COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

On the Wall and Online:

Graffiti Proximity Moderation on Political Amplification in London's COVID-Era Twitter Discourse

Master's Thesis

MSc in Business Administration and Data Science

Author: Caoimhe Gallahue Advisor: Rob Gleasure

May 15, 2026

University:	Copenhagen Business School
Degree:	MSc in Business Administration and Data Science
Department	Department of Digitalization and Innovation
Author:	Caoimhe Gallahue (175890)
Advisor:	Rob Gleasure
Title and subtitle:	On the Wall and Online: Graffiti Proximity Moderation on Political Amplification in London's COVID-Era Twitter Discourse
Description:	This thesis examines whether proximity to street art predicts the amplification of politicized content on Twitter, using geocoded COVID-era tweets from London, Belgium, and Denmark. Applying inoculation theory to spatial political environments, it tests whether blame-attributing political discourse amplifies more widely than non-politicized content and whether graffiti proximity moderates that gap. Mixed-effects modeling reveals that individual user characteristics (ICC = 0.629) account for the majority of amplification variance, reframing the spatial hypothesis as a user-composition effect.
Keywords:	Twitter; graffiti; street art; inoculation theory; political amplification; COVID-19; geocoded data; mixed effects; computational social science
Date:	May 15, 2026
Number of pages:	77
Number of characters:	177,414

Table of content

1 Introduction

Graffiti is simultaneously a loud and quiet art form. Whether you are from a large city inundated with graffiti messaging, or have seen scribbles on an electricity box in suburbia, we have certainly all seen and interacted with graffiti and street art in the wild. The sometimes jarring or obscene style of graffiti tagging carries a cultural subconscious bias of being lawless and rebellious. Certainly, graffiti has a long history of being the expression of unsanctioned voices in cities. Long before the internet made mass communication frictionless, political dissent found its medium in painted walls - going all the way back to cave paintings and ancient Roman graffiti, to more modern examples like revolutionary slogans in Paris in 1968, anti-apartheid murals in Soweto, the IRA and Loyalist street art of Belfast, and the now-globally legible visual language of graffiti that emerged from New York's subway system in the 1970s (Ross & Ferrell, 2019). What distinguishes graffiti from other forms of political communication is not only its visual character but its constitutive relationship with transgression. To leave a mark where one is not permitted to do so is, at minimum, to refuse the spatial ordering of public life, asserting that the person without a billboard budget, without editorial access, without institutional standing, also has something to say and a space to say it (Ferrell & Stewart-Huidobro, 2021; Mcauliffe, 2012).

That transgressive quality of graffiti has never been purely aesthetic. Graffiti and street art encode a particular theory of legitimacy. The political implications of this are not incidental. Studies of urban political geography have consistently documented that dense concentrations of graffiti and street art co-occur with spaces of political activation, labour organization, and civic resistance (Dovey et al., 2012; Young, 2013). These are not coincidental co-presences. They reflect a shared character of public space, a common understanding that the city is a medium of conversation, somewhere discussions are being had, rather than merely a backdrop for everyday life. Graffiti tags and murals are not purely decorative choices but claims to who the neighborhood belongs to and who has the right to speak within it.

Since graffiti is so present in major cities, many residents living in neighborhoods with high graffiti density may not even notice the persistent writing on the wall. So while often loud and in-your-face stylistically, given the density it occupies in many cities, graffiti is a silent pervasive noise. There is also a paradox of graffiti while being pervasive and ever present in many cities, it is an ephemeral art form. In the majority of cities, graffiti is illegal and many governments take frequent action to clean up and minimize graffiti pieces. These properties make graffiti and street art situated in an interesting space prime for political activation and messaging.

This thesis begins with that observation and asks whether it might extend further than the walls themselves: whether living in proximity to spaces of political-aesthetic expression might shape how people relate to political discourse when it migrates online. It is an unusual question that requires explanation. The connection between a spray-painted wall and a tweet is not self-evident. But the question becomes more tractable, and perhaps more urgent, when it is placed alongside a set of observations about the pathologies of online information environments and the growing recognition that the problem of harmful content is not entirely separable from the problem of informational passivity.

1.1 Amplification and the Emotional Logic of Political Discourse

Online political communication does not spread uniformly. Research on social media sharing behavior consistently shows that content generating moral outrage and indignation circulates more widely than neutral informational content (Brady et al., 2017). The relevant distinction is made between emotional orientations that are externalizing and those that are internalizing. Content that names a responsible party and assigns blame activates outrage, contempt, and indignation. Conversely, content that describes systemic risk or shared uncertainty activates anxiety and grief. These two modes of negative affect carry markedly different sharing signatures, with externalizing responses mobilizing, and internalizing responses withdrawing users and prompting little amplification. As the outrage/anxiety distinction established in Chapter 2 predicts, the same valence label (negative) covers two structurally different emotional registers with opposite sharing dynamics.

This thesis examines that distinction in a corpus of geocoded COVID-era tweets from London, Belgium, and Denmark. The primary amplification measure is the quote tweet: a rebroadcasting of content with implicit engagement, endorsement, or additional commentary. In London, tweets referencing named political figures in the context of pandemic blame attribution show a quote rate 7.4 percentage points higher than non-politicized COVID content. The analysis asks what predicts this gap and whether the spatial character of where a tweet was sent in proximity to graffiti is among the predictors.

1.2 Graffiti, Inoculation, and the Urban Political Environment

The second element of this thesis concerns where these tweets were sent. If amplification behavior varies with the geographic context of the tweeter, specifically, with whether graffiti proximity influences that context. The theoretical framework connecting physical political environments to online amplification behavior is inoculation theory, developed fully in Chapter 3. What follows is a brief account of the mechanism as it applies here.

Inoculation theory proposes that prior exposure to contested persuasive claims builds cognitive resistance to those claims when subsequently encountered at full force (McGuire, 1964; Van Der Linden et al., 2017). Experimental work on inoculation-based ”prebunking” has confirmed this effect across a range of contexts, including COVID-19 and information sharing (Cook et al., 2017; Roozenbeek & Van Der Linden, 2019). That literature has operated almost exclusively through deliberate, administered interventions in a controlled setting. However, less examined is whether ambient or environmental exposure to political messaging, as a feature of everyday urban life rather than an experimental condition, might generate an analogous epistemic orientation.

Graffiti is a plausible candidate for this function. Its transgressive character means graffiti messaging carries enough importance to make a risk to share on surfaces controlled by others. A resident moving daily through walls of political attribution and blame cannot mistake these messages for institutional communication. They are flagged as contested claims by the very circumstances of their controversial production. And it is precisely contested claims, encountered repeatedly, that inoculation theory predicts

(a) ”Sammen mod magt-eliten”, ”Together against the power elite” in København S

(b) ”Free Palestine” in København N

(c) ”Every thing will be ok” in Christiana

(d) ”We’re all just visitors” in København N

Figure 1: Graffiti images collected by the author in and around Copenhagen Metropolitan Area

will cultivate critical epistemic orientation. Chapter 3 develops why the transgressive character of graffiti is not incidental to this function but constitutive of it. Chapter 2 situates graffiti within the urban political geography literature as a spatial practice of claim-making. The empirical question this thesis addresses is whether any of this leaves a detectable trace in online amplification behavior.

The results are, as Chapter 5 reports, largely null at the level of spatial moderation. Proximity to street art does not significantly predict tweet amplification patterns once user-level heterogeneity is accounted for. The intraclass correlation in the mixed-effects models (ICC = 0.629) indicates that approximately 63% of the variance in whether a tweet gets quoted is attributable to stable individual characteristics rather than content type or location. What the analysis reveals is that high-amplification users cluster geographically in politically expressive urban areas, a finding that re-frames the question from *does graffiti exposure change amplification behavior?* to *why do politically primed users concentrate in politically expressive urban spaces?* That re-framing is the substantive theoretical contribution of the thesis.

1.3 Research Question and Scope

The central question driving this thesis is:
To what extent does proximity to street art predict the nature and amplification of politicized content on Twitter?

This question is operationalized through two related hypotheses. The first *H1* concerns the relationship between content type and amplification across national contexts. Specifically, whether politicized content, tweets attributing pandemic outcomes to named political figures, exhibits higher amplification rates than non-politicized content. The second *H2* concerns the spatial moderation of this relationship: whether any content-type amplification gap is attenuated (or amplified) by proximity to street art, with London as the primary site of analysis.

The emphasis on proximity rather than viewership is a methodological constraint that also has substantive implications. This thesis does not measure whether individuals have seen specific pieces of graffiti. It measures the distance between the geographic coordinates of a tweet and the nearest cataloged piece of graffiti in that city. This is a coarse operationalization, and its limitations are acknowledged throughout. Approximately 80% of the London tweets used here are geocoded only to the level of the bounding box of the named place from which they were sent, meaning that ”proximity” in this context is more accurately described as neighborhood-level or area-level proximity. The proximity measure is a spatial covariate, not a measure of individual exposure.

The thesis does not claim, and cannot claim, that individuals near graffiti have been inoculated by it. What it examines is whether being physically located in a politically expressive urban space is associated with distinctive amplification patterns in online political discourse. The theoretical interest lies in the association itself and in what the pattern of results says about the relationship between urban political environments and online behavior.

The geographic scope is London as the primary site, with Belgium and Denmark providing cross-

national comparison. London is the primary site for three reasons: it has the largest TBCOV-derived Twitter dataset with geocoding (116,391 tweets), it has the most detailed graffiti cataloged (835 pieces concentrated in politically distinctive urban areas such as Shoreditch, Brixton, and Camden), and it is the national context in which the amplification gap for politicized content is positive and statistically detectable at the descriptive level.

1.4 Relevance

The question driving this thesis sits in a gap between three literatures that have each approached part of the problem without addressing the whole of it.

Urban political geography has mapped the production of politically expressive space with care. Subculture has shaped who makes graffiti, what it claims, which communities it belongs to. What the tradition has not asked is what that environment does to inhabitants' behavior in other domains: whether living in proximity to concentrated political expression in the built environment shapes how people engage with political discourse online. That question lies outside the literature's primary concern, which is the politics of space rather than the spatiality of political behavior.

The political amplification literature has developed sophisticated accounts of what drives political content to spread. But the structural question of what kinds of communities, concentrated in what kinds of urban environments, produce amplification at scale has not been a central concern. User heterogeneity, when modeled at all, tends to appear as a robustness check rather than a first-order explanatory variable. The result is a body of work that attributes amplification to message properties while leaving the dominant source of variance — who the sender is — largely unmodeled.

The third literature, Inoculation theory, offers a mechanism that could, potentially connect the first two. If repeated exposure to political counter-claims produces durable epistemic engagement, and if graffiti-saturated urban environments provide such exposure, urban political geography and online amplification behavior should be linked through a specific psychological pathway. But the experimental inoculation literature has operated almost exclusively with administered interventions in controlled settings. Ambient inoculation, the kind produced by years of living among politically expressive walls, has not been empirically tested.

The contribution of this thesis is to occupy the intersection these three literatures share but have not jointly addressed. The results are, as Chapter 5 documents, null at the level of individual spatial moderation. But the null result is the contribution. The intraclass correlation of 0.629 — nearly two-thirds of the variance in whether a tweet is quoted is explained by stable user-level characteristics — demonstrates that content and spatial effects are secondary to who the sender is. Research that attributes amplification differentials to message framing or spatial proximity, the approach each of these literatures would naturally adopt, misidentifies the primary driver when user identity is left unmodeled. The finding shifts the question from *does graffiti change amplification behaviour?* to *why do high-amplification users cluster in politically expressive urban spaces?*. A question that is sociologically richer and that requires

exactly the convergence of urban geography, political communication, and behavioral research that no previous literature has attempted.

2 Background Literature

This section develops the conceptual argument and outlines the related work and background literature that motivates the study's theoretical framework. The literature presented, creates the story that political discourse on social media exhibits the structural properties of a virus. Understanding how it spreads, and who is prepared to engage with it requires treating the digital and the physical expressions as mirrors of each other rather than separate domains.

2.1 Social Media as A Reflection of Physical Space

The idea of social media as a self-contained information environment has been challenged by research demonstrating that online political behavior is deeply shaped by the physical spaces in which it occurs. Twitter for example is not an isolated platform, it encourages engagement and interactions with community as well as sharing of data like location. It is used by embodied, spatially located, individuals whose political opinions and social networks are constituted in part by the physical environments they inhabit.

Neighborhood effects tradition in political sociology is a well researched idea that the political character of one's residential environment shapes political attitudes and behavior independently of individual-level characteristics. Lazarsfeld et al. (1948) identified the political composition of one's immediate social context as a determinant of voting behavior. Contextual effects research confirmed that living in a predominantly Labour or Conservative ward in the United Kingdom shapes political participation and political knowledge over individual socioeconomic position (Cutts et al., 2007). Huckfeldt et al. (2002) showed that political information flows through social networks embedded in local space, meaning who you talk to about politics is determined by who you live near. Enos (2016) demonstrated that the racial composition of one's neighborhood shapes implicit racial attitudes through repeated exposure rather than deliberate socialization. In each case, the political character of space operates through spatial proximity rather than conscious engagement.

The neighborhood effects interpretation faces a persistent methodological challenge: the relationship between residential environment and political behavior may be produced by self-selection rather than contextual influence. Meaning politically engaged individuals may cluster in politically expressive neighborhoods because those environments match their pre-existing political identities, not because living there shapes their political behavior. Bishop (2008) documented this dynamic at scale, showing that partisan geographic sorting in the United States has intensified across decades as like-minded citizens increasingly cluster in residential communities that reflect and reinforce their prior beliefs. The correlation between neighborhood political character and individual political behavior is an artifact of sorting rather than a contextual effect. The identification challenge this creates, distinguishing exposure from

selection, is not unique to this study. It is the central problem in neighborhood effects research, and one that fixed-effects estimation and matched observational designs are specifically constructed to address (Enos, 2016). The present study's within-user estimation approach and coarsened exact matching are designed with this challenge explicitly in view, and the findings are interpreted with the observational equivalence of exposure and selection accounts in mind.

The extension of this theory to social media is the natural next step as we increasingly gather in online spaces. Similarly the neighborhood effects are empirically supported in social media spaces. Research on geolocated Twitter data has consistently found that online political discourse is not uniformly distributed across urban space but clusters in ways that reflect the political character of local areas (Barberá & Rivero, 2015). Politically expressive neighborhoods generate online political activity at higher rates (Jungherr, 2016). Not necessarily because politically engaged individuals use Twitter more, but instead because the political character of a space constitutes part of the information environment that residents carry with them, including into their online activity.

Within the neighborhood effects literature, the visual character of political space has received less attention than demographic composition or social network structure. However the available evidence suggests it operates through the same logic of priming. Research on political yard signs, campaign banners, and visual political displays indicated that the visible presence of political expression in local environments signals the political norms of that community and shapes residents' willingness to express their own political views (Makse & Sokhey, 2014; Noelle-Neumann, 1974). The visibility of political expression functions as a social norm cue. Areas where political expression is visually abundant facilitates further expression and likeliness of participation.

Street art and graffiti are the most persistent and visually saturated form of political expression in the urban built environment, and their relationship to online political behavior is the specific instantiation of the physical-digital nexus this study examines. Concentrations of politically engaged street art mark areas where political expression is normalized and where residents are continuously exposed to political counter-narratives. These areas are also areas with documented histories of political engagement, anti-institutional activism, and organized resistance to urban governance. The visual political culture of these spaces is not incidental, political history has shaped the dispositions of people who inhabit and move through these spaces. Politically engaging in physical and digital spheres is continuous with the political culture of the physical space they occupy.

2.2 Biological and Ecological Metaphors in Urban and Media Studies

The application of biological and ecological metaphors to social and information systems has a long intellectual history. The Chicago School sociologists of the 1920s adapted Darwinian competition and succession models to explain how urban populations develop across space, treating the city as an ecosystem in which social groups occupy different niches and compete for resources (Park et al., 1925). Dawkins (1976) extended the metaphor to cultural transmission, proposing that memes or ideas and practices

spread through populations according to the same logic of differential replication that governs biological evolution. Forms that are better suited to the cognitive and social environment of human minds and networks, spread more widely and persist longer. This is the theoretical ancestor of contemporary research on information virality.

Virality literature has extended and operationalized this biological metaphor with increasing precision. Information spreads through social networks in ways that mirror biological contagion. Growth is exponential in the early stages, cascade structures are shaped by network topology, and content reaches further when it generates emotional arousal. Zhao et al. (2012) applied a SIHR (Susceptible, Infected, Hibernator, Recovered) epidemic model, which treats users as moving through states of awareness and engagement, much like a population exposed to a pathogen, to rumor spread on Twitter. Vosoughi et al. (2018) showed that false news travels faster and further than true news, driven by novelty and emotional arousal rather than bot activity. Bakshy et al. (2012) found that network structure determines which information reaches which users. The biological metaphor in this context describes structural features of how information actually moves.

The biological metaphor is even embedded in the language of contemporary public health communication about COVID-19 information. The WHO's characterization of the COVID-19 information environment as an *infodemic*, or information epidemic, that overwhelmed public ability to sort through reliable guidance explicitly adopts the epidemiological language framing (Cinelli et al., 2020). An infodemic has its own transmission dynamics, its own superspreaders, and its own interventions. The WHO and subsequent researchers have applied epidemic modeling frameworks to the spread of COVID-19 misinformation, finding that false information spreads faster through social networks than the pathogen itself. Dissemination of information follows similar exponential growth curves at the onset of outbreak events, and responds to interventions, prebunking, platform labeling, network disruption, structurally analogous to quarantine and vaccination (Roozenbeek & Van Der Linden, 2019).

The viral metaphor has attracted empirical qualification, however. Goel et al. (2016), found that most content conventionally described as viral does not spread through true contagion chains. The majority of wide-reaching content instead spreads through broadcast dynamics: a single high-reach account exposes a large audience in one transmission event, without the branching cascade structure that biological contagion produces. True structural virality is empirically rare. This finding shifts the relevant question. If diffusion is predominantly broadcast-driven, what matters is not simply the transmission properties of content but the amplification decisions of high-reach accounts. The present study's focus on quote behavior as an amplification signal reflects this. A small number of highly active users account for a disproportionate share of political content diffusion — a pattern consistent with broadcast dynamics, and one that motivates the within-user analytical approach. The biological metaphor retains its utility not as a precise claim about diffusion mechanics but as a framework for the content properties that determine which messages high-reach accounts choose to amplify.

The ecological and biological metaphors are consequential for the present study. They make structural claims about what political discourse on social media is. If discourse is viral, then the relevant questions

are those that epidemiology asks: What are the properties of the pathogen that make it more or less contagious? What are the properties of the host that make them more or less susceptible? What are the features of network structure that determine how widely and quickly infection spreads? And, critically, what forms of prior exposure build resistance to or priming for subsequent infection? These are the questions that Chapter 3's theoretical framework is designed to answer.

2.3 Political Blame as the Engine of Amplification

Twitter's asymmetric follower-following structure, its public default, and its native amplification mechanisms (quote, retweet, like) constitute a network designed for the rapid, transmission of content across network boundaries (Jungherr, 2016). Political content generated by a small proportion of highly active users is systematically over-represented in the information streams of the broader population (Pew Research Center, 2019) found 6% of adults produce 73% of all tweets from US adults. This vocal minority means that political discourse on Twitter is not a representative random sample of public opinion, instead it is a sample produced by, and circulated among, a politically primed population whose amplification behavior is consistent across contexts.

Within the broader category of political discourse, blame attribution toward named individuals occupies a structurally distinctive utility. It is the form of political content that is most precisely engineered for rapid diffusion through social networks, whether deliberately or emergent. Understanding why requires an account of the emotional architecture that drives sharing behavior.

Brady et al. (2017), in an analysis of political tweets, demonstrated that each additional moral-emotional word in a tweet increased its diffusion by approximately 20%. The combination of *moral evaluation* and *emotional activation* — judgments of right and wrong paired with emotional arousal — generated diffusion rates that substantially exceeded either component alone. Blame-attributing content directed at named individuals is precisely this combination: it bundles moral condemnation and outrage into a person-targeted form whose shareable properties are developed theoretically in Chapter 3.

There is a distinction between externalizing and internalizing negative affect. Research on emotion and information sharing has consistently differentiated outrage, contempt, and indignation — emotions directed at an external agent held responsible for harm — from anxiety, grief, and fear — emotions directed at one's own vulnerability (Berger & Milkman, 2012; Brady et al., 2017). Notably, externalizing emotions motivate approach behavior. They direct attention toward the responsible agent and activate the impulse to warn, condemn, and mobilize. While internalizing emotions motivate withdrawal, directing attention inward and suppressing the impulse to share. Blame attribution content is by definition externalizing the negative affect. It explicitly names the agent, assigns the responsibility, and activates collective denunciation.

UK COVID Twitter discourse is structured asymmetrically around blame. Research on UK pandemic political communication has autopsied the rapid politicization of pandemic-related content from the first weeks of lockdown. Early debates about PPE procurement, testing capacity, and care home deaths gen-

erated waves of blame-attributing discourse whose targets were overwhelmingly politicians (Thelwall & Thelwall, 2020). Fetzer et al. (2021) showed that perceptions of government accountability in pandemic management were strong predictors of political anxiety and blame attribution in the UK context.

2.4 Information Sharing Forced as Politicization

A distinctive feature of the COVID-19 information environment was the systematic collapse of the boundary between informational and political content. This collapse was not simply a matter of political actors exploiting a health crisis; it was produced by structural features of the pandemic's relationship to state governance, public health infrastructure, and social media's design as an amplification system.

In ordinary circumstances, information about a respiratory illness would flow primarily through public health channels without acquiring the emotional architecture of blame-attributing political discourse. COVID-19 disrupted this routing. The pandemic's management was inseparable from specific decisions by specific political actors. These were concrete decisions whose consequences were directly experienced by the public, who created specific named targets for blame attribution in a way that other national events like financial regulation failures or environmental policy shortcomings typically do not.

The structural consequence was that health-informational content and political-blame content became entangled in ways that made disentangling them difficult for both platforms and users. A tweet about NHS capacity became a claim about government performance. The informational content of the pandemic was politically charged not because users chose to politicize it but because the pandemic's management had made the political unavoidable. (Imran et al., 2022) captured this in the TBCOV dataset. Tweet volume was closely coupled to political events rather than to epidemiological events like case count peaks. Twitter users were processing the pandemic primarily as a political phenomenon, not tracking the disease curve but the accountability curve.

This forced politicization had a specific consequence for how content diffused. As (Bail et al., 2018) showed, exposure to politically charged content on social media can increase polarization even among users who were not seeking political engagement. On Twitter, this dynamic was intensified during COVID by the concentration of political blame content in the most highly quoted material. The tweets that spread widest were not those providing the most accurate health information but those providing the most emotionally activating political attribution. Information sharing, structurally, was not a neutral act during the pandemic but instead channeled by the platform's amplification logic into the circulation of politicized content.

This dynamic is directly operationalized in the present study's contrast between politicized content (blame attribution toward named political figures) and non-politicized COVID content (discussion of the disease, its effects, and public health responses without political attribution). The predicted consequence of forced politicization is that politically framed content circulates more widely than neutral discourse. The infodemic is a crisis of the selective virality of politically structured information within an amplification system that rewards exactly those content properties.

2.5 Dissent: From Anonymous Physical Space to the Identified Online

For most of the twentieth century, political expression through graffiti was characterized by three properties that distinguished it from all other forms of political communication: it was anonymous, it was spatially bounded, and its illegality was part of its meaning. A tag on a wall did not carry its author's name, it was visible only to those who physically passed through that space, and its presence on a surface not owned by the author constituted an act of illegal appropriation that was itself a political statement (Ferrell & Stewart-Huidobro, 2021; Mubi Brighenti, 2010).

The anonymity of graffiti was of course a protective measure. Writers in New York's subway graffiti scene in the 1970s and 1980s understood the logic explicitly, a tag name pseudonym allowed participation in a visible countercultural practice without fear of the legal and social consequences of identification (Ross & Ferrell, 2019). The same logic governed the stencil traditions that European political street artists developed in the 1980s and 1990s. Graffiti artist's anonymity across a career of politically provocative work is a continued graffiti tradition. Protection that anonymous inscription affords to those whose speech, if attributed, would attract censure or prosecution.

The spatial boundedness of graffiti is equally consequential. A graffiti piece only is visible to those who walk past that wall/structure. It cannot be algorithmically amplified or virally distributed, its audience is determined by physical geography and the pedestrian flows of the city. This boundedness was, historically, both a limitation and a form of protection. By design, the community that encountered and responded to a piece was local, and the encounter was embedded in the physical context. Research on graffiti's function as counter-public information (Fraser, 1990) has emphasized this spatial grounding. Graffiti creates alternative public spheres that are genuinely local, serving communities whose perspectives are underrepresented in mainstream media, in the physical spaces where those communities live.

During the COVID-19 pandemic, this counter-public function operated in conditions that made it particularly salient. As Ryan (2021) documented, pandemic-related street art proliferated across London from the first weeks of lockdown, addressing the same actors and themes that appeared in the TBCOV Twitter data. Street art was, during lockdown, one of the few forms of political communication that could reach people confined to their local neighborhoods, functioning as genuinely local counter-information at a moment when the usual channels of political expression (rallies, canvassing, public meetings) were unavailable. The Graffiti Database (n.d.) dataset used in the present study captures 835 pieces across London concentrated in Shoreditch, Brixton, Camden, and Hackney Wick areas with the longest histories of politically engaged street art practice, and areas whose politically expressive character predates the pandemic but whose visual counter-commentary was actively generated during the study period.

Against this tradition, online political expression on Twitter represents a structural transformation rather than a simple migration. The dissent is somewhat the same, political blame, countercultural challenge to institutional authority, the assertion of community, but the conditions of its expression are radically different. Twitter is more visibly documented; every political tweet is attributed to a specific account, however pseudonymous users try. It is also networked, content does not stay in a place but travels

across network connections according to the amplification decisions of users who may be anywhere. And it is surveilled, the metadata of every tweet — timestamp, location, engagement, network connections — is available to platform operators, state actors, and researchers that can be traced to an individual.

The migration of dissent from anonymous physical space to the identified online has consequences that are directly relevant to the present study's theoretical framework. The threshold for political expression has risen, even as the reach has expanded.

This transformation is consequential for the ambient inoculation argument developed in Chapter 3. The graffiti tradition of anonymous, spatially bounded political expression constituted a form of community political practice that was low-cost, low-risk, and pervasive in its environmental presence. It was a mode of political engagement that did not require identification and did not expose participants to the costs of attributed political speech. The areas where this tradition was densest, where political expression saturated the visual environment, were areas where political discourse was continuously modeled, normalized, and practiced in the daily visual experience of residents. The question the present study asks is whether the ambient political priming generated by this physical tradition of dissent continues to shape political engagement behavior in the domain where dissent has migrated.

2.5.1 Street Art and the Production of Politically Expressive Urban Space

The graffiti and street art tradition that produced the physical environments this study examines has deep countercultural roots that give politically expressive urban space its particular character. In the United States, the emergence of subway graffiti in New York City in the early 1970s was inseparable from the conditions of post-industrial urban crisis: fiscal collapse, white flight, the abandonment of public infrastructure, and the structural exclusion of Black and Latino youth from mainstream cultural and economic life (Castleman, 1999; Ross & Ferrell, 2019). Writing on a subway car that traversed the entire city was a claim to presence in a city that had effectively denied it — a refusal of spatial and social invisibility (Ross & Ferrell, 2019).

London's street art scene is the heir to these traditions. Concentrated in East London — Shoreditch, Brick Lane, Hackney Wick — it has developed through interaction between subcultural graffiti practice, gallery-adjacent mural commissioning, and sustained political engagement ("London calling", 2016; Riggle, 2010). The density of politically engaged work in Shoreditch reflects the area's history as a site of anti-gentrification activism, immigration politics, and anti-austerity expression and these concerns are consistently visible in the street art that characterizes the neighborhood. The 835 pieces captured in the Graffiti Database (n.d.), are the contemporary reincarnation of a tradition whose political character is constitutive of urban space.

A dimension of this account is the transformation of London's street art areas through gentrification. The neighborhoods with the highest concentrations of politically engaged street art have undergone substantial residential displacement over the same period that produced their visual political culture. The working-class and artist communities whose political commitments generated the anti-institutional, anti-

austerity, and anti-gentrification content visible in these areas have been progressively displaced by the very processes that street art in those areas critiques ("London calling", 2016). Mcauliffe (2012) has documented this tension in the context of the creative city: graffiti produced as unauthorized counter-expression is progressively aestheticized, incorporated into tourism infrastructure, and deployed as a marker of neighborhood distinctiveness in ways that sever its connection to the political communities that originally produced it. The process of political neutralization is not confined to market-driven displacement. Lerner (2019), examining post-Soviet graffiti in Moscow, documents how hybrid states actively co-opt rather than suppress dissident inscription, incorporating oppositional street art into official cultural narratives in ways that neutralize its political content without removing its visual presence. The mechanisms differ, with displacement by capital in the London case versus absorption by the state in hybrid authoritarian contexts, but both demonstrate that graffiti's political charge is institutionally contingent.

The present study does not require residents of graffiti-dense areas to be the original producing community, or to be consciously aware of the political traditions. The ambient inoculation mechanism operates through environmental exposure to visible political counter-expression. What matters is that graffiti-dense areas are spaces in which political dissent is continuously visible, normalized by its pervasiveness, and available as context for whoever currently inhabits them. Whether that ambient exposure produces the predicted amplification effects is ultimately an empirical question, and one the study's design is positioned to address.

2.6 Synthesis

The themes reviewed in this chapter converge on a unified account of why the physical and digital dimensions of political engagement are continuous rather than separate, and why the viral characterization of political discourse motivates a specific theoretical framework for understanding that continuity.

Social media reproduces and extends the political character of physical space. The biological and ecological metaphors long applied to urban life and information systems establish the structural logic of diffusion that virality research has operationalized empirically. Blame-attributing political content is the most contagious form of political discourse, combining the moral-emotional architecture that drives sharing with the outrage-externalization that motivates amplification over internalizing anxiety. These properties make political discourse on social media structurally viral, not only as metaphor but as an accurate description of its diffusion dynamics. The COVID-19 pandemic accelerated this virality by collapsing the boundary between informational health content and political blame content, channeling information sharing into the amplification of politicized attribution. And the sites of political dissent have migrated from the anonymous, spatially bounded, low-cost inscription of graffiti to the identified, networked, high-exposure act of the political tweet, raising the cost of political expression.

Together, these themes generate a precise question. When political discourse is viral, the relevant axes of variation are: *what makes particular content more or less contagious*, and *what makes particular individuals or populations more or less prepared to engage with it*. The present study focuses on the

second axis: whether the ambient political environment of physically expressive urban space constitutes a form of prior exposure that primes political engagement online.

This is a question the existing literature does not address. Research on neighborhood effects has established that physical political context shapes political attitudes and behavior but has not examined the online dimension of this relationship. Research on Twitter political amplification has documented the structural dominance of highly engaged minorities but has not connected this to the physical environments they inhabit. Research on graffiti has established its function as counter-public communication but has not examined whether the communities formed around political street art exhibit distinctive online political behavior.

The theoretical framework that motivates the exposure interpretation, and that Chapter 3 develops in full, is inoculation theory. The background literature builds toward this connection: if political discourse is viral, then inoculation is the appropriate theoretical account of how prior exposure to political counter-narratives builds the attitudinal readiness that shapes amplification sharing. Chapter 3 develops its theoretical framing.

3 Theory

This chapter develops the theoretical argument from which the study's two hypotheses are derived. The central question: whether proximity to street art predicts the nature and amplification of politicized content on Twitter, is approached through the lens of inoculation theory. Inoculation theory sits in the psychological account of how prior exposure to political counter-narratives builds resistance to persuasion and primes subsequent political engagement. This chapter proposes inoculation theory as the primary theoretical candidate for explaining the relationship between physical political space and online political behavior, and derives two spatial predictions from it that the empirical analysis is designed to test.

A note on what the theoretical argument can and cannot establish is important to establish. The inoculation framework generates clear spatial predictions but the cross-sectional observational design places limits on how strongly those predictions can be tested. These constraints, developed fully in 3.4, shape the study's analytical strategy: the discriminating test is not whether a spatial correlation exists, but whether the amplification pattern survives within-user estimation that holds stable individual political engagement constant. The chapter argues for inoculation as the theoretically motivated prediction while 3.4 specifies what the data can and cannot establish in its support.

3.1 Inoculation Theory

3.1.1 Core Ideas and History of the Theory

Mcguires 1961 Inoculation theory poses information inoculation as a metaphor (McGuire & Papageorgis, 1961). In the same way medically, we can be injected with a weaker version of a virus and then inocu-

lated against disease, Mcguire argues that exposure to weakened persuasions can then lead individuals to develop resistance against stronger, future persuasive attacks. The receiver in that way becomes immune to attacking messaging that attempts to change their attitudes or beliefs.

Conventional inoculation messages have two main structural features: a forewarning and preemptive refutation (Banas, 2020). The forewarning warns the individual that their beliefs are vulnerable and may be challenged. The statement is typically at the start of the message and includes wording: "You have the right position but others will not agree". The preemptive refutation then introduces and refutes the potential counter-arguments before the persuasive message is encountered (Banas, 2020). Two or three counter-arguments may be introduced and refuted in the inoculation treatment.

The most important part of inoculation messaging is a perceived threat. In inoculation theory, the threat is the recognition that a position is vulnerable (Banas, 2020; McGuire & Papageorgis, 1961). Threat drives resistance. When a forewarning is explicitly used, it creates threat. Conversely when a forewarning is not used, two-sided messaging creates threat. The forewarning and threat are two different concepts in the theory, forewarning being a message feature, and threat being the message effect. Threat is required in inoculation messaging, whereas the forewarning and preemptive refutation are optional.

The inoculation metaphor captures an important aspect of the dynamics of political persuasion. Resistance is not a matter of holding correct information, but of having been prepared to encounter and process challenges to one's beliefs. An individual who has been repeatedly exposed to political counter-narratives even in weakened or inoculated doses is better equipped to engage critically with new political messages than one encountering them for the first time.

3.1.2 Similar Applications

The idea of inoculation has been used to contextualize the spread of misinformation. The context of the theory has been applied to areas like politics, health, and commerce.

Pfau et al. (1992) naturally extended Mcguire's inoculation metaphor into health strategy with an inoculation-informed antismoking campaign. Their study found inoculation promoted resistance to smoking initiation particularly in young adolescents with low self esteem. With attitudinal effects lasting for up to 20 months after inoculation pretreatment (Pfau & Bockern, 1994; Pfau et al., 1992). Building on these results Godbold and Pfau (2000) found that inoculation similarly promoted resistance to alcohol influences in adolescents and Parker et al. (2012) found inoculation conferred resistance to challenges to college students' condom use attitudes (Godbold & Pfau, 2000; Parker et al., 2012). The latter also found a cross-protection effect where inoculating against condom use in turn conferred against alcohol use.

The theory also naturally leads to application in the political sphere. The study tested inoculation as an antidote to system-based consequences of issue advertising. Results revealed that inoculation was able to protect viewers against the consequences of party-sponsored ads. However, with PAC-sponsored ads, efficacy was restricted to Republican viewers (Pfau et al., 2001).

The theoretical reach of inoculation expanded substantially with the rise of online misinformation.

Roozenbeek and Van Der Linden (2019) demonstrated that a game-based inoculation intervention produced measurable resistance to subsequent exposure to false news. Subsequent work extended this prebunking approach to COVID-19 specifically, showing that brief inoculation messages warning about the manipulation techniques used in pandemic misinformation reduced susceptibility to those messages (Roozenbeek & Van Der Linden, 2019). Importantly, these effects did not depend on conveying accurate information about the specific false claim; they operated by building a generalized critical scepticism toward manipulative communication strategies, an attitudinal readiness rather than a factual correction.

3.1.3 Ambient and Environmental Inoculation

Consider the case of COVID conspiracy graffiti in London in 2020. Anti-lockdown and anti-government messages, "SCAMDEMIC", "#PLANDEMIC", or visual parodies of Boris Johnson, appeared on walls in graffiti dense neighborhoods Shoreditch, Brick Lane, and Brixton. For a resident passing these messages repeatedly in their everyday movement, the threat component of inoculation is operating. The graffiti signals that the dominant narrative (government competence, the legitimacy of lockdown) is not secure, that there exists a politically active counter-claim. The individual may not be convinced by the message, or even consciously aware of it. The exposure nevertheless registers that their existing beliefs are vulnerable to challenge. This is the environmental analogue of McGuire and Papageorgis (1961) threat induction. In this case, it is not a researcher warning a participant that their attitude will be attacked, but a built environment that persistently communicates the same epistemic instability.

Figure 2: SCAMDEMIC Graffiti found in London, 2020 (LDN Graffiti, n.d.)

The vast majority of empirical work on inoculation involves deliberate, designed interventions. How-

ever, McGuire's original biological analogy encompasses spontaneous as well as artificial immunity: the immune system responds to ambient exposure as well as to vaccines (McGuire & Papageorgis, 1961). The question this study addresses is whether analogous ambient processes occur in political cognition. Similarly whether repeated, unsolicited exposure to political counter-narratives in one's physical environment can generate a form of passive political inoculation.

This ambient framing connects to a broader body of work on incidental political learning and environmental priming. Research on residential neighbourhood effects has consistently shown that the political character of one's immediate environment shapes political attitudes and behaviour independently of deliberate socialisation (Huckfeldt et al., 2002; Legewie & Schaeffer, 2016). The mechanism at work in these studies is repeated, low-level exposure that normalizes certain forms of political expression and activates corresponding cognitive schemas.

The present study leans on ambient inoculation spatially. Proximity to concentrations of street art and graffiti serves as a proxy for sustained exposure to political counter-narratives in the built environment. The cumulative effect of inhabiting a space where political expression is persistently visible, the study argues, constitutes a form of environmental inoculation, a gradual habituation to political counter-narratives that shapes how individuals encounter and respond to political content in other spaces, including online.

Inoculation theory is proposed as the leading theoretical framework precisely because of the pervasive, involuntary character of graffiti as an information stimulus. Unlike deliberate media consumption like choosing to read a newspaper, following a political account on Twitter, or watching a campaign broadcast, graffiti is encountered passively and repeatedly as part of ordinary movement through urban space. A resident in graffiti-dense areas does not opt into the political messages on the walls around them, the exposure is incidental, cumulative, and continuous. Inoculation framework is well suited because it is repeated, low-level exposure that can produce a durable change in political engagement readiness. Graffiti in politically dense urban zones is, in this sense, a naturalistic analogue to the weakened doses of the inoculation model. Instead of a single persuasive intervention, graffiti exposure is a constant drip of low-intensity political signals that, over time, build resistance to blame-attribution content. When political content is encountered online, it carries less novelty and urgency as the individual has already processed the arguments, and the emotional register. The impulse to amplify is attenuated.

3.2 Political Content Amplification and Blame Attribution

3.2.1 Moral Emotion and Online Diffusion

The empirical basis for why blame-attributing political content amplifies more than non-politicized discourse is established in Brady et al. (2017) demonstration that the combination of moral evaluation and emotional activation drives diffusion at rates substantially exceeding either component alone. Specifically the finding that blame-attributing content directed at named individuals is precisely this combination. The theoretical question addresses is not the baseline moral-emotional effect but what ambient inoculation adds to it, why the amplification advantage of political over non-political content should be

narrower in graffiti-proximate areas.

The outrage/anxiety distinction established in 2.3 maps directly onto the study's content comparison. Political blame-attributing content is outrage-inducing. Non-politicized discourse alternatively activates anxiety or grief instead. Internalized responses, while intensely negative, do not generate the same impulse toward external sharing. H1 follows from this: blame-attributing political content should circulate at higher rates than non-politicized discourse. Because it is emotionally structured in a way that maps onto the social functions of Twitter sharing.

The link between inoculation theory and the amplification predictions requires one additional step. Inoculation theory, in its classical form, predicts resistance to persuasion meaning the inoculated individual is less moved by a subsequent persuasive message. The relevant extension is that inoculation builds *resistance through habituation*. to political counter-narratives hardens existing attitudes and raises the cognitive and motivational threshold for subsequent political engagement (Compton, 2012; Roozenbeek et al., 2022). An individual who has been repeatedly exposed to political expression has already processed the emotional and cognitive content of this type of messaging. The moral-emotional signal that drives amplification depends in part on novelty and urgency while repeated prior exposure diminishes both.

In the context of Twitter, amplification of political blame-attribution content is driven by the urgency of that moral-emotional signal. For an individual already habituated to political counter-narratives through environmental exposure, this signal carries less urgency. The prediction follows that graffiti-proximate individuals' threshold for amplification is higher. Ambient inoculation therefore predicts a lower rate of political content amplification in graffiti-proximate areas.

The chain is therefore: ambient exposure to politically expressive street art → habituation to political counter-narratives → higher threshold for amplification of political content encountered online → attenuated amplification rates in graffiti-proximate areas. H2 tests whether this chain produces a detectable spatial pattern, whether the amplification advantage of politicized over non-politizised content is narrower in areas with greater proximity to street art than in areas further away.

3.3 Hypothesis

The theoretical arguments developed in this chapter generate two testable predictions:

H1: Politicized content will exhibit higher amplification rates across national contexts.

This prediction follows from the moral-emotional amplification literature: blame-attribution toward named political figures activates outrage more strongly than the systemic, issue-based framing of non-politicized COVID discourse. H1 is tested cross-nationally (London, Belgium, Denmark) as an assessment of whether the predicted amplification advantage of person-targeted political content is consistent across contexts.

H2: The higher amplification rates of politicized content will be further attenuated by proximity to street art.

This prediction follows from the ambient inoculation argument: areas with higher concentrations of street art and graffiti represent politically expressive urban spaces in which inhabitants receive continuous, low-level exposure to political counter-narratives. This exposure constitutes a form of environmental inoculation that builds resistance to the amplification impulse, habituating individuals to political blame-attribution content before it is encountered online. H2 predicts that the amplification advantage of politicized over non-politicized content is moderated by proximity to street art, specifically that this advantage is dampened in graffiti-proximate areas. The amplification gap between political and non-political content narrows as proximity to street art increases. H2 is tested primarily in London, where the graffiti dataset and tweet volume are sufficient for individual-level modeling.

3.4 Scope Conditions and Design Constraints

Three constraints on the theoretical claims available from the present study require explicit acknowledgment. Together, they define what the study can establish in support of the ambient inoculation argument and what it cannot.

Cross-sectional design and temporal ordering. Classical inoculation theory requires that exposure precede outcome, the inoculation must occur before the persuasive attack for the mechanism to operate. The present study is cross-sectional, meaning proximity to graffiti and amplification behavior are measured at the same point in time, and no information about individuals' exposure histories is available. The study cannot establish that any individual was exposed to nearby street art prior to the tweeting behavior measured. What it can establish is whether the spatial pattern of amplification behavior is consistent with the predictions of an ambient inoculation process or whether graffiti-proximate areas show the aggregate signature of politically primed populations. This is a test of the spatial correlate of the theory rather than of the mechanism itself.

User composition as a design constraint. The most significant scope condition concerns the composition of the Twitter user population in graffiti-proximate areas. Ambient inoculation predicts that the politically expressive character of graffiti-dense areas activates political engagement in the people who inhabit them. The built environment shapes online sharing. A design constraint on this prediction is that high-influence, politically engaged users may independently cluster in the same urban areas where graffiti concentrates, producing a spatial pattern consistent with the inoculation prediction without any causal role for the environment itself. The cross-sectional data cannot rule this out by design.

What the data can do is test the inoculation prediction using within-user variation. If ambient inoculation is the operative mechanism, the proximity effect should persist even when stable individual quoting propensity is held constant. The same person should amplify political content at lower rates when tweeting from a graffiti-proximate area than from a graffiti-distant one. If user composition is the explanation, the proximity effect should disappear once individual-level quoting propensity is absorbed.

The within-user estimation strategy is designed specifically to perform this discrimination. This is what makes the analytical approach in Chapter 4 a genuine test of the inoculation prediction rather than simply a description of a spatial correlation.

Individual vs. area-level inference. Approximately 80% of London tweets are geolocated via place bounding box rather than precise GPS coordinates, meaning proximity to graffiti reflects an area-level characteristic of the tweet rather than an individual's specific location. H2 is accordingly framed as an area-level claim: that areas near street art show different political content amplification patterns. This is consistent with the ambient inoculation argument, which operates at the level of neighborhood political character and cumulative environmental exposure rather than event-level proximity to a specific piece.

4 Methodology

4.1 Research approach and philosophy

The study employs a post-positivist quantitative approach. It treats social phenomena, including online amplification behavior, geographic proximity, and content type, as measurable and amenable to systematic analysis, while acknowledging that observed relationships are correlational and context-dependent rather than universal laws (Saunders et al., 2019).

The research question central to the exploration is:
To what extent does proximity to street art predict the nature and amplification of politicized content on Twitter?

To address this, the study adopts a computational social science approach, analyzing digital trace data in the form of geolocated tweets, as naturally occurring behavioral records. The design is observational and cross-sectional. Tweets and graffiti locations are treated as co-occurring in space. Proximity is a structural feature of that co-occurrence, not evidence of individual exposure. All findings are explicitly correlational.

This framing has two implications for interpretation. First, proximity effects should be understood as area-level patterns. Finding that certain content is more common near graffiti could reflect exposure to street art, or could reflect the character of the neighborhoods that graffiti tends to cluster in. Proximity is therefore treated as a geographic correlate, not a mechanism.

4.2 Data Collection

4.2.1 TBCOV Dataset

The primary dataset for exploration is the TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels. TBCOV is a large-scale social sensing dataset comprising

two billion multilingual tweets posted from 218 countries by 87 million users in 67 languages. The data was collected over a 14-month period from 1 February, 2020 to 31 March, 2021 by Imran et al. (2022). Alongside the raw tweet text, TBCOV provides pre-computed labels for sentiment, named entities, geolocation, and inferred gender, making it well-suited to a study that requires both spatial and content-level analysis without constructing those features from scratch.

The data is downloadable for each of the 218 countries available. For this paper, three locations were selected: the United Kingdom, Belgium, and Denmark. The United Kingdom provides the primary analysis context, specifically the Greater London area, while Belgium and Denmark serve as cross-national comparison cases. Selection was guided by the availability of sufficient geolocated tweets and the existence of corresponding graffiti data for each location.

Geographic filtering was applied to retain only tweets with usable location signals. Both GPS coordinates and place bounding box coordinates (a rectangular bounding box associated with a named place, such as a borough or neighborhood) were kept. Tweets geolocated only at country or region level were excluded using a maximum bounding box area threshold of 50 km², consistent with practices to exclude coarse spatial data (Graham et al., 2014). This threshold excludes large administrative units while retaining city and district-level place tags, which helps increase spatial precision for proximity analysis.

4.2.2 Graffiti Dataset

The Graffiti Database (Graffiti Database, n.d.) is a world-wide online graffiti archive. The collective project, supported by the Dutch Graffiti Library, includes over 44,000 images from 66 countries and 51 cities. With permission from the owner, this dataset was used for geolocation of graffiti pieces in London, Belgium, and Denmark. 835 pieces from London were used, 1,222 for Belgium, and 727 for Denmark. Each entry includes geographic coordinates, city, country, year, artist name, and piece name where available. This data serves as a spatial reference point for graffiti proximity calculation.

The Graffiti Database's intention is self-described: "The intention of this mostly graffiti related site is to share this huge collection of pieces, burners, tags and sketches with like minded souls" (Graffiti Database, n.d.). The collective database in turn should not be interpreted as an all-encompassing work. It is community-maintained rather than systematically collected, so coverage reflects what contributors chose to submit rather than the full distribution of street art in any given city. The site most prominently features large artistic graffiti pieces like murals, and may not capture smaller tags or pieces that are seen in dense cities. Graffiti is also an ephemeral artform that can be removed or created in short timeframes. There is no guarantee that a piece included in this database was present during the 2020 study period. It is impossible to find a comprehensive, live, maintained graffiti database without significant work creating one from scratch. Therefore, the Graffiti Database is used in this study as a spatial snapshot of graffiti expression rather than an exhaustive record. Limitations of this caveat are further explored in Chapter 6.

Access to the graffiti coordinate data is not publicly shared in this study. Graffiti remains illegal in many contexts, and the site owner expressed concern that precise location data could be used to notify

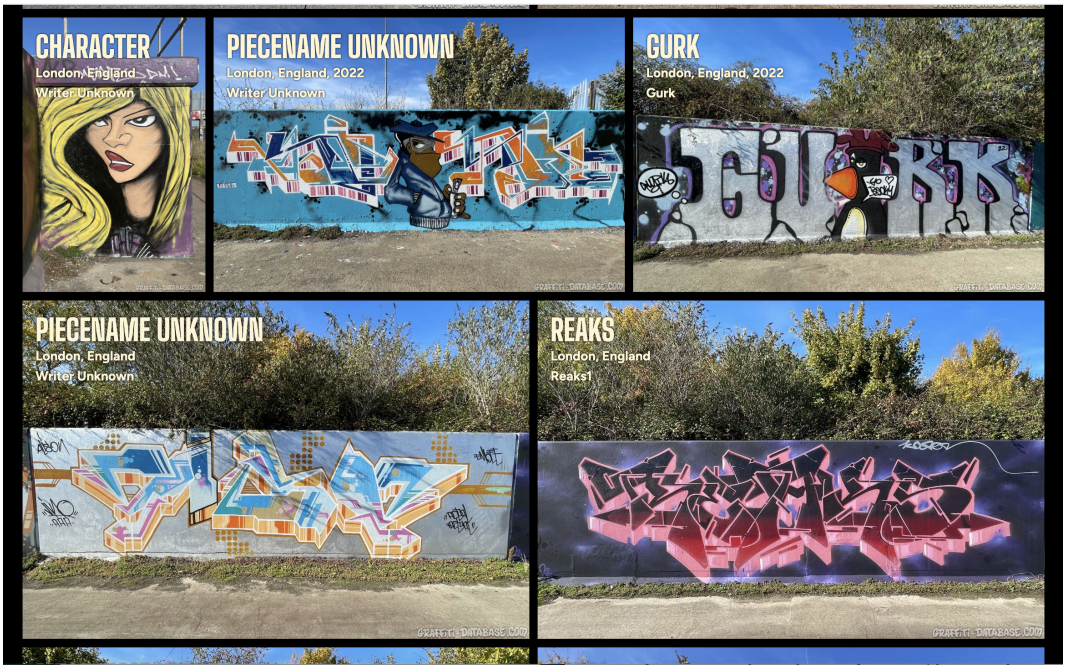

Figure 3: Screenshot from graffiti-database.com (Graffiti Database, n.d.)

law authorities. The graffiti piece coordinates are retained locally for analysis and are not included in any published dataset or appendix.

4.3 Data Preprocessing and Exploratory Data Analysis

Prior to analysis, the merged proximity datasets were subjected to a structured preprocessing and exploratory data analysis pass to document data quality, resolve ambiguities in geocoding, and define the cleaning decisions. No rows were removed at this stage, but instead problematic observations were flagged with binary indicators and handled analytically during data analysis.

4.3.1 Data filtering and geographic bounding

The TBCOV United Kingdom downloaded files cover United Kingdom territories England, Scotland, Wales, and Northern Ireland. To restrict the primary analysis to the Greater London area, a bounding box filter (51.28°N–51.69°N, −0.51°E–0.33°E) was applied, retaining only tweets whose resolved coordinates fall within Greater London. Two further geocoding filters were applied uniformly across all three datasets: (a) only tweets geolocated via GPS coordinates or place bounding box centroid were retained. Tweets located via user-stated location field or tweet text, were excluded because they carry no reliable spatial precision; (b) place bounding boxes larger than 50 km² were discarded before centroid computation, removing country and region-level bounding boxes that would have introduced coarse proximity estimates.

4.3.2 Data quality checks

No duplicate `tweet_id` or `(tweet_id, user_id)` pairs were found across any of the three datasets. Thus no de-duplication was required.

Three fields had significant missing values. The `county` field was missing for 90% of London rows, 68.2% of Belgian rows, and 100% of Danish rows and was subsequently dropped. The `city` field was missing for 5.5% of London rows, 91.9% of Belgian rows, and 86.9% of Danish rows. Given its heavy absence in Belgium and Denmark it was retained in the dataset but not used as an analytic variable. The `named_entities` field, containing pre-computed Named Entity Recognition (NER) output, was missing for 24.8% of London tweets, 19.4% of Belgian tweets, and 17.8% of Danish tweets. These represent tweets for which the TBCOV NER pipeline detected no entities mentioned. These entries receive a flag `no_entity = 1`, and are excluded only from entity-dependent analyses.

4.3.3 Exploratory findings

Several patterns observed during EDA directly informed analytic decisions.

Geocoding composition Geolocation is very rare in Tweet datasets. Location sharing is an opt-in setting, and reactive behaviors like quote tweeting, retweeting, and replying, are less likely to include geolocation data than original posts. As mentioned above, the cleaned datasets retain only tweets with GPS coordinates or place bounding box geolocation. The large majority of retained tweets are thus place bounding box-geolocated: approximately 80.6% of London tweets, 90.0% of Belgian tweets, and 83.5% of Danish tweets. The place bounding boxes are not precise coordinates but instead identify a neighborhood or location and geoencode that place. The datasets then have a bbox centroid problem, where multiple tweets tagged with the same named place share a single centroid coordinate, meaning the tweets-per-coordinate distribution is strongly right-skewed. This motivates the GPS-only robustness check discussed in the later section.

Proximity distribution The `nearest_graffiti_km` variable measures the distance from each tweet's coordinates to the nearest registered graffiti piece. The distribution is strongly right-skewed in all three countries, with a small number of tweets located far from any registered piece pulling the mean well above the median: London (median 1.01 km, mean 2.27 km), Belgium (median 1.65 km, mean 4.71 km), and Denmark (median 4.70 km, mean 5.30 km). The progressively larger gap between median and mean across the three countries reflects the sparser graffiti coverage in Belgium and Denmark relative to London, where the 835 registered pieces are more densely and evenly distributed across the city. Raw distance is not suitable for regression in this form. The long right tail violates the linearity assumption and would give disproportionate influence to a small number of distant observations. A log-transformation

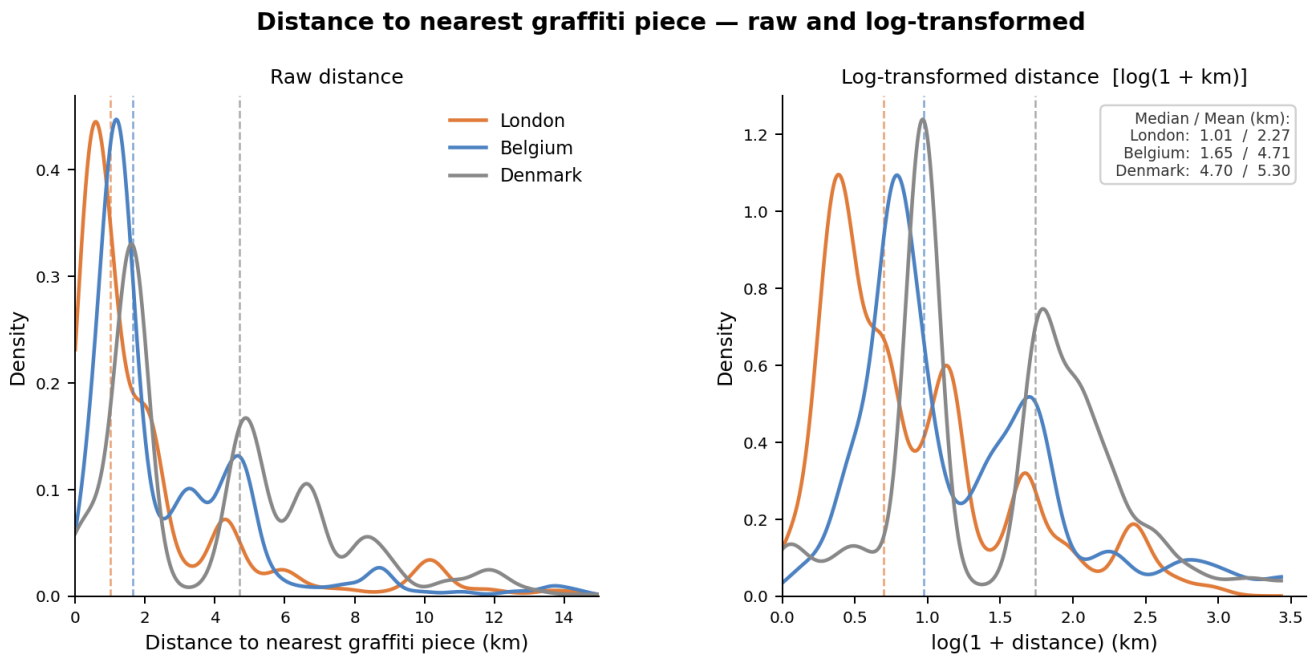

Figure 4: Log transformed distance

compresses the tail and produces approximately symmetric distributions across all three locations, making the variable suitable for use as a continuous predictor. The addition of 1 before logging ensures that tweets with a nearest piece at zero distance are handled without producing undefined values. Proximity distributions also differ substantially between GPS and bounding box tweet subsets within some countries. GPS tweets, which reflect precise posting locations, tend to show different distance profiles than bbox tweets, whose proximity is calculated from an area centroid. This divergence means the two subsets are not directly comparable on the proximity measure, and provides a further justification for running the GPS-only robustness check.

Heavy users Twitter activity is highly unequally distributed, a small number of prolific accounts can contribute a disproportionate share of posts. If those accounts are spatially clustered, they risk distorting area-level proximity signals. In the London dataset, the single most active user contributed 1,918 tweets, and the activity distribution is strongly right-skewed. Users with more than 50 tweets in the dataset were flagged as `is_heavy_user = 1`, a threshold chosen to capture accounts whose volume is high enough to plausibly influence aggregate statistics. An initial comparison of heavy and regular users found similar raw proximity and amplification profiles at the descriptive level. Median proximity to the nearest graffiti piece was identical across both groups at 1.01 km, and quote rates were nearly indistinguishable (31.6% for heavy users and 32.3% for regular users). However, raw aggregate comparisons do not capture whether heavy users concentrate within particular content types or interact differently with the political figure attribution signal. Given that a third of London rows come from accounts exceeding the threshold, the `is_heavy_user` flag is used to re-run the primary regressions on the regular-user subset

as a robustness check. Whether heavy users substantively alter the estimates is assessed in the results.

Sentiment and named entity composition . TBCOV creators, Imran et al. (2022), used XLM-Negative sentiment as the dominant classification model, because of its success in multilingual text. Across all three countries: 40.5% of London tweets, 39.2% of Belgian tweets, and 32.5% of Danish tweets are labelled negative. This is consistent with prior work on pandemic discourse, which has documented a persistent negativity bias in COVID-related social media content (Boon-Itt & Skunkan, 2020). Negative-labelled tweets also carry the highest classifier confidence scores in the TBCOV pre-computed labels, while neutral classifications carry the lowest, suggesting that neutral tweets are more ambiguous in their linguistic signal and should be interpreted with some caution. The overall prevalence of negative sentiment in the corpus contextualises the supplementary sentiment analysis, where politicized and non-politicized tweets are compared on the valence level.

Across all three datasets, `ORG` (organisation) and `GPE` (geopolitical entity) are the most frequent named entity labels, followed by `CARDINAL`, `PERSON`, and `DATE`. `COVID-Entity` is less frequent than these, appearing sixth in London and Belgium and sixth in Denmark, reflecting that the NER model flags specific virus and policy terminology rather than the broad topic of every tweet. `GPE` and `ORG` capture the institutional and geographic framing that characterises pandemic discourse; references to countries, cities, governments, and health authorities. `PERSON` entities are analytically central to H1 and H2. In London, the highest-frequency `PERSON` terms are political figures: boris (n=472), boris johnson (392), johnson (285), trump (192), and cummings (172), confirming that political blame attribution discourse is present in the corpus and concentrated around a small number of individuals. In Belgium, the top `PERSON` terms are predominantly international figures, including Trump, Biden, Bolsonaro, and Merkel, rather than domestic Belgian politicians. In Denmark, the NER model performs poorly on Danish-language text. The top `PERSON`-labelled terms are common Danish words ('tak', 'jeg', 'der', 'hvis') misclassified as named entities, meaning only a single reliable political figure, Frederiksen, can be extracted. This NER limitation directly shapes the political figure term lists used in H1 and H2, which were constructed from the highest-frequency genuinely political `PERSON` terms in each national dataset.

4.4 Measurement

4.4.1 Dependent variable: amplification

Tweet amplification is computed as whether a tweet is a quote tweet (`tweet_type == 'quote'`), yielding the binary variable `is_quote`. Understanding why this is the appropriate measure requires briefly explaining how Twitter's sharing mechanisms interact with geolocated data collection. Twitter offers two primary ways to share other's content: retweeting and quote tweeting. A retweet is a direct reshare with no added content, while a quote tweet embeds the original post and appends the sharing user's commentary. These two behaviours are treated differently by Twitter's API with respect to location metadata. When a user retweets, the API attributes the location of the original tweet rather than the retweeter's.

Location-filtered datasets, and TBCOV therefore systematically exclude retweets, as the retweeter's geographic position is never recorded. This is not a data quality problem but a structural feature of how Twitter handles retweet metadata, and it applies consistently across all geolocated COVID-19 tweet collections (Imran et al., 2022).

Quote tweets behave differently. Because a quote tweet is technically an original post, the user is composing new content that embeds another tweet, it carries its own location metadata at the time of posting. Quote tweets are therefore retained in the filtered dataset and represent the primary available measure of content amplification. `is_quote` captures deliberate amplification with commentary, a user has chosen to share content and respond to it, contextualise it, or add commentary. This is a meaningful measure of engaged amplification as it requires input from the sharer.

4.4.2 Independent variables: content type

TBCOV provides pre-computed NER tags for each tweet, stored as a list of dictionaries. Two content-type indicators were derived from these tags.

Politicized content (`has_pol_figure = 1`): the tweet contains at least one named entity whose normalized term matches a pre-specified list of political blame terms relevant to the pandemic period and geographic context. Matching is performed on term value across all NER label types, because the TBCOV NER pipeline assigns labels inconsistently to the same political referents across tweets. For example, collective terms such as 'tories' may be tagged `NORP`, `ORG`, or `PERSON` depending on context. The list was constructed to capture direct political blame discourse: tweets that name or collectively reference political actors held responsible for pandemic outcomes, whether as individuals or as a governing group. The list was derived by inspecting the top entity terms in each national dataset and is defined separately per country:

- **UK terms**: boris, boris johnson, bojo, johnson, johnsons, cummings, dominic cummings, matt hancock, hancock, raab, rishi sunak, sunak, patel, priti patel, keir starmer, starmer, sadiq khan, tory, tories, trump, donald trump, donaldtrump, biden, joe biden, joebiden, fauci, putin
- **Belgium additions:** de wever, frank vandenbroucke, vandenbroucke, sophie wilmes, sophie wilmès, wilmès, charles michel, macron, emmanuel macron, merkel, angela merkel, bolsonaro, jair bolsonaro, viktor orbn, viktor orban, orban
- **Denmark additions:** mette frederiksen, frederiksen, mette

The inclusion of collective terms ('tory', 'tories') reflects the discourse structure of UK COVID Twitter, where blame attribution operated both at the level of named individuals and at the level of the governing party as a collective agent. Tweets referencing 'tories' in this corpus function as blame-attributing discourse, assigning responsibility to a specific political actor, rather than as neutral institutional description. Generic party labels without blame framing (e.g. "Labour", "the Conservatives" used descriptively)

do not appear in the term list. Belgium-specific terms were expanded following inspection of the top entity terms in the Belgian dataset; the dominant figures were international (Trump, Merkel, Macron, Bolsonaro) rather than Belgian, reflecting the character of Belgian COVID discourse in TBCOV. Danish NER performance on Danish-language text was very poor. Common Danish words were misclassified as `PERSON` entities, meaning Frederiksen is the only reliably extracted Danish political figure. The `has_pol_figure` flag is re-derived from raw `named_entities` field at analysis time using this list, rather than relying on the pre-computed column in the cleaned CSVs. Robustness to term list variation is confirmed in Chapter 5.

Non-politicized content (`has_pol_figure = 0`, `no_entity = 0`): all tweets in the corpus that do not contain any political blame term and contain at least one detected entity. Because the entire corpus is drawn from TBCOV, no additional COVID-entity filter is applied. The distinction is one of framing: direct political blame versus all other COVID discourse. No-entity tweets (`no_entity = 1`) are excluded from this group for the same reason they are excluded from all entity-dependent analyses. Their high empirical quote rate is a structural artefact of short or commentary-only quote tweet text being entity-free, not a substantive content signal.

Organization entity content is not collapsed into the politicized group. `ORG`-labeled tweets cover a heterogeneous set of institutional actors from health authorities, media organizations, to political parties. They contain substantially lower quote rates than politicized content. Collapsing `ORG` into the politicized category would dilute the blame-attribution signal that is the conceptual core of the politicized operationalization. Label `has_ORG = 1` is retained as a control variable in the H2 regression.

4.4.3 Independent variables: proximity

Proximity to graffiti is measured at two levels. The continuous measure, `log_km`, is the log-transformed distance to the nearest registered graffiti piece correcting for the right skew in the raw distance distribution seen in Figure 4. The binary measure, `very_near`, flags tweets within 500 m of at least one graffiti piece (25.7% of London tweets, n = 29,902).

For analyses requiring a density measure, `graffiti_count_500m` counts the number of graffiti pieces within 500 m of each tweet's coordinates, computed via BallTree radius query on the full 835-piece London dataset. An equivalent count at 1,000 m (`graffiti_count_1000m`) is included in robustness checks. Because 74.3% of tweets have zero pieces within 500 m and the distribution is sharply right-skewed (max = 265), the density measure is log-transformed:

$$\texttt{log\_density\_500m} = \log(1 + \texttt{graffiti\_count\_500m}) \tag{1}$$

The 500 m threshold is consistent across binary and count measures by construction: `very_near = 1` and `graffiti_count_500m > 0` are exactly equivalent. The density variable adds information beyond

the binary by distinguishing tweets near a single piece from those embedded in a dense graffiti zone.

4.5 Data Analysis

4.5.1 H1: Content type and amplification

H1: Politicized content will exhibit higher amplification rates across national contexts.

Political attribution transforms a diffuse systemic crisis into a targeted, person-directed narrative. Blame-directed discourse is more emotionally activating and more amenable to expressive sharing, outrage at a named individual circulates in a way that geographic or institutional framing of the same crisis does not. Tweets referencing named political figures are therefore predicted to be quoted at a higher rate than the broader field of COVID discourse that makes no such attribution.

Descriptive test H1 is tested using a chi-square test of independence on the 2×2 contingency table of content type (politicized vs. non-politicized) by amplification (`is_quote = 1` vs. `0`), pooled across all three countries, then disaggregated by country. Cross-national consistency is treated as an exploratory test of scope rather than a primary prediction. Given the London-centric composition of the political figure list and the small number of politicized tweets in Belgium (n = 311) and Denmark (n = 18), replication in the secondary countries is informative but not required to confirm H1. Effect size is reported as the percentage-point gap in quote rates within each country.

Logistic regression To test whether the politicised–amplification association holds when controlling for other content characteristics and spatial proximity, binary logistic regression is estimated per country and pooled:

$$\begin{aligned}\text{logit}\left(P(\texttt{is\_quote} = 1)\right) = \beta_0 \\ + \beta_1 \cdot \texttt{has\_pol\_figure} \\ + \beta_2 \cdot \texttt{has\_COVID} \\ + \beta_3 \cdot \texttt{has\_ORG} \\ + \beta_4 \cdot \texttt{sentiment\_label} \\ + \beta_5 \cdot \log(\texttt{km}) \\ + \varepsilon\end{aligned} \tag{2}$$

The coefficient of interest is β_1 (`has_pol_figure`), reported as a log-odds coefficient and exponentiated odds ratio (OR > 1 = politicized more likely to be quoted). Controls capture the independent effects of COVID-entity content, organization-referencing content, tweet sentiment, and proximity to graffiti, each of which independently predicts amplification and may be correlated with political figure attribution. The pooled model includes country fixed effects (Belgium and Denmark indicators, London as

reference). Standard errors are clustered by `user_id` to account for within-user correlation. Per-country and pooled models are estimated on the same non-reply, non-null-entity sample used in the descriptive analysis.

Robustness checks Twitter activity is heavily concentrated in a small number of accounts: in the London sample, the top user contributed 1,918 tweets and 32.5% of rows come from users with more than 50 tweets. If these heavy users post a disproportionate share of political content and have different baseline quoting tendencies from regular users, the political amplification gap observed in the full sample could be a compositional artifact of that subset rather than a generalizable content-type effect. To test this, the regression is re-estimated on the non-heavy-user subsample (n = 69,171). If the gap holds among regular users, it cannot be attributed to prolific account activity.

The logistic regression with clustered standard errors treats repeated observations from the same user as independent after SE correction, but does not model within-user effects. Meaning it conflates compositional differences between users (some users quote more, some less) with content-type effects. To separate these, a mixed effects logistic regression with a random intercept by `user_id` is estimated using `lme4::glmer` in R:

$$\begin{aligned}\text{logit}\left(P(\text{is\_quote}=1)\right) = \beta_0 &+ \beta_1 \cdot \text{has\_pol\_figure} + \beta_2 \cdot \text{has\_COVID} + \beta_3 \cdot \text{has\_ORG} \\ &+ \beta_4 \cdot \text{sentiment\_label} + \beta_5 \cdot \log(\text{km}) + (1 \mid \text{user\_id})\end{aligned} \tag{3}$$

The random intercept absorbs each user's overall propensity to quote, allowing β_1 to estimate the within-user effect of content type: does the same user get quotes when tweets are politicized/have blame-attribution at a higher rate than their non-politicized tweets? This is a more demanding test than the population-level comparison. The intraclass correlation coefficient (ICC) is computed from the random effects variance and the logistic distribution residual variance ($\pi^2/3$), and quantifies what proportion of total quoting variability is attributable to user-level differences rather than content or spatial features. A high ICC implies that user identity is the dominant predictor of tweet amplification. Simpler models which ignore this nuance risk attributing user-composition differences to content-type effects. The model is fitted on the full London sample (n = 100,746; 20,565 users); convergence is achieved via the `bobyqa` optimizer.

Supplementary sentiment analysis. Politicized tweets are expected to be more negatively valenced than non-politicized tweets. Blame attribution towards named individuals generates intense outrage-driven content, whereas non-politicized COVID discourse is more informational in character and therefore less extreme. This is not a formal hypothesis but is reported as a descriptive finding to enrich the interpretation of H1. Sentiment scores (−1 negative, 0 neutral, +1 positive) are pre-computed in TBCOV. Group differences are tested with the Mann-Whitney U test (one-sided: politicized < non-politicized),

appropriate for an ordinal variable with a non-normal distribution.

4.5.2 H2: Proximity Moderates Content-Type Amplification

H2: The amplification advantage of politicized content will be attenuated by proximity to graffiti.

H2 posits a moderation effect: proximity to street art attenuates the relationship between politicized content and amplification. It is tested using binary logistic regression:

$$\begin{aligned}\text{logit}\left(P(\texttt{is\_quote} = 1)\right) = \beta_0 \\ + \beta_1 \cdot \texttt{has\_pol\_figure} \\ + \beta_2 \cdot \texttt{has\_COVID} \\ + \beta_3 \cdot \texttt{has\_ORG} \\ + \beta_4 \cdot \log(\texttt{km}) \\ + \beta_5 \cdot \texttt{very\_near} \\ + \beta_6 \cdot (\texttt{has\_pol\_figure} \times \texttt{very\_near}) \\ + \beta_7 \cdot (\texttt{has\_COVID} \times \texttt{very\_near}) \\ + \beta_8 \cdot \texttt{sentiment\_label} \\ + \varepsilon\end{aligned} \tag{4}$$

The key coefficient of interest is β_6 (`pol_figure × very_near`). A negative and significant β_6 would indicate that politicized content's amplification advantage is smaller in areas near graffiti than in areas further away, consistent with the predicted attenuation hypothesis. A positive β_6 would indicate the reverse: the amplification gap widens near graffiti, which would contradict the inoculation prediction and require an alternative explanation. β_7 (`COVID × very_near`) provides a within-model comparison for COVID-entity-framed content specifically (health, disease, and policy terminology). A null β_7 alongside a significant negative β_6 would indicate that the attenuation effect is selective or specific to political blame-attribution discourse rather than COVID-entity content generally. Note that `has_COVID` in the regression captures the NER `COVID-entity` label specifically and is not equivalent to the full non-politicized group used in H1. The model is estimated via maximum likelihood using `statsmodels.formula.api.logit`. Standard errors are clustered by `user_id` to account for within-user correlation. Cluster-robust SEs are more conservative than HC3 and avoid inflating test statistics due to user-level clustering. Results are reported as log-odds coefficients with p-values, odds ratios are provided for interpretability. Model fit is reported as AUC and McFadden pseudo-R^2.

To test whether the proximity effect reflects a dose-response relationship rather than a simple presence/absence threshold, a second model substitutes `log_density_500m` and its interaction with `has_pol_figure` for the binary `very_near` terms. A significant positive interaction in this model would indicate that amplification of political content increases with graffiti density, not merely with the pres-

ence of any nearby piece. The 1,000 m density model is run as a comparison to assess spatial decay.

Robustness checks For H2, the heavy user concern is more specific than in H1. If politically active heavy users are spatially concentrated near graffiti, the interaction term (`pol_x_verynear`) would absorb user type rather than a genuine content × location effect. A geographically clustered subset of prolific accounts could produce an apparent proximity interaction even if no such relationship exists among regular users. The regression is re-estimated on the non-heavy-user subsample (n = 69,171) to test whether the interaction survives their removal.

Then the same modeling logic applied in H1 is extended to H2 with a mixed effects logistic regression. A mixed effects logistic regression with a random intercept by `user_id` tests whether the proximity–content-type interaction persists once user-level quoting tendencies are absorbed:

$$\begin{aligned}\text{logit}\left(P(\texttt{is\_quote}=1)\right) = \beta_0 &+ \beta_1 \cdot \texttt{has\_pol\_figure} + \beta_2 \cdot \texttt{has\_COVID} + \beta_3 \cdot \texttt{has\_ORG} \\ &+ \beta_4 \cdot \log(\texttt{km}) + \beta_5 \cdot \texttt{very\_near} \\ &+ \beta_6 \cdot (\texttt{has\_pol\_figure} \times \texttt{very\_near}) \\ &+ \beta_7 \cdot (\texttt{has\_COVID} \times \texttt{very\_near}) \\ &+ \beta_8 \cdot \texttt{sentiment\_label} \\ &+ (1 \mid \texttt{user\_id})\end{aligned} \tag{5}$$

The key coefficient of interest remains β_6 (`pol_figure × very_near`). The mixed effects model tests whether, within the same user's tweet history, politically attributed tweets originating near graffiti are quoted at a higher rate. A non-significant β_6 in the random-intercept model would indicate that the proximity interaction observed in simpler models is a user composition effect rather than a genuine content-location interaction. A graffiti density extension replaces the binary `very_near` with `log_density_500m` and its interaction with `has_pol_figure`; robustness variants exclude heavy users (> 50 tweets). All mixed effects models are estimated with `lme4::glmer` in R using the bobyqa optimiser.

H2 is not tested in Belgium or Denmark as the primary analysis. A cross-national replication of H2 in Belgium was conducted as an exploratory comparison. Given the smaller sample size, sparser graffiti dataset, and differences in urban political geography between London and Belgian cities, non-replication is not treated as falsifying H2 but as informative scope evidence discussed in Chapter 5.

4.5.3 Causal Inference and Robustness Checks

The observational design means that associations between content type, proximity, and amplification may reflect user selection rather than causal mechanisms. Four additional analyses address this threat, all run on the London sample after the same reply and no-entity exclusions applied. They address two distinct threats in sequence. Coarsened Exact Matching (CEM) tests whether the H1 and H2 patterns

survive covariate balance correction. If they do, the results are not artifacts of observable differences between treated and control groups. The conditional logit then relaxes the random effects assumption entirely, absorbing all stable user characteristics and identifying from within-user variation alone. The distance decay and spatial autocorrelation analyses test the mechanistic prediction directly, a genuine proximity effect requires a dose-response relationship and spatial co-occurrence between graffiti and political amplification that should be visible in the data regardless of modeling approach.

Coarsened Exact Matching CEM (Iacus et al., 2012) was applied separately for H1 and H2. CEM is preferred over Propensity Score Matching (PSM) because it guarantees covariate balance within matched strata rather than on a scalar propensity score, and is less sensitive to model misspecification, PSM approximates a completely randomized experiment whereas CEM approximates the more efficient block randomized design (King & Nielsen, 2019). With four to five matching covariates, exact matching within coarsened bins was feasible without excessive sample loss.

H1 CEM The treatment variable is `has_pol_figure`. Matching covariates are `sentiment_label`, `is_heavy_user`, and `has_ORG`. Distance to graffiti (`log_km`) is deliberately excluded from the matching formula as it is the outcome-adjacent variable under investigation in H2, not a confounder of the content-type comparison. Two post-match regressions are estimated on the matched sample. A weighted GLM with user-clustered SEs (`sandwich::vcovCL`) and `lme4::glmer` with `(1|user_id)`, to allow comparison between population-level and within-user estimates after balance correction.

H2 CEM The treatment variable is `very_near`. Matching covariates are `has_pol_figure`, `has_COVID`, `has_ORG`, `sentiment_label`, and `is_heavy_user`, ensuring that near-graffiti and far-graffiti tweet groups are comparable in content composition and user type before the proximity interaction is re-estimated. The post-match regression replaces the binary `very_near` threshold with the continuous log-distance interaction (`pol_x_logkm = has_pol_figure` $\times$ `log_km`), avoiding an arbitrary 0.5 km cutoff and remaining consistent with the gradual ambient exposure implied by the inoculation framing. A negative `pol_x_logkm` coefficient would indicate that political content amplifies more as distance to graffiti decreases, consistent with the theoretically predicted direction. Both clustered SE and mixed effects specifications are estimated. Replication of the unmatched pattern after CEM rules out covariate imbalance as an explanation; absorption of the effect by the mixed effects random intercept after matching would indicate that the remaining signal is a between-user composition effect rather than a content or spatial mechanism.

Spatial Autocorrelation — Moran's I A proximity mechanism also requires that political amplification and graffiti density co-occur in the same parts of the city. Global Moran's I is computed for political quote rate, non-political quote rate, the amplification gap, and log graffiti density at the 1km² grid cell level, using KNN spatial weights (k = 5) and 999 Monte Carlo permutations. Bivariate Moran's I jointly

tests whether high-amplification cells neighbor high-graffiti cells. LISA cluster maps identify any significant local co-clustering. Estimation uses `esda` and `libpysal` (Python). A null or negative bivariate Moran's I would indicate that political amplification and graffiti concentration occur in different parts of London, directly contradicting a city-scale spatial exposure mechanism.

4.5.4 Supplementary Sentiment Analysis

Beyond the H1/H2 tests, the sentiment composition of the corpus is reported as a descriptive complement to the hypothesis results. TBCOV sentiment labels (−1 negative, 0 neutral, +1 positive) are not used as a predictor in any primary model, but provide a measure of emotional valence that can illuminate the character of the amplification patterns found. Two questions motivate this supplementary analysis.

The first is whether content types differ systematically in emotional valence. Blame-attributing discourse directed at named political figures is expected to be more intensely negative than non-politicized COVID content: accountability framing is emotionally activating in a way that informational or issue-based pandemic discourse is not. Mean sentiment and the share of negatively labeled tweets are reported by content group. The NHS group is separated from the broader non-politicized category because solidarity discourse around health-worker recognition carries a distinct affective register from health-informational content, and collapsing the two would obscure variation in the comparison baseline. Group differences are tested via Mann-Whitney U.

The second question is whether the political content amplification advantage is simply a negativity effect. That is, whether political tweets quote at a higher rate merely because they are more negative, rather than because of anything specific to political blame attribution. If amplification were proportional to negative sentiment, the ordering of content groups by quote rate should mirror their ordering by negative sentiment prevalence. A disproportionate amplification advantage in political content, beyond what negativity alone would predict, would indicate that the amplification signal is not reducible to valence. The comparison of sentiment profiles and quote rates across content groups provides a descriptive test of this interpretation.

4.5.5 Entity Clustering

As a complement to the regression-based hypothesis tests, an unsupervised clustering analysis explores whether interpretable groupings emerge from the joint distribution of entity type, graffiti proximity, and amplification. The feature matrix consists of 11 variables: row-normalised proportions of eight NER entity label types (`GPE`, `ORG`, `PERSON`, `LOC`, `NORP`, `DATE`, `CARDINAL`, `COVID-Entity`), `is_quote` (weighted ×3), `log1p(nearest_graffiti_km)` normalized (weighted ×3), and mapped sentiment (−1→0, 0→0.5, 1→1). The 3× weighting on amplification and proximity prioritises the research-relevant axes,whether a tweet was amplified and how close to graffiti it originated, while still allowing entity content to differentiate clusters with similar amplification and proximity profiles. Without upweighting, the eight entity dimensions would collectively dominate the embedding regardless of their substantive relevance to the

research question.

Dimensionality reduction precedes clustering: UMAP (`umap-learn`) with `n_neighbors=30`, `n_components=10`, `min_dist=0.0`, `n_epochs=200`. KMeans (`n_clusters=12`, `n_init=10`, `random_state=42`) is applied to the UMAP embedding. HDBSCAN was evaluated as an alternative but abandoned. At any `min_cluster_size`, approximately 45–50% of tweets were assigned to noise due to the large share of tweets with no entity features (flat regions in UMAP space). KMeans assigns every point and produces more interpretable clusters at the cost of assuming spherical cluster geometry.

Cluster characterization reports the dominant entity types, mean proximity, and quote rate per cluster. The analysis is exploratory and is reported to identify whether the regression findings are visible in the unsupervised structure of the data.

5 Results

5.1 Descriptive

5.1.1 Data Composition

The final cleaned dataset includes 134,760 tweets across three national contexts. 116,391 London entries, 16,012 Belgium entries, and 2,357 Denmark entries. London is the primary location for analysis, while Belgium and Denmark serve as secondary comparison cases. The size differences reflect the underlying data rather than any variation in collection or filtering, Denmark in particular yielded few tweets because the TBCOV corpus contains few geocoded Danish tweets.

Tweets were geolocated with two main mechanisms as described in Chapter 4. GPS coordinates of tweets were pulled from tweets where users enable location services or where the IP address can be geocoded (Zohar, 2021). This GPS geolocation is relatively scarce in datasets, the TBCOV dataset included 19% (n=22,582) in London, 10% (n=1,600) in Belgium, and 16.5% (n=388) in Denmark of tweets with GPS coordinates. The remainder were assigned to the centroid of a named place bounding box, at neighbourhood or borough level rather than a precise GPS point. This distinction is consequential for H2, where proximity to graffiti is an area-level measure for the majority of tweets rather than an individual one.

	London	Belgium	Denmark
Total tweets	116,391	16,012	2,357
GPS coordinates	22,582 (19.4%)	1,600 (10.0%)	388 (16.5%)
Place bounding box	93,809 (80.6%)	14,412 (90.0%)	1,969 (83.5%)
Replies excluded	15,645 (13.4%)	2,137 (13.3%)	374 (15.9%)
Working sample (non-reply)	100,746	13,875	1,983
Unique users	20,565	2,774	624

Table 1: Dataset composition by country. Working sample excludes replies.

Two exclusions were applied consistently across all three contexts. Replies were removed from the H1 and H2 analyses because they are typically directed at a specific user rather than broadcast to a general feed. Reply amplification behavior is structurally different from the original posts and would confound the comparisons the hypotheses require. Zade et al. (2024), find that the quote retweet interaction is used more broadly to broadcast ideas, while the reply is used to reframe the conversation. Additionally, tweets with no NER labels were excluded from entity-dependent analyses. Tweets with no NER label contain no identifiable political or COVID entities and cannot be assigned to either politicized or non politicized content category.

5.1.2 Content Type Distribution

Politicized tweets, those attributing pandemic outcomes to named political figures, make up a small fraction of each corpus: 2.6% in London (n=2,589), 2.4% in Belgium (n=340), and 1.0% in Denmark (n=20). Non-politicized entity-bearing content is more prevalent with 73,200 tweets in London, 10,751 in Belgium, and 1,630 in Denmark. The bulk of each corpus consists of tweets that mention organizations, places, dates, and other entities without necessary political attribution. Ambient COVID-era discourse rather than blame-directed content.

The named entity label distributions reflect this. In London and Belgium, `ORG` is the most frequently occurring label, followed by `GPE` and `PERSON`. `COVID-Entity` labels appear in a meaningful share of tweets but well below the organizational and geographic references that dominate both corpora. Denmark's distribution is structurally different: `PERSON` labels appear in 89.3% of tweets, far above London (26.1%) and Belgium (29.3%). This is almost certainly a NER failure rather than a genuine content difference, the model used for entity extraction performs poorly on Danish text and systematically mislabels terms as `PERSON` entities. This is why the politicized content operationalization is unreliable for Denmark, and why only Frederiksen could be extracted with any confidence. Denmark's n=18 politicized tweets should be read with that in mind.

The character of politicized content also differs across countries. In London, the political figure term list is dominated by domestic figures, reflecting a discourse focused on UK government accountability during the pandemic (see Appendix A.3). Belgium's list includes some domestic figures, but also required supplementing with international figures. Even with international figures included, the Belgian dataset only includes 340 politicized tweets. Suggesting that Belgian COVID discourse on Twitter was less anchored to domestic political attribution and more oriented toward global political commentary. Denmark's low number of political tweets in combination with it's NER failures prompts excluding it as a location in analysis.

5.1.3 Proximity Distribution

Distance to the nearest graffiti piece varies considerably across the three contexts. London tweets have a median distance of 1.01 km and a mean of 2.27 km, with a farthest long right tail extending to 23.19

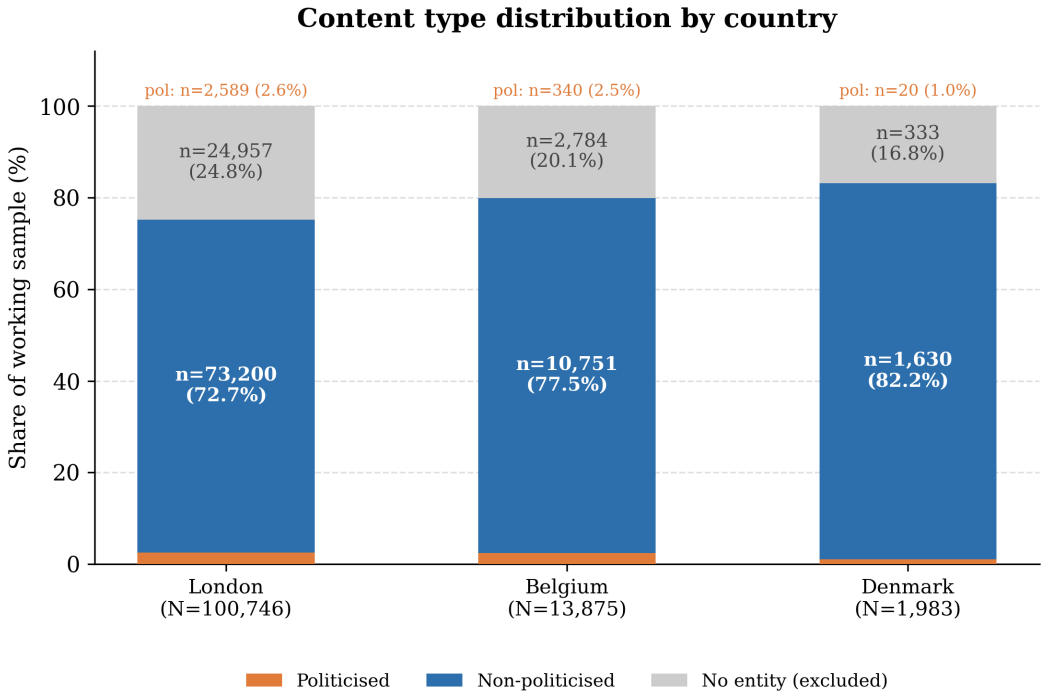

Figure 5: Content type distribution

km. Belgium's median is 1.65 km and Denmark's 4.70 km. Part of this difference reflects sparser graffiti coverage in those countries, but it also reflects a difference in analytical scope. London is treated as a city-level analysis, while Belgium and Denmark are analyzed at the country-level. Tweets in the latter two are spread across much larger geographic areas, so longer median distances are expected by design rather than being a signal about spatial graffiti availability.

In all three cases, the distribution is heavily right-skewed, driven by a minority of tweets located far from any recorded graffiti piece. A log transformation, `log(1 + nearest_graffiti_km)`, was applied throughout the regression analyses to correct for this, producing an approximately symmetric distribution. Log transform is commonly applied to prevent distant outliers from exerting undue leverage on the estimates (Gelman & Hill, 2006).

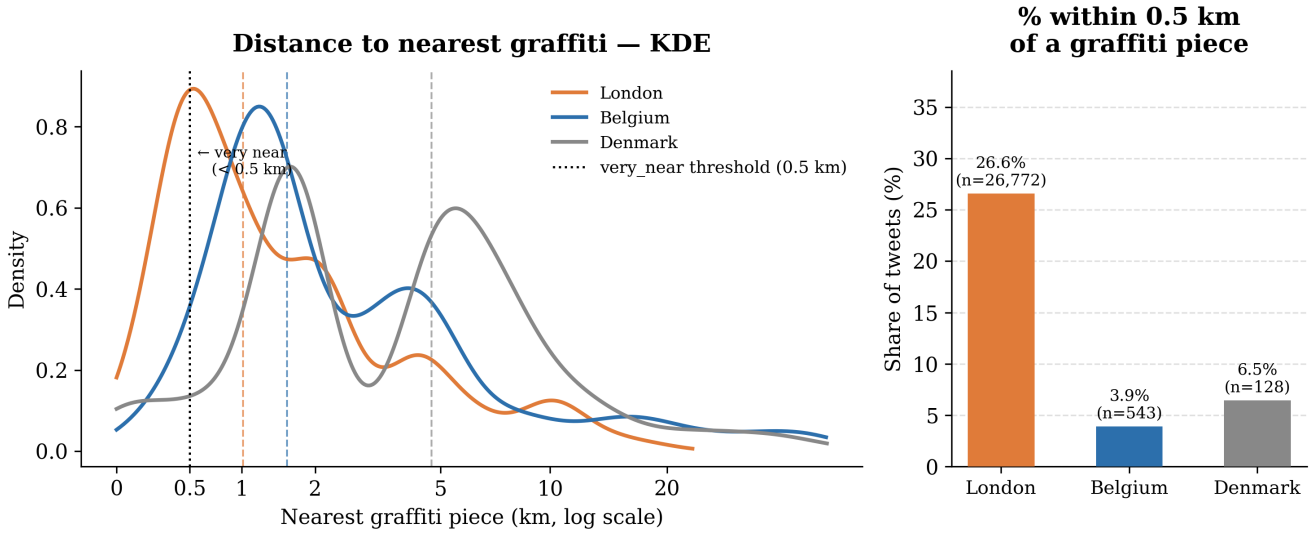

Figure 6: Proximity distribution

Within London, 25.7% of tweets (n=29,902) fall within 0.5 km of a graffiti piece. The inverse is equally interesting with 74.3% of London tweets having no recorded graffiti piece within 500 m, and 55.3% having none within 1 km. Graffiti commonly appears in clusters, the mean piece count within 500 m is 4.55, but the median is zero and the maximum is 265, reflecting a small number of very dense areas pulling the average upward. Graffiti coverage is concentrated in a handful of zones rather than distributed across the city, and the proximity measures used in H2 reflect that natural structure.

5.1.4 Sentiment Distribution

Negative sentiment is the dominant label across all three countries, though the balance varies by each country. London is the most negative at 40.5%, with neutral (33.4%) and positive (26.1%) labels making up the remainder. Belgium is close behind at 39.2% negative, while Denmark is somewhat less negative and more evenly split at 32.5% negative and 39.6% neutral. Positive sentiment is consistently low across all three datasets. The overall skew toward negative sentiment is consistent with prior work on pandemic discourse, where crisis framing, loss, and institutional criticism tend to dominate (Boon-Itt & Skunkan, 2020; Cinelli et al., 2020). Specifically COVID-era discourse was understandably negatively valenced, but additionally the United Kingdom sentiment tends to skew very negative (Fancourt et al., 2020). In contrast, Denmark consistently displays a very temperate political discourse, even in crisis times (Nielsen & Lindvall, 2021).

The more interesting pattern emerges when sentiment is broken down by content type, previewed here and tested formally in H1. Politicized tweets are substantially more negative than non-politicized content in both London and Belgium. The near-identical politicized means across two independently collected national corpora (−0.624 in both) is a notable cross-national consistency, suggesting the emotional reaction and blame attribution discourse is not London-specific. Denmark follows the same direction but the difference is not statistically significant, almost certainly a consequence of the NER limitations noted

in 5.1.3. With only 20 reliably extracted politicized tweets, the test is severely underpowered rather than indicative of a genuinely null effect.

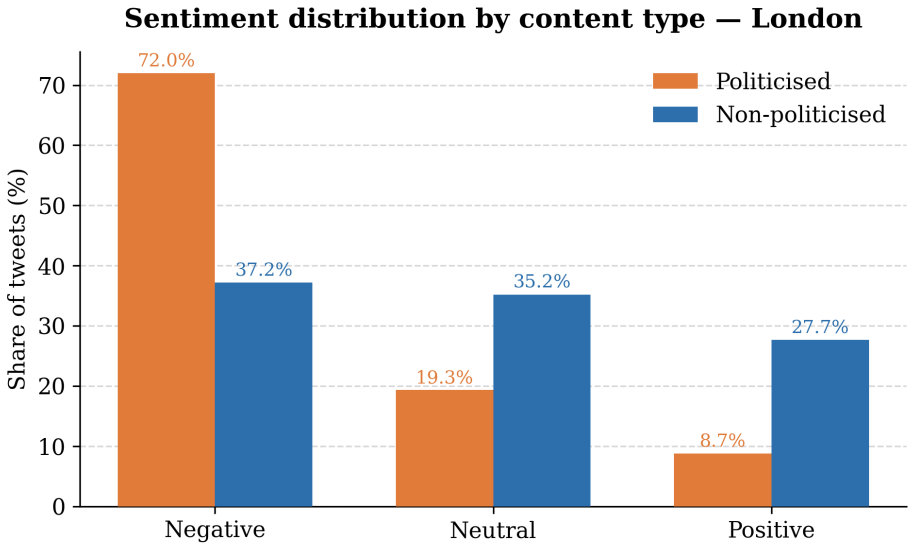

Figure 7: Sentiment distribution

5.1.5 Heavy-user Distribution

A subset of London users posted at high volume during the TBCOV 2020 collection period. 37,845 tweets, or 32.5% of the London corpus, come from accounts that contributed more than 50 tweets each, with the most prolific single user accounting for 1,918 tweets.

Heavy users do meaningfully shape the aggregate results. In the full London sample the logistic regression yields a non-significant OR of 0.972 ($p = 0.715$) for political content. Once heavy users are excluded, the OR rises to 1.202 ($p = 0.011$). The direction changes from null to significantly positive. The mixed effects model shows a comparable pattern, with a significant negative within-user effect (OR = 0.785, $p < 0.001$) in the full sample that attenuates to null once heavy users are removed (OR = 0.912, $p = 0.265$). Heavy users therefore contribute both volume and a distinctive quoting pattern that pulls against the political amplification signal in the full sample. All H1 and H2 regressions are run twice, with and without heavy users, to expose this dependence directly.

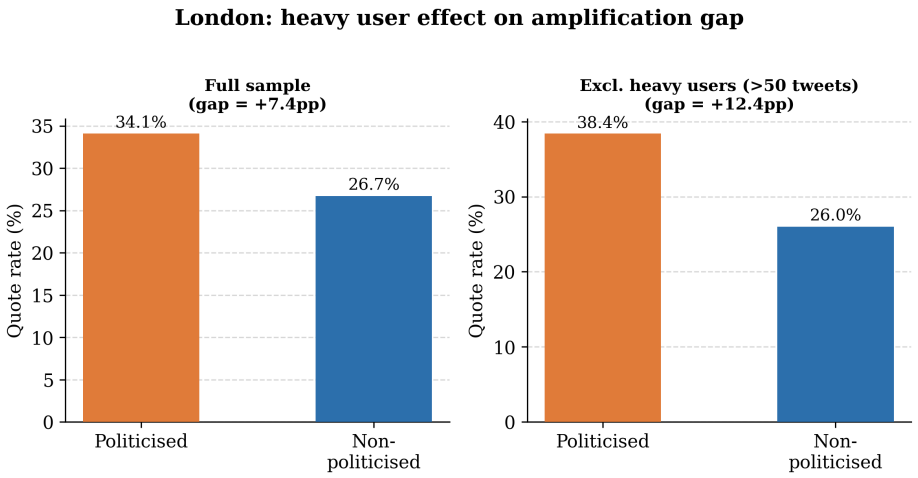

Figure 8: Heavy User distribution

5.2 Hypothesis Testing

5.2.1 Hypothesis 1: Politicized content will exhibit higher amplification rates across national contexts.

At the descriptive level, London shows a modest positive gap. Politicized tweets were quote retweeted at 34.1% compared to 26.7% for non-politicized content. Belgium and Denmark both reverse the direction, with non-politicized content quoting at higher rates in both cases. The pooled chi-square is statistically significant ($\chi^2 = 31.53$, $p = 1.96 \times 10^{-8}$), but the pattern is not consistent across countries and the London gap is itself modest once the non-politicized baseline includes all entity types rather than COVID content alone. Cross-national consistency does not hold at the descriptive level.

Country	Pol. quote rate	Non-pol. quote rate	Gap	n (pol / non-pol)
London	34.1%	26.7%	+7.4pp	2,589 / 73,200
Belgium	22.1%	35.0%	−12.9pp	340 / 10,751
Denmark	15.0%	31.3%	−16.3pp	20 / 1,630

Table 2: Descriptive quote rates by country and content type.

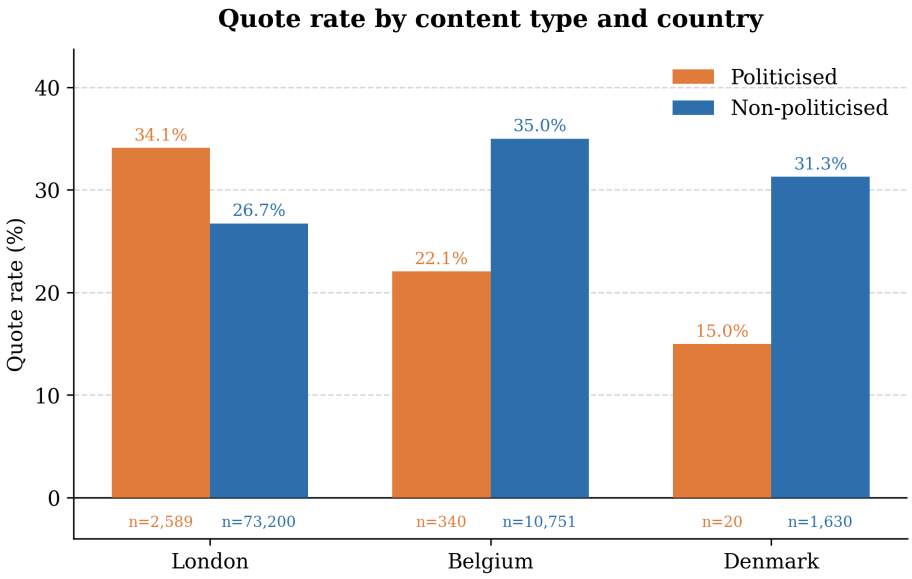

Figure 9: Politicized and non-politicized quote rates by country.

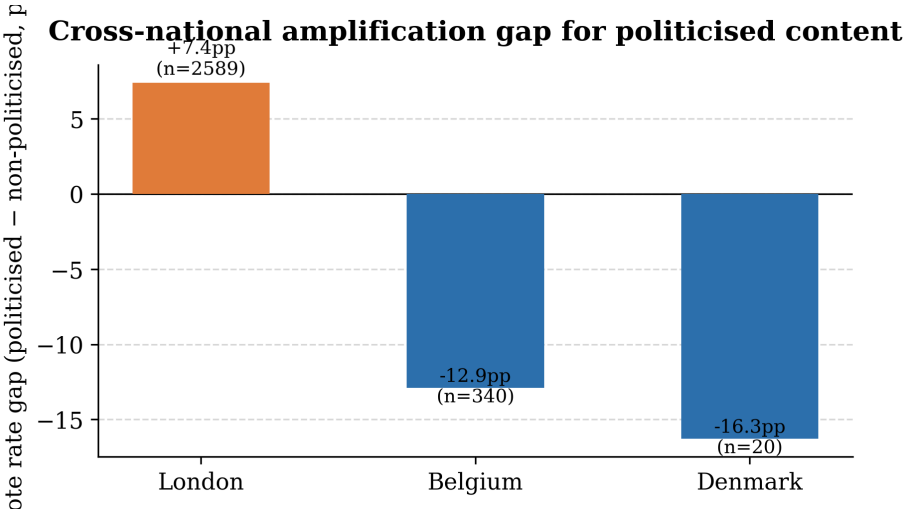

Figure 10: Amplification gap (pp) across countries and model specifications, from descriptive to mixed effects.

Logistic Regression A binary logistic regression controlling for content type, sentiment, and proximity (with standard errors clustered by user to account for within-user correlation) narrows the London gap further. In the full London sample the effect is not significant (OR = 0.972, $p = 0.715$). Excluding heavy users, a positive effect emerges (OR = 1.202, $p = 0.011$), suggesting that high-volume accounts suppress the signal in the full sample. Belgium and Denmark both produce negative odds ratios, though Belgium's estimate is imprecise given the small number of politicized tweets, and Denmark's significant negative result (OR = 0.289, $p = 0.004$) is based on only 20 politicized tweets and cannot be treated as reliable. The pooled model with country fixed effects produces no significant effect (OR = 0.905, $p = 0.318$).

Sample	n	Users	OR	95% CI	p	AUC
London (full)	100,746	20,565	0.972	[0.835, 1.132]	0.715	0.640
London (excl. heavy)	69,171	20,268	1.202	[1.043, 1.386]	0.011*	0.634
Belgium	13,875	2,774	0.559	[0.224, 1.392]	0.211	0.634
Denmark	1,983	624	0.289	[0.124, 0.673]	0.004**	0.606
Pooled (country FE)	116,604	23,957	0.905	[0.743, 1.101]	0.318	0.630

* $p < 0.05$; ** $p < 0.01$. SEs clustered by `user_id`.

Table 3: H1 logistic regression results. `is_quote` $\sim$ `has_pol_figure` + `has_COVID` + `has_ORG` + `sentiment_label` + `log_km`.

Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC-ROC). The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 − specificity) across all classification thresholds, with the AUC summarizing overall discriminative ability (Fawcett, 2006). A value of 0.5 indicates no discrimination beyond chance, while 1.0 indicates perfect separation. Across all H1 model specifications, AUC values were consistently modest, ranging from 0.606 (Denmark) to 0.640 (London full), reflecting the well-documented difficulty of predicting individual-level amplification behavior from content and proximity features alone. The near-identical curves across countries and samples indicate that model fit is stable and not driven by any single corpus. The modest AUC is consistent with the thesis's central finding: that user-level heterogeneity (ICC = 0.629) accounts for the majority of variance in quote behavior, leaving content type and spatial proximity as weak predictors at the individual tweet level.

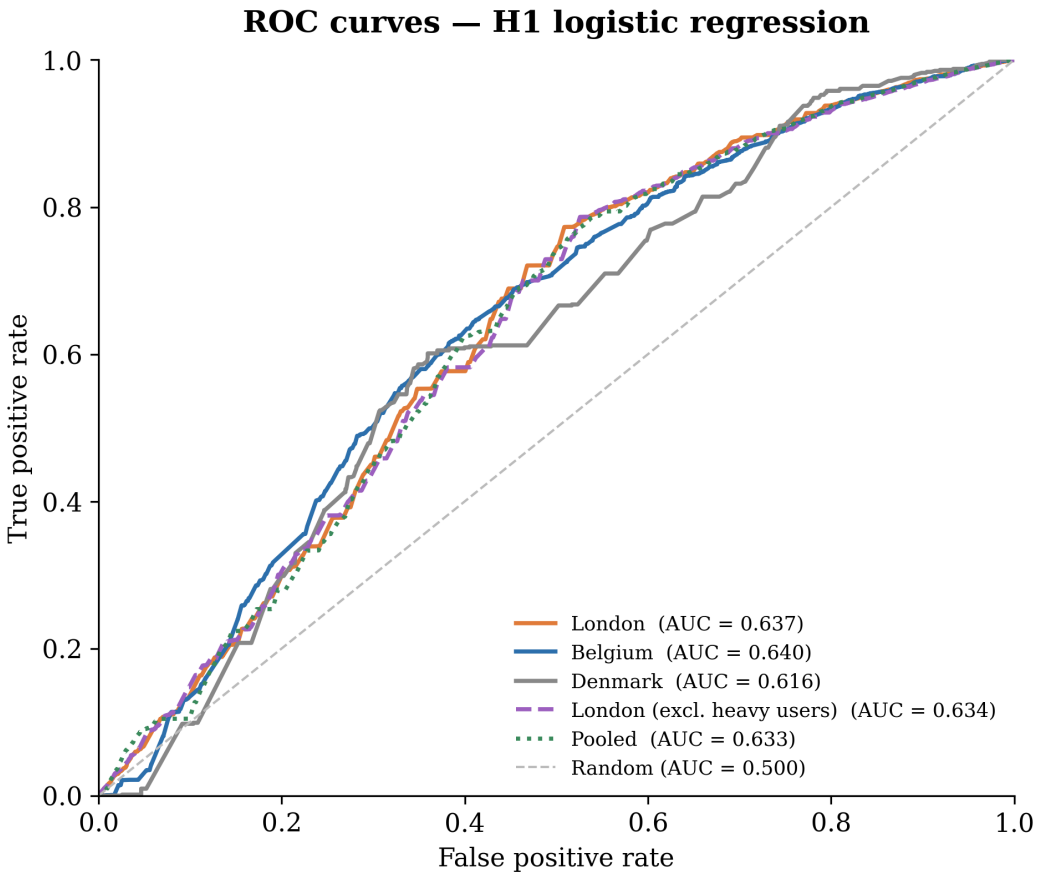

Figure 11: ROC curves for the London full and excl. heavy users models (AUC = 0.640 / 0.634).

Mixed Effects Logistic The logistic models cluster standard errors by user, which corrects for the fact that tweets from the same account are not independent observations. But clustering does not account for the underlying quoting tendencies of different users. Some users are structurally highly-quoted and their tweets circulate regardless of the content of the tweet. Clustering adjusts the uncertainty estimates around coefficients but leaves this between-user variance in the residuals. If high-quoted users also happen to post more political content, then the estimated effect of political content will absorb some of that user-level signal. The result looks like a political content effect but is partly an effect of the user posting.

A mixed effects model addresses this directly. Adding a random intercept per user, `(1|user_id)`, estimates a baseline quoting propensity for each account and partitions variance into what is stable across a user's tweets and what varies within them. The fixed effect for `has_pol_figure` then captures whether their political tweets are quoted more or less than their non-political tweets. This is a within-user comparison rather than a between-user one. The intraclass correlation coefficient (ICC) quantifies how much of the total variance in quoting sits at the user level versus the tweet level. An ICC of 0 would mean all variance is between tweets, content, timing, wording, and user identity is irrelevant. An ICC of 1 would mean quoting is entirely determined by who you are, with no variation across a given user's tweets. The London ICC is 0.629.

Sample	ICC
London (full)	0.629
London (excl. heavy users)	0.622

Table 4: Intraclass correlation coefficients from H1 mixed effects models.

Nearly two-thirds of the variance in whether a tweet gets quoted has nothing to do with the tweet. It is explained by characteristics of the account that posted it. This is a substantive finding of the study. Amplification on Twitter is far more a property of users than of content. This also introduces a direct implication for H1. If high-quoted users disproportionately post political content, then a model will attribute political framing to what is actually a consequence of user posting behavior. The mixed effects results confirm this. Once user random intercepts are included, the `has_pol_figure` coefficient reverses and becomes strongly significant.

Sample	n	Users	OR	95% CI	p
London (full)	100,746	20,565	0.785	[0.695, 0.886]	$<$0.001***
London (excl. heavy)	69,171	20,268	0.912	[0.775, 1.073]	0.265

*** $p < 0.001$. Random intercept: `(1|user_id)`.

Table 5: H1 mixed effects logistic regression results. `is_quote` $\sim$ `has_pol_figure` + controls + `(1|user_id)`.

Within the same user's output, political tweets are quoted less than non-political tweets. The contrast with the clustered SE model is stark. The direction of the mixed effects result is consistent across both samples and robust to heavy user exclusion. The positive signal in the clustered SE model when heavy users are excluded reflects the fact that removing high-volume accounts reduces some user-level confounding, but does not eliminate it.

Sentiment The strong sentiment divergence between content types, politicized mean -0.624 vs. non-politicized -0.098 (Mann-Whitney $p = 4.70\times10^{-211}$), highlights the question: is outrage what drives political content to circulate? A Baron-Kenny mediation analysis (Baron & Kenny, 1986) with sentiment as the mediator finds no evidence for this. Sentiment strongly predicts content type (path a: $\beta = -0.524$, $p < 0.001$) but does not predict quoting (path b: $\beta = -0.011$, $p = 0.627$), with the indirect effect negligible and non-significant (Sobel $z = 0.49$, $p = 0.627$). Controlling for sentiment changes the political content coefficient by less than 0.01.

A supplementary interaction between sentiment and content type reveals a more nuanced pattern. Political content quotes at lower rates than non-political content at every sentiment level. However, neutral political content performs especially poorly at 19.2% quoted, well below both negatively valenced political content (36.5%) and neutral non-political content (29.8%). Emotional charge does not explain

the gap between political and non-political content amplification. That gap is driven by user composition, not content valence.

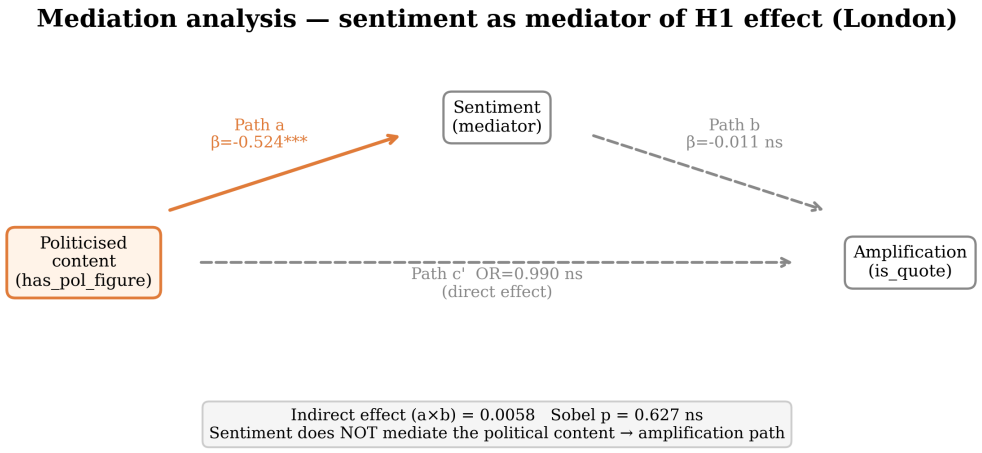

Figure 12: Baron-Kenny mediation path diagram: political attribution → sentiment → amplification.

H1 verdict *H1 is not supported as stated.* The descriptive gap in London (+7.4pp) and the positive logistic effect when heavy users are excluded (OR = 1.202, $p = 0.011$) show that politicized content amplifies more at the population level. Cross-national consistency does not hold with Belgium and Denmark both reversing the direction. Within-user mixed effects models reverse the London direction (OR = 0.785, $p < 0.001$), establishing that the population-level gap is carried by prolific, high-quoted users who disproportionately post political content. The mechanism driving this amplification is user composition, not content type.

5.2.2 Hypothesis 2: The higher rates will be attenuated by the proximity to graffiti.

Before testing H2 it is necessary to establish what `nearest_graffiti_km` is actually measuring. The London graffiti database contains 835 registered pieces, but these are not distributed evenly across the city. Only 54 of approximately 1,500 possible 1km^2 grid cells contain at least one piece. Of those 835 pieces, 37% (n=311) are concentrated in two adjacent cells in Shoreditch, with further clusters in South Bank/Waterloo (n=88) and Brixton/Stockwell (n=53). Seventy percent of all pieces fall within 5km of Charing Cross. Nearest-neighbour analysis confirms that pieces cluster tightly along specific walls and corridors, median distance between pieces is 3 metres, with 95.3% within 100 metres of another piece.

The practical implication is that `nearest_graffiti_km` is not measuring proximity to graffiti in general. It is largely measuring proximity to Shoreditch, South Bank, or Brixton. A tweet classified as `very_near` is predominantly a tweet posted from one of these three zones. As discussed previously, the database also skews toward professional mural-scale work rather than small graffiti tags. Meaning, it is not a comprehensive census of street-level political expression. This shapes how the proximity results should be interpreted throughout the analysis.

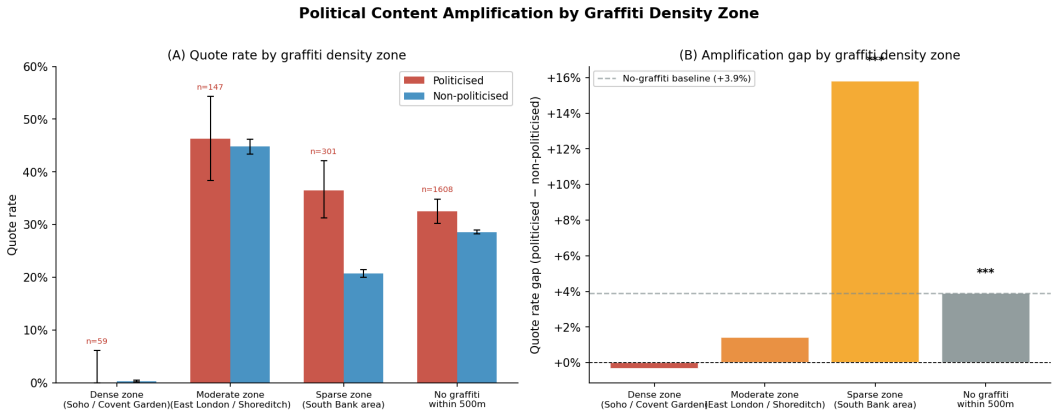

Figure 13: Quote rate partitioned by graffiti density zone. And amplification gap for each graffiti density zone.

Binary Regression Model The individual-level model regresses `is_quote` on content type, proximity, their interaction, and sentiment, with standard errors clustered by user. The main proximity effect is negative, tweets further from graffiti are less likely to be quoted ($\beta = -0.138$, $p = 0.130$), and the baseline quote rate is lower in near-graffiti areas ($\beta = -0.435$, $p = 0.004$). COVID-entity content is strongly suppressed overall ($\beta = -1.365$, $p < 0.001$).

Term	β	p	Interpretation
`has_pol_figure`	−0.049	0.544	Political content not significantly different overall
`has_COVID`	−1.365	<0.001***	COVID-entity content strongly suppressed
`log_km`	−0.138	0.130	Distance gradient, imprecise
`very_near`	−0.435	0.004**	Baseline quote rate lower near graffiti
`pol_figure × very_near`	+0.362	0.085	Direction positive, not significant
`COVID × very_near`	+0.035	0.825	No proximity effect for COVID content

** $p < 0.01$; *** $p < 0.001$. SEs clustered by `user_id`. AUC = 0.658.

Table 6: H2 binary logistic regression. `is_quote` ∼ `has_pol_figure` + `has_COVID` + `has_ORG` + `log_km` + `very_near` + `pol_figure`×`very_near` + `COVID`×`very_near` + `sentiment_label`.

The main effects tell a consistent story before the interaction is considered. Political content is not significantly different overall ($\beta = -0.049$), COVID-entity content is strongly suppressed ($\beta = -1.365$), and distance from graffiti is associated with lower quote rates ($\beta = -0.138$). The baseline quote rate is also lower in graffiti-proximate areas ($\beta = -0.435$), meaning tweets posted close to graffiti are generally less likely to be quoted than tweets posted further away.

The key result is the interaction term. `pol_figure × very_near` is positive but does not reach conventional significance ($\beta = +0.362$, $p = 0.085$, OR = 1.436). The direction is the opposite of H2, which predicted attenuation. Instead near graffiti, political content amplifies relatively **more** than non-political content. The equivalent interaction for COVID-entity content is null ($\beta = +0.035$, $p = 0.825$),

meaning the directional tendency is selective and applies to blame-attribution discourse but not to health or policy content.

A density extension replacing the binary `very_near` with continuous graffiti counts at 500m and 1km does not replicate the interaction once user clustering is applied, suggesting the binary threshold is the more robust operationalization and that the effect does not extend beyond 500m.

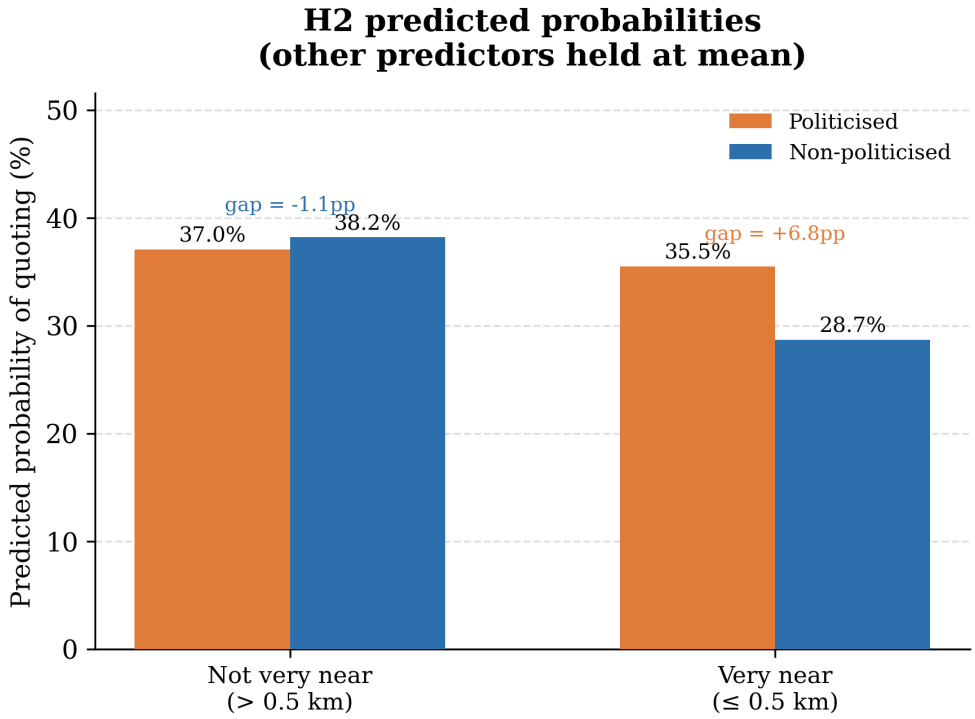

Figure 14: Predicted probability of `is_quote` by `very_near` and content type. Shows the positive but non-significant interaction direction.

Area-Level Ecological Regression The individual-level result carries a structural limitation previously mentioned but now explored. 80% of tweet locations are bounding box centroids, so proximity is an area-level measure rather than an individual one. The ecological regression addresses this directly by aggregating to 1km^2 grid cells and asking whether graffiti-dense cells show a larger amplification gap (Gelman et al., 2001). The analysis is restricted to cells with at least five political tweets (n=29 cells; n=38 at a lower threshold of three tweets), with weighted least squares and HC3 robust standard errors.

The design aggregates all London tweets to grid cells, then restricts to cells with at least five political tweets to ensure stable within-cell rate estimates. This yields 29 cells at the primary threshold and 38 at a lower threshold of three tweets. The outcome for each cell is the difference between the political quote rate and the non-political quote rate. The predictor is log-transformed mean graffiti density within 500m. Weighted least squares (WLS) is used with cell-level political tweet counts as weights, so cells with more political tweets contribute more to the estimate. HC3 robust standard errors correct for heteroskedasticity given the small sample size.

Outcome	β	95% CI	p	R^2
Amplification gap	+0.036	[+0.007, +0.064]	0.014*	0.215
Political quote rate	−0.049	[−0.159, +0.061]	0.380	0.231
Non-political quote rate	−0.085	[−0.189, +0.019]	0.110	0.499

* $p < 0.05$. WLS, weights = n_pol per cell, HC3 SEs. n = 29 cells ($\geq$5 pol tweets).

Table 7: H2 area-level ecological regression. Predictor: log1p(mean `graffiti_count_500m`) per 1km^2 cell.

The amplification gap widens with graffiti density (β = +0.036, p = 0.014), consistent across both cell-count thresholds. Neither the political quote rate nor the non-political quote rate individually reaches significance, but they move in the same direction, both declining in graffiti-dense areas. The non-political rate declines more steeply ($\beta = -0.085$) than the political rate ($\beta = -0.049$). The widening gap is therefore not driven by political content amplifying more in graffiti-dense areas, but by non-political content amplifying less. Health and informational discourse is relatively suppressed in politically expressive urban zones, and political content holds its position while that suppression occurs.

This is a different mechanism from what H2 anticipated, but it is consistent with the theoretical framing. Graffiti-proximate areas are spaces where political discourse is the ambient register, and in those spaces, content that does not fit that register circulates less freely. The ecological result is exploratory given the small number of cells, and Spearman rank correlations are non-significant at both thresholds (p = 0.096–0.176), indicating the WLS estimate is sensitive to the weighting scheme. It is reported as corroborating evidence for the individual-level direction rather than a standalone finding.

Mixed Effects The individual-level and area-level results both point in the same direction indicating a wider political amplification gap near graffiti. However, neither accounts for user-level quoting tendencies. As established above, the ICC of 0.629 means that nearly two-thirds of variance in quoting is between users rather than between tweets. The same confounding process that undermined H1 applies here: if politically engaged, high-quoted users are disproportionately located in graffiti-dense areas, then the proximity interaction in the logistic model may be recovering a spatial piece of that user population rather than a proximity mechanism.

A mixed effects model with a random intercept per user tests this directly. The within-user comparison again asks whether the same user's political tweets are quoted at different rates depending on whether they were posted near graffiti.

Specification	Key term	OR	95% CI	p
Binary `very_near` (full)	`pol_x_verynear`	1.226	[0.923, 1.627]	0.159
Binary `very_near` (excl. heavy)	`pol_x_verynear`	1.123	[0.768, 1.642]	0.549
Density 500m	`pol_x_density500`	1.111	[0.930, 1.327]	0.245

Random intercept: `(1|user_id)`. ICC = 0.622.

Table 8: H2 mixed effects logistic regression results. Key coefficient is the `pol_figure` $\times$ `proximity` interaction term in each specification.

Across all three specifications the proximity interaction loses significance once user random intercepts are absorbed. The odds ratios remain above 1.0, consistent with the individual-level direction, but none is distinguishable from noise. What appears as a spatial or content effect reflects the quoting tendencies of the users who inhabit those spaces, not a proximity mechanism operating on tweets directly.

H2 verdict *H2 is not supported as hypothesized.* The baseline clustered SE model shows a positive but non-significant tendency (OR = 1.436, $p = 0.085$), consistently in the opposite direction to the predicted attenuation. The area-level ecological regression shows a significant widening of the amplification gap in graffiti-dense cells ($p = 0.014$), driven by the relative suppression of non-political content rather than enhancement of political content. Both effects disappear in mixed effects models once user heterogeneity is modelled (OR = 1.226, $p = 0.159$). The aggregate pattern reflects user selection into politically expressive urban areas rather than a direct proximity mechanism.

5.3 Follow-up Testing and Exploration

5.3.1 Clustering Analysis

To complement the regression results, a KMeans clustering analysis examines whether tweet-level patterns of entity type, graffiti proximity, and amplification form interpretable clusters that give more context to the H1 and H2 findings. The feature matrix weights amplification and proximity more heavily than entity composition, reflecting their theoretical centrality. UMAP reduces the feature space to ten dimensions before KMeans assigns tweets to twelve clusters.

Binary Quote Rate Structure The most immediate and interpretively consequential feature of the clustering output is its near-binary structure. Every cluster either quotes at 0–5% or at 97–100%, there is no cluster with an intermediate quote rate. Cluster 6 (`PERSON` entities) quotes at 100.0%; Cluster 10 (Boris, Cummings, Johnson) at 97.2%; Cluster 9 (mixed temporal terms) at 99.4%; Cluster 1 (NHS-dominant) at 100.0%. All remaining clusters quote at 0–4.9%. See the full breakdown in Appendix C.

This binary structure is an unsupervised, data-driven restatement of the ICC = 0.629 finding from above. The UMAP embedding, which used amplification as a heavily-weighted input, has partitioned

the corpus along the same axis that the mixed effects model identified as dominant. Entity type and proximity are secondary separators that distinguish clusters within the amplified and organic groups from each other. The threshold from 0% to near-100% reflects user identity, not marginal content or spatial differences.

Political Clusters Near Graffiti Two clusters contain the most politically explicit content in the corpus and are simultaneously the closest to registered graffiti of any amplified cluster. Cluster 10 (dominant terms: Boris, Cummings, Johnson; n = 3,348; 97.2% quoted; mean distance 0.74 km) and Cluster 6 (dominant terms: Boris, with mixed `PERSON` co-occurrences; n = 3,032; 100.0% quoted; mean distance 0.74 km) share identical mean proximity and near-identical amplification rates. Together they account for 6,380 tweets (approximately 3% of the London corpus) at the intersection of political blame attribution, near-graffiti location, and high quote amplification.

This spatial pattern is consistent with the H2 reversal observed in the regression models. Political blame content is over-represented in graffiti-proximate areas relative to non-political content. But the clustering does not provide evidence for a causal mechanism. Clusters confirm the composition of these areas but does not reveal the process generating it. Clusters 6 and 10 are high-quote clusters near graffiti because high-influence political Twitter users inhabit Shoreditch and Brixton. Whether proximity is a cause, correlate, or consequence of their political engagement is not resolvable from this data.

Cluster 9 (mixed temporal entity terms; n = 4,822; 99.4% quoted; mean distance 0.86 km) represents a third near-graffiti amplified cluster, though its entity composition is less interpretable. Its temporal entity dominance suggests these are quote tweets where the primary named content is a date or time reference rather than a person or institution, consistent with high amplification among users who quote heavily regardless of content.

The NHS Cluster Contrast The theoretically richest contrast is between two NHS-dominant clusters with markedly different amplification and proximity profiles. Cluster 0 (dominant terms: NHS, Brexit, EU; n = 4,781; 0.0% quoted; mean distance 1.09 km) represents organic NHS expression originating relatively near graffiti-dense areas. Cluster 1 (dominant terms: NHS, and `ORG` entities; n = 7,200; 100.0% quoted; mean distance 2.08 km) represents amplified NHS discourse originating substantially further from graffiti.

NHS expression that circulates happens, on average, nearly a kilometer further from graffiti than NHS expression that does not circulate. This is in the same direction as the area-level ecological regression, which found that non-political content quote rates decline more steeply near graffiti than political content quote rates. The clustering recovers this same structure from the unsupervised joint distribution of entities, proximity, and amplification. Revealing NHS content near graffiti is original expression while NHS content that spreads originates from further away.

Cluster 7 adds a third dimension to the NHS picture. Its dominant terms span both political and NHS framing. This hybrid cluster quotes at 0.0%, with a mean proximity of 1.57 km. The failure of

mixed political-NHS content to amplify, despite its proximity to the high-amplification political clusters, is consistent with the regression interaction results. What drives amplification in political content is the specificity of blame attribution, not the presence of NHS themes as co-occurring framing.

Figure 15: Example of the street art existing in London in support of the NHS (COVID-19 Street Art Archive, 2020)

What the Clustering Adds The unsupervised analysis does not introduce new findings but provides a data-driven triangulation of the regression results from three directions. First, the binary quote rate structure confirms that amplification is a near-discrete user-level property, not a continuum driven by marginal content or spatial differences. Second, the concentration of amplified political clusters at 0.74 km confirms the composition of graffiti-proximate areas identified as the user selection confound. Third, the NHS contrast independently reproduces the ecological regression finding where non-political discourse is relatively suppressed in politically expressive urban spaces.

5.3.2 Proximity to Graffiti and Sentiment

An exploratory analysis was prompted by a consistent descriptive pattern of tweets near graffiti being less negative across all content types. It is not a test of H1 or H2 and is reported as a secondary observation.

Descriptive Pattern Mean sentiment varies by distance from graffiti. Tweets in the 0–0.5 km band (mean = −0.098, 35.3% negative) and the 5+ km band (mean = −0.032, 35.2% negative) are both less negative than the intermediate 0.5–2 km bands (mean = −0.131 to −0.140, 39.9–40.9% negative). A simple closer to graffiti means less negative sentiment interpretation is not supported by the raw data. The relationship is instead U-shaped.

Regression Results An OLS regression with standard errors clustered by `user_id` finds that the full-sample effect of `very_near` on sentiment is marginal (overall $\beta = 0.043$, $p = 0.086$). Splitting by content type reveals where the signal is concentrated:

Content type	n	β (`very_near`)	95% CI	p
Political (`has_pol_figure`)	2,589	0.088	[0.018, 0.157]	0.014*
COVID-entity (`has_COVID`)	14,506	0.057	[0.024, 0.091]	<0.001***
General (neither)	58,694	0.038	[−0.022, 0.099]	0.216

* $p < 0.05$; *** $p < 0.001$. SEs clustered by `user_id`.

Table 9: Effect of graffiti proximity (`very_near` < 0.5 km) on sentiment by content type.

Near-graffiti areas are associated with less negative sentiment specifically where the discourse is substantive — i.e political or COVID-entity content — but not in undirected general content. A linear mixed effects model with a random intercept by `user_id` (ICC = 0.245) confirms the effect holds within users. `very_near` $\beta = 0.034$, 95% CI [0.013, 0.056], $p = 0.0016$. `log_km` is also independently significant ($\beta = 0.021$, $p = 0.0011$), meaning the same person's tweets are measurably less negative when originating from a graffiti-proximate area.

A `very_near` × `has_pol_figure` interaction is non-significant ($\beta = 0.057$, $p = 0.102$), indicating the proximity-sentiment association is not specific to political content.

Interpretation These results suggest that graffiti-proximate areas are associated with a distinct affective character in online COVID discourse. This is marginally less negative in substantive content and this pattern holds within individual users. The effect is small ($\beta = 0.034$ on a -1 to $+1$ scale) but consistent. The same identification limits established in the causal inference section apply, there is no way to establish that proximity to graffiti produces less negative expression. The U-shaped distance pattern (5+ km also less negative) warrants caution: outer London tweets differ demographically and in content profile from inner London.

5.3.3 ZINB on Rehydrated GPS Subset

The main analysis uses a binary amplification proxy (`is_quote`) because the TBCOV source data does not include engagement counts like number of likes, comments, retweets, or quote retweets. A subset of London GPS-located tweets was rehydrated via the X API v2 to recover actual rehydrated text and `quote_count` values. Then a Zero-Inflated Negative Binomial (ZINB) regression was attempted on count outcomes.

The rehydrated sample contained $n = 1{,}990$ tweets, with `quote_count`: mean = 0.36, max = 114, and 94.4% zeros (n=1,878). Political figure tweets in this subsample: $n \approx 24$ (1.2%). The ZINB model for `quote_count` failed to converge with standard errors returned as NaN and coefficients implausibly large.

The plain Negative Binomial also failed to converge. Three structural limitations preclude inference: $n_{\text{pol}} \approx 24$ is far below the minimum needed for stable count model estimation; `very_near` is constant at 1 in the rehydrated sample (the rehydrator collected only tweets within 0.5km of a graffiti piece, so there is no comparison group); and 94.4% zeros at n = 1,990 produces extreme zero-inflation relative to sample size. The analysis is reported for methodological transparency. The binary `is_quote` proxy in the main 100,746-tweet sample remains the sole reliable amplification measure available.

5.4 Causal Inference

5.4.1 What Cannot Be Claimed Causally

The analyses in §5.2–5.5 are observational throughout. No causal identification strategy is available in this dataset. There is no easy randomization of proximity to graffiti, natural experiment, or instrumental variable that plausibly affects graffiti density without also affecting amplification through other pathways. Graffiti density and political content amplification are jointly determined by the same latent variable: the political character of a neighborhood and its residents. Shoreditch simultaneously concentrates graffiti and politically engaged, high-Twitter-influence users. The association between proximity and amplification cannot be separated from the prior selection of those users into that space.

The ICC of 0.629 makes this concrete. Nearly two-thirds of variance in whether a tweet is quoted is accounted for by between-user differences, not by the content of the tweet or where it is posted from. This is the central substantive finding. Any model that does not account for who posted the tweet is largely recovering a portrait of the user population rather than estimating a content or spatial effect. The H2 proximity interaction attenuates from OR = 1.436 (p = 0.085) to OR = 1.226 (p = 0.159) once user random intercepts are included. The parsimonious interpretation is that the clustered standard error result was absorbing user heterogeneity rather than detecting a proximity mechanism.

5.4.2 The User Selection Confound

The pattern across H1 and H2 is consistent with a single confounding process: high-influence users disproportionately post political content and disproportionately inhabit politically expressive urban areas. Under this account, the H1 descriptive gap reflects that prolific political posters are themselves high-quoted users, not that political framing generates amplification. Political content amplifying more near graffiti in simpler models reflects that graffiti-dense zones are precisely where high-influence political users are most concentrated. Once mixed effects models condition on individual quoting tendencies, both patterns disappear.

This is not equivalent to saying proximity to graffiti has no effect. It is saying the current data cannot distinguish a genuine proximity effect from spatial sorting of users. Ruling out user selection would require either longitudinal data tracking how the same user's amplification changes as they move relative to graffiti-dense areas, or an instrument for graffiti density that is uncorrelated with the political

composition of the local user population.

5.4.3 Causal Identification

The observational design cannot rule out the possibility that the associations observed in 5.2 and 5.3 are driven by unobserved confounders. The primary causal inference strategy employed in this study is Coarsened Exact Matching, reported in full in 5.4.4. CEM addresses the threat of covariate imbalance between politicized and non-politicized tweets by pre-processing the sample to ensure matched groups are comparable on observable characteristics before estimation. The conditional logit additionally eliminates all stable user characteristics through within-user fixed effects. Together they represent the strongest identification available within a cross-sectional observational design.

A quasi-experimental approach was also evaluated. In June 2020, Hackney Council removed a Black Lives Matter mural from Forest Road, Hackney, a dated, administratively motivated event within the study window that partially satisfies the exogeneity requirements for a difference-in-differences design. However, the design was not feasible to execute. Within 1km of Forest Road, the dataset contains too few tweets for any treatment comparison. The root constraint is geocoding, with many London locations being centroid box bound, street-level treatment assignment is not possible. Additionally, administrative records that might have established the precise removal date and other cleanup initiatives were requested from Hackney Council but are no longer available following a ransomware cyberattack in October 2020 (Information Commissioner's Office, 2024). One further concern is temporal confounding, the removal coincided with the peak of the Black Lives Matter protests, meaning national political activation was simultaneously increasing amplification across London boroughs.

The Hackney case illustrates the general constraint on causal identification in this data. A suitable exogenous event existed within the study window, but the geocoding limitations of TBCOV prevented its exploitation. Causal inference in this study therefore rests on the matching and within-user designs in 5.4.4 rather than on quasi-experimental variation in the built environment.

5.4.4 Directly Testing the User Composition Confound

The mixed effects models in 5.2 and 5.3 absorb user heterogeneity via random intercepts but do not rule out covariate imbalance as an alternative explanation: politicized and non-politicized tweets may differ on observable characteristics that drive amplification independently of content type. Three tests address this directly.

Tweet-Level Coarsened Exact Matching Coarsened Exact Matching pre-processes the sample by coarsening and exactly matching tweets on observed covariates before re-running the logistic models (Iacus et al., 2012). Any remaining difference in quote rates between matched groups cannot be attributed to imbalance on those covariates. For H1, tweets were matched on sentiment category, heavy-user status, and `has_ORG` (n = 75,789 matched tweets). For H2, matching additionally included `has_pol_figure`

and has_COVID, with the continuous pol_x_logkm interaction as the post-match key coefficient (n = 75,784).

The H1 signal survives matching, clustered SE OR = 1.277 (p = 0.001), and is stronger than the unmatched estimate, ruling out covariate imbalance as its source. Once user random intercepts are added, the effect collapses to OR = 1.040 (p = 0.537, ns; ICC = 0.700). The population-level advantage is real, but disappears within users. The H2 proximity interaction is null under CEM across both estimators.

User-Level Coarsened Exact Matching Tweet-level matching balances observable tweet characteristics but does not balance on the user characteristics that the selection confound identifies as the true driver of the H2 pattern. A second CEM matched users directly on baseline quoting propensity, tweet volume, share of political tweets, and typical proximity to graffiti. Of 6,139 treated and 11,693 control users, CEM retained 1,082 treated and 8,337 control users (n = 26,750 tweets from 9,419 matched users). After balancing on these user-level characteristics, the H2 proximity interaction remains non-significant across both specifications. The ICC in the user-matched mixed effects model rises to 0.950. Meaning when users are matched on their quoting propensity, nearly all remaining variance in amplification is between-user, confirming that user composition is what the unmatched models were recovering.

Conditional Logit The strongest test uses clogit() with strata(user_id), estimating entirely from within-user variation. All stable user characteristics are completely absorbed. Only users with at least one quoted and one non-quoted tweet contribute to estimation (3,215 of 20,565 users; 36,850 tweets). The H2 proximity interaction is OR = 1.068 (p = 0.490) in the full sample and OR = 0.963 (p = 0.796) excluding heavy users. The point estimate falls below 1.0, meaning the same user's political tweets posted near graffiti are quoted marginally *less* than their political tweets posted further away.

Hypothesis	Specification	n	OR	p	ICC
H1 — Treatment: has_pol_figure; matched on sentiment, heavy-user, has_ORG					
H1	Tweet-CEM, clustered SE	75,789	1.277	0.001**	—
H1	Tweet-CEM, mixed effects	75,789	1.040	0.537	0.700
H2 — Key term: pol_x_logkm / pol_x_verynear					
H2	Tweet-CEM, clustered SE	75,784	1.045	0.702	—
H2	Tweet-CEM, mixed effects	75,784	0.991	0.926	—
H2	User-CEM, clustered SE	26,750	1.379	0.683	—
H2	User-CEM, mixed effects	26,750	0.250	0.286	0.950
H2	Conditional logit (full)	36,850	1.068	0.490	—
H2	Conditional logit (excl. heavy)	—	0.963	0.796	—

** $p < 0.01$; all others ns. User-CEM matched on quoting propensity, volume, pol share, and proximity. Conditional logit: strata(user_id).

Table 10: CEM and conditional logit robustness results.

All three tests converge on the same conclusion. The H1 population-level signal is robust to covariate matching but not to user heterogeneity modeling. The H2 proximity interaction is null regardless of specification, matching strategy, or whether the comparison is made between tweets, between users, or within the same user's own tweet history. User identity is the dominant predictor of amplification.

5.4.5 Spatial Autocorrelation and Distance Decay

Two further analyses test whether the data are consistent with a genuine spatial mechanism.

Distance decay. A causal proximity mechanism predicts a monotonic dose-response: the amplification gap should be largest near graffiti and decay smoothly with distance. Quote rates were computed by content type within five distance bands:

Band	n_{pol}	Pol. rate	Non-pol. rate	Gap	p
0–0.5 km	507	35.1%	22.0%	+13.1pp	<0.001
0.5–1 km	469	27.9%	30.7%	−2.8pp	0.212 ns
1–2 km	416	36.1%	31.0%	+5.1pp	0.031*
2–5 km	518	33.2%	26.8%	+6.4pp	0.001**
5+ km	205	34.2%	24.2%	+9.9pp	0.001**

Table 11: Political and non-political quote rates by distance band. Chi-square tests within each band.

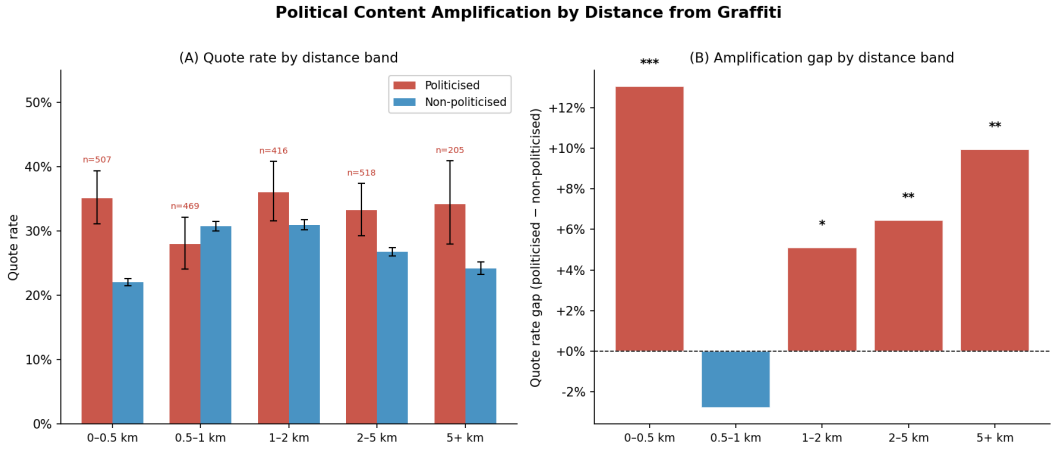

Figure 16: Political and non-political quote rates by distance band, with amplification gap annotated. Non-monotonic pattern is inconsistent with a smooth spatial exposure mechanism.

The pattern is non-monotonic. The near zone (0–0.5 km) shows the largest raw gap (+13.1pp), but this reverses at 0.5–1 km (−2.8pp, ns), and the outer zones (2–5 km, 5+) show stable gaps comparable in size to the inner London average. A genuine spatial exposure mechanism would predict the gap to be largest near graffiti and smallest at distance. The reversal at 0.5–1 km and the large gap at 5+ km are both inconsistent with a smooth proximity mechanism and consistent with the user composition interpretation.

Spatial autocorrelation. A spatial exposure mechanism further requires that graffiti density and political amplification co-occur in the same parts of the city. Global Moran's I was computed at the 1km^2 grid cell level ($n = 87$ cells with ≥ 1 political tweet; KNN weights $k = 5$; 999 Monte Carlo permutations):

Variable	Moran's I	p
Political quote rate	0.005	0.374 ns
Non-political quote rate	0.094	0.063 ns
Amplification gap	−0.004	0.419 ns
Log graffiti density	**0.235**	**0.003****

Table 12: Global Moran's I for key variables at the 1km^2 grid cell level.

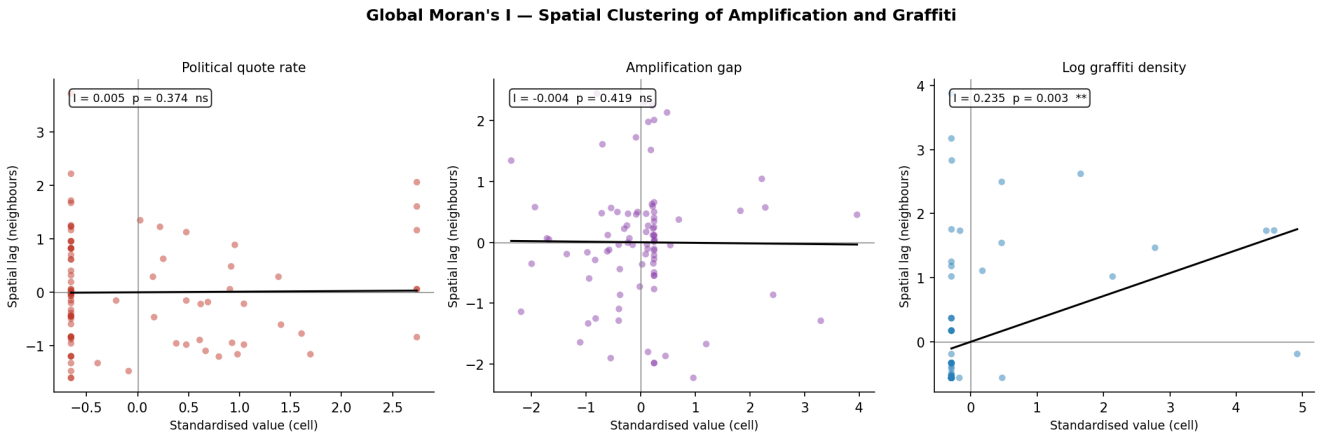

Figure 17: Moran scatter plot of political quote rate vs spatial lag, with Moran's I and p-value annotated.

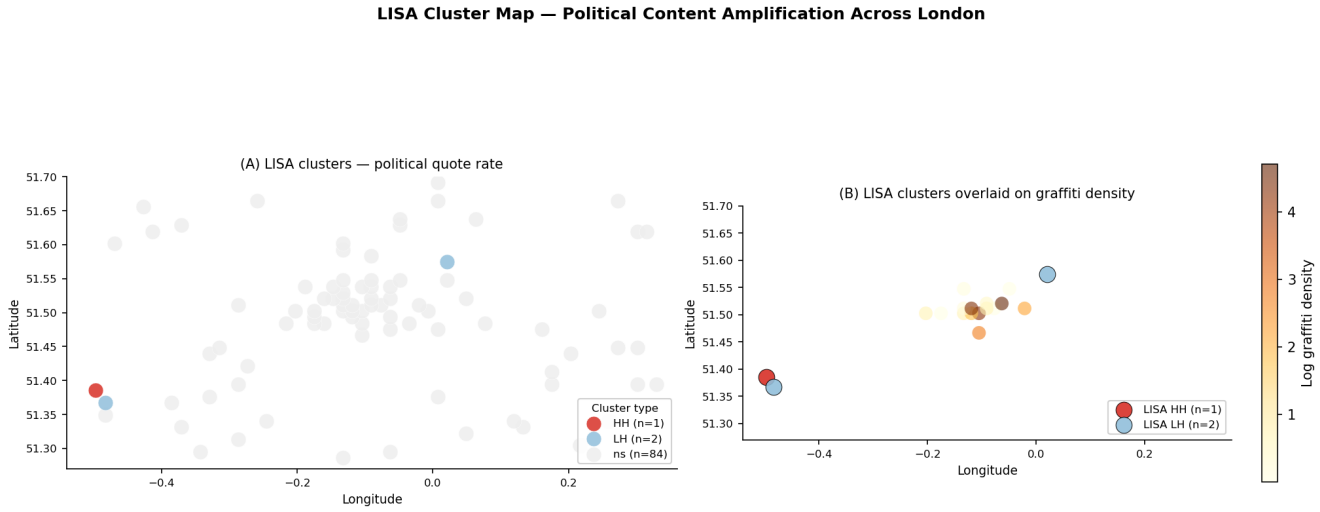

Figure 18: LISA cluster map at 1km^2 grid cell level: HH/HL/LH/LL cluster types for political quote rate × graffiti density.

Graffiti density clusters spatially, which is expected given the concentration in Shoreditch and Brixton. Political amplification, however, does not (I = 0.005, $p = 0.374$). Bivariate Moran's I (political quote rate $\times$ log graffiti density): I = -0.099, $p = 0.010$ is negative and significant. High-amplification grid cells tend to neighbor low-graffiti cells and visa versa. LISA analysis finds only 1 HH cluster and 2 LH clusters (84 of 87 cells non-significant). Graffiti concentration and political amplification are spatially segregated at the city level, directly contradicting a city-scale spatial mechanism for H2.

5.5 Robustness

5.5.1 Excluding Heavy Users

Heavy users (>50 tweets per user) account for a large amount of tweets in the London corpus. After their exclusion, the working sample is $n = 69{,}171$ across 20,268 unique users. The H1 descriptive gap widens from +7.4pp to +11.0pp (politicized 37.1% vs. non-politicized 26.1%). The H1 logistic model yields a significant positive effect (OR = 1.202, $p = 0.011$) in contrast to the null in the full sample. The H1 mixed effects model direction remains negative but attenuates to null (OR = 0.912, $p = 0.265$), indicating that the within-user reversal is driven by heavy users who post political content. For H2, the proximity interaction remains non-significant in both the clustered SE (OR = 1.304, $p = 0.093$) and mixed effects (OR = 1.123, $p = 0.549$) specifications. Heavy user exclusion does not alter the H2 conclusion.

5.5.2 GPS-Only Subset

The GPS-only subset ($n = 22{,}510$) contains only 104 tweets in the key interaction cell (`has_pol_figure` = 1, `very_near` = 1), causing complete separation in standard logistic regression. A penalized alternative (L1, $\alpha = 0.01$) yields a near-zero and unreliable coefficient. Full output is reported in Appendix A. H2 findings throughout rest on the full dataset and are framed as area-level claims accordingly.

5.5.3 Alternative Proximity Thresholds

Three proximity specifications were compared: the binary 500m threshold (`very_near`), continuous graffiti density within 500m, and continuous graffiti density within 1,000m. The binary threshold yields OR = 1.436 ($p = 0.085$) as reported in section 5.2. The density specifications do not replicate once user clustering is applied: `log_density_500m` $\times$ `pol_figure` $\beta = +0.116$, OR = 1.123, $p = 0.311$ (ns); `log_density_1000m` $\times$ `pol_figure` $\beta = +0.045$, OR = 1.046, $p = 0.531$ (ns). A dose-response interpretation is not supported with clustered standard errors. The binary threshold is the more robust operationalization.

5.5.4 Cross-National Comparison

The H1 and H2 specifications were run on Belgium with identical model formulae. Belgium H1 logistic: OR = 0.559, $p = 0.211$ (ns; wide CIs [0.224, 1.392] due to $n_{\text{pol}} = 311$). Belgium H2 proximity interaction

reverses direction relative to London: political content amplifies *less* near graffiti in Belgium. The non-replication is not treated as falsifying the London findings; substantive reasons for divergence include Belgian political discourse being dominated by international rather than domestic figures, the smaller and nationally dispersed sample, and different urban geography. The London findings are treated as London-specific rather than cross-nationally generalizable.

5.6 Central Finding

H1 is supported at the population level in London, political content amplifies more than non-political content descriptively (+7.4pp) and under covariate matching (CEM clustered SE: OR = 1.277, $p = 0.001$). The within-user mixed effects model reverses this (OR = 0.785, $p < 0.001$). The effect is carried by who produces political content, not by political blame attribution or framing. H2 is null within users, the conditional logit returns OR = 1.068 ($p = 0.490$), and the baseline clustered SE result disappears once user heterogeneity is modeled. ICC = 0.629–0.700 across specifications meaning user identity is the dominant predictor of amplification.

6 Discussion

6.1 Summary

This thesis explored the connection between proximity to street art and its ability to predict the nature and amplification of politicized content on Twitter. Results show politically expressive urban zones in London are where politically active, high-amplification Twitter users concentrate. The spatial co-occurrence of graffiti geography and political online discourse is a robust empirical finding. The analysis reveals that this correspondence is driven by who inhabits those spaces rather than what the spaces do to individual behavior in real time. Three parts summarize the empirical picture.

First, H1 is supported in London at the population level. In line with established literature on political virality, political content amplifies more than non-political content, most clearly when heavy users are excluded (OR = 1.202, p = 0.011) and under covariate matching (CEM clustered SE: OR = 1.277, p = 0.001). Cross-national consistency, however, does not hold; Belgium reverses the direction. The London signal is real and in the expected direction. The progression of model specifications establishes the mechanism. The descriptive gap of 7.4 percentage points reflects who produces political content rather than what political framing itself does to a tweet's probability of being shared. Once individual quoting tendencies are modeled by mixed effects (OR = 0.785, p < 0.001), political content amplifies less within the same user's output, and the effect collapses to null under CEM mixed effects (OR = 1.040, p = 0.537). H1 holds at the population level, and the within-user analysis reveals that the politically engaged user carries the effect rather than the content type alone.

Second, H2 is not supported in the predicted direction. Proximity to graffiti is associated with a

widening of the amplification gap, not an attenuation as hypothesized, but this finding is substantive. Political content amplifies relatively more in graffiti-proximate areas (OR = 1.436, p = 0.085 in the baseline model; $\beta = +0.036$, p = 0.014 in the area-level ecological regression). Once user heterogeneity is modeled the interaction attenuates to null (conditional logit OR = 1.068, p = 0.490). The reversal and its disappearance together document the spatial sorting pattern that is this thesis's core empirical contribution: high-influence, politically active Twitter users cluster in precisely the zones where political expression is most visibly inscribed in the built urban environment.

Third, and most importantly, the intraclass correlation of 0.629 is the thesis's central empirical finding. Nearly two-thirds of the variance in whether a tweet is quoted is explained by the identity of the user who posted it. Content type and spatial location are secondary confounds. The contribution of this study is to demonstrate that standard logistic regression, as used across much of the political virality literature, overstates content and spatial effects because it leaves this dominant source of variance unmodeled. The hypotheses were not supported, but the why is substantive and theoretically interesting, the physical–digital relationship this study set out to test is real in the aggregate but entirely compositional in origin.

6.2 Contributions

This thesis makes contributions at three levels: empirical, methodological, and in terms of data infrastructure. They are summarized here before the detailed interpretation of results.

Empirical Contributions The primary empirical contribution is the city-scale measurement of the spatial correspondence between street art geography and political amplification on Twitter. Using 100,746 geolocated tweets from the TBCOV dataset and a database of 835 registered graffiti pieces, this thesis demonstrates that politically engaged, high-amplification Twitter users disproportionately post from the same London zones where political expression is most visibly spoken on the walls through graffiti and street art. This co-occurrence is robust across descriptive, ecological, and cluster-based analyses, and has not previously been measured at this spatial resolution or scale. While the within-user null result establishes that the physical environment does not causally alter individual sharing behavior in real time, the aggregate spatial pattern is substantive. The physical and digital geographies of political engagement in London overlap in ways that reflect the underlying geography of politically engaged communities, and this overlap represents a novel empirical observation about how political communities organize and present across physical and digital ecosystems.

Methodological Contributions The second contribution is methodological, though it should be qualified by the dataset's limitations. The intraclass correlation of 0.629 illustrates concretely what is at stake when stable user-level differences are left unmodeled in Twitter amplification research. The progression from descriptive to clustered SE to mixed effects to conditional logit provides a example of the inferential consequences at each step, and the conditional logit design is well-suited to research questions that ask

whether a content or spatial feature changes amplification behavior for the same user. Whether the ICC of 0.629 is representative of Twitter amplification data more broadly is an open empirical question. The geocoded GPS-enabled subsample of TBCOV is not a random sample of Twitter users, and users who share location data are likely a self-selected, publicly engaged population whose quoting behavior may be more consistently patterned than average. The claim this study can support is that user heterogeneity was a first-order concern in this specific dataset, large enough to reverse the direction of content-type effects, and that the modeling approach adopted here is the appropriate response when repeated observations from the same users are available.

Data Contributions The third contribution is the construction of a linked spatial dataset joining geolocated Twitter records from TBCOV to registered graffiti piece coordinates from graffiti-database.com. No existing public dataset links these two sources, or any two Twitter and graffiti datasets, at spatial resolution or at this scale. The pipeline developed for this purpose covering geocoding of both GPS-located and bounding-box-centered tweets, proximity computation to graffiti pieces, and integration of named entity recognition outputs is extensible to other cities and periods covered by TBCOV, and to other databases of political expression in the built environment. The London, Belgium, and Denmark extracts produced by this pipeline represent a resource that future work in the area of physical–digital political communication can build upon directly.

6.3 Interpreting H1: User Composition, Not Content Effect

6.3.1 The Descriptive-to-Mixed-Effects Gap

The progression of H1 results across model specifications is itself informative, and the gap between the descriptive estimate and the mixed effects estimate is the thesis's most important methodological lesson. The descriptive quote rate gap of +7.4 percentage points in London represents a raw comparison between two groups of tweets that differ not only in content type but in who produces them, how often they tweet, and how widely followed they are.

Controlling for content covariates and proximity with clustered standard errors collapses the political content coefficient in the full London sample. This is the first reduction: compositional differences between politicized and non-politicized tweets explain much of the apparent gap. The second reduction comes from the mixed effects model, which adds random intercepts by user identity. Here the coefficient not only loses significance but changes sign OR = 0.785 ($p < 0.001$). Within the same user's output, political content is a less shareable type of post than non-political output. The descriptive gap was driven by who produces political content, not by anything political attribution itself does to a tweet's probability of being shared. Each step in this progression isolates a distinct source of variance. The progression as a whole demonstrates that models which ignore user identity are seeing a symptom of the user population, not estimating a content effect.

6.3.2 Why Political Content Amplifies Less Within Users

The mixed effects reversal is not a null result. It is a substantive finding that requires explanation. Two interpretations are consistent with it.

The first, is what might be called the identified expression hypothesis. Quote-tweeting a political blame post is a named, permanent, networked act of political endorsement. It appears on the quoting user's profile, is visible to all their followers, and constitutes a legible statement of political alignment with specific real-world consequences professional associations, follower composition, the possibility of being criticized or piled on. Quoting a neutral health information tweet carries none of these social costs: it is informational, low-stakes, and signals care rather than alignment. If the threshold for engagement is raised by the social cost of the act, political content should face higher barriers to circulation than non-political content from the same account which is precisely what the within-user mixed effects result shows. This interpretation connects to the structural distinction between anonymous physical dissent and identified online expression. Graffiti carries no reputational cost to its viewer, a quote tweet identifying the user as a political actor does. The within-user suppression of political amplification is consistent with this asymmetry in social exposure, even if the mechanism cannot be directly tested with the available data.

The second interpretation concerns selective audience activation. High-quoted users in the dataset tend to have large, politically diverse follower networks. When they post non-political content such as health information or institutional updates, their network can engage without political cost. However, when they post political blame content, only the politically aligned fraction of their network responds. Politically neutral or disengaged followers scroll past. This means that per-tweet, political content activates a narrower share of the available audience, even if the absolute volume of political quotes is high because the user posts prolifically. The superspreader heavy users driving the descriptive H1 gap, is consistent with this account. Prolific political content producers have established overtly political identities, so their political content is the expected output and their networks are pre-sorted accordingly. For ordinary users, the marginal social cost of political attribution is higher, because their networks are less pre-committed.

Neither interpretation can be definitively tested with the current data. Both require information about follower network composition and the social costs of identified political speech that the TBCOV metadata does not contain. But they are grounded in established theory, and the within-user reversal is the result that connects most directly to the thesis's own theoretical background, specifically the asymmetry between anonymous physical expression and identified online expression developed in Chapter 2.

6.3.3 The Heavy User Effect

The CEM analysis excluding heavy users adds important nuance to the H1 picture. When users who posted more than 50 tweets in the London corpus are removed, the logistic regression with clustered standard errors recovers a significant positive effect (OR = 1.202, p = 0.011), while the within-user mixed

effects estimate attenuates to null (OR = 0.912, p = 0.265). The CEM on the full sample, which balances tweet-level covariates before re-estimating the model, shows the same pattern. Clustered SE OR = 1.277 (p = 0.001), mixed effects OR = 1.040 (p = 0.537). The political content advantage survives covariate matching but not user heterogeneity modeling, in both the heavy-user-excluded and CEM specifications. This is the most favorable reading available for H1. Among users with comparable tweet profiles, political blame attribution is associated with higher amplification at the population level, but the same user does not quote their political content at a higher rate than their non-political content. The within-user reversal is concentrated in heavy users, politically prolific accounts whose followers selectively engage with only a fraction of their output.

What this implies for the mechanism is important. Heavy users are not distorting the data in a statistical sense, they represent real behavior, but they are the channel through which political content achieves aggregate visibility. They post political content at high volume, they attract high quote rates by virtue of their established status, and their volume inflates population-level quote rate comparisons between content types. The effect is not simply a content property that political attribution reliably generates. It is, in large part, a structural feature of who produces political blame content at scale on geolocated Twitter.

6.3.4 Cross-National Scope: Why Belgium Reverses

Belgium's consistent negative direction across all H1 specifications (OR = 0.559) is not adequately explained by its smaller political tweet count (n = 340) or by statistical imprecision. The confidence intervals are wide ([0.224, 1.392]), but the direction is consistent and holds across multiple model specifications. Something about the character of political discourse in non-UK contexts generates a different amplification dynamic.

The most substantive explanation is the nature of political attribution in each corpus. London COVID Twitter discourse was dominated by domestic political figures whose decisions were directly consequential for UK residents. The blame-attribution dynamic was personal and immediate, stating these were the named individuals responsible for lockdown rules, PPE procurement failures, and Downing Street gatherings. Belgian COVID Twitter, by contrast, was dominated by international figures referenced from a position of greater geographic and political distance. Tweets about Macron's France or Trump's America from Belgian Twitter users function differently as social acts, they are more likely to be informational reference than outrage-driven blame. If the sharing mechanism for political content is emotional activation through proximate blame attribution, the mechanism should operate more strongly for domestic political figures than for international ones. Belgium's discourse structure may simply lack the conditions under which outrage-driven sharing of political attribution content is activated.

The Belgium reversal is therefore not a replication failure. It specifies the boundary conditions under which the outrage-attribution amplification mechanism operates. The blamed actor must be politically consequential to the sharer's own life and context. Domestic blame attribution directed at leaders whose decisions shaped your lockdown, your job, your family's access to healthcare, activates outrage and the

impulse to share and comment on. Commentary on international figures observed at a remove activates something closer to political curiosity or informational engagement. The H1 mechanism is not about political content in general; it is about the specific emotional architecture of proximate, personalized blame attribution. Belgium makes this visible precisely because its COVID Twitter discourse lacked the domestic structure that London's possessed.

6.4 Interpreting H2: What the Reversal Means

6.4.1 The Predicted vs Observed Direction

H2 predicted that political content's amplification advantage would be attenuated in areas near graffiti. The empirical pattern shows the reverse: in simpler specifications, political content amplifies more near graffiti. The direction is consistent across multiple specifications before user heterogeneity is controlled, and it is a theoretically meaningful reversal rather than noise.

Two readings of this pattern are worth distinguishing. The first, interpretation is that the reversal is an artifact of user composition. High-quoted, politically active Twitter users concentrate in particular zones, Shoreditch, Brixton, South Bank, that happen to be zones with the densest registered graffiti. The apparent amplification advantage near graffiti is not the physical environment influencing how content spreads online. It is a portrait of who inhabits those spaces. Politically engaged, high-influence users whose tweet output is amplified because of their social capital, not because of their proximity to murals. The conditional logit result is the definitive test when controlling for all stable user characteristics, posting near graffiti does not predict higher political amplification. The reversal in simpler models is compositional throughout.

The second, exploratory reading connects the area-level ecological finding to a broader urban affect account. The ecological regression shows that both quote rates decline near graffiti, but non-political content declines more steeply ($\beta = -0.085$) than political content ($\beta = -0.049$). The widening gap is driven not by the amplification of political discourse in graffiti-dense areas but by the relative suppression of health and informational discourse. If the affective character of politically expressive urban spaces orients online expression away from health anxiety and toward political register (even if it does not actually amplify political content) the ecological pattern would follow. This reading is consistent with the exploratory sentiment finding reported in 5.8, the same areas show marginally less negative sentiment in substantive COVID discourse. Graffiti-dense zones may have a distinct collective affective character, one that shapes what users choose to post rather than how widely their posts travel. This is speculative and explicitly flagged as such, it is not a finding of H2 but a direction for subsequent theory.

6.4.2 Inoculation Theory Implications

The original theoretical framing of H2 drew on inoculation theory, proposing that ambient exposure to political counter-narratives in the built environment would reduce reliance on online amplification

to circulate those narratives. If the message has already been encountered on the walls in the form of graffiti, the need to re-share the message online is lower. The observed pattern runs in the opposite direction, though the empirical result is best understood as compositional rather than causal. What does this mean for the theoretical framework that motivated H2?

First, the theory is not falsified. Inoculation is a longitudinal process: prior exposure to a weakened form of a persuasive claim changes an individual's susceptibility when the full version of that claim is subsequently encountered. A point-in-time proximity measure — the distance from a tweet's coordinates to the nearest registered graffiti piece — is not a measure of prior exposure, cumulative residence time, or depth of engagement with political content in the built environment. The study measured a spatial correlate of the conditions under which inoculation *might* occur. The conditional logit confirms that tweeting near graffiti on a specific occasion does not alter amplification behavior in real time (OR = 1.068, $p = 0.490$). But this rules out an instantaneous spatial trigger, not the longitudinal mechanism the theory actually describes.

Second, and more substantively, the ICC of 0.629 is not in conflict with an inoculation account — it may be its empirical signature, read at the wrong level of analysis. Inoculation theory predicts that accumulated exposure to political counter-narratives produces durable changes in how individuals engage politically: greater critical processing, more robust epistemic engagement, a stable disposition toward active participation in political discourse. These are individual-level characteristics that persist across time and context. The ICC captures exactly this: the stable between-user variance that explains nearly two-thirds of amplification behavior, independently of any specific tweet's content or location. If inoculation has occurred, its output is a durable political engagement disposition — precisely what the between-user component of a mixed effects model is estimating. The mechanism, if real, looks like user-level ICC, not spatial proximity correlation.

This re-framing converts the spatial null into a specification error in the research design, not a theoretical failure. What the study was measuring, proximity to graffiti at tweet time, is orthogonal to what the theory requires, which is accumulated prior exposure across an individual's political biography. The high-amplification users who cluster in graffiti-proximate areas may carry the accumulated traces of prior exposure to politically expressive environments — through years of residence, daily commute, political socialization in activist communities whose physical presence is inscribed in these streets. Cross-sectional proximity data cannot recover that trajectory; it can only observe the endpoint. The spatial sorting this study documents — politically engaged users concentrating in politically expressive urban spaces — is consistent with a community-formation version of the inoculation account: environments that continuously model and normalise political counter-expression attract and retain individuals whose accumulated histories have produced exactly the stable engagement dispositions the ICC captures. The graffiti marks the community; the community is not produced by the graffiti.

6.4.3 The User Selection Confound as the Substantive Finding

The pattern across H1 and H2 is explained coherently by a single mechanism: high-influence, politically active users cluster in politically expressive urban areas. Shoreditch concentrates both registered graffiti and Twitter power users with large quote-generating follower networks. The correlation between graffiti proximity and political amplification is real in the descriptive data but is not a causal relationship between the built environment and online behavior. It is a consequence of who lives, works, and tweets in those areas.

This re-framing shifts the thesis's contribution from a null result to an empirical and methodological demonstration. The null at the within-user level (ICC = 0.629; conditional logit OR = 1.068, $p = 0.490$) establishes that simpler models of physical–digital correspondence overstate spatial effects because they leave the dominant source of variance unmodeled. This is not a negative finding about street art and online politics. It is a finding about the structure of Twitter amplification and a demonstration that cross-sectional spatial methods cannot distinguish environmental effects from the selection processes that produce spatial patterns in the first place.

The question the thesis opens, why high-influence political Twitter users cluster in graffiti-dense urban areas, connects directly to the neighborhood effects literature reviewed in Chapter 2. The residential sorting documented by Bishop (2008) at the national scale, and the local exposure mechanisms established by Enos (2016), together describe a world in which politically engaged individuals select into environments that match and reinforce their prior political identities. Whether the spatial co-occurrence of graffiti and political Twitter activity reflects prior residential sorting, the cultural attractiveness of politically transgressive neighborhoods to politically engaged individuals, or a feedback process in which visible political expression reinforces the concentration of political community, is not resolvable with this data. Each mechanism implies a different causal direction and a different intervention point. That the current study can document the pattern but not explain it is an honest characterization of what cross-sectional observational data can establish and a precise specification of what a stronger design would need to achieve.

6.5 The Sentiment–Amplification Paradox

6.5.1 Political Content Is Most Negative and Most Amplified

The descriptive pattern in Chapter 5 presents an apparent contradiction. Political content is the most negatively valenced category in the corpus and yet it amplifies more than non-political content at the aggregate level. COVID-entity content is moderately negative (mean −0.098) and amplifies at the lowest rates. NHS content is near-neutral and amplifies at intermediate rates. A simple negativity-drives-virality account predicts the opposite ordering.

The data do not directly resolve this paradox, because the TBCOV sentiment variable captures valence only and cannot distinguish between qualitatively different forms of negative emotional content.

The paradox is most parsimoniously explained by the outrage/anxiety distinction developed in Chapter 2. Political content's negativity externalities blame and drives sharing. COVID content's negativity is anxiety. It is internalized and does not circulate. The paradox dissolves once the undifferentiated negativity variable is replaced by that distinction.

This is a theoretically motivated reading of the pattern rather than a demonstrated empirical finding. The data are consistent with it but cannot confirm it. Directly testing the outrage/anxiety distinction would require an emotion-specific sentiment classifier capable of separating anger from fear within the negative category, which the pre-computed TBCOV sentiment labels do not support. This represents the primary open empirical question that future work building on this thesis should pursue.

6.5.2 Why Neutral Political Content Circulates Least

The strongest indirect evidence for the outrage/anxiety interpretation within the available data is the sentiment breakdown within the political content group. Neutral political content quotes at only 19.2%, a rate lower than negative political content (36.5%), lower than positive political content (37.9%), and lower than neutral non-political content (29.8%). If it were the negative valence of political content that drove its amplification, neutral political content should quote at an intermediate rate between negative and non-political content. Instead it quotes least of all the content groups in the corpus.

This pattern is consistent with a charge-based account of political amplification. What enables political content to circulate is emotional activation, not valence direction. Negative political content and positive political content are both emotionally charged, one is outrage, the other may be celebration, or solidarity. Neutral political content, factual attribution and descriptive reporting, lacks this emotional charge. It circulates at rates comparable to informational non-political content, not political content. The content is political by the operationalization criteria (it names political figures) but does not function as politically activating discourse in the amplification sense.

This is the closest the data come to distinguishing emotional activation from valence direction, but it remains indirect. A definitive test would require an emotion-specific classifier applied to the re-hydrated tweet content, a resource unfortunately not available in the current dataset. The neutral political content finding rules out a simple negativity-drives-virality account and is the within-data evidence most consistent with the outrage hypothesis, but it cannot confirm that the relevant emotion in negative political tweets is specifically outrage rather than another externalizing affect.

6.6 Implications

6.6.1 Implications for Practice

The findings of this thesis carry practical implications across three groups: social media practitioners and organizations seeking to communicate effectively online, platform regulators concerned with the governance of political amplification, and urban policymakers and developers making decisions about

politically expressive public space.

For Social Media Practitioners and Organizations The ICC of 0.629 is the most directly actionable finding in this thesis. The practical implication of this finding is that amplification is primarily a function of who tweets something, not what content is tweeted. For organizations seeking to achieve reach on social media, this means that optimizing message content, adjusting framing, emotional tone, or narrative structure, is a second-order strategy at best. The first-order strategy is identifying and cultivating relationships with users whose output is structurally amplified, regardless of content. This re-frames influencer strategy away from audience size as the primary selection criterion and toward the demonstrated amplification rate of a user's own output, a distinction the ICC makes quantitatively precise.

The sentiment finding adds a further practical dimension. Political blame attribution content achieves higher aggregate amplification than informational, health-oriented, or neutral content. For organizations communicating during crises, this creates a tension. Messages that diagnose institutional failure and attribute responsibility to named actors spread further than messages that provide information or appeal to shared concern. Public health communicators in particular face the implication that anxiety-framed health information is structurally disadvantaged relative to blame-framed political content in the same information environment. Awareness of this asymmetry should shape both the framing of public health messaging and the realistic expectations organizations hold about organic reach for informational content.

For Platform Regulators and Content Governance The user heterogeneity finding has direct implications for how platforms and regulators think about interventions targeting political amplification. The dominant paradigm in platform governance treats amplification as a content-level problem: harmful or misleading content spreads because of properties of the message — its emotional valence, its novelty, its confirmation of prior beliefs — and interventions accordingly target content through labeling, demotion, or removal. The ICC finding suggests this paradigm misidentifies the primary driver. If amplification is predominantly a function of stable user-level social capital rather than content properties, then content-level interventions address a secondary source of variance while leaving the primary one untouched.

This does not imply that content moderation is ineffective, but it does imply that its effectiveness is ceiling-limited by the degree to which the amplifying users, not the amplified messages, are the operative variable. Regulatory frameworks that focus exclusively on content — as the majority of current platform governance does — may therefore systematically underestimate the role of user network structure in shaping what circulates at scale. Approaches that complement content governance with attention to the network positions of high-amplification users, such as transparency requirements around the reach of political accounts, represent a direction that the evidence from this thesis supports.

For Urban Policymakers and Developers The spatial correspondence between graffiti-dense areas and politically engaged, high-amplification Twitter communities has implications for how cities think about politically expressive public space. Neighborhoods like Shoreditch, Brixton, and South Bank func-

tion as physical anchors for politically active communities whose online behavior extends the reach of political discourse far beyond the walls. Decisions about the preservation, removal, or permitting of political street art in these zones are therefore not purely aesthetic or public order questions. They are decisions about the character of the communities those spaces attract and sustain.

For developers and local authorities managing regeneration in graffiti-dense inner London neighborhoods, the data suggest a further implication. The spatial sorting documented in this thesis is consistent with the well-documented pattern of creative and politically active communities being displaced by the very gentrification processes their cultural presence helps initiate ("London calling", 2016). Whether the online political mobilization associated with these areas persists through gentrification, relocates with displaced communities, or dissipates, is a question this data cannot answer but that follows directly from the spatial sorting mechanism this thesis documents.

6.6.2 Reconsidering the Physical–Digital Link

The original theoretical claim, that physical political expression in the built environment shapes the amplification of online political content at the individual level, is not supported by this study's data. The conditional logit, which eliminates all stable user characteristics and identifies purely from within-user variation, establishes that the physical environment is not a meaningful predictor of whether a specific tweet is amplified. The spatial correspondence between graffiti-dense areas and political amplification is real in aggregate but is a consequence of sorting rather than exposure.

This does not render the theoretical question uninteresting. It re frames it. The more productive question is not "does the physical environment change how individuals behave online?" but "what types of users inhabit which spaces, and does that spatial sorting produce spatial structure in online discourse?" These are different questions with different empirical designs. The first requires longitudinal or quasi-experimental data that can establish prior exposure and trace subsequent behavior that the current cross-sectional design cannot answer. The second is a sociological question about the relationship between urban political geography and political community formation, why politically engaged Twitter users cluster in the same spaces where political expression is most visibly inscribed in the built environment, that the current data document but cannot explain.

The broader implication for physical–digital research is methodological, compositional effects can easily be mistaken for environmental effects in cross-sectional observational data. A robust test of environmental exposure requires individual-level longitudinal data or a natural experiment that exogenously varies the built environment independently of the characteristics of its inhabitants. The current study demonstrates the problem clearly enough to motivate such designs, even if it cannot execute them.

6.6.3 The ICC as a Theoretical Statement

The intraclass correlation of 0.629 is not merely a methodological nuisance that requires statistical correction. It is a finding about the structure of political amplification on Twitter. Nearly two-thirds of the

variance in whether a tweet is quoted is determined by who posted it, independently of what the content says or where it was posted from. The stable inter-user difference overwhelms the marginal effects of content framing, emotional valence, or spatial location.

This finding has implications for how content-level analyses of political virality should be interpreted in datasets with similar structure. Research that attributes amplification differentials to message properties is not necessarily wrong, but such effects operate against a background of stable individual differences that this dataset suggests can explain the majority of variance. Content type mattered in the descriptive results here since it largely disappeared once user identity was modeled. Whether the same pattern holds in datasets collected differently is not established by this study. What the study demonstrates is that when repeated observations from the same users are available at sufficient scale, the within-user comparison is the appropriate test of whether content properties drive amplification, and that the between-user comparison, which most cross-sectional studies are implicitly running, conflates user composition with content effects. That conflation was large enough in this dataset to reverse the sign of the key coefficient. Researchers working with geolocated or user-panel Twitter data should treat user heterogeneity as a first-order modeling concern rather than a robustness check.

6.7 Limitations

6.7.1 The Bounding Box Problem

Approximately 80.6% of London tweets in the analysis sample are geolocated via bounding box centroids, a characteristic inherited from the TBCOV source dataset and not a consequence of the study's design choices. This means that the proximity variable `nearest_graffiti_km` reflects the distance from an administrative area centroid to the nearest registered graffiti piece not the distance from a specific tweeted location to a specific graffiti piece for the majority of tweets. Multiple tweets can share identical centroid coordinates because they were tagged with the same named place. At scale, this collapses what appears to be individual-level spatial analysis into area-level ecological comparison.

The study addresses this limitation by explicitly framing H2 as an area-level claim rather than an individual-level one, reporting the area-level ecological regression as a parallel primary analysis, and providing a GPS-only robustness check that acknowledges the severe power limitation it imposes (n = 104 in the key interaction cell). Nevertheless, the bounding box problem is a structural constraint on what inferences are licensed. The inability to locate tweets at street-level resolution prevents fine-grained spatial analysis. The Forest Road Hackney event study illustrates this concretely, the spatial coarseness of TBCOV made it impossible to define a treatment group at meaningful scale. Future work in this area should prioritize data sources with GPS-level location precision.

6.7.2 Amplification as Binary

The binary `is_quote` variable captures whether a tweet was quoted but does not capture how widely it was quoted. A tweet quoted once and a tweet quoted five hundred times are treated equivalently in the primary analysis. The ZINB analysis attempted to address this by recovering actual quote counts for a re-hydrated GPS-located subsample, but failed due to near-complete separation in the political content cells ($n \approx 24$ political tweets) and the absence of a `very_near = 0` comparison group in the re-hydrated sample. The binary proxy therefore misses ceiling effects, whether political content near graffiti produces disproportionately viral tweets among the tweets that are shared at all, which remains an open empirical question.

The threshold question answered by binary logistic (does this tweet cross the zero-quote barrier?) is meaningful but incomplete. Virality in the strict sense requires distributional data that binary outcomes cannot capture. Future work with access to engagement count data at source, rather than requiring post-hoc re-hydration through a rate-limited API, would add a dimension to the amplification question that this study could not address.

6.8 Future Work

6.8.1 Longitudinal and Event-Study Designs

The most important extension of this research is a longitudinal design that tracks how individual users' amplification patterns change as they move into or out of graffiti-dense areas over time. The cross-sectional limitation established throughout this thesis, where proximity at tweet time cannot be distinguished from stable user characteristics associated with living in graffiti-proximate areas, can only be resolved with data that observes the same individuals before and after changes in their spatial relationship to political expression in the built environment. GPS-opt-in longitudinal Twitter samples, such as those constructed by tracking users who voluntarily disclose location continuously across time, would make this feasible for a sufficiently large cohort.

The Forest Road Hackney mural removal demonstrates both the promise and the current limitations of event-study approaches. The event satisfies several conditions for a quasi-experimental design, but the data created obstacle. TBCOV's spatial coarseness prevents the identification of a meaningful treatment group at street-level resolution, and the council's ransomware attack in October 2020 eliminated the administrative records that would have anchored the event with precision. A GPS-rich dataset with documented street art change events, mural installations and removals recorded with coordinates and dates, would make difference-in-differences designs viable. Crowdsourced street art registries with timestamped entries represent a natural data infrastructure for this purpose.

6.8.2 Comprehensive Graffiti Databases

The graffiti database used in this study is a curated record of professional and semi-professional registered pieces. A more complete spatial record of political street art, including tags, paste-ups, stickers, and distinct pieces, would allow proximity to be used against a genuinely comprehensive representation of political expression in the built environment. This would require field data collection at city scale, which is feasible via participatory GIS methods. A crowdsourced mapping platform with geographic and thematic tagging could produce a substantially more complete inventory than any registry-based source.

A comprehensive database would also permit the analytical separation of different kinds of street art that the current operationalization collapses together. Professional political murals by named artists, community-sanctioned wall art, and spontaneous political tags may have different relationships to the political communities that produce them and inhabit the surrounding spaces. Whether proximity to ephemeral tags produces different amplification patterns than proximity to large curated pieces is a question the current database cannot begin to address.

6.8.3 The Outrage/Anxiety Distinction as a Research Agenda

The sentiment–amplification paradox is the most tractable open question identified by this study's findings. The prediction is specific: negative emotional content of the outrage variety should predict amplification, while negative emotional content of the anxiety/fear variety should predict less amplification or no amplification advantage over neutral content. This cannot be tested with TBCOV's pre-computed three-class sentiment variable, which collapses these into a single negativity dimension.

A direct test would apply an emotion-specific classifier, for example, a RoBERTa model fine-tuned on emotional labels, or a domain-specific classifier trained on political and public health Twitter data, to a re-hydrated tweet sample. The study is feasible with existing methods. The re-hydration infrastructure built for the ZINB analysis in this thesis could support a larger GPS-located sample, and open-source emotion classification models are available. The prediction is sufficiently specific that it could be confirmed or falsified by a well-powered replication with an appropriate dependent variable. A positive result would move the outrage/anxiety interpretation from a theoretically consistent account to a demonstrated finding, with implications for both political communication theory and pandemic communication research.

7 Conclusion

This thesis asked whether proximity to street art predicts the nature and amplification of politicized content on Twitter. The short answer is that it does not, at least not in any way attributable to the built environment rather than to the people who inhabit it.

The longer answer is more interesting. The ICC of 0.629 reveals that who you are as a Twitter user explains amplification far better than what you tweet or where you tweet from. Nearly two-thirds of the variance in whether a tweet is quoted is determined by stable between-user differences that content-level

models cannot capture and spatial models cannot displace. The H1 descriptive gap of +7.4 percentage points in London reflects that prolific political blame content producers tend to be high-quoted users. The H2 association between graffiti proximity and amplification reflects that those same users cluster in politically expressive urban zones. Once user identity is controlled through mixed effects random intercepts, coarsened exact matching, and conditional logit fixed effects, both effects collapse. The content of the message or location it is tweeted from is not what drives sharing. The person is. Belgium's consistent reversal adds a boundary condition. The mechanism requires proximate domestic blame attribution, not political commentary about distant international figures.

The sentiment-amplification paradox adds a secondary finding the data can support but cannot fully resolve. Political content is the most negatively valenced category in the corpus and yet the most amplified at the aggregate level. COVID-entity content is moderately negative and amplifies least. This pattern is inconsistent with simple negativity-drives-virality accounts. It is consistent with a distinction between externalizing and internalizing emotional content. Outrage at named actors travels where anxiety about the pandemic does not. Because the pre-computed TBCOV sentiment variable cannot separate anger from fear within the negative category, this interpretation remains theoretically motivated rather than empirically demonstrated. It constitutes the primary open question for subsequent work.

The inoculation theory framework developed in Chapter 3 is not falsified by the null spatial result. The study rules out an instantaneous spatial trigger. Posting near graffiti does not alter amplification behavior in real time. It cannot test the longitudinal mechanism the theory actually describes. The stable between-user variance captured by the ICC may itself be the empirical signature of accumulated prior exposure, the durable engagement dispositions that inoculation predicts. Cross-sectional proximity data can observe the endpoint of that process but cannot recover the trajectory.

The consistent spatial sorting of politically engaged users into graffiti-proximate areas raises a question this thesis was not designed to answer but cannot avoid. Why do high-influence political Twitter users concentrate in precisely the spaces where political self-expression is most visibly inscribed in the built environment? Whether this reflects cause, consequence, or coincidence is not resolvable with cross-sectional observational data. Political communities may produce the expressive urban spaces that surround them, be drawn to those spaces, or simply co-locate through shared socioeconomic sorting. The most the data can establish is directional. The graffiti marks the community, the community is not produced by the graffiti. What the current study provides is the descriptive foundation, a precise specification of the confound that any stronger design must address, and a demonstration with direct implications for how political amplification research is conducted. Research that attributes sharing differentials to message framing or spatial location while leaving stable user identity unmodeled addresses a secondary source of variance. The spatial correspondence between graffiti geography and political amplification on Twitter is a novel empirical finding. Its source is compositional, and establishing that distinction is the contribution.

Declaration of use of GenAI Tools In line with CBS guidelines for the use of GenAI tools (CBS, 2026) the following is a declaration of AI use. Generative AI tools, namely Claude (Anthropic, 2026), and ChatGPT (OpenAI, 2026) were used for idea generation and conceptualization during the design of the thesis, specifically to brainstorm chapter structure and discuss potential methodological and analytic approaches. The tools also assisted with development of Python code underlying specific parts of the methodology described in Chapter 5. All code was reviewed, tested, and verified by the author before integrated. All analytical decisions, source attributions, interpretations, and conclusions are the author's own.

References

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The Role of Social Networks in Information Diffusion. *Proceedings of the 21st international conference on World Wide Web*, 519–528. https://doi.org/10.1145/2187836.2187907

Banas, J. A. (2020, September). Inoculation Theory. In *The International Encyclopedia of Media Psychology* (1st ed., pp. 1–8). Wiley. https://doi.org/10.1002/9781119011071.iemp0285

Barberá, P., & Rivero, G. (2015). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*, *33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Bishop, B. (2008). *The Big Sort: Why the Clustering of Like-minded America is Tearing Us Apart*. Houghton Mifflin. https://books.google.dk/books?id=NXN2AAAAMAAJ

Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health and Surveillance*, *6*(4), e21978. https://doi.org/10.2196/21978

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Castleman, C. (1999). *Getting up: Subway graffiti in New York* (8. print). MIT Press.

CBS. (2026, April). Generative artificial intelligence: Working with integrity as a CBS student [Accessed: 2026-05-15]. https://libguides.cbs.dk/c.php?g=684990&p=5136839

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, *10*(1), 16598. https://doi.org/10.1038/s41598-020-73510-5

Compton, J. (2012). Inoculation Theory. In *The SAGE Handbook of Persuasion: Developments in Theory and Practice* (pp. 220–236). SAGE Publications, Inc. https://doi.org/10.4135/9781452218410.n14

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence (E. Manalo, Ed.). *PLOS ONE*, *12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

COVID-19 Street Art Archive. (2020). Nhs street art, london. https://covid19streetart.omeka.net/items/show/191

Cutts, D., Fieldhouse, E., Purdam, K., Steel, D., & Tranmer, M. (2007). Voter Turnout in British South Asian Communities at the 2001 General Election. *The British Journal of Politics and International Relations*, *9*(3), 396–412. https://doi.org/10.1111/j.1467-856x.2006.00261.x

Dawkins, R. (1976). *The Selfish Gene* (30th anniversary ed). Oxford University Press.

Dovey, K., Wollan, S., & Woodcock, I. (2012). Placing Graffiti: Creating and Contesting Character in Inner-city Melbourne. *Journal of Urban Design*, *17*(1), 21–41. https://doi.org/10.1080/13574809.2011.646248

Enos, R. D. (2016). What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior. *American Journal of Political Science*, *60*(1), 123–142. https://doi.org/10.1111/ajps.12156

Fancourt, D., Steptoe, A., & Wright, L. (2020). The Cummings effect: Politics, trust, and behaviours during the COVID-19 pandemic. *The Lancet*, *396*(10249), 464–465. https://doi.org/10.1016/S0140-6736(20)31690-1

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferrell, J., & Stewart-Huidobro, E. (2021, October). *Crimes of Style: Urban Graffiti and the Politics of Criminality* (1st ed.). Routledge. https://doi.org/10.4324/9781003160731

Fetzer, T., Hensel, L., Hermle, J., & Roth, C. (2021). Coronavirus Perceptions and Economic Anxiety. *The Review of Economics and Statistics*, *103*(5), 968–978. https://doi.org/10.1162/rest_a_00946

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, (25/26), 56. https://doi.org/10.2307/466240

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A., Park, D. K., Ansolabehere, S., Price, P. N., & Minnite, L. C. (2001). Models, Assumptions and Model Checking in Ecological Regressions. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *164*(1), 101–118. https://doi.org/10.1111/1467-985X.00190

Godbold, L. C., & Pfau, M. (2000). Conferring Resistance to Peer Pressure Among Adolescents: Using Inoculation Theory to Discourage Alcohol Use. *Communication Research*, *27*(4), 411–437. https://doi.org/10.1177/009365000027004001

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The Structural Virality of Online Diffusion. *Management Science*, *62*(1), 180–196. https://doi.org/10.1287/mnsc.2015.2158

Graffiti Database. (n.d.). Graffiti database [Accessed: 2026-05-13]. https://graffiti-database.com/

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the World Are You? Geolocation and Language Identification in Twitter. *The Professional Geographer*, *66*(4), 568–578. https://doi.org/10.1080/00330124.2014.907699

Huckfeldt, R., Johnson, P. E., & Sprague, J. (2002). Political Environments, Political Dynamics, and the Survival of Disagreement. *The Journal of Politics*, *64*(1), 1–21. https://doi.org/10.1111/1468-2508.00115

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1–24. https://doi.org/10.1093/pan/mpr013

Imran, M., Qazi, U., & Ofli, F. (2022). TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels. *Data*, *7*(1). https://doi.org/10.3390/data7010008

Information Commissioner's Office. (2024, July). London borough of hackney reprimanded following cyber attack. https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2024%20/07/london-borough-of-hackney-reprimanded-following-cyber-attack/

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, *13*(1), 72–91. https://doi.org/10.1080/19331681.2015.1132401

King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, *27*(4), 435–454. https://doi.org/10.1017/pan.2019.11

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1948). The People's Choice. How the Voter Makes Up His Mind in a Presidential Campaign. In *The People's Choice* (pp. 3–11). https://doi.org/10.1007/978-3-658-50422-9_1

LDN Graffiti. (n.d.). *Inkie 1980 baby*. Retrieved May 14, 2026, from https://ldngraffiti.co.uk/streetart/streetartists/ink1980baby?pic=166081&piece=166133&filters=%7B%22authors%22:%5B162443%5D%7D

Legewie, J., & Schaeffer, M. (2016). Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict. *American Journal of Sociology*, *122*(1), 125–161. https://doi.org/10.1086/686942

Lerner, A. M. (2019). The co-optation of dissent in hybrid states: Post-soviet graffiti in moscow. *Comparative Political Studies*, *54*, 1757–1785. https://api.semanticscholar.org/CorpusID:211406245

London calling: Contemporary graffiti and street art in the UK's capital. (2016, March). In J. I. Ross (Ed.), *Routledge Handbook of Graffiti and Street Art* (0th ed., pp. 312–327). Routledge. https://doi.org/10.4324/9781315761664-35

Makse, T., & Sokhey, A. E. (2014). The Displaying of Yard Signs as a Form of Political Participation. *Political Behavior*, *36*(1), 189–213. https://doi.org/10.1007/s11109-013-9224-6

Mcauliffe, C. (2012). Graffiti or Street Art? Negotiating the Moral Geographies of the Creative City. *Journal of Urban Affairs*, *34*(2), 189–206. https://doi.org/10.1111/j.1467-9906.2012.00610.x

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology*, *62*(2), 327–337. https://doi.org/10.1037/h0042026

McGuire, W. J. (1964). Some Contemporary Approaches. In *Advances in Experimental Social Psychology* (pp. 191–229, Vol. 1). Elsevier. https://doi.org/10.1016/S0065-2601(08)60052-0

Mubi Brighenti, A. (2010). At the Wall: Graffiti Writers, Urban Territoriality, and the Public Domain. *Space and Culture*, *13*(3), 315–332. https://doi.org/10.1177/1206331210365283

Nielsen, J. H., & Lindvall, J. (2021). Trust in government in Sweden and Denmark during the COVID-19 epidemic. *West European Politics*, *44*(5-6), 1180–1204. https://doi.org/10.1080/01402382.2021.1909964

Noelle-Neumann, E. (1974). The Spiral of Silence a Theory of Public Opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). *The city: Suggestions for investigation of human behavior in the urban environment* (Reprint.). Univ. of Chicago Press.

Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's Efficacy With Young Adults' Risky Behaviors: Can Inoculation Confer Cross-Protection Over Related but Untreated Issues? *Health Communication*, *27*(3), 223–233. https://doi.org/10.1080/10410236.2011.575541

Pew Research Center. (2019, April). *Sizing Up Twitter Users* (tech. rep.). Pew Research Center.

Pfau, M., & Bockern, S. V. (1994). The Persistence of Inoculation in Conferring Resistance to Smoking Initiation Among Adolescents: The Second Year. *Human Communication Research*, *20*(3), 413–430. https://doi.org/10.1111/j.1468-2958.1994.tb00329.x

Pfau, M., Bockern, S. V., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communication Monographs*, *59*(3), 213–230. https://doi.org/10.1080/03637759209376266

Pfau, M., Park, D., Holbert, R. L., & Cho, J. (2001). The Effects of Party- and PAC-Sponsored Issue Advertising and the Potential of Inoculation to Combat its Impact on the Democratic Process. *American Behavioral Scientist*, *44*(12), 2379–2397. https://doi.org/10.1177/00027640121958384

Riggle, N. A. (2010). Street art: The transfiguration of the commonplaces. *The Journal of Aesthetics and Art Criticism*, *68*, 243–257. https://api.semanticscholar.org/CorpusID:191191515

Roozenbeek, J., & Van Der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 65. https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254

Ross, J. I., & Ferrell, J. (Eds.). (2019). *Routledge handbook of graffiti and street art* (First issued in paperback). Routledge, Taylor & Francis Group.

Ryan, H. E. (2021). The political work of graffiti during the Covid-19 pandemic: A view from Tottenham, London. *Visual Studies*, *36*(2), 133–140. https://doi.org/10.1080/1472586X.2021.1911677

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (Eigth edition). Pearson.

Thelwall, M., & Thelwall, S. (2020). A thematic analysis of highly retweeted early COVID -19 tweets: Consensus, information, dissent, and lockdown life [Version Number: 3]. https://doi.org/10.48550/ARXIV.2004.02793

Van Der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Young, A. (2013, November). *Street Art, Public City: Law, Crime and Urban Imagination* (0th ed.). Routledge. https://doi.org/10.4324/9780203796917

Zade, H., Williams, S., Tran, T. T., Smith, C., Venkatagiri, S., Hsieh, G., & Starbird, K. (2024). To Reply or to Quote: Comparing Conversational Framing Strategies on Twitter. *ACM Journal on Computing and Sustainable Societies*, *2*(1), 1–27. https://doi.org/10.1145/3625680

Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J., & Cui, H. (2012). SIHR rumor spreading model in social networks. *Physica A: Statistical Mechanics and its Applications*, *391*(7), 2444–2453. https://doi.org/10.1016/j.physa.2011.12.008

Zohar, M. (2021). Geolocating tweets via spatial inspection of information inferred from tweet meta-fields. *International Journal of Applied Earth Observation and Geoinformation*, *105*, 102593. https://doi.org/https://doi.org/10.1016/j.jag.2021.102593

Appendix

A TBCOV Dataset Reference

A.1 Field Descriptions

Table 13: TBCOV: Data Descriptor

Attribute	Type	Description
tweet_id	Int64	The integer representation of the unique identifier of a tweet. This number is greater than 53 bits and some programming languages may have difficulty/silent defects in interpreting it.
date_time	String	UTC time when the tweet was created.
lang	String	ISO-6391 Alpha-2 language code consisting of two characters.
user_id	String	Represents the id of the author of the tweet.
retweeted_id	Int64	If the tweet is a retweet, the retweeted_id represents the id of the parent tweet.
quoted_id	Int64	If the tweet is a quoted tweet, the quoted_id represents the id of the parent tweet.
in_reply_to_id	Int64	If the tweet is a reply to an existing tweet, the in_reply_to_id represents the id of the parent/original tweet.
sentiment_label	Int64	Represents the sentiment label values: -1 (negative), 0 (neutral), 1 (positive).
sentiment_conf	Float	Represents the confidence score of the sentiment classifier for a given sentiment label to a tweet.
user_type	String	Represents the inferred type of the user account. Personal accounts are coded as "PER" and accounts belonging to organizations are coded as "ORG".

Table 13 – continued

Attribute	Type	Description
gender_label	String	One character code representing the identified gender of the user where "F" represents female and "M" represents male user types.
tweet_text_named_entities	Dictionary array	Named-entities (person, organization, location, etc.) extracted from tweet text, provided as an array of dictionaries.
geo_coordinates_lat_lon	Float	GPS coordinates in the latitude, longitude format retrieved from the user's GPS-enabled device.
geo_country_code	String	Two character country code obtained through resolving the GPS coordinates (latitude, longitude).
geo_state	String	The name of the state/province obtained through resolving the GPS coordinates (latitude, longitude).
geo_county	String	The name of the county obtained through resolving the GPS coordinates (latitude, longitude).
geo_city	String	The name of the city obtained through resolving the GPS coordinates (latitude, longitude).
place_bounding_box	Float	Twitter provided bounding boxes representing place tags.
place_country_code	String	Two character country code obtained through resolving the place bounding boxes.
place_state	String	The name of the state/province obtained through resolving the place bounding boxes.
place_county	String	The name of the county obtained through resolving the place bounding boxes.
place_city	String	The name of the city obtained through resolving the place bounding boxes.

Table 13 – continued

Attribute	Type	Description
user_loc_toponyms	Dictionary array	Toponyms recognized and extracted from the user location field, provided as an array of dictionaries.
user_loc_country_code	String	Two character country code obtained through resolving the user location toponyms.
user_loc_state	String	The name of the state/province obtained through resolving the user location toponyms.
user_loc_county	String	The name of the county obtained through resolving the user location toponyms.
user_loc_city	String	The name of the city obtained through resolving the user location toponyms.
user_profile_description_toponyms	Dictionary array	Toponyms recognized and extracted from the user profile description field, provided as an array of dictionaries.
user_profile_description_country_code	String	Two character country code learned through resolving the recognized user profile description toponyms.
user_profile_description_state	String	The name of the state/province obtained through resolving the recognized user profile description toponyms.
user_profile_description_county	String	The name of the county obtained through resolving the recognized user profile description toponyms.
user_profile_description_city	String	The name of the city learned through resolving the recognized user profile description toponyms.
tweet_text_toponyms	Dictionary array	Toponyms recognized and extracted from the tweet full_text field, provided as an array of dictionaries.
tweet_text_country_code	String	Two character country code obtained through resolving the recognized tweet text toponyms.

Table 13 – continued

Attribute	Type	Description
tweet_text_state	String	The name of the state/province obtained through resolving the recognized tweet text toponyms.
tweet_text_county	String	The name of the county learned through resolving the recognized tweet text toponyms.
tweet_text_city	String	The name of the city learned through resolving the recognized tweet text toponyms.

A.2 Named Entity Types

Table 14: TBCOV: Named Entity Types

Entity Type	Description
PERSON	Name of a person. E.g., Peter Pan, Steve Jobs
ORG	Companies, agencies, institutes names, e.g., MIT, Microsoft, QCRI
GPE	Name of countries, cities, states, etc.
LOC	Non-GPE locations such as mountain ranges, water bodies, etc.
FAC	Represents buildings, airports, highways, etc.
PRODUCT	Objects, vehicles, foods, etc.
NORP	Nationalities or religious or political groups
LANGUAGE	Any named language, e.g., English, Arabic
DATE	Dates or periods, e.g., July 12, 2003
TIME	Times smaller than a day, e.g., five hours, 2 hours
QUANTITY	Measurements, as of weight or distance, e.g., 40 kg, several kilometers
CARDINAL	Numerals such as 8, five, ten
ORDINAL	First, second, third, etc.
PERCENT	Percentages, including % sign
EVENT	Named event names, e.g., hurricanes, battles, wars
MONEY	Monetary values and unit, e.g., ten cents
LAW	Named documents made into laws
WORK_OF_ART	Titles of books, songs, etc.
COVID-ENTITY	COVID-19 related terms such as corona, covid_19, coronavirus

The has_pol_figure flag (central to the H1/H2 operationalization) is not a direct NER output. It is derived by matching the term values in named_entities against the political figure term list defined in Chapter 4. Any entity label is eligible for matching, including NORP, ORG, and PERSON.

A.3 Political Figure Term List

Table 15: Political figure term list used to derive `has_pol_figure`. Matching is performed on the `term` value of each NER entity across all label types. A tick indicates the figure's terms were included in that country's term list.

Figure	`Matched terms`	**Lon.**	**BE**	**DK**
Boris Johnson	`boris, boris johnson, bojo, johnson, johnsons`	✓	✓	✓
Dominic Cummings	`cummings, dominic cummings`	✓	✓	✓
Matt Hancock	`matt hancock, hancock`	✓	✓	✓
Dominic Raab	`raab`	✓	✓	✓
Rishi Sunak	`rishi sunak, sunak`	✓	✓	✓
Priti Patel	`patel, priti patel`	✓	✓	✓
Keir Starmer	`keir starmer, starmer`	✓	✓	✓
Sadiq Khan	`sadiq khan`	✓	✓	✓
Conservative Party	`tory, tories`	✓	✓	✓
Donald Trump	`trump, donald trump, donaldtrump`	✓	✓	✓
Joe Biden	`biden, joe biden, joebiden`	✓	✓	✓
Anthony Fauci	`fauci`	✓	✓	✓
Vladimir Putin	`putin`	✓	✓	✓
Bart De Wever	`de wever`		✓	
Frank Vandenbroucke	`frank vandenbroucke, vandenbroucke`		✓	
Sophie Wilmès	`sophie wilmes, sophie wilmès, wilmès, wilmes`		✓	
Charles Michel	`charles michel, michel`		✓	
Emmanuel Macron	`macron, emmanuel macron`		✓	
Angela Merkel	`merkel, angela merkel`		✓	
Jair Bolsonaro	`bolsonaro, jair bolsonaro`		✓	
Viktor Orbán	`viktor orban, orban`		✓	
Mette Frederiksen	`mette frederiksen, frederiksen, mette`			✓

B Tweet Text Samples

The following examples are drawn from the rehydrated GPS-located London subset of the TBCOV dataset. All examples are real tweets from 2020–2021. Tweet text has been reproduced as-is.

B.1 Politicized tweets

Tweets where (`has_pol_figure = 1`)

Table 16: Example text from politicized tweets.

Date	Text (excerpt)	is_quote	Sentiment
2020-07-08	"Boris lying again. Surprise surprise. #BorisHasFailedTheNation"	1	−1
2020-04-21	"And so Britain's 'herd immunity' policy continues? Time for Donald Trump to identify the unnamed country he claims has responded catastrophically to coronavirus?"	1	−1
2020-07-28	"Are the workers party still siding with the Tories to keep Scotland shackled to Westminster?"	1	−1
2020-07-17	"Boris, you're a nice guy but this is totally irresponsible"	1	−1
2020-08-18	"The problem with this argument is history and reality. Labour always whinge that the Tories will sell off the NHS. History tells us the vast majority of private investment into the NHS came under the last Labour government"	1	−1
2020-03-06	"Coronavirus: US President Donald Trump cancels visit to CDC at last minute"	0	−1
2020-04-21	"@JoeBiden Can't wait till Trump actually leaves and a normal person is in the White house. You will be to educate the whole of US on Covid and make vaccines for all not just those who can afford them."	0	−1
2020-07-28	"Boris Johnson tells us that he's not surprised that 20% of calls to the London Ambulance Service (a few hundred a day) are Coronavirus related"	0	−1
2020-04-06	"You and All Tory MPs should set an example and spend a week doing a trial run. @BorisJohnson, NO EXCEPTIONS. We're all in it together, aren't we? Of course, Cummings can lead the way @10DowningStreet"	0	+1
2021-01-14	"So BoJo wants us to lose weight, commendable but how about you first deal with that other problem called the global pandemic that is covid-19 which is still killing people daily"	0	−1

B.2 Non-politicized COVID tweets

Tweets where (`has_COVID = 1`, `has_pol_figure = 0`)

Table 17: Non-politicized COVID tweets

Date	Text (excerpt)	is_quote	Sent.
2021-01-02	"FUCK COVID."	1	−1
2020-03-04	"Wondering — and worried — we may see a domino effect from hereonin over the next few weeks in events over fears of #Covid_19."	1	0
2020-03-04	"The results of my survey are in. Proof that facts and figures on Twitter can be relied on one hundred percent. #Covid_19 #CoronaOutbreak"	1	+1
2020-05-12	"Advice For People With Diabetes During Covid-19 Pandemic"	0	0
2020-05-25	"IFC releases guidance for companies on preventing reprisals against project opponents during COVID-19, encouraging zero tolerance approach"	0	0
2020-05-27	"How Can Fitness Professionals Thrive During Covid-19?"	0	0
2020-06-17	"Which Countries Are Putting The Most Effort in COVID-19 Research"	0	0

B.3 NHS Solidarity tweets

Table 18: NHS solidarity tweets.

Date	Text (excerpt)	is_quote	Sent.
2020-04-02	"THANK YOU NHS workers. Go blue to show your support. #coronavirus #clap #clapforourcarers #nhs #nhsappreciation"	0	+1
2020-04-02	"To the #amazing #nhs risking their lives for us! #stayhome #staysafe #protectthenhs #savelives"	0	+1
2020-04-02	"Help the NHS to help you!!! Adhere to government guidelines. #besafe #keepsafe #nhsworkersareheroes #nhsworker #socialdistancing"	0	+1
2020-03-30	"Hearing of the Glory of the Lord of the Universe, the mind becomes fearless. #covid #Coronavirus #weareallinthistogether #StayAtHome #ProtectTheNHS #SaveLives"	0	+1
2020-04-02	"#nhs #covid19 #oldstreet #streetphotography @ Old Street Roundabout"	0	0

C Clustering Analysis

C.1 Cluster Overview

Table 19: KMeans cluster profiles ($k = 12$). Mean km = mean distance to nearest graffiti piece; quote rate = proportion of tweets that are quote tweets. Clusters 6 and 10 are the analytically central clusters for H2.

Cluster	Dominant entity / theme	Mean km	Quote rate	Interpretation
0	NHS, Brexit, EU	1.09	0%	Organic NHS/political expression near graffiti; original posts
1	NHS (amplified)	2.08	100%	Amplified NHS content, farther from graffiti
2	—	—	∼0%	Low-amplification general content
3	General / varied	2.42	0%	Organic general posts, far from graffiti
4	—	—	∼100%	High-amplification general content
5	—	—	∼0%	Low-amplification content
6	Boris Johnson, Cummings	0.74	∼100%	Political figure discourse near graffiti — amplified
7	#ThankyouNHS hashtag	medium	∼0%	Organic NHS solidarity movement
8	—	—	∼0%	Low-amplification content
9	Mixed / temporal terms	0.86	99%	High-amplification near-graffiti cluster
10	Boris Johnson, Johnson	0.74	∼100%	Political figure discourse near graffiti — amplified
11	—	—	∼0%	Low-amplification general content

C.2 Illustrative tweets for key clusters

The following examples are matched by entity composition and proximity to illustrate the characteristic discourse in each cluster type.

Clusters 6 and 10 — political figure, near graffiti (< 1 km), quoted:

"Boris lying again. Surprise surprise. #BorisHasFailedTheNation"
2020-07-08 · `is_quote=1` · sentiment = −1 · km = 0.42

"Boris, you're a nice guy but this is totally irresponsible"
2020-04-19 · `is_quote=1` · sentiment = −1 · km = 0.42

"And so Britain's 'herd immunity' policy continues? Time for Donald Trump to identify the unnamed country he claims has responded catastrophically to coronavirus?"
2020-04-21 · `is_quote=1` · sentiment = −1 · km = 0.42

"Doesn't matter who he presents himself as because he is still Boris Johnson, the worst prime minister the U.K. has ever had."
2020-08-04 · `is_quote=1` · sentiment = −1 · km = 0.42

Cluster 0 — NHS/organic, near graffiti, not quoted:

"THANK YOU NHS workers. Go blue to show your support. #coronavirus #clap #clapforourcarers #nhs"
2020-04-02 · `is_quote=0` · sentiment = +1 · km = 0.02

"To the #amazing #nhs risking their lives for us! #stayhome #staysafe #protectthenhs #savelives"
2020-04-02 · `is_quote=0` · sentiment = +1 · km = 0.02

"Help the NHS to help you!!! Adhere to government guidelines. A great time of turbulence for the whole world."
2020-04-02 · `is_quote=0` · sentiment = +1 · km = 0.02

"#nhs #covid19 #oldstreet #streetphotography @ Old Street Roundabout"
2020-04-02 · `is_quote=0` · sentiment = 0 · km = 0.02

Cluster 1 — NHS, far from graffiti, quoted:

"Can't help but view this in a positive light. Surgeons working alongside us in #ICU not only have an excellent grasp of #CriticalCare as a core component of their training but the message on regional collaboration must not be lost."
2021-01-01 · `is_quote=1` · sentiment = +1 · km = 0.42

Non-politicised COVID type (low-amplification):

"Advice For People With Diabetes During Covid-19 Pandemic"
2020-06-08 · `is_quote=0` · sentiment = 0 · km = 0.00

"IFC releases guidance for companies on preventing reprisals against project opponents during COVID-19, encouraging zero tolerance approach"
2020-06-17 · `is_quote=0` · sentiment = 0 · km = 0.00

"Wondering — and worried — we may see a domino effect from hereonin over the next few weeks in events over fears of #Covid_19."
2020-03-04 · `is_quote=0` · sentiment = 0 · km = 0.42

D Most prominent graffiti images for each location (Graffiti Database, n.d.)

D.1 London pieces

(a) MByte, Writer Unknown, 2013

(b) Worth More, Ben Eine, 2013

(c) Still Life, Dan Kitchener, 2012

(d) SKIRE, Skire, 1997

(e) Fuel CCD, Fuel, 1991

(f) Frunch, Jim Vision, 2015

Figure 19: Examples of the most prominent graffiti pieces (closest proximity to tweets) identified in London from (Graffiti Database, n.d.)

D.2 Belgium pieces

(a) SEYB, Seyb, 2017

(b) Sport93, Writer Unknown, No date

(c) Character, Smug One, No date

(d) Characters, Roa, No date

Figure 20: Examples of the most prominent graffiti pieces (closest proximity to tweets) identified in Belgium from (Graffiti Database, n.d.)

D.3 Copenhagen pieces

(a) Bates, Bates, 2019

(b) Iris, Iris, 1995

(c) MOAS, Writer Unknown, 2019

(d) Spook, Writer Unknown, 2008

(e) Various pieces, Faze, Elo32, Frak, Geks, 2011

Figure 21: Examples of the most prominent graffiti pieces (closest proximity to tweets) identified in Copenhagen from (Graffiti Database, n.d.)

E Results

Finding	Result	Model	Verdict
H1: London descriptive gap	+7.4pp	—	Descriptive support
H1: London logistic (full)	OR=0.972, ns	Clustered SEs	Not supported
H1: London logistic (excl. heavy)	OR=1.202, p=0.011	Clustered SEs	Weak support
H1: Within-user	OR=0.785, $p<.001$	Mixed effects	**Reversed**
H1: CEM (clustered SE)	OR=1.277, p=0.001	CEM	Supported
H1: CEM (mixed effects)	OR=1.040, p=0.537	CEM	Not supported
H2: Individual-level	OR=1.436, p=0.085	Clustered SEs	Not supported
H2: Area-level	β=+0.036, p=0.014	Ecological WLS	**Dir. reversed**
H2: Within-user	OR=1.226, ns	Mixed effects	Not supported
H2: CEM (pol_x_logkm, clustered)	OR=1.045, ns	CEM	Not supported
H2: User-CEM	OR=1.379, ns	User-CEM	Not supported
H2: Conditional logit	OR=1.068, ns	User FE	Not supported
Distance decay	Non-monotonic	Chi-square	Inconsistent
Spatial autocorrelation	I=0.005, ns	Moran's I	Not clustered
Bivariate Moran's I	I=−0.099, p=0.010	Bivariate Moran's I	Neg. assoc.
Sentiment mediates H1	Sobel p=0.627	Mediation	Not supported
Emotional charge moderates	β=−0.249, p=0.011	Interaction	Supported
Proximity → less neg. sentiment (pol)	β=0.088, p=0.014	OLS	Exploratory
Proximity → less neg. sentiment (within-user)	β=0.034, p=0.002	Mixed effects	Exploratory

Table 20: Summary of all H1 and H2 findings across model specifications.