: Most similar words to “community”,  $N = 20$

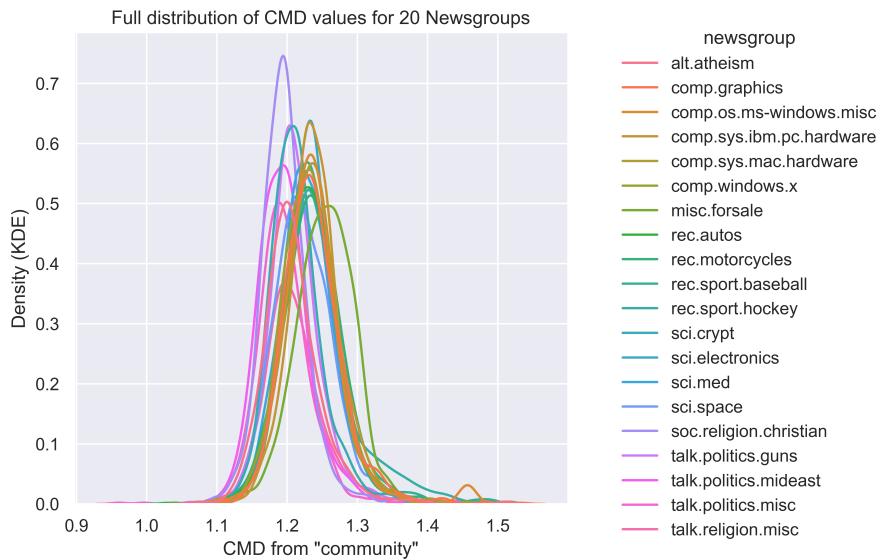


Figure 3.6: Full distribution of CMD values for 20 Newsgroups

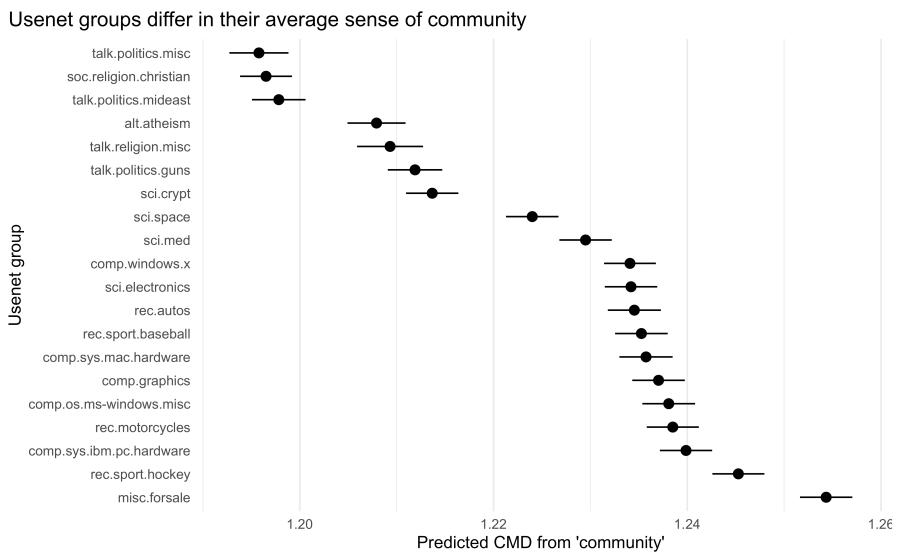


Figure 3.7: Usenet groups differ in their average sense of community

# 4 Community talk

Interactional contexts and consequences

Building on everyone's comments and Mako's offer of data, I'll use Reddit, not Twitter, as the starting point for Chapter 3. I really liked the point about thinking about group-level outcomes, though I definitely want to think more about measurable interaction-level outcomes (even if that's just, as someone mentioned, sustained conversation). I'll need to think about the comparative logic of case more – especially the intersections of place, identity, and activities/interests. I can think of a few example papers using different kinds of subreddits that I might draw on.

## 4.0.1 Short version

Chapter 1 shows that when actual people talk about community they often actually mean something very similar to the sociological concept of Gemeinschaft; chapter 2 explores the conditions under which people experience a sense of community in relation to an identity category. All of this provides evidence of the reality and salience of this sociological thing called “community” for LGBTQ people. But what happens when LGBTQ people seek community out, and express community in the process? I engage

with this question in Chapter 3, by using LGBTQ virtual communities to ask who produces community talk, and what happens when they do?

This chapter will make use of the direct measures of expressions of the social and cultural aspects of community developed in Chapter 1. This will be critical because, as chapter 2 is likely to show, for LGBTQ people, feeling connected to a community is not simply about conformity or sameness. I have chosen to develop my own measure to account for this, rather than reuse existing computational linguistic measures of belonging (Lucy and Bamman 2021), which equate belonging with conformity.

As in Chapter 2, a primary question is whether pro-community factors reinforce each other, or whether community becomes ambient and backgrounded once it has been created. In this chapter, the question is whether community talk is most prevalent among those who are most embedded or central in a group, perhaps performatively creating community for downstream and peripheral group members; or, is it the peripheral members of a group that engage in the most community talk in an effort to create their own sense of belonging? Those are two distinct possibilities for the sources and the potential impacts of expressions of community.

The ideal virtual community for studying community talk would have both network and interaction measures, the contents of conversations, and their eventual outcomes. I don't think perfect data are attainable; instead, different platforms have different downsides.

Based on initial feedback, I think Twitter comes closest to most of these measures. The problem, of course, is that "gay twitter" doesn't have clear, defined community boundaries. My main idea is to use an identity symbol as a proxy for community membership and a window into the network. (Alternatives like pride hashtags might be worth exploring but seem overly inclusive.)

The fallback plan here is Reddit, which lends itself to a slightly different set of questions; Reddit conveys less information about the embeddedness of individuals in networks of people, but rather in networks of groups. (It has the countervailing advantage that some LGBTQ-oriented groups on Reddit are broader and some are more specialized.)

#### **4.0.2 Long version**

Ultimately, “community” is something that people do – they build community, they create it. This chapter aims to study how the process of creating and invoking community plays out in social networks and social interactions. Social density of interactions and relations, not mere physical proximity, produces the social reality and the individual subjective experience of community, as well as observable expressions of the same.<sup>1</sup> But then in turn expressions of community help create the conditions for group-based social life. Building on the discursive analysis from chapter 1, and selecting a complementary context compared to the place-based analysis from chapter 2, I aim to use virtual communities as a site to study the relation between social density and expressions of community. However, the contemporary digital platforms that provide spaces for virtual communities are vast and varied, and I need help developing and crystallizing a concrete project that connects digital data sources to some facet of the underlying theoretical process. In this chapter proposal, I’ll lay out a few possibilities and articulate their tradeoffs.

The expression of community I plan to focus on is linguistic or textual, the kind of community talk or vocabulary of belonging that I explored in chapter 1. This could be explicit invocations of community and belonging that keyword methods can surface, or

---

<sup>1</sup>In conjunction with cultural or moral density.

broader language that more generally resonates with those concepts, of the kind that embeddings models can uncover (Stoltz and Taylor 2021). Text is key partly for measurement and analysis reasons; while image, voice, and video data are all also important aspects of the online social experiences that can create or sustain communities, the computational toolkit for using those types of data are less robust and more challenging to apply. In addition to the methodological motivation, there are also important theoretical reasons to focus on text and language, especially in the context of virtual communities that produce informal written language (McCulloch 2019). There are two ways to think about text: as something that matters in itself, or as a proxy for something else. What comes out of people’s mouths (or keyboards) isn’t the same thing as what’s in their heads. While one view might be that community-oriented language is only interesting as a proxy for the sense of community that individuals feel and perceive internally, in this case I’d argue for the importance of language on its own. This is because I think community talk can be performative, not merely expressive; what’s in people’s heads can’t do anything in the social world unless expressed in some way, and those expressions might have consequences regardless of what individuals really feel.<sup>2</sup>

A couple analytic considerations shape the possible research questions I might pose. First, do I focus on the causes or consequences of community talk? The structural/interactional formation of a community and the expressions of community that emerge from and in turn reinforce that social reality are a reciprocal, self-reinforcing process that unfolds over time.<sup>3</sup> Because of that, disentangling cause and consequence

---

<sup>2</sup>Of course, I can dream about having both survey measures of what someone perceives they are experiencing internally, and text measures of what they express in social contexts, in order to investigate to what degree those actually correspond, and whether “authentic” expressions operate differently from less authentic ones. I’ve fielded surveys through social media before, but that kind of measurement seems best left for future work. It would take resources I’m unlikely to have for my dissertation.

<sup>3</sup>i.e., this is a complex system. If need be, simulation might be a way to explore the dynamics better.

might be difficult. If possible, I want this chapter to focus on the consequences of invoking community, but I'm not committed to that direction yet. That focus would complement the work that chapter 1 and chapter 2 aim to do, by stepping through first what community means, then how community happens, and finally what community does. Still, it might be worth considering preceding contexts and downstream outcomes simultaneously if possible, depending on the exact form the data take.

Second, there is the matter of scale. There are two temporal scales or levels to consider when thinking about social density. Social density could be structural and relational, adhering in durable, culturally-recognized ties, like a friendship. Or it could be interactional, ephemeral, and activity-based, like a conversation. In a sense, of course, the latter coalesce into and constitute the former, or the former are an emergent, culturally-perceived (White [1965] 2008) property of the latter. The key analytic question is what each time scale lets me observe. It might be easier to observe the outcome of a conversation than to track the evolution of a friendship; it's possible to analyze networks dynamically but more likely I'll get a snapshot of a network than the full temporal dynamic. Interactions often unfold in the context of longstanding ties and/or durable groups, so taking that context into account is warranted regardless of my ultimate focal level of analysis.

With those analytic issues in mind, I'll sketch out a few potential research questions as possible scenarios, in order to make the possibilities inherent in different data sources more concrete. For instance, on the interactional level of a conversation, how might participants bring in or invoke community, whether implicitly or explicitly? People might be chatting back and forth in an amiable way that steadily builds positive emotional energy (Collins 2004), and that leads to more effusive or expressive community oriented language. Or, community might be invoked in contentious situations (perhaps by a

third party), in order to manage contention and steer people toward interactions with more positive emotional energy. In either case, what precedes those moments of heightened community-building? What's the outcome? Backing out further, to the relational level of a group, who are the people who consistently use high amounts of community-oriented language? Are they the most embedded? Accordingly, is combining network and discursive measures a way of identifying people who structurally *and* culturally play a key role in a group (Goldberg et al. 2016)? Or, are individuals on the margins the most effusive, to compensate for a lack of structural belonging and performatively create their own sense that they belong? I haven't committed yet to any one of those questions or the specific hypotheses they might entail; I expect to be able to refine and align at least one of those questions with some digital data source until they correspond well enough to constitute a feasible and interesting project.

Virtual communities provide a key site for observing how community emerge in interaction. Virtual communities are real sites for building community (Baym 1994; Driskell and Lyon 2002; Hampton and Wellman 2003; Rheingold 2000), often intertwined with individuals' offline lives as a kind of "augmented reality" (Jurgenson 2011; Orne 2017) rather than being completely distinct and separate (i.e., "digital dualism"). Moreover, the interactions that constitute virtual community building often happen *through* text (McCulloch 2019). By contrast, it's difficult to collect detailed social density data from offline interactions or ties alone (But not impossible, see Boessen et al. 2014). Another noteworthy aspect of digital spaces is that LGBTQ virtual communities are and have long been highly visible; queer people have been using digital technology to form connections, to "find" or "build" community with each other, since the virtual communities of the 1990s, like Usenet or the WELL (boyd 2014; Dame-Griff 2019; Rheingold 2000).<sup>4</sup>

---

<sup>4</sup>On sexuality and censorship online: one feature of Reddit and Twitter in particular vis-a-vis

Different digital platforms have distinct structures and affordances, which shape the ways people stage social interactions and the kinds of communities they build. Those affordances also shape the kinds of data I can collect, and ultimately mean I can learn different things from communities on different platforms. In addition, some platforms are better-studied and more convenient to analyze than others; those pragmatic considerations will also shape my case selection and project development. To make the possibilities and tradeoffs clearer, I'll discuss three platforms in detail: Reddit, Wikipedia, and Twitter. In each case, there are subcommunities of LGBTQ people to study, which have unique features but also reveal something about how each platform operates as a whole.

Reddit provides an obvious option for studying virtual communities, because the entire platform is structured around public groups (called *subreddits*). In other words, the key affordance of Reddit is the existence of groups in which to participate; these groups are the focal points for almost all interaction, which occurs in threads of posts and comments within subreddits. Posts and comments can be rated and ranked through upvotes and downvotes, providing a crowdsourced measure of quality (Medvedev, Lambotte, and Delvenne 2018). Direct structural ties between people are deemphasized; analyses of networks of Reddit users instead focus on the web of group affiliations (Olson and Neal 2015; Simmel 1971; Waller and Anderson 2019, 2021) created by ties of subreddit co-membership. In one sense, this group-based structure means that Reddit takes “groupness” for granted – i.e., it can appear to presume the existence of real community rather than showing how strong or cohesive a given community really is.

---

Facebook, Instagram, TikTok, Tumblr is that they have a more permissive stance toward sexuality (including pornography, but not necessarily other forms of sex work, especially post-SESTA/FOSTA (Blunt and Wolf 2020)). LGBTQ people frequently come up against the boundaries of censorship of sexuality on the latter set of platforms; they're on all of them, but that's a force pushing them toward the former.

Everyone who participates to any degree could be said to be a “member” of a “community”, which potentially drains those words of any deeper meaning. However, subreddits vary immensely, in size and activity level, but also in moderation efforts, adherence to local rules and norms, and other signals of distinctive group styles and subcultures. Users vary as well, in their levels of participation, how specialized and selective they are in engagement across different groups, and how embedded they are in within-group conversations.

Existing research leverages some of those variations. For instance, Zhang et al. (2017) characterize user engagement across a typology of community-level linguistic features; Lucy and Bamman (2021) use contextual word embeddings to study linguistic variation and conformity and identify semantically unique communities; and Waller and Anderson (2021) study polarization of subreddits through “community embeddings” based on co-membership. In the LGBTQ context, Reddit affords opportunities to observe community-oriented language across both large and generic (e.g. r/lgbt) and small and niche (e.g. r/LesbianGamers/) groups. Based on prior work, there are methods to operationalize individual trajectories, conversation “quality” proxies, and so on.

A second, perhaps less intuitively obvious option is Wikipedia. Wikipedia is a public collective representation of knowledge, but Wikipedia is also virtual community; people can come to have a strong sense of identity as Wikipedians (Pentzold 2011). This is a “peer production” community, rather than a purely social one. But Wikipedia mobilizes and socializes volunteer labor with strong norms and rules – to the point of being bureaucratic – because coordinating the collective production of knowledge is a massive, complicated endeavor (an endeavor the academic community can relate to). On Wikipedia, individual contributors can interact but are not linked by a durable and observable network structure; still, they vary substantially in their overall degree of

investment in Wikipedia as a project and community, measurable through time and labor. My primary interest in Wikipedia would be in the backstage interactional work that takes place on the “talk” pages associated with articles, where the outcome of those interactions is the production of public-facing representations. Especially for marginalized identity groups like LGBTQ people, the external stakes of representation are high; moreover, the well-known social biases of Wikipedia in terms of gender, nationality, and so on that emerge from its community of volunteers make clear that this representation is the outcome of a social process (Evans et al. 2020).

For instance, English-language Wikipedia has a “List of LGB people.” I’ve previously mapped the latent structure of the articles on this list, using document embeddings. As the t-SNE projection of these embeddings down into two dimensions reveals, these embeddings fractally map substantive domains (e.g., politics, sports, music), nationalities, and genders – grouping e.g. gay French writers together (Foucault alongside Barthes and Gide and Hocquenghem). Figure 8 shows a thematically labeled version of this map. (An interactive version of the figure which reveals more of its structure is hosted at [https://staff.washington.edu/cgilroy/lgb\\_wikipedia.html](https://staff.washington.edu/cgilroy/lgb_wikipedia.html)). Behind this structure, however, are real people collaborating (and contending) to produce it. They produce different representations with different valences than other Wikipedias in other languages (Park et al. 2020). Other researchers have curated data and produced tools like ConvоКit (Chang et al. 2020) for analyzing Wikipedia conversations; I might leverage and connect those tools to my own methods for surfacing community language and to the ultimate outcomes of the articles resulting from backstage conversations.

Twitter, the “model organism” of social media (Tufekci 2014), is the third major option. While numerous studies based on Twitter data already exist, the platform offers clear advantages for observing community in interaction: both durable friend and

follower ties as well as interactions like retweets, quote tweets, and replies are directly observable. In addition, the absence of a clear feature for demarcating the boundaries of groups is in some sense an advantage.<sup>5</sup> Unlike Reddit, the structure of Twitter doesn’t take “community” for granted, making Twitter a more flexible and less predetermined data source. While interactions are structured around feeds, timelines, and hashtags instead of groups, distinctive Twitter communities and subcultures do exist – “gay Twitter” among them.

But in order to study how sense of community and social density unfold in some slice of LGBTQ Twitter, I’ll need some sort of anchor. One idea I might propose is to use the adoption on Twitter of the rainbow flag, in emoji form (rainbow-flag-here), either as a window into a broader community, or merely as a way to demarcate a subset of Twitter users of intrinsic interest. While emoji in general can be thought of and analyzed in conjunction with informal written text (e.g., as a form of gesture, McCulloch 2019), rainbow-flag-here is not just any emoji. Rather, it’s a symbol, and a marker of identity. While new symbols proliferate constantly and variations abound, the six-stripe rainbow flag is widely-recognized to the point that it’s been formally adopted into the Unicode standard. Twitter users who adopt the rainbow-flag-here symbol as an identity marker in their bios thus constitute a visible subset of the set of LGBTQ people who use Twitter. And identity is grounds for building community: to the extent that those people belong to a distinct group or community we might call “gay Twitter”, I’d expect them to be connected to each other by durable friend/follower ties and interaction-based ties like replies, retweets, and reactions. These social connections should be disproportionately dense, relative to some (as-yet-undetermined) baseline, in

---

<sup>5</sup>Of course, there are distinct conceptual and methodological challenges around boundaries that aren’t pre-given. This is one reason many studies of community, social capital, and civil society focus on formal organizations and their membership rather than informal community-building, even though formal organizations are clearly only the tip of the iceberg.

a way that network community detection algorithms or clustering metrics would be able to pick out. And from those interactions, sense of community and group culture would be evident and measurable through the same computational text analysis toolkit I've discussed previously.<sup>6</sup> Compared with the previous potential cases, the outcomes of interactions over time are less clearly defined, but Twitter presents an opportunity to observe a dynamic system as it unfolds.<sup>7</sup>

However, those three platforms don't exhaust the possibilities for using digital spaces to understand how community operates. In particular, I doubt that they're the contemporary virtual spaces that display the strongest instances of community expression and interactional dynamics. For that, I'd look instead to platforms that facilitate chat or conversation in a fluid, informal way – a dynamic, rapid temporality lending itself more toward experiences of emotional energy typically associated with co-presence in physical space. This means Discord servers, Slack channels, even Twitch streams or Gather instances or Messenger group chats. Many of those examples are more private, backstage, sometimes ephemeral, and so also offer participants opportunities to express themselves differently; maybe in a less deliberately curated way and in a way that I would think more revealing of interactional dynamics.<sup>8</sup> Concerns about privacy, ethics, and access limit the feasibility of exploring these alternatives now, but I would not discount them entirely.

Reddit, Wikipedia, and Twitter present three distinct options for studying social density and community language in an LGBTQ context. These platforms are conventional, well understood, and publicly accessible. This has advantages: the ethical issues

---

<sup>6</sup>Some small fraction of users/tweets will be geolocated as well – which might present a point of connection to the chapter 2 project about place characteristics.

<sup>7</sup>Especially if I'm able to collect data over a long period of time.

<sup>8</sup>Some of these newer platforms start to bridge from informal written text to voice and video communication. That multimodality is no doubt an asset for creating a sense of community even as it renders measurement and observation much more complex.

and risks of potential additional harm are lessened when many researchers have already studied communities on these platforms, and the power dynamics of being an insider or an outsider to those spaces in varying degrees are perhaps easier to navigate as well. With these platforms, it's more likely I'll be able to construct a comprehensive quantitative picture of a process, and the ways these platforms fail to represent broader populations are also well known. Table 3.1 summarizes the relevant characteristics of Reddit, Wikipedia, and Twitter and how research projects focused on social dynamics and LGBTQ community might be developed from each of them.

:::

::: .cell-output-display

**Table 3.1: Key platforms for studying LGBTQ virtual communities**

Platform	LGBTQ Communities	Social density			Project anchor
		Ties	Interactions	Outcomes	
Reddit	LGBTQ subreddits (many)	Subreddit co- membership	Posts and comments on threads within subreddits	Conversation "quality", up- votes/downvotes	General vs specific LGBTQ subreddits(?)
Wikipedia	LGBTQ Wikipedians (language- specific)	None(?)	Talk page discussions	Conversation "quality", article representations	List of LGB People
Twitter	"gay Twitter" (porous boundaries)	Friends and followers	Tweets and replies, quote tweets, retweets	Conversation "quality", reactions, persistence of ties(?)	Symbolic anchor of rainbow-flag in bio

:::

::: .cell-output-display

# 5 Conclusion

[TODO]

# References

- Abascal, Maria, and Delia Baldassarri. 2015. “Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined.” *American Journal of Sociology* 121(3):722–82. doi: [10.1086/683144](https://doi.org/10.1086/683144).
- Albishre, K., M. Albathan, and Y. Li. 2015. “Effective 20 Newsgroups Dataset Cleaning.” Pp. 98–101 in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. Vol. 3. doi: [10.1109/WI-IAT.2015.90](https://doi.org/10.1109/WI-IAT.2015.90).
- Anderson, Benedict. [1983] 2016. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Arseniev-Koehler, Alina, and Jacob G. Foster. 2022. “Machine Learning as a Model for Cultural Learning: Teaching an Algorithm What It Means to Be Fat.” *Sociological Methods & Research* 51(4):1484–1539. doi: [10.1177/00491241221122603](https://doi.org/10.1177/00491241221122603).
- Baldassarri, Delia, and Maria Abascal. 2020. “Diversity and Prosocial Behavior.” *Science* 369(6508):1183–87. doi: [10.1126/science.abb2432](https://doi.org/10.1126/science.abb2432).
- Bamman, David, Chris Dyer, and Noah A. Smith. 2014. “Distributed Representations of Geographically Situated Language.” Pp. 828–34 in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics. doi: [10.3115/v1/P14-2134](https://doi.org/10.3115/v1/P14-2134).

- Baym, Nancy K. 1994. "From Practice to Culture on Usenet." *The Sociological Review* 42:29–52. doi: [10.1111/j.1467-954X.1994.tb03408.x](https://doi.org/10.1111/j.1467-954X.1994.tb03408.x).
- Berlant, Lauren, and Michael Warner. 1998. "Sex in Public." *Critical Inquiry* 24(2):547–66.
- Bérubé, Allan. 2011. *My Desire for History: Essays in Gay, Community, and Labor History*. edited by J. D'Emilio and E. B. Freedman. University of North Carolina Press.
- Bishop, Bill. 2009. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Houghton Mifflin Harcourt.
- Blunt, Danielle, and Ariel Wolf. 2020. *Erased: The Impact of FOSTA-SESTA & the Removal of Backpage*. Hacking//Hustling.
- Boessen, Adam, John R. Hipp, Emily J. Smith, Carter T. Butts, Nicholas N. Nagle, and Zack Almquist. 2014. "Networks, Space, and Residents' Perception of Cohesion." *American Journal of Community Psychology* 53(3-4):447–61. doi: [10.1007/s10464-014-9639-1](https://doi.org/10.1007/s10464-014-9639-1).
- Boutyline, Andrei. 2017. "Improving the Measurement of Shared Cultural Schemas with Correlational Class Analysis: Theory and Method." *Sociological Science* 4:353–93. doi: [10.15195/v4.a15](https://doi.org/10.15195/v4.a15).
- Boutyline, Andrei, Devin Cornell, and Alina Arseniev-Koehler. 2021. "All Roads Lead to Polenta: Cultural Attractors at the Junction of Public and Personal Culture." *Sociological Forum* n/a(n/a). doi: [10.1111/soef.12760](https://doi.org/10.1111/soef.12760).
- Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. n edition edition. Cambridge, Massachusetts London, England: The MIT Press.
- boyd, danah. 2014. *It's Complicated: The Social Lives of Networked Teens*. New Haven:

Yale University Press.

Brandt, Philipp, and Stefan Timmermans. 2021. “Abductive Logic of Inquiry for Quantitative Research in the Digital Age.” *Sociological Science* 8:191–210. doi: [10.15195/v8.a10](https://doi.org/10.15195/v8.a10).

Brekhus, Wayne. 2003. *Peacocks, Chameleons, Centaurs: Gay Suburbia and the Grammar of Social Identity*. University of Chicago Press.

Brint, Steven. 2001. “Gemeinschaft Revisited: A Critique and Reconstruction of the Community Concept.” *Sociological Theory* 19(1):1–23.

Brown-Saracino, Japonica. 2017. *How Places Make Us: Novel LBQ Identities in Four Small Cities*. 1st ed. Chicago ; London: University of Chicago Press.

Bruckman, Amy S. 2022. *Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge*. Cambridge: Cambridge University Press.

Butler, Judith. 2004a. “Bodies and Power Revisited.” Pp. 183–96 in *Feminism and the Final Foucault*, edited by D. Taylor and K. Vintges. University of Illinois Press.

Butler, Judith. 2004b. *Undoing Gender*. New York ; London: Routledge.

Calhoun, Craig. 1998. “Community Without Propinquity Revisited: Communications Technology and the Transformation of the Urban Public Sphere.” *Sociological Inquiry* 68(3):373–97. doi: [10.1111/j.1475-682X.1998.tb00474.x](https://doi.org/10.1111/j.1475-682X.1998.tb00474.x).

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases.” *Science* 356(6334):183–86. doi: [10.1126/science.aal4230](https://doi.org/10.1126/science.aal4230).

Chang, Jonathan P., Caleb Chiam, Liye Fu, Andrew Z. Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. “ConvoKit: A Toolkit for the Analysis of Conversations.” *Proceedings of SIGDIAL* 4.

Collins, Randall. 2004. *Interaction Ritual Chains*. Princeton, NJ: Princeton University

Press.

- D'Emilio, John. 1992. "Capitalism and Gay Identity." in *Making Trouble: Essays on Gay History, Politics, and the University*. New York: Routledge.
- Dai, Andrew M., and Quoc V. Le. 2015. "Semi-Supervised Sequence Learning."
- Dame-Griff, Avery. 2019. "Herding the 'Performing Elephants': Using Computational Methods to Study Usenet." *Internet Histories* 3(3-4):223–44. doi: [10.1080/24701475.2019.1652456](https://doi.org/10.1080/24701475.2019.1652456).
- Douglas, Mary. 1966. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. Routledge & K. Paul.
- Driskell, Robyn Bateman, and Larry Lyon. 2002. "Are Virtual Communities True Communities? Examining the Environments and Elements of Community." *City & Community* 1(4):373–90. doi: [10.1111/1540-6040.00031](https://doi.org/10.1111/1540-6040.00031).
- Durkheim, Émile. [1893] 1997. *The Division of Labor in Society*. New York: Free Press.
- Durkheim, Émile. [1912] 2001. *The Elementary Forms of Religious Life*. Oxford University Press.
- Eliasoph, Nina. 1998. *Avoiding Politics: How Americans Produce Apathy in Everyday Life*. Cambridge University Press.
- Eliasoph, Nina, and Paul Lichtenman. 2003. "Culture in Interaction." *American Journal of Sociology* 108(4):735–94. doi: [10.1086/367920](https://doi.org/10.1086/367920).
- Evans, Siân, Jacqueline Mabey, Michael Mandiberg, and Melissa Tamani. 2020. "15 What We Talk About When We Talk About Community." in ::wikipedia @ 20.
- Firth, John R. 1957. "A Synopsis of Linguistic Theory, 1930-1955." *Studies in Linguistic Analysis*.
- Fischer, Claude S. 1975. "Toward a Subcultural Theory of Urbanism." *American Jour-*

- nal of Sociology* 80(6):1319–41.
- Forstie, Clare. 2020. “Theory Making from the Middle: Researching LGBTQ Communities in Small Cities.” *City & Community* 19(1):153–68. doi: [10.1111/cico.12446](https://doi.org/10.1111/cico.12446).
- Foucault, Michel. 1998. “Friendship as a Way of Life.” Pp. 135–40 in *Ethics: Subjectivity and Truth*, edited by P. Rabinow. New York: The New Press.
- Frost, David M., and Ilan H. Meyer. 2012. “Measuring Community Connectedness Among Diverse Sexual Minority Populations.” *Journal of Sex Research* 49(1):36–49. doi: [10.1080/00224499.2011.565427](https://doi.org/10.1080/00224499.2011.565427).
- Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2017. “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes.”
- Ghaziani, Amin. 2014. *There Goes the Gayborhood?* Princeton University Press.
- Gieryn, Thomas F. 2000. “A Space for Place in Sociology.” *Annual Review of Sociology* 26:463–96.
- Gieseking, Jen Jack. 2020. “Mapping Lesbian and Queer Lines of Desire: Constellations of Queer Urban Space.” *Environment and Planning D: Society and Space* 0263775820926513. doi: [10.1177/0263775820926513](https://doi.org/10.1177/0263775820926513).
- Goldberg, Amir. 2011. “Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined.” *American Journal of Sociology* 116(5):1397–1436. doi: [10.1086/657976](https://doi.org/10.1086/657976).
- Goldberg, Amir, Sameer B. Srivastava, V. Govind Manian, William Monroe, and Christopher Potts. 2016. “Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness.” *American Sociological Review* 81(6):1190–1222. doi: [10.1177/0003122416671873](https://doi.org/10.1177/0003122416671873).
- Gonen, Hila, and Yoav Goldberg. 2019. “Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But Do Not Remove Them.”

- Halberstam, Jack. 2005. *In a Queer Time and Place: Transgender Bodies, Subcultural Lives*. NYU Press.
- Halperin, David M. 2012. *How To Be Gay*. Sew edition. Cambridge, Mass: Belknap Press.
- Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.”
- Hampton, Keith N., and Barry Wellman. 2018. “Lost and Saved . . . Again: The Moral Panic about the Loss of Community Takes Hold of Social Media.” *Contemporary Sociology* 47(6):643–51. doi: [10.1177/0094306118805415](https://doi.org/10.1177/0094306118805415).
- Hampton, Keith, and Barry Wellman. 2003. “Neighboring in Netville: How the Internet Supports Community and Social Capital in a Wired Suburb.” *City & Community* 2(4):277–311. doi: [10.1046/j.1535-6841.2003.00057.x](https://doi.org/10.1046/j.1535-6841.2003.00057.x).
- Hargittai, Eszter, and Aaron Shaw. 2015. “Mind the Skills Gap: The Role of Internet Know-How and Gender in Differentiated Contributions to Wikipedia.” *Information, Communication & Society* 18(4):424–42. doi: [10.1080/1369118X.2014.957711](https://doi.org/10.1080/1369118X.2014.957711).
- Held, Nina. 2017. “‘They Look at You Like an Insect That Wants to Be Squashed’: An Ethnographic Account of the Racialized Sexual Spaces of Manchester’s Gay Village.” *Sexualities* 20(5-6):535–57. doi: [10.1177/1363460716676988](https://doi.org/10.1177/1363460716676988).
- Hunzaker, M. B. Fallin, and Lauren Valentino. 2019. “Mapping Cultural Schemas: From Theory to Method.” *American Sociological Review* 0003122419875638. doi: [10.1177/0003122419875638](https://doi.org/10.1177/0003122419875638).
- Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. Random House.
- Jones, Jason, Mohammad Amin, Jessica Kim, and Steven Skiena. 2020. “Stereotypical Gender Associations in Language Have Decreased Over Time.” *Sociological Science* 7:1–35. doi: [10.15195/v7.a1](https://doi.org/10.15195/v7.a1).

- Jurgenson, Nathan. 2011. "Digital Dualism Versus Augmented Reality." Retrieved April 29, 2021 (<https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/>).
- Kelly, Brian C., Richard M. Carpiano, Adam Easterbrook, and Jeffrey T. Parsons. 2014. "Exploring the Gay Community Question: Neighborhood and Network Influences on the Experience of Community Among Urban Gay Men: Experience of Community Among Urban Gay Men." *The Sociological Quarterly* 55(1):23–48. doi: [10.1111/tsq.12041](https://doi.org/10.1111/tsq.12041).
- Kim, Jae Yeon. 2022. "How Other Minorities Gained Access: The War on Poverty and Asian American and Latino Community Organizing." *Political Research Quarterly* 75(1):89–102. doi: [10.1177/1065912920983456](https://doi.org/10.1177/1065912920983456).
- Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings." *American Sociological Review* 0003122419877135. doi: [10.1177/0003122419877135](https://doi.org/10.1177/0003122419877135).
- Kusner, Matt J., Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. 2015. "From Word Embeddings To Document Distances." 10.
- Lakoff, George. [1987] 2008. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.
- Lakoff, George, and Mark Johnson. [1980] 2008. *Metaphors We Live By*. University of Chicago Press.
- Lang, Ken. 1995. "Newsweeder: Learning to Filter Netnews." Pp. 331–39 in *Proceedings of the twelfth international conference on machine learning*.
- Lee, Monica, Amaç Herdağdelen, Minsu Park, and John Levi Martin. 2021. "Measuring Mohr Social Capital." *Poetics* 101596. doi: [10.1016/j.poetic.2021.101596](https://doi.org/10.1016/j.poetic.2021.101596).
- Levine, Jeremy R. 2017. "The Paradox of Community Power: Cultural Processes

- and Elite Authority in Participatory Governance.” *Social Forces* sf;sow098v1. doi: [10.1093/sf/sow098](https://doi.org/10.1093/sf/sow098).
- Levine, Jeremy R. 2021. *Constructing Community: Urban Governance, Development, and Inequality in Boston*. Princeton University Press.
- Lichterman, Paul. 1999. “Talking Identity in the Public Sphere: Broad Visions and Small Spaces in Sexual Identity Politics.” *Theory and Society* 28(1):101–41.
- Lizardo, Omar. 2017. “Improving Cultural Analysis: Considering Personal Culture in Its Declarative and Nondeclarative Modes.” *American Sociological Review* 82(1):88–115. doi: [10.1177/0003122416675175](https://doi.org/10.1177/0003122416675175).
- Lucy, Li, and David Bamman. 2021. “Characterizing English Variation Across Social Media Communities with BERT.” *Transactions of the Association for Computational Linguistics* 9:538–56. doi: [10.1162/tacl\\_a\\_00383](https://doi.org/10.1162/tacl_a_00383).
- Mannarini, Terri, and Angela Fedi. 2009. “Multiple Senses of Community: The Experience and Meaning of Community.” *Journal of Community Psychology* 37(2):211–27. doi: [10.1002/jcop.20289](https://doi.org/10.1002/jcop.20289).
- Martin, John Levi, and Alessandra Lembo. 2020. “On the Other Side of Values.” *American Journal of Sociology* 126(1):52–98. doi: [10.1086/709778](https://doi.org/10.1086/709778).
- Martin, John Levi, and Alessandra Lembo. 2021. “Response to Vaisey.” *Sociological Forum* n/a(n/a). doi: [10.1111/socf.12790](https://doi.org/10.1111/socf.12790).
- Marx, Karl. [1848] 1972. “The Communist Manifesto.” in *The Marx-Engels Reader*, edited by R. C. Tucker. New York, Norton.
- Mattson, Gregg. 2015a. “Bar Districts as Subcultural Amenities.” *City, Culture and Society* 6(1):1–8. doi: [10.1016/j.ccs.2015.01.001](https://doi.org/10.1016/j.ccs.2015.01.001).
- Mattson, Gregg. 2015b. “Style and the Value of Gay Nightlife: Homonormative Placemaking in San Francisco.” *Urban Studies* 52(16):3144–59. doi:

[10.1177/0042098014555630](https://doi.org/10.1177/0042098014555630).

Mattson, Gregg. 2020. “Small-City Gay Bars, Big-City Urbanism.” *City & Community* 19(1):76–97. doi: [10.1111/cico.12443](https://doi.org/10.1111/cico.12443).

McCulloch, Gretchen. 2019. *Because Internet: Understanding the New Rules of Language*. Penguin.

McMillan, David W. 1996. “Sense of Community.” *Journal of Community Psychology* 24(4):315–25. doi: [10.1002/\(SICI\)1520-6629\(199610\)24:4<315::AID-JCOP2>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1520-6629(199610)24:4<315::AID-JCOP2>3.0.CO;2-T).

McMillan, David W., and David M. Chavis. 1986. “Sense of Community: A Definition and Theory.” *Journal of Community Psychology* 14(1):6–23. doi: [10.1002/1520-6629\(198601\)14:1<6::AID-JCOP2290140103>3.0.CO;2-I](https://doi.org/10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I).

Medvedev, Alexey N., Renaud Lambiotte, and Jean-Charles Delvenne. 2018. “The Anatomy of Reddit: An Overview of Academic Research.”

Mendelsohn, Julia, Yulia Tsvetkov, and Dan Jurafsky. 2020. “A Framework for the Computational Linguistic Analysis of Dehumanization.” *Frontiers in Artificial Intelligence* 3. doi: [10.3389/frai.2020.00055](https://doi.org/10.3389/frai.2020.00055).

Meyer, Ilan H. 2020. “Generations: A Study of the Life and Health of LGB People in a Changing Society, United States, 2016-2019: Version 1.”

Meyer, Ilan H., and Soon Kyu Choi. 2020. *Differences Between LGB Democrats and Republicans in Identity and Community Connectedness*. Williams Institute, UCLA School of Law.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.”

Mohr, John W., Christopher A. Bail, Margaret Frye, Jennifer C. Lena, Omar Lizardo, Terence E. McDonnell, Ann Mische, Iddo Tavory, and Frederick F. Wherry. 2020.

- Measuring Culture*. Columbia University Press.
- Monk, Ellis P. 2022. “Inequality Without Groups: Contemporary Theories of Categories, Intersectional Typicality, and the Disaggregation of Difference.” *Sociological Theory* 07352751221076863. doi: [10.1177/07352751221076863](https://doi.org/10.1177/07352751221076863).
- Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. “Editors’ Comment: A Few Guidelines for Quantitative Submissions.” *American Sociological Review* 83(6):1281–83. doi: [10.1177/0003122418806282](https://doi.org/10.1177/0003122418806282).
- Nelson, Laura K. 2017. “Computational Grounded Theory: A Methodological Framework.” *Sociological Methods & Research* 0049124117729703. doi: [10.1177/0049124117729703](https://doi.org/10.1177/0049124117729703).
- Nelson, Laura K. 2021a. “Cycles of Conflict, a Century of Continuity: The Impact of Persistent Place-Based Political Logics on Social Movement Strategy.” *American Journal of Sociology* 127(1):1–59. doi: [10.1086/714915](https://doi.org/10.1086/714915).
- Nelson, Laura K. 2021b. “Leveraging the Alignment Between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South.” *Poetics* 101539. doi: [10.1016/j.poetic.2021.101539](https://doi.org/10.1016/j.poetic.2021.101539).
- Oldenburg, Ray. 1998. *The Great Good Place: Cafés, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. Marlowe.
- Olson, Randal S., and Zachary P. Neal. 2015. “Navigating the Massive World of Reddit: Using Backbone Networks to Map User Interests in Social Media.” *PeerJ Computer Science* 1:e4. doi: [10.7717/peerj-cs.4](https://doi.org/10.7717/peerj-cs.4).
- Orne, Jason. 2017. *Boystown: Sex and Community in Chicago*. Chicago ; London: University of Chicago Press.
- Park, Chan Young, Xinru Yan, Anjalie Field, and Yulia Tsvetkov. 2020. “Multilingual

Contextual Affective Analysis of LGBT People Portrayals in Wikipedia.”

Pechenick, Eitan Adam, Christopher M. Danforth, and Peter Sheridan Dodds. 2015.

“Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution.” *PLOS ONE* 10(10):e0137041. doi: [10.1371/journal.pone.0137041](https://doi.org/10.1371/journal.pone.0137041).

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. “[GloVe: Global Vectors for Word Representation](#).” Pp. 1532–43 in *Empirical methods in natural language processing (EMNLP)*.

Pentzold, Christian. 2011. “Imagining the Wikipedia Community: What Do Wikipedia Authors Mean When They Write about Their ‘Community’?” *New Media & Society* 13(5):704–21. doi: [10.1177/1461444810378364](https://doi.org/10.1177/1461444810378364).

Pew Research Center. 2013. “A Survey of LGBT Americans.” Retrieved April 29, 2021 (<https://www.pewresearch.org/social-trends/2013/06/13/a-survey-of-lgbt-americans/>).

Putnam, Robert D. 2007. “E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture.” *Scandinavian Political Studies* 30(2):137–74. doi: [10.1111/j.1467-9477.2007.00176.x](https://doi.org/10.1111/j.1467-9477.2007.00176.x).

Rheault, Ludovic, and Christopher Cochrane. 2020. “Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora.” *Political Analysis* 28(1):112–33. doi: [10.1017/pan.2019.26](https://doi.org/10.1017/pan.2019.26).

Rheingold, Howard. 2000. *The Virtual Community: Homesteading on the Electronic Frontier*. 2 edition. The MIT Press.

Rodriguez, Pedro L., and Arthur Spirling. 2020. “Word Embeddings: What Works, What Doesn’t, and How to Tell the Difference for Applied Research.” *Journal of Politics* 54.

- Rosch, Eleanor, and Carolyn B. Mervis. 1975. “Family Resemblances: Studies in the Internal Structure of Categories.” *Cognitive Psychology* 7(4):573–605. doi: [10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9).
- Sahlgren, Magnus. 2008. “The Distributional Hypothesis.” *Italian Journal of Linguistics* 20(1):33–53.
- Saussure, Ferdinand de. [1916] 1972. *Cours de linguistique générale*. Paris: Payot.
- Schnabel, Landon. 2018. “Sexual Orientation and Social Attitudes.” *Socius* 4:1–18. doi: [10.1177/2378023118769550](https://doi.org/10.1177/2378023118769550).
- Shalaby, Walid, and Wlodek Zadrozny. 2017. “Learning Concept Embeddings for Efficient Bag-of-Concepts Densification.” doi: [10.1007/s10115-018-1321-8](https://doi.org/10.1007/s10115-018-1321-8).
- Sia, Suzanna, Ayush Dalmia, and Sabrina J. Mielke. 2020. “Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics Too!”
- Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. 1st edition. Beijing ; Boston: O’Reilly Media.
- Simmel, Georg. 1971. *Georg Simmel on Individuality and Social Forms*. edited by D. Levine. Chicago, University of Chicago Press.
- Simmel, Georg. [1903] 1971. “The Metropolis and Mental Life.” in *Georg Simmel on Individuality and Social Forms, Heritage of sociology*, edited by D. Levine. Chicago, University of Chicago Press.
- Simmel, Georg. 2012. “The Dyad and the Triad.” in *Classical Sociological Theory*, edited by Craig J. Calhoun editor, Joseph Gerteis editor, editor James W. Moody, Steven Pfaff editor, and editor Indermohan Virk. Chichester, West Sussex ; Malden, MA: John Wiley & Sons.
- Soni, Sandeep, Lauren F. Klein, and Jacob Eisenstein. 2021. “Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers.” *Journal*

- of Cultural Analytics* 18841. doi: [10.22148/001c.18841](https://doi.org/10.22148/001c.18841).
- Spiro, Emma S., Zack W. Almquist, and Carter T. Butts. 2016. “The Persistence of Division: Geography, Institutions, and Online Friendship Ties.” *Socius: Sociological Research for a Dynamic World* 2:2378023116634340. doi: [10.1177/2378023116634340](https://doi.org/10.1177/2378023116634340).
- Stoltz, Dustin S., and Marshall A. Taylor. 2019. “Concept Mover’s Distance: Measuring Concept Engagement via Word Embeddings in Texts.” *Journal of Computational Social Science* 2(2):293–313. doi: [10.1007/s42001-019-00048-6](https://doi.org/10.1007/s42001-019-00048-6).
- Stoltz, Dustin S., and Marshall A. Taylor. 2020. *Cultural Cartography with Word Embeddings*. SocArXiv.
- Stoltz, Dustin S., and Marshall A. Taylor. 2021. “Cultural Cartography with Word Embeddings.” *Poetics* 88:101567. doi: [10.1016/j.poetic.2021.101567](https://doi.org/10.1016/j.poetic.2021.101567).
- Swidler, Ann. 1986. “Culture in Action: Symbols and Strategies.” *American Sociological Review* 51(2):273. doi: [10.2307/2095521](https://doi.org/10.2307/2095521).
- Tavory, Iddo. 2016. *Summoned: Identification and Religious Life in a Jewish Neighborhood*. University of Chicago Press.
- Taylor, Marshall A., and Dustin S. Stoltz. 2020. “Integrating Semantic Directions with Concept Mover’s Distance to Measure Binary Concept Engagement.” 18.
- Tönnies, Ferdinand. [1887] 2001. *Community and Civil Society*. edited by J. Harris. Cambridge ; New York: Cambridge University Press.
- Tufekci, Zeynep. 2014. “Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls.” *ICWSM* 14:505–14.
- Vaisey, Stephen. 2007. “Structure, Culture, and Community: The Search for Belonging in 50 Urban Communes.” *American Sociological Review* 72(6):851–73.
- van Loon, Austin, Salvatore Giorgi, Robb Willer, and Johannes Eichstaedt. 2022. “Re-

gional Negative Bias in Word Embeddings Predicts Racial Animus—but Only via Name Frequency.”

Vrana, Adele Godoy, Anasuya Sengupta, and Siko Bouterse. 2020. “16 Toward a Wikipedia For and From Us All.” in *::wikipedia @ 20*.

Wallach, Hanna M. 2006. “Topic Modeling: Beyond Bag-of-Words.” Pp. 977–84 in *Proceedings of the 23rd international conference on Machine learning - ICML '06*. Pittsburgh, Pennsylvania: ACM Press. doi: [10.1145/1143844.1143967](https://doi.org/10.1145/1143844.1143967).

Waller, Isaac, and Ashton Anderson. 2019. “Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms.” Pp. 1954–64 in *The World Wide Web Conference on - WWW '19*. San Francisco, CA, USA: ACM Press. doi: [10.1145/3308558.3313729](https://doi.org/10.1145/3308558.3313729).

Waller, Isaac, and Ashton Anderson. 2021. “Quantifying Social Organization and Political Polarization in Online Platforms.” *Nature* 1–5. doi: [10.1038/s41586-021-04167-x](https://doi.org/10.1038/s41586-021-04167-x).

Wellman, Barry. 1979. “The Community Question: The Intimate Networks of East Yorkers.” *American Journal of Sociology* 84(5):1201–31.

Wellman, Barry, and Barry Leighton. 1979. “Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question.” *Urban Affairs Quarterly* 14(3):363–90. doi: [10.1177/107808747901400305](https://doi.org/10.1177/107808747901400305).

Weston, Kath. 1995. “Sexual Imaginary and the Great Gay Migration.” *GLQ: A Journal of Lesbian and Gay Studies* 2(3):253–77.

White, Harrison C. [1965] 2008. “Notes on the Constituents of Social Structure. Soc. Rel. 10 - Spring '65.” *Sociologica* (1):0–0. doi: [10.2383/26576](https://doi.org/10.2383/26576).

Winer, Canton. 2020. “Solidarity, Disdain, and the Imagined Center of the Gay Imagined Community.” *Sociological Inquiry* n/a(n/a). doi: [10.1111/soin.12403](https://doi.org/10.1111/soin.12403).

- Zerubavel, Eviatar. 2002. "The Fine Line: Making Distinctions in Everyday Life." Pp. 223–32 in *Cultural Sociology*, edited by L. Spillman. Malden, MA: Wiley-Blackwell.
- Zerubavel, Eviatar. 2018. *Taken for Granted: The Remarkable Power of the Unremarkable*. Princeton ; Oxford: Princeton University Press.
- Zhang, Justine, William L. Hamilton, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. "Community Identity and User Engagement in a Multi-Community Landscape." 10.